Deep Learning Project

Haein Lee

June 9, 2023

1 Introduction

I have decided to analyze the electricity usage fees of apartment complexes in Gangnam-Gu, Seoul. I constructed a recurrent neural network model for predicting the electricity usage fees based on past data. I have done so for each apartment complex and hope to use that model as a way to examine the status quo and envision the future to find apartment complexes that are more wasteful, and thus need more intervention in energy conservation.

2 Data

I have chosen to analyze this topic because I am interested in various environmental policies, including those for energy conservation issues.

While looking through various public data resources, I found various open APIs provided by Korea's Ministry of Land, Infrastructure, and Transport (국토교통부). One provided a list of all apartment complexes inside a specified area, with their unique apartment codes, various detailed information such as the number of households, electricity contract methods, number of elevators, and many more. With the apartment codes gained from this API, one can request data about the energy usage fees of the apartment complex. The data returned contains information about private and common usage fees of heating, electricity, water, and more. The private usage refers to the usage fees of each individual household, all summed up together, and the common usage refers to the usage fee of the entire apartment complex.

I first queried for all apartments in Seoul, requesting region by region(Gu). However, it was before I had decided to make a prediction model using RNN. After learning about a similar case in class, even predicting energy production in an assignment, I decided to focus more on a certain region, so that data collection would not take too long, since longer time periods of data would be advantageous in constructing a good model. I randomly selected one region, Gangnam-Gu to focus on, and I have energy fee data from January, 2016 to December, 2022. Looking through all of the data, the category with the least number of empty values was electricity usage fee, so I decided to focus on that category for my prediction.

Each apartment complex is of different size, and there are many ways to normalize the data, so that it makes more sense to compare between apartment complexes. There were many examples of options, such as dividing by area of living space, but I decided on dividing by the number of households. In cases such as common usage fees, its burden would naturally be shared equally between each household, so I decided that it would make more sense to divide using the number of households.

In the experiment, I focused on one apartment complex, and used its common usage fees and private usage fees to construct two models that predicted future usage fees. But I plan on using the same modelling principles to model all apartment complexes in Gangnam-Gu, and also to make a model of the average usage fees in Gangnam-Gu, so that I may compare regions and use it to see if there are any correlations with regional energy policies. I can also use the model constructed for each apartment complex to see the status quo of energy usage in each apartment complex. I may use other statistical analysis methods, such as causal inference, to analyze the effects of certain variables on energy usage fees. This status quo model will be a good comparison point for simulated situations with different variable values.

3 Experiment

The structure of the dataset and model construction follows the structure of assignment 6. I have constructed two training datasets and two test datasets, for common usage fees and for private usage fees. The model architecture was mainly based on the models that were successful in assignment 6. However, I did try many different variations. I did not include the variations in the code file because there were too many trials to include in one file. The code file contains the final architecture I arrived at for both models.

For both common usage model and private usage model, I used the past 10 months worth of data as input, and used it to predict the next month's output. I had 84 months worth of data overall, and when I split the data into training and test sets, I had 60 months in the training set and the remaining 24 months in the test set. After many trials, for both cases, 10 months seemed to be the reasonable amount of past data for the model to learn of seasonal patters, but not too much data, as there weren't as much data to learn as there were in assignment 6.

The first model I constructed is for predicting common electricity usage fees. It consists of a 50 unit layer of LSTM, then a 0.2 dropout to prevent overfitting, another 50 unit layer of LSTM, and a final dense layer with a 1-node output. I trained the model for 300 epochs, with a batch size of 32.

