Journal of Computations & Modelling, vol.6, no.2, 2016, 67-84
ISSN: 1792-7625 (print), 1792-8850 (online)
Scienpress Ltd, 2016

Applying Neural Network for the Improvement of Forecasting Accuracy –Airlines Weekly Cargo Data Case–

Yuki Higuchi[1], Yuta Tsuchida[2], Tatsuhiro Kuroda[3], and Kazuhiro Takeyasu[4]

Abstract

In recent years, severe competition is executed both on getting air passengers and those of air cargos. The forecast of the number of taking-off and landing is expanding, while demand for air cargo is decreasing. Strict marketing is required in such fields. Forecasting the trend of air cargo is an essential item to be investigated in airlines. In order to make forecast for time series, the method of using linear model is often used. Forecasting using neural network is also developed. Reviewing past researches, there are many researches made on this. There is many room to improve in neural network, therefore we make focus on

[1] Faculty of Business Administration, Setsunan University.
E-mail: y-higuch@kjo.setsunan.ac.jp
[2] Osaka Prefecture University. E-mail: soil.y.paddy@gmail.com
[3] Osaka Prefecture University. E-mail: tkuroda@jp.qatarairways.com
[4] College of Business Administration, Tokoha University.
E-mail: takeyasu@fj.tokoha-u.ac.jp

Article Info: *Received* : March 12, 2016. *Revised* : April 17, 2016.
Published online : June 10, 2016.

them. We use time series data, and in order to make forecast, a new coming data should be handled and the parameter should be estimated based upon its data. This is a so-called on-line parameter estimation. In this paper, neural network is applied and Multilayer perceptron Algorithm is newly developed. The method is applied to the Airlines Cargo Data in the case of Weekly data. When the data do not behave regularly, we have to device a new method for the neural networks to learn the past data much more. Repeating the data into plural sections, we could make a neural network to learn the past data much more. The result is compared with the method of ARIMA model. Good results were obtained. The new method shows that it is useful for the time series that has various trend characteristics and has rather strong seasonal trend. The effectiveness of this method should be examined in various cases.

Mathematics Subject Classification: 92B20

Keywords: forecasting; neural network; time series analysis; ARIMA model

1 Introduction

In industry, how to make a correct forecasting such as sales forecasting is a very important issue. If the correct forecasting is not executed, there arise a lot of stocks and/or it also causes lack of goods. Time series analysis, neural networks and other methods are applied to this problem. There are some related researches made on this. Reviewing past researches, Kimura et al. (1993)[1] applied neural networks to demand forecasting and adaptive forecasting method was proposed. Baba et al. (2000) [2] combined neural networks and the temporal difference learning method to construct an intelligent decision support system for dealing stocks. Takeyasu et al. (2009)[3] devised a new trend removing method and imbedded a theoretical solution of exponential smoothing constant. As a whole, it can be said that an application to sales forecasting is rather a few.

In this paper, neural network is applied and Multilayer perceptron Algorithm is newly developed. The method is applied to the Airlines Cargo Data in the case of weekly data. When the data do not behave regularly, we have to device a new method for the neural networks to learn the past data much more. Repeating the data into plural sections, we could make a neural network to learn the past data much more. The result is compared with the method of ARIMA model. Good results were obtained.

2 The method for Neural Networks[2]

In this section, outline of multilayered neural networks and learning method are stated. In figure l, multilayered neural network model is exhibited. It shows that it consist of input layer, hidden layer and output layer of feed forward type. Neurons are put on hidden layer and output layer. Neurons receive plural input and make one output.

Now, suppose that input layer have input signals $x_i (i = 1,2, \dots, l)$, hidden layer has m neurons and output layer has n neurons. Output of hidden layer $y_j (j = 1,2, \dots, m)$ is calculated as follows. Here $x_0 = -1$ is a threshold of hidden layer.

$$y_j = f\left(\sum_{i=0}^{l} v_{ij} x_i\right) \tag{1}$$

$$f(x) = \frac{1}{1 + \exp(-x)} \tag{2}$$

When v_{ij} is a weighting parameter from input layer to hidden layer and (2) is a sigmoid function. $y_0 = -1$ is a threshold of output layer and has the same value in all patterns. The value of the neuron of output layer, $z_k (k = 1,2, \ldots, n)$ which is a final output of network, is expressed as follows.

