

A Proper Web Service Selection Method Based on Machine Learning Algorithms and Semantic Web

¹Muaz Gultekin, ²Ugur Ergul
¹Yalova University
²Yildiz Technical University

Abstract

Web Service is parts of program which serves over the web with HTTP protocol. Web Services use the protocol which is called SOAP (Simple object Access Protocol) produces XML output against invocations by made over HTTP. Since data transfer is done through these XML outputs and the platform has an independent structure. Today, web services have become indispensable elements of web technologies. The number of web services is increasing day by day and UDDI registers have millions of web services. But there is important problem still has an importance that users find web services which they request in UDDI registers. Seek out methods of desired web services are key-word based. Therefore, it does not give the desired result. In this paper, by using semantic architecture, we provide a framework to find the desired web services more easily. Also, we showed more meaningful results to users by separating into various classes of web services with our developed framework. Thanks to this framework QoS (quality of service) of each web service are evaluated and processed via ontological web language. Additionally the developed framework processed QoS parameters using machine learning algorithms such as Fuzzy C-Means and Gustafson-Kessel algorithms to serve desired result expressively to users. Thanks to this framework requested web services are found with % 99.2 accuracy rates.

Keywords: web service, web service discovery, semantic web service, machine learning

1. Introduction

Web service is a software component which uses SOAP through web with XML. In fact, SOAP is basic protocol based on XML which allows information exchange between applications by using open protocol standard [3]. At the same time, SOAP is a communication protocol. Web services are based on dispensed technologies and provide interoperability and standard way between several web services within institutional boundaries with the use of XML [1, 2].

Basic web service platform is compound of XML and HTTP. HTTP is the most common current internet protocol. XML provides usable language between different platforms and different programming languages. Web services uses XML and SOAP for coding and decoding data. Web services are web applications interfaces are exposed over protocols like HTML and XML. Web services are described in XML format by WSDL. WSDL is a language which provides XML format and model for identify syntax relative to web services. WSDL acts as a word relevant to UDDI [4].

The Web service description masks the details of implementation, but also makes sufficient information required for the service mutual effect. Figure 1 specifies the web service communication paradigm.

SOAP, UDDI and WSDL are used in different stages, detecting and connecting, in the Web

Services development cycle [6]. The Web Service Oriented Architecture (SOA) is specified in Figure 1, the model starts with the publishing operation, when a business makes a decision to present a Web Service.

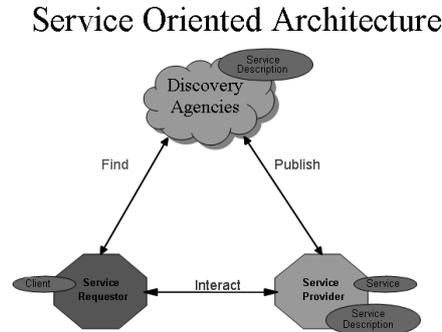


Figure 1. Services Oriented Architecture

2. Suggested Solution

Normal web service search engines are text based and most of the time they cannot provide searched item correctly also there is still a big controversy whether it's safe or not. In order to solve these problems various technics have been used and some of them have been successful. In this study, we primarily determined the Non-Functional quality parameters (QoS) of web services [7, 8]. By looking these quality parameters, we had some information about security and quality of web service. At the same time, some values of these parameters were very close. When we apply PCA technic some parameters have no effect on the result so we cast out these quality parameters. Features of web services restored in a file called WSDL. But in WSDL file, only functional parameters can be restored. Because we could not restore our quality parameters, we used OWL-S file method which was developed for web services [8, 9]. By doing that, we linked the web services to their own quality parameters. Last phase was the process of presenting the results effectively during the web service searches. In order to make results more effective, web services is separated to 5 different classes by using various classification algorithms.

2.1. Proposed Model Steps

In this study we primarily created a date set that includes non-functional properties of web services. We applied classifying algorithms to the date set. We listed generated results comparatively.

Because of we don't have labels of data's we primarily tagged examples of our data set applying K-Means, Fuzzy C-Means and Gustafon Kessel algorithms. Later we qualified our labeled data's with Meta learning ways and result listed in Table 2, Table 3 and Table 4.

