Economic Computation and Economic Cybernetics Studies and Research, Issue 4/2019; Vol. 53

Assistant Professor Kanagaraj VAITHEKI
E-mail: vaithekiphd2016@gmail.com
Department of Computer Science
Pondicherry University, Puducherry, India
Associate Professor Suresh JOSEPH
Department of Computer Science
Pondicherry University, Puducherry, India

HYBRID HAVERSINE-GENETIC GEOGRAPHICAL MODELLING FOR AN OPTIMAL WEB SERVICE SELECTION/RECOMMENDATION

Abstract. *Web Service selection and recommendation becomes major aspects of service providing the community. Quality of Service prediction through Collaborative Filtering (CF) or Matrix Factorization (MF) is the key process in WS recommendation. Few or absence of QoS records (cold start), integration of QoS factors (response time and throughput), extensible models utilization and inaccurate similarity measures are the major problems in CF/MF approaches. This paper focuses on integration of similarity values with the geographical distance and QoS values (minimum response time and maximum throughput) to provide efficient mapping relationship (user-service and service-user). Initially, the geographical model is constructed by Haversine distance formulation based on latitude/longitude values. The Genetic Algorithm (GA) based recommendation list construction provides best WS as the responses to user query. The experimental validation of proposed work regarding response time, memory space, computation time, precision, coverage and number of patterns mined against existing methods assure effectiveness in QoS prediction.*

Key words*: Collaborative Filtering, Euclidean Distance, Matrix Factorization, Response Time, Throughput, Quality of Services, Web Service Selection & Recommendation.*

JEL Classification: O30

1. Introduction

Web Services (WS) assembly shifts the uniform applications to dynamic business set-up process that is necessary for multi-organization aspects. The software development based on the assembling process is referred as Service Oriented Computing (SOC) that has the specific characteristics such as coupling-free architecture between user and the service provider (WS ability of hiding

DOI: 10.24818/18423264/53.4.19.14

implementation details), dynamic binding (WS invoking with respect to the applications flexibility), and the composability (creation of complex services). The abundant and a large number of users participation on the social networks introduces the additional information (feedback, interaction histories, and social relationships) with the functional/non-functional properties description. The inclusion of information in WS regarding the non-functional properties termed as Quality-of-Service (QoS) values and they regarded as the factor for suitable service selection. The property declaration in QoS (response time and throughput) depends on the two contexts such as user and service. The QoS evaluation by the single user is not suitable for the other users during the service selection process. The time and resource consuming nature of WS revocation make the acquiring QoS information as an impractical one.

The QoS measurement contains two stages such as either in server or client side. The consistent server-side measurement uncovers the shared feature of WS and the client-side measurement is difficult due to the diverse locations. The real-world WS evaluation in client-side is the critical task because of the excessive time, memory and resources consumption. The overloading effect of WS causes the limitation in right WS locating with the right time. The provision of intelligent/proactive WS to the users requires the combination of QoS attributes, preferences and user needs. The technique provides the efficient combination is referred as Collaborative Filtering (CF) and it also supported the WS recommendation with the automatic QoS attributes prediction. The representation of the user-service provider relationship in CF is the user-item matrix. The element in the user-item matrix represents the QoS factor of the user for the respective service without concerning the service provider information. Also, there are some missing QoS entries in the user-item matrix affects prediction performance of the relationship between the WS and providers. The recommendation of consumers for providers and vice versa initiates the research into the field of bidirectional WS recommendation with hybrid CF algorithms. The non-recommendation state of WS without required user interactions is termed as cold start problem. The CF-based QoS prediction is not in an easier way due to the following reasons diversity in user and QoS values, a limited number of service invocation records and more time / resources consumption for service invoking the process. The evolution of Matrix Factorization (MF) model-based CF improves the QoS prediction with the historical invocation records-based training. The user-perceived QoS determination with the few latent features affect the both user and service sides. Even though the MF-based methods achieved the good performance in QoS prediction, the entries in the matrix based on the user ratings of the respective services. The subjective nature of the element in the user-item matrix reflects the user preference on the item. The user-observed QoS values are affected by the factor called location. When the users and services are in close location, then the response time is small. Hierarchical Matrix Factorization (HMF), Extended Matrix Factorization (EMF) and the Integrated Matrix Factorization

DOI: 10.24818/18423264/53.4.19.14

(IMF) approaches supports the global matrix construction with users and services, missing QoS values prediction and real-time sharing of QoS selection. The problems observed in the traditional CF and MF models are non-reliability due to the QoS changes, sparse QoS nature, high computational cost and the limitation in relationship prediction between similarity value and geographical value. This paper discusses the selection based on similarity measurement, geographical value response time and throughput for WS selection. The technical contributions of proposed work are listed as follows:

- Locational information (latitude and longitude)-based Euclidean formulation computes the nearest neighbors supports the user-service interrelationship efficiently.
- Service distance formulation using the minimum response time and maximum throughput support the efficient QoS selection to the specified users
- Preprocessing prior to Genetic Algorithm (GA) reduces the number of computations that reduces the time consumption.
- GA application to the preprocessed dataset (distance, response time and throughput) selects the optimal best WS to the user.

