

Application in college practice teaching based on decision tree algorithm

Qun Liang

Nanning University

dahai2079@163.com

Project Name and No.: Research on Practical Teaching Mode Based on Innovation Ability Training-Taking Engineering Major as an example (2022XJJG44)

Abstract. When Chinese colleges and universities use information technology to improve management level, the internal system will set a variety of databases, such as employment information, educational administration information and so on. Although the system accumulates a large amount of data and information when it runs for a long time, the lack of information awareness and science and technology leads to the fact that managers can only obtain the surface information through simple statistical operations and cannot dig the hidden value. This paper mainly discusses decision tree algorithm and association rules, clarifies the application demand and teaching evaluation system of university practice teaching in the new era, and finally studies the application effect of decision tree algorithm in university practice teaching combined with practical cases.

Keywords: decision tree; Association rules; Efficient teaching; Teaching evaluation; Data mining.

1. Introduction

Nowadays, the social development of using information technology to produce and search data is increasingly high, a large number of databases are widely used in engineering development, scientific research, government office, commercial management and other fields, practice development has emerged more opportunities and challenges. Especially in the context of the era of big data, in the face of the explosive growth of information data, excessive information has become the main problem for researchers to discuss. How to find valuable content from it and improve the application efficiency of information is also an important topic for scholars around the world to study. [1-3]In the face of the challenge that people are blinded by data, but people are hungry and the development of knowledge, data mining technology emerges at the historic moment, and shows a strong vitality in the development of The Times. After the country's economic construction enters the brand new stage of socialist modernization, high quality higher education meets the challenge and opportunity of the new era. As the core task of higher education is teaching work, how to improve teaching quality is the key factor to promote the reform and development of colleges and universities, so the construction of efficient and reasonable education quality evaluation system is the main problem discussed by researchers. In foreign countries, a number of theories have been obtained from the actual teaching quality management activities, while in China, teaching level evaluation indexes of ordinary colleges and universities of the Ministry of Education are mainly studied, so as to build a personalized quality evaluation system.

Nowadays, Chinese colleges and universities have realized the importance of teaching evaluation and regard teaching quality evaluation as the basic task of education management. From the perspective of practical education management, the comprehensive evaluation of the teaching quality of college teachers mainly starts from four aspects: the first is the evaluation of students, the second is the evaluation of experts, the third is the peer evaluation, and the last is the self-evaluation of teachers, so as to obtain the final evaluation results. Many universities will apply the data mining technology in the university practice teaching evaluation system, integrate the data analysis and mining environment into the teaching evaluation management system, and form a management system with management, analysis, mining and other functions. However, there are still many problems in the teaching quality evaluation system during the application research. Finally, it

directly affects the mining effect of potential knowledge and the expression form of quality evaluation function. Nowadays, in the absence of scientific and technological methods and application means, facing the problems emerging in the evaluation system of time teaching quality in colleges and universities, how to improve the justice and safety of the final evaluation results has become the main issue discussed by scientific researchers. Data mining technology can extract or mine more valuable content from a large amount of data. Therefore, based on the understanding of the development status of practical education in colleges and universities, this paper, based on the research results of teaching quality evaluation at home and abroad, deeply discusses the evaluation results of efficient practical teaching quality with decision tree algorithm as the core, and defines the relationship between various indicators. And its influence on the composition of final evaluation results.

2. Methods

2.1 Data Mining

This concept refers to the extraction of interesting data knowledge from large databases, which is implicit and unknown, and mostly in the form of patterns, rules, laws, concepts, etc. In the development of modern technology innovation, data mining can be applied in any type of information storage, such as data warehouse, database system, Internet data, relational database and so on. According to the data mining flow chart shown in Figure 1 below, the specific steps involve the following points:[3-4]

