

Recommendation System for Web Mining: A Review

Purvi Dubey
RGPV
Indore
India

Pramod S. Nair, Ph.D.
RGPV
Indore
India

ABSTRACT

Today the huge amount of information is available online due to the acceptance and understanding of the possibilities of internet. This reason makes the World Wide Web as an important research area. In this paper we review some of the recommendation systems. We discuss by looking at its merits and demerits accordance to the applications.

Keywords

Information retrieval, web, information extraction.

1. INTRODUCTION

World Wide Web is very prevalent and important central for information retrieval in current age. The World Wide Web is holding huge amount of data at times this will lead to the problem information overload [1]. The different other problems that may be categorized as below:-

1.1 Finding Relevant Information

The search engine is used to search any information on the internet. The user input a keyword as a query and the search take place accordance to the keyword and display with the link to the pages where the keywords are found. The display of the page links is based on the result of the page rank algorithm.

1.2 Low Precision

The main reason for low precision is the irrelevancy data in the resulting pages of the search. The second problem is the disability of indexing all the available information on the internet. This problem is called as low recall. Due to the no indexing of the relevant data the search leads to low precision in the search result related to the query. This is related to knowing what customer exactly wants from their given query.

The recommendation systems are type of filtering system that suggest people which product to choose, recommend entertainment item like video, image, book, music, or people (dating site, twitter, face book). There are mainly three techniques of recommendation system, content based recommendation system, collaborative filtering recommendation system and hybrid recommendation system. The gravity recommendation system automatically learn user

browsing or shopping behavior on a given web site or platform. It uses collaborative filtering technique to improve quality of recommendation system by learning the test of user.

In this paper part II gives an oversight of web mining and information retrieval, web mining types and techniques. In part III we describe recommendation system, overview, content based recommendation system, collaborative filtering recommendation system and hybrid recommendation system and part IV concludes the paper.

The Web is a collection of inter-related elements and the method of extracting the coal or other fuel from earth is called mining, so the data mining is technique for retrieval information from web.

1.2.1 Information retrieval

Information retrieval is a process of automatically extraction relevant document while at same time extraction of non-relevant document. [3]. the main goal of information retrieval is indexing the text which is related to query and find useful document in collection. Latest information includes modeling, data visualization, filtering, document classification, document categorization e.t.c. [4].

1.2.2 Information extraction

Information extraction covert the collection of document into an information that is easier to understand and access [5]. The main aim of information extraction is to select relevant fact from document as unlike in the information retrieval used to select relevant document [6]. The main difference is the

Information extraction views the structure of representation of document while information retrieval views only text in document [7]. Information processing is like a preprocessing step which is held after information retrieval process and before data mining technique performed. Information extraction can use to modify the indexing process which is actually a work of information retrieval process.

1.2.3 Web mining classification

Figure1. Can show the classification of web mining technique, there are mainly three type of web mining:

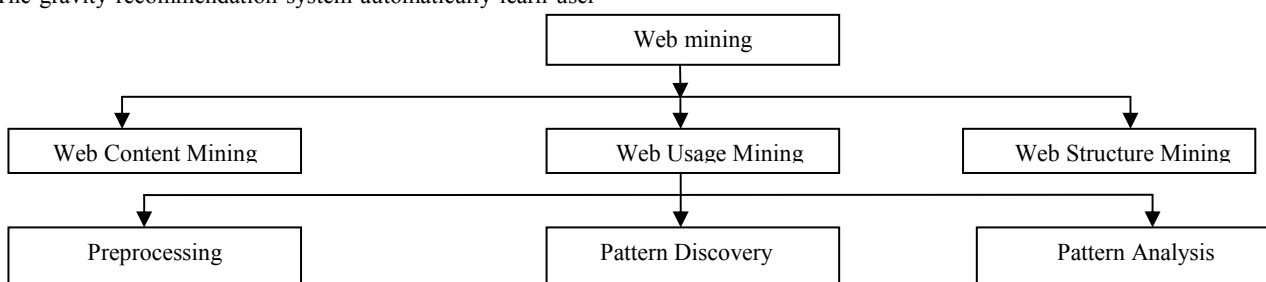


Figure: 1 Classification of Web Mining

- (i) **Web content mining:** - The web content mining can be affined to data mining. the mining of extraction of information from the content of the document. So it is called web content mining. Web content mining is related to data mining and text mining but having some difference also.

