Intelligent Computing Technology and Automation Z. Hou (Ed.) © 2024 The Authors. This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0). doi:10.3233/ATDE231249

A Model of Talent Management in Universities Based on Decision Tree and Correlation Analysis

Lisha WU^{a,1}, Yanbin LIU^a

^a Chongqing Three Gorges Vocational College, Wanzhou, 404100, China

Abstract. Traditional human resource evaluation methods rely too much on people's subjective judgment and experience. In order to overcome this defect, a talent management model based on decision tree and correlation analysis is proposed. The scheme uses the concept of decision tree, classifier and correlation analysis to preprocess the primary data. Then Apriori association rules and C4.5 algorithm are adopted to complete the combination prediction of the two, which reduces the efficiency of the algorithm to a certain extent when searching for approximate high reliability rules, and makes the improved decision tree algorithm have good scalability. Finally, the model is tested and analyzed through the actual data in human resource management in universities. The results show that the scheme can improve the prediction accuracy of the algorithm and provide a theoretical basis for scientific decision-making of leaders and functional departments.

Keywords. decision tree; correlation analysis; talent management; data mining; C4.5

1. Introduction

Human resource management is an important part of university management. It has been paid more and more attention in various fields. The important content of personnel management in colleges and universities is talent prediction and evaluation, which involves the factors such as the total number of talents, the education background of talents, the professional distribution of talents, the age of talents and the level of talents. Its main contents include: prediction of professional structure of talents: prediction of hierarchical structure of talents; The age structure prediction of talents; Prediction of the educational structure of talents; Prediction of the total demand for talents, etc. Human resource management faces massive data. There is an urgent need for a technology to find valuable knowledge, so as to formulate talent development strategies and provide decision support for them. The traditional talent identification method is realized by examining the qualitative indicators of the introduced personnel. It is difficult to ensure the comprehensiveness, scientificity and accuracy of the identification results by relying on manual t-operation. In view of the fact that many colleges and universities have established talent database systems, the application of data mining technology to analyze the previous talent data will help to improve the accuracy of talent identification and improve the identification efficiency [1]. Decision tree is a method of data classification and mining based on information theory. The

basic idea is to build a decision tree through a batch of known training data, and then use the built decision tree to predict the data. The establishment process of decision tree can be regarded as the generation process of data rules. Therefore, it can be considered that decision tree realizes the visualization of data rules, and its output results are easy to understand. The decision tree method has high accuracy, easy to understand results and high efficiency. Therefore, decision tree technology has the characteristics of data analysis efficiency, understandability and usability. Thus, applying decision tree technology to human market evaluation will make our work more effective, enable us to reasonably identify, employ, manage and retain people, and reduce the loss of human, material and financial resources [2].

This paper focuses on the application of decision tree algorithm in human resource management. Firstly, the related basic knowledge, namely data mining, common decision tree algorithm and association rule analysis, is explained. Then according to the real samples and requirements in practice, a decision tree algorithm model based on association rules is proposed, in which the management rule algorithm adopts Apriori and the decision tree algorithm is C4.5. It is preferred to use the data mining results to guide the various stages of human resource management, and then apply C4.5 to further sort the preprocessed standardized data to get the final human resource allocation scheme. The case analysis is an example of talent allocation in a university. The improved decision tree technology is used to analyze and process the existing data. The results show that our scheme can well predict the development of human resource management and provide theoretical reference for decision makers.

2. Principle Knowledge

2.1 Principles of Data Mining

Model is an abstract description of objects in the real world, and it is an indispensable part of human life; Model is very important for all fields of life, especially for data mining. Generally speaking, the model used in data mining is the structure of human knowledge. Model is an abstract description of objects in the real world, and it is an indispensable part of human life; Model is very important for all fields of life, especially for data mining. The model used in data mining is the structure of human knowledge, and its basic process is shown in Figure 1.



Figure 1. The process of data mining

2.2 Decision Tree Algorithm

Generally, a decision tree contains a root node, several internal nodes and several leaf nodes, so the decision tree is equivalent to a multi branch tree. Leaf nodes correspond to decision results, and each other node corresponds to an attribute test. The sample set contained in each node is divided into child nodes according to the result of attribute test When building the analysis model, first define the amount of knowledge information as I and moisture as E, and the calculation formula of the two is:

$$\begin{cases} I = \log_2 P i \\ E = \sum \log_2 (P i) \end{cases}$$
(1)

Supposing that a data set S contains x conditional attributes, of which a classification attribute is k, and there are n different discrete attribute values $k_1, k_1, ..., k_n$ in the k set. The records in the data set s can be divided according to the classification attribute k, and n different subsets can be obtained. If the number of all records in data set S is known and the classification attribute is known, the total entropy of data set s before division should be

$$E(s_1, s_2, ..., s_n) = -\sum_{i=1}^n P_i \log_2(P_i)$$
⁽²⁾

When establishing the decision tree analysis model, we first need to obtain and sort out the main sample data in human resources and management decision analysis, calculate the expected value of decision attributes and the information entropy of various elements of human resources, and output the analysis results after obtaining the human resources information gain results, so as to complete the human resources and university management decision analysis [3,4].