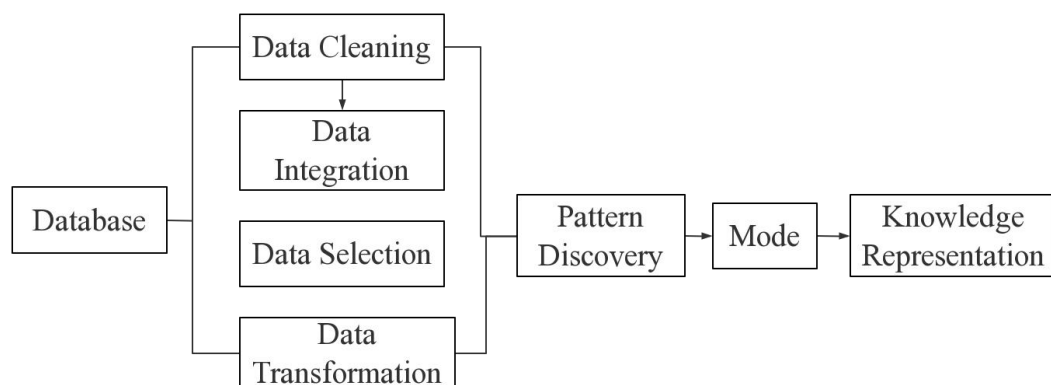


Figure 1 Flow chart of data mining

First, identify the business object. Clarifying the purpose of data mining is an important link in practical operation, which can ensure that the whole process will not appear blind problems.

Second, data preparation. Firstly, search all the internal and external data information related to the business object, and select the content suitable for mining application; Secondly, the quality of data information should be studied to make adequate preparation for subsequent analysis, and the types of mining operations should be defined. Finally, the data should be converted into an analysis model, and the corresponding content should be built based on the mining algorithm, so as to get the analysis model that is really suitable for the mining algorithm.

Third, data mining. Mining has been converted data, in addition to improving the selection of appropriate mining algorithm, other work can be completed automatically. Fourth, result analysis. The final results are analyzed and evaluated, often using visualization techniques.[5-6]

Fifth, knowledge assimilation. Integrate the knowledge gained from analysis into the organizational structure of the business information system.

2.2 Decision tree algorithm

The classification method based on decision tree is one of the most common supervised learning methods. In essence, the decision tree is a tree structure similar to a flow chart. Each internal node

represents an attribute test, the branch represents the actual result of the test, the leaf node represents a category, and the node at the highest level of the number represents the root node. The number of numbers affects the classification accuracy and classification size. Since the decision tree algorithm is based on examples, it will deduce the classification rules of the representation form of the decision tree from tuples without order and rules. It will use top-down recursion to compare attribute values in the internal nodes of the decision tree, and then branch down from this node according to different attribute values. The middle node of the algorithm is the category to be divided by learning. Among them, decision tree classification algorithm refers to greedy algorithm, which is classified and controlled by top-down regulations, as shown in Table 1 below:[7-8]

Table 1 ID3 version analysis

Algorithm: Generate a Decision tree based on a given data set (generate-decision-tree)
Input: training sample, all attributes are taken as discrete values, and the candidate attribute set that can be summarized is attribute-list
Output: Decision tree
Method:
(1) Create a node N
(2) If all samples in this node are in the same category C, then (3)
(3) Return N as a leaf node and mark it as class C
(4) If attribute-list is empty, then (5)
(5) N is returned as a leaf node and marked as the category with the largest number of categories in the samples contained by this node
(6) Select an attribute test-attribute with the maximum information gain from attribute-list
(7) Mark node N as test-attribute
(8) For each known value a_i in test-attribute, prepare to divide the sample set contained by node N
(9) Generate a branch from node N according to the test-attribute= a_i condition to represent the test condition
(10) Set s_i as the sample set obtained under the condition test-attribute= a_i
(11) If s_i is empty, the corresponding leaf node is marked as the category with the largest number of categories in the samples contained by the node
(12) Otherwise, the corresponding leaf node will be marked as the return value of Generate-Decision-tree

Based on the analysis in the above table, it is found that the core of this kind of algorithm is that when selecting attributes at all levels of the decision tree, information gain will be regarded as the selection standard of attributes, so that the maximum category information about the tested record can be obtained when testing each non-leaf node. The most common method is to detect all attributes and select the attribute with the maximum information gain to obtain the decision tree node. Different values of this attribute are used to build branches, and when recursing subsets of branches, this method is called to build branches of decision tree nodes until all subsets contain only the same class of data. The resulting decision tree can be used to classify new samples.

