Appliance Classification using Energy Disaggregation in Smart Homes

Sourodeep Bhattacharjee, Anirudh Kumar and Joydeb RoyChowdhury
Embedded Systems Lab
CSIR - Central Mechanical Engineering Research Institute
M.G. Avenue, Durgapur, West Bengal, India - 713209
Email: sourodeepbhattacharjee@gmail.com, anirudhkumar@cmeri.res.in and jrc@cmeri.res.in

Abstract—**In this work we have addressed the problem of ap- pliance classification and power consumption anomaly detection using energy disaggregation and machine learning techniques. The active power consumption data, received from a smart-meter, has been used as the only parameter for solving our problem. We have implemented a decision tree algorithm to classify appliances based on thresholds of their power consumption. Additionally, we have also proposed and implemented an algorithm for unusual fluctuation detection based on average magnitude of such fluctuations and an appliance quality recommender based on power-factor of the appliance. Initial results are promising as the classifier works correctly for 74% of instances, while the anomaly detector works correctly for 80% anomalies.**

Keywords—Energy Disaggregation, Non-Intrusive, Energy In- formatics, Machine Learning, Classification.

I. Introduction

This paper attempts to address the problem of energy disaggregation, using supervised machine learning techniques, based on real-time power consumption data of appliances. Energy disaggregation involves identifying power consumption of individual appliances from total power consumption of a collection of appliances.

The power consumption data is received from a smart digital energy meter [6] developed in our lab. Apart from the energy meter, our setup consists of ZigBee communication nodes, real-time energy consumption analyzer (developed in C# programming language), and off-the-shelf electrical appliances. Our goal is to create a framework for identifying currently running appliances, detecting unacceptable anomalies in the consumption data and finding out the power efficiency of appliances from their power factor.

With our experiments, we have demonstrated that decision tree algorithm can be effectively used for appliance classification. Moreover, we have also provided a method for detecting unacceptable anomalies in power consumption, without doing computationally expensive real-time analysis.

II. Background

Energy disaggregation is the process of identifying individual loads (electrical appliances or electronic devices) from the total power consumption data of a collection of appliances or devices [11]. Several policies have been applied to achieve energy disaggregation or non-intrusive load identification as summarized in [3], such as classifying devices based on electrical events by analyzing power consumption patterns as finite states.

TABLE I: Energy disaggregation and data frequency [10]

Data Frequency	Features Used for Disaggregation	Disaggregation Capabilities
1 hr - 15 min	Visually distinguishable patterns, time and duration for which the appliance was used	Continuous loads or time-dependent loads or loads that are related to outdoor temperature etc.
1 min - 1 s (1 Hz)	Steady state steps or power transitions	Appliances like Refrigerator, ACs, Heaters etc. less than 10 in number
1 - 60 Hz	Steady state steps or power transitions	10 - 20 types of appliances

Recent research trends in this area, indicate extensive use of Hidden Markov models to identify complex load patten signatures of devices. In our work, however, we have used decision trees for non-intrusive load classification. The reason is that our work is at exploratory stage and we are working with active power only, derived using smart energy-meter, developed in our lab, which currently transmits data at interval of 4 seconds. Please refer to table I [10] to realize the scope of our work.

Power factor of an appliance is ratio of real power that the appliance is consuming to apparent power in the electrical circuit. Power factor varies from 0 to 1. Devices with high power factor are considered to be of good quality, because low power factor devices unnecessarily draw more current than those with high power factor. This high current increases dissipation of electrical energy in the distribution system. In our work we have made an effort to filter out bad quality appliances by recommending the user of smart home to remove such appliances with low power factor.

The rest of the paper is organized as follows: section III deals with previous related work; section IV describes the hardware setup and the software we have developed; in section V, VI and VII we have expanded on the problem and the algorithms/methodologies employed to solve the same; section VIII contains the experiments, results and discussion; and we have concluded this paper in section IX.

III. Related Work

A previous work on applying supervised machine learning techniques for energy disaggregation uses Factorial Hidden

978-1-4799-3826-1/14/$31.00©2014 IEEE

Markov Model (FHMM) to classify devices [1]. The authors in this paper claim that using FHMM yields much better results than using Simple Means, where FHMM has an average accuracy of 47.7% while that of Simple Mean is 25.9% during testing the models. While the authors in this paper have used smart plugs and routers developed by Enmetric (http://www.enmetric.com) for power monitoring consumption data, we have performed our experiments without smart plugs. Another difference is that while the authors have used appliances with complex power consumption patterns, we have used not used such loads to conduct the experiments reported in our work.

