Research Article

Decision-Making Support for the Evaluation of Clustering Algorithms Based on MCDM

Wenshuai Wu ¹, Zeshui Xu ¹, Gang Kou ² and Yong Shi ^{3,4}

¹Business School, Sichuan University, Chengdu, Sichuan 610065, China

²School of Business Administration, Southwestern University of Finance and Economics, Chengdu, Sichuan 611130, China

³CAS Research Center on Fictitious Economy & Data Science, Chinese Academy of Sciences, Beijing 100190, China

⁴Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing 100190, China

Correspondence should be addressed to Zeshui Xu; xuzeshui@263.net

Received 28 October 2019; Revised 18 February 2020; Accepted 5 March 2020; Published 5 May 2020

Academic Editor: Dimitri Volchenkov

Copyright © 2020 Wenshuai Wu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In many disciplines, the evaluation of algorithms for processing massive data is a challenging research issue. However, different algorithms can produce different or even conflicting evaluation performance, and this phenomenon has not been fully investigated. The motivation of this paper aims to propose a solution scheme for the evaluation of clustering algorithms to reconcile different or even conflicting evaluation performance. The goal of this research is to propose and develop a model, called decision-making support for evaluation of clustering algorithms (DMSECA), to evaluate clustering algorithms by merging expert wisdom in order to reconcile differences in their evaluation performance for information fusion during a complex decision-making process. The proposed model is tested and verified by an experimental study using six clustering algorithms, nine external measures, and four MCDM methods on 20 UCI data sets, including a total of 18,310 instances and 313 attributes. The proposed model can generate a list of algorithm priorities to produce an optimal ranking scheme, which can satisfy the decision preferences of all the participants. The results indicate our developed model is an effective tool for selecting the most appropriate clustering algorithms for given data sets. Furthermore, our proposed model can reconcile different or even conflicting evaluation performance to reach a group agreement in a complex decision-making environment.

1. Introduction

Clustering is widely applied in the initial stage of big data analysis to divide large data sets into smaller sections, so the data can be comprehended and mastered easily with successive analytic operations [1–3]. The processing of massive data relies on the selection of an appropriate clustering algorithm, and the issue of the evaluation of clustering algorithms remains an active and significant issue in many subjects, such as fuzzy set, genomics, data mining, computer science, machine learning, business intelligence, and financial analysis [1, 4–6]. Computer scientists, economists, political scientists, bioinformatics specialists, sociologists, and many other groups usually debate the potential costs and benefits by analyzing these data for supporting decision-making [7]. However, the decision-making process is

extremely complex because of competing interests of multiple stakeholders and the intricacy of systems [8–10].

Clustering algorithms, which are unsupervised pattern-learning algorithms without a priori information, partition the original data space into smaller sections with high intergroup dissimilarities and intragroup similarities. Clustering can be used to process various types of massive data to uncover unknown correlations, hidden patterns, and other potentially useful information. However, Naldi et al. [11] pointed out different clustering algorithms sometimes produce different data partitions. In some situations, different algorithms produce different or even conflicting results. Therefore, the evaluation of clustering algorithms remains a significant task and a challenging problem.

Several validity measures for assessing clustering algorithms are presented successively, such as the Xie–Beni (XB)

index [12], the I-index [13], the CS index [14, 15], Dunn's index [16, 17], and the Davies–Bouldin (DB) index [18, 19]. These validity measures are often divided into the three categories of external, relative, and internal measures [20–22]. External measures compare the partitions produced by clustering algorithms with a given data partition [20, 22]. Relative measures compare partitions produced by the same clustering algorithm with discrepant data subsets or diverse parameters [22]. Internal measures depend on computing the property of the resulting clusters [22]. Brn et al. [20] stated that relative and internal measures fail in predicting and locating error produced by clustering algorithms, and external measures for evaluating clustering results perform more effectively. Therefore, in our empirical research, we will select external measures to evaluate and measure the performance of clustering algorithms.

The theorem of No Free Lunch (NFL) states that there exists no single model or algorithm which can get the best performance for a given domain problem [23–25]. It suggests that the evaluation of clustering algorithms is very complicated and challenging. Moreover, different clustering algorithms may produce different or conflicting partitions. The motivation of this paper fixates on the evaluation of clustering algorithms to reconcile different or even conflicting evaluation performance. Besides, the reconciliation of these differences or conflicts is an important problem which has not been fully investigated. In addition, the evaluation of clustering algorithms usually involves multiple criteria which are modeled as an MCDM problem. So, based on MCDM, this paper proposes a model, called decision-making support for evaluation of clustering algorithms (DMSECA), to evaluate and measure the performance of clustering algorithms and further to reconcile their differences or even conflicts among the evaluation performance of clustering algorithms during a complex decision-making process.

The proposed model consists of three steps. First, we apply the six most influential clustering algorithms to task modeling on 20 UCI data sets with a total of 18,310 instances and 313 attributes. Second, based on nine external measures, we employ four commonly used MCDM approaches to rank the performance of clustering algorithms over the 20 UCI data sets. Third, based on the eighty-twenty rule, we propose a decision-making support model to generate a list of algorithm priorities to identify the best clustering algorithm among 20 UCI data sets for secondary mining and knowledge discovery. Each MCDM method is randomly assigned to five UCI data sets.

The contributions of this article are threefold: first, our proposed DMSECA model can identify the best clustering algorithms for the given data sets by a generated list of algorithm priorities during a complex decision-making process. Second, the proposed model can reconcile those differences or even conflicts to achieve agreement in terms of the clustering algorithm evaluation. Third, based on the eighty-twenty rule, the expert wisdom is merged to propose a decision-making support model to carry out secondary knowledge discovery for information fusion in a complex decision-making environment.

The rest of this article is organized as follows. Section 2 reviews the related work. Section 3 describes some preliminaries, such as clustering algorithms, MCDM methods, and external measures. Section 4 proposes our model by merging the expert wisdom to reconcile disagreements among the clustering algorithms. Section 5 presents the data sets, provides the experimental design, shows empirical results, and discusses the significance of this work. Section 6 summarizes this article.

2. Related Work

Cluster analysis aims to classify elements into categories on the basis of their similarity [26]. In recent years, many clustering algorithms have been proposed [26–29]. The density peak clustering has been published by Rodriguez and Laio in Science [26]. In view of the low objectivity and accuracy because of the man-made factor, a density fragment clustering without peaks is proposed based on density peak clustering [30]. Jiang et al. [28] developed GDPC algorithm with an alternative decision graph based on gravitation theory and nearby distance to identify centroids and anomalies accurately. In order to overcome the defect of the original DPC in detecting anomalies and hub nodes, Jiang et al. [29] proposed an improved recognition method on the halo node for density peak clustering algorithm (halo DPC) [29]. The proposed halo DPC can improve the ability to deal with varying densities, irregular shapes, the number of clusters, outlier, and hub node detection [29].

Clustering ensemble has been increasingly popular in the recent years by consolidating several base clustering methods into a probably better and more robust one. Ali-zadeh et al. [31] presented a novel optimization-based method for the combination of cluster ensembles. Parvin and Minaei-Bidgoli [32] proposed a weighted locally adaptive clustering (WLAC) algorithm which is based on the LAC algorithm. Considering that some features have more information than the others in a dataset, Parvin and Minaei-Bidgoli [27] proposed a fuzzy weighted locally adaptive clustering (FWLAC) algorithm, which is capable of handling imbalanced clustering. Abbasi et al. [33] proposed a criterion to assess the association between a cluster and a partition which is called edited normalized mutual information, ENMI criterion. Mojarad et al. [34] presented a clustering ensemble method, named RCEIFBC, with a new aggregation function, which takes into account the two similarity criteria: (a) one of them is the cluster-cluster similarity and (b) the other one is the object-cluster similarity. Mojarad et al. [35] proposed an ensemble aggregator, or a consensus function, called as the robust clustering ensemble based on sampling and cluster clustering (RCESCC) algorithm in order to get better clustering results. Rashidi et al. [36] proposed a new clustering ensemble approach using a weighting strategy for performing consensus clustering by exploiting the cluster uncertainty concept. Bagherinia et al. [37] proposed a novel fuzzy clustering ensemble framework based on a new fuzzy diversity measure and a fuzzy quality measure to find the base clusterings with the best performance. In clustering ensemble, multiple clustering outputs can be combined to

produce better results in terms of consistency, robustness, and performance than the basic individual clustering methods.

The evaluation of clustering algorithms is an active issue in the fields such as machine learning, data mining, artificial intelligence, databases, and pattern recognition [11]. In a typical clustering scenario, three fundamental questions must be addressed: (i) identifying an effective clustering algorithm suitable for a given data set; (ii) determining how many clusters are presented in the data; and (iii) evaluating the clustering [38]. This article focuses on the first problem.

Several validity measures have been proposed to evaluate clustering algorithms. Yeung et al. [39] pointed out that the figure of merit (FOM) is used on microarray data, and different biological groups represent the clusters. Halkidi et al. [40] presented the Rand statistic to measure the proportion of pairs of vectors. Roth et al. [41] presented a stability measure to evaluate the partitioning validity and to choose the number of clusters. Chou et al. [14] presented a CS cluster relative measure to assess clusters with different sizes and densities. Žalik [42] presented a CO cluster-validity measure based on compactness and overlapping measures to estimate the quality of partitions. Chou et al. [43] presented an area measure to evaluate the initial cluster number based on the information of cluster areas. Wani and Riyaz [44] presented a new compactness measure using a novel penalty function to describe the typical behavior of a cluster. Azhagiri and Rajesh [45] proposed a novel approach to measure the quality of the cluster and can find intrusions using intrusion unearthing and probability clomp algorithm.

The validity measures are often divided into the types of internal, relative, and external measures [20, 21, 24, 25, 46]. Internal measures are based on computing properties of the resulting clusters, and these measures do not include additional information on the data [20, 25, 47]. Relative measures are based on the comparison of partitions produced by the same clustering algorithm with different data subsets or different parameters, and they do not demand additional information [20, 25, 39]. External measures compare the partitions produced by clustering algorithms with a given data partition [20, 25, 48]. These correspond to a kind of error measurement, so they can be supposed to offer improved correlation to the true error [20]. The results of Burn et al. [20] indicate that external measures for evaluating clustering results are more accurate than internal or relative measures. Thus, external measures are selected to assess the performance of clustering algorithms.

