International Journal of Engineering Sciences Paradigms and Researches [Volume 47, Issue: Special), March 2018]
www.ijesonline.com ISSN (Online): 2319-6564

An Evaluative Analysis of Data Mining to Protect Web Services

Ratikanta Sahoo
Raajdhani Engineering College, Bhubaneswar
ratikanta sahoo@rec.ac.in

Abstract

Web services and data mining are concepts that are explained in this paper. The techniques for protecting the online services are described here. This article discusses the many web services-related concerns. This paper explains the various data mining techniques used in online services. A web service is a way for two electronic devices to communicate with each other. Data mining techniques are used to improve data utilization and make web services secure.

Keywords: *Data Mining, Web Services, Web Services Technologies, Web Services Usage.*

Introduction

A Web service is a method of communication between two electronic devices over the World Wide Web. A Web service is a software function provided at a network address over the web or the cloud, it is a service that is "always on" as in the concept of utility computing.

The W3C defines a "Web service" as "a software system designed to support interoperable machine-to-machine interaction over a network". It has an interface described in a machine-process able format (specifically Web Services Description Language, known by the acronym WSDL). Other systems interact with the Web service in a manner prescribed by its description using SOAP messages, typically conveyed using HTTP with an XML serialization in conjunction with other Web-related standards." The W3C also states, "We can identify two major classes of Web services, REST-compliant Web services, in which the primary purpose of the service is to manipulate XML representations of Web resources using a uniform set of "stateless" operations; and arbitrary Web services, in which the service may expose an arbitrary set of operations.

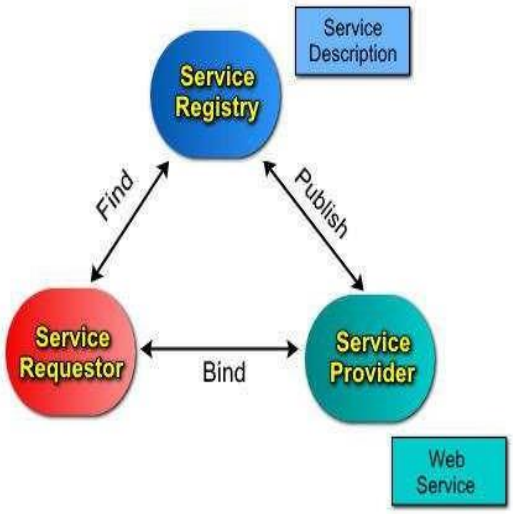

Figure 1: Web Services

Data Mining Techniques:

Association Mining

Web server access logs record all interactions between Web services and users. With the huge volume of Web service access, the amount of logs that can be collected makes a feasible source for identifying Web services with similar usage patterns.

The tasks are: (1) selection from the Web server log all entries related to the offered Web services, (2) extraction of a set of a client's interaction with a Web service, (3) calculation of client sessions with the Web service, and (4) application of an association mining algorithm to find Web services with similar usage patterns.

Clustering Mining

While previous approaches capture the intra-query relationships by clustering queries on a per query basis (Beeferman & Berger, 2000; Wen et al 2002), they omit the inter-query relationships that exist between queries submitted by a user in one search session. A better option is to group the similar search sessions and provide
suggestions of search terms from search sessions that belong to the same cluster.

The first task is to consolidate the data from the user query and Web server logs. This is done by matching the query recorded in the query log with the subsequent service descriptions viewed by the user recorded in the Web server log. The next task is to form search sessions to arrange a set of queries in sequence by a user to locate a particular service. Search session similarity now can be calculated based on the similarity of the set of search terms used and the set of service descriptions viewed between two search sessions. The Jaccard coefficient (Han & Kamber, 2001) can be used to calculate the similarity of the search terms and service descriptions sets.

Predictive Mining

Service providers have information such as the line of business, size of business, and what services their clients use. These can be used as inputs for predictive modeling operations and recommendations can be made to new clients. Inputs such as the interfaces, functionality, and security offered by the service, as well as the cost, and other resources required by the service can also be considered in analysis. Classification techniques such as decision trees can be used to build rules on service subscriptions. Since the only information service providers have about clients are those for billing purposes, the number of attributes available is small.
Consequently, the structure of the resulting decision tree will be relatively simple and easily comprehensible to a human analyst. To further enhance the success rate of recommendations, service providers can find dissociations among the services they offer. Dissociations capture negative relationships between services with rules such as $X \Rightarrow Z$; $X \wedge Y$ that is, the use of services X and Y implies that it is unlikely service Z will also be used, even though X and Z are often used (Teng, 2002). By incorporating these dissociations in the recommendation process, more specific recommendations can be made.

