Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 2, Issue. 7, July 2013, pg.224 – 230

RESEARCH ARTICLE

A Pervasive Data Analytic Web Service for Educational Data Mining

Shana. J[1], T. Venkatachalam[2]

[1]Department of MCA, Coimbatore Institute of Technology, Coimbatore, India

[2]Department of Physics, Coimbatore Institute of Technology, Coimbatore, India

[1] shana@cit.edu.in; [2] tvenkatachalam@cit.edu.in

Abstract— Educational institutions can apply data analytical techniques on the large amount of data that is generated within. This would result in deriving useful knowledge from the data and thus help in decision making. In this paper we suggest a framework for pervasive data analytic web service which can be integrated into any university management system or built as a standalone application easily. Being a web service allows it to be accessed by any type of application, from desktop to mobile phones inside the campus and can even be hosted onto the private cloud of the institution. The web service based model provides greater advantages over the tradition data mining systems. We have built a prototype of this model to work with k-means clustering algorithm on a sample student data to demonstrate the capabilities and limitations of the proposed framework. And we have developed three different types of client applications to make it truly pervasive.

Key Terms: - web service; educational data mining; clustering and SOA

I. INTRODUCTION

Large amount of data in educational institutions are generated in the form of personal information, academic data, fee payment data and much more. Educational data mining is one of the important areas that enable institutions to derive hidden and useful knowledge from student data. This would help the institution in many ways to improve it.

Analysis software for the education domain differs from mining the financial data or the business data in general. So the traditional commercial analysis software cannot satisfy the needs of the institution. Educational data mining generally emphasizes reducing learning into small components that can be analyzed and then influenced by software that adapts to the student [15]. As far as we know, there is no web service for educational data mining which allows an end-user to extract patterns and models by simply sending his data file without having to carry out the tedious job of selecting attributes, pre-processing and setting data mining algorithms. A service like this does not only offer non-expert data miners a tool for analysis but also facilitates the work of the expert data miners who can use it to obtain initial patterns easily and quickly. Commercially cloud based analytics have been offered now by companies like Amazon, Google and others. But they have many issues in terms of cost and security to be adopted by educational organizations. Our work aims at proposing a model that implements the data mining techniques used for data analysis as a service based architecture to be accessed by pervasive applications inside the campus. This would enable the faculty to access the service from any type of client application, irrespective of the language in which the service is built. In this paper we show how a web application, an android app and a console application can access the data analytic web service over the network. This would provide a greater flexibility to integrate even with the legacy client application or the latest kind of mobile application. Also the model can be extended to work on the private cloud

of the institution there by leveraging the power of cloud computing. This framework is applicable not only to educational domain but can also be used by any other organization which needs to analyze data. Of late cloud computing has taken a centre stage for it is able to deliver everything as a service and the organization need no longer buy expensive hardware or software to build the infrastructure needed. And web service is one of the many technologies that are helping cloud computing to achieve its objective. The web services approach uses standards-based interfaces for connecting data providers with data users. The network strives to go beyond searching and visualizing data to include data processing and analysis services to allow users to create new content. From a technical perspective, the web service technologies consist of a collection of standard protocols that enable the creation, distribution, discovery and integration of software components over the internet. Central to the Web service technologies are the concepts of "software as service" and "platform independence" [2]. This framework when implemented in the educational institution can benefit the faculty and the decision makers to analyze the student data more effectively.

This paper is organized as follows. Section II discusses the basic concepts of web service and its architecture. It also gives an introduction to data analytics and data mining. Section III is the literature survey of the work done in the related area. Section IV describes the features of the proposed system and the implementation of the prototype.

II. THE BACKGROUND

A. Web Service Architecture

Web services provide a standard means of interoperating between different software applications, running on a variety of platforms and/or frameworks. When we talk about a data mining service we understand "service" as a software product which offers a solution or gives an answer to the needs of a customer, this being either a person or another software application. So, there are at least two parties involved, the service provider and the service consumer, although a third party could exist, a service broker, which would act as the intermediary. The web services architecture is interoperability architecture. It identifies those global elements of the global web services network that are required in order to ensure interoperability between Web services [14].

From a technical perspective, the web service technologies consist of a collection of standard protocols that enable the creation, distribution, discovery and integration of software components over the Internet. Central to the web service technologies are the concepts of "software as service" and "platform independence."

