IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 2, February 2016

Data Mining Technique to Predict Missing Items and Find Optimal Customer for Beneficial Customer Relationship and Management

Saroja T.V.[1], Greeshma Nair[2], Kajal Nalawade[3], Yojana Prajapati[4]

Associate Professor, Computer Department, SSJCOE College, Dombivli, India [1]

Student, Computer Department, SSJCOE College, Dombivli, India [2, 3, 4]

Abstract: The aim of association rule mining is to find frequently co-occurring groups of items in transactional databases. The intention of this knowledge is for prediction purposes. This paper contributes a technique that uses the partial information about the contents of a shopping cart for the prediction of products that the customers wish to buy or are more likely to buy along with the already bought products. So this paper presents a technique called the "Combo Matrix" whose principal diagonal elements shows the association between items and looking to the principal diagonal elements, the customer can choose different items that can be bought with the purchased contents of the shopping cart and also reduces the rule mining cost. In this paper, we also propose a data mining and artificial technique to maintain the customer relationship between company and customers. For this purpose, we maintain a historical database and then we use data mining ARM technique to get the customer information from this database. Also we use Customer Relationship Management (CRM) systems which are developed and used to support marketing, customer interactions, preferences and data.

Keywords: Association rule mining, Prediction, Combo Matrix, Customer Relationship Management, data mining.

I. INTRODUCTION

Data mining is the extraction of hidden predictive information from databases. Generally, data mining is the process of analysis of data from different perspectives and summarizing it into useful information used to increase profit, costs, or both. For data analysis, data mining is used as analytical tool. Users are permitted to analyse data from different perspectives and summarize the relationships identified. For example, if the customer is buying mobile phone then we can suggest him to buy power bank along with the mobile phone. This prediction is done on the basis of previous knowledge of his and other customer's transactional database.

Customer relationship management (CRM) refers to plans, practices and technologies that organizations use to manage and analyse customer interactions and customer data. CRM systems also give detailed information on customer's personal information, history of purchase, buying preferences**.** Apriori is an algorithm which is used for learning association rules. It is designed to operate on databases which contains transactions (for example, collections of items that the customer purchases). Particle swarm optimization is a computational method that optimizes a problem by repeatedly trying to improve a candidate solution with a given measure of quality. For example, we take into consideration transaction of two customers say x and y. Suppose customer x logs into his account five times but he buys only one item and customer y logs into his account three times but he buys six items then on the basis of this knowledge and the algorithms we will give more discount to customer y as compared to customer x.

II. RELATED WORK

Customer relationship management is used in business strategy for maintaining good customer relationships.

The Graph based algorithm proposed [1], [2] efficiently solves the problem of mining association rules. We can draw a graph on the basis of large itemsets where each of these itemsets is numbered randomly and they are stored in database in the form of bit vectors.

A bit vector represents a transaction, where 1 means that the item is present and 0 means that the item is absent. Prediction of missing items [3], [4] uses flagged itemset trees (IT-TREE) concept for rule generation. An itemset tree A, is made up of root and a (possibly empty) set {A1, A2,…., Ak} each element of which is an itemset tree. The root is a pair [s, f(s)], where itemset is denoted by s and frequency is denoted by f(s).

III. EXISTING SYSTEM

One of the challenges of existing system is that it is time-consuming and difficult to design different recommendation algorithms for different recommendation purposes. It can be seen that most of these recommendation problems have some common properties, where a general framework is needed to merge the task of recommendation on the Web. Also most of the methods which currently exist are complicated and require tuning a large number of parameters.

Disadvantage: It is becoming much difficult to find contents that are relevant and also it is difficult to understand what user actually recommends.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 2, February 2016

IV. PROPOSED SYSTEM

With the aim of satisfying the information needs of Web users and also to improve the experience of the users in many Web applications we use Recommender Systems. This is a technique which will automatically predict the interest of an active user by collecting information about the rating from other similar users or products. The primary aim of collaborative filtering is that the users which are active will prefer those items which other similar users have already preferred. The proposed method consists of two stages: generating queries for the candidates and determining the generalization or specialization relations between these queries. The method initially uses a small set of linguistically motivated extraction patterns relevant to each entry from the query logs and then it applies a series of enhancement filters to rank the candidate attributes.
Advantages: Using this model one can get personalized recommendations. It is scalable to large datasets.

A. Modules:

1) Posting the opinion: This module is used to get the opinions from various people about business, e-commerce and products through online. The opinions may be of two types. They can be either direct opinion or comparative opinion. Direct opinion means to post a comment about the product and components of the product directly. Whereas comparative opinion means to post a comment by comparing two or more products. The comments can be positive or negative.

