Applied Soft Computing 14 (2014) 305–315

Contents lists available at ScienceDirect

Applied Soft Computing

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

Water leakage forecasting: the application of a modified fuzzy evolving algorithm

Lech Birek[a], Dobrila Petrovic[a,*], John Boylan[b]

[a] Coventry University, Priory Street, CV1 5FB Coventry, United Kingdom
[b] Buckinghamshire New University, Queen Alexandra Road, HP11 2JZ High Wycombe, United Kingdom

ARTICLE INFO

Article history:
Received 31 October 2012
Received in revised form 13 April 2013
Accepted 30 May 2013
Available online 10 June 2013

Keywords:
Fuzzy If-Then rules
Evolving fuzzy system
Forecasting
Evolving clustering
Leakage

ABSTRACT

This paper investigates the use of evolving fuzzy algorithms in forecasting. An evolving Takagi-Sugeno (eTS) algorithm, which is based on a recursive version of the subtractive algorithm is considered. It groups data into several clusters based on Euclidean distance between the relevant independent variables. The Mod eTS algorithm, which incorporates a modified dynamic update of cluster radii while accommodating new available data is proposed. The created clusters serve as a base for fuzzy If-Then rules with Gaussian membership functions which are defined using the cluster centres and have linear functions in the consequent i.e., Then parts of rules. The parameters of the linear functions are calculated using a weighted version of the Recursive Least Squares algorithm. The proposed algorithm is applied to a leakage forecasting problem faced by one of the leading UK water supplying companies. Using the real world data provided by the company the forecasting results obtained from the proposed modified eTS algorithm, Mod eTS, are compared to the standard eTS algorithm, exTS, eTS+and fuzzy C-means clustering algorithm and some standard statistical forecasting methods. Different measures of forecasting accuracy are used. The results show higher accuracy achieved by applying the algorithm proposed compared to other fuzzy clustering algorithms and statistical methods. Similar results are obtained when comparing with other fuzzy evolving algorithms with dynamic cluster radii. Furthermore the algorithm generates typically a smaller number of clusters than standard fuzzy forecasting methods which leads to more transparent forecasting models.

1. Introduction

Development and application of forecasting models to be used in real-world scenarios are often difficult tasks because of non-linear relationships between dependent and independent variables, measurement errors and incomplete datasets. In the last few decades, fuzzy logic has been successfully used to model complex, dynamic problems such as forecasting that would otherwise be difficult to accomplish using conventional mathematical approaches. The advantage of using fuzzy logic lays in its ability to express non-linear relations among variables, typically by combining several linear sub-models, expressed in the form of fuzzy If-Then rules. The rules are often generated based on experts' knowledge. This, however, requires the presence of an experienced specialist, and it is usually time consuming and not feasible for large scale applications.

One of the ways to overcome this problem is to use data clustering. The data points which share similar properties are grouped into clusters which in turn are a base for the fuzzy If-Then rules. The recent advances in the field of fuzzy clustering [1,2] have allowed for real-time generation and update of fuzzy If-Then rules. This is particularly helpful in situations where it is important to adapt the rule structure to changing conditions as well as being able to control the way the clusters are generated in the real time.

In this paper, evolving fuzzy systems are applied to the forecasting of water leakage for one of the leading water supply companies in the UK. This approach is particularly useful for this application as it is often difficult to determine the relationship between the dependent and independent variables and it is important to adapt the forecasting model to changing conditions. Having a number of regions of operation, it is also inconvenient to generate different forecasting models for each one of these regions. This paper is novel in three ways: (i) a fuzzy (Takagi-Sugeno) evolving algorithm is adapted in order to be suitable for a forecasting application, (ii) the generation of the cluster radii is conducted in a different manner as compared to other existing methods, and (iii) fuzzy and statistical forecasting methods are evaluated and compared in the context of water leakage for the first time.

* Corresponding author at: Computing Department, Faculty of Engineering and Computing, Coventry University, Priory Street, Coventry CV1 5FB, United Kingdom. Tel.: +44 24 77659181; fax: +44 24 76888585.
E-mail addresses: birekl@uni.coventry.ac.uk, lebirek@o2.pl (L. Birek), d.petrovic@coventry.ac.uk (D. Petrovic), John.Boylan@bucks.ac.uk (J. Boylan).

1568-4946/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.asoc.2013.05.021

The paper is organized as follows: Section 2 provides a literature review on fuzzy logic based forecasting methods. Section 3 presents the evolving Takagi-Sugeno (eTS) algorithm, introduces the modifications of the standard eTS algorithm and explains how the algorithm is used in forecasting. In Section 4 the case study is presented which introduces the leakage forecasting problem and points out some of the factors influencing the leakage. The application of the algorithm proposed, the results obtained and comparison with some other fuzzy forecasting methods, as well as some standard statistical methods such as Naive, Holt-Winters and linear regression are presented in Section 5. The paper closes with some concluding remarks and indications of future work.

2. Literature review

The application of fuzzy sets theory to forecasting has been of interest to many researchers around the world. One of the first applications of fuzzy sets was in forecasting tax revenue using the language patterns extracted from the available data [3]. Further work has been done in using language values in forecasting. In [4] Song and Chissom proposed the fuzzy time series approach; the method involved defining and partitioning the available data space into a universe of discourse which was fuzzified and used in the forecasting process. The method was applied and investigated in [5,6] and further modifications were introduced in [7,8].

