A Knowledge based approach for disaggregation of low frequency consumption data

Krishnan Srinivasarengan, Goutam Y G, and M Girish Chandra

Abstract Smart electric meters that are capable of measuring and transmitting the power consumed by a household are being installed across the world. The low frequency average consumption data generated by these smart meters are now available widely for application development. A key enabler for the applications is the disaggregation of the consumption data into its constituent components, which can be useful to a wide spectrum of target audience from end user to utilities. Disaggregation of active power data sampled once in a few seconds has been studied extensively in the literature. However, there are only limited attempts to use the once in a few minutes consumption data for disaggregation. This paper discusses some preliminary results obtained using a knowledge based approach for disaggregating the low frequency average consumption data into consumption classes. The approach utilizes the spatial and temporal characteristics extracted from the data to estimate class consumption. The initial results from this approach are promising and have inspired to develop a larger probabilistic framework, which is also described in this paper. This proposed framework will be useful in generalizing and scaling the disaggregation algorithms across data of different sampling rates.

Krishnan Srinivasarengan
Innovation Labs, Tata Consultancy Services, Whitefield, Bangalore, India. e-mail: krishnan.srinivasarengan@tcs.com

Goutam Y G
Innovation Labs, Tata Consultancy Services, Whitefield, Bangalore, India. e-mail: goutam.yg@tcs.com

M Girish Chandra
Innovation Labs, Tata Consultancy Services, Whitefield, Bangalore, India. e-mail: m.gchandra@tcs.com

1 Introduction

The smart electric meter installation across the world is getting accelerated through a number of government regulation and standards. Smart meters can generate two different types of data about the household consumption. A low sampling rate (about once in 15 minutes) of consumption/energy data and the other, a higher sampling rate (once in a few seconds) of active power and consumption data. The former, referred to as Advanced Metering Infrastructure (AMI) data, is accessed directly by the utility companies and used for various purposes, significant of them is the implementation of time-differential tariff. This data is also made available to the end user and to third party service providers upon request from the end user. The higher sampling rate data interface, also known as Home Area Network (HAN), requires an extra hardware in the form of a gateway, which can interface with the smart meter and access the data at the rate of once in a few seconds (typical being 10 seconds). This data can then be either processed in-home or transmitted through the home Internet connection for a remote processing. Disaggregating into the constituent appliances or appliance categories is considered one of the important analytics that can be done on smart meter data [1]. Studies have indicated that energy consumption breakdown feedback can help in improving energy savings in households [3]. Apart from energy savings, disaggregation can be useful to schedule appliances to take advantage of differential tariff, analyze appliance health through consumption analysis over the years, etc.

In this paper, we explore how consumption categorization can be performed on a 15-min sampled data. The algorithm whose results are presented, provide an illustration of the usefulness of the approach. It uses the knowledge about various consumption categories in the form of their temporal and spatial characteristics to estimate their consumption. The impact of the variation of the appliance characteristics and their impact on the accuracy on the consumption estimation is discussed. An outline of how this algorithm would fit into the proposed probabilistic framework for consumption estimation is given. In the recent times, a number of start up companies in the load disaggregation domain provide consumption categorization based on AMI data [1]. However to the best of our knowledge, the literature does not discuss the details of an approach to implement this.

In illustrating these ideas and works, the paper is organized as follows: Sec 2 touches upon a few relevant works which have tackled the problem of disaggregation. The knowledge based approach algorithm and inferences are discussed in Sec 3, which also briefly outlines the knowledge based approach algorithms. The results are discussed in the Sec 4. The Sec 5 provides a summary and the future outlook of the proposed approach.

[1] see for instance, Bidgely: www.bidgely.com

2 Relevant Works

Non-intrusive load Monitoring has been of interest since Hart's seminal work [4]. A good majority of the works (see [10], [11] and references thereof) have focused on using a number of specialized sensors and/or multiple electrical parameters. These features will not be provided by the standard smart meters being installed across households. Works focusing on using only the active power meter data are emerging in the recent times (see for instance [8], [5], [9] and references in [10], [11]). However, these works require the instantaneous active power data to be available, and use the patterns in the data to identify appliances and compute their consumption.