Fig 1 depicts the general architecture of web service.

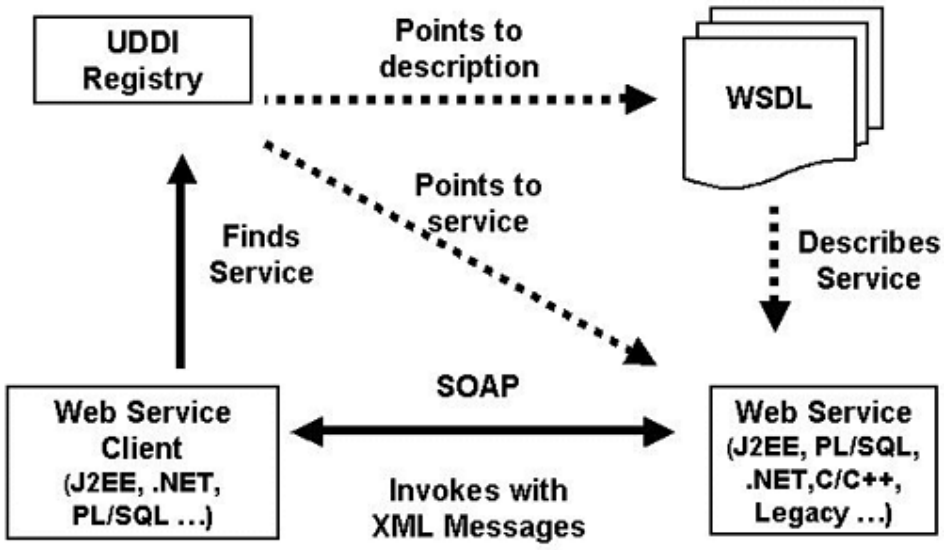

Fig 1 Components of a web service

The communications (operations) among the three Web service entities rely on XML and use Simple Object Access Protocol (SOAP). SOAP messages are commonly exchanged over HTTP, even though other bindings are possible. The service descriptions are published using the Web Services Description Language (WSDL), which provides information on how to use a Web service, including a description of the service methods and

binding information. UDDI is the service registry defined by the standard bodies for SOAP-based Web services. This web service is built using Axis2 web service engine [13, 20].

B. Data Analytics and Data Mining

Data analysis is a body of methods that help to describe facts, detect patterns, develop explanations, and test hypotheses. It is used in all of the sciences. It is used in business, in administration, and in policy [5]. It involves lot of statistical and machine learning techniques to extract useful information from huge database and data warehouse. Data Mining is a particular data analysis technique that focuses on modeling and knowledge discovery for predictive rather than purely descriptive purposes. Business Intelligence covers data analysis that relies heavily on aggregation, focusing on business information. In statistical applications, some people divide data analysis into descriptive statistics, exploratory data analysis (EDA), and confirmatory data analysis (CDA). EDA focuses on discovering new features in the data and CDA on confirming or falsifying existing hypotheses. Predictive analytics focuses on application of statistical or structural models for predictive forecasting or classification, while text analytics applies statistical, linguistic, and structural techniques to extract and classify information from textual sources, a species of unstructured data. All are varieties of data analysis.

Educational data mining is emerging as a research area with a suite of computational and psychological methods and research approaches for understanding how students learn. New computer-supported interactive learning methods and tools—intelligent tutoring systems, simulations, games—have opened up opportunities to collect and analyze student data, to discover patterns and trends in those data, and to make new discoveries and test hypotheses about how students learn. Data collected from online learning systems can be aggregated over large numbers of students and can contain many variables that data mining algorithms can explore for model building [15]. Many kinds of algorithms exist for association rule extraction, classification, prediction, clustering, trend analysis and deviation analysis in data mining. These algorithms are implemented in many tradition data analytic software. There is little knowledge about any educational data mining software being implemented as a web based service. This work highlights the benefits of web service oriented framework for data analytics.