$$z_k = f\left(\sum_{j=0}^{m} w_{jk} y_j\right) \tag{3}$$

When w_{jk} is a weighting parameter of Hidden layer through Output layer, Learning is executed such that v, w is updated by minimizing the square of "output –supervisor signal". Evaluation function is shown as follows.

$$E = \frac{1}{2}\sum_{k=0}^{n} (d_k - z_k)^2 \tag{4}$$

where d_k is a supervisor signal. Error signal is calculated as follows.

$$e_k = d_k - z_k \tag{5}$$

Δw_{jk}(Output layer) is calculated as follows.

$$\delta_k = e_k z_k (1 - z_k) \tag{6}$$

$$\Delta \mathrm{w}_{jk} = \eta y_j \delta_k \tag{7}$$

Therefore, weighting coefficient is updated as follows.

$$w_{jk}^{\mathrm{new}} = w_{jk}^{\mathrm{old}} + \Delta \mathrm{w}_{jk} \tag{8}$$

where η is a learning rate.

Δv_{ij} (Hidden layer) is calculated as follows.

$$\gamma_j = y_j(1 - y_j) \sum_{k=1}^{n} w_{jk}^{\text{new}} \delta_k \tag{9}$$

$$\Delta v_{ij} = \eta x_i \gamma_j \tag{10}$$

v_{ij} is updated as follows.

$$v_{ij}^{\text{new}} = v_{ij}^{\text{old}} + \Delta v_{ij} \tag{11}$$

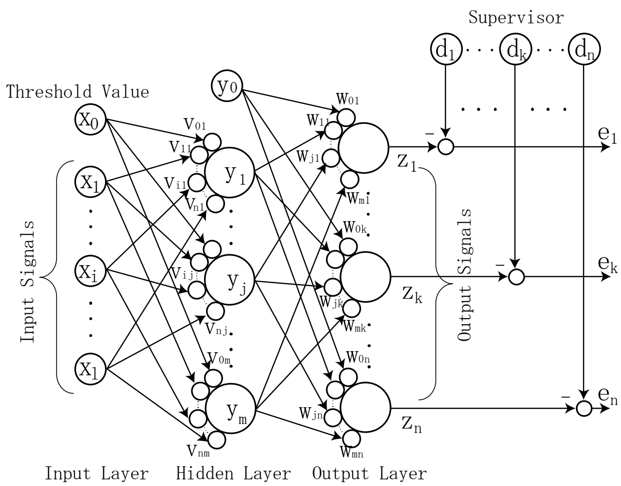

Figure 1: Multilayered neural network

3 An Application Method to the Time series

Now we apply neural networks to the forecasting of time series. Suppose there are M months' time series data. We use them as follows: Latter half N months' data for test, the first half $(M - N)$ months' data for learning.

3.1 Normalization

Data should be normalized because output is controlled by sigmoid function. We use time series this time, therefore data is normalized in the range [0:1]. We obtained max, min from 1 through $(M-N)$ months, which is a learning data period. We cannot grasp the test data range as the time of learning. Therefore estimated values $\widehat{\max}$, $\widehat{\min}$ are calculated as follows.

$$\widehat{\max} = \max \cdot \mu_{\max} \tag{12}$$

$$\widehat{\min} = \frac{\min}{\mu_{\min}} \tag{13}$$

Where $\mu_{\max}$, $\mu_{\min}$ are margin parameters. Set a_k as time series data, then a_k is normalized as follows.

$$X_k = \frac{a_k - \widehat{\min}}{\widehat{\max} - \widehat{\min}} \tag{14}$$

3.2 Forecasting method

Forecasting is executed as follows.

$$\hat{X}_k = F(X_{(k-l)}, X_{(k-l+1)}, \ldots, X_{(k-l+i)}, \ldots, X_{(k-1)}) \tag{15}$$

