Manual OE-LSTM

This document serves as a short manual for the OE-LSTM code as developed for the SSDBM paper submission Enhancing LSTM Prediction of Traffic Flow Data using Outlier Correlations.

The project contains a streamlined version of the code. The idea is that another project can use this code as an example and use their own input data to replace the input data for this project or add on top of the current code to improve it.

To run the FPD-LOF model for outlier detection, go to the FPD-LOF folder and start with the create\_fpd notebook to create the fpds. The data from the input\_data folder is used as input and the fpds are saved in the edited\_data folder. Then, to calculate the LOF scores, the perform\_lof notebook is used. The fpds from the edited\_data folder are used as input and the output is saved in the edited\_data folder as well. Finally, the stream\_merge notebook is used to merge the individual streams into one big stream per location and the output is saved in the output\_data file.

To enable easier handling of the data for the analysis, all the individual datasets are merged into one big dataframe. This is done in the dataframe\_creation folder -> create\_dataframe notebook. The completed dataframe is saved as a csv file in this folder.

The correlation analysis is performed in the calculated\_correlation folder -> correlation\_analysis notebook. This notebook used the complete dataframe as input and besides the correlation matrix, it outputs the correlation scores as a csv file in this folder.

The prediction is performed in the prediction folder. In this folder there are two prediction notebooks. The predictors notebook contains multiple experiments with different predictors and other parameters. The feature\_reduction\_prediction notebook uses the correlation scores csv file and uses it to select only features with a certain correlation threshold.

The OE-LSTM notebook contains the steps taken to enhance the multivariate LSTM with the output of the FPD-LOF analysis.