WEATHER TIME SERIES FORECASTING USING RECURRENT FUZZY NEURAL NETWORK

Thi Nguyen[1], Lee Gordon-Brown[2], Peter Wheeler[1], and Jim Peterson[1]

[1]Centre for GIS, School of Geography and Environmental Science,
Monash University, Victoria, Australia
Email: thi.nguyenthanh@gmail.com
[2]Department of Econometrics and Business Statistics
Monash University, Victoria, Australia

ABSTRACT

Weather forecasting becomes more and more indispensable to our lives and thus many approaches have been investigated so far to meet this high demand. Conventional methods proposed in the 1980s or earlier were mostly linear models and usually applied to deal with short-range prediction. With the development of information technology, many data mining techniques have been introduced aiming to improve the power and accuracy in prediction. In this paper, three daily meteorological time-series encompassing maximum temperature, minimum temperature, and rainfall spanning from 1981 to 1990 in Melbourne, Australia were utilized to verify the prediction potential of the incorporation of fuzzy sets and neural networks by deployment of the Recurrent Fuzzy Neural Network (RFNN). Good experimental results were achieved via the application of the RFNN model to weather time series forecasting.

1. INTRODUCTION

Weather forecasts and warnings are the most important services provided by the meteorological profession. Forecasts are used by government and industry to protect life and property and to improve the efficiency of operations, and by individuals to plan a wide range of daily activities. With the ongoing availability and increasing capacity of high performance computing, various methods have been proposed to satisfy the increasing demands of weather forecasting. Numerical Weather Prediction (NWP) method was considered as the best forecasting method for the day-to-day weather changes (Kalnay *et al.*, 1990; Roads, 1986; Ghil *et al.*, 1979). Marchuk (1970) presented numerical methods for weather forecasting several days in advance, which were based on a complete system of equations of hydrodynamics and thermodynamics, taking atmospheric moisture transfer and radiational effects into account. Klein *et al.* (1959, 1974) introduced the model that is usually called "the perfect prog method". The concurrent statistical relationship between the predictand and predictors is applied to model numerical output at, say, a projection of 36 hours to get an estimate of the predictand 36 hours after the data input time for the model. Another method is the analog forecasting that consists of searching for analogs to a present or preceding situation and then predicting weather for the forthcoming period based on the similar cases in the past. The forecast of 5-day mean temperature and 10-day precipitation totals for Hungary for the next month using this approach has been reported by Toth (1989) as a solution for long-range weather forecasting. As an advance in Model Output Statistics (MOS), which is an objective weather forecasting technique, Glahn and Lowry (1972) integrated MOS and screening regression to forecast the surface weather variables such as maximum temperature, probability of precipitation, surface wind, cloud amount and conditional probability of frozen precipitation. Vislocky and Fritsch (1997) developed a prototype advanced MOS system that finished in 20th place out of 737 original entrants or better than approximately 97% of the human forecasters who entered the 1996–97 National Collegiate Weather Forecast Contest in the USA. The prototype system uses an optimal blend of aviation and nested grid model (NGM) MOS forecasts, explicit output from the NGM and Eta model (Black, 1994) guidance, and the latest surface weather observations from the forecast site.

Each of the above methods has advantages and disadvantages, and what may be appropriate for one circumstance may not be appropriate for another. Generally, the major drawbacks of these measures, however, are their inherently linear characteristic and poor performance for long-range forecasting. With the development of information technology, many data mining techniques have been introduced to improve the power and accuracy of weather forecasting.

This paper presents an application of one of the most effective methods belonging to machine learning, namely Recurrent Fuzzy Neural Network to weather forecasting. This incorporation between fuzzy set and neural network (Abraham, 2001; Buckley *et al.*, 1994; Horikawa *et al.*, 1992; Jain *et al.*, 1998; Lin *et al.*, 2004; Mandic *et al.*, 2001; Medsker *et al.*, 2001) has been investigated for prediction using three daily meteorological time series in Melbourne, Australia over the period 1981 to 1990. In trying to predict the maximum temperature at the time point (t), two experimental scenarios were explored. The first scenario is that the predicted value (predictand) depends on its values in the past, i.e. predictors are maximum temperature at time points $(t-1)$, $(t-2)$, …, $(t-n)$. The second scenario is that the predictand depends on the values of parallel time series, i.e. predictors are maximum temperature, minimum temperature, and rainfall at time point $(t-1)$. The results of experiments are stated in section 3. Firstly, we describe the RFNN model applying for weather time series forecasting in the next section.

