

Research Article

Business Support System using Hybrid Classification Algorithm

Kritika Walinjkar[†], Fiyansh Shah^{†*} and Sonal Maskeen[†]

[†]Department of Computer Engineering, Atharva College of Engineering, Marve Rd, Malad (west), Mumbai-95, Maharashtra, India

Accepted 20 Feb 2015, Available online 25 Feb 2015, Vol.5, No.1 (Feb 2015)

Abstract

Cataloguing and patterns extraction from customer data is very important for business support and decision making. Timely recognition of newly emerging trends is needed in business process. Changing market trends need to be taken into consideration for predicting which products have more demand. This paper is about integrating two different algorithms, one is clustering algorithm, which is K-means and other is to find most frequent pattern i.e MFP which will help the back end of a company i.e production and inventory management unit to understand what product is selling more and which has a slow selling rate. In this way company can increase their profit by stocking the market with only those products that people buy.

Keywords: K-Means, Most Frequent Pattern (MFP), Data mining.

1. Introduction

Data is very important for every organization and business. Data that was measured in gigabytes and is now being measured in terabytes, and will soon approach the peta byte range. In order to achieve our goals, we need to fully exploit this data by extracting all the useful information from it.

Unfortunately, the size and complexity of the data is such that it is impractical to manually analyse, explore, and understand the data. As a result, important information is often unnoticed, and the prospective benefits of increased computational and data gathering capabilities are only partially realized.

Sale data classification has different market trends. Some clusters or segments of sale may be growing, while others are declining. The information produced is very useful for business decision making.

Decision can take place on the basis classification of Dead-Stock (DS), Slow- Moving(SM) and Fast-Moving (FM) of the sale. Segment by segment sales forecasting can produce very useful information.

The forecasting can be short, mid and long term. Long term forecasting may not produce precise predictions.

However, it is very useful in understanding market trends. It is easy to turn cash into stock, but the challenge is to turn stock into cash.

Effective inventory management enables an organization to meet or exceed customer's expectations of product availability while maximizing net profits and minimizing costs. Only through data

mining techniques, it is feasible to extract useful pattern and association from the stock data. Data mining techniques like clustering and associations can be used to find meaningful patterns for future predictions.

Clustering is used to generate groups of interrelated patterns, while association provides a way to get generalized rules of dependent variables. Patterns from a huge stock data on the basis of these rules can be obtained. The behaviour in terms of sales transaction is significant. The general term used for such type of analysis is called Market Basket Analysis. Typically there are various items placed in a market for selling, in which some of the product will have a fast selling rate, some will have a slow selling rate and some will be dead stock i.e. rarely selling items.

We consider a scenario of super store or supermarket in distributed environment, or internet based data processing environment. Decision making in business sector is considered as one of the crucial tasks. There is study for data mining for inventory item selection with cross selling considerations which is used for maximal-profit selling items. But our problem is finding out the selling power of the products in the market. (Vivek Ware and Bharathi H. N, 2014).

This is a useful approach to distinguish the selling frequency of items on the basis of the known attributes, e.g. we can examine that a "black coat of imperial company in winter season has high trade", and here are a few qualities related to this example, i.e. colour, type, company, season, and location.

Similarly we can predict that certain products of certain properties have what type of sale trends in different locations. Thus on the basis of this scenario

*Corresponding author: Fiyansh Shah

we can predict the reason of dead-stock, slow moving and fast moving items. Data mining techniques are best suited for the analysis of such type of classification, useful patterns extraction and predictions.

2. Data Mining

Data mining techniques provide people with new power to research and manipulate the existing large amount of data. Data mining process recognizes interesting information from the hidden data which can either be used for future prediction and/or for intelligently summarizing the details of the data. There are so many achievements of applying data mining techniques to various areas such as marketing, medical, financial and manufacturing. (Jiawan Han and Micheline Kamber, 2004).

In this paper, a proposed data mining application in the manufacturing domain is explained. The application demonstrate the capability of data mining techniques in providing important analysis such as launch analysis and slow turning analysis. Such analysis help in providing manufacturing market with base for more accurate prediction of future market demand.

