Appliance-Level Demand Identification Through Signature Analysis

W.A.K. Dhananjaya, R.M.M.R. Rathnayake, S.C.J. Samarathunga, C.I. Senanayake and N. Wickramarachchi
Department of Electrical Engineering
University of Moratuwa
Katubedda, Sri Lanka
kumudithdhananjaya@gmail.com, mrrmahesh@gmail.com, chanakascjs@gmail.com, chathurainduma@gmail.com and wick@elect.mrt.ac.lk

***Abstract* – Appliance specific load monitoring is very useful in energy management solutions that are becoming a challenging task with growing energy demand. It facilitates appliance recognition and load monitoring such that optimum resource utilization can be achieved by correct appliance scheduling. In this paper we present a study of non-intrusive load recognition using steady state appliance signatures for identifying commonly used household appliances. Current harmonics, active and reactive power components acquired from data loggers are used as appliance signature in this study. This analysis enables the capability of providing detailed information on appliances in use and consumers could benefit from customized energy management recommendations. Also, suppliers could implement smart metering technology introducing appliance level information as well. We propose algorithms for non-intrusive load recognition using combination of several methods and techniques. It was seen that a higher accuracy of identification could be achieved when a combination of techniques are used rather than using a single technique.**

Keywords- appliance signature; demand identification; harmonic analysis; signature analysis.

I. INTRODUCTION

Today, energy efficiency is a major public concern. It is known that energy can be saved through behavioral changes of the customers. Saving energy is not only financially profitable but also helps to reduce the pollutant emissions too. Load recognition and monitoring are essential in order to achieve right behavioral patterns. In this study signature analysis techniques are used for appliance-level load identification.

Signatures of different appliances can be used to identify each appliance when several appliances are working simultaneously. Parameters of current harmonics, active and reactive power components were selected as the signature of appliances in our study. These parameters of the test appliances were initially recorded and then appliances which are working at any time were identified by analyzing recorded appliance signatures.

Nonintrusive load monitoring (NILM) is the process of determining what the appliances are working at any time in the house as well as their individual energy consumption. This is achieved by analyzing changes in current and voltage profiles of that household. The primary benefit of NILM is that there is no need to install sensors for each appliance and no access to each individual appliance is needed. Here a single sensor was used that is installed at the point of customer unit of a household. So, it is a cost effective method compared to sensing individual appliances separately. Supply current and the voltage wave form were measured for this analysis and a Data logger (YOKOGAWA® CW240) was used to acquire measurements. Smart meters also provide information which are useful to energy management. But it measures energy consumption on a house-level and not on appliance-level. It is more complex to provide information on appliance-level such as which appliances have been switched on, how much energy is consumed by each appliance and how long they are operated. In order to provide that type of detailed information, appliance identification is essential.

In this paper household appliances are the main concern. So, as the test appliances both low power appliances as well as high power appliance were used. Appliances such as CFL lamps, incandescent lamps, fluorescent lamps, notebook computers, and electric fans were tested in the low power category. Heaters, electric irons, refrigerators and some kitchen equipment were tested in the high power category. Approximately the same signature characteristics are shown by some of them. So, variability and overlap of “signatures” are the challenges in this research area. And also while a high power appliance is working, it is difficult to identify weather a low power appliance is switched on or not. A significant change in current waveform cannot be seen at that type of situations. In order to get better performance, an algorithm was introduced by combining several methods as described later in this paper. It could be seen that some appliances can be identified with higher accuracy using one particular method than the others. So, when the combined algorithm is used it can improve the accuracy of the overall results. Our contribution to this research area is as follows.

978-1-4799-1740-2/15/$31.00 ©2015 IEEE

i. Methods are presented that can identify appliances which are in isolated operation.

ii. Few algorithms are presented that can be used to identify multiple appliances which are in parallel operation.

iii. Since our main focus is identifying household appliances, the proposed algorithm consists of the best combination of methods in order to achieve a higher accuracy level.

