

UNIVERSIDAD DE GRANADA



Departamento de Ciencias de la Computación
e Inteligencia Artificial

Programa de Doctorado en Tecnologías de la Información y la Comunicación

The Design, Implementation and Application of Z-number Linguistic Model

Tesis Doctoral

Yangxue Li

Granada, Junio de 2025

UNIVERSIDAD DE GRANADA



The Design, Implementation and Application of Z-number Linguistic Model

MEMORIA PRESENTADA POR

Yangxue Li

Para optar al grado de doctor por la Universidad de Granada dentro del
Programa de Doctorado Tecnologías de la Información y la Comunicación

Junio de 2025

DIRECTORES

Juan Antonio Morente-Molinera

Departamento de Ciencias de la Computación
e Inteligencia Artificial

Editor: Universidad de Granada. Tesis Doctorales
Autor: Yangxue Li
ISBN: 978-84-1195-892-9
URI: <https://hdl.handle.net/10481/108295>

La memoria titulada “*The Design, Implementation and Application of Z-number Linguistic Model*”, que presenta D. Yangxue Li para optar al grado de doctor, ha sido realizada dentro del Programa Oficial de Doctorado en “*Tecnologías de la Información y la Comunicación*”, en el Departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada bajo la dirección de los doctores D. Juan Antonio Morente-Molinera.

El doctorando, D. Yangxue Li, y los directores de la tesis, D. Juan Antonio Morente-Molinera, garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección de los directores de la tesis, y hasta donde nuestro conocimiento alcanza, en la realización del trabajo se han respetado los derechos de otros autores a ser citados cuando se han utilizado sus resultados o publicaciones.

Granada, Junio de 2025

El Doctorando Los directores

Fdo: Yangxue Li Fdo: Juan Antonio Morente-Molinera

Acknowledgements

With these final words, my journey comes to an end. Four years of studying abroad are drawing to a close—an experience that has been one of the most colorful and enriching chapters of my life. I am deeply grateful for every person I have met and every path I have walked along the way.

First and foremost, I would like to express my heartfelt gratitude to my supervisor, Professor Juan Antonio Morente-Molinera and Professor Enrique Herrera-Viedma, for granting me the invaluable opportunity to study at the University of Granada. Their patience, enthusiasm, dedication, and outstanding academic expertise have greatly benefited me. Throughout my studies and research, they have provided me with profound insights and constructive suggestions, significantly enhancing both my research capabilities and academic skills. I feel truly honored and privileged to be their student.

During my time in Granada, I was fortunate to meet many wonderful friends, including Xingli Wu, Ningyue Zhang, Yinfeng Du, Tianrui Sun, Shuli Yan, Han Wang, and Feixia Ji. We shared countless meals and embarked on many unknown journeys together. I am deeply grateful for the selfless support you have given me in both my studies and daily life. Additionally, I would like to express my sincere gratitude to the China Scholarship Council (CSC) for providing financial support, which allowed me to experience this wonderful academic and life journey at the University of Granada.

I would also like to express my gratitude to my master's advisor, Professor Yong Deng. Even after I graduated, he continued to provide me with patient guidance whenever I encountered difficulties in my research. His academic and life advice has been invaluable and will benefit me for a lifetime. Furthermore, I am grateful to my lab colleagues, Xiaozhuan Gao, Lipeng Pan, Qianli Zhou, Jie Zhao, and Fan Liu. We often engaged in academic discussions together, and I truly appreciate your unwavering support and assistance!

Finally, I would like to express my heartfelt gratitude to my parents for their unconditional support and love. Thank you for being my strongest and warmest refuge, allowing me to move forward fearlessly!

Abstract

Decision-making in the real world is often complicated by the uncertainty of information and its partial reliability. Not only might the exact value of a quantity be vague, but our confidence in that value can also be limited. Traditional approaches, such as probability theory and fuzzy set theory, each have their own limitations when handling the partial reliability of information. To overcome these limitations, Zadeh introduced Z-numbers, defined as ordered pairs $Z = (\tilde{A}, \tilde{B})$, integrating a fuzzy quantity with its associated reliability measure.

This thesis builds on the concept of Z-numbers to develop a comprehensive Z-number linguistic modeling framework, explicitly addressing the uncertainty in both value and reliability. The main contributions include:

(1) Z-number linguistic model: A formal model is established for representing linguistic information (such as expert statements or sensor observations) as Z-number linguistic terms. This structured representation simultaneously encodes fuzzy values and their reliability, laying theoretical groundwork for robust reasoning under uncertainty.

(2) Z-number-valued rule-based classification system (ZRBCS): Extending traditional fuzzy rule-based classification systems, ZRBCS incorporates Z-number conditions into rules. Unlike conventional fuzzy systems, each rule condition explicitly encodes uncertainty and reliability, enabling nuanced inference and improving classification performance by utilizing negative samples to determine reliability and adjust fuzzy partitions.

(3) Z-number-valued rule-based decision tree (ZRDT): Building on fuzzy rule-based decision trees, ZRDT integrates Z-numbers into the decision-tree framework. Information gain replaces fuzzy confidence for feature selection, and negative samples guide the refinement of fuzzy numbers for better data fitting. Experimental comparisons demonstrate that ZRDT achieves higher classification accuracy and produces a more compact decision tree than standard algorithms such as FRDT, PUBLIC, C4.5, and AdaBoost.NC.

(4) Z-number generation model: A novel nonlinear model, Maximum Expected Minimum Entropy (MEME), is proposed to generate Z-numbers directly from multiple probability distributions, eliminating reliance on expert input. Further, MEME is integrated into a Z-valuation rule-based (ZVRB) classification system, significantly enhancing decision-making performance under uncertainty. Experimental validation confirms that the ZVRB system outperforms classical classifiers and existing rule-based methods.

(5) Optimization via FURIA: The Fuzzy Unordered Rule Induction Algorithm (FURIA) is integrated as an optimization tool to refine the proposed Z-number models. FURIA improves accuracy and interpretability by dynamically adjusting membership functions, optimizing rule sets, and selecting predictive features. This demonstrates that the developed framework can adaptively learn from data, significantly enhancing overall performance.

Collectively, these contributions advance uncertainty modeling by explicitly managing both value uncertainty and information reliability, providing a robust and nuanced toolset for real-world decision-making beyond existing probabilistic and fuzzy frameworks.

Resumen

La toma de decisiones en el mundo real suele complicarse por la incertidumbre de la información y su confiabilidad parcial. No solo el valor exacto de una cantidad puede ser impreciso, sino que también nuestra confianza en dicho valor puede ser limitada. Los enfoques tradicionales, tales como la teoría de la probabilidad y la teoría de conjuntos difusos, presentan limitaciones propias al manejar la confiabilidad parcial de la información. Para superar estas limitaciones, Zadeh introdujo el concepto de Z-números, definidos como pares ordenados $Z = (\tilde{A}, \tilde{B})$, integrando una cantidad difusa con su medida de confiabilidad asociada.

Esta tesis se basa en el concepto de Z-números para desarrollar un marco exhaustivo de modelado lingüístico con Z-números, abordando explícitamente la incertidumbre tanto en el valor como en la confiabilidad. Las contribuciones principales incluyen:

(1) Modelo lingüístico basado en Z-números: Se establece un modelo formal para representar información lingüística (como opiniones de expertos u observaciones de sensores) mediante términos lingüísticos en forma de Z-números. Esta representación estructurada codifica simultáneamente valores difusos y su confiabilidad, estableciendo las bases teóricas para un razonamiento robusto bajo incertidumbre.

(2) Sistema de clasificación basado en reglas con Z-números (ZRBCS): Extendiendo los sistemas tradicionales de clasificación basados en reglas difusas, el ZRBCS incorpora condiciones basadas en Z-números dentro de las reglas. A diferencia de los sistemas difusos convencionales, cada condición de regla codifica explícitamente incertidumbre y confiabilidad, permitiendo una inferencia más matizada y mejorando el desempeño en clasificación mediante el uso de muestras negativas para determinar la confiabilidad y ajustar las particiones difusas.

(3) Árbol de decisión basado en reglas con Z-números (ZRDT): Construido sobre los árboles de decisión basados en reglas difusas, el ZRDT integra los Z-números en el marco del árbol de decisión. La ganancia de información reemplaza la confianza difusa para la selección de características, y las muestras negativas guían el refinamiento de los números difusos para una mejor adaptación a los datos. Las comparaciones experimentales demuestran que el ZRDT logra una precisión de clasificación superior y produce un árbol de decisión más compacto que algoritmos estándar como FRDT, PUBLIC, C4.5 y AdaBoost.NC.

(4) Modelo de generación de Z-números: Se propone un nuevo modelo no lineal denominado Máxima Entropía Mínima Esperada (MEME), para generar Z-números directamente a partir de múltiples distribuciones de probabilidad, eliminando la necesidad de opiniones expertas. Además, MEME se integra en un sistema de clasificación basado en reglas con Z-valoración (ZVRB), mejorando significativamente el desempeño en la toma de decisiones bajo incertidumbre. Las validaciones experimentales confirman que el sistema ZVRB supera a los clasificadores clásicos y a los métodos existentes basados en reglas.

(5) Optimización mediante FURIA: Se incorpora el algoritmo de inducción de reglas difusas

no ordenadas (FURIA) como herramienta de optimización para refinar los modelos propuestos basados en Z-números. FURIA mejora la precisión y la interpretabilidad mediante ajustes dinámicos de funciones de pertenencia, optimización del conjunto de reglas y selección de características predictivas. Esto demuestra que el marco desarrollado puede adaptarse y aprender dinámicamente de los datos, mejorando considerablemente su desempeño general.

En conjunto, estas contribuciones avanzan en el modelado de incertidumbre mediante la gestión explícita tanto de la incertidumbre en el valor como en la confiabilidad de la información, proporcionando herramientas robustas y matizadas para la toma de decisiones en contextos reales, superando así los marcos probabilísticos y difusos existentes.

Table of Contents

	Page
I PhD dissertation	1
1 Introduction	1
Introducción	5
2 Preliminaries	10
2.1 Fuzzy linguistic approach	10
2.2 Z-numbers	12
2.3 Fuzzy rule-based systems	14
3 Justification	16
3.1 Z-number linguistic modeling	16
3.2 Z-number-valued rule-based system	16
3.3 Advanced optimization	17
3.3.1 Z-number-valued rule-based decision tree	18
3.3.2 Optimization via FURIA	18
3.3.3 Z-number generation model	19
4 Objectives	20
5 Methodology	21
6 Summary	23
6.1 Z-number linguistic model	23
6.1.1 Z-number linguistic term set	23
6.1.2 Theoretical support	25
6.1.3 Operators	26
6.1.4 Multi-attribute group decision making procedures	27
6.2 Z-number-valued rule-based classification system	29
6.2.1 Z-number-valued if-then rule structure	29
6.2.2 Z-number-valued if-then rules generation	29
6.2.3 Procedure for classifying new patterns	31
6.3 Z-number-valued rule based decision tree	31

6.3.1	Structure of the Z-number-valued if-then rule	31
6.3.2	Z-number-valued rule extraction algorithm	32
6.3.3	New samples classification	34
6.4	Z-number generation model and its application in a rule-based classification system	35
6.4.1	Maximum expected minimum entropy model	35
6.4.2	Z-valuation rule-based classification system	37
6.5	Z-mixture-number if-then rule induction algorithm for classification tasks . .	39
6.5.1	Definitions	39
6.5.2	Rule learning	40
6.5.3	Inference engine	44
7	Discussion of results	45
7.1	Z-number linguistic model	45
7.2	Z-number-valued rule-based classification system	45
7.3	Z-number-valued rule-based decision tree	46
7.4	Z-number generation model and its application in a rule-based classification system	47
7.5	Z-mixture-number if-then rule induction algorithm for classification tasks . .	47
8	Concluding remarks	49
	Conclusiones	49
9	Future Works	51
II Publications: Published Papers		53
1	Z-number linguistic term set for multi-criteria group decision-making and its application in predicting the acceptance of academic papers	54
2	Z-number-valued rule-based classification system	87
3	Z-number-valued rule-based decision trees	109
4	Z-number generation model and its application in a rule-based classification system .	128
5	Z-mixture-number if-then rules induction algorithm for classification tasks	142
Bibliografía		155

Chapter I

PhD dissertation

1 Introduction

Decision-making in the real world is often complicated by the uncertainty of information and its partial reliability [AKKJG23]. In many fields—ranging from engineering and finance to healthcare and daily planning—we must make judgments based on data or expert opinions that are not entirely reliable or certain. Not only might the exact value of a quantity be vague, but our confidence in that value can also be limited. For example, consider a scheduling estimate: “the project will likely finish in about three months.” This statement embodies two layers of uncertainty: “about three months” is an imprecise (fuzzy) duration, while the qualifier “likely” indicates that the source is not entirely sure about this estimate. Handling this dual uncertainty (what the value is and how reliability it is) is a significant challenge. If a decision-maker ignores the partial reliability of information—for example, treating an optimistic guess with the same level of confidence as a well-validated fact—the resulting decisions may be suboptimal or even risky. There is a clear motivation to develop methodologies that can explicitly manage both uncertainty in values and uncertainty in reliability, providing more nuanced and robust support for real-world problem solving.

Dealing with uncertainty is not a new problem. Traditional approaches, such as probability theory and fuzzy set theory, have been the primary tools for modeling uncertainty. However, each has its own limitations when it comes to handling the partial reliability of information [Fel08, Zad96]. Probability theory quantifies uncertainty by assigning probabilities to events. While highly effective in domains with well-defined random processes, it assumes precise probability values and often requires abundant data or strong assumptions to derive those values. It struggles to directly incorporate vague linguistic information - expressions such as “likely” or “approximately” do not translate neatly into exact probability numbers. Moreover, probability theory does not inherently account for the reliability of the source; a probability is usually treated as an objective quantity, and representing the confidence in that probability (second-order uncertainty) requires complex extensions. In practice, this means purely probabilistic models may fall short when information is scarce, subjective, or provided in natural language.

Fuzzy set theory, introduced by Zadeh, takes a different approach by allowing elements to belong to a set with graded membership between 0 and 1 [Zad96]. This framework excels at capturing vagueness and linguistic ambiguity. For example, instead of a sharp cutoff for “fast” vs “slow,” fuzzy sets allow an element (say a speed of 50 km/h) to be partly in the set “fast” and partly not, reflecting the continuum of the concept. Fuzzy logic enables us to work with terms like “about three months” by defining a membership function that assigns high truth values

to durations near three months and lower values further away. The most common method of linguistic information modeling is the fuzzy linguistic approach [Zad75c, Zad75a, Zad75b]. It deals with qualitative aspects of linguistic information represented in qualitative terms by linguistic variables. It has been successfully applied in various fields, such as supplier selection [JC21], route optimization [DRC⁺23], logistics outsourcing problems [WWC⁺23], decision-making [MMKS⁺21, LZY⁺24, ZXL⁺24], consensus reaching process [ZZC⁺22], and risk analysis [ZDX⁺20, XWZ22, ZZL⁺23]. However, basic fuzzy set theory lacks a mechanism to express how reliable that fuzzy information is. In a fuzzy rule-based system, a rule might state “IF project duration is about three months THEN outcome is successful,” but it does not indicate whether the expert providing that rule is very confident in it or just guessing. In other words, fuzzy sets capture what the information is (in a graded sense) but not how confident we are in that information. This is a crucial shortcoming for modeling real-world uncertainty. As research progressed, it became apparent that none of the conventional uncertainty-handling concepts could simultaneously address fuzziness in data and the reliability of that data’s source. This gap has prompted the search for more comprehensive models that can handle these two facets of uncertainty together.

To overcome the above limitations, Zadeh introduced the concept of Z-numbers [Zad11]. The Z-number is a novel framework that integrates a fuzzy quantity with its associated reliability measure, thereby handling two aspects of uncertainty simultaneously. Formally, a Z-number is defined as an ordered pair $Z = (\tilde{A}, \tilde{B})$, where $\tilde{A}$ is a fuzzy number representing the constraint (or value) of an uncertain variable, and $\tilde{B}$ is a fuzzy number representing the reliability (or confidence) of that value [Zad11]. The component $\tilde{A}$ answers the question “What do we know (approximately)?”, while $\tilde{B}$ answers “How sure are we about $\tilde{A}$?”. This pairing allows Z-numbers to encode statements like our earlier example in a structured way: for the statement “the project will likely finish in about three months,” we can interpret $\tilde{A}$ as a fuzzy number “about three months” and $\tilde{B}$ as a fuzzy assessment of “likely” (high confidence). By encoding both aspects, Z-numbers provide a richer and more faithful representation of uncertain information than either probability theory or standard fuzzy sets alone [AKKJG23, JLZ24].

The advantage of Z-numbers lies in their ability to represent real-world situations where information is imprecise and not fully reliable. By introducing a second dimension for reliability, they effectively extend classical fuzzy numbers, significantly enhancing their expressive power. In fact, Z-numbers have been shown to closely mimic the way humans communicate and reason about uncertainty [BPP21]. They allow us to work with knowledge expressed in natural language (which is often laden with ambiguities and hedges of certainty) in a mathematically rigorous manner [CCWM23, ZLWH24, LLWAB23]. Z-numbers can represent both the inherent uncertainty of a quantity and the “unreliability” of the human judgment providing that quantity, making them exceptionally well-suited for realistic decision models. Thanks to this comprehensive modeling capability, Z-numbers have attracted substantial attention and have been applied in various domains where handling uncertainty and confidence is crucial, such as risk assessment [QWX24, ZHMY24, Li24], decision-making [HL22, HYSHK25, UAA25], control systems [ABG24, BK24] and artificial intelligence [HGK24, AAA⁺24]. This thesis builds on the concept of Z-numbers to develop a new framework for linguistic decision modeling that addresses the identified shortcomings of existing approaches.

The primary objective of this research is to develop a Z-number linguistic modeling framework that enables more reliable reasoning under uncertainty. In essence, the goal is to design and implement a model that can handle fuzzy, imprecise information and explicitly account for the confidence in that information, thereby improving decision-making in scenarios with uncertain and partially trusted data. To achieve this overarching goal, several specific objectives are pursued.

First, we aim to formally establish a Z-number linguistic model, which serves as a foundational framework for translating expert knowledge into Z-number linguistic representations and performing computations. This model directly addresses the limitation of traditional fuzzy systems by incorporating a reliability dimension for each piece of information, thus bridging the gap identified earlier. Second, we seek to implement this model in practical reasoning systems, notably by developing a Z-number-valued rule-based system and a Z-number-valued decision tree. By extending a fuzzy rule-based inference engine and decision tree methodology to work with Z-numbers, we demonstrate how existing decision-support tools can be enhanced to consider both uncertainty and information reliability. Third, an important objective is to ensure the model is usable with real data inputs; to that end, we propose a Z-number generation method that can take raw data and produce the corresponding Z-number form. This tackles the challenge of obtaining Z-number inputs and makes the framework more applicable in practice. Finally, we aim to optimize and validate the performance of the proposed Z-number models. We integrate a machine learning approach - specifically, the Fuzzy Unordered Rule Induction Algorithm (FURIA) [HH09] - to refine the rule base and decision criteria. By using FURIA's ability to learn accurate fuzzy rules from data, we plan to tune the Z-number-based system for better accuracy and efficiency. In summary, the research objectives encompass the development of a comprehensive Z-number linguistic framework and its implementation in decision models, along with methods for Z-number generation and optimization, all directed at overcoming current limitations in uncertainty handling.

To fulfill the above objectives, this thesis makes several key contributions in the field of uncertainty modeling and intelligent systems:

- **Z-number linguistic model:** The thesis establishes a formal Z-number linguistic model for knowledge representation. This model provides a structured way to represent and operate linguistic information (e.g. expert statements or sensor observations) as Z-number linguistic term, encoding both a fuzzy value and its reliability. It lays the theoretical groundwork for integrating Z-numbers into reasoning processes, ensuring that uncertainty and confidence are accounted for simultaneously in analysis and decision-making.
- **Z-number-valued rule-based classification system:** A Z-number-valued rule-based classification system (ZRBCS) is developed, extending traditional fuzzy rule-based classification systems (FRBCS). In a conventional fuzzy rule-based system, rules might handle vague terms but assume full trust in those rules. Here, each rule is enhanced to carry Z-number information in its conditions. This means the rules can capture statements like “IF condition is $\tilde{A}$ (with reliability $\tilde{B}$) THEN outcome is class C ”. The contribution includes design of rule generation methods and inference engine. In Z-number-valued rule generation, negative samples are used to determine reliability. During inference, reliability is utilized to adjust the fuzzy partition, aiming to achieve higher classification performance.
- **Z-number-valued decision tree:** This thesis proposes a Z-number-valued rule-based decision tree (ZRDT) that extends the fuzzy rule-based decision tree (FRDT) [WLPZ15] by incorporating Z-numbers, which model both uncertainty and partial reliability. In the ZRDT, information gain replaces fuzzy confidence for feature selection, and negative samples are used to generate a second fuzzy number that adjusts the first fuzzy number for better data fit. Comparisons with FRDT, PUBLIC, C4.5, and AdaBoost.NC demonstrate that ZRDT achieves higher classification accuracy with a more compact decision tree structure.
- **Z-Number Generation Model:** Leveraging the unique structure and robust representation of uncertainty and partial reliability inherent in Z-numbers, our work introduces a novel non-linear model called Maximum Expected Minimum Entropy (MEME) to generate Z-numbers

from a set of probability distributions. Unlike traditional methods that derive probability distributions from a given Z-number, the MEME model works in reverse—summarizing multiple probability distributions into a single Z-number without the need for expert input. This approach not only streamlines the generation process but also enriches the information contained in the Z-number. Furthermore, we integrate the MEME model into classification by introducing Z-valuation if-then rules, where the deterministic consequent of a fuzzy rule is replaced with an uncertain Z-valuation. This enhancement leads to the development of a Z-valuation rule-based (ZVRB) classification system, which significantly improves decision-making in uncertain environments. Experimental results validate that the ZVRB system outperforms both classical classifiers and existing rule-based systems.

- **Optimization via FURIA:** The thesis integrates the Fuzzy Unordered Rule Induction Algorithm (FURIA) as an optimization tool for the developed Z-number models. FURIA is a state-of-the-art algorithm for learning fuzzy rules that has demonstrated high accuracy and interpretability [HH09]. By applying FURIA within the Z-number rule-based system, the rule set can be optimized—for example, by adjusting membership functions or selecting the most predictive features during rule generation—to enhance overall performance. This study not only demonstrates how the Z-number framework can be combined with machine learning techniques, but also develops an empirically validated enhanced model. Essentially, the integration of FURIA helps optimize the Z-number linguistic model and its rule base, thereby improving decision-making outcomes. Furthermore, it serves as evidence that the proposed approach can learn from data and adapt dynamically.

Together, these contributions advance the state of the art in managing uncertain and partially reliable information. They illustrate a full pipeline from theory to application: defining the Z-number linguistic representation, building inference systems around it, enabling its use with real inputs, and optimizing the solution for practical performance. The work fills a significant gap by providing tools that account for uncertainty in both value and source reliability, which is a step beyond what existing probability or fuzzy systems offer.

The remainder of the thesis is structured as follows: Section 2 provides an overview of pertinent preliminary concepts. In Section 3, the motivation behind this work is discussed, highlighting the fundamental ideas and challenges found in current research. Section 4 outlines the primary objectives pursued in this study. Section 5 details the methodologies implemented throughout the thesis, while Section 6 reports on the experimental results and proposals. A critical discussion of these findings is presented in Section 7, and Section 8 summarizes the main conclusions drawn from this work. Finally, Section 9 explores potential directions for future research.

Introducción

La toma de decisiones en el mundo real se ve a menudo complicada por la incertidumbre de la información y su fiabilidad parcial [AKKJG23]. En muchos campos —que abarcan desde la ingeniería y las finanzas hasta la atención sanitaria y la planificación diaria— debemos emitir juicios basados en datos o en opiniones de expertos que no son completamente confiables o ciertos. No solo puede ser impreciso el valor exacto de una cantidad, sino que nuestra confianza en ese valor también puede ser limitada. Por ejemplo, consideremos una estimación de programación: “es probable que el proyecto se termine en aproximadamente tres meses”. Esta afirmación encarna dos niveles de incertidumbre: “aproximadamente tres meses” representa una duración imprecisa (borrosa), mientras que el calificativo “probable” indica que la fuente no tiene plena seguridad respecto a esta estimación. Abordar esta dualidad de incertidumbre (tanto en el valor como en su fiabilidad) es un desafío significativo. Si un tomador de decisiones ignora la fiabilidad parcial de la información —por ejemplo, considerando una conjetura optimista con el mismo nivel de confianza que un hecho bien validado— las decisiones resultantes pueden ser subóptimas o incluso arriesgadas. Existe, por tanto, una motivación clara para desarrollar metodologías que puedan gestionar de forma explícita tanto la incertidumbre en los valores como en la fiabilidad, proporcionando un apoyo más matizado y robusto para la resolución de problemas en el mundo real.

Abordar la incertidumbre no es un problema nuevo. Los enfoques tradicionales, tales como la teoría de la probabilidad y la teoría de conjuntos difusos, han sido las herramientas principales para modelar la incertidumbre. Sin embargo, cada uno posee sus propias limitaciones cuando se trata de gestionar la fiabilidad parcial de la información [Fel08, Zad96]. La teoría de la probabilidad cuantifica la incertidumbre asignando probabilidades a eventos. Aunque es altamente eficaz en dominios con procesos aleatorios bien definidos, asume valores precisos de probabilidad y, a menudo, requiere abundantes datos o suposiciones firmes para derivar esos valores. Además, se enfrenta a dificultades para incorporar directamente información lingüística imprecisa: expresiones como “probable” o “aproximadamente” no se traducen de manera exacta en números de probabilidad precisos. Por otro lado, la teoría de la probabilidad no tiene en cuenta de manera inherente la fiabilidad de la fuente; una probabilidad se trata generalmente como una cantidad objetiva, y representar la confianza en esa probabilidad (incertidumbre de segundo orden) requiere extensiones complejas. En la práctica, esto significa que los modelos puramente probabilísticos pueden resultar insuficientes cuando la información es escasa, subjetiva o se presenta en lenguaje natural.

La teoría de conjuntos difusos, introducida por Zadeh, adopta un enfoque diferente al permitir que los elementos pertenezcan a un conjunto con una pertenencia graduada entre 0 y 1 [Zad96]. Este marco destaca por su capacidad para capturar la vaguedad y la ambigüedad lingüística. Por ejemplo, en lugar de establecer un límite rígido entre “rápido” y “lento”, los conjuntos difusos permiten que un elemento (por ejemplo, una velocidad de 50 km/h) pertenezca parcialmente al conjunto “rápido” y parcialmente no, reflejando así el carácter continuo del concepto. La lógica difusa nos posibilita trabajar con términos como “aproximadamente tres meses” mediante la definición de una función de pertenencia que asigna valores de verdad altos a duraciones cercanas a tres meses y valores más bajos a aquellas que se alejan de ese umbral. El método más común para el modelado de información lingüística es el enfoque lingüístico difuso [Zad75c, Zad75a, Zad75b]. Este aborda los aspectos cualitativos de la información representados en términos lingüísticos mediante variables cualitativas. Se ha aplicado con éxito en diversos campos, tales como la selección de proveedores [JC21], la optimización de rutas [DRC⁺23], problemas de externalización logística [WWC⁺23], la toma de decisiones [MMKS⁺21, LZY⁺24, ZXL⁺24], el proceso de logro de consenso [ZZC⁺22] y el análisis de riesgos [ZDX⁺20, XWZ22, ZZL⁺23]. Sin embargo, la teoría básica de conjuntos

difusos carece de un mecanismo para expresar la fiabilidad de dicha información difusa. En un sistema basado en reglas difusas, una regla podría afirmar “SI la duración del proyecto es de aproximadamente tres meses, ENTONCES el resultado es exitoso”, pero no indica si el experto que formula esa regla tiene una alta confianza en ella o simplemente está adivinando. En otras palabras, los conjuntos difusos capturan lo que es la información (en un sentido graduado), pero no cuán confiados estamos en dicha información. Esta limitación resulta crucial a la hora de modelar la incertidumbre en situaciones reales. Con el avance de la investigación, quedó claro que ninguno de los conceptos convencionales para el manejo de la incertidumbre podía abordar de manera simultánea la imprecisión en los datos y la fiabilidad de la fuente de esos datos. Esta brecha ha impulsado la búsqueda de modelos más integrales que puedan gestionar conjuntamente estas dos facetas de la incertidumbre.

Para superar las limitaciones mencionadas, Lotfi Zadeh introdujo el concepto de Z-números [Zad11]. El Z-número es un marco novedoso que integra una cantidad difusa con su medida de fiabilidad asociada, abordando así dos aspectos de la incertidumbre de forma simultánea. Formalmente, un Z-número se define como un par ordenado $Z = (\tilde{A}, \tilde{B})$, donde $\tilde{A}$ es un número difuso que representa la restricción (o valor) de una variable incierta, y $\tilde{B}$ es un número difuso que representa la fiabilidad (o confianza) de ese valor [Zad11]. El componente $\tilde{A}$ responde a la pregunta “¿Qué sabemos (aproximadamente)?”, mientras que $\tilde{B}$ responde “¿Qué tan seguros estamos acerca de $\tilde{A}$?”. Esta combinación permite a los Z-números codificar de forma estructurada afirmaciones como el ejemplo anterior: para la afirmación “el proyecto probablemente se terminará en aproximadamente tres meses”, podemos interpretar a $\tilde{A}$ como un número difuso “aproximadamente tres meses” y a $\tilde{B}$ como una evaluación difusa de “probablemente” (alta confianza). Al codificar ambos aspectos, los Z-números proporcionan una representación más rica y fiel de la información incierta que la teoría de la probabilidad o los conjuntos difusos estándar por sí solos [AKKJG23, JLZ24].

La ventaja de los Z-números radica en su capacidad para representar situaciones del mundo real en las que la información es imprecisa y no completamente fiable. Al introducir una segunda dimensión para la fiabilidad, extienden de manera efectiva los números difusos clásicos, mejorando significativamente su capacidad expresiva. De hecho, se ha demostrado que los Z-números imitan de cerca la forma en que los humanos comunican y razonan sobre la incertidumbre [BPP21]. Permiten trabajar con el conocimiento expresado en lenguaje natural (a menudo cargado de ambigüedades y matices de certeza) de manera rigurosa desde el punto de vista matemático [CCWM23, ZLWH24, LLWAB23]. Los Z-números pueden representar tanto la incertidumbre inherente a una cantidad como la “falta de fiabilidad” del juicio humano que proporciona esa cantidad, lo que los hace excepcionalmente adecuados para modelos de decisión realistas. Gracias a esta capacidad de modelado integral, los Z-números han atraído una atención considerable y se han aplicado en diversos campos donde el manejo de la incertidumbre y la confianza es crucial, tales como la evaluación de riesgos [QWX24, ZHMY24, Li24], la toma de decisiones [HL22, HYSHK25, UAA25], los sistemas de control [ABG24, BK24] y la inteligencia artificial [HGK24, AAA⁺24]. Esta tesis se fundamenta en el concepto de Z-números para desarrollar un nuevo marco para el modelado de decisiones lingüísticas que aborda las deficiencias identificadas en los enfoques existentes.

El objetivo principal de esta investigación es desarrollar un marco de modelado lingüístico basado en Z-números que permita un razonamiento más confiable bajo incertidumbre. En esencia, el objetivo es diseñar e implementar un modelo que pueda manejar información difusa e imprecisa y que tenga en cuenta de manera explícita la confianza en dicha información, mejorando así la toma de decisiones en escenarios con datos inciertos y parcialmente confiables. Para lograr este objetivo global, se persiguen varios objetivos específicos. En primer lugar, se pretende establecer formalmente un modelo lingüístico basado en Z-números, el cual servirá como un marco fundamental para

traducir el conocimiento experto en representaciones lingüísticas de Z-números y realizar cálculos. Este modelo aborda directamente la limitación de los sistemas difusos tradicionales al incorporar una dimensión de fiabilidad para cada pieza de información, cerrando así la brecha identificada anteriormente. En segundo lugar, se busca implementar este modelo en sistemas de razonamiento práctico, en particular mediante el desarrollo de un sistema basado en reglas con valores en Z-números y un árbol de decisión con valores en Z-números. Al extender un motor de inferencia basado en reglas difusas y la metodología de árboles de decisión para trabajar con Z-números, se demuestra cómo las herramientas de apoyo a la decisión existentes pueden ser mejoradas para considerar tanto la incertidumbre como la fiabilidad de la información. En tercer lugar, un objetivo importante es asegurar que el modelo sea usable con entradas de datos reales; para ello, se propone un método de generación de Z-números que pueda tomar datos sin procesar y producir la correspondiente forma de Z-números. Esto aborda el desafío de obtener entradas en Z-números y hace que el marco sea más aplicable en la práctica. Finalmente, se pretende optimizar y validar el rendimiento de los modelos de Z-números propuestos. Se integra un enfoque de aprendizaje automático —específicamente, el Algoritmo de Inducción de Reglas No Ordenadas Difusas (FURIA) [HH09]— para refinar la base de reglas y los criterios de decisión. Al utilizar la capacidad de FURIA para aprender reglas difusas precisas a partir de datos, se planea ajustar el sistema basado en Z-números para lograr una mayor precisión y eficiencia. En resumen, los objetivos de la investigación abarcan el desarrollo de un marco lingüístico integral basado en Z-números y su implementación en modelos de decisión, junto con métodos para la generación y optimización de Z-números, todo ello orientado a superar las limitaciones actuales en el manejo de la incertidumbre.

Para cumplir con los objetivos anteriores, esta tesis realiza varias contribuciones clave en el campo del modelado de la incertidumbre y los sistemas inteligentes:

- **Modelo lingüístico basado en Z-números:** La tesis establece un modelo lingüístico formal basado en Z-números para la representación del conocimiento. Este modelo proporciona una forma estructurada de representar y operar con información lingüística (por ejemplo, declaraciones de expertos u observaciones de sensores) como términos lingüísticos en Z-números, codificando tanto un valor difuso como su fiabilidad. Sienta las bases teóricas para integrar los Z-números en los procesos de razonamiento, garantizando que la incertidumbre y la confianza se tengan en cuenta de manera simultánea en el análisis y la toma de decisiones.
- **Sistema de clasificación basado en reglas con valores en Z-números:** Se desarrolla un sistema de clasificación basado en reglas que utiliza Z-números, extendiendo la inferencia basada en reglas difusas tradicional. En un sistema convencional basado en reglas difusas, las reglas pueden manejar términos imprecisos, pero asumen plena confianza en dichas reglas. Aquí, cada regla se mejora para llevar información en Z-números en sus condiciones. Esto significa que las reglas pueden capturar afirmaciones como “SI la condición es $\tilde{A}$ (con fiabilidad $\tilde{B}$) ENTONCES el resultado es la clase C ”. La contribución incluye el diseño de métodos de generación de reglas y un motor de inferencia. En la generación de reglas con valores en Z-números, se utilizan muestras negativas para determinar la fiabilidad. Durante la inferencia, la fiabilidad se utiliza para ajustar la partición difusa, con el objetivo de alcanzar un mayor rendimiento en la clasificación.
- **Árbol de decisión basado en Z-números:** Esta tesis propone un árbol de decisión basado en reglas con valores en Z-números (ZRDT) que extiende el árbol de decisión basado en reglas difusas (FRDT) [WLPZ15] al incorporar Z-números, los cuales modelan tanto la incertidumbre como la fiabilidad parcial. En el ZRDT, la ganancia de información reemplaza a la confianza difusa para la selección de características, y se utilizan muestras negativas para

generar un segundo número difuso que ajusta el número difuso primario para lograr un mejor ajuste a los datos. Las comparaciones con FRDT, PUBLIC, C4.5 y AdaBoost.NC demuestran que el ZRDT alcanza una mayor precisión en la clasificación con una estructura de árbol de decisión más compacta.

- **Modelo de generación de Z-números:** Aprovechando la estructura única y la representación robusta de la incertidumbre y la fiabilidad parcial inherentes a los Z-números, nuestro trabajo introduce un novedoso modelo no lineal llamado Máxima Entropía Mínima Esperada (MEME) para generar Z-números a partir de un conjunto de distribuciones de probabilidad. A diferencia de los métodos tradicionales que derivan distribuciones de probabilidad a partir de un Z-número dado, el modelo MEME funciona de manera inversa: resume múltiples distribuciones de probabilidad en un único Z-número sin necesidad de aportes expertos. Este enfoque no solo agiliza el proceso de generación, sino que también enriquece la información contenida en el Z-número. Además, integramos el modelo MEME en la clasificación mediante la introducción de reglas del tipo “si-entonces” con Z-valoración, donde el consecuente determinista de una regla difusa se reemplaza por una Z-valoración. Esta mejora conduce al desarrollo de un sistema de clasificación basado en reglas de Z-valoración (ZVRB), que mejora significativamente la toma de decisiones en entornos inciertos. Los resultados experimentales validan que el sistema ZVRB supera tanto a los clasificadores clásicos como a los sistemas basados en reglas existentes.
- **Optimización mediante FURIA:** La tesis integra el Algoritmo de Inducción de Reglas No Ordenadas Difusas (FURIA) como herramienta de optimización para los modelos de Z-números desarrollados. FURIA es un algoritmo de vanguardia para el aprendizaje de reglas difusas que ha demostrado alta precisión e interpretabilidad [HH09]. Al aplicar FURIA dentro del sistema basado en reglas con Z-números, se puede optimizar el conjunto de reglas —por ejemplo, ajustando las funciones de pertenencia o seleccionando las características más predictivas durante la generación de reglas— para mejorar el rendimiento global. Este estudio no solo demuestra cómo se puede combinar el marco de Z-números con técnicas de aprendizaje automático, sino que también desarrolla un modelo mejorado validado empíricamente. Esencialmente, la integración de FURIA ayuda a optimizar el modelo lingüístico basado en Z-números y su base de reglas, mejorando así los resultados en la toma de decisiones. Además, sirve como evidencia de que el enfoque propuesto puede aprender a partir de datos y adaptarse de manera dinámica.

En conjunto, estas contribuciones avanzan el estado del arte en el manejo de información incierta y parcialmente fiable. Ilustran una cadena completa que va desde la teoría hasta la aplicación: definiendo la representación lingüística mediante Z-números, construyendo sistemas de inferencia en torno a ella, permitiendo su uso con entradas reales y optimizando la solución para un rendimiento práctico. Este trabajo llena una brecha significativa al proporcionar herramientas que tienen en cuenta tanto la incertidumbre en el valor como la fiabilidad de la fuente, lo cual representa un avance respecto a lo que ofrecen los sistemas probabilísticos o difusos existentes.

El resto de la tesis está estructurado de la siguiente manera: La Sección 2 presenta una visión general de los conceptos preliminares relevantes. En la Sección 3 se discute la motivación que subyace a este trabajo, destacando las ideas fundamentales y los desafíos presentes en las investigaciones actuales. La Sección 4 expone los objetivos principales perseguidos en este estudio. La Sección 5 detalla las metodologías implementadas a lo largo de la tesis, mientras que la Sección 6 informa sobre los resultados experimentales y las propuestas planteadas. En la Sección 7 se realiza una discusión crítica sobre estos hallazgos, y la Sección 8 resume las principales conclusiones derivadas

del trabajo. Finalmente, la Sección 9 explora posibles líneas para futuras investigaciones.

2 Preliminaries

This section presents a concise overview of the background knowledge related to the fuzzy linguistic approach, Z-numbers, and fuzzy rule-based systems.

2.1 Fuzzy linguistic approach

People typically rely on everyday language for thinking and communication, often drawing inferences conveyed in verbal form. This approach, grounded in fuzzy logic, enables computational systems to analyze and interpret language by employing fuzzy sets to capture uncertainty. Under this framework, the nuanced qualitative features of language are embodied as verbal values via linguistic variables. According to Zadeh, a linguistic variable is defined as one whose outcomes are expressed in words and phrases rather than numerical figures.

Fuzzy sets capture how elements belong to sets by means of membership functions. Let X be a variable on a given universe of discourse U , with u denoting a typical element of U . In other words, we can write

$$X = u, \quad (\text{I.1})$$

which indicates that X takes the value u , where $u \in U$.

Definition 1 (Fuzzy set [Zad96]). *A fuzzy set $\tilde{A}$ is specified by a mapping*

$$\mu_{\tilde{A}} : U \rightarrow [0, 1].$$

For each $u \in U$, the value $\mu_{\tilde{A}}(u)$ describes the degree to which u belongs to $\tilde{A}$, with $\mu_{\tilde{A}}(u) = 1$ representing full membership, and $\mu_{\tilde{A}}(u) = 0$ indicating total non-membership.

Definition 2 (Fuzzy restriction [Zad78]). *Consider a fuzzy set $\tilde{A}$ specified on a domain U , with its membership function denoted by $\mu_{\tilde{A}}(\cdot)$. In this setting, $\tilde{A}$ acts as a fuzzy restriction on the variable X , functioning as an soft constraint on the values assignable to X :*

$$X = u : \mu_{\tilde{A}}(u). \quad (\text{I.2})$$

Here, $\mu_{\tilde{A}}(u)$ indicates the extent to which $\tilde{A}$'s constraint is fulfilled once u is assigned to X .

Definition 3 (Possibility distribution [Zad78]). *Suppose X is a variable that takes values in some domain U . Let $\tilde{A}$ be a fuzzy restriction associated with X . Under this view, the statement “ X is $\tilde{A}$ ” gives rise to a possibility distribution Π . The corresponding distribution function $\pi(\cdot)$ coincides with the membership function of $\tilde{A}$, namely*

$$\pi(u) = \mu_{\tilde{A}}(u). \quad (\text{I.3})$$

Definition 4 (Possibilistic restriction [Zad11]). *A possibilistic restriction is formulated on a fuzzy set $\tilde{A}$ as follows:*

$$X \text{ is } \tilde{A}, \quad (\text{I.4})$$

where $\tilde{A}$ is a fuzzy set within the domain U , characterized by its membership function $\mu_{\tilde{A}}(\cdot)$. In addition, $\tilde{A}$ represents the possibility distribution of X , such that

$$\text{Poss}(X = u) = \mu_{\tilde{A}}(u). \quad (\text{I.5})$$

The measure of possibilistic uncertainty for a possibilistic distribution Π is defined as follows [HK82]. When Π is discrete,

$$U(\Pi) = \frac{1}{l_r} \sum_{k=1}^{r-1} (l_{k+1} - l_k) \log_2(|c(\Pi, l_{k+1})|), \quad (\text{I.6})$$

where $c(\Pi, l)$ denotes the l -cut of Π , defined as

$$c(\Pi, l) = \{u \in U \mid \pi(u) \geq l\}; \quad (\text{I.7})$$

and

$$L_\Pi = \{l_1 = 0, \dots, l_r \mid l_k = \pi(u), l_k \leq l_{k+1}\} \quad (\text{I.8})$$

represents the level set of Π .

Consider X as a discrete variable that takes values in the set U . Let $\{\pi(u_i)\}_{i=1}^n$ represent its possibility distribution, where $u_1 \leq u_2 \leq \dots \leq u_n$ are elements of U . Following [LL02], the expected value of X , denoted by $E[X]$, is defined as

$$E[X] = \sum_{i=1}^n u_i w_i, \quad (\text{I.9})$$

where each coefficient w_i is given by

$$w_i = \frac{1}{2} \left(\max_{j=1}^i \pi(u_j) - \max_{j=1}^{i-1} \pi(u_j), 0 \right) + \frac{1}{2} \left(\max_{j=i}^n \pi(u_j) - \max_{j=i+1}^n \pi(u_j), 0 \right). \quad (\text{I.10})$$

It follows that $w_i \in [0, 1]$ for all i , and $\sum_{i=1}^n w_i = 1$.

Definition 5 (Fuzzy number [Zad96]). *A fuzzy number is defined as a fuzzy set $\tilde{A}$ on $\mathbb{R}$ that meets the following criteria:*

- (1) *The membership function attains its maximum value, i.e.,*

$$\max_{u \in U} \mu_{\tilde{A}}(u) = 1;$$

- (2) *For any two points $u_1, u_2 \in U$ and for every $\lambda \in [0, 1]$, the membership value at the convex combination satisfies*

$$\mu_{\tilde{A}}(\lambda u_1 + (1 - \lambda)u_2) \geq \min(\mu_{\tilde{A}}(u_1), \mu_{\tilde{A}}(u_2));$$

- (3) *The support of $\tilde{A}$, denoted by $\text{supp}(\tilde{A})$, is a bounded subset of $\mathbb{R}$.*

Definition 6 (Fuzzy linguistic term set [Yag95, HHVM00]). *A fuzzy linguistic term set is denoted by $S = \{s_0, s_1, \dots, s_g\}$, which forms a finite and totally ordered collection of fuzzy linguistic terms arranged on a scale. Each term s_i (with $0 \leq i \leq g$) satisfies the following conditions:*

- (1) **Negation:** *The negation operator is defined by $\text{Neg}(s_i) = s_{g-i}$, reflecting the fact that the set comprises $g + 1$ terms.*
- (2) **Ordering:** *For any two terms, s_i is considered greater than s_j if and only if $i > j$.*

The semantics of these linguistic terms are conveyed through the membership functions of the associated fuzzy sets.

2.2 Z-numbers

Zadeh observed that real-world information is characterized not only by uncertainty but also by partial reliability [Zad11]. To capture both aspects simultaneously, he introduced the concept of Z-numbers [Zad11]. A Z-number is represented by an ordered pair of fuzzy numbers, which may be continuous or discrete.

Definition 7 (Z-number). *A Z-number is defined as an ordered pair of fuzzy numbers, namely, $Z = (\tilde{A}, \tilde{B})$, where the fuzzy number $\tilde{A}$ acts as a constraint on the possible values of a random variable X :*

$$X \text{ is } \tilde{A}, \quad (\text{I.11})$$

and the fuzzy number $\tilde{B}$, with membership function

$$\mu_{\tilde{B}} : [0, 1] \rightarrow [0, 1],$$

serves as a constraint on the probability measure associated with $\tilde{A}$:

$$P(\tilde{A}) \text{ is } \tilde{B}. \quad (\text{I.12})$$

In this context, the probability measure $P(\tilde{A})$ is given by

$$P(\tilde{A}) = \int_{\mathbb{R}} \mu_{\tilde{A}}(u) p(u) du, \quad (\text{I.13})$$

where $p(\cdot)$ denotes the probability density function of X , and it holds that $P(\tilde{A}) \in \text{supp}(\tilde{B})$.

If the centroids of $\tilde{A}$ and the expected value of probability distribution p coincide, the two distributions are considered compatible:

$$\int_U u p(u) du = \frac{\int_U u \mu_{\tilde{A}}(u) du}{\int_U \mu_{\tilde{A}}(u) du}. \quad (\text{I.14})$$

For a given Z-number, both components are either discrete or continuous. However, in practical applications, it is common to encounter scenarios where continuous and discrete attributes coexist. To accommodate such cases, Xian et al. introduced the concept of Z-mixture-numbers [XCLH21].

Definition 8 (Z-mixture-number). *Given a Z-number $Z = (\tilde{A}, \tilde{B})$, where the fuzzy numbers $\tilde{A}$ and $\tilde{B}$ may be either continuous or discrete, Z is called a Z-mixture-number.*

Based on the nature of its two components, Z-mixture-numbers can be classified into four types:

1. Both $\tilde{A}$ and $\tilde{B}$ are discrete;
2. Both $\tilde{A}$ and $\tilde{B}$ are continuous;
3. $\tilde{A}$ is discrete while $\tilde{B}$ is continuous;
4. $\tilde{A}$ is continuous while $\tilde{B}$ is discrete.

Moreover, a linguistic Z-number is an extension of Z-numbers that represents both components using fuzzy linguistic terms [WCZ17, WWTZ18].

Definition 9 (Linguistic Z-number [WCZ17, WWTZ18]). *Let U be a universe of discourse, and let*

$$S_1 = \{s_0, s_1, \dots, s_{g_1}\} \quad \text{and} \quad S_2 = \{s'_0, s'_1, \dots, s'_{g_2}\}$$

be two fuzzy linguistic term sets. A linguistic Z-number on X is defined as

$$Z = \{(x, A_{\phi(x)}, B_{\varphi(x)}) \mid x \in U\},$$

where $A_{\phi(x)} \in S_1$ serves as a fuzzy restriction on the potential values of the variable, and $B_{\varphi(x)} \in S_2$ provides a fuzzy restriction on the reliability of $A_{\phi(x)}$.

Let $Z_1 = (A_{\phi(1)}, B_{\varphi(1)})$ and $Z_2 = (A_{\phi(2)}, B_{\varphi(2)})$ be two linguistic Z-numbers, and let f and h be strictly monotonically increasing functions. Some linguistic Z-number operators are defined as follows [WCZ17, WWTZ18]:

- (1) $\text{neg}(Z_1) = \left(f^{-1}\left(f(s_{g_1}) - f(A_{\phi(1)})\right), h^{-1}\left(h(s_{g_2}) - h(B_{\varphi(1)})\right) \right);$
- (2) $Z_1 + Z_2 = \left(f^{-1}\left(f(A_{\phi(1)}) + f(A_{\phi(2)})\right), h^{-1}\left(\frac{f(A_{\phi(1)}) \cdot h(B_{\varphi(1)}) + f(A_{\phi(2)}) \cdot h(B_{\varphi(2)})}{f(A_{\phi(1)}) + f(A_{\phi(2)})}\right) \right);$
- (3) $\lambda Z_1 = \left(f^{-1}\left(\lambda f(A_{\phi(1)})\right), B_{\varphi(1)} \right), \quad \lambda \geq 0;$
- (4) $Z_1 \times Z_2 = \left(f^{-1}\left(f(A_{\phi(1)}) \cdot f(A_{\phi(2)})\right), h^{-1}\left(h(B_{\varphi(1)}) \cdot h(B_{\varphi(2)})\right) \right);$
- (5) $Z_1^\lambda = \left(f^{-1}\left(f(A_{\phi(1)})^\lambda\right), h^{-1}\left(h(B_{\varphi(1)})^\lambda\right) \right), \quad \lambda \geq 0.$

The comparison method is defined as follows [WCZ17, WWTZ18]:

- (1) If $A_{\phi(1)} > A_{\phi(2)}$ and $B_{\varphi(1)} > B_{\varphi(2)}$, then Z_1 is strictly greater than Z_2 , denoted by $Z_1 > Z_2$;
- (2) If $S(Z_1) > S(Z_2)$, or if $S(Z_1) = S(Z_2)$ and $A(Z_1) > A(Z_2)$, then Z_1 is greater than Z_2 , denoted by $Z_1 \succ Z_2$;
- (3) If $S(Z_1) = S(Z_2)$ and $A(Z_1) = A(Z_2)$, then Z_1 is equal to Z_2 , denoted by $Z_1 \approx Z_2$;
- (4) If $S(Z_1) = S(Z_2)$ and $A(Z_1) < A(Z_2)$, or if $S(Z_1) < S(Z_2)$, then Z_1 is less than Z_2 , denoted by $Z_1 \prec Z_2$.

Here,

$$S(Z_1) = f(A_{\phi(1)})g(B_{\varphi(1)}) \quad \text{and} \quad A(Z_1) = f(A_{\phi(1)})(1 - g(B_{\varphi(1)})).$$

The operations proposed by Wang et al. lack sufficient theoretical support from Z-number theory and exhibit certain limitations [WCZ17]. Firstly, if both linguistic Z-numbers contain s_0 , then according to the above operation rules, their addition cannot be computed. For example, when calculating $(s_0, s'_3) + (s_0, s'_5)$, the denominator becomes $f(s_0) + f(s_0) = 0$. Moreover, in the comparison method of linguistic Z-numbers, according to the third rule, $(s_0, s'_0) \approx (s_0, s'_6)$ holds because

$$S((s_0, s'_0)) = S((s_0, s'_6)) = 0 \quad \text{and} \quad A((s_0, s'_0)) = A((s_0, s'_6)) = 0.$$

This conclusion is counterintuitive.

2.3 Fuzzy rule-based systems

Fuzzy rule-based systems are designed to handle imprecise and uncertain information while making decisions based on a predefined set of fuzzy rules and fuzzy logic. In such systems, rules are typically expressed as “if-then” statements, where specific conditions trigger corresponding actions or outcomes. A standard fuzzy rule-based system consists of two main components: a knowledge base and an inference engine. The knowledge base contains the fuzzy rules, and the inference engine applies these rules to the input data. This framework offers several advantages, including clear interpretability, high modularity, robustness, and strong scalability.

Let's consider a classification problem. Suppose there is a dataset $D = \{(x_1, y_1), \dots, (x_n, y_n)\}$, comprising n samples and M features, where each sample $x_i = \{x_{i1}, \dots, x_{iM}\}$ is represented by its feature vector, x_{ij} denotes the value of feature f_j , and y_i is the class label of the i th sample. Our objective is to predict the class label y_i based on the given feature vector x_i . Suppose there are W different class labels, denoted by the set $\Omega = \{\theta_1, \dots, \theta_W\}$.

Definition 10 (Fuzzy rule). *A typical fuzzy rule in a fuzzy rule-based system for classification tasks can be expressed as follows:*

$$\text{Rule } R_l: \quad \text{If } x_{i1} \text{ is } \tilde{A}_{l1} \text{ and } \dots \text{ and } x_{iM} \text{ is } \tilde{A}_{lM}, \text{ then } y_i \text{ is } C_l \text{ with } CF_l. \quad (\text{I.15})$$

Here, $x_i = \{x_{i1}, \dots, x_{iM}\}$ denotes the input feature vector; $\tilde{A}_{l1}, \dots, \tilde{A}_{lM}$ are the antecedent fuzzy numbers associated with the M input features; y_i is the output variable; C_l is the consequent class label with $C_l \in \Omega$; and CF_l is the certainty grade of rule R_l , where $CF_l \in [0, 1]$.

Definition 11 (Compatibility grade with input pattern). *The compatibility grade of rule R_l with the input pattern x_i is defined using the product operator as*

$$\mu_l(x_i) = \prod_{j=1}^M \mu_{\tilde{A}_{lj}}(x_{ij}), \quad (\text{I.16})$$

where $\mu_{\tilde{A}_{lj}}$ is the membership function of the fuzzy number $\tilde{A}_{lj}$, for $j = 1, 2, \dots, M$.

Definition 12 (Compatibility grade with respect to a class). *The compatibility grade of rule R_l with respect to class θ_w is defined as:*

$$\psi_l(\theta_w) = \sum_{x_i \in \theta_w} \mu_l(x_i). \quad (\text{I.17})$$

In [INT92], Ishibuchi et al. introduced an uncomplicated fuzzy grid approach to segment a single dimension into fuzzy numbers. In this method, the dimension is evenly partitioned into K triangular fuzzy numbers, which are denoted by $\tilde{A}_1, \tilde{A}_2, \dots, \tilde{A}_K$. Each antecedent fuzzy number $\tilde{A}_{lj}$ is chosen from these K numbers. Consequently, the entire pattern space is divided into $L = K^M$ regions, with each corresponding to a distinct fuzzy if-then rule. The consequent class for a rule R_l is determined as $C_l = \theta_{w^*}$ if

$$\psi_{R_l}(\theta_{w^*}) = \max\{\psi_l(\theta_1), \psi_l(\theta_2), \dots, \psi_l(\theta_W)\}. \quad (\text{I.18})$$

If more than one maximum occurs in Eq. (I.18), then the consequent class is defined as empty, i.e., $C_l = \emptyset$.

The certainty grade of rule R_l is computed by

$$CF_l = \frac{\psi_l(C_l) - \bar{\psi}}{\sum_{k=1}^K \mu_l(\theta_w)}, \quad (\text{I.19})$$

with

$$\bar{\psi} = \frac{\sum_{\theta_w \neq C_l} \psi_l(\theta_w)}{K - 1}. \quad (\text{I.20})$$

For classifying a new pattern x_{new} , the rule R_{l^*} is selected as the one achieving the highest value of

$$\mu_l(x_{\text{new}}) \times CF_l,$$

that is,

$$\mu_{l^*}(x_{\text{new}}) = \max \left\{ \mu_1(x_{\text{new}}) \times CF_1, \mu_2(x_{\text{new}}) \times CF_2, \dots, \mu_L(x_{\text{new}}) \times CF_L \right\}. \quad (\text{I.21})$$

The new pattern is then assigned the class C_{l^*} associated with rule R_{l^*} . If several rules achieve the same maximum value in Eq. (I.21), the pattern is considered unclassifiable.

There exist numerous strategies for generating fuzzy rules and executing inference, among which FURIA stands out as a state-of-the-art method that balances accuracy with complexity. FURIA is built upon a modified version of the Repeated Incremental Pruning to Produce Error Reduction (RIPPER) algorithm and has shown notable improvements in rule induction performance [HH09, Coh95]. In this framework, the crisp rules derived from the modified RIPPER algorithm are transformed into fuzzy rules through a fuzzification process. A typical fuzzy rule produced by FURIA is formulated as

$$\text{Rule } R_l : \quad \text{If } x_{i\rho(1)} \text{ is } \tilde{A}_{l\rho(1)} \text{ and } \dots \text{ and } x_{i\rho(d_l)} \text{ is } \tilde{A}_{l\rho(d_l)}, \text{ then } x_i \text{ belongs to class } y_l \text{ with } RW_l. \quad (\text{I.22})$$

Here, for $j = 1, \dots, d_l$ with $d_l \leq m$, $\tilde{A}_{l\rho(j)}$ is an antecedent trapezoidal fuzzy set; $\rho(j)$ indicates the feature index used in the j -th condition; $y_l \in \Theta$ represents the output class label; and $RW_l \in [0, 1]$ denotes the rule weight.

3 Justification

Given the profound theoretical significance of Z-numbers, they inherently possess an advantage in handling information expressed in natural language. They offer greater representational capability than fuzzy logic in linguistic modelling. Consequently, applying Z-numbers can enhance the performance of systems that utilize linguistic modelling. The aim of this thesis is to develop a comprehensive Z-number linguistic model that encompasses both theoretical refinement and practical application enhancements.

3.1 Z-number linguistic modeling

Fuzzy linguistic modeling adeptly manages the inherent uncertainty in information by leveraging linguistic terms and fuzzy numbers. Among the various approaches for modeling linguistic data, the fuzzy linguistic approach is the most widely used, as it captures the qualitative characteristics of information through linguistic variables.

Moreover, when processing information, it is vital to consider not only its uncertainty but also partial reliability. Since the trustworthiness of data is closely linked to its source, unreliable information may misguide decision-makers, leading to erroneous choices and significant economic or resource losses. Thus, evaluating information reliability is crucial for sound decision-making. To more comprehensively represent both uncertainty and partial reliability, Zadeh introduced the concept of the Z-number. However, performing arithmetic operations on Z-numbers—especially continuous ones—involves solving multiple optimization problems to find appropriate hidden probability distributions, which can be computationally intensive and impractical for large datasets.

In order to mitigate these challenges, Wang et al. integrated linguistic terms with Z-numbers, defining linguistic Z-numbers as either a variant or a subclass of Z-numbers [WCZ17]. This extension models both components with fuzzy linguistic terms and introduces several associated operations. Thanks to their intuitive depiction of linguistic information and a relatively simple computational framework, linguistic Z-numbers have been successfully applied in various scenarios that rely on Z-number based data.

Nevertheless, despite the computational simplicity of the approach by Wang et al., it suffers from insufficient theoretical backing within Z-number theory and exhibits certain limitations. Therefore, there is a need to develop a new linguistic Z-number approach that not only conforms to intuitive expectations but also rests on a solid theoretical foundation in Z-number theory.

3.2 Z-number-valued rule-based system

Fuzzy rule-based systems have emerged as one of the most effective implementations of fuzzy sets and fuzzy logic, primarily due to their clarity and ease of interpretation. In practical scenarios, a major challenge in building such systems lies in deriving fuzzy if-then rules directly from numerical data. To formulate a fuzzy rule, one must identify the antecedent fuzzy numbers, the corresponding fuzzy classes for the conclusion, and the rule weight. The most straightforward approach is the fuzzy grid technique, which employs local representations to equally partition a dimension into K fuzzy linguistic terms (with K typically taking values such as 2, 3, 4, etc.). This method segments the pattern space into K^M unique fuzzy regions, where M denotes the number of features. The effectiveness of this approach is entirely contingent upon the quality of the fuzzy partitions. When these partitions are overly sparse, a significant number of patterns may be misclassified; conversely,

if the partitions are too dense, redundant rules may be generated due to an insufficient number of training samples.

The antecedent fuzzy numbers in fuzzy rules are pre-established and remain fixed throughout the training process, thereby fully determining the system's performance. By incorporating Z-numbers, this rigidity can be alleviated, leading to improved system performance. Analogous to fuzzy linguistic modeling, Z-number linguistic modeling enables the construction of rule-based systems where the rules are expressed in terms of Z-numbers. Figure 1 illustrates the architecture of a Z-number-valued rule-based system. Generally, such a system comprises two primary components: a knowledge base and a processing framework. The knowledge base retains all available information as Z-number-based if-then rules, while the processing framework conducts inference on the input data using these rules. Within the knowledge base, fuzzy partitions define the linguistic term sets associated with each variable and specify their corresponding membership functions. A scaling function is employed to translate between the universe of discourse—where the fuzzy sets are defined—and the domains of the system's input and output variables. Furthermore, a fuzzification interface converts crisp input data into Z-numbers for the inference engine, and a defuzzification interface translates the resulting output Z-numbers back into crisp values. The inference engine then derives several output Z-numbers from the Z-number inputs based on the Z-number-valued rules stored in the knowledge base.

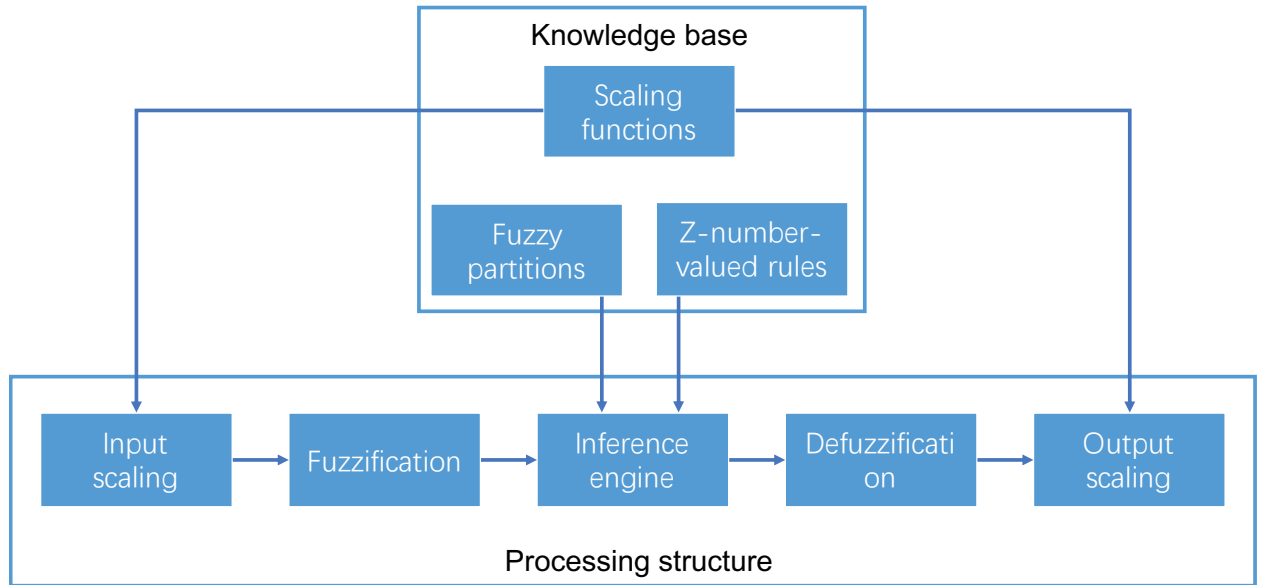


Figure 1: The structure of Z-number-valued rule-based systems.

3.3 Advanced optimization

Many experts argue that human recognition may be governed by a series of linguistic rules. In areas such as learning, object identification, understanding, and other cognitive functions, rule-based symbolic processing is considered one of the most effective means to elucidate human intelligence. To enhance system performance, boost interpretability, simplify complexity, and reduce costs, a number of advanced techniques have been developed for extracting fuzzy if-then rules from numerical data. Motivated by these innovations, this thesis introduces several novel algorithms designed to optimize Z-number-valued rule-based systems, thereby increasing accuracy while curtailing com-

plexity.

3.3.1 Z-number-valued rule-based decision tree

The FRDT represents a distinct form of fuzzy decision tree, characterized by an oblique, multi-layered structure [WLPZ15]. In this architecture, each node is associated with a fuzzy rule that outputs a unique class and involves several features. To maintain a compact tree, only one fuzzy rule per class is allocated in each layer. If a sample is not classified by the rules of the current layer, it is subsequently passed to the next layer. Unlike conventional fuzzy rules that take into account all features, those in the FRDT concentrate solely on the most influential ones. This selection is guided by the fuzzy confidence index, which assesses the rule’s validity. Overall, FRDT outperforms many other fuzzy rule-based algorithms in terms of both classification accuracy and the succinctness of its rules.

Inspired by this design, we propose the development of a Z-number-valued rule-based decision tree (ZRDT) to further boost the performance of systems based on Z-numbers. Given that Z-numbers offer a more comprehensive modeling framework than traditional fuzzy numbers, their integration is expected to enhance the efficacy of fuzzy rule-based decision trees. The structure of the proposed ZRDT mirrors that of the FRDT—an oblique, layered tree as illustrated in Fig. 2—with the key difference that the fuzzy rules are replaced by Z-number-valued rules.

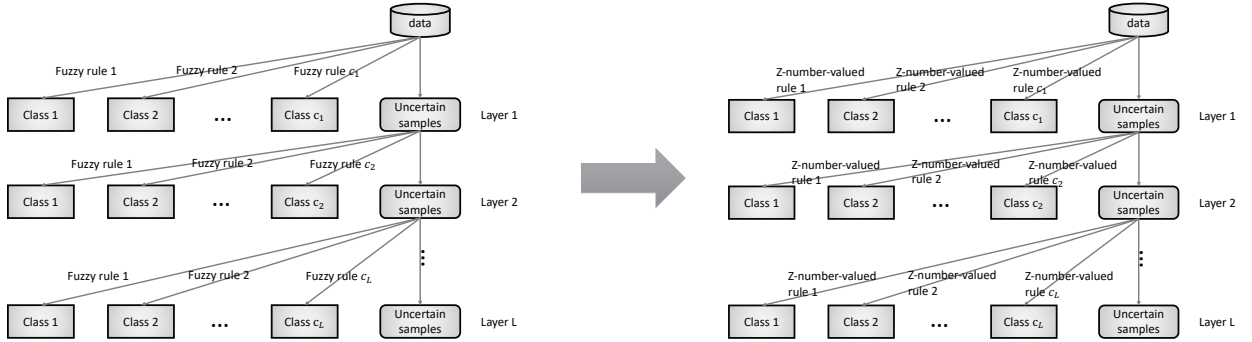


Figure 2: The structure of FRDT (left) and the structure of ZRDT (right).

3.3.2 Optimization via FURIA

Developed from the RIPPER algorithm, FURIA has demonstrated significant enhancements in rule induction performance. In contrast to conventional rules, those produced by FURIA are notably more adaptable in modeling classification boundaries. The fuzzy rules it generates are based on trapezoidal fuzzy numbers instead of the typical triangular ones, and they integrate sophisticated rule induction techniques inherited from RIPPER. By employing a 'one-vs-rest' strategy, FURIA constructs an unordered collection of rules for each class. Rather than relying on preset fuzzy partitions, it utilizes an approach that optimizes information gain to derive fuzzy antecedents, thereby reducing computational complexity while substantially boosting overall performance.

Moreover, several techniques developed in FURIA can be applied to a Z-number-valued rule-based system to further improve its efficiency. For instance, methods for generating fuzzy antecedents can help shorten the rule length, and advanced rule induction strategies can reduce the overall size of the rule base.

3.3.3 Z-number generation model

Central to the definition of Z-numbers is the probability measure associated with fuzzy numbers, which effectively links the two fuzzy components. This measure is computed through a underlying probability distribution that, although not directly observable, is indirectly constrained by a fuzzy restriction. In this framework, the second fuzzy number can be viewed as imposing a fuzzy restriction on the probability measure defined by the first fuzzy number. As Zadeh once noted, a Z-number can serve as a summary of a series of probability distributions [Zad11]. Researchers have proposed multiple techniques to infer the underlying probability distributions from a given Z-number, thereby deriving a set of possible distributions. Conversely, an interesting open question is whether a single Z-number could be employed to summarize a set of probability distributions—a problem that remains largely unexplored.

Addressing this challenge would bring Z-number-valued rules into closer alignment with the theoretical foundations of Z-number theory, thereby fully capitalizing on the advantages offered by Z-numbers. Moreover, if this issue can be resolved, Z-numbers can be generated directly from data without the need for expert intervention, which would yield significant practical benefits.

4 Objectives

- (1) **Develop a novel Z-number linguistic model.** The first objective is to establish a new linguistic model rooted in Z-numbers that is both intuitively sound and theoretically robust. Initially, a definition for Z-number linguistic terms will be introduced, wherein each term consists of two ordered fuzzy linguistic terms. Next, a complete set of these Z-number linguistic terms, arranged in ascending order, will be defined. Fundamental operations—such as comparison, negation, and aggregation—will then be formulated for these terms. Finally, the model's effectiveness and advantages will be demonstrated through its application to a multi-attribute group decision-making problem.
- (2) **Construct a Z-number-valued rule-based system.** The second objective involves designing a rule-based system where the rules are expressed in terms of Z-numbers. In the rule-generation process, the first fuzzy number in the antecedent remains identical to that used in classical fuzzy rule-based systems. However, the second fuzzy number is employed to adjust the first, ensuring a better alignment with training patterns. This second component also provides valuable information regarding the positions of negative samples, which aids in accurate classification. By incorporating this additional information, the proposed system goes beyond the capabilities of traditional fuzzy rule-based methods. Its performance will be validated through real-world applications.
- (3) **Develop a Z-number-valued rule-based decision tree (ZRDT).** In this objective, the structure of a decision tree based on Z-number-valued rules will be defined. During tree construction, initial fuzzy numbers derived from the mean values of the training data will serve as the first component of the antecedent. Then, a second fuzzy number—constructed from the negative samples covered by the rule—is used to adjust this initial value. For feature selection in rule generation, information gain may replace the fuzzy confidence index employed in FRDT. The effectiveness of this Z-number-valued decision tree will be confirmed through its application to practical problems.
- (4) **Formulate a Z-number generation model.** Given a set of probability distributions for a variable, this objective is to summarize them into a single representative Z-number. Here, the base values of the first fuzzy number correspond to the potential values that the variable can take. Since a higher expected reliability and a lower entropy imply clearer and more decisive information, the model will aim to maximize expected reliability while minimizing entropy. This optimization will yield a unique Z-number that can be automatically generated from data, without relying on expert knowledge. The resulting model will be integrated into the Z-number-valued rule-based system and evaluated using real-world scenarios.
- (5) **Enhance the Z-number-valued rule-based system using FURIA.** Finally, the techniques from FURIA for generating antecedents and inducing rules will be incorporated to optimize the Z-number-valued system. These methods are expected to shorten rule lengths and reduce the overall size of the rule base, thereby boosting system performance. The improved system will then be validated through application to practical problems.

5 Methodology

Based on the aforementioned objectives, the primary goal of this thesis is to develop theories and methods for Z-number linguistic modeling—building upon fuzzy linguistic modeling—and to explore its applications in rule-based systems. The primary methods employed in this study are as follows:

- (1) **Literature Review.** By reviewing the theoretical foundations and current research on Z-numbers, an understanding of their developmental background, application areas, and differences from traditional fuzzy numbers is attained. By reviewing the theories and previous studies on fuzzy linguistic modeling, the background of the research problem and the theoretical foundations of the study are established. Unresolved issues or underexplored areas in existing fuzzy linguistic modeling are identified, and the potential of incorporating Z-numbers to address these problems or provide valuable insights is considered. The methods used in existing fuzzy linguistic modeling are analyzed and evaluated for their effectiveness, and appropriate approaches are selected—or improved methods for Z-number linguistic modeling are proposed.
- (2) **Mathematical modeling.** Mathematical tools and models are employed to formalize the Z-number linguistic model. This entails defining new linguistic terms, a linguistic term set, operational rules, and solution algorithms for the Z-number linguistic model, thereby accommodating the simultaneous handling of uncertainty and reliability. The resulting mathematical models and algorithms are designed to enhance the computational efficiency and real-time data processing capabilities of the system.
- (3) **Algorithm development.** Algorithm development is central to the implementation of the Z-number-valued rule-based system, with a focus on creating efficient computational methods to process, analyze, and reason with input-output data. This involves developing new algorithms or enhancing existing ones to generate rules, perform classification or prediction tasks, and optimize the performance and accuracy of the decision-making process.
- (4) **Experiments and simulations.** Experiments and simulations are key methodologies for evaluating and validating the performance of the Z-number-valued rule-based system. By constructing a simulated environment, the system's responses under controlled conditions can be tested, its performance across various scenarios analyzed, and system parameters adjusted to achieve optimal performance.
- (5) **Case analysis.** Specific real-world application scenarios and their datasets are selected for case analysis, such as the application of the Z-number linguistic model and Z-number-valued rule-based systems in financial risk assessment, medical diagnosis, or supply chain management. By analyzing these cases, deeper insights into the performance and challenges of the Z-number linguistic model in practical applications can be obtained.
- (6) **Comparative studies.** Comparative studies play a crucial role in the research on Z-number-valued rule-based systems by contrasting various decision support systems, including traditional fuzzy systems and other uncertainty-based approaches. This approach evaluates the unique advantages and potential limitations of Z-number-valued rule-based systems. Through such studies, the effectiveness of Z-number-valued rule-based systems in handling complex and uncertain information can be demonstrated, thereby highlighting improvements over other methods.

Collectively, these research methods form a comprehensive research framework that supports both the development and implementation of Z-number-valued rule-based systems and ensures their theoretical rigor as well as practical effectiveness. The literature review establishes the theoretical foundation, whereas mathematical modeling and algorithm development construct the system framework. Experiments, simulations, and case analyses validate the system's performance, while comparative studies highlight its advantages over similar technologies. The entire research process aims to promote the widespread application and advancement of Z-number-valued rule-based systems in various real-world scenarios.

6 Summary

In this section, we provide an overview of the primary proposals introduced in this thesis and detail the key findings and results reported in the journal publications. The list of published and submitted works is presented below.

