Identification method for home electrical signal disaggregation

Luc Dufour, Dominique Genoud ,Gianluca Rizzo [1]
Bruno Ladevie and Jean-Jacques Bezian [2]

July 1, 2014

1 Institute of Information Systems, University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland
2 Mines Telecom, Albi, France
Corresponding Email:(Luc Dufour, Dominique Genoud, Pierre Roduit,Gianluca Rizzo)@hevs.ch,
(Bruno Ladevie and Jean-Jacques Bezian)@mines-albi.fr

ABSTRACT
In order to enable demand response schemes for residential and industrial users, it is crucial to be able to predict and monitor each component of the total power consumption of a household or of an industrial site over time. We used the cross-validation method which is a model validation technique for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. We exploit Non-Intrusive Load Monitoring (NILM) techniques in order to provide behavior patterns of the variables identified. This work presents a review Non-Intrusive Load Monitoring (NILM) techniques and describe the results of recognition patterns used for the identification of electrical devices.The proposed method has been validated on an experimental setting and using direct measurements of appliances consumption, proving that it allows achieving a high level of accuracy in load disaggregation.

KEYWORDS: *Data intelligence analysis; Microgrid; Energy information management; Advanced Metering Infrastructure;*

1. INTRODUCTION

The last decade has been characterized by a sharp increase in fuel prices and in energy demand. At the same time, energy consumption awareness has been raised over the impact which the greenhouse effect is having on the climate and on world economy. Furthermore, By 2020,there will be 7.5 billion on the globe and consumption will have increased by 75% (compared to 2000), equally split between developing and developed countries [0]. This means a 37.5% increase every 10 years. In Swiss,the electric consumption increased 120%in 50 years [0].
These factors have driven research of solutions for sustainability in energy production, distribution, storage, and consumption. In energy distribution, new "smart" solutions have been proposed, centered on the idea that exploiting properly data on power generation, distribution and consumption, a substantial increase in efficiency is achievable in power production and distribution.We can talk too about the objectives 3*20 in Europe : to decrease in 20% gas emissions with greenhouse effect, to reduce of 20% the energy consumption and to increase of 20% the production of renewable energy.
In 2010, the Chinese government is poised to invest more than 5.4 billion Euros in the development of Smart Grid technologies, while the United States has earmarked some 5.2 billion Euros. Meanwhile, the annual investment in Europe is estimated to be approximately 5 billion Euros [0].

At the user level, several load management and demand response techniques have been proposed, aiming at flattening the peaks of power consumption over time, and at adapting demand to variations in supply due to renewable.The techniques proposed to monitor appliance consumption, non-intrusive ones (called also NILM [0]) are of particular interest in households, since they do not require specialized, costly hardware and installation and maintenance of a sensor network. Moreover, they adapt over time in changes in households (such as changes in appliance number and type) without requiring new installations or reconfiguration of existing hardware and software.
The extensive deployment of smart meters which is planned in many countries for the near future will enable a large scale deployment of NILM techniques [0]. Such deployment will make available measurements of the total active and reactive power consumed, typically sampled at low frequencies, allowing non-intrusive load monitoring without the use of additional hardware.

NILM methods have been first proposed in [0], and they are typically structured in three phases: feature extraction, events detection, and events classification. They make use of a database of electric signatures of appliances, and they are based on the measure of the total active power consumed, sampled at frequency of one Hertz.
Later methods [0] [0] try to decrease the duration of the training period. Indeed, a fine granularity and a good accuracy in load disaggregation are crucial in order to enable useful feedback to users, to set up appropriate measures for changing consumption patterns, and to enable detection of anomalies and appliance malfunctioning. Many of the techniques proposed in order to overcome these drawbacks imply a substantially higher sampling frequency, and therefore expensive hardware [0] [0] [0].

