Research on Information of Elective Computer Courses for College Students Based on Data Mining Technology

Xi-hong ZHOU
School of Computer and Information Technology,
Northeast Petroleum University, Daqing 163318, China

Keywords: Data mining, Association rules, Data warehouse, Computer optional course.

Abstract. This paper briefly introduces the data mining technology, and in detail analyzes association rule mining algorithm and the application of computer optional course in higher education, which includes four steps: data preparation and selection, data preprocessing, selection of the mining algorithm and the implementation, description of the mining results, acquires some useful rules information and obtains certain good application results, these information can assist related departments to allocate teaching resources reasonably and make corresponding decisions.

Introduction

Public elective courses are interdisciplinary elective courses for all students in the university, aiming at satisfying students' interests and hobbies, broadening students' scope of knowledge, improving comprehensive quality and cultivating students' innovative ability [1]. Along with society's emphasis on computer technology, the elective computer courses, as the highly applied technical courses for all arts and sciences students, play an important role in training students to process information, so there are more and more computer elective courses, the same time more and more elective people. As a result, the information about computer elective courses has also increased dramatically. For huge amounts of data and information, data mining technology is used to obtain useful information from a large number of elective courses information, find out the information hidden behind the data, mine students' preference for computer elective courses, the ultimate goal is to improve the teaching quality of computer elective courses, at the same time, make elective courses play a real role in improving students' practical ability and increasing students' employment opportunities.

Data Mining Technology

Data Mining (DM) also known as Knowledge Discovery in Database (KDD). Specifically, data mining is the process of processing data in a database, including extract implicit, previously unknown, but potentially useful information from large, incomplete, noisy, fuzzy and random data[2]. In short, it can automates the detection of relevant patterns and find relevant useful data information in a database. By handling a large amount of daily business data, some valuable knowledge is extracted from these data or information, this improves the utilization rate of information, and raw data is the source of knowledge.

Association rule mining algorithm is an important data mining methods. The data analysis process of the DM can be divided into four steps: preparation and selection of data, preprocessing of data, selection and implementation of mining algorithm and description of mining results. Data mining is mainly about association rules and cluster analysis. The purpose of association rules is to find the frequent patterns, associations, dependencies, or causal structures that exist between sets of items or objects in a data warehouse. Association rule is to find the correlation or interdependence between one thing and other things.[3]. The concept was first proposed by Agrawal R et al.[4]. The so-called association rule refers to the relationship between objects.
Application of Data Mining in the Information Analysis of Elective Computer Course

The basic data is a large number of data in student elective system, which needs to be pretreated. In order to be able to explore different levels of information, different granularity needs to be considered, in addition, metadata and storage methods need to be further analyzed before data mining. We first analyzed the structure of the original database and then determined the subject of the data warehouse to be built:

1. Analyze the number of elective courses to determine the popularity degree of the courses.
2. Analyze the course credit of each course and determine how much credit students should consider when choosing courses.
3. Analyze the course selection rate and find out the students' satisfaction with the current course arrangement.
4. Analyze the situation of students choosing courses and understand the situation of students choosing current courses.

Data Preprocessing

A large number of data in the course selection system are basic data, which should be preprocessed. Data preprocessing includes three steps: data cleaning, data integration and data transformation.

After analysis and consideration of various requirements, the corresponding data warehouse is established.

1. Data cleaning
   The raw data in the data table was cleaned up, the data items that are vacant in the course selection information of students are relevantly searched and compared, and filled with the most "approximate" data. Redundant duplicate data and noise data are deleted. In the system, the data sources involved are: students' basic information, teachers' basic information, students' course selection information, curriculum information, teaching evaluation information and students' score sheets. In these information tables, there are many fields that are not needed in this mining, for example, the information of students' place of origin and age in the basic information table, and the information of teachers' basic information about their work units, etc. So the method of deleting redundant data is applied to delete redundant fields. Because the information of computer elective courses is analyzed, the information of other elective courses is also cleaned up by deleting directly. In the data of the course selection system, some students choose two identical courses at the same time, which is not allowed. We treat it as noise data.

   Data reduction: The number, school year, semester, course serial number, IP, operation time and operation type in the course selection database are removed; the age and gender in the teacher information are removed; the student number, name, specialty, grade, affiliated department, age, etc information are retained in the student information, and other information is removed; the course number, school number and achievement are only retained in the performance information.

   Only the student's number, course name and examination result field can be used to mine the correlation between courses. Finally, there are 11 specialized courses for data mining.

2. Data integration
   Design a consistent data stored procedure. Considering the needs of mining applications, a data storage model is established to integrate data from different data sources (MS Excel 2003, MS Access 2003, MS Sql sever 2000). After a series of data cleaning and data integration processes, data tables as shown in Table 1 can be extracted from selected data source tables.

