Exploration and Practice of the Teaching Reform of Data Warehouse and Data Mining under the Background of Big Data

Yue Junying

College of Computer Science and Technology, Inner Mongolia

Normal University, Hohhot, 010022, China

Keywords: Big Data, Data Warehouse, Data Mining

Abstract: Data warehousing and data mining are a new cross-cutting lesson in the era of Big Data. It requires the mining of a large number of structured and unstructured data sets. The data mining field needs to undergo radical changes, and the content and practice of data mining courses must be adjusted accordingly. This article analyzes the connotation and characteristics of big data, and the core curriculum of data mining, and discusses the new requirements and challenges of the "Data Warehouse and Data Mining" course in the era of Big Data.

1. Big Data Features

Gartner believes that big data is a massive, high growth rate and diversified information asset that requires a new processing model to achieve its discovery and optimization. Wikipedia believes that the amount of data involved in big data is so large that it cannot be acquired, managed, processed, and organized in a reasonable amount of time through mainstream software tools. It is a data resource that can play a supporting role in decision-making. Big data is all datasets including transactional and interactive datasets. Its size or complexity is beyond common technology. McKinsey believes that big data refers to the inability to collect and store content using traditional database software tools within a certain period of time. Management and analysis of the data collection, so we need to use data mining to analyze a large number of structured and unstructured data collections to provide useful data insights. [1]

Big data generally has four main features:

1) The volume of data is huge. Over time, the unit of measurement of data volume is from G, T, P to E.

2) Variety of data, the development of the Internet, Internet of Things, and sensor networks have made data types more complex, including not only traditional relational data types, but also web pages, video, audio, E-mail, etc. Raw, semi-structured, and unstructured data.

3) Velocity, in the face of rapid and dynamic streaming data, the speed of acquiring, storing, and excavating effective information is difficult to handle with traditional systems.

4) The value of value is low, and the exponential increase in the amount of data increases the difficulty in obtaining useful information. It is more difficult to quickly and efficiently discover potentially valuable patterns hidden in mass data.

2. Contents of Data Warehouse and Data Mining Course

"Data warehouse and data mining" as a subject with both theoretical and practical value has become one of the learning contents of many undergraduates majoring in computer science, information systems and so on. It can be seen that this subject is in today's science. The importance and the breadth of the application. As a frontier discipline, there are many technologies and methods for data warehouse and data mining that are open or still in the exploration phase.[2] Therefore, in addition to mastering basic concepts and methods, students should be aware of the many problems faced in the subject and the research directions to be expanded. This helps students to provide an extensive and appropriate overview of the subject to improve the ability of independent learning,

and to provide an opportunity for students who are interested in conducting in-depth research in this field. Since the course originally belonged to the professional courses offered by the graduate students, most of the teaching materials focused on the theoretical knowledge of the structure of the system, the principle of the algorithm, and the analysis and improvement of efficiency. The contents involved were mostly deep and many knowledge exceeded the acceptance of undergraduates. Scope; In addition, teaching materials have relatively few descriptions of related theories in practical applications, which is not conducive to arranging experimental teaching. Therefore, to achieve the teaching goals of the "Data Warehouse and Data Mining" curriculum, we must flexibly select and arrange [1-2] factors in the theoretical teaching and experimental teaching, taking into account factors such as how many learning hours, teaching conditions, and student acceptance.

