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. 221-226

Forecasting Time-Series by Kohonen Classification

Amaury Lendasse[1], Michel Verleysen[1], Eric de Bodt[2],
Marie Cottrell[3], Philippe Grégoire[4]

[1]Université catholique de Louvain, Electricity Dept., 3 pl. du Levant,
B-1348 Louvain-la-Neuve, Belgium.
[2]Université catholique de Louvain, IAG-FIN, 1 pl. des Doyens,
B-1348 Louvain-la-Neuve, Belgium.
[3]Université de Paris I – SAMOS, 90 rue de Tolbiac,
F-75634 Paris Cedex 13, France.
[4]University of Namur, CeREFim, 8 Rempart de la Vierge,
B-5000 Namur, Belgium.

Abstract. In this paper, we propose a generic non-linear approach for time series forecasting. The main feature of this approach is the use of a simple statistical forecasting in small regions of an input space adequately chosen and quantized. The partition of the space is achieved by the Kohonen algorithm. The method is then applied to a widely known time-series from the SantaFe competition, and the results are compared with the best ones published for this series.

1. Introduction

The goal of the time-series forecasting method presented in this paper is to capture the dynamics of a process observed through a set of data to simulate its future behaviour. The main feature of this approach is the use of a simple statistical forecasting in small regions of an input space adequately chosen and quantized. The model is local, contrasting to most common prediction methods that try to globally forecast the time-series.

In the second section of the paper, we present in a few words the Kohonen algorithm, used here as a vector quantization algorithm for the discretisation of the process dynamics. In the third section, we describe the proposed method. In the fourth one, we show its application to a widely used time-series coming from the Santa Fe competition.

2. The Kohonen self-organising map

The Kohonen map [6, 1, 2] is a well-known unsupervised learning algorithm which produces a map composed by a fixed number of units. A physical neighbourhood

relation between the units is defined and each unit is characterised by a parameter vector of the same dimension as the input space. After learning, each unit represents a group of individuals with similar features (this group is named Voronoi region of the unit). The correspondence between the features and the units approximately respects the input space topology: similar features correspond to the same unit or the neighbouring units. The final map is said to be a self-organised map which preserves the topology of the input space.

The Kohonen self-organising map is thus a vector quantization (VQ) method which has the supplementary property (compared to other VQ methods) that its respects the topology. Strictly speaking, we use only the VQ property of the Kohonen map in the following forecasting method. Nevertheless, we use this algorithm instead of a more conventional VQ method because of two reasons: first, the Kohonen algorithm seems to be faster than more conventional methods for the same quality of results (but the generality of this statement and theoretical justifications remain open questions). But secondly, what is more important for us, is that the parameter vectors resulting from the Kohonen algorithm are ordered, what greatly facilitates their interpretation (see for example [9]).

3. Forecasting Time Series

To forecast a time-series, we have to fit the corresponding process that is observed through a set of data. In this paper, we limit the problem to one-dimensional time-series (i.e. series with a scalar output). The initial data vector, denoted D, is therefore of length n, where n is the number of observations in time.

The first step of our approach is, as for a lot of classical time-series analysis methods, to choose a lagging order. The initial data vector, D, is then modified to incorporate the vector of the observed variable as well as the past realisations of this one. The new data matrix, P, is of order $[n-\lambda, \lambda]$, where λ is the lagging order. Rows of P are denoted $x_t = \{ x_{t,1} , \dots , x_{t,\lambda} \}$, where t is the time index.

The P matrix is then decomposed into a number of homogeneous clusters, using the Kohonen algorithm (a one dimensional Kohonen map is used). To each unit of the Kohonen map is associated, after learning, a number of individuals for which this unit is the winning one. The clusters are formed. For each unit, the parameter vector (called here "profile" because it represents the mean of the attached time slots of length λ in the series) $\bar{x}_i$ is calculated. As λ, the choice of the number of clusters will depends on the features of the analysed process and results from the quality of learning measured on validation sets. The observation of the homogeneity of the clusters (for example, the Fisher statistics) obtained after learning is a good indicator.

For each row x_t of the P matrix, the associated deformation is then computed. It is noted y_t and is computed according to $y_t = x_{t+1} - x_t$. On this basis, for each cluster of

the P matrix, a P_i matrix, composed by the y_t corresponding to the x_t of the cluster i, is formed.

For each P_i matrix, as for the P matrix, decomposition into a number of homogeneous clusters is realised, once again using the Kohonen algorithm (a one-dimensional Kohonen map is also used). The mean profiles of the formed clusters are then determined and denoted by $\bar{y}_j$.

