Data Based College Students' Comprehensive Quality Evaluation System

Li Li
School of Electronics and Information Engineering
Tongji University
Shanghai 201804, China
lili@tongji.edu.cn

Zhou Ou*
School of Electronics and Information Engineering
Tongji University
Shanghai 201804, China
18817366097@163.com
*Corresponding author

Shaoming Yao
School of Electronics and Information Engineering
Tongji University
Shanghai 201804, China
1333705@tongji.edu.cn

Qijun Chen
School of Electronics and Information Engineering
Tongji University
Shanghai 201804, China
qjchen@tongji.edu.cn

***Abstract*—As EQO(Essential Qualities Oriented) education has become consensus of society, the evaluation of college students' comprehensive quality stands at the vital position in the work of educational workers. First college students' comprehensive quality evaluation system is built based on some University's information system. Then framework of comprehensive quality evaluation system is designed. Finally, an application based on the framework is accomplish with the database technology and data mining technology. The complete of the application, which works as an auxiliary of related departments' decision-making, shows the process of realizing the framework.**

Keywords-comprehensive quality evaluation; data mining; Apriori; association rules

I. INTRODUCTION

With the rapid development of modern science and technology, society has higher criteria for talents. To adapt to the demands of new era, the Ministry of Education has paid more attention to the cultivation of college students' comprehensive quality. At present, the implementation of EQO education has become consensus of the whole society [1].

In the meantime, the information construction of colleges is advancing at an unprecedented speed. The campus local area network with advanced performance covers nearly all colleges. And the construction of data warehouse systems enhances the schools' management and decision-making ability [2].However, the amount of data stored by colleges are rising rapidly, which leads to a concern on how to use colleges' database system efficiently.

Big data era has a vital influence on our work, thinking and life. From Google's prediction of H1N1 to Taobao's accurate recommendation, big data era has been deeply into modern life. Data mining technology appears under such a background [3]. With data mining technology and related data analysis algorithms, we can deal with those data and find rules and patterns to help people with their decision-making.

In this paper, we mainly establish college students' comprehensive quality evaluation index system and framework of college students' comprehensive quality evaluation and complete the design by realizing an application which can assist related departments in their decision-making.

This paper is organized as follows: college students' comprehensive quality evaluation indexes are established in Section II. And the framework of comprehensive quality evaluation system is built in Section III. In Section IV, the content of the framework is implemented step by step, ultimately realizing the design of college students' comprehensive quality evaluation system by carrying out an application, which can be used as an auxiliary of related departments' decision-making. Section V concludes the work done and gives some outlooks to future work.

II. THE ESTABLISHMENT OF COLLEGE STUDENTS' COMPREHENSIVE QUALITY EVALUATION INDEX

A. Design Principles

To ensure the quality of comprehensive evaluation indexes, which is the basis of the comprehensive quality assessment, the system must follow some basic principles [4-5]:

1) Principle of comprehensiveness: Each evaluation index can comprehensively and accurately cover, and describe the characteristics of college students' comprehensive quality.

*2) Principle of hierarchy:*The principle of hierarchy reflects developmental evaluation thought.

3) Principle of representative: The optimized index can be in accordance with students' personality development and can typically reflect the college students' comprehensive quality.

4) Principle of independence: The index system must be independent of each other in the same level of indicators.

5) Principle of feasibility: Each evaluation index should be operable and effective.

B. The Establishment of a Comprehensive Quality Evaluation Index System

In this paper, we evaluate students' comprehensive quality based on some University's information system. As our evaluation objects are numerous and the complexity of data is high, we use hierarchical structure [4-5] to establish college students' comprehensive quality evaluation index system.

The college students' comprehensive quality evaluation index hierarchical structure model (Tab. 1) is divided into three levels according to the design principles of comprehensive quality evaluation index system.

III. DESIGN OF COMPREHENSIVE QUALITY EVALUATION SYSTEM FRAMEWORK

A. Design Idea

The design of comprehensive quality evaluation system framework is the core part of this paper. The framework is divided into three parts including the establishment of data layer, model layer and application layer.

The data layer is used to extract the relevant data samples, which depends on a specific application. And the model layer is designed to analysis the data samples and to get the information needed. And the construct of application layer is to meet the demand of users in an accurate and concise way.

Fig. 1 shows some details and a specific relationship between each layer.

