An Exploratory Study for Neural Net Forecasting of Retail Sales Trends Using Economic Indicators

Mike Orra and Gursel Serpen

Electrical Engineering and Computer Science Dept
University of Toledo
Toledo, Ohio 43606 USA

Abstract

This paper proposes the use of artificial neural networks (feed forward multi-layer perceptron and Elman recurrent networks) in forecasting sales trends at retail by analyzing industry and manufacturer specific metrics along with national economic indicators. Relevant data drivers were gathered based on consultations with the manufacturer as well as experts in the fields of economics and finance. Simulations were run using the proposed system to determine the amount of product sold at retail for the end of the present week, as well as how much product will sell at retail three months from today: three months is the required lead time for the manufacturer to fabricate the products being examined. The initial results of this study indicate that both feed-forward neural networks and Elman recurrent neural networks show potential in being able to forecast sales trends with reasonable accuracy.

Keywords: retail sales forecasting, neural network prediction, multilayer perceptron, Elman recurrent network, national economic indicators

1. Introduction

There are many factors which influence the flow of production in the manufacturing industry: availability and cost of materials, consumer demand, cost of labor, shipping and transportation, etc. Moreover, seasonal trends in the retail industry can also have significant impact on the production and shipment of goods. The aforementioned variables are complex functions of economic, social, and political environments. The ability to accurately forecast these conditions has tremendous impact on the determination of production quotas. A misquoted production schedule can result in a shortage in the production of goods, or a large surplus of inventory.

A sudden change in market conditions could lead to lost revenue and possibly declined market share for the firm. For example, the manufacturer could easily deplete its inventory in an attempt to satisfy sharply increasing market demands. Furthermore, the lag between the time taken to manufacture the product and the time at which it arrives at the marketplace could be on the order of weeks or months. During this time, the manufacturer would certainly not realize any revenue and may end up losing market share to competitors whose products are readily available. On the other hand, if a firm fails to realize a sudden decline in consumer demand, it may face a surplus of goods. This excess inventory can result in costly warehousing and maintenance fees. Corporations, on occasion, have had to resort to public clearance sales to purge their warehouses of overstocked goods. While the company can recover some revenue, it pales in comparison to the revenue that could have been realized if they had not produced more than was actually necessary. This is an excellent example of the losses that manufacturing firms face due to inaccurate production estimates that lead to excess inventory. These losses can be on the order of millions of dollars. Being able to accurately forecast product demand is critical for any manufacturing firm: the potential losses associated with poor or inaccurate production quotas makes investment in accurate forecasting systems well worthwhile.

Artificial neural networks (ANN) algorithms have shown promise in constructing data-driven prediction models for time-series data, as is found in retail. In particular, ANNs were reported to perform better or at least comparable to other time series models [3-5]: specifically a number of studies have reported good classification and predictive ability using feedforward multi-layer perceptron (MLP) neural networks [10, 11]. Furthermore, recurrent neural networks (RNN) were determined to be well-suited to short term forecasting [2, 9]. There are a number of successful studies on the application of ANNs to trends forecasting. Shanmugasundaram et al. [1] examined the management of inventory for a pharmaceutical company using a recurrent neural network and real-time recurrent learning for training. The company sought to maintain a 95% fulfillment rate of their products (that is there is a 95% chance of the product being in stock at any given time). The idea was to accurately predict consumer demand and adjust inventory accordingly. Each day was defined to be a data point. Sales records for two classes of drugs, "fast and slow moving drugs," for day *t* were given to the network and the sales for day *t+1* (the next day) were predicted by the network. The actual sales for day *t+1* were then provided to the network for continued learning. Various network configurations were tested to determine the most appropriate topology. The resulting recurrent network was able to follow trends fairly well for both fast and slow moving drugs. While the recurrent neural network was able to foresee upcoming trends in sales, it was not able to predict the actual sales figures with any significant accuracy. One possible explanation is the influence of external factors in sales that were not included as indicators.

A separate study [2] proposed the use of artificial neural networks in long-term electric load forecasting. Two network types were examined: a three-layer feed forward multi-layer perceptron network with back-propagation training, and a recurrent neural network. Factors such as gross national product (GNP), gross domestic product (GDP), index of industrial production, and the amount of CO_2 pollution were included among the primary indicators of electricity demands. The reported results of this study indicate that forecasts by the RNN network are relatively accurate for up to three years into the future.

