Fuzzy Clustering Based Highly Accurate Prediction Algorithm for Unknown Web Services

Ms. Mohini Shevatkar
Dept. of Computer Engineering
DYPSOE
Pune, India
mohininmadgulkar@gmail.com

Mr. Soumitra Das
Dept. of Computer Engineering
DYPSOE
Pune, India
soumitra.das@dypic.in

Abstract— In today's reality, the measure of web administrations is expansions on web, so that determination and suggestion of web administration are becoming more imperative. In the fields of E-commerce and other Web-based services recommendation systems are extremely significant. Recommendation system first of all searches the list of web services those having similar functionality, which is user wants. By using filtering, separated the required list and finally on the basis of past records of service provider select out the optimal web services and recommend to users. In this paper predicts that much not known Web services QoS values more precisely than other accessible approaches. Also, we proposed the QoS prediction by utilizing fuzzy clustering technique with ascertaining the clients similarity. Our methodology enhances the prediction accuracy, and this is confirmed by contrasting investigations with different techniques.

Keywords- *Fuzzy clustering; Pearson Correlation Coefficient (PCC); QoS; Recommender system; Web services*

*****

I. INTRODUCTION

Web services are mainly designed for interaction between computers over a network. The web services publication over network is increasing speedily. The users need to select web services from a huge set of applicant services. Hence useful and competent service suggestion is necessary to determining the most appropriate service element. For that reason, Quality of Service (QoS), which characterizes the non-functional properties of services, is applied as a base for service evaluation.

Since several web services have the related features, users have to construct service selecting judgment without any awareness about service contender. Some QoS properties are user independent, having matching values for diverse users while further QoS properties are user-dependent. For example, response time, invocation failure rate etc. The user dependent QoS values are indefinite for the user, if a service which has not been applied before is recommended by the recommender system to the user. So, predicting the indefinite QoS values is awfully essential to find out whether this service is suitable to recommend or not.

Collaborative Filtering (CF) is a widely used approach to construct many popular commercial recommender systems. Existing approaches focus on representing the response time or the throughput of web services. Thus, the performance of a specific service cannot broadly characterize by these quality aspects. To set up a score function and to produce multidimensional QoS recommendations to characterize the overall performance of services there is necessitate of providing web service users with a proficient and useful approach.

Our approach composes application of the fuzzy clustering in QoS prediction. In this, every pattern does not fit in to a correct cluster. This is near to the QoS prediction trouble where users or services have dissimilar characteristics and it is complex to determine the exact cluster. Here we get web service users as the clustering objects and improve the conventional FCM for web service recommendation. We are focusing on minimizing the fault of prediction in case of data sparsity, and developing the value of recommendation by offering a complete QoS characteristics standing effect.

II. LITERATURE SURVEY

S. Ran et. al. [6], talk about problems linked to this sluggish take up and argues that value of services is one of the causal factors. The paper proposes a new Web services detection replica in which the functional and non-functional requirements (i.e. quality of services) are taken into consideration for the service detection.

Z. Zheng et. al. [7], present a collaborative filtering method for forecasting QoS values of Web services and creating Web service recommendation by captivating reward of past usage knowledge of service users. In this, a collaborative filtering method is modelled to forecast Web service QoS values based on the gathered QoS data.

Y. Jiang et. al. [8], present an efficient modified collaborative filtering method for Web service recommendation. A key module is calculation of parallel measurement of Web services. Based on this, they built up an efficient modified Hybrid Collaborative Filtering (PHCF) method by integrating modified user-based algorithm and modified item-based algorithm.

M. Tang et. al. [9], propose a technique of location aware collaborative filtering to recommend Web services to users by

ISSN: 2454-4248
68 – 73

integrating locations of both services and users. Unlike the current user-based collaborative filtering for discovery of related users for a target user, instead of probing complete group of users, they focus on users actually close to the target user. After discovery, they apply the parallel measurement to forecast absent QoS values based on a fusion collaborative filtering method. Web service aspirants with the peak QoS values are recommended to users based on a real-world Web service QoS dataset.

Meng Zhang et. al. [5], goes for creating a more exhaustive web administration suggestion to clients with a novel way to deal with satisfy more precise forecast of obscure administrations QoS values. They perform the QoS expectation by utilizing fluffy grouping system with computing the clients comparability. Their methodology enhances the expectation exactness, and this is affirmed by contrasting analyses and different routines.

