RELEVANT FEATURE SELECTION FOR HOME APPLIANCES RECOGNITION

S. Houidi[1,2], F. Auger[1], H. Ben Attia Sethom[2], D. Fourer[1], L. Miègeville[1]

1. IREENA, CRTT, 37 Bd de l'Université, BP 406, 44602 Saint-Nazaire cedex, France.
e-mail : francois.auger@univ-nantes.fr,
2. Université de Tunis El Manar, École Nationale d'Ingénieurs de Tunis,
LR11ES15 Laboratoire des Systèmes Électriques, 1002, Tunis, Tunisia,
e-mail : houda.benattia@enicarthage.rnu.tn

Abstract - All over the world, the residential sector represents an important part in electrical energy consumption, and thus, is an opportunity to address substantial savings in terms of energy and money. In order to attempt this objective, a relevant knowledge of the appliances used in residential buildings is needed in order to better control or monitor energy consumption. This will be made possible through an effective automatic recognition of the home appliances. In this context, the main objective of this work is to be able to describe appliances as best as possible in order to recognize them individually, using features deduced from current and voltage measurements recorded at the grid connection point. In this paper, methods for selecting the most relevant features allowing the recognition of home appliances are proposed. The set up of a database of sampled measurements recorded on various home appliances types is also introduced.

Keyword - Home appliances, Non Intrusive Load Monitoring (NILM), automatic recognition, statistical data analysis, features selection.

1 INTRODUCTION

In a global trend of seeking for more power efficiency and more controllability over electrical energy consumption, home appliances recognition using voltage and current measurements at the grid connection point offers meaningful information to homeowners and grid managers to achieve a sustainable ascendency over domestic energy consumption. From the inhabitants' point of view, a relevant knowledge of home appliances in a non-intrusive way provides a breakdown of the energy expense per appliance to firstly identify greedy devices and then adjust and reduce energy consumption. From the grid manager point of view, receiving information about the appliances' usages helps to manage the energy distribution [1], and represents more possibilities of actions (such as off loading) in order to keep the grid stability and reliability, especially with the increasing integration of renewable energy sources. Therefore, recognizing the usage of each appliance is one of the core issues in the field of smart buildings energy management [2], [3]. The knowledge of the most relevant features is an important step to carry out such a purpose. Saitoh *et al.* [4] define six features for the home appliances recognition, whereas Kato *et al.* [5] use linear combination of features related to current. Both researches do not consider feature selection as a preprocessing step to home appliances recognition.

The aim of this work is the determination of a reduced number of discriminat features deduced from voltage and current measurements which enable an efficient automatic recognition of home appliances by a digital system using a low-cost processor with a low storage capacity. The first part of this paper concerns the acquisition of steady state voltage and current waveforms of electrical home appliances and the definition of the electrical features deduced from these measurements. Ninety features are extracted, related to the current and the power components. The second part of this paper deals with the application of feature selection algorithms to determine the most informative and relevant features that ensure a correct grouping of the recordings within classes of unique appliances. The third part of this paper discusses the results, before concluding and presenting future work at last.

2 ELECTRICAL FEATURES COMPUTED FROM CURRENT AND VOLTAGE RECORDINGS

2.1 HOME APPLIANCE PROFILING

The aim of this study is to explore the feasibility of recognizing home appliances from energy descriptors recorded in real-time. For this, the current and the voltage of a wide array of $K = 59$ home appliances of different kinds, ages, brands and power levels (such as fans, fridges, TVs, washer, etc.) have been recorded in steady-state conditions in a 50 Hz

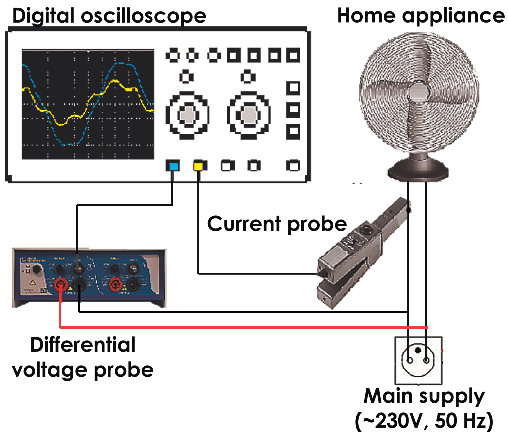

Fig. 1. Electrical assembly and instrumentation.

