The Application of Data Mining Technology in Data Service of Micro Service Architecture

Kun Liu¹², Hanjing Liu¹² and Xiaohan Ding¹²

Abstract

This paper introduces the application of data mining technology in data service. Data mining is a new technology, but its application time is long, and the application effect is obvious. Data mining technology can be used in enterprise customer service system, which can help enterprises to find potential customers, while retaining the most valuable customers. We pointed out the necessity of establishing data mining technology and service intelligence analysis system based on the combination of intelligence analysis and service characteristics and processes, the method of data mining, knowledge management ideas applied to intelligence analysis and service system.

Keywords: Data mining, technology, micro service architecture, data service

Introduction

With the rapid increase in the era of knowledge economy and information society, between the various disciplines cross penetration, on the one hand the amount of information explosion and information change accelerated, multitude of various types of information. On the other hand, people's demand for information has been developed from simple access to information and information, which is more personalized and professional. The traditional way of information work is difficult to grasp and deal with these complex information, it is more difficult to meet the growing demand for information. Change and innovation of information research and service work method is imminent, must through the information technology support, improve the ability of collecting, processing, analysis, processing and storage of information, broaden the scope of information services, accelerate the intelligence research work rhythm, shortened from intelligence gathering and intelligence to produce research results released by the cycle, in order to improve the corresponding the speed of information timeliness and intelligence services, improve the precision and reliability of information research achievements, in order to meet the needs of the information society and users.

With the explosive growth of network information, at the same time obtaining information in a convenient people also are plagued by large amounts of spam, a large number of duplicate information, relevant information and even wrong information filled in the search results, how to analyze and judge the real needs of users, to provide users with personalized custom service providers must become information problems. Data mining technology has become the best choice to solve this problem.

The advantage of data mining technology is prominent, so the scholars of various countries have carried out deep research on it. But so far, the definition of data mining technology is not

¹School of Information Science and Engineering, University of Jinan, Jinan 250022, PR China
²Shandong Provincial Key Laboratory of Network Based Intelligent Computing, University of Jinan, Jinan 250022, PR China
uniform. With the deep research and international cooperation, the data mining technology will deeply understand the data mining technology.

Service oriented architecture (SOA) is derived from early distributed computing. Driven by OMG and IONA, data services have become widely recognized norms. However, the commonly used relational database does not support SOA and object oriented programming, so it is difficult to deal with the data by SOA and process oriented programming. In order to use SOA for data access and processing, domestic and foreign research institutions have carried out a preliminary study. BEA has studied the use of data sources in the form of data services, and the establishment of enterprise data service layer for data management.

The Proposed Methodology

**Personalized data service and data mining technology.** Personalized data service refers to a kind of service which can meet the individual needs of information users, according to the specific service user's request, or through the analysis of the usage of individual users, and to provide its active users may need information service.

Data mining technology can obtain from large, incomplete, noisy, fuzzy and random data, in which the extraction of implicit, unknown, and potentially useful information and knowledge.

Personalized customization service is one of the effective methods to get the user's demand information. It obtains the user's personalized information through the user's customization, so as to understand the needs of users, to provide users with more accurate information services, improve customer satisfaction. Customized content includes information resources, interfaces and services.

Retrieval customization needs to fully support the user's personalized retrieval strategies, retrieval methods and retrieval results. The retrieval customization can be divided into the following aspects: the customization of personal search template, the customization of retrieval tools, the customization of retrieval expression mode, the customization of personal vocabulary, the customization of retrieval results, and the retrieval of historical analysis.

**Data mining process.** It is an important step of data mining to determine the purpose of mining, only the purpose is clear to develop effective models and methods, in order to ensure the success of data mining.

Select the data source: comprehensive consideration of the internal and external data of the service object, select the appropriate data source. Data preprocessing: on the basis of analysis
and Research on data integration, transformation, reduction, compression and other operations, and then determine the type of mining.

Select the appropriate data mining methods, construct data mining model to extract valuable knowledge and data from the target data, then to analyze and verify the results, adjust the data mining model, so as to ensure the reliability and practicability of the results.

The user's interest and preference, professional use habits are important elements of the development strategy of personalized information service, only to master these information to improve the information service of the pertinence, rationality, safety and operability. To achieve this purpose, the comprehensive application of various data mining techniques is needed.

Association rules. It is a description of the association between the data items in the database, according to the emergence of some items in a thing can be derived from another item in the same thing, that is, hidden in the data association or correlation between. Through the analysis of the correlation between the user set and the session set, we can find out the information of each user's interest, access pattern, etc. Association rules can be prepared in advance of the information that may be of interest to the service object, reduce the load time of information, and provide efficient service.

