

Feature Analysis to Human Activity Recognition

J. Suto, S. Oniga, P. Pop Sitar

Jozsef Suto*

Department of Information Systems and Networks
University of Debrecen, Debrecen, Hungary

*Corresponding author: suto.jozsef@inf.unideb.hu

Stefan Oniga

1. Department of Information Systems and Networks
University of Debrecen, Debrecen, Hungary
oniga.istvan@inf.unideb.hu
2. Department of Electronic and Computer Engineering
Technical University of Cluj-Napoca,
North University Center at Baia Mare, Baia Mare, Romania
stefan.oniga@cunbm.utcluj.ro

Petrica Pop Sitar

Department of Mathematics and Informatics
Technical University of Cluj-Napoca,
North University Center at Baia Mare, Baia Mare, Romania
petrica.pop@cunbm.utcluj.ro

Abstract: Human activity recognition (HAR) is one of those research areas whose importance and popularity have notably increased in recent years. HAR can be seen as a general machine learning problem which requires feature extraction and feature selection. In previous articles different features were extracted from time, frequency and wavelet domains for HAR but it is not clear that, how to determine the best feature combination which maximizes the performance of a machine learning algorithm. The aim of this paper is to present the most relevant feature extraction methods in HAR and to compare them with widely-used filter and wrapper feature selection algorithms. This work is an extended version of [1]^a where we tested the efficiency of filter and wrapper feature selection algorithms in combination with artificial neural networks. In this paper the efficiency of selected features has been investigated on more machine learning algorithms (feed-forward artificial neural network, k-nearest neighbor and decision tree) where an independent database was the data source. The result demonstrates that machine learning in combination with feature selection can overcome other classification approaches.

Keywords: human activity recognition, feature extraction, feature selection, machine learning.

^aReprinted and extended, with permission based on License Number 3958150787732 [2016] IEEE, from "Computers Communications and Control (ICCCC), 2016 6th International Conference on"

1 Introduction

This paper is an extended version of our previous work where only artificial neural networks have been used to the efficiency investigation of feature selection in the human activity recognition (HAR) problem [1]. However, in this paper by the involvement of more machine learning techniques we can examine the relation between classifier and feature selection methods.

The appearance of data mining was a milestone of modern biomedical applications. HAR is an interesting and rapidly expanding part of this area. In this type of problem, we want to determine

the activity of people from the information which comes from one or more accelerometer-based data collector devices. Generally, sensors are placed to different parts of the body and provide information about the functional ability and lifestyle of an observed person. Although many articles have been presented in this topic, some questions are unanswered yet. *Which feature selection algorithm generate the best feature combination that maximize the recognition rate of a machine learning algorithm? Does a general feature combination which similarly efficient independently of the person exist? Can a machine learning algorithm in combination with feature selection overcome other classification approaches?.* The aim of this study is to give reliable answers for the above questions.

The rapidly growing rate of elderly population in our society has a huge impact to health care systems. Obviously, everyone wants to stay in a familiar environment where they feel comfortable during the observation. This new challenge motivated the development of different kind of home care services, assistive systems and wearable sensor networks in order to increase the autonomy and life quality of an observed person [2,3]. Today, the miniaturized sensor technology (MEMS) makes it possible for a person to wear data acquisition devices on predetermined body segments. In the past decade, more research groups developed different kinds of devices for HAR purposes which ensure continuous observation in both indoor and outdoor environments [4]- [7]. Such a wearable sensor network was used to the construction of the WARD 1.0 database. It is a benchmark database to the HAR research which was collected at University of California, Berkeley [7]. This public data set gives an opportunity for qualitative comparison of existing HAR algorithms. The database contains information about 13 different activities which were acquired from 20 people aged between 19 and 75 years. During the data acquisition, 5 sensor nodes were placed at multiple body locations of each person: left and right forearm, waist, left and right ankle. In this study we used the data of only one sensor which has been placed on the right ankle because Ertugrul et al. and Oniga et al. demonstrated that one sensor is enough for appropriate HAR recognition while Preece et al. claimed that the ankle is an optimal placement for single sensor [8,28,31].

Beyond data acquisition, efficient algorithms are also necessary to interpret the collected data. Previous works have shown that, machine learning algorithms are efficient for human movement classification. Khan et al., Oniga et al. and Yang et al. used artificial neural network (ANN) to their HAR research and archived 97.9%, 95% and 99% recognition rates (with single sensor), respectively [9]- [12]. Duarte et al. and Preece et al. used k-nearest neighbor (kNN) method and measured 97.8% and 95% recognition rates [8,13]. Finally, Gao et al. and Maurer et al. reached similarly good results with decision tree 96.4%, 92.8% [4,5]. It was the main motivation to use those three machine learning methods in this study.

