A novel feature selection technique for improving wearable activity recognition

Oresti Banos, Choong Seon Hong, Sungyong Lee*
Department of Computer Engineering, Kyung Hee University
Email: {oresti,cshong,sylee}@khu.ac.kr

Abstract

Last technological advances in wearable sensors and machine learning are allowing for a new generation of human monitoring techniques, especially devised for the analysis of biomechanics and activity patterns. In this paper, a novel technique to improve the identification of daily physical activity is presented. Taking into account the importance of data featuring and the selection of the most important features for the subsequent pattern recognition stage, a new feature selection methodology based on a filter technique via a couple of two statistical criteria is presented. Satisfactory accuracy rates are achieved by using support vector machines, particularly for preprocessed inertial data from the wrist.

1. Introduction

The analysis of human physical activity is gaining much attention, especially in the health domain [6]. Diverse technologies are used for tracking people conducts, such as cameras, microphones or ambient sensors; however, given the capabilities of the new generation of mobile systems, inertial sensors are predominantly considered.

The recognition process consists of various stages from which feature extraction and selection prove to be of crucial importance [2]. This work presents a new technique for extracting the most relevant features given a feature set to be used in the activity recognition system. One of the most important characteristics of the proposed method is that it does not provide a rank for every individual feature but for every set of features, allowing for the synergic use of several features when considered altogether.

The rest of the paper is organized as follows. Section 2 describes the experimental setup, data preprocessing and featuring mechanism. Next, the proposed rank-based feature-set selection technique is presented. Section 4 shows the evaluation results and final remarks are given in Section 5.

2. Data preprocessing and feature extraction

The initial setup consists of a signal set corresponding to acceleration values measured by a group of sensors located in strategic different parts of the body (hip, wrist, arm, ankle, thigh), for four daily activities (walking, sitting, standing, running) [1].

The acceleration signals are frequently affected by noise and other sort of anomalies, thus preprocessing techniques are typically used. Filtering techniques normally entail loss information, so it is of interest to check their effect on the recognition process. Accordingly, two preprocessing techniques are considered here, respectively, a mean filter (Fig. 1.b) and a band pass filter (Fig. 1.c). The mean filtering is defined to remove the initial offset introduced in the original data acquired and the discontinuities associated to the sensors calibration changes between different monitoring sessions (Fig. 1.a). Band pass filtering also permits to remove the high frequency noise. Considering that a 20 Hz sampling is sufficient to assess habitual daily physical activity [5], an elliptic filter with 0.5Hz and 20 Hz cutoff frequencies is used for the last one.

Once the data has been processed, a parameter set made up of 861 features corresponding to a combination of statistical functions such as mean, kurtosis, mode, variance, etc., and magnitudes obtained from a domain transformation of the original data such as energy spectral density, spectral coherence or wavelet decomposition, among others.

In this stage, a mechanism to determine which features are the most important ones to discriminate among the activities is required.

3. Feature selection based on discrimination and robustness statistical criteria

To obtain a relevant group of variables from a given initial set is not a trivial task due to the huge number of possible feature combinations. In our experimental setup the sample space is represented by n = 861 features, so brute force techniques like 'branch and bound' (O(2n) convergence → $2^{861} \approx 1.5 \times 10^{259}$ possible permutations) or wrapper methods are impractical. In this section, we present an alternative method based on the concepts of discrimination and robustness for a complete set of features.

Let us define the sample range of a class as the set of values included between the maximum and the minimum value (both inclusive) that a feature or variable takes for this class. Given a group of samples (associated to every class) we rank its discriminant capability with respect to that class through the overlapping probability between this class and the others. This is calculated computing the number of samples from the analyzed class which are inside of the sample range defined by the others. For N classes and M samples for each class (let us suppose that this number is independent of the class), we define the overlapping probability of a set of samples as follows:

$$p(k) = \frac{1}{N-1} \sum_{n \neq k}^{N} \frac{m(k,n)}{M} \qquad (1)$$

with m(k,n) being the number of samples from the class k inside the sample range of class n.

We now carry out a thresholding process which allows us to define the feature analyzed as discriminative or not. This overlapping threshold takes values from 0 (the most restrictive, for cases with no overlapping between classes) to 1 (the most relaxed, when every sample from a class is inside the others). In general for a specific feature, if the analyzed class exceeds the threshold, the feature will be considered as no discriminant for this class.

Apart from the discriminant capacity of a feature or a set of features, a second characteristic accounting for the usability of this set of features for diverse sources is considered. E.g., a specific measure taken from the hip accelerometer can be very discriminative to distinguish between the activities walking and standing, but this very measure may not be that reliable when taken from the ankle accelerometer. Thus, the robustness criteria is here defined to categorize for how many sensors the feature is of relevance. In short, discriminant capacity says how useful a motion feature is in general, and robustness is how this depends on where the sensor is.

