



Efficient Data Extraction in Web Services

Sarthak Kabra, Uday Singh, V Prithvi
UG scholar
SRM University, Chennai
sarthak3594@gmail.com

R. Augustian Issac
Asst. Professor
SRM University, Chennai

Abstract— The actual data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown, interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection), and dependencies (association rule mining, sequential pattern mining). This usually involves using database techniques such as spatial indices. These patterns can then be seen as a kind of summary of the input data, and may be used in further analysis or, for example, in machine learning and predictive analytics. For example, the data mining step might identify multiple groups in the data, which can then be used to obtain more accurate prediction results by a decision support system. Neither the data collection, data preparation, nor result interpretation and reporting is part of the data mining step, but do belong to the overall KDD process as additional steps.

Keywords - web, data mining, algorithm, design

I. INTRODUCTION

Data mining is the process in which the potentially useful information and knowledge that is unknown before is extracted from a large amount of incomplete, unclear, obscure and random and practical application data. With the widespread application of information technology in various fields of the society, the primary data handled by data mining becomes more complex, structured or semi-structured, such as text, graph, image data, even the heterogeneous data

distributed in the network [2]. Mining algorithm plays an important role in data mining system. An effective data mining system necessarily has various mining algorithms. Generally speaking, these mining algorithms cannot be developed by one time, but rather updated continuously according to the development of algorithm and the requirement of application. It is difficult for the traditional data mining system to adapt to the needs and it is hard for the upgrade of algorithms library. Therefore, the research on how to dynamically add algorithm and automatically form the input and output interfaces according to different algorithms so as to realize the automatic invocation of data mining algorithms is the key issue which needs study for the development of algorithms library module in data mining system.

The introduction of Web Service technology in algorithms library module for data mining is good to realize the mining for heterogeneous data source. Meanwhile, by encapsulating various data mining algorithms through Web Service and saving the compressed algorithms to algorithms library, the data mining system and mining algorithm can be loosely interconnected and data mining algorithm can be invoked online through Web interface, which makes algorithm can be reused and convenient and realizes the dynamic management for algorithms library.

To this end, this paper introduces Web Service technology into data mining system, proposes the basic framework of



algorithms library module for data mining system based on Web Service, studies the description of algorithm metadata, and automatically builds Web interface components according to algorithm metadata, which tremendously reduces the complexity of the development for algorithms library in data mining system so that it is advantageous in cross platform, cross-language, deployment convenient as well as dynamically-managed algorithms library and can efficiently improve the maintenance, universality and flexibility of algorithms library for data mining system. [5] discussed about a method, Optimality results are presented for an end-to-end inference approach to correct (i.e., diagnose and repair) probabilistic network faults at minimum expected cost. One motivating application of using this end-to-end inference approach is an externally managed overlay network, where we cannot directly access and monitor nodes that are independently operated by different administrative domains, but instead we must infer failures via end-to-end measurements.

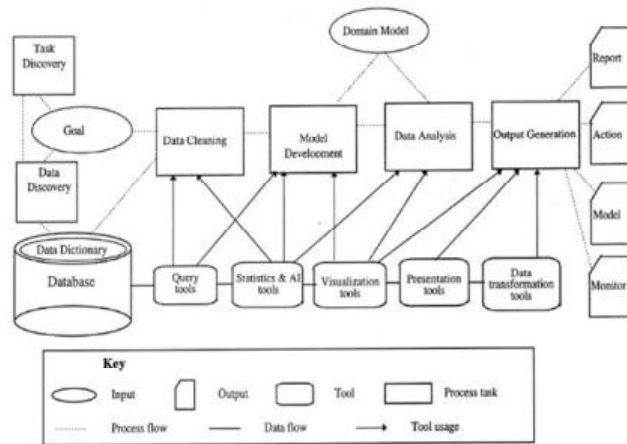
II. THE DESCRIPTION OF METADATA FOR DATA MINING ALGORITHM

In data mining system, the data has a complicated source and large in amount with various kinds, from structured data to non-structured data. We can say that data is the foundation for the operation of data mining system. In the process of data mining, it needs to use various universal or special algorithms, which can be provided either by the system or the exterior. As the important component in data mining system, algorithms library is responsible to supply and form mining modes for data mining system. With the constant improvement of traditional algorithms and the constant development of new algorithms, it needs to save and manage

various algorithms according to certain norms. In the meantime, various kinds of users are required to use the algorithms in line with certain norms. Therefore, in the development and maintenance process of algorithms library, it needs a powerful dynamic management function. For this purpose, the formulation of algorithm description and the basic norms for management is the key problems for the development of algorithms library. Metadata is the data about data and it is composed by the characters and symbols describing the resource features. Metadata can provide the standard and common description and method to various forms of digitalized information units and resource aggregate, which plays an important role for the organization and management for the resources. Describing the descriptive data and data structure in algorithms library of data mining system by using metadata can build the association relationship among figures, fields, data and terms in a convenient way. The uniform metadata pattern is good for sharing information and exchanges of information in different patterns. According to the actualized function of metadata, the input and output information for the algorithm so that the user can input parameter when using algorithm. Besides, the location and name of the algorithm should be provided for the convenience of the invocation of dynamic algorithm. For data mining system, the metadata is core in the system. It mainly aims to describe the basic attributes of the data and provide support to data mining. The whole process of the establishment and execution for data mining task is under the control of metadata. Therefore, the efficient management of the metadata will make the system uniform and strong. The description of the metadata for data mining algorithm by using XML language is good for the standardized management of



algorithm and the sharing operation of algorithm description, which will not only realize the flexible expansion for the data structure and effectively realize the loose connection among the data, algorithm and interface, but also solve the problem that the data structure and data application strongly depend on each other so as to actualize the mining for heterogeneous data source and support the building of flexible algorithms library for data mining.



