A Benchmark Study on Feature Selection for Human Activity Recognition

Katerina Karagiannaki[*‡], **Athanasia Panousopoulou**[‡], **Panagiotis Tsakalides**[*‡]

[*]Department of Computer Science, University of Crete
[‡]Institute of Computer Science, Foundation for Research and Technology-Hellas
Heraklion, Crete, GR-70013, Greece
{karayan, apanouso, tsakalid}@ics.forth.gr

http://dx.doi.org/10.1145/2968219.2971421

Abstract
Human Activity Recognition (HAR) currently confronts the challenge of interpreting massive data streams to a significantly smaller number of activities. Thus, feature selection should be treated as an inseparable aspect of the HAR chain. In this work we perform an integrated study on feature selection, considering: (a) the generation of an expanded HAR dataset; (b) the development of a software tool that covers the entire feature-level fusion chain; (c) the calculation of performance metrics that go beyond machine learning terms. The results yield guidelines on the preferable feature selection technique that should be considered for adoption in the HAR domain.

Author Keywords
Human Activity Recognition; Feature Selection; Data Analytics; Wearable Computing.

ACM Classification Keywords
I.5.2 [Pattern Recognition]: Design Methodology: Feature evaluation and selection

Introduction
Recent advances in ubiquitous sensing are causing a paradigm shift on the multi-disciplinary and emerging field of Human Activity Recognition (HAR). Massive volumes of daily generated raw information are currently available for interpret-

Figure 1: The deployment of the wearable sensors (●) for the FORTH-TRACE dataset.

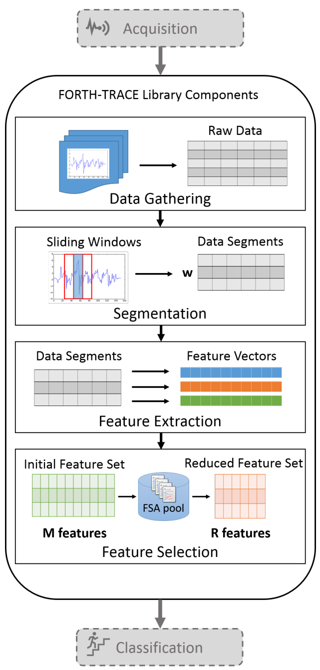

Figure 2: The FORTH-TRACE library (implemented in Java).

ing complex activities. Prior to the adoption of an appropriate classification mechanism, feature selection techniques [6] can offer substantial data compression, introducing limited or no information loss. Indeed, empirical studies on extensive datasets collected from diverse groups of individuals [3] conclude that dominant feature subsets can produce similar or better classification accuracy than the one achieved when the entire feature set is used. Despite the practical importance of incorporating feature selection in the HAR domain, limited efforts are recorded on the systematic calculation of dominant attributes as an inseparable aspect of the activity recognition process. In order to address this literature gap, in this work we perform evaluation studies on landmark feature selection algorithms (FSA) in the HAR domain, and we outline revealing remarks on their performance when applied to different types of HAR datasets.

The FORTH-TRACE Benchmark Framework

Feature selection is considered an important step of filtering redundant information, prior to applying any classification technique, especially for data streams characterized by high correlation. In order to study its effects in the HAR domain, in this work we adopt a two-step procedure, which entails: (a) the generation of an expanded, in terms of activities, HAR dataset (the FORTH-TRACE dataset); (b) the development of a modular library that covers the entire feature-level fusion chain (the FORTH-TRACE library).

The FORTH-TRACE dataset addresses the existing gap in publicly available HAR datasets on combining both inertial sensing from multiple locations on the human body with the capture of combined and short in duration activities. Specifically, the FORTH-TRACE dataset[1] utilizes 5 wearable sensors, and monitors acceleration, angular velocity, and changes in the magnetic field from both the upper and lower parts of the human body, as indicated at Figure 1. During the data collection, 15 participants performed a series of 16 activities[1]. Compared to the current state of the art on publicly available datasets for HAR applications, the FORTH-TRACE dataset emphasizes on capturing postural transitions by employing specialized wearable platforms[2] and an expanded set of inertial sensing modalities. This combination can yield more accurate interpretation of body motions, including the detection of sudden postural transitions. In addition, going beyond the current state of the art, the FORTH-TRACE dataset expands the traditional set of activities towards combined human actions (i.e., “Sit & Talk”, “Walk & Talk”, “Stairs Climbing (Up/Down) & Talk”), thereby better describing daily human routine.

The FORTH-TRACE library adopts a modular design on implementing the intermediate steps between the raw data collection and the classification of human actions, according to the Activity Recognition Chain (ARC) [2]. In contrast to generalized libraries for machine learning (e.g., Weka [5]), the FORTH-TRACE library considers a sequential interaction between its key components, while it can interchangeably be used for different types of HAR datasets, regardless of the number of modalities or locations involved. Referring to the architecture of the FORTH-TRACE library (Figure 2), the key parameter for the Segmentation component is the length w for the data tessellation into sliding windows of observations. In addition, the Feature Extraction Component incorporates the statistical attributes proposed in [9], in order to extract M features per data segment. Another important characteristic of the library is that it is extensible in terms of the feature selection technique employed for the calculation of the reduced feature set. Specifically, the FSA Pool provides different FSAs available for use, in order to calculate the R $(R < M)$ most representative features.

