Web-Based Data Mining in System Design and Implementation

Jianhu Gong[1*] and Jianzhi Gong[2]

[1]Computer Science and Engineering Department, Guangdong Peizheng College, Guangzhou 510830, P.R. China

[2]Faculty of Management, University of Wisconsin-Platteville, USA

Abstract: From the traditional perspective, the system design process of data mining should be continuously broke through, based on web platform, making further analysis of model build. This paper combine four parts of whole web mining system to make discussion, making data mining models of the web meet scientific requirements and making relevance and effectiveness of the data mining process are fully guaranteed. Hoping through the research process of this article, the majority of business users can get a positive impact on their development, and at the same time, hoping that it can provide scientific and effective theory and model evidence for further study of data mining system.

Keywords: Data mining, model investigation, system design, web platform.

1. INTRODUCTION

During the design process of Web data mining system, we first have to fully recognize that there is a necessary connection between building a database and data mining systems. As to the definition of data mining, data mining is the exploration of knowledge towards the data in the database to gain a wider range of information. The final purpose of data mining is to get the high value information through corresponding data processing.

During the discussion of this article, first combine the design of the system structure mined by web to make related discussion, and it includes research process of the overall structure of data mining system and the functional model of web mining. Then the following is the model design research and discussion of data mining system, which is carried out by combining with the model analysis of data mining system, the whole model design of web mining system, analysis system of multi-dimensional OLAP and the design of database's table four aspects to get the design principles of web mining system. From the main line study we can see that the research and discussion process of this article has tightness, making the design of web data mining system more reasonable and providing a strong foundation for the future conduction of exploring work in depth.

2. THE DESIGN OF SYSTEM STRUCTURE MINED BY WEB

2.1. Whole Structure of Data Mining System

Data mining's characteristics and purpose is mainly reflected in two aspects, the first is the process of exploring and discovering knowledge in databases, the second is the

amount of data the is relatively large in the process of data mining. And in this cumbersome process, it can be divided into four stages.

First, it is the process of corresponding arrangement of the databases, since the amount of data is relatively large and complicated in this process, thus, to a certain extent, the efficiency of data mining should continue to be improved in order to achieve the ultimate goal in the process of data mining. When carrying out preparatory work, the date experiences corresponding data sorting, filtering, transformation and some other series of work. In the process of data arrangement, the spare data should be effectively cut and combined and so on, and the existing improper data can be effectively excluded. Then the data-related information is effectively screened and the data is effectively aggregated based on its source, and it is the most effective way for the screening process. Finally, there is scientific adjustment for the screened data, which will conduct scientific convert of their existing data format to reach the specific requirements of data mining work.

The second is the core part of the effective data mining in database, and in this process it requires mathematical algorithms when dealing with data scientifically. When dividing the data into groups, the implicit information hiding in the data should be considered, and the data's category should be merged effectively, so that data mining results can be intuitively reflected [1].

Third, it is effective evaluation of the result obtained by the process of data mining, and this process is "model evaluation" process which we often refer to. The important purpose of evaluating this model is to make scientific examination for the data to get corresponding measure standards in the process of data mining and evaluation and to achieve the ultimate goal of making the process of data mining scientific and reasonable. From the data analysis process, the relationship between model evaluation and data mining module is a process of exchanging work, and search results

can be accurately positioned through effective references of the degree of their interest got from the process of exchanging work, so that the finding that the degree of interest can state filtering standards of data can be illustrated.

The final is the expressing process of the definition of new knowledge in the process of data mining, and the means of expression has its diversity, such as text, data, tables, images and other forms of expression. From the process discussed above, we can establish a corresponding diagram of data mining system, and the specific building processes are shown in Fig. (**1**):

2.2. The Function Model of Web Mining

From the definition of data mining, the fundamental purpose of the data mining process is to explore and re-appear knowledge in the database. The algorithm form of data mining commonly comprises data statistic and data analysis, and it retrieves the meaningful and valuable information in the database. But the data content of the database is alarming, so the retrieval process is more complicated, and the degree of difficulty and task is larger. Therefore, there is an obvious contrast between the process of data mining and traditional data search. Meanwhile, according to different existence of data, data mining models play an important role in its work.

The characteristics of the contraction of data mining is that it can build mining model with strong adaptability based on the needs of different data information, which is quite flexible. From the build characteristics of data mining model's point of view, the process of model build can be built correspondingly according to the requirements of different data, and at this time, the search of data mining models has a strong adaptability when searching the types and characteristics of data, so that the problems of data information can be solved effectively, but from the function model build mined by web point of view , web mining model can be divided into six specific steps to carry out [2].

