

Recognizing atomic activities with wrist-worn accelerometer using machine learning

Martin Gjoreski, Hristijan Gjoreski, Mitja Luštrek, Matjaž Gams
Mednarodna podiplomska šola Jožefa Stefana
Odsek za Inteligentne Sisteme, Institut Jožef Stefan
martin.gjoreski@ijs.si

ABSTRACT

In this paper we present a machine learning approach to activity recognition using a wristband device. The approach includes: data acquisition, filtering, feature extraction, feature selection, training a classification model and finally classification (recognizing the activity). We evaluated the approach using a dataset consisting of 10 everyday activities recorded by 10 volunteers. Even though the related work shows that with a wrist-worn device one should expect worse accuracy compared to devices worn on other body locations (chest, thigh and ankle), our tests showed that the accuracy is 72%, which is slightly worse compared to the accuracy of the thigh (82%) and ankle-placed devices (83%); and slightly better compared to a chest-placed device (67%). Additionally, by applying feature selection and increasing the window size, the accuracy increased by 5%.

Keywords

Activity recognition, wrist, accelerometer, machine learning, classification, feature extraction.

1. INTRODUCTION

With the recent trends and development in sensor technology (miniaturization – MEMS; connectivity - Bluetooth low energy and WiFi; battery - Li-ion) people get used to the idea of: wearing an additional device on themselves beside the telephone, or replacing an existing one – the wristwatch. These devices provide sensor data which can be used for extracting useful information about the user: how many calories are burned during the day, what types of activities are performed during the day (sedentary vs. dynamic ones), detecting alarming situations (e.g., falls), detecting behavioral changes, and similar. A service that emerged as an essential basic building block in developing such applications is Activity Recognition (AR). The activity of the user provides reach contextual information which can be used to further infer additional useful information [1][2][3]. Wristband devices are becoming popular mainly because people are more or less accustomed to wear watches and therefore this placement is one of the least intrusive placements to wear a device. Nowadays we are witnessing various types of fitness/health oriented wrist-worn devices, such as: FitBit¹, Empaica², Microsoft band³; and also in the last few years smartwatches are gaining attraction: Apple watch, Android wear wristwatches, Samsung Galaxy gear, etc.

¹ www.firbit.com

² www.epatica.com

³ <http://www.microsoft.com/microsoft-band/en-us>

In this paper we present a machine learning approach to activity recognition using a wristband device. The approach includes: data acquisition, filtering, feature extraction, feature selection, training a classification model and finally classification (recognizing the activity). It was evaluated on a dataset consisting of 10 everyday activities recorded by 10 volunteers. The results showed that with a wrist-worn device one can recognize much more activities than what is commonly used for (i.e., walking - step counter), running, lying - sleeping). Additionally, the accuracy is comparable even in some cases higher compared to devices worn on other body locations (chest, thigh and ankle), which are more established and commonly used for activity recognition tasks.

2. RELATED WORK

The most recent literature in AR field shows that wearable accelerometers are among the most suitable sensors for unobtrusive AR [7]. Accelerometers are becoming increasingly common because of their lowering cost, weight and power consumption. Currently the most exploited and probably the most mature approach to AR is with wearable accelerometers by using machine learning approach [18][16][17]. This approach usually implements widely used classification methods, such as Decision Tree, SVM, kNN and Naive Bayes.

For the sake of the user's convenience, AR applications are often limited to a single accelerometer. Numerous studies have shown that the performance of an Activity Recognition System strongly depends on the accelerometer placement (e.g., chest, abdomen, waist, thigh, ankle) and that some placements are more suitable (in terms of AR performance) for particular activities [4][6][5].

In the past the wrist was the least exploited placement for AR. Mainly because of our inclination towards frequent hand movements which negatively influence an AR system. The researchers usually were testing chest, waist, thighs (left and right) [18][19], ankles (left and right) and neck. The results vary a lot and cannot be compared through different studies (different datasets, different algorithm parameters, different approaches, etc.). In our previous work we also tested most of these locations on two datasets. On the first one, the results showed that all of the locations perform similarly achieving around 82% accuracy [8]. On the second dataset, where the experiments were more thorough (bigger dataset, improved algorithms) the results showed that thigh and ankle perform similarly (82% and 83% respectively) and achieve higher accuracy compared to the chest (67%) [9].

