Learning Systems for Electric Consumption of Buildings

Mario Berges, Ethan Goldman, H. Scott Matthews, Lucio Soibelman
{mberges, ejgoldma, hsm, lucio}@andrew.cmu.edu
Civil and Environmental Engineering Department
Carnegie Mellon University

Appeared in: *Proceedings of the 2009 ASCE International Workshop on Computing in Civil Engineering, Austin, Texas.*

Abstract

Individual appliances' electricity consumption is automatically disaggregated from a single custom metering system on the main feed to an occupied residential building. A data acquisition system samples voltage and current at 100 kHz, then calculates real and reactive power, harmonics, and other features at 20Hz. A probabilistic event-detector using the generalized likelihood ratio (GLR) matches human-labeled events to the time-series of features. Machine-learning classification was most successful with the 1-nearest-neighbor algorithm, correctly identifying 90% of the laboratory-generated training events and 79% of validation examples. The challenge of obtaining adequate training data for the real-world home leads to the development of the *Wire Spy*, a wirelessly-networked event detector with an inductive sensor which clamps to the cable of an appliance.

Introduction

Reducing energy consumption in buildings has both economical and ecological benefits, but even if motivated, building owners often lack clear direction. Because typical electricity meters and bills report usage infrequently and in aggregate, they are of little use in determining the best methods for conservation. A metering system that broke down electricity consumption by end use and gave feedback at more frequent intervals would aid users both in identifying worthwhile targets and in recognizing what actions were effective in reducing those loads. While the additional hardware needed to explicitly meter individual circuits and appliances would be prohibitively expensive, Non-Intrusive Load Monitoring (NILM) has shown promise in distinguishing the behavior of individual electrical loads by analyzing the composite consumption. However, applying this approach to an actual house presents new challenges beyond those of a more controlled laboratory environment.

Context

Out of the 102 quadrillion BTU (quads) of total annual primary energy consumption in the United States, 40% is used to generate electricity. Of this number, 36% is consumed by commercial buildings and 37% by the residential sector. (Energy Information Administration 2008)

The problem of estimating the electricity consumption of a particular building has many different solutions depending on the level of detail sought by the end-user. A wide spectrum of estimates can be obtained, each with an associated cost and degree of uncertainty. These go from national averages down to real-time measurements for individual appliances. The answers to any specific inquiry a user may have lie somewhere in this spectrum and it is the choice of uncertainty level and cost that determines the right solution.

Generally, the more detailed the information we want to obtain about a building's electricity consumption, the higher the price we have to pay to obtain it (Berges et al. 2008). Commercially available power meters can provide traditional power metrics at varying degrees of detail: the building's main electrical feed, circuit panels, individual circuits or individual outlets. These solutions have varying prices, and if deployed to estimate the total building consumption, their cost tends to increase with the level of detail obtained.

Given the relationship that exists between information and cost, there are various possible propositions to answer the problem, each of which will consist of a different combination of technology, granularity, and uncertainty. In the United States, just 12 types of appliances account for 80% of residential electricity consumption (Energy Information Administration 2007). Therefore, if instead of placing the most detailed measurement tools (i.e. outlet level power meters) on all appliances throughout the building, we only install them on those that are shown to be the biggest consumers, then we can start to make intelligent compromises about the cost and uncertainty.

Another possible solution is to install aggregate level power meters (e.g. total building consumption) and extract detailed information by post-processing it, taking advantage of the fact that the different electrical loads in a building have a distinct way of drawing power. We could measure the building's aggregate consumption (i.e. at the main feed), at a high enough rate that would allow specialized signal processing and machine learning algorithms to classify distinctive characteristics of the measured signals (voltage and current) that are associated with the operation of individual appliances. This disaggregation solution has a low cost and a high level of detail, and would be preferred over others.

Problem Description

In its simplest form the problem is to find a way to obtain high value, detailed information about electricity consumption in a building, at a low cost, in order to provide support for homeowners and facility managers during the decision making process around energy issues.