Demand identification creates several opportunities in energy sector because it enables detailed information in energy consumption, some of them are as stated below. Demand identification and load monitoring is useful to perform surveys of residential and commercial energy consumption. So, energy saving potential of particular household or organization can be identified. And also in smart grids there is demand response system. In that system there should be proper load monitoring and demand identification technique. Further, smart metering can be developed with this technology which can provide more details about the purpose of end-user energy utilization. And also it could be used to introduce itemized tariff system.

When previous work in this field is considered, there are two categories of appliance load monitoring based on the number of sensors used to gather appliance-level information, namely multi-sensors system and single-sensor system.

In multi-sensor system sensors should be installed in-line with every device. This system is used for intrusive load monitoring. This method has some drawbacks such as it fails to give an overall picture of consumption, high installation complexity and cost. So this method is not use that much.

In single-sensor system measurements are taken at a single point. This system mainly related with non-intrusive load monitoring (NILM). NILM approach has been used for real time recognition through a single sensor using parameters of real power, power factor, rms current, rms voltage, peak current, peak voltage etc [1]. Other method is to identify switching event (switch on or off) and appliance identification using change in appliance signature of apparent, reactive, and distortion power [2]. And also using transient features, load identification can be achieved [3]. Active power and reactive power components also have used for NILM studies [4]. In this paper first twelve current harmonics and power components are used. So, there are number of parameters and it is helpful to improve the accuracy.

The rest of the paper is organized as follows. Section II describes the methods by which measurements were obtained. In section III, the methods have been developed to identify individual appliances. Section IV describes the algorithms and the way they were combined to identify multiple appliances which are in parallel operation. Section V describes the results of the analysis as the conclusion.

II. Obtaining Measurements

Measurements were taken individually and different combinations of appliances by using a power analyzer. The measured parameters are the first twelve components of current harmonics, total current, active power, reactive power and apparent power.

The data acquired were initially stored internal memory of power analyzer and then transferred in to a computer database for further studies. First, parameters of individual appliances were measured to reconstruct their signatures individually. Then using these data, a database with appliance signatures is developed with MySQL®. After that, several combinations of appliances were measured to obtain their combined signals.

III. Identifying Individual Appliances

Second step of this research was developing algorithm to identify unknown individual appliances. Two algorithms were used for that purpose.

A. Artificial Neural Network (ANN) to identify individual appliances

ANN analysis is a pattern identification method which can be used in complex nonlinear systems. It involved following steps. Most important step of this analysis is training the system and the developed database is used for that. First twelve current harmonics, total current, active power, reactive power and apparent power is used to train the system. Then clustering has to be done manually before implementing the training algorithm. Neural Network Tool Box available in MATLAB® software is used for this. Manual clustering should be done according the required format of the Neural Network Tool Box. This shows in TABLE 1.

Then a target file of an unknown appliance is compared with the trained system and identified the best fit appliance with the target file.

Accuracy of this method depends on the number of available data in the initial database. Accuracy increases with the number of available records for single appliance.

B. Support Vector Machines (SVM) to identify individual appliances

SVM can classify exactly two data sets by recognizing patterns. It classify data by finding the best hyper plane that separates all the data points of one class form those from the other class. First the support vector machine was trained and then used the trained machine to classify new data.

TABLE 1
Sample Database and method of clustering

I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I	P	Q	S		CFL	FAN	LAP
Measured Data of CFL																	1	0	0
																	1	0	0
Measured Data of FAN																	0	1	0
																	0	1	0
Measured Data of LAP																	0	0	1
																	0	0	1