In this paper,we used and test the different works on this subject (non-intrusive load monitoring [0] [0]) by the cross validation method to validate the inputs vectors in our global predictor model. In our small microgrid (one residential building), we have electrical data, weather data, production of electricity by photovoltaic panel, electricity consumption by devices and the results of a thermal analysis. Our final aim is to predict the global energy consumption and to pilot intelligent the residential building(with or without renewable production) to provide the good knowledge's to the control management system. we focus on the electrical data and in particular the global load curve and the electrical consumption by devices. We chose three devices that we consider like interesting : the dishwasher, the heat pump and a freezer. This three devices can be moved in a predetermined time. For example, the heat pump can move for one hour without lose the comfort of users with the inertial mass of the building.
We realized three tests to generalize the problem of identification of devices. We have a classification for the test-set validation about 99% when we train and we test with the same house for three devices. For the others tests, we have a classification about 94% when we train the sames devices of the two houses for two weeks.We test on the two next weeks. Finally, we have a classification about 98% when we train the devices of one house for two weeks. We test with an other house for the same weeks.
The paper is organized as follows. In Section II we present our cross validation method and detail the algorithm for load disaggregation. In Section III we describe the setting used for the test and in Section IV we present the results of the different algorithms. Finally we conclude and discuss future directions of research in Section V.

2. CROSS VALIDATION METHOD

In this section, we describe the data required for our cross Validation method and the prediction model to classify the started devices in use the global power curve. The Cross-validation method can be used to compare the performances of different

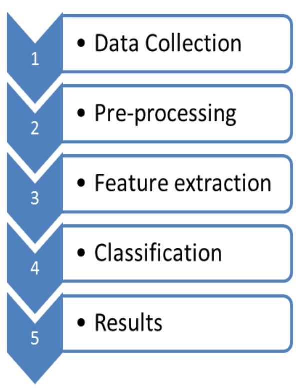

Figure 1: Cross Validation Method

predictive modeling procedures and to select the inputs variables.For the test set validation, we partitioning the sample of size n into training data set and a testing set. The model is built on the training set and validated on the testing set. The error is estimated by calculating a test, a measure and a score of performance of the model on the testing set.

Data Collection: Our method requires data for every second for the active power and reactive power by devices. This allows us to understand the dependency of the target variable to the input vectors. The target variable are the difference between two active power points and in a second part,the difference between two reactive power points.

Pre-Processing: We begin by filtering the interferences that we must identify on the global diagrams as well as on individual connected appliances diagrams.To reduce the effects of noise and transients on the measured data, we apply a median

filter of size 60 to the time series of measured values.We have empirically determined the optimal size of the filter via a set of tests, by attempting to balance the elimination of spurious load level variations and the performance of the event detection algorithm, in terms of average number of undetected events.
We use a median filter on each entry signal to remove high frequencies. This is used on the global load curve for each phase and on the different devices' signals. The threshold used depends mainly on the type of device that we want to detect.If the objective is to detect all devices, the active threshold must be the finest possible, but the variations detected can lead to errors corresponding to residual noise.

Feature Extraction: We have in inputs three values every second for the voltage, three for the amperage, three for the active and the reactive power, three for the power factor and the accumulated active energy and a time reference for the global load curve. We have too in inputs the active power and the time reference for the different devices connected inside the house. We use edge detector method , which consists on detecting variations of active and reactive power with regards to a predefined threshold. [0]. The different work on this subject and our analysis working enable to know the different characteristics to describe a device [0], [0] [0].
We have now a list of variations of active and reactive powers variations by phases which represent the different devices connected in the test house. A time level recognition is realized to group the potential vectors (/) for the three identified devices.

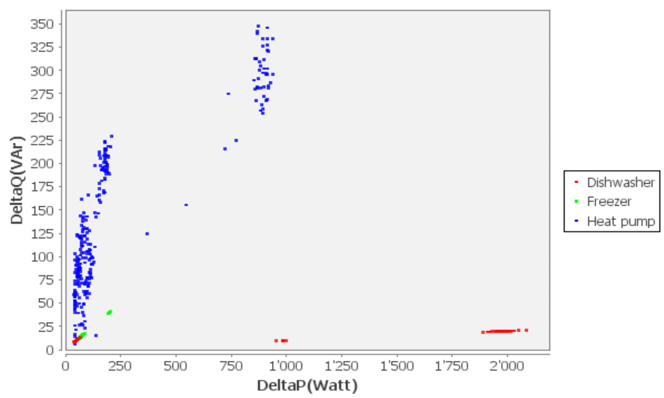

Figure 2: Active and Reactive power vectors detected on phase 1 for three devices, (Dishwasher,Freezer and Heat pump),11 February 2013 - 17 February 2013

