The Application of Data Warehouse and Data Mining Technology in Power System

Zhang Yi

Nanchang Normal University

Department of Mathematics and Computer Science

Abstract—In the past half century, the development of information technology, computer technology and network technology has deeply influenced the production and management of power system from all aspects. SCADA, EMS, GIS and other information systems appeared. But the existing information systems have some problems to be solved, such as data cannot be shared, the level of integration is low; vast amounts of data extract information characteristics are difficult; timely monitor and forecast of business is difficult. The electric power information factory based on data warehouse and data mining technology is the solution of the problems. This paper mainly studies the applications of the data warehouse technology and the data mining technology in power system. After introducing the subject background, this paper firstly introduces the overall architecture of information enterprises and factories, individual components and their mutual relations. The paper mainly expounds the decision support system, data warehouse and data mining technology.

Keywords-Data warehouse, Data mining technology, Enterprise information factory, Digital electric power system

I. INTRODUCTION

Power system is a huge one time energy system which is responsible for the production, transmission, distribution and use of electricity. The matching protection, control and command scheduling system are needed to maintain the normal operation of the system. If the electric power production, transmission, distribution and use are considered as a process of movement and change of energy, so the protection control and command scheduling of power system can be seen as a process of movement and change of power information. With the development of information technology, the traditional science which takes material and energy as the center has given way to the modern science which takes material, energy and information as the center.

This paper firstly briefly introduces the concept and system structure of enterprise information factory, focusing on decision support system, data warehouse and data mining technology. Aiming at a large amount of data in power system cannot be effectively used at present, based on data warehouse, the paper puts forward a solution which can sort, extract, purify and transfer the existing date and provide decision support with fast and effective data response

II. THE CONCEPT AND CHARACTERISTICS OF DATA WAREHOUSE

Due to the intensifying of market competition and the demand of information society development, extracting (retrieving and querying) information from the large amount of data to formulate marketing strategy is more and more important, which does not only require online services, but also involves a large amount of data for decision making. However, the traditional database system has been unable to meet the requirements, which embodies in three aspects: there are a large amount of historical data, auxiliary decision-making information involves a large amount of data of many departments and the date of different system is difficult to integrate; the ability to access the data is in a lack, the access performance on a large amount of data of which significantly decreases.

As the mature and parallel of C/S technology and the development of database, it is necessary to improve the efficiency and effectiveness of decision-making. The development trend of information processing technology is to extract data from a large number of transactional database, and clean and transfer the data into a new storage format, which is to aggregate the data in a special

format for decision-making purpose. With the development and perfection of the process, the special data storage which is used in DSS- Decision Support System is called Data Warehouse (DW).

The data structure of a typical data warehouse is shown in Fig .1.
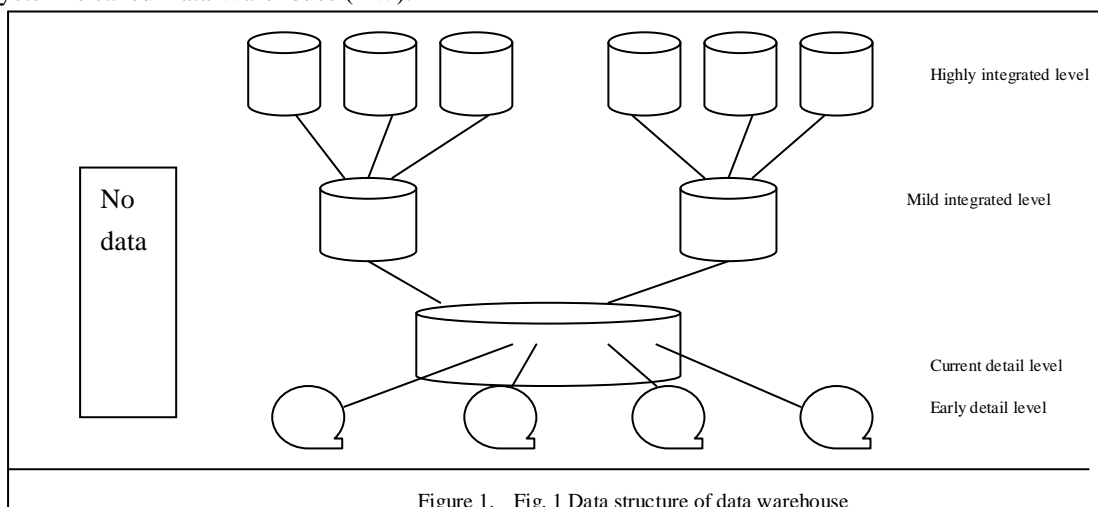

Figure 1.   Fig. 1 Data structure of data warehouse

The data in data warehouse is divided into four levels: early detail level, current detail level, mild integrated level and highly integrated level. After being integrated, source data firstly enters current detail level, and is further integrated according to the specific needs, and then enters mild integrated level and highly integrated level. The data of aging enters early detail level. It can be seen that there are different integrated levels in the data warehouse, which are generally referred to as "granularity". The larger the granularity is, the lower level of small details and the higher level of integrated.

