Proceedings of the Third International Conference on Machine Learning and Cybernetics, Shanghai, 26-29 August 2004

# UNIVARIATE TIME SERIES FORECASTING WITH FUZZY CMAC

**DA-MING SHI[1], JUN-BIN GAO[2], RAVEEN TILANI[1]**

[1]School of Computer Engineering, Nanyang Technological University, Singapore 639798
[2]School of Mathematics, Statistics and Computer Science, University of New England, Australia
E-MAIL: asdmshi@ntu.edu.sg, jbgao@turing.une.edu.au

**Abstract:**

**In financial and business areas, forecasting is a necessary tool that enables decision makers to predict changes in demands, plans and sales. This paper applies a novel Fuzzy Cerebellar-Model-Articulation-Controller (FCMAC) into univariate time-series forecasting and investigates its performance in comparison to established techniques such as Single Exponential Smoothing, Holt's Linear Trend, Holt-Winter's Additive and Multiplicative methods and the Box-Jenkin's ARIMA model. Experimental results from the M3 Competition data reveal that the FCMAC model yielded lower errors for certain data sets. The conditions under which the FCMAC model emerged superior are discussed.**

**Keywords:**

**Univariate time series forecasting; neural networks; fuzzy; CMAC**

## 1. Introduction

Forecasts can be used to enhance a variety of purposes in an organization, including production planning, budgeting, sales quota setting, and human resource planning [1]. The importance of sales forecasting for a firm has often been emphasized as more and more organizations recognize the importance of formal forecasting. Without a forecast, firms in the short run can only respond retroactively, which may result in lost orders, poor service and inefficient utilization of resources. In the long run, this misallocation of resources may have drastic consequences on the firm's ability to survive. Much literature has been written describing the attributes and factors of good forecasting practices and a large number of empirical studies have been conducted on forecasting in general or sales forecasting in particular [2].

The importance of forecasting has seen the emergence of companies specialising in forecasting packages and software. These range from stand-alone, specialized packages to modules in enterprise management systems. Many of these packages use popular forecasting methods and models. The methods include the Simple Exponential Smoothing method (SES), Holt's Linear Exponential Smoothing method (HLES), Holt-Winter's method with Additive Seasonality (HWA) and Multiplicative Seasonality (HWM). A poplar forecasting model is the Box-Jenkin's Autoregressive Integrated Moving Average (ARIMA) model [3].

Newer forecasting techniques and innovative modifications of present techniques have also emerged. These include pattern matching of historical data [4], which relies on old structures of data may be used for matching with current structures to generate a future prediction. Segmentation of time-series data [5] to enable further manipulation and extraction of significant information have also been explored successfully. These new techniques have been compared to the classical techniques and have yielded encouraging results indeed.

Emerging developments in neural networks for forecasting have also been observed, as neural networks can easily model any type of parametric or non-parametric process including automatically and optimally transforming the input data [6]. Encouraging results have emerged from various comparisons of neural networks with the above mentioned classical techniques [7,8]. However neural networks suffer certain drawbacks. Firstly it is not known if a training set is adequate or not and secondly, the network knowledge is not easily extracted and comprehended.

With this, important developments in neural network modelling in recent years include incorporating a fuzzy system into a neural net [9]. Based on the concepts of fuzzy sets and fuzzy logic, fuzzy systems encode the linguistics samples in a designated numerical matrix, which links input to output through fuzzy membership functions and sets of fuzzy rules. Therefore fuzzy rules are compact, efficient representations of human knowledge. Fuzzy rules therefore provide a human-like thinking ability which allows expert-knowledge to be incorporated into a system. This, coupled with the learning capabilities of neural networks harnesses the advantages of neural networks and fuzzy rules, leading to the advent of a fuzzy neural system [10]. The Cerebellar Model Articulation Controller (CMAC) first proposed by Albus [11,12], is a type of associative

**0-7803-8403-2/04/$20.00 ©2004 IEEE**

memory network that models how a human cerebellum would take inputs, organize its memory and compute outputs. The CMAC system has the advantages of fast learning, simple computation and local generalization, and can be realized by high-speed hardware. The main disadvantage of CMAC is that the input parameters grow exponentially with the input variables. To overcome this shortcoming, some researchers has combined fuzzy set theory with CMAC (FCMAC) [13].