III. THE CREATION OF DYNAMIC

The automatic creation technology of interface based on XML involves XML document described by the interface, listener of interface component, interface creator and the component library for programming language. To create an interface, it should first provide the description document for XML interface and listener of interface component. XML document for interface description can be compiled by XML editor or automatically created by other data sources. The description document for XML interface uses all the components and attributes in XML description interface, including the name, size, location, layout and the corresponding listener, etc. Listener of interface component is responsible for monitoring and handling operations of components in XML document described

by corresponding interface, such as button click operation. A series of operation brought by the component is realized by the item. In the analysing process of XML document described by the interface, the interface creator will dynamically add the listener according to the attribute value of the component in the document and bind it with corresponding components. When the operation of a certain component is triggered, the defined operation in its listener will be executed. Through analysing the description document for XML interface and invoking the constituent components in the component library for programming language, the interface creator binds corresponding listener for the component to automatically create the required user interface. The interface creator is the core for the automatic creation technology of the interface. It is in charge of analysing XML document for interface description, distinguishing the defined component information in the document, establishing corresponding component according to the component information and the used programming language, setting the attributes and binding the corresponding listener, adding the components into a component holder and finally returning the automatically created interface to the invoker. The creator mode is one kind of programming designs facing the object. The creator mode separates the complicated object's structure from expression so that the same building process can create different expressions.

To improve the universality of the system, the development mode of creator is introduced to make reasonable structure and performance for interface creator. Builder is to actualize the interface of Builder to construct and assemble the components of the product. It defines and clarifies the built expression and provides the interface for retrieving the product.



Director is to build an object that uses Builder interface. Product is the complicated object that is built. Concrete Builder creates the inner expression of the product and defines the assembly process, including defining the item of the component and assembling these components to the interface of the end product.

Since the invocation of Web Service is one part of button operation, Web Service can be introduced by the developer in advance and it can be used directly when it is needed. Because the address of Web Service is possible to change anytime and the content of Web Service will be updated constantly, it should adopt a dynamic invocation method for Web Service based on reflex mechanism. Thus, only by acquiring the location of service in the network and the name of the used method, the dynamic invocation of Web Service can be realized. The location information of the service and the name of the service method can be given through XML document.

IV. CHALLENGES AND CONSEQUENCES

There are many challenges faced in this approach. The concerns about the personal privacy have been increasing enormously recently especially when the internet is booming with social networks, e-commerce, forums, blogs.... Because of privacy issues, people are afraid of their personal information is collected and used in an unethical way that potentially causing them a lot of troubles. Security is a big issue. Businesses own information about their employees and customers including social security number, birthday, payroll and etc. However how properly this information is taken care is still in questions. There have been a lot of cases that hackers accessed and stole big data of customers from the big corporation such as Ford Motor Credit Company, Sony... with

so much personal and financial information available, the credit card stolen and identity theft become a big problem. Information is collected through data mining intended for the ethical purposes can be misused. This information may be exploited by unethical people or businesses to take benefits of vulnerable people or discriminate against a group of people. In addition, data mining technique is not perfectly accurate. Therefore, if inaccurate information is used for decision-making, it will cause serious consequence.

To overcome these challenges, we can propose many ideas such as, the research on how to dynamically add algorithm and automatically form the input and output interfaces according to different algorithms so as to realize the automatic invocation of data mining algorithms.

The introduction of Web Service technology in algorithms library module for data mining is good to realize the mining for heterogeneous data source. Designing a transparent system which decides and engages how and when information is accessed and by whom.

V. CONCLUSION

Compared with the traditional data mining system algorithms library, the introduction of Web Service technology into data mining system algorithm library realizes the separation of data, algorithm and interface. The loose and interconnected mode greatly reduces the complexity of the development for the algorithms library in data mining system and is convenient for the dynamic management of data mining algorithm library. Besides, based on the standard and agreement irrelevant to the platform, the system design adopting Web Service is featured in platform irrelevance and language irrelevance. The mining algorithm encapsulated by Web Service



can be realized by any languages and is separated from the binding of programming design language, which gives the maximum freedom to the algorithm design and improves the maintain ability, universality and flexibility of algorithm library. The algorithms library module in the data mining system built on the basis of Web Service technology is of critical significance to improving the efficiency of data mining.

REFERENCES

- [1] Ren Yanna, Lv Suhong, Wang Qiang "The Design of Algorithm for Data Mining System Used for Web Service ". Henan Agricultural University.
- [2] "The Principle and Algorithm of Data Mining" Wikipedia
- [3] David Tanir "Research And Trends in Data mining Technologies" Monash University, Australia.
- [4] S G Djorgovski "Data Mining And Exploration"
- [5] Christo Ananth, Mona, Kamali, Kausalya, Muthulakshmi, P.Arthy, "Efficient Cost Correction of Faulty Overlay nodes", International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE), Volume 1, Issue 1, August 2015, pp:26-28
- [6] M. Dunham, Data Mining: Introductory and Advanced Topics, Prentice-Hall, 2002.
- [7] Bill Palace "Technology Note Prepared For Management" Anderson Graduate School of Management at UCLA.
- [8] George H. John and Pat Langley "Statistic Versus Dynamic Sampling For Data Mining", Stanford University, CA.