ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (517-524), Month: April - June 2015, Available at: www.researchpublish.com

Web Service Recommendation via Quality of Service Information

[1]Rajeshree Mohalkar, [2]Kshiteeja Fadtare, [3]Shweta Todkar

[1, 2, 3] P.E.S. Modern College of Engineering, University of Pune, Pune, India.

Abstract: **Web services are integrated software components for the support of interoperable machine-to-machine interaction over a network. Web services have been widely employed for building service-oriented applications in both industry and academia in recent years. The number of publicly available Web services is steadily increasing on the Internet. However, this proliferation makes it hard for a user to select a proper Web service among a large amount of service candidates. An inappropriate service selection may cause many problems (e.g., ill-suited performance) to the resulting applications. In this paper, we propose a novel collaborative filtering-based Web service recommender system to help users select services with optimal Quality-of-Service (QoS) performance. This system provides a QoS-aware Web service recommendation approach. The basic idea is to predict Web service QoS values and recommend the best one for active users based on historical Web service QoS records. Thus we can improve the recommendation accuracy and time complexity compared with existing service recommendation algorithms. Proposed method uses enriched NLP protocols to get the recommendation from the user comments. System successfully merges web service ranking and user comments to provide best hybrid solution for proper recommendation.**

Keywords: **Web Service, Quality of Service (QoS), Recommendation, Collaborative Filtering, Pearson's Correlation**

I. INTRODUCTION

Selection of a high quality Web service among a large number of web services is not an ordinary task.Some developers implements their own services instead of using publicly available ones, which brings additional overhead to both time and resource. Use of an inappropriate service, may add potential risk to the business process. Hence, effective approaches to service selection and recommendation are necessary, which can help service users reduce risk and deliver high-quality business processes.Quality-of-Service (QoS) is the non-functional characteristics of Web services and considered as the key factor in service selection [1]. QoS is a set of properties including availability,response time, throughput etc. Among these QoS properties, values of some properties (e.g., response time, user-observed availability, etc.) need to be measured at the client-side [2].We propose a system which gives recommendation by the similarity measures of the captured response time of all web services to save time of the new user.This system ensures the application in web servers has a proper cooperation with one another by suggesting point of slowness. It also recommend web service via user Comments on the site, which can be read by web crawler.

II. LITERATURE SURVEY

Now a day's web services are becoming a prime factor to push the software's over the internet network. Here the quality of service of the system is totally dependent on how accurately the supporting web service is behaved. So it becomes a challenging task to select the web service for the service users of the software's. For some of the web services it is very difficult to identify the QoS, Hence it becomes a challenging to predicate its quality.

The QoS of web service is depends on number of factors such as time of response, probability of failure, throughput and etc. Since the environment of network is different, it may possible that service user will get different QoS metrics for the

same web service. Also it may possible that the same user will get different QoS measure for the different times. Because it is depends on the current cloud services, the end users demands for the multimedia and many other factors. It is very important for the user to predict the QoS of web service. Ecommerce websites makes use of past history of users to recommend the new things to users, but QoS service recommendation is not as easy as product recommendation.

Table.1 Literature survey

Sr. No	Paper Title	Year	Author	Result
1.	SPEEDING-UP PEARSON CORRELATION COEFFICIENT CALCULATION ON GRAPHICAL PROCESSING UNITS	2010	Logoglu K.B, Atek T. K.	Calculate similarity between two entities
2.	A latent model for collaborative filtering	2011	Helge Langseth,ThomasDyhre Nielsen	Model based on EM algorithm for collaborative filtering
3.	Towards web service selection based on QoSestimation	2012	S. Wang, Q. Sun, and F. Yang	Web Service Selection Approach based on QoS Estimation
4.	A Clustering-Based QoS Prediction Approach for Web Service Recommendation	2012	J. Zhu, Y. Kang, Z. Zheng, and M.R. Lyu	Enhancing the QoS prediction accuracy via clustering techniques
5.	CONTENT -BASED RECOMMENDATION SYSTEM BASED ON VAGUE SETS	2013	Yu Shi, Weiping Wang	Improves the quality and accuracy of the recommendation.

