sustainability

MDPI

Article

A Textual Data-Oriented Method for Doctor Selection in Online Health Communities

Yinfeng Du [1], Zhen-Song Chen [2], Jie Yang [3,*], Juan Antonio Morente-Molinera [4], Lu Zhang [1] and Enrique Herrera-Viedma [4]

[1] School of Economics and Management, Southwest Jiaotong University, Chengdu 610031, China
[2] School of Civil Engineering, Wuhan University, Wuhan 430072, China
[3] School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 611756, China
[4] Andalusian Research Institute in Data Science and Computational Intelligence, University of Granada, 18071 Granada, Spain
* Correspondence: yj@my.swjtu.edu.cn

Abstract: As doctor–patient interactive platforms, online health communities (OHCs) offer patients massive information including doctor basic information and online patient reviews. However, how to develop a systematic framework for doctor selection in OHCs according to doctor basic information and online patient reviews is a challenged issue, which will be explored in this study. For doctor basic information, we define the quantification method and aggregate them to characterize relative influence of doctors. For online patient reviews, data analysis techniques (i.e., topics extraction and sentiment analysis) are used to mine the core attributes and evaluations. Subsequently, frequency weights and position weights are respectively determined by a frequency-oriented formula and a position score-based formula, which are integrated to obtain the final importance of attributes. Probabilistic linguistic-prospect theory-multiplicative multiobjective optimization by ratio analysis (PL-PT-MULTIMOORA) is proposed to analyze patient satisfactions on doctors. Finally, selection rules are made according to doctor influence and patient satisfactions so as to choose optimal and suboptimal doctors for rational or emotional patients. The designed textual data-driven method is successfully applied to analyze doctors from Haodf.com and some suggestions are given to help patients pick out optimal and suboptimal doctors.

Keywords: online health communities; doctor selection; doctor influence; patient satisfactions; improved MULTIMOORA; selection criteria

Citation: Du, Y.; Chen, Z.-S.; Yang, J.; Morente-Molinera, J.A.; Zhang, L.; Herrera-Viedma, E. A Textual Data-Oriented Method for Doctor Selection in Online Health Communities. *Sustainability* **2023**, *15*, 1241. https://doi.org/10.3390/su15021241

Academic Editor: Norval Zhang

Received: 30 November 2022
Revised: 5 January 2023
Accepted: 5 January 2023
Published: 9 January 2023

1. Introduction

As doctor–patient interactive platforms, online health communities (OHCs), such as iWantGreatCare, Healthgrades, MedHelp, Vitals, Doximity, Haodaifu, Guahao, and Yelp have attracted millions of doctors and patients to communicate and interact more conveniently without temporal and spatial restrictions [1–4]. Patients enable to accomplish online medical search, online medical consultation, online medical appointment, and experience sharing in OHCs [5], while doctors have an access to share their medical knowledge and offer health service to more patients through OHCs. Existing literature on OHCs concentrates on tendency forecast of epidemic [6], knowledge-sharing [3], and group joining behavior [7]. Few works pay attention to evaluation and selection of doctors in OHCs [8–10]. Owing to medical information overload and the lack of other ways to get service quality information of doctors, it is more difficult for patients to select suitable doctors in OHCs [8]. Therefore, researching on doctor selection in OHCs would be a valuable and important task.

Usually, OHCs offer patients much information, such as doctor basic information, popular science knowledge, prescribing advice, and online patient reviews. In particular, doctor basic information and online patient reviews are undoubtedly significant basis for patients to pick out appropriate doctors. Nevertheless, existing literature [8–12], which mainly assesses doctors based on online patient reviews neglects the influence of doctor basic information on doctor selection. Doctor basic information, such as doctor title, number of honored the Annual Good Doctor, value of recommendation, satisfaction of online service, and number of online consultation is objective information, which can be quantified to explore the influence of each doctor. On the contrary, online patient reviews are relatively subjective judgements, which can be applied to analyze patient satisfactions regarding doctors. There are five challenges in doctor selection in OHCs. The first challenge is that how to develop a systematic framework to effectively and efficiently pick out optimal and suboptimal doctors for all patients. Actually, doctor titles are textual data, namely Assistant Doctor, Doctor, Attending Doctor, Associate Director Physician, and Director Physician. The number of times honored the Annual Good Doctor and the value of recommendation are positive numbers with minor deviation. Satisfaction of online service is expressed in the form of percentage. The number of online consultation is a number much larger than 10. Obviously, different types of results collected from basic doctor information should be unified and minor deviation among these five sets of values is necessary. Therefore, the second challenge is that how to quantify doctor basic information and characterize the relative influence of each doctor. In general, every online review implies the performance and ranking of attributes simultaneously [13]. Performance of attributes, being seemed as evaluation information, can be applied to analyze patient satisfactions, while ranking of attributes, being considered as position information, depicts the importance of attributes. Moreover, frequency of attributes in all online patient reviews also reflects the impact of frequency information on weights of attributes. So, how to obtain the importance of attributes through their frequency information and position information is the third challenge in this study.

Scholars explore patient satisfactions from different perspectives. Ref. [14–17] analyzed patient satisfactions in the light of empirical analysis. Researchers also explore patient satisfactions with some quantitative approaches. For example, Büyüközkan et al. [18] used questionnaire method and fuzzy analytic hierarchy process (FAHP) to evaluate health care service quality in Turkey from six dimensions (i.e., tangibles, responsiveness, reliability, assurance, empathy and professionalism). With the aid of online doctor–patient interaction from Haodf.com, Yang et al. [12] pointed out that online physician service quality has significantly affected the patient satisfaction, and the ratio of the number of positive ratings to the number of patients has been used to measure patient average satisfaction for a physician's overall service quality. Du and Liu [19] designed prospect probabilistic linguistic weighted Muirhead mean operator and improved prospect theory to assess the service quality of doctors, in which evaluation information was collected by market survey method. Distinctly, online review is a more time-saving and convenient way to collect massive evaluations when comparing with market survey techniques, which has been applied to opinion mining from online hotel reviews [20], analyzing effects of online reviews in a dual channel supply chain [21], dining sentiment analysis [22], analysis of passenger demands and passenger satisfactions for high-speed rail [23], service-quality measurement and improvements for hotels [24,25], online product recommendation [26], and identification and priority of user preferences [27]. Fortunately, several scholars explore patient satisfactions based on online patient reviews. For example, James et al. [9] applied a text mining methodology to textual feedbacks of physicians and related the textual commentary to their numeric ratings. Shah et al. [11] proposed a text mining approach to investigate the determinants of patient satisfaction and patient dissatisfaction across different types of diseases. Hu et al. [8] combined term frequency-inverse document frequency, intuitionistic fuzzy set (IFS), and VIKOR method to rank alternative doctors sourced from Haodf.com. Clearly, methods in prior works, such as FAHP [18], ratio-based method [12], improved prospect theory [19], and VIKOR method [8] are single decision-making approaches, which have poorer robustness than integrated methods, such as multiplicative multiobjective optimization by ratio analysis (MULTIMOORA) [28]. Lots of evidences [13,29–31] indicate that MULTIMOORA has excellent strengths in robustness, time-saving, simplicity, and mathematical computations. Probabilistic linguistic term set (PLTS) [32] is one of the most important linguistic forms, which is more suitable than IFS [8] to quantify unstructured textual reviews into structured data [10,33,34]. Since the advantages in expressing diversified preferences and their corresponding occurrence probabilities, Li et al. [10] used PLTS to quantify online patient reviews and developed PL-MULTIMOORA method to evaluate service quality of doctors. Although the method proposed by [10] has robustness, it still ignores the impact of psychological behavior of evaluators on decision-making outcomes. Actually, behavior characteristics of evaluators play an important role in the decision-making process. As one of the most influential psychological behavior decision theories, prospect theory (PT) [35,36] considers irrationality behavior of decision makers (DMs) in the whole decision-making process, which is more in step with incomplete rational behavior of DMs, especially in the condition of risk and uncertainty [19,37–39]. Therefore, the fourth challenge is how to evaluate patient satisfactions by integrating MULTIMOORA method and PT into probabilistic linguistic preference environment.

Usually, rational patients and emotional patients have different choices. Rational patients may prefer to choose suitable doctors in the light of doctor influence ranking (resulted from objective information), while emotional patients may consider patient satisfactions (resulted from relatively subjective judgements) more when they select appropriate doctors. So, the last challenge in this study is that how to design a reasonable selection criteria/rules for all patients to pick out optimal and suboptimal doctors based on doctor influence and patient satisfactions.

As aforementioned, this study mainly designs a systematic framework to select optimal and suboptimal doctors based on doctor basic information and online patient reviews no matter what patients are rational or emotional. Core contributions of this paper are outlined in the following. (1) A new textual data-oriented framwork is developed to select optimal and suboptimal doctors in OHCs with the aid of doctor basic information and online patient reviews. In particular, topics extraction, sentiment analysis and PLTSs are applied to excavate and quantify textual reviews; (2) Doctor basic information from OHCs is quantified and used to characterize the relative influence of doctors; (3) Frequency weights and position weights are fused to get the final importance of attributes. Specifically, frequency weights are obtained by a frequency-oriented formula while position weights are acquired through a position score-based formula; (4) An improved MULTIMOORA method considering psychological cognition of DMs (i.e., PL-PT-MULTIMOORA) is developed to analyze patient satisfactions; (5) According to doctor influence and patient satisfactions, selection criteria are introduced to help rational or emotional patients pick out the optimal and suboptimal doctors.