B. Introduction of Each Layer

1) Data layer. The data layer mainly completes the work of data sampling and data storage. Data warehouse of some University information system unifies data information of departments and provides convenience for related research. And Oracle database [6] is adopted and helps us realize the establishment of data layer based on a specific application in the application layer.

2) Model layer. This layer mainly introduces data mining technology. Data mining is the process of extracting valuable knowledge hidden in large volumes of raw data. It is a computer-assisted process of digging through numerous sets of data, which are large, incomplete, noisy, fuzzy and random, and extracting the meaning of the data [7-8].

3) Application layer. The application layer is mainly designed to provide customers with applications, which includes employment recommendation, students' behavior survey, graduate students selection, academic performance prediction and so on. And it ultimately shows the required information to the customers in an accurate and concise way. These applications can be used as an auxiliary of related departments' decision-making.

TABLE I. COMPREHENSIVE QUALITY EVALUATION INDEX SYSTEM BASE ON SOME UNIVERSITY'S INFORMATION SYSTEM

First level index	Second level index	Third level index
Academic quality	Ability of practice	Grades of practical courses
		Internship experience
		Social practice experience
	Professional performance	Grades of professional courses
		Grades of Minors
		The percentage rank
	Basic skills	Grades of basic courses
		Foreign language level test
		Computer level test
		Mandarin proficiency test
	Learning habits	Reading times in library
		Times of books borrowed
	Prizes	Scholarship
		Contest prize
		Prize outside the school
		Individual title of honor
	Innovation ability	Grades of innovative courses
		Papers published
		Innovative project attendance
Ideological and moral qualities		Disciplinary records
		Disposition records
		Owing Money records
		Grades of ideological courses
		Political status
Living habits	Diet status	Eating times
		Trading volume
		Consumption type
		Consumption site
		Card using times
	Family background	Subsidy information
Physical and psychological health	Physical health	Grades of physical courses
		Sports competitions
	Psychological health	Mental health status
		Psychological questionnaire

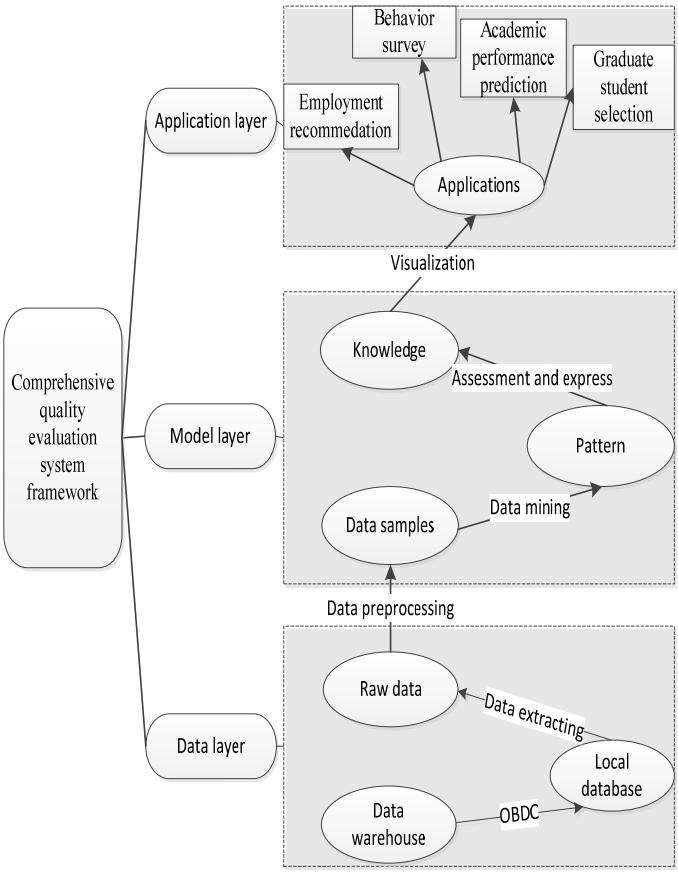

FIGURE I. DATA SAMPLES OF SEM_2013

Sex	Birthyear	Department	Book_borrow	Subsidies	Card_use_time	Consume	Scholarship
M	92	SEM	High	Very high	High	Very high	Zero
F	93	SEM	Middle	Zero	High	Middle	Excellent
M	93	SEM	Low	Zero	Middle	Very high	Zero
M	93	SEM	Low	Zero	High	High	Zero
M	93	SEM	Zero	Middle	High	Middle	Zero
F	93	SEM	Low	Zero	Middle	Zero	Low
F	93	SEM	Low	Low	Middle	Very low	Zero
M	93	SEM	Very low	Low	Very high	Middle	Middle
F	93	SEM	High	Zero	High	Very high	Zero

