Prediction of urban business district based on trajectory time series feature recognition

Qingyang LI; Jiahao Chen

Summary

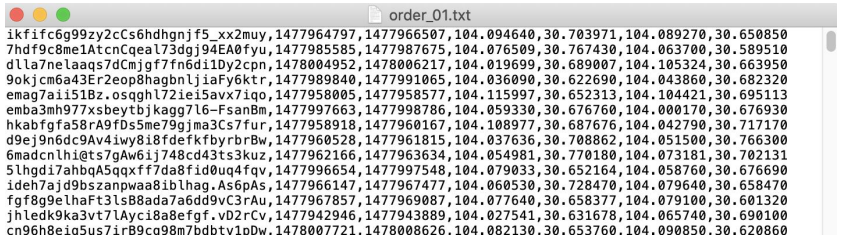
- Description and Goal: Business district refers to the urban functional area with commercial functions, and it has the functions of catering, entertainment and so on. However, this district is changed overtime due to human daily activities. The goal of our project is to predict this changes using deep neural network.

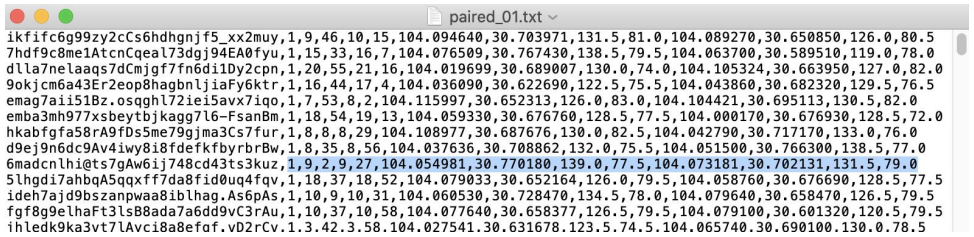
- Approach: In this project, we are going to use Convolution Long Short-Term Memory(Convolution LSTM) to be our neural network, and we will use some time series data feature recognition methods.

- Results: With our CLSTM model, we can predict the futures status of business areas with about 80% precision and 0.72 f1 score.

Implementation Details

- Data extraction and preprocessing: We get taxi data from DIDI, and Chengdu's POI information from Qingyang's University senior. The original data contains: ID of taxis, the time for passengers to get on and off, and latitude and longitude coordinates of each order. We translate the time data into (day,hour,min) form, and translate the latitude and longitude coordinates into 2-d coordinates.
- We associate the location with the nearest function area (from POI), and then we can build a heat map recording how many times of one area become a get on or get off place. We save this information as numpy matrix by generating npy files.
- before we processing: after we processing:

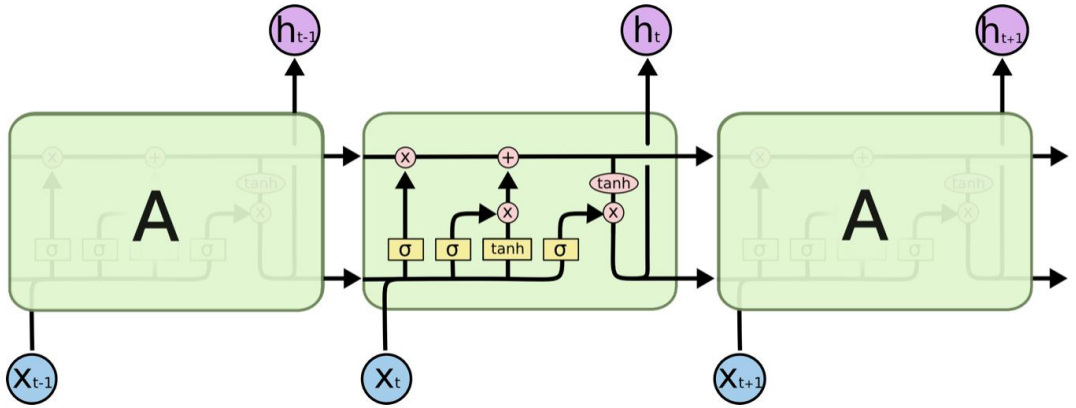




- We can see, before processing, the data contain: id, get on/off time, get on/off latitude and longitude. After processing, the data contains: id, day, get on hour and minute, get off hor and minute, get on/off latitude and longitude and get on/off XY coordinate.

