SSOAFD: A Brief Study on Service Oriented Architecture using Frequent Data Item sets

Dr. Ramalingam Sugumar
Department of Computer Science,
ChristhuRaj College,
Trichy, Tamil Nadu, Indina.

M.Appas Ali
Department of Computer Science,
Christhu Raj College,
Trichy, Tamil Naud, India,

Abstract— **The Service Oriented Architecture (SOA) is a model for building flexible, modular, and interoperable software applications. Concepts behind SOA are derived from component based software, the object oriented programming, and some other models. The SOA model allows the composition of distributed applications regardless of their implementation details, deployment location, and initial objective of their development. An important principle of service oriented architectures is, in fact, the reuse of software within different applications and processes. Service oriented architecture is essentially based on a collection of services. A service, along with its interface, must be defined in the most general way to allow utilization in different contexts and for different purposes. Once defined and deployed, services operate independently of the state of any other service defined within the system. SOA using the Frequent pattern mining techniques helpful to find interesting patterns in massive data. Prior domain knowledge leads to decide appropriate minimum support threshold. This review article show the service oriented architecture and different frequent pattern mining techniques based on apriori or FP-tree or user define techniques under different computing environments like parallel, distributed or available data mining tools, those helpful to determine interesting frequent itemsets with or without prior domain knowledge. Proposed review article helps to develop efficient service oriented architecture using frequent pattern mining techniques.**

Keywords— *Keywords- Architecute ,frequent, Itemsets, Apriori, Algorithms, techniques, pattern.*

I. INTRODUCTION

The development of information systems and computer technologies has enabled the automation of the activities in every field of the real-world this has induced a fast increase in the information available, the development of high volume data warehouses and finally, the emergence of Data Mining. The latter corresponds in a set of techniques and methods which from the data (typically stored in a data warehouse) extract usable knowledge in various fields such as environment, public health, pharmacy, biology, etc. However, the growing market draws attention to distributed Data Mining because data and software are geographically distributed over a network instead of being located in a single site. Moreover, the cost is another reason for the distribution. To optimize investment, users prefer to use components that respond to their specific needs. However, since the arrival of Web and cloud computing, distributed data is now much easier to access. Furthermore, distributed computing in heterogeneous environments has become much more feasible. At the same time, service-oriented architectures (SOA) are becoming one of the main paradigms

for distributed computing. SOA provides solutions for integrating diverse systems that support interoperability, loose coupling and reuse. To full-fill clients need one service invoke another services. It is possible that there is some evolution among these external services. Through an approach based on services, especially service-oriented architecture (SOA), integrated services can be defined to support the distributed data mining tasks in cloud and the Web. Such services can address most aspects taken into account in data mining and knowledge discovery in databases (KDD). Moreover, the most important SOA implementation is represented by web services. The popularity of Web services is mainly due to the fact that they are standardized (adoption of universally accepted technologies, such XML, SOAP, HTTP, WSDL, UDDI.). Web Services are simple, flexible and independent from both platforms and languages. Furthermore, their adoption by a number of communities, including the clooud community, indicates that the development of data mining applications based on Web services is likely be useful to an important user community. Such Web service is particularly met in business environments where time and data intensive transactions are performed between customers and offered services. [1]

2. Simple Object Access Protocol

SOAP is the protocol that is responsible for routing messages between the client and the server. It is a lightweight XML-based messaging protocol. SOAP is based on XML and thus it provides good interoperability between applications. SOAP implementations provided by vendors typically consist of two pieces: a client side Proxy that handles the SOAP message creation and result message cracking to return the result data, as well as a server piece that implements the Web Service logic. The server piece tends to be an application server that calls out to custom Web Service classes that is created on the server side and that contain the business logic of the Web Service. The server code essentially consists of simple methods to handle inputs and outputs via parameters and return values respectively. The logic in the actual method may contain any functionality. In essence it is the breaking of the business tier from the presentation tier [3].

3. Web Services in .NET

The .Net framework introduces Web Services as an integral part of the architecture, making it very easy to create and consume these services with minimal amounts of code written. In .NET framework, Web Services are featured as the new component architecture in the distributed age where not only

Internet exposure is handled through them but also common reusable business and application services. The .Net framework abstracts most of the internal logic that handles the remoting details of method calls over the wire and Visual Studio.Net builds support for Web Services directly into the development environment. Thus server side logic is made easily available to client applications. There are three major components that make up a web service.

