FIT5149 S1 2021 Assessment 2 Report Group No. XX

RESIDENTIAL ENERGY

APPLIANCE CLASSIFICATION

Author 1

Author 2

Author 3

MAY 28, 2021

**Abstract**

**Table of Contents**

[***1 Introduction 3***](#_heading=h.gjdgxs)

[***2 Pre-processing and features used 3***](#_heading=h.30j0zll)

[***3 Models 4***](#_heading=h.1fob9te)

[**3.1 Model 1 4**](#_heading=h.3znysh7)

[**3.2 Model 2 5**](#_heading=h.2et92p0)

[**3.3 Model X if any 5**](#_heading=h.tyjcwt)

[**3.4 Discussion of model difference(s) 5**](#_heading=h.3dy6vkm)

[***4 Experiment setups 6***](#_heading=h.1t3h5sf)

[***5 Experimental results 7***](#_heading=h.4d34og8)

[***6 Conclusion 8***](#_heading=h.2s8eyo1)

[***References 8***](#_heading=h.17dp8vu)

# 1 Introduction

This section should briefly introduce the project, for example,

* What is the project about?
* What is the goal of the project?
* How the project was done in your group? For example, how was the work allocated to each individual group member?

Please cite references properly, e.g., [1, 2, 3].

Electricity is one of our most irreplaceable resource. It is use in our routine life tasks, complicated applications such as space centre or satellite. Due to its importance, people in recent year are still working with the study of the nature of electricity in order to create a way generates electricity unprecedented in past.

Throughout history, human had hydroelectricity, thermal power, wind power, solar power and even nuclear power. The nation that we are living – Australia, is a thermoelectric country. And because each method of generating electricity has advantages and disadvantages, thermal power leads to the air pollution, environment degradation and also ozone depletion. However, there is no alternatives for Australian government.

Therefore, study how to reduce the electricity consumption is a work without delay. This research aims to analyse the usage of power for household appliances and predict the consumption for future application (when an appliance uses power in a specific minute) – to transmit electricity more rationally. For “testing the water”, our scope is only for 5 appliances including Air conditioner, Electric vehicle charger, Oven, Cloth washer and dryer in over 500,000 minutes (has been divided into training and testing data).

Because of the lack of time and human resources, the works for this project would be separated for 3 people. The works including:

* Study the data, exploratory data analysis: Ngoc Anh Duy Nguyen
* Implement model 1: Duy Tho Le
* Implement model 2: Minh Thai Nguyen

Then, we would summarise and write a report for each section. The whole process would be conducted in parallel with the goal is to maximize the evaluation. Therefore, some part in the research seemed not to be reasonable but it brought the higher result. We would discuss it in the particular part with our team’s work experience.

# 2 Pre-processing and features used

This section discusses

* The steps (if any) you used to pre-process data
* What are the feature set(s) used?
* If you generated new features, describe how were they generated? (Optional)
* Or any other you deem necessary.

As mentioned in the EDA step in Jupiter Notebook, in this part we would summarise our exploration.

Our given data are training and testing with 417720 and 105540 data records in respectively, each of data record represented for one minute and they are continuous (time-series data). Our core features are “load”, “hourofday”, “dayofweek”, and 5 columns contain predicted values for 5 appliances whether at that moment they are used or not. Some default statistical values provided to us including “dif”, “absdif”, “max”, “var”, “entropy”, “nonlinear”, “hurst”:

|  |  |
| --- | --- |
| **Name** | **Description** |
| Load | electrical load |
| Hourofday | what time the data record was captured in a day |
| Dayofweek | which day the data record was captured in a week |
| Dif | difference between two sequential load data points |
| Absdif | absolute value of diff |
| Max | maximum load in that moment |
| Var | variance of load over a neighbourhood time window of 30 minutes around each load data point |
| Entropy | the Shannon entropy that measures the "forecast ability" of a time series data |
| Nonlinear | the nonlinearity coefficient is used in Terasvirta 's nonlinearity test |
| hurst | the hurst is used as a measure of the long-term memory of a time series |

***Figure 1:*** *Description of variables*

Due to EDA process, we decided to keep the set of attributes as original because the new generated features using tsfresh library in Python are not useful since **some of them had low relation with the predicted values and some could be replaced by the default variables (because their correlation between them is high)** and we need to have **enough attributes to pass through the deep learning model**. Our working experience showed that although we have a potential set of features, if its size is too small for the algorithm, the result should be note acceptable. Therefore, we would use 10 default features for both of our models.

With the above intuition, we introduce Convolutional neural network (CNN) and Recurrent neural network (RNN). Detail of these algorithms would be discussed in the implement step and the comparison between them would be explained in detail in this report. The key idea of pre-processing is “deep learning” by scrolling the data by the size of window we prefer. There are 2 ways of scrolling: depth (for CNN) and breadth (for RNN). The window of rolling is based on the average running minutes of all 5 appliances for each running time (36 minutes).

For testing data, we were noticed that the order is not properly in series. However, due to these anomalies are very small so this phenomenon would not influence the result clearly. So, we would like to keep the order as normal.

# 3 Models

This section discusses the classifiers that your group used in the project, particularly those used in the comparison.

