MS-E2177 Operaatiotutkimuksen projektityöseminaari (V)

Project plan:
Disaggregation of electricity data

Client: Fortum

Mihail Douhaniaris
Samuel Marisa
Matias Peljo
Johan Salmelin

21.02.2017

1 Background

Motivation

Energy consumption data and monitoring hardware becoming more and more accessible has opened possibilities to develop new services and solve existing problems. Smart metering already has provided customers a chance to have an hourly contract, and make savings by following hourly prices and moving their consumption to cheaper hours. Emerging application is utilizing the increasingly popular 1Hz smart metering data, which provides a strong potential for software based disaggregation of the metered consumption data into specific appliances in the household.

It is shown that people tend to save energy and change their behaviour when they are provided real data from their actions or concrete recommendations like, “consider decreasing the peak consumption and your energy costs by going into sauna at 9pm instead of 8pm”.

The increasing use and demand for renewable energy generation requires a change in consumer habits, as they add weather dependent fluctuation to the energy system. These fluctuations can be compensated by developing the transparency between the consumers and the energy company. Developing tools and methods to guide consumer behaviour can decrease the peak demand and save costs, but also decrease the need of peak power generation which is often expensive and covered by carbon dioxide intensive generation.

As energy generation and distribution provider Fortum can make use of the disaggregated consumption data in many different ways.

Some use case examples of new ventures/services for Fortum that use disaggregation are:

- Your refrigerator uses 40% more energy than modern one, a new one would pay for itself in 2 years
- Your lights have turned on, are you expecting someone in your home?
- Your TV is on, your kids are back from school
- Possibility to better guide demand elasticity by concrete, appliance specific recommendations for consumers.
- Possibility to identify devices potential for IoT inspired automated control of demand.

High frequency vs low frequency

Disaggregation methods can be classified into two categories based on the sampling frequency of the electrical data. Low sampling frequency methods refer to disaggregation algorithms that typically use power data with a sampling rate of 1 Hz or under whereas high frequency sampling methods use power data with sampling rates ranging from tens of Hz up to several MHz. Low frequency sampling methods disaggregate the electricity signal based on visually observable patterns, duration, time of use or state transition of the appliances while high sampling frequency methods disaggregate the electricity signals based on different harmonic orders present in the signal. Typically high frequency sampling methods allow for higher disaggregation precision and are able to identify several distinct appliances, in the range of 20-40 separate appliances, whereas low frequency sampling methods are able to identify 5-10 separate appliances. However, the main drawback of high frequency disaggregation methods is the high cost and installation effort of the electrical sensors providing the electrical data [1]. In our project, the data provided by Fortum has a sampling frequency of 1 Hz and, as such, low frequency disaggregation methods are appropriate.

Real power vs reactive power

In an alternating current circuit, the phases of voltage and current can behave differently depending on the load. If there is only resistive load, the voltage and current are in same phase and the power present is real power. However, when capacitance and inductance are present then the voltage and current get out of phase. Then the product of voltage and current, i.e. the power, can change its sign during time. For example, when the load consists of elements that have only inductance and capacitance, the phase difference between current and voltage gets 90 degrees. This leads to situation, where the sign of power is positive half of the time and negative the other half of time. There is no net energy flow, and the power present is called reactive power. Some of the disaggregation methods can utilise the reactive power measurements as well, but it requires high sampling frequency and special measuring devices. Especially in the case of low sampling frequency, only the real power is observed. Hence, there are plenty of methods that are based only on real power measurements.

Supervised vs unsupervised vs semi-supervised learning

Another classification for disaggregation methods is based on the type of data used in the disaggregation algorithm. All disaggregation methods take an aggregated electricity signal as input with the aim of decomposing, or *disaggregating*, said signal into individual appliances' signals as accurately as possible. However, the way the model parameters are obtained differ between disaggregation methods. Some of the disaggregation methods require already disaggregated training data to determine the model parameters and to be able to disaggregate new aggregated signals. This disaggregated data can be obtained by installing sensors and monitoring each appliance separately. This kind of learning is called supervised learning. Supervised learning methods require the training data for each household separately in order to work well. The drawback is that this kind of training data can be expensive or difficult to obtain [2].

