

A Preliminary Study on the Use of Fuzzy Rough Set Based Feature Selection for Improving Evolutionary Instance Selection Algorithms

Joaquín Derrac¹, Chris Cornelis², Salvador García³, and Francisco Herrera¹

¹ Dept. of Computer Science and Artificial Intelligence, CITIC-UGR (Research Center on Information and Communications Technology),
University of Granada, 18071 Granada, Spain

`jderrac@decsai.ugr.es`, `herrera@decsai.ugr.es`

² Dept. of Applied Mathematics and Computer Science. Ghent University,
Gent, Belgium

`chris.cornelis@ugent.be`

³ Dept. of Computer Science. University of Jaén, 23071 Jaén, Spain
`sglopez@ujaen.es`

Abstract. In recent years, the increasing interest in fuzzy rough set theory has allowed the definition of novel accurate methods for feature selection. Although their stand-alone application can lead to the construction of high quality classifiers, they can be improved even more if other preprocessing techniques, such as instance selection, are considered.

With the aim of enhancing the nearest neighbor classifier, we present a hybrid algorithm for instance and feature selection, where evolutionary search in the instances' space is combined with a fuzzy rough set based feature selection procedure. The preliminary results, contrasted through nonparametric statistical tests, suggest that our proposal can improve greatly the performance of the preprocessing techniques in isolation.

Keywords: Fuzzy Rough Sets, Evolutionary Algorithms, Instance Selection, Feature Selection, Nearest Neighbor Classifier.

1 Introduction

Data reduction is a data preprocessing task which can be applied to ease the problem of dealing with large amounts of data. Its main objective is to reduce the original data by selecting the most representative information. In this way, it is possible to avoid excessive storage and time complexity, improving the results obtained by any data mining application. The best known data reduction processes are Feature Selection (FS)[9], Feature Extraction, Attribute Discretization, Instance Generation and Instance Selection (IS)[8].

Recently, Rough Set Theory (RST) [10] has been employed to tackle the FS problem. This approach can be enhanced with the use of fuzzy logic, obtaining methods which offer a greater flexibility and better potential to produce high-quality feature subsets than the crisp ones [3]. On the other hand, Evolutionary

Algorithms [6] have been successfully applied in IS problems due to the possibility of defining it as a search problem [5]. Both fields can offer suitable tools for enhancing the performance of machine learning methods.

In this work, a new hybrid approach considering both fuzzy-RST based FS and evolutionary IS is presented, EIS-RFS (Evolutionary IS enhanced by RST-based FS). A steady-state Genetic Algorithm (GA) is used to conduct the search of instances, whereas features are selected by a fuzzy-RST based method. The features selected are considered within the framework of the GA, thus modifying the environment in which the instances are chosen. At the end of its application, EIS-RFS considers the best subsets found to build a reduced version of the training set, well suited to be used as a reference set for the 1 Nearest Neighbor classifier (1-NN). This method is compared with their main components in isolation and the 1-NN classifier without preprocessing. The results achieved are contrasted by using nonparametric statistical techniques, reinforcing the conclusions obtained.

The rest of this work is organized as follows. Section 2 gives some background about evolutionary IS and fuzzy RST based FS. Section 3 describes the main characteristics of IFS-RTS. Section 4 describes the experimental study performed and shows the results achieved. Finally, Section 5 summarizes our conclusions.

2 Background and Related Work

This section is focused on two topics: IS and FS as data reduction techniques (Section 2.1), and the use of fuzzy RST for FS (Section 2.2).

2.1 Instance and Feature Selection

The IS objective is to isolate the smallest set of instances which enable a data mining algorithm to predict the class of a instance with the same quality as the initial data set [8]. By minimizing the data set, it is possible to reduce the space complexity and decrease the computational cost of the data mining algorithms that will be applied, improving their generalization capabilities.

