Short-Term Traffic Flow Prediction:

Final Presentation 5/3/2023

Oswald Garcia, Yifei Jin, Lydia King

Overview & Motivation

- Goal: Create a regression algorithm to use current traffic flow to predict short term (5 mins) traffic flow within a specific area.

- Motivation: Short-term traffic flow prediction based on current traffic flow is useful for e.g. warning drivers of an ongoing or impending traffic jam in time to change routes to avoid the jam, possibly even preventing the jam from occurring.

Research into Topic

- [1]. Zhang, Lun & Liu, Qiuchen & Yang, Wenchen & Nai, Wei & Dong, Decun. (2013). An Improved K-nearest Neighbor Model for Short-term Traffic Flow Prediction. Procedia - Social and Behavioral Sciences. 96. 653-662. 10.1016/j.sbspro.2013.08.076

Zhang et al used a modified kNN algorithm with k=18 to predict traffic flow 5min into the future, and obtained a Mean Absolute Percent Error of 9.5~9.8%.

→ We decided to first try the same approach.

Approach

- Obtain a dataset (everyone).

- Preprocess the dataset (Oswald).

- Implement kNN algorithm, tune k, evaluate performance on test set (Yifei).

- Investigate other algorithms such as MLP or RNN (Lydia): kNN has the disadvantages of being lazy (storage space and prediction time) and of only using the current traffic flow for prediction (not considering the previous several points like LSTM could do).

Dataset

- Used California Department of Transportation website (https://pems.dot.ca.gov)
- Chose a specific interstate highway, district, and vehicle detection sensor(VDS)
    - I 10-W
    - District 7
    - VDS 717129
- Observation Rate
- December 1, 2022 - February 28, 2023

Dataset

Out[46]:

	5 Minutes	Speed (mph)	Flow (Veh/5 Minutes)	# Lane Points	% Observed
0	12/1/2022 0:00	69.3	99	4	100.0
1	12/1/2022 0:05	67.3	71	4	100.0
2	12/1/2022 0:10	69.3	70	4	100.0
3	12/1/2022 0:15	67.7	104	4	100.0
4	12/1/2022 0:20	67.1	79	4	100.0

Out[50]:

	5 Minutes	Speed (mph)	Flow (Veh/5 Minutes)	# Lane Points	% Observed	Time of Day(Every 5 Mins)	Day of Week	Output
0	12/1/2022 0:00	69.3	99	4	100.0	0.0	0.0	71
1	12/1/2022 0:05	67.3	71	4	100.0	1.0	0.0	70
2	12/1/2022 0:10	69.3	70	4	100.0	2.0	0.0	104

Preprocessing

- Dropped columns labelled as: 5 Minutes, # Lane Points, and % Observation

Out[52]:

	Speed (mph)	Flow (Veh/5 Minutes)	Time of Day(Every 5 Mins)	Day of Week	Output
25914	60.4	133	282.0	5.0	124
25915	60.5	124	283.0	5.0	97
25916	63.2	97	284.0	5.0	93
25917	62.1	93	285.0	5.0	56
25918	65.9	56	286.0	5.0	32

25919 rows × 5 columns

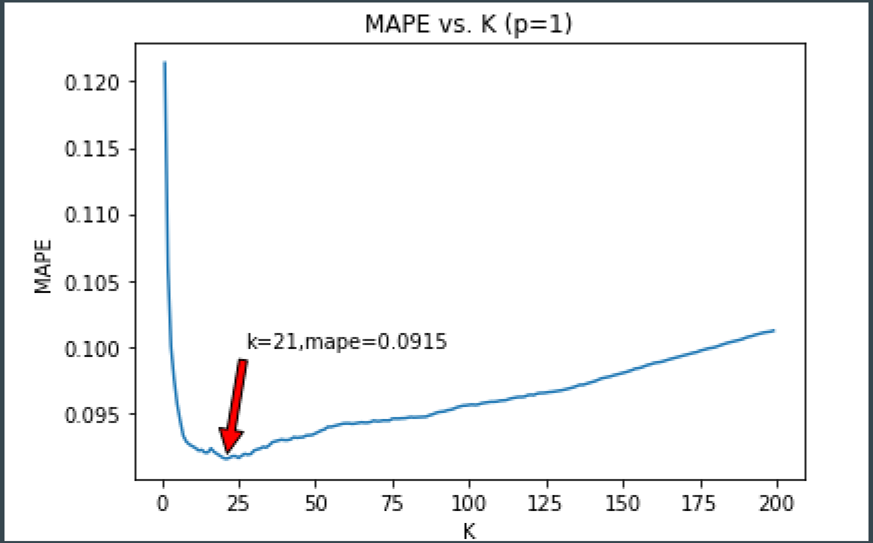
- Split dataset into training, validation, and test set
  - 8:1:1 (80% training, 10% validation, 10% test)
- Used StandardScaler from sklearn.preprocessing to normalize the features

