Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 4, Issue. 4, April 2015, pg.81 – 87

RESEARCH ARTICLE

Use of Data Mining Techniques to Improve the Effectiveness of Sales and Marketing

Aditya Joshi, Nidhi Pandey, (Professor) Rashmi Chawla, Pratik Patil

Department of Information Technology, Rajiv Gandhi Institute of Technology, Mumbai, India
aditya11.joshi@gmail.com, nidhipandey105@gmail.com, rashmi.chawla.be@gmail.com, pratikpatil786.pp@gmail.com

Abstract: Classification and patterns extraction from customer data is very important factors for business support and decision making. Identification of newly emerging trends is needed in business process. Sales patterns from inventory data indicate market trends and can be used in forecasting which has great potential for decision making, strategic planning and market competition. The system consists of two phases. In the first phase, we divide the stock data in three different clusters on the basis of product categories and sold quantities i.e. Dead-Stock (DS), Slow-Moving (SM) and Fast- Moving (FM) using K-means algorithm. In the second phase we will use Most Frequent Pattern (MFP) algorithm to find frequent patterns of item attributes in each category of products and also give sales trend in a compact form.

Keywords: Stock data, K-Means, Clustering, Most Frequent Pattern, Data Mining

I. INTRODUCTION

Data is very important for every organization and business. Data that was measured in gigabytes until recently, 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, useful information is often overlooked, and the potential 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 term, midterm and long term. Long term forecasting may not produce accurate predictions. However it is very useful in understanding market trends. It is easy to turn cash into inventory, but the challenge is to turn inventory into cash. Only through data mining techniques, it is possible 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 related 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. 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 ratio of sale", here we have basic property 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 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.

II. RELATED WORK

In recent years, it has been recognized that the partitioned clustering technique is well suited for clustering a large dataset due to their relatively low computational requirements. The time complexity of the partitioning technique is almost linear, which makes it widely used. The best known partitioning clustering algorithm is the K-means algorithm and its variants. This algorithm is simple, straightforward and is based on the firm foundation of analysis of variances. In addition to the K-means algorithm, several algorithms, such as Particle Swarm Optimization (PSO) is another computational intelligence method that has already been applied to image clustering and other low dimensional datasets.

Data mining techniques and clustering techniques have a greater scope in creating software which will not only enable companies and organizations in decision making, but also help in following areas:

- Sales forecasting.
- Identifying new and emerging trends.
- Maintaining accounts and inventory.
- Prospecting the potential customers by reports generated.

III. METHODOLOGY

Proposed System:

In this system we implement an algorithm for mining patterns of huge stock data to predict factors affecting the sale of products. In the first phase, we divide the stock data in three different clusters on the basis of sold quantities i.e. Dead-Stock (DS), Slow-Moving (SM) and Fast-Moving (FM) using K-means algorithm. In the second phase we proposed Most Frequent Pattern (MFP) algorithm to find frequencies of property values of the corresponding items. MFP provides frequent patterns of item attributes in each category of products. Cluster analysis is widely used in market research when working with multivariate data from surveys and test panels. Market researchers use cluster analysis to partition the general population of consumers into market segmentation, intra and inter stock patterns and to understand better association between them.

System Architecture:

Our proposed approach is a two phased model. First we generate clusters using K-Mean algorithm, and then MFP is designed for counting frequencies of items under their specified attributes. The block diagram of the whole process is given in figure. In phase-1 the first step is to collect sample data from real store inventory data. We have process the data to remove the noise first, so the incomplete, missing and irrelevant data are removed and formatted according to the required format.

