Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 50-1 (2017) 15865–15870

A multivariate symbolic approach to activity recognition for wearable applications

M. Terzi * A. Cenedese ** G.A. Susto **

* Human Inspired Technology Center, University of Padova
** Department of Information Engineering, University of Padova

Abstract: With the aim of monitoring human activities (in critical tasks as well as in leisure and sport activities), wearable devices provide enhanced usability and seamless human experience with respect to other portable devices (e.g. smartphones). At the same time, though, wearable devices are more resource-constrained in terms of computational capability and memory, which calls for the design of algorithmic solutions that explicitly take into account these issues. In this paper, a symbolic approach for activity recognition with wearable devices is presented: the Symbolic Aggregate approXimation technique is here extended to multi-dimensional time series, in order to capture the mutual information of different dimensions. Moreover, a novel approach to identify gestures within activities is here presented. The performance of the proposed methodology is tested on the two heterogeneous datasets related to cross-country skiing and daily activities.

Keywords: Activity Recognition, Machine Learning, Time Series Learning, Wearable Devices

1. INTRODUCTION

Activity Recognition (AR) is a prominent research area with applications to home automation (Belgioioso et al., 2014), gaming (Gowing et al., 2014), sport (Cenedese et al., 2016) and health care (Clifton et al., 2013). In particular, the rapid growth of IMUs (Inertial-Measurement Units) has allowed, in recent years, the development of compact sensor-equipped devices (e.g. smartwatches and smartphones), which lead efficient monitoring of human activities to be feasible and to have a strong impact on the quality of life (Clifton et al., 2013).

On the other hand, wearable devices present some limitations in terms of computational capability and memory, which force the algorithm design to be at the same time efficient and simple. Moreover, decision algorithms need to be portable, i.e., classification tasks have to be taken at a device-level (Cenedese et al., 2015).

It is important in this context to differentiate between *activities* and *gestures*. Gestures (also called in the following *atomic gestures*) are here considered as basis movements that compose an activity that is conversely completely characterized by one or more atomic gestures: for example, in swimming, a stroke (gesture) completely characterizes the style (activity).

Due to the vastness of application scenarios, it is helpful to categorize the AR problems into three main types:

- *continuous-repetitive* - activities that are continuous and composed by repeated gestures with a periodic behaviour within the same activity type;
- *continuous-spot* - continuous activities with non-repetitive gestures;
- *isolated* - activities composed by isolated gestures.

This work is focused on the *continuous-repetitive* type (Morris et al., 2014), that is typical of sports (e.g. rowing and swimming) and health monitoring applications.

AR problems are usually solved by means of Machine Learning (ML) approaches, however, the aforementioned restrictions on computation and memory capacity cause great limitations in choosing the ML algorithms to be employed. For instance, in (Cenedese et al., 2016), Relevance Vector Machines are chosen over popular Support Vector Machines in order to meet the parsimony constraint required by wearables in the ML algorithm complexity.

Within this context the above restrictions can otherwise be handled by adopting symbolic representation techniques (Rajagopalan and Ray, 2006) in the treatment of IMU-generated time series data; with symbolic approaches, time-series are mapped into *strings*, which implies dimensional and numerosity reduction. Moreover, symbolic representations allow avoiding one pre-processing phase, called *Feature Extraction*, which is common to AR solution and often critical in the selection of parameters to be retained in the models.

With these premises, the contributions of this paper are:

(1) a symbolic approach based on Symbolic Aggregate ApproXimation (SAX) (Lin et al., 2003) for AR. SAX is a popular symbolic approach intended for univariate time-series: since in many AR problems multiple IMU-generated time-series are available, we extend here the approach to multi-dimensional time series in order to exploit mutual information from multiple axes; the results of our experiments confirm that this multi-dimensional extension leads to superior accuracy w.r.t. univariate approaches.

Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2017.08.2333

(2) a procedure to extract atomic gestures and a classification model for Gesture Recognition (GR) is built directly on gestures; more interestingly, a model that is *invariant* to duration and amplitude warpings is depicted. Then, we perform AR starting from GR classification results, through a window-based approach;
(3) an *Event Identification* (EI) procedure is designed in order to detect time windows where activities to be identified are not performed; these time windows are labelled as *all the other movements* (AOM).

The remainder of the paper is organized as follows: Section 2 is dedicated to discuss related works, while in Section 3 univariate symbolic classification problem is illustrated; Section 4 discusses the multivariate extension of the SAX approach while the gesture extraction phase is depicted in Section 5; in Section 6 the experimental results are shown, while Section 7 is dedicated to final remarks.

2. AR PROCEDURES

As stated above, AR/GR problems are usually tackled by means of ML approaches (Morris et al., 2014); more precisely AR/GR problems are generally *classification* ones: the activity or gesture in exam has to be associated with one of the a-priori defined K possible classes of activities/gestures $\mathcal{C} = \{c_i\}_{i=1}^{K}$.

