Traffic prediction across multi-modal transportation modes using GNN

Mathis Konarski, Higher School of Economics

Supervised by: I. Makarov

June 22, 2022

Abstract

While moving inside a city is primordial, it is also often difficult because of congestion, lack of adapted mass transit option, unavailability of safe bike roads. This is due to the limited space available compared to the important number of peoples living inside cities.

City planning has always been a difficult task with a lot of factors, one of them is the prediction of travel time between the different modes of transportation. While at a world scale different crisis are threatening our lifestyle it is crucial to adapt our cities with the aim of avoid useless ride and unnecessary waiting time.

This research work aims to predict travel time and traffic demand inside a city by considering a multi modal scheme.

1 Literature Review

Over the years a multitude of techniques was used in order to improve the quality of traffic forecasting, the objective is to give better prediction than the historical average. The ARIMA model [1] is using time series to do so. We saw the interest of using machine learning methods with [2] using Support Vector Regression. [3] is showing the effectiveness of using deep learning methods, [4] is using Graph Neural Networks GNN and [5] establishes a first framework for the use of deep learning. [6] is showing the superiority of using travel time distribution instead of precise timing. [7] and [8] shows the possibility to predict respectively metro demand and pedestrian moves using similar techniques as traffic forecasting. We can also scale models as [9] shown by predicting traffic flow at highway stations.

[10] and [11] are able to capture spatial dependencies and short terms features. [12] is integrating long-term temporal dynamics.

[13] is showing the effectiveness of using together crowd flows and origin destination prediction in order to improve the results of a GNN. [14] uses the combination of solo and shared rides to improve taxi demand prediction. [15] is going further by combining bike and taxi rides.

2 Presentation

This paper is building a GNN for volume prediction, the number of persons starting and finishing a trip from one specific area during one period of time, it is based on a grid division from a city.

We are using both image convolution based on the demand volume and graph convolution [16] based on the previous flows. We are then combining the volume and flow resulting information and using an LSTM [17] [18] layer followed by a dense layer to make predictions for each mode of transportation. The prediction is finally refined with multi modal information through 2 dense layers. In order to predict the most important data, a targeted loss function based on Mean Squared Error is implemented.

This paper is aiming to show the effectiveness of using a graph-based approach for flow treatment rather than an image based approach. The second objective is to see the impact of multi-modality on training.

3 Datasets

As shown in [19] it is important to use datasets already used in similar traffic prediction problems.

We will use a trio of datasets with the NYC bike [20] and taxi records [21] from April to June 2016. The taxi records are decomposed between green cab and yellow cab that we will be considering differently. We are using data corresponding to Manhattan, the East coast of Jersey City and the Northwest part of Brooklyn where 95% of the bike trips are realized.

It represents 3.5 million bike trips and 24 million yellow taxi trips and 0.8 million green taxi trips. This will allow to validate the model on various data.

Each trip got a precise time and place of departure and arrival. Extra information such as the distinction between subscriber and customer for a bike trip or the number of passengers for a taxi trip is not considered.

4 Metrics

We are predicted the start and stop demand for each part of the city. Our results are based on a time step of half an hour using RMSE and MAPE. So for a 10x20 grid we have 400 demand predictions each half an hour. However most of the demand values are really low and so not interesting for understanding the demand evolution. In order to best represent the real needs we are scoring the demand for one area only when there are at least 10 users in the area for the time step.

As the minimal volume to be considered and exact area of study can differ between 2 implementations I will not been able to directly compare my results with other papers. However papers are usually giving the historical average score which will give a baseline to estimate the performance.

5 Models

The main part of the image based model is based on [5].

5.1 Processing of volume data

For each area of the grid we have the number of trips started and finished in the area. This volume data corresponds to the one we are aiming to predict.

The volume entry tensor is a 10x20 image with 2 channels (start volume, stop volume).

We are doing a batch normalization, then we are using 3 convolution layers with a kernel of 3x3 followed by a dense layer with ReLU activation.

5.2 Processing of flow data

For the image based neural network we are treating the flow data on the same way than the volume data.

For each area of the grid, the flow data consists of the number of trips finished inside that area grouped by starting area. So we are having a 10x20 image for each area with one channel. In total it represents a 10x20x10x20 tensor.

