Dette resumé er udgivet i det elektroniske tidsskrift
Artikler fra Trafikdage på Aalborg Universitet
(Proceedings from the Annual Transport Conference at Aalborg University)
ISSN 1603-9696
https://journals.aau.dk/index.php/td

Machine Learning Methods for Transportation under Uncertainty

Inon Peled, Francisco C. Pereira, Justin Dauwels, Yu Jiang, Kelvin Lee, Daniele Gammelli, Filipe Rodrigues and others; *Corresponding Author: inonpe@dtu.dk*
Technical University of Denmark

Abstract

Transportation is rife with uncertainty, e.g., due to sudden disruptions and incomplete knowledge. Properly modelling this uncertainty is thus crucial for effective Transport practitioning. Fortunately, Transportation is also a rich source of data, from which Machine Learning models can extract useful patterns. This extended summary succinctly describes several joint works on Machine Learning methods for handling uncertainty in Transportation. In these works, we find that recent technological advances can alleviate the degradation of data-driven prediction models under road incidents, for which we offer a dedicated framework. We also advise to explicitly model an inherent limitation in Transportation demand observations, for which we offer two non-parametric alternatives. For dynamic operation of shared mobility services, we demonstrate the benefits of preserving a full uncertainty structure of demand, and we also quantify the relationship between predictive quality and subsequent service optimization.

1. Introduction

Traffic Management Centers and Intelligent Transportation Systems rely on predictive models of short-term and long-term traffic. The models are often based on Machine Learning (ML) methods to automatically detect patterns in data. For instance, ML is widely used in Intelligent Transportation Systems, real-time incident management, autonomous mobility, international freight shipping, Public Transport, and fleet management.

In these and many other examples, *uncertainty* exists in the modelled phenomenon, whether because of incomplete knowledge or innate non-determinism. For example, highly crowded events can disrupt normal Public Transport usage, road incidents vary considerably in duration and impact, and inclement weather can hinder traffic inconsistently. Transportation data itself can also be noisy, e.g., because of omission, limited observability, inaccuracy or inherent bias. In the following works, we address some gaps in research on ML methods for Transportation under uncertainty. For brevity, we include only references to the summarized works, and note that these contain further references to other related works.

2. Simulation-based Model Adaptation during Incidents [1]

When an incident happens, the correlation structure between response and explanatory variables changes abruptly, in a manner which is unique to the incident characteristics. As such, there are advantages to treating incidents separately from other atypical conditions. At first sight, it may seem worthwhile to pregenerate sufficiently many incident simulations offline, and train an ML model on them. However, doing so is impractical, because every incident involves too many varying parameters.

Nowadays, however, In-Vehicle Monitor Systems (IVMS) give real-time access to incident details, allowing for real-time incident simulations that can be used for ad-hoc model training. For this, we propose the QTIP framework, as illustrated in Figure 2.1. QTIP adapts $M_{ordinary}$, a traffic model for ordinary conditions, into $M_{abnormal}$, an incident-specific model, regardless of specific model form. The adaptation uses data from real-time simulations, based on IVMS information from the incident, to cover a range of unknown incident parameters.

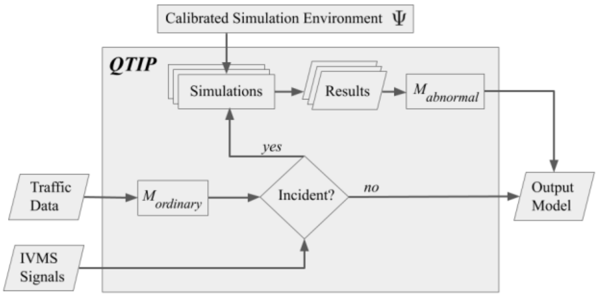

Figure 2.1: QTIP framework.

