The Regularity, Predictability, and Travel Behavior in Urban Transit Mobility

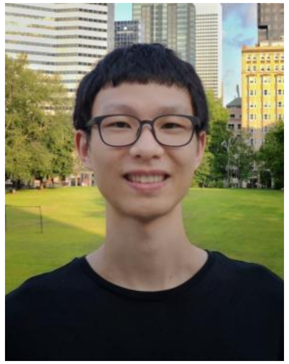
Dr. Zhanhong Cheng

Postdoctoral Researcher

McGill University

zhanhong.cheng@mcgill.ca

About me



Postdoc at McGill University 2022—current

Education

- Ph.D., McGill University, 2018-2022
- B.Eng. and M.S., Harbin Institute of Technology (China), 2012-2018

Research interest

- Spatiotemporal data modeling
- Machine learning in transportation
- Travel behavior and mobility
- Sustainable transportation

Opportunity at McGill University

Lijun Sun, Associate Professor
- 1-2 PhD students for 2024 Fall 2025 Winter!
- Welcome to apply for postdoc.
- See https://lijunsun.github.io/

Introduction

A mobility pattern refers to the **regular and repeated** movements and behaviors of individuals or groups in a given **geographical area over time.**

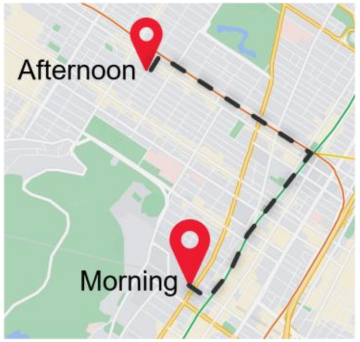
In this talk:

- Mobility patterns in unban transit system
    - Regularity
    - Power law
    - Chained travel
- Applications of mobility patterns
    - Trip destination inference
    - Passenger flow forecasting

Regularity

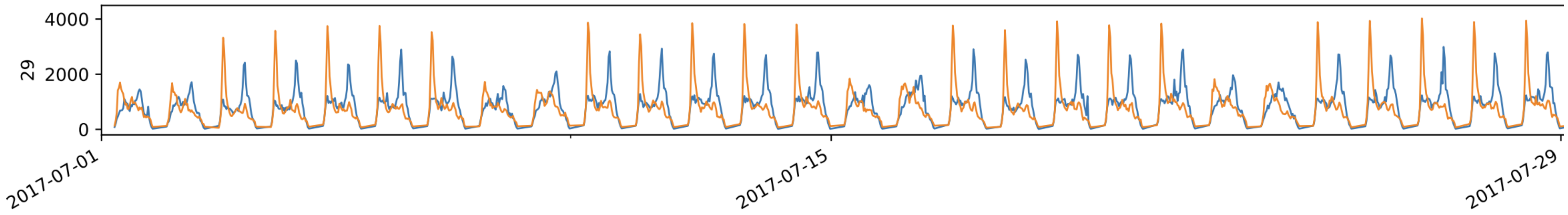
Individual regularity

An individual tends to repetitively visit similar locations at a similar time of the days/weeks.



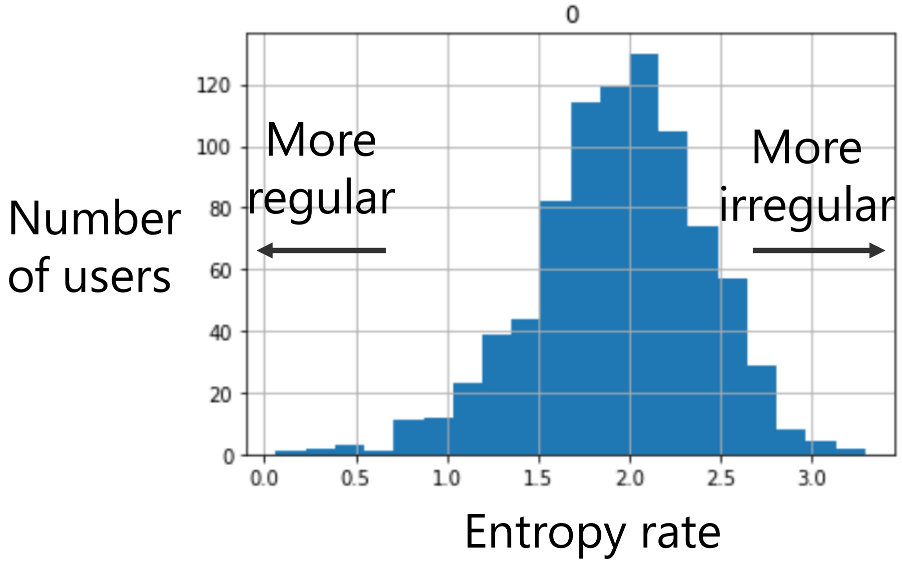
Aggregated regularity

The boarding/alighting flow at a metro station is similar every day and every week.



How to measure the regularity

- Using **entropy rate**[1] to measure the travel regularity of a users.

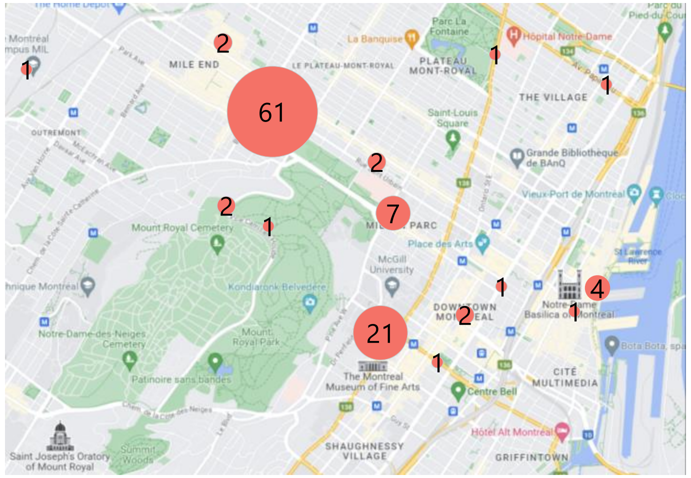


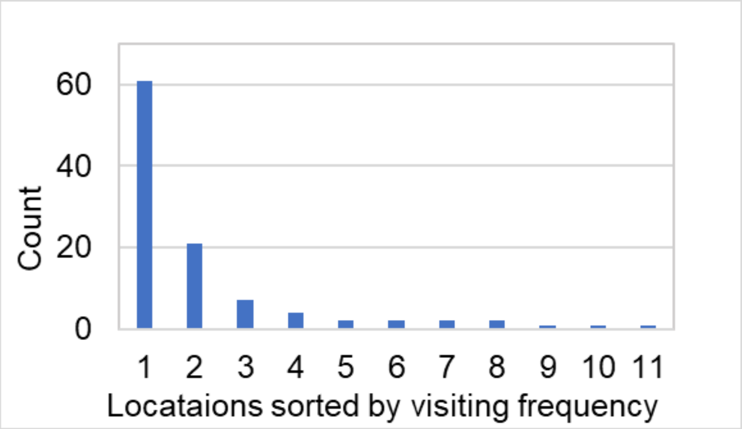
Predicting the next location

	Entropy rate	Prediction accuracy
1%	(0.0, 0.5]	0.992308
2%	(0.5, 1.0]	0.825574
12%	(1.0, 1.5]	0.712156
37%	(1.5, 2.0]	0.564097
37%	(2.0, 2.5]	0.480812
10%	(2.5, 3.0]	0.376798
1%	(3.0, 3.5]	0.278975

[1] Goulet-Langlois, G., Koutsopoulos, H. N., Zhao, Z., & Zhao, J. (2017). Measuring regularity of individual travel patterns. IEEE Transactions on Intelligent Transportation Systems, 19(5), 1583-1592.

