TUDelft

Nikola Nachev
N.Nachev@student.tudelft.nl

Effects of weather data on traffic flow predictions using an LSTM deep learning model

1. Introduction

- Accurate traffic predictions are essential as they can improve the traffic flow of urban cities
- The LSTM gives accurate predictions [1] as a base model
- Traffic volumes are sensitive to weather changes
- Precipitation has a correlation with traffic flow [2]
- Including weather data in prediction models can improve prediction accuracy [3] [4].
- Weather changes do not have an immediate impact on traffic flow
- Adding a lag to the precipitation data can be beneficial

2. Objective

The goal of this research was to show:

- How to make an LSTM model for accurate traffic predictions
- How that model can be extended with the addition of weather data
- How the usage of weather data in the model affects the prediction accuracy of the model.

3. Methodology

- Implement an LSTM model with optimized architecture and hyperparameters.
- Analyze the correlation between the weather and traffic data
- Add lag of different sizes to rain data and analyze the correlations and prediction accuracy effects
- Predict the sensors with the highest and lowest correlation with and without weather data
- Check if predicting multiple timestamps in the future improves accuracy more
- Predict the intersections with the highest and lowest correlation

4. Experimental setup

Baseline LSTM model architecture:

- Two LSTM layers with 500 units
- Look back window of 80, or 20 hours
- Root Mean Squared Error loss function
- Trained on 200 epochs with early stopping
- Min-Max normalized traffic and weather data

Used datasets:

- Traffic data from 128 sensors across 11 intersections from The Traffic Department of The Hague Municipality
- Precipitation and temperature data from OpenMeteo

Input data for predicting a single sensor vs multiple

$$S = \begin{bmatrix} w_1 & w_2 & \cdots & w_l \\ x_1 & x_2 & \cdots & x_l \end{bmatrix} \qquad M = \begin{bmatrix} w_1 & w_2 & \cdots & w_l \\ x_{11} & x_{12} & \cdots & x_{1l} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{ml} \end{bmatrix}$$

6. Conclusion

- The LSTM benefitted from the use of rain data with a 30-minute lag when predicting individual sensors
- Different sensors had different results
- A correlation analysis helped to identify potential sensors
- More significant accuracy improvement when predicting the next 10 timestamps, or 2.5 hours
- No improvement when adding temperature data or when predicting multiple sensors at once

7. Future work

- Predicting multiple sensors at once remains an open question
- A possible solution is to use 2D feature input data, which was not possible with Keras
- Not very significant accuracy improvements found
- More significant results are expected if the findings of this research are used on larger datasets

5. Results

- The traffic sensor data had a weak positive correlation with the rain and temperature data
- Highest correlation between the traffic data and the rain data with a 30-minute lag
- Including the lag gave 4.6% accuracy improvement

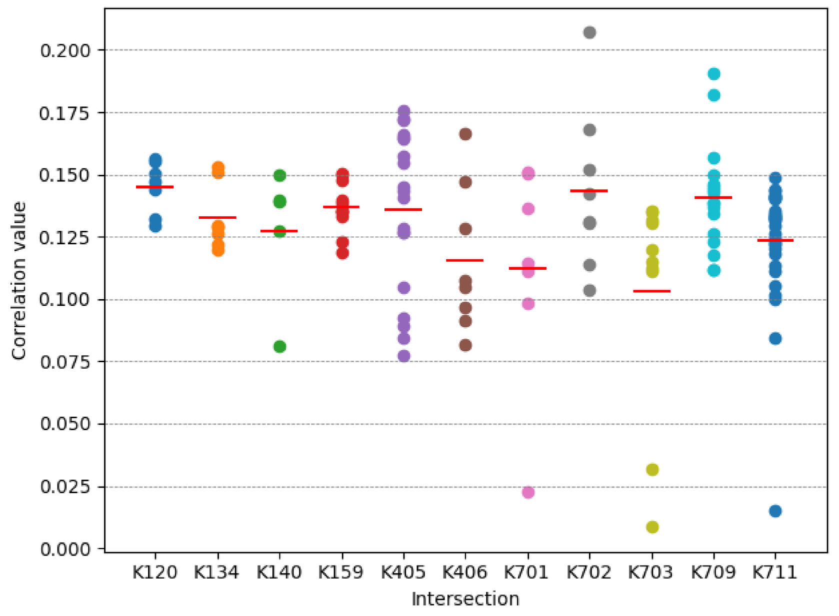

Figure 1: Correlation coefficients values for each sensor, grouped by intersection, with respect to precipitation feature data

Rain data lag	correlation mean
no lag	0.1278
15-min lag	0.1294
30-min lag	0.1296
45-min lag	0.1290

Table 1: Mean of correlation between traffic data and rain data

Test RMSE	HC with rain
no lag	3.13
30-min lag	2.99
% Improvement	4.6%

Table 2: Predicting the sensor with the highest correlation (HC) using the original and the lagged rain data

- Accuracy improved by 5.4 and 3% when adding rain to predict the highest and lowest correlated sensors
- Including temperature decreased accuracy
- Predicting the next 2.5 hours increased the accuracy by 7.3%
- Predicting the highest and lowest correlated intersections with rain did not improve accuracy

Test RMSE	Baseline	With Weather	% Improvement
HC with rain	3.15	2.99	5.4%
LC with rain	5.49	5.33	3%
HC with temperature	5.98	6.16	-3%
LC with temperature	2.23	2.26	-1.3%
HC with rain 2.5 hours	4.13	3.85	7.3%
HC intersection with rain	7.54	7.8	-3.3%
LC intersection with rain	8.34	8.25	1%

Table 3: RMSE results for predicting the sensors and intersections with the highest (HC) and lowest (LC) correlation for short-term and long-term predictions (2.5 hours)

REFERENCES

[1] ZHENG ZHAO1,WEIHAI CHEN1, XINGMINGWU1, PETER C. Y. ,CHEN2, AND JINGMENG LIU1, LSTM NETWORK: A DEEP LEARNING APPROACH FOR SHORT-TERM TRAFFIC FORECAST.

[2] PAUL A. PISANO AND LYNETTE C. GOODWIN, ARTERIAL OPERATIONS IN ADVERSE WEATHER. IN INSTITUTE OF TRANSPORTATION

[3] LYKOURGOS TSIRIGOTIS, ELENI I VLAHOGIANNI, AND MATTHEW G. KARLAFTIS, DOES INFORMATION ON WEATHER AFFECT THE PERFORMANCE OF SHORT-TERM TRAFFIC FORECASTINGMODELS?

[4] DA ZHANG AND MANSUR R. KABUKA, COMBINING WEATHER CONDITION DATA TO PREDICT TRAFFIC FLOW: A GRU BASED DEEP LEARNING APPROACH