

Human Physical Activity Recognition from Acceleration Data under Naturalistic Conditions

Abhijeet Parmar^{a,*}, Prashant Sinha^b

^a*Design Innovation Centre, University of Delhi, Delhi, India 110007*

^b*Cluster Innovation Centre, University of Delhi, Delhi, India 110007*

Abstract

Providing accurate and opportune information on peoples activities and behaviors is one of the most important tasks in pervasive computing. Innumerable applications can be visualized, for instance, in medical, security, entertainment, and tactical scenarios. In this study, we explore the detection of human activities (walk, run, jog, climb stairs up, climb stairs down, stand still, biking) from data acquired using single tri-axial accelerometer. We address the task of detecting activities irrespective of the user specific sensor position, orientation and body attachment. We compare the classification accuracies in case of orientation uncorrected data in different positions and orientations across multiple individuals. Also, we show that a single device and accelerometer data is sufficient to detect the activities. Additionally we also show significant activity detection accuracies when the user holds the device in the hand or carries it in the pockets, as in a real life scenario. Random Forest classifiers showed the best performance recognizing everyday activities with an overall accuracy rate of 96%. Finally, we present some open problems and ideas that, due to their high relevance, should be addressed in future research.

Keywords: Human Activity Recognition, Feature Extraction, Windowing, Pervasive Computing, Random Forest

1. Introduction

Providing and predicting information about user's intent, behaviour, and activities is one of the most important task in pervasive computing. Human Activity Recognition (HAR) is a problem of inferring the state of the user by exploiting the data obtained from sensors that are already present in many mobile devices. The information provided from HAR system can be utilised to provide contextual awareness which can further be utilised in innumerable applications. Context aware mobile apps may optimize their user experience based on the activity that the user is presently performing. It can also be used in various patient and baby monitoring systems to provide accurate and realtime activity data which may prove useful to the caretakers in case of emergencies Anguita et al. (2013). Activity information may also be used to remove false positives from pedometer data in fitness tracking applications for more accurate gait analysis.

2. Previous Work

Various studies have shown efficiency of machine learning methods for the solution of HAR task. Anguita et al.

(2013) proposes an energy efficient method of activity classification using fixed point arithmetic scheme for training a SVM. The classification accuracy of their scheme is 89% for 6 activity classes. Gupta and Dallas (2014) is the closest to the present study. They take a 6 seconds window to compute the features. The method involves further windowing to calculate the features such as mean trends and windowed mean difference. Their result had an accuracy of 97-98% using kNN and Naive Bayes classifier.

Banos et al. (2014) investigate the relation between window length and the classification accuracy, while Reyes-Ortiz et al. (2015) report success using signal processing and filters on the combined gyroscope and accelerometer data.

He and Jin (2009) use PCA and Cosine transform for the HAR task. They report 97% accuracy between 4 activity classes. Shoaib et al. (2013) also use gyroscope and accelerometer data in their study while claiming the decrease in classification accuracy due to the bias arising by shifting the position of the device.

3. Undertaken Challenge

3.1. Hardware Complexity Constraint

It is required that the task is solved solely through the data from a single triaxial accelerometer. This reduces the hardware complexity.

*Corresponding author

Email addresses: abhijeet@ducic.ac.in (Abhijeet Parmar), prashant@ducic.ac.in (Prashant Sinha)

3.2. Attitude Independent Recognition (Axis Constraint)

It is required that the task be solved without any prior knowledge of the location and orientation of the motion tracker. This follows from the fact that the motion tracker could be placed at any spot on the user, for example their pockets, their arm or in their hand.

3.3. Hardware Friendly Classification (Time Complexity Constraint)

It is required that the features that are generated from the sensor sample sequence must be computationally low cost. This is required for energy efficient implementation of the technique in various portable devices such as a fitness tracker or a mobile phone.

