Time-Series Prediction Using Self-Organising Mixture Autoregressive Network

He Ni and Hujun Yin

School of Electrical and Electronic Engineering, University of Manchester
he.ni@manchester.ac.uk, hujun.yin@manchester.ac.uk

Abstract. In the past few years, various variants of the self-organising map (SOM) have been proposed to extend its ability for modelling time-series or temporal sequence. Most of them, however, have little connection to, or are over-simplified, autoregressive (AR) models. In this paper, a new extension termed, self-organising mixture autoregressive (SOMAR) network is proposed to topologically cluster time-series segments into underlying generating AR models. It uses autocorrelation values as the similarity measure between the model and the time-series segments. Such networks can be used for modelling nonstationary time-series. Experiments on predicting artificial time-series (Mackey-Glass) and real-world data (foreign exchange rates) are presented and results show that the proposed SOMAR network is a viable and superior to other SOM-based approaches.

1 Introduction

Exchange rate forecasting has always been a challenging area of research that has received a great deal of attention. Since the break up of the Bretton-Woods system in 1973, the trend analysis in spot foreign exchange rates has been a recurrent theme among statisticians and econometricians during the last two decades.

A fundamental way is to use the economic theory to underline the structural relations between exchange rate and other variables and to use statistical methods to identify the correlations between the past data and future moves. Researchers have devoted a great deal of effort on these techniques in order to beat the random walk model. However, these econometric and time-series techniques cannot even outperforms the simplest random walk [1]. The reason is that most of the econometric models are linear and used under specific or strict assumptions. For instance, autoregressive maving average (ARMA) models assume a linear relationship between the current value of the variables and previous values of the variable and error terms. The mean and variance of variables need to be a constant overtime.

Due to the recent advances in computational intelligence and computer power, nonparametric models have been used extensively in the last few years with various successes. Exchange rate forecasted by Artificial Neural Networks (ANNs) provide strong evidence in term of out-of-sample forecasting achievements. Many comparison studies show that ANNs significantly outperform linear ARMA model and native random walk model [2–4]. The most widely used techniques so far are the multilayer perceptron (MLP), radial basis function (RBF) networks and recurrent networks. As a regressive method, support vector machines (SVM) have been proposed as a good alternative for MLP in time-series forecasting. SVMs are established on the theory of the structural risk minimisation principle.

The main problem in modelling financial time-series is their non-stationarity. That is the mean and variance of the time-series are changing over time, which implies that the variables switch their dynamics in different regions. It is particularly true in exchange rates due to the amount of nonconstant "information flow". Empirical studies [5] show that the distribution of daily returns[1] is approximately symmetric and leptokurtic (i.e., heavy tailed). One possible explanation for the heavy tailed distribution is that samples are independently distributed as a normal distribution whose mean and variance change over time. Many others argued that observed returns come from a mixture of normal distributions [6, 7]. It is not convincing for a single model to capture the dynamics of the entire time-series.

It is reasonable to assume that a time-series locally is a homogeneous model; at least it is true for most cases. A potential solution can be found using the "divide-and-conquer" principle, in which the entire model is divided into several smaller ones [4]. The solutions are then combined to make the final solution. The prediction is thus made only by the best fit local model.

Self-Organising Map (SOM) can be used to partition the input data to smaller regions by associating input data with their unique best-matching units. The area in the input space for which the reference vector is called Voronoi tesselation. Voronoi tesselations partition the input space into disjoint sets. Models can be created by locally fitting to the specific Voronoi tesselations. The topological relationship of local models are maintained as the pre-determined lattice. There are various successes on different applications. For example, Dablemont [8] applied SOM based local models with RBF network as regressor to predict the returns of the DAX30 index. Liu and Xu [9] used SOM based local models to perform PCA on the data from multi-modes. Cao [4] proposed a SVM experts system, which is also based on SOM local models, to predict time-series.

For the SOM to be used for modelling time-series, the consective input points have to be grouped into vectors to form temporal context by means of a window of a pre-fixed length. The information in between the vectors is however lost. Recently, increasing interest arises in the SOM for time-series or sequence processing. Typical methods include Temporal Kohonen Map (TKM) [10], the recurrent SOM (RSOM) [11].