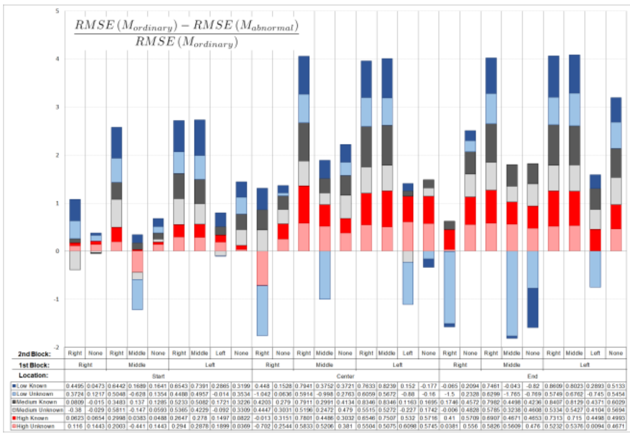

Figure 2.2: Performance of model obtained through QTIP.

We experiment QTIP in different incident scenarios in a simulated environment of the Hillerød Motorway, calibrated on data from the Danish Road Directorate. Each ”ground-truth” simulation involves one of three road demand levels (low, medium, high) and either 1 or 2 road blocks in one of three locations on the road, where each block is positioned in one of three lanes. For stable results, we replicate each scenario 5 times by stochastically perturbing the underlying Origin-Destination demand matrix. In all cases, QTIP is not given the demand, but only the 1 min average speed on the road, and optionally also the blocks' exact locations.

For models $M_{ordinary}$ and $M_{abnormal}$, we independetly experiment with Linear Regression, Deep Neural Networks, and Gaussian Processes. $M_{ordinary}$ is then obtained by training on multiple, incident-free simulations for each demand level. For each incident scenario, QTIP generates 100 ad-hoc incident simulations and uses their data to train a 2-piece $M_{abnormal}$, where the first piece pertains to the first 6 critical minutes of the incident, and the second piece pertains to the time until the incident is cleared.

Linear Regression turns out to be the best performing model. For this model type, Figure 2.2 shows the superior performance of QTIP's $M_{abnormal}$ vs. $M_{ordinary}$, as measured through Rooted Mean Squared Error. On average, $M_{abnormal}$ performs 28.74% better than $M_{ordinary}$. Our findings thus suggest that through QTIP, the long-standing problem of instantaneous model adaptation under incidents becomes more tractable. In particular, we find that pre-calibrated simulations can sufficiently account for variability in unknown incident parameters, thereby evading the major challenge of online calibration.

3. Modelling Latent Mobility Demand

Transport service providers rely on proper models of mobility demand, to make decisions coherently with user behavior and needs. Such models often use data from past service usage, which is inherently limited (”censored”) by available service supply, so that the actual demand is sometimes unknown (”latent”). We next study two non-parametric modelling methods for handling such censored data.

3.1 Censored Gaussian Processes [2]

A well-known, parametric model for censored regression is Tobit, where the latent variable y depends on the covariates X linearly with Gaussian noise, and censorship occurs at known thresholds. Previous works extend the Tobit likelihood to Gaussian Processes (GPs): a nonparametric modeling method for learning a distribution on functions from X to y. We further extend this method to time-varying and stochastic censorship thresholds. For the GP covariance matrix, we use a combination of several kernels (RBF, periodic and Matern) and infer their parameters via Expectation Propagation.

We experiment our Censored GP (CGP) methodology in several case studies: (1) a synthetic, sinusoidal dataset, where points near the sinusoidal peaks are censored to within 70%-80% of their original value. (2) Real-world measurements of acceleration during a motorcycle crash, for which we censored either 10%, 50% or 90% of the observations to within either 0%-33%, 34%-66%, 67%-100% of their original values. (3) Bike rentals and returns from a real-world bike-sharing provider in Copenhagen, where we treat rentals as demand, and choose which of them are censored per supply shortages; for $c = 0, 0.1, \dots, 1$, each censored observation is *c* times its original value. (4) Taxi pickups and dropoffs, where we treat pickups as demand, and censor them stochastically per available supply, again using $c = 0, 0.1, \dots, 1$.

