ASIA-PACIFIC
SOLAR RESEARCH CONFERENCE

Bruce Ho

Exploration of K-means Neural Network for Residential Load Profile Decomposition by Appliance

Bruce Ho[1], Anna Bruce[1,2], Navid Haghdadi[2,3]

[1]*School of Photovoltaic and Renewable Energy Engineering, UNSW, Sydney NSW 2052, Australia*

[2]*Centre for Energy and Environmental Markets, UNSW, Sydney NSW 2052, Australia*

[3]*School of Electrical Engineering and Telecommunications, UNSW, Sydney NSW 2052, Australia*

E-mail: bruce.ho98@gmail.com

Abstract

The operation of electricity industries around the world are currently undergoing significant changes as we transition to higher penetrations of renewable generation in our energy mix. Renewables introduce volatility to our electricity supply and therefore system flexibility is becoming increasingly important in helping to balance demand and supply over different timeframes. Currently, high barriers of entry exist in the residential sector to effectively participate in DR. Methods for estimating flexible appliance load profiles in the residential sector will be valuable for estimating the potential value of different residential DR resources and for informing participation by aggregators of residential DR. This may therefore help to reduce the barriers for capturing the DR potential within the residential sector.

This paper explores the potential of utilising a hybrid K-means Neural Network (NN) approach to decompose an aggregated residential load profile into separate appliance load profiles. Four different appliances are the focus of this study, including air conditioning (AC), hot water (HW), dryer and washing machine loads. A K-means NN approach is expected to improve results compared to other popular machine learning (ML) models by separating timeperiods that exhibit similar characteristics (e.g. cold or hot days, high or low demand) and training those timeperiods on separate NNs. The separate NNs can therefore learn the appliance operating characteristics under the different operating circumstances without needing to generalise its learning for all operating circumstances. The accuracy of K-means NN are evaluated against a Neural Network, Decision Tree and Support Vector Regression (SVR) for residential load decomposition using half-hourly appliance level data from ninety-two households collected over 6 months (between August 2013 and February 2014) during the Ausgrid Smart Grid Smart City trial. It is found that NN was the best performing model for decomposing AC, hot water, washing machine and dryer loads with a normalized root mean square error (NRMSE) of 4.15%, 6.82%, 8.75% and 9.29% respectively. K-means NN performed the second best with a NRMSE of 4.53%, 7.01%, 8.75% and 9.35% respectively. Further work is required using better data sets to improve the accuracy of the load decomposition and determine the value of a hybrid K-means NN approach.

1. Introduction

Demand response (DR), defined as modifying demand in response to price, incentives, or signals from the system operator (US - DoE, 2006) can provide system flexibility over multiple timescales. In the short-term (second to minutes), DR can provide frequency management services and contingency reserves. On the scale of minutes to hours, DR can be used for load shifting, curtailment for network congestion management or wholesale market arbitrage. Current barriers to residential DR include the cost of large-scale deployment of smart grid technologies, and a limited understanding of load behaviours of different household appliances that can effectively participate in DR, and therefore the DR resource available across different timeframes.

In Australia the residential sector contributes up to approximately 50% of peak demand in all state regions in the NEM (AEMC & EY, 2011). Generally the system peak demand occurs on extremely hot summer days with a large number of AC systems operating across a significant number of residential households (SA.GOV, 2019). The residential sector also contributes significantly to increased requirement for generator ramping over 1-2 hours in the morning and evening, and potential overgeneration during the middle of the day due to high penetration of rooftop photovoltaics (PV) installations (AEMO, 2018). This is also known as the "duck curve" phenomenon. These problems might be alleviated if DR potential within the residential sector can be effectively captured, and can work in conjunction with DR procured from other sectors as well.

To effectively utilise the DR potential within the residential sector, it is important to understand the characteristics of the underlying appliance load profiles which are often unknown due to limited metering of individual appliance loads. Since different appliances have different aggregated load profiles and different flexibility to offer DR across different timeframes. A better understanding of the appliance load profiles are important for utilities, aggregators and network operators to understand the DR potential. Furthermore, it may also allow for better forecasting of expected DR participation, resource planning and setting baselines to determine the amount of demand response provided when operating in a market.

Previous researchers have tried to estimate individual household appliance loads by monitoring the aggregate home level load data and processing its signal. This approach in identifying appliances without actually monitoring them is called Non-intrusive Load Monitoring (NILM) and has been validated in previous studies (Zhuang, Shahidehpour and Li, 2018). The method proposed in this paper has similar characteristics to (NILM) methods, in that it applies machine learning methods to household load profiles to discover appliance load profiles. There are however significant differences between the method's proposed in this paper compared to the literature in NILM. Firstly, the method presented in this paper uses a low resolution data of 30 minutes, applied only to load profiles aggregated over many households. On the other hand, NILM techniques are often focused on individual household loads using much higher resolution data. The use of high-resolution data enables NILM methods to characterise and detect the different signals of appliance loads related to their duty cycles, power draw when turning on/off, and patterns during operation. This is not possible with lower resolution data.

As residential DR matures, in the short- to medium-term there are likely to be increasing numbers of residential customers with smart grid technologies that enable the metering of key DR appliances. Such a sample of data would enable the model described in this paper to be extended beyond planning purposes, to providing near real-time estimates of the aggregated appliance load profiles within a selected region for operational purposes such as deploying DR.

In section 2 of this paper, a brief review of machine learning methods applied to load disaggregation is provided. Section 3 describes the methods used in this study, including the data used, data cleaning, feature selection, and the application of the ML models. Section 4 presents the results and discussion of the K-means NN, NN, Decision Tree and Support Vector Regression models for residential load decomposition of AC, Hot water, Dryer and Washing Machine loads.

