Pattern Recognition 45 (2012) 1076–1091

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

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

Classification trees for time series

Ahlame Douzal-Chouakria*, Cécile Amblard

LIG University of Joseph Fourier Grenoble 1, 38041 Grenoble cedex 9, France

ARTICLE INFO

Article history:
Received 17 June 2010
Received in revised form
20 July 2011
Accepted 12 August 2011
Available online 23 August 2011

Keywords:
Time series proximity measures
Supervised classification
Classification trees
Learning metric

ABSTRACT

This paper proposes an extension of classification trees to time series input variables. A new split criterion based on time series proximities is introduced. First, the criterion relies on an adaptive (i.e., parameterized) time series metric to cover both behaviors and values proximities. The metrics parameters may change from one internal node to another to achieve the best bisection of the set of time series. Second, the criterion involves the automatic extraction of the most discriminating subsequences. The proposed time series classification tree is applied to a wide range of datasets: public and new, real and synthetic, univariate and multivariate data. We show, through the experiments performed in this study, that the proposed tree outperforms temporal trees using standard time series distances and performs well compared to other competitive time series classifiers.

1. Introduction

Time series classification has been the subject of extensive research in the last several years. A first category of proposals consists of mapping time series to a new description space where conventional classifiers can be applied. Signal processing or statistical tools are commonly used to project time series into a given functional basis space. For instance, such projection can be performed by a Fourier or wavelet transform, a polynomial or an ARIMA approximation. Standard classifiers are subsequently applied on the fitted basis coefficients [1–5]. A second class of works proposes new heuristics, generally starting with the time series segmentation to extract prototypes that best characterize the time series classes. The prototypes, defined by such factors as a set of subsequences or regions of values, are subsequently described by a set of numerical features where standard classifiers can be applied [6–11]. A third category may be distinguished that consists of the hidden Markov models [12], which is frequently used for speech recognition and signal processing.

This paper focuses on a distance-based approach to extending classification trees to temporal data. We propose a new time series split criterion characterized by, on the one hand, the use of an adaptive metric to cover both behaviors and values proximities. This metric may change from one node to another according to the set of time series to be divided. On the other hand, the proposed split involves an automatic extraction of the most discriminating subsequences (i.e., segments of time series). We show, through the experiments performed, that the proposed tree outperforms temporal trees using standard time series distances and performs well compared to other competitive time series classifiers.

The rest of the paper is organized as follows. In the next section, we discuss two distance-based temporal trees proposed by Yamada et al. [13] and by Balakrishnan and Madigan [14]. In Section 3, the major metrics for time series are presented in a novel unified formalism. Section 4 presents the new time series classification tree, provides the main algorithms and discusses their complexity. In Section 5, the proposed classification tree is performed on six public and three new simulated datasets. The induced trees are compared to temporal trees using standard distances and are compared to other competitive time series classifiers.

2. Related works

In this section, we describe two temporal classification trees proposed by Yamada et al. in 2003 and by Balakrishnan and Madigan in 2006. Both works build binary classification trees in which internal nodes are labeled by one or two time series. Proposed classifiers are mainly based on new split tests to bisect the set of time series within internal nodes most effectively.

Yamada et al. [13] proposes two split tests. The first test, called the *standard-example* split test, uses an exhaustive search to select one existing time series (called the standard time series), leading to division with a maximum purity gain ratio. The first child node is composed of time series with a distance to the standard time series that is less than a given threshold, while the second child node contains the remaining time series. If more than one standard time series provides the largest value of the purity gain

* Corresponding author. Tel.: +33 4 56 52 00 68; fax: +33 4 56 52 00 22.
E-mail addresses: Ahlame.Douzal@imag.fr (A. Douzal-Chouakria), Cecile.Amblard@imag.fr (C. Amblard).

0031-3203/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2011.08.018

ratio, a class isolation criterion is used to select the split that exhibits the most dissimilar child nodes.

The second proposed split test, which is called the *cluster-example* split test, performs an exhaustive search for two standard time series. The bisection is constructed by assigning each time series to the nearest standard time series. Similarly, the purity gain ratio and the class isolation criterion are used to select the best split test. For both split tests, the dynamic time warping is used as the time series proximity measure.

Balakrishnan and Madigan [14] look for a pair of reference time series that best bisects the set of time series according to a clustering-goodness criterion. For this purpose, a k-means algorithm is used. This algorithm ensures a partitioning that optimizes clustering criteria, namely, the compactness and isolation of the clusters but not their purity. To alleviate this problem, the authors repeat the k-means clustering several times and select the partition that gives the highest Gini index. The centers of the clusters define the pair of reference time series for the split test. For the time series proximities, both the Euclidean distance and the dynamic time warping are used to compare the efficiency of the obtained classification trees.

In summary, the *cluster-example* split test of Yamada et al. [13] and the one proposed by Balakrishnan and Madigan [14] are highly similar. The former first looks for a set of time series bisections with the highest purity clusters (i.e., the highest Gini index) and picks the one optimizing some clustering criteria (i.e., maximizing the separability of the clusters), whereas the latter first looks for a set of splits that optimize clustering criteria (i.e., k-means criteria) and accordingly selects the one exhibiting the highest purity clusters (i.e., maximizing the Gini index). When giving priority to a clustering criterion instead of the purity of the clusters, the split test may fail to select bisections of lower clustering criteria but of higher purity.

