Research on Data Mining Method of Human Resource Job Matching under the Trend of Data Management

Chen Huang[a,*], Xuefei Liu[b]

School of Economics and Management, Yunnan Technology and Business University, Guanjun Street, Kunming, Yunnan, China

[a]283408358@qq.com, [b]14379401@qq.com

*Corresponding author

Keywords: Data Management, Human Resources, Job Matching, Data Mining

Abstract: Human resource management is a systematic project, and the construction of human resource management information system is the key to realize human resource information management. Under the trend of data management, in the fierce market competition, how to improve the core competitiveness of enterprises and improve the level of human resources management are the key elements. Based on corporate vision and strategy, the effective application of data mining technology in enterprise human resources management is beneficial to the reasonable matching of people and posts, giving full play to employees' work abilities and potentials, improving the performance of organizations and employees, and realizing the sustainable development of enterprises. In this paper, the method of information extraction is used for information resource scheduling and feature extraction of human resource management, and the standardized regular migration model is used for information registration of personnel and posts in human resource management, and the optimal scheduling design of human resource is carried out. The autocorrelation component mining method is used for human resource information mining, and the optimal modeling of human resource information management is realized.

1. Introduction

The core of human resource management is the so-called post matching. The human resource management strategy based on post matching adopted by enterprises can realize the effective human resource management of enterprises to a certain extent. The traditional matching theory of people and jobs focuses on selecting the right person for a specific job, which has certain defects [1]. People-organization matching emphasizes the matching, mutual influence and interdependence between employees' characteristics and organizational characteristics, which is conducive to enhancing employees' loyalty and improving organizational efficiency. It is not enough for an enterprise to accumulate human resources. It is necessary to allocate human resources effectively and rationally in order to maximize its benefits. Otherwise, the first resource will only make the enterprise produce greater internal friction [2-3].

Under the trend of data management, human resources are the first resource, and having a talent team with reasonable knowledge structure and strong innovation ability is the key to the success or failure of enterprise technological innovation and sustainable development. Therefore, the management of human resources, especially the scientific management of knowledge-based talents, has become the core content of enterprise technological innovation [4]. Through data mining, the related data clustering of human resources management information data is realized, and a high-dimensional information space reflecting the characteristics of human resources management information data is constructed. In the information space, information fusion method is used to process big data information, which improves the scheduling ability of human resources, thus realizing the optimal structure design of human resources.

  DOI: 10.25236/meeit.2024.018

2. Human Resource Management and Data Mining

2.1 Human resource management

The evolution of human resource management is adapted to the social, economic and cultural background, and in the final analysis, it is the product of production relations adapting to the development of productive forces. It has gone through the stages of personnel management, human resource management, human resource management development and contemporary human resource management theory-strategic human resource management [5].

The greatest characteristic of strategic human resource management is to combine human resource management with the realization of strategic objectives of organizational development. In the organizational strategic management, the original administrative and transactional management functions are being replaced by today's human resource management, which has changed from administrative management and transactional management to strategic management. In many literatures, there is no consensus on the definition of strategy, but it is various. Strategic human resource management focuses on the relationship between policies and practices of human resource management and the results of enterprise behavior, and all related research methods are strategic. It is the main contribution of strategic human resource management to explain the change of organizational behavior results [6].

2.2 Data mining

Data mining technology is a kind of deep-level data analysis method, which is a process of extracting potentially useful information and knowledge hidden in a large number of incomplete, noisy, fuzzy and random practical application data that people do not know in advance. According to different knowledge discovery methods, it mainly includes machine learning algorithms (such as inductive learning, rule induction, case-based learning, decision tree and genetic algorithm), statistical methods (such as regression analysis, cluster analysis, discriminant analysis and exploratory analysis), neural network methods (such as BP neural network algorithm and self-organizing neural network algorithm) and database methods. According to the different methods of discovery tasks, they can be divided into classification, clustering, time series prediction, association rule discovery and evolution rule discovery.