- (1) Y. Li, G. Kou, Y. Peng, and J. A. Morente-Molinera, Z-number linguistic term set for multi-criteria group decision-making and its application in predicting the acceptance of academic papers, *Applied Intelligence*, 54(21),10962-10981, 2024, <https://doi.org/10.1007/s10489-024-05765-8>.
- (2) Y. Li, E. Herrera-Viedma, I. J. Pérez, M. Barragán-Guzmán, and J. A. Morente-Molinera, Z-number-valued rule-based classification system, *Applied Soft Computing*, 137, 110168, 2023, <https://doi.org/10.1016/j.asoc.2023.110168>.
- (3) Y., Li, E. Herrera-Viedma, G. Kou, and J. A. Morente-Molinera, Z-number-valued rule-based decision trees, *Information Sciences*, 643, 119252, 2023, <https://doi.org/10.1016/j.ins.2023.119252>.
- (4) Y. Li, J. A. Morente-Molinera, J. R. Trillo, and E. Herrera-Viedma, Z-number generation model and its application in a rule-based classification system, *IEEE Transactions on Cybernetics*, 2024, DOI: 10.1109/TCYB.2025.3545195.
- (5) Y. Li, I. J. Pérez, F. J. Cabrerizo, and J. A. Morente-Molinera, Z-mixture-number if-then rules induction algorithm for classification tasks, submitted to *IEEE Transactions on Systems, Man, and Cybernetics*, 2024.

This section is organized into five parts based on the primary research objectives (Section 4). Subsection 6.1 introduces the establishment of the Z-number linguistic model. Subsection 6.2 presents the Z-number-valued rule-based classification system. Subsection 6.3 further proposes Z-number-valued rule-based decision trees. Subsection 6.4 introduces the Z-number generation Model. Finally, Subsection 6.5 optimizes the generation algorithm for Z-number-valued rules using some techniques in FURIA, improving system accuracy and reducing complexity.

6.1 Z-number linguistic model

The proposed Z-number linguistic model is built upon two essential components. The first component is a model for representing linguistic information using Z-numbers, while the second is a computational framework designed to aggregate such linguistic data. In addition, we examine the underlying probability density functions of Z-number linguistic terms and derive ranking rules based on their interval expectation values. This thorough analysis lays the theoretical foundation for our model.

6.1.1 Z-number linguistic term set

Definition 13. Let $S = \{s_0, s_1, \dots, s_m\}$ denote a finite, totally ordered set of fuzzy linguistic terms that serve as fuzzy restrictions, referred to as the restriction fuzzy linguistic term set. Likewise, let $R = \{r_0, r_1, \dots, r_n\}$ be another finite, totally ordered set of fuzzy linguistic terms representing fuzzy

constraints on reliability, known as the reliability fuzzy linguistic term set. A Z-number linguistic term is defined as the ordered pair

$$Z_{I(i,j)} = (s_i, r_j), \quad (\text{I.23})$$

where $s_i \in S$, $r_j \in R$, and $0 \leq i \leq m$, $0 \leq j \leq n$. In this context, s_i represents a fuzzy restriction on the values that the random variable X may assume, while r_j imposes a fuzzy constraint on the reliability of s_i . The index function $I(i, j)$ is defined as shown in Eq. (I.24).

Some key properties of linguistic Z-number terms are summarized below:

- (1) **Maximality:** The most positive Z-number linguistic term is (s_m, r_n) . For example, this could be interpreted as (Very High, Very Sure).
- (2) **Minimality:** The most negative Z-number linguistic term is (s_0, r_n) . For instance, it may represent (Very Low, Very High).
- (3) **Symmetry:** Consider two Z-number linguistic terms $(s_{\frac{m}{2}-k}, r_j)$ and $(s_{\frac{m}{2}+k}, r_j)$, where $k = 0, 1, \dots, \frac{m}{2}$ and $j = 0, 1, \dots, n$. There exists a symmetrical central term $(s_{\frac{m}{2}}, r_j)$ that balances the two. For example, the statements "Expert is very sure the speed is high" and "Expert is very sure the speed is low" illustrate such symmetry.
- (4) **Monotonicity:** Let (s_{i_1}, r_{j_1}) and (s_{i_2}, r_{j_2}) be two Z-number linguistic terms. Then:
 - If $i_1 > i_2$, then $(s_{i_1}, r_{j_1}) \succ (s_{i_2}, r_{j_2})$.
 - If $i_1 < i_2$, then $(s_{i_1}, r_{j_1}) \prec (s_{i_2}, r_{j_2})$.
 - If $i_1 = i_2$ and the terms are positive (i.e., $m/2 \leq i_1, i_2 \leq m$), then:
 - If $j_1 > j_2$, then $(s_{i_1}, r_{j_1}) \succ (s_{i_2}, r_{j_2})$.
 - If $j_1 < j_2$, then $(s_{i_1}, r_{j_1}) \prec (s_{i_2}, r_{j_2})$.
 - If $i_1 = i_2$ and the terms are negative (i.e., $0 \leq i_1, i_2 < m/2$), then:
 - If $j_1 > j_2$, then $(s_{i_1}, r_{j_1}) \prec (s_{i_2}, r_{j_2})$.
 - If $j_1 < j_2$, then $(s_{i_1}, r_{j_1}) \succ (s_{i_2}, r_{j_2})$.

Definition 14. For given sets $S = \{s_0, s_1, \dots, s_m\}$ and $R = \{r_0, r_1, \dots, r_n\}$, we can construct the complete set of $(m+1) \times (n+1)$ Z-number linguistic terms:

$$\{(s_0, r_0), (s_0, r_1), \dots, (s_0, r_n), \dots, (s_m, r_{n-1}), (s_m, r_n)\}.$$

By ranking all these terms from the smallest to the largest, we obtain a finite and totally ordered set

$$\Gamma = \{Z_0, Z_1, \dots, Z_{(m+1)(n+1)-1}\},$$

which is called the Z-number linguistic term set.

From the monotonicity properties of the Z-number linguistic terms, we can infer the order index function $I(i, j)$ for a Z-number linguistic term (s_i, r_j) as follows:

$$I(i, j) = \begin{cases} i(n+1) + j, & \text{if } s_i \text{ is positive;} \\ i(n+1) + n - j, & \text{if } s_i \text{ is negative.} \end{cases} \quad (\text{I.24})$$

6.1.2 Theoretical support

Owing to the inherent differences in their definitions, the order of Z-number linguistic terms diverges from that of conventional Z-numbers. In Z-number linguistic terms, the fuzzy terms are confined to a specific interval on the real number line; in contrast, the fuzzy numbers in traditional Z-numbers are defined over the entire real line. A basic probabilistic principle is that when the probability measure for one restriction fuzzy linguistic term declines, the measures for the other terms tend to increase. This behavior permits the utilization of interval expectation values derived from hidden probability density functions to rank Z-number linguistic terms effectively. By employing these interval expectation values, we can develop a ranking system that accurately reflects the comparative significance and reliability of fuzzy linguistic terms within the context of Z-number linguistic terms.

Consider two Z-number linguistic terms, denoted by

$$Z_{\phi(1)} = (s_{i_1}, r_{j_1}) \quad \text{and} \quad Z_{\phi(2)} = (s_{i_2}, r_{j_2}),$$

where the fuzzy components are given by

$$\begin{aligned} s_{i_1} &= (a_{i_1}^1, b_{i_1}^1, c_{i_1}^1), & r_{j_1} &= (a_{j_1}^2, b_{j_1}^2, c_{j_1}^2), \\ s_{i_2} &= (a_{i_2}^1, b_{i_2}^1, c_{i_2}^1), & r_{j_2} &= (a_{j_2}^2, b_{j_2}^2, c_{j_2}^2), \end{aligned}$$

with all these fuzzy numbers assumed to be triangular. Here, the index function $\phi(\cdot)$ maps indices from 0 to $(m+1)(n+1)-1$. We assume that the underlying (hidden) probability distributions associated with these terms are normal.

For each $k = 1, 2$, the probability density function is defined as

$$p_k(x) = \frac{1}{\sqrt{2\pi} \sigma_k} \exp \left(-\frac{(x - \mu_k)^2}{2\sigma_k^2} \right).$$

Based on Eq. (I.14), the expectation of each normal distribution can be expressed as

$$\mu_k = \int_{\mathbb{R}} x p_k(x) dx = \frac{\int_{\mathbb{R}} x \mu_{s_{i_k}}(x) dx}{\int_{\mathbb{R}} \mu_{s_{i_k}}(x) dx} = b_{i_k}^1, \quad k = 1, 2.$$

The standard deviation σ_k is determined using Eq. (I.13) by the relation

$$\mu_{r_{j_k}}(prob_k) = \mu_{r_{j_k}} \left(\int_{\mathbb{R}} p_k(x) \mu_{r_{j_k}}(x) dx \right).$$

Although the expectation of $p_k(x)$ on the entire real line is $b_{i_k}^1$, when the domain is restricted to the interval $[0, 1]$ (since the restriction fuzzy linguistic term set is defined on $[0, 1]$), the expectation is recalculated as

$$E(k)_{[0,1]} = \frac{\int_0^1 x p_k(x) dx}{\int_0^1 p_k(x) dx},$$

with $\int_0^1 p_k(x) dx$ acting as the normalization factor. Based on these interval expectation values, if $E(1)_{[0,1]} > E(2)_{[0,1]}$, then we conclude that $Z_{\phi(1)} \succ Z_{\phi(2)}$; if $E(1)_{[0,1]} < E(2)_{[0,1]}$, then $Z_{\phi(1)} \prec Z_{\phi(2)}$; and if $E(1)_{[0,1]} = E(2)_{[0,1]}$, then it follows that $Z_{\phi(1)} \approx Z_{\phi(2)}$.

Based on the convexity inherent in the normal distribution, one can infer that if $s_{i_1} > s_{i_2}$, then the interval expectation of p_1 will invariably exceed that of p_2 , independent of any changes in variance. Therefore, we conclude that:

- If $i_1 > i_2$, then $(s_{i_1}, r_{j_1}) \succ (s_{i_2}, r_{j_2})$;
- If $i_1 < i_2$, then $(s_{i_1}, r_{j_1}) \prec (s_{i_2}, r_{j_2})$.

When the probability measure of s_{i_k} decreases while its membership function remains unchanged, the variance of the associated normal distribution diminishes, reducing the probability density near the mean and consequently increasing the density further from the mean. For a positive restriction fuzzy linguistic term, this effect causes the probability densities of all negative restriction fuzzy linguistic terms to rise, which in turn lowers the expectation value over the interval $[0, 1]$. Hence, we conclude that:

- If $i_1 = i_2$, with s_{i_1} and s_{i_2} being positive (i.e., $m/2 \leq i_1, i_2 \leq m$), and if $j_1 > j_2$, then $(s_{i_1}, r_{j_1}) \succ (s_{i_2}, r_{j_2})$.
- Conversely, if $i_1 = i_2$, s_{i_1} and s_{i_2} are positive, and $j_1 < j_2$, then $(s_{i_1}, r_{j_1}) \prec (s_{i_2}, r_{j_2})$.

Similarly, when the probability measure of s_{i_k} decreases without any change in its membership function, the normal distribution's variance contracts, reducing the density near the mean and boosting the density in the tail regions. If s_{i_k} represents a negative restriction fuzzy linguistic term, this leads to an increase in the probability density of all positive restriction fuzzy linguistic terms, thereby raising the expectation value over $[0, 1]$. Therefore, we deduce that:

- If $i_1 = i_2$, with s_{i_1} and s_{i_2} being negative (i.e., $0 \leq i_1, i_2 < m/2$), and if $j_1 > j_2$, then $(s_{i_1}, r_{j_1}) \prec (s_{i_2}, r_{j_2})$.
- Conversely, if $i_1 = i_2$, s_{i_1} and s_{i_2} are negative, and $j_1 < j_2$, then $(s_{i_1}, r_{j_1}) \succ (s_{i_2}, r_{j_2})$.

6.1.3 Operators

Definition 15 (Negation Operator). *We introduce a negation operator for Z-number linguistic terms, defined by*

$$\text{Neg}(Z_i) = Z_{(m+1)(n+1)-1-i}.$$

Aggregation operators are essential tools for synthesizing disparate pieces of information into a single, unified result.

Definition 16 (Arithmetic average aggregation operator). *Let $\Gamma = \{Z_0, Z_1, \dots, Z_{(m+1)(n+1)-1}\}$ be a set of Z-number linguistic terms. Consider a collection of t terms $T = \{Z_{\phi(1)}, Z_{\phi(2)}, \dots, Z_{\phi(t)}\}$, where $t > 0$ and each $Z_{\phi(i)}$ belongs to Γ for $i = 1, 2, \dots, t$; here, $\phi(\cdot)$ denotes an index mapping function. The arithmetic average, denoted by $\bar{Z}$, is then defined as*

$$\bar{Z} = Z_{\text{round}(\sum_{i=1}^t \frac{1}{t} \phi(i))}, \quad (\text{I.25})$$

with $\text{round}(\cdot)$ representing the standard rounding operation.

Definition 17 (Average weighted aggregation operator). *Let $\Gamma = \{Z_0, Z_1, \dots, Z_{(m+1)(n+1)-1}\}$ be a set of Z-number linguistic terms. Consider a collection of t terms $T = \{Z_{\phi(1)}, Z_{\phi(2)}, \dots, Z_{\phi(t)}\}$, where $t > 0$ and each $Z_{\phi(i)}$ belongs to Γ for $i = 1, 2, \dots, t$. Here, $\phi(\cdot)$ serves as the index mapping function. Let $W = \{w_1, w_2, \dots, w_t\}$ be a corresponding set of weights satisfying $w_i \in [0, 1]$ and $\sum_{i=1}^t w_i = 1$. The weighted aggregated Z-number linguistic term $\bar{Z}$ is then defined by*

$$\bar{Z} = Z_{\text{round}(\sum_{i=1}^t w_i \phi(i))}, \quad (\text{I.26})$$

where $\text{round}(\cdot)$ denotes the standard rounding operation.

Definition 18 (Ordered weighted aggregation operator). Let $\Gamma = \{Z_0, Z_1, \dots, Z_{(m+1)(n+1)-1}\}$ be a set of Z-number linguistic terms. Consider a set of t such terms $T = \{Z_{\phi(1)}, Z_{\phi(2)}, \dots, Z_{\phi(t)}\}$, with $t > 0$ and $Z_{\phi(i)} \in \Gamma$ for $i = 1, 2, \dots, t$, where $\phi(\cdot)$ is the index function. Let $W = \{w_1, w_2, \dots, w_t\}$ be the associated weight set with $w_i \in [0, 1]$ and $\sum_{i=1}^t w_i = 1$. The ordered weighted aggregated Z-number linguistic term $\bar{Z}$ is defined as

$$\bar{Z} = Z_{\text{round}(\sum_{i=1}^t w_i \phi^*(i))}, \quad (\text{I.27})$$

where $\phi^*(i)$ denotes the i -th largest value among the set of indices $\{\phi(1), \phi(2), \dots, \phi(t)\}$.

6.1.4 Multi-attribute group decision making procedures

After establishing the Z-number linguistic term set, we illustrate its versatility by integrating it into a multi-attribute group decision-making (MCGDM) framework. This newly proposed Z-number linguistic term set provides a clear and efficient decision-making approach, thereby enhancing both the simplicity and overall effectiveness of the process.

Consider an MCGDM problem that employs Z-number linguistic terms. Let $A = \{a_1, a_2, \dots, a_l\}$ denote a set of l alternatives, and let $C = \{c_1, c_2, \dots, c_h\}$ represent a collection of h criteria, associated with a weight vector $W_c = \{w_{c,1}, w_{c,2}, \dots, w_{c,h}\}$, where each $w_{c,i} \in [0, 1]$ and $\sum_{i=1}^h w_{c,i} = 1$. Suppose there are g decision makers, denoted by $D = \{d_1, d_2, \dots, d_g\}$, with an associated weight vector $W_d = \{w_{d,1}, w_{d,2}, \dots, w_{d,g}\}$, where each $w_{d,i} \in [0, 1]$ and $\sum_{i=1}^g w_{d,i} = 1$. The objective is to rank the alternatives accordingly.

To perform the process, the following procedure is adopted:

Step 1: Establish the sets of restriction fuzzy linguistic terms $S = \{s_0, s_1, \dots, s_m\}$ and reliability fuzzy linguistic terms $R = \{r_0, r_1, \dots, r_n\}$, and subsequently derive the Z-number linguistic term set Γ . The ordering index for a Z-number linguistic term (s_x, r_y) , with $x = 0, 1, \dots, m$ and $y = 0, 1, \dots, n$, is determined by

$$I(x, y) = \begin{cases} x(n+1) + y, & \text{if } s_x \text{ is positive;} \\ x(n+1) + m - y, & \text{if } s_x \text{ is negative.} \end{cases} \quad (\text{I.28})$$

Step 2: Construct the decision matrix, in terms of Z-number linguistic terms, for each decision-maker.

For example, the decision matrix for decision maker d_k is illustrated in Table I.1. In this matrix, the term $Z_{\phi(i,j,k)} = (s_{i,j,k}, r_{i,j,k})$ denotes the rating assigned by d_k for alternative a_i with respect to criterion c_j , where $s_{i,j,k} \in S$ and $r_{i,j,k} \in R$. Here, $\phi(i, j, k)$ represents the index function.

Step 3: Determine the weights for both the criteria and the decision-makers.

Suppose the importance Z-number linguistic term set is denoted by Γ' , which is defined by the restriction fuzzy linguistic term set $S' = \{s'_1, s'_2, \dots, s'_{m'}\}$ and the reliability fuzzy linguistic term set $R' = \{r'_1, r'_2, \dots, r'_{n'}\}$. Experts evaluate the significance of the criteria as:

$$C_w = [Z'_{c,\phi_c(1)}, Z'_{c,\phi_c(2)}, \dots, Z'_{c,\phi_c(j)}, \dots, Z'_{c,\phi_c(h)}], \quad (\text{I.29})$$

where $Z'_{c,\phi_c(j)} = (s'_p, r'_q)$ is the importance Z-number linguistic term for criterion c_j with $s'_p \in S'$, $r'_q \in R'$, $0 \leq p \leq m'$ and $0 \leq q \leq n'$; here, $\phi_c(j)$ is the corresponding index function.

	c_1	c_2	$\dots$	c_j	$\dots$	c_h
a_1	$Z_{\phi(1,1,k)}$	$Z_{\phi(1,2,k)}$	$\dots$	$Z_{\phi(1,j,k)}$	$\dots$	$Z_{\phi(1,h,k)}$
a_2	$Z_{\phi(2,1,k)}$	$Z_{\phi(2,2,k)}$	$\dots$	$Z_{\phi(2,j,k)}$	$\dots$	$Z_{\phi(2,h,k)}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
a_i	$Z_{\phi(i,1,k)}$	$Z_{\phi(i,2,k)}$	$\dots$	$Z_{\phi(i,j,k)}$	$\dots$	$Z_{\phi(i,h,k)}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
a_l	$Z_{\phi(l,1,k)}$	$Z_{\phi(l,2,k)}$	$\dots$	$Z_{\phi(l,j,k)}$	$\dots$	$Z_{\phi(l,h,k)}$

Table I.1: Decision matrix of decision maker d_k .

Similarly, the significance of the decision-makers is assessed as:

$$D_w = [Z'_{d,\phi_d(1)}, Z'_{d,\phi_d(2)}, \dots, Z'_{d,\phi_d(k)}, \dots, Z'_{d,\phi_d(g)}], \quad (\text{I.30})$$

where $Z'_{d,\phi_d(k)} = (s'_p, r'_q)$ represents the importance of decision maker d_k (with the same conditions on s'_p and r'_q as above) and $\phi_d(k)$ is its index function.

The weight assigned to criterion c_j is computed as:

$$w_{c,j} = \frac{\phi_c(j)}{\sum_{i=1}^h \phi_c(i)}, \quad (\text{I.31})$$

and the weight of decision maker d_k is determined by:

$$w_{d,k} = \frac{\phi_d(k)}{\sum_{i=1}^g \phi_d(i)}. \quad (\text{I.32})$$

Step 4: Aggregate the Z-number linguistic terms for each alternative using a weighted average aggregation operator.

Specifically, the aggregated Z-number linguistic term for alternative a_i is defined as:

$$\overline{Z}_i = Z_{\sum_{k=1}^g w_{d,k} \sum_{j=1}^h w_{c,j} \phi(i,j,k)}. \quad (\text{I.33})$$

Step 5: Rank the alternatives based on the order indices of their aggregated Z-number linguistic terms.

This methodology first establishes the ordering for all individual ratings, and then applies weighting and aggregation to synthesize the data. As a result, it provides a clear depiction of the relative magnitude of the alternatives. Since both fuzzy and Z-number-based information are inherently qualitative, the emphasis is placed on their ranking rather than on precise numerical values [DFPR22]. Notably, by determining the ordering of Z-number linguistic terms prior to performing any calculations, the approach substantially reduces computational complexity compared to methods that compute complex Z-number values first and then order them. Additionally, simple arithmetic on ordering indices preserves the inherent ranking of the alternatives.

The journal paper concerning this part is:

Y. Li, G. Kou, Y. Peng, and J. A. Morente-Molinera, Z-number linguistic term set for multi-criteria group decision-making and its application in predicting the acceptance of academic papers, *Applied Intelligence*, 54(21),10962-10981, 2024, <https://doi.org/10.1007/s10489-024-05765-8>.

6.2 Z-number-valued rule-based classification system

Based on the robust capability of Z-numbers to represent and reason with uncertain information that exhibits partial reliability, we propose a Z-number-valued if-then rule-based classification system (ZRBCS). This system operates in two primary stages: first, it generates Z-number-valued if-then rules, and second, it uses these rules to classify new patterns. The rule generation process is further subdivided into three steps: the creation of the first fuzzy number, the derivation of the second fuzzy number, and the determination of corresponding certainty degrees.

6.2.1 Z-number-valued if-then rule structure

When we substitute the antecedent fuzzy numbers in Eq. (I.15) with Z-numbers, the resulting structure of a Z-number-valued if-then rule is expressed as follows.

Definition 19. For a classification problem involving W classes with an M -dimensional input vector $x_i = \{x_{i1}, x_{i2}, \dots, x_{iM}\}$, a Z-number-valued if-then rule is defined by

$$\begin{aligned} \text{Rule } R_l^Z : \quad & \text{If } x_{i1} \text{ is } Z_{l1} \text{ and } x_{i2} \text{ is } Z_{l2} \text{ and } \dots \text{ and } x_{iM} \text{ is } Z_{lM}, \\ & \text{then Class } C_l \text{ with certainty degree } CF_l, \end{aligned} \quad (\text{I.34})$$

where $i = 1, 2, \dots, n$, with n being the number of patterns; $l = 1, 2, \dots, L$, where L denotes the total number of rules; $Z_{lj} = (\tilde{A}_{lj}, \tilde{B}_{lj})$ is the antecedent Z-number corresponding to the j -th feature ($j = 1, 2, \dots, M$); C_l is the consequent class of the l -th rule, chosen from the set $\{\theta_1, \theta_2, \dots, \theta_W\}$; $CF_l \in [0, 1]$ represents the certainty degree of rule R_l^Z .

6.2.2 Z-number-valued if-then rules generation

To derive Z-number-valued if-then rules from training data, three key challenges must be addressed: constructing the first antecedent fuzzy number, generating the second antecedent fuzzy number, and determining the consequent class. Similar to the approach in FRBCS, the first fuzzy number and the consequent class are obtained using the simple fuzzy grid method and the maximum compatibility grade method, respectively. Essentially, the initial step in forming Z-number-valued if-then rules is to generate standard fuzzy if-then rules.

The second phase in generating Z-number-valued if-then rules involves identifying the negative training patterns associated with each fuzzy if-then rule. In cases where a negative pattern slightly exceeds the threshold (i.e., it lies at the intersection of two adjacent fuzzy numbers), a minor adjustment of the fuzzy number is sufficient to ensure that the pattern is no longer included.

Definition 20. For a given rule R_l , the set of negative training patterns is defined as

$$E_l = \{x_e \mid \forall j = 1, 2, \dots, M, \mu_{\tilde{A}_{lj}}(x_{ej}) \geq 0.5, y_e \neq C_l\}. \quad (\text{I.35})$$

Here, a membership degree of 0.5 corresponds to the point of intersection between two adjacent fuzzy numbers.

Next, the second component of the antecedent Z-number is generated. Specifically, for rule R_l , the second fuzzy number of Z_{lj} (denoted as $\tilde{B}_{lj}$) is defined as a triangular fuzzy number:

Definition 21. For rule R_l , the second fuzzy number of Z_{lj} , $\tilde{B}_{lj}$, is given by

$$\tilde{B}_{lj} = \begin{cases} \left(\min_{x_e \in E_l} \left(1.5 - \epsilon - \mu_{A_{lj}}(x_{ej}) \right), \right. \\ \quad \left. \frac{\sum_{x_{ej} \in E_l} \left(1.5 - \epsilon - \mu_{A_{lj}}(x_{ej}) \right)}{|E_l|}, \right. \\ \quad \left. \max_{x_e \in E_l} \left(1.5 - \epsilon - \mu_{A_{lj}}(x_{ej}) \right) \right), & \text{if } |E_l| \neq 0, \\ (1, 1, 1), & \text{if } |E_l| = 0, \end{cases} \quad (\text{I.36})$$

where $j = 1, 2, \dots, M$ and $|E_l|$ denotes the cardinality of the set E_l .

The complete procedure for generating Z-number-valued if-then rules is summarized in Algorithm 1.

Algorithm 1 Z-number-valued if-then rules generation algorithm.

Input: The training dataset $D = \{(x_i, y_i) | x_i \in \mathbb{R}^M, y_i \in \{\theta_1, \theta_2, \dots, \theta_W\}\}$, the number of fuzzy partitions K .

Output: A set of Z-number-valued if-then rules R^Z and the corresponding consequent classes set C

```

1:  $N \leftarrow K^M$ 
2:  $R^F, C = \text{Fuzziness}(T, K)$  // The simple fuzzy grid if-then rules generation,  $R_l^F = \{\tilde{A}_{l1}, \tilde{A}_{l2}, \dots, \tilde{A}_{lM}\}$ ,  $C = \{C_1, C_2, \dots, C_L\}$ ,  $l = 1, 2, \dots, L$ 
3:  $E \leftarrow L * \emptyset$ 
4: for  $x_i$  in  $D$  do
5:   for  $R_l^F$  in  $R^F$  do
6:     if  $\forall j = 1, \dots, M, \mu_{A_{lj}}(x_{ij}) \geq 0.5$  and  $y_i \neq C_l$  then
7:        $x_i$  add in  $E_l$ 
8:     end if
9:   end for
10: end for
11: for  $R_l^F$  in  $R^F$  do
12:   for  $j = 1$  to  $M$  do
13:      $\tilde{B}_{lj}$  calculated by Eq. (I.60)
14:   end for
15:    $R_l^Z \leftarrow \{(A_{l1}, B_{l1}), (A_{l2}, B_{l2}), \dots, (A_{lM}, B_{lM})\}$ 
16: end for
17:
18: return  $R^Z, C$ 

```

The Z-numbers can adjust them horizontally in the ZRBCS.

Definition 22. Let a Z-number be defined as $Z = (\tilde{A}, \tilde{B})$, where $\tilde{A}$ and $\tilde{B}$ are triangular fuzzy numbers, i.e., $\tilde{A} = (a_1, a_2, a_3)$ and $\tilde{B} = (b_1, b_2, b_3)$ with $a_1, a_2, a_3, b_1, b_2, b_3 \in \mathcal{R}$. The reliability fuzzy number $\tilde{B}$ is then used to adjust $\tilde{A}$ and generate a modified fuzzy number $\tilde{A}' = (a'_1, a'_2, a'_3)$, defined by

$$\begin{aligned} a'_1 &= a_2 - (a_2 - a_1) \times \text{Rel}(\tilde{B}), \\ a'_2 &= a_2, \\ a'_3 &= a_2 + (a_3 - a_2) \times \text{Rel}(\tilde{B}), \end{aligned} \quad (\text{I.37})$$

where the reliability measure is given by $Rel(\tilde{B}) = \frac{b_1+b_2+b_3}{3}$.

Definition 23. The Z-compatibility degree of a Z-number-valued if-then rule R_l^Z for an input pattern x_i is defined by

$$\mu_l^Z(x_i) = \mu_{\tilde{A}'_{l1}}(x_{i1}) \times \mu_{\tilde{A}'_{l2}}(x_{i2}) \times \cdots \times \mu_{\tilde{A}'_{lM}}(x_{iM}), \quad (\text{I.38})$$

where $\tilde{A}'_{lj}$ denotes the adjusted fuzzy number associated with the Z-number $Z_{lj} = (\tilde{A}_{lj}, \tilde{B}_{lj})$.

Definition 24. The Z-compatibility degree of a Z-number-valued rule R_l^Z with respect to a class θ_w is defined as

$$\psi_l^Z(\theta_w) = \sum_{x_i \in \theta_w} \mu_l^Z(x_i). \quad (\text{I.39})$$

The certainty factor CF_l for rule R_l^Z is calculated by

$$CF_l = \frac{\psi_l^Z(C_l) - \bar{\psi}^Z}{\sum_{w=1}^W \psi_l^Z(\theta_w)}, \quad (\text{I.40})$$

where the average Z-compatibility is given by

$$\bar{\psi}^Z = \frac{\sum_{\theta_w \neq C_l} \psi_l^Z(\theta_w)}{W - 1}.$$

6.2.3 Procedure for classifying new patterns

For a new pattern x_{new} , the classification is determined by selecting the rule $R_{l^*}^Z$ that satisfies

$$\mu_{l^*}^Z(x_{new}) = \max \left\{ \mu_1^Z(x_{new}) \times CF_1, \mu_2^Z(x_{new}) \times CF_2, \dots, \mu_L^Z(x_{new}) \times CF_L \right\}. \quad (\text{I.41})$$

If more than one rule achieves the maximum Z-compatibility value, the pattern is regarded as unclassifiable. Otherwise, the class of x_{new} is assigned as the consequent class C_{l^*} of the selected rule $R_{l^*}^Z$, that is, $x_{new} \in C_{l^*}$.

The journal article associated to this part is:

Y. Li, E. Herrera-Viedma, I. J. Pérez, M. Barragán-Guzmán, and J. A. Morente-Molinera, Z-number-valued rule-based classification system, *Applied Soft Computing*, 137, 110168, 2023, <https://doi.org/10.1016/j.asoc.2023.110168>.

6.3 Z-number-valued rule based decision tree

The architecture of the proposed Z-number-valued rule-based decision tree (ZRDT) is identical to that of the FRDT, which is constructed as an oblique, multi-layered tree; however, in ZRDT, the traditional fuzzy rules are substituted by Z-number-valued rules.

6.3.1 Structure of the Z-number-valued if-then rule

When the fuzzy numbers in the FRDT are replaced with Z-numbers, the formulation of a Z-number-valued if-then rule is defined as follows.

Definition 25. For an input sample $x_i = \{x_{i1}, x_{i2}, \dots, x_{iM}\}$, a Z-number-valued rule R_Z involving m selected features is expressed as:

$$\begin{aligned} R_Z : \quad & \text{IF } x_{i\rho(1)} \text{ is } Z_{\rho(1)} \text{ and } x_{i\rho(2)} \text{ is } Z_{\rho(2)} \text{ and } \dots \text{ and } x_{i\rho(m)} \text{ is } Z_{\rho(m)} \\ & \text{THEN } x_i \text{ belongs to class } y_{R_Z}, \end{aligned} \quad (\text{I.42})$$

where the indices $\rho(1), \rho(2), \dots, \rho(m)$ are chosen from $\{1, 2, \dots, M\}$ to indicate the selected features; $Z_{\rho(1)}, Z_{\rho(2)}, \dots, Z_{\rho(m)}$ are the corresponding antecedent Z-numbers, each represented as $Z_{\rho(k)} = (\tilde{A}_{\rho(k)}, \tilde{B}_{\rho(k)})$ with $\tilde{A}_{\rho(k)}$ and $\tilde{B}_{\rho(k)}$ being fuzzy numbers for $k = 1, 2, \dots, m$; and $y_{R_Z} \in \Omega$ denotes the consequent class.

Definition 26. The overall membership degree of a sample x_i with respect to all antecedent Z-numbers in rule R_Z is calculated as the average of the adjusted membership values:

$$\phi_{R_Z}(x_i) = \frac{1}{m} \sum_{k=1}^m \mu_{\tilde{A}'_{\rho(k)}}(x_{i\rho(k)}), \quad (\text{I.43})$$

where $\tilde{A}'_{\rho(k)}$ represents the fuzzy number $\tilde{A}_{\rho(k)}$ after being adjusted by $\tilde{B}_{\rho(k)}$.

6.3.2 Z-number-valued rule extraction algorithm

In the FRDT framework, feature selection relies on fuzzy confidence. Specifically, a rule that employs u features is accepted only if its maximum fuzzy confidence exceeds that of a rule utilizing $u + 1$ features by a margin of $\alpha = 0.02$. Although the authors present α as a fixed parameter, variations in its value can influence the results, and the choice of 0.02 lacks a solid theoretical basis. In contrast, the proposed ZRDT algorithm selects features based on the FOIL information gain [Qui90, QCJ93], which is defined as follows.

Definition 27. The FOIL information gain measures the enhancement a rule provides compared to the default rule for the target class. It is calculated by

$$IG(R) = p_R \times \left(\log_2 \left(\frac{p_R}{p_R + n_R} \right) - \log_2 \left(\frac{p}{p + n} \right) \right), \quad (\text{I.44})$$

where p_R and n_R are the numbers of positive and negative samples that satisfy rule R , respectively, and p and n denote the numbers of positive and negative samples covered by the default rule.

Since a rule that covers more positive samples yields a higher information gain—and because this gain is independent of the number of features used—the parameter α becomes unnecessary. In other words, a rule incorporating u features is acceptable if its maximum information gain exceeds that of a rule using $u + 1$ features.

Before presenting the full Z-number-valued rule extraction algorithm, it is important to introduce the algorithm for generating the second fuzzy number. In the context of a Z-number-valued rule within the ZRDT, the initial values of the antecedent Z-numbers are computed using Eqs. (I.45)–(I.48) below.

$$a_{jw} = \begin{cases} -\infty, & w = 1, \\ s_{j\sigma(w-1)}, & \text{otherwise;} \end{cases} \quad (\text{I.45})$$

$$b_{jw} = \begin{cases} -\infty, & w = 1, \\ s_{j\sigma(w)}, & \text{otherwise;} \end{cases} \quad (\text{I.46})$$

$$c_{jw} = \begin{cases} \infty, & w = W, \\ s_{j\sigma(w)}, & \text{otherwise;} \end{cases} \quad (\text{I.47})$$

$$d_{jw} = \begin{cases} \infty, & w = W, \\ s_{j\sigma(w+1)}, & \text{otherwise.} \end{cases} \quad (\text{I.48})$$

Here, s_{jw} represents the mean value of class θ_w for the j -th feature, computed as

$$s_{jw} = \frac{\sum_{x_i \in D_{\theta_w}} x_{ij}}{|D_{\theta_w}|}. \quad (\text{I.49})$$

To form a pure leaf node in the decision tree, we have developed a Z-number-valued rule extraction algorithm, denoted by $ZREA(D, \Lambda, MaxL, \delta)$, which leverages information gain (see Algorithm 2). In this context, r_{Zu} represents the Z-number-valued rule that achieves the highest information gain when considering u features. For instance, when $u = 1$, the rule consists of a single antecedent Z-number; we compute the corresponding Z-number for every feature and then select the one with the highest information gain as the antecedent. When $u = 2$, an additional feature is selected based on the initial rule r_{Z1} to further maximize the information gain. This iterative procedure continues until adding more features no longer results in an improvement in information gain.

Algorithm 2 The Z-number-valued Rules Extraction Algorithm ($ZREA(D, \Lambda, MaxL, \delta)$).

Input: The training data for W classes: D ; The initial fuzzy numbers for the features in D : $\Lambda = \{\tilde{A}_{jw}\}$ for $j = 1, \dots, M$ and $w = 1, \dots, W$; The maximum number of features to use: $MaxL$; The threshold parameter: δ .

Output: The set of Z-number-valued rules for all classes: $R_{Z1}, \dots, R_{ZW}$.

```

1: for each class  $\theta_w$  do
2:   Set the consequent of the rule:  $y_{R_{Zw}} \leftarrow \theta_w$ 
3:   Initialize  $v \leftarrow \min(MaxL, M)$ 
4:   for  $u = 1, 2, \dots, \min(MaxL, M)$  do
5:     Compute  $r_{Zu}$  and  $r_{Zu+1}$   $\{r_{Zu}$  is the rule with the maximum information gain when using
        $u$  features, with consequent  $\theta_w\}$ 
6:     if  $IG(r_{Zu}) > IG(r_{Zu+1})$  then
7:       Set  $v \leftarrow u$ 
8:       Exit the loop
9:     end if
10:  end for
11:  Assign  $R_{Zw} \leftarrow r_{Zv}$ 
12: end for
```

In the ZRDT, additional impure nodes at each layer are constructed similarly to those in the FRDT by employing a threshold δ to collect samples whose average membership degree across all antecedent Z-numbers in a rule R_Z is below this threshold. The pseudocode for the ZRDT algorithm, denoted as $ZRDT(D, MaxL, \delta)$, is provided in Algorithm 3. Each layer of the ZRDT

is designed to reclassify the uncertain samples passed from the previous layer; during this process, new fuzzy numbers are generated based on these uncertain samples, and fresh Z-number-valued rules are extracted. Initially, the entire training dataset is treated as the set of uncertain samples.

Algorithm 3 Z-number-valued Rule-Based Decision Tree Algorithm ($ZRDT(D, MaxL, \delta)$).

Input: Training data for c classes: D ; Maximum number of features allowed: $MaxL$; Threshold parameter: δ .

Output: A series of decision tree layers, $\Gamma_l = \{R_{Zl,1}, \dots, R_{Zl,c_l}, D_l^{imp}\}$, where $l = 1, 2, \dots, L$ denotes the layer index.

- 1: Initialize $l \leftarrow 0$ and set the initial impure sample set as $D_0^{imp} \leftarrow D$.
 - 2: **while** $D_l^{imp} \neq \emptyset$ **do**
 - 3: Compute the fuzzy numbers for each feature and class in D_l^{imp} to form Λ .
 - 4: Generate Z-number-valued rules for the $(l + 1)$ -th layer: $R_{Z(l+1),1}, \dots, R_{Z(l+1),c_{l+1}} \leftarrow ZREA(D_l^{imp}, \Lambda, MaxL, \delta)$.
 - 5: **for** each sample x in D_l^{imp} **do**
 - 6: **if** for all rule $R_{Z(l+1),w}$ in $\{R_{Z(l+1),1}, \dots, R_{Z(l+1),c_{l+1}}\}$, the average membership $\phi_{R_{Z(l+1),w}}(x)$ is less than δ **then**
 - 7: Add x to the impure sample set D_{l+1}^{imp} .
 - 8: **end if**
 - 9: **end for**
 - 10: **if** $D_{l+1}^{imp} = \emptyset$ or $D_{l+1}^{imp} = D_l^{imp}$ **then**
 - 11: Terminate the loop.
 - 12: **end if**
 - 13: Increment $l \leftarrow l + 1$.
 - 14: **end while**
-

6.3.3 New samples classification

After training, the model comprises a set of Z-number-valued rules. For a new input sample $x = \{x_1, x_2, \dots, x_M\}$, denote by $R_{Zl,w}$ (with $l = 1, \dots, L$ representing the layer index and $w = 1, \dots, W_l$ indexing the rule within the l th layer) the Z-number-valued rule whose consequent is $y_{l,w}$. Such a rule is formulated as:

$R_{Zl,w}$: IF $x_{\rho_{l,w}(1)}$ is $Z_{l,w\rho_{l,w}(1)}$ and $x_{\rho_{l,w}(2)}$ is $Z_{l,w\rho_{l,w}(2)}$ and $\dots$ and $x_{\rho_{l,w}(m_{l,w})}$ is $Z_{l,w\rho_{l,w}(m_{l,w})}$
 THEN x belongs to class $y_{l,w}$,

where the indices $\rho_{l,w}(1), \dots, \rho_{l,w}(m_{l,w}) \in \{1, 2, \dots, M\}$ indicate the features considered in rule $R_{Zl,w}$; W_l is the number of rules (or classes) in layer l ; $m_{l,w}$ is the number of features employed in rule $R_{Zl,w}$; and $y_{l,w} \in \Omega$ is its consequent class.

For classifying a new sample, the process begins at the first layer ($l = 1$). The average membership degree of the sample under each rule in a layer is calculated, and the sample is assigned to the rule R_Z^* that has the highest average membership degree, provided that this degree meets or exceeds the threshold δ :

$$\phi_{R_Z^*}(x) = \max\{\phi_{R_{Zl,1}}(x), \phi_{R_{Zl,2}}(x), \dots, \phi_{R_{Zl,W_l}}(x)\}. \quad (\text{I.50})$$

If $\phi_{R_Z^*}(x) \geq \delta$, the new sample is classified into the consequent class of R_Z^* . Otherwise, the sample is passed on to the next layer ($l \leftarrow l + 1$) for further evaluation.

The journal article associated to this part is:

Y., Li, E. Herrera-Viedma, G. Kou, and J. A. Morente-Molinera, Z-number-valued rule-based decision trees, *Information Sciences*, 643, 119252, 2023, <https://doi.org/10.1016/j.ins.2023.119252>.

6.4 Z-number generation model and its application in a rule-based classification system

Zadeh, who originally introduced the concept of Z-numbers, argued that a Z-number encapsulates the probability distributions defined over a variable [Zad11]. Specifically, for a Z-number $Z = (\tilde{A}, \tilde{B})$, one can derive a collection of probability distributions $p(u)$ for the variable X by applying the corresponding possibilistic restrictions and probability measures. Conversely, it is natural to ask whether a Z-number can be employed to summarize a set of probability distributions. In this section, we introduce the Maximum Expected Minimum Entropy (MEME) model, which is designed to generate a Z-number that effectively summarizes a given set of probability distributions.

6.4.1 Maximum expected minimum entropy model

Consider a collection of probability distributions $P = \{p_1, p_2, \dots, p_m\}$ defined on a discrete variable X . Our objective is to encapsulate these distributions using a discrete Z-number $Z = (\tilde{A}, \tilde{B})$, where $\tilde{A}$ and $\tilde{B}$ are characterized by the membership functions $\mu_{\tilde{A}}(u)$ and $\mu_{\tilde{B}}(v)$, respectively. Naturally, the support (base values) of $\tilde{A}$ is identical to the set of possible values of X . The determination of the membership degrees for $\tilde{A}$ as well as the assignment of base values and corresponding membership degrees for $\tilde{B}$ is achieved through a framework that combines two distinct optimization models.

The first of these models is the *Maximum Expected Value of Reliability* model. It is formulated as follows:

$$\begin{aligned}
 & \max E[V] \\
 & \text{subject to:} \\
 & v_j = \sum_{i=1}^n p_j(u_i) \mu_{\tilde{A}}(u_i), \\
 & \mu_{\tilde{B}}(v_j) = \frac{|\{v_k = v_j \mid k = 1, 2, \dots, m\}|}{\max(|\{v_k = v_{j'} \mid k = 1, 2, \dots, m\}|)}, \\
 & 0 \leq \mu_{\tilde{A}}(u_i) \leq 1, \\
 & 0 \leq v_j \leq 1, \\
 & 0 \leq \mu_{\tilde{B}}(v_j) \leq 1, \\
 & i = 1, 2, \dots, n, \quad j = 1, 2, \dots, m.
 \end{aligned} \tag{I.51}$$

Here, $E[V]$ represents the expected reliability value associated with $\tilde{B}$, and is computed as

$$E[V] = \sum_{j=1}^m v_j w_j, \tag{I.52}$$

where the variable V denotes the discrete reliability values, and the weight w_j is defined by

$$w_j = \frac{1}{2} \left(\max_{k=1}^j \mu_{\tilde{B}}(v_k) - \max \left(\max_{k=1}^{j-1} \mu_{\tilde{B}}(v_k), 0 \right) \right) + \frac{1}{2} \left(\max_{k=j}^m \mu_{\tilde{B}}(v_k) - \max \left(\max_{k=j+1}^m \mu_{\tilde{B}}(v_k), 0 \right) \right). \quad (\text{I.53})$$

It follows that $E[V] \in [0, 1]$.

The first constraint guarantees that $\tilde{A}$ adheres to the probability measure condition as dictated by the definition of Z-numbers. Meanwhile, the second constraint sets the membership degree of each v_j based on its frequency of occurrence. The remaining constraints are standard conditions imposed on Z-numbers.

However, relying solely on the optimization model described earlier does not yield a satisfactory Z-number. This is because that model produces a unique solution independent of the input probability distributions, which causes all membership degrees to become one and thus forces the expected reliability $E[V]$ to its maximum value of 1. As indicated in Eq. (I.52), any increase in each v_j directly raises $E[V]$. Moreover, the first constraint implies that increasing $\mu_{\tilde{A}}(u_i)$ leads to a higher v_j . In fact, when $\mu_{\tilde{A}}(u_i) = 1$ for all $i = 1, 2, \dots, n$, it follows that $v_j = \sum_{i=1}^n p_j(u_i) = 1$ for every $j = 1, 2, \dots, m$, which results in $E[V] = 1$.

Since the fuzzy number $\tilde{A}$ acts as a possibilistic constraint on the variable X , reducing its uncertainty is advantageous for accurately determining X 's values. To address this, we introduce an additional optimization model that minimizes the entropy of $\tilde{A}$. This second model, termed the *Minimum Entropy* model, is formulated as:

$$\begin{aligned} & \min U(\mu_{\tilde{A}}) \\ & \text{subject to: } 0 \leq \mu_{\tilde{A}}(u_i) \leq 1, \quad i = 1, 2, \dots, n, \end{aligned} \quad (\text{I.54})$$

where $U(\mu_{\tilde{A}})$ quantifies the possibilistic uncertainty (entropy) of $\tilde{A}$ and is defined by

$$U(\mu_{\tilde{A}}) = \frac{1}{l_r} \sum_{k=1}^{r-1} (l_{k+1} - l_k) \log_2 \left(\left| c(\mu_{\tilde{A}}, l_{k+1}) \right| \right), \quad (\text{I.55})$$

with

$$c(\mu_{\tilde{A}}, l) = \{ u \in U \mid \mu_{\tilde{A}}(u) \geq l \},$$

and

$$L_{\mu_{\tilde{A}}} = \{ l_1 = 0, \dots, l_r \mid l_k = \mu_{\tilde{A}}(u), l_k \leq l_{k+1} \}$$

being the level set of $\mu_{\tilde{A}}$.

By integrating Model (I.51) and Model (I.54), we obtain the MEME model:

$$\begin{aligned}
& \min \frac{U(\mu_{\tilde{A}})/\log_2(n) + 0.1}{E[V] + 0.1} \\
\text{subject to: } & v_j = \sum_{i=1}^n p_j(u_i) \mu_{\tilde{A}}(u_i), \\
& \mu_{\tilde{B}}(v_j) = \frac{|\{v_k = v_j \mid k = 1, 2, \dots, m\}|}{\max(|\{v_k = v_{j'} \mid k = 1, 2, \dots, m\}|)}, \\
& 0 \leq \mu_{\tilde{A}}(u_i) \leq 1, \\
& 0 \leq v_j \leq 1, \\
& 0 \leq \mu_{\tilde{B}}(v_j) \leq 1, \\
& i = 1, 2, \dots, n, \quad j = 1, 2, \dots, m,
\end{aligned} \tag{I.56}$$

where the constant 0.1 is introduced to prevent division by zero. Given that $E[V] \in [0, 1]$ and $U(\mu_{\tilde{A}}) \in [0, \log_2(n)]$, the factor $\log_2(n)$ effectively balances these two terms. Overall, the MEME model is designed to minimize the uncertainty of $\tilde{A}$ while simultaneously maximizing the expected reliability.

6.4.2 Z-valuation rule-based classification system

A novel type of rule, called a Z-valuation rule, is defined as follows:

$$\text{Rule } R_l^Z : \quad \text{IF } x_{i1} \text{ is } \tilde{A}_{l1} \text{ and } \dots \text{ and } x_{im} \text{ is } \tilde{A}_{lm} \quad \text{THEN } (y_l, \tilde{A}_l, \tilde{B}_l). \tag{I.57}$$

In this formulation, each $\tilde{A}_{lj}$ (for $j = 1, \dots, m$) represents an antecedent fuzzy set, while the consequent is given as the Z-valuation $(y_l, \tilde{A}_l, \tilde{B}_l)$. Here, y_l indicates the class label assigned by rule R_l^Z and $(\tilde{A}_l, \tilde{B}_l)$ constitutes a discrete Z-number.

When fuzzy rules are derived using the FURIA approach, the corresponding Z-valuation rules take on a slightly modified structure:

$$\text{Rule } R_l^Z : \quad \text{IF } x_{i\rho(1)} \text{ is } \tilde{A}_{l\rho(1)} \text{ and } \dots \text{ and } x_{i\rho(d_l)} \text{ is } \tilde{A}_{l\rho(d_l)} \quad \text{THEN } (y_l, \tilde{A}_l, \tilde{B}_l). \tag{I.58}$$

In this case, each antecedent $\tilde{A}_{l\rho(j)}$ (for $j = 1, \dots, d_l$) is a trapezoidal fuzzy set, with $d_l \leq m$ indicating the number of features involved in the rule. The index function $\rho(j)$ specifies the feature corresponding to the j -th antecedent fuzzy set. The rule's consequent remains the Z-valuation $(y_l, \tilde{A}_l, \tilde{B}_l)$, where y_l is the designated class label and $(\tilde{A}_l, \tilde{B}_l)$ represents a generalized discrete Z-number.

During the training phase, a collection of Z-valuation rules is generated and saved in the knowledge base. Leveraging the balance between accuracy and complexity achieved by FURIA, our method for extracting Z-valuation rules combines FURIA with the MEME model. In this hybrid approach, FURIA is used to derive a compact set of fuzzy rules, while the MEME model is responsible for computing the corresponding consequent Z-valuations.

For the l -th fuzzy rule R_l produced by FURIA (refer to Eq. (I.22)), every training sample that satisfies this rule is treated as representing an underlying probability distribution over the class

labels. Since the actual class labels for the training samples are known, we define the probability distribution for a sample x_i as follows:

$$p_i(\theta_w) = \begin{cases} 1, & \text{if } x_i \text{ belongs to } \theta_w, \\ 0, & \text{otherwise.} \end{cases} \quad (\text{I.59})$$

Thus, there are W distinct probability distributions. This framework allows us to obtain the Z-valuation $(y_l, \tilde{A}_l, \tilde{B}_l)$ using the MEME model. According to Model (I.56), the first constraint is reformulated as:

$$v_i = \sum_{w=1}^W p_i(\theta_w) \mu_{\tilde{A}_l}(\theta_w) = \mu_{\tilde{A}_l}(\theta_w), \quad \text{if } x_i \text{ belongs to } \theta_w. \quad (\text{I.60})$$

Consequently, there exist W distinct values for v , denoted as $\{v_{\theta_1}, v_{\theta_2}, \dots, v_{\theta_W}\}$, where each v_{θ_w} is equal to $\mu_{\tilde{A}_l}(\theta_w)$.

A key distinction from Model (I.56) is the approach used to determine the membership degrees for $\tilde{B}_l$. In this framework, the rule's coverage degree is integrated into the second constraint, which is formulated as:

$$\mu_{\tilde{B}_l}(v_{\theta_w}) = \frac{\sum_{x_i \in D_{\theta_w}} \phi_l(x_i)}{\max_{\theta'_w \in \Theta} \left(\sum_{x_i \in D_{\theta'_w}} \phi_l(x_i) \right)}, \quad (\text{I.61})$$

where $D_{\theta_w} = \{x_i \in D \mid x_i \text{ belongs to } \theta_w\}$ denotes the set of samples in T that are classified under θ_w , and $\phi_l(x_i)$ represents the coverage degree of sample x_i by rule R_l . This coverage degree is computed as:

$$\phi_l(x_i) = \prod_{j=1}^d \mu_{\tilde{A}_{l\rho(j)}}(x_{i\rho(j)}). \quad (\text{I.62})$$

In summary, the Z-valuation $(y_l, \tilde{A}_l, \tilde{B}_l)$ is obtained by solving the following MEME model:

$$\begin{aligned} & \min \frac{U(\mu_{\tilde{A}_l})/\log_2(W) + 0.1}{E[V] + 0.1} \\ \text{subject to: } & v_{\theta_w} = \mu_{\tilde{A}_l}(\theta_w), \\ & \mu_{\tilde{B}_l}(v_{\theta_w}) = \frac{\sum_{x_i \in D_{\theta_w}} \phi_l(x_i)}{\max_{\theta_w \in \Theta} \left(\sum_{x_i \in D_{\theta_w}} \phi_l(x_i) \right)}, \\ & 0 \leq \mu_{\tilde{A}_l}(\theta_w) \leq 1, \\ & 0 \leq \mu_{\tilde{B}_l}(v_{\theta_w}) \leq 1, \\ & w = 1, 2, \dots, W. \end{aligned} \quad (\text{I.63})$$

Thus, after optimizing using the MEME model, each fuzzy rule is converted into a Z-valuation rule. The process for converting fuzzy rules into Z-valuation rules via the DE algorithm is detailed in Algorithm 4.

The journal article associated to this part is:

Y. Li, J. A. Morente-Molinera, J. R. Trillo, and E. Herrera-Viedma, Z-number generation model and its application in a rule-based classification system, *IEEE Transactions on Cybernetics*, 2024, DOI: 10.1109/TCYB.2025.3545195.

Algorithm 4 Transformation of Fuzzy Rules into Z-valuation Rules

Input: The training dataset $D = \{x_1, x_2, \dots, x_n\}$; the fuzzy ruleset $RS = \{R_1, R_2, \dots, R_L\}$; the number of individuals N ; the mutation probability F ; the crossover probability CR ; and the maximum number of iterations $MaxItera$.

Output: The resulting Z-valuation ruleset $RS^Z = \{R_1^Z, R_2^Z, \dots, R_L^Z\}$.

- 1: **for** each rule $R_l \in RS$ **do**
- 2: Compute $\mu_{\tilde{B}_l}(v_{\theta_w}) = \frac{\sum_{x_i \in D_{\theta_w}} \phi_l(x_i)}{\max_{\theta_w \in \Theta} \left(\sum_{x_i \in D_{\theta_w}} \phi_l(x_i) \right)}$ for every $\theta_w \in \Theta$.
- 3: Initialize N individuals with candidate values $\{\mu_{\tilde{A}_l}(\theta_w) \mid \mu_{\tilde{A}_l}(\theta_w) \in [0, 1], \theta_w \in \Theta\}$.
- 4: Evaluate the fitness of these individuals.
- 5: Set iteration counter $itera \leftarrow 1$.
- 6: **while** $itera \leq MaxItera$ **do**
- 7: Perform mutation.
- 8: Execute crossover.
- 9: Conduct selection.
- 10: Increment $itera \leftarrow itera + 1$.
- 11: **end while**
- 12: Choose the best individual as the consequent Z-valuation for R_l , that is, $(y_l, \tilde{A}_l, \tilde{B}_l)$.
- 13: Form the Z-valuation rule R_l^Z by combining the antecedents of R_l with the consequent Z-valuation $(y_l, \tilde{A}_l, \tilde{B}_l)$.
- 14: Add R_l^Z to the Z-valuation ruleset RS^Z .
- 15: **end for**

6.5 Z-mixture-number if-then rule induction algorithm for classification tasks

In this section, we introduce a novel if-then rule called the Z-mixture-number rule. Unlike conventional Z-number rules that are either completely discrete or entirely continuous, the antecedents in these rules are hybrid—specifically, the first fuzzy number is defined continuously, while the second is represented discretely. This hybridization allows the Z-mixture-number rules to effectively manage both continuous and discrete sources of uncertainty. Next, we propose the Z-Mixture-number Rule Induction Algorithm (ZMRIA), which generates a set of such rules for each class label. In ZMRIA, the selection of antecedents and features is based on maximizing the difference between positive and negative Z-mixture-numbers. Finally, these rules are used to infer the class labels of new samples with unknown classifications.

6.5.1 Definitions

In contrast to the previously discussed Z-number-valued rules, the antecedents in Z-mixture-number rules are represented by Z-mixture-numbers. In these rules, the first fuzzy component is expressed as a trapezoidal fuzzy number, while the second fuzzy component is discrete. The l th Z-mixture-number rule for classification is formulated as follows:

$$\text{Rule } R_l^Z : \quad \text{IF } x_{i\rho(1)} \text{ is } Z_{l\rho(1)} \text{ and } \dots \text{ and } x_{i\rho(m')} \text{ is } Z_{l\rho(m')} \quad \text{THEN } x_i \text{ is classified as } C_l. \quad (\text{I.64})$$

Here, the input vector is given by $x_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}$. Each antecedent Z-mixture-number $Z_{l\rho(j)} = (\tilde{A}_{l\rho(j)}, \tilde{B}_{l\rho(j)})$ consists of a trapezoidal fuzzy number $\tilde{A}_{l\rho(j)} = (a_{l\rho(j)}, a_{l\rho(j)}, b_{l\rho(j)}, \bar{b}_{l\rho(j)})$ and a discrete fuzzy number $\tilde{B}_{l\rho(j)}$. The index function $\rho(j)$ (with $1 \leq j \leq m' \leq m$) indicates the

j th selected feature for the rule, y_i denotes the output variable, and $C_l \in \Omega$ is the consequent class label.

6.5.2 Rule learning

When learning a Z-mixture-number rule, the first step is to select the consequent class C_l from Ω . Additionally, each feature value x_{ij} in the training set $\{x_{i1}, x_{i2}, \dots, x_{im}\}$ is treated as a probability distribution defined by:

$$p_{ij}(u) = \begin{cases} 1, & \text{if } u = x_{ij}, \\ 0, & \text{otherwise.} \end{cases} \quad (\text{I.65})$$

Assume that for the j th feature the initial fuzzy number is $\tilde{A}_{lj}$. According to the definition of Z-numbers, the probability measure of $\tilde{A}_{lj}$ can be computed by

$$P_i(\tilde{A}_{lj}) = \sum_{u \in \{x_{1j}, \dots, x_{nj}\}} \mu_{\tilde{A}_{lj}}(u) p_{ij}(u) = \mu_{\tilde{A}_{lj}}(x_{ij}). \quad (\text{I.66})$$

If a sample x_i is labeled as C_l , it is considered a positive sample; otherwise, it is treated as a negative sample.

Definition 28. A positive Z-mixture-number $Z_{C_l}^{lj} = (\tilde{A}_{C_l}^{lj}, \tilde{B}_{C_l}^{lj})$ based on the proportion of positive samples is defined by:

$$\tilde{A}_{C_l}^{lj} = \tilde{A}_{lj}, \quad \mu_{\tilde{B}_{C_l}^{lj}}(v) = \frac{|\{x_i \mid x_i \in D_{C_l} \text{ and } P_i(\tilde{A}_{lj}) = v\}|}{|D_{C_l}|}, \quad (\text{I.67})$$

where $D_{C_l} = \{x_i \in D \mid x_i \text{ belongs to } C_l\}$ denotes the set of all samples in D with class label C_l , and $|\cdot|$ denotes the cardinality of a set.

Definition 29. For every other class label $\theta_w \in \Omega$ with $\theta_w \neq C_l$, a negative Z-mixture-number $Z_{\theta_w}^{lj} = (\tilde{A}_{\theta_w}^{lj}, \tilde{B}_{\theta_w}^{lj})$ is computed using the following formulas:

$$\tilde{A}_{\theta_w}^{lj} = \tilde{A}_{lj}, \quad \mu_{\tilde{B}_{\theta_w}^{lj}}(v) = \frac{|\{x_i \mid x_i \in D_{\theta_w} \text{ and } P_i(\tilde{A}_{lj}) = v\}|}{|D_{\theta_w}|}, \quad (\text{I.68})$$

where $D_{\theta_w} = \{x_i \in D \mid x_i \text{ belongs to } \theta_w\}$ is the set of samples in D with class label θ_w .

The resulting second fuzzy number may not exhibit convexity. In practice, however, non-convex fuzzy numbers can more effectively capture the actual data distribution, thereby enhancing classification accuracy.

For a classification rule, the primary objective is to select positive samples while filtering out negatives. Consequently, a greater disparity between the positive and negative Z-mixture-numbers is desirable. As observed from Eqs. (I.67) and (I.68), the first fuzzy number remains identical in both cases. Therefore, we assess the difference between positive and negative Z-mixture-numbers using the expectation of their second fuzzy numbers.

Definition 30. The positive-negative Z-mixture-number difference is defined as follows:

$$\begin{aligned} Diff(Z_{C_l}^{lj}) &= \sum_{v \in \mathcal{R}} v \mu_{\tilde{B}_{C_l}^{lj}}(v) \\ &\quad - \sum_{\theta_w \in \Omega, \theta_w \neq C_l} \frac{|D_{\theta_w}|}{|D| - |D_{C_l}|} \sum_{v \in \mathcal{R}} v \mu_{\tilde{B}_{\theta_w}^{lj}}(v), \end{aligned} \quad (\text{I.69})$$

where $\frac{|D_{\theta_w}|}{|D|-|D_{C_l}|}$ is the weight assigned to the negative Z-mixture-number $Z_{\theta_w}^{lj}$.

A central challenge in rule learning is to identify a fuzzy number $\tilde{A}_{lj}$ that maximizes the difference between the positive and negative Z-mixture-numbers. This problem can be formulated as an optimization task, as described below:

$$\begin{aligned}
& \max \quad \text{Diff}(Z_{C_l}^{lj}) \\
& \text{subject to: } \tilde{A}^{lj} = (\underline{a}_{lj}, a_{lj}, b_{lj}, \bar{b}_{lj}), \\
& \quad \underline{a}_{lj} \leq a_{lj} \leq b_{lj} \leq \bar{b}_{lj}, \\
& \quad \tilde{A}_{C_l}^{lj} = \tilde{A}_{lj}, \\
& \quad \mu_{\tilde{B}_{C_l}^{lj}}(v) = \frac{|\{x_i \mid x_i \in D_{C_l}, P_i(\tilde{A}_{lj}) = v\}|}{|D_{C_l}|}, \\
& \quad \tilde{A}_{\theta_w}^{lj} = \tilde{A}_{lj}, \quad \forall \theta_w \in \Omega, \theta_w \neq C_l, \\
& \quad \mu_{\tilde{B}_{\theta_w}^{lj}}(v) = \frac{|\{x_i \mid x_i \in D_{\theta_w}, P_i(\tilde{A}_{lj}) = v\}|}{|D_{\theta_w}|}.
\end{aligned} \tag{I.70}$$

Here, the objective is to maximize $\text{Diff}(Z_{C_l}^{lj})$ subject to the constraints that define the fuzzy number $\tilde{A}^{lj}$ and ensure consistency between the positive and negative Z-mixture-numbers. The first constraint requires that $\tilde{A}^{lj}$ be represented as $(\underline{a}_{lj}, a_{lj}, b_{lj}, \bar{b}_{lj})$ with the proper ordering of its components. The subsequent constraints ensure that the positive Z-mixture-number (associated with class C_l) and the negative Z-mixture-numbers (associated with all other classes $\theta_w \neq C_l$) share the same first fuzzy number $\tilde{A}_{lj}$, while their second fuzzy numbers are determined by the corresponding membership degrees calculated from the training data.

To enhance both precision and clarity, each Z-mixture-number serving as an antecedent is chosen from the complete set of feature-specific Z-mixture-numbers based on the maximum difference between positive and negative values observed among the samples already covered by preceding antecedents. Algorithm 5 exemplifies the rule learning process.

In Algorithm 5, the segment from lines 7 to 12 is devoted to identifying the feature that exhibits the greatest difference between positive and negative Z-mixture-numbers. In particular, the function $\text{findZnumber}(D^c, C_l, f_j)$ searches for the Z-mixture-number associated with feature f_j in dataset D^c (with C_l as the target class) that maximizes this difference. This function resolves the optimization problem described in Model (I.70), with the overall process further detailed in Algorithm 6. Moreover, lines 5 to 18 outline the procedure for constructing the antecedent portion of a rule—beginning with an empty rule, antecedents are iteratively appended until the termination condition is met. Additionally, lines 3 to 20 illustrate the generation of multiple rules with θ_w as the consequent; if the remaining dataset contains a number of samples from class θ_w above a certain threshold, a subsequent rule with θ_w as its consequent will be produced.

To derive the Z-mixture-number $Z = (\tilde{A}, \tilde{B})$, where $\tilde{A} = (\underline{a}, a, b, \bar{b})$ and $\tilde{B}$ is obtained via Eq. (I.67), Algorithm 6 is partitioned into two phases. In the first phase (lines 2–17), the values of a and b within $\tilde{A}$ are determined. The second phase (lines 18–35) focuses on estimating $\underline{a}$ and $\bar{b}$. Both phases are designed to maximize the computed positive-negative difference.

Algorithm 5 Algorithm for learning rules using Z-mixture-numbers.

Input: The training dataset $D = \{(x_1, y_1), \dots, (x_n, y_n)\}$;

Output: Z-mixture-number rule set $RS^Z = \{R_1^Z, \dots, R_L^Z\}$, where the antecedent set of rule R_l^Z is

$\{Z_{l\rho(0)}, \dots, Z_{l\rho(m')}\}$

```

1: for Each class label  $\theta_w \in \Omega$  do
2:    $l \leftarrow 0, D^{nc} \leftarrow D$ 
3:   while  $|D_{\theta_w}^{nc}| \geq 2$  do
4:      $l \leftarrow l + 1, D^c \leftarrow D^{nc}, C_l \leftarrow \theta_w, t \leftarrow 0$ 
5:     while  $|D^c| > 0$  and  $Acc < 1$  do
6:        $t \leftarrow t + 1, maxDiff \leftarrow 0$ 
7:       for Each feature  $f_j \in F$  do
8:          $Diff(Z_{C_l}^{lj}), Z_{C_l}^{lj} \leftarrow findZnumber(D^c, C_l, f_j)$ 
9:         if  $Diff(Z_{C_l}^{lj}) > maxDiff$  then
10:           $maxDiff = Diff(Z_{C_l}^{lj}), Z_{l\rho(t)} = Z_{C_l}^{lj}, \rho(t) = j$ 
11:        end if
12:      end for
13:       $D_n^c \leftarrow \{x_i \mid x_i \in D^c, \mu_{\tilde{A}_{l\rho(t)}}(x_{i\rho(t)}) > 0\}$ 
14:      if  $|D_n^c| = |D^c|$  then
15:        break
16:      end if
17:       $D^c \leftarrow D_n^c, Acc \leftarrow \frac{|D_{\theta_w}^c|}{|D^c|}$ 
18:    end while
19:     $D^{nc} \leftarrow D^{nc} - D^c$ 
20:  end while
21: end for

```

Algorithm 6 $findZnumber(D, C_l, f_j)$ **Input:** A training set $D = \{(x_1, y_1), \dots, (x_n, y_n)\}$, a designated class label C_l , and a feature f_j .**Output:** A Z-mixture-number $Z = (\tilde{A}, \tilde{B})$ along with its associated positive-negative difference $Diff(Z)$.

- 1: Initialize $\underline{a}$ and a with $-\infty$, and set b and $\bar{b}$ to $+\infty$; then define $\tilde{A} \leftarrow (\underline{a}, a, b, \bar{b})$.
- 2: Define $D' \leftarrow \{(x'_i, y'_i) \mid (x'_i, y'_i) \in D, x'_{i-1,j} \leq x'_{i,j} \leq x'_{i+1,j}\}$.
- 3: Set $maxDiff \leftarrow 0$ and $num \leftarrow 0$.
- 4: **for** each $x'_i \in D'$ **do**
- 5: Assign $p \leftarrow x'_{i,j}$ and initialize $DiffLR \leftarrow 0$.
- 6: Calculate $Diff_L(Z_L)$ with $\tilde{A}_L = (-\infty, -\infty, p, p)$ using Eq. (I.69) on D' .
- 7: Calculate $Diff_R(Z_R)$ with $\tilde{A}_R = (p, p, +\infty, +\infty)$ using Eq. (I.69) on D' .
- 8: Let $DiffLR \leftarrow \max(Diff_L(Z_L), Diff_R(Z_R))$.
- 9: **if** $DiffLR > maxDiff$ **then**
- 10: Update $maxDiff \leftarrow DiffLR$.
- 11: **if** $Diff_L(Z_L) > Diff_R(Z_R)$ **then**
- 12: Set $DiffLR \leftarrow Diff_L(Z_L)$ and update $b \leftarrow p$.
- 13: **else**
- 14: Set $DiffLR \leftarrow Diff_R(Z_R)$ and update $a \leftarrow p$.
- 15: **end if**
- 16: **end if**
- 17: **end for**
- 18: **if** $b \neq +\infty$ **then**
- 19: **for** each $x'_i \in D'$ with $x'_{i,j} > b$ **do**
- 20: Compute $Diff1(Z_1)$ for $\tilde{A}_1 = (\underline{a}, a, b, x'_{i,j})$ using Eq. (I.69) on D' .
- 21: **if** $Diff1(Z_1) > maxDiff$ **then**
- 22: Set $maxDiff \leftarrow Diff1(Z_1)$ and update $\bar{b} \leftarrow x'_{i,j}$.
- 23: Assign $Z \leftarrow Z_1$ and $Diff(Z) \leftarrow maxDiff$.
- 24: **end if**
- 25: **end for**
- 26: **end if**
- 27: **if** $a \neq -\infty$ **then**
- 28: **for** each $x'_i \in D'$ with $x'_{i,j} < a$ **do**
- 29: Compute $Diff2(Z_2)$ for $\tilde{A}_2 = (x'_{i,j}, a, b, \bar{b})$ using Eq. (I.69) on D' .
- 30: **if** $Diff2(Z_2) > maxDiff$ **then**
- 31: Update $maxDiff \leftarrow Diff2(Z_2)$ and set $\underline{a} \leftarrow x'_{i,j}$.
- 32: Assign $Z \leftarrow Z_2$ and $Diff(Z) \leftarrow maxDiff$.
- 33: **end if**
- 34: **end for**
- 35: **end if**

6.5.3 Inference engine

Assume that a total of L Z-mixture-number rules have been obtained and stored in the knowledge base. We denote this collection by

$$RS^Z = \{R_1^Z, R_2^Z, \dots, R_L^Z\}.$$

Definition 31. For a given sample $x_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}$, the coverage provided by rule R_l^Z (with $l = 1, 2, \dots, L$) is determined by

$$\varphi_{R_l^Z}(x_i) = \prod_{j=1}^{m'} \mu_{\tilde{A}_{l\rho(j)}}(x_{i\rho(j)}). \quad (\text{I.71})$$

In this framework, only the first fuzzy number of the Z-mixture-number is employed, while the second one is disregarded. This choice is motivated by the fact that the first fuzzy number is designed to optimize the second one, which is already in its optimal state. Thus, the first fuzzy number sufficiently represents the entire Z-mixture-number.

Definition 32. The certainty factor associated with rule R_l^Z is defined as [HH09]:

$$CF(R_l^Z) = \frac{2|D_{C_j}|/|D| + \sum_{x_i \in D_{C_j}} \varphi_{R_l^Z}(x_i)}{2 + \sum_{x_i \in D_{C_j}} \varphi_{R_l^Z}(x_i)}. \quad (\text{I.72})$$

For a new query sample $\dot{x} = \{\dot{x}_1, \dot{x}_2, \dots, \dot{x}_m\}$, its class label $\dot{y}$ is predicted according to

$$\dot{y} = \arg \max_{\theta_w \in \Omega} \left(\sum_{C_l = \theta_w} \varphi_{R_l^Z}(\dot{x}) CF(R_l^Z) \right). \quad (\text{I.73})$$

In situations where a new query sample is not covered by any existing rule, we adopt the rule stretching technique proposed in [HH09]. The essential idea is that the sequence of antecedents reflects their importance. For a Z-mixture-number rule R_l^Z , let the ordered set of antecedents be

$$\{Z_{l\rho(1)}, Z_{l\rho(2)}, \dots, Z_{l\rho(m')}\}.$$

If the $(j+1)$ th antecedent is the first that fails to cover the sample (implying that the antecedents $Z_{l\rho(1)}, Z_{l\rho(2)}, \dots, Z_{l\rho(j)}$ do cover it), then these antecedents combined with the rule's consequent form a new rule, denoted by $\hat{R}_l^Z$. Its certainty factor is given by

$$CF(\hat{R}_l^Z) = \frac{j}{m'} CF(R_l^Z), \quad (\text{I.74})$$

and the predicted class label for the new sample is determined by

$$\dot{y} = \arg \max_{\theta_w \in \Omega} \sum_{C_l = \theta_w} \varphi_{\hat{R}_l^Z}(\dot{x}) CF(\hat{R}_l^Z). \quad (\text{I.75})$$

The journal article associated to this part is:

Y. Li, I. J. Pérez, F. J. Cabrerizo, and J. A. Morente-Molinera, Z-mixture-number if-then rules induction algorithm for classification tasks, submitted to IEEE Transactions on Systems, Man, and Cybernetics, 2024.

7 Discussion of results

This section primarily discusses the results obtained from all the aforementioned stages of this thesis.

7.1 Z-number linguistic model

The concept of Z-numbers provides a robust framework for representing real-world data. Nonetheless, their practical use is frequently constrained by significant computational demands. In this study, we propose a novel Z-number linguistic model that integrates both Z-number linguistic terms and their corresponding term sets within a dedicated computational framework, thereby alleviating the computational burden. Additionally, an MCGDM model that employs these Z-number linguistic terms is introduced. To demonstrate the efficacy and advantages of the proposed approach, we perform an extensive comparison with current Z-number-based MCGDM techniques and standard machine learning clustering algorithms, using the ACL 2017 dataset.

Compared with other Z-number MCGDM approaches, the Z-number linguistic model provides several distinct benefits.

- (1) In decision-making scenarios, rather than relying on methods that depend on intricate hidden probability distributions or convert Z-numbers into fuzzy numbers, our approach simplifies the process by focusing on the ordering information of Z-number linguistic terms. Rather than computing exact Z-number values, it emphasizes establishing ranking relationships among alternatives. By arranging the Z-number linguistic terms and analyzing their order, subsequent operations—such as aggregation—become more straightforward, offering clear insights into the relative standings of alternatives.
- (2) Based on solid Z-number theory, our method is more theoretically robust and achieves excellent accuracy in practical applications.
- (3) The proposed Z-number linguistic model eliminates the need to define the explicit semantics of linguistic terms by concentrating solely on their ranking relations. This feature has the potential to substantially broaden the practical use of Z-numbers.

