International Research Journal of Multidisciplinary Science & Technology
www.irjmrs.com ISSN: 2455-930X

DATA MINING PROCEDURES FOR DECISION SUPPORT SYSTEM ON INVENTORY MANAGEMENT

MAI Navid and NH Niloy*
Ruhea College, Bangladesh
**Corresponding author's E-mail: niloynh1997@gmail.com*

ABSTRACT - *With the objectives are to get better decision making for improving sale, services and quality, which useful mechanism for business support, investment and surveillance. An approach is implemented for mining patterns of huge stock data to predict factors affecting the sale of products. For this 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 or Hierarchical agglomerative algorithm. Moreover, the Most Frequent Pattern (MFP) algorithm is implemented to find frequencies of property values of the corresponding items. MFP provides frequent patterns of item attributes and also gives sales trend in a compact form. Clustering and MFP algorithm can generate more useful pattern from large stock data which is helpful to get item information for inventory.*

Keywords: Stock data, Most Frequent Patterns, Clustering, Decision Making.

1. INTRODUCTION

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. The behavior in terms of sales transaction is significant. The general term used for such type of analysis is called Market Basket Analysis (Han *et al*., 2004). (Khan *et al*., 2011) (Gandhmal*et al*., 2011). It is easy to turn cash into inventory, but the challenge is to turn inventory 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. 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. Only through data mining techniques, it is possible to extract useful pattern and association from the stock data (Kim et al., 2006). Decision making in business sector is considered as one of the critical tasks. There is study for data mining for inventory item selection with cross selling considerations which is used for maximal-profit selling items (Khan *et al*., 2011). Typically there is lot of different items, placed in a market for selling, in which some of the product will be fast selling items, some will be slow selling items and some will be dead stock.

However the problem is finding out the selling power of the products in the market. 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 "Sinthetic Surat sadi of red color of type nylon in marriage season has high ratio of sale", here we have basic property related to this example, i.e. color, type, season, and Design. Data miningtechniques are best suited for the analysis of such type of classification, useful patterns extraction and predictions (Khan *et al*., 2011). So it can be predict that what products of certain properties have what type of sale trends in different locations. Thus on the basis of this scenario it can predict the reason of dead-stock, slow moving and fast moving items

2. BACKGROUND

In the current years of development in the field of data mining, it is considered that the partitioned clustering technique is well suited for clustering a large document dataset due to their relatively low computational requirements and increase in the gradual performance of the system. The time factor complexity of the partitioning technique is almost linear, because of which it is widely used. The best known partitioning clustering algorithm is the K-means algorithm and its variants (Artigan, 1975). In the field of data mining, the researchers always try to find innovative techniques so as to improve the performance of the extraction methods used in data mining as they usually use history of the different transactions done in finding the data as it will be useful for future use. As 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, other algorithms such as Particle Swarm Optimization (PSO) are another computational intelligence method that has already been applied to image clustering and other low dimensional datasets (Khan *et al*., 2011). This data collection can be used by them to predict the customer behavior and their interests L.K.Soon et al [27], compared

the execution performance of numerical and symbolic representation of using data in term of similar search. M. C. Lo (Han *et al*., 2004) considered a model for inventory decision support system [IDSS] in which ordering quantity, ordering cost, safety factor, lead time and backorder discounts are decision variables, the algorithm is applied to fine the optimal solution for the case where the lead time demands to follow a general distribution. Gandhmal*etal*. (2011) he proposed a technique based trading data mining approach for intra-stock mining whichusually perform concentrates on finding most appearing items for the stock time series data and inter trading mining which used to discover the different strong relationship among the several stocks. Gandhmal*et al*. (2011) generated a list of stocks which are influential to Kuala Lumpur Composite Index (KLCI), and then produce classification rules, which he denotes the inter-relationships among the stocks in terms of their trading performance with respect to KLCI. (Gandhmal*et al*., 2011) (Khan *et al*., 2011).

3. ARCHITECTURE

In this work an algorithm used for mining patterns of huge stock data to predict factors affecting the sale of products. In the first phase, it divides the stock data in three different clusters on the basis of sold quantities i.e. Dead Stock (DS), Small Growth (SG) and Fast-Growth (FG) using K-means algorithm or Hierarchical Agglomerative. In the second phase Most Frequent Pattern (MFP) algorithm is used to find frequencies of property values of the corresponding items (Khan *et al*., 2011).

