Title	Application of artificial neural networks in sales forecasting
Author(s)	Yip, DHF; Hines, EL; Yu, WWH
Citation	International Conference on Neural Networks Proceedings, Houston, USA, 9-12 June 1997, v. 4, p. 2121-2124
Issued Date	1997
URL	http://hdl.handle.net/10722/46579
Rights	Creative Commons: Attribution 3.0 Hong Kong License

Application of Artificial Neural Networks in Sales Forecasting

Devil H.F. Yip[1], E.L. Hines[2] and William W.H. Yu[1]

Department of Industrial and Manufacturing Systems Engineering,
University of Hong Kong, Pokfulam Road, Hong Kong[1]

Department of Engineering, University of Warwick, Coventry, CV4 7AL, United Kingdom[2]

Email : hfdyip@hkusua.hku.hk

Abstract

The aim of the work presented in this paper is to forecast sales volumes as accurately as possible and as far into the future as possible. The choice of network topology was Silva's adaptive back-propagation algorithm [1] and the network architectures were selected by Genetic Algorithms (GAs) [2]. The networks were trained to forecast from 1 month to 6 months in advance and the performance of the network was tested after training. The test results of artificial neural networks (ANNs) are compared with the time series smoothing methods of forecasting using several measures of accuracy. The outcome of the comparison proved that the ANNs generally perform better than the time series smoothing methods of forecasting. Further recommendations resulting from this paper are presented after the conclusions.

Keyword: Sales Forecasting, Artificial Neural Networks, Back-propagation, Genetic Algorithms.

1. Introduction

The ability of artificial neural networks (ANNs) to generalize allows for their utilization in the context of classification or prediction problems such as bond rating [3], stock price prediction [4] and currency exchange rate prediction [5]. Traditionally, statistical analysis models have been used to assist in the decision making process where an expert makes predictions based on a large number of cases. ANNs may have an advantage over other methods as they do not require specification of the functional model of the problem; they 'learn' the underlying model from the training cases.

This paper is based on the DTI indexes and the sales data of Glynwed Foundry Products Limited in the United Kingdom from January 1986 to October 1992. The aim is to use appropriate combinations of these inputs to forecast sales 1 month to 6 months in advance. The ANNs results are compared with the time series smoothing methods of forecasting (i.e. 3 months moving average, 12 months moving average, alpha equal to 0.5 and 0.8 exponential moving average). The network is trained to forecast from 1 month to 6 months in advance and the network performance is tested after training. The test results are compared with the time series smoothing methods of forecasting using several measures of accuracy (i.e. mean error, mean percentage error, mean absolute deviation, mean absolute percentage error and mean squared error). The choice of network topology was Silva's adaptive back-propagation algorithm [1] and the network architectures were selected by genetic algorithms [2].

2. Network architecture and training algorithm

One of the critical problems with using ANNs is in designing the network topology appropriate for a given problem. Design considerations include network architecture (number of hidden units, hidden layers, and interconnects), and the network parameters (learning rate, momentum, weight decay, initial weights, etc.,). Typically the topology and other parameters are determined through trials and experiments which could often be long and tedious. The temptation is to accept any solutions which may not be optimal. In some cases, the problem may appear insolvable after a long process of trials and experiments during which the right combination of parameters has not been reached.

GA researchers have shown that GAs are a suitable tool for determining the optimal topology and parameters of ANN for a specific problem [6-8]. GAs have the advantage of less likely to get stuck in a local optima. When the search space is large and there is no *a priori* knowledge about the parameter space, the optimization process may take a long time to converge.

The Back-propagation network has been adopted because it is capable of approximating arbitrary mapping given a set of examples and adaptive back-propagation (ABP) technique [9] is used to achieve fast convergence during training. Since the authors are using the ABP for training, the ANNs are restricted to a feed-forward Multi-Layer Perceptron (MLP). The MLP may be randomly or fully connected. In a fully connected MLP every node in a layer is connected only to all the nodes in a subsequent layer.