2. Background

2.1. Brief Review of Machine Learning, K-means and Neural Network

ML models such as Neural Networks (NN) are particularly well suited for regression or pattern recognition applications as they can capture non-linearities like human behaviour in response to a range of external factors (i.e. effect of temperature on AC usage). Their ability to "learn" or recognise patterns from data without using explicit instructions means that they can be more cost effective to deploy while achieving similar (or even better) results than a conventional algorithm that are explicitly programmed. ML models are also well-suited for dealing with the big data that comes with a smart grid future. There are various ML algorithms that can be applied for this application, such as Quantile Regression Forest (QRF) and Support Vector Regression (SVR), however the focus of this paper will be on a combination of two algorithms, K-means and NN.

K-means is an unsupervised algorithm that groups similar data points together based off the distance between the points (Al-Masri, 2015). The user defines the "k" number of unique clusters for the algorithm to solve for. A centroid is created for each unique cluster, which seeks to minimise the distance between the centroid and the data points. K-means is used to cluster similar time periods so that each cluster can be trained on separate a NN (this will be detailed in the method).

Neural network (NN) is the algorithm used to classify the appliance load profiles from the aggregated residential load profile using correlated parameters such as time of day. NN consist of multi-layer of interconnected neurons consisting of one input layer for each of the parameters used, one or more hidden layers and an output layer for the result of the model. Each neuron has an activation function which processes the input signal before passing it onto the next neuron down the chain. The input from one neuron to the next also gets multiplied by a weight which determines contribution of the previous neuron. This in turn determines its influence on the output signal or result. Hence, to train a NN, a cost function is applied to the output to obtain a measure of the error (actual vs predicted result). The error is passed backwards (known as backpropagation) through the network to adjust the weighting and minimise the output error.

3. Method

The method consists of 6 main steps as shown in *Figure 1*. Once the data has been obtained, a data quality assessment and data treatment (e.g. data filtering, cleaning and filling) is performed. The most relevant variables (or features) from the data are then selected. Data preparation for the ML model is then performed, which includes separating the training and testing data sets for evaluating the model, as well as any data normalisation. The final steps in the methodology includes building the ML model, and evaluating the ML model with the test data set.

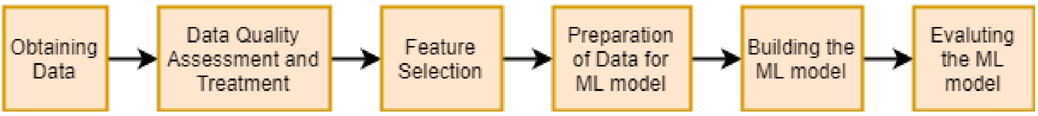


Figure 1 – Summary of the main process steps in the method

3.1. Summary of Data Utilised

The Smart Grid Smart City (SGSC) was part of a 4-year initiative that ran in 2010-2014 to study the financial feasibility and application of smart grid technologies in Australia. This trial was located across the Ausgrid network in New South Wales (NSW), and was designed to be representative of the variations in geography, climate, customer demographic and electricity network characteristics across Australia. The project involved over 17,000 electricity customers in 6 different geographical regions including Greater Newcastle area, Sydney CBD, Ku-ring-gai area, Newington area, Scone area and Nelson Bay area (Ausgrid and AEFI, 2014).

The (SGSC) data set was used to perform the study in this paper. This included 3 different data sets being utilised: The Home Area Network (HAN) Plug, the Customer Household Data, and the half-hourly Smart Meter Interval Reading made available on data.gov.au. Half-hourly Bureau of Meteorology (BoM) weather data recorded at Sydney Observatory Hill between 2013 to 2014 was used to provide variables for training the ML model. The key weather variables utilised included dry-bulb temperature (°C), relative humidity (%) and precipitation. Due to the geographical spread of customers in the SGSC trial, the weather parameters used to train the model may not be representative of the actual weather conditions each customer was subjected to, and hence may add to the error in the proposed model.

Table 1 – Description of Data used in this Study

Data	Description
Home Area Network (HAN) Plug	• Appliance measurements of 808 customers • In cumulative half-hourly kWh consumption format
Customer Household Data	• From survey of customers participating in the SGSC trial. • **Includes data on:** Location, presence of PV system, income group (high/med/low), gas and electricity usage group (high/med/low) and appliance ownership.
30min Smart Meter Interval Reading	• Total half-hourly consumption recordings of households. • Includes general load and controlled load data.
Weather Data	• BoM weather data recorded at Sydney Observatory Hill • Data used is between years 2013 – 2014; the same date range as the available data for SGSC data sets.