In addition to CGP, we also fit non-censored GP model (NCGP), and a censorship-aware GP model (NCGP-A) that first discards of censored observations. We then evaluate all models in terms of Rooted Mean Squared Error (RMSE) and Coefficient of Determination (R^2). Example results for the taxi case study appear in Figure 3.1. Overall, the experimental results from all case studies highlight how standard regression models are prone to bias under data censorship, whereas CGP can yield consistent predictions even under severe censorship. This further supports a key message of our work: it is better to deal with data censorship via more *knowledgeable* models than via common data cleaning techniques.

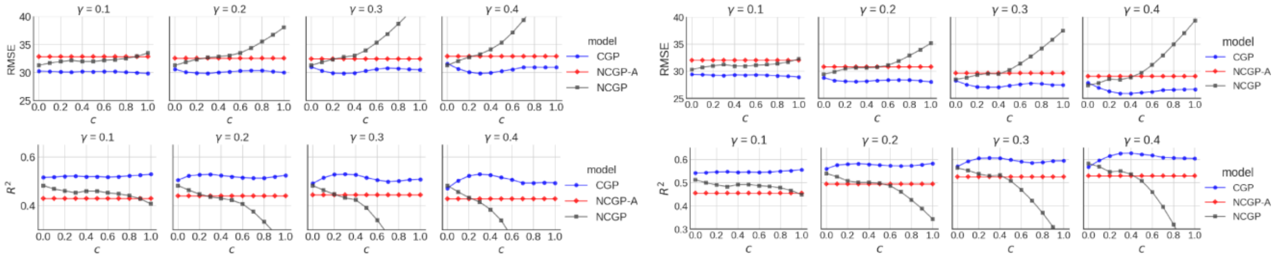

Figure 3.1: predictive quality for experiments with taxi data, when evaluating on either the entire test data (left) or only non-censored observations (right), where the actual values are known.

3.2 Censored Quantile Regression Neural Networks [3]

While Gaussian Processes allow for a flexible, non-parametric fit, they still impose a Gaussian assumption on the latent distribution and have somewhat limited scalability. To overcome these limitations, we next study Censored Quantile Regression Neural Networks (CQRNN), to estimate the latent distribution while accounting for data censorship.

For any $0 < \theta < 1$, the θ'th quantile of a Cumulative Distribution Function (CDF) is the smallest real number at which the CDF equals θ. In Quantile Regression (QR), multiple quantiles of the CDF are modeled, thereby allowing to approximate the full distribution. Censored Quantile Regression (CQR) extends on this by accounting for censorship thresholds through various techniques, and few CQR works use Neural Networks for nonparametric CQR. As far as we know, our work is the first to apply CQRNN to Transport.

As illustrated in Figure 3.2, we construct NNs independently for each θ, whose values we choose next in the experiments. Throughout the experiments, we vary the Hidden layers and the activation function of the Aggregate Outputs layer. For training, we also devise a dedicated log-likelihood function.

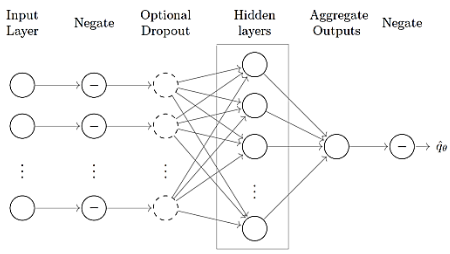

Figure 3.2: CQRNN architecture.

Our experiments begin with three commonly used synthetic baseline datasets, for which we use $\theta = 0.05, 0.5, 0.95$. We show that CQRNN outperforms censorship-unaware QRNN and nonparametric CQRNN yields better quantile estimates vs. parametric Tobit. Next, we experiment with two real-world datasets: the same bike-sharing data as in the CGP work, and daily rentals of shared Electric Vehicles. For these, we use $\theta = 0.05, 0.95$ and measure CQRNN performance for increasingly complex architectures. The results suggest that more complex architecture improves CQRNN performance, and demonstrate the advantage of modelling a full predictive distribution vs. estimating only several of its moments.

