ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 167-172

Recurrent SOM with Local Linear Models in Time Series Prediction

Timo Koskela, Markus Varsta, Jukka Heikkonen, and Kimmo Kaski

Helsinki University of Technology
Laboratory of Computational Engineering
P.O. Box 9400, FIN-02015 HUT, FINLAND

Abstract.
Recurrent Self-Organizing Map (RSOM) is studied in three different time series prediction cases. RSOM is used to cluster the series into local data sets, for which corresponding local linear models are estimated. RSOM includes recurrent difference vector in each unit which allows storing context from the past input vectors. Multilayer perceptron (MLP) network and autoregressive (AR) model are used to compare the prediction results. In studied cases RSOM shows promising results.

1. Introduction

In time series prediction the goal is to construct a model that can predict the future of the measured process under interest. Various approaches to time series prediction have been studied over the years [14]. Many different types of neural networks have been used in time series prediction, see e.g. [8] and [10]. Of linear methods autoregressive (AR) [1] models are frequently used. Different models can be divided to global and local models. In global model approach only one model is used to characterize the measured data. Local models are based on dividing the data set to smaller sets of data, each being modeled with a simple local model [9]. Creation of the local data sets is usually carried out with some clustering or quantization algorithm such as k-means, Self-Organizing Map (SOM) [13], [12] or neural gas [7]. Input to the model is usually provided by using a windowing technique to split the time series into input vectors. Typically input vectors contain past samples of the series up to certain length. In this procedure the temporal context between consecutive vectors is lost. One way of trying to avoid this is to include to the model memory that can store contextual information which exists between the consecutive input vectors.

Our approach in this study is to use Recurrent Self-Organizing Map (RSOM) [11] to store temporal context from the input vectors. The model consists of RSOM and local linear models that are each associated with a unit in the map. RSOM is used to cluster the time series into local data sets which belong to

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 167-172

certain unit and corresponding local model. Local model parameters are then estimated using the obtained local data sets. The rest of the paper is organized as follows: In the second section RSOM architecture and learning algorithm is introduced. In the third section different prediction cases are studied. In three cases of different time series results of RSOM are compared with linear and nonlinear global models (AR and MLP). Finally some conclusions are made.

2. Temporal Quantization with RSOM

Self-Organizing Map (SOM) [4] is a quantization method with topology preservation. Temporal Kohonen Map (TKM) [2] is a modification to the SOM that involves adding leaky integrators to the outputs of the map. Moving the leaky integrators from the unit outputs into the inputs gives rise to the Recurrent Self-Organizing Map (RSOM) [11].

2.1. Recurrent Self-Organizing Map

In the training algorithm of the RSOM an *episode* of consecutive input vectors $x(n)$ starting from a random point in the input space is presented to the map. The difference vector $y_i(n)$ in each unit of the map V_M is updated as follows:

$$y_i(n) = (1-\alpha)y_i(n-1) + \alpha(x(n) - w_i(n)) \ , \qquad (1)$$

where $y_i(n)$ is the leaked difference vector in unit i, $0 < \alpha \leq 1$ is the leaking coefficient, $x(n)$ is the input vector and $w_i(n)$ is the weight vector of the unit i. Each unit involves an exponentially weighted linear IIR filter with the impulse response $h(k) = \alpha(1-\alpha)^k$, $k \geq 0$, see Fig. 1. At the end of the episode (step n), the best matching unit b is searched by

$$y_b = min_i\{\|y_i(n)\|\} \ , \qquad (2)$$

where $i \in V_M$ and parallel vertical bars denote the Euclidean vector norm. Since the feedback quantity in RSOM is a vector instead of a scalar it also captures the direction of the error which can be exploited in weight update. The map is now trained with a slightly modified Hebbian training rule:

$$w_i(n+1) = w_i(n) + \gamma(n)h_{ib}(n)y_i(n) \ , \qquad (3)$$

where $i \in V_M$ and $\gamma(n)$, $0 \leq \gamma(n) \leq 1$, is a scalar valued adaptation gain. The neighborhood function, $h_{ib}(n)$, gives the excitation of unit i when the best matching unit is b. The winning unit is moved toward the linear combination of the sequence of input vectors captured in y_i. After updating all difference vectors y_i are set to zero, and a new random point from the input space is selected. The above scenario is repeated until the mapping is formed.

