Review of Web Recommendation System and Its Techniques: Future Road Map

Vidya Waykule, Prof Shyam S. Gupta

*Computer Sc. & Engineering Dept, Siddhant College of Enggineering ,
Sadumbare ,Chakan Talegaon Road,Taluka Maval,District Pune 412109*

Abstract-Recently the use of web recommendation techniques is growing worldwide with aim of providing the customized required data to end users. Different recommendation methods and solutions impose many research challenges to researchers. Web recommendation techniques are divided into two main types such as content based web recommendation system and collaborative web recommendation system. Basically web recommendation systems are building simultaneously with web. This system at the start based on the demographic, collaborative and content based filtering. At present such systems containing the social information. For the future, such systems will use the local, personal and implicit information over the Internet. In this paper, the main goal is to present the review of web recommendation systems and different techniques. Web recommendation systems recently suffered from different limitations. The current web recommendation techniques need to be more efficient in order to overcome the limitations of them.
Keywords: *Web Recommendation System, Content based system, collaborative recommendation system, association rule mining, sequential pattern mining.*

1. INTRODUCTION

Now days accessing of web data is daily used process in end user life. The log data is generated in which each end user attempts are collected over the web server. Such web log data having the customer address, customer name, date, time, server name, server IP, status code, method name, and includes many features like page name [1]. Users click events are organized in this work. Web site visitors and developers to find interesting web pages and recommended for directions to the facility must analyze historical information. The researcher's active in the field of data mining is an important issue for research. To get around this problem, the Web log mining plays an important role in web recommendation system [2]. Exploration and analysis of user behavior with the help of web log mining and data mining techniques to discover interesting patterns are an application. The web log mining process includes the tasks like data collection, preprocessing, and pattern discovery and pattern analysis is a process for as many tasks. Web site development for every human being is incredible. The change log data into business intelligence technology is used. Each phase has several important functions. In the first phase, data collection is collected separately server log files and convert them into a common file format. The second step, the full data from raw log data, performs preprocessing to find [17]. Such third stage

classification, clustering and association mining techniques applied to many pattern discoveries. It is help to collect a major pattern for web detection [9] [10]. Finally during the final phase patterns are counted and hence the best patterns are available.

In this paper we are presenting the review over the use of data mining concepts and advanced association rule mining algorithm. More specifically we are presenting the survey over web recommendation methods, different data mining based web recommendation methods, and related works etc. In next sections we are discussing the related works presented over web usage mining and web recommendation systems in section II, in section III survey of different types of web recommendation systems. In section IV different data mining techniques those are presenting for web usage mining.

2. RELATED WORKS

The concept of web usage mining is playing main role for identifying the web page requirements of end users through the web server. Generally the end users want to find the right web pages within the short duration of time. So the need of demand, the development is required to forecast the correct web pages from the web. Many techniques applied to the analysis of web log data, but researchers have been attracted by ARM. Preprocessing is for Web Usage Mining works basis. Preprocessing methods discussed the importance of this work; various techniques are compared and identified. Preprocessing techniques to preprocess a complete extraction of user patterns, web log files are proposed [1]. Data cleaning algorithms irrelevant web log files and remove entries from the log file filtering algorithm discards unselfish characteristics. Users are able to identify the session. Sanjay Gandhi et al also a full stream of data preprocessing techniques proposed for use. The preprocessing stage and search log data is collected from different data sources are used before meaningful patterns. Web mining valuable information from secondary data derived from user access logs. It is important for web site organization, improve business services, personalization web traffic and web recommendation. Web usage mining divided into three different phases and these are planned. Big web traffic data calculated & applied to web mining techniques for discovering an interesting pattern useful from traffic analysis.

During the 1993 by Agarwal et al [2] the new method introduced called ARM which is one of the best strategies

used to find out correlated item sets purchased together frequently. Frequent patterns are patterns that often appear in the data set. For example, a set of items such as milk and bread that are appear frequently together in the transaction data set.

Using expert knowledge Ali Mirza et al [3] proposed a pruning approach. This technique is used to enhance and maintain the level of accuracy and dramatically reduces the size of the tree. This technique is used to enhancing and retaining the level of accuracy and dramatically reduces the tree size. They handle large data sets and eliminate wrong decision. Many ARM variations such divisions, Tree guess as proposed by researchers over the years are Growth etc. Markov model FP Association mining algorithm is not the order of transactions. A navigation of web pages in this proposed order [4] [5]. Efficiently and accurately using the Poisson distribution algorithm frequent item sets discovered in a large database is uncertain. This process re-execution of the algorithm is eliminated and supports incremental mining. Symmetrical association rule mining method is a comprehensive educational web log data. Symmetrical mining, the support and confidence of association rule mining only calculate factors; they also evaluated interesting measures like lift, correlation or conviction. Other relevance measures such as Chi square, Cosine and odd rules and found a positive correlation between items that were set to learn.