Algorithm: kNN (Yifei)

- Using sklearn.neighbors library to train the K nearest Regressor
- Using MAPE metrics to evaluate the result of validation set and find out the best hyperparameter(K)
    - Comparing to MSE and RMSE, the MAPE is more likely to show how good the result is.
- Using both Euclidean Distance and Manhattan Distance to evaluate the result
- Use the KNN model with the best-adjusted parameters to evaluate the test set

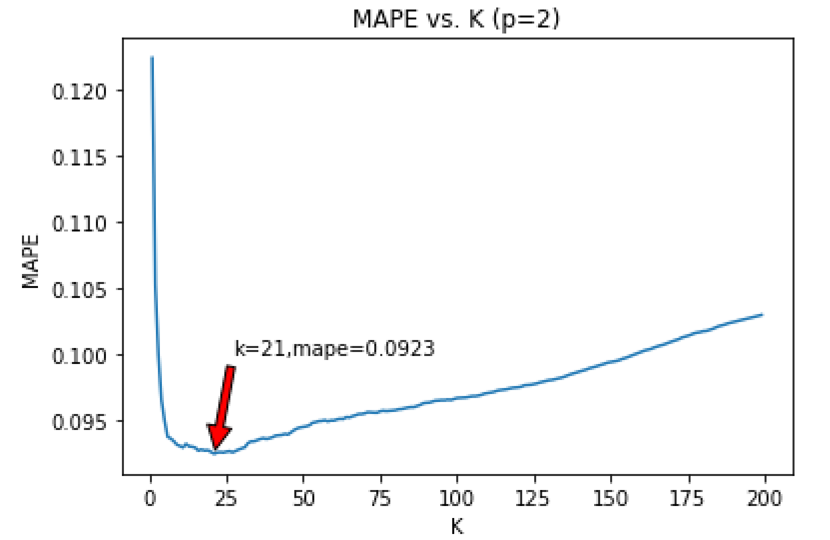
Evaluation & Results (Yifei)

- We find the best k = 21 using the manhattan distance in the validation set and the MAPE ≈ 0.0915



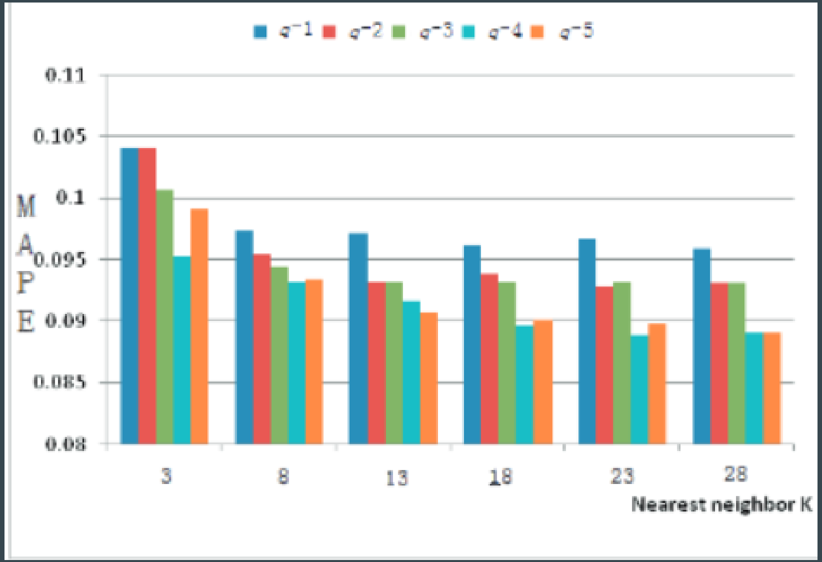
Evaluation & Results (Yifei) Cont.

- We find the best k = 21 using the euclidean distance in the validation set and the MAPE ≈ 0.0923



Evaluation & Results (Yifei) Cont.

- Cited from An Improved K-nearest Neighbor Model for Short-term Traffic Flow Prediction[1]



Evaluation & Results (Yifei) Cont.