The rest of the paper is outlined as follows. Some basic definitions and concepts are described in Section 2. Section 3 introduces a systematic framework for doctor selection. In Section 4, an application of 10 doctors is constructed and discussed to illustrate the feasibility and effectiveness of proposed method. Finally, Section 5 concludes the paper and elaborates on future studies.

2. Preliminaries

Some definitions and concepts about PLTSs and MULTIMOORA are displayed in this section.

2.1. Probabilistic Linguistic Term Sets

Since the significant advantage in presenting diversified preference information with the aid of natural language and occurrence probabilities, probabilistic linguistic term set (PLTS) [32] has become one of the effective ways to quantify online reviews and been applied to online reviews-based management problems, such as green enterprise ranking [40], assessments of web celebrity shop [41] and O2O takeaway [34], television selection from Amazon.com [42], selection of tourists [33], and user satisfaction analysis [13]. In the following, we introduce the PLTSs with unbalanced linguistic term sets.

Definition 1 ([43])**.** *Let* $S^{(\psi)} = \left\{ S_\alpha^{(\psi)} \middle| \alpha = -(\psi-1), -\frac{2}{3}(\psi-2), \ldots, 0, \ldots, \frac{2}{3}(\psi-2), (\psi-1) \right\}$ *be an unbalanced linguistic term set (UBLTS), a PLTS with UBLTS* $L(P)$ *can be denoted as:*

$$L(P) = \left\{ L^{(k)}\left(P^{(k)}\right) \middle| L^{(k)} \in S^{(\psi)}, r^{(k)} \in \alpha, P^{(k)} \geqslant 0, k = 1, 2, \ldots, \#L(P), \sum_{k=1}^{\#L(P)} P^{(k)} \leqslant 1 \right\}. \quad (1)$$

where $L^{(k)}\left(P^{(k)}\right)$ *represents the linguistic term* $L^{(k)}$ *associated with the probability* $P^{(k)}$, $r^{(k)}$ *is subscript of* $L^{(k)}$ *and* $\#L(P)$ *is the number of all linguistic terms in* $L(P)$.

For PLTSs, [43] also developed their generalized distance measure as follows.

Definition 2 ([43])**.** *Let* $L_1(P) = \left\{ L_1^{(k)}\left(P_1^{(k)}\right) \middle| k = 1, \ldots, \#L_1(P) \right\}$ *and* $L_2(P) = \left\{ L_2^{(k)}\left(P_2^{(k)}\right) \middle| k = 1, \ldots, \#L_2(P) \right\}$ *be arbitrary two PLTSs along with descending order and satisfy* $\#L_1(P) = \#L_2(P)$*. The generalized distance measure of* $L_1(P)$ *and* $L_2(P)$ *is given as:*

$$d_{gd}(L_1(P), L_2(P)) = \left[\frac{1}{\#L_1(P)} \sum_{k=1}^{\#L_1(P)} \left(\frac{1}{2\psi - 1} \left| r_1^{(k)} P_1^{(k)} - r_2^{(k)} P_2^{(k)} \right|^\lambda \right) \right]^{\frac{1}{\lambda}}. \quad (2)$$

where $\#L_1(P)$ *displays the number of linguistic items in* $L_1(P)$*. The precondition of generalized distance measure is that keeps all PLTSs with the same number of linguistic terms (i.e.,* $\#L_1(P) = \#L_2(P)$*).* $2\psi - 1$ *be cardinality number of linguistic terms in S.* λ *is a positive integer. Especially, if* $\lambda = 1$ *or* $\lambda = 2$*, then generalized distance measure will separately reduce to the Hamming distance or Euclidean distance.*

2.2. MULTIMOORA Method

Multiobjective optimization by ratio analysis (MOORA) [44] is first proposed by Brauers and Zavadskas in 2006 and extended as multiplicative MOORA (MULTIMOORA) by same authors in 2010 [28]. Compared with MOORA, MULTIMOORA has additionally considered the influence of full-multiplicative form (FMF) on decision-making outcomes. In general, MULTIMOORA is a multiple attribute decision-making (MADM) method with wonderful robustness in which three subordinate approaches, such as the ratio system (RS), reference point (RP), and FMF are simultaneously involved [10,31]. Let $X = (x_{ij})_{m\times n}$ $(i = 1, \ldots, m; j = 1, \ldots, l, \ldots, n)$ be evaluation matrix for m alternatives regarding n attributes. The former l attributes are beneficial attributes, while the others are cost attributes.

The four core steps of MULTIMOORA method are described as follows.

- Step 1: Acquire dimensionless value $\bar{x}_{ij}$ by Formula (3);
- Step 2: Compute three kinds of utility values (i.e., $U_1(a_i)$, $U_2(a_i)$ and $U_3(a_i)$) through RS, RP and FMF models;
- Step 3: Rank $U_1(a_i)$ and $U_3(a_i)$ in descending order while $U_2(a_i)$ in ascending order;
- Step 4: Determine the final ranking by dominance theory [28].

$$\bar{x}_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} \left(x_{ij}\right)^2}}. \quad (3)$$

$$\begin{cases} \text{(RS model.) } U_1(a_i) = \sum_{j=1}^{l} \bar{x}_{ij} - \sum_{j=l+1}^{n} \bar{x}_{ij}; \\ \text{(RP model.) } U_2(a_i) = \max_j |x_j - \bar{x}_{ij}|; \\ \text{(FMF model.) } U_3(a_i) = \prod_{j=1}^{l} \bar{x}_{ij} \Big/ \prod_{j=l+1}^{n} \bar{x}_{ij}. \end{cases} \tag{4}$$

where $x_j = \max_i \bar{x}_{ij}$.

3. Methodology

As doctor–patient interactive platforms, OHCs offer patients much significant information, such as doctor basic information, popular science knowledge, prescribing advice, and online patient reviews. In particular, doctor basic information and online patient reviews are undoubtedly important basis for patients to select appropriate doctors. Doctor basic information can be quantified to explore the influence of each doctor, while patient reviews can be applied to analyze the patient satisfactions regarding doctors. Therefore, this study takes doctor basic information and online patient reviews as data sources and attempts to systematically design a data-driven framework for doctor selection in OHCs, in which data collection, data processing, and multiple participant multiple attribute decision making (MPMADM) process are included in the framework. Specifically, data collection and data processing are prerequisites for MPMADM process. MPMADM process mainly involves quantifying relative influence of doctors, determining attributes, evaluating patient satisfactions, and selecting criteria/rules. The flowchart of textual data-driven method for doctor selection is presented in Figure 1.

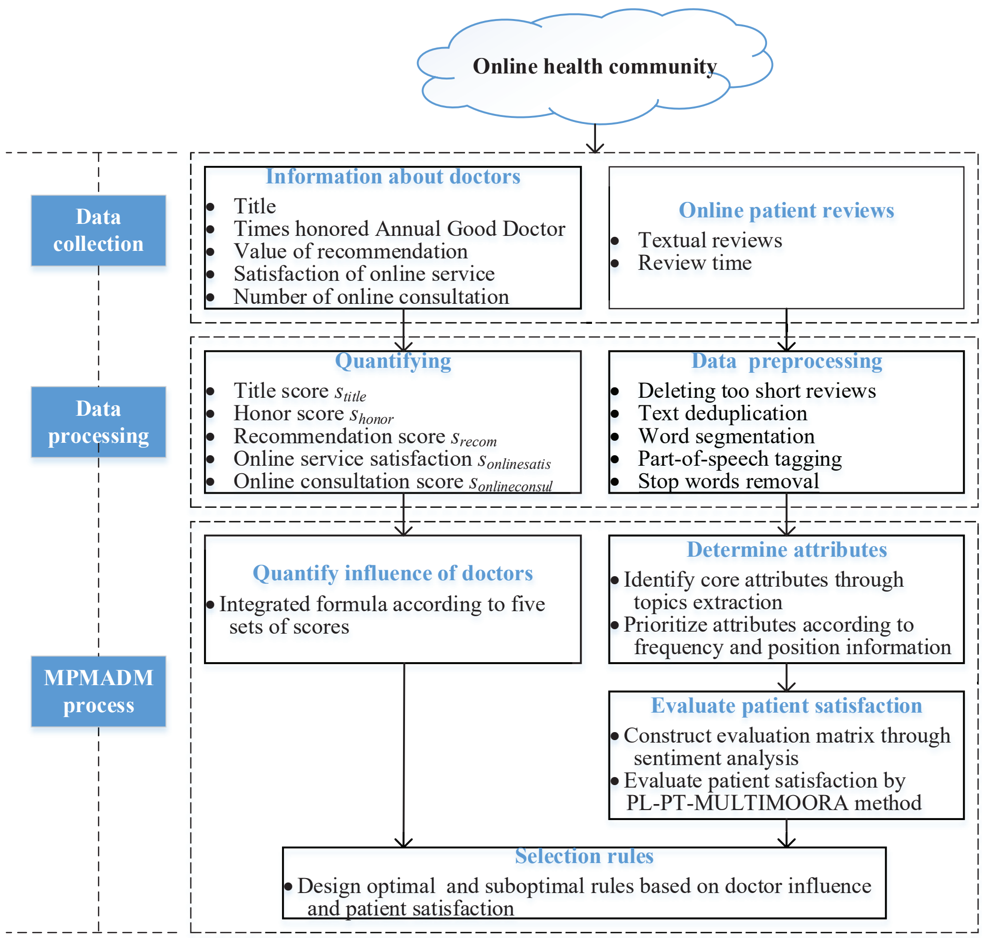

Figure 1. A textual data-driven framework for doctor selection in OHCs.

