Available online at www.sciencedirect.com
ScienceDirect
NEUROCOMPUTING
Neurocomputing 70 (2007) 2325–2329
www.elsevier.com/locate/neucom

Editorial

Time series prediction competition: The CATS benchmark

1. Introduction

Time series forecasting is a challenge in many fields. In finance, one forecasts stock exchange courses or stock market indices; data processing specialists forecast the flow of information on their networks; producers of electricity forecast the load of the following day. The common point to their problems is the following: how can one analyze and use the past to predict the future? Many techniques exist: linear methods such as ARX, ARMA, etc. [1,7], and nonlinear ones such as artificial neural networks [2–5,9,11].

In general, these methods try to build a model of the process that is to be predicted. The model is then used on the last values of the series to predict future ones. The common difficulty to all methods is the determination of sufficient and necessary information for a good prediction. If the information is insufficient, the forecasting will be poor. On the contrary, if information is useless or redundant, modeling will be difficult or even skewed.

In parallel with this determination, a prediction model has to be selected. In order to compare different prediction methods several competitions have been organized, for example: The Santa Fe Competition [11]; The K.U. Leuven Competition: Advanced Black-Box Techniques for Nonlinear Modeling: Theory and Applications [6]; The EUNITE competition [9].

After the competitions, their results have been published and the time series have become widely used benchmarks. The goal of these competitions is the prediction of the following values of a given time series (30–100 values to predict). Unfortunately, the long-term prediction of time series is a very difficult task, more difficult than the short-term prediction.

Furthermore, after the publication of results, the real values that had to be predicted are also published. Thereafter it becomes more difficult to trust in new results that are published: knowing the results of a challenge may lead, even unconsciously, to bias the selection of model; some speak about “data snooping”. It becomes therefore more difficult to assess newly developed methods, and new competitions have to be organized.

In the present CATS competition, the goal was the prediction of 100 missing values of the time series; they are grouped in 5 sets of 20 successive values. The prediction methods have then to be applied several times, allowing a better comparison of the performances. Twenty-four papers and predictions were submitted to the competition. Seventeen papers were accepted to IJCNN’04.

The papers that are published in this special issue have been selected according to two criteria:

- novelty of the proposed method;
- accuracy of the prediction.

In the following, we will summarize the previous prediction competitions in Section 2. We will present the CATS benchmark in Section 3. The results are described in Section 4.

2. Previous competitions

Several challenging time series competitions have been organized, and time series data sets have been collected.

2.1. Santa Fe time series competition

Six time series data sets were proposed: Data Set A within this competition: Laser-generated data, Data Set B: Physiological data, Data Set C: Currency exchange rate data, Data Set D: Computer-generated series, Data Set E: Astrophysical data, Data Set F: J.S. Bachs last (unfinished) fugue [11]. The main benchmark of the competition was the Data Set A recorded from a Far-Infrared-Laser in a chaotic state. From this physical system 1000 data points were given, and 100 points in the future had to be predicted by the participants. The winner of the competition was E.A. Wan, using a finite impulse response neural networks for autoregressive time series prediction.

2.2. K.U. Leuven time series prediction competition

The benchmark of the competition was a time series with 2000 data. The competition data were generated from a computer-simulated generalized Chuas circuit. The task was to predict the next 200 points of the time series. In total, 17 entries were submitted for the competition, and the winning contribution was made by J. McNames [10]. The strategy incorporated a weighted Euclideanmetric and a novel multi-step cross-validation method to assess model accuracy. A nearest trajectory algorithm was proposed as an extension to fast nearest neighbor algorithms [10].

0925-2312/$ - see front matter © 2007 Published by Elsevier B.V.
doi:10.1016/j.neucom.2007.02.013

2.3. EUNITE: EUropean Network on Intelligent TEchnologies for smart adaptive systems classification competition

The problem to be solved here was the forecasting of maximum daily electrical load based on half-an-hour loads and average daily temperatures (time period 1997–1998) [13]. Also included were the holidays for the same period of time. The actual task of each participant was to supply the prediction of maximum daily values of electrical loads for January 1999 (31 data values all together). The advantages of this benchmark were the length (around 35,000 points) and that the real data set allows to give further interpretation on the prediction result. The disadvantage was the specificity of the prediction with maximum of curves and the use of external inputs (temperatures). The winner of the competition was C.-J. Lin with a support vector machine method [13]. In total, 26 entries were submitted for the competition.