[1] https://github.com/karayan/FORTH_TRACE_DATASET

[2] http://www.shimmersensing.com/

Component	Experimental Parameters
Data gathering	Number of independent experiments per dataset: 10
Data Segmentation	Window length w: $\{$2, 5, 10, 20$\}$ (s) (50% overlapping factor)
Feature Extraction	Initial number of features M: 81(UCI-HAPT), 135 (FORTH-TRACE)
Feature Selection	SFFS: Correlation metric $\rightarrow$ Linear Relation Metric FSSA: k parameter for similar features clustering [7]$\rightarrow$ 0.8$\times M$ Relief-F: Contents of reduced feature set $\rightarrow$ 10% of the top ranked features GCNC: Maximum number for features per cluster discarded as redundant [8]$\rightarrow$ 75% of the original cluster size

Table 1: The experimental parameters of FORTH-TRACE library for the evaluation studies.

Currently, our study consists of four implementations originating from distinct FSA families addressed in literature, reflecting on: supervised FSA (Sequential Floating Forward Selection, SFFS, [6]); unsupervised FSA (Feature Selection based on Feature Similarity, FSSA, [7]); ranking-based FSA (Relief-F, [6, 5]); and unsupervised graph-based FSA (Graph Clustering with Node Centrality, GCNC, [8]).

Evaluation Studies

For the purposes of evaluation, we consider two HAR datasets, namely: (a) the UCI-HAPT dataset [1], which employs a smartphone attached on the waist and monitors 3-axial accelerometer and gyroscope, and (b) the herein proposed FORTH-TRACE dataset. Both datasets capture postural transitions, however they differ in the range of activities captured; opposed to UCI-HAPT, FORTH-TRACE classifies the combination of motion and talk as a separate activity.

Table 1 summarizes the execution parameters of the experiments, including the specific user-defined parameters per FSA. The evaluation metrics are: (a) the normalized value ($\overline{H} \in [$0,1$]$) of the representation entropy [7] for the reduced data pattern, with respect to its maximum value ($\log R$); (b) the classification accuracy ($CA \in [$0,1$]$) of a classifier using Gaussian-kernel Support Vector Machines (SVM) [4]; (c) the compression ratio $CR = (M - R)/M$, and (d) the execution time per FSA. The $\overline{H}$ metric quantifies the redundancy achieved in the final set; when $\overline{H}$ $\rightarrow$0 (1), the data patterns are highly relevant (complementary) to each other.

Experimental Results. Figures 3(a) and (b) present the CA metric with respect to the size of the segmentation window $w = \{$2, 5, 10, 20$\}$ s. It is interesting to note that for all FSAs considered and datasets examined, an increase of the segmentation window is accompanied by a degradation on the classification accuracy of the produced reduced data patterns. For instance, considering the UCI-HAPT dataset (Figure 3(a)) and the FSSA algorithm as w 2s $\rightarrow$ 10s, the value of CA drops from 0.91 to 0.85. This is due to the fact that short data segments result to a more coherent feature vector with respect to the associated activity label; human activities change over time and thus the wider the observation window becomes, the more activities will have to be reflected on the resulting feature vector.

Considering the best-case scenario for the parameter w (=2s) the optimal performance in terms of CA is offered by Relief-F for both datasets. By contrast, the worst performance is provided by FSSA (CA =0.91) and SFFS (CA =0.9) for the UCI-HAPT and FORTH-TRACE datasets respectively. Notably, the unsupervised GCNC provides reduced data patterns with solid classification accuracy for both the UCI-HAPT and FORTH-TRACE datasets. In conjunction to this remark, Figure 3(c) yields the $\overline{H}$ scores when FSSA or GCNC is applied on both datasets. These results highlight the importance of the additional sensing modality (i.e., magnetometer at the FORTH-TRACE dataset); the reduced feature patterns extracted from the FORTH-TRACE dataset have a higher value of $\overline{H}$ than those corresponding to the UCI-HAPT dataset. In addition, the GCNC algorithm yields better results in terms of quality compression, since the respective value $\overline{H}$ is constantly higher than the one extracted when the FSSA algorithm is applied. Finally, focusing on the FORTH-TRACE dataset, Table 2, enlists the results on all metrics considered.

