A Data Driven Educational Decision Support System

Yaqiong Zhu
Henan Finance University, Henan, China
`yaqiongzhu29103@163.com`

Abstract—To make use of the latest Internet technology to provide innovative support for teacher education, a comprehensive study is made on DDEDSS (data driven education decision support system) and DDEDSS software prototype is designed and developed as an auxiliary tool for educational decision-making. At the same time, the two sessions of education data is collected and tested in practice. The multidimensional summation and average are carried out for different time, different source regions, different gender learners and different subjects. The research results show that the system can evaluate learners' learning and further provide decision-making basis for curriculum optimization and class adjustment. It can be seen that DDEDSS is a frontier educational research, which has important research significance.

Keywords—educational decision support system, educational engineering, data analysis, data mining

1   Introduction

Educational decision-making is a term understood by the public. Its essence is to make decisions based on specific educational needs. Classical education decision-making is generally achieved through opinions and meeting discussion and combined with a certain number of data statistics and analysis. Although it is fast and convenient, because it is not based on the comprehensive system of data and data processing methods are relatively single, there are difficult obstacles in the objectivity, scientific nature and other aspects. With the rapid development of education information, various kinds of educational information platforms have been gradually stored various kinds of data and increased greatly. It is becoming a hot spot in the current educational field and an inevitable direction of development how to effectively integrate and utilize these data and make educational decisions, thus improving the objectivity, scientific nature, comprehensiveness and systematicness of decision-making. In educational decision-making based on data, it often involves massive data processing, which must be aided by information systems. This kind of system is generally called EDSS (Educational Decision Support System). If it emphasizes data driven, it is a DDEDSS (Data Driven Education Decision Support System).

DDEDSS will be studied from the aspects of the topic proposal, the current research situation at home and abroad, the research content, the theoretical basis, the research

method, the innovation point, and the research design. At the same time, the educational data warehouse, the education multidimensional data set and the educational data mining structure are successfully designed. The results show that the designed educational data warehouse, the education multidimensional data set and the educational data mining structure are in line with the basic requirements of the educational decision-making.

2   Literature Review

DDEDSS is the decision based on educational data, so educational data modeling is the foundation of DDEDSS. Scholars Byington and Butler proposed the basic framework of educational data modeling. The data is established based on the theme of "students", including the students' basic information, academic achievements, courses, educational institutions, awards and other data [1]. The whole educational system is centered on learning, so it is worthwhile to set up a data model based on students as the theme. Then, because the current learning performance is mainly reflected in the performance of the course, it is also desirable using the results of all subjects as the key data. However, it is not detained studied how these achievements form a relational table is suitable for multi-dimensional analysis and data mining needed for educational decision-making. Cheowsuwan et al., based on the specific requirements of the data warehouse, proposed the star education data model, which fully met the technical requirements of the database. It also took the students' curriculum performance as the major part, but it did not consider the hierarchical information of the year / term, the province / city / county, or the curriculum / course category of the multidimensional analysis. As a result, there is still a lack of supportive direction [2]. Scholar Chou et al. put forward a general structure of educational decision support system. The goal was to take the system as the auxiliary, to extract, excavate and analyze the statistical data of education, so as to provide timely and reliable information inquiry and statistical ability for educational decision-making level. On this basis, the decision support for education work (enrollment prediction, employment prediction, development trend, scientific research construction, etc.) was formed. The system did not specifically describe data modeling, nor is it a model centered on the learner's learning performance. For the most essential DDEDSS research, it was of little reference [3]. Hernes and Sobieska-Karpińska proposed the data modeling of talent quality evaluation [4]. The current DDEDSS mainly takes curriculum learning as the main source of data, so the model has a certain reference value as an extended application of DDEDSS in the evaluation of talent quality.

