AI for Traffic Prediction

Quentin Fruytier (260777209)
quentin.fruytier@mail.mcgill.ca
McGill University
Montreal, Quebec, Canada

Abstract

This research project aims to analyse and propose improvements to Deep Learning methods for traffic prediction. Traffic Prediction problem is a difficult and important problem impacting multiple industries. Previous methods have struggled to adequatly model the spatial and temporal dependency as well as the non-stationarity of the problem to yield good long and short term predictions. Recently, Deep Learning has become the state of the art method for traffic prediction as it is unmatched for discovering hidden patterns in large datasets. More specifically, Recurrent Neural Networks (RNN), Graph Convolutions and Encoder-Decoder frameworks have led to very succesful tools such as DCRNN [6], 3D-TGCRNN [10] and SLCNN [11]. Here, I take a deeper look at DCRNN by studying the effect of feature space's dimension on the METR-LA dataset and suggest improvements to the architecture.

CCS Concepts: • **Information systems → Spatial-temporal systems; Sensor networks**.

Keywords: Traffic Prediction, Deep Neural Network, Encoder-Decoder Framework, Stationarity, RNN, Graph Convolution, Attention Mechanism

1 Introduction

Traffic prediction is the task of forecasting real-time traffic information based on floating car data and historical traffic data, such as average traffic speed, time of the day, time of the week, underlying road network data, traffic incidents data or even online route queries data [5]. Traffic prediction is an important part of intelligent transportation systems (ITS) and is crucial to many applications including route scheduling problem, vehicle routing problem [2], congestion avoidance and headlights algorithms. Because, their tends to be too many vehicles for a road network with a limited capacity, traffic cannot always be avoided. But with accurate traffic predictions, it is possible to manage it.

Highly accurate traffic predictions is challenging due to the dynamic and complex environment of road networks. First, there is spatial dependency of the traffic data as the traffic any driver is experiencing depends on the section of the road network where the driver is located. Second, there is a temporal dependency of the traffic data as traffic speeds may be lower or higher depending on the time of the day (e.g. peak hours vs late at night) or time of the week (e.g. week-ends, vacations). Li et al. demonstrate that the spatial structure in the traffic is non-euclidean and directional. For instance, traffic is directional because congestion at a certain sensor in one direction does not imply congestion in the reverse direction. Furthermore, the euclidean distance between two road sensors is not a good way to model spatial dependency as the shortest road path length between the two points could be anything.

Previously, data-driven approaches like Auto-Regressive Integrated Moving Average (ARIMA) was used for this task. However, this time-series model relies on the stationarity assumption which is often violated by the traffic data. Indeed, the statistical properties causing traffic can change over time (e.g. road accidents, peak hours, etc.). Stationary processes are easier to analyze [8], meaning that a more complex approach is required for accurate traffic prediction.

In this paper, following the approach outlined in DCRNN, I model the underlying road network using a directed graph whose nodes are road sensors and edge weights denote proximity between the sensor pairs measured by the road network distance. The dynamics of the traffic flow are modeled using the sequence to sequence architecture and the diffusion convolution proposed by li et al. The contributions of this paper can be summarized as follows:

- I conduct extensive experiments on the METR-LA data set:
  - studying the effect multiple factors have on the prediction accuracy
  - the models predictive power across multiple time horizons
  - the models ability to follow trends in the traffic data through the day
  - the learning curve of the approach
- I propose and investigate improvements to the architecture for DCRNN.

The rest of this paper is organized as follows. I introduce several recent related works for traffic prediction in Sec.2. Following that, I give a detailed overview of the methodology and results of the experiments in Sec.3 and Sec.4. In Sec.5, I discuss and propose improvements to the Neural Network architecture. Finally, I conclude the paper in Sec.6.

2 Related Works

2.1 Traffic Forecasting

Due to the stochastic and non-linear nature of traffic, researchers have focused their attention on non-parametric methods for traffic prediction such as RF, SVR and more recently Long Short Term Memory Neural Network (LSTM) [7]. Liao et al. used LSTMs to study the effect of using highly correlated auxiliary information like online route queries on the accuracy of traffic predictions. They performed their experiments on the QTraffic data set with various modified network architectures of LSTMs in a sequence to sequence model. Finally, they demonstrated that the auxiliary information yielded more accurate predictions across all time horizons.