Location visiting frequency





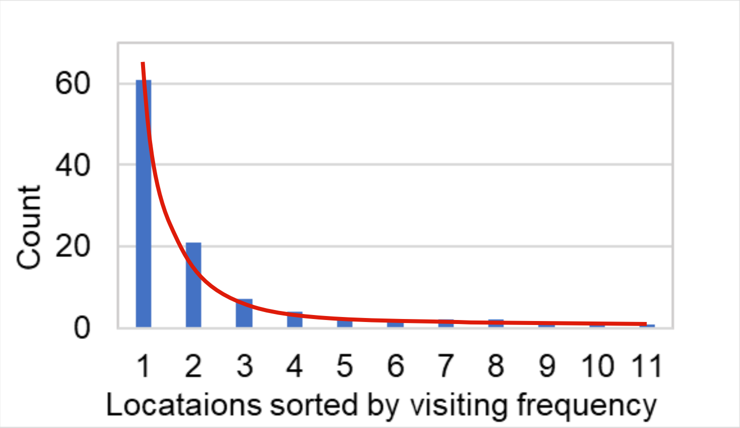
	longitude	latitude	counts
0	-73.779178	45.504268	61.0
1	-73.809165	45.511794	21.0
2	-73.683222	45.514318	7.0
3	-73.622383	45.496095	4.0
4	-73.674878	45.509570	2.0
5	-73.813745	45.510917	2.0
6	-73.706328	45.498340	2.0
7	-73.831310	45.463755	2.0
8	-73.616497	45.503281	2.0
9	-73.671235	45.507773	1.0
10	-73.632257	45.489807	1.0
11	-73.676664	45.440399	1.0
12	-73.589794	45.492502	1.0
13	-73.625907	45.500586	1.0
14	-73.740878	45.438603	1.0
15	-73.619936	45.497891	1.0

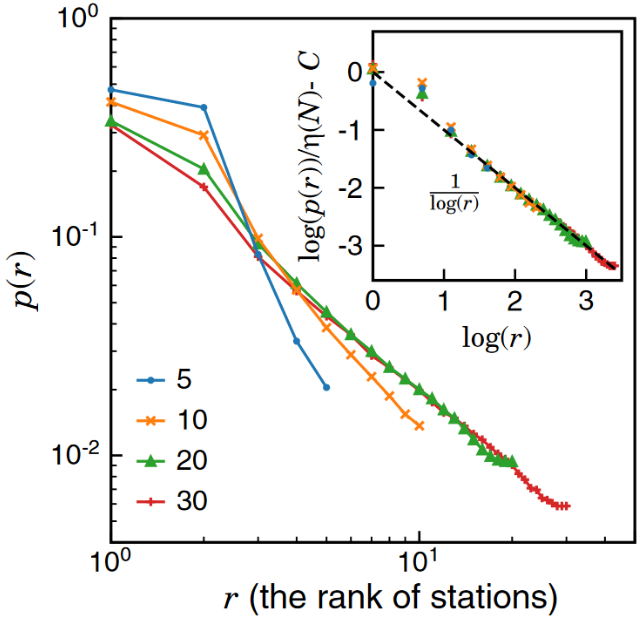
The top few locations dominate.

Power law

The frequency of a passenger visiting different stations follows a power law.

$$p(r) \propto r^{-\eta}$$

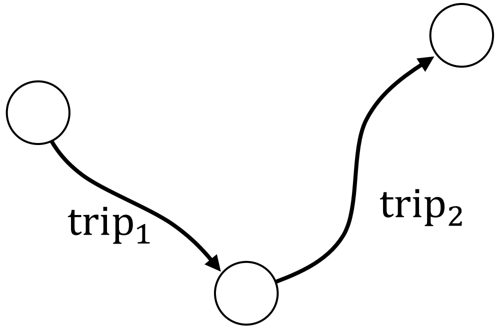




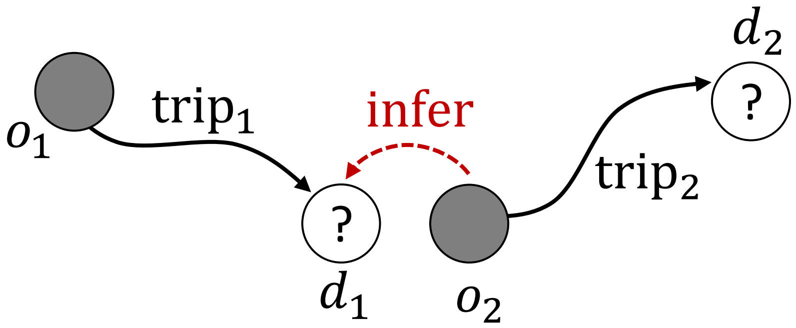
In addition to the power law. There is a **bi-central** mobility pattern in transit systems.

Chained travel

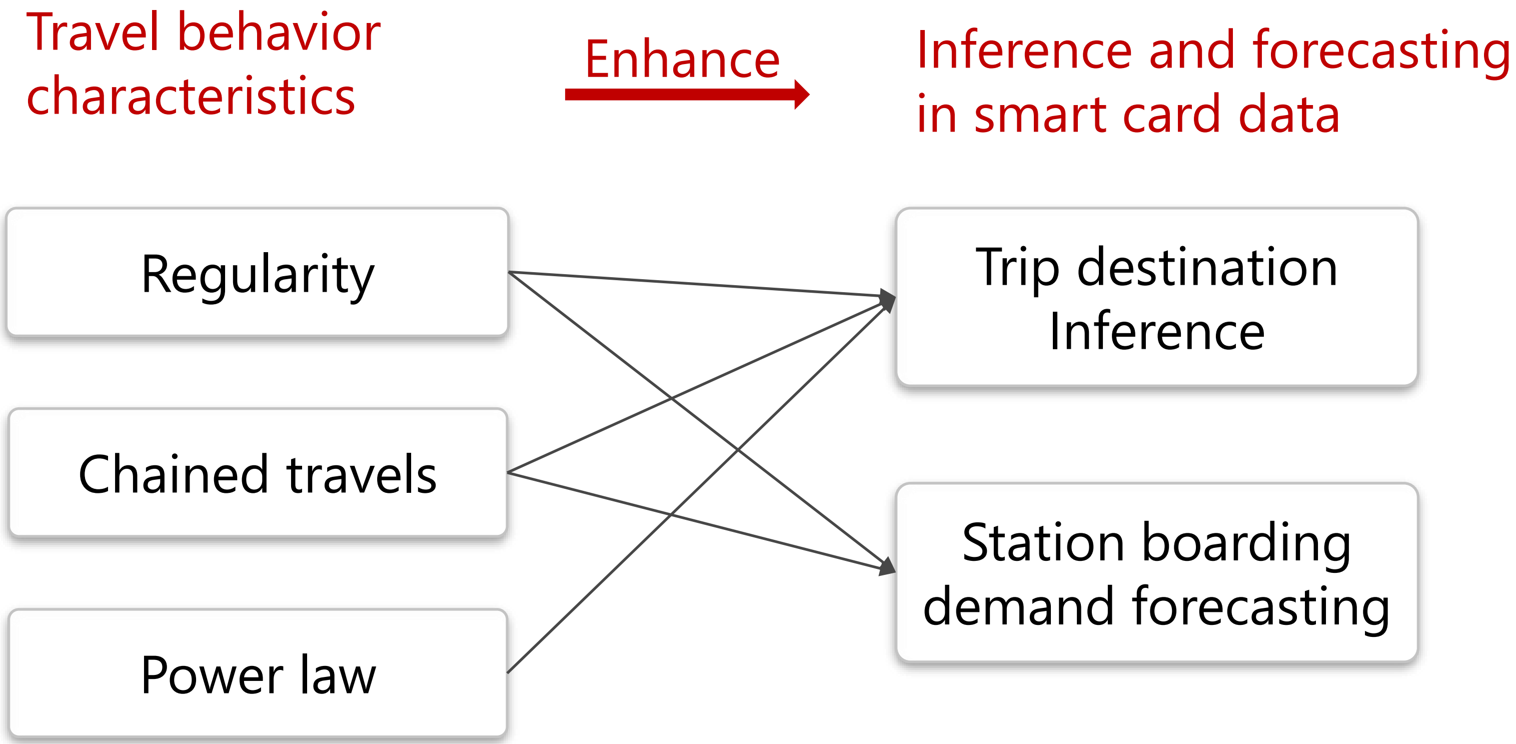
- The next trip starts at the end of the previous trip.



