Research and Implementation of Networked Student Management Information System in Colleges and Universities Based on Decision tree Algorithm

Huimin Liu, Wenjing Yu

School of Marxism, Beijing Jiaotong University, Beijing, China

Abstract. Decision tree is one of the special methods in digital mining. Using decision tree algorithm, the relationship between model and digital can be found in massive digital, and dynamic prediction can be made accordingly. Find out the deficiencies in student information management, find valuable hidden information in the huge digital related to student information, analyze it, obtain the useful information needed, and summarize its useful rules for promoting the management and development of colleges and universities, As the basis for analysis and resolution-making and research on the direction of reform, a large amount of digital and information can be effectively used. This paper uses the decision tree algorithm commonly used in machine learning to learn knowledge from the existing student information management system, explores the knowledge information between digital, extracts useful rules from it, uses these rules to find the key factors affecting information management, and optimizes students reasonably. Information management work, so as to improve the level of education management.

Keywords: Decision tree; Algorithm; Network management; Information system.

1. Introduction

In recent years, digital mining is considered to be one of the effective ways to solve the problem of rich digital but poor information. In short, it is the process of discovering hidden, unknown and valuable knowledge from massive digital [1]. Data mining has obvious advantages in analyzing massive digital. At present, it has been applied in the field of education. It is of great practical significance to apply digital mining technology to scientifically and deeply analyze student information digital, extract valuable information from these digital, and provide a reliable basis for the resolution-making of college managers [2]. By improving the level of student management, it is of great practical significance to further improve the level of running a higher vocational college and the quality of student training [3].

As a method of digital mining technology, decision tree algorithm has the advantages of high accuracy, fast classification speed, simple description, easy to understand and so on. This paper adopts C4.5 algorithm and uses Python as the development language. Python has clear syntax structure and rich digital types, such as lists, tuples, dictionaries, sets, queues, etc. the operation of these digital types is simple [4]. Add an intelligent service layer containing various machine learning algorithms, input the cleaned digital into the intelligent service layer, select different machine learning algorithms through different condition judgments, and output various rule information or different predictions and verifications, so as to make the original system more "smart" [5]. The purpose of this paper is to use the decision tree algorithm in digital mining to extract the digital characteristics and relationships in the student information system, and combine the digital analysis to form a reference student personalized information, which provides valuable digital support for education management, so as to effectively improve the quality and efficiency of education and management [6].

2. Decision tree algorithm

Decision tree is a basic classification and regression method. It is a tree structure, in which the internal nodes represent the features or attributes in the test digital set, and the leaf nodes are a resolution result. When the decision tree is used to deal with classification problems, the process is to classify digital sets by features or attributes, and the results can be regarded as a set of classification

rules summarized, or as an estimated conditional probability model. Decision tree algorithm usually includes three steps: feature selection; Generation of decision tree and pruning of decision tree, in which the generation of decision tree is the key point. In this paper, the basic classic ID3 algorithm is selected to study the student information management system, and other algorithms will be tried later [7].

Among all digital mining methods, the improved C4.5 decision tree algorithm uses different attribute values to divide the training set into different subsets, and the number of divided subsets is equal to the number of attribute values. In the process of decision tree generation, branches correspond to subsets one by one, leaf nodes represent the end point of each branch, and resolution rules are path rules from root nodes to leaf nodes [8]. The scale of the decision tree is too large, which is caused by too many branches and nodes, too many redundant rules and low classification accuracy in the decision tree. It is caused by the increase of resolution rules in the decision tree, and all these problems are caused by too many attribute values. The following improvement ideas are put forward to solve this problem: after calculating the information entropy of attribute values, if you want to show that the information carried by two attribute values is similar, then the attribute should contain the information entropy of two attribute values that are similar [9].

2.1 Construction of decision tree

Specific steps: find out the key features in the current digital set, divide the current digital set according to the features, get different digital subsets, and do the same processing on the digital subsets again until all the same types of digital are in one The specific process is shown in Figure 1.