Here, a simple illustration to address the QoS prediction for Web services. As shown in Table 1, there are response time (i.e. RT) records of three Web services w.r.t five users. The element $r_{i,j}$means the RT value of user i for service j, and "NA" represents the corresponding value not available at present. Assume user u3 has some interests on the third service, since there is no ready record in the table; he has to predicate the issue r3, 3 according to his own and others' service invocation records.

Table.2 Illustrating QoS prediction problem.

User	Service1 Response time	Service2 Response time	Service3 Response time
U1	0.4	1.6	NA
U2	0.9	NA	1.9
U3	2.8	3.5	??
U4	NA	3.0	4.0
U5	0.8	NA	0.9

How to estimate the missing value? Besides u3's existing records on other two services (i.e. r3, 1 and r3, 2), the available service invocation records of other four users also should be taken into consideration.With regard to prediction techniques, experiences tell us that collaborative filtering (CF) techniques can be viewed as a good choice.

The web service is grouped and distributed across a large network hence it is bit difficult for the service users to find out the data related with the execution of the service. If the precision of the prediction is to be improved then available web services records should get fully utilized. [3] Gives an approach to predicate the QoS of the web service with the help of collaborative filtering. Here the author collects the experiences of the users on web service and based on that data prediction of quality of web service is done. In this research similarity based prediction by using person correlation is

used. It is important to note that person based similarity can efficiently predict the result but it can take more time to compute and the performances also degrades if the datasets to be used are sparse.

Basically there are two types of predictions are used to predict the QoS of web service: statistical prediction and personalized prediction. Personalized recommendation involves collecting the information about the users who visited the site, manage and assess the historical behavior of the user and based on that a perfect matter is sent to the right users. Recommendation methods are also categorized in three types: rule based filtering; content based filtering and collaborative filtering.

Rule based filtering generates a utility functions that are specific to the users and thus applying this to the data under observation. In content based filtering for each user a profile is generated and the profile is depends on the description of the previously visited items by the user.

Collaborative filtering is one of the most widely acceptable and successful method used for recommendation purpose. In this method of prediction rating is done which is based on the past history of the user's selection. In the beginning of the 1990's collaborative filtering start to impress the online information world. That time Tapestry [4] a collaborative filtering system that works manually to do prediction on various domains. Later on automatic collaborative filtering system such as GroupLens is proposed. The method is fully automatic as the method is used to identify the articles in which users are interested.

[5] Did a survey on collaborative filtering methods. Baseline predicators is method which is not comes actually under collaborative filtering, it just taken as base for other methods. K-NN collaborative filtering is first automated CF method and first time it gets used in the GroupLens recommender system. Item–Item Collaborative Filtering is another CF method which is effective in many scenarios but often suffered from the scalability problem. Dimensionality Reduction used to reduce the dimensions of ranking which was problematic in previous methods. All the methods of the collaborative filtering are comes under three categories: hybrid methods. Memory based methods and the model based methods.

An experimental results show that the input of users, network condition and the performance of the web service can dramatically affects the QoS. [6] Implements a method for the prediction of QoS of web service based on the collaborative filtering and Pearson correlation. Here in this paper the missing QoS values can be predicted by making the combination of Pearson correlation and Slope One method. To minimize the prediction errors another strategies such as weight adjustment and smoothing are used. The experimental evolution shows that it achieves the higher prediction precision than the famous WsRec algorithm.

[7] Gives an approach on personalized prediction for the changing service management. Here pattern mining is done to extract the invocation patterns from the past history of the user visits. Here the way in which prediction done is dynamic and it makes use of collaborative filtering technique. [8] Proposed a theory of recommendation on dynamic content s by using bilinear models that are predictive in nature. The author concludes that the given approach is lightweight approach and also it is flexible with the personalized tasks. To do experiments they selects yahoo as a there datasets with the six well known prediction techniques.