Where $F(x)$ is a neural network and X_k is a kth month's data (input signal). The number of learning patterns is $(M-N)-l$. We vary l as $l = 1,2,\ldots,(M-N/2$. The relation of learning data and supervisor data is shown as Figure 2. In this figure, input data is shown by the broken line when X_8 is targeted for learning under l=4. Learning is executed recursively so as to minimize the square of $\hat{X}_k - X_k$, where $\hat{X}_k$ is an output.

$$\left(\hat{X}_k - X_k\right)^2 \to \varepsilon \tag{16}$$

This time, ε is not set as a stopping condition of iteration, but predetermined s steps are adopted for the stopping condition. Forecasted data $\hat{a}_k$ is reversely converted to Eq.(17) from Eq.(14) as follows.

$$\hat{a}_k = \hat{X}_k\left(\widehat{\max} - \widehat{\min}\right) + \widehat{\min} \tag{17}$$

3.3 Forecasting accuracy

Forecasting accuracy is measured by the following "Forecasting Accuracy Ratio (FAR)".

$$\mathrm{FAR} = \left\{1 - \frac{\sum_{k=M-N}^{N} |a_k - \hat{a}_k|}{\sum_{k=M-N}^{N} a_k}\right\} \cdot 100 \tag{18}$$

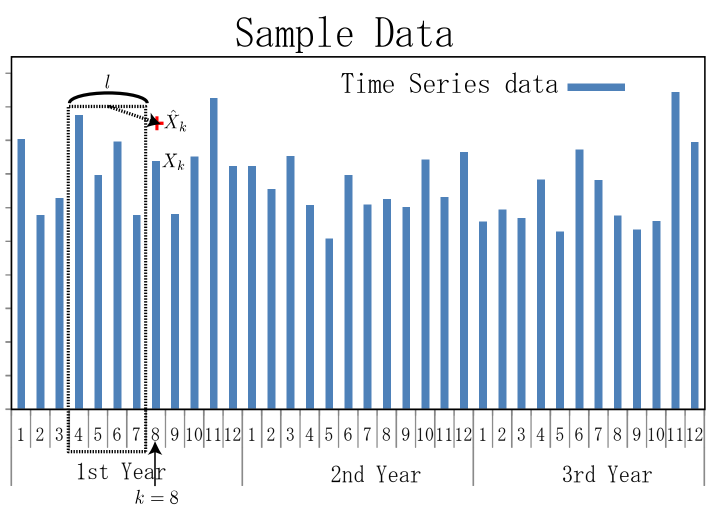

Figure 2: Choose the input data and supervisor for neural network

(ex: $l = 4, k = 8$)

4 A Newly Proposed Method

We have found that the mere application of neural networks does not bear good results when there is a big change of the data. Therefore we have devised a new method to cope with this. Repeating the data into plural section, we aim to make a neural network learn more smoothly. The concept of the change of data sampling is exhibited in Figure 3. Data is repeated τ times and after the learning, the value is taken average by τ in order to fit for the initial condition.

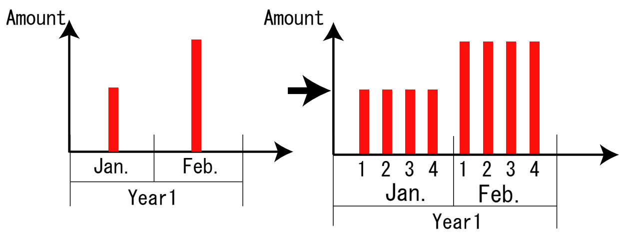

Figure 3: Change the time sampling (ex: $\tau = 4$)

5 ARIMA Model[6]

Using the delay operator Z^{-1} which means

$$Z^{-1}x_t = x_{t-1} \tag{19}$$

Define

$$A(Z^{-1}) = 1 + a_1 Z^{-1} + a_2 Z^{-2} + \cdots + a_p Z^{-p} \tag{20}$$

and

$$B(Z^{-1}) = 1 + b_1 Z^{-1} + b_2 Z^{-2} + \cdots + a_q Z^{-q} \tag{21}$$

then, ARMA model

$$x_t + a_1 x_{t-1} + \cdots + a_p x_{t-p} = e_t + b_1 e_{t-1} + \cdots + b_p e_{t-p} \tag{22}$$

is stated as

$$A(Z^{-1})x_t = B(Z^{-1})e_t \tag{23}$$

(p, d, q) order ARIMA model of d times differences from the original data is stated as

$$A(Z^{-1})(1 - Z^{-1})^d x_t = B(Z^{-1})e_t \tag{24}$$

The order of ARIMA model is determined by calculating AIC. AIC is calculated as follows.

$$AIC = -2lnL + 2(p + d + q) \tag{25}$$

where L is a Likelihood Function.