AMI systems will be ubiquitous in future and the data from these systems would be available by default. However, the literature is limited for the problem of disaggregation on low frequency AMI data. In [7], a technique based on discriminative sparse coding is used to disaggregate hourly-consumption data. The overcomplete dictionary capable of providing sparse representation and discriminative ability for different classes of appliances is learnt through the training data. The requisite sparse *activation* matrices for the classes leading to the disaggregation results are obtained through coordinate descent based approach. Even with very coarse data, the accuracy results reported appear to be useful in terms of energy consumption (47%) and classification (55%). In [2], the authors develop an approach for partial disaggregation on hourly whole-house data using the inverse modelling technique. They compare the categories of consumption and some key consumption related parameters against the external temperatures and develop models to represent these relationships.

3 Algorithm and Framework

Disaggregation of 15-min sampled data into consumption categories is a challenging task due to a number of reasons. First, the data has a number of subjective elements that depends on demographic, climatic and cultural factors, among others. For instance, the highest power consuming appliance category in North America (HVAC) is different from that of Europe (Cold Storage/Laundry)[2]. Second, due to the large sampling interval, multiple appliances' ON/OFF characteristics are captured into a single average consumption measurement, making separation difficult. The consumption classes can hence vary between different geographies. A typical set of consumption classes shall be as follows:

- Phantom Loads
- Cold Storage
- Laundry

[2] This was learned during a project interaction

- Evening Consumption
- Unknown High Power rating appliances

which are used in this paper. Phantom loads correspond to those consumption that are ON through out the day without any appreciable consumption changes (e.g. fire alarm, router, TV standby). Cold storage comprises of refrigerators and freezers which are used throughout, but show a variation in consumption. Evening consumption corresponds to consumption from low power appliances that show up prominently during the evenings due to collective usage. While this doesn't account for all the low power consumption appliances, it enables the end-user to understand consumption. The Laundry class accounts for both washer and dryer and the unknown high power class combines all other high power rating appliances, including microwave, stove etc.

Before we describe the components of the knowledge based disaggregation approach, the proposed larger probabilistic framework is discussed first due to the ease of illustration. The framework was arrived at after an exhaustive set of experiments using various datasets and different types of appliance characteristics. The experiences are discussed in the Sec 4.

3.1 Proposed Framework

The framework aims to estimate the consumption of different classes and weigh them with confidence measure obtained by Maximum a Posteriori (MAP) estimate. The key blocks of the framework are shown in the Fig. 1. In a nutshell, the framework does the following: the features are extracted from the data with which an estimate of the consumption of each classes are obtained. The MAP estimate of each class consumption is then computed absorbing any apriori information available.

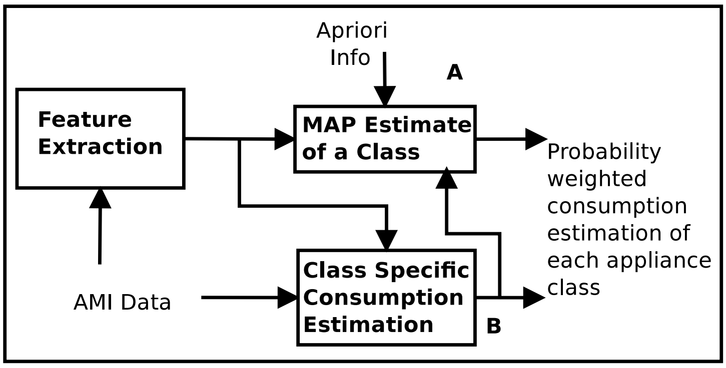

Fig. 1 Consumption Classification Framework

The Feature extraction involves processing the given data (96 samples per day). After a prolonged analysis of the data and their implications on the consumption estimation, the following categories of features were arrived at:

- Ordered Consumption extremas
- Rate of Change of Variation (Positive and negative) between local extremas
- Duration & Consumption between apparent ON & OFF
- Parameters of Probability Distribution (for learning)

