WILEY | Hindawi

Research Article

Human Resource Allocation Based on Fuzzy Data Mining Algorithm

You Wu [ID],[1] Zheng Wang,[1] and Shengqi Wang[2]

[1]*School of Public Administration, Xinhua College of Sun Yat-Sen University, Dongguan, China*
[2]*School of Business, GuiLin University of Electronic Technology, Guilin, China*

Correspondence should be addressed to You Wu; wuyou880@xhsysu.edu.cn

Data mining is currently a frontier research topic in the field of information and database technology. It is recognized as one of the most promising key technologies. Data mining involves multiple technologies, such as mathematical statistics, fuzzy theory, neural networks, and artificial intelligence, with relatively high technical content. The realization is also difficult. In this article, we have studied the basic concepts, processes, and algorithms of association rule mining technology. Aiming at large-scale database applications, in order to improve the efficiency of data mining, we proposed an incremental association rule mining algorithm based on clustering, that is, using fast clustering. First, the feasibility of realizing performance appraisal data mining is studied; then, the business process needed to realize the information system is analyzed, the business process-related links and the corresponding data input interface are designed, and then the data process to realize the data processing is designed, including data foundation and database model. Aiming at the high efficiency of large-scale database mining, database development tools are used to implement the specific system settings and program design of this algorithm. Incorporated into the human resource management system of colleges and universities, they carried out successful association broadcasting, realized visualization, and finally discovered valuable information.

1. Introduction

Data mining is the process of discovering and extracting hidden information or knowledge from large databases or data warehouses [1]. In a large enterprise or public institution, it is very important for the selection of talents and the formulation of organizational talent development strategies to grasp the composition and types of talents in the organization and determine which type of talent a certain employee belongs to [2]. Data mining technology can extract human resource information from many databases about the work situation of personnel in the organization [3]. Its purpose is to help analysts look for potential correlations between data and find neglected elements, and this information is very useful for predicting trends and in decision-making behavior. It is used to discover meaningful associations and related links between item sets in a large amount of data [4]. Human resource management faces massive amounts of data and urgently needs a technology to discover valuable knowledge [5]. The current database system can efficiently realize data entry, query, statistics, and other functions, but it cannot find the relationships and rules existing in the data [6].

Now, data mining has been applied to many fields as their problem-solving tools, such as finance, medicine, sales, stocks, telecommunications, manufacturing, health care, customer relations, and other fields. Diabat [7] puts forward the viewpoint of applying classification and prediction methods in data mining to solve the talent problem. Krömer et al. [8] proposed an improved K-means clustering algorithm and established a personnel management system clustering analysis model based on the improved K-means algorithm. Lin [9] introduced data mining into the corporate talent selection and retention system, using decision tree-based data mining technology, combined with empirical research, to find out the laws hidden in talent selection and

human resource management. Xiao and Feng [10] studied the role of data warehouse and data mining in human resource management and the development direction of the two in human resource management. Zhang [11] proposed to introduce data mining technology in performance evaluation, establish a decision tree classification model to screen talents, and improve the decision analysis ability of HRM. Devi and Amalraj [12] studied the association rules based on the Apriori algorithm, used the Apriori algorithm to discover the factors affecting the work performance of employees, and expounded the application of the association rules in the enterprise personnel management system. Chen et al. [13] combined the principles and methods of data mining and artificial intelligence, designed a data mining method combining cross decision tree and multi-cross-decision tree, and constructed a structural model of human resource intelligent fault diagnosis system. Delgado et al. [14] proposed a comprehensive evaluation method based on data mining and information theory, introduced decision tree method in human resource evaluation, and developed a model based on decision tree. Strohmeier and Piazza [15] proposed an improved density-based distributed clustering algorithm, applied the cluster analysis model based on the improved density clustering algorithm of performance evaluation, and developed and implemented a performance evaluation system. Jantan et al. studied the application of data mining in talent recruitment selection, using decision tree method and neural network method to deal with network recruitment [16]. Kale and Patil improved and expanded the genetic operator of the standard genetic algorithm, optimized the fuzzy clustering method based on this improved genetic algorithm, and applied it to the human resource management system [17]. It was pointed out that it can be used for training and recruitment, talent selection, formulating a competitive salary and welfare system, effective performance management system, and other aspects that use data mining and also pointed out the direction for further research on data mining [18]. They focused on the decision tree algorithm, analyzed and designed a human resource management system for schools, and designed a data mining system [19]. Data mining technology is used to quantify decision-making attributes and perform human resource analysis based on ID3 algorithm to provide reliable basic data [20]. In the process of evaluating the effectiveness of human resource management, a classification-based approach is introduced [21]. The association rule mining technology is applied to the electronic file management information system of university officials [22], and artificial neural network technology is applied to the evaluation process and integration with the system. It studies the necessity of establishing a human resource management system based on data mining technology [23–26]. It is based on this point that the solution of human resource management system began to pay attention to the application of data mining [27, 28], pointing out that coal companies do not pay attention to the prediction and planning of talents, the brain drain is serious, the talent market development is not sound enough, and the talents are ignored. Targeted measures for the shaping of coal corporate culture and human resource

management of coal companies were provided [29, 30]. The difficulties and current situation of enterprise human resource management are analyzed, and solutions and specific countermeasures are put forward. Countermeasures and suggestions to prevent the loss of human resources were proposed [31, 32]. Through the summary of salary incentive theory, several current salary incentive systems are briefly introduced, and the problems existing in the salary management of coal enterprises are analyzed [33, 34].

This article uses fast clustering method to realize data partition, uses improved FP-growth algorithm to realize association rule mining, uses incremental FP-growth mining algorithm to realize incremental data mining, and uses clustering method to cluster large-scale data. After partitioning, the local frequent itemsets are mined separately for each local database, and then the local frequent itemsets are combined to generate the global frequent itemsets, and the association rules are mined on the merged global frequent itemsets. At the same time, it is necessary to be able to mine incremental data. Later, this association rule mining technology was applied to the human resource management system, and the visual software development was carried out using Visual c++ 6.0 development tools and Oracle9i database technology to give concrete realization.