III. RELATED WORKS

Literature survey does not produce any concrete knowledge of web service based analytics. There are many web service based applications being developed for various functionalities in different domains. Similarly data mining techniques have been applied in educational data to derive useful patterns. Apart from some commercial data analytic service from Google's Big Query for enterprise there is no such service being used. The base idea came from the study of these papers. [7] describes an automatic advice for scientists in their data analysis using data analysis web service.[6] describes the development and application of web service for comparing US and global emission inventories.[9] specifies an on demand BI architecture on cloud for the healthcare sector. [8] Offers a wide range of techniques that support the transformation of data analysis including web service technology. [10] Proposes a novel infrastructure that offers complex and optimized query facilities for web services through combined invocation of different web services. Academic analytics is used to derived intelligent information from academic data and used in higher education in decision making. Many notable works have been done in the field of educational data mining. [18] have used Open Source software for predicting student's results using Naïve Bayesian Learner and Naïve Bayesian predictor.[17] listed some of the applications of data mining to higher education, and provided some case studies to showcase the application of data mining to the student retention problem.[19] has performed an empirical study to identify the factors that influence the result of students in a particular course in India from their academic data.[1] uses three data mining techniques, namely, classification trees, multivariate adaptive regression splines (MARS), and neural networks to study the student retention problem.

IV. PROPOSED SYSTEM AND ITS FEATURES

A. Architecture

The proposed system follows the three tier architecture as shown in figure 2. The Data tier holds the dataset that is to be analyzed. Here we have used MySql as the back-end. The Application or the business tier implements the analysis techniques (web service) and the presentation tier displays the result of the analysis. The Analytic web service is hosted in the web server and the three different client applications can access the service from anywhere.

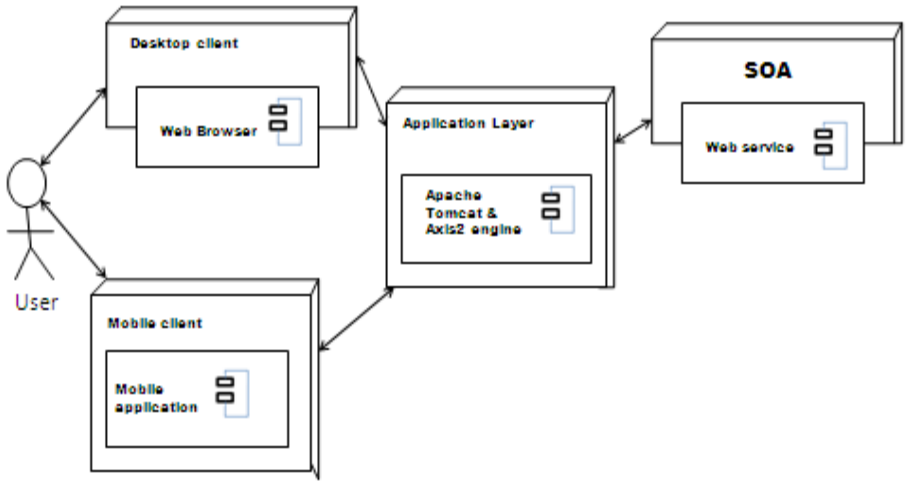

Fig 2 Components of the proposed system

The web service implements the core functionality of the system. The service reads the database and performs cluster analysis on it. Preprocessing of the data to remove noise required is also implemented in the service. The web service implements the K-means clustering algorithm depicted in Table 1. The web service is developed using Java and Axis2 API.

B. K-Means Clustering Algorithm

Clustering method is one of the data mining methods that allow grouping the data items according to similarity between them. It is an unsupervised learning algorithm. Table 1 shows the steps involved in the clustering process.

The K-means clustering algorithm has been implemented as a web service here. A sample of 250 student course data containing the internal and external marks is given as input to the web service ,with the number of clusters , k=2.The result of analysis (clustering) is displayed on to the screen in fig 4 and fig 5

Table 1 K-Means clustering algorithm

Input:

K: the number of clusters.

D: a dataset containing n objects.

Output:

A set of k clusters.

Method:

1. arbitrarily chooses k objects from D as the initial cluster
2. Re-assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster;
3. Update the cluster means, i.e., calculate the mean value of the objects for each cluster;
4. Repeat step 2 and step 3 until there is no change in clusters.