This paper is organized as follows: Section II describes the related works on web service in section II. Section III discusses the proposed Genetic Algorithm-based Web Service Selection. Section IV presents the performance analysis of GA-WSSR regarding the security, execution time and storage complexity. Finally, section V presents the conclusion.

2. Related Work

Kayastha and Baria (2015)surveyed the various QoS-based WS recommendation with the consideration of non-functional properties namely, response time, reliability and the availability. The prediction of user's navigation is the major objective of WS recommendation. Research studies focused on the development of WS recommendation algorithm (prefixspan) with the less memory consumption. Suneetha and Rani (2012)modified the prefixspan algorithm that incorporates the various constraints namely spending time and recent visiting. The weighted sequential patterns utilization in the extension of prefixspan algorithm efficiently constructed the recommendation model that consumed less memory. The page importance based on visiting frequency and the time spent is not included in the recommendation set. Wanaskar, et al. (2013)used the web usage login on adaptive association rule mining basis and presented the effective text mining technique. The addition of semantic knowledge to the improved algorithm increased the efficiency and the quality. The memory-based algorithms (rule mining) were suitable for minimum user ratings. But, the computation cost is increased with increase in a large number of users. Moreover,

DOI: 10.24818/18423264/53.4.19.14

an inappropriate WS selection leads to problems in the real-time applications. Chen, et al. (2014)employed the location information and QoS values to the user-service matrix to improve the recommendation accuracy. The atomic service composition and the added value logic exposure initiated the QoS-aware WS selection schemes. Parejo, et al. (2014) proposed the hybrid methods namely Greedy Randomized Adaptive Search Procedure (GRASP) and Path Relinking (PR) for minimum time consumption and maximum availability.

The quality of the recommendation system was decided by the user feedback. The accuracy of the results dependent upon the personal interest on the viewing content. The undesirable nature of user feedback due to the loading speed and the quality required the analysis of the influence of QoS. De Pessemier, et al. (2012)investigated the audio-visual quality parameters impact on the user feedback and it can be used as the additional feedback filter for better video recommendation performance. The accuracy improvement in WS composition dependent on the disparate functionalities with varying QoS. But, the increase in similar functionalities leads to the difficulty in optimal WS selection. Shehu, et al. (2014)discussed the WS composition concepts and reviewed the several field publications. The heterogeneous nature of future applications and the large size information maintenance from the convergence of the devices required the trade-off between the security and QoS. Nieto and Lopez (2014)discussed the impact of parameter-relationship increase on the security-QoS tradeoff assessment. They employed the Context-based Parametric Relationship Models (CPRM) for dynamic inputs generation with context basis. A large number of WS deployment for large size applications triggered the user-feedback-based service recommendation system. Several challenges such as sparsity, trustworthiness, and the cold-start problem were observed in similarity approaches-based service recommendation models. Deng, et al. (2015)discussed the effects of incorporation of a trust relationship in the social network with the user feedback. The model with the trust enhancement-based service recommendation scheme is referred as Relevance Trust Walker (RTW). Based on the functional requirements, the optimal WS selection is the challenging issue in the research studies. Gupta, et al. (2015)considered the trust rate parameter for each service provider that was selected by the binary search algorithm. The searching of optimal WS consumed more time in traditional methods. But, the inclusion of trust rate parameter reduced the time consumption efficiently.

The recommendation process was difficult due to a large number of published web services available for real-time applications. Kuang, et al. (2012)proposed the new method called Context-Aware Services Recommendations (CASR) with reference to the previous service invocation schemes to support the WS recommendation system. The service invocation records were clustered initially and then the proposed work selected the cluster with the most similar context. When CASR model extended to predict the user choices on the large size music contents, then the exploration of more

DOI: 10.24818/18423264/53.4.19.14

contexts was the difficult process for high-recommendation performance. Deng, et al. (2014)utilized the user's microblogs for emotions extraction under different granularity levels and the time windows. The correlation of user, music and user's emotion supported the music recommendation efficiently. The CASR models based on the assumption of identical users participation in QoS evaluation and that was not possible. The ignorance of diversity during the recommendation stage caused the limitations in optimal WS selection. Yu (2014)discussed the diversity inclusion instead of uniform QoS application in WS selection. The Relational Clustering-based Model (RCM)-based QoS evaluation scheme addressed the scarcity issues. The unpredictability internet environment and the sparse nature of QoS caused the challenges in accurate personalized QoS prediction. Zhu, et al. (2012)predicted the QoS values based on the landmark and clustering algorithms (User-Based Clustering (UBC) and Web Service-Based Clustering (WSBC)) for QoS prediction accuracy improvement. The WS invocations of clustering algorithms were expensive and time-consuming. The search of top-k similar services in QoS prediction did not consider the unbalanced data distribution. Xiong, et al. (2014)proposed the collaborative approach to the QoS prediction under unbalanced data distribution. They provided the sampling/resampling techniques on similar neighbor selection for the reduction of time consumption.

