Deep Neural Networks for Non-Intrusive Load Monitoring

By

OMAR AL ZEIDI

MONASH University

Faculty of Information Technology

A dissertation submitted to MONASH University in accordance with the requirements of the degree of MASTER IN DATA SCIENCE at the Faculty of Information Technology.

NOVEMBER 2018

Word count: ten thousand and four

Abstract

Non-Intrusive Load Monitoring NILM (a.k.a as energy disaggregation) is a form of blind-source separation problem, where the aim is to estimate appliance-by-appliance electricity consumption in a household from a single meter that measures total consumption. Researchers used variety of methods to solve this problem, including Combinatorial Optimization, variants of hidden Markov model (HMM) and more recently deep neural networks. This work use deep convolutional networks to disaggregate the consumption of five household appliance using sequence-to-point training. This study has three main contributions: (1) modification of NILM regression network to consider variant power signature and time-embedding to learn better features for disaggregation, (2) building unsupervised quantization model to detect appliance states, (3) reformulating NILM as classification problem to simplify the training and improve the accuracy. These modification achieved state-of-the-art performance by reducing the overall mean absolute error (MAE) and signal aggregate error (SAE) by 31% and 37%, respectively.

Dedication and acknowledgements

I would like to thank my supervisors Dr. Reza Haffari and Dr. Zahraa Abdallah for their continuous and endless support over the course of the past year. This thesis would not have been completed if not for their dedication, motivation and insightful ideas.

A special thank to my wife Noura, who took care of all the tiny details of my life when I was submerged into my studies and research. She was my best friend through the ups and downs of this journey. My love and thanks to my little piece of heaven, my son Al-Ezz, for coping with an occupied dad with long study hours.

This work is dedicated to my parents for whom I owe everything in my life. Your endless love and support made me what I am.

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: .... DATE: ....

Table of Contents

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

1.1 Background

The growth of industrial activity over the last century, accompanied by accelerated population growth, has put enormous stress on energy resources of the planet (IEA, 2017b). Electricity is one of the key energy forms that are widely demanded for industrial, commercial and residential usage. The International Energy Agency estimated the global electricity consumption at 61 terawatts per day in 2015, where 66% of that energy was generated from fossil fuels (IEA, 2017a, p. 60). The continuous depletion of these fossil fuels will profoundly impact the sustainability of electricity generation, a fact that mandate urgent actions to balance the supply-demand cycle. While alternative electricity sources, such as nuclear and renewables, are developing very slowly, maintenance of this balance mandates an efficient and intelligent use of available electricity.

Energy disaggregation is one of the emerging methods to build intelligent Energy Management Systems (EMS) for conservation and efficient consumption (Faustine et al., 2017). In principle, disaggregation is a form of blind-source separation where the aim is to reconstruct individual appliances power consumption from the house main reading. With reference to figure XX, the disaggregation objective is to infer the consumption profiles for microwave, washing machine and kettle from the main consumption profile (red). This approach is referred to as Non-Intrusive Load Monitoring (NILM), where the main power reading is the only input required for a model to reconstruct the individual appliances profiles. However, unlike many of the conventional blind-source separation problems, disaggregation seems to be a non-trivial task. A typical household contains many appliances with different consumption signatures and various activity durations. With many devices being active at the same point of time this task become more challenging.

An alternative approach to NILM is to apply Intrusive methods, where meters are installed at each appliance to record direct consumption measurements (Hart, 1992). While the latter method provides an accurate breakdown of consumption per appliances, their intrusive nature and the requirement for individual smart meter per appliance make them impractical and expensive (Zoha et al., 2012). In contrast, NILM methods are relatively cheap considering the requirement for a single smart meter only. Many countries are already changing their energy policies to mandate the installation of smart meters at every household. In 2013, the Victorian state government started an initiative to upgrade all the electricity network with smart meters. This meters provide real-time consumption feed to customers and utilities (DED, 2015). These policy changes bring high potential for NILM systems, where they can be deployed at a large scale, utilizing the existing infrastructure and available data sources. With such system, utilities can provide customers with itemized energy bill, where one can know the percentage of consumption per appliance in the house. This emphasis the urgent need to develop robust and accurate NILM models capable of of providing accurate energy disaggregation results.

1.2 Objective and Research Questions

The main objective of this project is to build NILM model based on Deep Neural Network architectures, which can outperform stat-of-the-art model proposed by (Zhang, Zhong, Wang, Goddard, & Sutton, 2016). Stemming from this objective we endeavor to answer the following three research questions: First, How to improve the accuracy of NILM models in the context of Deep Neural Network architecture. Secondly, what pre-processing and data augmentation techniques can be used for this specific case of energy disaggregation. Third
To achieve this objective and answer the research questions we build-up on the recent work by (Kelly & Knottenbelt, 2015a) and (Zhang, 2016). In order to make these models applicable to real-life scenarios, we train the networks on real data from the UK-DALE data set provided by (Kelly & Knottenbelt, 2015b). To enable benchmarking the results with previous work, the models are tested on 5 appliances: kettle, microwave, washing machine, fridge and dish washer, which are typically used for NILM evaluation. The results are benchmarked with (Kelly & Knottenbelt, 2015b) and (Zhang, 2016) based on two evaluation metrics: Mean Absolute Error (MAE) and Signal Aggregate Error (SAE).

This project was motivated by recent development in deep neural networks, which have revolutionized many machine learning domains such as machine translation and computer vision. Hence these techniques has the potential to provide outstanding results compared to conventional methods of energy disaggregation.

1.3 Thesis Outline

This thesis is organized in four chapters. The first chapter gives an introduction of the project, the objective, motivation and main research questions. Chapter 2 reviews the previous literature on energy disaggregation and NILM models. Section 2.2 highlights and compares the type of features that can be used for disaggregation while section 2.3 gives a comprehensive overview of the datasets that can be used for such task. Sections 2.4 and 2.5 provides an overview of the type of models used for NILM and highlights the advantages and disadvantages associated with each modeling approach. Based on preceding literature review, section 2.6 highlight the key knowledge gaps encountered in previous studies.

Chapter 3 is dedicated for detailing the methodology and implementation details of this thesis. It starts by setting a high level overview of the methodology adopted throughout the project. Section 3.2 highlights the main resource and tools used for conducting this work. Section 3.3 provides an overview of the UK-DALE dataset which is used for training, testing and benchmarking the results with previous studies. Since deep neural network models require extensive datasets for training, section 3.4 details the process of generating synthetic data to expand the training set size. Sections 3.6 and 3.7 explore our proposed features of smoothed power variant and time-embedding, which were used to augment the network input in an attempt to have better features for disaggregation. Section 3.8 highlights the pre-processing techniques conducted on the network input. Section 3.9 highlights our proposed approach for unsupervised consumption quantization method to transform the appliance consumption from continuous space into discrete space suitable for classification. Section 3.10 discuss the details of the experiments conducted in this study and their corresponding network architectures.

Chapter 4 focus on evaluating the results obtained from this study which starts by highlighting the key evaluation metrics used to assess the performance of our models in Section 4.1. Section 4.2 benchmarks the results with previous NILM work.

Building on the results obtained from previous sections, Chapter 5 provides a brief discussion of the results obtained from the two experiments conducted in this project. Section 5.2 highlight the real-life impacts of this study and how can they influence different aspects of the daily . Finally, Section 5.3 wraps up the outcomes of this study and highlight some of the limitation and potential improvements in future work.

CHAPTER 2

Literature Review

The research on Non-Intrusive Load Monitoring systems (NILM) was pioneered by the seminal work of George (Hart, 1992), which set the basis for modern NILM research.
As well as reviewing the recent related literature this chapter aims to highlight the main techniques, challenges and directions of NILM research.

2.1 General NILM Framework

In general, NILM energy disaggregation can be viewed as a form of optimization problem, where households aggregate power consumption is expressed as $P(t)$. Assuming $i = 1, 2 \ldots n$ appliances in a household where each appliance has a binary (on/off) state denoted by $S_t^{(i)}$ to indicate the i^{th} appliance state at time t. With prior knowledge of appliance mean power consumption, denoted by $P_t^{(i)}$, one can express the total household consumption at any time "t" as the summation of individual appliances:

$$P(t) = \sum_{i=1}^{n} S_t^{(i)}.P_t^{(i)} \tag{2.1}$$

The above could be re-framed as a minimization problem, to infer the appliance state vector S_t^* and consequently the appliances power consumption at each time step "t":

$$S_t^* = argmin_S \left| P(t) - \sum_{i=1}^{n} S_t^{(i)} P_t^{(i)} \right| \tag{2.2}$$

While the model suggested by equation (2.2) can be solved by any form of combinatorial optimization, gradient descent or even heuristic methods, it suffer from several limitations:

1. The model assumes a constant number of appliances (n) in a household, which is hardly the case in modern houses, where the number of appliances could change due to new purchases,

seasonal variation etc. This variability in appliances also inhibits any prior knowledge of appliances mean power consumption $P^{(i)}$.

2. Minor fluctuations in the total aggregate $P(t)$ might lead to entirely different state vector S_t^*. This is inherent from the intractable nature of equation (2.2) and the lack of exact solution.

In addition to these limitations, most modern appliances have multiple running states, rather than a binary (on/off) state. These different states have different corresponding power consumption $P^{(i)}$, which is not captured by the model.Figure 2.1 shows typical consumption patterns for four different appliances. Surveys on the consumption of appliances show that they can be categorized into four main types (Hart, 1992; Zoha et al., 2012):

i Type I Appliances: These are the most common appliances, which have two (on/off) states only. This includes appliances like kettles, lamps and TV screens.

ii Type II Appliances: In comparison to type I, these have multiple operating states with different power consumption. They are usually referred to as Finite State Machine (FSM) (Hart, 1992) and include appliances like fans, dish washers, dimmer lamps and stove burners.

iii Type III Appliances: have an infinite number of states. Hence, their power consumption is continuously varying with time. Appliances like power drillers and sewing machines are included in this category.

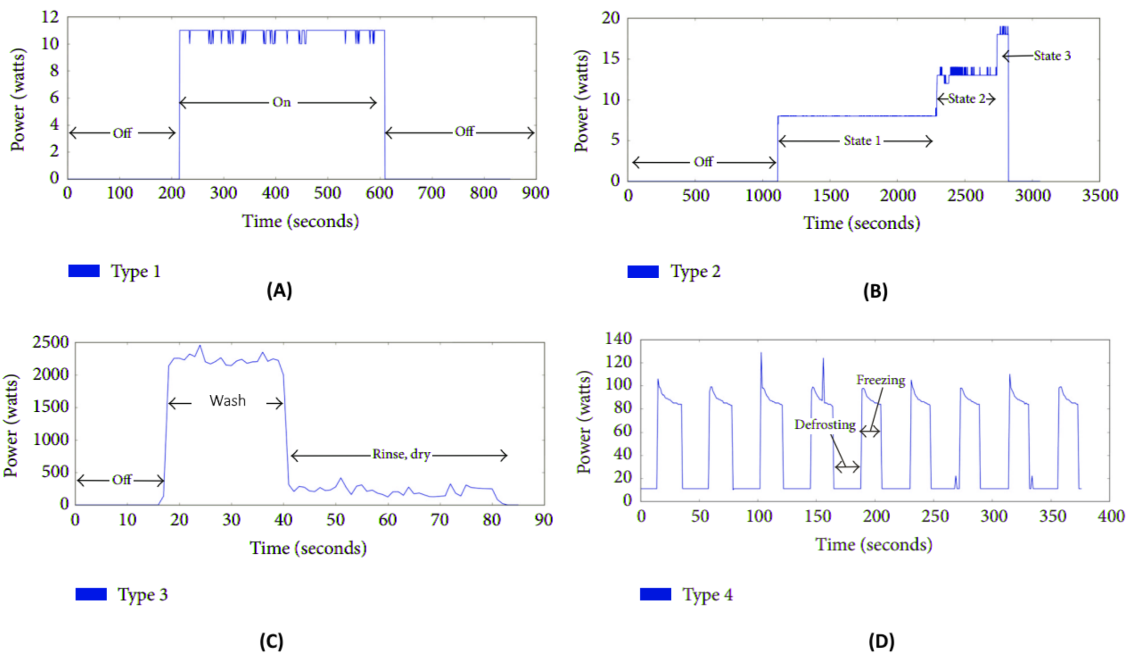
iv Type IV Appliances: which run all day long and draw constant power. This include appliances such as fridges, Wi-Fi routers, smoke detectors, CCTV cameras and refrigerators (Baranski & Voss, 2003).