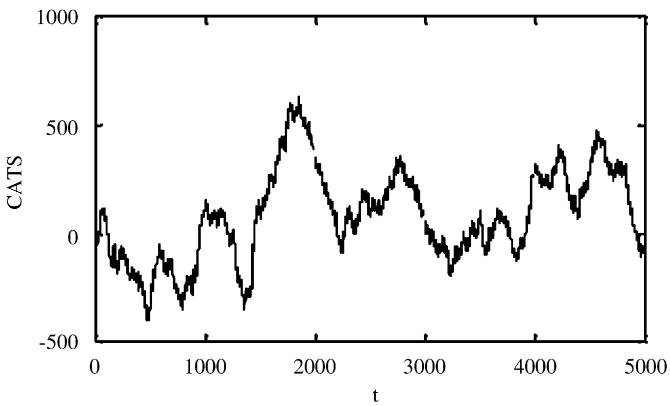

Fig. 1. CATS benchmark.

3. The CATS benchmark

The proposed time series is the CATS (competition on artificial time series) benchmark. This series is represented in Fig. 1.

This artificial time series is given with 5000 data, among which 100 are missing. The missing values are divided into five blocks:

- elements 981–1000;
- elements 1981–2000;
- elements 2981–3000;
- elements 3981–4000;
- elements 4981–5000.

The mean square error E_1 will be computed on the 100 missing values using

$$E_1 = \frac{\sum_{t=981}^{1000}(y_t - \hat{y}_t)^2}{100} + \frac{\sum_{t=1981}^{2000}(y_t - \hat{y}_t)^2}{100} + \frac{\sum_{t=2981}^{3000}(y_t - \hat{y}_t)^2}{100} + \frac{\sum_{t=3981}^{4000}(y_t - \hat{y}_t)^2}{100} + \frac{\sum_{t=4981}^{5000}(y_t - \hat{y}_t)^2}{100}. \quad (1)$$

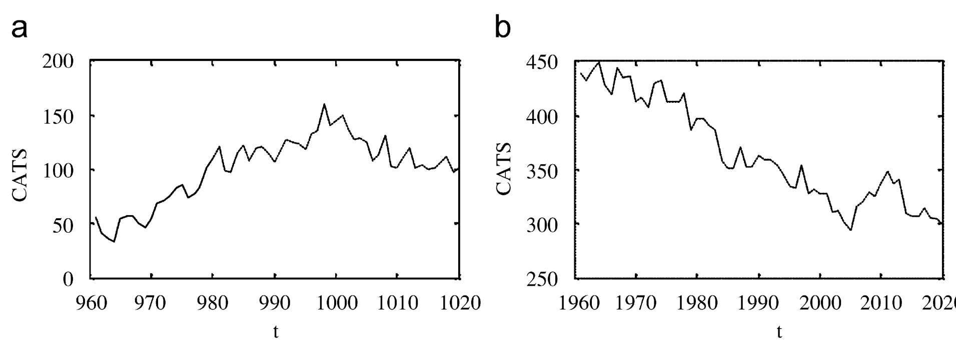

Fig. 2. Missing values: (a) 981–1000; (b) 1981–2000.

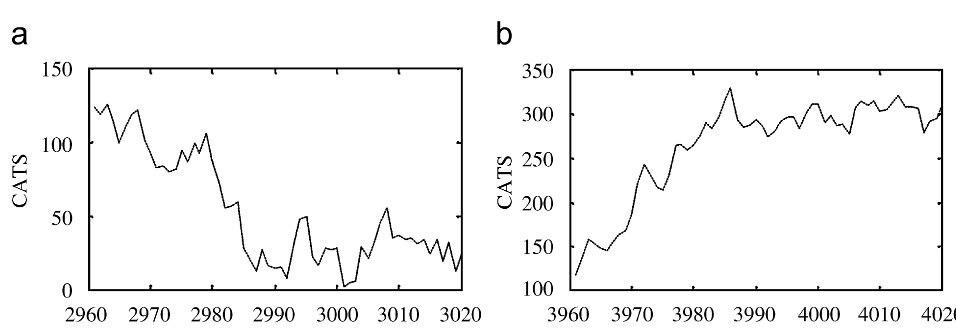

Fig. 3. Missing values: (a) 2981–3000; (b) 3981–4000.

