基于 SOA 的电信行业数据挖掘应用平台研究

于明凯[1]　勾学荣[2]　张　勖[2]　柴莎莎[1]

(1.北京邮电大学 网络优化研究中心 北京 100088；2.北京邮电大学 网络教育学院 北京 100088)

Research of SOA-Based Data Mining Application Platform in Telecommunications Industry

Mingkai Yu (Network Optimize Center, Beijing University of Posts and Telecommunications, Beijing 100088)

Xuerong Gou (Network Education School, Beijing University of Posts and Telecommunications, Beijing 100088)

Xu Zhang (Network Education School, Beijing University of Posts and Telecommunications, Beijing 100088)

Shasha Chai (Network Optimize Center, Beijing University of Posts and Telecommunications, Beijing 100088)

Abstract: Building data mining application platform in telecommunications industry can take full advantage of business data in telecommunication systems to explore the potential patterns in user behavior and business trends, thereby providing more effective decision support to help enterprises cope with increasingly fierce industry competitions. This paper discusses data mining architecture and business demand for data mining in telecommunications industry, and on this basis puts forward an SOA-based data mining application platform solutions in telecommunications industry with the characteristic of integration, flexibility, and customization.

Key words: data mining; SOA; telecommunications industry; application platform

1 Introduction

Every industry is facing fierce competitions whether it is at home or abroad. Correct trends analysis, and timely decision-making are key sectors to the survival and development of enterprises. China's telecommunications industry also faces fierce competitions. It needs to make use of the existing operating data, dynamically capture market opportunities and detect problems in enterprises to make correct decisions quickly.

With competition in domestic telecommunications market getting keener and keener, business model of telecom operators transform gradually from “technology-driven” to “market-driven” and “customer- driven”. Massmarketing has lost its edge facing the diversified, hierarchical and personalized needs of customers. Marketing concept of insight marketing and precision marketing through in-depth data analysis based on customer information, value and behavior has gradually been accepted by major telecom operators. They also desire a lot to mining value from data of various application systems, which can reduce marketing costs and improve marketing effectiveness.

Using intelligence technology in enterprise, especially building data mining application platform is a better way to solve these problems and meet needs of business decisions. Data mining is a key technology of business intelligence. It is the senior process to identify effective, innovative, potentially useful and ultimately understandable models from data sets. It focuses on analysis and mining of the potential pattern of things and can realize forecasting of problems and future. So it can provide a more effective help for decision-making of enterprises.

So what marketing and decision-making supports can be provided by data mining for telecommunications industry How to build a data mining application platform in telecom in order to give this industry better service. In this article, it does an in-depth study in these issues, and based on analysis of data mining architecture and business demand for data mining in telecommunications industry, puts forward an SOA-based data mining application platform solutions in telecommunications industry with the characteristic of integration, flexibility and customization.

2 Architecture and Business Demand Analysis of Telecom Data Mining

2.1 Related works

Data mining technology is a very crucial part of the data analysis and usage technology contained in business intelligence. It emerged with the development of relational database, emergence of data warehouse and as well as challenges of massive data. Data mining can extract implied particular, in the past unknown and potential valuable information from data[1]. It is a science about extracting useful information from large amounts of data or from a database [2].

In recent years, data mining technology develops very quickly, and has more and more extensive applications. Major data analysis and database companies also launched their own data mining software and solutions, such as IBM DB2 data mining tools and solutions[3], etc. At the same time, some scholars have begun to study the application of data mining to telecom industry, such as data mining and prediction telecommunications network[4], data mining and analysis of log in telecommunications network[5], telecommunications call center data mining[6] and so on.

But these solutions still have some problems in integration, flexibility, and customization. And there is also no overall solution about data mining platform in telecom yet.

2.2 Architecture of telecom data mining

Data mining includes key steps such as data acquisition, data management and data usage and so on, just like business intelligence system. Based on the current general data mining structure[7] and the characteristics of telecommunications industry, we put forward a data mining architecture in telecom industry as shown in Fig.1.