- A typical application of chained travel in inferring trip destinations.



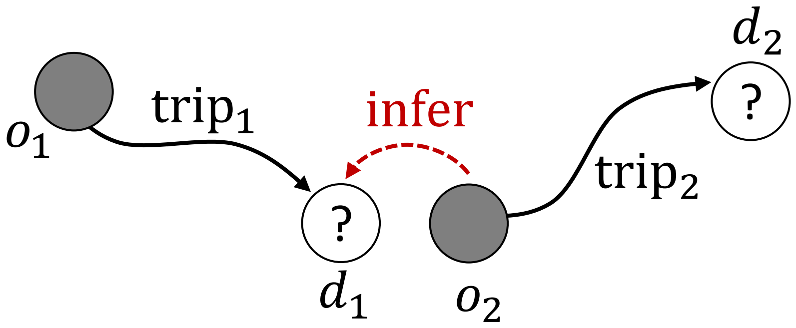
How mobility patterns help



Trip destination inference

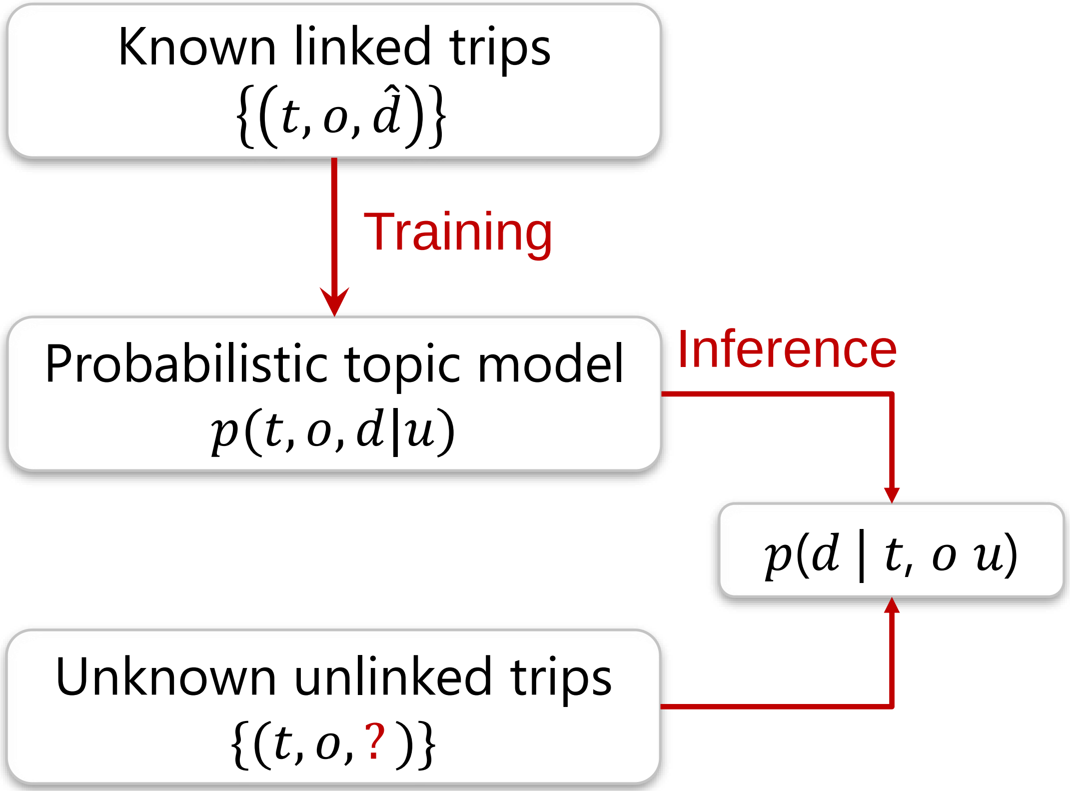
By chained travel:

- For **linked trips:** infer destinations by the origin of the next trip.



- For **unlinked trips**: lack an appropriate destination inference method.

Proposed method:



Cheng, Z., Trépanier, M., Sun, L., 2021. Probabilistic model for destination inference and travel pattern mining from smart card data. *Transportation*.

Probabilistic topic model

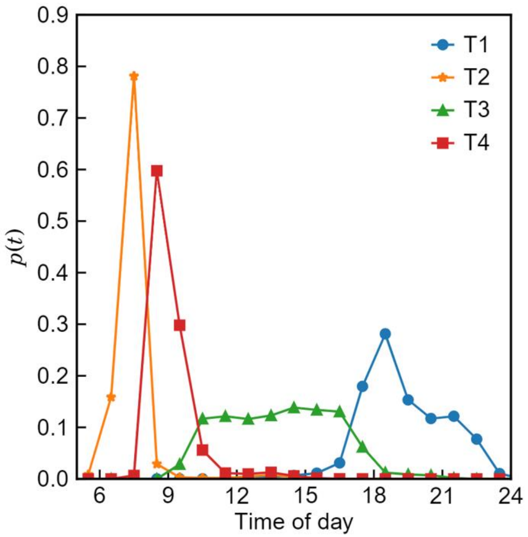
$$P(d \mid t, o; u) \propto P(t, o, d; u)$$

$$= \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{l=1}^{L} P(t|z_j^t) P(o|z_k^o) P(d|z_l^d) P(z_j^t, z_k^o, z_l^d; u)$$

Time topic distributions

$$P(d \mid t,o;u) \propto P(t,o,d;u)$$

$$= \sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{L} P(t|z_j^t)P(o|z_k^o)P(d|z_l^d)P(z_j^t,z_k^o,z_l^d;u)$$

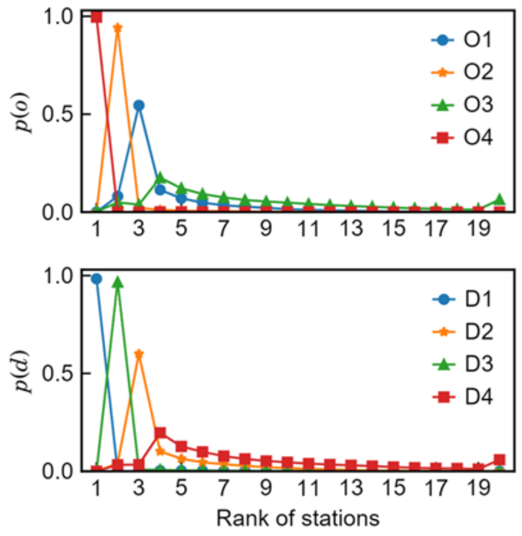


- Four **time topics** have clear semantic meanings.
  - T1: evening trips.
  - T2: early morning trips.
  - T3: afternoon trips.
  - T4: late morning trips.

