Human Resource Allocation Effectiveness Assessment Based on K-Means Algorithm

Xuan Wang[1, a, *]
[a]671045447@qq.com

[*] Corresponding author: 671045447@qq.com

[1]School of Management, Shanghai University, China

Abstract—Based on the modern human resource theory of "personnel quality" and the requirement of balanced staffing, this paper selects the K-Means clustering algorithm, which is simple and efficient in multidimensional data mining analysis with low time complexity and spatial complexity, as a tool for data mining analysis of human resources allocation in audit departments, taking into account the actual high professional quality requirements for personnel in audit operations and the use of data mining technology in human resources allocation analysis at home and abroad. The K-Means clustering algorithm, which is simple and efficient in multidimensional data mining analysis with low time complexity and space complexity, is selected as a tool for data mining analysis, thus clarifying the main direction of data mining analysis research on human resource allocation in audit departments.

Keywords- human resources; big data; K-Means algorithm

1. Preface

An important aspect of advances in computer information technology is the creation of data mining techniques. Historically, because decision makers did not have effective means to extract valuable knowledge from massive amounts of data, many important decisions relied on the intuitive feelings of decision makers and lacked a direct basis for data analysis, making the decision-making process somewhat flawed. The development of data mining technology has provided effective tools for people to convert data into knowledge [1].

Although the HR information system brings together a large amount of personnel information, it is obviously not enough for the staff of HR management department if it is limited to the completion of regular functions such as information collection. HRM department staff need to effectively integrate unit HR information and business information, conduct deeper data processing and data mining, discover some hidden patterns, trends and abnormalities and other information, re-integrate HR in a timely manner, and propose new targeted management measures. Through in-depth data analysis, HRM can take measures to fully explore the potential of various types of personnel, reduce the waste of human resources and seek greater benefits for the unit; it can also reflect the innovative working ability of HRM and improve the practical value of HRM work. Data mining technology of computer provides some new ideas and good solutions for solving the above problems [2].

BDEIM 2022, December 02-03, Zhengzhou, People's Republic of China

DOI 10.4108/eai.2-12-2022.2328746

2. Introduction of K-Means clustering algorithm principle

K-Means algorithm is a well-known algorithm in distance-based clustering algorithm. It uses distance as the evaluation index of similarity, and its principle is that if the vectors representing the points are close to each other in space, we can consider these points as a class. That is, for a given sample set, the sample set is divided into K classes according to the size of the distance between samples when performing the classification, and let the distance of points within a class be as small as possible, while the distance between classes is as large as possible [3].

The K-Means algorithm has a basic assumption for the data to be clustered: for each class, we can select a centroid such that the distance from all points in the class to the centroid is less than the distance to the centers of the other classes. However, in the actual classification, the data obtained often cannot directly meet such requirements, but can only be as close as possible, and the reasons for such differences are often inherent in the data itself, or the data can no longer be classified. The method of K-Means algorithm is to select K initial reference points according to the input parameter K. Based on the K reference points, all the points in the data set are divided into K classes, and the center of mass (the average of all the points in the class) of these K classes is used as the reference point for the next round of iteration. The iteration makes the selected reference points closer and closer to the real class center of mass, so the clustering effect becomes better and better [4].

3. Experiment and clustering effect analysis

Based on the analysis of HR information data and the study of the basic principles of K-Means algorithm, we need to verify through experiments whether HR-related variables can be clustered effectively, whether the clustering results are similar to the actual situation, and which variables have more influence on the clustering results, so as to provide HR management departments with priority directions when allocating human resources [5].

We mainly perform the validation through three experiments. The first experiment mainly establishes the experimental environment and performs machine learning first, and then randomly selects some variables without explicitly selecting the initial prime to verify the effect on the clustering effect in this case; the second experiment mainly verifies whether the variable selection and explicit initial prime affect the clustering effect by explicitly specifying the prime and using new variables; the third experiment mainly verifies whether the clustering effect is affected by reducing some of the variables to see if it affects the clustering effect, to verify whether there is a possible redundancy in variable selection under the premise of ensuring the clustering effect [6].

Python is a dynamic, object-oriented scripting language, and this paper uses the Python language architecture to establish the experimental environment, establish the analytical model by importing the dataset and K-Means analysis tool, conduct simulation training, and finally derive the clustering results [7].

The main steps of the experiment are.

(1) install the Python tool; (2) import the processed data into the dataset and set the title for the dataset; (3) import the K-Means analysis tool into the Python environment; (4) normalize the

imported data; (5) build the analysis model and set the clustering K value; (6) load the feature vector to be analyzed and perform the algorithm operation; (7) obtain the classification results and merge them with the dataset; (8) output Final results. The main procedure is as follows.

import the K-Means analysis tool into the python environment and import the dataset using pandas

from sklearn.cluster import KMeans

Import pandas as pd

#Read the data used for clustering and create the teaching data table

loan_data=pd.DataFrame(pd.read_csv('SJRY_data.csw',header=0))

loan_data.columns

Index(['XM','ZW','GT','WT','age','DT','zc','ZCJB','XL','ZY','JZJ'],dtype='object')

#Process the training set and normalize the analysis data features

loan_data_zs=1.0*(loan_data-loan_data.mean())/loan_data.std()

build the analytical model, load the feature vector "job title, time in the workforce and age" that needs to be clustered, the number of clusters is 3 loan=np.array(loan_data[['ZW','WT','age']])

clf=KMeans(n_clusters=3)

#Substitute the data into the clustering model

clf=clf.fit(loop)

#Perform simulation training to get predicted values

Cluster=clf_KMeans.fit_predict(X)

print(cluster)

#Add clustering result labels to the original data table

loan_data['label']=clf.labels_

#View the clustering results

print(loan_data)

From the clustering results (Table 1), it can be seen that the samples were classified into three categories.