2 Raw data pre-processing and classification

2.1 Feature extraction

Many machine learning application require feature extraction and feature selection; see for example [23,25]. Feature extraction can be seen as a data pre-processing step where different kinds of features will be extracted from the raw data. In the first step the raw data (time series) will be split into short intervals - windows. Usually, a window covers one or two seconds long time interval and its size depends on the sampling frequency. For instance, Preece et al. used 2 seconds long window with 50% overlapping while Gao et al. and Karantonis et al. used 1 second long window without overlapping [4,8,14]. In our case, a window contains 32 samples and there is a 50% overlapping between windows in the training phase and no overlapping in the test phase. This size covers 1.6 seconds time interval because the sampling frequency of the

Table 1: Most common feature extraction methods in HAR

Category	Feature	Abbreviation	References
Time	Mean	M	[4, 8]
	Variance	V	[4, 8]
	Mean absolute deviation	MAD	[5, 15]
	Root mean square	RMS	[5, 15]
	Zero Crossing Rate	ZCR	[4, 5]
	Interquartile Range	IQR	[5, 15]
	75'th percentile	PE	[5, 8]
	Kurtosis	KS	[15, 16]
	Signal magnitude area	SMA	[4, 15]
	Min-max	MM	[17, 18]
Frequency	Spectral energy	SE	[4, 8]
	Spectral entropy	E	[8, 15]
	Spectral centroid	SC	[4, 19]
	Principal frequency	PF	[8, 20]
Other	Correlation between axis	CORR	[4, 17]
	Autoregressive coefficients	AR1, AR2	[9, 19]
	Tilt Angle	TA	[9, 18]

WARD database is approximately 20 Hz.

After the windowing step, features will be extracted from each window. In previous activity classification studies, the researchers used different kinds of features from different domains. Most of them come from the time a frequency domains but exist some other types of features. Table 1 summarizes the most common features from the literature and their references. Some features are redundant and those were omitted from the table. For instance, standard deviation because variance is in the table. Moreover, wavelet features have been similarly omitted because time-frequency features are more efficient than wavelets, see [8]. At the rest of this subsection we give a short description to the methods in Table 1. In the equations T indicates the window size, F is the number of frequency components, i refers to the accelerometer dimensions (x, y, z), $a_i(t)$ is an element of time series and $A_i(f)$ is a frequency component.

Mean, variance, mean absolute deviation and root mean square are statistical indicators which give information about sample distribution.

ZCR measures sing changes along the time series. In the formula, the $I\{x\}$ function returns 1 if its argument is true and 0 otherwise.

$$ZCR_i = \frac{1}{T-1} \sum_{t=1}^{T-1} I\{a_i(t)a_i(t+1) < 0\}. \quad (1)$$

Quartiles ($Q1$, $Q2$ and $Q3$) divide an ordered time series into quarters. IQR is the difference between the upper and lower quartiles: $IQR = Q3 - Q1$. It measures the spread of a data set over a range. Percentiles are similar as quartiles, except that those divide a data set into arbitrary parts (given in percentage). Therefore, the 25th percentile is equal to the lower quartile ($Q1$) and the 75th percentile is equal to the upper quartile ($Q3$). In this study we used only the 75th percentile.

Kurtosis measures the peakedness of probability distribution of collected data.

$$KS_i = \frac{\frac{1}{T} \sum_{t=1}^T (a_i(t) - M_i)^4}{\left(\frac{1}{T} \sum_{t=1}^T (a_i(t) - M_i)^2\right)^2}. \quad (2)$$

SMA equals to the normalized sum of accelerometer components.

$$SMA = \frac{1}{T} \sum_{t=1}^T |a_x(t)| + |a_y(t)| + |a_z(t)|. \quad (3)$$

Min-max is the difference between the maximum and the minimum values of time series.

Generally, signal energy is the area under the squared signal. In this case, SE measures the sum of the squared frequency components. (Since, the spectrum is symmetric, in the following three formulas F can be replaced by F/2).

$$SE_i = \sum_{f=1}^F |A_i(f)|^2. \quad (4)$$

The entropy is a measure for uncertainty. The following formula is the entropy of the normalized power spectrum density.

$$E_i = - \sum_{f=1}^F \frac{|A_i(f)|^2}{\sum_{j=1}^F |A_i(j)|^2} \log_2 \left(\frac{|A_i(f)|^2}{\sum_{j=1}^F |A_i(j)|^2} \right). \quad (5)$$

Spectral centroid measures the average frequency, weighted by the amplitude of spectrum.

$$SC_i = \frac{\sum_{f=1}^F f |A_i(f)|}{\sum_{j=1}^F |A_i(j)|}. \quad (6)$$

Principal frequency refers to the most significant frequency component which has the highest amplitude (DC component has been omitted).

$$PF_i = \max_f |A_i(f)| \quad f \neq 0. \quad (7)$$

Autoregressive model is another representation of signal. In the model, an element in the time series can be estimated by a linear weighted sum of previous elements where the weights are the coefficients. In (8), (9) and (10) p indicates model order, ϕ_k is the AR coefficient, $\varepsilon(t)$ is the noise and τ_k refers to the autocorrelation. In Table 1, AR1 and AR2 abbreviations refer to the first and second autoregressive coefficients, respectively.

$$AR(p) \longrightarrow a_i(t) = \sum_{k=1}^p \phi_k a_i(t-k) + \varepsilon(t). \quad (8)$$

$$\phi = \begin{pmatrix} 1 & r_1 & r_2 & \cdots & r_{p-1} \\ r_1 & 1 & r_1 & \cdots & r_{p-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{p-1} & r_{p-2} & r_{p-3} & \cdots & 1 \end{pmatrix}^{-1} * \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_p \end{pmatrix}. \quad (9)$$

$$r_k = \frac{\frac{1}{T} \sum_{t=1}^{T-k} (a_i(t) - M_i)(a_i(t-k) - M_i)}{V_i}. \quad (10)$$