2.2 Convolution on Graphs

Because of the intricate spatial dependencies of the traffic data, graph convolutions have emerged as a powerful method for traffic prediction for their ability to model the spatial dynamics of the underlying road network. Li et al. proposed DCRNN [6], a sequence to sequence architecture of stacked RNN Layers with the diffusion convolution, a graph convolution they proposed. They achieve significantly better results on previous state of the art baselines on two benchmark data-sets. Zhang et al. propose Structure Learning Convolution (SLC) [11], a convolution approach on graphs where the adjacency matrix is also learned from the data. SLCNN achieves promising results performing slightly better than most state of the art approaches including DCRNN. More recently, Yu et al. propose 3D-TGCN [10] where instead of constructing the road graph based on spatial information, it is learned by comparing the similarity between time series for each road. Similarly to SLCNN, they achieve promising results performing slightly better than most state of the art approaches including DCRNN. These recent results suggest that it is more effective for traffic prediction to learn the adjacency matrix from the data rather than calculate it based on the spatial information.

2.3 Attention

One of the more important recent developments in Deep Learning has been argued to be Attention [3]. As a network architecture for time series data replacing the sequence to sequence architecture, it has been shown to be very effective for capturing the temporal dependencies in data [1]. Guo et al. proposed ASTGCN [4], an attention based spatial-temporal graph convolutional neural network model for traffic prediction. They achieve promising results on two real-world data-sets.

3 Methodology

A methodology similar to that of DCRNN (without scheduled sampling) is followed [6].

3.1 Traffic Forecasting Problem

The goal of traffic forecasting is to predict future traffic speed given historical traffic speeds from known sensors linked by the road network. The sensor network is represented as a weighted directed Graph G = (V,E,W), where V is the set of sensor nodes, E is the set of edges between the sensor nodes representing the road connections, W is an |V|x|V| weighted adjacency matrix representing the road distance between nodes. The traffic flow observed on G, X is a |V|x|P| vector where |P| is the number of features of each node (e.g. traffic speed, time of day, number of online route queries in a neighborhood of the sensor). Let $X^{(t)}$ represent the observed graph signal at time t. The goal of the traffic forecasting problem is to approximate the function h that maps T historical graph signals to T future graph signals, given a graph G:

$$[X^{(t-T+1)}, ..., X^{(t)} : G] \rightarrow [X^{(t+1)}, ..., X^{(t+T)} : G]$$

3.2 Spatial Dependency Modelling

The spatial dependency is modelled using bidirectional random walks to relate traffic flow to a diffusion process. The diffusion is used to characterize the reachability and impact road segments have on each other. Also, the bi-directionality allows to capture the influence of both the downstream and upstream traffic. The diffusion convolution is defined as follows:

$$X_{:,p*G} f_\theta = \sum_{k=0}^{K-1} (\theta_{k,1}(D_O^{-1}W)^k + \theta_{k,2}(D_I^{-1}W^T)^k) X_{:,p} \; for p \in 1, ..., |P|$$

where theta is a Kx2 filter of parameters to train, X is the Nx|P| graph signal at N nodes with P features for each node.

3.3 Temporal Dynamics Modeling

The Diffusion Convolutional Gated Recurrent Unit (DCGRU) is used to model the temporal dependency. It is a Gated Recurrent Unit (GRU) where the matrix multiplication is replaced with the diffusion convolution operation outline above. Furthermore, the sequence to sequence architecture is used for reading sequences of inputs and making a sequence of predictions [9]. Both the encoder and decoder are DCGRU with only two layers. During training, the sequence of historical data is fed into the encoder whose final states initialize the decoder. The decoder makes predictions based on previous ground truth predictions. However, during testing, the ground truth observations are replaced by predictions generated by the model itself. The model architecture to be implemented is shown in Fig.1.

3.4 Implementation

The implementation is written in python 3.7 google colab using TensorFlow 2.4.1. The Adam optimizer is used for optimization of the training weights with scheduled learning rate decay starting at $1x10^{-3}$. The code can be found at Implementation Code.