The main challenge in applying ML algorithms in AR problems is to translate the informative content contained in the IMU-generated time series into a static format that can be handled by ML classifiers (Susto et al., 2016; Cenedese et al., 2016). Typically, this is achieved with a flow chart of operations as in the scheme depicted in Figure 1: the pipeline contains two blocks, window extraction and feature extraction, aiming to translate the informative content in the classical form $X \in \mathbb{R}^{N\times p}$, where N and p are, respectively, the number of observations (time windows in this context) and the number of features extracted from each window.

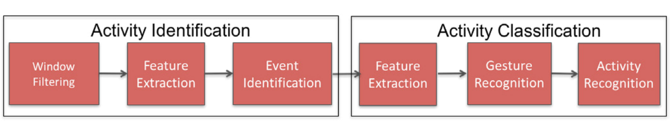

Fig. 1. Scheme of the classical ML approach to AR/GR problems.

The main drawback of using window-based methods is that there is no one-to-one correspondence between gestures and windows; in fact, gestures could have different shapes (at least locally) and warping in time and amplitude domains that dramatically change gestures duration and the window-based statistics (features) $\phi \in \mathbb{R}^p$. Moreover, the aforementioned approach has other two major issues: (i) single gestures are not isolated, which of course make the approach not feasible for GR problems; (ii) the feature extraction phase may be computationally expensive, causing this approach to be almost impracticable for wearable applications.

An alternative procedure suggests to directly compare raw signals with specific distance definitions and with the usage of a distance-based classifier (like Nearest Neighbor, NN (Susto et al., 2015)); classifying directly on the time-series allows to bypass the feature extraction phase, but the window extraction procedure is still required. In this sense, one of the most popular approaches for defining a distance between time series of different length is the Dynamic Time Warping (DTW) (Berndt and Clifford, 1994)): unfortunately, DTW is a dynamic programming technique that hardly meets computational complexity and memory requirements of wearable devices.

In this paper, instead, we adopt a symbolic approach, which is graphically summarized in Figure 2. The pro-

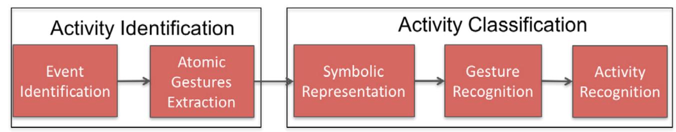

Fig. 2. Scheme of the symbolic approach to AR/GR problems adopted in this work.

posed method is not relying on window-based extraction procedures, but exploits a preliminary *atomic gesture* extraction phase. After the extraction phase, each gesture is then symbolized trough the SAX technique (briefly described in the following); afterwards, a classification model is built over the symbolic representation of a input collection $\mathcal{X} = \{X_i \in \mathbb{R}^{m\times p}\}_{i=1}^{N}$, where m is the dimensionality, p the feature cardinality and N represents the number of gestures. The *Activity Classification* phase will be detailed in Sections 3, 4 then in Section 5 the *Activity Identification* phase is described.

We here recall the SAX technique [1] , which mainly consists on 3 phases:

- signals standardization in order to obtain a zero mean and unit variance signal;
- Piecewise Aggregate Approximation (PAA), described afterwards;
- Symbolic mapping through discretization on amplitude domain.

After normalization, in the PAA phase, a signal $T = t_1, \dots, t_q$ is discretized on time in w frames in order to obtain a vector $\bar{T} = \bar{t}_1, \dots, \bar{t}_w \in \mathbb{R}^w$. Formally, the resulting i-th element $\bar{t}_i$ is defined by:

$$\bar{t}_i = \frac{w}{q} \sum_{j=\frac{q}{w}(i-1)+1}^{\frac{q}{w}i} t_j \tag{1}$$

Then, the SAX representation procedure (i.e. the discretization on amplitude domain) can be summarized as follows. Let a_i denote the i-th element of the alphabet $\mathcal{A}$, with $|\mathcal{A}| = \alpha$. The mapping from the PAA approximation to the correspondent word $\tilde{T} = \tilde{t}_1, \dots, \tilde{t}_w$ of length w is obtained as follow:

$$\tilde{t}_i = a_j \qquad \text{iff} \qquad \beta_{j-1} \le \bar{t}_i < \beta_j, \tag{2}$$

where $\{\beta_j\}_{j=1}^{\alpha-1}$ are break-points tuned to have symbols with equiprobable occurrence. One of the advantages of introducing the *SAX representation*, is that a new distance measure - which is a lower bound of euclidean distance - can be immediately defined. Let T and U be two time-series of the same length q and $\tilde{T} = \tilde{t}_1, \dots, \tilde{t}_q$ and $\tilde{U} =$

[1] We refer the interested reader to (Lin et al., 2003) for a more detailed treatment of the SAX technique.