Autoregressive Modeling of Wrist Attitude for Feature Enrichment in Human Activity Recognition

Priscila L.R. Aguirre[1], Leonardo A.B. Torres[2], and André P. Lemos[2]

[1] Graduate Program in Electrical Engineering - Universidade Federal de Minas Gerais - Av. Antônio Carlos 6627, 31270-901, Belo Horizonte, MG, Brazil
priaguirre@gmail.com,
[2] Dep. de Engenharia Eletrônica - Universidade Federal de Minas Gerais - Av. Antônio Carlos 6627, 31270-901, Belo Horizonte, MG, Brazil
leotorres@ufmg.br, andrepl@cpdee.ufmg.br

Abstract. The use of time-series from wrist worn accelerometers for Human Activity Recognition is investigated in this work. We employ, as features, coefficients of two-dimensional multivariate/vector autoregressive (AR) models obtained from raw acceleration signals and from estimated wrist attitude roll and pitch angles. It is shown that the simultaneous use of both types of models improves the overall accuracy about 20% when compared to recently published algorithms where only univariate AR models coefficients for each raw acceleration signal are employed.

Keywords: human activity recognition, wrist attitude, autoregressive models, SVM

1 Introduction

Human Activity Recognition (HAR) has sprouted, in the past couple of decades, as a field of high interest for applications from surveillance systems to patient monitoring systems (lifecare and health care). In such contexts, the automatic identification of activities of daily living has become relevant.

This field, which aims at extracting knowledge from sensor data, has two main data sources: video based information [13], and non-visual sensors' signals [12]. Even though the use of video has been widely researched, the sensors approach seem to be more promising, since it is less intrusive and associated processing techniques tend to require less computational power. Besides, it has been usual for people to carry devices, such as mobile phones and smartwatches, that have embedded various non-visual sensors, rendering the sensors data collection a potentially ubiquitous possibility [15].

Measured attributes can be usually related to user movements, environmental or physiological signals. Even though there might be some sensor fusion to improve recognition performance, tri-axial accelerometers signals have been the commonly used sensor data for ambulation activities (e.g. walk, sit, climb stairs). However, it has been reported that for some daily activities (e.g. eat, drink, brush teeth) the use of raw acceleration signals might not be the best option because these activities are rich in arm movements [12]. It is worth noticing that these more complex activities are usually not considered in many works on HAR based on accelerometer signals [8, 11, 9].

On the other hand, researchers have reported in [14, 3] possible improvements in the overall accuracy when working with features that depend on attitude estimation of wrist-worn devices. Following this approach, our research considers the feature extraction problem from wrist positioned accelerometer signals for recognition of daily activities that are rich in arm movements, by relying not only on raw acceleration signals, but also on estimated attitude angles.

The rest of this paper is organized as follows. In section 2, related work is considered. In section 3 presents the main goals and contributions of this paper. Section 4 presents the mathematical and physical background for the considered feature extraction and Section 5 describes implementation and experimental results. We discuss main contributions in Section 6.

2 Related Work

As shown in [15], accelerometers have been the most common sensors employed in human activity recognition problems using wearable sensors. The information extracted from sensor data is of two main types: time domain (TD) and frequency domain (FD) features. The most common TD features are mean, standard deviation, energy, and correlation between axes. Fast Fourier Transform (FFT) and Wavelet Transform coefficients are the most common FD features. TD features are most frequently used in real-time applications since they depend on shorter time windows and are computationally cheaper [15].

Alternatively, the use of coefficients from Autoregressive (AR) models have been considered in previous works as features for HAR problems [4, 8]. AR models can be used as powerful one-step ahead predictors, and they can be considered as mixed TD and FD representations for the dynamic behavior of a stochastic process.

[4] shows that better performance in activity recognition was attained when using AR coefficients over traditional TD features. The authors in [8] go further and present improvements by augmenting the feature vector with *signal magnitude area* (SMA) and the *tilt angle* (TA) defined as the angle between the z-axis and the gravitational vector $\boldsymbol{g}$. Their idea is that SMA will enable distinguishing dynamic and static activities (e.g. standing and walk), and TA allows distinguishing postures (e.g. standing and lying).