It is affined to data mining due to most of data mining techniques are applied into the web content mining. It is related to web text because the contents are texts. It is different in context of generally data used in web are mainly semi-structured or unstructured data, and data which is used in data mining generally structured data. It is also distinguish to the text mining in context of semi structured nature of web while text mining focused on unstructured texts.

- (ii) **Web structured mining:** - Web structure mining is used to identify the relation between web page linked by information. This linked connection pull data for search engine which is related to a search query.

Web structure mining mainly face two of problem is because of vast amount of data are:-

- (a) Irrelevant search results: - Relevance of search results are miss achieve due to search engine allow low precision criteria.
- (b) The second is lake of index, due to vast amount, and then the recall is low amount in web content mining.

The main purpose of the structured mining is to extract unknown relation between web pages.

- (iii) **Web usage mining:** - It is also called web log mining. The web usage mining is a process of retrieving pattern from web access log. The three processes which are used in Web usage mining is Preprocessing, Pattern discovery & analysis.

1.3 Data Preprocessing

This is the mosti important task of web usage mining. The below part will described the preprocessing part which is described in review paper. [8]

1. **Data fusion:** - in this step merging the log file from different servers. In case of different server must add the name of server and time clock of file which belongs to the server. All the files are then un-named due to privacy issue [25].
2. **Data Cleaning:** - The data cleaning process means removing the inconsequential data from database log. This process is used to reduce the size of data, if a request sent for a web page there are various attributes called and many contents loaded for that particular request. The result can be image and graphics, as we loaded for the HTML tags as our requirement, we only need data which is generated by user, and it is necessary to make sure that user request exist in server logs. So task which is generated by user should be ignored and remove from log files [26].
3. **User Identification:** - This step is after the data cleaning step. This is the main task for data preprocessing process. In this must identify the unique users from data log. If data log shows many user having same IP address, by tracking the agent in which the request have been

produced. Still differentiate for obtaining the different user sessions. If two users have common IP address then both of users considered as a single user [27].

- (i) **Session Identification:** - It is clear by the name, session identification indicates the number of time a user visited a web page. According to the time cracy, if we consider 30 minutes time limit for visit a specific web page, if any user is stay on that page more than 30 minutes then session will divided into sub sessions. This problem deals with the problem of caching from browser.
- (ii) **Path completion:** - As the name indicates the path means the pages visited from start to end. The main problem in these missing entries which deceive in the tracking request. But with the help of referrer one can easily set details what path is followed by proper tracking of web page request.
- (iii) **Formatting:** - This process is held after the data cleaning and the data preprocessing task. It is performed on access log. The preprocess data is basically formatted as need of algorithm of data mining, which are applied to retrieve information from access data.
- (iv) **Data summarization:** - The data summarization is most important process and it is last step of the data processing phase, in which data is stored in a relational database system for analysis and access in future. It reduces the dimensionality of data.

1.4 Pattern Discovery and Pattern Analysis

The pattern discovery is used after the data cleaning and user identification session from web access log. The analysis of preprocessed data is very useful to them who will perform business on the web. The tool used in pattern analysis is based on data mining algorithm, psychology. To pattern analysis different kind of algorithms are used association rule mining, path analysis, sequential pattern, clustering and classification.

1.4.1 Techniques of web mining

- (i) **Page rank:** - Page rank work by calculate the number of quality of link to a page for presumes how important a link or website is The definition of page rank is give: - we suppose a page has n pages (T1.....Tn) and the parameter d is damping factor (set from 0 to 1). The rank of page is given by:-

$$P_R(A) = (1 - d) + d \left[\sum_{i=1}^n \frac{P_R(T_i)}{P_R(T_i)} \right] \quad (1)$$

Where page rank PR (A) can be calculated by a simple iterative algorithm and it is correspond to the principle eigenvector of normalized link matrix of web. N is the number of document we have, link matrix is M, where the Mij entry is 1/ni, if there is a link form document j to document i, then we can compute the page rank on the graph which is dominant eigenvector of matrix A.