2. RECURRENT FUZZY NEURAL NETWORK FOR FORECASTING

2.1 The RFNN configuration

The neural network model diagrammed in Figure 1 is a structure for approximating the nonlinear function $F : R^N \rightarrow R^P$ where N and P are the number of inputs and outputs, respectively. There are four layers of RFNN in total; each layer consists of from one to some nodes, i.e. neurons, which are computational units. Let us denote $u_i^{(k)}$ and $o_i^{(k)}$ as input and output values of the i^{th} node in the k^{th} layer.

Layer 1

$$o_i^{(1)} = u_i^{(1)} = x_i(t),\ i = 1 \div N$$

Layer 2

$$o_{ij}^{(2)} = \exp\left[-\frac{\left(u_{ij}^{(2)} - m_{ij}\right)^2}{\left(\sigma_{ij}\right)^2}\right],\ i = 1 \div N,\ j = 1 \div M$$

where M is the number of fuzzy rules, m_{ij} and σ_{ij} are the centre and width of the Gaussian membership function, and

$$u_{ij}^{(2)}(t) = o_i^{(1)} + \theta_{ij} o_{ij}^{(2)}(t-1)$$

where θ_{ij} is the weight of the recurrent node.

$$o_{ij}^{(2)} = \exp\left[-\frac{\left\lfloor x_i(t) + \theta_{ij} o_{ij}^{(2)}(t-1) - m_{ij} \right\rfloor^2}{(\sigma_{ij})^2}\right], i = 1 \div N, \; j = 1 \div M$$

Layer 3

Operator AND is used to multiply outputs of layer 2 together.

$$o_j^{(3)} = \prod_{i=1}^{N} o_{ij}^{(2)} = \prod_{i=1}^{N} \exp\left[-\frac{\left[x_i(t) + \theta_{ij} o_{ij}^2(t-1) - m_{ij}\right]^2}{(\sigma_{ij})^2}\right], i = 1 \div N, \; j = 1 \div M$$

Layer 4

The nodes on layer 4 undertake the defuzzification function

$$y_k = o_k^{(4)} = \sum_{j=1}^{M} u_{jk}^{(4)} w_{jk} = \sum_{j=1}^{M} o_j^{(3)} w_{jk} = \sum_{j=1}^{M} w_{jk} \prod_{i=1}^{N} \exp\left[-\frac{\left[x_i(t) + \theta_{ij} o_{ij}^2(t-1) - m_{ij}\right]^2}{(\sigma_{ij})^2}\right]$$

All four types of parameters (m_{ij}, σ_{ij}, θ_{ij} and w_{jk}) need to be trained over the whole (N + N.M + M + P) nodes of the RFNN.