Classification: We use a random sampling of all active/reactive power vectors for the partitioning: the input table is split into two partitions,train and test data. To estimate the training data, we use three known classification algorithms : k-Nearest Neighbor(KNN) [0], Support Vector Machine (SVM) [0], and a decision tree(DT) [0].
The nearest neighbor algorithm derives the class labels or the numeric target values of new input objects from the most similar training examples, where similarity is measured by distance in the feature space. The prediction is computed by a majority vote of the nearest neighbors or by averaging their numeric target values.
The kNN allows setting the number of neighbors to be considered and if the distance should be used to weigh in on those neighbors. For the SVM, our analysis consists of two steps: one learning the SVM model offering the choice of a few well known kernels and the second one allowing to apply the model to a second data set.
The kNN and the SVM are very sensitive to the chosen distance function so we should make sure to normalize the data and use the exact same normalization procedure for both the training and test data.This can be achieved by using the normalizer function and Normalizer (Apply) function , which copies the settings from the first node. We then feed those two data tables into the K Nearest Neighbor node which adds a column with the predicted class to the test data.
The native SVM implementation currently offers three different kernels functions: a polynomial, hyper tangent and RBF kernel. The Kernel functions (or at least the settings of their respective parameters) are rather sensitive to the normalization of the input data. For instance, the sigma parameter of the RBF kernels controls the width of the basis functions in relation to the Euclidean Distance of the input vectors to the support vectors. Hence it is critical to adjust this parameter accordingly.
For the DT,a number of options are available like the information gain Gini index and Gain ratio, the pruning method like a minimum description length (MDL) [0], the number threads .

Results: We use a random sampling of all active/reactive power vectors for the partitioning: the input table is split into two partitions which depends of our different tests. this tests are explain in Experimental setting part.In order to evaluate the classification by scored, we define a contingency matrix representing the possible outcomes of the classification, namely the true positives (TP positive examples classified as positive), the True Negatives (TN negative examples classified as negative), the False Positives (FP negative examples classified as positive) and the False Negative (FN positive samples classified as negative). The recall is defined as TP/(TP+FN) and the precision is TP/TP+FP.

3. SETTING

In this section, we describe how our information system collects data, how the parameters are used in our analysis software and how the data set for prediction model.

Data Collection: We use the schneider electric system to collect active and reactive power by phase on the global electric meter [0]. This smart meter is the PowerLogic Series 800 PM810 of the Schneider Electrical Company [0]. As outputs, we have the amperage, voltage, active and reactive power and energy consumed with in a one second interval per phase. We use the same device to collect data from the principal Heat pump. We also have amperage, voltage, active and reactive power per second for each phase for the heat pump 1.
The different devices of houses are connected using the Geroco smart plugs. This smart meter collects the amperage, voltage, the active power and timestamps. The protocol of communication used is a Zigbee [0]. with this technology, we can define activation and dis-activation of one or many devices. An integrated pre-programmable code enables the recognition of the different devices.
The data travels through a modbus communication.These datas are stored on mini-pc in csv files and to send on server in HES SO at Sion. We use the open source data analysis software KNIME [0] to connect the database and process the data. At this point, we have two years of data (November 2012- November 2014) for the global charge charts and the devices connected. 10houses are today connected.

