ISSN:0975-9646

M.Thirumaran et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (2) , 2015, 1433-1438

Collaborative Web Service QoS Prediction with Multi-Criteria Decision Making Using CB-NIMF

M.Thirumaran[1], A.V.Arimathi[2]

[1, 2]*Department of Computer Science and Engineering, Pondicherry Engineering College, Puducherry-605014, India.*

***Abstract* – The composite web services on the World Wide Web has increased the demand of efficient web service quality evaluation approaches. To avoid the expensive and time-consuming composite services, this paper proposes a collaborative web quality-of-service (QoS) prediction approach by using the past web service usage experiences of service users. First, we apply the concept of user-collaboration for the web service QoS information sharing. Then from the collected QoS data, a criteria-based neighborhood integrated matrix factorization (CB-NIMF) approach is designed for accurate QoS value prediction. To improve the prediction, CB-NIMF can be extended for multi-criteria complex decision making over the WS-QoS using AHP (Analytical Hierarchical Processing).**

Key words — composite web service, user-collaboration, criteria-based matrix factorization, QoS prediction, multi-criteria

I. INTRODUCTION

Web services are independent software components realizing specific tasks, which can communicate with each other by exchanging messages. The process called service composition usually results in the creation of a new composite services that can be defined as the aggregation of other elementary or composite services. The Quality of service (QoS) deals with the nonfunctional characteristics of the web services. With the increase in composite services, QoS has become an important functionality equal to web service.

Many QoS-based approaches have been proposed for web service composition. Accurate QoS values of the composite services are required for these QoS-based approaches to work well. The QoS values of composite services are measured both at client-side as well as server-side. The QoS value measured at the client-side (e.g. response time, availability, throughput, execution time, etc) can vary widely according to the user environment whereas the QoS values measured at the server-side (e.g. price, popularity, etc) are usually identical for different users.

Moreover, real-world composite service evaluations at the client-side and at server-side are difficult since: 1) composite service invocations are chargeable because services are hosted by organization. 2) Service users are not experts on composite service evaluation. 3) It consumes more time for all service users to evaluate all web service candidates.

Without sufficient client-side and server-side QoS evaluations, it is impossible to obtain accurate web service QoS values and it is difficult for QoS based approaches, which employs these values as the input. To overcome all the above disadvantages, we propose a criteria-based neighborhood integrated matrix factorization for a collaborative web service value prediction. Here we are considering the past web service usage experience of the service users to predict the accurate QoS values both at client-side and at server-side.

Our proposed work consists of a Composite web service which is a combination of multiple sub service. Each sub service consists of a multiple QoS parameters which are mapped to the criteria based parameters like cost, performance, security, usability etc. Our CB-NIMF approach first finds the user similarities from a set of similar users. Both the local information and global information are used to avail a QoS value which fits our factor model to predict the personalized web service QoS value. The collaboration of different service users, who does not have any idea regarding the internal designs and implementation details or does not made any evaluation can effectively predict the QoS values.

Our approach overcomes the shortcomings of previous evaluation approaches by avoiding the expensive and time-consuming real-world composite web service invocations. This paper mainly focuses on providing accurate and personalized Qos values for the service users whereas others are complementary to various QoS based approaches which mainly focuses on mainly using the QoS values.

The contribution of this paper is three fold:

- First, we propose a CB-NIMF approach for personalized composite web service QoS value prediction. By systematically fusing the model-based and neighborhood-based collaborative filtering approaches to achieve higher prediction quality in our approach by using the past web service usage experience of service users.
- Second, we created a composite web services and extracted the web service QoS data set. Based on this web service QoS data set we found the QoS value prediction accuracy of our approach.
- Third on finding the accurate composite web service QoS values, we extended it for multi-criteria complex decision making over the WS-QoS using AHP (Analytical Hierarchical Processing) for finding the best fit web service according to the customer requirements.

The remainder of this paper is organized as follows:

Section 2 presents our collaborative web service QoS value prediction using our approach and also explain about AHP. Section 3 describes about metrics used, and Section 4 concludes the paper.

II. COLLABORATIVE WEB SERVICE QOS PREDICTION

Quality-of-Service deals with the nonfunctional characteristics of the web service. The performances experienced by the service users are more realistic at the client-side whereas QoS values provide good indication of the server capacities at the server-side. A wide range of QoS properties are considered in our work. Some of the commonly used web service client-side and server-side QoS properties include:

Response-time - the time duration for a service user to send a request and receive a response.