Data mining technology is one of the strong pillars in customer relationship management (CRM) and plays a vital role in business expansion.

By knowing customer behavior based on previous records, growing profits from cross-selling and other business strategies can be achieved. The behavior in terms of sales transactions is considered significant. Data mining on such transactions is called market basket analysis.

We consider the scenario of a manufacturing firm or a supermarket. Typically there are a lot of variations in the items offered, and the volume of transactions can be very large. For instance, Hedberg (1995) reports that the American supermarket chain Wal-Mart keeps about 20 million sales transactions per day. This enormous volume of data requires sophisticated methods in the analysis.

Effective Decision making in the business sector is considered one of the critical tasks in data mining.

3. Proposed Method

Our proposed approach is a two phased model. Firstly we make use of Flat Files and Online Transaction Processing to obtain customer data and then we combine it into Data Sets. Once we have the combined the data sets we apply the K-Means algorithm to divide the data sets into clusters. These clusters are formed on the basis of the attributes of a particular product.

After forming these clusters we need to classify the products as slow moving (SM), fast moving (FM) and dead stock (DS). Slow moving stock is considered to be a category of those products which have a slow selling rate. Fast moving stock is the category of products which have the highest sale or the highest demand

amongst consumers. Dead stock is the category of products that is completely dormant.

After the products have been categorized into slow moving, fast moving and dead stock we apply the Most Frequent Pattern Algorithm to extract frequent patterns. Once the patterns have been extracted we store it into Knowledge Discovery Database.

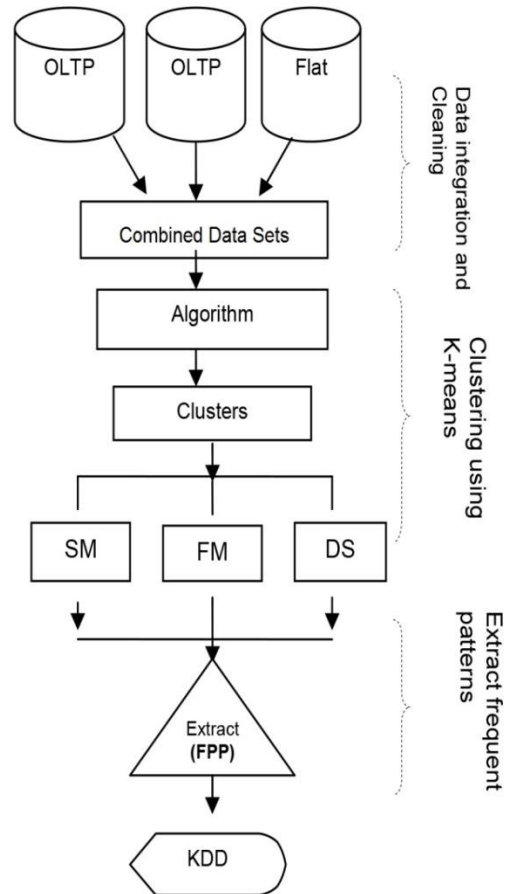


Fig.1 Block Diagram of Proposed Method

4. K-MEANS

Clustering is the process of splitting an assemblage of data points into a lesser number of clusters. For instance, the items in a supermarket are clustered in categories (butter, cheese and milk are grouped in dairy products). In our case, number of products of particular attribute sold will be partitioned in three clusters i.e slow moving, fast moving and dead stock. In general, we have n data points $x_i, i=1...n$ that have to be partitioned in k clusters. The goal is to assign a cluster to each data point. K-means is a clustering method that aims to find the positions $\mu_i, i=1...k$ of the clusters that minimize the distance from the data points to the cluster.

K-means clustering solves

$$\arg \min \sum_{i=1}^k \sum_{x \in c_i} d(x, \mu_i) = \arg \min \sum_{i=1}^k \sum_{x \in c_i} \|x - \mu_i\|^2$$

Where c_i is the set of points that belong to cluster i . The K-means clustering uses the square of the Euclidean distance $d(x, \mu_i) = \|x - \mu_i\|^2$.