More specifically, IS can be defined as follows: Let $(\mathcal{X}, \mathcal{A})$ be an information system, where $\mathcal{X} = \{x_1, \dots, x_n\}$ and $\mathcal{A} = \{a_1, \dots, a_m\}$ are finite, non-empty sets of instances and features. Then, let us assume that there is a training set TR which consists of N instances, and a test set TS composed of T instances ($TR \cup TS = (\mathcal{X}, \mathcal{A})$). Let $S \subseteq TR$ be the subset of selected samples that resulted from the execution of an IS algorithm, then we classify a new pattern T from TS by a data mining algorithm acting over the instances of S .

In the data mining field many approaches of evolutionary IS have been developed [7,5]. The interest in this field was increased by the study performed by Cano et al. [2], where a complete study of the use of evolutionary algorithms in IS is made. They concluded that evolutionary algorithms outperform classical algorithms both in reduction rates and classification accuracy.

On the other hand, in FS the goal is to select the most appropriate subset of features from the initial data set. It aims to eliminate irrelevant and/or redundant features to obtain a simple and accurate classification system [9]. Starting from the definition given for IS, FS can be defined as follows: Let us assume \mathcal{A}, \mathcal{X} , TR and TS have been already defined. Let $B \subseteq \mathcal{A}$ be the subset of selected features that resulted from the execution of a FS algorithm over TR , then we classify a new pattern from TS by a data mining algorithm acting over TR , employing as a reference only the features selected in B .

As with IS methods, a great number of FS methods have been developed recently. Some complete surveys, analyzing both classical and advanced approaches to FS, can be found in the literature [9]. Some advanced methods, combining both evolutionary IS and FS, have been also developed [4].

2.2 Fuzzy RST for FS

In rough set analysis [10], each attribute a in \mathcal{A} corresponds to an $\mathcal{X} \rightarrow V_a$ mapping, in which V_a is the value set of a over \mathcal{X} . For every subset B of \mathcal{A} , the B -indiscernibility relation R_B is defined as

$$R_B = \{(x, y) \in \mathcal{X}^2 \text{ and } (\forall a \in B)(a(x) = a(y))\} \tag{1}$$

Therefore, R_B is an equivalence relation. Its equivalence classes $[x]_{R_B}$ can be used to approximate concepts, in other words, subsets of the universe \mathcal{X} . Given $A \subseteq \mathcal{X}$, its lower and upper approximation with respect to R_B are defined by

$$R_B \downarrow A = \{x \in \mathcal{X} | [x]_{R_B} \subseteq A\} \text{ and } R_B \uparrow A = \{x \in \mathcal{X} | [x]_{R_B} \cap A \neq \emptyset\} \tag{2}$$

A *decision system* $(\mathcal{X}, \mathcal{A} \cup \{d\})$ is a special kind of information system, used in the context of classification, in which d ($d \notin \mathcal{A}$) is a designated attribute called the decision attribute. Its equivalence classes $[x]_{R_d}$ are called decision classes. Given $B \subseteq \mathcal{A}$, the B -positive region POS_B contains those objects from X for which the values of B allow to predict the decision class unequivocally:

$$POS_B = \bigcup_{x \in X} R_B \downarrow [x]_{R_d} \tag{3}$$

Indeed, if $x \in POS_B$, it means that whenever an instance has the same values as x for the attributes in B , it will also belong to the same decision class as x . The predictive ability with respect to d of the attributes in B is then measured by the following value (degree of dependency of d on B):

$$\gamma_B = \frac{|POS_B|}{|\mathcal{X}|} \tag{4}$$

Instead of using a crisp equivalence relation R to represent objects' indiscernibility, we can also measure it by means of a fuzzy relation R . Typically, we assume that R is at least a fuzzy tolerance relation (reflexive and symmetric).