Cluster analysis. Clustering is to classify a group of individuals into categories according to their similarity. Such as user clustering, web page clustering. Cluster analysis can be used to generate specific information aggregation for specific target users. When a user group visits the user, it can provide real-time and dynamic content.

Statistical analysis. Statistical analysis is used to calculate the user's most visited web pages, the average access time per page, the average length of the browsing path data, in order to obtain the basic information of the user access to the system. In the personalized information service system, user visits, long time, large-scale access for information, for a period of time the access frequency can be given access to user needs and habits, to provide personalized information is presented on the basis of the establishment and active information organization.

Sequential pattern. In the web log, the user's access is marked by the time period, after data cleaning and transaction identification is a discontinuous time series. Through sequential pattern research, we can provide users with the most interesting information and improve the user's adhesion.

Data service. Direct access is closely related to the platform and data source types, such as structured query language (SQL). The program is generally not accessed directly through the SQL language, but through the application interface on a specific platform.

Indirect access is a platform independent data access method, which usually involves the integration or conversion of data sources. Data warehouse is a kind of physical integration method. Another special method is based on the data access application services, which can be used to access the database through the establishment of service in a single application, also can be accessed through the establishment of service in the enterprise integration platform.
Figure 2. Data service.

**Service oriented architecture for data services.** The idea of data service is based on providing data as a service to the user, which provides a simple and efficient way to access data. It can integrate various heterogeneous data, including the SQL statement can be used to query the data, such as relational databases; cannot use query data, such as text or Web data; also includes to access through the Web service data. It is not the same as the traditional Web services to access the data through the packaging application, but the direct encapsulation of the underlying data.

The concept of data service is put forward by BEA in 2005, and the data service can be constructed directly on the SQL statement or stored procedure. The data service entity is generally associated with a single entity in a business application, which has some characteristics of the object, as well as some properties of the relational data model. Following the BEA, Microsoft, IBM, and Oracle also put forward the concept of data services.

In commercial applications, there are some complex data queries and other operations. In this case, you can use data services to encapsulate SQL statements, functions, or stored procedures directly. When dealing with some non-relational data sources, such as XML, XQuery and XSLT can be used to provide XML services, XQuery can provide data query, XSLT can provide data conversion.

Query data is the most commonly used option in data integration, and the XML form of its SOAP response is also the most complex, because it contains a large amount of data. In the Web environment, you can use DOM or SAX parsing and transformation in the server data in XML, you can also use XSLT or AJAX to XML form data converted directly or indirectly for the visualization results in a web page.

The data layer contains a variety of heterogeneous data sources used in enterprise applications, including relational databases, XML databases and other data services supported data formats.

**Data service layer.** The data service layer provides a single access point for other programs within the enterprise to access data, ensuring that data is always extracted from a protected single source. The data service layer provides the following functions: standard data access operations such as create, read, search and delete data, provide the data source, data mapping and analysis between service providers and the underlying data source model, cross database platform differences, use files and directories as data sources across differences in non-database structure. By building the data service layer to encapsulate all the functions of centralized data
service all code, can object to other services in the application provides all details of data, so as to produce a solution based on SOA.

The application can access the data service directly to obtain the data that also can be accessed through the SOA layer. In distributed enterprise applications, public information can be obtained by registering UDDI and calling data services. For enterprise related business data services, can be provided by the enterprise service bus (ESB) on the way to provide access to internal users.

With the rapid development of database technology and the wide application of database management system, more and more data are accumulated. There are a lot of important information hidden behind the explosion of data, and the intelligence work requires a higher level of analysis in order to make better use of the data. At present, the database system can efficiently realize the functions of data input, query the statistics and so on, but cannot find the relationship and rules in the data.

Conclusion

The key of data information service is to find the user's interest quickly and accurately, and use the data mining technology to push the information in time to the customer. Intelligence planning and information resources integration cannot do without the support of IT technology, only the application of advanced information system and information tools, in order to achieve continuity and intelligence collection system, improve the timeliness of information processing; modern mass information cannot do without the support of modern intelligent processing technology, so the establishment of the information research and service system based on data mining technology is very necessary. We analyze the application of data service entity, data service layer in SOA, and use an example to show that the data service can be integrated into the SOA environment. In the enterprise environment, data services using SOAP messages to send large amounts of data will take up a lot of time, after the completion of the data transfer will also consume a lot of time. We can further study the combination of data transfer and compression technology to reduce the amount of data transmission or improve the transmission speed of Web service messages. The security problem is always an important problem in data access. We can further study how to enhance the security of the service and control the authorized access to the data.

References