Methodology

Automated tools can aid in the creation of a Web service. For services using WSDL it is possible to either automatically generate WSDL for existing classes (a bottom-up strategy) or to generate a class skeleton given existing WSDL (a top-down strategy). A developer using a bottom up method writes implementing classes first (in some programming language), and then uses a WSDL generating tool to expose methods from these classes as a Web service.[6] This is often the simpler approach.
A developer using a top down method writes the WSDL document first and then uses a code generating tool to produce the class skeleton, to be completed as necessary. This way is generally considered more difficult but can produce cleaner designs

Use of Data Mining Technologies to Make Web Services Secure

Data mining techniques are increasingly in gathering data from various sources that can be used for web services. Additionally, data mining tools are now required to access a variety of standards and platforms. The solutions for interoperability by using XML and SOAP as means of Web services communication can assist data mining by standardizing importing data and information to XML format. Web services can offer assistance to data mining in integration of data coming from various sources.
The SOAP protocol enables data interaction on the Web and therefore makes the collection of data possible (Nayak & Seow, 2004). Some efforts have been made to implement these protocols, but in fact the full potential of these technologies has not yet been reached. An example is *Web services for DB2 Intelligent Miner* (http://www.alphaworks.ibm.com/tech/ws4im), "a collection of Web services that allow clients to describe and perform basic mining tasks using XML, XML Schema, and XPath on top of DB2 Intelligent Miner." Another example of a SOAP-based data mining solution is *XML for Analysis* (http://www.xmla.org/tdwievent.asp), an openindustry-standard Web service interface for online analytical processing and data mining functions. It provides "a set of XML Message Interfaces that use SOAP to define the data access interactions between a client application and an analytical data provider." The interfaces are aimed at keeping the client

programming independent from the mechanics of data transport, but at the same time providing adequate information concerning the data and ensuring that it is properly handled. This data mining solution is platform, programming language, and data source independent.

There are many data mining applications that use Web services technologies to implement them efficiently. An example is online banking, which has recently grown substantially as a Web service. As the online transactions increase, so does the possibility of fraud, in particular credit card fraud. Chiu and Tsai (2004) proposed a Web services-based collaborative scheme for participating banks to share their knowledge about fraud patterns. The participating banks share their individual fraud transactions and new transactions via a central location. The data exchange in this heterogeneous and distributed environment is secured with WSDL, XML, and SOAP. The frequent pattern mining is then applied on this integrated data for extracting more valuable fraud patterns to improve the fraud detection.

Other research introduces a dynamic data mining process system based on Web services to provide a dynamical and satisfied analysis result to the enterprise (Chiu & Tsai, 2005). Each data mining process (data pre-processing, data mining algorithms and visualization analysis) is viewed as a Web service operated on the Internet. The Web service for each activity provides its functionality. Depending on the user's requirement, the Web services are dynamically linked using the Business Process Execution Language for Web Service (BPEL4WS) to construct a desired data mining process. Finally, the result model described by the Predictive Model Markup Language (PMML) is returned for further analysis. PMML is an XML markup language defined for data mining functions and models (Wettschereck, & Muller, 2001) to make easy and efficient data models and result interpretation.

There are many application-oriented research studies as well. Zheng and Bouguettaya (2005) model the biological entities and the dynamic processes as Web services, and then propose a Web service mining approach to automatically discover the unexpected and potentially interesting pathways.

Use of Data Mining in Improving Web Services Usage

The previous section discusses a number of possible applications of data mining to assist the Web services. In this section we outline some existing works. The majority of work is in the direction of addressing the shortcoming of UDDI by finding relationships between search terms and service descriptions in UDDI.

Sajjanhar, Jingyu, and Yanchun (2004) have applied the regression function called singular value decomposition (SVD) to discover semantic relationships on services for matching best services. Their preliminary results show a significant increase in correct matching between service descriptions and the search terms after application of their algorithm with IBM UDDI. The matched results are not merely based on the number of matched keywords within the service descriptions. The algorithm evaluates the keyword global weights within the SVD procedure and aggregates services containing the highest global weight words to find semantic matched services.

Wang and Stroulia (2003) developed a method for assigning a value of similarity to WSDL documents. They use vector-space and WordNet to analyze the semantic of the identifiers of the WSDL documents in order to compare the structures of their operations, messages, and types, and to determine the similarity among two WSDL documents. This helps to support an automatic process to localize Web services by distinguishing among the services that can potentially be used and that are irrelevant to a given situation. Dong, Halevy, Madhavan, Nemes, and Zhang (2004) build a Web service search engine to support the similarity search for Web services along with keyword searching with utilizing clustering and association mining. Starting with a keyword search, a user can drill down to a particular Web service operation.

