Power Disaggregation for Low-sampling Rate Data

Jing Liao, Georgia Elafoudi, Lina Stankovic, Vladimir Stankovic
Department of Electronic and Electrical Engineering
University of Strathclyde, Glasgow, UK
Email: {jing.liao, georgia.elafoudi, lina.stankovic, vladimir.stankovic}@strath.ac.uk

***Abstract*—In this paper, we focus on energy disaggregation at low-sampling rates (at 6sec and 1min) and use only active power measurements for training and testing. Specifically, we develop two algorithms: one is a low-complexity, supervised approach based on Decision Trees and another is an unsupervised method based on Dynamic Time Warping. Both proposed algorithms share common pre-classification steps. These are benchmarked with a state-of-the-art Hidden Markov Model (HMM)-based approach. Experimental results using the REDD dataset as well as data collected from a real UK household, show that the two proposed methods outperform the HMM-based approach and are capable of disaggregating a range of domestic loads even when the training period is very short.**

I. Introduction

Disaggregating individual appliance usage from the total, aggregated energy consumption captured at the energy monitor, is referred to as non-intrusive appliance load monitoring (NALM) and was first proposed by Hart in [1]. Since [1], many NALM algorithms have been proposed to adapt to advances in sensor technology capturing energy measurands at a range of sampling rates, generally in the order of kHz. However, more recently, as smart meter deployments are underway across the world and with increased awareness of domestic energy efficiency, NALM algorithmic work is focusing on cost-effective equipment that samples data at rates under 1Hz. It is not only the cost of the sensing technology [2], but also computational and storage cost as well as implementation efficiency that are key drivers towards the wide deployment of low-sampling smart meters. However, so far, there are no widely available efficient solutions for NALM, with high accuracy, at low sampling rates [3].

For example, Kim *et. al* [4] propose four different methods for low-rate power disaggregation using (conditional) factorial Hidden Markov Model (HMM) and Hidden semi-Markov models. Using active power measurements collected every 3sec together with duration and time of use of appliances, accuracy ranging from 72% to 99% was obtained for 7 different houses with up to 10 appliances with an average accuracy of 83%. This method cannot disaggregate base load, such as, for example, refrigerator, and is prone to converge to a local minimum. In [5] a factorial HMM is used for disaggregation of active power load. The algorithm proposed in [5] is an unsupervised method that uses expert knowledge to set initial models for states of known appliances; however, the models' operation for reliable results depends on correctly setting the a priori-values for each state for each appliance, which in turn is strongly dependent on the particular aggregate dataset on which NALM is being performed. Indeed, a similar factorial HMM-based approach is tested in [6], where it is shown that the disaggregation accuracy drops for up to 25% when different houses are used to set the initial models compared to the case when the same house is used for building the models and testing. Results are reported for REDD dataset [6] with sampling rates of 1sec and 3sec.

In [7], an unsupervised Additive Factorial Approximate Maximum A-Posteriori (AFMAP) inference algorithm is proposed using differential factorial HMMs. First, all snippets of active power data are extracted using a threshold and modelled by an HMM; next the k-nearest-neighbor graph is used to build nine motifs that are treated as HMMs over which AFMAP is run. The results show average accuracy of 87.2% using 7 appliances and sampling rate of 60Hz.

In [8], the potential of using Dynamic Time Warping (DTW) for disaggregation is demonstrated, using energy consumption, rising edge count and Discrete Fourier Transform (DFT) as diagnostic features calculated over a window. It is shown that in contrast to speech applications of DTW, DFT is not needed, due to relatively small variations in the load within one window. The performance of the approach is unclear since only one-day's worth data is used for testing with sampling rate of 10sec in laboratory conditions.

In [9], a decision-tree (DT) classifier is used for pattern matching. However, in contrast to our work where we use active power samples at less than 1Hz, [9] uses transient features in addition to active and reactive power and collects samples of voltage and current at 100kHz. Using human-supplied labels corrected with the generalized likelihood ratio for event detection, the decision tree algorithm of [9] achieved accuracy of only 64% compared to the 1-nearest neighbour classifier with the Fourier regression coefficients that achieved 79% when tested in laboratory conditions with eight loads.