Lampinen and Oja proposed a method based on the SOM, where every unit represents an AR model with its reference vector as the model parameters [12]. The experiments conduced have shown the model can learn to distinguish textures from images. The method in fact is a multiple AR model with the parameters of component models forming topological orders.

[1] A simple logarithm difference transform

Here, we propose a similar multi-regressive model with a different winner selection rule to reflect the characteristics of homogeneous time-series.

The rest of paper is organised as follows. In section 2, we describe the proposed methodology. Section 3 will present the application of the proposed methodology for prediction of exchange rate. Finally, conclusions will be summarised in Section 4.

2 Methodology

The problem of predicting future value of a stochastic process is closely related to the problem of estimating the unknown parameters of a regressive model. We start from identifying parameters of artificial autoregressive(AR) models.

The target process is assumed to be generated by several independent stationary discrete autoregressive processes. It has many fields of applications, especially in econometrics and automatic control. A number of studies recent focus on modeling such non-stationary process. The model is based on the assumption that the underly process consists of several independent stationary AR processes, referred to as local models. The model can be considered as a mixture of these independent local models, or is regarded to be from one of these local models at a time. Such a model can be expressed as,

$$F(x_t|\digamma_{t-1}) = \sum_{i=1}^{K} \beta_{(i,\mathbf{x})} \Phi_i(x_t - \phi_{i0} - \phi_{i1}x_{t-1} - \ldots - \phi_{ip_i}x_{t-m_i}). \quad (1)$$

where Φ_i is i-th local AR model. In the mixture AR (MAR) model, $\beta_{(i,\mathbf{x})}$ are the mixing parameters; and in the latter case, considered in this paper, $\beta_{(i,\mathbf{x})}$ are selection functions, given as,

$$\beta_{(i,\mathbf{x})} = \begin{cases} 1 & \text{if } \mathbf{x} \in \Phi_i \\ 0 & \text{else} \end{cases} \quad (2)$$

where input vector $\mathbf{x}(t) = [x(t), x(t-1), \ldots, x(t-m_i)]^T$, $\digamma_{t-1}$ represents the information up to time $t-1$, $\{\phi_i\}$ are parameters, K is the number of AR processes and m_i is the order of AR process i.

2.1 Lampinen and Oja's Self-Organising AR models

Lampinen and Oja proposed a method called self-organizing AR map (SOAR) based on a self-organizing map of "neural" units for unsupervised segmentation and classification of 1D and 2D signals [12]. The SOAR models index each neural unit by i and each with weight vector w_i which signifying an AR process. The prediction is based on,

The error $e(t) = x(t+1) - \mathbf{x}^T\mathbf{w}$ is further smoothed by an exponential average over the recent estimation errors v_i has been used instead of immediate estimation errors e_i

$$v_i' = \beta e_i(t) + (1-\beta)v_i. \quad (3)$$

where β is a smoothing factor. The best matching unit is the one with smallest v'_i.

The winner and its neighborhood units update their weights according to

$$\mathbf{w}_i(t) = \mathbf{w}_i(t-1) + g(r)e(t)\mathbf{x}(t). \tag{4}$$

where $g(r)$ is a linearly decreasing adaptation rate. The model is shown working well in segmenting the image into texture classes, without priori knowledge about the number of classes or the class models. The authors also presented good experimental results on 1D artificial signal and 2D textures.

However the performance of the SOAR model in finding underlying AR processes is hampered by the inaccuracy and volatile error terms (i.e. the winning rule), despite being smoothed. It cannot guarantee a good estimation of the parameters of the underlying process. Fig. 1 shows a divergence of such a model for an AR(2) process. The initial weights were set randomly (upper figure) or to the true parameters (lower figure).