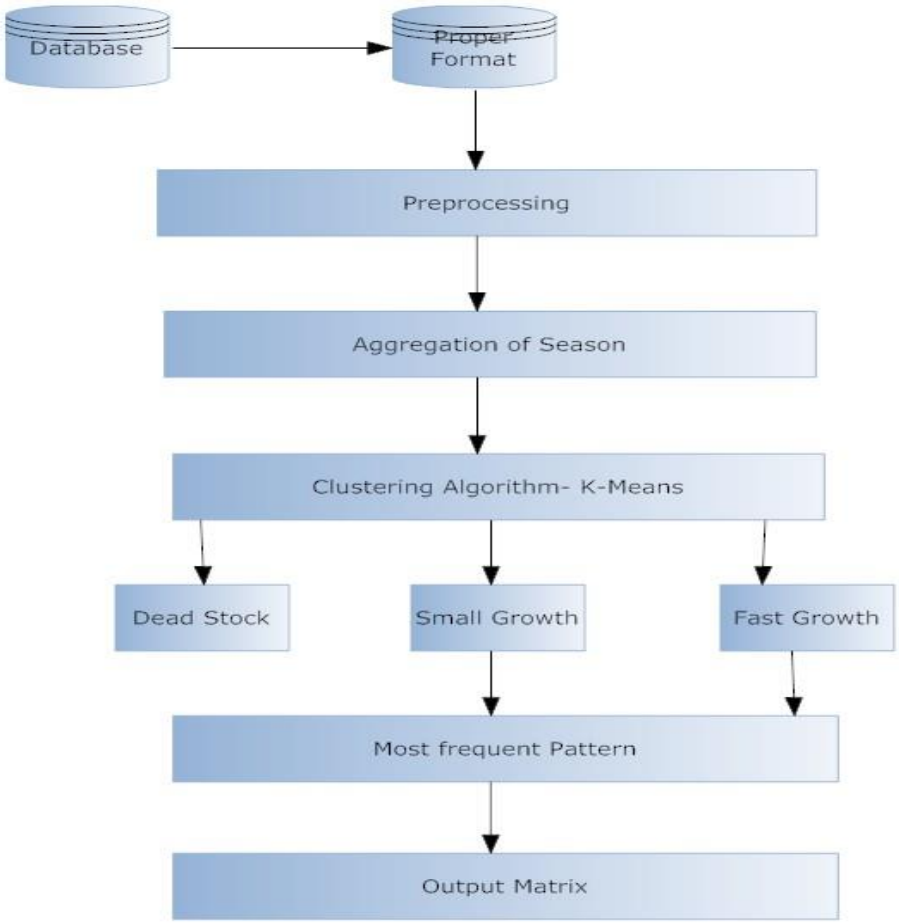


Fig.1: Implementation Diagram of Proposed System.

Our proposed system is divided into 2 distinct phases described as follows:

- The data is collected from the database and pre-processing is done on data. After the pre-processing is done k-mean clustering algorithm is applied to the data. Clusters are formed on bases of dead stock, slow moving stock and fast moving stock.
- Later in 2nd phase we apply the most frequent pattern (MFP) for analyzing the patterns.

K-MEANS CLUSTERING:

The inputs of this algorithm are the number of clusters to be formed i.e., k and the data to be clustered. The algorithm starts with an initial set of cluster centres, chosen at random or according to some heuristic procedure. In each iteration, each instance is assigned to its nearest cluster centre according to the Euclidean distance between the two. Then the

cluster centres are re-calculated. The centre of each cluster is calculated as the mean of all the instances belonging to that cluster. This algorithm aims at minimizing an objective function, in this case a squared error function. The objective function is:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2$$

Where $\left\| x_i^{(j)} - c_j \right\|^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster c_j centre, is an indicator of the distance of the *n* data points from their respective cluster centres.

Most Frequent Pattern [MFP] ALGORITHM:

Association rule mining is one of the most important and well defines technique for extract correlations, frequent patterns, associations or causal structures among sets of items in the transaction databases or other repositories. Association rules are widely used in various areas such as risk management, telecomm, market analysis, inventory control, and stock data. Apriori algorithm for strong association among the patterns is highly recommended. In this work we proposed a new algorithm MFP that is more efficiently generates frequent patterns and strong association between them. For this purpose a property matrix containing counted values of corresponding properties of each product has been used as shown below.

Let we have 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.