Correlation measures linear relationship between two axes. Actually, $CORR_{i,j}$ is the cosine of angle between normalized vectors. Therefore, $-1 \leq CORR_{i,j} \leq 1$. The sing indicates the

direction of correlation. If $CORR_{i,j}$ is near to ± 1 implies that there is strong linear correlation between vectors.

$$CORR_{i,j} = \frac{\sum_{t=1}^T (a_i(t) - M_i)(a_j(t) - M_j)}{\sqrt{\sum_{t=1}^T (a_i(t) - M_i)^2 (a_j(t) - M_j)^2}}. \quad (11)$$

Finally, tilt angle indicates the relative tilt of the sensor. It is the angle between z-axis and gravitational vector. In this survey, the absolute values of tilt angles have been summarized.

$$TA = \arccos\left(\frac{a_z(t)}{\sqrt{a_x(t)^2 + a_y(t)^2 + a_z(t)^2}}\right). \quad (12)$$

2.2 Feature selection

Feature selection, also called feature reduction, is the process of choosing a subset of original features according to a well-defined evaluation criterion. It is a frequently used dimensionality reduction technique which removes irrelevant and redundant features. This approach has more useful effects for real applications because it accelerates algorithms, improves the performance and simplifies the model. In contrast to other dimensionality reduction techniques like linear discriminant analysis (LDA) or principal component analysis (PCA) which are based on projection, feature selection does not alter the original representation of feature sets, see [21].

According to the training data which are may be tagged, untagged or partially tagged, there have been developed three categories of algorithms: supervised, unsupervised and semi-supervised feature selection where a tag refers to a given class. In addition, depending on the feature evaluation process, feature selection algorithms belong to three different groups: filter, wrapper and embedded. Filter algorithms calculate scores for all features and select features according to the score. Wrapper methods require a predefined classification technique and use its performance as ranking criteria of feature subsets. In embedded models, feature selection takes place at the training process. In this article we confine on the supervised category and particularly we are interested in filter and wrapper methods.

In real applications filter methods are frequently used for feature selection because they have some significant advantages. Firstly, those methods are independently applicable with any types of machine learning techniques. Secondly, filter methods are faster than wrappers. However, wrapper methods are more efficient than feature ranking algorithms in some cases because they take into consideration the classifier hypothesis. This also means that, wrapper techniques can handle feature dependencies. So, both types of feature selection methods have advantages and disadvantages [22]. Essentially, independently of categories and groups, the goal is to find the most appropriate hyperplane of the n-dimensional feature space in all cases where the sample distributions can be separable. For example, Fig. 1 illustrates a proper separation between six 2 dimensional sample distributions (as the colour indicates) where each distribution (samples) comes from different classes. Obviously it is an ideal case and the classification will be 100%. More information about feature selection and their application opportunities in bioinformatics can be found in [24, 25].

Today's, a reach literature exists concerning the feature (or variable) selection. During the last decades, there have been developed a large amount of filter algorithms. The proposed algorithms are based on different kinds of approaches such as statistical, information theory, rough set, etc. [23]. However, to the best of our knowledge, the number of articles which utilizes feature selection algorithm on HAR is very small. We found only two papers where the minimum redundancy maximum relevance and the correlation feature selection techniques have been utilised, see [5, 26]. Therefore, we collected the most relevant filter techniques and examined them on the HAR problem. Fortunately, Zhao et al. proposed a generally applicable repository to feature

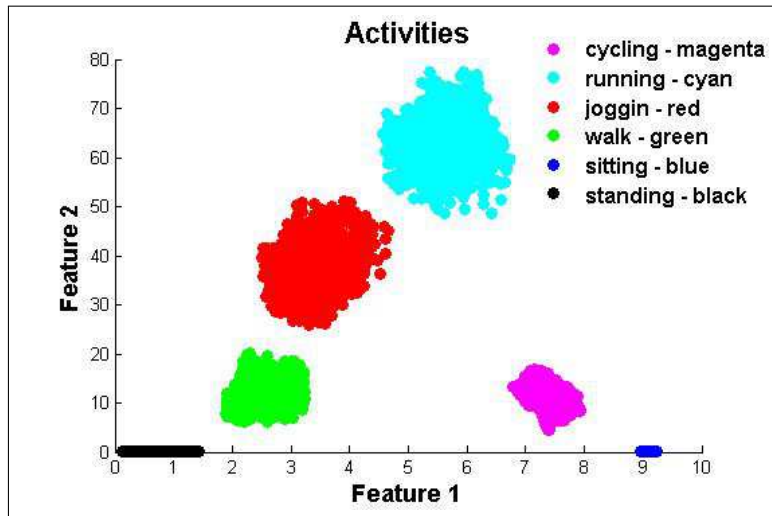


Figure 1: An ideal separation between sample distributions.

selection research which contains 13 supervised and unsupervised filter methods [27]. Moreover, the repository suggests references and implementations (in Matlab) to the algorithms which have been applied in this investigation. In addition, we similarly utilised the naive Bayesian wrapper method from [1].

So, in this work the following 9 feature selection methods have been tested: (1) Correlation feature selection (CFS); (2) Chi square (CHI); (3) Fast correlation-based filter (FCBF); (4) Fisher score (FIS); (5) Information gain (IG); (6) Kruskal-Wallis (KW); (7) Minimum redundancy maximum relevance (MRMR); (8) T-test; (9) Naive Bayesian (Bayes).