Unsupervised learning does not require this kind of appliance-level training data. In fact, unsupervised methods do not require any prior knowledge of appliances (except the number of appliances in some cases). Unsupervised learning determines the model parameters from aggregated signals and then the algorithm can disaggregate the signal. However, methods that utilise unsupervised learning are not able to assign labels to each appliance, since there is no prior information about the names of appliances. Therefore, some labelling process is required to join the disaggregated data and the names of the appliances together. The drawback of unsupervised methods is that they cannot distinguish between appliances well when there are many appliances in the household [2].

An alternative approach to purely supervised or unsupervised learning methods involves manipulating the aggregated smart-meter data into retrieving single-event appliance-level information. This is achieved by turning on each appliance sequentially while keeping others turned off. Such an approach is often called semi-supervised, in the sense that the installation and use of device-specific sensors is avoided [3].

2 Objectives

The project is aiming to both provide a study of energy disaggregation for Fortum, and an valuable learning experience for the project team of students with multidisciplinary backgrounds. During the project students will use and develop their engineering and scientific information acquisition skills, while providing a valuable deliverable for Fortum.

The main objective is firstly to provide Fortum further knowledge of the energy disaggregation. Throughout the project a major workforce is focused first on the literature review and later more thoroughly on the most relevant findings. The aim is to identify previous studies and implemented methods that are useful during the project and in the future development by Fortum.

Secondly the objective is to develop a proof of concept energy disaggregation model. The purpose of the model, model data and performance will be scoped together with the team members and Fortum in the beginning of the project based on the literature review and then modified further iteratively when needed.

3 Tasks

Introduction to assignment topic

In the beginning of the project the most important task is to fully understand the assignment given by Fortum through discussion and further study of the field of energy disaggregation. During this phase we will do a brief high-level literature review on the topic of disaggregation, including learning about use cases of power disaggregation, different approaches to power disaggregation and finding what open data sets exist.

Exploratory data analysis

In this phase we do simple calculations and plots of the data provided by Fortum in order to get acquainted with the data, find possible errors and understand how the three different power phases relate to different appliances.

Literature review

We begin reviewing different methods and algorithms used in power disaggregation in order to be able to compare different approaches and find the most suitable for our use. This literature review and model comparison is part of the assignment given by Fortum and, as such, after this phase we will be able to provide Fortum recommendations on different approaches for data collection as well as model validation and performance expectations, both for the scope of this project and for possible future endeavours.

Project plan development

We work on writing the project plan and on the presentation.

Model comparison

We analyze and compare the most promising methods identified during the literature review in order to be able to choose the most suitable ones for our project.

Model specification

Following the model comparison in the previous part, we specify the model we will begin implementing. The model will be formulated in a way that will allow us to work on the implementation iteratively, meaning we will begin with a simpler prototype and work our way from there towards more complex implementations.

Model implementation: Prototype

We begin working on the first version of our model. This version of the implementation will have very basic functionality, such as simple disaggregation with few appliances or low "true positive" disaggregations. The idea is that the prototype will serve as a draft for the final model and will help notice possible error or pitfalls in the model specification.

Model implementation: Prototype evaluation and improvement ideation

Following the iterative nature of our implementation plan, in this phase we will evaluate the finished prototype and what additional features and improvements should be implemented in the final version.

Model implementation: Model improvements

During this phase we will work on implementing the features and correcting the errors of our model identified in the previous phase. This phase will be planned so that different group members can work on different parts of the model.

Interim report development
We work on writing the interim report and preparing the presentation.

