International Journal of Latest Engineering Research and Applications (IJLERA) ISSN: 2455-7137

Volume – 02, Issue – 01, January – 2017, PP – 25-31

Web Recommendation System Based on Approach of Mining Frequent Sequential Patterns

Sampada Khorgade[1], Praful Sambhare[2]

[1](*Department of Computer Science, P. R. Pote College of Engg, Amravati,India)*

[2](*Assist Prof. Department of Computer Science, P. R. Pote College of Engg, Amravati,India)*

Abstract: Today the internet users are increasing day by day so too the websites are also in the increased number. Analyzing and modeling web navigation behavior is helpful in understanding demands of online users. New viewers must follow the sequence path of top viewed websites instead of surfing randomly. Recommendation is a process to analyze users past behavior to recommend new user for purchasing products or viewing certain websites. The main aim of the proposed system is to achieve this feature so this system is designed which recommends the new user to visit the most frequently visited websites. With the help of recommendation process the time of user is saved as the top visited frequently accessed websites are shown in different manner. Also for online purchasing this system provides association between various products. For identifying the patterns the Apriori algorithm is used which shows the frequent items and is designed to operate on databases containing transactions.

Keywords: Apriori algorithm, Association mining, Sequential Pattern, Web Recommendation System

1. INTRODUCTION

In this world of Information Technology, the internet has affected almost every aspect of our world. Since the number of web sites and web pages has increased rapidly, discovering and understanding web users surfing behavior are essential for the development of successful web monitoring and recommendation systems. The substantial increase in the number of websites presents a challenging task for masters to manage the contents of websites [1]. Analyzing and modeling web navigation behavior is supportive in accepting demands of online users. Recommendation is a process to analyze users past behavior to recommend new user for purchasing products or screening certain websites. It is extensively used by commercial websites to recommend some products and services to users.

The basic function of recommender system is to offer recommendation to the users based on items or information related to their importance. Web usage mining is the creation made for user while accessing the internet. Considering an example of shopkeepers who face many challenges where customer's needs are changing all the time. To build up acceptable marketing strategies to assure need of customer in order to carry on and gain a competitive advantage. Recommendation is one of useful thing that people like. If any food service provider uses such system which suggests food based on previous transaction then it will be time reducing activity for customer. Most of the people who bought burger will go for coke and the further food in some chain. That's why it is considerable that achieving this pattern has great blow to better advertisement and better management. Recommendation systems allow buyers to find what they need.

In data mining, association is useful for analyzing and predicting customer behavior. They play an important part in shopping basket data analysis. Association rule mining is primarily focused on finding frequent co-occurring associations among a collection of items. It is sometimes referred to as "Market Basket Analysis", since that was the original application area of association mining. Apriori is an algorithm for frequent item set mining and association rule learning over transactional databases.

2. Background History

2.1 Literature Survey

The GSP algorithm which was proposed by Agrawal and Shrikant [12] that makes multiple passes over the data, in this algorithm if the candidate sequence donot fit in the memory then it generates limited sequence which will be stored in the memory. The data is then scanned to count the support of that candidate sequence. Minimum support of candidates is deleted and the frequent item is stored in the memory.

The SPIRIT Algorithm was designed which uses the regular expressions which was constraint specification tool. On mined patterns it applies the user specified regular expression constraint. This algorithm uses the constraint based approach which is motivated by two main factors: Regular expression (RE) provide

simple and easy syntax for finding the sequential patterns. It also specifies the interesting and non-trivial pattern constraints.[13]

The SPADE Algorithm was proposed by M.J.Jaki [14] which completes the mining process in three steps by scanning the database. The horizontal formatting method i.e GSP that database sequence is transformed into vertical format consisting of id-lists. Additional computation time is required for transforming the layout into vertical format. This format also requires more storage capacity than that of original database.