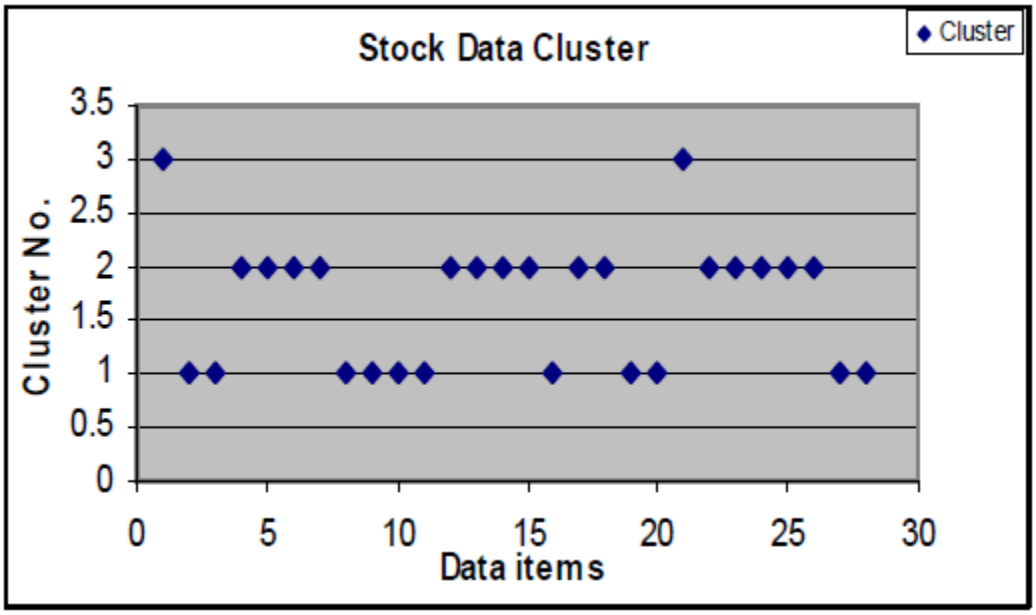
```
Input: Datasets (DS)
Output: Matrix
Frequent Property Pattern (FPP):
FPP (DS)
Begin
        for each item Xi in DS

            a.  for each attribute
                    i.  count occurrences
                        for Xi
                      C=Count (Xi)
                    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
```


IV. EXPERIMENTATION AND RESULTS

As discussed earlier, first we have transformed inventory data in required format by removing noise and any other inconsistencies which is then used for clustering. Clusters are formed from data on basis of the quantity sold. Clustering is done by means of K-Means Clustering Algorithm. The results obtained from both phases are as follows:

Phase One: In first phase, data is divided into three clusters by K-Means Clustering Algorithm. The Clusters obtained are Dead Stock, Slow Moving and Fast Moving. The Dead Stock cluster contains records of those products with small selling quantity. Slow Moving stock cluster contains records of products with medium sales. Fast Moving Stock Cluster contains records of products with large selling quantity.

Phase Two: In this phase MFP algorithm has been used to generate a property matrix, containing counted values of corresponding properties of each product. This procedure receives data sets from clusters. The first loop scans all the records of the data set. The inner loop counts occurrences of the attribute for a given item and placed in the MFP matrix. Finally maximum occurrences of attributes values within a row give a single pattern. On the basis of these patterns, we can say that why a certain product falls in particular cluster.



Cluster Results

V. CONCLUSION

In this system, the clustering association mining approach is used to classify stock data and find associated patterns of sale. From the experimental results it is clear that the approach is very efficient for mining patterns of huge stock data and predicting the factors affecting the sale of products. We formulate most frequent pattern of products using their known properties in inventory system. We identified the trends of selling products through their known attributes. Our technique is simple by using matrix and counting of attribute values. The limitation of study is, that it requires proper data format with specific attributes. In future we will extend our work to implement in sentiment analysis process and decision making from online customer reviews and blogs data.

REFERENCES

1. Aurangazeb Khan, Khairulllah Khan,behram B.Baharuddin "Mining Frequent Patterns Minning Of Stock Data Using Hybrid Clustering Association Algorithm" International Conference 2009.
2. Rachita Sony Krotha, Satish Muppidi, "Performance Evaluation of K-Means Algorithm and Enhanced Mid-point based K-Means Algorithm on Mining Frequent Patterns", 2013.
3. Neha Aggarwal,Kirti Agarwal " A Mid-point based K-mean clustering algorithm for Data Mining", IJCSE 2012.
4. http://www.roselladb.com/sales-trend- forecast.htm
5. http://en.wikipedia.org/wiki/cluster_analys
6. Matt Hartely "Using Data Mining to predict inventory levels" IEEE, 2005
7. Darken, C. Moody, J. Yale Comput. Sci., New Haven, " Fast adaptive k-means clustering" IEEE- 2002
8. M. Al-Noukari, and W. Al-Hussan, "Using Data Mining Techniques for Predicting Future Car market Demand" IEEE, 2008.
9. Houshmand M., "Improve the classification and sales management of products using multi-relational data mining" IEEE 2011.

