RELEVANT FEATURE SELECTION FOR HOME APPLIANCES RECOGNITION

Sarra HOUIDI, François AUGER, Houda BEN ATTIA SETHOM, Dominique FOURER, Laurence MIEGEVILLE

ELECTRIMACS 2017
July 4-6th 2017

Problem statement

Home electrical appliances' recognition from current and voltage measurements:

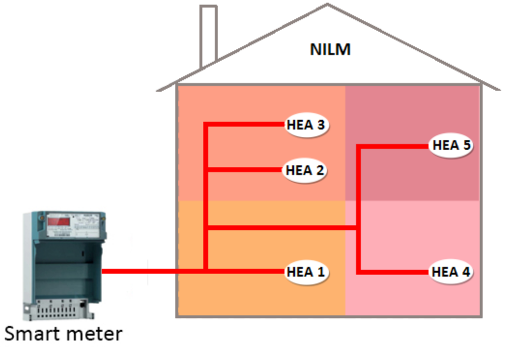
- **Which features to achieve this?**
- **Which are the most relevant?**

CONTENT

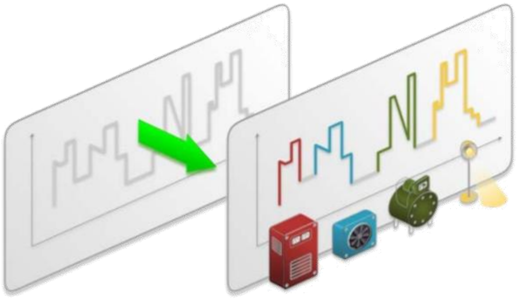
Non Intrusive Load Monitoring

Non-Intrusive Load Monitoring

Definition



- Process to estimate the energy consumed by individual **Home Electrical Appliances** (HEAs)
  → with **a single meter** in a house electrical panel.



Purposes:

- Partition of the load curve into its major components
- Breakdown of the energy expense per HEA

Non Intrusive Load Monitoring

Motivation and goals

- **Energy Conservation:** challenging issues due to population growth and HEAs' multiplication

Consumers

- Bills' awareness
- Monthly budget control and understanding
- Identification / Reparation / Replacement of energy « hogs »

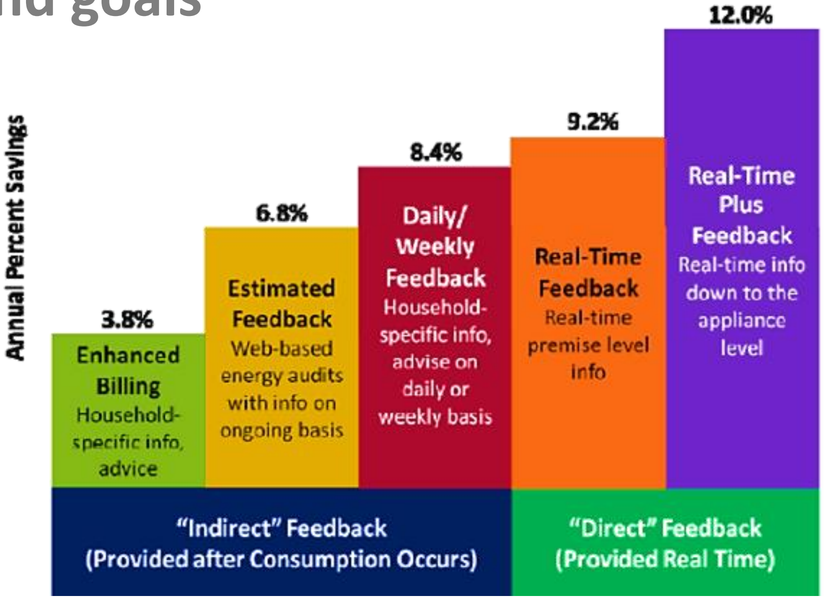
Utilities

- Customers' behavior prognosis
- Improve capacity planning
- Identification and verification of HEAs that could participate in DR

- **Political Context:** International directives for a better control electricity consumption through ***direct feedback***

Non Intrusive Load Monitoring

Motivation and goals



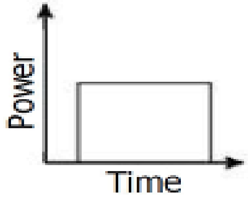
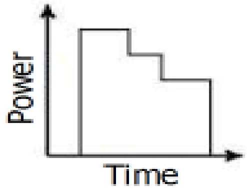
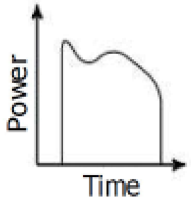
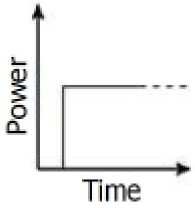
ACEEE energy savings resulting from 36 different studies between 2000-2010