Convolution LSTM Model

- Structure diagram

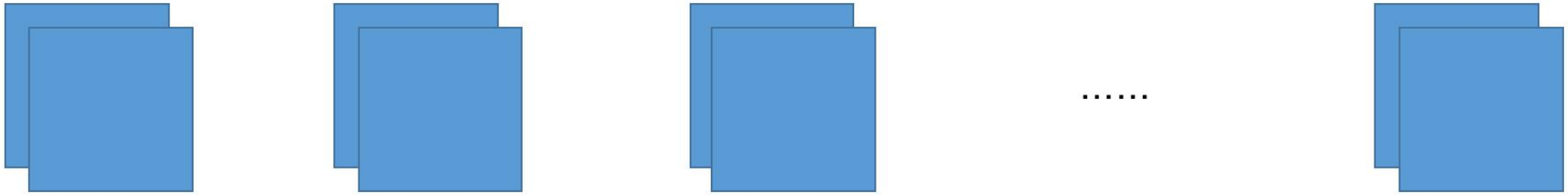


- mathematical model

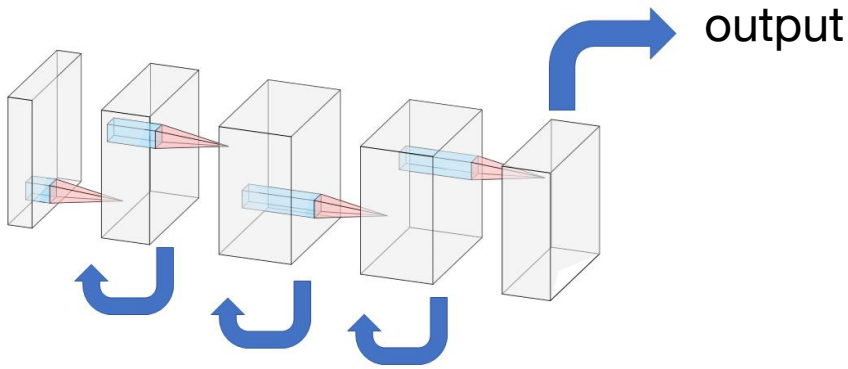
$$
\begin{aligned}
i_t &= \sigma(W_i \cdot [h_{t-1}; x_t] + b_i) \\
f_t &= \sigma(W_f \cdot [h_{t-1}; x_t] + b_f) \\
o_t &= \sigma(W_o \cdot [h_{t-1}; x_t] + b_o) \\
g_t &= \tanh(W_\top \cdot [h_{t-1}; x_t] + b_\top) \\
c_t &= i_t \odot g_t + f_t \odot c_{t-1} \\
h_t &= o_t \odot \tanh(c_t)
\end{aligned}
$$

Convolution LSTM Model

- input : each of the layer is 2*40*40

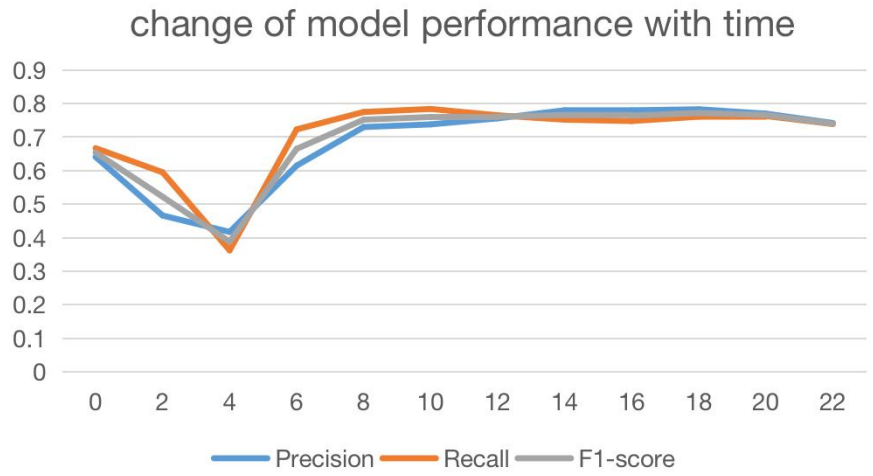


output of the model:

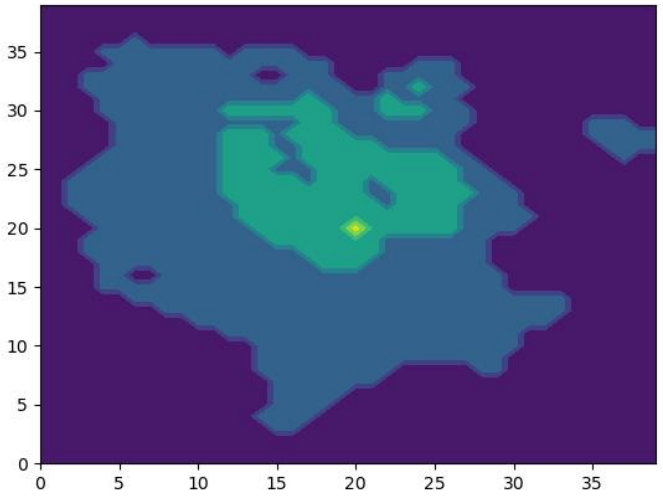


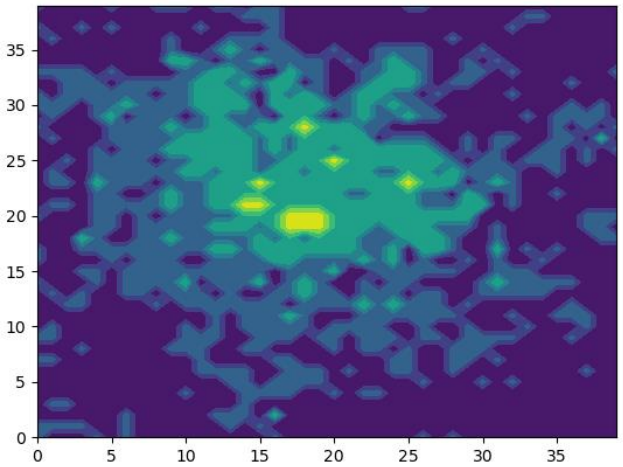
Result

We compare the model performance, using different number of days. And we finally choose 20 day's information as the input of our project.



Prediction





The left figure is the result of our prediction, and the right one is the ground truth.

References

- We found some papers relative to this project, and we are going to read and implement them. And by the time we actually working on the project, there may be more papers and models we will reference.

[1].Shi X, Chen Z, Wang H, et al. Convolutional LSTM Network: a machine learning approach for precipitation nowcasting[C]. neural information processing systems, 2015: 802-810.
[2].Zhang J, Zheng Y, Qi D, et al. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction[C]. national conference on artificial intelligence, 2016: 1655-1661.
[3].Zheng Y, Li Q, Chen Y, et al. Understanding mobility based on GPS data[C]. ubiquitous computing, 2008: 312-321.
[4]. Kwella B, Lehmann H. Floating car data analysis of urban road networks[C]. International Conference on Computer Aided Systems Theory. Springer, Berlin, Heidelberg, 1999: 357-367.