MACHINE LEARNING APPROACHES FOR THE COSMETIC SALES FORECASTING

[1]KAI-BIN LIN, [2]LAI-WANG WANG, [3]BI-MIN HSU, [4]MING-HUNG SHU

[1,2,4]Industrial Engineering and Management Department, National Kaohsiung University of Applied Science, Kaohsiung, Taiwan
[3]Industrial Engineering and Management Department, Cheng Shiu University, Kaohsiung, Taiwan
E-mail: [1]kk124228@gmail.com, [2]dorthytingkimo@gmail.com, [3]biminhsu@gmail.com,
[4]workmanshu@gmail.com

Abstract— In the contemporary information society, constructing an effective sales prediction model is challenging due to the sizeable amount of purchasing information obtained from diverse consumer preferences. Many empirical cases shown in the existing literature argue that the traditional forecasting methods, such as the index of smoothness, moving average, and time series, have lost their dominance of prediction accuracy when they are compared with the modern types of forecasting approaches, such as the neural network (NN) and support vector machine (SVM) models. To verify these findings, this paper utilizes the Taiwanese cosmetic sales data to examine three forecasting models, namely, the back propagation neural network (BPNN), least-square support vector machine (LSSVM), and auto regressive model (AR). The result concludes that the LS-SVM has the smallest mean absolute percent error (MAPE) and largest Pearson correlation coefficient (R^2) between model and predicted values.

Keywords— Machine Learning, Back Propagation Neural Network, Least Square Support Vector Machine, Autoregressive Model, Model Performances Comparing.

I. INTRODUCTION

Information technology and internet have been more and more sophisticated in Taiwan. This situation brings the diversity of the information and leading the companies to predict the customer behavior become more difficult. This research will focus on the cosmetic industry which is one of the famous industries in Taiwan. Generally, the main forecasting method for sales is discovered by using questionnaire then analyzed with basic statistical approaches. Machine learning, as important tool in data mining, and which provide many algorithm to predict and cause choosing suitable method has become a universal question. The forecasting results depend on the accuracy of prediction models. Although the basic statistical forecasting model, such as Exponential Smoothing approach, Linear Regression, etc. is simple and practical but the actual prediction accuracy is unstable. Because of machine learning methods offer many algorithms to predict for time series data instead of traditional forecasting methods. Thus, this study develops two machine learning methods, back propagation neural network (BPNN) and least square support vector machine (LSSVM) compare with Auto Regressive model (AR).

Machine learning techniques can be track back to the 1980s, which growth rapidly after the machine learning conference hold on Carnegie Mellon University. Since technology and machine learning has been developing for a long time, great quantity data sets are accumulated and can be used to examine algorithm. Several researchers have their own definition for machine learning, but all of these definitions are making computers teach themselves and improve performance with artificial intelligence [1], [2], [3], [4], [5]. There are many algorithms (such as: Neural Network, Support Vector Machine, etc.), all of them are developed to learn for classification or cluster, then proceeding forecasting base on learning results. Because of these methods has theirs advantage and disadvantage, several literatures apply different method in same data sets and measure the performance to find which method is more suitable for adopted data [6], [7]. In machine learning, artificial neural network (ANN) is one of the most common tools to solve classification and forecasting problems and which was first proposed by Mcculloch and Pitts in 1943. Their study created a neural network computing model based on mathematics and threshold logic algorithm. ANN is a calculating system, which simulates the processing system of animal neural network with computer and imitates its ability with many simple neurons [8], [9]. In 1990, Webos proposed back propagation neural network (BPNN) by combined error back propagation rule and ANN. To train multiple layer neural networks quickly is a difficult task, but BPNN provide the solution effectively [10], [11]. Support vector machine (SVM) is the other useful classification and prediction way in machine learning, which was first invented by Vapnik and Cotes in 1995. Creating a hyperplane is the main target of SVM which will cause either side of the data has a fairly uniform distinction [5], [12]. Suykens and Vandewalle proposed the concepts of least square support vector machine by integrated least square method and SVM. In general, LSSVM can be differentiated to classification and regression problem. The aim of classification is to distinguish the data point and find the maximum interval hyperplane. For regression problem, the main intent is forecasting after classification. [13], [14], [15],

[16]. The Auto Regressive model is included in auto regressive integrated moving average model (ARIMA), which was indicated by Box and Jenkin in 1970. The characteristic of ARIMA is able to integrate the non-stationary time series data and making it station with differencing. Furthermore, ARIMA can be distinguished to Auto Regressive model (AR), moving average (MA), and Auto Regressive model moving average (ARMA) with stationary. If ARMA model is applied in non-stationary time series and integrated by differencing, then ARMA will change to ARIMA model [17], [18], [19], [20].