Reliability – it is a probabilistic measure of correctly delivering services during a period of time.

Availability - it is measured in percentage in which it is accessible and available publically via internet.

Price – the services which are cheap and best at the server side.

Popularity – the most often used services.

Higher the service user contributes, higher the QoS value prediction accuracy is achieved in our approach. Collaborative web service QoS value can be made based on the collected QoS values.

The workflow for our approach is indicated in Fig 1. Our proposed work consists of a Composite web service (e.g. CS) which is a combination of multiple sub services (e.g. S1,S2 ect). Each sub service consists of a multiple QoS parameters which are mapped to the criteria-based parameters like cost, performance, security, usability etc. Further, a criteria-based user-item matrix is formed to predict the QoS values with respect to its QoS properties. The main advantage composite web service with criteria-based is that 1) it finds the best fit web service for the customer as per their requirements. 2) It has multiple selection option for the customer. 3) It has more scope etc.

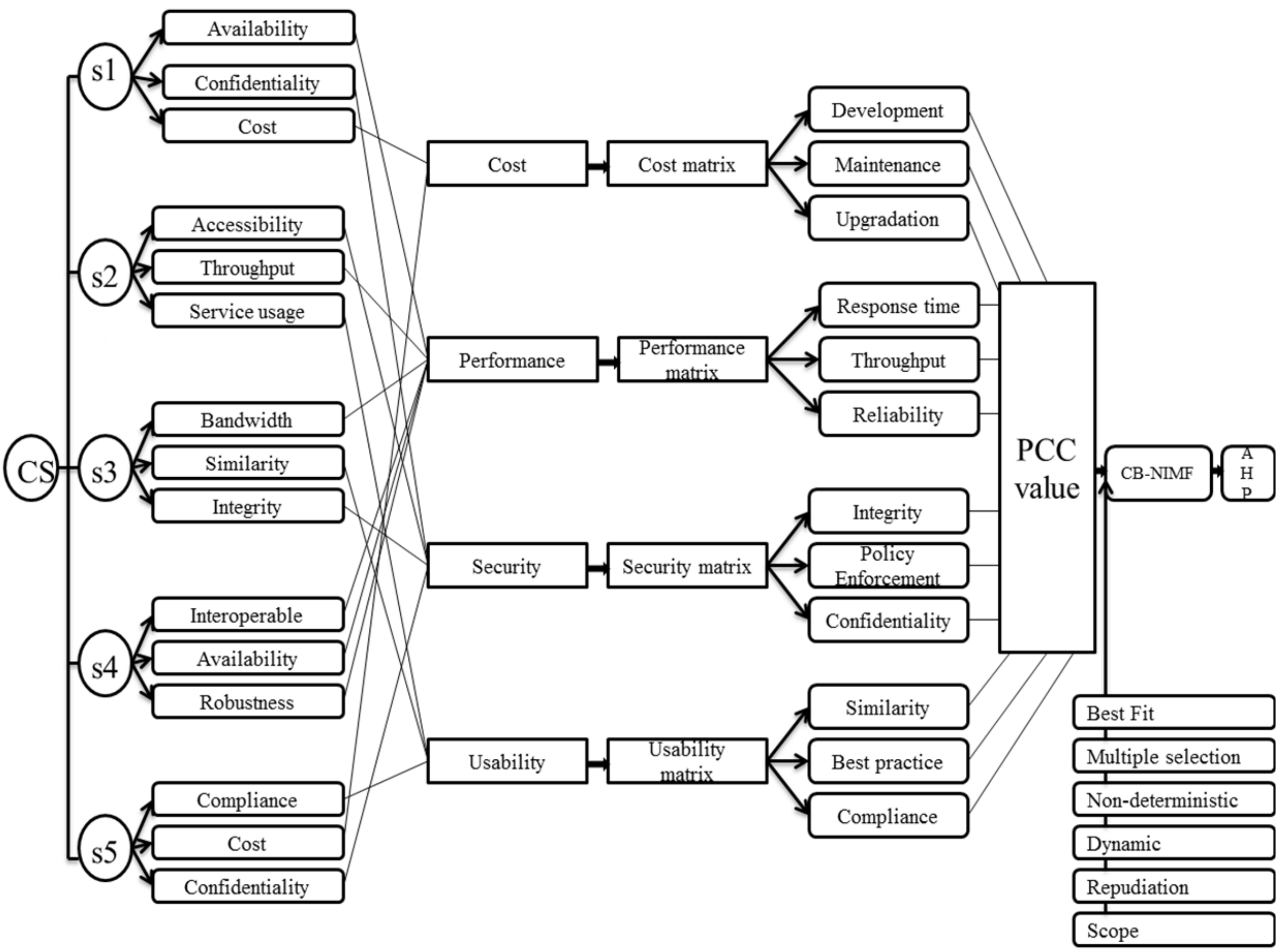

Fig1. Flow diagram for criteria-based prediction

	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10
u1	1.3	1.1	0.2	0.1	1.5			0.3	0.5	
u2	1.7	1.1	0.2	0.1	1.5				0.5	
u3	0.1	0.3	0.5	1.2				1.2		0.4
u4	1.5		1.4	0.1	1.1				0.6	
u5	0.2			0.2	1.5			0.5	0.9	
u6	0.1	0.3	0.5	1.6		0.5	0.2	0.3	1.7	
u7	1.3	0.1	0.2	0.1	1.5				0.5	
u8	1.3	1.0	0.2	0.1	1.5				0.5	
u9	0.1	0.3	0.5	1.2			1.2		0.4	
u10	1.5		0.4	0.1	1.1				0.6	
u11	0.2			0.2	1.5			0.5	0.9	
u12	0.5	0.3	0.5	1.6		0.5	0.2	0.3		0.8

Fig 2. User-Item matrix

2.1. DESCRIPTION OF THE PROBLEM

The process starts with a user-item matrix in which the entries in the matrix represent certain QoS property of a web service at client-side or at server-side. Consider the QoS property (e.g. response-time) of a web service (e.g. i_1 to i_{10}) observed by a service user (e.g. u_1 to u_{12}). The response-time value of each service user depends on the invoked web service. The two different user similarities can be calculated by analyzing their same web service QoS values. For similarity computation, here we use Pearson correlation coefficient (PCC). To specify the similarity between the two users u1 and u2 the PCC value ranges from [-1, 1]. If the PCC value is high there exist higher similarity and vice versa. If the users do not invoked any common services, then it is non-available. In this paper, the problem we study is how to find the missing QoS values in the user-item matrix. Using our approach, we can provide the accurate missing QoS values for making service ranking, finding the best fit web service according to the user requirement etc.