3. The Role of Data Mining in Matching People and Posts

3.1 Matching relation classification

Clustering algorithm based on data mining is used to divide the matching relationship between people and posts in different employee groups, avoiding the subjective randomness brought by traditional methods, thus embodying the differentiated management strategy for different employee groups in enterprises and maximizing the value of human resources.

3.2 Matching relation prediction

Grasping the dynamic characteristics of the matching relationship between people and posts is the key to the matching management of people and posts. By collecting and sorting real-time dynamic relationship data, data mining integrates and mines multi-factor data, and extracts the state characteristics of the matching relationship between people and posts. At the same time, it simulates and predicts the evolution process of the relationship by combining historical data and future prediction data, which effectively promotes the development of the matching relationship between people and posts in a favorable direction [7].

3.3 Employee segmentation and evaluation of matching between people and posts

By inputting the index characteristic value of the employed employees, the network can output the matching results, so as to help enterprises formulate appropriate recruitment strategies and select the best talents. The decision tree classification algorithm of data mining is used to subdivide the

different attribute indexes of employees, so as to determine which attributes will lead to the status of employees, so as to carry out more targeted job training, save the training cost of employees and achieve effective management performance.

4. Problems Existing in the Allocation of Human Resources in Chinese Enterprises

Human resources allocation is so important, but in many domestic enterprises and industries, there are more or less unreasonable human resources allocation. Especially, the problem of "overqualified" human resource allocation exists in many enterprises. Looking at some practices of human resources allocation in Chinese enterprises, the following human resources allocation problems deserve the attention of enterprise leaders and human resources departments:

4.1 Plan configuration or lead subjective configuration

The planned allocation can't keep up with the change of economic structure, which leads to the shortage of escape sequence and structure of human resources distribution and the surplus of human resources and production equipment in different units and departments. Such defects exist in many state-owned enterprises. In some enterprises, the allocation of human resources is completely decided by the leaders. In the case of planned allocation and subjective allocation of leaders, the rigid matching of posts coexists with job vacancies in many departments, which often makes the posts and departments that should be staffed unable to get the required talents, while the posts and departments that have reached the saturation of personnel have to accept some talents assigned by their superiors [8].

4.2 Poor allocation of human resources leads to serious internal friction in enterprises

These problems generally exist in Chinese enterprises: the organizational structure is unreasonable, the management level is not high, some members of the leadership team are not United, and they do not support each other in their work, but break down each other; The work between departments is disjointed or wrangling with each other. In particular, in enterprises with family management, the allocation of human resources is to use relatives instead of strongmen, and posts are set up for people instead of things, so they lack long-term planning for quick success and instant benefit, which weakens the internal cohesion of enterprises and makes the use of human resources inefficient.

4.3 Individual and position do not match

Due to the lack of scientific job analysis and talent evaluation methods in most domestic enterprises, the requirements of job responsibilities, tasks and positions on personnel are unclear. At the same time, the applicants can not be fully grasped from their knowledge, skills, abilities and personality, which makes it difficult to match individuals and positions in the recruitment stage. After recruiting talents, they are still managed, ignoring the training and development of existing talents, which makes the mismatch between individuals and posts persist.

5. Data Mining Method of Human Resource Post Matching

5.1 Establish the relevant index system of the evaluation object

Under the trend of data management, to evaluate the matching degree of human resources posts, we must rely on a complete and scientific evaluation index system to achieve [9]. In view of the differences of posts in real work, it is impossible to have a set of evaluation index system that is generally suitable for all posts. Therefore, before evaluating, each unit should first design a corresponding evaluation index system according to the actual characteristics of posts of different natures and levels, so as to prepare for the next evaluation. In this paper, the index system for evaluating the matching degree of human resources is shown in Table 1.

