ITEM SET MINING AND USER INTEREST MODEL FOR EFFECTIVE SHOPPING SYSTEM

P. KAVITHA¹, S. ANITHA², BISMI ILLYAS³, P. SIVABALAN⁴

¹,²Assistant Professor, New Prince Shri Bhavani College of Engineering and Technology, Gowrivakkam, Chennai, Tamilnadu, India.
³,⁴PG Student, Department of Computer Applications, New Prince Shri Bhavani College of Engineering and Technology, Gowrivakkam, Chennai, Tamilnadu, India.

¹kavitha@newprinceshribhavani.com, ²anitha@newprinceshribhavani.com
³bismiillyas1995@gmail.com, ⁴aaranirojanadaganram@gmail.com

Abstract: In the world of digital markets, online shopping is the new area of interest to common people for a variety of experience and the availability of selection from thousands of things in a simple mouse click. Online shopping helps customers to multiple accesses to similar products and specification without any constraints. The challenge in this is the selection of products as per one’s requirements from the various options available in the internet. The main purpose of this paper is to identify one’s aspects and analyze the tweets and thereby provide a good shopping experience in a most economic way. A Twitter like application is designed where user’s likes in this page and likes in the Purchase website monitored parallel. Purchase portal consists of two options – General purchase and Profile based purchase. In Profile based purchase, Items are displayed based on the Users interests that were monitored from the user’s online interest.

Keywords: Item set mining, Tweet stream clustering, Collaborative filtering

1. INTRODUCTION

With the development of technologies and digitalization, the whole world is a turned into a supermarket under a single roof. One of the main challenges in online shopping is the selection of the particular item from various available sites. Normally, a single site cannot satisfy user’s choice as multiple sites provides a range of offers on similar items with a lot of price variation. For example, to purchase a pair of shoes, a user to browse various E–commerce sites to compare products and user reviews of the products. However, a user faces numerous difficulties while purchasing as multiple choices are available when selecting services. Various service information, such as price and user rating, can affect user’s decision and choice of purchase. Due
to the explosive growth of Internet, users are flooded with a large amount of information. In this situation in order to enhance user experience and promote sales, recommendation system has become an indispensable part of online shopping sites.

Recommendation systems have become very common in recent years, and are utilized in various e-commerce sites for shopping. In the last decade scope of e-commerce has enlarged exponentially with millions of users and items because of the improvement due to digitalization. Due to the explosive growth of Internet, users are flooded with a large amount of information. In this situation in order to enhance user experience and promote sales, recommendation system has become an indispensable part of online shopping sites.

Frequent Itemset Mining is the most popular techniques to extract knowledge from data. However, these mining methods are not easy to use when they are applied to Big Data. With the latest improvements in this field of parallel programming provide many tools to solve these issues. In this paper, we are presenting a detailed survey of Hadoop, which helps in storing data and parallel processing. Here we have explored various Frequent Itemset Mining techniques on parallel and distributed environment with the aim to present a comparison of different frequent itemset mining techniques and help to develop efficient and convenient itemset mining techniques.

2. RELATED WORKS

The Advanced features of this project is to identify the customer’s interest with high end collaborative filtering thereby providing the high quality shopping experience with maximum monetary savings to the customer. The work [1][5], has focussed on Mining top-k frequent patterns in the presence of the memory constraint. In this the main focus is to retrieve top-k (closed) item sets in the presence of the memory constraint and specify the available upper memory size that can be utilized by mining frequent itemsets. To comply with the upper bound of the memory consumption, two efficient algorithms, called MTK and MTK Close, are devised for mining frequent item sets and closed item sets, respectively, without specifying the subtle minimum support. Instead, users only need to give a more human understandable parameter, namely the desired number of frequent (closed) itemsets k.[4]

Another approach is UP-Growth: An Efficient Algorithm for High Utility Itemset Mining. Mining high utility itemsets from a transactional database refers to the discovery of
itemsets with high utility like profits and they incur the problem of producing a large number of
candidate itemsets for high utility itemsets[2][8]. This large number of candidate itemsets
degraded the mining performance in terms of execution time and space requirement. To
overcome these issue an efficient algorithm, namely UP-Growth (Utility Pattern Growth), for
mining high utility itemsets with a set of techniques has been introduced.[10]In most of the cases
candidate set generation is costly, especially when there exist prolific patterns or long
patterns[11][9][8]. In order to make the data mining more reliable , a novel frequent pattern tree
(FP-tree) structure for storing compressed, crucial information about frequent patterns, and
develop an efficient FP-tree based mining method has developed[12].