A number of observations are in order as a result of these studies while noting that forecasting sales trends appears to be highly ill-defined: (1) artificial neural networks have the ability to predict with as much, if not more, accuracy than traditional statistical models, and (2) indirect economic data indicators might be highly beneficial to improve prediction accuracy rates. Accordingly, this paper proposes the inclusion of manufacturer and industry specific indicators along with national economic indicators in the training data set for time-series forecasting of consumer product (kitchenware) sales at retail using artificial neural networks.

2. Proposed Neural Predictor

In this exploratory research project, the idea is to determine if it is feasible to be able to forecast, with reasonable accuracy, how much inventory will sell in one week, and how much product will be in demand three months into the future. Discussions with the retail marketing manager of the product manufacturer* revealed that the following data are used as the primary indicators in their current forecasting system:

a) Number of retail stores carrying the product.
b) Price of the product for that week (in US dollars).
c) In-print advertisement status for that week.
d) Ending inventory (units).
e) Shipments for that week (units).

In addition to the above set of data, an effort was put forth to examine other potential economic factors that may have significant impact on the performance of retail sales for department stores and companies in house-wares manufacturing. A relatively extensive set of interviews with academicians, marketing professionals, and a literature survey indicated that the set of economic drivers to be included in training might be specified as follows:

f) Disposable personal income (billions of dollars, monthly, seasonally adjusted) [6].
g) Federal funds rate (percent, weekly) [6].
h) Consumer price index (index, monthly, seasonally adjusted) [6].
i) Consumer confidence index (index, monthly, seasonally adjusted) [8].
j) Household debt service payments as a percent of disposable income (percent, quarterly) [6].
k) Household financial obligations as a percent of disposable income (percent, quarterly) [6].
l) Total retail sales excluding food services (millions of dollars, monthly, seasonally adjusted) [7].
m) Department store retail sales (millions of dollars, monthly, seasonally adjusted) [7].

According to consultations with local experts in the fields of economics, finance and forecasting, drivers f through i represent the basic primary factors affecting the performance of the retail industry. Clearly, if the consumer confidence index is high and the average disposable personal income is also high, the retail industry is likely to benefit from these conditions and perform well. Similar logic and rationale can be applied to drivers j through m; past performance of the retail industry and measures of household debt and financial obligations may be able to provide forecasting systems with additional insight towards upcoming industry trends. Therefore, we believe that this set of indicators might provide a sufficient basis for this exploratory study.

Attributes a through m above were used as inputs to the neural predictors, which were implemented with a multi-layer feedforward neural network, and an Elman recurrent neural network. These network architectures will each have two output neurons: the first output neuron yields the predicted amount of product sold at retail for the end of the present week, and the second yields the predicted amount of product to be sold at retail three months (considered as 12 weeks in this study) from the present (three months is

* Name of the manufacturer is withheld in order to safeguard proprietary information and data.

the required lead time for the manufacturer to supply the specific product being studied herein).

3. Simulation Study

Mathwork's Matlab Neural Network Toolbox was employed for the simulation study [12,13]. The simulations in this study focused solely on one particular product produced by the manufacturer and were conducted for the following neural network choices, topologies, and parameter value settings in Table 1 (where terminology is as it appears in the Matlab Neural Network Toolbox [13]):

Neural Net	**MLP**	**Elman**
Input Range	[0.0,1.0]	[0.0,1.0]
Neurons in Input Layer	13	13
Number of Hidden Layers	1	1
Neurons in Hidden Layer	[1,26]	[1,26]
Neurons in Output Layer	2	2
Activation Function	tansig	tansig
Training Algorithm	traingdm	traingdx
Initial Weights & Biases	[0.0,1.0]	[0.0,1.0]
Performance Measure	MSE	MSE
Desired MSE Goal	0.01	0.01
Upper Bound on Epochs	10,000	10,000