However, with the penetration of the wrist-worn fitness trackers and smartwatches, it is to be expected that wrist sensor placement will be quite researched area. Recently, Trust et al. [12] presented a study for hip versus wrist data for AR. The models using hip data slightly outperformed the wrist-data models. Similarly, in the

study by Rosenberg et al. [13] for sedentary and activity detection, the models using hip data outperform the wrist models. In the study by Manini et al. [14] ankle data models achieved high accuracy of 95.0% that decreased to 84.7% for wrist data models. Shorter (4 s) windows only minimally decreased performances of the algorithm on the wrist to 84.2%. Ellis et al. [15] presented an approach for locomotion and household activities recognition in a lab setting. For one subset of activities the hip-data models outperformed the wrist data, but over all activities the wrist-data models produced better results. Garcia-Ceja et al. [20] presented person-specific activity detection for activities such as: shopping, showering, dinner, computer-work and exercise.

3. EXPERIMENTAL SETUP

3.1 Sensor equipment and experimental data

The sensor equipment consists of a Shimmer sensor platform. The sensors were placed on the chest, thigh, ankle and wrist with adjustable straps. The accelerometer data was acquired on a laptop in real-time via Bluetooth using frequency of 50 Hz. The data was manually labeled with the corresponding activity. Ten volunteers performed a complex 90-minute scenario which included ten elementary activities: lying, standing, walking, sitting, cycling, all fours, kneeling, running, bending and transition (transition up and transition down). These activities were selected as the most common elementary, everyday-life activities. In this paper, we are performing analysis only on the wrist-sensor data. The data from the other sensors (chest, thigh and ankle) has been extensively analyzed in our previous studies [6][7][9][10][11]. Nevertheless, the results presented in those studies provide valuable guidelines to which we are comparing. Overall, 1,000,000 raw-data samples per volunteer were recorded. These raw-data samples were transformed into approximately 7,000 data instances per volunteer. Figure 1 shows the instances' class distribution.

3.2 Experimental Method

Figure 2 shows the machine learning approach used in this research. It includes the following modules: data segmentation, data filtering, feature extraction, feature selection and building a classification model. The data segmentation phase uses an overlapping sliding-window technique, dividing the continuous sensor-stream data into data segments – windows. A window of a fixed size (width) moved across the stream of data. Once the sensor measurements are segmented, further pre-processing is

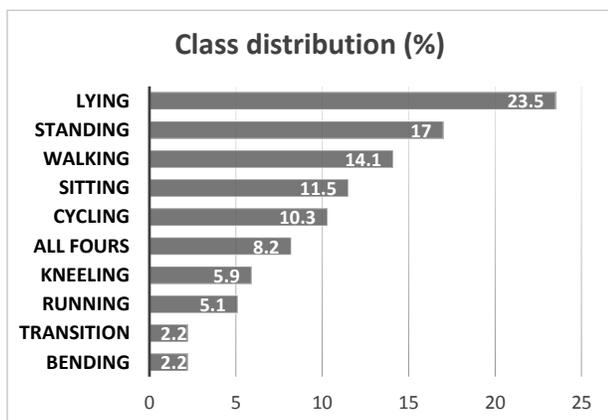


Figure 1. Class (activity) distribution

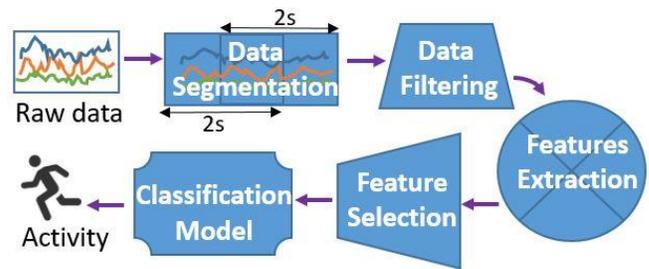


Figure 2. Activity recognition approach

performed using two simple filters: low-pass and band-pass. The feature extraction phase produces lowpass filtered features that measure the posture of the body, and band-pass filtered features that represent: the motion shape, the motion variance, the motion energy, and the motion periodicity [21]. The features extraction phase results in 53 extracted features. Since all of the features are extracted from one data source (wrist accelerometer), a high feature correlation is expected. For that reason the feature selection method is based on feature-correlation analysis which serves the purpose of removing correlated and “non-informative” features. Low informative features are considered those that have low information gain. The information gain evaluates the worth of a feature by measuring the information gain with respect to the class. Regarding the correlation of the features, we checked for Pearson’s correlation, which measures linear correlation between features, and Spearman correlation, which measures how well the relationship between two variables can be described using a monotonic function. The feature selection steps are:

- Rank features by gain ratio.
- Starting from the lowest ranked feature, calculate its correlation coefficients (Pearson and Spearman) with each of the features ranked above. If it has a correlation coefficient higher than 0.95 with at least one feature, remove it.
- Repeat step 2 until 50% of the features are checked.