GAs have been used to determine the most effective number of hidden layers, number of hidden nodes per layer and interconnects. For each jth genetic string or chromosome, the training of the network produces a fitness value given by:

$$f_j = \frac{1}{\sum\limits_{i=1}^{8} W_i F_i(X_{ij})}$$

Where W_i is the significance of weighting factor for the ith parameter,
F_i is a function for the ith parameter and,
X_{ij} is the ith parameter value.

The parameters of the fitness values included are: the overall error, the maximum number of nodes in the network, the number of links in the network, the number of epochs required to train the network, maximum number of epochs specified, the number of misclassified patterns during training/recall, the number of nodes out of the tolerance limit during training/recall, and the total number of training/recall patterns. Each gene represents a specific network topology and the fitness value indicates its performance.

Crossover and/or mutation follows each training process and the genes compete for survival based on their fitness values. The optimization process is subject to some specified conditions. The network has a maximum of 2 hidden layers each with a maximum of 15 hidden nodes; 10 generations with population size 50; crossover probability 0.9 and mutation probability 0.01.

3. ANNs in sales forecasting

Sales forecasting is a valuable management tool for determining the future magnitudes, timing, and possible effects of uncontrollable influences over a company's

future success. It is an estimate of the expected demand for a company's products, and it is an indispensable element of management planning for many major company activities, e.g. marketing, production, finance, research and development, personnel, capital investment determination, purchasing, inventory and warehousing. Forecasts are valuable only if and when they help managers to achieve their objectives in an uncertain environment.

The ANN has the advantage over other approaches of arbitrary non-linear mapping capability. The non-linear map is a smooth function which gives plausible generalizations (interpolations and extrapolations). The independent validation method automatically gauges how much relevant information exists in a given data set and treats it accordingly. This makes it robust to the problems of ill-conditioning, noise and data shortage. The whole process may be automated in computer software so that people with little knowledge of either forecasting or ANNs can prepare reasonable forecasts in a relatively short space of time. If there is no information in the data, the ANN will offer a noncommittal response. As with all scientific forecasting methods, the ANN can never predict the unpredictable. It can only generate useful forecasts if it has had access to data from which it can construct an appropriate model.

4. Experimental results

In this paper, two data sets (PROD1 and PROD2) are used to evaluate the performance of ANNs sales forecasting compared with the time series smoothing methods of forecasting. The ANNs with 36 inputs for data set 1 and data set 2 were applied to the data set using V-fold validation. V-fold validation is a technique in which a number of training patterns are kept for validation purposes during training, that is, the network will not be trained on this set of data. A different set of input patterns from the training set is retained each time for the same purpose and the network performance is measured using this set. The process continues until all patterns have been used cyclically for validation.

The networks were trained to forecast from 1 month to 6 months in advance and the performance of the network was tested after training. The ANNs results are compared with the time series smoothing methods of forecasting (i.e. 3 months moving average, 12 months moving average, alpha equal to 0.5 and 0.8 exponential moving average). The network is trained to forecast from 1 month to 6 months in advance and the network performance is tested after training. The test results are compared with the time series smoothing methods of forecasting using several measures of accuracy (i.e. mean

error, mean percentage error, mean absolute deviation, mean absolute percentage error and mean squared error). Some of the results are summarized below in tables 1 and 2, for products 1 and 2 respectively.

PROD1	ME	MPE	MAD	MAPE	MSE
3MMA	-326.75	-14.00%	4210.82	32.84%	1.365e+7
12MMA	-894.83	-20.37%	*3688.82*	32.35%	1.130e+7
α = 0.5	-210.82	-13.31%	4237.10	32.02%	1.454e+7
α = 0.8	-178.47	-12.13%	4620.43	34.00%	1.705e+7
1 month	329.00	*-5.84%*	3914.21	*28.31%*	1.209e+7
2 months	125.68	-8.08%	3889.00	29.68%	*1.078e+7*
3 months	-80.03	-9.70%	4637.08	33.63%	1.782e+7
4 months	64.05	-8.92%	4273.68	32.08%	1.530e+7
5 months	-50.40	-10.37%	4456.13	33.36%	1.636e+7
6 months	*-34.77*	-13.25%	5411.26	40.62%	2.235e+7