The main purpose of this paper is to provide three methodologies for predicting the sales of cosmetic and make a comparison of errors and performances between adopted machine learning algorithms and Auto Regressive model. Then find the most suitable model for cosmetics sales forecasting. In following sections, section 2 descripts the proposed model; section3 displays result and making some discussion. Finally, based on the outcome of analysis in section 3, section 4 gives conclusions and suggest further study.

II. METHODOLOGY

2.1. Study Design

The approaches of this paper are forecasting with three proposed model and make comparisons by evaluating model performances and errors after collecting Taiwanese cosmetic industry monthly sales. The structure of study design is demonstrated as following Fig.1. , and which can be distinguished to three part. Firstly, the process of BPNN was listed in left side of the figure. Secondly, the procedure of LSSVM was shown in middle of the figure. Thirdly, the AR(1) model which is named as first order auto regressive model is displayed in right side of the figure. Finally, giving summaries based on the results of comparing.

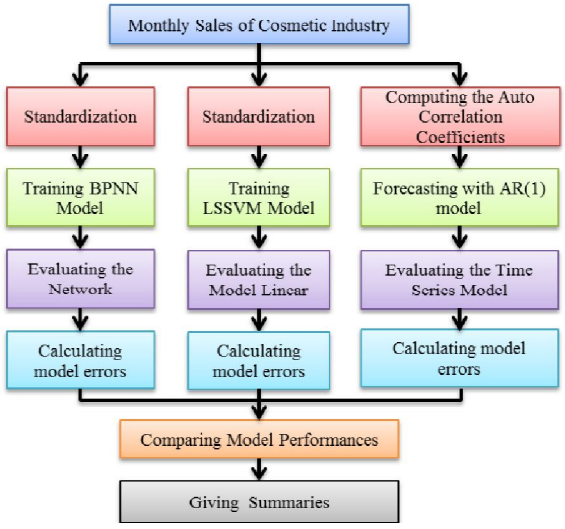


Fig.1. Structure of Methodologies

2.2. Architecture of BPNN model

BPNN can solve many difficult problems by cooperating with highly interconnected but simple computing neurons. It adopts logistic function as transforming function to sum the inputting weights. The definition of this function is represented as following equation:

$$y(v_i)=\tanh(v_i) \quad and \quad y(v_i)=\left(1+e^{-v_i}\right)^{-1} \qquad (1)$$

Where y_i is output of i the neuron, v_i is the sum of inputting weights. The former logistic function is hyperbolic tangent function and its range from -1 to 1. However, the latter function is similar to the former but range from 0 to 1.

Several literatures indicated that the hidden layers of net were determined by number of input and output layer, and which might be decided with trial and error method. In this paper, we try one hidden layer with 3, 6, 9 nodes and two hidden layer with 3, 6, 9 nodes. The results of attempt shows that using two hidden layers and each contain six nodes will give the best reliability for the network. Thus, the structure of proposed BPNN model is shown as Fig.2. We can see an input layer contains three nodes represent input vectors, an output layer includes a node as output vector, and two hidden layer each involve six hidden nodes. If we assuming the node of input layer is x, the node of hidden layer is y_j and y_k, the node of output layer is O, the connection weights between input layer and first hidden layer is W_{xyj}, the connection weights between two hidden layer is Wy_jy_k, the connection weights between second hidden layer and output layer is Wy_kO.

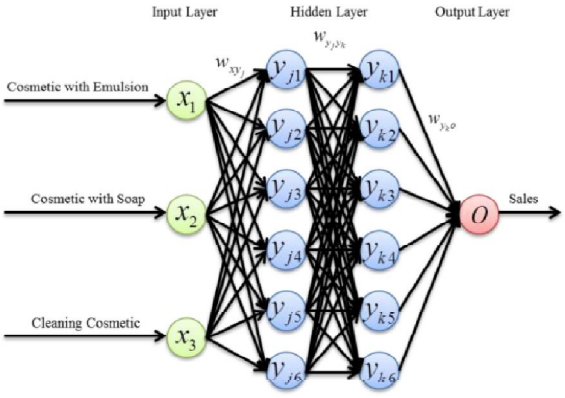


Fig.2. Architecture of Proposed BPNN Model

2.2. Mathematical Definition of LSSVM Model

Because of the limiting equation of general SVM is an inequality which may make calculating procedure become more complicate. However, LSSVM adopts the quadratic loss function of least squares method instead of the insensitive loss function of SVM. Moreover, the limited condition of LSSVM is change from inequality to equation, so that the solution process becomes linear equations, and which causes the speed of solving more quickly and efficiently.