Another paper [2] suggests a more flexible approach using Iterative Hidden Markov Models to create models of appliances as a result of which no other prior knowledge (sub-metering of individual appliances) is needed. The authors have tested their work using the REDD public dataset [1] and have implemented a power savings recommender system. Due to the uniqueness of their approach the authors have not compared the performance analysis of their algorithm with other related work. However the authors claim that their method is non-intrusive in nature as it does not include sub-metering of appliances.

Even though HMMs have expressive power, the authors in [3] claim that, additive factorial HMMs are difficult to analyze especially where the gross model consists of several HMMs. To address this issue, the authors have suggested Additive Factorial Approximate MAP algorithm; which falls under unsupervised learning, and attempts to reduce the number of posteriors allowed. The authors claim that their method works successfully due to the additive nature of power consumption. Other related work include [4] and [5].

IV. Energy Monitoring and Analysis Tools

This section elaborates on the hardware setup and the software used for appliance classification and anomaly detection in smart home environment.

The block diagram in Fig. 1 shows input appliances connected to an electrical outlet, which in turn is connected to the meter. The meter sends date-time stamped total power consumption data to a receiver module. The receiver module is connected to a computer, running the energy analysis software.

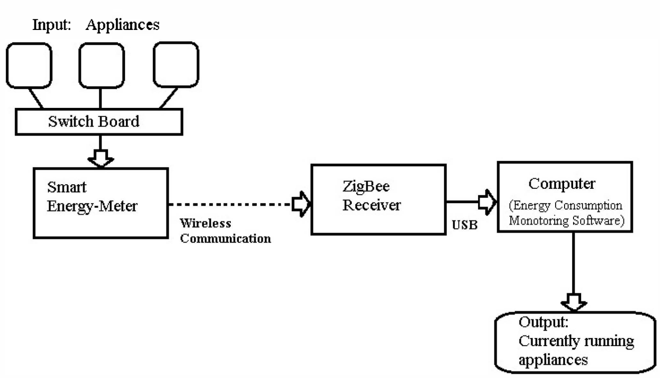

Fig. 1: Block Diagram of Hardware Architecture

A. Smart Energy-Meter and ZigBee Trans-Receivers

The smart energy-meter forms an important entity of the envisioned smart home. The salient components of the meter, pertinent to the scope of the this paper, includes a micro-controller based embedded system, a ZigBee communication module (XBee Pro S1 provided by Digi. www.digi.com), and a real-time clock (RTC). The active power consumption data is date-time stamped by the micro-controller using the RTC. This data is then transmitted by the meter, at 4 second intervals, to receiver modules.

The purpose of ZigBee trans-receiver module is to receive the data sent by the energy meter and preprocess it before forwarding it to a computer application using USB. This module can also be used to send specific queries to the energy-meter and interact with smart home actuators such as smart plugs. The module consists of an Arduino Mega 2560 micro-controller board (http://arduino.cc/), a XBee Pro S1 module and a bridge to connect the XBee to the Arduino. In our setup, the preprocessing step involved converting the raw data (delimited by spaces and new-line) into comma delimited values to facilitate data manipulation using standard and custom softwares.

B. Energy Consumption Monitoring Software

The desktop application software is primarily responsible for saving the data received to the computer's hard disk drive. It has real-time data analysis capabilities like visualization of load signatures, as illustrated in Fig. 3. The softwares supports two-way communication with the smart meter using the ZigBee module; this feature can be used to selectively turn off appliances. Additionally, the software is threaded to ensure reliable real-time performance. Developed in C# using Visual Studio 2013 as a Windows Forms desktop application, it is backward compatible to .NET Framework 3.5.

