

# Studies on Application of Data Mining Algorithms Based on XML Web Service

Qian Wan<sup>1</sup>, Qiangshan Zhang<sup>2</sup>

<sup>1</sup>College of Economics & Management, Huazhong Agricultural University, China

<sup>2</sup>Center of Network, Xinyang Vocational and Technical College, China

**Abstract:** Based on XML Web Service, the article constructs the management system of data mining algorithms, managing and calling mining algorithms, and dynamically adding mining algorithms according to different demands of algorithms. Thus the algorithms in data mining is effectively autocalled. Then the heterogeneous data can be effectively integrated through the data exchange platform.

**Keywords:** Data mining, Algorithms, XML Web Service, Data exchange.

## 1. INTRODUCTION

Data Mining is also called the knowledge discovery as a process in which useful data is extracted from large, vague, complex and potential sources, and then transformed into standard information output. With the rapid development of society, IT has been widely applied into various fields and becomes increasingly complicated, so its structure is also getting diversified, including texts, icons, picture data, even distributed data in network [1]. Data mining techniques can extract useful information from the original data and regenerate the unified and standard document to be used by anyone that needs it. In this process, mining algorithms play a major role. An effective data mining system covers a large number of mining algorithms, which is continuously exploited and improved with the development of technology and real demands. Therefore, in the mining system, a lot of mining algorithms needs to be managed and called, so an effective management system is necessary to be constructed to maintain the exclusive and shared library. Meanwhile, the new algorithms can also be added to the library for the better system [2]. The article, based on the Web Service Technology, studies how to manage and call mining algorithms, and dynamically add mining algorithm according to the demand of the auto call of mining algorithms.

## 2. RELEVANT KNOWLEDGE OF XML WEB SERVICE

XML is a source Markup Language. It defines a complete new document format and a method for saving data, and integrates data from different data sources. The other system can customize the new Markup Language and the supporting labels according to the actual demand. In this

way, various heterogeneous data in different platforms can be effectively transmitted and shared for application.

The structure of Web Service is essentially a service-oriented architecture used to solve the problem of heterogeneous data and the volatility of the data. The introduction of Web Service contributes to the realization of various heterogeneous and complex data source in data mining algorithm, and at the same time, a large amount of data mining algorithms can be encapsulated by Web Service. It is possible to call the mining algorithms online and compress the library to reuse algorithms available to dynamically manage the library.

### 2.1. XML Web SERVICE Architecture

With a view to constructing mining algorithms management system applicable for the distributed environment, the transfer and processing of data are taken as the whole center based on XML format's Web SERVICE. The basic architecture of system adopts three-tier architecture, respectively a presentation tier, a business-logic tier and data server tier. A presentation tier is the Web browser interface providing the interface for users and entire system; a business-logic tier, the most important tier, which affords various functions to be provided by the system; data server tier is mainly used to provide various XML documents generated by the connection and storage system of local or distant database. To enhance the security of system, it adopts three-tier structure, installing application in business tier computer, while there is no need to install application in users' computer. So, in the data mining process, the corresponding algorithms can be called repeatedly as it is more secure and applicable.

The distributed application based on XML Web Service takes the XSD pattern to define data dictionary. Data, element, property, type, etc. adapts to the XML pattern to define the language and the default constraining rule.

\*Address correspondence to these authors at the College of Economics & Management, Huazhong Agricultural University, China; E-mail: [zhangqs\\_xy@126.com](mailto:zhangqs_xy@126.com)

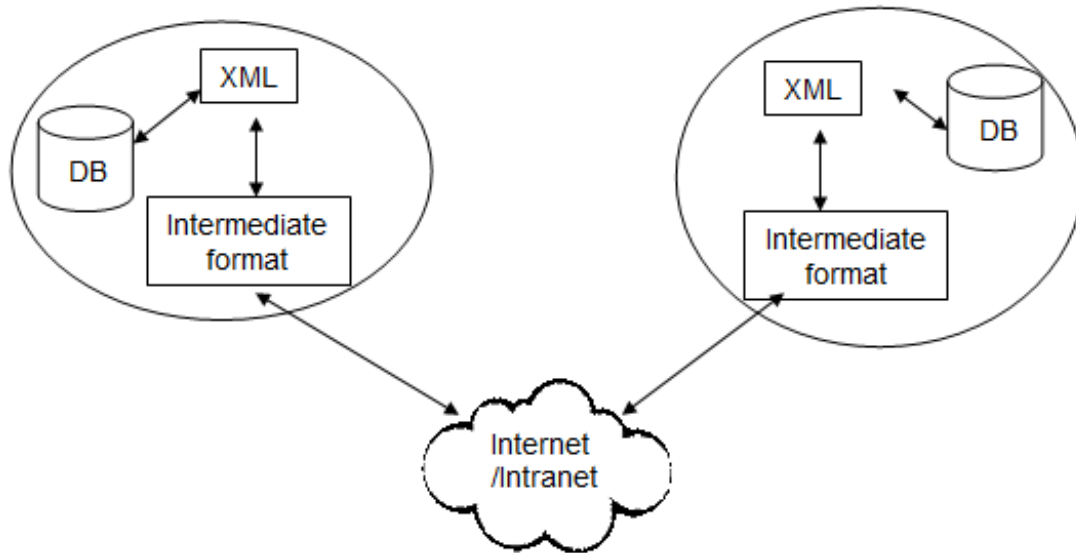


Fig. (1). Data mining engineering chart.