That is the reason why it has been widely used in the forecasting domain. Furthermore, LSSVM can be distinguished to classification problem and regression problem. Forecasting is belonging to regression problem in LSSVM, so we assuming *T* is *n* is sample size, in this situation *T* can represent as equation (2):

$$T=\left\{\left(x_1,y_1\right),\left(x_2,y_2\right),\cdots,\left(x_n,y_n\right)\right\} \quad (2)$$

where $x_i \in R^n$ is input vector, $y_i \in R$ is output variable corresponding to x_i.

From reviewing a large amount literature we find that the SVM model always compare with NN model, the reason might be both of them are calculate by minimizing error function. The minimization problem of LSSVM is represented as following equation:

$$\min\ J_2\left(w,b,e\right)=\frac{\mu}{2}w^Tw+\frac{\zeta}{2}\sum_{i=1}^{N}e_{c,i}^2 \quad (3)$$

in addition, equation (3) was limited by equation (4):

$$y_i[w^T\phi(x_i)+b]=1-e_{c,i}\quad i=1,2,\cdots,N. \quad (4)$$

if the binary target of LSSVM is $y_i=\pm 1$, then using $y_i^2=1$ can get following equation (5):

$$\sum_{i=1}^{N}e_{c,i}^2=\sum_{i=1}^{N}(y_ie_{c,i})^2=\sum_{i=1}^{N}e_i^2=\sum_{i=1}^{N}\left(y_i-\left(w^T\phi\left(x_i\right)+b\right)\right)^2 \quad (5)$$

where $e_i=y_i-\left(w^T\phi\left(x_i\right)+b\right)$ are the errors of model, these errors will let LSSVM have a meaning to corresponding data. Thus, the classifying equation of LSSVM is shown as equation (6):

$$J_2(w,b,e)=\mu E_w+\zeta E_D \quad (6)$$

where $E_D=\frac{1}{2}\sum_{i=1}^{N}e_i^2=\frac{1}{2}\sum_{i=1}^{N}\left(y_i-\left(w^T\phi\left(x_i\right)+b\right)\right)^2$ and $E_W=\frac{1}{2}w^Tw$, the μ and ζ must be the number of hyper Parameters normalized corresponding to the adjusted sum of squared errors (SSE). Moreover, the feasible solutions are depend on adjusted ratio $\gamma=\zeta/\mu$. Thus, the goal of adopting μ and ζ is give a traditional Baysian's probability to explain and get feasible solutions of LSSVM regression after construct Lagrangian multiplier.

Selecting kernel function is a necessary procedure before running the LSSVM. Kernel function can be differentiated to five types, general kernel function, linear kernel function, polynomial kernel function, sigmoid function, and Gaussian radial basis function kernel. In general, the format of kernel function can be represented as equation (7):

$$K\left(\vec{x_i},\vec{x_j}\right)=\phi\left(\vec{x_i}\right)\cdot\phi\left(\vec{x_j}\right) \quad (7)$$

where the Greek notation ϕ is a kernel function, $\phi(x)$ is the mapping of data in the other space. The general kernel function can extend many formats with different applications. For example, this paper will use the Gaussian radial basis function kernel as the kernel of LSSVM, and its function is displayed in equation (8):

$$K\left(\vec{x_i},\vec{x_j}\right)=e\left(-\frac{\left\|x_i-x_j\right\|^2}{\sigma^2}\right) \quad (8)$$

in the equation, $\vec{x_i}$ and $\vec{x_j}$ is the feature vectors of input space, which was given from training sample and test sample. $\left\|x_i-x_j\right\|^2$ as the square of Euclidean distance between two feature vectors, e is exponential, and σ is a free parameter.