4. Method

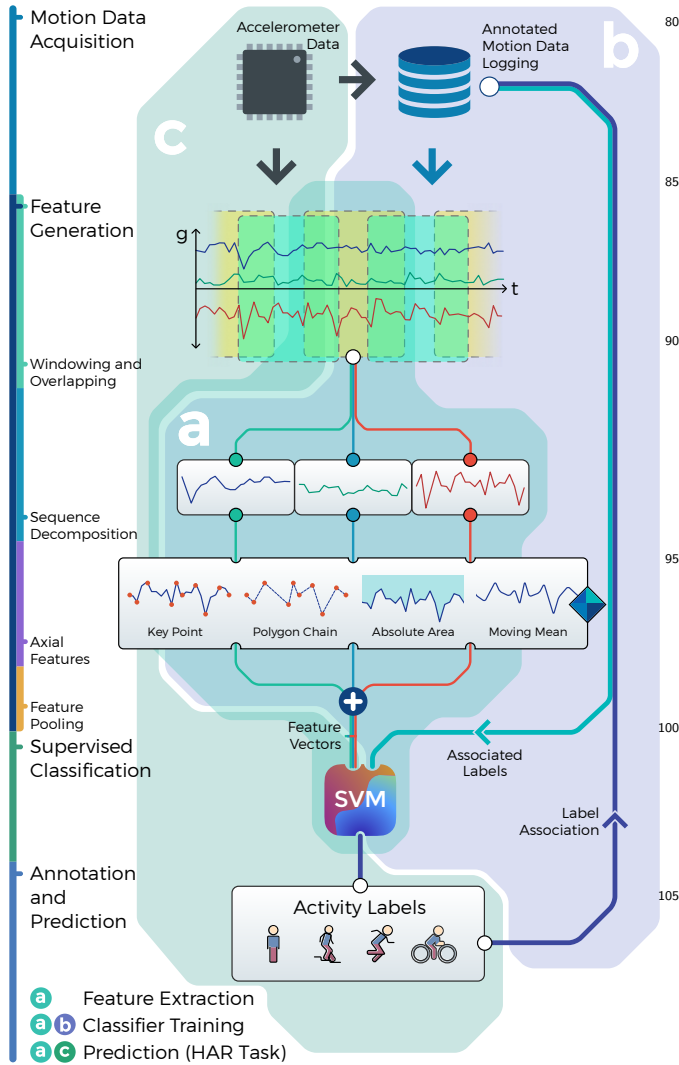


Figure 1: Methodology Overview

The task of human activity recognition (HAR) begins with data collection. However, the raw data in itself is

not representative enough to be used for carrying out the task of classification. Thus, we need to prepare the data by cleaning it and carrying out various data preprocessing activities. The preprocessed data still doesn't capture the temporal nature of the task. Hence data segmentation is performed to enhance the discriminative power of the data. Feature extraction, then, converts the signals into the most relevant and powerful features which are unique for the activity at hand (Figure 1). Finally classification algorithms can be trained on the refined data, and a reliable prediction model can be created.

4.1. Acquisition of Activity Dataset

Motion data was logged through motion tracking device and an iPhone app developed for the purpose at the sampling rate of 50 Hz while multiple human subjects performed various activities as mentioned in section 4.2. The location of the devices and the duration of activities were neither controlled nor taken into consideration. The subjects were also allowed to move or use the devices during the activity. These events were also not controlled and did not have any exclusive label. This was done to ensure that the dataset reflected the real-world usage pattern. The starting and ending timestamp of the activities were used to label the logged data.

We also obtained the public dataset Reyes-Ortiz et al. (2015) and Shoaib et al. (2014) to further increase the volume of our data. Only the accelerometer data was utilised from the obtained dataset.

4.2. Activity Classes

We define following activity classes to set the scope of the HAR task.

1. Stationary
 - (a) Sitting
 - (b) Standing
2. Active
 - (a) Walking
 - i. Typical
 - ii. Stair
 - A. Walking Upstairs
 - B. Walking Downstairs
 - (b) Running
 - i. Typical
 - ii. Jogging
 - (c) Biking

4.3. Windowing and Overlapping

Due to the nature of task, it is often very difficult to represent a particular activity solely through one sample point. Hence, it is recommended to utilise a sequence of signal samples to represent the activity instances Banos et al. (2014).