Table 1. Parameter values and topology settings

Typically there is much room for heuristics and trial-and-error in neural network computing. The number of hidden layers is 1 in most cases as much empirical evidence in the literature indicates as sufficient. On the other hand, the number of neurons in the hidden layer requires a hands-on trial-and-error based search process for determining a near-optimal value for the problem at hand. Similarly, substantial empirical evidence in literature suggests the type of training algorithms appropriate in many circumstances. Initial small values for network weights and biases determined through uniform random process is also appropriate. It is relevant to note that although empirical and heuristic value determination for the set of parameters might not be optimal (neural net parameter space is not explored), yet it saves incredible amount of time and effort in the process of training

Determining the (near) optimal number for the hidden layer neurons was realized through a rather meticulous empirical process. Each type network (i.e. MLP and Elman) ran simulations with a varying number of hidden layer neurons: from 1 to 26. The application software then analyzed the network's predictive error from each of the 26 trials to determine which of the 26 network configurations yielded the most accurate prediction for each of the four weeks under test. The application then determined which of the 26 network configurations yielded the most accurate prediction (least error) for each of the four sample points examined. The application applied the same procedure to determine which of the 26 network configurations yielded the most accurate prediction when examining the cumulative average over all four weeks. This was done to obtain some reference for how many hidden layer neurons would be effective for a given network topology in this application.

Data scaling is an important aspect of neural computing. In order to appropriately scale the data down to the range of [0,1] a linear model was used. The translation is mathematically defined as [4]:

$$s = \frac{v - \min(v_1, \ldots, v_n)}{\max(v_1, \ldots, v_n) - \min(v_1, \ldots, v_n)},$$

where s is the scaled down value, and v is the value to be scaled. Similarly, a translation process was employed to scale up the data, which is defined by an algebraic manipulation of the above equation [4]:

$$v = \min(v_1, \ldots, v_n) + (s)\left[\max(v_1, \ldots, v_n) - \min(v_1, \ldots, v_n)\right]$$

The household measures of financial obligations and debt are given as percentages. Thus the linear data scaling defined above need not be applied for them. Furthermore, in order to give a more accurate representation of the number of retail stores carrying the product under examination, the data was rescaled to account for the retailer's growth model (information provided by the manufacturer).

The networks were trained with the first sixty weeks of data, and were then prompted to predict the sales at retail for week k (output neuron 1) and week k+11 (output neuron 2) over the next four weeks. The results were then compared to the retailer's actual sales records for weeks 61 through 75. The data set gathered for this study (comprised of industry and national economic indicators) spanned from the summer of 2002 to the fall of 2003.

The simulation results obtained for the two networks indicate that the best performing topologies (in terms of averaged prediction error rate) are with 9 and 21 hidden layer neurons for the MLP and Elman networks, respectively. The respective prediction errors (averaged over the four-week periods) are 12.52% and 12.27% for the MLP and the Elman. Figure 1 presents normalized predicted and actual sales values over the two four-week periods (weeks 61 through 64 and 72 through 75). The MLP tends to underestimate for both the short (weeks 1 through 4) and long (weeks 12 through 15) term prediction, while the Elman delivers reasonably close estimates for the short term (weeks 1 through 3) and underestimates for the long term (weeks 12 through 15) for the most part. Long terms prediction by both networks are underestimates with practically the same values.

Searching over the best performing network topologies for a given week results in the figures in Tables 2 and 3. In the case of Elman network, the averaged error value reduces to 8.82% (from 12.27%) for network topologies with varying hidden layers, which suggests potentially a multi-network architecture rather than a single network with fixed hidden layer neurons.

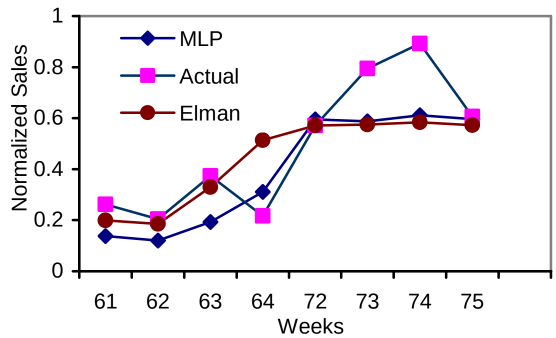

Figure 1. Prediction performances for MLP and Elman with 9 and 21 hidden neurons, respectively.