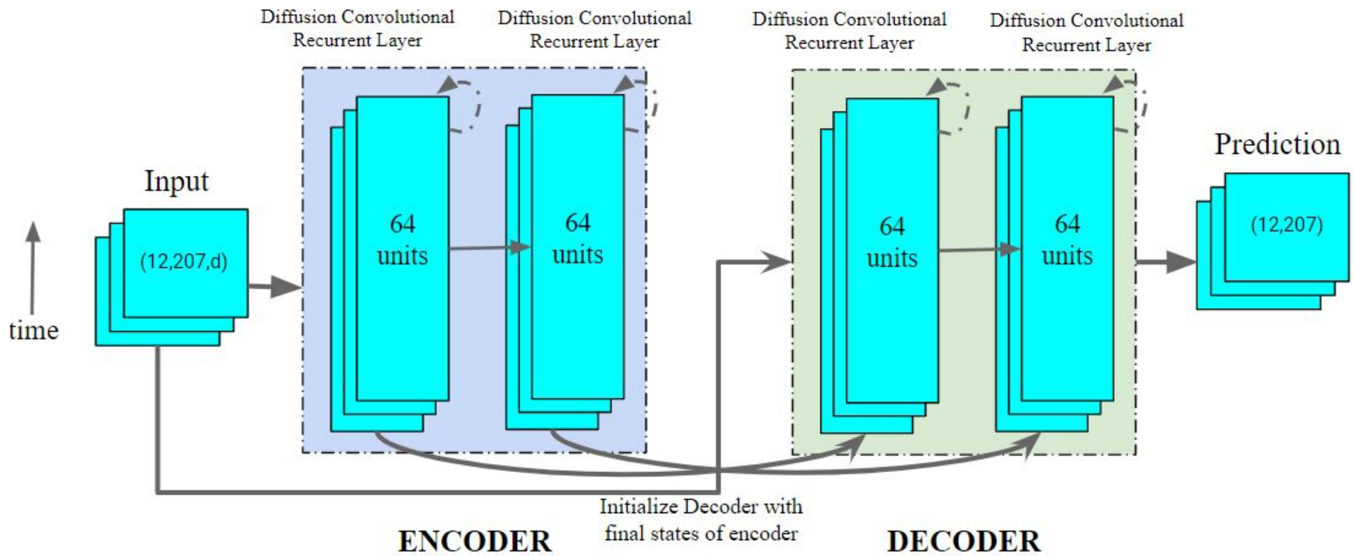

Figure 1. System Network Architecture to be implemented. The historical time series are fed into an encoder whose final states are used to initialize the decoder. The decoder makes predictions based on previous ground truths

4 Experiments

4.1 Datasets

I perform experiments on the METR-LA data sets from Los Angeles. It is a large-scale data set containing traffic speed information at 207 sensors over the month of Mar 1st 2012 to Jun 30th 2012.

4.2 Pre-processing

In total, I aggregate traffic speed readings into 34249 data points at 5 mins intervals. Each data point consists of 12 sequential historical traffic readings, time of day and time of week ranging from 5 mins to 60 mins into the future. Z-score normalization is used to normalize the data. Also, 70% of the data is used for training, while the other 30% is used for testing. The adjacency matrix is calculated according to DCRNN method [6]. The adjacency matrix is built from the pairwise road network distances between the sensors using thresholded Guassian kernels

$$W_{ij} = exp(-dist(v_i, v_j)^2/\sigma^2) \quad \text{if } dist(v_i, v_j) \leq \kappa, \;\; else \;\; 0$$

Where W_{ij} represents the edge weight between sensor v_i and v_j, dist(v_i,v_j) is the road distance between the two sensors from v_i to v_j. Sigma is the standard deviation and Kappa is the threshold.

4.3 Comparison and Results Analysis

Table 1 shows the comparison of different approaches for 15 mins, 30 mins and 60 mins in the future prediction against the base line FC-LSTM. FC-LSTM is an Encoder-Decoder framework using LSTM with peep-hole [6]. Both the encoder and decoder contain two recurrent layers. These methods are evaluated using the Mean Absolute Error (MAE) defined as:

$$Error = \sum |Actual - Prediction|$$

The implementation achieves the best performance across all time horizons suggesting the effectiveness of the diffusion convolution for modelling the spatial-temporal dependency of the traffic data. Interestingly, the implementation using only the traffic speed achieves the best result for 15 mins horizon. This is likely because for such a small time step, the prediction of the traffic is more stationary and the optimization reaches a better optimum with fewer training weights.

Fig.2 demonstrates the effect of using multiple factors for traffic prediction. It clearly shows that for short term predictions, it is preferable to use fewer features on top of traffic speed as the problem becomes more linear in nature. However, for long term predictions, it can be argued that the extra information from Time of the Week could yield better predictions for time horizons stretching past 60 mins.

