INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Analysis on Missing Item Prediction and Its Recommendation Based on Users Approach in Ecommerce

Mr.Himanshu M. Deulkar[1], Prof. R. R. Shelke[2]

[1]ME (CSE) , [2]Asst. Professor

Hvpm Coet ,Amt

Abstract: The Internet is one of the fastest growing areas of intelligence gathering. Due to the tremendous amount of data on internet, web data mining has become very necessary. Predicting the missing items form dataset is indefinite area of research in Web Data Mining. Current approaches use association rule mining techniques which are applied to only small item sets. Numbers of mechanisms were intended for frequent item sets but less attention has been paid that take the advantage of these frequent item sets for prediction purpose. In order to reduce the rule mining cost for large dataset & to provide online prediction efficiently, the proposed approach use novel method for predicting the missing items. The proposed approach extends advantages of prediction at a higher level of abstraction and reduced rule generation complexity by finding out a technique that will work on dissimilar approach.

I.   INTRODUCTION

With the popularity of network, E-commerce has gained fast development and accumulated a huge number of faithful online users all over the world [1]. Through E-commerce, users can browse, compare and select the product items that they like in a more convenient manner, which brings great facility to the Ecommerce users [2]. Today, many E-commerce companies (e.g., Amazon, eBay, Bestbuy) have provided various product items to their massive online users.

In proposed work, I wanted to make the next logical step by allowing any item to be treated as a class label—its value is to be predicted based on the presence or absence of other items. Put another way, knowing a subset of the shopping cart's contents, I want to "guess" (predict) the rest. Suppose the shopping cart of a customer at the checkout counter contains bread, butter, milk, cheese, and pudding. Could someone who met the same customer when the cart contained only bread, butter, and milk, have predicted that the person would add cheese and pudding? Implicitly or explicitly, this task stood at the cradle of this field; now that many practical obstacles (e.g., computational costs) have been reduced, I want to return to it.

In general, the traditional Collaborative Filtering (i.e., CF) based recommendation approaches can work very well, when the target user has one or more similar friends (i.e., user-based CF), or the target user's purchased and preferred product items own one or more similar product items. Different from the traditional CF-based recommendation approaches where I look for "similar friends" or "similar product items" directly, I first look for the target user's dissimilar "enemy" (i.e., antonym of "friend"), and furthermore, I look for the "possible friends" of E-commerce target user, according to "enemy's enemy is a friend" afterwards, the product items preferred by the target user's "possible friends" are regarded as the recommendation candidates for target user; likewise, for the product items preferred by target user, I first determine their "possibly similar product items" based on "enemy's enemy is a friend" rule of Structural Balance Theory, and regard them as the recommendation candidates for target user.

The Ecommerce recommendation problem is formalized and afterwards, the proposed work motivation is demonstrated. A recommendation approach over big rating data in E-commerce, is put forward. A set of experiments are designed and deployed in later section for validating feasibility in terms of recommendation accuracy, recall and efficiency. In Last Section, I summarize the whole proposed work and suggest the possible research directions in the future.

A.   Literature Review & Related Work

Association rle mining (ARM) in its original form finds all the rules that satisfy the minimum support and minimum confidence constraints. Most works allow the user to specify a QoS-based service selection policy: [4] it works its work is to get the result depending upon the surrounding searches or the users of same locations. [7] its objective is to get the most appropriate results by

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

implementing parameters.[3] and [6] proposes of implementing various patterns to retrieve data without generating any candidate keys.

The goal was to build a classifier using so-called class association rules. In classification rule mining, there is one and only one predetermined target, the class label. Most of the time, classification rule mining is applied to databases in a "table" format, with a predefined set of attributes and a class label. Attributes usually take a value out of a finite set of values (although missing values are often permitted). In our task, I do not have a predefined class label. In fact, all items in the shopping cart become attributes and the presence/absence of the other items has to be predicted. What is needed is a feasible rule generation algorithm and an effective method to use to this end the generated rules. For the prediction of all missing items in a shopping cart, our algorithm speeds up the computation by the use of the item set trees (IT-trees) and then uses DS theoretic notions to combine the generated rules.

