Web Service Recommendation Based on Semantic Analysis of Web Service Specification and Enhanced Collaborative Filtering

S. Subbulakshmi[1]([⊠]), K. Ramar[2], Ameena Shaji[1], and Parvathy Prakash[1]

[1] Department of Computer Science and Applications,
Amrita School of Engineering, Amritapuri, Amrita Vishwa Vidyapeetham,
Amrita University, Amritapuri, Kollam, India
`subbulakshmis@am.amrita.edu`,
`ameenashaji.009@gmail.com`,
`parvathiprakashlll@gmail.com`
[2] Department of CSE, Einstein College of Engineering, Tirunelveli, India
`kramar.einstein@gmail.com`

Abstract. With growing momentousness of Internet applications, digital world is overwhelmed with huge number of web services. To ease the job of selecting relevant WS in service composition process, recommendation system of Web Services is designed. It uses semantic analysis of WS along with enhanced collaborative filtering. Ontology based Semantic Analysis performed using Tversky Content Similarity Measure helps to identify most similar functionally relevant WS. The collaborative filtering process uses DBSCAN clustering and PCC similarity to identify highly collaborative WS, based on ratings given by experienced users. To overcome the existence of sparse data in WS ratings and to enhance filtering process, SVM Regression is implemented before collaborative filtering. Relative frequency method is applied to amalgamate collaborative and sematic similarity values of WS. The methodology is proved to produce more realistic, accurate and efficient WS recommendation. Future focus may be towards knowledge based filtering with real world contextual information.

Keywords: WS ontology description · Semantic content filtering · Enhanced collaboration · Efficient filtering of WS

1 Introduction

With the beginning of 21st century, the Internet began developing and expanding with magnificent velocity. The amount of information available on the Internet has got extremely prodigious, resulting in an information overfill. Performing complex business operations and providing quality information to the end users can be achieved through the use of selection of most relevant Web Services. Exiting systems [10] are not enough efficient to retrieve the actual web services desired by the user up to a limit.

Contextual based Recommender systems [9] are competent of solving this issue up to a great extent by selecting the services in par with the requirements of the end users. It has emerged as a powerful tool for reducing the complexity of information/services enormity. Web service recommendation is the process of identifying the measure of usefulness of the web services and proposing them to the user.

Web services provide a regulated manner to incorporate web applications using open standards over an Internet protocol back bone, using some platform and language independent interfaces meant for easily assimilating heterogeneous systems. Web Services provide interoperability between various applications. UDDI, WSDL and SOAP define standards for service discovery, description, and messaging protocols for web services respectively.

Many researchers are focused in adopting the content based and collaborative filtering approach in the process of selecting the web services. Collaborative filtering [15] mainly centres on identifying neighbourhoods of target user consisting of other users with similar interests or preferences. Collaborative recommenders rely on user profiles, usually represented as rating vectors. Examples of such applications includes recommending movies, tour destinations, music, games etc.

Existing content based filtering methods employs the exact keyword similarity measures for the selection of web services. Majority of the traditional methodologies focus on searching the existing UDDI registries or implements keyword based search process. This resulted in poor recommendation and also requires clear and correct queries from the user. Therefore, in this paper, we present a high performance recommender system of web services based on user preferences, which makes use of machine learning and data mining techniques like regression [2] and clustering coupled with the advantages of semantic analysis. It is able to provide the end user with the most relevant web service which delivers the most pertinent information.

An enhanced content and collaborative filtering approaches for web services is designed to select the most appropriate services which handles data sparsity, data overload and scalability issues of the existing system. The initial phase of the approach concentrates on content based filtering along with semantic analysis. It consists of two central tasks such as domain feature's similarity checking, and matching of input output parameters. An ontology is designed using necessary data extracted from the corresponding WSDL files of web services. Semantic based similarity calculations are performed on domain features and input-output parameters of the given web services using Tversky's Content Similarity Measure to reach at a set of highest associated web services.

Secondary phase of the approach mainly encloses three major stages - Data sparsity removal, clustering of similar items, and ranking of similar web services. Unrated or unobserved web services available in web may cause data sparsity. In order to tackle with the problems of data sparsity, we use SVM regression to fill in missing user ratings. Grouping of similar web services is achieved through DBSCAN. When a user inputs a search query, its corresponding cluster is identified and web services that fall on the identified cluster are then ranked using PCC.