Table 1   Results for PROD1 after calculating the forecasting performance

From table 1, the best overall results for PROD1 is 2 months ANN forecasting in advance, the second best is 1 month ANN forecasting in advance and the worst one is 6 months ANN forecasting in advance. The top three forecasts are obtained from ANN forecasting and there are three forecasts. They are the 12 months moving average, alpha equal to 0.5 exponential moving average and 5 months ANN forecasting in advance in rank 4. The ANN forecasting has thus performed better than the traditional time series methods for PROD1.

PROD2	ME	MPE	MAD	MAPE	MSE
3MMA	-652.02	-8.86%	7531.79	23.96%	4.405e+7
12MMA	-1219.23	-12.13%	7630.01	25.37%	4.286e+7
α = 0.5	-354.86	-8.11%	7594.96	23.85%	4.488e+7
α = 0.8	-358.88	-7.65%	8412.43	25.58%	5.418e+7
1 month	-202.84	-5.21%	7418.22	*22.52%*	5.147e+7
2 months	-498.96	-6.42%	8000.72	24.37%	4.619e+7
3 months	-250.73	-7.40%	8366.07	26.21%	5.126e+7
4 months	1183.17	*-1.87%*	9121.76	26.92%	5.719e+7
5 months	-288.51	-6.13%	*7349.14*	22.76%	*4.248e+7*
6 months	*30.28*	-5.28%	8484.92	25.63%	5.359e+7

Table 2   Results for PROD2 after calculating the forecasting performance

From table 2, the best overall ranking for PROD2 is 1 months ANN forecasting in advance, the second best is 5 months ANN forecasting in advance and the worst one is 4 months ANN forecasting in advance. The top two forecasts are obtained from ANN forecasting and in rank 5 there are two forecasts. They are 2 months and 5 months ANN forecasting in advance. The ANN forecasting has thus performed better than the traditional time series methods for PROD2.

5. Conclusions

In this study, two forecasting models were successfully developed using in house software and That is Silva's adaptive back-propagation was used for forecasting the two data sets (PROD1 and PROD2). The results suggest that it is sufficiently accurate to be used as a tool to develop an effective sales forecasting model. Furthermore, Silva's adaptive back-propagation algorithm was selected to develop the forecasting model [1] and the network architectures were selected by genetic algorithms [2]. Combinations of the inputs were used to forecast sales 1 month to 6 months in advance. The ANN results are compared with the time series smoothing methods of forecasting after training. The test results are compared with the time series smoothing methods of forecasting using several measures of accuracy (i.e. ME, MPE, MAPE, MAD and MSE). The results in the previous section show that the ANNs approach is likely to be able to do a better job in sales forecasting than time series smoothing methods of forecasting. At this point, it is worthwhile for us to assess the extent to which we have achieved the paper's aims.

Therefore, we conclude that the work was successfully carried out since its aims have been accomplished. We have shown that the forecasting models developed using the ANNs approach proved to perform better than traditional time series methods. The results of ANN forecast appear to be sufficiently accurate. ANNs need a long time to be properly trained during the implementation stage. In addition, the user has to equip himself/herself with adequate training and knowledge in ANNs in order to use the software package. Each product requires different configuration and learning parameters in order to obtain an optimal solution. For instance, the configuration and learning parameters for PROD1 are different from PROD2. This is still felt to be advantageous compared with the traditional approach because the results are better and the system is faster once the ANN has been trained.

6. Future trends in ANNs research

In the near future the development of truly parallel processing machines will increase the speed of operation of ANN models. Hardware implementations are already been developed by the research community but few have been fully commercialized so far. Along with ANN models, fuzzy logic is currently been explored extensively. The success of the integrated fuzzy neural networks has been reported in several studies [10-11]. The ANN is believed to overcome the knowledge acquisition problem of expert systems and fuzzy logic can

handle inaccurate, or some incomplete information problems without difficulty. Thus the integration of ANN and fuzzy logic is already proving to be a popular strategy.