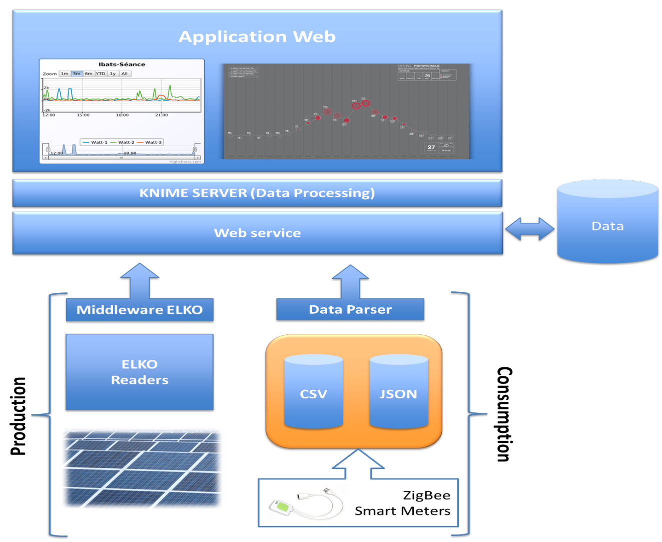

Figure 3: I-BAT information system architecture based No-SQL databases

Feature Extraction: We tested two thresholds for the active and reactive power to 50W/15VAR and 40W/5VAR to study the impact of transitional effects on electrical signals. If we use a median filter with a size of 60 and an edge detector of 40W/5VAR, we detect all variations for the three devices (heat pump, refrigerator/freezer, dishwasher). The active threshold, 40Watt or 50 Watt, is sufficient to detect a dishwasher, a freezer or a heat pump on/off events. But the reactive power threshold is too high to detect a fridge or freezer after the median filter of size 60 for the global load curve. That is why we used a 40W and 5VAR threshold in our methodology. The devices not detected with this active power threshold from the global load curve are the modem and the computers.

Classification: We used two similar Swiss houses to study the classification of devices for the energy management. We selected the same devices for the two houses : a heat pump, a dishwasher and a freezer. We realized three tests in order to evaluate our training set.
The training set vectors are normalized by Min-Max Normalization Method. The same model of normalization is applied for the test set. The three values normalized are the time reference, the active power and the reactive power events. For the kNN,

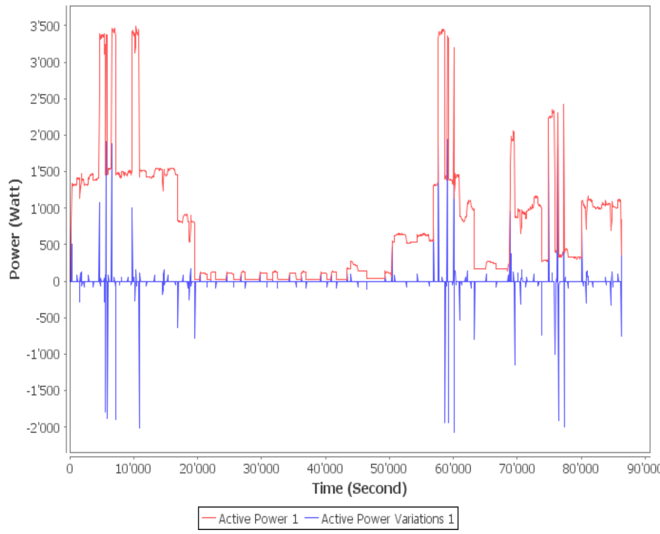

Figure 4: Active power variations on phase 1, 15 February 2013

we fix k = 3 and we used the weight neighbors by distance. For the DT, the quality measures for split calculation is the gini index. For the SVM, we don't case so we must to find an new plan in order to separate the different clusters. The kernel function is the radial basic function(RBF).

For the test 1, we used two random weeks (1 November 2012 at 15 November 2012) and we used a training set of 60% and a testing set of 40%. We chose this partitioning to balance the number of cycles studied by device.
For the test 2, we take the same weeks for the training set (1 November 2012 at 15 November 2012)and we take the two next weeks for the testing dataset (15 November 2012 at 30 November 2012). We train and we test for the two houses.
For the test 3, we take the two same weeks for the training set and the testing dataset (15 November 2012 at 30 November 2012) but we train on the house 1 and we test on the house 2.

- test 1 : 60 % for the training set, 40 % for the testing set , house 2
- test 2 : training set (house 1 + 2) for two weeks, the testing set (house 1 + 2) for two next weeks
- test 3 : training set (house 1), testing set (house2)

4. RESULTS AND DISCUSSION

In this section, we discuss of our results of the cross validation method. We checking if the active and reactive power events enable the identification a class of devices or a state of a device. We test if we can classify a device like a dishwasher with a dishwasher of an other house.
Test 1: We use two random weeks in the winter season to study the activation of the heat pump for our data set(1 November 2012 at 15 November 2012).We use a training set of 60% and a test set of 40%.