In the obtained discrete-time sample sequence, we defined a window of length having l samples for all the three

axes as activity instance. The windows were also overlapped to reduce any segmentation bias. The value of l , and its effect on classification accuracy is discussed in section 6.

4.4. Feature Engineering

To reduce the computational complexity and enhance the recognition characteristics, feature generation is required. Overall 5 features were calculated for each axis. To conform to the axis constraints as defined in section 3.2, the features from the three axes were combined to get a final feature vector.

To address the time complexity constraint as defined in section 3.3, all the features are calculated in time domain, since the frequency domain conversion using discrete Fourier transform (DFT) is a computationally intensive process.

From the figure 2 it can be observed that various activities are inherently periodic in nature. The features are calculated for all the three axes of the activity instances obtained after windowing and overlapping the data.

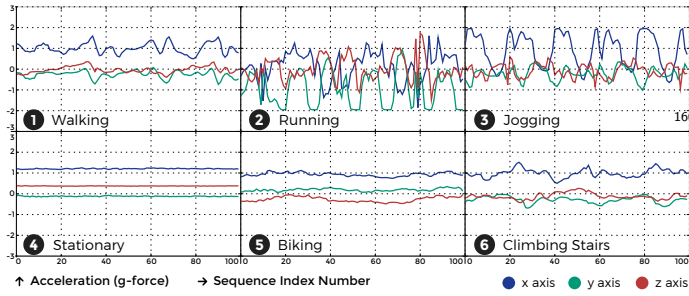


Figure 2: Example instance frames for various activities

4.5. Axial Features

Mathematically, the activity instance can be represented by a subsequence of discrete acceleration sample sequence.

Let S_{raw} be a sequence of ordered discrete-time acceleration sample sequence as obtained from the motion tracker. Here,

$$S_{raw} = (\langle a_t^x, a_t^y, a_t^z \rangle) \quad (1)$$

where a_t^x , a_t^y and a_t^z represent the accelerometer reading at instance t , in x , y and z axis respectively.

We define the activity instance S_{act} as the subsequence of S_{raw} having length l . From S_{act} , we further define decomposed axial discrete-time sample sequence by equation 2, and indexed sequence by 3.

$$S_{act}^k = (a_i^k) \quad (2)$$

$$S_{rmp}^k = (\langle i, a_i^k \rangle) \quad (3)$$

where $k \in \langle x, y, z \rangle$ and i is the sequence index.

The axial feature vector is defined in the equation 4.

$$F^k = \langle a_{auc}^k, a_{tss}^k, a_{kp}^k, a_{binnedkp}^k, a_{sm}^k \rangle \quad (4)$$

The feature vector components are discussed below.

Absolute area under linearly interpolated curve. Let $f_{[i,j]}^l$ denote the two-point form of a straight line segment joining arbitrary elements in S_{rmp}^k . Here,

$$f_{[i,j]}^l = \left(\frac{a_j^k - a_i^k}{j - i} (x - i) + a_i^k \right) \quad x \in [i, j] \quad (5)$$

We may obtain the linearly interpolated curve function f_{int} as a piecewise function of multiple line segments $f_{[i,j]}^l$. Hence, if l denotes the number of elements in sequence S_{rmp}^k , we get equation 6.

$$f_{int}^k = (f_{[i,i+1]}^l)_{i=0}^{l-1} \quad (6)$$

We take the absolute area under f_{int}^k as a feature, which is given by equation 7

$$a_{auc}^k = \int_0^l |f_{int}^k| dx \quad (7)$$

Total sum of squares. Let \bar{a}^k denote the mean of S_{act}^k . Then, we obtain the total sum of square for series S_{act}^k as:

$$a_{tss}^k = \sum_{n=0}^l (a_n^k - \bar{a}^k)^2 \quad \forall a_n^k \in S_{act}^k \quad (8)$$

The quantity obtained from equation 8 is used as a feature.

