經營科學 第9卷 第1號

A Comparative Study Between the Neural Network and The Winters' Model In Forecasting

Kim Wanhee*

1. Introduction

As the computer technology advances, the idea of building an artificial intelligence (AI) based system as a problem solving tool has drawn much attention among the researchers and practitioners in the various fields. Expert system and neural network, equivalently neural computing, represent two main streams in AI domain. These two scientific disciplines can be complementary. They, employing different approaches for a given problem, however, possess many fundamental differences.

Expert system is a logic-driven system which is built on a set of rules. The rules representing knowledge are used by an explicitly programmed inference machine. Therefore, an expert system, using its embedded reasoning process, takes a series of logical steps to reach a goal.

On the contrary, neural neural network is a data-driven system which needs sophisticated mathematics and statistical techniques such as linear programming and nonlinear regression analysis. It learns by examples. It develops its own problem-solving algorithm through analyzing various patterns of input vectors alone(unsupervised learning), or input and desired output vectors together(supervised learning). Although neural network can discover the complex fundamental function for generating an output vector, the reasoning procedure explaining how the conclusion is reached is hard, often impossible, to find due to the absence of any explicitly programmed logic.

A neural network generally consists of an input, several intermediate, and an output layers. Each of the layers contains a set of processing elements that are equivalent to

* CIM DIVISION SAMSUNG DATA System

the neurons in a human brain's nervous system. The processing elements in one layer are variously interconnected with those in other ones. When a given input vector in any arbitrary "N" space is exposed to a neural network, it statistically detects the fundamental function needed to map the given vector into an output vector in an arbitarary "M" space. The intermediate layers are used to refine the mapping procedure.

Neural network has several advantages over expert system with the conventional AI techniquse. They include distributed memory. fault tolerance, and parallel processing capability. These distinguishing features make neural network more attractive.

Fast growing rate of development in neural network has been indicated by many sources. For example, Edward Rosenfeld, the editor of intelligence newsletter, reported that in 1988 the number of companies worldwide involved in neural network exceeded 175, which was only about 24 in 1984. Furthermore, about 1,500 people attended the first International Neural Network Society conference in San Diego. During the conference period, more than 200 papers were presented(9).

In order for neural network to be conceived as a right problem-solving tool, its underlying theory as well as applicability should be proved and demonstrated. To date, no single research has been conducted on examining the performance of the neural network against well-known conventional approaches. Thus, this study sttempts to examine the applicability of neural network as a short-term forecasting model.

In detail, using two published time series with different characteristics, forecasts were made with a set of neural networks and Winters' additive models for the same period, Then, average forecast errors(AFE), mean absolute deviations(MAD), and absolute percentage errors(APE) were computed, and they served as the criteria on which neural network is compared with the Winters' model.

This parper is organized as follows. Section 2 illustrates several applications of neural networks. Section 3 presents the theoretical aspects of the major neural network paradigms as well as the structure of the back-propagation network used in the study. Section 4 describes the experiment including data analysis, modeling, and the performance criteria followed by the detailed discussion of the experimental results. Future research avenues including advantages and limitations of neural network are presented in the last section.

2. Applications of the Neural Networks.

Many succesful applications of neural network have been reported since the mid 1970s. They include text-to-speech conversion, economic and financial modeling, quality inspection, process and quality control, signal processing, and etc. Since it is our objective to list all of them, only a few are selected and briefly described. However, even this limited set may represent a broad spectrum of applications, reflecting the potential of neural networks.

Using the back-propagation network, Sejnowski and Rosenberg (6) built a system called Netalk. It is able to learn to convert ASCII text into understandable English Text. It babbles like an infant at first and gradually learns to distingusith between words. After about 16 hours of practice in matching text and phonemes (corresponding speech sound), it can talk at the level of a six-year-old child.

