Activity Classification Using Mobile Phone based Motion Sensing and Distributed Computing

Arkaitz ARTETXE[a,b,1] and Andoni BERISTAIN[a,b,c] and Luis KABONGO[a,b]
[a] *Vicomtech-IK4 Research Centre, Mikeletegi Pasealekua 57, 20009 San Sebastian, Spain*
[b] *Biodonostia Health Research Institute, P. Doctor Begiristain s/n, 20014 San Sebastian, Spain*
[c] *Computation Intelligence Group, Basque University (UPV/EHU) P. Manuel Lardizabal 1, 20018 San Sebastian, Spain*

Abstract. In this work we present a system that uses the accelerometer embedded in a mobile phone to perform activity recognition, with the purpose of continuously and pervasively monitoring the users' level of physical activity in their everyday life. Several classification algorithms are analysed and their performance measured, based for 6 different activities, namely walking, running, climbing stairs, descending stairs, sitting and standing. Feature selection has also been explored in order to minimize computational load, which is one of the main concerns given the restrictions of smartphones in terms of processor capabilities and specially battery life.

Keywords. smartphone, accelerometer, activity recognition, classification, machine learning

Introduction

According to the United Nations forecast, the population aged 60 years and over is expected to increase from 20 to more than 30 per cent by the year 2050 in the more developed regions, from 8 to 20 per cent in the less developed regions and from just 5 to 10 per cent in the least developed regions, making ageing of global population an increasingly relevant topic in most government's strategies. Additionally, according to the *Madrid International Plan of Action on Ageing*, as part of the United Nations work on ageing, the promotion of healthy nutrition habits and physical activity are two of the pillars to improve the quality of life of elderly people. Related to this plan, the *Oxford Institute of Ageing* in their *Global Aging Survey (GLAS)* report [15] remarks that suffering from illnesses and disability is one of the main concerns of elderly people. Another risk factor is obesity according to World Health Organization (WHO) report *10 facts on obesity*, where physical inactivity has been pointed as one of the possible causes. These two factors may be partially related to the increase in health care expenditure and therefore there is a need to foster physical activity.

[1] Corresponding Author: Arkaitz Artetxe; E-mail: aartetxe@vicomtech.org

There is an emerging movement called *Quantified Self* (QS), which consists in people recording daily living events that can be classified as inputs (weather, food intake, etc.), states (mood, blood oxygen levels, heart rate, etc.) and performance (mental and physical activities, etc.). The aim is to gain self-awareness and eventually to have an impact of determined daily life actions on people's health, therefore promoting healthy habits and self-care.

One of the catalysts of the QS movement has been the development of many smart wearable devices, miniaturized devices with sensing, computing and communication capabilities that a person can wear to automatically record different parameters. Recently these devices started to be shipped within mobile phones, but most of existing devices are designed to work in cooperation with a smartphone and, are autonomous in most cases. This movement is under the scope of *mHealth*, which brings the possibility to patients to monitor their vital signs or relevant parameters for better prevention, diagnosis, treatment and follow up.

Additionally, in the last years, the use of mobile phones with advanced features, so called smartphones, is becoming more widespread among individuals. Usually these devices incorporate various types of sensors including proximity sensors, GPS sensor, compass and accelerometer sensors to name a few, as well as computational and data communication capabilities.

The work presented in this paper has been driven by all these issues and trends and targets a specific challenge, the use of the triaxial accelerometer embedded in the majority of smartphones to perform activity recognition. The ultimate purpose of this work is to advance in the seamless monitoring of people's level of physical activity during their everyday life.

1. Related Work

Activity recognition based on accelerometers is a research topic that has been extensively studied in recent years. Since early 2000s several studies have shown the capacity of external wearable accelerometer sensors to recognize different physical activities [1][2][3]. The most extensive and thorough work on the subject is that of Bao et al. [1], in which a recognition algorithm capable of recognizing 20 different daily activities using 5 biaxial accelerometers is presented. That paper makes a review of the state of the art and makes stress upon the use of data recorded and annotated by the user in a real-world environment, in contrast to most studies that use data gathered in a controlled environment. This work concludes that the use of multiple accelerometers helps discriminating among activities, although notes that good performance can be achieved with only two accelerometers (placed in the thigh and wrist).

