Improving inventory performance with clustering based demand forecasts

Pradip Kumar Bala
Xavier Institute of Management, Bhubaneswar, India

Abstract

Purpose – The purpose of this paper is to develop a forecasting model for retailers based on customer segmentation, to improve performance of inventory.

Design/methodology/approach – 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 of demand. The proposed model has been applied for forecasting demands of eight SKUs for grocery items in a supermarket. Based on the proposed forecasting model, the inventory performance has been studied with simulation.

Findings – The proposed forecasting model with the inventory replenishment system results in the reduction of inventory level and increase in customer service level. Hence, the proposed model in the paper results in improved performance of inventory.

Practical implications – Retailers can make use of the proposed model for demand forecasting of various items to improve the inventory performance and profitability of operations.

Originality/value – With the advent of data mining systems which have given rise to the use of business intelligence in various domains, the current paper addresses one of the most pressing issues in retail management, as demand forecasting with minimum error is the key to success in inventory and supply chain management. The proposed forecasting model with the inventory replenishment system results in the reduction of inventory level and increase in customer service level. The proposed model outperforms other widely used existing models.

Keywords Supermarkets, Inventory management, Demand forecasting, Supply chain management, Data mining, Artificial intelligence, Logistics, Operations management

Paper type Research paper

1. Introduction

Supply chain management systems and intelligent systems for forecasting have grown 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 to 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

Journal of Modelling in Management
Vol. 7 No. 1, 2012
pp. 23-37

1746-5664
DOI 10.1108/17465661211208794

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 (SKUs) is in the range of a few thousand and in a large supermarket, it is generally more than 50,000 SKUs. 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, SPSS-Clementine 12.1.

2. Related work

Improved demand forecasting accuracy can result in monetary savings, greater competitiveness, enhanced channel relationships, and customer satisfaction (Moon *et al.*, 2003). The importance of accurate sales forecasts to efficient inventory management has long been recognized. Barksdale and Hilliard (1975) found that successful inventory management depends to a large extent on the accurate forecasting of retail sales. Thall (1992) and Agrawal and Schorling (1996) also pointed out that Accurate demand forecasting plays a critical role in profitable retail operations and poor forecasting results in understock 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 (Fildes *et al.*, 2009). 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 frequently employed forecasting model that results in a reasonably acceptable accuracy and it has been successfully tested in many practical applications. Equivalent ARIMA models have been implemented for the popular Winters additive and multiplicative exponential smoothing models with an improved accuracy (McKenzie, 1984; Bowerman and O'Connell, 1993).

The traditional methods which include ARIMA also are linear methods as they assume 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 (Huarng and Yu, 2006; Yu and Huarng, 2008, 2010; Yu *et al.*, 2009).

ARIMA has certain advantages over ANN models in terms of providing better understanding of the phenomenon. By analyzing the coefficients of the regressors, ARIMA provides the importance of each independent variable in relation to the dependent variable. Motivated by the advantages of both ARIMA and ANN, Aburto and Weber (2007) developed an additive hybrid forecasting model. The paper models the original time series by ARIMA and error associated with this model forms another time series which was modeled by using ANN. The final hybrid forecast is made by adding ARIMA-based forecast and ANN-based forecast. This outperforms pure ARIMA and pure ANN models.

Various data mining applications for inventory management have been suggested in various works. In Wong *et al.* (2005), a method to select inventory items from the association rules has been proposed for cross-selling consideration. 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 in Bala *et al.* (2010). 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 in Bala (2009b). Decision tree is inducted using data mining on sale transaction data of purchased 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 sale for investigating the impact of promotion has been used in retail sale (Bala, 2009a). Chang *et al.* (2009) 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 sale, decision tree is induced for various uses in customer relationship management (CRM).

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, described by Sun *et al.* (2004) 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 (Yan *et al.*, 2004). 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 in 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.* (2007), 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 sale with the potential of data mining demonstrated in other areas.

3. 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 frequency of shopping, monetary value of purchase, number of items purchased, etc. In retailing, segmentation of customers is done usually for CRM. It has been done for designing promotional offers, customer attraction, customer retention and customer development (Ngai *et al.*, 2009). To forecast the monthly tourist arrivals to Taiwan, Huarng *et al.* (2011) has proposed an innovative forecasting model to detect the clusters of regime switching properly with application of fuzzy time series model. Huarng *et al.* (2008) demonstrates 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 co-ordinate grids. Sometimes three-dimensional and one-dimensional grids are also used.

