Optimal Allocation Model of Human Resource Redistribution on Account of Decision Tree Algorithm

Xian Jiang and Xuebin Liu(✉)

Department of Economic Management, Tianfu Information Vocational College, Meishan 620564, Sichuan, China
344601319@qq.com

Abstract. The process of human resource redistribution is often very complicated, the recommendation accuracy is not high, and it cannot meet the current application requirements of human resources. In view of the typical problems existing in human resources, it is an important problem to improve the accuracy of human resources redistribution. In order to solve the problems, such as the error of human resource recommendation, the recommendation accuracy is not good, the recommendation direction is wrong, and the quality of the recommendation content is not high, this paper puts forward a study on the optimal allocation model of human resource redistribution based on decision tree algorithm. This article studies the redistribution of human resources based on decision tree algorithm optimized configuration model research, this paper introduces the concept of the decision tree and related content, this paper expounds the redistribution of human resources optimization allocation model study of theory knowledge, referred to the redistribution of human resources based on decision tree algorithm optimization allocation model for the optimization of human resources redistribution effect. Through data testing, the results show that the optimal allocation model of human resource redistribution based on decision tree algorithm is better than other current human resource recommendation methods.

Keywords: Decision Tree Algorithm · Human Resources · Redistribution Optimization · Optimal Allocation Model

1 Introduction

Many scholars at home and abroad have studied decision tree algorithm. In foreign studies, Pandhi Asg proposed to develop a decision tree clinical algorithm, which combined simple neurological physical examination with simple cognitive assessment, to distinguish the elderly with MILD cognitive impairment from normal elderly in primary care [9]. NancyP proposed an intelligent fuzzy time decision tree algorithm, which can detect intruders effectively by combining the extended decision tree algorithm with convolutional neural network. Experimental analysis using KDD cup data set and network

N. Radojević et al. (Eds.): ICAID 2022, AHIS 7, pp. 206–213, 2023.
https://doi.org/10.2991/978-94-6463-010-7_22

tracking data set proves the effectiveness of this method [8]. FayazSA proposed a distributed decision tree algorithm based on historical geographic data of Kashmir and its implementation. The machine learning decision tree algorithm is applied to the Kashmir data set and the accuracy is 81.54%. Distributed decision tree generates multiple trees based on partitions of the original data set, in which data is isolated according to sub-stations (42026, 42027, 42044) [4]. However, the research on the optimal allocation model of human resource redistribution based on decision tree algorithm in China is still in the preliminary stage, and there is still a certain gap compared with that in foreign countries.

The research on the optimal allocation model of human resources redistribution based on decision tree algorithm needs to be further improved. The following points must be started: First, strengthen the research on decision tree algorithm; Secondly, improve the application efficiency of decision tree algorithm; Finally, strengthen communication and cooperation with foreign countries.

2 Design and Exploration of Optimal Allocation Model of Human Resource Redistribution on Account of Decision Tree Algorithm

2.1 Decision Tree Algorithm

Decisiontree (DT) refers to the model of algorithm. The principle of this algorithm is like a tree, which is constantly branching like a branch, connecting the trunk of the tree from the root to the branches and leaves. This principle is essentially a classification problem and a regression problem, consisting of nodes and line segment edges. In the operation principle of decision tree, node splitting is explained. Node splitting is an intermediate part of the process. Data is divided from top to bottom and then processed by neurons to form more than two child nodes. Node splitting is processed according to the optimal characteristics of the training data. Optimal feature is the selection of optimal feature value after classification of training data after full training, which is conducive to the improvement of the efficiency of decision tree and elimination of features without classification ability. Commonly used criteria include information gain, information gain ratio and Gini index. The process of decision tree algorithm is as follows:

Step1 for the split node, the training set has multiple features, all of which will divide the training set, so that the gini coefficient can be obtained through the training set according to the corresponding eigenvalues. The principle of selection is to select the minimum value of multiple eigenvalues, as shown in Fig. 1 [7].