***ACEEE**: American Council for an Energy Efficient Economy

Non Intrusive Load Monitoring

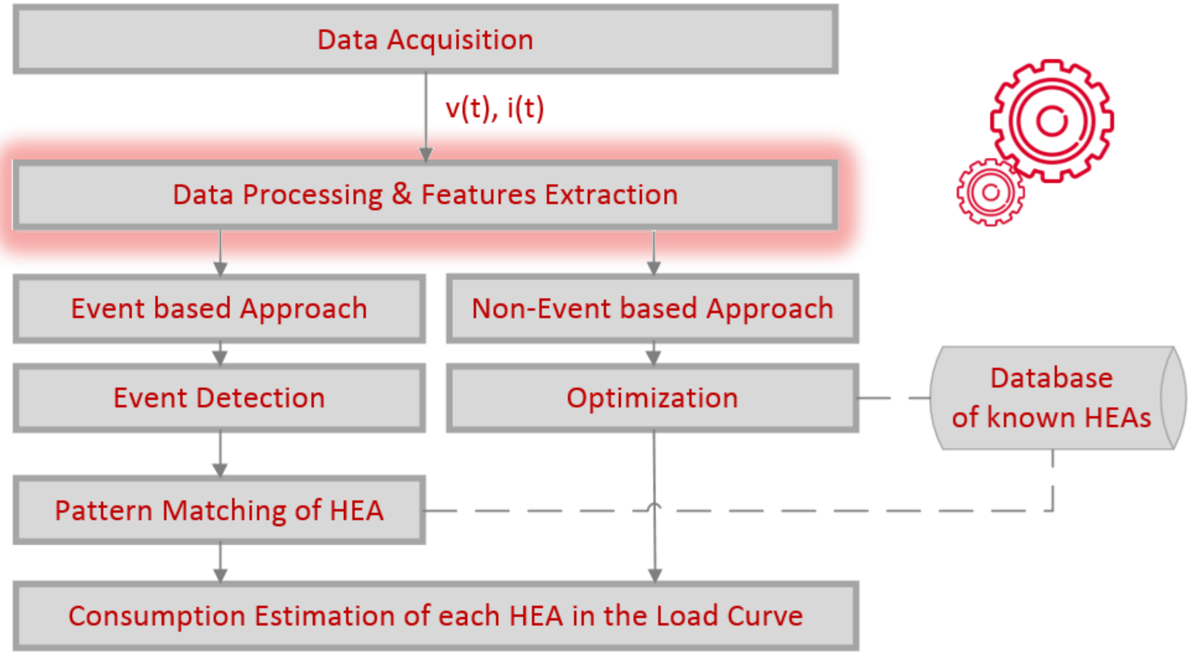
Challenges

- Different electrical behaviours: purely resistive / resistive inductive / harmonic polluting
- Multiple HEAs operation types:

ON/OFF	Finite State Machine	CVD	PCD
			
			

Non Intrusive Load Monitoring

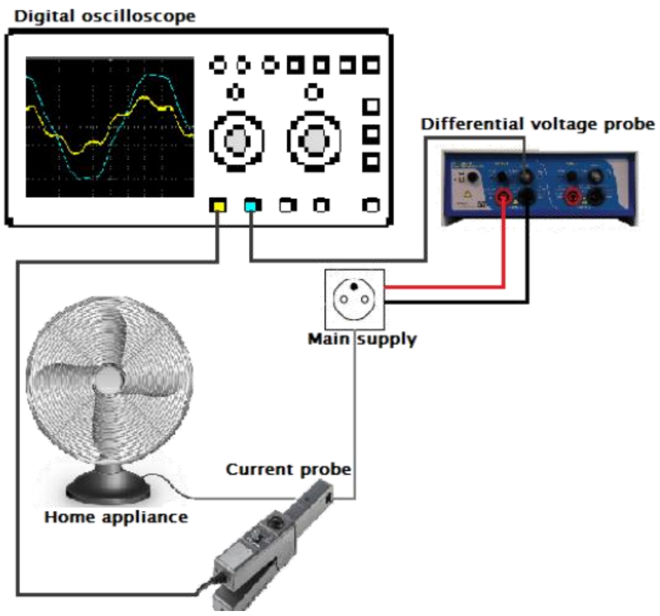
General framework



Home Electrical Appliances Database

Home Electrical Appliances Database

Setting up the HEAs Database



- **Acquisition of i(t) and v(t):**
  - in **steady-state** operating conditions
  - 2 and 6 periods
  - **59** types of HEAs
  - **sampling frequencies:**
    fs_1 = 250 kHz
    fs_2 = 50 kHz