Table 1 Evaluation index system of human resource matching degree

Target layer	General manager's comprehensive performance		
Main factor layer	Working attitude	Work performance 1	Work performance 2
Sub-factor layer	Team spirit at work	Completion of development volume	Social communicative competence
	Implementation of business plan	Completion of work indicators	Sensitivity to the needs of others
	To the company's development plan	Engineering quality situation	Adaptability to the environment
	Attention to strategy	Safety production situation	Ability to manage subordinates
	A sense of responsibility for work	Completion of monthly plan	Ability to think strategically
	Concern for the growth of employees		
	Concern for work efficiency		

Human resources management should make the best use of people's talents, make the best use of them, adapt personnel and give full play to the role of human resources. However, how to realize scientific and rational allocation is an important problem that needs to be solved urgently in human resource management for a long time. Training talents from within, providing opportunities and challenges to the competent people, creating tension and incentive atmosphere, is the driving force for the development of the company [10]. However, this is not to exclude the introduction of necessary external talents. This is a last-place elimination configuration mode) and two-way selection (Figure 1).

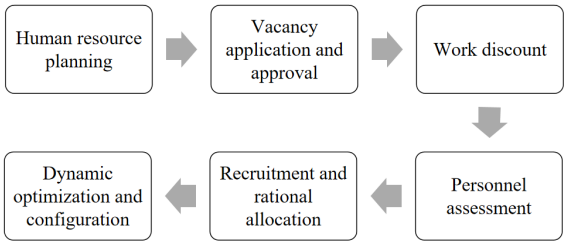

Figure 1 Personal-post dynamic matching model

In order to comprehensively investigate the overall adaptability of the appraisee to the post, it is necessary to evaluate the matching degree of the appraisee to the post in each sub-index, so an evaluation result set must be established for the above evaluation first. In this evaluation, the evaluation team is used to determine the weight coefficient by AHP after collective discussion according to the specific situation of each post in the unit, which is simple and objective. As the personnel in the evaluation team have different influences on the evaluation results, it is necessary to assign different weights to the results of various evaluation personnel, and give their evaluation weight distribution matrix $W=(w_1,w_2,\cdots,w_m)$, m is the number of people evaluated, so that their evaluation results have a certain degree of influence on the final result, but the final result is the result of their comprehensive evaluation.

5.2 Data mining of human resource management information

The method of information extraction is used to schedule and extract the information resources of human resource management, and the second-order invariant moment of the information flow model of human resource management information data is calculated [11]. For the nonlinear sequence $f(x,y)$ of human resource management information data, its $(p+q)$-order geometric moment is:

$$m_{pq}=\sum_{m=1}^{M}\sum_{n=1}^{N}x^{p}y^{p}f(x,y) \tag{1}$$

In the formula, (x,y) is the polar coordinate of nonlinear time series distribution of human resource management information data. According to the invariance principle of scale segmentation, the characteristic distribution function of nonlinear time series characteristics of human resource management information data is obtained as follows:

$$\mu_{pq}=\sum_{m=1}^{M}\sum_{n=1}^{N}(x-\dot{x})^{p}(y-\dot{y})^{q}f(x,y) \tag{2}$$

Through distributed mining of main feature vectors, it is found that the feature function of association rules of human resource management information data satisfies $I_j\in I(j=1,2,\cdots,n)$, and the rank of decision matrix satisfies $B_k=\left(r_{ij}^{(k)},0\right)_{m\times n}$.

A standardized regular migration model is constructed to schedule human resource management information data [12], and the regular training migration rule of human resource scheduling is described as follows:

$$V_0(k)=\begin{cases}\gamma(1)\gamma^{T}(1) & k=1\\ \dfrac{\left[\rho V_0(k-1)+\gamma(k)\gamma^{T}(k)\right]}{1+\rho} & k>1\end{cases} \tag{3}$$