2. Construction of Human Resource Allocation Model Based on Fuzzy Data Mining Algorithm

2.1. Human Resource Deployment Level. Data mining is to extract from a large amount of incomplete, noisy, fuzzy, random data, the process of not knowing in advance but potentially useful information and knowledge. Data mining is sometimes called knowledge mining, knowledge extraction, etc. In the data mining process, data mining algorithms are the most critical. Data mining technology can extract human resource information from numerous databases about the work situation of personnel in the organization. Using fuzzy mining algorithms and data extracted from the data warehouse, we can discover the types of cadres currently in the organization, and we can also determine that an employee belongs to these types. Since in actual data, the collected data are often not the numbers in the closed interval of [0, 1], so these raw data should be standardized and averaged first in Figure 1. For example, there are n samples in the sample set, and their average value is calculated as following formula:

$$u(i) = \frac{\sum_{i=1}^{n} u_i - 1}{\sum_{i=1}^{n} |u_i^2|}. \tag{1}$$

Then, calculate the standard deviation S-k of these raw data. Then, calculate the standardized value u of each data within the closed interval of [0, 1], and the following extreme value standardization formula must be used:

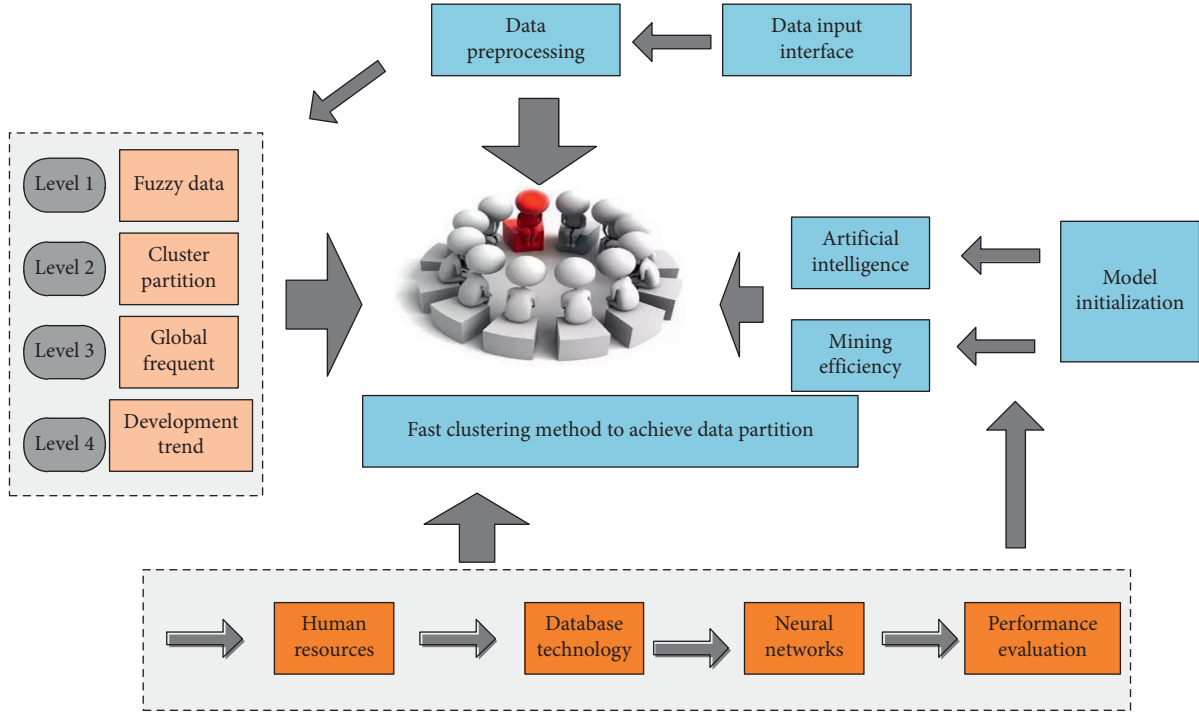$$u(k) = \frac{\sum_{i=1}^{n} u_i - k}{\sum_{i=1}^{n} |u_i^2|}. \tag{2}$$

FIGURE 1: Hierarchical structure of human resource deployment.

A hash-based algorithm for efficiently generating frequency sets was proposed by Park et al. Through experiments, it can be found that the main calculation for finding frequency sets is to generate frequent 2 item sets. Use this property to introduce hashing techniques to improve the method of generating frequent 2 item sets. The establishment of the modulus similarity relationship can be expressed as a similarity matrix, generally the form is as follows:

$$s(k) = \lim_{n \longrightarrow \infty} \sqrt{\frac{1}{n \sum (u(i,k) - u(k))}}. \qquad (3)$$

Association rule is a common technique in association analysis (the other is sequence mode) to find the correlation of different items that appear in the same event. D is a set of transactions, where each transaction T is a set of items and $T \in L$. Each transaction T has a unique identification TID. If the item set is X and $x \in T$, we say that transaction T contains X. An association rule is such a form of relationship: $x \Rightarrow Y$, X and Y are, respectively, called the premise and conclusion of the association rule $X \Rightarrow Y$. The other two concepts related to association rules are support and confidence. According to the definition, for an association rule $x \Rightarrow Y$, the transaction set D contains the number of transactions of item set x. The support number (frequency) of item set x is denoted as S. The formula determines whether there is a certain node s in T and a certain node k in u that make the association rule y exists. If it exists, switches nodes and adjusts node S to be after node p.

$$y(x) = \frac{|u(i,k) - u(k)||s(k)|}{|u(i,k) - u(k)| - |s(k)|}. \qquad (4)$$

The support degree of item set x is recorded as support (x), where f is the number of transactions in transaction set D, if support (x) is not. If it is less than the minimum support specified by the user, then x is called frequent itemsets, referred to as frequency sets (or large itemsets); otherwise, x is called infrequent itemsets, referred to as nonfrequency sets (or small itemsets). The item set x has a support degree of sup. If there is sup% transaction support item set x in D. The support degree of the association rule $X \Rightarrow Y$ is recorded as support, that is, the transaction in D contains $XU\ Y$ (both x and Y) percentage.

For each mode obtained, the average index is calculated, where s represents the total number of patterns, k represents the number of records in the warehouse that the pattern (i.e., the i pattern) is derived from, and p represents the total number of records from which the pattern is derived.

$$r(i,j) = \frac{\sum \min(u(i,k), u(j,k))}{\sum \max(u(i,k), u(j,k)) + \sum \min(u(i,k), u(j,k))}. \qquad (5)$$

2.2. Fuzzy Data Mining Algorithm.
In particular, with the popularization and development of the Internet and the emergence of the knowledge economy, various organizations put the organization, management, learning, and innovation of knowledge at the highest level. They also discover the connections and patterns among them, so as to objectively reflect the internal talent composition of the organization. The important position has aroused people's enthusiasm for data mining and knowledge discovery in one of the important technical fields. In the learning and research

process of data mining and knowledge discovery, data, information, and knowledge are three concepts that directly contact each other. In order to achieve a reasonable sample classification, its specific attributes should be quantified. The quantified attributes are called sample indicators. There are m indicators, which can be described by m-dimensional vectors. There are differences and connections between these three. In actual data mining, the transformation from data to knowledge is also such a transformation process, but it is realized by various algorithms and modes. The maximum tree method is adopted, that is, a special graph is constructed with all classified objects as vertices. The specific method is to first draw a certain i in the vertex set and then press i to connect the edges in order from large to small and require no loops until all vertices are connected, so that a maximum tree is obtained.