III. EXISTING SYSTEM ARCHITECTURE

A. *HAPA (Highly Accurate Prediction Algorithm)*

The user-based and item-based HAPAs can formulate forecast individually; To get better forecast correctness HAPA joins these two types of HAPAs.

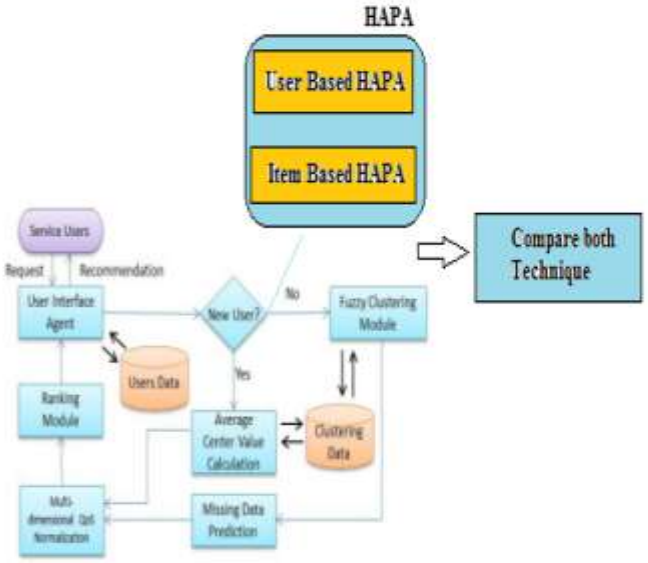

Figure 1. Architecture of HAPA [5]

Figure 1 shows a schematic architecture of web service recommender system. The role of the user interface manager is to accept the user's requirements and recommend services as per the outcome from the standing component. The manager must differentiate if the user is an obtainable active user or a new user, when service user asks a query for a web service to the user interface manager. To find out whether the user is an active user (means, before now user has an access record in database), the manager required to interrelate with the users record. In case if a user is an active user, system will apply the fuzzy clustering component. The role of this component is to regain the cluster linked information. This component takes the factors of grouping and resemblance as key contribution; and produces the membership values matrix and the cluster centers matrix. These two matrixes are key in to the subsequent component; absent data forecast which forecasts absent values of QoS properties. The QoS values will be normalized with the help of multi-dimensional QoS normalization in diverse techniques, and after that recorded into a single dimensional value on top of which the standing component workings. If the user does not have some access knowledge means he is not an active user, in that case we come across every cluster centers, calculate an average values for every service and move towards the normalization component and standing component for the recommendation which is trustworthy with the active user status.

IV. PROPOSED FRAMEWORK AND DESIGN

The main aim of project is Highly Accurate Prediction Algorithm (HAPA) which is Collaborative Filtering based algorithm, i.e., it actually includes user-based and item-based HAPAs. Both kind of HAPA can compose forecasts; though we constantly apply the grouping of two HAPAs to construct more correct forecasts.

User-based HAPA: In this, the unidentified QoS value is premeditated from every forecast values prepared by each related user.

Item-based HAPA: In this, an item is specifically a Web service. It applies related services to compose forecasts. Its mathematical principle is extremely related to that of user based HAPA.

A. *Problem Definition*

In existing system QoS data compile a set of non-functional characteristics; with each one is describing a definite feature of service quality. A number of QoS characteristics are user independent, having the same values for dissimilar users (e.g., popularity, price, and availability) whereas further QoS characteristics are user-dependent (e.g., invocation failure rate, response time). Forecasting the unidentified QoS values is not evaluated whether the service is appropriate recommendation.

B. *Goals and objectives*

For active users, who have had some web services access knowledge, the performance of web services they are going to access can be forecasted by making use of a fuzzy clustering algorithm. In active user condition, our system will make use of the fuzzy clustering component. The working of this component is to collect the cluster associated data. The factors of clustering and similarity are the key in for this component; the membership values matrix and the cluster centers matrix are produced out of this component.

C. Mathematical Model

Let S be the system which does recommend accurate web service to various users. The system presented in this paper, a recommendation system. By calculating most accurate similarity,

S = {SC, RSD, PQ, FPV}

Where,

SC: Similarity Calculation

RSD: Reciprocal of various similarities

PQ: Prediction Quality

FPV: Final Predicted value

Input: A WS-Dream dataset is used which is text file, describes real-world QoS assessment outcome from 339 users on 5,825 Web services.