In addition, the evaluation of clustering algorithms involves more than one criterion. Thus, it can be solved by MCDM methods. This differs from previous approaches. For example, Dudoit and Fridlyand [49] proposed a prediction-based resampling method to evaluate the number of clusters, and Sugar and James [50] chose the number of clusters by an information-theoretical approach. Peng et al. [51] developed an MCDM-based method to select the number of clusters. Peng et al. [52] also developed a framework to select the appropriate clustering algorithm and to further choose the number of clusters. Meyer and Olteanu [53] indicated clustering in the field of multicriteria decision aid (MCDA)

has seen a few adaptations of methods from data analysis, most of them however using concepts native to that field, such as the notions of similarity and distance measures. Besides, Chen et al. [54] pointed out the clustering problem is one of the well-known MCDA problems, and the existing versions of the K-means clustering algorithm are only used for partitioning the data into several clusters which do not have priority relations; therefore, Chen et al. [54] proposed a complete ordered clustering algorithm called the ordered K-means clustering algorithm, which considers the preference degree between any two alternatives. Mahdiraji et al. [55] presented marketing strategies evaluation based on big data analysis by a clustering-MCDM approach. This paper takes a new perspective by proposing a DMSECA model based on the MCDM method, merging expert wisdom by using the eighty-twenty rule to select the best clustering algorithms for the given data sets during a complex decision process. Furthermore, our proposed DMSECA model can reconcile different or even conflicting evaluation performance to reach a group agreement for information fusion in a complex decision-making environment.

The eighty-twenty rule is proposed by Pareto [56], who researches the wealth distribution in different countries. The eighty-twenty rule is based on the observation that, in most countries, about 80% of the wealth is controlled by about 20% of the people, which is called a “predictable imbalance” by Pareto [57]. The eighty-twenty rule has been expanded to many fields such as sociology and quality control [58]. In this work, the eighty-twenty rule is used to focus on the analysis on the most important positions of the rankings in relation to the number of observations for predictable imbalance. The truth is often in a few hands: the views of about 20% of the people represent more satisfactory rankings in the opinion of all participants.

The decision-making process is extremely complex because of competing interests of multiple stakeholders and the intricacy of systems [8–10]. In this paper, the proposed DMSECA model, based on MCDM methods and the eighty-twenty rule, presents a new perspective by merging expert wisdom to evaluate the most appropriate clustering algorithm for the given data sets and the proposed model can reconcile individual differences or conflicts to achieve group agreements among clustering algorithm evaluations in a complex decision-making environment.

3. Preliminaries

This section presents some elementary and preparatory knowledge. It first introduces several evaluation approaches in Section 3.1, and then, the classic MCDM methods are presented in Section 3.2; finally, the performance measures of clustering algorithms are described in Section 3.3.

3.1. Clustering Algorithms. Clustering is a popular unsupervised learning technique. It aims to divide large data sets into smaller sections so that objects in the same cluster are lowly distinct, whereas objects in different clusters are lowly similar [21]. Clustering algorithms, based on similarity

criteria, can group patterns, where groups are sets of similar patterns [54, 59, 60]. Clustering algorithms are widely applied in many research fields, such as genomics, image segmentation, document retrieval, sociology, bio-informatics, psychology, business intelligence, and financial analysis [61–64].

Clustering algorithms are usually known as the four classes of partitioning methods, hierarchical methods, density-based methods, and model-based methods [65]. Several classic clustering algorithms are proposed and reported, such as the K-means algorithm [66], k -medoid algorithm [67], expectation maximization (EM) [68], and frequent pattern-based clustering [65]. In this paper, the six most influential clustering algorithms are selected for the empirical study. These are the KM algorithm, EM algorithm, filtered clustering (FC), farthest-first (FF) algorithm, make-density-based clustering (MD), and hierarchical clustering (HC). These clustering algorithms can be implemented by WEKA [69].

The KM algorithm, a partitioning method, takes the input parameter k and partitions a set of n objects into k clusters so that the resulting intracluster similarity is high, and the intercluster similarity is low. And the cluster similarity can be measured by the mean value of the objects in a cluster, which can be viewed as the centroid or center of gravity of the cluster [65].

The EM algorithm, which is considered as an extension of the KM algorithm, is an iterative method to find the maximum likelihood or maximum a posteriori estimates of parameters in statistical models, where the model depends on unobserved latent variables [70]. The KM algorithm assigns each object to a cluster.

In the EM algorithm, each object is assigned to each cluster according to a weight representing its probability of membership. In other words, there are no strict boundaries between the clusters. Thus, new means can be computed based on the weighted measures [68].

The FC applied in this work can be implemented by WEKA [69]. Like the cluster, the structure of the filter is based exclusively on the training data, and test instances will be addressed by the filter without changing their structure.

The FF algorithm is a fast, greedy, and simple approximation algorithm to the k -center problem [67], where the k points are first selected as a cluster center, and the second center is greedily selected as the point farthest from the first. Each remaining center is determined by greedily selecting the point farthest from the set of chosen centers, and the remaining points are added to the cluster whose center is the closest [66, 71].

The MD algorithm is a density-based method. The general idea is to continue growing the given cluster as long as the density (the number of objects or data points) in the neighborhood exceeds some threshold. That is, for each data point within a given cluster, the neighborhood of a given radius must contain a minimum number of points [65]. The HC algorithm is a method of cluster analysis that seeks to build a hierarchy of clusters, which can create a hierarchical decomposition of the given data sets [66, 72].

3.2. MCDM Methods. The MCDM methods, which were developed in the 1970s, are a complete set of decision analysis technologies that have evolved as an important research field of operation research [73, 74]. The International Society on MCDM defines MCDM as the research of methods and procedures concerning multiple conflicting criteria, which can be formally incorporated into the management planning process [73]. In an MCDM problem, the evaluation criteria are assumed to be independent [75, 76]. MCDM methods aim to assist decision-makers (DMs) to identify an optimal solution from a number of alternatives by synthesizing objective measurements and value judgments [77, 78]. In this section, four classic MCDM methods: the weighted sum method (WSM), grey relational analysis (GRA), TOPSIS, and PROMETHEE II are introduced as follows.

3.2.1. WSM. WSM [79] is a well-known MCDM method for evaluating finite alternatives in terms of finite decision criteria when all the data are expressed in the same unit [80, 81]. The benefit-to-cost-ratio and benefit-minus-cost approaches [82] can be applied to the problem of involving both benefit and cost criteria. In this paper, the cost criteria are first transformed to benefit criteria. Besides, there is nominal-the-better (NB), when the value is closer to the objective value, the nominal-the-better (NB) is better.

The calculation steps of WSM are as follows. First, assume n criteria, including benefit criteria and cost criteria, and m alternatives. The cost criteria are first converted to benefit criteria in the following standardization process.

- (1) The larger-the-better (LB): a larger objective value is better, that is, the benefit criteria, and it can be standardized as

$$x'_{ij} = \frac{x_{ij} - \min_i x_{ij}}{\max_i x_{ij} - \min_i x_{ij}}. \quad (1)$$

- (2) The smaller-the-better (SB): the smaller objective value is better, that is, the cost criteria, and it can be standardized as

$$x'_{ij} = \frac{\max_i x_{ij} - x_{ij}}{\max_i x_{ij} - \min_i x_{ij}}. \quad (2)$$

- (3) The nominal-the-better (NB): the closer to the objective value is better, and it can be standardized as

$$x'_{ij} = 1 - \frac{|x_{ij} - x_{ob}|}{\max\{\max_i x_{ij} - x_{ob}; x_{ob} - \min_i x_{ij}\}}. \quad (3)$$

Finally, the total benefit of all the alternatives can be calculated as

$$A_i = \sum_{j=1}^k w_j x'_{ij}, \quad 1 \leq i \leq m, 1 \leq j \leq n. \quad (4)$$

The larger WSM value indicates the better alternative.

3.2.2. GRA. GRA is a basic MCDM method of quantitative research and qualitative analysis for system analysis [83]. Based on the grey space, it can address inaccurate and incomplete information [84]. GRA has been widely applied in modeling, prediction, systems analysis, data processing, and decision-making [83, 85–88]. The principle is to analyze the similarity relationship between the reference series and alternative series [89]. The detailed steps are as follows.

Assume that the initial matrix is R :

$$R = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \quad (1 \leq i \leq m, 1 \leq j \leq n). \quad (5)$$

(1) Standardize the initial matrix:

$$R' = \begin{bmatrix} x'_{11} & x'_{12} & \cdots & x'_{1n} \\ x'_{21} & x'_{22} & \cdots & x'_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ x'_{m1} & x'_{m2} & \cdots & x'_{mn} \end{bmatrix} \quad (1 \leq i \leq m, 1 \leq j \leq n). \quad (6)$$

(2) Generate the reference sequence x'_0 :

$$x'_0 = (x'_0(1), x'_0(2), \dots, x'_0(n)). \quad (7)$$

where $x'_0(j)$ is the largest and standardized value in the j th factor.

(3) Calculate the differences $\Delta_{0i}(j)$ between the reference series and alternative series:

$$\Delta_{0i}(j) = |x'_0(j) - x'_{ij}|, \quad \Delta = \begin{bmatrix} \Delta_{01}(1) & \Delta_{01}(2) & \cdots & \Delta_{01}(n) \\ \Delta_{02}(1) & \Delta_{02}(2) & \cdots & \Delta_{02}(n) \\ \vdots & \vdots & \cdots & \vdots \\ \Delta_{0m}(1) & \Delta_{0m}(2) & \cdots & \Delta_{0m}(n) \end{bmatrix} \quad (1 \leq i \leq m, 1 \leq j \leq n). \quad (8)$$

(4) Calculate the grey coefficient $r_{0i}(j)$:

$$r_{0i}(j) = \frac{\min_i \min_j \Delta_{0i}(j) + \delta \max_i \max_j \Delta_{0i}(j)}{\Delta_{0i}(j) + \delta \max_i \max_j \Delta_{0i}(j)}, \quad (9)$$

where δ is a distinguished coefficient. The value of δ is generally set to 0.5 to provide good stability.

