Predicting Taxi Demand with NYC Taxi Data

Team 1
Seongouk Kim, Sangjune Park, Sumin An, Minh Duc Nguyen, Thu Phuong Nguyen

September 23 2022

1 Description

In this project, we will train the model which predicts the taxi demand.

1.1 Dataset

For the data, we will use a NYC taxi data. This dataset is collected since 2009 and contains huge amount of data. Using entire data set may be too heavy to train in Google Colab environment or our PC, so we are considering to use some portion of the data. We are also considering to use the pre-processed dataset in Kaggle if possible.
Example of Kaggle dataset - NYC Yellow Taxi Trip Data

1.2 Model

Traffic information like taxi data contains both spatial information and temporal information. To predict the demand, we should handle both information. To do this, we will combine CNN for spatial information, and LSTM for temporal information.

2 Contributions

In implementing the model above, we must consider various factors such as the structures of the Convolution and LSTM layers, the order in which CNN or LSTM should be placed first in the architecture, and tuning hyperparameters. Thus, experimenting with diverse settings and analyzing each result to obtain the best one is the must-do in the project. By doing this, we will discover the impact of each factor on the performance of the model, hence, have a better understanding of the CNN, LSTM architectures, and Deep Learning development.

3 Role

Here is the table containing our team members' expected roles. However, this table is subject to change when we carry out the project.

Name	Role
Seongouk Kim	Data pre-processing
Sangjune Park	Implementing models
Sumin An	Implementing models
Minh Duc Nguyen	Analyzing results / visualization
Thu Phuong Nguyen	Presentation

Table 1: Team members' expected roles