To find the missing values in the user-item matrix, we can employ the observed QoS values of the web service by other service users for predicting the performance of the web service for the current user. The current user may not experience the similar QoS performance since they are under different network conditions and different geographic locations.

To overcome the above challenging web service QoS value prediction problem, we propose a CB-NIMF approach where we use both the local information of similar user and global information of all available user. Our approach consists of three phase. In phase 1, the user similarities are calculated using PCC and set of Top-k similar users for the current users are determined. Then in phase 2, using our proposed CB-NIMF approach, the missing values are found in the user-item matrix. In phase 3, CB-NIMF can be extended for multi-criteria complex decision making over the WS-QoS using AHP (Analytical Hierarchical Processing) for finding the best fit web service according to the customer requirements.

2.2 phase 1: Criteria Based – Neighborhood Similarity Computation

Consider a × b user-item matrix Z which consists of 'a' service users and 'b' web services. Client-side or server-side QoS property value of a web service' j' observed by service user 'i' are represented as Z_{ij} in the matrix. Z_{ij} = null, when the user 'i' did not invoked any web service 'j'. Using PCC we can compute the similarities between the service users in the user-item matrix with available web service QoS values. Here PCC is used since it can achieve higher performance than other.

With the following equation by employing PCC, we can find the similarity between the two users i and k based on their observed QoS values on the commonly invoked web services are as follows:

$$PCC(i,k) = \frac{\sum_{j\epsilon J}(Zij-\overline{Zi})(Zkj-\overline{Zk})}{\sqrt{\sum_{j\epsilon J}(Zij-\overline{Zi})^2}\sqrt{\sum_{j\epsilon J}(Zkj-\overline{Zk})^2}} \quad (1)$$

where j is then subset of web services that are invoked by both the user i and user k, Z_{ij} is the QoS value of web service j observed by the user i. $\bar{Z_i}$ and $\overline{Z_k}$ are the average QoS values of different web services observed by the service user i and k respectively. Similarity of two service user i and k can be calculated from the above definition where PCC (i,k) value ranges from [-1,1]. If PCC value is high, there exist higher user similarity and vice versa.