Assuming that for a qualitative attribute a , the classical way of discerning objects is used, that is, $R_a(x, y) = 1$ if $a(x) = a(y)$ and $R_a(x, y) = 0$ otherwise, we can define, for any subset B of \mathcal{A} , the fuzzy B-indiscernibility relation by

$$R_B(x, y) = \mathcal{T}(R_a(x, y)), a \in B \tag{5}$$

in which \mathcal{T} represents a t-norm. It can be seen that if only qualitative attributes are used, then the traditional concept of B-indiscernibility relation is recovered.

For the lower and upper approximation of a fuzzy set A in X by means of a fuzzy tolerance relation R , the formulas defined in (2) are paraphrased (given the Lukasiewicz implicator $\mathcal{I}(x, y) = \min(1, 1 - x + y)$ and the minimum t-norm $\mathcal{T}(x, y) = \min(x, y), x, y \in [0, 1]$) to define $R \downarrow A$ and $R \uparrow A$, for all y in X , by

$$(R \downarrow A)(Y) = \inf_{x \in X} \mathcal{I}(R(x, y), A(x)) \quad (R \uparrow A)(Y) = \sup_{x \in X} \mathcal{T}(R(x, y), A(x)) \tag{6}$$

Using fuzzy B-indiscernibility relations, the fuzzy B-positive region is defined by

$$POS_B(y) = \left(\bigcup_{x \in X} R_B \downarrow [\mathcal{X}_{R_d}] \right) (y) \tag{7}$$

Once we have fixed the fuzzy positive region, we can define an increasing $[0, 1]$ -valued measure to gauge the degree of dependency of a subset of features on another subset of features. For FS it is useful to phrase this in terms of the dependency of the decision feature on a subset of the conditional features:

$$\gamma_b = \frac{|POS_B|}{|POS_{\mathcal{A}}|} \tag{8}$$

3 EIS-RFS: Instance and Feature Selection Enhanced by Rough Set Theory

This section is devoted to describe EIS-RFS. Section 3.1 describes the steady-state GA employed for performing IS and the fuzzy RST based FS method. Section 3.2 shows the full model combining both techniques.

3.1 Basic Techniques of EIS-RFS

The IS component of EIS-RFS is guided by a steady-state GA where only two offspring are produced in each generation. Binary codification, binary tournament selection procedure, a two point crossover operator and a bit-flip mutation operator are considered. Concerning the fitness function, it should pursue both reduction and accuracy objectives. To do so, we will follow the proposal given in [2], where Cano et al. defined *AccRate* as the accuracy achieved by a 1-NN classifier when classifying the entire training set, using the currently selected subset as a reference and using leave-one-out as validation scheme, *RedRate* as the

reduction rate achieved over the currently selected instances, and a real-valued weighting factor, α , to adjust the strength of each term in the resulting fitness value. Equation 9 defines it, where J is an IS chromosome to be evaluated.

$$Fitness(J) = \alpha \cdot AccRate(J) + (1 - \alpha) \cdot RedRate(J) \quad (9)$$

Following the recommendations given in [2], EIS-RFS will employ a value $\alpha = 0.5$, which should offer an adequate trade-off between accuracy and reduction.

The fuzzy RST based FS method is taken from [3], where a hill climbing heuristic (quickreduct heuristic) is used for searching iteratively subsets of features maximizing the gamma measure (Equation 8). The similarity measure selected for quantitative values is

$$R_a(x, y) = \max \left(\min \left(\left(\frac{a(y) - a(x) + \sigma_a}{\sigma_a}, \frac{a(x) - a(y) + \sigma_a}{\sigma_a} \right), 0 \right) \right) \quad (10)$$

where x and y are two different instances belonging to the training set, and σ_a denotes the standard deviation of a . For nominal attributes we use the Value Difference Metric (VDM) [11], where two values are closer if they have more similar classifications (that is, more similar correlations with the output classes).

3.2 Hybrid Model for Simultaneous IS and FS

Once the two basic tools considered for performing IS and FS have been defined, the hybrid model which composes our approach can be described. Basically, it can be described as a steady-state GA for IS where, every time a fixed number of evaluations has been spent, a fuzzy-RST based FS procedure is applied to modify the features taken in consideration during the search.