3.1. Data Collection and Processing

OHCs provide patients with basic information about each doctor, such as doctor title, number of times honored the Annual Good Doctor, value of recommendation, satisfaction of online service, and number of online consultation; they are also great channels for patients to share their medical experiences and judgements. So, the aforementioned basic doctor information and online patient reviews would be important data sources for doctor influence analysis and patient satisfaction analysis. In this study, Python software is used to crawl these two kinds of data from OHCs.

Actually, doctor titles are textual data, namely Assistant Doctor, Doctor, Attending Doctor, Associate Director Physician, and Director Physician. Number of times honored the Annual Good Doctor and value of recommendation are positive numbers with minor deviation. Satisfaction of online service is expressed in the form of percentage. Number of online consultation is a number much larger than 10. Significantly, different types of results collected from basic doctor information should be unified and minor deviation among five values is necessary. Methods for quantifying basic doctor information will introduce in Section 3.2.1. Reviews that are less than five words will induce the sparsity of evaluation matrix, which is not conducive to obtaining an objective and accurate decision result. Therefore, comments that are too short should be filtered before text deduplication, word segmentation, and part-of-speech tagging and stop words removal. Suppose that total N patient reviews are mined; N_1 effective records with regard to m doctors are acquired through data preprocessing.

3.2. MPMADM Process

During the MPMADM process, the results from data collection and processing are utilized to quantify relative influence of doctors, determine attributes, and assess patient satisfactions regarding doctors. Based on doctor influence and patient satisfactions, selection rules are finally proposed to give some guidance for optimal and suboptimal alternatives.

3.2.1. Quantifying Influence of Doctors

Obviously, doctors from OHCs differ in several aspects, such as title, number of times honored Annual Good Doctor, value of recommendation, satisfaction of online service, and number of online consultation. Since all these indicators are highly related to the influence of each doctor, we can quantify them to explore doctor influence. Generally, doctor titles include five scales, namely Assistant Doctor, Doctor, Attending Doctor, Associate Director Physician, and Director Physician. They are separately scored with 1 to 5 and denoted as s_{title}. For example, if a doctor is titled with Attending Doctor, then his/her title score is 3. The number of times honored Annual Good Doctor and value of recommendation are originally scored with positive number with minor deviation, so the scores of these two indicators displayed as s_{honor} and s_{recom} keep their original scores. To reduce the deviation among five kinds of scores, satisfaction of online service and number of online consultation may be scored by ranking score method. That is, the formula $s(h) = m - h + 1$ is used to quantify the online service satisfaction or/and online consultation ranked with hth among m doctors. Scores for online service satisfaction and online consultation are respectively presented as $s_{onlinesatis}$ and $s_{onlineconsul}$. According to five sets of scores, an integrated formula is designed to calculate relative influence λ_i $(i = 1, 2, \ldots, m)$ for m doctors.

$$\lambda_i = \frac{s_{title} + s_{honor} + s_{recom} + s_{onlinesatis} + s_{onlineconsul}}{\sum_{i=1}^{m}(s_{title} + s_{honor} + s_{recom} + s_{onlinesatis} + s_{onlineconsul})} \tag{5}$$

3.2.2. Determining Attributes

Term Frequency-Inverse Document Frequency (TF-IDF) is an important technique for keyword extraction, which enables to depict the importance of a certain word to a certain article. The core idea of TF-IDF is that words that appear more frequently in an article and rarely in other articles tend to have higher values and could better describe the meaning of current article. TF-IDF gets better results than method that only considers word frequency. Hence, TF-IDF method is used to dig keywords, which are further clustered to several topics about patient evaluation. For convenience, attributes denoted as c_j $(j = 1, 2, \ldots, n)$ are the mined topics.

Patients prefer to evaluate doctors with the attributes they focused and the frequency of attributes in reviews could positively depict the importance of attributes. So, a frequency-oriented formula is defined to calculate the frequency weight $fre(c_j)$ of each attribute.

$$fre(c_j) = \frac{\sum_{h=1}^{n} \sum_{i=1}^{m} \lambda_i N_{ij}^{(h)}}{\sum_{j=1}^{n} \sum_{h=1}^{n} \sum_{i=1}^{m} \lambda_i N_{ij}^{(h)}} \tag{6}$$

where λ_i be the relative influence of doctor a_i $(i = 1, \ldots, m)$, $N_{ij}^{(h)}$ presents frequency of attribute c_j ranked hth in online reviews about doctor a_i.

Even if patients use the same attributes to assess their satisfactions regarding doctors, the order of attributes varies from patient to patient. The interesting phenomenon implies that the sequence of attributes appeared in each online review characterizes the importance of attributes in the patient's mind. This is also consistent with primacy effect [45], which states that the information presented first has a greater weight in impression formation than the information presented later. Hence, position score-based formula is proposed to depict the position weight $pos(c_j)$ of every attribute.

$$pos(c_j) = \frac{\sum_{h=1}^{n} \sum_{i=1}^{m} \lambda_i N_{ij}^{(h)} Q(h)}{\sum_{j=1}^{n} \sum_{h=1}^{n} \sum_{i=1}^{m} \lambda_i N_{ij}^{(h)} Q(h)} \tag{7}$$

where $Q(h)$ satisfying $Q(h) = n - h + 1$ be the score of attributes ranked h-th.

In order to completely reflect the influence of frequency information and position information on importance of attributes, an integrated formula is subsequently designed to fuse frequency weights and position weights. The final importance of attribute c_j $(j = 1, \ldots, n)$ is denoted as $\omega(c_j)$.

$$\omega(c_j) = \frac{fre(c_j) + pos(c_j)}{\sum_{j=1}^{n} (fre(c_j) + pos(c_j))} \tag{8}$$

3.2.3. Evaluating Patient Satisfactions for Doctors

Sentiment Knowledge Enhanced Pretraining (SKEP) is one of the important improvements of Bidirectional Encoder Representation from Transformers (Bert) in the field of sentiment analysis. SKEP not only incorporates emotional prior knowledge, it also enhances the pretrainning method for the emotion analysis task and adds a loss function about sentiment. Importantly, the use of SKEP can skew the attention mechanism toward both affective and attribute-affective word pairs. Owing to its outstanding performance in the sentiment analysis, SKEP model is utilized in this study to extract the sentiment degree regarding the performance of each attribute for m doctors. Five emotional scales (i.e., very dissatisfied, dissatisfied, general, satisfied, and very satisfied) are applied to accurately segment sentiment degree, which could be further expressed in the form of PLTS. PLTS is an important linguistic form, which can simultaneously depict the diversified emotional scales and their occurrence probabilities. Therefore, we can quantify the outcomes of sentiment analysis in the form of PLTS and get the original evaluation matrix $R = (r_{ij})_{m \times n}$. Where $r_{ij} = L_{ij}(P_{ij})$ $(i = 1, \ldots, m; j = 1, \ldots, l, \ldots, n)$ represents the performance of doctor a_i regarding attribute c_j. The former l attributes are beneficial attributes, the others are cost attributes.

Generally, MADM methods with multiple subordinate techniques, such as MULTIMOORA, have better robustness and get a more objective decision-making result. Classical MULTIMOORA ignores the impact of psychological cognition of DMs on decision-making outcome. PT is a behavior decision theory that considers bounded rationality of DMs in the whole decision process. Consequently, we try to extend traditional MULTIMOORA and PT into PLTS preference context and design a new bounded rationality-based MADM method with robustness (i.e., PL-PT-MULTIMOORA) to analyze the patient satisfactions regarding m doctors.

Suppose that $r_j^+ = \left\{ \left(L_j^{(1)}\left(P_j^{(1)}\right)\right)^+, \ldots, \left(L_j^{(k)}\left(P_j^{(k)}\right)\right)^+, \ldots, \left(L_j^{(\#L_{ij}(P_{ij}))}\left(P_j^{(\#L_{ij}(P_{ij}))}\right)\right)^+ \right\}$ and $r_j^- = \left\{ \left(L_j^{(1)}\left(P_j^{(1)}\right)\right)^-, \ldots, \left(L_j^{(k)}\left(P_j^{(k)}\right)\right)^-, \ldots, \left(L_j^{(\#L_{ij}(P_{ij}))}\left(P_j^{(\#L_{ij}(P_{ij}))}\right)\right)^- \right\}$ present the positive and negative ideal solutions of attribute c_j for all alternatives, respectively. For beneficial attribute c_j, $\left(L_j^{(k)}\left(P_j^{(k)}\right)\right)^+$ satisfies $\left(r_j^{(k)}\left(P_j^{(k)}\right)\right)^+ = \max_{1\le i\le m}\left\{r_{ij}^{(k)}\left(P_{ij}^{(k)}\right)\right\}$ and $\left(L_j^{(k)}\left(P_j^{(k)}\right)\right)^-$ satisfies $\left(r_j^{(k)}\left(P_j^{(k)}\right)\right)^- = \min_{1\le i\le m}\left\{r_{ij}^{(k)}\left(P_{ij}^{(k)}\right)\right\}$. However, $\left(r_j^{(k)}\left(P_j^{(k)}\right)\right)^+ = \min_{1\le i\le m}\left\{r_{ij}^{(k)}\left(P_{ij}^{(k)}\right)\right\}$ and $\left(r_j^{(k)}\left(P_j^{(k)}\right)\right)^- = \max_{1\le i\le m}\left\{r_{ij}^{(k)}\left(P_{ij}^{(k)}\right)\right\}$ are used for cost attributes where $\#L_{ij}(P_{ij})$ be the number of linguistic term in $L_{ij}(P_{ij})$, $r_{ij}^{(k)}$ be the subscript of linguistic term $L_{ij}^{(k)}$, and $P_{ij}^{(k)}$ be the occurrence probability of $L_{ij}^{(k)}$. If positive ideal solutions are chosen as reference points, all alternatives are considered as loss and DMs are risk lover. On the contrary, when negative ideal solutions are selected as reference points, all alternatives are seemed as gain and DMs prefer to risk aversion. Therefore, probabilistic linguistic prospect effect on gain $d\left(r_{ij}, r_j^-\right)$ and probabilistic linguistic prospect effect on loss $d\left(r_{ij}, r_j^+\right)$ are separately depicted with the aid of Euclidean distance. For all $i = 1, \ldots, m; j = 1, \ldots, l, \ldots, n$, probabilistic linguistic prospect value functions on gain and loss are defined as follows.

$$plpv(r_{ij}) = \begin{cases} plpv(r_{ij})^+ = \left(d\left(r_{ij}, r_j^-\right)\right)^\alpha, \text{ if } r_j^- \text{ be reference point;} \\ plpv(r_{ij})^- = -\theta\left(d\left(r_{ij}, r_j^+\right)\right)^\beta, \text{ if } r_j^+ \text{ be reference point.} \end{cases} \tag{9}$$

where α and β separately reflect concavity on gain and convexity on loss, and parameter θ be coefficient of loss aversion. They are separately valued with 0.89, 0.92, and 2.25 according to literature [36,46].