Detailed information, in this context, refers to metrics for electricity consumption at a finer granularity of time and/or power consumption than what the utility's monthly bill offers. For example, we could try to disaggregate the monthly totals by providing the daily (finer timescale) power consumption of the individual appliances (finer power increments) in the building. This would result in a richer information model, and is the main idea we explore in this paper.

Stated more precisely, our problem is two-fold. We would like to: 1) automatically disaggregate a buildings total electrical load into the individual appliances that compose it; and 2) use the disaggregated information to create a decision support system that would help users address their energy saving needs more effectively. We intend to achieve these goals in a non-intrusive fashion, maximizing the use of the existing infrastructure rather than imposing the need to install various new devices in the building, thus reducing the associated hardware and labor costs.

It has been previously shown, as we will present in the section that follows, that under certain conditions (i.e. on a laboratory setting) disaggregating the total load is an achievable task, and can be done with a relatively high degree of accuracy. However, the problem is much more challenging in the real world, where not only does the number and type of appliances increases, but also the measurements are susceptible to more noise and obtaining ground truth data becomes more difficult.

This paper will present some initial results we have obtained, as well as relevant findings resulting from our work towards the goal just described. Particularly, we will describe our experiences in learning from the building's main electrical feed. We present experimental results of our implementation of Non-Intrusive Load Monitoring (NILM) algorithms in a laboratory setup, the problems with obtaining ground truth data along with our solution for this, and some comments about the issues that arise when deploying these systems in the real world. The paper concludes with a discussion of the experimental findings, and a list of future tasks that need to be completed.

Previous Work

Before exploring the literature, we decided to investigate the available commercial solutions that were in the market. We found a wide range of products (Berges et al. 2008, Matthews et al. 2008), but very few addressed our specific problem. A variety of plug-through power meters are available, designed to monitor the consumption of individual appliances. However, within this group only a small portion provided a communication link that enabled the users to concurrently monitor all the meters in the building from a central location. On the other end of the spectrum there were numerous power meters designed to measure the total building consumption at the main feed. They use a variety of different methods to obtain their readings: some rely on the utility's power meter, others attach current transducers to the main electrical lines, etc. Nevertheless, these solutions had a very low reporting rate for our purposes, providing an updated power measurement no more than once a second.

There are also some intermediate solutions, which measure individual electrical circuits in the building. We have installed one of such systems for two electrical panels in a building at Carnegie Mellon University. We found that even though the acquired data had a higher level of detail than that which would be obtained from measuring the total building consumption, the hardware and labor costs far exceed the value of this information. We needed to trace all circuits to confirm the loads that were being served, and even after doing so we did not completely eliminate uncertainty. One of the reasons for this is that some circuits feed a variety of different, sometimes non-fixed loads (Berges et al. 2008).

Commercial products meeting our needs were virtually inexistent. We did find, however, a company that manufactures power meters which monitor the total electrical load of a building and can infer the operational schedule of large loads. This company, Enetics, is leveraging the results of research by George Hart (1992). Non-Intrusive Load Monitoring, as this type of monitoring is known in the literature, has been studied for the past twenty years, producing promising laboratory results but no significant incursion into the market. We then turned our search to the literature.

In Shaw et al. (2008) a transient event classification scheme is presented, which summarizes and implements algorithms and techniques discussed earlier (Norford and Leeb 1996, Laughman et al. 2003). The main idea described therein is that by acquiring high frequency samples of current and voltage at the main feed, and transforming these into power- and reactive power-like features, it is possible to detect changes in this new feature space that correspond to appliance state changes, and by carefully analyzing the transient associated with this event, it is possible to classify it as belonging to a particular appliance state change.