Figure 3 shows the results of the Person’s correlation analysis before (left) and after (right) the feature selection phase. On the figure there are two correlation matrices, 53x53 (left) and (35x35) right. Each row (column) represents different feature. Red color represents negative, blue color represents positive and the intensity of the color represents the absolute value of the correlation. This figure on one hand depicts the dimensionality reduction of 34% (from 53 features to 35 features), and on the other hand the correlation reduction (the intensities of the colors). On the left matrix some regions with high correlations are marked (with black rectangles) to present candidate features that the feature selection algorithm may delete. On the right matrix there is high correlation between some of the features even after the feature selection phase. These are features that have high gain ratio index. In each experiment we checked the accuracy with and without the feature selection phase. The experiments with feature selection phase achieved at least equal results and in some cases even slightly better results. Once the features are extracted (and selected), a feature vector is formed, and is fed into a classification model, which recognizes the activity of the user. The classification model is previously trained on feature vectors computed over training data. We tested several machine learning algorithms, Decision Tree, RF, Naive Bayes, and SVM with Leave-one-user out cross-validation.

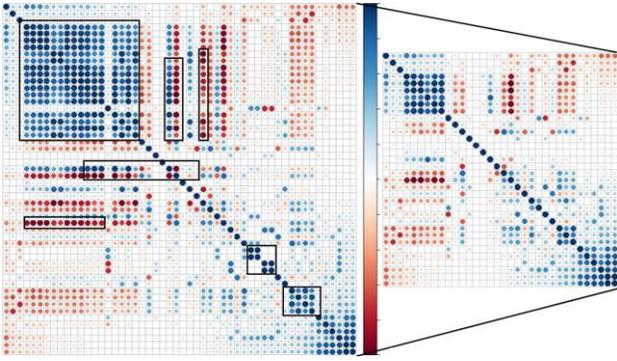


Figure 3. Person's correlation matrix before (left) and after (right) feature selection

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Wrist vs other sensor placement

First we wanted to know how well the machine learning models will perform when built using wrist-accelerometer compared to other body placements (ankle, chest and thigh). Here we present results without the feature selection phase in order to be comparable to our previous studies. Figure 4 shows accuracy comparison based on which sensor placement is used for building the machine learning model. For each study the same data is used with almost identical methodologies (same segmentation scheme, same number of features and same classifiers). From the figure we can see that for our dataset ankle or thigh sensor-placement provide better results than wrist and chest.

From now on we will report only on results achieved by the Random Forest (RF) classifier (which is not included in Figure 5 due to lack of information for the Ankle, Chest and Thigh accuracies) since with accuracy of 74% it performed best in our experiments.

Table 1 shows the confusion matrix, precision, recall and F1 score for each class obtained by the RF classifier. The F1 score for each of the activities shows that bending, kneeling and transition, are the three activities that are hard to recognize by the classifier. Standing and all fours are somewhat in the middle, whereas sitting, walking, lying, cycling and running are the activities that are recognized with a satisfying level.

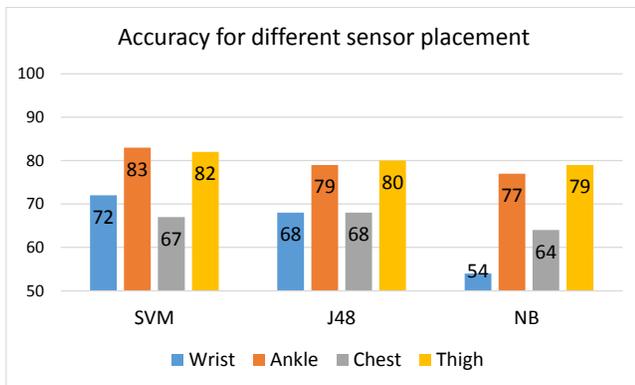


Figure 4. Accuracy for wrist vs other sensor placement

Table 1. RF confusion matrix and performance metrics (recall, precision and F1 score) per class

RF - Acc = 74%	1	2	3	4	5	6	7	8	9	10
Walking-1	8428	1067	42	43	305	7	4	113	114	4
Standing-2	298	8185	151	318	400	164	68	123	1381	1079
Sitting-3	6	300	6256	1489	0	9	0	68	10	95
Lying-4	2	303	1531	14889	61	21	0	38	71	90
Bending-5	99	590	1	23	798	13	0	10	26	56
Cycling-6	99	1110	0	8	117	5950	0	3	72	46
Running-7	44	410	4	19	0	0	3157	2	7	1
Transition-8	175	402	50	57	16	2	0	643	161	80
All_fours-9	123	1140	6	128	72	77	6	65	4283	96
Kneeling-10	52	1980	267	391	127	108	0	43	492	818
Recall	83	67	76	88	6	80	87	41	71	19
Precision	90	53	75	86	42	94	98	58	65	35
F1 score	87	59	76	87	11	87	92	48	68	25