Other approach include mining Frequent Patterns without Candidate Generation. A new
approach for efficiently mining high utility itemsets and removes unpromising items early from
the database, thus causing better upper bounds of utility values and reduce more itemsets. It can
reduce the size of data to be processed and the scan time. Junqiang Liu, et al.,[6],proposed Direct
Discovery of High Utility Itemsets without Candidate Generation. A high utility itemset growth
approach that works in a single phase without generating candidates. Our basic approach is to
enumerate itemsets by prefix extensions, to prune search space by utility upper bounding, and to
maintain original utility information in the mining process by a novel data structure[3][7]. Data
structure enables us to compute a tight bound for powerful pruning and to directly identify high
utility itemsets in an efficient and scalable manner. In other existing work the Isolated items
discarding strategy for discovering high utility itemsets is mentioned. Isolated Items Discarding
Strategy (IIDS), which can be applied to any existing level-wise utility mining method to reduce
candidates and to improve performance[15].In the existing system, data in their basic form,
while being informative, can also be overwhelming. It is a huge task to go through all these
millions of data which contains a lot of unwanted and redundant data with the increase in use of
internet throughout the world. The user’s information also increased substantially.[15][7][13]
The main challenge is to make out useful information from these raw data thereby developing a
user friendly software. In existing work, the some of the detriments are (i) Does not analyze
maximum purchased products (ii) Less accuracy due to large number of data.
3. ITEM SET MINING APPROACH

We are analyzing the overall transaction of all the users as per the available data and we are extracting maximum profit yielding purchase of frequency item set is analyzed. This process will exhibit the maximum profit based analysis. With help of collaborative filtering similarity of like between two users were analyzed and items based on similar tastes will be recommended to the other user.

Twitter like application is designed where users’ likes in this page and likes in the purchase website are monitored parallel. Purchase portal will have two options like general purchase and profile based purchase for obtaining the desirous result. In profile based purchase, items are displayed based on the user’s Interest as per the likes of the user. Related items and items which are purchased more often are also displayed to the user based on the user interest. Based on the taste of the user and reviews available the best product will be recommended to the user for a good shopping experience as per the budget of the user. The main advantages of these systems are as follows (i) Ease of analyzing the maximum purchased products with the help of available data. (ii) Can collect user likes while they give their registration (iii) More effective in pricing and selection of the products.

5. SYSTEM ARCHITECTURE

While collecting the user’s like we can analyse the data based on his/her interest and fetch the information to monitor the likes of user. These information will classified as personnel choice and the users choice related items can be displayed without time delay for a customized shopping. Figure 1 shows the architecture diagram of User Interest Shopping System.

Fig.1. Architecture diagram
6. SYSTEM IMPLEMENTATION

Introducing the modules used in the implementation are:

User Registration
In client side user can enter all details. Then user can login using particular username and password. All the inserted also updated items are added into the product list. Then select user wanted items then add all items into cart products with count of the each item. A warning message will display in dialogue box when the customer type the quantity above the constraint value mentioned in the database. All selected items are displayed in the cart product list and purchase the required items.

Twitter like Application:
User can register in twitter application and go for login by giving valid user name and password. If the user name and password is valid the user can login into home page. Once we login in home page the display of several products is to be done. Based on user interest he go for likes to the products. So this likes is going to monitor by server and stored in data base. These information giving input to hadoop server.

Purchase Portal
Consumer buying behaviour is the sum total of a consumer's attitudes, preferences, intentions and decisions regarding the consumer's behaviour in the marketplace when purchasing a product or service. The study of consumer behaviour draws upon social science disciplines of sociology, and economics. At this stage, the consumer will make a purchasing decision. The ultimate decision may be based on factors such as price or availability. For example, our consumer has decided to purchase a particular model of car because its price was the best she could negotiate and the car was available immediately.

Server
Server module will monitor the entire User’s information in their database and verify them if required. Also the Server will store the entire User's information in their database. Also the Server has to establish the connection to communicate with the Users. The Server will update the each User’s activities in its database. The Server will authenticate each user before they access the Application. So that the Server will prevent the Unauthorized User from accessing the Application.
Product Ranking

The product ranking is based on the feedback value we rate the promising items. Then find out the promising items. Candidate item sets can be generated efficiently with only two scans of database.

Mining high utility item sets from database refers to the discovery of item sets with high utility like profit. So the users can the feedback base product to purchase. This will be useful for the new user to by the product.

7. RESULTS AND DISCUSSIONS

Snapshot of the item set mining

Fig 2: Home Page

The home page for item set mining that is available to the user is shown in fig 2. User link is monitored by analyzing the likes and interest based on the purchase portal. Fig. 3 shows the items added to the cart. Fig. 4 shows the list of items purchased.
8. CONCLUSION

In this paper, we had analyzed the various aspects of online shopping as per the user data and the usage of Big Data Analysis of items Set Mining in this field. With the use of user’s requirement data, we can provide the most suitable products to the user in minimum time and cost. The user’s choices were given top priority and the most relevant items can be displayed for user satisfaction and a superior shopping experience. The user’s data also be maintained in a confidential manner to maintain the privacy of the user and safeguard the data for future use.

REFERENCES