network. A current probe with a 10 mV/A sensitivity and a differential voltage probe with a 1/100 attenuation have been used to measure current and voltage respectively. For the signal acquisitions, an 8-bit resolution digital oscilloscope (RIGOL DS1104Z) has been used to acquire 2 and 6 periods of current and voltage signals in steady-state operating conditions at the sampling frequencies $f_{s_1} = 250$ kHz and $f_{s_2} = 50$ kHz (see Fig.1). Thus, 8 periods have been observed for each appliance signature, resulting in a set of $n = 8 \times 59 = 472$ individuals. Current, voltage and power features were then computed based on the latest IEEE 1459-2010 standard for the definition of single phase physical components under non sinusoidal conditions [6], [7].

2.2 Features computation

From the recorded voltage and current signals, the Fourier coefficients v_{ak}, v_{bk}, i_{ak} and i_{bk}, are first computed by:

$$v_{ak} = \frac{2}{M}\sum_{m=0}^{M-1} v[m]\cos(2\pi m k f_0/f_s) \quad (1)$$

$$v_{bk} = \frac{2}{M}\sum_{m=0}^{M-1} v[m]\sin(2\pi m k f_0/f_s) \quad (2)$$

$$i_{ak} = \frac{2}{M}\sum_{m=0}^{M-1} i[m]\cos(2\pi m k f_0/f_s) \quad (3)$$

$$i_{bk} = \frac{2}{M}\sum_{m=0}^{M-1} i[m]\sin(2\pi m k f_0/f_s) \quad (4)$$

with $v[m] = v(m/f_s)$, $f_0 = 50$ Hz and $M = f_s/f_0$. These Fourier coefficients are computed up to $k = 15$ range, and are considered as the most significant. They can also be expressed as:

$$v_{ak} = \sqrt{2}V_k\cos(\varphi_k), \quad v_{bk} = \sqrt{2}V_k\sin(\varphi_k) \quad (5)$$

$$i_{ak} = \sqrt{2}I_k\cos(\psi_k), \quad i_{bk} = \sqrt{2}V_k\sin(\psi_k) \quad (6)$$

where V_k, φ_k and I_k, ψ_k are the RMS and phase shifting values of the k^{th}-harmonic component of the voltage and the current respectively, $\forall k \in \{1,\ldots,15\}$. From these Fourier coefficients, the following features can be computed:

- The RMS value of the k^{th} harmonic component of the voltage V_k and currents I_k, and their sum V_H and I_H:

$$V_k = \sqrt{\frac{v_{ak}^2 + v_{bk}^2}{2}}, \quad I_k = \sqrt{\frac{i_{ak}^2 + i_{bk}^2}{2}} \quad (7)$$

$$V_H = \sqrt{\sum_{k=2}^{15} V_k^2}, \quad I_H = \sqrt{\sum_{k=2}^{15} I_k^2} \quad (8)$$

- The RMS voltage V and current I:

$$V = \sqrt{V_1^2 + V_H^2}, \quad I = \sqrt{I_1^2 + I_H^2} \quad (9)$$

- The k^{th} harmonic component of the active, reactive and apparent powers P_k, Q_k, S_k, and their sum P_H, Q_H, S_H:

$$P_k = \frac{1}{2}(v_{ak}i_{ak} + v_{bk}i_{bk}) \quad (10)$$

$$Q_k = \frac{1}{2}(v_{ak}i_{bk} - v_{bk}i_{ak}) \quad (11)$$

$$P_H = \sum_{k=2}^{15} P_k, \quad Q_H = \sum_{k=2}^{15} Q_k \quad (12)$$

$$S_k = V_k \times I_k, \quad S_H = V_H \times I_H \quad (13)$$

- The active, reactive apparent and distorsion powers P, Q, S and D:

$$P = P_1 + P_H, \quad Q = Q_1 + Q_H \quad (14)$$

$$S = V \times I \quad (15)$$

$$D = \sqrt{S^2 - P^2 - Q^2} \quad (16)$$

- The voltage and current total harmonic distortion THD_V and THD_I:

$$\mathrm{THD}_V = \frac{V_H}{V_1}, \quad \mathrm{THD}_I = \frac{I_H}{I_1} \quad (17)$$