To avoid the influence of various dimensions on decision-making results, the probabilistic linguistic prospect values should be normalized through Formula (10).

$$\begin{cases} \overline{plpv}(r_{ij})^+ = \dfrac{plpv(r_{ij})^+}{\sqrt{\sum_{i=1}^m \left(plpv(r_{ij})^+\right)^2}}; \\ \overline{plpv}(r_{ij})^- = \dfrac{plpv(r_{ij})^-}{\sqrt{\sum_{i=1}^m \left(plpv(r_{ij})^-\right)^2}}. \end{cases} \tag{10}$$

Subsequently, three subordinate decision-making techniques, namely probabilistic linguistic prospect value ration system (PLPVRS) model, probabilistic linguistic prospect value reference point (PLPVRP) model, and probabilistic linguistic prospect value full multiplicative form (PLPVFMF) model are respectively designed to compute comprehensive prospect values for all alternatives. The detailed formulas for aforementioned models are described in the following.

$$\text{(PLPVRS model.) } CV_1(a_i) = \sum_{j=1}^{l}\left(\xi_j^+ \overline{plpv}\left(r_{ij}\right)^+ + \xi_j^- \overline{plpv}\left(r_{ij}\right)^-\right) - \sum_{j=l+1}^{n}\left(\xi_j^+ \overline{plpv}\left(r_{ij}\right)^+ + \xi_j^- \overline{plpv}\left(r_{ij}\right)^-\right). \tag{11}$$

where ξ_j^+ and ξ_j^- respectively denote probability weight functions regarding gain and loss, which satisfy $\xi_j^+ = \frac{\omega(c_j)^\tau}{\left(\omega(c_j)^\tau + (1-\omega(c_j))^\tau\right)^{1/\tau}}$, $\xi_j^- = \frac{\omega(c_j)^\delta}{\left(\omega(c_j)^\delta + (1-\omega(c_j))^\delta\right)^{1/\delta}}$. $\omega(c_j)$ denotes the importance of attribute c_j. Parameters τ and δ separately reflect risk attitude on gain and loss, which are valued with 0.61 and 0.69 according to Tversky and Kahneman [36].

$$\text{(PLPVRP model.) } CV_2(a_i) = \max_j \xi_j^+ \left| \overline{plpv}\left(r_{ij}\right)^+ - \left(\overline{plpv}\left(r_j\right)^+\right)^{max} \right| - \min_j \xi_j^- \left| \overline{plpv}\left(r_{ij}\right)^- - \left(\overline{plpv}\left(r_j\right)^-\right)^{min} \right|. \tag{12}$$

where if attribute c_j is beneficial, then $\left(\overline{plpv}(r_j)^+\right)^{max} = \max_{1\leqslant i\leqslant m} \overline{plpv}(r_{ij})^+$, $\left(\overline{plpv}(r_j)^-\right)^{min} = \min_{1\leqslant i\leqslant m} \overline{plpv}(r_{ij})^-$, otherwise, $\left(\overline{plpv}(r_j)^+\right)^{max} = \min_{1\leqslant i\leqslant m} \overline{plpv}(r_{ij})^+$, and $\left(\overline{plpv}(r_j)^+\right)^{min} = \max_{1\leqslant i\leqslant m} \overline{plpv}(r_{ij})^-$.

$$\text{(PLPVFMF model.) } CV_3(a_i) = \frac{\prod_{j=1}^{l}\left[\left|\overline{plpv}\left(r_{ij}\right)^+\right|^{\xi_j^+} / \left|\overline{plpv}\left(r_{ij}\right)^-\right|^{\xi_j^-}\right]}{\prod_{j=l+1}^{n}\left[\left|\overline{plpv}\left(r_{ij}\right)^+\right|^{\xi_j^+} / \left|\overline{plpv}\left(r_{ij}\right)^-\right|^{\xi_j^-}\right]}. \tag{13}$$

For PLPVRS and PLPVFMF models, their comprehensive values (i.e., $CV_1(a_i)$ and $CV_3(a_i)$) are both ranked in descending order. On the contrary, comprehensive values resulted from PLPVRP model (i.e., $CV_2(a_i)$) should be ranked in ascending order. The above three sets of ranking outcomes are fused through the following formula to get the patient satisfactions (i.e., $PS(a_i)$ $(i = 1, \ldots, m)$) regarding all doctors.

$$PS(a_i) = \frac{3(m+1) - \sum_{\zeta=1}^{3} R\left(CV_\zeta(a_i)\right)}{3m} \times 100\%. \tag{14}$$

where $R\left(CV_\zeta(a_i)\right)$ reflects the ranking of doctor a_i according to model ζ. $\zeta = 1, 2, 3$ is separately corresponding to PLPVRS, PLPVRP, or PLPVFMF model. Especially, dominance theory [28] should be used to rank doctors when they have the same patient satisfaction.

For every doctor a_i $(i = 1, \ldots, m)$, we obtain two kinds of rankings (i.e., R_i^{influ} and R_i^{satis}) according to doctor influence and patient satisfactions. The selection criteria for optimal and suboptimal alternatives are designed with the aid of two ranking results.

1. Optimal rule: If doctor a_i ranks first in both rankings of doctor influence and patient satisfactions (i.e., $R_i^{influ} = R_i^{satis} = 1$), then doctor a_i is the optimal alternative;
2. Suboptimal rule: For rational patients, doctor with the best ranking in doctor influence (i.e., $R_i^{influ} = 1$) should be regarded as suboptimal alternative, while emotional patients are prone to choose doctor with the best ranking in patient satisfactions (i.e., $R_i^{satis} = 1$)

From the selection criteria, it is clear that only doctor rank first in both of the doctor influence ranking and the patient satisfaction ranking, he/she would be the optimal doctor for patients. Since doctor influence is characterized based on objective data (i.e., doctor basic information), rational patients are likely to choose suitable doctor in the light of doctor influence ranking. Conversely, patient satisfactions determined through relatively subjective judgements (i.e., online patient reviews) would be more considered when emotional patients select appropriate doctor. That is, whether patients are rational or emotional, the described selection criteria enable to help them select optimal and suboptimal alternatives.

To understand above-mentioned decision-making processes easily and efficiently, the detailed steps of textual data-driven framework for doctor selection in OHCs are outlined in the following.

Step 1: Dig doctor basic information and online patient reviews from OHCs;
Step 2: Quantify the mined doctor basic information by the methods introduced in Section 3.2.1 and preprocess all online reviews;
Step 3: Fuse doctor basic information and characterize the relative influence of doctors through Formula (5);

Step 4: Identify the key attributes with the aid of TF-IDF technique;
Step 5: Determine the frequency weights, position weights and final importance of all attributes by Formulas (6)–(8), separately;
Step 6: Construct evaluation matrix in the light of SKEP technique and PLTSs;
Step 7: Obtain the positive and negative ideal solutions for each attribute;
Step 8: Acquire probabilistic linguistic prospect value functions on gain and loss through Formula (9);
Step 9: Normalize all probabilistic linguistic prospect values according to Formula (10);
Step 10: Calculate three kinds of comprehensive values (i.e., $CV_1(a_i)$, $CV_2(a_i)$ and $CV_3(a_i)$) by PLPVRS model, PLPVRP model and PLPVFMF model, respectively;
Step 11: Rank $CV_1(a_i)$ and $CV_3(a_i)$ in descending order while $CV_2(a_i)$ in ascending order;
Step 12: Compute the patient satisfactions for all doctors with the aid of Formula (14);
Step 13: Rank doctors according to their patient satisfactions. Especially, dominance theory is used to rank doctors when they have the same patient satisfaction;
Step 14: Make selection criteria to help patients pick out optimal and suboptimal doctors.

4. An Illustration of Proposed Method

In this section, we employ the aforementioned framework to analyze doctor influence and patient satisfactions in the light of doctor basic information and online patient reviews so as to give some guidance for optimal and suboptimal alternatives.

4.1. Case Description

Haodf.com is an earliest OHC in China that not only offers patients with professionally medical knowledge and information but also gives patients chances to freely share their medical experiences. For simplicity, we separately mine doctor basic information and online patient reviews for 10 doctors treating coronary heart disease from Haodf.com. 4420 records dated from January 2019 to January 2021 are totally crawled. 4412 effective records are then obtained through data preprocessing, such as removing the too short, duplicate, empty, and wrong comments.