The next algorithm is the SPAM algorithm which traverses the sequence tree in depth first search manner and checks the support of each sequence with min support count for each extended items. SPAM uses the vertical bitmap data structure and uses the depth first fashion strategy for mining sequential patterns. SPAM algorithm is similar to that of SPADE algorithm, but it uses bitwise operators rather than regular expression. In experimental results found that SPAM algorithm outperforms the SPADE by the factor of 2.5, while SPADE algorithm is 5 to 20 times more space efficient than SPAM.[15]

The scalable and efficient sequential pattern mining method being developed called Free Span which integrates the mining of frequent sequence with that of frequent patterns and uses projected database to confine the search and the growth of subsequence fragments. The Free Span was developed to substantially reduce the expensive candidate generation and uses frequent items to recursively pass the sequence database into projected databases while growing subsequence fragments in each projected database. Each projection partitions the database and makes smaller and more manageable units. The amount of sequence duplication may occur as the same sequence could appear in more than one projected database. However, the size of each projected database usually decreases rapidly with recursion. [16]

Wap mine is a pattern growth and tree structure-mining technique with tree structure. Here the sequence database is scanned only twice to build the WAP-tree from frequent sequences along with their support. The first happening for each item in a frequent item set is checked and then it tracked to mine the tree for frequent sequences. The WAP mine algorithm is reported to have better scalability than GSP. Although it scans the database only twice and can avoid the problem of generating candidates as in Apriori-based methods, as it recursively reconstructs numerous intermediate WAP-trees during mining so the problem of memory consumption occurs, as the number of mined frequent patterns increases. [17]

The prefix span (prefix-projected sequential pattern mining) algorithm which is both scalable, efficient and presented by Jian Pei representing the pattern-growth methodology, which finds the frequent items after scanning the sequence database once. The database is then projected according to the frequent items, into several smaller databases. Finally, the complete set of sequential patterns is found by recursively growing subsequence fragments in each projected database. Although the Prefixspan algorithm successfully discovered patterns employing the divide-and-conquer strategy, the cost of memory space might be high due to the creation and processing of huge number of projected sub-databases. [18]

2. Related Work

The sequential pattern mining problem was first addressed by Agrawal and Srikant [1] and was defined as follows: "Given a database of sequences, where each sequence consists of a list of transactions ordered by transaction time and each transaction is a set of items, sequential pattern mining is to discover all sequential patterns with a user-specified minimum support, where the support of a pattern is the number of data-sequences that contain the pattern". This form of discovery requires a different type of algorithm[6] and will be described separately from those algorithms that are based on the more general transaction oriented datasets.

The SPADE (Sequential Pattern Discovery using Equivalence classes) algorithm by Zaki[7] there are generally four types of recommendation systems[2]: content-based filtering[3], collaborative filtering, rule based filtering and hybrid approaches, content based filtering system performs recommendation in steps. First, an item profile consisting of a set of features is extracted for each item. Then based on features of items that are purchased by each user, user profiles are generated. Then similarity scores between user profiles and item profiles are calculated to finally recommend items with top similarity scores. The limitations of these systems are that recommendations are always similar to items that a user has already purchased, and it is difficult to recommend items for new users.[3]

Collaborative filtering systems are built on the assumption that a good way to find interesting content is to find other people who have similar interests and then recommend items rated highly by those similar users. There are two general classes of collaborative filtering algorithms. Memory based algorithms operate over the whole user database to make recommendations [5]. Model based collaborative filtering, in contrast, learns a model from the user database, which is then used for recommendations. Collaborative filtering recommendation systems, however, also have limitations that it is difficult to recommend items for new users, to recommend items which have not been rated before, and to recommend when rating information is insufficient.[4]

Rule based approach is a simple but popular way of recommendations. Rules are usually derived from database of previous transactions. Aggarwal[6] discovered localized association rules among items that are purchased together, which are helpful for target marketing. Cho[7] used data mining techniques such as web usage mining, decision tree induction, association rule mining to build a rule based recommendation system.

