UNIVERSITE PARIS 1 PANTHEON SORBONNE

CENTRE DE RECHERCHE

S.A.M.O.S

STATISTIQUE APPLIQUEE ET MODELISATION STOCHASTIQUE

Long term forecasting by combining Kohonen algorithm and standard prevision

M. Cottrell, B. Girard et P. Rousset

Prépublication du SAMOS n° 76
Mai 1997

90, rue de Tolbiac - 75634 PARIS CEDEX 13

Long Term Forecasting by Combining Kohonen Algorithm and standard prevision

Marie Cottrell, Bernard Girard and Patrick Rousset

Centre SAMOS, Université Paris 1 Panthéon Sorbonne
F-75634 Paris Cedex 13, France

Abstract. To forecast a complete curve is a delicate problem, since the existing methods (vectorial prevision, long-term forecasting) are difficult to use and often give disappointing results. We propose a new strategy that consists in dividing up the problem into three sub-problems: prediction of the mean value and of the standard deviation and estimation of the normalized curve (the profile). The mean value and the standard deviation are predicted by any classical method (linear or neural). As to the profile, it is estimated with the help of a previous classification. The results are very convincing and a real-world application is presented : the polish electrical consumption.

1 Introduction

We address a particular, but very usual case of long-term prediction. It is very well suited for the consumption curves, and more generally for all the cases where there are two time scales. The slow scale represents a fixed duration, which is divided into a constant number of time intervals, which defines the rapid scale. For instance, each curve corresponds to a day, and contains 24 values (one per hour) or 48 (one per half-hour). The goal is to predict at the same time all the values of the next curve with the same precision.

We could see this problem as a long-term prediction problem, but in that case the usual method consists in introducing the successive predicted values as new inputs to compute the future value, and the estimation accuracy is worst for the most remote values than for the nearest ones. That characteristic is not at all adapted to the concrete applications where we want the same forecasting quality along all the curve. As to the non linear prediction method (as multilayer perceptron model [10], [12], [13], [4]), one knows that the recurrence equation which defines the model can have several fixed points, which do not have any real meaning. In that case the long-term forecast values strongly depend on the initial values and have a chaotic behavior.

On another hand, we could handle the problem as a vectorial prediction problem, which consists in simultaneously predict all the values of the next curve. But actually, the mathematical frame for this kind of prediction is not completely clear so far, it is necessary to take into account the dependency between the values, and so on. In fact the parametric methods (linear or not) seem to be delicate to implement. For example, in [6], the authors propose a previous classification of all the days, and define as many perceptrons as classes. In [11], another mixed method is used which needs $I \times M$ multilayer perceptrons, where I is the number of values for a day and M is the number of classes that are identified. All these methods are somewhat complex. We propose here a semi-parametric alternative which is particularly well suited for the prediction of complete curves, as consumption curve, periodic phenomena, etc.

The paper is organized as follows : the section 2 is devoted to present the method and in section 3, we develop a real-world example that is the Polish electricity consumption.

2 The Method

The essential of this section reminds the definitions that we presented in [3] and [5], where artificial and real examples are studied.

Let us consider a time series $X(t,i)$, with a double indexation. The time is denoted by (t,i), where t is the slow scale (the day for example), and i is the rapid one (the hour for example). For each $t \in 1,T$, the second index takes its values between 1 and I, where I does not depend on t. The successive values are grouped according to the t index, and we set

$$X(t) = (X(t,i), i = 1, \ldots, I), t = 1, \ldots T \tag{1}$$

The time series is assumed to be stationary, or stationarized by a previous transformation.

For each curve $X(t)$, we define its mean value $\mu(t)$, its variance $\sigma^2(t)$, and its normalized *profile*,

$$P(t) = (P(t,i), 1 \leq i \leq I) = \left(\frac{X(t,i) - \mu(t)}{\sigma(t)}\right). \tag{2}$$

The prediction problem is then divided into three parts :

1. to forecast the mean value of the next day $t+1$, from the previous values
2. in the same way, to forecast the standard deviation of the next day,
3. to estimate the profile of the next day.

That is justified by the fact that for the consumption curves, it is often useful to distinguish the shape from the level and from the variability.

So the estimation of the complete curve $X(t+1)$ will be

$$\hat{X(t)} = \sigma\hat{(t)}\hat{P(t)} + \mu\hat{(t)} \tag{3}$$

The previsions of the mean and of the standard deviation are made by any method. We can use a linear method (ARMA model or ARMAX with exogenous variables), or a non linear model (a Multilayer Perceptron for example).