Table 1: Average Results by algorithms with Active power data set for two weeks, house 2

3-NN	SVM	DT
87.4 $\pm$1.6	81.8 $\pm$2.8	90.6 $\pm$3.1

Table 2: Average Results by algorithms with Active power and Reactive power data set for two weeks, house 2

3-NN	SVM	DT
95.7 $\pm$1.6	96 $\pm$2	94.1 $\pm$2.1

Table 3: Accuracy statistics by devices with Active power and Reactive power data set for two weeks, house 2, KNN algorithm

Devices	TP	FP	TN	FN	Recall	Precision	F-measure
Dishwasher	175	8	1207	40	0.814	0.956	0.879
Freezer	406	48	965	11	0.974	0.894	0.932
Heat pump	787	6	626	11	0.986	0.992	0.989

Table 4:Accuracy statistics by devices with Active power and Reactive power data set for two weeks, house 2,DT algorithm

Devices	TP	FP	TN	FN	Recall	Precision	F-measure
Dishwasher	177	38	1201	14	0.927	0.823	0.872
Freezer	417	0	996	17	0.961	1	0.98
Heat pump	779	19	606	26	0.968	0.976	0.972

Table 5:Accuracy statistics by devices with Active power and Reactive power data set for two weeks, house 2,SVM algorithm

Devices	TP	FP	TN	FN	Recall	Precision	F-measure
Dishwasher	146	64	1219	1	0.993	0.695	0.818
Freezer	401	0	966	63	0.864	1	0.927
Heat pump	799	20	591	20	0.976	0.976	0.976

We know that the three devices tested was very different. However, all these devices are a pump which has a same form : it is one or two "'on"' events and one or two "'off"' events in a cycle. In fact, the results show if the different algorithms enable to associate the pumps with the associated devices.

For this house, our results are suitable if we used the active power and the reactive power : the SVM is the best algorithm to classify the dishwasher and the heat pump. The DT is very interesting if we use only the active power, in particular to classify the freezer. The addition of reactive power events don't change the results of the DT algorithm. However, it's very important for the SVM because we have with an increase around 20%. between this two tests (with and without the reactive power)
We can conclude that the performed methods are suitable and that the active and reactive power are the pattern which described the different devices. Moreover, the composed signature can be an accurate description for each of the appliances in the data set. Nevertheless, the small number of electrical appliances might be a limitation that should be further considered.

Test 2: We used the same period for the training set (1 November 2012 at 15 November 2012) and we take the two next weeks for the testing set (15 November 2012 at 30 November 2012). We used the dishwashers, the freezers and the heat pumps of two test houses.

Table 6:Average Results by algorithms with Reactive and Active power data set for two weeks for the training set (1 November 2012 at 15 November 2012), two weeks for the testing set (15 November 2012 at 30 November 2012), house 1 +2

3-NN	SVM	DT
89.8 $\pm$2	93.3 $\pm$1.1	89.7 $\pm$3.6

Table 7: Accuracy statistics for two weeks for the training set (1 November 2012 at 15 November 2012), two weeks for the testing set (15 November 2012 at 30 November 2012), house 1 +2, KNN algorithm

Devices	TP	FP	TN	FN	Recall	Precision	F-measure
Dishwasher	839	208	8852	441	0.655	0.801	0.721
Freezer	2978	7686	6382	212	0.934	0.795	0.859
Heat pump	5472	75	4395	398	0.932	0.986	0.959

Table 8: Accuracy statistics for two weeks for the training set (1 November 2012 at 15 November 2012), two weeks for the testing set (15 November 2012 at 30 November 2012), house 1 + 2, DT algorithm

Devices	TP	FP	TN	FN	Recall	Precision	F-measure
Dishwasher	542	738	8949	111	0.83	0.423	0.561
Freezer	3070	120	6394	756	0.802	0.962	0.875
Heat pump	5663	207	4272	198	0.966	0.965	0.965

Table 9: Accuracy statistics for two weeks for the training set (1 November 2012 at 15 November 2012), two weeks for the testing set (15 November 2012 at 30 November 2012), house 1 +2, SVM algorithm

Devices	TP	FP	TN	FN	Recall	Precision	F-measure
Dishwasher	248	150	2922	0	1	0.623	0.768
Freezer	999	0	2115	206	0.829	1	0.907
Heat pump	1850	73	1380	17	0.991	0.962	0.976

The results are suitable for the freezers and the heat pumps, in particular with the SVM. The results decrease in this test for the dishwashers because the different algorithms confuse the dishwashers (in fact, the pump of the dishwasher) and the freezers.The pump of the dishwashers of the house 2 has more power events between 20Watt and 80 Watt. The active and reactive power of these pumps depends of the installation of the house and of the type of function (there are several modes in function of the temperature for the dishwasher). For the heat pumps, the different events (active and reactive power) are most important that a freezer. The reactive power is crucial to differentiate this two devices. The pump of the dishwasher is generally lower.

