World Academy of Science, Engineering and Technology
International Journal of Transport and Vehicle Engineering
Vol:16, No:01, 2022

Data Refinement Enhances The Accuracy of Short-Term Traffic Latency Prediction

Authors : Man Fung Ho, Lap So, Jiaqi Zhang, Yuheng Zhao, Huiyang Lu, Tat Shing Choi, K. Y. Michael Wong

Abstract : Nowadays, a tremendous amount of data is available in the transportation system, enabling the development of various machine learning approaches to make short-term latency predictions. A natural question is then the choice of relevant information to enable accurate predictions. Using traffic data collected from the Taiwan Freeway System, we consider the prediction of short-term latency of a freeway segment with a length of 17 km covering 5 measurement points, each collecting vehicle-by-vehicle data through the electronic toll collection system. The processed data include the past latencies of the freeway segment with different time lags, the traffic conditions of the individual segments (the accumulations, the traffic fluxes, the entrance and exit rates), the total accumulations, and the weekday latency profiles obtained by Gaussian process regression of past data. We arrive at several important conclusions about how data should be refined to obtain accurate predictions, which have implications for future system-wide latency predictions. (1) We find that the prediction of median latency is much more accurate and meaningful than the prediction of average latency, as the latter is plagued by outliers. This is verified by machine-learning prediction using XGBoost that yields a 35% improvement in the mean square error of the 5-minute averaged latencies. (2) We find that the median latency of the segment 15 minutes ago is a very good baseline for performance comparison, and we have evidence that further improvement is achieved by machine learning approaches such as XGBoost and Long Short-Term Memory (LSTM). (3) By analyzing the feature importance score in XGBoost and calculating the mutual information between the inputs and the latencies to be predicted, we identify a sequence of inputs ranked in importance. It confirms that the past latencies are most informative of the predicted latencies, followed by the total accumulation, whereas inputs such as the entrance and exit rates are uninformative. It also confirms that the inputs are much less informative of the average latencies than the median latencies. (4) For predicting the latencies of segments composed of two or three sub-segments, summing up the predicted latencies of each sub-segment is more accurate than the one-step prediction of the whole segment, especially with the latency prediction of the downstream sub-segments trained to anticipate latencies several minutes ahead. The duration of the anticipation time is an increasing function of the traveling time of the upstream segment. The above findings have important implications to predicting the full set of latencies among the various locations in the freeway system.

Keywords : data refinement, machine learning, mutual information, short-term latency prediction
Conference Title : ICCT 2022 : International Conference on Computational Transportation
Conference Location : Tokyo, Japan
Conference Dates : January 07-08, 2022