Other researchers have applied similar ideas to signals sampled at lower frequencies, such as readings from the utility's analog meter using an optical sensor (Baranski and Voss 2003) and 15 minute samples of the total power consumption (Prudenzi 2002). Incorporating other information sources like time of day has also proven to be effective (Farinaccio and Zmeureanu 1999). Some researchers have also tried to omit the event detection step, and proceed straightly to the classification of appliances based on analysis of the current in the frequency domain (Srinivasan 2006).

Learning from the Electric Mains

We now proceed to describe our experiments and present the results obtained so far, in our exploration of the use of aggregate power readings to estimate the operational schedule of appliances in the building. The results come from our implementation of a NILM system prototype, which we used to collect electricity readings from a laboratory setup where we had eight appliances connected to a power strip. We will also refer to the initial challenges we faced when redeploying the system in a real world scenario, acquiring data from a split-phase electrical panel feeding all the circuits of an occupied residential building. Specifically, we will present our solution to the problem of obtaining ground truth labels for the appliance state transitions occurring in a house.

The idea behind the experiments presented here is to capture power metrics over a period of time, during which the state changes for all electrical appliances of interest are also being recorded (either automatically or as supplied by a person). This data will later be transformed into a list of events (i.e. appliance state transition) that are composed of a time-stamp, a label, and a series of features associated with the transition. Finally, we train classification algorithms on a subset of these records, and test them on the remaining to evaluate their performance. A more detailed description of this process follows.

Hardware setup

The dataset on which we will base our analysis was obtained using a voltage transformer and current transducer connected to a custom-built 20A electrical outlet (Figure 1), to which a series of everyday use residential electrical appliances are connected. A data acquisition device (DAQ) acquires the aggregate analog voltage and current signals at a 100 kHz rate, converting them to digital signals and transmitting them to a computer via a USB cable. A LabVIEW program then makes use of the sampled signals to compute a number of traditional power metrics at a 20Hz rate: watts of real power (W), unsigned reactive power (VAR), RMS voltage (VRMS), amplitude of odd-numbered current harmonics, etc. For the experiments described in this paper, we focused only on the first two metrics (real and reactive Power).

Eight different appliances, as illustrated in Table 1, were monitored in this dataset, creating a total of 34 possible distinct state transitions (e.g. going from off to on, low to high, etc.). These transitions become the classes, as we frame the problem as a classification task. The dataset contains 483 labeled events distributed non-uniformly across the classes.

We will refer to this dataset as the noise-free (NF) dataset, because of the nature of how the samples were obtained: a laboratory environment where we had knowledge of the duty cycle of every individual load that composed the aggregate signal.

Disaggregation

The problem is in essence a multi-class classification problem. We want to extract, from the original power metrics series, a select number of features about the samples around each and every event (appliance state transition), and find a function that can map vectors in this new feature space to a finite set of classes which correspond to the possible state transitions.

Event Detection

The first part of the problem is the detection of the appliance state transitions that are of interest to us. We have human-supplied time-stamps for these datasets, but if this approach is to be effective in the future, we need to be able to detect those events automatically from the aggregate power metrics. In other words, for us to be able to classify state transitions we must first be able to recognize when they occur.

We have used changes in real power as indicators of the state transitions. However, characterizing changes in real power is not a straightforward task. A change of 5W, for instance, between one sample and the following can be an indication of: a) a state transition for one or more appliances; b) signal noise; c) part of the normal operation of a given appliance without it being a state transition. For this reason we decided to use a probabilistic model to detect such changes. More specifically, we used a modified version of the generalized likelihood ratio (GLR) presented in Luo et al. (2002).

