ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

Hybrid Recommendation System for Web Services using QOS

Reshma Gulhane[1], Dr.S.M.Chaware[2]

Student, Department of Computer Engineering, BSCOER, Narhe, Pune, India[1]

Professor, Department of Computer Engineering, BSCOER, Narhe, Pune, India[2]

ABSTRACT: Due to increasing amount of information in the web, searching proper information is becoming most challenging task. Search engine smartly does this thing to fulfill the user's requirement. But Search engines are designed to provide the probabilistic URL's for the given query but they never recommend the best among them. To achieve this recommendation of the web services need to be generate based on the quality of service (QOS) of the web services. Many recommendation systems are been introduced to do so. But due one or another reason they are not able to merge the recommendation with the technical parameter of the web services. Proposed system introduces an idea of recommending web services based on the QOS parameter and then this parameter will be merged with the collaborative filtering technique to yield best hybrid recommendation system for the web services.

KEYWORDS: Pearson Correlation, Colloborative filtering, recommendation, similarity measures, preference function.

I. INTRODUCTION

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.

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.

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

Table 1: illustrating QoS prediction problem.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

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. [1] 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 [2] a collaborative filtering system that works manually to do prediction on various domains. Later on automatic collaborative filtering system such as Group Lens is proposed. The method is fully automatic as the method is used to identify the articles in which users are interested.

[3] 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.

There is Basian network based method to predict the QoS is proposed, this system predicts the performance level of the service by analyzing the requirement level of the users. But here also there is hurdle to find out the performance level of the system. Number of QoS method are implemented that will not bother about the influence of the external factors on the performance of the service.

An experimental results show that the input of users, network condition and the performance of the web service can dramatically affects the QoS. [4] 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.

[5] 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. [6] 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

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

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.

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.

The rest of the paper is organized as follows. Section 2 discusses some related work and section 3 presents the design of our approach. The details of the results and some discussions we have conducted on this approach are presented in section 4 as Results and Discussions. Sections 5 provide hints of some extension of our approach as future work and conclusion.

II.LITERATURE SURVEY

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.

Lemmatization is a variant of stemming method. The workings of both the methods are same with a little difference in their behavior. Stemming makes use of stems but here the context of the sentence in which word is present is not identified. In contrast to this lemmatization makes use of lemma. Here the context of the sentence in which word is present is identified. In case of stemming morphological forms of words are assumed to have same meaning as like base word. [7] Did a survey on different stemming methods. Here advantages and disadvantages of each method are explained to propose a best method for the stemming.

Stemming algorithms are widely classified as shown in figure. In 1968 lovins [8] 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.

In 1980 Porter [9, 10] proposed a widely accepted, efficient Porter stemming algorithm. The algorithm considers almost 1200 prefixes of English language. It works on the principle that the long prefixes are generated by taking the combination of small prefixes. To accomplish the task Porter generates a snowball framework. It also has the same disadvantage as like Lovins i.e. large time consumption. [11] Gives a Corpus based mixed stemmer. Here author tries to overcome the little drawbacks of the famous stemmer Porter stemmer. The given stemmer takes Porter stemmer as its base of operation.

[12] Elaborates a similarity measure that is sensitive to the given query. Also author try to compare three similarity measures with the traditional similarity measures to find the best one. By conducting numerous experiments on the system author states that the given approach behaves efficiently over cosine coefficient. The main motto to do the given research is to check whether query sensitive measures are responsible for interdependent relationships of the documents.

[13] Present's a new approach for similarity measure where similarity between the two documents will be evaluated. The experimental evaluation of the system with different real word datasets shows that it behaves more efficiently than the other approaches. [14] Proposed a multi-view based similarity. Prior this research the similarity measure algorithms makes use of single point of view which degrades the performance of the system. Here in this research a measure having multi-views is proposed to increase the performance of the system. As the system get views from the multiple objects it will use to find the correct relationship between the objects of the cluster.

