MSECA: MICROBLOG SUBSPACE ENSEMBLE CLUSTER APPROACH FOR PREDICTING USER INTEREST ON SEMANTIC WEB LOGS RECOMMENDATION IN WEB MINING

[1]Velmurugan R, [2]Victor S P
[1]Associate Professor, [2]Dean & Associate Professor
[1]Department of Computer Science [PG],
[1]Kristu Jayanti College (Autonomous), Bangalore, India
[2]Department of Computer Science,
[2]St. Xavier's College (Autonomous), Tirunelveli, India

Abstract: Web search engines are designed to serve all users, independent of the special needs of any individual user. The objective of the project is to develop a personalized web search engine which considers users interest and generates search results based on the user's semantic profile from web links. The proposed system utilizes clustering and re-ranking algorithms in order to organize the web documents and provide an order to the results displayed to the user this paper proposed an automatic microblog subspace ensemble clustering algorithm (MSECA) based on entropy for discovering the interest pattern over users' web log with dynamic rank prediction process (DRP). We introduced the information entropy on the basis of clustering algorithm. Compared with traditional clustering algorithms, our method does not require any parameters specified by the end user. Meanwhile, it can discover the clusters in arbitrary shape and size. Experimental results over real-world dataset have fully demonstrated the advantage of our algorithm, which is effective in the problem of high-dimensional and non-informative prior's pattern link prediction. The framework also implies the keywords that help to incorporate user's current interest. Finally, experimental result shows the effectiveness of proposed search engine with commercial search engine in different criteria

***Index Terms* - web mining, clustering, behavior analysis, log analysis, Users' Interests Similarity**

1. INTRODUCTION

With the development of the Internet and the information updating technologies, there are more and more network resources and methods to acquire them. As a result, how to get the part one needs from so large amount of network resources becomes especially important. The personal service has successfully dealt with the problems such as multi information and semantic web links existing in the present Internet. User interest model is the key part of the personal service system and the accuracy of user interest description directly decides the quality of the personal service system from web links.

This incorporates with informational, cultural, social and evidential values to be specific. With the existence of various Search Engines like Google, Yahoo and many more, the users are tend to use them for retrieving their desired information from Web pages. But existing search engine cannot distinguish individual user's request. If user search for index terms as interest on links in search engine retrieves result related to both semantic and lexical terms. The same time if other user who is interested to know interest also get both the result on search term Index. Web users normally issue keyword queries to search engines to fetch relevant information on specific topics, as users may have diverse backgrounds and different expectations for a given query; some search engines try to personalize their results to better match the overall interests and preferences of an individual user. In most of existing search engine, search result is retrieved by evaluating relative importance of links. Ranking algorithm calculate rank of links by their in edge and out edge links. Higher the in edges or out edges, higher the rank is assigned to link.

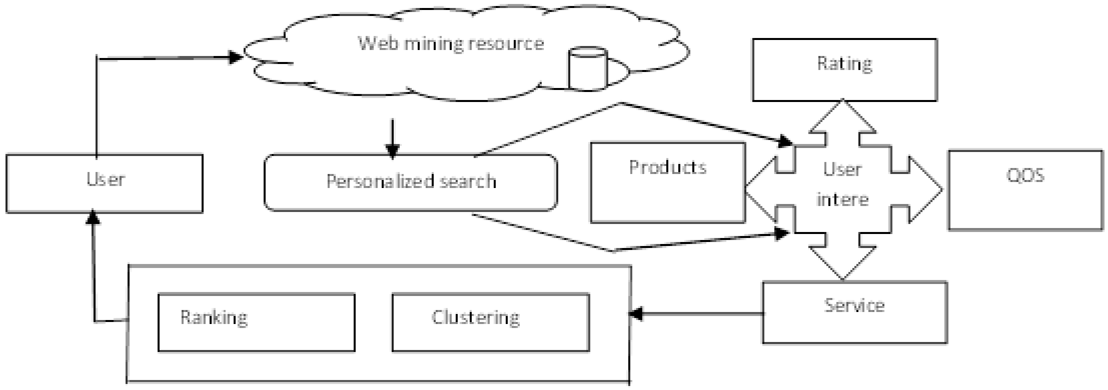

Figure 1: user interest on personalized search

In summary, although these algorithms have different advantages, but generally there are two issues. First, most clustering algorithms require some pre-given parameters such as the number, distance, density, etc. For people who are non-professional areas, it not only increased the difficulty of analysis, but also the quality of clustering is difficult to grasp. Second, many clustering algorithms have good clustering effect only on low-dimensional distribution of particular sample and poor performance on high dimensional cluster sample with some special distributions. These algorithms are unable to find some irregularly shaped clusters and pervasive poor. Therefore, in order to design a better algorithm for user interest pattern mining, we firstly need to enhance the adaptability of the algorithm, which means minimizing the involvement of users. Secondly, the user interest pattern is a kind of high dimension, irregular shape clustering. So it is necessary for the algorithm to find the clusters with any size and shape. To conquer the above issues, in this paper, we proposed a clustering algorithm based on entropy and sub space clusters. The algorithm does not need the user to specify any parameters in advance, and it can automatically clustering sample to identify potential clusters of any shape and size. Experimental results show that the algorithm is effective in the problem of high dimensional and non-informative prior's pattern recognition.

