Pattern Recognition Letters 32 (2011) 1659–1668

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Improving the classification accuracy of streaming data using SAX similarity features

Pekka Siirtola *, Heli Koskimäki, Ville Huikari, Perttu Laurinen, Juha Röning

Computer Science and Engineering Laboratory, P.O. BOX 4500, FI-90014, University of Oulu, Finland

ARTICLE INFO

Article history:
Received 16 April 2010
Available online 3 July 2011
Communicated by R.C. Guido

Keywords:
Activity recognition
Classification
Symbolic dynamics
SAX

ABSTRACT

The classification accuracy of time series is highly dependent on the quality of used features. In this study, features of new type, called SAX (Symbolic Aggregate approXimation) similarity features, are presented. SAX similarity features are a combination of the traditional statistical number-based and the template-based classification. SAX similarity features are obtained from the data of the time window by first transforming the time series into a discrete presentation using SAX. Then the similarity between this SAX presentation and predefined SAX templates are calculated, and these similarity values are considered as SAX similarity features. The functioning of these features was tested using five different activity data sets collected using wearable inertial sensors and five different classifiers. The results show that the recognition rates calculated using SAX similarity features together with traditional features are much better than those obtained employing traditional features only. In 20 tested cases out of 23, the improvement is statistically significant according to the paired *t*-test.

1. Introduction

In this paper, we present a novel method for the purpose of calculating features for real-time classification of streaming activity data using a sliding window-technique (Babcock et al., 2002). In this technique, the data are divided into certain fixed-length time windows, from which features are extracted. Classification is done on the basis of these features using a machine learning algorithm. Normally, the extracted features are basic features, such as time or frequency domain features, statistical numbers, or correlations. In this study, the features represent the similarity between the SAX presentation of the time window, which is a symbolic presentation of the time series, and predefined SAX templates. Previously, classification was done either using basic features or templates-based recognition, but the method presented here combines these techniques in a novel way.

The contributions of our paper are as follows:

- We propose features of a new type, called SAX similarity features.
- We use five different data sets and classifiers to show that a combination of SAX similarity features and basic features improves the classification accuracy; and
- We show that this improvement is statistically significant according to paired *t*-test.

SAX similarity features are calculated in three steps. First, the time window studied is divided into n equal-sized sub-windows, and from each n sub-window, a statistical number is extracted. This way, the length of the time window is compressed to the size of n. Secondly, the values obtained are transformed into a discrete representation to get the data into SAX form (Lin et al., 2007). Finally, the SAX similarity feature is extracted by calculating the similarity between the obtained SAX series and the predefined SAX template. This SAX template must also consist of n different symbols; the similarity can be calculated using a string matching measure. In theory, with this method, it is possible to calculate an infinite number of features by using different values of n and mapping the time series to a discrete representation with different vocabularies.

The accuracy and efficiency of SAX similarity features are shown with five different activity data sets and five different classifiers. The data sets were collected using wearable inertial sensors attached to subject's wrist or both wrists and the data sets included from four to eight different human actions and activities. Using these data sets, the functionality of SAX similarity features is tested and the statistical significancy of the results is shown using the paired *t*-test.

The article contains the following sections: Section 2 presents related work. Section 3 introduces SAX similarity features and methods for calculating them. Section 4 evaluates the performance

* Corresponding author. Tel.: +358 8 553 2994; fax: +358 8 553 2612.
E-mail addresses: pesiirto@ee.oulu.fi (P. Siirtola), hejunno@ee.oulu.fi (H. Koskimäki), villehui@ee.oulu.fi (V. Huikari), perttu@ee.oulu.fi (P. Laurinen), jjr@ee.oulu.fi (J. Röning).

0167-8655/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2011.06.025

and accuracy of the proposed method with five different data sets and classifiers. Finally, the discussion and conclusions can be found in Section 5.

2. Related work

Symbolic time series analysis (Lind and Marcus, 1995) extracts relevant information from signal to generate symbol sequences (Rajagopalan and Ray, 2006). It can for example be used to model motion; therefore, symbolic time series analysis can be applied to detect anomalities (Chin et al., 2005), for instance. In this article, it is employed to model human motion measured using inertial sensors. The method we used for the transformation is called SAX (Symbolic Aggregate approXimation) (Lin et al., 2007). SAX is a symbolic representation of a time series and it has been used in several problems applied in many different fields of science such as medicine (Keogh et al., 2006), gesture recognition (Stiefmeier et al., 2007), genetics (Tseng et al., 2007), behavior recognition (Hunter and Colley, 2007), robotics (Murakami et al., 2004) and economics (Lkhagva et al., 2006).

