ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 329-334

Time Series Forecasting using CCA and Kohonen Maps - Application to Electricity Consumption

A. Lendasse[1], J. Lee[2], V. Wertz[1], M. Verleysen[2*]

[1]Université catholique de Louvain, CESAME, 4 av. G. Lemaître
B-1348 Louvain-la-Neuve, Belgium, {lendasse, wertz}@auto.ucl.ac.be.
[2]Université catholique de Louvain, Electricity Dept., 3 pl. du Levant,
B-1348 Louvain-la-Neuve, Belgium, {lee, verleysen}@dice.ucl.ac.be.

Abstract. Using large regressors in non-linear time series forecasting makes the fitting of the model difficult. This paper shows how to reduce the size of regressors in order to improve the forecasting performances, using the Curvilinear Component Analysis as projection tool. The method is applied to the Polish electrical load forecasting.

1. Introduction

Time series forecasting is a challenge in many fields of application. In the financial context, one wants for example to forecast some stock exchange or currency indexes. Data processing specialists try to predict the flow of information on networks, and electricity producers try to have the best possible prediction of the electrical load usually one day ahead. The common point to these problems is the following question: how to analyse and use the past to forecast the future? Many techniques exist, such as for example the linear models (ARX, ARMA...) [1,2] and the non linear artificial neural networks [3]. In general, these methods try to build a model of the process, which is in turn used for further forecasting; so-called auto-regressive models relate the value(s) to predict to the past values of the series. The common difficulty to all methods is the determination of sufficient and necessary information for a good forecasting. If information is insufficient, the forecasting will be poor. On the contrary, if part of the information is useless or redundant, modelling (in particular non-linear modelling) will be difficult or even skewed.

In this paper, we will describe an original method for the selection of useful information from the past values of a series. Further forecasting will be achieved by using Kohonen maps. Finally, we will illustrate the proposed method by a traditional example of time series forecasting, the electric consumption in a country (we will use Polish data).

---

[*] Michel Verleysen is a Research Associate of the Belgian F.N.R.S.

2. Non linear forecasting

Non-linear forecasting [4] uses the same idea as linear forecasting. In the following, we will restrict ourselves to auto-regressive models, i.e. models using (only) past values of the series. Adding exogenous variables is straighforward.

Let us consider a time series known through N samples y_t (from time t = 1 to t = N). The model used to capture the dynamics of the process is of the form:

$$y_{t+1} = F(y_t, y_{t-1}, \ldots, y_{t-p}, \Theta) , \quad (1)$$

where Θ is the set of the parameters in the model. The vector including values y_t to y_{t-p} is called the *regressor*. It is obvious that the choice of the regressor and thus of p is capital. A too small value of p will lead to a poor model, while a too large one will lead to difficulties in fitting model F. Several methods exist to choose the regressor. For example, the regressor can be chosen so that it gives the best performances when used in a linear model. Another option is to initially select a sufficiently large regressor, and then apply pruning methods to the model F. However, both these possibilities are not fully satisfactory. The first one is obviously not optimised for a non-linear forecasting, while the second one is computationally intensive and sometimes fails on too complex models.

3. Determination of the regressor using CCA

Instead of selecting a regressor *among* past values of the series, we will *build* a regressor *from* past values. To that purpose, we will use a non-linear projection method, as detailed in the following.

First, we build an initial Y_t regressor of size p, which must be sufficiently large to contain all the information necessary for a good prediction

$$Y_t = \lfloor y_t, y_{t-1}, \ldots, y_{t-p+1} \rfloor , \quad (2)$$

Using this regressor, we work in a p-dimensional space in which information is most probably redundant. This redundancy of information can be viewed as (possibly non-linear) correlations between the coordinates in the p-dimensional space. In other terms, the real or intrinsic dimension of the dataset is lower than p, and will be referred to as d in the following. An intuitive definition of the intrinsic dimension is to view that the regressors (the data points) form a d-surface in the p-dimensional space.

Having this is mind, we will try to build new regressors of size d, by projecting the initial regressors on a d-dimensional space. Many projection techniques exist, among them the linear PCA. Using a linear method to compress the data when a non-linear forecasting technique will be used is however non-optimal. An interesting alternative to PCA is the Curvilinear Component Analysis (CCA), which is one of its non linear

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 329-334

extensions [5,6]. Figure 1 shows an example of non-linear projection with CCA. CCA resembles Sammon's mapping and Multi-Dimensional Scaling (MDS), but has several advantages discussed in [5].