Association rule mining is the most common content in data mining, and it is also an important subject in the current practical teaching and application research of colleges and universities. It can mine valuable knowledge describing the interconnection between data directions in a large number of data.[8-11]

2.3 Teacher teaching quality evaluation system

From the perspective of realizing teaching in colleges and universities, the overall system process is shown in Figure 2 below:

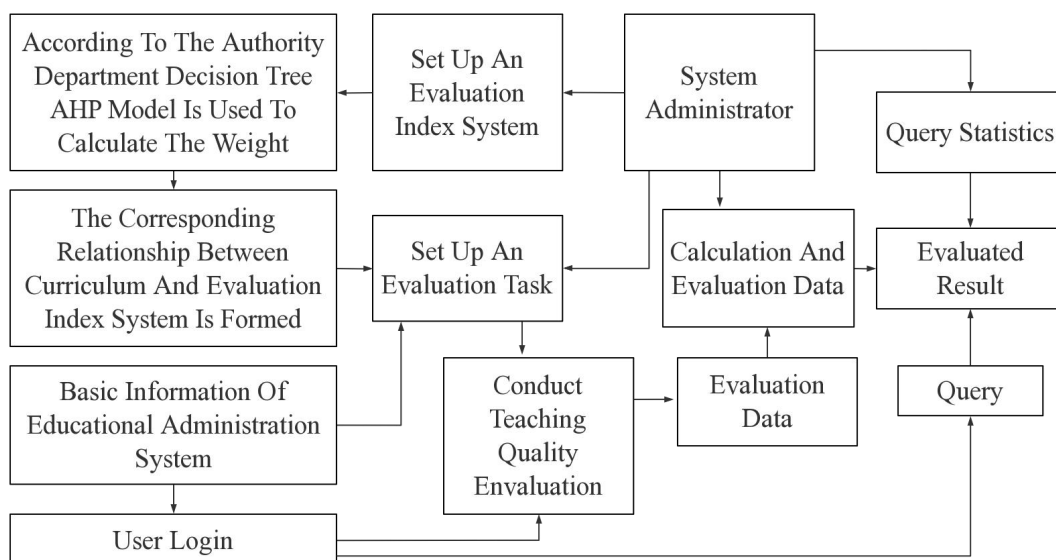


Figure 2 System flow chart

The administrator should set the corresponding course evaluation index, and the index weight is allocated and set based on the analytic hierarchy computing module. The system user login should enter the new interface according to the basic form information of the educational administration system, and finally complete the corresponding evaluation and query and other basic operations.

The research system in this paper mainly consists of five types of users, so five key functional modules are set. The specific structure is shown in Figure 3 below:

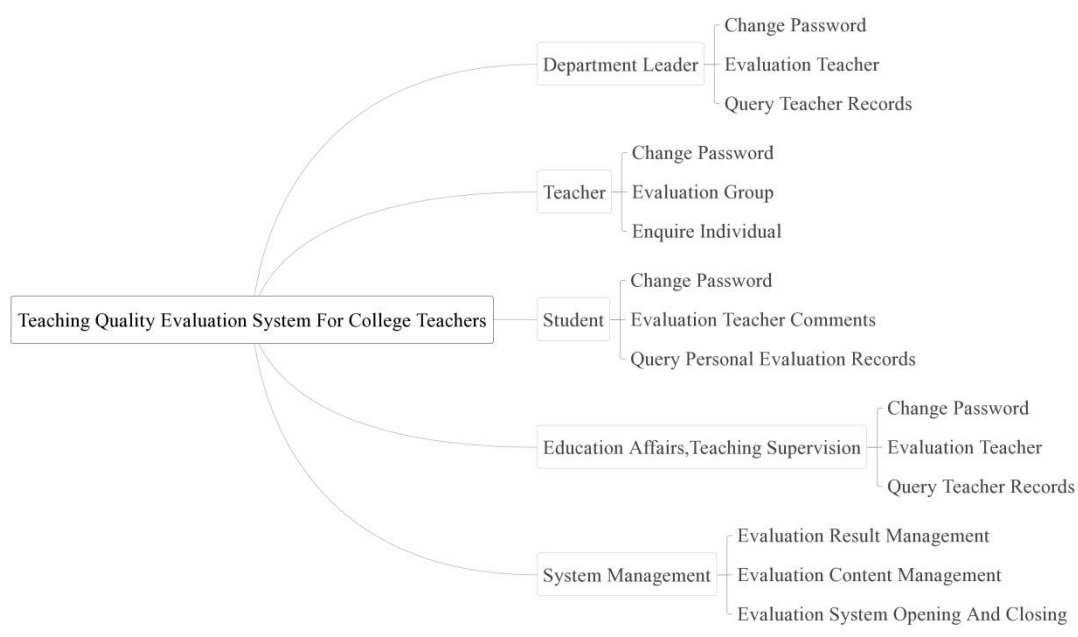


Figure 3. Functional modules of the evaluation system

First, the department of leadership function module. This module can help the leaders of each department in colleges and universities quickly enter their user names and passwords, enter the evaluation teaching module after effective verification, evaluate and analyze the teaching work of the teachers in this semester, and inquire the evaluation opinions of the teachers in this department;

Second, teacher function module. Teachers of each department can enter the corresponding module after inputting the user name and password, and focus on inquiring the evaluation results of leaders, peers, supervisors and students. The score obtained by teachers and full marks of the corresponding indicators are listed for each indicator.

