Evaluation of College Teaching Quality Based on Association Data Mining

Jinghong GUO¹

Preparatory Intelligent Manufacturing College of Minxi Vocational and Technical College, Longyan, 364021, China

Abstract. To find the valuable information hidden behind the teaching evaluation data, a scheme of applying the data mining technology based on association rules to teaching evaluation is proposed. We analyze the basic principle of association data mining and the needs of teaching quality evaluation in colleges, and establish a multidimensional comprehensive evaluation online analysis mining model. After data standardization, to reduce the generation of candidate item sets, Apriori algorithm is optimized, and the running speed of association rule mining is increased by an order of magnitude. Finally, the concrete construction methods of mining environment and association rules are provided through experimental analysis, which verifies the effectiveness of the teaching evaluation scheme proposed in this paper.

Keywords-teaching quality; Apriori; data mining; index system; evaluation.

1. Introduction

A scientific and reasonable teaching quality evaluation system is the foundation for ensuring teaching quality. In the era of big data, the traditional teaching quality evaluation system can no longer fully meet the development needs of education reform in the era of big data. Therefore, under the new technologies and ideas provided by the big data era, it is necessary to try to build a new teaching quality evaluation system based on data mining, keep up with the pace of the times, and actively promote teaching reform. Its purpose is to check and promote teaching and learning, and gradually establish a teaching quality analysis and evaluation system with the fundamental goal of promoting the common development of schools and teachers, so that the teaching quality evaluation data can truly become an important basis for scientific management and scientific decision-making of schools. Analyzing the teaching quality is an important means of teaching evaluation. Using advanced technology to analyze the data generated in the teaching link at multiple levels and angles, and using the analysis results to assist teaching decision-making is an inevitable requirement to ensure the teaching quality, improve the quality of students and teachers' comprehensive ability [1]. Analyzing and using these association rules can not only improve the teaching level of teachers, but also provide decision-making reference for teaching management departments, so as to improve the quality of education and teaching. On the one hand, it is not easy to find and use useful information in a huge number of databases so as not to waste available information; On the other hand,

information is related, which requires us not to stay on the surface of information. We should dig out the connections between information and find out the potential knowledge hidden behind information, to maximize the use value of information [2]. This paper first expounds certain basic theoretical knowledge of data mining and data warehouse, and focuses on the basic theory of association rules and classic Apriori algorithm. Firstly, combined with the actual needs of higher education evaluation, a multidimensional quality evaluation structure based on association data mining is provided. Then, aiming at the shortcomings of Apriori algorithm, which needs to traverse the data set many times, low efficiency and long time-consuming, an optimized Apriori algorithm is proposed. The frequent set of the scheme is combined with the database to find the final option of the user. Finally, the implementation process of the algorithm is given, and the wild goose array is carried out in the actual teaching case. The test and analysis results show that the prediction results of the mining results are good, which is helpful to improve the teaching quality and work level of colleges and universities as a whole.

2. Principles of Association Data Mining

2.1 Basic Concepts of Association Rules

The association model mainly describes the closeness or relationship of a group of data items. The mining of association rules can be formally defined as: let $i=\{i_1, i_2..., i_m\}$ be a set of all items, in which $i_k(k=1,2...,M)$ is called an item. The collection of items is called itemset, and the itemset containing K items is called k-itemset. A transaction T (transaction) is an itemset, which is a subset of I. each transaction is associated with a unique identifier TID. Different transactions together form transaction set D, which constitutes the transaction database for association rule discovery [3].

If itemset $X \subseteq T$, traction T supports itemset X. Then the association is the implication of following form: $X \Rightarrow Y$ where $X \subset I$, $Y \subset I$ and $X \cap Y \neq \phi$.

Association rule attributes have two main parameters: support and credibility Support is an important measure, because rules with low support may only appear by chance. From a business perspective, rules with low support are mostly meaningless. Therefore, support is usually used to delete meaningless rules. In addition, support also has an expected property, which can be used for the effective discovery of association rules. Confidence is reliable when reasoning through rules. For a given rule $X \Rightarrow Y$, the higher the confidence, the greater the possibility that y is included in the transaction of X. Confidence can also estimate the conditional probability of Y under a given X. The results of correlation analysis should be interpreted carefully. The inference made by association rules does not necessarily imply causality. It only means that the items in the pre and post rules obviously appear at the same time. On the other hand, causality requires knowledge about the attributes of causes and consequences in the data, and usually involves long-term relationships

2.2 Apriori Algorithm

In order to solve the problem of high frequency transmission electronic music, emphasizes the high frequency signal components are needed to produce the transmission spectrum is more equal modulation index, namely the compensation for the high frequency component of the input signal processing, the processing method is pree

Apriori is the core of association rule mining algorithm [4-8], and the specific algorithm is described as follows:

Algorithm Apriori: use layer by layer iteration to find frequent itemsets Input: transaction database d; Minimum support threshold.