(5) Calculate the value of grey relational degree b_i :

$$b_i = \frac{1}{n} \sum_{j=1}^n r_{0i}(j). \quad (10)$$

(6) Finally, standardize the value of grey relational degree β_i :

$$\beta_i = \frac{b_i}{\sum_{i=1}^n b_i}. \quad (11)$$

3.2.3. TOPSIS. Developed by Hwang and Yoon [90], TOPSIS is one of the classic MCDM methods to rank alternatives over multicriteria. The principle is that the chosen alternative should have the shortest distance from the positive ideal solution (PIS) and the farthest distance from the negative ideal solution (NIS) [91]. TOPSIS can find the best alternative by minimizing the distance to the PIS and maximizing the distance to the NIS [92]. The alternatives can be ranked by their relative closeness to the ideal solution. The calculation steps are as follows [93]:

(1) The decision matrix A is standardized:

$$a_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^m (x_{ij})^2}} \quad (1 \leq i \leq m, 1 \leq j \leq n). \quad (12)$$

(2) The weighted standardized decision matrix is computed:

$$D = (a_{ij} * w_j) \quad (1 \leq i \leq m, 1 \leq j \leq n). \quad (13)$$

where the V^* are the criteria weights, and $\sum_{i=1}^m w_j = 1$.

(3) The PIS V^* and the NIS V^- are calculated:

$$V^* = \{v_1^*, v_2^*, \dots, v_n^*\} = \left\{ \left(\max_i v_{ij} \mid j \in J \right), \left(\min_i v_{ij} \mid j \in J' \right) \right\}, \\ V^- = \{v_1^-, v_2^-, \dots, v_n^-\} = \left\{ \left(\min_i v_{ij} \mid j \in J \right), \left(\max_i v_{ij} \mid j \in J' \right) \right\}. \quad (14)$$

(4) The distances of each alternative from PIS and NIS are determined:

$$S_i^+ = \sqrt{\sum_{j=1}^n (V_i^j - V^*)^2} \quad (1 \leq i \leq m, 1 \leq j \leq n), \\ S_i^- = \sqrt{\sum_{j=1}^n (V_i^j - V^-)^2} \quad (1 \leq i \leq m, 1 \leq j \leq n). \quad (15)$$

(5) The relative closeness to the ideal solution is obtained:

$$Y_i = \frac{S_i^-}{S_i^+ + S_i^-} \quad (1 \leq i \leq m), \quad (16)$$

where when R is closer to 1, the alternative is closer to the ideal solution.

(6) The preference order is ranked.

The larger relative closeness indicates the better alternative.

3.2.4. PROMETHEE II. PROMETHEE II, proposed by Brans in 1982, uses pairwise comparisons and “values outranking relations” to select the best alternative [94]. PROMETHEE II can support DMs to reach an agreement on feasible alternatives over multiple criteria from different perspectives [95, 96]. In the PROMETHEE II method, a

positive outranking flow reveals that the chosen alternative outranks all alternatives, whereas a negative outranking flow reveals that the chosen alternative is outranked by all alternatives [51, 97]. Based on the positive outranking flows and negative outranking flows, the final alternative can be selected and determined by the net outranking flow [98]. The steps are as follows:

- (1) Normalize the decision matrix R :

$$R_{ij} = \frac{x_{ij} - \min x_{ij}}{\max x_{ij} - \min x_{ij}} \quad (1 \leq i \leq n, 1 \leq j \leq m). \quad (17)$$

- (2) Define the aggregated preference indices. Let $a, b \in A$ and

$$\begin{cases} \pi(a, b) = \sum_{j=1}^k p_j(a, b)w_j, \\ \pi(b, a) = \sum_{j=1}^k p_j(b, a)w_j, \end{cases} \quad (18)$$

where A is a finite set of alternatives $\{a_1, a_2, \dots, a_n\}$, k is the number of criteria such that $1 \leq k \leq m$, w_j is the weight of criterion j , and $\sum_{j=1}^k w_j = 1$ ($1 \leq k \leq m$). $\pi(a, b)$ represents how a is preferred to b over all criteria, and $p_j(a, b)$ represents how b is preferred to a over all criteria. $p_j(a, b)$ and $p_j(b, a)$ are the preference functions of the alternatives a and b .

- (3) Calculate $\pi(a, b)$ and $\pi(b, a)$ for each pair of alternatives.

In general, there are six types of preference function. DMs must select one type of preference function and the corresponding parameter value for each criterion [51, 98].

- (4) Determine the positive outranking flow and negative outranking flow. The positive outranking flow is determined by

$$\phi^+(a) = \frac{1}{n-1} \sum_{x \in A} \pi(a, x), \quad (19)$$

and the negative outranking flow is determined by

$$\phi^-(a) = \frac{1}{n-1} \sum_{x \in A} \pi(x, a). \quad (20)$$

- (5) Calculate the net outranking flow:

$$\phi(a) = \phi^+(a) - \phi^-(a). \quad (21)$$

- (6) Determine the ranking according to the net outranking flow.

Larger $\phi(a)$ is the more appropriate alternative.

3.3. Performance Measures. Burn et al. [20] proposed that external measures for evaluating clustering results are more effective than internal and relative measures. Accordingly, in this study, nine clustering external measures are selected for

evaluation. These are entropy, purity, microaverage precision (MAP), Rand index (RI), adjusted Rand index (ARI), F-measure (FM), Fowlkes–Mallows index (FMI), Jaccard coefficient (JC), and Mirkin metric (MM). Among them, measures of entropy and purity are widely applied as external measures in the fields of data mining and machine learning [99, 100]. The nine external measures are generated by a computer with an Intel core i5-3210M CPU @ 2.50 GHz with 8G memory. Before introducing external measures, the contingency table is described.

3.3.1. The Contingency Table. Given a data set D with n objects, suppose we have a partition $P = \{P_1, P_2, \dots, P_n\}$ by some clustering method, where $\cup_{i=1}^k P_i = D$ and $P_i \cap P_j = \emptyset$, for $1 \leq i \neq j \leq k$. According to the preassigned class labels, we can create another partition on $C = \{C_1, C_2, \dots, C_k\}$, where $\cup_{i=1}^k C_i = D$ and $C_i \cap C_j = \emptyset$, for $1 \leq i \neq j \leq k$. Let n_{ij} denote the number of objects in cluster P_i with the label of class C_j . Then, the data information between the two partitions can be displayed in the form of a contingency table, as shown in Table 1 [65].

The following paragraphs define the external measures. The measures of entropy and purity are widely applied in the field of data mining and machine learning [99, 100].

(1) *Entropy.* The measure of entropy, which originated in the information-retrieval community, can measure the variance of a probability distribution. If all clusters consist of objects with only a single class label, the entropy is zero, and as the class labels of objects in a cluster become more varied, the entropy increases [101]. The measure of entropy is calculated as

$$E = - \sum_i \frac{n_i}{n} \left(\sum_j \frac{n_{ij}}{n_i} \log \frac{n_{ij}}{n_i} \right). \quad (22)$$

A lower entropy value usually indicates more effective clustering.

(2) *Purity.* The measure of purity pays close attention to the representative class (the class with majority objects within each cluster) [102]. Purity is similar to entropy. It is calculated as

$$P = \sum_i \frac{n_i}{n} \left(\max_j \frac{n_{ij}}{n_i} \right). \quad (23)$$

A higher purity value usually represents more effective clustering.

(3) *F-Measure.* The F-measure (FM) is a harmonic mean of precision and recall. It is commonly considered as clustering accuracy [103]. The calculation of FM is inspired by the information-retrieval metric as follows:

$$F\text{-measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}, \quad (24)$$

$$\text{precision} = \frac{n_{ij}}{n_j}, \quad \text{recall} = \frac{n_{ij}}{n_i}.$$

TABLE 1: Contingency table.

		Partition C				Σ
		C_1	C_2	$\dots$	C_k	
Partition P	P_1	n_{11}	n_{12}	$\dots$	n_{1k}	N_1
	P_2	n_{21}	n_{22}	$\dots$	n_{2k}	N_2
	P_k	n_{k1}	n_{k2}	$\dots$	n_{kk}	n_k
	Σ	N_1	n_2	$\dots$	n_k	n

A higher value of FM generally indicates more accurate clustering.

(4) *Microaverage Precision*. The MAP is usually applied in the information-retrieval community [104]. It can obtain a clustering result by assigning all data objects in a given cluster to the most dominant class label and then evaluating the following quantities for each class [60]:

- (1) $\alpha(C_j)$: the number of objects correctly assigned to class C_j .
- (2) $\beta(C_j)$: the number of objects incorrectly assigned to class C_j .

The MAP measure is computed as follows:

$$\text{MAP} = \frac{\sum_j \alpha(C_j)}{\sum_j \alpha(C_j) + \beta(C_j)}. \quad (25)$$

A higher MAP value indicates more accurate clustering.

(5) *Mirkin Metric*. The measure of Mirkin metric (MM) assumes the null value for identical clusters and a positive value, otherwise. It corresponds to the Hamming distance between the binary vector representations of each partition [105]. The measure of MM is computed as

$$M = \sum_i n_i^2 + \sum_j n_j^2 - 2 \sum_i \sum_j n_{ij}^2. \quad (26)$$

A lower value of MM implies more accurate clustering.

In addition, given a data set, assume a partition C is a clustering structure of a data set and P is a partition by some clustering method. We refer to a pair of points from the data set as follows:

- (i) SS: if both points belong to the same cluster of the clustering structure C and to the same group of the partition P
- (ii) SD: if the points belong to the same clusters of C and to different groups of P
- (iii) DS: if the points belong to different clusters of C and to the same groups of P
- (iv) DD: if the points belong to different clusters of C and to different groups of P

Assume that a , b , c , and d are the numbers of SS, SD, DS, and DD pairs, respectively, and that $M = a + b + c + d$, which is the maximum number of pairs in the data set. The following indicators for measuring the degree of similarity between C and P can be defined.

(6) *Rand Index*. The RI is a measure of the similarity between two data clusters in statistics and data clustering [106]. RI is computed as follows:

$$R = \frac{(a + d)}{M}. \quad (27)$$

A higher value of RI indicates a more accurate result of clustering.

(7) *Jaccard Coefficient*. The JC, also known as the Jaccard similarity coefficient (originally named the “coefficient de communauté” by Paul Jaccard), is a statistic applied to compare the similarity and diversity of sample sets [107]. JC is computed as follows:

$$J = \frac{a}{(a + b + c)}. \quad (28)$$

A higher value of JC indicates a more accurate result of clustering.

