Human motion recognition using smartphone accelerometer

Carlos Tiana Gómez, Carles Ubach Granados

ETSEIB - Universitat Politècnica de Catalunya

Abstract

This paper presents an offline method for classifying locomotion type human activities into 6 separate groups: underground subway, train, walking, running, climbing stairs and the additional position of sitting down. The data is gathered with the accelerometers that are embedded into current smartphones, having the device placed into the subject's pocket. Data is treated to obtain meaningful features in order to impact the discrimination ability of a k-nearest neighbours clustering algorithm which is applied to classify the motion. At the end, results and conclusions are presented.

Keywords: Machine Learning, KNN, Human motion, Motion Recognition

1. Introduction

Smartphones have earned their name by providing much more than a regular cell phone would do. Their ever-increasing computing power, the inclusion of a plethora of sensors - light sensors, high-definition cameras, gyroscopes and accelerometers, among other - have extended their capabilities beyond communication. Moreover, the apparition of user-created applications (apps) extend their functionality and allow these kind of devices to perform many more tasks than those they were conceived for. Overall, the user can install the apps that best suit their needs to manage their daily life.

Nowadays, they are playing an important role in the exploration of novel information retrieval approaches directly from the users. Giving the sensors different use from its original function, known as opportunistic sensing, along with the high availability of these mass-marketed devices which nearly cover the entire population, seems an interesting area to be exploited for its application in human activity recognition.

2. State of the art

Human Activity Recognition (HAR) is an active research field in which methods for understanding human behavior are developed by interpreting attributes derived from various sources [1], [2], [3] (e.g. by sensing motion, location, physiological signals, weather and temperature etc.). It aims to identify the actions carried out by a person given a set of observations of him/herself and the surrounding environment. HAR has provided substantial contributions in human-centered areas of study such as Ambient Intelligence, Pervasive Computing and Active Assisted Living. These areas make use of HAR systems as an underlying technological tool which gathers behavioral information from people about their actions and environment during the course of their daily life and delivers context-aware data that can help to provide valuable services, products and technologies aiming to improve people's QoL.

A general representation of the principal components of a HAR process is shown in Figure 1. Many of the approaches found in literature, follow a regular structure with slight variations based on their application, sensors, and selected ML algorithms.

2.1. Activity

A HAR system is dependent on the set of activities to recognize as they can directly affect the way systems are designed and implemented. Different classifications can be done depending on the specificity of the study, as proposed in [4] activities could be classified in: Personal hygiene, mobility, feeding, communication and functional transfers. A general classification is used in this study in order to classify the

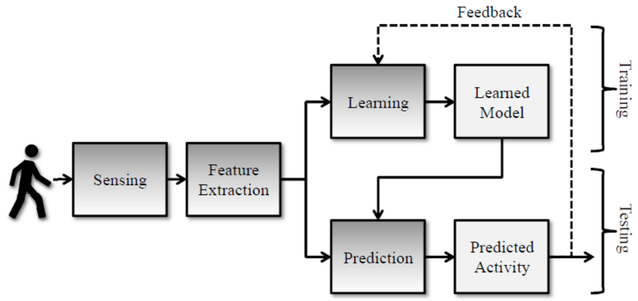

Figure 1: The Human Activity Recognition Process Pipeline with its four main blocks

wide range of documentation found in this field as well as classifying our study:

- Activities of daily living: This activities refer to people's daily self care activities, some examples could be: Watching TV, ironing, eating, showering, cleaning. As presented in [5] everyday activities are recognized with decision tree classifiers.

- Postural transition: A PT is an event with limited duration determined by its start and end times. Generally, this duration varies among healthy individuals. PTs are bounded by other two activities and represent the transition period between the two. In [6] a Gaussian Mixture Model (GMM) system was used for distinguishing between three postures (sitting, standing and lying) and five movements (sit-to-stand, stand-to-sit, lie-to-stand, stand-to-lie and walking).