Origin-destination topic distributions

$$P(d \mid t,o;u) \propto P(t,o,d;u)$$

$$= \sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{L} P(t|z_j^t)P(o|z_k^o)P(d|z_l^d)P(z_j^t,z_k^o,z_l^d;u)$$



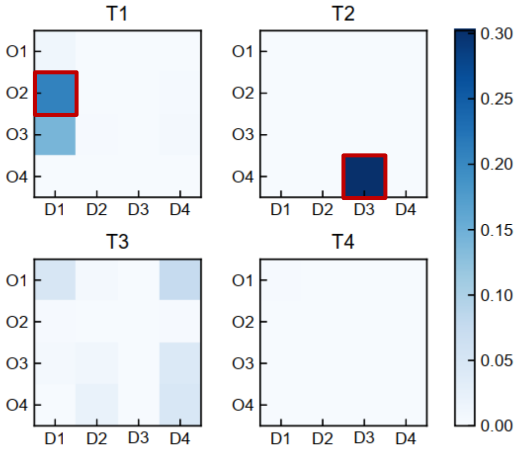
- For each user, we represent stations by their **rank** in **visiting frequency**.
  - Diverse spatial distribution → Similar behavioral regularities.
  - Improve destination inference accuracy.
- We can find the **power-law** property in the distribution of origin/destination topic distributions.

User-topic distributions

$$P(d \mid t,o;u) \propto P(t,o,d;u)$$

$$= \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{l=1}^{L} P(t|z_j^t)P(o|z_k^o)P(d|z_l^d)P(z_j^t,z_k^o,z_l^d;u)$$

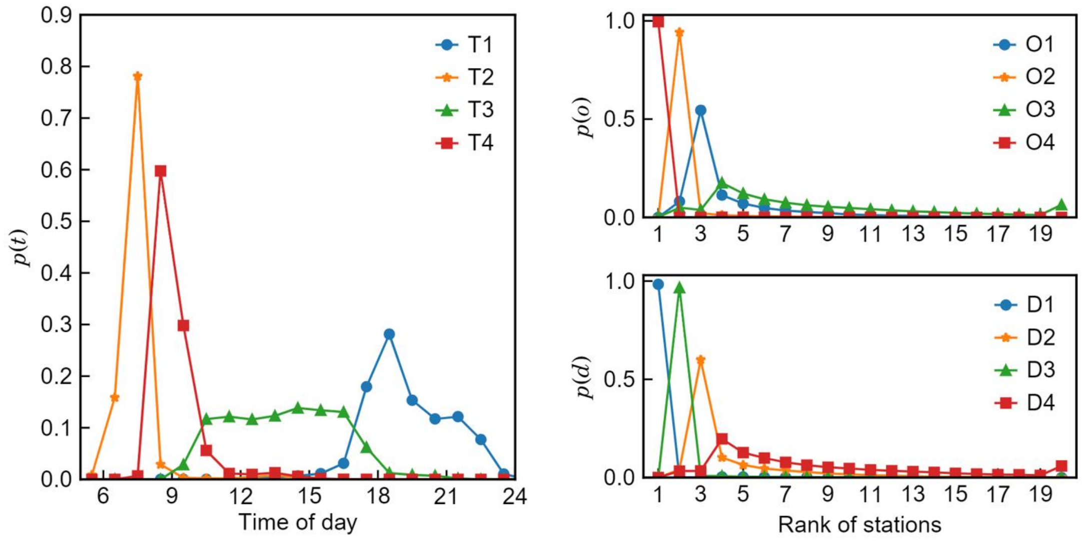
- Each user's travel behavior is characterized by a distribution over topics.
- The user on the left clearly has **two** types of frequent trips (probably a commuter).
- The user-topic distribution is a good feature for passenger clustering.



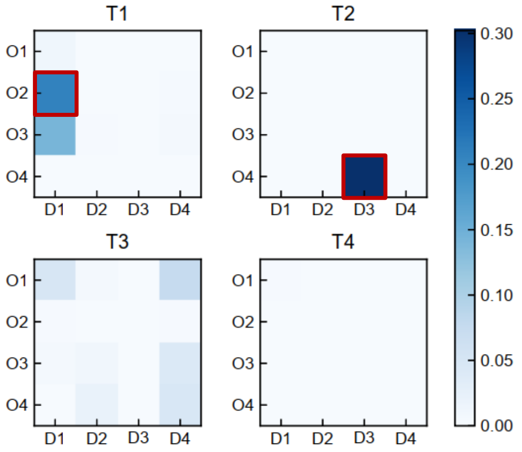
Probabilistic topic model

$$P(d \mid t, o; u) \propto P(t, o, d; u)$$

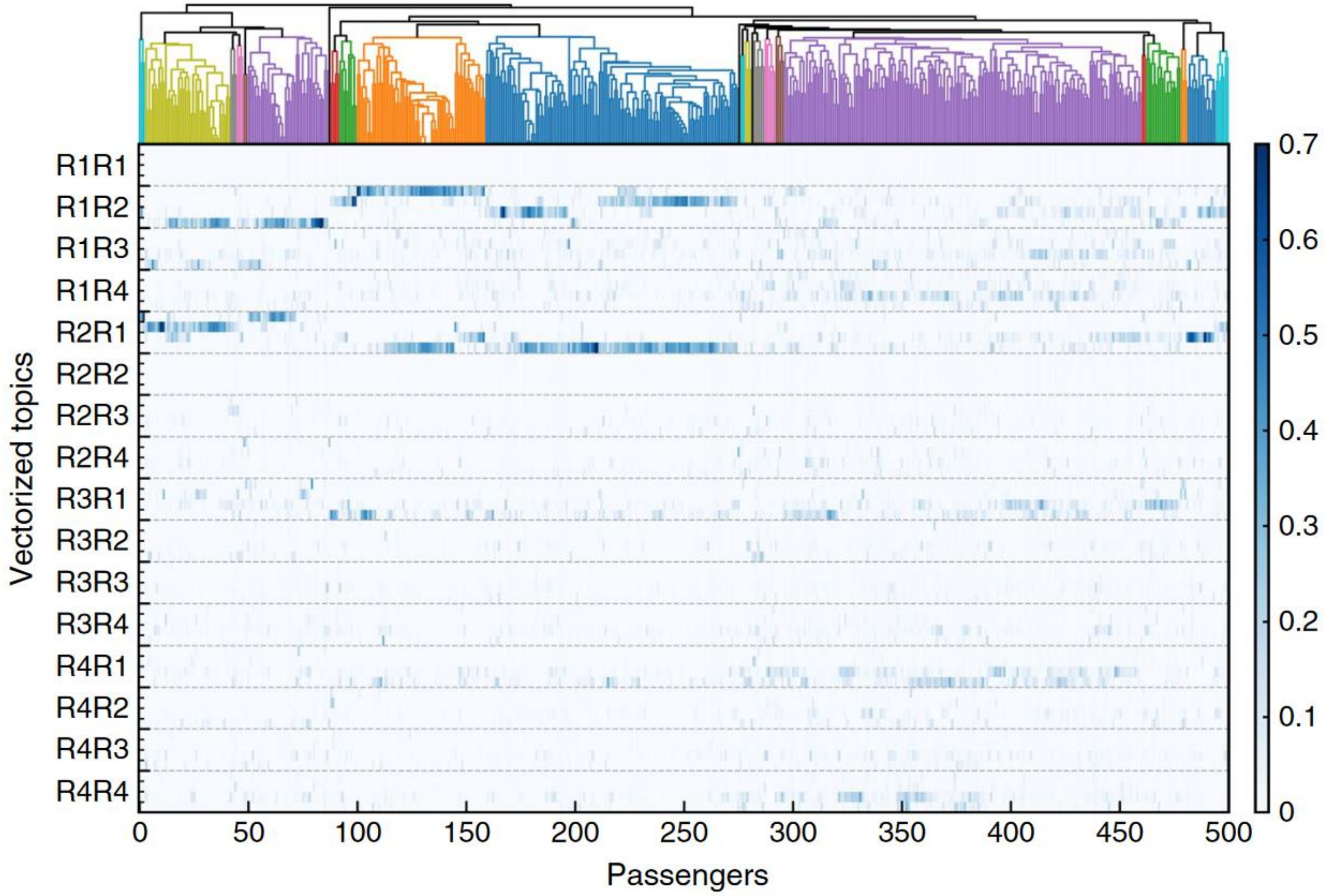
$$= \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{l=1}^{L} P(t|z_j^t) P(o|z_k^o) P(d|z_l^d) P(z_j^t, z_k^o, z_l^d; u)$$



- Estimate the destination with the largest probability given the origin and time.

