Airline In-flight Meal Demand Forecasting with Neural Networks and Time Series Models*

Young-Chan Lee**
E-mail : chanlee@family.sogang.ac.kr**

ABSTRACT

The purpose of this study is to introduce a more efficient forecasting technique, which could help result the reduction of cost in removing the waste of airline in-flight meals. We will use a neural network approach known to many researchers as the "Outstanding Forecasting Technique".

We employed a multi-layer perceptron neural network using a backpropagation algorithm. We also suggested using other related information to improve the forecasting performances of neural networks. We divided the data into three sets, which are training data set, cross validation data set, and test data set. Time lag variables are still employed in our model according to the general view of time series forecasting. We measured the accuracy of our model by "Mean Square Error" (MSE). The suggested model proved most excellent in serving economy class in-flight meals.

Forecasting the exact amount of meals needed for each airline could reduce the waste of meals and therefore, lead to the reduction of cost. Better yet, it could enhance the cost competition of each airline, keep the schedules on time, and lead to better service.

1. INTRODUCTION

The activities of all enterprises should be planned and performed on the basis of demand forecasting. The production of goods or service performed without forecasting could bring on many risks. Therefore, demand forecasting is probably the most fundamental factor in production planning and control. After forecasting is processed, the enterprise can aggregate production planning, capacity planning, scheduling, inventory planning and control, and etc.

Although it is very difficult to forecast under uncertain political, economic, and social changes, enterprises should first collect and analyze the demand data related to the market.

But forecasting is not always perfect and the real data might be different from that of the forecasted data. Like this, demand forecasting has the unfortunate problem in which we aren't sure how accurate the data is. Accuracy works as a decision factor in dispute over the necessity of forecasting. On one side, the most important factor which influences the results of forecasting

---

*) Please do not cite without authors' permission. This is only a draft.

**) Institute for Business Research, Sogang University, Seoul, Korea

is the selection of forecasting techniques. Even though many forecasting techniques are suggested, they may not all be suitable.

Specifically, most airline in-flight meal services are provided without forecasting the exact demand in need, therefore, may result in the delay of plane schedules and the waste of meals on account of false forecasting.

Meanwhile, Suh & Kwak's study[1998] presented a hybrid forecasting technique which combined both neural network and exponential smoothing as an excellent forecasting technique of time series data for airline in-flight meal demands. In addition, Suh & Kwak[1998] still used "Mean Absolute Error" (MAE) and "Mean Square Error" (MSE), which was the most general method in estimating the accuracy of forecasting technique, while they focused on reducing the error.

In this paper, we tried to show the improved forecasting performance in time series data when used with neural networks.

The remainder of this paper is organized as follows: In section 2, we reviewed on neural networks and found recent trends in forecasting. In section 3, we presented the methodology and the results. The last section is the conclusion.

2. LITERATURE REVIEW ON NEURAL NETWORKS

Recently, neural networks have been widely applied to various areas. It is shown that the accuracy of the model using neural networks has been quite excellent. We also tried to use neural networks as a forecasting technique.

Neural networks do have certain drawbacks in mathematical background, but it is still applied to many fields because of the variety in practical applications[Salchenberger et. al. 1992; Eddy et. al. 1993; Norman et. al. 1993; Chanda, 1994; Philipoom et. al. 1994; Jain et. al. 1995; Markham et. al. 1995]. It especially shows its excellence in time series forecasting which needs robustness against noise or variation[Gorr et. al. 1994; Hill et. al. 1996; Kim & Noh, 1996; Kwon & Golden, 1996; Jo, 1998; Suh & Kwak, 1998].

Up to now, many studies on neural networks had been published, but the robustness of neural networks did not reach one precise conclusion. In Kwon & Golden's study[1996], neural networks are shown that they are superior to traditional statistical models. Hill at al.[1996], Jain & Nag[1995], Markham & Ragsdale[1995], Philipoom, at al.[1994], and Salchenberger, at al.[1992] came up with the same result.

Kim & Noh[1996] conducted comparison studies between Box-Jenkins and neural networks. They showed that neural networks are superior to Box-Jenkins and that they had statistical significance. But it was also shown that there was no significant difference among other various neural network models.

On the other hand, Gorr at al.[1994] showed that neural networks are not significant with a statistical test.

But most of the results of these studies showed that the neural networks' accuracy depended on the number of nodes, sample size, and architecture of neural networks. Neural networks have some advantages in business problems. The first is its ease of application. Neural networks have the ability of internal pattern recognition by simple data learning process. The second is robustness. Even though data set is incomplete or distorted, neural networks can make good results using learning algorithm without additional data handling. The third is that it is useful to analyze nonlinear relations among input variables since neural networks consisting of numerous neurons are of parallel structures and nonlinear functions of higher degree.

