Imperial College London
Department of Computing

Evaluating an Assumption Based Framework for the purposes of Disaggregating Smart Meter Readings and User Profiling

by

Gregory Loyse

Submitted in partial fulfilment of the requirements for the
MSc Degree in Computer Science of Imperial College London
September 2012

Abstract

Text of the Abstract.

Acknowledgements

I would like to express my deepest thanks to:

- Francesca Toni, my project supervisor. I couldn't have had a more enthusiastic, proactive and fun supervisor.

- Fariba Sadri, my personal tutor. Fariba has been understanding and supportive. I am very grateful.

- My father Michel Loyse and my sister Angela Loyse. Without their support I wouldn't have been able to study for this masters.

Dedication

This project is dedicated to my sister Angela Loyse.

‘Do not assume anything Obi-Wan. Clear, your mind must be, if you are to discover the real villains behind this plot.’

Yoda

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Figure ??

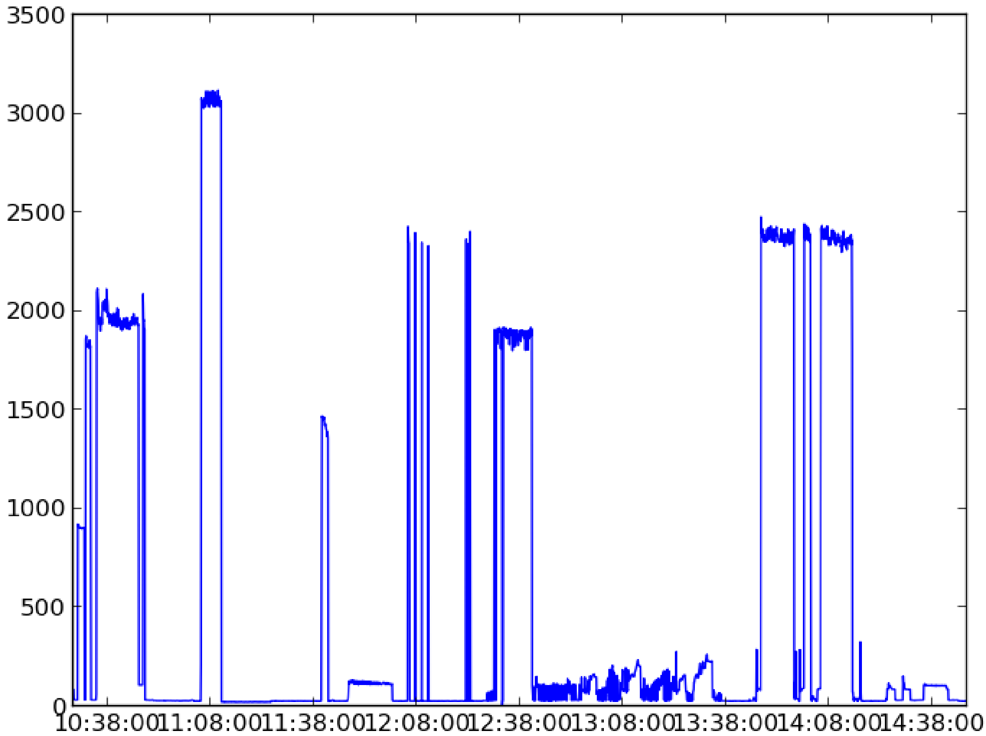

Figure 1.1: mylabel

1.1 Motivation

It is becoming increasingly important to better manage electricity consumption. Domestic electricity prices in the UK increased by 35 (in real terms) between 2003 and 2011 [15]. Energy price rises alone are projected to inflate the UK consumer prices index by 1.5 % in Q4 2011 [17], making energy one of the most dominant upward forces acting on UK consumer prices. This upward price trend is likely to continue as global energy demand, especially from non-OECD countries, contin- ues to grow. In 2000, China used half as much energy as the USA. In 2009, China overtook the USA to become the worlds largest energy user [11]. The International Energy Agency projects that Chinese consumption will increase by 75% between 2008 and 2035, and that Indian energy consumption will more than double over the same period (although India will still consume less energy in 2035 than either the USA or China) [11]. The UK is especially exposed to energy price rises. The UK makes extensive use of natural gas for heat and power generation. Between 1995 and 2004 the UK was not just self-sufficient for gas but it was also a net exporter of natural gas pumped from the North Sea [20]. Gas production from the North Sea peaked and began its terminal decline in 2000 hence the UK has been a net importer since 2004 [20], requiring us to buy increasing amounts of gas on the volatile global market. Gas production from the North Sea halved between 2000 and 2010 Text of the Background.

The public are broadly unaware of the finer details of their energy consumption. For example only 17%

[1] of people identified the most power hungry appliance in their home (the washing machine) . Pilot projects have demonstrated that mere knowledge of power consumption helps to bring consumption down. The DEHEMS European Union funded project showed an average reduction of 8%

The UK government has made it an explicit goal to increase the understanding of energy usage by users. It has set itself the objective that for every home and business in Great Britain to have a smart energy meters adapted to their type of usage.