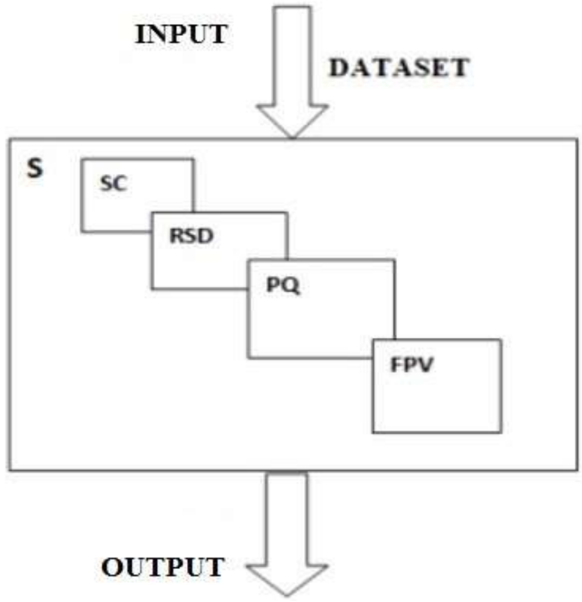

Figure 2. Mathematical model of HAPA

Processes:

1) *SC (Similarity calculation):* In existing system we use PCC method for calculating user and item similarity. PCC is used to determine the similarity among users or items. For user-based CF methods, PCC applies the subsequent equation to determine the similarity among two users u and v based on the Web services they usually called up [1]:

$$\text{sim }(u,v) = \frac{\sum_{i\in I}(r_{u,i}-r_u)(r_{v,i}-r_v)}{\sqrt{\sum_{i\in I}(r_{u,i}-r_u)2}\sqrt{\sum_{i\in I}(r_{v,i}-r_v)2}}$$

where I= Iu ∩ Iv is the subset of Web service formerly accessed by both user u and v. r(u,i) r is the QoS value of Web service i determined by user u, and ru stands for the mean QoS value of diverse Web services determined by user u. In the same way, for item-based CF, the PCC among two service items is considered as [1]:

$$\text{sim }(i,j) = \frac{\sum_{u\in U}(r_{u,i}-r_i)(r_{u,j}-r_j)}{\sqrt{\sum_{u\in U}(r_{u,i}-r_i)2}\sqrt{\sum_{u\in U}(r_{u,j}-r_j)2}}$$

where U =Ui ∩ Uj is the subset of users who have accessed both Web services i and j earlier, and ri stands for the mean QoS value of Web service i determined by diverse users.

2) *RSD (Reciprocal of similarity):* For pvUser and pvitem, MS represents the maximum similarity among users in KU and services in KI, indicated by ms (pvUser) and ms (pvitem), correspondingly [1].

$$\text{ms }(pv_{user}) = \max \{\text{sim}(u, v) | v \in KU\}$$

$$\text{ms }(pv_{item}) = \max \{\text{sim}(i, j) | j \in KI\}$$

For pvUser and pvitem, “as” represents the average similarity among users in KU and services in KI, indicated by as (pvUser) and as (pvitem), correspondingly [1].

$$\text{as }(pv_{user}) = \frac{\sum_{v\in KU} sim(u,v)}{|KU|}$$

$$\text{as }(pv_{item}) = \frac{\sum_{j\in KU} sim(i,j)}{|KI|}$$

For pvUser and pvitem, RSD corresponds to the reciprocal of the standard deviation of the similarities among users in KU and services in KI, indicated by rsd (pvUser) and rsd (pvitem), correspondingly [1].

$$\text{rsd}(pv_{user}) = 1/\sqrt{\frac{\sum_{v\in KU}[sim(u,v)-as(pv_{user})]2}{|KU|}}$$

$$\text{rsd}(pv_{item}) = 1/\sqrt{\frac{\sum_{v\in KI}[sim(i,j)-as(pv_{item})]2}{|KI|}}$$

