Analysis of Data Mining Based on Object Oriented Analysis Method

Jing Sun and Yu Sun
Department of Information, Yunnan normal university of China
Email: sunyu_km@hotmail.com

Abstract. Object oriented analysis is used to analyze the discipline of data mining. The subject of data mining is considered as a system, combining the concepts of class, attribute, method and relation between class and class in object oriented analysis method, and summarize the main task flow, pattern method, and some algorithm used in data mining. A systematic understanding of the content of data mining.

1. Data Mining

1.1. Concept of Data Mining
Data mining technology has been proposed in 1980s, and it is a new subject in database research at that time. However, data mining is widely used in all walks of life now, and it has become a foundation of various technologies. Data mining, as a branch of artificial intelligence (AI), provides a foundation for the development of AI. Data mining is popular in terms of "mining the intrinsic information in the data". This information can be different categories. It can be interrelated or different in the same group. By selecting different patterns to mine data, users can get the knowledge and information they need.

Data mining is a core step in the process of KDD. KDD (Knowledge discovery in databases) is knowledge discovery in databases [1]. In the basic course of the data mining is carried out by a series of operations on the target data to get the hidden pattern in the target data. [2]

Figure 1. Flow chart of KDD

1.2. The Process of Data Mining
Data mining, like KDD, has its own process of processing data. For the process of data mining, the process of human-machine interaction is repeated. For different data sources, the process of data mining is not exactly the same. Generally speaking, data mining generally includes the following five stages. [3]
Data preparation: For different tasks, data mining should deal with large amounts of data, resulting in large amount of data, complex data structure, redundant data, and vacant data and so on. Data preparation provides the foundation for data mining.

Data selection: If all data are processed, it will cause waste of manpower and material resources. So data selection is to select data sets related to this project task.

Data pretreatment: After data collection is set, data is preprocessed so that data can be used for us. The purpose of data preprocessing is to improve data quality.

Data mining process: according to the data information in the data warehouse, determine the task of data mining, select the appropriate analysis tool, and select the appropriate mining technology to deal with the data produced by the preprocessing.

Pattern assessment: From a business perspective, industry experts verify the correctness of data mining results.

Knowledge representation: the analysis information obtained from data mining is presented to the user in a visual manner, or stored in the knowledge base as a new knowledge for use by other applications.

1.3. Data Mining Model

1.3.1. Classification model. The classification model is based on the characteristics of the input data set, constructs the classification model, and maps the attribute sets of each data record to a pre defined class label through the classification model. [3]

1.3.2. Clustering model. The clustering mode divides the data set into groups (clusters). In clustering mode, the data between groups should be as similar as possible, and the data between different groups should be as varied as possible. The difference between clustering mode and classification mode is that the categories classified by clustering mode are unknown, while the classification of training set in classification mode is known.

1.3.3. Association model. The most typical example of association mode is shopping basket business. Association patterns are discovering associations between data sets. Association rules are implicit expressions like X to Y, where X and Y are disjoint data items. The association rules require the measurement of support degree and confidence, and the low support shows that the association is only an accidental result of data mining, and the confidence is a measure of reliability inference for the association rules.

2. Analysis of Data Mining by Object-Oriented Method

2.1. Concept of Object Oriented Analysis Method
The object oriented analysis method mainly identifies the relationships among attributes, classes, methods, classes and classes of the system. Object-Oriented Method is a methodology based on the concept of object. An object is a package composed of data and admissible operations, which has a direct correspondence with the objective entity, and a class of objects defines a group of objects with similar properties. Each inheritance is a way to share the attributes and operations of hierarchically related classes.

