

# An Hybrid Approach for Suitable Content Navigation Based on Weighted Clustering for Online Users

Suma H.C  
Dept. of CSE,  
VTU, CIT, Gubbi,

Harish T.A  
Assistant Professor, Dept. of CSE,  
VTU, CIT, Gubbi

Girish L  
Assistant Professor, Dept. of CSE,  
VTU, CIT, Gubbi

**Abstract** – Due to exponential growth of several contents which is created on the Web, Recommendation procedures need to be developed is becoming critical. Inestimable diverse classes of recommendations are done on the Web each day, which includes pictures, song, imageries, files recommendations, suggestion of queries, recommendation of tags, etc. No restriction that what categories of data sources are handled for the recommendations, basically these sources of data can be demonstrated in the custom of numerous categories of graphs. This paper provided the Common structure on mining the Web graphs for References using a hybrid method for appropriate content navigation built on weighted clustering for online users. First we recommend a collaborative filtering method to acquire the appropriate content from web records and creating the weighted clusters of related items for the user created queries. Secondly centered on the click through data study we achieve the interests and hidden semantic associations between customers and queries also queries and clicked Web data. And based on this clickthrough information the bipartite graph will be produced, which represents the relationship amongst the queries and URLs. Lastly we recommend a new query suggestion standard, personalized query recommendation created on weighted clustering for online users.

**Index Terms**-- Recommendations, query suggestion, collaborative filtering

## I. INTRODUCTION

Due to diverse and volatile development of Web data, it has become more critical to establish and exploit the data successfully and efficiently. Also the information which is developed by the customers is formless and freestyle, so it is impossible to mine the necessary data from the data sources. In order to full the needs of customers of the web by providing them essential information the recommendation method is organized extensively in the business.

The most effective method is collaborative filtering in order to design the recommender schemes, where the items will be recommended to the customers based on comparable tastes and preferences liked in the previous customers. [10]

Even though it's successful, but the presentation of Collaborative filtering approach is strongly inadequate by the lack of data caused from: (i) the vast amount of items far beyond customer's capacity to estimate even a slight portion of them; (ii) customers do not have preferences to rate the purchased or viewed items. The initial one is the uncertainty

which usually occurs in the regular language. Requests having ambiguous languages and it might complicate the systems and it does not accomplish the information requirements of customers. [11]

The next task is to implement the personalization structures. Personalization is vital because diverse users have diverse data needs. The adopting the personalization will clean out the unrelated data to an individual, and it will also offer more specific data that is progressively suitable to an individual's benefits. The last challenge is time consuming and ineffective to enterprise diverse recommendation processes for diverse recommendation jobs. Basically, most of these recommendation complications have common methods, where a centralized structure is required to combine the recommendation jobs in Web.

Aim at resolving the difficulties examined above, to propose a common structure for the recommendations on the data of Web. This structure is made upon the hybrid approach for suitable content navigation based on weighted clustering and collaborative system combination for the online users.

It is a general method, and it can be used to numerous recommendation tasks on the Web.

- It can propose hidden semantically related results to the original data needed.
- This model suggests a regular action for personalized suggestions.
- The scheduled suggestion algorithm is available to very huge data collections.

## II. LITERATURE REVIEW

### A. Traditional Recommendation Approaches

**Content-based filtering:** Content-based suggestion schemes work on user profiles which is formed initially. A profile contains the information about a customer and his taste centered on which stuffs he favors the most. Usually, when constructing a profile, recommender schemes make an analysis, to get preliminary information about a customer in demand to overcome the new-customer problem. [1]

In recommendation procedure the engine compares similar tastes of the similar users, and by that analysis the items will be recommended to the users. Content-based recommendation schemes typically use identifiers or

keywords for effective and improved filtering. Going into depth of approaches of collaborative filtering are classified into: customer based, model based and item-based methods.

**Collaborative filtering:** Collaborative filtering developed one of the greatest investigated methods of recommender schemes since this method was stated and defined by Hal Varian and Paul Resnick in 1997. [2] The impression of this is to discover the customers in a public that share obligations [3].

