CHAPTER - 5

PROPOSED QOS FORECASTING APPROACH FOR WEB SERVICE SELECTION

5.1 Preamble

Since the number of WS is increasing, QoS are used to describe the non-functional features of WS. Some quality properties of web service such as price, availability etc., do not depend on users and have the same values for diverse users. Normally, services providers or third-party registries (i.e. UDDI) offers the user independent QoS properties values. Also there are QoS properties such as response time, failure rate etc. which depend on users having diverse values for different users. Since the client side evaluation user-dependent QoS properties of WS are required in order to measure their performance, finding QoS values which are user dependent is a challenging task. Client-side WS evaluation requires real-world Web service invocations and encounters the following drawbacks:

- First, Web service invocations for the service users involve costs consuming the resources of the service providers. Few Web service invocations sometimes may be charged.
- Second, a large number of WS candidates have to be evaluated and there is a chance for some apt WS not discovered and involved by the service users in the evaluation list.
- Finally, mostly the service users may not be an expert to evaluate WS.

Hence, it is impossible to obtain the QoS properties values which are user-dependent without enough client-side evaluation. Therefore, it is difficult to get ideal Web service selection and recommendation.

In order to attack this challenge, in this research work a new approach for QoS forecasting is proposed. Most of the methods mentioned in the literature review are using Pearson-based similarity method. Even though a good prediction effect is provided by this method of finding similarity, it takes more run time and its performance is poor when the sparsity of the data set is high. Compared to the similarity based CF, Slope One method has shown a better predicting quality because of its high performance and simplicity. In this work, a new hybrid method for QoS forecasting is presented which combines Pearson and Slope One methods.
The experiments conducted reveals that the proposed hybrid method can achieve better forecasting accuracy when compared to the Slope One and Pearson based CF methods. The contributions of this work are as follows:

- A new Slope One method based prediction algorithm used along with Pearson similarity measurement is proposed which provides QoS values for Web service user.
- Weight adjustment and Statistically Process Control based smoothing strategies are employed to achieve better prediction accuracy.
- The performance analysis is performed on real-world dataset in order to prove the efficiency of the proposed method.

5.2 Proposed Framework for Forecasting QoS Values

It is obvious that the QoS data of services from different users can form a sparse matrix of service invocation records. So the prediction for a specific service’s QoS with respect to the current user is very necessary to assist the users so as to enable them to make a balanced decision with respect to selecting services. In this work, a hybrid forecasting method is proposed through systematically adopting the advantages both from Pearson correlation-based algorithm and Slope One algorithm.

5.2.1 The Overall Forecasting Framework

Let the total number of WS invoked by the active service user $u$ is denoted as IN. Usually for all $n$ service items i.e. WS in this work, the IN will be very low. So, in order to offer accurate QoS estimates for the rest of the service items with respect to the user ‘$u$’, other users’ invocation records should be fully considered for these services. Here, matrix M is considered as matrix consisting of the historic data about QoS with respect to ‘$m$’ number of users for ‘$n$’ number of service items. Similarly, the matrix consists of only partial QoS information for each service user. The density ‘$d$’ denotes the proportion of QoS data which exists in the matrix.

Based on the surveys made on CF techniques, it is found that the Slope One method suites well for the low density data, whereas Pearson centered CF achieves preferred forecasting results when the density data is high. Therefore, for adjusting the reference weight, in the proposed approach Slope One method is adopted for forecasting and Pearson correlation is
computed between services. A greater weight is allotted to the deviation of QoS between a service and the subject service for user ‘u’ when the relations are closer between those services.

Fig. 5.1 shows the methodology used in the proposed approach for WS QoS prediction. First, the historical QoS records of ‘n’ WS for ‘m’ users are collected and it is named as training data M. The training data is normally a sparse matrix as service users usually have only limited QoS records for all ‘n’ services. So, in order to provide useful information for QoS prediction, the matrix M is filled at the maximum. Secondly, a new Slope One algorithm using similarity values is proposed in order to forecast the null records in the training data set. With respect to Web service execution, there is a chance for training data to have certain unusual QoS records, particularly for the wide scale values of QoS attributes. To tackle this problem, finally a strategy called statistical process control [132] is used in order to regulate such exception data.

As a result, an enhanced training dataset is generated which has a promoted data density. Finally, also the proposed algorithm is used for forecasting QoS values for active user using the new training dataset and error analysis is used for measuring the quality of the prediction.

