A Model for Energy-Saving in an IoT Smarthome accounting for End-User Convenience

Alistair Francis Bowman Grevis-James

November 2019

Abstract

The preface pretty much says it all.
Second paragraph of abstract starts here.

Contents

1 Introduction

1.1 Background

1.1.1 Humankind, Technology & Development

Since the inception of the first home computers in the late 1970's (Press, 1993), modern society has become utterly dependent on and indeed, inexorably bound to digital technology. The rapid and widespread adoption of computational technology has led to the fastest rate of societal and economic development our species has ever experienced. One of the most salient manifestations of technical progress has been the widespread availability and adoption of Information and Communications Technology (ICT), including the rise of the global network of networks known as the Internet.

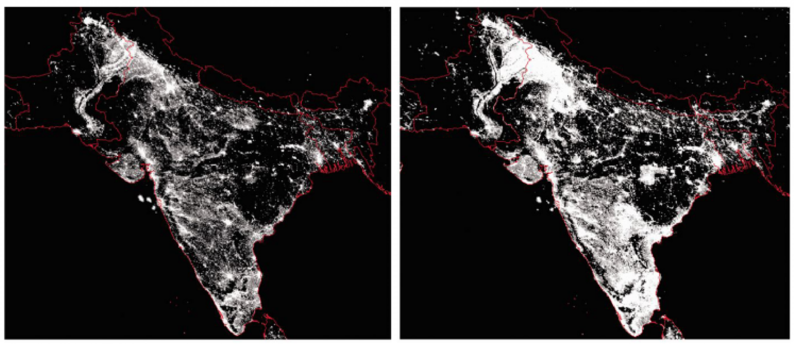

Figure 1: Satellite images of South Asia by night. Left (South Asia in 1994) Right (South Asia in 2010). Images are taken from Maxim Pinkovskiy and Xavier Sala-i-Martin (2016) - Lights, Camera ... Income! Illuminating the National Accounts-Household Surveys Debate. The Quarterly Journal of Economics.

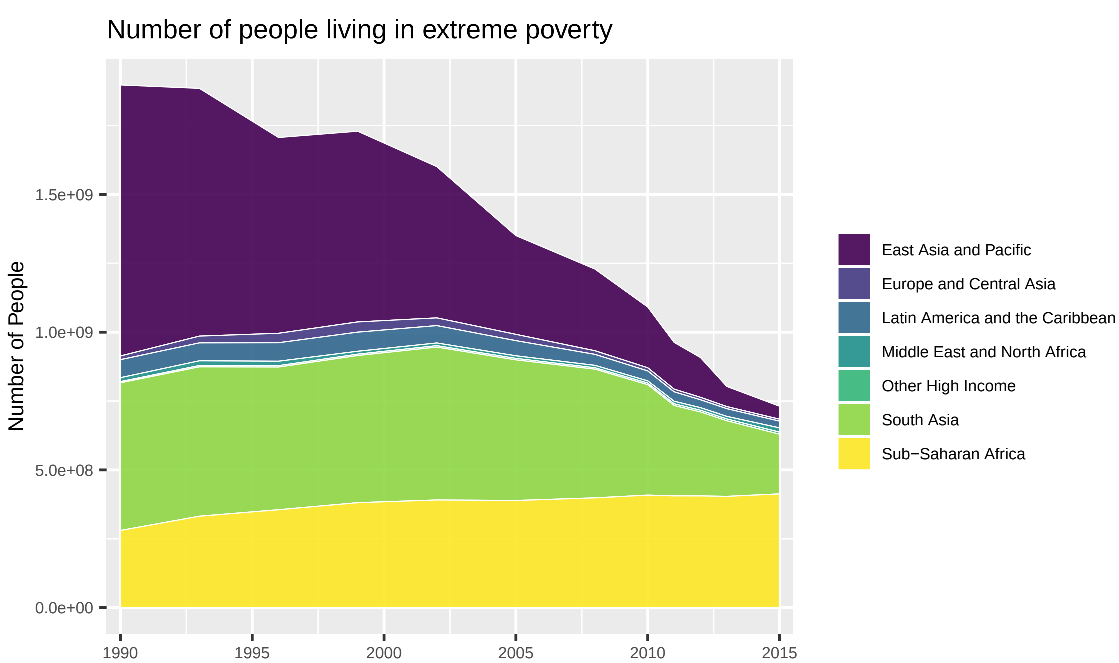

Figure 2: The number of people living in extreme poverty between the years of 1990 till 2015, segmented by region. Source: The World Bank

http://iresearch.worldbank.org/PovcalNet/

1.1.2 The Rise and Rise of the World Wide Web

According to the International Telecommunication Union (ITU) 2015 ICTs figures, Internet penetration has grown from just over 400 millions users (6 per cent of global population) in 2000 to 3.2 billion users in 2015 (43 per cent of global population), which includes around 2 billion users from developing countries (Dutta et al., 2015). ICTs bring a broad range of benefits and are recognised as a key to eradicating poverty and unemployment. They enable and facilitate the building a people-centred, inclusive and development-oriented Information Society, where everyone can create, access, utilize and share information and knowledge. This enables individuals, communities and peoples to achieve their full potential in promoting their sustainable development and improving their quality of life ("Information and communication technologies (ICTs) | Poverty Eradication," n.d.). In addition to a rapidly growing internet user-base both in the developing and developed world, the nature of internet usage has fundamentally changed. Once the purview of academics, engineers and computer scientists sending tiny packets of information back and forth, there are now some 2.5 quintillion bytes of data created each day by all manner of users, industry and sensors to name a few ("Data Never Sleeps 5.0 | Domo," n.d.).

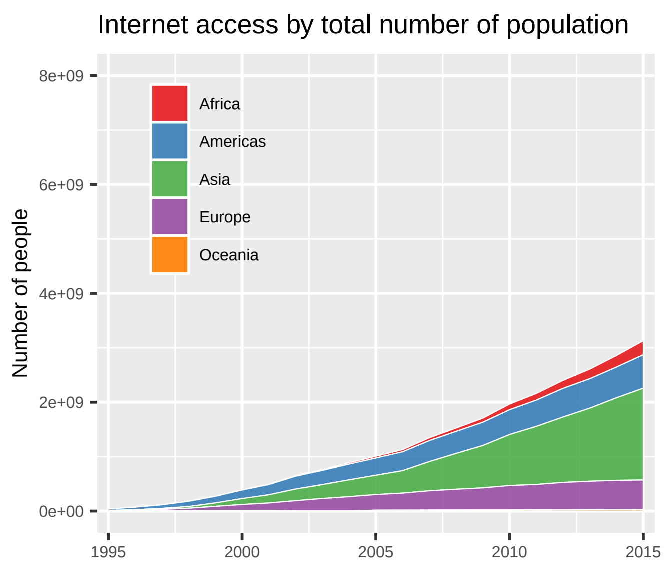

Figure 3: ADD TEXT, Source: The World Bank, World Development Indicators

http://data.worldbank.org/data-catalog/world-development-indicators/

1.1.3 Forging a New Technological Paradigm

Such rapid and widespread internet adoption has created a seemingly insatiable demand for exponentially greater computational power and digital storage capacity. This has led to a new and utterly ubiquitous technological paradigm; Cloud Computing. Cloud Computing is succinctly defined as; The practice of using a network of remote servers hosted on the Internet to store, manage, and process data, rather than a local server or a personal computer (https://www.dictionary.com/browse/cloud-computing).

As with the initial rise and widespread implementation of the internet, Cloud Computing itself acts as a facilitator for new technologies. One such example being the Internet of Things (IoT) paradigm. The Internet of Things can be surmised as the extension of the Internet and the Web into the physical realm, by means of the widespread deployment of spatially distributed devices with embedded identification, sensing and/or actuation capabilities (Daniele Miorandi, 2012). The Internet of Things (IoT) paradigm enables physical devices to connect and exchange information, and also allows objects to be sensed or controlled remotely through the internet (Bing Huang, 2018). IoT devices allow objects to be sensed or controlled remotely through the Internet (Luigi Atzori, 2010). IoT thus represents a convergence of real-world objects and digital objects into a unified cyber-physical system.