Despite tremendous research efforts, there is no evidence that current NALM algorithms are able to disaggregate appliances in a home environment with high accuracy using sampling rate above one second [3]. This paper focuses on NALM with sampling rates of 6sec or more, using active power measurements only. We propose two NALM methods that differ in the classification step. The first method is a supervised, decision tree (DT)-based NALM method and the second, unsupervised Dynamic Time Warping (DTW)-based NALM method. The proposed DT-based approach is a low-complexity approach, that can be trained using a very small dataset. It uses only active power changes during appliance state transitions to identify appliances while ignoring power fluctuations within each state. On the other hand, the DTW-based method is unsupervised and scalable in the sense that it does not require re-training of the entire dataset when a new appliance is added. In contrast to DT that only looks at the range of power changes, DTW exploits the entire appliance signature to perform template matching. The two methods share the same pre-processing step, to identify and remove corrupted data, and the same event detection algorithm; these significantly reduce the DTW library size and hence classification/pattern matching complexity.

We validate the proposed methods using real smart meter measurements for nine typical appliances acquired over a period of two months in a real UK household as well as the REDD dataset [6] and compare our results to a state-of-the-art HMM-based approach of [5] proposed for low sampling rates.

II. Proposed Disaggregation Algorithms

A. Notation and Problem Formulation

Let $\mathbf{M}$ be a set of all test appliances, which is assumed to be known. Let $p(t_i)$ be measured active power at time t_i. Without loss of generality, we denote $p(t_i)$ as $p(t_i) = p(iT) = p(i)$, where $T = t_i - t_{i-1}$ is the sampling interval.

For all $i = 1, \ldots, N$, the disaggregation task is to express $p(i)$ as $p(i) = \sum_{j=1}^{M} p_j(i) + n(i)$, where $p_j(i) \geq 0$ is the power load of appliance j at time instance iT and $n(i)$ is the measurement noise at time instance iT. Note that $p_j(i)$ is zero if the appliance is off. Measurement noise comprises acquisition and communications noise as well as consumption of all unknown appliances.

During the training phase, p_j's are estimated for some or all appliances using only the aggregate data. The obtained values are fed into a library of *appliance signatures*. Note that the training phase needs to be re-executed whenever appliances change (i.e., addition, change or removal of appliances).

B. Event Detection

If $|p_j(i) - p_j(i-1)| \geq W$, we say that the appliance j has *changed state* at time instant iT, where W is a threshold. For single-state (on-off) appliances, this means that the appliance was switched on or off. For multi-state appliances, such as washing machine, $|p_j(i) - p_j(i-1)| \geq W$ indicates that the appliance has transited from one state to another.

W is set low enough so that for all j, if $|p_j(i) - p_j(i-1)| \leq W$, appliance j did not change its state or, otherwise, it did. All appliances that operate below threshold W will not be detected. W is initialized during the training process based on the minimum state transition that needs to be detected and the maximum variation of the active power within one appliance state across all appliances' states, that is

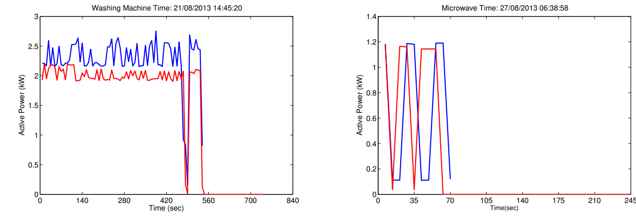
$$W = \max\{\min_{m \in \mathbf{M}} \mathbf{p_m}, \max_{\mathbf{m} \in \mathbf{M}} |\max(\mathbf{p_m}) - \min(\mathbf{p_m})|\}, \quad (1)$$

where $\mathbf{p_m}$ is a vector of active power readings of appliance m and W depends on the set of appliances in the dataset. W is kept fixed for the DT-based method. For the DTW-based method, in each iteration, one appliance is detected and its power load removed from the aggregated readings going from higher to lower consuming appliances. After the detected appliance is removed, W reduced using (1).