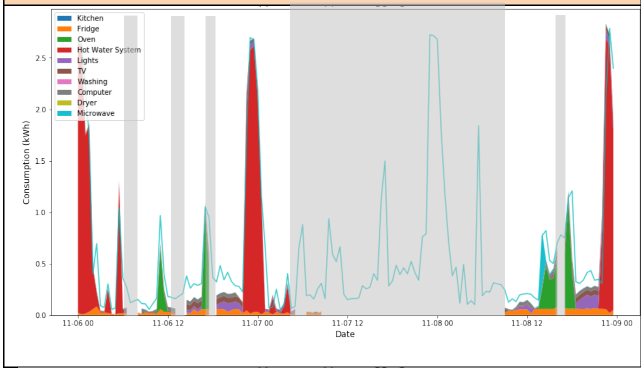
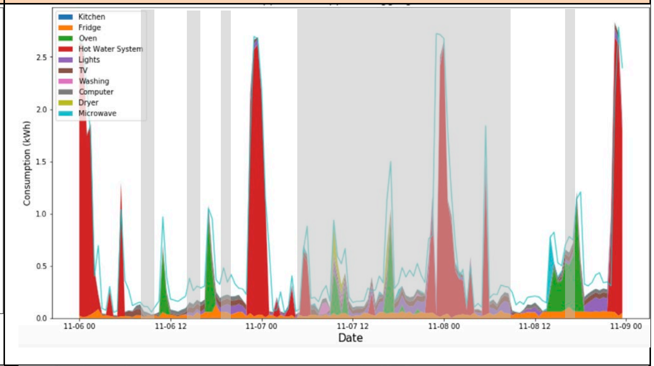
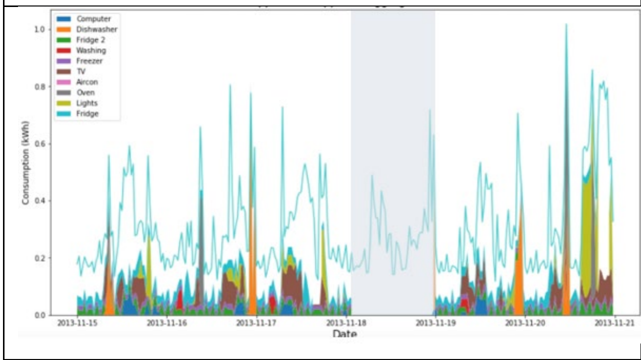
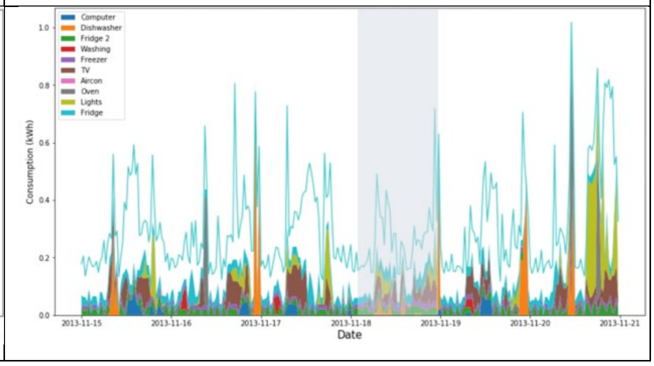
3.2. Data Cleaning, Filtering and Filling of the Smart Grid Smart City Household Data

Data cleaning and filling was applied to the Home Area Network (HAN) Plug Readings data, which contains the appliance load measurements. This was due to a large amount of low-quality data points flagged (e.g. unfeasible consumption values), as well as a significant amount of missing data for some appliances (observed to be more than 50% for some). The process of treating the data is important to enable a larger percentage of the already limited amount of data to be used to train the model. In addition, it removes the invalid data points that could negatively influence the model during training.

The data cleaning and filling was done using python data processing packages. The raw and cleaned data was compared to assure it was not significantly altered in the data processing. A summary of data preparation steps is outlined below:

1. Fill all missing values where the previous non-missing value is equal to the next non-missing value before converting the cumulative HAN data set to kWh per half-hour.
2. Remove all data sets with more than 50% missing data points.
3. Apply physical limits to appliance loads and remove data points outside of these physical limits. This includes:
   - loads must be greater or equal to 0.
   - Less than maximum load (4.61kWh / 30mins) from 30-minute interval data.
   - Each appliance also have physical limits applied based off Ausgrid's average appliance usage guide (Ausgrid, 2015), with an buffer of 30% above the average usage to avoid losing valid data.
4. Remove all data sets with more than 30% of missing values.
5. Fill in missing data using K-nearest neighbors (KNN) (with k = 3) data filling technique from the python package fancyimpute. KNN from fancyimpute takes in the different columns as variables (i.e. the different appliance data) to fill in the missing data. The smart metered load profile (shown by the light blue line in Table 1) was also included as a variable. KNN can make good approximations for the missing data point by observing the value of points closest to it, based on other variables.
6. Remove any dataset that does not have recordings earlier than 01/08/2013 and extends past 01/02/2014. This is done due to recording periods not the same for each customer. The dates have been selected to maximise the number of customers kept for analysis.

Table 2 – Example Comparison of Data with and without Data Filling with KNN for individual households (The total consumption is given by the Teal Blue line).

Before Filling Data with KNN	**After Filling Data with KNN**
	
	

After removing the low quality data points and performing data filtering, the number of customers were reduced from 808 to only 92 customers. A summary of these 92 customer details are shown in *Figure 1*.

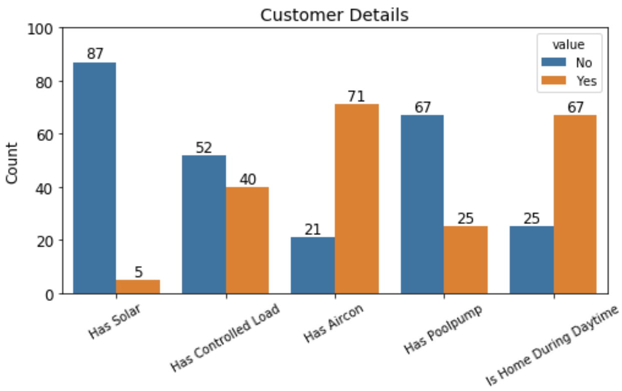


Figure 2 – Retained customer characteristics after filtering, based on SGSC survey data (Has Solar refers to having solar thermal).

For load disaggregation, the total household usage is required to be known (Adabi *et al.*, 2015). However, for households with solar net metering, an additional meter is required to measure total solar generation so that the total household usage can be calculated. Analysis of the 92 customers showed that 1 customer had PV generation. The customer with PV generation was included in the analysis as it provided valuable information on appliance loads while having minimal impact on the aggregated household load.