According to the research system of data mining (as shown in Figure 1), the main teaching contents of data mining related courses are given, including:
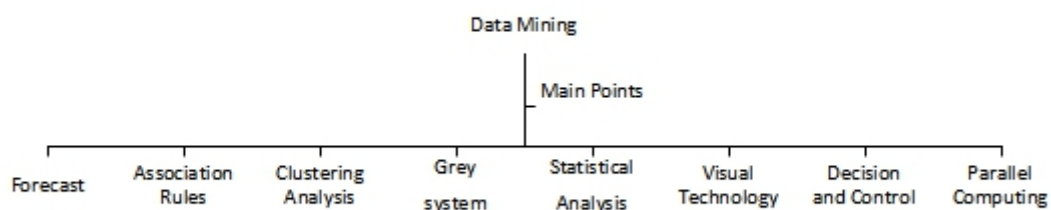

Figure 1 Research System of Data Mining

3. Problems in Data Warehouse and Data Mining Teaching

(1) "Data warehouse and data mining" as a professional elective course, comprehensive, practical, and in a continuous development process. Data warehouse and data mining involve a wide range of knowledge, involving various aspects such as mathematics, computer, signal processing, visualization technology, machine learning, algorithm analysis, program design, and statistical theory. The textbooks of the data warehouse and data mining curriculum are currently on the market. There are Chinese textbooks and English translation textbooks. However, most textbooks are vague, and they simply explain the basic concepts of data warehouse and data mining in principle. Many teachers think that the course content should be based on the data mining organizational framework elements - data, classification, clustering, and cutting-edge are the main lines, pointing out that a universal problem in current textbooks is that they have written knowledge materials, encyclopedias, even the development of computer graphics, digital image processing, database technology, statistical analysis, and algorithm design.[3] The contents of other basic course systems such as the operating system and the operating system are also included in the collection, and various related knowledge is described in detail, and the internal rules of data warehouse and data mining cannot be clearly represented, and the organizational elements of data mining are artificially fragmented. Teachers use such teaching materials and will face two major problems in the arrangement of teaching content: First, the content is too broad, the scope is too large, and the relevant content is all in the water, and the two are the many courses in the preparatory courses. It has already been studied, or will be described in detail in the follow-up course, resulting in a waste of student and teacher time and teaching resources.

(2) The course "Data Warehouse and Data Mining" is closely related to the database. When explaining the data warehouse part, students should first understand the process of data management technology development from the database to the data warehouse in order to understand the data warehouse technology and data mining technology. [3] The reason for this, and then the emergence of data mining this kind of deep-seated data analysis development process.

While explaining the relationship between data warehouse and data mining, it should be pointed out that the data warehouse system can meet the requirements of the data mining technology for the data environment and can be directly used as the data source of data mining. The data warehouse can provide enterprise managers with the data environment needed for decision analysis. Data mining can discover knowledge from a large amount of data and is a deep-level data analysis method. While teaching data mining, you should focus on the concepts and attributes of the methods used, rather than applying different data mining tools mechanically. Because data mining, in-depth understanding of mining methods, models and their working principles is the basic conditions for effective and successful use of data mining technology.

(3) Experimental teaching and ability training and practical application Disjointed data warehouse and data mining are highly practical disciplines. At present, experimental teaching of "Data Warehouse and Data Mining" has the phenomenon that experiment contents and students' ability to cultivate and practical application are disconnected. The experiment of school data warehouse and data mining is carried out through SQLSERVER's data mining platform. Using SQLSERVER's data mining platform to conduct experiments, the relevant experiments are completed simply by clicking the mouse. These basic fool-applicative skills are available to high school students who have had a little contact with computers. These experiments are far from the teaching purpose of data warehouse and data mining and the corresponding teaching goals. In particular, the use of a ready-to-use data mining platform for SQL Server will lead students to complete the layout of experimental tasks within a few minutes, resulting in 90 minutes of experimental courses, most of the time students have nothing to do, resulting in a lack of classroom discipline, most of the experiment time and experiment resources are wasted.[4] Moreover, the author believes that any course of information should be changed to change students' learning thinking, that is to change people's thinking into computer thinking, as the main goal, which is the core of any course experimental teaching. At present, experiment teaching of Data Warehouse and Data Mining cannot achieve this goal.