An extra step before classification is added in [9, 11] in order to reduce large within-class and low between-class variance and therefore improve separability. High within-class variance can be related to lack of firm attachment of the wearable accelerometer and its different positioning, as in the case of embedded sensors in phones that might be in shirt pockets, trousers front or back pockets, facing upwards, facing downwards, etc. Thus, linear discriminant analysis (LDA) is applied to the feature vector composed of AR coefficients and SMA so that recognition can be independent of accelerometer position. However, kernel discriminant analysis (KDA) proved to be more effective in reducing within-class variance, as shown in [7]. Another extra step in feature extraction was also applied in [5], where raw accelerometer signals were decomposed by computing Wavelet Transform prior to AR model building. In a work by Khan and co-workers [10] a complete smartphone-based activity recognition system is presented, relying on AR models coefficients extraction from accelerometer signals followed by the use of KDA.

In these previous works where AR models coefficients were considered as features to perform HAR, data is collected from one single tri-axial accelerometer either strapped to the user's chest or in a trouser pocket, and ambulation type activities were investigated (such as running, walking, climbing up stairs, going down stairs, standing still and jumping). More complex daily activities, such as eating, are not considered in such researches. With respect to classifiers, authors in [8, 9, 11, 5] have used artificial neural networks (ANN), while [4, 5] have opted for SVM following a one-versus-one approach.

3 Statement of the Contributions

Considering the lack of works that actually deal with complex daily activities, that are rich in arm movements, we investigate in this paper if HAR accuracy can be improved by describing these movements with the time evolution of the wrist *pitch* and *roll* attitude Euler angles using AR models coefficients. This can be considered as the application of prior nonlinear transformations to low-pass filtered acceleration signals aiming at the extraction of relevant information for HAR. The attitude angles resemble the TA variable proposed in [8], but they are a more complete description of the wrist attitude.

By representing the attitude angles in a multivariate/vector autoregressive model, instead of separate autoregressive models as done in previous works, we consider our system coupled and we are able to quantify the correlation between variables, which increases the level of information that can be used in HAR.

Finally, we present measurements of variability of our main results. Since the number of data trials is not high and the classifiers depend on training examples randomly chosen, one single result may be biased, if such variability factors are not taken into account.

4 Methodology

In this work, feature extraction is done in two stages: 1) attitude estimation based on Euler angles and 2) autoregressive modeling of estimated attitude signals.

4.1 Attitude Estimation

The measured acceleration can be represented as $\boldsymbol{a}_{\mathrm{m}} = \boldsymbol{a}_{\mathrm{t}} - \boldsymbol{g} + \boldsymbol{\nu}$, where $\boldsymbol{a}_{\mathrm{m}} = [a_{\mathrm{x}}\, a_{\mathrm{y}}\, a_{\mathrm{z}}]^{\top}$; $\boldsymbol{a}_{\mathrm{t}} \in \mathbb{R}^3$ is the translational acceleration; $\boldsymbol{g} \in \mathbb{R}^3$ is the gravity acceleration vector ($\|g\| = g_0 = 9.806\,\mathrm{m/s}^2$); and $\boldsymbol{\nu} \in \mathbb{R}^3$ represents additive zero mean measurement noise.

The next step, assuming that the contribution of the wrist translational acceleration is much smaller than that of the local gravity, and that the level of noise is negligible, the wrist roll ϕ and pitch θ orientation angles were computed as

$$\begin{aligned}
\phi &= \operatorname{atan2}\left(-a_{\mathrm{y}}, -a_{\mathrm{z}}\right), \\
\theta &= -\operatorname{asin}\left(-a_{\mathrm{x}}, \sqrt{(a_{\mathrm{x}})^2 + (a_{\mathrm{y}})^2 + (a_{\mathrm{z}})^2}\right).
\end{aligned} \tag{1}$$

Notice that, differently of what is commonly performed in previous works, the acceleration data was not filtered before computing the attitude angles. The reason is that this helps in providing numerically well-conditioned data to perform least squares based parameter estimation, as described below.

