Application of Fuzzy Clustering Maximum Tree Algorithm in Teaching Quality Evaluation

Bo Xu*

Zhuhai City Polytechnic, Zhuhai 519090, China

*Corresponding author e-mail: xubo@zhcpt.edu.cn

Abstract. The evaluation process of school teaching quality is more complicated. The traditional schemes mainly based on manual evaluation or expert evaluation. Due to the strong subjectivity, there is a big deviation between the evaluation results and the real results. To this end, a teaching quality evaluation program based on the largest tree of fuzzy clustering is proposed. After raw data collection, preprocessing, and data mining, accurate and objective evaluation results are finally obtained. Taking the original evaluation data set as the object, construct the fuzzy clustering maximum tree model, and obtain the importance matrix of the data set and the similarity coefficient between the input data variables. With the help of the fuzzy clustering maximum tree model, the importance of each factor in the evaluation of teaching quality can be objectively evaluated. Application examples also verify the superiority of the proposed evaluation algorithm.

Keywords: Fuzzy Clustering Maximum Tree, Teaching Quality Evaluation, Data Mining, Importance Matrix

1. Introduction

The teaching process is a relatively complex system engineering. From the perspective of the actors, it includes teachers, students, and other related educational work managers. From the perspective of the teaching object, it includes paper textbooks, electronic textbooks and various teaching equipment and tools. The teaching system also includes specific teaching methods and teaching methods [1]. The traditional view is that the teacher is the leader of the entire teaching process, so the level of teaching quality is closely related to the teacher's teaching ability and classroom grasping ability. It is undeniable that teachers play a central role in the improvement of teaching quality, but other factors, such as the overall quality and acceptance of students themselves, the advanced nature and functionality of teaching equipment, etc., will also have an important impact on teaching quality [2, 3]. Especially in the context of the Internet and big data, multimedia teaching and online teaching will have an important impact on teaching quality. A comprehensive, true, and scientific evaluation of teaching quality is one of the basic conditions for accurately grasping the progress and effects of teaching implementation and improving school teaching quality. The quality evaluation of the research teaching system needs to start from two aspects: evaluation methods and evaluation methods. First, from the perspective of teaching methods, most of the existing teaching evaluation methods use the most basic statistical analysis schemes, which cannot handle massive evaluation data; secondly, from the teaching method In terms of quality evaluation methods, the traditional manual scoring method is too subjective and cannot accurately feedback the true teaching quality status. In recent years, with the continuous improvement of computer software and hardware technology, machine learning algorithms have been widely used in various fields [4, 5]. For this reason, this paper introduces the fuzzy clustering maximum tree algorithm into the teaching quality evaluation process, and strives to improve A true and objective assessment of the real teaching quality of a school.
2. Data Mining for Education Quality Evaluation

Analyzed from a narrow perspective, teaching quality evaluation refers to the use of a specific evaluation index system to evaluate the teaching effect of teachers for a period of time; from a broad perspective, teaching quality evaluation is not only an evaluation of teachers’ teaching effects, but also research and evaluation of the function of the entire education system. Education quality evaluation is an indispensable and important link in education and teaching management. A good education quality evaluation is one of the basic conditions for perfecting the education system and improving teaching quality. In the context of the new network age, teachers’ teaching quality is affected by multiple factors, including teacher qualifications, teacher abilities, the purchase and use of networks and multimedia, and the running-in between teachers and students. In the context of the Internet and big data, the construction of the teaching quality evaluation index system and the design of specific indicators should be based on data mining, including the collection of raw data for teaching quality evaluation, data classification mining, decomposition, index selection and system construction. And analysis and research on the evaluation results.

