Mobility21

A USDOT NATIONAL UNIVERSITY TRANSPORTATION CENTER

Carnegie Mellon University

Penn UNIVERSITY of PENNSYLVANIA

THE OHIO STATE UNIVERSITY

CCAC

Dynamic Origin-Destination estimation (DODE) under incidents using individual trajectories data

PI: Sean Qian (ORCID: 0000-0001-8716-8989)

Lead Research Assistant: Haocheng Duan (ORCID: 0009-0001-5456-4659)

FINAL RESEARCH REPORT - September 1, 2023

Contract # 69A3551747111

All data generated from this project can be accessed from https://github.com/HaochengDuan/Incident_Route_Choice.git

1 Introduction

Non-recurrent traffic congestion caused by roadway construction work, planned events, and unplanned traffic incidents can create massive traffic tie-ups and can have equally large economic and environmental regional impacts. With the availability of various traffic data (real-time and historically archived), how to minimize incident-induced disruption to commuting traffic and its impact to the environment presents a big challenge to global cities and communities. While unplanned incidents require careful evaluation of traffic management response plans, guidelines to develop efficient response plans for real-time operations are often lacking. Consequently, there is a real need to study unplanned traffic incidents to understand human behaviors under those incidents, and learn valuable lessons to prepare public agencies to deal more effectively with large routine highway maintenance, reconstruction, big sports events, catastrophic vehicle crash and emergency situations.

In particular, this project develops a model that encapsulates travelers' route choice behavior under unplanned incidents. Travel behavior in terms of route choices under incidents is modeled based on a disutility function for individuals and the calibrated regional network model. This novel behavioral model can be further integrated into a process of Dynamic Origin-Destination Estimation (DODE) that calibrates dynamic network simulation under incidents.

In an off-line manner, we intend to GPS-based traces data to learn the disutility function of individuals' travel choice under incidents. When deployed in real-time, those initial disutility functions can be seen as the expected route choices of individual travelers under incidents, which can be further tuned and refined in real time provided with real-time data. Furthermore, the real-time simulation and DODE receive real-time traffic data feeds (INRIX or GPS traces) and calibrate the en-route route choices in the real time, corrects the forecast of incident-induced traffic congestion in the next hour, and computes the optimal traffic diversion ratios for pre-determined detour routes. Those research steps are left for future work. However, the first step is develop models for route choices under unplanned incidents and algorithms to learn those choices from data.

2 Problem Formulation

This study targets to utilize GPS trace data to deduce drivers' route choice behaviors under incidents. This section initially defines the parametric route choice behavior model. Subsequently, it presents an algorithm framework to tune the parameters in the model from raw GPS traces.

2.1 Route Choice Model

The route-choice model in this study is built based on the hybrid route-choice model (Qian, 2012) and the logit probability model. Specifically, the following assumptions are adopted:

Assumption 1: Drivers can be categorized into two groups: habitual drivers, who strictly follow their predetermined route choices based on previous knowledge and experience, and adaptive drivers, who may adjust their en-route route choices depending on real-time traffic information (including incidents).

Assumption 2: Both habitual and adaptive drivers follow the logit probability model. For habitual drivers, given the utility function of different routes, the probability of choosing the route p between the OD pair r-s is:

$$\mathrm{P}_{rs}^{H,p}(t) = \frac{\exp(-C_{rs}^{H,p}(t))}{\sum_{\pi' \in \Pi_{rs}} \exp(-C_{rs}^{H,\pi'}(t))} \tag{1}$$

where $C_{rs}^{H,p}$ denotes to the habitual drivers utility function of the route p between the OD pair r-s, and Π_{rs} denotes to possible route set between r-s. As for adaptive drivers, the probability of choosing the route p between a node pair r-s at time t is:

$$\mathrm{P}_{rs}^{A,p}(t, \mathbf{I}(t)) = \frac{\exp(-C_{rs}^{A,p}(t, \mathbf{I}(t)))}{\sum_{\pi' \in \Pi_{rs}} \exp(-C_{rs}^{A,\pi'}(t, \mathbf{I}(t)))} \tag{2}$$

where $C_{rs}^{A,p}$ is the utility function estimated by adaptive drivers, and $\mathbf{I}(t)$ is the incident information. Including $\mathbf{I}(t)$ term in the adaptive driver's utility function is because they can access to real-time traffic information.