4.2. MPMADM Process

Owing to various backgrounds, different doctors have different influence. We collect the basic information about 10 doctors denoted as $\{a_1, \ldots, a_{10}\}$ in Table 1. In the light of quantifying methods described in Section 3.2.1, five sets of scores are obtained, which can be used to calculate relative influence of doctors by Formula (5). The influence of 10 doctors are presented as $\lambda = \{0.122, 0.113, 0.099, 0.091, 0.077, 0.082, 0.116, 0.104, 0.111, 0.086\}$. The corresponding ranking of doctor influence is described as $R^{influ} = \{1, 3, 6, 7, 10, 9, 2, 5, 4, 8\}$.

Table 1. Basic information for 10 doctors.

Doctors	Titles	Times Honored Annual Good Doctor	Values of Recommendation	Satisfaction for Online Service	Number of Online Consultation
a_1	Director Physician	4	5	93%	36067
a_2	Director Physician	2	4.6	97%	8007
a_3	Attending Doctor	2	4.2	100%	6032
a_4	Director Physician	1	4.2	96%	5681
a_5	Associate Director Physician	1	3.8	95%	3811
a_6	Director Physician	0	4	100%	3261
a_7	Associate Director Physician	4	4.5	100%	6461
a_8	Director Physician	3	3.4	80%	14042
a_9	Director Physician	1	3.3	100%	10459
a_{10}	Associate Director Physician	1	4	100%	3801

TF-IDF technique is implemented through Python software to extract core keywords and topics from 4412 effective reviews. Seven topics, such as communication (c_1), therapeutic effect (c_2), process (c_3), beside manner (c_4), medical ethics (c_5), medical skill (c_6) and remedy (c_7) are totally considered as the attributes patients mainly concerned. All attributes are beneficial attributes. We collect the frequency of seven attributes ranked from 1st to 7th in the effective comments of each doctor, which is shown in Table 2. According to Section 3.2.2, frequency information and position information have positively affected importance of attributes. With the aid of Formulas (6)–(8), frequency weights, position weights and final importance for seven attributes are shown as $fre(c) = \{0.112, 0.077, 0.099, 0.138, 0.135, 0.226, 0.213\}$, $pos(c) = \{0.111, 0.071, 0.099, 0.136, 0.131, 0.235, 0.217\}$, and $\omega(c) = \{0.111, 0.074, 0.099, 0.137, 0.133, 0.23, 0.215\}$, separately.

Table 2. Ranking-frequency of attributes for 10 doctors.

		c_1	c_2	c_3	c_4	c_5	c_6	c_7		c_1	c_2	c_3	c_4	c_5	c_6	c_7
1st		52	33	119	83	115	362	199		70	31	54	51	53	64	127
2ed		62	53	78	109	130	157	144		33	29	26	37	22	29	54
3rd		45	56	56	60	54	70	77		11	3	5	6	11	9	15
4th	a_1	20	30	25	23	30	34	26	a_6	2	2	1	2	1	2	4
5th		10	18	4	14	19	7	9		0	1	0	0	0	0	0
6th		2	6	0	3	13	1	1		0	0	0	0	0	0	0
7th		0	2	0	1	2	0	0		0	0	0	0	0	0	0
1st		53	37	62	43	56	283	119		42	13	34	72	47	142	99
2ed		37	41	19	62	48	67	80		32	14	25	44	49	44	54
3rd		13	11	4	17	17	14	14		12	12	13	16	19	31	14
4th	a_2	4	2	1	1	2	0	3	a_7	6	7	7	10	11	6	5
5th		0	2	0	0	1	0	0		2	1	1	3	2	4	7
6th		0	0	0	0	1	0	0		1	1	2	2	1	0	3
7th		0	0	0	0	0	0	0		0	0	0	2	0	0	0
1st		106	37	75	119	86	94	179		27	12	34	51	40	132	92
2ed		73	31	14	59	48	40	89		20	19	24	33	27	46	51
3rd		16	23	12	11	21	10	23		6	6	4	16	13	18	16
4th	a_3	4	5	4	6	5	1	1	a_8	8	3	6	1	5	4	4
5th		1	1	0	1	5	0	1		1	1	2	2	6	0	1
6th		0	0	0	1	1	0	0		0	1	0	0	1	0	0
7th		0	0	0	0	0	0	0		0	0	0	0	0	0	0
1st		90	22	42	65	65	87	118		16	20	36	43	44	112	48
2ed		46	14	20	39	33	37	64		11	23	11	23	25	52	43
3rd		14	12	8	12	16	7	24		16	11	18	11	9	16	13
4th	a_4	2	1	1	0	4	3	3	a_9	4	11	6	5	2	8	5
5th		0	0	0	1	0	0	0		0	4	3	1	6	0	2
6th		0	0	0	0	0	0	0		0	2	0	0	0	1	0
7th		0	0	0	0	0	0	0		1	0	0	0	0	0	0
1st		78	32	53	84	43	37	124		32	15	27	49	52	59	64
2ed		43	24	13	43	25	30	63		17	16	7	26	24	16	42
3rd		7	6	10	14	7	6	24		6	5	4	7	8	10	14
4th	a_5	1	1	2	1	1	0	0	a_{10}	0	3	3	0	0	3	1
5th		2	0	1	0	1	0	0		1	0	0	1	0	0	0
6th		0	2	0	0	0	0	0		0	0	0	0	0	0	0
7th		0	0	0	0	0	0	0		0	0	0	0	0	0	0

Sentiment analysis is then conducted with the aid of SKEP model to analyze the performance of attributes for 10 doctors from very dissatisfied to very satisfied. The outcomes are described in Table 3. Owing to psychological cognition of DMs when they make decisions, a new probabilistic linguistic MADM technique with robustness described in Section 3.2.3 is applied to explore the patient satisfactions for 10 doctors. Specifically, Formulas (9) and (10) are used to construct normalized probabilistic linguistic prospect value matrices on gain and loss displayed in Table 4 when positive and negative ideal solutions are chosen as reference points. Subsequently, PLPVRS, PLPVRP and PLPVFMF models are employed to acquire three kinds of comprehensive prospect values regarding 10 doctors. The patient satisfactions can be determined by Formula (14), and the outcomes are shown in Table 5.

Table 3. Original evaluation matrix on seven attributes.

Alternatives	c_1					$\cdots$	c_7				
	Very Dissatisfied	Dissatisfied	General	Satisfied	Very Satisfied		Very Dissatisfied	Dissatisfied	General	Satisfied	Very Satisfied
a_1	0.094	0.026	0.110	0.267	0.503	$\cdots$	0.037	0.048	0.075	0.149	0.691
a_2	0.028	0.028	0.037	0.374	0.533	$\cdots$	0.023	0.028	0.023	0.148	0.778
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
a_{10}	0	0.071	0.036	0.375	0.518	$\cdots$	0.017	0.041	0.041	0.132	0.769

Table 4. Normalized probabilisitc linguistic prospect value matrices from online patient reviews.

	Normalized Probabilisitc Linguistic Prospect Value Matrix on Gain						
	c_1	c_2	c_3	c_4	c_5	c_6	c_7
a_1	0.082214	0.172325	0.250376	0.121265	0.396248	0.387601	0.149449
a_2	0.268784	0.494657	0.503995	0.36711	0.202007	0.409903	0.442598
a_3	0.386115	0.525269	0.277287	0.346523	0.266301	0.27505	0.348814
a_4	0.339517	0.278604	0.246425	0.17864	0.290961	0.307453	0.293522
a_5	0.286596	0.261087	0.163196	0.35643	0.258403	0.163084	0.333655
a_6	0.361572	0.198189	0.184051	0.380995	0.095785	0.297551	0.185944
a_7	0.445548	0.169281	0.442732	0.355125	0.425966	0.33065	0.326822
a_8	0.259456	0.173287	0.379017	0.377021	0.307629	0.325894	0.368626
a_9	0.280285	0.110106	0.23608	0.227771	0.48893	0.339316	0.101504
a_{10}	0.315242	0.44027	0.299635	0.325878	0.235636	0.25644	0.418383
	Normalized Probabilisitc Linguistic Prospect Value Matrix on Loss						
	c_1	c_2	c_3	c_4	c_5	c_6	c_7
a_1	−0.500357	−0.383388	−0.393324	−0.608849	−0.174315	−0.225048	−0.471744
a_2	−0.35205	−0.083392	−0.054437	−0.143083	−0.384971	−0.237742	−0.119176
a_3	−0.161245	−0.108648	−0.289383	−0.183933	−0.308344	−0.288987	−0.172727
a_4	−0.262888	−0.273556	−0.388906	−0.547793	−0.282291	−0.246153	−0.319252
a_5	−0.361448	−0.28933	−0.419787	−0.153579	−0.319921	−0.632579	−0.206948
a_6	−0.180084	−0.379998	−0.421442	−0.153981	−0.562147	−0.261162	−0.446177
a_7	−0.19967	−0.386542	−0.151339	−0.15314	−0.140195	−0.247503	−0.282548
a_8	−0.314739	−0.370514	−0.203264	−0.170199	−0.266323	−0.251848	−0.196453
a_9	−0.280867	−0.478187	−0.353754	−0.374858	−0.120487	−0.235867	−0.497214
a_{10}	−0.386916	−0.126307	−0.253271	−0.186245	−0.350982	−0.316302	−0.154751

Table 5. Results of three subordinate models and patient satisfactions.