1. **Initialization:** IS chromosomes are initialized randomly. Two different subsets are considered for the selection of the initial subset of features: The full set and the subset of features selected by the RST based FS method using as input the whole training set. The best performing subset is selected as the global subset of features of EIS-RFS.
2. **New IS generation:** An IS generation is carried out using the steady-state GA scheme. Note that, when evaluating a new chromosome, the 1-NN classifier used in the fitness function will only consider the selected features in the global subset of features of EIS-RFS.
3. **Update Features:** If the Stabilize phase has not been activated yet (see below), the procedure of update of features is called every time *UpdateFS* evaluations has been spent by the steady-state GA. This procedure consists of using the RST-based FS filter method, considering as input instance set the current best chromosome of the population. The new subset of features obtained is tested by applying it to a 1-NN classifier (considering as reference set only the current best subset of instances). If this subset performs better than the former, it is accepted as the global subset of features of EIS-RFS.

4. **Stabilize phase:** Changes in the current subset of features are only considered if the search is not near its end. Therefore, if the number of evaluations spent is higher than $\beta \cdot NEvaluations$, the stabilize stage is activated and no further changes in the subset of features selected are considered. This mechanism allows EIS-RFS to easily converge for hard problems, where the final subset of features is fixed before the end of the search. It allows EIS-RFS to focus its last efforts in optimizing the subsets of instances selected, performing a final refinement of the solutions achieved.
5. **Termination criterion:** The search process ends if EIS-RFS has spent $NEvaluations$. Otherwise, a new cycle of the algorithm begins.
6. **Output:** When the fixed number of evaluations runs out, the best chromosome of the population (a subset of instances) and the current global subset of selected features are returned as the output of EIS-RFS.

The resulting subsets of instances and features define a pruned version of the original training set. This set can be used as reference set by a 1-NN classifier to perform a faster and more accurate classification of new test instances.

4 Experimental Framework and Results

This section describes the experimental study conducted in this contribution. Data sets, comparison methods and parameters are detailed in Section 4.1. Results are shown and discussed in Section 4.2.

4.1 Experimental Framework

In our experiments, we have used 20 data sets taken from the KEEL-Datasets¹ [1] repository. Table 1 shows their main characteristics. For each data set, it is shown the number of instances, features and classes of the problem described. We have used a ten fold cross-validation (10-fcv) procedure as validation scheme.

As comparison methods, we have selected the techniques considered in the construction of EIS-RFS (a Steady-State GA for IS (IS-SSGA) and the fuzzy-RST based FS method (FS-RST)). The preprocessed data sets obtained as result of the application of them have been used as reference sets for a 1-NN, estimating accuracy. Furthermore, we have considered the inclusion of the 1-NN classifier as a baseline (using the full training set as reference). The parameters of each method are shown in Table 2.

Finally, we will employ the well-known Wilcoxon Signed-Ranks test for contrasting the accuracy results achieved. Further information about this test and other statistical procedures specifically designed for use in the field of Machine Learning can be found at the SCI2S thematic public website on *Statistical Inference in Computational Intelligence and Data Mining*².

¹ <http://www.keel.es/datasets.php>

² <http://sci2s.ugr.es/sicidm/>

Table 1. UCI Data sets used in our experiments

Data set	Instances	Features	Classes	Data set	Instances	Features	Classes
Australian	690	14	2	Housevotes	435	16	2
Balance	625	4	3	Iris	150	4	3
Bupa	345	6	2	Mammographic	961	5	2
Cleveland	303	13	5	Newthyroid	215	5	3
Contraceptive	1473	9	3	Pima	768	8	2
Ecoli	336	7	8	Sonar	208	60	2
German	1000	20	2	Tic-tac-toe	958	9	2
Glass	214	9	7	Wine	178	13	3
Hayes-roth	160	4	3	Wisconsin	699	9	2
Hepatitis	155	19	2	Zoo	101	16	7