- (ii) **Clustering:** - The clustering is a process in which search result is merging to a particular group according to typed query. This helps us to enter in to a document quickly.

(iii) *Markov Model* :- The markov model is oldest technique of web mining, in which describing the series of possible events and where the possibility of each event is depend upon only on the state which already got in the past event.

(iv) *Association Rule mining*: - This is very popular method for discovering relationship between variables in large database. It is used to identify the strong rules discover according to interestingness.

2. RECOMMENDATION SYSTEM

2.1 Oversight

Recommendation systems are defined as a software tool and techniques which providing advice for item to a user. The suggestions are like what music to listen, what online news to read etc. Recommendation system is used for finding the needed information from wider information available on the internet. Recommendation system mainly uses three approaches content based recommendation system, collaborative filtering recommendation system and hybrid recommendation system.

2.2 Content based Collaborative Filtering

The content based generally used the textual information such as- document, websites and movie description. The recommendation item is based on comparison between the user profile and the context of document. Item is represented as a set of terms. The user profile can be made by a analyzing the context of item which has been seen by user. The term can choose automatically or manually, when a term is assign automatically a method has to be chose that can extract from item. The term must assign like in which the user profile and item can compare in meaningful way. A learning algorithm is used which is based on seen items and can recommendation based on profile of user. Learning algorithm can be genetic algorithm or neural network. Neural network are slow as compare to another algorithm because many computation required in it. The relevance feedback and Bayesian classifier is faster in terms of speed.

An approach in which is hidden markov model (HMM) and classical content based collaborative filtering combine to propose a new algorithm called collaborative filtering recommendation using HMM [12].

The main concept is collaborative filtering obtain by information filtering and information retrieval. Content based recommended system is based on the user similarity rather than comment of item. Content based recommended system focused on making item and user profile, user profile contain information about users:- tastes, preferences and need and an item based profile contain information about the set of attribute of item. After that the profile of user and each profile of all item and recommended item according to preference of users.

1) *Hidden markov model*: - A HMM defines by a set of sets and alphabet of output symbol. Output symbol can be varying according to state transform from one to another. First of all initialize data from www.groupelens.org and collect from [http:// movielens.umn.edu](http://movielens.umn.edu). Then apply k-means algorithm with hamming distance, for example:- if hamming distance is between 1011101 & 1001001 is 2, then we use an algorithm called k-means similarity distance standard for similarity distance and similarity distance standard to cluster. After that train for each cluster which is stored in cluster results by applying

forward backward algorithm. Estimate the probability of each “string” from HMM by using forward algorithm then save the results in the cluster, repeat the process from step no. 2 to step no. 3 until cluster result no edit any more. According to this process combine movie using HMM after a user seen a movie can be recommended other movie to that user.

The main advantage of this algorithm is useful for the problem of information overloading. The very important application of recommended system used in e-commerce which can include video, books etc according to the interest of user. Another advantage is we can calculate the probability of one sequence produced by a HMM identifiably.

Main disadvantage of this approach is the average precision of algorithm is not good. And it is very tough to apply an all purpose recommendation algorithm for different backgrounds [12].

Tag based recommendation system uses the open directory project data about web page design an recommendation system which can suggest most important web page for each user which is based on profile of user and web page information from odp.

Tagging is used to produced recommendation of user, the main steps include first user modeling, as a primary step have to form an user model and used it to differentiate between users. For profile construction we utilize the tags. Next step is producing user profile from data which is tagged; and we will use the naive approach [13]. In this approach tag data is collected in a way that indicates the interest of the user accordingly to the user identity. Then calculate a weight $W(u_j, t_k)$ for each user (u_j, x_k) for find interesting tags for user and then filter it basic of recommendation system. The profile can be shown as follows:-

$U_j = \{(t, w_{1j}), (t_2, w_{2j}), (t_k, w_{kj}), \dots\}$ (1) where t_k is a tag, which is used by u for bookmark a web page.