Over the last few years the technical capabilities of smartphones have increased considerably. These devices usually include a built-in triaxial accelerometer, among other sensors. Recently, in conjunction with the popularization of these devices, several experiments have been conducted in order to recognize physical activities using smartphones [4][5][6][7][8][9][10]. This approach differs from previous works mainly in the fact that a single accelerometer is used instead of several. Another major difference is that the position of the mobile phone in the body is not fixed, which leads to new problems. Sun et al. [6] have addressed the issue of the position and orientation variability by training several SVM classifiers (one for each predefined position) and applying the corresponding classifier based on a previous position identification process. Yang [7] tackles the orientation problem by means of computing the horizontal and vertical components in acceleration. A similar approach is presented by Henpraserttae et al. [10] who uses a projection-based technique for device orientation transformation. However, the work of Mizell [22] shows that the vertical component can be inferred by means of the static acceleration that the gravity produces. Our approach follows Mizell's approach in order to effectively recognize physical activities regardless of the orientation of the device inside the pocket.

Regarding the algorithms used for the activity classification task, there are various approaches. For instance, different implementations of decision trees, such as the C4.5 algorithm family have been used, achieving high performance with low complexity [1][5][7][8][12]. Other classification algorithms included in Weka [20], such as k-Nearest Neighbour or Naive Bayes, have also been employed in physical activity recognition [1][7][8][10][12]. Authors like Sun et al. [6] or Krishnan et al. [13] have studied the use of Support Vector Machines for this same classification task, achieving an accuracy of 94.8% and 83.6% respectively. As stated by Lester et al. in [11] and applied by Mannini et al. in [14], Hidden Markov Models may be useful to recognize physical activities based on their capacity to capture the temporal regularities and smoothness of activities.

Unlike other works, ours presents two different classifiers corresponding to two different scenarii: I) an optimal classifier for ideal conditions of computation capacity and II) a classifier for limited computing capacity conditions.

2. Data acquisition

For data acquisition we used a mobile phone (LG Optimus L7) which contains an accelerometer that captures samples consisting of the acceleration on x, y and z axes (in m/s^2) plus the timestamp containing the time at which the sample was taken (in nanoseconds). The sampling frequency at which the accelerometer can operate is variable and depends greatly on the operating system (Android in this case). When setting the sampling rate of the sensor we have observed that despite the fact that the approximate sampling rate can be programmatically specified (Android actually permits choosing among four different rates), the operating system will ultimately decide the actual sampling frequency depending on the background running tasks, in order to adjust processor load.

In order to avoid consequent inconveniences we have defined two different methods. The first method lies in keeping the screen active while the program remains in the foreground (with high priority), preventing the OS from limiting its resources. Another solution is to run the recording applications in a freshly started device with no applications running on the background. Nevertheless, there is no guarantee that the OS will sustain the same sensing frequency when another application demands an intensive use of the smartphone resources.

In order to collect the data we have implemented an Android application with a simple user interface that allows the user to fill in his name and the task or activity he is about to perform. During the activity, the user puts the phone in his pocket and meanwhile, the application records all the data coming from the accelerometer. For our experiments we used the highest sampling frequency allowed by the device, i.e. 96 Hz.

3. Data Processing

Most of the previous work in the field of activity recognition using accelerometers has been done using devices placed in fixed positions of the body [1][2][3]. In these cases the orientation of the device is known and therefore the local coordinate system is also known. However, this is not possible with a mobile phone since the orientation is no longer fixed, each person may carry it in different locations and with different orientations, and therefore data varies significantly. Our experiments show that the same activity carried out by the same person generates very different acceleration patterns depending on the location (e.g. hip, arm, leg) of the device. Moreover, data vary significantly depending on the shape and size of the pocket, whether the trousers are tight or loose etc. In order to restrain mobile device position variability, we made the subjects to keep the mobile phone in their front pockets of their trousers.