Compared to conventional machine learning algorithms, the Z-number linguistic model offers additional advantages.

- (1) By utilizing Z-number linguistic terms, the model provides a more intuitive and human-aligned approach to data processing, making it especially effective for natural language processing tasks and for capturing the fuzzy aspects of human language.
- (2) Since the Z-number model explicitly accounts for both the uncertainty and reliability of variables during its formulation, it generally affords superior interpretability compared to many traditional black-box machine learning models.

7.2 Z-number-valued rule-based classification system

In this part, we extend traditional fuzzy if-then rules by embedding them within a Z-number-valued framework, thereby formulating Z-number-valued if-then rules. Based on this novel formulation, a

foundational rule-based classification system is constructed. This system capitalizes on the negative patterns in conventional fuzzy rule-based methods.

The first fuzzy numbers that constitute the antecedent Z-numbers are derived via the classical fuzzy rule extraction technique known as the fuzzy simple grid method, while the second fuzzy numbers capture information regarding negative patterns—such as maximum, average, and minimum values. During classification, these second fuzzy numbers modulate the first ones according to the identified negative patterns. Two experiments using several public datasets have been carried out to demonstrate both the learning and predictive capabilities of our classification system.

Relative to conventional fuzzy rule-based systems, the Z-number-valued rule-based system offers several advantages:

- (1) It not only manages data uncertainty but also quantifies the reliability of information. By simultaneously addressing these aspects, the system becomes more adept at processing ambiguous data and provides more robust decision support.
- (2) By integrating both the fuzziness and reliability of each rule, the system allows for flexible rule adjustments tailored to various scenarios during the formation of classification rules. This adaptability generally leads to enhanced classification accuracy, particularly when the input data is uncertain or varies in quality.
- (3) Since the system encapsulates both uncertainty and reliability within its rules, it achieves improved interpretability, enabling users to more clearly understand the underlying decision-making process.

7.3 Z-number-valued rule-based decision tree

In this part, we have developed and evaluated a novel ZRDT that advances the integration of Z-numbers within a FRDT framework. A key innovation of our approach is the use of negative samples to derive the second fuzzy number of Z-numbers. By leveraging insights from these negative samples, the first fuzzy number is dynamically adjusted to better conform to the training data, thereby overcoming the inflexibility seen in earlier models.

In addition, we enhance the ZRDT model by substituting fuzzy confidence with information gain during the feature selection process for each rule. This change not only streamlines the model by reducing the number of parameters required in the FRDT framework, but it also refines the selection of features to ensure that only the most pertinent ones are employed for each rule. Such methodological improvements contribute significantly to the overall efficiency and effectiveness of the decision-making process within the tree.

Our empirical evaluation of ZRDT across thirteen diverse datasets demonstrates its superior classification performance compared to the traditional FRDT—tested under three different parameter configurations. Moreover, the ZRDT model outperforms several established decision tree methods, including PUBLIC, C4.5, and AdaBoost.NC, underscoring its robustness and adaptability in handling various data types and classification scenarios.

Furthermore, the structural efficiency of ZRDT is evident in the reduced size of the generated decision tree. A smaller model not only implies lower complexity but also promotes greater interpretability and faster execution, which are crucial in practical applications. Maintaining high classification accuracy with a more compact structure represents a significant advantage, particularly in environments where computational efficiency and speed are paramount.

7.4 Z-number generation model and its application in a rule-based classification system

In this part, we have investigated the distinct strengths of Z-numbers in encoding and managing uncertain as well as partially reliable information, drawing on their intrinsic links to underlying probability distributions. By exploiting the inherent structure of Z-numbers, an innovative reverse engineering technique has been introduced, whereby a collection of probability distributions is succinctly summarize into a single Z-number via the proposed MEME model. This nonlinear approach enables the direct derivation of Z-numbers from raw data, thus removing the traditional dependence on expert intervention. Such progress underscores the potential of Z-numbers to revolutionize data-driven decision-making processes.

A major contribution of this research is the formulation of Z-valuation if-then rules within our classification framework. Unlike conventional fuzzy rules that typically yield a deterministic consequent, these Z-valuation rules provide an uncertain Z-valuation as their consequents, thereby enhancing the system's capacity to capture and express the inherent uncertainty in the data. This modification not only offers a more faithful representation of real-world information but also enriches the decision-making context by delivering a more nuanced interpretation of potential outcomes.

To validate the effectiveness of the ZVRB classification system, extensive experiments were conducted, comparing its performance against several established classical classifiers and contemporary rule-based systems. The empirical results demonstrate that the ZVRB model significantly outperforms its competitors, owing to its superior ability to accurately capture and utilize the uncertain nature of real-world data—a feature often oversimplified or overlooked in traditional classification approaches.

Moreover, the entirely data-driven nature of the ZVRB system marks a substantial shift towards more adaptive and robust classification methodologies. By leveraging the capabilities of Z-numbers, the system exhibits an exceptional aptitude for handling the variability and ambiguity prevalent in complex datasets, challenges that frequently impede the performance of conventional classification models.

7.5 Z-mixture-number if-then rule induction algorithm for classification tasks

In this part, we introduce the ZMRIA, a sophisticated classification system built on Z-mixture-number rules that integrates FURIA to boost performance. The incorporation of Z-mixture-numbers represents a significant breakthrough in rule-based classification, as it overcomes the shortcomings of previous models in managing heterogeneous data types. This approach, which unifies discrete and continuous variables within a single framework through Z-mixture-numbers, underscores our commitment to both versatility and comprehensive data analysis.

The procedure for generating Z-mixture-numbers from each sample's feature values is firmly rooted in the foundational definition of Z-numbers. By utilizing underlying probability functions during generation, ZMRIA adheres to the core principles of probabilistic and fuzzy logic that underpin Z-numbers, thereby enhancing the faithful representation of uncertainty within the classification process. This ensures that every rule produced by ZMRIA is closely aligned with the actual data distribution, allowing for more nuanced and accurate classification outcomes.

A notable innovation in ZMRIA is its feature selection mechanism, which addresses the common issue of combinatorial rule explosion—a problem that often results in overfitting and

increased computational costs in many rule-based systems. By applying a selection criterion based on the difference between positive and negative Z-mixture-numbers, ZMRIA prioritizes features that offer the greatest discriminatory power between classes. This targeted strategy not only boosts the algorithm's efficiency by reducing computational load but also delivers faster runtimes without sacrificing performance.

Extensive comparative experiments were conducted to assess ZMRIA's performance against ten other classifiers over a wide array of real-world datasets. This evaluation, which included three benchmark classifiers and seven rule-based classifiers, provided a robust foundation for performance comparison. The results were highly promising, showing that ZMRIA not only achieves superior classification accuracy but also operates with faster runtimes relative to seven of the compared methods. These outcomes validate the enhancements integrated into ZMRIA and highlight its potential as an advanced tool for tackling complex classification tasks across diverse domains.

8 Concluding remarks

In this section, we present the results obtained from the research conducted during this Ph.D. dissertation. This thesis has introduced a comprehensive framework based on Z-number theory to address real-world problems characterized by both uncertainty and partial reliability. Recognizing the limitations of traditional fuzzy linguistic modeling—which primarily addresses uncertainty without adequately considering the reliability of information—this work leverages the unique advantages of Z-numbers. By naturally integrating both aspects, the proposed approach provides a more robust and intuitive representation of linguistic information.

A novel Z-number linguistic model was established to represent complex real-world data more effectively. This model simplifies computations by focusing on the ordering of linguistic terms rather than engaging in complex probability distribution derivations, thereby retaining theoretical rigor while enhancing computational efficiency. Such a design directly addresses the challenge of handling large datasets without sacrificing accuracy or interpretability.

Building upon this foundation, a Z-number-valued rule-based classification system was developed. By incorporating a second fuzzy number to capture negative pattern information, the system dynamically adjusts its decision rules to better reflect the dual aspects of uncertainty and reliability. This approach not only improves classification performance compared to classical fuzzy rule-based systems but also enhances the overall transparency and adaptability of the decision-making process.

In addition, a novel Z-number-valued rule-based decision tree was designed. By utilizing negative samples to adjust initial fuzzy numbers and replacing traditional fuzzy confidence with information gain for feature selection, the resulting decision tree exhibits a more compact structure and faster execution. This advancement demonstrates that integrating Z-number theory into decision tree models can significantly enhance both efficiency and predictive performance.

A further contribution is the data-driven Z-number generation model, which summarizes a set of probability distributions into a unique Z-number. This reverse-engineering approach, based on maximizing expected reliability and minimizing entropy, bypasses the need for expert input while ensuring that the generated Z-numbers accurately reflect the underlying data characteristics.

Advanced optimization techniques, inspired by state-of-the-art fuzzy rule induction algorithms, were also incorporated to further refine the system. These enhancements have led to a reduction in model complexity and improved overall accuracy, thereby showcasing the potential of Z-number-based methodologies in practical, real-world applications.

In summary, the integration of Z-number theory with linguistic modeling, rule-based systems, decision trees, and data-driven methodologies not only enriches the theoretical landscape but also provides practical, efficient, and interpretable tools for decision-making under uncertainty. The results presented in this thesis pave the way for future research in large-scale, real-time, and complex environments, where the nuances of uncertainty and reliability are critical to successful outcomes.

Conclusiones

En esta sección, presentamos los resultados obtenidos a partir de la investigación realizada durante esta disertación doctoral. Esta tesis ha introducido un marco integral basado en la teoría de los Z-números para abordar problemas del mundo real caracterizados tanto por la incertidumbre como por una confiabilidad parcial de la información. Reconociendo las limitaciones de la modelización lingüística difusa tradicional —que aborda primordialmente la incertidumbre sin considerar adecuadamente la confiabilidad de la información—, este trabajo aprovecha las ventajas únicas de los Z-números. Al integrar de manera natural ambos aspectos, el enfoque propuesto ofrece una representación más robusta e intuitiva de la información lingüística.

Se estableció un novedoso modelo lingüístico basado en Z-números para representar de manera más efectiva datos complejos del mundo real. Dicho modelo simplifica los cálculos al centrarse en el ordenamiento de términos lingüísticos en lugar de incurrir en complejas derivaciones de distribuciones de probabilidad, manteniendo así el rigor teórico y mejorando la eficiencia computacional. Este diseño aborda directamente el desafío de gestionar grandes conjuntos de datos sin sacrificar la precisión ni la interpretabilidad.

Sobre esta base, se desarrolló un sistema de clasificación basado en reglas con valores Z-número. Al incorporar un segundo número difuso para captar la información de patrones negativos, el sistema ajusta dinámicamente sus reglas de decisión para reflejar de manera más precisa los dos aspectos: incertidumbre y confiabilidad. Este enfoque no solo mejora el rendimiento en la clasificación en comparación con los sistemas clásicos basados en reglas difusas, sino que también realza la transparencia y adaptabilidad global del proceso de toma de decisiones.

Además, se diseñó un novedoso árbol de decisión basado en reglas con valores Z-número. Mediante la utilización de muestras negativas para ajustar los números difusos iniciales y la sustitución de la tradicional confianza difusa por la ganancia de información en la selección de características, el árbol de decisión resultante exhibe una estructura más compacta y una ejecución más rápida. Este avance demuestra que la integración de la teoría de los Z-números en modelos de árboles de decisión puede mejorar significativamente tanto la eficiencia como el rendimiento predictivo.

Otra contribución es el modelo de generación de Z-números basado en datos, que resume un conjunto de distribuciones de probabilidad en un único Z-número. Este enfoque de ingeniería inversa, basado en la maximización de la confiabilidad esperada y la minimización de la entropía, elimina la necesidad de intervención experta, asegurando que los Z-números generados reflejen con precisión las características subyacentes de los datos.

También se incorporaron técnicas avanzadas de optimización, inspiradas en los algoritmos de inducción de reglas difusas de vanguardia, para refinar aún más el sistema. Estas mejoras han conducido a una reducción en la complejidad del modelo y a una mayor precisión global, demostrando así el potencial de las metodologías basadas en Z-números en aplicaciones prácticas del mundo real.

En resumen, la integración de la teoría de los Z-números con la modelización lingüística, los sistemas basados en reglas, los árboles de decisión y las metodologías impulsadas por datos no solo enriquece el panorama teórico, sino que también proporciona herramientas prácticas, eficientes e interpretables para la toma de decisiones en condiciones de incertidumbre. Los resultados presentados en esta tesis allanan el camino para futuras investigaciones en entornos complejos, en tiempo real y a gran escala, donde los matices de la incertidumbre y la confiabilidad son críticos para el éxito.

9 Future Works

Although this thesis has addressed several aspects of Z-number linguistic modeling, new challenges remain. For example, future work could focus on extending the Z-number-valued rule-based system to handle regression tasks, processing large-scale and high-dimensional data, integrating advanced machine learning techniques, and adapting the Z-number linguistic model for dynamic and real-time decision support systems.

1. **Dynamic and Real-Time Decision Support:** Future work should focus on extending the current approach to dynamic and real-time decision support systems. In rapidly changing environments, it is crucial to develop mechanisms that enable continuous model updating and adaptation. Exploring online learning techniques, real-time rule adjustment, and adaptive inference mechanisms will ensure that the system remains responsive and robust as new data become available.
2. **Extension to Regression Tasks:** Although this work has primarily focused on classification problems, future research could explore adapting the Z-number rule system for regression tasks. This would involve modifying the framework to handle continuous output variables and developing appropriate fuzzy operations and aggregation methods that capture the nuances of regression scenarios.
3. **Scalability for Large-Scale and High-Dimensional Data:** As real-world datasets continue to grow in size and complexity, it is imperative to address the computational challenges associated with large-scale and high-dimensional data. Future studies may focus on optimizing the model through techniques such as parallel processing, dimensionality reduction, and more efficient rule induction algorithms to enhance both performance and scalability.
4. **Integration with Deep Learning and Transfer Learning:** Another promising direction is to integrate the Z-number framework with advanced machine learning techniques, such as deep learning and transfer learning. By combining the interpretability and uncertainty modeling strengths of Z-numbers with the powerful representation learning capabilities of deep neural networks, researchers can develop hybrid models that offer improved robustness and accuracy. Additionally, transfer learning can facilitate the application of these models across diverse domains, especially in scenarios with limited labeled data.

Pursuing these directions will not only broaden the applicability of Z-number-based methodologies but also contribute to the development of more versatile, scalable, and adaptive decision-making systems capable of addressing a wider range of real-world challenges.

Chapter II

Publications: Published Papers

1 Z-number linguistic term set for multi-criteria group decision-making and its application in predicting the acceptance of academic papers

Y. Li, G. Kou, Y. Peng, and J. A. Morente-Molinera, Z-number linguistic term set for multi-criteria group decision-making and its application in predicting the acceptance of academic papers, *Applied Intelligence*, 54(21),10962-10981, 2024, <https://doi.org/10.1007/s10489-024-05765-8>.

- Status: **Published**.
- Impact Factor (JCR 2023): 3.4
- Subject Category: Computer Science, Artificial Intelligence, Ranking 78/197 (Q2).

Z-number linguistic term set for multi-criteria group decision-making and its application in predicting the acceptance of academic papers

Yangxue Li¹, Gang Kou^{2*}, Yi Peng^{3*},
Juan Antonio Morente-Molinera^{1*}

^{1*}Andalusian Research Institute in Data Science and Computational Intelligence, Dept. of Computer Science and AI, University of Granada, Granada, 18071, Spain.

²School of Business Administration, Faculty of Business Administration, Southwestern University of Finance and Economics, Chengdu, China.

³School of Management and Economics, University of Electronic Science and Technology of China, Chengdu, China.

*Corresponding author(s). E-mail(s): kougang@swufe.edu.cn;
pengyi@uestc.edu.cn; jamoren@decsai.ugr.es;
Contributing authors: liyangxueugr@gmail.com;

Abstract

Real-world information is often characterized by uncertainty and partial reliability, which led Zadeh to introduce the concept of Z-numbers as a more appropriate formal structure for describing such information. However, the computation of Z-numbers requires solving highly complex optimization problems, limiting their practical application. Although linguistic Z-numbers have been explored for their computational straightforwardness, they lack theoretical support from Z-number theory and exhibit certain limitations. To address these issues and provide theoretical support from Z-numbers, we propose a Z-number linguistic term set to facilitate more efficient processing of Z-number-based information. Specifically, we redefine linguistic Z-numbers as Z-number linguistic terms. By analyzing the hidden probability density functions of these terms, we identify patterns for ranking them. These patterns are used to define the Z-number linguistic term set, which includes all Z-number linguistic terms sorted in order. We also discuss the basic operators between these terms. Furthermore, we develop a multi-criteria group decision-making (MCGDM) model based on the Z-number linguistic term set. Applying our method to predict the acceptance of academic papers, we

demonstrate its effectiveness and superiority. We compare the performance of our MCGDM method with five existing Z-number-based MCGDM methods and eight traditional machine learning clustering algorithms. Our results show that the proposed method outperforms others in terms of accuracy and time consumption, highlighting the potential of Z-number linguistic terms for enhancing Z-number computation and extending the application of Z-number-based information to real-world problems.

Keywords: Z-numbers, Z-number linguistic term, ranking, decision-making, acceptance prediction

1 Introduction

Real-world information is often imperfect and uncertain. To manage such uncertainty, soft constraints on variable values, such as those provided by probability theory, are typically employed. However, probability theory exhibits some limitations when dealing with uncertain information. First, it requires precise and accurate probability values for events, which can be challenging or infeasible in certain situations. Second, probability theory often relies on statistical data modeling. When data is insufficient or unreliable, its application may be constrained. Lastly, probability theory emphasizes objective statistical information, potentially lacking the ability to effectively address subjective judgments and linguistic ambiguity. In such situations, fuzzy set theory provides a robust foundation for modeling real-world information. Fuzzy set theory allows for a more flexible expression of uncertainty. When it is not possible to describe the probability of an event with a precise value, fuzzy sets can be employed to depict its possibility. Additionally, fuzzy set theory does not rely on statistical data and can better handle imprecise information by modeling uncertainty through fuzzy concepts. Furthermore, fuzzy set theory is more suitable for handling subjective judgments and linguistic ambiguity, as it can manage vague language descriptions through fuzzy logic. The most common method of linguistic information modeling is the fuzzy linguistic approach[1–3]. It deals with qualitative aspects of linguistic information represented in qualitative terms by linguistic variables. It has been successfully applied in various fields, such as supplier selection [4], route optimization [5], logistics outsourcing problems [6], decision-making [7–9], consensus reaching process [10], and risk analysis [11–13].

Furthermore, it is essential to consider not only uncertainty but also the partial reliability of information when handling it. The reliability of information is closely related to its source, and unreliable information can mislead decision-makers, leading to incorrect decisions and significant economic and resource losses. Therefore, assessing the reliability of information is crucial in practical decision-making. In [14], researchers introduced the concept of basic uncertain linguistic information, which comprises a fuzzy linguistic variable and a crisp number between 0 and 1 representing the reliability level of the linguistic variable. When the fuzzy linguistic variable is replaced by an interval 2-tuple linguistic variable, this ordered pair is known as improved basic

uncertain linguistic information, as proposed in [15]. In [16] and [17], the authors aggregate collective opinions by considering the maximum values of consensus and confidence level (reliability) as two objectives. Ran et al. propose a method to manage linguistic group overconfidence behavior during the consensus-reaching process in group decision-making [18]. These methods use a crisp number to describe the reliability of information, which may distort the original meaning of linguistic information when expressing terms such as (*good, sure*). To better capture both the uncertainty and partial reliability of linguistic information, Zadeh introduced the concept of the Z-number [19]. A Z-number is composed of two ordered fuzzy numbers, $Z = (A, B)$, where the first fuzzy number, A , restricts the values that the variable can take, and the second fuzzy number, B , restricts the reliability of A [19]. However, performing arithmetic operations on Z-numbers, especially continuous Z-numbers, is complex and involves multiple optimization problems to find appropriate hidden probability distributions [20]. This process can be time-consuming and unsuitable for large datasets. Combining linguistic terms and Z-numbers, Wang et al. provided the definition of linguistic Z-numbers as a form of Z-numbers or a subclass of Z-numbers [21, 22]. Linguistic Z-number is an extension of Z-number, representing the two components of a Z-number using fuzzy linguistic terms. The authors also defined some operations of linguistic Z-numbers. Due to their intuitive representation of linguistic information and straightforward computational approach, linguistic Z-numbers find applications in various scenarios involving Z-number-based information [23–25].

Linguistic Z-numbers are well-suited for describing review scoring information, especially in the context of peer reviews for academic papers in journals and conferences. The peer review process is a crucial step in academic publishing, as it determines whether a paper will be accepted for publication. During this process, a manuscript is evaluated by experts in the field, who provide scores and comments on various aspects of the paper’s quality, such as clarity and originality, using natural language descriptions. Linguistic Z-numbers are well-suited to handle these natural language descriptions due to their ability to model linguistic quality assessments. Fuzzy linguistic terms effectively capture the uncertainty and imprecision inherent in qualitative reviews. Additionally, “reviewer confidence,” which indicates the level of certainty a reviewer has in their assessment, can be represented by the second fuzzy linguistic term in a linguistic Z-number, corresponding to reliability. The first component of the linguistic Z-number can represent aspects such as clarity. For example, a linguistic Z-number, (*Very clear, Quite sure*), combines a clarity score with a confidence level, providing a comprehensive representation of the paper’s evaluation. Hence, we can use linguistic Z-numbers to describe review scoring information, utilizing the calculations and inference of linguistic Z-numbers to predict the acceptance of academic papers. However, while the computational approach proposed by Wang et al. is straightforward, it lacks robust theoretical support from Z-number theory and has certain limitations.

To address these limitations and provide theoretical support from Z-numbers, the expectation values of hidden probability density functions are used to rank all possible linguistic Z-numbers. To distinguish it from the operational system proposed by Wang et al., we redefine linguistic Z-numbers as Z-number linguistic terms here, which

is defined based on two fuzzy linguistic term sets: restriction and reliability fuzzy linguistic term sets. A Z-number linguistic term is composed of two ordered fuzzy linguistic terms, denoted as $Z = (s, r)$. The first fuzzy linguistic term, denoted as s , is selected from the restriction fuzzy linguistic term set, while the second fuzzy linguistic term, denoted as r , is selected from the reliability fuzzy linguistic term set. We define an ordered set, known as the Z-number linguistic term set, by ranking all possible Z-number linguistic terms in ascending order. Using this ordering information, we establish basic operators between Z-number linguistic terms, including comparison, negation, and novel aggregation operators. These operators are designed for simplicity and efficiency, requiring only basic numerical calculations. Finally, we develop a MCGDM method based on the Z-number linguistic term set.

To demonstrate the effectiveness and advantages of the proposed MCGDM method, we conducted a comprehensive comparison with five existing Z-number-based MCGDM methods and eight machine learning clustering algorithms, using the PeerRead dataset for predicting academic paper acceptance. The results of this comparison reveal that our proposed MCGDM method consistently outperforms the other methods. Our approach, which utilizes Z-number linguistic terms, simplifies decision-making by relying on straightforward numerical calculations and focusing solely on the ordering relationships of linguistic terms, rather than defining their specific semantics. In contrast to other Z-number-based MCGDM methods, this results in a more efficient and practical decision-making process. Additionally, when compared to machine learning clustering algorithms, our MCGDM model offers greater interpretability and efficiency.

In summary, this paper makes several key contributions as follows.

- (1) The Z-number linguistic term set is defined, and some basic operators for Z-number linguistic terms are designed based on this set. Compared to existing methods, these aggregation methods are computationally simple.
- (2) The utilization of the interval expectation value of the hidden probability density function for ranking Z-number linguistic terms provides theoretical support for the definition of the Z-number linguistic term set.
- (3) A MCGDM method is designed based on the Z-number linguistic term set, which provides an excellent and simple approach to make decision.
- (4) The proposed MCGDM method is applied to the real-world problem of academic paper acceptance prediction. The superiority of the method is confirmed through comparisons with five existing Z-number-based MCGDM methods and eight traditional machine learning clustering algorithms.

This paper is organized as follows. We briefly introduce some background knowledge of fuzzy linguistic approach and Z-numbers in Section 2. The concepts of Z-number linguistic term and Z-number linguistic term set are defined in Section 3. In Section 4, we construct a MCGDM model based on Z-number linguistic terms. In Section 5, we conduct an experiment to validate the effectiveness and superiority of our proposed method using the ACL 2017 dataset. Finally, we make some conclusions in Section 6.

2 Preliminaries

In this section, we review some key concepts related to the fuzzy linguistic approach and Z-numbers, which are integral to defining the Z-number linguistic term set.

2.1 Fuzzy linguistic approach and its extension

Humans primarily use natural language for reasoning and communication in their daily lives, often drawing conclusions that are expressed in the form of natural language. The fuzzy linguistic approach is a method designed to enable computational systems to handle and reason with linguistic information by employing fuzzy sets to model information uncertainty. In this approach, the qualitative aspects of linguistic information are represented as linguistic values through linguistic variables. Zadeh defined a linguistic variable as one whose values are words and sentences in natural language, rather than numerical values [1–3].

In the fuzzy linguistic approach, once a linguistic variable is defined, it is essential to construct a term set that includes all possible linguistic values of the variable along with their associated semantics. This collection is referred to as a fuzzy linguistic term set, where each individual linguistic value is termed a fuzzy linguistic term.

Definition 1 (Fuzzy linguistic term set). *A fuzzy linguistic term set $S = \{s_0, s_1, \dots, s_g\}$ is a finite and totally ordered set, where all $s_i, 0 \leq i \leq g$ are the fuzzy linguistic terms which are distributed on a scale and satisfy the following properties [26]:*

- (1) *Negation operator $Neg(s_i) = s_j$, where $j = g - i$ ($g + 1$ is the granularity of the term set).*
- (2) *Comparison operator, $s_i > s_j$ if and only if $i > j$.*

The semantics of the terms are represented by membership functions of fuzzy numbers.

For example, a fuzzy linguistic term set with seven linguistic terms $S = \{s_0 : \text{Very Low} = (0, 0, \frac{1}{6}), s_1 : \text{Low} = (0, \frac{1}{6}, \frac{1}{3}), s_2 : \text{Slightly Low} = (\frac{1}{6}, \frac{1}{3}, \frac{1}{2}), s_3 : \text{Medium} = (\frac{1}{3}, \frac{1}{2}, \frac{2}{3}), s_4 : \text{Slightly High} = (\frac{1}{2}, \frac{2}{3}, \frac{5}{6}), s_5 : \text{High} = (\frac{2}{3}, \frac{5}{6}, 1), s_6 : \text{Very High} = (\frac{5}{6}, 1, 1)\}$, which is graphically shown in Figure 1.

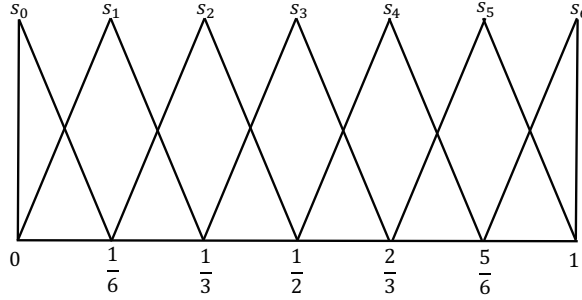


Fig. 1 A seven term set with semantics.

To address the limitations of fuzzy linguistic term set in information description and computation, which may result in information loss due to insufficient precision, Herrera and Martinez proposed 2-tuple fuzzy linguistic representation model [27].

Definition 2 (2-tuple fuzzy linguistic representation model [27]). *Let $S = \{s_0, s_1, \dots, s_g\}$ be a fuzzy linguistic term set, the linguistic information is represented by means of 2-tuple (s_i, α_i) , where $s_i \in S$ and $\alpha_i \in [-0.5, 0.5]$. Where α_i represents the offset of s_i on the coordinate axis.*

However, Guo et al. argue that the numerical values in 2-tuple fuzzy linguistic representation model may distort the original meaning of linguistic information [28]. Therefore, they proposed the concept of a double hierarchy linguistic term set, which consists of two hierarchy linguistic term sets. The first hierarchy linguistic term set is the main linguistic term set and the second hierarchy linguistic term set is a linguistic feature or detailed supplementary of each linguistic term in the first hierarchy linguistic term set.

Definition 3 (Double hierarchy linguistic term set [28]). *Let $S = \{s_0, s_1, \dots, s_g\}$ be the first hierarchy fuzzy linguistic term set, $O^i = \{o_0^i, o_1^i, \dots, o_\zeta^i\}$ be the second hierarchy fuzzy linguistic term set. Then, $S_O = \{s_{i < o_k^i} | i = 0, 1, \dots, g; k = 0, 1, \dots, \zeta\}$ be a double hierarchy linguistic term set, where $s_{i < o_k^i}$ is called double hierarchy linguistic term.*

2.2 Z-numbers

Since its introduction, the Z-number has garnered significant attention from researchers due to its robust capability to model both uncertainty and partial reliability of information. The mathematical framework based on Z-numbers is continually being refined [29]. Additionally, Z-numbers have found applications across various fields [30].

Definition 4 (Probability measure [19]). *Let A be a fuzzy set in a space, U , and X be a random variable, the probability measure of the fuzzy event "X is A", $Prob(A)$, is defined as the following formula.*

$$Prob(A) = \int_U p(u) \mu_A(u) du, \quad (1)$$

where $\mu_A(u)$ is the membership function of A . $p(u)$ is the hidden probability density function of X on U .

If the centroids of A and p are coincident, these distributions are compatible:

$$\int_U up(u) du = \frac{\int_U u \mu_A(u) du}{\int_U \mu_A(u) du}. \quad (2)$$

Definition 5 (Z-number [19]). *A Z-number, $Z = (A, B)$, is an ordered pair of fuzzy numbers, where the first fuzzy number A restricts the values that the variable X can take:*

$$X \text{ is } A, \quad (3)$$

and the second fuzzy number B restricts the probability measure of "X is A":

$$\text{Prob}(A) \text{ is } B. \quad (4)$$

The membership function of A is $\mu_A(u) : U \rightarrow [0, 1]$, and that of B is $\mu_B(v) : [0, 1] \rightarrow [0, 1]$.

In a Z-number, the second fuzzy number represents the reliability of the first fuzzy number. While Z-numbers are highly effective for describing uncertain information, performing arithmetic operations on them is complex. This complexity arises particularly with continuous Z-numbers, as it requires solving multiple optimization problems to determine the appropriate hidden probability distributions.

Assume there are two Z-numbers, $Z_1 = (A_1, B_1)$ and $Z_2 = (A_2, B_2)$, describing imperfect information about the values of variables X_1 and X_2 . The addition, subtraction, multiplication or division of these two Z-numbers, $Z_{12} = Z_1 * Z_2$, $*$ $\in \{+, -, \times, /\}$ can be calculated as the following steps [19, 31, 32].

- Step 1: Calculate the first fuzzy number A_{12} of resulted Z-number Z_{12} , $A_{12} = A_1 * A_2$. The computational approach employed here involves operations on fuzzy numbers.
- Step 2: Find the hidden probability density function of X_1 and X_2 , denoted as p_1 and p_2 . In the case of Z-numbers what we know are not p_1 and p_2 but restrictions on p_1 and p_2 . We can identify some probability functions satisfying the known restrictions. In terms of the membership functions of B_1 and B_2 , the restrictions may be expressed as:

$$\begin{aligned} \mu_{B_k}(\int_U p_k(u) \mu_{A_k}(u) du), \\ \int_U p_k(u) du = 1, \\ \int_U u p_k(u) du = \frac{\int_U u \mu_{A_k}(u) du}{\int_U \mu_{A_k}(u) du}, \\ k = 1, 2. \end{aligned}$$

- Step 3: Compute $p_{12} = p_1 * p_2$. The computational approach employed here involves operations on random variable.
- Step 4: Calculate the second fuzzy number B_{12} of resulted Z-number Z_{12} using the following formula.

$$\mu_{B_{12}}(\int_U p_{12}(u) \mu_{A_{12}}(u) du) = \sup(\min_{k=1,2}(\mu_{B_k}(\int_U p_k(u) \mu_{A_k}(u) du))).$$

As the result, $Z_{12} = Z_1 * Z_2$ is obtained as (A_{12}, B_{12}) .

Example 1. Given two Z-numbers, $Z_1 = (A_1, B_1)$ and $Z_2 = (A_2, B_2)$, the fuzzy numbers are triangular fuzzy numbers: $A_1 = (\frac{1}{3}, \frac{1}{2}, \frac{2}{3})$, $B_1 = (\frac{1}{3}, \frac{1}{2}, \frac{2}{3})$, $A_2 = (\frac{2}{3}, \frac{5}{6}, 1)$ and $B_2 = (\frac{1}{6}, \frac{1}{3}, \frac{1}{2})$. The sum of Z_1 and Z_2 is $Z_{12} = Z_1 + Z_2$, calculated by the following steps.

- Step 1: The first fuzzy number of Z_{12} is $A_{12} = A_1 + A_2 = (1, \frac{4}{3}, \frac{5}{3})$.

Step 2: We assume the hidden probability density functions, p_1 and p_2 , are normal distributions. Subsequently, we can uncover the underlying probability functions through the following nonlinear optimization problem:

$$p_k(u) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(u - \mu_k)^2}{2\sigma_k^2}\right), k = 1, 2,$$

subject to:

$$\mu_{B_k} \left(\int_U p_k(u) \mu_{A_k}(u) du \right),$$

$$\int_U u p_k(u) du = \mu_k = \frac{\int_U u \mu_{A_k}(u) du}{\int_U \mu_{A_k}(u) du}.$$

We can easily obtain, $\mu_1 = \frac{1}{2}$ and $\mu_2 = \frac{5}{6}$. For different base value of B_k , the value of σ_k is different, as shown in Table 1.

v_1	$\frac{1}{3}$	$\frac{5}{12}$	$\frac{1}{2}$	$\frac{7}{12}$	$\frac{2}{3}$
σ_1	0.1873	0.1439	0.1133	0.0898	0.0701
v_2	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{5}{12}$	$\frac{1}{2}$
σ_2	0.3931	0.2570	0.1873	0.1439	0.1133

Table 1 The obtained solutions σ_k for different base values of B_k .

Step 3: Compute $p_{12} = p_1 + p_2$. $\mu_{12} = \mu_1 + \mu_2 = \frac{4}{3}$ and $\sigma_{12} = \sqrt{\sigma_1^2 + \sigma_2^2}$. Then the values of σ_{12} are shown as Table 2.

$\sigma_1 \backslash \sigma_2$	0.1873	0.1439	0.1133	0.0898	0.0701
0.3931	0.4354	0.4186	0.4091	0.4032	0.3993
0.2570	0.3180	0.2945	0.2809	0.2723	0.2664
0.1873	0.2649	0.2362	0.2189	0.2077	0.2000
0.1439	0.2362	0.2035	0.1832	0.1696	0.1601
0.1133	0.2189	0.1832	0.1603	0.1446	0.1333

Table 2 The values of σ_{12} .

Step 4: The base values of B_{12} is calculated by

$$v_{12} = \int_U p_{12}(u) \mu_{A_{12}}(u) du.$$

The results are shown in Table 3. Therefore, the second fuzzy number B_{12} can be approximated as a triangular fuzzy number $B_{12} = (0.2913, 0.5126, 0.6825)$. In fact, this approximate second fuzzy number is highly inaccurate. According to Table 3, the base values with a membership degree of 0 include not only 0.2913 and 0.6825 but

$v_2 \backslash v_1$	$\frac{1}{3}$	$\frac{5}{12}$	$\frac{1}{2}$	$\frac{7}{12}$	$\frac{2}{3}$
$\frac{1}{6}$	0.2913	0.3019	0.3082	0.3122	0.3150
$\frac{1}{4}$	0.3837	0.4088	0.4248	0.4355	0.4430
$\frac{1}{3}$	0.4450	0.4854	0.5126	0.5315	0.5451
$\frac{5}{12}$	0.4854	0.5389	0.5765	0.6035	0.6234
$\frac{1}{2}$	0.5126	0.5765	0.6229	0.6569	0.6825

Table 3 The base values of B_{12} .

also 0.3019, 0.3082 and others. However, in the approximate second fuzzy number, the membership degrees for these base values are not equal to 0. The addition of Z_1 and Z_2 is $((1, \frac{4}{3}, \frac{5}{3}), (0.2913, 0.5126, 0.6825))$.

This example illustrates that Z-number computations involve solving numerous nonlinear optimization problems and often yield approximate results rather than exact ones. Achieving precise results for Z-number operations is both complex and challenging to implement. In practical decision-making scenarios, the extensive computations required can slow down the decision-making process and lead to significant resource consumption. Therefore, our goal is to propose simpler methods that provide approximate operational results, thereby enhancing decision-making efficiency.

2.3 Linguistic Z-numbers

Linguistic Z-number is an extension of Z-numbers, representing the two components of a Z-number using fuzzy linguistic terms [21, 22].

Definition 6 (Linguistic Z-number [21, 22]). Let X be a universe of discourse, $S_1 = \{s_0, s_1, \dots, s_{g_1}\}$ and $S_2 = \{s'_0, s'_1, \dots, s'_{g_2}\}$ be two fuzzy linguistic term sets. A linguistic Z-number Z on X is an object with the form:

$$Z = \{(x, A_{\phi(x)}, B_{\varphi(x)}) | x \in X\}, \quad (5)$$

where $A_{\phi(x)} \in S_1$ is a fuzzy restriction on the values that the variable can take, and $B_{\varphi(x)} \in S_2$ is a fuzzy restriction on the reliability of $A_{\phi(x)}$.

Let $Z_1 = (A_{\phi(1)}, B_{\varphi(1)})$ and $Z_2 = (A_{\phi(2)}, B_{\varphi(2)})$ be two linguistic Z-numbers, and f and h functions be strictly monotonically increasing functions, some linguistic Z-number operators are defined as follows [21, 22].

- (1) $neg(Z_1) = (f^{-1}(f(s_{g_1}) - f(A_{\phi(1)})), h^{-1}(h(s_{g_2}) - h(B_{\varphi(1)})))$;
- (2) $Z_1 + Z_2 = (f^{-1}(f(A_{\phi(1)}) + f(A_{\phi(2)})), h^{-1}(\frac{f(A_{\phi(1)}) \times h(B_{\varphi(1)}) + f(A_{\phi(2)}) \times h(B_{\varphi(2)})}{f(A_{\phi(1)}) + f(A_{\phi(2)})}))$;
- (3) $\lambda Z_1 = (f^{-1}(\lambda f(A_{\phi(1)})), B_{\varphi(1)}), \lambda \geq 0$;
- (4) $Z_1 \times Z_2 = (f^{-1}(f(A_{\phi(1)})f(A_{\phi(2)})), h^{-1}(h(B_{\varphi(1)})h(B_{\varphi(2)})))$;
- (5) $Z_1^\lambda = (f^{-1}(f(A_{\phi(1)})^\lambda), h^{-1}(h(B_{\varphi(1)})^\lambda)), \lambda \geq 0$.

The comparison method is defined as follows [21, 22].

- (1) If $A_{\phi(1)} > A_{\phi(2)}$ and $B_{\varphi(1)} > B_{\varphi(2)}$, then Z_1 is strictly greater than Z_2 , denoted as $Z_1 > Z_2$;

- (2) If $S(Z_1) > S(Z_2)$ or $S(Z_1) = S(Z_2)$ and $A(Z_1) > A(Z_2)$, then Z_1 is greater than Z_2 , denoted as $Z_1 \succ Z_2$;
- (3) If $S(Z_1) = S(Z_2)$ and $A(Z_1) = A(Z_2)$, then Z_1 equals Z_2 , denoted as $Z_1 \approx Z_2$;
- (4) If $S(Z_1) = S(Z_2)$ and $A(Z_1) < A(Z_2)$ or $S(Z_1) < S(Z_2)$, then Z_1 is less than Z_2 , denoted as $Z_1 \prec Z_2$.

Where $S(Z_1) = f(A_{\phi(1)})g(B_{\varphi(1)})$ and $A(Z_1) = f(A_{\phi(1)})(1 - g(B_{\varphi(1)}))$.

The operations proposed by Wang et al. lack theoretical support from Z-number theory and have some limitations [21]. Firstly, if both two linguistic Z-numbers contain s_0 , according to the above operation rules, their addition cannot be calculated. Such as when calculating $(s_0, s'_3) + (s_0, s'_5)$, the denominator $f(s_0) + f(s_0) = 0$. Moreover, In the comparison method of linguistic Z-numbers, according to the third rule, $(s_0, s'_0) \approx (s_0, s'_6)$ holds because $S((s_0, s'_0)) = S((s_0, s'_6)) = 0$ and $A((s_0, s'_0)) = A((s_0, s'_6)) = 0$. This conclusion goes against intuition.

3 Z-number linguistic term set

In this section, we introduce our novel Z-number-based linguistic computational model, which comprises two main components: (1) a linguistic information representation model utilizing Z-numbers, and (2) a computational model for aggregating linguistic information with Z-numbers. We also analyze the hidden probability density functions of Z-number linguistic terms, deriving ranking rules based on their interval expectation values. This analysis provides the theoretical support for our proposed model.

3.1 Basic concept

Definition 7. Let $S = \{s_0, s_1, \dots, s_m\}$ is a finite and totally ordered fuzzy linguistic term set, representing a set of fuzzy restrictions, referred to as the restriction fuzzy linguistic term set. Let $R = \{r_0, r_1, \dots, r_n\}$ be another finite and totally ordered fuzzy linguistic term set, representing a set of fuzzy restrictions on reliability, known as the reliability fuzzy linguistic term set. A Z-number linguistic term is defined as an ordered pair:

$$Z_{I(i,j)} = (s_i, r_j), \quad (6)$$

where $s_i \in S$, $r_j \in R$, $0 \leq i \leq m$ and $0 \leq j \leq n$. A Z-number linguistic term consists of two ordered fuzzy linguistic terms, s_i is a fuzzy restriction on the values that the random variable X can take and r_j is a fuzzy restriction on the reliability of s_i . $I(i, j)$ is an index function shown as Eq. (7).

The distinction between Z-number linguistic terms and other models, such as the 2-tuple fuzzy linguistic representation model or double hierarchy linguistic terms, lies in how supplementary information is represented. In the 2-tuple fuzzy linguistic representation model, a crisp number is used to provide supplementary information for a fuzzy linguistic term. Similarly, double hierarchy linguistic terms use an additional fuzzy linguistic term to describe supplementary information for the primary fuzzy linguistic term. In contrast, the second fuzzy linguistic term in a Z-number linguistic term does not act as supplementary information but rather represents the reliability of the first term. This distinction is significant because, unlike the former models

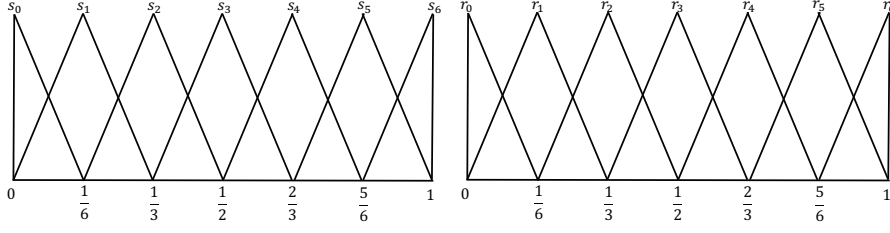


Fig. 2 Restriction fuzzy linguistic term set with its semantics. **Fig. 3** Reliability fuzzy linguistic term set with its semantics.

where the second component is used to enhance the precision of the fuzzy linguistic term set, the Z-number linguistic term model is designed to more accurately reflect the way human language processes information. For example, when experts provide judgments, they often express their confidence or uncertainty, which is directly modeled by the reliability term in a Z-number linguistic term.

The semantics of fuzzy linguistic terms are usually described by membership functions of fuzzy numbers. We consider balanced triangular membership functions represented by 3-tuples, $s_i = (a_i^1, b_i^1, c_i^1)$ and $r_j = (a_j^2, b_j^2, c_j^2)$. For example, $S = \{s_0 : \text{Very Low} = (0, 0, \frac{1}{6}), s_1 : \text{Low} = (0, \frac{1}{6}, \frac{1}{3}), s_2 : \text{Slightly Low} = (\frac{1}{6}, \frac{1}{3}, \frac{1}{2}), s_3 : \text{Medium} = (\frac{1}{3}, \frac{1}{2}, \frac{2}{3}), s_4 : \text{Slightly High} = (\frac{1}{2}, \frac{2}{3}, \frac{5}{6}), s_5 : \text{High} = (\frac{2}{3}, \frac{5}{6}, 1), s_6 : \text{Very High} = (\frac{5}{6}, 1, 1)\}$; $R = \{r_0 : \text{Very Not Sure} = (0, 0, \frac{1}{6}), r_1 : \text{Slightly Not Sure} = (0, \frac{1}{6}, \frac{1}{3}), r_2 : \text{Not Sure} = (\frac{1}{6}, \frac{1}{3}, \frac{1}{2}), r_3 : \text{Medium} = (\frac{1}{3}, \frac{1}{2}, \frac{2}{3}), r_4 : \text{Sure} = (\frac{1}{2}, \frac{2}{3}, \frac{5}{6}), r_5 : \text{Slightly Sure} = (\frac{2}{3}, \frac{5}{6}, 1), r_6 : \text{Very Sure} = (\frac{5}{6}, 1, 1)\}$, which is graphically shown in Figures 2-3. This is just an example, the specific semantics in the Z-number linguistic terms are not necessary; what matters is their ordering relationship.

Some properties of linguistic Z-number terms can be summarized as follows.

1. Maximality

The most positive Z-number linguistic term is (s_m, r_n) . For instance, (Very High, Very sure).

2. Minimality

The most negative Z-number linguistic term is (s_0, r_n) . For instance, (Very Low, Very High).

3. Symmetry

For Z-number linguistic terms $(s_{\frac{m}{2}-k}, r_j)$ and $(s_{\frac{m}{2}+k}, r_j)$, $k = 0, 1, \dots, \frac{m}{2}$, there is a symmetrical point $(s_{\frac{m}{2}}, r_j)$, $j = 0, 1, \dots, n$. For instance, "Expert is very sure the speed is high." and "Expert is very sure the speed is low."

4. Monotonicity

Assume two Z-number linguistic terms (s_{i_1}, r_{j_2}) and (s_{i_2}, r_{j_2}) ,

- if $i_1 > i_2$, then $(s_{i_1}, r_{j_1}) \succ (s_{i_2}, r_{j_2})$;
- if $i_1 < i_2$, then $(s_{i_1}, r_{j_1}) \prec (s_{i_2}, r_{j_2})$;
- if $i_1 = i_2$, s_{i_1}, s_{i_2} are positive, $m/2 \leq i_1, i_2 \leq m$, and $j_1 > j_2$, then $(s_{i_1}, r_{j_1}) \succ (s_{i_2}, r_{j_2})$;

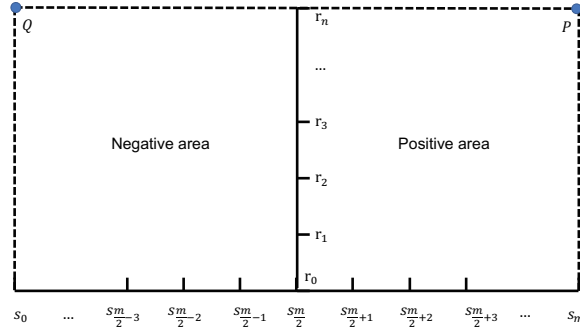


Fig. 4 m is a even number.

- if $i_1 = i_2$, s_{i_1}, s_{i_2} are positive, $m/2 \leq i_1, i_2 \leq m$, and $j_1 < j_2$, then $(s_{i_1}, r_{j_1}) \prec (s_{i_2}, r_{j_2})$;
- if $i_1 = i_2$, s_{i_1}, s_{i_2} are negation, $0 \leq i_1, i_2 < m/2$, and $j_1 > j_2$, then $(s_{i_1}, r_{j_1}) \prec (s_{i_2}, r_{j_2})$;
- if $i_1 = i_2$, s_{i_1}, s_{i_2} are negation, $0 \leq i_1, i_2 < m/2$, and $j_1 < j_2$, then $(s_{i_1}, r_{j_1}) \succ (s_{i_2}, r_{j_2})$;

In order to visually illustrate the aforementioned properties, we utilize a half-coordinate system, as shown in the Figure 4. In which, point P is the most positive Z-number linguistic term, while point Q is the most negative Z-number linguistic term. Point P and point Q are symmetry with respect to vertical axis.

Definition 8. For S and R , we can get a set of all $(m+1) \times (n+1)$ Z-number linguistic terms $\{(s_0, r_0), (s_0, r_1), \dots, (s_0, r_n), \dots, (s_m, r_{n-1}), (s_m, r_n)\}$. For that purpose, it is possible to rank all terms from smallest to largest to get a finite and ordered set $\Gamma = \{Z_0, Z_1, \dots, Z_{(m+1) \times (n+1) - 1}\}$, named as the Z-number linguistic term set.

From the monotonicity of the Z-number linguistic terms, we can infer the order index function $I(i, j)$ of a Z-number linguistic term (s_i, r_j) :

$$I(i, j) = \begin{cases} i(n+1) + j, & \text{If } s_i \text{ is positive;} \\ i(n+1) + n - j, & \text{If } s_i \text{ is negative.} \end{cases} \quad (7)$$

Example 2. Assume $m = n = 6$, then $S = \{s_0, s_1, s_2, s_3, s_4, s_5, s_6\}$ and $R = \{r_0, r_1, r_2, r_3, r_4, r_5, r_6\}$. According to the ranking value function of Z-number linguistic term, we can get:

$$\begin{aligned} \Gamma &= \{Z_0, Z_1, Z_2, \dots, Z_{48}\} \\ &= \{(s_0, r_6), (s_0, r_5), (s_0, r_4), (s_0, r_3), (s_0, r_2), (s_0, r_1), (s_0, r_0), \\ &\quad (s_1, r_6), (s_1, r_5), (s_1, r_4), (s_1, r_3), (s_1, r_2), (s_1, r_1), (s_1, r_0), \\ &\quad (s_2, r_6), (s_2, r_5), (s_2, r_4), (s_2, r_3), (s_2, r_2), (s_2, r_1), (s_2, r_0), \\ &\quad (s_3, r_0), (s_3, r_1), (s_3, r_2), (s_3, r_3), (s_3, r_4), (s_3, r_5), (s_3, r_6), \\ &\quad (s_4, r_0), (s_4, r_1), (s_4, r_2), (s_4, r_3), (s_4, r_4), (s_4, r_5), (s_4, r_6), \end{aligned}$$

$$\begin{aligned} & (s_5, r_0), (s_5, r_1), (s_5, r_2), (s_5, r_3), (s_5, r_4), (s_5, r_5), (s_5, r_6), \\ & (s_6, r_0), (s_6, r_1), (s_6, r_2), (s_6, r_3), (s_6, r_4), (s_6, r_5), (s_6, r_6) \}. \end{aligned}$$

3.2 Theoretical support

The ranking of Z-number linguistic terms differs from that of traditional Z-numbers due to the nature of their definitions. In linguistic Z-numbers, fuzzy linguistic terms are defined within a finite set of fuzzy terms over a specific interval, whereas in traditional Z-numbers, fuzzy numbers are defined over the real number line. A fundamental property of probability is that as the probability measure of one restriction fuzzy linguistic term decreases, the probability measure of the remaining restriction fuzzy linguistic terms increases. This property allows us to use the interval expectation values of hidden probability density functions to effectively rank Z-number linguistic terms. By leveraging these interval expectation values, we can establish a ranking that reflects the relative importance and reliability of the fuzzy linguistic terms in the context of Z-number linguistic terms.

For two Z-number linguistic terms, $Z_{\phi(1)} = (s_{i_1}, r_{j_1})$ and $Z_{\phi(2)} = (s_{i_2}, r_{j_2})$, where $s_{i_1} = (a_{i_1}^1, b_{i_1}^1, c_{i_1}^1)$, $r_{j_1} = (a_{j_1}^2, b_{j_1}^2, c_{j_1}^2)$, $s_{i_2} = (a_{i_2}^1, b_{i_2}^1, c_{i_2}^1)$, $r_{j_2} = (a_{j_2}^2, b_{j_2}^2, c_{j_2}^2)$ are triangular fuzzy numbers and $\phi(\cdot) = 0, 1, \dots, (m+1)(n+1) - 1$, we assume their hidden probability distributions are normal distributions.

$$p_k(x) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(x - \mu_k)^2}{2\sigma_k^2}\right), k = 1, 2. \quad (8)$$

According to Eq. (2), we can obtain the expectation value of normal distributions:

$$\mu_k = \int_R x p_k(x) dx = \frac{\int_R x \mu_{s_{i_k}}(x) dx}{\int_R \mu_{s_{i_k}}(x) dx} = b_{i_k}^1, k = 1, 2. \quad (9)$$

The value of σ_k can be obtained by Eq. (4):

$$\mu_{r_{j_k}}(prob_k) = \mu_{r_{j_k}}\left(\int_R p_k(x) \mu_{r_{j_k}}(x) dx\right). \quad (10)$$

The expectation value of p_k is $b_{i_k}^1$ on the real number line, but it changes on the specific interval. The restriction fuzzy linguistic term set always defined on the interval $[0, 1]$. The expectation value on $[0, 1]$ is defined as:

$$E(k)_{[0,1]} = \frac{\int_0^1 x p_k(x) dx}{\int_0^1 p_k(x) dx}, \quad (11)$$

where $\int_0^1 p_k(x) dx$ is the normalization coefficient. If $E(1)_{[0,1]} > E(2)_{[0,1]}$, then $Z_{\phi(1)} \succ Z_{\phi(2)}$; If $E(1)_{[0,1]} < E(2)_{[0,1]}$, then $Z_{\phi(1)} \prec Z_{\phi(2)}$; $E(1)_{[0,1]} = E(2)_{[0,1]}$, then $Z_{\phi(1)} \approx Z_{\phi(2)}$.

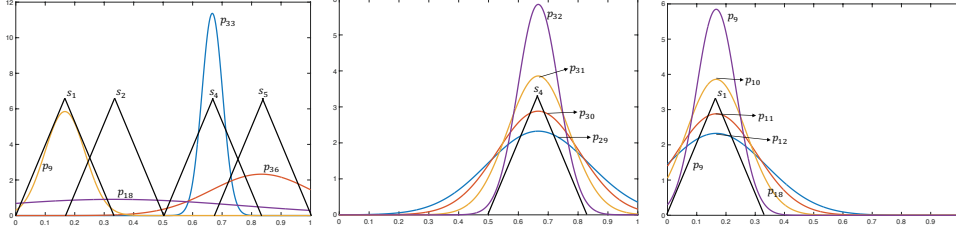


Fig. 5 The hidden probabilities in Example 3.

Fig. 6 The hidden probabilities in Example 4.

Fig. 7 The hidden probabilities in Example 5.

Based on the convexity property of the normal distribution function, it can be observed that when $s_{i_1} > s_{i_2}$, the interval expectation value of p_1 will always be greater than that of p_2 , regardless of how the variance changes. We can draw the conclusion that:

- if $i_1 > i_2$, then $(s_{i_1}, r_{j_1}) \succ (s_{i_2}, r_{j_2})$;
- if $i_1 < i_2$, then $(s_{i_1}, r_{j_1}) \prec (s_{i_2}, r_{j_2})$.

Example 3. Assume there are four Z-number linguistic terms $Z_9 = (s_1, r_4)$, $Z_{18} = (s_2, r_2)$, $Z_{33} = (s_4, r_5)$, and $Z_{36} = (s_5, r_1)$ shown in Figures 2 and 3, their hidden probability distributions are:

$$p_9(x) = \frac{1}{\sqrt{2\pi}\sigma_9} \exp\left(-\frac{(x - \frac{1}{6})^2}{2\sigma_9^2}\right),$$

$$p_{18}(x) = \frac{1}{\sqrt{2\pi}\sigma_{18}} \exp\left(-\frac{(x - \frac{1}{3})^2}{2\sigma_{18}^2}\right),$$

$$p_{33}(x) = \frac{1}{\sqrt{2\pi}\sigma_{33}} \exp\left(-\frac{(x - \frac{2}{3})^2}{2\sigma_{33}^2}\right),$$

$$p_{36}(x) = \frac{1}{\sqrt{2\pi}\sigma_{36}} \exp\left(-\frac{(x - \frac{5}{6})^2}{2\sigma_{36}^2}\right),$$

When $Prob_9 = \frac{2}{3}$, $\sigma_9 = 0.004654$ and $E(9)_{[0,1]} = 0.168053961$. When $Prob_{18} = \frac{1}{3}$, $\sigma_{18} = 0.191858$ and $E(18)_{[0,1]} = 0.336411382$. When $Prob_{33} = \frac{5}{6}$, $\sigma_{33} = 0.00123$ and $E(33)_{[0,1]} = 0.666666667$. When $Prob_{36} = \frac{1}{6}$, $\sigma_{36} = 0.029439$ and $E(36)_{[0,1]} = 0.782148588$. Their hidden probability density functions are shown in Figure 5. We can deduce from the figure that regardless of how the probability measure changes, the probability density function always remains convex. Therefore, $Z_9 \prec Z_{18} \prec Z_{33} \prec Z_{36}$.

When the probability measure value of s_{i_k} decreases while its membership function remains unchanged, the variance of the normal distribution will decrease to reduce the probability density near the mean. This results in an increase in the probability density at locations farther away from the mean. If s_{i_k} is a positive restriction fuzzy linguistic term, the probability density of all negative restriction fuzzy linguistic terms will increase and the expectation value on $[0, 1]$ will decrease. We can draw the conclusion that

- if $i_1 = i_2$, s_{i_1}, s_{i_2} are positive, $m/2 \leq i_1, i_2 \leq m$, and $j_1 > j_2$, then $(s_{i_1}, r_{j_1}) \succ (s_{i_2}, r_{j_2})$;
- if $i_1 = i_2$, s_{i_1}, s_{i_2} are positive, $m/2 \leq i_1, i_2 \leq m$, and $j_1 < j_2$, then $(s_{i_1}, r_{j_1}) \prec (s_{i_2}, r_{j_2})$.

Example 4. Four Z-number linguistic terms $Z_{29} = (s_4, r_1)$, $Z_{30} = (s_4, r_2)$, $Z_{31} = (s_4, r_3)$ and $Z_{32} = (s_4, r_5)$ shown in Figures 2 and 3, their hidden probability distributions are:

$$\begin{aligned} p_{29}(x) &= \frac{1}{\sqrt{2\pi}\sigma_{29}} \exp\left(-\frac{(x - \frac{2}{3})^2}{2\sigma_{29}^2}\right), \\ p_{30}(x) &= \frac{1}{\sqrt{2\pi}\sigma_{30}} \exp\left(-\frac{(x - \frac{2}{3})^2}{2\sigma_{30}^2}\right), \\ p_{31}(x) &= \frac{1}{\sqrt{2\pi}\sigma_{31}} \exp\left(-\frac{(x - \frac{2}{3})^2}{2\sigma_{31}^2}\right), \\ p_{32}(x) &= \frac{1}{\sqrt{2\pi}\sigma_{32}} \exp\left(-\frac{(x - \frac{2}{3})^2}{2\sigma_{32}^2}\right). \end{aligned}$$

When $Prob_{29} = \frac{1}{6}$, $\sigma_{29} = 0.029439$ and then $E(29)_{[0,1]} = 0.656055633$. When $Prob_{30} = \frac{1}{3}$, $\sigma_{30} = 0.019185$ and then $E(30)_{[0,1]} = 0.663588618$. When $Prob_{31} = \frac{1}{2}$, $\sigma_{31} = 0.010686$ and then $E(31)_{[0,1]} = 0.666438699$. When $Prob_{32} = \frac{2}{3}$, $\sigma_{32} = 0.004654$ and then $E(32)_{[0,1]} = 0.666666488$. Their hidden probability density functions are shown in Figure 6. We can observe that as the probability measure of s_4 increases, the probability of negative terms decreases accordingly. Therefore, $Z_{29} \prec Z_{30} \prec Z_{31} \prec Z_{32}$.

When the probability measure value of s_{i_k} decreases while its membership function remains unchanged, the variance of the normal distribution will decrease to reduce the probability density near the mean. This results in an increase in the probability density at locations farther away from the mean. If s_{i_k} is a negative restriction fuzzy linguistic term, the probability density of all positive restriction fuzzy linguistic terms will increase and the expectation value on $[0, 1]$ will increase. We can draw the conclusion that

- if $i_1 = i_2$, s_{i_1}, s_{i_2} are negation, $0 \leq i_1, i_2 < m/2$, and $j_1 > j_2$, then $(s_{i_1}, r_{j_1}) \prec (s_{i_2}, r_{j_2})$;
- if $i_1 = i_2$, s_{i_1}, s_{i_2} are negation, $0 \leq i_1, i_2 < m/2$, and $j_1 < j_2$, then $(s_{i_1}, r_{j_1}) \succ (s_{i_2}, r_{j_2})$;

Example 5. Four Z-number linguistic terms $Z_{12} = (s_1, r_1)$, $Z_{11} = (s_1, r_2)$, $Z_{10} = (s_1, r_3)$ and $Z_9 = (s_1, r_4)$ shown in Figures 2 and 3, their hidden probability distributions are:

$$\begin{aligned} p_{12}(x) &= \frac{1}{\sqrt{2\pi}\sigma_{12}} \exp\left(-\frac{(x - \frac{1}{6})^2}{2\sigma_{12}^2}\right), \\ p_{11}(x) &= \frac{1}{\sqrt{2\pi}\sigma_{11}} \exp\left(-\frac{(x - \frac{1}{6})^2}{2\sigma_{11}^2}\right), \\ p_{10}(x) &= \frac{1}{\sqrt{2\pi}\sigma_{10}} \exp\left(-\frac{(x - \frac{1}{6})^2}{2\sigma_{10}^2}\right), \end{aligned}$$

$$p_9(x) = \frac{1}{\sqrt{2\pi}\sigma_9} \exp\left(-\frac{(x - \frac{1}{6})^2}{2\sigma_9^2}\right).$$

When $Prob_{12} = \frac{1}{6}$, $\sigma_{12} = 0.029439$ and then $E(12)_{[0,1]} = 0.217851412$. When $Prob_{11} = \frac{1}{3}$, $\sigma_{11} = 0.019185$ and then $E(11)_{[0,1]} = 0.196921259$. When $Prob_{10} = \frac{1}{2}$, $\sigma_{10} = 0.010686$ and then $E(10)_{[0,1]} = 0.178545009$. When $Prob_9 = \frac{2}{3}$, $\sigma_9 = 0.004654$ and then $E(9)_{[0,1]} = 0.168053961$. Their hidden probability density functions are shown in Figure 7. We can observe that as the probability measure of s_1 increases, the probability of positive terms decreases accordingly. Therefore, $Z_9 \prec Z_{10} \prec Z_{11} \prec Z_{12}$.

3.3 Operators

Definition 9 (Negation operator). We define the negation operator over Z-number linguistic term as:

$$Neg(Z_i) = Z_{(m+1) \times (n+1) - 1 - i}. \quad (12)$$

Aggregation operators are often used to integrate different pieces of information and then derive the combined information. There are many aggregation operators of Z-numbers in the existing literature.

- (1) Z linguistic combined weighted averaging aggregation (ZLCWAA) operator in [33]:

$$\begin{aligned} ZLCWAA(Z_1, Z_2, \dots, Z_n) = (< [\sum_{i=1}^n \omega_i a_i, \sum_{i=1}^n \omega_i b_i, \sum_{i=1}^n \omega_i c_i, \sum_{i=1}^n \omega_i d_i], \\ & [\frac{\sum_{i=1}^n \omega_i \|\alpha_i\| \mu_a^i}{\sum_{i=1}^n \omega_i \|\alpha_i\|}, \frac{\sum_{i=1}^n \omega_i \|\alpha_i\| \mu_b^i}{\sum_{i=1}^n \omega_i \|\alpha_i\|}, \frac{\sum_{i=1}^n \omega_i \|\alpha_i\| \mu_c^i}{\sum_{i=1}^n \omega_i \|\alpha_i\|}, \frac{\sum_{i=1}^n \omega_i \|\alpha_i\| \mu_d^i}{\sum_{i=1}^n \omega_i \|\alpha_i\|}] >, \quad (13) \\ & (\frac{\sum_{i=1}^n e_i}{n}, \frac{\sum_{i=1}^n f_i}{n}, \frac{\sum_{i=1}^n g_i}{n}, \frac{\sum_{i=1}^n h_i}{n}; \min_i(w_{B_i}))), \end{aligned}$$

where $Z_i = (< [a_i, b_i, c_i, d_i], [\mu_a^i, \mu_b^i, \mu_c^i, \mu_d^i] >, (e_i, f_i, g_i, h_i; w_{b_i}))$. $< [a_i, b_i, c_i, d_i], [\mu_a^i, \mu_b^i, \mu_c^i, \mu_d^i] >$ is a discrete fuzzy number and $(e_i, f_i, g_i, h_i; w_{b_i})$ is a trapezoidal fuzzy number. $\|\alpha_i\| = (a_i + b_i + c_i + d_i)/4$ and ω_i is the weight of Z_i .

- (2) The generalized normal Z^+ -values power weighted average (GNZPWA) in [34]:

$$\begin{aligned} GNZPWA(z_1, z_2, \dots, z_n) = & (((\sum_{i=1}^n \omega_i \mu_i^\lambda)^{\frac{1}{\lambda}}), \\ & (\sum_{i=1}^n \omega_i \mu_i^\lambda)^{\frac{1}{\lambda}-1} \sqrt{\sum_{i=1}^n \omega_i \mu_i^{2\lambda-2} \sigma_i^2}, (\frac{\sum_{i=1}^n \omega_i \mu_i^\lambda e x_i^\lambda}{\sum_{i=1}^n \omega_i \mu_i^\lambda})^{\frac{1}{\lambda}}), \quad (14) \\ & ((\frac{\sum_{i=1}^n \omega_i \mu_i^\lambda e x_i^\lambda}{\sum_{i=1}^n \omega_i \mu_i^\lambda})^{\frac{1}{\lambda}-1} \sqrt{\frac{\sum_{i=1}^n \omega_i \mu_i^\lambda e x_i^{2\lambda-1} e n_i^2}{\sum_{i=1}^n \omega_i \mu_i^\lambda}}, \\ & (\frac{\sum_{i=1}^n \omega_i \mu_i^\lambda e x_i^\lambda}{\sum_{i=1}^n \omega_i \mu_i^\lambda})^{\frac{1}{\lambda}-1} \sqrt{\frac{\sum_{i=1}^n \omega_i \mu_i^\lambda e x_i^{2\lambda-1} h e_i^2}{\sum_{i=1}^n \omega_i \mu_i^\lambda}}), \end{aligned}$$

where Z-numbers are invoked into normal Z^+ -values, $z_i = ((\mu_i, \sigma), (ex_i, en_i, he_i))$ and ω_i is the weight of z_i .

- (3) The Z-number generalised weighted average (ZGWA) operator in [35]:

$$\begin{aligned}
ZGWA^q(Z_1, Z_2, \dots, Z_n) &= (\oplus_{i=1}^n \omega_i Z_i^q)^{\frac{1}{q}} \\
&= ((\sum_{i=1}^n \omega_i a_{ij}^q)^{\frac{1}{q}}, g^{-1}(\frac{1}{q} g(h^{-1}(\sum_{i=1}^n \omega_i h(g^{-1}(qg(\mu_A(a_{ij}))))))), \\
&\quad g^{-1}(\frac{1}{q} g(h^{-1}(\sum_{i=1}^n \omega_i h(g^{-1}(qg(b_{ij})))))), g^{-1}(\frac{1}{q} g(h^{-1}(\sum_{i=1}^n \omega_i h(g^{-1}(qg(\mu_B(b_{ij}))))))), \\
&\quad g^{-1}(\frac{1}{q} g(h^{-1}(\sum_{i=1}^n \omega_i h(g^{-1}(qg(F_{il}(a_{ij}))))))),
\end{aligned} \tag{15}$$

where Z-numbers are discrete, $Z_i = ((a_{i1}, a_{i2}, \dots, a_{is}), (\mu_A(a_{i1}), \mu_A(a_{i2}), \dots, \mu_A(a_{is})), (b_{i1}, b_{i2}, \dots, b_{it}), (\mu_B(b_{i1}), \mu_B(b_{i2}), \dots, \mu_B(b_{it})), (F_{i1}, F_{i2}, \dots, F_{it}))$. F_{il} is the cumulative distribution function with respect to b_{il} , $l = 1, 2, \dots, t$. However, in order to obtain a cumulative probability function, it is necessary to solve a linear optimization problem.

Definition 10 (Arithmetic average aggregation operator). Let $\Gamma = \{Z_0, Z_1, \dots, Z_{(m+1) \times (n+1) - 1}\}$ be a Z-number linguistic term set. Consider a set of t Z-number linguistic terms $T = \{Z_{\phi(1)}, Z_{\phi(2)}, \dots, Z_{\phi(t)}\}$ to be aggregated, where $t > 0$ and $Z_{\phi(i)} \in \Gamma$ for $i \in \{1, 2, \dots, t\}$. Here, $\phi(\cdot)$ is the index function. The arithmetic average for Z-number linguistic term $\bar{Z}$ of T is defined as:

$$\bar{Z} = Z_{\text{round}(\sum_{i=1}^t \frac{1}{t} \phi(i))}, \tag{16}$$

where $\text{round}(\cdot)$ is the usual round operation.

Example 6. Assume $m = 6$ and $n = 4$, $S = \{s_0, s_1, s_2, s_3, s_4, s_5, s_6\} = \{\text{absolutely bad, very bad, bad, medium, good, very good, absolutely good}\}$ be the restriction fuzzy linguistic term set. $R = \{r_0, r_1, r_2, r_3, r_4\} = \{\text{very not sure, not sure, medium, sure, very sure}\}$ be the reliability fuzzy linguistic term set. According to the definition of Z-number linguistic term set and the monotonicity of Z-number linguistic terms, we can get the Z-number linguistic term set Γ :

$$\begin{aligned}
\Gamma &= \{Z_0, Z_1, \dots, Z_{34}\} \\
&= \{(s_0, r_4), (s_0, r_3), (s_0, r_2), (s_0, r_1), (s_0, r_0), \\
&\quad (s_1, r_4), (s_1, r_3), (s_1, r_2), (s_1, r_1), (s_1, r_0), \\
&\quad (s_2, r_4), (s_2, r_3), (s_2, r_2), (s_2, r_1), (s_2, r_0), \\
&\quad (s_3, r_0), (s_3, r_1), (s_3, r_2), (s_3, r_3), (s_3, r_4), \\
&\quad (s_4, r_0), (s_4, r_1), (s_4, r_2), (s_4, r_3), (s_4, r_4), \\
&\quad (s_5, r_0), (s_5, r_1), (s_5, r_2), (s_5, r_3), (s_5, r_4), \\
&\quad (s_6, r_0), (s_6, r_1), (s_6, r_2), (s_6, r_3), (s_6, r_4)\}.
\end{aligned}$$

Consider two Z-number linguistic terms $T = \{Z_{\phi(1)}, Z_{\phi(2)}\} = \{(s_1, r_4), (s_5, r_1)\}$ to be aggregated. According to Eq. (7), $\phi(1) = 5$ and $\phi(2) = 26$. Thence, the arithmetic mean Z-number linguistic term $\bar{Z}$ of T is:

$$\bar{Z} = Z_{\text{round}(\frac{1}{2} \times 5 + \frac{1}{2} \times 26)} = Z_{16} = (s_3, r_1).$$

Definition 11 (Weighted average aggregation operator over Z-number linguistic terms). Let $\Gamma = \{Z_0, Z_1, \dots, Z_{(m+1) \times (n+1) - 1}\}$ be a Z-number linguistic term set. Consider a set of t Z-number linguistic terms $T = \{Z_{\phi(1)}, Z_{\phi(2)}, \dots, Z_{\phi(t)}\}$ to be aggregated, where $t > 0$ and $Z_{\phi(i)} \in \Gamma$ for $i \in \{1, 2, \dots, t\}$. Here, $\phi(\cdot)$ is the index function. $W = \{w_1, w_2, \dots, w_t\}$ be a associated weight set and satisfies $w_i \in [0, 1]$ and $\sum_{i=1}^t w_i = 1$. The weighted Z-number linguistic term $\bar{Z}$ of T is defined as:

$$\bar{Z} = Z_{\text{round}(\sum_{i=1}^t w_i \phi(i))}. \quad (17)$$

Example 7. Using the Z-number linguistic term set Γ and T in Example 6, assume the weights set is $W = \{w_1, w_2\} = \{\frac{1}{4}, \frac{3}{4}\}$. The weighted Z-number linguistic term $\bar{Z}$ of T is:

$$\bar{Z} = Z_{\text{round}(\frac{1}{4} \times 5 + \frac{3}{4} \times 26)} = Z_{21} = (s_4, r_1).$$

Definition 12 (Ordered weighted aggregation operator over Z-number linguistic terms). Let $\Gamma = \{Z_0, Z_1, \dots, Z_{(m+1) \times (n+1) - 1}\}$ be a Z-number linguistic term set. Consider a set of t Z-number linguistic terms $T = \{Z_{\phi(1)}, Z_{\phi(2)}, \dots, Z_{\phi(t)}\}$ to be aggregated, where $t > 0$ and $Z_{\phi(i)} \in \Gamma$ for $i \in \{1, 2, \dots, t\}$. Here, $\phi(\cdot)$ is the index function. $W = \{w_1, w_2, \dots, w_t\}$ be associated weights and satisfies $w_i \in [0, 1]$, $\sum_{i=1}^t w_i = 1$. The ordered weighted Z-number linguistic term $\bar{Z}$ of T is defined as:

$$\bar{Z} = Z_{\text{round}(\sum_{i=1}^t w_i \phi^*(i))}, \quad (18)$$

where $\phi^*(i)$ is the i -th largest of $\phi(i)$ values.

Example 8. Using the Z-number linguistic term set Γ and T in Example 6, assume the weights set is $W = \{w_1, w_2\} = \{\frac{1}{4}, \frac{3}{4}\}$. The ordered weighted Z-number linguistic term $\bar{Z}$ of T is:

$$\bar{Z} = Z_{\text{round}(\frac{1}{4} \times 26 + \frac{3}{4} \times 5)} = Z_{10} = (s_2, r_4).$$

From the expression, it is evident how much simpler and clearer our proposed aggregation operators are compared to general computation of Z-numbers and Eqs. (13)-(15).