- Results on the test set (parameter: k=21, p=1) slightly improved

```
print(mape_test)
```

```
0.08657253136771242
```

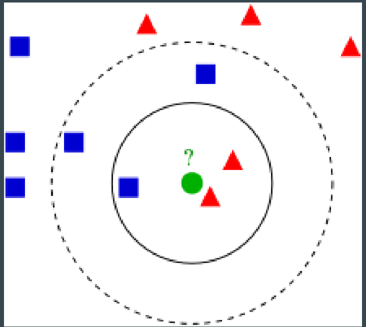
- The 8.6 percentage error seems to be an okay result, but what else did we do?

LSTM Model Implementation

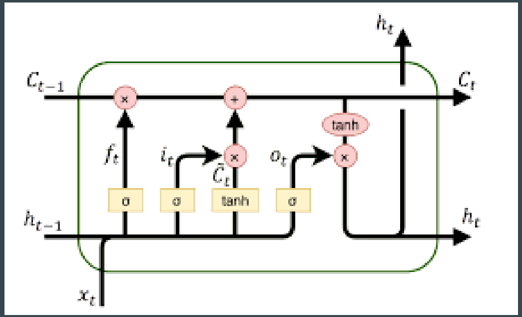
- The kNN model gives good results (MAPE < 9%) and surprisingly fast prediction (0.15ms/prediction on average on Yifei's machine), but all the training data must be stored for prediction.

- It may be advantageous (faster and smaller) to instead implement an eager algorithm such as LSTM.

- Several groups have successfully implemented traffic flow prediction using LSTM models, obtaining (respectively) MAPE of 6.49% over 15-min prediction time [2] and 26.4% when fitting the traffic flow curve for 24 hrs [3].