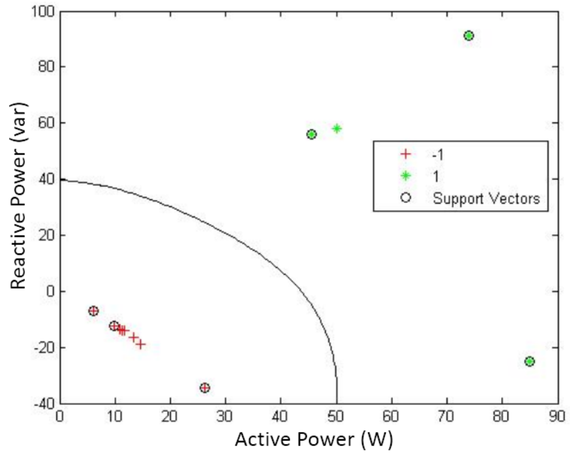

Fig. 1. SVM analysis of CFL and TV

Fig. 1 shows SVM analysis of a CFL and TV. It was plotted only using active power and reactive power because only two dimensional data can be plotted in a 2D plane. But in real analysis first twelve current harmonics, total current, active power, reactive power and apparent power were used.

Here only two data sets can be compared. That is the main problem of this method. Therefore this method is used with the ANN as an accuracy improving method. Top two outputs of the ANN is analyzed in SVM and the output is compared with the top result of the ANN.

IV. IDENTIFYING MULTIPLE APPLIANCES

This is the most important stage of this research. Here several equipment were identified which were turn on at the same time. To that two different analysis methods were used.

1) *State Based Analysis*
2) *Sequential Analysis*

In state based analysis four independent algorithms were developed to identify multiple appliances

1) *State Based Analysis*

In state based analysis, data in the signature library are rearranged at the initialization step. This will end up with the signatures of combined appliances and it fed to the algorithm. Then target measurement which want to analyze is taken from the data logger and it also fed to the algorithm. Identification techniques analyze states of appliances and the final output will be given as a matrix which contains the on/off state of each appliance. This process is illustrated in Fig. 2.

A. *Initialization of data for analysis*

After getting the data from MySQL® database to the MATLAB® workspace as a standardized matrix, it is required to rearrange those to be suitable to enter to the algorithms developed. This is called initialization and is includes following steps.

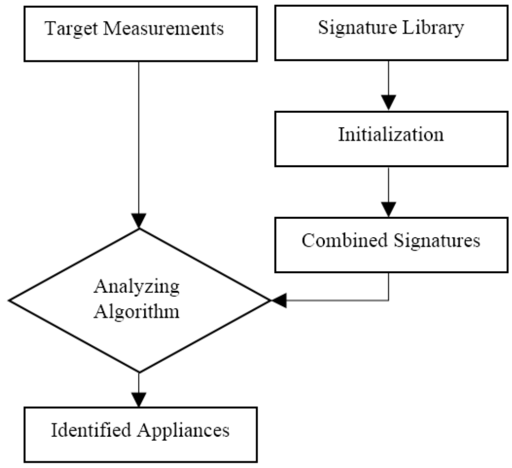

Fig. 2. State Based Analysis Process

a. Dividing and assigning data of individual appliances to separate matrices considering n number of harmonics from each appliance.

b. Getting i number of data from each harmonic of each appliance and assigning those into matrices in appliances vise. Hence each appliance has a $(i \times n)$ matrix including these data.

   $\mathrm{I_A} = [\mathrm{I_1}\ \mathrm{I_2}\ ...\ \mathrm{I_j}\ ...\ \mathrm{I_i}]'$) is the individual appliance matrix of appliance A, where $\mathrm{I_j}$ is the n dimensional vector of having first n harmonic components of appliance A.

c. Deriving possible combinations by performing following matrix multiplication in (1). This is showed for a number of appliances system.

