INTERNATIONAL JOURNAL FOR TRENDS IN ENGINEERING & TECHNOLOGY
VOLUME 4 ISSUE 1 – APRIL 2015 - ISSN: 2349 - 9303

An Improvised Fuzzy Preference Tree Of CRS For E-Services Using Incremental Association Rule Mining

Roshni.K[1]
[1]Arunai Engineering College, CSE,
kroshni30@gmail.com

Dhanalakshmi.S[2]
[2]Arunai Engineering College, CSE
dhanalakshmi1984@gmail.com

Abstract—Web mining is the amalgamation of information accumulated by traditional data mining methodologies and techniques with information collected over the World Wide Web. A Recommendation system is a profound application that comforts the user in a decision-making process, where they lack of personal experience to choose an item from the confound set of alternative products or services. The key challenge in the development of recommender system is to overcome the problems like single level recommendation and static recommendation, which are exists in the real world e-services. The goal is to achieve and enhance predicting algorithm to discover the frequent items, which are feasible to be purchasable. At this point, we examine the prior buying patterns of the customers and use the knowledge thus procured, to achieve an item set, which co-ordinates with the purchasing mentality of a particular set of customers. Potential recommendation is concerned as a link structure among the items within E-commerce website, which supports the new customers to find related products in a hurry. In Existing system, a fuzzy set consists of user preference and item features alone, so the recommendations to the customers are irrelevant and anonymous. In this paper, we suggest a recommendation technique, which practices the wild spreading and data sharing competency of a huge customer linkage and also this method follows a fuzzy tree- structured model, in which fuzzy set techniques are utilized to express user preferences and purchased items are in a clustered form to develop a user convenient recommendations. Here, an incremental association rule mining is employed to find interesting relation between variables in a large database.

Index Terms— Recommender System; Fuzzy Set; Fuzzy Tree-Structured User Preference; Incremental Association Rule Mining.

1 INTRODUCTION

WITH the recent explosive growth of the amount of content on the Internet, the users are facing difficulties to determine and utilize information and also content providers to sort out and catalogue documents. Data mining or Knowledge Discovery is the process of analyzing data from different perspectives and summarizing it into useful data. This data can be applied to increase revenue, cut costs or both. Online libraries, other large manuscript repositories and search engines are springing up so quickly; it is hard and costlier to sort out every document physically. In order to deal with this issue, we looking forward a robotic method of working with web documents. So that they can be easily browsed, structured and cataloged with negligible human intervention.

Developing improved methods of accomplishing machine learning techniques on this huge amount of non-tabular, semi-structured web data are thus extremely suitable. The goal of clustering is to split the given data set into groups called clusters, such that items in the same cluster are similar to each other and dissimilar items are in other clumps. In the classification we attempt to predefine the category by assigning a data item based on a model that is produced from pre-classified training data.

In actual operation, the bundling and organizing data are fall in the field of knowledge discovery in databases or data mining. Exercise Data mining techniques to a content of web page is termed as web content mining, which is a new sub-sphere of web mining, somewhat framed upon the sanctioned area of data rescue. The vector - space model is employed to present text and web document information for clustering and categorization. Each term of this model becomes a characteristic dimension. The values designated to each dimension is representing number of times a similar term is noticeable on it or it may be the weight of other frequency information , such as the number of times that term is appears on other documents. This model allows the use of traditional machine learning methodologies that understand with numerical feature vectors in a Euclidean feature space. Simply, it removes content such as order and document the terms appear.

Graphs are essential and efficient mathematical constructs for representing relationships and morphological data. Many problems are using graphs to screen out the compression, traffic flow analysis, resource allotment, etc. In extension to problems, a few algorithms can made to process graphs, the same would be wishing for many applications. The inquiry has been done in the space of graph similarity in order to carry out the additional information is accepted by graphical representations. For dealing with the graphs, mathematical framework is introduced. Few application domains like face and fingerprint recognition as well as exception detection in communication networks, similar to a graph techniques have been used.

This composition is developed with popular sub-categories of Data Mining: - "Market Basket Analysis (MBA)", which is a modeling technique providing vision into the customer purchasing patterns. A market basket is collected on the item-sets which buy in a single trip to the shop. MBA basically finds the relationship between the items purchased in this basket. The marketing tool is employed to figure out the frequent item sets in a large number of transactions. So it is called as “Frequent Item-set Mining”.