5.2.2 Proposed Forecasting Method
With respect to QoS forecasting framework, the proposed forecasting algorithm and statistical process control-based adjustment strategy plays significant roles to improve the precision and they are discussed in detail below.

Forecasting Algorithm
As mentioned earlier, Slope One based CF method shows better quality when the sparsity of data is high. As active users usually have only IN QoS records which are normally less than ‘n’ for ’n’ WS, the QoS values for active user is forecasted using item-oriented Slope One method. Normally, the basic Slope One forecasting method does not consider the similarity between items. In this proposed work, a new QoS forecasting algorithm using Slope One method considering the similarities between two items is introduced. That is, the service having the higher similarity value should be set the higher priority when the deviation is considered in Slope One method. Item-based Pearson correlation is used for measuring the similarity between two WS. For any two services ‘i’ and ‘j’, the similarities between ‘i’ and ‘j’ is calculated using the Eqn. 5.1.
where \( U = U_i \cap U_j \) is the set of users having QoS records for both service ‘i’ and service ‘j’ and \( \overline{r}_j \) denotes the average QoS value of service ‘i’ witnessed by other users.

The similarities between service j and other services invoked by user ‘u’, that is \( I_u - \{j\} \) are measured and this is used for forecasting a missing value \( r_{u,i} \) in the user-item matrix. The remaining items after removing the services with undesirable similarity to service ‘j’ are considered as the related services of ‘j’ with respect to user ‘u’, denoted as \( R(j|u) \). It is denoted as Eqn. 5.2.

\[
R(j|u) = \{i \mid i \in I_u, Sim(i,j) > 0, i \neq j\} \tag{5.2}
\]

Then, the forecasting based on similarity based Slope One algorithm is derived using the Eqn. 5.3.

\[
P\left(r_{u,j}\right) = \frac{1}{\text{card}(R(j|u))} + \sum_{l \in R(j|u)}(w_{i,l} \cdot dev_{j,l} + r_{u,l}) \tag{5.3}
\]
where \( w_{t,j} \) is an adjustment weight related to the similarity values between ‘\( j \)’ and another service ‘\( i \)’ (\( i \in R(j \mid u) \)) and it is computed using the Eqn. 5.4.

\[
\begin{align*}
    w_{t,j} &= \frac{\text{Sim}^\lambda (i,j)}{\sum_{k \in R(j \mid u)} \text{Sim}^\lambda (k,j)} \\
    \text{(5.4)}
\end{align*}
\]

where \( \lambda \) (=1, 2 or 3) is adjustment strength factor where higher value denotes a stronger adjustment. Meanwhile, \( \text{Sim}^\lambda (k,j) \) is \( \lambda \) power of \( \text{Sim}(k,j) \), i.e. \([\text{Sim}(k,j)]^\lambda\]. In a nutshell, the proposed algorithm is employed both in finding the missing values and for forecasting QoS values for an active user.

**Proposed Smoothing Strategy using Statistical Process Control**

The matrix consisting of original QoS values of services is denoted as \( M \) and the matrix along with forecasted missing values is denoted as \( M' \). There may be some exceptional QoS records for WS existing in \( M \) showing that the QoS value of a specific user is far away from the records of neighbor users and also the sparsity is high for the matric \( M \). As a result, there are chances for the filled matrix \( M' \) to have QoS items which are far from the common situation and may have bad impact on the forecasting stage. Thus, at first they are recognized from matrix \( M' \), and then they are smoothened adopting a heuristic strategy.

So, in this work the idea from statistical process control is utilized to overcome this unusual QoS data in \( M' \). Statistical process control is used in the industrial production process for real time monitoring using statistical analysis method. It is used to scientifically distinguish the exceptional variation from the normal random variation to provide early warning for production process to manager.