Second step is web page modeling in order to choose relevant tags of web page have to calculate which $w(p_i, t_k)$ for each paper (p_i, t_k) by user for odp tags. The profile of every page p_i include:- all tags which are marked for bookmark p_i .

Profile representation as follows:-
 $P_i = [(t_1, w_{1ij}), (t_2, w_{2ij}), \dots, (t_k, w_{ki})]$ where t_k is tag used for mark p by user u and $w_{k,i}$ is the weight of t_k for web page p_i and, weight is calculate :-

$W(i_l, t_j) = TAF(t_j, i_l) + Iif(t_j)$, where t_j and I_l represent i th web page of user and item tag respectively. $W(i_l, t_j)$ is calculated for all pair we form and $TAF(t_j, t_1)$ and $IB(t_s)$ are considered as not of tag occurrences.

The last step is recommendation ne items we have to find tag of web page which is

$Wp(trel) = \{(I_1, trel, Wrel, 1), (I_2, trel, Wrel, 2), \dots, (I_3, trel, wrel, i)\}$.

The main advantage of this approach is improve the performance of prediction and remove the drawbacks of search engine are not personalized.

The main disadvantages are we it depends on matching of words so as a basic step we have to make a user model and used it to differentiate between users [14].

2.3 Collaborative Filtering

The collaborative filtering build a model from user's past behavior after that it used the model to predict items. This technique imaging a user list of n user $u = \{u_1, u_2, u_3, \dots, u_N\}$ and a item list of m items $I = \{i_1, i_2, i_3, \dots, i_m\}$, and every user rated an item; the main aim is to predicting unwanted items and suggest top items (n) to the user.

Two approaches of collaborative filtering are user based & item based collaborative filtering. User based utilizes the similarity computed between active users and all other users. Item based collaborative filtering used the parallelism between two items. The measurement is like if two user who gave close rating to an item then it can be calculated as 1, and if two user who gave different ratings to same items, it can be calculated as 0.

The definition of similarity is given by Pearson co-relation coefficient:-

$$sim(u, v) = \frac{\sum_{i \in I_{uv}} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I_{u,v}} (r_{u,i} - \bar{r}_u)^2 \sum_{i \in I_{u,v}} (r_{v,i} - \bar{r}_v)^2}} \quad (1)$$

Where I_{uv} set of item rated b user u and user v. u and v are users. r_{ui} is defined as user rating u to item I and \bar{r}_u and the \bar{r}_v is the average rating provided to by user v and user u respectively.

There are mainly two algorithms which are used in item based collaborative filtering which is known as memory based algorithm and algorithm which are used in item based collaborative filtering called model based algorithm.

2.3.1 Memory based collaborative filtering algorithm

it use full database to make a prediction. After that statical method are applying. Prediction rating r is given by

$$r_{u,i} = \bar{r}_u + \frac{\sum_{v \in N(u)} sim(u,v) * (r_{v,i} - \bar{r}_v)}{\sum_{v \in N(u)} sim(u,v)} \quad (2)$$

Where user v belongs to the neighborhood of user u. and I_{uv} set of item rated b user u and user v. r_{ui} is user average rating to item I and \bar{r}_v and the \bar{r}_u is the average rating provided to by user v and user u respectively.

2.3.2 Model based collaborating filtering

The model based collaborating filtering, usually implementing a model of user ratings by technique of machine learning techniques like clustering, rule based approach and bayesian network.

The paper investigates the similarity calculations by adjusting the positive and negative similarity and transforming similarities in social network. The process is first similarity calculation which can by pearson co-relation us widely used to find similarity between two variables. It range from -1 to +1, many pearson correlation used the user similarity batter [15]. Second step is adjusting the pcc similarity, the method to adjust similarity is:-

$$Sim(u,a) = \begin{cases} sim(u,a)+1/2, & sim[u,a] > 0 \\ 0, & sim(u,a) = 0 \\ 0, & sim(u,a) < 0 \end{cases}$$

The similarity will be (0,5,1) and the negative similarity will be [0,1]. If the user similarities between two user is grater then 1 then they have similar connection.