3.3. *Feature Selection*

Temporal and weather data are typical features that impact appliance operation and are therefore used in regression models for load forecasting (Petrican *et al.*, 2018), (Su, Xu and Tang, 2018). The same regression models and hence similar features can also be used for decomposing load profiles, and are applied to the model in this paper. These features include:

1) **Load data:** aggregated general and controlled loads of the households,
2) **Climate data:** dry bulb air temperature, relative humidity, precipitation, daylight hours
3) **Datetime:** hour, day and month
4) **Day type:** weekend/weekday, public holiday and school holiday.

Datetime (hour, day and months) values are one-hot encoded into binary values of 1s and 0s by separating each unique hour, day and month into its individual column. This results in 67 unique features (24 for each hour, 12 for each month, and 31 for each day).

The importance of each feature can be determined using Mutual_info_regression from sklearn (python's ML package). Mutual_info_regression estimates mutual information between the features and the target variable utilising KNN to estimate the entropy of the data. This analysis excludes temporal variables such as hour, day and month as these variables have been one-hot encoded.

3.4. *Preparation of Data for Training*

For each of the four different appliances (AC, dryer, washing machine and hot water), only the households with the selected appliance metered out of the ninety two were included for training and testing of the ML model (summarised in **Table 3**).

Table 3 – Number of Households used for training/testing the Models for Different Appliance Types.

Appliance	AC	Dryer	Washing Machine	Hot Water
Number of Households	52	44	68	34

The data is split into 50% used for training and the other 50% for testing (e.g. for AC, 26 households for training and 26 households for testing). As the aim of this paper is to look at decomposing aggregated loads, the smart meter interval load data (i.e. general and controlled loads) is aggregated for both the train and test data sets. The load data is then combined with the climate, datetime and day type data to get the full set of variables used in the ML model. The data is then normalised between 0 and 1 to convert all variables (i.e. features) to a common scale, while keeping the general distribution of the data. This is important for ML models that rely on distances between data points such as k-means, by assigning the same importance to all variables. Normalisation of data has also shown to improve the accuracy of NN, and can speed up NN's convergence towards a local/global minimum (Sola and Sevilla, 1997).

The model is then trained on the training data set and tested on the testing data. The results of the model is assessed based on its performance at evaluting the aggregated appliance loads for the testing data, which is found by comparing the output of the model with actual aggregated appliance loads. Since the use of different groups of customers for testing and training would affect the results, the model is trained 100 times with the training and testing data randomly split. The results are then the average of the 100 iterations. A control was also implemented, which consisted of directly comparing the aggregated appliance load profiles of the test and training data sets (made possible due to the even split of data). Highly similar aggregated appliance load profiles for the test and train data set would yield better statistical results for the control. The ML model should at minimum perform better than the control (the results are detailed in the *Appendix*).

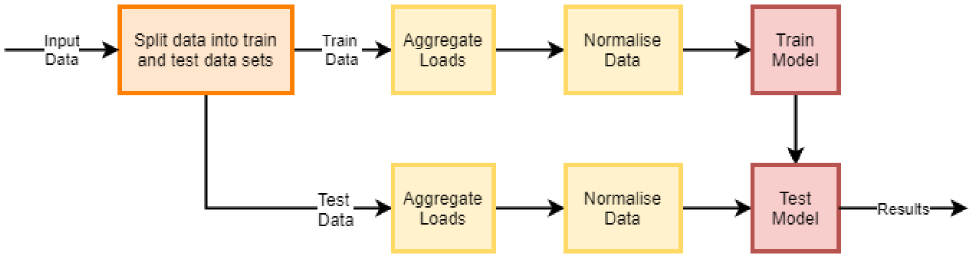


Figure 3 – Flow Diagram of Data Preparation for Training and Testing

3.5. Application of K-means to Pre-Cluster Similar Time Periods for Regression

K-means has been used to pre-cluster similar time periods, so that each cluster can be trained on a separate NN as shown in *Figure 4*. Training the NNs on separate clusters of similar load profiles allows for each NN to be specialised at identifying the appliance load profile of their cluster. Time periods are pre-clustered based on the features used for training as shown in section *3.3*. One or more features can be used for K-means clustering. The clusters chosen should ideally group together timeperiods which best reflects the operating characteristic of the aggregated appliance load profile. The choice of how many clusters, and the type of features used can be chosen by visual inspection of data. In *Figure 3*, which compares 2 clusters with 3 clusters, it can be seen that having 2 clusters is sufficient to provide indication of when AC usage is high, hence 2 clusters should be selected. There also exist more systematic and better methods to find the optimal number of clusters by evaluating the model performance on different number of clusters, however this method was not used in this paper. It is found that clusters formed by features with the highest dependency (as per *Figure 5*) often gave the best results. A summary of the K-means configuration used for the K-means NN in this paper is found in *Table 3*.

Table 4 – Features and Number of Clusters used in K-means Clustering.

Appliance	Features used for clustering	Number of Clusters
AC	Aggregated general load	2
Hot Water	Aggregated controlled load	2
Dryer	Aggregated general load	2
Washing Machine	Aggregated general load, daylight hours	2

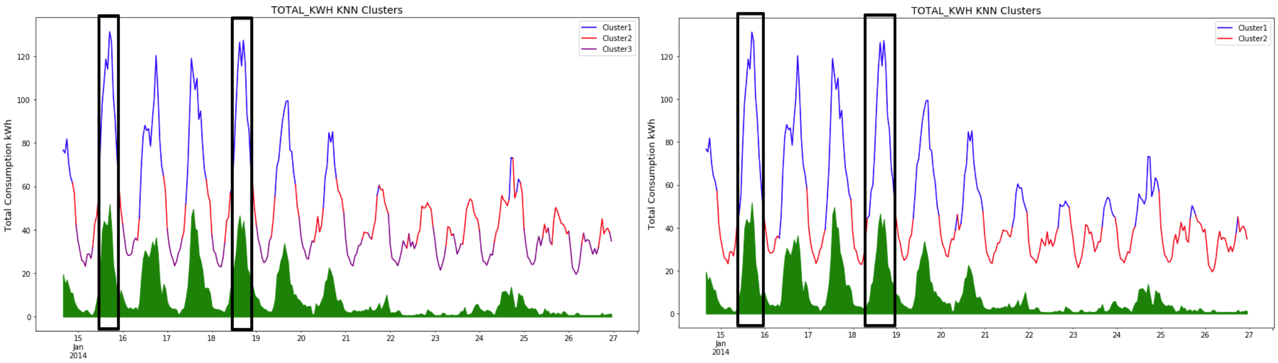


Figure 4 – K-means clustering (3 clusters on the left and 2 clusters on the right) of timestamps based on total aggregated household loads for determining AC load (shown in green).