Step2 continuously repeats Step1 to generate a decision tree. The condition for stopping repetition is to compare the non-pure pseudo-leaf node with parameter T, and stop until less than parameter T, as shown in the second and third layers of Fig. 1.

Step3 for impure pseudo-leaf nodes, the data classification model is produced, which is a data model MDT based on decision tree.

Step4 the algorithm ends, and the process is shown in Fig. 1.

As shown in Fig. 1, the basic principle of the decision tree algorithm is shown. Firstly, the first layer in the figure is taken as a split node, and the training model of this node all contains eigenvalues [4]. Then, the decision tree is divided into two parts according

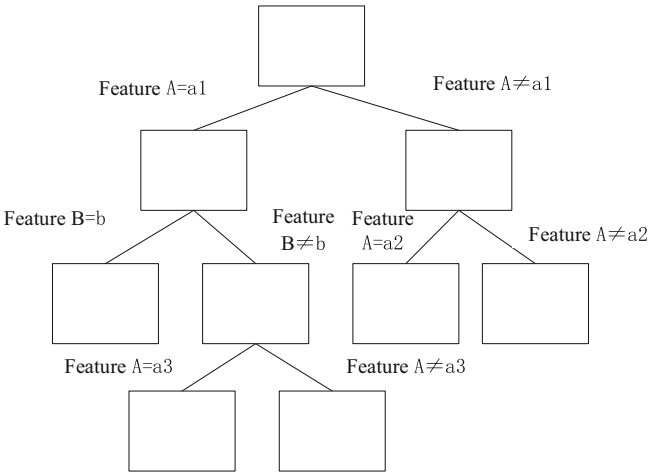

Fig. 1. Operation flow chart of decision tree algorithm

to the eigenvalues. When the decision tree is divided into two parts in the second layer, step1 is repeated until the constraint condition of parameter T is reached. The training set is divided according to the same principle of the third and fourth layers, and finally comes to the end of step4.

2.2 Research on Optimal Allocation Model of Human Resource Redistribution Based on Decision Tree Algorithm

2.2.1 Input Human Resource Data

The first thing that human resource system should do is to collect human data. Data collection is the basis of the optimal allocation model of human resources redistribution based on decision tree algorithm. Therefore, the quality of data is a very important aspect. This paper makes full use of decision tree algorithm to solve the major problems and challenges of the optimal allocation model of human resources redistribution [6].

2.2.2 Data Preprocessing

Human resource data preprocessing is the collection, processing and processing of the data generated by the optimal allocation model of human resource redistribution based on decision tree algorithm. The data contains various contents, such as the personal information of job seekers, job intention, job resources, and data of different industries. Make full use of these data to establish a human resource redistribution model, and then push effective results through the model. In practice, some data are continuous and some do not have continuity, and some data even produce bad results for the human resource redistribution model. Therefore, data preprocessing for the human resource redistribution model is needed. A series of data processing should be carried out for the obtained result set, such as eigenvalue selection, pseudo-data removal, data format conversion, data storage and so on [1].

2.2.3 Recommended Steps Based on Decision Tree Algorithm

Step1: Collect and preprocess the recommended data set of human resources.

Step2: Human resource data are divided into training samples and test samples according to certain rules.

Step3: construct the node N of the decision tree. If the data type is 1, it can be judged that the node is a leaf node. At the same time, if there is no attribute, the root node can also be judged as a leaf node.

Step4: according to the node attributes, carry out information gain, calculate and record the attribute value of information gain.

Step5: record the attribute of the root node as the highest information gain, and save the information.

Step6: give each node a new leaf node.

Step7: the splitting principle of decision tree is used to transform the recommendation algorithm.

Step8: remove the pseudo nodes in the decision tree to improve the overall efficiency of the decision tree.

Step9: Apply the recommendation principle and process the recommendation results of human resource data through the decision tree model.

