Predicting Traffic Indexes on Urban Roads based on Public Transportation Vehicle Data in Experimental Environment

Georgi Yosifov, PhD Candidate
Assoc. Prof. **Milen Petrov**, PhD
Sofia University “St. Kliment Ohridski”, Sofia, Bulgaria

July 2022, Computing Conference, London UK

computing
conference 2022

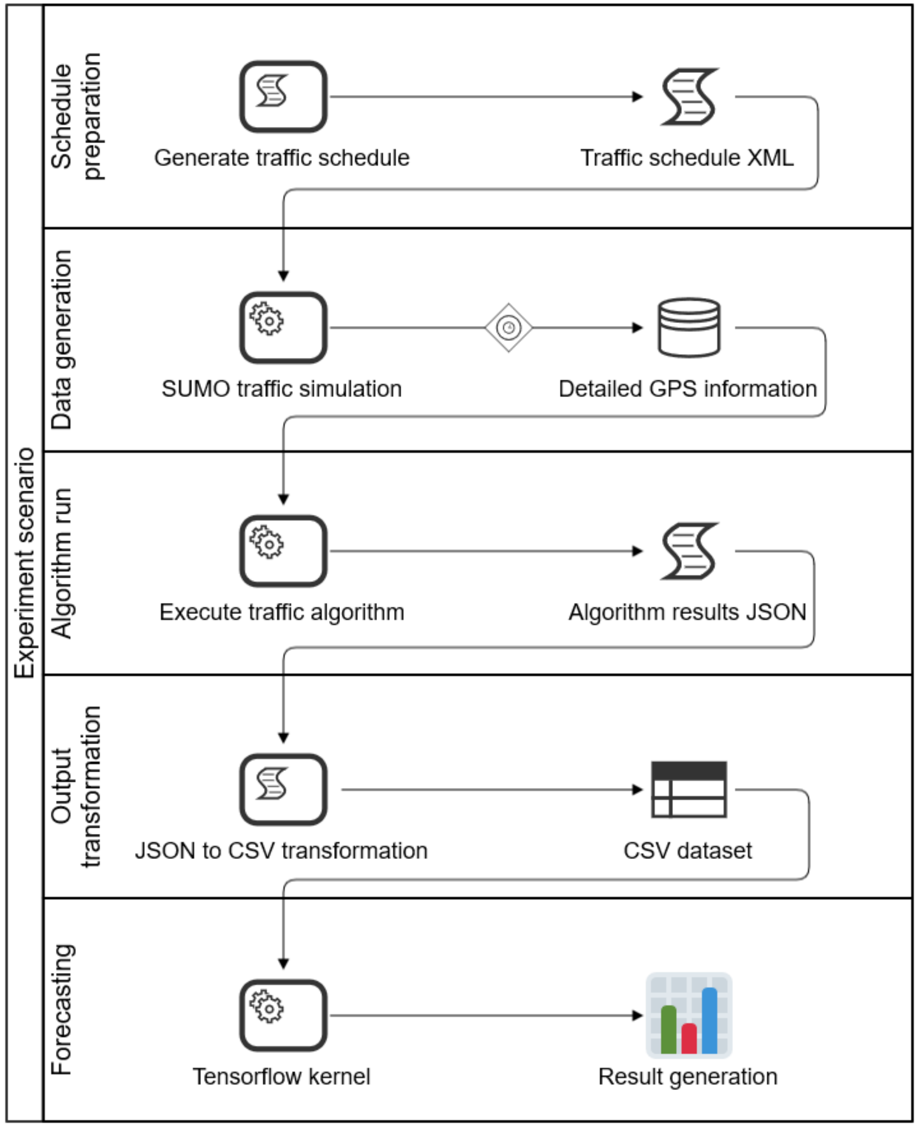
Overview

- Introduction
- Methodology
- Running and comparing the prediction models
- Conclusions
- Acknowledgements

