A Comparison of Traditional Forecasting Techniques and Neural Networks

Travis W. Gentry
Master of Business Administration Graduate, University of Wyoming

Bogdan M. Wiliamowski
Professor, Electrical Engineering Department, University of Wyoming

Larry R. Weatherford
Professor, Management and Marketing Department, University of Wyoming

ABSTRACT:
Neural networks represent an opportunity for businesses to make forecasts more accurately than traditional forecasting techniques.

Recent technological advancements in neural networks have begun to provide businesses with an array of opportunities in pattern recognition. Neural networks represent an opportunity to effectively solve many problems including scheduling, optimization, process control, and forecasting.

INTRODUCTION

Forecasting has been called both an art and a science. It is an ability to recognize patterns through a logical and analytical approach. With today's forecasting techniques, managers are able to understand the future better than managers of past eras (Makridakis and Wheelwright, 1989).

This study was designed to compare the short-term and long-term forecasting ability of particular traditional, quantitative forecasting methods and the forecasting ability of various neural network structures to predict the total passengers booked on a given airline.

DATA

The data was a time-series, provided by US Air, that represented the total passenger count on a specific flight. A total of thirty-six weeks of data were collected for this study.

The research was narrowed to studying the ability to forecast one variable, total passengers booked per week, for a short-term (one-week ahead) and long-term (three-weeks ahead). A sample of the data is given below:

01	245
02	456
03	362

The data is interpreted as week one, 245 passengers booked on all flights during week one; week two, 456 passengers booked on flights; and week three, 362 passengers booked on flights

TRADITIONAL FORECASTING METHODS

In the study, three methods of quantitative forecasting were used: moving averages (time series), exponential smoothing (time series), and regression (causal).

Simple Moving Averages

The assumption behind simple moving average (MA) models is that the average performance of the recent past is a good predictor of future performance (Eppen, Gould, and Schmidt, 1993). The simple moving average can be explained by equation (1) (Makridakis, 1981):

$$\hat{X}_{t+1} = \frac{X_t + X_{t-1} + X_{t-2} + X_{t-3} + \cdots + X_{t-N+1}}{N} \tag{1}$$

The forecasted value is represented as X_{t+1} and the current value and historic values are represented as X_t and X_{t-i} respectively. The key element in using moving average models is the proper selection of the number of past periods, N. The period values, N, ranged from two weeks to eight weeks in the study.

Exponential Smoothing

In the study, single exponential smoothing and exponential smoothing with trend techniques were used.

Single Exponential Smoothing

Single exponential smoothing or *exponentially weighted moving average* methods are similar to simple moving average techniques. Equation (2) illustrates the single exponential smoothing method (Makridakis, Wheelwright, and McGee, 1983).

$$\hat{X}_{t+1} = \alpha X_t + (1-\alpha)\hat{X}_t \tag{2}$$

The value of alpha (α) can range from zero to one. The higher the value of alpha, the more weight allocated to the most recent data observation. By iteratively solving for an optimum value of alpha during with the training set, the forecasting errors were minimized.

Exponential Smoothing with Trend

Exponential smoothing with trend is similar to single exponential smoothing except a trend component is introduced into the forecast. The trend component is defined below in equation (3) (Makridakis and Wheelwright, 1989).

$$T_t = \beta(\hat{X}_t - \hat{X}_{t-1}) + (1-\beta)T_{t-1} \tag{3}$$

Similar to the alpha value, beta (β) is a smoothing constant between zero and one. A higher value of beta indicates a larger portion of the most recent trend being added to the next period's forecast. The beta value can also be iteratively changed like alpha to minimize error values. The complete formula for exponential smoothing with trend is provided below in equation (4) (Krajewski and Ritzman, 1993).

$$\hat{X}_{t+1} = \alpha X_t + (1-\alpha)\hat{X}_t + T_t \tag{4}$$

Regression

Regression is a causal forecasting method that fits curves to the entire data set to minimize the forecasting errors. This study used three forms of regression techniques: linear, quadratic, and cubic.