4.2 Autoregressive Model Representation

To mathematically represent sequences of estimated attitude angles that could be associated to the execution of specific human activities, we consider $\mathbf{y}(k) = [\phi(k)\ \theta(k)]^{\top}$, and we find an AR model representation for the time evolution of the attitude angles given by the following *vector autoregressive model* (VAR):

$$\mathbf{y}(k) = A_1 \mathbf{y}\left(k-1\right) + \cdots + A_{n_y} \mathbf{y}\left(k - n_y\right) + \epsilon, \tag{2}$$

in which $A_i \in \mathbb{R}^{2\times 2}$ are the model coefficients; $n_y = 1, 2, \ldots$ is the autoregressive's model order; and $\epsilon \in \mathbb{R}^2$ is considered to be zero mean white Gaussian noise. This is a novelty of this work, since in previous works [8, 9, 11, 4, 5] the authors have used independent univariate AR models, one for each acceleration signal, instead of a vector/multivariate AR model that better captures the cross-correlation among signals.

The model order n_y was selected using the Akaike Information Criterion (AIC), such that $n_y = 4$ in (2) was found to be sufficient to explain each set of data, while whiteness tests for the residues has shown that they carry little information, which confirms the suitability of fourth order models [1] and that an ARMA model is not really necessary.

Due to its relative simplicity and robustness to noise, the least squares algorithm was used to estimate the model parameters in (2). These models were computed for the non-filtered data of $\phi(k)$ and $\theta(k)$ in order to avoid ill-conditioning in the computation of pseudo-inverses appearing in the least squares solution.

4.3 Signal Magnitude Area

Following the work in [8, 9, 11, 5], we also consider another component in the feature vector, namely the so-called Signal Magnitude Area, defined as

$$SMA = \frac{1}{N}\sum_{i=1}^{N}|x(i)| + |y(i)| + |z(i)| \tag{3}$$

for the tri-axial acceleration signals.

4.4 Building Feature Vectors

Features are therefore extracted from (2) and (3), and combined in order to compare the use of attitude angles for recognition of rich wrist movements.

Three different scenarios were considered: 1) feature vector composed of AR coefficients built from acceleration signal and SMA, aiming to reproduce results from [8]; 2) feature vector composed of VAR coefficients built from estimated attitude signals; and 3) feature vector composed of VAR coefficients built from acceleration signals, VAR coefficients built from estimated attitude signals and SMA, as depicted in Figure 1.

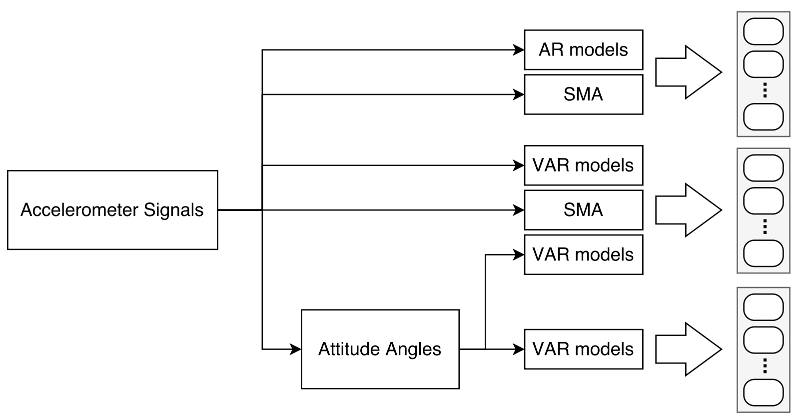


Fig. 1: Schematic describing the three feature vector components tested

5 Experimental Results

5.1 Dataset

The dataset used in this work was first presented in [2], and subsequently made available by the authors. The *Wearable Human Activity Recognition Folder*

(WHARF) dataset is a public collection of labeled accelerometer recordings of different Human Motion Primitives (HMP), which were defined as movements that describe an activity of daily living (ADL).

In this dataset tri-axial accelerometer measurements ranges from -1.5g to +1.5g, with 6 bits of digital resolution per axis and sampling rate of 32 Hz. The accelerometers were embedded in an ad-hoc sensing device attached to the right wrist of each one of the 17 volunteers (11 male, with age ranging from 19 to 81 years; and 6 female, with ages between 56 and 85 years), with x-axis pointing towards the hand, y-axis towards the left, and z-axis perpendicular to the hand's palm. Data transmission to the PC was wired, via a USB cable. We have considered a subset of the activities in this dataset, as listed in Table 1, ranging from ambulation activities (Lie down, Stand up and Walk) to activities that depend on arm movements (Comb hair, Drink glass and Pour water).

