International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:6, No:3, 2012

Consumer Product Demand Forecasting based on Artificial Neural Network and Support Vector Machine

Karin Kandananond

***Abstract*—**The nature of consumer products causes the difficulty in forecasting the future demands and the accuracy of the forecasts significantly affects the overall performance of the supply chain system. In this study, two data mining methods, artificial neural network (ANN) and support vector machine (SVM), were utilized to predict the demand of consumer products. The training data used was the actual demand of six different products from a consumer product company in Thailand. The results indicated that SVM had a better forecast quality (in term of MAPE) than ANN in every category of products. Moreover, another important finding was the margin difference of MAPE from these two methods was significantly high when the data was highly correlated.

***Keywords*—**Artificial neural network (ANN), Bullwhip effect, Consumer products, Demand forecasting, Supply chain, Support vector machine (SVM).

I. Introduction

DEMAND forecasting is critical to the efficiency improvement of supply chain system because each party of the supply chain will process the order according to the demand signal, As a result, the accuracy of demand forecasts will significantly improve the production scheduling, capacity planning, material requirement planning and inventory management. Without the accurate forecasting, there were many consequences including the bullwhip effect. The bullwhip effect occurs when the variabilities of demand in the supply chain are magnified as they moved up the chain. This will lead to inefficiency supply chain system. One of the major causes of the bullwhip effect is the lack of accuracy in the demand forecasting. Moreover, the demand forecasting for some products is considered challenging, especially consumer products, because of their complicated characteristics. Basically, there is a plenty of forecasting methods based on the historical demand data. Among these approaches are the machine learning method, an algorithm which has the potential to learn to make accurate predictions based on the previous observations. In this research, two machine learning methodologies, the artificial neural network (ANN) and support vector machine (SVM) were studied to compare the performance of each method in different scenarios.

Karin Kandananond is with Rajabhat University Valaya-Alongkorn, Thailand (e-mail: Kandananond@hotmail.com).

II. Literature Review

There is a number of works contributed to performance assessment of ANN and SVM methods. Bansal, Vadhavkar and Gupta [1] had identified the inventory patterns of a large medical distribution organization and elaborated a method to construct and choose an appropriate neural network for optimizing the inventory. The implementation led to the reduction of the total inventory by 50% in the organization while the customer satisfaction level was still high. Tay and Cao [2] compared the performance of SVM and ANN to forecast the financial time series. The historical data was based on five real future contracts collected from Chicago Mercantile Market). The results indicated that SVM had a better performance than the ANN. Kim [3] applied supporting vector machine (SVM) to predict stock price index and compared the performance with the one from ANN. The study showed that SVM approach outperformed the ANN method. Huang, Nakamori and Wang [4] also utilized SVM to predict the stock price index NIKKEI 225. SVM outperformed the ANN, Linear Discriminant analysis and Quadratic Discriminant Analysis. Hua and Zhang [5] had utilized support vector machines (SVMs) approach to forecast the demand of spare parts. The data used were spare parts from a petrochemical enterprise in China. It is obvious that the introduced method was able to forecast the demands of spare parts than the currently used methods. Gutierreza, Solis and Mukhopadhyay [6] applied artificial neural network method to forecast the lumpy demand and compared the performance of ANN to those using three traditional (single exponential smoothing, Croston's method, and the Syntetos–Boylan approximation). The results showed that it outperformed the remaining three methods significantly. According to the literature, there were some works conducted empirically to compare the performance of SVM and ANN method. However, each of these studies was based on the conclusion from a set of data with a single pattern. Therefore, sets of data with different patterns should be deployed in order to assess the performance of these two approaches.

III. Methodology

Two methods used in this study were ANN and SVM approaches;

A. ANN Method

The development of ANN models was based on studying the relationship of input variables and output variables. Basically, the neural architecture consisted of three or more

layers, i.e. input layer, output layer and hidden layer [7] as shown in Fig. 1. The function of this network was described as follows:

$$Y_j = f(\sum_i w_{ij} X_{ij}) \quad (1)$$

where Y_j is the output of node j, f (.) is the transfer function, w_{ij} the connection weight between node j and node i in the lower layer and X_{ij} is the input signal from the node i in the lower layer to node j.