1.1.4 The Bigger Picture

Considering the bigger picture of societal benefit, as seen in figure 2, from 1990 through 2015 the number of people living in extreme poverty has dropped by more than half. As evidenced by Figure 3, over the same time period, the percentage of people with access to the internet has move from around 1 per cent to an average of around 50 per cent (this trend is now moving exponentially as a function of time). And, figure 4 shows that over the last 100 years, the human population has experienced unprecedented growth from 1 billion individuals to in excess of 7 billion. Mankind have thus simultaneously increased our population, increased of technological development and decreased poverty (to name but a few, 'key performance indicators').

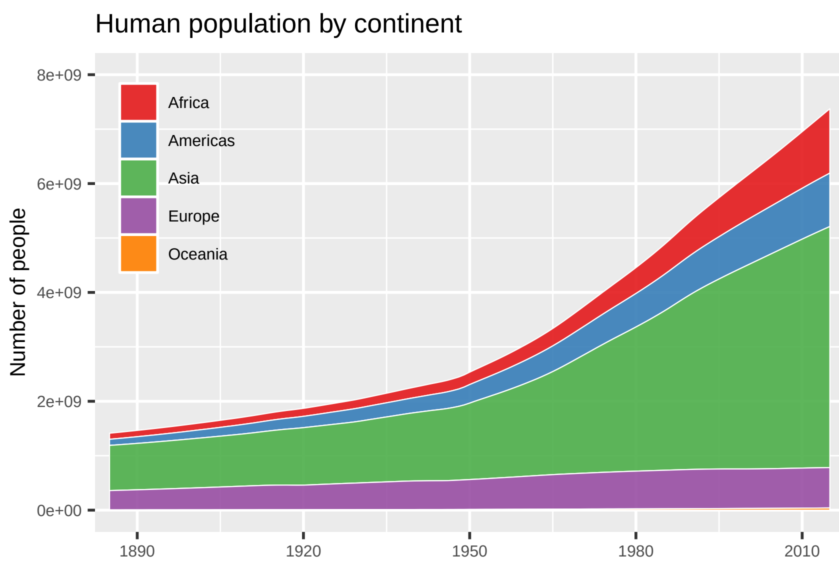

Figure 4: ADD TEXT, Source: The World Bank, World Development Indicators

The success of modern human society is not without consequence. All of the benefits our society has enjoyed from the development, production and deployment of technology, has required vast amounts of energy. This energy has, since the industrial revolution, primarily been derived from the burning of fossil fuels, as illustrated by Figure 5.

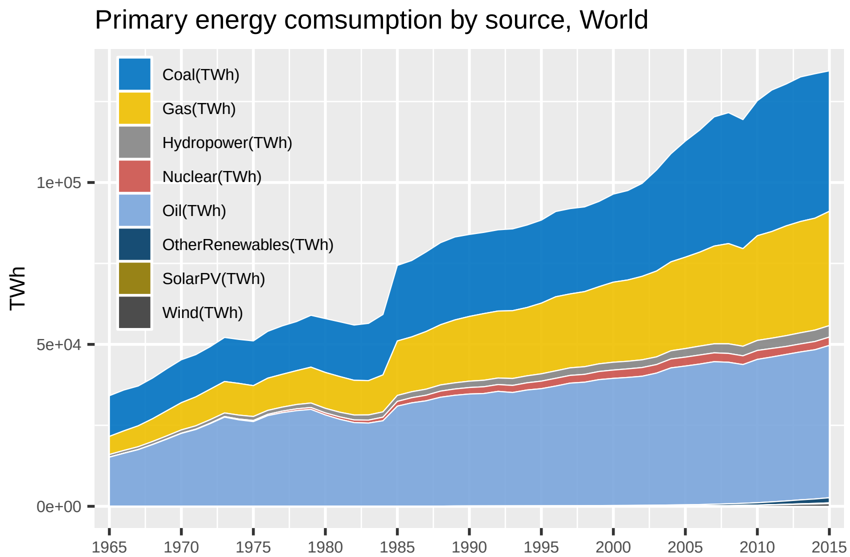

Figure 5: ADD TEXT, Source: Our World in Data

The International Panel on Climate Change (IPCC) finds that Human activities are estimated to have caused approximately 1.0°C of global warming above pre-industrial levels, with a likely range of 0.8°C to 1.2°C. Global warming is likely to reach 1.5°C between 2030 and 2052 if it continues to increase at the current rate (Intergovernmental Panel on Climate Change, 2018).

Climate change poses an existential threat to modern human civilisation, with warming of between 1.5°C and 2°C predicted to cause increases in mean temperature in most land and ocean regions, hot extremes in most inhabited regions, heavy precipitation changes including drought and precipitation deficits in some regions. Additionally, increases in ocean temperature as well as associated increases in ocean acidity and decreases in ocean oxygen levels are projected to reduce risks to marine biodiversity, fisheries, and ecosystems, and their functions and services to humans. Taken together, these effects will lead to risks of the health, livelihoods, food security, water supply, human security, and economic growth of mankind (Intergovernmental Panel on Climate Change, 2018 - OTHER REF?).

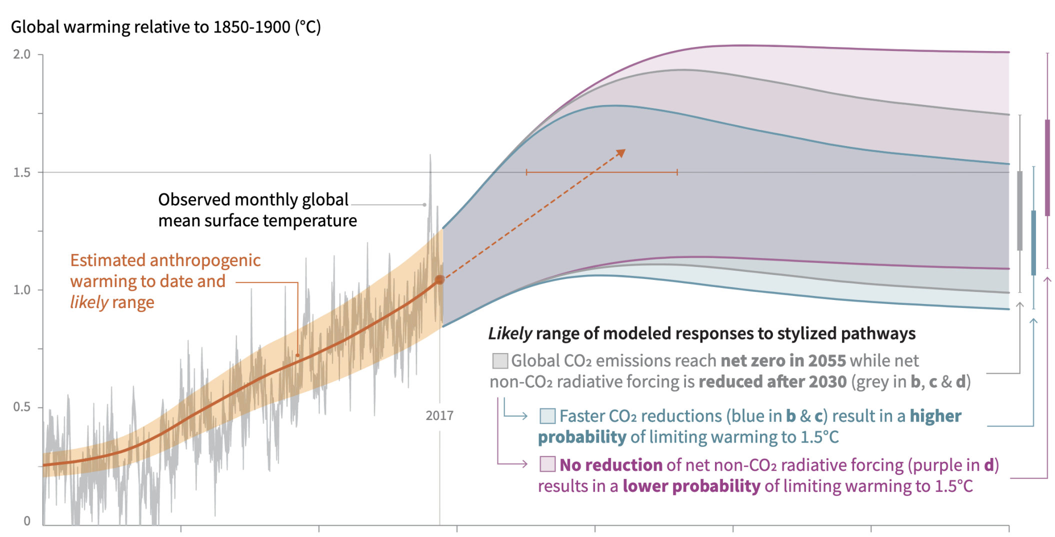

Figure 6: Atmospheric Changes with respect to Carbon Emissions and Global Warming - Observer and Projected. Source: Intergovernmental Panel on Climate Change, 2018

It is therefore imperative moving forward as a species that all steps are be taken to mitigate the emission of greenhouse gases and abate the advance of anthropomorphic climate change. The scale of the challenge is such that technology itself will prove critical in effectively combating this existential threat to civilisation. When considering energy consumption, the industrial sector (including the non-combusted use of fuels) currently consumes around half of all global energy and feedstock fuels, with residential and commercial buildings (29%) and transport (21%) accounting for the remainder (https://www.bp.com/en/global/corporate/energy-economics/energy-outlook/demand-by-sector.html).

```
energyBySector <- read_table2("/Users/alistairgj/Documents/GitHub/IoT_ResearchProject/IoT_Novembe
energyBySector <- gather(energyBySector, `Industry`, `Residential`, `CommercialAndPublicServices`
           `Transport`, `Other`, key = "Source", value = "MTOE")
```