3. Genetic Algorithm-Based QoS Prediction for Optimal Web Selection

This section presents the detailed description regarding the implementation of proposed Genetic Algorithm (GA) based initial service distance formulation. Figure 2 shows the flow of proposed method. Initially, the latitude and longitude information about the user is extracted for the distance estimation. Initially, the algorithm collects the latitude and longitude information about the user for distance estimation. The Haversine formula utilization directly estimates the minimum distance for the users. Then, an average distance for the overall users is calculated. Based on the average distance value, the user datasets are clustered through agglomerative manner to extract the most similar one. Then, by using the average distance estimation for each cluster forms the new cluster with minimum distance value. The cluster with the minimum distance value merged with the available services having a minimum response time and maximum throughput. The simultaneous extraction of services is mathematically interpolated by the GA–based optimal service distance formulation. The output from the sequential iterations in GA depicts the recommended service for minimum distance user. The prior distance estimation and the GA based service distance formulations avoid the unnecessary computations that lead to are duction in memory and computation time. The optimization provided in the computational steps and the similar cluster formation in proposed work increases the coverage space and the precision values. The proposed flow contains the following operational stages:

DOI: 10.24818/18423264/53.4.19.14

1. Haversine geographical modeling
2. Agglomerative Clustering
3. Merging
4. GA-based WS list formation

Table I lists the symbols and their descriptions used in the proposed work.

Table 1. Symbols and Description

Symbol	Description
∂	Distance
d_{lt}	Latitude deviation from requested value
d_{lg}	Longitude deviation from requested value
α_{u_i}	Latitude information of user
α_{req}	Requested latitude information
Ud_i	User distance
R	Radius of the earth
W_d	User attributes with the minimum distance
Sd_j	Service distance
W_m	Merged dataset
$F(Sd_i)$	Fitness function
W_{th}	Dataset with throughput
W_t	Dataset with response time values

DOI: 10.24818/18423264/53.4.19.14

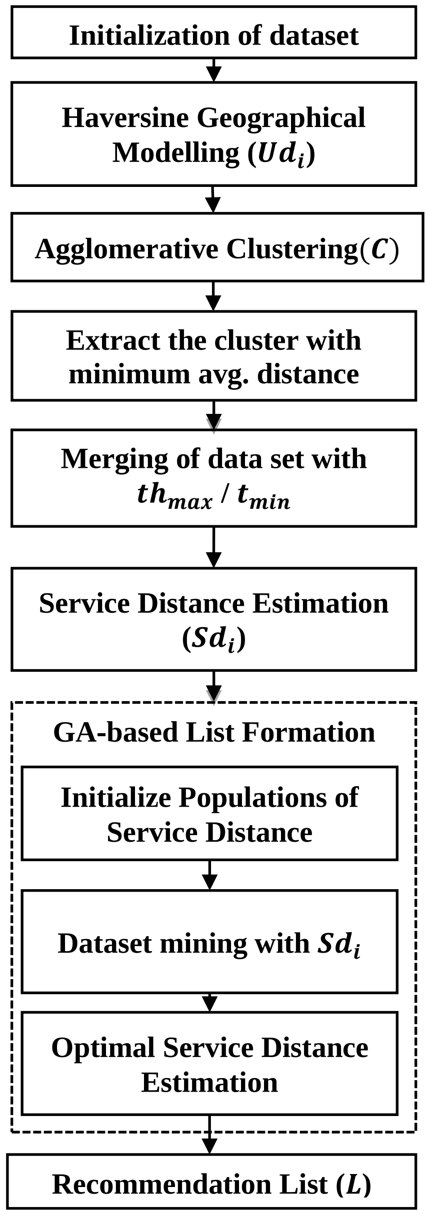


Figure 1. Flow diagram of proposed GA-WSSR

DOI: 10.24818/18423264/53.4.19.14

The proposed GA-based WSSR algorithm is described as follows:

GA-WSSR Algorithm

Input: *Fitness percentage f_p, user latitude and longitude values (α and β)*

Output: *Recommended Service List ℒ*

//Initialization

Initialize four dataset (W_u, W_s, W_t and W_r) and load into database

For *each user (u_i)*

$\alpha_{u_i} = \alpha\ of\ user\ u_i$ *//Latitude of each user*

$\beta_{u_i} = \beta\ of\ user\ u_i$ *//Longitude of each user*

End For

//Distance Formulation

For *each user(u_i)*

$$d_{lt} = \left|\alpha_{u_i} - \alpha_{req}\right|$$
$$d_{lg} = \left|\beta_{u_i} - \beta_{req}\right|$$
$$\partial = sin^2 {d_{lt}}/_2 + cos\, \alpha_{u_i} * cos\, \alpha_{req} * sin^2 {d_{lg}}/_2$$
$$Ud_i = 2 * R * atan2(\sqrt{\partial})$$