According to a recent issue of Fortune(4), a neural network for automobile motor inspection was developed and tested by Siemens, the West German electrical equipment maker. Previously workers used the sounds from the motors to identify defective ones, but that was rather tedious and performance dropped off quickly. Siemens now inspects the motors with neural networks, which turns out to be correct more than 90% of the time.

Widrow (8), the recipient of the Alexander Graham Bell Award, developed an adaptive network for noise reduction. His model keeps the signal-to-noise ratio high for the given state of the line. A slight variation of his achievement has been extended to build high-speed modems. Modems that transfer data at 9.6K bit/sec. use a neural network that learns to distinguish between noise and data (9).

The Graded Learning Network (GLN), developed by Hecht-Nielsen Neurocomputer, is a neural network architecture and training methodology. It eliminates extensive analytical work and modeling efforts in conventional approaches to process control, since it can provide accurate control solutions based on the actual performance of a system rather than on a model (7). During the training period, the network's trial on the given task is caught by the camera in the system. Then, the information from the camera is converted into the performance grade, which is fed into the neural network and used as a benchmark to improve its performance in the next trial. Since the learning takes place from an actual system, the GLN easily learns how to handle a given task when the system encounters unusual conditions.

Another implication of neural networks in manufacturing field was demonstrated at the San Diego Conference. Alware, Inc. in Cleveland presented a neural simulation of paper-tension control in a paper mill. The system performs the quality-control functions, determining how strong and thick paper will be.

Finally, neural networks have been most successfully implemented for financial services firms. Nestor, Inc. has installed a neural network system for beta-testing in a financial servieces firm, where it suggests recommendations concerning mortgage applications. Other financial services organization reports that its neural loan analysis system outperforms the company's human experts by 27% in terms of profitability(9). In addition, a recent issue of fortune reports that companies such as American Express that need to appraise credit risks say they are using neural networks.

3. The Neural Network

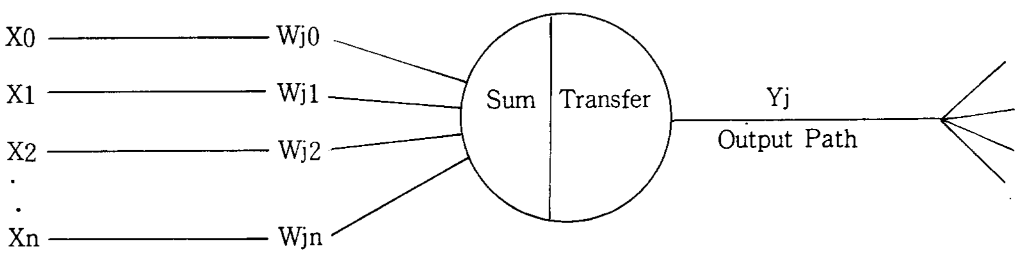
The scientists in the various fields have adopted different approaches to explain and model how the nervous system in a human's brain encodes/processes information and recognizes patterns. This results in more than a dozen paradigms in neural computing.

Earlier paradigms in neural computing include Rosenblatt's perception, Widrow's adaptive paradigm, Anderson's Brain-State-in-a-box (BSB), and Hopfield's hopfield network. They originated from the perspectives of neuro-physiologist, engineer, cognitive psychologist, and physicist.

Other paradigms, representing more recent developments, such as back-propagation, counter-propagation(CPN), Bi-directional Associative Memory(BAM), and Spatio-temporal Pattern Recognition(SPR) are the variations of the earlier ones. Each of the different paradigms has its own unique underlying theory, assumptions, and learning algorithms. All of them, however, employ a network, consisting of a series of layers with many processing elements.

In a neural network, the processing element(PE) represents the biological neuron. Typically, many input paths are linked to a PE, and the weighted sum of the values on these input paths is modified by a transfer function. This transfer function can be a threshold level that only passes information if the weighted sum reaches a certain level, or it can be a continuous function of the weighted sum. The output value of the transfer function is generally passed to the PE's output path. Figure 1 demonstrates such mechanics.