Tertiary phase focuses on combining outputs from previous filtering modules to constitute an improved and more accurate high quality recommender output. Relative frequency method is implemented for this purpose. The efficient enhanced

recommendation of web services can be widely used in the service composition where the service broker agent wants to automate the dynamic selection of the best services from the existing set of web service registries.

The rest of paper includes: the detailed study of the existing research work in Sect. 2 Related Works, illustration of the proposed Enhanced Recommendation System of WS in Sect. 3 System Methodology, the analysis report of Enhanced Collaborative Filtering and Sematic based Content Filtering approach in Sect. 4 Result and Analysis and the concluding part with directions for future improvements in Sect. 5 Conclusion.

2   Literature Review

A number of researches have already been done on recommendation system for web services. Yang et al. [3] gives the semantic similarity between web services through calculating the normalized google distance. It uses google massive terms and open google search engine to determine the normalized google distance between notions.

Lina Yao proposes an approach [1] that joins both Content based and Collaborative based methodologies by considering both appraisals and functionalities of web administrations utilizing a Probabilistic Generative Model. The idle inclinations are measurably assessed utilizing Expectation-boost calculation. To overcome information sparsity issue, data smoothing method is adapted. The system is further improved by content similarity and implicit user description aspect model.

An implementation of automated adaptive framework [5] for the WS coupled with optimisation of QoS based on quality specifications in the Web Services Ontology. Using this framework, the users are able to acquire a set of web services, by consuming the context information of users and services, which is further enhanced by the Quality factors of those web services.

Mingxin Gan proposes an approach [7] that relies on ontologies to determine the semantic similarity between tags. The system uses five categories of methods based on semantic distance, information content, properties of tags, ontology hierarchy [15], and hybrid methods. Semantic similarity is calculated by the length of the path from the leaf nodes to the root node. Tags are represented as collection of features, normalization and set theory functions are applied to estimate semantic similarity between tags.

An interactive composition approach [4] by Evren Sirin, using matchmaking algorithms is presented to help users to filter and select services while building the composition. The filtering and selection of services helps the users in the composition process. A travel recommendation system in Semantic Web using Ontology is designed [6] by Chang Choi. The Metadata is made by preference profile and transaction profile. The Travel Ontology is made by OWL Rule based on Description Logic.

3   Proposed Methodology

Every service available on Internet are provided by different web services. The core area of our work is an improved web service recommendation system that recommends the most relevant information available on web. The proposed methodology shown in Fig. 1

relies on two major phases Sematic Based Content (SCB) filtering and Enhanced Collaborative filtering (ECR) methods for the selection of relevant service. The key idea of our proposed approach is to recommend web services by selecting the semantically similar service with high user ratings provided by different users.
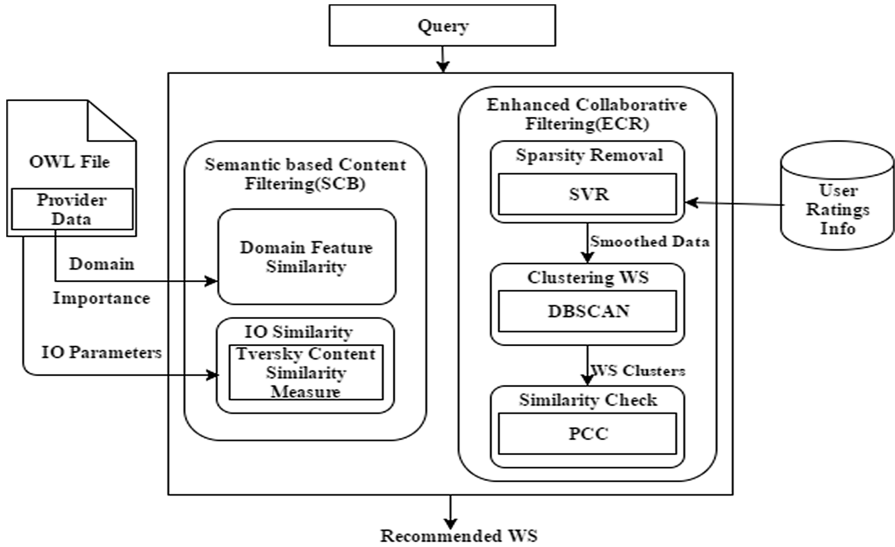

Fig. 1. Methodology for Web Service Recommendation

The design methodology implements three major features: removal of sparse data using regression methodology, improve efficiency by clustering of data and to accomplish more realistic selection by adopting the semantic based methodology. An Ontology of the web services is maintained as a repository to store the details of the services and their relationships to be used for the sematic retrieval of the required services. In essence, recommendation is based on an automatic dynamic selection of pertinent services, subject to the filtering process of web services collaboration based on the ameliorated ranking given by multiple users which exhibit similar preferences or behaviours.

