On curriculum relevance of the information management and information system major in the age of big data

Weiwei Wu*, Xiqin Ao, Fan Li

College of Information Engineering, Anhui Xinhua University, Hefei, Anhui, 230031, China

*Corresponding author’s e-mail: wuweiwei@mail.ustc.edu.cn

Abstract. This paper aims to study the curriculum relevance of the major of information management and information system in the age of big data by analyzing the curriculum scores of students via Pearson correlation coefficient and Apriori algorithm employing the software IBM Statistical Product and Service Solutions (SPSS) Statistics 26.0. The correlations between different courses are presented from the multiple perspectives by statistical description. The results could provide a theoretical basis to improve the training program of information management and information system major.

1. Introduction

With the development of computer and information communication technology, the age of big data has arrived [1-3]. The application of big data is increasingly penetrating into all aspects of society. The big data serves social and economic development and improve people's lives [4]. The demand for big data talent is growing explosively, and there is a large gap now. As one of the majors closely related to big data application, the major of information management and information system (IMIS) should take the opportunity of big data development to reform the training program of professional talents for cultivating more compound talents satisfying social needs [5]. The students of IMIS major mainly studies the principles of management, economy, system of management information, computer application and program design, principle and application of big data, quantity analysis method and have a comprehensive use knowledge analysis and the basic ability to solve some problems.

In recent years, most of universities in China have established their own educational administration management systems, and accumulate a large number of teaching data, which often contains many pieces of information, such as: when a subsequent course is completed, what are the several prerequisite courses with the greatest impact on them? Does the score distribution of students with several similar ID numbers almost be same? If it is true, what causes it? What is the influence on the teaching effect of the sequence of several courses in a semester?

It is of great significance for us to analyze the students' achievements, dig out the potential laws, and apply the results to the teaching affairs with the help of the software IBM Statistical Product and Service Solutions (SPSS) Statistics 26.0. This process could be beneficial to revise the training plan, optimize the teaching contents, and reform the teaching methods. It also can improve the teaching quality.

The paper is organized in the following way. In Section 2, the data source and basic statistical description analysis are presented. The curriculum relevance analysis based on Pearson correlation coefficient [6, 7] is derived in Section 3. The following section gives the course relevance analysis
based on the Apriori algorithm [8] via the software IBM SPSS Statistics 26.0. Section 5 is devoted to summary.

2. Data source and basic statistical description analysis

2.1. Data source
This paper takes the examination scores of all courses of 2018 graduates of students of Anhui Xinhua University as the empirical analysis object. The data comes from the academic affairs system of the university, and is guaranteed.

2.2. Data pre-processing
The students' examination scores of courses usually use the five-level system or hundred-mark system due to the different course types and grading standards. In order to facilitate the following statistical analysis, the five-level system scores will be converted into the hundred-mark system, such as: 95 points for excellent, 85 for good, 75 for medium, 65 for passing and 55 for failing. The score of absent students is 0, which belongs to the missing value. In order to ensure the accuracy of the research, the score of absent students is deleted.

2.3. Statistical analysis of original data
We have analyzed the distribution of the scores of the course examinations of the 2018's students of IMIS majors in the hundred-mark system using the software IBM SPSS Statistics 26.0. The results are shown in the following Table 1.

| Course                                    | Minimum | Maximum | Mean  | Standard deviation | Kurtosis | Deviation |
|-------------------------------------------|---------|---------|-------|--------------------|----------|-----------|
| Introduction to Information Management    | 53      | 93      | 71.18 | 8.231              | -.694    | .025      |
| Discrete Mathematics                      | 51      | 100     | 79.90 | 10.232             | -.438    | -.119     |
| Principles of Management A                | 56      | 95      | 70.23 | 9.525              | .008     | -.342     |
| Data Structure                            | 45      | 97      | 76.19 | 8.842              | -.093    | -.769     |
| Operating System                          | 60      | 99      | 75.56 | 12.846             | .061     | -.589     |
| System of Management Information A        | 65      | 100     | 80.04 | 7.264              | -.041    | .237      |
| ERP                                       | 61      | 98      | 77.73 | 9.903              | -.274    | -1.002    |
| Operational Research A                    | 55      | 98      | 79.39 | 11.563             | -.723    | .364      |
| College English                           | 69      | 88.5    | 80.5  | 5.075              | -.093    | -.69      |
| Computer Application and Program Design   | 54      | 97      | 83.31 | 5.543              | -.287    | -.003     |
| Ideological and Moral Cultivation and legal basis | 69 | 93 | 79.32 | 4.546 | .117 | .769 |
| College Chinese B                         | 73      | 98      | 87.49 | 4.453              | -1.007   | 1.762     |
| Marxism Basic Principle                   | 60      | 92      | 77.31 | 5.573              | -.452    | -.018     |
| Calculus                                  | 51      | 99      | 73.8  | 10.371             | .068     | -1.009    |
| Linear Algebra B                          | 43      | 100     | 76.29 | 15.835             | -.263    | -.529     |
| Probability theory and mathematical statistics B | 55 | 96 | 75.29 | 10.39 | .264 | -1.035 |
### Introduction to Mao Zedong Thought and Theoretical System of Socialism with Chinese Characteristics