$$(x, f(x)) = \frac{xf(x) + (1+x)f(x)}{xf(x) - (1+x)f(x)}. \quad (6)$$

To be precise, it is a "weighted" tree. Each side can be assigned a certain weight, namely r. However, due to different specific connection methods, this largest tree cannot be unique. Then, take the cut set for the largest number, that is, remove those edges with weight, and enter [0, 1]. In this way, a tree is cut into several subtrees that are not connected to each other.

$$g(i, j) = \frac{|f(i) + f(j)|}{|f(i) - f(j)|}. \quad (7)$$

Although the largest tree is not unique, after taking the cut set, the resulting subtrees are the same. These subtrees are the patterns found by induction in the data warehouse. Starting with a frequent pattern of length 1 (initial suffix pattern), construct its conditional pattern library (a "predatabase" consisting of the prefix path set that appears together with the suffix pattern in the FP-tree). Then, construct its conditional pattern tree and recursively dig on the tree. Pattern growth is realized by connecting the suffix pattern and the frequent pattern generated by the conditional pattern tree. From the function, we can see if a is a frequency set in database D, B is a conditional pattern library of a, and b is an item set in B. However, since the base of the power function is the objective function, it is not necessarily a positive number; it is also difficult to determine the power exponent, and it takes a long time to use the properties of the function or other information to determine the sufficient value that is beneficial to the current problem.

$$z(i, j) = \lim_{n \to \infty} \sqrt{\sum_{i,j=1}^{n} \frac{g(i, j) - g(i)}{g(i, j) + g(i)}}. \quad (8)$$

Agrawal et al. proposed an important method for mining association rules between item sets in a customer transaction database. Its core is a recursive algorithm based on the idea of two-stage frequency sets. The association rules are classified as single-dimensional, single-layer, and Boolean association rules. The basic idea of this algorithm is to first find out all candidate item sets and then compare these candidate item sets with the predefined minimum support, select the candidate item set greater than or equal to the minimum support as frequent item sets, and then generate strong items from the frequency set of association rules; these rules must meet the minimum support and minimum confidence. In this way, a large number of candidate sets may be generated among them, and the database may need to be scanned repeatedly. Support and confidence are two important concepts to describe association rules. The former is used to measure the statistical importance of the association rules in the entire data set, and the latter is used to measure the credibility of the association rules. It uses a divide-and-conquer strategy. The partitioning method first divides the data point set into k partitions and then starts from these k initial partitions and optimizes a certain criterion through repeated control strategies to achieve the final result.

In order to achieve global optimization, the partition-based method requires exhaustion of all partitions. Because searching all possible subspaces is computationally impossible, a certain iterative optimization shoulder method is often used. This means repeatedly relocating the category center of each category among the k categories and redistributing the objects in each category. The partition clustering algorithm has a fast convergence speed from Figure 2. The disadvantage is that it tends to identify clusters with similar convex distribution sizes and similar densities and cannot find clusters with more complex distribution shapes. It requires that the number of categories k can be reasonably estimated and the initial center Selection and noise will have a great impact on the clustering results. Since in actual data, the collected data are often not a number in the closed interval of [0, 1], so these raw data should be standardized, and the average value should be found first. When the amount of original data is large, the division method can also be combined so that a FP-tree can be placed in the main memory. The FP-growth method converts the problem of finding long frequent patterns into finding some short patterns recursively and then connects the suffixes. It uses the least frequent items as a suffix, providing good selectivity. This method greatly reduces the search overhead. Generally speaking, only association rules with high support and confidence can be interesting and useful association rules for users.

2.3. Model Variable Optimization Processing. (1) Preprocessing data: collect and purify information from data sources and store them, usually in a data warehouse. (2) Model search: use data mining tools to find models in the data. This search process can be automatically executed by the system. The original facts can be searched from the bottom-up to find a certain connection between them, and user interaction can also be added. The analysts take the initiative to ask questions and search from top to bottom to verify the correctness of the hypothesis. Many tools may be used to search for a problem. For example, neural networks, rule-based systems, basic case-based reasoning,
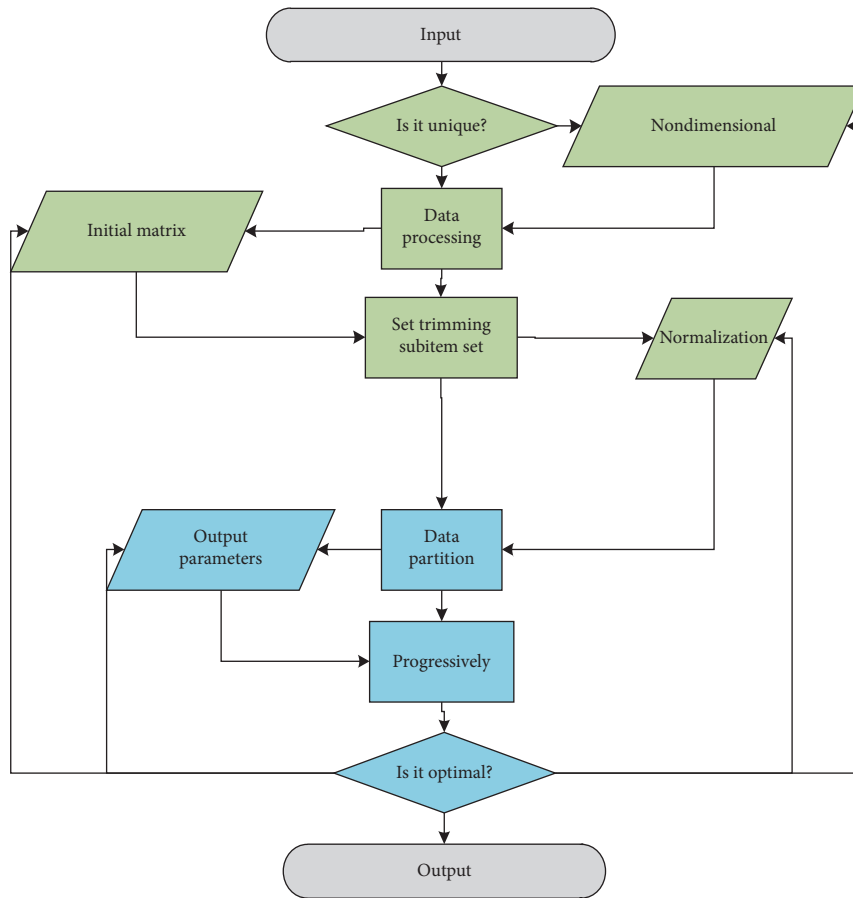
FIGURE 2: Fuzzy data mining algorithm flow.

machine learning, statistical methods, etc. (3) Evaluate the output results: the search process of data mining generally needs to be repeated many times because after the analysts evaluate the output results, they may form some new problems or require more refined queries on a certain aspect. (4) Generate the final result report. (5) Interpret the results report, interpret the results, and take corresponding measures based on the results. This is a manual process. By analogy, scan the entire database. In order to facilitate the traversal of the tree, a frequent item header table is established, and the node link of the header table entry points to the node with the same item name in the tree. Nodes with the same item name are linked together by node link.