Another area of fuzzy logic which found its application in forecasting has been fuzzy clustering, where historical data is grouped into clusters based on a distance measure. The clusters are then used to generate fuzzy If-Then rules which are applied in a forecasting process. Considerable research has been done in the application of the well-known fuzzy C-means algorithm. In [9], the fuzzy C-means algorithm is used to generate different lengths of intervals to be used in conjunction with the fuzzy time series approach described above. In [10], the extended fuzzy C-means algorithm was applied and used to help forecast newspaper demand in order to reduce newspaper losses. Use of fuzzy C-means algorithm to automatically build a fuzzy If-Then rule structure from the clusters was researched in [11]. Fuzzy If-Then rules were used to obtain forecasts based on calculated weights of the activated rules. The method was applied to temperature forecasting based on the previously recorded temperature and cloud density, as well as to the forecast of the Taiwan stock market index. The drawback of using fuzzy C-means, however, was that the method required setting the number of clusters in advance. This led to time consuming process of finding the number of clusters which would produce the best results. One of the solutions presented in [12] is to use the fuzzy possibilistic approach. Another solution, proposed by Yager and Filev [13], called the Mountain method, helped in estimating the initial coordinates of the cluster centres to be used in other clustering algorithms, such as C-means. Other clustering algorithms emerged directly from the method proposed, such as the subtractive clustering algorithm [14], used for identification of the Takagi-Sugeno fuzzy models from the available data.

The extension of the subtractive clustering algorithm called eTS, presented in [15] and [16], allowed for the dynamic update of the cluster structure by using a notion of the potential of each data sample. The resulting fuzzy If-Then rules, as well as the consequent parameters of the resulting Linear Equations, were also updated in a recursive way. Different versions and modifications of the algorithm and various applications have been researched since then. In [7], a simplified version of the eTS algorithm was introduced, which utilized scatter instead of the potential and included ageing of the rules. In [18], exTS algorithm was presented, which involved the recursive update of the cluster radii according to a data distribution. eTS+algorithm, presented in [2], further enhanced the eTS method with the on-line monitoring of the quality of generated clusters, on-line structure simplification and on-line input selection. More methods have been developed presenting different approaches, such as neural fuzzy systems (SaFIN [19] and DENFIS [20]), modification to standard algorithms, like C-means [21], Gustafson-Kessel [22] and Gath-Geva [23] clustering, and new ones, such as FLEXFIS algorithm [24]. Complexity reduction was addressed in [25], by analysing the approach of removing local redundancies during the training period. Evolving fuzzy systems were also applied and evaluated on classification problems [26]. The on-line evolving fuzzy classifiers presented there could be used with different model architectures. Some practical applications are described in [27–29]. A review of different fuzzy evolving approaches which emerged over the past decade is presented in [1].

One of the areas which did not receive enough attention is the application of evolving fuzzy algorithms to forecasting problems. In [28], the eTS algorithm was applied to a time series data set from a Neural Network (NN) competition. The method was successfully applied, but the accuracy of the proposed approach was not compared with any other method. In [30], the evolving fuzzy approach was also applied to a NN competition, with a focus on proposing different top-down approaches to improve the accuracy of a detailed forecast by aggregating daily reading into a weekly time-series. The results for the petrol volume sales estimation using the recursive Gustafson-Kessel clustering were presented in [31]. Those studies show that the evolving fuzzy methods can be used in forecasting applications. However, the obtained results are not compared with any of the widely accepted statistical methods in terms of the forecasting accuracy. Another issue is that the fuzzy evolving algorithms are very often applied to data generated by mathematical models. The lack of comparison and the use of the generated data makes it difficult to assess the usability of the fuzzy forecasting methods in the practical forecasting application.

This study aims at addressing those issues by focusing on the application of a Mod eTS algorithm to a real life problem of one year ahead multivariate water leakage forecasting. The data used consists of multiple independent and dependent variables, with visible seasonal events. The evolving fuzzy algorithm which has been used is based on the eTS with dynamic update of cluster radii. Although the dynamic radii update has already been explored in the previous extensions to eTS, the update process has been changed to accommodate to the features of this application – in particular to take into account the importance of the high leakage during certain seasonal periods, which has the highest influence on the average yearly leakage values. The performance of the proposed approach and other fuzzy forecasting and statistical methods are investigated and the results are compared.

3. Mod eTS algorithm

3.1. Fuzzy rules generation with subtractive clustering

Before introducing the modified evolving Takagi-Sugeno algorithm for forecasting, we will first focus on the standard, offline subtractive clustering algorithm. The subtractive clustering algorithm [14] is a fuzzy model-based identification method which uses a modification of Mountain clustering method proposed by Yager and Filev [13]. Mountain clustering was initially used to help in estimation of the number of cluster centres for the fuzzy C-means algorithm. In the fuzzy C-means, coordinates of the cluster centres are obtained as a result of minimizing the cost function which takes into account the distance between cluster centres and data samples and the corresponding membership degrees. This however, requires the number of clusters to be specified in advance.