Efficiency of Travel Time Prediction

Chiwei Yan[a,*], James Johndrow[b] and Dawn Woodard[c]

[a] Department of Industrial and Systems Engineering, University of Washington, Seattle
chiwei@uw.edu

[b] Department of Statistics and Data Science, Wharton School, University of Pennsylvania

[c] Uber Technologies

[*] Corresponding author

Extended abstract submitted for presentation at the 11th Triennial Symposium on Transportation Analysis conference (TRISTAN XI)
June 19-25, 2022, Mauritius Island

April 3, 2022

Keywords: travel time prediction, statistical efficiency, spatial granularity

1 INTRODUCTION

Geospatial (maps) technologies underlie a broad spectrum of modern mobile applications. For example, consumer-facing navigation applications (such as Google Maps and Waze) provide recommended routes along with associated times, as well as turn-by-turn navigation along those routes. Geospatial technologies are also the foundation of decision systems for ride-sharing (such as Uber, Lyft, Didi Chuxing and Ola) and delivery platforms (such as UberEats and Doordash). For example, riders on these platforms are presented with expected pickup time and time to destination, and drivers are provided with turn-by-turn navigation. Matching and pricing decisions on these platforms also heavily rely on mapping inputs to optimize for efficiency and reliability (Yan *et al.*, 2020).

An important geospatial technology is prediction of the time required for a driver (or biker, or pedestrian) to travel a *particular* route in the road network. Two general classes of approaches have been proposed for this travel time prediction problem: (1) approaches based on predicting traffic and travel time at the level of road segments and turns, and aggregating across the route ("*segment-based approaches*"); and (2) approaches that use generic information about the origin, destination, and route to predict the travel time ("*route-based approaches*"). Both of these methods leverage location data traces from past vehicle trips in the road network, typically gathered (with permission) from users of the particular application, such as a consumer-facing navigation service.

Though many variations of the these methods have been proposed in the literature and in practice, there has been no rigorous theoretical comparison of the accuracy of these two approaches. Empirical studies have in many cases drawn opposite conclusions. To fill this gap, we conduct first theoretical analyses comparing these two classes of methods in terms of their predictive accuracy as a function of the training data sample size (the statistical efficiency).

Summary of Contributions and Results

- **Modeling.** We introduce a modeling framework for the travel time estimation problem. This framework uses general priors on mean travel times and allows arbitrary road network and spatial correlation structure among segment travel times, and thus lays a theoretical foundation for analyzing the accuracy of different travel time estimation methods. Under this framework, we formally define a family of segment-based estimators and route-based estimators that resemble many practical estimators proposed in the literature and used in practice. Furthermore, we explicitly characterize the optimal estimators within each family in terms of minimizing integrated risks based on squared error loss.

- **Finite Sample Analysis.** Under any amount of samples, when segment travel times are non-negatively correlated spatially within the route, we show that the integrated risk (based on squared error loss) of the optimal segment-based estimator is always lower than that of a special case of the optimal route-based estimator where one is only allowed to use historical trips that traverse the exact same route as the predicting route. We show that the non-negative correlation assumption, though being not quite restrictive in practice, is necessary for the result to hold.
- **Asymptotic Analysis.** To achieve more general results regarding their comparison, we extend our analysis to an asymptotic setting where the number of trip observations grows as the size of the road network increases and trip observations are sampled randomly from a generic route distribution. We give conditions under which a large family of simple segment-based estimators can deliver integrated risk of smaller order than that of the optimal route-based estimator when the road network size is large enough. This family of estimators only requires information of the prior means of segment travel times and encompasses popular estimators used in practice. Intuitively, the conditions require that data accumulates faster on each segment of the route than on "similar" routes in a defined neighborhood of the predicting route times the length of the predicting route (number of segments on the route). We show that such conditions hold naturally in a grid network when considering a route-based estimator that uses all historical routes that are similar in origin and destination to those of the predicting route. We also give explicit rates of the integrated risk of this family of simple segment-based estimators as the grid size grows.