Besides the incident impact, the formulation of the utility function, $C_{rs}^{H,p}$ and $C_{rs}^{A,p}$, in equations (1) and (2) also considers time, route-overlapping, and location effects. Specifically, $C_{rs}^{H,p}$ is formulated as:

$$C_{rs}^{H,p}(t) = \widetilde{\alpha}\widetilde{\tau}_{rs}^{p}(t) + \widetilde{\beta}\eta_{rs}^{p} + \widetilde{\theta}\rho_{rs}^{p} \tag{3}$$

The first term $\widetilde{\tau}_{rs}^{p}(t)$ refers to the average historical travel time experienced by travellers. The second term refers to the C-logit factor, which reduces the overlapping path effects in the fundamental logit model (Cascetta et al., 1996). ρ_{rs}^{p} refers to the fixed location effect simulating that some drivers may prefer highways instead of minor roads, and $\widetilde{\alpha}$, $\widetilde{\beta}$, $\widetilde{\theta}$ are tun-able parameters. $C_{rs}^{A,p}$ is similar to $C_{rs}^{H,p}$, which is formulated as:

$$C_{rs}^{A,p}(t, \mathbf{I}(t)) = \alpha\tau_{rs}^{p}(t) + \beta\eta_{rs}^{p} + \theta\rho_{rs}^{p} + \sum_{a \in A} \delta_{rs}^{pa}\gamma(a, \mathbf{I}(t)) \tag{4}$$

where $\tau_{rs}^{p}(t)$ refers to the instantaneous travel time, and δ_{rs}^{pa} is a indicator function, equals to 1 if link a is on the route p of r-s and 0 otherwise. The last term, $\gamma(a, \mathbf{I}(t))$, is a learnable function that evaluates the driver's intuition of how the link a is affected by the incident at time t, and α, β, θ are tun-able parameters.

2.2 Optimization Object

Given habitual driver traces and adaptive driver traces, this study applies maximum likelihood estimation (MLE) to tune the parameters in (3) and (4) separately.

It is straightforward to write down the likelihood function of habitual drivers:

$$\underset{\widetilde{\alpha},\widetilde{\beta},\widetilde{\theta}}{maximize} \prod_{t \in T} \prod_{s \in N_S} \prod_{r \in N_R} \prod_{k \in V_{rs}^{H}(t)} \prod_{p \in \Pi_{rs}} \mathrm{P}_{rs}^{H,p}(t)^{y_{rs}^{H,p}(k)} \tag{5}$$

$$\text{S.T.} \quad (1) \quad \text{and} \quad (3)$$

where T denotes the analysis period, N_R refers to the origin node set, N_S refers to the destination node set, $V_{rs}^{H}(t)$ denotes habitual driver traces set (from r to s departing within the time period t), and $y_{rs}^{H,p}(k)$ is an indicator function, equals to 1 if traces k is on route p and 0 otherwise.

To maximize the (5) is to minimize its negative logarithm, thus the optimization objective can be further written as:

$$\underset{\widetilde{\alpha},\widetilde{\beta},\widetilde{\theta}}{minimize} - \sum_{t\in T} \sum_{s\in N_S} \sum_{r\in N_R} \sum_{k\in V_{rs}^{H}(t)} \sum_{p\in\Pi_{rs}} y_{rs}^{H,p}(k)\log(\mathrm{P}_{rs}^{H,p}(t)) \tag{6}$$

$$\text{S.T.} \quad (1) \quad \text{and} \quad (3)$$

The likelihood function employed to optimize parameters in the adaptive driver's utility function differs. This distinction arises because adaptive drivers often make en-route route choices, implying that their traces are not directly related to their initial route choice. Therefore, the likelihood function uses link-choice probability instead, as adaptive drives' traces still reflect their link choice at the intersection point. The link-choice probability can be easily derived from the route choice probability: for a driver with a destination s on the intersection point n_0, the probability of choosing the link $\overrightarrow{n_0n_1}$ is:

$$\mathrm{P}_{n_0s}^{\overrightarrow{n_0n_1}}(t,\mathbf{I}(t)) = \sum_{p\in\Pi_{n_0s}} \delta_{n_0s}^{p\overrightarrow{n_0n_1}} \mathrm{P}_{n_0s}^{A,p}(t,\mathbf{I}(t)) \tag{7}$$