3.1 Semantic Based Content Filtering (SCB)

According to W3C, the semantic web establishes a standard framework which enables to share or reuse data and services across application, organizations, enterprise, and community boundaries. Initial phase of our work is centred on Semantic Content based filtering approach which uses the ontology for identifying the web services with the required specifications given in the user query. Each web service consists of a WSDL file which defines how a service can be called, parameters required as its input/output data, domain name and other specifications. By consuming the materials available in the WSDL file, an ontology is populated with OWL, the ontology language for semantic web.

SCB similarity is calculated by using improved Tversky's Content Similarity Measure considering domain features and input-output parameters of services. Firstly,

the domain type and prominence of the service are expended to find the similar services. It is followed by an input-output parameter matching to analyse the similarity between user requirements (as query) and WS message descriptions. Thus, web services with highest similarity values are selected to produce the output of SCB filtering.

3.1.1 Web Service Ontology Creation

The SCB filtering starts with the creation of an ontology for web services. Ontology is the working model of objects belonging to a particular area of interest and their semantic relationships to each other. Ontologies are authored using Ontology Web Language (OWL), a set of knowledge representation languages built upon a W3C XML standard. Our system relies upon a web service ontology developed using Protégé and it defines a set of web services and their mutual semantic relationships. The specification of web services given by service providers residing in service registries are retrieved, to be used in the process of web services ontology creation.

Ontology created includes a set of classes and subclasses which depicts the relationship between different services. The service description like domain name, input/output parameters and importance of the services are included as data properties in the ontology. Figure 2 shows the sample of the ontology created in the system using protégé tool. For the selection of required services, the semantic details of web services are retrieved to calculate the similar web services.
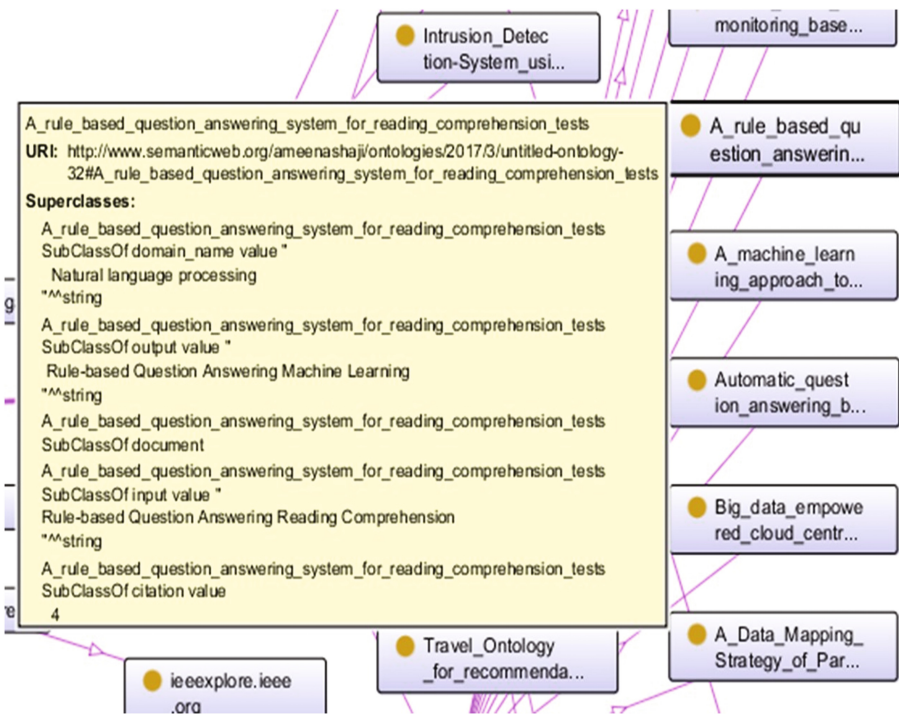

Fig. 2. Ontology of web services with service description

Similarity between user requirements and web service description is calculated using improved [7] Tversky Content Similarity Measure. Tversky defines a similarity measure according to the matching process, which generates a similarity value based on, not only common factors but also distinct features of web services.