This paper focuses on univariate time series forecasting which only uses historical data to generate a forecast. The structure of this paper is outlined as follows. In section 2, a brief introduction on forecasting techniques, such as SES, HLES, HWA, HWM and ARIMA, is given. In section 3, we describe the architecture and the construction of our proposed FCMAC. In section 4, our experiments are conducted on the well-known "M3-competition" data [14]. Section 5 provides our conclusions.

## 2. Forecasting Techniques

By studying the past behaviour of time series data, certain features may be identified and these may help in choosing an accurate forecasting method or model. Generally, plotting historical time series data may reflect one or more of the following features: (1) Trends - A trend is a gradual upward or downward shift in the level of the series or the tendency of the series values to increase or decrease over time. (2) Seasonal and non-seasonal cycles - A seasonal cycle is a repetitive, predictable pattern in the series values. A non-seasonal cycle is a repetitive, possibly unpredictable, pattern in the series values. (3) Pulses and steps - Many series experience abrupt changes in level. They generally come in two types: a sudden, temporary shift, or pulse in the series level; or a sudden, permanent shift, or step in the series level. (4) Intermittent demand - Some time series data is not complete due to gaps in recorded activity. This pattern is referred to as "intermittent demand" because it appears most often in series of purchase histories for high-ticket or uncommonly bought items.

The classical forecasting techniques used in our experiment are, SES, HLES, HWA, HWM and ARIMA. These popular techniques will provide us with a credible platform upon which we can compare the performance of our model.

Single Exponential Smoothing (SES) Method is a method in which smaller weights are assigned to older historical data and heavier weights to more recent data to reflect the relative significance the more recent data may have on the forecast. Simple exponential smoothing assumes that there is no trend or seasonal aspects in the data and that the level of the series changes slowly over time. The expression for SES is as follows [15]:

$$F_{t+1} = (1-\alpha)^t F_1 + \alpha \sum_{j=0}^{t-1} (1-\alpha)^j Y_{t-j} \tag{1}$$

where $F_{t+1}$ is the forecast at time t+1, $Y_{t-j}$ is the actual value at time t-j and $\alpha$ is the smoothing parameter.

Suppose that the data series is non-seasonal but does display a trend. The HLES method extends the SES method to estimate both the current level and the current trend. The expressions for HLES are as follows:

$$\begin{aligned} L_t &= \alpha Y_t + (1-\alpha)(L_{t-1} + b_{t-1}) \\ b_t &= \beta(L_t - L_{t-1}) + (1-\beta) b_{t-1} \\ F_{t+m} &= L_t + b_t m \end{aligned} \tag{2}$$

Here $L_t$ and $b_t$ are exponentially smoothed estimates of the level and linear trend of the series at time *t* respectively, whilst $F_{t+m}$ is the linear forecast from *t* for a forecast horizon *m*, and α and β are the smoothing parameters.

The Holt-Winter's Method with Additive (HWA) and Multiplicative (HWM) Seasonality methods cater to data with both trend and seasonal components. The seasonality is multiplicative if the magnitude of the seasonal variation increases with an increase in the mean level of the time series. It is additive if the seasonal effect does not depend on the current mean level of the time series and can simply be added or subtracted from a forecast that depends only on level and trend. The expressions for each type are as follows:

Additive seasonality,

$$\begin{aligned} L_t &= \alpha(Y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + b_{t-1}) \\ b_t &= \beta(L_t - L_{t-1}) + (1-\beta) b_{t-1} \\ S_t &= \gamma(Y_t - L_t) + (1-\gamma) S_{t-s} \\ F_{t+m} &= L_t + b_t m + S_{t-s+m} \end{aligned} \tag{3}$$

Multiplicative seasonality,

$$\begin{aligned} L_t &= \alpha \frac{Y_t}{S_{t-s}} + (1-\alpha)(L_{t-1} + b_{t-1}) \\ b_t &= \beta(L_t - L_{t-1}) + (1-\beta) b_{t-1} \\ S_t &= \gamma \frac{Y_t}{L_t} + (1-\gamma) S_{t-s} \\ F_{t+m} &= (L_t + b_t m) S_{t-s+m} \end{aligned} \tag{4}$$

where $S_t$ is estimate of the seasonality of the series at time $t$, the factor $s$ is the number of periods in one cycle of seasons (e.g. number of months or quarters in a year) and α, γ and β are the smoothing parameters.