Key point. We define the key points, S_{kp}^k , a subsequence of S_{rmp}^k , where the signal sequence attains relative maximum or minimum value (extrema values).

Due to noisy nature of the signal, the relative extrema are estimated by taking n neighbours of the element currently in consideration. The quantity n is defined as the 'order' of extrema function. This is done to ensure that small fluctuations are not accounted as key points.

Key point gradient. Let

$$G_{kp}^k = \left(\frac{a_j^k - a_i^k}{j - i} \right)_{i,j} \quad (9)$$

be a sequence defined from all consecutive elements i and j in key point sequence S_{kp}^k . Here, the elements of sequence G_{kp}^k denote the gradient of a straight line joining the key points.

We take the quantity from equation 10 as a feature.

$$a_{kp}^k = \text{Var}[G_{kp}^k] \quad (10)$$

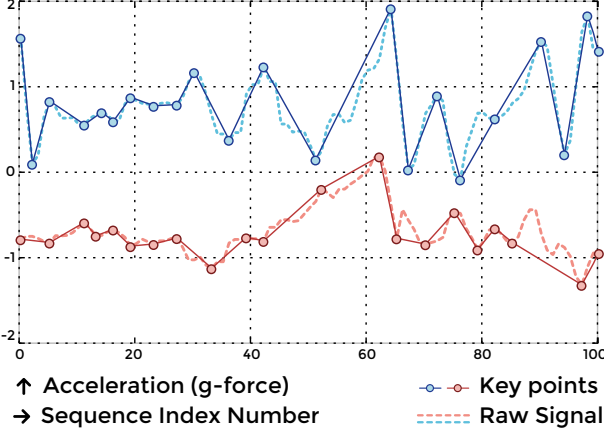


Figure 3: Sample key points selected from raw signal.

Binned key point gradient. We define gradient binning as the process of grouping straight lines on the basis of their gradient. The grouping ranges are obtained by dividing the I and IV quadrants (within the domain of arctan) of cartesian plane in, say, h equal segments. The binning ranges are defined by the following sequence:

$$G_{bin} = \left(\left(i, \left[\tan\left(\frac{\pi}{h}i\right), \tan\left(\frac{\pi}{h}(i+1)\right) \right] \right) \right)_{i=\lceil \frac{h}{2} \rceil}^{\lfloor \frac{h}{2} \rfloor} \quad i \in \mathbb{Z} \quad (11)$$

The binning function f_{bin} follows from G_{bin} where, the gradients are given index number i , if they fall in corresponding range.

Using f_{bin} , we define equation 12, where, a_i are the elements in G_{kp}^k .

$$G_{binnedkp}^k = (f_{bin}(a_i)) \quad (12)$$

We take the following quantity as a feature:

$$a_{binnedkp}^k = Var[G_{binnedkp}^k] \quad (13)$$

Variance of smoothened series. We use a simple rolling average on S_{act}^k to smooth out the short-term fluctuations in the sequence. We form a new sequence S_{smooth}^k having length $l - n$, where l is the window length and n is the order of rolling average, from S_{act}^k . For the elements e in S_{act}^k , we have

$$S_{smooth}^k = \left(\frac{e_m + e_{m-1} + \dots + e_{m-(n-1)}}{n} \right)_{m=0}^{l-n} \quad (14)$$

The following quantity is taken as a feature.

$$a_{sm}^k = Var[S_{smooth}^k] \quad (15)$$

4.6. Axial Feature Pooling and Feature Vector

To conform to the Axis Constraint, as defined in section 3.2, the axial feature vectors F^x , F^y , and F^z are combined (pooling) to obtain an axis independent feature vector.

The variance features, $\{a_{kp}^k, a_{binnedkp}^k, a_{sm}^k\}$ are pooled by taking the combined variance. We take arithmetic mean of feature $\{a_{auc}^k, a_{tss}^k\}$. Here, k denotes the three orthogonal axes x , y , and z as mentioned in section 4.5.