- Locomotion: This type of activities seem to be high interest types since plenty of literature is focused on the study of this activities. For instance, this study which aims to recognize some of them. Examples of locomotion activities are: Walking, riding, standing, laying down, falling. Some of the literature found is [7], which aims to recognize several activities such as walking, jogging, climbing stairs, sitting, and standing. Also [8] is another example, which aims to recognize: sitting, walking at different paces, climbing stairs at different paces and jogging.

- Sports and fitness: Some authors refer to recognizing different activities related to sport like jumping, weight lifting, climbing, swimming, etc. As an example [9] aims to classify the quality of execution for weight lifting exercises.

2.2. Sensing

From the four main blocks of a HAR system, sensing is responsible of gathering the sensor data from the available sources and process them. Generally, signal conditioning (e.g. reducing noise, digitizing, amplifying) is always required to adapt the sensed signals to the application requirements. The most popular sensor used is the 3-axis accelerometer, but some diversity was found depending on the approach. Some examples of 3-axis accelerometer are [6], [7], [10]. In addition to 3-axis accelerometer [2] uses the microphone. Gyroscope data is also used in some studies such as [8], which uses 3-axis accelerometer in addition to 3-axis gyroscope, or [11] who uses 2-axis accelerometer and 1-axis gyroscope.

2.3. Feature selection and extraction

The feature extraction process is in charge of obtaining meaningful features that describe the data to largely impact the discrimination ability of a learning algorithm. The extracted features turn into the input of the ML algorithm, either for learning the model or for the activity prediction of samples when the model already exists. Depending on the application, the features required for the extraction of relevant information may vary. The most commonly used are between others the mean, standard deviation, minimum and maximum ([5], [6], [7], [8]).

2.4. ML algorithm

Several ML approaches have been developed throughout the years for HAR. It has mostly been targeted through supervised learning algorithms although semi-supervised and unsupervised methods have also been proposed ([12], [13]).

Frequentist and Bayesian models have been well covered throughout HAR literature. They involve rule-based models such as Decision Tree and Random Forest - [14], [15] -, geometric approaches including k-Nearest Neighbours, Artificial Neural Networks and Support Vector Machines - [16], [17], [18] -, and probabilistic classification methods as for example Naive Bayes classifiers, and Hidden Markov Models - [19], [20] -.

Many of these ML approaches have demonstrated comparable performance in different works (e.g. [21]), although suggesting that the effectiveness and right selection of the algorithms can be linked to other aspects such as data structure and application ([22]). Other relevant aspects for ML algorithm selection include: energy consumption, memory requirements, interpretability and computational complexity, etc. As an example, decision trees could be preferred when the model interpretability is required and support vector machines for high performance applications.

3. Activity recognition

3.1. Methodology

As it has been previously stated, the data was gathered using the accelerometer present in today's smartphones. To do so, a third-party app - *AccDataRec* [23] - was used which allowed capturing the acceleration on the three axes, as show in Figure 2. Note that the Earth's gravity is taken into account, which will affect in different ways the acceleration recordings (depending on how aligned it is to one of the phone's axes). The phone was placed always on the front right pocket of the subject with the screen facing outwards and data was collected at the maximum frequency, 400 Hz, to ensure that no information was lost.

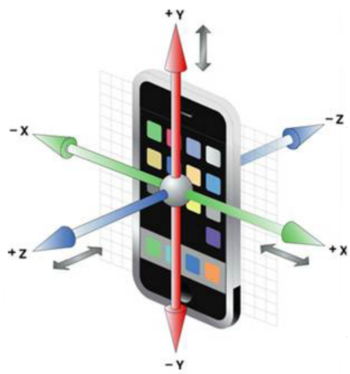

Figure 2: Acceleration axes orientation with respect to the smartphone

Data capturing sessions were performed by two individuals of different heights using two different smartphones (LG G6 and Xiaomi MiA2). The activities were:

- Walking: Walking at a steady pace, mainly on flat surfaces. It included straight lines and several turns each session.
- Running: Jogging at a regular pace, sometimes including short sprints. Mainly on flat surfaces, both straight lines and short turns.
- Climbing stairs: Climbing stairs. Some short walking intervals could be present between flights of stairs.
- Underground: Travelling by Metro, sitting down. Data only used since the subject was sitting down.
- Train: Travelling by Renfe, sitting down. As before, it is taken into account only measures where the subject is sitting down.
- Sitting down: Resting in a chair.