The mean square error E_2 will be computed on the 80 first missing values using

$$E_2 = \frac{\sum_{t=981}^{1000}(y_t - \hat{y}_t)^2}{80} + \frac{\sum_{t=1981}^{2000}(y_t - \hat{y}_t)^2}{80} + \frac{\sum_{t=2981}^{3000}(y_t - \hat{y}_t)^2}{80} + \frac{\sum_{t=3981}^{4000}(y_t - \hat{y}_t)^2}{80}. \quad (2)$$

This second error criterion is used because some of the proposed methods are using not only the data before a set of missing values to perform the prediction but also the data after the set. As such procedure is not possible in the case of the fifth set of missing values, error E_2 is used to assess the prediction on the first four blocks only. The mean square error E_1 is the only one that is used for the ranking of the submissions; the mean square error E_2 is used to give some additional information about the performances and the properties of these methods. The missing parts are given in Figs. 2–4, and the numerical values can be found in [6,14].

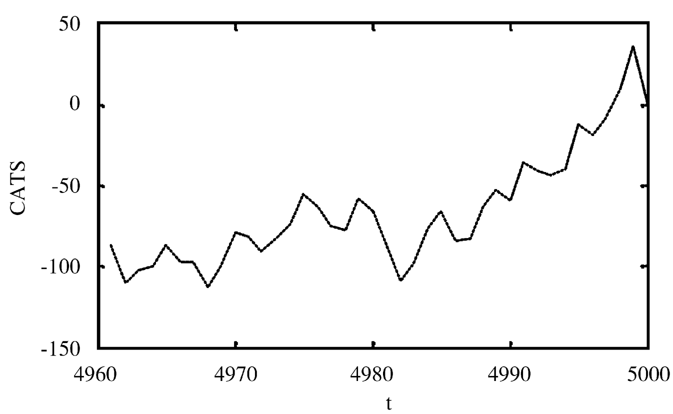

Fig. 4. Missing values 4981–5000.

4. Results of the competition

The 24 methods that were submitted to the competition are very different and give very dissimilar results. The results are summarized in [6]. The error E_1 is in a range between 408 and 1714. It is important to notice that some methods are very good for the prediction of the 80 first values but very bad for the last 20 ones. The results of the winner [8] are represented in Figs. 5–7.

The results of the winner on the first 80 values only [12] are represented in Figs. 8–10. The method that has been used by the winner of the competition is divided in two parts: the first sub-method provides the short-term prediction and the second sub-method provides the long-term one. Both sub-methods are linear, but according to the author better results could be obtained if the first sub-method was nonlinear. According to this author, the key of a good prediction is this division between two

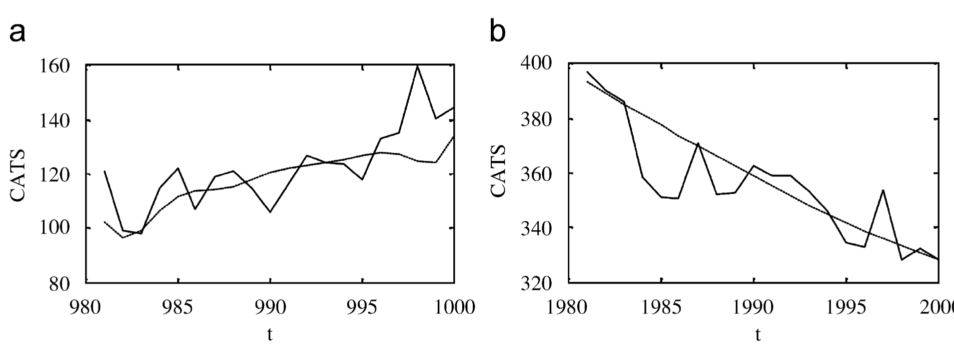

Fig. 5. Missing values: (a) 981–1000 (solid line) and their approximation in [8]; (b) 1981–2000 (solid line) and their approximation in [8].