4.7. Classification using Supervised Learning

Various supervised learning schemes were utilised for the motion classification. The classifiers were trained using the labelled feature vectors and the resulting classifier was then used for cross validation and prediction.

4.8. Realtime Activity Prediction

To implement the realtime activity classification, the triaxial accelerometer is sampled at a sampling rate of 50 Hz. These samples are stored in a buffer array and then the feature vector is calculated for this buffer array. The predicted activity label for this feature vector is obtained by fitting the vector into the trained classifiers.

5. Feature Classification

Since the HAR task requires that multiple activities are classified, a multi-class classifier is required. We explored various multi-class supervised-learning methods to evaluate the classification performance of the features as discussed in section 4.5. The methods that were evaluated are: Decision Tree Classifier (DTC), Random Forest Classifier (RFC), and Support Vector Machine (SVM) with RBF Kernel with varying parameters.

The performance as well as the relevance of the features are discussed in section 6.

6. Results and Discussions

6.1. Results

After calculating the features for the windows, and assigning the labels, we split the dataset into two segments with random cuts. First segment, having 75% the size of the dataset, was used for training the classifiers. The other segment, representing the 25% of the dataset, was used for testing purpose. The same data-subset was used for evaluating all the classifiers as mentioned in section 5.

Figure 4 contains the classification accuracy obtained by each classifier with respect to the chosen window length. Using the dataset as mentioned in section 6, we obtained a maximum classification accuracy of 96.32% using Random Forest Classifier closely followed by Support Vector Machine with 95.05% accuracy (Figure 5).

6.2. Discussions

Human activities are performed during relatively long periods of time compared to the sensors sampling rate. Besides, a single sample on a specific time instant does not provide sufficient information to describe the performed activity. Thus, activities need to be recognized in a time window basis rather than in a sample basis. Now, the question is: how do we compare two given time windows? It would be nearly impossible for the signals to be exactly identical, even if they come from the same subject performing the same activity. This is the main motivation for applying feature extraction methodologies Guyon and Elisseeff (2003) Chen and Lin (2006) to each time window: filtering relevant information and obtaining quantitative measures that allow signals to be compared. In general, two approaches have been proposed to extract features from time series data: statistical and structural. In our study, we have used a combination of both statistical and structural features.

In addition to efficient feature extraction, selection of window length that is dividing the measured time series in time windows is also crucial in increasing the classification accuracy. A key factor is, therefore, the selection of the window length because the computational complexity of any feature extraction method depends on the number of samples. Having rather short windows may enhance the feature extraction performance, but would entail higher overhead due to the recognition algorithm being triggered more frequently. Besides, short time windows may not provide sufficient information to fully describe the performed activity. Conversely, if the windows are too long, there might be more than one activity within a single time window. Different window lengths have been used in the study but the maximum classification accuracy is observed with window length of 2s. Of course, this decision is conditioned to the activities to be recognized and the measured attributes.

Figure 6 is the parallel coordinate visualisation Wegman (1990) of 50 feature vectors, chosen randomly, for each activity class as described in section 4.2. We used parallel coordinates to plot the components in order as presented in figure. The independence between individual features is clearly distinguished for the activity classes.

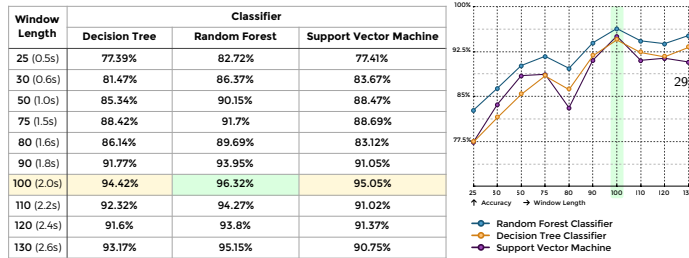


Figure 4: Comparative classification accuracy for varying window sizes.