2. Literature survey

The Web became popular in less than ten years and has grown exponentially to an estimated number of pages of over two billions [1]. This exponential growth poses a difficult scalability problem to Web search engines, particularly in the coverage of it interest and also in how to rank reasonably well large answers.

This process is crucial for search engines based on question-answering. Because of the short lengths of queries, approaches based on keywords are not suitable for query clustering [2, 3]. To overcome used a query clustering method that makes use of user logs which allow us to identify the documents the users have selected for a query.

In the context of web search engines, query expansion involves evaluating a user input typed into the search query area and expanding the search query to match additional documents [4,5]. Concept-based user profiling methods that are based on both positive and negative preferences. Exiting methods against our previously proposed personalized query clustering method [6]. Experimental results show that profiles which capture and utilize both of the user's positive and negative preferences perform the best. Personalized agglomerative clustering algorithm that is able to generate personalized query clusters [7]. To the best of the authors' knowledge, no previous work has addressed personalization for query suggestions [8, 9]. Subspace learning is one of the key techniques for feature extraction and dimensionality reduction that are prevalent in many pattern clustering problems.

The process of assigning a class label or mapping a user based on browsing history or on the basis of some other attribute with one of existing class [11]. The output of pattern discovery by is not always used in the form that people can understand so this output is transformed into the form that tedious to integrate. The central problem of defining a measure of complexity [13], specifically for spatial systems. Measure and argue that increasing information is equivalent to increasing complexity about the clusters.

Complex systems defy definition in that they are intrinsically unpredictable and depend on innovations that occur in the future and cannot be forecast [14]. A combination of qualitative and quantitative methods, Initiated construct and empirically validate a gold standard list of lexical features (along with their associated sentiment intensity measures) which are specifically attuned to sentiment in micro-blog-like contexts Collaborative Filtering is a technique used in almost all Recommendation Systems [15,16]. Recommendation Systems are active information filtering systems that attempt to present to the user information items the user is interested in.

The web log mining is the process of identifying browsing patterns by analysing the user's navigational behaviour [19, 20]. The web log files which store the information about the visitors of web sites is used as input for web log mining and pattern prediction process in social products [21]. Web content mining, web usage mining and web structure mining are supportive to cluster parts. As per the researcher's involvement [23], not at all begins our idea of search during search process. Tedious level incorporates the user think of knowledge relevance to put high efforts our search behaviour.

3. Implementation of micro-blog ensemble cluster

Web mining is the intelligent analysis of Web data with Web mining techniques, business organization can gain a better understanding of both the web and web users' preferences to help them run their business more efficiently. One kind of outcome of Web mining is Web browsing patterns. By the use of Web browsing patterns, business organizations can perform mass customization and personalization, adapt their Web sites, and further improve their marketing strategies, product offerings, and promotional campaigns. Therefore, Web browsing pattern mining has special meaning for business organizations. Thus, it has attracted much attention from data mining, machine learning, and other research communities for many years. Of the existed methods, some are non-sequential, such as association rule mining and clustering; and some are sequential, such as sequential or navigational pattern mining. Both approaches ignore the site topology and need to identify user sessions.

To propose an on-demand micro blog clustering framework for analyzing the functional requirements in a web link with dynamic rank prediction. Our approach is novel in that the objects to be clustered capture the domain's action themes at a primitive level, and the essential attributes are uncovered via semantic analysis. Initially micro blog ensemble the cluster to provide automatic support to complement domain analysis by quickly identifying important entities and functionalities. A second contribution is our recognition of stakeholders' different goals in cluster analysis in dynamic rank process, e.g., feature identification for users versus system decomposition for designers. We thus advance the literature by examining requirements clusters that overlap and those causing a minimal information loss, and by facilitating the discovery of product line variability's. A proof-of-concept example is presented to show the applicability and usefulness of our approach.