Most NILM models perform well on type I and II appliances. However, the performance degenerates with the other types. In fact, the performance varies with the type of appliance based on the features used for disaggregation. Next section highlights two types of features used by NILM models.

2.2 NILM Disaggregation Features

The type and number of features that can be captured from household consumption is determined by the acquisition frequency of the metering device. Industrial-grade meters are capable of capturing data at a high sampling rate, which provides detailed features including voltage-current waveforms, real power and reactive power measurements. On the other hand, most commercial meters are only capable of capturing low-frequency data at the scale of few measurements per second. Such low sampling rate can only capture real and reactive power with no capability of collecting voltage-current waveforms. In general, these collected features can be split into two main categories: Transient and Steady-State features.

Figure 2.1: A.Type I: binary state lamp, B. Type II: adjustable light, C.Type III: washing machine with variable speed motor, D. Type IV: fridge



Transient-State Features

Transient-state features are derived from high-frequency consumption data, which captures momentarily changes in power corresponding appliances activation and changes in state. With these features, the appliance activations tend to be more distinctive, since most appliances have unique transient-state signatures. Features extracted from the current-voltage waveforms enabled researchers to provide better disaggregation results for type III appliances, such as variable speed drives and power drills (Wichakool, Avestruz, Cox, & Leeb, 2009).
However, the high cost and complexity of the hardware needed for capturing consumption at such high sampling rate inhibits applications of these methods. As well as the acquisition cost, processing transient-state features is computationally expensive, considering the large volume of samples that get generated per second. Hence these features are not considered in this study and more emphasis is given to steady-state features.

Steady-State Features

These features are derived from the sustained power consumption throughout appliances steady-state operation. In general, Real Power (P) and Reactive Power (Q) are two of the main features derived from the steady-state operation. Since these features are long-lasting, they can be captured using commercial meters at few seconds resolution with least cost and installation complexity. The other main advantage is the low volume of data generated, which reduces the computational cost required for analysis.

2.3 Datasets

In addition to aggregate consumption, disaggregation algorithms require sub-metered consumption per individual appliance. This appliance-level consumption is essential to provide ground-truth data especially for neural networks which require this for training and estimating accurate model parameters. This section reviews four of the most popular NILM datasets, whereas Table 2.1 provides a summary of all publicly available datasets.

The Reference Energy Disaggregation dataset, shortly known as REDD, was the first and most popular dataset, released for NILM community in 2011 (Kolter & Johnson, 2011). REDD contains aggregate and appliance-level power consumption monitored on six houses in Massachusetts. This data was used in many of the early disaggregation models, e.g. Hidden Markov Models. This data covers one month of consumption only, which make it impractical for DNN models that require extensive datasets for training.

In 2012, BLUED dataset was released by Anderson et al. (2012), capturing aggregate power consumption from a single household in Pennsylvania, US. Rather than sub-metering appliance-level consumption, this dataset used events-tags to label changes in appliances state over time. Hence, for every appliance one can know the start and the end of the activity based on text-tags but no actual power consumption was recorded. This lack of ground-truth appliance consumption limits the types of models that can be used with this dataset and hence it was ruled out from the scope of this project.

In 2014, Imperial College released the UK Domestic Appliance Level Electricity dataset (UK-DALE), which monitored total household and appliance-level consumption for 2.5 years (Kelly & Knottenbelt, 2015b).The metering spanned over 5 households covering up to 54 appliances monitored with 6 seconds resolution. The aggregate household was provided at high resolution of 16kHz and low resolution of 1Hz. Amongst the available dataset, UKDALE was found to be the best match for the scope of this project for following reasons:

1. It contains both aggregate and appliance-level consumption at fine resolution, suitable for supervised learning methods.

2. Availability of consumption on multiple households make it possible to train and test appliance models on different houses, which should gives an indication of the model generalization to appliances not seen during training.

3. The fact that this data was used extensively used in recent NILM research enables comparing and benchmarking this work with previous studies.

Its worth to mention that there was a recent release of new NILM dataset from "Pecan St Inc", US, which covers consumption from over 700 houses (Parson et al., 2015). This is by far the most extensive publicly available dataset covering both aggregate and sub-metered data. However, using it is out of the scope of the project since it will need extensive hardware resources for training as well as the requirement for manual data wrangling to align the data source for disaggregation.

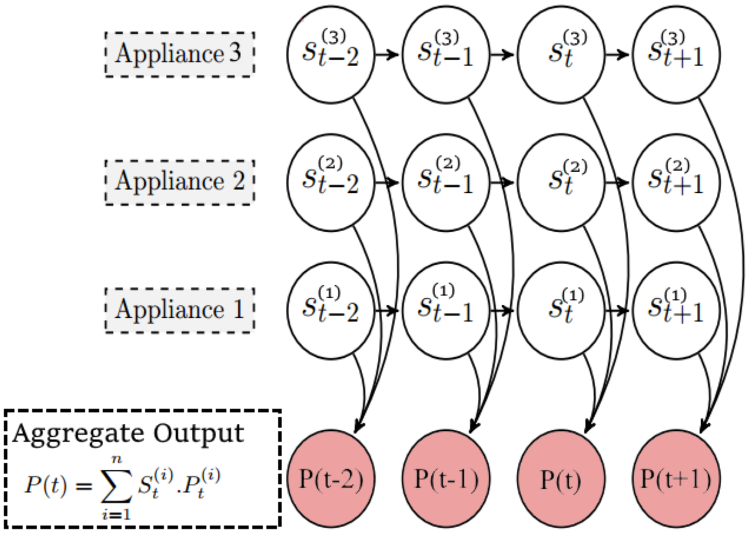
Table 2.1: Summary of the publicly available NILM datasets. Modified after (Faustine et al., 2017)

Dataset	**Location**	**Duration**	**No.of houses**	**Sensor/house**	**Aggregae Sampling Rate**	**Sub-metering Sampling Rate**
REDD	USA	3-19 days	6	24	15KHz	0.5Hz and 1Hz
BLUED	USA	8 days	1	Aggregated	12KHz	N/A
UK-DALE	UK	2.5 years	5	5-54 devices	16 kHz	6 sec
Dataport	USA	4 years	722	1 to 25	1 min	1 sec
BERDS	USA	1 year	1	4	20sec	20 sec
Smart	USA	3 months	3	21-26	1 sec	1 sec
DRED	Netherlands	6 months	3	12 appliances	1 sec	1 sec
Tracebase	Germany	N/A	15	158 devices	N/A	1 sec, 10 sec
AMPDS	Canada	1 year	1	19	1min	1min
AMPds2	Canada	2	1	21	1min	1min
iAWE	India	73 days	10	33 devices	1sec	1sec or 6sec
REFIT	UK	2years	20	11	8sec	8sec
GREEND	Austria/ Italy	1year	9	9	1 sec	1 sec
ECO	Switzerland	8months	6	N/A	1 sec	N/A
IHEPCDS	France	4 years	1	3	1min	1min
OCTES	Various	413months	33	N/A	7sec	NA
HES	UK	1 month - 1 year	251	13-51	2min	2 min
ACS-F1	Switzerland	2 & 1 hour	NA	100 devices	10sec	10 sec

2.4 Conventional NILM Models

Most conventional NILM techniques rely on variants of Hidden Markov Models (HMM) to infer appliance states from aggregate consumption. These models could be related to the general NILM framework highlighted in section 2.1. In HMM, appliance states $S_t^{(i)}$ are treated as a sequence of hidden variables defined by probability distribution function, whereas aggregate consumption

Figure 2.2: Representation of Factorial Hidden Markov Model (FHMM) used for NILM. Modified after (Zoha et al., 2012).



$P(t)$ is an observed variable corresponding to the combination of these hidden states. Figure 2.2 show representation of Factorial Hidden Markov Model (FHMM) developed by (Zoha, Gluhak, Nati, & Imran, 2013), which provided up to 90% disaggregation accuracy for low power type I appliances. Kolter and Jaakkola (2012) developed modified version using additive factorial hidden Markov model (AFHMM), which provided new disaggregation baseline with average accuracy of 71% considering 7 appliances of various types. Several other variants of FHMM has been implemented by Parson, Ghosh, Weal, and Rogers (2012) ,H. Kim, Marwah, Arlitt, Lyon, and Han (2011) and Zeifman and Roth (2011).

Despite their popularity FHMM suffer from several fundamental limitations (J. Kim, Le, & Kim, 2017):

1. The increase in model hidden states and number of appliance leads to an exponential increase in computational complexity, which in turn degrades the inference accuracy.

2. Despite the high accuracy with Type I appliances, HMM models show low performance with multiple state appliances.

3. Even with type I single-state appliances, these models experience difficulty in distinguishing appliances with similar consumption.

To avoid these limitations, researchers started looking into alternative NILM architectures.

2.5 Deep Learning NILM Models

Driven by the success of Deep Learning in conjugate machine learning domains, like computer vision (Krizhevsky, Sutskever, & Hinton, 2012), speech recognition (Graves, Mohamed, & Hinton, 2013) and machine translation (Cho et al., 2014), researchers started experimenting with these architectures in NILM application. Until recently, development of these models was inhibited by the lack enough data that can be used for training such deep architectures. Furthermore, the high computational cost for training has been a solid entry barrier for most researchers. The availability of dedicated datasets published by Kelly and Knottenbelt (2015b), Kolter and Johnson (2011), Anderson et al. (2012), Parson et al. (2015) and several others, along with the advancements in computational capabilities introduced by computer GPUs, made these types of models more feasible (Marques, 2016).

The interest in these models stems from the following facts about their underlying neural network:

1. They are flexible, allowing similar architectures to be used from other machine learning domains.

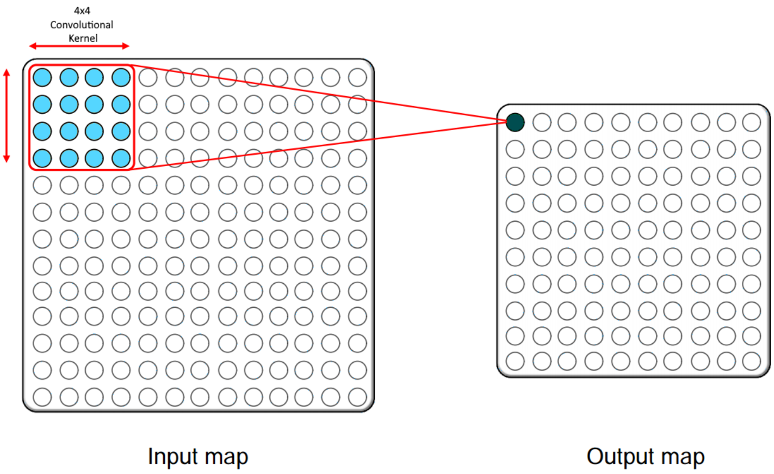
2. They allow automatic extraction of abstract, high-level features, which enables detecting appliance consumption from aggregate data with minimum prior knowledge about system latent variables, such as the number of appliances, number of states and mean consumption.

3. Despite the high computational cost during the initial training phase, once trained, the models can provide disaggregation results with minimal cost and with relatively good accuracy.

4. They are known to outperform state-of-the-art models in other machine learning domains.

The work by Kelly and Knottenbelt (2015a), presented the first novel and rigorous experiment to apply Deep Learning in NILM. More recent work by Zhang et al. (2016), J. Kim et al. (2017) and Krystalakos, Nalmpantis, and Vrakas (2018) followed up to cover the shortcomings in earlier models. Overall, all these studies were based on two architectures, namely deep convolutional networks (CNN) and recurrent neural networks (RNN). The choice to to use these two specific architectures stems from their ability to learn latent features from sequential data, which is the underlying data structure in NILM dataset. This project is focused on using CNN due to their ability to learn good feature representation for NILM and their lower training cost compared to the recurrent counterparts. Prior to going into the applications, it is important to have an overview of key components and underlaying principles of CNN. In contrast to conventional feed-forward networks, CNNs are distinguished by the presence of following components:

Figure 2.3: Schematic overview of convolution for 2D input map.