Introduction

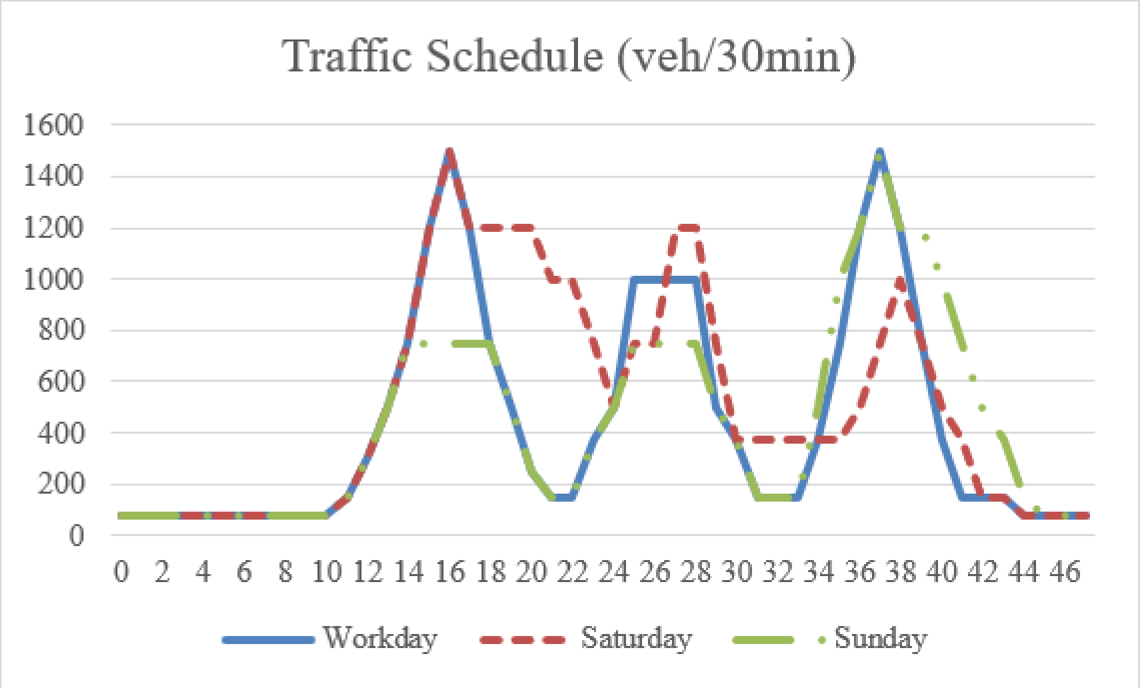
- The Traffic Indexes used in this paper are calculated based solely on positioning data gathered from periodic public transportation vehicles in an urban environment.
- A Traffic Index is defined as a discrete value from 0 to 5 which describes the traffic load for some 30-minute interval for a road segment or an entire city, where 0 is considered “no traffic” and 5 stands for “very heavy traffic”
- Simulation done by SUMO. The simulated data is chosen to be for a 1 km road stretch in Sofia, Bulgaria for a period of 365 days
- We will compare the performances of the different models used for making the predictions and finally we will present our conclusions

Methodology



computing conference 2022

First stage - generating a schedule



- Period of time - 365 days for 1 km of data
- Added ±10% randomness to the schedule
- Defaults: vehicle length - 4.3 m, minimum distance from other participants - 2.5 m, maximum acceleration - 2.9 m/s^2, deceleration - 3 m/s^2, emergency deceleration - 7 m/s^2, maximum speed 180 km/h, number of seats - 5, speed deviation - 0.1.
- With probability 0.5 - maximum speed of 80 km/h is defined. All other parameters are default.
- With a probability of 0.4 - lower acceleration parameters (2 m/s^2), and maximum speed of 40 km/h. They have an increased length of 5 m.
- With a probability of 0.1 - length of 6 meters, acceleration of 1.3 m/s^2, and a maximum speed of 30 km/h.