Let us make some remarks about the above proposed split tests. First, as for many distance-based approaches, the Euclidean distance and the dynamic time warping are considered for the time series proximities. These standard measures are values-based metrics and ignore the behaviors of the time series as discussed in Section 3. Second, the proposed splits use the same metric to divide all the nodes, but the peculiarities of the time series may change from one node to another. Finally, the time series distances are calculated using the whole time series values, even though the discrimination is determined by particular subsequences.

3. Time series metrics

We present, in a unified formalism, three categories of time series metrics. The first category relies on two standard values-based metrics: the dynamic time warping and the Euclidean distance. In the second category, we recall the definition of the correlation coefficient and the temporal correlation coefficient, which are used as behavior-based metrics. In the third category, we present a model to cover both behaviors and values components of time series. In particular, extensions of the Euclidean distance and of the dynamic time warping are provided to cover both behaviors and values proximities.

Let $S_1=(u_1,\ldots,u_p)$ and $S_2=(v_1,\ldots,v_q)$ be two time series of p and q values observed at the time instants $(t_1,\ldots,t_p)$ and $(t'_1,\ldots,t'_q)$, respectively. A mapping r between S_1 and S_2 is defined as a sequence of m pairs of observations $((u_{a_1},v_{b_1}),(u_{a_2},v_{b_2}),\ldots,(u_{a_m},v_{b_m}))$, with $a_i\in\{1,\ldots,p\}$, $b_i\in\{1,\ldots,q\}$, and $i\in\{1,\ldots,m-1\}$ obeying the order constraints:

$$a_1=1,\quad a_m=p,\quad a_{i+1}=a_i \text{ or } a_i+1 \text{ and,}$$

$$b_1=1,\quad b_m=q,\quad b_{i+1}=b_i \text{ or } b_i+1.$$

with $m\in[max(p,q),p+q-1]$. Let R be a subset of such mappings, possibly satisfying some additional constraints, and let $c(r)$ $(r\in R)$ be the mapping cost function measuring the distance between the coupled values in r. A unified formalism of the time series proximity measures, denoted *dUnif*, may be presented as an optimization problem minimizing the cost function $c(r)$ on the search space R:

$$dUnif_{(c,R)}(S_1,S_2)=\min_{r\in R}c(r). \tag{1}$$

3.1. Values-based metrics

For the cost function definition $c(r)=\sum_{i=1}^{m}|u_{a_i}-v_{b_i}|$, $dUnif_{(c,R)}$ (Eq. (1)) leads to the standard dynamic time warping [15]:

$$d_{Dtw}(S_1,S_2)=\min_{r\in R}\left(\sum_{i=1}^{m}|u_{a_i}-v_{b_i}|\right) \tag{2}$$

In the case of times series of the same length ($m=p=q$), and for the cost function definition $c(r)=(\sum_{i=1}^{m}(u_i-v_i)^2)^{1/2}$ minimized on $R=\{r_0\}$, $dUnif_{(c,R)}$ gives the Euclidean distance, with:

$$r_0=((u_1,v_1),(u_2,v_2),\ldots,(u_m,v_m)) \tag{3}$$

$$d_E(S_1,S_2)=c(r_0)=\left(\sum_{i=1}^{m}(u_i-v_i)^2\right)^{1/2} \tag{4}$$

The above cost functions $c(r)$ involve the differences between the aligned values, without allowance for the values neighborhoods. This characteristic can be illustrated by the following example. Let $S_i=(0,1,-3,-2)$, $S_j=(4,8,5,8)$, and $S_k=(2,-2,-1,-3)$ be the three time series given in Fig. 1. Note that S_i and S_j are close in behaviors (i.e., they increase or decrease simultaneously) and far apart in values, whereas S_i and S_k are close in values and opposite in behaviors (i.e., S_k increases when S_i decreases and vice-versa). Both the Euclidean distance and the dynamic time warping give S_i closer to S_k than to S_j with $d_E(S_i,S_k)=4.24<d_E(S_i,S_j)=15.13<d_E\ (S_j,S_k)=16.15$, and $d_{Dtw}(S_i,S_k)=6<d_{Dtw}(S_i,S_j)=29\leq d_{Dtw}(S_j,S_k)=29$.

3.2. Behavior-based metrics

Let us define two time series S_1 and S_2 to be similar in behavior if, during any observed period $[t_i,t_{i+1}]$, they increase or decrease simultaneously with the same growth rate. In contrast, they are considered to be opposite in behavior if, during any observed period $[t_i,t_{i+1}]$ in which S_1 increases, S_2 decreases and (vice-versa) with the same growth rate (in absolute value).

Until recently, many applications in different domains (e.g., speech recognition, system design control, functional MRI, micro-arrays and gene expression analysis) have used the Pearson correlation coefficient as a behavior proximity measure between signals [16–20]. Let us consider an equivalent formula for the

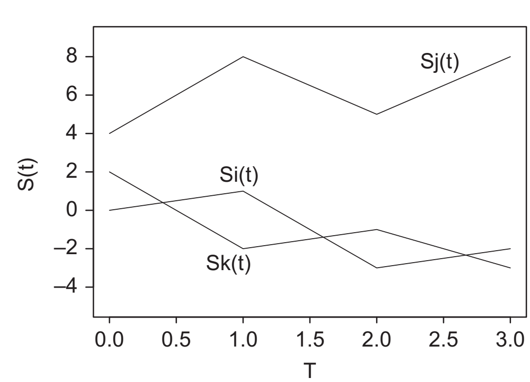

Fig. 1. Close on values and far on behavior vs. close on behavior and far on values.