The main novelty that we propose is the way to estimate the profile of any given day. The idea is to classify all the previous profiles (see (2)) into n classes. This number n has to be chosen with caution. If it is too small, the estimation will be too rough. If it is too large, the absolute frequency of each class will be too small and the clusters will not be significant.

Suppose that the T profiles are split up into n classes $(C_1, C_2, \ldots, C_n)$ represented by the centroids $W_1, W_2, \ldots, W_n$. Suppose that in the past, a given day, for example a working Monday of February (which is not a public holiday, nor an extra day), has been classified into one or several classes. Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be the number of occurrences of this kind of day in the classes $(C_1, C_2, \ldots, C_n)$. Note that α_j is 0 if and only if this day never belongs to the class C_j. Then the estimated profile of this given day is given by

$$\hat{P} = \frac{\sum_{j=1}^{n} \alpha_j W_j}{\sum_{j=1}^{n} \alpha_j} \tag{4}$$

After that the estimated curve for that day is computed according to formula (3).

The remaining point is to choose the clustering method in order to define the classes and the centroids. We propose to use the Kohonen algorithm [7], [8], [9], [1], [2] with a suited topology for many reasons. First it is an efficient algorithm, small time consuming, which provides classes and centroids in an easy way. Secondly, a neighborhood structure is defined between the classes, and that feature allows an interesting representation of the similarities of the classes. Since the Kohonen algorithm is "topology preserving", close inputs are classified into the same class or neighboring classes, and in fact a given kind of day (e.g. a working Monday of February) appears only in neighboring classes. This property is particularly useful.

For example, if a Monday of February belongs to a class that is far from the others, it is very probable that there is some error in the measured values or that it was a special day (public holiday, extra day, and so on). In fact the topological arrangement of the classes allows to decide if some exogenous qualitative variable has to be taken into account to compute the different types of days. This kind of control is very easy to realize with a Kohonen classification, and would be very tedious (or computationally impossible) with any other classification method.

3 Example: the Polish Electricity Consumption

We present an example to show how to implement the method. We deal with real Polish Power System Data, starting from 1/1/1986 to 31/12/1994 and kindly lent by Pr. Osowski from Warsaw Technical University. The data are hourly (expressed in 20 000 Mwh). As we do not know the particularities of the Polish calendar nor the meteorological variables, we cannot take into account the external temperature, the cloudiness, the extra-days, etc. The goal is here only explanatory.

So we have 3287 24-points daily curves, from which we compute 3287 daily normalized profiles. We consider a cylindrical Kohonen 10×10 network, where the left and right borders are neighboring. We train this network as usual, by taking the 3287 profiles as inputs, and renormalizing the code vectors at each step.

See in Fig. 1 the map of the centroids and the contents of the classes. Observe the continuity between one class and its neighbors.

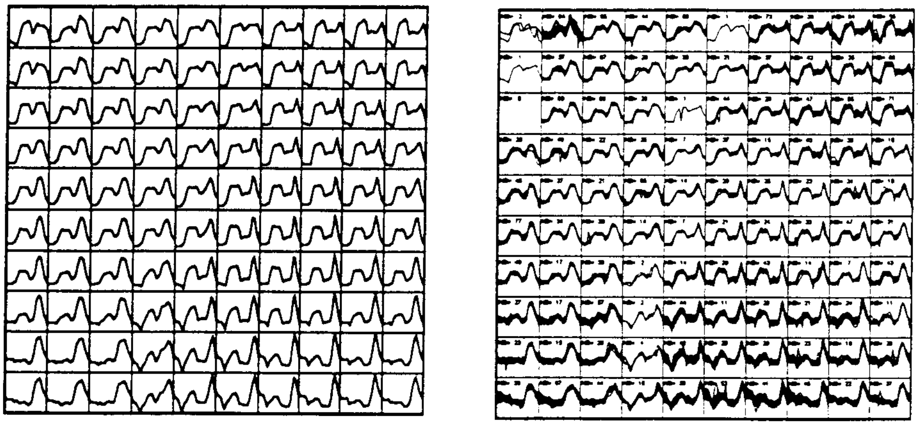

Fig. 1. The left map represents the code vectors, while the right shows the contents of the class

A hierarchical clustering shows in Fig. 2 how it is possible to gather the 100 Kohonen classes into 10 macro-classes, that can be more easily interpreted. In a