2.1.1. Applied learning Algorithm

2.1.1.1. K-Means Algorithm

K-means algorithm, one of the oldest clustering method is developed by JB MacQueen in 1967. (MacQueen, 1967). One of the most widely used unsupervised learning method of the K-means assignment mechanism allows belong to only one cluster of each data. Therefore, K-means algorithm is a sharp clustering algorithm. K-means is a method that represents based on the center point of the cluster [10]. Tends to take global clusters of equal size. K-means clustering method for the evaluation of the most widely used quadratic error criterion SSE. Clustering gives the best result which has the lowest SSE value.

The sum of squares of distance to the center points of cluster of objects to which they belong is calculated by equation [11]. The result of this criteria, k cluster units separate from each other to result in intense and targeted as possible is attempted. Algorithm, tries to identify k parts which will reduce the mean squared error function.

K-means algorithm, splits to k units cluster that k parameter given by the user to the algorithm and the data set consists of n units' data. Cluster similarity is measured by the mean value of the objects in the cluster, this is the centroid of the cluster [11].

2.1.1.2. Fuzzy C-Means Algorithm

Fuzzy c-means (FCM) algorithm is the most well-known of fuzzy clustering technique division and widely used method. Fuzzy c-means algorithm was introduced by Dunn in 1973 and was developed by Bezdek in 1981 (Höppner vd. 2000). Fuzzy c-means algorithm is a method based on the objective function. Fuzzy C-means method allows the object may belong to two or more clusters. In according to the principles of fuzzy logic each data, cluster to each of the ranging belongs to the membership value. Membership value to all classes of data should be that a total of "1". Objects which do not belong to cluster near the center of the cluster membership that do not belong to the other cluster will be greater than membership. Progress of objective function value close to the minimum required by the approach of the clustering process is completed. The algorithm has the following generalization of the least squares method to minimize the objective function by shift work (Höppner vd. 2000).

2.1.1.3. Gustafson Kessel Algorithm

The most common one fuzzy clustering methods Bezdek and Dunn's (1975) function called c-means based on the Fuzzy C-Means made a minimum of (FCM) algorithm. FCM algorithm is fuzzy state of k-means algorithm. FCM algorithm only finds phrases have the same shape and size. This is because often distance norm is Euclidean norm (circular blocks) and this norm is that can't adapt. If Gustafson-Kessel algorithm is an extended version of the standard FCM algorithm [12]. This algorithm has been developed for the detection of the different geometric figures in a data set and uses the Mahalanobis distance as the distance norm [13, 14].

Gustafson-Kessel algorithm is based on the type of c-means of the objective function to be optimized recursively [14].

2.1.2. Creating Dataset and Model

In this study we have defined 9 QoS parameters: Throughput, Response Time, Latency, Documentation, Availability, reliability, Successability, Compliance, Best Practices, Integrity, and Reputation.

Response Time: This an essential parameter of web service that sum of execution time and waiting. Execution time can be defined as time of service performing its functionality [15]. Waiting time considered as time the other activities that not include performing functionality. For instance message exchange between service requester and service provider. From the definition there is an uncertainty with this definition because of the network fluctuations. Instead of above definition for waiting time we can use the time elapsed to execution time can be considered as waiting time.

Reliability: can be thought as service provider's ability to successfully deliver requested service functionality. This ability can be calculated by the probability of the success in service execution [8, 7].is it can be formulated as ratio execution time and meantime between failures.

Availability: of web service means that the degree of service is operational and accessible when it is required for use. Formulated as proportion of service up time to service down time.

Latency: Time taken between service requests arrives and request is being serviced.

Throughput: the number of completed service requests over a time period.

Successability: can be thought as percentage of successful invocation.

Compliance: can be thought as coherency of services.

Documentation: can be thought as reputation of service.

Table 1. A Part of Dataset

Service Name	Response Time	Availability	Throughput	Successability	Reliability
AnalysisWSappLab	243.91	85	7.1	86	90.47
ReasonerWS	174	83	14.9	84	89.21
RRNAScanService	105.5	56	9.4	56	66.74
Codic_x0020_EngineWS	103.25	97	1.2	99	99.21
WSDLInteropTestR	626.75	66	4.3	67	59.41
XigniteInsider	186.56	96	8.8	99	87.56
QueryService	300.76	91	8.2	92	94.00
SRS	356.75	56	7.5	56	64.04

After decided QoS parameters we have evaluated these values for 2507 web service. A part of the data set listed in Table 1. After creating dataset we separated the data set into clusters

using unsupervised clustering algorithms such as K-means, Fuzzy C-means, Gustafson-Kessel methods which is explained in detail.