6 Numerical Example

6.1 Used Data

The JAL cargo data for 3 cases from 3rd October 2011 to 29th September 2012 were analyzed. These are the data of the Airline Z on Flight A from Osaka to Middle East, Flight B from New York to Middle East and Flight C from Manchester to Middle East. Here $M = 50$. First of all, graphical charts of these time series data are exhibited in Figure 4, 5 and 6.

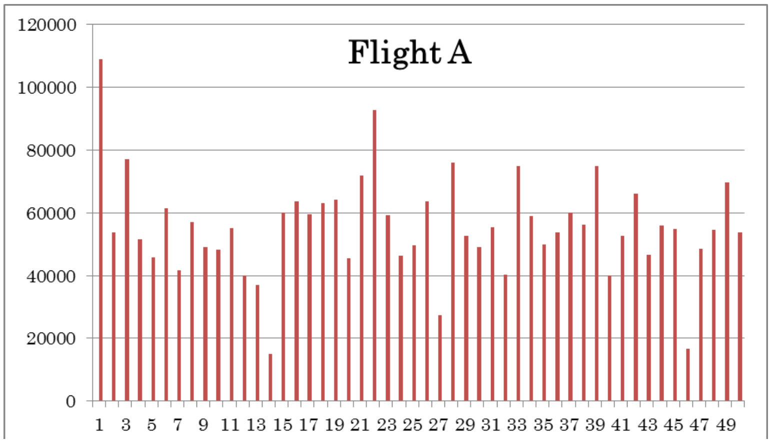

Figure 4: Original data of Flight A

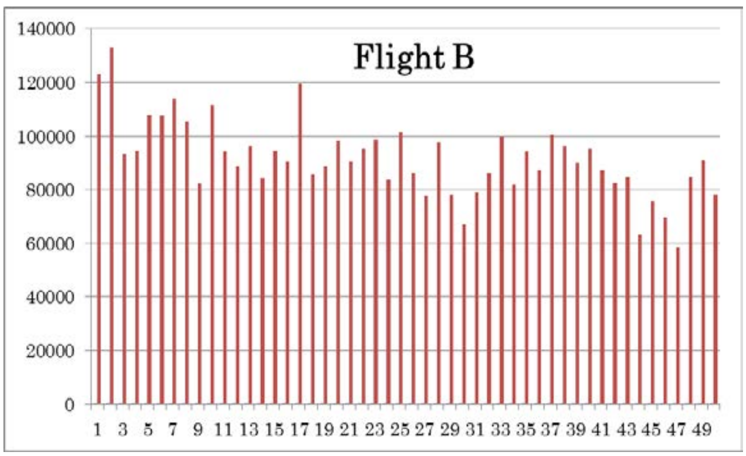

Figure 5: Original data of Flight B

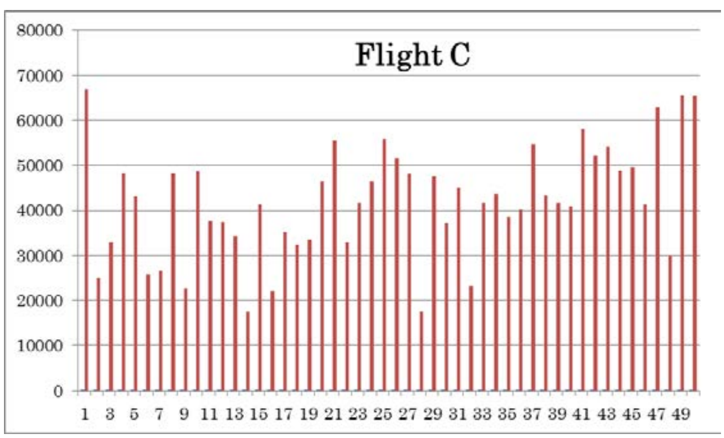

Figure 6: Original data of Flight C

Latter half data $N = 15$ are the data for test and the first half 35 data are the data for learning. $\mu_{\max}$ and $\mu_{\min}$ are set as follows.

$$\mu_{\max} = 1.1 \tag{26}$$

$$\mu_{\min} = 1.5 \tag{27}$$

Each maximum, minimum and estimated maximum, minimum data are exhibited in Table 1.