- **Instrumentation:**
  - 10 mV/A sensitivity current probe
  - 1/100 attenuation differential voltage probe
  - 8-bit resolution digital oscilloscope (RIGOL DS1104Z)

Home Electrical Appliances Database

HEA Profiling

- **Signature of a HEA recording *i*** over one period of T= 20 ms

$$\boldsymbol{x}_{\mathbf{i}} = \{\boldsymbol{f}_{i,1}(\vec{\boldsymbol{u}}), \boldsymbol{f}_{i,2}(\vec{\boldsymbol{u}}), \dots, \boldsymbol{f}_{i,j}(\vec{\boldsymbol{u}}), \dots, \boldsymbol{f}_{i,p}(\vec{\boldsymbol{u}})\}$$

  - $\vec{u}$: $voltage\ and\ current\ measurements$
  - $f_{i,j}$: j^{th} $f$unction of $\vec{v}$ of the ith HEA recording, where $j \in \{1, \dots, p\}$ and $i \in \{1, \dots, n\}$

$$X_{(n\times p)} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \boldsymbol{x_i} \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} f_{1,1}(\vec{u}) & f_{1,2}(\vec{u}) & \dots & f_{1,j}(\vec{u}) & \dots & f_{1,p}(\vec{u}) \\ f_{2,1}(\vec{u}) & f_{2,2}(\vec{u}) & \dots & f_{2,j}(\vec{u}) & \dots & f_{2,p}(\vec{u}) \\ & & & \vdots & & \\ \boldsymbol{f_{i,1}}(\vec{u}) & \boldsymbol{f_{i,2}}(\vec{u}) & \dots & \boldsymbol{f_{i,j}}(\vec{u}) & \dots & \boldsymbol{f_{i,p}}(\vec{u}) \\ & & & \vdots & & \\ f_{n,1}(\vec{u}) & f_{n,2}(\vec{u}) & \dots & f_{n,j}(\vec{u}) & \dots & f_{n,p}(\vec{u}) \end{bmatrix}$$

- *X*: HEAs database
- *n*: number of recordings = 59 × 8 = 472
- *p*: number of features = 90

Home Electric Appliances Database

HEAs' feature extraction

p=90 electrical features

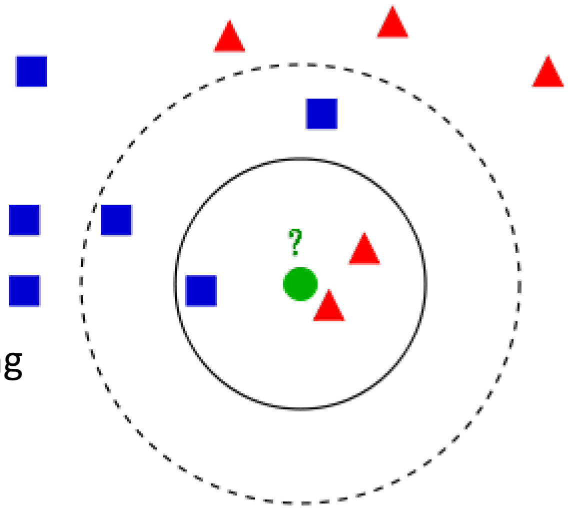
- Effective current $\boldsymbol{I}$ and its harmonics $\boldsymbol{I_k}$ for $k \in \{1,\ldots,15\}$ and $\boldsymbol{I_H}$ (A)
- Active power $\boldsymbol{P}$ and its harmonics $\boldsymbol{P_k}$ for $k \in \{1,\ldots,15\}$ and $\boldsymbol{P_H}$ (W)
- Reactive power $\boldsymbol{Q}$ and its harmonics $\boldsymbol{Q_k}$ for k ∈ {1,…,15} and $\boldsymbol{Q_H}$ (VAR)
- Apparent power S and its harmonics $\boldsymbol{S_k}$ for $k \in \{1,\ldots,15\}$ and $\boldsymbol{S_H}$, $\boldsymbol{S_N}$ (VA)
- Current harmonic distorsion $\mathrm{THD}_{\boldsymbol{I}}$
- Distorsion power $\boldsymbol{D}$, $\boldsymbol{D_I}$, $\boldsymbol{D_V}$ (VAD)
- Power factor $\boldsymbol{F_P}$, and its harmonics $\boldsymbol{F_{Pk}}$, for $k \in \{1,\ldots,15\}$
- Current crest factor $\mathrm{Fc}_{\boldsymbol{I}}$