3) *PQ (Prediction Quality):* If we separate the prediction quality into every index, then we have [1]:

$$\text{Qms}(pv_{user}) = \frac{ms(pv_{user})}{ms(pv_{user})+ms(pv_{item})}$$

$$\text{Qms}(pv_{item}) = \frac{ms(pv_{item})}{ms(pv_{user})+ms(pv_{item})}$$

ISSN: 2454-4248
68 – 73

$$Qas(pv_{user}) = \frac{as(pv_{user})}{as(pv_{user}) + as(pv_{item})}$$

$$Qas(pv_{item}) = \frac{as(pv_{item})}{as(pv_{user}) + as(pv_{item})}$$

$$Qrsd(pv_{user}) = \frac{rsd(pv_{user})}{rsd(pv_{user}) + rsd(pv_{item})}$$

$$Qrsd(pv_{item}) = \frac{rsd(pv_{item})}{rsd(pv_{user}) + rsd(pv_{item})}$$

The prediction quality of pvUser and pvitem afterward obtained as [1]:

$$Q(pv_{user}) = Qms(pv_{user}) + Qas(pv_{user}) + Qrsd(pv_{user});$$

$$Q(pv_{item}) = Qms(pv_{item}) + Qas(pv_{item}) + Qrsd(pv_{item});$$

4) *FPV (Final predicted value):* Thus, the final predicted value of ru is [1]:

$$pv = \frac{pv_{user} * Q(pv_{user}) + pv_{item} * Q(pv_{item})}{Q(pv_{user}) + Q(pv_{item})}$$

D. Software Architecture

Software architecture serves as the blueprint of Highly Accurate Prediction Algorithm with Fuzzy Clustering, describing the task allotments to facilitate the design and execution.

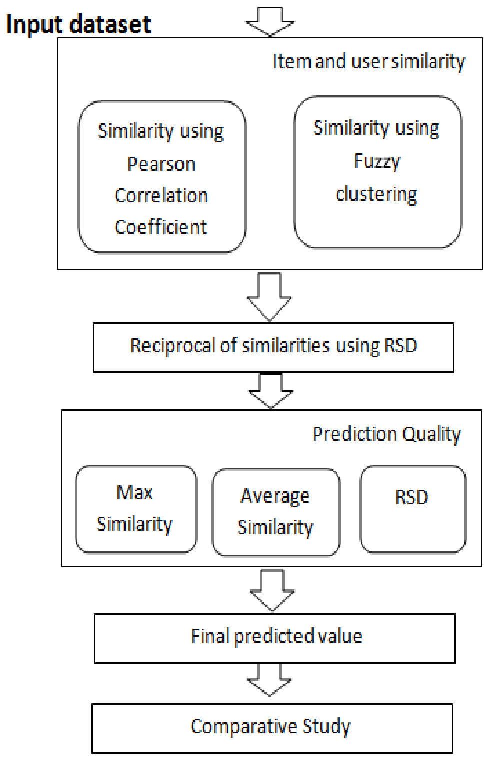

Figure 3. Architecture diagram of a Highly Accurate Prediction Algorithm for Unknown Web services with Fuzzy Clustering

The architecture is the key transporter of system capabilities such as modifiability, presentation, and safety; no one can be recognized by lacking a combining architectural vision. HAPA with FC is consisting of following modules. The description of individual module is explained with detail below.

As shown in architecture diagram in fig. 3, here we take a wsrec_dataset as an input. In existing system item and user similarity is evaluated using Pearson correlation coefficient and in our proposed we are using with that fussy clustering which gives more accurate similarity between users and items. After that we calculate reciprocal of all this similarity using reciprocal of standard deviation, and then we are going to calculate quality of predictions by calculating max Similarity, average similarity and RSD values. After calculating it for all user and item final predicted value is calculated. Finally we are going to do comparative analysis in between existing work and proposed work by using MAE and RMSE evaluation.

E. System Modules

1) *GUI Module:*

TABLE I. SOME THROUGHPUT VALUES IN WSDREAM

	WS_0	WS_1	WS_2	WS_3	WS_4	WS_5	WS_6	WS_7
U_{110}	0.798	5.134	7.623	2.583	2.684	4.065	0.337	47.24
U_{122}	0.318	4.347	6.437	2.171	2.322	0.52	0.321	43.816
U_{123}	0.318	4.489	6.795	2.309	2.293	0.571	0.324	43.337
U_{124}	0.328	4.362	7.009	2.229	2.247	0.579	0.321	43.673
U_{242}	0.965	13.513	19.672	7.042	7.017	2.676	1.661	$\boxed{112.277}$

This module is contribution of two different approaches:

1. First, an interface's status, which often controls its presence and actions.
2. Second, relationships among interface components and underlying data models.