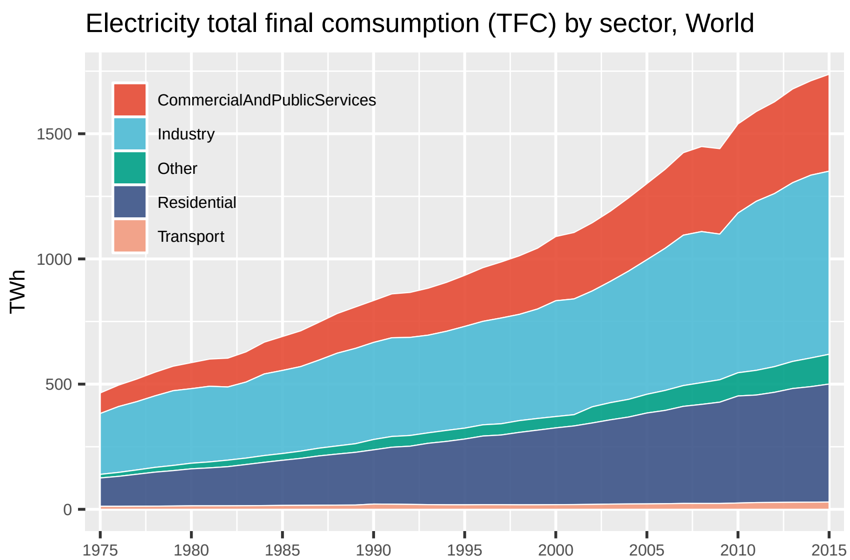

Figure 7: ADD TEXT, Source: Our World in Data

1.1.5 Optimising Global Energy Usage

Thus far, our staggering global achievements, including lifting hundreds of millions out of poverty, rapid global technological deployment and strong population growth has been inexorably linked to increased energy consumption. This in turn has led to perturbations in atmospheric chemistry, in the form of anthropogenic climate change (amongst many other environmental challenges), which fundamentally threatens our global achievements. It therefore stands that the key to continued human prosperity is to de-couple growth in energy demand from economic growth. In this work we will explore the possibility of reducing energy consumption via the optimization of services to end users in an IoT Smart Home.

1.1.6 The IoT Smart Home & Service Oriented Computing

As mentioned above, residential and commercial buildings account for 29% of energy demand globally (REF). In this work we propose that an avenue for reduced energy consumption is the optimization of existing services and utilities to end users in an IoT Smart Home. In this proposed smart home, the daily activities of the end user can be performed with the support of a personalised artificial intelligence (AI) system, such that the timing and manner of energy intensive activities is optimised to reduce overall energy consumption, whilst still considering the level of convenience afforded to the end user. This work uses sensor data analogous to what would be expected to be produced from a sensor rich IoT smart home and considers the interplay between end user convenience; predictive analytics; predictive service offering; energy consumption; demand response and smart electricity grids.

1.2 Aim

1.2.1 Research Questions

The aim of this work is to use sensor data from X to consider Y in the context of Z.

1.2.2 Research Contributions

The aim of this work is to use

2 Preliminaries

Significant growth in digital interconnectivity over the last 20 years has given the internet a pivotal role as an essential element of economic growth (Haseeb et al., 2019). Cloud computing and the IoT paradigm has enabled the association of previously disparate fields into a larger coherent framework. Namely, smart home appliances, demand-response, energy consumption, predictive analytics predictive service offering and end user convenience. This work proposes a framework for the association of sensor data into a model which can do X Y Z. **Mention internet in general?**

2.1 Related Work

2.1.1 Ubiquity of the World Wide Web and Implications for Energy Consumption

The International Telecommunication Union estimated about 3.2 billion people, or almost half of the world's population, would be online by the end of the 2015 ("Information and communication technologies (ICTs) | Poverty Eradication," n.d.). The impact of ICTs (i.e., internet usage and mobile cellular subscriptions), globalization, electricity consumption, financial development, and economic growth on environmental quality has been examined (Haseeb et al., 2019). By using 1994–2014 panel data of BRICS (Brazil, Russia, India, China & South Africa) economies, empirical results demonstrate that rise in both internet usage and mobile cellular subscription (ICTs) likely mitigates CO2 emissions in BRICS economies.

BRIDGE – Greater internet availability / connectivity facilitates the deployment of smart electricity grids

2.1.2 Smart Metering

A Smart meter is an electronic device that records consumption of electric energy and communicates the information to the electricity supplier for monitoring and billing. In the Unites States (for example) smart meters are a significant part of the larger Smart Grid infrastructure, and as far back as 2012, had been installed in over 25 million U.S. homes (Horne et al., 2015). Smart Meters transmit information about consumer electricity use to utility companies at vastly shorter time intervals than before and this information helps utility companies to coordinate power supply and demand, detect outages, implement time-of-use and dynamic pricing, and in other ways improve system efficiency and reliability. Additionally, these data are also becoming increasingly available, and sought-after, by end-users themselves. Indeed, the main purpose of provisioning smart metering data to end users is to encourage the use of less electricity, by better informing users of their consumption patterns (Jui-Sheng Chou, 2019). In the aggregate, these savings can significantly reduce national energy use and curb energy emissions while addressing pressing geopolitical and environmental concerns related to energy security and sustainability (Graab, 2011).

Since the widespread deployment of smart meter technology, there has been huge interest in the capability of these technologies with respect to the technical capacity of utility companies to manage demand (through demand response programs), incorporate renewable sources of electricity into the system, and increase the overall efficiency and reliability of the system (Horne et al., 2015). **It is worth noting, however, that the scenario described above relies on end user intervention.**

2.1.3 Electricity Demand Response, Smart Electricity Grids and Smart Home Data

Predicting and influencing residential energy use has been the subject to extensive study (**REF**). Literature indicates that factors such as occupant behaviour and socio-economic status are important **[15 from**

Cetin]. Nielsen attributed 36% of variation in energy consumption of homes to lifestyle and occupant behaviour, and 64% to socio-economic influences. This is exemplified by the work of De et al, who show that in developing nations, cooking consumes up to 90% of the overall residential energy consumption and is mainly based on non-renewable energy (International Energy Agency, 2006; De et al., 2013). Other factors such as climate zone, number of occupants, income level, age of home, and size of home have also been correlated with home energy use (K.S. Cetin, 2014). The has been much work both on using smart meter data to evaluate consumer behaviour, and the interplay between smart appliances and shifting households' electricity demand.

Kavousian et al identify the need for developing an analytical method that can leverage 15-minute or 30-minute interval energy consumption data produced via smart metering in order to improve the effectiveness of energy efficiency programs. They note that utilities spend millions of dollars annually to improve appliance energy efficiency. By way of example, in 2013 utilities in California US spent $80 M USD on appliance and plug load efficiency programs, the highest expenditure among all utility energy efficiency programs (**CPUC, 2013 FIND REF**). A need for analytical methodologies that can process smart meter data and allow energy reduction savings to be identified is demonstrated. Using a smart meter dataset of 4231 households, containing information on electricity consumption at 30-minute intervals, with aggregated local weather data, the authors use smart meter data to rank residential appliance efficiency. Various control methodologies are embedded into the analytical process, taking into account building type, building size (e.g., apartment, detached), household type (e.g., de facto, single, dependents), respondent age and heater type (e.g., electric or gas). The data set included 120+ household variables, many of which were highly correlated, model selection techniques were used to successfully reduce dimensionality. A set of load profiles were compiled based on the results (where normalized load was plotted as a function of hour of day). It was determined that household behaviour and demographic can be used to generate positive and negative energy efficiency coefficients with respect to load profiles. This work has implications for energy grid planning – for example in high-density housing versus suburban housing.