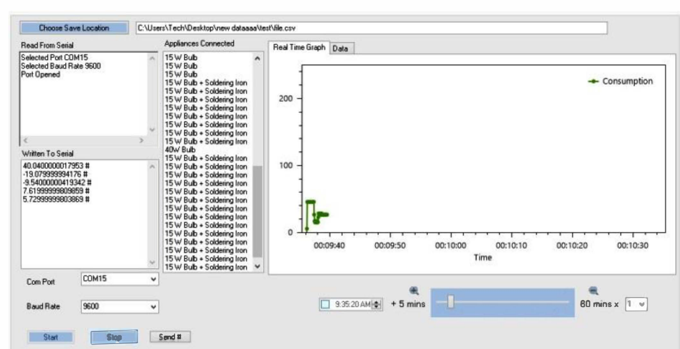

Fig. 2: Energy Consumption Monitoring Software

V. Appliance Classification

The primary goal of this paper is to classify appliances that are currently running using energy disaggregation and machine learning. Our focus is on resistive bulbs and soldering iron, and the challenges posed when classifying them. To better understand the problem please refer to the appliances and load signatures in Fig. 4 and Fig. 5. The supplied voltage from power station varies from 230 V to 250 V and hence the consumption is not always constant. Furthermore the power

TABLE II: Variation in Power Consumption (at 220 - 250 V and at 50 Hz)

Load	Power Rating (W)	Mean Power (W)	Standard Deviation
L1	15	14.88	1.08
L2	40	41.46	2.89
L3	11	10.25	0.96

TABLE III: Variation in Power Consumption of L3 (at 220 - 250 V and 45 - 55 Hz)

Supplied Voltage	Mean Power	Standard Deviation
220	8.98	0.10
225	9.36	0.09
230	9.8	0.10
235	10.25	0.12
240	10.7	0.13
245	10.86	0.2
250	11.7	0.17

consumption rating of the appliances are not accurate and is variable as demonstrated in table 2 where we have varied the voltage supply using a programmable AC source made by Chroma (www.chromausa.com) and measured it using a power analyzer made by YokoGawa (www.yokogawa.com). We have also varied the voltage supply frequency from 45 Hz to 55 Hz, for the soldering iron (L3), for supply voltage varying from 220 - 250 Hz in table III.

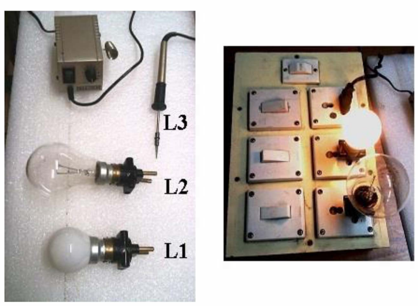

Fig. 3: Appliances

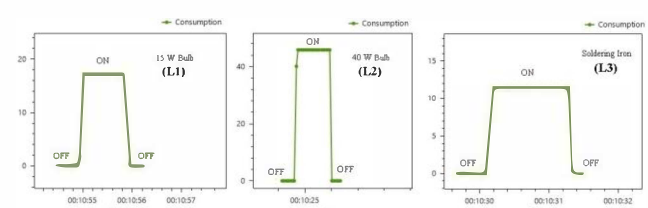

Fig. 4: Power Consumption Patterns

With these factors in mind our objective is to classify such loads in a reliable manner. In such situations using Hidden Markov Models may become a computationally expensive solution for both training and real-time testing, more so when the signatures are apparently similar to each other due to the low data frequency. Hence, we have restricted our study to basic machine learning algorithm such as C4.5 [8] Decision Trees.

TABLE IV: True Positive Rate and False Positive Rate

True Positive Rate	False Positive Rate	Class
1.000	0.000	40 W Bulb
1.000	0.000	15 W Bulb
1.000	0.000	Soldering Iron
1.000	0.013	40 W + 15 W Bulb
1.000	0.000	15 W Bulb + Soldering Iron
0.909	0.000	40 W Bulb + Soldering Iron
1.000	0.000	15 W Bulb + 40 W Bulb + Soldering Iron

A. Training Models using Prior-Knowledge

In order to classify appliances we have gathered power consumption data (consisting of power consumption of appliances and respective labeling) for individual appliances and in combination over a period of time in a particular day. We have included the normal variations in power consumption for a particular appliance in the database so that even when the supplied voltage fluctuates we should be able to classify the loads correctly. We have used WEKA [7] Machine Learning workbench for training models.