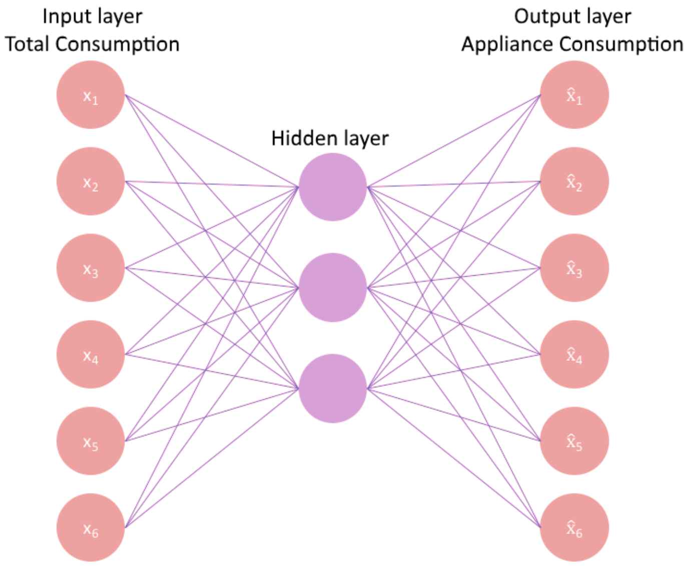
- **Convolutional Layers:** These are composed of kernels, which are small filters passed over the input to extract local features from large input. Figure 2.3 shows schematic illustration of 2D 4x4 convolutional kernel used on image recognition and computer vision applications. The kernel is strided across the height and width of the input map. The output of this process is referred to as local receptive field, which is essentially lower level representation of local features in the input map. By stacking multiple convolution layers one can obtain fine-grained feature map that autonomously learns the underlying latent structure in the input.Unlike in computer vision, NILM data is 1D time series, where 1D convolution is used to learn underlaying features required to capture appliance signature.

- **Pooling Layers:** With multiple stacks of convolution layers, the number of network parameters expands, which requires extensive memory resources. Pooling layers act on down-sampling convolution output, normally by taking mean or max over the output. This process reduces the number of parameters the networks has to learn and hence reduce the memory requirements and speed up the training.

Kelly and Knottenbelt (2015a) used CNN to build de-noising auto-encoder (DAE) architecture for disaggregation. The work was built up on applications from image processing domains, where the original objective is reconstruction of grainy photographs (Vincent, Larochelle, Bengio, & Manzagol, 2008). In its simplest form, DAE is a neural network that attempts to reconstruct its input through a sparse number of features in the hidden layer, as shown in figure 2.4. By increasing the number of hidden layers, the network is able to learn high-level abstraction of the input data.

While conventional DAEs require artificial corruption of the input data to train the network, in NILM applications, aggregate consumption is considered as the noisy input, where the objective is to reconstruct the clean power demand of the target appliance corrupted by consumption of

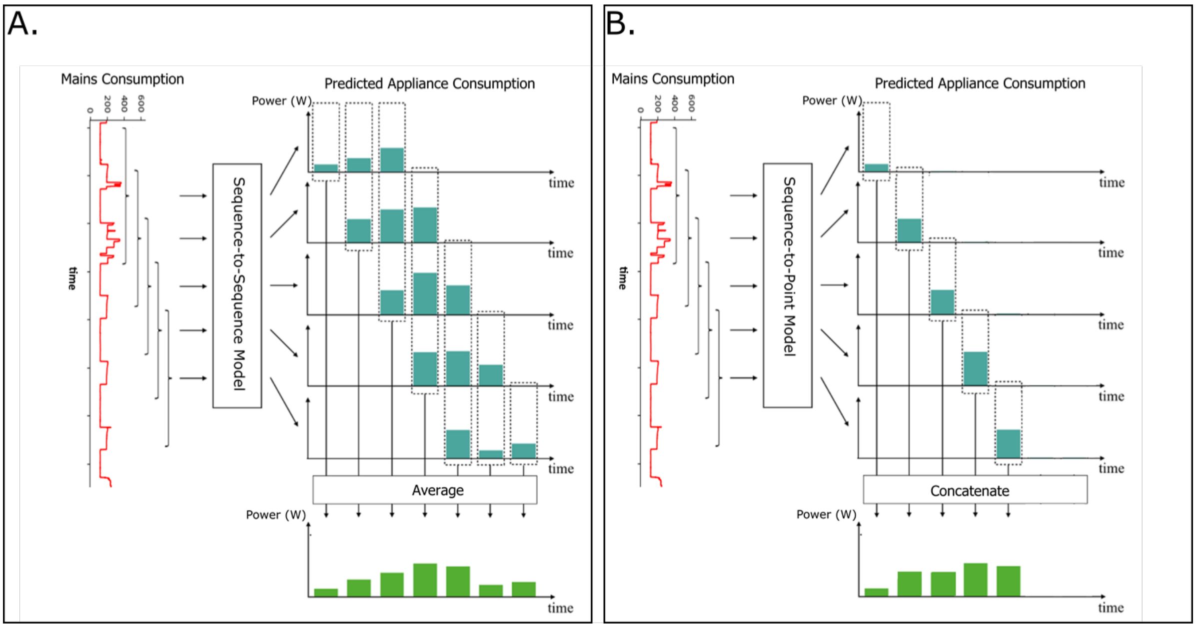
Figure 2.4: Schematic overview of convolution for 2D input map.



other appliances (Kelly & Knottenbelt, 2015a). The network was trained by feeding a fixed length sequence of the aggregate consumption as input where hidden convolution layers attempt to reconstruct the corresponding consumption of the target appliance. In principle, the network learns the best non-linear regression to map from a a sliding window of mains input to an equal length window of appliance consumption. Since this is formed as regression problem the loss is defined by Root Mean Squared Error. During disaggregation the network outputs overlapping windows of the target appliance consumption where the final result is obtained by taking the mean over those windows, as shown in figure 2.5. This approach is referred to as "sequence-to-sequence" learning and has been used successfully used in single channel blind source separation problems (Grais, Sen, & Erdogan, 2013). The main disadvantage of this approach is that the final output of the target tend to underestimate the true consumption due to averaging across overlapping windows.

To overcome the limitations in sequence-to-sequence method, Zhang et al. (2016) proposed a alternative approach of sequence-to-point learning, which has been widely used to model the distribution in images and speech. The principle behind this approach is that the network is fed a fixed-length window of the main consumption whereas the output is the mid-point corresponding to each window. By doing so, the network learns a sequence to point mapping and eliminates any overlap in the output. Figure 2.5 show comparison between these two approaches. In the right figure, the model gives a single point prediction for each window of the mains consumption. The final concatenated output gives the prediction for the whole consumption duration without the need for averaging or ensembling as in the sequence-to-sequence approach.

Figure 2.5: A. Sequence-to-Sequence disaggregation. B. Sequence-to-Point disaggregation.



2.6 Knowledge Gaps

Stemming from the previous literature review, following knowledge gaps were discovered:

1. Most of the previous studies focused on sequence-to-sequence approach for disaggregation, whereas conjugate fields like single channel blind source separation (BSS) show improved results using sequence-to-point learning.

2. All previous models used mains raw power as the only input to the network to learn disaggregation. One can use different variants of the raw power series to augment the network input and learn better features for disaggregation. Researchers in time-series analysis and audio processing use many variants of the raw input series to improve networks performance.

3. There has not been any work delineating the effect of using consumption time-stamps as a feature for disaggregation. Intuitively, some appliances are known to be active at certain times of the day. For example, microwaves normally show increased activity around meal times. Hence time-stamps could present a useful feature to predict the appliance state based on the time of the day.

4. NILM has always been tackled as a regression problem where the network is trained to predict the appliance power as real-value number. However, most household appliances are known to have discrete consumption states where disaggregation could be tackled as classification task, where the objective is to predict the appliance states rather than their consumption profiles.

This project aims to fill these knowledge gaps through the work conducted in the subsequent chapters of this thesis.

CHAPTER 3

METHODOLOGY

3.1 Overview

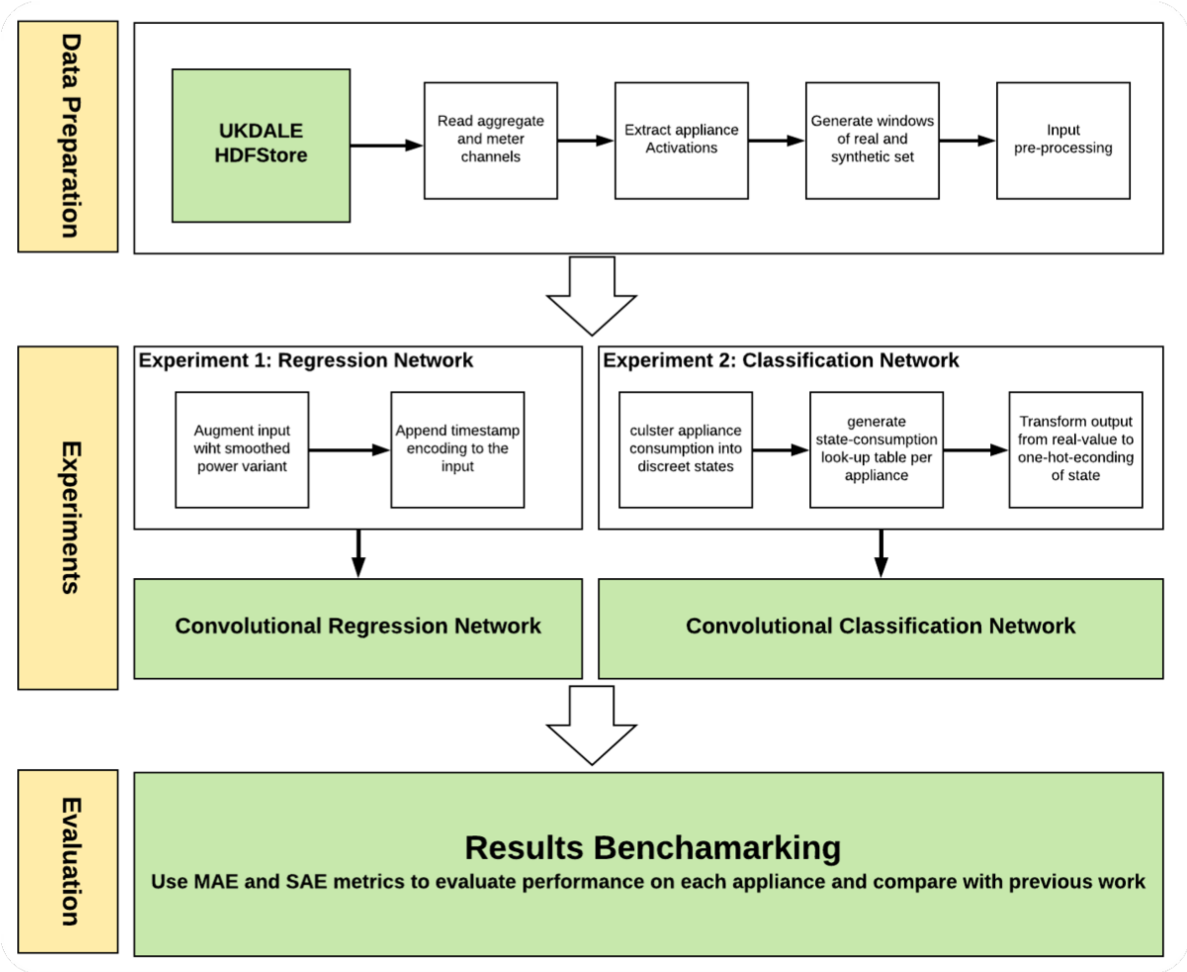
This chapter details the methodology adopted to address the two knowledge gaps and limitation encountered in previous NILM work. To tackle these challenges, this project utilize convolutional neural network (CNN) architectures to explore two experimental settings for NILM. Figure 3.1, show the overall work flow adopted in the methodology of this work. The work is based on two experiments that aim to improve on the previous NILM work conducted by Kelly and Knottenbelt (2015a) and Zhang et al. (2016). The regression network experiment focuses on augmenting the network input with new features that could aid disaggregation performance, whereas the classification experiment tries to reformulate NILM from the conventionally used regression setting to a more confined classification setting. Subsequent sections of this chapter detail the procedures undertaken to conduct these experiments.