A simplistic view of the MAP estimation process could be as follows. If Ω_i is the random variable representing the appliance class and $\boldsymbol{f}$ represent the feature set, then the MAP estimate is given by:

$$\Pr(\Omega_i|\boldsymbol{f}) = \frac{\Pr(\Omega_i, \boldsymbol{f})}{\Pr(\boldsymbol{f})} \tag{1}$$

$$\propto \Pr(\Omega_i, \boldsymbol{f}) \tag{2}$$

since the extracted featured are constants. The above representation assumes that the features have a representation which can uniquely identify classes. However, some of the features (e.g., the consumption between apparent ON & OFF) can only represent the combination of consumption classes active at that time. Hence the above representation can be rewritten as:

$$\Pr(\bar{\Omega}_i|\boldsymbol{f}) \propto \Pr(\bar{\Omega}_i, \boldsymbol{f}) \tag{3}$$

where, $\bar{\Omega}_i$ stands for the class combination i. Further,

$$\Pr(\bar{\Omega}_i, \boldsymbol{f}) = \Pr(\boldsymbol{f}|\bar{\Omega}_i) \times \Pr(\bar{\Omega}_i) \tag{4}$$

where, $\Pr(\boldsymbol{f}|\bar{\Omega}_i)$ is the likelihood of the consumption class combination. One approach to compute the likelihood would be in the form of:

$$\Pr(\boldsymbol{f}|\bar{\Omega}_i) = \frac{1}{\sum_j w_j^i |\boldsymbol{f}_j^i - \boldsymbol{f}_j|} \tag{5}$$

where, w_j^i is the weight associated with each feature, which can be variable depending upon the particular class combination i, $\boldsymbol{f}_j^i$ is the assumed feature value (either assumed or learned) for the jth feature for the class combination i and $\boldsymbol{f}_j$ is the extracted feature. The second term on the right hand side of (4) is a prior probability of the presence of the combination of consumption class. A typical prior probability could be given based on the Time of Day (T_D) and Day of the Week (D_w) correlation with the usage of each consumption class and hence that of their combination. For the sake of simplicity, if we assume that the timing correlation of each appliance class is independent of each other, we get,

$$\Pr(\bar{\Omega}) = \Pr(\Omega_1) \times \Pr(\Omega_2) \times ...\Pr(\Omega_n) \tag{6}$$

$$= \Phi_1(T_D, D_w) \times \Phi_2(T_D, D_w) \times ...\Phi_n(T_D, D_w) \tag{7}$$

The prior probability and the features corresponding to each class $\boldsymbol{f}_j^i$ are the components that would enable the framework to learn from the history of power data, contextual information based inputs (e.g. temperature based analysis as in [2]), any specific end-user feedback and so on. This MAP estimate of a class combination can be marginalized to obtain a confidence measure for individual consumption class. This could be enabled using a factor graph approach with Sum-Product or Max-Sum algorithm. Another key aspect of this MAP estimate (Block A) is the possible interaction with the consumption estimation (Block B) as indicated in the Fig. 1. This can lead to two: First, if a consumption class i is dependent on that of class j, then the MAP estimate algorithm corresponding to i can include the consumption estimate of j as a dependency factor. Second, in a scenario where the consumption measurement itself is the feature, the MAP estimate gives the confidence measure on the class combination. If the confidence measure is low, there is a scope to improve by iteratively considering different combination of the consumption classes to arrive at more accurate results.

The consumption estimation block is an essential component of the proposed framework. This block can be implemented through a variety of different techniques centered around pattern recognition, discrimination etc. In this paper, we have used a simplistic knowledge based approach to estimate the consumption of different classes. In case the accuracy is not good enough, there is always a scope, within the proposed broader framework, to improvise the results as discussed in the previous paragraph.

3.2 Knowledge driven Consumption Estimation Algorithm

The algorithm derives its steps from the exhaustive analysis of data to identify the pertinent features. The key components of the knowledge based algorithm can be summarized as follows:

- Correlating features with consumption classes
- Estimating the boundaries of consumption class usage
- Estimating the consumption of a consumption class

These components would wary depending upon the type of consumption class being estimated. For instance, the always ON classes like phantom loads and cold storage are assumed to be present throughout the day and hence wouldn't have a defined boundary. The features required by these classes are restricted to the consumption extremas at various times of the day. For instance, the simplest way to estimate the phantom load is by considering the minima of the whole-day data. However, this could be erroneous due to overlap with cold storage appliances like refrigerator/freezer. The estimation for cold storage consumption assumes that almost no other appliances apart from the phantom loads are likely to be used between 2AM to 5AM. The accuracy of these classes could however be improved when compared over different days and appropriate inferences about the underlying characteristics could be made.