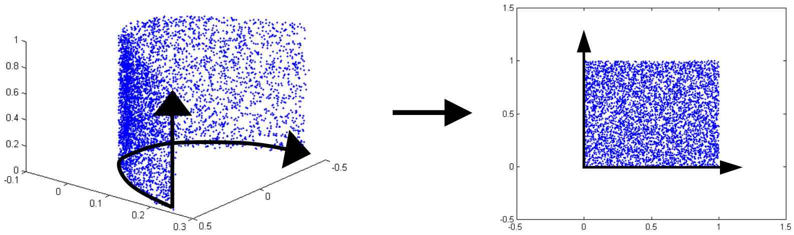

Fig. 1: Projection carried out by CCA from R^3 to R^2.

Each regressor given by (2) is thus projected using CCA on a new *state vector*

$$Z_t = [z_1, z_2, \ldots, z_d]. \tag{3}$$

The forecasting model is then build of the state vector instead of the initial regressor:

$$y_{t+1} = F(z_1, z_2, \ldots, z_d, \Theta) \ , \tag{4}$$

Reducing the size of the input vector to the forecasting model has a determining advantage for the subsequent forecasting. For most non-linear models F (multi-layer perceptrons for example), more inputs mean more parameters, which, for a specific task, decreases performances in the context of the bias-variance dilemma. It is also well known that learning in larger non-linear models is more difficult. Of course, on the contrary, some information will be lost in the projection process. Our method is based on the assumption that this loss will be largely compensated by the increase in the forecasting performances due to the use of a smaller input vector.

4. Artificial Example

Figure 2 shows an artificial series generated by equation

$$g(y) = \sin(y) + \sin(2y) \tag{5}$$

sampled at 0.25 intervals between values $y = 0$ and $y = 160$. We use initial regressors of dimension 3. Figure 3 shows the space of the initial regressors; it is clearly visible that the intrinsic dimension of the series is 1. We will thus project the 3-dimensional regressors on a 1-dimensional space, using CCA. Figure 4 shows the value to predict y_{t+1} in the initial series as a function of the unique state value z_t. It clearly shows that a simple 1-dimensional forecasting model will be sufficient to capture the dynamics of the series. Note that the two small sets of points further from the data in Figure 4 are due to the limitations of the non-linear projection.

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 329-334

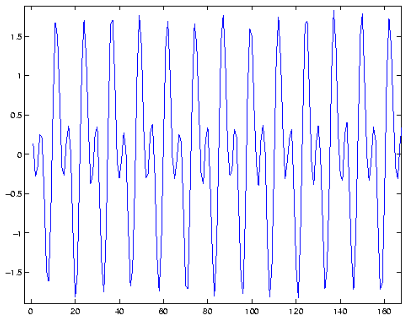

Fig. 2: An artificial time series.

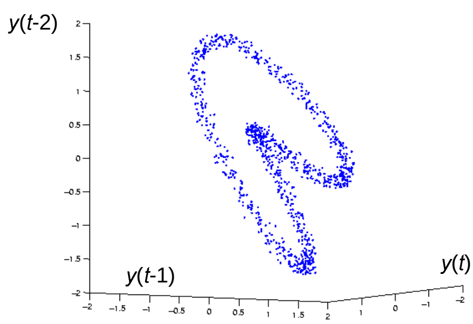

Fig. 3: Space of initial regressors (dimension = 3) for the series illustrated in Figure2

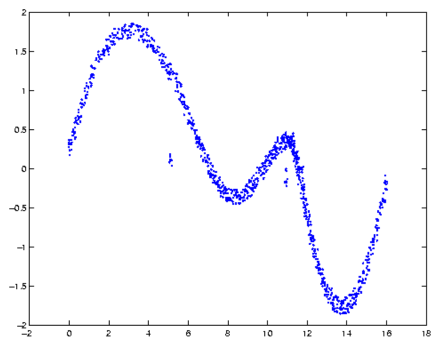

Fig. 4: Value to predict y_{t+1} according to new regressor Z (only one state variable in this example)..

This series can be easily modelled using a RBF network (Radial Basis Function Network) with 5 Gaussian kernels [7]. The MSE (Mean Square Error) obtained in this case is 0.12. In comparison the MSE obtained with the basic regressor is 0.26 with a linear model and 0.11 using a RBF with 25 Gaussian kernels.