$$\begin{pmatrix} 0 & 0 & 0 & \dots \\ 0 & 0 & 1 & \dots \\ 0 & 1 & 0 & \dots \\ 0 & 1 & 1 & \dots \\ 1 & 0 & 0 & \dots \\ 1 & 0 & 1 & \dots \\ 1 & 1 & 0 & \dots \\ 1 & 1 & 1 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}_{2^a \times n} \times \begin{pmatrix} [I_A]_{i\times n} \\ [I_B]_{i\times n} \\ [I_C]_{i\times n} \\ \vdots \end{pmatrix}_{a\times 1} = \begin{pmatrix} [0]_{i\times n} \\ [I_C]_{i\times n} \\ [I_B]_{i\times n} \\ [I_B + I_C]_{i\times n} \\ [I_A]_{i\times n} \\ [I_A + I_C]_{i\times n} \\ [I_A + I_B]_{i\times n} \\ [I_A + I_B + I_C]_{i\times n} \\ \vdots \end{pmatrix}_{2^a \times 1} \quad (1)$$

These three matrices are termed Combinations matrix $[C]$, Harmonic components of individual appliances $[H]$, and matrix of possible combinations $[P]$ respectively. In matrix $[H]$ a cluster of i vectorsare included for each appliance. And hence 2^a clusters are in matrix $[P]$. With this step, the initialization process will be over and can be moved to the appliance identification techniques.

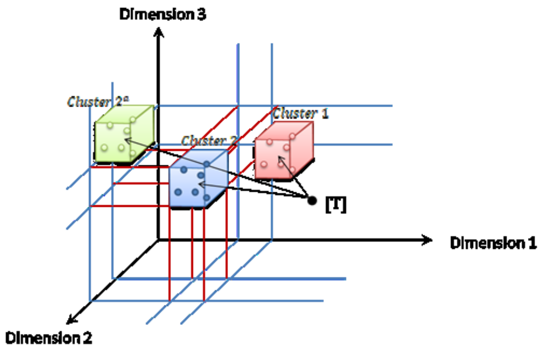

Fig. 3. Method STD

B. *Standard mathematical technique [Method STD]*

This method is the simplest method that is used for appliance identification, and this was used as the bottom line to maintain the accuracies of other methods more than this. Vector $[T]$ which should be analyzed will be created by getting first n harmonic components of the combination of appliances which have to be identified separately. Each cluster got in matrix $[P]$ and the vector $[T]$, was imaginary plotted in a n dimensional vector space. Fig. 3 shows the simplified plot of n dimensional vector space in a 3 dimensional vector space. Then the geometrical distance from the vector $[T]$ to each cluster was calculated by using (2).

$$[D_{STD}] = \sqrt{\sum_1^n (A-T)^2} \tag{2}$$

This can be done by two ways either as calculation of distance to the midpoints of each cluster or to the margin of the cluster. Hence $[A]$ represents the midpoint or the margin of each cluster. But in both cases it seems almost same accuracy. Then it can be say that, the vector $[T]$ is belongs to the cluster which located at a least distance from $[T]$. Then the unknown combination of vector $[T]$ shall be same as the combination of that cluster.

C. *Agglomerative clusters from data [Method ACD]*

In previous *Method STD*, which cluster that it belongs was known when a vector of matrix $[P]$is plotted on the vector space. But in this method all the vectors of matrix $[P]$ are plotted on a vector space without considering the relevant cluster of each. Then in MATLAB® the function *clusterdata* [5] can clusters the given data into given numbers of clusters. Fig. 4 shows a clustered plot that done by MATLAB® according to the prevalence of the data simplified to a three dimensional vector space. This prevents errors due to small variations in data and gives a slight higher accuracy than *Method STD*.