If multiple customers have identical or nearly similar rated stuffs in common, then they will be having comparable tastes. Such customers form a cluster or a so termed as neighborhood. A customer gets references to select things which are not appraised previously, but that were previously positively appraised by customers in their neighborhood. Collaborative filtering is broadly adopted in e-commerce. Recommendations can be done to the users and it can be books, songs, images etc. [4]

**Hybrid recommendation methods:** For improved outcomes some recommender schemes associate diverse methods of collaborative methodologies and content centered methods. Using this we can avoid certain restrictions and difficulties of pure recommender schemes. The grouping of methods can progress in diverse ways: 1) implementation of different algorithms and assembling the outcomes. 2) Develop some guidelines of content filtering in collaborative method. 3) Employ some guidelines of collaborative methods in content centered methodology. 4) Generate a united recommender scheme which brings together both methods. Robin Burke driven out the classification of hybrid recommender schemes which categorizing them. [5]

**Cross-lingual Methodologies:** The recommender scheme created on cross-lingual method lets the customers obtain recommendations to the stuffs that have explanations in languages they don't communicate and recognize. Yang, Chen and Wu introduced a method for a cross lingual summary collection reference. The foremost hint is to plan together text and keywords in diverse languages into a particular feature space, that is to say possibility dissemination over hidden topics. From the explanations of things the scheme analyses keywords than converts them in one well-defined language by vocabularies. After that, by collaborative or additional filtering, the scheme gives recommendations to customers [8].

### III. EXISTING SYSTEM

#### *A. Challenges and Issues of Existing Recommendation Approaches*

The previous challenge is ineffective and time consuming to propose diverse recommendation methods to diverse recommendation jobs.

Additionally, maximum of existing approaches are complex and need large Amount of constraints. In this paper, directing at resolving these difficulties, we suggest a universal framework for the suggestions on the web data as deliberated below. The earlier technique is for only slight quantity of data and this idea is too accomplished for bigger set of records.

#### *B. Issues of existing system*

**Scalability:** Due to quick growing of web data users and the items, the method requires huge amount of resources to process the necessary information and to recommend the appropriate items to web customers. Only to determine the users activities and there similar behaviors most of the resources is wasted. This drawback can be overcome by combining multiple types of filtering techniques and physical advancement of the systems.

**Trust:** The people with some short profiles may not be that relevant as the peoples of those who have rich profiles. The matter of faith rises towards appraisals of a definite user. The difficulty might be resolved by distribution of priorities to the customers.

**Privacy:** The most critical problem is privacy. In order recommend accurately and correctly, the system must include full information about the user. Naturally, the query of trustworthiness, safety and privacy of the given data rises. Many online systems propose effective security of confidentiality of the customers by using specific algorithms and databases.

**Cold-start:** To recommendations to new customers is very problematic when there profile is almost empty and they haven't preferred any items so there nature is mysterious to the scheme. This is so-called the cold start tricky. In certain systems this difficulty is resolved with review when generating a profile. Things can also have this problem once they are fresh in the scheme and haven't remained preferred earlier. Both of these complications can be similarly resolved with hybrid methods.

**Scarcity:** In online works that contain an enormous quantity of customers and things there are nearly always users that have preferred just a few things. By collaborative and extra methods recommender schemes usually create areas of users consuming their profiles. If a customer has estimated just few things then it's tough to define their taste and they might be connected to the incorrect neighborhood. Scarcity is the difficult of absence of evidence. [6]

### IV. PROPOSED SYSTEM

In this paper we implement one method define, examine, implement and promote the frequently used technique for web mining i.e. An hybrid approach for suitable content navigation based on collaborative system, weighted clustering and query suggestion combination for the online users.[4]