Third, student function module. After entering the user name and password, the students enter the evaluation teaching module after verification, browse all their elective courses and teachers in this semester, and then evaluate the teacher's teaching effect according to the specific situation, and put forward the corresponding suggestions. It should be noted that each student has only one opportunity to evaluate the designated teacher, and repeated evaluation of a teacher is not allowed;

Fourth, teaching management module. Users such as the teaching Affairs Office and the teaching supervision group can effectively evaluate the teachers in colleges and universities according to the specific situation of teaching inspection and teaching supervision, put forward suggestions on the teaching management of teachers, directly inquire the final results of all teachers, and clarify the ranking of all teachers in the teaching evaluation.

Fifth, system administrator module. Administrators maintain the entire network operation and basic functions. Is the basic guarantee for the operation of the teaching quality evaluation system of college teachers.

3. Result analysis

3.1 Data Processing

After clarifying the basic concept and main process of decision tree algorithm, combined with the main requirements of practical teaching management in colleges and universities in the new era, this paper focuses on discussing the relationship between the basic situation of college teachers and the final evaluation results, so as to build an excellent teacher model. Before the experimental study, the data obtained in the new annual student evaluation of a university were selected to discuss the teaching evaluation results of 67 teachers in the university. In order to ensure excellent results in this study, invalid data should be processed and vacant data should be filled in the period of data cleaning. Finally, data information should be converted into standardized format after integrated processing, so as to lay the foundation for subsequent mining. At the same time, the total value of teaching evaluation obtained by the weighted average of student evaluation and peer evaluation should be replaced by grade, which aims to find the excellent teacher model. Therefore, the final teaching evaluation results can also be simplified into the content shown in Table 3 below:[12]

Table 2 Contents after data processing

Name	Gender	Age	Professional title	Educational background	Whether excellent or not
A	Male	≥ 45	Auxiliary height	Master's degree	Yes
B	Female	30-45	Auxiliary height	Master's degree	Yes
C	Male	< 30	Intermediate	Doctor	No
...	...	...	...	...	...

3.2 Build a decision tree

After effective processing of experimental data, 67 pieces of evaluation data were obtained, including age, gender, name, educational background, professional title and other contents. Different attributes included specific classification and value range. ID3 algorithm was used to construct and analyze, and statistical results as shown in Table 3 were finally obtained:

Table 3 Statistical results of data

Attribute name	Total sample	Attribute sample	The "Yes" class	The No class
General classification	/	67	33	34
Professional title -High	67	3	3	0
Professional title - Vice senior	67	19	13	6
Professional title - Intermediate	67	36	16	20
Professional title - Primary	67	9	1	8
Gender - Male	67	35	18	17
Gender - Female	67	32	15	17
Education background - Doctor	67	10	6	4
Education background - Master	67	27	12	15
Education background - Bachelor's Degree	67	30	15	15
Age <30	67	13	2	11
30 ≤ Age <45	67	31	15	16
Age ≥45	67	23	16	7

Firstly, the following formula should be used to calculate the information entropy of the analysis subset:

$$\begin{aligned}
 I(E_1) &= 0 \\
 I(E_2) &= -\frac{13}{19} \log_2 \frac{13}{19} - \frac{6}{19} \log_2 \frac{6}{19} = 0.8998 \\
 I(E_3) &= -\frac{16}{36} \log_2 \frac{16}{36} - \frac{20}{36} \log_2 \frac{20}{36} = 0.9912 \\
 I(E_4) &= -\frac{1}{9} \log_2 \frac{1}{9} - \frac{8}{9} \log_2 \frac{8}{9} = 0.5034
 \end{aligned}$$

Secondly, the expected entropy of professional title is calculated and analyzed according to the following formula:

$$E(\text{Professional Title}) = \frac{3}{67} \times I(E_1) + \frac{19}{67} \times I(E_2) + \frac{31}{67} \times I(E_3) + \frac{14}{67} \times I(E_4) = 0.8552$$