End For

$d_u = {1}/_{k} \sum_{i=0}^{k} Ud_i$ *// Average distance*

// Form the cluster (C)

For $i = 1 \dots n$

$$c_i = \{d_u\}$$

End For

$$C = \{c_1, c_2 \dots\dots c_n\}$$

// Agglomerative Cluster Formation

Form the clusters with the minimum distance measures

Set $t = n + 1$

While *($sizeof(C)$>1)* ***do***

$Cd_u = c_i - c_{i+1}$ *// Distance measurement*$min(mCd_i, mCd_{i+1}) = min\, Cd_i\ \forall c_i\ in\ C$

// Minimum distance computation within clusters

Remove mCd_i, mCd_{i+1} *from* C

// Remove the minimum values from the clusters

Add $\{mCd_i, mCd_{i+1}\}$ *to* C

// Include the set with min distance to the clusters

$$t = t - 1$$

End While

$$C = \{C_1, C_2 \dots\dots, C_n\}$$

DOI: 10.24818/18423264/53.4.19.14

$$C_i = \{d_i\}$$

***For** each cluster C_i*

$$aC_i = {}^{1}\!/_{n} \sum_{i=0}^{n} d_i$$

End For

$$M_c = min(aC_i)$$

// Minimum Clusters having minimum distance

W_d=user attributes from M_c

//Merging

$k_n = no.of\ user\ count\ in\ W_d$ *//Nearest neighbors*

$U_i.size() = k_n$*// List formation according to neighbors*

$$W_{temp} = W_d \cup W_r \cup W_t$$

// merging of data set with distance, response time and throughput

$< U_{it} >= k_n\ of W_{temp}$ *//Merging List formation*

*// **Dataset prediction with maximum throughput (th_{max})/Minimum response time (t_{min})***

***For** each service (S_i)*

$$t_{min} = min(r_i)$$
$$W_{l1} \leftarrow (W_{temp})_{t_{min}}$$
$$th_{max} = max(th_i)$$
$$W_{l2} \leftarrow (W_{temp})_{th_{max}}$$

$W_{l1} \cup W_{l2} \Rightarrow W_m$ *//Merging of datasets with th_{max} and t_{min}*

$$(Sd_i)_{W_m} = \sqrt{\left(t_{ui} - t_{avg}\right)^2 + (th_{ui} - th_{avg})^2}$$

// Service distance estimation

//Genetic Algorithm

Initialize population distance (Sd_i) from $W_m$dataset

Get fitness percentage f_p

$f = 1 - (f_p * 0.001)$*/Fitness value initialization*

***For** ($int\ i = 0; i < f; i++$)*

Mined Sd_idataset from $W_m$dataset //Selection

Compute average Sd_i from mined $W_m$dataset //crossover

$$Sd_i = (Sd_i + Sd_{avg})/2$$

$\mathcal{L} \leftarrow (W_m)_{Sd_i} \cup W_s$ *// Recommendation List*

End For

DOI: 10.24818/18423264/53.4.19.14

3.1 Haversine Geographical Modelling

The mobile user's location prediction is the initial stage of service recommendation. The increase in availability of users and services may cause a number of computational steps to select the optimal user with minimum distance value. The location prediction normally depends on the transmission/reception of signals that initiates positioning technology with certain considerations. Global Positioning System (GPS) is one of the approaches to locating the user. The GPS data contain the latitude and longitude information about the user. Hence, the reverse geocoding is applied to convert the GPS data into the human-readable form. The coordinates of GPS data available in degrees, minutes and seconds. The general formulation of latitude is expressed as

$$Latitude(in\ decimal\ degs) = degrees + \left(\frac{minutes}{60}\right) + \left(\frac{seconds}{3600}\right) \quad (1)$$

With the declaration of minimum/maximum longitude bound values (-5.004101, -4.952431) and the wide value of map (500 px), the coefficients (a, b) are computed to state the longitude in decimal value as follows:

$$\left.\begin{aligned} -5.004101a + b &= 0 \\ -4.952431a + b &= 500 \end{aligned}\right\} \quad (2)$$

The solutions of equation (2) describe the coefficient values as $a = 9676.9$ and $b = 48423.66$ that are helpful to predict the longitude of the particular point by using the following equation

$$x = 9676.9 * longitude + 48423.6 \quad (3)$$

The major assumption for the Haversine implementation is that the earth is in spherical shape rather than ellipsoidal. The mathematical formula for Haversine distance estimation is expressed as

$$\partial = \sin^2 {}^{d_{lt}}/_{2} + \cos\alpha_{u_i} * \cos\alpha_{req} * \sin^2 {}^{d_{lg}}/_{2} \quad (4)$$

Where, ∂ – distance

d_{lt} – Latitude deviation from requested value

d_{lg} – Longitude deviation from requested value

α_{u_i} – Latitude information of user

α_{req} – Requested latitude information

The user distance estimation from the coordinate value (4) is mathematically expressed as

$$Ud_i = 2 * R * \text{atan}^2(\sqrt{\partial}) \quad (5)$$

Where, R – Radius of the earth – 3961 miles or 6371 km.