2.3. Statistical Procedure of Auto Regressive Model

Auto Regressive model is one of time series model which use its historical data in the past as independent variables. Assuming the independent variables is previous term, called first order auto regressive model AR (1) and which was defined as following equation:

$$Y_t=\beta_0+\beta_1Y_{t-1}+\varepsilon_t \quad (9)$$

where $\varepsilon_t\sim WN(0,\sigma^2)$ and $E\left(\varepsilon_tY_{t-j}\right)=0,\quad \forall j>0$, and β_1 is named as autoregressive coefficients. By repeated iterations can obtain equation (10):

$$Y_t=\beta_0\sum_{j=0}^{\infty}\beta_1^j+\sum_{j=0}^{\infty}\beta_1^j\varepsilon_{t-j}+\lim_{k\to\infty}\beta_1^kY_{t-k} \quad (10)$$

when $\left|\beta_1\right|<0$, then $\lim_{k\to\infty}\beta_1^kY_{t-k}=0$, and we can receive equation (11):

$$Y_t=\mu+\sum_{j=0}^{\infty}\beta_1^j\varepsilon_{t-j} \quad (11)$$

thus, the expectation is:

$$E\left[Y_t\right]=\mu=E\left[Y_{t-j}\right] \quad (12)$$

and variable is:

$$\begin{aligned}Var\left(Y_t\right)&=\sigma^2\left[1+\beta_1^2+\beta_1^4+\cdots\right]\\&=\frac{\sigma^2}{1-\beta_1^2}<\infty,\quad if.\quad |\beta|<1\end{aligned} \quad (13)$$

then, the condition of stationary is $|\beta|<1$.

If we want forecast the value of k th term in t th term. By recursive way, we can adapt Y_{t+k} to equation (14):

$$\begin{aligned}Y_{t+k}&=\beta_1Y_{t+k-1}+\varepsilon_{t+k},\\&=\beta_1\left[\beta_1Y_{t+k-2}+\varepsilon_{t+k-1}\right]+\varepsilon_{t+k},\\&=\beta_1^2Y_{t+k-2}+\beta_1\varepsilon_{t+k-1}+\varepsilon_{t+k},\\&\vdots\\&=\beta_1^kY_t+\beta_1^{k-1}\varepsilon_{t+1}+\cdots+\beta_1\varepsilon_{t+k-1}+\varepsilon_{t+k}\end{aligned} \quad (14)$$

in the situation of minimizing mean squared error, we can use the Information collection $\Omega_t = \{Y_t, Y_{t-1}, \cdots\}$ as conditional expectation, and the concept can be written as following equation:

$$E\left(Y_{t+1}\middle|\Omega_t\right) - E_t\left(Y_{t+1}\right) \tag{15}$$

the equation (15) is optimal forecasting formula for Y_{t+1}. In other word, the linear extension of projecting Y_t to Ω_t can be used to predict Y_{t+1}, and the concepts can be referred as following equation:

$$E\left(Y_{t+1}\middle|H_t\right) = \sum_{j=0}^{\infty} \beta_j Y_{t+j} \tag{16}$$

in equation (16), $H_t = S_p\left[Y_t, Y_{t-1}, \cdots\right]$ is the extension of linear space. Once the joint probability distributions of Y_t is follow normal distribution, then the projected linear will equal to the conditional expectation and the equation is shown as follow:

$$E\left(\varepsilon_{t+1}\right) = 0 \tag{17}$$

where ε_{t+k} is a exogenous shocks. It can be projected to the information collection of t th term, then will obtain 0. Thus, consider a time series data $\{Y_t\}_{t=1}^{T}$ and historical data is $\{Y_1, Y_2, \cdots Y_T\}$. Base on the above description, the equation of forecasting $T+k$ th term is shown as follow:

$$E\left(Y_{T+k}\middle|\Omega_T\right) = E_T\left(Y_{T+k}\right) = \beta_1^k Y_T \tag{18}$$

in equation (18), the β_1 is unknown parameter so we need to find the estimator of β_1 and construct the predictive function of Y_{T+k} with historical data $\{Y_1, Y_2, \cdots Y_T\}$. The estimation equation is demonstrated as following equation (18):

$$E_T\left(Y_{T+k}\right) = \beta_1^k Y_T \tag{18}$$

where $E_T\left(Y_{T+k}\right)$ is the equation which established with random samples. The subscript $(T+k)$ and (T) are the forecasting value of $T+k$ in T th term.