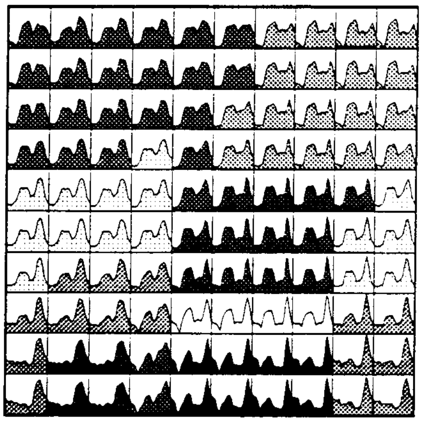

Fig. 2. The grey levels indicate the macro-classes that can be easily interpreted

rough way, by examining the contents of the classes, one can see that the Sundays are located at the bottom, the Mondays at the top in the middle, the summer week-days at the right top corner, the winter week-days at the left top corner and at the left, and so on. At the center, we find April, May and September. We can also see that the Saturdays are spread over the map. In fact the status of the Saturday cannot be taken into account, since some of them are worked like week-days, but we do not know which. In any case, this interpretation is not used to compute the estimate profile.

The next step is the making of a calendar to calculate for each day the number of its occurrences in the classes. For example for a worked Wednesday of November, we have the following distribution (the first number is the class, the second one the number of occurrences): 3(3), 4(4), 5(2), 12(5), 13(5), 14(3), 22(2), 23(11), 24(2), 34(1). This leads to the estimated profile according to formula (4).

On another hand, the predictions of the daily mean and standard deviation are achieved by ARMA method. The two models are :
$(I - B)(I - B^7)\mu(t) = (I - 0.9B)(I - 0.05B - 0.11B^2 - 0.1B^3)\varepsilon(t)$
and $(I - B^7)(I - B)\sigma(t) = (I - 0.8B^7)(I - 0.8B)\varepsilon(t)$.

We compare with a global ARMA model, defined by the following equation:

$$(I - B)(I - B^{168})(I - 0.2B^{24})(I - 0.13B^{48})Y(h) = (I - 0.2B)\varepsilon(h)$$

which explains one half-hour h from the past ones. But for the prediction, we cannot take into account the just past half-hours, since, for example in the afternoon of a day, we have to estimate all the curve of the next day. Then we compute the estimated curve by using the successive estimated values.

See in Fig. 3, two examples of comparisons, between the actual values and the predicted ones (by the ARMA model and the neural method).

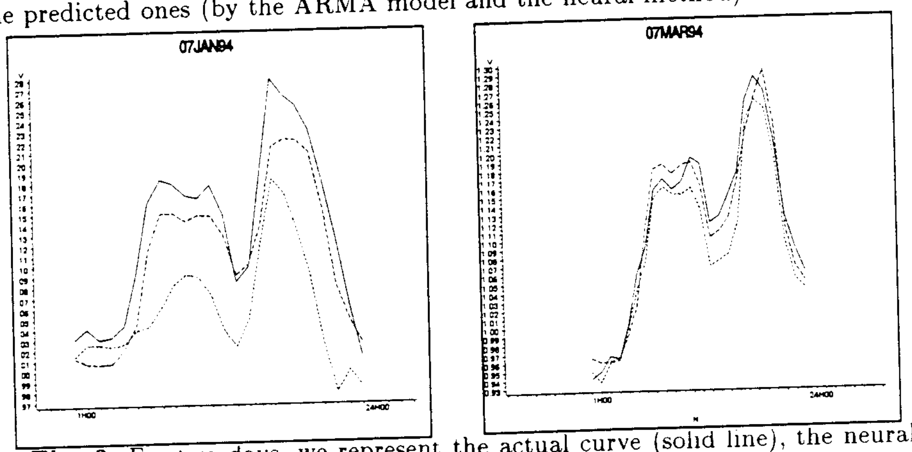

Fig. 3. For two days, we represent the actual curve (solid line), the neural forecast one (large dotted line), the ARMA forecast one (small dotted line). The neural model swallows the change of structure from January 1 to January 7, while the ARMA model depends on the same day of the previous week in a too strong way.

4 Conclusion, Performances

The performances could be improved if we could use explanatory exogenous qualitative variables. That can help to better define the type of days. In fact, according to the epoch of the changing of hour for example, or of the holidays, it can be interesting to use some days of the previous month together with the next month, etc. In any case here are some the indexes that we use to measure the performance of the model. There are two first order error indices: the mean value of the coeficient of variation $I_1 = \frac{1}{N}\sum_t \frac{1}{\mu(t)}\sqrt{\frac{1}{24}\sum_j (X(t,j)-\hat{X}(t,j))^2}$, and the mean value of the relative error $I_2 = \frac{1}{24N}\sum_{t,j}\frac{\|X(t,j)-\hat{X}(t,j)\|}{X(t,j)}$. We get $I_1 = 0.034$ and $I_2 = 0.029$.

