Expert Systems With Applications 54 (2016) 294–303

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Unsupervised event detection and classification of multichannel signals

Angel Mur[a,*], Raquel Dormido[a], Jesús Vega[b], Sebastian Dormido-Canto[a], Natividad Duro[a]

[a] Department of Computer Sciences and Automatic Control, UNED, Juan del Rosal 16, 28040 Madrid, Spain
[b] National Fusion Laboratory by Magnetic Confinement, CIEMAT, Complutense 40, 28040 Madrid, Spain

ARTICLE INFO

Keywords:
Classification
Event characterization
Event detection
Multichannel signal
Temporal sequence

ABSTRACT

In this paper, we present a new unsupervised method to classify a set of Multichanel Signals (*MC*) with unknown events. Each signal is characterized by a sequence of events where the number of events, the start time and the duration between events can change randomly. The proposed method helps in the classification and event detection of the *MC* signals by an expert which usually becomes a tedious and difficult task. To this end, first, the problem of classification of *MC* signals characterized by a succession of events is analyzed by transforming the *MC* signals into a set of temporal sequences of easy interpretation. The algorithm detects events by means of an optimal unsupervised classification. It is not necessary to know the nature of the events and formulate hypotheses regarding their behavior. Then, a set of multichannel electromyographic (*EMG*) signals with events is generated. These *MC* signals are used to test the proposed method.

1. Introduction

Temporal sequence refers to a sequence of happenings or events in a time interval. An event is something that takes place at a time t_E and it reflects some kind of change in a temporal evolution. An important characteristic of temporal sequences is the order in which events take place in time. Other important characteristic is the duration d_E between t_E and the next event.

A temporal sequence is normally represented by a series of nominal symbols from a particular alphabet. Every event is characterized by a symbol. In this way, an event with symbol E is described by 2 elements $(E, t_{E,})$. The 3 elements $(E, t_{E,}\ d_E)$ represent a state.

A temporal sequence S made of u events E^i for $i = 1, \ldots, u$ can be described by a u-tuple of states $S = \langle (E^1, t_E^1, d_E^1), (E^2, t_E^2, d_E^2), \ldots, (E^u, t_E^u, d_E^u) \rangle$ or $S = \langle (E^1, t_E^1), (E^2, t_E^2), \ldots, (E^u, t_E^u) \rangle$. For simplicity, we will use $S = \langle E^1, E^2, \ldots, E^u \rangle$.

Temporal sequences appear in different domains such as engineering, medicine and finance among others (Mitsa, 2010).

The Web usage can also be analyzed by temporal sequences where events are related to the pages accessed by the users as well as the time and duration of visits. Other example could be the sequence of activities (eating, walking...) that people carry out for a specific period of time. Both examples show temporal sequences whose events can be found easily through observation.

In general, the number and temporal locations of events can be unknown within a temporal sequence. This means that events have to be recognized and located inside the sequence. It occurs, for example, in the monitoring of a physical system with Q sensors for a period of time $[T_1, T_2]$. In this case, every sensor represents a channel and simultaneously provides a signal describing part of the whole information. Signals of this type that are generated by multiple sensors are called Multichannel (*MC*) signals. In other words, a *MC* signal refers to a set of signals that show cross-channel similarity or correlation.

A *MC* signal with Q channels is represented by a vector $X(t) = [C_1(t)C_2(t)\ldots C_Q(t)]$ where $C_q(t)$ is the signal of the channel q for $q = 1,2\ldots Q$.

A *MC* signal $X(t)$ with events can also be described, in compact form, as a temporal sequence S.

In this paper, we are interested in classifying a set of different *MC* signals in which not only the number of events is variable but also their start time t_E and duration d_E between events are unknown and they can change randomly. This kind of classification is important in domains such as Bioengineering, Volcanology, Geophysics, Nuclear Fusion, etc. where *MC* signals are present. The classification criterion used in this paper establishes that two *MC* signals belong to the same group if they have the same behavior

* Corresponding author. Tel.: +34 913987192.
E-mail addresses: a.r.m.g@outlook.fr (A. Mur), raquel@dia.uned.es (R. Dormido), jesus.vega@ciemat.es (J. Vega), sebas@dia.uned.es (S. Dormido-Canto), nduro@dia.uned.es (N. Duro).

http://dx.doi.org/10.1016/j.eswa.2016.01.014
0957-4174/

for a period of time $[T_1, T_2]$. Two *MC* signals have the same behavior if they present the same events in the same order. The start time t_E of each event and the duration d_E between events may be different.

For example, if for the interval [12:00 p.m., 14:00 p.m.] a person H_1 walks 1 h, cooks for 20 min, eats for 20 min, sleeps 20 min, other person H_2 sleeps 30 min, cooks for 30 min, walks 30 min, eats for 30 min other person H_3 walks 30 min, cooks for 15 min, eats for 25 min, sleeps 50 min, etc. In this example and according to the criterion mentioned above, only H_1 and H_3 have the same behavior.

Classification techniques are one of the most important domains in data mining. There are two types of classification procedures: supervised classification and unsupervised classification (Mitsa, 2010).

An unsupervised method cannot classify a set of *MC* signals with unknown events (using the same temporal interval), due to the random nature of t_E and d_E. This method selects a feature vector (a η-dimensional vector of numbers) and the *MC* signals with a similar feature vector are classified into the same group. But, when t_E and d_E change randomly, in general, the feature vectors found for the *MC* signals with the same behavior are significantly different and consequently these *MC* signals will not be classified together.

In *MC* signals with unknown events, we do not know the number, order, start time, duration and name of the events. As the events are not known, any classification method cannot address the problem in a supervised way since it is necessary a prior knowledge of the events characteristics. Furthermore, a supervised method using known events is not able to detect when new events appear and they have not been considered a priori.