4 The MCGDM model based on Z-number linguistic term set

Once the Z-number linguistic term set have been defined, we demonstrate their utility and flexibility by incorporating them into a MCGDM framework in this section. The proposed Z-number linguistic term set offers a straightforward and efficient approach for decision-making, enhancing both the simplicity and effectiveness of the decision-making process.

For a MCGDM problem with Z-number linguistic terms, let $A = \{a_1, a_2, \dots, a_l\}$ be a set of l alternatives and let $C = \{c_1, c_2, \dots, c_h\}$ be a set of h criteria with weights vector $W_c = \{w_{c,1}, w_{c,2}, \dots, w_{c,h}\}$, $w_{c,i} \in [0, 1]$ and $\sum_{i=1}^h w_{c,i} = 1$. Assume there are g decision makers $D = \{d_1, d_2, \dots, d_g\}$ with weights vector $W_d = \{w_{d,1}, w_{d,2}, \dots, w_{d,g}\}$ where $w_{d,i} \in [0, 1]$ and $\sum_{i=1}^g w_{d,i} = 1$. The goal is to determine the rank of alternatives.

4.1 Information preprocessing

In real-life decision-making scenarios, the assessment information may not always be presented as Z-number linguistic terms. Therefore, it is essential to convert the evaluation information provided by decision-makers into the Z-number linguistic term format to effectively utilize the proposed framework.

The most straightforward approach is to predefine both the restriction fuzzy linguistic term set and the reliability fuzzy linguistic term set. During the assessment process, decision-makers select a linguistic term from the restriction fuzzy linguistic term set to evaluate an alternative based on a specific criterion. They then choose a linguistic term from the reliability fuzzy linguistic term set to represent their confidence in the assessment.

Example 9. *Regarding the originality of the academic paper, there are five assessment options: {Very clear, Understandable, Mostly understandable, Hard to resolve, Confusing}. Additionally, there are five options representing the reviewers' confidence levels in their assessments of the originality of the academic paper: {Definitely correct, Quite sure, Pretty sure, Can't be sure, Not my area}. If a reviewer is a highly renowned expert in the field of this academic paper, then he or she might choose "Definitely correct" in the reviewer's confidence level.*

If decision-makers do not directly provide their confidence levels or assessment reliability, we can infer the reliability of the assessments using available information, such as the degree of hesitancy and similarity measures [36].

Example 10. *Using the restriction and reliability fuzzy linguistic term set shown in Figure 2 and Figure 3. Assume an expert's evaluation of the efficiency of a new energy source is "Very High". The similarity between this evaluation and evaluations from other experts is 0.8. Then we can use "Sure" to describe the reliability of this evaluation. Therefore, we can use the Z-number linguistic term (Very High, Sure) to evaluate the efficiency of the new energy source.*

4.2 MCGDM procedures

To conduct the MCGDM process, the following steps are followed:

- Step 1: Determine the restriction fuzzy linguistic term sets $S = \{s_0, s_1, \dots, s_m\}$ and the reliability fuzzy linguistic term set $R = \{r_0, r_1, \dots, r_n\}$ and obtain the Z-number linguistic term set Γ . The order index function $I(x, y)$ of a Z-number linguistic term

(s_x, r_y) , $x = 0, 1, \dots, m$, $y = 0, 1, \dots, n$, is

$$I(x, y) = \begin{cases} x(n+1) + y, & \text{if } s_x \text{ is positive;} \\ x(n+1) + m - y, & \text{if } s_x \text{ is negative.} \end{cases} \quad (19)$$

Step 2: Obtain the Z-number linguistic term decision matrix of each decision-maker.

The decision matrix of decision maker d_k is shown as Table 4. The Z-number linguistic term $Z_{\phi(i,j,k)} = (s_{i,j,k}, r_{i,j,k})$ represents the rating given by decision maker d_k for alternative a_i under criteria c_j , where $s_{i,j,k} \in S$, $r_{i,j,k} \in R$. The function $\phi(i, j, k)$ is the index function.

	c_1	c_2	$\dots$	c_j	$\dots$	c_h
a_1	$Z_{\phi(1,1,k)}$	$Z_{\phi(1,2,k)}$	$\dots$	$Z_{\phi(1,j,k)}$	$\dots$	$Z_{\phi(1,h,k)}$
a_2	$Z_{\phi(2,1,k)}$	$Z_{\phi(2,2,k)}$	$\dots$	$Z_{\phi(2,j,k)}$	$\dots$	$Z_{\phi(2,h,k)}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
a_i	$Z_{\phi(i,1,k)}$	$Z_{\phi(i,2,k)}$	$\dots$	$Z_{\phi(i,j,k)}$	$\dots$	$Z_{\phi(i,h,k)}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
a_l	$Z_{\phi(l,1,k)}$	$Z_{\phi(l,2,k)}$	$\dots$	$Z_{\phi(l,j,k)}$	$\dots$	$Z_{\phi(l,h,k)}$

Table 4 Decision matrix of d_k .

Step 3: Calculate the weights of criteria and decision-makers.

Assuming that the importance Z-number linguistic term set is Γ' , which is determined by restriction fuzzy linguistic term set $S' = \{s'_1, s'_2, \dots, s'_{m'}\}$ and reliability fuzzy linguistic term set $R' = \{r'_1, r'_2, \dots, r'_{n'}\}$. Then, the importance of criteria is evaluated by experts as:

$$C_w = [Z'_{c,\phi_c(1)}, Z'_{c,\phi_c(2)}, \dots, Z'_{c,\phi_c(j)}, \dots, Z'_{c,\phi_c(h)}], \quad (20)$$

where $Z'_{c,\phi_c(j)} = (s'_p, r'_q)$ is the importance Z-number linguistic term of criteria c_j and $s'_p \in S'$, $r'_q \in R'$, $0 \leq p \leq m'$, $0 \leq q \leq n'$. $\phi_c(j)$ is the index function.

The importance of decision-makers is evaluated by experts as:

$$D_w = [Z'_{d,\phi_d(1)}, Z'_{d,\phi_d(2)}, \dots, Z'_{d,\phi_d(k)}, \dots, Z'_{d,\phi_d(g)}], \quad (21)$$

where $Z'_{d,\phi_d(j)} = (s'_p, r'_q)$ is the importance Z-number linguistic term of criteria c_j and $s'_p \in S'$, $r'_q \in R'$, $0 \leq p \leq m'$, $0 \leq q \leq n'$. $\phi_d(k)$ is the index function.

The weight of criteria c_j is defined as:

$$w_{c,j} = \frac{\phi_c(j)}{\sum_{i=1}^h \phi_c(i)}. \quad (22)$$

The weight of the decision maker d_k is defined as:

$$w_{d,k} = \frac{\phi_d(k)}{\sum_{i=1}^g \phi_d(i)}. \quad (23)$$

Step 4: Calculate the aggregated Z-number linguistic term of alternative by using weighted average aggregation operator.

The aggregated Z-number linguistic term of alternative a_i is defined as:

$$\overline{Z}_i = Z_{\sum_{k=1}^g w_{d,k} \sum_{j=1}^h w_{c,j} \phi(i,j,k)}. \quad (24)$$

Step 5: Ranking the alternatives based on the order index of aggregated Z-number linguistic terms.

This model first determines the ordering information for all ratings, then weights and aggregates this information. This process results in a clear understanding of the relative size relationships among the alternatives. Both fuzzy information and Z-number-based information are qualitative in nature, and our primary focus is on their ordering relationships rather than their numerical values [37]. Establishing the ordering relationship of Z-number linguistic terms before performing calculations significantly reduces computational complexity compared to ordering results after complex Z-number calculations. Furthermore, performing simple calculations on ordering information does not alter the size order between them.

5 MCGDM application for predicting the acceptance of academic papers

In this section, we apply the designed MCGDM method to predict the acceptance of academic papers using the ACL 2017 dataset from the PeerRead dataset, which provides peer review scores for various aspects [38]. The dataset consists of 137 papers and 275 reviews, with each review containing scores for nine aspects: "impact", "substance", "meaningful comparison", "appropriateness", "soundness correctness", "originality", "recommendation", "clarity", and "reviewer confidence", as shown in Table 5. Among the papers in the dataset, 99 were accepted.

Aspect	5	4	3	2	1
Impact	Enabling	Useful	Potentially useful	Documentary	No usable
Meaningful comparison	Precise and complete	Mostly solid	Somewhat helpful	Only partial awareness	Little awareness
Substance	More than most publications	Appropriate	Leaves open questions	Work in progress	Seems thin
Appropriateness	Certainly	Probably	Unsure	Probably not	Certainly not
Soundness correctness	Very apt	Generally solid work	Fairly reasonable work	Troublesome	Fatally flawed
Originality	Surprising	Creative	Respectable	Pedestrian	Have been done
Recommendation	I'd fight to get it accepted	I'd like to see it accepted	Borderline	Leaning against	Poor
Clarity	Very clear	Understandable	Mostly understandable	Hard to resolve	Confusing
Reviewer confidence	Definitely correct	Quite sure	Pretty sure	Can't be sure	Not my area

Table 5 How ACL 2017 reviewers assign aspect scores to reviewed papers.

Fuzzy numbers provide a more appropriate way to express aspect scores compared to crisp numbers. For example, fuzzy terms such as "certainly", "probably",

"unsure", "probably not", and "certainly not" offer natural representations. Among these aspects, "reviewer confidence" reflects the confidence level of the reviewer's scores and comments. Therefore, the scores of an aspect (excluding "reviewer confidence") and "reviewer confidence" can be combined to form a Z-number linguistic term. For instance, if a reviewer assigns a score of "probably" for "appropriateness" and "quite sure" for "reviewer confidence", these two scores form a Z-number linguistic term for "appropriateness" as ("probably", "quite sure"). Since there are nine aspects with 5-level scores in the ACL 2017 dataset, we can obtain eight different Z-number linguistic terms for each reviewer of a paper. The cardinality of the fuzzy linguistic term set is determined by the number of score levels for each aspect. By aggregating these Z-number linguistic terms, we can derive the reviewer's overall evaluation of a paper. Subsequently, by aggregating the total evaluations of all reviewers, we can obtain the overall evaluation of the paper. Finally, based on this ranking and the expected number of accepted papers, a decision can be made to accept or reject the paper.

5.1 Numerical example

To demonstrate our method, we select four submitted academic papers from the ACL 2017 dataset: paper #768 (a_1), paper #178 (a_2), paper #355 (a_3), and paper #288 (a_4). Assume the conference aims to select the best one for publication, employing three reviewers: reviewer #1 (d_1), reviewer #2 (d_2), and reviewer #3 (d_3). These reviewers evaluate the academic papers according to eight aspects: impact (c_1), meaningful comparison (c_2), substance (c_3), appropriateness (c_4), soundness and correctness (c_5), originality (c_6), recommendation (c_7), and clarity (c_8), along with their confidence levels. Reviewers use 5-level scores, as shown in Table 5, to assess the quality of each paper under each aspect.

Step 1: Determine the restriction fuzzy linguistic term set $S = \{s_0, s_1, \dots, s_m\}$ and the reliability fuzzy linguistic term set $R = \{r_0, r_1, \dots, r_n\}$ and obtain the Z-number linguistic term set Γ .

Based on the Table 5, we can see that each aspect has five levels of evaluation. The restriction fuzzy linguistic term set for each aspect (excluding "reviewer confidence") is $S = \{s_0, s_1, s_2, s_3, s_4\}$ and the reliability fuzzy linguistic term set is $R = \{r_0, r_1, r_2, r_3, r_4\}$. We can obtain the Z-number linguistic term set Γ according to S and R :

$$\begin{aligned}\Gamma &= \{Z_0, Z_1, Z_2, \dots, Z_{24}\} \\ &= \{(s_0, r_4), (s_0, r_3), (s_0, r_2), (s_0, r_1), (s_0, r_0), \\ &\quad (s_1, r_4), (s_1, r_3), (s_1, r_2), (s_1, r_1), (s_1, r_0), \\ &\quad (s_2, r_4), (s_2, r_3), (s_2, r_2), (s_2, r_1), (s_2, r_0), \\ &\quad (s_3, r_0), (s_3, r_1), (s_3, r_2), (s_3, r_3), (s_3, r_4), \\ &\quad (s_4, r_0), (s_4, r_1), (s_4, r_2), (s_4, r_3), (s_4, r_4)\}.\end{aligned}$$

Step 2: Obtain the Z-number linguistic term decision matrix of each decision-maker.

According to the evaluation criteria in Table 5, reviewers rated four academic papers in nine aspects. The review reports of each paper are shown in Tables 6 and 7. The corresponding Z-number linguistic terms decision matrices are shown in Tables 8-10.

Aspect	Paper #768			Paper #178		
	reviewer #1	reviewer #2	reviewer #3	reviewer #1	reviewer #2	reviewer #3
Impact	3	3	3	3	3	3
Meaningful comparison	4	4	4	3	3	3
Substance	4	1	3	3	4	4
Appropriateness	5	5	5	4	5	5
Soundness correctness	3	3	3	4	3	3
Originality	4	4	4	3	2	3
Recommendation	4	2	2	2	2	2
Clarity	4	3	2	4	4	3
Reviewer confidence	2	5	4	3	4	4

Table 6 The review reports of paper #768 and paper #178.

Aspect	Paper #355			Paper #288		
	reviewer #1	reviewer #2	reviewer #3	reviewer #1	reviewer #2	reviewer #3
Impact	3	3	3	3	3	3
Meaningful comparison	3	3	3	3	3	3
Substance	4	4	4	4	2	4
Appropriateness	5	5	5	5	4	5
Soundness correctness	5	5	5	5	5	5
Originality	5	5	5	5	5	5
Recommendation	4	4	4	2	2	5
Clarity	4	4	4	2	3	5
Reviewer confidence	3	5	4	4	4	4

Table 7 The review reports of paper #355 and paper #288.

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8
a_1	$Z_{11} = (s_2, r_1)$	$Z_{16} = (s_3, r_1)$	$Z_{16} = (s_3, r_1)$	$Z_{21} = (s_4, r_1)$	$Z_{11} = (s_2, r_1)$	$Z_{16} = (s_3, r_1)$	$Z_{16} = (s_3, r_1)$	$Z_{16} = (s_3, r_1)$
a_2	$Z_{12} = (s_2, r_2)$	$Z_{12} = (s_2, r_2)$	$Z_{12} = (s_2, r_2)$	$Z_{17} = (s_3, r_2)$	$Z_{17} = (s_3, r_2)$	$Z_{12} = (s_2, r_2)$	$Z_7 = (s_1, r_2)$	$Z_{17} = (s_3, r_2)$
a_3	$Z_{12} = (s_2, r_2)$	$Z_{12} = (s_2, r_2)$	$Z_{17} = (s_3, r_2)$	$Z_{22} = (s_4, r_2)$	$Z_{22} = (s_4, r_2)$	$Z_{22} = (s_4, r_2)$	$Z_{17} = (s_3, r_2)$	$Z_{17} = (s_3, r_2)$
a_4	$Z_{13} = (s_2, r_3)$	$Z_{13} = (s_2, r_3)$	$Z_{18} = (s_3, r_3)$	$Z_{23} = (s_4, r_3)$	$Z_{23} = (s_4, r_3)$	$Z_{23} = (s_4, r_3)$	$Z_6 = (s_1, r_3)$	$Z_6 = (s_1, r_3)$

Table 8 The Z-number linguistic terms decision matrix of d_1 .

Step 3: Calculate the weights of criteria and decision-makers. Firstly, determine the restriction fuzzy linguistic term set $S' = \{s'_1, s'_2, \dots, s'_{m'}\}$ and reliability fuzzy linguistic term set $R' = \{r'_1, r'_2, \dots, r'_{n'}\}$ of the importance and obtain the Z-number linguistic term set Γ' of importance.

The importance restriction fuzzy linguistic term set is $S' = \{s'_0 : \text{very unimportant}, s'_1 : \text{unimportant}, s'_2 : \text{medium}, s'_3 : \text{Importance}, s'_4 : \text{Very Important}\}$ and the importance reliability fuzzy linguistic term set $R' = \{r'_0 : \text{very not sure}, s'_1 : \text{not sure}, s'_2 : \text{medium}, s'_3 : \text{sure}, s'_4 : \text{very sure}\}$.

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8
a_1	$Z_{14} = (s_2, r_4)$	$Z_{19} = (s_3, r_4)$	$Z_0 = (s_0, r_4)$	$Z_{24} = (s_4, r_4)$	$Z_{14} = (s_2, r_4)$	$Z_{19} = (s_3, r_4)$	$Z_9 = (s_1, r_4)$	$Z_{10} = (s_2, r_4)$
a_2	$Z_{13} = (s_2, r_3)$	$Z_{13} = (s_2, r_3)$	$Z_{18} = (s_3, r_3)$	$Z_{23} = (s_4, r_3)$	$Z_{13} = (s_2, r_3)$	$Z_6 = (s_1, r_3)$	$Z_6 = (s_1, r_3)$	$Z_{18} = (s_3, r_3)$
a_3	$Z_{14} = (s_2, r_4)$	$Z_{14} = (s_2, r_4)$	$Z_{19} = (s_3, r_4)$	$Z_{24} = (s_4, r_4)$	$Z_{24} = (s_4, r_4)$	$Z_{24} = (s_4, r_4)$	$Z_{19} = (s_3, r_4)$	$Z_{19} = (s_3, r_4)$
a_4	$Z_{13} = (s_2, r_3)$	$Z_{13} = (s_2, r_3)$	$Z_6 = (s_1, r_3)$	$Z_{18} = (s_3, r_3)$	$Z_{23} = (s_4, r_3)$	$Z_{23} = (s_4, r_3)$	$Z_6 = (s_1, r_3)$	$Z_{13} = (s_2, r_3)$

Table 9 The Z-number linguistic terms decision matrix of d_2 .

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8
a_1	$Z_{13} = (s_2, r_3)$	$Z_{18} = (s_3, r_3)$	$Z_{13} = (s_2, r_3)$	$Z_{23} = (s_4, r_3)$	$Z_{13} = (s_2, r_3)$	$Z_{18} = (s_3, r_3)$	$Z_6 = (s_1, r_3)$	$Z_6 = (s_1, r_3)$
a_2	$Z_{13} = (s_2, r_3)$	$Z_{18} = (s_3, r_3)$	$Z_{18} = (s_3, r_3)$	$Z_{23} = (s_4, r_3)$	$Z_{13} = (s_2, r_3)$	$Z_{13} = (s_3, r_3)$	$Z_6 = (s_1, r_3)$	$Z_{13} = (s_2, r_3)$
a_3	$Z_{13} = (s_2, r_3)$	$Z_{13} = (s_2, r_3)$	$Z_{18} = (s_3, r_3)$	$Z_{23} = (s_4, r_3)$	$Z_{23} = (s_4, r_3)$	$Z_{23} = (s_4, r_3)$	$Z_{18} = (s_3, r_3)$	$Z_{18} = (s_3, r_3)$
a_4	$Z_{13} = (s_2, r_3)$	$Z_{13} = (s_2, r_3)$	$Z_{18} = (s_3, r_3)$	$Z_{23} = (s_4, r_3)$	$Z_{23} = (s_4, r_3)$	$Z_{23} = (s_4, r_3)$	$Z_{23} = (s_4, r_3)$	$Z_{23} = (s_4, r_3)$

Table 10 The Z-number linguistic terms decision matrix of d_3 .

The importance values of criteria and decision makers are:

$$\begin{aligned}
C_w &= [Z'_{c,\phi_c(1)}, Z'_{c,\phi_c(2)}, Z'_{c,\phi_c(3)}, Z'_{c,\phi_c(4)}, Z'_{c,\phi_c(5)}, Z'_{c,\phi_c(6)}, Z'_{c,\phi_c(7)}, Z'_{c,\phi_c(8)}] \\
&= [(unimportant, sure), (medium, medium), (medium, sure), \\
&\quad (important, sure), (unimportant, sure), (very unimportant, sure), \\
&\quad (very important, very sure), (important, medium)] \\
&= [Z'_6, Z'_{12}, Z'_{13}, Z'_{18}, Z'_6, Z'_1, Z'_{24}, Z'_{17}].
\end{aligned}$$

$$\begin{aligned}
D_w &= [Z'_{d,\phi_d(1)}, Z'_{d,\phi_d(2)}, Z'_{d,\phi_d(3)}] \\
&= [(important, sure), (very important, sure), (important, very sure)] \\
&= [Z'_{18}, Z'_{23}, Z'_{19}].
\end{aligned}$$

According to Equation (22) and (23), the weights of criteria and decision makers are: $w_{c,1} = \frac{6}{6+12+13+18+6+1+24+17} = 0.0619$, $w_{c,2} = \frac{12}{6+12+13+18+6+1+24+17} = 0.1234$, $w_{c,3} = \frac{13}{6+12+13+18+6+1+24+17} = 0.1340$, $w_{c,4} = \frac{18}{6+12+13+18+6+1+24+17} = 0.1856$, $w_{c,5} = \frac{6}{6+12+13+18+6+1+24+17} = 0.0619$, $w_{c,6} = \frac{1}{6+12+13+18+6+1+24+17} = 0.0103$, $w_{c,7} = \frac{24}{6+12+13+18+6+1+24+17} = 0.2474$, $w_{c,8} = \frac{17}{6+12+13+18+6+1+24+17} = 0.1753$. $w_{d,1} = \frac{18}{18+23+19} = 0.3$, $w_{d,2} = \frac{23}{18+23+19} = 0.3833$, $w_{d,3} = \frac{19}{18+23+19} = 0.3167$.

Step 4: Calculate the aggregated Z-number linguistic term of alternative by using weighted average aggregation operator.

The aggregated Z-number linguistic terms of alternatives are:

$$\begin{aligned}
\overline{Z}_1 &= Z_{\sum_{k=1}^4 w_{d,k} \sum_{j=1}^3 w_{c,j} \phi(1,j,k)} \\
&= Z_{13.6307},
\end{aligned}$$

$$\overline{Z}_2 = Z_{\sum_{k=1}^4 w_{d,k} \sum_{j=1}^3 w_{c,j} \phi(2,j,k)}.$$

$$= Z_{13.8248},$$

$$\begin{aligned}\overline{Z}_3 &= Z_{\sum_{k=1}^4 w_{d,k} \sum_{j=1}^3 w_{c,j} \phi(3,j,k)} \\ &= Z_{18.4422},\end{aligned}$$

$$\begin{aligned}\overline{Z}_4 &= Z_{\sum_{k=1}^4 w_{d,k} \sum_{j=1}^3 w_{c,j} \phi(4,j,k)} \\ &= Z_{15.0606}.\end{aligned}$$

Step 5: Ranking the alternatives based on the order index of aggregated Z-number linguistic terms.

According to the aggregated Z-number linguistic terms, the ranking of submitted papers is as $a_3 \succ a_4 \succ a_2 \succ a_1$. Hence, paper #355 (a_3) is the best academic paper for publication. In fact, this paper was accepted while the other three papers were rejected. This example illustrates that our method does not require computing the membership functions or the hidden probability distributions of the Z-numbers. Consequently, there is no need to define the semantics or solve optimization problems to obtain the hidden probability distributions. This significantly simplifies the computational process and enhances the efficiency of decision-making.

5.2 Analysis and comparison

The proposed Z-number linguistic term construction offers an effective and straightforward approach to predict the acceptance of submitted academic papers. Using the ACL 2017 dataset, which does not provide information on aspect weights and reviewer weights, we applied an arithmetic mean aggregation operator, achieving a prediction accuracy of 73.72%. When assigning a weight of 0.4 to the criterion "recommendation" and 0.086 to the other criteria, the accuracy increases significantly to 86.86%. This demonstrates the robustness and potential of the proposed method for enhancing decision-making accuracy in academic paper evaluations.

When the weights of each aspect are varied from 0.0 to 1.0, with the weights of other aspects remaining equal, the prediction accuracy changes accordingly, as illustrated in Figure 8. The figure highlights "recommendation" as the most crucial aspect, as increasing its weight consistently leads to improved prediction accuracy. When the weight of "recommendation" is increased to 0.4, the accuracy peaks at 86.86%. Conversely, "soundness correctness" and "originality" are found to be less important for the acceptance of academic papers, with the highest accuracies occurring when their weights are set to 0. The aspects of "impact," "substance," "appropriateness," "meaningful comparison," and "clarity" influence the prediction results, with their accuracies increasing as their weights are increased, and then decreasing after reaching their peak. "Recommendation" is particularly important because it reflects the overall evaluation of the paper, integrating all the scores from the other aspects. In practice, journals or conferences often base their acceptance or rejection decisions primarily on

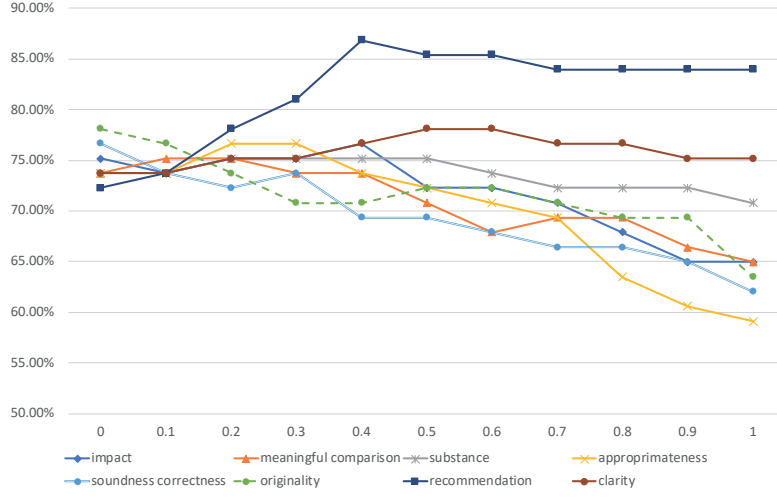


Fig. 8 Trend of accuracy when aspects weights change.

the "recommendation" score, as it provides a comprehensive summary of the paper's quality.

Several Z-number-based Multi-Criteria Group Decision-Making (MCGDM) methods have been proposed to address practical problems, such as Z-TOPSIS [39], COPRAS-Z methodology [40], the MCGDM method proposed by Xian et al. [33], Z-VIKOR method [41], and the MCGDM method proposed by Peng et al. [35]. However, these methods have primarily been tested on simulated problems and have not been validated on real-world datasets. We applied these methods, along with our proposed method, to predict the acceptance of academic papers using the ACL 2017 dataset. The results are detailed in Table 11. The semantics of Z-numbers used in five existing MCGDM methods are shown in Figures 2 and 3. Figures 2 and 3 illustrate the semantics of Z-numbers used in these MCGDM methods, with some methods utilizing continuous Z-numbers and others using discrete Z-numbers. For methods employing continuous Z-numbers, we discretized the membership functions before application. As shown in Table 11, our proposed method achieves the best prediction results on the ACL dataset when no additional weight information is provided. In terms of runtime, most methods exhibit similar performance, except for Peng et al.'s method [35], which has a significantly longer runtime due to numerous optimization operations required to compute hidden probability distributions. Compared to existing Z-number-based information processing methods, our proposed Z-number linguistic term set calculation is simpler and more efficient, providing a reasonable way to obtain decision results. This approach eliminates the need to define specific semantics for linguistic terms, focusing solely on their ordering relationships. This advancement has the potential to significantly expand the practical applications of Z-numbers.

In decision-making, the precise numerical values of alternatives are often less critical than their relative rankings, as decision-makers typically focus on identifying the

top alternatives. Unlike existing methods that rely on complex hidden probability distributions or convert Z-numbers into fuzzy numbers for decision-making, our approach simplifies the process by focusing on the ordering information of Z-number linguistic terms. Our method emphasizes determining the ordering relationships between alternatives rather than deriving specific Z-number values. By ranking all Z-number linguistic terms and analyzing their ordering information, subsequent steps such as aggregation become straightforward. This approach provides clear insights into the relative rankings of alternatives. Due to its straightforward calculation process and emphasis on intuitive ordering information, the multi-criteria group decision-making model based on the Z-number linguistic term set is highly suitable for large-scale decision-making scenarios. It is particularly advantageous for situations involving numerous criteria and a large number of decision-makers, offering a practical solution for managing complex group decision-making tasks.

Method	Processing method	Discrete or continuous	Accuracy	Running time
Z-TOPSIS [39]	Convert to fuzzy numbers	continuous	44.53%	0.1321
COPRAS-Z [40]	Convert to fuzzy numbers	continuous	69.34%	0.1103
Xian et al.'s method ($\beta = 0.5$) [33]	Computation with Z-numbers	mixture	63.50%	0.1218
Peng and Wang's method [34]	Cloud model	continuous	59.12%	0.1616
Peng et al.'s method [35]	Probability distribution	discrete	69.34%	1134.2576
Our method	Z-number linguistic term set	don't need semantics	73.72%	0.1095

Table 11 The clustering accuracy rates (%) of five existing MCGDM methods and our method.

In recent years, machine learning algorithms have demonstrated significant success in prediction tasks. In order to highlight the advantages of Z-number linguistic terms, eight commonly used machine learning clustering algorithms, including k-means, mini batch k-means, birch, spectral clustering, agglomerative clustering, density-based spatial clustering of applications with noise (DBSCAN), optics and mean shift are applied to the ACL 2017 dataset for binary clustering: accepted papers and rejected papers. The clustering accuracy and runtime are shown in the Table 12. If the method's parameters are not specifically indicated, the default parameters in scikit-learn (sklearn) are used. As shown in Table 12, our method achieves the best prediction results and nine methods have similar running times. We can conclude that our method is interpretable and more efficient compared to machine learning clustering algorithms.

Algorithm	Accuracy rate (%)	Running time (second)
k-means	58.39%	0.165092
Mini batch k-means	58.39%	0.1485
Birch	58.39%	0.1616
Spectral clustering	58.39%	0.4055
Agglomerative clustering	58.39%	0.1530
DBSCAN (eps=0.99,min_samples=9)	59.85%	0.1532
Optics (eps=0.99,min_samples=9)	57.66%	0.1491
Mean shift (bandwidth=None)	71.53%	0.2306
Our method	73.72%	0.1095

Table 12 The clustering accuracy rates (%) and runtimes (second) of clustering algorithms and our method.

6 Conclusions

The concept of Z-numbers provides a comprehensive framework for describing real-world information. However, their practical application can be hindered by computational complexity. This paper addresses this challenge by introducing a Z-number-based linguistic computational model, which includes the concepts of Z-number linguistic terms and Z-number linguistic term sets, alongside a computational model. This approach reduces the computational complexity associated with Z-numbers. Moreover, a MCGDM model that uses the Z-number linguistic terms has been presented. To validate the effectiveness and superiority of the proposed method, a comprehensive comparison is conducted with five existing Z-number-based MCGDM methods and eight machine learning clustering algorithms based on the ACL 2017 dataset. The results show that the proposed MCGDM method outperforms other methods. Compared to other Z-number MCGDM methods, our approach using Z-number linguistic terms offers a straightforward way to obtain decision results through simple numerical calculations. Furthermore, compared to machine learning clustering algorithms, our proposed MCGDM model is interpretable and more efficient.

It is important to note that the model we propose is expert-driven, meaning it relies on experts providing their judgments and evaluations. We will further explore data-driven models in the future.

Declarations

Acknowledgements

This research has been partially supported by grants from the National Natural Science Foundation of China (#71910107002), the China Scholarship Council (CSC), grant PID2022-139297OB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF/EU. Moreover, it is part of the project C-ING-165-UGR23, co-funded by the Regional Ministry of University, Research and Innovation and by the European Union under the Andalusia ERDF Programme 2021-2027.

Competing interests

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Authors contribution statement

Yangxue Li: conceptualization, methodology, formal analysis, software, validation, writing-original draft, investigation, data curation, visualization. **Gang Kou:** Writing - Review & Editing, visualization, supervision. **Yi Peng:** Writing - Review & Editing, visualization, supervision. **Juan Antonio Morente-Molinera:** Writing - Review & Editing, supervision, funding acquisition.

Ethical and informed consent

Not applicable.

Data availability

The data set analyzed during the current study is available in the github; https://github.com/allenai/PeerRead/tree/master/data/acl_2017.

References

- [1] Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning–i. *Information sciences* **8**(3), 199–249 (1975)
- [2] Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning–ii. *Information sciences* **8**(4), 301–357 (1975)
- [3] Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning–iii. *Information sciences* **9**(1), 43–80 (1975)
- [4] Jing, F., Chao, X.: Fairness concern: An equilibrium mechanism for consensus-reaching game in group decision-making. *Information Fusion* **72**, 147–160 (2021)
- [5] Dong, Y., Ran, Q., Chao, X., Li, C., Yu, S.: Personalized individual semantics learning to support a large-scale linguistic consensus process. *ACM Transactions on Internet Technology* **23**(2), 1–27 (2023)
- [6] Wang, S., Wu, J., Chiclana, F., Ji, F., Fujita, H.: Global feedback mechanism by explicit and implicit power for group consensus in social network. *Information Fusion*, 102205 (2023)
- [7] Morente-Molinera, J.A., Kou, G., Samuylov, K., Cabrerizo, F., Herrera-Viedma, E.: Using argumentation in expert’s debate to analyze multi-criteria group decision making method results. *Information Sciences* **573**, 433–452 (2021)
- [8] Luo, N., Zhang, Q., Yin, L., Xie, Q., Wu, C., Wang, G.: Three-way multi-attribute decision-making under the double hierarchy hesitant fuzzy linguistic information system. *Applied Soft Computing* **154**, 111315 (2024)
- [9] Zheng, Y., Xu, Z., Li, Y., Pedrycz, W., Yi, Z.: Bi-objective optimization method for large-scale group decision making based on hesitant fuzzy linguistic preference relations with granularity levels. *IEEE Transactions on Fuzzy Systems* **32**(8), 4759–4771 (2024)
- [10] Zhang, H., Zhu, W., Chen, X., Wu, Y., Liang, H., Li, C.-C., Dong, Y.: Managing flexible linguistic expression and ordinal classification-based consensus in large-scale multi-attribute group decision making. *Annals of Operations Research*, 1–54 (2022)
- [11] Zhang, H., Dong, Y., Xiao, J., Chiclana, F., Herrera-Viedma, E.: Personalized individual semantics-based approach for linguistic failure modes and effects analysis with incomplete preference information. *Iise Transactions* **52**(11), 1275–1296

(2020)

- [12] Xiao, J., Wang, X., Zhang, H.: Exploring the ordinal classifications of failure modes in the reliability management: An optimization-based consensus model with bounded confidences. *Group Decision and Negotiation* **31**(1), 49–80 (2022)
- [13] Zhou, M., Zhou, Y.-J., Liu, X.-B., Wu, J., Fujita, H., Herrera-Viedma, E.: An adaptive two-stage consensus reaching process based on heterogeneous judgments and social relations for large-scale group decision making. *Information Sciences*, 119280 (2023)
- [14] Chen, Z.-S., Martinez, L., Chang, J.-P., Wang, X.-J., Xiong, S.-H., Chin, K.-S.: Sustainable building material selection: A qfd-and electre iii-embedded hybrid mcgdm approach with consensus building. *Engineering Applications of Artificial Intelligence* **85**, 783–807 (2019)
- [15] Chen, Z.-S., Yang, L.-L., Chin, K.-S., Yang, Y., Pedrycz, W., Chang, J.-P., Martínez, L., Skibniewski, M.J.: Sustainable building material selection: An integrated multi-criteria large group decision making framework. *Applied Soft Computing* **113**, 107903 (2021)
- [16] Chen, Z.-S., Zhu, Z., Wang, X.-J., Chiclana, F., Herrera-Viedma, E., Skibniewski, M.J.: Multiobjective optimization-based collective opinion generation with fairness concern. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* (2023)
- [17] Chen, Z.-S., Zhu, Z., Wang, Z.-J., Tsang, Y.: Fairness-aware large-scale collective opinion generation paradigm: A case study of evaluating blockchain adoption barriers in medical supply chain. *Information Sciences* **635**, 257–278 (2023)
- [18] Ran, Q., Chao, X., Cabrerizo, F.J., Herrera-Viedma, E.: Managing overconfidence behaviors from heterogeneous preference relations in linguistic group decision making. *IEEE Transactions on Fuzzy Systems* **31**(7), 2435–2449 (2023) <https://doi.org/10.1109/TFUZZ.2022.3226321>
- [19] Zadeh, L.A.: A note on z-numbers. *Information Sciences* **181**(14), 2923–2932 (2011)
- [20] Li, Y., Herrera-Viedma, E., Pérez, I.J., Xing, W., Morente-Molinera, J.A.: The arithmetic of triangular z-numbers with reduced calculation complexity using an extension of triangular distribution. *Information Sciences* **647**, 119477 (2023)
- [21] Wang, J.-q., Cao, Y.-x., Zhang, H.-y.: Multi-criteria decision-making method based on distance measure and choquet integral for linguistic z-numbers. *Cognitive Computation* **9**, 827–842 (2017)
- [22] Wang, J., Wang, J.-q., Tian, Z.-p., Zhao, D.-y.: A multihesitant fuzzy linguistic

- multicriteria decision-making approach for logistics outsourcing with incomplete weight information. *International Transactions in Operational Research* **25**(3), 831–856 (2018)
- [23] Chen, B., Cai, Q., Wei, G., Mo, Z.: Novel aczel-alsina operations-based linguistic z-number aggregation operators and their applications in multi-attribute group decision-making process. *Engineering Applications of Artificial Intelligence* **124**, 106541 (2023)
 - [24] Liu, F., Liao, H., Wu, X., Al-Barakati, A.: Evaluating internet hospitals by a linguistic z-number-based gained and lost dominance score method considering different risk preferences of experts. *Information Sciences* **630**, 647–668 (2023)
 - [25] Zheng, Q., Liu, X., Wang, W., Han, S.: A hybrid hfacs model using dematel-oreste method with linguistic z-number for risk analysis of human error factors in the healthcare system. *Expert Systems with Applications* **235**, 121237 (2024)
 - [26] Yager, R.R.: An approach to ordinal decision making. *International Journal of Approximate Reasoning* **12**(3-4), 237–261 (1995)
 - [27] Herrera, F., Martínez, L.: A 2-tuple fuzzy linguistic representation model for computing with words. *IEEE Transactions on fuzzy systems* **8**(6), 746–752 (2000)
 - [28] Gou, X., Xu, Z.: Double hierarchy linguistic term set and its extensions: the state-of-the-art survey. *International Journal of Intelligent Systems* **36**(2), 832–865 (2021)
 - [29] Xu, B., Deng, Y.: Information volume of z-number. *Information Sciences* **608**, 1617–1631 (2022)
 - [30] Aliev, R.A., Pedrycz, W., Guirimov, B., Huseynov, O.H.: Clustering method for production of z-number based if-then rules. *Information Sciences* **520**, 155–176 (2020)
 - [31] Aliev, R.A., Alizadeh, A.V., Huseynov, O.H.: The arithmetic of discrete z-numbers. *Information Sciences* **290**, 134–155 (2015)
 - [32] Aliev, R.A., Huseynov, O.H., Zeinalova, L.M.: The arithmetic of continuous z-numbers. *Information Sciences* **373**, 441–460 (2016)
 - [33] Xian, S., Chai, J., Guo, H.: Linguistic-induced ordered weighted averaging operator for multiple attribute group decision-making. *International Journal of Intelligent Systems* **34**(2), 271–296 (2019)
 - [34] Peng, H.-G., Wang, J.-Q.: A multicriteria group decision-making method based on the normal cloud model with zadeh’s z-numbers. *IEEE Transactions on Fuzzy Systems* **26**(6), 3246–3260 (2018)

- [35] Peng, H.-g., Wang, X.-k., Zhang, H.-Y., Wang, J.-q.: Group decision-making based on the aggregation of z-numbers with archimedean t-norms and t-conorms. *Information Sciences* **569**, 264–286 (2021)
- [36] Liu, P., Zhang, X., Pedrycz, W.: A consensus model for hesitant fuzzy linguistic group decision-making in the framework of dempster–shafer evidence theory. *Knowledge-Based Systems* **212**, 106559 (2021)
- [37] Dubois, D., Faux, F., Prade, H., Rico, A.: Qualitative capacities: Basic notions and potential applications. *International Journal of Approximate Reasoning* **148**, 253–290 (2022)
- [38] Kang, D., Ammar, W., Dalvi, B., Zuylen, M., Kohlmeier, S., Hovy, E., Schwartz, R.: A dataset of peer reviews (peerread): Collection, insights and nlp applications. *arXiv preprint arXiv:1804.09635* (2018)
- [39] Yaakob, A.M., Gegov, A.: Interactive topsis based group decision making methodology using z-numbers. *International Journal of Computational Intelligence Systems* **9**(2), 311–324 (2016)
- [40] Chatterjee, K., Kar, S.: A multi-criteria decision making for renewable energy selection using z-numbers in uncertain environment. *Technological and Economic Development of Economy* **24**(2), 739–764 (2018)
- [41] Shen, K.-w., Wang, J.-q.: Z-vikor method based on a new comprehensive weighted distance measure of z-number and its application. *IEEE Transactions on Fuzzy Systems* **26**(6), 3232–3245 (2018)

2 Z-number-valued rule-based classification system

- Y. Li, E. Herrera-Viedma, I. J. Pérez, M. Barragán-Guzmán, and J. A. Morente-Molinera, Z-number-valued rule-based classification system, *Applied Soft Computing*, 137, 110168, 2023, <https://doi.org/10.1016/j.asoc.2023.110168>.
- – Status: **Published**.
 - Impact Factor (JCR 2023): 7.2
 - Subject Category: Computer Science, Artificial Intelligence, Ranking 27/197 (Q1).
 - Subject Category: Computer Science, Interdisciplinary Applications, Ranking 16/170 (Q1).

Z-Number-Valued Rule-Based Classification System

Yangxue Li^a, Enrique Herrera-Viedma^{a,*}, Ignacio Javier Pérez^a, Mónica Barragán-Guzmán^a, Juan Antonio Morente-Molinera^{a,*}

^aAndalusian Research Institute in Data Science and Computational Intelligence, Dept. of Computer Science and AI, University of Granada, 18071 Granada, Spain.

Abstract

The fuzzy rule-based classification system (FRBCS) is a popular tool for classification problems due to its interpretability and comprehensibility. As an extension of fuzzy numbers, the concept of Z-number is a more appropriate formal structure to describe uncertain and partially reliable information. A Z-number is an ordered pair of fuzzy numbers, where the second fuzzy number describes the reliability of the first one. As a result of its representation capability, it can receive better classification results. However, there is still a gap in the application of Z-numbers to classification problems due to their high computation complexity. To take advantage of the Z-number, we design a simple way to make Z-numbers apply to classification problems. Use the second fuzzy number to adjust the first fuzzy number to fit the training data. Then we create a kind of Z-number-valued if-then rule by extending the fuzzy if-then rule. In addition, a Z-number-valued rule-based classification system (ZRBCS) is developed, including two main processes: rule generation and new pattern classification. The developed system can cover more information than the classic fuzzy rule-based system, which can improve classification effects. The proposed ZRBCS is compared with classical FRBCS with/without certain degrees and three classical classification algorithms. According to statistical tests, ZRBCS is superior to FRBCS and two other algorithms.

© 2011 Published by Elsevier Ltd.

Keywords: Z-numbers, fuzzy modeling, Z-number-valued if-then rule, rule-based system, classification

1. Introduction

The fuzzy rule-based classification system (FRBCS) is one of the most successful applications of fuzzy sets and fuzzy logic, because of its interpretability and comprehensibility. In classification problems, the main issue for constructing a fuzzy rule-based system is to generate fuzzy if-then rules from numerical data. A fuzzy if-then rule for a classification problem can be written as follows:

If input pattern is antecedent fuzzy numbers then consequent class.

In order to generate a fuzzy rule, we need to determine its antecedent fuzzy numbers, consequent fuzzy classes, and rule weight. The simplest method is the simple fuzzy grid method [1], which is based on local representations, and equally divides a dimension into K , $K = 2, 3, 4, \dots$, triangular fuzzy numbers. These triangular fuzzy numbers divide the pattern space into M^K different fuzzy divisions, where M is the number of features. Its classification performance totally depends on fuzzy partitions. If the fuzzy partitions are too sparse, many patterns may be misclassified. But if the fuzzy partitions are too dense, many useless rules may be generated since the lack of training patterns. In Ref. [1], the certain grades of rules are used to improve the classification performance. This simple heuristic approach requires neither time-consuming iterative computation nor complicated learning procedures. Many different FRBCSs

*Corresponding authors.

Email addresses: viedma@decsai.ugr.es (Enrique Herrera-Viedma), ijperez@decsai.ugr.es (Ignacio Javier Pérez), monicabarragan@ugr.es (Mónica Barragán-Guzmán), jamoren@decsai.ugr.es (Juan Antonio Morente-Molinera)

have been proposed to further improve the performance and interpretability of the system, such as neuro-fuzzy rule-based classification system [2], rule-based granular classification [3, 4], genetic-fuzzy algorithms [5], fuzzy rule-based clustering system [6].

The antecedent fuzzy numbers in FRBCS are set at the beginning and will not be changed during the training process. These preset fuzzy numbers completely determine classification performance. The use of Z-numbers can change this situation and improve the classification performance of fuzzy system. As an extension of fuzzy numbers, Z-number is a more appropriate formal structure for representing real-world information than fuzzy numbers, which can describe both uncertainty and partial reliability of information. A Z-number consists of two fuzzy numbers, $Z = (A, B)$, where A is a fuzzy restriction on the values of the variable, and B is a fuzzy restriction on the reliability measure of A [7]. However, due to there being many optimization problems in computing the underlying probability density functions, Z-numbers have high computation complexity, making them inefficient to be applied to classification problems. And the underlying probability density functions are counterintuitive in classification problems. In this paper, we design a simple way to make Z-numbers apply to classification problems. Use the second fuzzy number to adjust the first fuzzy number to fit the training data better. The larger the reliability, the larger the pattern space covered by the first fuzzy number. On the contrary, the smaller the reliability, the smaller the pattern space covered by the first fuzzy number. Based on this, we extend the FRBCS to the Z-number-valued framework. There are two problems associated with it: generation of Z-number-valued if-then rules from training patterns and classification of new patterns.

The aim of this paper is to construct a basic Z-number-valued rule-based classification system (ZRBCS). We define the basic form of the Z-number-valued if-then rule:

If input pattern is antecedent Z-numbers then consequent class.

The first fuzzy number of an antecedent Z-number is the same as the antecedent fuzzy number of the FRBCS. It is possible to adjust the first fuzzy number to fit training patterns better by using the second fuzzy number of an antecedent Z-number. The second fuzzy number describes information about negative patterns' positions. By adjusting the first fuzzy number according to their specific positions, negative patterns can be correctly classified. More information is considered in the proposed method than in the classical FRBCS.

The main contributions of this paper can be summarized as follows.

- (1) Designed a simple method to make Z-numbers applicable to classification problems, avoiding complex optimization problems with Z-numbers and fitting data better.
- (2) The structure of Z-number-valued if-then rule for classification problems is defined. The antecedent fuzzy numbers in the fuzzy rule are replaced with Z-numbers.
- (3) The Z-number-valued if-then rule generation method is proposed. We make full use of the position information of the negative patterns by the FRBCS to generate the second fuzzy number of the antecedent Z-number.
- (4) The classification process of the new patterns is presented. Using the second fuzzy number, the fuzzy partitions are adjusted to fit the training patterns better.

In order to evaluate the effectiveness of ZRBCS, three experiments are conducted to compare ZRBCS to FRBCS with/without certain degrees and three classical classification algorithms, including support vector machine(SVM) [8], kernel classifier [9] and cam weighted distance nearest neighbor Classifier (CamNN)[10]. The results of the first and the second experiments show that ZRBCS is superior to FRBCS with/without certain degrees. The result of the third experiment demonstrates that ZRBCS is superior to SVM and CamNN.

This study is set out as follows: Section 2 reviews some representative FRBCSs, and the existing mathematical framework and applications of Z-numbers. Section 3 recalls some background knowledge of Z-numbers, the FRBCS without certain grades and with certain grades. Section 4 details the construction procedures of the Z-number-valued if-then rule-based classification system. In Section 5, two experiments on 20 public datasets are performed to analyze and compare the learning and predictive abilities of ZRBCSs and FRBCSs. Finally, we conclude this study and provide some suggestions for future works in Section 6.

2. Related works

This section introduces some representative FRBCS and the mathematical framework and applications of Z-numbers.

2.1. Review of fuzzy rule-based classification system

It is believed that human recognition practice may adopt a set of linguistic rules. In learning, object recognition, understanding, and other activities, rule-based symbolic processing may be the most effective tool to explain human intelligence [11, 12]. The use of fuzzy rules in classification systems can help achieve a balance between interpretability and accuracy. Several researchers have proposed methods for generating and learning fuzzy if-then rules from numerical data. The simplest method is the simple fuzzy grid method [1]. Afterwards, Ishibuchi et al. extended this work later by using the idea of sequential partitioning of the feature space into fuzzy subspaces until a predetermined stopping criterion is satisfied [13]. After that, Ishibuchi et al. suggested a genetic-algorithm-based method to minimize the number of significant fuzzy rules [14]. Ref. [15] discussed that a neuro-fuzzy model can learn fuzzy rules and membership functions from training data. Ishibuchi and Nakaskima designed a fuzzy rule-based system with triangular fuzzy sets [16], in which a genetic algorithm is used to generate rules. They also proposed a heuristic algorithm to compute rule weights. Ref. [17] presented a neuro-fuzzy scheme for realizing a fuzzy rule-based classifier along with feature selection. The network is used to learn the important features and the classification rules. In Ref. [18], authors proposed a support vector learning approach to construct fuzzy classifiers. The fuzzy rule-based classification system is constructed directly from the given training samples using the support vector learning approach. Ishibuchi and Yamamoto proposed two heuristic methods for rule weight specification [19]. Eghbal et al. also proposed mansoori rule weight for fuzzy rule-based systems to improve performance [20]. Ref. [21] discussed a fuzzy association rule-based classification method for high-dimensional problems. In this method, a search tree is employed to generate fuzzy association rules. Hu et al. designed granular fuzzy rule-based models [4]. In Ref. [22], authors extended FRBCS to be able to solve big data classification problems. Niu et al. designed a novel fuzzy rule-based classification model for fuzzy granular rule learning [23]. The methods of rule generation, reasoning and rule weights of some representative FRBCSs are concluded in Table 1.

Ref.	Rule generation	Fuzzy reasoning	Rule weights
Ishibuchi et al. [1]	fuzzy grid method	winner rule	certain grades
Ishibuchi et al. [13]	sequential subdivision, fuzzy grid method	winner rule	certain grades
Ishibuchi et al. [14]	multiple fuzzy rule tables, genetic algorithm	winner rule	certain degrees
Nauck and Kruse [15], Mitra and Hayashi[24]	neuro-fuzzy model	neuro-fuzzy model	no weights
Ishibuchi and Nakaskima [16]	genetic algorithm	winner rule	heuristic algorithm
Chen et al. [18]	support vector learning	input output mapping	no weights
Ishibuchi and Yamamoto [19]	fuzzy grid method	winner rule	penalized certainty factor
Eghbal et al. [20]	uzzy grid method	winner rule	Mansoori rule weight
Alcalá-Fdez et al.[21]	search tree	fuzzy associated rule	no weights
Hu et al. [4]	granular computing	input output mapping	no weights
Sanz et al. [25]	apriori algorithm	fuzzy associated rule	certain grades

Table 1: The methods of rule generation, reasoning and rule weights of some representative FRBCSs.

2.2. Review of Z-numbers

The Z-number has been proven to be significant by many researchers. For example, in decision-making problems, Z-numbers can be useful to identify the unreliable information and utilize the partially reliable information. A lot of proposed Z-number-based decision-making methods which are based on traditional fuzzy decision-making methods have explained the superiority of Z-numbers. Such as the Z-number-based Analytical Hierarchy Process (Z-AHP) method [26], Best Worst Method based on Z-number (Z-BWM) [27], Z-number based technique for order preference by similarity to ideal solution (Z-TOPSIS) [28], Z-number based vlskriterijum-ska optimizacija I kompromisno resenje (Z-VIKOR) [29, 30], decision-making based on Z-relation equation [31]. Furthermore, due to its modeling features, the Z-number has native advantages for natural language processing. Banerjee and Pal [32] extended the basic Z-numbers into a tool for level-2 computing with words (CWW) and consequently subjective natural language

understanding, then use this extension to precise knowledge-frames. Banerjee and Pal [33] proposed augmentation of the Z-numbers for machine-perception encapsulation. In addition, Z-numbers have been successful applied to solve real-world problems, including dynamic control systems [34], safety analytics [35], risk assessment [36], medical diagnosis [37], data envelopment analysis [38], social media evaluation [39], clustering [40], and other fields.

The development of the mathematical frame of Z-numbers can build a firm basis for applications of Z-numbers. General formulas for addition, standard subtraction, square, square-root, and ranking of discrete Z-numbers were established by Aliev et al. [41]. After that, Aliev et al. extended the arithmetic of discrete Z-numbers into continuous Z-numbers [42]. Based on the arithmetic of Z-numbers and differential evolution optimization, Aliev et al. developed the Z-number-based linear programming (Z-LP) model [43]. Aliev extended the concept of specificity to Z-numbers and suggested operations on Z-number by using the horizontal membership concept [44]. Furthermore, Aliev et al. defined the formulation of Z-valued If-Then rule-based interpolation for approximate reasoning [45]. Kang et al. designed the definition of the total utility of Z-number [46]. Mazandarani and Zhao [47] introduced the concepts of Z-differentiability, Z-integral, and Z-Laplace transform of a Z-number-valued function and constructed the framework of calculus for Z-number, Z-calculus.

3. Preliminaries

This section recalls the definition of the Z-number, the structures of the FRBCS without certain grades and with certain grades proposed in [1].

3.1. Z-numbers

Definition 1. A Z-number, Z , has two components, (A, B) . The first component A is a type-1 fuzzy number with membership function $\mu_A : U \rightarrow [0, 1]$, restricting on values of the variable, and the second component B is a type-1 fuzzy number with membership function $\mu_B : [0, 1] \rightarrow [0, 1]$ restricting on the reliability of A [7].

We only considers Z-numbers based on type-1 fuzzy numbers in this paper.

Definition 2. A ordered triple (Y, A, B) is viewed as a Z-restriction on random variable Y [7]:

$$Prob(Y \text{ is } A) \text{ is } B, \quad (1)$$

where $Y \text{ is } A$ is a fuzzy restriction, $Poss(Y = u) = \mu_A(u)$, $Poss(\cdot)$ is a possibility measure, μ_A is the membership function of A , and $u \in \mathcal{R}$; $Prob(Y \text{ is } A)$ is the probability measure of A , given by:

$$Prob(Y \text{ is } A) = \int_{\mathcal{R}} \mu_A(u) p_Y(u) du, \quad (2)$$

where p_Y the underlying probability density of Y . If A and p_Y are consistent, the centroids of A is same with the expected value of p_Y :

$$\int_{\mathcal{R}} u p_Y(u) du = \frac{\int_{\mathcal{R}} u \mu_A(u) du}{\int_{\mathcal{R}} \mu_A(u) du} \quad (3)$$

The underlying probability density functions are not considered in this paper. The reason for this is not only that computing them involves many optimization problems, increasing the difficulty of the algorithm, but also that they are counterintuitive in classification problems.

Example 1. Let us consider two Z-numbers: $Z_1 = (A_1 = (2, 5, 7), B_1 = (0.2, 0.3, 0.4))$, $Z_2 = (A_2 = (4, 7, 9), B_2 = (0.7, 0.8, 0.9))$. Obviously, the second one is more reliable than the first one. We often assume the probability density functions (pdfs) underlying them are normal pdfs [42]. When the reliability value is 0.3, the underlying pdf of Z_1 is $p_{Y1} = \mathcal{N}(\frac{14}{3}, 6.6902)$. When the reliability value is 0.8, the underlying pdf of Z_2 is $p_{Y2} = \mathcal{N}(\frac{20}{3}, 0.5940)$. These normal pdfs are depicted in Fig. 1. Although the reliability of the second Z-number is more reliable, the pdf of the first one covers more space (for the same variable, its probability density value of the first pdf is greater than that of another), even covering $Y = 8$ to 10. This means that more patterns will be covered by a less reliable Z-number, which is counterintuitive.

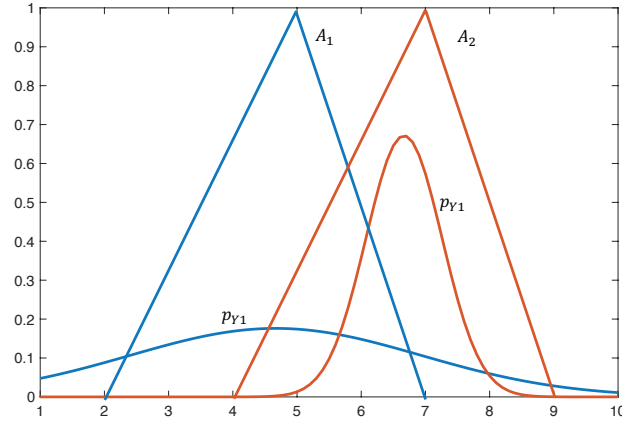


Figure 1: The underlying pdfs of two Z-numbers.

3.2. Fuzzy rule-based classification systems without certain grades

Definition 3. For a W -class classification problem with M -dimensional inputs, the fuzzy if-then rule without certain grades is written as:

$$\text{Rule } R_j: \text{ If } x_{i1} \text{ is } A_{j1} \text{ and } x_{i2} \text{ is } A_{j2} \text{ and } \dots \text{ and } x_{iM} \text{ is } A_{jM} \text{ then Class } C_j, \quad (4)$$

where $i = 1, 2, \dots, n$, n is the number of patterns; $j = 1, 2, \dots, N$, N is the number of rules; $x_i = \{x_{i1}, x_{i2}, \dots, x_{iM}\}$ is the input vector; A_{jm} is the antecedent linguistic value, $m = 1, 2, \dots, M$; C_j is the consequent class of j -th rule, $C_j \in \{c_1, c_2, \dots, c_W\}$.

Ishibuchi et al. [1] proposed the simple fuzzy grid method to partition a dimension into some fuzzy numbers. The dimension is uniformly divided into K triangular fuzzy subsets: $A_1^K, A_2^K, \dots, A_k^K, \dots, A_K^K$, $k = 1, 2, \dots, K$. The antecedent fuzzy subsets A_{jm} is one of $\{A_1^K, A_2^K, \dots, A_k^K, \dots, A_K^K\}$. Therefore, $N = K^M$ different fuzzy if-then rules partition the pattern space into N subspaces.

Definition 4. The compatibility grade of rule R_j with the input pattern x_i is defined by product operator as

$$\mu_j(x_i) = \mu_{A_{j1}}(x_{i1}) \times \mu_{A_{j2}}(x_{i2}) \times \dots \times \mu_{A_{jM}}(x_{iM}), \quad (5)$$

where $\mu_{A_{jm}}$ is the membership function of fuzzy number A_{jm} , $j = 1, 2, \dots, N$, $m = 1, 2, \dots, M$.

Definition 5. The compatibility grade of rule with the class c_ω is defined as:

$$\psi_j(c_\omega) = \sum_{x_i \in c_\omega} \mu_j(x_i) \quad (6)$$

The consequent class of R_j , $C_j = c_{\omega^*}$, if

$$\psi_j(c_{\omega^*}) = \max(\psi_j(c_1), \psi_j(c_2), \dots, \psi_j(c_W)) \quad (7)$$

If there are multiple maximum values in Eq. (7), the consequent class $C_j = \emptyset$.

For a new pattern x_{new} , it is classified by the single winner rule R_{j^*} , if

$$\mu_{j^*}(x_{new}) = \max(\mu_1(x_{new}), \mu_2(x_{new}), \dots, \mu_N(x_{new})). \quad (8)$$

The class of this new pattern is the consequent class of R_{j^*} , C_{j^*} . When multiple rules take the maximum value in Eq. (8), the new pattern cannot be classified, considered as an unclassifiable pattern.

Example 2. Let us consider a 3-class classification problem with two-dimensional input shown in Fig. 2. There are three classes noted by red inverted triangles, green points, and blue squares, in which the solid points are training patterns and the hollow points are testing patterns. When $K = 2, 3, 4, 5$, the pattern space is divided into 4, 9, 16, 25 fuzzy partitions (subspaces) shown as Figs. 3a, 3b, 3c and 3d, respectively. The consequent classes of subspaces are marked by the corresponding colors.

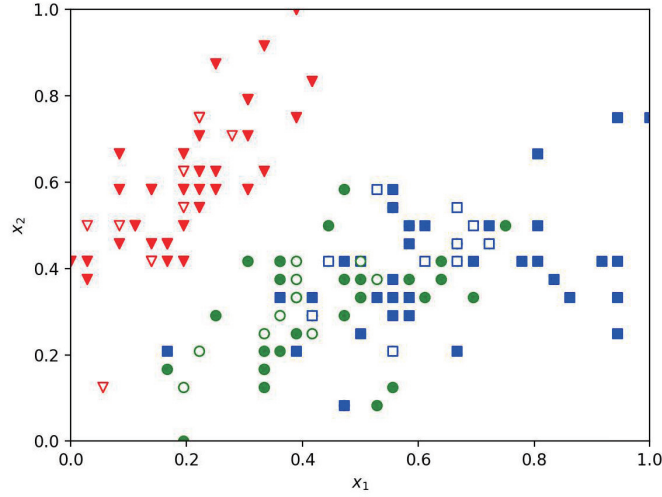


Figure 2: An example of 3-class classification problem with two-dimensional input.

3.3. Fuzzy rule-based classification systems with certain grades

Definition 6. For a classification problem with a M -dimensional input vector $x_i = \{x_{i1}, x_{i2}, \dots, x_{iM}\}$, the fuzzy if-then rule without certain grades is written as:

$$\text{Rule } R_j : \text{ If } x_{i1} \text{ is } A_{j1} \text{ and } x_{i2} \text{ is } A_{j2} \text{ and } \dots \text{ and } x_{iM} \text{ is } A_{jM} \text{ then Class } C_j \text{ with } CF_j, \quad (9)$$

where $i = 1, 2, \dots, n$; $j = 1, 2, \dots, N$, N is the number of rules; A_{jm} is the antecedent linguistic value, $m = 1, 2, \dots, M$; C_j is the consequent class of j -th rule, $C_j \in \{c_1, c_2, \dots, c_W\}$; CF_j is the certain grade of rule R_j and $CF_j \in [0, 1]$.

The certain grade is determined as:

$$CF_j = \frac{\psi_j(C_j) - \bar{\psi}}{\sum_{\omega=1}^W \psi_j(c_\omega)}, \quad (10)$$

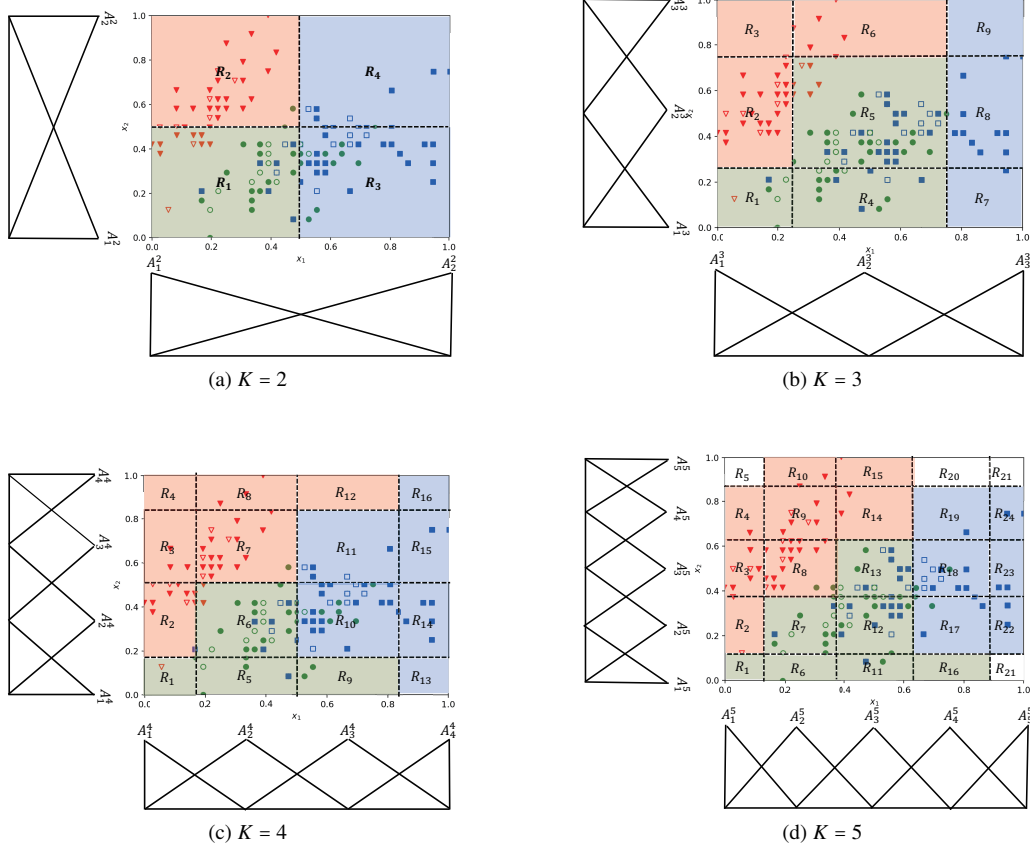
where

$$\bar{\psi} = \frac{\sum_{c_\omega \neq C_j} \psi_j(c_\omega)}{W - 1}. \quad (11)$$

For a new pattern x_{new} , it is classified as:

$$\mu_{j^*}(x_{new}) \times CF_{j^*} = \max(\mu_1(x_{new}) \times CF_1, \mu_2(x_{new}) \times CF_2, \dots, \mu_N(x_{new}) \times CF_N). \quad (12)$$

The class of this new pattern is the consequent class of R_{j^*} , C_{j^*} . If multiple rules take the maximum value in Eq. (12), the new pattern cannot be classified, considered as an unclassifiable pattern.

Figure 3: Fuzzy partitions of each fuzzy rules when $K = 2$ (a), 3 (b), 4 (c), 5 (d).

4. Z-number-valued rule-based classification system

Based on the Z-number's powerful ability to represent and reason with uncertain and partially reliable information, we develop a Z-number-valued if-then rule-based classification system (ZRBCS). As shown in Fig. 4, the ZRBCS involves two main procedures: generating Z-number-valued if-then rules and classifying new patterns. The procedure of generating Z-number-valued if-then rule includes the first fuzzy numbers generation, the second fuzzy number generation and certain degrees determination.

4.1. Z-number-valued if-then rule structure

When we replace the antecedent fuzzy numbers in Eq. (4) with Z-numbers, the structure of the Z-number-valued if-then rule can be defined as follow.

Definition 7. For a W -class classification problem with a M -dimensional input vector $x_i = \{x_{i1}, x_{i2}, \dots, x_{iM}\}$, the Z-number-valued if-then rule is written as:

$$\text{Rule } R_j^Z: \text{ If } x_{i1} \text{ is } Z_{j1} \text{ and } x_{i2} \text{ is } Z_{j2} \text{ and } \dots \text{ and } x_{iM} \text{ is } Z_{jM} \text{ then Class } C_j, \quad (13)$$

where $i = 1, 2, \dots, n$, n is the number of patterns; $j = 1, 2, \dots, N$, N is the number of rules; $Z_{jm} = (A_{jm}, B_{jm})$ is the antecedent Z-number; $m = 1, 2, \dots, M$; C_j is the consequent class of j -th rule, $C_j \in \{c_1, c_2, \dots, c_W\}$.

Similarly, when we replace the antecedent fuzzy subsets in Eq. (9), the structure of the Z-number-valued if-then rule with certain degree can be written as follows.

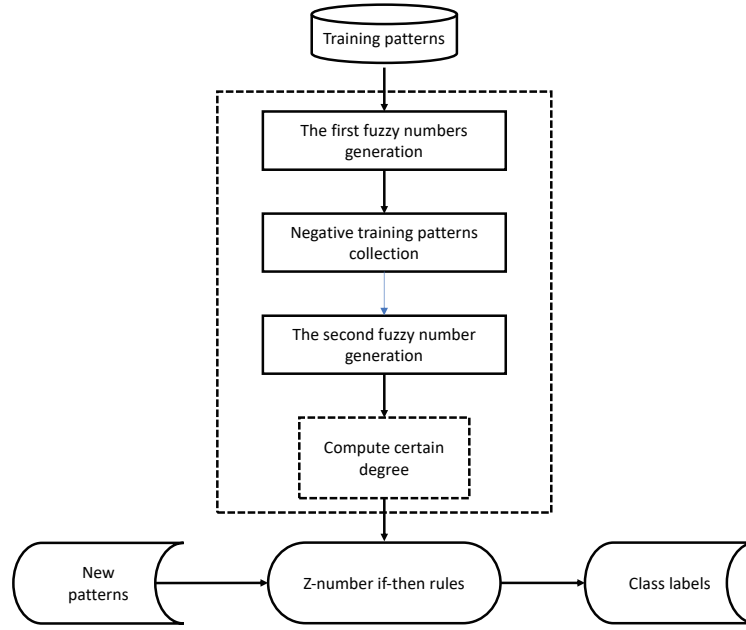


Figure 4: Z-number-valued rule-based classification system.

Definition 8. For a W -class classification problem with a M -dimensional input vector $x_i = \{x_{i1}, x_{i2}, \dots, x_{iM}\}$, the Z-number-valued if-then rule is written as:

$$\text{Rule } R_j^Z: \text{ If } x_{i1} \text{ is } Z_{j1} \text{ and } x_{i2} \text{ is } Z_{j2} \text{ and } \dots \text{ and } x_{iM} \text{ is } Z_{jM} \text{ then Class } C_j \text{ with } CF_j, \quad (14)$$

where $i = 1, 2, \dots, n$, n is the number of patterns; $j = 1, 2, \dots, N$, N is the number of rules; $Z_{jm} = (A_{jm}, B_{jm})$ is the antecedent Z-number, $m = 1, 2, \dots, M$; C_j is the consequent class of j -th rule, $C_j \in \{c_1, c_2, \dots, c_W\}$; CF_j is the certain degree of rule R_j and $CF_j \in [0, 1]$.

4.2. Z-number-valued if-then rules generation

To generate Z-number-valued if-then rules from training patterns, three issues need to be addressed: the generation of the first antecedent fuzzy number, the generation of the second antecedent fuzzy number, and the determination of the consequent class. As with the FRBCS, the first antecedent fuzzy number and consequent class are generated using the simple fuzzy grid method and maximum compatibility grade method, respectively. The first step of generating Z-number-valued if-then rules is creating fuzzy if-then rules.

Let us consider the two-dimensional classification problem in Example 2. Let $K = 4$ and each dimension is divided into $K = 4$ triangular fuzzy numbers: A_1^4, A_2^4, A_3^4 and A_4^4 . The pattern space is partitioned into 16 subspaces as shown in Fig. 3c. The fuzzy if-then rule represented by the sixth fuzzy partition is:

$$\text{Rule } R_6^Z: \text{ If } x_{i1} \text{ is } A_2^4 \text{ and } x_{i2} \text{ is } A_2^4 \text{ then Class } c_2,$$

The consequent class of the sixth rectangle is c_2 , but there are some red inverted triangles and blue squares are misclassified as c_2 . Our aim is correctly to classify these patterns. In the FRBCS without certain degrees, the pattern space is uniformly divided into some fuzzy partitions. However, the patterns are not uniformly distributed. The certain degree is used to adjust these partitions vertically. The Z-numbers can adjust them horizontally in the ZRBCS.

Definition 9. For a Z-number $Z = (A, B)$, where A and B are triangular fuzzy numbers $A = (a_1, a_2, a_3)$ and $B = (b_1, b_2, b_3)$, $a_1, a_2, a_3, b_1, b_2, b_3 \in \mathcal{R}$, we can use the reliability fuzzy number B to adjust A to obtain a new fuzzy number $A' = (a'_1, a'_2, a'_3)$, defined as:

$$\begin{aligned} a'_1 &= a_2 - (a_2 - a_1) \times \text{Rel}(B), \\ a'_2 &= a_2, \\ a'_3 &= a_2 + (a_3 - a_2) \times \text{Rel}(B). \end{aligned} \quad (15)$$

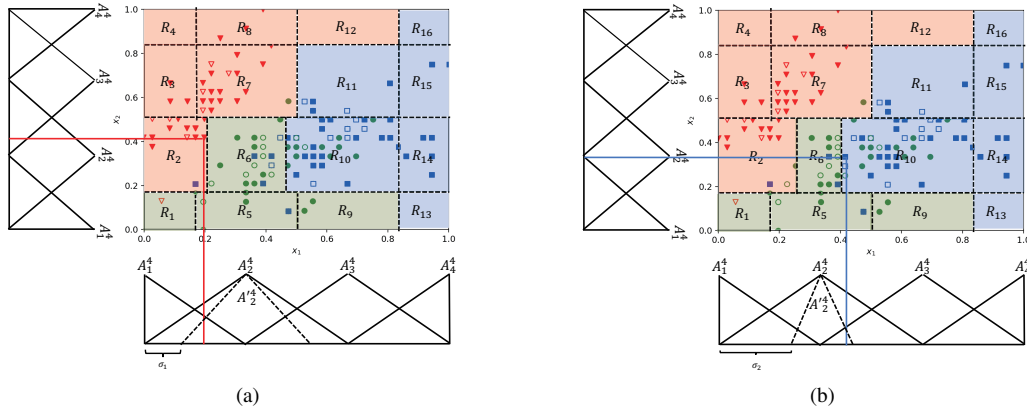


Figure 5: (a) A negative pattern that is a little over the threshold. (b) A negative pattern that is a lot over the threshold.

Where $Rel(B) = (b_1 + b_2 + b_3)/3$ is the reliability value.

We can use the negative patterns to determine the second fuzzy number. If a negative pattern is just a bit over the threshold (the intersection of this fuzzy number with its adjacent fuzzy number), the fuzzy number just needs to adjust a bit so that it doesn't cover the pattern. For example in Fig. 5a, the 6th fuzzy partition covers some negative points, including some red triangles and some blue squares. If we adjust this fuzzy partition so that it does not cover these negative points. The fuzzy subset A_2^4 needs to be reduced by σ_1 to $A_2'^4$ to not cover the red point marked in the figure. Whenever a negative pattern is a lot over the threshold, the fuzzy number has to be adjusted by a lot to not cover it. In Fig. 5b, the fuzzy subset A_2^4 has to be reduced by σ_2 to $A_2'^4$ to not cover the blue point marked in the figure. It is obvious that σ_2 is greater than σ_1 . The closer the membership degree is to the threshold, the closer the value $1.5 - \epsilon - \mu_{A_2^4}(x)$ is to $1 - \epsilon$, the smaller the adjustment needed for the fuzzy number. Where ϵ is the infinitesimals.