2 Brief Overview of Model and Analysis

We give a brief overview of our model, analysis, and two of the key results. Note that many results are omitted or presented in a simplified way due to page limit, the complete set of results can be found at https://arxiv.org/abs/2112.09993.

We consider a setting where we are given N historical trips on an arbitrary road network $(\mathcal{V}, \mathcal{S})$ where $\mathcal{V}$ is a vertex set and $\mathcal{S}$ is an edge (road segment) set. Let $y_1, \ldots, y_N$ be the routes for each trip. Let $[N]$ denote the set $\{1, \cdots, N\}$. Each route consists of a sequence of road segments $s \in \mathcal{S}$. Note that road segment here is defined generally that could include a segment along with a particular direction of traversing that segment and a following turn direction. To simplify the notation, we use $y_{[N]} := \{y_n\}_{n \in \{1,\ldots,N\}}$. Let $T_{n,s}$ be the travel time on segment $s \in y_n$ for the n^{th} observed trip. We give the following assumption regarding the generative process of $T_{n,s}$.

Assumption 1 *We make the following assumptions about $T_{n,s}$,*

1. *For the n^{th} trip, $\{T_{n,s}\}_{s \in y_n}$ are drawn from a distribution with means $\{\theta_s\}_{s \in y_n}$ and covariances $\{\sigma_{st}\}_{s,t \in y_n}$, where $\sigma_{ss} = \sigma_s^2$ is the variance of the travel time on segment s.*
2. *For any $n \neq n'$ and any $s \in y_n, t \in y_{n'}$, $T_{n,s}$ and $T_{n',t}$ are independent conditional on θ_s, θ_t.*
3. *$\{\theta_s\}_{s \in \mathcal{S}}$ are drawn independently from a distribution with population mean θ and variance τ^2.*

We first consider a family of *segment-based estimators.* Intuitively, this family of estimators uses segment-level traversal data to estimate mean segment travel times and then aggregate over all the segments on a route to obtain an estimate for the mean route travel time. One typical way of estimating mean segment travel time is by solving a regularized regression (e.g., ridge regression) problem: $\min_{\hat{\theta}_s} \sum_{n:s\in y_n}(\hat{\theta}_s - T_{n,s})^2 + \lambda(\hat{\theta}_s - \theta)^2$ where $\lambda \geq 0$ is a regularization parameter that helps to "shrink" the estimate towards the prior mean θ if there is not much traversal data on segment s. Let $N_s := |\{y_n : s \in y_n, n \in [N]\}|$ denote the sample size of traversals on segment s. It is not hard to check that $\hat{\theta}_s$ here has a closed form: $\hat{\theta}_s = (\lambda/(\lambda + N_s))\theta + (N_s/(\lambda + N_s))(\sum_{n:s\in y_n} T_{n,s}/N_s)$. Generalizing this observation, we now define a family of segment-based estimators $\hat{\Theta}_y^{(\text{seg})}$.

Definition 1 (Segment-Based Estimator) *A segment-based estimator takes the form*

$$\hat{\Theta}_y^{(\text{seg})} := \sum_{s\in y} \hat{\theta}_s,$$
$$\hat{\theta}_s := (1 - \phi_s(N_s))\theta + \phi_s(N_s)\frac{\sum_{n:s\in y_n} T_{n,s}}{N_s},$$

for some $\phi_s(N_s) : \mathbb{Z}_{\geq 0} \mapsto \mathbb{R}$ such that $\phi_s(0) = 0$ and define $\phi_s(0)/0 = 0$, for all $s \in y$.