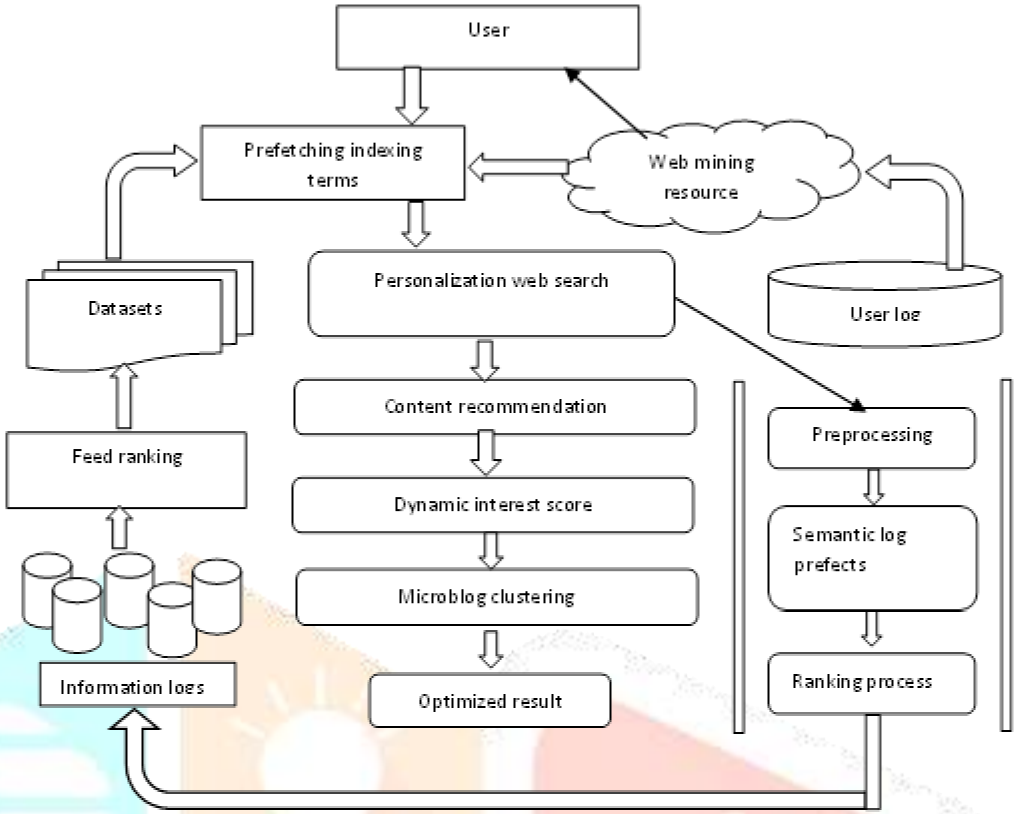

Figure 2 Architecture diagram for proposed system

Clustering analysis to mine the Web is quite different from traditional clustering due to the inherent difference between Web usage data clustering and classic clustering. Therefore, there is a need to develop specialized techniques for clustering analysis based on Web usage data. Some approaches to clustering analysis have been developed for mining the Web access logs.

A) **Content based recommendation**

In this approach, a model of user ratings is first developed, and is used to provide recommendations by comparing representations of content contained in an item with representations of content that the user is interested in. Algorithms in this category use probabilities and calculate the expected value of a user prediction based on the representations of interest of the user through ratings in his profile. In order to provide better recommendations, several approaches combine collaborative filtering with content based recommendation. First ranks web pages through a topic filter and then sends them to user's profiles to obtain relevance feedback. They consider recommendation as a classification problem and use social and Content-Based Information in Recommendations

B) **User recommendation search**

The recommendations implies the parameters are important to take into account for ranking, when considering each of the terms of a retrieved webpage that match the user's profile. We then talk about how to calculate a score for these matching terms based on these parameters and subsequently calculate an aggregate personalization score for each page, based on which the results are ranked.

Thus while these terms are not keywords used to describe a category, they are terms that a specific user frequently associates, and hence considers as important as far as that category is concerned. Thus by including these user specific words we can more accurately define the user's categories of interest. To project micro blog cluster ensemble Algorithm to find frequent search links of words across pages in the user's browsing history and bookmarks. For example a web document can contain 2 types of words. Those that match keywords of categories: k1, k2...kj and also unmatched words: um1, um2, ...umk. Since we want to incorporate user specific words in their respective categories, they must themselves be unmatched words that frequently occur with keywords of categories in the user's browsed pages. Thus we only consider frequent links of the form (um, k1, k2, kj) that contain one unmatched word, frequently occurring with category keywords searched as web links.

