Knowledge-Based Systems 105 (2016) 29–37

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Dynamic optimization of fuzzy cognitive maps for time series forecasting☆

Jose L. Salmeron[a,b,c,*], Wojciech Froelich[d]

[a] Data Science Lab, Universidad Pablo de Olavide, Seville, Spain
[b] Universidad Autónoma de Chile, Chile
[c] Univerzita Hradec Králové, Czech Republic
[d] Institute of Computer Science, University of Silesia, Sosnowiec, Poland

ARTICLE INFO

Article history:
Received 21 January 2016
Revised 21 April 2016
Accepted 24 April 2016
Available online 26 April 2016

Keywords:
Fuzzy cognitive maps
Forecasting time series
Machine learning

ABSTRACT

In this paper we propose a new approach to learning fuzzy cognitive maps (FCMs) as a predictive model for time series forecasting. The first contribution of this paper is the dynamic optimization of the FCM structure, i.e., we propose to select concepts involved in the FCM model before every prediction is made. In addition, the FCM transformation function together with the corresponding parameters are proposed to be optimized dynamically. Finally, the FCM weights are learned. In this way, the entire FCM model is learned in a completely new manner, i.e., it is continuously adapted to the current local characteristics of the forecasted time series. To optimize all of the aforementioned elements, we apply and compare 5 different population-based algorithms: genetic, particle swarm optimization, simulated annealing, artificial bee colony and differential evolution. For the evaluation of the proposed approach we use 11 publicly available data sets. The results of comparative experiments provide evidence that our approach offers a competitive forecasting method that outperforms many state-of-the-art forecasting models. We recommend to use our FCM-based approach for the forecasting of time series that are linear and tend to be trend stationary.

1. Introduction

Forecasting time series is an important problem that has been investigated for many years by numerous researchers. In spite of significant achievements in this domain, the topic still raises active interest. One of the lesser-known prediction models that can be applied to forecasting time series is the Fuzzy Cognitive Map (FCM). In this paper we investigate the FCM model and propose a new approach to learning it. The objective of our research is to improve the forecasting accuracy of the FCM model.

FCM is a soft computing technique mixing Artificial Neural Networks and Fuzzy Logic [1]. It is a signed fuzzy weighted digraph consisting of nodes and arcs between them. The nodes (concepts) are the fuzzy sets that describe the modeled problem and the weighted arcs represent causal dependencies between the concepts. FCM is a nonlinear model because the cumulative impact of causal concepts (parents within the graph) on the effect concept (child node) is transformed by a nonlinear transformation function. An advantage of the FCM model over Artificial Neural Networks is that FCMs can be interpreted easily by humans, and each FCM node and arc has a specific meaning known to the expert [2–4]. FCM can be interpreted as a fuzzy granular model [5].

Extensive research has been conducted with the purpose of improving the forecasting accuracy demonstrated by the FCM. It is possible to distinguish two general approaches to the problem. The first one assumes that the historical data used for learning FCMs are generated from existing FCMs [6]. In this case, the learned FCM mimics the source data in the long-term horizon. In the second approach, a problem with forecasting real-world time series is addressed. An FCM model has been effectively applied to forecast real-world univariate [7–9] and multivariate [10–13] time series. In this paper we use FCMs to forecast univariate time series.

It is known from the literature that the recognition of appropriate regressors is a key factor in determining forecasting accuracy [14]. When using FCMs for the forecasting of univariate time series, the concepts of the FCM play a role similar to that of the regressors in auto-regressive models. For this reason, the selection of concepts included within the FCM is an important problem

☆ This manuscript is a collaborative effort.

* Corresponding author. Tel.: +34954349063.
E-mail addresses: salmeron@acm.org (J.L. Salmeron), wojciech.froelich@us.edu.pl (W. Froelich).
URL: http://eps.upo.es/salmeron (J.L. Salmeron)

http://dx.doi.org/10.1016/j.knosys.2016.04.023
0950-7051/

influencing the efficiency of forecasting [8,9]. In all previous works, the structure of the FCM was either assumed by an expert or generated indirectly while learning the FCM's weights. In the latter case, all concepts with the corresponding arcs were taken into account during the learning process. Therefore, the final trained FCM also contained all concepts and arcs, although the values of weights were very small for many of the arcs. To solve this issue, filtering of weights was performed that led to the removal from the FCM's structure of those arcs for which the absolute value of weights was below a certain threshold [2,8,9]. As a consequence of that filtering, certain concepts were evaluated as less important and were subsequently removed from the FCM. For the first time in this paper, we propose to select the concepts to be included in the FCM dynamically without the need to perform the previously used filtering.

In studies [2,8,15], the sliding window approach was used to learn the FCM model. The length of the window was assumed to be the parameter, constant for the entire learning part of time series. The experiments revealed that the performance of forecasting largely depended on the assumed window size [2,8]. To avoid the laborious procedure of optimizing this parameter, and to adjust this parameter to the local characteristics of time series, we propose to optimize it using a population-based algorithm.