The idea of this proposed method is triggered by the fact that internet users are not getting desired web services even though they spend more time to identify such one. Generally the best web service is identified by the two main parameters

- Web service whose transaction time is less
- Web service which is been using by the most of the users

So by keeping these things in the mind we leverage our idea of finding the best web services by the combination of web service response time and collaborative recommendation of other users.

So we weave a scenario where we are developing some 4 to 5 web service for flight booking services. Where every user first creates his user profile and then book the flight to his desired destinations. So for every transaction carried out by the user is recorded with its response time at the web server. Then after getting some threshold number of response times our system will generate the recommendation purely based on response time of the user. This we call as Fresh recommendation.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (517-524), Month: April - June 2015, Available at: www.researchpublish.com

Then user enjoys this fresh recommendation and then goes ahead at the end to give his opinion in the survey. Then this opinion will be used for the collaborative recommendation. Then by finally merging and measuring the weight of the two recommendations, system will suggest the best web service to the user.

In data mining pre-processing is at state of heart as the information to be processed is never presents in proper forms. Also the pre-processing is the basic need of the natural language processing systems. Hence today's data mining research area it grabs a lot of attention. Pre-processing includes stemming, stop word removal etc. Out of those stemming is an important method. Stemming is a process where infected and derived forms of words are brings to their base form without changing the meaning of the word. Also stemming proves its usefulness in information retrieval system, because it increases the no of documents to be retrieved.

Stemming algorithms are widely classified as shown in figure. In 1968 lovins [9] proposed a lovins stemming algorithm where it considers almost 294 suffixes, 35 rules of transformations and 29 constraints to bring down the idea into reality. As Lovins stemming algorithm is a single stemmer algorithm hence it removes only one suffix from the word at a time. But one of the biggest disadvantages of the same is its time required for the operations. Because of large time consumption it lags behind others.

Correlation is used to find the interpersonal association between the two things Correlational analysis is one of the best method to analyses the data in medical and research area. Many times it is required to find whether there is any relationship between the two things or not? To what degree they bind with each other? , at the same time correlation coefficients plays an important role. This correlation coefficient can be zero, positive or negative. It can go up to the any ranges. Zero coefficient indicates that there is no relationship is exists between the two variables. Positive sign represents a positive response while negative is used for negative marking. The stronger the association the more strongly will be the relationship between the variables.

Basically three most popular coefficients are existed: Pearson's coefficient (r), Spearman's rho coefficient (r_s), and Kendall's tau coefficient (τ) [10] explains the person correlation. In 1938 Kendall's proposed a tau coefficient. He proposed it in such way that it can be used as an alternative for the rho coefficient. In general rho is bit easier to calculate than tau. The main advantage of the rho coefficient is that it has slightly better statistical properties also the way in which it show the direct relationship between the variables to be analysed.

Pearson correlation was developed in 1846 by Bravais, but for the very first time it is described by the Karl Pearson in 1896. When he proposed the idea he assumes normality of the variables being analysed. In 1904 Spearman introduced a new coefficient known as rho to find the linear relationship between the two things. By using this it get possible to measure the relationship that cannot be measure by the quantitative measures.