[15] Gives a new similarity measure which is based on the phrases. Traditional VSD models do not bother about the position of the words in the documents also the context of the sentence is get ignored. Because of this

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

drawback of the VSD models SDT model gets evolved that keeps the sequential information about the characteristics of the words. Here TF-IDF measures are used to accomplish the task and also to get the high degree of accuracy between the documents of the same cluster.

K means clustering algorithm is one of the best and widely acceptable algorithm used for the purpose of clustering. In [16] a new similarity measure is stated that can be applied on the well-known k means algorithm. Here to the k means algorithm parameter values are need to be passed, the experimental evaluation of the system with different datasets conclude that the given approach obtains the best result if the intermediate parameters are used. [17] Took a one survey on all the existing similarity measure approaches used for the document clustering purpose. It shows that some measures are really good compare to others but they are time consuming hence get lagged in the series. Also the advantage of one with respect to others gets explained for the simplicity of the readers of the document.

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) [18] 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 analyzed.

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 analyzed. 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.

[19] Proposed a web service recommender system that makes use of collaborative filtering. Here author developed the system by using java language in real world environment. To evaluate the performance the system almost gathered 21197 public webs from the internet. More than 150 users give 1.5 million test results. Hence this real world experiments shows that the system achieves high performance over another systems.

Xi Chen [20] introduces a method which is based on the technique of visualization system where it fuses three factors like

- First, Author combine the model-based and memory based CF (Collaboration filtering)algorithms for web service recommendation, which significantly improves the recommendation accuracy and time complexity compared with previous service recommendation algorithms.
- Second, they design a visually rich interface to browse the recommended web services, which enables a better understanding of the service performance.
- Finally, they conduct comprehensive experiments to evaluate our approach by employing real-world web service QoS data set. More than 1.5 million real-world web service QoS records from more than 20 countries are used in our experiments.

In this method if user wants to recommend same services of the region to the other user of different locality then system always yields wrong outcomes. And the system recommends based on the rich interface rather than considering any QOS parameters.

III.PROPOSED METHODOLOGY

In this section, we describe our framework for hybrid recommendation system for web services using QoS parameter with the below mentioned steps as shown in figure 1.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

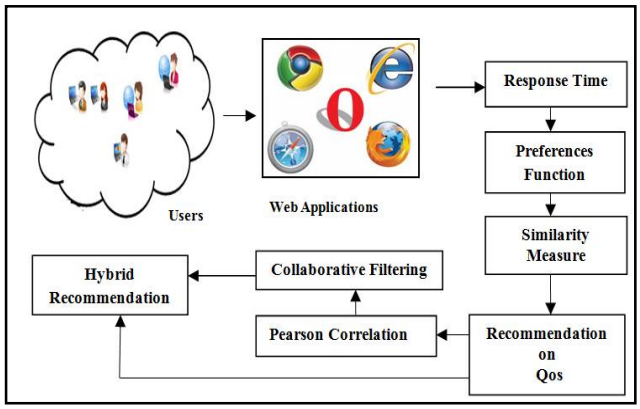

Figure 1: Overview of the proposed work

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, 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.
Step 3: In this step similarity between the two web services are calculated using Kendal rank correlation similarity equation as mention in the equation no 2.
Step 4: Here in this step based on the similarity values obtained from the previous step web service indexing is performing to rank the services. And the best web service is recommended to the user purely on the basis of the web service performance.
Step 5: Here in this step the user opinion will be identified using other user opinion this is known as the collaborative filtering. Then the opinion collected over a matrix will be fed to the Pearson correlation to identify the best recommendation.
Step 6: Then by using the matrix of the user opinion 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)
1: calculate sum of square of column 1 as SS1
2: calculate sum of square of column 2 as SS2
3. calculate square of mean of column 1 as m1
4. calculate square of mean of column 2 as m2
5.calcualte square root of SS1-m1 as SQ1
6. calculate square root of SS2-m2 as SQ2
7. calculate denominator as DR as SQ1 * SQ2
8. Calculate sum of column 1 as sum1
9. Calculate sum of column 2 as sum2
10. Calculate product of sum1 and sum2 as TP
11. Calculate Mean product as MP as TP/ N
12. Calculate sum of product of all rows as PS
13. calculate nominator as NR as MP*PS
14. calcualte pearson coefficient as NR/DR
15. return pearson coefficient