Learning performance data is also dominant, reflecting the idea of learning as the main work. Konstantinidis and Bamidis proposed a DDEDSS data framework based on learners' learning performance, which was the most systematic case of a learning performance data framework in these studies. However, the framework took into account the educational environment in the United States too much and needed to be combined with local utilization of Chinese education [5]. Scholars Mccoy and Rosenbaum put forward the overall structure of educational decision support system. The framework was very large, basically from the logic of software implementation and decomposing functional modules, such as browser, server, local area network, database management

system, model base management system, knowledge base management system, and method library management system. It was suitable for DDEDSS developers and users to understand the whole system [6]. The data platform of the framework is Microsoft SQL Server 2000. The data sources include the enrollment scale in the educational statistics system, the number of students, the number of graduates, the number of teachers and staff, the construction of scientific research, and the public dictionaries in the fixed assets and systems. The realization interface of enrollment analysis and school running condition analysis are also introduced, but there is no specific analysis result. Scholar Mokin et al. put forward the overall structure of educational decision support system, but they did not describe the data platform used. The framework was hierarchical description of database, data warehouse, OLAM and DLAP [7].

DDIDSS data processing refers to the method of data processing based on computer in the whole DDEDSS system, especially data warehouse, multidimensional analysis and data mining method. From the perspective of data complex calculation, it also involves the basic research of describing statistics, inference statistics, mining algorithms, and probability. The DDEDSS data processing mainly explored the application research of data warehouse, multidimensional analysis and data mining based on database technology. Scholars Shi et al. carried out systematic application research on the necessary data warehouse technology, data mining technology and related algorithms in the educational decision support system, and made a related research on remote education as the object. They pointed out that these technologies mainly included data warehouse, data analysis, and data mining [8]. The integration function of database data to data warehouse is provided by the database platform system (such as SQI Server). Based on the platform, DDEDSS researchers mainly design a reasonable DDEDSS data model for application and development.

To sum up, all documents in DDEDSS simply span DDEDSS's three fields: education, data and computer. Therefore, the background of the information age is considered to excavate the interoperability of the three fields of education, data and computer across DDEDSS. In addition, the whole idea of DDEDSS is found, so that DDEDSS can be implemented logically and mutually.

3   Method

3.1 Principles and engineering framework of the educational system

The framework of education system principle and engineering is the inheritance and expansion of the framework of system principle and engineering, as shown in Figure 1. In essence, it can be called the "educational principle and engineering" of the concept of information interaction system.

The elements of the educational system and their relationships in Figure 1 are more instructive if they are further iterated into the granularity of the elements and their relationships for the various educational systems. It also lays the core foundation for the modeling of educational data. In the current information age, the educational system is defaulted as the interactive system of educational information. It is necessary to analyze

the elements of the educational system from simple to complex, and analyze the relationship of these elements from the perspective of information interaction.

When analyzing the elements of the educational system, the interactive relationship between the two protagonists must first be abstracted from the local character, as shown in Figure 2.

In Figure 2, educators, learners, and interactions remain unchanged. Education information is further concretely translated into educational content, educational methods and educational media according to current customs. The, Figure 2 will be transformed into Figure 3.
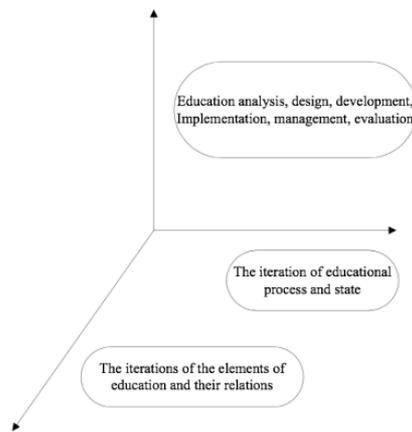
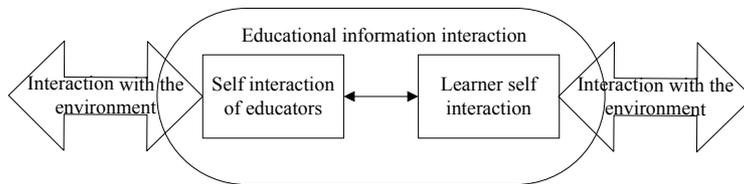**Fig. 1.** The principle of education and the schematic diagram of engineering

Fig. 2. The iterative starting point of the elements of education and its relationship
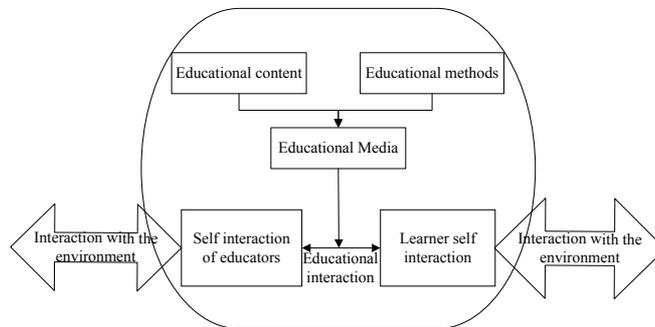

Fig. 3. One of the iterations of the elements of education and their relationship

The iterative granularity shown in Figure 3 is still abstract. After further iteration, it is the ideal iteration granularity for all kinds of education.