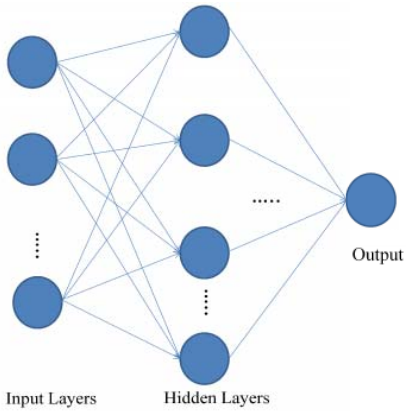

Fig. 1 The architecture of a neural network

B. SVM Method

Support Vector Machine (SVM) was a classification method which was based on the construction of hyperplanes in a multidimensional space. As a result, it was allowed different class labels to be differentiated. Normally, SVM was utilized for both classification and regression tasks and it was able to handle multiple continuous and categorical variables. The purpose of the regression task of SVM was to find a function f (such that y = f(x) + noise) which was able to predict new cases. This was achieved by training the SVM model on a sample set, i.e., training set, a process that involved the sequential optimization of an error function [8]. There were two types of SVM model for regression, type 1 and 2. For regression type 1, the objective function was the minimization of the error function.

$$\min \frac{1}{2} w^T w + C\sum_{i=1}^{N} \xi_i + \sum_{i=1}^{N} \xi^*{}_i$$

s.t.

$$w^T \phi(x_i) + b - y_i \le \varepsilon + \xi^*{}_i$$

$$y_i - w^T \phi(x_i) - b_i \le \varepsilon + \xi^*{}_i$$

$$\xi_i, \xi^*{}_i \ge 0, i = 1,..., N, \varepsilon \ge 0$$

Similarly, objective function of the regression type 2 was

$$\min \frac{1}{2} w^T w - C[v\varepsilon + \frac{1}{N}\sum_{i=1}^{N} (\xi_i + \xi^*{}_i)]$$

s.t.

$$w^T \phi(x_i) + b - y_i \le \varepsilon + \xi^*{}_i \quad y_i - w^T \phi(x_i) - b_i \le \varepsilon + \xi^*{}_i$$

$$\xi_i, \xi^*{}_i \ge 0, i = 1,..., N, \varepsilon \ge 0 .$$

There were four types of kernels (ϕ), linear, polynomial, radial basis function (RBF) and sigmoid, used for SVM models. Among these kernels, RBF was the most frequently used kernel because of their localized and finite responses across the entire range of the real x-axis [8]. The functions of these kernels were shown as follows:

$$\phi = \begin{cases} x_i * y_i .....Linear \\ (\gamma x_i X_j + coefficient)^{number} .....Polynomial \\ \exp(-\gamma |X_i - x_j|^2) .....RBF \\ \tanh(\gamma x_i X_j + coefficient) .....Sigmoid \end{cases}$$

IV. Research Procedures

Monthly data of six different consumer products, cooking aids, shower gel, body lotion, dishwashing liquid, deodorant and fabric detergent, from January 2009 to August 2011 (32 cases) was used to assess the performance of the designated methods. The time series plots of each product demand were shown in Fig. 2, 3, 4, 5, 6 and 7 in order to study the pattern and trend of the data.