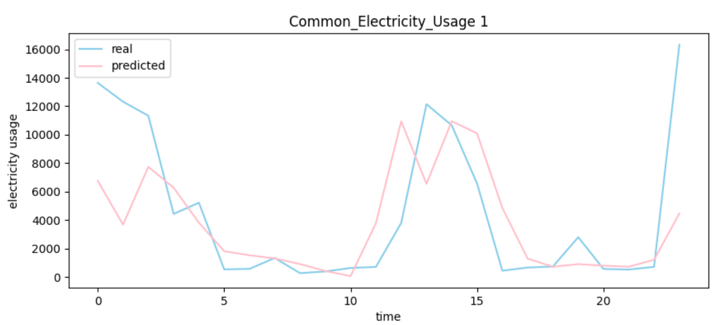
Though I cannot list all trials of tuning this model, throughout the tuning process I gained some insight about a model for this data. Common usage data is not as big as the private usage data, and also has lower variance. Thus, patterns were simpler and more regular, only needing a quite shallow and simple structure for modelling. It could also handle larger batch size when training. The batch size did not matter too much, as long as it was not too large or small.

The second model I constructed is for predicting private electricity usage fees. It consists of a 80 unit layer of LSTM, then a 0.2 dropout to prevent overfitting, a 50 unit layer of LSTM, and another dropout of 0.2, then another 50 unit layer of LSTM, with a final dense layer with a 1-node output. I trained the model for 300 epochs, with a batch size of 16.

Private usage data is bigger in scale than the common usage data, and has larger variance. Private electricity usage has been decided by an amalgamation of a large number of households' decisions, rather than a central decision by the apartment complex, like for common usage, thus it contained more random-looking patterns. The models had a very hard time predicting this data overall. But a deeper model structure seemed to work better with this data that contains more variation within the same time frame, compared to the common usage data. Thus, I increased a layer, and added another dropout layer. It performed slightly better with a batch size of 16, compared to other values such as 32, 24, or 8.

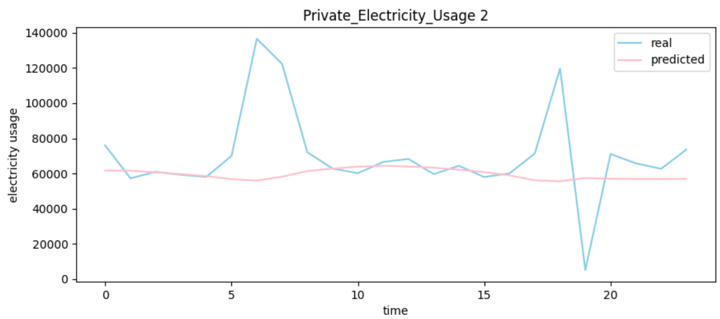
4 Conclusion and Discussion

4.1 Predicting Common Electricity Usage



The common usage model was successful at predicting some patterns of the data. It isn't perfect, but I did test this with a quite narrow time frame, so that should be considered.

4.2 Predicting Private Electricity Usage



The private usage model was quite unsuccessful at capturing the peaks and troughs of the data. With a more volatile and high-variance data, the model structure doesn't seem to be able to handle it proficiently.

The collection and preparation of data took me quite a long time, as it takes a while to request and receive each query from an API. Thus, I wasn't able to spend as much time as I would have liked on the modelling aspect. I would have liked to use some different structures, maybe some layers utilising simple RNN layers.My experiments were somewhat satisfactory regarding common usage data, that has lower variance, but unsatisfactory for private usage data, that has higher variance.

Overall, I succeeded at constructing models for common and private electricity usage fee data for a particular apartment complex in Gangnam-Gu. I plan on proceeding with this for more apartment complexes in Gangnam-Gu, as I have the data. I hope that these models would provide a better idea of the status quo than some existing time-series prediction models. I further plan on running these models multiple times, hopefully getting enough data to construct a sort of prediction error region. Though the nature of deep learning models make these models a black-box model of sorts, they have great proficiency in capturing the current state of the world and extrapolating it.

5 References

- STA3140 - 01 Deep Learning Lectures and Lecture Notes
- https://www.data.go.kr/data/15058453/openapi.do 국토교통부_공동주택 기본 정보제공 서비스
- https://www.data.go.kr/data/15057332/openapi.do 국토교통부_공동주택 단지 목록제공 서비스
- https://www.data.go.kr/data/15059469/openapi.do 국토교통부_공동주택 관리비(개별사용료)정보제공서비스