The substantial influence of the transformation function on the accuracy of forecasting has been theoretically and practically recognized in numerous papers devoted to FCMs [11,15,16]. However, none of these papers applied dynamic selection to the transformation function, dependent on the considered learning period, we propose exactly that in this paper.

To the best of our knowledge, all existing works considered the following approach to learning FCMs. Independent of the length and the other characteristics of the learning part of time series, the FCM model always contained the same number of concepts and the same transformation function. The weights of the FCM were learned using the entire part of the learning data. In all previous works, the learning algorithm tried to adapt FCM weights to the characteristics of the entire learning part of time series. This means that the influence of the first step and the last step of the learning part of time series on the resulting FCM were the same. In this paper, we propose to optimize not only the weights, but all elements of FCMs with respect to the most recent part of the learning data set. The length of this part of time series is also optimized. The method proposed in this paper makes the FCM a locally working forecasting model, instead of the global approach that was until now applied in the literature.

The contributions of this study are as follows:

- We propose retraining FCM dynamically, adapting it to the local characteristics of the forecasted time series.
- We propose optimizing the FCM model by selecting which concepts of the FCM should be included within its structure. Thus, after mapping the FCM to the variables representing the lagged time series, the order of the predictive model is optimized.
- We propose optimizing the selection of the FCM's transformation function using a pool of functions. After the function is selected, its parameters and thus its shape are optimized.

To learn the FCM model, including all of the aforementioned elements, we apply the following population-based algorithms: a real-coded genetic algorithm, particle swarm optimization, simulated annealing, artificial bee colony and differential evolution.

We validate and evaluate the proposed FCM-based model using 11 publicly available data. It is obvious that none of the existing forecasting models is the best for all possible data. For that reason we recognize a single criterion by which to identify time series, for which the application of FCM brings more benefits in comparison to competitive forecasting models. After numerous experiments it

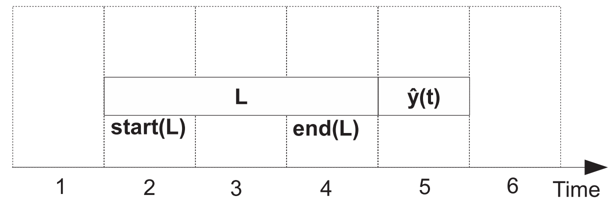

Fig. 1. Exemplary sliding window.

turned out that our FCM-based approach is recommended for the forecasting of time series that are linear and tend to be trend stationary.

The remainder of this paper is organized as follows. First, in Section 2 we provide background knowledge on time series forecasting. In Section 3 we introduce the theory of Fuzzy Cognitive Maps together with an overview of the population-based optimization algorithms that are used to learn FCM. In Section 4 we describe the proposed methodology of learning FCM applied to one-step ahead forecasting of time series. The experimental evaluation of the proposed approach is described in Section 5. Section 6 concludes the paper.

2. Preliminaries of time series

Let $y \in \Re$ be a real-valued variable whose values are observed at a discrete time scale $t \in [1, 2, \ldots, n]$, where $n \in \aleph$ is the length of the considered period. A time series is a sequence $\{y(t)\} = \{y(1), y(2), \ldots, y(n)\}$. The goal of one-step ahead forecasting is to calculate $\hat{y}(t) = M(\{y(t-1)\})$, where: $\hat{y}(t)$ denotes the forecast and M is the forecasting model. The challenge is to select and train such a model M that produces the lowest absolute values of forecasting errors calculated as $e(t) = \hat{y}(t) - y(t)$.

The concept of a sliding window can be applied when the most recent part of the time series is used for the training of the predictive model. The sliding window is defined as a subsequence $L \subset \{y(t)\}$ of the considered time series. It begins at time $start(L)$ and finishes at $end(L)$. We assume that $end(L) = t-1$ is the last time step at which the variable y is observable. During the period $t \in [start(L), t-1]$, the predictive model FCM is learned. A single one-step ahead forecast is made as $\hat{y}(t)$. The value of $start(L)$ is a parameter. The time step $t = end(L) + 1$, at which point the forecast is made, moves forward as time flows. An exemplary sliding window is depicted in Fig. 1.

The series of individual forecasting errors $\{e(t)\}$ is generated while performing $n - length(L)$ learn-and-test trials. To assess all forecasting errors accumulated over the considered period of time, diverse error measures can be applied. For the purpose of this paper, we use only one of them that is widely used in the literature: Mean Absolute Percentage Error (MAPE), given as formula (1). Besides its popularity, the reason for the selection is that the MAPE is scale-independent and intuitively easy to interpret as a sum of individual errors related to the actual values of time series.

$$MAPE = \frac{1}{n} \cdot \sum_{t=1}^{n} \left| \frac{e(t)}{y(t)} \right| \cdot 100\%, \text{ for } y(t) \neq 0 \tag{1}$$

There are several statistical tests that determine the characteristics of time series. One of the objectives of using these tests is to support the selection of the best forecasting model dependent on the results of the tests. It is worth mentioning that such a selection method is never definitive; comparative experiments in terms of forecasting accuracy are always required. For the purpose of this research, before every group of experiments we provide a corresponding table including numerical values of the following tests.