Step-1: to collect database (cash memo) form storekeeper and put it into proper format (excel sheet). Also collect inventory and put in separate excel sheet.

Step-2: In this data preprocessing is done by filling missing value either by global constant or by average.

Step-3: After that aggregation has to perform. Season-wise aggregation (collection) is performed.

Step-4: clustering algorithms has to be performed. 1. K-mean Algorithm or 2.Hierachical Agglomerative. Any one of them has to be performed.

Step-5: after any one clustering algorithm performed. 3 clusters are generated as small growth cluster, fast growth clusters and dead stock clusters.

Step-6: perform Most Frequent Pattern algorithm on 1. Fast growth cluster and 2. Slow growth clusters. Show the result in matrix format.

4. K-MEANS

K-means (Darken *et al*., 2002) is a typical clustering algorithm and has used for classification of data for decades. Proximity is usually measured by some sort of distance; the most commonly being used is the Euclidean distance (Khan *et al*., 2011) l Dist (i, j) = sqrt (∑ (xit - xjk) ^2) …….. (1) k = 1

The main idea is to define k centroids, one for each cluster. These centroids should be placed in a cunning way because of different location causes different result. This algorithm aims at minimizing an objective function, in this case a squared error function. The objective function is k k J = ∑ ∑ || xj j - cj ||2

…… (2) j = 1 i = 1 Where || xj j - cj ||2 is a chosen distance measure between a data point (j) I x and the cluster centre j. c, is an indicator of the distance of the n data points from their respective cluster centers. The steps of the K-mean algorithm (Han and Kamber, 2004) are as described below: Step 1: Place K points into the space represented by the objects that are being clustered; these points represent initial group centroids. Step 2: Assign each object to the group that has the closest centroid. Step 3: When all objects have been assigned, recalculate the positions of the K centroids. Step 4: Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated. B. Hierarchical Agglomerative Hierarchical clustering (Han *etal*., 2004) is an agglomerative (top down) clustering method. As its name suggests, the idea of thismethod is to build a hierarchy of clusters, showing relations between the individual members and merging clusters of data based on similarity. In the first step of clustering, the algorithm will look for the two most similar data points and merge them to create a new "pseudo-datapoint", which represents the average of the two merged datapoints. Each iterative step takes the next two closest datapoints (or pseduo-datapoints) and merges them. This process is generally continued until there is one large cluster containing all the original datapoints. Hierarchical clustering results in a "tree", showing the relationship of all of the original points. C. Most Frequent Pattern (MFP) 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 (Khan *et al*., 2011). Apriori algorithm (Han *et al*., 2004) for strong association among the patterns is highly recommended. 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.

5. 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 yij 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.

6. IMPLEMENTATION

Data in its original format never confirm to the required shape for data mining. It needs to be transformed, integrated, and aggregated so that the mining process can effectively perform on it. There is a need to process the data before it used in the knowledge discovery (KDD) process. Being data quality a key issue with data mining as 50% to 80% of mining experts often spend their time on data quality, the pre-processing in data mining have a key importance. (Khan et al., 2011) Customer buying details (Cash Memo) are stored in the excel sheet as shown in the figure. Here filed taken as Product Id (PID), Product Name (PNAME), Product Color (Color), Type of design product has (Design), Product Prize Range (Prize Range), How many products are sold on that day (Volume), total bill number of a day (Bill numbers), Total cash of day (Cash Total), date, Month and season. Where 6 Season are considered as summer, winter, Mansoon, Marriage, Cell, and Festival. An Inventory is also recorded when stock is came to shop for future use.

7. DATA CLEANING AND AGGREGATION:

Missing values can be replaced by either average value or global constant as per preprocessing techniques. Select appropriate attribute on which preprocessing going has to done.

