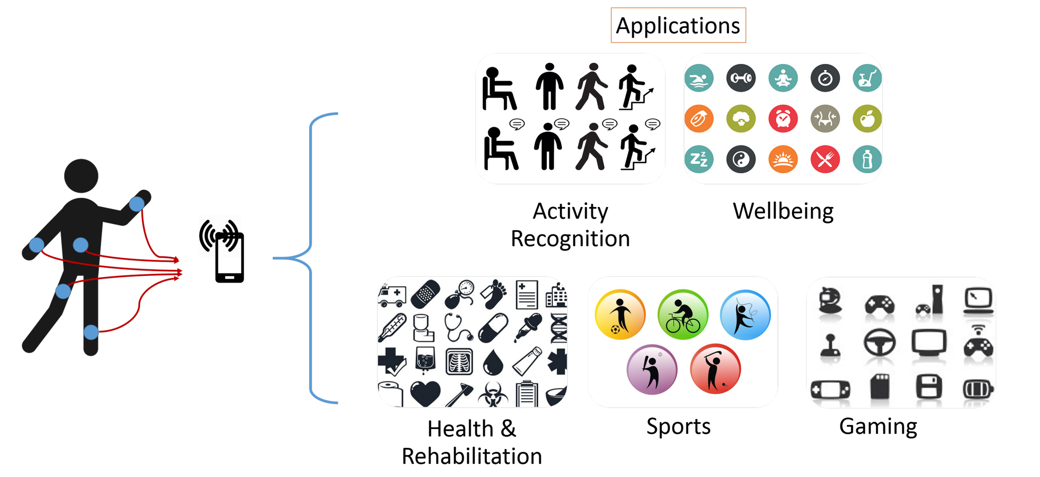
AN ONLINE FEATURE SELECTION ARCHITECTURE FOR HUMAN ACTIVITY RECOGNITION

Katerina KARAGIANNAKI, Athanasia PANOUSOPOULOU and Panagiotis TSAKALIDES

Foundation for Research and Technology – Hellas , Institute of Computer Science, Heraklion, Crete, Greece, GR-70013
University of Crete, Department of Computer Science, Heraklion, Crete, Greece, GR-70013

{karayan, apanouso, tsakalid}@ics.forth.gr

Motivation - Online Human Activity Recognition



Online Human Activity Recognition

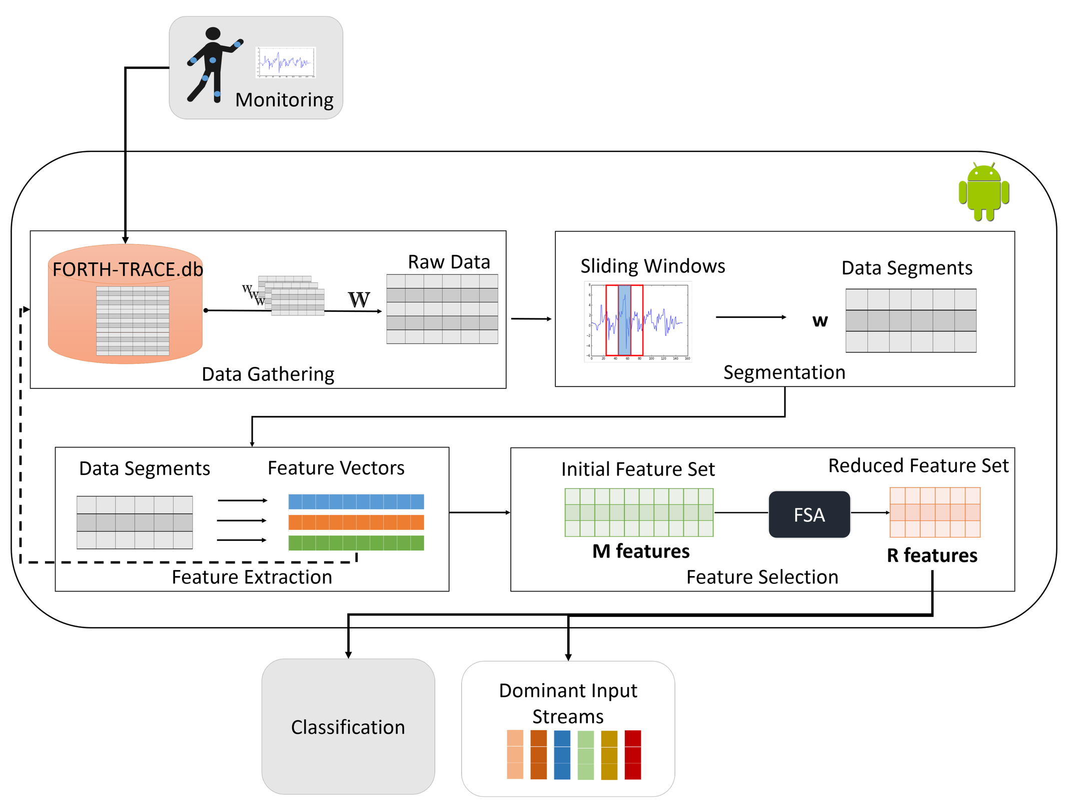
- Massive Streams of daily generated data
- Constantly alternating activities