Conclusions

In this work, we studied the performance of feature selection in the HAR domain, considering both traditional approaches (i.e. SFFS, Relief-F) as well as state-of-art algorithms (i.e. graph-based GCNC). By the means of a novel dataset and the appropriate software tool that mediates be-

	Left Wrist			
Algorithm	$\overline{H}$	**CA**	**Time (s)**	**CR**
SFFS	1	0.86	3.19	0.98
FSSA	0.91	0.99	2.56	0.8
Relief-F	0.87	0.99	1.77	0.9
GCNC	**0.93**	**0.97**	**0.85**	**0.96**
	Right Wrist			
Algorithm	$\overline{H}$	**CA**	**Time (s)**	**CR**
SFFS	1	0.91	3.17	0.98
FSSA	0.91	0.99	2.54	0.8
Relief-F	0.88	0.99	1.76	0.9
GCNC	**0.97**	**0.97**	**0.85**	**0.97**
	Torso			
Algorithm	$\overline{H}$	**CA**	**Time (s)**	**CR**
SFFS	1	0.9	3.17	0.98
FSSA	0.92	0.99	2.54	0.8
Relief-F	0.84	0.99	1.75	0.9
GCNC	**0.95**	**0.97**	**0.87**	**0.97**
	Right Thigh			
Algorithm	$\overline{H}$	**CA**	**Time (s)**	**CR**
SFFS	0.99	0.92	3.82	0.98
FSSA	0.88	0.99	2.57	0.8
Relief-F	0.86	0.99	1.79	0.9
GCNC	**0.97**	**0.92**	**0.9**	**0.97**
	Left Ankle			
Algorithm	$\overline{H}$	**CA**	**Time (s)**	**CR**
SFFS	1	0.86	3.19	0.98
FSSA	0.89	0.99	2.55	0.8
Relief-F	0.78	0.99	1.76	0.9
GCNC	**0.96**	**0.96**	**0.85**	**0.97**

- All FSA: Highly compressed reduced feature set ($CR \geq$ 0.8)
- Relief-F: Poor performance in terms of $\overline{H}$ ($\leq$ 0.87).
- GCNC: Optimal performance in terms of execution time ($\leq$ 0.9s) and data compression ($CR \geq$ 0.96)
- SFFS: Poor performance in terms of execution time ($\geq$ 3.17s)

Table 2: FSA performance for the FORTH-TRACE dataset and w =2s.

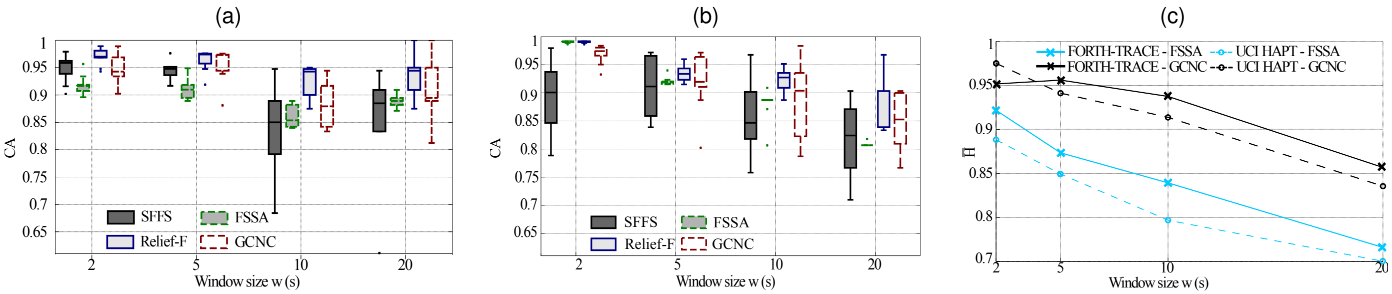

Figure 3: Experimental Results w.r.t. the window size w: (a)-(b) the value of CA w.r.t window size w for the UCI-HAPT dataset (a), and the FORTH-TRACE dataset (b); (c) the value of $\overline{H}$ when FSSA and GCNC are applied on the UCI-HAPT and the FORTH-TRACE datasets.

tween the raw data acquisition and the classification, we studied different metrics on different types of HAR datasets. The combination of all results leads to the observation that the unsupervised graph-based approach yields the optimal combination of all metrics and types of datasets examined.

Acknowledgment

This work was funded by the DEDALE project, contract no. 665044, within the H2020 Framework Program of the European Commission.

REFERENCES

1. D. Anguita, et al. 2013. A Public Domain Dataset for Human Activity Recognition using Smartphones. In Proceedings of *ESANN 2013*. 437–442.
2. A. Bulling, et al. 2014. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)* 46, 3 (2014), 33.
3. N. A. Capela, et al. 2015. Feature Selection for Wearable Smartphone-Based Human Activity Recognition with Able bodied, Elderly, and Stroke Patients. *PLoS ONE* 10, 4 (04 2015), 1–18.
4. C.-C. Chang and C.-J. Lin. 2011. LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.* 2, 3, Article 27 (May 2011), 27 pages.
5. M. Hall, et al. 2009. The WEKA Data Mining Software: An Update. *SIGKDD Explorations* 11, 1 (2009).
6. H. Liu and H. Motoda. 2007. *Computational methods of feature selection*. CRC Press.
7. P. Mitra, et al. 2002. Unsupervised feature selection using feature similarity. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24, 3 (2002), 301–312.
8. P. Moradi and M. Rostami. 2015. A graph theoretic approach for unsupervised feature selection. *Engineering Applications of Artificial Intelligence* 44 (2015), 33–45.
9. M. Zhang and A. A Sawchuk. 2011. A feature selection-based framework for human activity recognition using wearable multimodal sensors. In *Proceedings of BodyNets 2011*.