| Course                                                   | Mean | Standard Deviation | Kurtosis | Skewness | Score |
|----------------------------------------------------------|------|--------------------|----------|----------|-------|
| Outline course of modern Chinese history                | 70   | 95                 | 80.37    | 5.101    | .639  |
| C++                                                      | 52   | 100                | 79.59    | 8.352    | -.214 |
| Human Resource Management A                             | 60   | 98                 | 78.57    | 8.372    | -.106 |
| Principle and Application of Database                   | 65   | 99                 | 82.86    | 9.834    | -.206 |
| Microeconomics B                                        | 40   | 95                 | 75.13    | 8.395    | .836  |
| Macroeconomics B                                        | 61   | 98                 | 76.98    | 15.394   | -.295 |
| Accounting Principles A                                 | 55   | 98                 | 77.38    | 10.402   | -.406 |
| Network A                                                | 62   | 96                 | 77.43    | 9.835    | -.491 |
| Logistics and Supply Chain Management                   | 68   | 99                 | 85.29    | 6.326    | -.683 |
| Information Economics                                    | 52   | 93                 | 69.39    | 9.295    | -.109 |

The above Table 1 tells us that the average scores of the courses always are 70-80, the standard deviation is less than 10 and the kurtosis and deviation are around 0. The data show that the most courses' scores subject to normal distribution and can reflect the students' mastery of the course. So it is reasonable to take the score data as the object of subsequent research and analysis.

### 3. Pearson correlation coefficient of curriculum

#### 3.1. Pearson correlation coefficient of computer courses

Correlation analysis is to describe the strength of the relationship between things through quantitative indicators. The research of this paper is about the correlation between the scores of various courses, which belongs to the continuous variable. Pearson correlation coefficient is used to investigate the correlation between two courses. The analysis results on Pearson correlation coefficient of computer courses using the software IBM SPSS Statistics 26.0 are shown in Table 2, which contains six courses as Computer Application and Program Design (denoted by A), C++ (B), Data Structure (C), Principle and Application of Database (D), Operating System (E) and Network (F).

**Table 2. Pearson correlation coefficient of computer courses.**

|       | A    | B    | C    | D    | E    | F    |
|-------|------|------|------|------|------|------|
| A     | 1    | .187 | .303 | .257 | .274 | -.006|
| B     | .187 | 1    | .315 | .363 | .405 | .355 |
| C     | .303 | .315 | 1    | .662 | .654 | .177 |
| D     | .257 | .363 | .662 | 1    | .739 | .086 |
| E     | .274 | .405 | .654 | .739 | 1    | .042 |
| F     | -.006| .355 | .177 | .086 | .042 | 1    |

Since course Network has little in relation with other computer courses, the scatter diagram of the other five computer courses as A-E will be given in the following figure.
Figure 1. Scatter diagram for computer courses.

It can be found from Table 2 and Figure 1 that there is a certain correlation between computer courses, and course Principle and Application of Database (D) have the strongest correlation with other computer courses. Table 2 shows that course Network (F) has relatively weak correlation with other computer courses.

3.2. Pearson correlation coefficient of specialized courses

Pearson correlation coefficients of specialized courses employing the software IBM SPSS Statistics 26.0 are shown in followed Table 3, which contain five courses as Introduction to Information Management (denoted by G), Principles of Management (H), ERP (I), System of Management Information (J) and Operational Research (K).

Table 3. Pearson correlation coefficient of specialized courses.

|     | G    | H    | I    | J    | K    |
|-----|------|------|------|------|------|
| G   | 1    | .511*| .314*| .389*| .298*|
| H   | .511*| 1    | .572*| .629*| .454*|
| I   | .314*| .572*| 1    | .531*| .505*|
| J   | .389*| .629*| .531*| 1    | .617*|
| K   | .298*| .454*| .505*| .617*| 1    |

Since course ERP is independent with other specialized courses, the scatter diagram of the other four specialized courses as G, H, J and K will be given in the following figure.

Figure 2. Scatter diagram for specialized courses.
Table 3 and Figure 2 give the results as there is a certain correlation between each other of the kind of specialized courses, as courses Operation Research has the strongest correlation with course Management Information System, and the weakest correlation with course Introduction of Information Management.