In which,

$$\gamma(k)=\hat{M}(k)-\hat{H}(k)\hat{x}(k|k-1) \tag{4}$$

The registration information function of employees and positions in HR management information data is written as follows:

$$Sim(X,Y)=Cos(X,Y)=\frac{C(X)\cdot C(Y)}{|C(X)|\cdot|C(Y)|} \tag{5}$$

Formula (5) represents a set of similarity characteristic parameters of resource information distribution, and human resource information scheduling is carried out by principal component analysis, and the correlation accompanying quantity of human resource scheduling is defined as:

$$S_{i,j}(t)\frac{p_{i,j}(t)-sp_{i,j}(t)}{p_{i,j}(t)} \tag{6}$$

$T_{i,j}(t)$ is defined as the non-cooperative game function of the optimal structure distribution of human resources, and the calculation formula is:

$$T_{i,j}(t)=\frac{\left|p_{i,j}(t)-\Delta p(t)\right|}{p_{i,j}(t)} \tag{7}$$

$U_{i,j}(t)$ is defined as the end-to-end channel allocation measurement value of human resource scheduling, and the calculation formula is:

$$U_{i,j}(t)=\exp\left[-b\left[z_i(t)-z_j(t)\right]^{2}\right] \tag{8}$$

In which: $p_{i,j}(t)$ is the information throughput of a single group of HR information scheduling

nodes at time t; $sp_{i,j}(t)$ is the average value of information amplitude of extracted human resources. Through the design of standardized regular migration model, the information registration of personnel and posts in human resource management is carried out [13].

5.3 Application analysis of decision tree algorithm in human resources

Decision tree technology is the main technology for classification and prediction, and decision tree learning is an inductive learning algorithm based on examples. It focuses on reasoning the classification rules of decision tree representation from a group of unordered and irregular cases. It adopts top-down recursion, compares the attribute values at the internal nodes of the decision tree, judges the branches down from the nodes according to different attributes, then prunes them, and finally draws a conclusion at the leaf nodes of the decision tree. ID3 is a decision tree algorithm proposed earlier and widely used. It forms the decision tree through the selection window, which uses the mutual information in information theory to find the attribute field with the largest amount of information in the training set, establishes a node of the decision tree, and then establishes the branches of the tree according to the different values of the attribute field; repeats the process of establishing the lower nodes and branches of the tree in each branch subset.

In ID3 algorithm, the attribute selection of decision nodes is mainly accomplished by using the concept of entropy in information theory. In this attribute selection method, the decision attribute with the maximum information gain (or the maximum entropy decreasing direction) is selected as the current node. The node attributes selected in this way can ensure that the decision tree has the minimum number of branches, which makes the redundancy of the decision tree minimal. Figure 2 shows the generation of decision tree.

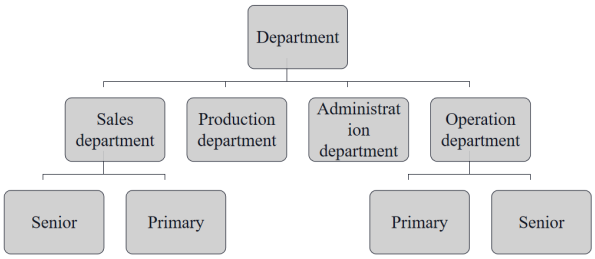

Figure 2 Decision tree generation

Because the attribute "department" has the largest information gain value, it is selected as the root node of the decision tree. Repeat the above steps for each branch. For example, for the branch "attribute department =' sales department'", the subset given in the following table can be generated, and the above attribute selection operation can be repeated for this subset. A complete decision tree can be obtained by performing the same operations on other branches in turn. After that, the decision tree can be pruned.

In general decision trees, information is used as a parameter to evaluate the quality of node splitting. In SLIQ algorithm, $gini$ index is used instead of information. $gini$ index has better performance than information and is convenient to calculate. For data sets containing n classes, S and $gini(S)$ are defined as:

$$gini(S)=1-\sum p_j * p_j \quad (9)$$

p_j is the frequency of class j data in S, and the smaller $gini$ is, the greater Information Gain is.

Different from general decision tree, SLIQ adopts binary search tree structure, and every node needs to calculate the best splitting scheme first, and then execute splitting.