3.2 Resources and Tools

3.2.1 Hardware Resource

The models were trained on two Nvidia Tesal P100 GPU with 16GB memory. The devices were equipped with CUDA and cuDNN libraries which optimize the GPUs for neural network operations such as back-propagation and convolution. In addition to the GPUs, a 12-core Intel E5-2680V3 processor was used for pre-processing the input batches simultaneously. This hardware architecture was available as part of M3 MASSIVE cluster provided by MONASH University.

Figure 3.1: Methodology overview.



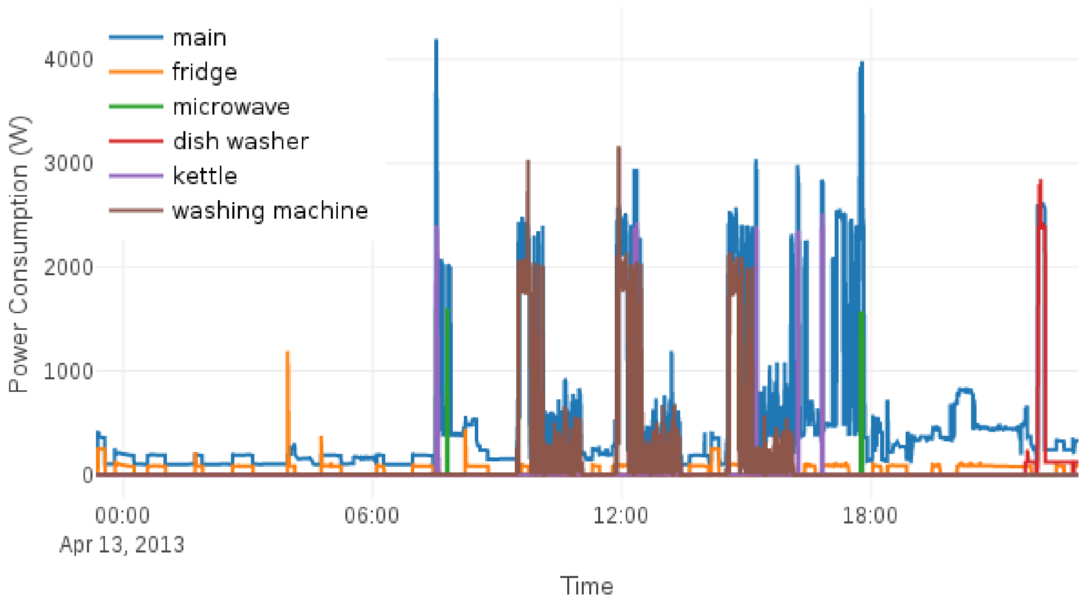
3.2.2 Keras

Keras is a high-level neural network API built on top of the popular Tensorflow library. The library provides modular and user friendly structure that allow for quick easy and fast prototyping of neural network. It is tightly integrated with python NumPy, allowing efficient matrix operations.

3.2.3 NILMTK

NILMTK* is an open source tool-kit designed specifically for energy disaggregation. The library builds-up on Pandas and HDFS store to provide a universal API interface to wide range of NILM dataset. While the library has ready algorithms for disaggregation, processing and benchmarking, it was used here for the purpose of loading and handling the large UK-DALE dataset. Most of the pre-processing and time series manipulation was done at a lower level with Pandas and NumPy packages.

Figure 3.2: Sample of the consumption profiles for a typical consumption day from house 1.



3.3 UK-DALE Dataset

The work in this project is based on the UK-DALE dataset provided by Kelly and Knottenbelt (2015b). The data includes mains and appliance level consumption from five households in the UK. Table 3.1 shows general summary statistics of the dataset. As can be seen in the table, monitoring duration varies across houses. In general house 1 has the longest duration and the highest number of monitored appliances. As the data was transmitted over radio signal there were periods of lost communication as highlighted in the dropout rate. Proportion of energy sub-metered measures the ratio between the mains to the sum of sub-meters reading, which varies from 80% at houses 1 and 2 to 19% at house 3.

Figure 3.2 show typical consumption profile for a day at house 1. Preliminary investigation shows two main concerns with the consumption data. First, most appliance show noisy profile attributed to acquisition errors, which resulted some spikes that exceeds the mains profile. This is evident in the washing machine and the dish washer spikes at 12:00 and 22:00 respectively, figure 3.2. This can resolved by setting a maximum power threshold when loading the appliance data . A second issue arises from the slight time misalignment between main and appliance meters, which can be resolved by manual adjustment of the profiles.

In addition to the temporal variability in consumption,seen in figure 3.2, consumption varies significantly across the houses, as illustrated by the consumption histograms in 3.3. Figures 3.4 shows the daily percentage of consumption relative to total consumption of one week, which reflects different consumption behavior at daily granularity.

Figure 3.3: Histogram of main consumption over 1 week duration.

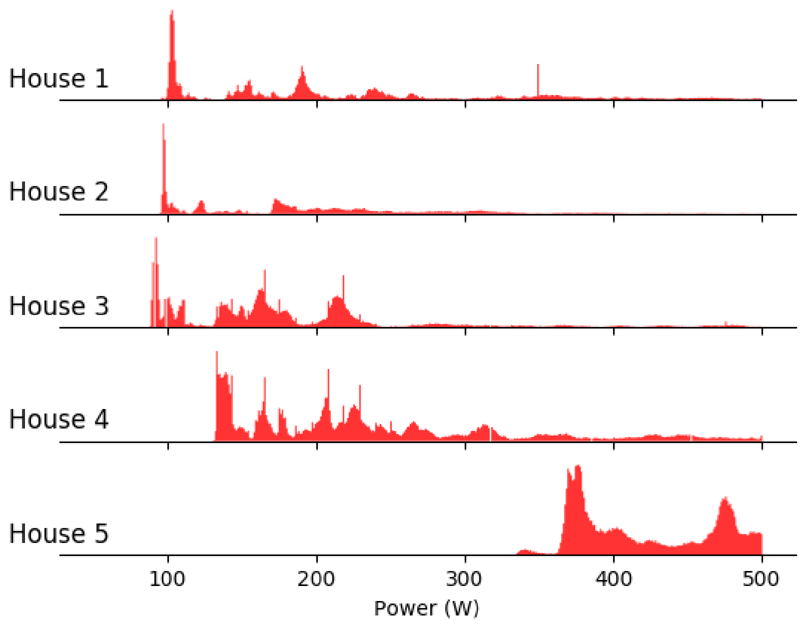


Figure 3.4: Mains power consumption density in the five houses over 1 week duration.

Table 3.1: General summary of the UK-DALE dataset, adapted from (Kelly & Knottenbelt, 2015b)

House	**1**	**2**	**3**	**4**	**5**
Building type	end of terrace	end of terrace		mid-terrace	flat
Number of occupants	4	2		2	2
Total number of meters	54	20	5	6	26
Number of site (mains) meters	2	2	1	1	2
Sample rate of mains/meters	16 kHz & 6 sec	16 kHz & 6 sec	6 sec	6 sec	16 kHz & 6 sec
Date of first measurement	9/11/2012	17/02/2013	27/02/2013	9/03/2013	29/06/2014
Date of last measurement	5/01/2015	10/10/2013	8/04/2013	1/10/2013	13/11/2014
Total duration (days)	786	234	39	205	137
Total uptime for mains meter (days)	655	140	36	155	131
Uptime proportion	0.83	0.6	0.93	0.75	0.96
Average mains energy consumption per day (active kWh)	7.64	7.17			13.75
Correlation of sum of submeters with mains	0.96	0.86	0.47	0.55	0.9
Proportion of energy submetered	0.8	0.68	0.19	0.28	0.79
Mean dropout rate (ignoring large gaps)	0.02	0.02	0.02	0.02	0.02

3.4 Appliance Activity Extraction

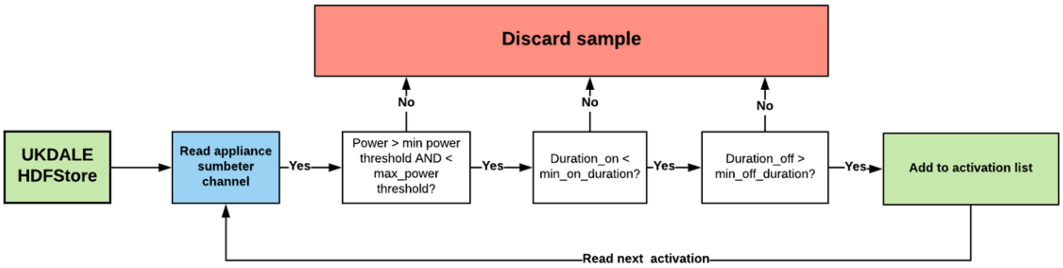
This work focuses on disaggregating the consumption for five appliances: kettle, microwave, fridge, washing machine and dish washer. These particular appliances were chosen for the following reasons:

1. They account for high proportion of consumption in a typical house.

2. They have a variety of consumption profiles ranging from Type I in a kettle to Type III in a washing machine.

3. All the five appliances were available at multiple houses, which make it possible to train on at least 2 houses and test on a different house. Table 3.2 shows the houses used for training and testing the networks in this study.

Figure 3.5: Work-flow to extract appliance activations.



NILMTK was used to extract the appliances activation based on the parameters shown in Table 3.3. The maximum power threshold and minimum on duration are set to eliminate spurious spikes caused by acquisition errors. The minimum off duration helps to capture appliance with short off durations, like the short pauses of a washing machine between wash, dry and rinse cycles.The overall workflow of activation extraction is highlighted in figure 3.5.

Table 3.2: Split of houses between testing and training set.

Appliance	Training houses	Testing house
Kettle	1, 2, 3, 4	5
Fridge	1, 2, 4	5
Washing machine	1, 5	2
Microwave	1, 2	5
Dish washer	1, 2	5

Table 3.3: Parameters used for extracting appliances activation.

Appliance	On power threshold (W)	Maximum power (W)	Minimum On Duration (sec)	Minimum Off Duration (sec)
Kettle	2000	3100	12	0
Fridge	50	300	60	12
Washing machine	20	2500	1800	160
Microwave	200	3000	12	30
Dish washer	10	2500	1800	1800

3.5 Synthetic Data

Using mixture of synthetic and real data is common procedure in training neural networks. It is used to increase the size of the training set in order to improve robustness and generalization of neural networks. This work uses 75%:25% mixture of real to synthetic data.
The synthetic data is built by stacking multiple activation from the five chosen appliance to generate aggregate a profile resembling the mains consumption. All the activations used were randomly sampled from a pool activation series extracted from the five houses.
The data generation process starts by creating two empty vectors where one holds the input of the network and the other holds output. The size of these vectors is dictated by the window size used for network input. The activations are added randomly with target appliance having 50% chance of appearing in this synthetic window. This probability is chosen to ensures a well balanced set between on and off training samples. The other four appliances are considered as distractors and their activations are added based on 25% probability. All activations can appear anywhere in the aggregate window, with exception of the target activation that has to be added at a location where it is fully contained in the window.

Kelly and Knottenbelt (2015a) criticized this approach of random synthesis on the basis that it ignores a lot of the structure present in real aggregate data. Their argument was based on the observation that activity of appliances like the kettle and toaster tend to be highly associated and appear within few minutes from each other. However, we see these observations to be more related to personal consumption behavior that might vary between households and/or geographic regions. A random synthesis process helps to generate a training set free of any trends associated consumer behaviors, which in turn helps with network generalization.

3.6 Smoothed Power Variant (ΔP)

As highlighted in chapter 2, the raw mains series tend to be heavily corrupted with acquisition noise, especially on the onset of appliance activations or transitions between different states. This is caused by meters limitation in capturing high frequency alternating current at low sampling rate. The noise can negatively impact the network learning by masking the genuine appliance signatures and introducing. In time series analysis domain, researchers normally consider a smoothed variant of the raw series to eliminate such noise, using well established methods like simple moving average (SMA) or exponential smoothing (Hansun, 2013). These methods will definitely help in smoothing out noise, but they will also tend to blur out appliance activation and transient-state signatures, which are key features used by the the network to learn the appliance consumption patterns. These appliance activation and transient-state signatures are very important for type II, III and IV devices, where an appliance shows multiple state transitions in one running cycle.