3.6. Building and Training the Neural Network Structure

TensorFlow from Google's opensource ML library was used to construct the NN. A NN that consisted of 4 dense layers as summarised in *Table 4***,** was used to model all 4 appliances. A dense layer in this situation refers to a fully interconnected layer, whereby all neurons in the previous layer are connected to all neurons in the next layer. The structure of the NN was chosen by testing various structures, and selecting the NN that gave the best results for determining the AC loads. Since not all configurations have been tested, there may be alternative structures for K-means NN model that could perform better than those presented in this paper.

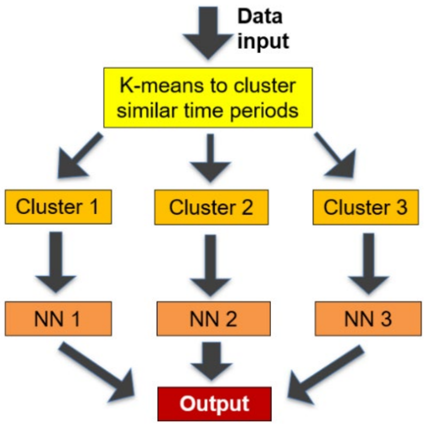


Figure 5 – Example Overview on the workings of K-means Neural Network with 3 clusters.

Forty percent of neurons in the first hidden layer are dropped out (i.e. these neurons are deactivated) during training to reduce the chance of the NN from overfitting (Brownlee, 2018a). The weights of the output layer are randomly initialised by a uniform distribution ranging from -0.5 to 0.5. Random initialisation allows a different starting point for the NN, and has been shown to prevent a NN from being stuck in the same local minima during training (Brownlee, 2018b).

Table 5 – Neural Network Configuration and Structure

Layer	Description
Input Layer (Dense)	76 input features
First Hidden Layer (Dense)	108 neurons with tanh activation.
Second Hidden Layer (Dense)	36 neurons with tanh activation
Output Layer (Dense)	1 neuron with linear activation.

3.7. Error Metrics

The following error metrics shown in **Table 6** are used to evaluate the performance of the ML models in this paper.

Table 6 – Summary of Error Metrics used

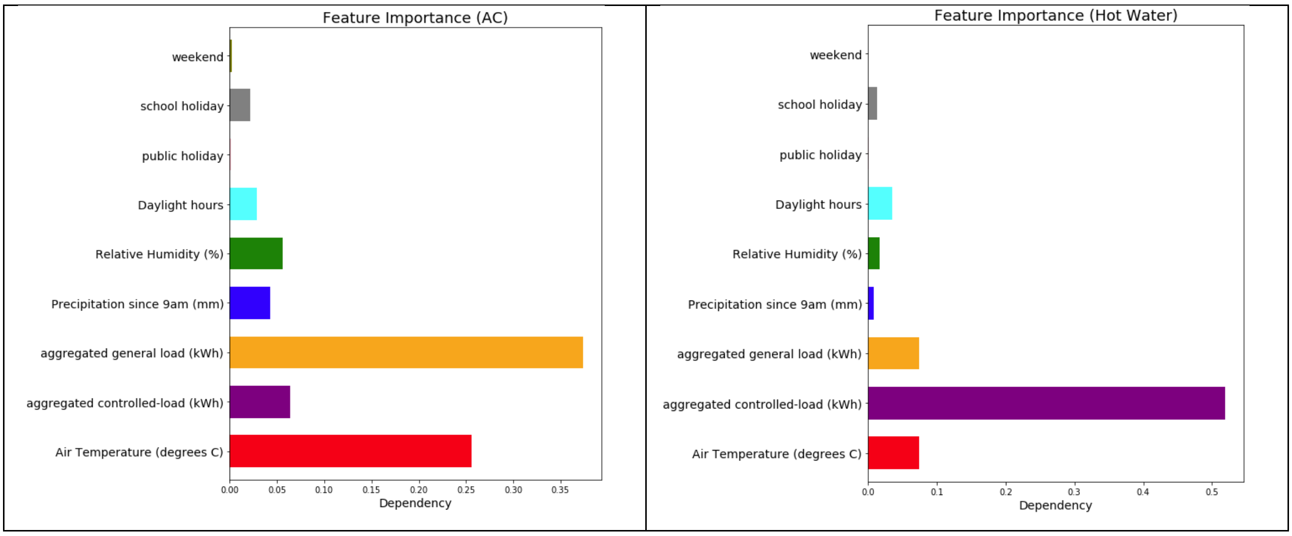
Mean Squared Error $(MSE) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	**Normalised Root Mean Squared Error** $(NRMSE) = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{y_{max} - y_{min}}$
$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$	**Mean Bias Error (MBE)** = $\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)$
Where: $y_i = The\ Predicted\ value\ returned\ by\ the\ model$ $\hat{y}_i = The\ Actual\ value$ $\bar{y}_i = The\ mean\ of\ the\ y\ values$	

4. Results and Discussion

4.1. Correlation with Feature Dependency and Accuracy of the Model

Figure 5 shows the importance of each of the features in determing the target variable. The results indicate that air temperature and aggregated total household load are best predictors for AC load, while total aggregated controlled load is the best predictor for hot water loads. The high correlation between controlled load and hot water load is expected as many hot water systems in NSW are ripple controlled (Ausgrid, 2016). Dryer loads have the highest correlation with aggregated general load, closely followed by controlled load and precipitation. Washing machine loads have the highest correlation with aggregated general load and daylight hours. In comparison to AC and hot water loads, both dryer and washing machine loads have low correlation with the selected features (i.e. the features have a dependency metric of less than 0.15).