2.3 Classification

A classification system has to use an algorithm that is capable to learn and tolerate errors which come from noise. Previous studies have shown that artificial neural network (ANN), k-nearest neighbor (kNN) and decision tree (DT) are well applicable for HAR, see for example [5, 11, 13]. Therefore, to the efficiency measurement of selected features those classifiers have been applied.

As the name indicates, in feed-forward networks the input data go through all layers where the incoming data will be modified according to the weights and biases of a layer. The ANN theory gives some advices to architecture construction but general rules do not exist.

Finding the right architecture of an ANN for a specific purpose is a time-consuming task because it requires lots of simulations. Our ANN architecture design is based on the work of Oniga et al. where the authors demonstrated that a simple feed-forward ANN with only one hidden and one output layers is enough to HAR [28]. Therefore, in this study, some feed-forward ANNs were generated with the same architecture as in their work. The number of neurons on the input, hidden and output layers are equal to I , $2I$ and C respectively where I indicates the number of inputs and C is the number of activities (classes). The activation functions on the hidden and output layers were sigmoid and linear and the training algorithm was the Levenberg-Marquardt.

The kNN classifier generation is based on the work of Duarte et al. [13]. In this research, the authors reached approximately 98% recognition rate by a 1NN classifier with the Euclidean distance metrics. Finally, we used the default decision tree in Matlab without any modification.

3 Results

At the beginning of the investigation, 7 volunteers were selected from the WARD database with different and similar ages and their raw data were the input of the feature extraction step in the training and test phases: (Subject 1) 19 years old; (Subject 2) 75 years old; (Subject 3) 27 years old; (Subject 4) 29 years old; (Subject 5) 20 years old; (Subject 6) 29 years old; (Subject 7) 34 years old.

In Table 1 two features are one dimensional while others are multi-dimensional according to the axis of the accelerometer sensor. Therefore, multi-dimensional features were separated into one dimensional. This step generated 50 one dimensional features from the 17 feature extraction methods. After feature extraction, we got a feature matrix (*windows x features*) where rows contain the features of the windows. This matrix was the input of each feature selection method and the selected feature vectors were the input of the classifiers. In this study we applied the first 5 and 6 selected features because Gao et al. and Oniga et al. demonstrated that approximately 5 or 6 features are enough to an effective classification [4,28]. Finally, according to the selected features the performance of the classifiers was measured. Table 2-10 contain the selected features by the methods in selection order while the measured recognition rates (in percentage) can be seen in Table 11-17.

Table 2: Selected features by CFS

Subject	Features
Subject 1	$M_x, M_z, MAD_x, RMS_x, RMS_y, RMS_z$
Subject 2	$M_x, M_y, MAD_z, RMS_x, RMS_y, RMS_z$
Subject 3	$M_x, M_y, V_z, MAD_y, RMS_x, RMS_z$
Subject 4	$M_x, M_y, M_z, RMS_y, RMS_z, IQR_x$
Subject 5	$M_x, M_y, M_z, V_x, V_z, RMS_x$
Subject 6	$M_x, M_y, M_z, MAD_y, RMS_x, RMS_z$
Subject 8	$M_x, M_y, M_z, MAD_y, RMS_x, IQR_y$

Table 3: Selected features by CHI

Subject	Features
Subject 1	$PE_y, PE_z, MM_z, MM_y, PE_x, SMA$
Subject 2	$PE_y, MM_x, PE_x, MM_y, MM_z, RMS_x$
Subject 3	$PE_y, MM_x, MM_z, MM_y, PE_x, PE_z$
Subject 4	$PE_y, MM_x, MM_y, PE_z, MM_z, PE_x$
Subject 5	$MM_y, PE_y, MM_x, MM_z, PE_z, PE_x$
Subject 6	$MM_y, PE_y, MM_z, MM_x, PE_x, PE_z$
Subject 8	$PE_y, MM_x, MM_y, PE_x, MM_z, PE_z$

Table 4: Selected features by FCBF

Subject	Features
Subject 1	$RMS_x, PE_y, RMS_y, M_z, RMS_z, SC_x$
Subject 2	$SE_x, PE_x, SE_z, RMS_y, M_y, SMA$
Subject 3	$PE_y, SMA, RMS_x, M_y, RMS_z, MM_x$
Subject 4	$2PE_y, SE_y, SE_x, M_z, SC_x, MAD_x$
Subject 5	$RMS_x, PE_y, M_x, TA, E_y, SC_x$
Subject 6	$SMA, RMS_z, PE_x, PE_y, M_x, MAD_z$
Subject 8	$PE_y, RMS_x, M_z, MAD_y, SC_x, IQR_x$

Table 5: Selected features by FIS

Subject	Features
Subject 1	$RMS_x, MAD_x, M_x, PE_x, SC_z, E_y$
Subject 2	$SE_z, RMS_x, M_x, TA, PE_x, M_z$
Subject 3	$RMS_x, PE_x, M_x, MAD_x, E_y, RMS_z$
Subject 4	$MM_x, MAD_x, MAD_y, MM_y, SC_x, SC_z$
Subject 5	$E_y, SC_z, RMS_x, SC_y, E_z, M_x$
Subject 6	$RMS_x, PE_x, M_x, SE_x, SC_y, MM_z$
Subject 8	$MM_x, MAD_y, RMS_x, IQR_y, V_y, SE_x$