Clustering is the action of grouping same elements together with a strong criteria and threshold values. The items which are grouped into a cluster are very closely related to each other and those in different clusters don’t exhibit a close relationship. So it can be described as Intra cluster dissimilarity, it should be as low as possible and Inter cluster similarity should also be very low. Incremental association rule mining is the action of seeking the relationship between the items with criteria of customer purchasing patterns. Depending on the transactions performed by the customer the frequent items and relationship between those items were found out.

2 RELATED WORK

In general, there is a encompassing literature on measuring the similarity between recommendation systems. This section reviews several related work in order to explore the strengths and limitations of previous methods, and to spot the difficulties in E-services.

Greg Linden, Brent Smith, Jeremy York [1] to generate a list of recommended items, a recommendation algorithm is used. This algorithm finds the set of customers who purchased and the items rated are overlie the other customers purchased and rated items. But the recommendation is not scalable over very large customer set and product catalogs.

GediminasAdomavicius, Alexander Tuzhilin [2] Advanced recommendation algorithm is used to advance the user behavior representing methods and the sequence of an items is to be recommended. Capabilities of recommendation systems were improved, but not accurate recommendations due to marginal number of ratings.

SilvanaAciar, Debbie Zhang, Simeon Simoff, and John Debenham [3] to tackle the difficulty of using a customer view regarding the products, articulated online in free-form text, an informed recommender is created to produce product recommendations. The Textual information is used to make recommendations. Recommendations are produced based on review comments.

Zan Huang, Hsinchun Chen [4] Collaborative filtering recommendation algorithm is compared with E-commerce data sets. A Meta level guideline is developed to recommend a suitable recommendation algorithm to demonstrate certain characteristics in the given applications. Six types of representative CF algorithms and different E-commerce was evaluated to assess the algorithm's effectiveness with additional data.

A. C. M. Fong, Baoyao Zhou, S. C. Hui, *Se*, Guan Y. Hong [5] Web recommender systems has become trendy for several consumers who not only do procure online, but also finds related information on products and services prior to the commit to purchase. User behavior knowledge base was constructed, which uses the fuzzy logic. It represents the real-life sequential concepts and significant resources for cyclic pattern-based web access activities.

Alexandros Nanopoulos [6] Item recommendation in collaborative tagging system problems is considered, so three-mode tensor model was proposed to capture the three-way correlations between users, tags and items. To improve the quality of recommendations, multi way analysis was used to expose latent correlations.

Hyea Kyeong Kim, Young U. Ryu, Yoonho Cho, and Jae Kyeong Kim [7] nowadays the size of the consumer networks becomes extremely large, critical scalability problems occurs in the traditional global processing method of CF, but it may not be helpful in real-time environment. In preference-based customer network, customer recommendation system was used in local processing method to form a content recommendation of social network.

Qi Liu, Enhong Chen, HuiXiong, Chris H. Q. Ding, Jian Chen [8] this paper presents an iExpand, which develop the consumer latent interests in developing an item-oriented model-based collaborative framework. IExpand model was used to capture every user’s interests and then personalized ranking strategy was developed to predict a user’s likely interest development.

3 PROPOSED ALGORITHM

Here, we have to develop a fuzzy model for classifying the customers in order to provide dynamic recommendations accordingly. Items and customers have been clustered to derive customer based recommendations. Incremental association rule mining is used to detect frequent items which are purchased by the customers. This section will discuss issues relating to recommendation systems and various other implementation issues.

A. USER INTERFACE

In the industrial design, field of human–machine interaction plays an important role. It is the space where interaction between humans and machines occurs. The main goal of interaction will be effective operation at the user interface. The user can manipulate a system with the help of inputs and should perform either login or register operation for to proceed further stages.

B. CLUSTERING TRANSACTION HISTORY

To produce a clustered set of transactions, transaction history database has been used. The cluster transactions are derived from finding the frequent items which are purchased by customers. This transaction contains the previous transactions made by customers. This transaction consists of two phases; they are allocation phase and refinement phase.

Allocation Phase: In the allocation phase, each transaction ’T’ is read in sequence. The transaction ’T’ can be assigned to an existing cluster or a new cluster will be created to accommodate ’T’ for minimizing the total cost of clustering. For each transaction, the initially allocated cluster identifier is written back to the database. The decision of whether to include the transaction in one of the existing clusters or to create a new one is made by calculating the cost of clustering. The cost consists of intra-cluster dissimilarity and inter-cluster similarity explained in system design.