Based on the framework of "system principles and Engineering", the framework of "DDEDSS system principles and engineering" can be inherited and expanded. That is, the iteration of "analysis, design, development, implementation, management and evaluation of DDEDSS elements and their relationships, processes and states" is used as an iterative starting point and a main line of iteration for the understanding and transformation of DDEDSS, as shown in Figure 1. In essence, it can be called the "DDEDSS principle and engineering" framework of information interaction system.

3.2 DDEDSS elements and the relationship

As shown in Figure 1, iteration is explored until approaching DDEDSS principles and engineering. The goal or result of DDEDSS elements and their relationships must be clear.

In the current information age, the DDEDSS system can be regarded as an information interaction system. The protagonist elements are abstracted, and then the information interaction between the two protagonists is abstracted, as shown in Figure 4.

In Figure 4, the two elements of "DDEDSS" and "education system" are maintained. The elements of "DDEDSS information interaction" are iterated into "DDEDSS content", "DDEDSS method", "DDEDSS media", "DDEDSS interaction" in accordance with the relationship among information content, method, and media, as shown in Figure 5.

Figure 5 is slightly abstract, for which the six elements continue to carry out the iterative depth of the elements of the usual information interaction system. Further iterations can be used as the ideal granularity of the DDEDSS elements and their relationships.
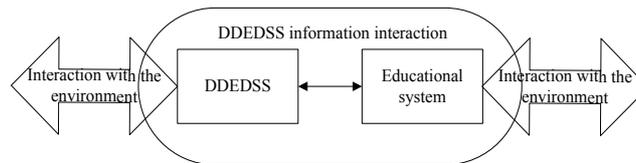

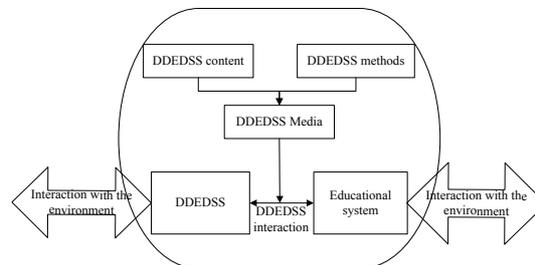**Fig. 4.** The iterative starting point of the DDEDSS elements and their relationships

Fig. 5. One of the iterative granularity of DDEDSS elements and their relationships

4 Results and discussion

4.1 DDEDSS process and state

The iterative framework for defining the elements and their relationships of the DDEDSS system requires further iteration of the DDEDSS process and its state, until the target or result required by DDEDSS is approximated.

The DDEDSS process and its state mainly highlight the educational data collection, warehousing, analysis, mining, data education decision making, and submission of educational decision suggestions for the DDEDSS system. Therefore, the further iteration can be used as the ideal state of the DDEDSS process and its state iteration.