MLP Network Weeks	Optimal Number of Hidden Neurons	Predictive Error (%)
k & *k*+11	9	7.38
k+1 & *k*+12	9	14.51
k+2 & *k*+13	9	23.05
k+3 & *k*+14	26	2.64
Average		11.89

Table 2. Best Performing MLP topologies

Elman Network Weeks	Optimal Number of Hidden Neurons	Predictive Error (%)
k & *k*+11	4	1.96
k+1 & *k*+12	13	9.45
k+2 & *k*+13	13	14.46
k+3 & *k*+14	5	9.40
Average		8.82

Table 3. Best performing Elman topologies

Both the Elman and the MLP networks perform reasonably well as potential contenders for serving as prediction algorithms for this data set. Simulation results suggest that the performance for both networks depend on the number of neurons in the hidden layer among others: there is not a single hidden layer neuron count that leads to superior performance for the entire time-series data for either network. The results for fixed hidden layer neuron count as in Figure 1 indicate that the Elman performs somewhat better for short term prediction considering the average values over the four-week period covering weeks 61 through 64. The MLP and the Elman project very similar performance profiles for long term prediction. On the other hand, results presented in Tables 2 and 3 suggest that a hybrid architecture, where multiple predictor networks with differing hidden layer counts operate on the same data in parallel, might be appropriate to address dependency on the hidden layer neuron count for either network while noting that revision to such architecture might be dictated by other yet undiscovered dependencies on remaining parameters.

The network performance can only improve further in the presence of continually incoming data stream given that a continuous training algorithm is applied to keep the network in online learning mode, while observing that over-fitting does not occur.

4. Conclusions

The results of this feasibility study clearly suggest artificial neural networks might offer a viable choice in time series prediction of retail sales trends. However, further extensive testing is needed to validate or invalidate the findings in a statistically significant framework. We plan to further the study reported herein using neural networks as well as statistical models for a comparative context.

5. References

[1] J. Shanmugasundaram, M. Prasad, S. Vadhavkar, A. Gupta, “Use of Recurrent Neural Networks for Strategic Data Mining of Sales,” MIT Sloan Working Paper, no. 4347-02, http://ssrn.com/abstract=300679, 2002.

[2] B. Kermanshahi, H. Iwamiya, “Up to year 2020 Load Forecasting using Neural Nets,” Electrical Power and Energy Systems, vol. 24 pp. 789-797, 2002.

[3] V. Desai, J. Crook, G. Overstreet, “A Comparison if Neural Networks and Linear Scoring Models in the Credit Union Environment,” The European Journal of Operational Research, vol. 95, pp. 24-37, 1996

[4] G. C. Vasconcelos, P.J. Adeodato, D.S.M. Monteiro, “A Neural Network Based Solution for the Credit Risk Assessment Problem,” Proceedings of the Fourth Brazilian Conference on Neural Networks, pp. 269-274, 1999

[5] G. Zhang, M. Hu, B. Patuwo, D. Indro, “Artificial Neural Networks in Bankruptcy Prediction: General Framework and Cross-Validation Analysis,” The European Journal of Operational Research, vol. 116, pp. 16-32, 1999

[6] Source of Data: http://research.stlouisfed.org/fred2/

[7] Source of Data: http://www.census.gov

[8] Source of Data: http://www.sca.isr.umich.edu/main.php

[9] S. Ho, M. Xie, T. Goh, “A Comparative Study of Neural Network and Box-Jenkins ARIMA Modeling in Time Series Prediction,” Computers & Industrial Engineering, vol. 42, pp. 371-375, 2002.

[10] J. Zahavi, N. Levin, “Issues and Problems in Applying Neural Computing to Target Marketing,” Journal of Direct Marketing, vol. 11 no. 4 pp. 63-75, 1997.

[11] T. Al-Saba, I. El-Amin, “Artificial Neural Networks as Applied to Long-Term Demand Forecasting,” Artificial Intelligence in Engineering, vol. 13, pp. 189-197, 1999.

[12] http://www.mathworks.com/products/neuralnet/

[13] H. Demuth, M. Beale, *Neural Network Toolbox User's Guide*, Version 4, MathWorks, Inc. Natick, MA., 2000.