Biomedical Signal Processing and Control 10 (2014) 50–57

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Using mutual information in supervised temporal event detection: Application to cough detection

Thomas Drugman*

TCTS Lab, University of Mons, 31 Boulevard Dolez, 7000 Mons, Belgium

ARTICLE INFO

Article history:
Received 10 July 2013
Received in revised form
12 December 2013
Accepted 2 January 2014
Available online 29 January 2014

Keywords:
Event detection
Biomedical engineering
Cough detection
Mutual information
Supervised learning

ABSTRACT

A large number of biomedical and surveillance applications target at identifying specific events from sensor recordings. These events can be defined as rare and relevant occurrences with a limited duration. When possible, human annotation is available and developed techniques generally adopt the standard recognition approach in which a statistical model is built for the event and non-event classes. However, the goal is not to detect the event in its complete length precisely, but rather to identify the presence of an event, which leads to an inconsistency in the standard framework. This paper proposes an approach in which labels and features are modified so that they are suited for time event detection. The technique consists of an iterative process made of two steps: finding the most discriminant segment inside each event, and synchronizing features. Both steps are performed using a mutual information-based criterion. Experiments are conducted in the context of audio-based automatic cough detection. Results show that the proposed method enhances the process of feature selection, and significantly increases the event detection capabilities compared to the baseline, providing an absolute reduction of the revised event error rate between 4 and 8%. Thanks to these improvements, the audio-only cough detection algorithm outperforms a commercial system using 4 sensors, with an absolute gain of 26% in terms of sensitivity, while preserving the same specificity performance.

1. Introduction

This paper addresses the problem of automatic event detection from time series. A proper method for event detection is of interest in various biomedical, surveillance and signal-based applications involving a sensor-based monitoring of any phenomenon. These applications encompass the characterization of molecular events [1], cough detection [2], monitoring of biomedical measures (sleep apnea [3], muscle activity [4], etc.), seismic event detection [5,6], anomaly detection [7], meteorological changes, traffic regulation [8], or event detection in social streams [9]. This paper proposes a new method of event detection based on mutual information in the context of supervised learning. The problem positioning is more precisely presented in Section 1.1. Since the validity of the proposed approach is illustrated in the frame of audio-based cough detection, the background on this issue is given in Section 1.2. The structure of the paper is finally described in Section 1.3.

1.1. Temporal event detection

The detection of events from time series data is a problem which gained interest from the research community [10–12]. In most cases, studies refer to the issue of unsupervised event detection, in which the underlying phenomenon is ill-understood, making human annotation impossible [10,13]. In such a context, the goal is to identify the time points at which the system behavior change occurs. This is referred to as the *change-point detection* problem [10]. This is typically achieved by considering probability distributions from data in the past and present intervals, and by inspecting whether these two distributions are significantly different [12]. Semi-supervised learning has also been addressed in [14] for the detection of rare and unexpected events. This method can be applied when collecting a sufficient amount of labeled training data for supervised learning is practically infeasible (e.g. because manually annotating such a large corpus would be too time-consuming, and consequently too expensive).

On the other hand, there is a large number of applications for which human annotation is available and which target at identifying specific events from sensor recordings [2,5,8]. Events can then be defined as rare and relevant occurrences, generally with a limited duration. For such an issue, vectors of features characterizing the signal are extracted at a constant sampling rate. More precisely, the signal is windowed in so-called *frames* where

* Tel.: +32 65 374749.
E-mail address: thomas.drugman@umons.ac.be

1746-8094/$ – see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.bspc.2014.01.001

a short-term analysis is performed [15]. For each frame, a set of characteristics (also called features, measurements or attributes) are extracted. The whole signal is analyzed by shifting the frames by a constant delay, consecutive frames possibly overlapping [15]. Based on these sequences of frames, developed techniques generally adopt the standard approach in which a statistical model is built for each class to be identified (presence or not of an event) [16]. This approach nonetheless suffers from a main drawback: models built at learning stage are based on the whole event duration. However, the goal is not to detect the event in its complete length precisely, but rather to identify the presence of an event, i.e. to detect an event trigger.

This induces a dramatic change from the typical training formalism: instead of building models so as to minimize the error rate at the *frame level*, the supervised learning has to focus on the detection ability at the *event level*. An extreme case to illustrate this concept would be a classification system which identifies correctly only one single frame among the several contained in each event. In the conventional framework, this would be characterized by low performance since the majority of the frames contained in the event are missed, although it leads to a perfect discrimination at the event level since the event has been properly detected.