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
$CV_1(a_i)$	−0.212	0.292	0.194	−0.053	−0.109	−0.122	0.224	0.131	−0.087	0.124
ranking	10	1	3	6	8	9	2	4	7	5
$CV_2(a_i)$	0.079	0.034	0.025	0.042	0.069	0.084	0.044	0.041	0.092	0.034
ranking	8	3	1	5	7	9	6	4	10	2
$CV_3(a_i)$	0.378	2.92	1.774	0.731	0.643	0.554	1.904	1.349	0.585	1.387
ranking	10	1	3	6	7	9	2	5	8	4
$PS(a_i)$	16.667%	93.333%	86.667%	53.333%	36.667%	20%	76.667%	66.667%	26.667%	73.333%
R_i^{satis}	10	1	2	6	7	9	3	5	8	4

It is obvious that rankings of doctor influence and patient satisfactions are separately displayed as $R^{influ} = \{1, 3, 6, 7, 10, 9, 2, 5, 4, 8\}$ and $R^{satis} = \{10, 1, 2, 6, 7, 9, 3, 5, 8, 4\}$. Since different doctors rank first in these two rankings, the optimal rule described in Section 3.2.3 fails. Patients could pick out suitable suboptimal doctor according to their characteristics. For rational patients, doctor a_1 would be the suboptimal doctor. Differently, emotional patients are likely to select doctor a_2 as his/her suboptimal alternative.

4.3. Comparison Analysis and Discussions

To illustrate the effectiveness of proposed method, we construct 2 sets of comparative experiments in this subsection.

(1) Whether attribute weight methods are more effective than existing approaches?

Owing to frequency information and position information are both mined to determine final weights of attributes, we compare the designed method with a frequency-driven method (i.e., TF-IDF [47]) and a position score-oriented method (i.e., Borda count [48]). The results are depicted in Table 6.

Table 6. Ranking of attribute weights resulted from different methods.

Methods	Ranking
TF-IDF [47]	$c_5 \succ c_7 \succ c_6 \succ c_3 \succ c_4 \succ c_1 \succ c_2$
Frequency-based formula in this paper	$c_5 \succ c_7 \succ c_6 \succ c_3 \succ c_4 \succ c_1 \succ c_2$
Borda count [48]	$c_5 \succ c_7 \succ c_6 \succ c_3 \succ c_4 \succ c_1 \succ c_2$
Position score-oriented formula in this paper	$c_5 \succ c_7 \succ c_6 \succ c_3 \succ c_4 \succ c_1 \succ c_2$
Integrated formula in this paper	$c_5 \succ c_7 \succ c_6 \succ c_3 \succ c_4 \succ c_1 \succ c_2$

Clearly, TF-IDF and Borda count approach both acquire the same ranking of attribute weights with the proposed methods according to Table 6. This important result illustrates that our designed frequency-based formula and position score-oriented formula are feasible and reasonable. Although TF-IDF and Borda count approach can get the same results, they only consider the influence of frequency information or position information on importance of attributes. As described in Section 3.2.2, however, both frequency information and position information positively affect attribute weights to some extent. Comparing with TF-IDF and Borda count approach, the proposed methods for attribute weights have significant strengths, which not only consider the influence of frequency information and position information on importance of attributes at the same time but also acquire reliable and feasible results.

(2) Whether PL-PT-MULTIMOORA method is more appropriate than existing approaches?

We further compare PL-PT-MULTIMOORA with four MADM approaches, in which two are single decision-making methods (i.e., TOPSIS [32] and VIKOR [8]) and the others are multiple subordinate methods (i.e., PL-MULTIMOORA [10] and PL-PT-MULTIMOORA [37]). The outcomes are presented in Table 7 and the final rankings of five decision-making methods are depicted in Figure 2.

The results in Figure 2 show that different decision-making methods have different decision-making results. Though the rankings obtained by five decision-making methods are different, the ranking results of TOPSIS [32], VIKOR [8], PL-MULTIMOORA [10] and PL-PT-MULTIMOORA (proposed in this paper) are relatively similar. This phenomenon demonstrates the effectiveness of designed method. TOPSIS [32], VIKOR [8] are single decision-making approaches, which have worse robustness than multiple subordinate methods. So, the methods from literature [32] and [8] are not appropriate for this case. Among three PL-MULTIMOORA approaches (i.e., PL-MULTIMOORA [10], PL-PT-MULTIMOORA [37] and PL-PT-MULTIMOORA (proposed in this paper)), original data instead of normalized data are used in the RP model of PL-MULTIMOORA method [10], which is clearly inconsistent with the traditional MULTIMOORA. Conversely, PL-PT-MULTIMOORA methods designed by [37] and this paper utilize normalized data into three subordinate models (i.e., RS, RP and FMF models). Besides, PL-MULTIMOORA [10] ignored the influence of psychological cognition of DMs on decision-making outcomes. Hence, method from Li et al. [10] is also not the best approach for this case. Both Liu and Li [37] and this study develop PL-PT-MULTIMOORA methods, but obvious differences are summarized as follows: (1) Different reference points: PL-PT-MULTIMOORA [37] chose the mean as reference point so that zero elements maybe present in evaluation matrix. Since positive and negative ideal solutions are regarded as reference points in this paper, the evaluation matrix in this study has no zero elements. (2) Different formulas about three subordinate models: owing to different selection of reference points, prospect values on gain or loss are used to construct three models in PL-PT-MULTIMOORA [37], while both prospect values on gain and loss are simultaneously considered into three formulas in PL-PT-MULTIMOORA (proposed in this study). (3) Different integrated ways: both numerical results and ranking results of three models are considered into final value determination method in PL-PT-MULTIMOORA [37], which is more complicated and time-consuming. This paper develops a formula to fuse rankings resulted from three models and get the final value (i.e., patient satisfaction). Only alternatives with the same final value should be further compared through dominance theory. That is, an integrated way in this paper is more simple and efficient. (4) Different usage of attribute weights: Liu and Li [37] used attribute weights twice during the decision-making process. One is to calculate prospect value, and the other is to get three kinds of collective values on three models. However, attribute weights are only employed into three models in this paper. More importantly, the ranking result from this paper is more credible and effective than that from [37] since ranking obtained by this study is more similar to that from other methods.

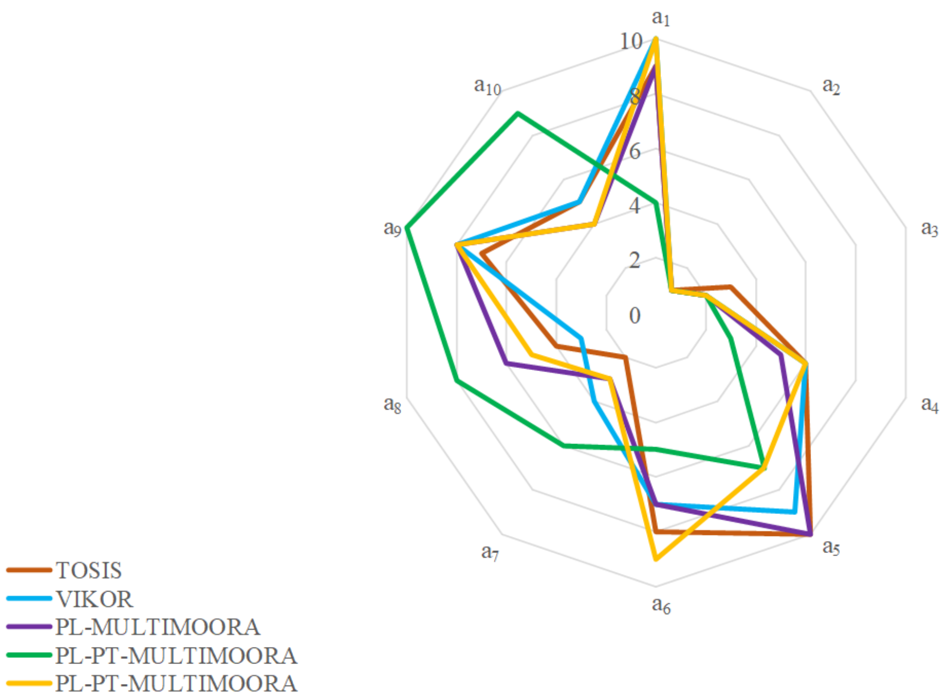

Figure 2. Final ranking results of five decision-making methods.

Characteristics for five decision-making techniques are further analyzed in Table 8 to highlight the superiority and advantages of designed method. In addition to method from [8,32], the other techniques believe that integrated decision-making method which includes three subordinate approaches has better robustness. Compared with methods from [10], PL-PT-MULTIMOORA methods designed by [37] and this study use normalized data to compute three subordinate methods. Although psychological cognition of DMs is also considered by [37], Liu and Li [37] reused attribute weights during the decision-making process.