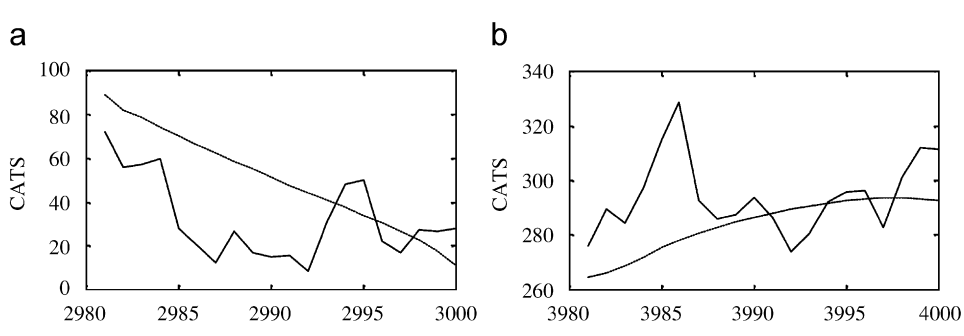

Fig. 6. Missing values: (a) 2981–3000 (solid line) and their approximation in [8]; (b) 3981–4000 (solid line) and their approximation in [8].

subproblems. More details about the different methods can be found in [6].

5. CATS special issue

The eight papers that are included in this special issue have been selected according to different criteria. First, they provide good or very good predictions. It is important to remember that mainly the presented methods have been tested on only one time series. Therefore, it can be dangerous to give much importance to the ranking. Secondly, the selection has been done according to the originality of the methods. Thirdly, the diversity has been taken into account.

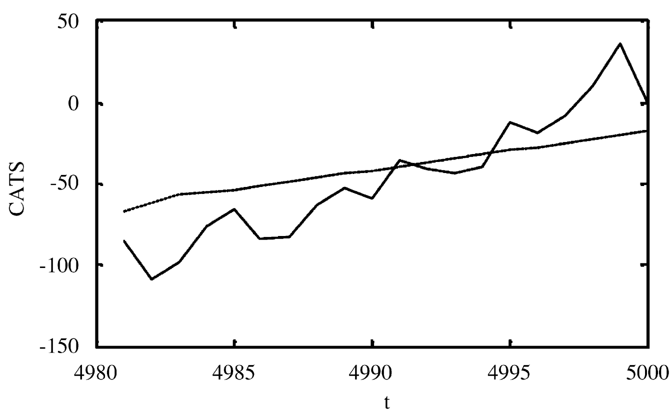

Fig. 7. Missing values 3981–4000 (solid line) and their approximation in [8].

The selected papers can been divided into five categories:

- Bayesian methods, methods related to Kalman filters' and methods based on other filtering techniques. The papers that can be classified in this category are: Xiao Hu *et al.*, Shuichi Kurogi *et al.*, Federico Palacios-Gonzalez, P.F. Verdes *et al.*
- Methods based on recurrent neural networks. The papers that can be classified in this category are: Igor Beliaev *et al.*, Xindi Cai *et al.*, Xiao Hu *et al.*
- Vector Quantization. The paper that can be classified in this category is Geoffroy Simon *et al.*
- Fuzzy logic. The paper that can be classified in this category is L.J. Herrera *et al.*
- Ensemble methods. The paper that can be classified in this category is P.F. Verdes *et al.*

It is surprising but interesting that so many different methods are proposed to preform time series prediction. Some of the papers can be classified in several categories, for example P.F. Verdes *et al.* or Xiao Hu *et al.*

Unfortunately, it should be added that several methods or papers that were submitted have been rejected due to the poor performances that are obtained; even if the proposed methods are very similar to the ones that are published in this special issue. That shows that the quality of the prediction does not depend only on the method but also on

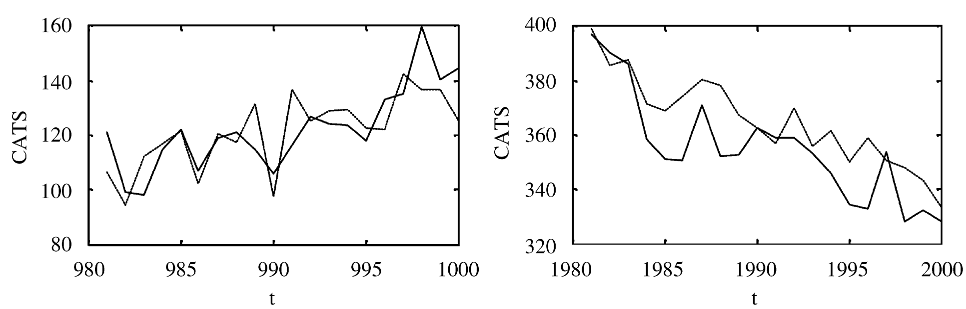

Fig. 8. Missing values: (a) 981–1000 (solid line) and their approximation in [12]; (b) 1981–2000 (solid line) and their approximation in [12].