The separation of the Always ON or slow changing low power appliances and that of high power rating appliances is established through an adaptive quantization procedure. This would track the changes in power consumptions such that only low amplitude changes are accommodated. The resulting data could help in identifying the boundaries of consumption of high power appliance usage. In case of high power appliances, the Laundry analysis assumes that the washer and dryer usage would be close to each other temporally. The analysis hence looks for a pattern with a large ON time correlated to appropriate time of day. The accuracy of the high power appliances could be improved by training them over several days and through inputs from the end-user.

These inferences are captured through an algorithm and is used to disaggregate the low frequency data into its constituent consumption classes. The flow of a sample algorithmic implementation used to generate the results in this paper shall be as follows:

1. Separate the time series data into low power and high power region using an adaptive quantization procedure
2. Phantom load as a function of minimum value of the time series
3. Cold Storage as the Phantom load subtracted value of consumption during night time
4. Laundry as the high power consumption region of sufficiently long duration with appropriate time of day correlation. Any temporal distance between washer and dryer usage are accommodated through an extended search in the adjacent high power regions.
5. The high power regions negating the laundry forms the Unknown high power appliance class

4 Results & Discussion

4.1 Data Generation Approach

The data used for illustrating the algorithm is generated by averaging a higher frequency active power data. This is due to the lack of ground truth with the available 15-minutes data as well as the availability of annotated 10-seconds sampled data through a number of open data sets (e.g. REDD [6]) and from a pilot project. This helps in evaluating the algorithm's performance in a quantitative manner. The data is generated in a bottom-up way, starting from creating a base load by fusing phantom loads and evening extra consumption. Variations in these loads are brought by varying the amplitudes of these loads. To this base data, the plug data available for appliances like washer, dryer, dishwasher, refrigerator, heater are added. The temporal positions of these appliances are varied during the addition to enable a truly representative power data of a household. This hybrid, 10-second sampled data is averaged over windows of 15 minutes length to generate the data necessary for analysis. This generated data represents the first difference data of the AMI data available from a smart meter. A typical 10-second sampled generated data is shown in Fig. 2 and its corresponding 15 minutes sampled data is shown in Fig. 3. It is to be noted that the 15-minutes sampled data represents the first difference of the typical cumulative consumption data of a smart meter.

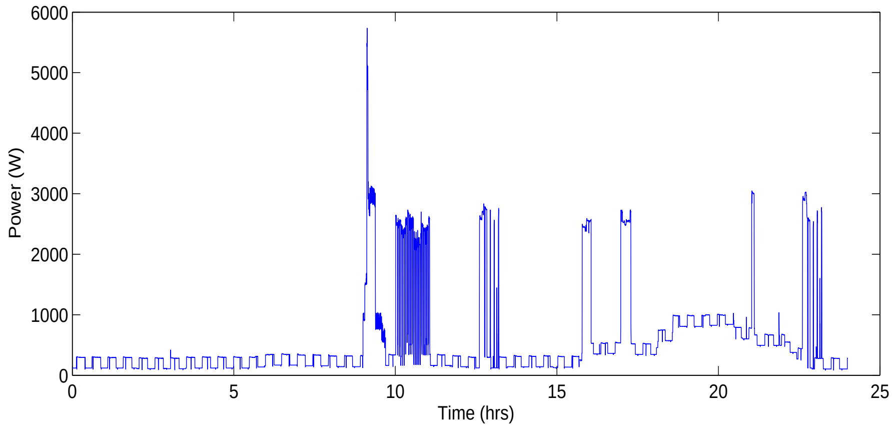

Fig. 2 An instance of generated 10-second sampled data

4.2 Results

The results obtained through the knowledge based approach are summarized in Table 1. The data for the three houses given in the table were generated by varying the characteristics of underlying appliances in the 10-seconds data and their time of day usage. The results were generated for several days for the three households and the statistics (minimum, maximum and average) of the accuracies of the consumption estimation are provided in the Table 1.