B. C 4.5 Decision Tree Algorithm

Invented by Quinlan, the C4.5 decision tree algorithm is a widely used algorithm for its simple implementation, transparency and robustness, even though it is prone to over-fitting the training data. The algorithm, well described in [9], involves creating a tree by selecting an attribute as root node and generating branches based on the number of possible values of that attribute. This process is repeated until we receive same classification for all instances at a node. The attributes are split based on information gain as a result of the splitting. WEKA provides a Java implementation of the C4.5 algorithm which they have named J48. In this paper we have used the model generated by the same J48 algorithm.

C. Model Generation

For generating the model, we have created the database with 15 minutes pattern for each isolated appliance and in combination at different times of the day to take into account the variation in supply voltage and minor fluctuations. This data was then fed to WEKA and the J48 algorithm was run and tested using 10-fold cross validation in which the data was split into 10 partitions and model was generated using 9 of them and tested with the remaining one. This process was repeated 10 times. We have received an accuracy of 99% with 10-fold cross validation for the J48 algorithm. The true positive rate and the false positive rate of the classifier is given in table IV.

The resulting code works on threshold values of power consumption. The model is derived from extensive dataset making the code robust in its detection mechanism. The decision tree derived from the model is demonstrated in pseudocode 1.

Pseudo code 1 Appliance Classification

```
GET_APPLIANCE_STATE ( load )
{
if(load<1)
{
PRINT ( "No Appliances Connected")
}
if(load <= 57.23 AND load > 1)
{
 if(load <= 28.61)
{
 if(load <= 17.17)
{
if(load <= 11.45)
PRINT ("Soldering Iron")
else if(load > 11.45)
PRINT ("15 W Bulb")
}
else if(load > 17.17)
PRINT ("15 W Bulb + Soldering Iron") }
else if (load> 28.61)
{
if(load <= 45.78)
PRINT ("40W Bulb");
else if(load > 45.78)
PRINT ("40W Bulb + Soldering Iron ");
}
 else if(load > 57.23)
{
if (load<= 62.94)
PRINT ("40 + 15 W Bulbs");
else if (load> 62.94)
PRINT ("Soldering Iron + 40 W Bulb
+ 15 W Bulb");
}
```

VI. Power Consumption Anomaly Detection

Some power consumption anomalies manifesting as fluctuations, as shown in Fig. 5, are highly undesirable. These fluctuations can cause to harm to sensitive devices like computers. Hence, these incidents must be detected as soon as possible, to identify and rectify the cause, viz. a loose connection between the electrical outlet and the plug of the device.

To address the abnormal fluctuation problem we have incorporated a function in our software, which detects the fluctuations close to real-time. The basis of the detection comprises of empirically identifying a threshold level of those fluctuations which can be classified as abnormal. The next step is to identify pattens such as sharp drops followed by rise in the consumption (not necessarily in sequence), beyond the acceptable threshold and classify the incident as abnormal. The pseudocode 2 is able to detect persistent fluctuations which last for 3 data instances.

VII. Device Quality Recommendation

We have also included power factor of various devices in our algorithms so as to gain a better insight into wastage of