The discussion should include

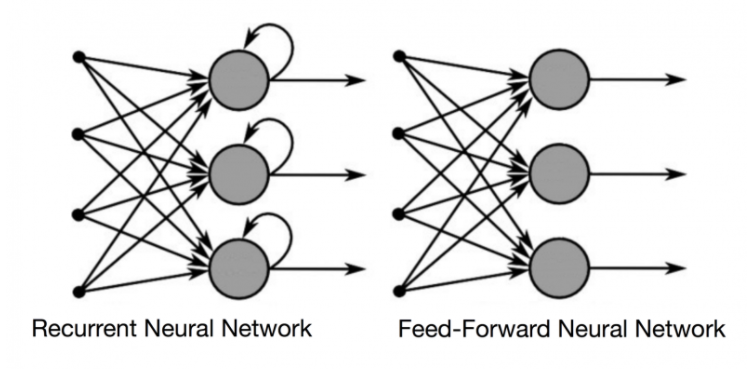
* A description of each of the models. For example,
  + the basic idea of each model,
  + how the model is constructed and works,
  + what are the assumptions made by each model if any
* The difference between the models
  + the advantage(s) and disadvantage(s) of the models
* Or any other aspects of the models that you think should be discussed.

Please replace the title of each subsection below with the name of the model your group used in the development.

## **3.1 Model 1: LSTM**

**3.1.1. Model idea**

While reading a story, human brains understand the current content based on the understanding that is in their memory from previous words. The traditional neural networks do not have the ability to connect previous knowledge to the current task. However, the recurrent neural network (RNN) can address this problem due to its internal memory, and it is ideally suited for machine learning issues involving sequential data. The below figure compares the difference between the structure of the RNN and the Feed Forward Neural Network:



Clearly, a feed-forward neural network simply takes into account the present input, so it cannot remember anything and is not suitable for time series problems.

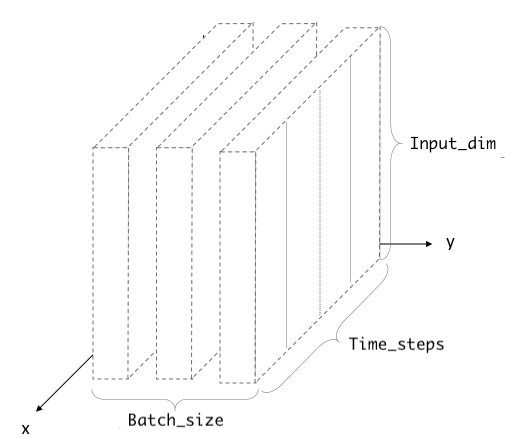
In RNN, the decision of the current input is made by considering what has been learned from the previous data. Therefore, our team decided to use LSTM ( a special case of RNN) to build a model for the given problem.

LSTM, which stands for Long Short Term Memory networks, has the ability to remember inputs for a long time. The LSTM can write, read and even remove the memory information, much like a computer’s memory.

**3.1.2. Model development**

After preprocessing data, we built five LSTM networks to predict the output of five features independently. The LSTM network is supported by Keras and the input data for the first layer must be a 3-dimension shape (M, T, N), where:

* M: number of samples
* T: sequence length (timestep)
* N: number of features



Firstly, we concatenate the training set and testing set into a big data frame. As we want to use T-minute windows of input data for predicting the target classes, T-1 last training records must be prepended to the beginning of the test set.

Secondly, we construct the 3D input using the NumPy array in Python for both the training and testing input data. The training set is further divided into two smaller subsets, one for training and one for validation.

Thirdly, 5 distinct LSTM networks will be constructed to predict the classes of five target variables. The final layer of each network would have only 1 neuron, to predict the class “0” or “1” with the activation function to be “sigmoid”. The number of epochs, learning rate, lambda in L2 regularization, and the dropout rate will be carefully tuned based on the f1 score on the training and validation set.

Finally, the trained models will be used to perform the classification tasks on the test set. The output would be saved in CSV format and submit to Kaggle for performance evaluation.

**3.1.3. Model assumption**

Because LSTM networks are most suitable for time series data, we do not use any technique to shuffle the input data. Moreover, As the prediction is made in both current and previous data, then we need to append the T-1 last training records to the beginning of the testing set. According to Edstem, the “load” data column in the test set is not ordered by time and it may affect the ability of the LSTM network to predict the target classes.

https://shiva-verma.medium.com/understanding-input-and-output-shape-in-lstm-keras-c501ee95c65e

https://www.kaggle.com/szaitseff/classification-of-time-series-with-lstm-rnn#-3.-Data-Pre-processing-for-LSTM-Model

ref: https://colah.github.io/posts/2015-08-Understanding-LSTMs/

## **3.2 Model 2: LSTM 5**

**3.2.1. Model idea**

Similar to the first model, we also used LSTM networks to perform classification for the target variables.

**3.2.2. Model development**

After getting the 3D input array for the training, validation, and testing set, we

## **3.3 Model X if any**

## **3.4 Discussion of model difference(s)**

# 4 Experiment setups

This section discusses how the experiments were setup. For example,

* What are the parameter settings for the models, if there is any?
* How was the cross-validate setup, if you used cross-validation?
* If you used training-validation approach on the provided training dataset, how your setup the split?
* How was the F1 score computed?

For the LSTM we choose T (timesteps window) to be 30 minutes after trying different values of T. Next,,

# 5 Experimental results

This section should discuss the comparison of different models with different feature set(s) (under difference settings if necessary). You could report your results in table(s) and discuss the results.

# 6 Conclusion

This section concludes the report with, for example,

* What is the optimal classifier? What is the best set of features used by that classifier?
* Any lessons learned from this project in regard to the classification task?
* Any suggestions for the future work?

# References

|  |  |
| --- | --- |
| [1] | T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado and J. Dean, “Distributed representations of words and phrases and their compositionally,” *Advances in Neural Information Processing Systems 26,* pp. 3111-3119, 2013. |
| [2] | G. A. Miller, “WordNet: a lexical database for English,” *Communications of the ACM,* vol. 38, no. 11, pp. 39-41, 1995. |
| [3] | J. Pennington, R. Socher and C. Manning, “GloVe: Global Vectors for Word Representation,” *the 2014 Conference on Empirical Methods in Natural Language Processing,* pp. 1532-1543, 2014. |
|  |  |