(8) *Fowlkes and Mallows Index*. The Fowlkes and Mallows index (FMI) was proposed by Fowlkes and Mallows [108] as an alternative for the RI. The measure of FMI is computed as follows:

$$\text{FMI} = \sqrt{\frac{a}{a+b} \cdot \frac{a}{a+c}}. \quad (29)$$

A higher value of FMI indicates more accurate clustering.

(9) *Adjusted Rand Index*. The adjusted Rand index (ARI) is the corrected-for-chance version of the measure of RI [106]. It ranges from -1 to 1 and expresses the level of concordance between two bipartitions [109]. A value of ARI closest to 1 indicates almost perfect concordance between the two compared bipartitions, whereas a value near -1 indicates almost complete discordance [110]. The measure of ARI is computed as

$$\text{ARI} = \frac{a - ((a + c) + (a + b)/M)}{((a + c) + (a + b)/2) - ((a + c) + (a + b)/M)}. \quad (30)$$

A higher value of ARI indicates more accurate clustering.

3.4. Index Weights. In this work, the index weights of the four MCDM methods can be calculated by AHP. The AHP method, proposed by Saaty [111], is a widely used tool for modeling unstructured problems by synthesizing subjective and objective information in many disciplines, such as politics, economics, biology, sociology, management science, and life sciences [112–114]. It can elicit a corresponding priority vector according to pair-by-pair comparison values [115] obtained from the scores of experts on an appropriate scale [116]. AHP has some problems, for example, the priority vector derived from the eigenvalue method can violate a condition of order preservation proposed by Costa and Vansnick [117]. However, AHP is still a classic and important approach, especially in the fields of operation research and management science [118]. AHP has the following steps:

- (1) Establish a hierarchical structure: a complex problem can be established in such a structure, including the goal level, criteria level, and alternative level [119, 120].
- (2) Determine the pairwise comparison matrix: once the hierarchy is structured, the prioritization procedure starts for determining the relative importance of the criteria (index weights) within each level [119, 121, 122]. The pairwise comparison values are obtained from the scores of experts on a 1–9 scale [116].
- (3) Calculate index weights: the index weights are usually calculated by the eigenvector method [120] proposed by Saaty [111].
- (4) Test consistency: the value of 0.1 is generally considered the acceptable upper limit of the consistency ratio (CR). If the CR exceeds this value, the procedure must be repeated to improve consistency [119, 121].

4. The Proposed Model

Clustering results can vary according to the evaluation method. Rankings can conflict even when abundant data are processed, and a large knowledge gap can exist between the evaluation results [123] due to the anticipation, experience, and expertise of all individual participants. The decision-making process is extremely complex. This makes it difficult to make accurate and effective decisions [124]. As mentioned in Section 1, the proposed DMSECA model consists of three steps. They are as follows.

The first step usually involves modeling by clustering algorithms, which can be accomplished using one or more procedures selected from the categories of hierarchical, density-based, partitioning, and model-based methods [65]. In this section, we apply the six most influential clustering algorithms, including EM, the FF algorithm, FC, HC, MD, and KM, for task modeling by using WEKA 3.7 on 20 UCI data sets, including a total of 18,310 instances and 313 attributes. Each of these clustering algorithms belongs to one of the four categories of clustering algorithms mentioned previously. Hence, all categories are represented.

In the second step, four commonly used MCDM methods (TOPSIS, WSM, GRA, and PROMETHEE II) are applied to rank the performance of the clustering algorithms over 20 UCI data sets based on nine external measures as the input, computed in the first step. These methods are highly suitable for the given data sets. Unsuitable methods were not selected. For example, we did not select VIKOR because its denominator would be zero for the given data sets. The index weights are determined by AHP based on the eigenvalue method. Three experts from the field of MCDM are selected and consulted as the DMs to derive the pairwise comparison values completed by the scores of experts. We randomly assign each MCDM method to five UCI data sets. We apply more than one MCDM method to analyze and evaluate the performance of clustering algorithms, which is essential.

Finally, in the third step, we propose a decision-making support model to reconcile the individual differences or even

conflicts in the evaluation performance of the clustering algorithms among the 20 UCI data sets. The proposed model can generate a list of algorithm priorities to select the most appropriate clustering algorithm for secondary mining and knowledge discovery. The detailed steps of the decision-making support model, based on the 80-20 rule, are described as follows.

Step 1. Mark two sets of alternatives in a lower position and an upper position, respectively.

It is well known that the eighty-twenty rule reports that eighty percent of the results originate in twenty percent of the activity in most situations [58]. The rule can be credited to Vilfredo Pareto [56], who observes that eighty percent of the wealth is usually controlled by twenty percent of the people in most countries [57]. The implication is that it is better to be in the top of 20% than in the bottom of 80%. So, the eighty-twenty rule, introduced in Section 5, can be applied to focus on the analysis of the most important positions of the rankings in relation to the number of observations for predictable imbalance. The eighty-twenty rule indicates that the twenty percent of people, who are creating eighty percent of the results, which are highly leveraged. In this research, based on the expert wisdom originating from the twenty percent of people, the set of alternatives is classified into two categories, where the top of 1/5 of the alternatives is marked in an upper position, which represents more satisfactory rankings from the opinion of all individual participants involved in the algorithm evaluation process. The bottom of 1/5 is in a lower position, which represents more dissatisfactory rankings from the opinion of all individual participants. The element marked in the upper position is calculated as follows:

$$x = \frac{n * 1}{5}, \quad (31)$$

where n is the number of alternatives. For instance, if $n = 7$, then $x = 7 \times 1/5 = 1.4 \approx 2$. Hence, the second position classifies the ranking, where the first and second positions are those alternatives in the upper position, which are considered as the collective group idea of the most appropriate and satisfactory alternatives.

Similarly, the element marked in the lower position is calculated as

$$x = \frac{n * 4}{5}, \quad (32)$$

where n is the number of alternatives. For instance, if $n = 7$, then $7 * 4/5 = 5.6 \approx 6$. Thus, the sixth position classifies the ranking, where the sixth and seventh positions in the lower positions are considered collectively as the worst and most dissatisfactory alternatives.

Step 2. Grade the sets of alternatives in the lower and upper positions, respectively.

A score is assigned to each position of the set of alternatives in the lower position and upper position, respectively.

The score in the lower position can be calculated by assigning a value of 1 to the first position, 2 to the second position, ..., and x to the last position. Finally, the score of each alternative in the lower position is totaled, marked d .

Similarly, the score in the upper position can be calculated by assigning a value of 1 to the last position, 2 to the penultimate position, ..., and x to the first position. Finally, the score of each alternative in the upper position is totaled, marked b .

Step 3. Generate the priority of each alternative.

The priority of each alternative f_i , which represents the most satisfactory rankings from the opinions of all individual participants, can be determined as

$$f_i = b_i - d_i, \quad (33)$$

where a higher value of f_i implies a higher priority.

5. Experimental Design and Results

We now present an experiment on 20 UCI data sets. This is designed to test and verify our proposed DMSECA model for performance evaluation of clustering algorithms in order to reconcile individual differences or even conflicts in the evaluation performance of clustering algorithms based on MCDM in a complex decision-making environment. The experimental data sets, experimental design, and experimental results are as follows.

5.1. Data Sets. A total of 20 datasets are applied for performance evaluation of clustering algorithms in the experiment. They are originated from the UCI repository (<http://archive.ics.uci.edu/ml/>) [125]. These 20 datasets, which include the structures and characteristics of data set characteristics, attribute characteristics, number of instances, number of attributes, and area, are including Liver Disorders Data Set (<http://archive.ics.uci.edu/ml/datasets/Liver+Disorders>), Wine Data Set (<http://archive.ics.uci.edu/ml/datasets/Wine>), Teaching Assistant Evaluation Data Set (<http://archive.ics.uci.edu/ml/datasets/Teaching+Assistant+Evaluation>), Wholesale Customers Data Set (<http://archive.ics.uci.edu/ml/datasets/Wholesale+customers>), Haberman's Survival Data Set (<http://archive.ics.uci.edu/ml/datasets/Haberman%27s+Survival>), Balance Scale Data Set (<http://archive.ics.uci.edu/ml/datasets/Balance+Scale>), Contraceptive Method Choice Data Set (<http://archive.ics.uci.edu/ml/datasets/Contraceptive+Method+Choice>), Page Blocks Classification Data Set (<http://archive.ics.uci.edu/ml/datasets/Page+Blocks+Classification>), Breast Tissue Data Set (<http://archive.ics.uci.edu/ml/datasets/Breast+Tissue>), Blood Transfusion Data Set (<http://archive.ics.uci.edu/ml/datasets/Blood+Transfusion+Service+Center>), and Yeast Data Set (<http://archive.ics.uci.edu/ml/datasets/Yeast>). Table 2 concludes data information of these datasets. These datasets summarize a total of 18,310 instances and 313 attributes from a variety of disciplines, such as life sciences, business, physical sciences, social sciences, and CS/Engineering. The data sets have a variety of data structures. Their

sizes range from 100 to 4,601, the number of attributes from 3 to 60, and the number of classes from 2 to 10.

5.2. Experimental Design. In this section, the experimental design is described in detail to examine the feasibility and effectiveness of our proposed DMSECA model. The DMSECA model can be verified by applying the four MCDM methods introduced in Section 3.2 to estimate the performance of the clustering algorithms for the 20 selected public-domain UCI machine learning data sets. Each MCDM method is randomly assigned to five UCI data sets. The experimental design can be implemented as follows:

Input: 20 UCI data sets.

Output: rankings of evaluation performance of clustering algorithms to generate a list of algorithm priorities in order to select the best clustering algorithm and reconcile individual disagreements among their evaluations.

Step 1: prepare target data sets: data preprocessing to delete class labels of the original data sets.

Step 2: obtain clustering solutions: obtain clustering solutions of the six classic clustering algorithms introduced in Section 3.1 by WEKA based on the target data sets.

Step 3: calculate the values of nine external measures of each data set.

Step 4: obtain the weights of external measures. In this paper, the weights of external measures are obtained by AHP based on the eigenvalue method, which is scored by three invited and consulted experts.

Step 5: use WSM, TOPSIS, PROMETHEE II, and GRA to generate rankings of evaluation performance of clustering algorithms. Each MCDM method is randomly assigned to one of the five UCI data sets. The four MCDM methods are implemented using MATLAB 7.0, using the external measures as the input.