Because RSOM is trained with the y's it seeks to minimize quantization criterion that differs from the criterion with TKM. Nevertheless the resolution of RSOM is limited to the linear combinations of the input vectors with different responses to the operator in the unit inputs.

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 167-172

2.2. Local Model Estimation

Figure 2. shows the procedure for building the models and evaluating their prediction abilities with testing data [5]. Time series is divided to training and testing data. Input vectors to RSOM are formed by windowing the time series. For model selection purposes 4-fold cross-validation [3] was used. The best model according to cross-validation is trained again with the whole training data. This model is then used to predict the test data set that has not been presented to the model before.

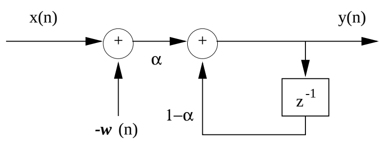

Figure 1: Schematic picture of an RSOM unit which acts as a recurrent filter.

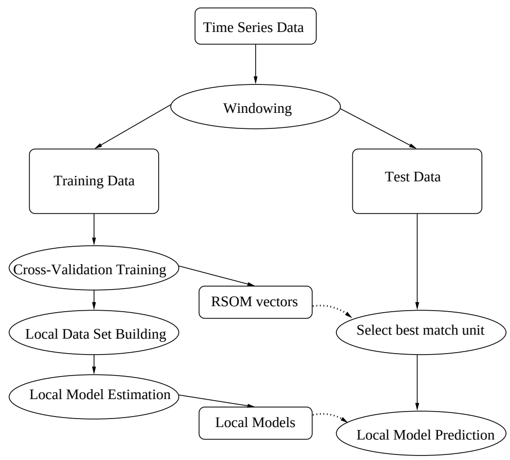

Figure 2: Building of the local models.

3. Case Studies

Three different time series were studied to compare RSOM with local linear models to MLP and AR models. The prediction task was in all cases one-step prediction. The same cross-validation scheme was used for all models.

For the RSOM with local linear models free parameters were input vector length p, time step between consecutive input vectors s, number of units n_u and the leaking coefficient α of the units in the map giving rise to model $RSOM(p, s, n_u, \alpha)$. In the studied cases parameters were varied as $n_u \in \{5, 9, 13\}$, $s \in \{1, 3, 5\}$ and $\alpha \in \{1.0, 0.95, 0.78, 0.73, 0.625, 0.45, 0.40, 0.35\}$ corresponding to episode lengths 1 ... 8. Input vector length p was varied differently in the three cases as described later. The regression models were estimated using the least squares algorithm in MATLAB 5 statistics toolbox using the data for which the corresponding RSOM unit was the best matching unit.

The MLP network was trained with Levenberg-Marquardt learning algorithm implemented with MATLAB 5 neural networks toolbox. An *MLP(p,s,q)* network with one hidden layer, p inputs and q hidden units was used. Variation of parameters p and s were chosen to be the same as in RSOM models, while q was varied as $q \in \{3, 5, 7, 9\}$.

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 167-172

AR(p) models with p inputs were estimated with MATLAB 5 using the least-squares algorithm. The order of the AR model was varied as $p \in \{1, \ldots, 50\}$. Results of the AR model serve as an example of the accuracy of a global linear model in the current tasks.

3.1. Mackey-Glass Chaotic Series

Mackey-Glass time series (Fig 3.) is produced by a time-delay difference system of the form [6]:

$$\frac{dx}{dt} = \beta x(t) + \frac{\alpha x(t-\gamma)}{1 + x(t-\gamma)^{10}} \tag{4}$$

where $x(t)$ is the value of the time series at time t. This system is chaotic for $\gamma > 16.8$. The time series was constructed with parameter values $\alpha = 0.2$, $\beta = -0.1$ and $\gamma = 17$ and it was scaled between [-1,1]. From the beginning of the series 3000 samples was selected for training, and the rest 1000 samples were used for testing. For RSOM and MLP models length of the the input vector was varied as $p \in \{3, 5, 7\}$.