In this module, user simply gives text file of WS-Dream dataset as input for prediction techniques and pre-processing on it.

rtMatrix.txt - 339 * 5825 user-item matrix of response-time.
tpMatrix.txt - 339 * 5825 user-item matrix for throughput.

2) *Prediction Value Calculation:* Final prediction value is calculated by calculating various quality values like Similarity, Reciprocal Similarity, Quality Prediction etc. in this module.

3) *Accuracy Calculation:* In this module accuracy of result is calculated by mean absolute error (MAE) and root mean squared error (RMSE).

 ISSN: 2454-4248
68 – 73

F. *Secrete Sharing Algorithm*

This fuzzy clustering method taken with reference from paper "A Web Service Recommendation Approach Based on QoS Prediction Using Fuzzy Clustering".

According to the problem definition, we have U = {u1, u2, u3, · · ·, uv} as the set of v users that have provided a set of services access data, and ui offers Qi = {qi1, qi2, qi3, · · ·, qiw}, here we take as the ith pattern. So Q = {Q1, Q2, Q3, · · ·, Qv} is a set of v patterns match up to v users, and every pattern is a w-dimensional vector equivalent to w services.

The key design of fuzzy clustering is to construct k cluster centers, and every pattern has a membership value to every cluster center. Thus our intention of this component is to find out the membership degree of every user for each cluster center.

The membership function in FCM should assure the subsequent surroundings [5]:

$$\begin{cases} \sum_{i=1}^{k} \mu_{ij} = 1, \forall j = 1, \dots, v \\ 0 < \sum_{j=1}^{n} \mu_{ij} < n, \forall i \\ \mu_{ij} \epsilon [0,1] \end{cases}$$

In equation, μij corresponds to the membership value of the *jth* pattern for the *ith* cluster. The fuzzy clustering is an iterative method, it loops till the objective function getting a minimum, and the objective function is evaluated by applying the following equation in the original algorithm as stated.

FCM is a data clustering technique which had been broadly used in pattern determination. It is launch by Bezdek and it seeks at clustering several patterns into clusters while reducing the objective function [5]:

$$J = \sum_{j=1}^{n} \sum_{i=1}^{c} \mu_{ij}^{m} \, (d_{ij})^2$$

Where $(d_{ij})^2 = \|C_i - Q_j\|^2$ represents the Euclidean distance between ith cluster and jth pattern in the original article. n represents the number of patterns; c stands for the number of clusters; m represents the weighting exponent; µij is the membership value that illustrates the membership degree of that the jth pattern to the ith cluster center; and dij is a distance measure function to determine the distance among the jth pattern and the ith cluster center.

V. Result

We used two datasets to evaluate our approach. There are two matrices used for response time and throughput. First is graph shows accuracy for RT and second is for accuracy for TP which shows that our proposed method achieves highest accuracy.