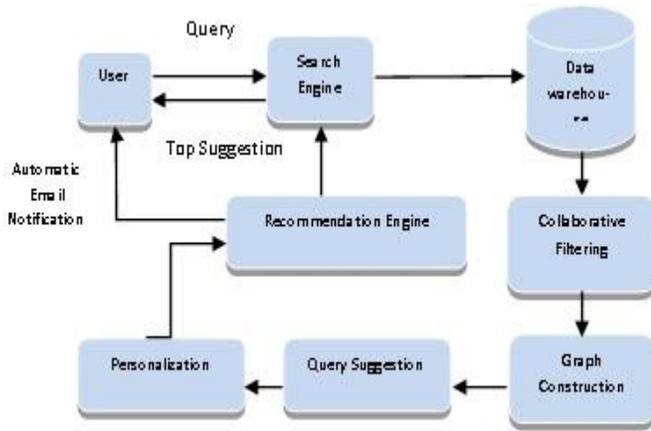


Figure 1: Recommendation System Architecture

This method can be straightforwardly used in recommendation schemes and it is scalable. This technique gives high accuracy, and it gives weight to the pages which are visited by the customers. On the basis of the number of times the user visited the weight will be assigned. Approach includes providing the weight is clickthrough data analysis, and based on that weight graph is constructed and it is used for appropriate query suggestion.

### V. MODULE DESCRIPTION

The modules included are:

- Collaborative filtering
- Click through data analysis
- Query suggestion
- Automatic email notification

The user enters the query in the search to obtain the essential data. Built on the customer query by mining the huge web data source some related data will be filtered. The technique used to filter the required data referred as collaborative filtering.

#### A. collaborative filtering

User based and item based approaches are the two types of collaborative filtering. To predict the preferences of the customer's user based approach is adopted. Similarly item based approaches used to identify the preferences of an active customer based on the other similar items which he prefers. [7]

Following figures displays the customer and item based approaches correspondingly:

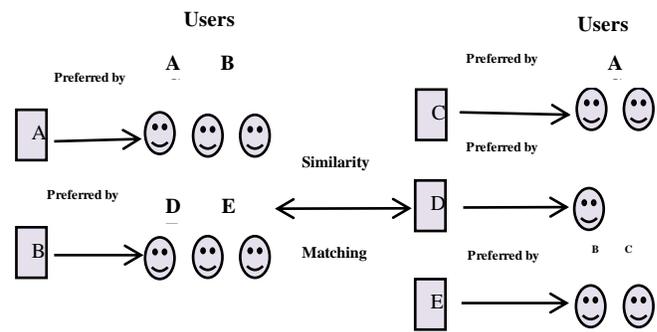


Figure 2: User based approach

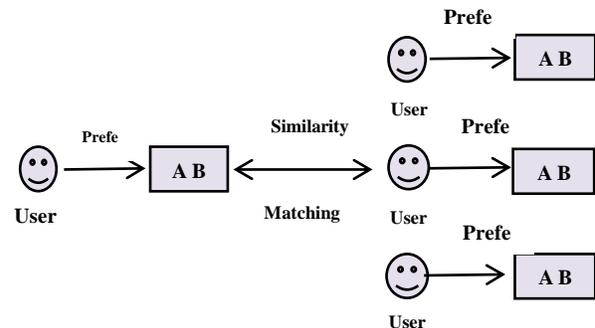


Figure 3: Item based approach

This approach will gain higher performance than other methods. The filtering of the data sets mainly based on two concepts.

- Posting the opinion
- Rating Prediction

**Posting the opinion:** Based on the opinions from various users about trade, import/export and products through the online, the filtering of appropriate data items can be done successfully. Opinion refers to posting the remarks about the modules and its characteristic which is preferred by the customers. Both positive and negative remarks will be included in this. [8]

**Rating the prediction:** The filtering of the items also based on this rating concept, where the customers will rate the items by analyzing it properly. Rating also depends on user's satisfaction on that preferred items. So on the basis of the ratings of the active users the collaborative filtering can be done. [8]

#### B. Clustering

Collaborative filtering of appropriate information is can be achieved successfully based on the posting the opinion and rating the prediction of the active users in the recommended system. Clustering is jobs of combination a set of items in such an approach that items present in the identical cluster are more related than the items existing in the another cluster. [9] Similarly here clustering of items can be done based on active users opinions and ratings so that it will be very useful for the different users who exists in the recommended system. By analyzing the clusters the new users can prefer the items efficiently. The method of clustering is shown below:

### C. Query Suggestion

The process of recommending the most similar queries to the customers of the web, a successful efficient method, and query suggestion has been adopted by certain commercial search engines. Query recommendation purpose is to recommend full demands that have been expressed by earlier customers so that query reliability and consistency are conserved in the recommended queries. [11]

The technique of query recommendation will comprise of several varieties of data sources like documents, scripts, etc. for such recommended queries. This suggestion must fully satisfy the customers.