Table 7. The Ranking results for 10 doctors resulted from five methods.

		a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
TOPSIS [32]	$CI(a_i)$ [1]	−1.016	0	−0.322	−0.762	−1.613	−1	−0.29	−0.478	−0.851	−0.57
	Ranking	9	1	3	6	10	8	2	4	7	5
VIKOR [8]	$S^+(a_i)$ [1]	0.664	0.323	0.358	0.544	0.592	0.593	0.367	0.399	0.571	0.417
	$S^-(a_i)$ [1]	0.19	0.086	0.093	0.124	0.217	0.178	0.109	0.08	0.201	0.102
	$f(a_i)$	0.9	0.023	0.099	0.486	0.894	0.755	0.17	0.111	0.805	0.219
	Ranking	10	1	2	6	9	7	4	3	8	5
PL-MULTIMOORA [10]	$U_1(a_i)$ [1]	0.311	0.327	0.322	0.317	0.308	0.313	0.319	0.316	0.311	0.318
	Ranking	9	1	2	5	10	7	3	6	8	4
	$U_2(a_i)$ [1]	0.011	0.012	0.015	0.012	0.035	0.013	0.012	0.013	0.012	0.016
	Ranking	1	3	8	4	10	7	5	6	2	9
	$U_3(a_i)$ [1]	0.31	0.326	0.321	0.317	0.307	0.313	0.319	0.316	0.311	0.318
	Ranking	9	1	2	5	10	7	3	6	8	4
	Final ranking	9	1	2	5	10	7	3	6	8	4
PL-PT-MULTIMOORA [37]	$TS_1(a_i)$ [1]	−0.015	0.012	0.007	−0.005	−0.038	−0.016	0.03	−0.021	−0.139	−0.052
	Ranking	5	2	3	4	8	6	1	7	10	9
	$TS_2(a_i)$ [1]	0.038	0.03	0.031	0.032	0.068	0.038	0.102	0.105	0.097	0.098
	Ranking	5	1	2	3	6	4	9	10	7	8
	$TS_3(a_i)$ [1]	0.351	0.378	0.374	0.362	0.322	0.35	0.275	0.234	0	0.173
	Ranking	4	1	2	3	6	5	7	8	10	9
	Final ranking	4	1	2	3	7	5	6	8	10	9
PL-PT-MULTIMOORA (proposed in this paper)	$CV_1(a_i)$ [1]	−0.212	0.292	0.194	−0.053	−0.109	−0.122	0.224	0.131	−0.087	0.124
	Ranking	10	1	3	6	8	9	2	4	7	5
	$CV_2(a_i)$ [1]	0.079	0.034	0.025	0.042	0.069	0.084	0.044	0.041	0.092	0.034
	Ranking	8	3	1	5	7	9	6	4	10	2
	$CV_3(a_i)$ [1]	0.378	2.92	1.774	0.731	0.643	0.554	1.904	1.349	0.585	1.387
	Ranking	10	1	3	6	7	9	2	5	8	4
	Final ranking	10	1	2	6	7	9	3	5	8	4

[1] All utility values in Table 7 are rounded to three significant digits since the layout limitations.

To sum up, the proposed method considering the impact of psychological cognition of DMs on decision-making outcomes has robustness and gets more reliable results.

Table 8. The characteristics analysis for five decision-making techniques.

	Fuse Several Subordinate Methods	Compute Utility Values Based on Normalized Data	Consider Psychological Cognition	Reuse Attribute Weights
TOPSIS [32]	No	No	No	No
VIKOR [8]	No	No	No	No
PL-MULTIMOORA [10]	Yes	No [1]	No	No
PL-PT-MULTIMOORA [37]	Yes	Yes	Yes	Yes
PL-PT-MULTIMOORA (Proposed in this paper)	Yes	Yes	Yes	No

[1] Not all subordinate methods are computed based on normalized data.

4.4. Managerial Implications and Limitations

OHCs, being doctor-patient interactive platforms, not only provide patients with objective and basic information about doctors but also encourage patients to share their experience-oriented online reviews. Therefore, this paper conduces to mine and quantify doctor basic information and online patient reviews to systematically design a textual data-driven framework for doctor selection in OHCs. Specifically, this paper contributes to characterize relative influence of doctors through quantifying doctor basic information, such as doctor title, number of times honored Annual Good Doctor, value of recommendation, satisfaction of online service, and number of online consultation. For determination of attributes, this study uses topics extraction technique to identify core attributes and prioritizes attributes according to their frequency information and position information. Both frequency and position information have positive impact on importance of attributes, so this study dedicates to obtain more reliable attribute weights. For patient satisfactions, this method digs original evaluations with the aid of sentiment analysis technique and proposes an improved MULTIMOORA (i.e., PL-PT-MULTIMOORA) method to analyze patient satisfactions on doctors. The designed method also promotes to acquire more effective and credible patient satisfactions. More importantly, the selection criteria help patients choose the optimal and suboptimal doctors, regardless of whether patients are rational or emotional. In particular, optimal rule requires that only doctor who ranks first in both doctor influence ranking and patient satisfaction ranking should be selected as optimal doctor for patients. According to suboptimal rule, rational patients prefer to choose suitable doctor in the light of doctor influence ranking while patient satisfactions are more concerned when emotional patients select appropriate doctor. That is, no matter what patients are rational or emotional, the described selection criteria enable to help patients pick out optimal and suboptimal doctors.

This work has the following limitations. Linear ordering approaches [49–52] developed by Prof. Hellwig are interesting and meaningful. Researching on the combination of traditional MADM techniques with Hellwig's approach will be a meaningful work. Knowledge sharing is an important content in OHCs. Although knowledge sharing has not been considered in this study, it is also an interesting and meaningful topic to study in future. For example, we can compare doctor selection of patients before and after knowledge sharing and then explore the influence of knowledge sharing on doctor selection of patients.

5. Conclusions

In the era of Big Data, OHCs offer patients much useful information including doctor basic information and online patient reviews. Therefore, in this paper, we pay attention to doctor selection in OHCs based on doctor basic information and online patient reviews.

The detailed strengths of designed method are displayed as follows: (1) A data-driven method is systematically developed to pick out optimal and suboptimal doctors for all patients; (2) Doctor basic information is quantified and employed to characterize relative influence of doctors by an integrated formula; (3) Topics extraction and sentiment analysis are separately used to dig core attributes and original evaluations; (4) Frequency weights and position weights are respectively determined by a frequency-oriented formula and a position score-based formula, which are fused to get final importance of attributes; (5) A bounded rationality-based method with robustness (i.e., PL-PT-MULTIMOORA) is designed to analyze patient satisfactions on doctors; (6) Selection criteria are introduced to help rational or emotional patients select the optimal and suboptimal doctors.

In the future, we will use the proposed method to deal with other management problems, such as recommendation of tourist attractions, user satisfaction analysis of products, evaluation of project, and so on. We will also focus on the impact of knowledge sharing in OHCs on doctor selection in our future work.

Author Contributions: Conceptualization,Y.D., Z.-S.C., J.A.M.-M., L.Z. and E.H.-V.; methodology, Y.D. and J.Y.; formal analysis, Y.D.; data curation, L.Z.; writing—original draft preparation, Y.D. and J.Y.; writing—review and editing, Y.D., Z.-S.C., J.Y., J.A.M.-M., L.Z. and E.H.-V.; project administration, Z.-S.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the National Natural Science Foundation of China [Nos. 72171182, 71801175, 71871171, and 72031009], the Project of Service Science and Innovation Key Laboratory of Sichuan Province [No. KL2105], and the Project of China Scholarship Council [Nos. 202107000064 and 202007000143]. Besides, this research was also partially supported by the project PID2019-103880RB-I00 funded by MCIN/AEI/10.13039/501100011033, by FEDER/Junta de Andalucía-Consejería de Transformación Económica, Industria, Conocimiento y Universidades/ Proyecto B-TIC-590-UGR20, and by the Andalusian government through project P20 00673.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Alemu, E.N.; Huang, J. HealthAid: Extracting domain targeted high precision procedural knowledge from online communities. *Inf. Process. Manag.* **2020**, *57*, 102299. [CrossRef]
2. Jie, M.G.; Gao, G.; Agarwal, R. The creation of social value: Can an online health community reduce rural–urban health disparities? *MIS Q.* **2016**, *40*, 247–263.
3. Meng, F.B.; Zhang, X.F.; Liu, L.B.; Ren, C.C. Converting readers to patients? From free to paid knowledge-sharing in online health communities. *Inf. Process. Manag.* **2021**, *58*, 102490. [CrossRef]
4. Zhao, H.; Fu, S.; Chen, X. Promoting users' intention to share online health articles on social media: The role of confirmation bias. *Inf. Process. Manag.* **2020**, *57*, 102354. [CrossRef]
5. Yan, Z.J.; Wang, T.M.; Chen, Y.; Zhang, H. Knowledge sharing in online health communities: A social exchange theory perspective. *Inf. Manag.* **2016**, *53*, 643–653. [CrossRef]
6. Huang, W.S.; Cao, B.L.; Yang, G.; Luo, N.Z.; Chao, N.P. Turn to the Internet First? Using Online Medical Behavioral Data to Forecast COVID-19 Epidemic Trend. *Inf. Process. Manag.* **2021**, *58*, 102486. [CrossRef]
7. Qiao, W.X.; Yan, Z.J.; Wang, X.H. Join or not: The impact of physicians' group joining behavior on their online demand and reputation in online health communities. *Inf. Process. Manag.* **2021**, *58*, 10263. [CrossRef]
8. Hu, J.H.; Zhang, X.H.; Yang, Y.; Liu, Y.M.; Chen, X.H. New doctors ranking system based on VIKOR method. *Int. Trans. Oper. Res.* **2020**, *27*, 1236–1261. [CrossRef]
9. James, T.L.; Calderon, E.D.V.; Cook, D.F. Exploring patient perceptions of healthcare service quality through analysis of unstructured feedback. *Expert Syst. Appl.* **2017**, *71*, 479–492. [CrossRef]
10. Li, Y.; Zhang, Y.X.; Xu, Z.S. A decision-making model under probabilistic linguistic circumstances with unknown criteria weights for online customer reviews. *Int. J. Fuzzy Syst.* **2020**, *22*, 777–789. [CrossRef]
11. Shah, A.M.; Yan, X.B.; Tariq, S.; Ali, M. What patients like or dislike in physicians: Analyzing drivers of patient satisfaction and dissatisfaction using a digital topic modeling approach. *Inf. Process. Manag.* **2021**, *58*, 102516. [CrossRef]
12. Yang, H.L.; Guo, X.T.; Wu, T.S. Exploring the influence of the online physician service delivery process on patient satisfaction. *Decis. Support Syst.* **2015**, *78*, 113–121. [CrossRef]
13. Du, Y.F.; Liu, D.; Morente-Molinera, J.A.; Herrera-Viedma, E. A data-driven method for user satisfaction evaluation of smart and connected products. *Expert Syst. Appl.* **2022**, *210*, 118392. [CrossRef]
14. Chen, S.Q.; Guo, X.T.; Wu, T.S.; Ju, X.F. Exploring the online doctor-patient interaction on patient satisfaction based on text mining and empirical analysis. *Inf. Process. Manag.* **2020**, *57*, 102253. [CrossRef]
15. Hao, H.J.; Zhang, K.P. The voice of Chinese health consumers: A text mining approach to web-based physician reviews. *J. Med. Internet Res.* **2016**, *18*, e4430. [CrossRef] [PubMed]
16. Lu, N.J.; Wu, H. Exploring the impact of word-of-mouth about Physicians' service quality on patient choice based on online health communities. *BMC Med. Inform. Decis. Mak.* **2016**, *16*, 151. [CrossRef]
17. Wu, H.; Lu, N.J. Service provision, pricing, and patient satisfaction in online health communities. *Int. J. Med. Inform.* **2018**, *110*, 77–89. [CrossRef]
18. Büyüközkan, G.; Çifçi, G.; Güleryüz, S. Strategic analysis of healthcare service quality using fuzzy AHP methodology. *Expert Syst. Appl.* **2011**, *38*, 9407–9424. [CrossRef]
19. Du, Y.F.; Liu, D. An integrated method for multi-granular probabilistic linguistic multiple attribute decision-making with prospect theory. *Comput. Ind. Eng.* **2021**, *159*, 107500. [CrossRef]
20. Hu, Y.H.; Chen, Y.L.; Chou, H.L. Opinion mining from online hotel reviews-a text summarization approach. *Inf. Process. Manag.* **2017**, *53*, 436–449. [CrossRef]
21. Yang, W.J.; Zhang, J.T.; Yan, H. Impacts of online consumer reviews on a dual-channel supply chain. *Omega* **2021**, *101*, 102266. [CrossRef]
22. Tian, G.; Lu, L.; McIntosh, C. What factors affect consumers' dining sentiments and their ratings: Evidence from restaurant online review data. *Food Qual. Prefer.* **2021**, *88*, 104060. [CrossRef]

