N

Neural Network Time Series Forecasting Using Recency Weighting

Brad Morantz
Science Applications International Corporation, USA

Thomas Whalen
Georgia State University, USA

G. Peter Zhang
Georgia State University, USA

INTRODUCTION

In building a decision support system (DSS), an important component is the modeling of each potential alternative action to predict its consequence. Decision makers and automated decision systems (i.e., model-based DSSs) depend upon quality forecasts to assist in the decision process. The more accurate the forecast, the better the DSS is at helping the decision maker to select the best solution.

Forecasting is an important contributor to quality decision making, in both the business world and for engineering problems. Retail stores and wholesale distributors must predict sales in order to know how much inventory to have on hand. Too little can cause lost sales and customer dissatisfaction—If too much is on hand, then other inventory problems can occur (i.e., cash flow, ad valorem tax, etc.). If the goods are perishable, it could most certainly be a financial loss.

Items that occur over time, as in the number of cars sold per day, the position of an airplane, or the price of a certain stock are called “time series.” When these values are forecast, the accuracy can vary, depending on the data set and the method. This subject has been greatly discussed in the literature and many methods have been presented. Artificial neural networks (ANN) have been shown to be very effective at prediction.

Time series forecasting is based upon the assumption that the underlying causal factors are reflected in the lagged data values. Many times, a complete set of the causal factors either is not known or is not available. Predictions are made based upon the theory that whatever has occurred in the near past will continue into the near future. Time series forecasting uses past values to try and predict the future.

A slight modification to this concept is the application of recency. What happened more recently is closer to the current situation than the more distant past. The older data still contain knowledge, it just is not as important (or as correct) as the newest information. Things change, life is dynamic, and what used to be may be no more or may be to a different extent.

Modification of the training algorithm of a neural network forecaster to consider recency has been proven on real economic data sets to reduce residual by as much as 50%, thereby creating a more accurate model which would allow for better decision making.

BACKGROUND

Time series forecasting assumes that the near future will resemble the not too distant past. Many time series forecasters assume linearity in the data. Some also assume that whatever it is that is driving the outcome will continue to do so in the same way. Some researchers have acknowledged that the underlying processes are nonlinear but claim that their method is adequately robust and that the error from this naïve assumption is minimal. Depending on the accuracy required and the time span, the method might be adequately robust for the intended purpose.

It has been shown that no single linear forecasting method is globally the best (Makridakis et al., 1982). Real-world events rarely happen linearly (Lin, Yu, Gregor, & Irons, 1995). In fact, it is unreasonable to assume a priori that any particular data set is linear (Zhang, Patuwo, & Hu, 1998). Yet, many of the prediction models are based upon linear methodology. In some cases, if the forecast horizon is short enough and the model adequately robust, then linearity is assumed and the error is small.

In the last few decades, models like ANNs and others have evolved to cope with nonlinear models (Lin et al., 1995; Zhang, 2001). ANNs are limited mathematical emulations of biological neural networks, which use a relatively larger number of parameters to model input to output relationships (Abdi, Valentin, & Edelman, 1999). ANNs are capable of modeling nonlinear functions, are not affected by multicollinearity, and are less sensitive to noise than their traditional statistical counterparts. Neural networks have been shown to be universal function approximators. They map input to output patterns. ANNs are data driven and require few if any a priori rules or information about the model. They learn underlying relationships, even if unknown or hard to measure (Zhang et al., 1998). ANNs have become a powerful tool for forecasting.

These methods currently do not consider the possibility that the underlying factors might drift or change over time and are most useful when conditions remain constant (Bowerman & O'Connell, 1993). But some processes do have changes in their causal factors, occurring over time. Some of these occur slowly while others occur quickly or are even in the form of system shocks. Two or more of these influences can be occurring simultaneously. The solution has been to typically limit the forecast to short horizons. Qualitative forecasting is often used when the existing data patterns might change. These changes in data patterns are quite often based on expert opinion (Bowerman & O'Connell, 1993).

Training is a process by which the parameters of the model are estimated. Research has shown that the performance of a forecast method is affected by the training method and upon which data it was trained. Estimating the parameters of the model allows the model to determine the input-output relationship (Hanke, Wichern, & Reitsch, 2001). These learned relationships are the basis for the forecast that it will produce. Therefore, which data is used to estimate these parameters does affect the performance and accuracy of the prediction model. The performance of the neural network is a result of the training process that assigned the weights (Lawrence, Back, Tsoi, & Giles, 1997). These methods rely totally on the historical data that was used to train the models (Kolarik & Rudorfer, 1994). Some of the more common training methods for neural networks are: back propagation, genetic algorithms, reduced gradient, steepest descent, and simulated annealing.

Ordinary least squares (OLS) regression has many parallels to neural networks. One of the assumptions of autoregression is constant error variance or homoscedasticity. Heteroscedasticity, or non-constant error variance, can degrade the performance of the model. Weighting is an accepted, although not widely used, method to correct this. Transforming some or all of the variables is a common way of attempting to correct this problem (Neter, Kutner, Nachtsheim, & Li, 2005). Weighted least squares regression is another approach, where the objective function is multiplied by the weighting matrix. This is effective in estimating model parameters when the data points differ in quality from one another. The weight determines how much each observation influences the estimation of the model parameters (National Institute of Standards [NIST], 2006). ANNs are limited mathematical emulations of biological neural networks, which use a relatively larger number of parameters to model input to output relationships (Abdi et al., 1999). The weight determines how much each observation influences the estimation of the model parameters (NIST, 2006). Assigning the weights requires skill, knowledge, experience, and some luck (Render & Stair, 1988).

MAIN FOCUS OF THE ARTICLE

Recency weighting is introduced here as a combination of rolling and moving windows. Like rolling windows, older information is retained to add knowledge to the model. As in weighted regression, this older information is reduced in importance, as some rules may have changed or causal factors may have become less influential. The moving window concept is maintained, because as new data is added some of the older data is discarded. This combination of rolling and moving windows and weighted regression is called "recency weighted" (RW).

Assigning a weight of zero is the same as not including it in the training set. In the situation where the older data has less effect on the data, the variance becomes larger, and hence the weight smaller. As the weight approaches zero, it has the effect of attenuating the influence or removing that observation from the training set. In summary, recency method gives more

Related Content