The last step to characterise the analysed process is to establish the empirical frequencies of $\bar{y}_j$ conditionally to $\bar{x}_i$ by counting. They are denoted by $P(\bar{y}_j | \bar{x}_i)$.

Once this characterisation procedure using two stages of Kohonen maps is completed, the forecasting procedure takes the following form:

- choice of a starting point x_t (for example, the initial individual in the P matrix);
- determination of the winning $\bar{x}_i$ using ArgMin_i ($\| x_t - \bar{x}_i \|$);
- picking at random an $\bar{y}_i$ according to the conditional probabilities $P(\bar{y}_j | \bar{x}_i)$;
- computation of x_{t+1} by $x_t + \bar{y}_j$;
- iteration of the procedure for simulating the dynamics of the process on a specific time horizon;
- the procedure will be iterated and the results will be averaged.

4. Application to Santa Fe Data

The objective of this section is to test the presented method on a widely recognised benchmark, the data series D from the Santa Fe competition [8].

The characteristics of this series are:
- relatively high-dimensional dynamics;
- long data sets, (100,000 points);
- no background information;
- quite stationary.

The lagging order of this series reported in the literature (for non-linear forecasting methods) [8] is about 20. Therefore, we are using the last 20 values of the series to predict the next value. The P matrix is of order [99980, 20].

The number of clusters in the Kohonen algorithm is chosen through the observation of the homogeneity of the clusters, by the Fisher statistics and visual inspection of profiles. In this case, it will be included between 200 and 300.

For each row x_t of the P matrix, the associated deformation is then computed. On this basis, for each cluster of the P matrix, a P_i matrix is formed. For each P_i matrix, decomposition into a number of homogeneous clusters is realised again through the

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. 221-226

Kohonen algorithm. The number of clusters is chosen according to the minimum quantization error on a validation set, not used for training. The optimal number of clusters in the deformation space is between 3 and 10.

Hence, we predicted the 15 and 30 next values of the series and compared them both with the true ones and with the best results of the Santa Fe's competitors.

The evaluation is measured by the *normalised mean squared error*:

$$\mathrm{NMSE}(N) = \frac{\sum_{k\in T}(\mathrm{observation}_k - \mathrm{prediction}_k)^2}{\sum_{k\in T}(\mathrm{observation}_k - \mathrm{mean}_T)^2} \approx \frac{1}{\hat{\sigma}_T^2}\frac{1}{N}\sum_{k\in T}(x_k - \hat{x}_k)^2$$

where k = 1 … N holds for the N (here 15 or 30) values in the targets set T, and mean_T and $\hat{\sigma}_T^2$ denote the sample average and sample variance of the observed values (targets) in T. A value of NMSE = 1 corresponds to simply predicting the average.

The series predicted with 200 clusters in the input space and 10 clusters in each deformation space is illustrated in Figure 1. For these parameters, we obtain NMSE(15) = 0.4957 and NMSE(30) = 0.6384.

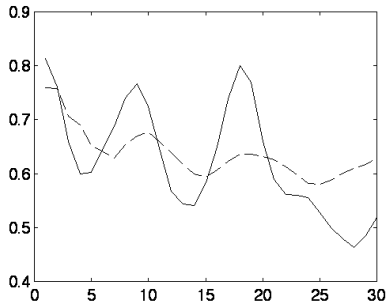

Fig.1: comparison between the true (continuous line) and the predicted value (dash line).

To improve the accuracy of the forecast, we tried to learn the dynamics of the series on the last 10,000 points instead of the whole 100,000 points set. The results are illustrated in Figure 2 and summarised in Table 1.

Clusters in input space / cl. in deformations spaces	NMSE(15)	NMSE(30)
100 / 5	0.5244	0.6134
200 / 3	0.3834	0.5542
200 / 5	0.3921	0.5342
250 / 4	0.5542	0.7135

Table 1: summary of the main results.