Step 7: This is the last step to provide the hybrid recommendation where our system integrates the recommendation provide by the collaborating filter with the recommendation got from the QOS of the system. Here our system check

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

for the valid integrity of the system based on the balanced weight of the two systems and then provide a best recommendation systemThe whole proposed system is expressed mathematically in the below model.

Mathematical Model

1. Let S={ } be as system for Web Services Recommendation
2. Identify Input as I_n={ T_n, O_p, R_q}
   Where T_n= User Transactions
   O_p= User Opinion
   R_q= User Request for recommendation
   S= {I_n}
3. Identify Ras Output i.e. Recommendation
   S= {I_n, R}
4. Identify Process P
   S= { I_n, R, P}
   P= { R_t, P_f, P_{cr}, C_f, M_r, O_p }
   Where R_t = Response time Detection
   P_f =Preference Function
   P_{cr}= Pearson Correlation
   C_f = Collaborative Filtering
   M_r = Merging Recommendation
   O_p = Optimizing preferences
5. S = { I_n, R, R_t, P_f, P_{cr}, C_f, M_r, O_p }

The union of all subset of S Gives the final proposed system.

(A) PREFERENCE FUNCTION

$$\Psi(i,j) = \sum_{V \epsilon N(u)^{i,j}} W_v (q_{v,i} - q_{v,j}) \quad \text{----------(1)}$$

$\Psi(i,j)$ = Preference function
$N(u)^{i,j}$ = Subset of similar user
V= Similar user of the current web service u
W_V = Weighting factor
q_v= Quality of service of two web services

(B) KENDAL RANK CORRELATION SIMILARITY MEASURE

$$Sim(u,v) = \frac{C-D}{N(N-1)/2} \quad \text{----------(2)}$$

Where N = Number of services
C= number of concordant pairs between two lists
D= number of discordant pairs
u,v= Similar users

(C) PEARSON CORRELATION

$$r = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{[N\sum x^2 - (\sum x)^2][N\sum y^2 - (\sum y)^2]}} \quad (3)$$

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
$\sum x^2$ = sum of squared x scores
$\sum y^2$ =sum of squared y scores

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

IV.RESULTS AND DISCUSSIONS

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:

$MAE = \sum_{i,j} | r_{i,j} - r'_{i,j} | / N$

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.

On different trials of our experiment it yields different NMAE as listed in table 2.

No. Runs	Hybrid Recommendation	Collaborative Filtering Recommendation
1	0	0.9
2	3.6	4.5
3	2	8
4	1.8	2
5	1.88	8
6	2.8	16.3

Table 2: Comparative NMAE Values.

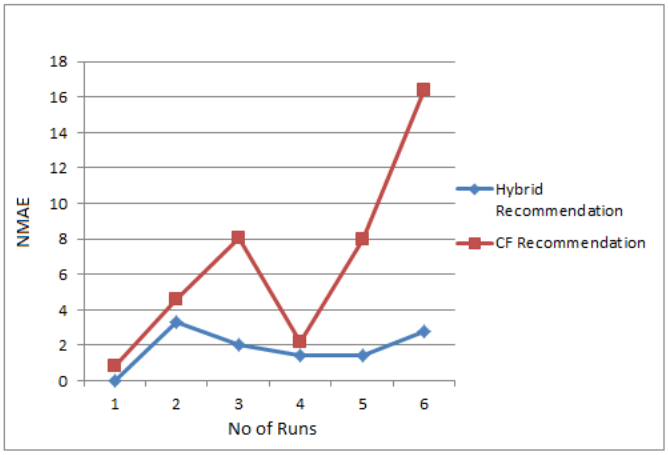

Figure 2 : Performance Graph

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.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