Linear Regression

$$\hat{X}_t = a + bt \tag{5}$$

Quadratic Regression

$$\hat{X}_t = a + bt + ct^2 \tag{6}$$

Cubic Regression

$$\hat{X}_t = a + bt + ct^2 + dt^3 \tag{7}$$

It should be noted that as higher order polynomial models are used, the overall degree of error will be reduced, but the actual forecasting ability may also be reduced. For this reason, the cubic regression is the highest order forecasting tool used.

MEASURING FORECASTING ACCURACY

A fundamental concern in forecasting is the measure of forecasting error for a given data set and a given forecasting method. Accuracy can be defined as "goodness of fit" or how well the forecasting model is able to reproduce data that is already known (Makridakis and Wheelwright, 1989).

This study used three standard error measures: mean squared error (MSE), mean absolute percent error (MAPE), and mean absolute deviation (MAD).

Mean Squared Error (MSE)

As a measure of dispersion of forecast errors, statisticians have taken the average of the squared individual errors. The smaller the MSE value, the more stable the model. However, interpreting the MSE value can be misleading, for the mean squared error will accentuate large error terms. Equation (8) describes the mean squared error measurement.

$$MSE = \frac{\sum_{t=1}^{n} (X_t - \hat{X}_t)^2}{n} \tag{8}$$

Mean Absolute Percent Error (MAPE)

MAPE is regarded as a better error measurement than MSE because it does not accentuate large errors. Equation (9) illustrates the MAPE formula.

$$MAPE = \frac{\sum_{t=1}^{n} \left(\frac{\left| X_t - \hat{X}_t \right|}{X_t} \right)}{n} \times 100\% \tag{9}$$

Mean Absolute Deviation (MAD)

The final accuracy measurement is the mean absolute deviation. This error measurement is the average of the absolute value of the error without regard to whether the error was an overestimate or underestimate (Krajewski and Ritzman, 1993).

$$MAD = \frac{\sum_{t=1}^{n} \left| X_t - \hat{X}_t \right|}{n} \tag{10}$$

NEURAL NETWORKS

Two basic neural network structures were considered: a one layer neural network and a functional link network. Many training algorithms can be used, but for these two structures, a modified regression technique was used (Anderson and Wiliamowski, 1995). This technique leads to immediate convergence.

Input Data Transformation

The training sets of inputs and outputs were subjected to a mathematical, Gaussian distribution designed to introduce a element of decay into the data. The equation takes the exponential value of the negated squared difference between the most recent observation (X_n) and past observations (X) divided by a selected value of sigma (σ). The Gaussian weight is placed on all training input and output data of the neural network structures.

$$Gaussian\_Weight = \exp\left(\frac{-\left(X_n - X\right)^2}{\sigma} \right) \tag{11}$$

The purpose of the Gaussian distribution function was to place more emphasis on current data and less emphasis on older data. The initial data preparation is essential for the success of the neural network. The preparation is often an overlooked process that significantly impacts the neural network's accuracy.

Network Structures

There is literally an unlimited amount of neural network structures that can be designed. The structures can vary in the number of layers, the number of neurons, the numbers of inputs and outputs, type of activation functions, feedback information, etc. Combinations of these variables give designers a variety of options to implement the neural network for thousands of different applications.

This study was interested in comparing the basic power of the neural network in forecasting applications. Two structures were used: a single layer, feedforward neural network and a variable order, feedforward, functional link neural network.

Single Layer, Feedforward Network

In some cases, a one layer neural network can be reduced to one neuron problems. Therefore, the first network used in the study was a simple, one layer, one neuron design. The neuron's inputs were the past eight observations (X_{n-1}, X_{n-2},... X_{n-8}).

Functional Link Network

Single layer neural networks are easier to train than complex, multi-layer networks. However, single layer networks can only solve linearly separable problems. To avoid working with complex, multi-layer networks and training algorithms to introduce nonlinear separation abilities, the functional link network can be used (Pao, 1989). The functional link works by introducing nonlinear terms into a neural network. The network can be reduced to a single layer, which also increases the speed and ease of training.