2.2. Use Case Analysis
In this paper, data mining is used to analyze the situation of data mining. For data mining system, there is usually only one participant, data preparation, selection, preprocessing, mining and other processes.
Table 1. Analysis of data mining system

| Numbered | class                    | input         | output                                                      |
|----------|--------------------------|---------------|-------------------------------------------------------------|
| 1        | Data preparation         | massive data  | Extracting valuable data sets                               |
| 2        | Data selection           | Data set      | Target data set related to the project task                 |
| 3        | Data pretreatment        | Target dataset| Preprocessing dataset                                       |
| 4        | Data mining process      | Preprocessing dataset | useful analysis information (mode).                       |
| 5        | Pattern assessment       | Pattern       | Correct knowledge                                           |
| 6        | Knowledge representation | Correct knowledge | Visual knowledge                                      |

Figure 2. Data mining condition diagram

2.3. Class Analysis

2.3.1. Class identification. According to the flow chart of data mining, we can see that the main classes in data mining are: data preparation module, data selection module, data preprocessing module, data mining module, model evaluation module and knowledge representation. This method is used to identify classes in data mining systems in order to better understand the attributes and methods in these modules.

For data mining module, it is divided into different mode processing methods. In this paper, three models are analyzed as sub classes of data mining class, including three sub categories: classification mode, clustering mode and association mode.
2.3.2. **Attribute in the class.** Data mining is a hidden knowledge mining from massive, incomplete, fuzzy and random data, so the data source is a necessary attribute in the process of data mining.

For the whole system of data mining:
- The property of data preparation is that the massive data source that it inputs is its attribute.
- The property of data selection is the set of data extracted by data preparation.
- The property of data preprocessing is the target data set determined by data selection.
- The property of data mining is preprocessed data obtained through data preprocessing.
- The attribute of the mode evaluation is the analysis information (pattern) obtained through data mining.
- The attribute of knowledge representation is the correct knowledge acquired through pattern assessment.

For the three subclasses of data mining:
- The attributes of classification model are input data set, training set, test set and class label.
- The attribute of the clustering model is the data set and the centroid.
- The attributes of association model are data set, support degree and confidence level.

2.3.3. **The method of class.** In data mining system, different classes correspond to different methods of operation of attributes. The following six methods and three subclasses corresponding to this article are introduced. For classification mode, clustering mode and association mode, the corresponding algorithm is a series of operations for attributes.

For the whole system of data mining:
- Data preparation: The methods include determining project objectives, data collection and acquisition, and data integration.
- The methods of data selection are: abstract the feature information needed in the data analysis, select the appropriate information collection method, and store the information that needs to be stored in the database.
- The methods of data preprocessing includes data cleaning, data integration, data specification, data transformation.
- The method of data mining process: determine the tasks of mining; choose mining technology; choose the appropriate mining tools. The corresponding algorithm is:
  - Classification mode: decision tree algorithm, neural network algorithm, naive Bayes algorithm, etc.
  - Association pattern: Apriori algorithm and so on
  - Clustering mode: K mean algorithm, agglomerative hierarchical clustering algorithm, DBSCAN algorithm, etc.
- The methods of pattern evaluation is to verify the correctness of data mining results and eliminate redundant or irrelevant patterns.
- The method of knowledge representation is to present the correct information obtained by the model to the user in a visual way, or to store it in the knowledge base as a new knowledge.

2.3.4. **The relation of class.** For a data mining system, it must include data preparation, data selection, data preprocessing, data mining, pattern evaluation, and knowledge representation. Each module is dependent on the relationship, and data mining includes a variety of analysis functions such as classification mode, association pattern, and cluster pattern and so on, so the data mining process can be the parent of these analysis methods.
2.3.5. Class Diagram

![Class Diagram of Data Mining]

3. References

[1] K. P. Soman 2009 *Insight into Data mining Theory and Practice* (Beijing: Machinery Industry Press)

[2] Han Qiuming, Li Wei and Li Huafeng 2009 *Application examples of data mining technology* (Beijing: Machinery Industry Press)

[3] Pang-Ning Tan, Michael Steinbach and Vipin Kumar 2011 *Introduction to Data Mining* (Beijing: The Posts and Telecommunications Press) P89-347

[4] Wang Guanghong and Jiang Ping 2004 A summary of data mining *JOURNAL OF TONGJI UNIVERSITY* 32