FIGURE II. ANALYSIS THROUGH HORZONTAL AND VERTICAL COMPARISON

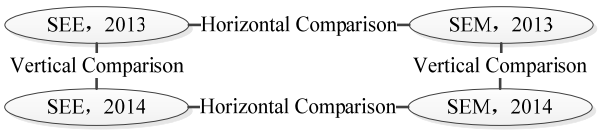

FIGURE III. DATA SAMPLES OF SEM_2013

TABLE II. BASIC STATISTICS OF DATA SAMPLES

	Female	Male	Total	Scholarship	Subsidy
SEM	145	181	326	92	72
SEE	109	362	471	159	128

IV. THE DESIGN AND IMPLEMENTATION OF COMPREHENSIVE QUALITY EVALUATION SYSTEM

In this chapter, we will put the college students' comprehensive quality evaluation system into practice by taking students' behavior survey as an example and show the process of realizing the framework.

A. The Establishment of Data Layer

Generally, data warehouse adopts three-layer structure [7-8]. Local database is established by connecting the data warehouse of school's information system. However, the raw data extracted cannot be directly used, which must be preprocessed. Data preprocessing includes data cleaning, data integration, reduction and transformation [7-8]. After preprocessing, the data can be analyzed by data mining tools. We choose the students of grade 11 from SEE(School of Electricity) and SEM(School of Economy Management) as data samples. In horizontal direction, we search for the difference between academies. And in vertical direction, we search for the changing of students in different years. Basic statistics are shown in Tab. 2.

Fig. 2 shows a portion of data samples of SEM_2013 (students of SEM in 2013). The attributes chosen include sex, year of birth, book renal times, subsidy information, campus card use times, consumption and scholarship information.

B. The Establishment of Model Layer

1) The introduction of association rule mining. Mining association rules stands for the process of finding all of the valuable associations existing in the data sets made up of a great number of data that meet certain requirements of minimum support and the minimum confidence [9-10].

Association rules mining is normally a two- step process. The first step is to find all the frequent item sets. And the second step is to generate strong association rules. The rules that meet both the requirements of support and confidence are named strong rules. There are plenty of association rule mining algorithms, such as Apriori, FP-Tree, and FP-Growth.

2) The Introduction and Realization of Apriori Aglorithm. Apriori algorithm is one of the most classic algorithms of mining association rules. The iteration of layer by layer is used. The 1_item set is obtained if the support is satisfied. Similarly, the (k+1)_item set is explored by the k_item set until item set no longer satisfies the minimum support. Apriori algorithm can be described in the following pseudocode [9-10].

1. L_1 =find_frequent_1-itemsets(D);
2. For(k=2; L_{k-1} != Ø; k++){
3. C_k =apriori_gen(L_{k-1});
4. For each affair t ∈ D{
5. C_t =subset(C_k,t);
6. For each candidate c∈ C_t
7. c.count++; }
8. L_k ={c ∈ C_k | c.count>=min_sup}
9. Return L= all the frequent item sets;

TABLE III. ANALYSIS REPORT

	Association rules	Confidence	Conclusions
SEE_2013	A2=93 A3=Low =>A7=Zero A7=Fair =>A3=High / Very high	0.92 0.33 / 0.16	Getting scholarship and the time of borrowing books are positive correlated (92% students who seldom borrow books fail to get a scholarship. And from the students who get a scholarship, more than half of them prefer borrowing books). In the year of 2014, students seem to borrow more books in the library. In SEM department, the correlation between books and scholarship is less apparent. Measures should be taken to courage students to borrow more books.
SEE_2014	A2=93 A3=Low =>A7=Zero A7=Fair =>A3=High	0.91 0.41	
SEM_2014	A2=93 A3=Low =>A7=Zero	0.88	

FIGURE IV. THE OBTAINMENT OF ASSOCIATION RULES

```
Card_use_times=Middle Consume=High 101 ==> Sex=M 93     conf: (0.92)
Consume=Very high 113 ==> Sex=M 104     conf: (0.92)
Subsidies=Zero Consume=Very high 71 ==> Sex=M 65     conf: (0.92)
Book_borrow=Very low Consume=High 54 ==> Sex=M 49     conf: (0.91)
Birth_year=93.0 Consume=Very high 53 ==> Sex=M 48     conf: (0.91)
Birth_year=92.0 Card_use_times=Middle 73 ==> Sex=M 66     conf: (0.9)
```