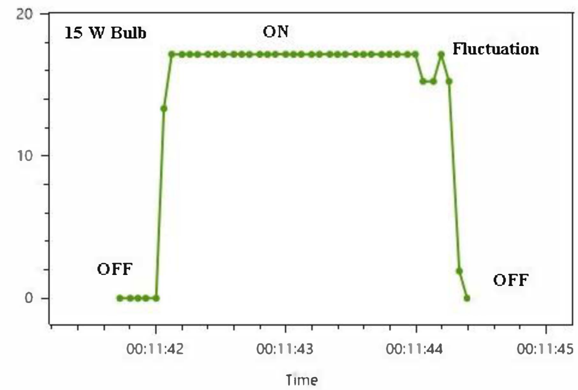

Fig. 5: Fluctuation in Power Consumption

Pseudocode 2 Abnormal Fluctuation Detection

```
DETECT_FLUCTUATION (PREVIOUS_CONSUMPTION,
          PRESENT_CONSUMPTION,
    NEXT_CONSUMPTION )
{

PREVIOUS_CHANGE =
  PREVIOUS_CONSUMPTION – PRESENT_CONSUMPTION

    NEXT_CHANGE
      = PRESENT_CONSUMPTION – NEXT_CONSUMPTION

    while (NEXT_CHANGE is negligible
              AND Number_of_iterations <= 3)
   {
        PRESENT_CONSUMPTION = NEXT_CONSUMPTION
        UPDATE (NEXT_CONSUMPTION)
        UPDATE (NEXT_CHANGE)
   }
  if((PREVIOUS_CHANGE < 0 AND NEXT_CHANGE > 0)
OR
  (PREVIOUS_CHANGE > 0 AND NEXT_CHANGE < 0))
  {

   if( absolute_value_of(PREVIOUS_CHANGE) >
   Threshold_LOW   AND
  absolute_value_of(PREVIOUS_CHANGE)
   <  Threshold_HIGH )
  {
   PRINT "ABNORMAL FLUCTUATION DETECTED"
  }}}
```

energy incurred, when using devices with low power factor. For example refer to Fig. 6, which shows a high quality branded Compact Fluorescent lamp (CFL) is consuming more power while a poor quality lamp is consuming less power. With present setup, the poor quality lamp should be used to save power. However, the poor quality lamp has very bad power factor (approximately 0.45) rating while the high quality lamp has a power factor close to .93. We have included a feature in the smart-meter that sends power factor data to the receiver, which is then shown to the user as a recommendation on the quality of the device, where above 85% power factor is shown as a high quality appliance and below that is considered as poor quality appliance.

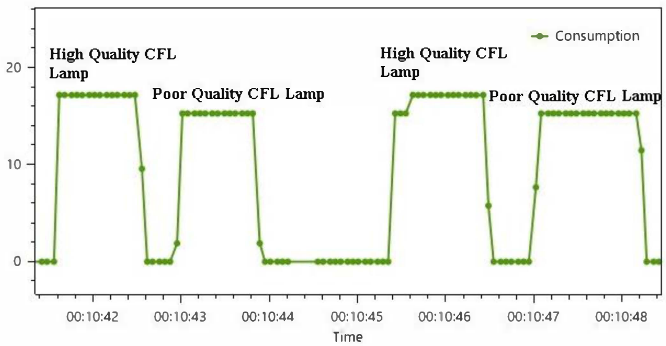

Fig. 6: Quality Variation of Same Appliance Type

VIII. Experiments and Results

For experimenting we have measured the percentage of correctly classified appliance in real time. The C4.5 model was implemented in C# as a function in the energy monitoring software. We have turned the appliances on and off several times and in all possible eight combinations. To ensure robustness of the methodology and repeatability, we have also tested the model using different bulbs of same power rating and attached the setup in different electrical outlets in our lab.

We have found that the number of correctly classified instances returned by the model is 74% for approximately 500 test cases. The apparently low performance level, despite the simple setup, is attributed to several factors. Firstly, we have used a very simple algorithm to classify appliances in on-line mode. More sophisticated algorithms where the time series pattern can be incorporated, are expected to perform better at the cost of time complexity and maintainability. Secondly, it takes 4 seconds for the meter to send the latest reading, hence if appliances are switch very quickly then we get errors due to the low resolution of the data. This is a hardware limitation and the problem can be addressed by having a separate micro-controller to sample and send the data.

And lastly loose connections between plugs and electrical outlets, cause sparks which give aberrations in the data, although these are detected by the Abnormal Fluctuation Detection Module 80% times for 100 test cases. The same can be used in tandem with the Appliance Classifier to establish a confidence level of the detection. The device quality recommendation is 100% correct for the CFLs, where we have experimented separately (not as a part of appliance classifier) using two sets of high and low power factor CFLs, each containing two lamps.

IX. Conclusion and Future Work

In this paper we have addressed the problem of appliance classification using energy disaggregation in real-time using C4.5 algorithm. For experiments we have used three loads consisting of a 15 W bulb, a 40 W bulb and a soldering iron. The collective consumption data is transmitted by the smart energy-meter to zigbee receiver modules connected a PC via USB. A software on the PC is responsible for the archiving the data and analyzing it in real-time. The same software contains a C4.5 decision tree model which displays the currently running appliances and detects abnormal data, as well as, provides recommendation on the quality of the appliance.