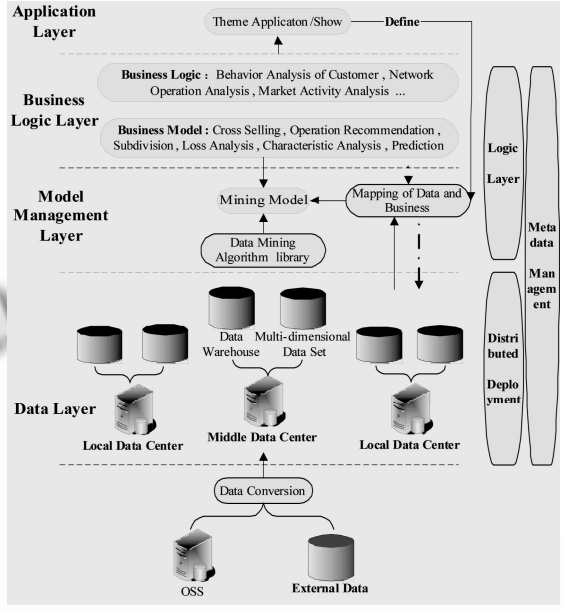

Fig.1 Architecture of telecom data mining

As shown above, the process of data mining in telecommunication industry is:

(1) Using the data (such as networks, business, billing, cost and other data) in telecom operation support system(OSS) and external data (such as marketing, and other data), after data conversion step, to generate the entire network-level data warehouse or small theme-oriented multi-dimensional data sets, which is similar to data mart. These form the data layer of telecom data mining.

(2) Then setting up data mining model for different business applications using different data. This process is the logic layer of data mining, which is divided into two components: business logic layer and model management layer. Through the establishment of model management layer, data mining algorithms are persistent to provide the core algorithm components for data mining applications, and it is easy to store and manage data mining-generated models and model results data. Business logic layer establishes the match relations among business demand, application and data mining algorithms. Then using mapping model between data and business, we can find

the corresponding data of business requirements and applications, combine data and algorithms to generate data mining model, finally conduct model calculations.

(3) The top-level is application layer. In this layer, we can show rules data or forecast data generated by data mining model to the end users through release application. Users can estimate data mining model and support decision-making by adopting model results. At the same time, users can manage all levels of data mining, definite data mining process using the application layer, thus guide the implementation of data mining tasks. This can help to achieve customization of data mining platform.

2.3 Demand analysis of telecom data mining

In this bottom-up structure of telecommunications data mining displayed in Fig.1, understanding of business is very crucial, because it guides the whole process of data mining. To be precise, business experts play main roles as technical experts play a supporting role in the process of data mining projects, this is the key issue about success or failure of data mining projects, or whether it is up to the expected results. Simply speaking, the business problem in the telecommunications industry data mining is to find customers' behavior and provide targeted services through the excavation of a large number of detailed records in telecommunications business systems. This will not only enhance the service levels to increase customers' loyalty, but more importantly can help to find services users that contribute highly to enterprise profits, and excavate its rules in order to increase sales, thereby improving enterprise's profitability.

There are longitudinal data mining tasks in telecom such as:

(1) Looking for the law of users' call time. Such as what time is the peak period of telephone calls? What is the time of different types of telecommunications services And so on.

(2) Law of the calls of customers divided by market segments(such as personal clients, governments departments, corporate clients). Whether same type customers after market segmentation have similar call model or not.

(3) Relationship between location and call of mobile business users. A clear and visual effect to understanding this relationship is telephone direct marketing can be more efficient in some industries (such as cross-selling of insurance products in banking industry).

(4) Mode of international long-distance call. Because the international long-distance call has high profit margins, understanding it's calling patterns and taking corresponding measures will play a significant role in telecom business performance.

(5) Cross-selling problem. For example, we first get the fragments of users through customers and their customized business. If a user customizes the business phone mailbox and other services, then users who belong to the same class and customize the same services except business phone mailbox, have a high probability to customize mobile phone mailbox service. So there has a large success rate to market phone-mail business service to these users.

To consider telecommunications data mining tasks horizontally, they will across telecommunications enterprise resource management, network optimization management, user behavior analysis, operations analysis, sales analysis and many other fields, resulting in tasks including the off-grid analysis, customer changes, channel changes, etc.

These tasks can be concluded as the below figure:

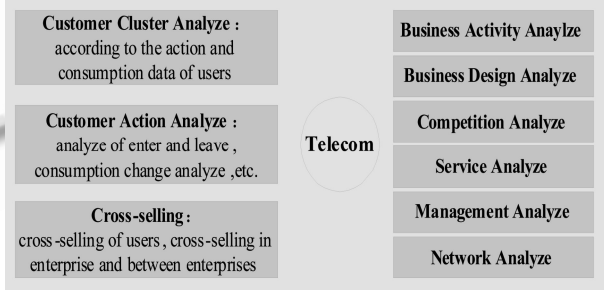

Fig.2　Data mining application in Telecom

From above we can see that, although the implementation process and architecture of data mining for telecom industry are relatively clear, but business and users play a leading role in data mining, and business and users have diversity, these lead to diverse and complex data mining tasks. At the same time, telecom mining data may come from a wide range of systems such as business operation system, investigation and analysis system, etc. So it is still quite difficult to implement data mining in

telecom, especially implement a flexible platform. Data mining involves a wide range of users, including business analysts, data mining analysts, software developers and data management staff, systems will become more complex because of collaboration problem between team members, and implementation cycle of data mining will become longer.