In the real world, we meet nonlinear higher degree systems correlated with each input variable. Therefore, neural networks are profitable in modeling with complex and correlated variables.

Meanwhile, neural networks have some defects. First, if the data is not sufficient enough or there isn't proper learning function among the data, then a satisfied solution can't be found. Second, it is hard to explain why this result can be taken for the result of neural networks. The reason is that the result is generated from repetitive calculation of many weights and input patterns, and the weight is determined by complicated learning activity functions. Third, neural networks can use up much time and cost during the learning data selection, processing and analysis. Fourth, neural networks have problems in processing time. Connections between each node means multiplication, so the total processing time relates highly with the numbers of connection. Therefore, a small increase in unit numbers can cause a large increase in the total processing of time. But fortunately, these problems have been recently solved by developing application programs.

On one side, many studies show that neural networks are superior to other techniques when data is sufficient enough. This means that neural networks are inferior when there is lack of data. Therefore, in case of insufficient data for one reason or another, it is not recommended to use neural networks.

Classification problems using neural networks such as discriminant analysis, can be solved effectively by using highly correlated data with factor analysis. But in demand forecasting of time series data using neural networks, many researchers don't use this kind of highly correlated data but only time series data itself. Like this, most time series analyses related to neural networks simply focus on comparing two techniques. Kim & Noh[1996] also simply compare time series analysis with neural network models. Suh & Kwak[1998] showed the usefulness and application possibility of hybrid neural networks which combined both the neural network and traditional time series model.

Recently, the hybrid neural networks were considered as an effective model for the improvement of forecasting performance. Jo's study[1998] is a good example showing improvement of forecasting performances in neural networks. Jo[1998] designed a hybrid neural network model by adding virtual series to time series data. In this study, we focused on improving the forecasting performance by considering additional time series data as input variables.

So to speak, we are able to improve the performance of neural networks by using highly correlated data such as classification problems. Following this logic, we applied it to demand forecasting using time series data.

We compared and analyzed two cases of forecast results. One is considering highly correlated data to input variables, the other is not.

3. ANALYSIS OF DEMAND FORECASTING RESULTS

In this study, the data of airline in-flight meals from April 1^{st} to October 15^{th} in 1997 were used for forecasting. To improve the performances of demand forecasting using neural networks, we employed a multi-layer perceptron model that had backpropagation algorithm, and then we put 1 as the number of hidden layer.

In this study, we largely classified three kinds of models. The first model is called the 'simple neural network model', and it assumes that the past amount of airline meals in demand affects future one. The demand one days ago affects the day (lag 1), the demand two days ago affects the day (lag 2), the demand three days ago affects the day (lag 3), the demand one weeks ago affects the day (lag 7). Therefore, number of input node is 1 to 4.

The second model is called the 'hybrid neural network model'. It adds the forecasting data by exponential smoothing as an input node, which is superior to other forecasting technique such as Box-Jenkins in Suh & Kwak's study[1998]. Therefore, the number of the input node is from 2 to 5.

The third model is called the 'perfect neural network model'. It adds two other airlines, which is related to original airline data. We considered two airlines to America, and so add lag 1 data and forecasting data by exponential smoothing as input nodes. Therefore, each model has additional 4 input nodes to second model and the number of input node is from 6 to 9.

Considering time lag, we lost the data of 7 days. We used the data of 45 days as a test data of total data of 191 days. We randomized the rest of the data of 146 days for the reduction of data loss, and then, the data of 116 days was used as training data while the other data of 30 days were used as cross validation data. <Table 1> presents the description of the data as above mentioned.

<Table 1> Data description

Total data(198)	Usee data(191)	Training data	116
		Validation data	30
		Test data	45
	Excepted data(7)	Time lag	7

We processed 2000 iterations for model learning. In this process, we adopted stopped training (or early stopping, optimal stopping) for cross validation. After this training, forecasting results were produced by using weights that were calculated by activity function when training was stopped. We employed hyperbolic tangent (tanh) as an activity function for tanh is the most desirable transfer function[Brown et al. 1994]. This nonlinear transfer function is defined by:

$$\tanh(x) = \frac{1-e^{-2x}}{1+e^{-2x}}$$

Input layer has nodes of each independent variable and is connected to a hidden layer. It does not matter how many numbers of nodes there are in the hidden layer, but generally the # of nodes in hidden layer is less than or equal to twice the # of the input node. These hidden nodes are connected to output nodes and our model has one output node because we consider it one dependent variable. We considered that the # of nodes in the hidden layer is twice the # of input nodes according to general view.