III. PROPOSED SYSTEM

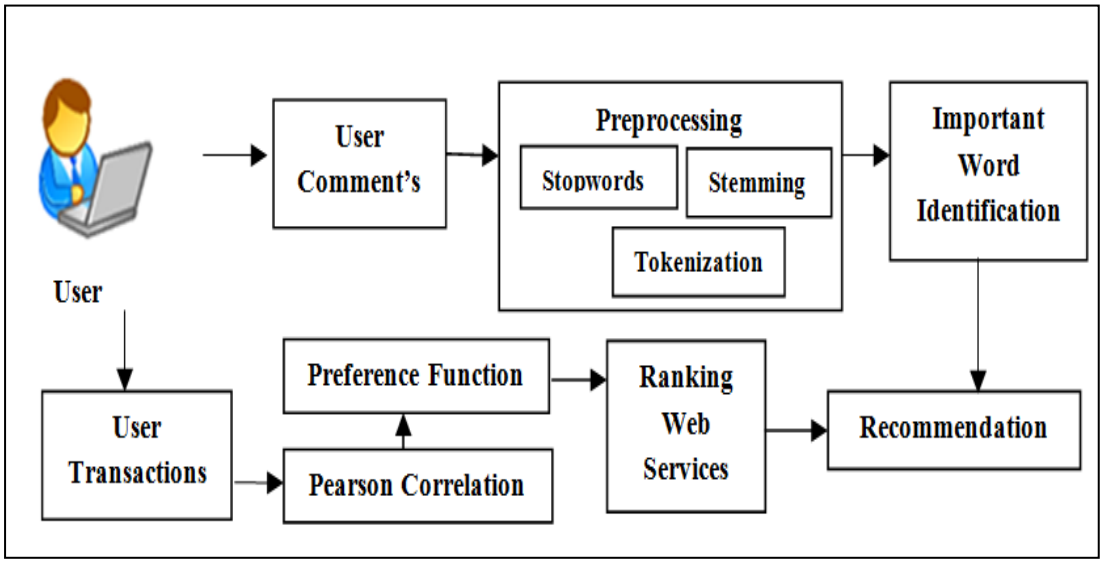

Fig.1 Architecture design of proposed system

System works as follow:

Step 1: This is the initial step our project where many users are accessing the different web applications and perform the transactions. Where on each transaction the response time is captured at the server's end and store in the database.

Step 2: Here in this pair of response time of two services is feed to the preference function as shown in equation 1, where this function actually calculates the linear aggregation of the two services. This gives the output in the form of real number. A non-zero value indicates the strength between the two services whereas zero means there is no preference between two services. And this can calculate using the equation no 1.

$\Psi(i,j) = \sum Wv\ (qv,i - qv,j)$ ----------(1)

$V \epsilon N(u)i,j$

$\Psi(i,j)$ = Preference function

$N(u)i,j$ = Subset of similar user

V= Similar user of the current web service u

WV = Weighting factor

qv = Quality of service of two web services

Step 3: Then by using the matrix of the user preferences a correlation is calculated using Pearson correlation stated in the equation 3.and this can also be shown as algorithm.

Algorithm 1: Pearson Correlation

// input: Two parameter matrix of N rows and 2 columns and Let matrix be M

// output: Pearson factor (i.e. in between 0 to 1)

Step 1: calculate sum of square of column 1 as SS1

Step 2: calculate sum of square of column 2 as SS2

Step 3: calculate square of mean of column 1 as m1

Step 4: calculate square of mean of column 2 as m2

Step 5: calculate square root of SS1-m1 as SQ1

Step 6: calculate square root of SS2-m2 as SQ2

Step 7: calculate denominator as DR as SQ1 * SQ2

Step 8: calculate sum of column 1 as sum1

Step 9: calculate sum of column 2 as sum2

Step 10: calculate product of sum1 and sum2 as TP

Step 11: calculate Mean product as MP as TP/ N

Step 12: calculate sum of product of all rows as PS

Step 13: calculate nominator as NR as MP*PS

Step 14: calculate Pearson coefficient as NR/DR

Step 15: return Pearson's coefficient

And this can be shown in the following equation

$$r = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{[N\sum x2 - (\sum x)2][N\sum y2 - (\sum y)2]}} \quad \text{(2)}$$

N = number of pairs of scores

$\sum xy$ = sum of the product of paired scores

$\sum x$ = sum of x scores

$\sum y$ = sum of y scores

$\sum x2$ = sum of squared x scores

$\sum y2$ = sum of squared y scores

Step 4: In this step web services are ranked based on the correlation factor. And then these ranked web services are used by the user and then system is asking users to enter their opinion on completion of a transaction in the form of a survey. And then this data is further accessed for the pre-processing as stated in the next step.