An *event* occurs whenever an appliance changes its state. Edge detection is used to detect events by comparing $|p(i) - p(i-1)|$ with W. We say that the event started at time l_s and ended at l_e, if an appliance changed its state at l_s and l_e, such that

$$|[p(l_s) - p(l_s - 1)] + [p(l_e) - p(l_e - 1)]| \leq C,$$

where C is a parameter smaller than W.



(a) Washing Machine. (b) Microwave.

Fig. 1. Examples of appliance identification. DTW library signatures are in blue and detected signatures in red. Sampling rate is 6sec.

C. Feature Extraction

For the DT-based method, from each identified event window, we extract the first increasing edge at the start of the event window and the last decreasing edge at the end, that is, $|p(l_s) - p(l_s - 1)|$ and $|p(l_e) - p(l_e - 1)|$, respectively. During the training phase, for each appliance m, the *range* of $\mathbf{p_m}$ is estimated by finding the maximum and minimum of $|p_m(l_s) - p_m(l_s - 1)|$ and $|p_m(l_e) - p_m(l_e - 1)|$ over the entire training set. Note that this significantly reduces the library size, since for each appliance only two values are stored.

During training, the DTW-based method collects different instances of signatures that are stored in the library. An example of the identification of washing machine and microwave for 6sec experiments is shown in Fig. 1.

D. Classification and Pattern Matching

Classification using the DT-based method is straightforward and is based on finding a unique path in the trained decision tree model for the extracted rising and falling edges, that is, $|p(l_s) - p(l_s - 1)|$ and $|p(l_e) - p(l_e - 1)|$, respectively.

For the unsupervised DTW-based method, classification is performed via pattern matching. Given two windows of possibly different lengths, DTW performs non-linear mapping of one window to another by minimizing the distance between them via dynamic programming. DTW is more effective in our problem than conventional clustering, such as k-means, because the duration of an appliance usage at different times can significantly vary due to a change in appliance settings.

DTW algorithm performs pattern matching as follows: Let $p(1), \ldots, p(m)$ and $q(1), \ldots, q(n)$ be two events. Let $d(x, y) = |x - y|$ for any two readings x and y. Let $D(0, 0) = 0$ and $D(i, 0) = D(0, i) = \infty$, for all $i > 0$. $D(i, j)$ is the accumulated DTW distance between points $p(1)$ to $p(i)$ and points $q(1)$ to $q(j)$, and $D(n, m)$ is the final distance between the two vectors.

After event detection and feature extraction, the best match is found in terms of minimizing $D(n, m)$ distance between the testing event and all signatures in the database. Note that the output of the DTW classification is a soft value - $D(n, m)$ -

TABLE I. COMPARISON BETWEEN THE THREE METHODS USING REDD DATA.

	PR(%)			*RE*(%)			F_M(%)		
House	DTW	DT	HMM	DTW	DT	HMM	DTW	DT	HMM
House 1	85.01	92.12	79.29	79.72	64.98	62.72	82.28	76.20	70.04
House 2	89.86	68.49	51.80	84.39	81.42	62.79	87.04	74.39	56.77
House 6	98.86	75.68	99.30	81.22	88.61	74.74	89.17	81.63	85.29

Algorithm 1 DTW pattern matching: Find a distance between two vectors of possibly unequal lengths.

function DTW(p, q)
  $n = length(p)$, $m = length(q)$
  for $i = 1, \dots, n$ **do**
    for $j = 1, \dots, m$ **do**
      $d(i, j) = p(i) - q(j)$
      $D(i, j) = d(i, j) + \min\{D(i-1, j), D(i-1, j-1), D(i, j-1)\}$
    end for
  end for
  Output $D(n, m)$
end function

the minimum distance between the testing event and the best matching window. This minimum distance is compared to a threshold (*dtw threshold*), which is the acceptable minimum distance between two windows, in order to classify them as a match. This threshold is appliance specific and it depends on the state variations of each appliance, e.g., *dtw threshold* is higher for washing machine than toaster.

III. EXPERIMENTAL RESULTS

A. Experimental Methodology and Evaluation Metrics

We evaluate the proposed disaggregation methods through the following two experiments: (1) evaluation using three houses from the REDD database downsampled to 1min resolution; (2) experiments using actual smart meter data from a UK test house at 6sec sampling rate. We compare all our results with the state-of-the-art HMM-based method of [5] designed for low-sampling rates.