The users with the minimum average distance are extracted and form the agglomerative clusters for respective services.

DOI: 10.24818/18423264/53.4.19.14

3.2 Agglomerative Clustering

The mathematical tool to discover specific structures from the dataset is termed as clustering. The most widely used clustering algorithm is agglomerative clustering and it has more complexity for large size dataset. The selection of point and merging of the two clusters with the smallest inter-cluster distance until the clustering is satisfied are the iterative processes in the agglomerative clustering. The hierarchical clusters formed by the agglomerative clustering contains the following properties:

- Nested formation of clusters generated in earlier stages with the later stages
- Clusters are valuable for discovery

The algorithmic steps for agglomerative clustering are listed as follows:

1. Assign object to each cluster
2. Evaluate the distance between the clusters
3. Construct the distance matrix
4. Search the pair of clusters with the shortest distance
5. Remove the pairs from the distance matrix
6. Compute the new distance of formed new clusters with other clusters
7. Repeat from step 1 until the distance matrix contains single element

In this paper, the set of users is initialized as $(u_1, u_2, \ldots, u_n)$ and the Haversine geographical modelling provides the distance between the nearest clusters. Then, the average distance is computed for each cluster and the minimum distance clusters are extracted. Based on the distance measurement process, the diversities in agglomerative clustering are single linkage (shortest distance between the two cluster objects), complete linkage (longest distance between the cluster objects), and average linkage (average distance between the cluster objects. This paper utilizes the single linkage property to create the cluster with minimum distance.

3.3 Merging

The cluster formed from the agglomerative process contains the user attributes (user ID, IP address, and country/ latitude information) with minimum distance. The group formation with minimum distance, minimum response time and maximum throughput is required for best service recommendation and such integration process refers merging. There are two major processes in the merging implementation.

1. Based on nearest neighbor analysis, merging list is formed
2. The integrated dataset is formed with minimum response time and throughput.

The user attributes in agglomerative clusters are arranged as dataset (W_d). For each attribute, the corresponding services are extracted with the constraints minimum response time and maximum through put. The service dataset includes only the response time and the throughput values. Hence, the service having the minimum response time and maximum throughput is integrated to the user attributes for best service selection. The mathematical formulation of best service selection is modeled as

DOI: 10.24818/18423264/53.4.19.14

service distance Sd_i estimation that contains Euclidean distance between the response time with the average response time and throughput with the average throughput as follows:

$$(Sd_i)_{W_m} = \sqrt{\left(t_i - t_{avg}\right)^2 + (th_{ui} - th_{avg})^2} \tag{6}$$

Where, t_i – current response time

t_{avg}- Average response time

th_{ui}– Throughput for current service

th_{avg}– Average throughput

The computed service distance metrics for all the services considered as the initial population for Genetic Algorithm (GA) to create the list of recommended web services.

3.4 Genetic Algorithm-based WS list formation

The search algorithm that has the capability to discover the good solutions for large and complex search space is referred as Genetic Algorithm (GA). The chromosome initialization, fitness function selection, and the genetic operations are the sub-processes for best WS selection. The encoding of the problem into the genome structure is the prior stage in GA. The integer array containing the abstract services represents the genome structure. The each object from the merged dataset contains the index to the concrete services matched to the abstract services. The GA implementation depends on the fitness function formulation with the specific constraints. The fitness function formulation depends on the reduction of response time and maximum throughput. The constraint for the fitness function is

$$Cl(Sd_i) \leq 0, \quad i = 1, \dots, n \tag{7}$$

The distance required to satisfy the constraint is expressed as

$$D(Sd_i) = \sum_{i=1}^{n} Cl(Sd_i) y_i \tag{8}$$

Where, y_i- 0 for $Cl(Sd_i)) \leq 0$

y_i- 1 for $Cl(Sd_i)) > 0$

The fitness function formulation for the genome is expressed as

$$F(Sd_i) = \frac{w1(Response\ time\ (t_i(Sd_i))}{w2(Throughput\ (th_i(Sd_i))} + w3(D(Sd_i)) \tag{9}$$

Where, $w1, w2$ and $w3$ – real positive weight of different fitness factors. The stopping criteria for this GA-based WS recommendation completes only if all the populations should satisfy the fitness computation. The quality of each chromosome depends on the fitness measure. The reproduction of chromosome through the following genetic operations namely, selection, mutation and cross over. The implementation steps of GA based WS selection and recommendation are listed as follows:

Step 1: Initialize the genetic parameters (fitness function)

Step 2: Generate the initial population

Step 3: While (fitness function satisfaction)

DOI: 10.24818/18423264/53.4.19.14

Compute fitness for each service distance
Rank individuals and assign their fitness
Randomly select the individuals based on the selection
Probability
Adopt the crossover and mutation operators to form the
New population
Collect the pop size/2 individuals and the pop size/2
Individuals in the temporary population to form new data
Step 4: Select the best individual in current population

The best individuals from the GA are the required web services recommended to the user. The best service selection and the prior cluster formation with the minimum distance reduce the computational burden and memory consumption.