Additionally, the integration of expert systems and neural networks, called connectionist expert systems [12-13] has been shown to be able to complement for the weakness of each model in some application. The fact that ANN does not have an explanation capability for the outcomes which the network derives has been a weakness of the model. However, the explanation system is a major part of an expert system. Hence the desire to integrate the ANN with the expert system so that each system can complement the other. In terms of ANN design, an object oriented approach has been proposed recently [14]. A node in an ANN model is supposed to have a simple and similar process to other nodes, it is therefore a good idea to apply object oriented design to network design. This is an attractive approach for the network design which is being explored.

Acknowledgment

References

[1] W. Schiffmann, M. Joost and R. Wernre, *'Comparison of Optimized Backpropagation Algorithms'*, European Symposium on Artificial Neural Networks ESANN'93 Proceeding, April 1993, pp. 97-103.

[2] P. Arena, R. Caponetto, L. Fortuna and M.G. Xibilia, *'M.L.P. Optimal Topology via Genetic Algorithms'*, International Conference on ANNs and GAs, Innsbruck, Austria, April 1993, pp. 670-674.

[3] S. Dutta and S. Shekhar, *'Bond Rating: A Non-Conservative Application of Neural Network'*, International Conference on Neural Networks, San Diego, CA July 24-27, 1988.

[4] E. Schoneburg, 'Stock Price Prediction using Neural Networks', Neurocomputing Journal, Vol. 2, No. 1, June 1990, pp. 17-27.

[5] A.N. Refences, M. Amema-Barac, L. Chu and S.A. Karoussos, 'Currency Exchange Rate Prediction and Neural Network Design Strategies', Neural Computing and Applications Journal, Jan 1993, pp. 46-58.

[6] J.D. Schaffer, D. Whitley and L.J. Eshelman, *'Combinations of Genetic Algorithms and Neural Networks: A Survey of the State of the Art'*, International Workshop on Combinations of Genetic Algorithms and Neural Networks, pp. 1-37, 1992.

[7] N.J. Radcliffe, *'Genetic Set Recombination and Its Application to Neural Network Topology Optimization'*, Neural Computing and Applications, Vol. 1, pp. 67-90, 1993.

[8] Y. Ikuno, H. Kawabata, Y. Shirao, M. Hirata, T. Nagahara and Y. Inagaki, *'Application of an Improved Genetic Algorithm to the Learning of Neural Networks'*, IEICE Transactions Fundamentals, Vol. E77-A, No. 4, pp. 731-735, 1994.

[9] F.M. Silva and L.B. Almeida, *'Acceleration Techniques for the Backpropagation Algorithm'*, Lecture Notes in Computer Science: Part 412, Springer-Verlag, 1990.

[10] E. Cox, 'Integrating Fuzzy Logic into Neural Nets', AI Expert, Vol. 7, No. 6, Jun 1992, pp. 43-47.

[11] F.S. Wong, P.Z. Wang, T.H. Goh and B.K. Quek, 'Fuzzy Neural Systems for Stock Selection', Financial Analysts Journal, Vol. 48, No. 1, Jan/Feb 1992, pp. 47-52.

[12] S.I. Gallant, 'Connectionist Expert Systems', Communications of the ACM, Vol. 31, No. 2, Feb 1988, pp. 152-169.

[13] G. Bradshaw, R. Fozzard and L. Ceci, 'A Connectionist Expert System that Actually Works', Advances in Neural Information Processing Systems, edited by Touretzky, D.S., 1989, Vol. 1, San Mateo, Morgan Kauafmann, pp. 248-255.

[14] J.P. Williams, 'Neural Network Design: An Object Oriented Approach', PCAI, May/June 1992, pp. 23-27.