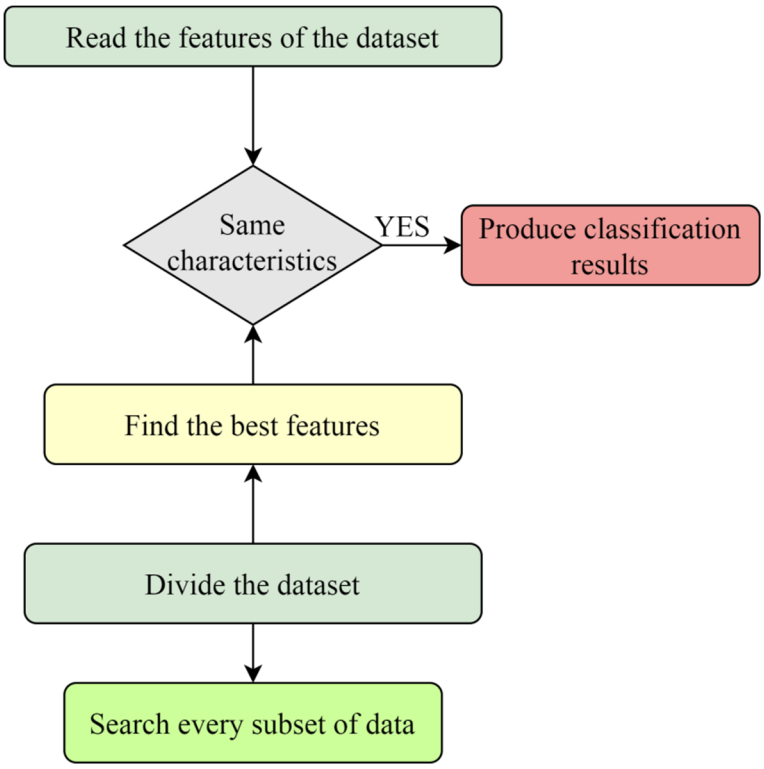

Figure 1. Constructing decision tree algorithm

2.2 Information Gain

As can be seen from the above algorithm, how to measure the best features to divide the digitalset is a key issue. The ID3 algorithm refers to the change of information before and after dividing the digital set as information gain. By calculating the information gain, you can understand the information gain obtained by dividing the digital set by each feature value, and comparing the feature with the highest information gain is the best choice at present. According to the feature with the

largest information gain, the current digital set can be divided, this feature becomes a branch on the spanning tree, and so on, a decision tree can be constructed [10]. Calculate the information gain rate of the attribute and the information entropy of each attribute value in the attribute; compare the attribute value of each attribute to determine whether there is an attribute value pair whose information entropy is within the threshold range, and compare the cosine of the two attribute value pairs. The similarity value is calculated. If it is greater than the threshold value of 0.9, it indicates that the similarity of the two vectors is very high; the new attribute value vector can be obtained by combining the two attribute value vectors, and the new subset and new attribute value are represented by a new vector. The new attribute is composed of adding a new attribute value after deleting the original attribute value in the attribute that participated in the comparison.

2.3 General application process of decision tree

The application process of decision tree is as follows: (1) Collect digital: it can be in any way, such as from the digitalbase. (2) Clean the digital: Discrete the digital into numerical digital to generate the corresponding training digital set. (3) Training digital: The generated training digital set is used to construct a decision tree according to the above method. (4) Test algorithm: Use the test digital set to estimate the error rate (applicable to supervised learning algorithms).

According to this process, the decision tree model is established as shown in Figure 2:

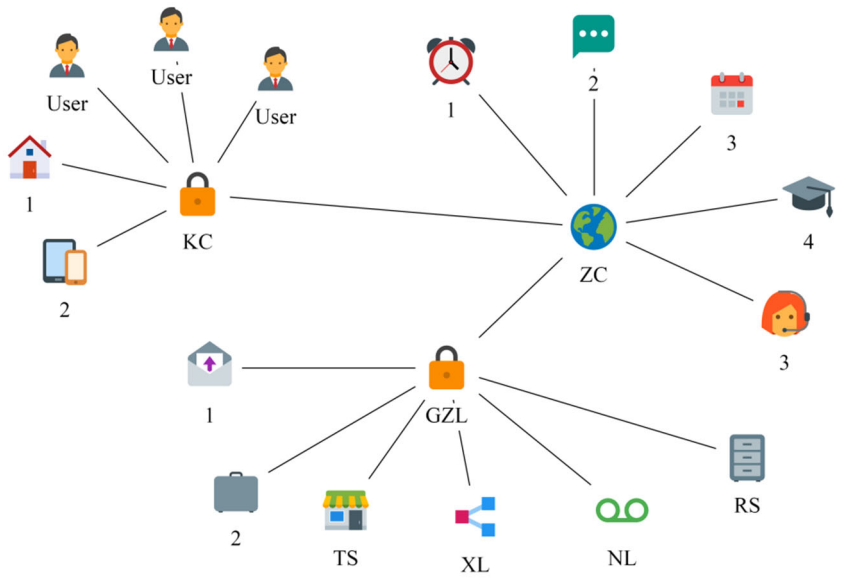

Figure 2. Decision tree model

3. Application of networked student information management system

3.1 Student information management system requirements

In the new era, the information management system of college students is required to conform to the characteristics of college management. Besides improving the management efficiency, it is also required to extract effective information to assist management resolutions. Information management of students at school: This is a common concern in information management of colleges and universities. Recruitment companies want to learn about college students' performance in school through the internet, and students themselves pay attention to the submission and modification of information, which leads to frequent system clicks. Traditional college students' information management methods have some disadvantages, such as low efficiency, slow update, inconvenient query, etc. Therefore, it is inevitable for modern enterprises to develop and research a personnel management system with convenient operation. Student achievement information management: the traditional paper management is inefficient, and the searching method is inflexible. It often takes a

lot of time to look up the names of thousands of people to find a certain grade of a certain class student. While the student information management system is convenient for inputting and counting students' related digital, it should also enable students to make horizontal and vertical inquiries about their scores and total scores of various subjects at any time. Graduate tracking management: it can track and manage the graduates' whereabouts in our school, and save whether they are employed or not, so as to know the basic situation of graduates in the enterprise and facilitate the analysis of the relationship between graduates' specialty and employment; Or inquire about the rewards and punishments between schools, so that we can objectively evaluate each graduate on these attributes and discover their inherent laws. It is required that the system can conveniently query the basic information of each graduate and the results of each semester in school at any time, so as to provide resolution-making guidance digital management for students' graduation employment and education mode.

3.2 Application of decision tree algorithm in student comprehensive assessment

Decision trees are the most popular data mining technique because of their fast training performance, high accuracy, and easy-to-understand patterns. It first processes the data, uses an inductive algorithm to generate readable rules and decision trees, and then uses the decision tree to analyze the new data. Essentially a decision tree is the process of classifying data through a series of rules. In this paper, the decision tree classification theory is used to construct a decision tree for students' performance evaluation to classify their comprehensive quality. The use of the model stage is to use the classification model formed by the decision tree to classify or predict new samples. This paper uses the previously established performance evaluation model to classify the comprehensive quality of the student data of a university's information management department, as shown in Figure 3.

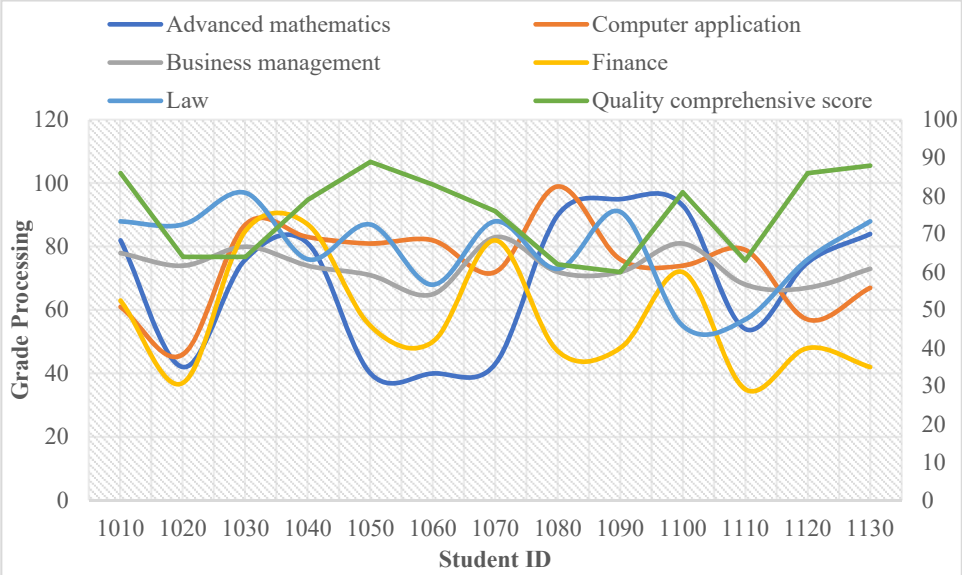

Figure 3. Data set processing

In the data selection and processing stage, students' basic information data, examination score data and comprehensive quality evaluation data from the digital campus digital platform are used, and the basic information data is from the basic information table. Students' professional attributes can be obtained by associating the class code in the basic information table with the unit information table. The comprehensive quality evaluation number comes from the comprehensive quality evaluation field in the student graduation information table. The digitalization of students' comprehensive quality evaluation is maintained by the students' comprehensive quality evaluation management system under the digital campus platform. After the number is selected, the basic information number, examination score number and comprehensive quality evaluation number are integrated through the

student ID to form the original number set. In order to ensure digital quality and obtain instructive digital mining results, digital preprocessing is needed, that is, filtering the numbers that do not meet the requirements, such as irrelevant numbers, noisy numbers, incomplete numbers, etc. After digital preprocessing, the attributes of the digital set include politics, major type, school performance, employment, English level, basic course performance, professional course performance and comprehensive quality evaluation results. Among them, the political outlook, professional type, school performance, employment, English level, basic course scores and professional course scores are the examination attributes, and the comprehensive quality evaluation results are the classification attributes.