The second step of generating Z-number-valued if-then rules is to collect the negative training patterns for each fuzzy if-then rule.

Definition 10. For the rule R_j , its set of negative training patterns is defined as:

$$E_j = \{x_e | \forall m = 1, 2, \dots, M, \mu_{A_{jm}}(x_{em}) \geq 0.5, x_e \in C_j\}, \quad (16)$$

where 0.5 is the membership degree of a pattern that is on the intersection of two adjacent fuzzy numbers.

Based on the above analysis, we can define the second antecedent fuzzy number as:

Definition 11. For the rule R_j , the second fuzzy number of Z_{jm} , B_{jm} is a triangular fuzzy number:

$$B_{jm} = \begin{cases} \left(\min_{x_e \in E_j} (1.5 - \epsilon - \mu_{A_{jm}}(x_{em})), \right. \\ \quad \frac{\sum_{x_{em} \in E_j} (1.5 - \epsilon - \mu_{A_{jm}}(x_{em}))}{|E_j|}, \\ \quad \left. \max_{x_e \in E_j} (1.5 - \epsilon - \mu_{A_{jm}}(x_{em})), \right) & |E_j| \neq 0, \\ (1, 1, 1), & |E_j| = 0, \end{cases} \quad (17)$$

where $m = 1, 2, \dots, M$, $|E_j|$ is the cardinality of E_j .

The algorithm of Z-number-valued if-then rules generation can be concluded as Algorithm 1.

Algorithm 1 Z-number-valued if-then rules generation algorithm.

Input: The training dataset $T = \{(x_i, c^i) | x_i \in \mathbb{R}^M, c^i \in \{c_1, c_2, \dots, c_M\}\}$, the number of fuzzy partitions K .

Output: A set of Z-number-valued if-then rules R^Z and the corresponding consequent classes set C

```

1:  $N \leftarrow K^M$ 
2:  $R^F, C = \text{Fuzziness}(T, K)$  // The simple fuzzy grid if-then rules generation,  $R_j^F = \{A_{j1}, A_{j2}, \dots, A_{jM}\}$ ,  $C = \{C_1, C_2, \dots, C_N\}$ ,  $j = 1, 2, \dots, N$ 
3:  $E \leftarrow N * \emptyset$ 
4: for  $x_i$  in  $T$  do
5:   for  $R_j^F$  in  $R^F$  do
6:     if  $\forall m = 1, \dots, M, \mu_{A_{jm}}(x_{im}) \geq 0.5$  then
7:        $x_i$  add in  $E_j$ 
8:     end if
9:   end for
10: end for
11: for  $R_j^F$  in  $R^F$  do
12:   for  $m = 1$  to  $M$  do
13:      $B_{jm}$  calculated by Eq. (17)
14:   end for
15:    $R_j^Z \leftarrow \{(A_{j1}, B_{j1}), (A_{j2}, B_{j2}), \dots, (A_{jM}, B_{jM})\}$ 
16: end for
17:
18: return  $R^Z, C$ 

```

Example 3. Considering the previous two-dimensional classification problem with $K = 4$ shown in Fig. 3c in Example 2, the first rule is:

Rule R_1^Z : If x_{i1} is (A_1^4, B_{11}) and x_{i2} is (A_1^4, B_{12}) then Class c_2 .

Its set of misclassified training patterns is: $E_1 = \emptyset$. Thus the second fuzzy numbers B_{11} and B_{12} are $(1, 1, 1)$. The complete first Z-number-valued if-then rule is:

Rule R_1^Z : If x_{i1} is $((0, 0, 0.3333), (1, 1, 1))$ and x_{i2} is $((0, 0, 0.3333), (1, 1, 1))$ then Class c_2 .

The 7-th rule is:

Rule R_7 : If x_{i1} is (A_2^4, B_{71}) and x_{i2} is (A_3^4, B_{72}) then Class c_1 .

Its set of misclassified training datapoints is: $E_7 = \{(0.4444, 0.5), (0.4722, 0.5833)\}$. According to the Eq. (17), The second fuzzy number of Z_{71} is:

$$\begin{aligned}
 B_{71} &= (\min(1.5 - \epsilon - \mu_{A_2^4}(0.4444), 1.5 - \epsilon - \mu_{A_2^4}(0.4722)), \frac{(1.5 - \epsilon - \mu_{A_2^4}(0.4444)) + (1.5 - \epsilon - \mu_{A_2^4}(0.4722))}{2}, \\
 &\quad \max(1.5 - \epsilon - \mu_{A_2^4}(0.4444), 1.5 - \epsilon - \mu_{A_2^4}(0.4722))) \\
 &= (\min(0.8333 - \epsilon, 0.9166 - \epsilon), \frac{0.8333 - \epsilon + 0.9166 - \epsilon}{2}, \max(0.8333 - \epsilon, 0.9166 - \epsilon)) \\
 &= (0.8333 - \epsilon, 0.8750 - \epsilon, 0.9166 - \epsilon).
 \end{aligned}$$

The second fuzzy number of Z_{72} is:

$$\begin{aligned}
 B_{72} &= (\min(1.5 - \epsilon - \mu_{A_3^4}(0.5), 1.5 - \epsilon - \mu_{A_3^4}(0.5833)), \frac{(1.5 - \epsilon - \mu_{A_3^4}(0.5)) + (1.5 - \epsilon - \mu_{A_3^4}(0.5833))}{2}, \\
 &\quad \max(1.5 - \epsilon - \mu_{A_3^4}(0.5), 1.5 - \epsilon - \mu_{A_3^4}(0.5833))) \\
 &= (\min(0.9999 - \epsilon, 0.75 - \epsilon), \frac{0.9999 - \epsilon + 0.75 - \epsilon}{2}, \max(0.9999 - \epsilon, 0.75 - \epsilon)) \\
 &= (0.75 - \epsilon, 0.8750 - \epsilon, 0.9999 - \epsilon).
 \end{aligned}$$

The complete 7-th Z-number-valued if-then rule is:

Rule R_7^Z : If x_{i1} is $((0, 0.3333, 0.6667), (0.8333 - \epsilon, 0.8750 - \epsilon, 0.9166 - \epsilon))$ and x_{i2} is $((0.3333, 0.6667, 1.0), (0.75 - \epsilon, 0.8750 - \epsilon, 0.9999 - \epsilon))$ then Class c_1 .

The other 14 Z-number-valued if-then rules are generated by the same way.

4.3. ZRBCS with certain grades

In the FRBCS, the certain grades are calculated by Eq. (10). The certain grades in the ZRBCS are calculated in the same way. Only the fuzzy numbers adjusted by the second fuzzy number are used. For a Z-number-valued rule R_j ,

Rule R_j^Z : If x_{i1} is Z_{j1} and x_{i2} is Z_{j2} and $\dots$ and x_{iM} is Z_{jM} then Class C_j with CF_j . (18)

It is first necessary to introduce a few relevant definitions before discussing the calculation method of certain grades.

Definition 12. The Z-compatibility grade of the Z-number-valued if-then rule R_j with the input pattern x_i is defined as

$$\mu_j^Z(x_i) = \mu_{A'_{j1}}(x_{i1}) \times \dots \times \mu_{A'_{jM}}(x_{iM}), \quad (19)$$

where A'_{jm} is the adjusted fuzzy number of the Z-number $Z_{jm} = (A_{jm}, B_{jm})$.

Definition 13. The Z-compatibility grade of Z-number-valued rule R_j with class c_ω is defined as:

$$\psi_j^Z(c_\omega) = \sum_{x_i \in C_j} \mu_j^Z(x_i). \quad (20)$$

CF_j can be calculated by:

$$CF_j = \frac{\psi_j^Z(C_j) - \bar{\psi}^Z}{\sum_{\omega=1}^W \psi_j^Z(c_\omega)}, \quad (21)$$

where

$$\bar{\psi}^Z = \frac{\sum_{c_\omega \neq C_j} \psi_j^Z(c_\omega)}{W - 1}. \quad (22)$$

4.4. New patterns classification procedure

For a new pattern x_{new} , it is classified by rule R_{j^*} , if

$$\mu_{j^*}^Z(x_{new}) = \max(\mu_1^Z(x_{new}), \mu_2^Z(x_{new}), \dots, \mu_M^Z(x_{new})). \quad (23)$$

Algorithm 2 New pattern classification algorithm.

Input: A new pattern $x_{new} = \{x_{new,1}, x_{new,2}, \dots, x_{new,M}\}$, a set of Z-number-valued if-then rules R^Z and its corresponding consequent classes set C and certain degrees CF .

Output: The class of the new pattern, c_{new} .

```

1: for  $R_j^Z$  in  $R^Z$  do
2:    $\mu_j^Z \leftarrow 1$ 
3:   for  $m = 1$  to  $M$  do
4:      $\mu_j^Z \leftarrow \mu_j^Z \times \mu_{A'_{jm}}(x_{new,m})$ 
5:   end for
6: end for
7:  $j^* = \operatorname{argmax}_{j=1,2,\dots,N}(\mu_j^Z)$ 
8: if  $\max_{j=1,2,\dots,N, j \neq j^*}(\mu_j^Z) = \max_{j=1,2,\dots,N}(\mu_j^Z)$  then
9:    $c_{new} \leftarrow \text{misclassified}$ 
10: else
11:    $c_{new} \leftarrow C_{j^*}$ 
12: end if
13: return  $c_{new}$ 

```

If there are multiple rules that have the maximum Z-compatibility values, the new pattern cannot be classified, considering as an unclassifiable pattern. Else the class of x_{new} is the consequent class of $R_{j^*}^Z$, $x_{new} \in C_{j^*}$. The algorithm of the new pattern classification can be concluded as Algorithm 2.

Example 4. A new pattern $x_{new} = \{0.1944, 0.6249\}$ input the ZRBCS generated in Example 3. The Z-compatibility grade of rules are:

$$\mu_2^Z(x_{new}) = 0.0521;$$

$$\mu_3^Z(x_{new}) = 0.3646;$$

$$\mu_7^Z(x_{new}) = 0.4490.$$

The other Z-compatibility grades are zero. The 7-th rule R_7^Z takes the maximum Z-compatibility grade, thus this new pattern is c_1 , a red inverted triangle.

The certain degrees of 2-th, 3-th and 7-th rules are $CF_2 = 0.8412$, $CF_3 = 0.9963$ and $CF_7 = 0.7994$. The Z-compatibility grade with CF of rules are:

$$\mu_2^Z(x_{new}) \times CF_2 = 0.0438;$$

$$\mu_3^Z(x_{new}) \times CF_3 = 0.3633;$$

$$\mu_7^Z(x_{new}) \times CF_7 = 0.3589.$$

The 3-th rule R_3^Z takes the maximum value. The consequent class of R_3^Z also is c_1 , a red inverted triangle.

5. Performance evaluation

This section conducts three experiments comparing ZRBCS and FRBCS with and without certain grades, and some classical classification algorithms.

Dataset	#Patterns	#Dimensions	#Classes
Banknote	1372	4	2
Breast	569	5	2
Bupa	345	6	2
Diabetes	768	8	2
Ecoli	336	7	8
Glass	214	9	6
Haberman	306	3	2
HTRU-2	17898	8	2
Ionosphere	351	34	2
Iris	150	4	3
Liver	345	6	2
Newthyroid	215	5	3
Occupancy	2665	5	2
Pageblocks	5473	10	5
Pendigits	10992	16	10
Pima	768	8	2
Rice	3810	7	2
Segment	2310	19	7
Tranfution	748	4	2
Vehicle	846	18	4

Table 2: The description of used datasets.

5.1. Experimental set-up

The performance of the proposed ZRBCS is evaluated by three experiments over 20 datasets. These datasets vary in the number of classes, dimensions, and patterns. The description of these datasets is shown in Table 2, where ‘#Patterns’ is the number of patterns, ‘#Dimensions’ is the number of dimensions, and ‘#classes’ is the number of classes.

For evaluating generalization ability, 10-fold cross-validation (10CV) is used. In it, the patterns are divided into 10 folds, each of them contains 10% patterns. Nine folds are used to train the model, and the other is used to test it. The average accuracy of 10 folds can show the performance of the model. All algorithms are implemented in KEE (knowledge extraction based on evolutionary learning) software [48]. KEEL is an open source Java software tool that can be used for a large number of different knowledge data discovery tasks. Compared to Python, Java has higher speed and efficiency due to it being a compiled language. But as an interpreted language, Python has simpler, more concise syntax than Java. The details of other algorithms can be found on KEEL. Five different fuzzy simple grid partitions ($K = 2, 3, 4, 5, 6$) are used to make the comparison, where each dimension is homogeneously divided into K triangular fuzzy numbers in the same manner.

5.2. Comparison without certain degrees

Table 3 and Table 4 show the classification accuracy rates of the ZRBCS in comparison with FRBCS without certain degrees over the training data and the testing data, respectively. Bold-faced results indicate that the current method is superior. As listed in Table 3, the ZRBCS shows better learning ability over most datasets, meaning Z-numbers can indeed improve the learning ability of the FRBCS. When $K = 2, 3, 4, 5$ and 6, it performs better on predictive ability than FRBCS in 16, 17, 17, 18 and 17 out of 20 datasets. It also does not perform poorly in 17, 18, 18, 19 and 18 out of 20 datasets, respectively. We can conclude that ZRBCS outperforms FRBCS on most datasets. In order to find the superior approach statistically, we use nonparametric tests for multiple comparisons, taking into account the ranks over the test data. We use the Wilcoxon signed-rank test to compare the ZRBCS with FRBCS for each number of partitions, and $\alpha = 0.05$ with two-side. The statistic values and probability (p) values for each number of partitions are shown in Table 5. The critical value can be found in the table of critical values for the Wilcoxon signed-rank test. Clearly, the Wilcoxon signed-rank test rejects all hypotheses that there are statistically significant differences between the two methods. As a result, we can conclude from the statistical study in Table 5 that the ZRBCS is superior to FRBCS without certain degrees.

	K=2		K=3		K=4		K=5		K=6	
	FRBCS	ZRBCS	FRBCS	ZRBCS	FRBCS	ZRBCS	FRBCS	ZRBCS	FRBCS	ZRBCS
Banknote	64.07%	80.29%	69.55%	76.08%	72.64%	81.37%	79.69%	89.39%	94.84%	97.50%
Breast	37.26%	37.26%	62.35%	85.59%	79.77%	85.26%	84.57%	87.97%	80.98%	87.82%
Bupa	57.13%	59.26%	58.65%	59.58%	62.09%	64.28%	69.27%	68.63%	78.32%	76.46%
Diabetes	53.76%	62.28%	71.12%	70.10%	78.95%	81.65%	85.88%	88.99%	94.85%	93.88%
Ecoli	61.28%	57.31%	76.49%	80.06%	82.90%	83.90%	90.91%	90.67%	94.41%	94.68%
Glass	57.38%	58.83%	57.32%	50.96%	67.08%	69.47%	71.59%	78.09%	85.93%	83.55%
Haberman	73.60%	73.64%	71.57%	71.20%	71.72%	70.88%	74.51%	73.57%	75.74%	74.11%
HTRU-2	92.22%	94.01%	92.91%	96.23%	94.64%	96.73%	95.56%	97.28%	95.80%	96.92%
Ionosphere	97.59%	98.35%	98.45%	98.42%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Iris	60.52%	65.33%	92.00%	94.59%	85.41%	87.63%	84.96%	85.26%	96.96%	97.56%
Liver	57.13%	59.13%	58.78%	59.45%	62.45%	63.96%	69.34%	68.34%	78.74%	77.17%
Newthyroid	83.72%	80.36%	87.96%	91.11%	94.16%	94.88%	95.35%	97.26%	94.01%	94.06%
Occupancy	86.33%	86.43%	83.56%	86.37%	97.90%	97.86%	96.93%	97.58%	96.48%	96.37%
Pageblocks	89.66%	90.31%	91.16%	91.48%	92.67%	92.88%	91.06%	92.83%	92.48%	92.98%
Pendigits	85.34%	89.74%	96.67%	98.44%	99.47%	99.78%	99.887%	99.892%	99.97%	99.98%
Pima	58.28%	64.42%	41.70%	54.76%	76.30%	80.64%	87.44%	87.51%	94.27%	92.51%
Rice	45.39%	87.53%	57.02%	78.54%	48.49%	64.11%	69.46%	83.79%	68.07%	81.01%
Segment	76.44%	66.96%	82.36%	85.65%	95.94%	94.30%	94.66%	94.33%	96.41%	96.91%
Tranfufusion	40.24%	40.78%	51.77%	49.82%	69.31%	70.23%	52.42%	57.32%	72.67%	73.19%
Vehicle	56.50%	64.33%	80.98%	77.96%	95.17%	95.42%	98.50%	98.57%	99.68%	99.79%

Table 3: Classification accuracy rate (%) of the FRBCS and ZRBCS without certain degrees over training data.

	K=2		K=3		K=4		K=5		K=6	
	FRBCS	ZRBCS	FRBCS	ZRBCS	FRBCS	ZRBCS	FRBCS	ZRBCS	FRBCS	ZRBCS
Banknote	64.07%	80.25%	69.39%	76.02%	72.52%	81.27%	79.52%	89.28%	94.39%	97.30%
Breast	37.26%	37.26%	61.33%	84.72%	78.20%	83.99%	82.25%	86.47%	78.03%	86.48%
Bupa	55.87%	57.85%	53.80%	56.12%	51.81%	52.66%	49.58%	51.02%	53.93%	57.04%
Diabetes	51.96%	60.03%	62.78%	63.57%	62.50%	67.84%	61.33%	67.19%	64.07%	64.20%
Ecoli	58.02%	55.32%	69.06%	73.54%	70.57%	73.27%	77.40%	77.39%	74.12%	74.39%
Glass	51.44%	55.28%	47.76%	41.23%	52.49%	55.53%	53.09%	57.70%	63.32%	56.76%
Haberman	72.87%	73.19%	69.92%	70.26%	66.04%	66.06%	64.71%	65.71%	60.48%	61.47%
HTRU-2	92.21%	93.95%	92.81%	96.14%	94.50%	96.71%	95.37%	97.13%	95.44%	96.56%
Ionosphere	64.96%	66.10%	63.79%	63.79%	56.98%	56.98%	54.40%	54.40%	46.43%	46.43%
Iris	60.00%	64.67%	91.33%	94.67%	84.00%	86.67%	82.00%	84.67%	90.00%	91.33%
Liver	55.92%	59.09%	52.18%	53.94%	48.99%	51.88%	50.20%	54.55%	51.93%	54.80%
Newthyroid	83.72%	80.48%	86.10%	88.90%	90.78%	91.26%	92.64%	94.46%	89.33%	90.28%
Occupancy	86.15%	86.23%	83.56%	86.30%	97.90%	97.86%	96.85%	97.64%	96.44%	96.29%
Pageblocks	89.60%	90.26%	90.72%	91.19%	92.05%	92.36%	90.19%	92.03%	91.21%	91.92%
Pendigits	84.08%	88.63%	94.88%	97.03%	97.64%	98.02%	95.62%	95.65%	90.40%	90.44%
Pima	56.10%	62.13%	33.84%	49.22%	59.88%	66.01%	63.03%	66.03%	64.17%	65.60%
Rice	45.54%	87.51%	56.90%	78.48%	48.27%	64.20%	68.87%	83.46%	67.30%	80.42%
Segment	74.72%	66.54%	79.48%	84.03%	91.60%	90.09%	89.65%	89.74%	90.35%	90.95%
Tranfusion	40.23%	40.90%	52.01%	51.07%	69.50%	70.57%	50.38%	56.00%	71.00%	71.67%
Vehicle	46.00%	58.51%	58.50%	60.27%	61.49%	63.14%	59.91%	60.27%	62.88%	63.11%

Table 4: Classification accuracy rate (%) of the FRBCS and ZRBCS without certain degrees over testing data.

K	Statistic value	p-value	Critical value	Hypothesis
$K = 2$	30	0.00285	52	Rejected
$K = 3$	20	0.00058	52	Rejected
$K = 4$	20	0.00058	52	Rejected
$K = 5$	10	9.61658e-05	52	Rejected
$K = 4$	20	0.00058	52	Rejected

Table 5: Wilcoxon signed-rank test of the accuracy rate for comparing ZRBCS to FRBCS without certain degrees.

	K=2		K=3		K=4		K=5		K=6	
	FRBCS	ZRBCS	FRBCS	ZRBCS	FRBCS	ZRBCS	FRBCS	ZRBCS	FRBCS	ZRBCS
Banknote	52.4	64.1	75.7	63.6	85.1	80.3	96.9	85.3	110	95.4
Breast	36.3	44	46	41.5	53.4	60	64.9	65.2	79.9	61.5
Bupa	28.1	35.2	43.2	39.4	52.4	56.3	64.1	64.3	73.4	70.5
Diabetes	58.6	72.7	106.8	94.8	177.6	173.9	237.3	236.5	308.8	293.2
Ecoli	34.8	41.7	52.8	51.3	68.8	69.3	78.7	78.2	90.8	88
Glass	33.2	40.9	41.9	38.1	49.8	51.6	57.7	58.3	65.8	63.1
Haberman	30.7	37.9	39.2	33.7	42.9	43.3	48.1	47	50.4	45.2
HTRU-2	3520.4	3536	4455.6	3816.6	5676.9	4571.8	5467.7	4552.7	6677.2	5250.2
Ionosphere	306.3	316.3	304.6	295.1	408.6	429.8	389.7	362	414.5	409.3
Iris	34.1	36.2	37	30.4	42.8	45.5	47.3	47.3	50.9	42.2
Liver	33.2	36.8	46.2	40.3	55.7	58.6	64.6	66.1	73.5	68.3
Newthyroid	31.6	35.2	38.9	33.6	45.3	47.8	48.1	49.3	50.9	43.1
Occupancy	105.9	106.4	103.2	94.6	115.7	104.8	105.4	123.8	109.3	128.1
Pageblocks	542.6	552.6	619.8	500.5	756.8	548	813.3	563.1	881.5	559.1
Pendigits	7260	4078.9	16241.2	10169	24758.6	18622.8	32568	25414.4	36001.2	29833.2
Pima	103.8	119.5	126	116.8	198.7	207.5	264.6	252.4	314.2	298.7
Rice	335.9	319.3	472.5	392.9	562	396.4	637.5	442.5	708.2	496.5
Segment	343	295	695.8	557	1112.5	1004.3	1471.3	1303.6	1718.7	1568.4
Tranfusion	40.4	50.1	44.6	41	49.2	49.7	52.8	52	56.9	50.5
Vehicle	240.1	273.2	479.8	481.2	779.4	766.9	871	837	924.7	915.8
Average	658.57	504.6	1203.54	846.57	1754.61	1369.43	2172.45	1735.05	2438.04	2019.015

Table 6: Running time (in millisecond) of the FRBCS and ZRBCS without certain degrees.

For both FRBCS and ZRBCS, the algorithm finds a rule for each training pattern and iterates all patterns to compute the consequent class of this rule. Therefore, the time complexity of FRBCS without certain degrees is $O(n^2 * M * K)$. And ZRBCS needs to iterate all patterns to compute the second number, which means its time complexity is higher than FRBCS, $O(n^3 * M * K)$. Table 6 lists the average running time (including training time and testing time) of FRBCS and ZRBCS without certain degrees over 20 datasets. It can be seen that the proposed ZRBCS is faster than FRBCS without certain degrees.

5.3. Comparison with certain degrees

Certain degrees adjust the fuzzy partitions vertically, and Z-numbers adjust them horizontally. On the basis of ZRBCS, certain degrees further improve the system's learning and prediction capabilities. Table 7 and Table 8 show the classification accuracy of the FRBCS and ZRBCS with certain degrees over the training data and testing data, respectively. Bold-faced results indicate that the current method is superior to the other one. As listed in Table 7, the ZRBCS with certain degrees shows better learning ability than FRBCS over all datasets. This model performs better on predictive ability in 18, 18, 18, 12 and 11 out of 20 datasets. It is not worse in 19, 19, 19, 15 and 15 datasets, respectively, when $K = 2, 3, 4, 5$ and 6 . The ZRBCS outperforms FRBCS for most datasets. To compare the results statistically, the Wilcoxon signed-rank test with $\alpha = 0.05$ with two-side is used to compare these two methods on predictive ability. The statistic values and p-values are shown in Table 9. The test rejects all hypotheses that there are statistically significant differences between the two methods. Therefore, we can conclude that ZRBCS is superior to FRBCS with certain degrees.

For both FRBCS and ZRBCS, the algorithm finds a rule for each training pattern and iterates all patterns to compute the consequent class and certain degree of this rule. Therefore, the time complexity of FRBCS with certain degrees is $O(n^2 * M * K)$. And ZRBCS needs to iterate all patterns to compute the second number, thus its time complexity is higher than FRBCS, $O(n^3 * M * K)$. Table 10 lists the average running time (including training time and testing time) of FRBCS and ZRBCS with certain degrees over 19 datasets. It can be seen that FRBCS is faster than the proposed ZRBCS with certain degrees.

	K=2		K=3		K=4		K=5		K=6	
	FRBCS	ZRBCS	FRBCS	ZRBCS	FRBCS	ZRBCS	FRBCS	ZRBCS	FRBCS	ZRBCS
Banknote	76.68%	84.95%	95.12%	96.05%	94.91%	96.31%	96.39%	97.16%	99.34%	99.42%
Breast	75.81%	81.57%	90.18%	90.49%	88.87%	91.62%	91.68%	92.42%	91.97%	93.40%
Bupa	58.00%	58.16%	59.87%	62.68%	67.79%	71.43%	73.53%	74.04%	75.91%	77.49%
Diabetes	65.76%	67.29%	75.49%	78.82%	80.87%	82.48%	85.79%	86.46%	90.19%	90.31%
Ecoli	72.06%	74.21%	77.74%	83.53%	86.04%	88.82%	92.92%	93.59%	92.89%	93.52%
Glass	53.43%	59.40%	71.55%	62.93%	65.73%	71.50%	79.86%	79.34%	78.61%	81.31%
Haberman	73.75%	73.71%	74.26%	74.84%	75.42%	77.41%	79.08%	79.05%	79.56%	79.99%
HTRU-2	93.36%	95.33%	95.60%	96.78%	96.36%	96.98%	97.11%	97.52%	97.57%	97.66%
Ionosphere	91.23%	93.67%	97.66%	97.97%	99.49%	99.49%	99.43%	99.43%	100.0%	100.0%
Iris	67.48%	78.15%	93.78%	94.89%	89.26%	93.11%	96.15%	96.07%	98.15%	97.56%
Liver	58.00%	58.16%	59.68%	61.80%	68.02%	71.47%	73.43%	74.17%	75.84%	77.62%
Newthyroid	79.54%	82.43%	86.15%	90.13%	91.63%	94.42%	93.44%	94.94%	93.54%	94.94%
Occupancy	85.44%	85.80%	93.03%	96.11%	97.05%	97.77%	97.89%	97.86%	97.79%	97.79%
Pageblocks	89.99%	90.83%	91.93%	92.71%	92.75%	93.29%	93.72%	94.25%	94.62%	95.04%
Pendigits	85.13%	91.71%	98.18%	98.53%	99.35%	99.50%	99.79%	99.82%	99.96%	99.97%
Pima	65.29%	66.19%	75.22%	77.37%	80.02%	80.66%	84.62%	85.07%	88.79%	89.00%
Rice	91.12%	91.82%	88.26%	91.59%	91.99%	92.39%	91.55%	91.99%	92.83%	92.87%
Segment	64.85%	69.10%	85.66%	88.67%	91.25%	91.63%	94.11%	93.87%	96.43%	96.52%
Tranfusion	76.46%	76.75%	76.59%	77.12%	77.06%	78.02%	77.45%	77.87%	78.06%	79.65%
Vehicle	41.84%	47.60%	69.88%	70.19%	84.84%	86.01%	92.84%	93.42%	98.24%	98.23%

Table 7: Classification accuracy rate (%) of the FRBCS and ZRBCS with certain degrees over training data.

	K=2		K=3		K=4		K=5		K=6	
	FRBCS	ZRBCS	FRBCS	ZRBCS	FRBCS	ZRBCS	FRBCS	ZRBCS	FRBCS	ZRBCS
Banknote	76.60%	85.13%	94.97%	95.92%	94.68%	96.21%	96.36%	97.08%	99.20%	99.27%
Breast	75.57%	81.55%	89.63%	90.34%	88.58%	91.03%	90.52%	91.05%	90.34%	91.92%
Bupa	57.89%	57.89%	57.87%	58.74%	61.34%	62.46%	59.42%	60.56%	57.03%	58.12%
Diabetes	65.63%	66.54%	72.79%	75.14%	72.92%	72.66%	71.48%	71.62%	69.02%	68.37%
Ecoli	69.96%	71.47%	73.22%	79.19%	73.86%	76.81%	79.20%	79.47%	77.69%	77.09%
Glass	50.50%	56.98%	64.36%	55.35%	52.66%	57.93%	61.36%	61.13%	56.97%	60.94%
Haberman	73.53%	72.54%	73.19%	74.17%	71.23%	73.82%	72.54%	70.59%	70.25%	70.25%
HTRU-2	93.40%	95.31%	95.59%	96.78%	96.36%	96.95%	97.05%	97.46%	97.44%	97.51%
Ionosphere	65.52%	68.67%	64.65%	64.94%	56.70%	56.70%	54.40%	54.40%	45.86%	45.86%
Iris	67.33%	78.00%	92.67%	94.67%	87.33%	89.33%	95.33%	95.33%	93.33%	92.00%
Liver	57.98%	58.27%	58.54%	60.01%	61.77%	62.96%	62.04%	60.87%	56.83%	57.97%
Newthyroid	79.57%	81.88%	84.70%	88.42%	89.85%	93.10%	91.26%	92.64%	90.26%	92.12%
Occupancy	85.48%	85.63%	93.09%	96.17%	97.04%	97.71%	97.82%	97.82%	97.71%	97.71%
Pageblocks	89.95%	90.81%	91.76%	92.47%	92.38%	93.00%	93.13%	94.25%	93.86%	94.32%
Pendigits	84.87%	91.35%	97.52%	97.85%	97.88%	98.03%	95.88%	95.90%	90.44%	90.45%
Pima	65.11%	65.89%	73.06%	74.88%	72.53%	73.18%	71.76%	72.14%	69.40%	69.53%
Rice	91.13%	91.92%	88.19%	91.52%	92.02%	92.36%	91.31%	91.76%	92.76%	92.60%
Segment	64.72%	68.61%	84.03%	87.01%	88.74%	89.05%	90.30%	90.00%	91.60%	91.47%
Tranfusion	76.07%	76.74%	76.61%	76.61%	76.87%	77.67%	76.47%	76.34%	77.14%	78.61%
Vehicle	39.61%	45.51%	63.01%	63.37%	65.97%	66.92%	63.10%	63.11%	62.40%	62.40%

Table 8: Classification accuracy rate (%) of the FRBCS and ZRBCS with certain degrees over testing data.

K	Statistic value	p-value	Critical value	Hypothesis
$K = 2$	10	9.61658e-05	52	Rejected
$K = 3$	10	9.61658e-05	52	Rejected
$K = 4$	10	9.61658e-05	52	Rejected
$K = 5$	45	0.08956	52	Rejected
$K = 4$	42.5	0.13361	52	Rejected

Table 9: Wilcoxon signed-rank test of the accuracy rate for comparing ZRBCS to FRBCS with certain degrees.

	K=2		K=3		K=4		K=5		K=6	
	FRBCS	ZRBCS	FRBCS	ZRBCS	FRBCS	ZRBCS	FRBCS	ZRBCS	FRBCS	ZRBCS
Banknote	55.6	81.3	74.3	101	87.5	116.5	94.6	136.3	109.7	139.5
Breast	40.1	56	47.7	67.4	59.1	66.3	64.5	79.9	73.1	81.2
Bupa	33.8	40.2	46.1	51.1	56.5	61.8	63	67.8	75.5	79.2
Diabetes	66.1	93.7	100	123.8	184	202.5	237.2	261.2	324.5	345.2
Ecoli	35.6	44.1	50.5	56.9	66	78	78.4	81.8	94	107.7
Glass	31.6	42.5	38.8	45.8	47.8	54.7	53.9	60.2	60.5	63.7
Haberman	36.7	42.4	39.6	47.9	46.4	48.1	48.3	52.3	54	54.4
HTRU-2	3564.7	7635.4	4444.3	8837.5	5626.8	11205.5	5428.1	11437.4	6685.8	13123.5
Ionosphere	306	329.1	316.7	332.8	441.2	437.7	379.4	401.6	437.3	436.8
Iris	33.7	36.3	38.2	41.1	42.9	43.6	46.4	52.9	51	52.7
Liver	33.5	40.9	43.9	50.6	56.7	62.2	66.6	71.4	75.7	79.1
Newthyroid	31.2	35.5	38.6	41.2	44.1	51.6	50.2	53	51.3	52.6
Occupancy	110.2	171.4	104.7	151.7	116.2	158.1	125.6	165.9	130.5	178.7
Pageblocks	543.5	1099.7	616.7	1239.2	750.5	1395.7	810.6	1410.1	881	1459.1
Pendigits	6895.8	10620.3	16092.3	20190.9	24885.2	30204	32085.8	39141.9	36378.8	44502.1
Pima	107.1	128.1	125.8	150	196.8	224.9	264.2	280.7	324.6	344.9
Rice	351.1	517	475.2	774.3	582.7	885.6	639.2	932.9	723.8	1041.5
Segment	304.1	474.2	628.9	802.4	1055.7	1179.1	1408.7	1579.7	1765	2039.2
Tranfusion	40.9	61.8	48.6	62.1	49	63.3	51.9	65	56	71.9
Vehicle	124.8	169.6	332.3	377.1	708.2	667.6	828.1	833.6	932.7	999.2
Average	637.31	1085.98	1185.16	1677.24	1755.17	2360.34	2141.24	2858.28	2464.24	3262.61

Table 10: Running time (in millisecond) of the FRBCS and ZRBCS with certain degrees.

	SVM	kernel	CamNN	FRBCS	ZRBCS
Banknote	100.00% (1)	76.97% (5)	99.42% (2)	99.20% (4)	99.27% (3)
Breast	84.35% (5)	89.28% (4)	89.99% (3)	90.52% (2)	91.92% (1)
Bupa	63.16% (2)	68.92% (1)	54.43% (5)	61.34% (4)	62.46% (3)
Diabetes	64.84% (5)	72.39% (3)	68.36% (4)	72.92% (2)	75.14% (1)
Ecoli	61.96% (5)	81.58% (1)	80.69% (2)	79.20% (4)	79.47% (3)
Glass	70.01% (1)	35.57% (5)	46.42% (4)	64.36% (2)	61.13% (3)
Haberman	70.25% (4)	72.23% (3)	66.37% (5)	73.53% (2)	74.17% (1)
HTRU-2	96.44% (5)	97.28% (4)	97.36% (3)	97.44% (2)	97.51% (1)
Iris	97.33% (1)	91.33% (5)	94.67% (4)	95.33% (2.5)	95.33% (2.5)
Liver	63.50% (2)	66.37% (1)	55.66% (5)	62.04% (4)	62.96% (3)
Newthyroid	93.96% (1)	88.40% (4)	87.45% (5)	91.26% (3)	93.10% (2)
Occupancy	97.90% (3)	98.31% (1.5)	98.31% (1.5)	97.82% (4.5)	97.82% (4.5)
Pageblocks	92.60% (4)	95.60% (1)	70.36% (5)	93.86% (3)	94.32% (2)
Pendigits	61.09% (5)	99.39% (1)	97.21% (4)	97.88% (3)	98.03% (2)
Pima	63.41% (5)	70.32% (3)	69.27% (4)	73.06% (2)	74.88% (1)
Rice	68.82% (5)	87.69% (4)	90.63% (3)	92.76% (1)	92.60% (2)
Tranfusion	71.25% (4)	77.13% (3)	67.38% (5)	77.14% (2)	78.61% (1)
Average rank	3.5	2.8611	3.8056	2.78	2.0556

Table 11: Classification accuracy rates (%) of SVM, kernel, CamNN, FRBCS and ZRBCS over testing data.

Statistic(χ_F)	Critical value	p-value	Hypothesis
3.8834	2.507	0.0067	Rejected

Table 12: The Friedman test result with $\alpha = 0.05$.

5.4. Synthetic comparison

To demonstrate the advantages of the system in a more comprehensive and visual way, Fig. 6 shows the classification accuracy rate of each dataset over testing data. The test accuracy for ZRBCS is always higher than that for FRBCS for most datasets. The ZRBCS with certain degrees obtains the highest classification accuracy rate in 16 out of 20 datasets. And we can also see that the learning ability of the model increases as the value of K increases. The predictive ability of the model does not always increase, especially for data sets with more features. This is caused by the limited number of training patterns. For smaller data sets ($M < 10$), we suggest using a partition number ($K = 5$); in other cases, a partition number ($K = 3$) is recommended to achieve a better balance between accuracy and complexity.

To further demonstrate the superiority of the proposed method, we further compare the ZRBCS with three classical classification algorithms, including support vector machine(SVM) [8], kernel classifier [9] and cam weighted distance nearest neighbor Classifier (CamNN)[10]. Since two datasets Ionosphere and Segment cannot be trained and tested by CamNN, there are 18 datasets to be compared. The classification accuracy rates over testing data are shown in Table 11. In which, the classification accuracy rates of FRBCS and ZRBCS are taken from the optimal values in Table 8. The values in parentheses mark the ranks of the classifiers: The better the performance, the higher the rank (in case of ties, average ranks are assigned). The last row lists the average rank of each classifier. It is shown that the proposed ZRBCS is superior to other methods. Nonparametric tests based on average ranks are used to compare classifiers statistically. First, the Friedman test is carried out to determine whether the results obtained by different methods present significant differences. Table 12 shows the result of Friedman test with a level of significance $\alpha = 0.05$. It can be seen that the statistic value is greater than the critical value, which means there are significant differences among the results. After that the Nemenyi test is used to compare the proposed ZRBCS with other ones, and the results are shown in Fig. 7. It is shown that the proposed ZRBCS is not significantly better than the methods kernel, but is significantly better than the methods SVM and CamNN. The superiority of the ZRBCS over FRBCS is discussed in the Section 5.2 and Section 5.3.

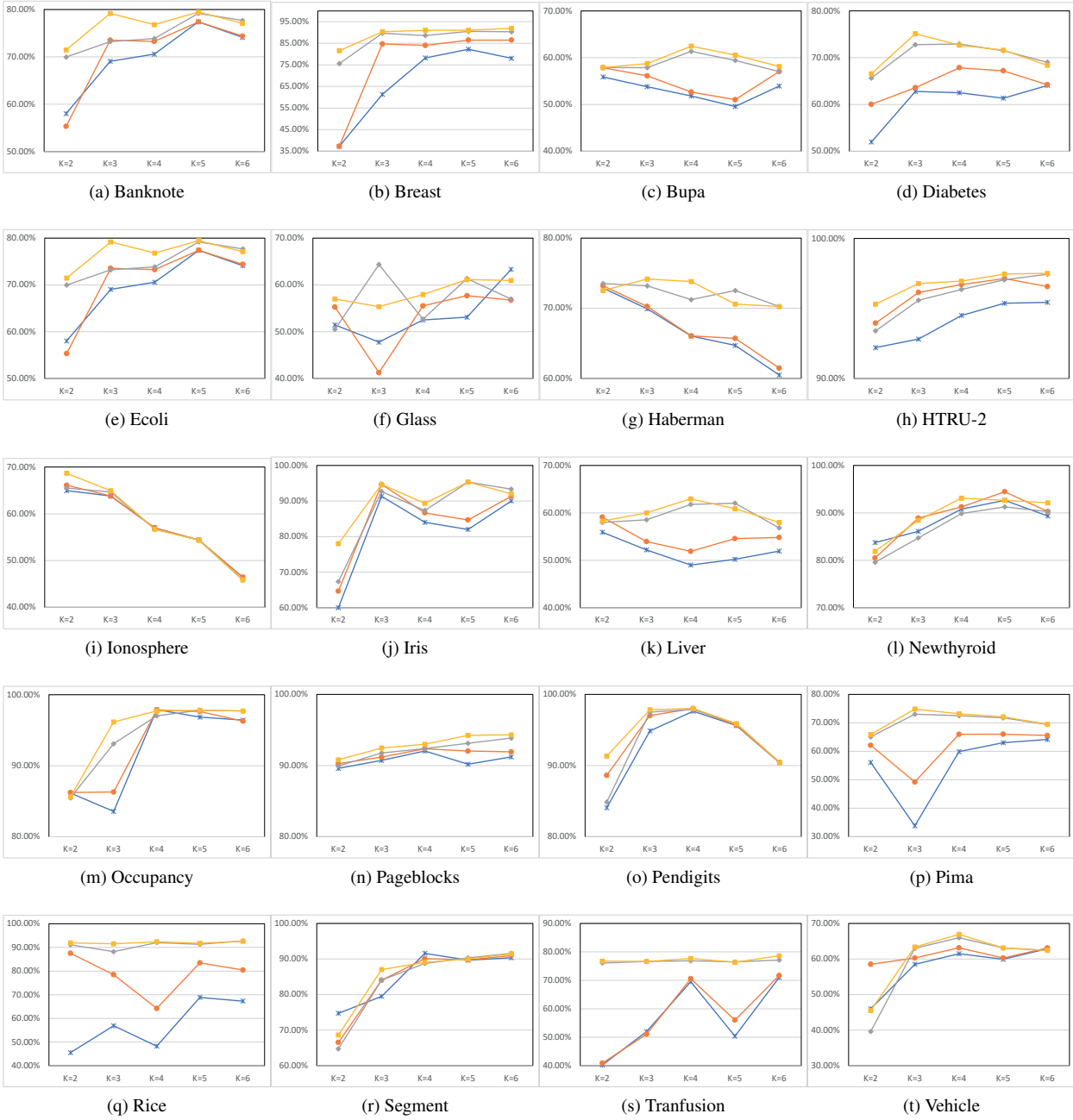
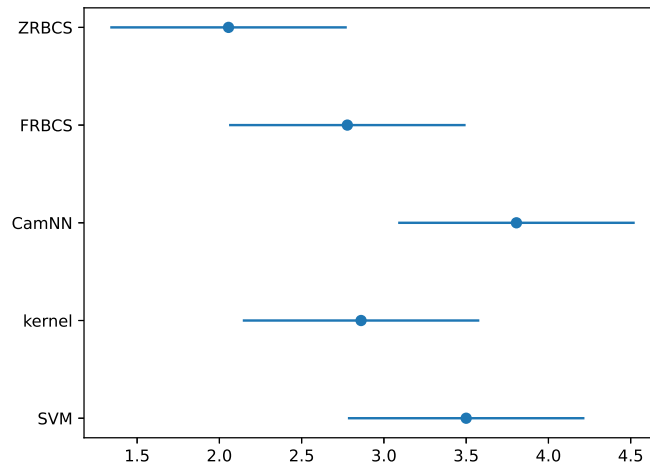


Figure 6: Classification accuracy rate (%) of FRBCS and ZRBCS over testing data. The symbol 'x' denotes the FRBCS without certain degrees, '•' denotes the ZRBCS without certain degrees, '♦' denotes the FRBCS with certain degrees, and '■' denotes the ZRBCS with certain degrees.

6. Conclusions

The Z-number is more suitable for describing uncertain and partially reliable information due to its model feature. In this paper, we extend the fuzzy if-then rules in the Z-number-valued framework to propose Z-number-valued if-then rules. Then based on the suggested form of the Z-number-valued if-then rule, a basic Z-number-valued rule-based classification system is established. The proposed system is based on the negative patterns of the fuzzy rule-based

Figure 7: The Nemenyi test result with $\alpha = 0.05$.

system. The first fuzzy numbers of the antecedent Z-numbers are generated by the classical fuzzy rule extraction method, the fuzzy simple grid method. In the second fuzzy numbers of antecedent Z-numbers, information about negative patterns is stored, such as maximum, average, and minimum values. When classifying, the second fuzzy number can adjust the first fuzzy number based on negative patterns. Three experiments of 20 datasets have been conducted to demonstrate the learning and predictive abilities and running time of our classification system, and the results show that our method achieves better performance than the basic fuzzy rule-based system and two classical classification algorithms, SVM and CamNN.

Due to the proposed system in this paper just being a basic Z-number-valued rule-based system, it must be noted that the proposed ZRBCS, like FRBCS, cannot handle incomplete and unbalanced data. We can use some auxiliary means to further improve classification accuracy, interpretability and handle incomplete and unbalanced data, such as genetic algorithms and clustering algorithms, and to reduce the dimensionality based on this basic Z-number-valued rule-based system in further work.

Acknowledgements

This work was supported by the project B-TIC-590-UGR20 co-funded by the Programa Operativo FEDER 2014-2020 and the Regional Ministry of Economy, Knowledge, Enterprise and Universities (CECEU) of Andalusia, by the China Scholarship Council (CSC), by the project PID2019-103880RB-I00 funded by MCIN / AEI / 10.13039/501100011033 and by the Andalusian government through project P20_00673.

References

- [1] H. Ishibuchi, K. Nozaki, H. Tanaka, Distributed representation of fuzzy rules and its application to pattern classification, *Fuzzy sets and systems* 52 (1) (1992) 21–32.
- [2] K. Siminski, An outlier-robust neuro-fuzzy system for classification and regression, *International Journal of Applied Mathematics and Computer Science* 31 (2).
- [3] J. Niu, D. Chen, J. Li, H. Wang, A dynamic rule-based classification model via granular computing, *Information Sciences* 584 (2022) 325–341.
- [4] C. Fu, W. Lu, W. Pedrycz, J. Yang, Rule-based granular classification: A hypersphere information granule-based method, *Knowledge-Based Systems* 194 (2020) 105500.
- [5] Y. Farjami, K. Abdi, A genetic-fuzzy algorithm for spatio-temporal crime prediction, *Journal of Ambient Intelligence and Humanized Computing* (2021) 1–13.
- [6] E. G. Mansoori, Frbc: A fuzzy rule-based clustering algorithm, *IEEE transactions on fuzzy systems* 19 (5) (2011) 960–971.

- [7] L. A. Zadeh, A note on z-numbers, *Information Sciences* 181 (14) (2011) 2923–2932.
- [8] C. Cortes, V. Vapnik, Support-vector networks, *Machine learning* 20 (3) (1995) 273–297.
- [9] G. J. McLachlan, *Discriminant analysis and statistical pattern recognition*, John Wiley & Sons, 2005.
- [10] C. Y. Zhou, Y. Q. Chen, Improving nearest neighbor classification with cam weighted distance, *Pattern recognition* 39 (4) (2006) 635–645.
- [11] M. Woźniak, A. Zielonka, A. Sikora, Driving support by type-2 fuzzy logic control model, *Expert Systems with Applications* 207 (2022) 117798.
- [12] M. Woźniak, A. Zielonka, A. Sikora, M. J. Piran, A. Alamri, 6g-enabled iot home environment control using fuzzy rules, *IEEE Internet of Things Journal* 8 (7) (2020) 5442–5452.
- [13] H. Ishibuchi, K. Nozaki, H. Tanaka, Efficient fuzzy partition of pattern space for classification problems, *Fuzzy Sets and Systems* 59 (3) (1993) 295–304.
- [14] H. Ishibuchi, K. Nozaki, N. Yamamoto, H. Tanaka, Selecting fuzzy if-then rules for classification problems using genetic algorithms, *IEEE Transactions on fuzzy systems* 3 (3) (1995) 260–270.
- [15] D. Nauck, R. Kruse, A neuro-fuzzy method to learn fuzzy classification rules from data, *Fuzzy sets and Systems* 89 (3) (1997) 277–288.
- [16] H. Ishibuchi, T. Nakaskima, Improving the performance of fuzzy classifier systems for pattern classification problems with continuous attributes, *IEEE Transactions on industrial electronics* 46 (6) (1999) 1057–1068.
- [17] D. Chakraborty, N. R. Pal, A neuro-fuzzy scheme for simultaneous feature selection and fuzzy rule-based classification, *IEEE Transactions on Neural Networks* 15 (1) (2004) 110–123.
- [18] Y. Chen, J. Z. Wang, Support vector learning for fuzzy rule-based classification systems, *IEEE Transactions on fuzzy systems* 11 (6) (2003) 716–728.
- [19] H. Ishibuchi, T. Yamamoto, Rule weight specification in fuzzy rule-based classification systems, *IEEE transactions on fuzzy systems* 13 (4) (2005) 428–435.
- [20] E. G. Mansoori, M. J. Zolghadri, S. D. Katebi, A weighting function for improving fuzzy classification systems performance, *Fuzzy sets and systems* 158 (5) (2007) 583–591.
- [21] J. Alcalá-Fdez, R. Alcalá, F. Herrera, A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning, *IEEE Transactions on Fuzzy systems* 19 (5) (2011) 857–872.
- [22] M. Elkano, M. Galar, J. Sanz, H. Bustince, Chi-bd: A fuzzy rule-based classification system for big data classification problems, *Fuzzy Sets and Systems* 348 (2018) 75–101.
- [23] J. Niu, D. Chen, J. Li, H. Wang, Fuzzy rule-based classification method for incremental rule learning, *IEEE Transactions on Fuzzy Systems* 30 (9) (2021) 3748–3761.
- [24] S. Mitra, Y. Hayashi, Neuro-fuzzy rule generation: survey in soft computing framework, *IEEE transactions on neural networks* 11 (3) (2000) 748–768.
- [25] J. Sanz, M. Sesma-Sara, H. Bustince, A fuzzy association rule-based classifier for imbalanced classification problems, *Information Sciences* 577 (2021) 265–279.
- [26] A. Azadeh, M. Saberi, N. Z. Atashbar, E. Chang, P. Pazhoeshfar, Z-ahp: A z-number extension of fuzzy analytical hierarchy process, in: 2013 7th IEEE International Conference on Digital Ecosystems and Technologies (DEST), IEEE, 2013, pp. 141–147.
- [27] H. Aboutorab, M. Saberi, M. R. Asadabadi, O. Hussain, E. Chang, Zbw: The z-number extension of best worst method and its application for supplier development, *Expert Systems with Applications* 107 (2018) 115–125.
- [28] R. A. Krohling, A. G. Pacheco, G. A. dos Santos, Todim and topsis with z-numbers, *Frontiers of Information Technology & Electronic Engineering* 20 (2) (2019) 283–291.
- [29] K.-w. Shen, J.-q. Wang, Z-vikor method based on a new comprehensive weighted distance measure of z-number and its application, *IEEE Transactions on Fuzzy Systems* 26 (6) (2018) 3232–3245.
- [30] A. Garg, S. Das, J. Maiti, S. K. Pal, Granulized z-vikor model for failure mode and effect analysis, *IEEE Transactions on Fuzzy Systems* (2020) 1–1doi:10.1109/TFUZZ.2020.3037933.
- [31] R. A. Aliev, B. G. Guirimov, O. H. Huseynov, R. R. Aliyev, Z-relation equation-based decision making, *Expert Systems with Applications* 184 (2021) 115387.
- [32] R. Banerjee, S. K. Pal, On z-numbers and the machine-mind for natural language comprehension, in: *Fifty Years of Fuzzy Logic and its Applications*, Springer, 2015, pp. 415–457.
- [33] R. Banerjee, S. K. Pal, Z*-numbers: Augmented z-numbers for machine-subjectivity representation, *Information Sciences* 323 (2015) 143–178.
- [34] R. H. Abiyev, Dynamic system control using z-number based fuzzy neural network, in: *International Conference on Intelligent and Fuzzy Systems*, Springer, 2020, pp. 964–972.
- [35] I. M. Jiskani, F. Yasli, S. Hosseini, A. U. Rehman, S. Uddin, Improved z-number based fuzzy fault tree approach to analyze health and safety risks in surface mines, *Resources Policy* 76 (2022) 102591.
- [36] M. Abbaspour Onari, S. Yousefi, M. Jahangoshai Rezaee, Risk assessment in discrete production processes considering uncertainty and reliability: Z-number multi-stage fuzzy cognitive map with fuzzy learning algorithm, *Artificial Intelligence Review* 54 (2) (2021) 1349–1383.
- [37] D. Wu, X. Liu, F. Xue, H. Zheng, Y. Shou, W. Jiang, A new medical diagnosis method based on z-numbers, *Applied Intelligence* 48 (4) (2018) 854–867.
- [38] R. Yazdanparast, R. Tavakkoli-Moghaddam, R. Heidari, L. Aliabadi, A hybrid z-number data envelopment analysis and neural network for assessment of supply chain resilience: a case study, *Central European Journal of Operations Research* 29 (2) (2021) 611–631.
- [39] Z. Bobar, D. Božanić, K. Djurić, D. Pamučar, Ranking and assessment of the efficiency of social media using the fuzzy ahp-z number model-fuzzy mabac, *Acta Polytechnica Hungarica* 17 (3) (2020) 43–70.
- [40] R. A. Aliev, W. Pedrycz, B. Guirimov, O. H. Huseynov, Clustering method for production of z-number based if-then rules, *Information Sciences* 520 (2020) 155–176.
- [41] R. A. Aliev, A. V. Alizadeh, O. H. Huseynov, The arithmetic of discrete z-numbers, *Information Sciences* 290 (2015) 134–155.

- [42] R. A. Aliev, O. H. Huseynov, L. M. Zeinalova, The arithmetic of continuous z-numbers, *Information Sciences* 373 (2016) 441–460.
- [43] R. A. Aliev, A. V. Alizadeh, O. H. Huseynov, K. Jabbarova, Z-number-based linear programming, *International Journal of Intelligent Systems* 30 (5) (2015) 563–589.
- [44] R. Aliev, Operations on z-numbers with acceptable degree of specificity, *Procedia computer science* 120 (2017) 9–15.
- [45] R. A. Aliev, W. Pedrycz, O. H. Huseynov, S. Z. Eyupoglu, Approximate reasoning on a basis of z-number-valued if–then rules, *IEEE Transactions on Fuzzy Systems* 25 (6) (2017) 1589–1600.
- [46] B. Kang, Y. Deng, R. Sadiq, Total utility of z-number, *Applied Intelligence* 48 (3) (2018) 703–729.
- [47] M. Mazandarani, Y. Zhao, Z-differential equations, *IEEE Transactions on Fuzzy Systems* 28 (3) (2019) 462–473.
- [48] J. Alcalá-Fdez, L. Sanchez, S. Garcia, M. J. del Jesus, S. Ventura, J. M. Garrell, J. Otero, C. Romero, J. Bacardit, V. M. Rivas, et al., Keel: a software tool to assess evolutionary algorithms for data mining problems, *Soft Computing* 13 (3) (2009) 307–318.

3 Z-number-valued rule-based decision trees

- Y., Li, E. Herrera-Viedma, G. Kou, and J. A. Morente-Molinera, Z-number-valued rule-based decision trees, *Information Sciences*, 643, 119252, 2023, <https://doi.org/10.1016/j.ins.2023.119252>.
 - Status: **Published**.
 - Impact Factor (JCR 2023): 8.1
 - Subject Category: Computer Science, Information Systems, Ranking 13/158 (Q1).

Z-number-Valued Rule-Based Decision Trees

Yangxue Li^a, Enrique Herrera-Viedma^{a,b,*}, Gang Kou^c, Juan Antonio Morente-Molinera^{a,*}

^aAndalusian Research Institute in Data Science and Computational Intelligence, Dept. of Computer Science and AI, University of Granada, 18071 Granada, Spain.

^bDepartment of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia

^cSchool of Business Administration, Southwestern University of Finance and Economics, Chengdu 611130, P. R. China

Abstract

As a novel architecture of a fuzzy decision tree constructed on fuzzy rules, the fuzzy rule-based decision tree (FRDT) achieved better performance in terms of both classification accuracy and the size of the resulting decision tree than other classical decision trees such as C4.5, LADtree, BFTree, SimpleCart and NBTree. The concept of Z-number extends the classical fuzzy number to model both uncertain and partial reliable information. Z-numbers have significant potential in rule-based systems due to their strong representation capability. Firstly, this paper designs a Z-number-valued rule-based decision tree (ZRDT) and provides the learning algorithm. The information gain is used to replace the fuzzy confidence in FRDT to select features in each rule. Additionally, we use the negative samples to generate second fuzzy Z-numbers that adjust the first fuzzy numbers and improve the model's fit to the training data. The proposed ZRDT is compared with the FRDT with three different parameter values, the FRDT with information gain in terms of learning ability, predictive ability, and size of the produced decision tree. As a result of the statistical tests, the proposed ZRDT has the highest classification performance with a relatively small size for the reduced decision tree.

Keywords: Z-numbers, rule-based system, decision trees, Z-number-valued rule, classification, information gain

1. Introduction

Decision trees [7, 36, 14] are a common classification method, which is a machine learning algorithm with simple logic. The model of the decision tree is based on a series of if-then-else rules obtained from training data. It is easy to implement, highly interpretable, fully compatible with human intuitive thinking, and able to treat large-scale data. Many decision trees have been introduced to solve classification problems, such as Iterative Dichotomies 3 (ID3) [10], Successor of ID3 (C4.5) [8], classification and regression tree (CART) [5], Chi-squared automatic interaction detector (CHAID) [29], multivariate adaptive regression spline (MARS) [25], generalized, unbiased, interaction detection and estimation (GUIDE), conditional inference trees (CTREE) [23, 15], and so on.

The classical decision tree is also considered a classifier that is very unstable to small changes in the training data, in other words, a method with high variance. With the elasticity of the fuzzy set formalism, fuzzy logic can improve this problem. A variant of classical decision tree inductive learning using fuzzy set theory is proposed, called fuzzy decision trees or soft decision trees [34, 19]. After that, some advanced fuzzy decision trees are introduced, such as optimization of fuzzy decision trees [30], flexible fuzzy decision trees (FlexDT) [16], multi flexible fuzzy decision tree (MFlexDT) [9], Chi-square-based multi flexible fuzzy decision trees (Chi-MFlexDT) [17] and fuzzy rule-based decision trees (FRDT) [31].

The FRDT is a special architecture of fuzzy decision trees, which is an oblique tree assembled by layers. Each node is a fuzzy rule with a unique consequent class, which is written as:

IF input sample is antecedent fuzzy numbers THEN consequent class.

*Corresponding authors.

Email addresses: viedma@decsai.ugr.es (Enrique Herrera-Viedma), jamoren@decsai.ugr.es (Juan Antonio Morente-Molinera)

Each class corresponds to only one fuzzy rule in a layer in order to minimize the size of the decision tree. If a sample cannot be classified in the fuzzy rules of this layer, it will be sent to the next layer. Compared to other decision tree algorithms, the FRDT exhibits superior performance in terms of both accuracy and the size of trees generated. However, the FRDT can be improved. First, the initial fuzzy numbers cannot fit the training data well. The antecedent fuzzy numbers of fuzzy rules for a layer are selected from a series of initial fuzzy numbers that consist of the average values from training data. The intersection region between adjacent initial fuzzy numbers is equally divided. This method does not fit the training data well. Moreover, the features selection approach cannot select the most appropriate features. A fuzzy rule does not consider all features but uses fuzzy confidence to select a few suitable ones from them. And the fuzzy confidence value of a fuzzy rule considered some features must be larger than that considered these features and another feature. Thus, the authors have to set a penalty coefficient to select the most suitable features. When two fuzzy rules, one taking into account one more feature than another, have a fuzzy confidence difference greater than this coefficient, the fuzzy rule with the greater confidence difference will be adopted. However, fuzzy confidence has not been shown to be directly related to the effect of the decision tree.

For the first problem, the concept of Z-numbers can be used to extend the fuzzy rule. A Z-number is an ordered pair of fuzzy numbers, where the first one restricts the values of the variable, and the second one measures the reliability of the first [35]. Because of the special form of the Z-number, it can describe information with both uncertainty and partial reliability. Uncertain and partially reliable information is the most common information in the real world. Since the introduction of Z-number, it has demonstrated its powerful representative ability in many fields, such as decision making [3, 20], information fusion [26], failure analysis [33, 11], linguistic information processing [6, 28] and game theory [12]. In this paper, we define the form of the Z-number-valued if-then rule written as:

IF input sample is antecedent Z-numbers THEN consequent class.

The second fuzzy number can adjust the first one to fit the training data better. If the fuzzy rules are replaced by the Z-number-valued rule, the FRDT is extended to a new decision tree structure, named the Z-number-valued rule-based decision tree (ZRDT). We also provide the learning algorithm of the ZRDT: Firstly, a series of initial fuzzy numbers are obtained by the average values from training data as the first fuzzy numbers of the antecedent Z-numbers. Then the features in a rule are selected by information gain. The second fuzzy numbers are structured by negative samples covered by the rule. If a rule does not cover any negative samples, we believe that this rule is totally reliable, and its second fuzzy numbers should be 1 (crisp numbers are a special form of fuzzy numbers).

For the second problem, information gain can replace the fuzzy confidence in FRDT to select features for a rule. Information gain is a common index to measure the impurity of a node in decision tree algorithms [24, 32]. In addition information gain, GINI index [27, 18], gain ratio [4, 13], and misclassification rate [1] are common in decision tree algorithms. Information gain is based on entropy, which is the difference between the entropy of a class and the conditional entropy of a class for the selected feature. In this paper, features are chosen so as to maximize the FOIL's information gain [21, 22], which is a measure of improvement of the rule in comparison with the default rule for the target class. In this case, we can discard the penalty coefficient parameter for comparing the fuzzy confidences.

The main contributions of this paper can be summarized as follows.

- (1) The fuzzy rule-based decision tree learning algorithm is improved by using the information gain to select features.
- (2) Proposed the Z-number adjustment method to make the Z-numbers fit training data better than fuzzy numbers and define the structure of the Z-number-valued rule used in rule-based decision trees.
- (3) Provided the Z-number-valued rule-based decision tree (ZRDT) and provided its learning algorithm, the Z-number-valued rule extraction algorithm.
- (4) Some experiments are used to evaluate the learning ability, predictive ability, and size of the produced decision trees of the ZRDT compared with the FRDT and improved FRDT.

This study is set out as follows: Section 2 recalls some background knowledge of FRDT and the definition of Z-numbers. Section 3 details the algorithms of the ZRDT, including the second fuzzy number calculation algorithm, the Z-number-valued rules extraction algorithm, and the ZRDT overall algorithm. In Section 4, some experiments using 19 well-known benchmark datasets are performed to analyze and compare the learning ability, predictive ability, and size of the produced decision tree of the FRDT, improved FRDT, and ZRDT. Finally, we concluded this study in Section 5.

2. Preliminaries: fuzzy rule-based decision trees and Z-numbers theory

In this section, we review some fundamentals of the FRDT and the Z-numbers theory. Furthermore, In order to help understand the meaning of some notations and terminologies used in this paper, we briefly describe a few in Table 1.

Notation	Description
$X = \{x_i\}_{i=1}^{\chi}$	The dataset X with χ samples, where x_i is the i -th sample.
$x_i = \{x_{ij}\}_{j=1}^n$	The i -th sample x_i with n features, where x_{ij} is the value of sample with respect to j -th feature.
$Y = \{y_w\}_{w=1}^c$	The set of classes in dataset.
R	The fuzzy rule.
y_R	The consequent class of rule R .
$y_{l,w}$	The w -th class of layer l .
$A_{\rho(k)}$	The k -th antecedent fuzzy number of the fuzzy rule, where $\rho(k) \in \{1, \dots, n\}$.
$A_{jw} = (a_{jw}, b_{jw}, c_{jw}, d_{jw})$	The w -th fuzzy number for j -th feature.
s_{jw}	The mean value of class y_w on j -th feature.
$inf i$	The positive infinite value.
$-inf i$	The negative infinite value.
m	The number of considered features of the rule.
$m_{l,w}$	The number of considered features of rule $R_{Zl,w}$.
$FConf(R)$	The fuzzy confidence of fuzzy rule R .
$\phi_R(x)$	The average membership degree of sample x for rule R .
$\mu_{A_{\rho(k)}}(x_{i\rho(k)})$	The membership function of $x_{i\rho(k)}$ belonging to fuzzy number $A_{\rho(k)}$.
X_{y_R}	The set of samples belonging to class y_R .
X^{imp}	The additional impure node contains a set of uncertain samples.
α	The penalty coefficient parameter used to select features.
δ	The threshold parameter.
$l = 1, \dots, L$	The layer index, where L is the number of layers.
R_Z	The Z-number-valued rule.
$R_{Zl,w}$	The Z-number-valued rule in layer l with consequent class $y_{l,w}$.
$Z_{\rho(k)} = (A_{\rho(k)}, B_{\rho(k)})$	The k -th antecedent Z-number of the rule.
$A'_{\rho(k)}$	The adjusted fuzzy number of $Z_{\rho(k)} = (A_{\rho(k)}, B_{\rho(k)})$.
$IG(R)$	The FOIL information gain of rule R .
$Z_{l,w\rho(k)}(k_{l,w})$	The $k_{l,w}$ -th antecedent Z-number of rule $R_{Zl,w}$.
c_l	The number of classes of layer l .
$X_{l,w}$	The set of samples belonging to class $y_{l,w}$ in layer l .

Table 1: Descriptions of some notations used in this paper.

2.1. Fuzzy rule based decision tree

The FRDT is a special architecture of fuzzy decision trees. Unlike the tree structure of a traditional decision tree, FRDT is an oblique tree structure assembled by layers as shown in Fig. 1 [31]. Each layer includes one and only one pure leaf node for a class. It also includes one additional impure node used to contain the samples that are not covered by the current layer. If the impure node is not empty and different from the impure node of the previous layer, the next layer will be developed to classify the containing samples until the new impure node is empty or the same as the impure node of the previous layer. A pure leaf node of each layer corresponds to a fuzzy rule, which involves multiple features (variables) rather than a single feature considered at a node in traditional decision trees. The feature selection of a fuzzy rule is based on the index of fuzzy confidence, which measures the validity of a fuzzy rule. Some of the definitions and algorithms in FRDT are described in detail below.

A dataset X has χ samples and n features, $X = \{x_1, \dots, x_i, \dots, x_\chi\}$, where $x_i = \{x_{i1}, \dots, x_{ij}, \dots, x_{in}\}$, and x_{ij} is the value of x_i with respect to j -th feature. The dataset X contains c different classes, denoted as a set $Y = \{y_1, \dots, y_c\}$.

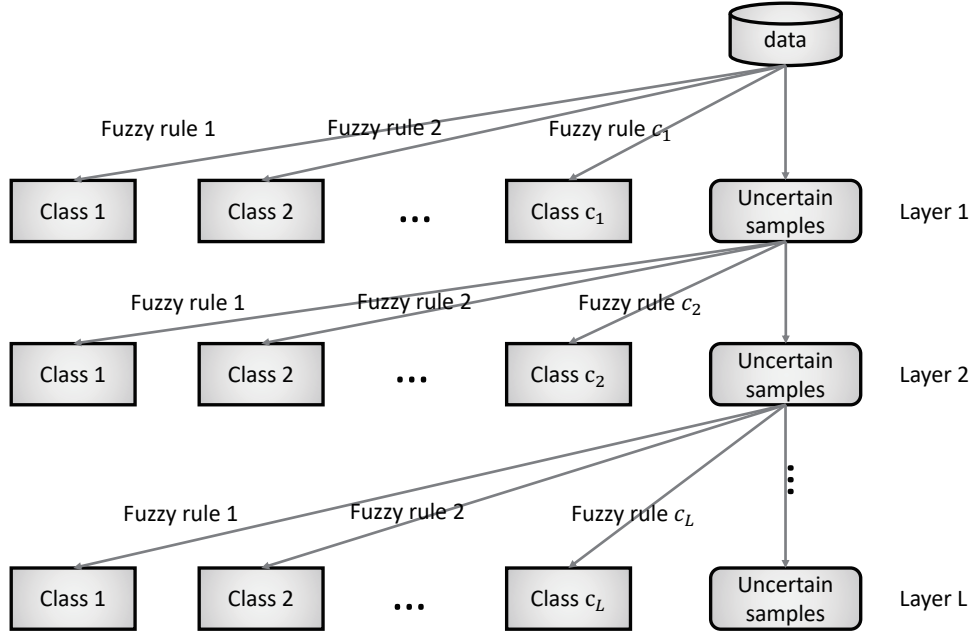


Figure 1: The overall structure of FRDT.

Definition 1. For a input sample $x_i = \{x_{i1}, \dots, x_{in}\}$, the formulation of a fuzzy rule R considered m features can be written as Eq. (1) [31].

$$R: \text{IF } x_{i\rho(1)} \text{ is } A_{\rho(1)} \text{ and } \dots \text{ and } x_{i\rho(m)} \text{ is } A_{\rho(m)} \text{ THEN } x_i \text{ belong to class } y_R, \quad (1)$$

where $\rho(1), \dots, \rho(m)$ are elements of $\{1, \dots, n\}$ and indicate the considered features; $A_{\rho(1)}, \dots, A_{\rho(m)}$ are the antecedent fuzzy numbers in $\rho(1)$ -th, $\dots$, $\rho(m)$ -th features respectively; $y_R \in Y$ is the consequent class of the fuzzy rule.

Definition 2. The fuzzy confidence of fuzzy rule R is defined as Eq. (2) [31].

$$FConf(R) = \frac{\sum_{x \in X_{y_R}} \phi_R(x)}{\sum_{x \in X} \phi_R(x)}, \quad (2)$$

$$\phi_R(x) = \frac{\sum_{k=1}^m \mu_{A_{\rho(k)}}(x_{i\rho(k)})}{m}, \quad (3)$$

where $\phi_R(x)$ is the average membership degree of a sample x for all antecedent fuzzy numbers in rule R ; $\mu_{A_{\rho(k)}}(x_{i\rho(k)})$ is the membership degree of $x_{i\rho(k)}$ belonging to fuzzy number $A_{\rho(k)}$; X_{y_R} is the set of all samples belonging to class y_R .

The FRDT also provides an algorithm to generate the initial fuzzy numbers used in fuzzy rules. For j -th feature, the number of initial fuzzy numbers equals the numbers of classes c of data to be processed. Let A_{jw} is the w -th ($w \in \{1, \dots, c\}$) fuzzy number defined for the j -th feature. A_{jw} is a trapezoidal fuzzy number, denoted as Eq. (4) [31].

$$A_{jw} = (a_{jw}, b_{jw}, c_{jw}, d_{jw}), \quad (4)$$

where $a_{jw}, b_{jw}, c_{jw}, d_{jw}$ are real numbers. These parameters are given by the mean value of all samples belonging to class y_w about j -th feature, ($w = 1, \dots, c$). Let the set $\{s_{j1}, \dots, s_{jc}\}$ is the mean values set of all classes on j -th feature, where s_{jw} is the mean value of class y_w on j -th feature calculated by Eq. (5).

$$s_{jw} = \frac{\sum_{x_i \in X_{y_w}} x_{ij}}{|X_{y_w}|}. \quad (5)$$

Then the parameters of fuzzy numbers can be given by Eqs. (6)-(9).

$$a_{jw} = \begin{cases} -infi, & w = 1, \\ s_{j\sigma(w-1)}, & \text{otherwise;} \end{cases} \quad (6)$$

$$b_{jw} = \begin{cases} -infi, & w = 1, \\ s_{j\sigma(w)}, & \text{otherwise;} \end{cases} \quad (7)$$

$$c_{jw} = \begin{cases} infi, & w = c, \\ s_{j\sigma(w)}, & \text{otherwise;} \end{cases} \quad (8)$$

$$d_{jw} = \begin{cases} infi, & w = c, \\ s_{j\sigma(w+1)}, & \text{otherwise.} \end{cases} \quad (9)$$

Where $\sigma(w)$ is the index function that means $s_{j\sigma(w)}$ is the w -th smallest element of $\{s_{j1}, \dots, s_{jc}\}$; $infi$ is the positive infinite value and $-infi$ is the negative infinite value. The generated fuzzy numbers of $\{s_{j1}, \dots, s_{jc}\}$ are portrayed in Fig. 2.

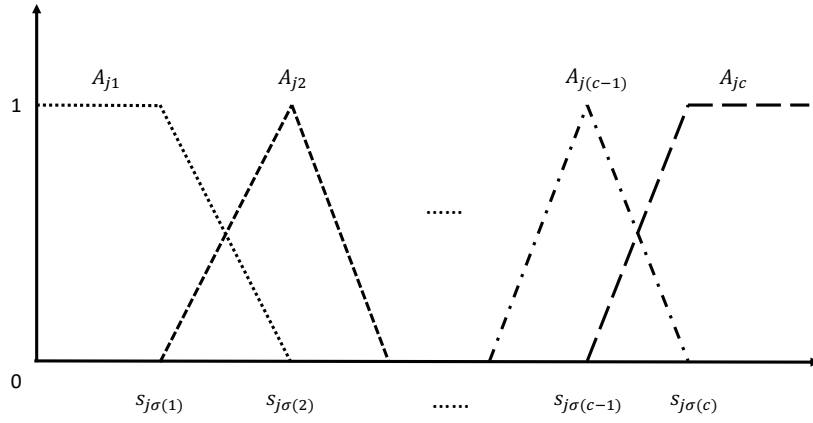


Figure 2: The generated initial fuzzy numbers on j -th feature.

For building the layers of the tree, FRDT provides the association rules extraction algorithm (AREA) to extract only one fuzzy rule for each class [31]. The fuzzy confidence is used as a criterion to select several promising features with high fuzzy confidence value. The specific algorithm process is shown in Algorithm 1. Where the penalty coefficient $\alpha = 0.02$ is a certain parameter.

After constructing the pure leaf nodes of a layer, the additional impure node contains a set of uncertain samples of that layer, denoted as X^{imp} , satisfied $\forall x \in X^{imp}, \forall R_w \in \{R_1, \dots, R_c\}, \phi_{R_w}(x) < \delta \in [0, 1]$. To summarize, the complete algorithm of FRDT is as Algorithm 2 [31].

2.2. Concept of Z-numbers

Definition 3. A Z-number is an ordered pair of fuzzy number,

$$Z = (A, B), \quad (10)$$

where A is the first fuzzy number with membership function $\mu_A(u) \rightarrow [0, 1]$ and $u \in \mathcal{R}$ constants values that the variable can take; B is the second fuzzy number with membership function $\mu_B(v) \rightarrow [0, 1]$ and $v \in [0, 1]$ that measures the reliability of the first fuzzy number.