The next step is transforming similarities in social network the system of social network tries to have user to find user means user having the similar tastes. Om order to make the similar relationship of user the user graph $G(v,e)$ where v is the user and edge is this similarity between users. And then prediction is used in which user based collaborating filtering make new prediction by find users with similar ratinf to the active user and weighted combination of ratings. After that top recommendation suggest to the user in which list of item to active user. Clustering the item that relevant to the choosen item.[16]

The main advantage of this approach is the prediction and accuracy is good as compare to traditional method. And similarity relation make similar user can be found as neighbor [17].

Another approach in which the collaborative filtering technique and search based technique used to generate recommendation for items unlike as product attributes uses given by user which will apply a collaborative filtering to find out similarity user means who have smellier behavior in past , based on product they have choose in past. The two method are used which are CFF round robin and CFAg query technique the CFF round robin technique combine the collaborating filtering with round robin [17]. Second method is CFAg query in which first derive more detailed products attribute based on preference of neighborhood and then a new query derive by aggregates the derive product attribute. The main advantage is it is used for unrated product to recommended to the user such as :- caars, houses and electronic device) another one is both proposed techniques gives batter result compared to collaborative filtering and basic search approach.

The main disadvantage is the result shows only of most viewed products. And frequency based technique give worst result as compare to model based collaborative filtering [19].

A technique for producing personalized web page recommendation. First, from user browsing log, we recognize users who find the interesting pages before others, we called them early adepter [30]. In the previous approaches, the cluster of similar users has been used. In this approach, the log to build a weighted, directed graph. In the graph, node represents user and edge represents that the user on both node access the same page.

The model encodes the latent temporal patterns on the basis of user visits to the web pages. We found that some works are faster and better in finding new pages. According to browsing activities of users, we can recognize the new pages early and we can suggest these pages to the other users whose browsing activity are similar as early adaptor. The model is based on information networks for example: - In twitter, a user follows other users. In this network, people encourage each other by post and re-post of messages (either short or long). Much research has been developed to finding influential user and distinguish parameter by user encourage each other.

2.4 Hybrid Model

There are some limitations to the content based recommended system and collaborative filtering recommended system so we have to merge the characteristics of content based recommended system and collaborative filtering recommended system, which is called the hybrid model.

- (i) *The cold star problem*: - It is also known as new user problem, suppose if a user is new, then it is difficult to suggest any item, because no one item is rated by the

user. So it is difficult to recommendation item to new user.

- (ii) *The problem with new item:* - Collaborating filtering having a problem with new item. A new item even if having a high rating, it can't be recommendation because it is not rated by maximum number of user.
- (iii) *Sparsity of data:* - Collaborative filtering having a main technical limitation, it means the web having a large number of item, so the user who are very active can give rating to few items available in database. Popular item can be rate, by only a few numbers of users which is called data sparsity problem.
- (iv) *Feature extraction:* - content based recommendation system not supported the feature extraction.

The Demographic recommender system, in this categorization of user and item based on particular attribute and after that recommendation process is based on demographic categorization [20]. The approach is for a unique hybrid recommendation approach by merging rating feature and demographic information about items. It is usually based on cascading hybrid recommendation technique [21]. Which build a model of item based on rating given to the item, features and demographic information evaluation of algorithm by data set of movie lens and film –trust.

Step:1 Calculate similarity between items by rated data, demographic and feature data after that save this information, adjust cosine similarity [22] between two or more items and is used for estimating the similarity of rating data and vector similarity [23] between two item that is used for estimating similarity using feature vector and demographic data.

Step:2 Find boosted_{sim} which is defined as:-

$$\max = \text{argmax } u(m_i, n_j) \quad \forall m_i \in M^I, \quad \forall n_j \in N^T$$

Where, the equation maximize utility means reducing in mean absolute error of all vector (M^I) over set of items (M^J) in the training set.

Boosted similarity boosted_{sim} is given by a combination of F_{sim} , R_{Dsim} and D_{Dsim} over the set of items as follows:-

$$\text{Boostedsim}(nt, ci) = F_{sim} + \beta \times R_{Dsim} + \gamma \times D_{Dsim}.$$

Where α and β represent relative impact of similarity and $\alpha + \beta + \gamma = 1$ without the generality loss.