Table 6: Selected features by IG

Subject	Features
Subject 1	$ZCR_z, ZCR_x, MM_x, E_x, E_z, AR2_x$
Subject 2	$AR2_z, KS_z, PF_z, AR1_y, ZCR_y, CORR_y$
Subject 3	$ZCR_z, IQR_z, E_x, AR1_x, AR2_x, AR2_y$
Subject 4	$ZCR_y, ZCR_z, E_x, E_y, AR1_x, AR2_x$
Subject 5	$KS_z, ZCR_x, ZCR_z, ZCR_y, E_x, AR1_x$
Subject 6	$AR1_z, AR2_y, PF_z, AR2_z, ZCR_y, ZCR_z$
Subject 8	$AR2_y, ZCR_z, E_x, E_y, AR1_x, AR2_x$

Table 7: Selected features by KW

Subject	Features
Subject 1	$E_y, SC_y, SC_z, SC_x, IQR_x, MAD_x$
Subject 2	$SC_z, SC_y, E_y, IQR_y, KS_z, MAD_y$
Subject 3	$SC_y, SC_z, IQR_x, IQR_y, MM_y, MAD_x$
Subject 4	$SC_x, SC_y, SC_z, IQR_x, IQR_z, MAD_x$
Subject 5	$E_y, E_z, SC_y, SC_z, SC_x, IQR_y$
Subject 6	$SC_y, SC_z, E_y, E_z, MAD_x, PE_y$
Subject 8	$SC_y, SC_z, SC_x, M_y, IQR_y, IQR_z$

Table 8: Selected features by MRMR

Subject	Features
Subject 1	$PE_y, TA, AR2_y, AR2_z, AR2_x, AR1_z$
Subject 2	$PE_y, TA, SC_z, SC_y, E_y, E_z$
Subject 3	$MM_x, TA, AR2_z, AR1_z, AR2_x, AR2_y$
Subject 4	$PE_x, TA, AR2_y, AR2_z, AR2_x, AR1_z$
Subject 5	$MM_y, TA, AR2_z, AR2_y, AR2_x, AR1_z$
Subject 6	$MM_y, TA, AR2_z, AR2_y, SC_z, E_z$
Subject 8	$PE_y, TA, AR2_z, AR2_y, AR2_x, AR1_z$

Table 9: Selected features by T-test

Subject	Features
Subject 1	$MM_x, E_x, E_z, AR2_x, SE_z, RMS_z$
Subject 2	$ZCR_z, RMS_z, TA, SE_y, SE_z, RMS_y$
Subject 3	$IQR_z, E_x, AR1_x, AR2_x, AR2_y, RMS_z$
Subject 4	$E_x, E_y, AR1_x, AR2_x, ZCR_y, ZCR_z$
Subject 5	$E_x, AR1_x, RMS_y, RMS_z, SE_y, SE_z$
Subject 6	$TA, SE_z, RMS_z, ZCR_y, AR2_x, RMS_y$
Subject 8	$E_y, E_z, AR1_x, AR2_x, ZCR_z, ZCR_y$

Table 10: Selected features by Bayesian

Subject	Features
Subject 1	$PE_z, SMA, CORR_y, KS_x, CORR_x, MM_y$
Subject 2	$RMS_y, M_y, MM_z, KS_x, PE_z, PE_x$
Subject 3	$PE_y, MM_x, SMA, PF_y, IQR_x, E_y$
Subject 4	$PE_y, MM_x, PE_z, MM_y, M_x, PF_y$
Subject 5	$PE_y, TA, CORR_y, MM_x, PF_y, MM_z$
Subject 6	$PE_x, PE_y, PF_y, PE_z, MM_x, MM_y$
Subject 8	$PE_y, PE_x, MM_x, IQR_x, SMA, PF_x$

Table 11: Recognition rates for subject 1

	ANN		1NN		DT	
	n=5	n=6	n=5	n=6	n=5	n=6
CFS	93.1	94.7	98.4	99.7	94.2	95.9
CHI	84.9	92.6	100	100	98.6	98.2
FCBF	86.2	90.7	99.7	99.9	90.9	96.3
FIS	87.2	93.4	91.9	94.6	90.1	91.8
IG	18.0	24.7	12.5	12.5	18.0	24.7
KW	83.9	89.0	97.4	99.6	87.2	89.7
MRMR	87.9	90.8	95.9	98.7	93.5	95.0
T-test	12.7	51.7	43.5	77.4	44.0	73.9
Bayesian	92.2	93.0	96.3	98.3	80.7	94.3

Table 12: Recognition rates for subject 2

	ANN		1NN		DT	
	n=5	n=6	n=5	n=6	n=5	n=6
CFS	96.5	97.8	100	100	85.1	86.6
CHI	96.6	97.9	99.9	100	97.6	85.8
FCBF	82.5	85.5	99.7	99.9	97.1	97.8
FIS	71.3	84.7	81.7	83.4	94.0	94.6
IG	63.2	69.0	95.9	88.0	73.6	74.1
KW	89.9	92.2	99.8	99.9	93.2	94.1
MRMR	94.4	94.3	100	100	95.1	96.5
T-test	43.1	72.8	98.5	99.8	93.6	96.0
Bayesian	96.4	98.6	98.7	99.4	96.9	97.5