(1) Data collection and preparation

At present, with the exception of some remote areas and underdeveloped areas in the central and western regions, most colleges and universities and primary and secondary schools have their own local area network systems, computer management systems, and database systems, but they vary in the degree of system perfection. The evaluation of teaching quality not only assesses the overall scores of students in the mid-term and the end of the term, but also includes some classroom teaching indicators, indicators of students’ comprehensive ability, and comprehensive consideration of teaching. After all relevant information is entered into the database; a specific data file is formed. These database files are the basis for establishing an index evaluation system. The collected raw data cannot be used directly, because the raw data contains a large amount of incomplete data, redundant data, and erroneous data, and can enter the database management system after preprocessing steps such as classification and cleaning. The sorting and preprocessing of the original evaluation data will affect the accuracy of subsequent data mining. Therefore, big data analysis software tools such as Hadoop framework or Spark framework are usually used in data preprocessing [6].

(2) Data mining processing

The teaching evaluation data contains a huge amount of information within a period of time after sorting, and the use of specific algorithms or model tools can dig out the useful information inside. Traditional evaluation models mostly use manual evaluation methods such as scoring and scoring. This method is based on qualitative evaluation and lacks the scientific and timeliness of content evaluation. The index value obtained after real data mining is a quantitative research method, which is relatively fairer. On this basis, referring to the school leadership score, teacher score, student score and parent score, the teaching quality evaluation result is obtained. It is more objective and true, and to a large extent avoids the subjective biases and assumptions of the judges. The establishment of the index system is more complicated. It can be based on the AHP analytic hierarchy process to first decompose all the indexes layer by layer to clarify the hierarchical relationship and subordination relationship between the indicators. The index system constructed under the data mining model is not a static index evaluation system, but a dynamic evaluation model. All indicators should be dynamically adjusted and updated according to the actual situation of the school. First, determine the first-level indicators in the hierarchical mode, and conduct systematic research on each first-level indicator, and finally form a tree-shaped indicator system. The weight matrix can be constructed with the refined secondary and tertiary evaluation indicators to facilitate the verification of optimal indicator consistency [7-8].

In terms of the selection of specific evaluation methods, this paper chooses the more efficient fuzzy clustering maximum tree model in machine learning and artificial intelligence algorithms to process the evaluation indicators and raw data in a fuzzy and dynamic manner. The fuzzy evaluation method is suitable for the case of a small number of samples, and the maximum decision tree algorithm can also obtain more accurate deduction results in terms of index deduction and data depth mining [9].

(3) Analysis and discussion of evaluation results

After the basic quantitative evaluation results are obtained through data fuzzing and in-depth mining, the relevant teaching quality assessors can also refer to the artificial scoring results for further comparative analysis. If there is a big discrepancy, there may be an unreasonable phenomenon in the indicator selection method, and the indicator system is being further adjusted. Under the dual evaluation of quantitative analysis and qualitative analysis, more objective and formal evaluation results can often be obtained [10].

3. Fuzzy Clustering Maximum Tree Algorithm and Its Application in Teaching Quality Evaluation

(1) Construction of fuzzy clustering maximum tree algorithm model
Fuzzy theory came into being earlier and can be traced back to the 1960s. It was put forward by the American cybernetic scholar Zadeh. Once the theory was put forward, it caused a strong reaction in the academic community. Various optimized and improved versions emerged one after another, and the fuzzy theory was also widely used in many fields. Fuzzy clustering algorithm is a classic early machine learning algorithm, which establishes hierarchical relationships based on the amount of change or the close relationship between the input data, and then obtains accurate clustering results and classification results in a small sample environment. This paper incorporates the maximum tree algorithm on the basis of the fuzzy clustering algorithm. The maximum tree can improve the performance of the classic fuzzy clustering algorithm without the need to multiply the matrix in the clustering. The specific steps are as follows:

First, classify the universe of discourse that contains data set objects:

$$Z = \{z_1, z_2, \ldots, z_i, \ldots, z_n\}$$  \hspace{1cm} (1)

Each object is composed of three indicators, and a set of importance degrees of the teaching quality evaluation index system is established according to the AHP analytic hierarchy process:

$$U = \{u_1, u_2, \ldots, u_j, \ldots, u_k\}$$  \hspace{1cm} (2)