Web Service in this architectures is located in business facade layer, once it is in the middle layer: data visit and business logic and after correctly creating objects related to business logic, the representation tier can visit the business logic to expand distributed application program to Internet by Web SERVICE to achieve useful information by algorithms. In this way, the application program can provide the transparent access to the business logic in the middle tier. Web SERVICE not only provides service for application program, but also serves each other. The architecture of the application program renders it possible to realize the flexible form of end-user interface for the distributed application to be fulfilled [3].

**2.2. XML Platform for Data Exchange**

To realize the data exchanges among heterogeneous systems and the data mining process to effectively share the data and exchange data, this article designs a platform for data exchange based on XML to realize data mining by algorithms. After producing the useful information, we integrate the information to generate the common and standard XML document for users.

The data exchange platform sets XML format as a standard format to effectively transform different systems of application data which derives from data mining and is shared by generating standard document. XML as the center of platform users' view, is defined by XSL or XSLT, and uses XML Schema or DTD to construct legal XML document, so the process of data mining is realized. The process of data mining can be illustrated by Fig. (1).

In the platform of data exchange, IML manager, database manager, information center console and security mechanism effectively cooperate and use mining algorithms. The system completes data mining from network function knot to Intranet or Internet to achieve the homogeneous, heterogeneous data exchange and shared goal.

**3. ALGORITHM STUDY ON THE APPLICATION**

In the process of data mining, the important question is how to repeatedly call mining algorithms to effectively mine data and manage mining algorithms. We encapsulate a large number of data algorithms by introducing Web SERVICE, and compress library by Web Service to call mining algorithms and then to make algorithms to be reused for effectively managing the library dynamically.

**3.1. Application Process of Algorithms**

Mining library is involved with the algorithms register and the algorithms search, including the basic information of each algorithm provider. Library accepts the claim of algorithms application and preserves all other information of mining algorithms, thus is widely used by library. Meanwhile, mining library also accepts the algorithms claim to inquire the published algorithms whose condition is satisfied in library and return the relevant information [4]. In the process of data mining, when some algorithm is needed to be called, the system adds algorithms, and gains information based on XML data exchange platform. If the information doesn't contain the useful information, the system will redefine metadata, and choose proper mining algorithms, return and add algorithms until the useful information is obtained, and preserve the information to generate the standard XML document. Then the system will upload the standard document to server to be called by users to finish mining process. The data mining process is shown in Fig. (2).

In this process, algorithms generate the Algorithm Library through the Web SERVICE encapsulation. Algorithms metadata is described by XML standard language to realize useful data's extraction, utilization, sharing by means of the data exchange platform to reduce the complexity of library module. Thus the maintenance capacity of library will be greatly improved and the dynamic management of library will be realized.

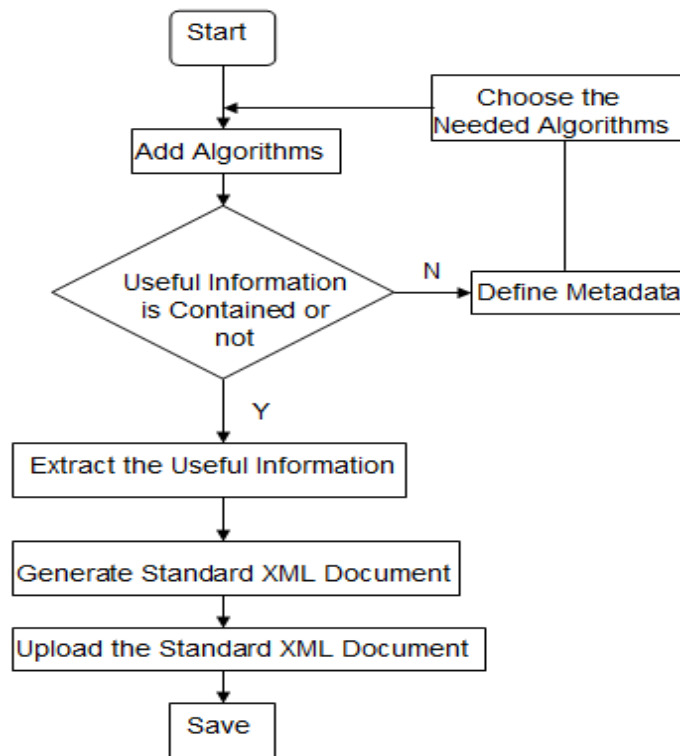


Fig. (2). Data Mining Process Chart.