Once the user's similarity are calculated and based on the PCC values, a set of Top-K similar users can be identified. Limited number of similar users may differ in using services. So, this can be ignored using traditional Top-K algorithm which includes dissimilar users with negative PCC values.

2.2.1 ALGORITHM

The Top-K algorithm (a selection algorithm) is used to set the similar users.

For a service user i, a set of similar users T(i) can therefore be identified by the following equation

$$T(i) = \{ k \mid k \in \text{Top-k}(i), PCC(i,k) > 0, i \neq k \} \quad (2)$$

where Top-k = similar users to the current user i

PCC (i,k) = PCC similarity value between user i and user k

2.2.2 SELECTION ALGORITHM

Step 1: Starts the process by comparing first two elements of an array and swaps if necessary.

Step 2: If element one greater than element two, it swaps.

Step 3: Again first element is compared with third element and swaps if necessary.

Step 4: The process continues until first element and last element of an array is compared.

Step 5: End of the process.

It is not mean that user k is in the Top-k neighbors of user i does not indicate that user i is also in the Top-k neighbors of user k. With this information we can find our CB-NIMF model for the QoS value prediction.

2.3 Phase 2: Criteria Based – Neighborhood Integrated Matrix Factorization

A familiar approach to predict the missing values to a factor model that fits in the user-item matrix. Consider an a×b user-item matrix Z, a rank-l matrix X = $U^T V$ is employed in matrix factorization method to fit it, where U € $IZ^{l\times a}$ and V € $IZ^{l\times b}$. The low dimensional matrix U and V are unknown in the above definition and need to be estimated. This feature represents clear meaning. A user's web service QoS values corresponds to a linear combination of the vectors with user-specific coefficients. Each column of U performs as a "factor vector" for a user and each column of V is a linear predictor for a web service, predicting the entries in the corresponding column of the user-item matrix Z based on the "factors" in U. The length of the "factor vector" is called its dimensionality. By adding the constraints of the norms of U and V to penalize large values of U and V, where we have following optimization problem.

$$\min \mathcal{L}(Z,U,V) = \frac{1}{2} \sum_{i=1}^{a}\sum_{j=1}^{b} I_{ij}^{Z} (Z_{ij} - U_i^T V_i)^2 + \frac{\lambda U}{2}\|U\|_F^2 + \frac{\lambda V}{2}\|V\|_F^2 \tag{3}$$

where I_{ij}^{Z} is the indicator function, that is equal to 1 if the user u_i invoked web service v_j. If the user does not invoke any service then it is represented as 0. $\|.\|_F^2$ Denotes the Frobenius norm, and λU and λV are two parameters. The sum-of-squared-errors objective function with quadratic regularization terms minimizes the optimization problem in the above equation (3). It also has the probabilistic interpretation with the Gaussian observation noise.

The global information of all QoS values in the user-item matrix for all predicting missing values is utilized by the above approach. This approach is effective at estimating overall global information that relates simultaneously to all users or items. But this model is poor at identifying association with a small set of closely related users or items instead neighborhood models performs better. The web service QoS value in the user-item matrix is sparse. Hence, neither a factorization method nor neighborhood based approaches provide optimal results. Here we employ a parameter to fuse both global information and local information. The idea is that every time when factorizing a QoS value, we treat it as the ensemble of a user's information and the user's neighbor's information. The neighbors of the current user can be obtained by employing (2). Hence, we can minimize the following sum-of-squared-errors objective functions with quadratic regularization terms:

$$\mathcal{L}(\mathrm{Z,S,U,V}) = \frac{1}{2}\sum_{i=1}^{a}\sum_{j=1}^{b} I_{ij}^{Z} \Big(Z_{ij} - \big(\alpha\, U_i^T V_i + (1-\alpha)\sum_{k\epsilon T(i)} S_{ik}\; U_k^T V_j\big)\Big)^2 + \frac{\lambda U}{2}\|U\|_F^2 + \frac{\lambda V}{2}\|V\|_F^2 \tag{4}$$