2) Image Recommendation Technique**:** Image recommendation is one of the interesting applications on web. Its main aim is to recommend interesting images to users based on their preference. At first this system first asks the user to rate some products on the basis of their likes and dislikes and then it will recommend images to the users based on the tastes of the users.

3) Collaborative Filtering**:** User-based approaches predict the ratings of active users based on the ratings of their similar users. Items similar to those chosen by the active user are predict by Item-based approaches.

4) Ranking Approach: In this we rank items according to the rating variance of neighbors of a particular user for a specific item. Different ranks are existed for an approach that can improve recommendation diversity other with topmost predicted rating values to a user with suggesting items.

V. PSEUDO CODE

A. ALGORITHM STEPS-:

1) Algorithm to Generate Key Item
Item: name of the item to be sold.
Key=1;
IndexMatr:matrix which stores name of the item.
for i=1: length(Item)
IndexMatr (i, 1) =item (i);
IndexMatr (i, 2) =key++;
end

2)Combo Matrix initialization
InitializeComboMat (IndexMatr)
Combo Mat: matrix to represent association graph
for i=1: length (IndexMatr)
for j=1: length(indexMatr)
ComboMat(i,j)=0;
end
end

3)To predict the items
predict (ComboMat, threshold, key_Index)
threshold:minimum value required by pair
key_Index: list of item unique key which are purchased by the customer
predict _Item: it represent an array of item to be predicted, which is initially NULL
product: it contains each item key_Index that is selected by customer.
pair: it contains list of item which will be stored in diagonal element of Combo matrix of product.
pair_Index: it will contain individual item key index of each pair.
edge_Value: it is used for storing the edge value between different pair.
for i=1: length (key_Index)
product=key_Index (i);
pair = ComboMat (product, product);
for j=1: length (pair)
pair_Index=pair (j);
edge_Value= ComboMat (product, pair_Index);
If edge_Value >= threshold
Predict_Item=predict_Item U pair_Index
end
end
end
return predict_Item;

4)Combo Matrix Updation
UpdateComboMat (key_Index)
key_Index: name of item key_Index that the customer has purchased.
for i=1:length(key_Index)
producti=keyIndex(i);
for j=1:length(keyIndex)
productj=keyIndex (j);
if (producti==productj)
ComboMat(producti,productj)U=key_Index
else
ComboMat(producti,productj) +=1;
end
end
end

5)Transaction algorithm
do Transaction (item)
item: name of item that customer has selected
threshold=2;
key_Index: Null
for k=1:length(item)
for i=1:length (indexMatr)
product=indexMatr(i, 1);

DOI 10.17148/IJARCCE.2016.52114

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 2, February 2016

```
if product==item(k)
key_Index=key_Index U indexMatr(i, 2);
break;
end
end
end
predict_Key=Predict(ComboMat, threshold, key_Index)
if (choice)
item=item U newSelecteditem from the predict_Key
doTransaction(item);
else
updateComboMat (key_Index)
end
return;
```

B. Apriori Algorithm

Pass 1

1. To generate candidate item_sets in A_1
2. To save frequent item_sets in B_1

Pass k

To generate candidate item_sets in A_k from frequent item_sets in B_{k-1}

1. Join B_{k-1} p with B_{k-1}q, as given below: **insert into** A_k **select** p.item$_1$, p.item$_2$, . . . , p.item$_{k-1}$, q.item$_{k-1}$ **from** B_{k-1} p, B_{k-1}q
   where p.item$_1$ = q.item$_1$, . . . p.item$_{k-2}$ = q.item$_{k-2}$, p.item$_{k-1}$ < q.item$_{k-1}$
2. Now generate all (k-1) subsets from the candidate item_sets in A_k
3. Now prune all candidate item_sets from A_k where some (k-1) subset of the candidate item_set is not in the frequent item_set B_{k-1}
4. Next step is to scan the transaction database and to determine the support for each candidate item_set in A_k
5. Now save the frequent item_sets in B_k

VI. ARCHITECTURE DESIGN

A. THE STRUCTURE OF CRM SYSTEM**:** The customer information is stored in the field Customer ID and their date of attainment is stored in the date column and the number of transactions done by the specific customer is stored in the field Number of transactions. The particular customer has given company revenue which is called profit.