Data preprocessing is responsible for making necessary preparations for the data source to be mined. Data preprocessing may generally include eliminating noise, deriving and calculating missing data, eliminating duplicate records, and completing data type conversion (such as converting continuous value data into discrete data for symbol induction or converting discrete data into continuous value type to facilitate neural networks, etc.). The data preprocessing in this article is to preprocess the personnel information data: transform the personnel information data according to the mining requirements in Figure 3. That is, the data represented by letters in the personnel information

data are transformed into data represented by numbers. The advantage of this preprocessing is that it simplifies program processing and also ensures the independence of the program and the source data. The data preprocessing in this article is to create a correspondence table to convert the correspondence of letters in personnel information data into numbers (corresponding codes of letters), which is unique and fixed. This kind of corresponding code is processed in the program. After the program is processed and the association rules are mined, the corresponding code will be converted into the letter in the corresponding personnel information data through the correspondence table and displayed. What the user sees is only the letters in the input personnel information data and the association rules shown in the alphabet in the personnel information data displayed by the mining results. The specific method is to first draw a certain i in the vertex set and then press r; i connect the edges in order from large to small and require no loops until all vertices are connected, so that a tree is obtained the largest tree. This algorithm restricts the clustering of high-dimensional data to high-dimensional subspaces, instead of adding new dimensions that are combined by certain dimensions in the original space. In the actual clustering, a density-based approach is adopted. In order to estimate the density of data points, the bottom-up method is used to divide each dimension of the space into
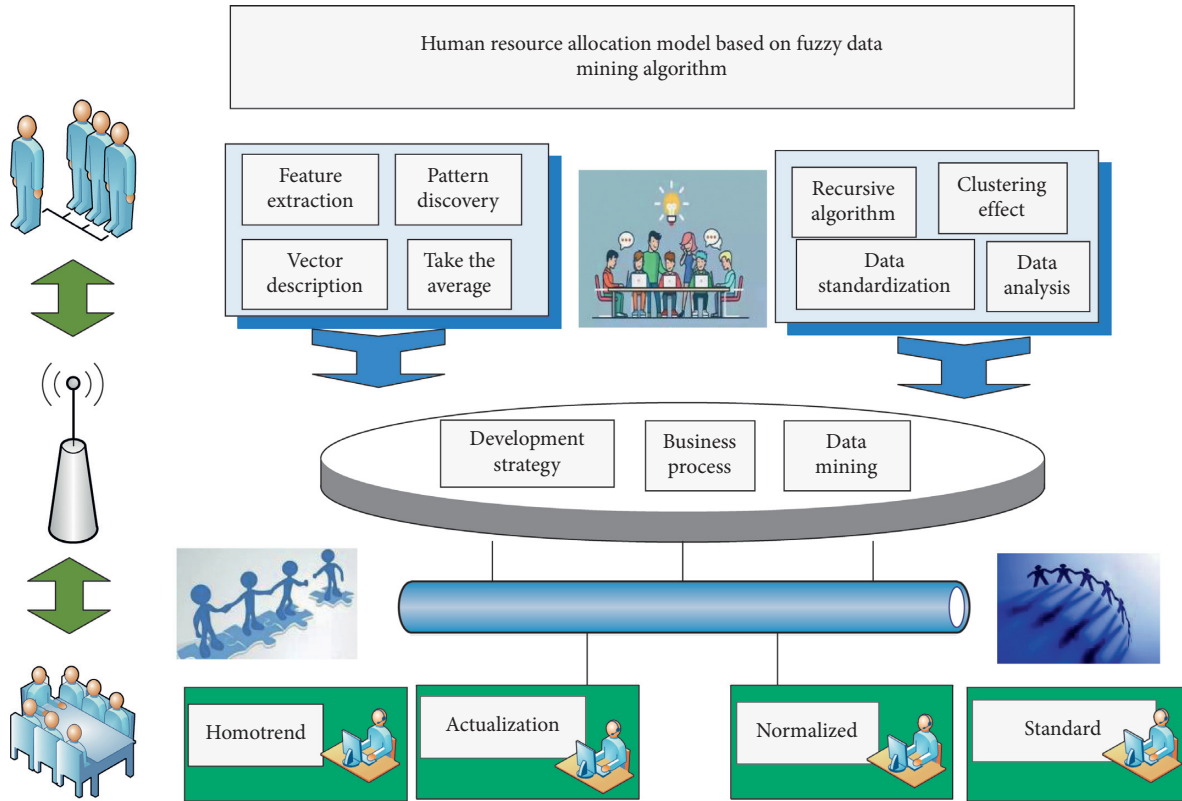
FIGURE 3: Human resource deployment model framework based on fuzzy data mining algorithm.

equal-length intervals. Because the volume of each grid cell is the same, data can be obtained from a certain grid cell. Then, use these densities to identify suitable spaces. The data points are separated by the density function, and the connected high-density areas in the space are merged. For the sake of simplicity, the clusters are limited to supercuboids parallel to the coordinate axes, and the expressions are used to express these clusters.

Each cluster can be represented by a set of overlapping rectangular parallelepipeds. Two input parameters are used to partition the subspace and identify dense units. Among them, a is the number of intervals, which determines the subspace division method, and another input parameter is the density threshold. If the frequent items $\langle f, O, a, b, m \rangle$ of transaction 200 are inserted into the FP-tree in the order of $\langle f, e, a, m, b \rangle$, as shown in Figure 4, we can see that the node m in the original FP-tree (Figure 4) is compressible; the improved FP-tree method is obviously better than the original FP-tree method. Although the largest tree is not unique, after taking the cut set, the resulting subtrees are the same. These subtrees are the patterns discovered by induction in the data warehouse. Therefore, in order to solve the optimization and compression problem of the same frequency items mentioned above, this article adopts an improved FP-tree method. This method is to improve the insert-tree function in FP-tree construction to the improved insert-tree function, where p is the frequent item table. The first element in P is the list of remaining elements, and T is the transaction that needs to be inserted currently.
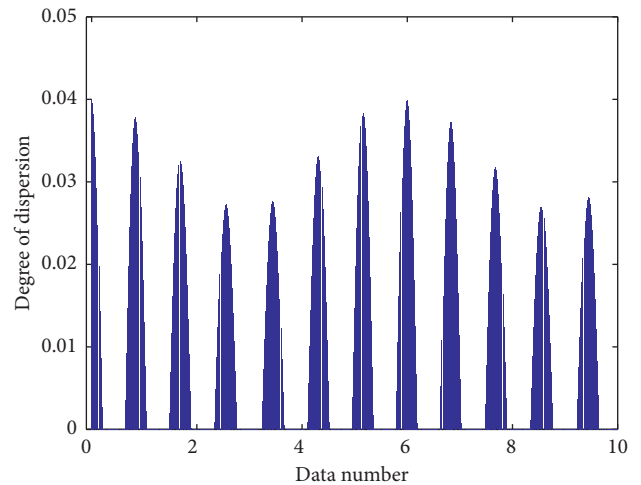

FIGURE 4: Columnar distribution of the degree of dispersion of fuzzy data.

