Online Empirical validation of network learning with taxi GPS data from Wuhan, China

Susan Jia Xu[1], Qian Xie[1], Joseph Y.J. Chow[1*], Xintao Liu[2]

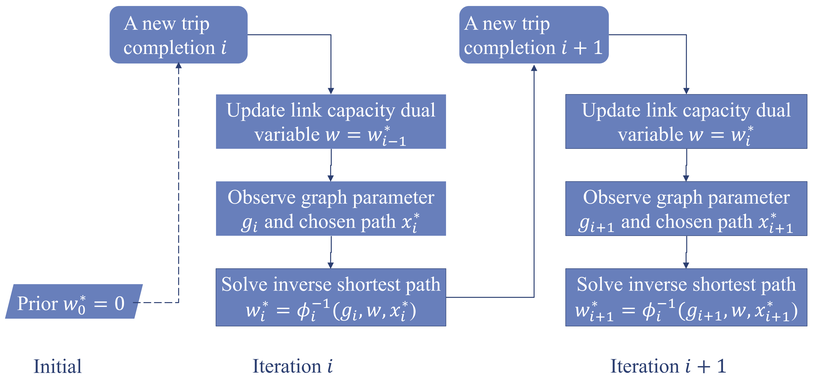
[1]Depatment of Civil and Urban Engineering, Tandon School of Engineering, New York University *Corresponding author: **joseph.chow@nyu.edu**

[2]Depatment of Land Surveying and Geo-informatics, The Hong Kong Polytechnic University

Introduction

- Our prior research developed a statistically cheap method to monitor transportation network performance by using only a few groups of agents without forecasting the population flows.
- The "multi-agent inverse optimization" (MAIO) method infers congested links' capacity dual variables using samples of agents' inverse shortest path problems.
- This empirical study validates the MAIO method by using real taxi trajectory data to simulate online monitoring environment and learn network parameters.
- This work demonstrates how taxi trajectory data quantifies and explains the change of network states (e.g. congestion effects).

Multi-agent Inverse Optimization Method



Proposed Validation Experiment Design

1. Initiate with values of link capacity dual variables equal to zero for all links in the study network.
2. Starting at 5:00AM, and every 5 minutes thereafter until 9:00AM,
   a. For all the trajectories that arrived in that period, identify OD pairs.
   b. Run the path reconstruction algorithms to get real-time travelers' choices for each of the OD pairs (in this step, the traveler's choice is assumed as the shortest path).
   c. Compare the predicted route and the actual route chosen. (Criteria 1)
   d. Run MAIO to update the dual variables based on the reconstructed path.
   e. Compute the correlation between real travel times and estimated travel time. (Criteria 2)

Data Processing Framework

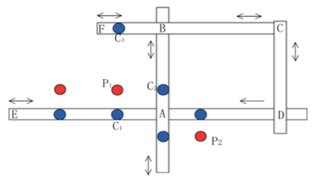
Hotspot Identification → Trip Extraction → Road Network Abstraction → Path Reconstruction

Hotspot Identification

Road Network Abstraction

Taxi ID	Timestamp	Longitude	Latitude	Status	dist(P, O)	dist(P, D)	inBetween	Start/End
9018	5/6/2014,7:00:01	114.26105	30.543255	1	805	4973	0	0
9018	5/6/2014,7:01:01	114.26089	30.546388	1	457	5077	1	1
9018	5/6/2014,7:03:01	114.257895	30.548533	1	350	5423	1	0
9018	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
9018	5/6/2014,7:16:01	114.312165	30.538913	1	5100	647	1	0
9018	5/6/2014,7:17:01	114.311593	30.53681	1	5111	411	1	-1
9018	5/6/2014,7:19:01	114.311227	30.535638	0	5118	282	0	0

Trip Extraction



Matching-link Candidates[1]

Best-fit Path

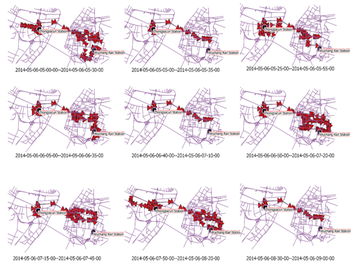
Case Study – Wuhan Downtown

Wuhan, China

855 nodes
2833 links

132 samples observed for OD 1
48 samples observed for OD 2

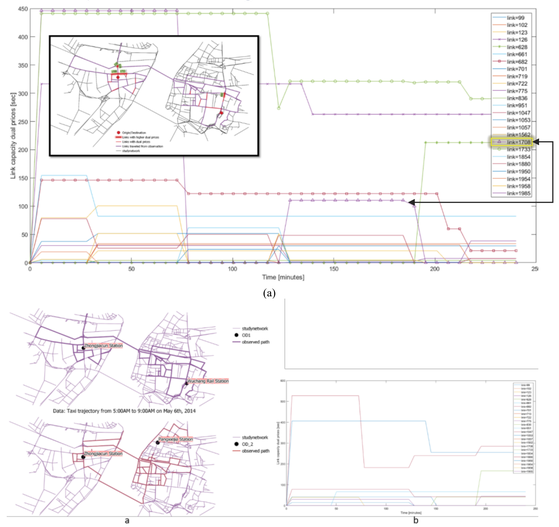
53 different path taken for OD 1
29 different path taken for OD 2



Samples of route diversions for one OD

Case Study Results

Sensitivity to network changes



(a) Trajectories of link dual variables with one OD
(b) Updated dual variables after adding new OD

25 links with positive dual variables

Link 1078: highest, then drop to 0 ⇒ an incident in the earlier spike

Link 1733: sustained ⇒ heavy usage under recurrent congestion effect

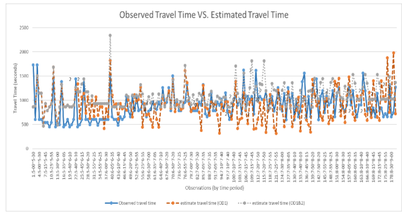
MAIO can estimate dual variables (congestion effects) in real-word urban network

Temporal profile of link dual variables look quite different

Link 1078: more accurate ⇒ effectiveness of MAIO depends on OD sampling

More route observations are considered, more information on dual variables is provided

Accuracy improvement on estimation



Estimated travel time and real travel time of 180 observed routes from single and two OD pairs

Estimated travel time is calculated as free flow travel time plus estimated dual variables on traveled links

Correlations between the observed and estimated travel time are 0.23 and 0.56

MAIO provides a good fit to real observations, even samples from only two ODs

Conclusions and Future Work

- Our earlier study only provided a theoretical argument and numerical illustrations. This work fills in the research gap by conducting an empirical validation experiment with real route observations obtained from taxi GPS trajectory data.
- Network parameters like link capacity dual variables can be updated using only samples of individual route observations, without estimating the total link or path flows.
- Future work should implement the system in a real-world setting using GIS tools and monitor with predefined thresholds to set alerts for dual variables in an online dashboard.
- Monitoring a network before, during, and after a disaster to quantify the impact of dual price increases due to capacity degradation can also be meaningful.
- Due to user privacy issues in GPS data sharing , we can consider using a blockchain or a differential privacy oriented database to anonymize GPS data.

Funding: *NSF* grant *CMMI-1652735*

[1] Zhao, Xiangmo, et al. "Advanced topological map matching algorithm based on D–S theory." *Arabian Journal for Science and Engineering* 43.8 (2018): 3863-3874.