- The voltage and current distorsion powers D_V and D_I:

$$D_V = S_1 \times \mathrm{THD}_V, \quad D_I = S_1 \times \mathrm{THD}_I \quad (18)$$

- The non fundamental apparent power S_N:

$$S_N = \sqrt{D_I^2 + D_V^2 + S_H^2} \quad (19)$$

- The voltage and current crest factors for $m \in [0, M-1]$:

$$F_{cv} = \frac{\max|v[m]|}{V}, \quad F_{ci} = \frac{\max|i[m]|}{I} \quad (20)$$

- Finally, the global, fundamental and harmonic power factors F_p and F_{pk}, computed by:

$$F_p = \frac{P}{S}, \quad F_{pk} = \frac{P_k}{S_k} \quad (21)$$

The features related to voltage are not considered, since they depend more on the power network than on the home appliance. As a result, $p = 90$ features can be computed at each voltage period. To perform an automatic recognition process with significant accuracy and speed, using a processor with cost and memory limitations, this number of features should be smaller. A selection of the most informative features is therefore considered.

3 Feature selection

3.1 Heuristic search

The dataset is stored in a 472×90 normalized matrix $\bar{X}$, where the 90 features detailed earlier are in columns and have been normalized with a zero mean and a unit standard deviation. Each class k ($k \in \{1, \ldots, K\}$) is composed of $n_k = 8$ acquisitions of the same home appliance. Given these 90 features and an Euclidean-based K-Nearest Neighbours (K-NN) classifier [8] with K_{NN} nearest neighbours, the aim is to find the optimal subset of $d \leq p$ features achieving a maximal performance of the classifier corresponding to the highest classification success rate (an individual is well classified when its K_{NN} closest neighbours are the other observed periods of the same home appliance, by excluding the tested individual). the number of nearest neighbours $K_{\text{NN}} = 7$ was chosen because it is the closest odd number to n_k so that for each neighbourhood there is a majority vote. A heuristic forward greedy search strategy has been applied to seek the best subset among all the possible $\binom{p}{d}$ feature subsets. It consists in adding features one at a time until the maximal classification success rate is reached. Increasing the size of the feature vector is theoretically expected to provide more discrimination power. Thus, the classifier performance depends on the number of features and on the classifier complexity. It should be an exponential function of the feature set dimension [9]. However, in practice, a subset of features makes the classification task more efficient. Fig. 2 shows the classifier success rate as a function of the used number of features. From this figure, the best classification rate of 99.4% is obtained using a subset of $d = 36$ features, which are ranked from the highest to lowest discrimination power:

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

According to Yu and Liu's definitions of relevant, irrelevant and redundant features [9], the relevant ones are those which increase this classification success rate. They correspond to I (which has single-handedly brought an accuracy increase of 85%), I_3, S_3, S_{13}, S_{15}, P_9, D_I and P_H. A decrease of the classification success rate is observed from the 77^{th} feature. Indeed, additional features which correspond to F_{pk} are redundant, so they take the role of relevant features such as S_k and P_k, and might possibly add more noise than useful information in the signature of home appliances. The flatness of the curve can be explained by the addition of irrelevant features such as the even order harmonic features, which are close to zero due to the symmetry of the recorded signals. They have no influence on the classification success rate. Notwithstanding the obtained result, it is based on a heuristic approach that leads to a solution that relies on the particular K-NN classifier which depends on the chosen metric. Indeed, the classification success rate as a function of the number of features changes depending on an Euclidean-based distance or a City Block-based distance [10] (see Fig.2) for measuring the distance between the individuals and the K_{NN} nearest neighbours. Moreover, this feature selection procedure is exhaustive by generating a high computational cost. Therefore, other approaches independent of the classification process, used as a pre-processing step are considered in the next section, based on statistical tools, intrinsic characteristics of the data and effectiveness regarding computational cost. We focus on two methods for selecting features: the unsupervized Principal Components Analysis (PCA) method [11], which relies on the dataset measure of dispersion, and the supervized Inertia Ratio Maximization using Feature Space Projection (IRMFSP) method [15], which uses class labels and inter-class inertia to minimize the redundancy among the selected features.