In other words, $\Theta_y^{(\text{seg})}$ is the summation of segment-level estimators $\hat{\theta}_s$ that are constructed using a weighted average of the sample mean and the population mean of the prior distribution, where the weight is sample size dependent. One would typically expect the weights $\{\phi_s(N_s)\}_{s\in y} \in [0, 1]$ to converge to 1 as the sample size N_s grows to infinity. Such behavior ensures consistency of the estimator, though this is not required for our analysis. The aforementioned estimator based on ridge regression takes the form of $\phi_s(N_s) = N_s/(\lambda + N_s)$. Another example of such $\phi_s(N_s)$ is the Bayes estimator $\phi_s(N_s) = \tau^2 N_s/(\tau^2 N_s + \sigma_s^2)$ when the segment travel times are independently distributed, i.e., $\sigma_{st} = 0, \forall s \neq t$ (see e.g., Gelman *et al.* (2013)). Note that one could also consider a more general family of segment-based estimators that uses traversal data on other road segments in $\mathcal{S}$ other than s to predict $\hat{\theta}_s$, e.g., the Bayes estimator under the general correlation structure. Here we are going to put a disadvantage to the segment-based estimator by excluding such options in Definition 1. In other words, the accuracy of the segment-based estimators we analyze can be considered as upper bounds. Importantly, from a practical point of view, such simplification of segment-based estimators is also aligned with industry practice of implementing segment-based methods (Derrow-Pinion *et al.*, 2021).

We then consider a family of *route-based estimators* which uses route-level traversal data to estimate the travel time on a new route y. Denote $\delta(y) \subset y_{[N]}$ as a subset of historical routes which represents the neighbor of route y. For example, $\delta(y)$ can be historical routes that share the same or similar origin and destination (but possibly with different sequence of segments) as y. These neighboring routes are representative observations to estimate travel time on route y. Let $M_{\delta(y)} = \sum_{n=1}^{N} \mathbf{1}\{y_n \in \delta(y)\}$ be the sample size of traversals within route y's neighborhood, and $|y|$ be the number of segments traversed on route y. Similar to the segment-based estimator, one can consider an estimator $\hat{\Theta}_y$ by solving the ridge regression problem: $\min_{\hat{\Theta}_y} \sum_{n:y_n\in\delta(y)}(\hat{\Theta}_y - \sum_{s\in y_n} T_{n,s}) + \lambda(\hat{\Theta}_y - |y|\theta)^2$ where $\lambda \geq 0$ is a regularization parameter. The estimator will shrink to the prior mean of the travel time on route y, $|y|\theta$, if there is not much traversal data in $\delta(y)$. Again, $\hat{\Theta}_y$ has a closed form: $\hat{\Theta}_y = (\lambda/(\lambda + M_{\delta(y)}))|y|\theta + (M_{\delta(y)}/(\lambda + M_{\delta(y)}))(\sum_{n:y_n\in\delta(y)} \sum_{s\in y_n} T_{n,s}/M_{\delta(y)})$. This motivates the following formal definition of route-based estimators.

Definition 2 (Route-Based Estimator) *A route-based estimator takes the form*

$$\hat{\Theta}_y^{(\text{route})} := (1 - \phi_y(M_{\delta(y)}))|y|\theta + \phi_y(M_{\delta(y)})\frac{\sum_{n:y_n\in\delta(y)} \sum_{s\in y_n} T_{n,s}}{M_{\delta(y)}},$$

for some $\phi_y(M_{\delta(y)}) : \mathbb{Z}_{\geq 0} \mapsto \mathbb{R}$ such that $\phi_y(0) = 0$ and define $\phi_y(0)/0 = 0$.

Our first main result investigates the comparison of the two optimal estimators $\hat{\Theta}_y^{*(\text{seg})}$ and $\hat{\Theta}_y^{*(\text{route})}$ under a special case where $\delta(y)$ is chosen to be the set of historical routes that are exactly the same as y, i.e., $\delta(y) = \{y_n : y_n = y\}$. The optimal estimator minimizes the integrated risk of the estimator Θ_y, defined as $R\left(\hat{\Theta}_y \,\middle|\, y_{[N]}\right) := \mathbb{E}\left[\left(\hat{\Theta}_y - \Theta_y\right)^2 \,\middle|\, y_{[N]}\right]$. Minimizing the integrated risk is equivalent to minimizing the predictive mean squared error, which is often used in evaluating various predictive models in practice.

For any route y, when the covariances of the segment travel times within the route are non-negative, $\sigma_{st} \geq 0, \forall s \neq t \in y$, we prove that the optimal segment-based estimator $\hat{\Theta}_y^{*(\text{seg})}$ always has a weakly lower integrated risk based on squared loss compared to that of the optimal route-based estimator $\hat{\Theta}_y^{*(\text{route})}$.