For addressing the orientation related sensitivity, we followed the approach by Mizell [22], as stated previously. Following this idea, it is possible to estimate the vertical component of acceleration in world coordinates, including its magnitude and sign as well as its horizontal magnitude.

In order to extract the vertical acceleration attributable to gravity, a vector v is calculated by computing the average absolute acceleration within a reasonable time interval. Let $a_i = (x_i, y_i, z_i)$ be the acceleration vector at a given point of the sampling interval. Projecting a_i onto the normalized v vector, $\hat{v}$, we get the length and sign (positive or negative depending on the direction) of the vertical component. The projection vector p_i is obtained by multiplying the latter by $\hat{v}$. The horizontal component h_i can be calculated by subtracting the projection vector p_i from a_i. In this case we only know that h_i is orthogonal to p_i but we have no way to know its direction. Therefore, the horizontal component is just a magnitude and has no sign.

4. Feature Extraction

As the result of the data processing method described before, the absolute horizontal component and signed vertical components of acceleration are calculated out of the accelerometer data. The set of features will be extracted out of this two dimensional vector of vertical and horizontal components.

When calculating the features we use a window of 512 samples with 50% overlap. Several works [1][12][14] have shown the validity of using a sliding window with 50% overlap. In addition, with a sampling rate of 96 Hz we get a window of approximately 5.33 seconds. This window size is similar to that used in [1] or [14] and has proven to be large enough to capture the most significant patterns of the activities. The window of 512 samples is not accidental, since a window of $2^{\log_2 N}$ allows to compute Fast Fourier Transform (FFT) efficiently.

For each window a set of temporal domain and frequency domain features are calculated. These features are calculated for both vertical and horizontal components. In the time domain we calculate some features such as arithmetic mean (mean), standard deviation (sd) and median absolute deviation (mad). We also compute the zero crossing rate, setting the zero level in the average, what we call the mean crossing rate (mcr). Interquartile range (iqr) and 25^{th}, 50^{th} and 75^{th} percentiles are also calculated.

In the frequency domain we compute the FFT and then we calculate the entropy and energy of both components. The feature set is completed with the correlation between the vertical and horizontal components.

The complete feature set is composed of 21 features, namely: *h_mean*, *v_mean*, *h_sd*, *v_sd*, *h_mcr*, *v_mcr*, *h_mad*, *v_mad*, *h_25percentile*, *v_25percentile*, *h_50percentile*, *v_50percentile*, *h_75percentile*, *v_75percentile*, *h_iqr*, *v_iqr*, *h_energy*, *v_energy*, *h_entropy*, *v_entropy* and *corr_hv*.

5. Classification

In this work several classification algorithms have been evaluated and compared.

Random Forest [16] is a classifier consisting of multiple decision trees trained using randomly selected feature subspaces. **C4.5** [18] is a classification algorithm that generates a decision tree using the concept of information entropy in order to select the attribute in each node. **Naive Bayes** [17] is a classifier based on Bayes' theorem assuming that, given the class variable, there is independence between a given feature and any other feature of the feature set. **Multinomial Logistic Regression** [19] is a model that allows establishing a relationship between a set of independent, continuous or discrete variables and a dependent variable. **Multilayer Perceptron** is an Artificial Neural Network formed by multiple layers that is able to distinguish data that are not linearly separable.

Initially, we trained a classifier for identifying all the different activities in the whole dataset and the achieved precision was sufficient for the task. Nevertheless, in order to implement the classifier locally into a smartphone platform, with implies restrictions in the computational and storage capabilities, we also propose an optimization of the process: the implementation of a different classifier that would be computationally less expensive while preserving an acceptable precision level. In order to achieve this goal we propose a two-level classifier presented next.

