Volume : 1 | Issue : 2 | November 2011 ISSN - 2249-555X

Research Paper Computer Science

Performance analysis of Human Resource Processes using Fuzzy data mining approach

* R. Lakshmi Devi ** Dr. R. Amalraj

* Assistant Professor, Bharathi Women's College, Broadway

** Associate Professor, Sri Vasavi College, Erode

ABSTRACT

An efficient tool should be required to analyze the performance of Human Resource (HR) processes to make decision and strategic plan in an organization. The effective mathematical tool to deal with the vagueness and uncertainty, fuzzy data mining is considered as a highly desirable tool being applied to many application areas. This paper briefly explains the basic fuzzy data mining theory and proposed fuzzy data mining algorithm. We present and justify the capabilities of fuzzy data mining technology in the evaluation of human resource in a college environment by proposing a practical model for improving the efficiency and effectiveness of human resource management.

Key word : Human Resource, Performance Analysis

Introduction

Data is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.

Most of the companies collect and refine massive quantities of data. Data mining techniques can be implemented rapidly on existing software and hardware platforms to enhance the value of existing information resources, and can be integrated with new products and systems as they are brought on-line. Also it is possible to use the Data Mining approach in environments like Institution to enhance the effectiveness of the students. Data mining techniques are the result of a long process of research and product development. This evolution began when business data was first stored on computers, continued with improvements in data access, and more recently, generated technologies that allow users to navigate through their data in real time. Data mining takes this evolutionary process beyond retrospective data access and navigation to prospective and proactive information delivery.

The Scope of Data Mining

Data mining derives its name from the similarities between searching for valuable business information in a large database. Like finding linked products in gigabytes of store scanner data and mining a mountain for a vein of valuable ore. Both processes require either sifting through an immense amount of material, or intelligently probing it to find exactly where the value resides. Given databases of sufficient size and quality, data mining technology can generate new business opportunities by predicting trends and behaviors automatically. Data mining automates the process of finding predictive information in large databases. A typical example of a predictive problem is targeted marketing. Data mining uses data on past promotional mailings to identify the targets most likely to maximize return on investment in future mailings. Other predictive problems include forecasting bankruptcy and other forms of default, and identifying segments of a population likely to respond similarly to given events.

Data mining also discovers the previously unknown patterns automatically. Data mining tools sweep through databases and identify previously hidden patterns in one step. An example of pattern discovery is the analysis of retail sales data to identify seemingly unrelated products that are often purchased together. Other pattern discovery problems include detecting fraudulent credit card transactions and identifying anomalous data that could represent data entry keying errors.

Fuzzy data mining methods can mean data mining methods that are fuzzy methods as well. This process can also mean approaches to analyze fuzzy data. In some sense, the later ones are fuzzy methods as well but the conceptions are different. Fuzzy data means imprecise, vague, uncertain, ambiguous, inconsistent, and/or incomplete data. Therefore, the source of uncertainty is the data themselves. It is very important to develop methods that are able to handle this kind of data because data from several information sources might be fuzzy logic.

Fuzzy data Mining algorithm for Human resource Management

The human resources of an organization consist of all people who perform its activities. Human resource management (HRM) is concerned with the personnel policies and managerial practices and systems that influence the workforce. In broader terms, all decisions that affect the workforce of the organization concern the HRM function. The activities performed by HRM professionals fall under five major domains like Organizational design, Staffing, Performance Management and Appraisal, Employee and Organizational Development and Reward Systems, Benefits.

Acquiring human resource capability should begin with organizational design and analysis. Organizational design involves the arrangement of work tasks based on the interaction of people, technology and the tasks to be performed in the context of the objectives, goals and the strategic plan of the organization. HRM activities such as human resources planning, job and work analysis, organizational restructuring, job design, team building, computerization, and worker-machine interfaces fall under this domain. The HR practices in an educational institution signify the Library, infrastructure, Welfare schemes, Distance education etc. In this paper, we considered the fuzzy data mining algorithm to improve the HR policies of an Institution.

Algorithm

In order to apply the fuzzy data mining algorithm for performance evaluation, we firstly introduce a set of relevant fuzzy cluster formulas and describes how to calculate relationship between records. The aim of cluster analysis is the classification of network connections based on similarities among them and organizing objects into groups. A cluster can be defined as set of objects that are more similar to other ones than to other clusters. Similarity between two groups is often defined by means of distance based upon the length from a data vector to some prototypical object of the cluster.