The aim is to begin a national roll out beginning in 2014 and to be completed by 2019. This rollout is to be achieved with the cooperation of the energy industry. It will replace over 53 million gas and electricity meters.

The present project inserts itself into this push by providing a reasoning mechanism to derive

[1] Europa's Article "Digital Agenda: EU funded project helps citizens compare and reduce their energy consumption via TV, PC and social networks applications"

information about energy consumption.

Currently however, there are two main aims of deriving information. One is by using a smart meter as aforementioned. The other is by employing an Energy Monitoring Unit. A smart meter is the more accurate of the two options, the monitoring unit only delivering estimated consumption. However a monitoring unit is easy to install and readily available. In this project we will be using a monitoring unit.

The aim of this project is to attempt to profile users based solely on energy readings from an energy monitoring unit.

As the reasoning will be based on a centralised consumption it will remain valid regardless of whether the input is from a smart meter or an energy monitoring unit. A decoupled architecture will ensure that the system can work on both with minor if any modifications.

1.2 Smart Meter hardware and Data source

A large part of this work was spent in data capture and storage.

The constraints. requirements

The energy monitor chosen is marketed as Current Costs EnviR. This unit was chosen due to its ease of installation and the clear xml output that it can generate.

You can attach the device to a computer via usb and thus collect discrete xml messages periodically, usually every 6 seconds.

1.3 Objectives

1.4 Report outline

Contributions here.

Chapter 2

Background Theory

2.1 previous work in disaggregation

disaggregation has a long history.

Previous work on disaggregation algorithms tend to focus on advance hardware. see jacks' report.

2.2 Abduction

Abductive reasoning will be used to derive results from the system. At this stage the hope is that Jiefei Mas re-implementation of Asystem (a standard abduction reasoning system) will provide off the shelf abductive reasoning for this project. [*]

Abduction is a logical inference mechanism that flows from description to hypothesis.

It is non-monotonic. This means that later descriptions could affect the truth status or hypothetical conclusions that previously held true.

It is intuitive. It is elegant in the sense that it deals with the unknown in terms of probability.

$\langle P, A, IC \rangle$

where: P is a normal logic program. It is a model, a description, of the problem domain. It is a set of rule clauses that match observations as a head to possible explanations as the tail. Thus the

[*] Jiefei Ma's homepage

set of tails represent all the possible hypothesises of the model. And, the set of heads represent all the observations that the program understands and can take into account.

IC, the set of integrity constraints are simple first order clauses. These will filter out the set of candidate explanations to provide a solution.

G are the observations that need explaining they are ground predicates. In our program they will be derived from signal processing. They must belong to the subset of heads.[??]

E are the set of The program is solved by a combination of backwards reasoning and integrity checks.

Once an explanation has been chosen, then this becomes part of the theory which can be used to draw new conclusions. We don't use this aspect in our project. But see future work.

[eg program]

It is efficient in that it deals very efficiently with combinatorial explosion. A good adductive theory will in effect result in a strongly guided search that only considers relevant answers.

Furthermore non-montonicity can be exploited. Unfortunately this aspect couldn't be explored to the full in this project. More on this in future work section.

Abduction has been used in Medical diagnosis... to the best of my knowledge it hasn't been used in the context of signal processing and disaggregation nor has it been used for profiling users.

Thus this report has an academic component that will evaluate abduction in these contexts.

2.3 convolution

As will be discussed later an exploration of how to apply calculus, and integration in particular, to signal processing inspired the find of this.

convolution

There are other aspects that could be looked into.

auto correlation cross correlation

Chapter 3

Current Cost setup and raw data

3.1 Setup and Raw data

The format of the data received is the following:

Raw data was captured and stored in csv files. xml was considered but the overhead [give calculations] of $timestamp, digit/n$ compared to $< capture >< time > timestamp < /time >< watt > 1234 < /watt >< /capture >$ would have resulted in much larger files. simple csv was deemed sufficient.

It was only as the project evolved that the above became apparent. So there are 3 types of raw data that were captured. Each corresponding to a change of mind as to what was optimal. Therefore a few scripts were necessary to transform the data into a homogenous set.

The final capture script corresponds to how the energy should be.

I decided to take the time from the receiving computer itself as the time of the actual device would need to be configured by the user himself and therefore is more likely to be wrong.

3.1.1 Capturing and Parsing Data

The general output of the Energy Monitor and thus the input to our system has this form:

serial port data

$\langle msg \rangle$

$\langle src \rangle CC128 - v1.29 \langle /src \rangle$

$\langle dsb\rangle 00097\langle /dsb\rangle$

$\langle time\rangle 08:03:42\langle /time\rangle$

$\langle tmpr\rangle 18.5\langle /tmpr\rangle$

$\langle sensor\rangle 0\langle /sensor\rangle$

$\langle id\rangle 01657\langle /id\rangle$

$\langle type\rangle 1\langle /type\rangle$

$\langle ch1\rangle\langle watts\rangle 00184\langle /watts\rangle\langle /ch1\rangle$

$\langle /msg\rangle$

Given a certain installation the data never changes. Much of the information is redundant. For my program the only variables that are interesting are the watt and time readings. Potentially reasoning based on temperature could be useful, but this information could be obtained from more reliable database sources later. Furthermore the time provided by the smart meter is dependent on the user correctly configuring the device. I felt that a more accurate approach would be for the receiving computer to provide the timestamp.