While we have had some success utilizing the human-supplied labels and correcting them with the automatic event detector, it is time-consuming to generate a large enough set of labeled events on which to train the algorithms. To date, we have used a two-person procedure where one person switches appliances on and off (or into other states, if possible) and announces each change as they perform it. The second person uses the metering system interface to record each event in real-time. In order to correct small time delays between the actual event and the human-recorded timestamp, the event detection algorithm is used to shift human-labeled events to match the exact time at which the event was detected. As the event detector is permitted to adjust the time-stamp of a label by up to four seconds, a delay of at least 10 seconds between generated events is enforced to prevent ambiguity. This rate is the upper limit on acquiring labeled events; in practice an hour's work typically yields closer to 100 labeled events. As this is a fairly tedious process, it is unlikely that anyone without an extraordinary interest in the outcome would have the patience to persist long enough to acquire an adequate set of labeled events. In short, this is not a practical approach for end-users to train the system on the appliances in their house. We will present our solution to this particular problem in a later section.

Feature Extraction

Once we have established an appropriate time-stamp and label for all events, the next step is to extract features from a signal composed of samples in the neighborhood of each event. We capture transients from the real and reactive power of all available voltage sources, and concatenate them into the signal that is presented in the graph.

For the NF dataset, we decided to concentrate our efforts on extracting some simple features. The first and simplest we will call the "delta" metric and is simply the difference between the post- and pre-event window average real power. Other features tested include regularized regression coefficients using linear and non-linear basis functions of different orders. We also experimented with different window sizes for the captured transients. Our search was guided by the average 10-fold cross validation accuracy of a 1-Nearest Neighbor classifier.

Pattern Matching

The final step is to present the feature vectors along with their associated classes (i.e. appliance state transition) to a classification algorithm in order to train it. From the original NF dataset, 449 events were chosen for training and selecting the right parameters for the feature extraction algorithms, as shown in the previous section. The remaining examples were reserved for validation of the algorithms in the end.

We tested four different classifiers on each of the feature sets: 1-Nearest Neighbor (1-NN), Gaussian Naïve Bayes (GNB), Decision Trees (DT) and Multiclass Adaboost (MultiBoost).

Experimental Results

Each algorithm performed differently, but the best results were obtained using the 1-NN. During training, the average 10-fold cross validation accuracy of this classifier was 90%, using the Fourier regression coefficients. In other words, by randomly showing the algorithm 45 examples for it to compare against the 405 remaining, on average, this algorithm only misclassified less than 5 out of the 45. The other algorithms performed with lower accuracy: DT – 85%, GNB – 83%, MultiBoost – 76%.

Using the 44 examples reserved for validation, we obtained a better estimate of the true accuracy. The events in these validation sets were chosen to be approximately uniformly distributed across all possible state transitions (i.e. almost one example per class). The highest accuracy was obtained, again, using the 1-NN classifier and the Fourier coefficients, with a 79% accuracy on labeling those unseen examples. Again, the rest of the algorithms showed lower accuracy: GNB – 67%, DT – 64%, MultiBoost – 51%.

Automatic Labeling

While these results are encouraging, they require multiple labeled examples for each class (on the order of one hundred classes for a typical house) for training purposes. Besides the issues of synchronizing timestamps from human-labeled training examples, as presented earlier, an additional challenge comes from appliances that change states on their own. One example of this is appliances with thermostats, such as refrigerators, water heaters, and some electric ranges. In order to label events from these appliances it is either necessary to manually force them to switch on and off, for example by manipulating the thermostat of a refrigerator, or to listen for cues to a state change. Worse, the human-labeled timestamp associated with turning a burner off might fall during a period where the burner is not drawing any power, so the recorded "event" does not correspond to any change in electrical consumption. Further, there are appliances like dishwashers and clothes washers that have multiple states that they pass through. Simply labeling the "on" and "off" events and assuming a steady-state power level to calculate total energy consumption will lead to a significant error.

These challenges point to the need for reliable ground truth with which to train the classification algorithm. One planned approach was to use an experimental plug-through meter with wireless communication capabilities called the *Jiga –watt*. However, some appliances can't be metered with plug-through devices, either because they have an atypical plug or no plug at all. Unfortunately, these are often large loads which are both important contributors to the overall consumption of the home and also tend to have indirectly-driven events that are hard to manually tag.