Since clusters can formally be seen as subsets of the data set, one possible classification method can be whether the subsets are fuzzy or crisp (hard). Hard clustering methods are based on classical set theory, and it requires an object that either does or does not belong to a cluster. Fuzzy clustering methods (FCM) allow objects to belong several clusters simultaneously with different degrees of membership.

The data set (R) is partitioned into r fuzzy subsets. In many real situations, fuzzy clustering is more natural than hard clustering, as objects on the boundaries between several classes are not forced to fully belong to one of the classes. However, they are assigned to membership degrees between 0 and 1 indicating their partial memberships.

Data Normalization

The data are typically observations of some phenomenon. Each object consists of m measured variables, grouped into an m-dimensional column vector and set of n objects is denoted by . The data set can be expressed as a matrix.

$$\mathbf{R}=\begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1m} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_{n1} & \alpha_{n2} & \cdots & \alpha_{nm} \end{bmatrix} \qquad \text{..... (1)}$$

The data set is normalized by subtracting its mean and divided by standard deviation.

$$\alpha^{new}_{ik} = \left| \frac{\alpha_{ik} - \overline{\alpha_k}}{S_k} \right| \qquad i=1,2,\ldots,n,k=1,2,\ldots,m \qquad \text{.....(2)}$$

Where $\overline{\alpha_k}$ and S_k is the mean value and standard deviation. | | is the absolute operation.

α^{new}_{ik} may not be in the interval [0,1]. Hence it will be still be modified as $\alpha_{ik}^{modified}$ in order to come in the interval [0,1].

$$\alpha_{ik}^{modified} = \frac{\alpha_{ik}^{new} - \min_{1\le i\le n}\{\alpha_{ik}^{new}\}}{\max_{1\le i\le n}\{\alpha_{ik}^{new}\} - \min_{1\le i\le n}\{\alpha_{ik}^{new}\}} \qquad (k=1,2,\cdots,m) \qquad \text{.....(3)}$$

Generation of Correlation Coefficient

In order to cluster objects, we build a fuzzy similar matrix $\mathbf{\Phi}$ to determine the correlation coefficient between α_i and α_j $r_{ij} = \Phi(\alpha_i, \alpha_j)$ The general form of similar matrix $\mathbf{\Phi}$ is as follows:

$$\Phi=\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix} \qquad \begin{matrix} 0 \le r_{ij} \le 1 \\ i,j=1,2,\ldots,n \end{matrix} \qquad \text{.....(4)}$$

Where the above mentioned auto correlation matrix is a square matrix. When the values of the elements are maximum then there is a strong correlation between and and if it is zero then it is completely independent. A commonly used similarity metric is the cosine of the angle between two objects. In this paper we calculate them by:

$$r_{ij} = \frac{\sum_{k=1}^{m} |\alpha_{ik} - \overline{\alpha_i}||\alpha_{jk} - \overline{\alpha_j}|}{\sqrt{\sum_{k=1}^{m}(\alpha_{ik} - \overline{\alpha_i})^2}\sqrt{\sum_{k=1}^{m}(\alpha_{jk} - \overline{\alpha_j})^2}} \qquad \text{.....(5)}$$

$$\overline{\alpha_i} = \frac{1}{m}\sum_{k=1}^{m}\alpha_{ik} \;, \quad \overline{\alpha_j} = \frac{1}{m}\sum_{k=1}^{m}\alpha_{jk} \qquad \text{.....(6)}$$

Application in Performance Evaluation

The above mentioned steps are used in this paper to measure the effectiveness of the student's performance in an education institution. The input samples are collected from students based on some categories. In this experiment, we collect 12 evaluation records. Each record has four features: they are learning level, innovation ability, independence work ability, reading efficiency. Each feature has four states of 1, 2, 3 and 4. They are the corresponding score of each feature. We can interpret them as worse, general, better, best. The sample space is given in Table.1 and portrayed in Figure.1.