Karjalainen considers the manner in which smart meter data is communicated to end users, with the principal purpose of the paper being to study what kind of electricity consumption feedback consumers understand and prefer (Karjalainen, 2011). It is noted that the most effective feedback tools for engaging households in reducing energy consumption are both computerized and interactive. In this work a series of options for feedback is gathered, for example consumption (kHh), power (W), cost ($), environmental impact (e.g., kg CO2), total for household, disaggregation by appliance, real time, hour, day, and so on. Examples of feedback options are shown in Figure **X**, below.

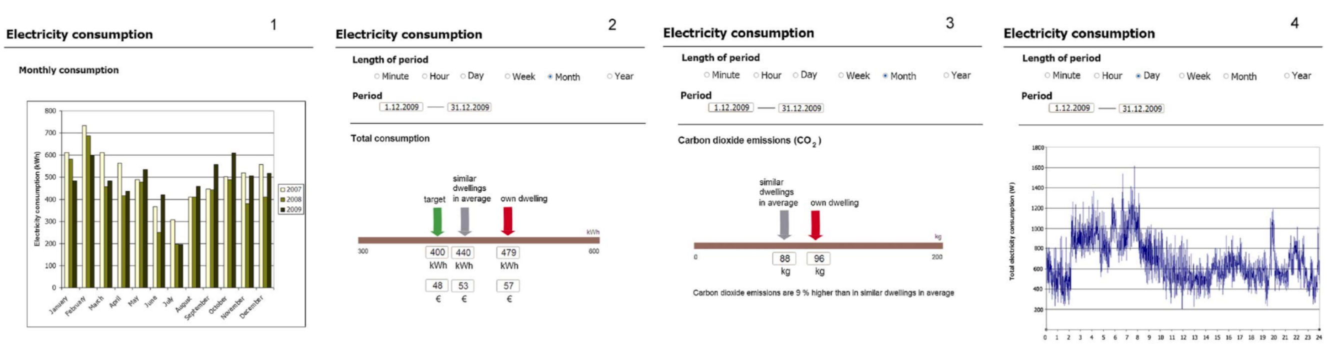

Figure 8: ADD TEXT

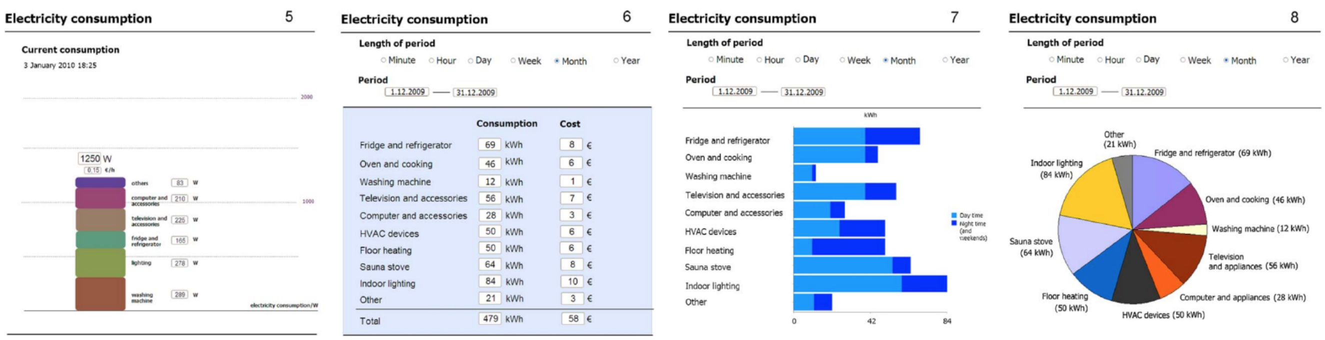

Figure 9: ADD TEXT

The results of qualitative participant interviews clearly showed that while some consumers are very interested in saving household electricity, other consumers show only a little interest (**NOTE they were asked about saving electricity, not about saving money or convenience**). When specifically asked to list what measures (if any) were typically taken at home to realize reductions in energy usage, some respondents listed numerous measures, while others just said they turn lights off in rooms that are empty. The author also found some participants were unaware of differences in stand-by versus active modes of operation for electrical appliances, resulting in practising inefficient energy saving measures in the home environment. When trialling feedback prototypes to participants, two main issues were encountered: (1) many people are not familiar with scientific units and do not understand the difference between W and kWh and (2) many people do not understand how carbon dioxide emissions are related to electricity consumption. It is perhaps surprising that the overall most popular prototype was '6', below. Perhaps because unlike the other prototypes, this clearly (the most clearly) articulates cost. The concept of convenience was absent entirely from consideration.

In the domain of supply and demand economics, consumer behaviour and smart appliances Kobus et al investigated if households can shift their electricity demand to times when electricity is abundantly available (Charlotte B.A. Kobus, 2015). Using a household electrical monitoring system (EMS) coupled to a smart appliance (smart washing machine), photovoltaic cells and the electricity grid, they were able to show that households can shift 10–77% of the electricity demand of their washing machine. This longitudinal study was conducted via the participation of 50 Dutch households over a period of one year. By utilizing an EMS which shows appliance status, dynamic tariff information and current status of household electricity usage, participants are able to schedule smart washing machine use in such a way that cost is minimized.

Figure 10: ADD TEXT

Their work is in response to two major challenges to domestic electricity supply and demand. **REMOVED: It is well observed that over electricity grids, demand shows a pattern over time, influenced by natural and social circumstances. This poses a great challenge when con-**

sidering renewable energy sources like PV – people are generally not at home when local electricity is abundantly produced (for example the daylight hours of Monday till Friday) The first being that the amount of distributed renewable electricity generation is increasing with time (for example, as more households install photovoltaic (PV) panels), and the second that electricity demand will continue to significantly increase moving into the foreseeable future. These developments pose great challenges to 'traditional' power systems, where supply follows demand entirely. Smart grids are proposed as a potential solution to facilitate the affordable introduction of cleaner electricity producing and consuming technologies.

Cetin et al consider electricity usage of appliance, as when aggregated, this accounts for approximately 30 per cent of electricity used in the residential building sector (K.S. Cetin, 2014). This, together with small appliances, home electronics and lighting, account for more than 2/3 of total residential electricity use. Cetin et al also highlight that influencing 'time of use' is becoming increasingly important to control the stress on today's electrical grid infrastructure. The authors seek to determine when refrigerators, clothes washers, clothes dryers and dishwashers are predominantly used (and thus consume energy) and what causes variation in their use. Using disaggregated energy data from 40 homes over a period of 1-year, normalized load profiles for the four target appliances were generated, as a percentage of daily electricity load.