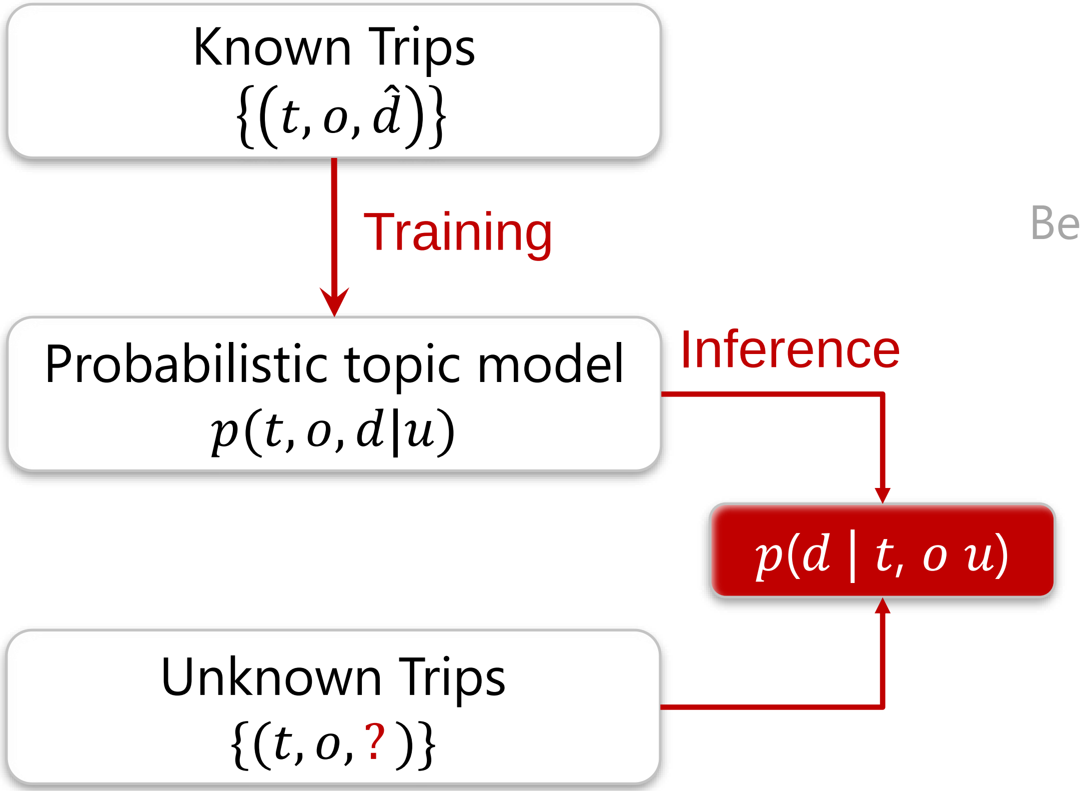


Passenger clustering



Top: the dendrogram of the hierarchical clustering on 500 passengers.
Bottom: the feature matrix for the clustering

Inference accuracy



Destination inference accuracy of unlinked trips in 10,000 passengers

Benchmark models

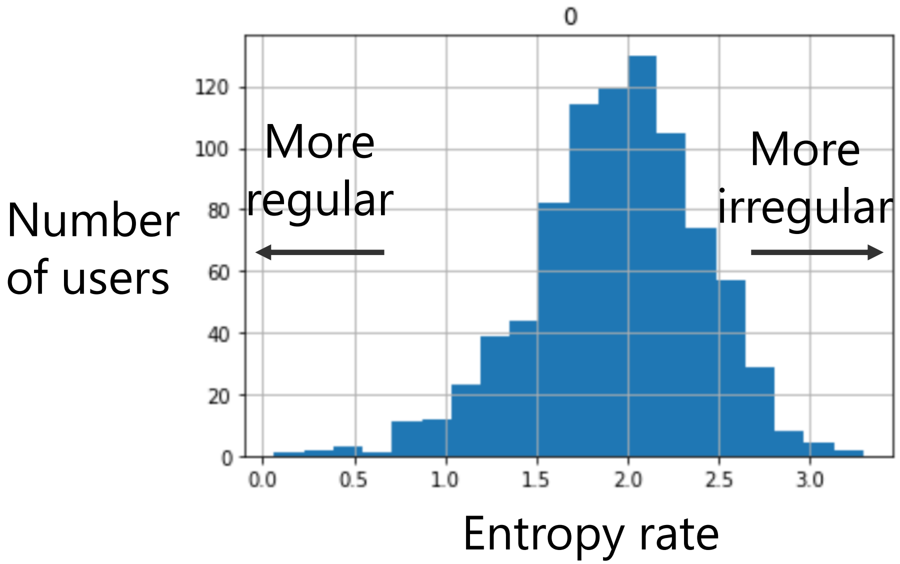
Methods	Accuracy
SO	49.63%
ST	43.02%
SOT_O	48.93%
SOT_T	44.19%
Kernel-based [1]	50.51%
Rank topic	51.43% (±0.14%)
No-rank topic	31.14% (±0.20%)

- The proposed topic model improves the destination inference accuracy.
- Representing stations by their **ranks** improves the inference accuracy.
- The topic model can also be used to analyze passengers' travel behavior patterns.

[1] He and Trépanier (2015)

Inference accuracy

- Using **entropy rate** to measure the travel regularity of a users.

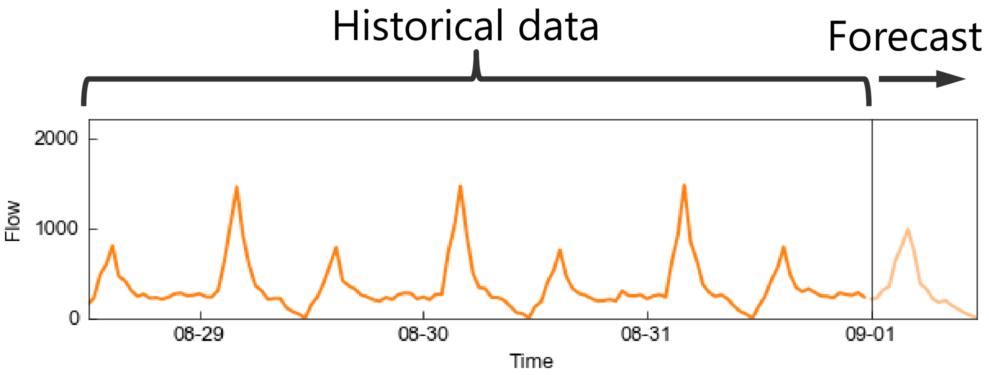


Predicting the next location

	Entropy rate	Prediction accuracy
1%	(0.0, 0.5]	0.992308
2%	(0.5, 1.0]	0.825574
12%	(1.0, 1.5]	0.712156
37%	(1.5, 2.0]	0.564097
37%	(2.0, 2.5]	0.480812
10%	(2.5, 3.0]	0.376798
1%	(3.0, 3.5]	0.278975