3.1 **Preprocessing**

It's a process of the procedure of splitting a search string of written language into its phrases. Textual content statistics includes block of characters called tokens. It's miles used to do away with noisy, inconsistent and incomplete facts. For doing the classification, textual content preprocessing and function extraction is an initial phase. So the files are being separated as tokens and were used for in addition processing. Elimination of stop words are the phrases which are needed to be filtered i.e. can be before or after natural language processing. Forestall words are phrases which contain little informational. Various equipment mainly keep away from to put off those stop phrases a good way to aid phrase search. Numerous collections of words may be chosen as prevent words for any purpose.

Step1: load web link dataset Ds.
Step2: Identify No of index terms N.
Step3: Initialize Occurrence relevant links OM.
Step4: Identify related term of lexical words with link T in Ds.
Step5: for each term links Ii in Ds
For (subspace inter link == Ds)
Noc=Count Number of occurrence in all records linksT.
OM[Ii]=Noc related mean group.
Group similar Li
End for
Store Noc in Occurrence link present in words.
End.
Step6: End.

The information extraction method identifies key words and relationships within the text. It does this by looking for predefined sequences in the text a, an, and…etc of stop words, a process called similar zed relevance case measure among words. The preprocess infers the relationships between all the identified places, people, and time to give the user with meaningful information as index terms on links.

3.2 Microblog subspace ensemble cluster

In the clustering process, the algorithm doesn't require any parameters associated with the cluster, which reduces the involvement of user and the results are more objective. The algorithm makes clustering automatically; the number of invalid cluster samples in the results is usually less than 5, indicating that the algorithm is less sensitive to noise. With the changes of the duration days, the number of effective clusters remains stable, it proving that clustering result of the algorithm has a high accuracy.

Input : Term set Ts, Semantic micro objective O
Output : Semantic Bound Measure SBM, Semantic Closeness measure, micros cluster o.
Start
Read Term set ts.
Read semantic microcluster links term index O.
For each term Ti from Ts
Compute semantic bound measure sbm = Nc/Tn. (5)
Nc = Number of microblog contains Ti index terms
Tn- Total Number of terms present in related links.
End
For each class C
Compute web relative closeness measure scm =
∫ sbm / Number of terms from links in other class
End
Choose the top closure microblog links C = O(Max(Scm))
Identify match case elements Ne = Σ Terms (Ts) ∄O (c)
Stop

Since the clustering process is based on the similarity between stabilities of the clusters from micro subspace ensemble groups, it is possible to find out various classifications with irregular shapes in space of related links.

3.3 Dynamic user interest prediction

Tracking Dynamic Change of Interest In order to do this we have to first define a time period over which we consider the web pages viewed by the user relevant. For our experiment we considered a 90 day period. Thus the user profile will contain contributions from pages considered interesting by the user in only the last search term word from the current day, with contributions from pages before that being removed. Thus if the user's interest were to change over time, the tree would reflect these changes by adding contributions to categories recently viewed and removing contributions from categories no longer found interesting.

3.4 Adding user interest predictive elements

The next step is creating the user specific profile, by adding weights to the categories that are of interest to the user. The user submits pages that he finds interesting through bookmarks, based on which his interests are understood. Since several topics are interrelated, each page may belong to several categories. We therefore carry out a ensemble cluster of each page, and calculate the percentage of relevance to each category. These percentages are the contribution each interesting page makes to the category nodes. Thus the weight to a category over time signifies the extent of the user's interest in the topic. Since these weights change dynamically every time a new web page is added, the current interests of a user can be portrayed in the tree by keeping weights from only those pages that the user has viewed recently. User interest is also introduced into the micro ensemble by another very important technique, in which user specific words that are frequently associated with a topic, other than the category keywords are incorporated into the relevant topic node. These additional keywords that the user frequently associates with a topic, such as names of important people, organizations, or a specialized terminology, etc. can be used to more accurately define the user's categories of interest.

```
Data: inverse category frequency, weight of word given categorist wenblink (wi)
Result: classification of indexing in predicted web link Cj
 Initialization; term ICF as cluster frequency, Term frequent TF ,ensemble cluster weight EW
For each keyword wi in document do
For each category Cj which has wi keyword do
relevance[Cj ] = relevance[Cj ] + ICF(Cj ) * P(wi, C) * TF(wi) * EW(wi);
End
End
```

The micro cluster Algorithm is used to extract these associated words from the user's web history in the form of frequent item sets containing the user specific word with web links and category keywords. The dynamic resemble the feed ranking is later further populated with the user specific words of similar users through collaborative filtering using cluster ensemble micro subspace clustering algorithm.