3.1.2 Content Similarity Measure

This approach is an efficient method used in determining information-theoretic similarity values. Unlike other models this model determines the similarity between the matching-features and then it evaluates the impact of non-matching features of those web services to assess the similarity between them.

Algorithm to calculate the similarity between given web service W_k with all other web services in the ontology:

Improved Tversky's Similarity for WS

---

1.   Identify the domain name d_k and set of properties p_k of the web service W_k
2.   Select the set of web service S whose domain name is equal to d_k
3.   For every web service W_i in S, do the following:
    a   Retrieve the input/output property pi of web service W_i
    b   Calculate the Improved Tversky's similarity $IT_{Sim(i)}$
       between p_k and p_i ,

$$T_{Sim(w_i, w_k)} = IT_{sim} + C_{Sim}$$

    c   Retrieve and add the importance value I_i of W_i with $T_{Sim(i)}$.

3.1.3 Improved Tversky Content Similarity Measure

The improved similarity method implements the Tversky's normalization with the set-theory functions intersection ($p_i \cap p_k$), difference (p_i / p_k) and the Cosine similarity functions. The standard formulation is given as:

$$IT_{Sim} = \frac{|p_i + p_k|}{|p_i \cap p_k| + \mu |p_i/p_k| + (\mu - 1)|p_k/p_i|}. \tag{1}$$

$$or \ 0 \leq \mu \leq 1$$

$$C_{Sim} = \frac{p_i \cap p_k}{p_i + p_k - (p_i \cap p_k)}. \tag{2}$$

where p_k and p_i corresponds to the description sets of web service W_k and W_i and μ is a function [7] that defines the relative importance of the non-common features. The semantic relationships maintained in ontology is used to determine the relative weightage for properties of web services. Thus, by dynamically assigning accurate value for μ and by aggregating the importance of the service, this method is able to select content based similar web services.

3.2   Enhanced Collaborative Filtering (ECR)

Generally Collaborative filtering is used in recommendation system to identify popular items among peer users with the help of the ratings given by different users. We have adopted an Enhanced Collaborative (ECR) filtering process, which employs sparsity removal and clustering methods to select popular web services. Sparsity removal is used to fill the missing values in the user ratings data set with SVM regression methods. DBSCAN clustering method is implemented to group related services so that only those services with basic features could be considered for collaborative calculation.

3.2.1   SVM Regression

SVM Regression is used for predicting missing values. Rated web services are contained in trained file and unrated web services are contained in tested file. Unrated web services in the test file are rated using ratings contained in train file. SVR derives a function $f(x)$ which has less deviation between observed and predicted training samples. It also minimizes the error which is a combination of training error and a regularization term that controls the complexity of the hypothesis space.

Algorithm for Sparsity Removal

---

1.   Actual dataset with user ratings is divided into training set and testing set.
2.   Data set is classified and predicting using SMOreg classification and prediction.
3.   Classification is done using *classifier.getClassification* and class labels are assigned.
4.   Maximum marginal plane is identified in order to maximize the prediction accuracy.
5.   Real values are predicted to fill the sparse data with *evaluation.NumericPrediction*

3.2.2   DBSCAN

Density-Based Spatial Clustering groups all services of data set into service of clusters and noise. The key idea of clustering is to identify whether the minimum number of services are present within the given radius i.e., the density in the neighbourhood has to exceed some threshold value.

Input: User ratings filled using SVM regression and Eps value.

Algorithm for Clustering Web Services

---

1.   Select an arbitrary web service w_a
2.   Recognize all web service's density reachable from web service w_a with Eps and MinPts
3.   If w_a is a core service, a cluster is formed.
4.   If w_a is a border service, no web services are density reachable from w_a and DBSCAN visits the next web service of the dataset.
5.   Continue the process until all of the web services have been processed.

