Developing a Forecasting Model for Retailers Based on Customer Segmentation using Data Mining Techniques

Kayalvizhi Subramanian[1], Gunasekar Thangarasu[2]

[1]Faculty of Built Environment, Linton University College, Negeri Sembilan, Malaysia
[2]Department of Computer and Information Sciences, Universiti Teknologi Petronas, Perak, Malaysia

ABSTRACT

The purpose of this paper is to develop a forecasting model for retailers based on customer segmentation, to improve the performance of inventory. The research makes an attempt to capture the knowledge of segmenting the customers based on various attributes as an input to the demand forecasting in a retail store. The paper suggests a data mining model which has been used for forecasting demand. The proposed model has been applied for forecasting for grocery items in a supermarket. Based on the proposed forecasting model, the inventory performance has been studied by simulation. Hence, the proposed model in the paper results in improved performance of inventory. Retailers can make use of the proposed model for demand forecasting of various items to improve the inventory performance and profitability of operations. With the advent of data mining systems which have given rise to the use of business intelligence in various domains.

Keywords: *Forecasting, Data mining, Artificial Intelligence, Supermarkets, Inventory*

1. INTRODUCTION

Supply chain management systems and intelligent systems for forecasting grew significantly during the last two decades. However, the growth of these two has mostly taken place independently. On one hand, we have very sophisticated supply chain management systems and on the other, we have very sophisticated forecasting systems. However, we rarely come across the combination of two sophisticated systems. At the same time, retailing has gone through a period of unprecedented change as customers' demands and competition amongst the retailers have intensified in the last 25 years in most countries. Over this period, the retail industry has seen a transition from manual merchandise control systems of the computerized systems. Retailers with the sophisticated computerized systems for better forecasting and improved inventory management have an edge over the others in terms of profitability. Initially, these sophisticated systems were being used only by the supermarkets. Gradually, other retailers found it necessary to remain competitive in their businesses. India is also not an exception to this movement and it has witnessed a sea change in retail business in the last ten years. In a typical retail outlet of grocery items, number of stock keeping units is in the range of a few thousand and in a large supermarket. Retailers buy these items from a large number of distributors and sometimes directly from manufacturers. For each item, inventory managers are to decide when to purchase, how much to purchase and from whom to purchase. Efficient forecasting for future demand is the key to success for inventory management. Future demand of an item depends on a large number of factors and it has been a challenging task for the retailers to predict the future demand. The current research work suggests a data mining-based business intelligence model for demand forecast and its application in enhancing supply chain performance in an Indian retail outlet. The model suggests the use of clustering-based segmentation of the customers as an input to forecasting. Based on customers' demographical profiles and other details, segmentation of customers is done in clusters using data mining software.

2. Literature Review

Improved demand forecasting accuracy can result in monetary savings, greater competitiveness, enhanced channel relationships, and customer satisfaction [1]. The importance of accurate sales forecasts to efficient inventory management has long been recognized. Barksdale and Hilliard [2] found that successful inventory management depends to a large extent on the accurate forecasting of retail sales. Thall [3] and Agrawal and Schorling [4] also pointed out that Accurate demand forecasting plays a critical role in profitable retail operations and poor forecasting results in under stock or overstock that directly affects profitability and competitive position of the retailer.

The most common practice of forecasting demand in supply chain planning involves the use of a statistical software system which incorporates a simple univariate forecasting method, such as exponential smoothing, to produce an initial forecast [5]. The common practices and various literatures include time series decomposition, exponential smoothing, time series regression and autoregressive and integrated moving average (ARIMA) models. Out of these models, seasonal ARIMA model has been the most frequently employed forecasting model that results in a reasonably acceptable accuracy and it has been successfully tested in many practical applications [6 and 7].

The traditional methods which include ARIMA also are linear methods as they assume a linear relationship between independent and dependent variables. This problem is overcome by nonlinear methods. Artificial neural network (ANN) model is one popular nonlinear model which is extensively used for forecasting demand. ANN models use multi-layer perceptron (MLP) which does not need to assume stationary data in time series. Neural network-based fuzzy time series models have been used to improve forecasting for the stock index in Taiwan [8, 9 and 10].

Various data mining applications for inventory management have been suggested in various works. A method to select inventory items from the association rules has been proposed for cross-selling consideration [11]. This gives methodology to choose a subset of items which can give the maximal profit with the consideration of cross- selling effect. Relevance of association rule mining in the context of multi-item inventory replenishment has been

discussed [12]. The paper shows with a case that inventory costs can be reduced with the implementation of data mining-based replenishment policy. A decision tree-based model for inventory replenishment in retail stores has been proposed. The decision tree is inducted using data mining on sale transaction data of purchase items with the demographic profile and other details of the customers. However, these models do not talk about forecasting and relevance of accuracy in forecasting in the performance of inventory in supply chain management. A decision tree-based application in retail sales for investigating the impact of promotion has been used in retail sales. Chang [13] suggest decision tree-based classification to analyze the customers' behavior in order to form the right customers' profile and business growth model under internet and e-commerce environment. In retail sales, the decision tree is induced for various uses in customer relationship management.