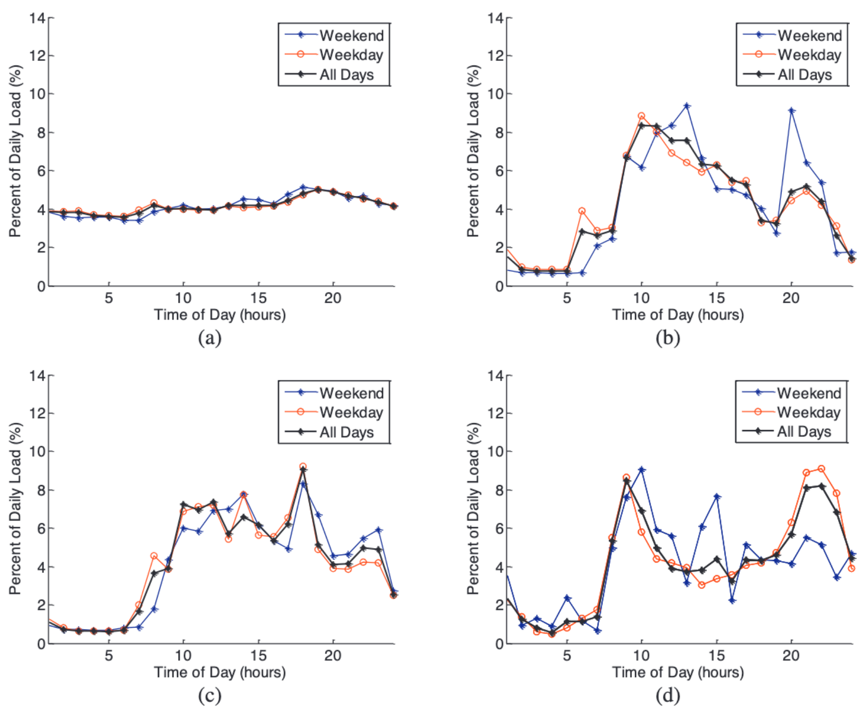

Figure 11: ADD TEXT

It was found that the refrigerator had the most consistent consumption profile across all homes surveyed. Influencing factors for the refrigerator were correlated to both indoor and outdoor temperature, however effect was found to be minimal. The clothes washer and dryer were found to have the greatest variation in normalised energy use by hour, with the greatest period of use from 9am until 2pm. The dishwasher had distinct peaks in load profile at 9am and 10pm. The authors found that user-dependent appliance use patterns vary more between homes and between days than automated appliances, weekday and weekend use patterns of appliances are similar, and that electricity use varies more between houses during peak use times of day than during low-use times. Murray et al further explore the topic of home appliance energy consumption using smart meter data, specifically pertaining to residential activities around food storage and preparation (D.M. Murray, 2018), with the aim of providing a model that can easily be applied to existing smart meter energy datasets. The authors use real-time energy consumption data from microwaves and ovens, from which user behaviour / desired outcome can be implicitly associated (for example, oven usage on a weekday between the hours of 6pm and 8pm can be associated with the behaviour of preparing dinner). Consideration is given to the concept of the RisingEdge (when an appliance is switched on or moves to a high consuming state) and FallingEdge (when an appliance is switched off or moves to a low consuming state) of power consumption. Accurate energy consumption models for major cooking appliances are successfully constructed, further re-enforcing the value and widespread applicability of residential smart meter data.

In this work,

This supports ___ that smart meters represent a valuable source of data that can be used for a variety of purposes, beyond just real-time monitoring. For example the model constructed by Murray et al does not rely on difficult to obtain parameters such as food type, temperature and weight. it has been shown that …

- Authors are able to use smart meter data to build accurate energy consumption models of major cooking appliances
- Smart =
- Using consumption data from real homes, the authors are able to create scalable mathematical models to account for energy consumption associated with food preparation
- Indeed, from an energy-efficiency perspective, targeted energy feedback can be provided on consumers appliance usage habits and activities (Stankovic et al., 2016) leading to informed energy savings programmes and retrofit strategies for replacing appliances.
- And if this variation can be influenced??

Demand Response allows for the management of demand side resources in real-time; i.e. shifting electricity demand according to fluctuating supply. When integrated into electricity markets, Demand Response can be used for load shifting and as a replacement for both control reserve and balancing energy. These three usage scenarios are compared based on historic German data from 2011 to determine that load shifting provides the highest benefit: its annual financial savings accumulate to €3.110M for both households and the service sector. This equals to relative savings of 2.83% compared to a scenario without load shifting.

[@article{feuerriegel2016}] discuss electricity demand response - allowing the management of demand side resources in real-time. **Based on Darby and McKenna (REF), we define demand response and a household action (automated, manual, or both) due to which electricity use is shifted in time in response to a price signal of other stimuli. This can result in more efficient usage of the available sustainable electricity, like self-consumption of on-site PV electricity (REF) and peak demand reductions (REF) - all from @article{kobus2015}**

2.1.4 Predictive Modelling and Forecasting

In the domestic predictive energy consumption space Basu et al consider home automation systems linked via a communication network to enable interaction, data collation and control of appliances remotely by end users (Kaustav Basu, 2012). The potential competing priorities of energy savings versus comfort optimization for home occupants is discussed. The objective of this work is to propose a learning system that is able to help the home automation system compute an energetic plan that is also satisfactory to user requests. Taking into account correlation between appliances, a time-series based multi-label classifier is used to predict appliance usage up to one hour into the future.

The attribute construction technique of Knowledge Extraction applied. This process aims to extract novel attributes from underlying substructures in the training instances in the form of sub-events, for example periodicity in data. This Knowledge Extraction process is similar to the implicit associations used by Murray et al, in the analysis of energy consumption for food preparation (D.M. Murray, 2018). The substructures are then fed to a propositional learner. The proposed model is trained in an iterative manner and attempts to take into account all the possible information based on consumption data, time of the event and meteorological information. Time is specifically model time as a periodic variable, segmenting on hour of day and day of week, noting that this takes into account the periodic nature of human behaviour. Using BR1, LP, CC1, CC2 and MLk machine learning algorithms, the precision, recall and accuracy for a variety of electricity-consuming appliances is determined (where each appliance constitutes a target variable over a set of iterations). In the evaluation phase they find that user behaviour toward an appliance is highly variable and the predictability of an appliance is dependent on the regularity of usage patterns of the inhabitants. It is specifically noted that *it is therefore very difficult for now to propose a generic methodology of appliance prediction for private houses.*

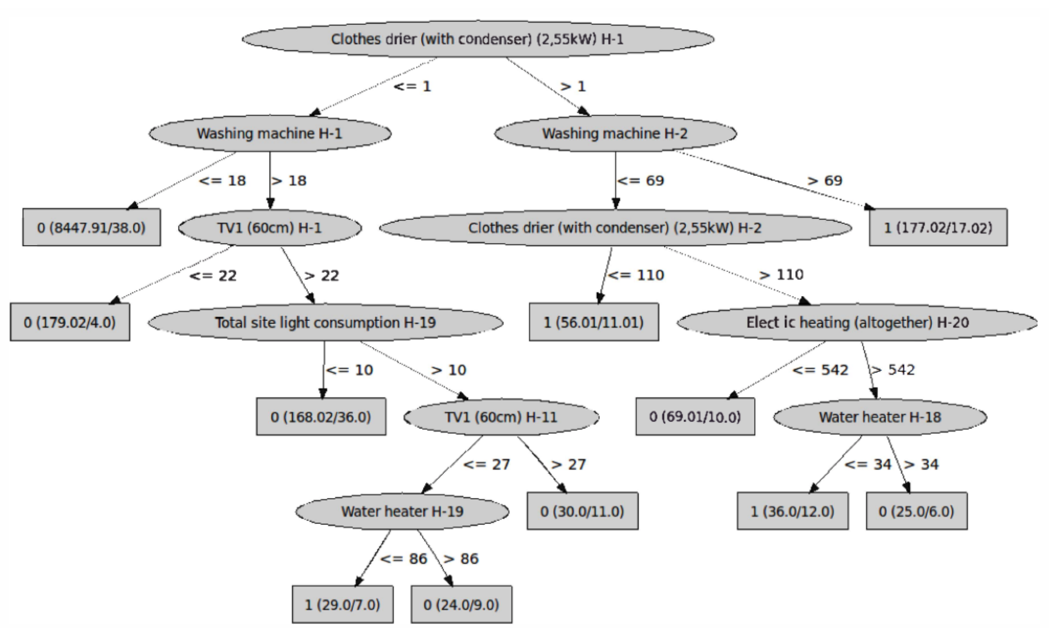

Figure 12: ADD TEXT

Chou et al consider global energy consumption in the residential housing sector in the context of domestic energy information system (EIS) and smart meter system (smart grids) technologies (Jui-Sheng Chou, 2019). They note two influential factors on overall residential energy consumption, the first being the type and number of electrical appliances and the second being the usage of these appliances by occupants. It is proposed that a major challenge for people who are willing to save energy at home is a lack of information about their energy consumption. To test this hypothesis, the authors develop a web-based energy information management system for the power consumption of home appliances that monitors the energy load of a home, analyses its energy consumption based on machine learning, and then sends information to various stakeholders. Interaction with end-users in the home is achieved via energy dashboards and emails. The authors propose that end-users of this system can use forecast information and anomalous data to enhance the efficiency of energy usage in their buildings especially during peak times by adjusting the operating schedule of their appliances and electrical equipment. **In order to achieve the desired model (???)**

During this study, an EIS with 5 main components was installed in an experimental building. The parts were as follows; (1) the internal communication network, (2) the data management infrastructure, (3) the automated prediction system, (4) the web-based system and dashboard and (5) the early warning system. Data from both the smart meter and from sensors was used to compile the model, including timestamps (YYYY-MM-DD hh:mm:ss), outdoor temperature (ºC), total building energy consumption (kWh), aggregated energy consumption data (e.g., second floor lighting) and individual appliances. The backend of the application was notably cloud-hosted.