Figure 1. A Processing Element

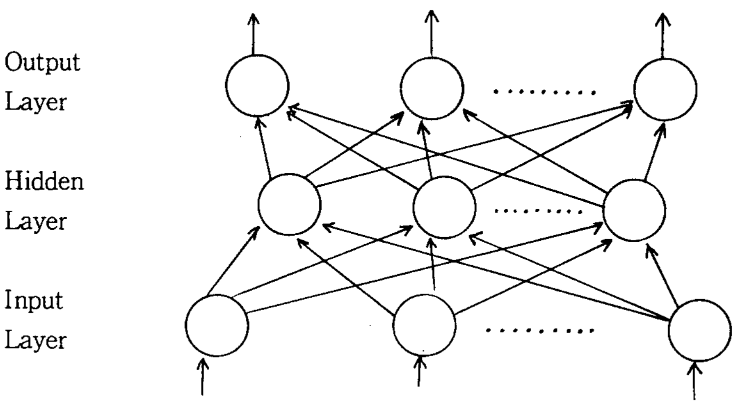


$$1 = \Sigma_i W_{ji} X_i$$

$$Y_i = f_{(i)} \text{ Transfer}$$

A neural network consists of many inter connected processing elements. Figure 2 shows a simple neural net architecture. There is an input layer where data is presented to the network, and an output layer which holds the response of the network to a given input. Hidden layers are distinct from the input and output layers.

Figure 2. A Neural Network Architecture



In this paper, the back−propagation method, a supervised learning technique using a gradient descent in an error variable, is employed because its underlying properties have led to several successful applications for encoding and data compression(1,2) and signal processing(3), In a back−propagation network, the error is computed by comparing an output value to a desired or expected value. A gradient of the error can be found by differentiating the accumulated errors with respect to the weights used to estimate an output value.

The basic architecture of the neural network in this study was originally developed by NeuralWare., Inc. In detail, the network consists of four layers−input, two hidden, and output layers. The input layer contains ten processing elements which correspond to neurons in a human's brain. The first hidden layer, containing sixteen processing elements. is divided into two halves. The left half uses the sigmoid transfer function. The right half uses a sine transfer function. The second hidden layer, which has nine processing elements, uses the sigmoid transfer function. For the learning and recall, the network uses a cumulative delta rule.

4. Experiment

4.1 Data Analysis, Modeling, and Measurements

Two sets of time series have been chosen from Business Statistics published by U.S. Department of Commerce. The first time series are U.S. domestic beer production quantities. The second one are factory sales(from plant in the U.S.) of new passenger cars. Each time series have 216 observations of the monthly data from 1969 to 1986.

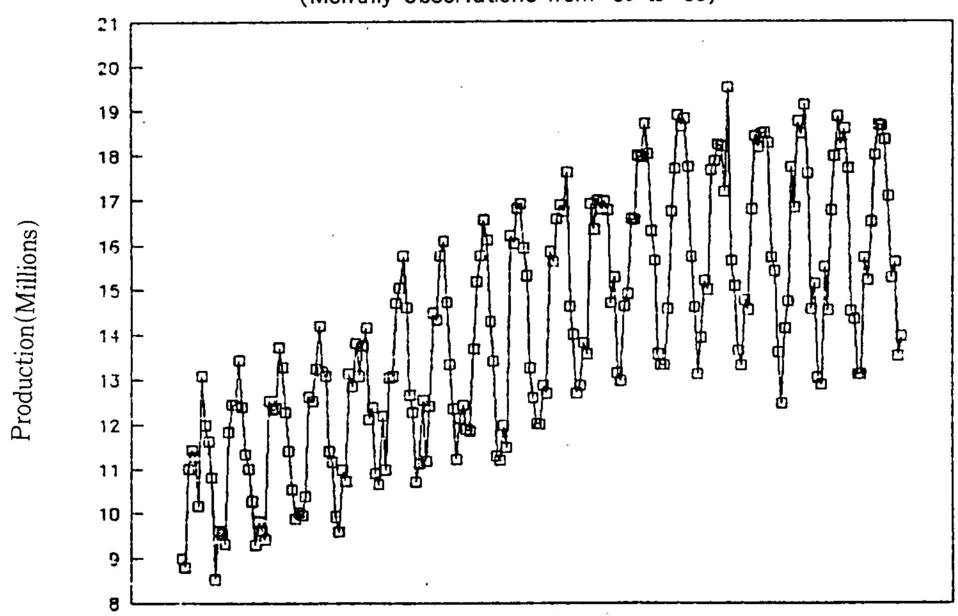
In order to figure out the characheristics of the time series, graphical analyses have been conducted. The plots in figures 3 and 4 reveal that strong monthly seasonalities exist in both time series. In addition, there is a monotonically increasing linear trend in the first time series, which is not true in the second one. The second one shows somewhat erratic pattern, exclusive of seasonalities, due to the twice of oil crisis taken place between 1969 and 1986.