Algorithm 1 The association rules extraction algorithm ($AREA(X, \Lambda, MaxL, \alpha)$).

Require: The c -classes training data: X ;

The initial fuzzy numbers of features on X : $\Lambda = \{A_{jw}\}, j = 1, \dots, n, w = 1, \dots, c$;

Maximum number of features used: $MaxL$;

The penalty coefficient: α .

Ensure: The fuzzy rules of all classes: $R_1, \dots, R_c$.

```

1: for each class  $y_w$  do
2:    $y_{R_w} \leftarrow y_w$ 
3:    $v \leftarrow \min(MaxL, n)$ 
4:   for  $u = 1, 2, \dots, \min(MaxL, n)$  do
5:     Compute  $r_u$  and  $r_{u+1}$  { $r_u$  is the fuzzy rule that has maximum fuzzy confidence when considering  $u$  features
      and its consequent class is  $y_w$ .}
6:     if  $FConf(r_u) - FConf(r_{u+1}) > \alpha$  then
7:        $v \leftarrow u$ 
8:       break
9:     end if
10:  end for
11:   $R_w \leftarrow r_v$ 
12: end for

```

Algorithm 2 Fuzzy rule based decision tree algorithm ($FRDT(X, MaxL, \alpha, \delta)$).

Require: The c -classes training data: X ;

Maximum number of features used: $MaxL$;

The penalty coefficient: α ;

The threshold: δ .

Ensure: The layers of decision tree: $\Gamma_l = \{R_{l1}, \dots, R_{lc}, X_l^{imp}\}, l = 1, 2, \dots, L$ is the l -th layer of the tree.

```

1:  $l \leftarrow 0, X_0^{imp} \leftarrow X$ 
2: while  $X_l^{imp} \neq \emptyset$  do
3:    $\Lambda \leftarrow$  the initial fuzzy numbers of features for each class in  $X_l^{imp}$ 
4:    $R_{(l+1)1}, \dots, R_{(l+1)c_{l+1}} \leftarrow AREA(X_l^{imp}, \Lambda, MaxL, \alpha)$ 
5:   for Each sample  $x$  in  $X_l^{imp}$  do
6:     if  $\forall R_{(l+1)w} \in \{R_{(l+1)1}, \dots, R_{(l+1)c_{l+1}}\}, \phi_{R_{(l+1)w}}(x) < \delta$  then
7:       Add  $x$  to  $X_{l+1}^{imp}$ 
8:     end if
9:   end for
10:  if  $X_{l+1}^{imp} = \emptyset$  or  $X_{l+1}^{imp} = X_l^{imp}$  then
11:    break
12:  end if
13:   $l \leftarrow l + 1$ 
14: end while

```

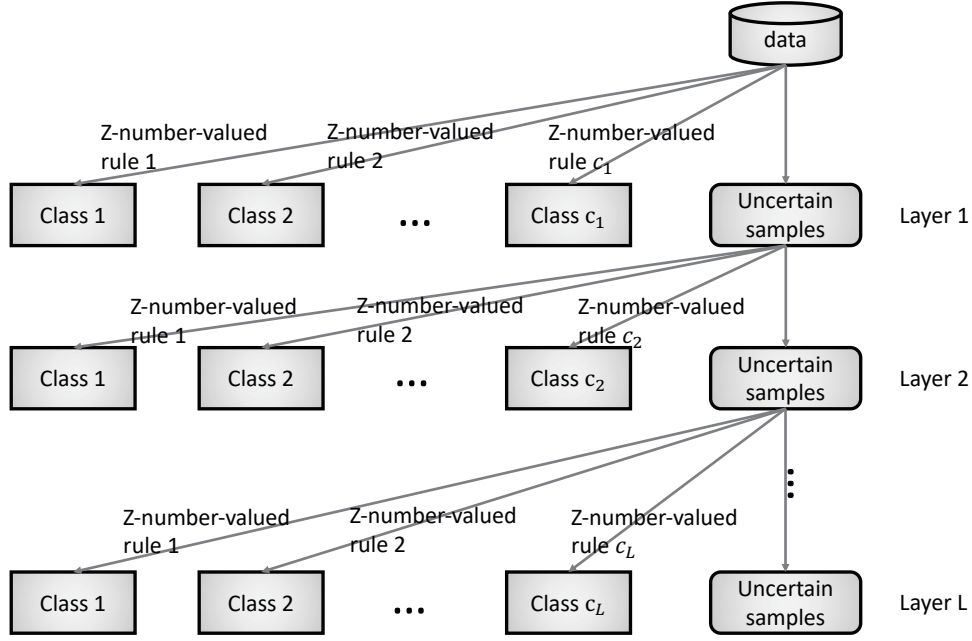


Figure 3: The overall structure of ZRDT.

3. Z-number-valued rule based decision tree (ZRDT)

The structure of the proposed ZRDT is the same as FRDT, which is an oblique tree assembled by layers. The fuzzy rules in FRDT are replaced with Z-number-valued rules. The overall structure of ZRDT is shown in Fig. 3.

3.1. Z-numbers adjustment

In the FRDT, the fuzzy numbers used in fuzzy rules are generated by mean values. Their coverage is fixed. When the distribution ranges of samples from different classes are relatively different, this method causes the fuzzy numbers of classes with smaller distribution ranges to cover more negative samples.

Example 1. In a two-class classification problem shown in Fig. 4, the mean values in this feature of class 1 and class 2 are 4.2292 and 8.7033, respectively. Then the fuzzy number generated by class 1 is $A_1 = (-\text{infi}, -\text{infi}, 4.2292, 8.7033)$ and the fuzzy number generated by class 2 is $A_2 = (4.2292, 8.7033, \text{infi}, \text{infi})$. The area between the mean values of the two classes is not equally divided into two classes. Class 1 covers a wider area than class 2, but A_2 covers a wider area than its samples. This results in two samples of class 1 being misclassified as class 2.

We use Z-numbers to solve this problem. For a Z-number, the second fuzzy number measures the reliability of the first fuzzy number. Then we can use the second fuzzy number to adjust the first fuzzy number and control its coverage.

Definition 4. For a Z-number $Z = (A, B)$, $A = (a_1, a_2, a_3, a_4)$ and $B = (b_1, b_2, b_3, b_4)$ are two trapezoidal fuzzy numbers, $a_1, a_2, a_3, a_4, b_1, b_2, b_3, b_4 \in \mathcal{R}$. After using B to adjust A the result is a fuzzy number $A' = (a'_1, a'_2, a'_3, a'_4)$, defined as Eq. (11).

$$\begin{aligned}
 a'_1 &= a_2 - (a_2 - a_1) \times \text{Rel}(B), \\
 a'_2 &= a_2, \\
 a'_3 &= a_3, \\
 a'_4 &= a_3 + (a_4 - a_3) \times \text{Rel}(B),
 \end{aligned} \tag{11}$$

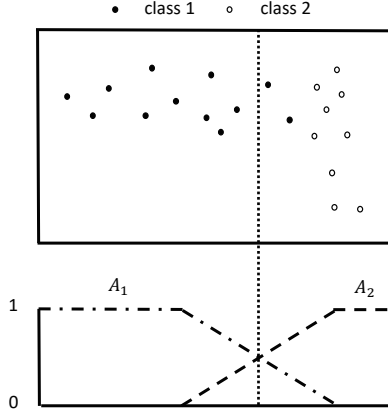


Figure 4: A two-class classification problem with fuzzy numbers.

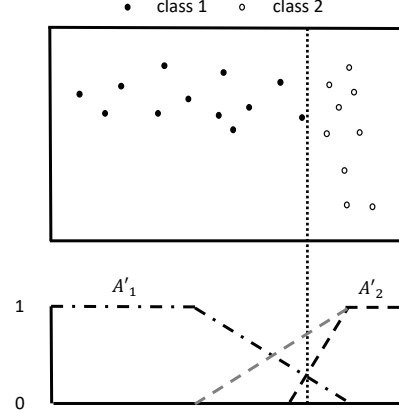


Figure 5: A two-class classification problem with adjusted fuzzy numbers.

where $Rel(B)$ is the reliability value of A , defined as Eq. (12).

$$Rel(B) = \frac{y_B^*(b_3 + b_2) + (b_4 + b_1)(1 - y_B^*)}{2},$$

$$y_B^* = \frac{(b_3 - b_2)/(b_4 - b_2) + 2}{6}. \quad (12)$$

When $b_1 = b_4$, $Rel(B) = b_1$; when $b_1 \neq b_2 = b_3 \neq b_4$, B is a triangular fuzzy number and $Rel(B) = (b_1 + b_2 + b_4)/3$.

Example 2. In the two-class classification problem in Example 1, we assume the reliability value of the fuzzy number about class 1 is 1 and the reliability value of the fuzzy number about class 2 is 0.4. After adjustment, the fuzzy numbers become $A'_1 = (-\text{infi}, -\text{infi}, 4.2292, 8.7033)$ and $A'_2 = (6.9136, 8.7033, \text{infi}, \text{infi})$ shown in Fig. 5. All samples are correctly classified.

3.2. Structure of Z-number-valued if-then rule

When the fuzzy numbers of Eq. (1) are replaced by Z-numbers, the structure of Z-number-valued if-then rule can be defined as follows.

Definition 5. For a input sample $x_i = \{x_{i1}, \dots, x_{in}\}$, the formulation of a Z-number-valued rule R_Z considered m features can be written as Eq. (13).

$$R_Z : \text{IF } x_{i\rho(1)} \text{ is } Z_{\rho(1)} \text{ and } \dots \text{ and } x_{i\rho(m)} \text{ is } Z_{\rho(m)} \text{ THEN } x_i \text{ belong to class } y_{R_Z}, \quad (13)$$

where $\rho(1), \dots, \rho(m) \in \{1, 2, \dots, n\}$ and indicate the considered features; $Z_{\rho(1)}, \dots, Z_{\rho(m)}$ are the antecedent Z-numbers in $\rho(1)$ -th, $\dots$, $\rho(m)$ -th features, respectively; $Z_{\rho(k)} = (A_{\rho(k)}, B_{\rho(k)})$, $A_{\rho(k)}$ and $B_{\rho(k)}$ are fuzzy numbers, $k = 1, 2, \dots, m$; $y_{R_Z} \in Y$ is the consequent class of the Z-number-valued rule.

Definition 6. The average membership degree of a sample x for all antecedent Z-numbers in rule R_Z can be calculated by Eq. (14).

$$\phi_{R_Z}(x) = \frac{\sum_{k=1}^m \mu_{A'_{\rho(k)}}(x_{i\rho(k)})}{m}, \quad (14)$$

where $A'_{\rho(k)}$ is the fuzzy number of $A_{\rho(k)}$ adjusted by $B_{\rho(k)}$.

3.3. Z-number-valued rule extraction algorithm

The selection of features in the FRDT is based on fuzzy confidence. A rule considered u features will only be used if its maximum fuzzy confidence value is larger than that of a fuzzy rule considered $u + 1$ features by $\alpha = 0.02$. Although the authors state that α is a certain parameter, changes in the value of this parameter have an impact on the results. The determination of the value of α , 0.02, is also unsupported. In the proposed ZRDT algorithm, the FOIL information gain [21, 22] is used to select features defined as follows.

Definition 7. The FOIL information gain is a measure of improvement of the rule in comparison with the default rule for the target class, and can be calculated by Eq. (15) [21, 22].

$$IG(R) = p_R \times (\log_2(\frac{p_R}{p_R + n_R}) - \log_2(\frac{p}{p + n})), \quad (15)$$

where p_R and n_R are the numbers of positive and negative samples covered by rule R , respectively; p and n are the numbers of positive and negative samples covered by default rule.

The more positive samples covered by the rule, the greater the value of information gain. And the value of information gain is not related to the number of features considered in this rule. Therefore, we can discard the parameter α . That is, a rule with u features can be used if its maximum information gain is larger than that of a rule with $u + 1$ features.

Before giving the complete Z-number-valued rule extraction algorithm, the second fuzzy number generation algorithm must be mentioned. For a Z-number-valued rule in the ZRDT, the first numbers of antecedent Z-numbers are calculated by Eqs. (6)-(9). The algorithm of calculating second fuzzy numbers of antecedent Z-numbers $SeFC(X, R, \delta)$ is shown as Algorithm 3.

In the algorithm $SeFC(X, R, \delta)$, N_X is the set of negative samples covered by rule R . The formula $\gamma_{\rho(k)} \leftarrow 1 + \delta - \epsilon - \mu_{A_{\rho(k)}}(x_{\rho(k)}^N)$ calculates the distance of the membership degree of $A_{\rho(k)}$ about the negative sample with respect to $\rho(k)$ -th feature from the threshold. The closer the value of $\mu_{A_{\rho(k)}}(x_{\rho(k)}^N)$ is to δ , the closer the value of $\gamma_{\rho(k)}$ is to $1 - \epsilon$, the smaller the adjustment needed for the fuzzy number $A_{\rho(k)}$.

Example 3. In a two-class classification problem shown in Fig. 6, the circled negative sample is only a bit over the threshold. A_1 only needs to be adjusted τ_1 to A'_1 in order not to cover this sample. In contrast, the circled negative sample in Fig. 7 is much over the threshold, A_2 should be adjusted τ_2 to A'_2 in order to exclude the sample. Thus, we believe that the closer the negative samples are to the threshold, the higher the reliability of the fuzzy number.

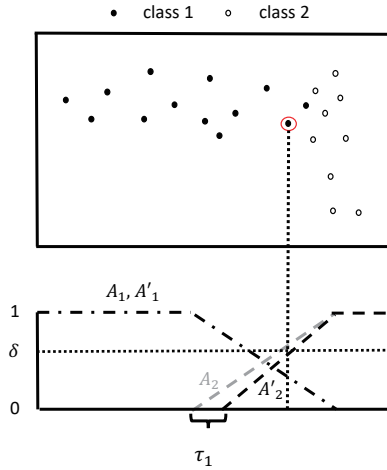


Figure 6: A two-class classification problem with fuzzy numbers.

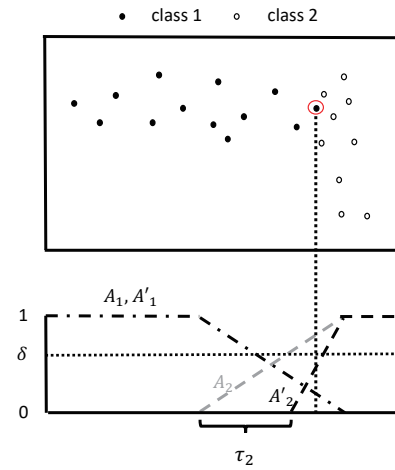


Figure 7: A two-class classification problem with adjusted fuzzy numbers.

Algorithm 3 The second fuzzy number calculation algorithm ($SeFC(X, R, \delta)$).

Require: The c -classes training data: X ;

The fuzzy rule: R , with antecedent fuzzy numbers: $A_{\rho(k)}$, $k = 1, \dots, m$;

The threshold: δ .

Ensure: The Z-number-valued rule: R_Z , with antecedent Z-numbers: $Z_{\rho(k)} = (A_{\rho(k)}, B_{\rho(k)})$, $k = 1, \dots, m$.

```

1:  $N_X \leftarrow \emptyset$ 
2: for Each sample  $x$  in  $X$  do
3:   if  $\phi_R(x) \geq \delta$  and  $y_x \neq y_R$  then
4:     Add  $x$  to  $N_X$ 
5:   end if
6: end for
7: for Each antecedent fuzzy number  $A_{\rho(k)}$  in  $R$  do
8:    $\min_{\rho(k)} \leftarrow 1$ ,  $mean_{\rho(k)} \leftarrow 0$ ,  $\max_{\rho(k)} \leftarrow 0$ 
9:   for Each sample  $x^N$  in  $N_X$  do
10:    if  $\mu_{A_{\rho(k)}}(x^N_{\rho(k)}) > \delta$  then
11:       $\gamma_{\rho(k)} \leftarrow 1 + \delta - \epsilon - \mu_{A_{\rho(k)}}(x^N_{\rho(k)})$  { $\epsilon$  is the infinitesimals.}
12:       $mean_{\rho(k)} \leftarrow mean_{\rho(k)} + \gamma_{\rho(k)} / |N_X|$ 
13:      if  $\gamma_{\rho(k)} > \max_{\rho(k)}$  then
14:         $\max_{\rho(k)} \leftarrow \gamma_{\rho(k)}$ 
15:      end if
16:      if  $\gamma_{\rho(k)} < \min_{\rho(k)}$  then
17:         $\min_{\rho(k)} \leftarrow \gamma_{\rho(k)}$ 
18:      end if
19:    end if
20:  end for
21:   $B_{\rho(k)} = (\min_{\rho(k)}, mean_{\rho(k)}, mean_{\rho(k)}, \max_{\rho(k)})$ 
22: end for
23: Compute the information gain of  $R$  and  $R_Z$ :  $IG(R)$ ,  $IG(R_Z)$ .
24: if  $IG(R) \geq IG(R_Z)$  then
25:    $B_{\rho(k)} = (1, 1, 1, 1)$  for  $k = 1, \dots, m$ 
26:   Return  $R_Z$ .
27: else
28:   Return  $R_Z$ .
29: end if

```

In order to build the pure leaf node of the tree, we design the Z-number-valued rule extraction algorithm $ZREA(X, \Lambda, MaxL, \delta)$ based on information gain, which is shown as Algorithm 4. Where r_{Zu} is the Z-number-valued rule that has maximum information gain when considering u features. If $u = 1$, the Z-number-valued rule has only one antecedent Z-number. We will generate the Z-numbers corresponding to all features, and then choose the one with the highest information gain as the antecedent Z-number of this rule. If $u = 2$, we will choose another feature on the basis of r_{Z1} so that the information gain is maximized. Use this way until the increase in features does not contribute to the increase in information gain.

Algorithm 4 The Z-number-valued rules extraction algorithm ($ZREA(X, \Lambda, MaxL, \delta)$).

Require: The c -classes training data: X ;

The initial fuzzy numbers of features on X : $\Lambda = \{A_{jw}\}, j = 1, \dots, n, w = 1, \dots, c$;

Maximum number of features used: $MaxL$;

The threshold: δ .

Ensure: The Z-number-valued rules of all classes: $R_{Z1}, \dots, R_{Zc}$.

```

1: for each class  $y_w$  do
2:    $y_{R_{Zw}} \leftarrow y_w$ 
3:    $v \leftarrow \min(MaxL, n)$ 
4:   for  $u = 1, 2, \dots, \min(MaxL, n)$  do
5:     Compute  $r_{Zu}$  and  $r_{Zu+1}$  { $r_{Zu}$  is the Z-number-valued rule that has maximum information gain when considering  $u$  features and its consequent class is  $y_w$ .}
6:     if  $IG(r_{Zu}) > IG(r_{Zu+1})$  then
7:        $v \leftarrow u$ 
8:       break
9:     end if
10:  end for
11:   $R_{Zw} \leftarrow r_{Zv}$ 
12: end for

```

3.4. The overall algorithm of ZRDT

The additional impure node of each layer in ZRDT is generated in the same way as FRDT, relying on a threshold δ to collect the samples whose average membership degree for all antecedent Z-numbers in rule R_Z is below the threshold. The pseudocode of ZRDT $ZRDT(X, MaxL, \delta)$ is shown in Algorithm 5. The mission of each layer of ZRDT is to classify the uncertain samples of the previous layer, where new fuzzy numbers will be generated based on the uncertain samples, and then new Z-number-valued rules will be generated. The initial uncertain samples set is the entire training data set. Afterward, the structure of the ZRDT in Fig. 3 can be visualized in Fig. 8. Where $R_{Zl,w}$, $l = 1, \dots, L$, $w = 1, \dots, c_l$, is the Z-number-valued rule in the l -th layer whose consequent class is $y_{l,w}$. $X_{l,w}$ is a set of samples that are classified as class $y_{l,w}$ by rule $R_{Zl,w}$.

3.5. New samples classification

After training we can obtain a model consisting of Z-number-valued rules as Fig. 9. For a new input sample $x = \{x_1, \dots, x_n\}$, $R_{Zl,w}$, $l = 1, \dots, L$, $w = 1, \dots, c_l$, is the Z-number-valued rule in the l -th layer with consequent class $y_{l,w}$:

$$R_{Zl,w} : \text{IF } x_{\rho_{l,w}(1)} \text{ is } Z_{l,w\rho_{l,w}(1)} \text{ and } \dots \text{ and } x_{\rho_{l,w}(m_{l,w})} \text{ is } Z_{l,w\rho_{l,w}(m_{l,w})} \text{ THEN } x \text{ belong to class } y_{l,w}, \quad (16)$$

where $\rho_{l,w}(1), \dots, \rho_{l,w}(m_{l,w}) \in \{1, \dots, n\}$ and indicate the considered features in rule $R_{Zl,w}$; c_l is the number of classes of layer l ; $m_{l,w}$ is the number of considered features in rule $R_{Zl,w}$; $y_{l,w} \in \{y_1, \dots, y_c\}$. The new sample can be classified by rule $R_{Zl,w}^*$, if $\phi_{R_{Zl,w}^*}(x) \geq \delta$ and

$$\phi_{R_Z^*}(x) = \max(\phi_{R_{Zl,1}}(x), \phi_{R_{Zl,2}}(x), \dots, \phi_{R_{Zl,c_l}}(x)), \quad (17)$$

The average membership degree for this new sample will be calculated under each rule starting from the first layer, $l = 1$. If the maximum average membership degree of rules in this layer is over the set threshold, the new sample will

Algorithm 5 Z-number-valued rule based decision tree algorithm ($ZRDT(X, MaxL, \delta)$).**Require:** The c -classes training data: X ;Maximum number of features used: $MaxL$;The threshold: δ .**Ensure:** The layers of decision tree: $\Gamma_l = \{R_{Zl,1}, \dots, R_{Zl,c_l}, X_l^{imp}\}$, $l = 1, 2, \dots, L$ is the l -th layer of the tree.

```

1:  $l \leftarrow 0, X_0^{imp} \leftarrow X$ 
2: while  $X_l^{imp} \neq \emptyset$  do
3:    $\Lambda \leftarrow$  the fuzzy numbers of feature for each class in  $X_l^{imp}$ 
4:    $R_{Z(l+1),1}, \dots, R_{Z(l+1),c_{l+1}} \leftarrow ZREA(X_l^{imp}, \Lambda, MaxL, \delta)$ 
5:   for Each sample  $x$  in  $X_l^{imp}$  do
6:     if  $\forall R_{Z(l+1),w} \in \{R_{Z(l+1),1}, \dots, R_{Z(l+1),c_{l+1}}\}, \phi_{R_{Z(l+1),w}}(x) < \delta$  then
7:       Add  $x$  to  $X_{l+1}^{imp}$ 
8:     end if
9:   end for
10:  if  $X_{l+1}^{imp} = \emptyset$  or  $X_{l+1}^{imp} = X_l^{imp}$  then
11:    break
12:  end if
13:   $l \leftarrow l + 1$ 
14: end while

```

be classified as the class that is the consequent class of the rule which has the maximum average membership degree. If the maximum average membership degree of rules in this layer is below the set threshold, the new sample will be sent to the next layer, $l = l + 1$.

4. Experimental studies

In this section, some experiments are used to evaluate the learning ability, predictive ability and the structural advantage of the proposed ZRDT. The FRDT is an effective fuzzy rule-based model for supervised classification problems, but it needs to preset two parameters: α and δ . The proposed ZRDT uses information gain to reduce the parameter α . The experiments focus on three aspects: the improvement of the proposed ZRDT in terms of learning and prediction ability and the size of the decision tree.

4.1. Experimental setup

Nineteen well-known benchmark datasets from the UCI repository are used to evaluate the performance of the ZRDT. These datasets differ in the number of classes, features, and samples. The details of these datasets are described in Table 2, where 'Samples' is the number of samples, '#Features' is the number of features and '#Classes' is the number of classes.

In order to evaluate generalization ability, 10-fold cross-validation (10CV) is used. Samples are divided into 10 folds, each containing 10% samples. The model is trained with nine folds and is tested with another fold. The average accuracy of 10 folds can provide insight into the model's performance. There are five algorithms used to compare their learning ability, predictive ability, and size of produced decision tree (all algorithms are implemented on software KEEL [2]):

- (1) FRDT (0.01): The FRDT with parameters $\alpha = 0.01$ and $\delta = 0.6$. The value of α is smaller than that recommended in [31].
- (2) FRDT (0.02): The FRDT with parameters $\alpha = 0.02$ and $\delta = 0.6$. The value of α is recommended in [31].
- (3) FRDT (0.03): The FRDT with parameters $\alpha = 0.03$ and $\delta = 0.6$. The value of α is larger than that recommended in [31].

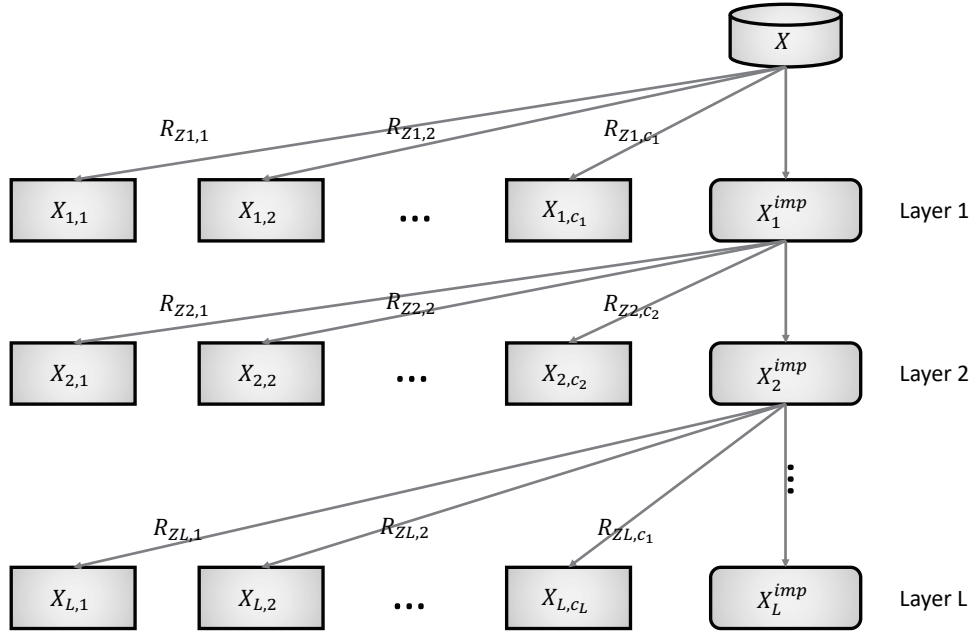


Figure 8: The specific Z-number-valued rules classification tree.

- (4) IGFRDT: The FRDT with information gain features selection and parameter $\delta = 0.6$. The information gain can find the most suitable features.
- (5) ZRDT: The proposed ZRDT with $\delta = 0.6$. The Z-numbers make the model fit training data better and the information gain can find the most suitable features.

4.2. Performance comparison

The classification accuracy rates of five algorithms over training data and testing data are listed in Table 3 and Table 4, respectively. The results in bold show that the current method is the most superior. In most datasets, the ZRDT outperforms other algorithms except for Haberman, Ionosphere, Iris, and occupancy datasets. To analyze the differences between the classifiers more statistically, the Holm test is used to compare a classifier with the superior performance classifier. The performance of two classifiers is significantly different if the hypothesis is rejected.

```

IF  $x_{\rho_{1,1}}(1)$  is  $(Z_{1,1\rho_{1,1}}(1))$  and ... and  $x_{\rho_{1,1}}(m_{1,1})$  is  $(Z_{1,1\rho_{1,1}}(m_{1,1}))$  THEN  $x$  belongs to class  $y_{1,1}$ 
ELSE IF  $x_{\rho_{1,2}}(1)$  is  $(Z_{1,2\rho_{1,2}}(1))$  and ... and  $x_{\rho_{1,2}}(m_{1,2})$  is  $(Z_{1,2\rho_{1,2}}(m_{1,2}))$  THEN  $x$  belongs to class  $y_{1,2}$ 
...
ELSE IF  $x_{\rho_{1,c_1}}(1)$  is  $(Z_{1,c_1\rho_{1,c_1}}(1))$  and ... and  $x_{\rho_{1,c_1}}(m_{1,c_1})$  is  $(Z_{1,c_1\rho_{1,c_1}}(m_{1,c_1}))$  THEN  $x$  belongs to class  $y_{1,c_1}$ 
ELSE
  IF  $x_{\rho_{2,1}}(1)$  is  $(Z_{2,1\rho_{2,1}}(1))$  and ... and  $x_{\rho_{2,1}}(m_{2,1})$  is  $(Z_{2,1\rho_{2,1}}(m_{2,1}))$  THEN  $x$  belongs to class  $y_{2,1}$ 
  ELSE IF  $x_{\rho_{2,2}}(1)$  is  $(Z_{2,2\rho_{2,2}}(1))$  and ... and  $x_{\rho_{2,2}}(m_{2,2})$  is  $(Z_{2,2\rho_{2,2}}(m_{2,2}))$  THEN  $x$  belongs to class  $y_{2,2}$ 
  ...
  ELSE IF  $x_{\rho_{2,c_2}}(1)$  is  $(Z_{2,c_2\rho_{2,c_2}}(1))$  and ... and  $x_{\rho_{2,c_2}}(m_{2,c_2})$  is  $(Z_{2,c_2\rho_{2,c_2}}(m_{2,c_2}))$  THEN  $x$  belongs to class  $y_{2,c_2}$ 
  ELSE
    ...
    ELSE
      IF  $x_{\rho_{L,1}}(1)$  is  $(Z_{L,1\rho_{L,1}}(1))$  and ... and  $x_{\rho_{L,1}}(m_{L,1})$  is  $(Z_{L,1\rho_{L,1}}(m_{L,1}))$  THEN  $x$  belongs to class  $y_{L,1}$ 
      ELSE IF  $x_{\rho_{L,2}}(1)$  is  $(Z_{L,2\rho_{L,2}}(1))$  and ... and  $x_{\rho_{L,2}}(m_{L,2})$  is  $(Z_{L,2\rho_{L,2}}(m_{L,2}))$  THEN  $x$  belongs to class  $y_{L,2}$ 
      ...
      ELSE IF  $x_{\rho_{L,c_L}}(1)$  is  $(Z_{L,c_L\rho_{L,c_L}}(1))$  and ... and  $x_{\rho_{L,c_L}}(m_{L,c_L})$  is  $(Z_{L,c_L\rho_{L,c_L}}(m_{L,c_L}))$  THEN  $x$  belongs to class  $y_{L,c_L}$ 

```

Figure 9: The obtained Z-number-valued rules from ZRDT.

Dataset	#Samples	#Features	#Classes
Banknote	1372	4	2
Bupa	345	6	2
Glass	214	9	6
Haberman	306	3	2
Heart-statlog	270	13	2
Ionosphere	351	34	2
Iris	150	4	3
Mfeat-fourier	2000	76	10
Mfeat-karhunen	2000	64	10
Mfeat-zernike	2000	47	1
Monk-2	432	6	2
New-thyroid	215	5	3
Occupancy	2665	5	2
Optdigit	5620	64	2
Pageblocks	5473	10	5
Pendigits	10992	16	10
Vehicle	846	18	4
Wine	178	12	3
Wisconsin	699	9	2

Table 2: The description of used datasets.

The Holm test is a nonparametric test which is based on the relative performance of classifiers in terms of their ranks: The better the performance, the higher the rank (in case of ties, average ranks are assigned); see Table 5. Let K be the number of classifiers and N be the number of datasets. Let r_k^n be the rank of classifier k on dataset n , $k = 1, 2, \dots, K$ and $n = 1, 2, \dots, N$. The test z-value for compared classifier k is calculated by:

$$z = \frac{aveR_k - \min(aveR_1, aveR_2, \dots, aveR_K)}{\sqrt{K(K+1)/(6N)}}, \quad (18)$$

where $aveR_k = \frac{\sum_{n=1}^N r_k^n}{N}$ is the average rank of classifier j ; $\min(aveR_1, aveR_2, \dots, aveR_K)$ is the average rank of the highest performance classifier; $k \neq \operatorname{argmin}_i(aveR_i)$. The z-values are used to find the corresponding probability values (p-values) of all classifiers: $p_1, p_2, \dots, p_{K-1}$. Then we store the p-values in order lowest-to-highest: $p_{\rho(1)}, p_{\rho(2)}, \dots, p_{\rho(K-1)}$, where $\rho(k)$ is the index function satisfying $p_{\rho(k-1)} \leq p_{\rho(k)} \leq p_{\rho(k+1)}$. If $p_{\rho(k)} < \frac{\alpha}{K-k}$, reject the null hypothesis that the performance of classifier $\rho(k)$ and the best performance classifier are not significantly different and continue to the next step, otherwise end.

In this study, the average ranks of classifiers are: $aveR_1 = 3.7895$, $aveR_2 = 3.6316$, $aveR_3 = 3.9727$, $aveR_4 = 2.3421$ and $aveR_5 = 1.2631$. The number of classifiers is $K = 5$ and the number of datasets is $N = 19$. The certain pre-specified significance level is $\alpha = 0.05$. In the Table 6, we list the z-values and p-values of each classifier. We can see that the Holm procedure rejects all hypotheses. We can conclude that the ZRDT performs significantly better than other algorithms at the significance level 0.05.

4.3. Decision tree size comparison

Since the number of considered features is different for different rules, the number of nodes cannot represent the size of the decision tree in detail. In this study, we summarize the number of all considered features of a decision tree. For example, a small decision tree has two rules: the first one includes two features and the second one includes three features. We think the size of this decision tree is five. We summarized the average sizes of decision trees of 10 folds in previous experiments in Table 7. We can see that the sizes of the decision trees generated by IGFRDT are consistently smaller than those generated by other algorithms. However, the decision trees generated by ZRDT are only a bit larger than the decision trees generated by IGFRDT.

Dataset	FRDT (0.01)	FRDT (0.02)	FRDT (0.03)	IGFRDT	ZRDT
Banknote	87.77%	87.81%	87.77%	%88.56%	90.44%
Bupa	66.12%	68.79%	68.76%	67.71%	72.27%
Glass	62.37%	60.61%	51.35%	63.92%	67.05%
Haberman	74.36%	74.00%	73.35%	74.40%	74.36%
Heart-statlog	85.06%	87.45%	87.28%	88.27%	88.35%
Ionosphere	89.27%	89.59%	88.51%	81.36%	88.70%
Iris	98.52%	97.48%	97.48%	97.93%	97.93%
Mfeat-fourier	53.88%	27.78%	17.62%	63.94%	65.49%
Mfeat-karhunen	56.50%	30.35%	17.41%	64.83%	66.92%
Mfeat-zernike	95.42%	96.37%	96.94%	97.46%	98.29%
Monk-2	88.89%	88.89%	89.75%	97.22%	97.22%
New-thyroid	94.99%	95.45%	95.55%	96.07%	96.28%
Occupancy	97.94%	97.94%	97.94%	97.93%	97.93%
Optdigit	56.18%	65.96%	65.96%	91.22%	95.48%
Pageblocks	76.10%	82.19%	79.60%	92.92%	95.22%
Pendigits	52.83%	48.97%	31.11%	59.39%	62.31%
Vehicle	50.96%	50.79%	48.86%	59.43%	60.52%
Wine	96.38%	96.82%	96.19%	98.38%	98.63%
Wisconsin	94.67%	95.14%	95.29%	96.68%	97.60%

Table 3: Classification accuracy rate (%) of five algorithms over training data.

Dataset	FRDT (0.01)	FRDT (0.02)	FRDT (0.03)	IGFRDT	ZRDT
Banknote	87.76%	87.98%	87.76%	88.48%	90.31%
Bupa	62.25%	64.25%	64.29%	63.93%	65.51%
Glass	53.66%	49.17%	45.46%	55.51%	58.28%
Haberman	73.47%	73.15%	70.84%	73.47%	73.81%
Heart-statlog	74.07%	80.00%	82.59%	81.48%	81.85%
Ionosphere	86.62%	86.04%	86.06%	81.19%	88.89%
Iris	97.33%	94.67%	94.67%	95.33%	95.33%
Mfeat-fourier	50.05%	27.80%	17.35%	59.30%	60.70%
Mfeat-karhunen	52.15%	29.25%	17.00%	61.40%	62.90%
Mfeat-zernike	95.00%	96.15%	96.10%	97.10%	97.70%
Monk-2	88.96%	88.96%	89.49%	97.27%	97.27%
New-thyroid	90.71%	93.53%	93.51%	93.94%	94.42%
Occupancy	97.86%	97.86%	97.86%	97.90%	97.90%
Optdigit	56.03%	66.14%	66.14%	90.94%	95.16%
Pageblocks	75.64%	81.65%	79.33%	92.95%	94.68%
Pendigits	52.37%	48.16%	31.04%	59.22%	61.97%
Vehicle	49.19%	49.18%	48.00%	57.44%	57.94%
Wine	94.38%	96.01%	91.60%	95.49%	95.49%
Wisconsin	93.99%	93.99%	94.27%	95.28%	95.56%

Table 4: Classification accuracy rate (%) of five algorithms over testing data.

Dataset	FRDT (0.01)	FRDT (0.02)	FRDT (0.03)	IGFRDT	ZRDT
Banknote	4.5	3	4.5	2	1
Bupa	5	3	2	4	1
Glass	3	4	5	2	1
Haberman	2.5	4	5	2.5	1
Heart-statlog	5	4	1	3	2
Ionosphere	2	4	3	5	1
Iris	1	4.5	4.5	2.5	2.5
Mfeat-fourier	3	4	5	2	1
Mfeat-karhunen	3	4	5	2	1
Mfeat-zernike	5	3	4	2	1
Monk-2	4.5	4.5	3	1.5	1.5
New-thyroid	5	3	4	2	1
Occupancy	4	4	4	1.5	1.5
Optdigit	5	3.5	3.5	2	1
Pageblocks	5	3	4	2	1
Pendigits	3	4	5	2	1
Vehicle	3	4	5	2	1
Wine	4	1	5	2.5	2.5
Wisconsin	4.5	4.5	3	2	1

Table 5: Ranks of classifiers.

k	Classifier	z-value	p-value	critical value $\alpha/(K - k)$	Hypothesis
1	FRDT (0.03)	5.2838	1.2653e-07	0.0125	Rejected
2	FRDT (0.01)	4.9247	8.4491e-07	0.0167	Rejected
3	FRDT (0.02)	4.6169	3.8951e-06	0.025	Rejected
4	IGFRDT	2.1032	0.03544	0.05	Rejected

Table 6: The Holm test on performance.

Dataset	FRDT (0.01)	FRDT (0.02)	FRDT (0.03)	IGFRDT	ZRDT
Banknote	7.1	9.6	11.1	7.9	8.7
Bupa	21.1	30.5	42.3	15.3	19.8
Glass	82.6	126.4	120.8	25.6	27.3
Haberman	5.2	6.9	13.5	4.7	4.7
Heart-statlog	23.6	38.0	47.8	16.1	17.3
Ionosphere	27.8	37.6	36.2	3.0	5.8
Iris	11.4	9.7	11.5	6.5	6.5
Mfeat-fourier	296.0	327.6	201.0	73.8	80.9
Mfeat-karhunen	246.0	388.4	325.7	81.4	87.3
Mfeat-zernike	32.7	34.7	37.9	10.3	11.7
Monk-2	10.0	10.0	9.5	5.0	5.0
New-thyroid	18.4	28.9	32.8	7.6	8.7
Occupancy	11.0	11.0	11.0	3.6	3.6
Optdigit	64.6	110.4	110.4	22.7	31.6
Pageblocks	93.6	132.8	149.7	25.6	38.3
Pendigits	335.0	643.8	441.9	71.3	74.4
Vehicle	79.1	112.7	107.3	21.3	27.8
Wine	16.7	24.5	48.9	12.4	12.5
Wisconsin	32.8	36.4	45.2	16.0	22.0

Table 7: The average size of decision trees.

k	Classifier	z-value	p-value	critical value $\alpha/(K - k)$	Hypothesis
1	FRDT (0.03)	6.6175	3.6517e-11	0.0125	Rejected
2	FRDT (0.02)	6.0533	1.4193e-09	0.0167	Rejected
3	FRDT (0.01)	3.7448	0.00018	0.025	Rejected
4	ZRDT	1.5389	0.1238	0.05	Accepted

Table 8: The Holm test on size.

We also use the Holm test to analyze the significance differences between sizes of these decision trees. Firstly, we determine the rank of each decision tree on each dataset: The smaller the size, the higher the rank. We can obtain the average rank of each decision tree: $aveR_1 = 3.0789$, $aveR_2 = 4.2631$, $aveR_3 = 4.5526$, $aveR_4 = 1.1579$ and $aveR_5 = 1.9473$. The z-values and p-values of decision trees are listed in Table. The Holm test accepted the last hypothesis which is there are no differences between the sizes of decision trees generated by ZRDT and IGFRDT at the significance level 0.05. Therefore, the ZRDT has the highest classification performance while the size of the decision tree is relatively small.

5. Conclusions

The concept of Z-number is a more adequate mathematical form for describing both uncertain and partially reliable real-world information. In this study, we extend the fuzzy rule-based decision tree in a Z-number-valued framework to propose the Z-number-valued rule-based decision tree called ZRDT. We pioneer the use of negative samples to generate the second fuzzy number of Z-numbers. The Z-numbers can adjust the created decision tree to fit the training data better. We also use information gain to replace the fuzzy confidence to select features for each rule. Thus, not only can ZRDT select the most appropriate features for each rule, but it also reduces one parameter in FRDT. Some experiments on 19 datasets show that the ZRDT significantly outperforms the original FRDT and the improved FRDT in learning ability, predictive ability. For the size of the produced decision tree, the decision tree generated by FRDT is the smallest. However, after statistical test, there is no significant size difference between the decision trees generated by ZRDT and FRDT. Therefore, the ZRDT has the highest classification performance while the size of the decision tree is relatively small.

Acknowledgements

This work was supported by the project B-TIC-590-UGR20 co-funded by the Programa Operativo FEDER 2014-2020 and the Regional Ministry of Economy, Knowledge, Enterprise and Universities (CECEU) of Andalusia, by the China Scholarship Council (CSC), by the project PID2019-103880RB-I00 funded by MCIN / AEI / 10.13039 / 501100011033 and by the Andalusian government through project P20_00673.

References

- [1] Aghaei, S., Azizi, M.J., Vayanos, P., 2019. Learning optimal and fair decision trees for non-discriminative decision-making, in: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 1418–1426.
- [2] Alcalá-Fdez, J., Sanchez, L., Garcia, S., del Jesus, M.J., Ventura, S., Garrell, J.M., Otero, J., Romero, C., Bacardit, J., Rivas, V.M., et al., 2009. Keel: a software tool to assess evolutionary algorithms for data mining problems. *Soft Computing* 13, 307–318.
- [3] Aliev, R.A., Guirimov, B.G., Huseynov, O.H., Aliyev, R.R., 2021. Z-relation equation-based decision making. *Expert Systems with Applications* 184, 115387.
- [4] Bhargava, N., Sharma, G., Bhargava, R., Mathuria, M., 2013. Decision tree analysis on j48 algorithm for data mining. *Proceedings of international journal of advanced research in computer science and software engineering* 3.
- [5] Brodley, C.E., Utgoff, P.E., 1995. Multivariate decision trees. *Machine learning* 19, 45–77.
- [6] Chai, J., Su, Y., Lu, S., 2022. Linguistic z-number preference relation for group decision making and its application in digital transformation assessment of smes. *Expert Systems with Applications* , 118749.
- [7] Charbuty, B., Abdulazeez, A., 2021. Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends* 2, 20–28.
- [8] Damanik, I.S., Windarto, A.P., Wanto, A., Andani, S.R., Saputra, W., et al., 2019. Decision tree optimization in c4. 5 algorithm using genetic algorithm, in: *Journal of Physics: Conference Series*, IOP Publishing. p. 012012.

- [9] Isazadeh, A., Mahan, F., Pedrycz, W., 2016. Mflexdt: multi flexible fuzzy decision tree for data stream classification. *Soft computing* 20, 3719–3733.
- [10] Jin, C., De-Lin, L., Fen-Xiang, M., 2009. An improved id3 decision tree algorithm, in: 2009 4th international conference on computer science & Education, IEEE. pp. 127–130.
- [11] Jiskani, I.M., Yasli, F., Hosseini, S., Rehman, A.U., Uddin, S., 2022. Improved z-number based fuzzy fault tree approach to analyze health and safety risks in surface mines. *Resources Policy* 76, 102591.
- [12] Kang, B., Chhipi-Shrestha, G., Deng, Y., Hewage, K., Sadiq, R., 2018. Stable strategies analysis based on the utility of z-number in the evolutionary games. *Applied Mathematics and Computation* 324, 202–217.
- [13] Kavitha, R., Kannan, E., 2016. An efficient framework for heart disease classification using feature extraction and feature selection technique in data mining, in: 2016 international conference on emerging trends in engineering, technology and science (icetets), IEEE. pp. 1–5.
- [14] Lee, C.S., Cheang, P.Y.S., Moslehpour, M., 2022. Predictive analytics in business analytics: decision tree. *Advances in Decision Sciences* 26, 1–29.
- [15] Lim, T.S., Loh, W.Y., Shih, Y.S., 2000. A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. *Machine learning* 40, 203–228.
- [16] Liu, X., Pedrycz, W., 2007. The development of fuzzy decision trees in the framework of axiomatic fuzzy set logic. *Applied Soft Computing* 7, 325–342.
- [17] Mahan, F., Mohammadzad, M., Rozekhani, S.M., Pedrycz, W., 2021. Chi-mflexdt: Chi-square-based multi flexible fuzzy decision tree for data stream classification. *Applied Soft Computing* 105, 107301.
- [18] Mathan, K., Kumar, P.M., Panchatcharam, P., Manogaran, G., Varadharajan, R., 2018. A novel gini index decision tree data mining method with neural network classifiers for prediction of heart disease. *Design automation for embedded systems* 22, 225–242.
- [19] Olaru, C., Wehenkel, L., 2003. A complete fuzzy decision tree technique. *Fuzzy sets and systems* 138, 221–254.
- [20] Peng, H.g., Wang, X.k., Zhang, H.Y., Wang, J.q., 2021. Group decision-making based on the aggregation of z-numbers with archimedean t-norms and t-conorms. *Information Sciences* 569, 264–286.
- [21] Quinlan, J.R., 1990. Learning logical definitions from relations. *Machine learning* 5, 239–266.
- [22] Quinlan, J.R., Cameron-Jones, R.M., 1993. Foil: A midterm report, in: *European conference on machine learning*, Springer. pp. 1–20.
- [23] Rokach, L., Maimon, O., 2005. Top-down induction of decision trees classifiers-a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 35, 476–487.
- [24] Sanjay, K., Danti, A., 2019. Detection of fake opinions on online products using decision tree and information gain, in: 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), IEEE. pp. 372–375.
- [25] Singh, S., Gupta, P., 2014. Comparative study id3, cart and c4. 5 decision tree algorithm: a survey. *International Journal of Advanced Information Science and Technology (IJAIST)* 27, 97–103.
- [26] Song, X., Qin, B., Xiao, F., 2021. Fr-kde: a hybrid fuzzy rule-based information fusion method with its application in biomedical classification. *International Journal of Fuzzy Systems* 23, 392–404.
- [27] Tangirala, S., 2020. Evaluating the impact of gini index and information gain on classification using decision tree classifier algorithm. *International Journal of Advanced Computer Science and Applications* 11, 612–619.
- [28] Tian, Z.p., Nie, R.x., Wang, J.q., Luo, H., Li, L., 2020. A prospect theory-based qualiflex for uncertain linguistic z-number multi-criteria decision-making with unknown weight information. *Journal of Intelligent & Fuzzy Systems* 38, 1775–1787.
- [29] Tso, G.K., Yau, K.K., 2007. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy* 32, 1761–1768.
- [30] Wang, X., Chen, B., Qian, G., Ye, F., 2000. On the optimization of fuzzy decision trees. *Fuzzy Sets and Systems* 112, 117–125.
- [31] Wang, X., Liu, X., Pedrycz, W., Zhang, L., 2015. Fuzzy rule based decision trees. *Pattern Recognition* 48, 50–59.
- [32] Yang, F.J., 2019. An extended idea about decision trees, in: 2019 International Conference on Computational Science and Computational Intelligence (CSCI), IEEE. pp. 349–354.
- [33] Yousefi, S., Valipour, M., Gul, M., 2021. Systems failure analysis using z-number theory-based combined compromise solution and full consistency method. *Applied Soft Computing* 113, 107902.
- [34] Yuan, Y., Shaw, M.J., 1995. Induction of fuzzy decision trees. *Fuzzy Sets and systems* 69, 125–139.
- [35] Zadeh, L.A., 2011. A note on z-numbers. *Information Sciences* 181, 2923–2932.
- [36] Zhou, H., Zhang, J., Zhou, Y., Guo, X., Ma, Y., 2021. A feature selection algorithm of decision tree based on feature weight. *Expert Systems with Applications* 164, 113842.

4 Z-number generation model and its application in a rule-based classification system

- Y. Li, J. A. Morente-Molinera, J. R. Trillo, and E. Herrera-Viedma, Z-number generation model and its application in a rule-based classification system, IEEE Transactions on Cybernetics, 2024, DOI: 10.1109/TCYB.2025.3545195.
 - Status: **Published**.
 - Impact Factor (JCR 2023): 9.4
 - Subject Category: Automation & Control Systems, Ranking 4/84 (Q1).
 - Subject Category: Computer Science, Artificial Intelligence, Ranking 16/197 (Q1).
 - Subject Category: Computer Science, Cybernetics, Ranking 2/32 (Q1).

Z-Number Generation Model and its Application in a Rule-Based Classification System

Yangxue Li, Juan Antonio Morente-Molinera, Jose Ramón Trillo, Enrique Herrera-Viedma *Fellow, IEEE*

Abstract—Due to their unique form and powerful capability in handling uncertainty and partial reliability of information, Z-numbers have achieved success in various fields. Zadeh previously stated that a Z-number can be regarded as a summary of probability distributions. Researchers have proposed various methods for determining the underlying probability distributions from a given Z-number. Conversely, can a Z-number be used to summarize a set of probability distributions? This problem is currently unexplored. In this paper, we propose a nonlinear model called Maximum Expected Minimum Entropy (MEME) for generating a Z-number from a set of probability distributions. Through this model, Z-numbers can be generated from data without requiring expert knowledge. Additionally, we applied the MEME model to classification problems, introducing a new if-then rule form called Z-valuation if-then rules. It replaces the deterministic consequent part of a fuzzy rule with an uncertain Z-valuation, further summarizing uncertain information in the consequent part of a rule. Based on the Z-valuation rules, we propose a Z-valuation rule-based (ZVRB) classification system. This system aims to enhance decision-making processes in situations where uncertainty is a key factor. To validate the effectiveness of the ZVRB classification system, we conducted two experiments comparing it with well-known and state-of-the-art classic classifiers, as well as rule-based classification systems. The results show that the ZVRB model is significantly superior to other classifiers and rule-based classification systems.

Index Terms—Z-number, fuzzy rule-based system, pattern classification, probability distribution, optimization.

I. INTRODUCTION

Due to factors such as incomplete information, measurement errors, and cognitive biases, uncertainty is an inherent aspect of many real-world problems, challenging our ability to make precise and confident decisions. Handling uncertainty is a crucial task in many disciplines and fields. Various theories and methods are employed for this purpose, such as probability theory [1], fuzzy set theory [2], [3], Dempster-Shafer evidence theory [4], [5], random permutation set [6] and Z-numbers [7]. In which, Z-numbers can handle not only uncertainty in information but also address partial reliability.

A Z-number is an ordered pair of fuzzy numbers, $(\tilde{A}, \tilde{B})$, where $\tilde{A}$ plays the role of a fuzzy restriction on the value which the random variable X can take, $\tilde{B}$ is referred to

as certainty (reliability) of $\tilde{A}$ [7]. Leveraging the reliability component, Z-numbers enhance the characterization of a variable's fuzzy number by specifying a degree of certainty or confidence. Due to the powerful capability of Z-numbers in describing uncertain and partially reliable information, research on Z-numbers has attracted numerous researchers and has found successful applications in various fields, such as decision-making [8], computing with words [9], risk assessment [10], system control [11], expert system [12] and machine learning [13], [14]. At the core of the definition of Z-numbers is the probability measure of fuzzy numbers, and the two fuzzy components of a Z-number are connected through it. The probability measure is defined through an underlying probability distribution. The underlying probability distribution is not known. However, what is discernible is a fuzzy constraint imposed upon it. The second fuzzy number can be interpreted as a fuzzy restriction on the probability measure associated with the first fuzzy number. Zadeh has stated that a Z-number is, in effect, a summary of probability distributions [7]. The operations of Z-numbers exhibit a close correlation with the operations of probability distributions [15], [16]. The first step is always to find the underlying probability distributions.

Researchers have proposed various methods for determining the underlying probability distributions from the restriction provided by a Z-number. For discrete Z-numbers, the maximum entropy principle is often used to determine the underlying probability distribution [17], [18]. Wang and Yang proposed a linear program to acquire the underlying probability distribution with minimum variance [19]. In [20], a genetic algorithm was subsequently employed to deduce the underlying probability distribution. Chu and Xiao used a nonlinear programming model to get the underlying probability distributions [21]. For continuous Z-numbers, to uncover the underlying probability distribution, it is common to assume that the underlying probability distributions are normal distributions [16], [22]. For triangular continuous Z-numbers, Li et al. used an extended triangular distribution as the form of the underlying probability distribution [23]. In addition to methods that extract information from known Z-numbers, there are also some approaches that attempt to generate Z-numbers using available information. In [24], the authors devised a methodology for generating a Z-number by incorporating the ordered weighted averaging operator and the maximum entropy principle from a discrete fuzzy number, taking into account the element of preference. However, this approach is inconsistent with the definition of Z-numbers. Tian and Kang modified Kang et al.'s method [25], but a significant portion of the information is still provided by experts. In [26],

This work was supported by the China Scholarship Council (CSC), and by the project PID2022-139297OB-I00 funded by MCIN / AEI / 10.13039/501100011033.

Manuscript received April 19, 2021; revised August 16, 2021.

The authors are with the Department of Computer Science and Artificial Intelligence, Andalusian Research Institute in Data Science and Computational Intelligence, DaSCI, University of Granada, 18071 Granada, (e-mail: liyangxueugr@gmail.com, jamoren@decsai.ugr.es, jrtrillo@ugr.es, viedma@decsai.ugr.es).

the authors used the logistic regression algorithm to obtain the probability distributions and then generated Z-numbers. In these methods for generating Z-numbers, the first fuzzy number of the generated Z-number was given, and these methods focused on generating the second fuzzy number.

Given a Z-number, it is possible to derive a set of probability distributions. Conversely, can a Z-number be used to summarize a set of probability distributions? In this paper, a nonlinear model called the Maximum Expected Minimum Entropy (MEME) is proposed for generating a Z-number from a set of probability distributions. Here, the expected value refers to the expected value of the reliability fuzzy number of a Z-number, and naturally, a higher expected value of reliability is desirable. The entropy refers to the entropy of the first fuzzy number, and unequivocally, smaller entropy indicates clearer information and is more conducive to decision-making. Through this model, Z-numbers can be generated from data without any expert knowledge. In addition, we applied the MEME model to classification problems. We introduced a new if-then rule form called Z-valuation if-then rules. It replaces the deterministic consequent part of a fuzzy rule with an uncertain Z-valuation. A Z-valuation is composed of a variable and a Z-number that on this variable [7]. Based on Z-valuation rules, we proposed a Z-valuation rule-based (ZVRB) classification system. This novel system aims to enhance decision-making processes in situations where uncertainty is a key factor. The system is fundamentally composed of two main components: a knowledge base and an inference engine. The knowledge base is responsible for storing the rules, while the inference engine utilizes the fuzzy rules from the knowledge base to determine the class label of input data. The Z-valuation rules generated by Fuzzy Unordered Rule Induction Algorithm (FURIA) [27] and the MEME model are stored in the knowledge base. Since FURIA achieves a favorable balance between accuracy and complexity, it is employed to generate concise and a few fuzzy rules, while the MEME model is used to transform the fuzzy rules to Z-valuation rules. To reduce overfitting caused by the fuzzy rules generated by FURIA, we constrained the number of antecedents in the rules. In inference engine, the predicted class label is with the maximum weighted membership degrees among all class labels. For a new sample not covered by any rules, the existing rules are pruned, and then the MEME model is used again to compute its consequent Z-valuation.

The ZVRB classification system extends traditional fuzzy rule-based systems by incorporating Z-numbers. Fuzzy rule-based classification systems have been widely used for handling uncertainty and imprecision in various domains [28], [29]. However, they have several limitations. First, the consequent part of the traditional fuzzy rule is a single class label, which does not capture any uncertainty about the classification outcome. Second, traditional fuzzy rules often use a rule weight in the range $[0, 1]$ to reflect the rule's importance or confidence. But the rule weight provides only a scalar measure lacking the ability to represent uncertainty about the rule's importance or confidence. Finally, traditional fuzzy rules cannot model second-order uncertainties, such as uncertainty about the uncertainty itself. By replacing the consequent single class label with a Z-number, the ZVRB offers richer representation

of uncertainty in the consequent. It allows modeling of uncertainty about multiple classes, expressing degrees of possibility for each potential class. The first fuzzy component of the consequent Z-number representing the possibility distribution over class labels. The second one representing the reliability or confidence in the assertion that the class label is the first one. Moreover, the Z-number framework naturally incorporates second-order uncertainty, representing uncertainty about the uncertainty. ZVRB systems have broad potential applications in domains where handling both uncertainty and reliability. They can enhance decision-making processes in fields like medical diagnosis, financial risk assessment, environmental modeling, engineering fault detection, and artificial intelligence. By incorporating Z-numbers into fuzzy rules, these systems effectively manage uncertain and imprecise data along with associated confidence levels, leading to more nuanced, accurate, and trustworthy results.

To validate the effectiveness of the ZVRB classification system, we conducted two experiments using 28 different benchmark datasets. The first experiment compared the ZVRB classification system with seven well-known and state-of-the-art classic classifiers, such as K-nearest neighbors [30], XGBoost [31] and support vector machine [32]. The second experiment compared the ZVRB classification system with seven classic and state-of-the-art rule-based classification systems, including Z-number-valued rule-based classification system [33], Z-number-valued rule-based decision tree [34] and FURIA [27]. The statistical test results show the ZVRB classification system is significantly superior to other classic classifiers and rule-based classification systems.

In summary, the principal contributions of this paper can be outlined as follows.

- (1) Proposed a model to generate a Z-number from a set of probability distributions, named Maximum Expected and Minimum Entropy model. This model fills the gap in the reverse generation relationship between Z-numbers and probability distributions. Through this model, Z-numbers can be generated from data without any expert knowledge.
- (2) Designed a new if-then rule form called Z-valuation if-then rule. It replaces the deterministic consequent part of a fuzzy rule with an uncertain Z-valuation. It can further summarize uncertain information in consequent part of a rule.
- (3) Introduced a Z-valuation rule-based classification system and designed the rules generation and inference algorithms for it. This system aims to enhance decision-making processes in situations where uncertainty is a key factor, as opposed to systems with deterministic consequences.

The structure of the remaining sections of the paper is as follows. Section II reviews the background knowledge used in this paper. Section III provides a detailed introduction to the proposed MEME model and includes a numerical example for better understanding and validation. Section IV introduces the designed ZVRB classification system and provides corresponding algorithms. Section V validates the effectiveness of

the system through two experiments. Finally, conclusions are drawn in Section VI.

II. PRELIMINARIES

In this section, some background knowledge used in this paper are briefly reviewed, including fuzzy set, possibility distribution, Z-number and fuzzy rule-based classification system.

A. Fuzzy set and possibility distribution

The variable X is defined to take values in a universe of discourse U , with the generic element denoted as u , and

$$X = u, \quad (1)$$

satisfying that X is assigned the value u , $u \in U$.

Definition 1 (Fuzzy set [2]): A fuzzy set $\tilde{A}$ can be viewed as a mapping $\mu_{\tilde{A}}(\cdot)$ from U to $[0, 1]$. Each element u belongs to $\tilde{A}$ to degree $\mu_{\tilde{A}}(u)$, where $\mu_{\tilde{A}}(u) = 1$ means total membership and $\mu_{\tilde{A}}(u) = 0$ means total non-membership.

Definition 2 (Fuzzy restriction [35]): Let $\tilde{A}$ be a fuzzy set defined on the universe of discourse U with the membership function denoted as $\mu_{\tilde{A}}(\cdot)$. In this context, $\tilde{A}$ serves as a fuzzy restriction on the variable X when it acts as a flexible constraint on the permissible values assignable to X :

$$X = u : \mu_{\tilde{A}}(u). \quad (2)$$

Here, $\mu_{\tilde{A}}(u)$ denotes the degree to which the constraint embodied by $\tilde{A}$ is fulfilled when u is assigned to the variable X .

Definition 3 (Possibility distribution [35]): Consider a variable X with values in the universe U , where $\tilde{A}$ functions as a fuzzy restriction linked to X . In this context, the proposition "X is $\tilde{A}$ " is related to a possibility distribution Π , and the possibility distribution function, denoted as $\pi(\cdot)$, is defined to numerically coincide with the membership function of $\tilde{A}$, $\pi(u) = \mu_{\tilde{A}}(u)$.

Definition 4 (Possibilistic restriction [7]): A possibilistic restriction is defined on a fuzzy set $\tilde{A}$,

$$X \text{ is } \tilde{A}, \quad (3)$$

where $\tilde{A}$ represents a fuzzy set within the universe U with the membership function $\mu_{\tilde{A}}(\cdot)$. $\tilde{A}$ is the function of the possibility distribution of X , denoted as $Poss(X = u) = \mu_{\tilde{A}}(u)$.

The measure of possibilistic uncertainty of a possibilistic distribution Π is defined as follows [36]. When Π is discrete,

$$U(\Pi) = \frac{1}{l_r} \sum_{k=1}^{r-1} (l_{k+1} - l_k) \log_2(|c(\Pi, l_{k+1})|), \quad (4)$$

where $c(\Pi, l)$ is l -cut of Π defined as $c(\Pi, l) = \{u \in U | \pi(u) \geq l\}$; $L_\Pi = \{l_1 = 0, \dots, l_r | l_k = \pi(u), l_k \leq l_{k+1}\}$ is a level set of Π .

Consider a discrete variable X with values in U , whose possibility distribution function is $\pi(u_i)$, $i = 1, 2, \dots, n$, $u_1 \leq u_2 \leq \dots \leq u_n \in U$. The expected value of X , $E[X]$ is defined as follows [37].

$$E[X] = \sum_{i=1}^n u_i w_i, \quad (5)$$

where the weight w_i is given by

$$w_i = \frac{1}{2} \left(\max_{j=1}^i \pi(u_j) - \max_{j=1}^{i-1} \pi(u_j), 0 \right) + \frac{1}{2} \left(\max_{j=i}^n \pi(u_j) - \max_{j=i+1}^n \pi(u_j), 0 \right). \quad (6)$$

The weights satisfy $w_i \in [0, 1]$ and $\sum_{i=1}^n w_i = 1$.

B. Z-number and Z-valuation

Definition 5 (Probability restriction [7]): A probability restriction is defined on a probability distribution $p(\cdot)$,

$$X \text{ is } p, \quad (7)$$

where $p(\cdot)$ is the probability distribution of X , $Prob(u) = p(u)$.

Definition 6 (Probability measure [38]): If X is a random variable, then "X is $\tilde{A}$ " is a fuzzy event. If $\tilde{A}$ is discrete, the probability of this event $Prob(X \text{ is } \tilde{A})$ can be expressed as:

$$Prob(X \text{ is } \tilde{A}) = \sum_{i=1}^n \mu_{\tilde{A}}(u_i) p(u_i), \quad (8)$$

where $p(u)$ is the underlying (hidden) probability density of X .

Definition 7 (Discrete Z-number [7]): A discrete Z-number is an ordered pair of discrete fuzzy numbers, $Z = (\tilde{A}, \tilde{B})$, where $\tilde{A}$ is a discrete possibilistic restriction on the variable X with possibility distribution $\mu_{\tilde{A}}(u) : \{u_1, u_2, \dots, u_n\} \rightarrow [0, 1]$. $\tilde{B}$ is a discrete possibilistic restriction on the probability of "X is $\tilde{A}$ " with possibility distribution $\mu_{\tilde{B}}(v) : \{v_1, v_2, \dots, v_m\} \rightarrow [0, 1]$. $\tilde{B}$ also can be viewed as the reliability measure of $\tilde{A}$.

Definition 8 (Discrete Z-valuation [7]): A discrete Z-valuation, $(X, \tilde{A}, \tilde{B})$, is defined on a discrete Z-number $(\tilde{A}, \tilde{B})$:

$$Prob(X \text{ is } \tilde{A}) \text{ is } \tilde{B}, \quad (9)$$

where $Prob(X \text{ is } \tilde{A}) = \sum_{i=1}^n \mu_{\tilde{A}}(u_i) p(u_i)$. The underlying probability distribution, $p(\cdot)$, is not known. However, what is ascertainable is a constraint on $p(\cdot)$. Essentially, a Z-number can be perceived as a summarization of p [7].

It must be noted that the discrete Z-numbers used in this paper do not entirely adhere to the definition of Z-numbers, as we overlook two of its conditions when using fuzzy numbers: the normality and convexity. We employ generalized discrete Z-numbers, which are based on discrete fuzzy subsets rather than discrete fuzzy numbers. The maximum value of membership degrees is not necessarily 1, and the membership degree values do not satisfy convexity, as the base values are unordered.

C. Fuzzy rule-based classification system and FURIA

A fuzzy rule-based classification system is a technology employed for data classification and decision support [39], [40]. It harnesses fuzzy logic and a set of fuzzy rules to address issues involving fuzziness and uncertainty. It primarily consists of two components: a knowledge base and an inference engine. The knowledge base is responsible for storing the fuzzy rules, while the inference engine utilizes the fuzzy rules from the

knowledge base to determine the class or label of input data. Fuzzy rules are rules defined based on fuzzy sets, and they are used to map input data to different classes or labels. Each fuzzy rule typically follows an "if-then" format. Constructing a fuzzy rule-based classification system involves generating a set of fuzzy rules from training data to be stored in the knowledge base and designing an inference engine.

For a classification problem, let's assume that the training dataset is denoted as $T = \{x_1, x_2, \dots, x_n\}$. This dataset contains n training samples, where each sample comprises m feature values, denoted as $x_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}$, and one class label. There are a total of C different class labels, denoted as $\Theta = \{\theta_1, \theta_2, \dots, \theta_C\}$. A fuzzy if-then rule can be written as:

$$\text{Rule } R_l : \text{If } x_{i1} \text{ is } \tilde{A}_{l1} \text{ and } \dots \text{ and } x_{im} \text{ is } \tilde{A}_{lm} \quad (10) \\ \text{then } x_i \text{ belong to class } y_l \text{ with } RW_l.$$

Where $\tilde{A}_{lj}$ is an antecedent fuzzy set, $j = 1, \dots, m$; $y_l \in \Theta$ is the output class label and $RW_l \in [0, 1]$ is the rule weight.

There are various existing methods for generating fuzzy rules and performing inference, and FURIA is one of state-of-the-art methods that strike a balance between accuracy and complexity. FURIA is an algorithm derived from the Repeated Incremental Pruning to Produce Error Reduction (RIPPER) algorithm, and it has shown several improvements in the performance of rule induction tasks [27], [41]. FURIA's fuzzy rules are generated through the process of fuzzifying the final crisp rules obtained from a modified version of the RIPPER algorithm. A fuzzy rule generated by FURIA can be written as follows.

$$\text{Rule } R_l : \text{If } x_{i\rho(1)} \text{ is } \tilde{A}_{l\rho(1)} \text{ and } \dots \text{ and } x_{i\rho(d_l)} \text{ is } \tilde{A}_{l\rho(d_l)} \quad (11) \\ \text{then } x_i \text{ belong to class } y_l \text{ with } RW_l.$$

Where $\tilde{A}_{l\rho(j)}$ is an antecedent trapezoidal fuzzy set, $j = 1, \dots, d_l$, $d_l \leq m$; $\rho(j)$ is an index function indicating the index of feature considered in the j -th antecedent fuzzy set of the rule; $y_l \in \Theta$ is the output class label and $RW_l \in [0, 1]$ is the rule weight.

III. MAXIMUM EXPECTED MINIMUM ENTROPY MODEL

The proponent of Z-numbers, Zadeh, stated that a Z-number serves as a summary of probability distributions on the variable [7]. Given a Z-number, $Z = (\tilde{A}, \tilde{B})$, it is possible to derive a set of probability distributions, $p(u)$, for the variable, X , based on the possibilistic restriction and the probability measure. Conversely, can a Z-number be used to summarize a set of probability distributions? In this section, we introduce a Maximum Expected Minimum Entropy (MEME) model to obtain a Z-number for summarizing a set of probability distributions. In this paper, we exclusively focus on scenarios where the variables are discrete.

Given a set of probability distributions on a discrete variable X , $P = \{p_1, p_2, \dots, p_m\}$, where our objective is to summarize them using a discrete Z-number, $Z = (\tilde{A}, \tilde{B})$, with membership functions $\mu_{\tilde{A}}(u)$ and $\mu_{\tilde{B}}(v)$, respectively. Naturally, the base values of $\tilde{A}$ are the possible values that the variable can take. For the membership degrees of $\tilde{A}$ and the base values of

$\tilde{B}$ along with their membership degrees, they are obtained through a model combining two optimization models. The first model is maximum expected value of reliability model:

$$\max E[V]$$

subject to:

$$v_j = \sum_{i=1}^n p_j(u_i) \mu_{\tilde{A}}(u_i), \\ \mu_{\tilde{B}}(v_j) = \frac{|\{v_k = v_j | k = 1, 2, \dots, m\}|}{\max(|\{v_k = v_j | k = 1, 2, \dots, m\}|)}, \\ 0 \leq \mu_{\tilde{A}}(u_i) \leq 1, \\ 0 \leq v_j \leq 1, \\ 0 \leq \mu_{\tilde{B}}(v_j) \leq 1, \\ i = 1, 2, \dots, n, \\ j = 1, 2, \dots, m, \quad (12)$$

where $E[V]$ is the expected value of reliability ($\tilde{B}$) calculated by:

$$E[V] = \sum_{j=1}^m v_j w_j, \quad (13)$$

where V indicates the discrete variable of reliability. The weight w_j is given by:

$$w_j = \frac{1}{2} \left(\max_{k=1}^j \mu_{\tilde{B}}(v_k) - \max_{k=1}^{j-1} \mu_{\tilde{B}}(v_k), 0 \right) \\ + \frac{1}{2} \left(\max_{k=j}^m \mu_{\tilde{B}}(v_k) - \max_{k=j+1}^m \mu_{\tilde{B}}(v_k), 0 \right). \quad (14)$$

It is evident that $E[V] \in [0, 1]$.

The first constraint is the probability measure of $\tilde{A}$ according to the definition of Z-numbers; The second constraint is to consider the frequency of occurrence of the value v_j as its membership degree. The rest are some standard constraints for Z-numbers.

However, using only this optimization model is insufficient to obtain a satisfactory Z-number. Since this optimization model produces only one result regardless of the input probability distributions, where all membership degrees are one, leading to the maximum expected value of reliability $E[V] = 1$. According to Eq. (13), as the values of v_j increase, the value of $E[V]$ increases. Furthermore, based on the first constraint, as the value of $\mu_{\tilde{A}}(u_i)$ increases, the calculated v_j value also increases. It can be easily observed that when $\mu_{\tilde{A}}(u_i) = 1, \forall i = 1, 2, \dots, n$, $v_j = \sum_{i=1}^n p_j = 1, \forall j = 1, 2, \dots, m$, resulting in the maximum expected value $E[V] = 1$.

The first fuzzy number $\tilde{A}$ is a possibilistic restriction on the variable X , therefore a smaller uncertainty in $\tilde{A}$ is more favorable for determining the values of the variable. Hence, we need the second optimization model to constrain the entropy of $\tilde{A}$. The second model is minimum entropy model:

$$\min U(\mu_{\tilde{A}})$$

subject to:

$$0 \leq \mu_{\tilde{A}}(u_i) \leq 1, \\ i = 1, 2, \dots, n, \quad (15)$$

where $U(\mu_{\tilde{A}})$ is the measure of possibilistic uncertainty (entropy) of $\tilde{A}$, given by:

$$U(\mu_{\tilde{A}}) = \frac{1}{l_r} \sum_{k=1}^{r-1} (l_{k+1} - l_k) \log_2(|c(\mu_{\tilde{A}}, l_{k+1})|), \quad (16)$$

where $c(\mu_{\tilde{A}}, l)$ is l -cut of $\mu_{\tilde{A}}$ defined as $c(\mu_{\tilde{A}}, l) = \{u \in U | \mu_{\tilde{A}}(u) \geq l\}$; $L_{\mu_{\tilde{A}}} = \{l_1 = 0, \dots, l_r | l_k = \mu_{\tilde{A}}(u), l_k \leq l_{k+1}\}$ is a level set of $\mu_{\tilde{A}}$.

Combining Model (12) and Model (15) gives rise to the Step 2: MEME model:

$$\min \frac{U(\mu_{\tilde{A}})/\log_2(n) + 0.1}{E[V] + 0.1}$$

subject to:

$$\begin{aligned} v_j &= \sum_{i=1}^n p_j(u_i) \mu_{\tilde{A}}(u_i), \\ \mu_{\tilde{B}}(v_j) &= \frac{|\{v_k = v_j | k = 1, 2, \dots, m\}|}{\max(|\{v_k = v_{j'} | k = 1, 2, \dots, m\}|)}, \\ 0 &\leq \mu_{\tilde{A}}(u_i) \leq 1, \\ 0 &\leq v_j \leq 1, \\ 0 &\leq \mu_{\tilde{B}}(v_j) \leq 1, \\ i &= 1, 2, \dots, n, \\ j &= 1, 2, \dots, m, \end{aligned} \quad (17)$$

where 0.1 is intended to prevent division by zero. Since the expected value of reliability ranges from $E[V] \in [0, 1]$, and the entropy value ranges from $U(\mu_{\tilde{A}}) \in [0, \log_2(n)]$, $\log_2(n)$ is to balance these two aspects effectively. The MEME model aims to minimize the uncertainty of $\tilde{A}$ while satisfying the maximum reliability. A numerical example is used to illustrate Step 3: Mutation.

