A Survey on Aware Web Service Recommendation

Ashwini V Rajguru[1] Kalpana T Kanade[2]
[1,2]Everest College of Engineering, Aurangabad

Abstract— To Find Correct web service we need to have a QoS-based approach for an effective web service recommendation. Web Services are one kind of software component which is designed to support machine to machine interaction over a web. With increasing use and adoption of Web services on the World Wide Web, Quality-of-Service (QoS) is becoming important for describing nonfunctional characteristics of Web services. In this paper, we present a collaborative filtering approach for prediction of QoS values of Web services and making Web service recommendation [1] by taking advantages of past usage experiences of service users. We first propose a user-collaborative mechanism for past Web service QoS information collection from different service users. Then, based on the collected QoS data, a collaborative filtering approach is designed to predict Web service QoS values.

Key words: Qos, Service Recommendation, Collaborative Filtering, Web Services

I. Introduction

WEB services are software components designed to support interoperable machine-to-machine interaction over a network, usually the Internet. Web service employs WSDL (Web Service Description Language) for interface description and SOAP (Simple Object Access Protocol) for exchanging structured information. Web services have been widely employed by both enterprises and individual developers for building service oriented applications. When developing service oriented applications, developers first design the business process according to requirements, and then try to find and reuse existing services to build the process. Collaborative filtering is valuable in e-commerce, and for direct recommendations for music, movies, news etc. Collaborative filtering aims at learning predictive models of user preferences, interests or behavior from community data, i.e. a database of available user preferences. We describe a new model-based algorithm designed for this task, which is based on a generalization of probabilistic latent semantic analysis to continuous-valued response variables.

When developing service-oriented applications, developers first design the business process according to requirements, and then try to find and reuse existing services to build the process. Currently, many developers search services through public sites like Google Developers (developers.google.com), Yahoo! Pipes (pipes.yahoo. com), programmable Web (programmableweb.com), etc. However, none of them provide location-based QoS information for users. Such information is quite important for software deployment especially when trade compliance is concerned. Some Web services are only available in EU, thus software employing these services cannot be shipped to other countries.

We propose a novel collaborative filtering-based Web service recommender system to help users select services with optimal Quality-of-Service (QoS) performance. Our recommender system employs the location information and QoS values to cluster users and services, and makes personalized service recommendation for users based on the clustering results. Compared with existing service recommendation methods, our approach achieves considerable improvement on the recommendation accuracy.

II. Literature survey

Web services are software components designed to support interoperable machine to machine interaction over a network. The adoption of web services as a delivery mode in business has fostered a new paradigm from the development of monolithic applications to the dynamic setup of business process. In recent years, web services have attracted wide attentions from both industry and academia, and the number of public web services is steadily increasing.

“Web services and e-services have been announced as the next wave of Internet-based business applications that will dramatically change the use of the Internet”. This is an eloquent sentence stated by Casati et al.[2][3][4][5][6] that may not be far from the real expectations of web services. Before the existence of this technology, the use of services from the World Wide Web was through a browser and a Web server using the HTTP protocol. Nowadays, this new model of interaction in the web given by web services has involved a primary subject of study which has had several contributions in the last few years. Several organizations have been putting a big effort for standardization of this technology. The current results of these efforts are the web services as we understand nowadays:

The main actors of this new model are web service providers and web service consumers. Web Service providers are the ones who develop the service and publish the Web Service Description Language (WSDL). A WSDL is an XML interface used to invoke the service in a programming-language independent manner

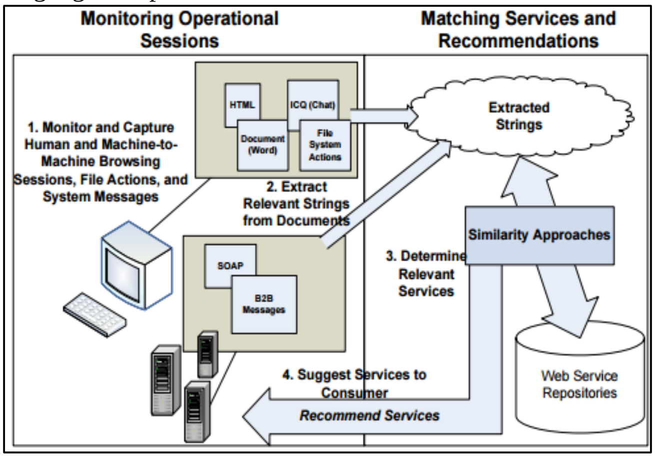

Fig. 1: Web Service Recommendation Scenario

QoS is defined as a set of user-perceived properties including response time, availability, reputation, etc. Currently, it’s not practical for users to acquire QoS information by evaluating all the service candidates, since conducting real-world web service invocations is time consuming and resource-consuming.