The sum-squared errors gained for one-step prediction task are shown in Table 1. *MLP*(3,1,7) model gives the smallest cross-validation error but fails to predict the test set accurately. For the *AR*(2) model the results are opposite. AR model does not model here the underlying phenomena, instead it predicts the next value of the series using mainly the previous value. *RSOM*(3,1,5,0.95) gives moderate accuracy for both cross-validation and test data sets. With the test set, however, the error is smaller than with MLP network.

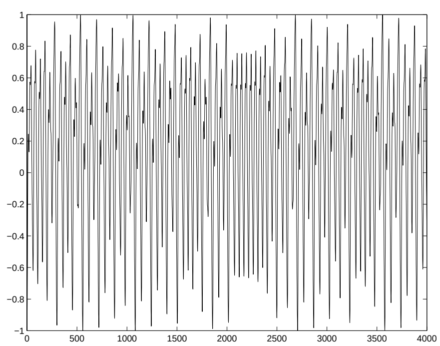

Figure 3: Mackey-Glass time series

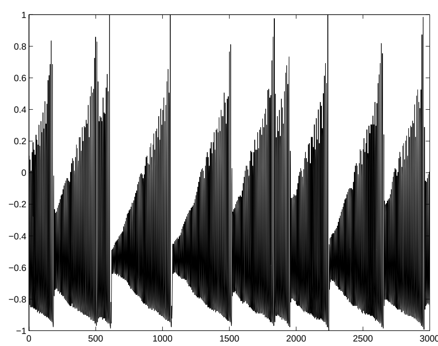

Figure 4: Laser time series

3.2. Laser Series

Laser time series [14] (Fig 4.) consists of measurements of the intensity of an infrared laser in a chaotic state. The data is available from an anonymous ftp server [1]. From the beginning of the series first 2000 samples were used for training, and the rest 1000 samples were used for testing. Both series were

[1]ftp://ftp.cs.colorado.edu/pub/Time-Series/SantaFe/ containing files A.dat (first 1000 samples) and A.cont (as a continuation to A.dat 10000 samples)

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 167-172

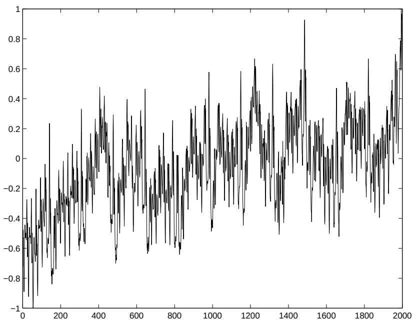

Figure 5: Electricity consumption time series

Table 1. Prediction Errors for Mackey-Glass Time Series.

	CV Error	Test Error
RSOM(3,1,5,0.95)	6.5556	3.1115
MLP(3,1,7)	0.3157	3.7186
AR(2)	4.1424	1.600

Table 2. Prediction Errors for Laser Time Series.

	CV Error	Test Error
RSOM(3,3,13,0.73)	14.6995	7.3894
MLP(9,1,7)	4.9574	0.9997
AR(12)	69.8093	29.8226

Table 3. Prediction Errors for Electricity Consumption Time Series.

	CV Error	Test Error
RSOM(8,1,13,0.73)	18.0673	2.6735
MLP(8,1,9)	7.6007	1.4345
AR(30)	6.5698	2.1059

scaled between [-1,1]. For RSOM and MLP models length of the the input vector was varied as $p \in \{3, 5, 7\}$.

The sum-squared errors gained for one-step prediction task are shown in Table 2. The laser series is highly nonlinear and thus the errors gained with $AR(12)$ model are considerably higher than for other models. The series is also stationary and almost noiseless, which explains the accuracy of the MLP(9,1,7) model predictions. In this case $RSOM$(3,3,13,0.73) gives results that are better than with AR model but worse than with MLP model.

3.3. Electricity Consumption Series

Electricity consumption series (Fig. 5.) contains measured load of an electric network. Measurements contain hourly consumption of electricity over a period of 83 days (2000 samples). The series was scaled between [-1,1]. For the training 1600 samples were selected, and the rest 400 samples were used for testing. For RSOM and MLP models length of the the input vector was varied as $p \in \{4, 8, 12\}$.