Following a traditional research design for evaluating a forecasting model on a collection of time series(4), we divided each series into two samples. Each of the first samples, containing the monthly observations from 1969 to 1985, is used to locate the optimum smoothing constants as well as seasonal indices for Winters' additive models and train the neural networks. Due to the existence of strong monthly seasonalities, twelve seasonal indices are computed for each Winters' model. For each of the observations of 1986, which constitute the second samples, a set of forecasts has been made by each model.

When making the examples for training a neural network, special consideration should be given because what a neural network does is determined by the examples it is fed. With the existence of seasonalities, it may be reasonable to use a set of homogeneous observations to make an example for the training. That is, the observations in an example should be of the same month.

Figure3

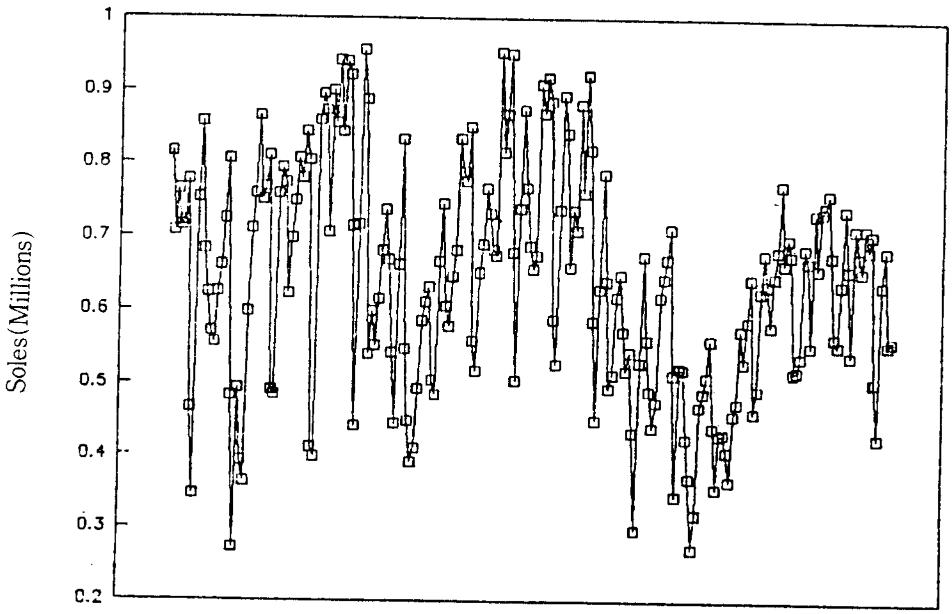
U.S. Beer Production
(Monthly Observations from '69 to '85)



Accordingly, a set of eleven consecutive observations of the same month, the first ten observations for an input set and the remaining observation for a desired output of the given input set, is used as an example for training purpose. For example, domestic beer production quantities of every January from 1969 to 1978 constitute an input set. For the given input set, the desired output should be the production quantity of January of 1979. By doing so, 12 sets of 7 examples, one set for each month, have been made from each time series.