III.   THE MATHEMATICAL MODEL AND ALGORITHMS OF DATA MINING

From the perspective of knowledge discovery, data mining can be divided into two categories: "verification-driven" and "discovery-driven".

Verification driven data mining technology uses conventional technologies, such as structured query language (SQL) and online analytical processing (OLAP). The analyst firstly makes dry period setting, and then uses conventional technologies to explore and hope to verify the assumptions. Discovery driven data mining technology discovers new assumptions which are the unknown hidden patterns by using machine learning, statistics and other various algorithms. A narrow concept of data mining actually refers to this approach.

Discovery driven data mining analysis is generally divided into Description analysis and Prediction analysis. Descriptive analysis is used to understand the characteristic of data which already exist in the system, and predictive analysis is to estimate the future of the system based on description analysis.

Some prediction models are trained by the historical data whose target variable values have been known training. This training sometimes refers to as the guidance learning, for it is to make it "learn" by giving some known answers (known results and data). Corresponding, there is also learning without a guidance, such as the description data mining (before operation, algorithm does not know anything about the data).

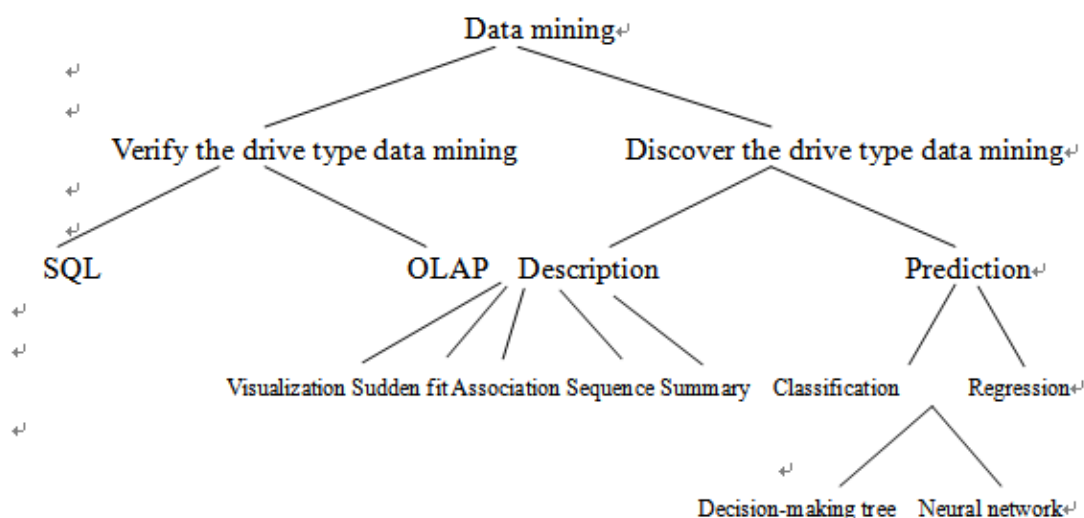Fig .2 is a simple classification of data mining method.

Figure 2. The classification of the data mining technology

IV. DATA WAREHOUSE CONSTRUCTION OF POWER SYSTEM

The data source of the data warehouse of power system mainly comes from EMS system of power system, electricity business data, geographic information system, etc. EMS system includes the grid real-time data of SCADA system, saving the operation mode of power grid, and the real-time operation parameters, such as entire network load, trend distribution, central voltage, system frequency, etc.; Electricity business data includes user information, sell electricity, electricity prices, measurement and other data; Geographic information system includes users, location of the power equipment; Other data sources include economic conditions, weather conditions, data input by handwork and so on.
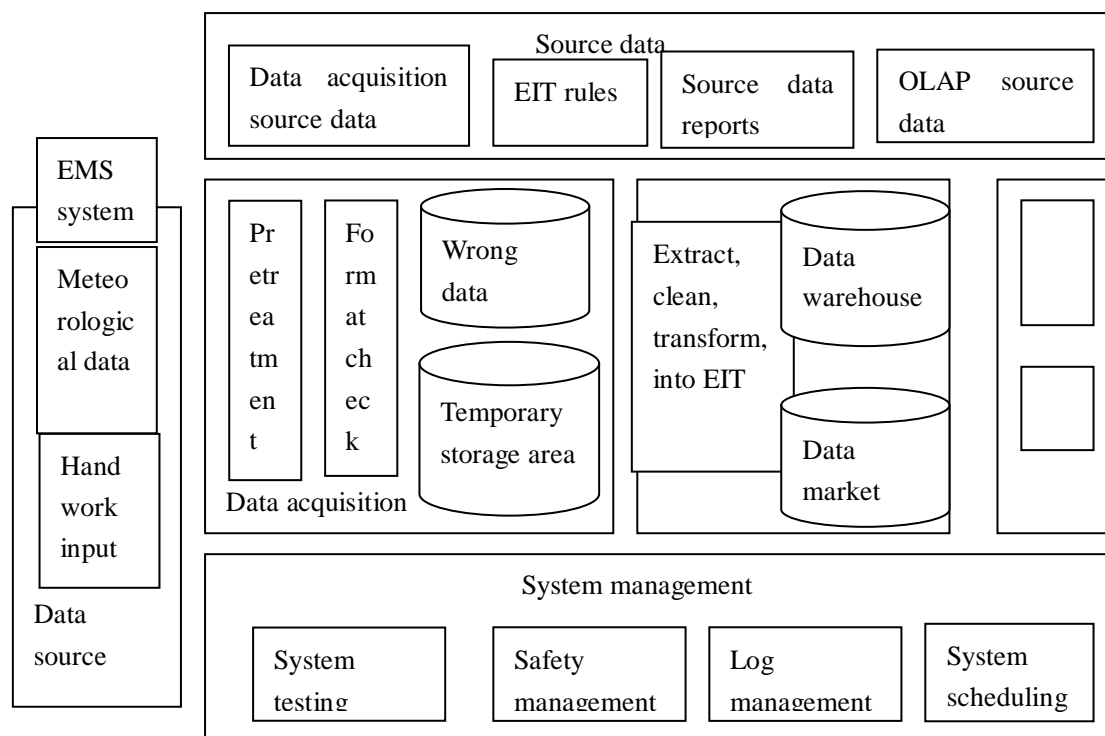