Output: frequent itemset L in D.

method:

1) L1 = find_frequent_1_itemsets(D); // Find the set L1 of frequent 1-itemsets

2) For (k = 2; LK-1 $\neq \emptyset$; k++) {// generate candidates and prune

3) Ck = aproiri_gen(Lk-1,min_sup);

4) For each transaction t $\in d\{// \text{scan D for candidate count}\}$

5) Ct = subset(Ck,t); // Get a subset of T

- 6) for each candidate $c \in Ct$
- 7) c.count++; // Support count
- 8) }

9) Lk={c \in CK | c.count \geq min | sup} / / returns the set of items in the candidate set that are not less than the minimum support

10) }

11) return $L = \bigcup kLk$; // All frequent sets

3. Evaluation System of College Teaching Quality Based on Association Data Mining

3.1 System Process and Basic Framework

Most components of the system in this paper constitute a data warehouse system. This system starts from the existing operating system, and part of the background processing is set up to support the data warehouse, so as to access and use the user tools of the data in the data warehouse. The overall framework of the system is shown in Figure 1.



Figure 1. Overall structure of the teaching quality evaluation system.

The underlying architecture of the system is mainly composed of three parts: data source, data extraction module and data source monitoring module. Functionally, it supports role-based login, teacher information management, student management, administrator review, etc. Through the data exchange platform, the teaching link data, student evaluation data, teaching practice data and other information are extracted in real time to make the teaching evaluation indicators diversified and distributed. Get through ERP, OA, CRM and other different business information systems, integrate and summarize university data, and use data visualization to meet the needs of different people for complex analysis, related query, real-time processing and construction cost control [9].

3.2 Data Acquisition and Processing

Data acquisition is a process to meet the needs of statistics, analysis and mining, collect and obtain various data, and realize the complete collation of large-scale evaluation data [10]. The data acquisition steps of the teaching quality analysis and evaluation system are shown in Figure 2. The teaching collection module collects data from the outside of the system and inputs it to an interface inside the system. The content of data collection includes teaching quality information data and teacher file data.



Figure 2. Data acquisition process.

Data preprocessing refers to certain processing of data before the main processing. For example, before converting or enhancing most of the geophysical areal observation data, the irregular distribution survey network is first converted into a regular network through interpolation, which is conducive to computer operation. The main tasks of data preprocessing include:

(1) Data audit can be divided into accuracy audit, applicability audit, timeliness audit and consistency audit;

(2) Data screening, and the errors found in the audit process should be corrected as far as possible;

(3) Data sorting: arrange the data in a certain order.

Through the above data preprocessing operations, the system has a complete,

clean and suitable data output for mining, to ensure that the subsequent data mining can produce useful rules.

3.3 Improved Apriori Scheme

According to the analysis above demand, this article proposes an optimized Apriori algorithm. The principle idea is to add a database D_k when creating the frequency itemset L_{k-1} each time. D_k is used to save k-frequent set and the sets of all transactions belonging to it, which is recorded as E_i , $E_i = \{t_1, t_2, ..., t_q\}$. Therefore, when L_{k-1} connects C_k which is counting the support, the scanning of original databased D is no more needed. E_i will be scanned instead and their intersections of each sub-set can be acquired. The number of transactions in the intersections is the support number of such candidate itemset. Finally, the candidate set with minimum support is deleted to leave the k-frequency itemset.

The improvement process of this process is shown in Figure 3.



Figure 3. A case of improved Ariori algorithm

4. Empirical Analysis

In the experiment, 1000 teachers were selected as sample data, and the evaluation indicators include teacher number, name, gender, education background, professional title, teaching age and early evaluation level in the past three years. The system registers and discovers system resources through basic configuration, establishes resource grouping, restricts the use of system resources during the execution of data tasks through restriction configuration, and specifies the coping strategies and rules for the change of system resource status and the occurrence of events during the execution of tasks through policy configuration. The results from each filled data set are synthesized to produce the final statistical inference, which takes into account the uncertainty caused by data filling. The consolidated data is shown in table 1.

Code	Name	Sex	Birthdate	Education	Title	Teaching age	Teaching research projects	Quality
1	Kang Shi	Male	1979.4	Doctor	associate professor	20	Provincial level	Excellent
2	Hang Long	Female	1986.10	Master	professor	14	Municipal level	Good
3	Li Sizhen	Female	1980.3	Doctor	lecturer	8	School level	Moderate
4	Ma Hu	Male	1975.4	Master	associate professor	12	Provincial level	Excellent
5	Wang Gexi	Female	1985.2	Master	lecturer	6	Provincial level	Excellent
6	Tan Qian	Male	1980.3	Doctor	professor	9	School level	Bad
7	Bi Wulu	Female	1984.1	Undergraduate	associate professor	5	None	Good
8	Xi Fanfan	Female	1982.2	Master	lecturer	18	Municipal level	Moderate
9	Si Xiang	Male	1988.12	Master	lecturer	4	School level	Excellent
10	Wu Shuang	Male	1979.11	Doctor	professor	11	Municipal level	Moderate

Table 1. Teacher information and evaluation level by normalization

If existing multi table data sources are defined using physical layer joins, they will still work as before. We associate the logical table with the migrated table, but we cannot downgrade the logical table. If you want to add more tables to the data model, you can consider changing the data source to use relationships. In order to realize these functions, ETL tools will expand some functions, such as workflow, scheduling engine, rule engine, script support, statistical information, etc. Create the required triggers on the table to be extracted [11]. Generally, three triggers of insertion, modification and deletion are established. Whenever the data in the source table changes, the changed data is written into a temporary table by the corresponding trigger. The extraction thread extracts data from the temporary table, and the extracted data in the temporary table is marked or deleted. The specific process is shown in Figure 4.