In time-aware recommendation is introduced, where time is considered as an important factor for predicting product quality. However, work only discusses the objective quality prediction, without considering the subjective preferences of different users. Matrix factorization technique is introduced in [8] to realize the recommendation purpose; however, if the user-product rating matrix is very sparse, the recommendation effect is not as good as expected (e.g., overfitting problem). In [6], a Monte Carlo algorithm named MCCP is brought forth to measure different users' personalized preferences towards different product items. According to MCCP, user target 's similar friends can be found by trust propagation; and afterwards, the missing product item quality could be predicted based on the obtained similar friends. Generally, MCCP can work very well if user target has similar friends. However, as introduced previously in this proposed work, I only focus on the specific recommendation situations when user target does not have similar friends; therefore, prediction accuracy and recall of MCCP are not as good as expected, which has been validated by the experiments. In our previous work , a recommendation approach USER BASED PARAMETER is put forward, for dealing with the specific recommendation scenarios where user target has no similar friends and the product items liked by user target do not have similar product items. While user based parameter approach has two shortcomings. First, only "enemy's enemy is a friend" rule is recruited in user based parameter. Second, user based parameter only adopts user-based CF recommendation, while neglects item-based CF recommendation as well as their integration. Therefore, the recommendation eff   ect of user based parameter is not as satisfactory as expected. In view of the shortcomings of above approaches, I put forward a novel product item recommendation approach User based parameter. Through "enemy's enemy is a friend" and "enemy's friend is an enemy" rules in Structural Balance Theory, user based parameter can make full use of the valuable structural balance information hidden in user-product purchase network, and further make precise product item recommendation. Moreover, User based parameter integrates both user-based CF recommendation and item-based CF recommendation; therefore, the recommendation recall could be improved. Finally, through a set of experiments deployed on MovieLens-1M dataset, I validate the feasibility of User based parameter in terms of recommendation accuracy, recall and efficiency.

II.   EXISTING SYSTEMS

Existing approaches in this area use item set trees and fast algorithms. These approaches employ association rule mining techniques. The first approach uses Item set Trees to establish the association rules between the items. The uncertainty in the occurrence is measured using Bayesian techniques and Dempster's-Shafer theory. In this method the authors generate all high support and high confidence rules using item set trees. Then consequents of all these rules are combined to give an estimated completion of shopping cart. This technique proves to appear better than the traditional techniques in association rule mining. However the disadvantage of this approach is that the rule generation complexity increases greatly with the increase of the average length of the transaction and with the number of distinct items. Yet another method to predict missing items uses Boolean vector and the relational AND operation to discover frequent item sets [3] [6] without generating candidate items directly generate the association rules. Association rules are used to recognize the correlation among a set of items in database. At beginning, Boolean Matrix is generated by transforming the database into

Boolean values (either 0's or 1's). The frequent item sets are generated from the Boolean matrix. The association rules generated from the frequent item set for prediction. The next item set i.e the content of incoming shopping cart will also be represented by a Boolean vector and "AND" operation is performed with each transaction vector to generate the association rules. Lastly the rules are combined by Dempster's rule of combination to get the predictions. The advantages of this technique are that it doesn't generate candidate item sets, it uses only a single pass over the database, the memory consumption is low and the processing speed is more as compared to the previous technique. The disadvantage of this approach is that the use of a Boolean matrix cannot handle huge amount of data. This disadvantage restricts the use of this technique in online applications where huge amount of data is generated.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

The major drawback of the above mentioned technique is the rule generation complexity. Generating rules from a huge amount of data involves a lot of high memory and time complexity. In DS – ARM technique the rule generation complexity increases by huge amount as the average transaction length increases and in the technique using fast algorithms though the rule generation complexity is lower the data structures used are not capable of handling huge amounts of data. In addition to these techniques another proposed work discusses a graph based approach towards association rule mining. This proposed work suggests an algorithm called Combo Matrix algorithm which predicts missing items using associative classification mining. The data structure used to store a graph is an adjacency matrix with a slight modification wherein the diagonal elements contain the list of vertices the adjacent to the given vertex. The advantages of this approach are reduced rule generation complexity however; the data structure used and the algorithm proposed do not perform well for large data sets since an adjacency matrix is used thereby having space complexity of $O|V2|$ where V is the number of vertices and equal to the number of items present in the database. The method proposed in this proposed work aims at reducing the rule generation complexity by classifying the items and then constructing a graph from it and using the graph for prediction purpose.