4.2 Effects of window size on classification performance

In our activity recognition approach there is a "data segmentation" phase where an overlapping sliding-window is used for transforming the continuous data stream into a data segments over which the features are calculated. The reported results in the previous experiments (Section 3.1) are achieved using a window of 2s with an overlap of 1s. That means for predicting the activity at time T, we are taking accelerometer data starting from T-2s to T. The next prediction is at time T+1s and we are taking data starting from T-1s to T+1s, and so on. Basically we are predicting activity once per second by analyzing the data from the previous two seconds.

In these experiments we wanted to study the effects of the window size on the performance of the RF classifier. Moreover we wanted to see if choosing a shorter window can improve the accuracy of the short-duration activities, such as bending and transition, and the other way around (if choosing a longer window can improve the accuracy of the long-duration activities, e.g., standing). Table 2 shows the summarization of these experiments. The second row presents the window which is used for the experiments, starting from 1s window with 0.5s overlap, all the way to 10s window with 8s overlap. For each experiment the F1 score per activity and the overall accuracy of the RF classifier is reported. This table presents several observations:

- A short window of 1s with 0.5s overlap does not improve the performance of the classifier for the short-duration activities. Better performance is achieved when longer windows are used.
- Only for the activity Running a shorter window (2s with 1s overlap) produces better performance (precision, recall and F1 score) than the other window lengths. For all the other activities the longer the window the better the performance of the classifiers.
- The overall accuracy increases by increasing the window. However, this increase is statistically important only for the increases from 1(0.5) to 2(1) and from 2(1) to 4(2). Also, the number of instances is highest for the windows size 1 (around 7000 per person), for window size 2 it is around 5000 per person, and for the rest of the window sizes the number of instances is equal i.e., around 3200 per person.

Table 2. RF classification performance for varying window

Random Forest		Data (overlap) window - seconds						
Metrics	Activity	1(0.5)	2(1)	4(2)	6(4)	8(6)	10(8)	
F1 score	Walking	83.8	86.6	90.4	91.1	90.9	91.4	
	Standing	55.3	60.0	64.4	65.6	66.0	66.8	
	Sitting	71.2	75.1	75.9	77.9	77.4	77.4	
	Lying	84.8	86.5	87.7	88.7	88.9	89.2	
	Bending	12.7	12.2	13.1	14.4	14.7	14.4	
	Cycling	81.4	85.0	88.4	89.1	89.8	90.0	
	Running	96.2	97.2	97.1	97.0	96.7	96.9	
	Transition	32.9	47.3	61.0	61.2	62.4	61.7	
	All fours	63.6	69.0	71.4	72.9	73.4	74.7	
	Kneeling	21.4	23.8	25.1	26.1	24.3	23.7	
Accuracy		70.8	74.4	77.5	78.6	78.8	79.1	

5. CONCLUSION

The high correlation between the features allowed for reducing the feature dimensionality by 34% (from 53 features to 35) while keeping the classifier performance. For that we removed features that have low information gain and high correlation.

Wrist accelerometer data produces slightly worse classifying performance than thigh and chest accelerometer data. The most problematic activities (from the 10 we analyzed) are bending, kneeling and transition. The results for the other activities are somewhat expected, except for the activity standing which is mixed by the classifier with almost all of the other activities. We hypothesize that during the data collecting scenario the volunteers were frequently moving their hands (while talking to each other), so the classifiers sees these hand movements as a movement of the whole body. Regarding the size of the window in the segmentation phase, it should be noted that for a longer window size the features are calculated over bigger data segments which may slightly increase the computational complexity. Window of 4s with 2s overlap may be the best tradeoff between computational complexity and classifier performance.

In these experiments each instance (activity) is treated independently of the previous activity, whereas in reality we rarely change our activity every 2s (2s is the predicting frequency for the highest achieved results - that is window size of 4, 6, 8 or 10 seconds). For future work we may use higher level features that provide information about the dependency of the instances [11].

Acknowledgements. The authors would like to thank Simon Kozina for his help in programming the AR algorithm. This work was partly supported by the Slovene Human Resources Development and Scholarship funds and partly by the CHIRON project - ARTEMIS JU, grant agreement No. 2009-1-100228.