4. Developing a forecasting model based on clustering

Based on the discussion in the above sections, a forecasting model has been developed for retail merchandises. 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 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, ARIMA (or, seasonal ARIMA) with predictors are used for forecasting. Predictors for daily and weekly forecasting have been identified separately and they are listed in Section 5.
- *Step 7.* To forecast the overall demand of the item/SKU, forecasts for various segments are summed up.

5. Application of the forecasting model based on clustering

The proposed model has been applied for forecasting demands of eight SKUs for grocery items in a supermarket located in the eastern part of India. Proposed model based on clustering and ARIMA has been first applied with no consideration of seasonality and then with the consideration of seasonality. Various other forecasting models based on ARIMA and ARIMA with neural network have also been applied. Analysis of demand forecasts for an SKU, 5 kg packet of *Atta* (a popular coarse flour consumed in Indian subcontinent) of Ashirvad brand of ITC Ltd has been given below.

Existing practice of forecasting follows 14-day moving average technique for predicting daily demand of the SKU. The retail store was facing frequent cases of sale failures as well as high average of daily stock. To examine the efficacy of daily forecasting, weekly forecasting has been examined for the SKU. For both daily and weekly forecasting, proposed model and other models have been used for forecasting.

For daily demand forecasting, following models have been used. The codes of the respective models used in the subsequent part of the article are given in the brackets:

- 14-day moving average (DM-0).
- Best ARIMA (DM-1).
- Best seasonal ARIMA (DM-2).
- Best ARIMA with predictors/regressors (DM-3).
- Best seasonal ARIMA with predictors/regressors (DM-4).
- Neural network on the errors of the best ARIMA with predictors/regressors (DM-5): Neural network is used to estimate error in demand forecasting using ARIMA. The same has been done in other neural network models mentioned below here.
- Neural network on the errors of the best seasonal ARIMA with predictors/regressors (DM-6).
- Clustering-based best ARIMA with predictors/regressors (DM-7).
- Clustering-based best seasonal ARIMA with predictors/regressors (DM-8).

Similarly, for weekly demand forecasting, following models have been used. The codes of the respective models used in the subsequent part of the article are given in the brackets:

- Best ARIMA (WM-1).
- Best seasonal ARIMA (WM-2).
- Best ARIMA with predictors/regressors (WM-3).

- Best seasonal ARIMA with predictors/regressors (WM-4).
- Neural network on the errors of the best ARIMA with predictors/regressors (WM-5).
- Neural network on the errors of the best seasonal ARIMA with predictors/regressors (WM-6).
- Clustering-based best ARIMA with predictors/regressors (WM-7).
- Clustering-based best seasonal ARIMA with predictors/regressors (WM-8).

Clustering-based models (DM-7, DM-8, WM-7 and WM-8) are based on the proposed forecasting model. Other models are based on the existing practices and available literature. Models in DM-5, DM-6, WM-5 and WM-6 are based on the work of Aburto and Weber (2007) which has also been discussed in Section 2 of this paper.

A total of 22 attributes for customers were included in the data. Classes were predefined as:

- *Class 1*. people purchasing one unit;
- *Class 2*. people purchasing two units; and
- *Class 3*. people purchasing three or more units.

With respect to these classes, the model of feature selection in the data mining software, SPSS Clementine 12.1, was performed on the transaction data. With the feature selection model using Pearson ratio with p-value of 0.05 for finding important factors, following are the five profile descriptions of the customers qualifying with threshold p-value and these have been considered as most dominant features for the classification defined:

(1) *Gender*. This refers to the gender of the customer with attribute values of "M" for male and "F" for female.

(2) *Income*. This refers to the monthly income (in rupees) of the customer.

(3) *Number of children*. This refers to number of children the customer is having. The variable is set to be an integer.

(4) *Level of education*. This refers to the different types of education of the buyer. It is given the value of "1" for those with education below undergraduate level, "2" for those with graduation in nonprofessional courses and "3" with graduation in professional courses.

(5) *Domicile of the province*. If a customer belongs to the province in which the supermarket is located, the corresponding input is "yes" and if the customer does not belong, the corresponding input is "no".

Based on the above five attributes, clustering was performed using the same data mining software, SPSS-Clementine 12.1. Clustering of sale transaction data was done based on the total of six inputs – five inputs as obtained in feature selection above and another input is the number of units purchased for the SKU. Two-step module in SPSS-Clementine 12.1 has been used for clustering and three clusters obtained are as given below:

Cluster-1
1244 Records

(1) Gender:
- M (80.23 percent):
  M 80.23 percent.
  F 19.77 percent

(2) Income:
- 20K-30K (70.50 percent):
  Below 20K – 4.34 percent.
  20 K-30 K – 70.50 percent.
  Above 30K – 25.16 percent.