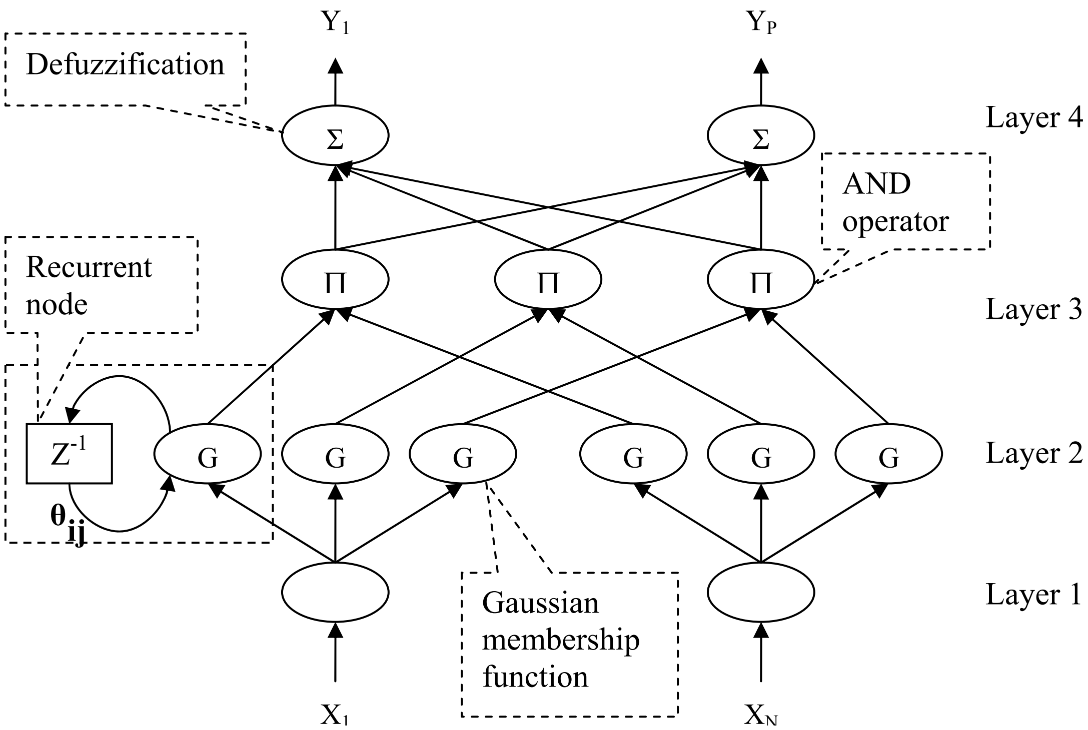

Figure 1. The four layers in the RFNN configuration. Each node in layer 2 at time point (t) contains the previous information of itself at time point (t - 1) and is termed a "recurrent" node.

2.2 Training process

Supervised gradient descent learning was utilized to tune the parameters relying upon the aim of minimizing the squared error function:

$$E(x)=\frac{1}{2}(f(x)-F(x))^2=\frac{1}{2}\left(y-o^{(4)}\right)^2$$

where $f(x)=y$ is the real value and $F(x)=o^{(4)}$ is the value computed from the RFNN.

The parameters are updated via the formula:

$$\xi(t+1)=\xi(t)-\mu_t\frac{\partial E}{\partial \xi}$$

where μ_t and $\xi(t)$ are learning rate and parameter value, respectively, at iteration *t*.

The momentum technique was also integrated in the parameters tuning process to increase the convergent speed (Qian, 1999). The learning fomula with momentum is as follows:

$$\xi(t+1)=\xi(t)-\mu_t\frac{\partial E}{\partial \xi}+\varepsilon.\Delta\xi(t)$$

where ε is momentum coefficient.

2.3 RFNN for weather forecasting

Three daily meteorological time series in Melbourne, Australia consisting of maximum temperature, minimum temperature and rainfall (Figure 2, 3, and 4, respectively) were used to verify the RFNN approach under two scenarios.

Scenario 1: The maximum temperature at time point *(t)* was considered as the function F1 of three determinants consisting of maximum temperature, minimum temperature and rainfall at time point *(t – 1)*.

$$\text{max\_temp(t)} = \text{F1}[\text{max\_temp(t - 1), min\_temp(t - 1), rainfall(t - 1)}]$$

Scenario 2: The maximum temperature at time point *(t)* was considered as the function F2 of its values in the past, at time points *(t – 1)*, *(t – 2)*, …, *(t – n)*. A broad range of *n* has been tested in order to find the best *n* for approximation. Results indicated that the higher the value of *n*, the more time-consuming is the training process whilst accuracy is not improved and that the cases of *n* from 3 to 7 show nearly same accuracy. Hence, *n = 3* is chosen as the most effective in terms of accuracy and processing time.

$$\text{max\_temp(t)} = \text{F 2}[\text{max\_temp(t - 1), max\_temp(t - 2), max\_temp(t - 3)}]$$

The 3650 data samples were partitioned into a training set and a validating set. The validating set accounted for exactly 10% of the total (365 samples at the end of the period 1981 to 1990).