Calculating distances from vector $[T]$ to the each cluster was done as same as the *Method STD* by using (3). Then as same as that method, the vector $[T]$ was considered as, belongs to the cluster and has same combination pattern which has least distance from it.

$$[D_{ACD}] = \sqrt{\sum_1^n (A-T)^2} \tag{3}$$

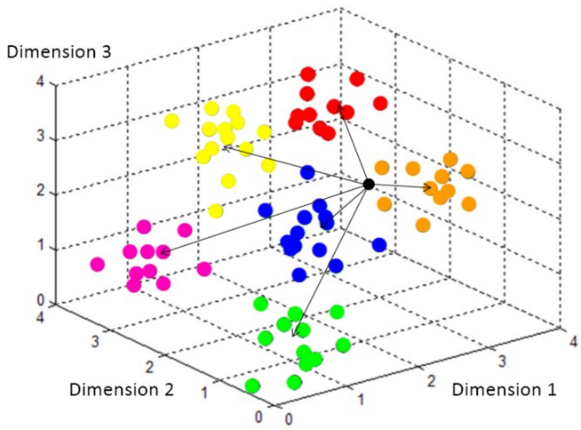

Fig. 4. Method ACD

D. *Procrustes Analysis [Method PRO]*

This method uses the function *procrustes* [6] which use Procrustes Analysis in MATLAB® and compares the distribution of the harmonic components of vector $[T]$ with the average of each and every cluster already developed in matrix $[P]$. This analysis matches landmark data to calculate the best shape-preserving Euclidian transformations. For that averages of each harmonic of every cluster were plotted in a two dimension vector space against the harmonic number as showed in Fig. 5. Magnitudes of each harmonic component of vector $[T]$ were also plotted in same vector space against harmonic number. Then a transformation was developed using by only shifting and considering to minimize the distance between the target shape $[T]$ and the transformed shape clusters at $[P]$ as measured by the least squares criterion. [7]

The *dissimilarity measure* d gives a number between 0 and 1 describing the difference between the target shape and the transformed shape. By inspecting this d, the cluster which has least difference with the target was selected as the cluster which target belongs to.

E. *Nearest Neighbors Method [Method KNN]*

In this method also, the plotting mechanism of known vectors in n dimensional vector space is same as the *Method ACD*. After plotting, there is a function in MATLAB® called *knnsearch* [8] to find the k number of vectors near to the target vector $[T]$. In Fig. 6 it shows the known vectors plotted in an n dimensional vector space and two different target vectors plotted in the same vector space. Their nearest neighbor vectors are marked by the each circles. The output of *knnsearch* is the vector numbers of the given number of nearest vectors.

Then the algorithm can identify the cluster that the majority of nearest vectors belongs to and it will be assigned as the cluster target vector belongs to. [9]

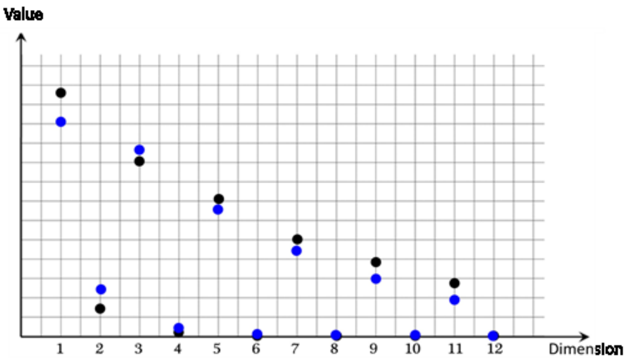

Fig. 5. Method PRO

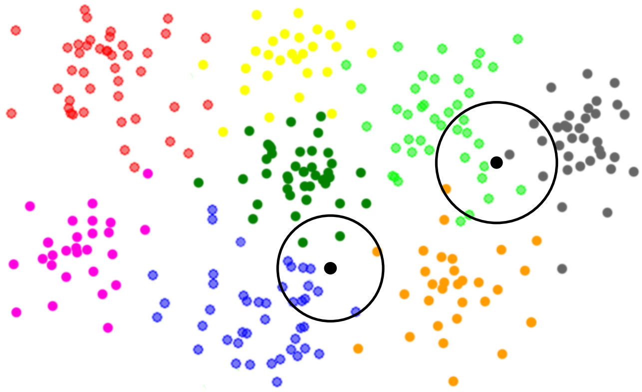

Fig. 6. Method KNN

F. Basic accuracies and accuracy optimization

Accuracies of each method when they identify five types of appliances are shown in TABLE 2. The last column shows the average accuracies of each method and the last raw shows the average accuracies of identifying each appliance. This is a result of an experiment done using five appliances for a period of 72 minutes.