A personalized product recommendation is an enabling mechanism to overcome information overload occurred when shopping in an Internet marketplace. Collaborative filtering has been known to be one of the most successful recommendation methods, but its application to e-commerce has exposed well-known limitations such as sparsity and scalability, which would lead to poor recommendations. In this paper[8] Jae Kyeong Kim suggests a personalized recommendation methodology by which we are able to get further effectiveness and quality of recommendations when applied to an Internet shopping mall. The suggested methodology is based on a variety of data mining techniques such as web usage mining, decision tree induction, association rule mining and the product taxonomy.

C. L. Huang and W. L. Huang [9] proposed a sequential pattern based system that predicts customer's time invariant purchase behavior for food items in a super market. Hybrid approach combines multiple techniques to overcome the limitations of individual systems. For example, a linear combination of rule based and collaborative filtering recommendations considers customers purchase sequences over time as well as their purchase data for latest period. [10]

In Web-based services of dynamic content systems face the difficulty of timely identifying new items of high-quality and providing recommendations for new users. In[11] Wei Chu, Seung-Taek Park propose a feature-based machine learning approach to personalized recommendation that is capable of handling the cold-start issue effectively. We maintain profiles of content of interest in which temporal characteristics of the content, e.g. popularity and freshness, are updated in real-time manner. We also maintain profiles of users including demographic information and a summary of user activities. Based on all features in user and content profiles, we develop predictive bilinear regression models to provide accurate personalized recommendations of new items for both existing and new users.

3. System Architecture

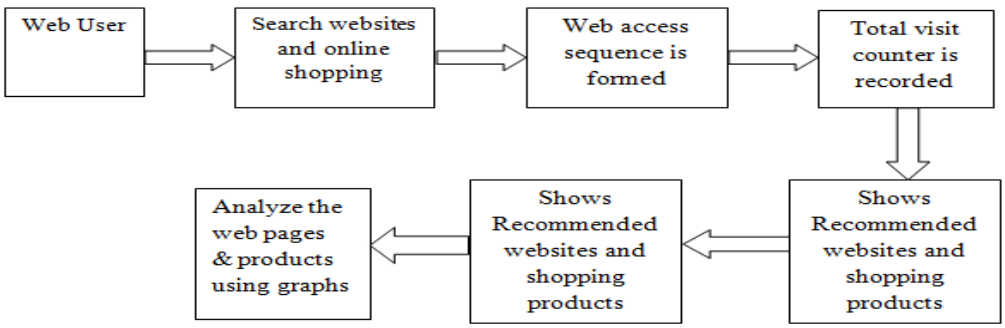

Fig.3.1 System Architecture

First the web user searches over the web pages in a transaction and on moving toward the next page the system recommends the user to visit next frequently accessed webpage after the current page is finished. This system is going to provide the total visit counts of the past visited user who have visited frequently. During visiting the web pages the web access sequence is generated through which the top viewed page count is incremented. The purpose of mining association methods is to find out which web pages are usually visited together. List of Engineering College websites is provided with their total visit count and if same visit is seen then user can see the placement and fees structure of respective colleges before visiting to that link and on the basis of that parameter user can then visit over there.

During online shopping the system provides distinct ranked products which mean system gives detailed count of all purchased products. Using these criteria the new web users can easily match up their needs, they can easily able to decide which products are frequently purchased by past users. Association between every products are listed which becomes an easy way of identifying which product to purchase. Using Apriori algorithm the result is shown in form of the graphs which is an easy way of identifying the products. Apriori can be used to find frequent patterns & determine association rules which highlight general trends in the database.

FLOWCHART

User needs to enter the user name and email id and starts the new transaction, after that user can either perform the online shopping transaction or can access the websites. If user selects to visit the websites then system recommends the new user to visit the most frequently accessed website by showing the total visit count. The web access sequence is formed while accessing the web pages. User can visit towards the websites which also shows total visit counter through which user can finds the mostly accessed website.