Second stage - SUMO simulation

- Once we have generated a new SUMO schedule file, we can run the simulation software.
- As a result, an XML file is produced containing the exact positions of all vehicles for each second. The size on disk of the file is 172GB.

Third stage - Running the algorithm on the generated data

- Applying the algorithm for evaluating the traffic indexes as per our previous research
- Generating output JSON file with indexes for every 30 minutes of the 365 days

Fourth stage - Transformation of the generated data

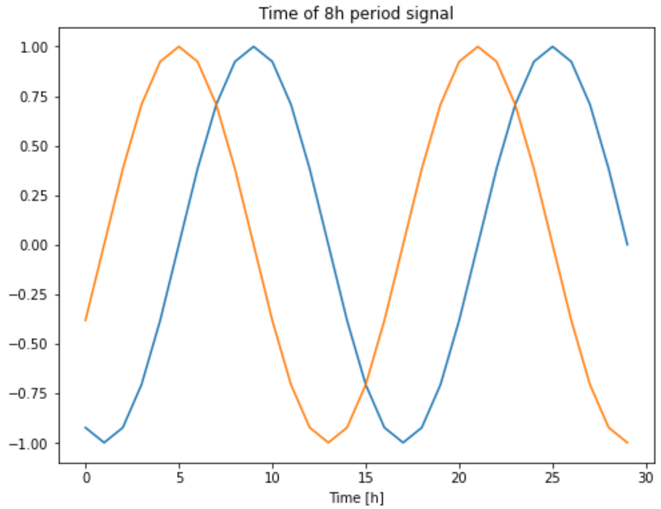
- In order to make a dataset usable by the machine learning (ML) models, a special software tool was created that reads the output of the algorithm and transforms it into a CSV file.

Fifth stage - Machine learning platform

- Prediction of the Traffic Index for the next 30 min time interval on a certain road segment.
- Prediction of Traffic Indexes in several consecutive time intervals based on a set of measurements in the past for a certain road segment.

Running the predictions – Data preparation

- The number of rows is 13,869
- Transforming the time as signal – using the UNIX time and calculating the signals for cos and sin for 24 hours and 8 hour periods
- We separate the dataset to training and the test data. The ratio chosen is 80 to 20. By comparing their mean absolute errors (MAEs), we can test our models for overfitting
- We use TensorFlow as framework