However, everything goes easy as the arrival of SOA (Service Oriented Architecture) technical standards.

3 SOA-Based Data Mining in Telecom

3.1 Related works

SOA is the application of service science in IT architecture. Service science[8] is put forward by IBM in 1998, it hopes to bring together ongoing work in computer science, operations research, industrial engineering, business strategy, management sciences, social and cognitive sciences and legal sciences to develop the skills required in a services-led economy. It can create new technology and market to provide more valuable service.

IBM defines SOA concept as follows[9]:

Service-Oriented Architecture (SOA) is an architectural style for creating an enterprise IT architecture that exploits the principles of service-orientation to achieve a tighter relationship between the business and the information systems that support the business.

SOA has the following characteristics [9]:

(1) It enhances the relationship between enterprise architecture and the business.

(2) It allows the building of composite applications as a set of integrated services.

(3) It provides flexible business processes.

It can be said that SOA provides a loosely coupled standard for interconnection of systems with different structures, it can integrate and assemble separate information systems as services, and it is an approach to integrate business as a group linked services. Service is obviously core of Service Oriented Architecture. Confined to the definition of service has no much meaning, we should focus on the key concepts behind services, including [9]:

(1) Business alignment: Services are not based on IT capabilities, but on what the business needs. Services business alignment is supported by service analysis and design techniques.

(2) Specifications: Services are self-contained and described in terms of interfaces, operations, semantics, dynamic behaviors, policies, and qualities of service.

(3) Reusability: Services reusability is supported by services granularity design decisions.

(4) Agreements: Services agreements are between entities, namely services providers and consumers. These agreements are based on services specification and not implementation.

(5) Hosting and discoverability: As they go through their life cycle, services are hosted and discoverable, as supported by services metadata, registries and repositories.

(6) Aggregation: Loosely-coupled services are aggregated into intra- or inter-enterprise business processes or composite applications.

These combined characteristics show that SOA is not just about "technology", but also about business requirements and needs.

At the same time, IBM puts forward SOA foundation architecture of SOA as shown in Fig.3 which gradually becomes industry standard of SOA [9].

With ESB (Enterprise Service Bus) as the connect point of entry, SOA foundation architecture provides interaction services, process services, information services, partners services, business application services, access services and other services, we can choose different services to build SOA application according to different needs of different enterprises.

3.2 SOA-based telecom data mining

The maturity of SOA standards and technology provides a good solution for the implementation of data mining platform in telecommunications industry.

Data mining platform in telecommunications industry will involve diverse and complex mining tasks, but each task is a relatively fixed and clear instance of data mining process, SOA architecture can provide process services to manage data mining business process. And although the data may come from operation system, investigation system, analysis system and such multiple systems, but SOA provides information services and access services to meet the needs of connection between heterogeneous systems. We are able to packet all levels

and all parts of data mining platform into services based on SOA services concept, thus we can use the package and integrate these services to implement data mining process.

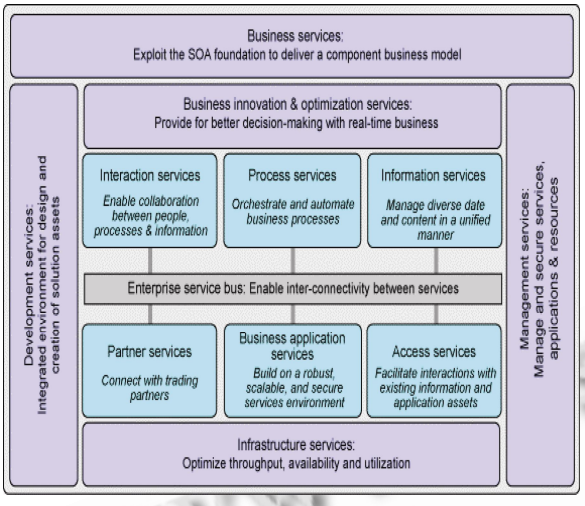

Fig.3 SOA foundation architecture

SOA-based telecommunications data mining lies in the definition and classification of services. After research and analysis, combined with telecom data mining architecture in unit 2.1, we believe that telecommunications data mining involves coarse-grained services as shown in Fig.4.

