Physica A 444 (2016) 680–688

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Using string invariants for prediction searching for optimal parameters

Marek Bundzel [a], Tomáš Kasanický [b], Richard Pinčák [c]

[a] Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

[b] Institute of Informatics, Slovak Academy of Sciences, Slovak Republic

[c] Institute of Experimental Physics, Slovak Academy of Sciences, Slovak Republic

HIGHLIGHTS

- We have developed a novel prediction method based on string invariants.
- The method does not require learning but a small set of parameters must be set to achieve optimal performance.
- We have implemented an evolutionary algorithm for the parametric optimization.
- We have tested the performance of the method on artificial and real world data.
- We compared the performance to statistical methods and to a number of artificial intelligence methods.

ARTICLE INFO

Article history:
Received 15 July 2015
Available online 29 October 2015

Keywords:
String theory and string invariants
Evolutionary optimization
Artificial intelligence

ABSTRACT

We have developed a novel prediction method based on string invariants. The method does not require learning but a small set of parameters must be set to achieve optimal performance. We have implemented an evolutionary algorithm for the parametric optimization. We have tested the performance of the method on artificial and real world data and compared the performance to statistical methods and to a number of artificial intelligence methods. We have used data and the results of a prediction competition as a benchmark. The results show that the method performs well in single step prediction but the method's performance for multiple step prediction needs to be improved. The method works well for a wide range of parameters.

1. Introduction

The string theory was developed over the past 25 years and it has achieved a high degree of popularity and respect among the physicists [1]. The prediction model that we have developed transfers modern physical ideas into the field of time series prediction. The physical statistical viewpoint proved the ability to describe systems where many-body effects dominate. The envisioned application field of the proposed method is econophysics but the model is certainly not limited to applications in economy. Bottom-up approaches may have difficulties to follow the behavior of the complex economic systems where autonomous models encounter intrinsic variability. The primary motivation comes from the actual physical concepts [2,3].

We have named the new method the Prediction Model Based on String Invariants (PMBSI). PMBSI is based on the approaches described in Ref. [4] and extends the previous work. In Ref. [5] we have performed comparative experimental

E-mail addresses: marek.bundzel@tuke.sk (M. Bundzel), kasanicky@neuron.tuke.sk (T. Kasanický), pincak@saske.sk (R. Pinčák).

http://dx.doi.org/10.1016/j.physa.2015.10.050
0378-4371/

analysis aimed to identify the strengths and the weaknesses of PMBSI and to compare its performance to Support Vector Machine (SVM). PMBSI also represents one of the first attempts to apply the string theory in the field of time-series forecast and not only in high energy physics. We describe briefly the prediction model below.

PMBSI needs several parameters to be set to achieve the optimal performance. We have implemented an evolutionary algorithm to find the optimal parameters. The implementation is described below. We show the previously achieved results and compare them to the results achieved with evolutionary optimized parameters. We have also tested PMBSI on 111 time series used in a 2008 time series forecast competition. Thus we could compare its performance to an extensive range of methods.

2. State of the art

Linear methods often work well and may well provide an adequate approximation for the task at hand and are mathematically and practically convenient. However, the real life generating processes are often non-linear. Therefore plenty of non-linear forecast models based on different approaches have been created (e.g. GARCH [6], ARCH [7], ARMA [8], ARIMA [9] etc.). Presently, the perhaps most used methods are based on computational intelligence. A number of research articles compare Artificial Neural Networks (ANN) and Support Vector Machines (SVM) to each other and to other more traditional non-linear statistical methods. Tay and Cao [10] examined the feasibility of SVM in financial time series forecasting and compared it to a multilayer Back Propagation Neural Network (BPNN). They showed that SVM outperforms the BP neural network. Kamruzzaman and Sarker [11] modeled and predicted currency exchange rates using three ANN based models and a comparison was made with ARIMA model. The results showed that all the ANN based models outperform ARIMA model. Chen et al. [12] compared SVM and BPNN taking auto-regressive model as a benchmark in forecasting the six major Asian stock markets. Again, both the SVM and BPNN outperformed the traditional models. SVM implements the structural risk minimization—an inductive principle for model selection used for learning from finite training data sets. For this reason SVM is often chosen as a benchmark to compare other non-linear models. Many nature inspired prediction methods have been tested. Egrioglu [13] applied Particle Swarm Optimization on fuzzy series forecasting. LIU et al. [14,15] applied ANFIS and evolutionary optimization to forecast TAIEX. So far no non-linear black box method reached significant performance superiority over others.

3. Prediction model based on string invariants

The original time-series (τ) is converted as follows

$$\frac{p(\tau+h)-p(\tau)}{p(\tau+h)} \tag{1}$$

where h denotes the lag between $p(\tau)$ and $p(\tau+h)$, τ is the index of the time series element. On financial data, e.g. on the series of the quotations of the mean currency exchange rate, this operation would convert the original time-series into a series of returns. Using the string theory let us first define the 1-end-point open string map

$$P^{(1)}(\tau,h)=\frac{p(\tau+h)-p(\tau)}{p(\tau+h)}, \quad h\in\langle 0,l_s\rangle, \tag{2}$$

where the superscript (1) refers to the number of endpoints and l_s to the length of the string (string size). l_s is a positive integer. The variable h may be interpreted as a variable which extends along the extra dimension limited by the string size l_s. A natural consequence of the transform, Eq. (2), is the fulfillment of the boundary condition

$$P^{(1)}(\tau,0)=0, \tag{3}$$

which holds for any τ. To enhance the influence of rare events a power-law Q-deformed model is introduced

$$P^{(1)}(\tau,h)=\left(1-\left[\frac{p(\tau)}{p(\tau+h)}\right]^{Q}\right), \quad Q>0. \tag{4}$$

The 1-end-point string has defined the origin and it reflects the linear trend in $p(\cdot)$ at the scale l_s. The presence of a long-term trend is partially corrected by fixing $P^{(2)}(\tau,h)$ at $h=l_s$. The open string with two end points is introduced via the nonlinear map which combines information about trends of p at two sequential segments

$$P^{(2)}(\tau,h)=\left(1-\left[\frac{p(\tau)}{p(\tau+h)}\right]^{Q}\right)\left(1-\left[\frac{p(\tau+h)}{p(\tau+l_s)}\right]^{Q}\right), \quad h\in\langle 0,l_s\rangle. \tag{5}$$

The map is suggested to include boundary conditions of *Dirichlet type*

$$P^{(2)}(\tau,0)=P_q(\tau,l_s)=0, \quad \text{at all } \tau. \tag{6}$$