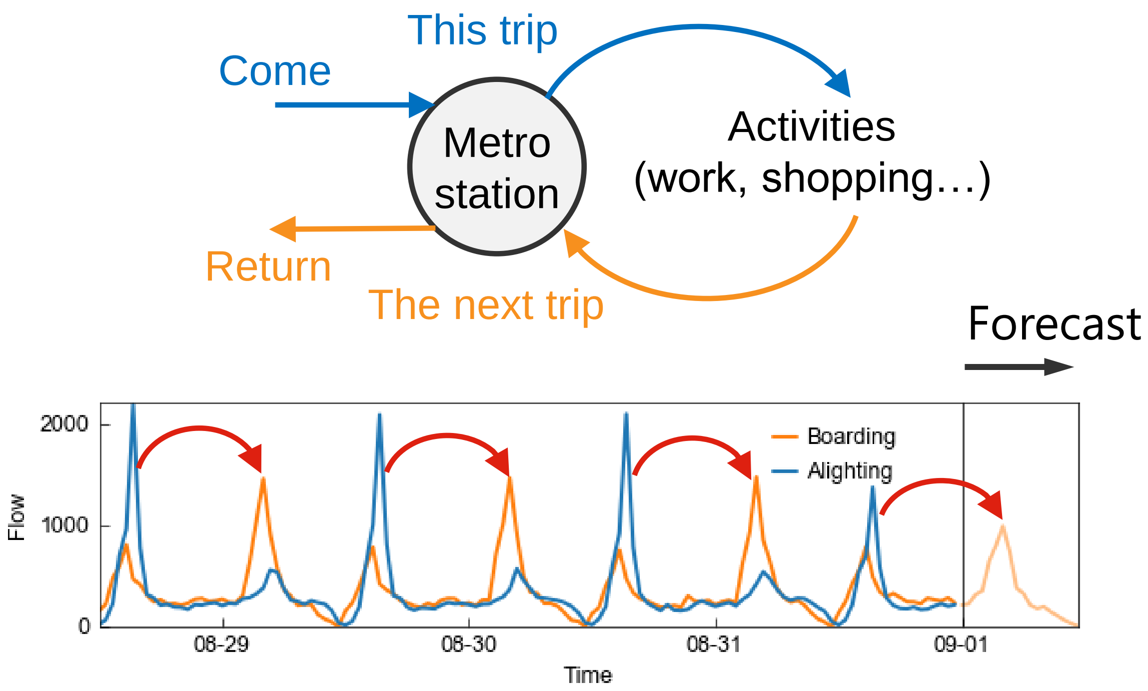
Boarding flow forecasting

Traditional approaches:
Learn from time series



- Local correlations.
- Overlook the generative mechanism of boarding flow.

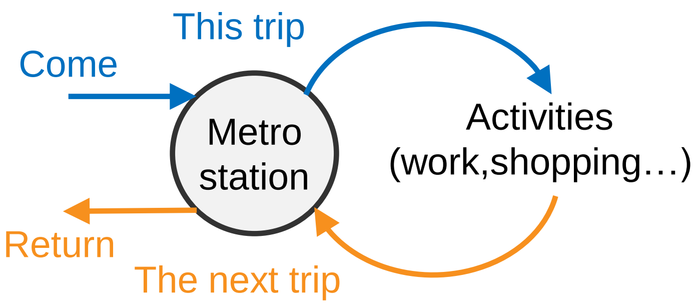
Proposed approach:
Using travel regularity and chained trips



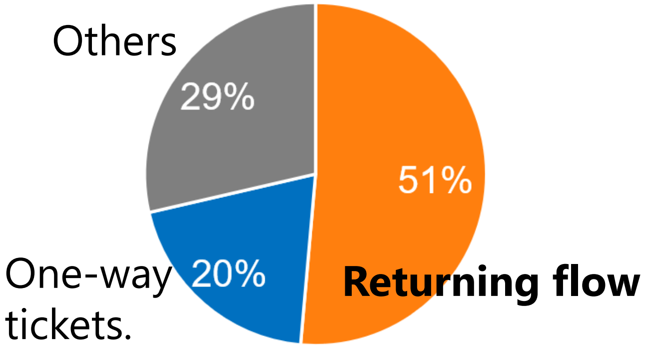
- Long-range correlations.
- Capture the trip generative mechanism.

Cheng, Z., Trépanier, M., Sun, L., 2021. Incorporating travel behavior regularity into passenger flow forecasting. *Transportation Research Part C: Emerging Technologies*.

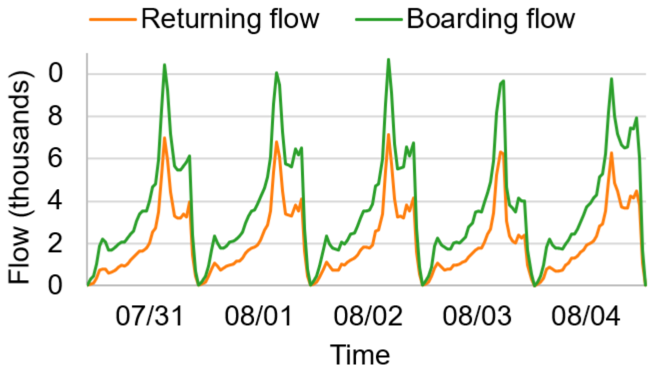
Returning flow



Returning flow r_t^s: the number of people who finish their activities and start their return trips at time t by the same station s.



- **More than 50%** of metro trips in Guangzhou are returning trips.