In this paper, we propose a new unsupervised method able to classify a set of *MC* signals with unknown events. To this end, first events are detected, then *MC* signals are transformed into a set of n sequences and finally the classification is performed.

To validate the proposed method a set of *EMG MC* signals with known events is used. The method detects events in these *EMG MC* signals by considering they are unknown. The good results obtained allow us to extend the use of this method to other type of *MC* signals with unknown events.

References such as Antunes and Oliveira (2001), Mitsa (2010), Leela Sandhya Rani, Naga Deepthi, and Rama Devi (2013), and Shahnawaz, Ranjan, and Danish (2011) provide a survey of the most significant techniques to deal with temporal sequences, in particular temporal sequences classification.

Xing, Pei, and Keogh (2010) present a review about sequence classification: feature based methods (Aggarwal, 2002; Ji, Bailey, & Dong, 2007; Littlestone, 1988; Ye & Keogh, 2009), sequence distance based methods (Kaján et al., 2006; Needleman & Wunsch, 1970; Xi, Keogh, Shelton, Wei, & Ratanamahatana, 2006), model based methods (Cheng, Carbonell, & Klein-Seetharaman, 2005; Durbin, Eddy, Krogh, & Mitchison, 1998; Srivastava, Desai, Nandi, & Lynn, 2007; Yakhnenko, Silvescu, & Honavar, 2005), early classification on sequences (Diez, González, & Boström, 2001), (Xing, Pei, Dong, & Philip, 2008) and semi-supervised learning on sequences (Weston et al., 2005).

Most of the work on event detection focuses on the change detection in data streams. It is based on certain windows to define and study changes in distributions (Muthukrishnan, van den Berg, & Yihua, 2007). Cormode and Muthukrishnan (2004), Gehrke, Ganti, Gehrke, and Ramakrishnan (1999), Kifer, Ben-David, and Gehrke (2004) and Krishnamurthy, Sen, Zhang, and Chen (2003) use two windows with a fixed size and estimate the change of distribution between them. Bartlett, Ben-David, and Kulkarni (2000), Gama, Medas, Castillo, and Rodrigues (2004), Widmer and Kubat (1996) use windows of varied sized but this method is computationally expensive. Ho (2005) proposes a martingale framework for change detection where an adaptive window is used. Bifet, Gavalda, & Gavaldà (2007) use exponential histograms to detect change of various scales. Muthukrishnan et al. (2007) detect changes without a window specification. It provides a new sequential change detection algorithm that improves the use of the Sequential Probability Ratio Test (Wald, 2004) to decide a change. Hidden Markov Models (*HMM*) (Dias & Ramos, 2014; Li, Fang, & Huang, 2015; Rabiner, 1989) are capable of modeling sudden changes in a signal as well as segment the signal into informative states. They can also classify signals in a supervised and unsupervised way. The main drawback of *HMM* is that they are models and consequently their accuracy and capacity of being generic is affected.

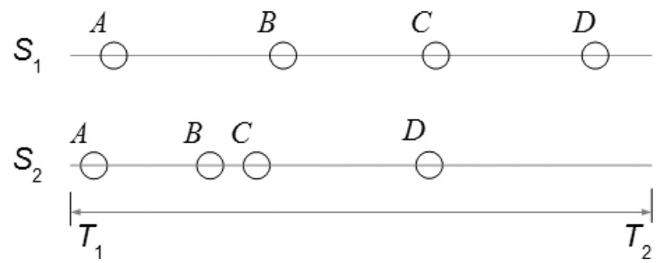

Fig. 1. Two temporal sequences with 4 events and the same behavior.

In our work, event detection and classification are related. Unsupervised classifications are used to detect and characterize events on a set of *MC* signals and subsequently, the events characterization allows classifying the *MC* signals.

In Section 2, the problem of classification of *MC* signals in which there are events whose start time t_E and duration d_E can change randomly is analyzed. In Section 3, we present the new method to detect events and classify the *MC* signals. In Section 4, a set of *EMG MC* signals made of events with random t_E and d_E is generated. Then, the proposed method is tested using these signals. Finally in Sections 5 and 6, the discussion and the conclusions of the paper are respectively presented.

2. Analysis of the classification of multichannel signals with events

A classic method of unsupervised classification (*CC*) for *MC* signals is built on the following steps: (1) the selection of a working window with significant duration, (2) the selection of a feature vector for each *MC* signal on that window, and (3) the application of an unsupervised classification method to the feature vectors chosen.

In this section the problem of classification of temporal sequences with unknown events is analyzed. As a particular case this analysis is valid for *MC* signals. The classification criterion of the *MC* signals mentioned above consists of grouping *MC* signals that show the same behavior in an interval $[T_1, T_2]$. For example, Fig. 1 shows two temporal sequences S_1 and S_2 with four events *A*, *B*, *C* and *D*. Both sequences have the same behavior (because they present the same events in the same order) and consequently, S_1 and S_2 are classified into the same group.

In Fig. 2, there are four temporal sequences S_1, S_2, S_3, S_4 with four events *A*, *B*, *C*, *D* in an interval $[T_1, T_2]$. According to the defined criterion, the temporal sequences are classified into two groups $G_1 = \{S_1, S_3\}$ and $G_2 = \{S_2, S_4\}$. The duration and the start time of the events of each group is the same. And therefore, a *CC* applied to the window $[T_1, T_2]$ would coincide with the result of our criterion, since the feature vectors are similar in each group.

A different situation is shown in Fig. 3. In this case, there are four temporal sequences S_1, S_2, S_3, S_4 with four events *A*, *B*, *C*, *D*