4. Predictive Optimization of Demand-Responsive Public Transport

Whereas the itineraries of traditional bus services are often pre-fixed, the itineraries of future Public Transport (PT) will be dynamically adapted in real-time, e.g., per fluctuations in mobility demand. Our next two works deal with using mobility demand predictions for effective PT route optimization.

4.1 Framework for Optimized Routing based on Demand Forecasts [4]

We first devise and evaluate an online predictive optimization framework for dynamic operation of shared transit services, as illustrated in Figure 4.1. On the travel demand side, the framework estimates marginal distributions of Origin-Destination pairs via Quantile Regression and joins them via a Gaussian copula. On the supply side, the framework uses a linear programming formulation to stochastically optimize the service per the joint distribution.

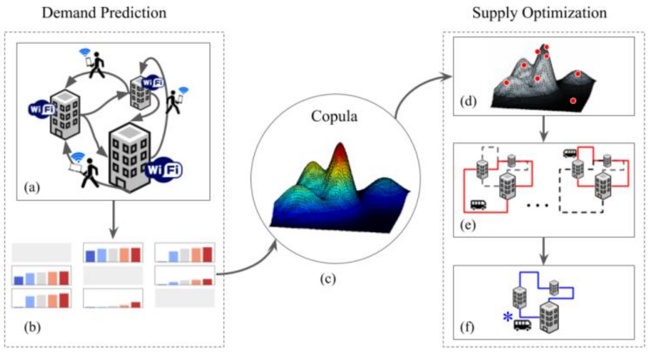

Figure 4.1: The proposed predictive optimization framework.

Using QR, we preserve a full structure of uncertainty in the demand, which in turn allows for more informed optimization of the service. Further, by stochastically constructing prediction-based demand scenarios, our optimization method is optimal in expectation. Also, contrary to most existing studies on scenario-based optimization, our framework generates the scenarios dynamically in real time.

We experiment the framework in a case study of a hypothetical shuttle service, where mobility demand is reflected by people's movements in the Lyngby campus of the Danish Technical University. These movements are reconstructed from actual data on WiFi usage in the campus. On the demand prediction side, we experiment with $\theta = 0.05, 0.1, \dots, 0.95$ and several QR model types: Historical Percentiles, Linear QR, Deep QR, and Gradient Boosting QR. We fit these models either independently for each OD pair and θ, or as multivariate models for multiple OD pairs and θ's together. We then obtain that multivariate Gradient Boosting QR is the best demand prediction model.

On the service optimization side, we devise a linear programming formulation that corresponds to a minimum cost flow, as illustrated in Figure 4.2. Given a joint demand distribution, we solve this formulation for independent samples from the distribution, and finally pick the most frequent solution as the output of the framework. We compare the quality of this output against two conventional optimization strategies – median estimates and robust (i.e., worst-case) optimization – and obtain that our framework mostly outperforms these point estimation methods. This suggests that route optimization can utilize a full structure of demand uncertainty better than just point estimates of this uncertainty.

4.2 Effects of Prediction Inaccuracy on Subsequent Optimization [5]

Uncertainty in prediction models yields residuals, namely, differences between predictions and actually observed values. Although residuals are often assumed to be identically and Normally distributed, they can sometimes be non-Gaussian and vary considerably in their central moments. Also, depending on data size and quality, it might also be impractical to fit a model with Gaussian residuals. While multiple methods have been devised for detecting deviations from normality, few works study the impact of such deviations in the context of predictive Transport optimization.

To address this gap, we conclude with a work on the relationship between accuracy in mobility demand predictions and effective operation of demand-responsive PT. We do so using an experimental case study on PT trips in Copenhagen, based on actual Rejsekort records, wherein we purposely refraining from using any specific prediction models. Instead, we simulate errors in demand predictions by sampling from a variety of distributions, and use the inaccurate predictions to simulate and optimize demand-responsive PT fleets. For service optimization, we use a formulation applicable to Mobility-on-Demand without assuming pre-booking, so that demand is observed through real-time boarding and alighting.