However, when unsatisfied, instead of modifying the keywords, the user can query for Web service operations according to the most similar and semantically associated keywords suggested by the engine using the data mining techniques. Gombots et al. (2005) attempt to apply data mining to Web services and their interactions in order to analyze interactions between Web service consumers and providers. They toss a new term, "WSIM Web services interaction mining," to analyze the log data to acquire additional knowledge about a system. They identify three levels of abstraction with respect to WSIM: the operation level, the interaction level, and the workflow level. On the Web service operation level, only one single Web service and its internal behavior is examined by analyzing a given log output of the Web service. On the Web services interaction level, one Web service and its "direct neighbors" Web services (that the examined WS interacts with) are examined. This analysis reveals interesting facts about a Web service's interaction partners, such as critical dependencies. On the highest level of abstraction the Web service workflow level the large-scale interactions and collaborations of Web services which together form an entire workflow are examined. This details the execution of the entire process: what is the general sequence of execution of various operations?

Malek et al. (2004) apply data mining in security intrusions detection while the Web services are in use. They show the impact of mining in detecting security attacks that could cripple Web services or compromise confidential information. They determine the relevance of different log records and define the attack signature with the use of sequential pattern mining with logs. Then they discover the highly compact decision rules from the intrusion patterns for pattern searching that help to describe some safeguard against the attacks.

Conclusion

The data mining tasks that find applications in Web services mining include value prediction, similar time sequence analysis, deviation analysis, classification, clustering, and association mining. These applications range from delivering business value that can be used by management for strategic decision making, to providing technical benefits that target specialist end users. Further testing is required to identify the real value of the applications. Additionally, because some applications such as search term suggestion require real-time responses, techniques for providing results efficiently need to be developed. These may include new algorithms for scheduling the processing of requests and delivery of responses to multiple users so the information is returned as at close to real time as possible.

References

[1] Benatallah, B., Hacid, M., Leger, A., Rey, C., & Toumani, F. (2004). On automating Web service discovery. VLDB Journal. Berlin: Springer-Verlag.

[2] Berardi, D., Grüninger, M., Hull, R., & McIlraith, S. (2004). Towards a first-order ontology for Semantic Web services. In Proceedings of the W3C Workshop on Constraints and Capabilities for Web Services.

[3] Berners-Lee, T., Hendler, J., & Lassila, O. (2001, May). The Semantic Web. Scientific American.

[4] Bernstein, A., & Klein, M. (2002). Discovering services: Towards high precision service retrieval. In Proceedings of the CaiSE workshop on Web Services, E-Business, and the Semantic Web (WES): Foundations, Models, Architecture, Engineering and Applications, Canada.

[5] Chiu, C.-C., & Tsai, C.-Y. (2004). A Web services-based collaborative scheme for credit card fraud detection. In Proceedings of the IEEE International Conference on E-Technology, E-Commerce, and E-Services (EEE 2004) (pp. 177-181).

[6] Chiu, C.-C., & Tsai, M.-H. (2005). A dynamic Web service-based data mining process system. In Proceedings of the 5th International Conference on Computer and Information Technology (CIT'05) (pp. 1033-1039).

[7] Clark, D. (2002). Next-generation Web services. IEEE Internet Computing, 6(2), 12-14. Cooley, R., Mobasher, B., & Srivastava, J. (1997). Web mining: Information and pattern discovery on the World Wide Web. In Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97), Newport Beach, CA.

[8] Curbera, F., Duftler, M., Khalaf, R., Nagy, W., Mukhi, N., & Weerawarana, S. (2002). Unraveling the Web services web: An introduction to SOAP, WSDL, and UDDI. IEEE Internet Computing, 6(2), 86-93

[9] Devore, J.L. (1995). Probability and statistics for engineering and the sciences (4th ed.). New York: Duxbury Press.

[10] Dong, X., Halevy, J., Madhavan, E., Nemes, & Zhang. (2004). Similarity search for Web services. VLDB

[11] Fayyad, U.M., Piatesky-Shapiro, G., Smyth, P., & Uthurusamy, R. (1996). Advances in knowledge discovery and data mining. London: AAAI/MIT Press.

[12] Gombotz, R., & Dustdar, S. (2005, September). On Web services workflow mining. In Proceedings of the BPI Workshop (LNCS), co-located at BPM, Nancy, France. Berlin: Springer-Verlag.

[13] Han, J., & Kamber, M. (2001). Data mining: Concepts and techniques. San Francisco: Morgan Kaufmann.

[14] Hoschek, W. (2002). Web service discovery processing steps. In Proceedings of the IADIS International Conference WWW/Internet 2002, Lisbon, Portugal.

[15] Johnson, J.T. (2003). State of the Web services world. Retrieved April 17, 2003, from http://www.computerworld.com.au/index.php?secid=1398720840&id=622609517&eid=-180