(3) Number of children:
- 2 (65.76 percent):
  0 – 15.84 percent.
  1 – 12.54 percent.
  2 – 65.76 percent.
  3 or more – 5.87 percent.

(4) Level of education:
- 2 (48.55 percent):
  1 – 5.79 percent.
  2 – 48.55 percent.
  3 – 45.66 percent.

(5) Domicile of the province:
- No (57.07 percent):
  Yes – 42.93 percent.
  No – 57.07 percent.

(6) Number of units purchased:
- 2 (55.87 percent):
  1 – 30.06 percent.
  2 – 55.87 percent.
  3 – 11.74 percent.
  4 – 2.33 percent.

Cluster-2

3295 Records

(1) Gender:
- F (61.24 percent):

  M – 38.76 percent.

  F – 61.24 percent.

(2) Income:
- 20K-30K (53.35 percent):

  Below 20K – 10.41 percent.

  20K-30K – 53.35 percent.

  Above 30K – 36.24 percent.

(3) Number of children:
- 1 (49.47 percent):

  0 – 6.19 percent.

  1 – 49.47 percent.

  2 – 40.39 percent.

  3 or more – 3.95 percent.

(4) Level of education:
- 3 (74.72 percent):

  1 – 3.40 percent.

  2 – 21.88 percent.

  3 – 74.72 percent.

(5) Domicile of the province:
- Yes (56.87 percent):

  Yes – 56.87 percent.

  No – 43.13 percent.

(6) Number of units purchased:
- 1 (64.80 percent):

  1 – 64.80 percent.

  2 – 21.70 percent.

  3 – 12.59 percent.

  4 – 0.91 percent.

Cluster-3
5879 Records

(1) Gender:
- F (83.18 percent):

  M – 16.82 percent.

  F – 83.18 percent.

(2) Income:
- Above 30K (91.67 percent):

  Below 20K – 0.94 percent.

  20K-30K – 7.40 percent.

  Above 30K – 91.67 percent.

(3) Number of children:
- 2 (72.78 percent):

  0 – 0.43 percent.

  1 – 25.46 percent.

  2– 72.78 percent.

  3 or more – 1.33 percent.

(4) Level of education:
- 3 (92.11 percent):

  1 – 0.54 percent.

  2 – 7.35 percent.

  3 – 92.11 percent.

(5) Domicile of the province:
- No (92.99 percent):

  Yes – 7.01 percent.

  No – 92.99 percent.

(6) Number of units purchased:
- 3 (89.13 percent):

  1 – 0.71 percent.

  2 – 9.54 percent.

  3 – 89.13 percent.

  4 – 0.61 percent.

As per Step 5 of the proposed forecasting model, database is to be segregated based on this clustering in three parts – first part is given by the records in cluster 1, second part consists of the records in cluster 2 and third part consists of the records in cluster 3. On each segregated part, ARIMA is to be used for forecasting using the models suggested in the proposed schemes of DM-7, DM-8, WM-7 and WM-8. Predictors/regressors, as used in ARIMA, for daily and weekly forecasting have been identified as given below.

For daily forecasting, following predictors/regressors have been used. These are some relevant descriptions by binary variables (if true, than assigned value is "1", otherwise "0") of a day for the purpose of forecasting:

- Payment day indicator. First, second, third, fourth, fifth and last day of the month.
- Holiday indicator (including weekend).
- Festival/special day indicator.
- Pre-holiday indicator.
- Pre-festival/special day indicator.
- School vacation (summer/winter) indicator.
- Mega-offer day indicator.

For weekly forecasting, following predictors/regressors have been used. These are some relevant descriptions of a week for the purpose of forecasting. The basic features in the description of week and day remain same. In week, it is an aggregate effect of the days:

- Number of payment days (First, second, third, fourth, fifth and last day of the month).
- Number of holidays (including weekend).
- Number of festival/special days.
- Number of pre-holidays.
- Number of pre-festival/special days.
- Number of school vacation (summer/winter) days.
- Number of mega-offer days.

The performance of each model has been computed using mean absolute percentage error (MAPE) and normalized mean square error (NMSE) and their expressions are as given below (Makridakis *et al.*, 1998):

$$\mathrm{MAPE} = \frac{1}{n}\sum\left|\frac{X_k - X_{ok}}{X_k}\right|, \quad \text{for } k = 1, 2, \ldots, n, \text{and}$$

$$\mathrm{NMSE} = \frac{\sum(X_k - X_{0k})^2}{\sum(X_k - X_0)^2}, \quad \text{for } k = 1, 2, \ldots, n,$$

where, n is the number of periods, X_k is the forecast value of kth period, X_{0k} is the actual value of kth period and X_0 is the average of actual values of all the periods.