• The Web Service on the Server side

• The client application calling the Web Service via a Web Reference

• A WSDL Web Service description that describes the functionality of the Web Service

A Web Service in .Net consists of a .asmx page that either contains a class that provides the Web Service functionality or references a specific external class that handles the logic in an external class file. Classes are standard .Net classes and the only difference is that every method that is exposed to the Web is prefixed with a [WebMethod] attribute. Once the .asmx page has been created, the Web Service is ready for accessing over the Web. .Net provides a very useful information page about the Web Service showing all the methods and parameters along with information on how to access the Web Service over the Web.

4. Frequent Itemset Mining

Frequent itemsets play an essential role in many data mining tasks that try to find interesting patterns from databases, such as association rules, correlations, sequences, episodes, classifiers, clusters and many more of which the mining of association rules is one of the most popular problems. The original motivation for searching association rules came from the need to analyse. Since their introduction in 1993 by Argawal, the frequent itemset and association rule mining problems have received a great deal of attention. Within the past decade, hundreds of research papers have been published presenting new algorithms or improvements on existing algorithms to solve these mining problems more efficiently. Frequent Itemset Mining (FIM) is a fundamental research topic in data mining. [2]

5. Client Application

Client applications can be any type of application from a Web backend aggregating data to display custom content to clients to a Fat Client application running Windows forms. The process of connecting a client application in Visual Studio.Net is always the same though you set up a Web Reference, add the Web Reference namespace and then simply call the methods of the Web Service. The method call actually calls a proxy object, which invokes the remote Web Service. The proxy base class contains all the black box magic that performs the SOAP call over the wire and the proxy class simply calls work methods in the base class. The rest of the paper is organized as, section 2 provides literature study of the existing approaches, and section 3 provides the conclusion of this study. [3].

II.   RELATED WORK

This paper first investigates the data mining applications on centralized medical databases, and how they are used for diagnostic and population health, then introduces distributed databases. The integration needs and issues of distributed medical databases are described. Finally the paper focuses on data mining studies on distributed medical databases [4].

This paper the commercial application of data mining is focused in terms of mobile computing and its management's services. And hence it's being brought into the consideration that the focusing of data mining techniques and its applications in mobile computing. Now a day's tracing the location of mobile is quit vital and important so this problem can be overcome through writing an appropriate algorithm and application development so it can be helpful to trace and capture smoothly and easily mobile computing management by applying Data management techniques and its approaches. According to the application and algorithm mobile location is traced out through mapping depending upon their classes and category and identified mobile network denoted as mobile reporting map and then mobile devices gives their current position. [5]

The paper presents a change management framework for a citizen-centric healthcare service platform. A combination between Petri nets model to handle changes and reconfigurable Petri nets model to react to these changes are introduced to fulfil healthcare 4goals. Thanks to this management framework model, consistency and correctness of a healthcare processes in the presence of frequent changes can be checked and guaranteed. [6]

A review of the literature related to the use of SOA in Industrial Automation Systems is given to set up a context for the discussion of the proposed in the above paper SOA IEC 61499 formal model. The presented, in the above paper, formal model and the execution environment architecture are discussed towards a better understanding of the potentials for the exploitation of the SOA paradigm in the industrial automation domain. SOAP and Web Services even though introduced in some PLCs have considerable performance overhead that is a big barrier in their use. The use of these technologies at the integration level of the device software constructs, greatly increases the performance overhead as well as the complexity at this level with questionable benefits regarding flexibility. Other technologies provide feasible solutions to this level of integration. [7]

This paper proposes a novel framework based on Divide-and-Conquer (D&C) for cost estimation for building SOA-based software. By dealing with separately development parts, the D&C framework can help organizations simplify and regulate SOA implementation cost estimation. Furthermore, both cost estimation modeling and software sizing work can be satisfied respectively by switching the corresponding metrics within this framework. Given the requirement of developing these metrics, this framework also defines the future research in four different directions according to the separate cost estimation sub-problems. [8]

The SOA based systems software, the successes software factors, the components of it and the rules of each of the component developers. Then defined the functional and nonfunctional requirements attributes and the importance of

them in the maintenance process. Finally present the importance of maintenance process in the SOA based systems and give some approaches in three issues in maintenance (the analysis influencing in the whole system, the understanding of services attributes, and the testing of services.) and explained each other. Finally the SOA maintenance topic still need other efforts to enhance the services maintenance process and it still a big space for researchers to support this area of research with new effective and creative approaches. [9]

In this paper through studying challenges of information systems in electronic city and with concentrating on advantages of service oriented architecture, a new architecture for integration of systems in electronic city and overcoming the challenges of information systems security to providing accurate information and efficient services to consumers. [10]

This paper is to describe the importance of SOA in telemedicine through distributed system architecture design and implementation, which is developed in .Net platform using external web services. The architecture of telemedicine system which have developed is comprises of three layers that are Presentation Layer, Business Logic Layer and Data Layer. [11]