- Improve consumer satisfaction by providing real-time services that enable end-users to monitor the energy consumption easily.
- Energy Information System - personalised terminal in domestic home which provides real-time or daily updating of hourly energy consumption data allows users to address building performance issues that are otherwise difficult to identify.
- The data stream from the smart meters arrives at 15 min intervals, generating 96 data points daily for each smart meter.
- The hardware of the DBMS consists of the communication system, the smart meter system, the environmental sensor system, and the server system, and the software programs that are used to process the data written in MATLAB. A DBMS stores real-time data that are retrieved from the smart grid metering infrastructure, including real-time energy consumption data, information about appliances, data from temperature and humidity sensors, and analytical results, including predicted electricity consumption, and data on electricity-saving alternatives. Moreover, the DBMS stores electrical parameters such as voltage, current, power, frequency, and power factor.
- The proposed prediction model (i.e.,SARIMA-MetaFA-LSSVR) consists of stage 1, involving univariate linear technique, and stage 2, involving optimized multivariate nonlineartechnique tofit the energy consumption data. SARIMA is adopted to handle the linear component, whereas Meta FA-LSSVR is used to model the nonlinear part of energy consumption data. In particular, the MetaFA is employed for automatically tuning LSSVR hyper parameters, thus improving the overall prediction performance.
- Output
- A web-based system to display energy consumption that is measured in real time with previously predicted values.
- Early-warning system integrated into web-based system to detect anomalous power consumption and provide warnings (notifications) thereof to users.
- The real-time prediction model in web-based system is designed and implemented based on our previous work[20,33,49], which developed a novel sliding window metaheuristic optimization-based machine learning system to analyze time-series data that are generated by a smart grid to efficiently forecast energy consumption one day in advance.

2.1.5 Smart Home and Energy Efficiency

[@article{bober2009distributed}NOT IN LIB] - As recently as 2009 it was posited that "The proposed model allows for introduction of power priorities for consumer's electrical equipment by the importance of their functions. The relevance of the functions carried out by the data device is evaluated by each consumer individually. The relevance of functions and priorities assigned to power mode / groups of electrical equipment can be changed over time."
'Haseeb et al [@article{haseeb2019does}NOT IN LIB] have (through review) shown that at a macro scale, global adoption of internet reduces energy consumption' + At the micro scale (that is, the behaviour of individuals…)

2.1.6 IoT and End User Convenience

The term Internet-of-Things (IoT) is used as an umbrella keyword for covering various aspects related to the extension of the Internet and the Web into the physical realm, by means of the widespread deployment of spatially distributed devices with embedded identification, sensing and/or actuation capabilities (Daniele Miorandi, 2012). The IoT paradigm is fundamental to this work, representing the confluence of multiple technological advancements (Luigi Atzori, 2010) including, ubiquity of internet access, the availability of high-performance internet connectivity, inexpensive consumer electronics with embedded sensing and control systems, automation, real-time analytics, machine learning, commodity sensors and embedded systems. In the IoT paradigm, digital and physical entities can be linked, by means of appropriate information and communication technologies, to enable a whole new class of applications and services (Daniele Miorandi, 2012).

One consequence resulting from the widespread deployment of consumer electronics with IoT capability is the evolution of the Internet from interconnecting computers to interconnecting things. Figure X, below, shows XYZ.

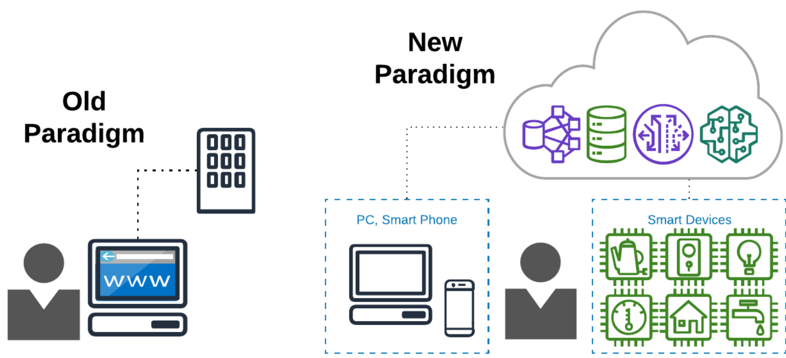

Figure 13: ADD TEXT

2.1.7 Heading???

Huang et al propose a novel service mining framework to personalize services in an IoT-based smart home (Bing Huang, 2018). This work considers the notion of personal 'convenience' as a driving force behind the provisioning of services to end users. That is, in an IoT smart home, where everything is interconnect, can services (for example, the switching on or off of a light) be automatically served to users such that their level of effort (to interact with their surroundings) will be diminished, and thus their level of convenience will be increased?

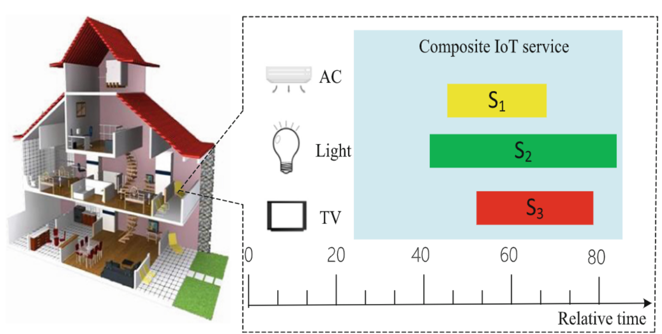

Figure 14: ADD TEXT

Using this framework, IoT services can be composed in a periodic fashion, for the convenience of the end user / consumer.

- Use intelligence to compose IoT services for the convenience of the end user / consumer

The input for this work was data from domestic IoT services, with a corresponding timestamp. The end user in this scenario performs daily activities by interacting with IoT services, these interactions are recorded as IoT service event sequences. A novel PCMiner algorithm (Periodic Composite IoT Service Miner)

Periodic composite IoT services can be loosely defined as the composite IoT services' repeating occurrence at certain locations with regular time intervals

The output for this work was an IoT service model and a composite IoT service model (based on spatio-temporal features). A periodic composite IoT service model to represent the regularity of composite IoT services occurring at a certain location at a certain time interval

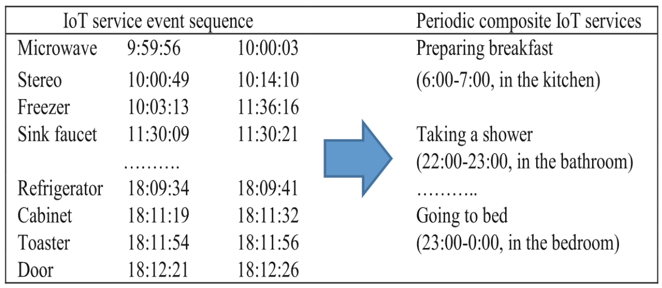

Figure 15: ADD TEXT

Critical to note that in this work model composition is reliant on temporal, proximity and duration features.