The proposed models along with the existing model in practice and other models available in literature have been compared. Performance has been compared on training dataset of 36 months and test data set of 12 months. Comparison of daily forecasting and weekly forecasting has been shown separately in Tables I and II, respectively.

It is found that in both daily and weekly forecasting, the proposed models perform better than other models. Overall, weekly forecasting proves better than the weekly forecasting.

6. 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 days and the purchase order has to be sent at least L

Model	Description of the model for daily forecasting	Training set MAPE	Training set NMSE	Test set MAPE	Test set NMSE
DM0	14-day moving average (existing practice)	51.47	0.8889	54.71	0.9256
DM1	ARIMA	49.61	0.8758	53.89	0.9178
DM2	Seasonal ARIMA	48.27	0.8803	52.47	0.9022
DM3	ARIMA with predictors	47.9	0.8696	50.03	0.8977
DM4	Seasonal ARIMA with predictors	45.45	0.8575	50.01	0.8877
DM5	Neural network on ARIMA with predictors	44.29	0.7635	48.8	0.8105
DM6	Neural network on seasonal ARIMA with predictors	44.16	0.7623	47.28	0.8111
DM7	Clustering-based ARIMA with predictors	32.16	0.5336	36.25	0.5669
DM8	Clustering-based seasonal ARIMA with predictors	32.11	0.5227	36.25	0.5522

Table I. Comparison of daily forecasting models

Model	Description of the model for weekly forecasting	Training set MAPE	Training set NMSE	Test set MAPE	Test set NMSE
WM1	ARIMA	38.70	0.7835	41.44	0.8012
WM2	Seasonal ARIMA	38.67	0.7789	41.09	0.7995
WM3	ARIMA with predictors	34.56	0.6742	37.75	0.6621
WM4	Seasonal ARIMA with predictors	31.39	0.6664	35.68	0.6656
WM5	Neural network on ARIMA with predictors	24.21	0.5656	26.32	0.5443
WM6	Neural network on seasonal ARIMA with predictors	21.95	0.4960	22.17	0.4540
WM7	Clustering-based ARIMA with predictors	17.92	0.3425	17.96	0.3701
WM8	Clustering-based seasonal ARIMA with predictors	14.01	0.3272	14.65	0.3375

Table II. Comparison of weekly forecasting models 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 standard normal distribution table which depends on 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. Both, the average demand (m_0) and the standard deviation (s) are estimated based on the used forecasting model for demand forecast. Using one year's sales data, simulation has been done for the inventory level of the product applying the replenishment model as discussed using the proposed models.

For 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:

$$\text{Standard deviation of demand during } (P+L)\text{days}$$
$$= \text{Standard deviation of daily forecast demand} \times \sqrt{(P+L)}$$

$$\text{Safety Stock} = z \times [\text{standard deviation of demand during } (P+L) \text{ Days}]$$

$$\text{Target or Desired Inventory Level} = T$$
$$= (\text{Average daily forecast demand}) \times (P+L) + \text{Safety stock}$$
$$= m_0 \times (P+L) + \text{Safety stock}$$

$$\text{Order Quantity}, Q = T - \text{Inventory Level}$$

For using 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:

$$\text{Average daily demand} = \text{Average weekly forecast demand}/7$$

$$\text{Standard deviation of daily demand} = \frac{\text{Standard deviation of weekly forecast demand}}{\sqrt{7}}$$

$$\text{Standard deviation of demand during } (P+L) \text{ days}$$
$$= \text{Standard deviation of daily demand} \times \sqrt{(P+L)}$$

$$\text{Safety Stock} = z \times [\text{standard deviation of demand during } (P+L) \text{ Days}]$$

$$\text{Target or Desired Inventory Level} = T$$
$$= \text{Average daily demand} \times (P+L) + \text{Safety stock}$$

$$\text{Order Quantity}, Q = T - \text{Inventory Level}$$

Comparison of inventory performance based on the four proposed forecasting models (DM7, DM8, WM7, WM8) outperforming other models and the existing forecasting model (DM0) has been shown in Table III. It is found that weekly forecasting models, WM7 and WM8, outperform their counterparts in daily forecasting, DM7 and DM8. However, all four outperform the existing practice. With respect to inventory level given by reaching days and sales failure both, WM8 is found to be slightly better than WM7.

