Innovative Design and Intelligent Manufacturing
L.C. Jain et al. (Eds.)

doi:10.3233/FAIA241091

Evaluation Model of Higher Mathematics Teaching Quality Based on Improved ID3 Algorithm

Jiang WANG[a,1]
[a]*College of General Education, Chongqing Three Gorges Vocational College, Chongqing 404100, China*

Abstract. In order to obtain effective information from accumulated data and make awards and punishments for teachers such as appointment, promotion, or bonus increase, and provide convincing basis for this decision, an evaluation model of higher mathematics teaching quality based on improved ID3 algorithm is proposed. This paper introduces the definition and classification of data mining, and introduces the ID3 algorithm of decision tree in detail. According to the ID3 algorithm, a large number of teaching evaluation data samples collected in colleges and universities are analyzed to obtain information gain on different attributes and generate the final decision tree, which can be converted into a set of if then rules. Generate rules and decision trees, and then analyze and predict the new data. The results show that A1 teaching content attribute is the most important in this evaluation system. From here, we can get a fair and objective evaluation. Secondly, teaching methods are also important indicators. Conclusion: Through data modeling, we can discover rules and patterns, extract valuable information, and avoid irrationality in current teaching quality evaluation. The results of example verification and analysis show the effectiveness of this method. Provide reasonable and scientific decision support for teaching quality evaluation, so as to improve teaching quality and improve teaching results.

Keywords. data mining; Teaching quality evaluation; ID3 algorithm; Decision tree

1. Introduction

Since the reform and opening up, China has entered a stage of rapid development in all fields. In the process of development of the demand for talents is increasing day by day, high-quality talent reserves to a certain extent related to a country's comprehensive competitiveness in the world, the popularization of higher education is related to the cultivation of talents and reserves. In order to narrow the gap between China and foreign education level, comprehensively improve the quality of national, China began in 1999 to expand the enrollment scale of colleges and universities [1-3]. Expansion of college enrollment in a certain extent to promote China's economic structure of change, but also to promote the realization of the process of sustainable economic development. China in this unprecedented speed of development, the quality of education as a prerequisite for training and reserves of talent, inevitably become the focus of national attention and

[1] Corresponding Author：Jiang WANG cqsx_wj@163.com

discussion, the ability of universities and colleges and their quality of education is also subjected to a huge test and challenge [4-7]. The development of higher education must focus on improving quality, deepen its reform in teaching, so as to achieve the goal of perfecting the education quality assurance system. After repeated research and modification of the "National Medium and Long-term Education Reform and Development Plan" was finally considered and adopted, the core idea of the plan is that information technology can be used as a tool to promote the development of education, improve the teaching assessment of institutions of higher learning, and realize the informationization of education [8-10]. In this case, the education sector after discussion and research immediately after the release of the "ten-year development plan for education informatization", the plan clearly points out the need to promote the modernization of education-by-education informatization. In order to promote the development of modern higher education, the integration of scientific and technological means, it is necessary to modernize and reform the content, means and methods of education, in order to effectively improve the quality of education [11-13]. The quality of education in colleges and universities will affect the development of education in China and the cultivation and reserve of talents, on the other hand, the quality of education depends on the professionalism and teaching level of the teaching team in colleges and universities, and also represents the effectiveness of education reform, therefore, to promote the modernization of education and promote the development of education is the primary issue at present. Therefore, the modernization of education and the promotion of education development are the primary issues at present

It is important to provide a convincing basis for decision-making, to be objective and fair, to mobilize teachers' motivation through reasonable rewards and punishments, and to help school administrators to exercise appropriate supervision and control over teachers so as to improve the quality of education and teaching.

2. Research methodology

2.1. Data mining

2.1.1. Definition of Data Mining

With the development of science and technology, computers, networks, databases and other technologies are widely used in day-to-day management, all walks of life have accumulated a large amount of information and data, access to databases and query operations, is far from being able to meet the requirements. People need to get more important information behind these data from the massive data, such as the overall characterization of the data, attempting to discover the correlation between the events, as well as the trend of development for prediction.