To treat this data we are first flattening the last 2 channels , corresponding to the flow by area, and using a dense layer with ReLU activation on them. This is giving a 10x20xn tensor.

This tensor is processed the same way than the volume data, with a batch normalization, 3 convolution layers and a dense layer.

However this approach is bringing 2 problems, the first one is the size of the flow tensor who is mainly empty. The second problem is that this representation isn't natural, we can't understand it by a simple visualisation, which lead to the impossibility to exploit it in depth. It is not possible to do a convolution giving the same importance to the starting and finishing point of the trip without creating extra tensors with redundant data.

To treat these problems we are using a GNN. The flow data is treated as a graph with weighted edges. Each node represents an area, a directed edge connects 2 points if there is at least a trip from the first point to the second one finishing on the considered period. Each edge is then weighted with the number of trips.

We are doing 3 edges convolutions based on the model developed in [16]. Those convolutions are giving us an information tensor for each node. We are then normalizing those tensors followed by a dense layer with ReLU activation.

However this approach is losing the spatial information, this is counterbalanced by the fact that closer nodes will share more trips.

5.3 LSTM process

Following the volume and flow data processing we are getting 2 tensors of same dimension for each period.

The 2 tensors are multiplied together and transformed by a dense layer with ReLU activation. We are then applying the LSTM process.

5.4 Final prediction

The output of the LSTM process is directly used to make the prediction via a dense layer with a hyperbolic tangent activation. The output size of this dense layer is 400. The prediction is then unflattened for practical use into a 10x20x2 tensor.

5.5 Multi modal predictions combination

Based on the prediction and in order to study the impacts between each mode of transportation a multi modal prediction model is build.

This model consists of one dense layer with ReLU activation and one dense layer with hyperbolic tangent activation.

5.6 Loss

Our models are all based on Mean Squared Error loss.

As explained in the Metrics section, we are focusing on the areas with significant demand. To represent this and avoid that the model focus on predicting area with 0 to 1 user or predict a lower demand, we are giving more weight when the demand is superior to 10 trips, starting and finishing demand are treated separately.

Concretely we are dividing the loss by 100 where the demand is too low.

6 Experimentation

The experimentations are conducted using April, May, and June 2016 data from [20] and [21]. The training period is 70 days, and the validation period is 21 days, i.e. 1008 periods of test.

The area considered is a box around Manhattan within coordinates (40.68, 40.78), (-74.04, -73.94) divided into a 10x20 grid.

On all models we are using a batch size of 8, a LSTM window size of 3 days (144 time periods).

The number of neurons for the volume convolution is 2 by area, for the flow convolution it is 4 by area for both GNN and Image based model.

The resulting tensor from volume and flow process got 4 neurons as the LSTM input size. The LSTM output size is 32.

Inside the final combination neural network, the hidden layer size is 64.

Each dense layer is preceded by a dropout layer with a dropout rate of 0.5 as shown in [22].

Dataset	Model	Start RMSE	Start MAPE	Stop RMSE	Stop MAPE
NYC bike	Historical average	17.44	0.4470	22.08	0.5955
	Image based Model	**9.26**	0.2619	9.12	0.2581
	GNN model	9.38	**0.2573**	**8.98**	**0.2490**
	Combined GNN model	10.25	0.2828	10.20	0.2494
NYC green taxi	Historical average	12.11	0.4577	9.47	0.6518
	Image based Model	6.33	0.3091	3.04	0.2013
	GNN model	**6.00**	0.2775	**2.97**	**0.1951**
	Combined GNN model	**6.00**	**0.2685**	2.98	0.1966
NYC yellow taxi	Historical average	52.65	0.5160	60.97	0.4988
	Image based Model	30.75	0.3811	27.42	0.3723
	GNN model	**30.32**	**0.3670**	**26.24**	**0.3623**
	Combined GNN model	30.75	0.3813	29.38	0.3709

7 Conclusion

In this paper we show the effectiveness of using a graph approach to handling the flow information compared to an image based approach.

This approach in addition to be more efficient is making the model more understandable, reduce pre-processing time and largely reduce the memory needs. However it increases the time required for the model to learn and predict.

We also tried an approach about the multi-modality interactions, but the results are not conclusive. It slightly improves performances only for the green taxi more probably due to the fact that more data are available for the yellow taxi prediction with similar behaviours than thanks to an efficient multi-modality training.