Table 2. Parameter specification for the algorithms tested in the experimentation

Algorithm Parameters	
EIS-RFS	NEvaluations: 10000, Pop. size: 50, Cross. prob.: 1.0, Mutat. prob.: 0.005 per bit, α : 0.5 <i>MaxGamma</i> : 1.0, UpdateFS: 100, β : 0.75
IS-SSGA	NEvaluations: 10000, Pop. size: 50, Cross. prob.: 1.0, Mutat. prob.: 0.005 per bit, α : 0.5
FS-RST	<i>MaxGamma</i> : 1.0
1-NN	-

4.2 Results and Analysis

Table 3 shows the results measured by accuracy, ratio of reduction over instances (Reduction (IS)) and ratio of reduction over features (Reduction (FS)) in test data. For each data set, the best result in accuracy is highlighted in **bold**.

As can be seen in the table, EIS-RFS achieves the best average accuracy result in test phase. For contrasting this fact, we have carried out a Wilcoxon Signed-Ranks test, the results of which are summarized in Table 4.

Reading Tables 3 and 4, we can make the following analysis:

- In accuracy, EIS-RFS outperforms the rest of algorithms on 14 of 20 data sets, and has the best average result. This superiority is remarked as significant by the results of the Wilcoxon Signed-Ranks test, the results of which show that EIS-RFS outperforms all the comparison methods with a level of significance $\alpha = 0.01$. This is a strong result, which supports the fact that EIS-RFS clearly outperforms all the other techniques in accuracy.
- Concerning reduction in instances' space, EIS-RFS achieves slightly better rates than IS-SSGA. Therefore, our approach is able to effectively reduce the training set without harming the accuracy of the 1-NN rule. Moreover, its accuracy is even increased, exceeding the IS-SSGA method. Concerning reduction in features' space, it shows a similar behavior as FS-RST, although the features selected are different (which is one of the reasons of the accuracy improvement of EIS-RFS over FS-RST).

Table 3. Results obtained

Measure	Accuracy				Reduction (IS)		Reduction (FS)	
	Data set	EIS-RFS	IS-SSGA	FS-RST	1-NN	EIS-RFS	IS-SSGA	EIS-RFS
Australian	85.66	85.65	81.45	81.45	0.8872	0.8799	0.1571	0.0000
Balance	85.92	86.40	79.04	79.04	0.8464	0.8686	0.0000	0.0000
Bupa	65.72	61.14	62.51	61.08	0.8502	0.8644	0.0000	0.1274
Cleveland	55.16	52.82	52.51	53.14	0.9014	0.9171	0.0462	0.3908
Contraceptive	45.42	44.54	42.63	42.77	0.7637	0.7530	0.0667	0.0360
Ecoli	82.14	80.38	76.58	80.70	0.8882	0.9077	0.1286	0.2286
German	70.80	70.40	67.90	70.50	0.8014	0.7914	0.2350	0.1450
Glass	67.35	67.10	74.50	73.61	0.8718	0.8791	0.0444	0.0168
Hayes-roth	80.86	69.15	76.07	35.70	0.8544	0.8384	0.2500	0.1000
Hepatitis	82.58	79.33	79.50	82.04	0.9262	0.9226	0.5368	0.4263
Housevotes	94.48	93.79	90.78	91.24	0.9387	0.9410	0.3500	0.0188
Iris	96.00	94.67	93.33	93.33	0.9511	0.9481	0.1250	0.0000
Mammographic	80.65	79.50	75.76	76.38	0.8322	0.8229	0.0000	0.3396
Newthyroid	96.77	98.16	97.23	97.23	0.9473	0.9571	0.0600	0.0000
Pima	74.80	72.26	70.33	70.33	0.7911	0.8187	0.0000	0.0000
Sonar	80.76	75.45	81.69	85.55	0.8899	0.8595	0.2900	0.7183
Tic-tac-toe	78.29	78.71	73.07	73.07	0.8655	0.7917	0.0000	0.0000
Wine	97.19	92.68	95.49	95.52	0.9451	0.9538	0.3308	0.5231
Wisconsin	96.42	96.13	95.57	95.57	0.9103	0.9027	0.0444	0.0000
Zoo	96.39	94.22	96.50	92.81	0.8634	0.8714	0.2125	0.2750
Average	80.67	78.63	78.12	76.55	0.8763	0.8745	0.1439	0.1673