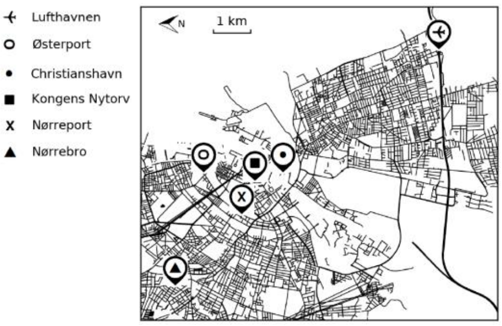

Figure 4.3: Stations in prediction accuracy case study.

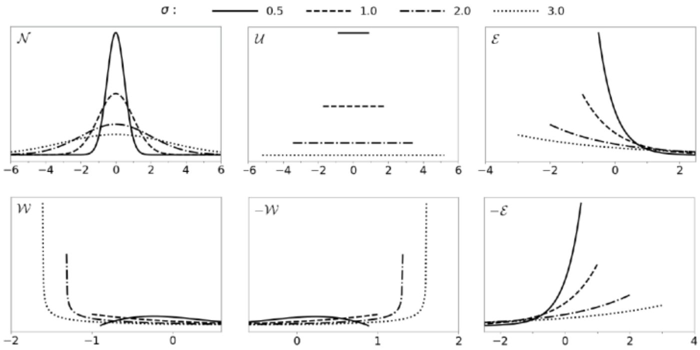

Figure 4.4: Error distributions for different standard deviations (σ).

In this case study, the observations are hourly Origin-Destination counts between 6 main PT stations, as in Figure 4.3. First, we experiment with several error distributions – Gaussian, Uniform, Exponential and Weibull – that cover a wide range of central moments, as illustrated in Figure 4.4. We shift all distributions to have zero mean and then experiment with different standard deviations. For each distribution, we obtain predictions by drawing samples and adding them to the ground truth observations. Finally, for each set of demand predictions, we simulate demand-responsive PT with varying no. of buses, bus capacity, and percentage of dynamically routed buses.

The results suggest that the optimized fleet performance is mainly affected by the presence of infrequently large prediction errors and the skew of the error distribution. In particular, non-Gaussian error distributions with negative skew and no bias can yield better fleet optimization vs. a Gaussian error distribution with the same standard deviation. For this case study, we also obtain that dynamic routing reduces trip time by at least 23% vs. classic, static routing – an estimated yearly gain of 6M DKK in Value of Travel Time.

References

[1] Peled, I., Kamalakar, R., Lima Azevedo, C. M., Pereira, F. C. *QTIP: Quick simulation-based adaptation of traffic model per incident parameters*. Journal of Simulation, 2020. DOI: 10.1080/17477778.2020.1756702

[2] Gammelli, D., Peled, I., Rodrigues, F., Pacino, D., Kurtaran, H. A., Pereira, F. C. *Estimating latent demand of shared mobility through censored Gaussian Processes.* Transportation Research Part C: Emerging Technologies, 2020. DOI: 10.1016/j.trc.2020.102775

[3] Peled, I., Rodrigues, F., Pereira, F. C., *Modeling Censored Mobility Demand through Quantile Regression Neural Networks*. arXiv, 2020.

[4] Peled, I., Lee, K. Jiang, Y., Dauwels, J., Pereira, F. C. *Online Predictive Optimization Framework for Stochastic Demand-Responsive Transit Services*. arXiv, 2019. arXiv:1902.09745

[5] Peled, I., Lee, K. Jiang, Y., Dauwels, J., Pereira, F. C. *Curb Your Normality: On the Quality Requirements of Demand Prediction for Dynamic Public Transport*. arXiv, 2020. arXiv:2008.13443