To avoid the lose of this key information, we use variant power signature (ΔP) introduced by J. Kim et al. (2017). The main idea here is to generate two exponentially weighted moving averages of the power, each with different reflection rate (α). Then, difference of these two series is taken to get the ΔP power variant. The reflection rate α is a float ranging from 0 to 0.99. A lower rate will make the series smoother by putting more weight on earlier readings in the series. Hence, gradual changes and steady-state patterns will be maintained, whereas transient-state and momentarily noise will be smoothed out. Conversely, a higher reflection rate will put more weight on recent readings, which preserves transient-state signatures with slight smoothing of the noise. By taking the the difference of these two smoothed series one can get a noise free signal while still maintaining the transient-state signature and appliance activation transitions.

J. Kim et al. (2017) empirically found that reflection rates α of 0.1 and 0.01 are appropriate for capturing signature of most household appliances. Algorithm 1 shows the procedure for generating the smoothed power SP where P is the raw series and α is the reflection rate. Using two smoothed series SP^1 and SP^2 with reflection rate α, algorithm 2 shows the procedure to generate the variant power ΔP. Figure 3.6, shows the corresponding series generated by these two algorithms. By using this noise-free ΔP power variant as an additional input feature, the network is expected to learn better representation for predicting the appliance consumption.

Algorithm 1: Generate exponentially weighted moving average (EWMA) from raw power series.

Input: Raw power $P = p_1, p_2, \ldots p_T, SmoothingFactor\ \alpha$
Output: Smoothed series $SP = sp_1, sp_2, \ldots sp_T$

```
SP ← 0
for i ← 0 to T do
    d ← p_i − sp_{i−1} ;
    sp_i ← sp_{i−1} + α · d
return SP
```

Algorithm 2: Generate ΔP power variant from two smoothed series $SP^1 and SP^2$.

Input: Smoothed Series $SP^1 = sp^1_1, sp^1_2, \ldots sp^1_T$, $SP^2 = sp^2_1, sp^2_2, \ldots sp^2_T$
Output: $\Delta P = \delta p_1, \delta p_2, \ldots \delta p_T$

```
for i ← 0 to T do
    δp_i ← sp^1_i − sp^2_i
return ΔP
```

Figure 3.6: A. Raw power series showing noisy outliers, B. SP^1 smoothed series using reflection rate $\alpha = 0.1$, C. SP^2 smoothed series using reflection rate $\alpha = 0.01$, D. ΔP Power variant used for augmenting network input.

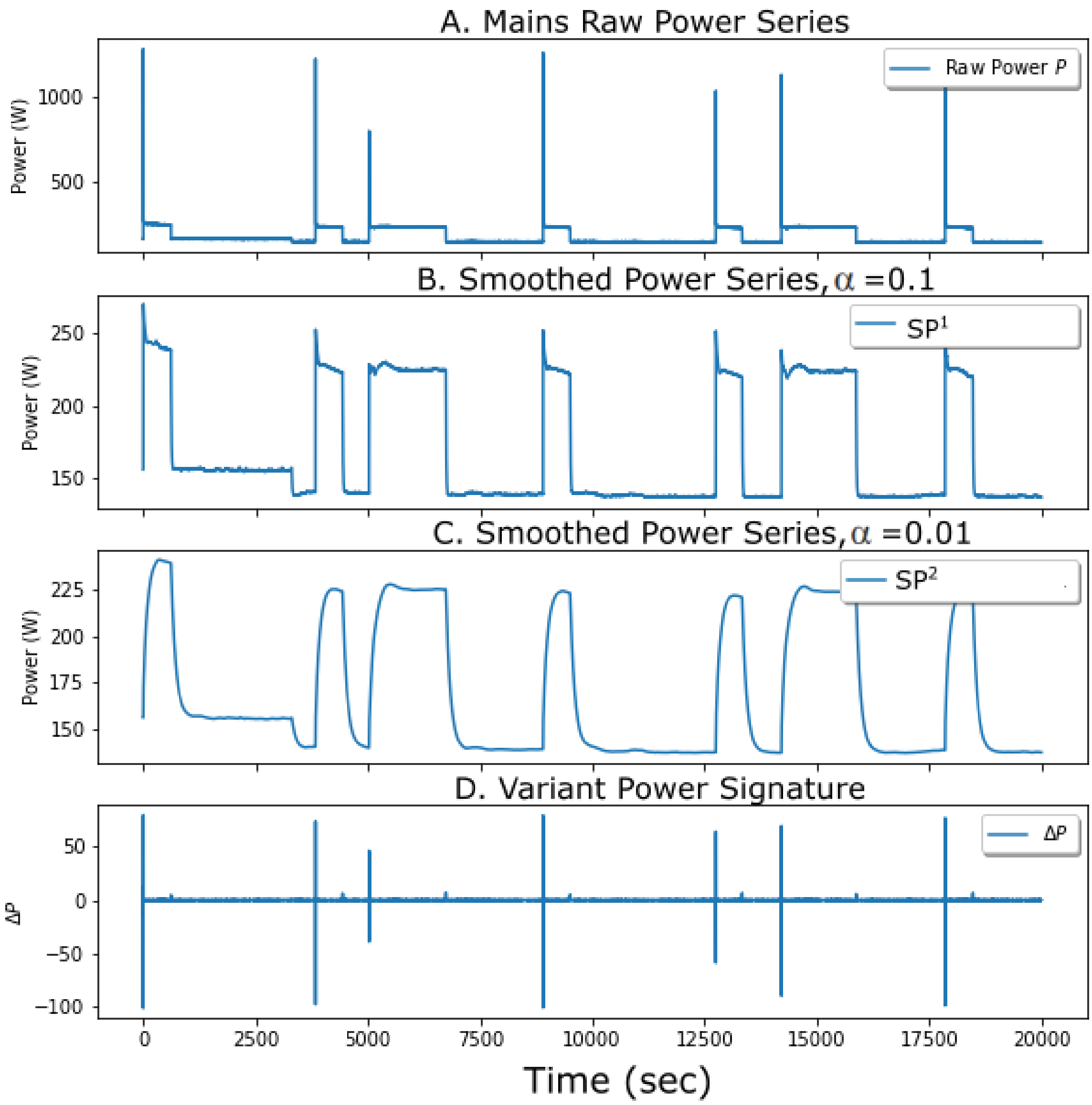
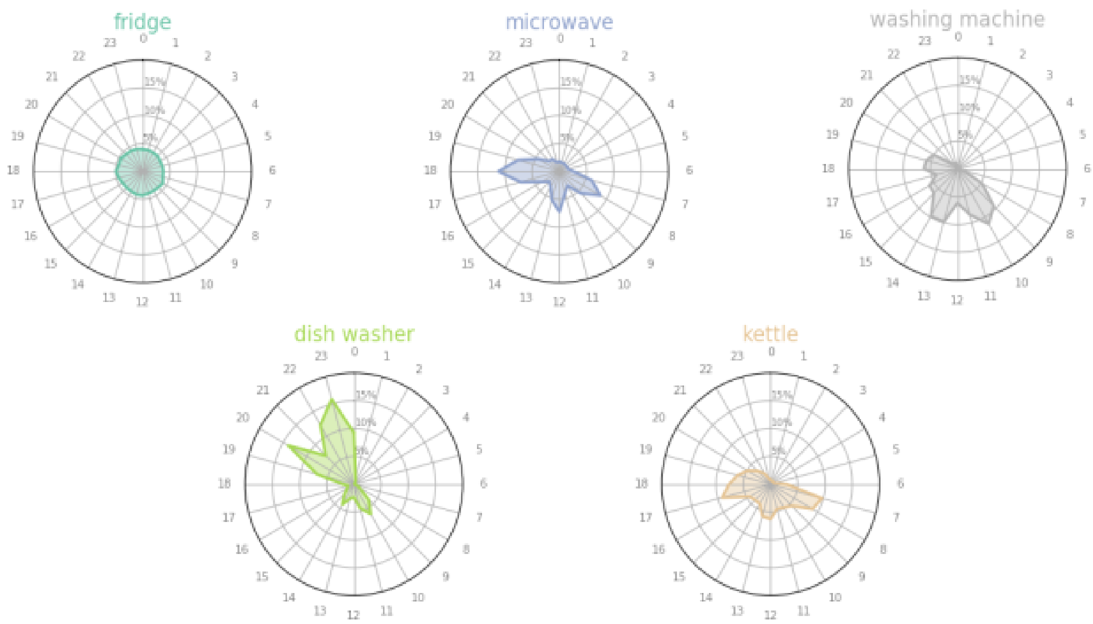


Figure 3.7: Radar chart of appliance percentage power consumption per hour of the day out of the total appliance consumption. This is based on the total monitoring period of the appliance from the five households of the study.



3.7 Time Embedding

Intuitively, one knows that certain appliances are more likely to be active at certain times of the day. To validate this intuition we visualize the percentage consumption per hour of the day relative to the total appliance consumption over the total monitoring duration. Figure 3.7 shows this temporal distribution of power over the five houses of the study across the 24 hours of the day. From this figure one can see three distinct peaks in the kettle and microwave plots, at 8am, 12pm and 6pm. These peaks are associated with meal times and reflect higher tendency of consumers to use the appliances at these times. Similarly, the dish washer shows significant peaks from 8-11pm, reflecting a general tendency to use the appliance at end of the day. An Appliance, like fridge, which runs all day long show uniform distribution with no distinct peaks.

The idea here is to use this time information as an input feature to aid the network in learning a better mapping from mains to the appliance consumption. This can be done by converting the mid-point hour time-stamp of the mains window to one-hot encoding and then augment it to the network input. Using this features the network can infer appliance consumption as a function of hour of the day.
One challenge that might arise from this approach, is the possibility of over-fitting to the residents consumption behaviors rather than the genuine appliance signature. Chapter 4 discuss some of the mitigation measures implemented to avoid this problem.

3.8 Normalization

To provide a scale-invariant data to the neural network, z-score normalization was applied on the input and target output. Mean and standard deviation were calculated from the sub-meter data of each appliance in the training set, Table 3.4. Target output is normalized for the regression network only, whereas we use a quantization method for the classification network output.

Table 3.4: Normalization parameters for each appliance network.

Appliance	**Mean**	**Standard Deviation**
Kettle	700	1000
Fridge	200	400
Washing machine	400	700
Microwave	500	800
Dish washer	700	1000

3.9 Appliance Power Quantization

Most household appliances fall under type I or II appliances. Type I is the simplest, where at any point of time, the appliance can either be on or off. Type II appliances have finite number of states with discreet consumption levels. Figure 3.8 shows the consumption pattern for a washing machine, a typical type II appliance. Despite the noisy signal one can easily identify three distinct consumption states. This observation sets the motivation to transform NILM from a regression problem into a more confined classification task. This transformation can be done by quantizing the network target output from continuous real-value space into finite, discreet consumption levels which reflect appliance states.