3. Application and Analysis of Human Resource Allocation Model Based on Fuzzy Data Mining Algorithm

3.1. Data Mining Preprocessing. (1) Choosing the index elements to be considered and alternative objects. Select the four index elements of "learning level," "innovative ability," "independent working ability," and "work efficiency." A total of 10 objects are divided, that is, domain $U = \{$cadre 1, cadre

2, cadre 3, cadre 4, cadre 5, cadre 6, cadre 7, cadre 8, cadre 9, and cadre 10}. The average vector is $u'k = \{3.7, 2.2, 2.4, 1.2\}$ $(k = 1, 2, 3, 4)$. The standard deviation vector is $S\text{-}k = \{1.676, 1.187, 1.428, 0.748\}$ $(k = l, 2, 3, 4)$. The standardized matrix l is calculated from the formula (4). Set $i = 0.81$, which is divided into 9 categories: {cadres l, cadres 7}, and the rest are in one category. Take $i = 0.65$. It was divided into 7 categories: {cadre 1, cadre 7}, {cadre 2, cadre 9}, {cadre 3, cadre 8}, and the rest belong to one category. Take $i = 0.48$, which is divided into 3 categories: {cadre 2, cadre 9} (a cadre with a slightly weaker ability), {cadre 4} (a cadre with a weaker ability), and the rest are in the first category (a cadre with strong cadres). Take $i = 0.31$. The sample x to be predicted is the n fuzzy subsets of the sample in the universe u. Compare them with the classified patterns in the data warehouse to find the closeness between them. All are divided into one category. (4) Taking entry $= 0.48$ as an example, calculate the average index of each mode:

In order to simplify the process, the relevant human resource data should be transformed into numerical data. First, discretize the continuous data in the relevant personnel information, such as discretizing the age attribute into segments: below 30, 30–40, 40–50, 50–60 four situations (characteristics); the papers are published first divided into three situations (attributes) and then discretized into SCI: 0 articles, 1–3 articles, 3 or more situations (characteristics): core: 0 articles, divination 3 articles, and 3 or more articles (nature); general: 0 articles, 3 articles, 3 articles or more (nature) and so on. Second, the discretized properties of each attribute are regarded as an item, and the corresponding number is used to represent an item. For example, the four properties under the age attribute can be represented by the four numbers 1, 2, 3, and 4. Thus, a correspondence table is established, that is, a one-to-one correspondence code between various attributes and numbers in the personnel information, and the code is unique and fixed. Finally, the corresponding code is processed in the program. After the program is processed and the association rules are mined, the corresponding code is converted into the corresponding attribute in the personnel information through the correspondence table and displayed. What the user sees is only the input personnel information and the association rules represented by the personnel attributes displayed in the mining results. The association rules we want to mine are strong association rules (association rules that meet the minimum support and minimum confidence). At the same time, because the rules are generated by frequent itemsets, each rule automatically meets the minimum support and so the frequent items. The set generation association rules only need to meet the confidence level.

The correspondence table and its corresponding code (number) are transparent to the user. As shown in Figure 5. It can be seen that the establishment of a correspondence table between the source data and the processed data is equivalent to the establishment of a mapping relationship between them. For example, the assessment of a new cadre data $x = $ (good 5, generally 1.5, strong 3, not too high 1), compare with the above classification model, and find its closeness to each category. For A class, inner product $= 0.58$,

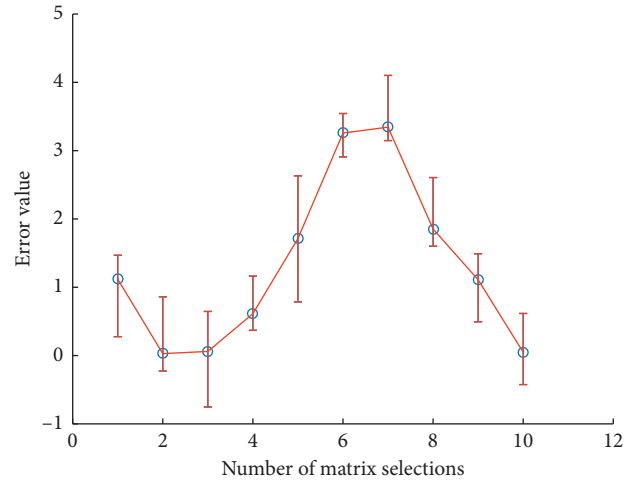FIGURE 5: Error bar graph of human resource matrix.

outer product $= 0.69$, and closeness $= 0.445$. For class B, inner product $= 0.75$, outer product $= 0.50$, and closeness $= 0.625$. For class C, inner product $= 0.64$, outer product $= 0.48$, and close to degree 0.58. It can be obtained that the degree of closeness to class B is the largest, and the cadre belongs to the weaker cadre.

3.2. Cluster Analysis of Human Resources. This section is a detailed expansion of the model processing level, involving the implementation of the algorithm for the evaluation of work performance indicators, the implementation of the Apriori algorithm for the evaluation of ability and quality indicators, the realization of the fuzzy matrix of the indicator evaluation of professional ethics and work attitude, and the k of the work evaluation index, which means clustering algorithm realization and algorithm realization of comprehensive performance appraisal results of employees. When mining a large database, it is unrealistic to construct a memory-based FP-tree. Therefore, we need to divide the transaction data in the large database to form multiple partial (sub) databases, and then on each partial (sub) database. Because this article involves more matrix operations, in order to facilitate analysis and research, the Matlab coding rules are used to program the program.

Combined with the characteristics of work performance indicators that can be directly quantified, and the data require real-time and accurate data. The first-level indicator weight refers to the weight of each first-level indicator in the corresponding first-level indicator system, the second-level indicator weight refers to the weight of the second-level indicator in the corresponding first-level indicator, and the total weight refers to the weight of the secondary indicator in the entire secondary indicator system and is equal to the product of the weight of the secondary indicator and the weight of the corresponding primary indicator. For quantitative indicators, the assessment results are specific values.