This technique is adopted in this work to pick out the abnormal QoS values so as to achieve better forecasting accuracy. Firstly, for matrix \( M' \), item \( r_{u,i} \) i.e., the QoS of service \( i \) for user \( u \) is checked for its exceptionality using the rule 5.5.

\[
\text{isAbn}(r_{u,i}) = \begin{cases} 
    \text{true}, & \mu_i - \Theta \cdot \sigma_i < r_{u,i} < \mu_i + \Theta \cdot \sigma_i \\
    \text{false}, & \text{otherwise}
\end{cases}
\]

where \( \mu_i \) is the average QoS value of service \( i \) (\( 1 \leq i \leq n \)), and \( \sigma_i \) is the standard deviation of service \( i \)'s QoS records from different users. \( \Theta \) is a positive integer to regulate the normal range of QoS value and it is normally assigned a value of 3 in most applications of statistical process control.
When an abnormal QoS is detected through the above approach, this isolated item is smoothed before the forecasting step. Here, an approach called “small amplitude shift” is proposed for smooth out treatment. Suppose \( r_{u,i} \) is an abnormal issue according to the rule 5.5. Then the smoothing action is performed using the Eqn. 5.6 and the adjusted value is denoted as \( \bar{r}_{u,i} \).

\[
\bar{r}_{u,i} = \mu_i - \Theta \cdot \sigma_i, \quad r_{u,i} < \mu_i + \Theta \cdot \sigma_i \\
\mu_i + \Theta \cdot \sigma_i, \quad r_{u,i} > \mu_i + \Theta \cdot \sigma_i \\
r_{u,i}, \quad otherwise
\] (5.6)

Here, the lower or upper boundaries are used to substitute the unusually low or high QoS records respectively.

### 5.3 Experiments and Evaluation

In order to validate the effectiveness of the proposed algorithm for QoS prediction, experiments are conducted on a publicly available dataset, which is released by [39] and has been widely adopted in the current researches. The data set has 5825 service invocation records about 339 users, and QoS attributes include response time and throughput.

The experiments are conducted on a partial QoS records consisting of 100 WS and 150 users from the original data. Then, this new dataset is randomly divided into two parts such as training dataset and test dataset. The training dataset includes 100 users and the remaining 50 users are treated as test or active users. In order to satisfy the real condition, two new sparse datasets are constructed after removing some records from training data matrix whose data densities are set as 10% and 15% respectively as normally the recorded QoS values for a user occupy only a few part of all 100 services.

Similarly, for active users only IN (=5, 10 or 20) QoS records are considered for every user. The records removed from test data set are treated as real data for evaluating the forecasting quality. The main parameters considered in the experiments are:

- ‘\( m \)’ which is the total number of users (150 is considered in this work where 100 users are considered as training users and the rest 50 users are considered as test users)
- ‘\( n \)’ which is the total number of services (100 is the value considered in this work)
• ‘d’ is the density of data in training matrix
• ‘IN’ is the total number of known QoS records for every active user (5,10 & 20 are the values considered in this work)
• ‘ߣ’ is the adjustment strength factor (values 1,2 and 3 is considered in this work)

Evaluation Criteria
Generally RS adopts Mean Absolute Error (MAE) method to evaluate the forecasting result. It is defined as the average of difference values between the forecasted QoS and the real record which is computed by Eqn. 5.7.

\[
MAE = \frac{1}{TN} \sum_{u,s} \left| P(r_{u,s}) - r_{u,s} \right|
\]  

(5.7)

where \( P(r_{u,s}) \) is the forecasted QoS value of service \( s \) witnessed by user \( u \), \( r_{u,s} \) is the actual QoS value of services w.r.t. user \( u \), and TN is the total number of forecasts. As the range of service’s QoS value may vary from each other, MAE is not neutral enough to return the accuracy of forecasting algorithm. Hence, the Normalized MAE (NMAE) is adopted as a metric to compare the forecasting quality of three algorithms and it is computed as Eqn.5.8. (smaller the NMAE value, more accurate forecasting)

\[
NMAE = \frac{MAE}{\sum_{u,s} P(r_{u,s})} \frac{1}{TN}
\]  

(5.8)

Evaluation of Results
WSRec [39] and basic Slope One algorithm [17] are considered as baseline in this for comparison purpose and they are implemented in the experiments. All three algorithms are coded using java language under Windows platform and tested on the same dataset as described earlier. The settings for WSRec algorithm is the same as mentioned in [39]. The three algorithms are compared for the QoS attributes response time and throughout.

The experiment is repeated for 50 times for each case of density and IN value and the average NMAE metrics are recorded. The observed results (i.e. NMAEs) on QoS attribute response time are shown in Fig. 5.2. It is clear that the accuracy of the forecasting by the proposed algorithm outperforms WsRec and basic SlopeOne algorithm in many cases. For the values (i.e. 10% and 15%) of density, the proposed algorithm (ߣ =3) achieves minimum NMAE
values for almost all cases. To conclude, the forecasting accuracy of the proposed algorithm is better than WsRec and basic Slope One methods for almost all situations.