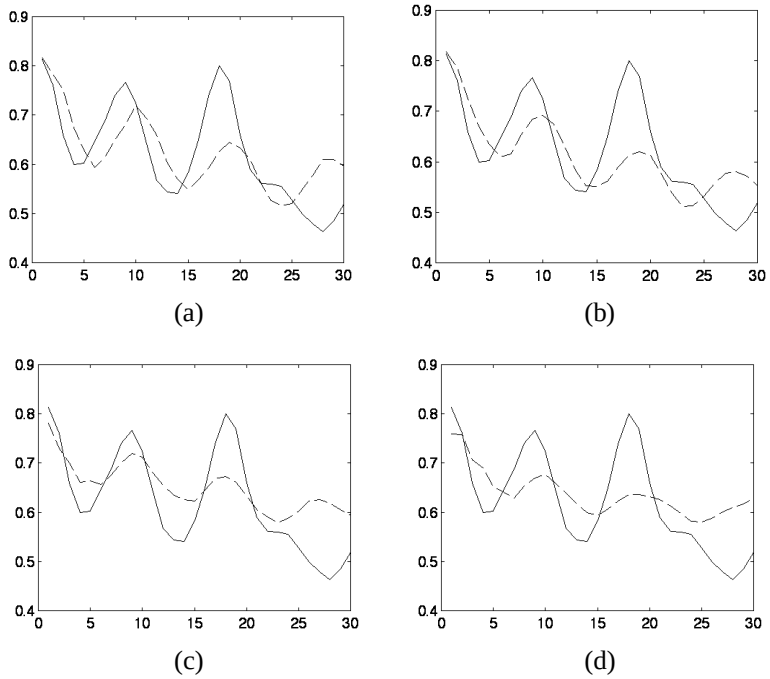

Fig.2: (a) 100 clusters in the input space and 5 in the deformations spaces (b) 200 clusters in the input space and 3 in the deformations spaces (c) 200 clusters in the input space and 5 in the deformations spaces (d) 250 clusters in the input space and 4 in the deformations spaces.

Table 2 shows a comparison between our method (indicated as "Kohonen" in the table) and the best results reported for this series [8].

Method	Type	NMSE(15)	NMSE(30)
conn	…-30-30-1	0.086	0.57
tree	k-d tree; AIC	1.3	1.4
conn	Recurrent, 4-4c-1	6.4	3.2
conn	1-30-30-1; lags 20,5,5	7.1	3.4
linear	36 AR(8), last 4k pts.	4.8	5.0
conn	Feedforward	17.	9.5
Kohonen	**200 / 5 clusters**	**0.39**	**0.54**

Table 2: comparison between the main methods.

It must be mentioned that the first method listed in Table 2 (which remains the best one) implies a computationally extremely heavy learning on thousands of perceptrons, while the other methods, including our proposal, remain computationally reasonable.

Similar results have been obtained with other standard benchmarks.

5. Conclusion

In the proposed method for time series forecasting, an adequately chosen space is divided into clusters, in order to apply a simple prediction methods on each of these clusters. The division (quantization) has been realised by the Kohonen algorithm. The proposed algorithm gives one of the best forecasts reported for this series in the literature.

The power of the method is linked to an adequate choice of its three main parameters: the lagging order λ, the number of clusters in the input space and the number of clusters in each deformation space. Once the lagging order is chosen as in any other forecasting method (application context, correlation analysis,…), the two other parameters are easily adjusted by using learning validation sets.

6. Acknowledgements

Michel Verleysen is Research Associate of the Belgian National Fund for Scientific Research (FNRS). Amaury Lendasse is working under a grant from the UCL – FSR.

References

1. Cottrell M., Fort J.C., Étude d'un algorithme d'auto-organisation, *Annales de l'Institut Poincaré*, Vol. 23, 1, p. 1-20, 1987.
2. Cottrell M., Fort J.C., Pagès G., Two or three things that we know about the Kohonen algorithm, *Proc of ESANN*, M. Verleysen ED., D Facto, Brussels, 1994.
3. Cottrell M., de Bodt E., Verleysen M., Kohonen maps versus vector quantization for data analysis, in *Proc of ESANN*, M. Verleysen ED., D Facto, Brussels, 1997.
4. Cottrell M., de Bodt E., Grégoire Ph., Simulating Interest Rate Structure Evolution on a Long Term Horizon : A Kohonen Map Application, *Proceedings of Neural Networks in The Capital Markets*, Californian Institute of Technology, World Scientific Ed., Pasadena, 1996.
5. Cox J.C., Ingersoll J.E., Ross S.A., *A Theory of the Term Structure of Interest Rates*, Econometrica 53, p. 385-407, 1985.
6. Kohonen T., Self-organising Maps, *Springer Series in Information Sciences*, Vol. 30, Springer, Berlin, 1995.
7. Walter J., Ritter H., Schulten K., Non-linear prediction with self-organising maps, *Proc. of IJCNN*, San Diego, CA, 589-594, July 1990.
8. A.S.Weigend and N.A.Gershenfeld, Eds., Time Series Prediction: Forecasting the Future and Understanding the Past, Reading, MA: Addison Wesley, 1994.
9. Cottrell M., Girard B. et Rousset P., Long Term Forecasting by Combining Kohonen Algorithm and standard prevision, *Proc. of ICANN'97*, Lausanne, 1997.