Analysis of Mathematics Curriculum Ability Development Based on Computer Data Mining

FengMing Zhao*
Sichuan Vocational and Technical College, Suining, Sichuan, China, 629000

*Corresponding author e-mail: Fengmingzhao@163.com

Abstract. In this paper, the data mining techniques is used to analyze the mathematics curriculum ability development. Firstly, a data warehouse of students' academic achievement is established. The comparative concept description method in data mining techniques is used to analyze various abilities of students of different genders from different majors over the years, and the corresponding reform measures for the teaching contents and approaches are proposed.

Keywords: Data Mining, Computer Mathematics, Ability Development, Teaching Research

1. Introduction
In recent decades, because of the broad application of Computer Mathematics in computer science and technology and many other subjects[1], computer mathematics has developed rapidly. As the mathematical basis of computer science and technology, computer mathematics is a required curriculum for computer science and technology majors. At the same time [2-3], computer mathematics is a branch of modern mathematics, which has the same important status as some of the primary curriculums in traditional applied mathematics[4]. Therefore, the curriculum of computer mathematics plays a vital role in the curriculum system of computer science and technology and mathematics-related majors [5].

Data mining is a process of extracting hidden, unknown, and potentially useful information from a large number of incomplete, noisy, fuzzy, and random data[6]. The main functions of data mining are concept description, comparative concept description, association analysis, classification and prediction, and cluster analysis. In this paper, the comparative concept description method is mainly used to compare the data characteristics of the two types of study and give a summary of the comparison results. The two types of data set analyzed include target and comparative data set. The main steps are data preprocessing, data warehouse construction, target class and comparison class
determination, attribute generalization and attribute deletion, class comparison query, concept feature interpretation, and evaluation.

2. Data preprocessing and data warehouse establishment

“The so-called computer data mining is what people can do based on a large number of data. Meanwhile, it is also the source for people to get new cognition and improve creativity. Most of the computer data mining can also change the relationship between citizens and governments and the institutions of market organizations.” the definition of computer data mining, The Internet data center of the United States also has a corresponding meaning. It believes that computer data mining is a new technical framework to obtain similar value from a large amount of data through the approaches of discovery, capture, and analysis. In a word, computer data mining is a large-scale, vibrant and diverse data collection

For simplicity, only mathematics teaching in higher vocational colleges is discussed in this paper:

\[ S = \{(x_1, y_1), K, (x_n, y_n)\} \in \mathbb{R}^n \times \{+1, -1\} \]  

(1)

Where \((x, y)\) indicates optimization problem of regularized dynamic functions, which can be expressed as follows:

\[ \min_{w \in \Omega} \Phi(w), \Phi(w) = r(w) + \frac{1}{n} \sum_{i=1}^{n} f_i(w) \]  

(2)

Where \(w \in \mathbb{R}^n\), \(r(w)\) stands for a regularization term, a dynamic function \(f_i(w)\) from data \(x_i\). Caused by.

Assuming that the dynamic function is dynamic, the objective function \(\Phi(w)\) has strong convexity. Many researchers have studied the solution of the optimization problem (2), among which the gradient descent method is the simplest first-order optimization method

\[ w_{i+1} = w_i - \eta_i g(w_i) \]  

(3)

Where \(g(w_i)\) stands for the total gradient of all data objective functions \(\Phi(w)\) at \(w_i\), \(\eta_i\) is the learning step.

The data source of this paper comes from the final examination results and the computer performance of nearly 1080 students majoring in information and Computing Science in the past three years. The main attributes are student number (ID), name (name), gender (gender), major (Major), score (score). The scores are composed of the test score, basic set, algebra, graph theory, logic theory, program and total scores.

For the total score, in order to facilitate the concept description, the continuous data is discretized and the concept is stratified. Scores are divided into intervals \{100-90, 89-80, 79-70, 69-60, 59-0\}, and data are grouped according to \{excellent, very good, good, normal, fair\}. Thus, a two-dimensional (2D) data warehouse is established. The DMQL data mining primitive can be expressed as follows:

The cub discretemath is defined [ID, name, gender, major, total_score].

\[ \text{total}\_\text{score}=\text{sum}(\text{score}); \]
define dimension score (test_score, program_score);
define dimension test_score (basic_set, algebra, graph_theory, logic_theory).

In the total score, the proportion of the examination score is 80%, and the proportion of the computer score is 20%. For the scores of set, algebra, graph theory and mathematical logic in the examination results, the proportion is 15%, 30%, 30%, and 25%, respectively. For the general questions with multiple chapters, the proportion of each chapter is extracted and summarized. The full score of each chapter is 12, 24, 24, 20. Table 1 is a two-dimensional view after data preprocessing.