ID	QUERY	URL	NO. OF CLICKS
120	facebook	http://www.facebook.com	199
120	facebook	http://en.wikipedia.org/wiki/Facebook	20
1020	Apple iphone	http://www.apple.com/iphone/	170

Table 1: Clustering

### D. Click through data analysis

In this click through data analysis method, the main aim is to improve the web search results. Click through data analysis included to identify the most popular sites or products. The process is analyzing the active user's activities. The activities can include the preferred items frequently by the users and the number of sites which they are visited. How many times the customer clicked the same links based on their query and how many times they preferred same items are recorded, also how much the active user is satisfied can be analyzed based on their opinion and ratings. This process is referred as click through data analysis.

ITEMS	RATINGS
Apple	Best
Sony	Best
Samsung	Best

ITEMS	RATING	ITEMS	RATINGS
Micromax	good	MTS	poor
Nokia	Good		
Carbon	Good		

Table 2: Example of click through data

The number of clicks obtained through implementing the efficient query suggestion algorithm.

### E. Graph construction

To construct the Query-URL graph, consider the query set  $Q=\{q_1,q_2,q_3,q_4\}$ , URL set  $L=\{l_1,l_2,l_3\}$  and edge  $E=\{q_i,l_j\}$ . The edge  $E=\{q_i,l_k\}$  can be generated if and only if when the customer  $u_i$  clicks the URL  $l_k$  after receiving the query  $q_i$ . By analysing the click through data the bipartite graph will be generated.

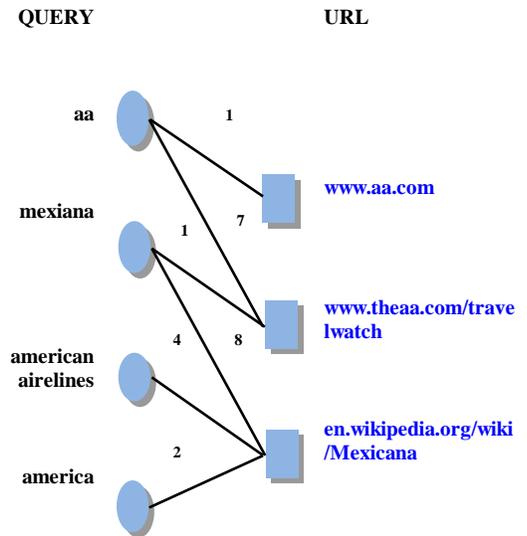


Figure 4: Graph Construction

The values existing in the edges states how many periods the customer clicked the URL. Finally through this relevant information the suggestions can be done to the new users who visit the recommended systems. The recommendation process can be implemented using the most popular method that is query suggestion. [10]

### F. Query Suggestion Algorithm

After the creation of the graph the query suggestion algorithm can be easily designed.

**Algorithm 1.** Query Suggestion Algorithm

1. A constructed graph  $G = (V^+ \cup V^*, E)$  comprises of query set  $V^+$  and URL set  $V^*$ . The ends are assigned some values by the click through data analysis technique.
2. For a given request  $V^+$ , graph will be created by depth first search. When related to the predefined quantity if the sum of requirements is high the search will be finalized.
3. As observed above, set preliminary value  $\alpha = 1$ , and without damage of simplicity, set the initial cost of query  $q$
4.  $f_q(0) = 1$ . Start the procedure using  $f(1) = e^{\alpha R} f(0)$ .
5. Outcome of Top-K requests with the major values in vector  $f(1)$  as the recommendation.

### G. Query Suggestion Results

The proposed method to suggest the query is through click through data analysis. A sample for query suggestion outcome is that if user requires a method like java the references such as virtual machine, sun micro system are suggested and then future the particulars of the enterprise name are recommended.