Where $T(i)$ is a set of Top-K similar users of user u_i and S_{ik} is the normalized similarity score between user u_i and useru_k, which can be calculated by

$$S_{ik} = \frac{PCC\,(i,k)}{\sum_{k\epsilon T(i)} PCC\,(i,k)} \tag{5}$$

A local minimum of the objective function is given by (4) can be performed by gradient descent in U_i , V_j:

$$\frac{\delta\mathcal{L}}{\delta U_i} = \alpha \sum_{j=1}^{b} I_{ij}^{R}\, V_i \left(\left(\alpha\, U_i^T + (1-\alpha)\sum_{k\epsilon T(i)} S_{ik}\; U_k^T V_j\right) - Z_{ij}\right) + (1-\alpha)\sum_{\rho\epsilon\beta(i)}\sum_{j=1}^{b} I_{\rho j}^{Z}\; S_{\rho(i)}\, V_j \left(\left(\alpha U_\rho^T V_j + (1-\alpha)\sum_{k\epsilon T(\rho)} S_{\rho k}\; U_k^T V_j\right) - Z_{\rho j}\right) + \lambda_u\, U_{i,} \tag{6}$$

$$\frac{\delta\mathcal{L}}{\delta V} = \sum_{i=1}^{a} I_{ij}^{Z} \left(\left(\alpha U_i^T V_j + (1+\alpha)\sum_{k\epsilon T(i)} S_{ik}\; U_k^T V_j\right) - Z_{ij}\right) \times \left(\alpha U_i + (1-\alpha)\sum_{k\epsilon T(i)} S_{ik}\; U_k^T\right) + \lambda_V\, V_{j.} \tag{7}$$

Consider the values of response-time as given in Fig 2, ranges from 0 – 20s. After predicting the composite web service QoS values using our CB-NIMF approach, 96 percent of the response-time values are smaller than 2s. Our CB-NIMF approach can be directly employed on different QoS properties without any modifications.

2.4 Phase 3: Multi-criteria decision making using Analytical Hierarchical Processing (AHP)

The Analytic Hierarchy Process (AHP) is a structured technique for organizing and analyzing complex decisions, based on mathematics and psychology. In short, it is a method to derive ratio scales from paired comparisons. It consists of an overall goal, a group of options or alternatives for reaching the goal, and a group of factors or criteria that relate the alternatives to the goal. The criteria can be further broken down into sub-criteria, sub sub-criteria, and so on, in as many levels as the problem requires. The input can be obtained from actual measurement such as price, weight etc or from subjective opinion such as satisfaction feelings and preference. The ratio scales are derived from the principal Eigen vectors and the consistency index is derived from the principal Eigen value.

2.4.1 Pair-wise Comparison

Once the hierarchy has been constructed, the participants analyze it through a series of pair-wise comparisons that derive numerical scales of measurement for the nodes. The criteria are pair-wise compared against the goal for importance. The alternatives are pair-wise compared against each of the criteria for preference. The comparisons are processed mathematically, and priorities are derived for each node.

2.4.2 Eigen Value and Eigen vector

For comparison matrix, we compute priority vector which is the normalized Eigen vector of the matrix. The normalized principal Eigen vector is also called priority vector. Since it is normalized, the sum of all elements in priority vector is 1. The priority vector shows relative weights among the things that we compare. Principal Eigen value is obtained from the summation of products between each element of Eigen vector and the sum of columns of the reciprocal matrix.

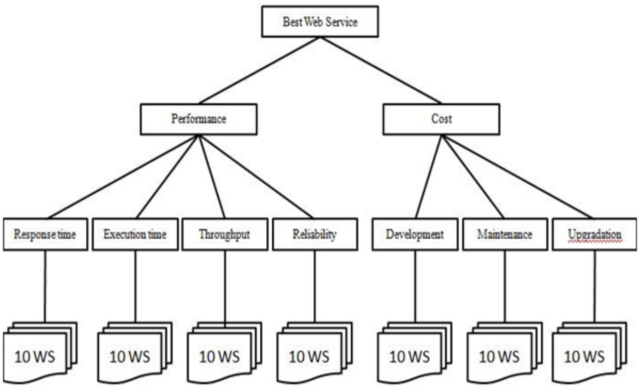

Fig 3.Analytical Hierarchical Processing

In the Fig 3, our aim is to find the best fit web service according to the customer criteria. The criteria here are performance and cost. The performance related QoS parameters are response-time, execution-time, throughput and reliability. Likewise for cost, the related QoS parameters are development cost, maintenance cost and up gradation cost. Here, we considered 10 sets of composite web services. Among these 10 sets of web services, we can find the best suit web service according to the customer criteria. We can find the best fit web service related to performance or cost. Even we can find a best suit service among performance as criteria with respect to response-time or with execution time etc.