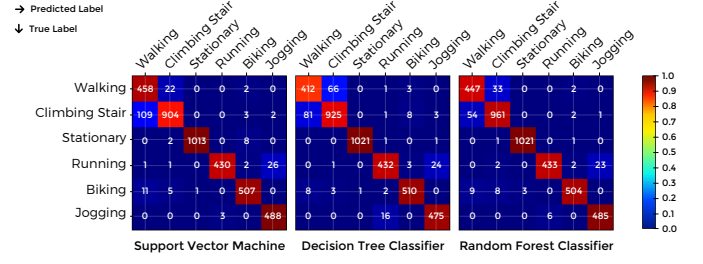


Figure 5: Confusion Matrix for window length $l = 100$.

The feature vectors are plotted in both Linear and Logarithmic scale to improve visibility of features that attain value close to zero, since the features are not normalised.

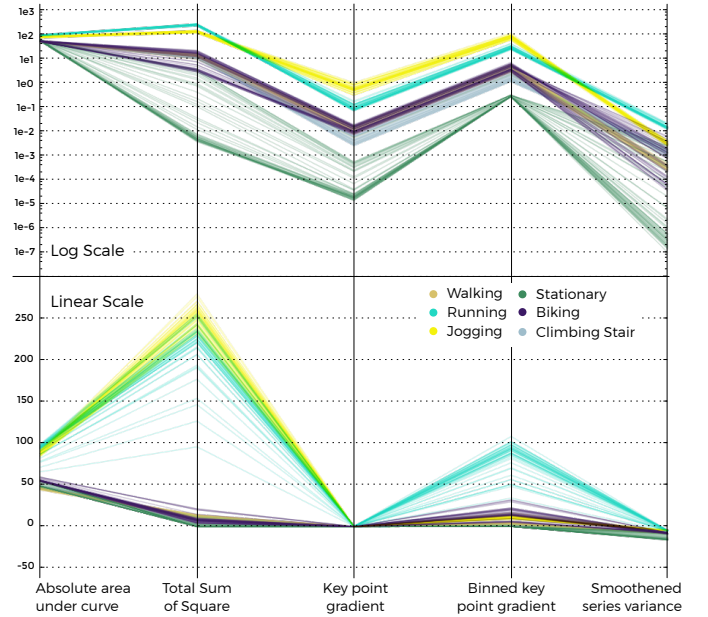


Figure 6: Parallel coordinates plot of feature vectors.

Similarly, figure 7 is the RadViz Visualisation Shargo et al. (2008) Bertini (2005) of 400 feature vectors chosen randomly for each activity classes. Distinct clusters of activity classes further support the conclusion made from the Parallel Coordinate plot of feature vector. Confusion between similar activities, such as Jogging and Running, and Walking and Climbing is also visible in the plot whenever the cluster of classes overlap. However, it should be noted here that only a small percentage of feature vectors overlap, as evident from section b in figure 7.

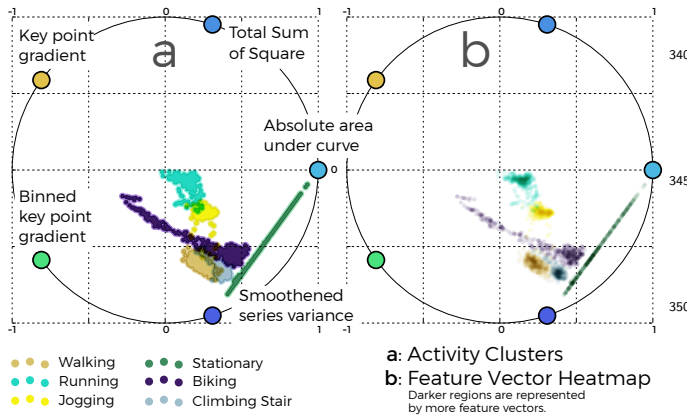


Figure 7: RadViz visualisation of feature vectors.