3 Explore the Effect of Optimal Allocation Model of Human Resource Redistribution on Account of Decision Tree Algorithm

Suppose the dataset where, $D = \{(x_i, y_i)\}_{i=1}^{n}$ $x_i \in R^{\mathrm{m}}$, $\mathrm{y}_i \in \{1, 2, 3...K\}$ is the node's flag. The data set can be classified and temporarily divided into K category. It can be known that the probability of samples of K category is. Then the corresponding Gini index of the data set under this condition is:

$$\mathrm{Gini}(p) = \sum_{k=1}^{K} p_k(1 - p_k) = 1 - \sum_{\mathrm{k}=1}^{K} p_k^2 \tag{1}$$

Approximate estimation P_k is made for D in the process of operation, and the value is $\frac{|C_\mathrm{k}|}{|D|}$, which C_k means the number of k-th data in D, and the number of training samples. The Gini index corresponding to D is:

$$G\mathrm{ini(D)} = 1 - \sum_{k=1}^{K} \left(\frac{C_k}{D}\right)^2 \tag{2}$$

When the decision tree splits (K = 2), when the node splits, the feature A is obtained, and then the Gini index is calculated. D needs to judge whether a can be evaluated according to the eigenvalue, and is divided into two parts, the formula is as follows:

$$D_1 = \{(x_i, y_i) \in D | A(x_i) = a\}, \mathrm{D}_2 = \mathrm{D} \tag{3}$$

Therefore, the Gini index of the set under the condition of features is defined as:

$$G\mathrm{ini(D,A)} = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2) \tag{4}$$

*G*ini(D,A) is the uncertainty of D after node splitting under condition A = a. The value of the uncertainty affects the value of the training set. If the uncertainty is large, the uncertainty of the training set will be larger.

Detailed analysis:

First, the generation of incomplete decision tree. Splitting nodes is a process of data generation. Eigenvalues of the splitting nodes are collected, and gini index is calculated according to all eigenvalues. The principle is to select the corresponding optimal feature and splitting node from the minimum eigenvalue. According to the value, the decision tree is divided into two parts, which is a split of nodes [5]. The process of the decision tree algorithm is repeated until it is less than the impure pseudo node. The decision tree has many nodes, and each node can be evaluated, and each node must generate two parts. The operation of the decision tree continues until the condition no longer meets the data model condition.

Second, the linear and nonlinear problem of training data set. The given training data set may be linearly indivisible at the beginning, but since the decision tree has been sorted out in the early stage, the division of nodes can be approximately regarded as a linear relationship for the data processing of decision tree on nodes [10].

Thirdly, the weight of the scale T of the impure pseudo-leaf node. An important part of decision tree algorithm is nodes, so it is necessary to consider the node size T. If T is large, the corresponding decision tree will have shorter branches and fewer leaves. In this case, the classification effect of decision tree is often poor. When T value is very small, the corresponding decision leaves and branches are deeper, and there are more classification nodes, so the complexity of decision tree will be higher. Therefore, the size of the t-value is often important, which determines the style of the decision tree model [2].

4 Investigation and Analysis of Optimal Allocation Model of Human Resource Redistribution on Account of Decision Tree Algorithm

In this paper, 10 randomly selected data sets are processed by decision tree model, as shown in Table 1. The model was used to test the data for 100 times, and the mean value was obtained to record the results, while preventing various restrictive factors of the experiment. The data model adopts the current popular device configuration: CPU Core I5-121, 3.40 GHz, and memory 8.0 GB.

In Table 1, as shown in Table 1, the four items in The first row are The data set, Number of training samples, The Number of test samples and Feature dimension. The left column is the 10 test data sets that were tested. The data for all 10 test datasets are recorded in Table 1.

The research institute of optimal allocation model of Human Resource redistribution uses the decision tree algorithm to test the error rate, and its effect is shown in Table 2.