Theorem 1 *When $\delta(y) = \{y_n : y_n = y\}$, for any route y such that $\sigma_{st} \geq 0$ for all $s \neq t \in y$, and any historical routes $y_{[N]}$, the integrated risk of the optimal segment-based estimator is at least as low as that of the optimal route-based estimator:*

$$R\left(\hat{\Theta}_y^{*(\text{seg})} \,\middle|\, y_{[N]}\right) \leq R\left(\hat{\Theta}_y^{*(\text{route})} \,\middle|\, y_{[N]}\right).$$

We then give our main asymptotic result. We compare a large family of simple segment-based estimator $\hat{\Theta}_{y_p}^{(\text{seg})}$ to the optimal route-based estimator $\hat{\Theta}_{y_p}^{*(\text{route})}$ with a sequence of routes $\{y_p\}_{p\in\mathbb{N}}$ within a growing network of size p and growing sample size N. The family of simple segment-based estimators *only* requires information of the prior means of segment travel times and encompasses popular estimators used in practice. We give conditions under which the segment-based estimator is more accurate when the size of the road network gets larger. The key quantities determining the efficiency of estimators are the rates at which training data accumulates on particular road segments and on particular routes. We denote the probability that a specific road segment s is traversed by a randomly generated route $Y \sim \mu$ as $q_s := \mathbb{P}[s \in Y]$. Similarly, we define $q_{\delta(y)} := \mathbb{P}[Y \in \delta(y)]$ as the probability that a particular route within a neighborhood $\delta(y)$ is sampled.

Theorem 2 *Under the data-rich regime with $Nq_{\delta(y_p)} = \omega(1)$, if $\lim_{p\to\infty} \max_{s\in y_p} |y_p| q_{\delta(y_p)}/q_s = 0$ then any segment-based estimator with $\phi_s(N_s) = 1 - \mathcal{O}(1/\sqrt{N_s})$ for all road segments in y_p, dominates the optimal route-based estimator:*

$$\lim_{p\to\infty} \frac{R\left(\hat{\Theta}_{y_p}^{(\text{seg})}\right)}{R\left(\hat{\Theta}_{y_p}^{*(\text{route})}\right)} = 0.$$

An alternative sufficient condition is that both $\lim_{p\to\infty} \max_{s\in y_p} q_{\delta(y_p)}/q_s = 0$ and $\sigma_{st} \geq 0, \forall s, t \in y_p, \forall p \in \mathbb{N}$.

Under the data-scarce regime $Nq_{\delta(y_p)} = \mathcal{O}(1)$, the route-based estimator is always inconsistent. However, it may hold that $\min_{s\in y_p} Nq_s = \omega(|y_p|)$, in which case any segment-based estimator with $\phi_s(N_s) = 1 - \mathcal{O}(1/\sqrt{N_s})$ for all road segments in y_p is consistent.

We show that the conditions in Theorem 2 hold naturally in a grid network. In summary, we recommend the use of segment-based approaches if one has to make a choice between the two methods. Our work highlights that the accuracy of travel time prediction is driven not just by the sophistication of the model, but also the spatial granularity at which those methods are applied.

References

Derrow-Pinion, Austin, She, Jennifer, Wong, David, Lange, Oliver, Hester, Todd, Perez, Luis, Nunkesser, Marc, Lee, Seongjae, Guo, Xueying, Wiltshire, Brett, *et al.* 2021. ETA Prediction with Graph Neural Networks in Google Maps. *arXiv preprint arXiv:2108.11482*.

Gelman, Andrew, Carlin, John B, Stern, Hal S, Dunson, David B, Vehtari, Aki, & Rubin, Donald B. 2013. *Bayesian data analysis*. CRC press.

Yan, Chiwei, Zhu, Helin, Korolko, Nikita, & Woodard, Dawn. 2020. Dynamic pricing and matching in ride-hailing platforms. *Naval Research Logistics (NRL)*, **67**(8), 705–724.