C. Snapshots of the prototype application

The Data Analytic web service was developed and hosted in Apache Tomcat Web Server. Figure 3 shows the screen shot of the Web Service Description Language (WSDL) file for the hosted web service. It is an XML file describing the web service and the service it provides. It is this file that the client application makes use of while connected to the web service. Three client applications were developed in different languages and platforms which have to connect to the service and send the request for the kind of analysis to be done. Figure 4 shows the Java console application displaying the two cluster values. Figure 5 shows the C#.NET application web form and Figure 6 shows the Android application displaying the cluster values. All the three applications use the same WSDL file to access the data analytic web service. Here the connecting network is LAN.

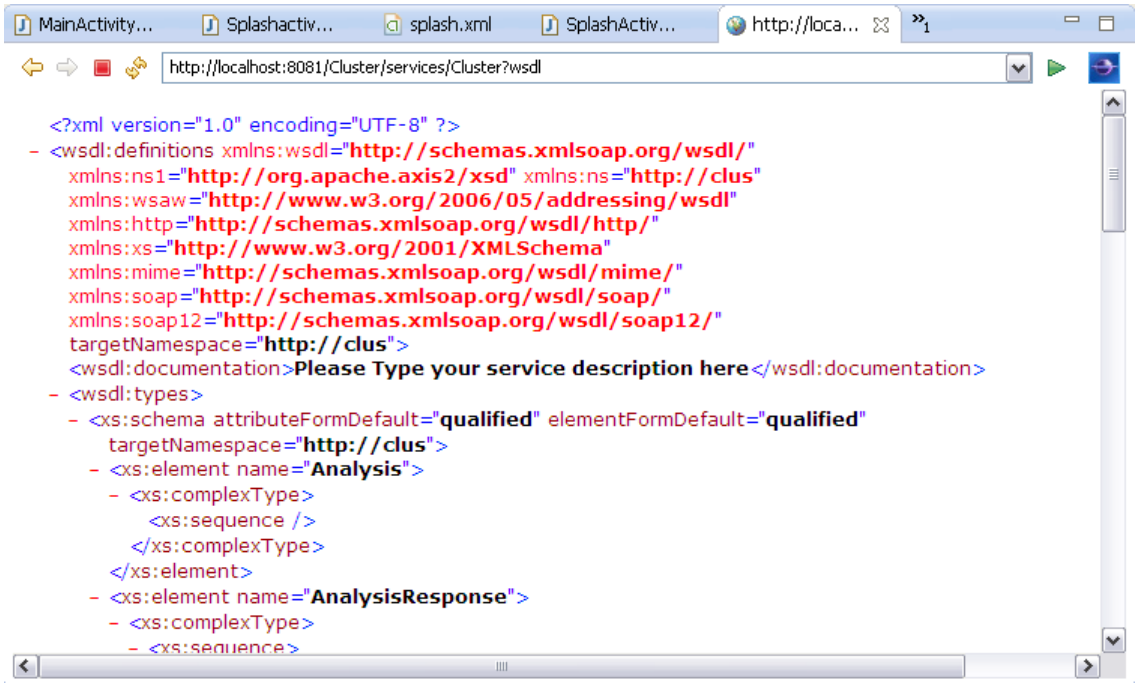

Fig 3: Screen shot showing the WSDL file for the hosted web

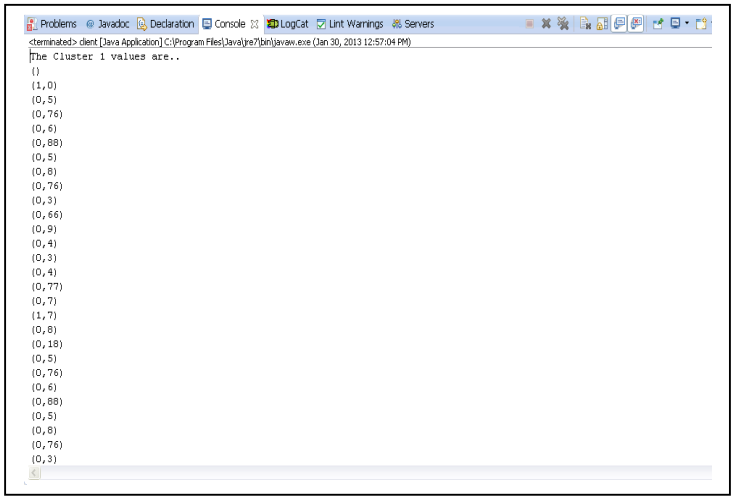

Fig 4: Java Client Application displaying the cluster values

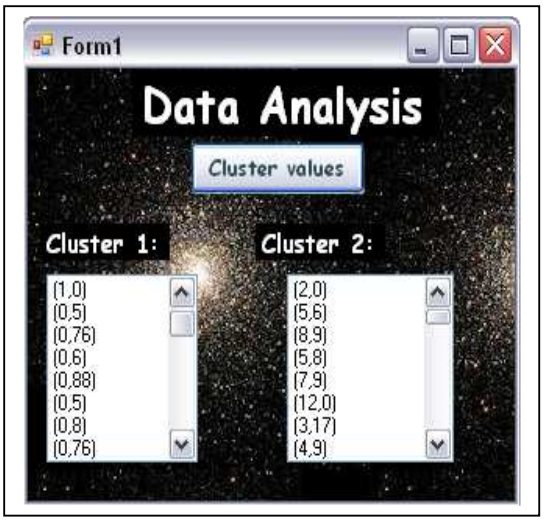

Fig5: Android Client Application showing the cluster values

ACKNOWLEDGEMENT

I would like to extend my thanks to my under-graduate students Shiji, Soubi and Karthik kumar who helped me with the implementation and deployment.

REFERENCES

[1] C. Ho Yu, S. Digangi, A. Jannasch and C. Kaprolet, "A Data Mining Approach for Identifying Predictors of Student Retention from Sophomore to Junior Year",Journal of Data Science,307-325, 2010.
[2] R.J.D. Baker, K. Yacef, " The State of educaional Data Mining in 2009: A review and future Versions", Journal of Educational Data Mining, 2009.
[3] C.Romero, S. Ventura, P.G. Espejo, c. Hervasi, "Data Mining to Classifiy Student", Proceedings of International Conference on Educational Data Mining, 2008.
[4] Hsing, K. , Cheng, Qian C.T., Leon, Z. "Web Services and Service-Oriented Application Provisioning: An Analytical Study of Application Service Strategies", IEEE Transactions on engineering Management, Vol 53, No. 4, Nov 2006
