Using Apriori Algorithm To Improve Crm For Shopping Malls

P.Sravanthi
Assistant professor,
Anurag Engineering College,
Kodad.

M. Madhavi
Assoc Professor,
Anurag Engineering College,
Kodad.

Abstract

At present, customer orientation has been one of the major concerns of commercial companies. Customer Relationship Management (CRM), influences the relationship of customers with commercial companies which inevitably contributes in the investments and profits of the company. Data mining discovers the frequent item sets of customers' needs. The purpose of this research which has been done on the data base of a Shopping mall is to pattern the behaviours of customer's interests, in time intervals using the time series analysis. The patters change according to festival offers, seasonal discounts, different offers e.t.c. Discovering these patterns by taking into account the number of frequent item sets will be able to meet the demands of the customers properly. In this case the customer is changed to a regular customer and will increase the profits in the long term.

Keywords: Data Mining, Customer Relationship Management (CRM), Time Series Analysis.

1. Introduction

Shopping malls are important not just to customers, but also employees and indeed to many others because of the investments of their pensions. Shoppers tend to follow the provision of attractive shopping areas. Improving shopper satisfaction can lead to changes in customer population, investments, and profits.

The main focus of most industries is on customer services. Discovering the real and main needs of Customers and providing them with proper services according to their needs, forms the basic rules of customer relationship management. Quick and in-time attention to the basic needs of the customers is the most important factor in reducing the costs and increasing the economic benefits.

In the present world, due to numerous number of shopping malls in various locations, a great load of data which indicates interestingness of people in different trends(based on seasons, climate, tours and trips) , is a good source in discovering the rules and the relationships between shopping malls and these preferences. By knowing these relationships, providing a more appropriate way and considering the future needs, it is possible both to satisfy the customers and also to reduce the costs of the products. On the other side, terms of attracting the customers in the first and second cycle of market's lifespan become operational. In this research we will try to focus on those parts which relates to the customer relationship management to be relevant with the operations of products and prices'.

2. Data Mining

Data Mining refers to extracting knowledge form large amounts of data. Knowledge discovery as a process of an iterative sequence of following steps

Data cleaning: To remove noise and inconsistent data.
Data integration: Where multiple data sources may be combined.
Data selection: Where data relevant to the analysis tasks are retrieved from the database.
Data transformation: Where data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations.

Data Mining: An essential process where intelligent methods are applied in order to extract data patterns.

Pattern evaluation: To identify the truly interesting patterns representing knowledge based on some interestingness measures.

Knowledge presentation: Where visualization and knowledge representation techniques are used to present the mined knowledge to the user.

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both.

Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. In data mining frequent item sets play a main role to obtain the relevant data to improve CRM.

3. CRM

It is a widely implemented model for managing a company's interactions with customers, clients, and sales prospects. It involves using technology to organize, automate, and synchronize business processes—principally sales activities, but also those for marketing, customer service, and technical support. The overall goals are to find, attract, and win new clients, service and retain those the company already has, entice former clients to return, and reduce the costs of marketing and client service. Customer relationship management describes a company-wide business strategy including customer-interface departments as well as other departments. Measuring and valuing customer relationships is critical to implement this strategy.

CRM(Customer Relationship Mangement) provides presenting a single image of the organization; Understanding the customers' likes and dislikes; Anticipating customer needs and addressing them proactively; and Recognizing when customers are dissatisfied and taking corrective action.

4. Time Series Analysis

Time series *analysis* comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. **Time series** *forecasting* is the use of a model to predict future values based on previously observed values. Time series are very frequently plotted via line charts.

The below chart indicates that the customers can purchase the goods frequently or some times may not. So the best-fit line is indicated likethis.
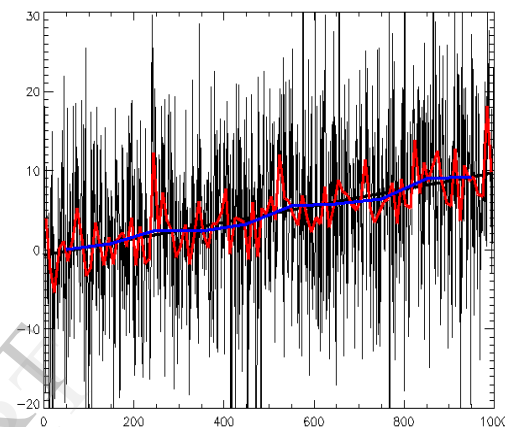