3.2 Principal Components Analysis

PCA is a statistical procedure which converts a set of n individuals described by p correlated features into a set of n individuals described by $q \leq p$ linearly decorrelated features called principal components which are linear combinations of the origi-

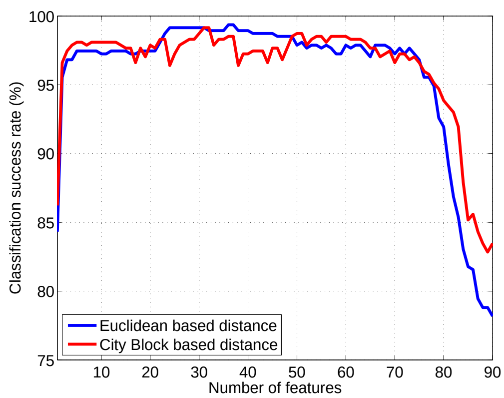

Fig. 2. Classification success rate as a function of the number of features.

nal features and contribute to the dispersion or inertia dataset [11]. Let $Y = \bar{X}^T\bar{X}$ be the $p \times p$ covariance matrix of $\bar{X}$, Λ the diagonal matrix of its p eigenvalues λ_j , and L the $p \times p$ matrix whose columns are the orthonormal eigenvectors of V such as $Y = L\Lambda L^T$. The subspace dimension q is then chosen to construct the submatrix $L_{p\times q}$ from L by retaining the q eigen vectors that represent more than 80% of the dataset cumulative inertia. The inertia $\mathcal{I}$ can be defined as the dispersion of the dataset on a principal component and can be computed by:

$$\mathcal{I}(j) = \frac{\lambda_j}{\sum_{j_a=1}^{p} \lambda_{j_a}} \times 100 \qquad (22)$$

and the cumulative inertia corresponds to:

$$\mathcal{I}_c(q) = \sum_{j=1}^{q} \mathcal{I}(j) \qquad (23)$$

The transformation $T_{n\times q} = \bar{X}L_{p\times q}$ maps the new coordinates of $\bar{X}$ to a new space of $q < p$ principal components which are uncorrelated over the dataset. Since each principal component is a linear combination of the original features, it is possible to determine which of the original features are the most correlated to a principal component by computing the correlation matrix W:

$$W_{p\times q} = \bar{X}^T T = \bar{X}^T\bar{X}L_{p\times q} = YL_{p\times q} \qquad (24)$$

The closer a correlation coefficient $w_{i,j}$ is to ±1, the more the original feature is correlated to the principal component [12].

A PCA is applied on the home appliances dataset. The importance of principal components is reflected by the cumulative inertia $\mathcal{I}_c$ or by the proportion of the total inertia "explained" by them. In our case, the cumulative inertia of the first $q = 4$ components corresponds to 87% of the total inertia (when $q = 3$, $\mathcal{I}_c = 78\%$ and when $q = 5$, $\mathcal{I}_c = 92\%$). To select the subset of d features, the correlation between a principal component and a feature is estimated. In what follows, a unit correlations circle $\mathcal{C}$ with center $\mathcal{O}$ having coordinates $(0, 0, 0, 0)$ is considered in the subspace of dimension $q = 4$. Features which are retained are the closest ones to the circonference of $\mathcal{C}$, and explain the largest part of the dataset inertia. Fig.3 depicts the distance of each of the 30 farthest features to $\mathcal{O}$ (for sake of clarity). Let $d = 16$ be the number of the selected features (highlighted in green in the figure) for which the distance from $\mathcal{O}$ is greater or equal to 0.95. The farthest a feature coordinates is to $\mathcal{O}$, the better this feature can be constructed from the four components. Conversely, the closer to $\mathcal{O}$, the less important a feature is for the four components. The features selected through PCA are for