Refinement Phase: In the refinement phase, the small large ratio (SL ratio) of all the transactions is calculated as follows.

SLR= (no. of small items) / (no. of large items).

The SL ratio of each transaction, thus calculated is then compared with the SLR threshold. If the SLR of the transaction exceeds the threshold, then the transactions are moved to the excess pool. The above process is explained in detail at the system design.

C. INCREMENTAL ASSOCIATION RULE MINING

The transaction history database will contain the preceding transactions made by the customers. The particulars consist of customer identification, the set of objects bought along with the transaction identification. This phase has two sub phases they are, original database history and updating frequent & promising frequent item sets.

Original Database History: A dynamic database may permit to insert new transactions. This may, not only invalidate existing association rules, but also activates new association rules. Maintaining association rules for a dynamic database is an important issue. Thus, a new algorithm is to deal with such updating situation is proposed.

Updating frequent and promising frequent item sets: When a new transactions are added to an original database, an old frequent k-item could become an infrequent k-item and an old promising frequent k-item could become a frequent k-item. This introduces all existing association rules would become Weaker. To deal with this problem, all k-items must be updated when new transactions are added to an original database.

D. RECOMMENDATIONS

In this section, all the recommendations are separated out, according to the types of Customers whom their Purchase items, a preferable range of Products and wish items are alike to the particular customer type Set. Then, recommended items are supplied to the customer from the matching type set.

4 SYSTEM DESIGN

Figure 1 demonstrates the framework of our proposed approach. The system architecture reveals the step-by-step process of generating a dynamic recommendation.

A. User Interface

The first step in the proposed approach is to authenticate the customer which aims to classify the customer. Allowing the users to manipulate a system. The user will perform either login or register operation. During the user manipulations, the users are classified by user preferences. The customer details are stored into the database. After these operations get over he will go to the next stage. The transaction database will commence its action in this stage.

B. Combined Mining

In this section, it consists of three parts. They are combined rule cluster, combined rule pair and combined association rule mining. To generate a dynamic recommendations for the customers, then follow the below steps

Combined Rule Cluster: In this part, transactions database is taken as input. The aim of the combined rule cluster is to produce the cluster set of transactions. Each transaction t is read in sequence. To minimize the total cost of cluster or new cluster, each transaction 'T' can be assigned to an existing cluster or a new cluster will be created to accommodate 'T'. For each transaction, cluster identifier is initially allocated and written back to the database. A new cluster will be create depends on either calculating the cost of clustering or transaction in one of the existing clusters will decide. The cost consists of intra-cluster dissimilarity and inter-cluster similarity which are calculated as follows.

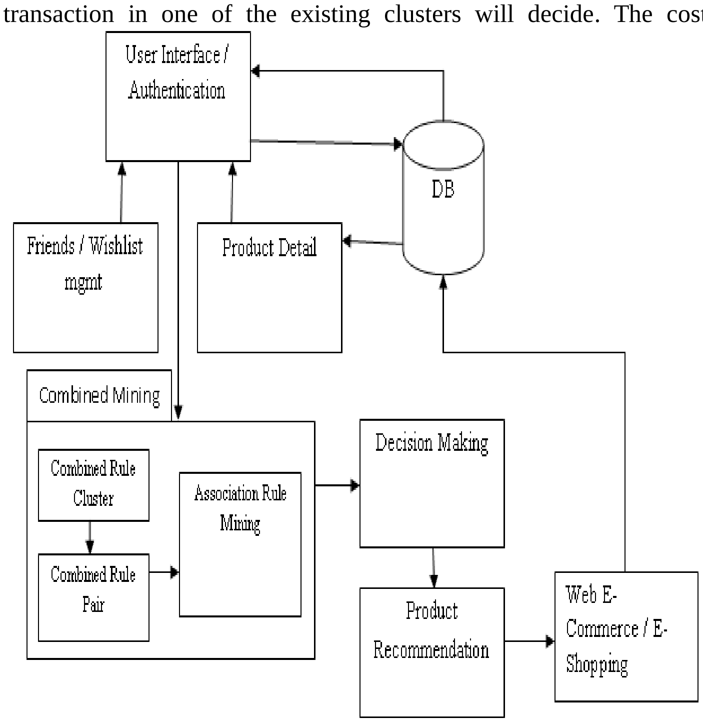

Fig. 1. System Architecture

Intra-cluster dissimilarity: Intra-cluster dissimilarity tells us how different the transactions are within a cluster.

$$Intra\ (U)= |Ukj=1\ sm\ (Cj, E)|$$

Where
Intra (U) – Intra cluster dissimilarity
Sm-small items
Cj – j the cluster
E – Maximum ceiling

The maximum ceiling is the maximum number of transactions that might contain an item to call it a small item. Thus intra cluster dissimilarity is the union of distinct small items present in all the clusters.