DBSCAN applies Euclidean distance to find distance between two web services. If the Cartesian coordinates of user ratings for two web services are $w_a = (w_{a1}, w_{a2}, \ldots, w_{an})$ and $w_b = (w_{b1}, w_{b4}, \ldots, w_{bn})$ in Euclidian n space, the distance (d) from w_a to w_b, or from w_b to w_a is given by the Pythagorean rule:

$$d(w_a, w_b) = \sqrt{\sum_{i=1}^{n} (w_{ai} - w_{bi})^2}. \qquad (3)$$

where n is the number of users and w_{ai} is the rating given by i^{th} user to web service w_a and also w_{bi} is the rating given by i^{th} user to web service w_b.

3.2.3 Pearson Correlation Coefficient

The final step Collaborative Filtering process is PCC which is a quite famous algorithm used for selection of candidate items. PCC measures the strength of linear association between two variables, where $r = 1$ means a perfect positive correlation and the value $r = -1$ means a perfect negative correlation. The selection of the highly collaborative web services among the cluster is effectively computed using this algorithm.

The Correlation between the web service W_k queried by the user with all other web services W_i in the cluster, to which the queried web service belongs is implemented using PCC equation:

$$Cor(W_k, W_i) = \frac{\sum_{j=1}^{n}(W_{kj} - \overline{w_k})(W_{ij} - \overline{w_i})}{\sqrt{\sum_{j=1}^{n}(W_{kj} - \overline{w_k})^2} \sqrt{\sum_{j=1}^{n}(W_{ij} - \overline{w_i})^2}}. \qquad (4)$$

where W_{kj} and W_{ij} are the ranks given by n users for web service W_k and W_i.

4 Result and Analysis

The simulation of the system is tested with the sample data set which includes the ontology owl file with web service descriptions and the ratings csv file with the user ratings provided by different users for a set web services. The online user should rate each web service available according to the level up to which he/she is satisfied with that WS. Ratings can range from one to five. Unobserved user ratings are assumed as 0, i.e. the sparse data. Only those sparse data are filled up using SVM Regression, it is done as the pre-processing step. Sample results of filling the sparse data is shown in Tables 1 and 2.

Table 1. Sample user ratings data set with sparse data

WS/ User ID	Admin@ .com	aishu@ .com	ajith@ .com	akhil@ .com	akhilv@ .com	akshaya@ .com	amee@ .com
w1	4	4	3	3	4	0	0
w2	3	3	4	5	3	0	0
w3	0	2	1	4	0	0	0
w4	2	1	2	3	2	1	2
w5	0	0	4	1	0	3	2

Table 2. Sample user rating data set with filled values after pre-processing

WS/ User ID	Admin@ .com	aishu@ .com	ajith@ .com	akhil@ .com	akhilv@ .com	akshaya@ .com	amee@ .com
w1	4	4	3	3	4	5	1
w2	3	3	4	5	3	4	1
w3	4	2	1	4	4	1	1
w4	2	1	2	3	2	1	2
w5	1	3	4	1	3	3	2

Table 3. Results of SCB and ECR filtering.

WS	TSim	CSim	ITSim	ECR	Nor. ECR	RF
w1	1.376	0.153	3.764	0.22	2.361	0.139
w2	1.4	0.166	1.783	0.209	2.284	0.069
w3	1.506	0.222	1.864	0.598	5	0.086
w4	1.764	0	2.882	0.025	1	0.073
w5	1.636	0.265	4.961	0.123	1.684	0.178
w6	1.733	0.3	4.016	0.355	3.304	0.186
w7	1.969	0.3	2.134	0.377	3.457	0.087
w8	2.173	0.25	4.211	0.095	1.489	0.15

The results of semantic based and collaborative filtering are shown in Table 3. Relative Frequency (*RF*) method is used for combining the SCB and ECR filtering outputs together for a better recommender result. The proportion of total possible number of events to the count of the favourable events is termed as relative frequency. Relative Frequency (*RF*) method is used for combining the SCB and ECR filtering outputs together for a better recommender result. The proportion of total possible number of events to the count of the favourable events is termed as relative frequency. RF functions for Similarity and Collaborative filtering are given by:

$$Relative\ Frequency_{(W_k)} = \frac{(Sim_i + Cor_i)}{\sum_{i=1}^{n}(Sim_i + Cor_i)}. \tag{5}$$

where Sim_i is the similarity score, Cor_i is the correlative score of the kth web service and n is the number of web services. The above function denotes relative frequency as a proportion.