Table 1: The maximum value and the minimum value

Data	1 to 35 months	Estimated
	Maximum	
	Minimum	
Flight A	108,965.0	119,861.5
	15,060.0	10,040.0
Flight B	133,218.0	146,539.8
	58,412.0	44,672.0
Flight C	66,815.0	73,496.5
	17,560.0	11,706.7

6.2 Condition of Experiment

Condition of the neural network's experiment is exhibited in Table 2. Experiment is executed for 10 patterns $(l = 1, 2, \ldots, 10)$ and the Forecasting Accuracy Ratio is calculated based on the results.

Table 2: The experiment of neural network

Name	Parameter	Value
The number of neurons in hidden layer	m	$2, 4, 8, 16$
The number of output	n	1
The learning rate	η	0.035
Learning steps	s	1000

6.3 Experimental results for $\tau = 1$ and $\tau = 8$

Now, we show the experimental results executed by the method stated in 3.2. The Forecasting Accuracy Ratio is exhibited in Table 3 through 5. Minimum score among 8 cases is painted in color for each case. In Flight A, the case $\tau = 1$ was the best. In Flight B, the case $\tau = 4$ was the best. In Flight C, the case $\tau = 2$ was the best. Forecasting results for each case are exhibited in Figures 7 through 9.

Table 3: The result for Neural network of Flight A

Forecasting Accuracy Ratio							
$\tau = 1$	$\tau = 2$	$\tau = 3$	$\tau = 4$	$\tau = 5$	$\tau = 6$	$\tau = 7$	$\tau = 8$
84.59	84.19	82.31	78.94	79.65	78.24	76.59	77.31
l=9 *m=16*	*l=1* *m=4*	*l=3* *m=2*	*l=6* *m=4*	*l=2* *m=2*	*l=5* *m=4*	*l=7* *m=16*	*l=5* *m=4*

Table 4: The result for Neural network of Flight B

Forecasting Accuracy Ratio							
$\tau = 1$	$\tau = 2$	$\tau = 3$	$\tau = 4$	$\tau = 5$	$\tau = 6$	$\tau = 7$	$\tau = 8$
89.36	90.83	90.51	91,12	89.52	90.21	88.54	90.26
l=5 *m=8*	*l=3* *m=8*	*l=4* *m=4*	*l=7* *m=4*	*l=2* *m=2*	*l=4* *m=4*	*l=2* *m=4*	*l=2* *m=2*

Table 5: The result for Neural network of Flight C

Forecasting Accuracy Ratio							
$\tau = 1$	$\tau = 2$	$\tau = 3$	$\tau = 4$	$\tau = 5$	$\tau = 6$	$\tau = 7$	$\tau = 8$
81.94	82.91	82.75	82.17	81.42	82.33	81.3	81.54
l=4 *m=4*	*l=4* *m=2*	*l=4* *m=2*	*l=4* *m=2*	*l=2* *m=2*	*l=7* *m=2*	*l=7* *m=8*	*l=3* *m=2*