The evaluation metrics used are precision (PR), recall (RE) and F-Measure (F_M) [10] defined as:

$$PR = TP/(TP + FP) \tag{2}$$

$$RE = TP/(TP + FN) \tag{3}$$

$$F_M = 2 * (PR * RE)/(PR + RE), \tag{4}$$

where true positive (TP) presents the correct claim that the appliance was used, false positive (FP) represents an incorrect claim that an appliance was used, and false negative (FN) indicates that the appliance used was not identified.

For each dataset, all three tested algorithms always use the same amount of data for training and testing. The HMM-based method [5] requires prior initialization of the model using expert knowledge (state variances, mean value for each state and state transition probabilities), which was carried out in our experiments either using the information from [5], or were generated during training.

During the experiments, the threshold W is set to 60W for DT method. Power variation for both methods is set to $C = 5\% - 20\%$ of the first rising power edge of the testing event, where a small C (percentage-wise) is used for larger power increases.

We use House 1, House 2 and House 6 in the REDD database with known appliances: refrigerator, washer dryer, dishwasher, air conditioner, toaster and microwave. Five weeks of data are used, where the first week is used for training. Our results are shown in Table I, where the obtained accuracies are calculated using (2)-(4).

Table I shows that DTW and DT in both Houses 1 and 2 outperform HMM, which is mainly due to the HMM's low recognition accuracy for all the appliances except refrigerator. This also explains the good performance of HMM for House 6, where the number of refrigerator events is high compared to the events of the other appliances.

B. Disaggregation in the Test House

The obtained results using 6sec resolution in the UK-based test house are shown in Table II, where only 20% of the dataset is used for training. The HMM-based method could not detect any TV activity. This is probably due to the fact that TV is usually on for a long time, thus there was no instance in the training set when TV was running alone, with no other appliance running in parallel. Other problems associated with detecting TV activity are that TV power values vary widely, as well as the duration of TV run-time from one day to the next. On the other hand, the DT-based and DTW-based proposed approaches can successfully be trained even when appliance usage overlaps. All other HMM results are similar to those reported in the literature showing average success rate between 60% and 80%. The DT-based method consistently outperforms HMM, and the DTW-based method is always better than HMM except for the boiler, which has very low power consumption. As expected, the DTW-based method is the best performer for those appliances that have multiple operation states, such as washing machine and microwave.

IV. DISCUSSION AND CONCLUSION

The two proposed approaches share the same pre-processing and event detection steps, thus the performance difference comes purely from different classification steps. The proposed DT-based method is a low-complexity approach that requires the least storage and computational resources. It operates purely on the rising and falling active power edges at the start and end of events, and requires only a small sample of training data to build a DT model. It is a supervised approach since it is trained on labeled aggregate data.

TABLE II. COMPARISON BETWEEN THE THREE METHODS USING 6SEC COLLECTED DATA.

	PR(%)			*RE*(%)			F_M(%)		
Appliances	DTW	DT	HMM	DTW	DT	HMM	DTW	DT	HMM
Boiler	0	89.52	94.52	0	96.91	73.40	0	93.07	82.63
Microwave	98.33	87.01	64.90	69.21	95.04	69.50	81.24	90.95	67.12
Toaster	69.16	87.50	67.21	96.10	71.01	58.57	80.43	78.40	62.60
Kettle	95.04	100	94.23	95.83	87.76	50.00	95.43	93.48	65.33
TV	100	81.48	0	50	68.75	0	66.67	74.58	0
Oven	100	41.18	100	100	87.50	62.50	100	56.00	76.92
Washing Machine	100	88.89	0	100	100	0	100	94.12	0
Hair Dryer	50.00	66.67	25.00	66.67	50.00	25.00	57.13	57.14	25.00
Pancake maker	29.51	50.00	42.86	48.65	50.00	75.00	36.74	50.00	54.55
Total	88.32	87.64	75.35	80.60	87.64	57.95	84.28	87.64	65.52

The proposed DTW-based method needs a library of appliance signatures, hence requires more storage space; it is computationally more expensive than the DT-based method as it performs pattern matching between the extracted window and all signatures present in the library. The complexity is thus driven by the number of signatures stored in the library. In the above experiments, for most appliances only one signature is kept, and there was no appliance with more than three signatures. Since training is done on unlabeled aggregate data, it is an unsupervised approach. It is also scalable, since in contrast to HMM and DT, it does not need to be retrained when a new appliance is added.