We also use a quadratic error I_3, to compare to the quadratic corrected total error per half-hour I_4, and per day I_5.

We have $I_3 = \frac{1}{24N}\sum_{t,j}(X(t,j) - \hat{X}(t,j))^2 = 0.0004$, to be compared to $I_4 = \frac{1}{24N}\sum_{t,j}(X(t,j) - X(.,j))^2 = 0.0312$ and to

$I_5 = \frac{1}{24N}\sum_{t,j}(X(t,j) - X(t,.))^2 = 0.0063$.

References

1. Cottrell M., Fort J.C.: Etude d'un algorithme d'auto-organisation. Ann. Inst. Henri Poincaré, **23**, 1, (1987), 1-23.
2. Cottrell M., Fort J.C. and Pagès G.: Two or three things that we know about the Kohonen algorithm. Proc. of ESANN 94, M.Verleysen Ed., Editions Quorum, (ISBN 2-9600049-1-4), (1994), 235-244.
3. Cottrell M., Girard B., Girard Y., Muller C.and Rousset P.: Daily Electrical Power Curves : Classification and Forecasting Using a Kohonen Map. From Natural to Artificial Neural Computation, Proc. IWANN'95, J.Mira, F.Sandoval eds., Lecture Notes in Computer Science, Vol.930, Springer, (1995), 1107-1113.
4. Cottrell M., Girard B., Girard Y., Mangeas M., Muller C.: Neural modeling for time series: a statistical stepwise method for weight elimination. IEEE Transactions on Neural Networks, **6**, 6, (1995), 1355-1364.
5. Cottrell M., Girard B., Rousset P.: Forecasting of curves using a Kohonen classification. To appear in Journal of Forecasting, (1997).
6. Garcia Tejedor A., Cosculluela M., Bermejo C. and Montes R.: A neural system for short-term load forecasting based on day-type classification. Proc. ISAP 94, (1994).
7. Kohonen T.: Self-organized formation of topologically correct feature maps. Biol. Cyb., **43**, (1982), 59-69.
8. Kohonen T.: Self-organization and associative memory. Springer, New York Berlin Heideberg, Third edition, 1989.
9. Kohonen T.: Self-Organizing Maps. Vol. 30, Springer, New York Berlin Heiderberg, 1995.
10. Lapedes A.S. and Farber R.: Non linear signal processing using neural networks : Prediction and System Modeling. Los Alamos National Laboratory Technical Report, 1987.
11. Muller M., Cottrell M., Girard B., Girard Y., Mangeas M.: A neural network tool for forecasting French electricity consumption. Proc. WCNN 94, San Diego, (1994), I-360-365.
12. Weigend A.S., Huberman B.A. and Rumelhart D.E.: Predicting the Future : a Connectionist Approach. Int. J. Neural Systems, 1, 3, (1990), 193-209.
13. Weigend A.S. and Gershenfeld N.A.: Times Series Prediction. Proceedings Volume XV, Santa Fe Institute, Addison-Wesley, New-York, 1994.

Prépublications du SAMOS (depuis 1995)

1995

40 - Claude BOUZITAT, Gilles PAGES
Pour quelques images de plus..., 12 p.

41 - Jean-Claude FORT, Gilles PAGES
About the Kohonen algorithm: Strong or Weak Self-organisation? 15 p.

42 - Marie COTTRELL, Patrick LETREMY
Classification et analyse des correspondances au moyen de l'algorithme de Kohonen: application à l'étude de données socio-économiques. 10 p.

43 - Joël CHADOEUF, Xavier GUYON, Jian-Feng YAO
Sur l'ergodicité de l'estimation par Restauration-Estimation de modèles incomplètement observés. 11 p.

44 - Jean-Gabriel ATTALI, Gilles PAGES
Approximation of functions by perceptrons, a new approach. 11 p.

45 - Marie COTTRELL, Bernard GIRARD, Yvonne GIRARD, Corinne MULLER, Patrick ROUSSET
Daily electrical power curves : classification and forecasting using a Kohonen map. 8 p.

46 - Fabienne COMTE, Cécile HARDOUIN
Regression on log-regularized periodogram for fractional models at low frequencies. 19 p.