According to the model construction, the evaluation and analysis of this type of indicators are programmed in the form of a sequential structure. In the process of program

compilation, the maximum-minimum method is used to normalize the data. Therefore, the maximum value of each group of indicators is 1, so the optimal value is of course chosen as 1, and 0 is chosen as the worst value for the same reason in Figure 6. In generating association rules, frequent itemsets need to be generated for all nonempty subsets; this is the process of generating all combinations. If the frequency set is large, the number of combinations will increase exponentially, and the number will be very large. Set the index quantity of an employee's event V1 under the best performance to be (1, 1, 1, 1, 1, 1), the index quantity of the event V2 under a good condition is (0.9, 0.9, 0.9, 0.9, 0.9, 0.9), the index quantity of event V3 in general is (0.7, 0.7, 0.7, 0.7, 0.7, 0.9. 7), and the index quantity of the event V4 under qualified is (0.6, 0.6, 0.6, 0.6, 0.6, 0.6), so select employee 1 as the event V0 to be judged. The quantity is (1, 1, 0.95, 0.85, 0.75, 0.95). From the results of the cluster analysis, it can be seen that category 1 academic level and professional skills are considered good, but work ability and innovation ability scores are low, which should be regarded as general technical personnel and need further incentives; category 2 scores in all aspects are very average, it should be regarded as key and employees, pay attention to retaining these employees; category 3 is very outstanding in work ability and innovation ability, but the academic level and professional skills are not high; they should be ordinary mine workers and focus on further development and training to improve them; category 4 work ability is very good, innovation ability is lacking, academic level and professional skills are average, we must focus on strengthening innovation ability.

From Figure 7, V0 and V2 are grouped into one category, and the rest are in the same category. Therefore, the performance result of "employee 1" is good. In the same way, the performance scores of "employee 2" to "employee 10" are good, excellent, excellent, excellent, excellent, excellent, excellent, good, and good. From the results in the example, it can be seen that there is a certain degree of distinction in the evaluation of employees' work attitudes, and they are mainly concentrated between excellent and good. This is closer to the actual situation and expected goals of employee performance evaluation. The only thing with strong subjectivity is when choosing the index quantities of Vl, V2, V3, V4, such as mandatory calibration of V2 to (0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9); to avoid being too subjective, it can be adjusted through statistical calculations. In order to facilitate analysis and processing, "normal work," "overtime," and "plus points" are combined into "positively related indicators," and "sick leave," "incidental leave," and "absenteeism" are combined into "negative related indicators".

3.3. Example Results and Analysis.
Each submodule of the above system can realize the visual analysis and display of the original data and processing results and provide users with intuitive visual perception. There are two methods for index calibration of fuzzy similarity matrix: the angle cosine method and the minimum-maximum method. However, in practical applications, because the amount of information can reach tens of thousands, the number of search rules will
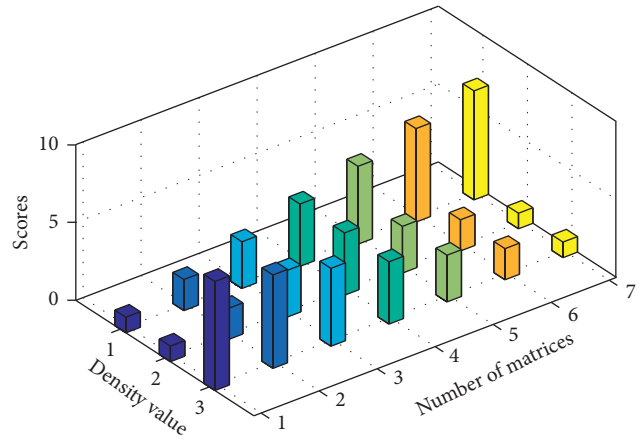

FIGURE 6: Three-dimensional histogram of human resource deployment.
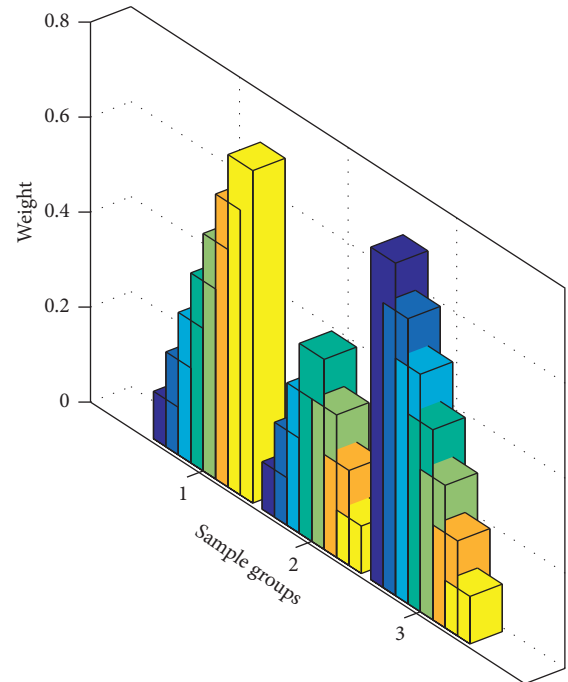

FIGURE 7: Weight distribution of fuzzy data mining algorithm.

be very large, so the minimum support and minimum confidence settings should not be too small. From the analysis of the above evaluation results, it can be seen that the attendance status of the 30 employees of branch E is divided into 4 categories, namely, 13 excellent, 14 good, 2 qualified, and 1 unqualified. In view of this, we use the clustering method to divide the database horizontally, so that the data in each subset are as similar as possible, so that the local frequent itemsets mined are relatively concentrated in each subset, and finally, the global frequent itemsets are reduced when they are merged. At the same time, the workload of local frequent itemsets mining is also reduced a lot. The classification of the results in this way generally meets the expectations of the evaluation, that is, the classification is reasonable and can give the performance

evaluation results of the personnel. Employees with excellent performance include: employees 1, 5, 6, 9, 11, 12, 14, 15, 17, 20, 21, 26, 29; employees with good performance include: employees 2, 3, 4, 7, 8, 10, 13, 16, 18, 19, 23, 24, 25, 27, 28.

According to the characteristics of coal enterprise employees in Figure 8, the four variables of "work ability," "innovation ability," "education level," and "professional skills" are selected. The work ability scoring standard is as follows: very strong: 95 points, strong: 85 points, strong: 75 points, 65 points for not too strong, and 50 points for not too strong. Innovative scoring standards is as follows: 95 points are very high, 85 points are high, 75 points are high, 65 points are good, and 50 points are average. The grading standard of academic level is as follows: very good: 90 points, good: 80 points, good: 70 people, and not very good: 55 points. Professional skills scoring criteria are as follows: very strong: 90 points, strong: 80 points, strong: 70 points, not very strong: 70 points, general: 60 points, and not strong: 50 points. After the overall analysis and testing of group A, the minimum support is set to 0.8. When the minimum confidence is set to 0.9, the search association rules are 97, which just meets all the needs of employees for predicting and judging. If an employee's performance appraisal results for the first-level indicators are excellent, good, qualified, and good; the quantification matrix is [95, 85, 75, 85], then his final performance score is 88.19, which is between excellent and good, and closer to good, so we finally judge his performance result to be good.