Combining both criteria we obtain a quality ranking

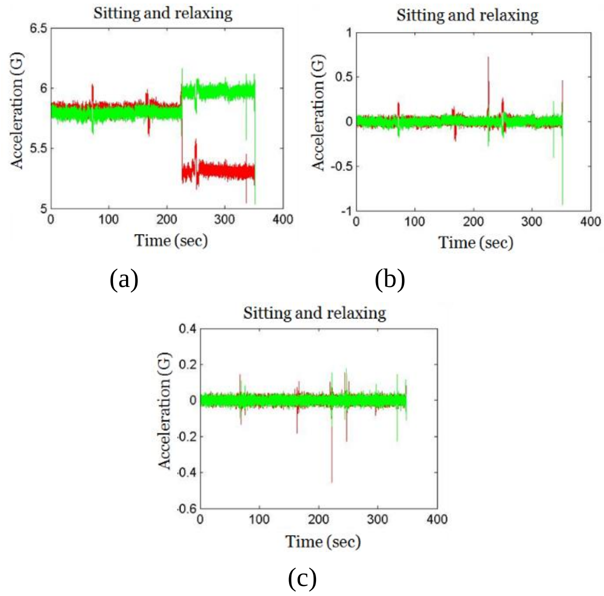

Figure 1. Acceleration signals from the wrist sensor: a) Original data with a 5.8G offset and a discontinuity at second 220 approx., b) Mean filtered data and c) BP filtered data.

procedure capable of grouping features in different stages. For the sake of simplicity, let us suppose a recognition system with 4 classes and 5 sources; features will be classified in groups defining a ranking (see table 1). For instance, features that discriminate 4 classes in every source will be added to group #1 (the best). Group #13 will be completed with features that classify 2 classes (the same) in 3 sources at least. This example is extensible to any classes and sources.

Table 1. Example for 4 classes and 5 sources of quality feature set (ranking) based on discriminant and robustness criterions.

Discriminant	Robustness	Quality group
4	5	#1
	4	#2
	3	#3
	2	#4
	1	#5
3	5	#6
	4	#7
	3	#8
	2	#9
	1	#10
2	5	#11
	4	#12
	3	#13
	2	#14
	1	#15
1	5	#16
	4	#17
	3	#18
	2	#19
	1	#20
0	–	#21

4. Results

Most remarkable features (set #1 and #2 primarily) are geometric mean for amplitude signal, autocorrelation and some wavelets coefficients obtained through a 3-level Daubechies decomposition. For classification, support vector machines [7] are considered, employing a RBF kernel implementation with hyper-parameters γ and C automatically tuned using a grid search. A 10-fold CV method is used for training and testing. Results are showed in Figure 2.

Clearly, the preprocessing is needed to optimize the recognition capability of the system, with particular remarkable results for the band pass (BP) approach. In fact, the mean filtering, that could be reasonably interpreted as a low pass filter (to remove the DC bias or 0Hz component) determines an important improvement with respect to the unprocessed data. Notwithstanding, the BP filter also permits the removal of the frequency components above 20Hz. This demonstrates that there are irregularities and abnormalities in upper frequencies that complicates the discrimination task.

Although good results are obtained in general for the BP filtering approach, we want to note that these results are achieved for each sensor separately, i.e., no information from other sensors is needed to accomplish the recognition task. Considering that accuracy rates above 95% are attained for all the sensors, any placement will work efficiently for the considered activities, being especially interesting the wrist sensor location for its unobtrusive properties.

Although a strict comparison with other studies cannot be made since the data and the number of classes may differ, in [1] an 89% classification accuracy was reached, 93% in [3] and 87% in [4].

5. Conclusions

In this work we have very briefly shown a direct application of ranking selection methods used on daily physical activity automatic recognition. An efficient classification method requires a productive and limited feature set, being necessary a selection process since the initial set is quite huge. We have defined a feature selector based on statistical discrimination and robustness criteria, focused on low computational time and resources, defining a real alternative to other selection processes.

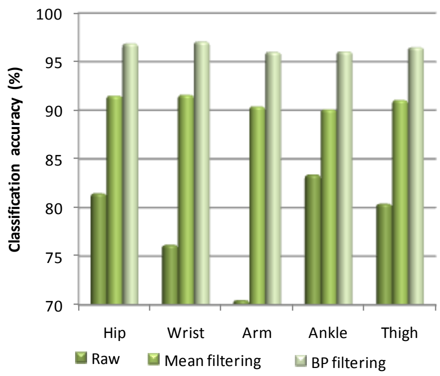

Figure 2. Accuracy rates for the three processing approaches and considered sensor.

The importance of an adequate preprocessing stage has been also showed, demonstrating that singularities and irregularities affect physical activity monitored data. The wrist location stands out for its efficiency and unobtrusively.

Acknowledgements

This work was supported by the Industrial Core Technology Development Program (10049079, Develop of mining core technology exploiting personal big data) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea).

References

[1]Bao, L., Intille, S.S.: Activity Recognition from User-Annotated Acceleration Data. In: Activity Recognition (Pervasive Computing) ». LNCS, vol. 3001, pp. 1--17. Springer, Heidelberg. (2004)

[2]Banos, O.; Damas, M.; Pomares, H.; Prieto, A.; Rojas, I. Daily living activity recognition based on statistical feature quality group selection. Exp. Syst. Appl. 2012, 39, 8013–8021.

[3]Bonomi, A.G., Goris, A.H.C., Yin, B., Westerterp, K. R.: "Detection of Type, Duration, and Intensity of Physical Activity Using an Accelerometer". Medicine & Science in Sports & Exercise. vol. 41, pp. 1770--1777. (2009)

[4]Ermes, M., Pärkka, J., Mantyjarvi, J., Korhonen, I.: "Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions". IEEE Trans. Inf. Technol. Biomed. vol. 12, pp. 20--26. (2008)

[5]Mathie, M.J., Coster, A.C.F., Lovell, N.H., Celler, B.G.: "Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement". In: Physiol. Meas., vol. 25, pp. 1--20. (2004)

[6] Swan, M. "Health 2050: the realization of personalized medicine through crowdsourcing, the quantified self, and the participatory biocitizen," Journal of Personalized Medicine, vol. 2, no. 3, pp. 93–118, 2012.

[7]Vapnik, V. "Statistical Learning Theory", John Wiley & Sons. (1998)