Output: Spatio-temporal proximity model

- Filter out insignificant spatio-temporal IoT service relationships Categories (e.g., objective):
- Quantify the correlation strength of among IoT services from time and location aspects
- IoT devices are highly heterogeneous in terms of supporting infrastructure ranging from networking to programming abstraction
- Hiding (abstracting) complex and diverse supporting infrastructure
- Proximity and correlation strength

Service Mining on the Web @article{zheng2009}

In the domain of Service Mining and Service Recognition, Zheng et al aim to provide end users with value-added services through the composition of existing services (output = composed web services). They attempt to craft a Service Mining Tool * More generalised * Technology = Service mining tool, web service mining * Output = composed web services * Aims at providing value-added services through composing existing services * It is essential to be able to proactively discover opportunities for composing useful services, even when the goals are unspecified at the moment, or simply hard to imagine or unknown * Much like the easy access to a glut of data has provided a fertile ground for data mining research, there is an expectation that an increase in Web services' availability will also spur both the need and opportunities to break new ground on Web Services Mining * We define Web Service Mining as a bottom-up search process aimed at the proactive discovery of potentially interesting and useful web services from existing services.

Two Main Challenges * Combinatorial Explosion * Evaluation of interestingness and usefulness

2.1.8 Summary

In summary there is lots of variability and this informs the construction of our model – it must be robust. By stripping the time variable this greatly simplifies modelling. As mentioned in the work of Murray et al (Appliance electrical consumption modelling at scale using smart meter data) smart home data contains implicit associations, the example given being that, if a person is using the oven between the hours of 6pm and 8pm on a weekday, it can be assumed they are preparing dinner – there is no need to explicitly determine this. In a similar way, we **believe** that we can implicitly obtain spatial and temporal information about how a series of appliances (electrical and non-electrical) are interrelated.

As demonstrated above, there has been a vast swathe of work done in the areas of smart meter data and predictive analytics, X, Y & Z. What is, however, of significant note, is the absence of holistic consideration with respect to energy consumption, smart home data, predictive analytics, IoT and end user convenience. As demonstrated above, numerous studies have found that creating and maintaining behaviour (such as monitoring a smart meter and then intervening) is ... shit.

Due to n = 1, it would be a fair criticism to say that the methodology employed cannot claim to be robust... this is fair. However, the Kavousian et al thoroughly analyse the effect of household variables on overall energy efficiency. Our analysis (n=1) has no such ... Take for example Consumer preferences for feedback on household electricity consumption (REF). The work considers, in isolation, the concept of saving electricity. For example during participant interviews, the line of questioning is framed around electricity saving NOT cost or convenience.

1. The cost of the sensors is negligible (EXCEPT for discussion later RE top 5 features)
2. The energy requirement for the sensors is negligible
3. In our proposed model, computation is performed in the cloud
4. An app is available to interact with the smart home
5. All of the electrical appliances in our smart home can be remote controlled to some extent

- What about, instead of responding to the average time, the antagonistic AI responded to price signals in the market?

2.2 The Data Set

The datasets were created during the thesis *Activity Recognition with End-User Sensor Installation in the Home* by Randy Joseph Rockinson, Submitted to the Program of Media Arts and Sciences, School of Architecture and Planning, in partial fulfilment of the requirement for the degree of Master of Science in Media Arts and Sciences at the Massachusetts Institute of Technology (MIT) February 2008 (REF).

The aim of Rockinson was to considered the effect of end user versus professional installation of a sensor array in the home - on the basis that, if installation of sensors is to be considered as a high initial cost, and a barrier to entry for end users wanting this technology, is there a difference if a professional versus an end user performs the installation?

Between 80-100 reed switch sensors where installed in two single-person apartments collecting data about human activity for two weeks. The sensors were installed in everyday objects such as drawers, refrigerators, containers, etc to record opening-closing events (activation deactivation events) as the subject carried out everyday activities.

3 Data Preprocessing & Visualisation

3.1 Importing & Preprocessing the Activities Meta Data

The dataset `S1Activities.csv` was imported into the interactive development environment. These data contains a tabulated summary of Heading, Category, Subcategory and a corresponding code. After importation, the dataset has dimensionality of [3, 33], with `Heading`, `Category` & `Subcategory` present as non-null objects, as seen in Table **??** below. The attribute `Code` (which codefies the unique set of Heading, Category) was imported as an index value. At this time, the activities data will not be kept in it's native state and will not be subject to preprocessing.

Table 1: The S1 activities dataset

	Heading	Category	Subcategory
1	Employment related	Employment work at home	Work at home
5	Employment related	Travel employment	Going out to work
10	Personal needs	Eating	Eating
15	Personal needs	Personal hygiene	Toileting
20	Personal needs	Personal hygiene	Bathing

3.2 Importing & Preprocessing the Sensor Meta Data

The dataset `S1sensors.csv` was imported into the interactive development environment. These data contains a tabulated values for Sensor ID, Room and Sensor Activity Type, with no header row present in the original dataset. After importation, the dataset has dimensionality of [3, 76], with header 0, 1 & 2 corresponding to SensorID, Room & Sensor Activity Type, respectively, as seen in Table 2. All attributes are nominal, and were imported as dtype str, accordingly. Attribute attribute 1 & 2 contain degenerate values (e.g., Bathroom & Light Switch). This will be addressed in the subsequent data preprocessing. The preprocessing of the sensor data is a critical step in our analysis. Careful consideration of the data, including the presence of duplicates. This is because if we dont have a sufficient understanding of where and why duplicates exist, we will not be able to satisfactorarily preprocess them. Failure to do so we mean that there is potentail degeneracy in our source dataset, leading to unknown issues with our downstream analysis.

Table 2: The S1 sensor meta data

0	1	2
100	Bathroom	Toilet Flush
101	Bathroom	Light switch
104	Foyer	Light switch
105	Kitchen	Light switch
106	Kitchen	Burner

Column [1] & Column [2] of the sensor data will be concatenated, whitespace will be removed, all text will be cast to lowercase and a final whitespace strip will be performed. The python script `S1sensorsPreprocessing.py` is run perform several preprocessing steps in these data. The script concatenates the attributes `dsS1Sensors[1]` and `dsS1Sensors[2]`, with an underscore. Whitespace is then stripped and all string values are coerced to lowercase. This newly created attribute is then added to the dataframe, as seen below (REF). Additionally, the attributes `0`, `1` & `2` are renamed `subActNum`, `room` & `activity`, respectively.

- Data types

- IF a sub-act requires electricity

Table 3: The first iteration of processed S1 sensor meta data

subActNum	room	activity	concat
100	Bathroom	Toilet Flush	bathroom_toiletflush
101	Bathroom	Light switch	bathroom_lightswitch
104	Foyer	Light switch	foyer_lightswitch
105	Kitchen	Light switch	kitchen_lightswitch
106	Kitchen	Burner	kitchen_burner
107	Living room	Light switch	livingroom_lightswitch

The function `getUniqueValues.py` is invoked to provide a means of capturing a list of unique values in a given column of a dataset. The newly created `dsS1Sensors` is then checked for duplicates in two attributes, `subActNum` & `concat`. The number of unique values in `dsS1Sensors.subActNum` is found to be 76, demonstrating that this attribute contains a set of completely unique values (recall n(rows) = 76). The number of unique values in `dsS1Sensors.concat` is found to be 41, demonstrating that despite the concatenation methodology, their are still duplicate values in the dataframe. These duplicate values warrant further investigation. The function `length(unique())` was used to list counts with their corresponding values.

- `length(unique(py$dsS1Sensors$subActNum))` = 76
- `length(unique(py$dsS1Sensors$concat))` = 41

COMMENT

Upon compilation of the above summary list, and with reference to the original work (REF) is was determined that these values result from multiple sensors with extremely similar functionality. For example, kitchen_burner has a value of n=4 - this is because on the burner in the apartment under investigation, there were 4 individual burners present. Similarly, kitchen_cabinet has a value of n=15, indicating that for the various cabinets in the apartment, each were given sensors. On the one-hand, this level of granularity may provide fertile grounds for advanced analysis, HOWEVER, for the purposes of this research project, such values will serve to increase the dimensionality of the overall dataset. High dimensionality can lead to difficulties with plotting, ... ML (REF) and thus IN A SUBSEQUENT PREPROCESSING exercise these values will be collapsed down to have n=1.