After analysis, it is found that among the 2% of employees with large differences in performance appraisal results in Figure 9, most of them are subjective deviations caused by unreasonable weight settings and large differences in manual scoring. The human resource data mining system based on association rules can be divided into five relatively independent modules to achieve, namely, data preprocessing module, mining frequent collection module, association rule generation module, display process information module, and incremental mining module. Therefore, the data mining system using this performance appraisal can reflect the accuracy of the performance appraisal results to a certain extent, and timely correct some unreasonable phenomena caused by subjective factors.

Mining is performed separately to generate local frequent itemsets, and then the local frequent itemsets are merged to form a global frequent itemsets, and finally the association rules are mined from the global frequent itemsets. The expressions and solutions are as follows: The situation where the same previous item has different subsequent items: if the technical job is called intermediate, two rules corresponding to excellent performance and good performance are generated. Treatment method: select those with relatively large support and confidence; if support and confidence are still equal, select those with poor performance, which can be seen in Figure 10. The same latter item corresponds to multiple subitem sets: for example, the performance score is good, which is derived from the two previous items of "technical job title is intermediate and the skill level is senior worker" and "technical job title is intermediate"; another example is the performance score.
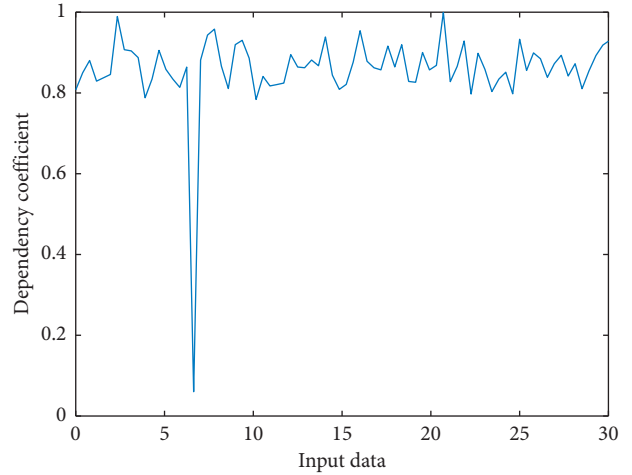

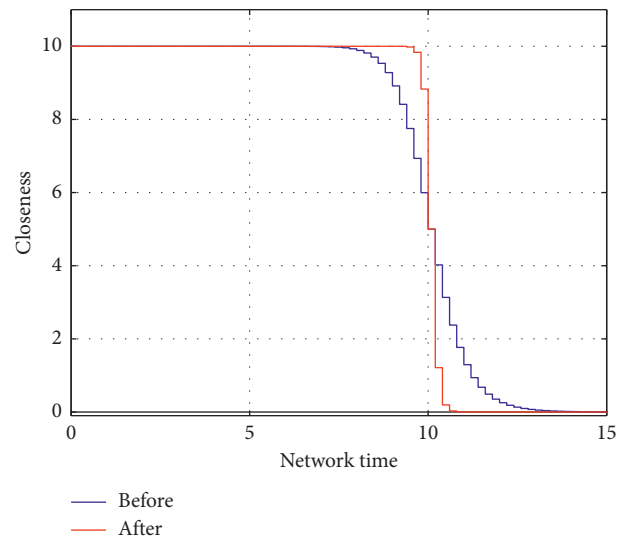FIGURE 8: Dependence of human resource data with input value.

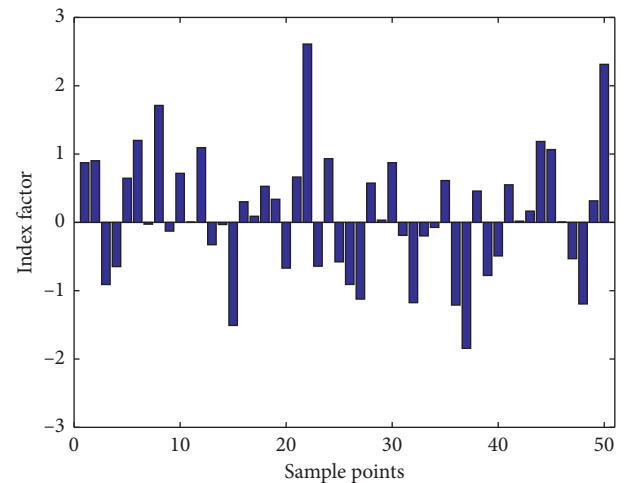FIGURE 9: Human resource deployment closeness curve before and after model simulation.

FIGURE 10: The distribution of exponential factors of human resources samples.

4. Conclusion

Data mining can discover the talent patterns and laws within the organization. This article designs a data mining method that can perform cluster analysis on the existing talents in the organization to discover the types of talents in the organization, which can make talent decisions and talents for decision makers. Aiming at the problem of excessive subjectivity in the performance evaluation process, this article uses data mining tools such as cluster analysis, association rule analysis, and fuzzy comprehensive evaluation analysis to model and parameterize the subjective judgment as much as possible and use multiangle and multilevel data, count, and summarize to get more objective and accurate results. 1. Aiming at the feature of easy collection of work performance indicators, the method is used to achieve ideal point ranking, that is, to find benchmark employees and to compare and rank the employees to be evaluated and the benchmark employees, so as to give reasonable performance evaluation results. 2. For the ability and quality indicators, Apriori association rule analysis is used to first obtain the probability of causal connection and then to transform the ability and quality into performance to make a prediction, thereby giving a reasonable result of the evaluation. 3. For work attitude indicators, because they are products of strong subjectivity in the initial stage of collecting data, use the method of 360 degrees plus fuzzy clustering evaluation, and through data analysis and processing, remove those subjective variables with excessive deviations and make them as smooth as possible. The influence of subjective factors makes the final result more real and objective. 4. For work attendance indicators, the k-means clustering algorithm is adopted to enable it to be naturally classified into four categories to meet the needs of the assessment to distinguish "excellent, good, qualified, and unqualified" classification. 5. Aiming at the hierarchical and hierarchical characteristics of the performance appraisal index system of group A. A data mining model based on analytic hierarchy process was proposed to calculate the weights of indicators at different levels and finally obtain comprehensive results, which facilitate the feedback and announcement of the evaluation results for using data mining to carry out performance evaluation has obvious advantages.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no known conflicts of interest or personal relationships that could have appeared to influence the work reported in this study.

References

[1] H. Jing, "Application of fuzzy data mining algorithm in performance evaluation of human resource," *Computer Science-Technology and Applications*, vol. 2, no. 9, pp. 343–346, 2020.

[2] M. Huang, L. Tsou, and S. Lee, "Integrating fuzzy data mining and fuzzy artificial neural networks for discovering implicit knowledge," *Knowledge-Based Systems*, vol. 19, no. 6, pp. 396–403, 2020.

[3] H. Wu, "Early warning management of enterprise human resource crisis based on fuzzy data mining algorithm of computer," *Machine Learning and Big Data Analytics for IoT Security and Privacy*, vol. 12, no. 1, pp. 640–645, 2019.