3) The obtainment of strong association rules. We use Eclipse(An extensible development platform based on Java) as simulation tool in combination with WEKA(Data Mining software in Java) to analyze the data samples. After simulating the data of SEE_2013, strong association rules are found in Fig. 4. Similarly, samples of SEE_2014, SEM_2013, SEM_2014 are analyzed in the same pattern. With the rules, we can dig up the information our customers concern.

4) Analysis of strong association rules. Tab. 3 numbers the attributes. We can analyze the rules by comparison. For instance, the rules {A1=M 362 => A7=Zero 293 conf:(0.81)} and {A1=F 109 => A7=Zero 68 conf:(0.62)} shows a greater ratio of getting scholarship among girls than boys. By comparison, girls seem to be more diligent than boys.

TABLE IV. ATTRIBUTE CODING

A1	A2	A3	A4	A5	A6	A7
Sex	Birth_year	Book_borrow	Subsidies	Card_use	Consume	Scholarship

By horizontal and vertical comparison, potential knowledge can also be found. For example, the rule {A7=middle 67=> A3=High 49 conf:(0.73)} taken from SEE_2013 and {A7=middle A2=93 56=> A3=High 38 conf:(0.68)}taken from SEE_2014 show students read fewer books than last year. So the library should take measures to win more readers. And the rule {A5=High A6=Middle 78=> A7=Zero 62 conf:(0.79)} taken from SEE_2013 and {A5=High A6=Middle 91=> A7=Zero 62 conf:(0.68)} taken from SEM_2013 show connections between consume and scholarship. And students from SEM consume more but there is less influence on their academic achievement.

C. The Establishment of Application Layer

The function of application layer is to make conclusions and display the results to the customers which can help with their decision-making. After analyzing the strong association rules, we dig up some useful information. Tab. 3 is a sample of analysis reports.

In this report, we provide related customers with conclusions which show advice analyzed by horizontal and vertical comparison with association rules.

V. CONLUSION

In this paper, we first complete establishment of comprehensive quality evaluation index system based on design principles. Then we construct the comprehensive quality evaluation system framework that includes data layer, model layer, and application layer. Next we choose students' behavior survey to show process of realizing the framework. We use Apriori as our mining algorithm to dig up potential association rules. After analyzing rules from horizontal and vertical direction, we display conclusions to customers with analysis reports. In future work, a more concise and direct way to display the conclusion will be considered and related application software will be designed.

ACKNOWLEDGEMENT

I would like to express my gratitude to all related people and departments that support the research and provide the data.

REFERENCES

[1] Shuhan Yang, Jingbo Xu, "Mainly talks about the college quality education," Journal of northeast agricultural university (social science edition) ,vol.21 , pp.1-2, February 2012

[2] Zengguo Zhou, Chunlan Li. University informatization construction status analysis and construction principle, 4th ed., vol.8. Automation office , 2008, pp:15-18

[3] Viktor Mayer-Schönberger, Big Data:A Revolution That Will Transform How We Live, Work, and Think, Hodder Export, 2012

[4] Jian Liu, Hongyan Zhu, Chun Liu, "College students' comprehensive quality evaluation index system and its data acquisition research," Journal of shandong institute of education, vol.108, 2005, pp. 3

[5] Guanghong Wang, Ping Jiang, "Survey of Data Mining,"Journal of Tongji University, vol.2, pp.3, 2004

[6] Oracle China LTD.Data warehouse solution.Oracle China coLTD , 2000

[7] Jiawei Han, Micheline Kamber, Jian Pei. Data Mining:Concepts and Techniques, MorganKaufmannPublishers, 2012

[8] Hans-Peter Kriegel, "Karsten M. Borgwardt, Peer Kroger, et al. Future trends in datamining," Data Mining and Knowledge Discovery, Febrary 2007, vol 15, pp.87-97

[9] CHEN M, HAN J, YU P, "Data mining: an overview from a database perspective," IEEE Transactions on Knowledge and Data Engineering, vol.8, 1996, pp. 866-881

[10] Guodong Ma, "University library circulation based on WEKA data correlation analysis.," The library work and study, 2012, pp. 4-5