When considering the lack of any prior knowledge about an optimal or near−optimal number of training for a given neural network, evaluating a limited number of neural networks was felt reasonable. Accordingly, with the same basic architecture, 15 neural networks, increasing the number of training by 1,000 were generated for each of the time series. Next, using each of the 15 neural networks, a set of forecasts for the 12 months as well as the corresponding average forecast error(AFE) and mean absolute deviation(MAD) were obtained. The neural network with the smallest MAD is chosen and used for the main performance comparison against the Winters' model.

Figure 4 Factory Sales of New Passenger Cars (Monthly Observations from 69' to '85)



For the comparative analysis, three criteria, average forecast error(AFE), mean absolute deviation(MAD), and absolute percentage error(APE), are employed and they are formalized as follows :

Notations :

i = time period index(i=1, 2,…, n)

Ai = actual observation of time period i

Fi = forecast of time period i,

$$AFE_i = \frac{\sum_{i=1}^{i=n} (F_i - A_i)}{n}$$

$$MAD = \frac{\sum_{i=1}^{i=n} |F_i - A_i|}{n}$$

$$APE_i = \frac{|F_i - A_i|}{A_i} \times 100$$

4.2 Experimental Results

Tables 1 and 3 summarize the MADs from the two Winters' models and 30 neural networks. Tables 2 and 4 show the actual figures in the two time series together with the corresponding forecast errors, their absolute values, and the percentage of absolute forecast errors. In the tables, W represents the Winters' model and Nn represents the neural network which was trained "n" thousand times. For the first time series, N12 performs the best among the neural networks, and N9 for the second time series.

As we can see from table 1 and figure 5, the neural network gives the smallest MAD of 500,000 barrels when it was trained 12,000 times, It is not certain if the result is global optimum since the maximum number of training was set to 15,000 for experimental purpose.

Table 1. MAD(U.S. beer production)

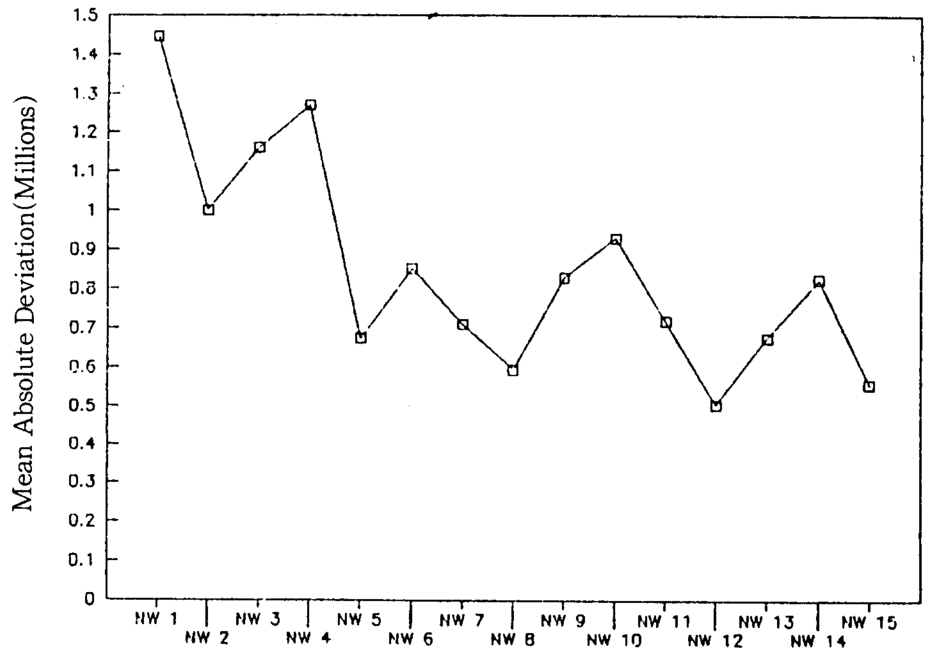
	Neural Network														
W	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	N13	N14	N15
55	145	100	116	128	67	85	71	59	83	93	72	50	67	82	56