4. Performance Analysis

This section discusses the effectiveness of proposed Genetic Algorithm-based Web Service Selection and Recommendation (GA-WSSR) by using the validation of Root Mean Square Error (RMSE), computational time, memory consumption, precision, coverage and number of mined patterns against distance values, no. of mined patterns, support values respectively. The experimental analysis shows that the proposed GA-WSSR assures the effectiveness in QoS prediction.

4.1 Dataset

Two real data sets used to validate the proposed GA-WSSR that contains 339 users and 5825 services. The dataset includes the affiliation records of the services and the geographical information of the users. The response time records of the followed by the service invocation are arranged in one dataset and the throughput records are arranged in another dataset. The random selection of large size QoS values forms the testing dataset and the remaining forms the training dataset. E.g. 95 % are testing and the remaining 5% are the training dataset.

4.2 RMSE

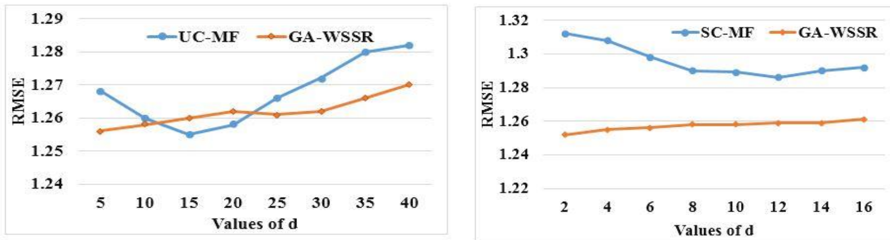
The squared difference estimation between the actual (qij) and predicted class ($\widetilde{q_{ij}}$) over the testing data size (ρ_T) is referred as Root Mead Square Error (RMSE). The mathematical formulation is expressed as

$$RMSE = \sqrt{\frac{1}{\rho_T}(q_{ij} - \widetilde{q_{ij}})^2} \tag{10}$$

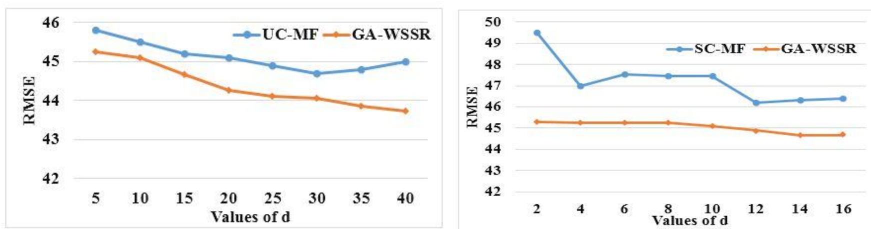
Figure 2 shows the experimental variations of $RMSE$ for the various latent factors. In response time dataset, the RMSE analysis for UC-MF and SC-MF models shows that the increase in latent factors gradually decreases the RMSE values. For the low values of latent factors (5, 2), the RMSE of the UC-MF and the SC-MF are 1.268 and 1.312. The proposed GA-WSSR offers the RMSE values 1.256 and 1.252. The comparison

DOI: 10.24818/18423264/53.4.19.14

yields the GA-WSSR reduces the RMSE by 0.94 and 4.57 % respectively. Similarly, for the higher values of latent factors (40, 16) the GA-WSSR provides 0.93 and 2.4 % reduction respectively.



(a) (b)



(c) (d)

Figure 2. RMSE analysis for response time datasets in (a) UC-MF (b) SC-MF and throughput datasets in (c) UC-MF and (d) SC-MF.

The RMSE values of UC-MF and the SC-MF schemes are 45.8 and 49.5 for the minimum latent factors 5 and 2 respectively. They offer 45 and 46.4 for large latent factors 40 and 16 in throughput dataset. The service and user context aware models based on the service distance formulation in GA-WSSR offers 1.2, 2.84 % reduction compared to UC-MF for minimum (5) and maximum (40) latent factors respectively. Similarly, the RMSE reduction in GA-WSSR compared to SC-MF are 8.48 and 3.7 % for minimum (2) and maximum (16) latent factors respectively.