Figure 1 gives several examples illustrate (7): $\mathrm{P}_{AD}^{\overrightarrow{AB}}(t,\mathbf{I}(t)) = \mathrm{P}_{AD}^{A,1}(t,\mathbf{I}(t))+\mathrm{P}_{AD}^{A,3}(t,\mathbf{I}(t))$; $\mathrm{P}_{AD}^{\overrightarrow{AC}}(t,\mathbf{I}(t)) =$

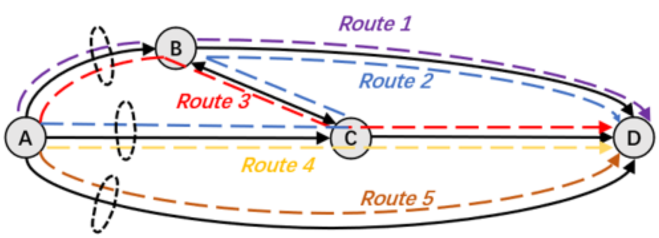


Figure 1: Link Choice Probability

$\mathrm{P}_{AD}^{A,3}(t,\mathbf{I}(t)) + \mathrm{P}_{AD}^{A,4}(t,\mathbf{I}(t))$; $\mathrm{P}_{AD}^{\overrightarrow{AB}}(t,\mathbf{I}(t)) = \mathrm{P}_{AD}^{A,5}(t,\mathbf{I}(t))$.

Thus, the likelihood function of adaptive drivers can be written as:

$$\underset{\alpha,\beta,\theta,\vec{\gamma}}{maximize} \prod_{t\in T} \prod_{s\in N_S} \prod_{r\in N_R} \prod_{k\in V_{rs}^{A}(t)} \prod_{n_a\in v_{rs}^{k}(t)} \prod_{t'\in T} \prod_{n_b\in N_E(n_a,s)} \mathrm{P}_{n_as}^{\overrightarrow{n_an_b}}(t',\mathbf{I}(t'))^{y_{n_as}^{\overrightarrow{n_an_b}}(k)} \tag{8}$$

$$\text{S.T.} \quad (2), \quad (4) \quad \text{and} \quad (7)$$

where $v_{rs}^{k}(t)$ denotes the intersection point the trace k passed, t' is the time the trace k passed n_a, and $N_E(n_a,s)$ refers to the node set that are the downstream node of the first edge on the route from n_a to s. Take the negative logarithm of (8), the objective function can be written as:

$$\underset{\alpha,\beta,\theta,\vec{\gamma}}{minimize} - \sum_{t\in T} \sum_{s\in N_S} \sum_{r\in N_R} \sum_{k\in V_{rs}^{A}(t)} \sum_{n_a\in v_{rs}^{k}(t)} \sum_{t'\in T} \sum_{n_b\in N_E(n_a,s)} y_{n_as}^{\overrightarrow{n_an_b}}(k)\log(\mathrm{P}_{n_as}^{\overrightarrow{n_an_b}}(t',\mathbf{I}(t'))) \tag{9}$$

$$\text{S.T.} \quad (2), \quad (4) \quad \text{and} \quad (7)$$

2.3 Real-world Data Adaptation

The above optimization problem would be easy to solve given perfect GPS traces (meaning the origin, destination, and entire traces are given). However, most real-world GPS traces are sparse without OD information. Thus, inferring origin, destination, and which route the driver actually selected would be difficult. Another potential problem is that due to the complexity of the large-scale network, the number of routes generated by permutations for each OD pair is excessive, making (6) and (9) challenging to optimize. Therefore, this study proposes a novel algorithm designed not only to reconstruct the traces but also to infer the route set for each origin-destination (OD) pair, ensuring a calculable number of routes.