Test 3: We want to generalize our classification problem. We used the same weeks for the training and testing set (1 November 2012 at 15 November 2012). We train with the three data set of three devices from the house 1 and we test with three devices of the house 2.

Table 10: Average Results by algorithms with Reactive and Active power data set for two weeks for the training set (1 November 2012 at 15 November 2012) House 1 , two weeks for the testing set (1 November 2012 at 15 November 2012), house 2

3-NN	SVM	DT
92.2 $\pm$1.2	93.5 $\pm$1.1	75.2 $\pm$2.7

Table 11:Accuracy statistics with Active power and reactive power data set for two weeks for the training set (1 November 2012 at 15 November 2012) House 1 , two weeks for the testing set (1 November 2012 at 15 November 2012), house 2, KNN algorithm

Devices	TP	FP	TN	FN	Recall	Precision	F-measure
Dishwasher	119	0	1231	0	1	1	1
Freezer	407	14	925	4	0.99	0.985	0.978
Heat pump	806	4	526	14	0.983	0.992	0.989

Table 12: Accuracy statistics with Active power and reactive power data set for two weeks for the training set (1 November 2012 at 15 November 2012) House 1 , two weeks for the testing set (1 November 2012 at 15 November 2012), house 2, DT algorithm

Devices	TP	FP	TN	FN	Recall	Precision	F-measure
Dishwasher	119	0	1231	0	1	1	1
Freezer	407	14	925	4	0.99	0.985	0.978
Heat pump	806	4	526	14	0.983	0.992	0.989

Table 13: Accuracy statistics with Active power and reactive power data set for two weeks for the training set (1 November 2012 at 15 November 2012) House 1 , two weeks for the testing set (1 November 2012 at 15 November 2012), house 2, SVM algorithm

Devices	TP	FP	TN	FN	Recall	Precision	F-measure
Dishwasher	119	0	1231	0	1	1	1
Freezer	407	14	925	4	0.99	0.985	0.978
Heat pump	806	4	526	14	0.983	0.992	0.989

The results decrease in this test because the different models confuse the dishwashers (in fact, the pump of the dishwasher) and the freezers.The pumps of the dishwashers of the house 2 has more power variations (between 20Watt and 80 Watt). So, For the heat pumps, the different variations are most important that the dishwasher and a freezer even if in sometimes the pumps of the heat pump can be confuse with the two others pumps.

5. CONCLUSION

We propose a cross validation method which uses active and reactive power to differentiate three devices : the heat pumps,the dishwashers and the freezers. We can conclude that the performed methods are suitable and that the active and reactive power are the pattern which described this different devices. Moreover, the composed signature can be an accurate description for each of the appliances in the data set. Nevertheless, the small number of electrical appliances might be a limitation that should be further considered.

Our different tests show that the problem of recognition of devices may be generalize : we can use the devices of a house to train a model of an other house. Our results are suitable if we used the active power and reactive power and the SVM is the best algorithm to classify the dishwasher and the heat pump. We can notify that this two devices are the most interesting to pilot the electrical consumption. For example, the maximum total flexible energy is 3.4 Kwh when the two heat pumps and the dishwasher are started. In decentralized electrical production, like in a microgrid, this potential of flexible energy is important.

We have more than 96% percent of classification if we train and test on the same house. We have 92% if we train the devices of the devices of an other houses. The results are encouraging and we must focus our job on the association between the different pumps detected and a device. The pumps of the dishwashers and the heat pumps are more powerful than that of the freezer. The reactive power of the pumps of the heat pump are more powerful than those of the refrigerator and dishwasher. It's normal in the residential case but it's may be different with an other activity sector like a restaurant.

Finally, we want deployed this methodology for 200 houses and test the classification method in on-line for the next paper.