Table 1: Motion primitives from WHARFs's dataset

Motion primitives	Number of trials
Comb hair	31
Drink glass	115
Lie down	28
Pour water	100
Stand up	112
Walk	100

5.2 Feature Extraction

Sequences of roll and pitch were extracted by applying equations (1) to accelerometer measurements for each trial from the dataset. The number of samples in each trial ranges from 200 to 1200, depending on the activity carried out. Then, parameters of model (2) were estimated by least squares algorithm [1] in batch mode for each trial. No windowing was applied to the signals.

Since $A_i \in \mathbb{R}^{2\times 2}$ in (2), and having defined $n_y = 4$, each trial renders a VAR model with 16 estimated parameters, which are used to compose the feature vectors as described in Section 4.4.

Differently from the VAR case, but following the same approach in [4, 8, 11, 9] for comparison purposes, three independent third order AR models were also identified for each accelerometer signal. SMA values for accelerometer signals were also computed to complete the feature vector, as illustrated in Figure 1.

5.3 Classification

For our supervised problem, we opted to use linear *support vector machines* (SVMs) in the classification stage, due to robustness and requiring tuning of just one hyper-parameter, the penalty factor C.

The multi-class problem was investigated by applying the *one-against-one* approach, in which a classifier is trained for each pair of classes and the final decision is made by a voting system for all pairwise classifiers.

5.4 Results

The process of classification was run 50 times in a Monte Carlo loop, in which input feature vectors were randomly reshuffled at each iteration and equally split into two sets: 70% for training, and the remaining 30% for testing. With this methodology we were able to estimate average performance measures and quantify their variability. At each resampling, the number of trials for each activity, both for training and testing, was kept the same. Data imbalance, however, was not treated.

In order to increase performance and avoid overfitting, values were assigned to hyper-parameter C by grid-search, following recommendations in [6]. For each of the three scenarios depicted in Figure 1, such procedure was applied, rendering minimum test error rates at $C = 2^{13.5}$ for the accelerometer feature vector, $C = 2^{12}$ for the attitude feature vector, as shown in Figure 2, and $C = 2^{8}$ when combining both feature vectors.

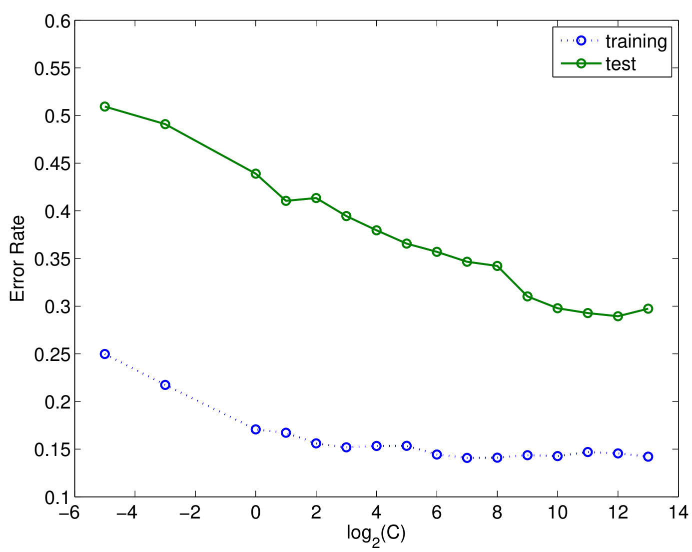


Fig. 2: Average training and test error rate for attitude feature vector

Results are expressed as the average of true positive rate (TPR) (also called *recall*), which measures the proportion of correctly labeled cases in a given class, and the corresponding standard deviation. Total accuracy figures, considering all properly labeled cases, were also computed. Results for the first feature vector are presented in Table 2. When using the second type of features incorporating attitude angles, improvements could be observed in activities rich in arm movements such as drink glass and pour water, as shown in Table 3. However, a much worse result was found for the comb hair activity.