After clustering dataset we have applied Meta learning techniques. We listed generated results comparatively Table 2, Table 3 and Table 4.

We have created 10 ensemble trees of bagging, random forest and random sub space. After that we have applied K-Means, Fuzzy C-Means and Gustafson Kessel methods for clustering and the result is listed in Table 2. We have applied the same operations for 50 and 100 trees in which results are listed Table 3. and Table 4.

Table 2. 10 Trees Ensemble

	K-Means	Fuzzy C- Means	Gustafson Kessel
Bagging	95.01	99.36	90.47
Random Sub Space	93.09	96.00	89.21
Random Forest	95.92	97.84	92.29

Table 3. 50 Trees Ensemble

	K-Means	Fuzzy C- Means	Gustafson Kessel
Bagging	95.84	99.33	90.49
Random Sub Space	93.43	98.17	90.87
Random Forest	97.12	98.50	93.04

Table 4. 100 Trees Ensemble

	K-Means	Fuzzy C- Means	Gustafson Kessel
Bagging	96.01	99.32	90.59
Random Sub Space	95.76	99.30	91.70
Random Forest	97.19	98.64	92.84

3. Conclusion

Efficiently finding desired web service among huge web service stack is still an important issue [3]. In this study a Novel Model is developed to find desired web service. At the beginning 9 QoS parameters are evaluated. These parameters are non-functional that's why owl-s file used to kept these values. But this process still not give desired and meaningful information to users. Some machine learning algorithms (K-Means, Fuzzy C-Means and Gustafson Kessel) applied to the service stack to label each web service. After that we use Meta learning techniques to classify web services. Results listed in Table 2, Table 3 and Table 4. According to evaluated values the best techniques is creating 50 trees with bagging and using Fuzzy C-Means unsupervised learning algorithms.

References

- [1] Bray, T., Paoli, J., Sperberg-McQueen, C. M., Maler, E., & Yergeau, F., 1998. Extensible markup language (XML). World Wide Web Consortium Recommendation REC-xml-19980210. <http://www.w3.org/TR/1998/REC-xml-19980210>.
- [2] Mitra, N., & Lafon, Y. Soap version 1.2 part 0: Primer. W3C recommendation, 24, 12, 2003.
- [3] UDDI Browser. <http://www.uddibrowser.org>. visit date: 02.12.2013
- [4] Menasce, D. (2002). QoS issues in web services. *Internet Computing, IEEE*, 6(6), 72-75.
- [5] WSDL file: <http://www.w3.org/TR/wsdl>, visit date: 15.12.2013.
- [6] Liu, Y., Ngu, A. H., & Zeng, L. Z. (2004, May). QoS computation and policing in dynamic web service selection. In *Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters* (pp. 66-73). ACM.
- [7] Mani A, Nagarajan A. Understanding quality of service for Web services. http://www.ibm.com/developerworks/library/wsquality.html?S_TACT=105AGX52&S_CMP=cn-a-ws. Visit date: 01.12.2013.
- [8] M. Paolucci et al. "Semantic Matching of Web Services Capabilities", *The Semantic Web- ISWC 2003: 1st Int'l Semantic Web Conf.*, LNCS 2342, Springer-Verlag, 2002, p.333.
- [9] Gultekin, M. Akbas, A. "QoS Based Semantic Web Service Selection Using Machine Learning Algorithm", ICECCO, October 2012.
- [10] T. Kanungo, D.M. Mount, N.S. Netanyahu, C. Piatko, R. Silverman, and A.Y. Wu, *The Analysis of a Simple k-means Clustering Algorithm*, Technical Report CAR-TR-937, Center for Automation Research, Univ. of Maryland, College Park, Jan. 2000.
- [11] Bradley, P., & Fayyad, U. Refining initial points for k-means clustering. *Proc. 15th International Conference on Machine Learning*, 1998
- [12] C. Cortes, V. Vapnik, *Introduction to Machine Learning*, pp. 273-297, 1995.
- [13] Tom M. Mitchell, *Machine Learning*, Singapore, McGraw-Hill, 1997.
- [14] Quinlan, J.R. *Decision trees*. *Machine Learning*, 1986.
- [15] Yang S W, Shi M L. A model for Web service discovery with QoS Constraints. *Chinese Journal of Computers*, pp. 589-594, 2005.