Step 6: achieve consensus. The consensus on different or even conflicting individual rankings of evaluation performance of clustering algorithms can be achieved by using the proposed decision-making support model in the third step, which merges expert wisdom.

Step 7: generate a list of algorithm priorities. The list can reconcile individual disagreements among the evaluation performance of clustering algorithms.

Step 8: end.

5.3. Experimental Results. This section gives the obtained results by testing the proposed DMSECA model on the 20 UCI datasets including a total of 18,310 instances and 313 attributes to reconcile those individual differences or conflicts among evaluation performance of clustering algorithms. The six clustering algorithms, nine external measures, and four MCDM methods are applied to illustrate and explain our model. The experimental results are as follows.

TABLE 2: Data information of the 20 data sets.

Data sets	No	Area	Number of instances	Number of attributes	Number of classes
Liver Disorders	1	Life sciences	345	7	2
ZOO	2	Life sciences	101	17	2
Pima Indians Diabetes	3	Life sciences	768	8	2
Wholesale Customers	4	Business	440	8	2
Haberman's Survival	5	Life sciences	306	3	2
Wine	6	Physical sciences	178	13	3
Balance Scale	7	Social sciences	625	4	3
Breast Tissue	8	Life sciences	106	10	6
Ecoli	9	Life sciences	336	8	8
Fertility	10	Life sciences	100	10	2
Ionosphere	11	Physical sciences	351	34	2
Iris	12	Life sciences	150	4	3
Teaching Assistant Evaluation	13	Other	151	5	3
Blood Transfusion	14	Business	748	5	2
Spambase	15	CS/Engineering	4,601	57	2
Page Blocks Classification	16	CS/Engineering	5,473	10	5
Sonar	17	Physical sciences	208	60	2
Contraceptive Method Choice	18	Life sciences	1,473	9	3
Dermatology	19	Life sciences	366	33	6
Yeast Data	20	Life sciences	1,484	8	10
Total			18,310	313	70

First, the values of the nine external measures of the 20 data sets can be obtained using the selected six clustering algorithms. The process is implemented according to Steps 1–3 in Section 5.2. To facilitate understanding, we have selected the Ionosphere data set as an example to explain the computational process. The initial values of the nine external measures, which are provided in Table 3, are standardized by equations (1)–(3) to transform cost criteria to benefit criteria. The standardized data are presented in Table 4. We highlight the optimal result of each external measure in boldface. It is clear that no clustering algorithm obtains the optimal results for all external measures. This supports the NFL theorem.

Second, the rankings of the clustering algorithms on the 20 data sets computed by SWM, TOPSIS, GRA, and PROMETHEE II are presented in Tables 5–8, respectively. The four MCDM methods are implemented using MATLAB 7.0 using the external measures, such as Purity, En, FM, and Rand as the input based on Tables 3 and 4. Each group of five UCI data sets can be processed by one of the four MCDM methods, which are randomly assigned. The measure weights of each expert applied in WSM, TOPSIS, GRA, and PROMETHEE II are obtained by AHP based on the eigenvalue method. The final index weights of three experts can be obtained by the weighted arithmetic mean to aggregation, which has been a widely used aggregation algorithm in the decision problems. The final index weights for the nine external measures, in the order given in Tables 4 and 5, are 0.1893, 0.1820, 0.0449, 0.0930, 0.0483, 0.1264, 0.1234, 0.1159, and 0.0769, respectively.

The results in Tables 5–8 do not enable us to identify and determine a regular pattern of evaluation performance of the clustering algorithms. The results indicate that various MCDM methods generate conflicting rankings. On the basis

of these observed results, secondary mining and knowledge discovery are proposed to reconcile these disagreements.

Finally, a decision-making support model based on the eighty-twenty rule for secondary mining and knowledge discovery is applied to reconcile individual disagreements. This model includes three steps as follows.

In Step 1, mark two sets of alternatives in a lower position and an upper position, respectively. According to equations (31) and (32), in the upper position, we know that $n=6$, and then $x = 6 \times 1/5 = 1.2 \approx 2$. Thus, the second position classifies the ranking, where the first and second positions are those alternatives in the upper position. Similarly, in the lower position, we have $x = 6 \times 4/5 = 4.8 \approx 5$. Hence, the fifth position classifies the ranking, where the fifth and sixth positions are those alternatives in the lower position. The two sets of alternatives in the lower position and upper position can be marked, and they are presented in boldface in Table 9, based on Tables 5–8.

In Step 2, grade the sets of alternatives in the lower and upper positions, respectively, according to Step 2 in Section 4. The scores of alternatives in the upper position d_i can be totaled. Similarly, the scores of alternatives in the lower position d_i can be totaled. Then, the results are presented in Table 10 for the 20 UCI data sets.

In Step 3, the priority of each alternative is computed by equation (33), and the calculation results are reported in Table 10.

5.4. Discussion and Analysis. The results in Tables 5–8 indicate that different MCDM methods produce different or even conflicting individual rankings. Thus, it is difficult for DMs to identify the best clustering algorithms for the given data sets. Table 10 reports a list of algorithm priorities. The

TABLE 3: Initial values of the nine external measures for the Ionosphere data set.

	Purity	En	F-m	Rand	ARI	Jaccard	FM	MAP	M
EM	0.9003	0.0331	0.1109	0.5897	0.0001	0.5689	0.7411	0.9003	0.4839
FF	0.6638	0.0506	0.3859	0.8091	0.0011	0.7705	0.8747	0.6638	0.3089
FC	0.9117	0.0296	0.0999	0.5954	0.0001	0.5774	0.7484	0.9117	0.4818
HC	0.6439	0.0356	0.4020	0.8177	0.0012	0.7785	0.8819	0.6439	0.2982
MD	0.8746	0.0408	0.1339	0.5783	0.0001	0.5502	0.7250	0.8746	0.4877
KM	0.9117	0.0299	0.0994	0.5983	0.0001	0.5791	0.7502	0.9117	0.4807

TABLE 4: Standardized values of the nine external measures for the Ionosphere data set.

	Purity	En	F-m	Rand	ARI	Jaccard	FM	MAP	M
EM	0.1748	0.1670	0.1579	0.1589	0.1666	0.1596	0.1619	0.1748	0.1608
FF	0.1514	0.1655	0.1833	0.1816	0.1667	0.1803	0.1757	0.1514	0.1778
FC	0.1761	0.1672	0.1570	0.1595	0.1666	0.1604	0.1627	0.1761	0.1610
HC	0.1495	0.1668	0.1850	0.1825	0.1667	0.1812	0.1765	0.1495	0.1788
MD	0.1721	0.1663	0.1598	0.1578	0.1666	0.1578	0.1604	0.1721	0.1605
KM	0.1761	0.1672	0.1570	0.1597	0.1666	0.1606	0.1628	0.1761	0.1611

TABLE 5: Rankings of WSM for the five assigned UCI data sets.

	ZOO		Balance Scale		Teaching Assistant Evaluation		Spambase		Yeast data	
	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank
EM	0.1677	2	0.1701	1	0.1547	6	0.1650	6	0.1719	2
FF	0.1653	5	0.1651	3	0.1684	4	0.1652	4	0.1790	1
FC	0.1677	2	0.1648	5	0.1727	1	0.1695	1	0.1644	5
HC	0.1638	6	0.1701	1	0.1595	5	0.1652	4	0.1560	3
MD	0.1676	4	0.1650	4	0.1721	3	0.1656	3	0.1645	4
KM	0.1679	1	0.1648	5	0.1727	1	0.1695	1	0.1643	6

TABLE 6: Rankings of TOPSIS for the five assigned UCI data sets.

	Pima Indians Diabetes		Wholesale Customers		Wine		Ecoli		Ionosphere	
	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank
EM	0.0866	6	0.1792	4	0.1859	5	0.1991	2	0.1797	3
FF	0.1102	5	0.1019	6	0.0661	6	0.3061	1	0.1427	5
FC	0.2019	1	0.2053	1	0.1870	1	0.1315	5	0.1858	2
HC	0.2019	1	0.1028	5	0.1870	1	0.0962	6	0.1406	6
MD	0.1974	4	0.2055	1	0.1870	1	0.1335	4	0.1646	4
KM	0.2019	1	0.2053	1	0.1870	1	0.1336	3	0.1865	1

TABLE 7: Rankings of the GRA for the five assigned UCI data sets.

	Breast Tissue		Fertility		Iris		Contraceptive Method Choice		Dermatology	
	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank
EM	0.1672	4	0.1379	4	0.1325	6	0.1850	3	0.1771	3
FF	0.1378	6	0.2142	2	0.1712	2	0.1366	5	0.1643	5
FC	0.1804	3	0.1362	6	0.1712	2	0.1857	1	0.1811	1
HC	0.1499	5	0.2321	1	0.1825	1	0.1229	6	0.1214	6
MD	0.1819	2	0.1416	3	0.1712	2	0.1842	4	0.1750	4
KM	0.1828	1	0.1379	4	0.1712	2	0.1857	1	0.1811	1

TABLE 8: Rankings of the PROMETHEE II for the five assigned UCI data sets.

	Liver Disorders		Haberman's Survival		Blood Transfusion Service Center		Page Blocks Classification		Sonar	
	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank
EM	0.1654	5	0.1133	6	0.1088	6	0.1252	6	0.1644	3
FF	0.1688	1	0.1766	4	0.1815	4	0.1867	5	0.1618	4
FC	0.1667	3	0.1780	1	0.1906	1	0.1413	3	0.1609	5
HC	0.1645	6	0.1780	1	0.1906	1	0.2371	1	0.1749	2
MD	0.1679	2	0.1762	5	0.1380	5	0.1685	2	0.1770	1
KM	0.1667	3	0.1780	1	0.1906	1	0.1413	3	0.1609	5

TABLE 9: Rankings of four MCDM methods for a total of 20 UCI data sets.