Step: 3 estimating the $P_{ma, nt}$ for an active user r_{naun} target item n_i .

$$P_{ma, nt} = \sum_{i=1}^k \frac{\text{Boostedsim}(nt, ci) \times r_{ma, ci}}{\sum_{i=1}^k (|\text{Boostedsim}(nt, ci)|)}$$

The Problem related with this system is scalability, data scarcity, Cold-star problem which affect coverage and poor quality recommendations. This approach results state of art recommendation system algorithm and eliminate recorded problem with recommended system. The algorithm gives batter result as compared to state of art recommendation algorithm.

A hybrid recommendation system [24] produced which calculate item attribute similarity, user rating similarity, user demography similarity and temporal information. It used hybrid structure system and model based system and memory based system, model based system and temporal information used for recommendation process to recommend faster.

The hybrid model combines the memory based and model based collaborating filtering algorithm, firstly it make model of a user by combining rating and attribute similarity, temporal and demographic information. Using the neighbor selection and recommendation based on the rating matrix, current time and temporal hybrid model.

In this approach firstly identify user identity the neighbor of user in time t and then use neighbor set to produce recommendation:-

(A.) *Estimating the neighbor set:* - It is related to compute similarity to user. Several similarity algorithm are available like Pearson correlation, cosine vector similarity, in this approach used hybrid model which is temporal aware, and then it calculate similarity of features vector of user a and user b in the time t.

$$\text{Sim}(a, b, t) = \frac{\sum_{k=1}^{p+q} |(Hakt - Hat) \times (Hbkt - Hbt)|}{\sqrt{\sum_{k=1}^{p+q} (Hakt - Hat)^2 \times \sum_{k=1}^{p+q} (Hbkt - Hbt)^2}}$$

Where, Hakt is the interest score for feature k in time t for user a and user b. Hat and Hbt is average rating of user a and user b with time t.

(B.) *Rating can be predicted as follows:-*

$$P(a, j, t) = \left\{ \frac{Rat \sum_{i \in u} \text{sim}(a, i, t) \times (Rijt - rit)}{\sum_{j \in u} \text{sim}(a, j, t)} \right\}$$

Where, P(a, j, t) is rating prediction for item i with time t. And u is nearest neighbor set for user a at time t. R_{it} is rating of neighbor J for content i and Rat is rating of user a in the time t.

The approach deal with scalability, sparsity and cold star problems. And produce batter result as compare to existing recommendation system. The accuracy is improved by memory based collaborating filtering and the scalability is improved by model based collaborative filtering. Computation time is also low in this approach.

The geographic location is not mention and time stamping has to improve in this approach.

2.5 Applications of recommendation system

Recommendation system is used in many applications. In the author describe a technique where a geographical location can be traced on the basis of user movements [9] and Helping user to selecting the traveling behavior, by use of available online information. A touristic-area-season modeled and developed in social event [10]. twitter also use TWITOBI a recommendation system for suggest topk-user follow and topk tweet to read web page recommendation [11] can be described as a application in which a graph based approach is used to discover interest of topic.

3. CONCLUSION

We have discussed about web mining and its type, techniques. We also mentioned about the recommendation system its type and we have described the demerits of the content based and collaborative filtering recommendation system that can be removing by combining the techniques which is called hybrid model. We have to discuss the application of recommendation system.