[5] Belqasmi, J. Singh, S.Y.B. Melhem, R.H. Glitho, "Soap Based vs Restful Web Service", IEEE Internet Computing, IEEE Computer Society, 2012.
[6] J.H. Levine, "Introduction to Data Analysis",Macintosh HD:DA:DAIX: Volume I:006, 1996.
[7] S. Falke, G, Stella, T.Keating, "Web Service for comparative Data Analysis of Emission Inventories",2007
[8] M.J. Kollingbaum, K. Cai, K , T.J. Norman "Web Service Support for Scientific Data Analysis", 2004
[9] M. Farber, M.Cameron, C. Ellis and Sullivan, "Massive Data Anlaytics and the Cloud", A whitepaper, Booz Allen Hamilto publications, 2006.
[10] B. Indrajit , R. Anandhi, "Healthcare Data Analytics on the Cloud", Online Journal of Health and Allied Sciences, 2011.
[11] M. Quzzani and A. Bouguettaya , "Efficient Access to Web Service", IEEE Computer Society,2004
[12] G. Zheng, A. Bouguettaya, "Service Mining on the Web", IEEE Transactions on Services Compting,2009.
[13] J.Han, M.Kamber,"Data Mining – Concepts and Techniques" ,Morgan Kaufmann Publisher, 2006.
[14] B. Suresh, "Web Service Using Apache Axis2",IBM devloper works, November 2010.
[15] "Web Service Architecture", WC3working group note 11, Feb 2004.
[16] M. Bienkowski, M. Feng, B. Means, "Enhancing Teaching and Learning through Educational Data Mining and Learning Analytics",Centre for technology in learning, SRI International, Oct 2012.
[17] J. Luan and A. M. Serban, "Data mining and its application in higher education in Knowledge Management: Building a Competitive Advantage in Higher Education", New Directions for Institutional

Research. Jossey-Bass, 2002.
[18] M. Sharma and M. Mavani, "Development of Predictive Model in Education System:Using Naïve Bayes Classifier", International Conference and Workshop on Emerging Trends in Technology,ACM, 2011.
[19] J. Shana and T.Venkatachalam, "Identifying Key Performance Indicators from Student Course Data",International Journal of Computer Applications,2011
[20] http://www. axis.apache.org/axis2/java/core.htm