4. Curriculum relevance analysis based on Apriori algorithm

Apriori algorithm is the most typical and widely used algorithm in association rules area. With the help of SPSS molder, Apriori algorithm is used to mine the relevance between courses. The results of correlation analysis of some courses are shown in following Table 4.

| Item 1                                      | Item 2                                      | Support (%) | Confidence (%) |
|---------------------------------------------|---------------------------------------------|-------------|----------------|
| Dynamic Website Design and Architecture B   | Course Design of Dynamic Website Design and Architecture B | 41.56       | 99.00          |
| @.NET                                       | Dynamic Website Design and Architecture B   | 19.23       | 84.00          |
| Principle and Application of Database       | Operating System                            | 21.74       | 71.99          |
| Course Design of @.NET                      | Course Design of Dynamic Website Design and Architecture B | 27.93       | 73.57          |
| Operating System Principle                  | C++                                         | 13.28       | 67.20          |
| Data Structure                              | Operating System Principle                  | 14.85       | 66.50          |
| Data Structure                              | Principle and Application of Database       | 15.09       | 73.40          |
| Operating System Principle & Principle and Application of Database | C++                                         | 14.55       | 66.50          |
| Operating System Principle & Principle and Application of Database | Data Structure                              | 12.04       | 65.90          |
| Linear Algebra B                           | Discrete Mathematics                        | 19.85       | 98.00          |
| Calculus & Linear Algebra B                 | Probability theory and mathematical statistics B | 11.02       | 82.60          |
| Principles of Management A                  | System of Management Information A          | 3.60        | 95.00          |
| System of Management Information A          | ERP                                         | 8.00        | 60.00          |
| Operational Research A                      | ERP                                         | 8.25        | 40.00          |
| System of Management Information A          | Operational Research A                      | 8.10        | 41.80          |

Based on the results of data association mining in above Table 4, the following conclusions are worth to point out as:

1. The score of one course is closely related to its course design. If the course is well mastered, it will always achieve good results on the course design. For example, there are 36.36% of the students which have the excellent scores of Dynamic Website Design and Architecture B and the related course design at the same time;

2. There is a great correlation between each other of the computer courses since all the computer courses belong to the technology courses. When setting up the courses, we should consider whether the relevant prerequisite courses have been completed. For example, 18.18% of the students have excellent scores both on @. Net and Dynamic Website Design and Architecture B. 80% of the
students whose scores are excellent in @. Net are also excellent in Dynamic Website Design and Architecture B;

(3) Mathematics courses share great relevance because they all reflect students' mathematical thinking ability. For example, the proportion of students having the excellent scores in both Linear Algebra B and Discrete Mathematics is 18.18%. 80% of the students whose score of Linear Algebra B is excellent also have excellent score of Discrete Mathematics. The reason may be that Linear Algebra B is the preparatory course of Discrete Mathematics.

(4) The relevance between the specialized courses is relatively weak since the specialized courses reflect only one aspect of ability. For example, the proportion of students having excellent scores both in System of Management Information A and Principles of Management A is 7.27%. Only 50% of students having excellent scores in System of Management Information A are also excellent in Principles of Management A. It does not need to check the relationship between the two ones of the kind of courses.

5. Summary
Firstly, we collected the course scores in four years of 2018 graduates of IMIS major through the school educational administration system. Then the data preprocessing and basic statistical description analysis were completed. The results show that most of the course scores subject to normal distribution, which can reflect the students’ real mastery of the courses. The curriculum relevance of IMIS major via Pearson correlation coefficient employing the software IBM SPSS Statistics 26.0 contain the conclusion as there is a strong correlation between each other of the same kind of courses. Finally, the Apriori algorithm was used to analyze the curriculum relevance. It is found that the relevance between each other of computer courses and mathematics courses is relatively large, while that between each other of specialized courses is relatively weak. These conclusions can provide scientific basis for the curriculum of IMIS major and the reform of professional training program for cultivating more compound talents for the needs of the age of big data.

Acknowledgments
This work was supported by National Training Program of Innovation and Entrepreneurship for Undergraduates under Grant No. 201912216098 and Anhui Province Training Program of Innovation and Entrepreneurship for Undergraduates under Grant No. AH201812216067.

References
[1] Wu, C., Xue, G.L., Li, J., Yau, K.A., Qadir, J. (2019) Computational intelligence for Internet of things in the big data era (Part I). IEEE Comput. Intell. M., 14(4): 11-88.
[2] Wu, C., Xue, G.L., Li, J., Yau, K.A., Qadir, J. (2020) Computational intelligence for Internet of things in the big data era (Part II). IEEE Comput. Intell. M., 15(1): 22-23.
[3] Wei, W., Guizani, M., Ahmed, S.H., Zhu, C. (2020) Special section on integration of big data and artificial intelligence for Internet of things. IEEE. T. Ind. Inform., 16(4): 2562-2565.
[4] Karim, S., Soomro, T.R., Aqil Burney, S.M. (2018) Spatiotemporal Aspects of Big Data. Appl. Compu. System., 23(2): 90-100.
[5] Wu, D., Wang, X.Y., Shao, X.X. (2010) Research on the mode of college-enterprise cooperation on talents-training in information management and information system. Compu. Educ., 2(1): 49-66.
[6] Pearson, R. G., Dawson, T.P., Berry, P.M., Harrison, P.A. (2002) Species: a spatial evaluation of climate impact on the envelope of species. Ecol. Model., 154(3): 289-300.
[7] Dereceli, C. (2017) Examination of studying approaches of students at school of physical education and sports in terms of different variables. J. Educ. Training Stud., 5(4): 252.
[8] Schober, P., Boer, C., Schwarte, L.A. (2018) Correlation coefficients: appropriate use and interpretation. Anesth. Analg., 126(5): 1763-1768.