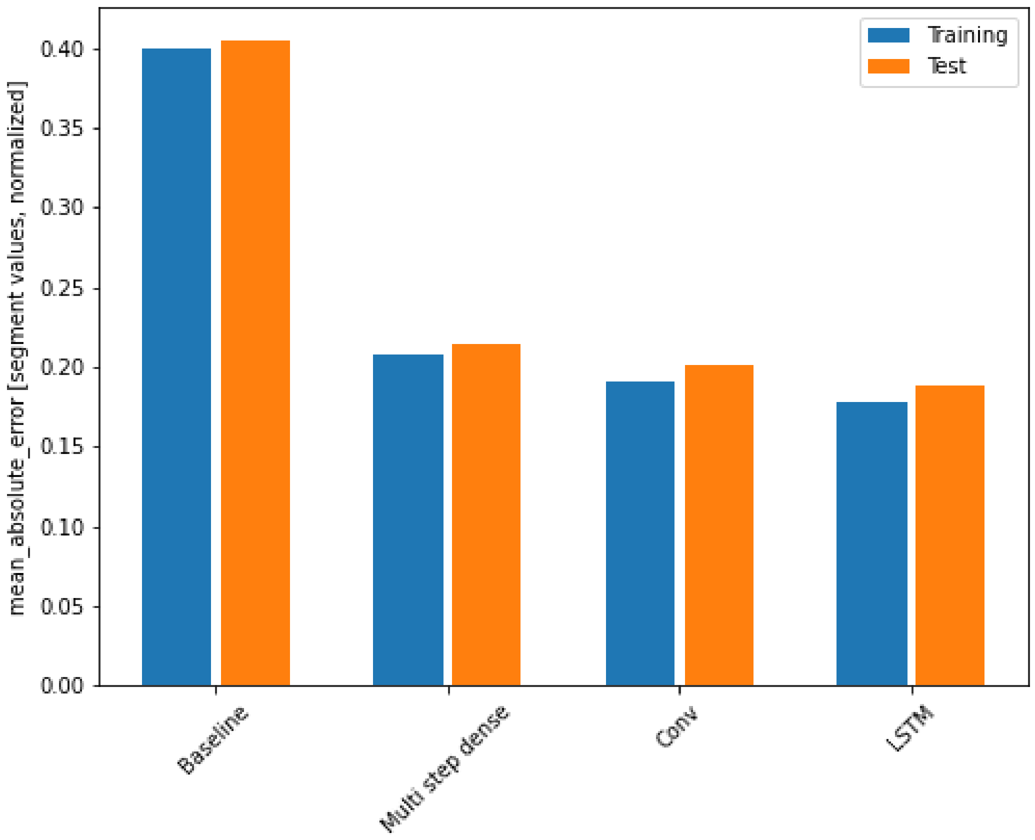


Single-step models

Single-step models provide predictions for just one Traffic Index in the future.

- Baseline - take the value in the last interval and return it as a forecast for the next, assuming that there will not be sharp declines and in-creases in the level of traffic in the two adjacent time intervals.
- Dense Neural Network – a neural network with several consecutive strongly connected layers. To make a prediction, the neural network uses the 16 previous states of the index as input.
- Convolutional Neural Network (CNN) – Similar to the previous model, but with the introduction of one Conv1D layer with 16 neurons.
- Recurrent neural network (RNN) – is a neural network known to work well with time series, as it maintains an internal state from one time step to the next. We are using Long Short-Term Memory (LSTM) network with one layer of 32 neurons.

Single-step models



Single-step models

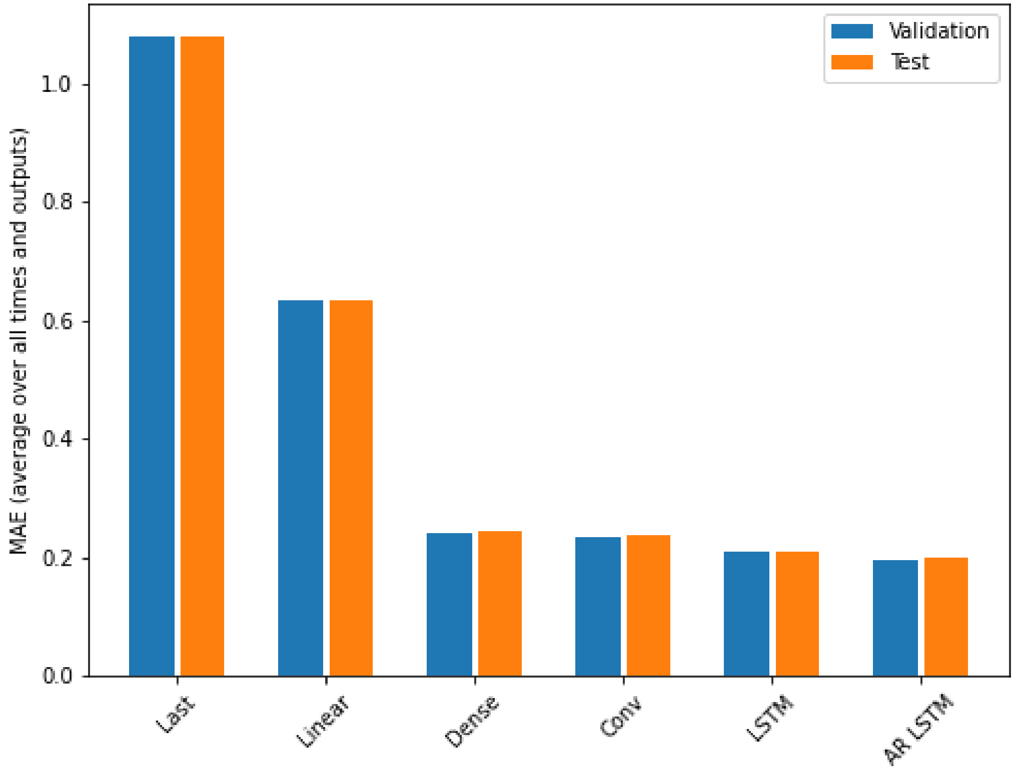
Model	Mean absolute error
Baseline	0.4052
Multi step dense	0.2146
Conv	0.2013
LSTM	0.1881