In Table 2, as shown in Table 2, the 2 items on the horizontal axis refer to the error rate and fault tolerance rate, and the 10 items on the vertical axis refer to the randomly selected 10 data sets of the study on the optimal allocation model of human resource redistribution based on the decision tree algorithm, respectively Image, Spambase, Magic

Table1. Decision tree dataset test table

The data set	Number of training samples	Number of test samples	Feature dimension
Image	5600	4500	28
Spambase	2300	1398	56
Magic Gamma Telescope	8790	9870	11
Madelon	560	1890	490
Banana	7890	980	8
Germen	1450	4320	14
Diabetis	2908	4570	7
Dota2DataSet	90345	12090	109
Credit Card Client	8700	17650	21
Cod_rna	9807	16700	43

Table 2. Error rate table of decision tree algorithm

The data set	Error rate*1000	Fault tolerance
Image	0.010*1000	94.21
Spambase	0.011*1000	95.42
Magic Gamma Telescope (MGT)	0.019*1000	95.64
Madelon	0.009*1000	97.51
Banana	0.012*1000	95.71
Germen	0.011*1000	92.71
Diabetis	0.013*1000	94.71
Dota2DataSet	0.021*1000	99.71
Credit Card Client	0.031*1000	98.12
Cod_rna	0.019*1000	96.23

Gamma Telescope (MGT), Madelon, Banana, Germen, Diabetis, Dota2DataSet, Credit Card Client, Cod_rna. Table 2 shows that the error rate of the study on the optimal allocation model of human resource redistribution based on decision tree algorithm is 0.010, 0.011, 0.019, 0.009, 0.012, 0.011, 0.013, 0.021, 0.031 and 0.019 respectively. The fault tolerance is 94.21, 95.42, 95.64, 97.51, 95.71, 92.71, 94.71, 99.71, 98.12 and 96.23.

Figure 2 is the data graph form 2, Fig. 2 horizontal error rates, and fault tolerance rate, said based on decision tree algorithm is the optimal allocation of human resources redistribution model research measure, ordinate 10 is a random sample of 10 based on decision tree algorithm is the optimal allocation of human resources redistribution