Centres and widths (m_{ij} and σ_{ij}) of the Gaussian membership functions are set up based respectively on mean and standard deviation of the data series whereas remaining weights (θ_{ij} and w_{jk}) are initialized randomly.

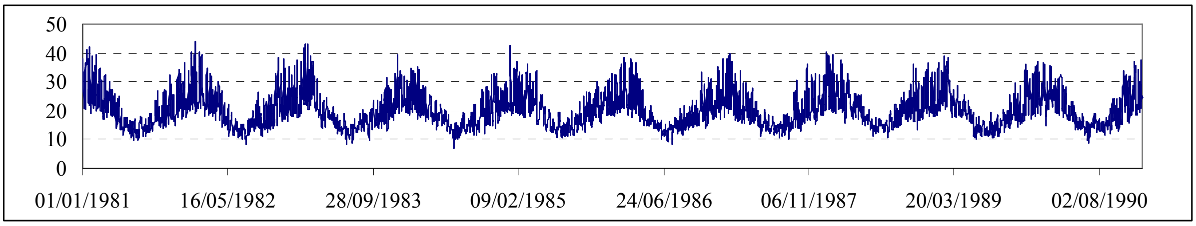

Figure 2. Daily maximum temperature measured in degrees Celsius (1981-1990).

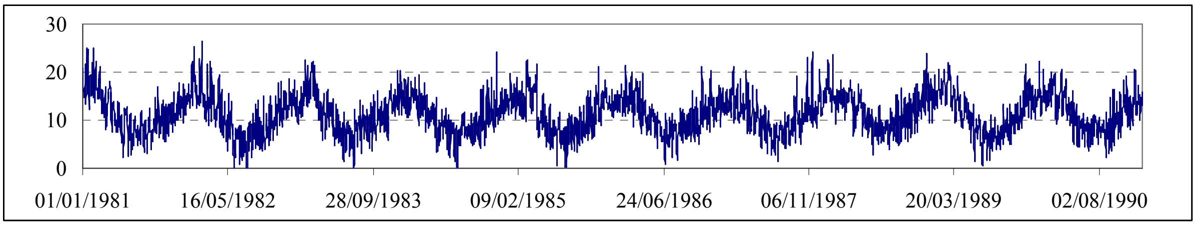

Figure 3. Daily minimum temperature measured in degrees Celsius (1981-1990).

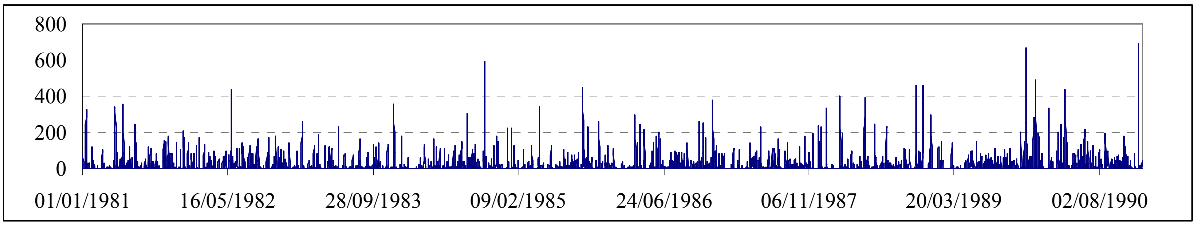

Figure 4. Daily rainfall gauged in milimetres (1981-1990).

3. RESULTS AND CONCLUSIONS

The testing results were displayed in Table 1 for both scenario 1 and scenario 2. All experiments were with the RFNN configuration: learning rate: 10^{-5}, momentum coefficient: 0.5, number of iterative cycles: 300, number of testing data samples: 365.

Table 1. Testing results performing on scenario 1 and scenario 2 for maximum temperature forecasting.