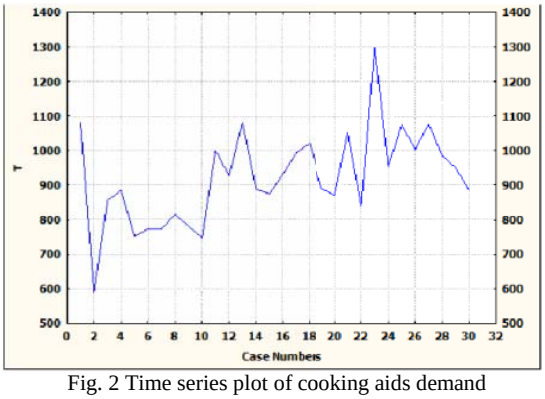

Fig. 2 Time series plot of cooking aids demand

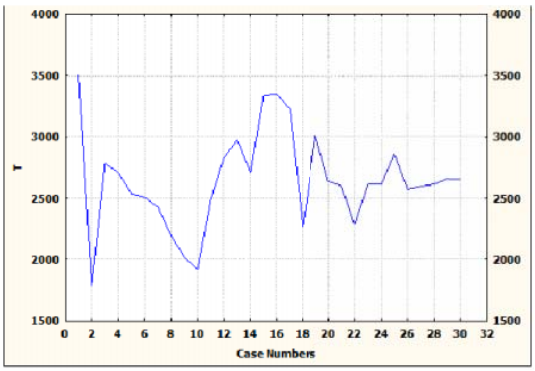

Fig. 3 Time series plot of shower gel demand

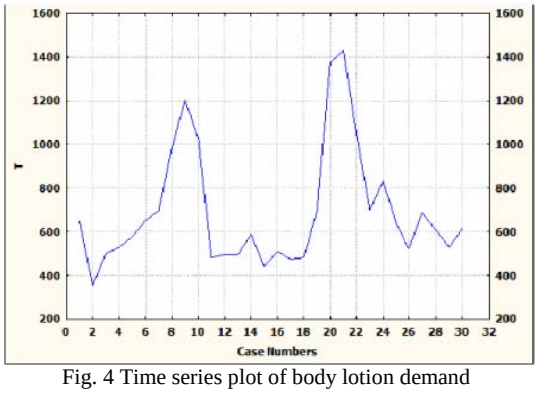

Fig. 4 Time series plot of body lotion demand

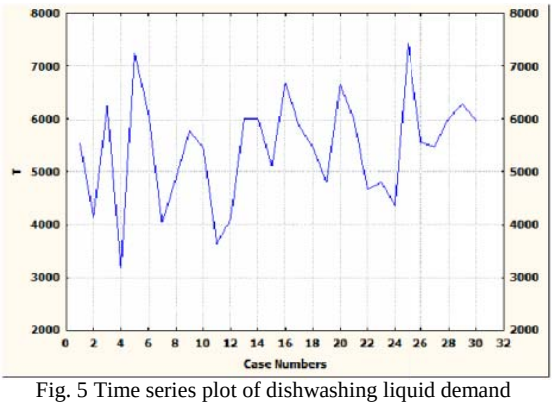

Fig. 5 Time series plot of dishwashing liquid demand

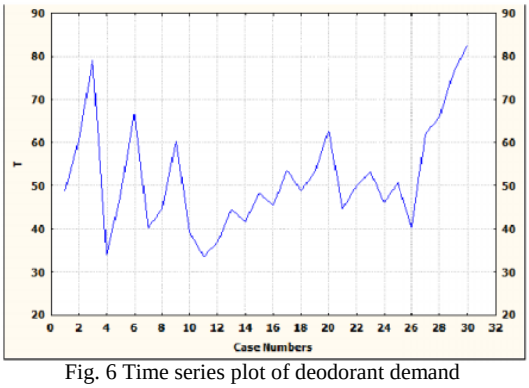

Fig. 6 Time series plot of deodorant demand

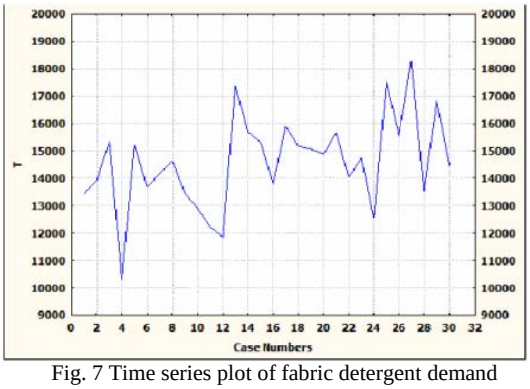

Fig. 7 Time series plot of fabric detergent demand

The characteristic of each sample data was analyzed by considering correlogram and time series plots and the results was shown in Table I.

TABLE I
DATA CHARACTERISTICS

Product Category	Correlation/Patterns
Cooking aids	Highly positive correlated (lag 2)
Shower gel	Highly positive correlated (lag 7)
Body lotion	-Highly positive and negative correlated (lag 1,5,6,7,8,11,12,13) -Seasonal pattern (cyclic)
Dishwashing liquid	Not correlated
Deodorant	Not correlated
Fabric detergent	Not correlated

After the tested data was chosen, two proposed methods, ANN and SVM, were utilized to construct models based on the sample data to forecast the demand of these six products using a statistical package, STATISTICA version 8. The performance of these approaches was justified by considering their error measurement, mean absolute percentage error (MAPE).

V. RESULTS

A. ANN

Two most popular neural network architectures, multilayer perceptrons (MLP) and radial basis function (RBF), were utilized for the regression purpose [7]. The inputs for training were the historical demand at t-8, t-7,…, t-1 while the amount of networks to train was set at 200 while the top performed five networks were retained for each type of products. The network with the best performance was kept to forecast the demand of each category (time: t). The results were shown in Table II.

TABLE II
MODEL, HIDDEN ACTIVATION AND OUTPUT ACTIVATION.