4.3 Precision & Coverage

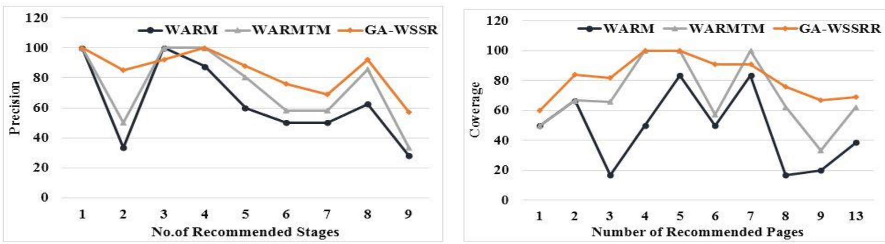
The factors to validate the recommendation performance of proposed GA-WSSR system are precision and coverage. The precision specifies the number of correct recommendations to the overall recommendations. The mathematical formulation of precision includes the recommendation set (R_{set}) and the session (T_{set}) as follows:

$$precision = \frac{R_{set} \cap T_{set}}{R_{set}} \quad (11)$$

DOI: 10.24818/18423264/53.4.19.14

$$coverage = \frac{R_{set} \cap T_{set}}{R_{cset}} \quad (12)$$

Coverage defines the proportion of relevant recommended documents to the correct recommended set. The performance of precision and coverage for the number of recommended pages is tested in this section. The number of recommended pages are varied from 1 to 9 as in Figure 3 and the corresponding precision and coverage are plotted.



(a) (b)

Figure 3. (a) Precision and (b) Coverage analysis

The increase in recommended pages results in abrupt transformations for existing Weighted Association Rule Mining (WARM) and Weighted Association Rule Mining-Text Mining (WARMTM) methods. The suitable best WS recommendation improves the coverage and precision values. The precision performance of GA-WSSR is 50 to 100 % and the coverage performance is 60 -100 % of the various number of recommended pages. The precision and coverage performance of GA-WSSR are 41.53 and 9.42 % better than the WARMTM due to the best services recommended.

4.4 Computation Time

The time required for W-prefix span algorithms-based WS selection is tested with the variations of support values. The algorithm effectiveness depends on the less time consumption. Conventionally, the W-prefix span algorithm offers the minimum time consumption values as 1.5 secs for low support values and 0.15 sec for high support values respectively compared to the prefix-span algorithm. Figure 4 depicts the comparative analysis of proposed GA-WSSR with the existing prefix span and weighted prefix span algorithm.

DOI: 10.24818/18423264/53.4.19.14

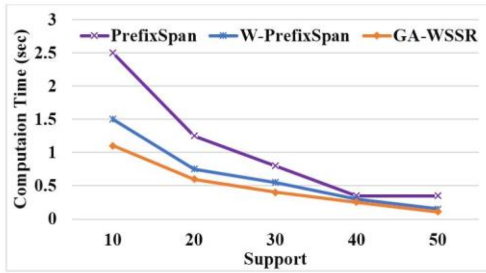


Figure 4. Computation time analysis

The best WS selection and recommendation via geographical modeling in proposed work reduce the time consumption by 26.67 % for minimum and maximum values respectively.

4.5 Memory Consumption

The memory consumed by the W-prefix span algorithms-based WS selection is tested with the variations of support values. The algorithm effectiveness depends on the less memory consumption. Conventionally, the W-prefix span algorithm offers the minimum memory consumption values as 3800 and 3400 kB for low and high-support values respectively compared to the prefix-span algorithm (5000, 4300 kB). Figure 5 depicts the comparative analysis of proposed GA-WSSR with the existing prefix span and weighted prefix span algorithms.

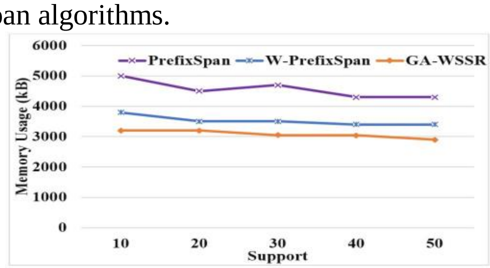


Figure 5. Memory consumption analysis

The best WS selection and recommendation via geographical modelingavoid the unnecessary computations in WS selection that reduces the memory consumption by 15.79 and 14.71 % for minimum and maximum support values respectively.

4.6 Number of patterns mined

The number of patterns derived from the WS selection and recommendation algorithm is more to state its effectiveness. Fig. 6 shows the variations of a number of patterns over the various support values. The W-prefix span algorithm mined a number of patterns (1560, 10) than the prefix-span algorithm (250, 5) for minimum and maximum support values respectively.