VS



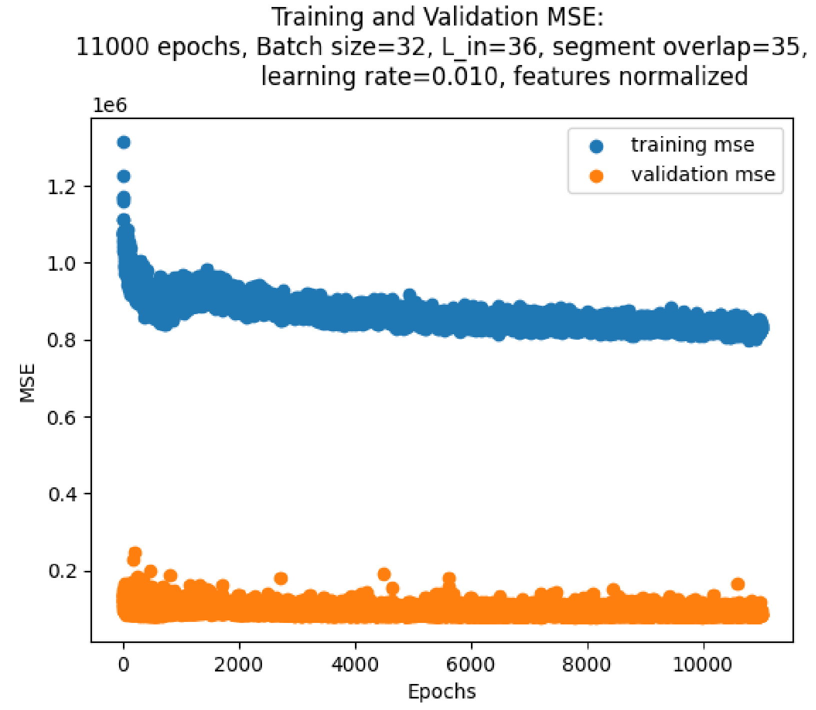
LSTM Tests

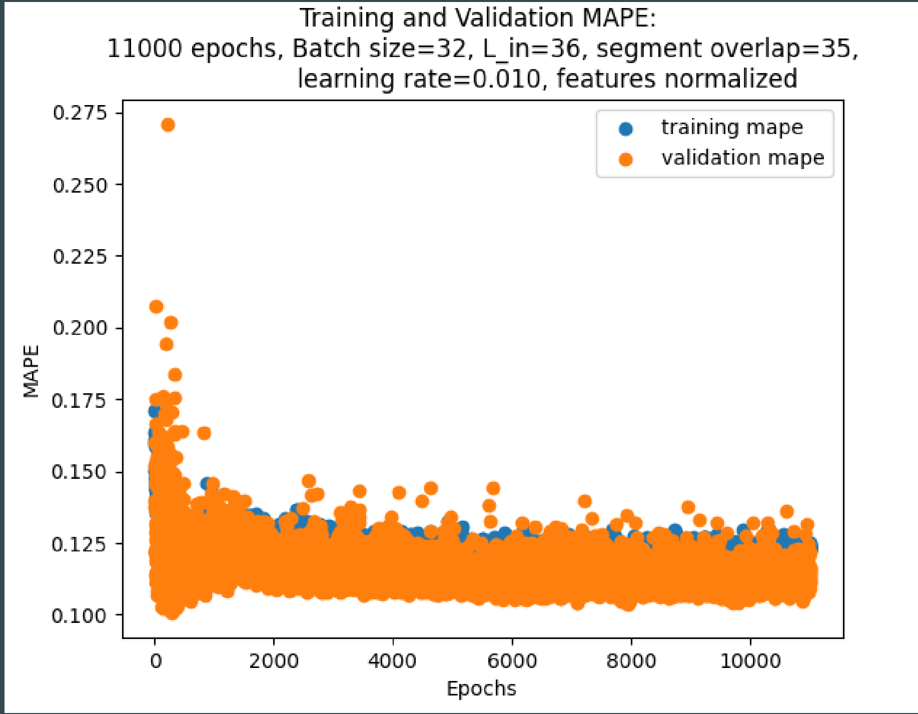
- Unnormalized, unshuffled data is divided into segments with adjustable L_in and overlap
- Segment output is the output of the final sample in the segment (flow from the next time)
- Segments and outputs are divided into shuffled train/val/test splits
- Segments are normalized using the mean and standard deviation of all points in all segments from the training set
- Network structure is similar to HW7: 2 LSTM layers, 1 fully-connected layer (512 neurons), 1 dropout layer (p=0.1), 1 fully connected layer for output (1 neuron)
- Adam optimizer is used with MSE loss criterion and learning rate = 0.01
- Final parameters used: L_in=36 (model 3 hour dependencies), overlap=35, batch_size=32

```
==========================================================================================
Layer (type:depth-idx)                   Output Shape              Param #
==========================================================================================
Lstm                                     [32, 36, 1]               --
├─LSTM: 1-1                              [32, 36, 512]             3,162,112
├─Linear: 1-2                            [32, 36, 512]             262,656
├─Dropout: 1-3                           [32, 36, 512]             --
├─Linear: 1-4                            [32, 36, 1]               513
==========================================================================================
Total params: 3,425,281
Trainable params: 3,425,281
Non-trainable params: 0
Total mult-adds (G): 3.65
==========================================================================================
Input size (MB): 0.02
Forward/backward pass size (MB): 9.45
Params size (MB): 13.70
Estimated Total Size (MB): 23.17
==========================================================================================
```

Final LSTM Test Results

11000 epochs, Batch size=32, L_in=36, hop=1 (overlap=35), lr=0.01
(first 6 MSE and MAPE points were much higher and not plotted to see more detail)
Final test MAPE = 11.88%





Conclusions and Future Directions

- The kNN model (8.66% MAPE) performs better than the LSTM model (11.88% MAPE)
- Potentially could improve LSTM performance by making architecture more flexible or training longer (overfitting was not observed within 11000 epochs)
- Could increase segment length L_{in} to model dependencies longer than 3 hours
- Could keep both kNN and LSTM models and average their outputs to get the final output
- Could try predicting farther into the future: 15-minute or 30-minute prediction times would be more useful to drivers

References

[1]. Zhang, Lun & Liu, Qiuchen & Yang, Wenchen & Nai, Wei & Dong, Decun. (2013). An Improved K-nearest Neighbor Model for Short-term Traffic Flow Prediction. Procedia - Social and Behavioral Sciences. 96. 653-662. 10.1016/j.sbspro.2013.08.076

[2] Y. Tian and L. Pan, "Predicting short-term traffic flow by long short-term memory recurrent neural network," *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)*, 2015.

[3] Q. Chu, G. Li, R. Zhou, and Z. Ping, "Traffic flow prediction model based on LSTM with Finnish dataset," *2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP)*, 2021.

[4] "Traffic Prediction: How Machine Learning Helps Forecast Congestions and Plan Optimal Routes." *AltexSoft*, 27 Jan. 2022, www.altexsoft.com/blog/traffic-prediction/.

[Slide 1] https://www.latimes.com/opinion/story/2020-04-27/los-angeles-streets-freeways-keep-traffic-like-it-is-now

[Slide 1] https://www.123rf.com/photo_162614713_black-is-driving-a-car-she-is-stressed-by-heavy-traffic.html

[Slide 5] https://www.i10exitguide.com/roadnews/california-i-10-corridor-project-approved-construction-begins-early-2020/

[Slide 13] https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

[Slide 13] https://apmonitor.com/do/index.php/Main/LSTMNetwork