Unsurprisingly, there is a strong correlation between the accuracy of the K-means Neural Network model and the dependency of the features used. As shown in *Table 5* and *Figure 6***,** a much higher accuracy in determining AC and hot water loads can be achieved than dryer and washing machine loads according to the R-squared value of the model.



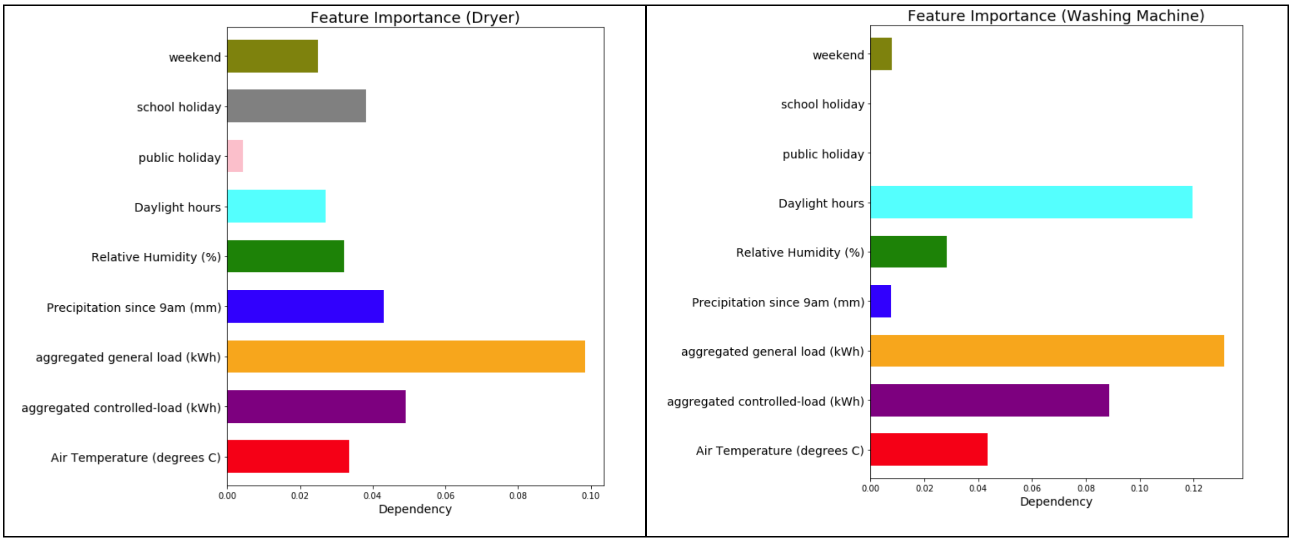
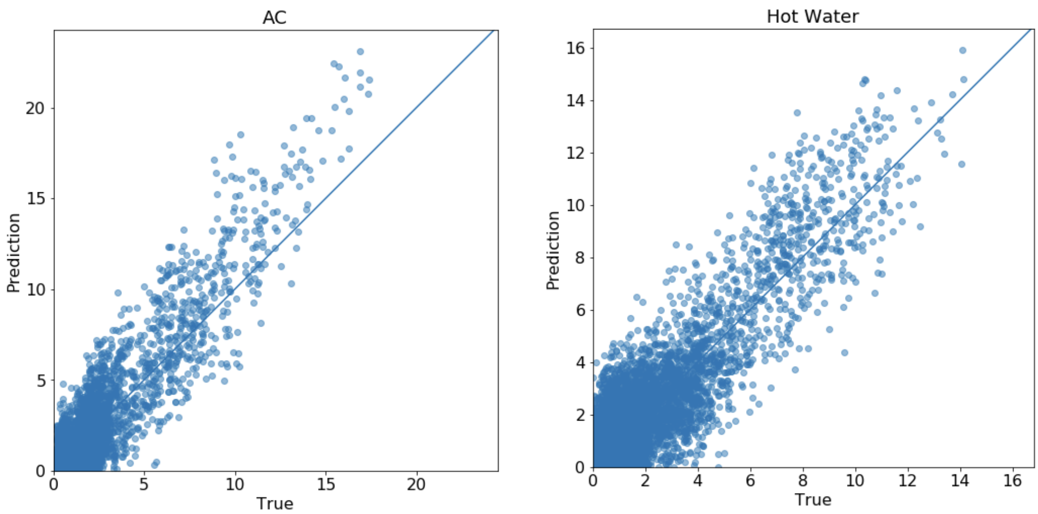


Figure 6 - Dependency of features to the target variable.

Table 7 – Average Performance of K-means Neural Network in Classifying the Aggregated Appliance Loads over 100 iterations.