The ARIMA processes are a class of stochastic processes used to analyze time series data. ARIMA forecasts are based on linear functions of the sample observations. The aim is to find the simplest models that provide an adequate description of the observed data. This is sometimes known as the principle of parsimony. Each ARIMA process has three parts, namely the Autoregressive (AR) part, Integrated (I) and the Moving Average (MA) part [15]. The (AR) part of the model describes how each observation is a function of the previous observations. There are 3 main phases involved in identifying an ARIMA model prior to a forecast being made. In the first phase, the identification phase, the data is prepared by making it stationary and stabilizing its variance. The appropriate model is then selected after examining the autocorrelation and partial autocorrelation plots of the data. In the estimation and testing phase, the model parameters are estimated by maximizing the likelihood with respect to the parameters. The parameters are then tested by checking if the residuals are of white noise. If so, the model can be used to generate the forecasts. Otherwise, the residuals contain a certain structure that should be studied and refined in the first phase.

## 3. Fuzzy CMAC Model

The structure of the FCMAC neural network is shown in Figure 1. The input vectors in the input space S are a number of sensors in real world. The input space consists of all possible input vectors. FCMAC then maps the input vector into points in the associative memory A after fuzzification. As shown in Fig. 1, two close inputs will have overlaps in A and the closer the inputs, the more overlapped they will be in A. Likewise, two distant inputs will have no overlap in A. Since the practical input space is extremely large, in order to reduce the memory requirement, A is mapped onto a much smaller physical memory in P through hash coding. Hence, any input presented to FCMAC will generate same number of physical memory locations as the number of points in A. The output Y will then be the summation of the content of the memory locations in P.

It is shown that the associative mapping within the CMAC network assures nearby points in the input space generalize while distant points not. Moreover, since the mapping from P to O is linear but from S to P is nonlinear, the nonlinear nature of the CMAC network performs a fixed nonlinear mapping from the input vector to a many-dimensional output vector.

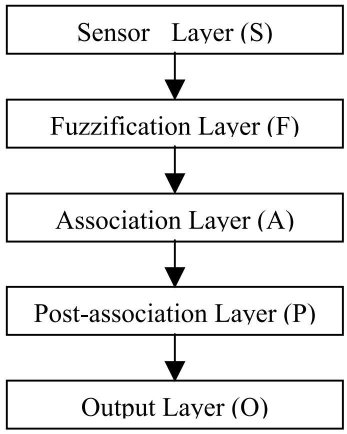


Figure 1. FCMAC Model

CMAC has the advantages of local generalisation, output superposition, easy hardware implementation, incremental learning and faster learning because of the limited number of computations per cycle. However, CMAC suffers from inherent disadvantages like its inefficiency in storing data storage, since the memory size required grows exponentially with respect to the number of input variables and its weak performance in classifying inputs which are similar and highly overlapping.

The motivation of advancing classical CMAC to FCMAC is to increase the learning capability for the model. The introduction of fuzzy membership in respective fields has the effect of smoothing the network output and increasing the approximation ability in function approximation. The FCMAC structure also reduces the memory requirement by a great deal as compared to the original CMAC. Given the input data $x^o$, the output of the simplified fuzzy reasoning $y^\tau$ is derived by the following equation:

$$\omega^p = \prod_{j=1}^{n} \mu_{A_j\,p}(x_j^o), \tag{5}$$

$$y^\tau = \frac{\sum_{p=1}^{N} \omega^p \times w^p}{\sum_{p=1}^{N} \omega^p} \tag{6}$$

where $\mu_{A_j\,p}(x_j^o)$ is a membership value, and $\omega^p$ is the total membership value of the antecedent part.