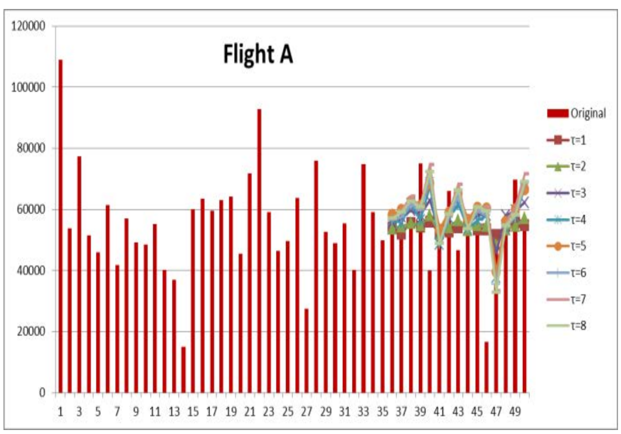

Figure 7: The result of Flight A

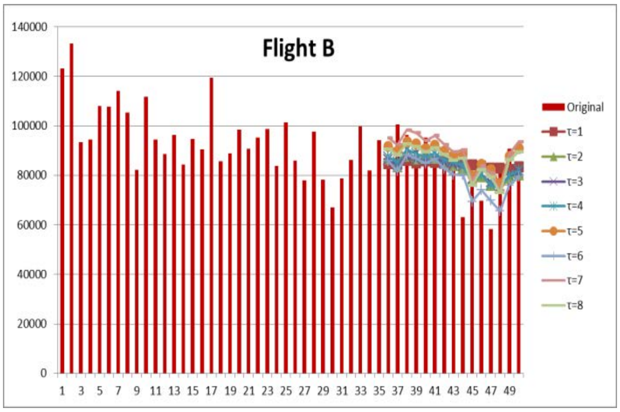

Figure 8: The result of Flight B

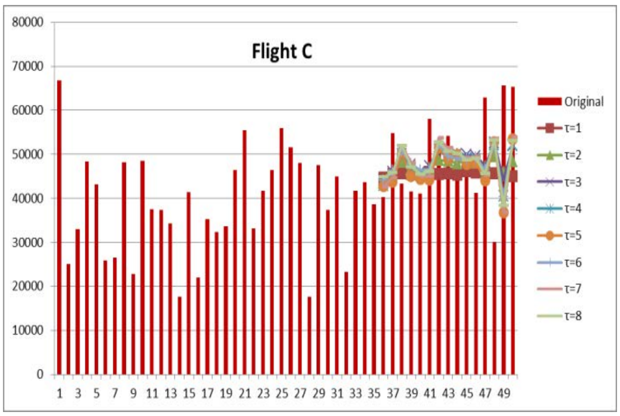

Figure 9: The result of Flight C

6.4 Forecasting Result of ARIMA

Forecasting Accuracy Ratio and the order of ARIMA model are exhibited in Table 6.

Table 6: The result for ARIMA: Forecasting Accuracy Ratio and its order

ARIMA	Flight		
	A	B	C
Forecasting Accuracy Ratio	80.11	87.70	83.54
Order	0,2,2	0,2,2	0,2,2

7 Remarks

Now, we compare with both results. In Table 7, both results are stated and compared. Their comparison is shown in Figure 10, 11 and 12.

Table 7: Comparison of the both results

	Forecasting Accuracy Ratio	
	Previous Method(ARIMA)	Proposed Method
Flight A	80.11	**84.59** $(\tau = 1, l = 9, m = 16)$
Flight B	87.70	**91.12** $(\tau = 4, l = 7, m = 4)$
Flight C	**83.54**	82.91 $(\tau = 2, l = 4, m = 2)$

In Flight A, the case $\tau = 1$ was the best. In Flight B, the case $\tau = 4$ was the best. In Flight C, the case $\tau = 2$ was the best. Next we compared the proposed method with ARIMA model.