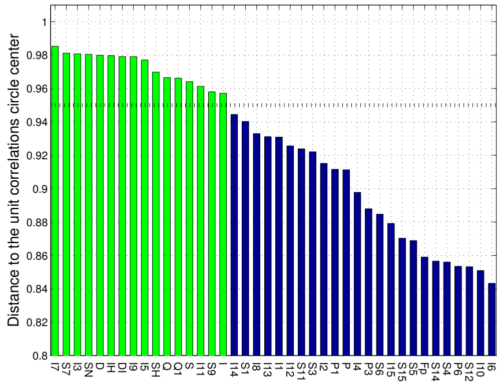

Fig. 3. Distance between each feature and the center of the unit correlations' circle .

the most odd-order harmonic and describe the nonlinearity of the home appliances current signals and thus, can be justified by the structure of the power supply included in home appliances [7].

3.3 Inertia Ratio Maximisation using Features Space Projection

The purpose of this algorithm is to maximize the relevance of the features subset for the recognition task while minimizing the redundancy between the selected features [13]. It starts by iteratively selecting features that maximize the ratio r between inter-class inertia and the total inertia. For a specific feature $j \in \{1, \ldots, p\}$, the ratio $r(j)$ is defined as:

$$r(j) = \frac{\sum_{k=1}^{K} n_k(\mu_{j,k} - \mu_j)^2}{\sum_{i=1}^{n} (\bar{x}_{ij} - \mu_j)^2} \qquad (25)$$

where μ_j is the center of gravity of the feature j over all the dataset (equal to 0 in our case since a normalized matrix $\bar{X}$ is considered), $\mu_{j,k}$ is the center of gravity of the feature j for data belonging to class k. $\bar{x}_{ij}$ is the normalized value of feature j affected to the individual i. Taking into account the fact that a feature with a high value of r may bring the same information as an already selected feature and is therefore redundant, the IRMFSP iteratively performs a Gram-Schmidt orthogonalization process [14] after the selection of each new feature f_j as follows:

$$f_{j_2}^{(m+1)} = f_{j_2}^{(m)} - ((f_{j_2}^{(m)})^T . f_j^{(m)}). f_j^{(m)}, \ \forall j_2 > j \qquad (26)$$

where $f_{j_2}^{(m)}$ is the j_2^{th} selected vector of feature at the iteration m. This process (ratio maximization followed by space projection) is repeated until the gain of adding a new feature becomes too small. This gain is measured by the ratio $r(l)$ obtained at the l^{th} iteration to the one at the first iteration [15]. Fig. 4 shows the $d = 32$ first features with a ratio $r \geq 0.5$ selected by the IRMFSP method. A stopping criterion

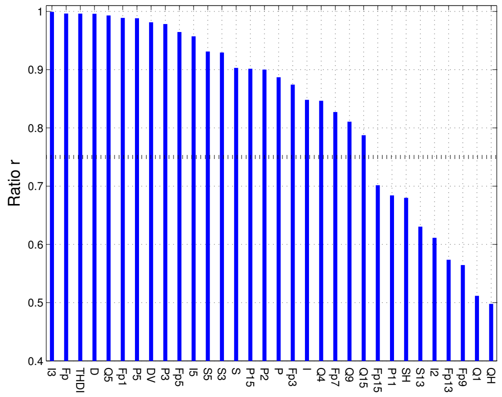

Fig. 4. Ratio r of each feature.

of $t = \frac{r(l)}{r(1)} < 0.01$ has been chosen. The features with a high value of r are those for which the classes are well separated. Most of the 21 features with a ratio $r > 0.75$ are odd-order harmonic features related to power components and current.