In such a way the suggestions for our associated test requests like name, the distribution of the item and the

particulars of a domain are all formed. It is clearly represented that the query proposal result produced by our technique is as worthy as those created by a general search engine. [9] The below table represents the query suggestions results:

Queries	Suggestions		
	Top1	Top2	Top3
Java	Sun java	Java download	Virtual machine
Apple	Itunes	Ipod	Apple ipod
Fitness	Excercise	Fitness Magzine	Women fitness

Table 3: query suggestion results

It shows our recommendation method not only propose the queries that are related to the trial queries but also delivers hidden semantically related outcome for the user query .

### ***H. Automatic email notification***

Users will logout from the recommended system after satisfying by obtaining the relevant information. The present data of the recommended system will be updated simultaneously through the frequent update of ratings and opinions of the existing users. So by analyzing the overall update progress of the products the automatic email notification will be sent to the particular users who preferred the particular products. So they the customer will get the updated day by day and very effectively useful to them to get efficient and reliable information about the products. The automatic email generation can be adopted using the SMTP protocol.

### ***I. Advantages of the proposed system***

- It is a universal technique, and it can be employed to numerous recommendation jobs on the Web.
- It will offer latent semantically appropriate outcomes to the original data need.
- This system delivers a usual dealing of personalized suggestions.
- The designed recommendation method is accessible to very bulky datasets.

## **VI. CONCLUSION**

In this paper, we suggested a new query suggestion methodology based on the hybrid combination of the collaborative filtering, clustering and query suggestion. This is an overall structure which can fundamentally be implemented to peak of the Web graphs for the suggestions jobs, such as suggesting the query, modified recommendations, email notification etc. The produced suggestions are semantically associated to the feedbacks of the active users. The experimental analysis on numerous huge scale data sources of the web shows the favorable.

## **REFERENCES**

- [1]. Dubois, D., Hüllermeier, E. and Prade, H., “A Systematic Approach to the Assessment of Fuzzy Association Rules,” Data Mining and Knowledge Discovery Journal, Vol. 13, No. 2, (2006), 167 – 192.
- [2]. Paul Resnik, Hall Varian, “Recommender System”, Communication of the ACM, Vol 40, No.3, 1997
- [3]. Gabor Takacs, Istvan Pillaszy, Bottyan Nemeth, and Domonkos Tikk, “Scalable Collaborative Filtering Approaches for Large Recommender Systems,” The Journal of Machine Learning Research, vol. 10, pp. 623-656, 2009
- [4]. Salvatore Orlando, Raffaele Perego and Claudio Silvestri, “A new algorithm for gap constrained sequence mining”, In Proceedings of the ACM Symposium on applied computing, pp. 540 – 547, Nicosia, Cyprus, 2004.
- [5]. Robin Burke, “Hybrid Recommender System :Survey and experiments”, User modelling and interaction
- [6]. B. Sarwar, G. Karypis, J. Konstan, and J. Reidl, “Item-based collaborative filtering recommendation algorithms,” in Proceedings of the 10th international conference on World Wide Web, ser. WWW '01. New York, NY, USA: ACM, 2001, pp. 285–295
- [7]. J.S. Breese, D. Heckerman, and C. Kadie, “Empirical Analysis of Predictive Algorithms for Collaborative Filtering,” Proc. 14th Conf. Uncertainty in Artificial Intelligence (UAI), 1998.
- [8]. M. Kanimozhi, K. Logeshwari, K. Saranya, D. Sowbha, Mr. R. Velumani, “Clustering For Collaborative Filtering Application in Web Recommendations”, International Journal of Computer Science and Mobile Computing, IJCSMC, Vol. 2, Issue. 4, April 2013, pg.217 – 222
- [9]. Mrs. Shweta Shinde, Prof. Dr. V.H. Patil, “Mining Web Graph For Query Recommendation”, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 2, Issue 3, May – June 2013
- [10]. Rupesh Patel, Gaurang Diwakar, “Web Graph of Queries for Recommendations”, International journal of advanced research in computer science and management studies, volume 3, issue 4, April 2015.