Since Iron and Heater are resistive loads with only fundamental current harmonic component, those could be identified with 100% accuracy in all four methods. Fan and Laptop have the maximum accuracies for *Method ACD* while CFL in *Method PRO*. Therefore by combing these four methods together, highest accuracy can be achieved and that accuracy optimization step was done as follows.

Each method gives a number relevant to the particular combination. Using following combination matrix C shown in (4), we can identify the availability of the each appliance at that particular instance.

$$C = \begin{bmatrix} 0 & 0 & 0 & \dots \\ 0 & 0 & 1 & \dots \\ 0 & 1 & 0 & \dots \\ 0 & 1 & 1 & \dots \\ 1 & 0 & 0 & \dots \\ 1 & 0 & 1 & \dots \\ 1 & 1 & 0 & \dots \\ 1 & 1 & 1 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{4}$$

TABLE 2
Accuracy Comparison

	Iron	Fan	Laptop	CFL	Heater	Average
STD	100.0%	91.67%	83.33%	65.28%	100.0%	88.05%
ACD	100.0%	93.06%	84.72%	66.67%	100.0%	88.89%
PRO	100.0%	88.89%	77.78%	73.61%	100.0%	88.05%
KNN	100.0%	91.67%	79.17%	66.67%	100.0%	87.50%
Average	100.0%	91.32%	81.25%	68.05%	100.0%	

For every method an availability matrix for each appliance was obtained for every instance. And from the previous studies the method which suitable for each appliance type was known. Then the availability of each appliance was extracted from the matrix got from most suitable method for that particular appliance. Then the final matrix was developed by combining those matrices. For the previous experiment 90.27% of accuracy was obtained when the accuracy is optimized by using this technique.

2) Sequential Analysis

In this method taking measurements were started from the ALL OFF state. Here an assumption was made that only a single appliance changes its operating state (ON/OFF) per one minute due to logging device which can only keep a single recode per minute with harmonic data. This limitation is the reason for the above assumption.

From the difference of two ancillary measurements the data of the appliance which has changed its operating state can be obtained. This difference then fed into the ANN which has been trained with the individual appliances data. Fig. 7 shows this process.

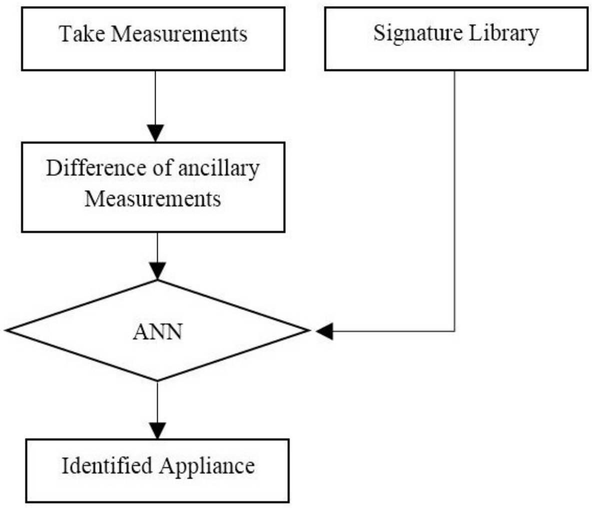

Fig. 7. Sequential Analysis Process

Hear two training methods were used. Observations of the results showed that each training method identifies specific type of appliances with more accuracy than the other method. For an example *fitnet* [10] method identifies CFL with more accuracy and *patternnet* [11] method identifies Laptop with more accuracy. Inputs to the ANN are analyzed using both of this training method and the result with the greater probability is chosen as the final result.