To overcome this problem, we have developed a hardware-based event detector called the *Wire Spy* to automatically label on/off cycles for a single appliance. Because it clamps onto the appliance's power cable it is not necessary to even unplug the appliance, allowing it to monitor appliances with non-standard plugs, no plugs, or that are inconvenient to unplug. It is battery-powered and transmits the events over an 802.15.4 wireless connection to the main data acquisition computer, so it can be deployed anywhere within the 50-100m range. This allows us to easily move the Wire Spy between appliances once enough events have been recorded.

The Wire Spy senses appliance activity through a clamp-on current inductor made by modifying a CR Magnetics 3110 current transformer (CT). Rather than sensing the current flow through a single conductor, the CT is affixed to the whole appliance power cable, including both conductors (and sometimes a ground wire as well). Because the current flows in the two conductors are of equal magnitude and opposite directions, their respective magnetic fields will nearly cancel one another out, and the CT will register ~0A, even when the appliance is drawing a significant current. However, by amplifying and processing the minute signal from the CT in this configuration, it is possible to detect a value which is roughly proportional to the actual current flow to the appliance. We hypothesize that the existence of this "shadow" of the current is due to the small distance between the conductors. This places one of the conductors slightly closer to the inductor, and thus it induces a slightly stronger current in proportion to the actual flow.

The Wire Spy uses an embedded AVR® 8-bit microcontroller to record user-set on/off thresholds, digitize the sensor values, and perform simple signal processing. For each appliance, the user "trains" the Wire Spy on a single example each of the "off" and "on" states, from which it derives thresholds. The microcontroller then records a peak-to-peak amplitudes at 20Hz and stores the last 16 samples. If the average value of the newer half corresponds to a different threshold than that of the older half, it triggers the event detector.

When an event is detected, a notification is sent from the Wire Spy to the main data acquisition computer, which is running the whole-house meter. This notification contains a shared appliance ID (input by the user), as well as the pre- and post-event signal averages. The whole-house metering software records the event along with its timestamp. Currently this approach has been developed specifically for binary (on/off) devices, but early testing suggests that the values associated with these states are clustered tightly enough that the Wire Spy should be able to monitor appliances with multiple states, such as a window fan with off/high/low settings.

Discussion and Conclusions

We have shown that non-intrusive load monitoring is achievable in the laboratory, even when using very simple features to describe the power transients associated with every appliance state transition. However, moving this system to the real world introduces various different new challenges.

One such obstacle is the issue of obtaining ground truth data reliably and in an automated fashion. We presented the Wire Spy as a possible solution that could not only facilitate the training stage for the users, but could also become a stand-alone application to use in addition to a whole-building power meter and obtain appliance-level consumption information.

Future Work

We plan to further explore the use of the system in a real world setting, with data that is currently being collected in an occupied residential building. Our intention is to experiment with the features and algorithms presented in this paper, with the possible additions of others given that we expect to be dealing with noisier transients.

Additionally, we intend to account for the structure that exists in the sequential ordering of the appliance states. For example, a light bulb can only turn on if it was previously in the off state. Finally, we would like to investigate the use of other sources of information, to enhance the disaggregation and continue with the non-intrusive approach to solving the problem.

Acknowledgements

The authors would like to gratefully acknowledge the support of the Robert Bosch LLC Research and Technology Center. Additional support was provided by the Pennsylvania Infrastructure Technology Alliance (PITA).

References

Baranski, M., and Voss J. "Nonintrusive appliance load monitoring based on an optical sensor." In *Power Tech Conference Proceedings, 2003 IEEE Bologna*, 4:8 pp. Vol.4, 2003. doi:10.1109/PTC.2003.1304732.

Berges, Mario, Ethan Goldman, H. Scott Matthews, and Lucio Soibelman. "Training Load Monitoring Algorithms on Highly Sub-Metered Home Electricity Consumption Data ." *Tsinghua Science & Technology* 13 (2008): 406-411.