Appliance	R-squared (R^2)	NRMSE (%)	MSE (kWh)	MBE (kWh)	Max Error (kWh)
AC	0.766	4.53	1.07	0.0520	8.10
Hot Water	0.684	7.01	1.22	0.0613	6.72
Dryer	0.0379	9.35	0.0938	0.00542	2.93
Washing Machine	0.217	8.75	0.0177	0.00290	1.18



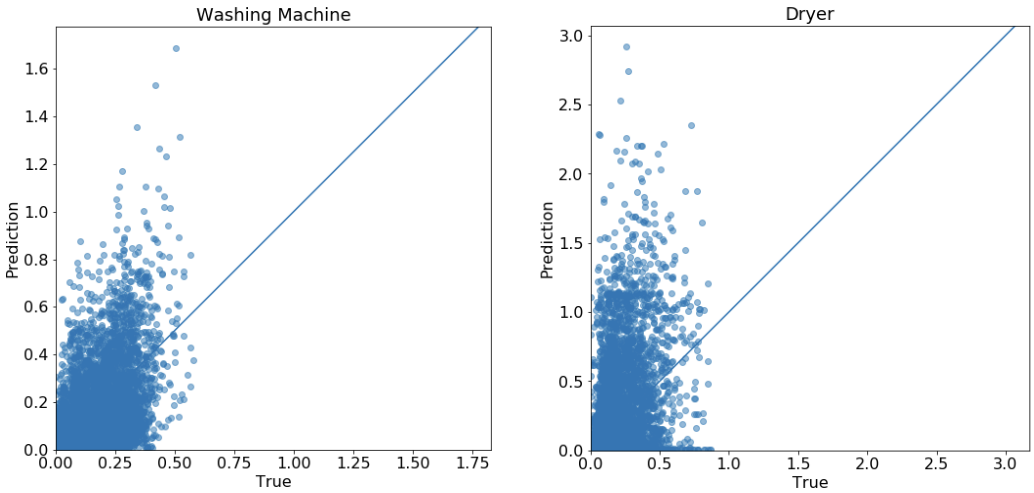


Figure 7 – Best Performing K-means NN Model for Predicted vs Actual Aggregated Appliance Loads for four different appliances.

4.2. Comparison of K-means Neural Network with other Machine Learning Models

Three other popular ML regression models (Neural Network, Decision Tree Regression and Support Vector Regression) have been used to comparatively evaluate the accuracy of the K-means Neural Network. The Neural Network (NN) has the same structure as the NN used in the hybrid K-means NN as shown in *Table 4***.** It is found that a simple NN performed the best of all the methods overall as illustrated in *Table 6*. K-means NN was second best overall, performing only slightly worse than the NN. Decision Tree was the worst performer for all appliances.

One factor that might have limited the accuracy of the K-means NN is that individual NNs in the hybrid model were only able to observe data from the cluster that they were trained on, limiting the individual NNs from “learning” about the correlations between the feature variables and the overall appliance load profiles. This therefore resulted in the individual NNs being less robust and more prone to overfit as they are trained on less data. Hence, a potential future improvement to the K-means NN would be to add the different clustered timeperiods as another variable/feature to be trained on a single NN.

Table 8 – Comparisons of NRMSE of Different Machine Learning Models.

Appliance	K-means NN (%)	NN (%)	Decision Tree (%)	SVR (%)
AC	4.53	4.15	5.15	5.11
Hot Water	7.01	6.82	8.67	7.27
Dryer	9.35	9.29	13.25	10.89
Washing Machine	8.75	8.75	12.46	9.74

4.3. Limitations of this study

One limitation of this study is that the SGSC home appliance data contains only 7 months of quality data (01/08/2013 to 01/02/2014). Since the duration of the data is less than 1 whole year, it prevents the ML models from properly training on seasonal load differences. This makes temporal variables like month less relevant and could reduce the accuracy of the models. Another possible limitation of load disaggregation is that the demand post solar generation is required to be known. If solar generation is not separated from the model it can create ambiguity in determining if patterns from the aggregated load are caused by loads switching on, or due to the reduction in solar generation from passing clouds.

The accuracy of the ML models explored is limited to the extent to which the sample of customers used for training are representative of consumption behaviours of the population. It is therefore expected that as more customers install smart appliances that have monitoring, the sample and hence accuracy of the proposed method will increase. Another limitation of this study is that the ML models are only tested on the SGSC data set. It is likely that there are biases with the types of customers that opted into participating in the SGSC trial. This bias could remove some of the variability in load profiles typically experienced in the electricity network, and hence the models could be performing better than expected.

5. Discussion and Conclusion

This paper aimed to explore a hybrid K-means NN for decomposition of aggregated household loads into aggregated appliance loads. Using the SGSC Home Area Network data along with total household load, a selection of weather and SGSC survey variables, ML methods were applied to load decomposition of 4 different appliance loads (AC, hot water, dryer and washing machine). The results indicate that a simple NN was able to outperform the hybrid K-means NN, however the K-means NN was able to outperform SVR and decision tree regression.

While the paper aimed to explore potential for the K-means NN to improve on NN, better results were not achieved. A likely reason for the lower performance is that training NNs in separate clusters may have resulted in overfitting, and have prevented the NNs from properly learning the correlations between the appliance usage and selected features.

Overall, results showed that AC loads can be most accurately disaggregated. This finding is significant as AC loads offers one of the highest DR potential, and are one of the biggest contributors to peak loads during summer in Australia. Hot water loads can also be determined fairly accurately, although this may be attributed to known controlled load profiles which many hot water systems are apart of. Decomposing for dryer and washing machine loads had extremely poor accuracies, and likely suggest that the features used in this study are not highly correlated with these load profiles.

Further work, particularly using better quality data is needed to improve the accuracy of load disaggregation. Exploration into different ML methods for clustering apart from k-means (e.g. Decision Tree, Support Vector Machines) may offer further improvements in the accuracy of load decomposition. The different clusters can also be trained on different ML algorithms to capitalise on the different strengths of the ML algorithms. Other ML methods that do not involve a pre-clustering clustering step should also be explored.