2.4. Performance Criteria

In order to evaluate the performances of proposed model, we need some criteria to measure. The methods of forecasting error measurement, max error (ME), mean absolute percent error (MAPE), and mean square error (MSE) will be applied to evaluate the adopted model in this paper and their mathematical equation are shown as equation (18) ~ (20). Furthermore, the accuracy standards of MAPE for predicting model are represented in Table 1.

$$ME = \max\left|y_f \quad y_r\right| \tag{18}$$

$$MAPE(\%) = 100 \times \frac{\sum_{k=1}^{N}\left|\frac{y_r - y_f}{y_r}\right|}{N} \tag{19}$$

$$MSE = \frac{\sum_{k=1}^{N}\left(y_r - y_f\right)^2}{N} \tag{20}$$

in equation (18) ~ (20), y_r is actual value, y_f is forecasting value in k th, and N is sample size.

To verify the proposed model whether suitable enough or not, the square of Pearson product moment correlation coefficient (R^2) will be used in evaluating model performance too. R^2 is very important in linear performance measuring. Its definition can be defined easily as equation (21) and **Table 2** demonstrates the evaluating criterion of R^2.

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \tag{21}$$

where SSR is sum of square regression, SST is sum of square total, and sum of square error is define as SSE.

Table 1: Accuracy Standards of MAPE

MAPE (%)	Descriptions
Under 10%	Accurate forecasting
Between 10% and 20%	Good forecasting
Between 20% and 50%	Feasible forecasting
Over 50%	Inaccurate forecasting

Table 2: Evaluating Criterion of R^2

Performance	Positive correlation	Negative correlation
Scarce	-0.09 ~ 0.0	0.0 ~ 0.09
Weak	-0.3 ~ -0.1	0.1 ~ 0.3
Medium	-0.5 ~ -0.3	0.3 ~ 0.5
Strong	-1.0 ~ -0.5	0.5 ~ 1.0

III. RESULTS AND DISCUSSION

3.1. Data Set

Data are collect from Taiwanese cosmetic industrial which are accumulated by department of statistics, ministry of economic affairs, Republic of China. In the data base, we find 396 records each containing three components for the input vector and one output vector (sales). Fig.3. descripts the time series from January 1982 to December 2014. We can see the sales have significant increase but fluctuation between 1982 and 2000. After 2005, a substantial decrease happened until 2010 and rise slowly before 2015. Moreover, from the historical data and calendar, we find that the sales of February always decline suddenly. The reason might be consumer behavior change on Chinese New Year, this problem are ignored in this paper. To verify the normality of data sets is also important for data analysis. Fig.4. demonstrates the distribution of adopted data and it seen likes a straight line, namely it is follow normal distribution.

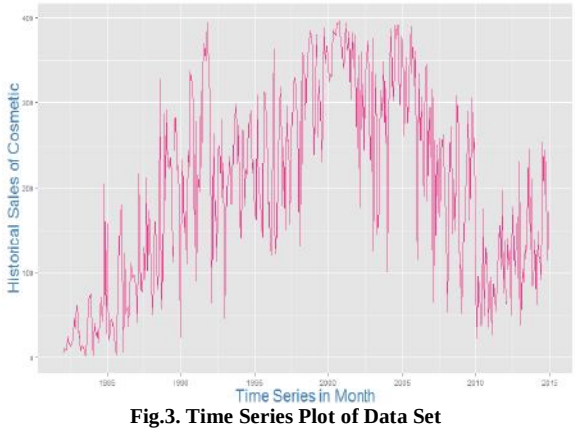


Fig.3. Time Series Plot of Data Set

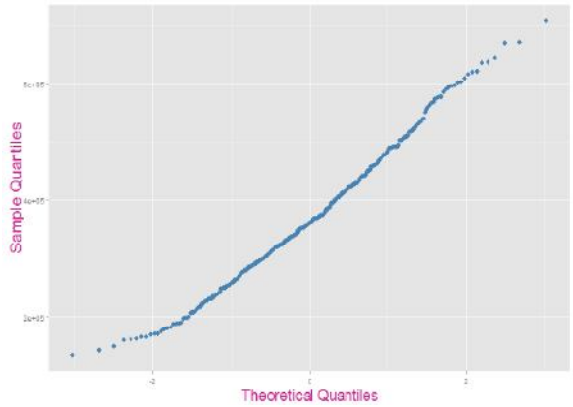


Fig.4. Normality of Data Set

3.2. Results from BPNN model and Actual Values

The implementing process is to forecast the sales in 2015 with historical data which are accumulated from 1982 to 2014. For the variables of BPNN, we find three types of cosmetic, emulsion cosmetic, soap cosmetic, cleaning cosmetic can be employed as input vector and the predicted sales have been selected as output vector. In the network, two hidden layer and each contain six nodes has been adopted. The results of BPNN was compared with actual values and shown as Fig.5. , we can see the line of BPNN is fluctuant and sometimes is very close to the line of actual values. For example, an overlapping happens in February and May. Furthermore, it is very near to actual values in June, September, and October.