Energy Information Administration. "Annual Energy Review 2007." (2008) 10 December 2008 <http://www.eia.doe.gov/aer/pdf/aer.pdf>.

Energy Information Administration. "End-Use Consumption of Electricity by End Use and Appliance." (2007) 18 November 2008 <http://www.eia.doe.gov/emeu/recs/recs2001/enduse2001/enduse2001.html>.

Farinaccio, Linda, and Radu Zmeureanu. "Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses." *Energy and Buildings* 30, no. 3 (August 1999): 245-259.

Hart, G.W. "Nonintrusive appliance load monitoring." *Proceedings of the IEEE* 80, no. 12 (1992): 1870-1891.

Laughman, C., Kwangduk Lee, R. Cox, S. Shaw, S. Leeb, L. Norford, et al. "Power signature analysis." *Power and Energy Magazine, IEEE* 1, no. 2 (2003): 56-63. doi:10.1109/MPAE.2003.1192027.

Luo, Dong, Leslie Norford, Steven Shaw, and Steven Leeb. "Monitoring HVAC Equipment Electrical Loads from a Centralized Location - Methods and Field Test Results." *ASHRAE Transactions* 108, no. 1 (2002): 841-857.

Matthews, H. Scott, Lucio Soibelman, Mario Berges, and Ethan Goldman. "Automatically Disaggregating the Total Electrical Load in Residential Buildings: a Profile of the Required Solution." In *Intelligent Computing in Engineering (ICE08) Proceedings*. Plymouth, 2008.

Norford, Leslie K., and Steven B. Leeb. "Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load-detection algorithms." *Energy and Buildings* 24, no. 1 (1996): 51-64.

Prudenzi, A. "A neuron nets based procedure for identifying domestic appliances pattern-of-use from energy recordings at meter panel." In *Power Engineering Society Winter Meeting, 2002. IEEE*, 2:941-946 vol.2, 2002.

Shaw, S.R., S.B. Leeb, L.K. Norford, and R.W. Cox. "Nonintrusive Load Monitoring and Diagnostics in Power Systems." *Instrumentation and Measurement, IEEE Transactions on* 57, no. 7 (2008): 1445-1454.

Srinivasan, D., W.S. Ng, and A.C. Liew. "Neural-network-based signature recognition for harmonic source identification." *Power Delivery, IEEE Transactions on* 21, no. 1 (2006): 398-405.

Figures and Tables

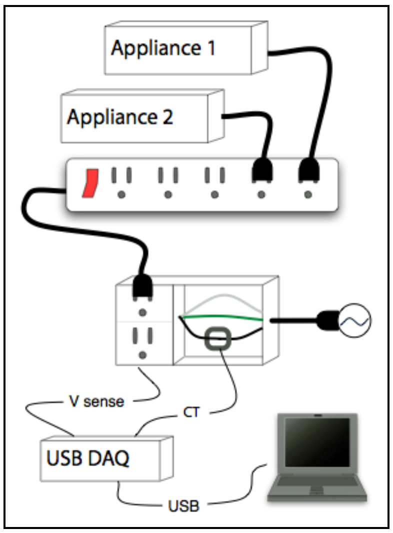

Figure 1: Diagram for the Noise-Free (NF) experiment setup.

Table 1: List of appliances and possible states for the Noise-Free dataset

APPLIANCE	STATES
1 : Radio/Casette	0: unplugged, 1: clock, 2: radio, 3: cassette
2: Light bulb 1	0: off, 1: on
3: Light bulb 2	0: off, 1: on
4: Fan	0: off, 1: low, 2: high
5: Small Fridge	0: off, 1: on
6: Microwave	0: unplugged, 1: stand-by, 2: cooking
7: Toaster	0: off, 1: normal, 2: bagel, 3: frozen
8: Hairdryer	0: unplugged, 1: plugged-in, 2: low, 3: high