We conducted an analysis using *Neurosolutions ver. 3.0* for demand forecasting with neural networks.

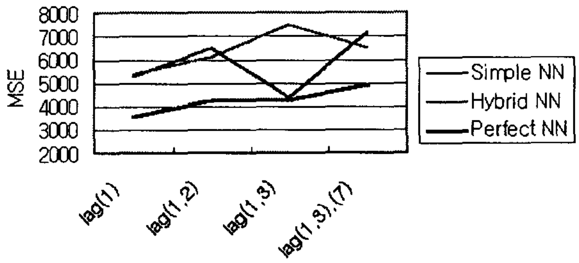
As you see in <Table 2> of airline UA807, the perfect neural network model (L model) has the least error during training epochs. It means that the weights calculated by the L model are the best. The L model contains input nodes of all time lag data, forecasting data by exponential smoothing and also contains lag 1 data and forecasting data using the same airline to America UA808 and UA 826 by exponential smoothing.

<Table 2> Comparison of neural network models at 807E

807E			Data Sets					
			Training Data		Validation Data		Test Data	
Models	Input Noes		Iterate	NMSE	Iterate	NMSE	MAE	MSE
Simple Neural Networks	lag(1)	A	2000	0.0947	1545	0.1190	59.0573	5259.7275
	lag(1-2)	B	2000	0.0688	549	0.1426	68.0232	6471.5474
	lag(1-3)	C	2000	0.0621	7	0.1415	**51.0659**	**4371.1245**
	lag(1-3), (7)	D	2000	0.0434	1510	0.1430	67.442	7110.2612
Hybrid Neural Networks	lag(1), **Exp.**	E	2000	0.0873	1879	0.1494	**54.1973**	**5361.0044**
	lag(1-2), **Exp.**	F	2000	0.0609	556	0.1435	64.2420	6125.2446
	lag(1-3), **Exp.**	G	2000	0.0505	1471	0.0990	65.0600	7464.6646
	lag(1-3), (7), **Exp.**	H	2000	0.0280	763	0.1123	63.5964	6517.3833
Perfect Neural Networks	lag(1), Exp., **Cor.**	I	2000	0.0181	5	0.1275	**45.9372**	**3552.6597**
	lag(1-2), Exp., **Cor.**	J	2000	0.0096	3	0.1294	51.1397	4264.7446
	lag(1-3), Exp., **Cor.**	K	2000	0.0069	5	0.1344	50.8001	4304.6714
	lag(1-3), (7), Exp., **Cor.**	L	2000	**0.0010**	6	0.1486	52.9493	4903.6475

It is necessary to divide data set as training data set, cross validation data set and test data set to minimize underfitting or overfitting problems. The training data set is for fitting data, cross validation data set for preventing data from underfitting and overfitting, and test data set for forecasting, respectively. We computed the cross validation error periodically during training, and stopped training when the cross validation error started to increase.

Test results using the weights of cross validation showed that I model had the least mean absolute error and mean square error in UA807. So perfect neural network model produced better forecasting performances than other models. Comparing three models in respect to the least error of UA807, forecasting accuracy has an order of perfect neural network model, simple neural network model, and hybrid neural network model. But comparing three models in respect to the same time lag, forecasting accuracy has an order of simple neural network models, hybrid neural network models, and perfect neural network models except for lag(1) and lag(1-3). As you see in <Figure 1>, forecast error is becoming smaller from simple neural network model → Hybrid neural network model → perfect neural network model.



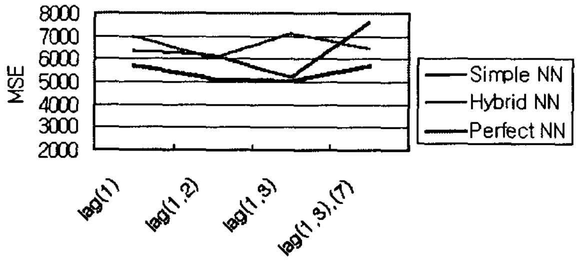
<Figure 1> Comparison of neural network models at each time lag in 807E

In the training process of airline UA808, as shown in <Table 3>, the prefect neural network model (K model) has the least forecasting error as similar to UA 807. K model contains input nodes of all time lag data except for lag 7, forecasting data by exponential smoothing, and also contains lag 1 data and forecasting data using the same airline to America UA807 and UA826 by exponential smoothing.