(4) Focus on the content, ignoring the guiding role of the course in the follow-up courses Data warehouse and data mining as an entry-oriented curriculum, the core goal of teaching is to guide students to understand the basic principles of data mining and the main application direction of data mining. Currently in the era of big data, the demand for data processing professionals in the society can be divided into three categories: the underlying data collectors, data analysis and processing personnel, and analysis results visualization personnel. [4] The first priority of the employer's needs for various types of personnel is divided into three levels: first, data processing personnel engaged in data collection, data acquisition, and data cleanup; second, data analysts who perform data analysis and draw relevant conclusions; It is an extension worker engaged in visualization technology and data graphic expression. At present, many teachers in the "data warehouse and data mining" teaching, the lack of guidance between the curriculum knowledge and follow-up professional course learning, is too much emphasis on the teaching of specific knowledge, such as classification methods, clustering methods and other specific knowledge content teaching.

4. Teaching Reform Direction of Data Warehouse and Data Mining

(1) Reform of course content, reinforcement of main content of teaching content, and adjustment of teaching content in combination with current development direction Teaching material is a crucial part of teaching work. It is the basis of teaching. In many teaching reform articles, teachers have proposed various kinds of teaching materials. Problems and gave corresponding reasonable suggestions and opinions. [3] Data warehouse and data mining is a science that is currently under development. There are various new technologies applied in various fields. It is very difficult to describe data warehouse and data mining contents in a comprehensive and systematic way through relatively stable teaching materials. It is not realistic. This requires teachers to give more play to subjective initiative and find a main line of teaching in the ever-developing technology and relatively stable teaching materials. This requires teachers to not only read the textbooks but also read a large number of research papers on data warehouses and data mining in recent years. This is

complementary to the research work of teachers.

(2) According to the specific teaching content, various teaching methods are adopted in the teaching method, not only the teaching of "Data Warehouse and Data Mining", but the teaching method for almost all courses is a single teaching method. This method makes it easier for teachers to get away from the actual theory and teaching, and it is easier for students to completely break away from the teaching link. After the course is over, they do not know what to do. In terms of the specific course of "Data Warehouse and Data Mining," the author believes that teaching methods that emphasize both teaching and experimentation are better. Among them, the lecture method should use blackboards for teaching. The principle knowledge of the course content should be explained clearly first [4].

(3) The teaching data warehouse and data mining combined with the practical application of various industries at present has its extensive application in current production and life practice. For example, current medical system disease diagnosis, business system price analysis, communication system flow analysis, and so on, must all use data mining methods for processing. Teachers should combine these widely used and relatively mature data mining techniques in the teaching process, so that students can understand that the data mining technology is on their own, not far away. This will enable students to understand that what they have learned has a wide range of applications in their future work, and is willing to invest more enthusiasm, more time and energy in the data warehouse and data mining learning [4].