**Table 1. 2D view after data preprocessing**

| ID   | Name          | Gender | Major                | Basic_set | Algebra | Test_score | Graph_theory | Logic_theory | Program_score | Total |
|------|---------------|--------|----------------------|-----------|---------|------------|--------------|--------------|---------------|-------|
| I0201| Li Fang       | female | computer science     | 12        | 20      | 22         | 16           | 13           | 83              |       |
| I0202| Wang Zhigang  | male   | computer science     | 11        | 14      | 24         | 14           | 17           | 80              |       |
|      | ...           |        | ...                  | ...       | ...     | ...        | ...          | ...          | ...            | ...   |
| average| Li Li Jie    | female | computer science     | 9         | 14      | 17         | 14           | 15           | 69              |       |
| I1101| Jian Jun Bao  | male   | Information and      | 12        | 20      | 22         | 16           | 13           | 83              |       |
|      | ...           |        | Computing Science    |           |         |            |              |              |                 |       |
| I1102| ...           |        | ...                  | ...       | ...     | ...        | ...          | ...          | ...            | ...   |
|      | Li Fang       | female | Information and      | 11        | 14      | 24         | 14           | 17           | 80              |       |
|      | ...           |        | Computing Science    |           |         |            |              |              |                 |       |
| average| Wang Zhigang | male   | Information and      | 10        | 18      | 16         | 16           | 12           | 70              |       |
|      |               |        | Computing Science    |           |         |            |              |              |                 |       |

3. Description of category concept

3.1. Professional Category Analysis

a) The target and comparison class are determined as two different majors of computer science and Information Computing Science in the attribute major. b) For attribute generalization and attribute deletion, the attributes of name and gender are not considered because the characteristics of two different specialties are analyzed. c) Carry out data mining query processing and relevant dimension analysis. d) Carry out synchronous generalization on the target class to form a comparative relationship, as shown in Table 2.

**Table 2. Professional category analysis**

| Category                | Generalization relation of object class (Computer Science) | Generalizations of comparative classes (Information and Computing Science) |
|-------------------------|------------------------------------------------------------|----------------------------------------------------------------------------|
|                         | Average | Excellent rate/% | failure rate/% | Average | Excellent rate/% | failure rate/% |
| Basic_set               | 9       | 23               | 5              | 10      | 24               | 4              |
| Algebra                 | 14      | 5                | 19             | 18      | 10               | 11             |
| Graph_theory            | 17      | 17               | 8              | 16      | 14               | 10             |
| Logic_theory            | 14      | 12               | 7              | 16      | 13               | 6              |
| Program_score           | 15      | 14               | 5              | 12      | 11               | 8              |
| Total                   | 69      | 14               | 9              | 70      | 12               | 7              |

Table 2 shows that compared with information and computing science, students majoring in computer science and technology have higher scores in graph theory and computer application. This
suggests that the students in this major have stronger abilities in mathematical application and computer programming, while students majoring in algebra system and mathematical logic have poor mastery of the contents of the chapters because this chapter is mainly about formula proving, logical reasoning and mathematical operation, where a sound mathematical foundation, a strong ability of mathematical operation and logical reasoning are required. Hence, teachers need to supplement the basic mathematical knowledge properly in the teaching process of computer science and technology, and at the same time, reduce some confusing contents of mathematical proof properly. However, students majoring in information and computing science have a good foundation in mathematics because they have studied mathematics analysis, advanced algebra, and other curriculums, but they are weak in mathematics application ability. In the teaching process, they should fully cultivate their knowledge in this area and strengthen their guidance in programming and computer application.

3.2. Gender category analysis

a) It should be ensured that target class and the comparison class are male and female in the attribute gender. b) For attribute generalization and attribute deletion, as the characteristics of two different majors are analyzed, the attributes of student number, name, and major are not considered. c) Carry out data mining query processing and relevant dimension analysis. d) At the same time, generalizations are made on the target classes to form a comparing relationship. As shown in Table 3.

| Generalizations of objective class (male) | Generalizations of the comparisonive category (female) |
|------------------------------------------|------------------------------------------------------|
| Average | Excellent rate/% | failure rate/% | Average | Excellent rate/% | failure rate/% |
| basic_set | 8 | 20 | 8 | 10 | 24 | 3 |
| Basic_set | 12 | 6 | 23 | 18 | 11 | 8 |
| Algebra | 17 | 17 | 9 | 15 | 14 | 10 |
| Graph_theory | 13 | 8 | 15 | 15 | 5 |
| Logic_theory | 15 | 13 | 7 | 13 | 11 | 10 |
| Program_score | 65 | 8 | 11 | 73 | 14 | 3 |
| Total | | | | | |