Algorithm 1 Trace-route Reconstruction by GPS Data

```
group GPS points by ANONID
spatial-joint GPS points with the consolidated network
for each trace do
    match to nearest origin and destination
    connect unconnected time-sequential matched links by the shortest distance
end for
for each OD pair do
    maximum cluster number = number of different traces
    put the same traces into one cluster
    for i from maximum cluster number down to 1 do
        compute the average Jacobian similarity between each cluster
        combine the clusters with the largest average Jacobian similarity
        for traces in the OD pair do
            compute the average trace-cluster Jacobian similarity
        end for
        compute overall trace-cluster Jacobian similarity, plot cluster number-similarity graph
    end for
    determine the cluster number by finding the 'knee' point on the cluster number-similarity graph
    with a reasonable cluster number, compute the 'mean route' within each cluster
end for
return the reasonable route set between each OD-pair
```

3 Preliminary Results

This section initially presents the outcomes of numerical experiments to validate the efficacy of the proposed optimization algorithm. Additionally, a preliminary real-world GPS matching result is showcased to demonstrate the capability of our solution framework in processing real-world GPS trace data.

3.1 Numerical Experiments

The numerical experiments were conducted in SUMO on a 2×2 network with intersections controlled by traffic lights (see Figure 2). Various sets of coefficients were pre-defined, from which corresponding GPS traces were generated. The data extracted from generated GPS trace data was subsequently fed into (6) and (9), enabling us to tune coefficients for comparison with the pre-defined ones. The optimization results are tabulated in Table 1, from which we can observe that the MLE is efficient to optimize the coefficients. This means that given reasonable route choice model assumption and good enough data, MLE is supposed to be able to solve both (6) and (9). Figure 3 further demonstrates the convexity of the optimization objective using two 2-parametric examples.

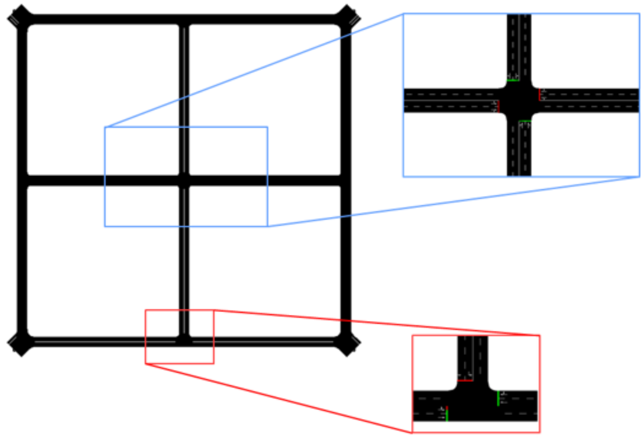


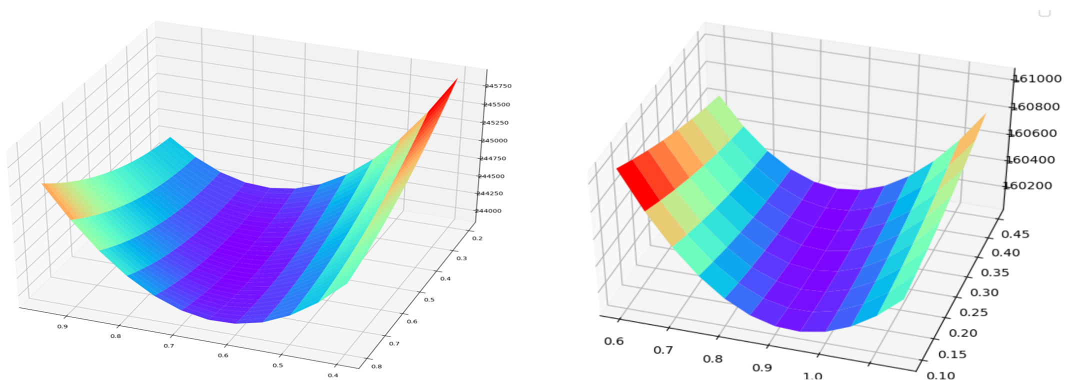
Figure 2: Mini-Network

Table 1: Ground-Truth Optimized Results Comparison

Coefficients	α	β	θ	γ	$\widetilde{\alpha}$	$\widetilde{\beta}$	$\widetilde{\theta}$
ground-truth	0.7	0.5	-	-	0.7	0.3	-
optimized results	0.684	0.523	-	-	0.701	0.291	-
ground-truth	0.9	-	0.5	-	0.9	-	0.3
optimized results	0.890	-	0.513	-	0.907	-	0.294
ground-truth	0.9	-	-	0.3	0.7	-	-
optimized results	0.882	-	-	0.310	0.693	-	-
ground-truth	0.7	0.3	-	0.5	0.7	0.3	-
optimized results	0.688	0.289	-	0.517	0.705	0.303	-