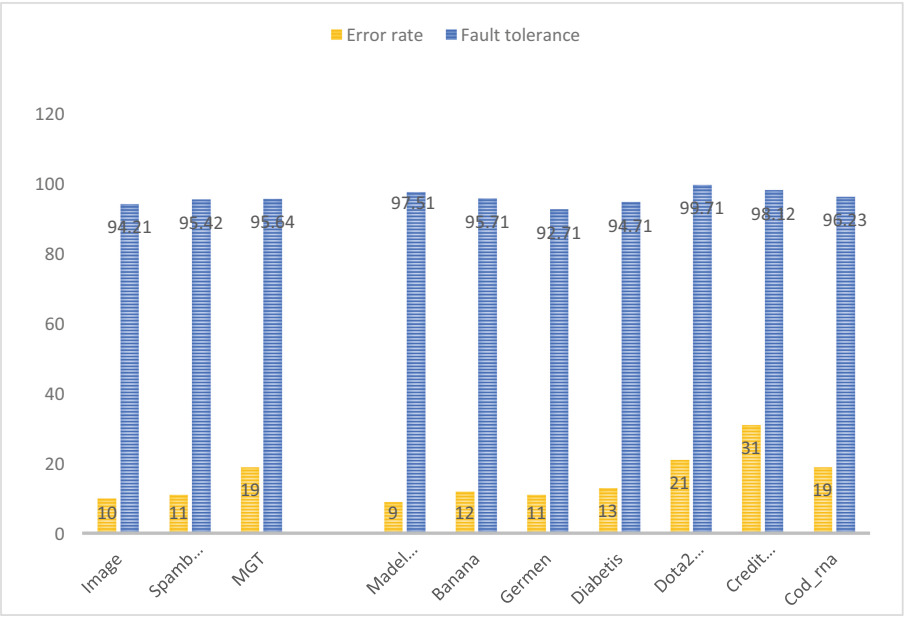

Fig. 2. Error rate diagram of decision tree algorithm

model research data sets, Image, Spambase, Magic Gamma Telescope (MGT), Madelon, Banana, Germen, Diabetis, Dota2DataSet, Credit Card Client, Cod_rna. Figure 2 shows that the error rates of test scores of the study on optimal allocation model of human resource redistribution based on decision tree algorithm are 0.010, 0.011, 0.019, 0.009, 0.012, 0.011, 0.013, 0.021, 0.031 and 0.019 respectively. The fault tolerance is 94.21, 95.42, 95.64, 97.51, 95.71, 92.71, 94.71, 99.71, 98.12 and 96.23.

From the above test data, compared with the optimal allocation model of human resource redistribution based on decision tree algorithm, the research is very excellent in error rate and error tolerance rate. The research on the optimal allocation model of human resource redistribution based on decision tree algorithm is bound to realize its great value in the field of human resource redistribution optimization.

5 Conclusions

This paper introduces an important decision tree algorithm, which firstly calculates the Gini coefficient. Before calculating this value, the eigenvalues of the decision tree are sorted out. By sorting out each eigenvalue, the smallest eigenvalue is selected and the corresponding node is selected. This node is the optimal node corresponding to the minimum eigenvalue [3]. The decision tree is usually split and divided into two parts of the subsystem at this optimal node. The process is then repeated until the limit is reached. The research on the optimal allocation model of human resources redistribution based on decision tree algorithm in China has made great efforts in the research and development of decision tree algorithm, and has made special breakthroughs in the problems and

obstacles encountered, which provides a set of reference examples for the optimization of human resources allocation.

References

1. Akrout B (2021) A new structure of decision tree based on oriented edges gradient map for circles detection and the analysis of nano-particles. Micron 145(5):103055
2. Aliyan E, Aghamohammadi M, Kia M, Heidari A, Shafie-khah M, Catalão JPS (2020) Decision tree analysis to identify harmful contingencies and estimate blackout indices for predicting system vulnerability. Electr Power Syst Res 178:106036.1-106036.8
3. Benton DC, Ferguson SL, Holloway A (2020) Task shifting: A High-level analysis of scholarship. J Nurs Regul 11(2):4–11
4. Fayaz SA, Zaman M, Butt MA (2021) To ameliorate classification accuracy using ensemble distributed decision tree (DDT) vote approach: an empirical discourse of geographical data mining. Procedia Computer Science 184(1):935–940
5. Ghiasi MM, Zendehboudi S, Mohsenipour AA (2020) Decision tree-based diagnosis of coronary artery disease: CART model. Comput Meth Program Biomed 192(6):105400
6. Jumin E, Basaruddin FB, Yusoff Y et al (2021) Solar radiation prediction using boosted decision tree regression model: a case study in Malaysia. Environ Sci Pollut Res 28(21):26571–26583
7. Maulana A (2021) Prediction of student graduation accuracy using decision tree with application of genetic algorithms. IOP Conf Ser Mater Sci Eng 1073(1):012055
8. Nancy P, Muthurajkumar S, Ganapathy S et al (2020) Intrusion detection using dynamic feature selection and fuzzy temporal decision tree classification for wireless sensor networks. IET Commun 14(5):888–895
9. Pandhitas G, Sutrisna B, Wibowo S et al (2020) Decision tree clinical algorithm for screening of mild cognitive impairment in the elderly in primary health care: Development, test of accuracy, and time-effectiveness analysis. Neuroepidemiology 54(3):1–8
10. Zarkami R, Hesami H, Sadeghi R (2020) Prediction of the abundance of artemia parthenogenetica in a hypersaline wetland using decision tree model. Wetlands 40(6):1967–1979. https://doi.org/10.1007/s13157-020-01332-2