3.2. Feature extraction

The collected data needs to be first treated in order to extract meaningful measures. The measurements include the moment of placing the phone on the pocket just after starting recording and taking it out before finishing, thus, those periods needed to be cut off. An example of this process can be seen in Figure 3.

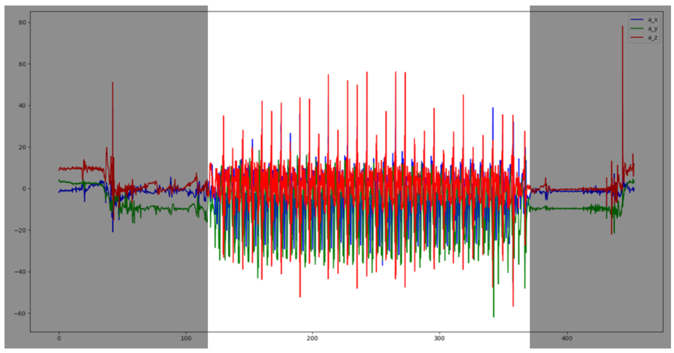

Figure 3: Extraction of meaningful measures from raw data

Once the data obtained was deemed appropriate, it was cut into 5 second windows, computed on 2.5 second intervals. Therefore, each window had half of its values shared with the previous time window. The 5 seconds allow detecting what kind of activity is being performed while, at the same time, avoid misclassification errors due to wrong readings or noise. By computing half of each window again it is taking into account previous values, and thus continuity of the activity.

For each time window, the first thing computed was the modulus of the acceleration for each measurement:

$$a = \sqrt{a_x^2 + a_y^2 + a_z^2}$$

Then, for the 4 accelerations (a_x, a_y, a_z and a), 6 features were computed, being:

- Mean
- Standard deviation
- Minimum value
- Maximum value
- Average distance between peaks
- Average Absolute Difference (AAD): The absolute average distance from the points to the mean value.

This resulted in a total of 24 features for each 2.5 seconds, that encompassed readings of 5 seconds. The distance between peaks was computed automatically detecting the greater peaks (both maximum and minimum values) and selecting the other peaks as those points that were above/below a certain threshold imposed as a relative distance to these extreme values.

As a means of visualization tool, a Principal Component Analysis (PCA) was performed. Its aim was to determine the degree of separability and/or confusion between classes. Since it was used as an exploratory analysis, all components were taken into account. Dimensionality reduction was not considered because the chosen classifier was simple enough to deal with all features. A representation of the data projected on the first 3 components of the PCA can be appreciated in Figure 4. It can be easily seen that all classes are easily separable except *walking* and climbing *stairs*, where some confusion will be expected. It makes sense, since they are very similar activities. It is surprising, however, the distance between *metro* and *renfe*, being them similar means of transport; as well as the scatter of a a few of the points of metro - those are clearly wrong, unfiltered measures -.

Activity recognition using accelerometer data is a common field of study and, thus, many solutions are

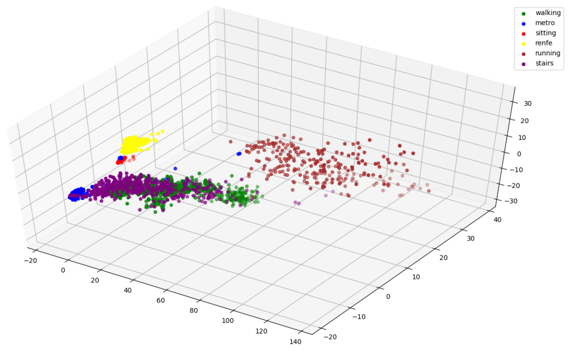

Figure 4: PCA representation of the data, first 3 components

available. Several authors, such as [24] and [25] have proved that Support Vector Machines (SVM) work well; and [5], [7] and [25] have also tried, among other techniques, decision trees. Since there is plenty of literature on more advanced techniques, it was deemed interesting to face the problem using a simpler classifier. It presents several benefits: the training period is smaller, it is more intuitive to understand - so, if results are not the expected, an explanation is easier to find -, allows for a fast classification, and not as much data is needed.