Step 5: This is the step where pre-processing is conducted, where string is processed to its basic meaning words by the following four main activities: Sentence Segmentation, Tokenization, Removing Stop Word, and Word Stemming.

i. Sentence segmentation is boundary detection and separating source text into sentence.

ii. Tokenization is separating the input query into individual words.

iii. Stop word removal: In any document narration the conjunction words does not play much role in the meaning of the document, so by discarding these words (like: is, the, for, an) from the documents which greatly reduces the overhead of processing.

iv. Stemming: Many of the elongated words in the English language generally fail to provide proper meaning in the given scenario and also they increases the computational time. So it is necessary to bring the words to their base form by replacing its extended.

Step 6: Term weight.-The most repetitive words in text are obviously the important words. So system identifies the list of most repeated words and considers some top n elements (where n is user defined) as the important word for text to store in vector. And this can be extract as in below shown algorithm:

Algorithm 2: To find top words

Step 0: Start

Step 1: Read string

Step 2: divide string into words on space and store in a vector V

Step 3: Identify the duplicate words in the vector and remove them

Step 4: for i=0 to N (Where N is length of V)

Step 5: for ith word of N check for its frequency

Step 6: Add frequency in List Called L

Step 7: end of for

Step 8: return L

Step 9: stop

Step 7: After identifying top words from the user comments then system weighs the words for maximum occurrence from the important word data set. (Like: good,bad,best, awesome etc.). And then finally recommend the web services which are quoted with large number of good words.

IV. RESULT AND DISCUSSION

To show the effectiveness of proposed system some experiments are conducted on java based windows machine using Apache tomcat as the server. Mean Absolute Error (MAE) metric is widely employed to measure the prediction quality of collaborative filtering methods, which is defined as:

$MAE= \sum i,j \mid r\ i,j - r'i,j \mid / N$

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (517-524), Month: April - June 2015, Available at: www.researchpublish.com

Where ri,j denotes the expected QoS value of Web service item j observed by service user i, r'i,j denotes the predicted QoS value, and N denotes the number of predicted values. Since different QoS properties of Web services have different value ranges, similar to [10], we use the Normalized Mean Absolute Error (NMAE) metric to measure the prediction quality of our hybrid collaborative filtering method. We define our NMAE to be the standard MAE normalized by the mean of the expected QoS values as follows:

$$NMAE = MAE / (\sum_{i,j} r_{i,j} / N)$$

Where smaller NMAE value means higher prediction quality.