3.5 Ranking relevance score

General topics are common to a lot of people, while specific topics are much more unique to a person. Thus if a page matches a very specific topic such as natural meaning of interest on web links at level in different stage as tree construction, the page should be scored a lot higher as it matches a user's very specific interest on web links. However if the page matches the much more general topic "sports" at different meaning level as sub tree, then it is not of very specific interest to the user and the match is not as important to a user's interest as a category deeper in the tree.

```
Data: inverse category frequency, relevance of word with category, level of category, category weight Result: rank of document
Initialization;
for each document Dm in results do
    for each keyword wi in document Dm do
        for each category Cj which has wi keyword
            do score[Dm] = score[Dm] + rel(C, wi) * level(Cj ) / max Level InSubtree(Cj ) *  percentage Weight[Cj ];
        end
    end
end
sort Documents();
```

Hence the depth of the matched category is important while scoring the document, and indicates how specific it is to the user's interest, and is given by: S(C) = depth of matched category max depth in sub tree clusters.

3.6 Evaluation Precision, Recall and F-measure analysis

Precision and recall are the 2 measures which are used in information retrieval and pattern recognition to test how well an algorithm works to predict the interest on web links.

Precision =(number of interesting pages retrieved)/(total number of pages retrieved)

Recall=(number of interesting pages retrieved)/(total number of interesting pages)

The traditional F-measure or balanced F-score (F1 score) is the harmonic mean of precision and recall and is used in the statistical analysis of Binary classification, as a measure of a test's accuracy on web link produce user frequent level of access.

F-measure=(2 × Precision × Recall)/(Precision + Recall)

Higher the precision implies better quality of algorithm in retrieving relevant pages whereas recall gives quantitative analysis. An algorithm is better if both the precision and recall values are high.

4. Results and discussion

The results are carried out to test by UCI web link repository with search engine framework. The proposed personalized webs search be implemented using micro blog based clustering algorithm and dynamic rank prediction (DRP) algorithm to deduce a set of related categories for each user query based on the retrieval history of the user session. The proposed method has produces efficient results on micro-blog clustering and improves the performance also. Parameters are tabulated given below.

Table 4.1: Details of Data set

Parameter	Value
Number of service	10
User logs	15
Datasets used	Web resources

Above Table 4.1, shows the details of data set being used to evaluate the performance of the proposed multi attribute opinion rate support measure based approach. The performance of SLPC is evaluated through clustering accuracy (cs), precision rate, and recall rate and time complexity.

Clustering accuracy (cs) $=\sum_{k=0}^{k=n} \times \frac{\text{Retrived number of interest terms cluster(Cds)predictedlinks}}{\text{Total related datsetsTr)from search links}}$

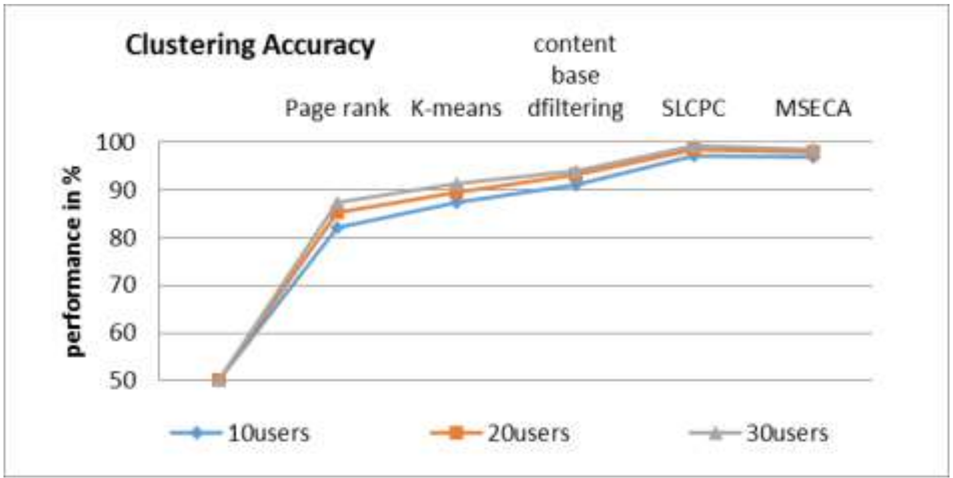

Figure 4.3: Comparison on Clustering Accuracy

Above Figure 4.3, shows the comparison of clustering accuracy and shows that the proposed method has produces higher clustering accuracy than other methods.

Table 4.2 comparison of clustering accuracy

	Impact of Clustering accuracy in %				
Methods/number of records	**Page rank**	**K-means**	**Content based filtering**	**SLCPC**	**MSECA**
10users	82.2	87.3	91.1	96.1	96.8
20 users	85.4	89.5	93.2	97.5	97.9
30 users	87.4	91.3	94.1	98.1	98.4

Above Table 4.2, shows the comparison of clustering accuracy produced 10 users as 96.8%, 20 users as 97.9% and 30 user's s as 98.4 % shows that the proposed approach has produces higher clustering accuracy.