V.CONCLUSION AND 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 to apply Pearson correlation to give proper hybrid recommendation to the user which is the blend of fresh recommendation from the system and collaborative 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

I would like to to express my gratitude to all those who helped me to complete this work want to thank my guide Dr.S.M.Chaware for his continuous help and generous assistance .He helped in abroad range of issues from giving me direction , helping to find the solutions, out lining the requirements and always having the time to see me.
I would like to thank our colleagues who helped me time to time from preparing report and giving good suggestions. I also extend sincere thanks to all the staff members of Department of Computer Engineering for helping us in various aspects.

REFERENCES

[1] Chengying Mao and JifuChen ”QoS Prediction for Web Services Based on Similarity-Aware Slope One Collaborative Filtering”
[2] 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.
[3] 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
[4] Chengying Mao and JifuChen “QoS Prediction for Web Services Based on Similarity-Aware Slope One Collaborative Filtering”
[5] Li Zhang1,2, Bin Zhang1 , Claus Pahl 2 , Lei SXu2 , Zhiliang Zhu1 “Personalized Quality Prediction for Dynamic Service Management based on Invocation Patterns “
[6] Wei Chu “Personalized Recommendation on Dynamic Content Using Predictive Bilinear Models” Social Networks and Web 2.0 / Session: Recommender System
[7] Ms. Anjali Ganesh Jivani “A Comparative Study of Stemming Algorithms”
[8] J. B. Lovins, “Development of a stemming algorithm,” Mechanical Translation and Computer Linguistic., vol.11, no.1/2, pp. 22-31, 1968.
[9] Porter M.F. “An algorithm for suffix stripping”. Program. 1980; 14, 130-137.
[10] Porter M.F. “Snowball: A language for stemming algorithms”. 2001.
[11] Xu Jinxi and Croft Bruce W. “Corpus-based stemming using co-occurrence of word variants”. ACM Transactions on Information Systems. Volume 16, Issue 1. 1998, 61-81.
[12] AnastasiosTombros and C.J. van Rijsbergen,“Query-sensitive similarity measures for information retrieval”
[13] Yung-Shen Lin, Jung-Yi Jiang, and Shie-Jue Lee, Member, IEEE “A Similarity Measure for Text Classification and Clustering “ IEEE
[14]GaddamSaidi Reddy and Dr.R.V.Krishnaiah,” Clustering Algorithm with a Novel Similarity Measure”, IOSR Journal of Computer Engineering (IOSRJCE), Vol. 4, No. 6, pp. 37-42, Sep-Oct. 2012.
[15] H. Chim and X. Deng, “Efficient phrase-based document similarity for clustering”, IEEE Transactions on Knowledge and Data Engineering, Vol. 20, No. 9, pp. 1217 – 1229, 2008
[16] J. Kogan, M. Teboulle and C. K. Nicholas, “Data driven similarity measures for k-means like clustering algorithms”, Information Retrieval, Vol. 8, No. 2, pp. 331–349, 2005.
[17] Survey on Similarity Measure for Clustering P. H. Govardhan, Prof.K . P. Wagh, Dr. P.N. Chatur
[18] Jan Hauke, Tomasz Kossowski” COMPARISON OF VALUES OF PEARSON’S AND SPEARMAN’S CORRELATION COEFFICIENTS ON THE SAME SETS OF DATA “
[19] Zibin Zheng, Hao Ma, Michael R. Lyu, and Irwin King “WSRec: A Collaborative Filtering Based Web Service Recommender System “
[20] Xi Chen, Zibin Zheng, Member,XudongLiu,Zicheng Huang, and Hailong Sun, Member “Personalized QoS-Aware Web Service Recommendation and Visualization”, IEEE TRANSACTIONS ON SERVICES COMPUTING, VOL. 6, NO. 1, JANUARY-MARCH 2013.