23. Chen, Z.S.; Liu, X.L.; Chin, K.S.; Pedrycz, W.; Tsui, K.L.; Skibniewski, M.J. Online-review analysis based large-scale group decision-making for determining passenger demands and evaluating passenger satisfaction: Case study of high-speed rail system in China. *Inf. Fusion* **2021**, *69*, 22–39. [CrossRef]
24. Park, J.; Lee, B.K. An opinion-driven decision-support framework for benchmarking hotel service. *Omega* **2021**, *103*, 102415. [CrossRef]
25. Zhang, C.X.; Xu, Z.S.; Gou, X.J.; Chen, S.X. An online reviews-driven method for the prioritization of improvements in hotel services. *Tour. Manag.* **2021**, *87*, 104382. [CrossRef]
26. Chen, Z.S.; Yang, L.L.; Rodríguez, R.M.; Xiong, S.H.; Chin, K.S.; Martínez, L. Power-average-operator-based hybrid multiattribute online product recommendation model for consumer decision-making. *Int. J. Intell. Syst.* **2021**, *36*, 2572–2617. [CrossRef]
27. Du, Y.F.; Liu, D.; Duan, H.X. A textual data-driven method to identify and prioritise user preferences based on regret/rejoicing perception for smart and connected products. *Int. J. Prod. Res.* **2022**, *60*, 4176–4196. [CrossRef]
28. Brauers, W.K.M.; Zavadskas, E.K. Project management by MULTIMOORA as an instrument for transition economies. *Technol. Econ. Dev. Econ.* **2010**, *16*, 5–24. [CrossRef]
29. Brauers, W.K.M.; Zavadskas, E.K. Robustness of MULTIMOORA: A method for multi-objective optimization. *Informatica* **2012**, *23*, 1–25. [CrossRef]
30. Chakraborty, S. Applications of the MOORA method for decision making in manufacturing environment. *Int. J. Adv. Manuf. Technol.* **2011**, *54*, 1155–1166. [CrossRef]
31. Wu, X.L.; Liao, H.C.; Xu, Z.S.; Hafezalkotob, A.; Herrera, F. Probabilistic linguistic MULTIMOORA: A multicriteria decision making method based on the probabilistic linguistic expectation function and the improved Borda rule. *IEEE Trans. Fuzzy Syst.* **2018**, *26*, 3688–3702. [CrossRef]
32. Pang, Q.; Wang, H.; Xu, Z.S. Probabilistic linguistic term sets in multi-attribute group decision making. *Inf. Sci.* **2016**, *369*, 128–143. [CrossRef]
33. Darko, A.P.; Liang, D.C. A heterogeneous opinion-driven decision-support model for tourists' selection with different travel needs in online reviews. *J. Oper. Res. Soc.* **2022**. [CrossRef]
34. Liang, D.C.; Dai, Z.Y.; Wang, M.W. Assessing customer satisfaction of o2o takeaway based on online reviews by integrating fuzzy comprehensive evaluation with ahp and probabilistic linguistic term sets. *Appl. Soft Comput.* **2021**, *98*, 106847. [CrossRef]
35. Kahneman, D.; Tversky, A. Prospect theory: An analysis of decision under risk. *Econometrica* **1979**, *47*, 263–291. [CrossRef]
36. Tversky, A.; Kahneman, D. Advances in prospect theory: Cumulative representation of uncertainty. *J. Risk Uncertain.* **1992**, *5*, 297–323. [CrossRef]
37. Liu, P.D.; Li, Y. An extended MULTIMOORA method for probabilistic linguistic multi-criteria group decision-making based on prospect theory. *Comput. Ind. Eng.* **2019**, *136*, 528–545. [CrossRef]
38. Ying, C.S.; Li, Y.L.; Chin, K.S.; Yang, H.T.; Xu, J. A new product development concept selection approach based on cumulative prospect theory and hybrid-information MADM. *Comput. Ind. Eng.* **2018**, *122*, 251–261. [CrossRef]
39. Dong, Y.C.; Luo, N.; Liang, H.M. Consensus building in multiperson decision making with heterogeneous preference representation structures: A perspective based on prospect theory. *Appl. Soft Comput.* **2015**, *35*, 898–910. [CrossRef]
40. Liao, H.C.; Wu, X.L. DNMA: A double normalization-based multiple aggregation method for multi-expert multi-criteria decision making. *Omega* **2020**, *94*, 102058. [CrossRef]
41. Liang, D.C.; Dai, Z.Y.; Wang, M.W.; Li, J.J. Web celebrity shop assessment and improvement based on online review with probabilistic linguistic term sets by using sentiment analysis and fuzzy cognitive map. *Fuzzy Optim. Decis. Mak.* **2020**, *19*, 561–586. [CrossRef]
42. Wu, X.L.; Liao, H.C. Modeling personalized cognition of customers in online shopping. *Omega* **2021**, *104*, 102471. [CrossRef]
43. Du, Y.F.; Liu, D. A novel approach to relative importance ratings of customer requirements in QFD based on probabilistic linguistic preferences. *Fuzzy Optim. Decis. Mak.* **2021**, *20*, 365–395. [CrossRef]
44. Brauers, W.K.M.; Zavadskas, E.K. The moora method and its application to privatization in a transition economy. *Control Cybern.* **2006**, *35*, 445–469.
45. Li, C. Primacy effect or recency effect? A long-term memory test of Super Bowl commercials. *J. Consum. Behav.* **2010**, *9*, 32–44. [CrossRef]
46. Abdellaoui, M. Parameter-free elicitation of utility and probability weighting functions. *Manag. Sci.* **2000**, *46*, 1497–1512. [CrossRef]
47. Xu, X.H.; Yang, X.; Chen, X.H.; Liu, B.S. Large group two-stage risk emergency decision-making method based on Big Data analysis of social media. *J. Intell. Fuzzy Syst.* **2019**, *36*, 2645–2659. [CrossRef]
48. Yang, J.M.; Shi, B.S. Joining method in group appraising. *Syst. Eng.-Theory Pract.* **1992**, *12*, 49–51.
49. Hellwig, Z. *Zastosowanie Metody Taksonomicznej do Typologicznego Podziału Krajów ze Względu na Poziom ich Rozwoju i Strukturę Wykwalifikowanych kadr [Procedure of Evaluating High Level Manpower Data and Typology of Countries by Means of the Taxonomic Method], Przegląd Statystyczny, tom 15, z. 4*; 1968; pp. 307–327.
50. Hellwig, Z. Procedure of evaluating high-level manpower data and typology of countries by means of the taxonomic method. In *Towards a System of Human Resources Indicators for Less Developed Countries*; 1972; pp. 115–134.

51. Hellwig, Z. Wielowymiarowa analiza porównawcza i jej zastosowanie w badaniach wielocechowych obiektów gospodarczych [Multivariate Comparative Analysis and Applications in Research of Multifeature Economic Objects]. In *Metody i Modele Ekonomicznomatematyczne w Doskonaleniu Zarządzania Gospodarką Socjalistyczną*; Welfe, W., Ed.; PWE: Warszawa, Poland, 1981; pp. 46–68.
52. Walesiak, M. Visualization of linear ordering results for metric data with the application of multidimensional scaling. *Ekonometria* **2016**, 9–21.