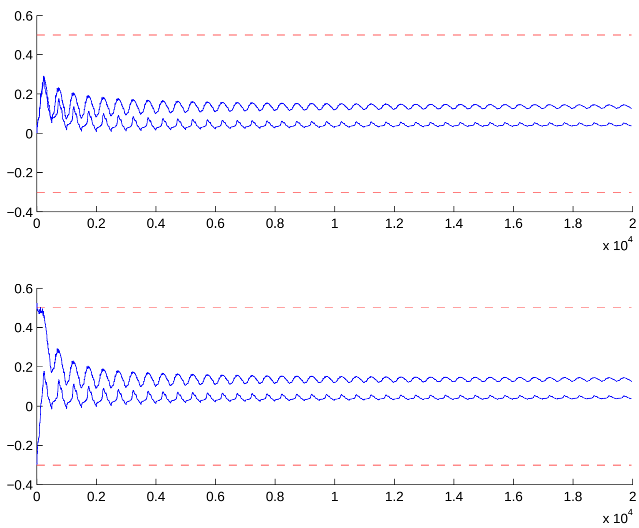

Fig. 1: Estimates of an AR(2) parameters by the SOAR with initial weights set randomly (upper panel), and to the true parameters (lower panel).

2.2 Self-Organising Mixture AR models (SOMAR)

As we assume that the stochastic process is characterised by white noise corruption. As a sufficient condition, the modeling error should be close to the white noise if the modelling is following a "correct" path. Therefore, we investigate the autocorrelation of the error instead of the error itself. In order to obtain sufficient estimation information, we hereby use a small batch input or a patch.

The modelling error is a discrete time-series of length p, the batch size, $\{e(1), e(2), \ldots, e(p)\}$, with mean μ and variance σ^2, an estimate of the auto-correlation coefficient $R(k)$ at lag k can be obtained as

$$R(k) = \frac{1}{(p-k)\sigma^2} \sum_{t=1}^{p-k} (e(t) - \mu)(e(t+k) - \mu). \tag{5}$$

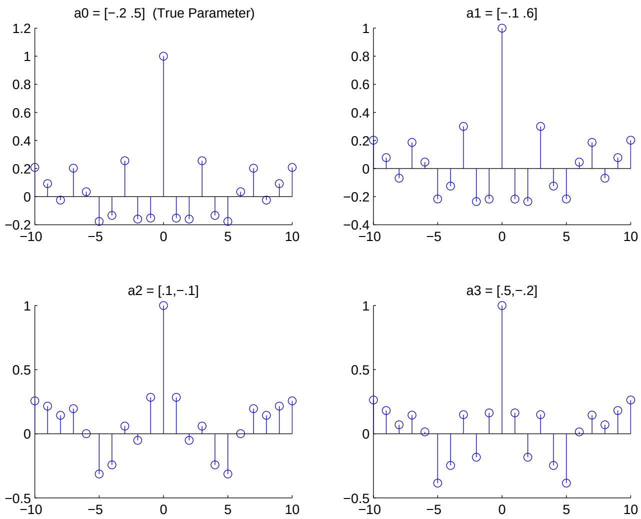

Fig. 2: Autocorrelation of the modelling errors for models $\mathbf{a}_0$ - $\mathbf{a}_3$.

Fig. 2 shows the autocorrelations of a set of the modelling errors from a patch of 20 points. The generating parameters are $\mathbf{a}_0$=[-.2, .5]. We randomly test it on three sets of parameters $\mathbf{a}_1$=[-.1, .6], $\mathbf{a}_2$=[.1, -.1] and $\mathbf{a}_3$=[.5, -.2], their correlations are plotted in Fig. 2.

Here we use the sum (of the absolute value) of autocorrelation coefficients (SAC) as the similarity measure,

$$v_i = \sum_{j=-p}^{p} |R_i(j)|. \tag{6}$$

The SAC values for these four cases are 3.8832($\mathbf{a}_0$), 4.3504($\mathbf{a}_1$), 4.5224($\mathbf{a}_2$), and 4.5963($\mathbf{a}_3$) respectively. As we can see that model $\mathbf{a}_1$ is closer to model $\mathbf{a}_0$ that to models $\mathbf{a}_2$ and $\mathbf{a}_3$, so are their SAC values.

In the proposed method, firstly a fixed number of consecutive input vectors are used to make a patch input. Analogy to the SOM algorithm, we choose the winner for that patch input according to the SAC, $v_i, i = 1, 2, \ldots, N$, i is the index of local model and N is the total number of the local models.