4 Comparison of the Feature subsets

In this section, the classification success rates when performing a K-NN classifier are compared with the different subsets of features: the ones obtained by the PCA method, by the IRMFSP method and the subset composed of active and reactive powers P and Q, which is commonly used in the literature for home appliances recognition [2],[5].

4.1 Experimental Results

The different feature subsets were determined from the original dataset according to the two methods detailed in Section 3. A new matrix for the PCA subset of features and other matrices for the IRMFSP subset of features and the P,Q subset of features were computed. The three matrices differ only in the amount of columns (features). An Euclidean distance based K-NN classifier was then performed on the matrices. The classification success rate is measured by using an 8-fold cross validation. This experiment splits randomly the data set into 8 equal subsamples of 59 individuals (no collision between the subsamples). One subsample is kept for validating the data, while the remaining 7 subsamples are used for training. This process is repeated 8 times, until all subsamples have been used as validation [16]. The classification success rates are averaged across to produce a single result. Table 1 shows the classification success rate across the home appliances dataset for the complete feature set and the feature subsets determined earlier. A comparison of the obtained results shows that the best accuracy is achieved by the PCA subset of features. Moreover, when looking at the size of the subsets it can be noticed that the subset of features selected by the PCA method is relatively small. Fig. 5 shows the classification success rate brought by the addition of each of the feature retained by the different methods, using the same approach as in Subsection 3.1 for the same number of features. The heuristic search and PCA subsets are exponential functions of the number of features which shows that the addition of features brings discrimination power to the subset.

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

feature subset	subset size	classification success rate
Complete feature set	90	78.2 %
Heuristic search feature set	36	99.4 %
IRMFSP feature set	32	89.8 %
PCA feature set	16	97.4 %
P, Q feature set	2	84.4 %

4.2 Discussion

In these experiments, a selection of features was performed in accordance with the physical framework by taking into account the real voltage and the current measurements and a covering broad array of home appliances for the features computation, with a view to engage in valid target setting. The PCA method for feature selection has provided the best result as it is based on dispersion measurement, which goes along with our aim. However, this work was only based on the use of the K-NN classifier (what is among the simplest methods). Therefore the performance of the retained feature subset should also be validated by other types of classifiers. In our experiments, we limited our scope to the selection of features using methods related to inertia notions. Nevertheless, the combinations of individually good features does not necessarily lead to good classification performance. In other words, “the d selected features are not the best d features for classification” [9]. Other methods taking into account the dependencies between features should also be studied [17].

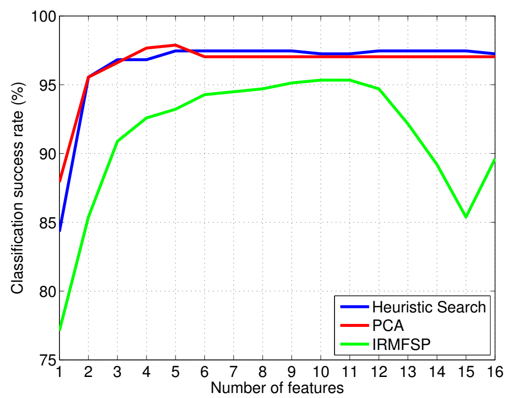

Fig. 5. Classification success rate as a function of the number of features selected by the different methods for $d \in \{1, \ldots, 16\}$ features.

5 Conclusion and future work

In this paper, we considered the selection of relevant features for home appliances recognition. First, we introduced an extraction of energy descriptors for home appliances recognition, using electrical current and voltage signals of a covering broad array of different types of home appliances. Then, methods for selecting the most informative features in a preprocessing step are detailed since the number of extracted features should be smaller to perform home appliances recognition on a processor with cost and memory limitations. In our experiments, the proposed methods were evaluated on a home appliances dataset and showed that the feature subset determined by the PCA method achieves an accurate classification rate with a small features' number. Two conclusions can safely be drawn from this study. Firstly, the performance of a designed classifier can be improved by the use of an optimal subset of features. Secondly, the retained features in the optimal subsets can be justified by the power supply topologies included in home appliance. In future works, the recognition of simultaneously connected home appliances' classes and their state change detection using transient signals will be investigated.