First, make effective references according to the specific requirements of data mining and give scientific analysis for the issues. Second, make full preparation for the relevant date needed in the process of systemic data analysis. Third, make effective analysis of the data of systemic research. Fourth, build individualized model of data mining according to the results of the analysis of systemic data. Fifth, give an individualized and effective test to its build process of data mining. Sixth, give updating and disposing according to test results. Through the build process discussed above, we can summarize the functional model of data mining as visualization, the specific construct shown in Fig. (**2**).
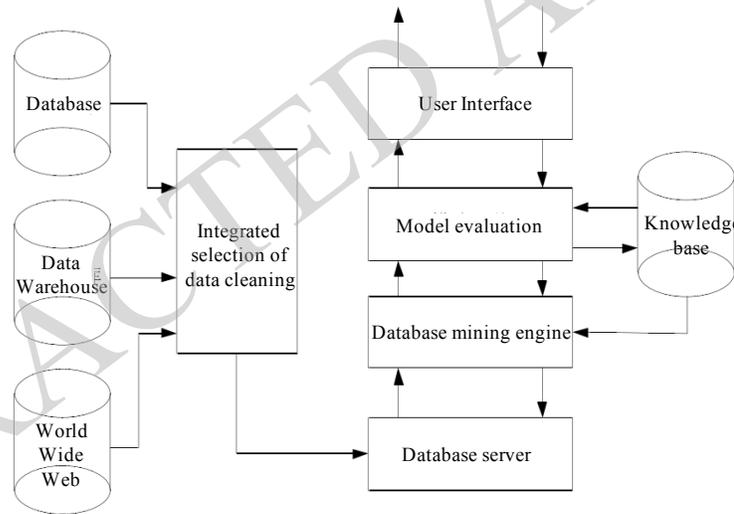
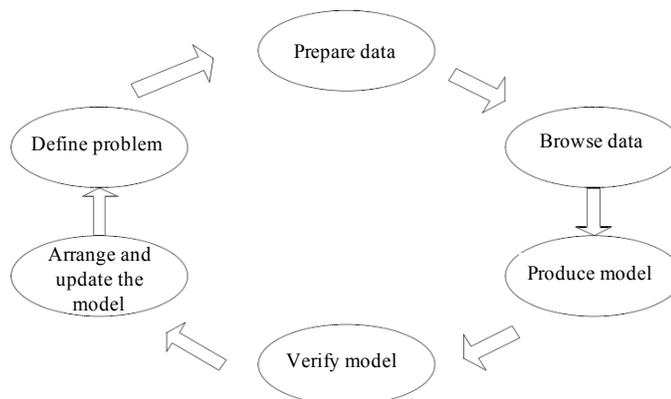**Fig. (1).** A typical diagram of data mining system.

Fig. (2). The functional module chart of data mining system.

3. MODEL DESIGN OF DATA MINING SYSTEM

3.1. Model Analysis of Data Mining System

What can be seen from the build process of system model in Fig. (**3**) is the data mining system of user's behavior is only one part in building of the whole data mining system. However, it requires the relative system to operate and make further analysis in turn, in order to carry out the specific analysis of the use's behavior. Fundamentally, the model analysis of data mining system still retains data analysis principle of the traditional model, through collecting the user's behavior information effectively, and carrying out data processing in the process of date arranging, achieving a comprehensive analysis. This is the core concept in the build of all the data model and is an important way to exclude, collate, analyze, and detect the large volume of information accordingly, making the analysis of data, to some extent, with higher effectiveness.

The so-called process of data extraction is the process of carrying out extraction and test on the data samples. Statistics and data during the process of statistics and arrangement of data, we can see that there are many irregularities and uncertainties among the data, these factors should be excluded effectively in the process of data extraction. The data we obtained in the process of model building of data mining system usually is not directly taken from various text documents, as well as one part is extracted directly from the database [3].

The process of data conversion is to convert extractive data respectively, and the expressions of the data format may have some differences in this process, existing corresponding inconsistence with the requirement of input data model. And by converting the model enables data to be further cleaned, classified and even merged, making mid-term preparation for the analyzing process of the data mining module.

From the general point of view, the process of data mining is made up of seven important aspects, the first is the process of data clustering, making effective polymerization according to the common characteristics of each type of data, and then followed by the association, classification, regression, prediction, sequence analysis and deviation analysis. These important steps are the results of the analysis of data model, and they also are critical components for the protection of their effectiveness.