Researchers in signal processing domains use various quantization techniques such as rounding, binning or more complex Fourier transformation. However most of these methods focus on noise reduction, down-sampling or distortion optimization of electrical signal. In this work we adopt an "unsupervised quantization algorithm", which is based on clustering technique to serve a dual purpose of: (1) detecting the optimum number of states for an appliance, (2) converting power consumption into discreet state classes. Figure 3.9 show the workflow of the algorithm.
The quantization process starts by extracting the appliance activation from the sub-meter training data. Next, k-mean clustering is applied to the power points, considering two to five clusters, where the number of clusters denotes the distinct appliance states. Silhouette score is used to decide the the best clustering model and the optimum number of clusters, where the centroids represent the average consumption for the respective state. The final the output of the workflow is a quantizer model that can translate new power points to one of the appliance states based on Euclidean distance measure to the state centroid. For convince, we also generate a state-power

Figure 3.8: Consumption states of a washing machine

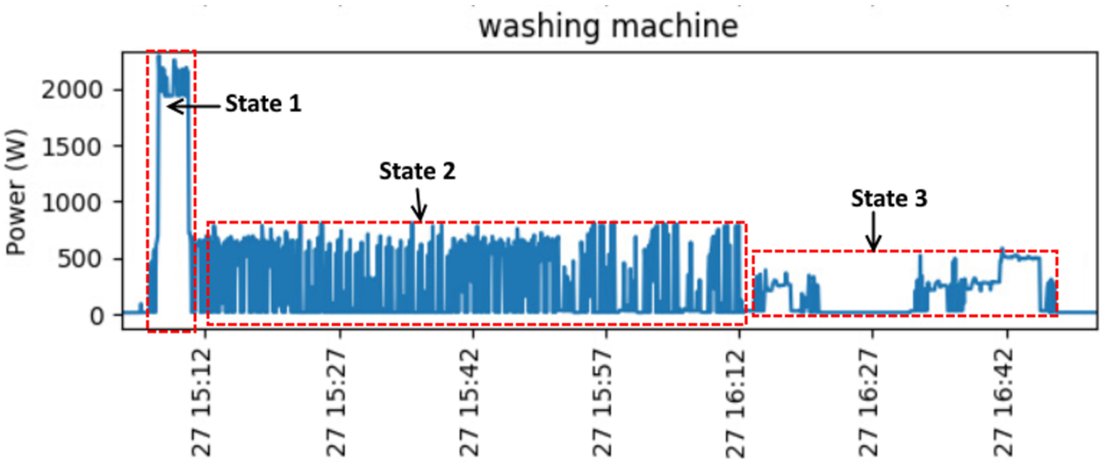
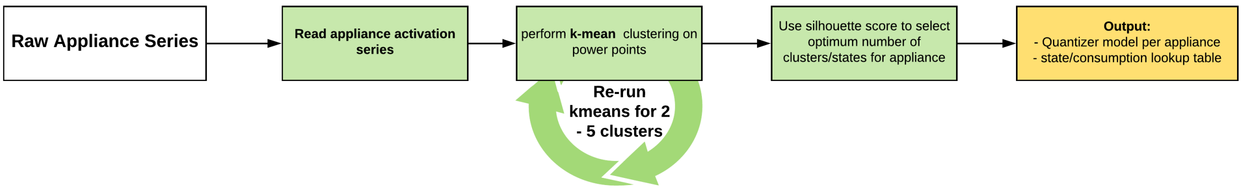


Figure 3.9: Unsupervised quantization workflow for appliance consumption.

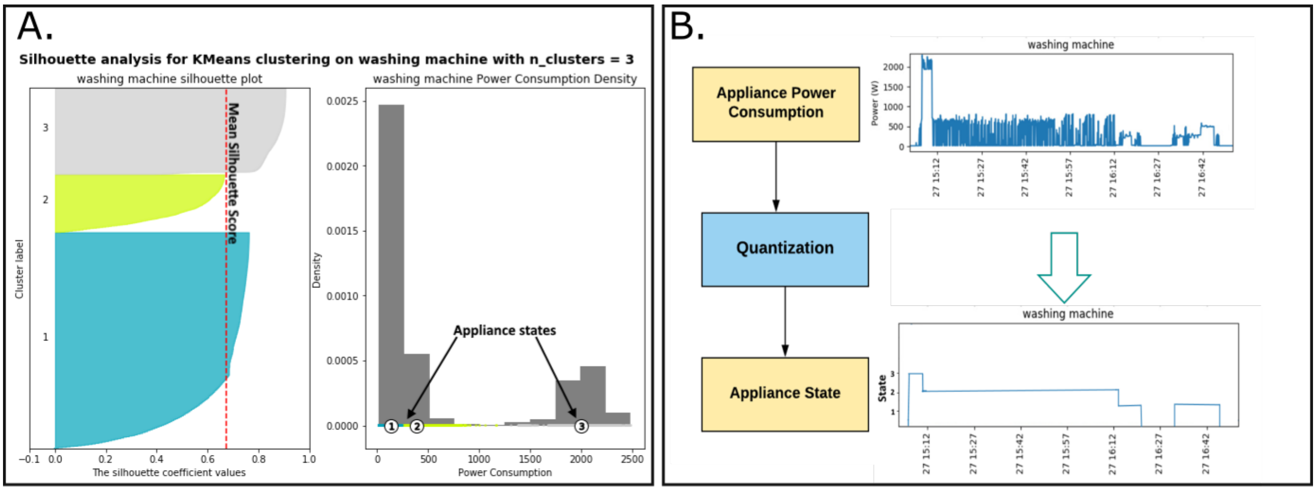


look up table for back-conversion of states to power. Figure 3.10 shows sample results of this quantization process. As will be illustrated in the next section,this quantizer model will be used as a pre-processing for the classification network target output.

3.10 Neural Network Architecture

The networks considered in this study are trained based on Sequence-to-Point mapping in comparison to the Sequence-to-Sequence method adopted in the work by (Kelly & Knottenbelt, 2015a). The main advantage in the latter that it avoids the need to ensemble or average the predictions of overlapping windows. To avoid memory limitation, the network is trained of a fixed-length window of the mains, where the window size is set to 10 minutes for all the appliances. The training set was obtained by sliding the window one step at a time on both mains and appliance consumption. For each window of the mains, the mid-point of the corresponding appliance window is considered as the target to enable the sequence-to-point mapping.

Figure 3.10: A. Silhouette analysis to decide the optimum number of states of washing machine. B. Sample quantization results on washing machine.



3.10.1 Regression Experiment

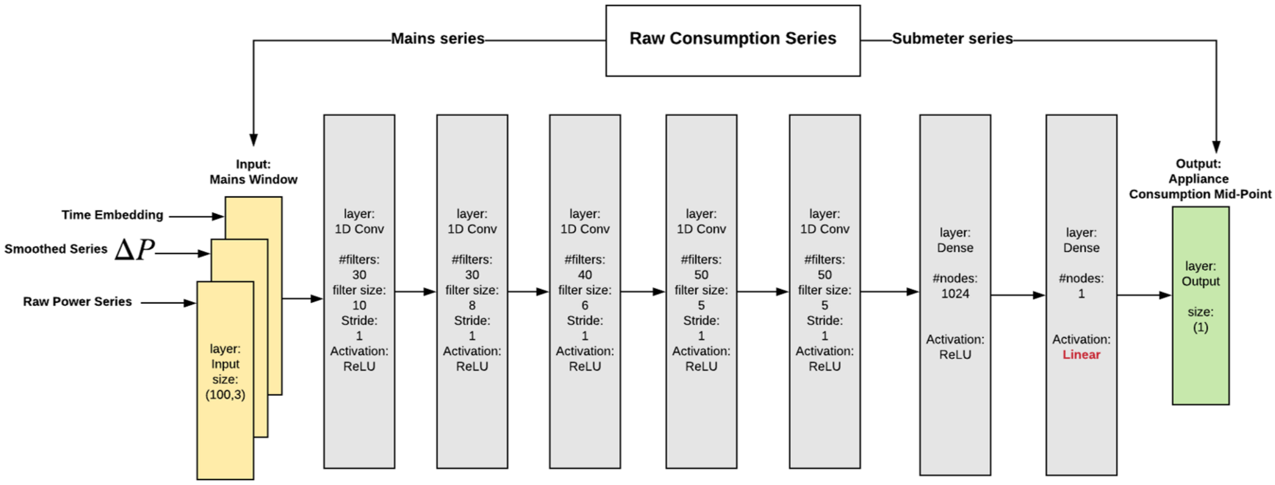
As shown in previous chapter, all previous NILM work relied on the raw mains aggregate power as the only of input for disaggregation. In conjugate fields of time series analysis and speech recognition researchers tend to augment the input with processed variant of the raw series, such as auto-correlation series or exponential smoothed series as well as many other variants. This process of augmentation has been proved to help networks to learn better features to predict the target output. Section 3.6 explains the process of using exponentially smoothed power variant (ΔP) to remove acquisition noise and provide clear signature of appliance state transition. In addition to the power variant, section 3.7 shows the process of creating an additional feature out of the mains consumption time-stamp.

Neural network architecture used for this experiment is adapted from the work of Zhang et al. (2016). Rather then using the raw power series, we consider augmenting the network input with the smoothed power variant (ΔP) and the time-embedding. Architecture details are show in figure 3.11.

3.10.2 Disaggregation Classification Experiment

All the previous studies tackled NILM disaggregation as regression task where the network is trained on MSE loss to map from a sequence of real valued input to a real valued output. Training on MSE yields a large model with non-convex error surface, which is prone to settling on local minima. Section 3.9 show that majority of the household appliance have finite number of states with discreet consumption value per state. Based on this fact, we formulate NILM into classification task where the target is to predict discreet output corresponding to appliance running state rather than a real valued consumption. To accommodate this reformulation the

Figure 3.11: Neural network architecture for the augmented regression experiment.



network requires three key changes:

1. The target output of the network need to be converted from real-value space into discreet space suitable for classification.

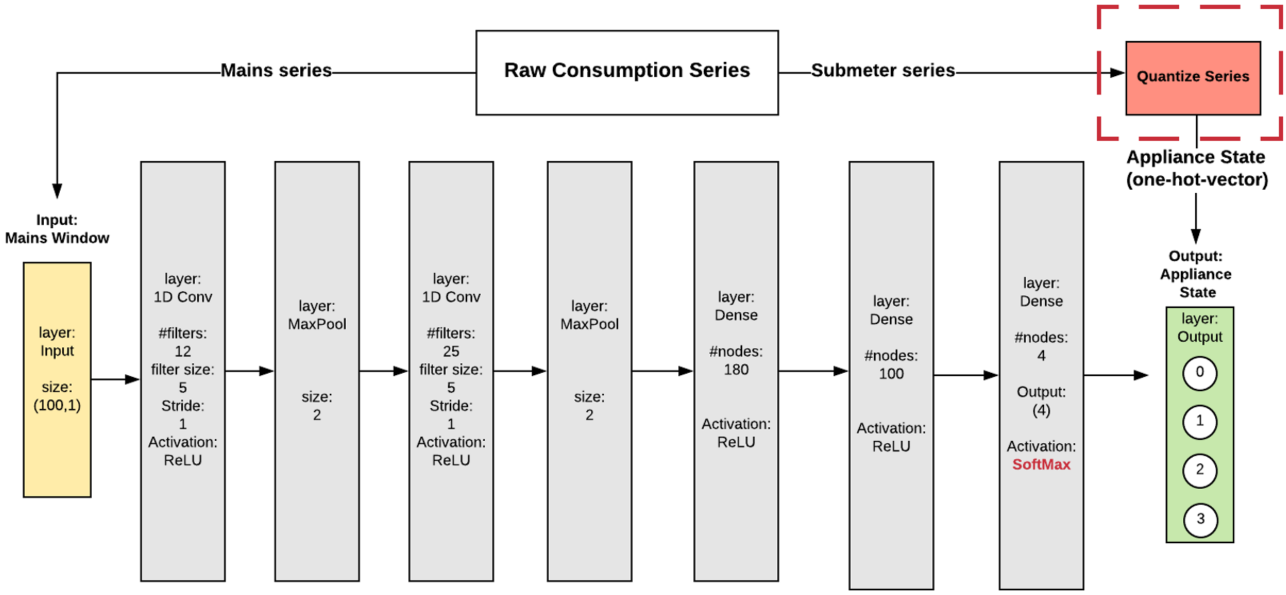
2. The last layer activation has to be changed from linear to soft-max activation, to enable detecting one of multiple possible running states.

3. Change the loss function from MSE to categorical cross-entropy which suits the classification objective.

The first requirement is attained through the quantization work-flow explained in section 3.9. To streamline the process the quantizer model is plugged as a preprocessing step of the network target output. The other two requirement are easily attained by modifying the objective function and last network layer in Keras library. As a proof-of-concept, we adapt the LenNet architecture from computer vision research. The architecture and convolutions are modified to consider 1D sequential data compared to original 2D setting for images. Figure 3.12 show schematic representation of this architecture.

During disaggregation the appliance sates are converted back to power using the state-power look-up table of the quantizer model, which enables comparing the performance with other disaggregation models.

Figure 3.12: Neural network architecture for the augmented regression experiment.