6. References

[1] K. Basu, V. Debusschere, S. Bacha, "Non Intrusive Load Monitoring : A Temporal Multi-Label Classification Approach", IEEE Transactions on Industrial Informatics, Vol. 11, No. 1, pp. 262–270, 2015.

[2] M. Zeifman, K. Roth, "Non intrusive appliance load monitoring : Review and outlook", IEEE Transactions on Consumer Electronics, Vol. 57, No. 1, pp. 76–84, 2011.

[3] G. Kalogridis, C. Efthymiou, S.Z. Denic, R. Cepeda, "Privacy for smart meters : Towards undetectable appliance load signatures", Proceeding of the first IEEE Int. Conf. on Smart Grid Communications (SmartGridComm), pp. 232–237, USA, 2010.

[4] T. Saitoh, T. Osaki, R. Konishi, K. Sugahara, "Current sensor based home appliance and state of appliance recognition", SICE Journal of Control, Measurement, and System Integration, Vol. 3, No. 2, pp. 86–93, 2010.

[5] T. Kato, H. Sang Cho, D. Lee, T. Toyomura, T. Yamazaki, "Appliance Recognition from Electric Current Signals for Information-Energy Integrated Network in Home Environments", Proceeding of the 7^{th} Int. Conf. on Smart Homes and Health Telematics ICOST, pp.150–157, France, 2009.

[6] IEEE standard 1459, "IEEE standard definition for the measurement of electric power quantities under sinusoidal, non sinusoidal, balanced and unbalanced conditions", Revision of IEEE Std 1459-2000, pp.1–40, 2010.

[7] A. E. Emanuel, "On the Assessment of Harmonic Pollution", IEEE Transactions on Power Delivery, Vol. 10, No. 3, pp.1693–98, 1995.

[8] X. Wu, V. Kumar, "Top 10 algorithms in data mining", Survey paper, IEEE Int. Conf. on Data Mining (ICDM) for knowledge and information systems, Vol. 14, Issue 1, pp. 22–25, Italy, 2008.

[9] L. Yu, H. Liu, "Efficient feature selection via analysis of relevance and redundancy", Journal of Machine Learning Research, Vol. 5, pp. 1205–1224, 2004.

[10] K. Chomboon, P. Chujai, P. Teerarassamee, K. Kerdprasop, N. Kerdprasop, "An empirical study of distance metrics for k-Nearest Neighbour algorithm", Proceedings of the 3^{rd} Int. Conf. on Industrial Application Engineering, pp. 280–285, Japan, 2015.

[11] I. Guyon, A. Elisseeff, "An introduction to variable and feature selection", Journal of Machine Learning Research 3, pp. 1157–1182, 2003.

[12] C. Boutsidis, M. Mahoney, P. Drineas, "Unsupervised feature selection for principal components analysis", Proceedings of the 14^{th} ACM SIGKDD Int. Conf. on knowledge discovery and data mining, USA, 2008.

[13] S. Essid, G. Richard, B. David, "Musical instrument recognition based on class pairwise feature selection", Proceedings of th 5^{th} Int. Conf. Society for Music Information Retrieval (ISMIR), Japan, 2004.

[14] G. Peeters, "Automatic classification of large musical instrument databases using hierarchical classifiers with inertia ratio maximization", 115^{th} AES convention, USA, 2003.

[15] D. Fourer, J-L. Rouas, P. Hanna, M. Robine, "Automatic timbre classification of ethnomusicological audio recordings", Proceedings of the 15^{th} Int. Conf. Society for Music Information Retrieval (ISMIR), Taiwan, 2014.

[16] S. Makonin, F. Popowich, "Non intrusive load monitoring (NILM) performance evaluation : A unified approach for accuracy reporting", Energy Efficiency, Vol. 8, No. 4, pp. 809–814, 2015.

[17] S. Alelyani, J. Tang, H. Liu, "Feature selection for clustering: A review", Charu Aggarwal and Chandan Reddy, CRC Press, 2013.