The complete iteration of the overall functional requirements of the system is completed in depth with the protagonist use case as the main line of iteration. The protagonist use case is often represented in the UML model diagram in Figure 6 in the field of system analysis. All use case iterations in this DDEDSS (teacher) protagonist use case are shown in tree structure shown in Figure 7 until the DDCDSS practice requirements are met.
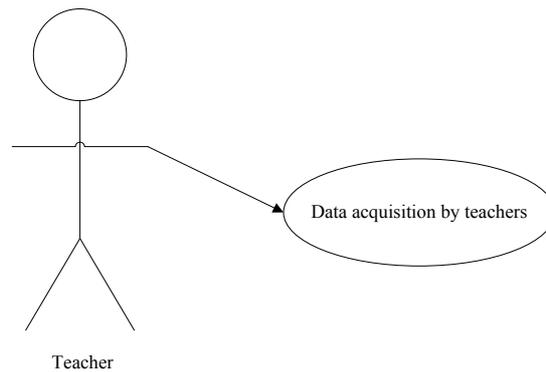
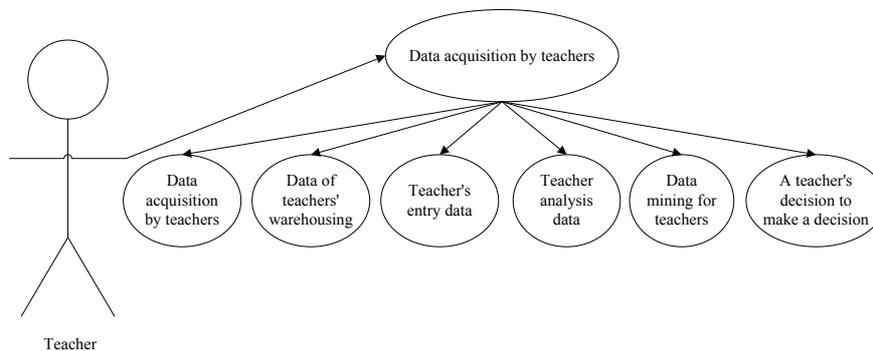**Fig. 6.** The header image of online-journals.org

Fig. 7. The header image of online-journals.org

As shown in Figure 7, the tree structure forms the namespace of the system protagonist use case, such as "all the use cases of the teacher's protagonist and teacher analysis data", and the node number "." connects the levels. Because the protagonist use cases in the tree structure hierarchy are combinatorial relations, the leaf protagonist use case is the protagonist use case that really needs to be implemented. The description of the protagonist use case is best in the form of subject-predicate- object, which is convenient for the maintenance and communication, such as the teacher acquisition data, the teacher mining data, and the system administrator warehousing data. These use cases are combined in time order to get the system process and its state. It can also be described in UML as shown in Figure 8. Of course, there may be the process and its state of some branches.
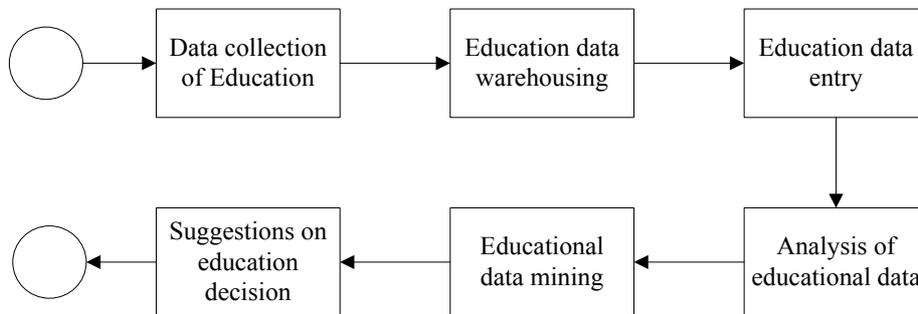

Fig. 8. The header image of online-journals.org

The above DDEDSS process and its state will eventually be mapped to the specific realization of educational data collection, educational data multidimensional analysis, and educational data mining. At the same time, the DDEDSS software system is developed, and the software metaphor of the process and its state is involved. This also reflects the guidance of DDEDSS system analysis, DDEDSS data processing and DDEDSS software development.

Through the DDEDSS engineering, people analyze, design, develop, implement, manage and evaluate the DDEDSS elements and their relationships, as well as DDEDSS processes and their status. In each of the six links, the DDEDSS complete process and its state iteration are embodied, so as to realize DDEDSS optimally.

DDEDSS is finally implemented in software metaphor, so DDEDSS engineering is integrated into DDEDSS software engineering research. The analysis, design, development, facilities, management and evaluation of DDEDSS project are concentrated in a table as an overview (six elements of DDEDSS, education system, DDEDSS content, DDEDSS method, DDEDSS media, and DDEDSS interaction are transverse, and DDEDSS process and its state are vertical).

4.2 Data collection, analysis and mining

DDEDSS data methods must be combined with specific computer software tools. When choosing a computer software tool, if it is used as a data processing tool, it is

suggested to choose the widely used tool software such as Exce1 and SPSS. If it is used as a development platform, the database systems such as Access, SQI.Server, Oracle, and My SQL are often selected. Considering the DDEDSS's overall solution, ease of use and mainstreaming, SQLServer2008 database system is chosen.

