Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Procedia Computer Science 22 (2013) 60 – 69

Procedia
Computer Science

17th International Conference in Knowledge Based and Intelligent Information and Engineering Systems - KES2013

Applying 2^k Factorial Design to assess the performance of ANN and SVM Methods for Forecasting Stationary and Non-stationary Time Series

Karin Kandananond*

Rajabhat University Valaya-Alongkorn, Prathumthani, Thailand 13180

Abstract

The performance of artificial neural network (ANN) and support vector machine (SVM) method for forecasting time series data is still an open issue for discussions among many authors in the literature. Hence, the purpose of this study is to characterize the capability of these two methods under the autocorrelation structure of time series and the most appropriate model is chosen. In this research, the performance of ANN and SVM is compared with respect to the autoregressive integrated moving average (ARIMA) structure. Two classes of ARIMA models, ARMA (1, 1) and IMA (1, 1), are utilized to represent stationary and non-stationary processes while the performance index of each learning method is the forecasting errors computed after each learning cycle. In order to deliver the right conclusions, the statistical analysis is conducted and the conclusions are drawn by utilizing the factorial design of experiment. The results indicate that these two machine learning methods have a different performance under the specific scenario of autocorrelation. When processes are stationary, the ANN might be a better choice than the SVM method. However, it turns out to be that the SVM has obviously outperformed the ANN for non-stationary cases.

Keywords: Artificial Neural Network (ANN); Autoregressive Integrated Moving Average (ARIMA); Factorial Design of Experiment; Non-stationary Process; Stationary Process; Support Vector Machine (SVM)

1. Introduction

Time series forecasting is critical for improving the performance of predicting a product demand. For this reason, the accuracy of demand forecasting greatly improves the whole process of production planning, namely

* Corresponding author. Tel.: +66-2529-3829; fax: +66-2909-3036.
E-mail address: kandananond@hotmail.com.

doi:10.1016/j.procs.2013.09.081

scheduling, capacity planning, material requirement planning and inventory management. This study focuses on two types of forecasting methods, artificial neural network (ANN) and support vector machine (SVM). There are many studies regarding the performance comparison of these two machine learning methods. However, the benchmark results do not portray a clear picture of the superior method. Moreover, they also lead to the controversial arguments because these studies are dealing with the specific sets of data, e.g., sets of time series data from different sources. As a result, the implementation of these results might be valid for a case by case basis only.

2. Literature Review

Mostly, the performance of any forecasting models is assessed by calculating the forecasting errors while the models are trained or developed from the historical data. Among the popular methods are the artificial neural network (ANN) and support vector machine (SVM). These machine learning methods are widely used to forecast different types of data including time series. Consequently, there are many authors trying to assess the capability of these two methods in order to identify the superior method. According to the literature, the main tool is the empirical study and the data used are varied from the actual data to the simulated one. For the real-life data, Tay and Cao [1] had utilized the ANN and SVM methods to forecast the financial time series and the historical data was based on five real future contracts collected from Chicago mercantile market. Similarly, Huang, Nakamori and Wang [2] also deployed both methods to predict the stock price index of Japan, NIKKEI 225. Kim [3] had applied the ANN and SVM to forecast the stock price index by considering the prediction performance for both training data and holdout data. Moreover, these methods were also utilized to predict the corporate bankruptcy by Shin, Lee and Kim [4]. Another popular forecasting methods was the autoregressive integrated moving average (ARIMA) utilized in parallel with the ANN and SVM by Pai, Hong and Lin [5].

For a precise conclusion, the design of experiment methodology was carried out by Lela, Bajic and Jozic [6] to draw a statistical conclusion from the performance assessment of the ANN, SVM and regression analysis. These three methods were used to predict the surface roughness of a milling process and the central composite design (CCD) was deployed to select the most appropriate method to model the response.

In addition to the comparison of these methods, another interesting aspect of the study was the utilization of autocorrelation structure as a basis to compare the performance of different forecasting methods. Lachtermacher and Fuller [7] utilized the Box-Jenkins model to study the relationship between the autocorrelation (the lag components) and the complexity of ANN structure. According to their study, each lag of autocorrelation structure was deployed to represent a unit of input for the ANN. Hwang [8] conducted a study to assess the performance of ANN method when processes were stationary by using the ARMA model as a benchmark. The randomized complete block design (RCBD) method was deployed as an experimental design to carry out the data analysis. This study leads to the profound understanding of how the ANN performs at the different degree of autocorrelation.