Our proposed FCMAC model has been fuzzified with the Truth Value Restriction (TVR) scheme. It uses clustering for the self-organizing phase and the back-propagation learning algorithm for the parameter-learning phase. Similar to CMAC, LMS approach is adopted to train the FCMAC-TVR network. The weights between the post-association layer and the response unit are initialized to zero and then updated through training. The output for the input data $x_i$, is derived by

$$y_i^{\tau} = \sum_{k=1}^{M} y_i^{\tau(k)} \quad (7)$$

where $y_i^{\tau(k)}$ is the output of the $k^{th}$ layer for the input data $x_i$.

The updating of the weights is stopped when the error converges to a predefined threshold.

## 4. Experimental Results

We are now in the position to compare the performance of the FCMAC network with the classical forecasting techniques. Our experiments were conducted using benchmark data from the M-3 competition data series. The M-3 Competition was conducted in 1997 by the *International Journal of Forecasting*. It compared a range of forecasting techniques across a range of measures on a holdout set. 3003 sets of historic time series data were collected to cover as wide of a range of data types as possible (e.g., microeconomic, macroeconomic, industrial, financial and demographic) and included monthly, quarterly and annual series. In order to ensure that enough data were available to develop an adequate forecasting model it was decided to have a minimum number of observations for each type of data. This minimum was set as 14 observations for yearly series (the median length for the 645 yearly series is 19 observations), 16 for quarterly (the median length for the 756 quarterly series is 44 observations), 48 for monthly (the median length for the 1428 monthly series is 115 observations) and 60 for ‘other’ series (the median length for the 174 ‘other’ series is 63 observations). The competition attracted forecasting experts from academic institutions such as the Wharton School, Case Western Reserve, INSEAD and the Imperial College , who tested their forecasting techniques with this data set. The relative performances were then tabulated and published in reputable journals. The M-3 competition data set has since been widely available and used by researches to analyse the performance of forecasting techniques.

The experiment was conducted using a fully automated program designed to obtain the ideal parameters of each method and model, based on a holdout set. Sample data sets from the M-3 series were taken to represent monthly, quarterly and annual historical data. In addition, a series of forecast horizons for each time aspect were also included. The results are as follows:

There are a variety of accuracy measures available in the forecasting literature, the frequently used measures are Mean Absolute Deviation (MAD), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). These error statistics were used in our study. The results are as follows:

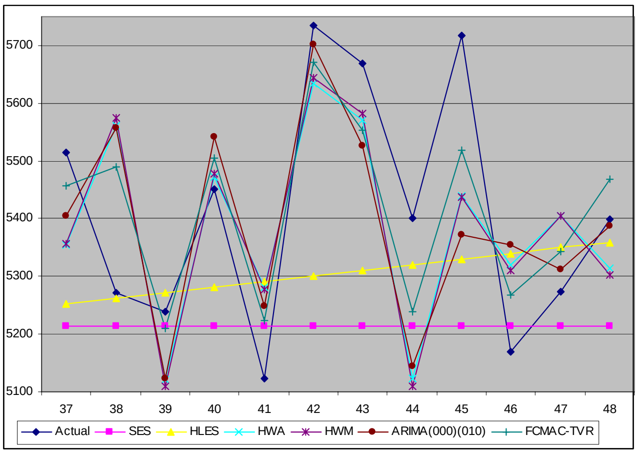


Figure 2. Performance over holdout data of 12 months for series N2071 at time origin 36

Figure 2 indicates that the proposed FCMAC-TVR model outperforms all the other methods in all forecasting horizons. The HWA method comes out the second best due to the fact that the simpler methods of SES and HLES cannot do very well with long forecast horizons. Why the HWA performs better than the HWM is because of the additive seasonality component inherent in the test data. The ARIMA model performs better than SES and HLES, as it is able to model a wider range of data patterns and even has its equivalent to SES and HLES. Over the respective holdout periods of 12 values for each figure, it is obvious that the data pattern does not show any stationarity in mean or variance or a very significant semblance of a consistent trend or seasonality. Under such circumstances, the novel approach of clustering inputs in the FCMAC-TVR model resulted in a model that can better adapt to irregular patterns in the data as compared to the classical techniques. This is especially so when the data undergoes a differencing and a log-transformation prior to being input into the system. However if the data exhibits significant and consistent characteristics (stationarity, trend and seasonality), the FCMAC-TVR's performance is unlikely to have a very significant superiority in performance.