The corresponding relations are abstracted based on the DDEDSS method level and the data method level supported by SQL Server2008. The application research is carried out in accordance with the three levels of DDEDSS data acquisition, warehousing, integrated warehouse - educational data multidimensional analysis - educational data mining (ignoring the emotional communication and philosophic thinking methods that SQL Server does not yet support). At the same time, the DDEDSS educational data warehouse, educational multi-dimensional data set and educational data mining structure are studied and designed.

4.3 Educational data acquisition and data warehouse design

The educational data model is the basis of data acquisition, warehousing, and integration. When data are collected based on the educational data model, the design problem of data entry needs to be considered, that is, the data collected must be able to be mapped into the designed computer database, and then input the data into the computer database.

When designing the educational database, it must be based on the educational data model, guided by the object-oriented engineering theory. The final granularity of the objects and their relationships are analyzed, and then the objects are described with the data. When describing objects with data, it is necessary to further consider data item / field, data type, data value, data structure / data table, and (data table) relationship. Then, based on the principle of OLTP (On-Line Transaction Processing) database, OLTP database is established. Then, data can be stored in the database based on objects.

DDEDSS research is limited to the study cycle and cannot get data from zero. It is designed to obtain these historical data directly from the educational system of the selected schools as the original data. The original data is obtained from an independent college of Normal University, and the performance score is the core of these historical data. Due to data normalization requirements, only two (eight years) data are available.

4.4 Educational data analysis and multi-dimensional data set design

Based on the education data warehouse, the OLAP education multidimensional data set can be set up to carry out data analysis (that is, data description statistics in the field of statistics). Data analysis obtains the required data from large amount of data, which is also called correspondence, so data analysis is often referred to as information analysis and information acquisition to reflect the abstract level of information on the data. Data multidimensional analysis, also known as analytic hierarchy process (AHP), is the focus of current educational data analysis technology.

In order to realize the analysis of educational information, it is supposed to first design and construct the education multidimensional data set (Cube) based on the educational data warehouse. The objective value of this education multidimensional data set

should include the average score, the number of students selected, the highest score and the lowest score, and the analysis dimension mainly includes the time (year and term), source place (provinces and cities), courses (568 courses), course categories (five courses), and students' gender (male and female).

In order to suit the general public's understanding, the model transformation shown in Figure 10 is also used. Figure 10 shows the multidimensional data set designed for three dimensions: total score and exam time, learner source area and gender.
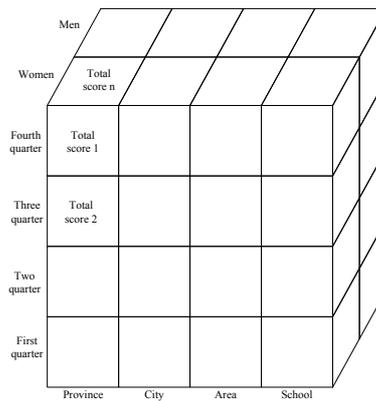

Fig. 9. Model representation of multidimensional datasets

In summary, data multi-dimensional analysis, compared with common software such as EXCEL, has the advantages of interactive multi-dimensional analysis (including four dimensions of province, city, gender, and all courses). Through the free combination of various dimensions, it can be used to assist in the implementation of the following educational decision objectives: multi-dimensional summing up and summing of the achievements of the five classes of different time (year and term), different source areas (province and city), different gender learners (male and female), different subject categories (Philosophy, love, science, technology and practice). It can be used to evaluate learners' learning and provide basis for educational decision-making such as curriculum optimization and class adjustment.

More educational data multi-dimensional analysis function is after the design of education multi-dimensional data set in SQL Server. It must develop B/S mode DDEDSS software, integrate the education multi-dimensional data set into DDEDSS software, so that all kinds of users use IE browser as an auxiliary tool for educational decision-making.