Feature selection is another data mining technique which has been widely used for learning customers' purchase behavior. Feature selection is the technique used in data mining for identification of the fields which are the best for prediction [14] as a critical process. This step helps with both data cleansing and data reduction, by including the important features and excluding the redundant, noisy and less informative ones [15]. There are two main stages to feature selection. The first is a search strategy for the identification of the feature subsets and the second is an evaluation method for testing their integrity, based on some criteria. Classification of customers with predefined groups or classes is done after finding the best features (or, attributes) using feature selection. Classification of customers after performing feature selection has several advantages. According to John et al [16] first, there might be a significant improvement over the performance of a classification by reducing the number of bands to a smaller set of informative features. Second, with a smaller number of bands, the processing time is greatly reduced. Third, in certain cases lower- dimensional datasets would be more appropriate, where a limited amount of training data is available.

In view of the several applications of data mining in retailing, it motivates towards devising models for demand forecasting in retail sales with the potential of data mining demonstrated in other areas.

3. Developing a Forecasting Model

3.1 Segmenting Customers using Clustering

Managing customers as an asset requires measuring them and treating them according to their true value. Customer segmentation is generally done by using the clustering techniques of data mining. Different customers' attributes have been used for segmentation by clustering technique in retailing. These attributes include customers' demographic profile and patterns in shopping behavior, like the frequency of shopping, the monetary value of purchase, number of items purchased, etc. In retailing, segmentation of customers is done usually for customer relationship management. Many authors demonstrate the advantages of applying a K-means clustering technique to analyze a time series for determining structural changes low to high or, high to low in the Taiwan Stock Exchange Capitalization Weighted Stock Index.

.

The main trends for developing models to segment customers include K-means, Kohonen map, two-step models. K-means clustering technique does the clustering in K groups where the similarity amongst the entities within a cluster is very high and a reasonably high inter-cluster distance is maintained at the same time. The two-step node uses two steps for clustering. The first step makes a single pass through the data to compress the raw input data into a manageable set of sub-clusters. The second step uses a hierarchical clustering method to progressively merge the sub- clusters into larger and larger clusters. Two-step has the advantage of automatically estimating the optimal number of clusters for the training data. It can handle mixed field types and large data sets efficiently. Kohonen map is a neural network-based clustering, where the clusters are represented by the nodes in two- dimensional coordinate grids. Sometimes three- dimensional and one-dimensional grids are also used.

3.2 Developing a forecasting model based on Clustering

Based on the discussion in the above sections, a forecasting model has been developed for retail merchandise. As a prerequisite to the segmentation of customers, the details of the customers are recorded along with the amount of purchase. The methodology involved in the proposed forecasting model can be described in the following steps:

Step 1 : Exhaustive list of demographic details of the customers and other details depicting the purchase behavior is prepared. These details are used as attributes to describe customers. It is to be noted that all these attributes are not equally important in describing an intended behavior of customers.

Step 2 : Construction of classes of customers is done for the item/SKU considered for demand forecasting. For the purpose of demand forecasting, classes are to be described based on the units of purchase for the SKU. For example, two classes of customers may be – those who purchase one unit and those who purchase two units or, more.

Step 3 : Based on the target classes, feature selection is performed on the database to select top few dominant attributes for the purpose of classification.

Step 4 : Based on the important attributes obtained using feature selection, clustering is performed for the customers.

Step 5 : The original database is segregated based on the segmentation described by the clusters obtained. Each cluster is to be treated as a separate database representing a unique segment.

Step 6 : For each segment of customers, seasonal ARIMA with predictors are used for forecasting. Predictors for daily and weekly forecasting have been identified separately.

Step 7 : To forecast the overall demand of the item, forecasts for various segments are summed up

3.3 Inventory control with the forecast models

For a periodic review policy of inventory replenishment, four proposed forecasting models outperforming other models and the existing forecasting model have been compared with respect to two performance indicators:

1. Inventory level given by reaching days of inventory (measured as inventory/daily sales average); and
2. Customer service (indicated by percentage of days with sales failure).

In fact, customer service level is inversely proportional to "percentage of days with sales failure". In periodic review policy using the daily forecasting, review of the inventory levels is done for the products every P day and the purchase order has to be sent at least L days (known as "lead time") before

the delivery date. The desired inventory level (T) has to be determined every period by the equation, $T = m_0 + Zs$, where m_0 is the average demand during (P+L) days, Z is obtained from a standard normal distribution table which depends on the desired service level and s is standard deviation of the demand during (P+L) days, (P+L) is known as protection period and Zs is the safety stock.