Our contribution in this work is a user-friendly intelligent software (for energy consumption monitoring), along with the basic methodology for classifying loads, detecting abnormal fluctuations in the power data and a framework that encourages the use of power efficient appliances. Power wastage minimization is of prime concern for both producer and consumers, and hence algorithms such as anomalous power consumption detection and power factor based appliance quality recommendation should be an integral component of smart home controlling policy. Experimental results are promising when compared to other benchmark work, although it is not justified to do a stringent comparison due to the vast difference in scale of architecture and experimental setup. Our claim is that it should be sufficient to use power consumption threshold based, tree like algorithms for the kind of loads we have used.

Possible future work will involve using complex loads such as microwave ovens, smart-plugs, high resolution data and sophisticated machine learning algorithms. We also plan to create an online repository for high resolution real data, which can be accessed and used for further research by others.

References

[1] Kolter, J. Zico, and Matthew J. Johnson. “Redd: A public data set for energy disaggregation research.” In proceedings of the SustKDD workshop on Data Mining Applications in Sustainability, pp. 1-6. 2011.

[2] Parson, Oliver, Siddhartha Ghosh, Mark Weal, and Alex Rogers.“Non-Intrusive Load Monitoring Using Prior Models of General Appliance Types.” In AAAI. 2012.

[3] Kolter, J. Zico, and Tommi Jaakkola. “Approximate inference in additive factorial HMMs with application to energy disaggregation.” In International Conference on Artificial Intelligence and Statistics, pp. 1472-1482. 2012.

[4] Kolter, J. Zico, Siddharth Batra, and Andrew Ng. ”Energy disaggregation via discriminative sparse coding.” In Advances in Neural Information Processing Systems, pp. 1153-1161. 2010.

[5] Shao, Huijuan, Manish Marwah, and Naren Ramakrishnan. ”A temporal motif mining approach to unsupervised energy disaggregation.” In Proceedings of the 1st International Workshop on Non-Intrusive Load Monitoring, Pittsburgh, PA, USA, vol. 7. 2012.

[6] RoyChowdhury, Joydeb. “A prepaid smart card operated electronic energy meter with online load optimizer for solar power application.” WIPO Patent 2008059528, issued May 23, 2008.

[7] Hall, Mark, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. “The WEKA data mining software: an update.” ACM SIGKDD Explorations Newsletter 11, no. 1 (2009): 10-18.

[8] Quinlan, J. Ross. “C4. 5: Programs for machine learning (morgan kaufmann series in machine learning).” Morgan Kaufmann (1993).

[9] Witten, Ian H., and Eibe Frank. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2005.

[10] Carrie Armel, K., Abhay Gupta, Gireesh Shrimali, and Adrian Albert. “Is disaggregation the holy grail of energy efficiency? The case of electricity.” Energy Policy (2012).

[11] Hart, George William. ”Nonintrusive appliance load monitoring.” Proceedings of the IEEE 80, no. 12 (1992): 1870-1891.

BIOGRAPHY

Sourodeep Bhattacharjee has completed Bachelor of Technology in Computer Science (2010) from West Bengal University of Technology and Master of Science in Computer Science (2012) from University of Windsor, Windsor, Ontario, Canada. He is currently employed as Senior Research Fellow with CSIR-Cetral Mechanical Engineering Research Institute (C.M.E.R.I) Embedded Systems Laboratory . His research interests revolve around machine learning, artificial intelligence, and energy informatics.

Anirudh Kumar is Scientist of Embedded Systems Lab, CSIR-C.M.E.R.I. as well as PhD student of ACSIR, C.M.E.R.I. He received his M.Tech. degree in Mechatronics from ACSIR, C.M.E.R.I. His B.Tech. degree was in Computer Science from Madan Mohan Malviya Engineering College, Uttar Pradesh. His research interest is in the domain of Real-Time Systems.

Dr. Joydeb Roy Chowdhury is the Senior Principal Scientist of Embedded Systems Lab, CSIR-C.M.E.R.I. He has completed PhD from NIT, Durgapur in Electronics, Instru- mentation and Control Engineering and M.E. from Jadavpur University, Kolkata. He has received Lock Heed Martin (USA) India Innovation Award - 2007 for development of Smart Card Operated Prepaid Energy Meter with Automatic Load Controller for Solar Power Application. His research interest is in the area of Real-Time Embedded Systems.