4.5 Educational data mining and mining structure design

In addition to the above data analysis and the acquisition of information, it can also be based on the principle of data mining, to find the function or correlation law between the data from massive data. These rules are often referred to as knowledge, so data mining is often called knowledge discovery (the term commonly used in the field of

artificial intelligence), to embody knowledge as an abstract level of data and information. It can also be called inferential statistics (the term commonly used in the field of statistics) and data mining (the term commonly used in computer science). The word data mining is mainly selected, which is a hot topic in the field of data technology and data intelligence.

The essence of data mining is the calculation of the correlation between massive data. The multi-dimensional analysis mentioned above is the calculation of functional relations. Functional relation is a deterministic mathematical relation, and the correlation is a mathematical relationship which is not completely deterministic, but a basic trend inference.

4.6 Educational data acquisition and data warehouse design

A series of data has a certain relationship, but there is no functional relationship with exact rule, which is called the correlation, also known as the similarity. It is often measured by the distance or correlation coefficient between the data.

There are various formulas for calculating distance and correlation coefficients, and each has its applicable data. For example, Mahalanobis distance d_{ij} is calculated as follows:

$$d_{ij} = \sum_{k=1}^{n} \left| X_{ik} - X_{jk} \right|. \tag{1}$$

In Formula (9), X_i and X_j are:

$$X_i = \begin{bmatrix} x_{i1} & x_{i2} & ... & x_{in} \end{bmatrix}. \tag{2}$$

$$X_j = \begin{bmatrix} x_{j1} & x_{j2} & ... & x_{jn} \end{bmatrix}. \tag{3}$$

From this, it can be seen that the smaller the distance d_{ij}, the higher the correlation of the two data; on the contrary, the lower the correlation.

Clustering, which first specifies the group number needs to be divided, and then seeks the algorithm to scientifically divide the data points into the required groups according to the distance relationship. As shown in Figure 10, all data points are basically divided into five groups.
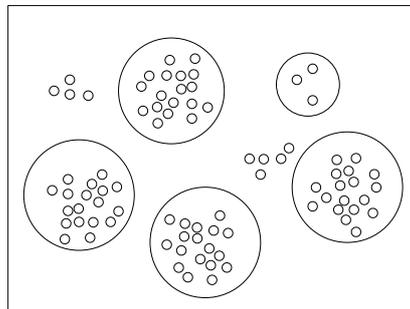

Fig. 10. Clustering schematic diagram

Figure 11 is a line representation of the results of classification / regression, and it can be found that all data are divided into two categories. Figure 12 is a regression formula for the results of classification / regression.

$$y = 0.5x + 5.\qquad(4)$$

The regression formula of another category is:
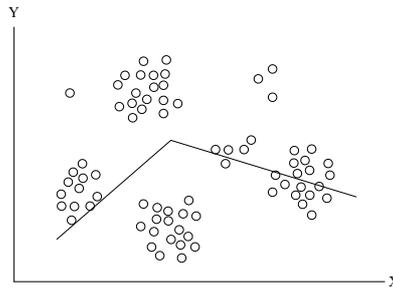
$$y = 0.25x + 8.75\qquad(5)$$

Fig. 11. Line representation of the results of classification/regression
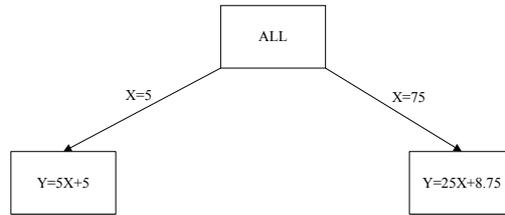

Fig. 12. Expression of regression formula for the results of classification/regression

Selected Curriculum Category ID: Predict Only. That is to say, the course category is designed to be predicted, so as to predict the probability of choosing five courses in different cities and genders. City: input attribute (Input). That is, city is set as an attribute of learners, which is used to predict the parameters in elective courses. Gender: input attribute (Input). That is, gender is set as an attribute of learners, which is used to predict the parameters of elective courses. Selected Curriculum ID: key attribute (Key), which is used to uniquely identify a record. The design of educational data mining structure and educational data mining model will be verified in DDEDSS software at last, so as to assist decision-making.

In male learners' tendency to choose courses, the tendency of male learners to choose courses can be seen. The category of "(Gender)=male" can show that the probability of male learners to choose 1 (practice class) is 11.17%, the probability of choosing 2 (technical class) is 3.64%, the probability of choosing 3 (science class) is 5.03%, the probability of choosing 4 (love class course) is 24.33%, and the probability of choosing 5 (Philosophy class) is 5.83%.