7. Conclusion

The focus of this research was to explore and analyze flexibility in various constraints, specifically the position, orientation and attachment of the mobile sensor device in a human activity recognition task. The results from our study provide evidence to the fact that activity recognition is possible in spite of variation in the aforementioned constraints. Our findings also indicate that neither device orientation information nor orientation correction appear to be necessary, as we are able to correctly classify the activities across all subjects without orientation correction. Moreover, by using simple features, we can achieve high classification accuracies, above 95% irrespective of how the device is carried and placed by the subject. This method is suitable for a distributed system where users can report data, in real-time or in batches, and this information can be processed and used to act according to the individual context.

The results of the present research have been submitted for publication in International Journal of Human-Computer Studies.

References

- Anguita, D., Ghio, A., Oneto, L., Parra, X., Reyes-Ortiz, J., 2013. Energy Efficient Smartphone-Based Activity Recognition using Fixed-Point Arithmetic. J. UCS URL: <http://www.jucs.org/jucs/19/9/energy-efficient-smartphone-based-jucs/19/09/1295/1314/anguita.pdf>.
- Banos, O., Galvez, J., Damas, M., Pomares, H., Rojas, I., 2014. Window size impact in human activity recognition. Sensors URL: <http://www.mdpi.com/1424-8220/14/4/6474/htm>.
- Bertini, E., 2005. Springview: Cooperation of radviz and parallel coordinates for view optimization and clutter reduction. Coordinated and Multiple Views in Exploratory Visualization URL: http://ieeexplore.ieee.org/xpls/abs/_all.jsp?arnumber=1508218.
- Bradie, . A Friendly Introduction to Numerical Analysis. Pearson Education. URL: https://books.google.co.in/books/about/A_Friendly_Introduction_to_Numerical_Analysis.html?id=aaaayS9azQUC&pgis=1.
- Chen, Y., Lin, C., 2006. Combining SVMs with various feature selection strategies. Feature extraction URL: http://link.springer.com/chapter/10.1007/978-3-540-35488-8_13.

- Gupta, P., Dallas, T., 2014. Feature selection and activity recognition system using a single triaxial accelerometer. Biomedical Engineering, IEEE Transactions URL: http://ieeexplore.ieee.org/xpls/abs/_all.jsp?arnumber=6780615.
- Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. The Journal of Machine Learning Research URL: <http://dl.acm.org/citation.cfm?id=944968>.
- He, Z., Jin, L., 2009. Activity recognition from acceleration data based on discrete cosine transform and SVM. Systems, Man and Cybernetics, 2009. SMC 2009 URL: http://ieeexplore.ieee.org/xpls/abs/_all.jsp?arnumber=5346042.
- Reyes-Ortiz, J., Oneto, L., Samà, A., Parra, X., Anguita, D., 2015. Transition-Aware Human Activity Recognition Using Smartphones. Neurocomputing URL: <http://www.sciencedirect.com/science/article/pii/S0925231215010930>.
- Sharko, J., Grinstein, G., Marx, K., 2008. Vectorized radviz and its application to multiple cluster datasets. Visualization and Computer Graphics URL: http://ieeexplore.ieee.org/xpls/abs/_all.jsp?arnumber=4658161.
- Shoaib, M., Bosch, S., Incel, O.D., Scholten, H., Havinga, P.J.M., 2014. Fusion of smartphone motion sensors for physical activity recognition. Sensors (Basel, Switzerland) 14, 10146–76. URL: <http://www.mdpi.com/1424-8220/14/6/10146/htm>, doi:10.3390/s140610146.
- Shoaib, M., Scholten, J., Havinga, P., 2013. Towards physical activity recognition using smartphone sensors. URL: <http://eprints.eemcs.utwente.nl/23843/>.
- Wegman, E., 1990. Hyperdimensional data analysis using parallel coordinates. Journal of the American Statistical Association URL: <http://www.tandfonline.com/doi/abs/10.1080/01621459.1990.10474926>.