Figure 1. Random Data Plus Trend, with best-fit line in Time Series
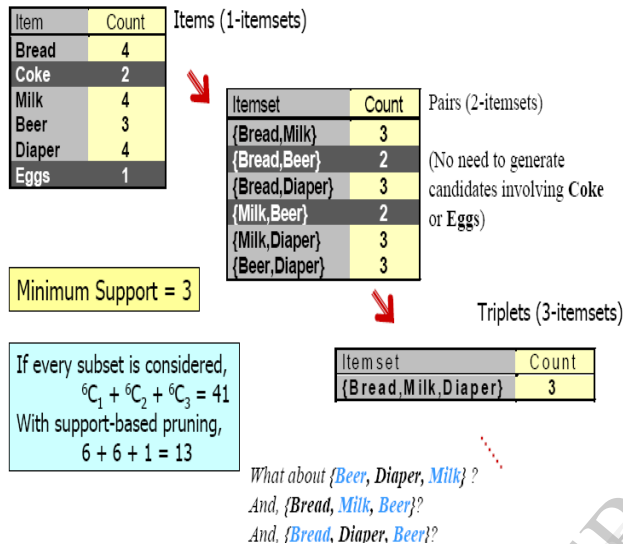
5. Frequent Item set Mining

A set of items, subsequence, substructures, etc., that occurs frequently in a data set is called a frequent item sets. We can identify the frequent item sets using **Apriori** algorithm, it is based on a property that "All nonempty subsets of a frequent item set must also be frequent".

A two-step process is followed, consists of join and prune actions.

Join step: To find *Lk,* a set of candidate *k*-item sets is generated by joining *Lk-1* with itself. This set of candidates is denoted *Ck*. Apriori assumes that items with in the item sets are sorted in lexicographic order i.e $li[1]<li[2]<\ldots.<li[k-1]$. The join, Lk-1 and Lk-1 is performed when members of Lk-1 is joinable if their (k-2) items are in common. That is, members l1and l2 are joined if $(l1[1]=l2[1])^\wedge\ldots\ldots^\wedge(l1[k-1]<l2[k-1])$. The

condition l1[k-1]<l2[k-1] simply ensures that no duplicates are generated.

Prune step: Initially, scan DB once to get frequent 1-itemset. Generate length (k+1) candidate item sets from length k frequent item sets. Test the candidates against DB. Terminate when no frequent or candidate set can be generated

Algorithm

Ck: Candidate item set of size k

Lk : frequent item set of size k

L1 = {frequent items};

for (*k* = 1; *Lk* !=∅; *k*++) **do begin**

  Ck+1 = candidates generated from *Lk*;

  for each transaction *t* in database do

   increment the count of all candidates in *Ck+1*

 that are contained in *t*

  Lk+1 = candidates in *Ck+1* with min_support

  end

return ∪*k Lk*;

  end

Step 1: **self-joining *Lk-1***

insert into *Ck*

select *p.item1, p.item2, …, p.itemk-1, q.itemk-1*

from *Lk-1 p, Lk-1 q*

Where *p.item1=q.item1, …, p.itemk-2=q.itemk-2, p.itemk-1 < q.itemk-1*

Step 2: **pruning**

for all **item sets c in Ck** do

for all **(k-1)-subsets s of c** do

if *(s is not in Lk-1)* **then delete** *c* **from** *Ck*

The mined data will contain the customers' information which can be used by the CRM. CRM includes the methodologies, strategies, software, and web-based capabilities that help an enterprise to organize and manage customer relationships and also identify the customers' behaviour, with the help of frequent item set mining. we can easily identify the customer trends in different season's i.e in the summer season customers usually prefer the cotton clothes, in the winter woollen clothes and in the rainy , raincoats, umbrellas etc are more demanded by the customers.

Time series analysis is going to compare the trends in different seasons. Time series based database consists of sequences of values obtained over repeated measurements of time. The values are typically measured at equal time intervals (eg: summer, winter, rainy, festival offers). Time series based databases are popular in many applications, such as stock market analysis, economic and sales forecasting, budgetary analysis, utility studies, inventory studies, yield projections, workload projections, process and quality control, observation of natural phenomena(such as atmosphere, temperature, wind, earthquake), scientific and engineering experiments, and medical treatments.

6. Conclusion

Thus we can conclude that time series based databases are playing an important role in developing customer relationships which improve investments and profits to all sectors of an enterprise .Even this can be extended to college organizations ,software, politics, agriculture.

7. References

[1] Pang-Ning Tan, Michael Steinbach and Vipin Kumar (2006)." Introduction to Data Mining", Addison Wesley",

ISBN: 0-321-32136-7

[2] M.H. Nikbakhsh Tehrani, (2001). "Electronic Commerce", Tehran: Isiran institute.

[3] Bose and Sugurmaran (2003), "Application of Knowledge Management Techniques inCustomer relationship

Management", Journal of Knowledge and Process Management, Vol.10, No. 1. pp.3 - 17.

[4] Lisaand Pritscher and Hans Fey en, (2004). "Data Mining And Strategic Marketing In TheAirline Industry",

Switzerland, CH-8058 Zurich-Airport.

[5] S.M.T. Roohani Rankoohi, (2004). "Fundamental Concepts of Databases", Tehran,Jelveh.

[6] Giudici. Paolo (2003). "Applied Data Mining Statistical Methods for Business and Industries", Wiley. London

[7,8] M. Kantardezik, (2006). "Data Mining", Babol: Uloome Rayane

[9] FROST, R.A. (1987). "An Introduction to knowledge Base Systems", Collins. London.

[10] A. Adel, (2001). "Statistic and Application in Management", Tehran: SAMT.

.