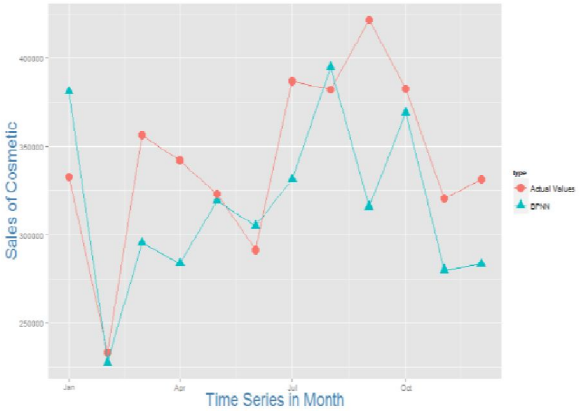


Fig.5. Forecasting Values of BPNN Model versus Actual Values

3.3. Results from LSSVM model and Actual Values

The input and output vectors of LSSVM model are same as the variables of BPNN model. Furthermore, the Gaussian Radial basis function kernel is selected as kernel of model. The results of LSSVM was compared with actual values and demonstrated in Fig.5. , the line of LSSVM is also fluctuant but always close to the line of actual values, an example are given from overlapping occurs in June. Furthermore, it is very near to actual values in January, May, July, November and December. In average, it seems a little better than BPNN model.

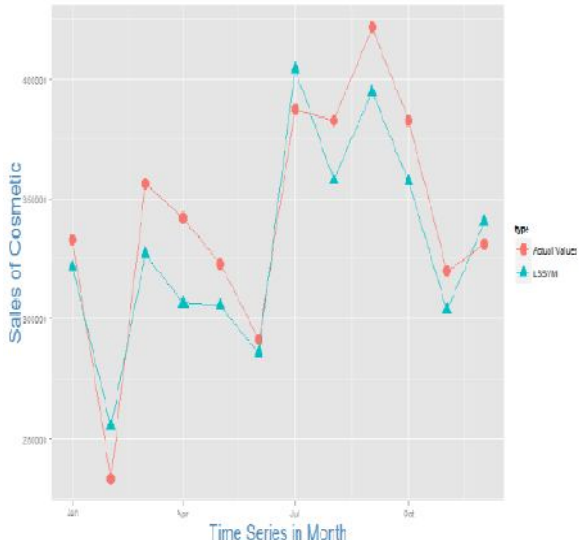


Fig.6. Forecasting Values of LSSVM Model versus Actual Values

3.4. Results from AR(1) model and Actual Values

According to the time series chart (Fig.3.), we can understand that the trend of sales changing by time. Before discussing the forecasting results of AR(1) model, we must judge the statationary of time series with auto correlation coefficient (ACF) and verify the auto correlation of AR(1) with the scatter plot. Firstly, in Fig.7., an auto correlation chart has been represented which revealed the time series of adopted data is stationary. Moreover, base on Box and Jenkins (1970) who is the original authors of ARIMA model, their study indicated that if ACF under 0.5, adopt the auto regressive model to forecast will become very inaccurracy. The ACF of this series is equal to 0.7 and satisfy the condition (Over 0.5). Secondly, the auto correlation scatter plot has been displayed in Fig.8. X axis is Y_{t-1} as explanatory variable which meaning the previous term of Y_t, Y axis is Y_t as dependent variable which meaning the current term. Finally, the results of AR(1) was compared with actual values and displayed in Fig.9. we can see the line of AR(1) model is also fluctuant and never close to the line of actual values, thus we can infer that the AR(1) is worse than the others model in this data set. The reason might be the stationary is not enough and maybe adopting the AR(2) or higher order model can solve this problem.