Finally, the information gain results of professional titles are calculated and analyzed according to the following formula:

$$\text{Gain (Professional Title)} = I(T) - E(\text{Professional Title}) = 0.1446$$

3.3 Final Result

Under the condition of large data set and more training data, using decision tree algorithm for classification analysis can get a larger data model, which will eventually increase the difficulty of data analysis and application. Therefore, in order to facilitate the efficient practice of teaching management application understanding, it is necessary to construct classification rules. The final experimental results show that teachers over 45 years old have richer teaching experience, better control of classroom teaching management, and ultimately achieve better teaching effect and higher actual teaching evaluation. Teachers between 30 and 45 years old generally have higher education, have rich innovative teaching methods, have accumulated certain teaching experience, and the actual teaching evaluation results are also high. On the other hand, although young teachers under 30 years old have mastered a lot of theoretical knowledge and advanced technology, they lack

high-quality teaching skills and management experience, and cannot fully control the whole process of the classroom, resulting in low teaching evaluation. It can be seen that the rational application of decision tree algorithm in the practical teaching management of colleges and universities can not only help colleges and universities grasp more information and data, but also get more valuable content in the mining research.[13-15]

Conclusion

To sum up, in the modern educational innovation and development, applying the decision tree algorithm in data mining to efficient practical teaching management can scientifically evaluate the teaching quality of professional teachers and clarify the main factors affecting the quality of classroom teaching, which can ensure the rationality and practicability of teaching evaluation and provide effective basis for the education management of colleges and universities.

References

- [1] LiYuan. Research on Application of Decision Tree Algorithm in Practical Teaching of Public Sports in Colleges and Universities [J]. Journal of Xi 'an Physical Education University, 2011, 28(6):4.
- [2] Yafang Yin, ZhaoyangSun. Research on Application of Decision tree Algorithm in Practice Teaching. Computer and Digital Engineering, 2018, 046(006):1078-1082,1088.
- [3] YanYuan. Application of Decision Tree Algorithm in Teaching Evaluation System of Colleges and Universities [J]. Journal of Zhejiang Ocean University (Natural Science Edition), 2006, 25(004):440-444.
- [4] ZhimingXiao. Research on Application of Decision tree Algorithm in Teaching evaluation of Colleges and Universities [J]. Light Industry Science and Technology, 2008, 24(11):164-165.
- [5] ZengyanPeng. Research on Application of R-rpart Decision Tree Algorithm in Grade Prediction -- A Case Study of "College Computer Foundation" Course [J]. Contemporary Educational Practice and Teaching Research: Electronic Edition, 2017(2):2.
- [6] PingHu. Analysis of ID3 Decision Tree Algorithm and Comprehensive Teaching Evaluation of Colleges and Universities [J]. Journal of Tongling University, 2012(004):011.
- [7] JunWANG. Application of Decision Tree in University Classroom Teaching Evaluation [J]. Computer Knowledge and Technology: Academic Edition, 2020, 16(36):2.
- [8] Ping Hu. Analysis of ID3 Decision Tree Algorithm and Comprehensive Teaching Evaluation of Colleges and Universities [J]. Journal of Tongling University, 2012, 11(4):2.
- [9] MinZhong. Application of Decision Tree Algorithm in University Teaching Information System [J]. Journal of Wuhan Polytechnic, 2021, 33(2):3.
- [10] JianwenDong, YichunZhang, YanHu. Design and Application of Learning Outcome Prediction Model Based on Decision Tree Algorithm: A Case study of a university calculus Course [J]. Journal of Guangzhou Radio and Television University, 2022, 22(1):9.
- [11] NaDeng, XiaofanBian, WeiHuang, et al. Data mining and its application in Teaching practice [J]. Journal of Agricultural University of Hebei: Agriculture and Forestry Education Edition, 2004, 6(4):4.
- [12] HongyingYu. Effect evaluation of Business English Practice Teaching based on Decision Tree [J]. Science and Technology Bulletin, 2016(2):4.
- [13] Zhizhen Liu. Business English Practice Teaching Effectiveness Evaluation based on Decision Tree [J]. China Training, 2016(14):2.
- [14] HuanFang, XianwenFang, JuanGuo, et al. Study on the teaching method of Classification and prediction curriculum driven by case [J]. Journal of Changchun Normal University, 2018, 37(8):6.
- [15] ZhenLin and WeiWang. Research on Optimization of Data Mining Algorithm Based on Decision Tree [J]. Modern Computer (Professional Edition), 2012, 000(019):11-14.