Features Selection

Features Selection

Objectives

- Determination of the ($d \leq p$) most informative and relevant features for a correct grouping of n_k= 8 recordings of the same HEA within classes $k \in \{1, \ldots, 59\}$
    - Dataset stored in a 472×90 normalized matrix $\bar{X}$
    - Euclidean-based K-Nearest Neighbours (K-NN) classifier

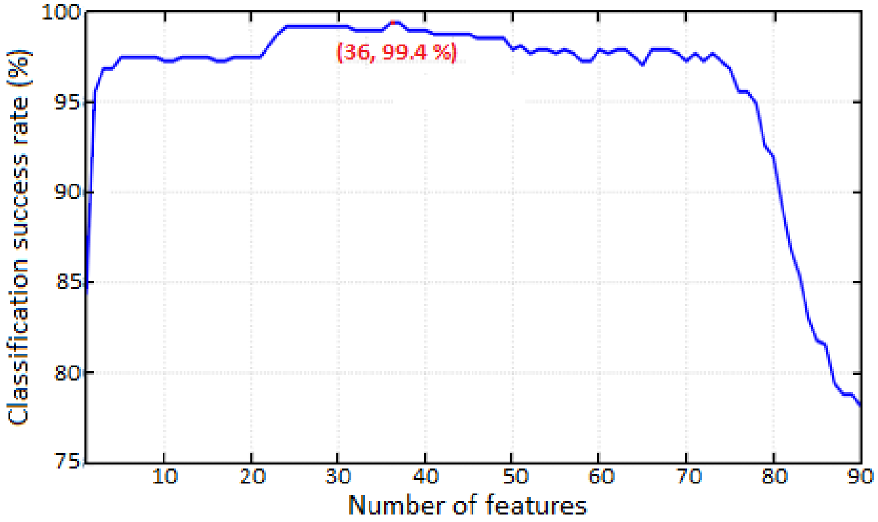


- 3 methods for selecting the most relevant subsets of features among the 90 extracted:
    - **Heuristic Forward Greedy Search**
    - **Principal Component Analysis** (PCA)
    - **Inertia Ratio Maximization using Feature Space Projection** (IRMFSP)

Features Selection

Heuristic Forward Greedy Search method

- Addition of features one at a time until the maximal classification success rate is reached



- Classification success rate= **99,4%**
- for *d*= **36** features:

$I, I_3, I_1, S, S_3, S_1, D_I, S_N, I_H, P_3, I_5, \mathrm{D}, P_7, S_H, S_7, Q_7, P_1, S_2, P, Q_1, Q_3, I_6, S_{15}, P_9, D_V, I_7, I_2, Q, I_4, P_2, S_{10}, I_8, I_9,, Q_4, Q_H, P_4$

- High computational cost
- Different result for other K-NN metrics

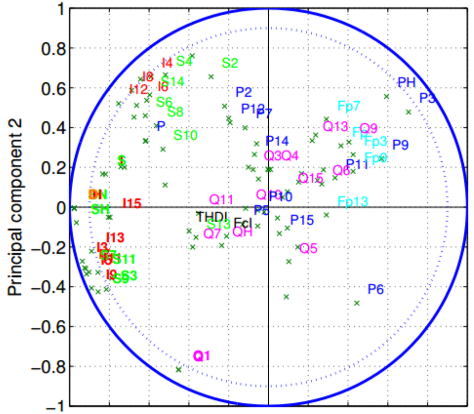
Features Selection

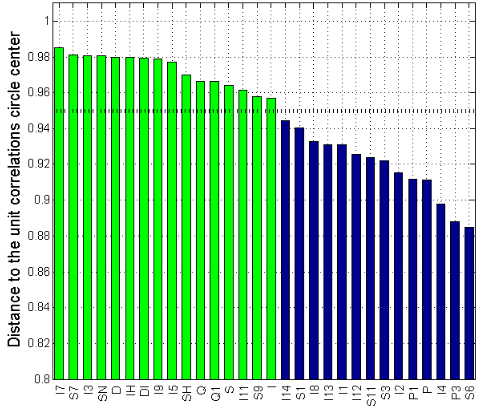
PCA method

- **Principal components**

p correlated features → Statistical conversion → *q<p* linearly decorrelated synthetical features

☞ *q = 4 principal components*





- Determination of the *d* most **correlated** original features to the *q = 4* principal components
  ☞ *d = 16*

→The closest ones to the circonference of the unit correlation circle *C*.