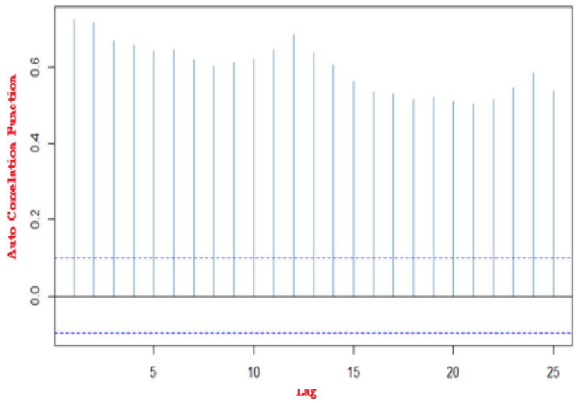


Fig.7. Auto Correlation Chart

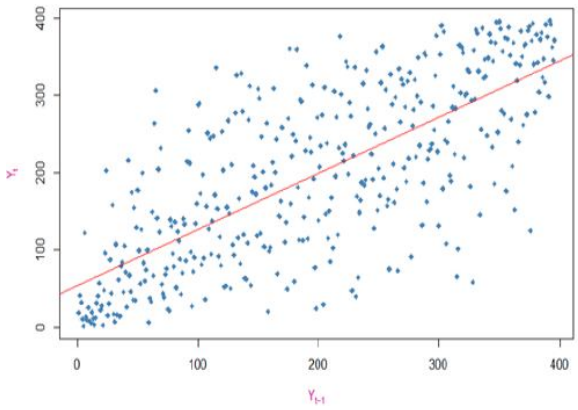


Fig.8. Scatter Plot between Y_t and Y_{t-1}

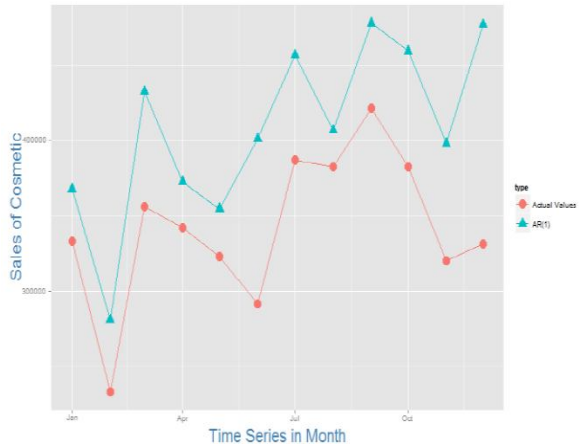


Fig.7. Forecasting Values of AR(1) Model versus Actual Values

3.5. Comparison between Proposed Models

If we put all of the predicted values together and compare with actual values, then we can get Fig.9. It seems that the LSSVM model provides the greatest accuracy between proposed models in average. To verify this inference, there are various methods are available for evaluating model performance. Table 3 represents the statistical measures and performance criteria for each forecasting model. It appears that the conclusion can be summarized. For the error evaluations criteria (such as: Me, MSE, and MAPE), the LSSVM model has minimal error. Furthermore, in the performance comparison, the LSSVM model has maximal R^2. Base on the results of analysis, we can find that the LSSVM is more accurate than the others adopted model.

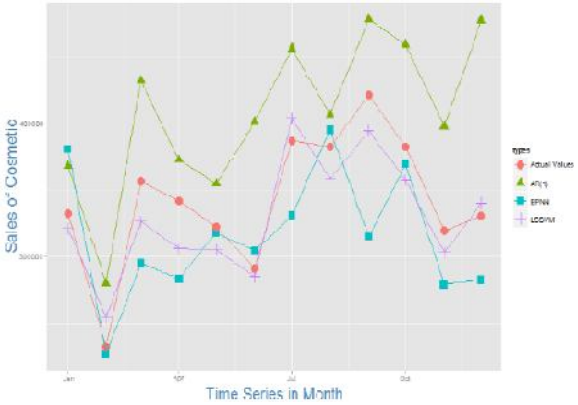


Fig.9. Predicted Values of Proposed Models Compare with Actual Values

Table 3: Comparison between Proposed Models with Criteria

Criteria	BPNN		LSSVM		AR(1)
	Training set	Test sets	Training set	Test sets	Data Sets
ME	28475	26261	19468	17132	56259
MSE	9.33×10^8	3.17×10^8	4.38×10^8	2.94×10^8	6.11×10^8
MAPE	21.78%	12.56%	12.81%	9.21%	17.13%
R^2	0.7837	0.8628	0.8512	0.9084	0.8076

CONCLUSIONS

The objective of this paper is to compare machine learning and time series approaches for predicting cosmetic sales data in terms of the prediction error and R^2. According to the analyses, we conclude that the least square support vector machine (LSSVM) provides the most preferable prediction benefits, namely, the minimal forecasting error and maximal performance. For the further studies, we suggest to design an experiment to discuss the error measuring and visualizing with response surface method (RSM) to achieve better results.