CHAPTER 4

RESULTS AND EVALUATION

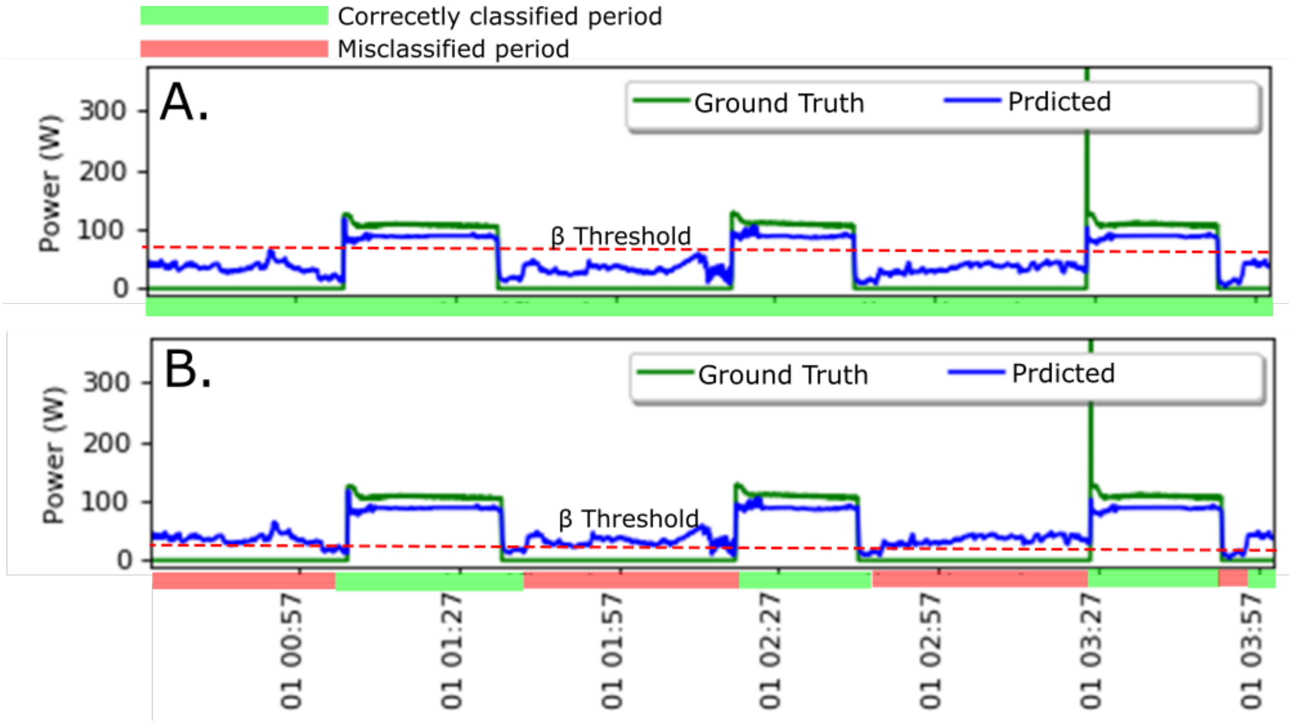
4.1 Evaluation Metrics

Early NILM studies used used binary classification metrics to evaluate the disaggregation results, considering a binary (ON/OFF) appliances scenario. Disaggregated power points get classified as positive or negative cases based on empirically selected power threshold (β). This power threshold is selected per appliance to indicate transition between on and off states. A point gets classified as positive if ($P_t > \beta$) and negative if ($P_t \leq \beta$) where P_t is the appliance predicted consumption at time point (t). Once the results are project into this binary space, standard classification metrics,such as recall, precision, accuracy and F1-score, are used to evaluate the performance. When it comes to benchmarking NILM models, this evaluation approach has to fundamental caveats:

1. The classification metrics are heavily influenced by the choice of the power threshold (β), which is selected empirically in most studies. Using two different β thresholds on the same model will yield completely different performance metrics, which make it difficult to benchmark the result across different models. Figure 4.1 illustrates this problem for disaggregated consumption of a fridge. By using high β threshold -around 75 W-, the model correctly classify the whole duration in plot A. Setting β to a lower threshold, majority of the duration get misclassified, despite using the same model and having the same results.

2. These metrics do not take into account the power difference between prediction and ground-truth consumption, hence they do not provide an accurate measure of the model performance. Figure 4.2 illustrates a case where two different models score the same results on the classification metrics, despite the obvious difference in their power approximation to the ground-truth consumption.

Figure 4.1: Sample disaggregation results of a fridge with corresponding classification outcome at x-axis. A. high power threshold cut-off, B. lower power threshold cut-off



To avoid these caveats, this project uses Mean Absolute Error (MAE) and Signal Aggregate Error (SAE) as the two metrics to evaluate models performance and benchmark the results with previous studies. MAE provides an average error measure considering the absolute difference prediction at each time point in the disaggregated series:

$$MAE = \frac{1}{T}\sum_{t=1}^{T} |\hat{P}_t - P_t| \tag{4.1}$$

where P_t is the appliance ground-truth consumption, $\hat{P}_t$ is the model prediction and T is the length of the disaggregated series.//

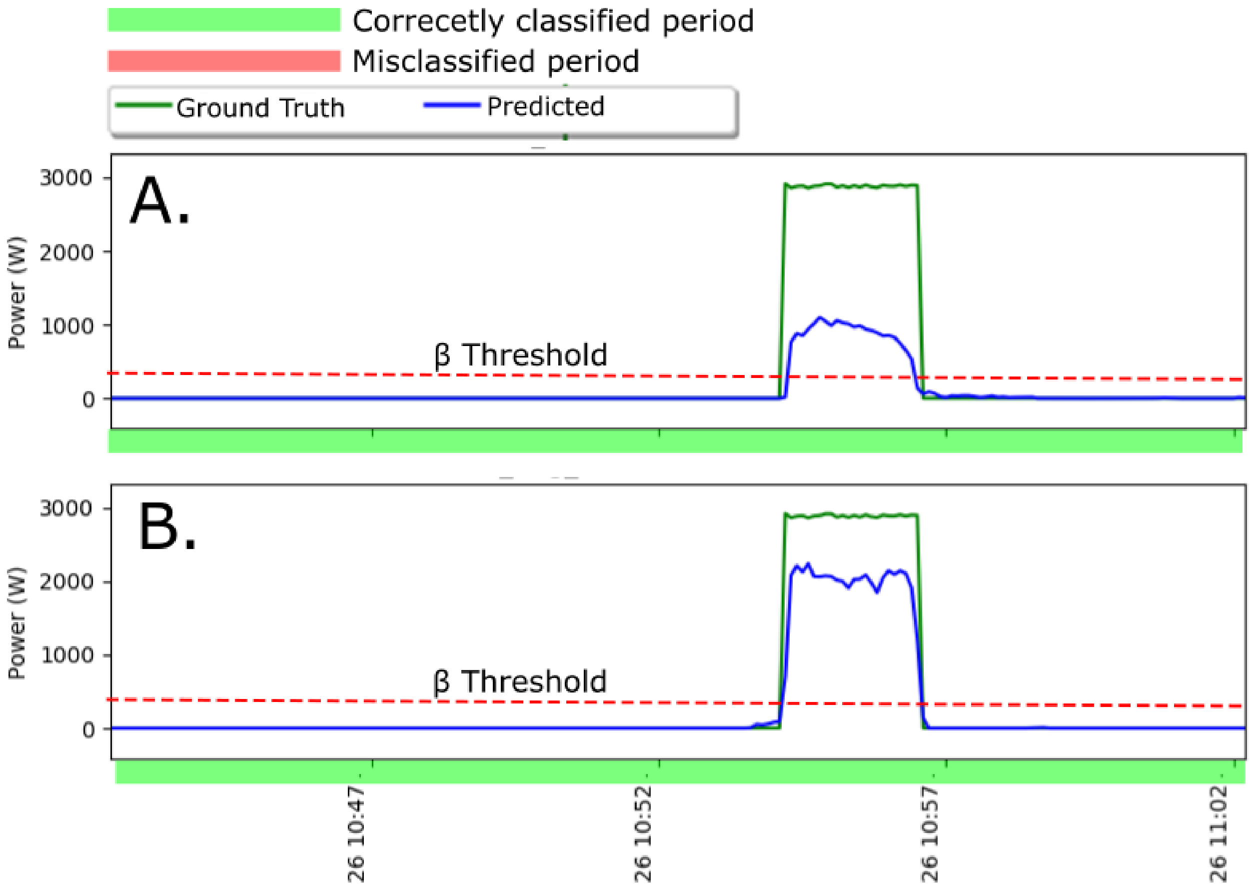
SAE is a specific metric used for NILM and it measures the overall difference ratio in consumption prediction to the actual appliance consumption:

$$SAE = \frac{|\hat{r} - r|}{r} \tag{4.2}$$

where $\hat{r}$ is the total predicted consumption ($\sum_{t=1}^{T} \hat{P}_t$) and r is the total actual consumption ($\sum_{t=1}^{T} P_t$).

These two metrics provide a consistent measure to evaluate NILM models without any discrepancy inherit from the power threshold cut-off in the classification metrics.

Figure 4.2: Sample disaggregation results of a kettle using two different models A and B. A. gives lower estimation of ground-truth consumption. B. give closer estimation to the ground-truth. Despite their estimation difference, A and B will have same accuracy, recall, precision and F1-score measure.



4.2 Experimental Results

The results of the two experiments conducted on this study are benchmarked with four NILM models, as highlighted in table 4.1. The work by Kolter and Jaakkola (2012) is based on Additive Factorial Hidden Markov Model (AFHMM), which provides a fair baseline for non-neural NILM models. Kelly and Knottenbelt (2015a) results are based on convolutional de-noising auto-encoder architecture trained as Seq-2-Seq model. The Seq-2-Seq and Seq-2-Point models by Zhang et al. (2016) are based on deep convolutional networks and represent state-of-the-art results.

Table 4.1 shows that our classification network provides the best results compared to all the other models, with respect to MAE, consistently across the five appliances. The regression network, augmented with ΔP power variant and time-embedding, performs worse than the classification network, but still better than all the other models for all appliances apart from

Table 4.1: Performance metrics of "Regression Network" and "Classification Network" experiments compared to models from previous studies. Bold numbers show best score amongst the other models.

Metric	Model	Kettle	Microwave	Fridge	Dish Washer	Washing Machine	Overall
MAE	AFHMM (Kolter et al. 2012)	47.380	21.180	42.350	199.840	103.240	82.798 ± 64.5
	Seq2Seq (Kelly et al. 2015)	13.000	14.559	38.451	237.960	163.468	93.488 ± 91.11
	Seq2Seq (Zhang et al. 2016)	9.220	13.619	24.489	32.515	10.153	17.999 ± 9.06
	Seq2Point (Zhang et al. 2016)	7.439	8.661	20.894	27.704	12.663	15.472 ± 7.72
	Regression Network	7.021	7.470	21.816	18.951	11.041	13.26 ± 6.05
	Classification Network	**6.139**	**5.015**	**20.222**	**13.959**	**7.615**	**10.59 ± 5.72**
SAE	AFHMM (Kolter et al. 2012)	1.060	1.040	0.980	4.500	8.280	3.172 ± 2.89
	Seq2Seq (Kelly et al. 2015)	0.085	1.348	0.502	4.237	13.831	4.001± 5.12
	Seq2Seq (Zhang et al. 2016)	0.309	**0.205**	0.373	0.779	0.453	0.424 ± 0.2
	Seq2Point (Zhang et al. 2016)	0.069	0.486	**0.121**	0.645	0.284	0.321± 0.22
	Regression Network	0.078	0.450	0.210	0.183	**0.210**	0.226 ± 0.12
	Classification Network	**0.066**	0.280	0.145	**0.150**	0.368	**0.202 ± 0.11**

the fridge. Overall, the classification network provides 31% improvement on MAE compared to state-of-the-art Seq-2-Point model by Zhang et al. (2016). In terms of SAE, the classification network outperformed the Zhang et al. (2016) Seq-2-Point models for the kettle and dish washer. The regression network provided the best SAE results for the dish washer, whereas Seq-2-Point and Seq-2-Seq gave the best results for the microwave and fridge, respectively. Overall, the classification network improves SAE by 37% compared to state-of-the-art model.

CHAPTER 5

Discussion and Conclusion

5.1 Discussion

The benchmarking results in the previous chapter showed that the two experiments conducted in this study have outperformed stat-of-the-art models for most of the appliance. The improvement in the regression network is attributed to the augmented features of the power variant and time-embedding, where the rest of the hidden layer are identical to the Seq-2-Point model by Zhang et al. (2016) . Appliances that show skewed temporal distribution, figure 3.7, showed significant improvement in MAE, such as the microwave and dishwasher. On the contrary, appliances with uniform temporal distribution, like the fridge, seemed to not gain much benefit from the time-embedding. In fact, the network tries to leverage the time-embedding by learning posterior probability of $p(P|t)$, where P is the appliance consumption and t is hour of the day. In the case of fridge, where P is independent, uniformly distributed variable, there is no dependency with time and hence no additional information gain from time-embedding.