Analysis of precision rate

Precision, Pr is defined as the proportion of total number of relevant URL links and total number of retrieved URL links, where R is the relevant URL links calculated manually and A is the total number of retrieved URL links.

$$\text{Precision, (Pr)} = \frac{Relavant\ links\ (R)}{Total\ number\ of\ retrived\ Links\ (A)} \times 100$$

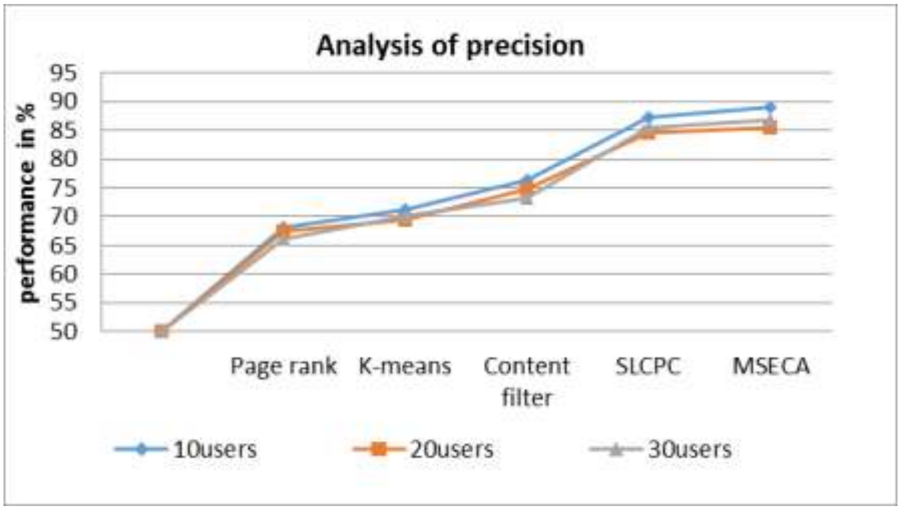

Figure 4.4: Comparison on precision rate

Above Figure 4.4, shows the comparison of precision rate produced by different methods and the proposed method has produces higher performance rate than other methods.

Table 4.3: comparison of precision rate

	Impact of precision in %				
Methods/number of users	**Page rank**	**K-means**	**Content based filter**	**SLCPC**	**MSECA**
10 users	68.2	71.2	76.3	87.3	89.1
20 users	76.4	69.4	74.8	84.6	85.4
30 users	66.2	70.2	73.2	85.5	86.8

The Table 4.3, shows the comparison of precision ratio produced 10 users as 89.3%, 20users as 85.4% and 30 users as 86.8 % shows that the proposed approach produces higher performance ratio.

Analysis of recall

Recall, Rc is defined as the proportion of total number of retrieved URL links within the relevant URL links and the total relevant URL links with paged ranking.

Recall, (Rc) = $\frac{total\ retrived\ from\ relavant\ links\ (RA)}{relavant\ links(R)}$ x 100

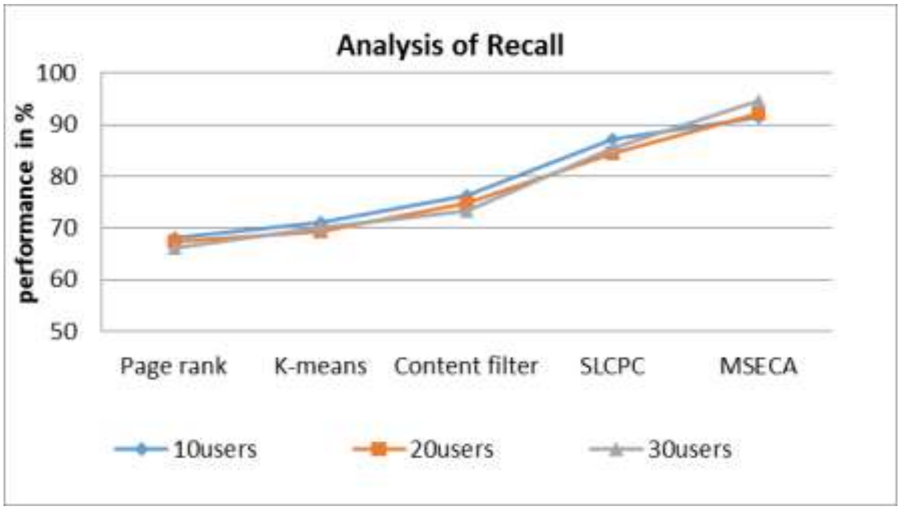

Figure 4.4: Comparison on recall

Above Figure 4.4, shows the comparison of false recall ratio produced by different methods and the proposed method has produces higher performance other methods.

