ARTICLE IN PRESS
Pervasive and Mobile Computing ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

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

Fast track article

Probabilistic time-series segmentation

Haik Kalantarian *, Majid Sarrafzadeh

Department of Computer Science, University of California, Los Angeles, CA 90024, United States

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Time-series segmentation
Pervasive computing
Signal processing
Machine learning

ABSTRACT

Among the major challenges in the realization of practical health monitoring systems is the identification of short-duration events from larger signals. Time-series segmentation refers to the challenge of subdividing a continuous stream of data into discrete windows, which are individually processed using statistical classifiers to recognize various activities or events. In this paper, we propose a probabilistic algorithm for segmenting time-series signals, in which window boundaries are dynamically adjusted when the probability of correct classification is low. Our proposed scheme is benchmarked using an audio-based nutrition-monitoring case-study. Our evaluation shows that the algorithm improves the number of correctly classified instances from a baseline of 75%–94% using the RandomForest classifier.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

This paper addresses the issue of efficient *time-series segmentation and classification*: important topics in real-time embedded systems such as those used in wearable health and fitness monitoring devices which collect and process continuous sensor data. To understand why time-series segmentation is important requires a basic understanding of sensor systems and statistical classifiers. In this section, we begin with some of these preliminaries.

1.1. Sensing

Real-time wearable sensor systems have become very popular in recent years, and address a variety of health needs ranging from fitness, health monitoring, and object tracking. Such systems can range from a simple activity monitor such as the Jawbone and Misfit [1,2] to more complex examples such as wearable diet-monitoring devices [3,4]. These diet monitors, which are the motivating use case of our work, characterize eating habits from audio data through the analysis of chew and swallow sounds acquired from a wearable microphone.

Regardless of their specific function, wearable devices typically acquire signals from sensors such as accelerometers, gyroscopes, and microphones, as individuals go about their normal daily activities. Subsequently, various algorithms are used to identify actions of interest from the continuous stream of real-time data.

1.2. Classification

Various techniques have been proposed to identify activities of interest from time-series sensor data. One example of such a system was proposed by Alshurafa et al. in [5], in which the authors extract statistical features from sensor data, and

* Corresponding author.
E-mail address: kalantarian@cs.ucla.edu (H. Kalantarian).

http://dx.doi.org/10.1016/j.pmcj.2017.03.005
1574-1192/© 2017 Elsevier B.V. All rights reserved.

Please cite this article in press as: H. Kalantarian, M. Sarrafzadeh, Probabilistic time-series segmentation, Pervasive and Mobile Computing (2017), http://dx.doi.org/10.1016/j.pmcj.2017.03.005.

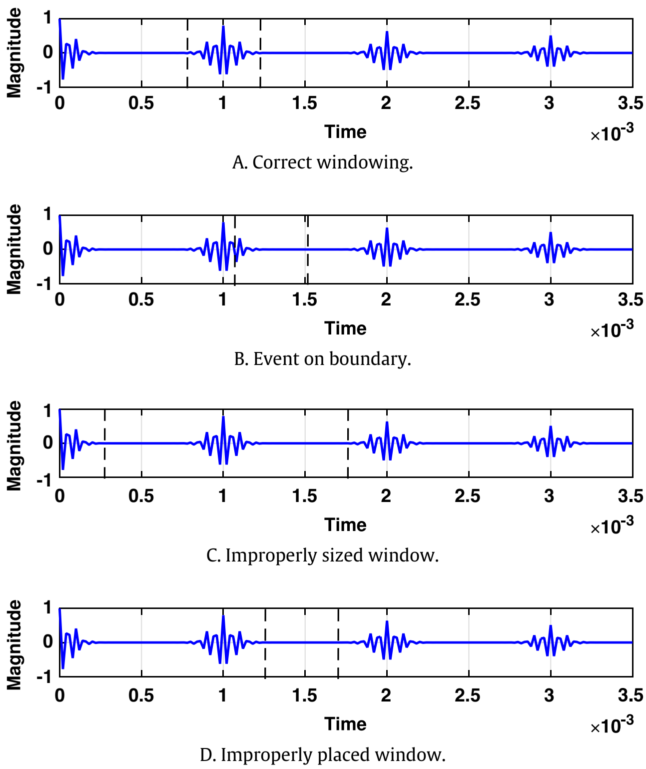

Fig. 1. This figure shows four potential problems when attempting to identify an action from a large discrete signal using a windowing approach. The boundaries of the window are represented by the dashed lines.

determine the activity being performed using machine learning tools. However, in the case of a 24 h stream of data it is not practical to assign a single class label to the whole dataset. Expectedly, the subject may be running for one hour, walking for one hour, and sitting for the rest of the day. Therefore, prior to classification, the data must be *segmented* into smaller windows that are each assigned a separate class label.

1.3. Segmentation

The efficient segmentation of time-series data requires that we select correct start indices and window sizes. Fig. 1 shows various challenges associated with windowing a signal. In this figure, the vertical dashed lines represent the boundaries of our window, and the perturbations in the otherwise constant signal are the events of interest.

In Fig. 1A, a correct windowing approach is shown; the boundaries of the window are selected such that the event of interest is centered, with minimal margins on both sides. In Fig. 1B, the window boundaries are selected such that the event of interest is bisected. Thus, there is no single window which holistically contains the representative features of the event that we wish to detect. Fig. 1C, the window is improperly sized. This may be problematic because the distinguishing features of the event in question may be averaged together with other irrelevant properties that are associated with unrelated actions or events. Fig. 1D shows another case in which an empty portion of the signal is windowed.

We can improve our ability to segment a signal with the empirical observation that a higher degree of classification confidence is typically associated with correct windowing. Intuitively, this is because our notion of an ‘ideal’ window segment is based on our training data, which is generally manually annotated and postprocessed. That is, a classification model that reports 99% confidence that a window is associated with *Class A* suggests a high likelihood that we have selected boundaries that best preserve the distinguishing attributes of the event in question. Otherwise, the window size or indices may require adjustment.

1.4. Classification confidence

Most classifiers assign distinct class labels based on a particular set of input features. However, for our purposes we are interested in knowing the *probability* that a classification is correct, rather than the class itself. The issue of deriving

Download English Version:

https://daneshyari.com/en/article/4957381

Download Persian Version:

https://daneshyari.com/article/4957381

Daneshyari.com