3.2 Real-World Traces Matching

The real-world GPS traces were collected in the DMV (DC-Maryland-Virginia) area from Jan 1st, 2020-Feb 28th, 2020. After cleaning, around 130,000 traces are kept for analysis. Figure 4 presents the results of the trace-route reconstruction algorithm for two OD pairs, where black dots are raw GPS points, red points are the origin and destination, and blue lines are reconstructed routes. It can be found that most GPS points matched with the reconstructed routes, while some distinct noises are also ignored. The figure 5 illustrates adjusting the incident cost to 5 minutes on the



(α, β) ground truth: (0.7, 0.5)

(α, γ) ground truth: (0.9, 0.3)

Figure 3: Convexity of the Objective Function

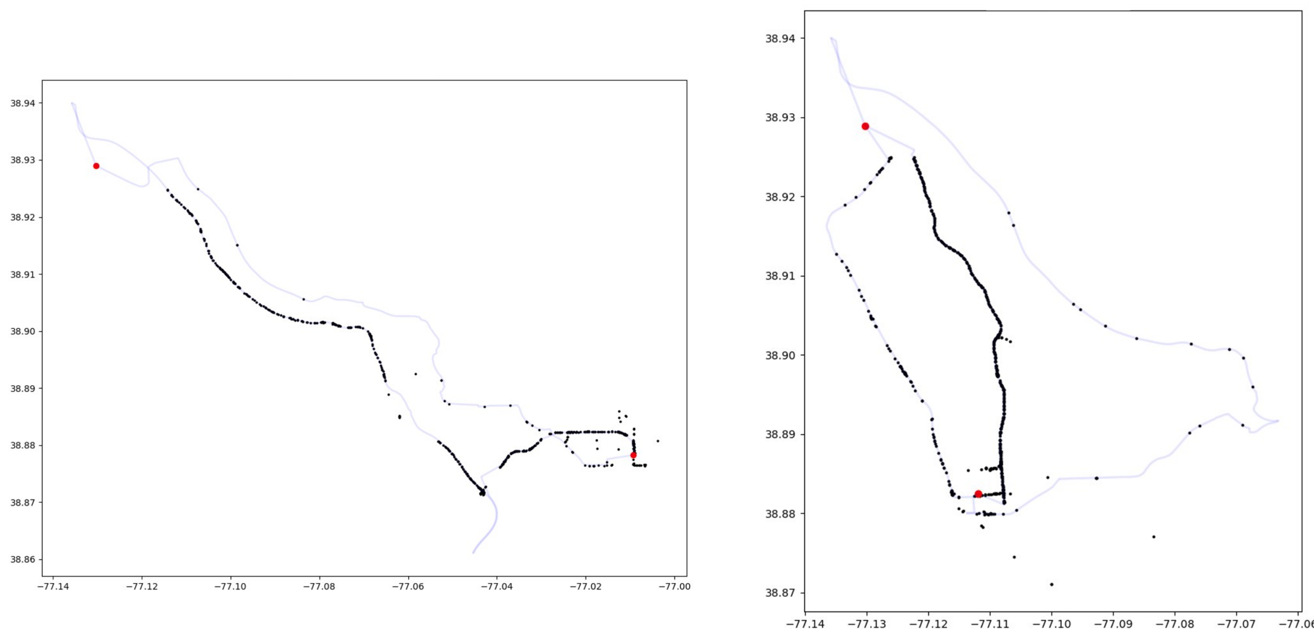


Figure 4: Demonstration of Route Clustering Results

link where the incident happened and keeping it at 0 on the other links. Though these results are initial findings from the grid-search method, the route with the highest probability based on these preliminary outcomes aligns with the real traces, which do not align with the time-dependent shortest path based on recurrent travel data (yellow line).