However, the flexibility and extensibility of web services has involved a new interaction model and development approach for the construction of software. As we can see in Figure 1, the client (which might be software itself) sends a request to a web service. This web service is in turn the client of other web services. Interaction between them is done through the SOAP protocol. In this new scenario, each web service is responsible of doing a small task, so following the SOA principles, a web service can be developed by combining a set of small pieces of software.

III. MATHEMATICAL MODEL

A. Input and Outputs of the System

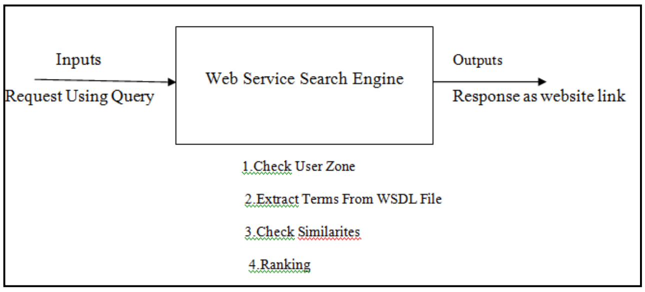

Fig. 2: Input and Outputs of the System

P ={U,Q,T,A}
U=set of users
Q = query
T= time limit
Constraints:
T and single user query.

1) Algorithms

A1: Single user gives his query to system. Mapping of U1 to U and Q1 to Q.
A2: Check user zone.
A3: Extract terms from WSDL file.
A4: Check similarity and Ranking.
A5: Total process time P time<=T.

B. Motivational Scenario

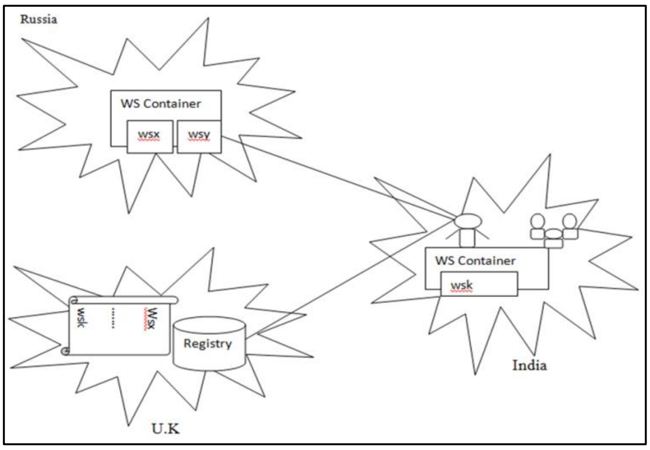

Fig. 3: Alice's Situational Problem

In this section, we present a webservice searching scenario to show the research problem of this paper. As figure2 depicts, Alice is a software engineer working in India. She needs a mobile number validation service to filter out numbers. So she does process of searching on the service registry located in the UK, and she gets a list of recommended services of the service average response time, in the ascending order. After that, Alice tries the first two services which are provided by the Russian company and finds that the response time is much higher than her expectation. She then realizes that the service ranking is based on the evaluation conducted by the registry in UK, and the response time of the same service may vary greatly due to the different user location, functionality and user network conditions, etc. Alice then turns to her colleagues in India for suggestion. They suggest her to try the service k, provided by a local company though ranked lower in the previous recommendation list. After trying it, Alice thinks that service k has a good performance and meets her requirements.

The problem that she faced is the of India service that meets her both functional and non-functional requirements. As Alice needs to try the recommended service so nearby one, the current way of finding a suitable web service is rather in efficient. To solve this problem we propose am or e accurate approach for service recommendation with consideration of the region factor. Moreover, we will try to provide more in formative and user friendly interface for browsing the recommendation results rather an a list, which will help users to know more about the overall performance of the recommended services, so as to build users trust in the recommendations.

On the basis of collected QoS data, we are going to design our system approach as two–phase process. In the first phase, we will divide the users into different regions based on their physical locations and historical QoS experience on web services. In the second phase, we will find similar users for the current user and will make QoS prediction for the unused services. And after that services with the best predicted QoS will be recommended to the current user.