Multi-step models

Unlike single-step, multi-step models can predict several steps forward in the time series – 4 hours ahead.

- Baseline - The baseline here returns the last known result as subsequent predictions. It is expected that we get worse results here than the single-step model's baseline, as the probability of traffic volume changing in period of 4 hours is higher.
- Linear model - a neural network that uses the last reported feature as input and predicts what the next 8 will be based on a linear projection.
- Dense model - The structure of this neural network differs from the linear one in that there is a layer with 512 neurons between the input and output layers. However, this configuration is like the linear one in that, like it, it only accepts the last feature and uses it to determine the next 8 Traffic Indexes.
- CNN - For our experiment, it is configured to work with the last 16 records, and again to predict the next 8. The convolutional layer of the network is composed of 256 neurons, and the activation used is ReLU.
- LSTM - The network is configured to accumulate input data for a period of 48-time intervals backwards and to generate information for the next 4 hours (8 Indexes).
- Autoregressive RNN – in this model the result is generated in steps and each generated step is added as input for the generation of the next one.

Multi-step models



computing
conference 2022

Multi-step models

Model	Mean absolute error
Last	1.0802
Linear	0.6338
Dense	0.2429
Conv	0.2371
LSTM	0.2093
AR LSTM	0.1985

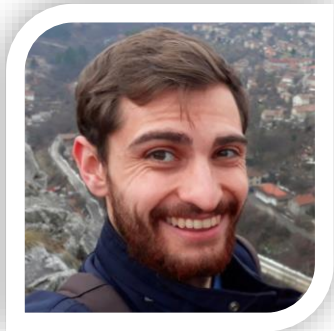
Conclusion

- In this paper we have shown how we can predict the Traffic Indexes for the next 30 minutes or the next 4 hours. For this purpose, we created a system that generates SUMO traffic schedules based on user defined parameters. Then, based on a one-year schedule, we created a traffic simulation with the specialized SUMO software. We processed the output generated by the simulation and ran an algorithm for calculating Traffic Indexes for every half hour of that year in ten separate segments. By selecting one of these segments and taking the Traffic Indexes calculated sequentially, we compiled a time series dataset to use with different ML prediction models. We investigated 4 single-step and 6 multi-step models. Of the single-step models, the one with the smallest MAE was the RNN (LSTM), and of the multi-step ones, the auto-regressive RNN gave best results.
- These results could help in the creation of urban traffic control systems, as well as to be used in routing and navigation software solutions. The benefits would be great for both administrations responsible for urban mobility or individuals and businesses.

Acknowledgment

The research reported here was partially supported by "An innovative software plat-form for big data learning and gaming analytics for a user-centric adaptation of technology enhanced learning (APTITUDE)" - research projects on the societal challenges – 2018 by Bulgarian National Science Fund with contract №: KP-06OPR03/1 from 13.12.2018 and project FNI-SU-80-10-152/05.04.2021, FNI project of Sofia University "St. Kliment Ohridski" (Bulgaria) "Challenges of developing advanced software systems and tools for big data in cloud environment (DB2BD-4)".

Thank you, any questions?



Georgi Yosifov, PhD Candidate

gkjosifov@fmi.uni-sofia.bg

Thesis Subject: “Aided decision making for public transportation optimizations using Big Data”

Assoc. Prof. Milen Petrov, PhD

milenp@fmi.uni-sofia.bg

Sofia University “St. Kliment Ohridski”