This problem is not trivial (in fact it is NP-hard), so the K-means algorithm only hopes to find the global minimum, possibly getting stuck in a different solution. K-means algorithm is used to solve the k-means clustering problem and works as follows,

First, decide the number of clusters k . Then:

1. Initialize the center of the clusters.
2. Attribute the closest cluster to each data point.
3. Set the position of each cluster to the mean of all data points belonging to that cluster.
4. Repeat steps 2-3 until convergence.

5. Most Frequent Pattern

Association rule learning is a popular and well researched method for discovering interesting relations between variables in large databases. It is suggested to identify strong rules discovered in databases using different measures of interestingness. It is widely used in various areas such as risk management, telecomm, market analysis, inventory control, and stock data. In data mining, association rules are useful for analyzing and predicting customer behavior. They play an important part in shopping basket data analysis, product clustering, and catalog design and store layout. For strong association among the patterns Apriori algorithm is highly recommended. In this work a new algorithm MFP, which more efficiently generates strong associations and frequent patterns in the clusters is used.

For this reason a property matrix comprising calculated values of analogous properties of each product has been used as shown in Figure 2.

Let's have a set X of N items in a dataset having set Y of attributes. This algorithm counts maximum of each attribute values for each item in the dataset.

```

MFP Algorithm: Let we have set  $X$  of  $N$  items in a Dataset having set  $Y$  of attributes. This algorithm counts maximum of each attribute values  $y_{ij}$  for each item in the dataset.

Input: Datasets (DS) Output: Matrix
Frequent Property Pattern (FPP):

FPP (DS)
Begin
    for each item  $X_i$  in DS
        a. for each attribute
            i. count occurrences for  $X_i$ 
               C=Count ( $X_i$ )
            ii. Find attribute name of C
                Mi=Attribute (Ci)
        next [End of inner loop]
    b. Find Most Frequent Pattern
        i. MFP=Combine(Mi)
    next [End of outer loop]
End
    
```

Fig.2 Proposed Algorithm for Frequent Pattern Extraction

6. Future Work

The proposed hybrid classification algorithm will prove its mettle in managing the inventory and to provide a necessary support to the business. After analyzing the customer purchase data of the market, we can move to analyzing the sentiments of the customer in the future. This sentiment analysis will be done by getting the customer reviews and web posts from the internet about the products sold online. This will help in understanding the consumers effectively.

Conclusion

In this paper, we have addressed the problem of managing the inventory and stock data. We have proposed the hybrid classification algorithm to narrow down the huge amount of data into more manageable set of data by identifying frequent patterns in the sales of the products. The K-means and MFP algorithm together are going to prove efficient in managing the data and increase the productivity of the company.

Acknowledgement

We would like to thank our project guide Professor Sonal Maskeen for her enormous co-operation and guidance. All the inputs given by her have found a place in the project. The technical guidance provided by her was more than useful. We are also grateful to our head of the department, Professor Mahendra Patil who has guided us. Lastly we would not be able to conduct this research without the infrastructure and facilities provided by Atharva College of Engineering.

References

<http://www.roselladb.com/sales-trendforecast.html>
 Vivek Ware and Bharathi H. N (2014), Decision Support System for Inventory Management using Data Mining Techniques, International Journal of Engineering and Advanced Technology
 Jiawan Han and Micheline Kamber (2004), Data Mining Concepts and Techniques, *The Morgan Kaufmann Series*.
 Darken, C. Moody (2002), Fast adaptive k-means clustering, *Institute of Electrical and Electronics Engineers journal*.
<http://www.onmyphd.com/?p=k-means.clustering>
http://en.wikipedia.org/wiki/cluster_analysis
 Brijs, Bart, Gilbert, Koen, Geert (2000), A Data Mining Framework for Optimal Product Selection in Retail Supermarket Data: The Generalized PROFSET Model, *KDD '00 Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp 300-304.
 M. Al-Noukari, and W. Al-Hussan (2008), Using Data Mining Techniques for Predicting Future Car market Demand, *IEEE*.