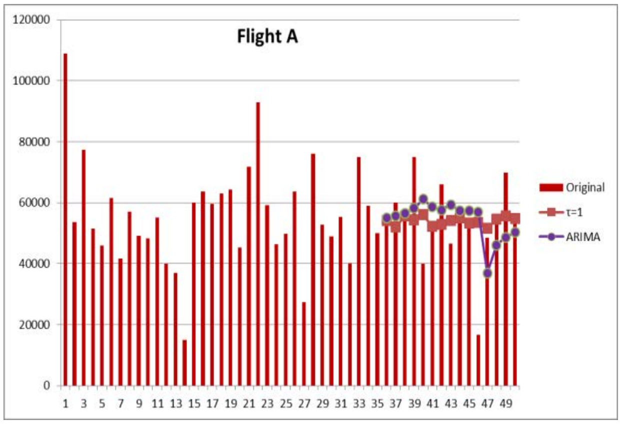

Figure 10: The result of Flight A: ARIMA used and $\tau = 1$

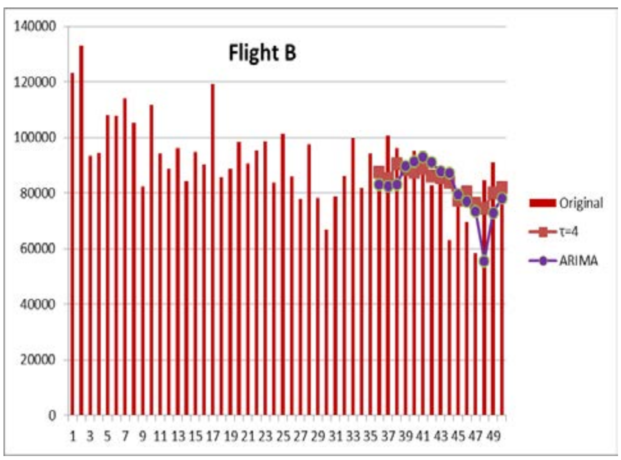

Figure 11: The result of Flight B: ARIMA used and $\tau = 4$

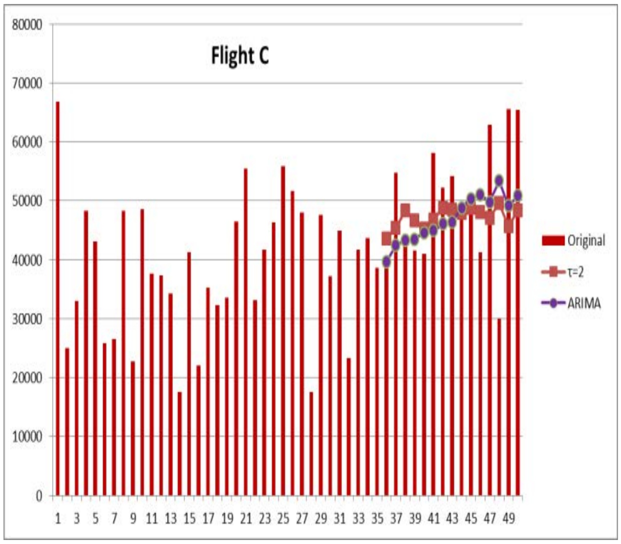

Figure 12: The result of Flight C: ARIMA used and $\tau = 2$

In two cases out of three, the newly proposed method had better forecasting results than those of ARIMA model. This means the proposed method is rather effective for these data.

8 Conclusion

In this paper, neural network was applied and Multilayer perceptron Algorithm was newly developed. The method was applied to the Airlines Passengers and Cargo Data in the case of weekly data. When the data do not behave regularly, we have to device a new method for the neural networks to learn the past data much more. Repeating the data into plural sections, we could make a neural network to learn the past data much more. The result was compared with the method of ARIMA model. In the numerical example, two cases out of three had a better forecasting accuracy than the ARIMA model. Various cases should be examined hereafter.

References

[1] Aritoshi Kimura, Ikuo Arizono and Hiroshi Ohta, An Application of Layered neural networks to demand forecasting, Japan Industrial Management Association, **44**(5), (1993).

[2] N. Babo and H. Suto, Utilization of artificial neural networks and the TD-learning method for constructing intelligent decision support systems

[3] Kazuhiro Takeyasu, Keiko Imura and Yuki Higuchi, Estimation of Smoothing Constant of Minimum Variance And Its Application to Shipping Data With Trend Removal Method, *Industrial Engineering and ManagementSystems*, 8(4), 257=263,2009.12

[4] Tatsuhiro Kuroda, Yuki Higuchi and Kazuhiro Takeyasu, A Hybrid Method to Improve Forecasting Accuracy Utilizing Genetic Algorithm –An Application to the Airlines Passengers and Cargo Data, *International Journal of Information Technology and Network Application* (IJITNA), Forthcoming.

[5] Edgar Sanchez-Sinencio, Clifford Lau, (editors), Artificial neural networks: Paradigms, applications, and hardware implementations, *IEEE press, Piscataway*, New Jersey, 1992.

[6] K. Takeyasu, Y. Ishii, Y. Higuchi and K. Nagata, Time Series Analysis-Forecasting and its Applications, *Izumi Syuppan*, (2012).