4. Results and Validation

4.1. Metrics

In this case, kNN models with k values of 3, 5 and 7 were studied. To compare them, the following measures were computed:

- No split: Accuracy score using the whole dataset as training set and validation set
- 50% split: Accuracy score, using half of the dataset as training set and the other half as validation set.
- 10-fold Cross validation. Mean Accuracy and f-1 scores after a 10-fold cross validation procedure. Standard deviations for each of them also provided.
- Classification report. The Python module *sklearn* includes a classification report for the model studied, which includes precision, recall, and f-1 score. It uses a 50% split between training set and validation set as well.
- Confusion Matrix. As before, it uses half the set for training and the other half for validation

Dataset Split	Measure	Number of Neighbours k = 3	k = 5	k = 7
No Split	Accuracy	0.9802	0.9738	0.9675
50% Split	Accuracy	0.9561	0.9562	0.9505
10-Fold Cross Validation	Accuracy	0.95 ±0.10	0.94 ±0.11	0.94 ±0.11
	f-1 score	0.96 ±0.07	0.95 ±0.08	0.95 ±0.08

Table 1: Score results for k = 3, 5, 7

4.2. Results

Table 1 displays the classification scores for the 3 values of k used. It is checked that the fewer number of neighbours considered, the better. When only 3 of them are taken into account, accuracy and f-1 (which measures the degree of confusion of the model) are the highest. Moreover, their standard deviation is the lowest, yielding more consistent results. Despite this, it is remarkable how close all these scores are and, so far, how well is performing the classifier.

Tables 2, 3 and 4 show the confusion matrices of the 3 considered classifiers. In this case, both k = 3 and k = 5 present the same results: a total of 29 confusion cases between climbing Stairs and Walking (skewed towards confusing Walking with Stairs) and 2 cases between Metro and Sitting. The first one was already expected, as it was a conclusion that was derived from the PCA analysis (recall Figure 4). This results further confirm that those 2 activities are very similar. The confusion between metro and sitting, albeit small, is present in all classifiers. Once again, could be due to the similarities between both activities - data gathered while travelling by metro was taken sitting down - or caused by the points observed on the PCA that were far away from the metro point cloud.

	Walk	Run	Met	Ren	Sit	Sta
Walking	104	0	0	0	0	12
Running	0	99	0	0	0	0
Metro	0	0	84	0	0	0
Renfe	0	0	0	90	0	0
Sitting	0	0	2	0	102	0
Stairs	17	0	0	0	0	197

Table 2: Confusion Matrix k = 3

	Walk	Run	Met	Ren	Sit	Sta
Walking	108	0	0	0	0	8
Running	0	99	0	0	0	0
Metro	0	0	84	0	0	0
Renfe	0	0	0	90	0	0
Sitting	0	0	2	0	102	0
Stairs	21	0	0	0	0	193

Table 3: Confusion Matrix k = 5

	Walk	Run	Met	Ren	Sit	Sta
Walking	108	0	0	0	0	8
Running	0	99	0	0	0	0
Metro	0	0	84	0	0	0
Renfe	0	0	0	90	0	0
Sitting	0	0	3	0	101	0
Stairs	24	0	0	0	0	190

Table 4: Confusion Matrix k = 7

Finally, the last metric presented are the ones provided for each class by *sklearn* in Tables 5, 6 and 7. It is interesting to note the high values of all the presented results. Across the three classifiers and three scores presented, the Walking class is the one that most underperforms. This is in line with the findings of the confusion matrix, suggesting that sometimes it is misclassified as Stairs, being this last class the second-to-worst. The model with the most number of neighbours - k = 7 - is the one that, once again, yields the worst results. However, it is not as clear in the case of the other two models. The first one, k = 3, presents worse precision but better recall; whereas the second one, k = 5, the opposite. It makes sense, then, that both have the same f-1 score. In this case, it is actually not that clear which one (precision or recall) is more important, i.e., it has the same relevance misclassifying Walking as Stairs as doing so with Stairs and Walking. Therefore, both k = 5 and k = 3 present similar performance. Given that k = 3 is more averaged - it has less difference between precision and recall -, it was considered a better solution.

