Forecasting Traffic Congestion States based on Motorway Grid Cells using Floating Car Data

Karen Schulz

Darmstadt University of Applied Sciences, Faculties Computer Science & Mathematics and Sciences

Topic relevance

Warning road users of upcoming traffic congestion can save time, money and CO_2 emissions and requires forecasting traffic congestion on road segments. The motorway network is congestion-prone, has a huge throughput and is therefore suited as study area. Models developed based on floating car data (FCD) can generally be deployed in wide spatial areas since FCD has broad spatial coverage.

Existing approaches predominantly forecast congestion states on road segments. Using FCD, single FCD observations are usually assigned to road segments with the help of a so-called map-matcher. A seldomly utilized approach assigns FCD to grid cells, which is a lot simpler but too general for segment-wise forecasts and should therefore be further developed.

Research questions

Q1 How can FCD be related to motorway segments in a grid-based setting?

How behaves the forecasting methodology in different dimensions?

Q2 Develop models for a single segment or segments in the whole grid?
Q3 Which features are important?
Q4 Which forecasting time is valuable?
Q5 Vary the evaluation performances between this work and related work?

Q6 Is the computational effort of a grid-based segment gathering approach lower than the effort of the map-matcher approach?

Literature findings

Study comparison obstacles were encountered due to no available benchmark FCD set as well as the usage of numerous unique study setups in related works that influenced the outcome of the evaluation results. As a consequence, the results of existing studies could not be reliably compared.

Data sets

The utilized FCD set had 12,000,000,000 observations in the examined time period between August 2019 and February 2020 in the observed North-Rhine Westphalia region. Figure 1 provides a sample FCD data set.

Id	*Datetime exact*	*Latitude*	*Longitude*	*Velocity (km/h)*	*Heading (°)*
1	2019-08-01 00:00:23	51.404660	7.473666	80	225
1	2019-08-01 00:01:15	51.402817	7.467361	91	80
1	2019-08-01 00:02:01	51.401985	7.465269	94	221
	...				
2	2019-08-01 00:00:58	51.654835	7.035406	107	93
	...				

Fig. 1: Sample FCD set

Gathered ground truth labels had been generated by thresholding the average velocity of the motorist population at specific points on motorways and were classified as *free-flowing* (99.7%) or *congestion* (0.3%). The generated data corpus consisted of 45,000,000 observations from 1,000 motorway segments spread across North-Rhine Westphalia.

Methodology

The FCD processing framework and the model development framework relied on a distributed infrastructure. The generated data corpus included features and ground truth labels and used the composite index of distinct five-minute time intervals, the cell id, and the binary direction class variable. Figure 2 shows how the proposed direction distinctive grid-based approach solves the issue of assigning FCD observations to motorway segments. Per cell, the main directions of travel (arrows) are obtained though the heading's mode and its counterpart.