5.1. Two-level hierarchical classifier

A two-level hierarchical classifier divides the classification process into two stages. The first stage consists of 4 classes, namely: *sit*, *stand*, *run* and *other* that groups other activities. The second stage performs a refinement in the classification of activities within the class *other* of the previous stage, namely: *walk*, *upstairs* and *downstairs*. When a new sample needs to be processed, the first classifier, which consists of a pruned decision tree, is applied from the very beginning. Its execution is very fast with a low computational cost. If the sample has been classified as *other*, the second classifier, which can be either a Random Forest or a Multilayer Perceptron, is applied. Its execution is slower than the first one and its computational cost is higher, but its accuracy is also better.

Considering the system performance, another parameter to take into account is the features ensemble used by the classifier. The calculation of the features also has a considerable impact on the computational cost. Since the Fast Fourier Transform complexity for a reasonable number of samples is quite high ($N \log N$), we have excluded frequency domain features in order to minimize this cost. Also, we have reduced the number of features to the minimum subset of relevant ones in the first-stage classifier.

The second stage classifier is more complex and relies on additional features. They are computed only when the first stage classifier has replied *other*. This way, we obtain a fast classifier yet very discriminative on the first stage and a heavier but more precise on the second one, resulting in a significant reduction of the computational cost while preserving a good level of accuracy.

6. Results

In our experiment we collected data from 8 male volunteers aged between 19 and 35. All the users were explained how to use the application and then they were told to put the phone in their trousers' front pocket (without any further indication) and perform certain activity. Once the activity was completed the user removed the phone from his pocket and ended the recording. The recorded data was stored in the device's internal memory and automatically uploaded to a server as soon as the phone was connected to a Wi-Fi connection. In order to skip the useless data produced during the placement and removal of the phone from the pocket, the first and last 6 seconds of each recording were removed.

In total about 1.14 hours of data were collected with around 10 minutes of total recordings among 8 subjects, each individual recording consisting of between 1 and 3 minutes of the 6 following activities: Sit, Stand, Walk, Run, Upstairs and Downstairs. As seen in Section 4, we used a window of 512 samples (approximately 5.33 seconds) with a 50% overlap. Therefore, from the collected data we obtain a dataset of 1443 instances (295 Sit, 251 Stand, 299 Walk, 274 Run, 184 Upstairs and 140 Downstairs instances). Using this data we have evaluated several classifiers included in Weka [20] and compared their results. The classifiers that were evaluated are those presented in Section 5, namely Random Forest, C4.5, Naive Bayes, Logistic Regression and Multilayer Perceptron. All the tests were conducted using 10-fold cross-validation.

When carrying out the experiments, different combinations of the features presented in Section 4 were used. Firstly, we have used a feature vector that includes all computed features, which we call full feature set. Secondly, we have used only the time domain feature. Thirdly, and in order to make a comparison between features in the spatial and frequency domain, we have also used a feature set consisting solely of frequency domain features. Lastly, we have selected the most significant features according to some feature selection algorithms discussed below. Table 1 shows the features contained in each of the above mentioned feature sets.

Table 1. Features contained in each feature set

	Feature Set			
Feature	**Full features**	**Temporal Domain**	**Frequency Domain**	**Selected**
h_mean	X	X		
v_mean	X	X		
h_sd	X	X		
v_sd	X	X		
h_mcr	X	X		X
v_mcr	X	X		X
h_mad	X	X		
v_mad	X	X		

h_25percentile	X	X		
v_25percentile	X	X		X
h_50percentile	X	X		
v_50percentile	X	X		
h_75percentile	X	X		
v_75percentile	X	X		
h_iqr	X	X		
v_iqr	X	X		
h_energy	X		X	
v_energy	X		X	
h_entropy	X		X	X
v_entropy	X		X	
corr_hv	X	X		

In the context of this work, the feature set that we want to obtain is not targeted to a particular predictive model, but to each of the above mentioned classifiers. For that reason in this paper we have utilized a filter algorithm for feature selection. Since an exhaustive search is impractical due to space dimensionality, we used heuristics, following a genetic search approach. As for filter metric we applied a consistency-based approach.