In most of the studies, SAX is employed to reduce computational cost when finding motifs (Xi et al., 2007), which are templates that can be used to find similar time series or images (Silva et al., 2007; Xi et al., 2007). Motifs and other template-based classification methods provide a good method for classifying instances and finding patterns. However, they are not the best method for classifying streaming human activity data, as patterns in that kind of data can vary over time, and especially between persons, unlike patterns of protein data (see Minnen et al., 2006; Siirtola et al., 2009b). Instead, other classification methods such as *k*NN (Pirttikangas et al., 2006; Van Laerhoven et al., 2008), SVM (Ravi et al., 2005; Suutala et al., 2007), LDA and QDA (Boughorbel, 2010; Ward et al., 2005), classification trees (Bao and Intille, 2004; Ermes et al., 2008; Siirtola et al., 2009a), and Naive Bayes (Tapia et al., 2004; Yang et al., 2010) have been used to classify human activity recognition data.

In this study, the experiments were done using five different human activity data sets. Previous human activity recognition studies can be divided into two categories: studies where recognition is done using data from sensors attached to only one place (Ravi et al., 2005; Siirtola et al., 2009a; Van Laerhoven and Cakmakci, 2000) and studies where sensors are located in several places (Bao and Intille, 2004; Ermes et al., 2008; Krishnan et al., 2009; Yanga et al., 2008). Both procedures have their own advantages and disadvantages; because more sensors produce more information about movement, naturally, using several sensors in several places provides more accurate results than only one sensor only. On the other hand, if the aim is to produce a commercial product for the masses, of course, as few sensors as possible should be used: people will not use the product unless the sensors are easy to wear, not time-consuming, and do not disturb the user. The purpose of this study was to present methods and results that can be used in real life, consequently, only one or two sensors were employed.

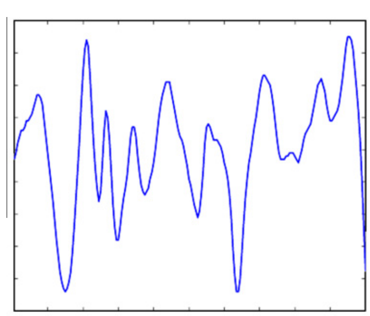
Normally, recognition of activities consists of recognizing 2–20 specific activities. Sometimes also a *null*-activity, i.e. an activity that includes non-specified actions, is recognized. Tool usage and *null*-activity recognition using inertial sensors were presented in our earlier article (Koskimäki et al., 2009). Using a microphone together with inertial sensors, in its turn, was presented in (Ward et al., 2006). Again, if the purpose is to produce a commercial product, the recognition of the *null*-activity is important because a system that can recognize all the possible activities cannot be developed. Therefore, if the *null*-activity is not recognized, the system classifies incorrectly all the activities it is not trained to recognize. In this study, three out of five data sets include *null*-activity, and the tool recognition data collected using only one wrist-worn sensor are the same that were used in (Koskimäki et al., 2009).

An extensive literature review revealed that SAX was used in this study for feature calculation in a novel way. In addition, the results show that the approach improves classification results in activity recognition data sets.

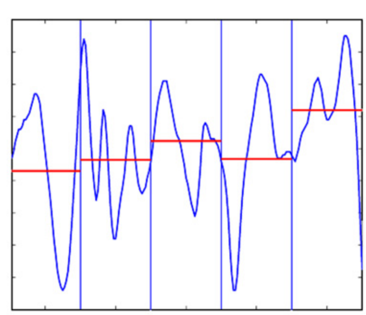
3. Methods

This article presents a novel method for calculating features used for real-time classification of streaming data. The presented method combines template-based recognition and traditional feature based recognition in which the information of time series is summarized as one single value. Here, traditional features are called *basic features*. Features like that are usually simple and fast to calculate; therefore, they are useful in real-time applications. In this study, the classification accuracy of the presented features is compared to that of using basic features and how accurately activities can be recognized with both feature sets, basic features and presented features. The basic features employed in this study were:

- Simple statistical values: mean, standard deviation, median, quartiles, minimum, maximum.
- Frequency domain features: sums of smaller sequences of Fourier-transformed signals, number and ratio of zero crossings.
- Correlation features between different signals.



(a) Time series S from which the features are calculated.



(b) Time series S is compressed so that it can be represented as a vector containing five values.

Fig. 1. Compressing time window so that it can be represented as a vector containing n statistical values.