ACKNOWLEDGMENTS

This work was partially supported by the National Science Council of Taiwan under the grant number: MOST104-2221-E-151-018-MY3.

REFERENCES

[1] P. Langley, "Elements of Machine Learning", Morgan Kaufmann Publish, San Francisco, CA, USA, 1995.
[2] T.M. Mitchell, "Machine Learning", McGrawHill, New York, USA, 1997.
[3] E. Alpaydin, "Introduction to Machine Learning", Cambridge, MA: The MIT Press, 2004.
[4] M. Sewell, "Machine Learning", University College London, UK, 2006.
[5] B. Lantz, "Machine Learning with R", PACKT Publish, Birmingham, UK, 2013.
[6] I.C. Yeh, "Modeling of Strength of High Performance Concrete Using Artificial Neural Networks", Cement and Concrete Research, Vol. 28, No. 12, pp. 1797–1808, 1998.
[7] K. Kandananond, "Applying 2^k Factorial Design to Assess the Performance of ANN and SVM Methods for Forecasting Stationary and Non-stationary Time Series", Procedia Computer Science, Vol. 22, pp. 60–69, 2013.

[8] W.S. Mcculloch, and W. Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity", Bulletin of Mathematical Biophysics, Vol. 5, pp. 115–133, 1943.
[9] T. Fukuda, and T. Shibata, "Theory and Applications of Neural Networks for Industrial Control Systems". IEEE Transactions on Industrial Electronics, Vol. 39, No. 6), pp. 472–489, 1992.
[10] P.J. Werbos, "Back Propagation through Time: What It Does and How to Do It". Proceedings of the IEEE, Vol.78, No.10, pp. 1550–1560, 1990.
[11] F.P. Fang, and X.F. Fang, "Multivariate Forecasting Mode of Guangdong Province Port Throughput with Genetic Algorithms and Back Propagation Neural Network". Procedia-Social and Behavioral Sciences, Vol. 96, pp. 1165–1174, 2013.
[12] V. Vapnik, and C. Cortes, "Support-Vector Network", Machine Learning, Vol.20, pp. 273–297, 1995.
[13] J.A.K. Suykens, and J. Vandewalle, "Least Squares Support Vector Machine Classifiers", Neural Processing Letters, Vol. 9, No.3, pp. 293–300, 1999.
[14] M. Ghaedi, M.R. Rahimi, A.M. Ghaedi, I. Tyagi, S . Agarwal, and V.K. Gupta, "Application of Least Squares Support Vector Regression and Linear Multiple Regression for Modeling Removal of Methyl Orange onto Tin Oxide Nanoparticles Loaded on Activated Carbon and Activated Carbon Prepared from Pistacia Atlantica Wood", Journal of Colloid and Interface Science, Vol. 461, pp. 425–434, 2016.
[15] J. Liu, and T.Y. Tang, "LS-SVM Based Substation Circuit Breakers Maintenance Scheduling Optimization" International Journal of Electrical Power and Energy Systems, Vol. 64, pp. 1251–1258, 2015.
[16] H. Zheng, and H.F. Lu, "A Least-Squares Support Vector Machine (LS-SVM) Based on Fractal Analysis and CIELab Parameters for the Detection of Browning Degree on Mango", Computers and Electronics in Agriculture, Vol. 83, pp. 47–51, 2012.
[17] G. Box, and G. Jenkins, "Time Series Analysis: Forecasting and Control", Holden Day Publish. San Francisco, LA, USA.1970.
[18] J. Hassan, "ARIMA and Regression Models for Prediction of Daily and Monthly Clearness Index", Renewable Energy, Vol. 68, pp. 421–427.
[19] C.Q. Yuan, S. F. Liu, and Z.G. Fang, "Comparison of China's primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM(1,1) model", Energy, Vol. 100, No. 1, pp. 384–390.
[20] B.Z. Zhu, and Y.M. Wei, "Carbon Price Forecasting with a Novel Hybrid ARIMA and Least Squares Support Vector Machines Methodology", Omega, Vol. 41, No. 3, pp. 517–524.

★★★