3.2. The Whole Model Design of Web Mining System

During the building process of mining system of the users' interest, by combining the corresponding representation in Fig. (**4**), the specific process of description and the framework are as follows:
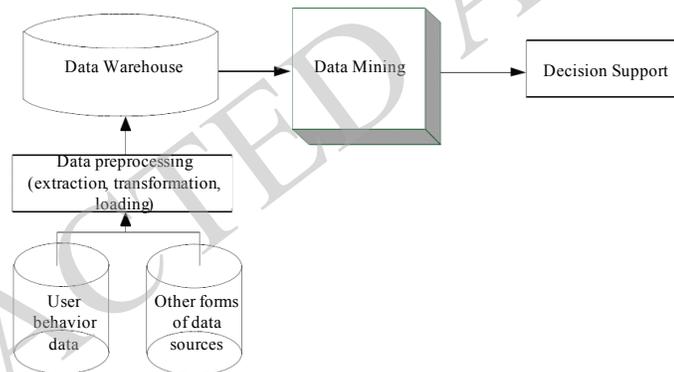

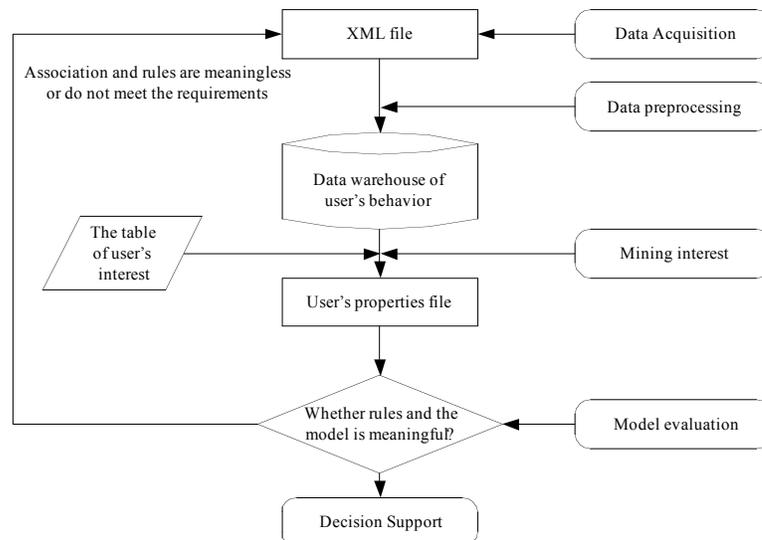**Fig. (3)**. Model of data mining system of users' behavior.

Fig. (4). Model of data mining system of users' behavior.

From the above figure, it can reflects the direction of web mining system model, where record of the user's network production, browsing the web site and other behaviors can be observed and analyzed. Thus the users' specific data characteristics can be fully summarized and it provides reference in the building of data mining model [4]. At the same time, integrating the characteristics of the user's data effectively, exploring the existing problem of the development direction of its data statistics effectively and providing important theory and data reference for the direction of its development in later period.

From the above figure, it can be clearly observed that data acquisition module, the module of data mining of the users' interest and data processing module comprised a web mining system and in the process of the following discussion these four modules are correspondingly presented.

Firstly, it is the data acquisition module. The so-called data acquisition module is the procedure of acquiring data efficiently, and its main manifestations is to make corresponding data collecting and collation through the record of the user's own daily behavior. Through this means provide effective reference for the process of users' data mining and count the valuable data information. This is a fundamental part for the building process of web mining system, and also important channels and ways of users' data source.

Secondly, it is the data pre-processing module. Literally, we can briefly understand this module that the so-called data preprocessing module is to get the results of the data collection module through the data, carry out preferential and targeted data preprocessing. And the main function of this module is to carry out preliminary screening and purification to the data, so as to provide an effective basis for the subsequent data processing.

The third one is the user's data mining module. Data mining module is the core part of the overall model building of web mining system, and it is specific process of effective analysis and synthesis for user data, and the produce results of the data reflect corresponding problem, then come to a targeted data mining conclusion, which is comprehensive statistics tables of the nature of the user's own interest.

The fourth one is the verification module of interest. The main role of this module is to carry out effective test according to the results produced in the before module and to pro-

vide effective protection for the pertinence of data analysis processes, making the data mining results and problems reach high agreement, which is specific instruction process of generating data validation.

3.3. Multidimensional OLAP Analysis

Using Web log file to create multidimensional view, and thus making multi-dimensional OLAP analysis, in OLAP process, we solve the difficulties encountered during pre-processing by using detailed analysis of layer by layer, meta-analysis, diced and sliced analysis techniques.

OLAP process can convert internal data structure of a Web log files, and can be used to find the first N users, as well as the highest number of times of visiting the page, which can easily identify potential customers and markets.