Rank	Data set						
	ZOO	Balance Scale	Teaching Assistant Evaluation	Spambase	Yeast Data	Pima Indians Diabetes	Wholesale Customers
1	KM	EM	FC	FC	FF	KM	FC
2	FC	HC	KM	KM	EM	FC	KM
3	EM	FF	MD	MD	HC	HC	MD
4	MD	MD	FF	FF	MD	MD	EM
5	FF	FC	HC	HC	FC	FF	HC
6	HC	KM	EM	EM	KM	EM	FF

Rank	Data set						
	Wine	<i>Ecoli</i>	Ionosphere	Breast Tissue	Fertility	Iris	Contraceptive Method Choice
1	FC	FF	KM	KM	HC	HC	KM
2	KM	EM	FC	MD	FF	KM	FC
3	MD	KM	EM	FC	MD	FC	EM
4	HC	MD	MD	EM	KM	FF	MD
5	EM	FC	FF	HC	EM	MD	FF
6	FF	HC	HC	FF	FC	EM	HC

Rank	Data set					
	Dermatology	Liver Disorders	Haberman's Survival	Blood Transfusion Service	Page Blocks Classification	Sonar
1	FC	FF	KM	KM	HC	MD
2	KM	MD	FC	FC	MD	HC
3	EM	FC	HC	HC	KM	EM
4	MD	KM	FF	FF	FC	FF
5	FF	EM	MD	MD	FF	FC
6	HC	HC	EM	EM	EM	KM

TABLE 10: Priority of each alternative.

Position Score	1st	2 nd	b_i	5 th	6th	d_i	f_i	Ranking
	2	1		1	2			
EM	1	2	4	3	7	17	-13	6
FF	3	1	7	6	3	12	-5	4
FC	5	6	16	4	1	6	10	2
HC	3	2	8	4	6	16	-8	5
MD	1	3	5	3	0	3	2	3
KM	7	6	20	0	3	6	14	1

TABLE 11: Statistical analysis of rankings for all 20 UCI data sets.

Algorithm	Ranking					
	1	2	3	4	5	6
EM	1	3	4	3	2	7
FF	3	2	1	5	6	3
FC	9	3	3	0	4	1
HC	8	1	1	1	3	6
MD	3	4	3	8	2	0
KM	12	1	3	1	2	1

rankings of the clustering algorithms are 6, 4, 2, 5, 3, and 1, which are in line with EM, FF, FC, HC, MD, and KM. Thus, the best clustering algorithm for the given data sets is the KM algorithm. In addition, we conduct a statistical analysis of the rankings obtained for the 20 UCI data sets to compare the results generated by our proposed model. The analysis results are reported in Table 11.

In Table 11, the number of each position ranking can be determined according to Tables 5–8. For example, for ranking 1 of the upper position, the numbers of clustering algorithms are 1, 3, 9, 8, 3, and 12, respectively, and the rankings of the clustering algorithms are 6, 4/5, 2, 3, 4/5, and 1 corresponding to EM, FF, FC, HC, MD, and KM, respectively. However, the rankings of the lower positions are

ignored. When making decisions, the overall situation affected by the decision-making process should be considered to the maximum extent. In this work, we establish two sets of alternatives in the lower and upper positions. After the rankings of the lower position are fully considered, the rankings of the clustering algorithms are 6, 4, 2, 5, 3, and 1, respectively. These results are basically the same, which shows that our proposed model is feasible and effective.

Therefore, in this paper, from an empirical perspective, the effectiveness of our proposed model is examined and verified using six clustering algorithms, nine external measures, and four MCDM methods on 20 UCI data sets, including a total of 18,310 instances and 313 attributes. Moreover, our proposed model merges expert wisdom using the eighty-twenty rule, which reports that eighty percent of the results originate from twenty percent of the activity [58] and indicates that the twenty percent of people who are creating eighty percent of the results are highly leveraged. Thus, based on the expert wisdom originating from the twenty percent of the people, the set of alternatives is classified into two categories, where the top of 1/5 of the alternatives is marked in an upper position, and the bottom of 1/5 is marked in a lower position. The empirical results also verify our proposed model and confirm its ability to reduce and reconcile individual differences among the performance of clustering algorithms by employing a list of algorithm priorities in a complex decision environment.

6. Conclusions

Data clustering is often widely applied in the initial stage of big data analysis. Clustering analysis can be used to examine massive data sets of a variety of types to uncover unknown correlations, hidden patterns, and other potentially useful information. However, Naldi et al. [11] pointed out that different clustering algorithms may produce different data partitions. Furthermore, the NFL theorem states that there exists no single algorithm or model that can achieve the best performance for a given domain problem [23–25]. Therefore, the focused question becomes how to select the best clustering algorithms for the given data sets.

The decision-making process is extremely complex because of competing interests of multiple stakeholders and the intricacy of systems [8–10]. This paper proposes a DMSECA model to estimate the performance of clustering algorithms in selecting the most satisfactory clustering algorithm according to the decision preferences of all individual participants during a complex decision-making process. The proposed model has been designed to reconcile individual disagreements in the evaluation performance of clustering algorithms. The studies have shown that the DMSECA model, which is based on the eighty-twenty rule, can generate a list of algorithm priorities and an optimal ranking scheme that is the most satisfactory according to the decision preferences of all individual participants involved in a complex decision-making problem. An experimental study involves the use of 20 UCI data sets, including a total of 18,310 instances and 313 attributes, six clustering

algorithms, nine external measures, and four MCDM methods in order to test and examine our proposed model.

The feasibility and effectiveness of the proposed model are illustrated and verified by carrying out a statistical analysis of rankings for a total of 20 UCI data sets to allow for a comparison of the results with those generated by our proposed model. The results are basically the same as the rankings of the clustering algorithms produced by our proposed DMSECA model. The empirical results show that our proposed model cannot only identify the best clustering algorithms for the given data sets but also can reconcile individual differences or even conflicts to achieve group agreement among the evaluation performance of clustering algorithms in a complex decision-making environment. Finally, a decision-making support model is proposed by merging expert wisdom for secondary knowledge discovery, based on the 80-20 rule, in order to focus the analysis on the most important positions of the rankings in relation to the number of observations for predictable imbalance.

In future work, a decision support system including data space, method space, model space, and knowledge space will be further developed, which can deal with much more methods/models/algorithms, such as general clustering, theory of subspace clustering, fuzzy clustering, and density peak clustering, in order to form a robust and effective algorithm selection and evaluation framework for improving the universality of the application.

Data Availability

The data used to support the findings of this study are included within the article, and a total of 20 datasets are originated from the UCI repository (<http://archive.ics.uci.edu/ml/>).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was partially supported by Grants from Fund for Less Developed Regions of the National Natural Science Foundation of China (#71761014), the State Key Program of the National Natural Science Foundation of China (#71532007 #71932008, and #91546201), the General Program of the National Natural Science Foundation of China (#71471149), Major Project of the National Social Science Foundation of China (#15ZDB153), and the Postdoctoral Science Foundation Project of China (#2016M592683).

References

- [1] Z. Xu, J. Chen, and J. Wu, "Clustering algorithm for intuitionistic fuzzy sets," *Information Sciences*, vol. 178, no. 19, pp. 3775–3790, 2008.
- [2] W. Hang, K. S. Choi, and S. Wang, *Synchronization Clustering Based on Central Force Optimization and its Extension for Large-Scale Data Sets*, Elsevier Science Publishers B.V., Amsterdam, Netherlands, 2017.