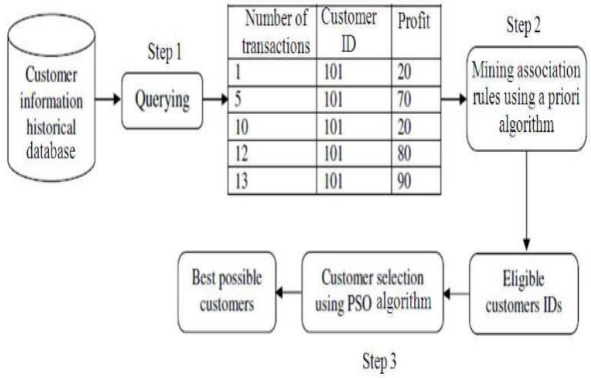


Fig: Structure of CRM system

B. RECOMMENDATION SYSTEM**:**

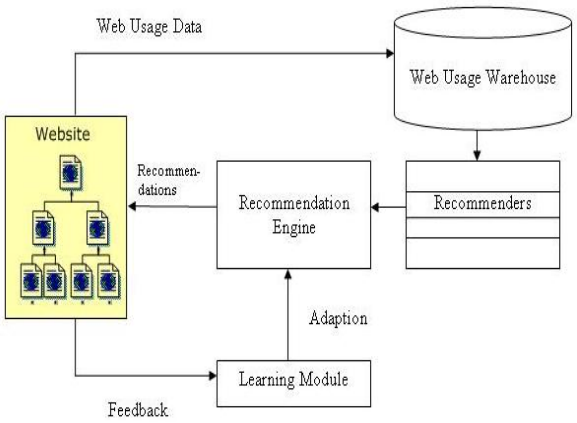
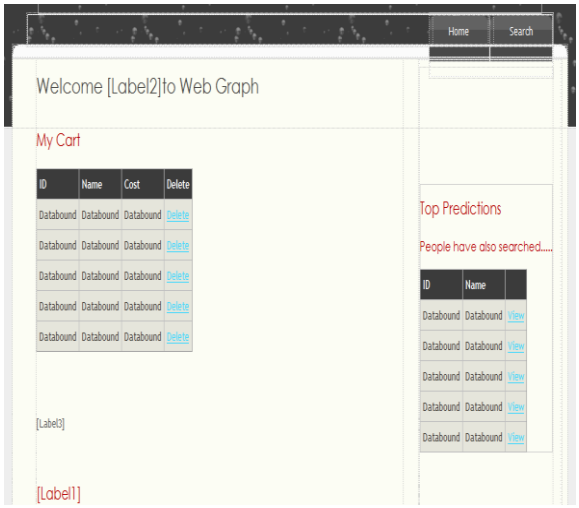
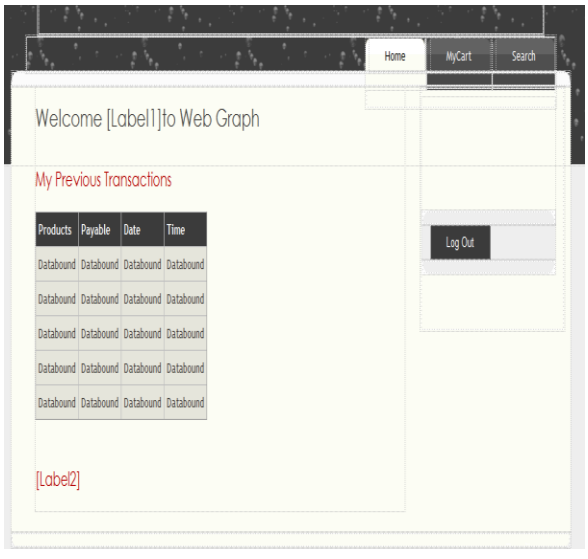


Fig: Architecture of recommendation system

VII. DISCUSSION OF RESULT

Results from our approach has an efficient performance. It shows the transaction history of the customer and also predicts items to the customer on the basis of similar user's history. It gives discounts to suitable customers by considering their history of transaction.





DOI 10.17148/IJARCCE.2016.52114

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 2, February 2016

VIII. CONCLUSION

We have studied various algorithms for transactional database which improved the execution time and reduce consumption of memory. The CRM system that we have used provides attractive offer to the customers where they have frequently visited and also provides high profit to the company. It is user friendly and more flexible.

REFERENCES

[1] Yen S.J. and Chen A "An efficient approach to discovering knowledge from large database". In proc. Of the IEEE/ACM International Conference on parallel distributed Information system, Pages 8-18, 1996.

[2] Lee K.L, G Lee and Chen A.L.P. Efficient Graph based Algorithm for discovering and maintaining association rules in large database.

[3] M. Kubat, Hafez. A, Raghavan V.V, J.R. Lekkala and hen W.K.: "Itemset Trees for Targeted Association Querying," IEEE Trans .Knowledge and Data Eng., vol. 15, no. 6, pp. 1522-1534, Nov./Dec.2003.

[4] Kasun Wickramaratna," Predicting missing Items in Shopping Carts", IEEE Transaction on Knowledge and data engineering, Vol. 21, No. 7, July 2009.

DOI 10.17148/IJARCCE.2016.52114