Experiments demonstrate that both proposed methods are successful even for small training data sets, which is not the case for probabilistic approaches such as those based on HMM. Indeed, HMM-based approaches require state variances and probabilities of state transitions which are difficult to estimate using a small training set. The DTW-based approach performs better with the lower resolution dataset (1min) than the DT-based method because it is very sensitive to fluctuations present in the 6sec dataset. The DTW-based method has difficulty recognizing low load appliances, such as boiler, but its strength lies in detecting multi-state appliances such as oven and washing machine. The DT-based approach performs very well for the appliances whose operation time significantly varies and tends to be long, such as TV, which is not the case with the HMM-based approach. Interestingly, the DTW-based approach shows good performance for the TV demonstrating that it is robust to fluctuations in event duration.

In summary, experimental results using REDD data and a real household based in the UK (test house) demonstrate the competitiveness of the proposed solutions with respect to a state-of-the-art HMM-based approach. Indeed, the proposed DT-based and DTW-based methods showed a success rate of over 84% in the UK test house with nine appliances as opposed to 65% obtained with the HMM-based approach.

Future work: 1) robustness of the two algorithms will be tested with respect to missing data, very limited training set, mismatched models, and the case when only manufacturer's manual's are used, 2) incorporating other sensors, such as temperature, light intensity, to improve disaggregation accuracy, 3) test the proposed approaches for disaggregating gas and water consumption.

ACKNOWLEDGMENT

This work is supported in part by the UK Engineering and Physical Sciences Research Council (EPSRC) projects REFIT EP/K002708 and APAtSCHE EP/K002368, under the Transforming Energy Demand in Buildings through Digital Innovation (BuildTEDDI) funding programme. The authors would like to thank A. Seeam for his help with data collection, O. Parson for sharing his HMM code, and J. Kolter and M. Johnson for making the REDD database available.

REFERENCES

[1] G. W. Hart, "Nonintrusive Appliance Load Data Acquisition Method", *MIT Energy Laboratory Technical Report*, Sept. 1984.

[2] K.C. Armel, A. Gupta, G. Shrimali, and A. Albert, "Is disaggregation the holy grail of energy efficiency? The case of electricity," *Energy Policy*, vol. 52, pp. 213–234, 2013.

[3] M. Zeifman and K. Roth, "Nonintrusive appliance load monitoring: Review and outlook," *IEEE Trans. Consumer Electronics*, vol. 57, no. 1, pp. 76–84, Feb. 2011.

[4] H. Kim, M. Marwah, M. Arlitt, G. Lyon, and J. Han, "Unsupervised disaggregation of low frequency power measurements," in *Proc. 11th SIAM Int. Conf. Data Mining,* Mesa, AZ, April 2011.

[5] O. Parson, S. Ghosh, M. Weal, and A. Rogers, "Non-intrusive load monitoring using prior models of general appliance types," in *Proc. the 26th Conf. Artificial Intelligence* , Toronto, CA, July 2012.

[6] J. Kolter, and M. Johnson, "REDD: A public data set for energy disaggregation research," in *Workshop on Data Mining Applications in Sustainability (SIGKDD)*, San Diego, CA, 2011.

[7] J. Kolter, and T. Jaakkola, "Approximate inference in additive factorial HMMs with application to energy disaggregation," in *J. Machine Learning*, vol. 22, pp. 1472–1482, 2012.

[8] F. Kupzog, T. Zia, and A.A. Zaidi, "Automatic electric load identification in self-configuring microgrids," in *Proc. IEEE Africon*, Nairobi, Kenya, Sept. 2009.

[9] M. Berges, E. Goldman, H. S. Matthews, and L. Soibelman, "Learning systems for electric consumption of buildings," in *Proc. 2009 ASCE Int. Workshop Computing in Civil Engineering*, Austin, TX, 2009.

[10] D.L. Olson and D. Delen, *Advanced Data Mining Techniques*, Springer, 2008.