In 2000 [8], the weighted association rule mining method first proposed by Wang et al. Depth-first search method and the pages they visit and maintained order to improve prediction accuracy mining frequent items is set using graph traversal path. A weighted sequential pattern mining algorithm Joong et al also proposed. This procedure is identified the time interval and maintains the navigation order that leads to produce interesting sequential patterns. Steaming Association Rule (SAR) mining algorithm combined the weighted association mining and divide-and-conquer technique. Compared to traditional mining algorithm the author has to improve prediction accuracy, rule accuracy and reduce the database scanning time. It is handled large log data and eliminated redundancy that leads to navigation order of web pages.

The Mohbubul et al. [7] proposed a weighted access pattern tree (WAP-tree) algorithm to access patterns effectively from log data. WAP-tree recursively calculated the trees to reconstruct an intermediate trees i.e. starting with suffix sequences and ending with prefix sequences.

Avrillia et al. [4] proposed a new sequential mining algorithm called FLAME (Flexible and Accurate Motif Detection). This method is addressed several frequent constraints.

Shrivastva Neeraj et al [9] proposed a new integrated method Closed Sequence generator mining (CSGM) to combine sequential generators and closed sequential patterns. This algorithm scanned the database only once and discovering non-redundant sequential association rules from sequential datasets with high accuracy, small memory and time.

3. WEB RECOMMENDATION METHODS
3.1. Traditional Recommendation Methods
3.1.1. Content-Based Filtering
As we defined in abstract of this paper as well, this method is works using the profiles of users that are created at the beginning. The end user profile has information about a user and his taste which is based on how user rates the items. Recommended procedure, the engine is already positive he did not rate the items rated by the user with the items and looked for equality comparison. Positive rated ones that are most similar to those goods will be recommended to the user. Most of recommendation systems based on content filtering locksmith efficient and better use tags or keywords [10]. In this case the profiles of other users are not required and the recommendations are based on personal information, because they do not affect user recommendations. User based, item based & model based approaches: Collaborative filtering methods are going into the details; we can distinguish the most popular approach [10]. Figure 1 showing the basic diagram for web recommendation system as architecture.

3.1.2. Collaborative Filtering
In 1997, this method mentioned Paul Resnick and Hal Varian. This was described by collaborative filtering recommender systems since become one of the most researched techniques. The idea of collaborative filtering is a community that shares the users praises it. Two users have the same or nearly the same in common rated items, they have similar tastes. A user he / she will not be rated in front, but that's already positive his / her neighborhood is rated by the users to select the item becomes recommendations. Collaborative filtering is widely used in e-commerce. Customer rate books, songs, movies and in the future can get recommendations about those issues. Collaborative filtering over some documents (scientific works, articles, and documents among journals) is used in the browsing [11].
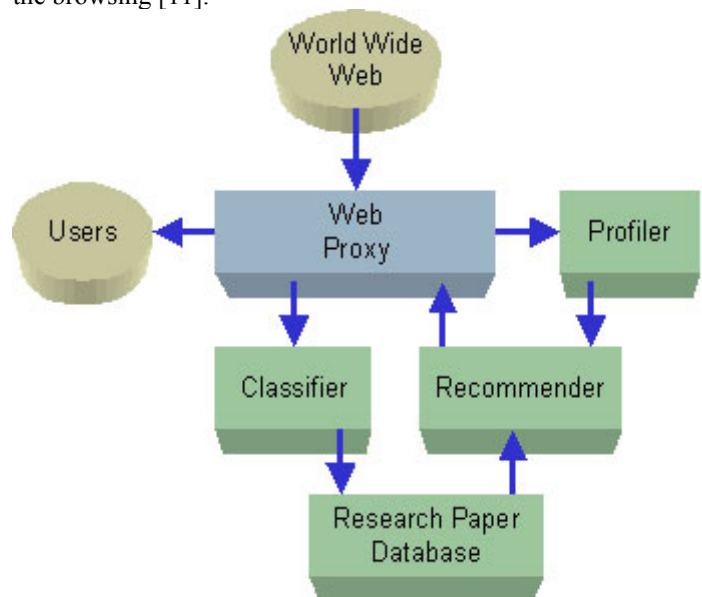

Figure 1: Architecture of web recommendation system

3.1.3. Hybrid Recommendation Approaches

This is nothing but the combination of content based method and collaborative method. For better results some recommender systems combine different techniques of collaborative approaches and content based approaches. The combination of approaches can proceed in different ways [13]:

- Involvement in implementation of another algorithms & results;
- Content based filtering rules in the use of collaborative approaches;
- Content based collaborative filtering approach in the use of certain terms;
- An integrated approach that brings together both create recommender system.