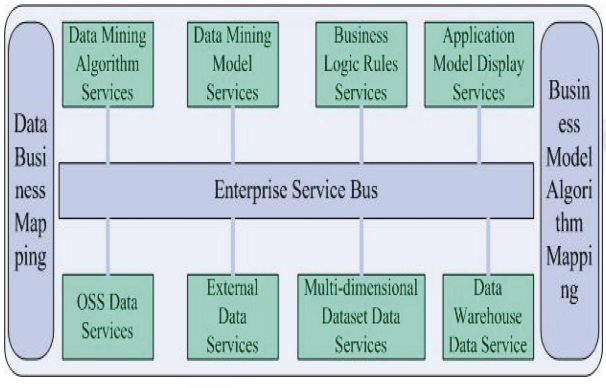

Fig.4 Classification of services in telecom data mining

As shown above, OSS data services and external data services correspond to the original data part of telecom data mining architecture, and adding the multi-dimensional data sets data services and data warehouse data services, these constitute the data layer of data mining. Over the data layer, data mining engine provides data mining algorithms and mining model services. At the same time, business logic and model rules engine provide rules services which record the mapping relationship between business logic, model and algorithm. Finally, result data generated by data mining model services is perceived and obtained by users by the application model displaying services.

How to implement and deploy SOA-based data mining platform after distinguishing every services in data mining? It is SCA (Service Component Architecture) and SDO (Service Data Objects) that really put SOA architecture into implementation.

The former is service component architecture, namely, service component model. Service component model is the realization of SOA thinking and service-based integration. It provides an executable model which can put individual service component into service network. It is a new programming model which has nothing to do with the language, and provides a unified call method, thus enables customers to package and call different types of components, such as POJO, EJB, process components, artificial interactive components, etc. through a standard interface. So when building enterprise applications, users no longer directly face the technical details of specific level, but build applications by the way of service components. This approach also gives users an enterprise application with good layered architecture which can separate business logic and IT logic clearly and which is not only easy to be built, but also easy to be changed and deployed.

And SDO (service data objects) is SCA's close partner. It is the basis of service integration, namely data integration. SDO provides heterogeneous data access and processing capabilities. Its first objective is to free developers from the underlying technology of processing data and thus make them more concern about business logic. The second purpose is to support majority of different programming languages, so that services can be implemented by using any language. SDO can achieve data transfer in integration process, so as to provide data foundation for service integration.

We are able to package and integrate above telecom data mining services using SCA and SDO, as shown in Fig. 5:

As shown, all parts of telecom data mining have

been packaged as SCA service components, so that it can be very convenient to calls various services, and data with the form of SDO can circulate easily between various components. We first deploy a process module, in which package data mining process in SCA component. The process can bind and call various parts of data mining services packaged in SCA components, and can easily achieve collaboration of data mining users. The calling steps are as following:

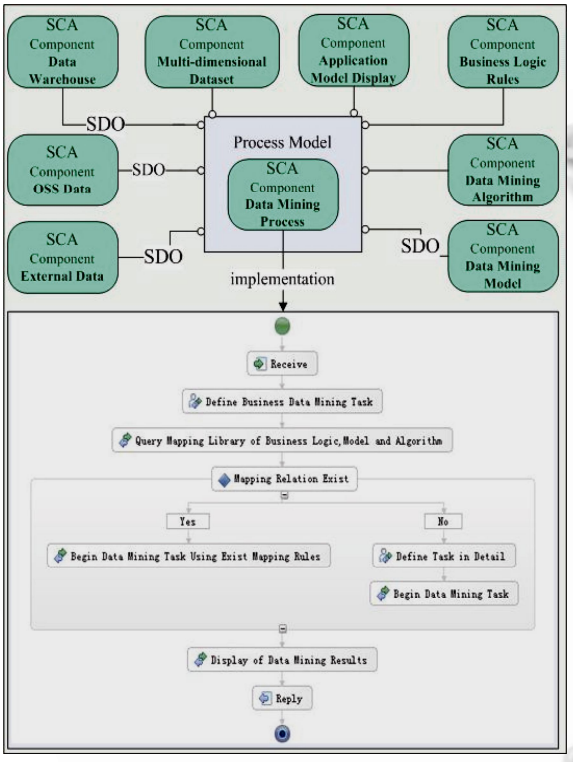

Fig.5 SOA-based telecom data mining process

(1) First, receive data mining tasks in form of natural language which are submitted by business analysts, and start data mining task process.