Features Selection

IRMFSP method

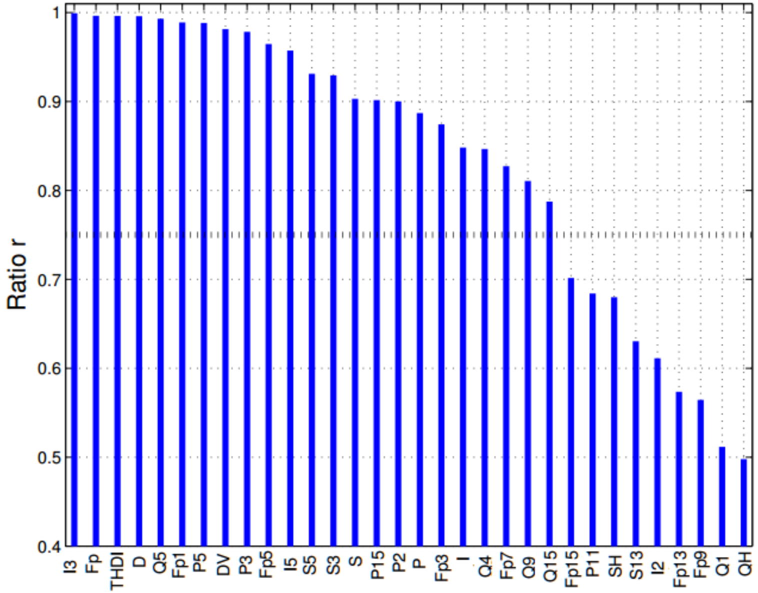
- **Maximization of the features subset relevancy** for the recognition task by:
  - Selection of features that maximize the ratio *r*

$$r(j) = \frac{\sum_{k=1}^{K} n_k (\mu_{j,k} - \mu_j)^2}{\sum_{i=1}^{n} (\overline{x_{ij}} - \mu_j)^2}$$

μ_j: center of gravity of the feature $j \in \{1, \dots, p\}$

$\mu_{j,k}$: center of gravity of feature j for data belonging to class $k \in \{1, \dots, K\}$

$\overline{x_{ij}}$: normalized value of feature j affected to individual $i \in \{1, \dots, n\}$



- **Minimization of features subset redundancy** by performing an iterative Gram-Schmidt orthogonalization process

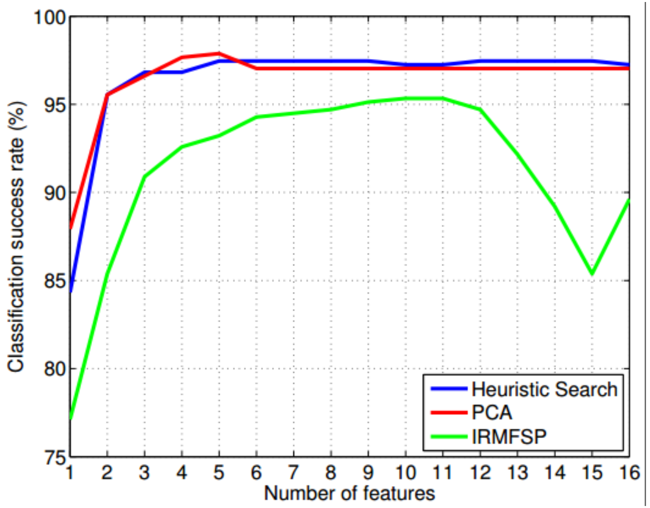
Features Selection

Comparison of the feature subsets

- Performance of the K-NN classifier across the home appliances dataset for different feature subsets.

	Subset size	Classification success rate
Complete feature set	90	78,2 %
Heuristic search set	36	99,4 %
PCA set	16	97,4 %
IRMFSP set	32	89,8 %
P, Q feature set	2	84,4 %

- Heuristic search and PCA subsets: **exponential functions** of the number of features → the addition of features brings **discrimination power** to the subset.



Conclusions

Prospectives

Conclusions

Non-Intrusive Load Monitoring

- Definition
- Goals and motivations
- Challenges

HEA database

- 472 HEAs recordings of one steady-state current and voltage period
- Extraction of p = 90 features

K-NN

Features selection

- Heuristic greedy search
- PCA
- IRMFSP

→Improvement of the performance of a designed classifier by the use of an **optimal subset of features**
→Justification of the retained features in the optimal subsets by the **power supply topologies** in HEAs
→Feature selection methods related to **inertia** notions

Prospectives

- Examination of other feature selection methods
- Validation of the performance of the retained feature subset by other types of classifiers
- Recognition of simultaneously connected HEAs
- HEAs' transient signals investigation

Thank you for your attention