Among the different parameter settings we tried, we finally decided to run the genetic algorithm with a maximum of 500 generations, a crossover probability of 0.6, a mutation probability of 0.001 and a population size of 50. These values had proven their efficiency in previous feature selection and optimisation problems such as [21]. 10-fold cross-validation was used in 10 independent executions using different seeds. As a result we obtain a feature set shown in the last column of Table 1, containing the following features: *h_mcr*, *v_mcr*, *v_25percentile* and *h_entropy*.

Results of our experiments are shown in Table 2 where the prediction accuracy of each classifier according to the feature subset used can be observed.

Table 2. Accuracy of each classifier according to the feature set

Feature Set	Random Forest	C4.5	Naive Bayes	Logistic Regression	Multilayer Perceptron
Full	98.26	98.4	86.62	92.93	96.11
Temporal Domain	98.68	98.61	86.83	92.51	97.71
Frequency Domain	89.32	86.07	83.16	82.95	81.98
Selected	99.09	**99.23**	81.21	80.73	82.25

When the full feature set is used the best results are obtained by the C4.5 Decision Tree with a 98.4% accuracy. Random Forest performs slightly worse with a 98.26% accuracy. The rest of classifiers perform well, achieving accuracies above 90% in all cases except Naive Bayes, which obtains the worst result with 86.62% accuracy, 0.883 of precision and 0.866 of recall.

Using the feature set containing only time domain features the results are slightly better than using the full feature set. In this case the best performance is achieved by Random Forest, followed by C4.5 with 98.68% and 98.61% accuracy respectively.

The feature set consisting of the attributes of the frequency domain provides the poorest results. It can be argued that these attributes by themselves are not sufficient for a correct classification.

When using the feature set selected by means of feature selection algorithms we get the best results. Random Forest and C4.5 obtain the best results with 99.09% and 99.23% accuracy respectively. Nevertheless Logistic Regression and Multilayer

Perceptron perform significantly worse with this feature set when compared to full and temporal domain feature sets.

Table 3. Confusion matrix for Random Forest (Selected feature set)

	walk	**run**	**down**	**up**	**sit**	**stand**
walk	298	1	0	0	0	0
run	0	273	1	0	0	0
down	0	0	138	2	0	0
up	5		3	176	0	0
sit	0	0	0	0	295	0
stand	0	0	0	1	0	250

Table 4. Confusion matrix for C4.5 (Selected feature set)

	walk	**run**	**down**	**up**	**sit**	**stand**
walk	298	0	0	1	0	0
run	0	273	1	0	0	0
down	1	0	136	3	0	0
up	2	0	3	179	0	0
sit	0	0	0	0	295	0
stand	0	0	0	0	0	251

As seen in the confusion matrices shown in tables 3 and 4, *walk*, *downstairs* and *upstairs* are the most difficult activities to distinguish since their motion signature is very similar. Figure 1 shows two samples of walking and downstairs activities, where their similarity can be appreciated.