Table 4. Wilcoxon Signed-Ranks Test results

Comparison	R^+	R^-	P-value
EIS-RFS vs IS-SSGA	188	22	0.0010
EIS-RFS vs FS-RST	183	27	0.0023
EIS-RFS vs 1-NN	174	36	0.0083

These results confirm the benefits of the hybridization of fuzzy-RST based FS and evolutionary IS techniques as a whole, highlighting EIS-RFS as a suitable preprocessing method for greatly reducing the size of the training set (>87%, on average) and enhancing the accuracy of the 1-NN classifier.

5 Conclusions

In this work we have presented EIS-RFS, a novel approach for simultaneous IS and fuzzy-RST FS. This approach includes the features selected by a fuzzy-RST based FS method inside a evolutionary IS search process, thus combining the benefits of both techniques into a stand-alone, yet accurate, procedure.

The experimental results show that our approach improves the accuracy achieved by the considered techniques in isolation, whereas reduction rates are maintained. Nonparametric statistical procedures confirm that EIS-RFS can be considered as a suitable tool for optimizing the 1-NN classifier.

Acknowledgements

This work was supported by Project TIN2008-06681-C06-01. J. Derrac holds a FPU scholarship from Spanish Ministry of Education. Chris Cornelis would like to thank the Research Foundation - Flanders for funding his research.

References

1. Alcalá-Fdez, J., Fernández, A., Luengo, J., Derrac, J., García, S., Sánchez, L., Herrera, F.: Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. *Journal of Multiple-Valued Logic and Soft Computing* (2010) (in press)
2. Cano, J.R., Herrera, F., Lozano, M.: Using evolutionary algorithms as instance selection for data reduction in KDD: An experimental study. *IEEE Transactions on Evolutionary Computation* 7(6), 561–575 (2003)
3. Cornelis, C., Jensen, R., Hurtado, G., Slezak, D.: Attribute selection with fuzzy decision reducts. *Information Sciences* 180, 209–224 (2010)
4. Derrac, J., García, S., Herrera, F.: IFS-CoCo: Instance and feature selection based on cooperative coevolution with nearest neighbor rule. *Pattern Recognition* 43(6), 2082–2105 (2010)
5. Derrac, J., García, S., Herrera, F.: A survey on evolutionary instance selection and generation. *International Journal of Applied Metaheuristic Computing* 1(1), 60–92 (2010)
6. Eiben, A.E., Smith, J.E.: *Introduction to Evolutionary Computing*. Natural Computing. Springer, Heidelberg (2003)
7. García, S., Cano, J.R., Herrera, F.: A memetic algorithm for evolutionary prototype selection: A scaling up approach. *Pattern Recognition* 41(8), 2693–2709 (2008)
8. Liu, H., Motoda, H. (eds.): *Instance Selection and Construction for Data Mining*. The Springer International Series in Engineering and Computer Science. Springer, Heidelberg (2001)
9. Liu, H., Motoda, H. (eds.): *Computational Methods of Feature Selection*. Chapman & Hall/Crc Data Mining and Knowledge Discovery Series. Chapman & Hall/CRC, Boca Raton (2007)
10. Pawlak, Z., Skowron, A.: Rudiments of rough sets. *Information Sciences* 177, 3–27 (2007)
11. Wilson, D.R., Martinez, T.: Improved heterogeneous distance functions. *Journal of Artificial Intelligence Research* 6, 1–34 (1997)