This work is based on automobiles study and will help the sellers and customers in making decisions. The objective is to find the important selling factors that affect the relevant sale of vehicles by using the association rule mining algorithm. Most famous algorithm of association rule mining is Apriori is used for knowledge discovery. Research work will improve the existing Apriori algorithm and will reduce some of the drawbacks of the existing algorithm. [12]

An improved algorithm in this paper with a aim of minimizing the temporal and spatial complexities by cutting off the database scans to one by generating compressed data structure bit matrix(b_matrix) and by reducing redundant computations for extracting regular itemsets using top down method. theoritical analysis and experimental results shows that improved algorithm is better than classical apriori algorithm. [13]

This algorithm encountered dense data due to the large number of long patterns emerge, this algorithm's performance declined dramatically. In order to find more valuable rules, this paper proposes an improved algorithm of association rules, the classical Apriori algorithm. Finally, the improved algorithm is verified, the results show that the improved algorithm is reasonable and effective, can extract more value information. [14]

This paper indicates the limitation of the original Apriori algorithm of wasting time for scanning the whole database searching on the frequent itemsets, and presents an improvement on Apriori by reducing that wasted time depending on scanning only some transactions. The paper shows by experimental results with several groups of transactions, and with several values of minimum support that applied on the original Apriori and our implemented improved Apriori that our improved Apriori reduces the time consumed by 67.38% in comparison with the original Apriori, and makes

the Apriori algorithm more efficient and less time consuming. [15]

In this paper, implement three variations of Apriori algorithm using data structures hash tree, trie and hash table trie. Trie with hash technique on MapReduce paradigm. To emphasize and investigate the significance of these three data structures for Apriori algorithm on Hadoop cluster, which has not been given attention yet. Experiments are carried out on both real life and synthetic datasets which shows that hash table trie data structures performs far better than trie and hash tree in terms of execution time. Moreover the performance in case of hash tree becomes worst. [16]

This work proposes FDM, a new algorithm based on FP-tree and DIFFset data structures for efficiently discovering frequent patterns in data. FDM can adapt its characteristics to efficiently mine long and short patterns from both dense and sparse datasets. Several optimization techniques are also outlined to increase the efficiency of FDM. An evaluation of FDM against three frequent itemset data mining algorithms, dEclat, FP-growth, and FDM* (FDM without optimization), was performed using datasets having both long and short frequent patterns. The experimental results show signi_cant improvement in performance compared to the FP-growth, dEclat, and FDM* algorithms. [17]

In this paper Improved Apriori algorithm which will help in reducing multiple scans over the database by cutting down unwanted transaction records as well as redundant generation of sub-items while pruning the candidate item sets. The performance of this algorithm is analyzed against the FP Growth algorithm in which there is no generation of candidate set. [18]

The algorithm decreases pruning operations of candidate 2-itemsets, thereby saving time and increasing efficiency. For the bottleneck: poor efficiency of counting support, proposed algorithm optimizes subset operation, through the transaction tag to speed up support calculations. Algorithm Apriori is one of the oldest and most versatile algorithms of Frequent Pattern Mining (FPM). Its advantages and its moderate traverse of the search space pay off when mining very large databases. The algorithm improves Apriori algorithm by the way of a decrease of pruning operations, which generates the candidate 2-itemsets by the apriori-gen operation. Besides, it adopts the tag-counting method to calculate support quickly. So the bottleneck is overcome. [19]

This paper presents a load balancing technique designed specifically for parallel publications applications running on multicore applications. This architecture provides a hardware parallelism through cores inside the CPU. It increased performance low cost as compare to single core machines attracts HPC high performance computing connectivity. [20].

A distributed association rule mining algorithm on Spark named as Adaptive-Miner which uses adaptive approach for finding frequent patterns with higher accuracy and efficiency. Adaptive-Miner uses an adaptive strategy based on the partial processing of datasets. Adaptive-Miner makes execution plans

before every iteration and goes with the best suitable plan to minimize time and space complexity. Adpative-Miner is a dynamic association rule mining algorithm which change its approach based on the nature of dataset. Therefore, it is different and better than state-of-the-art static association rule mining algorithms and conduct in-depth experiments to gain insight into the effectiveness, efficiency, and scalability of the Adaptive-Miner algorithm on Spark. [21]

III. CONCLUSION

Data mining services in SOA are key elements for practitioners who need to develop knowledge discovery applications that use large and remotely dispersed datasets and computers to get results in reasonable times and improve their competitiveness. In this study, we address the definition and composition of services for implementing knowledge discovery applications on SOA model using frequent item set mining. In this survey, we presented an in depth analysis of a many approaches of SOA and frequent itemset mining.