Then the winner and its neighbours adapt their weights by

$$\mathbf{w}_i(t) = \mathbf{w}_i(t-1) + h(r,t)\eta(t)e(t)\mathbf{x}(t). \tag{7}$$

where $h(r,t)$ is the neighborhood function and $\eta(t)$ is a decreasing adaption parameter. Gaussian neighborhood function is used, and linearly decreasing learning rate is used,

$$h(r,t) = e^{-\left(\frac{r}{2\delta(t)}\right)^2}. \tag{8}$$

$$\eta(t) = \eta_0 \frac{\tau_0}{\tau_0 + \tau_1 t + \tau_2 t^2}. \tag{9}$$

The neighborhood function is a useful feature for SOM alike techniques for avoiding the training process being trapped to local minima, and for forming topology among the nodes.

Here we show a simple example. In total 1,250 consecutive points were constructed by two AR(2) processes $\mathbf{a}_1 = [.5, -.2]$ and $\mathbf{a}_2 = [.4, -.3]$. The signal consists of 5 consequent 250 point long segments. Each segment was generated randomly by one of those two AR processes. The training set is the first 1000 points, shown in Fig. 3 and the testing set is the other 250 points. The sets were divided into 50-point patches.

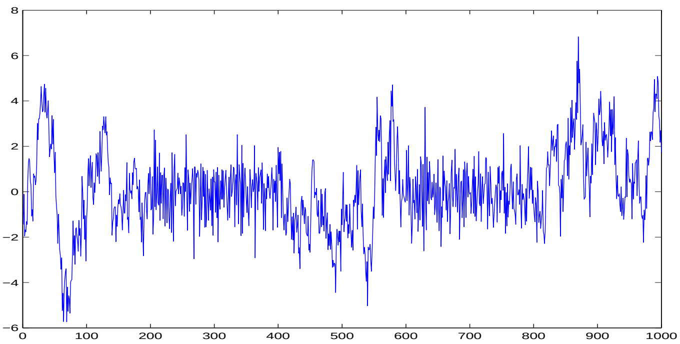

Fig. 3: Training set: 1,000 consecutive points generated by two AR(2) processes $\mathbf{a}_1 = [0.5, -0.2]$ and $\mathbf{a}_2 = [0.4, -0.3]$.

The training set was used to train the SOMAR network, the results of the weights are shown in Fig. 4. The trained SOMAR network was tested on the testing set. The results of prediction is illustrated on Fig. 5.

3 Experimental Results

In the section, we present experiments on the artificial data (Mackey-Glass data) and the foreign exchange rate data in respect to the ability of the propose method in characterising the dynamics of non-linear, non-stationary time-series.

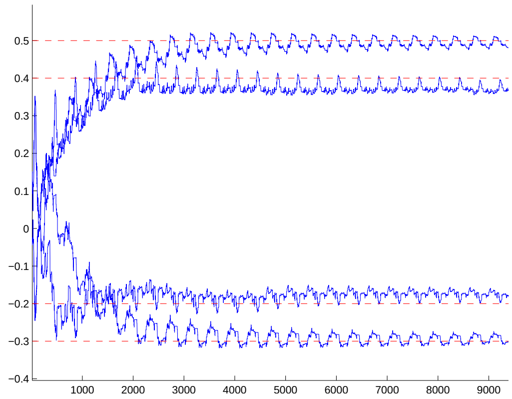

Fig. 4: Parameter estimation of two AR(2) processes $\mathbf{a}_1 = [0.5, -0.2]$ and $\mathbf{a}_2 = [0.4, -0.3]$.

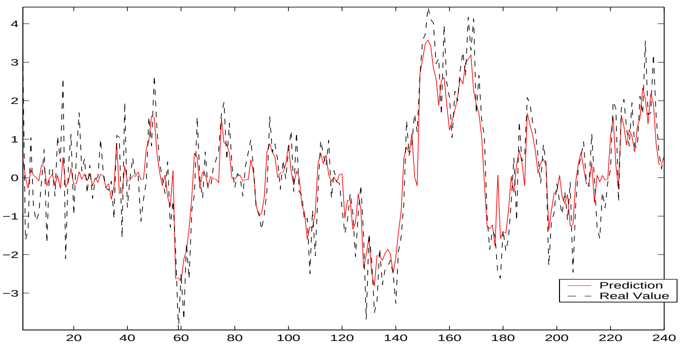

Fig. 5: The dash line represents the original data points, the red solid line represents the prediction by the proposed SOMAR network.