In conclusion, the empirical study is the most popular approach used to compare the performance of different forecasting methods. However, a major flaw is that these studies are based on the historical data from different case studies. For this reason, the empirical results might be valid only for a specific case (a set of data used to construct or train a model) and might not be applied to different scenarios. Moreover, the number of studies regarding the relationship between data structure and the appropriate learning method is also limited. As a result, the design of experiment and a standard model with a certain structure tends to be an interesting approach to draw the conclusions. A certain type of the data structure representing the autocorrelated time series is the autoregressive integrated moving average (ARIMA) model which is a stochastic difference equation frequently deployed to model the autocorrelated observations. Box, Jenkins and Reinsel [9] suggested that a specific form of ARIMA model, i.e., ARIMA (1, 1, 0) or ARMA (1, 1), is an appropriate choice to represent a type of autocorrelation, stationary processes. On the other hand, non-stationary processes should be modelled by ARIMA (0, 1, 1) or IMA (1, 1).

In this study, the ANN and SVM method are benchmarked by comparing their performance head-to-head under the same set of historical data simulated by utilizing the standard data models, ARMA (1, 1) and IMA (1, 1), the subclasses of autocorrelation, to represent stationary and non-stationary processes. After both machine learning methods are applied to the observations, the forecasting errors are calculated to be the performance index of each method under different scenarios. Finally, the statistical design of experiment is used to analyze the forecasting errors and come up with the most suitable method for each scheme.

3. Methods

The forecasting methods used in this study are two machine learning methods, ANN and SVM, while the underlying structure of observations is based on the ARIMA model. The performance measurement of each method is computed as the mean absolute percentage error (MAPE) and the experimental design method used is the 2^k factorial design;

3.1. ANN

The development of ANN models is based on studying the relationship between input variables and output variables. Basically, the neural architecture consists of three or more layers, i.e., input layers, output layers and hidden layers as shown in Fig 1. The function of this network was described as follows:

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

where Y_j is the output of node j, f (.) is the transfer function, w_{ij} is 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.

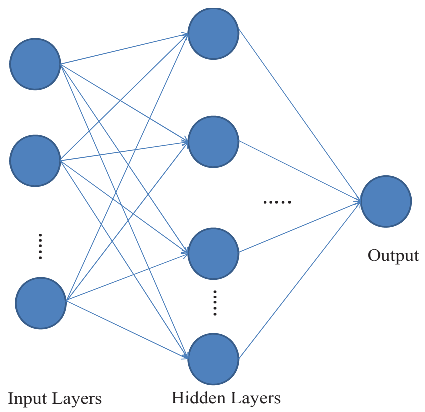

Fig. 1. Architecture of a neural network

3.2. SVM

SVM is a classification method which is based on the construction of hyperplanes in a multidimensional space. As a result, it allows different class labels to be differentiated. Normally, SVM is utilized for both classification and regression tasks and it is able to handle multiple continuous and categorical variables. The purpose of the regression task of SVM is to find a function f (such that y = f(x) + noise) which is able to predict new cases. This can be 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. There are two types of SVM models for the regression purpose, type 1 and 2. For regression type 1, the objective function is 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, the objective function of regression type 2 is

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

The regression type 2 also shares the same constraints as the regression type 1. For the SVM model, there are four types of kernels (ϕ), linear, polynomial, radial basis function (RBF) and sigmoid. Among these kernels, the RBF is the most frequently used kernel because of their localized and finite responses across the entire range of the real x-axis. The functions of these kernels are shown as follows:

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

4. Research Procedures

The analysis is categorized into two cases, stationary and non-stationary processes, according to the nature of autocorrelation. The stationary situation is represented by a class of an ARIMA model, ARMA (1, 1). After a number of data cases are simulated based on the ARMA (1, 1) equation while the IMA (1, 1) is utilized to model the non-stationary scenarios. The observations of a process are considered from period 1 to 100 (t = 1, 2, 3,…, 100) and the process output (Y_{t+1}) is equal to

$$Y_{t+1} = T + N_{t+1} \tag{2}$$

The source of autocorrelation is process disturbances, characterized by the ARIMA model, ARIMA (1, 1, 0) and ARIMA (0, 1, 1), as shown in (3) and (4):

$$N_{t+1} = \phi N_t + a_{t+1} - \theta a_t \text{ ; } -1 < \phi < 1, -1 < \theta < 1, \tag{3}$$

$$N_{t+1} = N_t + a_{t+1} - \theta a_t \text{ ; } -1 < \theta < 1, \tag{4}$$

The historical data at the time t-50, t-49,…, t-1 are utilized to predict the observation at time t while 50 training vectors are used to train the database. The number of training cycles used is 20,000 cycles and the forecasting error (MAPE : minimum average percentage error) is calculated. The structure of how ANN works for time series forecasting is shown in Fig 2.