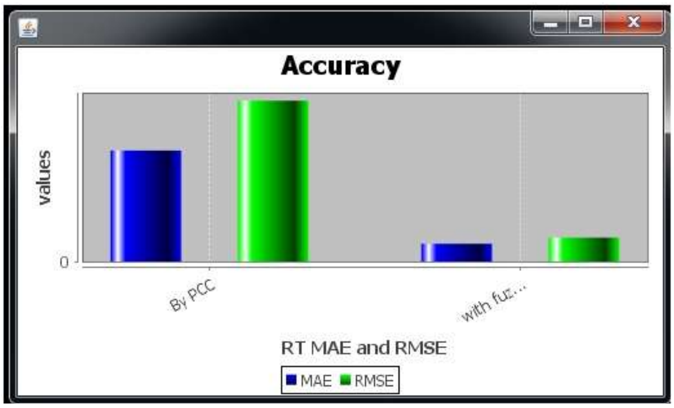

Figure 4. Response time graph

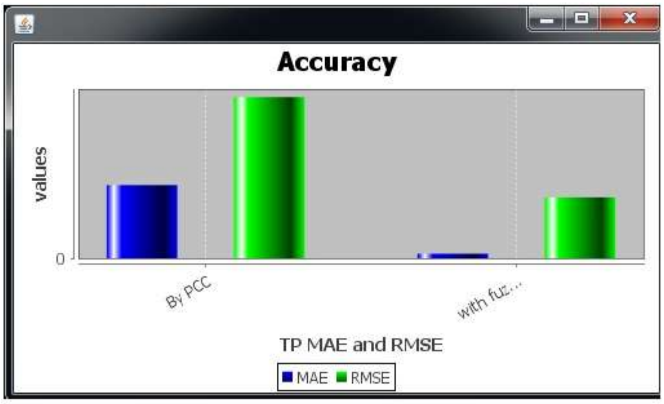

Figure 5. Throughput graph

VI. Conclusion

We have used CF to anticipate obscure QoS values. Entirely talking, our methodology is basically distinctive from conventional CF which is not appropriate to objective information expectation. QoS data is related to a different characteristic. It is hard to split all QoS data determined by distinct users into sets or clusters on a number of precise controls. In our approach, we advance the clustering method based on the FCM algorithm by joining PCC calculation. We use the value of similarity among user' data and cluster centers instead of using Euclidean distance. In other words, we use the PCC as the distance measure function to evaluate the membership function, it forecasts the missing QoS values more precisely in case of data sparsity. So, we get more precise similarity values, based on these factors, we proposed our HAPA. The prediction accuracy of HAPA was shown to outperform that of several existing QoS prediction techniques. There are still various parts of QoS properties which can be considered later on, for example, system environment the security, and so forth. Additionally, incorporates considering clients inclinations, for example, requesting that clients choose the measurement of QoS properties that he minds more. Such, we can enhance the heaviness of measurements in the general quality score computation.

ACKNOWLEDGMENT

It gives me a great pleasure and huge satisfaction to present this paper of topic "**Fuzzy Clustering Based Highly Accurate Prediction Algorithm for Unknown Web Services**" which is the result of steady support, expert guidance and focused direction of my guide **Mr. Soumitra Das** to whom I convey my deep sense of gratitude and humble thanks, for his valuable guidance.

REFERENCES

[1] You Ma, S. Wang, Patrick C.K. Hung, Ching-Hsien Hsu, Qibo Sun, and F. Yang, "A Highly Accurate Prediction Algorithm for Unknown Web service QoS Values" IEEE Trans, Services Computing, vol. 6, 2015. *(references)*

[2] X. Chen, Z. Zheng, X. Liu, Z. Huang, and H. Sun, "Personalized qos aware web service recommendation and visualization," Services Computing, IEEE Trans, Services Computing, vol. 6, pp. 35-47, 2013.

[3] J.Wu, L. Chen, Y. Feng, Z. Zheng, M. C. Zhou, and Z.Wu, "Predicting quality of service for selection by neighborhood based collaborative filtering," IEEE Trans, Systems, Man, and Cybernetics: Systems, vol. 43, pp. 428-439, 2013.

[4] L.-J. Zhang, J. Zhang, and H. Cai, Services Computing. Springer and Tsinghua Univ, 2007.

[5] M. Zhang, X. Liu, R. Zhang, H. Sun, "A Web Service Recommendation Approach Based on QoS Prediction Using Fuzzy Clustering" IEEE Ninth International Conference on Services Computing, 2012. *(references)*

[6] S. Ran, "A model for Web services discovery with QoS," ACM Sigecom exchanges, vol. 4, no. 1, pp. 1-10, 2003.

[7] Z. Zheng, H. Ma, M. R. Lyu, and I. King, "QoS-aware Web service recommendation by collaborative filtering," IEEE Trans, Services Computing,vol. 4, pp. 140-152, 2011.

[8] Y. Jiang, J. Liu, M. Tang, and X. Liu, "An effective Web service recommendation method based on personalized collaborative filtering," Proc. Ninth Int'l Conf. Web Services (ICWS '11), pp. 211-218, 2011.

[9] M. Tang, Y. Jiang, J. Liu, and X. Liu, "Location-aware collaborative filtering for qos-based service recommendation," Proc. 10th Int'l Conf. Web Services (ICWS '12), pp. 202-209, 2012.

[10] MX. Chen, X. Liu, Z. Huang, and H. Sun, "Regionknn: A scalable hybrid collaborative filtering algorithm for personalized Web service recommendation," Proc. Eighth Int'l Conf. Web Services (ICWS '10), pp. 9-16, 2010.

[11] L. Shao, J. Zhang, Y. Wei, J. Zhao, B. Xie, and H. Mei, "Personalized qos prediction forWeb services via collaborative filtering," Proc. Fifth Int'l Conf. Web Services (ICWS '07), pp. 439-446, 2007.