Back to Basics: Simplifying Non-Intrusive Appliance Load Monitoring Using Combinatorial Optimization

Author1
Indraprastha Institute of Information Technology
India

Author2
Twentieth Century Fox
Springfield, USA
Email: homer@thesimpsons.com

James Kirk
and Montgomery Scott
Starfleet Academy
San Francisco, California 96678-2391
Telephone: (800) 555–1212
Fax: (888) 555–1212

***Abstract*—Non-Intrusive appliance load monitoring (NIALM) is the process of disaggregating the overall electricity usage into constituent appliances. In this paper we extend the Combinatorial Optimization (CO) approach for disaggregation, which was originally proposed in the seminal work on NIALM, in following two ways: 1) Breaking the problem into subproblems and reducing the state space; 2) Applying additional constraints backed by sound domain expertise. We evaluate our approach using REDD dataset and show practical problems which need to be solved while dealing with the dataset. We also propose a metric for evaluating NILM, which we believe overcomes many shortcomings of commonly used metrics.**

I. Introduction

- Motivate the importance of energy consumption in building
- Motivate that appliance level information is crucial - detailed feedback and optimized decision making [1]
- Challenges with getting appliance level information - introduce NIALM [2]
- introduce your proposed approach
- Enumerate the contributions

Primary contributions of our work are:

- Fill it up with 2-3 crisp points

Open source implementation of the proposed work is released for comparative analysis with other NIALM approaches as an IPython notebook[1]. We believe this is the first extensive release of a generic NIALM

II. Related work

NIALM has been well studied in the recent past and survey papers [3], [4], [5] present its classification across various dimensions. Following are three important classification dimensions:

- **Frequency of data collection**: Approaches such as harmonic analysis require data to be sampled at more than a thousand samples a second. Whereas approaches
- **Supervised/Unsupervised**:

[1]http://www.ipython.org

When you do the comparison, bring up how is your work different rather than just saying X did A and Y did B.

- Classification of different NIALM approaches - High/Low frequency, Time/Frequency domain analysis, supervised/unsupervised [3], [4], [5].
  For a more detailed overview the reader is referred to the above mentioned survey papers.
- Discuss the modeling approaches that are used
  - Additive Factorial HMM
  - Difference HMM [6]
- Datasets used: Recent datasets have spurred this field
  - REDD [7]
  - Blued [8]
  - Smart* [9]

III. NIALM

Discuss in brief the NIALM problem

A. Terminologies/ Notations

Borrow the notation used by Parson and Hart.

- Input: Aggregate power sequence: $x = \{x_1, .., x_T\}$
- Infer: Power draw by constituent appliance: $y_n = \{y_{1,n}, ..y_{T,n}\}$
- Time slice: $t \in 1, ..T$
- Appliance: $n \in 1, ..N$
- Appliance state: $z_n = \{z_1, ..z_T\}$

B. NIALM using combinatorial optimization

This approach resembles subset sum problem and tries to minimize the difference of total observed power from the sum of various possible subsets coming from various combinations of appliances in different states. For each appliance we assume **K** states and at a given time, an appliance can only be in a single state. This is given as: $z_{t,n,k} \in \{0, 1\}$ and,

$$\sum_{k=1}^{k=K} z_{t,n,k} = 1$$

The power drawn by each appliance is given by:

$$\mu_n = \{\mu_{n,1}, ..\mu_{n.K}\}$$

Thus, CO can be formulated as:

$$z_t = argmin_{z_t} |x_t - \sum_{n=1}^{N} \sum_{k=1}^{K} z_{t,n,k} \mu_{n,k}|$$

- Statespace is K^N
- We assign different loads to different mains, N_i loads to $Mains_i$, $\sum_1^p N_i = N$. Now different state spaces are K^{N_1}.... We can define the overall state space as $\max K^{N_i}$
  As a practical example, two mains, 20 appliance, state space before = 2^{20}. After = 2^{10}. Exponential reduction in state space.

Highlight what is the simplification you are bringing forth.

IV. Name of algorithm as section label

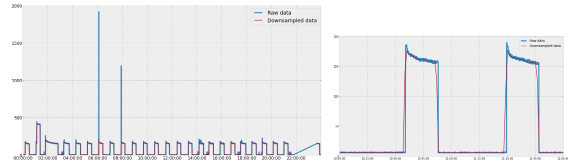
Think of appropriate name for your approach **Flowchart-TO go as image/flow diagram**

1) Data Preprocessing
   a) Aligning Mains and Appliance Level data
   b) Down sampling to 1 minute resolution
2) Load Assignment based on
   a) Appliance usage times
   b) Appliance Periodicity
   c) Appliance Power threshold
   d) Correlation amongst different appliance usage
3) Data Recalibration using the following
   a) Step changes occurring in Mains vs Appliances
   b) Isolating single appliance usage
4) Appliance states identification using clustering (Possibly talk about unbalanced data, but leave it for future work)
5) State space creation
6) Applying CO for different mains
7) Find energy distribution by appliance and assign weights (To be used in results)