For the case of the classification network, the improvement was attained by simplifying the learning from regression with complex MSE objective function into a simpler classification problem with cross-entropy objective function. This modification also enable using simpler, shallow network architecture without negative impact on the performance. Table 5.1 show the number of trained parameter for the regression and classification experiments.

Table 5.1: Number of trained parameters on each experiment network architecture.

Architecture	Number of Trained Parameters
Regression Network	5,159,449
Classification Network	119,180

Since the classifier experiment involves quantizing the appliance consumption and projecting it into a confined, finite space, the network is very prone to over-fitting to the appliances seen during training. Two measure are applied to mitigate this risk:

1. Dropout layers are added to the architecture, which provides a mean for regularization to prevent over-fitting.

2. The networks are tested on appliance from different houses than the ones used for training, which ensures that the network is robust and generalizes well for unseen appliances.

Figure XX, show disaggregation results for the five appliances, considering (Zhang et al., 2016) Seq-2-Point model, regression and classification experiments. Each of the models is trained for 50 epochs.

5.2 Project Impact

The main driver for this research is the forecasted imbalance between electric demand and supply due to the reliance on rapidly depleting fossil fuels for electricity generation. This situation is expected to exacerbate in the next 30 years, with electric consumption predicted to double (Larcher & Tarascon, 2015). This work will help to alleviate some of these challenges and will have significant impact from various perspectives:

1. From consumer perspective, disaggregated household consumption could be used to promote conservation through itemized energy bills. A recent study showed that consumers reduced their energy consumption by an average 7.7%, based on appliance-consumption feedback (Bidgely, 2015).

2. From environmental perspective, energy saving translates to indirect reduction in CO2 emission, through reduction of fossil fuels used for electricity generation. In the UK for example, where electricity generation accounts for 34% of CO2 emissions, a saving of 1GWh corresponds a reduction of 560 tonnes in CO2 emissions (DECC, 2016).

3. From utility perspective, disaggregation could be used to highlight electricity usage patterns and consequently facilitate proactive planning of generation capacity to meet with consumers demand. For example, disaggregation of heating and cooling appliance will help utilities to forecast seasonal changes in demand patterns. From a daily pattern perspective, this enable utilities to tailor tariffs based on time of day consumption to minimize demand peaks. In Victoria, a regulation was issued in December 2014 mandating utilities to replace all traditional meters with smart meters and hence customers were provided with better electricity pricing schemes based on time of day usage (DED, 2015).

5.3 Conclusion and Future Work

This project adopted two deep convolutional neural network (CNN)architecture to improve NILM performance. The first architecture was formulated as regression network where the input was augmented with smoothed power variant and time-embedding to improve the network learning. Experiment based on this architecture showed that network can learn better prediction using time-embedding for appliance that has time-dependent consumption. Second architecture was based on a modified setting of the LeNet architecture from computer vision domain and was formulated to address NILM as classification task. To achieve this reformulation we developed an unsupervised quantization algorithm, capable of detecting appliance states with their respective power consumption. This architecture outperformed state-of-the-art results providing an overall 31% and 37% improvement on MAE and SAE, respectively.

The experiments conducted in this project were limited to five household appliance, since it was intended as a proof-of-concept only. To deploy these models in real-life, one has to consider training more networks on wide variety of appliances. To do this we consider using a larger dataset such as the "Pecan St" data, which provides consumption from over 700 household. Such large dataset enables training more robust networks that are able to generalize to wide range of appliances,households and consumption patterns.

References

Anderson, K., Ocneanu, A., Benitez, D., Carlson, D., Rowe, A., & Berges, M. (2012, August). BLUED: A Fully Labeled Public Dataset for Event-Based Non-Intrusive Load Monitoring Research. In *Proceedings of the 2nd KDD Workshop on Data Mining Applications in Sustainability (SustKDD).* Beijing, China.

Baranski, M., & Voss, J. (2003, June). Nonintrusive appliance load monitoring based on an optical sensor. In *2003 IEEE Bologna Power Tech Conference Proceedings,* (Vol. 4, pp. 8 pp. Vol.4–). doi: 10.1109/PTC.2003.1304732

Bidgely. (2015, April). *PG&E Pilot Yields 7.7% Energy Savings.* Retrieved 2018-04-10TZ, from `http://www.bidgely.com/blog/pge-pilot-yields-7-7-energy-savings/`

Cho, K., Merrienboer, B. v., Gulcehre, C., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *CoRR*, *abs/1406.1078*. Retrieved from `http://arxiv.org/abs/1406.1078`

DECC. (2016). *Provisional estimates of UK Greenhouse Gas emissions for 2015, including quarterly emissions for 4th quarter 2015* (Technical Report). UK: UK Government. Retrieved 2018-04-10TZ, from `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/511684/20160331_2015_Provisional_Emissions_Statistics.pdf`

DED. (2015). *Smart meter installation* (Technical Report). Australia, Victoria: Department of Economic Development, Jobs, Transport and Resources. Retrieved 2018-04-10TZ, from `http://www.smartmeters.vic.gov.au/installation`

Faustine, A., Mvungi, N. H., Kaijage, S., & Michael, K. (2017, March). A Survey on Non-Intrusive Load Monitoring Methodies and Techniques for Energy Disaggregation Problem. *arXiv:1703.00785 [cs].* Retrieved from `http://arxiv.org/abs/1703.00785` (arXiv: 1703.00785)

Grais, E. M., Sen, M. U., & Erdogan, H. (2013, November). Deep neural networks for single channel source separation. *arXiv:1311.2746 [cs].* Retrieved 2018-10-26TZ, from `http://arxiv.org/abs/1311.2746` (arXiv: 1311.2746)

Graves, A., Mohamed, A.-r., & Hinton, G. E. (2013). Speech Recognition with Deep Recurrent Neural Networks. *CoRR*, *abs/1303.5778*. Retrieved from `http://arxiv.org/abs/1303.5778`

Hansun, S. (2013, November). A new approach of moving average method in time series analysis. In *2013 Conference on New Media Studies (CoNMedia)* (pp. 1–4). doi: 10.1109/CoNMedia.2013.6708545

Hart, G. W. (1992, December). Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, *80*(12), 1870–1891. doi: 10.1109/5.192069

IEA. (2017a). *Key World Energy Statistics* (Technical Report). Paris: International Energy Agency. Retrieved 2018-08-04, from https://www.iea.org/publications/freepublications/publication/KeyWorld2017.pdf

IEA. (2017b). *World Energy Outlook 2017* (Technical Report). Paris: International Energy Agency. Retrieved 2018-10-04, from http://www.iea.org/media/weowebsite/2017/Chap1_WEO2017.pdf

Kelly, J., & Knottenbelt, W. (2015a). Neural NILM: Deep Neural Networks Applied to Energy Disaggregation. In (pp. 55–64). ACM Press. Retrieved 2018-03-12TZ, from http://dl.acm.org/citation.cfm?doid=2821650.2821672 doi: 10.1145/2821650.2821672

Kelly, J., & Knottenbelt, W. (2015b, March). The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. *Scientific Data*, *2*, 150007. Retrieved 2018-05-31TZ, from http://arxiv.org/abs/1404.0284 (arXiv: 1404.0284) doi: 10.1038/sdata.2015.7

Kim, H., Marwah, M., Arlitt, M., Lyon, G., & Han, J. (2011, April). Unsupervised Disaggregation of Low Frequency Power Measurements. In *Proceedings of the 2011 SIAM International Conference on Data Mining* (pp. 747–758). Society for Industrial and Applied Mathematics. Retrieved 2018-05-31TZ, from https://epubs.siam.org/doi/abs/10.1137/1.9781611972818.64 doi: 10.1137/1.9781611972818.64

Kim, J., Le, T.-T.-H., & Kim, H. (2017). Nonintrusive Load Monitoring Based on Advanced Deep Learning and Novel Signature. *Computational Intelligence and Neuroscience*, *2017*, 1–22. Retrieved 2018-03-13TZ, from https://www.hindawi.com/journals/cin/2017/4216281/ doi: 10.1155/2017/4216281

Kolter, J. Z., & Jaakkola, T. (2012, March). Approximate Inference in Additive Factorial HMMs with Application to Energy Disaggregation. In *Artificial Intelligence and Statistics* (pp. 1472–1482). Retrieved 2018-03-14TZ, from http://proceedings.mlr.press/v22/zico12.html

Kolter, J. Z., & Johnson, M. J. (2011). REDD: A Public Data Set for Energy Disaggregation Research. *In proceedings of the SustKDD workshop on Data Mining Applications in Sustainability*, 6.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 25* (pp. 1097–1105). Curran Associates, Inc. Retrieved from http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

Krystalakos, O., Nalmpantis, C., & Vrakas, D. (2018). Sliding Window Approach for Online Energy Disaggregation Using Artificial Neural Networks. In *Proceedings of the 10th Hellenic Conference on Artificial Intelligence* (pp. 7:1–7:6). New York, NY, USA: ACM. Retrieved 2018-08-23TZ, from `http://doi.acm.org/10.1145/3200947.3201011` doi: 10.1145/3200947.3201011

Larcher, D., & Tarascon, J.-M. (2015, January). Towards greener and more sustainable batteries for electrical energy storage. *Nature Chemistry*, *7*(1), 19–29. doi: 10.1038/nchem.2085

Marques, P. P. (2016). *Applications of Deep Learning techniques on NILM* (PhD Thesis). Universidade Federal do Rio de Janeiro.

Parson, O., Fisher, G., Hersey, A., Batra, N., Kelly, J., Singh, A., ... Rogers, A. (2015, December). Dataport and NILMTK: A building data set designed for non-intrusive load monitoring. In *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)* (pp. 210–214). doi: 10.1109/GlobalSIP.2015.7418187

Parson, O., Ghosh, S., Weal, M., & Rogers, A. (2012, July). Non-Intrusive Load Monitoring Using Prior Models of General Appliance Types. In *Twenty-Sixth AAAI Conference on Artificial Intelligence.* Retrieved 2018-05-31TZ, from `https://www.aaai.org/ocs/index.php/AAAI/AAAI12/paper/view/4809`

Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P.-A. (2008). Extracting and Composing Robust Features with Denoising Autoencoders. In *Proceedings of the 25th International Conference on Machine Learning* (pp. 1096–1103). New York, NY, USA: ACM. Retrieved 2018-06-06TZ, from `http://doi.acm.org/10.1145/1390156.1390294` doi: 10.1145/1390156.1390294

Wichakool, W., Avestruz, A. T., Cox, R. W., & Leeb, S. B. (2009, December). Modeling and Estimating Current Harmonics of Variable Electronic Loads. *IEEE Transactions on Power Electronics*, *24*(12), 2803–2811. doi: 10.1109/TPEL.2009.2029231

Zeifman, M., & Roth, K. (2011, January). Nonintrusive appliance load monitoring: Review and outlook. In *2011 IEEE International Conference on Consumer Electronics (ICCE)* (pp. 239–240). doi: 10.1109/ICCE.2011.5722560

Zhang, C. (2016). *A Deep Learning Approach to Identifying Household Appliances from Electricity Data* (PhD Thesis). University of Edinburgh.

Zhang, C., Zhong, M., Wang, Z., Goddard, N., & Sutton, C. (2016, December). Sequence-to-point learning with neural networks for nonintrusive load monitoring. *arXiv:1612.09106 [cs, stat]*. Retrieved 2018-08-23TZ, from `http://arxiv.org/abs/1612.09106` (arXiv: 1612.09106)

Zoha, A., Gluhak, A., Imran, M. A., & Rajasegarar, S. (2012, December). Non-Intrusive Load Monitoring Approaches for Disaggregated Energy Sensing: A Survey. *Sensors*, *12*(12), 16838–16866. Retrieved 2018-03-15TZ, from `http://www.mdpi.com/1424-8220/12/12/16838` doi: 10.3390/s121216838

Zoha, A., Gluhak, A., Nati, M., & Imran, M. A. (2013, April). Low-power appliance monitoring using Factorial Hidden Markov Models. In *2013 IEEE Eighth International Conference on Intelligent Sensors, Sensor Networks and Information Processing* (pp. 527–532). doi: 10.1109/ISSNIP.2013.6529845