Table 1: The Sample Space

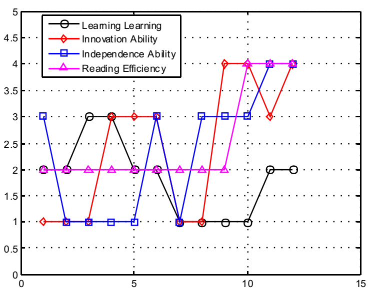
Sample	Features			
	Learning Level	Innovation Ability	Independence Ability	Reading Efficiency
α_1	2	1	3	2
α_2	2	1	1	2
α_3	3	1	1	2
α_4	3	3	1	2
α_5	2	3	1	2
α_6	2	3	3	2
α_7	1	1	1	2
α_8	1	1	3	2
α_9	1	4	3	2
α_{10}	1	4	3	4
α_{11}	2	3	4	4
α_{12}	2	4	4	4

We get the average value vector and the standard deviation vector is as follows

$$\overline{\alpha_k} = \{1.8, 2.4, 2.3, 2.5\}$$

$$S_k = \{0.718,\ 1.311, 1.231,\ 0.905\} \qquad k=1,\ 2, 3,4$$

Figure 1: Sample space



ISSN - 2249-555X

The normalization matrix is given in below Table. 2.

Table 2: Normalization matrix.

0.00	0.83	0.00	0.00
0.00	0.83	0.67	0.00
1.00	0.83	0.67	0.00
1.00	0.00	0.67	0.00
0.00	0.00	0.67	0.00
0.00	0.00	0.00	0.00
0.67	0.83	0.67	0.00
0.67	0.83	0.00	0.00
0.67	1.00	0.00	0.00
0.67	1.00	0.00	1.00
0.00	0.00	1.00	1.00
0.00	1.00	1.00	1.00

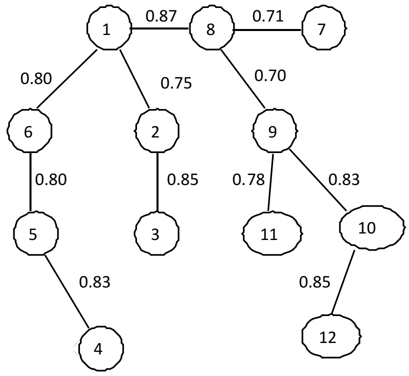
The fuzzy similar matrix is given in Table 3.

Table 3: Correlation coefficient

1.00											
0.75	1.00										
0.66	0.85	1.00									
0.54	0.66	0.77	1.00								
0.60	0.75	0.66	0.88	1.00							
0.80	0.60	0.54	0.72	0.80	1.00						
0.62	0.83	0.71	0.55	0.62	0.50	1.00					
0.87	0.62	0.55	0.45	0.50	0.70	0.71	1.00				
0.63	0.45	0.41	0.58	0.63	0.81	0.50	0.70	1.00			
0.53	0.38	0.35	0.50	0.53	0.69	0.41	0.58	0.83	1.00		
0.61	0.46	0.42	0.57	0.61	0.76	0.38	0.53	0.64	0.78	1.00	
0.57	0.42	0.40	0.53	0.57	0.71	0.35	0.50	0.71	0.85	0.92	1.00

The maximal tree is shown in Fig 1. When λ =0.78, the samples can be divided into four clusters: {α_1, α_4, α_5, α_6, α_8} (A cluster: better), {α_2, α_3} (B cluster: general), {α_7} (C cluster: worse), {α_9, α_{10}, α_{11}, α_{12}} (D cluster: best). Then we can obtain the average index of each cluster. The raw data of human resource evaluation is {3, 1, 3, 2}. Compare the data with each cluster and calculate the proximal values. Then we can determine that the sample belong to A cluster, that is the evaluation of human resource is better.

Figure 2: Graphic representation of maximal tree



Conclusions

This paper has been an effort in providing the motivation towards advancing the traditional human resources appraisal and selection via data mining technology. The motivation and need for data mining is given in a detailed manner. The main idea of this analysis has been organized into a model, which targets the superior advantages of fuzzy data mining in performance evaluation process. It can help teachers to improve the effectiveness of the student. The results obtained are similar to practical one and gives high amount of effectiveness. Fuzzy data mining theory has been applied to increasing domains and its advantage had been also more and more evident since it appeared.

REFERENCES

Doug Alexander. Data Mining. Retrieved from http://www.laits.utexas.edu/~norman/BUS.FOR/course.mat/Alex/ | Balazs Feil, Janos Abonyi. Introduction to Fuzzy Data Mining Methods. Retrieved from http://www.irma-international.org/viewtitle/20349/ | H John Bernadin, Human Resource Management, Fourth Edition, McGraw Hill, 2007 | R. J. Rummel. Understanding Correlation. Retrieved from http://www.mega.nu/ampp/rummel/uc.htm