DOI: 10.24818/18423264/53.4.19.14

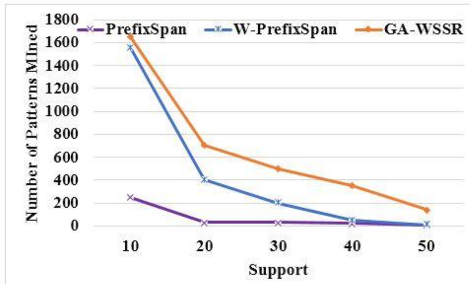


Figure 6. Mined Patterns Analysis

The best WS selection and recommendation via geographical modelingextract number of patterns with less computational steps. The proposed GA-WSSR offers 5.77 and 13 % number of patterns for minimum and maximum support values 10 and 50 respectively compared to W-prefix span algorithm.

Conclusion

This paper discussed the issues in the QoS prediction for WS selection and recommendation. The diversity in QoS records introduced the problems in similarity measurement and the factors integration under CF/MF models. The use of ensemble model increased the computational complexities to achieve the required prediction accuracy. This paper proposed the integral framework of similarity values with the geographical distance/QoS values (response time and throughput). Initially, the distance is computed and the datasets with the minimum distance are grouped together. Then, the integration of data set with the minimum response time and maximum throughput (merging) supported the mapping of user-service context aware QoS prediction. Besides, the GA-based recommendation list formation assured the best WS selection. The experimental verification of proposed algorithm with the existing methods regarding the response time, memory, computation time, coverage and the number of patterns mined assured the effectiveness of QoS prediction in optimal WS selection.

REFERENCES

[1]**Chen X., Zheng Z., Yu Q. & Lyu M.R. (2014),** ***Web Service Recommendation via Exploiting Location and Qos Information***. IEEE transactions on parallel and distributed systems, Vol. 25,No. 7, pp. 1913-1924;

[2]**De Pessemier T., De Moor K., Ketykó I., Joseph W., De Marez L. & Martens L. (2012),** ***Investigating the Influence of Qos on Personal Evaluation Behaviour in a Mobile Context***; Multimedia tools and applications, Vol. 57,No. 2, pp. 335-358;

DOI: 10.24818/18423264/53.4.19.14

[3]**Deng S., Huang L. & Xu G. (2014),** ***Social Network-based Service Recommendation with Trust Enhancement*****;** Expert Systems with Applications, Vol. 41,No. 18, pp. 8075-8084;

[4]**Deng S., Wang D., Li X. & Xu G. (2015),** ***Exploring User Emotion in Microblogs for Music Recommendation****;* Expert Systems with Applications, Vol. 42, No. 23, pp. 9284-9293;

[5]**Gupta R., Kamal R. & Suman U. (2015),** ***A QoS-aware Optimal Selection Scheme for Web Services with a Trusted Environment****;* CSI Transactions on ICT, Vol. 3,No. 1, pp. 13-21;

[6]**Kayastha R.R. & Baria J. (2015),** ***A Survey on Web Service Selection and Ranking Methods***; International Journal of Computer Applications, Vol. 117,No. 16;

[7]**Kuang L., Xia Y. & Mao Y. (2012),** ***Personalized Services Recommendation based on Context-aware QoS Prediction;*** 19th International Conference on Web Services (ICWS), 2012 IEEE pp. 400-406;

[8]**Nieto A. & Lopez J. (2014),** ***Security and QoS Tradeoff Recommendation System (SQT-RS) for Dynamic Assessing CPRM-based Systems*****;** Proceedings of the 10th ACM symposium on QoS and security for wireless and mobile networks, pp. 25-32;

[9]**Parejo J.A., Segura S., Fernandez P. & Ruiz-Cortés A. (2014),** ***Qos-aware Web Services Composition Using Grasp with Path Relinking;*** Expert Systems with Applications, Vol. 41,No. 9, pp. 4211-4223;

[10]**Shehu U., Epiphaniou G. & Safdar G.A. (2014),** ***A Survey of QoS-aware Web Service Composition Techniques*****;** International Journal of Computer Applications;

[11]**Suneetha K. & Rani M.U. (2012),** ***Web Page Recommendation Approach Using Weighted Sequential Patterns and Markov Model*****.** Global Journal of Computer Science and Technology, Vol. 12,No. 9;

[12]**Wanaskar U., Vij S. & Mukhopadhyay D. (2013),** ***A Hybrid Web Recommendation System Based on the Improved Association Rule Mining Algorithm;*** arXiv preprint arXiv:1311.7204;

[13]**Xiong W., Li B., He L., Chen M. & Chen J. (2014),** ***Collaborative Web Service QoS Prediction on Unbalanced Data Distribution;*** IEEE International Conference on Web Services (ICWS), 2014 pp. 377-384;

[14]**Yu Q. (2014),** ***QoS-aware Service Selection via Collaborative QoS Evaluation;*** World Wide Web, Vol. 17,No. 1, pp. 33-57;

[15]**Zhu J., Kang Y., Zheng Z. & Lyu M.R. (2012),** ***A Clustering-based QoS Prediction Approach for Web Service Recommendation;*** 15th IEEE International Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing Workshops (ISORCW), 2012 pp. 93-98.

DOI: 10.24818/18423264/53.4.19.14