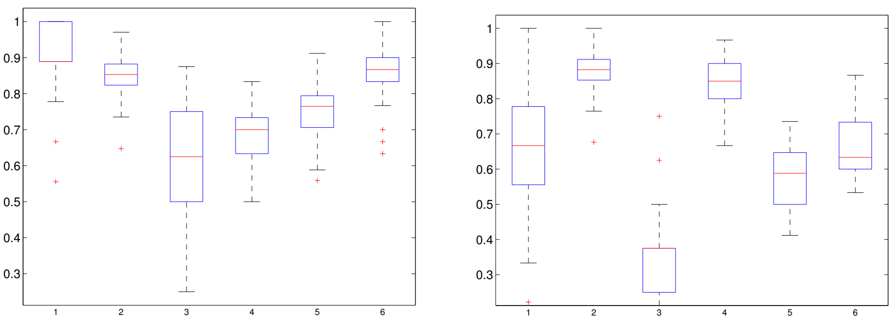
Table 2: Average true positive rate (TPR) (%), related standard deviation (σ_{TPR}) (%) and total accuracy for AR models of acceleration signals and SMA as features.

Activities	TPR	σ_{TPR} (%)
Comb hair	89.11	12.67
Drink	85.06	6.08
Lie down	61.75	15.89
Pour water	69.20	6.95
Stand up	75.29	7.51
Walk	85.33	7.02
Total Accuracy	78.51 ± 2.96	

Table 3: Average true positive rate (TPR) (%), related standard deviation (σ_{TPR}) (%) and total accuracy for VAR models of attitude signals as features.

Activities	TPR	σ_{TPR} (%)
Comb hair	68.44	18.37
Drink	87.94	5.91
Lie down	34	16.40
Pour water	83.33	7.18
Stand up	58.29	7.95
Walk	66.73	9.08
Total Accuracy	71.46 ± 3.76	

The TPR results in Tables 2 and 3 are also presented as box plots in Figure 3, where one can notice the striking low averaged TPR associated with the comb hair activity (activity number 1), when using attitude angles based features. This indicates that the VAR models parameters were not sufficiently discriminative in this case. By investigating the time series of attitude angles, we noticed approximately periodic dynamics for each volunteer. We conjecture that the linear structure of our VAR model should be changed to better capture this oscillatory behavior.



(a) acceleration features (b) attitude features

Fig. 3: Box plots for TPR results presented in Tables 2 and 3, respectively. Groups one through six correspond, respectively, to activities *comb hair*, *drink*, *lie down*, *pour water*, *stand up* and *walk*.

Finally, the simultaneous use of VAR models coefficients for attitude angles together with VAR models for the acceleration signals increased considerably the true positive rate average, and it has also reduced classification variability as observed in Table 4 and Figure 4. Part of this improvement is related to the fact that multivariate/vector AR models take into account the coupling among all three acceleration signals. It is important to notice that even for the comb hair activity, which had a low TPR rate when used alone in the classification stage, as shown in Table 3, the simultaneous use of attitude and acceleration based features was effective in increasing the performance of the original classification procedure, i.e. apparently the information added by the attitude angles was indeed new.

6 Conclusions

In this work, we have considered feature enrichment for human activity recognition from tri-axial accelerometer sensor data by computing attitude angles and corresponding parameters of vector/multivariate autoregressive (VAR) models. The underlying assumption is that these models are able to correctly capture the time evolution of the measured and computed signals.

The use of VAR coefficients obtained from the estimated attitude angles in fact has led to improvements in the recognition rate of activities that are rich in

Table 4: Average true positive rate (TPR) (%), related standard deviation (σ_{TPR}) (%) and total accuracy for VAR models of acceleration signals and attitude signals and SMA combined as features.

Activities	TPR	σ_{TPR} (%)
Comb hair	99.56	2.17
Drink	82.71	7.43
Lie down	60	15.41
Pour water	84.80	6.84
Stand up	60.94	8.11
Walk	91.80	4.01
Total Accuracy	79.71 ± 2.88	

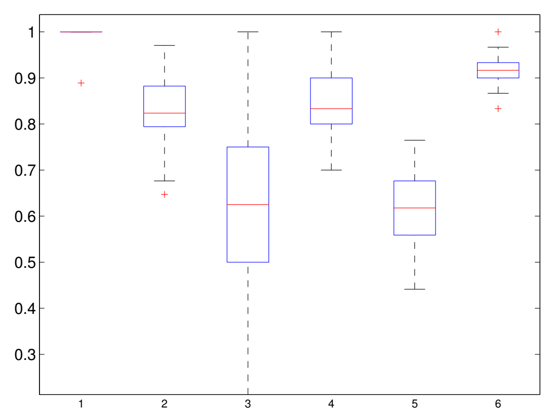


Fig. 4: Box plot for TPR results presented in Table 4. Groups one through six correspond, respectively, to activities *comb hair*, *drink*, *lie down*, *pour water*, *stand up* and *walk*.

arm movements, such as "drink glass" and "pour water". These features alone, however, were found to be inappropriate for full body motion activities, in the sense that the direct use of AR models for the acceleration signals were more effective. On the other hand, by combining accelerometer signals and attitude angles we were able to improve the recognition accuracy for each activity, while reducing the associated performance variability of the linear SVM classifier.