REFERENCES

[1] Mrs. C. Beulah Christalin Latha , Dr. (Mrs.) Sujni Paul , Dr.E.Kirubakaran , Mr. Sathianarayanan "A Service Oriented Architecture for Weather Forecasting Using Data Mining", Int. J. of Advanced Networking and Applications,Volume: 02, Issue:02, Pages:608-613, 2010.

[2] Jayant Kayastha,N. R. Wankhade, "A Survey Paper on Frequent Itemset Mining Techniques", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 12, December 2016.

[3] Mohammed Sabri, Sidi Ahmed Rahal, "APESS - A Service-Oriented Data Mining Platform: Application for Medical Sciences", I.J. Information Technology and Computer Science, 2016.

[4] Yasemin Atilgan and Firat Dogan, "Data Mining on Distributed Medical Databases: Recent Trends and Future Directions", IT Revolutions 2008, LNICST 11, pp. 216–224, 2009. © ICST Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering 2009

[5] Sayyada Sara Banu, Mohammed Waseem Ashfaque, Dr.Perumal Uma, Quadri S.S Ali Ahmed, "Data mining techniques on Mobile computing Management and Service Oriented Architecture of web Services" , International Journal of Electrical, Electronics and Computer Systems (IJEECS), ISSN (Online): 2347-2820, Volume -3, Issue-2 2015

[6] Sabri MTIBAA, Moncef TAGINA, "Managing Changes in Citizen-Centric Healthcare Service Platform using High Level Petri Net" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. XXX, No. XXX, 2011

[7] Kleanthis Thramboulidis, Member, IEEE , "Service-Oriented Architecture in Industrial Automation Systems - The case of IEC 61499: A Review

[8] Zheng Li, Jacky Keung, "Software Cost Estimation Framework for Service-Oriented Architecture Systems using Divide-and-Conquer Approach"

[9] Hamza Naji1 Mohammad Mikki,"A Survey of Service Oriented Architecture Systems Maintenance Approaches "International Journal of Computer Science & Information Technology (IJCSIT) Vol 8, No 3, June 2016.

[10] Dr. Gurpreet Singh, Er. Sonia Jassi , "Implementation and evaluation of optimal algorithms for computing association rule learning", International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 6 Issue 7, Page No. 22128-22133 July 2017.

[11] Shalini Dutt, Naveen Choudhary & Dharm Singh, "An Improved Apriori Algorithm based on Matrix Data Structure", Global Journal of Computer Science and Technology: C Software & Data Engineering Volume 14 Issue 5 Version 1.0 Year 2014.

[12] Jiao Yabing , "Research of an Improved Apriori Algorithm in Data Mining Association Rules", International Journal of Computer and Communication Engineering, Vol. 2, No. 1, January 2013.

[13] Mohammed Al-Maolegi1, Bassam Arkok , "An Improved Apriori Algorithm For Association Rules" International Journal on Natural Language Computing (IJNLC) Vol. 3, No.1, February 2014.

[14] Sudhakar Singh ,Rakhi Garg, P.K. Mishra,"Performance Analysis of Apriori Algorithm with Different Data Structures on Hadoop Cluster", International Journal of Computer Applications (0975 – 8887) Volume 128 – No.9, October 2015.

[15] George GATUHA, Tao JIANG, "Smart frequent itemsets mining algorithm based on FP-tree and DIFFset data structures", Turkish Journal of Electrical Engineering & Computer Sciences, 2017

[16] Sangita Chaudhari, Mayur Borkhatariya, Apurva Churi, Mohini Bhonsle, " Implementation and Analysis of Improved Apriori Algorithm", International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 5, Issue 2, March 2016

[17] Darshan M. Tank," Improved Apriori Algorithm for Mining Association Rules", I.J. Information Technology and Computer Science, 2014.

[18] Prantik Pancholi, Shital Khairnar, Jyoti kamble , Amol Jadhao, " MACH: Performance Enhancement in Multi-core Processor using Apriori Algorithm with file Chunking", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056, Volume: 03 Issue: 04 | April-2016.

[19] Sanjay Rathee, Arti Kashyap, "Adaptive-Miner: an efficient distributed association rule mining algorithm on Spark", J Big Data 5:6 https://doi.org/10.1186/s40537-018-0112-0, 2018.

[20] Sajjad Hashemi, Seyyed Yasser Hashemi, "A Novel Service Oriented Architecture For Integration of Information Systems In Electronic City", International Journal Of Scientific & Technology Research Volume 1, Issue 11, December 2012.

[21] Asadullah Shaikh, Muniba Memon , Nasrullah Memon , Muhammad Misbahuddin "The Role of Service Oriented Architecture in Telemedicine Healthcare System", International Conference on Complex, Intelligent and Software Intensive Systems.