C. Phase 1: Region Creation

We are going to group users, which are closely located with each other and likely to have same QoS values. Each user member belongs to exactly one region, while regions are going to be internally coherent, but clearly different from each other. The region creation phase will be designed as a three step process. In the first step, we will put users with the similar IP addresses into small regions and then we will extract region features. In the second step, we will calculate the similarity between the different regions. In the last step, we will aggregate highly correlated regions to form a certain number of large regions[7]. Details of these three steps are presented in following sections from 1 to 3 respectively

1) Region Feature Extraction

For each region, we are going to use region center as the main feature to reflect the average performance of web services, observed by the region users. The center will be defined as median vector of all the RTT vectors associated with the region users. The element X of the center will be the median RTT value of service, which will be observed by the users from the region. Median will be the numeric value separating the higher half of a sample from the lower half. The average quality will be observed by region users. We are also going to pay importance to the variation of the service performance. From large number of QoS records, we will discover that the service response time usually varies from region to region. Some services may have unexpected long response time or they may be unavailable to some regions. Being inspired by the three sigma rule[8],which is often used to test outlier, we are going to use a method to distinguish for region–sensitive services, which are having unstable performance for different regions, and which will be another important region feature besides the region center. These to of non-zero RTTs of service will be collected from the users of all regions, which will be a sample from the population of service's response time.

To calculate theme an and the standard deviation of the population, we are going use two robust measures: median and median absolute deviation (MAD). MAD is defined as the median of the absolute deviations from the sample's median.

$$MAD=median_i(|R_{i,s}-median_j(R_{j,s})|) \quad i=1,\ldots\ldots,k, j=1,\ldots\ldots,k \qquad (1)$$

Based on the median and MAD,the two estimators can be calculated by,

$$\mu=median_i(R_{i,s})I= 1,\ldots k \qquad (2)$$

- Definition 1(Region Sensitivity): The sensitivity of region M is the fraction between the number of sensitive services in region M over the total number of services.
- Definition 2 (Sensitive Region): Region M is a sensitive region if its region sensitivity exceeds the sensitivity threshold λ.

By the above definitions, we can identify services has drastically fluctuating response time and those regions where the fluctuation occurs, is an important feature for service QoS prediction and recommendation.

2) Region Similarity Computation

Determining whether the two regions are similar is an important step before region aggregation. The similarity of two regions M and N will be measured by region centers m and n. Pearson correlation coefficient (PCC) is widely used in recommender systems to calculate the similarity of two users. PCC value ranges from-1to1. A positive and negative value of PCC indicates that the two users have similar, opposite preferences respectively. PCC will compute the similarity between two regions M and N based on following formula

$$Sim(m,n)=\frac{\sum_{s\in S(n)\cap S(m)}(R_{m,s}-\overline{R_m}).(R_{n,s}-\overline{R_n})}{\sqrt{\sum_{s\in S(n)\cap S(m)}(R_{m,s}-\overline{R_m})^2}\cdot\sqrt{\sum_{s\in S(n)\cap S(m)}(R_{n,s}-\overline{R_n})^2}}.$$

Where S(m) is the set of services invoked by users in region M, and S(n)is the set of services invoked by users in region N. $R_{m,s}$ will be the RTT value of service S, which will be provided by the region centre m. R_m and R_n will represent the average RTT of all the services of centre m and n, respectively. The PCC often overestimates the similarity between two no similar regions, because of very similar RTT, PCC only considers the RTT difference of the co-invoked services by both regions. It often overestimates the similarity of two regions that are not similar, but happen to have a few co-invoked services with very similar RTTs[10].We are going to assume the accuracy of prediction that can be improved, if we add a correlation significance weighting factor to reduce the overestimated similarity. We will use the following adjusted PCC equation to calculate the similarity between two regions:

$$Sim'(m,n)=\frac{|S(m)\cap S(n)|}{|S(m)\cup S(n)|}Sim(m,n) \qquad (3)$$

Where, $|S(m)\cap S(n)|$ is the number of web services invoked by both regions, and $|S(m)\cup S(n)|$ is the number of web services invoked either by region *M* or by region *N*.

3) Region Aggregation

Due to the user's limited use of web services, we will get limited QoS record forming a very sparse QoS dataset. In this case, it will be difficult to find similar users and to predict the QoS values of the unused web services for the active user. To solve this problem, on the basis of region features, we propose a region aggregation method. As shown in algorithm 1, the region aggregation approach is a bottom-up hierarchical clustering algorithm [9].The input is a set of small regions r_1; r_l.Each region will consist of users with similar locations. The algorithm will successively aggregate pairs of the most similar, no sensitive regions until the stopping criterion.