7. Conclusions

The proposed business intelligent system for demand forecasting proves to give more accurate prediction for future demands compared to the existing models and practices in 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. However, the model in this paper improves both the performance indicators simultaneously. Reduction of reaching days implies reduction in inventory level and reduction in sale failure shows increase in service level. Aburto and Weber (2007) had shown that ARIMA with neural network for errors outperforms the models based on pure ARIMA or pure neural network. The model proposed in this work outperforms all these models. Various information of payment days, holidays, festival days, etc. are also taken into account by the model while predicting consumers' behavior in purchasing. In the present day order, where, most of the large supermarkets are taking resort to data mining for various uses in CRM (Bala, 2009a; Chang *et al.*, 2009), 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.

Model/existing practice		Description of the model	Inventory performance indicator: Reaching days (average of daily inventory/daily sale)	Number of sales failure days, i.e. without products)	Sales failure (percentage of days without products)
Actual findings	DM0	14-day moving average (existing practice)	32.05	105	28.77
Simulation results	DM7	Decision tree-based ARIMA with predictors	24.82	68	18.63
	DM8	Decision tree-based seasonal ARIMA with predictors	23.73	62	16.99
	WM7	Decision tree-based ARIMA with predictors	18.26	23	6.30
	WM8	Decision tree-based seasonal ARIMA with predictors	17.04	20	5.48

Table III. Performance of inventory replenishment systems based on various forecasting models

Clustering of the customers or the records is most important feature in the model suggested. Strong natural groups may not always come for a particular product in doing clustering. Features or attributes selected using "feature selection" plays a big role in the quality of clusters obtained. Some of the important features may be chosen for clustering and those may result in an appropriate clustering. Care should be taken while identifying the attributes prior to feature selection. In this context, the proposed model is appropriate for items for which a substantial proportion of sale is attributed strongly to a particular profile or profiles. For such cases, the suggested model can be extended further to understand the dynamics of migration and immigration of people with specific profiles of interest in a locality. Using the weak clusters from the sale records for demand forecasting provides scope for further work.

References

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.

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.

Bala, P.K. (2009a), "A data mining model for investigating the impact of promotion in retailing", *Proceedings of IEEE International Advance Computing Conference, Patiala, India, March 6-7*, pp. 670-4.

Bala, P.K. (2009b), "Data mining for retail inventory management", in Ao, S.I. and Gelman, L. (Eds), *Advances in Electrical Engineering and Computational Science*, LNEE Series, Vol. 39, Springer, New York, NY, pp. 587-98.

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.

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.

Bowerman, B.L. and O'Connell, R.T. (1993), *Forecasting and Time Series: An Applied Approach*, 3rd ed., Duxbury Press, Belmont, CA.

Chang, C.W., Lin, C.T. and Wang, L.Q. (2009), "Mining the text information to optimizing the customer relationship management", *Expert Systems with Applications*, Vol. 36 Nos 2/1, pp. 1433-43 (Part-1).

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.

Huarng, K.-H. and Yu, T.H.K. (2006), "The application of neural networks to forecast fuzzy time series", *Physica A: Statistical Mechanics and its Applications*, Vol. 363 No. 2, pp. 481-91.

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.

Huarng, K.-H., Yu, T.H.K. and Sole Parellada, F. (2011), "An innovative regime switching model to forecast Taiwan tourism demand", *The Service Industries Journal* (*Special Issue on Tourism Services*), Vol. 31 No. 10, pp. 1603-12.

McKenzie, E. (1984), "General exponential smoothing and the equivalent ARIMA process", *Journal of Forecasting*, Vol. 3 No. 3, pp. 333-444.

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.

Ngai, E.W.T., Xiu, L. and Chau, D.C.K. (2009), "Application of data mining techniques in customer relationship management: a literature review and classification", *Expert Systems with Applications*, Vol. 36 No. 2, pp. 2592-602 (Part-2).

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

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

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.

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

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.

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.

Yu, T.H.K., Huarng, K.-H. and Rianto, Rapon (2009), "Neural network-based fuzzy auto-regressive models of different orders to forecast Taiwan stock index", *International Journal of Economics and Business Research*, Vol. 1 No. 3, pp. 347-58.

About the author

Pradip Kumar Bala is an Associate Professor in Xavier Institute of Management, Bhubaneswar (XIMB), India. He received his PhD, M.Tech and B.Tech (Hons) from IIT, Kharagpur, India. His research interests include inventory management, quality management, maintenance management and data mining. His several papers on purchase dependence in retail inventory management and retail promotion have been published in reputed international journals and conferences. One of his papers was awarded Certificate of Merit at the World Congress on Engineering – 2008, organized by IAENG in London. Pradip Kumar Bala can be contacted at: p_k_bala@rediffmail.com