3.2 Modern Recommendation Methods

3.2.1. Context-Aware Approaches

End user environment context as well as status information about the information that he/she is under what circumstances such description alone in ratings is not detailed information about the items rating more recommendations from can user's play a crucial role Were returned by get. Some recommendations are more appropriate for a user in the evening and in the morning does not match her preferences and it is out of the heat when it is cold and entirely another when he/she want a thing. Association governance intra transaction relationship between sequential patterns represent relationships while transactions. Sequential patterns often resulted in a specific order that item sets a scene, a set of same item all items in a single transaction when price is considered or are within a time interval [13]. The recommender systems that pay attention and utilize such information in giving recommendations are called context-aware recommender systems [12].

Locksmith recommendation of context aware systems are one of the biggest problems is the reference information. Formed in the user / her to fill out a form asking him to interact with and create a survey that can be received clearly. The rating review process without making it complicated to obtain information regarding the most desirable.

3.2.2. Semantic Based Methods

Most of the item description, recommender systems, and in the rest of the web users are presented in the web as a literal. A very important part of understanding the structure of the text is recommended that's why. Traditional text mining by a computer or a recommender system, a user may not under stood that the textual description and syntactic analysis shows that based approach. Analyzes were based on the meaning that there was a reason to make a new text mining technique. Recommender systems with such techniques are called semantic based recommender systems [14].

3.2.3. Cross-Domain Based Methods

Finding similar users and collaborative recommender systems neighborhood is a precision manufacturing process is an important part of the recommendation. The similarities of two users based on their appreciation of items are discovered.

3.2.4. Peer-to-Peer Methods

Recommender systems are de-centralized P2P approach. Each peer group with similar interests relates only to other peers and can get recommendations from users of that group. A peer recommendation can be given on the basis of history. Recommender systems can decentralize the scalability problem.

3.2.5. Cross-lingual Methods

Based approach to cross- lingual recommender system users do not understand the language they speak and describe the item to get recommendations for places. Yang, Chen and Wu recommends an approach to a cross -lingual news group purposed. The main idea is to say that latent topics also feature a probability distribution, in various languages in space to map the text and keywords.

3.3 Challenges and Problems of Recommendation Methods

3.3.1. Cold-Start Problem: When you create a profile in a few recommender systems have solved this problem with the survey. They are new in the system and when the item has not previously been declared a cold start it can be. Both of these problems can be solved with the hybrid approach.

3.3.2. Believe: With a brief history of the people's voice in their rich history, which is as the voice of those that may not be relevant? The issue of trust arises to evaluate a particular client. The problem can be solved by the users for the distribution of preferences.

3.3.3. Scalability: With the increase of number of users and items, and recommendations for the formation of information processing systems need more resources. This problem also filters and systems are solved by combining different types of physical improvement. Many parts of the computations in order to accelerate the assurance of online recommendations can be applied offline.

3.3.4. Sparsity: Users and a large amount of items that online stores those users have rated only a few items are almost always there. Collaborative recommender systems using other methods to access their profiles, users typically create neighborhood. If a user has rated only a few items, it is very difficult to determine his taste and he / she may be wrong neighborhood.

3.3.5. Privacy: The most important issue is privacy. To get the most accurate and recommendation systems, demographic data, and the data about the location of a particular user with the most amount of information possible about the user, should receive. Naturally, the information's reliability, security and privacy questions arise. Many online stores by using special algorithms and programs provide effective protection of users' privacy [14].

4. EXISTING WEB RECOMMENDATION TECHNIQUES

In this section we are discussing the different techniques presented for efficient web usage mining such as Association Rule Mining, Weighted Association Rule Mining, Sequential Pattern Mining and Dynamic Programming techniques. We are discussing the features and limitations of these methods in this section.

4.1. Association Rule Mining

Association Rule Mining is one of the most important techniques in web mining. Association rule mining technique is used to find the frequently visited web pages from the user access sequences and constructs a set of rules based on those visits.

The ARM has two separate phases:

(i). to find the frequent item sets

(ii). Determine the rules form these item sets.