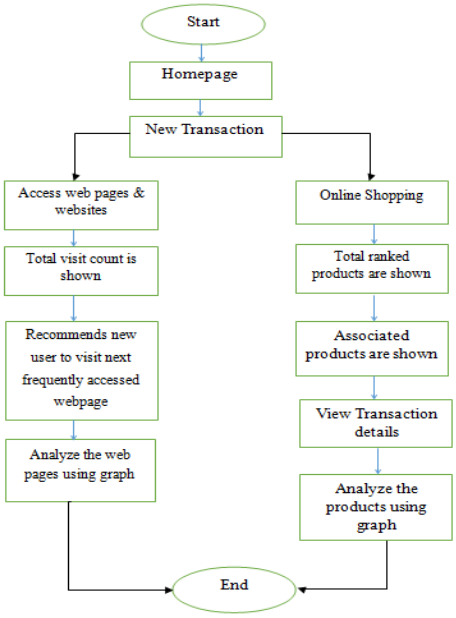

Fig 3.2: Flowchart of Proposed design

4. Result Analysis

The proposed work builds in dot net framework uses sql server r2 2008 for backend process. This work is based on providing frequently used web pages and the websites to the new web users. The web access sequence is generated as user visits the web pages. Also on visiting the websites the total visitors count is displayed and if the count is same then new user will decide on the basis of fees and placement criteria on which website to visit. The proposed work provides the recommendation list to visit the next frequently accessed webpage.

Also when user will purchase products online then total rank of every product is listed so that new user can easily track over there. Using Apriori algorithm the frequent items sequence can find out. Also the association between every product is shown through which user can easily predict which product to purchase. The graph of most frequently visited web pages and the product which purchased is shown for analyzing it. The experimental result of proposed work is shown below in the following graphs.

Products	**Sale rate**		
Mobile-Power bank	6	7	7
Mobile-Headphones	8	8	7
Milk-Cornflakes	8	9	9
Milk-Bread	5	5	4
Oil-Flour	3	4	3
Dal-Oil	5	6	5

Table 4.1: Associated products with their distinct item sale

From the above result we can see that first item is the ‘mobile-power bank’ whose highest sale rate is count-7, similarly the next product is the mobile-headphones whose highest count among three transactions is 8. So from these purchased items we can conclude that past users have purchased the mobile-headphones in a high quantity. Similarly we can find with the other products so the association between various products is beneficial.

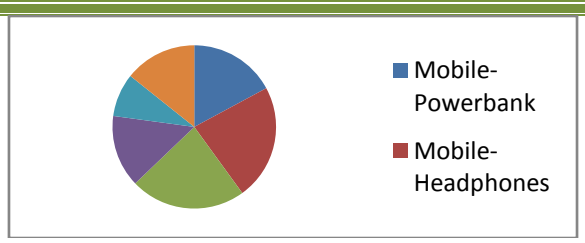

Fig 4.1: Graph of associated product

The above Pie chart indicates the various purchased products with its associated total items. Total associated products is shown as per user has purchased it, also it indicates the total items with its highest sale rate.

Algorithm	**Frequent item set**		
Sequential Pattern based approach	**7**	**5**	**5**
Apriori approach	**4**	**3**	**4**

Table 4.2: Comparison factor of purchased frequent product

Suppose the given a set of events E = {a, b, c, d, e, f} which may represent product accessed by users, a web access sequence database four users may have four records: [T1,<abdac>],[T2,<eaebcac>],[T3,<babfaec>],[T4,<abfac>]. From this the records the frequent items identified is ab whose count is 2. The above table also shows the frequent items whose count and the respective algorithm is shown.