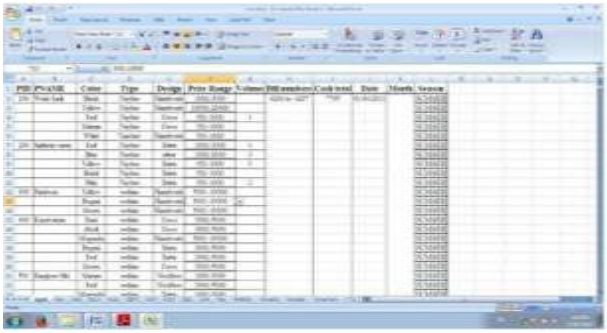

Figure 1.Snapshot of Proper Data in Excel Sheet

Figure 2. Snapshot of Items in Fast Growing Cluster

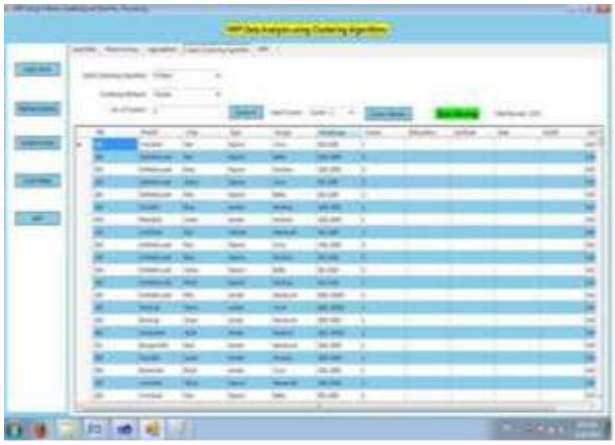

Figure 3.Snapshot of Items in Slow Growing Cluster

Figure 4. Snapshot of Items in Dead Stock Cluster

Figure 5.Snapshot of MFP on Fast Growing Cluster

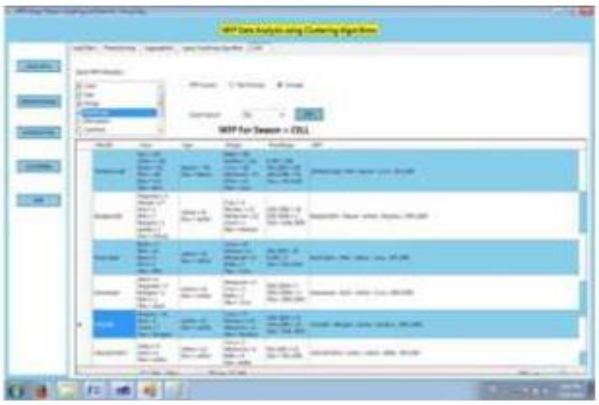

Figure 6. Snapshot of MFP on Small Growing Cluster

Here volume attribute is selected (figure 3) and missing values are replaced by global constant zero. Aggregation will here collect data by season wise attribute. Select the attribute on which aggregation has to do. Now form this we can tell for example consider the summer season in that Sinthetic Surat sadi of type Naylon with color red and having cross design is frequently taken by customer so u can purchase this type in summer so that storekeeper can get maximum profit.

8. RESULTS

Table 1 Number of Items Present in the Cluster Table 1 will show records in the clusters. Reason behind this is K-mean initially itself choose 3 central points to create cluster. Hierarchical Agglomerative choose cluster on basis of similarity. So we can't predict number of records in the any cluster in case of Hierarchical Agglomerative. Meaning of dead stock clusters is that not sold items, Fast moving clusters frequently large number going items and slow moving clusters are item which are sold more but not large in number. Now these clusters will help to shop owner get knowledge about items. Now storekeeper can make idea like using cell to sold dead stock or slow moving items. On cluster generation it gets item selling status. Now it will find pattern. Most frequent pattern will help now. It will tell which attribute of item has more frequency in selling. This gives more information about that item. Using present inventory in shop, owner can decide about inventory.

Table 1. Number of Items Present in the Cluster

Algorithm	Total Records	Fast Moving	Slow Moving	Dead Stock
K-Mean	23856	51	2170	21635
Hierarchical Agglomerative	23856	46	2175	21635

Table 2 is showing pattern of fast moving items using K-mean algorithm. Now form this we can tell for example consider the summer season in that Sinthetic Surat sadi of type Naylon with color red and having cross design and prize range 501-1000 is frequently taken by customer so he can purchase this type in summer so that storekeeper can get maximum profit.