III. ANALYSIS OF PROBLEM

Two major problems complicate the task: first, how to identify the relevant rules in a computationally efficient manner; second, how to combine (and quantify) the evidence of conflicting rules.

Furthermore, I need to be aware of the circumstance that the presence of an item might suggest the absence of other items. For example, if the shopping cart contains chips, cookies, cashews, the customer may not buy nuts. I was therefore interested in rules such as (chips;cookies;cashews):nut, where nuts means that no nuts will be added to the cart. Classical association mining usually ignores this aspect, perhaps because negated items tend to increase significantly the total number of rules to be considered; another reason can be that rules with mutually contradicting consequents are not so easy to combine. With all these issues in mind, I narrow down the space of association rules by the following guidelines:

For a given item set s, rule antecedents should be subsumed by s.

The rule consequent is limited to any single "unseen" item (presence or absence of the unseen item). In essence, this proposed work addresses the following tasks: Given a transaction with the item set s I, find the set of matching rules (entailed by the training data set) that are of the form) ij;j¼ 1;n, such that s and ij = 2 s, and exceed the user-supplied minimum support, s, and minimum confidence, c, thresholds. Then, devise a method to combine the matching rules that have mutually contradicting consequents and reach a decision on which other items would be added to the transaction

IV. DESIRED IMPLICATIONS

The mechanism reported in this proposed work focuses on one of the tasks in association mining: based on incomplete information about the contents of a shopping cart, can I predict which other items the shopping cart contains? Our literature survey indicates that, while some of the recently published systems can be used to this end, their practical utility is constrained, for instance, by being limited to domains with very few distinct items. Bayesian classifier can be used too, but I was not aware of any systematic study of how it might operate under the diverse circumstances encountered in association mining.

REFERENCES

[1] Lianyong Qi, Xiaolong Xu, Xuyun Zhang, Wanchun Dou, Chunhua Hu, Yuming Zhou, Jiguo Yu , "Structural Balance Theory-based E-commerce Recommendation over Big Rating Data" 2016 IEEE

[2] Kasun Wickramaratna, Miroslav Kubat,and Kamal Premaratne,"Predicting Missing Items in Shopping Carts" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 21, NO. 7, JULY 2009

[3] Junqiang Liu, Member, Ke Wang, and Benjamin C.M. Fung,"Mining High Utility Patterns in One Phase without Generating Candidates" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 28, NO. 5, MAY 201

[4] R. R. Shelke, Dr. R. V. Dharaskar, Dr. V. M. Thakare, "Analysis of User Movement Behaviour Pattern Mining Approaches for Mobile Environments", International Journal of Advanced Research in Computer Science, Volume 5, No. 4, April 2014, ISSN No. 0976-5697.

[5] M. Mayilvaganan,D. Kalpanadevi"Comparison of Classification Techniques for predicting the Cognitive Skill of Students in Education Environment" 2014 IEEE International Conference on Computational Intelligence and Computing Research

[6] Junqiang Liu,Ke Wang, and Benjamin C.M. Fung"Mining High Utility Patterns in One Phase without Generating Candidates" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 28, NO. 5, MAY 2016.

[7] R. R. Shelke, Dr. R. V. Dharaskar, Dr. V. M. Thakare, "DATA MINING FOR MOBILE DEVICES USING WEB SERVICES", International Conference on Industrial automation And Computing (ICIAC - 12th & 13th April 2014) , Jhulelal Institute of Technology, Nagpur

[8] Jos´e Cazal "Predictive models of economic systems based on data mining" International Workshop on Data Mining with Industrial Applications 2015

[9] An Approach for Predicting the Missing Itemsfrom Large Transaction Database" IEEE Sponsored 2nd International Conference on Innovations in Information

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Embedded and Communication Systems ICIIECS 1

[10] An Improved Collaborative Filtering Recommendation Algorithm not Based on Item Rating" Proc. 2015 IEEE 14th Int'l ConI. on CognltlVelnlormatlcs & Cognitive Computing IICCrCC'151

[11] Predicting Missing Items in a Shopping Cart using Apriori Algorithm" International Journal of Engineering Trends and Technology (IJETT) – Volume 20 Number 4 – Feb 2015

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 ◯ (24*7 Support on Whatsapp)