Table 3 shows that there is a big difference between the scores of male and female students in the total scores and the contents of each chapter. Generally speaking, girls' performance in this curriculum is better than boys', girls' excellent rate is high, and the failure rate is low, while boys' failure rate is high, and the excellent rate is low, and much basic knowledge is not properly mastered, which reflects the lack of learning attitude and enthusiasm. Female students have a better understanding of conceptual content, but a lower ability of knowledge application and computer operation, which is mainly reflected in the lower scores of graph theory and computer operation. Hence, in the curriculum of teaching, teachers should not only consider the common problems of students but also pay attention to the gender differences of students, to improve the quality of students scientifically and comprehensively.

4. Reform of teaching contents and approaches

(1) In the teaching content, given the traditional basic content of the set theory, which is introduced in
detail and slightly different in almost every curriculum of information and computing science major, this part of the content focuses on solving practical application problems with the method of set theory. For this part of content, computer major focuses on Deepening the understanding of basic concepts. In the part of mathematical logic, by transforming proposition formula and logic formula into normal form, the logical calculation abilities of students can be strengthened. Through the study of propositional logic reasoning theory, we can achieve the goal of cultivating students' rigorous thinking habits and improving their logic reasoning ability. This part of graph theory focuses on the understanding of basic concepts and the handling of practical problems. In order to improve this part of the theory, we should understand the research approaches of graph theory through the understanding of relevant theorems and their proving ideas. In dealing with this part of the algebra system, the major of information and computing science can reduce the number of learning hours, because some of the algebra system contents in computer mathematics are covered in the required curriculum catalog listed in its teaching plan. For computer science and technology majors, we focus on the understanding of group theory in this paper, especially the concepts of algebraic system, group, subgroup, cyclic group, transformation group, normal subgroup, etc. Through this part of the content of teaching, so that the computer professional students experience rigorous mathematical thinking approaches. In the reform of teaching content, we should not only focus on the prominent contradiction between the content and the short class hours, but also use our brains to deal with the related content of each chapter. We can use the method of drawing inferences from one instance to make it comprehensive, which saves time and effort, and achieves the effect of half the effort.

(2) In terms of teaching approaches, due to the difference of teaching objects and development objectives, it is imperative to establish an optimized teaching mode that adapts to certain conditions in the teaching process. During the teaching process, attention should be paid to help students understand the relationship between the knowledge of Computer Mathematics and related subjects, the crucial basic position and role of Computer Mathematics in these subjects, guide students consciously to use the theory to contact practical problems, improve the pertinence and usability of Computer Mathematics for professional curriculums, and make students feel “useful” after learning.

5. Conclusions
The data on computer mathematics final examinations and computer performance for years are used to explore and identify the knowledge mastering and ability development of students in the two majors for this curriculum based on the data mining techniques, which is conducive to performing the reform of teaching contents and approaches in a targeted manner. In this way, the students can master the knowledge of computer mathematics curriculum more properly, and various abilities can be developed to lay a better mathematical foundation for the curriculum study and professional development in the future.

References
[1] Lee, Chun - Yi, & Chen, Ming - Jang. (2016). Developing a questionnaire on technology - integrated mathematics instruction: a case study of the ama development curriculum in xinjiang and taiwan. British Journal of Educational Technology, 47(6), 1287-1303.
[2] Mingqi Lv, Ling Chen, Zhenxing Xu, Yinglong Li, & Gencai Chen. (2015). The discovery of
personally semantic places based on trajectory data mining. Neurocomputing, 173, 1142-1153.

[3] Lai Hui, Shuai Li, & Zhou Zongfang. (2013). The model and empirical research of application scoring based on data mining approaches. Procedia Computer Science, 17, 911-918.

[4] Mei Jiang, Xindan Gan, Chaofeng Wang, & Zhuo Wang. (2011). Research of the intrusion detection model based on data mining. Energy Procedia, 13, 855-863.

[5] Meiyian Du. (2016). The application of data warehouse model based on the integration of emerging technologies in business data mining. Journal of Computational and Theoretical Nanoscience, 13(12), 9581-9585.

[6] Michael Nowlin. Henry James and the Making of Modern African American Literature[J]. The Henry James Review, 2018, 39(3):282-292.