Let the D with a separate transaction database. D = (P1, P2, Pn} be a set of n distinct Web pages a Union is a P2= P1 ⇒ implications where P1= P1⇒ P ⊂ p, P2 and P1 (P2) = ∅ ∩ P2 or P2 where P1. A Web page is called as a result page is called the antecedent page P2. Rule interestingness in support, is measured by the confidence and lift [17].

Both support and confidence are fractions between. The support is a measure of statistical implication, whereas confidence is used to measure the strength of the rule. The rule is said to be "interesting" if its support and confidence are greater than user defined thresholds, Support minimum and Confidence minimum respectively. A pattern gets a score of 1 if it satisfies both of the threshold conditions and gets a score of 0 otherwise. The goal is to find all rules with a score of 1. Confidence alone may not be enough to assess the descriptive interest of a rule. A rule with high confidence occurs with chance. Those types of rules can be contacted by determining whether the antecedent and consequent are statistically independent.

$$Lift(P1 \Rightarrow P2) = \frac{Confidence(P1 \Rightarrow P2)}{Support(p2)}$$

It ranges within [0, and] values close to 1 indicate that P1 and P2 are independent and rule is not interesting. P1, P2 evidence so far about the value of 1 provides information that indicates. These three factors are used to determine the interestingness of rules. These measures are generally the application of pattern discovery phase dependent.

4.2. Weighted Association Rule Mining (WARM)

Weighted Association Rule Mining navigates each page "1" to "P " is assigned an integer value between. P on the first tour, the next tour page 1 to page last visited this happen and continues until one is assigned. Visit page overweight initial priority signal is acquired. This technique eliminates problems in existing association mining method and also handles both static and dynamic web pages. This requires only one scan of the data set. This in turn eliminates data redundancy, maintains command and navigation of web pages.

Let the Sample log data S have Transactions T={T$_1$,T$_2$,…,T$_n$} with set of pages P={P$_1$,P$_2$,…P$_n$} and a set of positive real number of weights W={W$_1$,W$_2$,…,W$_n$} attached with each visited page P. Let's take a pattern of the form P$_1$P$_3$P$_2$. In the weighted representation this pattern is represented as P$_1$=3, P$_3$=2 and P$_2$=1. First visited page has the highest priority (P$_1$=3), intermediate pages hold highest priority less than one and last visited page contains the one (P$_2$=1). The visitor's web page sequences are retained based on the weights assigned to the visited web page order. The proposed algorithm Weighted Association Rule Mining is scalable and efficient in discovering significant relationships between web pages.

4.3. Sequential Pattern Mining

Association rule intra transaction relationship between sequential patterns represent relationships while transactions. Sequential patterns often resulted in a specific order that item sets a scene, set the same item all items in one transaction is considered worth the time or are within a time interval.

Let *D* be a database [26] of transactions, I= {i$_1$, i$_2$, ik} be a set of k distinct attributes called item sets. A sequence S=<t1, t2, tm> is an ordered list. The length l(s) is the total number of items in the sequence. S ordered list item that consists of a set of sequences S 2, S 1.... SN Is set. Whose support exceeds a predefined minimum support extracting sequential patterns in mining sequential patterns? There are a large number of sequences; users find interesting sequential patterns have different interests and needs. MSP is defined by applications.

4.4. Dynamic programming Technique

Web log data on the scale and nature of the scan are unable to process the data in the data base. Then a new technique to divide several sub- datasets and independent sub- set of log data is essential to resolve. Divide and conquer method is applied in the current association mining techniques. It repeated the same sub division to solve problems and not independent and conquer algorithm, is disabled. These introduced the redundancy, increase processing time and memory space.

To overcome these problems, Dynamic Programming is an algorithm design technique for optimization problems. Dynamic programming approach is used to solve each sub problems only once and the results are stored in a table. And the alternation of the data set required for the scan time is not efficient. This calculation requires only linear time. Dynamic programming techniques to solve complex sub-problems found suitable. There are three basic components of the dynamic programming algorithm.

- The recurrence relation (for defining the value of an optimal solution)
- The tabular computation (for computing the value of an optimal solution)
- The trace back (for delivering an optimal solution).

Let the sequence of k = k1 <K2 <• • <n KN For each key

finding of likelihood of confusion, the solution keys. We expect at least a binary search tree with the cost of exploration (BST) to build. The actual cost is equal to the number of items subject to investigation. The key is satisfied , then the cost = depth (a) 1 , where we reach the target node to find an optimal path costs can trace back through the grid in the BST T. of depth (to) = depth to find a relative path . DP is both time and space complexity storage costs for all nodes and computation cost at each node is constant.