(2) Then call the application layer services, transmit tasks to data mining analysts to give a simple definition with the form of human tasks.

(3) The tasks which have been well defined go back to data mining process, automatically go to the business logic rules services to match corresponding rules, find correct data mining algorithms and business data to begin data mining operation. If they find no corresponding rules, the tasks go back to data mining analysts as human tasks to give a detailed definition, including corresponding relationship between tasks, business logic, and data mining algorithms. After definition, tasks will return back to the process, call data mining model services to start data mining operation.

(4) Data mining tasks which have entered data mining model SCA, can call data services SCA to get business data in form of SDO, and combine the data with data mining algorithms to begin calculation. Eventually generates excavation results which can be stored in model server.

(5) Finally call application layer service component to display excavation result to business analysts. Valuable data mining rules which are confirmed by business analysts will be added to the business logic rules library.

Various telecom data mining users can collaborate effectively through such a structure, which accelerates the pace of mining implementation. At the same time, all parts of data mining are packaged as standard services, which reduce integration complexity and improve reusability of various systems. The adoption of SCA, can easily achieve business logic rules which play a role in enhancing the flexibility of data mining application platform. And SDO shields business and data, makes it easy to define data mining task and implement data preparation, which make business users customize new data mining demands conveniently and flexibly.

3.3 Solution comparison

SOA-based data mining solution in telecom has an obvious advantage over traditional ones. We give a comparison between traditional and SOA-based solutions as shown in the following table.

Table 1 Solution comparison

Characteristic	Traditional	SOA-based
Integration	Medium	High, can take full use of existing systems
Flexibility	Low	High
Customizability	Low	High
Development Time	Long, more than half a year	Short, just a few weeks
Implement Cost	High	Low
Re-use ration	Low, it is separated.	High, give a standard to reuse service

To sum up, solution of SOA-based data mining application platform in telecommunications industry meets the needs of integration, flexibility and customizability well.

4 Conclusions and Outlook

Data mining application platform is critical to telecommunications industry, it can make full use of business data, analyze and excavate contained rules and predict development trend of business and users, thus provide guidance for business decision-making to cope with fierce competition. Based on research of telecom data mining architecture and mining demands, this paper puts forward an SOA-based telecom data mining application platform solution that allows building a integrated, flexible and customized telecom data mining platform become a reality. However, something remains to be optimized, such as there are still some shortcomings in performance of an SOA-based system, and because data mining tasks are based on huge amounts of data, speed of SOA-based data mining system will further slow down. And mappings between business logic and data mining algorithm are stored in rule engine, this knowledge database is not necessarily complete, but is perfecting in the usage process of platform, so the flexibility of telecom data mining platform also has limitations, needs further research and practice in the future.

References

1 Frawley W, Piatetsky-Shapiro G, Matheus C. Knowledge Discovery in Databases: An Overview. AI Magazine, 2001:213－228.

2 Hand D, Mannila H, Smyth P. Principles of Data Mining. MIT Press, Cambridge, MA, 2001.

3 Baragoin C, Andersen CM, Bayer S, Bent G, Lee J, chommer C. Mining your own business in telecoms using DB2 Intelligent Miner for Data. International Technical Support Organization, July 2001.

4 Sasisekharan R, Seshadri V. Weiss SM. Data mining and forecasting in large-scale telecommunication networks. IEEE Expert, 1996,11(1):37－43.

5 Zhang XL, Gong WJ, Narita TS. Business Intelligence in Telecommunication Enterprises:A Case Study of Log Data Analysis. Machine Learning and Cybernetics, 2006 International Conference on Aug, 2006: 1259－1264.

6 He SG, Li L, Qi ES. Study on the Continuous Quality Improvement of Telecommunication Call Centers Based on Data Mining. Service Systems and Service Management, 2007 International Conference, 2007,6:1－5.

7 Springer Berlin, Heidelberg. An architecture for distributed enterprise data mining. High-Performance Computing and Networking, Lecture Notes in Computer Science, 1999,1593:573－582.

8 Hefley B, Murphy W. Service Science, Management and Engineering. Series: Service Science: Research and Innovations in the Service Economy.

9 Portier B. IT Architect, IBM. SOA terminology overview: Service, architecture, governance, and business terms. 15 Nov 2006 Updated 24 May 2007. IBM developerworks

10 Du HT, Zhang B, Chen DF. Design and Actualization of SOA-based Data Mining System. Computer-Aided Industrial Design and Conceptual Design, 2008 (CAID/CD 2008):338－342.