3.1 Mackey-Glass data

To further investigate the capabilities of the proposed SOMAR network, we apply it on a consecutive 600 points Mackey-Glass data, a dynamic system defined by the differential equation:

$$\frac{dx}{dt} = \beta x(t) + \frac{\alpha x(t-\delta)}{1 + x(t-\delta)^{10}}. \tag{10}$$

with the parameter values $\delta = 17, \alpha = 0.2, \beta = -0.1$. We assume the Mackey-Glass data consists of a number of unknown AR processes. In this experiment, the input is the Mackey-Glass series grouped in every 15 points $\mathbf{x}(i) = [x(i), x(i+1), \ldots, x(i+14)]$. We prefixed the order of AR process to 14 in favor of the results of BIC in a previous study [13]. Experiments with other value have been implemented without any significant difference. The result of prediction of Mackey-Glass data shows on Fig. 6.

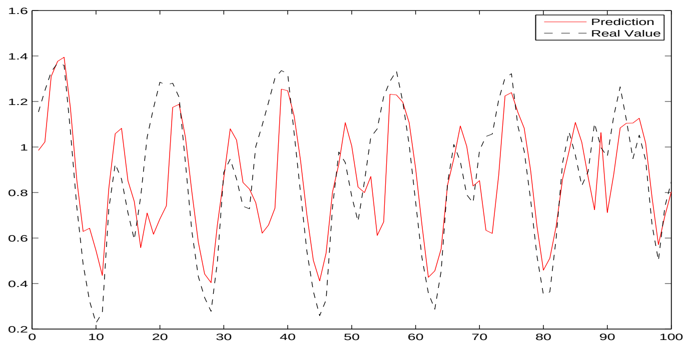

Fig. 6: The dash line represents the original data points, the red solid line represents the prediction by the SOMAR network on Mackey-Glass data.

3.2 Foreign exchange rate data

The data was retrieved from the PACIFIC Exchange Rate Service provided by W. Antwiler at UBCs Sauder School of Business. It consists 15 years' daily exchange rates excluding weekends and bank holidays when currency markets were closed. The proposed SOMAR network was trained on 3,000 consecutive data points and the performance of prediction was tested on the following 200 data points. Both the training and testing sets were windowed with the length of 15 points to form input vectors.

For a comparison with other SOM-based methods, we conducted two types of tests as follows.

Predicted FX return The correct prediction percentage, which is a criterion to check whether the prediction is made on the right direction (i.e. we calculate how many percents predicted returns[2] have the same signs as their corresponding actual returns), shown in Fig. 7.

Predicted FX price Mean-Square-Error between the testing exchange rates and predicted ones, shown in Fig. 8.

Table 1: Overall predicted FX returns and prices of various methods on the foreign exchange rate data.

	SOMAR	SOAR	VSOM	RSOM
FX return (%)	66.30	54.01	51.00	51.80
FX price	0.0450	0.0601	0.0625	0.0695

[2] We applied the price-return convert(i.e. $x'_t = \ln \frac{x_{t+1}}{x_t}$ here the x_t is the scalar values of the original data at the time t) to the original data.

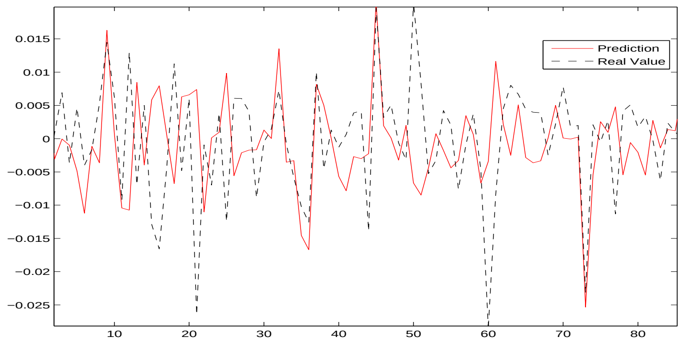

Fig. 7: Predicted returns spanned over 85 days. The dash line represents the FX returns, the solid line represents the prediction by the SOMAR network.