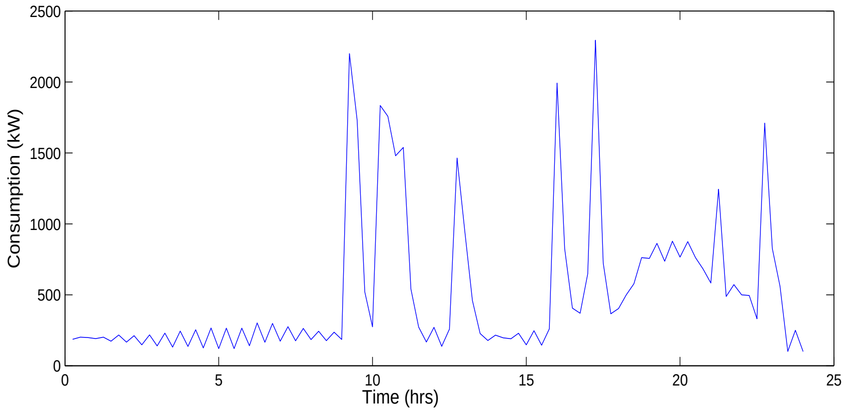

Fig. 3 An instance of generated 15-minutes sampled data

Table 1 Summary of Results

Consumption Class	House 1			House 2			House 3		
	min	**max**	**avg**	**min**	**max**	**avg**	**min**	**max**	**avg**
Phantom Load	0.33	0.97	0.79	0.51	0.91	0.79	0.5	0.92	0.68
Cold Storage	0.34	0.96	0.73	0.23	0.75	0.50	0.13	0.88	0.64
Evening Consumption	0.23	0.94	0.55	0.24	0.67	0.43	0.14	0.66	0.40
Laundry	0.81	0.99	0.86	0.63	0.98	0.87	0.39	0.99	0.86
Unknown High Power	0.73	0.94	0.85	0.56	0.90	0.77	0.52	0.98	0.81

4.3 Discussion

A number of assertions could be made based on the correlation between the accuracy and the characteristics of the underlying appliances/consumption classes. For instance, the accuracy of phantom loads and cold storage depends heavily on the usage of unknown high power appliances during the night. The evening consumption depends on the accuracy of the phantom loads and cold storage. The Laundry accuracy is sensitive to any significant difference in the time of day usage between washer and dryer. The Laundry accuracy and other high power rating appliances are interdependent and are affected by the temporal closeness of their operations. These inter-dependencies should be incorporated into the larger framework to improve the accuracy of the results.

5 Concluding Remarks

A probabilistic and knowledge based framework to disaggregate low frequency smart meter data into consumption classes has been discussed. The results obtained only by utilizing the knowledge based component of the framework looks promising and has opened up several important issues to be addressed in future. This includes the tight coupling between the blocks of the proposed framework. Further, the algorithm should be evaluated on large amount of diverse data with more complex user and appliance characteristics. Work is underway to develop the complete framework and also look at ways to improve the consumption estimation approach.

References

1. C. Armel, A. Gupta, G. Shrimali, and A. Albert. Is disaggregation the holy grail of energy efficiency? the case of electricity. Technical report, Precourt Energy Efficiency Center, Stanford, 05 2012.
2. B. J. Birt, G. R. Newsham, I. Beausoleil-Morrison, M. M. Armstrong, N. Saldanha, and I. H. Rowlands. Disaggregating categories of electrical energy end-use from whole-house hourly data. *Energy and Buildings*, 50:93–102, 2012.
3. J. Froehlich. Promoting energy efficient behaviors in the home through feedback: The role of human-computer interaction. In *Proc. HCIC Workshop*, volume 9, 2009.
4. G. Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870–1891, 1992.
5. H. Kim. *Unsupervised disaggregation of low frequency power measurements*. PhD thesis, University of Illinois, 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. Z. Kolter, S. Batra, and A. Ng. Energy disaggregation via discriminative sparse coding. In *Advances in Neural Information Processing Systems*, pages 1153–1161, 2010.
8. O. Parson, S. Ghosh, M. Weal, and A. Rogers. Nonintrusive load monitoring using prior models of general appliance types. In *26th AAAI Conference on Artificial Intelligence*, 2012.
9. K. Srinivasarengan, Y. Goutam, M. Chandra, and S. Kadhe. A framework for non intrusive load monitoring using bayesian inference. In *Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2013 Seventh International Conference on*, pages 427–432. IEEE, 2013.
10. M. Zeifman and K. Roth. Nonintrusive appliance load monitoring: Review and outlook. *Consumer Electronics, IEEE Transactions on*, 57(1):76–84, 2011.
11. A. Zoha, A. Gluhak, M. Imran, and S. Rajasegarar. Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey. *Sensors*, 12(12):16838–16866, 2012.