Unit : 10,000 of barrels

Case 1 : The first time series(U.S. beer production)

Figure 5 clearly indicates that performance of the neural network is not exactly proportional to the number of training. Rather, it fluctuates along with a series of continuous U-shaped curves which are gradually moving downward. This behavior is very consistent with the underlying nature of back-propagation network, since it repeatedly uses gradient descent to find a least mean squared error approximation.

Figure 5. Learning Progress : U.S. Beer production



For the given time series, N12, the best neural network, and the Winters' model result in the MADs of 500,000 and 550,000 barrels, respectively. The difference between the two MADs is somewhat significant, since it accounts for 10% of the MAD with N12.

Neither the Winters' model nor N12 produce any systematic forecast errors. The distributions of the forecast errors on the 3rd and 4th colums in table 2 show about the right mixture of positive and negative forecast errors

N12 outperforms the Winters' model in terms of the absolute percentage error. The largest ones with the two models are 7.8% and 7.6%, respectively. N12, however, for 11 out of 12 months, produces the forecast whose absolute percentage error(APE) is less than or equal to 4% of the corresponding month's actual observation. Compared to N12, the Winters' model tends to generate more deviated forecasts from the actual observations.

Table 2 Winters' Model vs N12 : (U.S. beer production)

Month	Act. Prod.	FE		AFE		APE	
		W	N12	W	N12	W	N12
1	1,571	−122	21	122	21	7.8	1.3
2	1,521	− 87	46	87	46	5.7	3.0
3	1,651	29	58	29	58	1.8	3.6
4	1,799	− 86	−59	86	59	4.8	3.3
5	1,867	− 75	−58	75	58	4.0	3.1
6	1,865	− 50	−75	50	75	2.7	4.0
7	1,833	3	−15	3	15	0.1	0.8
8	1,706	82	58	82	58	4.8	3.4
9	1,526	52	27	52	47	3.4	3.1
10	1,562	− 32	21	36	21	2.3	1.3
11	1,353	24	103	24	103	1.8	7.6
12	1,397	− 17	43	17	43	12.	3.1
Mean		23	16	55	50	3.4	3.1

Case 2 : The Second Time Series(U.S. sales of new passenger cars)

Figure 6, displaying the changes in the performance of the neural network, 1) supports our previous discussion concerning the relationship between the network's performance and the number of training, and 2) indicates N9 as the best neural network. Here, the issue of an optimality is left question again due to the limited number training.

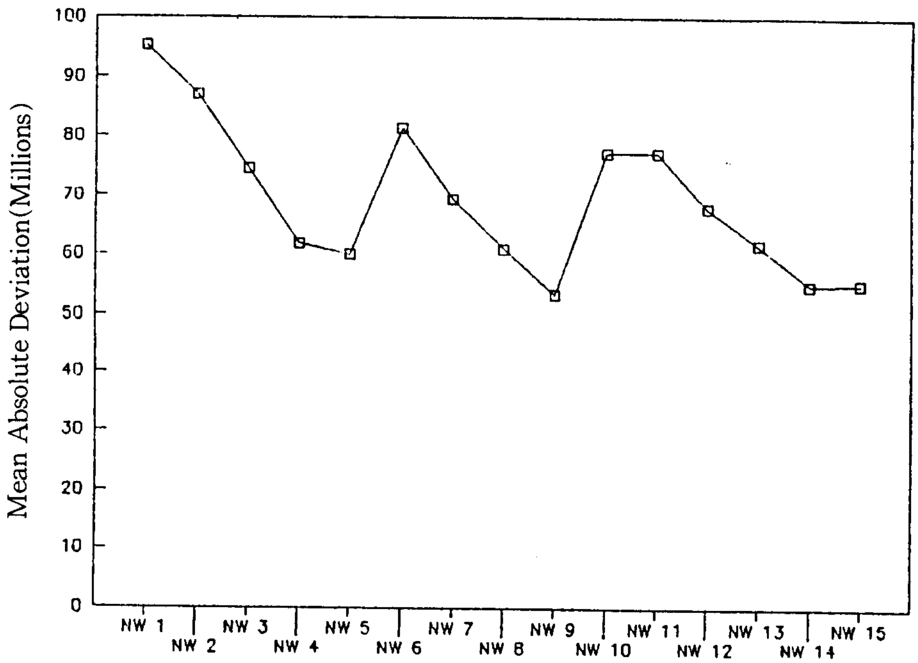
Table 3. MAD(New Passenger Cars)