Table 4.5: comparison of recall

	Impact of recall in %				
Methods/number of records	**Page rank**	**K-means**	**Content based filter**	**SLCPC**	**MSECA**
10 Users	68.2	71.2	76.3	87.3	91.3
20 Users	67.4	69.4	74.8	84.6	92.2
30 Users	66.2	70.2	73.2	85.5	94.6

The above table 4.5 shows the comparison of recall page rank value that produce higher performance compared to other methods.

Analysis of Time complexity

$$\text{Time complexity (Tc)} = \sum_{k=0}^{k=n} \times \frac{\text{prediction of clustering Accuracy(cs)+false classification(Fcr)}}{\text{Time taken(Ts)}}$$

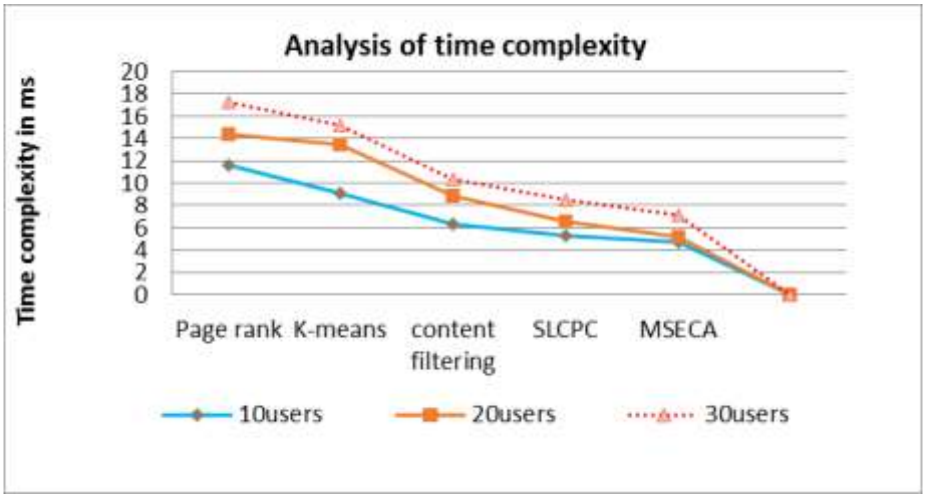

Figure 4.5: Comparison on time complexity

Above Figure 4.5, shows the comparison of time complexity produced by different methods and shows that the proposed approach has produced less time complexity than other methods.

Table 4.4: comparison of time complexity

	Impact of time complexity in mille seconds (ms)				
Methods/number of records	**Page rank**	**K-means**	**Content based filter**	**SLCPC**	**MSECA**
10 Users	11.6	9.1	6.3	5.3	4.7
20 Users	14.4	13.4	8.8	6.6	5.2
30 Users	17.2	15.2	10.3	8.5	7.1

Above Table 4.4 shows the comparison of time complexity proposed prefect clustering produced 10users as 5.2(ms), 20users as 6.6(ms) and 30users as 7.1(ms) shows that the proposed approach has produced less time complexity.

5. Conclusion

In this paper an efficient microblog based semantic algorithm and Dynamic ranking prediction (DRP) been presented to process the personalized web search. The method, preprocess the web resource based on the use interest to identify the most index terms and search logs, opinions. Then the method initializes the cluster with set of points and for each input data point of web links, the method computes the prefect user log support measure. Based on the measure a session time relevant link has been selected and indexed. The same is used to perform intelligence generation and produces efficient results on clustering and intelligence generation. Also the method reduces the time complexity as well. Our proposed system improves the search performance based on the user interest wants to deals time complexity and requirement analyzing the results of the experiments performed; we can conclude that our approach is efficient in reducing the running time without sacrificing the recommendation quality in most cases. This establishes that our method is scalable and it can be used to deal with even bigger datasets.

REFERENCES

[1] R. Baeza-Yates and C. Castillo, "Relating web characteristics with link based web page ranking," String Processing and Information Retrieval, pp. 21-32, 2001.

[2] J. R. Wen, J. Y. Nie, and H. J. Zhang, "Query Clustering Using User Logs," ACM Transactions on Information Systems (TOIS), vol. 20, no. 1, pp. 59-81, 2002.