4. REFERENCES

- [1] P.Maes."Agent that reduce work and information overload"Communication of ACM,30-40,1994.
- [2] C. Chakrabarti, "Data mining for hyper text: A tutorial survey" ACM MIGKDD Explorations, 1-11, 2000.
- [3] C.J rijsbergen, "Information retrieval" buttor-worths, 1979.
- [4] R. Baeza-yates and e. Berthjer ribejro-neto "Modern information retrieval" Addison –Wesley Longman publishing company, 1999.
- [5] J. Cowie and W. Lehnert, "Information extraction", communication of ACM, 80-91, 1996.
- [6] M.T. Pazienza, editor, "Information extraction: A multidisciplinary approach to an emerging information technology", Springer, volume 1299 of lecture notes in computer science information, 1997.
- [7] Y. Wilks "Information extraction as a core language technology", Springer, volume 1299, 1-9, 1997.
- [8] Vijayshri. Losawar, Dr. Madhuri Joshi, "Data processing in web usage mining", ICAIES, Singapor, July 15-16, 2012.
- [9] Quercia, D.; Lathia, N.; Calabrese,F.; Di Lorenzo, G.; Crowcroft, J., "Recommendation social events from mobile phone location data ", IEEE 10th conference on, vol.,no.,971-976,13-17 Dec,2010.
- [10] Qi Liu; Young Ge; Zhongmou Li; Enhong Chen; Hui Xiong, "Personalized Travel Package Recommendation ", IEEE 11th conference on, vol.,no.,407-416, 11-14 dec, 2011.
- [11] Qingyan Yang; Ju Fan; Jianyong Wang ; Lizhu Zhou, , "Web page Recommendation via collaborating filtering and Topic aware Markov Model", IEEE 10th International conference on, Vol.,no.,1145-1150,13-17 Dec. 2010.
- [12] Huj ci;fejcai;zhifany ciao "content based filter recommendation algorithm using HMM " IEEE fourth international conference, aug 2012.
- [13] Emichlmayer and s.cryzer "Learning user profile from tagging data and leveraging them for personalized information access"workshop on tagging and metadata for social information organization, 237-240. May 2007.
- [14] Davood,F.g; fate,;o."tag based recommendation system for social bookmarking sites " IEEE/ACM international conference, 26 aug 2012.
- [15] D.billsus m; pozzani "Learning collaborative filtering information filter " 15th conference on learning, 45-53,1998.
- [16] J> breese, D.heclreman and C.ladie "Empherical analysis of algorithm for collaborating filtering" 14th conference on uncertainty in artificial intelligence, 43-52,1998.
- [17] Jing wang; jian Yin," Enhancing accuracy of user based collaborative filtering recommendation system in social network " International conference on (ICJEM) 20-21, 2012.
- [18] K.srikumar and b. Bhonker "Personalized product selection internet learing" journal of electronic commerce vol;5,nov,216-227,2004.
- [19] A bdollan N; yue Xo, "Integrating collaborative filtering and search based technique for personalized online product recommendation " IEEE 11th conference on 11-dec-2011.
- [20] Ghazan, M.A, Prugel-Bennett, "A scalable, accurate hybrid recommender system" knowledge discovery and data mining, Third international conference IEEE, 9-10 Jan, 2010.
- [21] R. Burke, "Hybrid recommender system : survey and experiments" user modelling and user adapted interaction ACM, vol.12, no.4,331-370,2002.
- [22] B. Surwar, G. Karypis, T. Konath and J. Reidl, "Item based collaborating filtering recommendation algorithm" in proceedings of the the 10th international conference on www.ACM, New york, NY, usa, 285-295,2001.
- [23] J.S. S. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," Morgan Kaufmann, 43-52, 1998.
- [24] Ullah, F; sarwar, G.; Suns chong Lee, Yun Kyung Park; Kyenong Deok, Moon; Jin Tac Kim, "Hybrid recommendation system with temporal information" international conference on IEEE, 1-3 feb,2012.
- [25] Sanjay Tiwari, Renu Tilwani, "Web mining and pre processing of web usage data" international journal of advanced research in computer science and software engineering, vol.3, issue 10,629-634, October, 2013.
- [26] Aye, T.T., "Web log cleaning for mining of web usage patterns" computer research and development (ICCRD), 3rd international conference IEEE, vol.2,490-494, 11-13, march 2011.
- [27] Priynaka Patel, Mitixa parmar, "A Review on user session identification through web server log" International journal of computer science and information technology, India, volume 5(1), 146-148, 2014.
- [28] V. Chitraa, Dr. Antony selvadoss thanamani, "A novel technique for session identification in web usage mining processing" international journal of computer application, vol.34, 2011.
- [29] V. Chitraa, Dr. Antony selvadoss thanamani, "An efficient path completion technique for web log mining " international conference on computational intelligence and computing research, India, 2010.
- [30] Ida mele and franceco bonch "The early adeptor graph and its application to web page recommendation" ACM, pp.1686,2012.