Neurocomputing 167 (2015) 52–60

Contents lists available at ScienceDirect

Neurocomputing

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

Features and models for human activity recognition

Silvia González [a], Javier Sedano [a], José R. Villar [b], Emilio Corchado [c], Álvaro Herrero [d,*], Bruno Baruque [d]

[a] Instituto Tecnológico de Castilla y León, Burgos, Spain
[b] University of Oviedo, Campus de Viesques s/n 33204 Gijón, Spain
[c] Departamento de Informática y Automática, University of Salamanca, Spain
[d] Department of Civil Engineering, University of Burgos, Spain

ARTICLE INFO

Article history:
Received 19 March 2014
Received in revised form
28 November 2014
Accepted 26 January 2015
Available online 8 May 2015

Keywords:
Human activity recognition
Genetic fuzzy finite state machine
Feature domain reduction
Feature selection
Information correlation coefficient

ABSTRACT

Human Activity Recognition (HAR) is aimed at identifying current subject task performed by a person as a result of analyzing data from wearable sensors. HAR is a very challenging task that has been applied in different areas such as rehabilitation and localization. During the past ten years, plenty of models, number of sensors and sensor placements, and feature transformations have been reported for this task. From this bunch of previous ideas, what seems to be clear is that the very specific applications drive to the selection of the best choices for each case.

Present research is focused on early diagnosis of stroke, what involves reducing the feature space of gathered data and subsequent HAR, among other tasks. In this study, an Information Correlation Coefficient (ICC) analysis was carried out followed by a wrapper Feature Selection (FS) method on the reduced input space. Additionally, a novel HAR method is proposed for this specific problem of stroke early diagnosing, comprising an adaptation of the well-known Genetic Fuzzy Finite State Machine (GFFSM) method.

To the best of the author's knowledge, this is the very first analysis of the feature space concerning all the previously published feature transformations on raw acceleration data. The main contributions of this study are the optimization of the sample rate, selection of the best feature subset, and learning of a suitable HAR method based on GFFSM to be applied to the HAR problem.

1. Introduction

This research aims at developing a solution for the early diagnosis of stroke and the rehabilitation of elder people after a disruptive event: an injury due to a falling, a seizure onset, etc. In this context, only a small subset of activities are to be identified among those that a human being can usually do and hence, the recognition of those activities is simplified. On the other hand, activity recognition devices help to improve the mobility of elder people during rehabilitation, so technology is enhancing the quality of life for both elder and injured people.

Every human being performs different activities during the day and Human Activity Recognition (HAR) targets their identification. Though walking recognition is nowadays a clear-cut task [1], the recognition of other activities is not. It is difficult due to the fact that there are many different activities that a person may perform and some of them could even co-occur at the same time (talking to another person or eating a sandwich while walking, reading and being seated). Additionally, a wide spread of feature transformations and HAR methods have been applied up to now [2–4]. Many techniques of data gathering, including video-images, are being used but tridimensional accelerometers are the data sources for the majority of previous HAR studies.

The main problem to be solved in this research is the early diagnosis of stroke onsets. During such episodes, the upper limbs are the parts of the body that best reflect the differences regarding normal behavior. According to this idea, two triaxial accelerometers are usually placed on the subject's wrists. The hypothesis is that with these sensors we would be able to recognize an onset due to the differences in the movement patterns. However, these movement patterns will depend on the task that the subject is carrying out. Therefore, early diagnosis of stroke also includes HAR. Interesting enough to mention, the activities to be identified also depends on the focused population; generally speaking, the older you are the lower the amount of activities you perform during everyday life. Additionally, the quality of the movement may be rather different depending on the age of people; the younger the faster. Thus, the target population define the different

* Corresponding author.

http://dx.doi.org/10.1016/j.neucom.2015.01.082
0925-2312/

activities that may be excluded for recognition and the age reduces the amount of movement; movement for the elderly is lower than that for younger people while performing the same activities.

A light wearable device might be enough for the data gathering, pre-processing and analyzing provided the number of activities to be recognize is kept small. Previous studies have analyzed the window size for pre-processing the continuous data flow coming from data sources [5]. Once the data is properly pre-processed, they must be analyzed.

To do so, one of the most interesting HAR techniques that can be deployed in embedded devices is the Genetic Fuzzy Finite State Machine (GFFSM) [6], which also handles expert knowledge with high accuracy. To this end, this study is focused on enhancing the HAR by using 3DACC sensors in wearable devices. The aim of this research is two-fold: on the one hand, the selection of the 3DACC transformations is analyzed and a feature subset is chosen; on the other hand, the improvements on the GFFSM model obtained due to the reduced feature subset are explored. The study also makes use of this GFFSM model for HAR, but additionally it tackles several issues identified as relevant in the HAR literature, such as those concerning the diversity of transformations from the acceleration raw data and how to reduce the dimensionality and the different methods for cross validation used so far.

The remainder of the present paper is organized as follows. Next section deals with the challenging task of HAR, including an overview of the input feature domain, as well as an in-depth review of the HAR literature. Section 3 introduces the proposed method and its different stages: a two-step Feature Selection (FS) and a HAR modeling by means of the GFFSM. Subsequently, Section 4 is devoted to evaluate and discuss the experimentation carried out. Finally the main conclusions from the obtained results as well as the future work are drawn in Section 5.

2. A review of human activity recognition

After several years of study, a wide spectrum of features, calculated as transformations from raw acceleration data, have been proposed for HAR. A set of features is chosen for each one of the applied methods according to different criteria. This section describes the main features from those that have been proposed in the literature up to now; afterwards, the previous work on HAR is analyzed and compared.