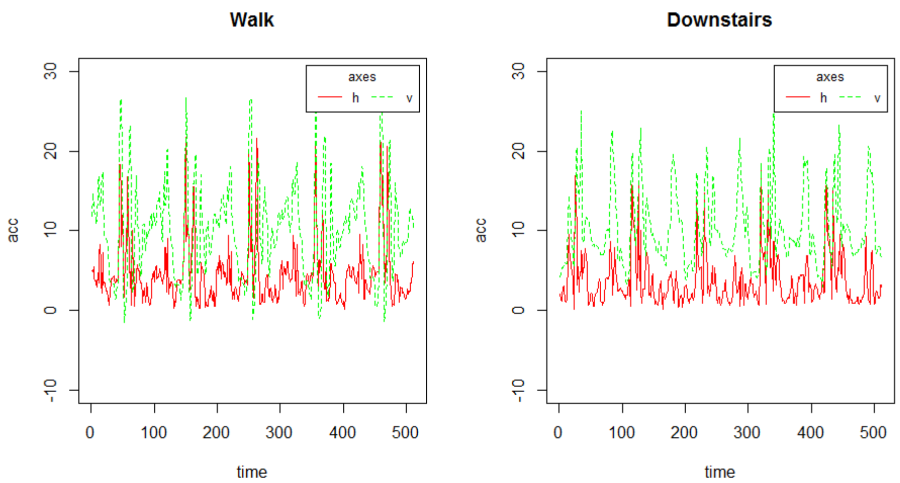

Figure 1. Horizontal and vertical components of walk and downstairs activities

6.1. Two-level classifier

In a first stage we have trained a single classifier to classify all activities using the full dataset. As exposed in Section 5.1, we have also followed a two-level hierarchical classifier approach. According to this approach, in the first level we have built a classifier to predict activities of type run, sit and stand, while the rest of activities (downstairs, upstairs and walk) are encompassed in a new class named *other*. This first-level classifier must have as high accuracy as possible, but it is also desirable to be computationally efficient in both prediction and feature extraction tasks. For this reason our first-level classifier is a pruned decision tree that uses only 3 features and achieves an accuracy of 99.93%.

Instances classified as *other* are passed to the second level classifier, which is specialized in classifying the activities named downstairs, upstairs and walk. We used a Multilayer Perceptron, which despite using all the temporal domain features and being computationally more expensive (in both training and execution time), has an accuracy as high as 99.51%. The global accuracy of the hierarchical two-level classifier is 99.72%.

7. Conclusions and future work

This paper has presented a work on automatic daily physical activity recognition using the accelerometers available on most of current smartphones. The proposed approach has been implemented using off-the-shelf equipment, proving its feasibility and practical application. In the future it could be ported to other smartphone platforms like iOS or Windows Mobile in order to cover potentially more people on the user end. Also a server-side middleware API based on REST web services for interaction with other applications would help other developments to arise on the developer and service provider end.

Several contributions have been presented for the different steps involved in the recognition process. A sensor data pre-processing algorithm has been used to permit smartphone orientation independence for classification. Two scenarios have been proposed. First, the best case scenario, without technical limitations, using all the features available. And second, a more constrained smartphone scenario, where battery life and processing capabilities are limited. After the pre-processing, a set of features have been obtained. Then, feature selection algorithms have been used adapted to each of the two scenarios.

Finally, several classification algorithms have been tested using different metrics and variations in the process, such as a comparison between the use of different feature sets.

This work is expected to increase the set of recognizable physical activities which may have a distinctive motion pattern like cycling or driving a vehicle or may be recognizable by other means. To that end, it is expected to make use of the data captured by other sensors available in current smartphones (e.g.: camera, microphone, proximity sensor). For example, using GPS to detect vehicle movement or terrain topography or using the gyroscope for more precise tracking. The use of *wearable sensor devices* which can communicate with the smartphone, such as an ECG or pulsometer will also be considered. The combination of such heterogeneous data will require the extensive use of data selection and fusion algorithms, which we want to explore, in order to maximize the accuracy while minimizing the number of features used.

There are at least two practical scenarios for these improvements. First, to be able obtain a precise monitoring of energy consumption or metabolic equivalent (MET) based on the activities recognized. And secondly, to monitor the activities carried out by the elderly people, fostering self-care and independent living under a ubiquitous but minimally-intrusive supervision.

8. Acknowledgements

This work has been supported by the research project ELDERBASK (2012-2015) funded by the Basque government under the ETORTEK programme. The authors want to thank Ph.D. Iñigo Barandiaran and Ph.D. Grégory Maclair for their collaboration.