<Table 3> Comparison of neural network models at 808E

808E			Data Sets					
			Training Data		Validation data		Test Data	
Models	Input Nodes		Iterate	NMSE	Iterate	NMSE	MAE	MSE
Simple Neural Networks	lag(1)	A	2000	0.0493	3	0.0978	63.6558	6417.9395
	lag(1-2)	B	2000	0.0455	2	0.0982	62.7859	6195.9771
	lag(1-3)	C	2000	0.0368	2	0.0950	**57.3638**	**5232.5596**
	lag(1-3), (7)	D	2000	0.0263	1	0.0924	69.9925	7616.1089
Hybrid Neural Networks	lag(1), **Exp.**	E	2000	0.0479	1121	0.0977	66.8971	6960.6313
	lag(1-2), **Exp.**	F	2000	0.0444	10	0.0990	**61.9347**	**6033.6870**
	lag(1-3), **Exp.**	G	2000	0.0325	436	0.0981	67.9369	7093.3667
	lag(1-3), (7), **Exp.**	H	2000	0.0276	6	0.0915	64.7215	6448.6104
Perfect Neural Networks	lag(1), Exp., **Cor.**	I	2000	0.0068	491	0.0926	60.6225	5755.7935
	lag(1-2), Exp., **Cor.**	J	2000	0.0040	12	0.1048	**55.0114**	5114.5107
	lag(1-3), Exp., **Cor.**	K	2000	**0.0026**	19	0.0971	55.5418	**5038.7451**
	lag(1-3),(7), Exp., **Cor.**	L	2000	0.0028	133	0.1019	61.4292	5766.7134

Forecasting results using the weights of cross validation shows that the J model has the least mean absolute error and that the K model has the least mean square error in UA808. This means that perfect neural network model produces better forecasting results than other models. Comparing three models in respect to the least error in UA808, forecasting accuracy has an order of perfect neural network model, simple neural network model, and hybrid neural network model. But comparing the three models in respect to the same time lag, forecasting accuracy has an order of simple neural network models, hybrid neural network models, and perfect neural network models except for lag(1) and lag(1-3) like UA807. As you see in <Figure 2>, forecasting error decreases from the simple neural network model to the perfect neural network model.

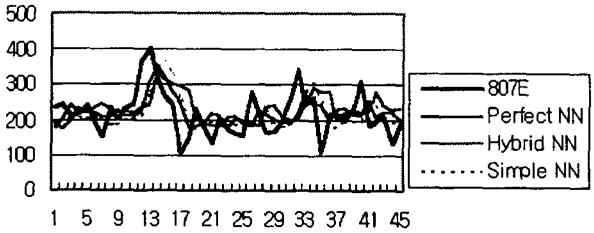


<Figure 2> Comparison of neural network models at each time lag in 808E

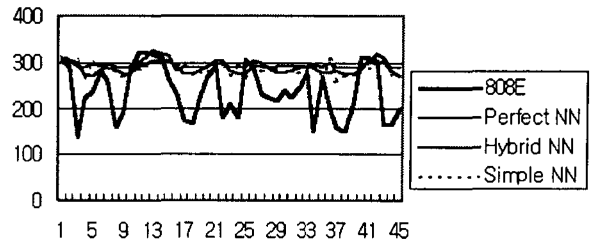
The summary of forecasting is presented in <Table 4>. As you see in the summary table, perfect neural network model is better than other neural network models in many cases. So we are able to forecast better by using various correlated data to original time series data. The comparison of forecasting results is presented in <Figure 3> for UA807, <Figure 4> for UA808, respectively.

<Table 4> Results of model comparison at each airline

Airlines	Comparison rule	Model comparison	Example model
807E	Minimum error	Perfect<Simple<Hybrid	I<C<E
	The same time lag	Perfect<Hybrid<Simple	J<F<B, L<H<D
808E	Minimum error	Perfect<Simple<Hybrid	K<C<F
	The same time lag	Perfect<Hybrid<Simple	J<F<B, L<H<D



<Figure 3> Comparison of forecasting results at 807E



<Figure 4> Comparison of forecasting results at 808E

4. CONCLUSIONS

Although neural networks have some advantages in time series forecasting problem by using highly correlated data, most studies on demand forecasting show only the difference of forecast error between neural networks and traditional time series analysis.

This study shows the possibility of application in neural networks using other highly correlated time series data. To forecast demand, we considered simple neural network model, hybrid neural network model adding forecasting data of other highly correlated time series data to simple neural network, and perfect neural network model which added lag 1 data and forecasting data by time series analysis to hybrid neural network model.