ACKNOWLEDGMENTS

This research was undertaken as part of the I-BAT project and has been funded by The Ark Energy (Project No. 712-07). We would like to thank all the contributions from the partners : HES-SO (Milene Fauquex, Jerome Treboux and Calixte Mayoraz), Icare Institute (Francois Chabbey), Alro Communication SA (Frederic Revaz),Geroco SA (Eric Nussbaum), Sierre Energie (Laurent Antille, Jean-Marc Salamin, and Christophe Darioly)

REFERENCES

[1]Thomas F. Garrity, VP, Sales and Business Development, Siemens Power available.
URL = https://w3.energy.siemens.com/cms/us/whatsnew/Documents/Getting20%Smart$_{Garrity.pdf}$

[2]IEEE (2012) OFEN, Statistique suisse de l'lectricit en 2012. *The Edison Foundation Institute for Electric Efficiency*. URL = http://www.strom.ch/fr/dossiers/graphiques-electricite.html.

[3]Reported by the Gerson Lehrman Group. URL = http://e360.yale.edu/content/digest.msp?id=2252.

[4]Zoha, A., Gluhak, A., Imran, M.A., Rajasegarar, S. (2012) Non-Intrusive Load Monitoring Approaches for Disaggregated Energy Sensing: A Survey. *Sensors*. vol. 12, no. 12, pp. 16838-16866.

[5]IEEE (2006) Utility-scale smart meter deployments, plans and proposals. *The Edison Foundation Institute for Electric Efficiency*.

[6]Hart, G.W. (1992) Nonintrusive appliance load monitoring. *Proceedings of the IEEE*. vol. 80, no. 12, pp. 1870-1891.

[7]Farinaccio, L. (1999) The disaggregation of whole-house electric load into the major end-uses using a rule-based pattern recognition algorithm. *Concordia University*, PhD Thesis.

[8]Marceau, M.L., Zmeureanu, R. (2000) Nonintrusive load disaggregation computer program to estimate the energy consumption of major end uses in residential buildings. *Energy Conversion and Management*. vol. 41, no. 13, pp. 1389-1403.

[9]Srinivasan, D. and Ng, W. S. and Liew, A.C. (2006) : Neural-network-based signature recognition for harmonic source identification, Vol. 21,no. 1,pp. 398-405.

[10]Patel, S., Robertson, T., Kientz, J., Reynolds, M., Abowd, G. (2007) At the Flick of a Switch: Detecting and Classifying Unique Electrical Events on the Residential Power Line. *UbiComp 2007: Ubiquitous Computing. Lecture Notes in Computer Science*. Vol. 4717, Springer Berlin Heidelberg, pp. 271-288.

[11]El Guedri, Mabrouka (2009) Caractérisation aveugle de la courbe de charge électrique: Détection, classification et estimation des usages dans les secteurs résidentiel et tertiaire

[12]Luan, S. W., Teng, J. H., Chan, S. Y., Hwang, L. C. (2009). Development of a smart power meter for AMI based on ZigBee communication. *In Power Electronics and Drive Systems, PEDS 2009. International Conference on*. pp. 661-665, IEEE.

[13]Froehlich, J., Larson, E., Gupta, S., Cohn, G., Reynolds, M., Patel, S. (2011). Disaggregated end-use energy sensing for the smart grid. *Pervasive Computing, IEEE*. Vol. 10, no. 1, pp. 28-39.

[14]T. Cover,P. Hart : Nearest Neighbor Pattern Classification(1967), journal=The Edison Foundation Institute for Electric Efficiency,*The Edison Foundation Institute for Electric Efficiency* pp. 21-27.

[15]Fast Training of Support Vector Machines using Sequential Minimal Optimization, URL = http://research.microsoft.com/en-us/um/people/jplatt/smo-book.pdf.

[16]John Shafer, Rakesh Agrawal, Manish Mehta,SPRINT, A scalable parallel classifier for data
URL = http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.104.152rep.pdf.

[17]Power meter series 800 data sheet,
URL = http://www2.schneider-electric.com/resources/sites/SCHNEIDER ELECTRIC/content/live/FAQS/140000/FA140345/en US/PM80020DataSheet.pdf.

[18]Power meter series 800 user guide,
URL = http://download.schneider-electric.com/files?p File Id=27600253p File Name=63230-500-225A2 PM800 User Guide EN.pdf.

[19]Knime,URL = http://knime.com/.

[20]Ecowizz,URL = https://www.ecowizz.net/.