References

[1] BAO, Ling; INTILLE, Stephen S. Activity recognition from user-annotated acceleration data. In *Pervasive computing*. Springer Berlin Heidelberg, 2004. p. 1-17.
[2] LEE, Seon-Woo; MASE, Kenji. Activity and location recognition using wearable sensors. *IEEE pervasive computing*, 2002, vol. 1, no 3, p. 24-32.
[3] MANTYJARVI, Jani; HIMBERG, Johan; SEPPANEN, Tapio. Recognizing human motion with multiple acceleration sensors. In *Systems, Man, and Cybernetics, 2001 IEEE International Conference on*. IEEE, 2001. p. 747-752.
[4] BIEBER, Gerald; VOSKAMP, Jörg; URBAN, Bodo. Activity recognition for everyday life on mobile phones. In Universal Access in Human-Computer Interaction. Intelligent and Ubiquitous Interaction Environments. Springer Berlin Heidelberg, 2009. p. 289-296.
[5] KWAPISZ, Jennifer R.; WEISS, Gary M.; MOORE, Samuel A. Activity recognition using cell phone accelerometers. ACM SigKDD Explorations Newsletter, 2011, vol. 12, no 2, p. 74-82.
[6] SUN, Lin, et al. Activity recognition on an accelerometer embedded mobile phone with varying positions and orientations. In Ubiquitous intelligence and computing. Springer Berlin Heidelberg, 2010. p. 548-562.
[7] YANG, Jun. Toward physical activity diary: motion recognition using simple acceleration features with mobile phones. In Proceedings of the 1st international workshop on Interactive multimedia for consumer electronics. ACM, 2009. p. 1-10.
[8] LONGSTAFF, Brent; REDDY, Sasank; ESTRIN, Deborah. Improving activity classification for health applications on mobile devices using active and semi-supervised learning. In Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2010 4th International Conference on-NO PERMISSIONS. IEEE, 2010. p. 1-7.
[9] BERCHTOLD, Martin, et al. Actiserv: Activity recognition service for mobile phones. In *Wearable Computers (ISWC), 2010 International Symposium on*. IEEE, 2010. p. 1-8.
[10] HENPRASERTTAE, Apiwat; THIEMJARUS, Surapa; MARUKATAT, Sanparith. Accurate activity recognition using a mobile phone regardless of device orientation and location. In *Body Sensor Networks (BSN), 2011 International Conference on*. IEEE, 2011. p. 41-46.
[11] LESTER, Jonathan, et al. A Hybrid Discriminative/Generative Approach for Modeling Human Activities. In *IJCAI*. 2005. p. 766-772.
[12] RAVI, Nishkam, et al. Activity recognition from accelerometer data. In *AAAI*. 2005. p. 1541-1546.
[13] KRISHNAN, Narayanan C.; PANCHANATHAN, Sethuraman. Analysis of low resolution accelerometer data for continuous human activity recognition. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*. IEEE, 2008. p. 3337-3340.
[14] MANNINI, Andrea; SABATINI, Angelo Maria. Machine learning methods for classifying human physical activity from on-body accelerometers. *Sensors*, 2010, vol. 10, no 2, p. 1154-1175.
[15] LEESON, George W; HARPER, Sarah. Some descriptive findings from the Global Ageing Survey (GLAS). Investing in Later Life. Oxford Institute of Ageing. 2008 - Research Report 108.
[16] HO, Tin Kam. Random decision forests. In Document Analysis and Recognition, 1995. Proceedings of the Third International Conference on. IEEE, 1995. p. 278-282.
[17] ZHANG, Harry. The optimality of naive Bayes. *A A*, 2004, vol. 1, no 2, p. 3.
[18] QUINLAN, John Ross. *C4. 5: programs for machine learning*. Morgan kaufmann, 1993.
[19] HOSMER, D. W. & LEMESHOW, S. (2000) Applied Logistic Regression, Wiley series in probability and statistics, John Wiley & Sons, 2000. Second Edition.
[20] HALL, Mark, et al. The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 2009, vol. 11, no 1, p. 10-18.
[21] CHTIOUI, Younes; BERTRAND, Dominique; BARBA, Dominique. Feature selection by a genetic algorithm. Application to seed discrimination by artificial vision. *Journal of the Science of Food and Agriculture*, 1998, vol. 76, no 1, p. 77-86.
[22] MIZELL, David. Using gravity to estimate accelerometer orientation. In *2012 16th International Symposium on Wearable Computers*. IEEE Computer Society, 2003. p. 252-252.