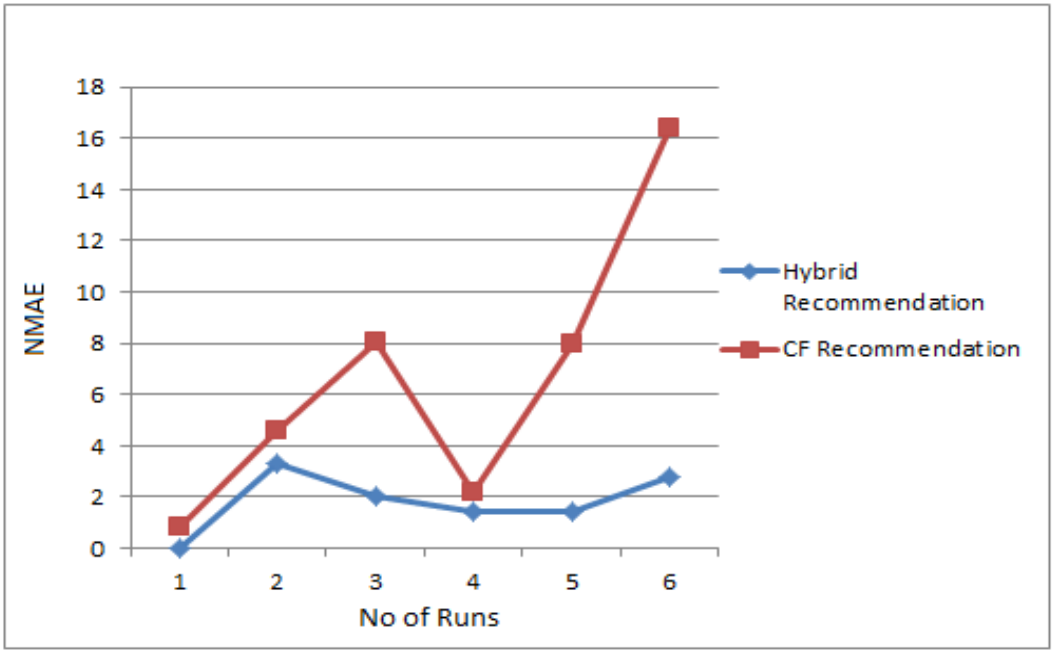

Fig.2 Performance Graph

Parameters:

On x-axis: Number of Runs

On y-axis: NMAE

The plot in the figure 2 clearly indicated that hybrid recommendation of our system clearly indicates the lesser NMAE, This means our system over performs than the traditional collaborative filtering.

V. CONCLUSION & FUTURE SCOPE

Proposed method successfully applies the similarity measure on captured response time for the transaction of the respective web services. So our system first provides the recommendation based on the performance parameter of response time and this we can consider as purely technical. Then users can use web services and finally they can tag their opinion for each of the web services. Our system successfully captures these user opinions which are in the form of textual data and then applying proper NLP rules on them system identifies the best web services for recommendation.

The proposed system can be enhance to consider more technical parameters of the web services like probability of failure, throughput to enrich the recommendation system.

VI. ACKNOWLEDGEMENT

We would like to express the deepest appreciation to our guide Prof.Mrs.DeeptiNirwal and our Head of Computer Engineering Department Mrs.SuhasiniItkar. We thank our parents who have shown tremendous support and inspired us with their blessings. We would like to express our gratitude with a word of thanks to all of those who have directly or indirectly helped us by giving beneficial information and knowledge.

REFERENCES

[1] L.-J. Zhang, J. Zhang, and H. Cai, Services Computing. New York, NY, USA: Springer-Verlag, 2007, Tsinghua University Press.

[2] N. Thio and S. Karunasekera, ‘‘Automatic Measurement of a QoS Metric for Web Service Recommendation,’’ in Proc. Australian Softw. Eng. Conf., 2005, pp. 202-211.

[3] Chengying Mao and JifuChen ”QoS Prediction for Web Services Based on Similarity-Aware Slope One Collaborative Filtering”

[4] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, “Using collaborative filtering to weave an information tapestry,” Communications of the ACM, vol. 35, no. 12, pp. 61–70, 1992.

[5] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, “GroupLens: an open architecture for collaborative filtering of netnews,” in ACM CSCW ’94, pp. 175–186, ACM, 1994

[6] Chengying Mao and JifuChen “QoS Prediction for Web Services Based on Similarity-Aware Slope One Collaborative Filtering”

[7] Li Zhang1,2, Bin Zhang1 , Claus Pahl 2 , Lei SXu2 , Zhiliang Zhu1 “Personalized Quality Prediction for Dynamic Service Management based on Invocation Patterns “

[8] Wei Chu “Personalized Recommendation on Dynamic Content Using Predictive Bilinear Models” Social Networks and Web 2.0 / Session: Recommender System

[9] J. B. Lovins, “Development of a stemming algorithm,” Mechanical Translation and Computer Linguistic., vol.11, no.1/2, pp. 22-31, 1968.

[10] Jan Hauke, Tomasz Kossowski” COMPARISON OF VALUES OF PEARSON’S AND SPEARMAN’S CORRELATION COEFFICIENTS ON THE SAME SETS OF DATA “