[3] Z. Zhang and O. Nasraoui, "Mining search engine query logs for social filtering-based query recommendation," Applied Soft Computing, pp. 1326-1334, 2008

[4] H.-r. Kim and P. K. Chan, “Personalized search results with user interest hierarchies learnt from bookmarks,” in Advances in Web Mining and Web Usage Analysis. Springer, 2006, pp. 158–176

[5] C. Biancalana and A. Micarelli, Social Tagging in Query Expansion: A New Way for Personalized Web Search, in Proceedings of the 2009 International Conference on Computational Science and Engineering - Volume 04, 1060- 1065, 2009.

[6] Kenneth Wai-Ting Leung and Dik Lun Lee. ”Deriving Concept-Based User Profiles from search Engine Logs,” IEEE Trans. Knowledge and Data Eng., vol. 22, no. 7, July 2010.

[7] K.W.-T. Leung, W. Ng, and D.L. Lee, “Personalized Concept-Based Clustering of Search Engine Queries,” IEEE Trans. Knowledge and Data Eng., vol. 20, no. 11, pp.1505-1518, Nov. 2008.

[8] Y. Fu, S. Yan, and T. Huang, “Correlation metric for generalized feature extraction,” IEEE Trans. Pattern Anal. Mach. Intell., pp.2229–2235, 2008.

[9] X. Tao, Y. Li, and N. Zhong, “A personalized ontology model for web information gathering,” Knowledge and Data Engineering, IEEE Transactions on, vol. 23, no. 4, pp. 496–511, 2011

[10] K. Jajuga, A. Sokolowski and H. Bock, Classification, clustering, and data analysis: recent advances and applications: Springer Science & Business Media, 2012.

[11] Poonamkaushal, “Prediction of User’s Next Web Page Request By Hybrid Technique”, International Journal of Emerging Technology and Advanced Engineering, Vol.2, Issue.3, pp.338-342, 2012.

[12] H. Jinjing, "Mining Users Potential Interested in Personalized Information Recommendation Service," Journal of Modern Information, vol. 4, pp. 7-11, 2013.

[13] M. Batty, R. Morphet, P. Masucci, and K. Stanilov, "Entropy, complexity, and spatial information," Journal of geographical systems, vol. 16, pp. 363-385, 2014.

[14] C. J. Hutto and E. Gilbert, “VADER: A parsimonious rule-based model for sentiment analysis of social media text,” in Proceedings of the Eighth International Conference on Weblogs and Social Media, ICWSM, Ann Arbor, Michigan, USA, June 1-4, 2014

[15] K. Abhishek, S. Kulkarni, V. Kumar, N. Archana and P. Kumar, "A Review on Personalized Information Recommendation System Using Collaborative Filtering," International Journal of Computer Science and Information Technologies (IJCSIT), vol. 2, no. 3, pp. 1272-1278, 2011

[16] D. Zhou, S. Lawless, and V. Wade, Improving search via personalized query expansion using social media, Information Retrieval, 2012, 15(3-4): 218-242.

[17] M. R. Bouadjenek, H. Hacid, and M. Bouzeghoub, Social networks and information retrieval, how are they converging? A survey, a taxonomy and an analysis of social information retrieval approaches and platforms, Information Systems, 2016, 56: 1-18.

[18] B. Bakariya, G.S. Thakur, "Effectuation of Web Log Preprocessing and Page Access Frequency using Web Usage Mining", International Journal of Computer Sciences and Engineering, Vol.1, Issue.1, pp.1-5, 2013

[19] A. Vishwakarma, and K.N. Singh, “A survey on web log mining pattern discovery”, IJCSIT – International Journal of Computer Science and Information Technologies, pp: 7022-7031, 2014.

[20] S. Parvatikar, B. Joshi, “Analysis of user behavior through web usage mining”, ICAST – International Conference on Advances in Science and Technology, pp: 27-31, 2014.

[21] C. Biancalana, F. Gasparetti, A. Micarelli, and G. Sansonetti, Social semantic query expansion, ACM Trans. Intell. Syst. Technol., 2013, 4(4): 1-43.

[22] Marie Fernandes, “Data Mining: A Comparative Study of its Various Techniques and its Process”, International Journal of Scientific Research in Computer Science and Engineering,Vol.5, Issue.1, pp.19-23, 2017.

[23] A. D. Kasliwal, and G. S. Katkar, “Web Usage mining for Predicting User Access Behavior”, IJCSIT International Journal of Computer Science and Information Technologies, Vol. 6 (1), 2015.

[24] V. Neha, Patil et al, “Prediction of Web Users Browsing Behavior: A Review”, International Journal of Computer Science and Mobile Computing, Vol.4,pp: 209-212, 2015.

[25] S. Khan, Y. Singh and A. Kumar. Sachan, “Web Mining Approach in Analysing User Behavior and Interest for Website Modification”, International Journal of Advanced Research in Computer Science and Software Engineering, Vol 5, 2015.