2.1. From acceleration data to the input feature space

Nowadays, the most common sensor applied in HAR is the triaxial accelerometer. Data gathered from this type of sensor, known as raw data (RD, a_i^x, a_i^y and a_i^z; $a_{i,j \in \{x,y,z\}}$ for the sake of brevity), should be decomposed in the gravity acceleration (G) – that is due to each gravity, g_i^x, g_i^y and g_i^z or $g_{i,j \in x,y,z^-}$ and the BA – which is due to the human movement, b_i^x, b_i^y and b_i^z or $b_{i,j \in x,y,z^-}$. The ability of BA to discriminate among different human gestures is documented in [7]. Nevertheless, the literature includes the use of a wide variety of transformations (the most interesting ones are described below), where w stands for the window size – if needed–, and sub-indexes $i \in \{1, \ldots, N\}$ and $j \in \{x, y, z\}$ stand for the number of the sample and the axis, respectively. It is worth mentioning that all these features computed on each one of the possible signals (RD, BA, and G) would generate a feature space with more than 190 features, whose processing and analysis are very challenging tasks indeed.

The following features have been previously applied to HAR:

1. The *mean, deviation and higher momentum statistics* values for the RD [8] or for the BA [9,7], and the RD *mean absolute deviation* $MAD_j = \frac{1}{w}\sum_{i=1}^{w} |a_{i,j} - m_j|$ [10,8], where m_j is the mean value of $a_{i,j}$.
2. The *Root Mean Square* $RMS_j = \sqrt{\frac{1}{w}\sum_{i=1}^{w} |a_{i,j}^2|}$ [10].
3. The *sum of the absolute values* of the BA [11] $sBA_i = \frac{1}{w}\sum_{t=i}^{i+w}\sum_{j \in \{x,y,z\}} |b_{t,j}|$, the *vibration of the sensor* (Δ) [9] $\Delta_i = \frac{1}{w}\sum_{t=i}^{i+w}\sum_{j \in \{x,y,z\}} a_{t,j}^2 - g_{t,j}^2$ and the *tilt of the body* ($tilt_i = \frac{1}{w}\sum_{t=i}^{i+w} |a_i^y| + |a_i^z|$[6]. The two former transformations were designed to detect whether the sensor registers no movement at all, as fixed to an steady object, while the latter is used whenever the sensor axes match with the body axes.
4. The *Signal Magnitude Area* $SMA = \frac{1}{w} \cdot \sum_{i=1}^{w}(|b_i^x| + |b_i^y| + |b_i^z|)$ [9,12,7] discriminating between gravity acceleration and BA.
5. The *Amount of Movement* $AM_i = \sum_{v=\{x,y,z\}} |\max_{t=i+1}^{i+w}(b_t^v) - \min_{t=i+1}^{i+w}(b_t^v)|$ [6] calculated as the maximum difference between the values of BA within the sliding window.
6. The Delta coefficients for estimating the first order time derivative of each of the G signal components [12]: $\Delta g_t^{\{x,y,z\}} = \sum_{d=-D}^{D} d \cdot g_{t+d}^{\{x,y,z\}} / \sum_{d=-D}^{D} d^2$, where the shift D is parameterized to the algorithms and $g_t^{\{x,y,z\}}$ stands for each of the three axis G components.
7. The Shifted Delta Coefficients (SDC) for estimating the first order time derivative of the BA signal components in the vicinity of the current timestamp [12]: $\Delta b_{t+i \cdot P}^{\{x,y,z\}} = \frac{\sum_{d=-D}^{D} d \cdot b_{t+i \cdot P+d}^{\{x,y,z\}}}{\sum_{d=-D}^{D} d^2}$, where $b_t^{\{x,y,z\}}$ stands for each one of the three axis BA components, N is the number of base features from which they are calculated, D stands for the same as in the delta calculations, P is the distance between samples and K is the number of samples taken.
8. The Average Energy (AE) [13,9,7] calculated as the sum of the squared discrete FFT component magnitudes of the signal in a window of a fixed size. This features allows to discriminate between static and dynamic activities. Although it is calculated for each axis, the aggregation or the average over the three axes is often used [7].
9. The correlation between axes [13] calculated for each pair of axes as the ratio of the covariance and the product of the standard deviations. This feature is useful to discriminate one dimensional activities if the sensor is properly placed. As stated in [7], this feature allows the discrimination between walking and climbing stairs.
10. The Intensity of the Movement (InMo) [14], which is the mean first derivative of the raw acceleration data, $InMo_t^{v \in \{x,y,z\}} = \frac{1}{w}\sum_{i=0}^{w-1} \cdot |a_{t-i}^v - a_{t-i-1}^v| / \Delta x_t$. Δx_t represents the time between samples, which can be ignored if the sampling rate is kept constant. The window size is given by the value of w.
11. The Time Between Peaks (TBP) [8], time in milliseconds between peaks in the sinusoidal waves associated with the frequency response of most activities (for each axis).
12. The Binned Distribution [7,8] as stated by the authors, this measure is used with sliding windows of size w. For each window the range should be calculated as maximum-minimium; then, the range is divided into 10 equal size bins; finally, it is recorded what fraction of the w values falls within each of the bins. This approach is named as Relative Binned Distribution (RBD). In this study, it is proposed the Absolute Binned Distribution (ABD) that is calculated using the lower and upper acceleration values as the range to be divided in bins.

In many of the solutions, sliding windows (with or without shifting) are proposed and the typical window size converges to the samples within a period of 2 s. Features are typically normalized to 0-mean 1-standard deviation and/or scaled to the interval