- [3] M. Abavisani and V. M. Patel, "Multi-modal sparse and low-rank subspace clustering," *Information Fusion*, vol. 39, pp. 168–177, 2018.
- [4] X. Zhang and Z. Xu, "Hesitant fuzzy agglomerative hierarchical clustering algorithms," *International Journal of Systems Science*, vol. 46, no. 3, pp. 562–576, 2015.
- [5] Y. Wang, Z. Sun, and K. Jia, *An Automatic Decoding Method for Morse Signal Based on Clustering Algorithm*, Springer International Publishing, Berlin, Germany, 2017.
- [6] C. Zhang, L. Hao, and L. Fan, "Optimization and improvement of data mining algorithm based on efficient incremental kernel fuzzy clustering for large data," *Cluster Computing*, vol. 22, no. S2, pp. 3001–3010, 2018.
- [7] X. Yang, Z. Xu, and H. Liao, "Correlation coefficients of hesitant multiplicative sets and their applications in decision making and clustering analysis," *Applied Soft Computing*, vol. 61, pp. 935–946, 2017.
- [8] J. C. Ascough II, H. R. Maier, J. K. Ravalico, and M. W. Strudley, "Future research challenges for incorporation of uncertainty in environmental and ecological decision-making," *Ecological Modelling*, vol. 219, no. 3–4, pp. 383–399, 2008.
- [9] Z. Xu and N. Zhao, "Information fusion for intuitionistic fuzzy decision making: an overview," *Information Fusion*, vol. 28, pp. 10–23, 2016.
- [10] Z. Xu and H. Wang, "On the syntax and semantics of virtual linguistic terms for information fusion in decision making," *Information Fusion*, vol. 34, pp. 43–48, 2017.
- [11] M. C. Naldi, A. C. P. L. F. Carvalho, and R. J. G. B. Campello, "Cluster ensemble selection based on relative validity indexes," *Data Mining and Knowledge Discovery*, vol. 27, no. 2, pp. 259–289, 2013.
- [12] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 8, pp. 841–847, 1991.
- [13] U. Maulik and S. Bandyopadhyay, "Performance evaluation of some clustering algorithms and validity indices," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 12, pp. 1650–1654, 2002.
- [14] C. H. Chou, M. C. Su, and E. Lai, "A new cluster validity measure and its application to image compression," *Pattern Analysis and Applications*, vol. 7, pp. 205–220, 2004.
- [15] S. Sriparna and M. Ujjwal, "Use of symmetry and stability for data clustering," *Evolutionary Intelligence*, vol. 3, no. 3–4, pp. 103–122, 2010.
- [16] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *Journal of Cybernetics*, vol. 3, no. 3, pp. 32–57, 1973.
- [17] S. Mahallati, J. C. Bezdek, D. Kumar, M. R. Popovic, and T. A. Valiante, "Interpreting cluster structure in waveform data with visual assessment and Dunn's index," *Frontiers in Computational Intelligence*, Springer, Cham, Switzerland, pp. 73–101, 2017.
- [18] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, 1979.
- [19] V. Bolandi, A. Kadkhodaie, and R. Farzi, "Analyzing organic richness of source rocks from well log data by using SVM and ANN classifiers: a case study from the Kazhdumi formation, the Persian Gulf basin, offshore Iran," *Journal of Petroleum Science and Engineering*, vol. 151, pp. 224–234, 2017.
- [20] M. Brun, C. Sima, J. Hua et al., "Model-based evaluation of clustering validation measures," *Pattern Recognition*, vol. 40, no. 3, pp. 807–824, 2007.
- [21] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering," *ACM Computing Surveys (CSUR)*, vol. 31, no. 3, pp. 264–323, 1999.
- [22] Y. Abdullahi, B. Coetzee, and L. van den Berg, "Relationships between results of an internal and external match load determining method in male, singles badminton players," *Journal of Strength and Conditioning Research*, vol. 33, no. 4, pp. 1111–1118, 2019.
- [23] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 67–82, 1997.
- [24] G. Kou and W. Wu, "An analytic hierarchy model for classification algorithms selection in credit risk analysis," *Mathematical Problems in Engineering*, vol. 2014, no. 1, Article ID 297563, 2014.
- [25] D. G. Guillén and A. R. Espinosa, "A meta-analysis on classification model performance in real-world datasets: an exploratory view," *Applied Artificial Intelligence*, vol. 31, no. 9–10, pp. 715–732, 2018.
- [26] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.
- [27] H. Parvin and B. Minaei-Bidgoli, "A clustering ensemble framework based on selection of fuzzy weighted clusters in a locally adaptive clustering algorithm," *Pattern Analysis and Applications*, vol. 18, no. 1, pp. 87–112, 2015.
- [28] J. Jiang, D. Hao, Y. Chen, M. Parmar, and K. Li, "GDPC: gravitation-based density peaks clustering algorithm," *Physica A: Statistical Mechanics and Its Applications*, vol. 502, pp. 345–355, 2018.
- [29] J. Jiang, W. Zhou, L. Wang, X. Tao, and K. Li, "HaloDPC: An improved recognition method on halo node for density peak clustering algorithm," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 33, no. 8, 2019.
- [30] X. Tao, *Research and Improvement on Density Peak Clustering Algorithm and Application for Earthquake Classification*, Changchun: Jilin University of Finance and Economics (D), Changchun, China, 2017, In Chinese.
- [31] H. Alizadeh, B. Minaei-Bidgoli, and H. Parvin, "Optimizing fuzzy cluster ensemble in string representation," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 27, no. 2, 2013.
- [32] H. Parvin and B. Minaei-Bidgoli, "A clustering ensemble framework based on elite selection of weighted clusters," *Advances in Data Analysis and Classification*, vol. 7, no. 2, pp. 181–208, 2013.
- [33] S.-o. Abbasi, S. Nejatian, H. Parvin, V. Rezaie, and K. Bagherifard, "Clustering ensemble selection considering quality and diversity," *Artificial Intelligence Review*, vol. 52, no. 2, pp. 1311–1340, 2019.
- [34] M. Mojarad, H. Parvin, S. Nejatian, and V. Rezaie, "Consensus function based on clusters clustering and iterative fusion of base clusters," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 27, no. 1, pp. 97–120, 2019.
- [35] M. Mojarad, S. Nejatian, H. Parvin, and M. Mohammadpoor, "A fuzzy clustering ensemble based on cluster clustering and iterative fusion of base clusters," *Applied Intelligence*, vol. 49, no. 7, pp. 2567–2581, 2019.
- [36] F. Rashidi, S. Nejatian, H. Parvin, and V. Rezaie, "Diversity based cluster weighting in cluster ensemble: an information theory approach," *Artificial Intelligence Review*, vol. 52, no. 2, pp. 1341–1368, 2019.
- [37] A. Bagherinia, B. Minaei-Bidgoli, M. Hossinzadeh, and H. Parvin, "Elite fuzzy clustering ensemble based on

- clustering diversity and quality measures,” *Applied Intelligence*, vol. 49, no. 5, pp. 1724–1747, 2019.
- [38] S. Saha and S. Bandyopadhyay, “Some connectivity based cluster validity indices,” *Applied Soft Computing*, vol. 12, no. 5, pp. 1555–1565, 2012.
- [39] K. Y. Yeung, D. R. Haynor, and W. L. Ruzzo, “Validating clustering for gene expression data,” *Bioinformatics*, vol. 17, no. 4, pp. 309–318, 2001.
- [40] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, “On clustering validation techniques,” *Journal of Intelligent Information Systems*, vol. 17, no. 2-3, pp. 107–145, 2001.
- [41] V. Roth, M. Braun, T. Lange, and J. M. Buhmann, *Stability-Based Model Order Selection in Clustering with Applications to Gene Expression Data*, Springer, Berlin, Germany, 2002.
- [42] K. R. Žalik, “Cluster validity index for estimation of fuzzy clusters of different sizes and densities,” *Pattern Recognition*, vol. 43, no. 10, pp. 3374–3390, 2010.
- [43] C. H. Chou, Y. X. Zhao, and H. P. Tai, “Vanishing-point detection based on a fuzzy clustering algorithm and new clustering validity measure,” *Journal of Applied Science and Engineering*, vol. 18, no. 2, pp. 105–116, 2015.
- [44] M. A. Wani and R. Riyaz, “A new cluster validity index using maximum cluster spread based compactness measure,” *International Journal of Intelligent Computing and Cybernetics*, vol. 9, no. 2, pp. 179–204, 2016.
- [45] M. Azhagiri and A. Rajesh, “A novel approach to measure the quality of cluster and finding intrusions using intrusion unearthing and probability clomp algorithm,” *International Journal of Information Technology*, vol. 10, no. 3, pp. 329–337, 2018.
- [46] F. Azuaje, “A cluster validity framework for genome expression data,” *Bioinformatics*, vol. 18, no. 2, pp. 319–320, 2002.
- [47] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley, New York, NY, USA, 2002.
- [48] E. R. Dougherty, J. Barrera, M. Brun et al., “Inference from clustering with application to gene-expression microarrays,” *Journal of Computational Biology*, vol. 9, no. 1, pp. 105–126, 2002.
- [49] S. Dudoit and J. Fridlyand, “A prediction-based resampling method for estimating the number of clusters in a dataset,” *Genome Biology*, vol. 3, Article ID research0036.1, 2002.
- [50] C. A. Sugar and G. M. James, “Finding the number of clusters in a dataset,” *Journal of the American Statistical Association*, vol. 98, no. 463, pp. 750–763, 2003.
- [51] Y. Peng, Y. Zhang, G. Kou, and Y. Shi, “A multi-criteria decision making approach for estimating the number of clusters in a data set,” *PLoS One*, vol. 7, no. 7, Article ID e41713, 2012.
- [52] Y. Peng, Y. Zhang, G. Kou, J. Li, and Y. Shi, “Multi-criteria decision making approach for cluster validation,” in *Proceedings of the International Conference on Computational Science*, pp. 1283–1291, Omaha, NE, USA, 2012.
- [53] P. Meyer and A.-L. Olteanu, “Formalizing and solving the problem of clustering in MCDA,” *European Journal of Operational Research*, vol. 227, no. 3, pp. 494–502, 2013.
- [54] L. Chen, Z. Xu, H. Wang, and S. Liu, “An ordered clustering algorithm based on K-means and the PROMETHEE method,” *International Journal of Machine Learning and Cybernetics*, vol. 9, no. 6, pp. 917–926, 2018.
- [55] H. A. Mahdiraji, E. Kazimieras Zavadskas, A. Kazemini, and A. Abbasi Kamardi, “Marketing strategies evaluation based on big data analysis: a CLUSTERING-MCDM approach,” *Economic Research-Ekonomska Istraživanja*, vol. 32, no. 1, pp. 2882–2898, 2019.
- [56] V. Pareto, *Cours d'Economie Politique*, Droz, Geneva, Switzerland, 1896.
- [57] B. Franz, *Pareto*, John Wiley & Sons, New York, NY, USA, 1936.
- [58] R. Cirillo, “Was Vilfredo Pareto really a “precursor” of fascism?” *The American Journal of Economics and Sociology*, vol. 42, no. 2, pp. 235–246, 2006.
- [59] R. Xu and D. WunschII, “Survey of clustering algorithms,” *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 645–678, 2005.
- [60] J. Wu, J. Chen, H. Xiong, and M. Xie, “External validation measures for K-means clustering: a data distribution perspective,” *Expert Systems with Applications*, vol. 36, no. 3, pp. 6050–6061, 2009.
- [61] Z. Wang, Z. Xu, S. Liu, and J. Tang, “A netting clustering analysis method under intuitionistic fuzzy environment,” *Applied Soft Computing*, vol. 11, no. 8, pp. 5558–5564, 2011.
- [62] S. Askari, “A novel and fast MIMO fuzzy inference system based on a class of fuzzy clustering algorithms with interpretability and complexity analysis,” *Expert Systems with Applications*, vol. 84, pp. 301–322, 2017.
- [63] Q. Li, M. Guindani, B. J. Reich, H. D. Bondell, and M. Vannucci, “A Bayesian mixture model for clustering and selection of feature occurrence rates under mean constraints,” *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 10, no. 6, pp. 393–409, 2017.
- [64] A. K. Paul and P. C. Shill, “New automatic fuzzy relational clustering algorithms using multi-objective NSGA-II,” *Information Sciences*, vol. 448–449, pp. 112–133, 2018.
- [65] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, San Francisco, CA, USA, 2nd edition, 2006.
- [66] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, 2nd edition, 2005.
- [67] D. S. Hochbaum and D. B. Shmoys, “A best possible heuristic for thek-center problem,” *Mathematics of Operations Research*, vol. 10, no. 2, pp. 180–184, 1985.
- [68] G. Fayyad and T. Krishnan, *The EM Algorithm and Extensions*, Wiley-Interscience, Hoboken, NJ, USA, Second edition, 2008.
- [69] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, “The WEKA data mining software,” *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.
- [70] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.
- [71] S. M. Kumar, “An optimized farthest first clustering algorithm,” in *Proceedings of the 2013 Nirma University International Conference on Engineering*, pp. 1–5, Ahmedabad, India, November 2013.
- [72] S. Dasgupta and P. M. Long, “Performance guarantees for hierarchical clustering,” *Journal of Computer and System Sciences*, vol. 70, no. 4, pp. 351–363, 2005.
- [73] Y. Peng and Y. Shi, “Editorial: multiple criteria decision making and operations research,” *Annals of Operations Research*, vol. 197, no. 1, pp. 1–4, 2012.
- [74] S. Hamdan and A. Cheaitou, “Supplier selection and order allocation with green criteria: an MCDM and multi-objective optimization approach,” *Computers & Operations Research*, vol. 81, pp. 282–304, 2017.
- [75] J. L. Yang, H. N. Chiu, G.-H. Tzeng, and R. H. Yeh, “Vendor selection by integrated fuzzy MCDM techniques with