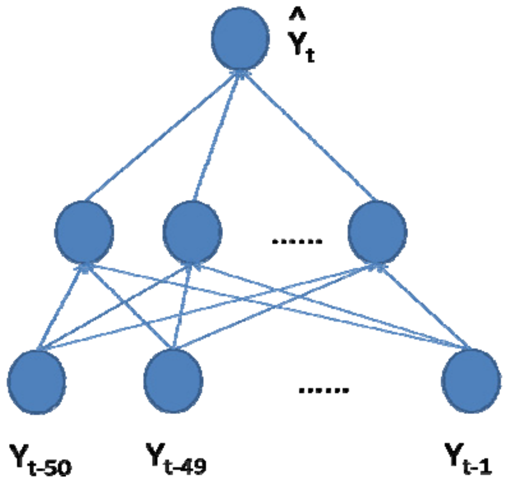

Fig. 2. ANN prediction

Afterwards, these results are compared empirically by utilizing 2^k factorial design. For stationary processes, three factors, A, B and C, are assigned to three influential factors, AR parameter (ϕ), MA parameter (θ) and methods used respectively while each factor is set at the low and high level as shown in Table 1. Similarly, all factors and their levels for the experimental study for non-stationary processes are shown in Table 2.

Table 1. Factors and levels for empirical study (stationary process)

Factor	Low	High
A (AR parameter; ϕ)	-0.9	0.9
B (MA parameter; θ)	-0.9	0.9
C (Types of Methods)	ANN	SVM

Table 2. Factors and levels for empirical study (non-stationary process)

Factor	Low	High
A (MA parameter; θ)	-0.9	0.9
B (Types of Methods)	ANN	SVM

The implementation of ANN and SVM method is possible by utilizing a statistical package, STATISTICA version 8, and the best algorithm for each learning method is automatically selected by the package. For the ANN, the neural network architecture used is the multilayer perceptron (MLP) and the training algorithm of the MLP network employed to build models is the Broyden-Fletcher-Goldfarb-Shanno (BFGS). Though the forecasting model based on the SVM approach is the regression type 1 with C=10.0, epsilon = 0.1 and the kernel is a radial basis function with gamma = 0.1.

5. Research Results

After each machine learning method is applied to forecast the time series data, the errors from each treatment (in term of MAPEs) are calculated and shown in Table 3. The design matrix is based on the 2^3 factorial with 5 replicates. The design of experiment package, Design Expert version 8, is utilized to analyze the empirical results. Before the analysis is finalized, the transformation (inverse) is deployed to ensure that the residuals satisfy all the independent and identically distributed (i.i.d.) conditions. After the transformation is performed, the analysis of variance (ANOVA) in Table 4 shows that the following factors; AR parameter:ϕ (A), MA parameter: θ (B) and forecasting methods (C), are the main factors affecting the forecasting errors (MAPE). Moreover, the results also point out that the interaction (ABC) does have a significant effect on the MAPE and ϕ has contributed the highest effect on the response followed by the method used and θ respectively.

Table 3. Forecasting errors for stationary case

Order	ϕ	θ	Method	MAPE
1	-0.9	-0.9	ANN	0.30259
2	-0.9	-0.9	ANN	0.2816
3	-0.9	-0.9	ANN	0.79831
4	-0.9	-0.9	ANN	1.65862
5	-0.9	-0.9	ANN	1.29721
6	0.9	-0.9	ANN	0.00992
7	0.9	-0.9	ANN	0.00959
8	0.9	-0.9	ANN	0.00971
9	0.9	-0.9	ANN	0.01055
10	0.9	-0.9	ANN	0.01087
11	-0.9	0.9	ANN	0.4072
12	-0.9	0.9	ANN	0.28205
13	-0.9	0.9	ANN	0.19434
14	-0.9	0.9	ANN	0.3145
15	-0.9	0.9	ANN	0.31658
16	0.9	0.9	ANN	0.01029
17	0.9	0.9	ANN	0.00896
18	0.9	0.9	ANN	0.01045

Table 3. Forecasting errors for stationary case (continued)