Explanation of above steps

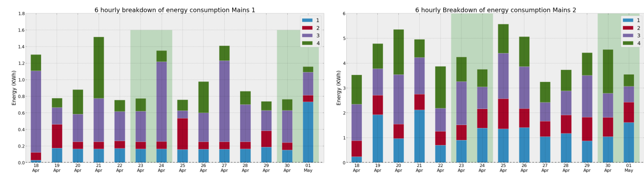
A. Data Pre processing

- Aligned data
- Down sampling and why it is needed
  - Sub metered data collected sometimes at 3s sometimes at 4 s
  - Getting rid of transients, also suggested by Hart
  - Reducing power fluctuations occurring due to voltage fluctuations. Put figure showing reduction in transients and signal smoothing



(a) Filtering startup transients (b) Filtering voltage fluctuations and oscillations

Fig. 1: Effect of downsampling appliance data



(a) Mains 1 (b) Mains 2

Fig. 2: 6 hourly energy usage breakdown Home 2

B. Load assignment

Draws inspiration from work by Parson et. al [6]. From prior knowledge we divide the loads into two different categories: Periodic such as refrigerator and non periodic such as Television.

V. Evaluation

A. About Dataset

We use REDD dataset [7] for validating our algorithms. This dataset contains power and voltage data for mains (2 phases) as well as appliances from 6 homes in Boston area collected in the summer of 2011. The data is made available as raw, high frequency (sampled at 15 KHz) and low frequency (Mains at 1 Hz, appliances at .3 Hz). Considering the practical implications of residential smart meter installation, we believe that low frequency data represents the most realistic scenario and thus we use this data for analysis. Figure 2 shows 6 hourly breakdown of energy consumption across the different mains in Home 2 .

B. Evaluation Metric

Commonly used metrics such as accuracy, sensitivity and specificity can be misleading when applied to NILALM. Figure **??** jkj

- Image of confusion matrix
- Show confusion matrix and argue that accuracy can be misleading, state 0 dominates in all appliances and will most probably always be predicted correctly
- Hart pointed towards residual power as an evaluation metric, REDD paper talks about % of energy recovered
- METRIC 1: Given an appliance having m states, we propose power weighted appliance accuracy, as follows $ApplianceAccuracy = \frac{\sum_{i=1}^{m} Accuracy(i) * Power(i)}{Power(i)}$

- METRIC 2: In this we also take into consideration the amount of time an appliance is in a particular state
- Overall Accuracy : Based on energy weights of different appliance, for instance .4* Fridge +.2*Light+... This is important since it shows the relative importance of larger loads
- Switch continuity

C. Empirical Analysis

We analyze data from Home 2 of the REDD dataset and believe that the same analysis can be easily repeated across multiple homes. We do this to compare our results against the Factorial Hidden Markov approach used by Kolter et al. [7] on the same home.

- Table on Cluster assignment
- Table on Results
- Table on on /off periods
- Table on switch continuity

VI. Conclusion

The conclusion goes here. We also provide mains load assignment of all 6 homes from REDD to further the research in this direction.

VII. Future Work

- Applying model on noisy datasets
- 2 D CO (when Real and Reactive Power are known)
- Factoring in Time of Day etc.
- Factoring in Appliance Correlation
- Factor in switch continuity, essentially leads to Factorial HMM
- Distributed NILM
- Adaptive Learning

Acknowledgment

The authors would like to thank TCS Research and Development for supporting the first author through PhD. fellowship. We would also like to thank NSF- DEITy for funding the project.

References

[1] S. Darby, "The effectiveness of feedback on energy consumption," *A Review for DEFRA of the Literature on Metering, Billing and direct Displays*, vol. 486, 2006.

[2] G. W. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.

[3] K. Carrie Armel, A. Gupta, G. Shrimali, and A. Albert, "Is disaggregation the holy grail of energy efficiency? the case of electricity," *Energy Policy*, 2012.

[4] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey," *Sensors*, vol. 12, no. 12, pp. 16 838–16 866, 2012.

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

[6] O. Parson, S. Ghosh, M. Weal, and A. Rogers, "Non-intrusive load monitoring using prior models of general appliance types," in *26th AAAI Conference on Artificial Intelligence*, 2012.

[7] J. Z. Kolter and M. J. Johnson, "Redd: A public data set for energy disaggregation research," in *proceedings of the SustKDD workshop on Data Mining Applications in Sustainability*, 2011, pp. 1–6.

[8] A. Filip, "Blued: A fully labeled public dataset for event-based non-intrusive load monitoring research."

[9] S. Barker, A. Mishra, D. Irwin, E. Cecchet, P. Shenoy, and J. Albrecht, "Smart*: An open data set and tools for enabling research in sustainable homes," in *The 1st KDD Workshop on Data Mining Applications in Sustainability (SustKDD)*, 2011.