For numerical continuous field split form $A \le v$. Therefore, numerical fields can be sorted first, assuming that the sorted result is $v_1, v_2, \cdots, v_n$, because splitting will only occur between two nodes, so there are $n-1$ possibilities. Generally, the midpoint $(v_i + v_i + 1)/2$ is taken as the split point, and

different split points are taken in turn from small to large. The one with the largest Information Gain index ($gini$ minimum) is the split point. Because each node needs to be sorted, this operation is extremely costly.

For discrete fields, let $S(A)$ be all possible values of A, and the split test will take all subsets S' of S. Looking for the $gini$ index when splitting into S' and $S-S'$ blocks, and getting the minimum $gini$ is the best splitting method. Obviously, this is a process of traversing all subsets of a set S, which needs to be calculated for $2|S|$ times, and the cost is also very high.

Decision tree provides an algorithm basis for human resource management decision makers to find key attributes. Its value lies in that it can scientifically and accurately evaluate existing data and predict trends according to algorithm rules. In terms of data structure, once the system gets a piece of data with complete information, the data will be classified into the category it belongs to according to the preset algorithm of the system, which becomes the attribute determination of the class label attribute, and the model is constructed by analyzing the database tuple described by the attribute. From the level of business decision-making, if you want to know what attribute is the key factor to determine and distinguish other marked attributes when a certain marked attribute is specific, you must use the decision tree algorithm. In this way, human resource decision makers can better identify and judge business events, and then make corresponding more scientific predictions and decisions.

6. Conclusion

Whether the allocation of human resources is reasonable or not has great influence on the short-term performance and long-term development of enterprises, so it should be paid enough attention to. Under the trend of data management, enterprises should also follow the relevant theories and methods of human resource allocation after completing the recruitment of talents, so that talents can be matched with posts. At present, data mining technology and its application are a hot research topic in the world, and have been applied well in many industries, especially in marketing, which initially reflects its superiority and development potential. In the field of human resources information management, the only way to realize the development and management of human resources is to apply data mining technology and artificial intelligence technology comprehensively.

References

[1] Liu yi. Research on the direction of enterprise human resource management reform under the background of data mining. Think tank era, no. 036, pp. 67-167, 2019.

[2] Qi Wenzhu. Analysis of the quality needs of human resource managers based on data mining. Modern Economic Information, no. 02, pp. 28+30, 2020.

[3] Ren Ranran. Research on Job Recommendation of AI Company Based on Data Mining. Value Engineering, no. 34, pp. 42-44, 2017.

[4] Mu Chunsheng. On the application of data mining technology in human resource management. New Business Weekly, no. 012, pp. 175, 2018.

[5] Lai Huaqiang, Wang Sanyin, Zhong Chonggao. Prospect of data mining application in the field of human resource management-taking the empirical analysis of turnover management based on grey relational model as an example. Jiangsu Business Theory, no. 008, pp. 42-47, 2018.

[6] Jiang Peng. Research and design of employment recommendation algorithm for college students based on convolutional neural network. Digital Design. CG World, vol. 008, no. 014, pp. 23-24, 2019.

[7] Fang Zhong, Zhang Xu. Application of data mining in information management of human resources. University of Science and Technology, no. 022, pp. 13, 2017.

[8] Liu Junqun. Analysis of personal factors and post matching of employees in state-owned petrochemical enterprises and countermeasures. Management, no. 16, pp. 170, 2017.

[9] Sheng Liwen. On the application of data mining technology in human resource management. Finance and Economics(Academic Edition), no. 030, pp. 131-132, 2018.

[10] Sun Jintao. Application of data mining technology in enterprise human resource management. Science and Technology Wind, no. 024, pp. 251, 2019.

[11] Fan Yang. human resource information management based on data mining technology. information and communication, no. 004, pp. 152-153, 2017.

[12] Liu Yuzhong. Application of data mining technology in enterprise human resource management. Chinese and foreign entrepreneurs, vol. 648, no. 22, pp. 101-102, 2019.

[13] Zhong Yilin. application of data mining technology in enterprise human resource management. modern economic information, no. 17, pp. 89, 2019.