In cloud computing environment, data storage infrastructure is the center of storage which is composed of a large number of store devices in different positions. These decentralized storage devices make use of their distributed systems to aggregate different resources into a high-reliability whole. Data mining algorithm is to use these file systems to effectively access between different devices. Users implement the cloud data access operation through the cloud service interface.

### 3.2. A Case of One Type Metadata's Definition

In practice, we can design a sample, and set data mining step according to the actual situation to clean the data which is collected before, to remove outliers, data smoothing (shift average), to establish the character pattern of the sample collection  $\{(m_i, y_i)\}_{i=1,2,\dots,n}$ , to structure classifier  $f(m)$ , and to meet  $f(m_i) = y_i$ . From this sample characteristics, model is drawn to remove the secondary feature model for principal component analysis to establish quality classification model (decision tree) and correlation analysis. The test sample set is examined to test the result of the mining. Finally, output decision tree or related analysis results are obtained and thus the process of data mining is completed.

Supposing that the total data point is  $m$ , the total range is divided into  $n$ , to ensure the area of each interval divided into equal in dividing, each interval of data point can be thought to be roughly equal. That is  $m/n$ . Therefore, the required range of final calculation can be determined by comparing the value of  $m/n$  and  $k$ . The process of data screening for metadata is as follows:

Input:  $k$  and  $w$ ;

Output: Data Partitioning which will be used as the data source of MapReduce.

1.  $p = \{K / (m / n)\}$ ;
2. Scan metadata table, get line 1 to  $p$ ;
3. For  $i=1$  to  $p$ ;
4. For  $j=1$  to  $q$ ;
5.  $V_{min}[i][j] = t_{min}/m$
6.  $V_{min}[i][j] = t_{min}/m$
7. end for;
8. compare all pairs of  $(V_{min}[i][j], V_{max}[i][j])$ ;
9. if  $V_{max}[i][j] \leq V_{min}[i][j]$ ;
10. delete data partitioning  $s$ ;
11. else;
12. output  $s$ ;
13. end for;

A simple defining process can be constructed to make the defining method easy in the process of data mining and bring it into full role in the Common Language Runtime (CLR). The following case shows two methods named MyApp. The main method is to program the entry point, and adding method only returns the sum of two parameters.

```

using System;
public class MyApp {
public static int Main(){
int ValueOne = 10;
int ValueTwo = 20;
Console.WriteLine("The Value is: {0}",
Add(ValueOne, ValueTwo));
return 0;
}
public static int Add(int One, int Two){
return (One + Two); }
}

```

## CONCLUSION

When XML Web SERVICE architecture is introduced into the data mining library, the problem of heterogeneous data and variability can be solved and the method can contribute to realizing the mining of various heterogeneous and complex data source. Meanwhile, a large number of data mining algorithms can be encapsulated by Web SERVICE, and we can also call the mining algorithms online and compress the library for the repeated use of the algorithms. By doing so, we can dynamically manage the library. Since the library is

encapsulated into Web SERVICE and stored in separate locations. The algorithms can be separated from the program design language. It can be of great significance: on the one hand, the freedom of algorithmic design can be effectively achieved. On the other, the system expenses at the Client can be reduced.

## CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

## ACKNOWLEDGEMENTS

This work was supported by a grant from Higher Education of Henan Province Fund for Young Leading Teachers (No. 2011GGJS-292).

## REFERENCES

- [1] J. Xu, Y. Lu, and Y. Wu, "Clients Relationship Management System Based on Web. Service," *Computer Engineering*, no. 19, pp. 26-29, 2009.
- [2] K. Luo, M. Lin, and D. Xi. "Review on Classification Algorithm in Data Mining," *Computer Engineering*, no. 1, pp. 3-4, 2005.
- [3] Q. Zhang, L. Bai, and Z. Yan. "Design of Digital Campus Data Exchange Platform Based on XML," *Journal of Jiling Architectural and Civil Engineering Institute*, no. 1, pp. 62-64, 2013.
- [4] L. Jia, "Design of Data Mining System algorithms Based on Web Service," *Computer Knowledge and Technology*, no. 9, pp. 2951-2954, 2013.

Received: May 26, 2015

Revised: July 14, 2015

Accepted: August 10, 2015

© Wan and Zhang; Licensee Bentham Open.

This is an open access article licensed under the terms of the (<https://creativecommons.org/licenses/by/4.0/legalcode>), which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.