## 5. Conclusions

This paper introduces a novel FCMAC with TVR inference scheme for univariate time series forecasting. This neural network has the characteristic of high learning speed and localization. TVR inference scheme makes the system more transparent and less rigid. It also gives the network a consistent rule base and a strong theoretical foundation. The proposed model uses DIC for self-organizing phase and the back-propagation learning algorithm for the parameter-learning phase. Our experiment investigates the performance of the FCMAC-TVR model in comparison to popular time-series forecasting techniques such as SES, HLES, HWA, HWM and ARIMA. The experiment was conducted using the M3 competition data and the results showed that the FCMAC-TVR model outperforms the other techniques.

## References


[1] J.T. Mentzer and J.E. Cox, Jr., Familiarity, application and performance of sales forecasting techniques, *Journal of Forecasting*, **3:** 27–36, 1984.

[2] H. Winklhofer, A. Diamantopoulos and S. F. Witt, Forecasting practice: A review of the empirical literature and agenda for future research, *International Journal of Forecasting,* 12: 193–221, 1996.

[3] Box G E P and Jenkins G M, "*Time Series Analysis*, Forecasting and Control", San Francisco: Holden Day, 1970.

[4] S. Singh and E. Stuart, *"A pattern matching tool for time-series forecasting",* Proceedings of the International Conference on Pattern Recognition, Vol.1, Iss., Pages:103-105, 1998

[5] F. Chung, T. Fu, R. Luk and V. Ng, *"Evolutionary time series segmentation for stock data mining",* Proceedings of the IEEE International Conference on Data Mining (ICDM), Vol., Iss., Pages: 83- 90, 2002.

[6] S. D. Balkin and J. K. Ord, *"Automatic neural network modeling for univariate time series",* International Journal of Forecasting, Vol. 16, Issue 4, Pages 509-515, 2000.

[7] C. Chu and G. P. Zhang, *"A comparative study of linear and nonlinear models for aggregate retail sales forecasting",* International Journal of Production Economics, Volume 86, Issue 3, Pages 217-231, 2003.

[8] G. Zhang, B. Patuwo and M. Hu, *"Forecasting with artificial neural networks: the state of the art",* International Journal of Forecasting, Vol. 14, 35-62, 1998.

[9] C. Wen and M. Yao, *"A combination of traditional time series forecasting models with fuzzy learning neural networks",* Proceedings of the International Conference on Machine Learning and Cybernetics, Vol. 1, Pages 21- 23, 2002.

[10] T. Frantti and P. Mähönen, *"Fuzzy logic-based forecasting model",* Engineering Applications of Artificial Intelligence, Volume 14, Issue 2, Pages 189-201, 2001.

[11] J.S. Albus, *"A new approach to manipulator control: the cerebellar model articulation controller (CMAC)",* Trans. Of ASME Journal of Dynamic Systems, Measurements, and Control, Vol. 97, No.3, pp. 220-227, 1975.

[12] J.S. Albus, "*Data storage in the cerebellar model of articulation controller (CMAC)",* Trans. Of ASME Journal of Dynamic Systems, Measurements, and Control, Vol 97, No. 3, pp. 228-233, 1975.

[13] J. Ozawa, I. Hayashi and N. Wakami, "*Formulation of CMAC-Fuzzy system*", IEEE International Conference on Fuzzy Systems, pp. 1179-1186, San Diego, 1992.

[14] M. Hibon M. and S. Makridakis, "*The M3 Competition: results, conclusions and implications",* International Journal of Forecasting, Volume 16, Issue 4, Pages 451-476, 2000.

[15] S. Makridakis, S.C. Wheelwright and R.J. Hyndman, *"Forecasting methods and applications (3rd ed.)",* New York: Wiley, 1998