	Neural Network														
W	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	N13	N14	N15
52	95	87	75	62	60	81	69	61	53	77	77	67	61	55	55

Unit : 10,000 of barrels

As table 4 shows, the Winters' model and N9 produce the MADs of 52,000 and 53,000, respectively. The difference of 1,000 accounts for only 1.9% of the MAD with the winters' model. The figures in the third and fourth columns of table 4, however, clearly show that N9 produces less balanced results than the winters' model. In detail, the average forecast errors with the Winters' model and N3 are 0 and −31. Moreover, further investigation on the distributions of the forecast errors reveals that N9 results in negative forecast error for 10 out of 12 observations, while the Winters' model does the equal numbers of positive and negative forecast errors.

Figure 6. Learning Progress : New Passenger Cars



Due to the erratic trend pattern embedded in the time series, neither models perform well with respect to absolute percentage error. Both models generate may large outliers. The largest absolute percentage errors made by the Winters' model and the neural are 21.4% and 30.5%. When counting the number of APEs which exceed 10%, the Winters' model and N9 report 4 and 9, respectively.

So far, we have compared the two models on the basis of only three performance criteria. When considering the applicability of the two models in the context of real world manufacturing, however, at least two more criteria are worth mentioning. First, without having a neural networks on a chip, simulating neural network in software seems to be the logical choice. In this regard, some researchers express their concern about the thime requirements for the simulations. Because short-term forecasting models should be built or modified easily within a reasonable amount of time so that they can make forecasts for a large number of items. When the simulation for the experiment was run on IBM PS/60 compatible PC without a math co-processor, it took about 20 minutes to train a network for 15,000 times. This supports our claim for neural network.

Table 4 Winters' Model vs Neural Network : (U.S. passenger cars)

Month	Act. Prod.	FE		AFE		APE	
		W	N9	W	N9	W	N9
1	713	−122	− 67	122	67	17.1	9.4
2	675	20	− 27	20	27	3.0	4.0
3	655	140	− 1	140	1	21.1	0.1
4	713	− 67	−103	67	103	9.4	14.4
5	685	55	− 27	55	27	8.0	3.9
6	706	0	− 89	0	89	0.0	12.6
7	505	− 21	49	21	49	4.2	9.7
8	426	− 14	130	14	130	3.3	30.5
9	637	− 6	− 65	6	65	0.9	10.2
10	684	40	− 97	40	97	5.8	14.2
11	556	58	− 8	58	8	10.4	1.4
12	561	− 82	− 69	82	69	14.6	12.3
Mean		0	− 31	52	53	8.2	10.2

Second, the costs associated with generating forecasts should be mentioned. The costs refer to the ones for maintaining the data sets and purchasing relevant software. It is obvious that the Winters' model requires less computer storage space than the neural network. In detail, the Winters' model requires the space for the seasonal indices, smoothing constants, and the latest observation, while the neural network does all of the previous observations. This cost, however, would become marginal when considering the current rate of development in hardware. Also, several inexpensive neural network simulation software are currently available, making the use of neural network economical.