In both daily and weekly forecasting, following data has been used:
P= 7 days
L=3 days
Service level = 90 percent,

The computation for daily forecasting has been shown below:
Standard deviation of demand during (P+L) days
= Standard deviation of daily forecast demand
$\sqrt{}$
Safety Stock = Z X [Standard deviation of demand during (P+L) Days]
Target level =T
= (Average daily forecast demand) X (P +L) + Safety stock
= m_0 X (P+L) + Safety stock
Order Quantity, Q =T – Inventory level

For use weekly forecasting, as the values of P and L remain same given in days, it is required to convert the weekly figures into daily figures. The computation for weekly forecasting has been shown below:

Average daily demand
= Average weekly forecast demand / 7
Standard deviation of daily demand =
$$\overline{} \over \sqrt{}$$
Standard deviation of demand during (P+L) days
= Standard deviation of daily demand $\sqrt{}$
Safety Stock = Z X [standard deviation of demand during (P+L) Days]
Target Inventory level =T
= Average daily demand X (P+L) + Safety stock
Order Quantity, Q=T –Inventory Level

Table 1: Retail forecasting results

	Inventory performance indicator		
	Daily sale	Sales failure days	Sales failure days (%)
Findings	34.09	108	29.22

4. Conclusion
The proposed business intelligence system for demand forecasting proves to give more accurate predictions for future demands compared to the existing models and practices in the supermarket. This helps inventory managers to better manage their supply chain performance by reducing reaching days and service level simultaneously. Reaching day as a measure of inventory level is generally reduced successfully by the retailers at the cost of service level in most of the places. In the present day order, where, most of the large supermarkets are taking resort to data mining for various uses in customer relationship management, enhancing inventory management with the use of clustering of records after identifying the important features will be an additional application for increasing profitability of operations.

5. References
1. Moon, M. A., Mentzer, J. T. and Smith, C. D. (2003), "Conducting a sales forecasting audit", International Journal of Forecasting, Vol. 19 No.1, pp. 5-25.

2. Barksdale, H. C. and Hilliard, J. E. (1975), "A cross- spectral analysis of retail inventories and sales", Journal of Business , Vol. 48 No. 3, pp. 365-82.

3. Thall, N. (1992), "Neural forecasts: a retail sales booster", Discount Merchandiser, Vol. 23 No. 10, pp. 41-2.

4. Agrawal, D. and Schorling, C. (1996), "Market share forecasting: an empirical comparison of artificial neural networks and multinomial logit model", Journal of Retailing, Vol. 72 No. 4, pp.383-407.

5. Fildes, R., Goodwin, P., Lawrence, M. and Nikolopoulos, K. (2009), "Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning", International Journal of Forecasting, Vol. 25 No. 1, pp. 3-23.

6. H. J. Zimmermann, "Fuzzy Sets, Decision Making and Expert Systems, Boston: Kluwer; 1987.

7. Bowerman, B. L. and O'Connell, R. T. (1993),Approach, 3rd ed., Duxbury Press, Belmont, CA..

8. Huarng, K.-H., Yu, T. H. K. and Kao, T.-T. (2008), "Analyzing structural changes using clustering techniques", International Journal of

Innovative Computing Information and Control, Vol. 4 No. 5, pp. 1195-201.

9. Yu, T. H. K. and Huarng, K.-H. (2010), "A neural network-based fuzzy time series model to improve forecasting", Expert Systems with Applications, Vol. 37 No. 4, pp. 3366-72.

10. Yu, T. H. K. and Huarng, K.-H. (2008), "A bivariate fuzzy time series model to forecast the TAIEX", Expert Systems with Applications, Vol. 34 No. 4, pp. 2945-52.

11. Wong, W., Fu, A. W. and Wang, K. (2005), "Data mining for inventory item selection with cross-selling considerations", Data mining and Knowledge Discovery, Vol. 11 No. 1, pp. 81-112.

12. Bala, P. K., Sural, S. and Banerjee, R. N. (2010), "Association rule for purchase dependence in multi-item inventory", Production Planning & Control, Vol. 21 No. 3, pp. 274-85.

13. Chang, C. W., Lin, C. T. and Wang, L. Q (2009), "Mining the text information to optimizing the customer relationship management", Expert System with Applications, Vol. 36, Nos. 21, pp. 1433-43..

14. Sun, Z., Bebis, G. and Miller, R. (2004), "Object detection using feature subset selection", Pattern Recognition, Vol. 37, pp. 2165-76.

15. Yan, L., Wolniewicz, R. and Dodier, R. (2004), "Predicting customer behavior in telecommunications", IEEE Intelligent Systems, Vol. 19 No. 2, pp. 50-8.

16. Aburto, L. and Weber, R. (2007), "Improved supply chain management based on hybrid demand forecasts", Applied Soft Computing, Vol. 7 No. 1, pp. 136-44.