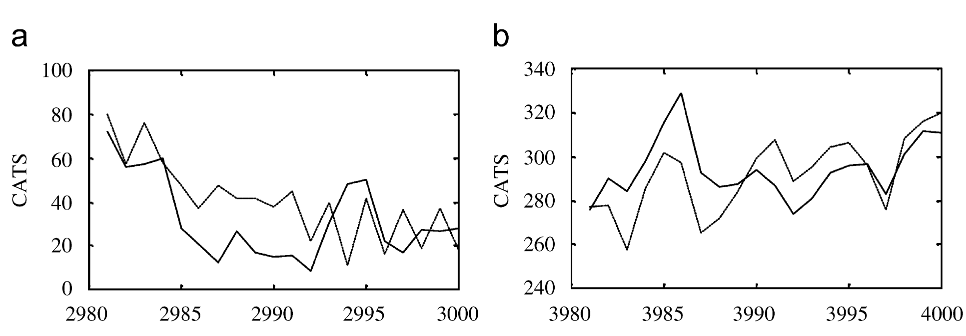

Fig. 9. Missing values: (a) 2981–3000 (solid line) and their approximation in [12]; (b) 3981–4000 (solid line) and their approximation in [12].

Fig. 10. Missing values 3981–4000 (solid line) and their approximation in [12].

the quality of the methodology and the knowhow of the user. For this reason, we have asked the authors:

- to emphasis the methodological explanation in their papers,
- to share their know-how and their experience in the field of time series prediction.

For all these reasons, we believe that this special issue would give a wide enough overview of the time series prediction field.

References

[1] G. Box, G. Jenkins, Time Series Analysis: Forecasting and Control, Cambridge University Press, Cambridge, 1976.

[2] M. Cottrell, B. Girard, M. Mangeas, C. Muller, Neural modeling for time series: a statistical stepwise method for weight elimination, IEEE Trans. Neural Networks 6 (1995) 1355–1364.

[3] M. Cottrell, B. Girard, P. Rousset, Forecasting of curves using a kohonen classification, J. Forecast. 17 (1998) 429–439.

[4] A. Lendasse, E. de Bodt, V. Wertz, M. Verleysen, Nonlinear financial time series forecasting—application to the bel20 stock market index, Eur. J. Econ. Soc. Syst. 14 (2000) 81–91.

[5] A. Lendasse, J.A. Lee, V. Wertz, M. Verleysen, Forecasting electricity consumption using nonlinear projection and self-organizing maps, Neurocomputing 48 (2002) 299–311.

[6] A. Lendasse, E. Oja, O. Simula, M. Verleysen, Time series prediction competition: the cats benchmark, in: Proceedings of IJCNN 2004, July 2004, pp. 1615–1620.

[7] L. Ljung, System Identification Theory for User, Prentice-Hall, Englewood Cliffs, NJ, 1987.

[8] S. Sarkka, A. Vehtari, J. Lampinen, Time series prediction by Kalman smoother with cross validated noise density, in: Proceedings of IJCNN 2004, July 2004, pp. 1653–1657.

[9] J. Sjoberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P. Glorennec, H. Hjalmarsson, A. Juditsky, Non linear black box modeling in system identification: an unified overview, Automatica 33 (1997) 1691–1724.

[10] J. Suykens, J. Vandewalle, The K.U. Leuven time-series prediction competition, Kluwer Academic Publishers, 1998, pp. 241–253.

[11] A. Weigend, N. Gershenfeld, Times Series Prediction: Forecasting the Future and Understanding the Past, Addison-Wesley, Reading, MA, 1994.

[12] J. Wichard, M. Ogorzalek, Time series prediction with ensemble models, in: Proceedings of IJCNN 2004, July 2004, pp. 1625–1629.

[13] Available from ⟨http://neuron.tuke.sk/competition/index.php⟩.

[14] Available from ⟨http://www.cis.hut.fi/lendasse/competition/competition.html⟩.

Amaury Lendasse, Erkki Oja
Olli Simula
Helsinki University of Technology, Adaptive Informatics Research Centre, P.O. Box 5400, 02015 Espoo, Finland
E-mail addresses: amaury.lendasse@hut.fi (A. Lendasse), erkki.oja@hut.fi (E. Oja), olli.simula@hut.fi (O. Simula)

Michel Verleysen
Universite catholique de Louvain, Machine Learning Group, 1348 Louvain-la-Neuve, Belgium
E-mail address: verleysen@dice.hut.fi