Goals and Contribution

Feature Selection for HAR

- ✔ Reduce raw signals into discriminative features
- ✔ Accurate interpretation of Human Activities
- **Software Framework** for Online Feature Selection on Mobile Devices
- Explore online aspects of popular Feature Selection Algorithms
- Analyze FSA performance with respect to variation of activities

Proposed Online Feature Selection Architecture



Online Feature Selection library

- Extension of **FORTH-TRACE library** [1] in Android Devices
- **Dynamic data processing and storage**
  $\Rightarrow$ Database integration to load and store data
- **Dynamic data gathering**
  $\Rightarrow$ Temporal windows of size W to simulate online scenario
- **Online** Feature Extraction and Feature Selection

Experimental Setup

Dynamic Data Gathering

- FORTH-TRACE Dataset [1] (duration of activities: 18 min)
- Temporal windows $W \in \{2, 3, 4, 5\}$ (min)
- Generating $P \in \{9, 6, 5, 4\}$ partitions

FORTH-TRACE DATASET

https://github.com/spl-icsforth/FORTH_TRACE_DATASET

Segmentation

- Sliding windows of length $w = 2s$

Feature Extraction

- Statistical Features [6]

Feature Selection

- **Unsupervised**: Feature similarity for redundancy reduction - FSSA (2002) [3]
- **Ranker**: Relief-F (1997) [4]
- **Graph-Based**: Graph Clustering with Node Centrality - GCNC (2015) [5]

Post-Processing Evaluation Metrics

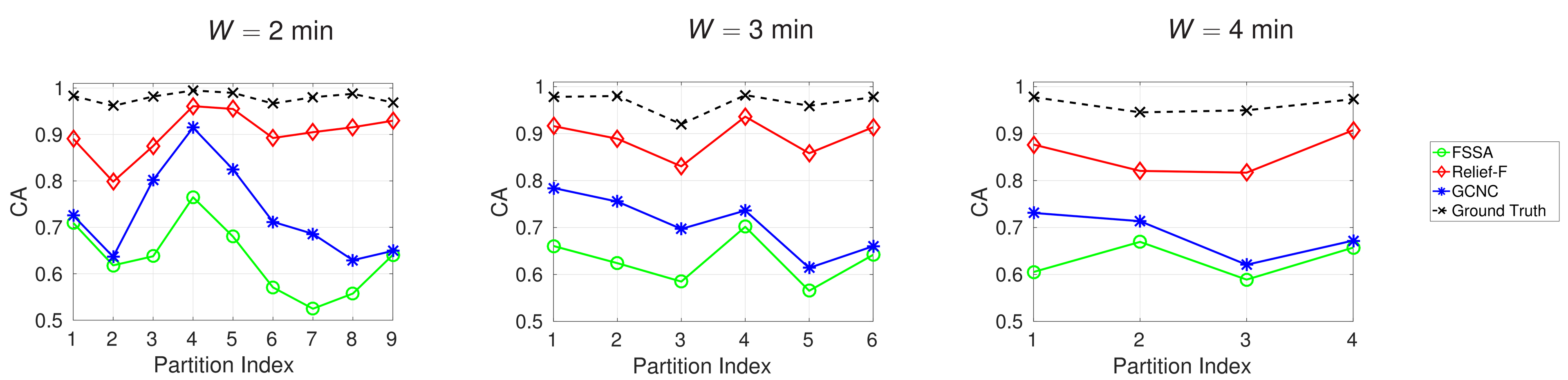
- Normalized Representation Entropy ($\bar{H}_r$) [3]
  - $\bar{H}_r \in [0, 1]$
  - Measures the amount of redundancy in feature set
- SVM classification accuracy (CA) [2]
  - Split feature set into Train (70%) and Test (30%) sets
  - Gaussian-Kernel SVM classifier