Example 1: Suppose there are ten probability distributions related to the variable X , $P = \{p_1, p_2, \dots, p_{10}\}$, as shown in Table I. Our objective is to generate a Z-number, $Z = (\tilde{A}, \tilde{B})$, to summarize these probability distributions. As both the calculation of expected values and entropy values involve sorting operations, using traditional gradient-based optimization methods can be challenging. Therefore, we employ the differential evolution (DE) algorithm [42] to find the optimal values of Model (17).

	u_1	u_2	u_3	u_4	u_5
p_1	0.08	0.23	0.51	0.05	0.13
p_2	0.03	0.18	0.49	0.0	0.3
p_3	0.16	0.18	0.54	0.11	0.01
p_4	0.20	0.09	0.45	0.15	0.11
p_5	0.08	0.23	0.51	0.05	0.13
p_6	0.16	0.18	0.54	0.11	0.01
p_7	0.08	0.23	0.51	0.05	0.13
p_8	0.04	0.01	0.47	0.21	0.27
p_9	0.05	0.17	0.46	0.25	0.07
p_{10}	0.23	0.18	0.56	0.0	0.03

TABLE I: Ten probability distributions related to the variable X .

Step 1: Population initialization.

Randomly generate N ($N = 100$) individuals, denoted as $IN = \{in_1, in_2, \dots, in_{100}\}$, in which each

individual includes five membership degrees, $in_q = \{\mu_{\tilde{A}}(u_1), \mu_{\tilde{A}}(u_2), \dots, \mu_{\tilde{A}}(u_5)\}$, $q = 1, 2, \dots, 100$, satisfying $\mu_{\tilde{A}}(u_i) \in [0, 1]$. For example, an individual in_1 is as follows.

$$\begin{aligned} \mu_{\tilde{A}}(u_1) &= 0.69, & \mu_{\tilde{A}}(u_2) &= 0.24, & \mu_{\tilde{A}}(u_3) &= 0.3, \\ \mu_{\tilde{A}}(u_4) &= 0.43, & \mu_{\tilde{A}}(u_5) &= 0.09. \end{aligned}$$

Step 2: Individual fitness calculation.

After generating the individuals, the values of v_j , $j = 1, 2, \dots, 10$ can be calculated using the first constraint in Model (17). Using the above individual mentioned in Step 1, the values of v_j are

$$\begin{aligned} v_1 &= 0.3, v_2 = 0.24, v_3 = 0.36, v_4 = 0.37, v_5 = 0.3, \\ v_6 &= 0.36, v_7 = 0.3, v_8 = 0.29, v_9 = 0.33, v_{10} = 0.37. \end{aligned}$$

The values of $\mu_{\tilde{B}}(v_j)$ can be calculated using the second constraint in Model (17). The results are

$$\begin{aligned} \mu_{\tilde{B}}(0.24) &= 0.33, \mu_{\tilde{B}}(0.29) = 0.33, \mu_{\tilde{B}}(0.3) = 1.0, \\ \mu_{\tilde{B}}(0.33) &= 0.33, \mu_{\tilde{B}}(0.36) = 0.67, \mu_{\tilde{B}}(0.37) = 0.67. \end{aligned}$$

Then the expected value of $\tilde{B}$ is

$$E[V] = 0.3133.$$

The entropy of $\tilde{A}$ is

$$U(\mu_{\tilde{A}}) = 1.06.$$

Therefore, the fitness of in_1 is $Fit(in_1) = \frac{U(\mu_{\tilde{A}})/\log_2(5) + 0.1}{E[V] + 0.1} = 1.35$.

Step 3: Mutation.

Randomly select three individuals, denoted as in_{r1} , in_{r2} and in_{r3} , to calculate the mutated individual, mu_q , $q = 1, 2, \dots, 100$, as follows.

$$mu_q = in_{r1} + F(in_{r2} - in_{r3}), \quad (18)$$

where F is the mutation probability ($F = 0.8$ here). The mutated individuals are denoted as $MU = \{mu_1, mu_2, \dots, mu_{100}\}$.

Step 4: Crossover.

Randomly select an individual in IN , denoted as in_r , and randomly select a mutated individual in MU , denoted as mu_r , to calculate the crossover individual, cr_q , $q = 1, 2, \dots, 100$, as follows.

$$cr_q = \begin{cases} in_r, & rand(0, 1) < CR \\ mu_r, & else. \end{cases} \quad (19)$$

Where CR is the crossover probability ($CR = 0.5$ here).

Step 5: Selection.

Calculate the fitness of crossover individuals, then compare them to the original individuals, retaining the ones with lower fitness values for the next generation.

Step 6: Check if the stopping criterion has been met.

The stopping criterion is typically the maximum number of iterations (100 here). When the maximum number of iterations is reached, select the individual with the lowest

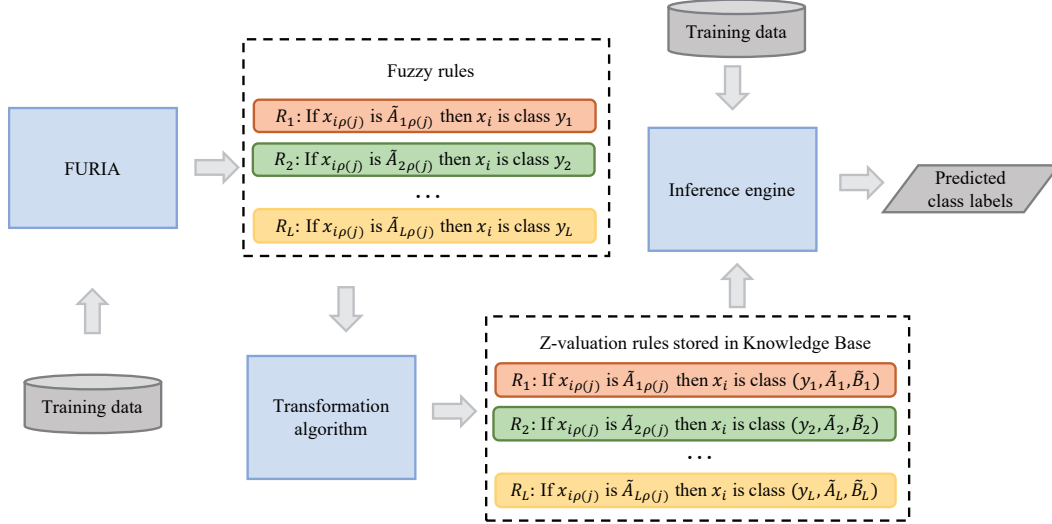


Fig. 1: The structure of the proposed ZVRB classification system.

fitness as the output; otherwise, proceed to Step 3. In this example, the output Z-number is $Z = (\tilde{A}, \tilde{B})$, where

$$\tilde{A} = \frac{0.001}{u_1} + \frac{0.0003}{u_2} + \frac{1}{u_3} + \frac{0.002}{u_4} + \frac{0.001}{u_5},$$

$$\tilde{B} = \frac{0.33}{0.46} + \frac{0.33}{0.47} + \frac{0.67}{0.49} + \frac{1}{0.51} + \frac{0.67}{0.54} + \frac{0.33}{0.56}.$$

The Z-number $Z = (\tilde{A}, \tilde{B})$ coincides with the probability distributions in Table I. First, in all distributions, u_3 has the highest or a significantly high probability compared to others. The highest membership degree in $\tilde{A}$ is at u_3 . Moreover, the reliability measure $\tilde{B}$ aligns with the probability measures of $\tilde{A}$, with higher membership degrees assigned to more frequently occurring probability measures.

Even though we start from numerical probability distributions, the Z-number generated retains its linguistic interpretability by summarizing the key information in human-understandable terms. $\tilde{A}$ can be interpreted as a linguistic term such as "approximately u_3 " or "most likely u_3 ". $\tilde{B}$ can be expressed using terms like "medium confidence" or "moderate certainty". The output Z-number, $Z = (\tilde{A}, \tilde{B})$, can be interpreted as "The variable X is approximately u_3 with medium confidence".

IV. Z-VALUATION RULE-BASED CLASSIFICATION SYSTEM

In this section, we introduce a novel rule-based classification system termed the Z-valuation rule-based (ZVRB) classification system. Z-valuation if-then rules replace the deterministic consequences of fuzzy rules with uncertain Z-valuations. The system primarily consists of two components: a knowledge base and an inference engine. The system's workflow is illustrated in Figure 1. First, fuzzy rules are generated from the training data by FURIA, then these fuzzy rules are transformed into Z-valuation rules through the transformation algorithm. Finally, the inference engine classifies the test data using the Z-valuation rules.

A. Z-valuation if-then rule

The Z-number-valued if-then rule defined in [43] is the form of

$$\text{Rule } R_l^Z : \text{If } x_{l1} \text{ is } Z_{l1} \text{ and } \dots \text{ and } x_{lm} \text{ is } Z_{lm} \text{ then } x_i \text{ belong to class } y_l \text{ with } RW_l. \quad (20)$$

Where Z_{lj} is an antecedent Z-number, $j = 1, 2, \dots, m$; $y_l \in \Theta$ is the output class label and $RW_l \in [0, 1]$ is the rule weight. The antecedent Z-numbers are not suitable for being used as Z-valuations. This is because a Z-valuation operate on the potential values of a variable X , and the value of variable X can only take on one value in an observation. As the probability of one possible value increases, the probabilities of other possible values decrease. The different possible values are mutually exclusive. However, for rules, samples with different feature values of the same feature may all potentially belong to the same class. These different feature values are not mutually exclusive. But the class labels of samples are mutually exclusive. A sample cannot belong to two classes simultaneously. Therefore, the consequent part of the rule can serve as a Z-valuation. Hence, a new form of rules, termed Z-valuation rules, is defined as follows:

$$\text{Rule } R_l^Z : \text{If } x_{l1} \text{ is } \tilde{A}_{l1} \text{ and } \dots \text{ and } x_{lm} \text{ is } \tilde{A}_{lm} \text{ then } (y_l, \tilde{A}_l, \tilde{B}_l). \quad (21)$$

Where $\tilde{A}_{lj}$ is an antecedent fuzzy set, $j = 1, \dots, m$; $(y_l, \tilde{A}_l, \tilde{B}_l)$ is a Z-valuation, where y_l is a variable about the class label of R_l^Z , $(\tilde{A}_l, \tilde{B}_l)$ is a discrete Z-number.

The above form is the Z-valuation rule based on general fuzzy rules. If it is based on fuzzy rules generated by FURIA, then Z-valuation rules can be written in the following form.

$$\text{Rule } R_l^Z : \text{If } x_{i\rho(1)} \text{ is } \tilde{A}_{l\rho(1)} \text{ and } \dots \text{ and } x_{i\rho(d)} \text{ is } \tilde{A}_{l\rho(d)} \text{ then } (y_l, \tilde{A}_l, \tilde{B}_l). \quad (22)$$

Where $A_{l\rho(j)}$ is an antecedent trapezoidal fuzzy set, $j = 1, \dots, d_l$, $d_l \leq m$; $\rho(j)$ is an index function indicating the index of feature considered in the j -th antecedent fuzzy set of the rule; $(y_l, \tilde{A}_l, \tilde{B}_l)$ is a Z-valuation, where y_l is a variable about the class label of R_l^Z , $(\tilde{A}_l, \tilde{B}_l)$ is a generalized discrete Z-number.

B. Knowledge base

During the model training phase, a set of Z-valuation rules are generated and then stored in the knowledge base. Since FURIA achieves a favorable balance between accuracy and complexity, the Z-valuation rule generation primarily relies on FURIA and the MEME model. FURIA is employed to generate concise and a few fuzzy rules, while the MEME model is used to generate the consequent Z-valuations for the fuzzy rules.

For l -th fuzzy rule generated by FURIA R_l shown in Eq. (10), each sample covered by this rule can be considered as an underlying probability distribution about the class label. Since the class labels of the training samples are known, the probability distribution for sample x_i is as follows.

$$p_i(\theta_c) = \begin{cases} 1, & \text{if } x_i \text{ belongs to } \theta_c, \\ 0, & \text{otherwise.} \end{cases} \quad (23)$$

Therefore, there are a total of C different probability distributions. This allows us to obtain the Z-valuation, $(y_l, \tilde{A}_l, \tilde{B}_l)$, based on the MEME model. According to Model (17), the first constraint can be written as:

$$v_i = \sum_{c=1}^C p_i(\theta_c) \mu_{\tilde{A}_l}(\theta_c) = \mu_{\tilde{A}_l}(\theta_c), \text{ if } x_i \text{ belongs to } \theta_c. \quad (24)$$

There are a total of C different values of v , denoted as $\{v_{\theta_1}, v_{\theta_2}, \dots, v_{\theta_C}\}$, where $v_{\theta_c} = \mu_{\tilde{A}_l}(\theta_c)$.

One difference from Model (17) lies in the method of calculating membership degrees of $\tilde{B}_l$. In this model, the coverage degree of the rule is added in the second constraint as follows.

$$\mu_{\tilde{B}_l}(v_{\theta_c}) = \frac{\sum_{x_i \in X_{\theta_c}} \phi_l(x_i)}{\max_{\theta_c' \in \Theta} (\sum_{x_i \in X_{\theta_c'}} \phi_l(x_i))}, \quad (25)$$

where $X_{\theta_c} = \{x_i \in X | x_i \text{ belongs to } \theta_c\}$; $\phi_l(x_i)$ is the coverage degree of x_i covered by the rule R_l , calculated by:

$$\phi_l(x_i) = \prod_{j=1}^d \mu_{\tilde{A}_{l\rho(j)}}(x_{i\rho(j)}). \quad (26)$$

Overall, the Z-valuation $(y_l, \tilde{A}_l, \tilde{B}_l)$ can be derived based on the following MEME model.

$$\min \frac{U(\mu_{\tilde{A}_l})/\log_2(C) + 0.1}{E[V] + 0.1}$$

subject to:

$$\begin{aligned} v_{\theta_c} &= \mu_{\tilde{A}_l}(\theta_c), \\ \mu_{\tilde{B}_l}(v_{\theta_c}) &= \frac{\sum_{x_i \in X_{\theta_c}} \phi_l(x_i)}{\max_{\theta_c' \in \Theta} (\sum_{x_i \in X_{\theta_c'}} \phi_l(x_i))}, \\ 0 &\leq \mu_{\tilde{A}_l}(\theta_c) \leq 1, \\ 0 &\leq \mu_{\tilde{B}_l}(v_{\theta_c}) \leq 1, \\ c &= 1, 2, \dots, C. \end{aligned} \quad (27)$$

In this way, after optimization through the MEME model, a fuzzy rule transforms into a Z-valuation rule. Algorithm 1 summarizes the process of transforming fuzzy rules into Z-valuation rules (using DE algorithm). Next, we illustrate the process using a classification example for E. coli dataset.

Algorithm 1 Transformation of fuzzy rules into Z-valuation rules.

Input: The training dataset $T = \{x_1, x_2, \dots, x_n\}$; the fuzzy ruleset $RS = \{R_1, R_2, \dots, R_L\}$; the number of individuals N ; the mutation probability F ; the crossover probability CR , the maximum number of iterations $MaxItera$.

Output: Z-valuation ruleset $RS^Z = \{R_1^Z, R_2^Z, \dots, R_L^Z\}$.

```

1: for Each rule  $R_l \in RS$  do
2:    $\mu_{\tilde{B}_l}(v_{\theta_c}) = \frac{\sum_{x_i \in X_{\theta_c}} \phi_l(x_i)}{\max_{\theta_c' \in \Theta} (\sum_{x_i \in X_{\theta_c'}} \phi_l(x_i))}$ ,  $\theta_c \in \Theta$ .
3:   Initialize  $N$  individuals with  $\{\mu_{\tilde{A}_l}(\theta_c) | \mu_{\tilde{A}_l}(\theta_c) \in [0, 1], \theta_c \in \Theta\}$ .
4:   Compute the fitness of the individuals.
5:    $itera \leftarrow 1$ 
6:   while  $itera \leq MaxItera$  do
7:     Mutation
8:     Crossover
9:     Selection
10:     $itera \leftarrow itera + 1$ 
11:   end while
12:   Select the best individual as the consequent Z-valuation of rule  $R_l$ ,  $(y_l, \tilde{A}_l, \tilde{B}_l)$ .
13:    $R_l^Z \leftarrow$  the antecedents of  $R_l$  and the consequent Z-valuation  $(y_l, \tilde{A}_l, \tilde{B}_l)$ .
14:    $R_l^Z$  adds in  $RS^Z$ 
15: end for

```

Example 2: The E. coli dataset from UCI repository is a multiclass classification dataset having eight features, where seven numerical features, *mecg*, *gvh*, *lip*, *chg*, *aac*, *alm1* and *alm2*, are utilized and the feature *sequence name* is omitted. There are eight classes *cp*, *im*, *imS*, *imL*, *imU*, *om*, *omL*, *pp*.

Suppose we obtain a fuzzy rule R_1 generated by FURIA as follows.

Rule R_1 : If *alm1* is $(-\infty, -\infty, 0.38, 0.39)$
and *mecg* is $(-\infty, -\infty, 0.54, 0.56)$
then belong to class *cp*.

We can transform it into a Z-valuation rule using the following steps.

Step 1: Compute the membership degrees of $\tilde{B}_1$.

$$\begin{aligned} \mu_{\tilde{B}_1}(v_{cp}) &= 1, \mu_{\tilde{B}_1}(v_{im}) = 0.0208, \mu_{\tilde{B}_1}(v_{imS}) = 0, \\ \mu_{\tilde{B}_1}(v_{imL}) &= 0, \mu_{\tilde{B}_1}(v_{imU}) = 0, \mu_{\tilde{B}_1}(v_{om}) = 0.0104, \\ \mu_{\tilde{B}_1}(v_{omL}) &= 0, \mu_{\tilde{B}_1}(v_{pp}) = 0.0208. \end{aligned}$$

Step 2: Population initialization. Randomly generate N ($N = 100$) individuals with $\{\mu_{\tilde{A}_1}(\theta_c) | \mu_{\tilde{A}_1}(\theta_c) \in [0, 1], \theta_c \in \Theta\}$. Such as an individual is

$$\mu_{\tilde{A}_1}(cp) = 0.7552, \mu_{\tilde{A}_1}(im) = 0.1285,$$

$$\begin{aligned}\mu_{\tilde{A}_1}(imS) &= 0.9032, \mu_{\tilde{A}_1}(imL) = 0.7361, \\ \mu_{\tilde{A}_1}(imU) &= 0.1507, \mu_{\tilde{A}_1}(om) = 0.2377, \\ \mu_{\tilde{A}_1}(omL) &= 0.0176, \mu_{\tilde{A}_1}(pp) = 0.7239.\end{aligned}$$

Step 3: Compute the fitness of the individuals. Using the initialized individuals and the first constraint of Model (27), the base values of $\tilde{B}_1$, v_{θ_c} , can be obtained as follows.

$$\begin{aligned}v_{cp} &= \mu_{\tilde{A}_1}(cp) = 0.7552, v_{im} = 0.1285, v_{imS} = 0.9032, \\ v_{imL} &= 0.7361, v_{imU} = 0.1507, v_{om} = 0.2377, \\ v_{omL} &= 0.0176, v_{pp} = 0.7239.\end{aligned}$$

Then the expected value of $\tilde{B}_1$ is $E[V] = 0.7486$. The entropy of $\tilde{B}_1$ is $U(\mu_{\tilde{A}_1}) = 1.6345$. The fitness of this individual is $\frac{U(\mu_{\tilde{A}_1})/\log_2(8)+0.1}{E[V]+0.1} = 0.8286$.

The remaining steps are identical to Steps 3 through 6 in Example 1. After 100 iterations, we obtained the final membership degrees of $\tilde{A}_1$.

$$\begin{aligned}\mu_{\tilde{A}_1}(cp) &= 0.9578, \mu_{\tilde{A}_1}(im) = 0.0056, \mu_{\tilde{A}_1}(imS) = 0.0012, \\ \mu_{\tilde{A}_1}(imL) &= 0.0005, \mu_{\tilde{A}_1}(imU) = 0.0048, \mu_{\tilde{A}_1}(om) = 0.0014, \\ \mu_{\tilde{A}_1}(omL) &= 0.0032, \mu_{\tilde{A}_1}(pp) = 0.0032.\end{aligned}$$

Thus, the fuzzy rule R_1 is transformed into a Z-valuation rule R_1^Z :

$$\begin{aligned}\text{Rule } R_1^Z : & \text{ If } aml1 \text{ is } (-\infty, -\infty, 0.38, 0.39) \\ & \text{ and } mcg \text{ is } (-\infty, -\infty, 0.54, 0.56) \\ & \text{ then } (y_1, \tilde{A}_1, \tilde{B}_1),\end{aligned}$$

where y_1 is a variable about the class label; $(\tilde{A}_1, \tilde{B}_1)$ is a discrete generalized Z-number as the follows.

$$\begin{aligned}\tilde{A}_1 &= \frac{0.9578}{cp} + \frac{0.0056}{im} + \frac{0.0012}{imS} + \frac{0.0005}{imL} \\ &+ \frac{0.0048}{imU} + \frac{0.0014}{om} + \frac{0.0032}{omL} + \frac{0.0032}{pp}; \\ \tilde{B}_1 &= \frac{0.0104}{0.0014} + \frac{0.0208}{0.0032} + \frac{0.0208}{0.0056} + \frac{1}{0.9578}.\end{aligned}$$

The output Z-number can be interpreted as "The class label of R_1^Z is approximately cp with high confidence."

We can observe that the Z-numbers here do not fully align with the standard definition of discrete Z-numbers, but rather represent a generalized form of discrete Z-numbers. The first component $\tilde{A}_1$ does not satisfy the normality and convexity properties of a fuzzy number. This is because the base values of $\tilde{A}_1$ are determined by the output class labels of the dataset. The dataset's class labels are always not ordered, then $\tilde{A}_1$ is not a fuzzy number but a fuzzy subset. Meanwhile, $\tilde{B}_1$ is a standard discrete fuzzy number, satisfying the normality and convexity properties of fuzzy numbers. If the outputs of the dataset are numerical and can be ordered, we can incorporate normality and convexity as constraints into the model (27), thereby generating discrete Z-numbers that strictly conform to the definition.

Algorithm 2 Inference algorithm.

Input: The new sample x_e ; the Z-valuation ruleset $RS^Z = \{R_1^Z, R_2^Z, \dots, R_L^Z\}$; the number of individuals N ; the mutation probability F ; the crossover probability CR , the maximum number of iterations $MaxItera$.

Output: The predicted class label of x_e , y^* .

```

1:  $max \leftarrow -\infty$ 
2: for Each class label  $\theta_c \in \Theta$  do
3:    $\delta(\theta_c) \leftarrow 0$ 
4:   for Each rule  $R_l^Z \in RS^Z$  do
5:      $\delta(\theta_c) \leftarrow \delta(\theta_c) + \phi_l(x_e)\mu_{\tilde{A}_l}(\theta_c)$ 
6:   end for
7:   if  $\delta(\theta_c) > max$  then
8:      $max \leftarrow \delta(\theta_c)$ 
9:      $y^* \leftarrow \theta_c$ 
10:  end if
11: end for
12: if  $max \leq 0$  then
13:    $max \leftarrow -\infty$ 
14:   for Each rule  $R_l^Z \in RS^Z$  do
15:      $j' \leftarrow 0$ 
16:     A new empty rule  $r_l^Z$ 
17:     for The antecedent  $\tilde{A}_{l\rho(j)}$  of rule  $R_l^Z$  do
18:       if  $\mu_{\tilde{A}_{l\rho(j)}}(x_{e\rho(j)}) < 0$  then
19:         break
20:       end if
21:        $j' \leftarrow j' + 1$ 
22:        $\tilde{A}_{l\rho(j)}$  adds in  $r_l^Z$  as  $j'$ -th antecedent.
23:     end for
24:     if  $j' > 0$  then
25:       Calculate the consequent Z-valuation,
        $(y_l, \tilde{A}'_l, \tilde{B}'_l)$ , of rule  $r_l^Z$  by Model (27).
26:       for Each class label  $\theta_c \in \Theta$  do
27:          $\delta(\theta_c) \leftarrow \delta(\theta_c) + \phi_l(x_e)\mu_{\tilde{A}'_l}(\theta_c)$ 
28:       end for
29:     end if
30:     if  $\delta(\theta_c) > max$  then
31:        $max \leftarrow \delta(\theta_c)$ 
32:        $y^* \leftarrow \theta_c$ 
33:     end if
34:   end for
35: end if

```

C. Inference engine

Assuming that the knowledge base stores L Z-valuation rules, denoted as the set $RS = \{R_1^Z, R_2^Z, \dots, R_L^Z\}$. For a new sample $x_e = \{x_{e1}, x_{e2}, \dots, x_{em}\}$ without class label, its class label can be predicted as y^* , obtained by:

$$y^* = \operatorname{argmax}_{\theta_c \in \Theta} (\delta(\theta_c)). \quad (28)$$

Where $\delta(\theta_c) = \sum_{l=1}^L \phi_l(x_e)\mu_{\tilde{A}'_l}(\theta_c)$ is the weighted sum of membership degrees of θ_c among all rules. The second fuzzy number, $\tilde{B}'_l$, is not used in the inference engine because $\tilde{A}_l$ has already reached the maximum reliability level through the maximum expected value of reliability model.

In the case where a new sample is not covered by any existing rules, meaning for any rule $\phi_l(x_e) = 0$, $l = 1, 2, \dots, L$, the antecedents of the rule will be pruned until at least one antecedent covers the sample. The remaining antecedents become a new rule, but it is not stored in the knowledge base. Subsequently, the MEME model is employed to deduce the consequent Z-valuation for this rule, which is then used to predict the class label of the sample. The inference algorithm is illustrated as Algorithm 2.

V. EXPERIMENTS

In this section, two experiments were conducted to validate the effectiveness and superiority of the proposed ZVRB classification system compared to commonly used classification algorithms and rule-based classification systems. The first subsection introduced the datasets, the validation methods, and statistical test approach used for comparative experiments. The second subsection compared the classification performance of the ZVRB classification system with seven commonly used classifiers. The third subsection compared the classification performance of the ZVRB classification system with seven rule-based classification systems.

A. Experimental setup

In the experiments, we selected 28 classification benchmarks from the UCI and KEEL repositories, each characterized by varying sample sizes, feature dimensions, and the number of class labels. Table II provides information about these datasets, where *#Samples* represents the number of samples, *#Features* indicates the number of features, and *#Classes* represents the number of class labels.

To assess the generalization ability of the model, we employed the 10-fold cross-validation (10-CV) technique. In this method, the dataset is partitioned into 10 subsets, referred to as folds. The model undergoes training 10 times, with each iteration utilizing 9 folds for training and the remaining fold for testing. This process yields 10 classification accuracies, and their average is taken as the final evaluation metric. Thus, for the 28 datasets, we can obtain 28 classification accuracies using this approach.

Merely comparing the average value of these 28 accuracies is insufficient for assessing the superiority or inferiority of the models. Specifically, we utilized the Wilcoxon signed-rank test to determine if there were significant differences between the accuracies from the compared two models. The Wilcoxon signed-rank test is a non-parametric test used to determine whether two dependent values are selected from populations with the same distribution. Assuming there are two sets of data, calculate the differences for each pair of data and then sort these differences based on their absolute values. Assign ranks to the sorted differences, ignoring cases where the difference is zero. Calculate the sum of ranks for positive differences and negative differences separately. Based on the smallest of the sum of ranks for positive and negative differences, compute the Wilcoxon signed-rank test statistic. The larger the deviation of this statistic from zero, the more likely it is to reject the null hypothesis. The null hypothesis is

TABLE II: The information about 28 datasets used in the experiments.

Name	<i>#Samples</i>	<i>#Features</i>	<i>#Classes</i>
analcatadata_authorship	841	70	4
breast_cancer	569	5	2
cleveland	303	13	5
coil2000	9822	85	2
column_3C	310	6	3
contraceptive	1473	9	3
crx	653	15	2
diabetes	768	8	2
ecoli	336	7	8
flare	1066	11	6
haberman	306	3	2
heart-statlog	270	13	2
housevotes	232	16	2
HTRU_2	17898	8	3
ilpd	583	10	2
ionosphere	351	34	2
led7digit	500	7	10
liver-disorders	345	6	2
mammographic	830	5	2
occupancy	2665	5	2
optdigits	5620	64	2
ozone-level-8hr	2534	72	2
page-blocks	5473	10	5
Rice_OC	3810	7	2
spectfheart	267	44	2
tic-tac-toe	941	9	2
transfusion	748	4	2
vehicle	846	18	4

that there is no significant difference between the two sets of data. By comparing with a critical value or p-value, a decision is made on whether to reject the null hypothesis. The significance level is 0.05 here.

B. Comparisons with classic classifiers

We selected seven well-known classic classifiers for comparison with the ZVRB classification system, including K-Nearest Neighbors (KNN) [30], C45 decision tree [44], hybrid Decision Tree/Genetic Algorithm (DTGA) [45], Support Vector Machine (SVM) [32], RIPPER [41], MultiLayer Perceptron with Conjugate Gradient (MLPCG) [46] and XGBoost [31] classifiers. Each classifier underwent 10-CV experiments on the same datasets to ensure the reliability of the results. All classifiers were implemented on the KEEL software platform [47]. The parameters for all classifiers are outlined as follows.

C45:

- Confidence: 0.25,
- The minimum samples per leaf: 2.

DTGA:

- Confidence: 0.25,
- Threshold: 10.
- The minimum samples per leaf: 2,
- Genetic algorithm: GA-LARGE-SN,
- Number of total generations: 50,
- Number of chromosomes: 200,
- Crossover probability: 0.8,
- Mutation probability: 0.01.

KNN:

- K: 1,
- Distance function: Euclidean.

MLPCG:

- topologymlp: 10.

RIPPER:

- grow_pct: 0.66,
- K: 2.

SVM:

- Kernel type: radial basis function
- C: 100,
- Epsilon: 0.001,
- Degree: 1,
- Gamma: 0.01,
- Coef0: 0.0,
- nu: 0.1,
- p:1.0,
- Shrinking: 1.0.

XGBoost:

- eta:0.3,
- round=10.
- max_depth:6,

ZVRB:

- Number of optimization: 2,
- Number of iterations: 50,
- Number of folds: 3,
- Crossover probability: 0.5,
- Number of populations: 50,
- Mutation probability: 0.8.
- Optimization technique: DE,
- Maximum number of antecedents per rule: 4,
- The minimum samples covered per rule: 2.

Table III lists the classification accuracies of eight classifiers on 28 datasets, with the maximum accuracy for each dataset highlighted in bold. The blank space in the table indicates that the classifier is unable to process this dataset. The last row provides the average accuracy for each classifier across the 28 datasets. It is important to note that due to the stochastic nature of the DE optimization algorithm, the results of ZVRB are not identical in every run. Therefore, the presented results here are the averages of 10 runs. To clearly depict the distribution of accuracies for each classifier across different datasets, a boxplot for the eight classifiers are shown in Figure 2. Each box represents a classifier, with the upper and lower edges indicating the upper and lower quartiles, and the middle line representing the median. Small circles denote outliers.

TABLE III: Classification accuracy rates (%) of eight classifiers on 28 datasets.

Dataset	C45	DTGA	KNN	MLPCG	RIPPER	SVM	XGBoost	ZVRB
analcata_data_authorship	93.58	91.55	99.41	98.69	94.41	37.69	97.03	96.43
breast_cancer	91.04	91.75	92.09	94.20	88.74	84.35	92.09	92.01
cleveland	51.10		53.43	50.81	42.20	50.46	54.49	55.36
coi2000	93.97	94.01	89.67	92.79	92.79	90.91	93.85	94.03
column_3C	81.61	79.03	76.77	81.29	76.77	76.13	82.58	83.48
contraceptive	52.82	53.49	44.06	54.18	52.21	55.47	55.94	54.96
crx	86.22	85.45	81.62	82.54	82.84	54.67	86.36	87.00
diabetes	75.13	75.26	69.66	76.17	69.93	64.84	75.12	75.90
ecoli	79.22	79.21	78.30	77.50	72.01	61.31	82.43	83.17
flare	74.30	74.02	64.63	74.68	67.63	74.76	74.30	74.83
haberman	70.91	70.91	67.63	72.55	50.65	69.94	70.88	74.16
heart-statlog	75.19	77.41	75.56	70.74	71.11	59.63	78.52	81.22
housevotes	96.11	96.11	91.32	95.71	95.27	96.11	95.69	96.12
HTRU_2	97.94	97.86	97.04	9.16	97.74	96.44	97.93	97.97
ilpd	67.07	68.59	63.45	69.47	62.93	73.07	67.07	68.65
ionosphere	87.17	87.48	60.96	64.10	89.44	93.17	90.87	89.17
led7digit	71.60	69.40	40.60	70.60	49.00	74.40	73.00	72.84
liver-disorders	66.97	64.32	63.14	67.52	63.43	63.50	68.71	69.99
mammographic	84.10	83.86	73.98	81.57	77.71	84.34	81.57	83.86
occupancy	98.24	98.31	98.72	97.94	97.94	97.90	98.72	98.36
optdigits	98.08	97.88	89.82	89.82	97.92	93.31	98.51	98.15
ozone-level-8hr	92.30	91.28	92.19	92.42	90.69	93.69	94.28	94.34
page-blocks	96.93	96.46	95.69	96.29	96.55	92.60	97.46	97.14
Rice_OC	92.36	92.36	88.79	92.86	91.42	68.82	92.18	92.81
spectfheart	76.79	74.19	70.03	74.86	75.68	79.42	77.56	79.81
tic-tac-toe	84.55	77.98	73.07	88.62	97.08	81.63	98.12	98.75
transfusion	78.61	78.47	61.64	78.88	49.07	71.25	76.86	78.80
vehicle	74.69	71.74	70.10	82.62	68.67	52.95	74.93	72.84
average	81.74	82.16	75.83	77.81	77.21	74.74	83.11	83.65

From Figure 2, it can be observed that ZVRB has the highest median, upper and lower quartiles. Moreover, it achieved the highest classification accuracy in 14 out of the 28 datasets.

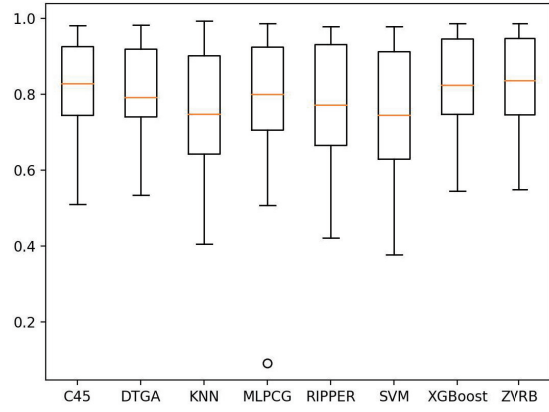


Fig. 2: The boxplot for the seven common classifiers and ZVRB classification system.

TABLE IV: The results of Wilcoxon signed-rank tests.

Classifier	Statistic value	p-value	Critical value	Hypothesis
C45	24	5.67734e-06	116	Rejected
DTGA	0	8.29809e-06	107	Rejected
KNN	11	4.09782e-07	116	Rejected
MLPCG	69	0.00155	116	Rejected
RIPPER	3	3.72529e-08	116	Rejected
SVM	44	0.00011	116	Rejected
XGBoost	104	0.02322	116	Rejected

Performing Wilcoxon signed-rank tests comparing the classification results of the seven classic classifiers individually with those of ZVRB, the test results are presented in Table IV. From the results, it can be observed that all null hypotheses were rejected, indicating statistically significant differences. In this scenario, it can be concluded that the classification accuracy of the ZVRB model is significantly superior to other classic classifiers.

C. Comparisons with rule-based Classifier

We selected seven well-known and state-of-the-art rule-based classification systems for comparison with the ZVRB classification system. They are Fuzzy Rule-Based (FRB) classification system proposed by Chi et al. [48], FURIA [27], Fuzzy Rule-based Decision Tree (FRDT) [49], Fuzzy Association Rule-based Classification method for High-Dimensional problems (FARC-HD) [50], Ensemble Belief Rule-Based model (En-BRB) [51], Z-number-valued Rule-Based (ZRB) classification system [33] and Z-number-valued Rule-based Decision Tree (ZRDT) [34]. Each model underwent 10-CV experiments on the same datasets. All models were implemented on the KEEL software platform [47]. The parameters for all models are outlined as follows.

FRB:

- Number of linguistic labels per feature: 3,
- T-norm: product,
- Rule weight: penalized certainty factor,
- Fuzzy reasoning method: winning rule.

FURIA:

- Number of optimization: 2,
- Number of folds: 3,
- The minimum samples covered per rule: 2.

FRDT:

- Maximum number of features: 5,
- Threshold: 0.6, • α : 0.02.

FARC-HD:

- Minimum support:0.05, • Parameter alpha: 0.15,
- Population size: 50, • Bits per gen:30.
- Number of linguistic Values: 5,
- Maximum confidence: 0.8,
- Maximum depth of the trees: 3,
- Parameter K of the prescreening: 2,
- Maximum number of evaluations: 1500.

En-BRB:

- Number of weak BRB: 20, • Mutation probability: 0.5,
- Iteration number: 200, • Cross probability: 0.8.
- Population size: 50,
- Optimization technique: DE,
- Number of reference values: 4,
- Number of antecedent attributes for each weak BRB: 2.

ZRB:

- Number of labels per feature: 3,
- T-norm: product,
- Rule weight: penalized certainty factor,
- Fuzzy reasoning method: winning rule.

ZRDT:

- Maximum number of features: 5,
- Threshold: 0.6.

ZVRB:

- Number of optimization: 2, • Number of iterations: 50,
- Number of folds: 3, • Crossover probability: 0.5,
- Number of populations: 50, • Mutation probability: 0.8.
- Optimization technique: DE,
- Maximum number of antecedents per rule: 4,
- The minimum samples covered per rule: 2.

TABLE V: Classification accuracy rates (%) of eight classifiers on 28 datasets.

Dataset	FRB	FURIA	FRDT	FARC-HD	En-BRB	ZRB	ZRDT	ZVRB
analcata_data_authorship	72.07	96.31	87.99	94.89	78.47	72.07	95.02	96.43
breast_cancer	89.63	91.74	90.34	92.10	90.87	90.34	89.64	92.01
cleveland		56.40						55.36
coil2000	75.63	93.93	93.76	94.03	94.03	75.45	86.55	94.03
column_3C	59.03	82.58	81.29	78.39	77.74	67.42	80.00	83.48
contraceptive	39.31	54.72	40.94	53.09	46.71	41.21	47.39	54.96
crx	81.62	86.82	84.37	86.68	78.24	81.17	86.21	87.00
diabetes	72.79	75.14	73.58	74.35	71.23	75.14	72.66	75.90
ecoli	72.09	82.18	63.71	80.72	64.06	77.75	64.28	83.17
flare	38.46	74.77	44.47	74.67	44.84	45.22	64.17	74.83
haberman	73.87	72.23	62.56	73.54	73.19	73.53	73.20	74.16
heart-statlog	52.59	80.37	83.70	84.07	79.26	52.59	81.48	81.22
housevotes	41.36	96.54	96.97	95.25	88.77	41.36	96.97	96.12
HTRU_2	95.68	97.87	97.37	97.86	96.82	96.74	96.17	97.97
ilpd	70.32	68.96	55.39	68.59	71.19	69.81	60.68	68.65
ionosphere	64.65	89.75	86.33	92.30	83.76	64.94	88.89	89.17
led7digit	63.20	72.80	23.00	70.60	49.00	63.20	31.80	72.84
liver-disorders	58.54	65.81	63.50	66.35	64.66	60.01	60.62	69.99
mammographic	79.64	83.37	83.01	80.24	82.05	79.40	84.58	83.86
occupancy	93.09	98.27	97.86	97.67	92.87	96.17	97.90	98.36
optdigits	45.34	99.09	58.47	98.19	89.82	45.34	95.16	98.15
ozone-level-8hr	92.54	93.84	70.52	93.88	93.69	92.54	81.61	94.34
page-blocks	91.91	97.42	84.16	94.41	90.32	92.23	90.19	97.14
Rice_OC	88.19	92.57	92.62	92.62	92.26	91.52	92.49	92.81
spectfheart	66.34	78.69	70.33	79.77	79.42	66.34	71.10	79.81
tic-tac-toe	0.00	98.54	70.78	92.70	66.18	0.00	70.67	98.75
transfusion	76.61	78.74	71.52	77.41	77.01	76.61	63.63	78.80
vehicle	60.77	70.20	48.71	67.49	55.58	60.77	57.94	72.84
Average	67.23	83.20	73.23	83.40	76.74	68.48	77.07	83.65

Table V displays the classification accuracies of the eight classifiers on 28 datasets with the highest accuracy for each dataset bolded. The blank spaces in the table indicate that the classifier is unable to process this dataset. The last row presents the average accuracy for each classifier across the 28 datasets. It should be noted that the results for ZVRB presented here are the average accuracy over 10 runs. To clearly depict the distribution of accuracies for each classifier across different datasets, a boxplot for the eight classifiers are shown in Figure 3. From Figure 3, it can be observed that the

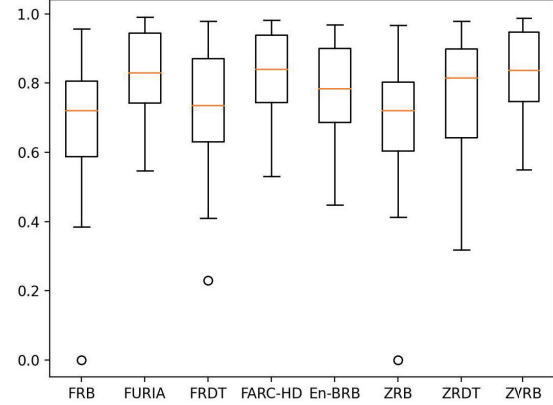


Fig. 3: The boxplot for the seven fuzzy rule-based classifiers and ZVRB classification system.

median, upper quartile, and lower quartile of ZVRB are all the highest. Additionally, out of the 28 datasets, ZVRB achieves the highest classification accuracy in 18 datasets. Wilcoxon signed-rank tests were conducted to compare the classification results of the seven rule-based classifiers individually with those of ZVRB, and the results are shown in Table VI. From the results, all null hypotheses were rejected, indicating a statistically significant difference. In this scenario, it can be concluded that the classification accuracy of the ZVRB model is significantly superior to other rule-based classifiers.

TABLE VI: The results of Wilcoxon signed-rank tests.

Classifier	Statistic value	p-value	Critical value	Hypothesis
FRB	2	4.47034e-08	107	Rejected
FURIA	106	0.02635	116	Rejected
FRDT	16	2.51829e-06	107	Rejected
FARC-HD	46	0.00101	107	Rejected
Ensemble-BRB	10	2.62975e-05	107	Rejected
ZRB	3	7.45058e-08	107	Rejected
ZRDT	13	1.31130e-06	107	Rejected

VI. CONCLUSIONS

Z-numbers possess powerful capabilities in representing and handling uncertain and partially reliable information. The underlying probability distributions are a crucial component of the definition of Z-numbers and serves as the link between two fuzzy numbers. This paper discusses the reverse relationship between Z-numbers and probability distributions, namely, summarizing a set of probability distributions using a

Z-number. To fill this gap, a nonlinear model called MEME is proposed. Through this model, Z-numbers can be generated from data without any expert knowledge. Therefore, Z-numbers can be employed to address classification problems, introducing a new if-then rule form called Z-valuation if-then rules. These rules replace the deterministic consequent part of a fuzzy rule with an uncertain Z-valuation, allowing for the further summarization of uncertain information in the consequent part of a rule. Based on the Z-valuation rules, a Z-valuation rule-based classification system is proposed. This system is entirely data-driven. To validate the effectiveness of the ZVRB classification system, two experiments were conducted comparing the ZVRB model with some well-known and state-of-the-art classic classifiers and rule-based classification systems. The results demonstrate that the ZVRB model is significantly superior to other classic classifiers and rule-based classification systems.

The ZVRB model relies on FURIA to generate the antecedents of rules, then its performance is subject to some limitations of the FURIA. In future work, we plan to further investigate the Z-valuation antecedents to replace the fuzzy antecedents.

Java implementation of the ZVRB classification system, running under the KEEL software platform, can be downloaded at <https://github.com/LiYangxue0410/ZVRB.git>.

REFERENCES

- [1] W. Feller, *An introduction to probability theory and its applications*. John Wiley & Sons, 2008, vol. 2.
- [2] L. A. Zadeh, "Fuzzy sets," in *Fuzzy Sets, Fuzzy Logic, And Fuzzy Systems: Selected Papers by Lotfi A Zadeh*. World Scientific, 1996, pp. 394–432.
- [3] Q. Zhou, W. Pedrycz, Y. Liang, and Y. Deng, "Information granule-based uncertainty measure of fuzzy evidential distribution," *IEEE Transactions on Fuzzy Systems*, p. 10.1109/TFUZZ.2023.3284713, 2023.
- [4] Z.-W. Zhang, Z.-G. Liu, A. Martin, and K. Zhou, "Bsc: Belief shift clustering," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 3, pp. 1748–1760, 2022.
- [5] Q. Zhou, É. Bossé, and Y. Deng, "Modeling belief propensity degree: Measures of evenness and diversity of belief functions," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 5, pp. 2851–2862, 2023.
- [6] Y. Deng, "Random permutation set," *International Journal of Computers Communications & Control*, vol. 17, no. 1, p. 4542, 2022.
- [7] L. A. Zadeh, "A note on z-numbers," *Information Sciences*, vol. 181, no. 14, pp. 2923–2932, 2011.
- [8] N. M. F. H. N. B. Alam, K. M. N. Ku Khalif, N. I. Jaini, and A. Gegov, "The application of z-numbers in fuzzy decision making: the state of the art," *Information*, vol. 14, no. 7, p. 400, 2023.
- [9] K.-W. Shen, J.-Q. Wang, and T.-L. Wang, "The arithmetic of multidimensional z-number," *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 2, pp. 1647–1661, 2019.
- [10] I. M. Jiskani, F. Yasli, S. Hosseini, A. U. Rehman, and S. Uddin, "Improved z-number based fuzzy fault tree approach to analyze health and safety risks in surface mines," *Resources Policy*, vol. 76, p. 102591, 2022.
- [11] B. Shang, Z. Chen, Q. Ma, and Y. Tan, "A comprehensive mortise and tenon structure selection method based on pugh's controlled convergence and rough z-number mabac method," *Plos one*, vol. 18, no. 5, p. e0283704, 2023.
- [12] N. Dhiman, M. M. Gupta, D. P. Singh, V. N. Mishra, and M. K. Sharma, "On z-intuitionistic fuzzy fractional valuations for medical diagnosis: An intuitionistic fuzzy knowledge-based expert system," *Fractal and Fractional*, vol. 6, no. 3, p. 151, 2022.
- [13] R. A. Aliev, W. Pedrycz, B. Guirimov, and O. H. Huseynov, "Clustering method for production of z-number based if-then rules," *Information Sciences*, vol. 520, pp. 155–176, 2020.
- [14] R. A. Aliev, W. Pedrycz, B. G. Guirimov, and O. H. Huseynov, "Acquisition of z-number-valued clusters by using a new compound function," *IEEE Transactions on Fuzzy Systems*, vol. 31, no. 1, pp. 279–286, 2020.
- [15] R. A. Aliev, A. V. Alizadeh, and O. H. Huseynov, "The arithmetic of discrete z-numbers," *Information Sciences*, vol. 290, pp. 134–155, 2015.
- [16] R. A. Aliev, O. H. Huseynov, and L. M. Zeinalova, "The arithmetic of continuous z-numbers," *Information Sciences*, vol. 373, pp. 441–460, 2016.
- [17] Q. Liu, H. Cui, Y. Tian, and B. Kang, "On the negation of discrete z-numbers," *Information Sciences*, vol. 537, pp. 18–29, 2020.
- [18] Y. Li, D. Pelusi, Y. Deng, and K. H. Cheong, "Relative entropy of z-numbers," *Information Sciences*, vol. 581, pp. 1–17, 2021.
- [19] Y. Yang and J.-q. Wang, "Smaa-based model for decision aiding using regret theory in discrete z-number context," *Applied Soft Computing*, vol. 65, pp. 590–602, 2018.
- [20] Q. Jia and E. Herrera-Viedma, "Pythagorean fuzzy sets to solve z-numbers in decision-making model," *IEEE Transactions on Fuzzy Systems*, vol. 31, no. 3, pp. 890–904, 2022.
- [21] J. Chu and X. Xiao, "Benefits evaluation of the northeast passage based on grey relational degree of discrete z-numbers," *Information Sciences*, vol. 626, pp. 607–625, 2023.
- [22] D. Qiu, Y. Xing, and R. Dong, "On ranking of continuous z-numbers with generalized centroids and optimization problems based on z-numbers," *International Journal of Intelligent Systems*, vol. 33, no. 1, pp. 3–14, 2018.
- [23] Y. Li, E. Herrera-Viedma, I. J. Pérez, W. Xing, and J. A. Morente-Molinera, "The arithmetic of triangular z-numbers with reduced calculation complexity using an extension of triangular distribution," *Information Sciences*, vol. 647, p. 119477, 2023.
- [24] B. Kang, Y. Deng, K. Hewage, and R. Sadiq, "Generating z-number based on owa weights using maximum entropy," *International Journal of Intelligent Systems*, vol. 33, no. 8, pp. 1745–1755, 2018.
- [25] Y. Tian and B. Kang, "A modified method of generating z-number based on owa weights and maximum entropy," *Soft Computing*, vol. 24, no. 20, pp. 15 841–15 852, 2020.
- [26] F. Bilgin and M. Alci, "Generating z-number by logistic regression," in *2021 7th International Symposium on Electrical and Electronics Engineering (ISEEE)*. IEEE, 2021, pp. 1–6.
- [27] J. Hühn and E. Hüllermeier, "Furia: an algorithm for unordered fuzzy rule induction," *Data Mining and Knowledge Discovery*, vol. 19, pp. 293–319, 2009.
- [28] X. Yuan, M. J. Liebelt, P. Shi, and B. J. Phillips, "Cognitive decisions based on a rule-based fuzzy system," *Information Sciences*, vol. 600, pp. 323–341, 2022.
- [29] F. Aghaeipoor, M. Sabokrou, and A. Fernández, "Fuzzy rule-based explainer systems for deep neural networks: From local explainability to global understanding," *IEEE Transactions on Fuzzy Systems*, vol. 31, no. 9, pp. 3069–3080, 2023.
- [30] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.
- [31] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.
- [32] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, pp. 273–297, 1995.
- [33] Y. Li, E. Herrera-Viedma, I. J. Pérez, M. Barragán-Guzmán, and J. A. Morente-Molinera, "Z-number-valued rule-based classification system," *Applied Soft Computing*, vol. 137, p. 110168, 2023.
- [34] —, "Z-number-valued rule-based classification system," *Applied Soft Computing*, vol. 137, p. 110168, 2023.
- [35] L. A. Zadeh, "Fuzzy sets as a basis for a theory of possibility," *Fuzzy sets and systems*, vol. 1, no. 1, pp. 3–28, 1978.
- [36] M. Higashi and G. J. Klir, "Measures of uncertainty and information based on possibility distributions," *International journal of general systems*, vol. 9, no. 1, pp. 43–58, 1982.
- [37] B. Liu and Y.-K. Liu, "Expected value of fuzzy variable and fuzzy expected value models," *IEEE transactions on Fuzzy Systems*, vol. 10, no. 4, pp. 445–450, 2002.
- [38] L. A. Zadeh, "Probability measures of fuzzy events," *Journal of mathematical analysis and applications*, vol. 23, no. 2, pp. 421–427, 1968.
- [39] Z.-J. Zhou, G.-Y. Hu, C.-H. Hu, C.-L. Wen, and L.-L. Chang, "A survey of belief rule-base expert system," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 8, pp. 4944–4958, 2019.
- [40] X. Hu, Y. Shen, W. Pedrycz, X. Wang, A. Gacek, and B. Liu, "Identification of fuzzy rule-based models with collaborative fuzzy clustering," *IEEE transactions on cybernetics*, vol. 52, no. 7, pp. 6406–6419, 2021.

- [41] W. W. Cohen, "Fast effective rule induction," in *Machine learning proceedings 1995*. Elsevier, 1995, pp. 115–123.
- [42] S. Das and P. N. Suganthan, "Differential evolution: A survey of the state-of-the-art," *IEEE transactions on evolutionary computation*, vol. 15, no. 1, pp. 4–31, 2010.
- [43] K. Shen, X. Wang, D. Qiao, and J. Wang, "Extended z-mabac method based on regret theory and directed distance for regional circular economy development program selection with z-information," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 8, pp. 1851–1863, 2020.
- [44] J. R. Quinlan, *C4. 5: programs for machine learning*. Elsevier, 2014.
- [45] D. R. Carvalho and A. A. Freitas, "A hybrid decision tree/genetic algorithm method for data mining," *Information Sciences*, vol. 163, no. 1–3, pp. 13–35, 2004.
- [46] M. F. Møller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural networks*, vol. 6, no. 4, pp. 525–533, 1993.
- [47] J. Alcalá-Fdez, L. Sanchez, S. Garcia, M. J. del Jesus, S. Ventura, J. M. Garrell, J. Otero, C. Romero, J. Bacardit, V. M. Rivas *et al.*, "Keel: a software tool to assess evolutionary algorithms for data mining problems," *Soft Computing*, vol. 13, no. 3, pp. 307–318, 2009.
- [48] Z. Chi, H. Yan, and T. Pham, *Fuzzy algorithms: with applications to image processing and pattern recognition*. World Scientific, 1996, vol. 10.
- [49] X. Wang, X. Liu, W. Pedrycz, and L. Zhang, "Fuzzy rule based decision trees," *Pattern Recognition*, vol. 48, no. 1, pp. 50–59, 2015.
- [50] J. Alcalá-Fdez, R. Alcalá, and F. Herrera, "A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning," *IEEE Transactions on Fuzzy systems*, vol. 19, no. 5, pp. 857–872, 2011.
- [51] Y. You, J. Sun, Y.-w. Chen, C. Niu, and J. Jiang, "Ensemble belief rule-based model for complex system classification and prediction," *Expert Systems with Applications*, vol. 164, p. 113952, 2021.



ences at the University of Granada (Spain). Master studies during 2022/2023 the Master's Degree in Mathematics. From 2019 to the present PhD student.

José Ramón Trillo PhD student of Intelligent Decision-Making Systems within the PhD Programme in Information and Communication Technologies at the University of Granada (Spain). Undergraduate studies from 2012 to 2018 in Mathematics at the University of Granada (Spain) and undergraduate studies in Computer Sciences from 2012 to 2018 at the University of Granada (Spain). During 2017/2018, he studied at the University of Salerno (Italy). Master studies during the academic year 2018/2019 the Master's Degree in Computer Sciences at the University of Granada (Spain). Master studies during 2022/2023 the Master's Degree in Mathematics. From 2019 to the present PhD student.



Yangxue Li received a bachelor's degree from the School of Network Security, Chengdu University of Technology, in 2018 and a master's degree from the Institute of Fundamental and Frontier Science, University of Electronic Science and Technology of China, in 2021. She is currently pursuing a Ph.D. degree in Department of Computer Science and Artificial Intelligence, School of Computer and Telecommunication Engineering, University of Granada, Spain. Li's research interests include Dempster-Shafer evidence theory, fuzzy sets

theory and Z-numbers.



Juan Antonio Morente-Molinera was born in Granada, Spain, in 1986. He received the Ph.D. degree in information and communications technology from the University of Granada (UGR), Granada, Spain, in 2015 and the UGR best thesis prize in 2019. He is currently associate professor in the Department of Computer Science and Artificial Intelligence, on the University of Granada. He is a member of the Soft Computing and Intelligent Information Systems (SCI2S) Research Group. His h-index is 21 and he has published more than 40 papers in

JCR journals. He is working in the decision making field trying to solve the challenges that the new era of ubiquitous computing brings to the decision support systems field. His research interests include group decision making, decision support systems, fuzzy ontologies, consensus models, linguistic modeling, and sentiment analysis.



Enrique Herrera-Viedma (Fellow, IEEE) is Full Professor in Computer Science and AI at University of Granada (UGR) from 2007. Now, he is Vice-President for Research and Knowledge Transfer at the UGR from June 2015. He has been Vice-President for Publications in the IEEE System, Man and Cybernetic Society (2019-2020) and currently he is the new VP for Cybernetics till 2025.

He received the UGR Young Researcher Career Award in 2006, several UGR excellence research paper awards (2003, 2010, 2019 y 2020), the 2011

IEEE Transactions on Fuzzy Systems Outstanding Paper Award from the IEEE Computational Intelligence Society, the 2016 Andrew P. Sage Best Transactions Paper Award, which recognizes authors of the best paper among all papers published in the IEEE Transactions on Systems, Man, and Cybernetics: Systems, and the 2014-Best Paper HERBERT SIMON AWARD of journal Int. Journal of Information Technology and Decision Making.

He is a Fellow of the Institute of Electrical and Electronic Engineers and also of International Fuzzy Systems Association for contributions to fuzzy decision making systems and linguistic modelling in 2021. He is also a member of the prestigious Spanish Think Tank "Royal Institute Elcano" supported by the large Spanish enterprises and the Spain Government.

He has published more than 560 peer-reviewed scientific publications, including 350 papers in JCR journals (More than 200 in Q1, more than 50 in D1). He has been identified as Highly Cited Researcher in both categories, Computer Science and Engineering by Clarivate Analytics, in the years 2014, 2015, 2016, 2017, 2018, 2019, only in Computer Science in 2020. According to Web of Science his h-index is 85 (more than 23000 citations received) and according to Google Scholar his h-index is 101 (more than 41000 citations received). Currently, he has 50 Highly cited papers indexed in Web of Science.

5 Z-mixture-number if-then rules induction algorithm for classification tasks

- Y. Li, I. J. Pérez, F. J. Cabrerizo, and J. A. Morente-Molinera, Z-mixture-number if-then rules induction algorithm for classification tasks. Submitted to IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2024
 - Status: **Submitted**.
 - Impact Factor (JCR 2020): 9.309
 - Subject Category: Computer Science, Human-Computer Interaction, Ranking 3/89 (Q1).
 - Subject Category: Engineering, Control and Systems Engineering, Ranking 8/232 (Q1).

Z-mixture-number if-then rule induction algorithm for classification tasks

Yangxue Li, Ignacio Javier Pérez, Francisco Javier Cabrerizo, Juan Antonio Morente-Molinera

Abstract—Fuzzy techniques are commonly employed methods for addressing classification problems. Various fuzzy classification algorithms have been proposed. Among them, fuzzy rule-based classifiers exhibit stronger interpretability and traceability. Recently, an extension of fuzzy rule-based classifier, called Z-number-valued rule-based classifier, has been proposed to handle uncertainty and partial reliability in classification problems. However, existing Z-number-valued rules not only deviate from the definition of Z-numbers but also fail to handle cases where both continuity and discreteness are present simultaneously. This paper introduces a novel Z-number-valued rule-based classification system called Z-Mixture-number if-then Rule Induction Algorithm (ZMRIA). It formulates Z-mixture-number if-then rules utilizing Z-mixture-numbers. By using the feature values of each sample as the underlying probability function, ZMRIA learns each Z-mixture-number antecedent of the rules in order. It avoids the problem of combinatorial rule explosion by maximizing the positive-negative Z-mixture-numbers difference to select features. Through comparative experiments on classification accuracy and runtime with some state-of-the-art methods across 46 real-world datasets, the test results demonstrate that ZMRIA outperforms other methods in terms of classification accuracy and is faster in runtime compared to most classifiers.

Index Terms—Z-number, fuzzy number, Z-mixture-number, rule-based system, pattern classification, probability distribution

I. INTRODUCTION

In the contemporary information age, there is a growing trend towards increasing complexity and uncertainty in data. As one of the most effective ways to deal with uncertain information, fuzzy techniques have been extensively and deeply researched in numerous domains, such as pattern recognition [1], image processing [2], decision support [3], expert systems [4] and control [5]. Especially for classification problems, fuzzy techniques, employing tools such as fuzzy set theory, fuzzy logic, and fuzzy inference, enable better handling of ambiguity and uncertainty inherent in data, thereby enhancing the accuracy and robustness [6], [7].

There exist various recent fuzzy classification algorithms that have been proposed. Fuzzy decision tree is a decision

tree model based on fuzzy logic and fuzzy set theory, allowing decision rules to possess fuzziness [8], [9]. Fuzzy support vector machine is an extended model of support vector machine that utilizes fuzzy set theory to map data into high-dimensional space and maximizes the margin between fuzzy sets for classification [10]. Fuzzy neural network is a classification method that integrates fuzzy set theory with neural network technology [11], [12]. It leverages the learning capability of neural networks and the representational power of fuzzy sets to perform fuzzy classification on data. Fuzzy rule-based classifier is a classification method based on fuzzy logic [13]–[15]. It utilizes a series of fuzzy rules to describe the relationship between features and classes and employs fuzzy inference for classification purposes [16], [17]. In comparison to other methods, fuzzy rule-based classifiers typically exhibit stronger interpretability and traceability. This is because fuzzy rule-based classifiers can facilitate a clearer interpretation and understanding of the model's output results, aiding users in gaining insights into the classification process. Moreover, in fuzzy rule-based classifiers, due to the formalized description of fuzzy rules, they can provide clear reasoning paths, enabling accurate traceability and interpretation of the model's decision-making process.

To enhance the accuracy and interpretability of classification, some extensions of fuzzy rule-based classifiers have been proposed. Yang et al. extended fuzzy rules using the Dempster-Shafer evidence theory, introducing belief rule structures [18]. This approach is capable of capturing fuzzy, imprecise, and incomplete causal relationships. Recently, Li et al. expanded fuzzy rules using Z-number theory, introducing Z-number-valued rule-based rules for classification tasks. This framework can handle not only the fuzziness and uncertainty inherent in problems but also the partial reliability within them [19], [20]. The concept of Z-numbers is a more comprehensive concept for describing real-world information compared to fuzzy numbers [21]. Z-numbers can simultaneously describe both the uncertainty and partial reliability of information. A Z-number is an ordered pair of fuzzy number, where the first number is a fuzzy constraint on values that the variable can take, and the second fuzzy number measures the reliability of the first one [22]. The two fuzzy numbers in Z-numbers are not individual entities; rather, they are connected through underlying probability functions within the mathematical model of a Z-number. However, in the existing Z-number-valued rules, the second fuzzy number of the Z-number is used to adjust the first one, which deviates from the definition of Z-numbers. Moreover, due to limitations in the definitions of continuous and discrete Z-numbers, existing methods cannot

This work was supported by the China Scholarship Council (CSC), and by the grant PID2022-139297OB-I00 funded by MICIU / AEI / 10.13039/501100011033 and by ERDF/EU, and by the project C-ING-165-UGR23, co-funded by the Regional Ministry of University, Research and Innovation and by the European Union under the Andalusia ERDF Programme 2021-2027.

The authors are with the Department of Computer Science and Artificial Intelligence, Andalusian Research Institute in Data Science and Computational Intelligence, DaSCI, University of Granada, 18071 Granada, (e-mail: liyangxueugr@gmail.com, jamoren@decsai.ugr.es).

Corresponding author: Juan Antonio Morente-Molinera

handle situations where both continuity and discreteness are present simultaneously.

This paper aims to introduce a novel Z-number-valued rule-based classification system, namely Z-Mixture-number if-then Rule Induction Algorithm, or ZMRIA for short. Xian et al. proposed the concept of Z-mixture-numbers to general continuous Z-numbers, discrete Z-numbers, and cases where the two fuzzy numbers in a Z-number are of different types, either discrete or continuous [23]. We utilize Z-mixture-numbers as antecedents for formulating Z-mixture-number if-then rules. Their first fuzzy numbers are trapezoidal fuzzy numbers and the second ones are discrete fuzzy numbers. Subsequently, its induction algorithm is designed. The consequent part of the rules is initially determined, consisting of each class label. To learn each Z-mixture-number antecedent of the rule, we employ the feature values of each sample as underlying probability functions. For each class label, we generate a Z-mixture-number. The generation method is fully consistent with the definition of Z-numbers. If this class is the consequent class for the rule, the Z-mixture-number is referred to as a positive Z-mixture-number; otherwise, it is termed a negative Z-mixture-number. To obtain the first fuzzy number, we maximize the difference between positive and negative Z-mixture-numbers, named the positive-negative Z-mixture-numbers difference, or positive-negative difference. Moreover, to avoid the problem of combinatorial rule explosion, we maximize the positive-negative difference to select features.

To assess the performance of ZMRIA, we conducted comparative experiments on classification accuracy and running time with some state-of-the-art methods across 46 real-world datasets, including three benchmark classifiers and seven rule-based classifiers. To provide clear and robust comparative results, we employed a non-parametric statistical test for pairwise compare the classification performance of classifiers. The test results indicate that ZMRIA outperforms the other ten methods in terms of classification accuracy. On the other hand, in terms of runtime, ZMRIA exhibits shorter running time compared to seven out of the ten classifiers.

In summary, the main contributions of this paper can be summarized as follows:

- (1) Introduced the concept of Z-mixture-numbers in if-then rules to handle the cases where the two fuzzy number types are different.
- (2) Utilized the feature values of each sample as underlying probability functions to generate Z-mixture-numbers for each class label, and introduced the concepts of positive and negative Z-mixture-numbers. Additionally, designed a method to compute the difference between them.
- (3) Proposed a novel and more efficient method for solving classification tasks, namely the Z-mixture-number if-then rule induction algorithm.
- (4) Conducted comprehensive performance evaluations on 46 real-world datasets, comparing classification accuracy and running time with some state-of-the-art methods, and validated the results using a non-parametric statistical test.

This paper is arranged as follows. Section II provides a brief review of background knowledge pertaining to fuzzy sets, Z-numbers, and fuzzy rules. Section III describes in detail

the proposed ZMRIA and provides corresponding algorithms. Section IV shows and discusses the results that are obtained on 46 real-world datasets. Finally, conclusions are drawn in Section V.

II. PRELIMINARIES

In this section, a brief review of background knowledge pertaining to fuzzy sets, Z-numbers, and fuzzy rules is provided.

A. Fuzzy number

Fuzzy sets describe the membership relationship between elements and sets through the utilization of membership functions [24]. The membership functions can be either continuous or discrete.

Definition 1 (Discrete fuzzy number): Let U be a classical set of objects, called universe, whose generic elements are denoted as u . Membership in a fuzzy subset $\tilde{A}$ of U is viewed as a characteristic function $\mu_{\tilde{A}}$ from U to interval $[0, 1]$. $\tilde{A}$ is a discrete fuzzy number if its support is finite, i.e. there exist $u_1, \dots, u_n \in \mathcal{R}$ with $u_1 \leq u_2 \leq \dots \leq u_n$, such that $\text{supp}(\tilde{A}) = \{u_1, \dots, u_n\}$ and there exist natural numbers s, t with $1 \leq s \leq t \leq n$ satisfying the following conditions.

- (1) $\mu_{\tilde{A}}(u_i) = 1$ for any natural number i with $s \leq i \leq t$;
- (2) $\mu_{\tilde{A}}(u_i) \leq \mu_{\tilde{A}}(u_j)$ for each natural number i, j with $1 \leq i \leq j \leq s$;
- (3) $\mu_{\tilde{A}}(u_i) \geq \mu_{\tilde{A}}(u_j)$ for each natural number i, j with $t \leq i \leq j \leq n$.

Definition 2 (Continuous fuzzy number): A continuous fuzzy number is a fuzzy set $\tilde{A}$ on $\mathcal{R}$ which possesses the following properties [24]:

- (1) $\max_{u \in U} (\mu_{\tilde{A}}(u)) = 1$;
- (2) $\mu_{\tilde{A}}(\lambda u_1 + (1-\lambda)u_2) \geq \min(\mu_{\tilde{A}}(u_1), \mu_{\tilde{A}}(u_2))$, $u_1, u_2 \in U$, $\lambda \in [0, 1]$;
- (3) The support of $\tilde{A}$, $\text{supp}(\tilde{A})$, is bounded.

The trapezoidal fuzzy number is a specific form of continuous fuzzy number characterized by a shape resembling a trapezoid.

Definition 3 (Trapezoidal fuzzy number): A trapezoidal fuzzy number can be represented by a quadruple $\tilde{A} = (\underline{a}, a, b, \bar{b})$, $\underline{a} \leq a \leq b \leq \bar{b}$, with membership function $\mu_{\tilde{A}}(u)$,

$$\mu_{\tilde{A}}(u) = \begin{cases} \frac{u - \underline{a}}{a - \underline{a}}, & \underline{a} \leq u \leq a, \\ 1, & a \leq u \leq b, \\ \frac{\bar{b} - u}{\bar{b} - b}, & b \leq u \leq \bar{b}. \end{cases} \quad (1)$$

B. Z-number and Z-mixture-number

Zadeh proposed that real-world information characterizes not only uncertainty but also partial reliability [22]. Therefore, he introduced the notion of Z-numbers to simultaneously describe the uncertainty and partial reliability of information [22]. A Z-number consists of two ordered fuzzy numbers, which means Z-numbers can be either continuous or discrete.

Definition 4 (Z-number): A Z-number is an ordered pair of fuzzy numbers, $Z = (\tilde{A}, \tilde{B})$, where $\tilde{A}$ is a fuzzy number

playing a role of fuzzy constraint on values that a random variable X can take,

$$X \text{ is } \tilde{A},$$

and $\tilde{B}$ is a fuzzy number with a membership function $\mu_{\tilde{B}} : [0, 1] \rightarrow [0, 1]$, playing a role of a fuzzy constraint on the probability measure of $\tilde{A}$,

$$P(\tilde{A}) \text{ is } \tilde{B},$$

where $P(\tilde{A}) = \sum_{i=1}^n \mu_{\tilde{A}}(u_i) p(u_i)$ if the Z-number is discrete; If it is continuous, $P(\tilde{A}) = \int_{\mathcal{R}} \mu_{\tilde{A}}(u) p(u) du$. $p(\cdot)$ is the underlying probability distribution on the random variable X satisfying $P(\tilde{A}) \in \text{supp}(\tilde{A})$.

The two components of a discrete or continuous Z-number are either both discrete or both continuous. However, in practical problems, situations where both continuous and discrete attributes coexist are not uncommon. In order to generate various types of Z-numbers, Xian et al. introduced the concept of Z-mixture-numbers [23].

Definition 5 (Z-mixture-number): Given a Z-number $Z = (\tilde{A}, \tilde{B})$, where $\tilde{A}$ and $\tilde{B}$ can be described using either continuous fuzzy numbers or discrete fuzzy numbers, then Z can be referred to as a Z-mixture-number.

According to the different types of the two components of a Z-mixture-number, there are four types of Z-mixture-numbers: both $\tilde{A}$ and $\tilde{B}$ are discrete, both $\tilde{A}$ and $\tilde{B}$ are continuous, $\tilde{A}$ is discrete and $\tilde{B}$ is continuous, $\tilde{A}$ is continuous and $\tilde{B}$ is discrete.

C. Fuzzy rule-based system

Fuzzy rule-based systems are designed to process vague and uncertain information and make decisions based on predefined fuzzy rules and fuzzy logic. The fuzzy rule-based system operates based on a set of fuzzy rules that define how the system should behave or make decisions. These fuzzy rules typically take the form of "if-then" statements, where specific conditions trigger particular actions or outcomes. A fuzzy rule-based system mainly consists of two components: a knowledge base and an inference engine. The knowledge base stores fuzzy rules, while the inference engine is responsible for applying these fuzzy rules to the data or inputs provided by the system. Fuzzy rule-based systems offer numerous advantages, such as transparency and ease of understanding, high modularity and robustness, and strong scalability. They find wide applications in fields like expert systems [25], fuzzy control [26], pattern recognition [13], data mining [27] and beyond. Especially for the classification problems, fuzz rule-based systems enhance classification accuracy and interpretability while addressing the inherent fuzziness and uncertainty present in practical scenarios.

Let's consider a classification problem. Suppose there is a dataset $D = \{(x_1, y_1), \dots, (x_n, y_n)\}$ containing n samples, and m features, $F = \{f_1, \dots, f_m\}$, where $x_i = \{x_{i1}, \dots, x_{im}\}$ represents the feature vector of the i th sample, x_{ij} is the feature value of feature f_j , and y_i denotes the class label of the i th sample. Our objective is to predict the corresponding class label y_i based on the given feature vector x_i . Suppose there are K different class labels, denoted as the set of class labels

$\Omega = \{\omega_1, \dots, \omega_K\}$. A typical fuzzy rule in a fuzzy rule-based system for classification tasks can be written as follows.

$$\text{Rule } R_l : \text{If } x_{i1} \text{ is } \tilde{A}_{l1} \text{ and } \dots \text{ and } x_{im} \text{ is } \tilde{A}_{lm} \text{ then } y_i \text{ is } C_l. \quad (2)$$

Where $x_i = \{x_{i1}, \dots, x_{im}\}$ is the input variable; $\tilde{A}_{l1}, \dots, \tilde{A}_{lm}$ are the antecedent fuzzy numbers associated with the input variables; y_i is the output variable; C_l is the consequent class label, $C_l \in \Omega$.

In addition to typical fuzzy rules, some extensions of fuzzy rules, such as belief rules and Z-number-valued rules, have been proposed to improve classification accuracy and interpretability. Belief rules are an extension of fuzzy rules that combines fuzzy logic and the Dempster-Shafer evidence theory [18], [28]. The rule structure of a belief rule is as follows.

$$\text{Rule } \hat{R}_l : \text{If } x_{i1} \text{ is } \alpha_{l1} \text{ and } \dots \text{ and } x_{im} \text{ is } \alpha_{lm} \text{ then } y_i \text{ is } \{(\omega_1, \beta_{l,1}), \dots, (\omega_K, \beta_{l,K})\}. \quad (3)$$

Where $x_i = \{x_{i1}, \dots, x_{im}\}$ is the input variable; $\alpha_{l1}, \dots, \alpha_{lm}$ are referential values; y_i is the output variable; $\beta_{l,k}$ is the belief degree of class label ω_k satisfying $\beta_{l,k} \in [0, 1]$ and $\sum_{k=1}^K \beta_{l,k} \leq 1$. The referential values can be assigned either as a fuzzy membership function or a numerical value through a utility-based equivalence transformation technique and optimization methods [29], [30].