References

Adabi, A. *et al.* (2015) “Status and challenges of residential and industrial non-intrusive load monitoring,” *2015 IEEE Conference on Technologies for Sustainability, SusTech 2015*. IEEE, pp. 181–188. doi: 10.1109/SusTech.2015.7314344.

AEMC & EY (2011) “AEMC Power of Choice: Rationale and drivers for DSP in the electricity market – demand and supply of electricity,” (December). Available at: http://www.aemc.gov.au/Media/docs/EPR0022-Rationale-and-drivers-for-DSP-in-the-electricity-market-Final-Report-EY---for-publication-pdf-1295dc06-199b-4b09-b9d7-8e52ca564492-2.PDF.

AEMO (2018) “AEMO observations: Operational and market challenges to reliability and security in the NEM,” (March). doi: 10.1002/da.22743.

Al-Masri, A. (2015) *How Does k-Means Clustering in Machine Learning Work?*, *Medium*. Available at: https://towardsdatascience.com/how-does-k-means-clustering-in-machine-learning-work-fdaaaf5acfa0 (Accessed: September 22, 2019).

Ausgrid (2015) *Appliance usage guide*. Available at: https://www.ausgrid.com.au/-/media/Documents/energy-use/Appliance-energy-use-guide.pdf.

Ausgrid (2016) "Ausgrid Demand Management Ausgrid Demand Management Hot water load control trials," (August), pp. 1–35. Available at: www.aer.gov.au.

Ausgrid and AEFI (2014) "Smart Grid, Smart City: Shaping Australia's Energy Future," *Ausgrid*, (July), pp. 277–280.

Brownlee, J. (2018a) *A Gentle Introduction to Dropout for Regularizing Deep Neural Networks*. Available at: https://machinelearningmastery.com/dropout-for-regularizing-deep-neural-networks/ (Accessed: June 15, 2019).

Brownlee, J. (2018b) *Why Initialize a Neural Network with Random Weights?* Available at: https://machinelearningmastery.com/why-initialize-a-neural-network-with-random-weights/ (Accessed: June 10, 2019).

Petrican, T. *et al.* (2018) "Evaluating forecasting techniques for integrating household energy prosumers into smart grids," *Proceedings - 2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing, ICCP 2018*. IEEE, pp. 79–85. doi: 10.1109/ICCP.2018.8516617.

SA.GOV (2019) *SA.GOV.AU - Managing peak demand*. Available at: https://www.sa.gov.au/topics/energy-and-environment/energy-supply/managing-peak-demand (Accessed: April 19, 2019).

Sola, J. and Sevilla, J. (1997) "Importance of input data normalization for the application of neural networks to complex industrial problems," *IEEE Transactions on Nuclear Science*, 44(3 PART 3), pp. 1464–1468. doi: 10.1109/23.589532.

Su, F., Xu, Y. and Tang, X. (2018) "Short-and mid-term load forecasting using machine learning models," *CIEEC 2017 - Proceedings of 2017 China International Electrical and Energy Conference*. IEEE, (51507193), pp. 406–411. doi: 10.1109/CIEEC.2017.8388482.

United States Department of Energy - DOE (2006) "Benefits of Demand Response in Electricity Markets and Recommendations for Achieving Them," *U.S. Department of Energy*, (February), p. 122. doi: citeulike-article-id:10043893.

Zhuang, M., Shahidehpour, M. and Li, Z. (2018) "An Overview of Non-intrusive Load Monitoring Approches, Business Applications, and Challenges," (Lm), pp. 4291–4299.

6. Acknowledgements

The authors would like to thank Peter LeGras for his assistance in reviewing this paper.

7. Appendix

Table 9 –Comparisons of Average Performance of Machine Learning Models for Aggregated AC Load

Model Type	R-squared (R^2)	NRMSE (%)	MSE (kWh)	MBE (kWh)	Max Error (kWh)
K-means NN	0.766	4.53	1.07	0.0520	8.10
NN	0.806	4.15	0.892	0.0557	7.23
Decision Tree	0.697	5.15	1.39	0.0548	9.48
SVM	0.700	5.11	1.35	0.276	7.41

Control	0.493	6.35	2.09	0.00682	13.0

Table 10 – Comparisons of Average Performance of Machine Learning Models for Aggregated Hot Water Load

Model Type	R-squared (R^2)	NRMSE (%)	MSE (kWh)	MBE (kWh)	Max Error (kWh)
K-means NN	0.684	7.01	1.22	0.0613	6.72
NN	0.700	6.82	1.16	0.0586	6.44
Decision Tree	0.524	8.67	1.90	0.0849	8.38
SVM	0.666	7.27	1.32	0.374	6.33
Control	0.119	11.73	3.43	0.00312	11.7

Table 11 –Comparisons of Average Performance of Machine Learning Models for Aggregated Dryer Load

Model Type	R-squared (R^2)	NRMSE (%)	MSE (kWh)	MBE (kWh)	Max Error (kWh)
K-means NN	0.0379	9.35	0.0938	0.00542	2.93
NN	0.0514	9.29	0.0923	0.00880	2.94
Decision Tree	-1.01	13.25	0.194	0.0192	3.12
SVM	-0.315	10.89	0.127	0.172	2.94
Control	-0.818	12.49	0.165	0.00328	3.16

Table 12 – Comparisons of Average Performance of Machine Learning Models for Aggregated Washing Machine Load

Model Type	R-squared (R^2)	NRMSE (%)	MSE (kWh)	MBE (kWh)	Max Error (kWh)
K-means NN	0.217	8.75	0.0177	0.00290	1.18
NN	0.226	8.75	0.0175	0.00338	1.18
Decision Tree	-0.595	12.46	0.0361	0.00742	1.32
SVM	0.0358	9.74	0.0218	0.0530	1.20
Control	-0.389	11.53	0.0304	-0.00378	1.29