Figure 3. The structure of the data warehouse of power system

The type of data source may be various types of database, text or other binary data; The data source position can also be scattered distribution, because the source data and data warehouse have different location, set up a data acquisition layer which is used to check data package delaying, losing and retransmissing. Data collection will sent the correct data into the temporary storage area, and the wrong data into the wrong data area.

ETL process extracts data from the temporary storage area and transform, integrate, clean the data, after which, processes the data into a data warehouse; If the data warehouse involves less professional and theme, the data can only be processed into a number of data markets.

In application layer, the application subsystems or modules obtain and process data from data warehouse according to the need. Reports can be generated, such as CUBE which can conduct OLAP analysis or analysis of data mining.

V. CONCLUSION

Review the study of this paper, the following conclusions can be obtained:

1. It is necessary to stand at the height of a higher place to set up information system. From the global look of enterprise information, consider the relationship between data warehouse, data mining technology and other components. Different users have different requirements of information system and each component has the scopes which it is good at.

2. Like banking and retail, power system has a vast amounts of data, and with the improvement of distribution network automation, the system data will be a geometric progression growth, so use data warehouse in electric power system and its related technology is imperative.

3. Data warehouse cannot be built in a short period of time. Its technology is developing and the data warehouse itself is a solution, but not a special software. Its construction process is evolutionary.

REFERENCE

[1] Sharma R, Shah K, Shirahatti Y, et al. Data Mining: Concepts and Techniques[J]. Building A Data Warehouse, 2010, 32(43):619–630.

[2] Han J, Kamber M, Pei J. Data mining : concepts and techniques[M]// Morgan Kaufmann Publishers, 2001.

[3] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2008). The weka data mining software: an update. Sigkdd Explor Newsl, 11(1), 10-18.

[4] Steinbach, Michael. Introduction to data mining =[M]// China Machine Press,, 2010:236-238.

[5] Data, Chd. "Data Mining: Concepts And Techniques." San Francisco 29.S1(2010):1 - 18.

[6] Berkhin, Pavel. "Survey Of Clustering Data Mining Techniques." Grouping Multidimensional Data 43.1(2002):25--71.

[7] Han J, Kamber M. Data Mining: Concepts and Techniques[J]. Data Mining Concepts Models Methods & Algorithms Second Edition, 2000, 5(4):1 - 18.

[8] Berkhin P. A Survey of Clustering Data Mining Techniques[M]// Grouping Multidimensional Data. Springer Berlin Heidelberg, 2006:25-71.

[9] Jiang Yunyan Cheng Changsheng College of Computer Science and Technology, University D, Shanghai. Data Mining Approaches for Intrusion Detection[J]. Proceedings of Usenix Security Symposium, 1998, 16(4):18 - 20.

[10] Liu H, Motoda H. Feature Selection for Knowledge Discovery and Data Mining[J]. Springer Berlin, 1998, 5476. ISBN 978-3-642-01306-5. Springer Berlin Heidelberg(4):xviii.

[11] Shafer J C, Agrawal R, Mehta M. SPRINT: A Scalable Parallel Classifier for Data Mining[C]// Proceedings of the 22th International Conference on Very Large Data Bases. Morgan Kaufmann Publishers Inc., 1996:544-555.

[12] Pyle D. Data preparation for data mining[J]. Applied Artificial Intelligence: An International Journal, 2003, 17(5):375-381.