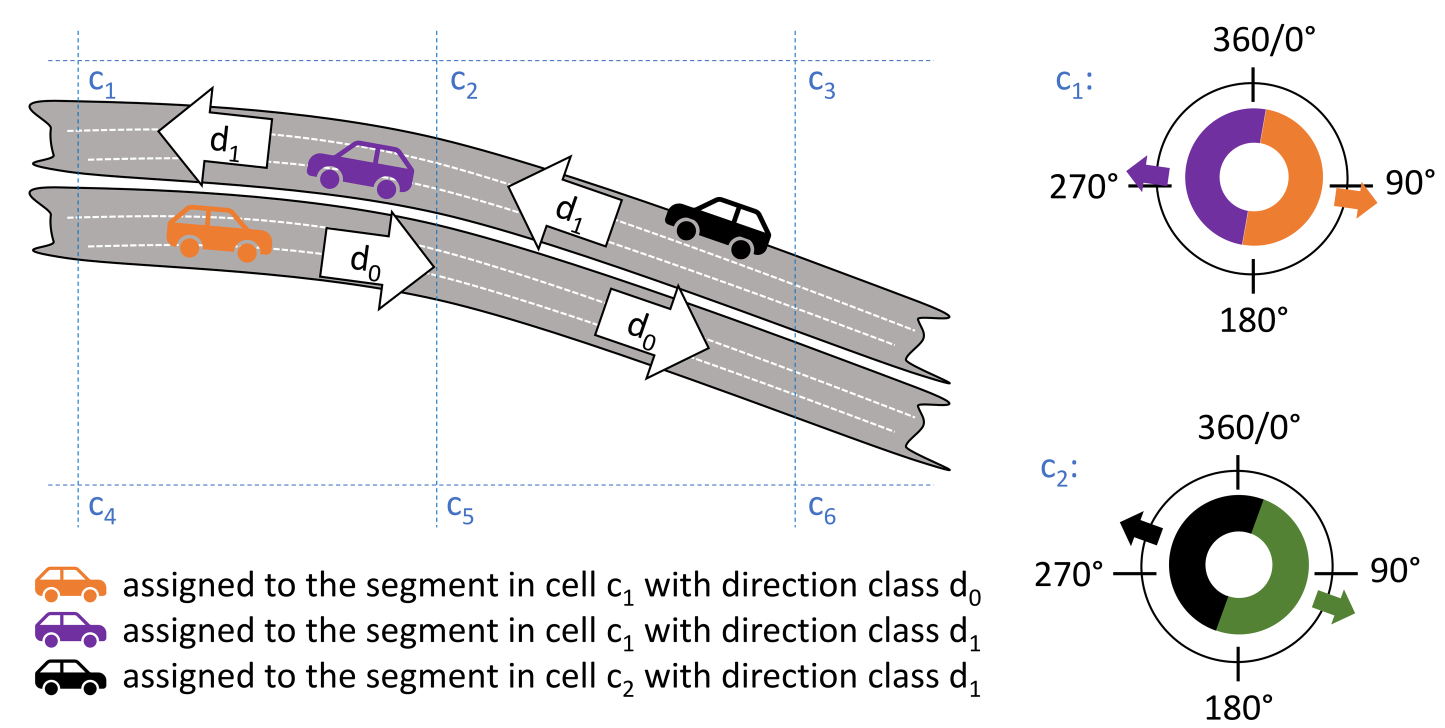

Fig. 2: Segment assignment to FCD obs. through the direction distinctive grid-based approach

The data corpus was split into a train and a test data set. Using cross-validation, temporal train-test splits were applied for models developed and deployed based on a single segment. So-called whole grid models had spatial splits based on distinct segments. The training data set was subsequently oversampled and fed into a random forest model. The bookmaker informedness served as the evaluation metric since it is unbiased against different class proportions.

Experimental results

Figure 3 shows one of the four examined motorway segments that had high *congestion* shares ($> 1\%$) along with its evaluation performance. Results show that the whole grid model had a superior five-minute forecasting performance in the observed segment.

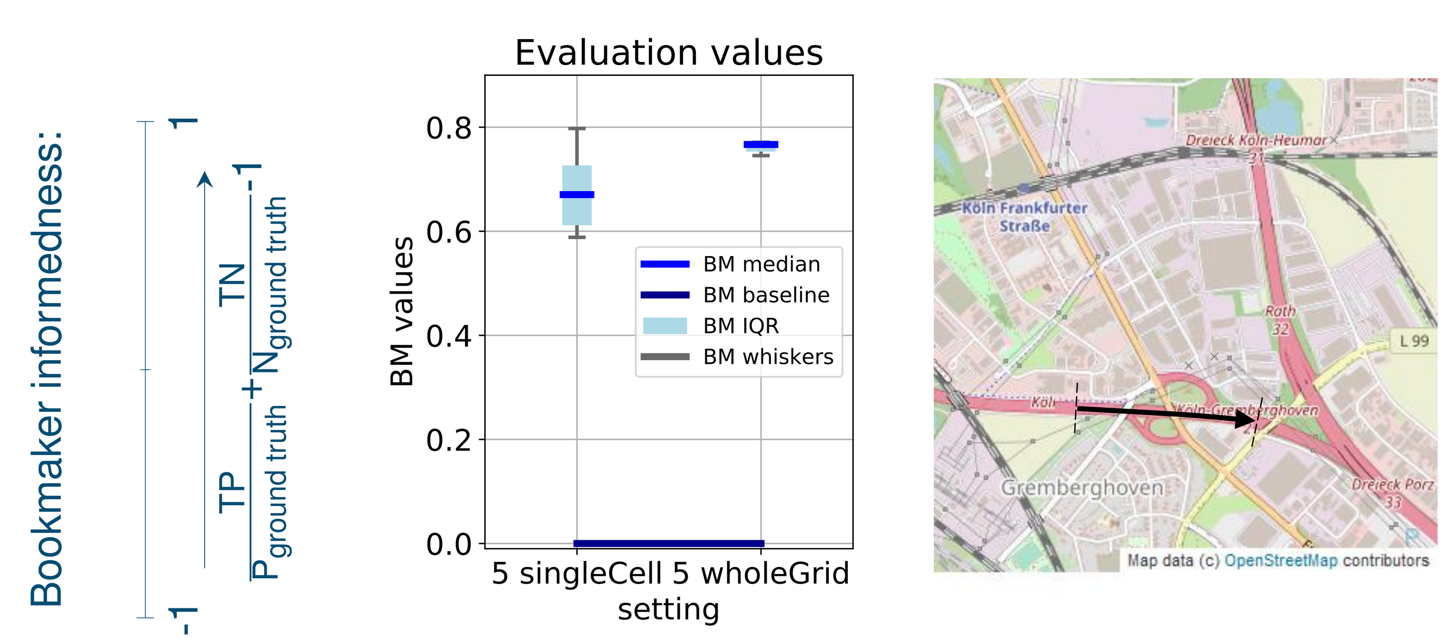

Fig. 3: Single segment model example

Velocity-related features (e.g. median velocity and 85% velocity percentile) and traffic count-related features (e.g. traffic count and traffic count of distinct motorists) played an important role in forecasting congestion states in the whole grid setting.