8 Perspectives

It would be possible to use an attention layer [23] as in [5] or [15] implementations. Using an attention layer would allow to reduce the LSTM window while using older data focused on the hour of the prediction.

Using a GNN allows to use a more advanced splitting of the city, such splitting can be realized based on [24] work. However the results from this splitting will not be directly comparable with ours.

During the multi modal predictions we are creating one model who aim to predict a really important volume of information. This can be avoided by using one extra model for each mode of transportation. The data from the other modes of transportation will be combined via an attention mechanism or a direct tensor multiplication. This process would allow to make extra steps on the extra modes of transportation such as convolution.

As it is hard to find corresponding data for one city on one period of time, the combined prediction is based on taxi and bike sharing while these 2 modes of transportation are answering different needs. It would be interesting to have extra mode of transportation such as public buses, metro ridership, pedestrian movements and transits with personal cars and bikes. It will allow to understand the interactions between each mode of transportation and how they can be combined efficiently.

References

[1] B.M. Williams and L.A. Hoel. Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoretical basis and empirical results. 2003.

[2] R. Chen, C. Liang, W. Hong, and D. Gu. Forecasting holiday daily tourist flow based on seasonal support vector regression with adaptive genetic algorithm. 2015.

[3] Y. Lv, Y. Duan, W. Kang, Z. Li, and F. Wang. Traffic flow prediction with big data: A deep learning approach. 2015.

[4] Y. Li, R. Yu, C. Shahabi, and Y. Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. 2018.

[5] Huaxiu Yao, Xianfeng Tang, Hua Wei, Guanjie Zheng, and Zhenhui Li. Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. 2019.

[6] Xiucheng Li, Gao Cong, Aixin Sun, and Yun Cheng. Learning travel time distributions with deep generative model. 2019.

[7] Pengfei Chen, Xuandi Fu, and Xue Wang. A graph convolutional stacked bidirectional unidirectional-lstm neural network for metro ridership prediction. 2020.

[8] Menghang Liu, Luning Li, Qiang Li, Yu Bai, and Cheng Hu. Pedestrian flow prediction in open public places using graph convolutional network. 2021.

[9] Tianpu Zhang, Weilong Ding, Tao Chen, Zhe Wang, and Jun Chen. A graph convolutional method for traffic flow prediction in highway network. 2021.

[10] Ahmad Ali, Yanmin Zhu, and Muhammad Zakarya. Exploiting dynamic spatio-temporal graph convolutional neural networks for citywide traffic flows prediction. 2021.

[11] Kunpeng Zhang, Zijian Liu, and Liang Zheng. Short-term prediction of passenger demand in multi-zone level: Temporal convolutional neural network with multi-task learning. 2021.

[12] Guangyin Jin, Min Wanga, Jinlei Zhang, Hengyu Shaa, and Jincai Huang. Stgnn-tte: Travel time estimation via spatial–temporal graph neural network. 2022.

[13] Senzhang Wang, Hao Miao, Hao Chen, and Zhiqiu Huang. Multi-task adversarial spatial-temporal networks for crowd flow prediction. 2020.

[14] Jintao Ke, Siyuan Feng, Zheng Zhu, Hai Yang, and Jieping Ye. Joint predictions of multi-modal ride-hailing demands: A deep multi-task multi-graph learning-based approach. 2021.

[15] Haihui Xu, Tao Zou, Mingzhe Liu, Yanan Qiao, Jingjing Wang, and Xucheng Li. Adaptive spatiotemporal dependence learning for multi-mode transportation demand prediction. 2022.

[16] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. 2016.

[17] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. 1997.

[18] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning, Chapter 10: Sequence Modeling: Recurrent and Recursive Nets.* MIT Press, 2016. http://www.deeplearningbook.org.

[19] Filipe Rodrigues. On the importance of stationarity, strong baselines and benchmarks in transport prediction problems. 2022.

[20] Citibike. New york bike data records https://ride.citibikenyc.com/system-data.

[21] NYC. New york taxi data records https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page.

[22] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. 2012.

[23] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. 2017.

[24] A. Rodriguez and A. Laio. Clustering by fast search and find of density peaks. 2014.