2.1.2. Classification for Data Mining

There are various data mining classification methods, which can be divided into classification and predictive model discovery, data summarization, clustering, association rule discovery, sequence pattern discovery, dependency or dependency model discovery, anomaly and trend discovery, etc. according to mining tasks. According to mining methods, they can be categorized into machine learning methods,

statistical methods, neural network methods and database methods. Machine learning includes inductive learning methods, case-based learning, genetic algorithms and so on. Statistical methods include regression analysis, discriminant analysis, cluster analysis, exploratory analysis and so on. Neural network methods include forward neural networks, self-organizing neural networks, etc. Database analysis methods include multi-dimensional data analysis methods, attribute-oriented induction methods, etc.

2.2. ID3 algorithm

ID3 algorithm is a decision tree classification algorithm based on information entropy. The core of this algorithm is to select attributes on nodes at all levels in the decision tree, and use the highest information gain as the test attribute of the node, so that when each non leaf node is tested, the maximum category information about the tested sample set can be obtained. After the sample set is divided into subsets by using this attribute, the entropy value of the system is minimum. It is expected that the average path of the non leaf node to each subsequent leaf node is the shortest, so that the average depth of the generated decision tree is smaller, thus improving the classification speed and accuracy .

Tree generation algorithm (ID3):

Let S be a set of s data samples, assume that the decision attribute has m different values, and define m different classesC_i, $(i = \{1, \cdots, m\})$, s_1It's classC_iThe number of samples in the The expected information for categorizing a sample can be given by the following equation (1).

$$I(s_1, s_2, \cdots, s_m) = -\sum_{i=1}^{m} \frac{s_1}{s} \log_2 \frac{s}{s} \tag{1}$$

If attribute A is taken as the root of the decision tree, attribute A has v different values $\{a_1, a_2, a_3, \cdots, a_v\}$, which divides S into v subsets $\{S_1, S_2, \cdots, S_v\}$, of which$s_j$Including such samples in S, which have the value aj on A, these subsets correspond to the branches grown from the nodes containing the set S. For a given subsets$_j$, with equation (2).

$$I(s_{1j}, s_{2j}, \cdots, s_{mj}) = -\sum_{i=1}^{m} \frac{s_{ij}}{s} \log_2 \frac{s_{ij}}{s} \tag{2}$$

The entropy of the subset divided by A is given by formula (3):

$$H(A) = \sum_{j=1}^{v} \frac{s_{1j}+\cdots+s_{mj}}{s} I\left(s_{1j} + \cdots +_{s_{mj}}\right) \tag{3}$$

The smaller the entropy, the higher the purity of the subset partition. The information gain obtained by branching on attribute A is formula (4):

$$H(A) = \sum_{j=1}^{v} \frac{s_{1j}+\cdots+s_{mj}}{s} I\left(s_{1j} + \cdots +_{s_{mj}}\right) \tag{4}$$

D3 selects the attribute A with the largest Gain (A) as the root node of the sample set. The sample subset of each branch recursively uses the ID3 method to establish the decision tree node and branch until the sample subset belongs to the same class. This

method makes the average depth of the generated decision tree minimum and has a fast speed, so a decision tree is generated.

3. Analysis of results

The decision tree method of data mining has been successfully applied in the information management of teachers' business records in the mathematics department of a college. Here, we only discuss the data mining of classroom teaching of the lecturers, and do not involve the teachers' scientific research and titles.

3.1 Data pre-processing

It is roughly divided into three steps: data selection and integration, data preprocessing, and data transformation. For details, please refer to the literature.

There are three copies of the table of teaching quality assessment index system for teachers in the department, one for teachers' peer review, one for departmental leaders' review, and one for students' review, and the contents of the three copies of the table are as shown in Table 1, in a similar format, with a total of two levels of indexes, and each item has a score of 10 points.

Table 1. Teachers' quality of teaching assessment form (II)(Form for Department Head and Faculty Evaluation) Department: Higher Mathematics: Evaluator: Score.