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 329-334

5. Forecasting Time Series using Kohonen Maps

Instead of using MLP or RBF as forecasting model, we use Kohonen maps [8] built on vectors formed by concatenating the state vector (or the regressor) and the value to predict:

$$x_t = [z_1, z_2, \ldots, z_d, y_{t+1}] , \tag{6}$$

Centroids C_i resulting from the Kohonen map learning are thus made of two parts: the first one C_{i1} (dimension d) corresponds to the state vector or regresor, and the second one C_{i2} corresponds to the forecasting. These centroids form our model. When a value is to be predicted, centroid C_w is selected whose distance between its first part C_{w1} and the regressor is minimal. The second part C_{i2} of the centroid is then used as prediction.

6. Application to Electricity Consumption

The series we will study represents the normalized average daily load of electric consumption in Poland [9], as shown in Figure 5 (left). One sees well there the seasonal variation that is nearly sinusoidal. On a smaller scale (Figure 5 right), the electric consummation shows another cycle, which is high for the days of the week and low the weekend.

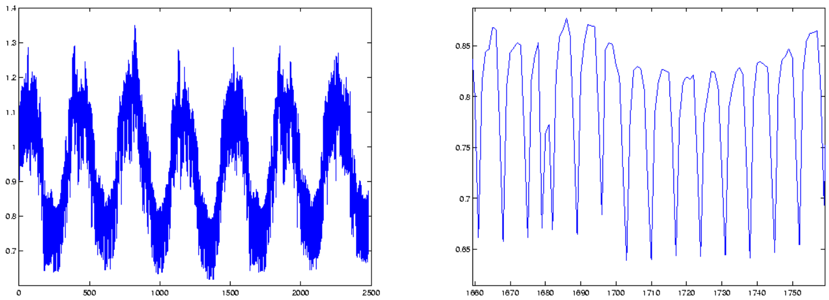

Fig. 5: Electric consumption in Poland.

The series is randomly divided in a training set (two third of the data) and a test set (one third of the data); it was experimentally verified that holdout or leave-one-out performance evaluations do not affect significantly the results, because of a sufficient number of data available.

If we use a linear model, a very good regressor is formed by the 8 last values of the series. This initial regressor is then projected to a 4-dimensional space using CCA (the dimension of the projection space has been experimentally determined here to keep 98% of the variance of the initial data, but could also result from an intrinsic dimension estimation). We then quantify the space formed with the projection by a SOM with 20x20 centroids and use this map for the forecasting. The MSE obtained on the test set is 0.0018. In comparison, the MSE obtained without projection is

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 329-334

0.0019. The improvement is only about 5% here, but remains limited because of the inaccuracies of the projection (as in Figure 5). We hope to improve these results through ameliorations on the projection.

7. Conclusion

The preliminary results presented in this paper show that the CCA method can be used to reduce the dimensionality of a data set, here the set of regressors in the problem of times series forecasting. The aim of this reduction is to facilitate a further forecasting by non-linear methods. Preliminary results obtained with a Kohonen forecasting model show that performances on a real dataset are slightly improved, despite inaccuracies in the projection method.

Acknowledgements

The authors would like to thank Dr. Patrick Rousset and Prof Marie Cottrell (Univ. Paris I, SAMOS-MATISSE) for providing the dataset used in this paper.

References

[1] Box G.E.P., Jenkins G.: "*Time Series analysis: Forecasting and Control*" Cambridge University Press, 1976.

[2] Ljung L.: "*System Identification - Theory for User*", Prentice-Hall, 1987.

[3] Weigend A. S., Gershenfeld N.A.: "*Times Series Prediction: Forecasting the future and Understanding the Past*", Addison-Wesley Publishing Company, 1994.

[4] Sjoberg J., Zhang Q., Ljung L. et al., "*Non linear black box modelling in system identification: A unified overview*", Automatica 33: (6) 1691-1724, 1995.

[5] Demartines P., Herault J., *"Curvilinear component analysis: A self-organizing neural network for non linear mapping of data sets"* IEEE Transaction on Neural Networks 8: (1) 148-154 January 1997.

[6] Demartines P. and Hérault J. *"Vector Quantization and Projection"* In A. Prieto, J. Mira, J. Cabestany, eds, International Workshop on Artificial Neural Networks, Vol. 686 of Lecture Notes in Computer Sciences, pp. 328-333. Springer-Verlag, 1993.

[7] Verleysen M., Hlavačkova K.: "*An Optimised RBF Network for Approximation of Functions*". In: Proc of European Symposium on Artificial Neural Networks, Brussels (Belgium), April 1994.

[8] Kohonen T., "*Self-organising Maps, Springer Series in Information Sciences*", Vol. 30, Springer, Berlin, 1995.

[9] Cottrell M, Girard B, Girard Y, et al. "*Daily electrical power curves: Classification and forecasting using a Kohonen map"*, Natural to Artificial Neural Computation 930: 1107-1113, 1995.