In female learners' tendency to choose courses, the tendency of female learners to choose courses can be viewed. The category of "(Gender)=female" can show that the probability of female learners to choose 1 (practice class) is 5.87%, the probability of choosing 2 (technical class) is 2.5$%, the probability of choosing 3 (science class) is 42.24%, the probability of choosing 4 (love class course) is 43.92%, and the probability of choosing 5 (Philosophy class) is 5.40%.

The comparison of the course selection tendency of male and female learners show that boys are more interested in the courses of practice and technology, while girls are more interested in the course of love, which is basically the same as the reality. As a normal college, it is possible to encourage students to choose more affective classes and lay a foundation for the emotional communication between teachers and students in the future development of education. For girls, they should also be reminded to over-come their fear of hardship and fear of fatigue. They should choose more practical and technical courses and improve their practical ability.

More educational data mining functions will be studied in the software development section. The reason is that after the design of the data mining structure and mining model in SQL Server, the B/S mode DDEDSS software must be developed to integrate the educational data mining model into the DDEDSS software. As a result, all kinds of users can use the software as an auxiliary tool for educational decision-making after the login of IE browser.

5 Conclusion

Educational data acquisition, storage, integrated warehousing - data analysis - data mining is used as the main line of logic, and the application of DDEDSS data processing is studied. Based on the educational data of the two sessions (eight years), preliminary verification is made. The research shows that SQL Server 2008 basically meets the functions required by DDEDSS, and it has the advantages of ease of use and main-streaming, so it will be used as the method and tool software of this subject. However, as a programming support for data services, the built-in data multi-dimensional analy-sis, data mining, and other programming calls are rarely referred to, which means the challenge of the development of DDEDSS software. At the same time, it is found that the five levels of data acquisition, data analysis / information interaction, data mining / knowledge discovery, emotion mining and philosophy mining are unified with the five levels of practice, technology, science, sentiment and philosophy in the perspective of the subject system. The feasibility of "information and interaction system" is further verified.

6 References

[1] Byington, J., Butler, C. (2016). Utilization of compustat tapes in decision sup-port systems education. Data Processor for Better Business Education, 25(1): 9-10.

[2] Cheowsuwan, T., Rojanavasu, P., Srisungsittisunti, B., Yeewiyom, S., Cheow-suwan, T., Rojanavasu, P. (2017). Development of data warehouses and decision support systems for

executives of educational facilities in northern thailand to increase educational facility management capacity. International Journal of Geoinformatics, 13(2): 35-43.

[3] Chou, W., Tien, P., Lin, F., Chiu, P. (2017). Application of visually based, com-puterised diagnostic decision support system in dermatological medical educa-tion: a pilot study. Postgraduate Medical Journal, 93(1099): 256. https://doi.org/10.1136/postgradmedj-2016-134328

[4] Hernes, M., Sobieska-Karpińska, J. (2016). Application of the consensus method in a multiagent financial decision support system. Information Systems and e-Business Management, 14(1): 167-185. https://doi.org/10.1007/s10257-015-0280-9

[5] Konstantinidis, S., Bamidis, P. (2016). Why decision support systems are im-portant for medical education. Healthcare Technology Letters, 3(1): 56. https://doi.org/10.1049/htl.2015.0057

[6] Mccoy, C., Rosenbaum, H. (2017). Unintended and shadow practices of decision support system dashboards in higher education institutions. Proceedings of the Association for Information Science & Technology, 54(1): 757-758. https://doi.org/10.1002/pra2.2017.14505401145

[7] Mokin, B., Mokin, V., Mokina, Y. (2016). Decision support system for the use of funds received from higher education institution paid services. Actual Problems of Economics, 177(3): 372-383.

[8] Shi, T., Tan, M., Young, B., Lye, D., Ng, T. (2017). Youtube video educational package increased acceptance of antibiotic clinical decision support system rec-ommendations. Open Forum Infectious Diseases, 4(Suppl 1): 263-S263. https://doi.org/10.1093/ofid/ofx163.579

7 Author

Yaqiong Zhu is with Henan Finance University, Henan 451464, China (yaqiongzhu29103@163.com).