4 Conclusion and Future Works

This study introduces a hybrid model designed to capture a driver's route choice behavior during incidents. We've developed two distinct solution algorithms, both grounded in MLE, to fine-tune the coefficients for adaptive drivers and habitual drivers. Additionally, the study details how

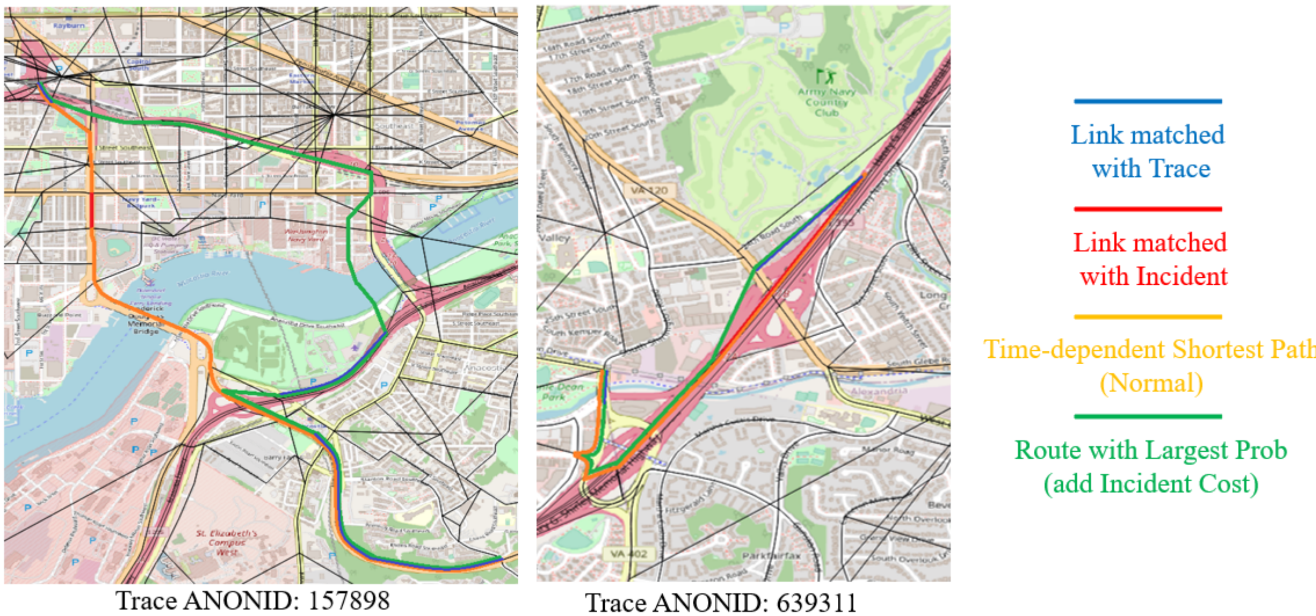


Figure 5: Trace Matching Results under Incidents

to incorporate real-world sparse GPS trace data into this algorithm. The numerical tests and initial findings underscore the potential of our solution framework. However, several areas require further exploration: (1) Currently, for real-world GPS trace data, this study assumed that the habitual drivers are those who always stick to the recurrent time-dependent shortest path. We need to investigate if there is a more reasonable way to distinguish real-world habitual driver traces and adaptive driver traces. (2) Our current model doesn't account for the topological impact on incident costs. For instance, incidents on neighboring links might deter adaptive drivers from using a particular route. We aim to design a more intricate incident cost parameter to reflect this. (3) To solve the MLE optimization problem, we currently apply the grid-search method, as it is difficult for gradient descent to converge. It's essential to determine if the log-likelihood function for real-world GPS traces remains convex. (4) We intend to incorporate the refined coefficients into a dynamic traffic assignment. By using system-level speed, we hope to further validate or re-calibrate the coefficients.

References

Qian, S. A Hybrid Route Choice Model for Dynamic Traffic Assignment. *Networks and Spatial Economics* **2012**, *13*, 183–203.

Cascetta, E.; Nuzzolo, A.; Russo, F.; Vitetta, A. A modified logit route choice model overcoming path overlapping problems. Specification and some calibration results for interurban networks. Transportation and Traffic Theory. Proceedings of The 13th International Symposium On Transportation And Traffic Theory, Lyon, France, 24-26 July 1996. 1996.