Order	ϕ	θ	Method	MAPE
19	0.9	0.9	ANN	0.00982
20	0.9	0.9	ANN	0.01042
21	-0.9	-0.9	SVM	0.3892
22	-0.9	-0.9	SVM	0.43917
23	-0.9	-0.9	SVM	1.0164
24	-0.9	-0.9	SVM	3.02584
25	-0.9	-0.9	SVM	4.06409
26	0.9	-0.9	SVM	0.0097
27	0.9	-0.9	SVM	0.00928
28	0.9	-0.9	SVM	0.009692
29	0.9	-0.9	SVM	0.010974
30	0.9	-0.9	SVM	0.011288
31	-0.9	0.9	SVM	0.47379
32	-0.9	0.9	SVM	0.27011
33	-0.9	0.9	SVM	0.20713
34	-0.9	0.9	SVM	0.80621
35	-0.9	0.9	SVM	0.50432
36	0.9	0.9	SVM	0.01128
37	0.9	0.9	SVM	0.0138
38	0.9	0.9	SVM	0.01276
39	0.9	0.9	SVM	0.01212
40	0.9	0.9	SVM	0.01273

Table 4. Analysis of variance

Source	SS	df	MS	F-Value	p-value
Model	86443.0445	7	12349.01	532.2085	< 0.0001
A-phi	85025.9473	1	85025.9473	3664.386	< 0.0001
B-theta	123.555063	1	123.555063	5.324886	0.0276
C-method	299.97986	1	299.97986	12.92831	0.0011
AB	256.141994	1	256.141994	11.03902	0.0022
AC	231.106188	1	231.106188	9.960046	0.0035
BC	255.902967	1	255.902967	11.02872	0.0023
ABC	250.411094	1	250.411094	10.79203	0.0025
Pure Error	742.506436	32	23.20333		
Total	87185.5509	39			

For the conclusions, the cube plot (Fig 3) representing the ABC interaction obviously illustrates that the

MAPEs resulted from the application of the ANN method are significantly lower than the ones from the SVM approach. This result signifies that the ANN method might be preferred to SVM when the process is stationary. Moreover, it also reveals the effect of the autocorrelation on the forecasting performance and it might be elaborated as follows; the forecasting errors at ϕ = +0.9 is significantly higher than the ones at ϕ = -0.9 no matter which type of methods are used. However, the result turns out to be different when ϕ is equal to 0.9 and the SVM method is utilized. Moreover, when θ is highly positive (θ = 0.9), the errors are much lower than the ones with highly negative value of θ (θ = -0.9) except the case of ϕ = 0.9 and method = SVM.

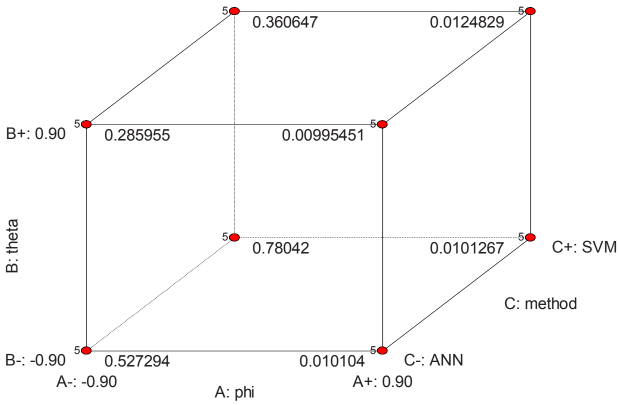

Fig. 3. Cube plot of the interaction ABC (MAPE).

For non-stationary case, the 2^k factorial design with 5 replicates is employed to study the relationship of all factors. The experimental results are shown in Table 5 while the ANOVA is illustrated in Table 6.

Table 5. Forecasting errors for non-stationary case

Order	θ	Method	MAPE
1	-0.9	ANN	0.094206
2	-0.9	ANN	0.09031
3	-0.9	ANN	0.08185
4	-0.9	ANN	0.13383
5	-0.9	ANN	0.11744
6	0.9	ANN	0.09149
7	0.9	ANN	0.088
8	0.9	ANN	0.1163
9	0.9	ANN	0.0891
10	0.9	ANN	0.09568
11	-0.9	SVM	0.020776

Table 5. Forecasting errors for non-stationary case (Continued)

Order	θ	Method	MAPE
12	-0.9	SVM	0.02426
13	-0.9	SVM	0.03913
14	-0.9	SVM	0.0221
15	-0.9	SVM	0.0372
16	0.9	SVM	0.06677
17	0.9	SVM	0.06258
18	0.9	SVM	0.07606
19	0.9	SVM	0.06662
20	0.9	SVM	0.06456

