Simple kNN-Method for Times Series Prediction

D'yakonov Alexander G.

***Abstract*—This document describes a simple approach to times series prediction at 2006/07 Forecasting Competition for Neural Networks & Computational Intelligence.**

I. Introduction

Now there is a huge number of various methods of time-series forecasting [1-4]. Naturally, there is no universal method. In each concrete problem it is necessary to choose the most adequate algorithm, and then to adjust its parameters. For data from competition [5] we have decided to take the most simple method applied in data analysis: nearest neighbour [6,7]. Below stages of adjustment of this method are described. We shall note, that the greatest interest causes its comparison with other methods (it becomes known after end of competition).

II. Method

Let $\tilde{f} = (f_1, \dots, f_n)$ be the time-series data. It is required to continue $\tilde{f}$ up to

$$\tilde{g} = (f_1, \dots, f_n, f_{n+1}, \dots, f_{n+t}) .$$

We shall assume, that such continuation is determined by "background" $(f_{n-l+1}, \dots, f_n)$, i.e. it is necessary to find a part $(f_{k-l+1}, \dots, f_k)$, which after some transformation A becomes similar on $(f_{n-l+1}, \dots, f_n)$:

$$\| A(f_{k-l+1}, \dots, f_k) - (f_{n-l+1}, \dots, f_n) \| \to \min_{k,\tilde{a}} ,$$

where $\tilde{a}$ is parameter of transformation A. Having determined k, we find the nearest neighbour to our background. Required continuation (forecast) will be

$$A(f_{k+1}, \dots, f_{k+l}) .$$

It was the idea of a method. Generally we search some nearest neighbours. Forecast will be their linear combination

$$\sum_k c_k A(f_{k+1}, \dots, f_{k+l}), \ \sum_k c_k = 1 .$$

The basic problem is a choice of transformation $A : \mathbf{R}^l \to \mathbf{R}^l$ and parameter l.

In more general case it is possible to consider transformation $A : \mathbf{R}^{lr} \to \mathbf{R}^l$ and to search for r nearest neighbours $\tilde{h}_1, \dots, \tilde{h}_r$:

$$\| A(\tilde{h}_1, \dots, \tilde{h}_q) - (f_{n-l+1}, \dots, f_n) \| \to \min_{\tilde{h}_1, \dots, \tilde{h}_q, \tilde{a}} .$$

Besides it is possible to search not only in the given time-series data $\tilde{f}$, but also in the transformed vectors $A_1\tilde{f}, \dots, A_q\tilde{f}$, where $A_1, \dots, A_q$ are some transformations. For example, linear operators, compression, stretching, nonlinear transformations have been used in this competition.

The most universal operator (which not bad worked for all competition series) has appeared the most simple operator A:

$$A(x_1, \dots, x_l) = (a_1 x_1 + a_2, \dots, a_1 x_l + a_2) ,$$
$$\tilde{a} = (a_1, a_2) .$$

Thus, the least squares method has been used for searching nearest neighbours.

However it is interesting, that the nearest neighbour forecast has lost the non-nearest neighbour forecast! For example, the final variant has been received by a linear combination of forecasts on the third, fourth and fifth neighbours (3,4,5NN).

Definition of an optimum value l has appeared very uneasy problem. It is connected with the instability of a method concerning change of parameter l. The forecast has been carried out with values $l = 10$ (when $n < 60$) and $l = 20$ (otherwise).

Let's note, that the considered method shows quite good enough results on long time series, because finding of the most adequate neighbours demands knowledge enough history (in which we search for similar parts).

The method shows absolutely inadequate results on very short time series. Partially it is the result of necessity to take small values l.

III. Conclusion

The basic conclusions can be made after end of competition and the publication of results. As in our experiments comparison of the method with others (MLP, TLRNN, ENN, SVR, Decision&Regression Trees, Fuzzy Predictors) was not made. The basic purpose of research was a choice of parameters of model, transformations A, $A_1, \dots, A_q$, proximity measure between parts of time series.

Acknowledgment

The author thanks challenge organizers for running this interesting competition.

A. G. D'yakonov is with the Moscow State University, Moscow, Russia, (e-mail: dxla@km.ru).

References

[1] A.S. Weigend, *Time series prediction: forecasting the future and understanding the past*. Proceedings of the NATO Advanced Research Workshop on Comparative Time Series Analysis held in Santa Fe, New Mexico, May 14 - 17,1992, Reading, Addison-Wesley.

[2] J.S. Armstrong (Ed.), *Principles of forecasting: a handbook for researchers and practitioners.* Boston; London, Kluwer Academic.

[3] S.F. Crone, P.C. Graffeille, An evaluation framework for publications on artificial neural networks in sales forecasting. Ic-Ai '04 & Mlmta'04, Vol 1 And 2, Proceedings. Athens, C S R E A Press.

[4] J.A. Suykens, J. Vandewalle, Nonlinear Modeling: advanced black-box techniques, Boston, Kluwer Academic Publishers.

[5] www.neural-forecasting-competition.com

[6] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*, 2001 John Wiley & Sons, Inc.

[7] B.V. Dasarathy, *Nearest neighbour (NN) Norms: NN pattern classification techniques*. IEEE Computer Society Press, 1991.