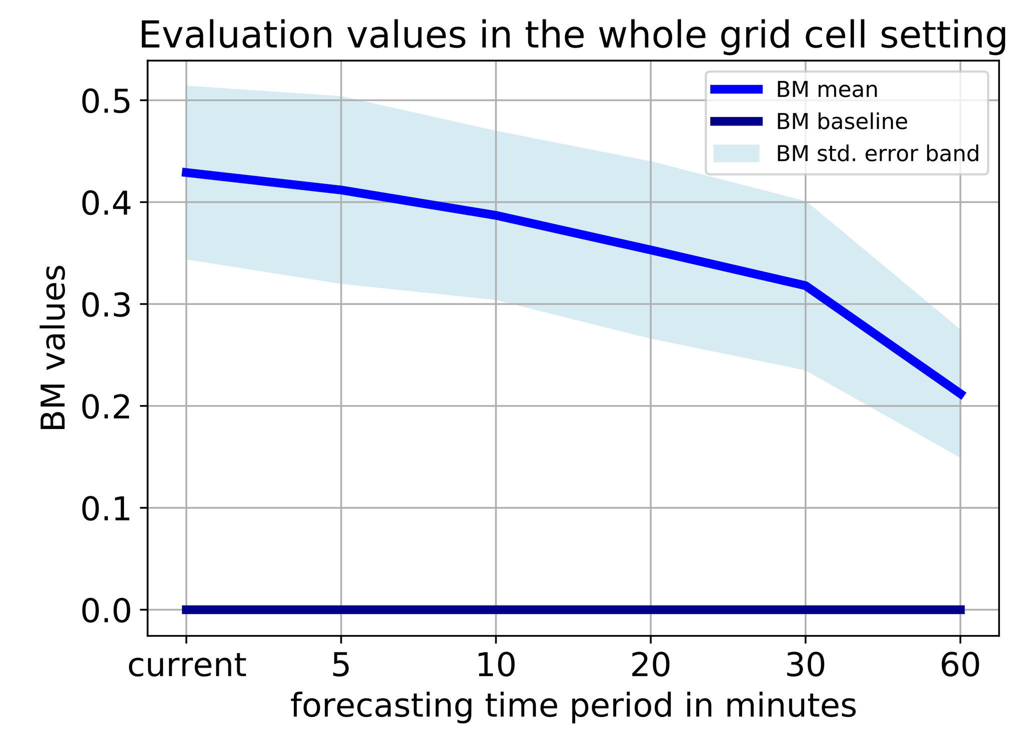

Fig. 4: Whole grid forecasting performance

Figure 4 shows the forecasting performance for time periods of up to 60 minutes in the whole grid setting. The models outperform the baseline that constantly forecasted only one congestion state.

No reliable comparison to evaluation performances of related work was possible since no scientific work with a similar study setup could be identified (cf. literature findings).

The computational time of the map-matcher and the direction distinctive grid-based segment assignment approach was compared on a smaller FCD sample (2.6 million obs.). The computational time was lowerable by more than 30% through the direction distinctive grid-based approach. Another advantage is no reliance on the additional map-matching tool that would need additional tuning and maintenance.

Conclusion

Q1 ✓ The proposed direction distinctive grid-based approach relates FCD to motorway segments

How behaves the forecasting methodology in different dimensions?

Q2(✓) Indication: Whole grid model for highly congested segments
Q3 ✓ Important features: Velocity-related and traffic count-related
Q4 ✓ Valuable forecasting time of up to 60 minutes
Q5 ? No reliable comparison with existing works possible, additionally map-matching the huge data corpus out of scope

Q6(✓) Indication: Computational effort lowerable by $>$30% & no additional tool needed

Improving the direction class could enhance the feature's exactness and also the forecasting performance. Establishing a metric that accounts for the various dimensions of different study setups should be pursued to relate model performances of scientific approaches.

Acknowledgements

This master thesis was developed in close cooperation with the data analytics team of the company urban mobility innovations (`https://www.umi.city/`). The VIZ.NRW data set and details regarding it were gladly received from Volker Gronau and Tilo Voigt of Landesbetrieb Straßenbau North-Rhine Westphalia (`https://www.strassen.nrw.de/de/`) and utilized for extracting ground truth labels.