Model implementation: Model refinement
In this phase we will focus our efforts in correcting any issues that might rise in the previous phase as well as refining and optimizing our code before finalizing the implemented model.

Verification and validation of model
In this phase we will assess and verify the performance of the implemented model. This includes disaggregation performance and the number of identified appliances. Additionally, our implementation will be validated together with Fortum, in order to ensure that the output of our project meets the needs and requirements of Fortum. The details of this phase will be discussed with the client, as it will need some additional data collection.

Disaggregated data business case ideation
We generate new ideas about possible new business cases and ventures using Fortum's disaggregated data.

Final report development
We work on writing the final report and preparing the presentation.

4 Schedule
The critical dates of the project specified by the course timetable are listed on Table 1.

Table 1: *Critical dates of the project.*

Date	Description
2017-02-22	Project plan deadline
2017-02-24	Project plan presentation
2017-04-18	Interim report deadline
2017-04-21	Interim report presentation
2017-05-17	Final report deadline
2017-05-19	Final report presentation & opponent work

Additionally, the tasks specified in the project plan are scheduled on a weekly resolution and can be seen in Figure 1.

5 Resources

Our project team consists of four students majoring in Operations Research, Energy Technology and Industrial Engineering and Management, which makes our group relatively heterogenous. Our team members have additional experience, e.g., in Computer Science, Stochastics and Machine Learning. Our aim is to utilise the specialties of each group member while keeping the workload evenly distributed. Project manager of our team is Mihail.

Our contacts inside Fortum includes Juhani Rantaniemi and Teppo Luukkonen. Frequent discussions with our Fortum contacts ensure that our project meets the interests of Fortum. Fortum has provided us with large amounts of aggregated household data and some device signatures from a single household.

The course personnel consisting of Professor Ahti Salo and course assistant Joonas Laihanen give feedback of our progress and provide us with suggestions and contacts relating to our tasks. The opponent team (Kesko) provides valuable feedback throughout the course, as well.

As students, our project team has academic access to scientific journals as well as licence to use different computing softwares. Additionally, resources such as open source software or open power disaggregation datasets may play an important role in our project.

6 Risks

The risks we have identified for this project are listed on Table 2. The occurrence probability of each risk is evaluated as low, medium or high. Additionally, we describe the impact of each risk on the project, as well as the taken precautions to minimize the risk.

Table 2: *The risks of the project.*

Risk	**Probability**	**Impact**	**Precautions**
Team member inactivity	Low	Imbalanced workloads between members	Transparency in scheduling and working together
Team member dropout	Low	Workload grows too large for the remaining members	Transparency in scheduling
Project scope proves to be too wide	Low	Workload grows too large and implementation may remain incomplete	Narrowing the scope iteratively along the course together with Fortum
Model performance not satisfactory	Medium	Low value creation for Fortum	Thorough model comparison and scoping together with Fortum
Model proves to be too complex for the scope of the course	High	Incomplete implementation and no value creation for Fortum	Start with simpler models and build up iteratively
Confidential information spread	Low	Legal consequences for Fortum and students	Data is kept only at specified storage places

7 References

[1] Energy Disaggregation (presentation), Armel C., Precourt Energy Efficiency Center, Stanford, December 2011

[2] An unsupervised training method for non-intrusive appliance load monitoring, Parson O., Ghosh S., Weal M. and Rogers A., Artificial Intelligence, Volume 217, December 2014

[3] A Hybrid Signature-based Iterative Disaggregation algorithm for Non-Intrusive Load Monitoring, Cominola A., Giuliani M., Piga D., Castelletti A. and Rizzoli A. E., Applied Energy, Volume 185, Part 1, January 2017

[4] Neural NILM: Deep Neural Networks Applied to Energy Disaggregation, Kelly, J. and Knottenbelt, W., Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments, 55-64, September 2015

[5] The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes, Kelly J. and Knottenbelt W., Scientific Data 2, Article number: 150007, March 2015