	Precision	Recall	f-1 score
Walking	0.86	0.90	0.88
Running	1.00	1.00	1.00
Metro	0.98	1.00	0.99
Renfe	1.00	1.00	1.00
Sitting	1.00	0.98	0.99
Stairs	0.94	0.92	0.93

Table 5: Classification report k = 3

	Precision	Recall	f-1 score
Walking	0.84	0.93	0.88
Running	1.00	1.00	1.00
Metro	0.98	1.00	0.99
Renfe	1.00	1.00	1.00
Sitting	1.00	0.98	0.99
Stairs	0.96	0.90	0.93

Table 6: Classification report k = 5

	Precision	Recall	f-1 score
Walking	0.82	0.93	0.87
Running	1.00	1.00	1.00
Metro	0.97	1.00	0.98
Renfe	1.00	0.97	0.99
Sitting	1.00	0.98	0.99
Stairs	0.96	0.89	0.92

Table 7: Classification report k = 7

Therefore, the chosen classifier is a kNN with 3 neighbours. As it was stated, its better results in accuracy and f-1 score along with a less standard deviation (more uniform results), and, especially, the least confusion between the classes Walking and Running deemed it as the best classifier among the ones studied.

4.3. Validation

To check the results obtained, chosen classifier was tested with new data that had not been used before. Additionally, this new data was not cut - it included the time period of putting in and taking out the phone from the pocket -, which adds noise (since it is not a class).

Accuracy was tested with 4 different kinds of motion, being the results displayed at Table 8. It is observed that results are very satisfactory: 3 out of 4 of them present an accuracy of 85% or over, whereas the other, Walking, was expected to yield worse results due to the aforementioned confusion with Stairs. However, it is surprising the performance of the classifier taking into account that there is noise added.

Class	Accuracy
Walking	0.8126
Running	0.8473
Metro	0.9139
Renfe	0.9237

Table 8: Validation measures

5. Conclusions

In this paper a simple classifier has been trained to identify 6 different human activities using smartphone's accelerometer data. A first visual approach of the data was performed which allowed for a better understanding of the problems that later on arose. The aim of using a k-nearest neighbours classifier was to prove the effectiveness of simpler and fast machine learning methods against more complex state of the art techniques. Moreover, this allows for a better understanding the problems the classifier is facing and why results are the ones obtained, what could be improved and, over all, how and where could they be improved - a much more challenging task more sophisticated classifiers such as with neural networks and support vector machines - .

As far as results go, the classifier yielded consistently good performance measures, despite the similarities between some of the considered activities. Therefore, the classification task can be deemed a success.

References

[1] D. M. Karantonis, M. R. Narayanan, M. Mathie, N. H. Lovell, B. G. Celler, Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring, IEEE 10 (2006) 156–167.

[2] P. Lukowicz, J. A. Ward, H. Junker, M. Stäger, G. Tröster, A. Atrash, T. Starner, Recognizing workshop activity using body worn microphones and accelerometers, in: International conference on pervasive computing, Springer, pp. 18–32.

[3] R. Poppe, Vision-based human motion analysis: An overview, Computer vision and image understanding 108 (2007) 4–18.

[4] B. Bruno, F. Mastrogiovanni, A. Sgorbissa, T. Vernazza, R. Zaccaria, Human motion modelling and recognition: A computational approach, in: 2012 IEEE (CASE), IEEE, pp. 156–161.