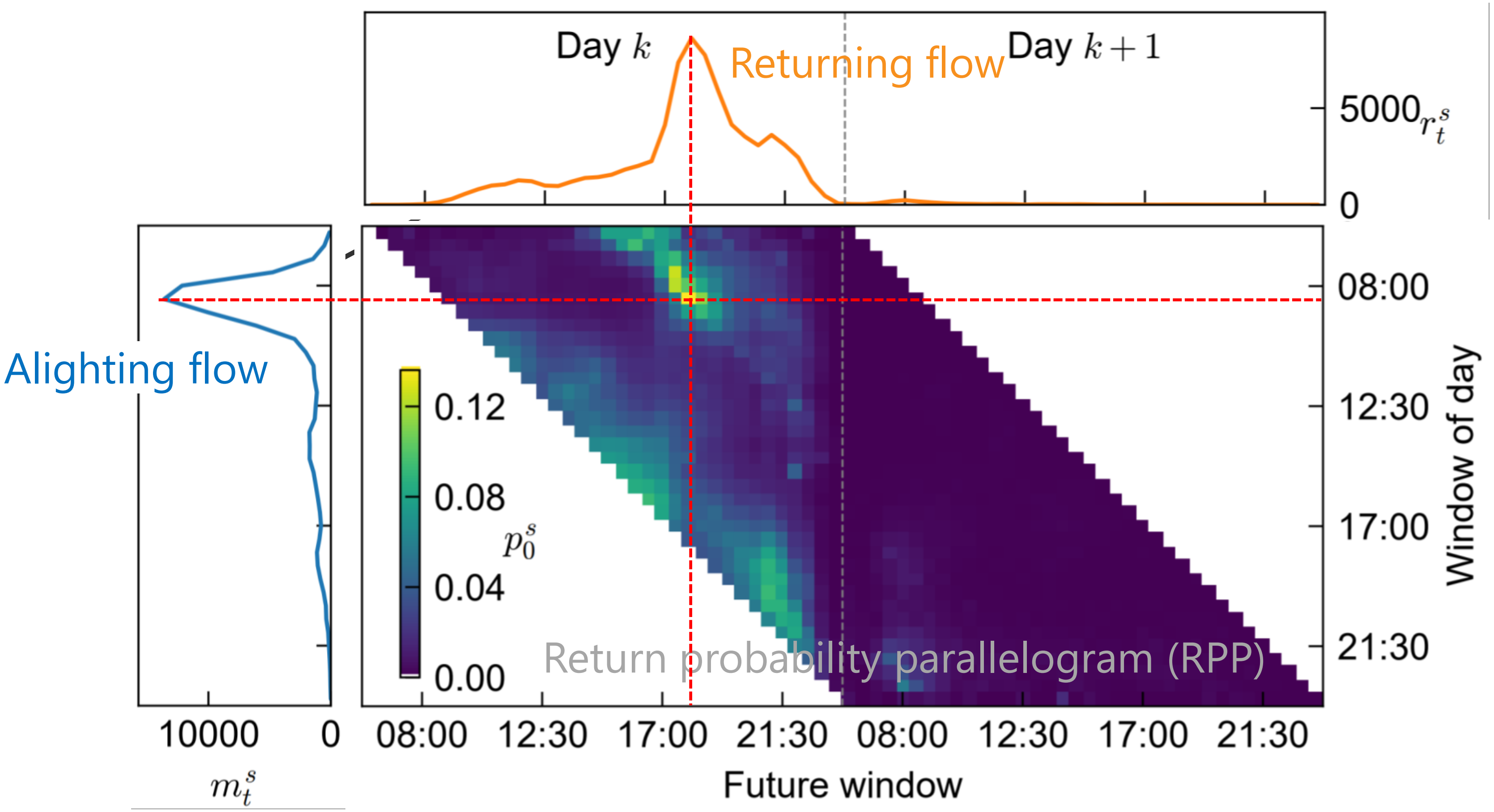


- Returning flow is **highly correlated** with boarding flow.

(1) A baseline boarding flow forecasting model: $\hat{y}_{t+1}^s = f(y_{1:t}^s)$.
(2) Using returning flow as a covariate in the forecasting: $\hat{y}_{t+1}^s = f\,(y_{1:t}^s, r_{t+1}^s)$

- **The forecasting of (2) is better than (1).**
- **How to obtain r_{t+1}^s?**

Forecast the future returning flow



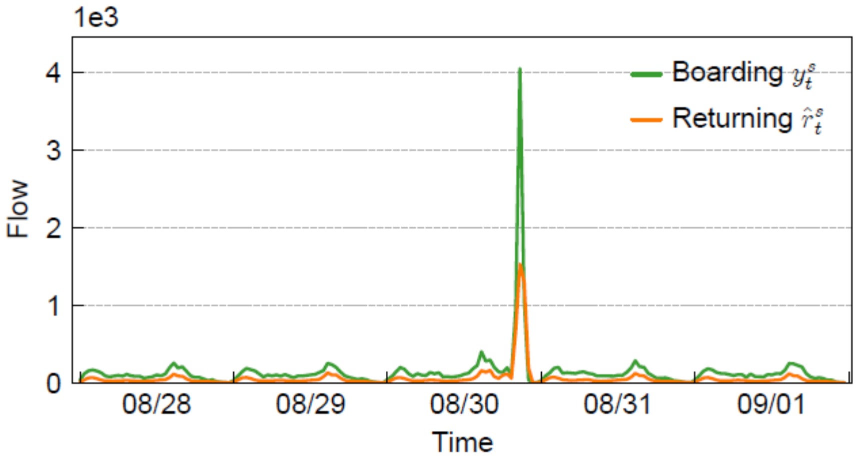
Estimate future returning flow: $\hat{r}_{t+1}^s = \sum_{h=1}^{H} m_{t-h+1}^s p^s(\tau_{\text{boarding}} = t+1 | \tau_{\text{alighting}} = t-h+1)$

The effect of using the returning flow

Multi-step boarding flow forecasting of a station.

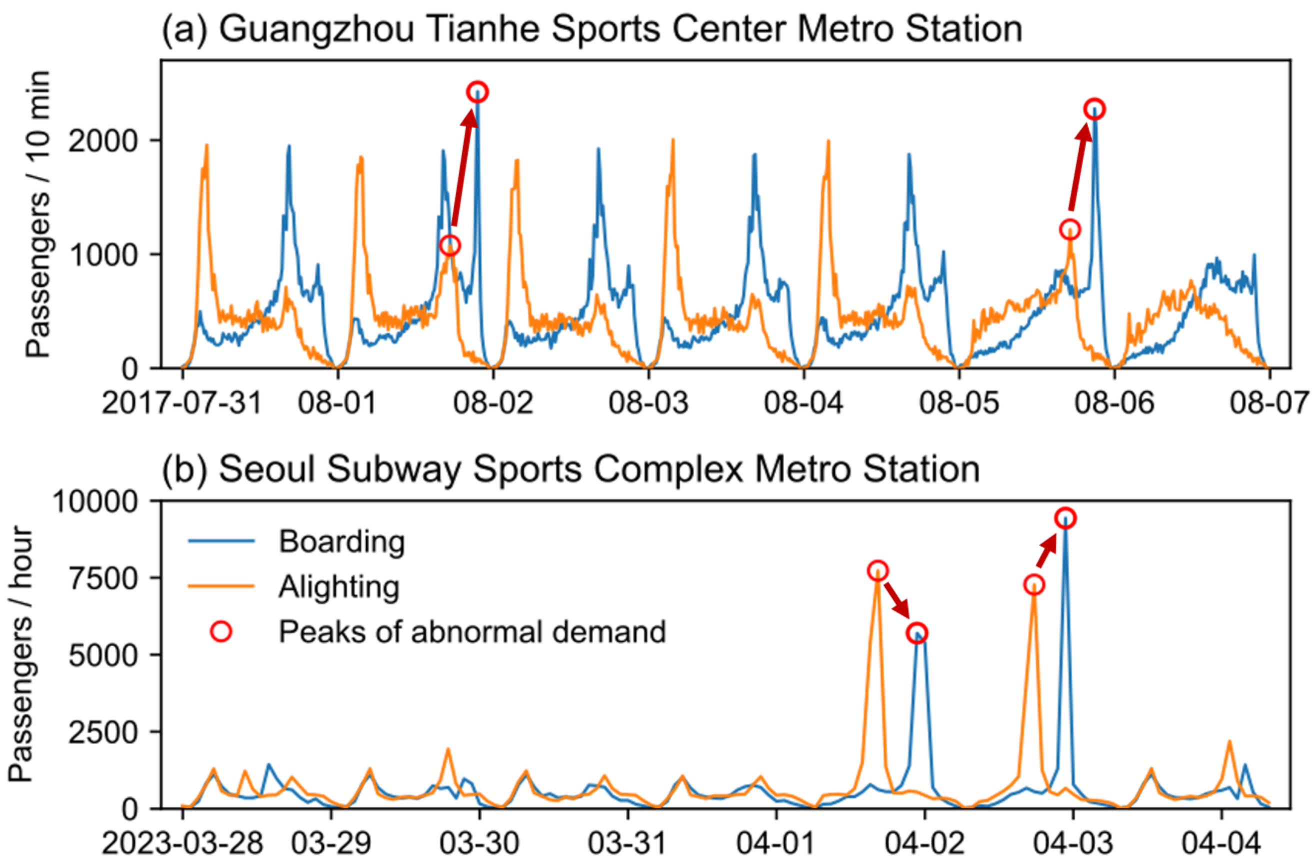
	Station	Model	30min (1 step)		1hour (2 step)		2hour (4 step)		3hour (6 step)	
			RMSE	SMAPE	RMSE	SMAPE	RMSE	SMAPE	RMSE	SMAPE
Business area	(a) Tiyu Xilu	SARIMA	334.06	12.92%	435.16	15.30%	547.35	19.88%	586.61	19.23%
		SARIMA + $\hat{r}^s_{t+n}$	**290.43**	**9.87%**	**328.90**	**12.60%**	**403.20**	**10.82%**	**465.09**	**11.71%**
Residential area	(b) Luoxi	SARIMA	**75.94**	10.64%	79.53	11.70%	88.49	12.14%	90.16	12.51%
		SARIMA + $\hat{r}^s_{t+n}$	76.10	**10.60%**	**79.43**	**11.65%**	**88.49**	**12.08%**	**89.97**	**12.38%**

1) Behavior-based method significantly improves long-range (multi-step) forecasting.
2) More effective for stations in business areas (Because of short activity duration).
3) Shed new light on forecasting under special events.