Dynamic programming is generalized to k dimensions, where the time and space complexity is k 0(l) for a hypercube with length l. Dynamic programming technique is used in the proposed methodology.

V. CONCLUSION AND FUTURE WORK

In this paper we have studied the different aspects of web recommendation system and also discussed the different challenges and limitations of existing web recommendation techniques. In this paper we have studied the techniques like Association Rule Mining, Weighted Association Rule Mining, Sequential Pattern Mining and Dynamic Programming techniques. After studying this methods we have identified the one challenging problem of such web recommendation systems is that pages which are recently added or rarely visited by end user. Such pages are not generally included into the set of recommendation. Thus we need to have efficient method which can able to add these pages to the recommendation page set. For the future work we will suggest to present the improved method for association rule mining in order to overcome the above said problem as well as other limitations associated with existing methods.

VI. REFERENCES

[1] D. Dubois, E. Hüllermeier and H. Prade, "A Systematic Approach to the Assessment of Fuzzy Association Rules," *Data Mining and Knowledge Discovery Journal*, Vol. 13, No. 2, 2006, pp. 167-192. doi:10.1007/s10618-005-0032-4

[2] Agrawal R, Imielinski T & Swami A N, (1993) "Mining association rules between sets of items in large databases", *ACM SIGMOD International Conference on Mgt. of Data*, Vol.22, Issue 2, pp.207-216.

[3] Ali Mirza Mahmood & Mrithyumjaya Rao Kuppa,(2012) " A novel pruning approach using expert knowledge for data-specific pruning", *Engineering with Computers* , Vol.28, pp.21–30, 2012.

[4] Anitha A & Krishnan N,(2011) "Dynamic Web Mining Frameworks for E-learning Recommendations using Rough Sets and Association Rule Mining", *International Journal of Computer Applications*,12(11) pp 36-41.

[5] Avrilia Floratou, Sandeep Tata & Jignesh M. Patel,(2011) " Efficient and Accurate Discovery of Patterns in Sequence Data Sets", *IEEE Transactions on Knowledge and Data Engineering,* Vol. 23, No. 8, pp. 1154-1168.

[6] Kerf, Richard, (1985) " Depth-First iterative-deepening : An optimal admissible tree search", *Artificial Intelligence*,Vol. 27, Issue 1, pp.97-109.

[7] Mahbubul Arefin Khan et al., (2012) "Pattern Finder–Efficient Framework for Sequential Pattern Mining", *4th International Conference on Computer Modeling and Simulation*, pp.130-133, vol.22.

[8] Wang W, Yang J & Yu P, (2000) "Efficient mining of weighted association rules", *Proceeding of the sixth ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp.270-274.

[9] Shrivastva Neeraj & Lodhi Singh Swati, (2012) "Approach to Recover CSGM Method with Higher Accuracy and Less Memory Consumption using Web Log Mining", *Journal of Engineering Sciences,* Vol.1,issue 1, pp.83-87.

[10] D. Dubois, E. Hüllermeier and H. Prade, "A Systematic Approach to the Assessment of Fuzzy Association Rules," *Data Mining and Knowledge Discovery Journal*, Vol. 13, No. 2, 2006, pp. 167-192. doi:10.1007/s10618-005-0032-4

[11] P. Resnik and H. Varian, "Recommender System," *Com- munication of the ACM*, Vol. 40, No. 3, 1997.

[12] G. Takacs, I. Pilaszy, B. Nemeth and D. Tikk, "Scalable Collaborative Filtering Approaches for Large Recom- mender Systems," *The Journal of Machine Learning Re- search*, Vol. 10, 2009, pp. 623-656.

[13] S. Orlando, R. Perego and C. Silvestri, "A New Algo- rithm for Gap Constrained Sequence Mining," *Proceed- ings of the ACM Symposium on Applied Computing*, Nicosia, 14-17 March 2004, pp. 540-547,

[14] R. Burke, "Hybrid Recommender System: Survey and Experiments," *User Modeling and User-Adapted Interac- tion*, Vol. 12, No. 4, 2002, pp. 331-370

[15] Baltrunas and X. Amatriain, "Towards Time-Dependant Recommendation Based on Implicit Feedback," *Work- shop on Context-Aware Recommender Systems*, New York, October 2009.

[16] Y. Shavitt, W. Ela and W. Udi, "Building Recommender Syterm Using Peer to Peer Shared Content," *CIKM'*10, Toronto, 25-29 October 2010.

[17] Mr. Sachin S. Patil, "Improved Association Rule Mining Algorithm for Web Recommendation System", International Journal of Engineering & Extended Technologies Research (IJEETR), Vol. 1-2012.