We compared and analyzed the accuracy of forecasting for each model. Through this study, several important contributions are provided as follows.

Firstly, Suh & Kwak's study[1998] showed that the hybrid neural network model is superior to the simple neural network model. In this study, however, we showed that the simple neural network model is superior to the hybrid neural network model in some case. When we considered the same time lag, forecast performance among three models depended on each combination of input variables. Mostly hybrid neural network models are superior to simple neural network models, and perfect neural network models are superior to hybrid neural network models.

Secondly, perfect neural network models are superior to other neural network models, so it is desirable to use forecasting technique adding various data correlated to original time series data. Last of all, in forecasting the exact amount of meals, each airline could reduce the waste of meals, and therefore, leads to the reduction of cost. Moreover, it enhanced the cost competitiveness of each airline, keeps the schedules on time, and leads to offering overall services of good quality.

All in all, we have the following limits.

Firstly, we focused only on the demand of airline meals, so we could not consider the possibility of usefulness or application in other fields.

Secondly, we used 'Mean Absolute Error' (MAE) and 'Mean Square Error' (MSE) to measure with accuracy. If the order of accuracy was different in respect to forecast error, we preferred

MSE to MAE. Therefore, we could not remove the possibility of different results derived from other forecasting error measurements.

Last of all, the activity function of neural network models was considered as a hyperbolic tangent, and the number of the hidden layer as 1. The number of nodes in the hidden layer was equal to twice the number of input nodes. So, we could not consider other neural network architectures.

REFERENCES

[1] Brown, M. & C. Harris, *Neuro-Fuzzy Adaptive Modeling and Control*, Prentice-Hall, 1994.

[2] Chanda, G. D., G. S. Dispensa, & S. Ghose, "Comparing the predictive performance of a neural network model with some traditional market response models," *International Journal of Forecasting*, 1994, pp. 235-244.

[3] Drury, D. H., "Issues in forecasting management," *Management International Review*, 1990, pp. 317-329.

[4] Duarte, T. & R. Taffler, "Neural Networks and Empirical Research in Accounting," *Accounting and Business Research*, 1996, pp. 347-355.

[5] Eddy, P., M. Y. Hu, & M. S. Hung, "Two-Group Classification Using Neural Networks," *Decision Sciences*, 1993, pp. 825-845.

[6] Gorr, W. L., N. Daniel, & J. Szczypula, "Comparative study of artificial neural network and statistical models for predicting student grade point averages," *International Journal of Forecasting*, 1994, pp. 17-34.

[7] Hill, T., M. O'Connor, & W. Remus, "Neural Network Models for Time Series Forecasts," Management Science, 1996, pp. 1082-1092.

[8] Jain, B. A. & B. N. Nag, "Artificial Neural Network Models for Pricing Initial Public Offerings," *Decision Sciences*, 1995, pp. 283-301.

[9] Jo, T. C., "Neural Approach to Time Series Prediction by Generating Virtual Series," *Proceedings of the 1998 Fall Conference of the Korean Institute of Industrial Engineers*, 1998, pp. 1210-1214.

[10] Kim, S. H. & H. Noh, "A comparative Study of ARIMA and Neural Network Models: Case Study in Korean Corporate Bond Yields," *Proceedings of KORMS '96*, 1996, pp. 19-22.

[11] Kwon, O. & B. Golden, "A Neural Network Model for Predicting Atlantic Hurricane Activity," *Proceedings of KORMS '96*, 1996, pp. 39-42.

[12] Markham, I. S. & C. T. Ragsdale, "Combining Neural Networks and Statistical Predictions to Solve the Classification Problem in Discriminant Analysis," *Decision Sciences*, 1995, pp. 229-241.

[13] Norman, P. & S. Wang, "Application of the Back Propagation Neural Network Algorithm with Monotonicity Constraints for Two-Group Classification Problems," *Decision Sciences*, 1993, pp. 60-75.

[14] Philipoom, P. R., R. L. Paul & L. Wiegmann, "Using Neural Networks to Determine Internally-Set Due-Date Assignments for Shop Scheduling," *Decision Sciences*, 1994, pp. 825-851.

[15] Salchenberger, L. M., C. E. Mine, & N. A. Lash, "Neural Networks: A New Tool for Predicting Thrift Failures," *Decision Sciences*, 1992, pp. 899-916.

[16] Suh, C. J. & S. H. Kwak, "Airline meal demand forecasting with neural network and time series," *Journal of the Korean Production and Operations Management Society*, Vol. IX, No. 2, 1998, pp. 67-95.