Inter-cluster similarity: Inter-cluster similarity, on the other hand briefs us on the pair wise similarity between transactions present in different clusters. As their purpose is simple, these parameters need to be kept to a minimum for the clustering to be efficient. The incoming transactions are first assigned to one of the existing clusters or a new cluster is created to accommodate the incoming transaction. The decision on whether or not to create a new cluster is based on the cost parameter, i.e., a new cluster is created to accommodate the transaction if it reduces the overall cost of clustering.

$$Inter\ (U) = \Sigma kj=1|La\ (Cj, S)| - |Ukj=1\ La\ (Cj, S)|$$

Where
Inter (U) – Inter cluster dissimilarity
La – Large items
Cj – j the cluster
S – Minimum support

Minimum support indicates the minimum number of transactions in which an item should be present to claim to be a large item. The total cost is calculated by the following formula

$$Cost = w * Intra\ (U) + Inter\ (U)$$

Where
w - Item
Intra (U) - Intra cluster dissimilarity
Inter (U) – Inter cluster similarity

A new transaction is first put into each of the existing clusters and the cost is calculated for each cluster. Then a new cluster is created to accommodate the transaction and the cost is calculated. The transaction is then finally assigned to the cluster with the lowest cost value as follows.

For every new non-clustered transaction or every cluster ‘c’, a transaction is assigned to the cluster ‘c’, and then the cost is calculated and compared with the best cost. If the cost is improved, then the current cost is assigned to the best cost and best cluster. Therefore a new cluster is created for the current transaction and the cost of cluster is also calculated.

Combined rule pair: The SLR is calculated as explained in refinement phase. An attempt is then made to accommodate these transactions in a different cluster and examine the SLR of these transactions in the new cluster doesn’t exceed the threshold. If not these transactions are deemed outliers and they are eliminated from consideration. The process is explained as; calculate the S-L ratio of every transaction, then moves all the transactions whose S-L ratio exceeds the threshold to the excess pool. Shuffle the transactions in the excess pool to different clusters such that the S-L ratio value stays below the threshold. Delete the remaining transactions from the excess pool. The clustering process is thus complete, incorporating both the allocation and refinement phases.

Combined rule mining: Incremental association rule mining is used to produce frequent item sets and promised frequent item sets. As explained in original database transactions, a new algorithm assumes the statistics of new transactions slowly changed from original transactions. According to the assumption, the statistics of old transactions, obtained from previous mining, can be utilized for approximating the new transactions. Therefore, Support count of item sets gathered from previous mining may slightly differs from support count of item sets after inserting new transactions into an original database that contains old transactions. The new algorithm uses maximum support count of 1-item sets gathered from previous mining to calculate infrequent item sets of an original database that will capable of being frequent item sets when new transactions are placed into the original database. With maximum support count and maximum size of new transactions, now it allowed to pushing into an original database, support count for infrequent item sets that will be qualified for frequent item sets, i.e. min_PL is shown in equation:

$$Min\_sup_{DB} - [(maxsupp\ /\ total\ size) * inc\_size] <= min\_PL < min\_sup_{DB}$$

Where in_sup (DB) is minimum support count for an original database, maxsupp is a maximum support count of item sets, the current size is a number of transactions of an original database and inc_size is a maximum number of new transactions. Here, a promising frequent item sets is defined as follows.

A promising frequent item set is an infrequent item set that satisfies the equation. In this paper, apriori algorithm is applied to find all possible frequent k- item sets and promising frequent k-item sets. Apriori scans all transactions in original database for each iterate with two step process, which are joining and prune step. Unlike typical apriori algorithm, items in both frequent k- item sets and promising frequent k-item sets can be joined together in the joint step. For a frequent item, its support count must be higher than a user-specified minimum support count threshold and for a promising frequent item; its support count must be higher than min_PL but less than the user-specified minimum support count.