RESULTS

The forecasting methods were compared against each other using MSE, MAPE, and MAD. Of the three accuracy measurements, MAPE provides the most accurate and fair comparison of forecasting methods.

For the one-week ahead forecast, the traditional methods were compared against the neural network structures. The first order, functional link neural network provided the best forecasting results. Table 1 summarizes the results.

	Error Calculation for Training Set				
	M.A.=2	Exponential Smoothing (M.A.D.)	Cubic Regression	NN First Order Functional Link	Single Layer Multiple Input
M.S.E.	8547.31	7311.09	5567.92	8575.36	5199.82
M.A.P.E.	23.33%	21.87%	19.84%	22.57%	18.90%
M.A.D.	73.55	67.42	61.87	66.45	57.27
	Error Calculation for Hold Out Sample				
M.S.E.	2486.17	2788.22	6148.44	3109.02	2721.00
M.A.P.E.	18.17%	20.42%	23.09%	**16.57%**	17.29%
M.A.D.	40.67	46.20	59.28	39.67	41.67

Table 1: One Week Ahead Forecast Results

Table 2 illustrates the comparison of the three-week ahead forecasts. The single layer, multiple input neural network structure yielded the best forecast.

	Error Calculation for Training Set (3 weeks ahead)				
	M.A.=3	Exponential Smoothing w/ Trend (M.S.E.)	Cubic Regression	NN First Order Functional Link	Single Layer Multiple Input
M.S.E.	11875.18	9916.18	5375.08	8575.36	5199.82
M.A.P.E.	31.82%	27.80%	19.33%	22.57%	18.90%
M.A.D.	95.07	86.99	60.84	66.45	57.27
	Error Calculation for Hold Out Sample (3 weeks ahead)				
M.S.E.	2573.13	8249.07	5911.95	7682.67	2687.33
M.A.P.E.	21.29%	40.28%	23.41%	25.50%	**17.30%**
M.A.D.	44.39	78.89	57.94	64.67	41.67

Table 2: Three Week Ahead Forecast Results

CONCLUSION

The results of the forecasting comparison illustrate how the most basic neural network (a single neuron) can outperform the traditional forecasting methods of moving averages, exponential smoothing, and regression. The first order functional link network yielded the best results for the one-week ahead forecast. The single layer, multiple input, feedforward network provided the best performance for three-week ahead forecast.

The results illustrate a fraction of the power neural networks can bring not only to forecasting, but all areas in business. There is a large potential for their use in databased marketing, production optimization, and the list goes on.

REFERENCES

Anderson, T., Wiliamowski, B. M., (1995). World Congress on Neural Networks, Washington, D.C.

Eppen, G.D., Gould, F. J., Schmidt, C. P., (1993). *Introductory Management Science*, Prentice Hall, New Jersey, 4th ed., p. 787.

Krajewski, L. J., Ritzman, L. R., (1993). *Operations Management, Strategy and Analysis*, Addision-Wesley Publishing Company, Inc., p. 436.

Makridakis, S. et al., (1981). "Accuracy of Extrapolation (Time Series) Methods: Result of a Forecasting Competition, *Journal of Forecasting*, June, Vol. 1, p. 143.

Makridakis, S., Wheelwright, S. C., (1989). *Forecasting Methods for Management*, John Wiley & Sons, New York, p. 3.

Makridakis, S., Wheelwright, S. C., McGee, V. E., (1983). *Forecasting: Methods and Applications*, John Wiley & Sons, New York, p. 86.

Pao, Y. H., (1989). *Adaptive Pattern Recognition and Neural Networks*, Addison-Wesley Publishing Co., Reading, Mass.

INTELLIGENT ENGINEERING SYSTEMS THROUGH ARTIFICIAL NEURAL NETWORKS

VOLUME 5

FUZZY LOGIC AND EVOLUTIONARY PROGRAMMING

Editors:

Cihan H. Dagli
Metin Akay
C. L. Philip Chen
Benito R. Fernández
Joydeep Ghosh