[4] W. Tai and C. Hsu, "A realistic personnel selection tool based on fuzzy data mining method," *Information Sciences*, vol. 2, no. 6, pp. 190–193, 2018.

[5] S. Bhattacharya and V. Bhatnagar, "Fuzzy data mining: a literature survey and classification framework," *International Journal of Networking and Virtual Organisations*, vol. 11, no. 3, pp. 382–408, 2018.

[6] E. Petry and L. Zhao, "Data mining by attribute generalization with fuzzy hierarchies in fuzzy databases," *Fuzzy Sets and Systems*, vol. 16, no. 15, pp. 2206–2223, 2019.

[7] M. Diabat, "Fuzzy data mining for autism classification of children," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 7, pp. 11–17, 2018.

[8] P. Krömer, S. Owais, and J. Platoš, "Towards new directions of data mining by evolutionary fuzzy rules and symbolic regression," *Computers & Mathematics with Applications*, vol. 6, no. 2, pp. 190–200, 2019.

[9] H. Lin, "Personnel selection using analytic network process and fuzzy data envelopment analysis approaches," *Computers & Industrial Engineering*, vol. 59, no. 4, pp. 937–944, 2019.

[10] C. Xiao and W. Feng, "Application of data mining on enterprise human resource performance management," *Information Management, Innovation Management and Industrial Engineering*, vol. 20, no. 10, pp. 151–153, 2018.

[11] H. Zhang, "Fuzzy evaluation on the performance of human resources management of commercial banks based on improved algorithm," *Power Electronics and Intelligent Transportation System*, vol. 20, no. 1, pp. 214–218, 2019.

[12] R. Devi and R. Amalraj, "Performance analysis of human resource processes using fuzzy data mining approach," *IEEE Access*, vol. 3, no. 22, pp. 119–127, 2019.

[13] W. Chen, M. Panahi, and H. Pourghasemi, "Performance evaluation of GIS-based new ensemble data mining techniques of adaptive neuro-fuzzy inference system (ANFIS) with genetic algorithm (GA), differential evolution (DE), and particle swarm optimization (PSO) for landslide spatial modelling," *Catena*, vol. 2, no. 157, pp. 310–324, 2019.

[14] G. Delgado, V. Aranda, and J. Calero, "Using fuzzy data mining to evaluate survey data from olive grove cultivation," *Computers and Electronics in Agriculture*, vol. 65, no. 1, pp. 99–113, 2020.

[15] S. Strohmeier and F. Piazza, "Domain driven data mining in human resource management: a review of current research," *Expert Systems with Applications*, vol. 4, no. 7, pp. 2410–2420, 2020.

[16] H. Jantan, R. Hamdan, and Z. Othman, "Intelligent techniques for decision support system in human resource management," *Decision Support Systems*, vol. 2, no. 10, pp. 261–276, 2020.

[17] S. Kale and S. Patil, "Data mining technology with fuzzy logic, neural networks and machine learning for agriculture," *Data Management, Analytics and Innovation*, vol. 2, no. 19, pp. 79–87, 2020.

[18] M. Sumathi and G. Anitha, "Energy efficient wireless sensor network with efficient data handling for real time landslide monitoring system using fuzzy data mining technique,"

Journal of Mobile Network Design and Innovation, vol. 8, no. 3, pp. 179–193, 2019.

[19] V. Shahhosseini and M. Sebt, "Competency-based selection and assignment of human resources to construction projects," *Scientia Iranica*, vol. 18, no. 3, pp. 163–180, 2018.

[20] S. Bhattacharya and V. Bhatnagar, "Critical parameters for fuzzy data mining," *Research Methods: Concepts, Methodologies, Tools, and Applications*, vol. 2, no. 15, pp. 1–18, 2019.

[21] R. Santhosh and M. Mohanapriya, "Generalized fuzzy logic based performance prediction in data mining," *Materials Today: Proceedings*, vol. 20, no. 12, pp. 22–29, 2018.

[22] A. Azar, M. Sebt, and P. Ahmadi, "A model for personnel selection with a data mining approach: a case study in a commercial bank," *Journal of Human Resource Management*, vol. 11, no. 1, pp. 1–10, 2018.

[23] H. Jantan, A. Hamdan, and Z. Othman, "Knowledge discovery techniques for talent forecasting in human resource application," *Academy of Science, Engineering and Technology*, vol. 2, no. 50, pp. 775–783, 2019.

[24] T. Chen and Y. Chang, "A study on the rare factors exploration of learning effectiveness by using fuzzy data mining," *Journal of Mathematics, Science and Technology*, vol. 13, no. 6, pp. 2235–2253, 2017.

[25] Y. Han, Z. Geng, and Q. Zhu, "Energy efficiency analysis method based on fuzzy DEA cross-model for ethylene production systems in chemical industry," *Energy*, vol. 8, no. 3, pp. 685–695, 2018.

[26] H. Jantan, R. Hamdan, and Z. Othman, "Towards applying data mining techniques for talent management," *Computer Engineering and Applications*, vol. 2, no. 11, pp. 12–19, 2019.

[27] J. Fang, C. Liu, T. E. Simos et al., "Neural network solution of single-delay differential equations," *Mediterranean Journal of Mathematics*, vol. 17, no. 1, pp. 1–15, 2020.

[28] J. Wang, Y. Huang, T. Wang, C. Zhang, and Y. H. Liu, "Fuzzy finite-time stable compensation control for a building structural vibration system with actuator failures," *Applied Soft Computing*, vol. 93, p. 106372, 2020.

[29] F. Orujov, R. Maskeliūnas, R. Damaševičius, and W. Wei, "Fuzzy based image edge detection algorithm for blood vessel detection in retinal images," *Applied Soft Computing*, vol. 94, p. 106452, 2020.

[30] G. Dong, W. Wei, X. Xia, M. Woźniak, and R. Damaševičius, "Safety risk assessment of a Pb-Zn mine based on fuzzy-grey correlation analysis," *Electronics*, vol. 9, no. 1, p. 130, 2020.

[31] L. Ding, S. Li, H. Gao et al., "Adaptive partial reinforcement learning neural network-based tracking control for wheeled mobile robotic systems," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 7, pp. 2512–2523, 2018.

[32] J. Yang, J. Zhang, C. Ma, H. Wang, J. Zhang, and G. Zheng, "Deep learning-based edge caching for multi-cluster heterogeneous networks," *Neural Computing and Applications*, vol. 32, no. 19, pp. 15317–15328, 2020.

[33] A. Onan, S. Korukoğlu, and H. Bulut, "Ensemble of keyword extraction methods and classifiers in text classification," *Expert Systems with Applications*, vol. 57, pp. 232–247, 2016.

[34] E. P. Sarabi and S. A. Darestani, "Developing a decision support system for logistics service provider selection employing fuzzy MULTIMOORA & BWM in mining equipment manufacturing," *Applied Soft Computing*, vol. 98, no. 71, p. 106849, 2020.