The above table shows the detailed result of SCB and ECR filtering and the integrated results using Relative Frequency Method. Moreover, graphical representation of the above results are also shown in Figs. 3 and 4.
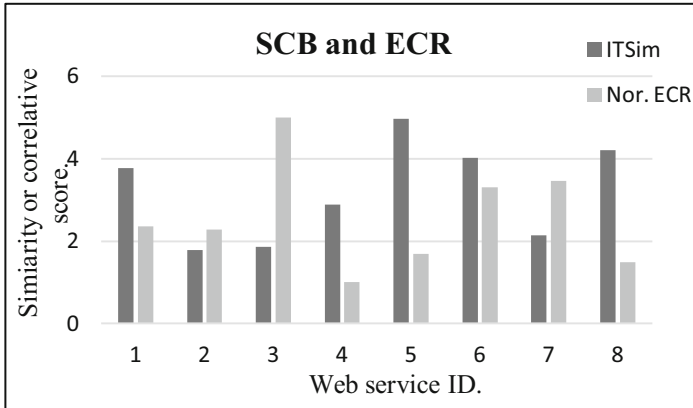
Fig. 3. Similarity values of SCB and ECR of web service.

The above graph reveals that the services are given higher ranks only when both SCB and ECR values are relatively higher as Web Service 6, 5, 8. If the services fails to score higher values in any of the filtering methods, they are given least importance in the process of selection like services 2, 4 shown in the above graph.
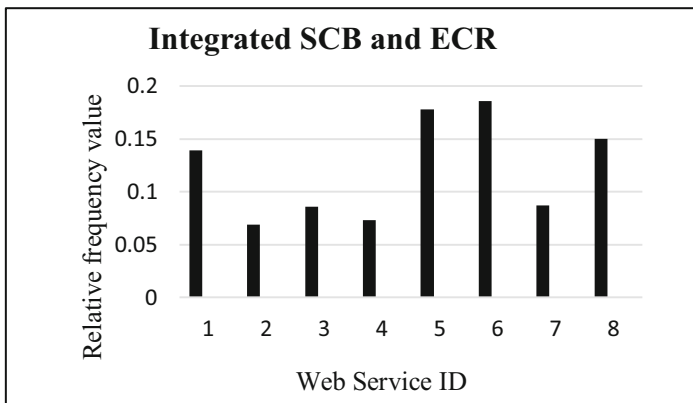

Fig. 4. Integrated Outcome of SCB and ECR filtering by applying RF.

The final ranking and recommendation of web services with regard to the given web service is shown in Fig. 5. The results of the implemented system revealed to be closest to the user expectation as the results produced is better than recommendation made without sparsity removal. The execution time of the system is reduced to greater extent as the clustering methodology is performed before PCC for web services.
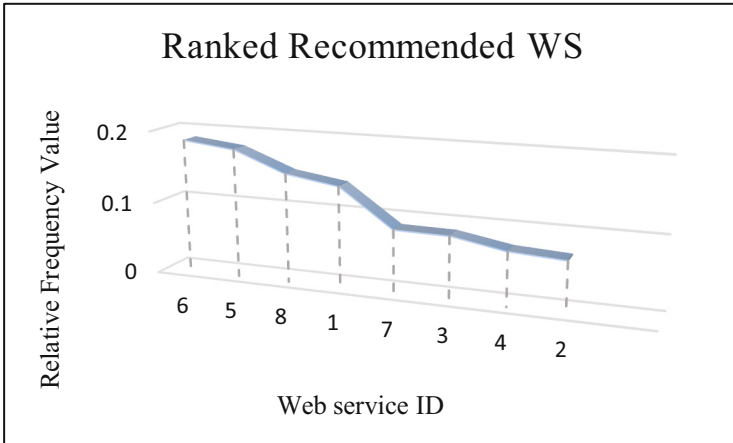
Fig. 5. Ranking of Web Services.

The elaborated web service selection process with the semantic and enhanced collaborative filtering methodology could be used in the process of service collaboration by broker services for selecting the most appropriate services in par with their requirements.