In addition, the use of VAR models for the acceleration signals, instead of multiple univariate AR models, as it is usually considered in the recent literature, seems to be valuable to improve the overall accuracy by capturing the cross-correlation among signals during the execution of the activities.

Notes and Comments. Part of the results presented in this paper were obtained through research on a project titled "HAR-HEALTH: Reconhecimento de Atividades Humanas associadas a Doenças Crônicas", sponsored by Samsung Eletrônica da Amazônia Ltda. under the terms of Brazilian federal law No. 8.248/91. This work has been supported by the Brazilian agency CAPES.

References

1. Aguirre, L.A.: In: UFMG, E. (ed.) Introdução a Identificação de Sistemas: Técnicas Lineares e Não Lineares: Teoria e Aplicação, chap. 5, pp. 221–252 (2015)
2. Bruno, B., Mastrogiovanni, F., Sgorbissa, A., Vernazza, T., Zaccaria, R.: Human motion modelling and recognition: A computational approach. In: Automation Science and Engineering (CASE), 2012 IEEE International Conference on. pp. 156–161. IEEE (2012)
3. Dong, Y., Hoover, A., Scisco, J., Muth, E.: A new method for measuring meal intake in humans via automated wrist motion tracking. Applied psychophysiology and biofeedback 37(3), 205–215 (2012)
4. He, Z.Y., Jin, L.W.: Activity recognition from acceleration data using ar model representation and svm. In: Machine Learning and Cybernetics, 2008 International Conference on. vol. 4, pp. 2245–2250. IEEE (2008)
5. He, Z.: Activity recognition from accelerometer signals based on wavelet-ar model. In: Progress in Informatics and Computing (PIC), 2010 IEEE International Conference on. vol. 1, pp. 499–502. IEEE (2010)
6. Hsu, C.W., Chang, C.C., Lin, C.J.: A practical guide to support vector classification (May 2016)
7. Khan, A.M., Lee, Y.K., Lee, S., Kim, T.S.: Human activity recognition via an accelerometer-enabled-smartphone using kernel discriminant analysis. In: Future Information Technology (FutureTech), 2010 5th International Conference on. pp. 1–6. IEEE (2010)
8. Khan, A.M., Lee, Y.K., Kim, T.S.: Accelerometer signal-based human activity recognition using augmented autoregressive model coefficients and artificial neural nets. In: Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE. pp. 5172–5175. IEEE (2008)
9. Khan, A.M., Lee, Y.K., Lee, S.Y., Kim, T.S.: A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer. IEEE transactions on information technology in biomedicine 14(5), 1166–1172 (2010)

10. Khan, A.M., Siddiqi, M.H., Lee, S.W.: Exploratory data analysis of acceleration signals to select light-weight and accurate features for real-time activity recognition on smartphones. Sensors 13(10), 13099–13122 (2013)
11. Khan, A., Lee, Y., Lee, S.: Accelerometer's position free human activity recognition using a hierarchical recognition model. In: e-Health Networking Applications and Services (Healthcom), 2010 12th IEEE International Conference on. pp. 296–301. IEEE (2010)
12. Lara, O.D., Labrador, M.A.: A survey on human activity recognition using wearable sensors. IEEE Communications Surveys and Tutorials 15(3), 1192–1209 (2013)
13. Poppe, R.: A survey on vision-based human action recognition. Image and Vision Computing 28(6), 976–990 (2010)
14. Ramos-Garcia, R.I., Muth, E.R., Gowdy, J.N., Hoover, A.W.: Improving the recognition of eating gestures using intergesture sequential dependencies. IEEE journal of biomedical and health informatics 19(3), 825–831 (2015)
15. Shoaib, M., Bosch, S., Incel, O.D., Scholten, H., Havinga, P.J.: A survey of online activity recognition using mobile phones. Sensors 15(1), 2059–2085 (2015)