Z-number-valued rules are another extension of fuzzy rules, where the antecedents are in the form of Z-numbers [19]. A Z-number-valued rule can be written as follows.

$$\text{Rule } R_l^Z : \text{If } x_{i1} \text{ is } Z_{l1} \text{ and } \dots \text{ and } x_{im} \text{ is } Z_{lm} \text{ then } y_i \text{ is } C_l. \quad (4)$$

Where $x_i = \{x_{i1}, \dots, x_{im}\}$ is the input variable; $Z_{l1} = (\tilde{A}_{l1}, \tilde{B}_{l1}), \dots, Z_{lm} = (\tilde{A}_{lm}, \tilde{B}_{lm})$ are antecedent Z-numbers; y_i is the output variable; C_l is the consequent class label, $C_l \in \Omega$. However, the inference logic of the proposed Z-number-valued rules is not consistent with the definition of Z-numbers. According to the definition of Z-numbers, the second fuzzy number is a fuzzy constraint on the probability measure of the first fuzzy number. In the Z-number-valued rules proposed by Li et al., the second fuzzy number plays an adjusting role for the first fuzzy number, which differs from the original definition of Z-numbers. In this paper, we explore the significant role of probability measures in Z-number-valued rules.

III. Z-MIXTURE-NUMBER IF-THEN RULE INDUCTION

In this section, a new Z-number-valued if-then rule is proposed, referred to as the Z-mixture-number if-then rule. Its antecedents are not the traditional discrete or continuous Z-numbers, but Z-mixture-number, where the first fuzzy number is continuous and the second fuzzy number is discrete. The Z-mixture-number rules can simultaneously handle continuous and discrete uncertain information. Subsequently, we present the Z-Mixture-number Rule Induction Algorithm (ZMRIA). In this algorithm, we generate a series of Z-mixture-number rules for each class label. We utilize the maximum difference

for positive and negative Z-mixture-numbers for antecedents determination and feature selection. Finally, samples with unknown class labels are inferred using these generated Z-mixture-number rules to predict their class labels.

A. Definitions

In contrast to the previously mentioned Z-number-valued rules, the antecedents of the Z-mixture-number rules are Z-mixture-numbers, where their first fuzzy numbers are trapezoidal fuzzy numbers, and the second fuzzy numbers are discrete. The l th Z-mixture-number rule for classification tasks can be written as follows.

Rule R_l^Z : If $x_{i\gamma(1)}$ is $Z_{l\gamma(1)}$ and ... and $x_{i\gamma(m')}$ is $Z_{l\gamma(m')}$ then y_i is C_l . (5)

Where $x_i = \{x_{i1}, \dots, x_{im}\}$ is the input variable; $Z_{l\gamma(1)} = (\tilde{A}_{l\gamma(1)}, \tilde{B}_{l\gamma(1)}), \dots, Z_{l\gamma(m')} = (\tilde{A}_{l\gamma(m')}, \tilde{B}_{l\gamma(m')})$ are the antecedent Z-mixture-numbers associated with the input variables, where $\tilde{A}_{l\gamma(j)} = (\underline{a}_{l\gamma(j)}, a_{l\gamma(j)}, b_{l\gamma(j)}, \bar{b}_{l\gamma(j)})$ is a trapezoidal fuzzy number, and $\tilde{B}_{l\gamma(j)}$ is a discrete fuzzy number; $\gamma(j)$, $1 \leq j \leq m' \leq m$, is an index function indicating the j th selected features in the rule; y_i is the output variable; C_l is the consequent class label, $C_l \in \Omega$.

When learning a Z-mixture-number rule, the first step involves selecting the consequent class label of the rule, denoted as C_l , $C_l \in \Omega$. Each feature value of each sample in the training set $\{x_{i1}, \dots, x_{im}\}$ can be considered as a probability distribution as follows.

$$p_{ij}(u) = \begin{cases} 1, & u = x_{ij}, \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

Assume the first fuzzy number for the j th feature is $\tilde{A}_{lj}$. According to the definition of Z-numbers, we can compute the probability measure of $\tilde{A}_{lj}$ by

$$P_i(\tilde{A}_{lj}) = \sum_{u=\{x_{i1}, \dots, x_{inj}\}} \mu_{\tilde{A}_{lj}}(u) p_{ij}(u) = \mu_{\tilde{A}_{lj}}(x_{ij}). \quad (7)$$

If the class label of x_i is C_l , then we refer to x_i as a positive sample. Conversely, it is referred to as a negative sample.

Definition 6: A positive Z-mixture-number $Z_{C_l}^{lj} = (\tilde{A}_{C_l}^{lj}, \tilde{B}_{C_l}^{lj})$ based on the proportion of positive samples can be obtained by the following formulas.

$$\tilde{A}_{C_l}^{lj} = \tilde{A}_{lj}, \quad \mu_{\tilde{B}_{C_l}^{lj}}(v) = \frac{|\{x_i | x_i \in D_{C_l}, P_i(\tilde{A}_{lj}) = v\}|}{|D_{C_l}|}, \quad (8)$$

where D_{C_l} is the set of all samples in set D with class label C_l , and $|\cdot|$ is the cardinality of a set.

Definition 7: For each other class label $\omega_k \in \Omega$ and $\omega_k \neq C_l$, a negative Z-mixture-number $Z_{\omega_k}^{lj} = (\tilde{A}_{\omega_k}^{lj}, \tilde{B}_{\omega_k}^{lj})$ based on the proportion of corresponding negative samples can be computed by the following formulas.

$$\tilde{A}_{\omega_k}^{lj} = \tilde{A}_{lj}, \quad \mu_{\tilde{B}_{\omega_k}^{lj}}(v) = \frac{|\{x_i | x_i \in D_{\omega_k}, P_i(\tilde{A}_{lj}) = v\}|}{|D_{\omega_k}|}. \quad (9)$$

The resulting second fuzzy number does not necessarily satisfy convexity. In practical applications, non-convex fuzzy

numbers can more flexibly accommodate the actual distribution of data, thereby improving classification accuracy.

For a rule used in a classification task, its objective is to select positive samples while eliminating negative ones. Therefore, the larger the difference between positive Z-mixture-number and negative Z-mixture-numbers, the better. From equations (8) and (9), it can be observed that the first fuzzy numbers of positive Z-mixture-number and negative Z-mixture-number are the same. Therefore, we use the expectation of the second fuzzy numbers to measure the difference between the positive and negative Z-mixture-numbers.

Definition 8: The difference between positive and negative Z-mixture-numbers, called positive-negative Z-mixture-numbers difference, is defined as follows.

$$\begin{aligned} Diff(Z_{C_l}^{lj}) &= \sum_{v \in \mathcal{R}} v \mu_{\tilde{B}_{C_l}^{lj}}(v) \\ &\quad - \sum_{\omega_k \in \Omega, \omega_k \neq C_l} \frac{|D_{\omega_k}|}{|D| - |D_{C_l}|} \sum_{v \in \mathcal{R}} v \mu_{\tilde{B}_{\omega_k}^{lj}}(v), \end{aligned} \quad (10)$$

where $\frac{|D_{\omega_k}|}{|D| - |D_{C_l}|}$ is the weight of the negative Z-mixture-number $Z_{\omega_k}^{lj}$.

The key to rule learning lies in finding a fuzzy number $\tilde{A}_{lj}$ that maximizes the positive-negative Z-mixture-numbers difference. The above problem can be summarized as an optimization problem, as follows.

$$\begin{aligned} \max \quad & Diff(Z_{C_l}^{lj}) \\ \text{subject to: } & \tilde{A}_{lj} = (\underline{a}_{lj}, a_{lj}, b_{lj}, \bar{b}_{lj}), \\ & \underline{a}_{lj} \leq a_{lj} \leq b_{lj} \leq \bar{b}_{lj}, \\ & \tilde{A}_{C_l}^{lj} = \tilde{A}_{lj}, \\ & \mu_{\tilde{B}_{C_l}^{lj}}(v) = \frac{|\{x_i | x_i \in D_{C_l}, P_i(\tilde{A}_{lj}) = v\}|}{|D_{C_l}|}, \\ & \tilde{A}_{\omega_k}^{lj} = \tilde{A}_{lj}, \omega_k \in \Omega, \omega_k \neq C_l, \\ & \mu_{\tilde{B}_{\omega_k}^{lj}}(v) = \frac{|\{x_i | x_i \in D_{\omega_k}, P_i(\tilde{A}_{lj}) = v\}|}{|D_{\omega_k}|}. \end{aligned} \quad (11)$$

B. Rule learning

To improve accuracy and interpretability, each antecedent Z-mixture-number of the rule is selected from Z-mixture-numbers for all features based on the largest positive-negative Z-mixture-numbers difference among the samples covered by previous antecedents. Algorithm 1 demonstrates the process of rule learning.

In Algorithm 1, lines 7 to 12 involve searching for a feature with the maximum positive-negative Z-mixture-numbers difference, where $findZnumber(D^c, C_l, f_j)$ is a function that searches for the Z-mixture-number with the maximum positive-negative difference for feature f_j on dataset D^c with C_l as the target class. This function is designed to resolve the optimization problem illustrated in Model (11), with the algorithm process outlined in Algorithm 2. Lines 5 to 18 are used to generate the antecedent part of a rule. It starts with an empty rule and gradually adds antecedents until the stopping criteria are met. Lines 3 to 20 generate a series of

Algorithm 1 Z-mixture-number rule learning algorithm.

Input: The training dataset $D = \{(x_1, y_1), \dots, (x_n, y_n)\}$;
Output: Z-mixture-number rule set $RS^Z = \{R_1^Z, \dots, R_L^Z\}$,
 where the antecedents set of rule R_l^Z is $\{Z_{l\gamma(0)}, \dots, Z_{l\gamma(m')}\}$

- 1: **for** Each class label $\omega_k \in \Omega$ **do**
- 2: $l \leftarrow 0$, $D^{nc} \leftarrow D$
- 3: **while** $|D_{\omega_k}^{nc}| \geq 2$ **do**
- 4: $l \leftarrow l + 1$, $D^c \leftarrow D^{nc}$, $C_l \leftarrow \omega_k$, $t \leftarrow 0$
- 5: **while** $|D^c| > 0$ and $Acc < 1$ **do**
- 6: $t \leftarrow t + 1$, $maxDiff \leftarrow 0$
- 7: **for** Each feature $f_j \in F$ **do**
- 8: $Diff(Z_{C_l}^{lj}), Z_{C_l}^{lj} \leftarrow findZnumber(D^c, C_l, f_j)$
- 9: **if** $Diff(Z_{C_l}^{lj}) > maxDiff$ **then**
- 10: $maxDiff = Diff(Z_{C_l}^{lj})$, $Z_{l\gamma(t)} = Z_{C_l}^{lj}$,
 $\gamma(t) = j$
- 11: **end if**
- 12: **end for**
- 13: $D_n^c \leftarrow \{x_i | x_i \in D^c, \mu_{\tilde{A}_{l\gamma(t)}}(x_{i\gamma(t)}) > 0\}$
- 14: **if** $|D_n^c| = |D^c|$ **then**
- 15: **break**
- 16: **end if**
- 17: $D^c \leftarrow D_n^c$, $Acc \leftarrow \frac{|D_n^c|}{|D^c|}$
- 18: **end while**
- 19: $D^{nc} \leftarrow D^{nc} - D^c$
- 20: **end while**
- 21: **end for**

rules with the consequent class being ω_k . If the remaining dataset contains a number of samples belonging to class ω_k exceeding a threshold, the next rule with the consequent being ω_k will be generated.

To obtain the Z-mixture-number $Z = (\tilde{A}, \tilde{B})$, where $\tilde{A} = (\underline{a}, a, b, \bar{b})$, $\tilde{B}$ is computed by Eq. (8), Algorithm 2 is divided into two stages. The first stage, from line 2 to line 17, is for learning a and b in $\tilde{A}$. The second stage, from line 18 to line 35, is for learning $\underline{a}$ and $\bar{b}$ in $\tilde{A}$. Each stage aims to maximize the positive-negative difference. Next, an example is used to demonstrate the process of rule learning in detail.

Example 1: Consider a two-dimensional three-class classification problem, with 18 training samples and one testing sample, as illustrated in Figure 2. According to Algorithm 1, we first generate the first rule R_1^Z for class 1.

$$R_1^Z : \text{If } x_{i\gamma(1)} \text{ is } Z_{1\gamma(1)} \text{ and } \dots \text{ and } x_{i\gamma(m')} \text{ is } Z_{1\gamma(m')} \\ \text{then } y_i \text{ is class 1.}$$

For the first antecedent Z-mixture-number, $Z_{1\gamma(1)}$, it is selected from Z-mixture-numbers generated from all features with the maximum positive-negative difference. According to the Algorithm 2, we first obtain the Z-mixture-number $Z_{11} = (\tilde{A}_{11}, \tilde{B}_{11})$, for the feature f_1 . A_{11} is initialized to $(-\infty, -\infty, +\infty, +\infty)$. Sort the training samples based on the feature values of f_1 , then calculate the positive-negative difference of $\tilde{A}_L = (-\infty, -\infty, p, p)$ and $\tilde{A}_R = (p, p, +\infty, +\infty)$ separately if $p = x'_{i1}$. Taking $x'_{51} = -1.3494$ as an example,

Algorithm 2 $findZnumber(D, C_l, f_j)$

Input: The training dataset $D = \{(x_1, y_1), \dots, (x_n, y_n)\}$; the target class label C_l ; the feature f_j .
Output: A Z-mixture-number $Z = (\tilde{A}, \tilde{B})$; its positive-negative difference $Diff(Z)$.

- 1: $\underline{a}, a \leftarrow -\infty$; $b, \bar{b} \leftarrow +\infty$; $\tilde{A} \leftarrow (\underline{a}, a, b, \bar{b})$
- 2: $D' \leftarrow \{(x'_i, y'_i) | (x'_i, y'_i) \in D, x'_{i-1j} \leq x'_{ij} \leq x'_{i+1j}\}$
- 3: $maxDiff \leftarrow 0$, $num \leftarrow 0$
- 4: **for** $x'_i \in D'$ **do**
- 5: $p \leftarrow x'_{ij}$, $DiffLR \leftarrow 0$
- 6: Compute $Diff_L(Z_L)$ when $\tilde{A}_L = (-\infty, -\infty, p, p)$ according to Eq. (10) based on dataset D' .
- 7: Compute $Diff_R(Z_R)$ when $\tilde{A}_R = (p, p, +\infty, +\infty)$ according to Eq. (10) based on dataset D' .
- 8: $DiffLR \leftarrow \max(Diff_L(Z_L), Diff_R(Z_R))$
- 9: **if** $DiffLR > maxDiff$ **then**
- 10: $maxDiff \leftarrow DiffLR$,
- 11: **if** $Diff_L(Z_L) > Diff_R(Z_R)$ **then**
- 12: $DiffLR \leftarrow Diff_L(Z_L)$, $b \leftarrow p$
- 13: **else**
- 14: $DiffLR \leftarrow Diff_R(Z_R)$, $a \leftarrow p$
- 15: **end if**
- 16: **end if**
- 17: **end for**
- 18: **if** $b \neq +\infty$ **then**
- 19: **for** $x'_i \in D'$ and $x'_{ij} > b$ **do**
- 20: Compute $Diff1(Z_1)$ when $\tilde{A}_1 = (\underline{a}, a, b, x'_{ij})$ according to Eq. (10) based on dataset D' .
- 21: **if** $Diff1(Z_1) > maxDiff$ **then**
- 22: $maxDiff = Diff1(Z_1)$, $\bar{b} \leftarrow x'_{ij}$
- 23: $Z \leftarrow Z_1$, $Diff(Z) \leftarrow maxDiff$
- 24: **end if**
- 25: **end for**
- 26: **end if**
- 27: **if** $a \neq -\infty$ **then**
- 28: **for** $x'_i \in D'$ and $x'_{ij} < a$ **do**
- 29: Compute $Diff2(Z_2)$ when $\tilde{A}_2 = (x'_{ij}, a, b, \bar{b})$ according to Eq. (10) based on dataset D' .
- 30: **if** $Diff2(Z_2) > maxDiff$ **then**
- 31: $maxDiff = Diff2(Z_2)$, $\underline{a} \leftarrow x'_{ij}$
- 32: $Z \leftarrow Z_2$, $Diff(Z) \leftarrow maxDiff$
- 33: **end if**
- 34: **end for**
- 35: **end if**

$\tilde{A}_L = (-\infty, -\infty, -1.3494, -1.3494)$. From Figure 2, it can be observed that

$$P_1(\tilde{A}_L) = P_2(\tilde{A}_L) = \dots = P_5(\tilde{A}_L) = 1,$$

$$P_6(\tilde{A}_L) = P_7(\tilde{A}_L) = \dots = P_{18}(\tilde{A}_L) = 0,$$

In samples where the probability measure is greater than 0, x'_2 belongs to class 3, while others belong to class 1. Then we can get the positive Z-mixture-number $Z_{L\omega_1}$ based on Eq. (8).

$$\tilde{A}_{L\omega_1} = (-\infty, -\infty, -1.3494, -1.3494),$$

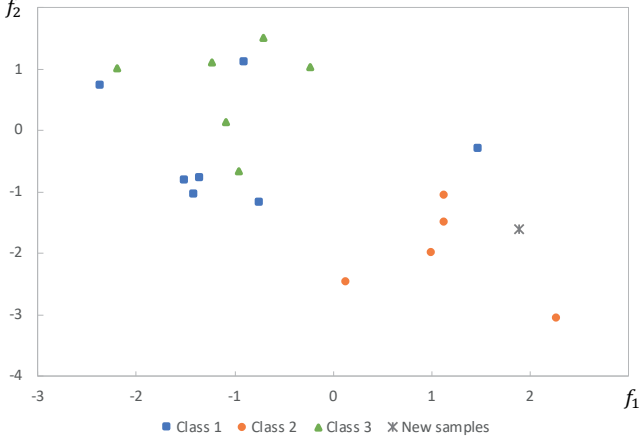


Fig. 1: A two-dimensional three-class classification problem.

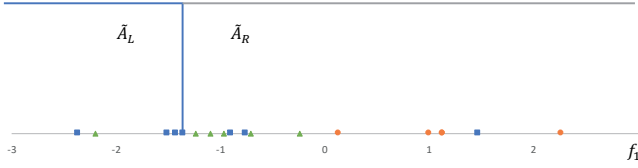


Fig. 2: If $p = x'_{51} = -1.3494$.

$$\mu_{\tilde{B}_{L\omega_1}}(1) = \frac{4}{7+1} = 0.5,$$

$$\mu_{\tilde{B}_{L\omega_1}}(0) = \frac{3}{7+1} = 0.375,$$

where 1 is included to avoid division by zero errors. Similarly, two negative Z-mixture-numbers, $Z_{L\omega_2}$ and $Z_{L\omega_3}$, are

$$\tilde{A}_{L\omega_2} = (-\infty, -\infty, -1.3494, -1.3494),$$

$$\mu_{\tilde{B}_{L\omega_2}}(1) = 0, \quad \mu_{\tilde{B}_{L\omega_2}}(0) = 0.8333;$$

$$\tilde{A}_{L\omega_3} = (-\infty, -\infty, -1.3494, -1.3494),$$

$$\mu_{\tilde{B}_{L\omega_3}}(1) = 0.1428, \quad \mu_{\tilde{B}_{L\omega_3}}(0) = 0.7142.$$

Then the positive-negative difference of Z_{ω_1} is

$$\begin{aligned} Diff(Z_{L\omega_1}) &= (0.5 + 0.125) - \left(\frac{5 \times (0 + 0.1667)}{18 - 7} \right. \\ &\quad \left. + \frac{6 \times (0.1428 + 0.1428)}{18 - 7} \right) \\ &= 0.3933, \end{aligned}$$

where 0.125, 0.1667 and 0.1428 are to compensate for adding 1 to the previous divisor.

Similarly, we can obtain the Z-mixture-number $Z_{R\omega_1}$, and its positive-negative difference $Diff(Z_{R\omega_1}) = -0.2970$. The corresponding positive-negative differences when p takes on the values of other sample feature values are shown in Table I. From Table I, it can be seen that the maximum positive-negative difference is $Diff(Z_{L\omega_1}) = 0.4096$. At this point, the value of p is -0.7454 . Therefore, the value of p is assigned to b , obtaining the fuzzy number after the first stage, $\tilde{A}_{11} = (-\infty, -\infty, -0.7454, +\infty)$.

TABLE I: The corresponding positive-negative differences when p takes on the value of x'_{i1} .

i	x'_{i1}	$Diff(Z_{L\omega_1})$	$Diff(Z_{R\omega_1})$
1	-2.3591	0.0963	0
2	-2.1932	0.0183	-0.1250
3	-1.5006	0.1433	-0.0470
4	-1.4166	0.2683	-0.1720
5	-1.3494	0.3933	-0.2970
6	-1.2287	0.3154	-0.4220
7	-1.0946	0.2375	-0.3441
8	-0.9619	0.1596	-0.2662
9	-0.8972	0.2846	-0.1883
10	-0.7454	0.4096	-0.3133
11	-0.7076	0.3317	-0.4383
12	-0.2339	0.2537	-0.3603
13	0.1403	0.1780	-0.2824
14	1.0044	0.1022	-0.2067
15	1.1317	0.0265	-0.1309
16	1.1368	-0.0492	-0.0551
17	1.4800	0.0757	0.0205
18	2.2800	0	-0.1044

As b is a real number while a is not, the second stage is for learning the value of $\tilde{b}$. First, assume the value of $\tilde{b}$ to be each feature value greater than b , as shown in Figure 3, then compute the corresponding positive-negative differences. Select the feature value with the maximum positive-negative difference as the value of $\tilde{b}$. When $\tilde{b} = x'_{13} = 0.1403$, $\tilde{A}_{11} = (-\infty, -\infty, -0.7454, 0.1403)$. Then

$$P_1(\tilde{A}_{11}) = P_2(\tilde{A}_{11}) = \dots = P_{10}(\tilde{A}_{11}) = 1,$$

$$P_{12}(\tilde{A}_{11}) = 0.9574, \quad P_{13}(\tilde{A}_{11}) = 0.4225.$$

Where $x'_2, x'_7, x'_7, x'_8, x'_{11}, x'_{12}$ belong to class 3, and others belong to class 1. Then we can get the positive Z-mixture-

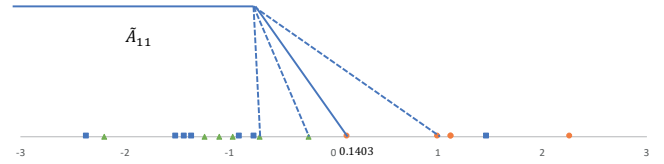


Fig. 3: The second stage of Algorithm 2.

number $Z_{11\omega_1}$ based on Eq. (8).

$$\tilde{A}_{11\omega_1} = (-\infty, -\infty, -0.7454, 0.1403),$$

$$\mu_{\tilde{B}_{11\omega_1}}(1) = 0.75, \quad \mu_{\tilde{B}_{11\omega_1}}(0) = 0.125.$$

Similarly, the negative Z-mixture-number $Z_{11\omega_2}$ is

$$\tilde{A}_{11\omega_2} = (-\infty, -\infty, -0.7454, 0.1403),$$

$$\mu_{\tilde{B}_{11\omega_2}}(1) = 0, \quad \mu_{\tilde{B}_{11\omega_2}}(0) = 0.8333.$$

The negative Z-mixture-number $Z_{11\omega_3}$ is

$$\tilde{A}_{11\omega_3} = (-\infty, -\infty, -0.7454, 0.1403),$$

$$\mu_{\tilde{B}_{11\omega_3}}(1) = 0.5714, \quad \mu_{\tilde{B}_{11\omega_3}}(0.9574) = 0.1428,$$

$$\mu_{\tilde{B}_{11\omega_3}}(0.4225) = 0.1428, \quad \mu_{\tilde{B}_{11\omega_3}}(0) = 0.$$

The positive-negative difference is

$$Diff(Z_{11\omega_1}) = -0.7008.$$

TABLE II: The corresponding positive-negative differences when p takes on the value of x'_{i1} .

i	x'_{i1}	$Diff(Z_{11})$
11	-0.7076	0.4096
12	-0.2339	-0.1279
13	0.1403	-0.7008
14	1.0044	-1.3349
15	1.1317	-1.9785
16	1.1368	-2.6226
17	1.4800	-3.3099
18	2.2800	-4.0268

The corresponding positive-negative differences for different values of $\bar{b}$ are shown in Table II. It can be observed that the feature value with the maximum positive-negative difference is -0.7076. Therefore, $Z_{11} = ((-\infty, -\infty, -0.7454, -0.7076), (\frac{0.75}{1} + \frac{0.125}{0}))$ with positive-negative difference 0.4096. At this point, the second discrete fuzzy number of Z_{11} does not conform to the general definition of a fuzzy number. We can normalize it to make it conform to the general definition, $Z_{11} = ((-\infty, -\infty, -0.7454, -0.7076), (\frac{1}{1} + \frac{0.1667}{0}))$. This normalization does not affect the final results.

For the feature f_2 , the Z-mixture-number can be obtained using the same method as $Z_{12} = ((-1.5093, -1.1811, +\infty, +\infty), (\frac{1}{1}))$ with positive-negative difference 0.3030 that is less than 0.4096. Therefore, the first Z-mixture-number of the first rule is $Z_{1\gamma(1)} = Z_{11}$ and $\gamma(1) = 1$. Using the same method, we can generate the remaining Z-mixture-numbers for the first rule, as shown as follows.

R_1^Z : If x_{i1} is $((-\infty, -\infty, -0.7454, -0.7076), (\frac{1}{1} + \frac{0.1667}{0}))$

and x_{i2} is $((-\infty, -\infty, -0.7757, -0.6649), (\frac{1}{1} + \frac{0.5}{0}))$

then y_i is class 1.

All learned complete Z-mixture-number rules are listed in Table III. These rules are stored in the knowledge base.

C. Inference engine

Let's assume that L Z-mixture-number rules have been learned and stored in the knowledge base, denoted as the set $RS^Z = \{R_1^Z, R_2^Z, \dots, R_L^Z\}$.

Definition 9: The coverage of one rule, R_l^Z , $l = 1, 2, \dots, L$, for one sample $x_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}$, can be calculated using the following formula.

$$\varphi_{R_l^Z}(x_i) = \prod_{j=1}^{j=m'} \mu_{\bar{A}_{l\gamma(j)}}(x_{i\gamma(j)}). \quad (12)$$

Here, we only utilize the first fuzzy number of the Z-mixture-number and do not employ the second fuzzy number. This is

because the first fuzzy number is generated with the objective of optimizing the second fuzzy number, which is optimal at this stage. Hence, using the first fuzzy number here can effectively represent the entire Z-mixture-number.

Definition 10: The certainty factor of rule R_l^Z is defined as follows [31].

$$CF(R_l^Z) = \frac{2|D_{C_j}|/|D| + \sum_{x_i \in X_{C_j}} \varphi_{R_l^Z}(x_i)}{2 + \sum_{x_i \in X_{C_j}} \varphi_{R_l^Z}(x_i)}. \quad (13)$$

For a new query sample $\dot{x} = \{\dot{x}_1, \dot{x}_2, \dots, \dot{x}_m\}$, its class label $\dot{y}$ is predicted by the follows.

$$\dot{y} = \underset{C_l = \omega_k}{\operatorname{argmax}}_{\omega_k \in \Omega} (\sum \varphi_{R_l^Z}(\dot{x}) CF(R_l^Z)). \quad (14)$$

For a new query sample not covered by any rule, we utilized the rule stretching method proposed in [31]. The idea is that the ordering reflects the importance of the antecedent. For one Z-mixture-number rule, R_L^Z , the ordered antecedents set is $\{Z_{l\gamma(1)}, Z_{l\gamma(2)}, \dots, Z_{l\gamma(m')}\}$. We assume that the $j+1$ th antecedent is the first one that does not cover the new sample, meaning that antecedents $Z_{l\gamma(1)}, Z_{l\gamma(2)}, \dots, Z_{l\gamma(j)}$ all cover the sample. Then these antecedents and the consequent of R_l^Z form a new rule $\hat{R}_l^Z$. Its certainty factor is $CF(\hat{R}_l^Z) = \frac{j}{m'} CF(R_l^Z)$. The predicted class label of this new query sample is

$$\dot{y} = \underset{C_l = \omega_k}{\operatorname{argmax}}_{\omega_k \in \Omega} \sum \varphi_{\hat{R}_l^Z}(\dot{x}) CF(\hat{R}_l^Z). \quad (15)$$

Example 2 (Continue with Example 1): The new sample in Figure 1 is $\dot{x} = \{1.8829, -1.6145\}$. The certainty factors are listed in Table III. The coverage of the rules for the new sample and the certainty factors are

$$\varphi_{R_1^Z}(\dot{x}) = \varphi_{R_2^Z}(\dot{x}) = \varphi_{R_4^Z}(\dot{x}) = \varphi_{R_5^Z}(\dot{x}) = 0, \varphi_{R_3^Z}(\dot{x}) = 1.$$

We can conclude that the class label of $\dot{x}_1$ is $\dot{y} = \text{Class 2}$.

IV. EXPERIMENTS

To analyze the performance of ZMRIA, An experiment is conducted under the framework of the KEEL software [32]. First, the experimental setup is introduced, followed by the presentation of the experimental results. We analyzed the performance of ZMRIA from two aspects: classification accuracy and runtime.

A. Experimental setup

We collected 46 benchmark datasets from the UCI and KEEL repositories. These datasets have varying numbers of samples, ranging from 90 to 17898; different numbers of features, ranging from 3 to 85; and different numbers of class labels, ranging from 2 to 9, as shown in Table IV.

TABLE III: All learned Z-mixture-number rules.

R_l^Z	Complete rule	$CF(R_l^Z)$
1	If x_{i1} is $((-\infty, -\infty, -0.7454, -0.7076), (\frac{1}{1} + \frac{0.1667}{0}))$ and x_{i2} is $((-\infty, -\infty, -0.7757, -0.6649), (\frac{1}{1} + \frac{0.5}{0}))$ then y_i is class 1.	0.7962
2	If x_{i2} is $((-0.6649, -0.2982, +\infty, +\infty), (\frac{1}{1}))$ and x_{i2} is $((-\infty, -\infty, 0.7163, 1.0068), (\frac{1}{1} + \frac{0.5}{0}))$ then y_i is class 1.	0.5556
3	If x_{i1} is $((-0.2339, 0.1403, +\infty, +\infty), (\frac{1}{1}))$ and x_{i2} is $((-\infty, -\infty, -1.0650, -0.2982), (\frac{1}{1}))$ then y_i is class 2.	0.7936
4	If x_{i2} is $((-0.7757, -0.6649, +\infty, +\infty), (\frac{1}{1}))$ and x_{i1} is $((-2.3591, -2.1932, +\infty, +\infty), (\frac{1}{1}))$ and x_{i1} is $((-\infty, -\infty, -0.9619, -0.8972), (\frac{1.0}{1.0}, \frac{0.5}{0.0}))$ then y_i is class 3.	0.7778
5	If x_{i2} is $((0.7163, 1.0203, +\infty, +\infty), (\frac{1}{1}))$ and x_{i1} is $((-0.8972, -0.7076, +\infty, +\infty), (\frac{1}{1}))$ then y_i is class 3.	0.6667

Where ‘#samples’ denotes the number of samples, ‘#features’ represents the number of features, and ‘#classes’ represents the number of classes.

TABLE IV: The information about 46 datasets used in the experiments.

Name	#samples	#features	#classes
appendicitis	106	7	2
automobile	159	25	6
bands	365	19	2
banknote	1372	4	2
breast cancer	569	5	2
bupa	345	6	2
cleveland	303	13	5
coil2000	9822	85	2
column_2C	310	6	2
column_3C	310	6	3
credit	690	14	2
german	1000	20	2
glass	214	9	7
haberman	306	3	2
heart statlog	270	13	2
hepatitis	90	19	2
housevotes	232	16	2
HTRU_2	17898	8	2
ilpd	583	10	2
ionsphere	351	34	2
iris	150	4	3
liver disorders	345	6	2
lymphography	148	18	4
mammographic	830	5	2
marketing	6876	13	9
mfeat-zernike	2000	47	2
monk_1	556	6	2
new-thyroid	215	5	3
optdigits	5620	64	2
page-blocks	5473	10	5
pima	768	8	2
qsar	1055	41	2
rice_OC	3810	7	2
ring	7400	20	2
SAheart	462	9	2
satimage	6435	36	7
seed	210	7	3
segment	2310	19	7
sonar	208	60	2
spambase	4601	57	2
tae	151	5	3
tic-tac-toe	941	9	2
transfusion	748	4	2
vehicle	846	18	4
wdbc	569	30	2
zoo	101	16	7

To evaluate the generalization ability of models to unseen data, we used the 10-fold cross-validation (10CV) technique. In this method, the original dataset is randomly partitioned into 10 equally sized folds. The modeling process is repeated 10 times. In each iteration, one of the 10 folds is held out as the testing set, while the remaining 9 folds are used as the training set. The model is trained on the training set, and its performance is evaluated on the testing set. After completing all 10 iterations, the classification accuracy obtained from each fold are averaged to compute the overall performance of the model.

To demonstrate the advantages of ZMRIA, we compared its classification performance with three benchmark classifiers including C45 decision tree, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM), as well as seven rule-

based classifiers: The Fuzzy Rule-Based (FRB) classification system is based on the method in [33] and the rule weights are computed by the method in [34]; The Extended Belief Rule-based (EBRB) classification system [35] is an extension of belief rule-based inference methodology [18]; The Fuzzy Unordered Rule Induction Algorithm (FURIA) is an algorithm derived from the RIPPER algorithm [36]; The Fuzzy Association Rule-based Classification method for High-Dimensional problems (FARC-HD) is a classification method based on fuzzy association rules, specifically designed for high-dimensional problems [37]. The Fuzzy Rule-based Decision Tree (FRDT) is a decision tree with oblique geometric structure, and its each node is a fuzzy rule; The Z-number-valued Rule-Based (ZRB) classification system is an extension of FRB classification system [19]; The Z-number-valued Rule-based Decision Tree (ZRDT) is an extension of FRBDT [20]. The parameter configurations for these methods and ZMRIA are presented in Table V. All models are implemented on KEEL software [32].

TABLE V: The parameter configurations for methods.

Model	Parameters
C45	Confidence: 0.25 Minimum samples per leaf: 2
KNN	K:1 Distance function: Euclidean
SVM	Kernel type: radial basis function C: 100 Coef0: 0.0 Epsilon: 0.001 nu: 0.1 Degree: 1 p: 1.0 Gamma: 0.01 Shrinking: 1.0
FRB	Number of linguistic labels per feature: 3 T-norm: product Rule weight: certainty factor Fuzzy reasoning method: winning rule
EBRB	Number of utility values per feature: 3 Transformation method: utility-based transformation methods Rule weight: Rule inconsistency
FURIA	Number of optimization: 2 Number of folds: 3 Minimum samples per rule: 2
FARC-HD	Number of linguistic labels per feature: 5 Minimum Support: 0.05 Maximum Confidence: 0.8 Depth of the trees (Depthmax): 3 Parameter K of the prescreening: 2 Maximum number of evaluations: 15000 Population size: 50 Parameter α : 0.15 Bits per gen: 30 Rule weight: certainty factor
FRDT	Maximum number of features: 5 Threshold: 0.6 α : 0.02
ZRB	Number of linguistic labels per feature: 3 T-norm: product Rule weight: certainty factor Fuzzy reasoning method: winning rule
ZRDT	Maximum number of features: 5 Threshold: 0.6
ZMRIA	Minimum samples per rule: 2

B. Classification accuracy

First, we compared the classification accuracy of ZMRIA with other classifiers. Table VI lists the classification accuracies, with the highest accuracy highlighted in bold. The blank spaces indicate that the classifier cannot classify the corresponding dataset due to its inability to handle missing data. The last row represents the average classification accuracy rate for each classifier. From the Table VI, it can be seen that ZMRIA achieved the maximum accuracies on 14 out of 46 datasets, and maximum average accuracy. To clearly describe the distribution of classification accuracy for each classifier across different datasets, a boxplot for the eleven classifiers are shown in Figure 4. Where the box in the middle of the plot

TABLE VI: Classification accuracy rates (%) of eleven classifiers on 46 datasets.

Dataset	C45	KNN	SVM	FRB	EBRB	FURIA	FARC-HD	FRDT	ZRB	ZRDT	ZMRIA
appendicitis	84.09	81.91	86.91	85.82	84.91	83.09	84.82	75.64	84.82	77.18	87.64
automobile	78.58	79.21	33.96	57.79	83.67	75.54	77.42	33.38	57.79	65.50	80.54
bands	65.14	72.25	63.57	67.98	70.93	69.05	67.39	62.19	67.70	56.15	70.17
banknote	98.40	99.85	100.00	94.97	98.62	98.91	99.64	87.90	95.92	90.31	98.98
breast cancer	91.04	92.09	84.35	89.63	89.99	91.74	91.22	90.34	90.34	89.64	92.63
bupa	66.97	63.14	63.50	58.54	61.18	65.81	65.49	63.50	60.01	60.62	69.84
cleveland	54.43	53.14	51.16	-	-	56.40	-	-	-	-	54.45
coil2000	93.97	89.67	90.91	75.63	89.12	93.93	94.03	93.76	75.45	86.55	93.86
column_2C	82.90	80.32	80.65	70.00	77.42	83.23	84.52	77.10	71.94	80.32	83.87
column_3C	81.61	76.77	76.13	59.03	68.71	82.58	80.00	81.29	67.42	80.00	82.58
credit	93.62	80.58	100.00	97.68	99.13	88.41	93.19	87.97	97.68	86.81	94.64
german	70.70	69.30	65.80	18.80	64.90	73.20	72.70	70.50	18.80	63.90	73.30
glass	69.63	71.60	69.20	60.74	64.09	71.45	66.88	53.29	60.78	55.28	72.03
haberman	70.91	67.63	69.94	73.87	71.90	72.23	71.90	62.56	73.53	73.20	72.59
heart-statlog	75.19	75.56	59.63	52.59	72.22	80.37	82.96	83.70	52.59	81.48	80.37
hepatitis	82.50	81.25	82.50	23.75	83.75	85.00	87.50	78.75	23.75	80.00	86.25
housevotes	96.11	91.32	96.11	41.36	79.69	96.54	95.25	96.97	41.36	96.97	96.54
HTRU_2	97.94	97.14	96.64	95.68	96.26	97.85	97.85	97.50	96.74	96.13	97.97
ilpd	67.07	63.45	73.07	70.32	68.27	68.96	68.77	55.39	69.81	60.68	70.32
ionosphere	87.17	60.96	93.17	64.65	90.87	89.75	91.16	86.33	64.94	88.89	89.75
iris	96.00	93.33	97.33	92.67	95.33	95.33	95.33	94.67	94.67	95.33	94.00
liver-disorders	66.97	63.14	63.50	58.54	61.18	65.81	65.49	63.50	60.01	60.62	69.84
lymphography	76.33	74.90	83.81	12.71	70.81	83.10	81.71	39.81	12.71	77.05	83.24
mammographic	84.10	73.98	84.34	79.64	78.43	83.37	81.20	83.01	79.40	84.58	83.73
marketing	31.25	27.78	34.35	20.77	29.26	29.04	25.73	26.44	22.31	20.17	31.33
mfeat-zernike	98.35	99.40	90.00	98.80	99.15	98.80	98.90	96.00	98.80	97.85	99.10
monk_1	98.02	81.30	88.47	68.86	75.71	100.00	99.46	66.55	68.86	66.55	100.00
new-thyroid	93.98	97.21	93.96	84.24	82.36	94.46	93.01	93.53	88.44	94.42	95.39
optdigits	97.74	89.82	93.29	45.41	98.29	98.75	98.56	56.32	45.41	95.12	98.84
page-blocks	96.93	95.69	92.60	91.91	90.46	97.42	95.21	84.16	92.23	90.19	97.31
pima	75.40	69.78	63.80	73.17	72.13	74.86	75.26	73.05	75.78	73.18	76.30
qsar	83.51	83.12	86.26	78.39	83.80	84.27	85.13	53.09	78.58	81.34	85.41
rice_OC	92.36	88.79	68.82	88.19	91.97	92.57	92.49	92.62	91.52	92.49	92.60
ring	90.27	75.24	50.49	55.57	52.39	94.07	93.86	66.53	87.49	66.68	92.69
SAheart	71.61	65.58	59.10	72.94	70.33	72.08	69.05	68.40	71.86	66.23	70.98
satimage	86.51	90.63	51.98	48.31	85.14	88.66	83.67	78.65	57.31	79.21	88.95
seed	92.38	94.29	91.43	91.90	90.95	92.38	92.38	90.95	91.90	90.00	93.81
segment	97.14	14.29	89.83	86.32	90.78	97.32	92.99	68.35	86.62	84.03	97.45
sonar	78.36	86.93	78.36	61.14	73.55	80.24	77.88	71.55	61.14	74.48	81.26
spambase	92.57	90.11	86.57	71.81	73.68	93.24	91.20	85.26	78.55	91.33	93.00
tae	54.42	41.83	56.38	55.13	51.79	43.12	57.79	41.75	53.13	42.50	54.46
tic-tac-toe	84.55	73.07	81.63	0.00	98.33	98.54	96.87	70.78	0.00	70.67	98.02
transfusion	78.61	61.64	71.25	76.61	75.94	78.74	78.61	71.52	76.61	63.63	78.74
vehicle	75.06	69.62	52.14	61.59	68.45	70.58	69.16	48.70	62.42	59.34	74.23
wdbc	92.25	95.25	63.62	93.14	97.01	96.31	95.07	94.89	93.14	95.60	95.25
zoo	94.00	92.00	95.00	65.27	90.00	95.00	93.00	85.09	65.27	87.09	95.00
Average	82.32	76.87	76.21	66.49	79.17	83.09	83.37	73.40	68.12	77.10	84.11

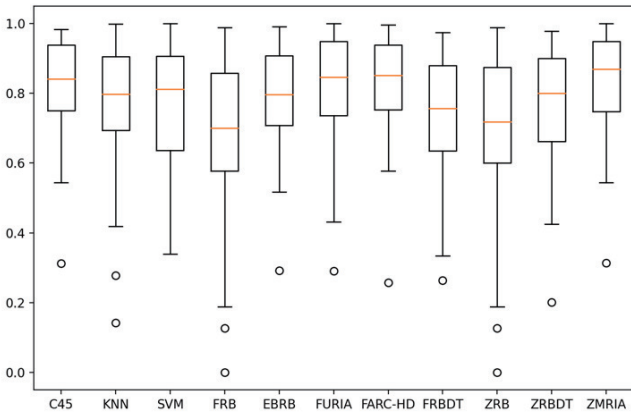


Fig. 4: The boxplot for the eleven classifiers.

TABLE VII: The results of Wilcoxon signed-rank tests.

Classifier	Statistic value	p-value	Critical value	Hypothesis
C45	74	1.2699e-08	361	Rejected
KNN	82	2.7712e-08	361	Rejected
SVM	115	5.5386e-06	361	Rejected
FRB	29	9.8850e-11	343	Rejected
EBRB	90	1.1528e-07	343	Rejected
FURIA	180	0.00116	361	Rejected
FARC-HD	170	3.8338e-05	343	Rejected
FRDT	27	7.1679e-11	343	Rejected
ZRB	31	1.3500e-10	343	Rejected
ZRDT	32	1.5717e-10	343	Rejected

represents the interquartile range, which spans from the first quartile to the third quartile. The length of the box indicates the spread of the middle 50% of the distribution. The median is represented by a horizontal line inside the box. It marks the midpoint of the distribution. The whiskers extend from the edges of the box to the minimum and maximum values within a certain range from the quartiles. Individual data points

that fall beyond the whiskers are considered outliers and are plotted as individual points or asterisks. They represent values that are significantly different from the majority of the data. From the box plots, it is evident that ZMRIA achieved the highest median and third quartile.

However, these results do not provide sufficient support to draw strong conclusions in favor or against any classifier. To obtain robust evidence, we relied on a statistical test of the results. The Wilcoxon signed-rank test is a non-parametric statistical test used to compare the differences between two related or paired samples. It can be employed to assess whether two sets of related observations have statistically significant differences. The basic principle of the Wilcoxon signed-rank test is to convert the absolute values of the differences into ranks, and then assess whether the sum of ranks indicates a statistically significant difference between the samples. The test is applied to paired data, where each observation in one sample is paired with a corresponding observation in the

other sample. For each pair of observations, calculate the difference between the two values. Discard pairs where the difference is zero. Rank the absolute differences from smallest to largest, ignoring the signs. Ties are handled by assigning the average rank. Calculate the sum of the ranks of the positive or negative differences. These sums are referred to as the test statistic W^+ and W^- , respectively. The final test statistic is $T = \min(W^+, W^-)$. We can compare the obtained test statistic to the critical value from the Wilcoxon signed-rank table or use statistical software to obtain the p-value. If the obtained test statistic exceeds the critical value (or if the p-value is less than the chosen significance level, 0.05 here), we reject the null hypothesis. Otherwise, we fail to reject the null hypothesis. The test results of classification accuracy for C45, KNN, SVM, FRB, EBRB, FURIA, FARC-HD, FRDT, ZRB and ZRDT compared to ZMRIA are shown in Table VII.

From the test results, it can be observed that all null hypotheses were rejected, indicating that the classification

TABLE VIII: Average model running times in milliseconds

Dataset	C45	KNN	SVM	FRB	EBRB	FURIA	FARC-HD	FRDT	ZRB	ZRDT	ZMRIA
appendicitis	40.1	35.2	29.9	57.2	115.6	57.1	515	50.8	56	53.2	65.4
automobile	62.7	49.9	53	174.8	464	126.5	14767.5	93.6	183	101.1	145.6
bands	95.7	56.3	79.6	313.6	810.3	256.2	6581.4	100.7	328.5	113.3	232.6
banknote	91.4	78.7	66.1	110.9	1004.6	258.8	2976.5	81.6	133	94.5	211
breast cancer	63.5	52.7	87.4	81.3	464.7	178.2	1253.7	71.4	85.4	86.7	158.3
bupa	57.7	43.5	66.8	70.7	300.7	153.9	1212	62.3	76.2	77.9	131.5
cleveland	70.6	49.5	78	-	-	193.6	-	-	-	-	206.2
coil2000	41705.5	4173.7	26628.5	199339	1000913.2	20327.7	241538.4	878	209615.9	1757.7	7695.4
column_2C	53.6	53.1	60.9	82.1	318.1	133.8	1006.9	70.4	87.7	76.1	131.6
column_3C	55.8	47.1	64.9	85.4	324.4	160.3	1376.4	79.7	90.2	86	160.9
credit	102.4	69.1	159.9	403.9	949.5	274.7	8052.1	87.7	424.4	92.7	236.9
german	113.4	98.8	247.8	1197.9	2329.4	316.8	24647	137.6	1224.1	126.8	276.8
glass	59.6	42.6	52.3	67	236.9	132	3639.2	91.3	70.5	78.7	162.8
haberman	43.6	39.4	70.4	72.7	181.7	92.8	766	61.1	69.6	62	110.4
heart statlog	62.5	48.5	60.7	223.2	409.6	136.9	2905.6	80.4	229	87.1	131.7
hepatitis	45	41.8	32.8	100	169.5	61.5	1308.3	60	104.5	63.5	64.8
housevotes	45.9	44.4	40.9	211.2	378.3	69.6	1382.5	71.5	211.4	70.2	67.1
HTRU_2	3999.3	2424.2	25498.6	5602.1	109291.2	14184.2	26249	455.2	10583.9	511	4014.2
ilpd	83.9	58.2	112.9	110.1	706.6	255	1900.6	98.5	127.2	93.7	194.7
ionosphere	121.5	65	58.9	343.8	975.5	249.2	8601.6	92.7	360.6	93.3	225.9
iris	40.3	36.6	28.8	58.9	135.6	65.7	331.2	59.8	63.1	59.1	69.8
liver disorders	56.9	43	63.9	69.4	297.8	151.2	1216.7	86.7	77.2	69	130.7
lymphography	48.9	45.3	41.2	180.4	299.9	85.3	3833	95.6	185.9	81.2	86
mammographic	77.1	56.8	177.9	87.9	582.5	153.3	2750.4	81.1	111.4	79.5	186.4
marketing	2152.7	529.5	9257.7	4762.3	25370	3099	102116.4	478.7	6329.1	474.7	2152.9
mfeat-zernike	649.5	244.4	1476.4	6360.2	22707.3	1028.5	38952.8	220.5	6429.8	286.9	1251.9
monk_1	50	45.6	80.3	150.2	346.1	129.5	3319.6	57.3	157	61.2	74.7
new-thyroid	46	37.9	38.2	61	199.1	76.9	571.9	65.4	65.6	61.1	82.2
optdigits	3975.2	1106.5	26740.2	72825.4	154813.4	2823.8	294315.1	840	77459.8	746.6	2070.4
page-blocks	1434.6	365.4	26104	753.1	16339.1	1906.9	17980.7	342	1271.4	386.8	4496
pima	85.8	58	166.1	163.9	849.7	312.6	5456.5	107.9	186.2	131.3	249.8
qsar	341.4	125.3	229.1	883.9	5567.5	904.4	29813.3	196.3	948	229.7	513.6
rice_OC	291.8	228.3	5344.2	571.3	7639.6	1219.9	10245.2	225.6	815.9	255.8	694.4
ring	6679.9	727.6	19396.9	6225	61952.9	21725.6	67217	347.4	9564.5	419	6283.7
SAheart	68.2	53.1	88.5	141.7	542.9	187.7	3973.1	72.6	147.3	87.2	153.8
satimage	3678.6	869.3	37244.5	3246	130967.1	19021.4	307817.2	663	4032.3	955.9	17868.5
seed	48.4	40.9	36	67.3	225	81.3	829.4	57.2	72.3	68.6	89.8
segment	428.7	199.6	1300.3	717.7	8949.6	1423.5	36430.8	493.5	828.4	407.1	1722.8
sonar	110.1	70.3	57.3	443.3	1421	246.2	164218.7	95.2	448	124.9	202.3
spambase	5695.1	761.1	7300.1	4292.2	125903.4	4555.9	149301.1	487.2	7268.3	772.7	3953.9
tae	48.3	37.4	45.3	53.4	125	77	919	52.6	57.8	59.8	82.2
tic-tac-toe	67.6	68.3	169.3	580.1	894.6	184.7	8509.4	99.6	578.9	129.9	121.5
transfusion	61.4	53.1	125.5	75.1	516.9	160.7	932.9	64.9	91.6	75.8	162
vehicle	152.5	78.4	256.1	362.6	2241	594.4	20254.5	134.4	385.4	161.5	436.9
wdbc	129.5	79	137.8	554.2	2136	309.5	9749.5	102.1	556.2	131.4	267.3
zoo	44.7	37.2	36.3	80.9	159.7	51.2	2299.8	64.6	83.7	70.3	48.5
Average	1596.45	294.99	4119.39	6942.54	37589.47	2134.67	36311.88	184.79	7606.13	224.72	1262.51

accuracy of ZMRIA is significantly better than the other ten models on these 46 datasets.

C. Running Time

In addition to the model's predictive accuracy, running time is also a crucial factor influencing the practical application of the model. Table VIII lists the running times in milliseconds for all compared models. Whereas the last row represents the average runtime across the 46 datasets. From the average runtime perspective, ZMRIA is faster than C45, SVM, FRB, EBRB, FURIA, FARC-HD and ZRB, but slower than KNN, FRDT and ZRDT. Comparing the runtime of C45 and ZMRIA, we can observe that C45 is faster than ZMRIA on most datasets. However, on some datasets with large sample sizes and high dimensionality, such as coil2000, optdigits, and spambase datasets, ZMRIA is faster than C45. This is even more pronounced for SVM, as ZMRIA is faster than SVM on datasets that are concentrated in large sample sizes and high dimensions, such as coil2000, HTRU_2, marketing, mfeat-zernike, optdigits, page-blocks, ring, satimage and spambase datasets. In FRB, ZRB and EBRB classification systems, it is highly likely to encounter the issue of combinatorial rule explosion, where the number of rule in a fuzzy rule-based system grows exponentially with the increase in the number of features. This phenomenon leads to a dramatic increase in system complexity and runtime when the system handles high-dimensional data. While ZMRIA performs feature selection during rule generation, effectively avoiding the issue of combinatorial rule explosion. Comparing the FURIA and ZMRIA models, in most datasets, ZMRIA is faster than FURIA because ZMRIA has one less optimization step than FURIA. Although FARC-HD uses a genetic algorithm for rule selection to avoid the issue of combinatorial rule explosion, the genetic algorithm requires multiple iterations to achieve satisfactory results. On the other hand, ZMRIA selects features by comparing positive-negative differences, avoiding the need for multiple iterations. Therefore, it runs faster than FARC-HD.

V. CONCLUSIONS

In this paper, we introduced a new rule-based classification system based on Z-numbers called Z-mixture-number rule induction algorithm (ZMRIA). It improves existing Z-number-valued rule-based systems in three aspects. Firstly, to handle both discrete and continuous cases simultaneously, Z-mixture-numbers are introduced into the rules. Then, the method of using the feature values of each sample as underlying probability functions to generate Z-mixture-numbers adheres to the definition of Z-numbers. Lastly, to avoid the issue of combinatorial rule explosion, we utilized the criterion of positive-negative Z-mixture-numbers difference to select features. These improvements not only resulted in a rule learner with outstanding classification performance but also reduced the runtime. To evaluate the performance of ZMRIA, we conducted comparative experiments with ten other classifiers across a large number of real-world datasets, including three benchmark classifiers and seven rule-based classifiers.

The results indicate that ZMRIA outperforms other methods in terms of classification accuracy and is faster in runtime compared to seven classifiers.

Future research can further explore and optimize the performance of ZMRIA to address a wider range of application scenarios and data types.

REFERENCES

- [1] B. Gohain, R. Chutia, and P. Dutta, "Distance measure on intuitionistic fuzzy sets and its application in decision-making, pattern recognition, and clustering problems," *International Journal of Intelligent Systems*, vol. 37, no. 3, pp. 2458–2501, 2022.
- [2] S. Song, Z. Jia, J. Yang, and N. Kasabov, "Image segmentation based on fuzzy low-rank structural clustering," *IEEE Transactions on Fuzzy Systems*, vol. 31, no. 7, pp. 2153–2166, 2023.
- [3] L. Pan, Y. Deng, and K. H. Cheong, "Quaternion model of pythagorean fuzzy sets and its distance measure," *Expert Systems with Applications*, vol. 213, p. 119222, 2023.
- [4] M. B. Castañeda, L. E. C. Velázquez, S. S. R. Aguilar, J. S. Romero, and V. A. C. Escamilla, "Expert system through a fuzzy logic approach for the macroscopic visual analysis of corroded metallic ferrous surfaces: Knowledge acquisition process," *Expert Systems with Applications*, vol. 214, p. 119104, 2023.
- [5] N. Wang, F. Tao, Z. Fu, and S. Song, "Adaptive fuzzy control for a class of stochastic strict feedback high-order nonlinear systems with full-state constraints," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 1, pp. 205–213, 2020.
- [6] Y. Zheng, Z. Xu, and X. Wang, "The fusion of deep learning and fuzzy systems: A state-of-the-art survey," *IEEE Transactions on Fuzzy Systems*, vol. 30, no. 8, pp. 2783–2799, 2021.
- [7] Z. Bian, Q. Chang, J. Wang, W. Pedrycz, and N. R. Pal, "Takagi-sugeno-kang fuzzy systems for high-dimensional multi-label classification," *IEEE Transactions on Fuzzy Systems*, pp. 1–16, 2024.
- [8] F. Mahan, M. Mohammadzad, S. M. Rozekhani, and W. Pedrycz, "Chimflxdt: Chi-square-based multi flexible fuzzy decision tree for data stream classification," *Applied Soft Computing*, vol. 105, p. 107301, 2021.
- [9] D. Zhang, Y. Xu, Y. Peng, C. Du, N. Wang, M. Tang, L. Lu, and J. Liu, "An interpretable station delay prediction model based on graph community neural network and time-series fuzzy decision tree," *IEEE Transactions on Fuzzy Systems*, vol. 31, no. 2, pp. 421–433, 2022.
- [10] G. Chen and S. Li, "Research on location fusion of spatial geological disaster based on fuzzy svm," *Computer Communications*, vol. 153, pp. 538–544, 2020.
- [11] M.-J. Hsu, Y.-H. Chien, W.-Y. Wang, and C.-C. Hsu, "A convolutional fuzzy neural network architecture for object classification with small training database," *International Journal of Fuzzy Systems*, vol. 22, pp. 1–10, 2020.
- [12] B. Cao, J. Zhao, Z. Lv, Y. Gu, P. Yang, and S. K. Halgamuge, "Multiobjective evolution of fuzzy rough neural network via distributed parallelism for stock prediction," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 5, pp. 939–952, 2020.
- [13] J. Niu, D. Chen, J. Li, and H. Wang, "Fuzzy rule-based classification method for incremental rule learning," *IEEE Transactions on Fuzzy Systems*, vol. 30, no. 9, pp. 3748–3761, 2021.
- [14] J. A. Sanz and H. Bustince, "A wrapper methodology to learn interval-valued fuzzy rule-based classification systems," *Applied Soft Computing*, vol. 104, p. 107249, 2021.
- [15] X. Zhu, D. Wang, W. Pedrycz, and Z. Li, "Design and development of granular fuzzy rule-based models for knowledge transfer," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 2, pp. 704–715, 2022.
- [16] P. Singh, N. R. Pal, S. Verma, and O. P. Vyas, "Fuzzy rule-based approach for software fault prediction," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 5, pp. 826–837, 2016.
- [17] L. Jiao, T. Denoeux, and Q. Pan, "A hybrid belief rule-based classification system based on uncertain training data and expert knowledge," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 46, no. 12, pp. 1711–1723, 2015.
- [18] J.-B. Yang, J. Liu, J. Wang, H.-S. Sii, and H.-W. Wang, "Belief rule-base inference methodology using the evidential reasoning approach-rimer," *IEEE Transactions on systems, Man, and Cybernetics-part A: Systems and Humans*, vol. 36, no. 2, pp. 266–285, 2006.

- [19] Y. Li, E. Herrera-Viedma, I. J. Pérez, M. Barragán-Guzmán, and J. A. Morente-Molinera, "Z-number-valued rule-based classification system," *Applied Soft Computing*, vol. 137, p. 110168, 2023.
- [20] Y. Li, E. Herrera-Viedma, G. Kou, and J. A. Morente-Molinera, "Z-number-valued rule-based decision trees," *Information Sciences*, vol. 643, p. 119252, 2023.
- [21] R. A. Aliev, A. Alizadeh, R. R. Aliyev, and O. H. Huseynov, *Arithmetic of Z-numbers, the: theory and applications*. World Scientific, 2015.
- [22] L. A. Zadeh, "A note on z-numbers," *Information Sciences*, vol. 181, no. 14, pp. 2923–2932, 2011.
- [23] S. Xian, J. Chai, T. Li, and J. Huang, "A ranking model of z-mixture-numbers based on the ideal degree and its application in multi-attribute decision making," *Information Sciences*, vol. 550, pp. 145–165, 2021.
- [24] L. A. Zadeh, "Fuzzy sets," in *Fuzzy Sets, Fuzzy Logic, And Fuzzy Systems: Selected Papers by Lotfi A Zadeh*. World Scientific, 1996, pp. 394–432.
- [25] O. Tuncer and H. A. Cirpan, "Adaptive fuzzy based threat evaluation method for air and missile defense systems," *Information Sciences*, vol. 643, p. 119191, 2023.
- [26] J.-W. Xing, C. Peng, and X. Xie, "Switched fuzzy control for nonlinear systems via fuzzy-rule-dependent adaptive event-triggered mechanism," *IEEE Transactions on Fuzzy Systems*, vol. 32, no. 3, pp. 870–882, 2024.
- [27] R. Li, Z. Feng, and H. Guo, "Data mining and analysis algorithm of smart city network information resource description framework based on fuzzy association rules," *Journal of Testing and Evaluation*, vol. 51, no. 3, pp. 1386–1397, 2023.
- [28] A. Zhang, F. Gao, M. Yang, and W. Bi, "A new rule reduction and training method for extended belief rule base based on dbscan algorithm," *International Journal of Approximate Reasoning*, vol. 119, pp. 20–39, 2020.
- [29] W. Zhu, L. Chang, J. Sun, G. Wu, X. Xu, and X. Xu, "Parallel multipopulation optimization for belief rule base learning," *Information Sciences*, vol. 556, pp. 436–458, 2021.
- [30] L. Chang, J. Sun, J. Jiang, and M. Li, "Parameter learning for the belief rule base system in the residual life probability prediction of metalized film capacitor," *Knowledge-Based Systems*, vol. 73, pp. 69–80, 2015.
- [31] J. Hühn and E. Hüllermeier, "Furia: an algorithm for unordered fuzzy rule induction," *Data Mining and Knowledge Discovery*, vol. 19, pp. 293–319, 2009.
- [32] J. Alcalá-Fdez, L. Sanchez, S. Garcia, M. J. del Jesus, S. Ventura, J. M. Garrell, J. Otero, C. Romero, J. Bacardit, V. M. Rivas *et al.*, "Keel: a software tool to assess evolutionary algorithms for data mining problems," *Soft Computing*, vol. 13, no. 3, pp. 307–318, 2009.
- [33] Z. Chi, H. Yan, and T. Pham, *Fuzzy algorithms: with applications to image processing and pattern recognition*. World Scientific, 1996, vol. 10.
- [34] H. Ishibuchi and T. Yamamoto, "Rule weight specification in fuzzy rule-based classification systems," *IEEE transactions on fuzzy systems*, vol. 13, no. 4, pp. 428–435, 2005.
- [35] A. Calzada, J. Liu, H. Wang, and A. Kashyap, "A new dynamic rule activation method for extended belief rule-based systems," *IEEE Transactions on knowledge and data engineering*, vol. 27, no. 4, pp. 880–894, 2014.
- [36] W. W. Cohen, "Fast effective rule induction," in *Machine learning proceedings 1995*. Elsevier, 1995, pp. 115–123.
- [37] J. Alcalá-Fdez, R. Alcalá, and F. Herrera, "A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning," *IEEE Transactions on Fuzzy systems*, vol. 19, no. 5, pp. 857–872, 2011.



Yangxue Li received a bachelor's degree from the School of Network Security, Chengdu University of Technology, in 2018 and a master's degree from the Institute of Fundamental and Frontier Science, University of Electronic Science and Technology of China, in 2021. She is currently pursuing a Ph.D. degree in Department of Computer Science and Artificial Intelligence, School of Computer and Telecommunication Engineering, University of Granada, Spain. Li's research interests include Dempster-Shafer evidence theory, fuzzy sets

theory and Z-numbers.



He is Associate Editor of Mathematics journal and author of more than 20 papers published in international journals. Three of his papers are classified as highly cited as well being in the top 1% of the most cited papers in its field. His h-index is 16 according to the Web of Science with more than 2.800 citations received.



He is Associate Editor of IEEE Transactions on Cybernetics, Journal of Intelligent & Fuzzy Systems, and International Journal of Information Technology & Decision Making. He has published extensively in leading international journals, and several of his papers (currently 10) are classified as Highly Cited Papers in Essential Science Indicators from Clarivate Analytics. His h-index is 33 and he presents more than 5500 citations in Web of Science from Clarivate Analytics. He has been identified as a Highly Cited Researcher by Clarivate Analytics from 2018 to 2022. His research interest includes fuzzy decision-making, decision support systems, consensus models, linguistic modelling, aggregation of information, digital libraries, web quality evaluation, and bibliometrics. Dr Cabrerizo's awards and honors include the ISDA 2009 Conference Best Paper Award, the ICSEE 2015 Conference Best Paper Award, the ICCCC2022 Conference Best Paper Award, the 2014 Herbert Simon Award for outstanding contribution to the International Journal of Information Technology & decision-making, and the 2016 Andrew P. Sage Best Transactions Paper Award, which recognizes the authors of the best paper published annually in IEEE Transactions on Systems, Man, and Cybernetics: Systems. In addition, in 2022, he was elected as IAITQM Fellow.



Juan Antonio Morente-Molinera was born in Granada, Spain, in 1986. He received the Ph.D. degree in information and communications technology from the University of Granada (UGR), Granada, Spain, in 2015 and the UGR best thesis prize in 2019. He is currently associate professor in the Department of Computer Science and Artificial Intelligence, on the University of Granada. He is a member of the Soft Computing and Intelligent Information Systems (SCI2S) Research Group. His h-index is 21 and he has published more than 40 papers in JCR journals. He is working in the decision making field trying to solve the challenges that the new era of ubiquitous computing brings to the decision support systems field. His research interests include group decision making, decision support systems, fuzzy ontologies, consensus models, linguistic modeling, and sentiment analysis.

Ignacio Javier Pérez was born in Granada, Spain, in 1984. He received the M.Sc. and Ph.D. degrees in computer science from the University of Granada, Spain, in 2007 and 2011, respectively. He is currently a Professor with the Department of Computer Sciences and Artificial Intelligence, University of Granada. His current research interests include group decision making, consensus models, linguistic modeling, aggregation of information, information retrieval, digital libraries, web quality evaluation, recommender systems, e-health, and social media.

Bibliography

- [AAA⁺24] Abbasi S. N., Ashraf S., Akram M., Jana C., Simic V., and Pamucar D. (2024) A pythagorean fuzzy z-number-based neutrality aggregation model for ai-enabled energy efficiency management. *Applied Soft Computing* 161: 111753.
- [ABG24] Aliev R. A., Babanli M., and Guirimov B. G. (2024) Z-number based neural network structured inference system. *Information Sciences* 671: 120341.
- [AKKJG23] Alam N. M. F. H. N. B., Ku Khalif K. M. N., Jaini N. I., and Gegov A. (2023) The application of z-numbers in fuzzy decision making: the state of the art. *Information* 14(7): 400.
- [BK24] Babanli K. M. and Kabaoglu R. O. (2024) Synchronization of fuzzy-chaotic systems with z-controller in secure communication. *Information Sciences* 657: 119988.
- [BPP21] Banerjee R., Pal S. K., and Pal J. K. (2021) A decade of the z-numbers. *IEEE Transactions on Fuzzy Systems* 30(8): 2800–2812.
- [CCWM23] Chen B., Cai Q., Wei G., and Mo Z. (2023) Novel aczel-alsina operations-based linguistic z-number aggregation operators and their applications in multi-attribute group decision-making process. *Engineering Applications of Artificial Intelligence* 124: 106541.
- [Coh95] Cohen W. W. (1995) Fast effective rule induction. In *Machine learning proceedings 1995*, pp. 115–123. Elsevier.
- [DFPR22] Dubois D., Faux F., Prade H., and Rico A. (2022) Qualitative capacities: Basic notions and potential applications. *International Journal of Approximate Reasoning* 148: 253–290.
- [DRC⁺23] Dong Y., Ran Q., Chao X., Li C., and Yu S. (2023) Personalized individual semantics learning to support a large-scale linguistic consensus process. *ACM Transactions on Internet Technology* 23(2): 1–27.
- [Fel08] Feller W. (2008) *An introduction to probability theory and its applications*, volumen 2. John Wiley & Sons.
- [HGK24] Hosseini S., Gordan B., and Kalkan E. (2024) Development of z number-based fuzzy inference system to predict bearing capacity of circular foundations. *Artificial Intelligence Review* 57(6): 146.
- [HH09] Hühn J. and Hüllermeier E. (2009) Furia: an algorithm for unordered fuzzy rule induction. *Data Mining and Knowledge Discovery* 19: 293–319.

- [HHVM00] Herrera F., Herrera-Viedma E., and Martínez L. (2000) A fusion approach for managing multi-granularity linguistic term sets in decision making. *Fuzzy sets and systems* 114(1): 43–58.
- [HK82] Higashi M. and Klir G. J. (1982) Measures of uncertainty and information based on possibility distributions. *International journal of general systems* 9(1): 43–58.
- [HL22] Hu Z. and Lin J. (2022) An integrated multicriteria group decision making methodology for property concealment risk assessment under z-number environment. *Expert Systems with Applications* 205: 117369.
- [HYSHK25] Haseli G., Yazdani M., Shaayesteh M. T., and Hajiaghaei-Keshteli M. (2025) Logistic hub location problem under fuzzy extended z-numbers to consider the uncertainty and reliable group decision-making. *Applied Soft Computing* 171: 112751.
- [INT92] Ishibuchi H., Nozaki K., and Tanaka H. (1992) Distributed representation of fuzzy rules and its application to pattern classification. *Fuzzy sets and systems* 52(1): 21–32.
- [JC21] Jing F. and Chao X. (2021) Fairness concern: An equilibrium mechanism for consensus-reaching game in group decision-making. *Information Fusion* 72: 147–160.
- [JLZ24] Jin X., Liao H., and Zhang Z. (2024) A z-number-based three-way decision method with classification-based state determination for the evaluation of new energy enterprises. *Applied Soft Computing* 167: 112489.
- [Li24] Li D. (2024) A linguistic z-number-based dual perspectives information volume calculation method for driving behavior risk evaluation. *Expert Systems with Applications* 257: 124992.
- [LL02] Liu B. and Liu Y.-K. (2002) Expected value of fuzzy variable and fuzzy expected value models. *IEEE transactions on Fuzzy Systems* 10(4): 445–450.
- [LLWAB23] Liu F., Liao H., Wu X., and Al-Barakati A. (2023) Evaluating internet hospitals by a linguistic z-number-based gained and lost dominance score method considering different risk preferences of experts. *Information Sciences* 630: 647–668.
- [LZY⁺24] Luo N., Zhang Q., Yin L., Xie Q., Wu C., and Wang G. (2024) Three-way multi-attribute decision-making under the double hierarchy hesitant fuzzy linguistic information system. *Applied Soft Computing* 154: 111315.
- [MMKS⁺21] Morente-Molinera J. A., Kou G., Samuylov K., Cabrerizo F., and Herrera-Viedma E. (2021) Using argumentation in expert’s debate to analyze multi-criteria group decision making method results. *Information Sciences* 573: 433–452.
- [QCJ93] Quinlan J. R. and Cameron-Jones R. M. (1993) Foil: A midterm report. In *European conference on machine learning*, pp. 1–20. Springer.
- [Qui90] Quinlan J. R. (1990) Learning logical definitions from relations. *Machine learning* 5: 239–266.
- [QWX24] Qiu D., Wang C., and Xie J. (2024) A novel todim-based multi-attribute decision making method under information described by z-numbers for selecting online b&b. *Information Sciences* 663: 120315.

- [UAA25] Ullah I., Akram M., and Allahviranloo T. (2025) Multi-criteria decision making with hamacher aggregation operators based on multi-polar fuzzy z-numbers. *Information Sciences* 694: 121707.
- [WCZ17] Wang J.-q., Cao Y.-x., and Zhang H.-y. (2017) Multi-criteria decision-making method based on distance measure and choquet integral for linguistic z-numbers. *Cognitive Computation* 9: 827–842.
- [WLPZ15] Wang X., Liu X., Pedrycz W., and Zhang L. (2015) Fuzzy rule based decision trees. *Pattern Recognition* 48(1): 50–59.
- [WWC⁺23] Wang S., Wu J., Chiclana F., Ji F., and Fujita H. (2023) Global feedback mechanism by explicit and implicit power for group consensus in social network. *Information Fusion* page 102205.
- [WWTZ18] Wang J., Wang J.-q., Tian Z.-p., and Zhao D.-y. (2018) A multihesitant fuzzy linguistic multicriteria decision-making approach for logistics outsourcing with incomplete weight information. *International Transactions in Operational Research* 25(3): 831–856.
- [XCLH21] Xian S., Chai J., Li T., and Huang J. (2021) A ranking model of z-mixture-numbers based on the ideal degree and its application in multi-attribute decision making. *Information Sciences* 550: 145–165.
- [XWZ22] Xiao J., Wang X., and Zhang H. (2022) Exploring the ordinal classifications of failure modes in the reliability management: An optimization-based consensus model with bounded confidences. *Group Decision and Negotiation* 31(1): 49–80.
- [Yag95] Yager R. R. (1995) An approach to ordinal decision making. *International Journal of Approximate Reasoning* 12(3-4): 237–261.
- [Zad75a] Zadeh L. A. (1975) The concept of a linguistic variable and its application to approximate reasoning–ii. *Information sciences* 8(4): 301–357.
- [Zad75b] Zadeh L. A. (1975) The concept of a linguistic variable and its application to approximate reasoning–iii. *Information sciences* 9(1): 43–80.
- [Zad75c] Zadeh L. A. (1975) The concept of a linguistic variable and its application to approximate reasoning–i. *Information sciences* 8(3): 199–249.
- [Zad78] Zadeh L. A. (1978) Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and systems* 1(1): 3–28.
- [Zad96] Zadeh L. A. (1996) Fuzzy sets. In *Fuzzy Sets, Fuzzy Logic, And Fuzzy Systems: Selected Papers by Lotfi A Zadeh*, pp. 394–432. World Scientific.
- [Zad11] Zadeh L. A. (2011) A note on z-numbers. *Information Sciences* 181(14): 2923–2932.
- [ZDX⁺20] Zhang H., Dong Y., Xiao J., Chiclana F., and Herrera-Viedma E. (2020) Personalized individual semantics-based approach for linguistic failure modes and effects analysis with incomplete preference information. *Iise Transactions* 52(11): 1275–1296.

- [ZHMY24] Zhang Z., Hosseini S., Monjezi M., and Yari M. (2024) Extension of reliability information of z-numbers and fuzzy cognitive map: Development of causality-weighted rock engineering system to predict and risk assessment of blast-induced rock size distribution. *International Journal of Rock Mechanics and Mining Sciences* 178: 105779.
- [ZLWH24] Zheng Q., Liu X., Wang W., and Han S. (2024) A hybrid hfacs model using dematel-oreste method with linguistic z-number for risk analysis of human error factors in the healthcare system. *Expert Systems with Applications* 235: 121237.
- [ZXL⁺24] Zheng Y., Xu Z., Li Y., Pedrycz W., and Yi Z. (2024) Bi-objective optimization method for large-scale group decision making based on hesitant fuzzy linguistic preference relations with granularity levels. *IEEE Transactions on Fuzzy Systems* 32(8): 4759–4771.
- [ZZC⁺22] Zhang H., Zhu W., Chen X., Wu Y., Liang H., Li C.-C., and Dong Y. (2022) Managing flexible linguistic expression and ordinal classification-based consensus in large-scale multi-attribute group decision making. *Annals of Operations Research* pp. 1–54.
- [ZZL⁺23] Zhou M., Zhou Y.-J., Liu X.-B., Wu J., Fujita H., and Herrera-Viedma E. (2023) An adaptive two-stage consensus reaching process based on heterogeneous judgments and social relations for large-scale group decision making. *Information Sciences* page 119280.