Table 2. MFP on Fast Growing Item using K-mean

Season	Product Name	Color	Type	Design	Prize Range
Winter	-	-	-	-	-
Summer	SintheticSurat	Red	Naylon	Cross	501-1000
Mansoon	SintheticSurat	Red	Naylon	Cross	0-500
Festival	SintheticSurat	Red	Naylon	Handwork	501-1000
Cell	SintheticSurat	Red	Naylon	Neckless	0-500
Mariage	SintheticSurat	Red	Naylon	Batta	0-500

9. CONCLUSION

Hierarchical (Hierarchical Agglomerative) and Partitional (K-Mean) Clustering have key differences in running time, assumptions, input parameters and resultant clusters. Typically, partitional clustering is faster than hierarchical clustering. Hierarchical clustering requires only a similarity measure, while partitional clustering requires stronger assumptions such as number of clusters and the initial centers. Hierarchical clustering does not require any input parameters, while partitional clustering algorithms require the number of clusters to start running. Hierarchical clustering returns a much more meaningful and subjective division of clusters but partitioned clustering results in exactly k clusters. The hierarchical clustering method, though simple, often encounters difficulties regarding the selection of merge or split points. Such a decision is critical because once a group of objects is merged or split, the process at the next step will operate on the newly generated clusters. Moreover, the method does not scale well, because each decision to merge or split requires the examination and evaluation of a good number of objects or clusters. Both of these choose their initial points randomly but in case of hierarchical it is seen that he decide the points itself because splitting and merging is dynamic at that points. So while implementing various result got when hierarchical agglomerative algorithms is used. Sometimes it gives 46 items in fastgrowth, sometimes 78 items in fast growth. Sometimes it gives same result as such given by K-mean. So according to items present in the cluster pattern are also

changed. The new algorithm MFP that is more efficiently generates frequent patterns and strong association between them. It does just calculate the frequency count which is easy to understand than apriori algorithm. The effectiveness of this procedure could be improved further on larger database

10. REFERENCES

[1] Khan,A., Baharudin,B., Khan, K.A. 2011. "Mining Customer Data for Decision Making using new Hybrid Classification algorithm" in journal of theoretical and applied Information Technology Vol 27 no.1.

[2] Gandhmal, D., Parihar, R. and Argiddi, R. 2011. "An Optimized Approach to Analyze Stock market using Data Mining Technique" in International Conference on Emerging Technology Trends(ICETT).

[3] Hilage, T. and Kulkarni, R. V. 2012. "Review of Literature on Data Mining" IJRRAS 10 (1) ,.

[4] Apte, C., Liu, B., and Pednault, E. P.D. and Smyth, P. 2002. "Business Application of Data Mining " ,Communication of the ACM August 2002/Vol. 45, No. 8

[5] Kim, S. J., Yun, D. S. and Chang, B.S. 2006. "Association Analysis of Customer Services from the Enterprise Customer Management System" ,ICDM-2006.

[6] Han,J., and Kamber, M. 2004. "Data Mining Concepts and Techniques" 2nd edition 2004

[7] Padhy,N., Mishra, P. and Panigrahi, R. 2012. "The Survey of Data Mining Applications And Feature Scope", International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.3.

[8] Soon, L.K. and Lee,S.H. 2007. "Explorative Data Mining on Stock Data Experimental Results and Findings", pringer- ADMA 2007, LNAI 4632, pp. 562–569.

[9] Moody,D.C., Yale, J., 2002. Comput. Sci., New Haven, "Fast adaptive k-means clustering" IEEE-2002.

[10] Berry and Linoff,. 1997. "data mining techniques: for marketing, sales and customer support", John Eilry #Sons, inc.

[11] Fayyad,U., Piatetsky-Shapiro, G. and Smyth,P. 1996. "From Data Mining to Knowledge Discovery in Databases", AI Magazine Volume 17 , Number 3.

[12] Gupta, S., Kumar, D. and Sharma, A. 2011. "Performance Analysis of Various Data Mining Classification Techniques on Healthcare Data", International Journal of Computer Science & Information Technology (IJCSIT) Vol 3, No 4.

[13] Liao, S.H., Ho, H.H. and Lin, H. 2008. "Mining stock category, association and cluster on Taiwan stock market", Expert Systems with Applications Volume 35 , Issue 1-2.

[14] Thomas, P. Macredie. 1999. "Knowledge Discovery and Data Mining".

[15] Artigan, J. A. 1975. Clustering Algorithms. Ohn Wiley and Sons, Inc., New York, NY.

[16] Noukari, M.A. and Hussan, W.A. 2008. "Using Data Mining Techniques for Predicting Future Car market Demand" IEEE.