2.3 Association Analysis

Association analysis is the process of discovering the hidden relationships and laws between objects from large-scale data, also known as association rule learning. It is used to find the relationship between items in the dataset. According to the mining association relationship, the information of one attribute can be inferred from the information of another attribute. When the confidence reaches a certain threshold, the rule can be considered to be true. Association analysis includes two processes: finding frequent itemsets from data sets and generating association rules from frequent itemsets. The common association analysis algorithms are listed in the table 1.

Table 1. Common association analysis algorithms	
---	--

Algorithm	Description
Apriori	Apriori uses the iterative method of layer by layer search to find the relationship between item sets in the database to form rules. Its process consists of connection and pruning. The concept of itemset in this algorithm is the set of items.
FP-Tree	FP Tree algorithm only scans the database twice. Compared with Apriori algorithm, it has no candidate set, and directly compresses the database into a frequent pattern tree, through which association rules are generated.

Eclat	Eclat algorithm adopts the method of vertical data representation, and stores the data according to the item set. In this way, the support of the k+1 order itemset can be obtained directly from the set operation of the transaction identifiers of its two k-order subsets
Grey correlation	The significance of grey correlation degree method refers to that in the process of system development, if the change situation of the two factors is consistent, that is, the degree of synchronous change is high, then the two factors can be considered to be highly correlated; On the contrary, the correlation between the two is small.

L. Wu and Y. Liu / A Model of Talent Management in Universities

3. Talent Management Model Based on Decision Tree and Correlation Analysis

3.1 Decision Trree Based on Association Rules

In C4.5, pessimistic pruning will be performed when the decision tree is constructed to improve the generalization ability of the decision tree. Since scholars have proposed improvements from many aspects. This paper attempts to improve its classification performance through the combination of decision tree algorithm and association rules, and then gives specific methods in their links, feature extraction and data transmission, to fit the training data perfectly and better adapt to the standardized data set. The main flow of the algorithm is shown in Figure 2.



Figure 2. Decision tree based on association rules

3.2 Data Processing

To apply these numerical attribute values to association rules, it is discretized by classification decision tree Using the principle of classification decision tree, attribute values are divided into leaf nodes, and data induction is carried out from top to bottom. Set several discrete dividing points within the value range, divide the value range into some discrete intervals, and finally use different symbols or integer values to represent the data values in each sub interval to form a task related basic table Then analyze and generalize each attribute, set a generalization threshold for all attributes or set a threshold for each attribute, and construct a classification sample model. At this time, the data sample model is an effective data set that has been compressed or generalized,

as shown in table 2. When building a decision tree, continuous attributes can be used many times, that is, they can be used as the partition attributes of subsequent nodes.

 Table 2. The normalized dataset after processing

NO.	Name	Title	Sex	Education	Age≤30
5	Li Da	Junior	Female	master	Y
8	Zhang Xun	Junior	Male	undergraduate	Ν
12	Si Biyi	Senior	Male	master	Y
21	Han Yingying	Junior	Female	undergraduate	Y
39	Yan Shali	Senior	Female	undergraduate	Ν
77	Dong Yong	Senior	Male	master	Ν

3.3 New Attributes Generation Based on Association Rules

On the basis of Apriori algorithm, the improvement degree parameter of rules is used as the measurement standard of importance [5]. Once the frequent itemsets are found by the transactions in database d, it is straightforward to generate strong association rules from them. For confidence, the medium conditional probability is expressed by the count of item set support. For association rule a=>b, the importance calculation formula is computed as

$$\operatorname{im\,portance}_{A \Rightarrow B} = Log_2 \frac{p(B \mid A)}{P(B \mid nolA)}$$
(3)

When the result is 0, there is no correlation between A and B; When the result is greater than 0, a and B are positively correlated; If the result is less than 0, a and B are negatively correlated. Calculate the importance of each attribute, and sort the results by size, as shown in Figure 3.



Figure 3. Final decision tree

4. Experimental Analysis

The universities are gradually strengthening human resource management while trying to improve teaching quality and promote the all-round development of students under the educational concept of cultivating people with virtue. Select all the staff of a university as the research object of this paper. Technical structure: there are 1735 professional and technical personnel, including 358 with senior professional titles; There are 956 qualified teachers. Perform human resource analysis and evaluation of enterprise employees within the school, and form 8000 basic data. In the way of random selection, the relevant data information of 500 employees is extracted as the number of leaf nodes, and the analysis work is carried out by using the analysis method in this paper and the traditional analysis method.