- independent and interdependent relationships," *Information Sciences*, vol. 178, no. 21, pp. 4166–4183, 2008.
- [76] B. Wang and Y. Shi, "Error correction method in classification by using multiple-criteria and multiple-constraint levels linear programming," *International Journal of Computers, Communications & Control*, vol. 7, no. 5, pp. 976–989, 2012.
- [77] J. He, Y. Zhang, Y. Shi, and G. Huang, "Domain-driven classification based on multiple criteria and multiple constraint-level programming for intelligent credit scoring," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 6, pp. 826–838, 2010.
- [78] Y. Shi, L. Zhang, Y. Tian, and X. Li, *Intelligent Knowledge: A Study beyond Data Mining*, Springer, Berlin, Germany, 2015.
- [79] L. Zadeh, "Optimality and non-scalar-valued performance criteria," *IEEE Transactions on Automatic Control*, vol. 8, no. 1, pp. 59–60, 1963.
- [80] P. C. Fishburn, *Additive Utilities with Incomplete Product Set: Applications to Priorities and Assignments*, Operations Research Society of America (ORSA), Baltimore, MD, USA, 1967.
- [81] E. Triantaphyllou, *Multi-Criteria Decision Making: A Comparative Study*, Kluwer Academic Publishers, Dordrecht, Netherlands, 2010.
- [82] E. Triantaphyllou and K. Baig, "The impact of aggregating benefit and cost criteria in four MCDA methods," *IEEE Transactions on Engineering Management*, vol. 52, no. 2, pp. 213–226, 2005.
- [83] J. Deng, "Control problems of grey systems," *Systems and Control Letters*, vol. 1, pp. 288–294, 1982.
- [84] J. Deng, *Grey System Book*, Windsor: Science and Technology Information Services, Albany, NY, USA, 1988.
- [85] W. Wu, G. Kou, and Y. Peng, "Group decision-making using improved multi-criteria decision making methods for credit risk analysis," *Filomat*, vol. 30, no. 15, pp. 4135–4150, 2016.
- [86] W. Wu and Y. Peng, "Extension of grey relational analysis for facilitating group consensus to oil spill emergency management," *Annals of Operations Research*, vol. 238, no. 1–2, pp. 615–635, 2016.
- [87] D. Liang, A. Kobina, and W. Quan, "Grey relational analysis method for probabilistic linguistic multi-criteria group decision-making based on geometric bonferroni mean," *International Journal of Fuzzy Systems*, vol. 20, no. 7, pp. 2234–2244, 2017.
- [88] E. Onder and C. Boz, "Comparing macroeconomic performance of the union for the mediterranean countries using grey relational analysis and multi-dimensional scaling," *European Scientific Journal*, vol. 13, pp. 285–299, 2017.
- [89] J. Deng, "Introduction to grey theory system," *The Journal of Grey System*, vol. 1, no. 1, pp. 1–24, 1989.
- [90] C. L. Hwang and K. Yoon, *Multiple Attribute Decision Making*, Springer-Verlag, Berlin, Germany, 1981.
- [91] G. R. Jahanshahloo, F. H. Lotfi, and M. Izadikhah, "Extension of the TOPSIS method for decision-making problems with fuzzy data," *Applied Mathematics and Computation*, vol. 181, no. 2, pp. 1544–1551, 2006.
- [92] S. J. Chen and C. L. Hwang, *Fuzzy Multiple Attribute Decision Making: Methods and Applications*, Springer-Verlag, Berlin, Germany, 1992.
- [93] S. Opricovic and G.-H. Tzeng, "Compromise solution by MCDM methods: a comparative analysis of VIKOR and TOPSIS," *European Journal of Operational Research*, vol. 156, no. 2, pp. 445–455, 2004.
- [94] J. P. Brans and B. Mareschal, "PROMETHEE methods," in *Multiple Criteria Decision Analysis: State of the Art Surveys*, J. Figueira, V. Mousseau, and B. Roy, Eds., pp. 163–195, Springer, New York, NY, USA, 2005.
- [95] C. Hermans and J. Erickson, "Multicriteria decision analysis: overview and implications for environmental decision making," *Advances in the Economics of Environmental Resources*, vol. 7, pp. 213–228, 2007.
- [96] H. Kuang, D. M. Kilgour, and K. W. Hipel, *Grey-based PROMETHEE II with Application to Evaluation of Source Water Protection Strategies*, Information Sciences, 2014.
- [97] J. P. Brans and B. Mareschal, "How to decide with PROMETHEE," 1994, <http://www.visualdecision.com/Pdf/How%20to%20use%20promethee.pdf>.
- [98] J. P. Brans and P. Vincke, "Note-A preference ranking organisation method," *Management Science*, vol. 31, no. 6, pp. 647–656, 1985.
- [99] M. Steinbach, G. Karypis, and V. Kumar, "A comparison of document clustering techniques," in *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Boston, MA, USA, August 2000.
- [100] Y. Zhao, G. Karypis, and U. Fayyad, "Hierarchical clustering algorithms for document datasets," *Data Mining and Knowledge Discovery*, vol. 10, no. 2, pp. 141–168, 2005.
- [101] E. Rendón, I. Abundez, A. Arizmendi, and E. M. Quiroz, "Internal versus external cluster validation indexes," *International Journal of Computers and Communications*, vol. 5, no. 1, pp. 27–34, 2011.
- [102] Y. Zhao and G. Karypis, "Empirical and theoretical comparisons of selected criterion functions for document clustering," *Machine Learning*, vol. 55, no. 3, pp. 311–331, 2004.
- [103] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*, Addison-Wesley, Boston, MA, USA, 1999.
- [104] N. Slonim, N. Friedman, and N. Tishby, "Unsupervised document classification using sequential information maximization," in *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval—SIGIR'02*, Tampere, Finland, August 2002.
- [105] B. Mirkin, *Mathematical Classification and Clustering*, Kluwer Academic Press, Dordrecht, Netherlands, 1996.
- [106] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *Journal of the American Statistical Association*, vol. 66, no. 336, pp. 846–850, 1971.
- [107] S. Jaccard, "Nouvelles recherches sur la distribution florale," *Bulletin de la Société vaudoise des sciences*, vol. 44, pp. 223–270, 1908.
- [108] E. B. Fowlkes and C. L. Mallows, "A method for comparing two hierarchical clusterings," *Journal of the American Statistical Association*, vol. 78, no. 383, pp. 553–569, 1983.
- [109] L. Hubert and P. Arabie, "Comparing partitions," *Journal of Classification*, vol. 2, no. 1, pp. 193–218, 1985.
- [110] D. Badescu, A. Boc, A. Baniré Diallo, and V. Makarenkov, "Detecting genomic regions associated with a disease using variability functions and Adjusted Rand Index," *BMC Bioinformatics*, vol. 12, no. S-9, pp. 1–10, 2011.
- [111] T. L. Saaty, *The Analytic Hierarchy Process*, McGraw-Hill, New York, NY, USA, 1980.
- [112] W. Wu, G. Kou, Y. Peng, and D. Ergu, "Improved ahp-group decision making for investment strategy selection," *Technological and Economic Development of Economy*, vol. 18, no. 2, pp. 299–316, 2012.
- [113] S. Tyagi, S. Agrawal, K. Yang, and H. Ying, "An extended Fuzzy-AHP approach to rank the influences of socialization-

- externalization-combination-internalization modes on the development phase,” *Applied Soft Computing*, vol. 52, pp. 505–518, 2017.
- [114] I. Takahashi, “AHP applied to binary and ternary comparisons,” *Journal of the Operations Research Society of Japan*, vol. 33, no. 3, pp. 199–206, 2017.
 - [115] C.-S. Yu, “A GP-AHP method for solving group decision-making fuzzy AHP problems,” *Computers & Operations Research*, vol. 29, no. 14, pp. 1969–2001, 2002.
 - [116] M. Kamal and A. H. Al-Subhi, “Application of the AHP in project management,” *International Journal of Project Management*, vol. 19, no. 1, pp. 19–27, 2001.
 - [117] C. A. Bana e Costa and J. C. Vansnick, “A critical analysis of the eigenvalue method used to derive priorities in AHP,” *European Journal of Operational Research*, vol. 187, no. 3, pp. 1422–1428, 2008.
 - [118] T. Ertay, D. Ruan, and U. Tuzkaya, “Integrating data envelopment analysis and analytic hierarchy for the facility layout design in manufacturing systems,” *Information Sciences*, vol. 176, no. 3, pp. 237–262, 2006.
 - [119] M. Dağdeviren, S. Yavuz, and N. Kılınç, “Weapon selection using the AHP and TOPSIS methods under fuzzy environment,” *Expert Systems with Applications*, vol. 36, no. 4, pp. 8143–8151, 2009.
 - [120] M. P. Amiri, “Project selection for oil-fields development by using the AHP and fuzzy TOPSIS methods,” *Expert Systems with Applications*, vol. 37, no. 9, pp. 6218–6224, 2010.
 - [121] X. Yu, S. Guo, J. Guo, and X. Huang, “Rank B2C e-commerce websites in e-alliance based on AHP and fuzzy TOPSIS,” *Expert Systems with Applications*, vol. 38, no. 4, pp. 3550–3557, 2011.
 - [122] Y. Peng, G. Kou, G. Wang, W. Wu, and Y. Shi, “Ensemble of software defect predictors: an AHP-based evaluation method,” *International Journal of Information Technology & Decision Making*, vol. 10, no. 1, pp. 187–206, 2011.
 - [123] P. Domingos, “Toward knowledge-rich data mining,” *Data Mining and Knowledge Discovery*, vol. 15, no. 1, pp. 21–28, 2007.
 - [124] G. Kou and W. Wu, “Multi-criteria decision analysis for emergency medical service assessment,” *Annals of Operations Research*, vol. 223, no. 1, pp. 239–254, 2014.
 - [125] A. Frank and A. Asuncion, *UCI Machine Learning Repository*, University of California, School of Information and Computer Science, Irvine, CA, USA, 2010, <http://archive.ics.uci.edu/ml>.