The number of nodes in				Average accuracy on testing data for		Approximate training duration
Layer 1 (N)	Layer 2 (N*M)	Layer 3$^{(*)}$ (M)	Layer 4 (P)	Scenario 1	Scenario 2	
3	60	20	1	84.37%	86.11%	2 min. 27 sec.
3	150	50	1	82.81%	83.15%	5 min. 57 sec.
3	240	80	1	76.28%	81.56%	9 min. 36 sec.

$^{(*)}$ The number of nodes in the layer 3 corresponds to the number of fuzzy rules: M.

The average accuracy is defined by the following formula:

$$average\_accuracy(\%) = \left(1 - \frac{\sum_{i=1}^{m} \frac{\left|estimated_i - observed_i\right|}{observed_i}}{m}\right) \times 100(\%)$$ where m is the number of testing samples; $estimated_i$ and $observed_i$ are estimated (predicted) and observed (real) values at the i^{th} testing data sample, respectively.

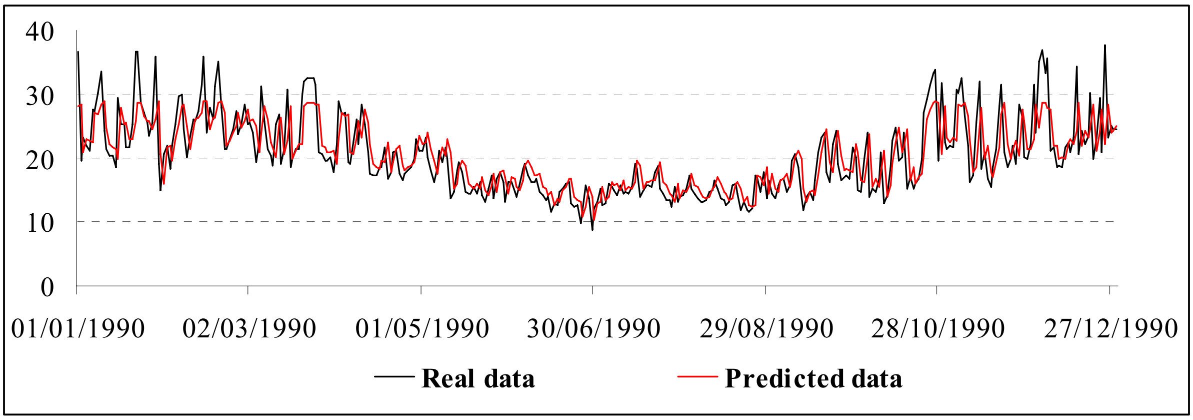

Figure 5. Scenario 2: Predicted data in comparison with real data in the case of 20 fuzzy rules for maximum temperature forecasting performing on 365 data samples at the end the period.

The training duration is in proportion to the number of fuzzy rules. More fuzzy rules, the more the number of parameters need to be tuned. The high number of fuzzy rules in RFNN is not surely congruent with the high accuracy of forecasting, especially in the case part of parameters of the model is initialized randomly.

Comparing the experimental results, scenario 2 always issued the better outcomes than scenario 1. This is contradictory with the assumption that the forecast should be more accurate if there are more relevant inputs modelled. Yet, in this circumstance, we could realize that the rainfall data is very noisy and nonlinear and hence the impact of rainfall on temperature is not useful during the deployment of the neural network model.

The approach has opened to the forecaster the new measure of benefiting from the advantages of the RFNN for prediction. Moreover, the online weather forecast system can be thought to commence using this approach since the time of constructing the model is just around some minutes to meet the expected precision. The long-term forecast is another strong point of the model when 1-year period of forecast can be reached with the nearly same accuracy over the period.

4. REFERENCES

Abraham A., 2001. Neuro-Fuzzy Systems: State-of-the-Art Modeling Techniques, *Connectionist Models of Neurons, Learning Processes, and Artificial Intelligence, Springer-Verlag Germany, Jose Mira and Alberto Prieto (Eds), Spain*, pp. 269-276.

Black, T. L., 1994. The new NMC mesoscale Eta model: Description and forecast examples. *Weather Forecasting*, vol. 9, pp. 265–278.