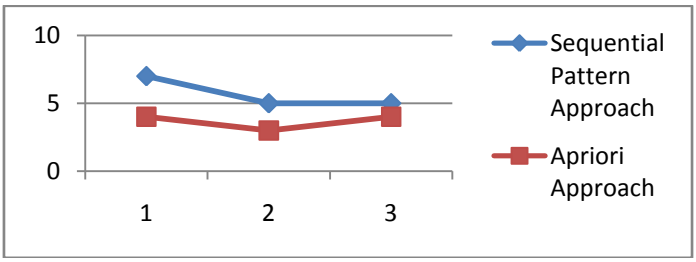

Fig 4.2: Comparison graph

From the above graph the comparison factor between sequential pattern based and Apriori based is shown. On X-axis total frequent item set from purchased product is taken and on Y-axis incremental count has been taken.

College Name	Fees Structure
Bharati Vidyapeeth, Pune	95,000
YCCE, Nagpur	74,646

Table 4.3: College name with its fees structure

The above table shows the name of colleges with its fees structure it is provided if the visit count of past users is same for some colleges then the fees parameter will provides helpful for moving towards respective colleges depending upon their needs.

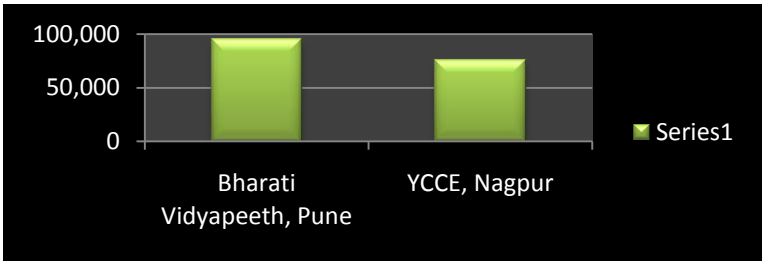

Fig 4.3: Graph generated based on fees parameter

From the above graph the comparison factor between sequential pattern based and Apriori based is shown. On X-axis total frequent item set from purchased product is taken and on Y-axis incremental count has been taken.

Name of College	Placement Criteria		
	2014	2015	2016
Ramdeo baba College, Nagpur	610	1003	1316
MIT, Pune	310	321	340

Table 4.4: College with Placement Criteria

The above graph is generated of placement criteria which show its result and on viewing this user can easily able to identify it. On X-axis year count is shown and on Y-axis total count of students is shown.

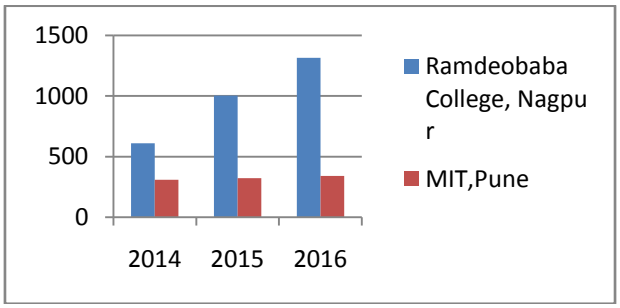

Fig 4.4: Graph of Placement Criteria

5. Conclusion

The proposed system provides the recommendation of websites and the shopping products to the user. Through which user can easily navigate over the suggested website. Common navigation behaviors of the users can be used to improve the actual design of web pages and for making other modifications to a web site. The contribution of this proposed work is to provide or recommend the websites which has visited frequently and also some parameters are provided if visited count of websites is same. Using Apriori algorithm the frequent items of products is provided, with their associated products which has purchased frequently. The proposed system uses this algorithm for finding the frequent set of items. Using those items the associated products can be recommended easily.