Tier 1 indicators	secondary indicators		distribution of points	Score
pedagogical content	Correctness	Lectures are correct and conceptually clear	4	
	conforming to the syllabus	Meets the requirements of the syllabus, is focused and in-depth	2	
	Advanced	Academic developments and new advances in the discipline	1	
	Theory to practice	Practice Teaching	3	
teaching attitude	The status of lesson planning	Well-prepared, well-organized and focused lesson plans	3	
	Teaching and educating	Responsible and dedicated	3	
	Compliance with teaching discipline	Teach according to the lesson plan and do not arrive late	2	
	Organization of teaching and learning		2	
Teaching methods	omit		10	
Teaching Effectiveness	omit		10	
Evaluation ratings				

Collect three survey scoring tables, summarize and calculate the daily teaching quality of each teacher, and calculate the average. The teaching quality evaluation form of an annual assessment of 49 teachers is discussed. In order to simplify the problem, the four first level indicators in the above table are taken as the fields (attributes) in the database single table (jx table name), which are A1, A2, A3, A4 respectively. Scoring refers to the evaluation index assigned to the department, which is divided into three grades: excellent (9-10), good (7-8.9), and medium (other). The final result (evaluation grade) is the fifth field (A5 attribute) of the single table. The grades in A1, A2, A3, A4 are divided into: 1, excellent 2, good 3, and medium. Data conversion is carried out according to this requirement. The daily teaching quality data of teachers in the department are obtained (see Table 2).

Table 2. Daily teaching quality of teachers in the department

ID	pedagogical content	teaching attitude	Teaching methods	Teaching Effectiveness	Evaluation ratings
1	good	good	good	excellent	good
2	good	good	good	good	good
3	good	good	good	middle	good
4	excellent	excellent	excellent	excellent	excellent
……					
23	good	excellent	middle	good	good
24	good	excellent	excellent	excellent	excellent
25	good	excellent	middle	good	good
26	good	middle	middle	middle	middle

3.2 ID3 algorithm builds decision tree

To calculate the information gain of all attributes, first calculate the expected information (information entropy) of the sample for A5 attributes.

Analyzing the data in the table, the numbers of good, excellent and moderate students are 11, 26 and 12 respectively.

I(11,26,12)＝1.46604

The conditional information entropy and information gain for each attribute are calculated below.

- For attribute A1, the attribute value is "excellent", and there are 13 people in total, of which category attribute A5 is excellent, good, and the number of people in the middle is 10,3,0. s11=10,s21=3,s31=0; The attribute value is "good", and there are 19 people in total, of which the category attribute A5 is excellent, good, and the number of people in the middle is 1,15,3. s12=1,s22=15,s32=3; The attribute value is "medium", and there are 17 people in total. The category attribute A5 is excellent, good, and the number of people in the middle is 0,8,9. s13=0,s23=8,s33=9。

Query SQL statement: SelectJx IDfromJxwhere (((jx. A5="medium") and ((jx. A1)="excellent") show that A1 is excellent, and A5 is the people in the middle.

I(10,3,0)=0.77935

I(1,15,3)=0.91328

I(0,8,9)=0.9975

H(A1)=13/49∗I(10,3,0)+19/49∗I(1,15,3)+17/49∗I(0,8,9)=0.906967

- For attribute A2, the attribute value is "excellent", with a total of 10 people, among which the number of people in the category attribute is excellent, good, and medium is 2,7,1. The attribute value is "good", and there are 26 people in total. The number of people with excellent category attribute, good category attribute, and medium category attribute is 7,13,6. The attribute value is "medium", and there are 13 people in total. Among them, the category attribute is excellent, good, and the number of people in the middle is 2,6,5.

H(A2)=1.41833

- For attribute A3, H (A3)=1.16766. Item

- For attribute A4, H (A4)=1.207373. Item

Their information gains are.
Gain(A1)=0.55907
Gain(A2)=0.04771
Gain(A3)=0.29838
Gain(A4)=0.258667

Gain (A1) is taken as the root node according to ID3 algorithm. Because A1 has the largest information gain. First, classify the attribute by A1 to get the following decision tree. Then continue to use the above algorithm recursion.

The A1 attribute value is the division of "excellent" subtree. Among the 13 records, the desired information I (10,3,0)=0.77935.

- The calculation of conditional entropy for A2 values of "excellent", "good", "medium", category attributes of "excellent", "good", and "medium" is as follows:

I(1,0,0) =0;I(7,1,0)=0。54356;I(2,2,0)=1H(A2)=0.64219。

- Calculate the condition entropy of A3 attribute. The number of people whose A3 value is "excellent", category attribute is "excellent", "good" and "medium" is 3,0,0 respectively. The number of A3 "good", category attribute "excellent", "good" and "medium" is 7,0,0. A3 is "medium", the category attribute is "excellent", "good", and the number of "medium" is 0,0,3.