Fig.3 shows the testing MAE vs the training MAE. We see that the difference between the two curves increases over time suggesting that the model is over-fitting the data. This is also suggested in Fig.4 as the training MAE has already long leveled off after 100 epochs.

Fig.5 and Fig.6 shows the implementation's ability to follow traffic trends throughout the day. We notice that the model is correctly able to predict significant changes in traffic 30 minutes before they happen as seen in Fig.5. Moreover, Fig.6 shows that the predictions are smooth.

5 Future Work

For future work, I plan to study the effects of additional auxiliary parameters utilizing QTraffic [7] datasets' online

Table 1. Performance comparison of different approaches for traffic speed forecasting. The implemented model performs the best according to the MAE error across all time horizons.

Time Horizons	Metric	FC-LSTM[6]	Implementation		
			Just Traffic Data	**Traffic Data + Time of Day**	**Traffic Data + Time of Day + Time of Week**
15 mins	**MAE**	3.44	2.92	2.94	3.05
30 mins	**MAE**	3.77	3.41	3.39	3.48
60 mins	**MAE**	4.37	4.12	3.98	4.02

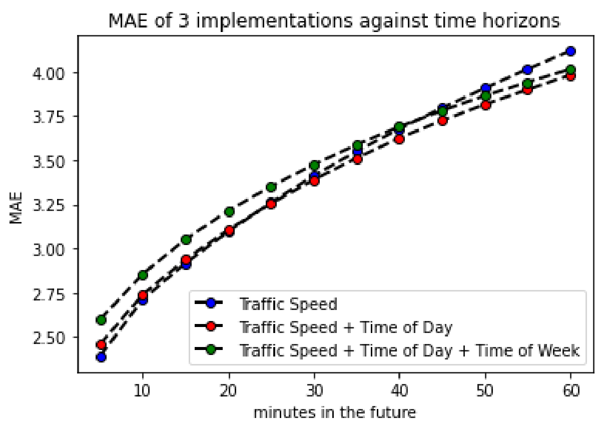

Figure 2. MAE of 3 implementations against Time Horizons

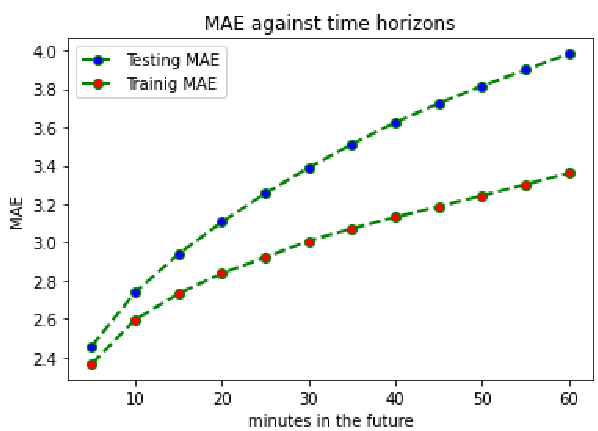

Figure 3. MAE against Time Horizons

route queries data. Furthermore, as Attention is coming up as the state of the art for dealing with sequences of time-series data [4], adding an attention mechanism instead of the sequence to sequence architecture could prove beneficial. Finally, as shown by Yu et al. and Zhang et al. learning the adjacency matrix instead of calculating it from spatial data greatly improves performance.

6 Conclusion

In this paper, I investigated a variation of the DCRNN [6] Deep Learning Approach for Traffic Forecasting Problem.