After established the view by using time dimension, resource and files dimension and tool dimension and established after-dimensional view, people can perform multi-dimensional analysis.

Contrary to drill-down analysis, Meta-analysis refers to the process from the similarity to the particularity. Slice analysis refers to the subset gained by setting up a dimension of a multidimensional array. Diced analysis is also similar to slice analysis, and it refers to a conclusion gained through setting a range of dimension members on one dimension of a multidimensional array [6].

3.4. Table Design of System Database

In the process of table design of the system database, we first combine the design flow of Fig. (**4**) to make related research process, and take it as a main table design program and the specific process is as follows:

During the research process of this part, it mainly conduct corresponding statistics of behavior time on internet of the customers, attributes and the parallel rule, so that customer feedback process can form effective table. This is an important part of acquiring the users' interest behavior data, providing references for effective discrimination of its users' interest.

Users' interest information is stored in the users' feedback form and this information lists the feedback identification, information description, file identification and reference

Table 1.   Users' feedback form.

Field Name	Field Type	Field Properties	Field Description
Feedback identification	Int	Non-empty	Distinguish between different entries
File identification	Int	Non-empty	Identify the file source
Description of information	VARCHAR	Non-empty	Representation of user interest
Reference value	Tinyint	Non-empty	Interest Reference

values of these four aspects of information.

It mainly used to distinguish the different the feedback information, and the source file is to be marked with the identification of documents to describe the users' interest which is the function of information describing and finally interest value fed back by users defined as reference value. As it can be seen, the four elements of the structure can be represented by Table **1**.

4. DESIGN PRINCIPLES OF WEB MINING SYSTEM

From the direction of development of the times, the using category of Internet technology is gradually expanding and the efficiency of corporate Internet technology is rising. The relationship between time and efficiency is gradually being more valued by enterprise, and they even develop a strong dependence on the processing function of Internet's data [7]. The process of users' data mining is the key to improving business productivity and sales capabilities, and the mining process of users' basic characteristics is gradually become the focus of attention, which is also an important factor in the survival and development of enterprises.

In this paper, during the design process of the web mining system, first combine the users' interest characteristics to development corresponding recipient process, the data collection modules form the basis of the system construction, and according to different data sources can carry out effective initial treatment, making the data with a high effectiveness, and its value can be initially reflected. Combining users' data mining module for data analysis, making the results of data analysis can be effectively guaranteed.

From the perspective of the overall development of the social, the use of Internet technology can gradually decrease the barriers among people, and increase openness of personal information. From this level, the demand for business users can be further satisfied, building a broad platform for mutual understanding between businesses and customers. For the process of mining users for business apart from the Internet all the means can not de conducted efficiently, from the perspective of data mining, data mining model is able to carry out effective analysis of customers' data with a large amount of information, making the process of data analysis more objective which can explain the corresponding problem.

5. CONCLUSION

This article above is research and discussion processes on the topic of the system design and implementation of data mining based on web. This process combined model analysis of data mining system, overall model design of web mining system, multi-dimensional OLAP analysis system and tables designed of the database this four aspects to start work , thus obtaining the corresponding design principles of web mining system. The combination makes the idea more closely, the build of data mining system more compact which guarantee the accuracy and authenticity of data processing.

CONFLICT OF INTEREST

The author confirms that this article content has no conflict of interest.

ACKNOWLEDGEMENT

REFERENCES

[1]   Y. Zhao, "a Web-based data mining system of electric power marketing", *Journal of Henan Polytechnic University: Natural Science*, vol. 29, no. 6, pp. 784-787, 2010.

[2]   W. J. Wu, "explore solutions on data mining system of a membership-based retail trade", vol. 15, pp. 23-24, 2011.

[3]   J. X. You, "Web-based data mining acquisition and application of website knowledge - take Dianping.com as an example", *Journal of Shanghai University: Natural Science*, vol. 20, no. 3, pp. 261-273, 2014.

[4]   D. H. Shan, "Study on the application of data mining technology in the Internet Retrieval", *Bulletin of science and technology*, vol. 30, no. 3, pp. 161-164, 2014.

[5]   Q. Gao, "Study on general framework of clustering in data mining", *Science Technology and Engineering*, vol. 16, pp. 112-118, 2014.

[6]   B. Shen, "complex networks and data mining: comparison and integration of the research paradigm", *Complex systems and Complexity Science*, vol. 11, no. 1, pp. 48-52, 2014.

[7]   S. Zhou, "The application of data mining in the management of journal client relationship", *Technology and publishing*, vol. 3, pp. 29-31, 2014.