As explained in updating frequent and promising frequent items, it updates all old items. The size of an updated database increases when new transactions are placed into an original database. Thus, min_PL must be recalculated in order to associate with the new size of an updated database. min_PL (update) is computed as the follows:

$$Min\_PL_{U\ db} = min\_sup_{DB\ U\ db} - (max\ supp\ /\ total\ size * inc\_size)$$

Then, if any k-item has support count greater than or equal to min_sup (DBUdb), this item set is moved to a frequent k-item of an updated database. In the other case, if any k-item has support count less than min_sup (DBUdb) but it is greater or equal to min_PL (update), this k-item is moved to a promise frequent item sets of an updated database. The following algorithms are developed to update frequent and promising frequent k-terms of an updated database. Then a decision process will be taken place.

C. Product Recommendation

After decision making, the recommendation is clarified that Includes the Product rate, preferable range of products and changes in the Wish List of Customer type set. Lastly recommended items are supplied to the customers. To maintain the transaction history, all the transactions are stored into the databases.

5 DISCUSSIONS

Dynamic product recommendations are provided to each customer for helping to purchase an item. The customers are classified by the clustering process and threshold value was found. Incremental association rule mining is used to find frequently purchased items and promising frequent items. The transaction history of the customers is stored in the database for producing dynamic recommendation.

6 RESULTS

To generate the dynamic recommendations, we used incremental association rule mining and SLR is calculated. Transaction History of the customers is maintained in the databases. To classify the customer, cluster process has been used. In cluster process, we had discussed the intra cluster dissimilarity and the inter cluster similarity. After cluster formation, the threshold value is found by calculating the SLR. Based on the threshold value of items, recommendations are generated.

7 CONCLUSION

With the help of Incremental Association Rule Mining and Transaction Clustering, we introduced a method to design an improved and well structured website design for an E-shop in the design phase. Taking for granted that the two inceptions, least support and assurance will not change. The promising frequent algorithm can guarantee to find frequent item sets. Efficient clustering process is used for data items to reduce the SL ratio in each group. The process is able to cluster the data items very efficiently. This process not only incurs an execution time but also Guides the clustering results to a very good quality.

REFERENCES

[1]Greg Linden, Brent Smith, and Jeremy York, “Amazon.com Recommendations Item-to-Item Collaborative Filtering”, IEEE Computer Society, Jan 2003.

[2]Gediminas Adomavicius, and Alexander Tuzhilin, A., “Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions.” IEEE Transactions on Knowledge and Data Engineering, Vol. 17, No. 6, June 2005.

[3]SilvanaAciar, Debbie Zhang, Simeon Simoff, and John Debenham, “Informed Recommender: Basing Recommendations on Consumer Product Reviews.”, IEEE Computer Society , June 2007.

[4]Zan Huang, Daniel Zeng and Hsinchun Chen, “A Comparison of Collaborative-Filtering Recommendation Algorithms for E-commerce”, IEEE Computer Society, October 2007.

[5]A. C. M. Fong, Baoyao Zhou, S. C. Hui, Guan Y. Hong and The Anh Do, “Web Content Recommender System based on Consumer Behavior Modeling.”, IEEE Transactions on Consumer Electronics, Vol. 57, No. 2, May 2011.

[6]Alexandros Nanopoulos, “Item Recommendation in Collaborative Tagging Systems.” IEEE Transactions on Systems, Man and Cybernetics-Part A: Systems and Humans, Vol. 41, No. 4, July 2011.

[7]Hyea Kyeong Kim, Young U. Ryu, Yoonho Cho, and Jae Kyeong Kim, “Customer-Driven Content Recommendation over a Network of Customers.” IEEE Transactions on Systems, Man and Cybernetics-Part A: Systems and Humans, Vol. 42, No. 1, January 2012.

[8]Qi Liu, Enhong Chen, HuiXiong, Chris H. Q. Ding, and Jian Chen, “Enhancing Collaborative Filtering by User Interest Expansion via Personalized Ranking.” , IEEE Transactions on Systems Man and Cybernetics-Part B: Cybernetics, Vol. 42, No. 1, February 2012.

[9]Q. Shambour and J. Lu, “A trust-semantic fusion-based recommendation approach for e-business applications,” Decision Support Systems, vol. 54, pp. 768-780, 2012.

[10]X. Yuan, J.-H. Lee, S.-J.Kim, and Y.-H.Kim, “Toward a user-oriented recommendation system for real estate websites,” Information Systems, vol. 38, pp. 231-243,2013.