Table 6. Analysis of variance

Source	SS	Df	MS	F-Value	p-value
A-theta	0.008427	1	0.008427	14.33121	0.0016
B-Method	0.055196	1	0.055196	93.86492	<0.0001
AB	0.013818	1	0.013818	23.499	0.0002
Pure Error	0.009409	16	0.000588		
Cor Total	0.08685	19			

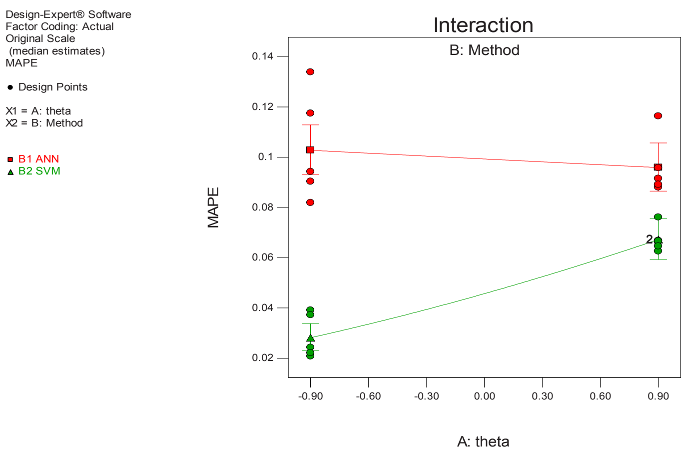

Fig. 4. Interaction plot AB (MAPE).

The interaction plot AB (Fig 4) signifies that the SVM method has outperformed the ANN for every values of theta. The difference of MAPEs when the SVM is used at the negative value of theta is significantly higher than the one at the positive value of theta. However, the gap difference is narrowed when θ is highly positive.

6. Conclusions

The ARIMA model is utilized to simulate the autocorrelation, both stationary and non-stationary processes. Two machine learning methods, ANN and SVM, are trained by a number of data and the forecasting errors are calculated as a mean to measure the performance of these methods. In order to achieve the conclusions statistically, the 2^k factorial design of experiment was deployed to compare the performance of the ANN and SVM method. Two parameters of the ARIMA equations, θ and ϕ, are adjusted at different levels to study the effect of these factors on the performance of the two forecasting methods. For stationary case, the statistical analysis shows that the AR parameter (ϕ) has the highest effect on the forecasting capability of both learning methods. Overall, the ANN generally performs better than the SVM in almost every treatments of the experiment. When the process is non-stationary, it is observed that the SVM approach turns out to be the superior method over the ANN method. In conclusion, according to the empirical study, the best method for every scenarios does not exist but it is important to select the right method for the specific type of autocorrelation structure underlying the observations.

References

[1] Tay FEH, Cao L. Applications of Support Vector Machines in Financial Time Series Forecasting. *Omega* 2001; **29**: 309–317.

[2] Huang W, Nakamori Y, Wang, S-Y. Forecasting Stock Market Movement Direction with Support Vector Machine. *Comput.Oper. Res*. 2005: **32**: 2513-2522.

[3] Kim K-J. Financial Time Series using Support Vector Machines. *Neurocomputing* 2003: **55**: 307–319.

[4] Shin KS, Lee TS, Kim HJ. An Application of Support Vector Machines in Bankuptcy Prediction. *Expert. Syst. Appl*. 2005: **28**: 127-135.

[5] Pai PF, Hong WC, Lin CS. Forecasting Electric Load by Support Vector Machines with Genetic Algorithms. *J. Adv. Comput. Intell. Informat.* 2005: **9**: 134-141.

[6] Lela B, Bajic D, Jozic S. Regression Analysis, Support Vector Machines, and Bayesian Neural Network Approaches to Modeling Surface Roughness in Face Milling. *Int. J. Adv. Manuf. Technol*. 2009: **42**: 1082–1088.

[7] Lachtemacher G and Fuller JD. Backpropagation in Time-series Forecasting. *J. Forecasting* 1995: **14**: 381–393.

[8] Hwang HB. Insights into Neural-Network Forecasting of Time Series Corresponding to ARMA (p, q) Structures. *Omega* 2001; **29**; 273-289.

[9] Box GEP, Jenkins GM, Reinsel GC. *Time Series Analysis*. New York: John Wiley & Sons; 2011.