Table 13: Recognition rates for subject 3

	ANN		1NN		DT	
	n=5	n=6	n=5	n=6	n=5	n=6
CFS	95.0	96.0	88.5	99.8	79.0	87.1
CHI	93.7	94.6	100	100	99.7	99.8
FCBF	96.2	96.1	99.0	99.9	88.8	90.0
FIS	94.0	89.4	76.2	93.9	92.6	94.0
IG	62.3	65.4	62.2	62.9	62.2	63.1
KW	88.1	89.7	99.4	99.4	97.1	97.4
MRMR	82.6	83.9	79.5	79.7	93.0	93.0
T-test	62.3	65.5	62.2	66.4	62.2	70.6
Bayesian	95.3	96.3	99.9	99.9	97.8	98.9

Table 14: Recognition rates for subject 4

	ANN		1NN		DT	
	n=5	n=6	n=5	n=6	n=5	n=6
CFS	92.1	94.6	99.9	100	94.2	97.2
CHI	92.2	93.4	100	100	98.9	99.5
FCBF	62.6	84.5	84.2	87.8	88.9	89.5
FIS	75.0	87.4	71.0	77.1	72.1	76.7
IG	57.9	66.7	51.5	62.5	57.9	66.7
KW	82.3	83.0	68.1	68.9	65.7	66.2
MRMR	82.4	86.9	85.2	89.7	88.3	88.6
T-test	62.7	66.8	62.5	66.5	62.8	66.8
Bayesian	93.7	95.0	100	100	99.4	99.6

Table 15: Recognition rates for subject 5

	ANN		1NN		DT	
	n=5	n=6	n=5	n=6	n=5	n=6
CFS	90.7	95.5	99.6	99.5	96.9	96.7
CHI	92.3	94.5	100	100	99.5	99.6
FCBF	93.1	94.2	70.0	74.1	84.0	62.1
FIS	85.4	92.6	93.5	97.5	92.6	94.2
IG	38.4	47.9	26.2	59.3	38.4	58.3
KW	82.8	87.8	88.2	95.6	82.8	88.7
MRMR	85.8	88.2	62.7	74.4	44.1	49.6
T-test	59.5	61.1	82.0	92.3	76.3	87.7
Bayesian	94.3	96.5	70.6	82.9	46.1	50.3

Table 16: Recognition rates for subject 6

	ANN		1NN		DT	
	n=5	n=6	n=5	n=6	n=5	n=6
CFS	93.0	97.6	100	100	97.7	97.8
CHI	93.8	96.9	100	100	99.5	99.6
FCBF	92.2	95.3	100	100	97.3	97.8
FIS	58.7	62.3	83.1	91.8	88.6	91.5
IG	46.9	67.3	58.1	70.4	53.8	56.6
KW	81.4	91.9	85.5	95.7	81.9	90.6
MRMR	82.6	85.3	87.4	90.7	84.8	88.6
T-test	66.0	68.1	93.6	99.5	83.6	90.4
Bayesian	94.5	96.5	100	100	99.7	97.7

Table 17: Recognition rates for subject 8

	ANN		1NN		DT	
	n=5	n=6	n=5	n=6	n=5	n=6
CFS	92.5	93.8	99.2	99.7	91.4	91.9
CHI	92.4	89.8	100	100	99.7	99.5
FCBF	91.8	95.7	94.7	98.8	93.7	96.7
FIS	88.0	70.5	98.3	99.9	98.2	98.4
IG	47.9	57.3	42.5	48.6	47.9	52.9
KW	87.8	91.0	83.3	87.7	80.0	81.6
MRMR	84.4	87.0	96.7	99.0	94.6	95.0
T-test	47.9	57.9	42.5	46.9	47.9	57.9
Bayesian	95.8	96.8	100	100	99.5	99.6

4 Discussion

Table 18: Average recognition rates

	ANN		1NN		DT	
	n=5	n=6	n=5	n=6	n=5	n=6
CFS	93.3	95.7	97.9	99.8	91.2	93.3
CHI	92.3	94.2	100	100	99.1	97.4
FCBF	86.3	91.7	92.5	94.3	91.5	90.0
FIS	79.2	82.9	85.1	91.2	89.7	91.6
IG	47.8	56.9	49.8	57.7	50.3	56.6
KW	85.2	89.2	88.8	92.4	84.0	86.9
MRMR	85.7	88.1	86.8	90.3	84.8	86.6
T-test	50.6	63.4	69.3	78.4	67.2	77.6
Bayesian	94.6	96.1	95.1	97.2	88.6	91.1

The results in Table 2-17 provide large amount of information about features, feature selection and classification algorithms. Moreover, Table 18 shows the average recognition accuracies where the best rates are highlighted by red. According to the results, we summarized our observations into the following points:

- A generally applicable feature set does not exist because the selected features for the subjects are just partially overlapping.
- More features do not guarantee performance improvement, see CHI and IG in Table 12.
- The performance of some feature selection methods is person dependent. For instance, in the case of subject 2 the machine learning algorithms with MRMR reached higher recognition rate than in the case of subject 3.
- The performance of machine learning algorithms is feature selection and person dependent. For example, in the case of subject 5, the MRMR algorithm reached higher recognition rates with ANN while in other cases, the 1NN or DT were better. In addition, the 1NN produced better results than ANN and DT with CHI in all cases.
- As Table 18 shows, a relation exists between machine learning and feature selection algorithms. The 1NN and DT reached the highest recognition rate with the CHI while the ANN in combination with the Bayesian was the most efficient.
- Already 5 features are enough for good classification, because the difference of recognition rates with 5 and 6 features is rather small.
- Table 18 shows that, the 1NN produced the best result however the decision of the 1NN is more slower than in the case of ANN or DT, see for example [4]. Therefore, the usage of DT with CHI can be a good decision when the calculation capacity is strongly limited.
- With only one training, the performance of the ANN is not as good as we expected. The outcome of an ANN training depends on more hyperparameters, therefore a right ANN construction requires hyperparameter search and more trainings. However, it is a very time-consuming process.
- Since CHI used only time domain features, in this case frequency domain features are negligible.