Creation of JSON Catalogues PRIOR to dup removal - why? Because even if a key-value pair cannot be matched it will simply be ignored **Prior to dupe removal** As this work is largely concerned with energy usage in the home, the sub-activities will be categorized based on their energy requirement. That is, if a sub-activity requires an input of energy beyond what the end user alone can provide, it will be classified as `energyReq` = true. Whereas, if a sub-activity is able to be performed through only interaction with the end user, it will be classified as `energyReq` = false. By way of example, the sub-activity `bathroom_toiletflush` will have an `energyReq` equal to false, while the sub-activity `bathroom_lightswitch` will have an `energyReq` equal to true. Each row (n=76) needs to be inspected manually to determine if the activity requires electricity.

Function `reqEnergy_containSpecialChar.py` is run - After visual inspection, uses `reqEnergy = 'ligh|burn|mach|toas|freez|dvd|lamp|washer|dry|exh|disp|frig|oven|hot|shower|micro'` to find subActivities which require energy input beyond that of the end user - Checked below for SPECIAL CHAR - Special CHARs removed

Table 4: Demo table

	subActNum	room	activity	concat	reqEnergy	specialChar
58	82	Office/study	Drawer	office/study_drawer	FALSE	TRUE
68	92	Office/study	Light switch	office/study_lightswitch	TRUE	TRUE

```
#%run -i reqEnergy_containSpecialCharClean.py
dsS1Sensors['concat'].replace('office/study_drawer','study_drawer',inplace=True)
dsS1Sensors['room'].replace('Office/study','Study',inplace=True)
dsS1Sensors['concat'].replace('office/study_lightswitch','study_lightwitch',inplace=True)
dsS1Sensors['subActNumConcat'] = 'subActNum_' + dsS1Sensors['subActNum'].astype(str)
dsS1Sensors.drop(columns=['specialChar'], inplace=True)
```

Table 5: Demo table

	subActNum	room	activity	concat	reqEnergy	subActNumConcat
59	82	Study	Drawer	study_drawer	FALSE	subActNum_82

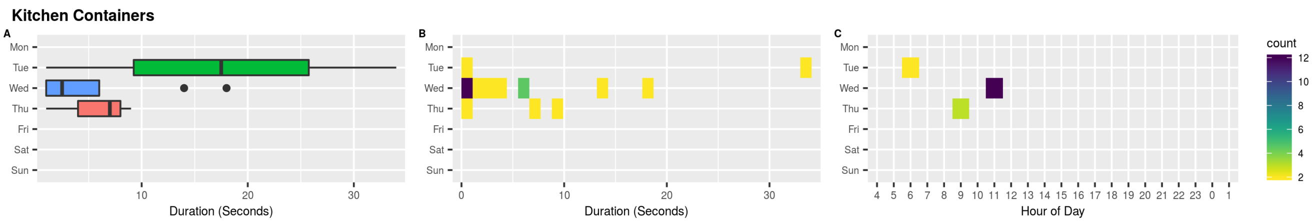

Figure 16: A caption

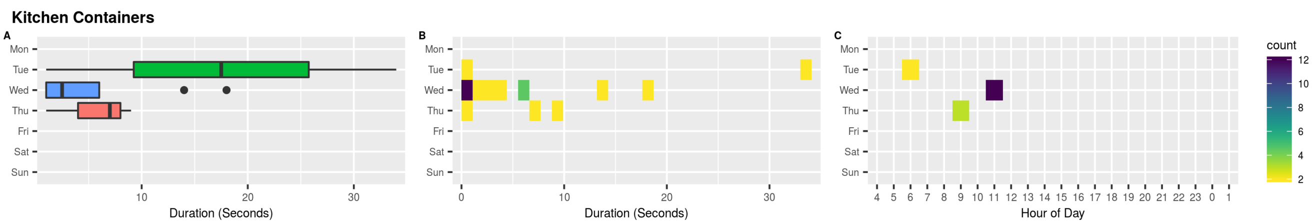

Figure 17: A caption

```
%run -i reqEnergy_containSpecialCharClean.py
```

Test for in 19 we see XYZ

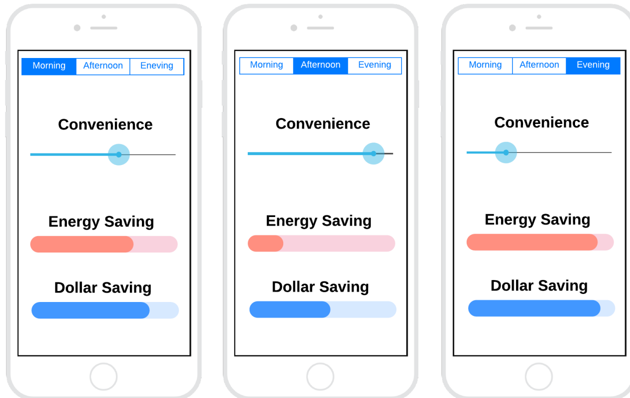

Figure 18: ADD TEXT

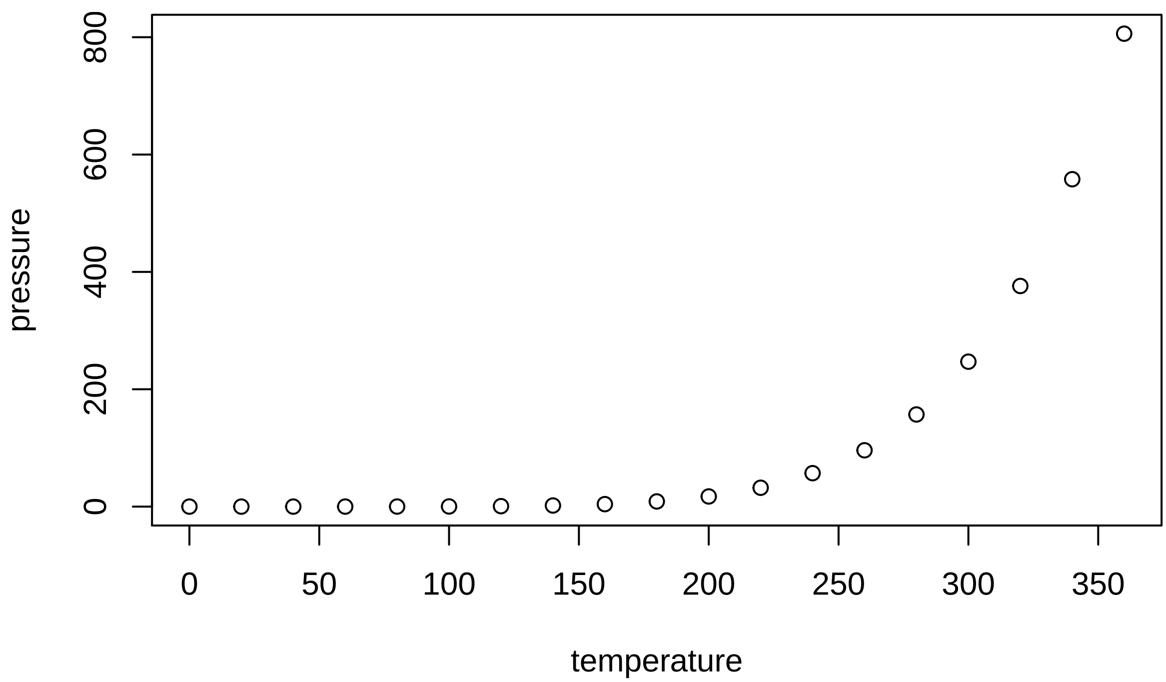

Figure 19: This is a caption

Reduce image border (not working???) https://holtzy.github.io/Pimp-my-rmd/