Sort the importance of the indicators from to to, and the evaluation values corresponding to the importance degrees are odd sequences, and the original matrix obtained can be expressed as:

$$\begin{bmatrix}
    z_{11} & z_{12} & \cdots & z_{1m} \\
    z_{21} & z_{22} & \cdots & z_{2m} \\
    \vdots & \vdots & \ddots & \vdots \\
    z_{n1} & z_{n2} & \cdots & z_{nm}
\end{bmatrix}_{n \times m}$$  \hspace{1cm} (3)

The fuzzy clustering maximum tree algorithm is based on the overall evaluation requirements of all elements in the fuzzy cluster analysis. Secondly, standardize the filtered index data and establish a similarity matrix corresponding to the fuzzy set. Any two variables in the matrix $z_i$ and $z_j$ the similarity coefficient between $r_{ij}$ to express the similarity coefficient with the help of Hamming distance method $r_{ij}$. The process is described as follows:

$$d(z_i, z_j) = \sum_{k=1}^{m} |z_{ik} - z_{jk}|$$

$$r_{ij} = 1 - \eta d(z_i, z_j)$$  \hspace{1cm} (4)

Among them, $\eta$ is a system parameter that controls the Hamming distance of the data variable. The value range is between 0 and 1. Solve the similarity coefficient between any two variables in the same way, and use the fuzzy clustering maximum tree algorithm to construct the similarity matrix $R$. Expressed as:

$$R = \left[ r_{ij} \right]_{m \times n}$$  \hspace{1cm} (5)

Taking each element in the universe as a fixed point, construct a tree structure according to the size of the similarity coefficient, and give the corresponding threshold range until each fixed point is connected to each other, and the tree structure formed at this time is maximized. Refer to the system parameter value to exclude the threshold range exceeding $[0, 1]$ the tree-like structure is simplified. After the fuzzy clustering maximum tree structure model is formed, the collected and preprocessed raw data is input into the model, a set of quantified evaluation values can be obtained, and then the actual teaching quality evaluation results can be obtained.

(2) Application and effect evaluation of algorithm

Taking a high school in Y city as the research object, the school has 152 teachers and 1,684 students. The evaluation period is from January 2019 to December 31, 2020. The algorithm performance test environment
is the school’s campus network with a bandwidth of 200M. Other test environments are shown in Table 1 below:

**Table 1. Test environment of fuzzy clustering maximum tree algorithm**

|                  | Parameter value               |
|------------------|-------------------------------|
| Web environment  | Campus network 200M bandwidth |
| Database environment | SQL2008                     |
| server           | Contact AWS                   |
| Client hardware  | Intel i7, RAM 16G, hard disk 2T |
| Client software  | WINDOWS10                     |

First, perform a functional test of the system to verify the effectiveness and applicability of the algorithm. Taking the user login function as an example, the test is as follows:

**Table 2. User login function test case**

| Test items       | Test content                                      |
|------------------|---------------------------------------------------|
| User login       | Check whether the system works normally           |
| Testing purposes | The matching degree of fuzzy algorithm and hardware |
| Test content     | Can log in normally, and whether other functions are normal |
| Test steps       | Enter user name, password, enter test data, and compare actual results with expected results |

Secondly, the performance test of the evaluation system and the fuzzy clustering algorithm is carried out from two aspects: the data processing efficiency of the algorithm and the evaluation accuracy of the algorithm. The test period is 2 years. Excluding the winter and summer vacations of two months each year, the effective data statistics month of each year is 8 months. For the comparison of the data processing efficiency of each effective statistics month, see Table 3:

**Table 3. Comparison of data processing efficiency**

| Statistics monthly | Total amount of data | Fuzzy cluster evaluation scheme | Traditional Expert Evaluation Program |
|--------------------|----------------------|---------------------------------|---------------------------------------|
|                    |                      | Data processing time h          |                                       |
| 2019.3             | 156415               | 3.56                            | 15.25                                 |
| 2019.4             | 164842               | 3.65                            | 21.56                                 |
| 2019.5             | 135854               | 3.18                            | 18.15                                 |
| 2019.6             | 146695               | 3.26                            | 17.44                                 |
| 2019.9             | 154412               | 3.69                            | 16.59                                 |
| 2019.10            | 102564               | 2.86                            | 12.56                                 |
| 2019.11            | 169897               | 4.10                            | 21.55                                 |
| 2019.12            | 163529               | 4.02                            | 18.47                                 |
| 2020.3             | 154461               | 3.68                            | 16.98                                 |
| 2020.4             | 133567               | 3.31                            | 15.85                                 |
| 2020.5             | 125958               | 3.18                            | 16.52                                 |
| 2020.6             | 103554               | 2.97                            | 13.47                                 |
| 2020.9             | 145287               | 4.15                            | 12.60                                 |
| 2020.10            | 163985               | 5.13                            | 15.68                                 |
| 2020.11            | 158451               | 5.02                            | 22.59                                 |
| 2020.12            | 136584               | 4.19                            | 19.81                                 |

From the analysis of the time-consuming situation of statistical data processing, in the process of processing each effective statistical monthly data, the fuzzy clustering maximum tree algorithm has a very big advantage in data processing efficiency compared with traditional manual or semi-manual data processing schemes. Finally, use MATLAB simulation software to verify the accuracy of teaching evaluation in 2019 and 2020, as shown in Figure 1 and Figure 2 below:
Figure 1. Analysis of the deviation of the accuracy of teaching quality evaluation in 2019

Figure 2. Analysis of the deviation of the accuracy of teaching quality evaluation in 2020

From the simulation results, in terms of the accuracy deviation of the basic data evaluation in each month, the deviation of the fuzzy clustering maximum tree algorithm and the theoretical value is small, while the maximum accuracy deviation of the traditional evaluation method approaches 2.5.

4. Conclusions
Regularly assessing the teaching quality of schools is one of the effective means to improve teaching quality. This article applies the fuzzy clustering maximum tree algorithm in machine learning algorithms to teaching quality assessment, which can effectively improve the efficiency and accuracy of data processing, and the results of application examples It shows that compared with the traditional evaluation scheme, the data processing efficiency and accuracy deviation control of the fuzzy clustering maximum tree algorithm have certain advantages.

References
[1] Tan Wei, Gu Xiaoqing. Research on the Mixed Teaching Mode and Effect Evaluation Index for Open Education. China Audio-visual Education, 2019, (02):126-130.
[2] Zhou Jiushi, Bao Jiansheng. Evaluation and Enlightenment of American PACT Teacher Teaching Quality Evaluation System. Educational Science, 2018, 34(05):83-90.
[3] Wang Sufang. Exploration of the Quality Evaluation System of Chinese-foreign Teacher Cooperative Teaching. China Adult Education, 2018, (11):88-90.
[4] Tang Bo, Kong Jianyi, Zeng Liangcai, et al. Research on "Internet +" Blended Teaching. Development and Evaluation of Higher Education, 2018, 34(03):90-99.
[5] Zhang Yaqing. Auxiliary teaching quality evaluation model based on active learning support vector machine. Modern Electronic Technology, 2019, 42(07):112-114.
[6] Chen Yanping. Teaching exploration and practice of "natural language processing" course [J]. Science and education Wenhui (first ten issues), 2020 (12): 135-136
[7] Lin Li. The development of natural language processing technology under the background of artificial intelligence [J]. Electronic world, 2020 (12): 135-136.
[8] Ma Zhibin. Research on improved technology of natural language processing based on word join [J]. Science and technology communication, 2020,12 (20): 145-147.
[9] Han Yi. Construction of online teaching resources recommendation model based on keywords [J]. Automation technology and application, 2019,38 (09): 170-173.
[10] Zhou Zhengguo. Application of reduction algorithm based on variable precision rough set in online education resource retrieval [J]. Information technology and informatization, 2018 (12): 84-85.