6. References

[1] Manisha Valera GTU, Uttamchauhan "An Efficient Web Recommender System based on approach of mining Frequent Sequential Pattern from Customized Web Log Preprocessing", *ICCCNT 2013 IEEE*

[2] Haiyun Lu, "Recommendations Based on Purchase Patterns" *International Journal of Machine Learning and Computing, Vol.4, No. 6, December 2014*

[3] C. L. Huang and W. L. Huang, "Handling sequential pattern decay: developing a two-stage collaborative recommendation system," *Electronic Commerce Research and Applications, vol. 8, no. 3, pp. 117-129, 2009.*

[4] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical Analysis of Predictive Algorithm for Collaborative Filtering," *in Proc. the 14thConference on Uncertainty in Artificial Intelligence, 1998.*

[5] J. Delgado and N. Ishii, "Memory based weighted-majority prediction for recommender systems," presented at *ACM SIGIR „99 Workshop on Recommender Systems: Algorithms and Evaluation, 1999.*

[6] Dr. Sunita Mahajan , Prajakta Pawar and Alpa Reshamwala ,"Performance Analysis of Sequential Pattern Mining Algorithms on Large Dense Datasets" *International Journal of Application or Innovation in Engineering & Management (IJAIEM).*

[7] C. C. Aggarwal, C. Procopiuc, and P. S. Yu, "Finding localized associations in market basket data," *IEEE Transactions on Knowledge and Data Engineering, vol. 14, no. 1, pp. 51-62, 2002.*

[8] Y. H. Cho, J. K. Kim, and S. H. Kim, "A personalized recommender system based on web usage mining and decision tree induction," *Expert systems with Applications, vol. 23, no. 3, pp.329-342, 2002.*

[9] C. L. Huang and W. L. Huang, "Handling sequential pattern decay: developing a two-stage collaborative recommendation system," *Electronic Commerce Research and Applications, vol. 8, no. 3, pp. 117 129, 2009.*

[10] D. R. Liu, C. H. Lai, and W. J. Lee, "An hybrid of sequential rules and collaborative filtering for product recommendation," *Information Sciences, vol. 179, no. 20, pp. 3505 3519, 2009.*

[11] Manan Parikh, Bharat Chaudhari, Chetna Chand," Comparative Study of Sequential Patterns mining algorithm", *International Journal of Application or innovation in engineering lap Yip,and Management(IJAIEM).*

[12] Minghau Zhang, Ben Kau, Chi, Da'vid Cheung, "GSP based efficient algorithm for mining frequent sequences", *HKU CS IS Tech Report TR 2002-05.*

[13] Minos N.Garofalakis, Rajeev Rastogi, Kyuseok Shim "SPIRIT: Sequential Pattern Mining with Regular Expression Constraints", *Proceedings of the 25thVLDBConference, Edinburgh, Scotland, 1999.*

[14] Zaki, M.J.,"SPADE: An efficient algorithm for mining frequent sequences", *Machine Learning 42(1/2), 31-60 (2001)'*

[15] Jay Ayres, Johannes Gehrke, Tomi Yiu, and Jason Flannick, "Sequential PAttern Mining using A Bitmap Representation" *SIGKDD '02 Edmonton, Alberta, Canada.*

[16] Prof. Kirti S. Patil, "Overview of Sequential Pattern Mining Algorithms" *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*

[17] Jiawei Han,Jian Pei, "FreeSpan:Frequent Pattern-Projected Sequential Pattern Minig" *supported in part by Natural Science and Engineering Research Council Of Canada.*

[18] Resul Das, Ibrahim TURKOGLU, "Extraction of Interesting Patterns Through Association RuleMining for Improvement of Website Usability," *Journal of Electrical & Electronics Engineering, Vol.9, No. 2, 2009.*

[19] Ezeife and Y. Lu. "Mining web log sequential patterns with position coded pre-order linked wap-tree." *International Journal of Data Mining and Knowledge Discovery (DMKD) Kluwer Publishers, 10(1):5–38, 2005.*

[20] Jian Pei,Jiawei Han, "Mining Access Patterns Efficiently from Web Logs" *supported in part by Natural Science and Engineering Research Council Of Canada (grant NSERCA3723).*

[21] Dheeraj Kumar Singh, Varsha Sharma, Sanjeev Sharma "Graph based Approach for Mining Frequent Sequential Access Patterns of Web pages", *International Journal of Computer Applications (0975 – 8887) Volume 40– ssssNo.10, February 2012.*