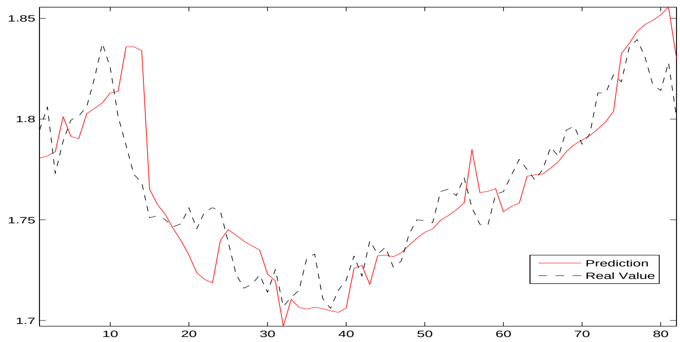

Fig. 8: Predicted rates spanned over 85 days. The dash line represents the FX prices, the solid line represents the prediction by the SOMAR network.

The results from two tests are compared to the vector SOM, SOAR and Recurrent SOM, in Table 1. It can be seen that the SOMAR outperforms other temporal SOM models. The experiments show that SOMAR is a good alternative method to cope with the nonstationarity and multiple underlying processes time-series.

4 Conclusions

A new approach to tackling nonstationarity of real-world time-series has been proposed by using the self-organising mixture autoregressive (SOMAR) model. The model consists of local autoregressive (AR) models and is organised and learnt by a self-organising map, so forming topologically ordered local regressive models. The proposed autocorrelation-based similarity measure makes the network effective and more robust compared to the error-based or Euclidean-based measures. The experiments show that the proposed model can correctly detect and uncover underlying AR models. They also show that the proposed method outperforms other SOM-based methods in modelling and prediction of nonstationary foreign exchange rates time-series.

References

1. Meese, R.A. and Rogoff, K.: Empirical exchange rate models of the seventies: do they fit out of sample?, Journal of International Economics 14 (1983) 3-24.
2. Hann,T.H. and Steurer, E.: Much ado about nothing? Exchange rate forecasting: Neural networks vs. linear models using monthly and weekly data, Neurocomputing 10 (1996) 323-339.
3. Chen, A.-S. and Leung, M.T.: Regression neural network for error correction in foreign exchange forecasting and trading, Computers and Operations Research 31 (2004) 1049-1068.
4. Cao, L.J.: Support Vector Machines Experts for Time Series Forecasting. Neurocomputing, Vol.51 (2002) 321-339
5. Diebold, F.X.: Empirical Modeling of Exchange Rate Dynamics, New York, Springer-Verlag. (1988)
6. Clark,P.K.: A Subordinate Stochastic Process Model With Finite Variance for Speculative Price, Economietrica, (1973) 135-155
7. Mandelbrot, B. and Taylor, H.: On the Distribution of Stock Price Differences, Operations Research, 15 (1969) 1057-1062.
8. Dablemont, S., Simon, G., Lendasse, A., Ruttiens, A., Blayo, F., Verleysen, M.: Time series forecasting with SOM and local non-linear models - Application to the DAX30 index prediction. WSOM'2003 proceedings - Workshop on Self-Organizing Maps Hibikino (Japan), 11-14 September (2003) 340-345.
9. Liu, Z. and Xu, L.: Topological local principal component analysis. Neurocomputing 55 (2003) 739-745.
10. Chappell, G. and Taylor, J.: The temproal Kohonen map, Neural Networks 6 (1993) 441-445.
11. Koskela, T.: Time Series Prediction Using Recurrent SOM with Local Linear Models. Helsinki University of Technology (2001)
12. Lampinen, J. and Oja, E.: Self-Organizing Maps for Spatial and Temporal AR Models. Proc. 6th SCIA, Scandinavian Conference on Image Analysis. Helsinki, Finland, (1989) 120-127.
13. Ni, H. and Yin, H.: Self-Organising Maps and Local MLP Models for Exchange Rate Prediction. Proceedings of the 12th Chinese Automation Computing Society Conference in the UK, Loughborough, England