Chapter 8
A New Approach and Its Applications for Time Series Analysis and Prediction Based on Moving Average of $n^{th}$-Order Difference

Yang Lan\textsuperscript{1} and Daniel Neagu\textsuperscript{2}

\textsuperscript{1} Department of Computing, School of Computing, Informatics and Media, University of Bradford, Bradford, BD7 1DP, UK
Y.Lan@bradford.ac.uk

\textsuperscript{2} Department of Computing, School of Computing, Informatics and Media, University of Bradford, Bradford, BD7 1DP, UK
D.Neagu@bradford.ac.uk

Abstract. As a typical problem in data mining, Time Series Predictions are widely applied in various domains. The approach focuses on series of observations, with the aim that, using mathematics, statistics and artificial intelligence methods, to analyze, process and make a prediction on the next most probable value based on a number of previous values. We propose an algorithm using the average sum of $n^{th}$-order difference of series terms with limited range margins, in order to establish a way to predict the next series term based on both, the original data set and a negligible error. The algorithm performances are evaluated using measurement data sets on monthly average Sunspot Number, Earthquakes and Pseudo-Periodical Synthetic Time Series.

1 Introduction

The importance of time for human activities has been emphasized from early times of civilization. Historical data analysis has been related to agriculture (sun and moon cycles, weather) and safety (earthquakes, floods). Nowadays, given technological advances in computational power and memory storage, it becomes a functionality of immediate use for industrial or economical processes.

Time series prediction proposes algorithms for which past data (mainly finite observation sequences of data points related to uniform time intervals) are used to generate models to forecast future data points of the series. It is widely applied in various domains from finance (stock markets) and economy (electricity consumption), meteorology, signal processing to disaster warning (river floods, earthquakes) and solar activity forecasting. The time series analysis was based originally on tools including mathematical modeling, time-frequency analysis (Fourier and wavelet transformations) but started using in the last years machine learning methods such as Artificial Neural Networks (ANNs) (time-delay
From a procedural perspective, using computational approaches may first require mathematical analysis to describe and breakdown the initial time series problem into simpler sub-problems for further computational modeling. A well-known approach in time series understanding and prediction is Auto-Regressive Moving Average (ARMA) which comes with the advantage of addressing auto-regressive terms and moving average terms.

A historical main constraint in using mathematical series models for prediction was the fact that the performance of the model is related to the length of data series, but nowadays is not anymore an issue from neither computational nor data storage and processing points of view. However, most machine learning methods face the difficulty of requiring a priori knowledge about the problem at hand. On the other hand, results of some traditional methods applied in time series analysis can not satisfy the demand of specific applications. We intend to address these drawbacks for the restricted problem of pseudo-periodical series with limited boundaries by a two-step approach: we propose hereby a new algorithm to approximate the time series terms using the moving average of $n^{th}$-order difference of already known values and intend to address later the problem of error of approximation by a hybrid model. Therefore future work is proposed to identify as accurately as possible a general approximation by use of a supervised-learning model to forecast a further approximation error if found necessary.

We propose an algorithm for efficient mining of pseudo-periodical time series with direct applications to sunspot number, earthquake and pseudo-periodical synthetic time series prediction, by exploring some interesting properties related to moving average of first-order difference for bounded time series. A further generalization to the use of the sum of $n^{th}$-order difference to increase forecast performances and a hybrid approach to combine the results of the moving average of $n^{th}$-order difference of time series with a supervised-learning model of the error of value approximation are also proposed. We study the possibility that pre-processing of time series combined with a priori knowledge and hybrid models can increase prediction performances for time series, even for mining noisy data. The results highlight our proposed algorithm’s efficiency in mining bounded pseudo-periodical patterns in time series with direct applications in sunspot time series prediction, earthquake time series prediction and pseudo-periodical synthetic time series prediction.

The following section introduces the notations and definitions on bounded pseudo-periodical time series. Section 3 describes terms and proofs used in our approach, error approximation with the use of ANNs and our algorithm. Section 4 proposes a way to define the suitable level order $n$ of difference operator and index $m$ for increasing the prediction precision. Case studies on the monthly average of sunspot number time series prediction, earthquake time series prediction and pseudo-periodical synthetic time series prediction are described in sections 5, 6 and 7. Section 8 shows the prediction results comparison between the algorithm we propose, Linear Regression (LR) method and Auto-Regression