[5] L. Bao, S. S. Intille, Activity recognition from user-annotated acceleration data, in: A. Ferscha, F. Mattern (Eds.), Pervasive Computing, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 1–17.

[6] F. R. Allen, E. Ambikairajah, N. H. Lovell, B. G. Celler, Classification of a known sequence of motions and postures from accelerometry data using adapted gaussian mixture models, Physiological Measurement 27 (2006) 935–951.

[7] J. R. Kwapisz, G. M. Weiss, S. A. Moore, Activity recognition using cell phone accelerometers, SIGKDD Explor. Newsl. 12 (2011) 74–82.

[8] W. Wu, S. Dasgupta, E. E. Ramirez, C. Peterson, G. J. Norman, Classification accuracies of physical activities using smartphone motion sensors, Journal of medical Internet research 14 (2012) e130.

[9] E. Velloso, A. Bulling, H. Gellersen, W. Ugulino, H. Fuks, Qualitative activity recognition of weight lifting exercises, in: Proceedings of the 4th Augmented Human International Conference, ACM, pp. 116–123.

[10] N. Ravi, N. Dandekar, P. Mysore, M. L. Littman, Activity recognition from accelerometer data, in: Aaai, volume 5, pp. 1541–1546.

[11] S.-W. Lee, K. Mase, Activity and location recognition using wearable sensors, IEEE 1 (2002) 24–32.

[12] M. Stikic, K. Van Laerhoven, B. Schiele, Exploring semi-supervised and active learning for activity recognition, in: 2008 12th IEEE, IEEE, pp. 81–88.

[13] D. Wyatt, M. Philipose, T. Choudhury, Unsupervised activity recognition using automatically mined common sense, in: AAAI, volume 5, pp. 21–27.

[14] B. Coley, B. Najafi, A. Paraschiv-Ionescu, K. Aminian, Stair climbing detection during daily physical activity using a miniature gyroscope, Gait & posture 22 (2005) 287–294.

[15] M. Ermes, J. Parkka, L. Cluitmans, Advancing from offline to online activity recognition with wearable sensors, in: 2008 30th Annual International Conference of the IEEE, IEEE, pp. 4451–4454.

[16] Z. He, L. Jin, Activity recognition from acceleration data based on discrete consine transform and svm, in: 2009 IEEE, IEEE, pp. 5041–5044.

[17] A. M. Khan, Y.-K. Lee, S.-Y. Lee, T.-S. Kim, Human activity recognition via an accelerometer-enabled-smartphone using kernel discriminant analysis, in: 2010 5th international conference on future information technology, IEEE, pp. 1–6.

[18] U. Maurer, A. Smailagic, D. P. Siewiorek, M. Deisher, Activity recognition and monitoring using multiple sensors on different body positions, Technical Report, CARNEGIE-MELLON UNIV PITTSBURGH, 2006.

[19] E. M. Tapia, S. S. Intille, W. Haskell, K. Larson, J. Wright, A. King, R. Friedman, Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor, in: 2007 11th IEEE, IEEE, pp. 37–40.

[20] C. Zhu, W. Sheng, Human daily activity recognition in robot-assisted living using multi-sensor fusion, in: 2009 IEEE, IEEE, pp. 2154–2159.

[21] A. Mannini, A. M. Sabatini, Machine learning methods for classifying human physical activity from on-body accelerometers, Sensors 10 (2010) 1154–1175.

[22] D. H. Wolpert, W. G. Macready, et al., No free lunch theorems for optimization, IEEE 1 (1997) 67–82.

[23] Capture accelerometer data, 2014. Available at: http://tiny.cc/c7xz7y.

[24] L. Sun, D. Zhang, B. Li, B. Guo, S. Li, Activity recognition on an accelerometer embedded mobile phone with varying positions and orientations, in: Z. Yu, R. Liscano, G. Chen, D. Zhang, X. Zhou (Eds.), Ubiquitous Intelligence and Computing, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 548–562.

[25] S. Wang, C. Chen, J. ma, Accelerometer based transportation mode recognition on mobile phones, volume 4, pp. 44–46.