Table 1 Clustering results

Name	ZW	GT	WT	age	DT	ZC	ZCJB	XL	ZY	JZJ	label
2	4	14	34	56	33	2	3	2	2	0	0
4	3	7	33	56	33	2	3	2	2	0	0
8	3	12	39	57	12	4	2	2	1	0	0
12	5	6	41	60	29	6	4	2	1	0	0
14	2	15	40	58	15	4	2	2	1	0	0

											0
1	4	16	31	53	31	2	3	2	1	1	1
3	4	7	31	52	31	6	4	2	1	0	1
6	4	7	21	43	15	3	4	5	3	1	1
9	3	6	31	53	31	2	3	2	1	0	1
11	4	6	21	44	14	3	4	5	3	1	1
13	2	18	31	53	31	3	4	2	1	0	1
……											1
5	2	2	10	33	10	2	3	5	1	1	2
7	4	3	14	37	14	3	4	2	1	1	2
10	1	3	7	30	7	2	3	2	1	1	2
15	1	3	9	32	9	2	3	5	1	1	2
22	1	3	13	36	9	5	3	5	1	1	2
59	1	3	9	29	9	2	3	5	1	1	2
62	4	6	17	39	17	3	4	2	1	1	2
73	2	3	20	40	9	2	3	2	0	1	2
……											2

The clustering results were compared with the number of three-year principal review projects, and the results of the comparison showed that there was no significant correlation between the clustering results and the number of three-year principal review projects. We conducted a correlation analysis of the time of participation in work and age through the data statistical software, and found that there is a large correlation between these two variables, while the position and the time of participation in work alone cannot effectively explain their inevitable association with the amount of workload. In addition, the initial masses chosen randomly also greatly reduced the effect of clustering. The choice of variables with a large distribution of values for age is also fragmented, which can also have a negative impact on the effectiveness of clustering.

Using traditional analysis methods, HRM personnel usually need to create various tables of personnel distribution first, and then rely on personal feelings to measure various aspects, without the scientific nature of the data, and it takes at least 5-7 days to conduct an analysis. After using K-Means algorithm for clustering, HRM department personnel only need to use the data of HRIS for calculation, and in the case that the existing is not integrated into the system, including the time of data supplementation and processing, it only takes half a day to conduct the most complicated analysis, which is much more efficient.

In addition, the complexity of the traditional analysis method increases geometrically when the dimensionality of the analysis increases, and the simplicity and efficiency of the K-Means algorithm, with its low time and space complexity, can effectively overcome this problem and make the analysis much more efficient. Theoretically, K-Means algorithm can increase the dimensionality of analysis infinitely, but considering the effect of clustering, the actual use should not use too many dimensions and not produce too many classes.

4. Conclusions

In this paper, we first briefly introduce the basic principle of K-Means algorithm, analyze several aspects that need to be focused on, and verify through three experiments that the selection

of variables and the initial center of mass will have a greater impact on the clustering effect, and verify that the three variables of job level, age group, and title level have a greater impact on the number of completed audit projects, and when conducting human resource allocation, priority should be given to consider these three factors. The experiments also proved that the K-Means algorithm is reasonable and effective in conducting human resource allocation. The clustering results are then presented by using the construction of two-dimensional point plots to make the clustering results more intuitive. The main innovation of this chapter is that it realizes the practical application of K-Means algorithm in the data mining analysis of human resource allocation in audit department and proves the feasibility of mining analysis[9].

This paper provides directions and ideas for conducting HR data mining analysis by introducing some basic concepts and technical methods of HRM basic theories and data mining; the HR allocation analysis module and analysis process are designed through the analysis of HRMS functional modules; the analysis index system is constructed by using the original functions of HRMS, and the analysis index The data was exported to EXECL table for data organization, so as to prepare for data mining analysis; through several experiments, it was concluded that K-Means algorithm is feasible in human resource allocation analysis and can achieve the expected results, which provides data analysis tools to support the design of human resource allocation analysis module and the construction of analysis indicators. The main innovations of this paper are: the expansion of the original HRM system software functions and the innovation of K-Means algorithm in the application of human resource allocation work in audit department [10].

References

[1] Liu Guojun, Chen Shaoye, Wang Fengshuo. Library catalog [M]. Beijing: Higher Education Press, 1957. 15-18.

[2] Yang Jing. Research on human resource management of Gansu Lanye Construction Supervision Co., Ltd. in the context of big data [D]. Lanzhou University of Technology, 2021.

[3] Li Hui. The innovation path of human resource management in the digital economy[J]. Business Development Economy, 2022(04):12-15.

[4] Chen Ce. Exploring human resource management reform in the era of big data [J]. Modern rural science and technology, 2022(02):116-168.

[5] Yu Xiaomei. Research on the transformation of human resource sharing center of Group H in the context of artificial intelligence [D]. Qingdao University, 2021.

[6] Zhu Lixin. Exploring the application of information technology in human resource management [J]. China Market, 2022(05):193-194.

[7] Ou Xiangqing. Exploring the innovation of enterprise human resource performance management under big data [J]. Modernization of shopping malls, 2022(02):70-72.

[8] Xin Xia. Optimizing human resource management in institutions in the network economy [J]. Human resources, 2022(02):48-49.

[9] He Xuan-Sen, He Fan, Xu Li, Fan Yueping. Determination of the optimal number of clusters for the K-Means algorithm[J]. Journal of the University of Electronic Science and Technology, 2022, 51(06):904-912.

[10] Wang Chao, Wang Yongshun, Di Fan. Fast automatic fuzzy C-mean clustering algorithm for color image segmentation[J]. Advances in Laser and Optoelectronics,2022,59(22):82-88.