Figure 4. ETL preparation data

Apriori algorithm is used to find all frequent itemsets, that is, itemsets whose support is not lower than the threshold set by the user; The rules that satisfy the minimum trust degree of users are constructed by using frequent itemsets; Then, L1 is used to generate the candidate itemset C2, and the items in C2 are determined. The frequent 2-itemset is mined, and the cycle continues until no more frequent k-itemsets are found; According to this method, the whole frequent data item set M is finally obtained. After data mining, the preliminary association rules that can be obtained are shown in table 2.

Rule	Age	Title	Education	Confidence level	Support degree
А	21-30			13.7	16.2
В	31-40			22.1	29.8
С	41-50			44.5	40.5
D	51-60			49.8	22.4
E		associate		12.3	18.6
		professor			
F		professor		19.7	23.1
G		lecturer		35.6	37.5
Н			Doctor	39.8	40.1
Ι			Undergraduate	16.4	72.4
J			Master	61.7	45.6

 Table 2.
 Preliminary association rules

According to the evaluation index, the basic state database is established through the collection of internal business system and external Internet data of colleges and universities, including basic functions such as data collection and verification, data adjustment and reporting, data retrieval, data reporting, data reuse, and son on, to provide data support for the evaluation and analysis of teaching quality From the above analysis, it is not difficult to see that teachers' academic qualifications and teaching attitude have a strong correlation with the evaluation score. Data mining can timely and effectively monitor the quality of school classroom teaching, stimulate teachers and students to fear the classroom together, solve classroom problems such as scripted teaching and less interaction, and improve the teaching quality of the training process.

5. Conclusion

In the face of the continuous improvement of the requirements of colleges and universities for classroom teaching quality, a teaching evaluation system integrating evaluation management, classroom evaluation, student evaluation of teaching, teacher evaluation of learning, questionnaire survey, evaluation center, mobile evaluation of teaching and so on is of great importance. This paper mainly analyzes the evaluation method of college teaching quality based on association data mining, and uses the improved Apriori algorithm to mine frequent patterns of teaching evaluation data, so as to improve the evaluation effect. Through the specific case analysis, it shows that this scheme has better operation efficiency and speed than the traditional management rules, and can provide strong data support for the system, so as to better serve the teaching evaluation of colleges and universities.

References

- Guo X L, Guo P, Li F. Implementation of Data Mining Technology in analysis and appraisal system of teaching quality. Journal of Northeast Dianli University, 2006, 26(3): 70-73
- [2] Xie Qing, Wu Xiaojun The application of data mining technology based on association rules in improving course teaching quality China Science and education innovation guide, 2009, 5: 175-176
- [3] Ling L. Application of Data Mining Based on Association Rules in the Teaching Evaluation System of University. Journal of Guangdong Polytechnic Normal University, 2008, 12: 41-44
- [4] Xie Q L. Data Mining in Teaching Quality Evaluation Based on Association Rules. Modern Computer, 2008, 6: 25-28
- [5] Jepsen D M, Varhegyi M M, Teo S T. The association between learning styles and perception of teaching quality. Education and Training, 2015, 57(5):575-587
- [6] Abdurrahman, Jamil A, Anwar S, et al. ASSOCIATION BETWEEN QUALITY TEACHING AND LEARNING OUTCOMES. Academic Research International, 2012, 1 :664-671
- [7] Kupersmith, Joel. Quality of care in teaching hospitals: a literature review. Academic Medicine Journal of the Association of American Medical Colleges, 2005, 80(5):458-66
- [8] Torre D M, Simpson D, Sebastian J L, et al. Learning/Feedback Activities and High-Quality Teaching: Perceptions of Third-Year Medical Students during an Inpatient Rotation. Academic Medicine, 2005, 80(10):950-954
- [9] Oni O D, Aina O, Ojini F, et al. Quality of life and associated factors among poststroke clinic attendees at a University Teaching Hospital in Nigeria. Nigerian Medical Journal Journal of the Nigeria Medical Association, 2016, 57(5):290-298
- [10] Cj A, Japm A, Mpl B, et al. Management of male anterior urethral strictures in adults. Results from a national survey among urologists in Spain - ScienceDirect Actas Urológicas Españolas (English Edition), 2020, 44(2):71-77
- [11] Zhang Y, Tian Y, Yao L, et al. Individual differences matter in the effect of teaching presence on perceived learning: From the social cognitive perspective of self-regulated learning. Computers & Education, 2022, 179:104427