In parallel, measures derived from the Information Theory [17] have been extensively used in machine learning. Among others, the usefulness of mutual information (MI) for selecting the most relevant features in a given classification task has been proven [18]. This efficiency is nonetheless also impaired if the traditional formalism aiming at detecting events in their whole duration is considered. Indeed, the relevance of a feature at the frame level does not necessarily imply its relevance at the event level, and vice versa.

The goal of this paper is precisely to investigate how MI can be used to alleviate the aforementioned drawbacks by localizing the relevant regions of interest in each event and by synchronizing features. Some concepts of the proposed approach and all our experimental results will be illustrated in the context of a particular application: audio-based cough detection.

1.2. Automatic cough detection

Cough is the commonest reason for which patients seek medical advice to the general practitioner (around 20% of consultations for children below 4 years old), the pediatrician and the pneumologist (for whom chronic cough represents one-third of consultations). The impact of cough, notably chronic coughing, on life quality can be important [19].

In order to evaluate the cough severity, a subjective assessment is possible by making use of cough diaries, quality-of-life questionnaires or relying on a visual analog scale [20]. However, it has been shown that the subjective perception of cough is only slightly correlated with objective measurements of its severity [21]. Medical literature on this topic therefore underlines the lack of a tool allowing the automatic, objective and reliable quantification of this symptom [19]. This latter step is notably required prior to any correct evaluation of possible treatments.

Some approaches have been recently proposed to address the automatic detection of cough [2]. These systems generally couple various sensors to the audio signal [2]: accelerometer, chest impedance belt, contact microphone, ECG, respiratory inductance plethysmography, etc. Although reported results are encouraging, there is currently neither standardized methods nor adequately validated, commercially available and clinically acceptable cough monitors [2,19]. Besides, following the patient in ambulatory and 24-h-long conditions (while preserving his daily habits) remains an open problem. As a result, cough quantification in the majority of hospitals is still nowadays performed by a tedious task of manual

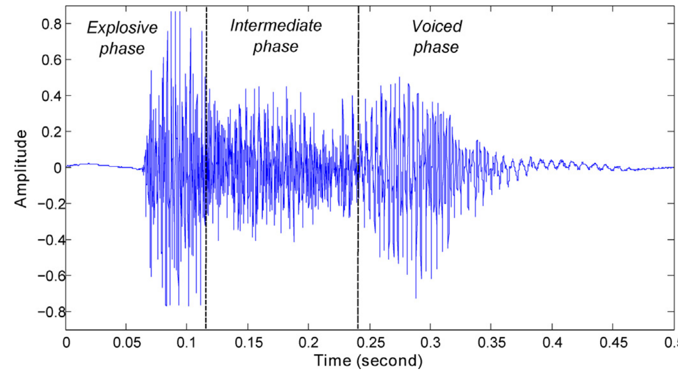

Fig. 1. Waveform of a typical cough sound with three phases.

counting from audio recordings, or for validation by comparison using simultaneous video recordings.

For respiratory physiologists, cough is three-phase expulsive motor act characterized by an inspiratory effort, followed by a forced expiratory effort against a closed glottis and then by opening of the glottis and rapid expiratory airflow [19]. As shown in Fig. 1, the acoustics of the cough sound is manifested by three phases, where the last one is optional [22]: an explosive phase, an intermediate period whose characteristics are similar to a forced expiration, and a voiced phase. At this point, it can then be understood that even for the detection of short events like cough: (i) it might not make sense to try to detect the cough event in its complete duration, (ii) as the signal properties vary across the duration of an event, the segments where features are particularly discriminative may not coincide. For example, some features might be relevant for detecting the explosive phase, while others would characterize the voiced phase. This might be particularly true when features arise from different sensors which might not be synchronous. This paper aims at addressing both of these issues.

1.3. Structure of the paper

This paper is structured as follows. Section 2 describes the proposed approach based on information localization inside events, and feature synchronization. The experimental protocol is detailed in Section 3. Results of our evaluation are reported in Section 4 and the paper is concluded in Section 5.

2. Proposed approach

The general workflow of the proposed approach is presented in Fig. 2. The method starts with a sequence of feature vectors and with the initial event labels (resulting from the manual annotation). The algorithm consists of an iterative process aiming at localizing the relevant information inside the events, and at synchronizing features with each other and with labels. The motivation behind these steps is the following. First of all, the relevant segments of the events, i.e. the portions of events which are the most distinguishable from other classes, are only a partial component of the whole event duration and have to be located. Secondly, features extracted from the time signal may characterize different aspects of this latter signal, which may occur at different instants. Besides, in some applications, features might even arise from various sensors, which strengthens this issue. For these reasons, features have to be synchronized such that their relevant segments emerge at the same time, which is expected to enhance the event discrimination capabilities of the classifier.