- The result does not show strong relation between selected features and measured recognition rates for subjects of similar age.

Conclusion

In this article, we investigated the performance of feature selection methods on the HAR problem in combination with three well-known machine learning algorithms. To the survey two external sources have been utilised: WARD 1.0 database was the data source and the feature selection methods derived from an open source repository. At the beginning of the article, the most common feature extraction methods from time, frequency and other domains have been collected from the literature. Thereafter we selected 7 volunteers with different ages from the WARD and applied the feature extraction methods on their data. The selected features were the input of the ANN, 1NN and DT classifiers and the recognition rates have been archived.

Duarte et al., Gao et al., Maurer et al., Khan et al., Oniga et al. and Yang et al. used similar research conditions and environment to their study (they used single tri-axial accelerometer attached to one part of the body; they applied machine learning algorithm(s); they split time series into windows; etc.) and reached 95%, 97.9%, 99%, 97.8%, 96.4% and 92.8% recognition rates, respectively [4, 5, 9, 10, 12, 13]. As Table 10-16 demonstrate, we reached approximately 100% recognition rates in each cases. It is better than previous results and clearly indicates the efficiency of the feature selection and machine learning combination.

Pinardi et al., Su et al. and Yang et al. are also used the WARD database in their research [7, 29, 30]. They archived 97.8%, 98.5% and 93.5% recognition rates with five sensors and different kinds of classifiers (majority voting, distributed sparsity and support vector machine). Against their works, we showed that less sensors are enough for good classification and a general classifier with an appropriate feature selection algorithm can overcome other classification approaches in the HAR problem.

Bibliography

- [1] Suto, J.; Oniga, S.; Pop Sitar, P. (2016); Comparison of wrapper and filter feature selection algorithms on human activity recognition, *Computers Communications and Control (ICCCC), 2016 6th International Conference on*, IEEE Xplore, e-ISSN 978-1-5090-1735-5, DOI: 10.1109/ICCCC.2016.7496749, 124-129.
- [2] Chernbumroong, S.; Cang, S.; Atkins, A.; Yu, H. (2013); Elderly activities recognition and classification for applications in assisted living. *Expert Systems with Applications*, ISSN: 0957-4174, DOI: 10.1016/j.eswa.2012.09.004 40(5):1662-1676.
- [3] Sebestyen, G.; Tirea, A.; Albert, R. (2012); Monitoring human activity trough portable devices. *Carpathian Journal of Electronic and Computer Engineering*, ISSN 2343-8908, 5(1):101-106.
- [4] Gao, L; Bourke, A.K.; Nelson, J. (2014); Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems. *Medical Engineering & Physics*, ISSN: 1350-4533, DOI: 10.1016/j.medengphy.2014.02.012, 36(6):779-785.
- [5] Maurer, U.; Smailagic, A.; Siewiorek, D,P; Deisher, M. (2006); Activity recognition and monitoring using multiple sensors on different body positions. *International Workshop on Wearable and Implementable Body Sensor Networks*, ISBN: 0-7695-2547-4, DOI: 10.1109/BSN.2006.6, Cambridge, USA, 112-116.

-
- [6] Orha, I.; Oniga, S. (2015); Wearable sensor network for activity recognition using inertial sensors, *Carpathian Journal of Electronic and Computer Engineering*, ISSN 2343-8908, 8(2):3-6.
- [7] Yang, A.Y.; Jafari, L.; Systry, S.S.; Bajcsy, R. (2009); Distributed recognition of human actions using wearable motion sensor networks. *Journal of Ambient Intelligence and Smart Environments*, ISSN 1876-1372, DOI: 10.3233/AIS-2009-0016, 1(1):1-5.
- [8] Preece, J.S.; Goulermas, J.Y.; Kenney, L.P.J.; Howard, D. (2009); A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. *IEEE Transactions on Biomedical Engineering*, ISSN: 1558-2531, DOI: 10.1109/TBME.2008.2006190, 56(3):871-879.
- [9] Khan, A.M.; Lee, Y.K.; Lee, S.Y.; Kim, T.S. (2010); A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer. *IEEE Transactions on Information Technology in Biomedicine*, ISSN: 1558-0032, DOI: 10.1109/TITB.2010.2051955, 14(5):1166-1172.
- [10] Oniga, S.; Suto, J. (2015); Optimal recognition method of human activities using artificial neural networks. *Measurement Science Review*, ISSN 1335-8871, DOI: 10.1515/msr-2015-0044, 15(5):323-327.
- [11] Oniga, S., Suto, J. (2014); Human activity recognition using neural networks. *15th International Carpathian Control Conference*, ISBN: 978-1-4799-3528-4, DOI: 10.1109/CarpathianCC.2014.6843636, Velke Karlovice, Czech Republic, 403-406.
- [12] Yang, J.Y.; Wang, J.S.; Chen, Y.P. (2008); Using acceleration measurements for activity recognition: an effective learning algorithm for constructing neural classifiers. *Pattern Recognition Letters*, ISSN: 0167-8655, DOI: 10.1016/j.patrec.2008.08.002, 29(16):2213-2220.
- [13] Duarte, F.; Lourenco, A.; Abrantes, A. (2014); Classification of physical activities using a smartphone: evaluation study using multiple users. *Procedia Technology*, ISSN: 2212-0173, DOI: 10.1016/j.protcy.2014.10.234, 17(1):239-247.
- [14] Karantonis, D.M.; Narayanan, M.R.; Mathie, M.; Lovell, N.H.; Celler, B.G. (2006); Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. *IEEE Transactions on Information Technology in Biomedicine*, ISSN: 1558-0032, DOI: 10.1109/TITB.2005.856864, 10(1):156-167.
- [15] Lara, D.O.; Labrador, M.A. (2013); A survey on human activity recognition using wearable sensors. *IEEE Communications Survey & Tutorials*, ISSN: 1553-877X, DOI: 10.1109/SURV.2012.110112.00192, 15(3):1192-1209.
- [16] Godfrey, A.; Conway, R.; Meagher, D.; O'laighin, G. (2008); Direct measurement of human movement by accelerometry. *Medical Engineering & Physics*, ISSN: 1350-4533, DOI: 10.1016/j.medengphy.2008.09.005, 30(10):1364-1386.
- [17] Bayat, A.; Pomplun, M.; Tran, D.A. (2014); A study on human activity recognition using accelerometer data from smartphones. *Procedia Computer Science*, ISSN: 1877-0509, DOI: 10.1016/j.procs.2014.07.009, 34(1):450-457.
- [18] Kavanagh, J.J.; Menz, B.H. (2008); Accelerometry: a technique for quantifying movement patterns during walking. *Gait Posture*, ISSN: 0966-6362, DOI: 10.1016/j.gaitpost.2007.10.010, 28(1):1-15.