Also the NMAE values of three algorithms on QoS attribute throughput are shown in Fig. 5.3. It is clear to see that the proposed algorithm ($\lambda = 3$) has improvement compared to both WsRec and basic Slope One methods. It is clearly seen that the predication error of proposed method is reduced when the $\lambda$ value is increased.

Based on the experimental results, it is concluded that the proposed forecasting algorithm is a better choice compared to WsRec and basic Slope One methods for forecasting QoS values for services particularly when the user service record matrix is sparse.

As mentioned earlier, the Slope One-based method is generally used when the training data matrix is having high sparsity. However, Pearson correlation based method shows its merit when the data density is high. As discussed earlier, the missing data in training matrix is filled by the proposed algorithm. The influence of missing value supplement with two types of approaches is discussed below.

![Figure 5.2: NMAE values for Slope One, WsRec and proposed algorithms for the response-time attribute (d= 10%)](image-url)
Figure 5.3: NMAE values for Slope One, WsRec and proposed algorithms for the response-time attribute ($d = 15\%$)

Figure 5.4: NMAE values for Slope One, WsRec and proposed algorithms for the throughput attribute ($d = 10\%$)
Figure 5.5: NMAE values for Slope One, WsRec and proposed algorithms for the throughput attribute ($d = 15\%$)

The density of training matrix during the process of missing value addition is denoted as $d^0$ where $d^0 > d$, and the borderline to consider filling approach is indicated as $\rho$. Then a two-stage filling approach is recommended for filling the missing data when the matrix is comparatively sparse (i.e. $d^0 \leq \rho$) and when the density of the training matrix reaches a certain degree (i.e. $d^0 > \rho$), Pearson method is employed for forecasting the missing values.

For evaluating the result of the proposed two-stage filling design, different training matrixes are formed for various values of $\rho$ for the proposed forecasting algorithm. The observed results are shown in Figs. 5.6, 5.7, 5.8 and 5.9. It is finally witnessed that the proposed two stage filling method has some perfection for some conditions, but it is not so noticeable. When the density is 10%, the fluctuating tendencies of forecasting error for two QoS attributes are highly consistent. From the overall point of view, NMAE basically decreases along with the growth of $\rho$’s value. However, NMAE has a little drop at the point of $\rho=0.6$. As a consequence, the best boundary point for this case is 0.6. While considering when density=15%, the variation tendency of forecasting error for attribute response-time is very similar to the first case of this attribute;
just the current variation is too low. There is an exceptional case for attribute throughput when \( \rho = 0.2 \), the corresponding NMAE value is suddenly low. Meanwhile, the prediction error value has a relatively high value at point \( \rho = 0.3 \). Subsequently, it has a small reduction at first and then gradually takes off. The optimal boundary point in this case can be considered as 0.5 for most situations.

Figure 5.6: NMAE values for response-time attribute for n boundary values (\( \rho \)) and density =10%  

Figure 5.7: NMAE values for response-time attribute for n boundary values (\( \rho \)) and density =15%
On the whole, the two-stage filling strategy proposed in this work has a small improvement with respect to prediction error. Considering the selection of boundary point, the value in the domain from 0.5 to 0.6 is worth considering in practice.
5.4 Chapter Summary

In this research work, a new hybrid approach is proposed which combines the Pearson similarity between WS into Slope One CF for solving QoS forecasting issue. In spite of giving similar weights for every WS, Pearson method is used as a weight in order to regulate for distinguishing the deviation among WS. Also a statistical process control-based smooth out method is presented for adjusting the exceptional data for improving the forecasting accuracy. In the practical sides, the basic Slope One and the WsRec algorithms are also executed along with the proposed approach for evaluation purpose. The evaluation is conducted on the publicly available dataset and the experimental results show that the proposed hybrid algorithm produces better accuracy with respect to forecast than the other two methods. The statistical process control-based smooth out method efficiently handles the noise data by reducing the forecasting inaccuracy. Additionally, a two-stage filling method is proposed, and the suitable borderline point for altering filling methods is also recommended in this work.

After the unknown QoS values are forecasted, the WS are ranked using multiple QoS values. Next chapter proposes new approaches for rank aggregation and vector-based ranking of WS considering users’ requirements.