The sum-squared errors gained for one-step prediction task are shown in Table 3. The series contains 24 hours long cycle and also slower trend and noise in the form of measurement errors. $AR(30)$ model is found to reach quite acceptable results, due to the fact that model includes the whole 24 hour cycle. As the results with MLP(8,1,9) model show, nonlinear model can reach better predictions with a shorter window length. In this case $RSOM$(8,1,13,0.73) model does not give any improvement due to the insufficient input vector length used in model estimation.

4. Conclusions

Time series prediction using Recurrent SOM with linear regression models has been studied. For the selected prediction tasks this scheme gives promising results. Due to the selection of RSOM parameters its prediction accuracy did not reach in all cases accuracy of the AR model. However, in the case of the highly nonlinear laser series RSOM model gave considerably better prediction results than linear models. In the studied cases MLP seems to perform better than RSOM. This is mainly due to the selected one-step prediction problem. Another reason is the linear models used with RSOM.

RSOM model has several attractive properties in the study of time series. Perhaps the most important is the visualization possibilities of the map. Another is the ability to find temporal features from the data with an unsupervised learning algorithm. In this study we used RSOM that has the same feedback structure in all the units. It is possible, however, to allow the units of RSOM to have different recurrent structures. Such extensions of RSOM will be studied in the near future.

Acknowledgments

This study has been funded by *The Academy of Finland.*

References

[1] G. Box, G. Jenkins, and G. Reinsel. *Time Series Analysis: Forecasting and Control.* Prentice Hall, Englewood Cliffs, New Jersey, 1994.

[2] G.J. Chappell and J.G. Taylor. The temporal Kohonen map. *Neural Networks*, 6:441–445, 1993.

[3] L. Holmström, P. Koistinen, J. Laaksonen, and E. Oja. Neural and statistical classifiers—taxonomy and two case studies. *IEEE Trans. Neural Networks*, 8(1):5–17, 1997.

[4] T. Kohonen. *Self-Organization and Associative Memory.* Springer-Verlag, Berlin, Heidelberg, 1989.

[5] T. Koskela, M. Varsta, J.Heikkonen, and K. Kaski. Time series prediction using RSOM with local linear models. Technical Report B-15, Helsinki University of Technology, Lab. of Computational Engineering, 1997.

[6] M. Mackey and L. Glass. Oscillations and chaos in physiological control systems. *Science*, pages 197–287, 1977.

[7] T. Martinetz, S. Berkovich, and K. Schulten. 'Neural-Gas' network for vector quantization and its application to time-series prediction. *IEEE Trans. Neural Networks*, 4(4):558–569, 1993.

[8] M. Mozer. Neural net architectures for temporal sequence processing. In A. Weigend and N. Gershenfeld, editors, *Time Series Prediction: Forecasting the Future and Understanding the Past*, pages 243–264. Addison-Wesley, 1993.

[9] A. Singer, G. Wornell, and A. Oppenheim. A nonlinear signal modeling paradigm. In *Proc. of the ICASSP'92*, 1992.

[10] A. Tsoi and A. Back. Locally recurrent globally feedforward networks: A critical review of architectures. *IEEE Transactions on Neural Networks*, 5(2):229–239, 1994.

[11] M. Varsta, J. Heikkonen, and J. Del Ruiz Millán. Context learning with the self-organizing map. In *Proc. of Workshop on Self-Organizing Maps*, pages 197–202. Helsinki University of Technology, 1997.

[12] J. Vesanto. Using the SOM and local models in time-series prediction. In *Proc. of Workshop on Self-Organizing Maps*, pages 209–214. Helsinki University of Technology, 1997.

[13] J. Walter, H. Ritter, and K. Schulten. Non-linear prediction with self-organizing maps. In *Proc. of Int. Joint Conf. on Neural Networks*, volume 1, pages 589–594, 1990.

[14] A. Weigend and N. Gershenfeld, editors. *Time Series Prediction: Forecasting the Future and Understanding the Past.* Addison-Wesley, 1993.