III. Metrics

In comparison with other with collaborative filtering methods, we use root-mean-squared-error (RMSE) and mean absolute error (MAE) to measure our prediction quality. RMSE is defined as

$$RMSE = \sqrt{\frac{\sum_{ij} (Z_{ij} - \widehat{Z_{ij}})}{N}} \tag{8}$$

where Z_{ij} represents observed QoS of the web service j observed by the user i, $\widehat{Z_{ij}}$ is the predicted QoS value, N is the number of predicted values.

Mean absolute error is defined as

$$MAE = \frac{\sum_{ij} |Z_{ij} - \widehat{Z_{ij}}|}{N} \tag{9}$$

The mean absolute error is the average over the verification sample of the absolute values of the difference between a prediction result and the corresponding observations. The MAE is the linear score in which all the individual differences are weighted equally in the average.

TABLE 1
PERFORMANCE COMPARISION (a Smaller MAE or RMSE value Mean a Better Performance)

QoS properties	Methods	Matrix Density = 10%		Matrix Density = 20%	
		MAE	RMSE	MAE	RMSE
Response-time (0 – 20s)	NIMF	0.4854	1.2745	0.4357	1.1678
	CB-NIMF	0.2783	1.0296	0.2680	1.0485
Throughput (0 – 1000 kpbs)	NIMF	16.0542	45.9409	13.7099	41.1689
	CB-NIMF	13.2695	40.9390	10.5491	36.4855

IV. Conclusion And Future Work

Based on the past usage experience of other similar users we predicted the composite web service QoS. We propose a CB-NIMF approach for making personalized QoS value prediction. This approach fuses both the model-based and neighborhood-based collaborative filtering approaches for predicting higher accuracy. Our experimental analysis shows the effectiveness of our approach. Finally this CB-NIMF is extended for multi-criteria complex decision making over WS-QoS using AHP to find the best fit web service for the customer according to their requirements.

The CB-NIMF approach in this paper can be employed to client-side QoS properties as well as server-side QoS properties. We have conducted experimental studies on QoS properties like response-time, throughput, availability, execution time etc. Moreover, we planned to apply our approach to the cloud computing environment where the collection of QoS values becomes easier since the user's application invokes web services on the running cloud environment.

REFERENCES

[1] An approach for composite Qos parameter based Web Service selection. Rajeev Pratap Singh, K.K. Pattanaik, Atal Bihari Vajpayee – Indian Institute of Information Technology, Gwalior, India – Procedia Computer Science, Volume 19,September 2013.

[2] Qos – Aware Dynamic composition of Web Services using Numerical Temporal Planning. Guobing Zou, Qiang LU, Yixin Chen, Senior Member, IEEE, Ruoyng Huang, You Xu, and Yang Xiang. IEEE Transactions on Services Computing, September 2012.

[3] Qos – Aware Web Service Composition using GRASP with Path Relinking. Jose Antonio Parejo , Sergio Segura , Pablo Fernandez , Antonio Ruiz – Cortes. Experts Systems with Application , Volume 41, July 2014.

[4] Personalized Web Service Recommendation via Normal Recovery collaborative Filtering. Huifeng Sun ; Zibin Zheng ; Juliang Chen ; Lyu , M.R. Services Computing, IEEE transaction On volume 6 , 2013.

[5] Collaborative Web Service QoS Prediction via Neighborhood Integrated Matrix Factorization. Zibin Zheng ; Shenzhen Res. Inst., Chinese Univ. of Hong Kong, Shenzhen,China; Hao Ma ; Lyu, M.R. ; King, I. Sun Service Computing, IEEE Transactions 2013.

[6] Stochastic Evaluation of Qos in Service – Based system. Bruneo.D; Distefano.S; Longo.F; Scarpa.M; Parallel and Distributed Systems, IEEE Transactions 2013.

[7] Evaluating probablistic matrix Factorization on Data set. Ruslan Salakhutdinov & Andriy Mnih., Advances in neural information processing systems. pp. 1257 – 1267, 2007.

[8] Collaborative Web Service QoS Prediction via Neighborhood Integrated Matrix Factorization. Zibin.Zheng; Shenzhen Res.Inst.,Chinese Univ.of Hong Kong,Shenzhen,China; Hao Ma;Lyu,M.R.;King,I. IEEE Transactions September 2013.