Product Category	ANN Model	Hidden Activation	Output Activation
Cooking aids	RBF (8-7-1)	Gaussian	identity
Shower gel	RBF (8-7-1)	Gaussian	identity
Body lotion	MLP (8-4-1)	Tanh	exponential
Dishwashing Liquid	MLP (8-8-1)	Exponential	identity
Deodorant	RBF (8-7-1)	Tanh	logistic
Fabric detergent	RBF (8-7-1)	Exponential	logistic

TABLE III
TRAINING ALGORITHM AND MAPE

Product Category	Training Algorithm	MAPE
Cooking aids	RBFT	0.300
Shower gel	RBFT	0.303
Body lotion	BFGS 3	0.5082
Dishwashing Liquid	BFGS 19	0.2346
Deodorant	BFGS 1	0.1955
Fabric detergent	BFGS 1	0.03724

The results in Table II pointed out that the number of hidden layers was ranged from 4 to 8 layers. According to MAPE, the MLP architecture might be suitable for the data set which was not correlated or the data with an identified pattern (in this case, cyclical pattern). On the other hand, RBF might

be preferred when the data was highly correlated or without any specific pattern. The hidden neuron activation functions of the retained five networks were Gaussian, tangent hyperbolic (tanh) and exponential while the exponential, identity (the activation of the neuron is passed on directly as the output) and logistic were assigned to the output neuron activation functions. Moreover, the training algorithm of the MLP network employed to build the models was the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [7] with the number of cycles used to train the model which was ranged from 1 to 19 cycles. The results in Table 3 signified that the MAPEs of three products whose data were not correlated (dishwashing liquid, deodorant and fabric detergent) were lower than the ones with correlated characteristics (cooking aids, shower gel and body lotion).

B. SVM

Similar to ANN case, the indicators used for SVM application were the historical data at t-8, t-7,…, t-1 to predict the demand at time t. The forecasting model based on SVM approach was the regression type 1 with C=10.0, epsilon = 0.1 and the kernel was radial basis function with gamma = 0.5. The number of support vectors and MAPE from the prediction for each category of products was illustrated in Table 4.

TABLE IV
NUMBER OF SUPPORT VECTORS AND MAPE

Product Category	Number of Supporting vectors	MAPE
Cooking aids	20	0.1055
Shower gel	16	0.1015
Body lotion	14	0.1886
Dishwashing liquid	2	0.0116
Deodorant	2	0.0148
Fabric detergent	2	0.0086

According to table IV, it indicated that the data with high correlation needed more number of support vectors than the ones with no correlation. The similar results also reflected on the MAPE since the MAPE of SVM models for dishwashing liquid, deodorant and fabric detergent (no correlation) was significantly lower than the ones with highly correlated characteristics.

VI. CONCLUSION

The forecasting errors (MAPE) from two methods were concluded as follows:

TABLE V
MAPES FROM ANN AND SVM MODELS

Product Category	MAPE	
	ANN	SVM
Cooking aids	0.300	0.1055
Shower gel	0.303	0.1015
Body lotion	0.5082	0.1886
Dishwashing Liquid	0.2346	0.0116
Deodorant	0.1955	0.0148
Fabric detergent	0.03724	0.0086

According to Table V, it signified that SVM outperformed ANN in every category of products. However, it was important to note that the performance of these two methods was significantly different only when the data set was not correlated (dishwashing liquid, deodorant and fabric detergent). However, the gap was highly likely to be narrowed when the sample data was highly correlated (cooking aids, shower gel and body lotion).

REFERENCES

[1] K. Bansal, S. Vadhavkar, and A. Gupta, “Neural Networks Based Forecasting Techniques for Inventory Control Applications,” *Data Min Knowl Disc*, vol. 2, no. 1, pp. 97-102, 1998.
[2] F.E.H. Tay and L. Cao, “Application of Support Vector Machines in Financial Time Series Forecasting,” *Omega*, vol. 21, pp. 309-317, 2001.
[3] K. Kim, “Financial Time Series Forecasting using Support Vector Machines,” *Neurocomputing*, vol. 55, no. 1-2, pp. 307-319, 2003.
[4] W. Huang, Y. Nakamoria and S.-Y. Wang, “Forecasting Stock Market Movement Direction with Support Vector Machine,” *Comput Oper Res* , vol. 32, no.10, pp. 319-326, 2005.
[5] Z. Hua and B. Zhang, “A Hybrid Support Vector Machines and Logistic Regression Approach for Forecasting Intermittent Demand of Spare Parts,” *Appl Math Comput*, vol. 181, pp. 1035-1048, 2006.
[6] R.S. Gutierrez, A.O. Solis and S. Mukhopadhyay, “Lumpy Demand Forecasting using Neural Networks,” *Int J Product Econ*, vol. 111, pp. 409-425, 2008.
[7] K. Kandananond, “"Forecasting Electricity Demand in Thailand with an Artificial Neural Network Approach," *Energies*, vol. 4, pp. 1246-1257, 2011.
[8] StatSoft, Inc. (2011). Electronic Statistics Textbook. Tulsa, OK: StatSoft. WEB: http://www.statsoft.com/textbook/.