3. Data generalization
   According to the requirement of data mining, the data items are transformed and the data are discretized. The purpose of this excavation is to analyze the number of elective courses of various courses and determine the popularity degree of heat and cold of various courses of computer elective courses. In the course data, classify the computer elective courses: K1-computer hardware; K2-computer drawing and animation production class (such as PHOTOSHOP, FLASH); K3-programming class (such as C# programming, C++ programming); K4-web page and database
design class; K5-computer network technology class; Similarly, the percentage system of performance in the course data is also discretized by Bin method to get 1 - (90-100); 2 - (80-89); 3 - (70-79); 4 - (60-69); 5 - (0-59).

Select appropriate fields in the database, A for courses; B for grades; C for departments; D for gender; E for teachers' titles (1-professor, 0-associate professor); F for credits; G for grades. Determine the range of values for each field such as A:3-3 grade, 2-2 grade, 1-1 grade; C:01-geoscience, 02-petroleum engineering, 03-petrochemical, 04-machinery, 05-electrical automation, 06-civil, 07-computer, 08-management, 09-electronic science, 10-mathematics, 11-humanities, 12-foreign languages; D:0-male, 1-female; E: the same as C; F:(0-1), (1-2) (2-3); G: K1-computer hardware; K2-computer graphics and animation production class (such as PHOTOSHOP, FLASH); K3-programming class (such as C# programming, C++ programming); K4-web page and database design class; K5-computer technology overview class; H: 1-excellent, 2-good, 3-medium, 4-pass, 5-fail; discrete partial data as table 1 below.

Table 1. Discretization of partial selection course data.

| A   | B  | C   | D  | E  | F  | G  |
|-----|----|-----|----|----|----|----|
| K1  | 1  | 11  | 0  | 1  | 3  | 1  |
| K1  | 3  | 6   | 0  | 1  | 2  | 3  |
| K2  | 2  | 2   | 0  | 1  | 2  | 2  |
| K5  | 1  | 5   | 1  | 0  | 2  | 3  |
| K3  | 3  | 9   | 0  | 0  | 3  | 1  |
| K2  | 2  | 3   | 1  | 0  | 3  | 2  |
| K4  | 2  | 3   | 0  | 0  | 3  | 2  |
| K5  | 3  | 8   | 1  | 1  | 3  | 3  |
| K4  | 1  | 6   | 0  | 1  | 3  | 2  |

**Mining Association Rules by Apriori Algorithms**

Data analysis and mining mainly extracts the corresponding decision rules from the selected analysis data for the school management to make corresponding decisions. FP-tree algorithm is used for data mining [5]. The specific steps are as follows:

1. Input database, minimum support minsup = 10%, minimum confidence minconf = 40%;
2. Scanning the database and finding 1-candidate set, C1;
3. Choose the item set whose support is greater than the minimum support from the candidate set Ck, namely Lk;
4. Extend Lk into K+1 dimension set, namely Lk+1;
5. Scanning the database, extracting the item sets larger than the minimum support in Ck+1, namely Lk+2;
6. When Lk+i is not empty, Lk+i is taken as the candidate set of next association, namely Ck+i, and step 3 is changed.
7. Calculate the confidence of each association in the maximum strength set L, and extract rules with confidence greater than minimum confidence and support greater than minimum support, that is, effective association rules.

The process and results of association mining are shown in Table 2.

Table 2. Partial mining results.

| Rule | Support degree(%) | Confidence degree(%) |
|------|-------------------|----------------------|
| R1   | 23.4              | 54.3                 |
| R2   | 24.6              | 56.3                 |
| R3   | 17.7              | 43.6                 |
| R4   | 29.4              | 48.6                 |
| R5   | 17.3              | 49.7                 |
| R6   | 20.6              | 49.2                 |
Summary
Based on the accumulation of a large amount of data information in the computer course selection system of colleges and universities, this paper preprocesses the original data from the perspective of data mining, combines data warehouse with data mining technology, and uses the FP-tree algorithm of association rules to mine a large number of data, and obtains some useful association rules. This research also puts forward some good suggestions for the actual teaching management.

Acknowledgement
This research was financially supported by the Heilongjiang Institute of Higher Education Committee.

References
[1] Zhang Xi, How to make public elective courses become the highlight of university teaching, J. Research on Science Education. 2 (2007) 156.
[2] Lu Qicheng, Zou Ping, Research and application progress of data mining, J. Journal of Kunming University of Technology. 27 (2002)62-66.
[3] Jia Caiyan, Ni Xianjun, Review of Association Rule Mining, J. Computer Science. 30 (2003)145-148.
[4] Ma Tinghuai, Zhang Haisheng, Zeng Zhengshank, Mining Association Rules with Concluding Domain, J. Computer Engineering. 29 (2003)16-17.
[5] Liu Duyu, Yang Jinhao, Zhong Shouming, A review of association rule mining, J. Journal of Chengdu University.25 (2006)20-23.