Online Performance of Library

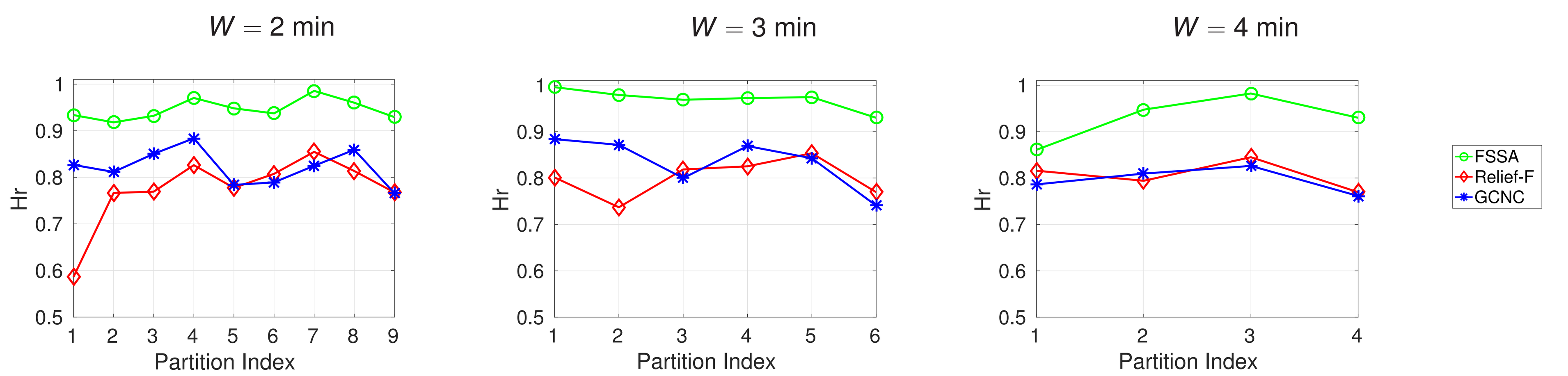
- Execution time per component
- Energy requirements of Android device

Experimental Results

Classification Accuracy w.r.t. the data chunk size W (mean values from all sensor locations)



Representation Entropy w.r.t. the data chunk size W (mean values from all sensor locations)



Inconsistent FSA performance over different temporal windows W due to variation of activity labels within different partitions P

Online HAR applications involve short data chunks!

Impact of activity labels to the Classification Accuracy

Classification Accuracy and percentage (%) of activity labels across different partitions for $W = 2$ min

Left Wrist

Partition Index	2	4	5	8
GCNC	0.66	0.81	0.84	0.68
Relief-F	0.78	0.97	0.97	0.88
FSSA	0.70	0.58	0.71	0.62

Right Thigh

Partition Index	2	4	5	8
GCNC	0.69	0.96	0.80	0.67
Relief-F	0.72	0.96	0.95	0.95
FSSA	0.62	0.84	0.74	0.54

Activity	**2**	**4**	**5**	**8**
stand	18.64	3.39	16.1	16.1
sit	41.53	-	-	-
sit & talk	26.27	-	-	-
walk	-	94.92	-	-
walk & talk	-	-	81.36	28.81
climb	-	-	-	-
climb & talk	-	-	-	50.85
postural transitions	13.56	1.69	2.54	4.23

Distribution of Activity Labels and CA

- ✔ Single primary activity $\Rightarrow$ optimal CA
- ✗ More activities $\Rightarrow$ less accurate predictions

Run-time aspects of Feature Selection library

Execution time(s) of the library components per W

W	2	3	4	5
Acquisition	1.2	1.36	1.73	1.7
Segmentation	0.73	0.95	1.14	1.16
Feature Extraction	**46.62**	**84.28**	**116.71**	**146.23**
FS: FSSA	**49.57**	**69.81**	**85.74**	**105.49**
FS: Relief-F	4.3	7.36	9.75	11.71
FS: GCNC	21.52	21.74	23.06	28.25

- **Most time consuming module:** Feature Extraction
- **FSA execution**:
- ✗ Relief-F (supervised): fast $\Rightarrow$ not suitable for online realizations
- ✔ GCNC (graph-based): adequate time and performance for integration with online architectures

Energy consumption of online library

Energy Requirement: $\sim$ 20.3 Joule

Outperforms:

- Adaptive Accelerometer Activity Recognition (A3R) algorithm [7] $\Rightarrow$ 100 Joule
- Shimmer3 Gesture Recognition application [8] $\Rightarrow$ 29.2 Joule

Reference List & Contact Details

1. Karagiannaki, et al. 2016. A benchmark study on feature selection for human activity recognition. ACM UbiComp 2016.
2. C.-C. Chang and C.-J. Lin. 2011. LIBSVM: A Library for Support Vector Machines.
3. P. Mitra, et al. 2002. Unsupervised feature selection using feature similarity.
4. I. Kononenko, et al. Overcoming the myopia of inductive learning algorithms with RELIEFF
5. P. Moradi and M. Rostami. 2015. A graph theoretic approach for unsupervised feature selection.
6. M. Zhang and A. A Sawchuk. 2011. A feature selection-based framework for human activity recognition using wearable multimodal sensors.
7. Z. Yan, V. Subbaraju, et al. 2012. Energy-efficient continuous activity recognition on mobile phones: An activity-adaptive approach.
8. Shimmer sensing, http://www.shimmersensing.com/

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

@spl_icsforth

http://spl.edu.gr/