47 - Fabienne COMTE, Cécile HARDOUIN
Regression on log-regularized periodogram under assumption of bounded spectral densities: the non fractional and the fractional cases. 14 p.

48 - Patrick ROUSSET
Prévision des courbes demi-horaires au moyen d'une classification de Kohonen. 25 p.

49 - Smail IBBOU et Marie COTTRELL
Multiple Correspondence analysis of a crosstabulations matrix using the Kohonen algorithm. 6 p.

50 - Philippe JOLIVALDT
Schémas de discrétisation pour la simulation et l'estimation d'un CAR(2): une étude expérimentale. 22 p.

51 - Philippe JOLIVALDT
Utilisation de méthodes implicites pour la simulation et l'estimation de modèles CAR(2) . 14 p.

1996

52 - Samuel BAYOMOG
Estimation of a Markov field dynamic. 14 p.

53 - Morgan MANGEAS et Jian-feng YAO
Sur l'estimateur des moindres carrés des modèles auto-régressifs fonctionnels. 19 p.

54 - Marie COTTRELL, Florence PIAT, Jean-Pierre ROSPARS
A Stochastic Model for Interconnected Neurons. 17 p.

55 - Marie COTTRELL, Jean-Claude FORT, Gilles PAGES
Two or three mathematical things about the Kohonen algorithm. 31 p.

56 - Marie COTTRELL, Bernard GIRARD, Patrick ROUSSET

Forecasting of curves using a Kohonen classification. 14 p.

57 - Jean-Claude FORT, Gilles PAGES
Quantization vs Organization in the Kohonen S.O.M. 5 p.

58 - Eric de BODT, Marie COTTRELL, Michel LEVASSEUR
Réseaux de neurones en finance.33 p.

59 - Marie COTTRELL, Eric de BODT, Emmanuel HENRION, Ismaïl IBBOU, Annick WOLFS, Charles Van WYMEERSCH
Comprendre la décision à l'aide d'une carte de Kohonen. Une étude empirique. 16 p.

60 - Marie COTTRELL, Eric de BODT
Understanding the leasing decision with the help of a Kohonen map. An empirical study of the Belgian market. 5p.

61 - Marie COTTRELL, Eric de BODT, Philippe GREGOIRE
The relation between interest rate shocks and the initial rate structure: an empirical study using a Kohonen map. 16p.

62 - Marie COTTRELL, Eric de BODT, Philippe GREGOIRE
A kohonen map representation to avoid misleading interpretation. 8p.

63 - Marie COTTRELL, Eric de BODT
Analyzing shocks on the interest rate structure with Kohonen map. 6p.

64 - Marie COTTRELL, Eric de BODT, Philippe GREGOIRE
Simulating interest rate structure evolution on a long term horizon. A kohonen map application. 5p.

65 - Fabienne COMTE, Cécile HARDOUIN
Log-regularized periodogram regression. 22p.

66 - Jian Feng YAO
Simulation et optimisation sous contrainte par une dynamique de Metropolis. 6p.

67 - Morgan MANGEAS, Jian-feng YAO
On least squares estimation for nonlinear autoregressive processes. 16p.

68 - Carlo GAETAN, Xavier GUYON
Simulation des modèles de Gibbs spatiaux par chaine de Markov. 28p.

69 - François GARDES, Patrice GAUBERT, Patrick ROUSSET
Cellulage de données d'enquêtes de consommation par une méthode neuronale. 41p.

70 - Serge IOVLEFF, José R. León
High-Frequency approximation for the Helmholtz equation :a probabilistic approach. 14p.

1997

71 - Xavier Guyon et Jian-feng Yao
On description of wrong parametrisation sets of a model . 23p.

72 - Sandie Souchet
Schéma de discrétisation anticipatif et estimation du paramètre d'une diffusion. 36p.

73 - M. Cottrell, J.C. Fort, G. Pagès
Theoretic aspects of the SOM algorithm 22p.

74 - M. Benaïm, J.C. Fort & G. Pagès
Convergence of the one-dimensional Kohonen algorithm. 23p.

75 - C. Bouton & G. Pagès
About the multidimensional competitive learning vector quantization algorithm with constant gain. **36p**.

76 - M. Cottrell, P. Rousset
The Kohonen algorithm: a powerful tool for analyzing and reprenting multidimensional quantitative **and** qualitative data 12p.

77 - M. Cottrell, B. Girard et P. Rousset
Long term forecasting by combining Kohonen algorithm and standard prevision . 6p.