D. Phase 2: QoS Value Prediction

After clustering the users in limited regions in previous steps, we can easily predict QoS values for rest unused web services. The service experience of users in a region will be represented by the region center. With the compressed QoS data searching neighbors and making prediction for an active user can be computed quickly. In our approach, similarity between the active users and users of a region will be computed by the similarity between the active user and the region center. Moreover, it will be more reasonable to predict the QoS value for active users based on their regions. Users in the same region are more likely to have similar QoS experience on the same webservice, especially on that region–sensitive ones. To predict the RTT value for the active user a on an unused services, we will take the following steps.

We will find the IP address of the user and its region. If no region found, a new region will be created. A prediction will be generated from the region center if the QoS values observed is different, if not, we will compute the similarity between the active user and each region center, that has evaluate d service s and then will find up to km ost similar centers. If the active user's region center will have the RTT value of s, the prediction will be computed using below equation

$$\hat{R}_u(s_i)=\overline{R_u}+\frac{\sum_{j=1}^{k}(R_{c_j}(s_i)-\overline{R_{c_j}})Sim'(u,c_j)}{\sum_{j=1}^{k}Sim'(u,c_j)}$$

Where, $R_{uj,s}$ will be the RTT of services provided by center c_j.The prediction will consist of two parts. One will be the RTT value of the region center of the active user $R_{center,\ s}$, which will denote the average QoS observed by these region users. The other part will be the normalized weighted sum of the deviations of the services, RTT from the average RTT observed by the most similar neighbors. So, by this method of prediction, we can predict about the QoS values of the unused webservices.

IV. Conclusion

In this paper, we are going present an innovative approach to web service recommendation and visualization. Different from the previous work, our algorithm will employ the characteristic of QoS by clustering users into different regions. After this, based on the region feature, are fined nearest– neighbor algorithm will be proposed to generate QoS prediction. The final service recommendations will be put on a map to reveal the underlying structure of QoS space, so it will help the users to accept the recommendations.

In this paper, our recommendation approach will be considered for the correlation between QoS records and users physical location, by using IP addresses, which will achieve a good prediction performance.

In this way initially, we are going to do the region creation to obtain the location of the service users, and after that QoS prediction and finally we are going to use trust worth in essential evaluation model, to filter out the results, to gain the trust and confidence upon the results.

REFERENCES

[1] XiChen, ZibinZheng, XudongLiu, ZichengHuang, and HailongSun" Personalized QoS Aware WebService Recommendation and Visualization". January-March2013 IEEE Transactions on Services Computing.
[2] L.Shao,J.Zhang,Y.Wei,J.Zhao,B.Xie,andH.Mei,"Personalized QoS Prediction for WebServices via Collaborative Filtering,"Proc.Int'l Conf .WebServices, pp. 439-446,2007.
[3] Z.Zheng, H.Ma, M.R.Lyu, and I.King, "WSRec: A Collaborative Filtering Based WebService Recommendation System," Proc. Int'l Conf. WebServices, pp. 437-444, 2009.
[4] W.Rong, K.Liu, and L.Liang, "Personalized Web Service Ranking via User Group Combining Association Rule," Proc. Int'l Conf. Web Services, pp. 445-452, 2009.
[5] R.M.Sreenath and M.P.Singh, "Agent-Based Service Selection," J.WebSemantics.vol.1, no.3, pp.261-279, 2003, doi:10.1016/j.websem.2003.11.006.
[6] X.Chen, X.Liu, Z.huang, and H.Sun, "Region KNN: A Scalable Hybrid Collaborative Filtering Algorithm for Personalized WebService Recommendation," Proc. Int'lConf. WebServices, pp.9-16, 2010
[7] J.L.Herlocker, J.A.Konstan, and J.Riedl, "Explaining Collaborative Filtering Recommendations," Proc. ACM Conf. Computer Supported Cooperative Work, pp. 241-250, 2000.
[8] C.D. Mining, P. Raghavan, and H. Schutze, an Introduction to Information Retrieval Cambridge Univ., 2009.
[9] M.R.McLaughlinandJ.L.Herlocker,"ACollaborativeFilteringAlgorithmandEvaluationMetricThatAccurately Model the User Experience," Proc.Ann.Int'l ACM SIGIR Conf., pp.329-336, 2004.