Buckley, J., and Hayashi, Y., 1994. Fuzzy Neural Networks: A Survey, *Fuzzy Set. Syst.*, vol. 66, no. 1, pp. 1-13.

Freeman, J. A. and Skapura, D. M. 1991. *Neural Networks: Algorithms, Applications, and Programming Techniques*. Addison-Wesley Publishing Company, Inc., USA.

Ghil, M., Halem, M., and Atlas, R., 1979. Time-Continuous Assimilation of Remote-Sounding Data and Its Effect on Weather Forecasting. *Monthly Weather Review*, vol. 107, pp. 140-171.

Glahn, H. R. and Lowry, D. A., 1972. The Use of Model Output Statistics (MOS) in Objective Weather Forecasting, *Journal of Applied Meteorology*, pp. 1203-1211.

Horikawa, A., Furuhashi, T. and Uchikawa, Y., 1992. On Fuzzy Modeling Using Fuzzy Neural Networks with the Back-Propagation Algorithm. *IEEE Transactions On Neural Networks*, vol. 3, no. 5, pp. 801-806.

Jain, L. C. and Martin, N. M., 1998. *Fusion of Neural Networks, Fuzzy Systems and Genetic Algorithms: Industrial Applications*. CRC Press LLC., Florida, USA.

Kalnay, E., Kanamitsu, M., and Baker, W. E., 1990. Global Numerical Weather Prediction at the National Meteorological Center, *Bulleting American Meteorological Society*, vol. 71, no. 10, pp. 1410-1428.

Kasabov, N., 1996. *Foundations of Neural Networks, Fuzzy Systems and Knowledge Engineering*, The MIT Press, CA, MA.

Klein, W. H. and Glahn, H. R., 1974. Forecasting Local Weather by Means of Model Output Statistics, *Bulleting American Meteorological Society*, vol. 55, no. 10, pp. 1217-1227.

Klein, W. H., Lewis, B. M. and Enger, I., 1959. Objective prediction of five-day mean temperatures during winter, *Journal of Applied Meteorology*, vol. 16, pp. 672-682.

Lee, C. H., and Teng, C. C., 2000. Identification and control of dynamic systems using recurrent fuzzy neural networks, *IEEE Trans. on Fuzzy Systems*, vol. 8, no. 4, pp. 349-366.

Lin, C. T., Chang, C. L., and Cheng, W. C., 2004. A recurrent fuzzy cellular neural network system with automatic structure and template learning, *IEEE Trans. on Circuits and Systems I*, vol. 51, no. 5, pp. 1024-1035.

Lorenz, E., 1977. An experiment in nonlinear statistical weather forecasting, *Monthly Weather Review*, vol. 105, pp. 590-602.

Madan, M. G., Liang, J., and Noriyasu, H., 2003. *Static and Dynamic Neural Networks*. John Wiley and Sons, Inc., New York, USA.

Mandic, D. P. and Chambers, J. A., 2001. *Recurrent Neural Networks for Prediction.* John Wiley & Sons, Ltd., West Sussex, England.

Marchuk, G. I., 1970. Numerical Methods of Weather Forecasting, *Oceanography and Atmospheric Science: Meteorology – Storming Media*, 396 pp.

Medsker, L. R. and Jain, L. C. 2001. *Recurrent Neural Networks: Design and Applications*. CRC Press LLC, Florida, USA.

Qian, N., 1999. On the momentum term in gradient descent learning algorithms, *Neural networks,* vol. 12, pp. 145-151.

Roads, J. O., 1986. Forecasts of time averages with a numerical weather prediction model, *Journal of Atmospheric Science*, vol. 43, pp. 871-892.

Toth, Z., 1989. Long-Range Weather Forecasting Using an Analog Approach, *Journal of Climate*, vol. 2, pp. 594-607.

Vislocky, R. L. and Fritsch, J. M., 1997. Performance of an Advanced MOS System in the 1996–97 National Collegiate Weather Forecasting Contest, *American Meteo. Society*, pp. 2851-2857.