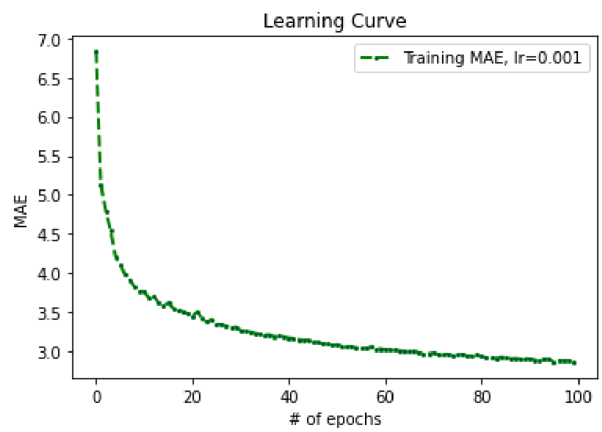

Figure 4. Learning Curve

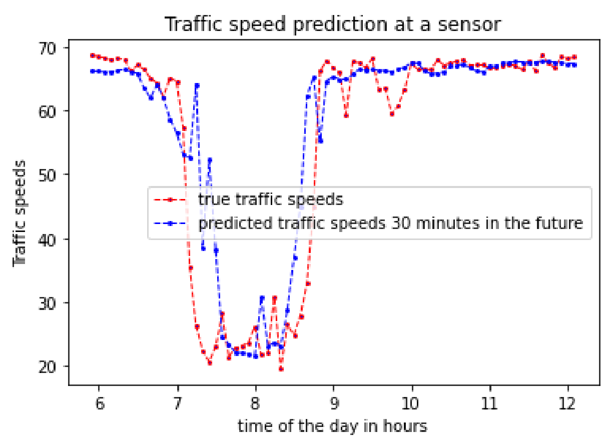

Figure 5. Traffic prediction visualization

The approach consists in formulating the traffic prediction problem as a spatiotemporal forecasting task and proposes a diffusion convolution for capturing the spatial dependency of the data. More specifically, I studied the effect of multiple features on the prediction accuracies using the METR-LA dataset. The experiments releaved that while increasing external factors for prediction does not help for short term predictions, it shows significant improvements for long term predictions. For future improvements, I plan to test the approach with more impactful external factors and different

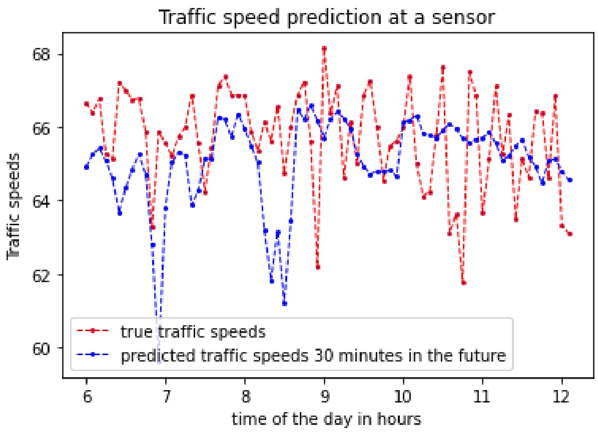

Figure 6. Traffic prediction visualization

datasets as well as adding an Attention mechanism to the architecture and learning the adjacency matrix.

References

[1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2016. Neural Machine Translation by Jointly Learning to Align and Translate. arXiv:1409.0473 [cs.CL]

[2] Kamil Bujel. 2018. Improving Operations with Route Optimization. https://towardsdatascience.com/improving-operations-with-route-optimization-4b8a3701ca39. Accessed: 2021-01-25.

[3] DeepLearningAI. 2018. *C5W3L07 Attention Model Intuition*. https://www.youtube.com/watch?v=SysgYptB198 Accessed: 2021-04-10.

[4] Shengnan Guo, Youfang Lin, Ning Feng, Chao Song, and Huaiyu Wan. 2019. Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting. https://doi.org/10.1609/aaai.v33i01.3301922

[5] Joao Laranjeira. 2020. What is traffic prediction and how does it work? https://www.tomtom.com/blog/road-traffic/road-traffic-prediction/. Accessed: 2021-03-15.

[6] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. 2018. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. arXiv:1707.01926 [cs.LG]

[7] Binbing Liao, Jingqing Zhang, Chao Wu, Douglas McIlwraith, Tong Chen, Shengwen Yang, Yike Guo, and Fei Wu. 2018. Deep Sequence Learning with Auxiliary Information for Traffic Prediction. arXiv:1806.07380 [cs.CV]

[8] Shay Palachy. 2019. Stationarity in time series analysis. https://towardsdatascience.com/stationarity-in-time-series-analysis-90c94f27322. Accessed: 2021-01-25.

[9] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to Sequence Learning with Neural Networks. arXiv:1409.3215 [cs.CL]

[10] Bing Yu, Mengzhang Li, Jiyong Zhang, and Zhanxing Zhu. 2019. 3D Graph Convolutional Networks with Temporal Graphs: A Spatial Information Free Framework For Traffic Forecasting. arXiv:1903.00919 [cs.LG]

[11] Qi Zhang, Jianlong Chang, Gaofeng Meng, Shiming Xiang, and Chunhong Pan. 2019. Spatio-Temporal Graph Structure Learning for Traffic Forecasting. arXiv:1704.08165