Through the mining of the achievement decision tree model, we found that there are 75. 06% of the students have excellent overall quality, 22. 59% of the students have a good overall quality, a total of 97.65%; 35.61% of the students who pass the subject have an excellent overall quality, 38.22% of the students have a good overall quality, a total of 73.83%. Therefore, we can predict that students with good academic performance should also have better overall quality, and the accuracy of analyzing the data set with this method is increased by 6.19%.

3.3 Overall frame design of information management system

The application of data mining technology based on decision tree algorithm in student comprehensive evaluation system realizes a more scientific student comprehensive evaluation system adapted to the times. It is helpful for student managers to recommend graduates and employers to select talents. It is also helpful for leaders to assess classes and head teachers to guide students. By using clustering and decision tree tools to analyze students' grades, this paper shows that it is feasible and necessary to introduce data mining technology into the field of education. If we popularize this idea, we can apply it to all aspects of the school, such as classroom teaching evaluation, analysis of students' course selection, analysis of the use of school equipment, students' employment decision-making and so on, so as to make the management and decision-making of the school "based on evidence". As far as schools are concerned, the model of batch data prediction can be used to carry out targeted employment guidance courses for different groups of people, which is of great significance in the current situation of extremely scarce educational resources. At the same time, students' employment situation can be observed as a whole according to the results of employment prediction, and courses that greatly promote employment can be carried out in time according to the distribution of majors in schools, so as to reduce the courses that are no longer applicable at present. Thus, a student information management system with perfect functions, convenient use and data sharing is formed. It can provide comprehensive academic management and employment guidance management based on data mining technology, so as to meet the requirements of the current environment for student information management.

4. Conclusions

The advantages of decision tree learning are that the computational complexity is not high, the output results are easy to understand, it is not sensitive to the absence of intermediate values, and it can handle irrelevant digital. The disadvantage is that it may cause over-matching problems. This paper starts from the theoretical basis of the system platform, and through the design pattern, studies the program structure that is easy to maintain and expand with high cohesion and low coupling among the functional modules in the network student information management system, and uses the decision tree model to realize the digitalbase design. According to the requirements and functions of the system, an in-depth analysis is carried out, focusing on the detailed introduction of the design and implementation of the system. Through the application of WEB development technology and design patterns and other related technologies, the system has high security, reliability and scalability. It is easy to maintain and upgrade, and supports massive digital processing and a large number of concurrent users.

References

[1] Zhong Min. A Preliminary Study on the Application of Resolution Tree Algorithm in University education Information System [J]. Journal of Wuhan Engineering Vocational and Technical College, 2021, 33(2):3.

[2] Hou Haixia. Application of digital mining based on decision tree algorithm in student information system [J]. Journal of Changzhou Institute of Technology, 2019, 25(4):5.

[3] Li Ying. Research on Information System Data Mining Based on Resolution Tree Algorithm [J]. Information Technology, 2022, 46(2):6.

[4] Wu Shanfeng, Lu Xia. Design of sports curriculum analysis and management system based on decision tree algorithm [J]. Modern Electronic Technology, 2019, 42(3):4.

[5] Zhou Dan, Shen Lidi. Research on the application of C4.5 decision tree construction algorithm [J]. China High-tech Enterprises, 2018(4):2.

[6] Kuang Jihong. Research on the application of digital mining in the performance analysis of educational administration system [J]. Jilin University, 2017(3):10.

[7] Wang Zhiruo, Hu Yunqin. Analysis and application of decision tree C4.5 algorithm in student achievement management system [J]. Electronic Production, 2019(8Z):1.

[8] Wei Jinri, Li Xueping. Application of decision tree algorithm based on K-nearest neighbor similarity in student employment management [J]. Coastal Enterprises and Technology, 2019(4):12.

[9] Wang Lihong. Application of decision tree algorithm in student achievement analysis [D]. Northwest Normal University, 2018(6):11.

[10] Qi Yunfei, Li Yongqiang. Application of decision tree digital mining algorithm in student evaluation analysis [J]. Journal of Henan Vocational and Technical College, 2021, 23(5):3.