Take the representative data in the data set as the input training set for model creation, which includes both the number of employees and the number of modeling variables. However, a small example helps to illustrate its internal operation. The main goal of the decision tree is to divide these employees into subgroups with "purest" category separation. This means that its leaf nodes should mainly include employees whose target variable turnover is either all "yes" or all "no". Employees are divided into positive examples and negative examples: resigned employees (P) and non resigned employees (N) [6]. At initial time the number of instances of class P and N are 6 and 9,

then $P(u_1) = \frac{6}{15}$ and $P(u_2) = \frac{9}{15}$. According to the information entropy equation of decision, the information entropy of given sample is $I(6,9) = \sum_{i=1}^{2} P_i \log_2^{P_i} = 0.945 bits$.

Select attribute gender as the root node and lead out two branches. At this time, the set is divided into two subsets to generate a decision tree containing two leaf nodes. Then the information gain is:

Gain (title) =
$$0.598 - 0.281 = 0.317bits$$

Gain (education) = $0.598 - 0.453 = 0.145bits$
Gain (age) = $0.598 - 0.98 = 0.000bits$

Prune while constructing the decision tree. All decision tree construction methods will stop the process of creating branches only when the entropy cannot be further reduced. In order to avoid over fitting, a threshold can be set, and the amount of entropy reduction is less than this threshold. Even if the entropy can still be reduced, the process of creating branches can be stopped. Use two-thirds of the data to construct a decision tree, and then use this decision tree to classify the test set. The accuracy obtained is the accuracy of the decision tree.

In this module, the personnel files of employees are mainly added, modified, deleted and printed. Technically, it adopts the currently popular Java technology framework, mainly including: springboot, Shiro, mybatis plus, Druid, redis, mysql, thymeleaf and other technology framework integration. Use ide to import this project. Pats of the key codes of this process are described as follows:

import dao.CourseDao; import java.io.IOException; import java.io.PrintWriter; import javax.servlet.ServletException; import javax.servlet.http.HttpServlet; import javax.servlet.http.HttpServletRequest; import javax.servlet.http.HttpServletResponse;

```
public class CourseServlet extends HttpServlet {
  public void doGet(HttpServletRequest request, HttpServletResponse response)
  throws ServletException, IOException {
    doPost(request, response);
  }
}
```

In terms of functional design, the HR system is divided into five parts: scientific research business management, scientific research portal, scientific research evaluation and decision-making, promotion center and support platform. In terms of business structure, the system is based on eight core business systems, including organizational structure, personnel management, teacher training, salary and welfare, human resource planning, talent recruitment, post employment and performance appraisal, and integrates financial, scientific research, teaching and other business systems to serve the staff, leaders, secondary units and relevant functional departments of the University. Configure relevant parameters according to the above requirements, and the final result of software test is shown in Figure 4. It can be seen that the human resources of the school have been well configured, and the catalogue is standardized and accurate, which can reflect the characteristics of various talents and facilitate viewing and searching.



Figure 4. Final generation interface by software test

5. Conclusion

. . .

This paper combines association rules with decision tree algorithm to form a new classification algorithm The scheme first finds all frequency sets, and the frequency of these item sets is at least the same as the predefined minimum support. At the same time, it can improve the efficiency of decision tree learning. The result of classification using a feature is not very different from that of random classification. In the case analysis, the program is used in the talent management of colleges and universities for test and analysis. The results show that this research method not only emphasizes the objectivity of dealing with problems, but also makes full use of the experience of practical workers. Therefore, the results of the model have strong practical value.

Acknowledgments

This work was supported by Shandong Research Project of teaching reform in undergraduate colleges and universities in 2018, under grant no.M2018X033.

References

- [1] Hu Yongxiang. Application of decision tree classification in personnel management in colleges and universities. Golden Mountain, 2011, 1:60-61
- [2] Sun H, Yi-Jie L I, Qing-Jiang LI. A Research of Human Resource Value Based on Decision Tree. Agricultural Equipment & Vehicle Engineering, 2007, 4:56-58
- [3] Huang Chen, Liu Xuefei. Human resource and enterprise management analysis of decision tree algorithm. Development of human resources, 2021, 16: 94-96
- [4] Vaiman V, Scullion H, Collings D. Talent management decision making. Management Decision, 2012, 50(5):925-941
- [5] Brown S D, Myles A J. Decision Tree Modeling ScienceDirect. Comprehensive Chemometrics (Second Edition), 2020, 1: 625-659
- [6] Tang L, Kim C Y, Martin J G, et al. Length of Stay Predicts Risk of Early Infection for Hospitalized Patients Undergoing Central Venous Port Placement. Journal of Vascular and Interventional Radiology, 2020, 31(3): 454-461