(4) Search for professional information on the frontier of the discipline In the final stage of the experiment course, arrange for students to search for and search for professional materials, mainly by arranging some of the most cutting-edge themes related to this stage of learning so that students can find out for themselves so as to broaden their knowledge. The vision enables students to understand the current problems faced by the discipline in research and application fields, which are more successful and effective methods, and which are the issues that need further development.
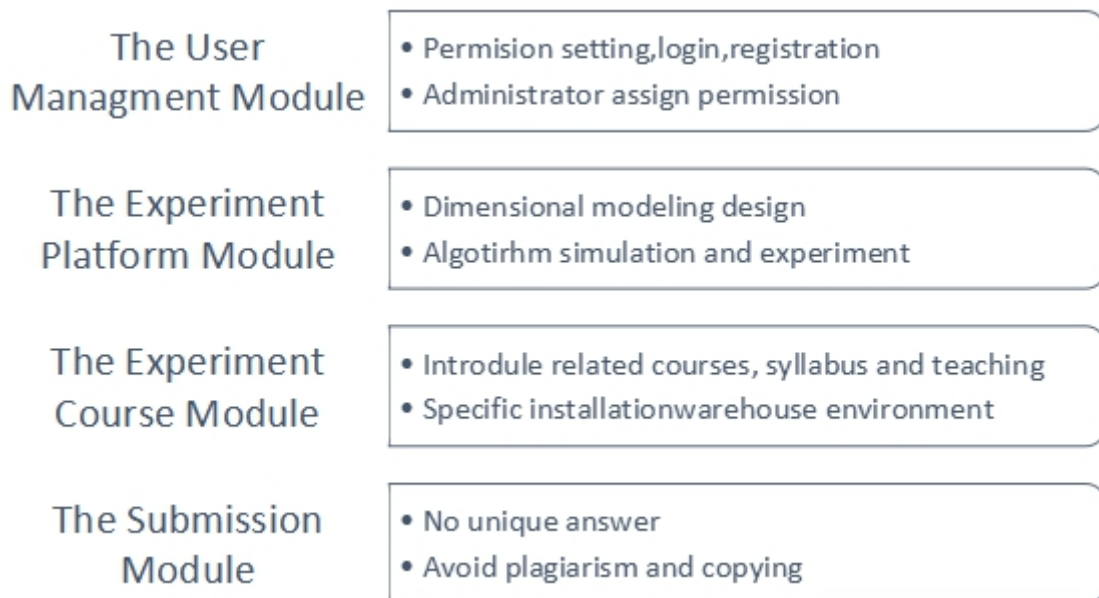

Figure 2 Data Warehouse and Data Mining Experiment Platform

5. The Following Changes Noted in the Data Mining Related Curriculum Teaching under the Five Data Environment

(1) Data preprocessing: In addition to using data warehouses to load traditional data, for the unstructured data involved in big data analysis, the integrity of the input data and the related ETL (Extraction-Transformation-Loading) should be guaranteed.

(2) Data storage mechanism: The realization of structured multidimensional data organization and management through multidimensional cubes is based on relational data models and relational

databases. When unstructured data needs to be considered in the analysis of big data, a distributed file system should be used to access the data in the file system in the form of streaming, providing access to high-transmission-rate applications (such as Hadoop and others) with large data sets. Open source distributed system infrastructure).

(3) data mining algorithm processing capabilities: In the face of increasing data size, the efficiency of data mining algorithms needs to be solved to improve the effectiveness and scalability of the algorithm.

(4) Data Mining Algorithm Processing Efficiency: The continuous increase of the data size leads to longer analysis and processing time, and the time-efficient requirements for information processing under the big data condition are higher and higher. A simple and effective artificial intelligence algorithm and new problems should be established. Solution.

6. Conclusion

The arrival of the era of big data has brought new challenges to the research and teaching of data mining. It is necessary to innovate from data preprocessing, data storage mechanisms, data mining algorithm processing capabilities and efficiency in order to meet the needs of knowledge management and intelligent decision-making in big data environments. Since data warehouse and data mining belong to a new discipline and the development history is not long enough, teaching contents must be arranged for the undergraduates' learning characteristics and the original knowledge system to inspire students' interest in learning and further improve classroom teaching. The effect is to make students have a deeper understanding and understanding of the overall structure, concepts, and technologies of data warehouse and data mining.

References

[1] Zhao Shujun, Yuan Xiaobo. Discussion on the Concept of Data Mining Technology[J]. Science and Technology Vision, 2013(33):113+141.

[2] Huang Meiliang. Discussion on the Teaching Practice of "Data Warehouse and Data Mining"[J]. Computer age, 2012(12):58-59+61.

[3] Li Mingjiang, Tang Ying, Zhou Lijun. Data Mining Technology and Application[J]. China New Communications, 2012(22):70-71+78.

[4] Hu Jianjun. Discussion on Undergraduate Teaching of Data Warehouse and Data Mining[J]. Journal of Guangxi Academy of Sciences, 2007(03):81-82+86.