5. Conclusions

As the overall experimental results indicate, the neural network produced better results than the Winters' model only for the first time series. For the second time series, neural network produced less desirable forecasts than the Winter's model. Although neither models performed well because of the underlying irregularity in the second time series, the performance gap between the two models was somewhat sizeable.

Apparently, it is somewhat early to conclude that neural network can be in place of Winters' model for short-term forecasting because of the contradictory results present in this study. Furthermore, this preliminary study has a limited scope in that only two sets of time series were employed and the neural networks were constructed with the same basic architecture. Thus, there is no quarantee that the neural network tested in the paper will give the similar results on other time series. Nonetheless, the promising results of the neural network for the first time series in this research suggest to consider the neural network as an alternative short-term forecasting technique to Winters' model, which is one of the best conventional time series forecasting models.

In order for the neural network to be accepted as a truly applicable short-term forecasting technique, further empirical research should be conducted to provide practical guidelines for using neural network. First of all, future comparative studies should test a variety of networks based on different paradigms on a collection of many different time series. This will help us to identify the best paradigm for a given time series and the characteristics of time series that have impact on the performance of a neural network.

Once the practical guidelines for matching paradigm and time series have been established, further investigation should answer the following questions : 1) how many layers and processing elements in each layer are sufficient? ; 2) when limited time is available, how many times of training should be made? ; 3) if a network results in one good answer, should it be worthy of continuing search for the better one? ; and 4) how many examples should be made for a certain time series?

There are two frequently cited fundamental limitations associated with neural network. They are lack of convergence to a minimum error and reasoning process. Sometimes, we may have to train a neural network over a long period of time without converging to a minimum error. In order to guarantee a convergence to a global minimum, the necessary learning laws should be provided. Although research is being undertaken for the problem, only limited findings are available.

Another fundamental problem that prohibits neural network from its wide use involves the lack of reasoning process. To date, neural network can not tell us how it reaches to a conclusion. This missing link often creates some difficulty in using it. For example, a neural loan analysis system tells us only whether a given application should be rejected or not. Therefore, the users may have to guess the system's reasoning process in order to give explanations for the rejected applications. However, these explanations may become inconsistent over a period of time.

Finally, little knowledge is available about how big task a neural network can handle within a reasonable amount of time. However, when we limit our focus on assessing its plausibility as a good short−term forecasting technique, it is a promising one in several aspects.

REFERENCES

1. Ackley, D. H., Hinton, G. E., and Sejnowski, T. J., "A Learning Algorithm for Boltzmann Machies", *Cognitive Science, Chaper* 9 : pp. 147−169.
2. Cittrell, G. W., Munro, P., and Zispser D., "Learning Internal Representations from Gray−Scale Images : An Example of Extensional Programming", *Proceedings of gth Annual Conference of the Cognitive Science Society*, pp. 461−473.
3. Lapedes, A. and Farber, R., "Non−Linear Signal Processing Using Neural Networks : Prediction and System Modeling", *Los Alamos National Laboratory Report*, LA−UR−87−2662.
4. Makridakis, s. et. al., "The Accuracy of Extrapolation(Time Series) Methods : Results of a Forecasting Competition", *Journal of forecasting*, 1 Jam.−March, (1982), pp. 111−153.
5. O'Reilly Brian, "Computers That Think like People", *Fortune*, February 27, (1989).
6. Sejnowski, T. J. and rosenberg, C. R., "Parallel Networks That Learn to Pronounce English Text", *Complex Systems*, 1 : 145−168, (1987)
7. Tom Inglesby, "No Clowning Around, Neural Networks Can Help Manufacturing", *Manufacturing Systems*, October (198).
8. Widrow, B., "Pattern Recognition and Adaptive Control", *Applications and Industry*, September (1964), pp. 269−277.
9. Williamson, M., "Neural Networks : Glamour and Glitches", *Computerworld*, February 15, (1988).