5   Conclusion

The process of selecting and recommending relevant web services from a wide variety of available choice is an area of concern in Service Oriented Computing. Most current recommendation approaches focus on either UDDI registries, or keyword-dominant, QoS-based Web service search engines that have limitations such as reduced recommendation performance and dependence on detailed, precise search queries from the user. Our combined approach simultaneously considers user ratings similarities along with semantic content of Web services. As the methodology also considers filling of missing user ratings, the recommender output is much better with improved accuracy. Our approach only considers web services semantics. Semantic analysis of keywords can be incorporated as a part of future advancement.

References

1. Yao, L., Sheng, Q.Z., Ngu, A.H.H., Yu, J., Segev, A.: Unified collaborative and content-based web service recommendation. IEEE Trans. Serv. Comput. **X**(X) (2014)
2. Song, X., Zhou, T., Zhang, H.: Support vector regression estimation based on non-uniform lost function. In: IEEE Engineering in Medicine and Biology 27th Annual Conference, Shanghai, 2005, pp. 1127–1130 (2005)

3. Yang, H., Fu, P., Yin, B., Ma, M., Tang, Y.: A semantic similarity measure between web services based on google distance. In: 35th IEEE Annual Computer Software and Applications Conference, pp. 14–19 (2011)

4. Sirin, E., Hendler, J., Parsia, B.: Filtering and selecting semantic web services with interactive composition techniques. IEEE Intell. Syst. **19**, 42–49 (2004). doi:10.1109/MIS.2004.27

5. Subbulakshmi, S., Ramar, K., Renjitha, R., Sreedevi, T.U.: Implementation of adaptive framework and WS ontology for improving QoS in recommendation of WS. In: Rodriguez, J.C., Mitra, S., Thampi, S., El-Alfy, E.S. (eds.) Intelligent Systems Technologies and Applications 2016. ISTA 2016. AISC, vol. 530, pp. 383–396. Springer, Cham (2016). doi:10.1007/978-3-319-47952-1_30

6. Choi, C., Cho, M., Kang, E., Kim, P.: Travel ontology for recommendation system based on semantic web. In: Conference: Advanced Communication Technology, ICACT March 2006, pp. 624–627 (20006)

7. Gan, M., Dou, X., Jiang, R.: From ontology to semantic similarity: calculation of ontology-based semantic similarity. Sci World J 2013, Article ID 793091 (2013)

8. Hu, Y., Peng, Q., Hu, X., Yang, R.: Time aware and data sparsity tolerant web service recommendation based on improved collaborative filtering. IEEE Trans. Serv. Comput. **8**(5), 782–794 (2015)

9. Basu, C., Hirsh, H., Cohen, W.: Recommendation as classification: using social and content based information in recommendation. In: Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI-98), pp. 714–720 (1998)

10. Smita, A.: Recommender system review. Int. J. Comput. Appl. **71**(24), 0975–8887 (2013)

11. Bahramy, F., Crone, S.F.: Forecasting foreign exchange rates using support vector regression. An empirical evaluation of mean reversion using bollinger bands. In: IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr), pp. 34–41 (2013)

12. Yao, L., Sheng, Q.Z., Segev, A., Yu, J.: Recommending web services via combining collaborative filtering with content-based features. In: IEEE 20th International Conference on Web Services, pp. 42–49 (2013)

13. Maximilien, E.M., Singh, M.P.: A framework and ontology for dynamic web services selection. IEEE Internet Comput. **8**(5), 84–93 (2004)

14. Lecue, F.: Combining collaborative filtering and semantic content based approaches to recommend web services. In: Proceedings of the 4th IEEE International Conference on Semantic Computing (ICSC 2010), Pittsburgh, PA, USA (2010)

15. Linden, G., Smith, B., York, J.: Amazon.com recommendations item-to-item collaborative filtering, pp. 76–80. IEEE Computer Society, Washington, DC (2003)

16. Sajeev, G.P., Ramya, P.T.: Effective web personalization system based on time and semantic relatedness. In: International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, pp. 1390–1396 (2016)

17. Gaurav, S., Jithendranath, Y., Adil, A., Yadav, S., Kasturi, B.: A study to assess and enhance educational specific search on web for school children. In: Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing, IEEE ICSC 2015, pp. 260–263. Institute of Electrical and Electronics Engineers Inc. (2015)