H（A3）＝0。

- Calculate A4 attribute conditional entropy, H (A4)=0.67667.

Gain（A2）＝0.137159
Gain（A3）＝0.77935,
Gain（A4）＝0.102678

In order to increase the readability and comprehensibility of the decision tree, it is necessary to prune the decision tree. However, for test data, pruning the decision tree will inevitably lead to an increase in the error rate, we design a maximum allowable error rate to obtain a pruned decision tree. In this example, we use the first pruning method to "prune" the tree by stopping the construction of the tree in advance (e.g., by deciding not to split or divide a subset of the training samples at a given node).

4. Conclusion

At present, in the field of data mining, there are many models to solve the classification problem, and the most extensive classification model is still the decision tree algorithm. In the process of classification, the decision tree method does not need to set any parameters artificially, which is more suitable for the requirements of knowledge discovery; The decision tree classification method does not need any additional information other than the test data set, which ensures that the decision tree has a higher classification speed and a very good classification accuracy compared with other classification methods. ID3 algorithm is the most influential algorithm of the decision tree. The decision tree technology ID3 algorithm is applied to the teaching quality evaluation of higher vocational education, student achievement prediction, teacher and teacher management, etc., giving fair, just and objective evaluation criteria. The school uses this hidden information to develop the existing value, provide decision-making support for the teaching management department, guide daily teaching and management work, Improve the management level, better carry out teaching work, guide the development of higher vocational colleges with reasonable ideas and scientific attitude, and improve the school running level of the whole school.

References

[1] Huang, Q., & Zhao, Y. (2023). Application of intelligent teaching resource organisation model and construction of performance evaluation model. Applied Mathematics and Nonlinear Sciences, 8(2), 241-250.
[2] Dou, F. (2022). Evaluation Method of College English Education Effect Based on Improved Decision Tree Algorithm. Journal of Information Processing Systems, 18(4), 500-509.
[3] Mao, L., & Zhang, W. (2021). Analysis of entrepreneurship education in colleges and based on improved decision tree algorithm and fuzzy mathematics. Journal of Intelligent & Fuzzy Systems, 40(2), 2095-2107.
[4] Yin, X. (2023). Application analysis of data mining based on improved decision tree in English flipped classroom teaching. International Journal of Networking and Virtual Organisations, 28(2-4), 171-183.
[5] Wang, B., Kong, W., & Zhao, P. (2021). An air quality forecasting model based on improved convnet and RNN. Soft Computing, 25(14), 9209-9218.
[6] Guo, Y. (2021). University Classroom Teaching Model Based on Decision Tree Analysis and Machine Learning. Mobile Information Systems, 2021, 1-12.
[7] Fang, C. (2021). Intelligent online English teaching system based on SVM algorithm and complex network. Journal of Intelligent & Fuzzy Systems, 40(2), 2709-2719.
[8] Yang, Y. (2021). The evaluation of online education course performance using decision tree mining algorithm. Complexity, 2021, 1-13.
[9] Wu, F., Liu, X., Wang, Y., Li, X., & Zhou, M. (2022). Research on evaluation model of hospital informatization level based on decision tree algorithm. Security and Communication Networks, 2022, 1-9.
[10] Hou, J. (2021). Online teaching quality evaluation model based on support vector machine and decision tree. Journal of Intelligent & Fuzzy Systems, 40(2), 2193-2203.
[11] Gao, B. (2022). Research and implementation of intelligent evaluation system of teaching quality in universities based on artificial intelligence neural network model. Mathematical Problems in Engineering, 2022, 1-10.
[12] Zhao, X., & Zheng, C. (2021). Fuzzy evaluation of physical education teaching quality in colleges based on analytic hierarchy process. International Journal of Emerging Technologies in Learning (iJET), 16(6), 217-230.
[13] Yang, Y. (2021). Quality evaluation method of a mathematics teaching model reform based on an improved genetic algorithm. Scientific Programming, 2021, 1-10.