-
- [19] Shoaib, M.; Bosch, S.; Incel, O.D.; Scholten, H.; Havinga, P.J.M. (2015); A survey of online activity recognition using mobile phones. *Sensors*, ISSN 1424-8220, DOI: 10.3390/s150102059, 15(1):2059-2085.
- [20] Suto, J.; Oniga, S.; Buchman, A. (2015); Real time human activity monitoring. *Annales Mathematicae et Informaticae*, ISSN 1787-6117, 44(1):187-196.
- [21] Cheng, C.H.; Wang, P.S.P. (2005); *Handbook of Pattern Recognition and Computer Vision*, 3th ed., World Scientific, ISBN 978-981-4505-21-5.
- [22] Liu, H.; Motoda, H.; Setiono, R.; Zhao, Z. (2010); Feature selection: an ever evolving frontier in data mining, *4th Workshop on Feature Selection in Data Mining*, ISSN 1533-7928, Hyderabad, India, 4-13.
- [23] Liu, H.; Motoda, H. (2008); *Computational Methods of Feature Selection*, CRC Press Taylor & Francis Group, ISBN 978-158-488-878-9.
- [24] Hall, M.A.; Smith, L.A. (1999); Feature selection for machine learning: Comparing a correlation-based filter approach to the wrapper, *Florida Artificial Intelligence Symposium*, Florida, ISBN 978-1-57735-756-8, USA, 235-239.
- [25] Saeys, Y.; Inza I.; Larranaga P. (2007); A review of feature selection techniques in bioinformatics, *Bioinformatics*, DOI: 10.1093/bioinformatics/btm344, ISSN 1460-2059, 23(19):2507-2517.
- [26] Jatoba, C.L.; Grobmann, U.; Kunze, U.; Ottenbacher J.; Stork, W. (2008); Context-aware mobile health monitoring: Evaluation of different pattern recognition methods for classification of physical activity. *30th Annual International IEEE EMBS Conference*, ISBN: 978-1-4244-1814-5, DOI: 10.1109/IEMBS.2008.4650398, Vancouver, Canada, 5250-5253.
- [27] Zhao, Z.; Morstatte, F.; Sharma, S.; Alelyani, S.; Anand, A.; Liu, H. (2011); *Advancing feature selection research- ASU feature selection repository*. Technical Report, Arizona State University, http://featureselection.asu.edu/old/featureselection_techreport.pdf
- [28] Oniga, S.; Suto, J. (2016); Activity recognition in adaptive assistive systems using artificial neural networks. *Elektronika ir Elektrotechnika*, ISSN: 2029-5731, DOI: <http://dx.doi.org/10.5755/j01.eee.22.1.14112>, 22(1):68-72.
- [29] Pinardi, S.; Bisiani, R. (2010); Movement recognition with intelligent multisensor analysis, a lexical approach. *6th International Conference on Intelligent Environments*, ISBN 978-1-60750-639-3, DOI: 10.3233/978-1-60750-638-6-170, Kuala Lumpur, Malaysia, 170-177.
- [30] Su, B.; Tang, Q.; Wang, G.; Sheng, M. (2016); The recognitions of human daily actions with wearable motion sensor system, *Lecture Notes in Computer Science: Transactions on Edutainment XII*, ISBN 978-3-662-50544-1, DOI: 10.1007/978-3-662-50544-16, 9292(1):68-77.
- [31] Ertugrul, O.F.; Kaya, Y. (2016); Determining the optimal number of body-worn sensors for human activity recognition. *Soft Computing*, ISSN 1433-7479, DOI: 10.1007/s00500-016-2100-7, 20(2):1-8.