Since this is a sequential analysis, results of the each state depends on the results of the previous states. Therefore maintaining high accuracy is a crucial factor. For an example when a CFL is turned on and if the system identifies it as laptop turning-on event, then when laptop is turn on actually system will record it as laptop turn off. So ON/OFF recode state always depend on the previous recode.

V. CONCLUSION

Six basic algorithms have been developed in this research paper for the appliance-level demand identification problem. It seems that resistive appliance like Iron and Heater could be identified with almost 100% accurately while other appliance have less accuracy around 65% to 90%. Let's consider the appliances except IRON and HEATER. The accuracy of a selected appliance is varying with the method used for the identification. The *Method PRO* is based on a identifying the pattern of harmonic content variation and hence it can identify appliance which have high harmonic content like CFL with higher accuracy. *Method KNN* has higher accuracy when identifying FANs while *Method ACD* has that for FAN and LAPTOP. Therefore using a combination of algorithms together for the identification process can increase the accuracy compared to a single method.

State Based Analysis and Sequential Analysis were planned to develop separately and combine together after achieving the high accuracy in both methods separately.

Also by differentiating the identification algorithms by using various methods for the process can increase the reliability of the entire system. Therefore, all of the methods when combined to declare the final output, more than 90% of accuracy for up to six appliances can be achieved.

REFERENCES

[1] A.G. Ruzzelli, C. Nicolas, A. Schoofs, and G.M.P. O'Hare, "Real-Time Recognition and Profiling of Appliances through a Single Electricity Sensor," in 2010 7th Annual IEEE Communications Society Conference on Sensor Mesh and Ad Hoc Communications and Networks (SECON), 2010, pp-1 - 9

[2] Markus Weiss, Adrian Helfenstein, Friedemann Mattern, and Thorsten Staake, "Leveraging smart meter data to recognize home appliances," IEEE International Conference on Pervasive Computing and Communications (PerCom 2012), Lugano, Switzerland, 2012, pp- 190-197.

[3] Hsueh-Hsien Chang. (2012, November). Non-Intrusive Demand Monitoring and Load Identification for Energy Management Systems Based on Transient Feature Analyses. *Energies (19961073)* [Online]. 5(11), p4569. Available: http://www.mdpi.com/1996-1073/5/11/4569/pdf

[4] Ahmed Zoha , Alexander Gluhak , Muhammad Ali Imran and Sutharshan Rajasegarar. (2012, December). Non-Intrusive Load Monitoring Approaches for Disaggregated Energy Sensing: A Survey. *Sensors* [Online]. 12(Special Issue on Ubiquitous Sensing), 16838-16866. Available: http://www.mdpi.com/1424-8220/12/12/16838/htm

[5] clusterdata., MathWorks, [online] 2014, http://in.mathworks.com/help/stats/clusterdata.html (Accessed: 11thNovember 2014)

[6] procrustes., MathWorks, [online] 2014, http://in.mathworks.com/help/stats/procrustes.html (Accessed: 14thNovember 2014)

[7] Procrustes Analysis., MathWorks, [online] 2014, http://in.mathworks.com/help/stats/procrustes-analysis.html#br3txrn-4 (Accessed: 14thNovember 2014)

[8] knnsearch., MathWorks, [online] 2014, http://in.mathworks.com/help/stats/knnsearch.html (Accessed: 15thNovember 2014)

[9] Classification Using Nearest Neighbors., MathWorks, [online] 2014, http://in.mathworks.com/help/stats/classification-using-nearest-neighbors.html#bsehylk (Accessed: 22ndNovember 2014)

[10] fitnet, MathWorks, [online] 2014, http://in.mathworks.com/help/nnet/ref/fitnet.html (Accessed: 3rdDecember 2014)

[11] patternnet, MathWorks, [online] 2014, http://in.mathworks.com/help/nnet/ref/patternnet.html (Accessed: 3rdDecember 2014)