A parser written in python will translate this into the ultimate chosen knowledge representation.

I tried a variety of implementations based on parsing the xml. This included creating a tree or using a event based (sax) approach. In the end using regular expressions seemed to offer the most straightforward concise and efficient solution. The data never changes.

3.1.2 The data

Over the course of slightly more than two weeks I accumulated a sufficient volume of real time data. This data is included in the project files.

I thought I was being clever in stripping out the extra information, but it turns out that a full datestamp makes the code clearer and I should have left the full datestamp.

As the program to register the readings evolved with my understanding of data capture the raw source data changed slightly. There are quite a few scripts that I had to code to homogenise the data in order to prepare it for future work.

With this data the project was ready to begin.

3.2 appliances

we won't be dealing with light switching.

all the light switches in my home are energy efficient. Old bulb types can be 60Watts or more.

Although some chandelier light might behave like appliances.

[graph of appliances]

Chapter 4

Signal Processing and Disaggregation

Originally there was no intention to enter the realm of disaggregation. However reliable disaggregation techniques just don't exist currently. If this project was to be more than a theoretical exercise and furthermore be testable it would need to deal with real data. Thankfully this was quite a pleasant area to work in. It involved a lot of research and thinking which meant refreshing my maths that haven't been visited since high school.

Unexpectedly a role for adductive reasoning was also found here.

I didn't manage to find any examples of Logic based programming being derived from signal processing. So this also proved to be an exciting and potentially innovative exercise in translating between discrete signals and a logic based framework.

4.1 Preparing the signal

Ideally we would like to be able to identify sharply defined state changes in the energy consumption. This would mean that electricity consumption would change instantaneously. [insert graph here]

This is of course not possible. The current cost meter already gives us a filter of values of sorts. The discrete signals received, usually one every 6 seconds, so effectively a sampling rate of roughly 0.6Hz. This sampling rate is totally inadequate to trace the finer detail of a devices state change.

For example [graph]

Thus the resolution at which a devices' signature can be identified will be affected. Our approach must take that into account. The changes will be more sudden and there is no telling what has happened in-between the two measurements.

This brings us to our first assumption. We will assume that this frequency rate is enough to supply a unique enough signature for each device.

4.1.1 Using a histogram to guide the analysis

awef awef awef

4.1.2 slicing away shadows consumption

Bar switching off the mains, (and therefore our measuring device) it is difficult to bring a households' electricity consumption to zero. This is due to devices being hard-wired and appliances on standby.

Experimentally I found that even when having unplugged everything I could my flat's energy consumption never descended below 28Watts. Furthermore at this level there was little or no variation.

This brings us to our second assumption. We will assume that every household will have a shadow consumption.

4.1.3 smoothing algorithms

A smoother signal ought to lead to clearer analysis.

A variety of algorithms were trialled for the purpose of smoothing the input.

4.2 Identifying appliances

4.2.1 Prototype

To get a feel for the data.

Home grown algorithm here.

4.2.2 Calculus 1 - differentiation

It is easy to find a maximum of a signal. What is more difficult is identifying local maxima or even plateaus.

A variety of techniques were considered. Hill climbing...

The idea of using basic differentiation was an idea I spent considerable time exploring.

[graph]

[pseudo code]

4.2.3 Calculus 2 - integration

It was whilst revising my calculus and in particular integration that it struck me that integration might offer an answer.

The area below a signal is a very accurate defining characteristic of that signal.

[pseudo code] [graph]

s

4.2.4 messiness

A measure of how messy the signal is could lead the way into identifying appliances

4.2.5 appliance identification characteristics

Finally there are certain characteristics that make an appliance identifiable.

These also introduce a new set of assumptions for each appliance

4.3 Wrapping all together and Abduction

4.3.1 combining above to for full signature characterics identification

The final algorithm will combine all the above notions into one set. For each event a set of characteristics will be recorded.

This will be translated into a table which identifies predicates.

Thus the data is readied for adductive based analysis.

4.3.2 Abduction

Using abduction to identify each appliance.

Chapter 5

Abduction

5.1 implementation considerations

I did research other adductive systems...

At my supervisors' suggestion I will be using Jeifei's system. Indeed my research of current systems suggest that it is the best in that it is currently maintained.

Jeifei also cordially corresponded with me and accepted to meet with me to discuss my project.

5.2 Abduction and Disaggregation

5.2.1 appliance eg

This is great stuff... blah blah blah [figure] [abductive program]

results.

5.3 Abduction and User Profiling

5.3.1 reasoning about users

In order to establish a use profile all we need is to create timelines of profiles. Each profile will be distinguishable if there is at least one differentiating defining characteristic.

[picture of user profile]

This is also great stuff...

5.3.2 the program

[insert program here]

5.3.3 results

with artificial data.

with real data.

Chapter 6

Conclusion

6.1 Summary of Thesis Achievements

Summary.

6.2 Applications

Applications.

6.3 Future Work

Future Work.

Bibliography