6. REFERENCES

[1] Gregory, D.A., Anind, K. D., Peter J. B., Nigel, D., Mark, S. and Pete, S. Towards a better understanding of context and context-awareness. 1st International Symposium Handheld and Ubiquitous Computing, pp. 304-307, 1999.

[2] Vyas, N., Farrington, J., Andre, D. and Stivoric, J. I. Machine learning and sensor fusion for estimating continuous energy expenditure. Innovative Applications of Artificial Intelligence Conference, pp. 1613-1620, 2012.

[3] Gjoreski, H., Kaluža, B., Gams, M., Milić, R. and Luštrek, M. Ensembles of multiple sensors for human energy expenditure estimation. Proceedings of the 2013 ACM international joint

conference on Pervasive and Ubiquitous computing, Ubicomp, pp. 359-362, 2013.

[4] Atallah, L., Lo, B., King, R. and Yang, GZ. Sensor Placement for Activity Detection Using Wearable Accelerometers. In Proceedings BSN (2010), 24–29.

[5] Cleland, I., Kikhia, B., Nugent, C., et al. Optimal Placement of Accelerometers for the Detection of Everyday Activities. Sensors. 2013; 13 (7).

[6] Gjoreski, H., Luštrek and M., Gams, M. Accelerometer Placement for Posture Recognition and Fall Detection. 7th International Conference on Intelligent Environments (IE). 2011; 47–54.

[7] Gjoreski, H. et al. Competitive Live Evaluation of Activity-recognition Systems. IEEE Pervasive Computing, Vol:14 , Issue: 1, pp. 70 – 77 (2015).

[8] Gjoreski, H. Master Thesis, 2011. Adaptive Human Activity Recognition and Fall Detection Using Wearable Sensors. Jozef Stefan International Postgraduate School.

[9] Kozina, S., Gjoreski, H., Gams, M. and Luštrek, M. Three-layer activity recognition combining domain knowledge and meta-classification. Journal of Medical and Biological Engineering, vol. 33, no. 4. p.p. 406-414, (2013).

[10] Gjoreski, H., Gams, M. and Luštrek, M. Context-based fall detection and activity recognition using inertial and location sensors. Journal of Ambient Intelligence and Smart Environments, vol. 6, no. 4, p.p. 419-433 (2014).

[11] Gjoreski, H., Kozina, S., Luštrek, M and M. Gams. Using multiple contexts to distinguish standing from sitting with a single accelerometer. European Conference on Artificial Intelligence (ECAI), 2014.

[12] Trost, S.G., Zheng, Y. and Weng-Keen Wong. Machine learning for activity recognition: hip versus wrist data. *Physiol Meas.* 2014 Nov; 35(11):2183-9. doi: 10.1088/0967-3334/35/11/2183.

[13] Rosenberger, M. E., et al. Estimating Activity and Sedentary Behavior From an Accelerometer on the Hip or Wrist. *Med Sci Sports Exerc.* 2013 May; 45(5): 964–975. doi:10.1249/MSS.0b013e31827fd9c.

[14] Mannini, A. et al. Activity recognition using a single accelerometer placed at the wrist or ankle. *Med Sci Sports Exerc.* 2013 November; 45(11): 2193–2203. doi:10.1249/MSS.0b013e31829736d6

[15] Ellis K. et al. A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. *Physiol. Meas.* 35 (2014) 2191–2203. doi:10.1088/0967-3334/35/11/2191

[16] Wu H, Lemaire ED and Baddour, N. Activity Change-of-state Identification Using a Blackberry Smartphone. *Journal of Medical and Biological Engineering*, 2012; 32: 265–272.

[17] Lai C, Huang YM, Chao HC, Park JH. Adaptive Body Posture Analysis Using Collaborative Multi-Sensors for Elderly Falling Detection. *IEEE Intelligent Systems*. 2010; 2–11.

[18] Kwapisz JR, Weiss GM, Moore SA. Activity Recognition using Cell Phone Accelerometers. *Human Factors*. 2010; 12:74–82.

[19] Ravi N, Dandekar N, Mysore P, Littman ML. Activity Recognition from Accelerometer Data. In Proceedings of the 17th conference on Innovative applications of artificial intelligence. 2005; 1541–1546.

[20] Garcia-Ceja, E., Brena R. F., Carrasco-Jimenez J. C. and Garrido, L. Long-Term Activity Recognition from Wristwatch Accelerometer Data. *Sensors* 2014, 14, 22500-22524;

[21] Tapia, E. M. Using Machine Learning for Real-time Activity Recognition and Estimation of Energy Expenditure. Ph.D. Thesis, Massachusetts Institute of Technology, 2008.