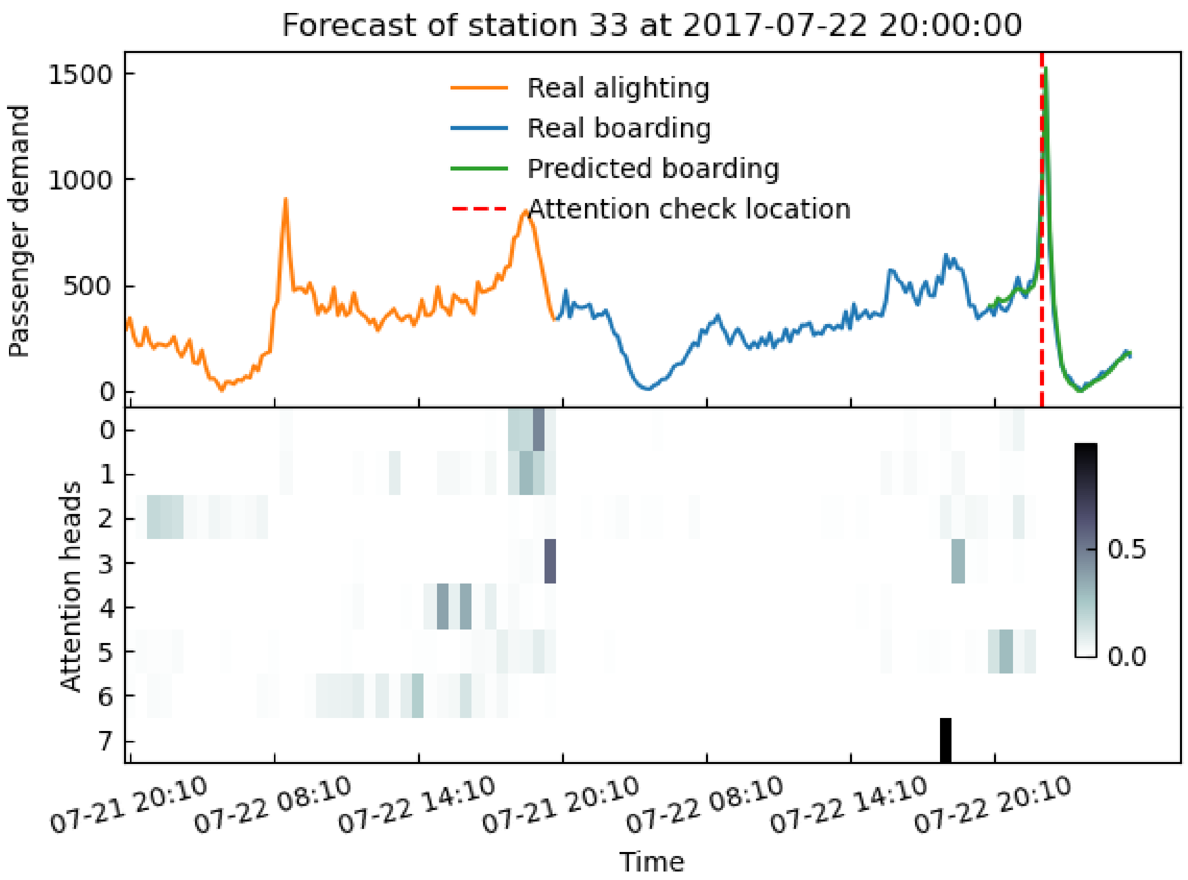
Estimated returning flow in an event.

Passenger flow during events



Irregular surges of metro passenger boarding demand

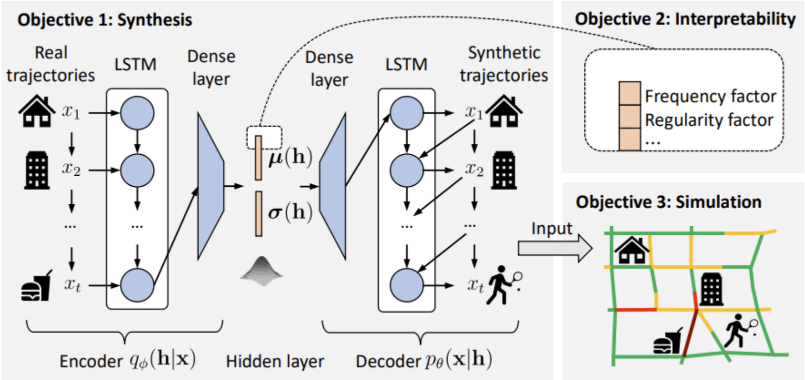
Passenger flow forecast during events



Interpreting boarding flow forecasting with attention weights in a Transformer model.

Future directions

- **Individual mobility prediction**: for better trip planning recommendation.
- **Address the privacy concern**: even a few records of individuals' spatiotemporal locations can uniquely identify a person [1].
- **Mobility synthesis**: generate fake but realistic individual trajectories. The synthesized mobility dataset can be published without no privacy concerns.



- **Generative models for urban mobility**: generate full trajectories based on partially observed trajectories, e.g., generate full trajectories based on data only from metros.

[1] Gao, J., Sun, L., & Cai, M. (2019). Quantifying privacy vulnerability of individual mobility traces: A case study of license plate recognition data. Transportation research part C: emerging technologies, 104, 78-94.

The body of this talk is based on research during my Ph.D. study.

Supervisors:
Prof. Lijun Sun
Prof. Martin Trépanier

Thank you!
Questions?

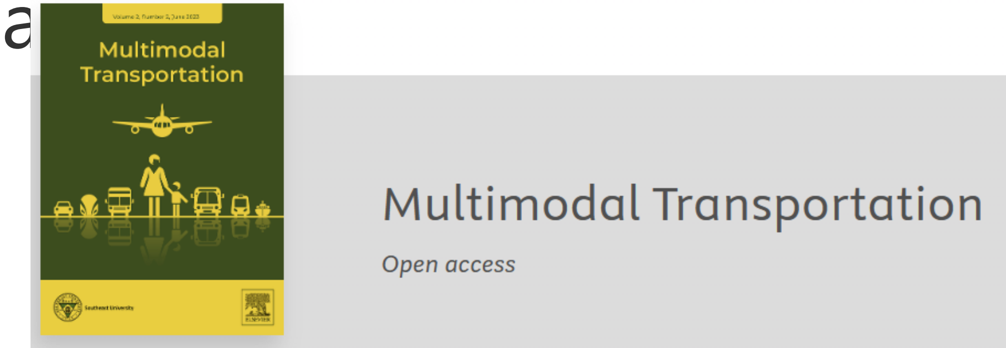
zhanhong.cheng@mcgill.ca

McGill Engineering Doctoral Awards (**MEDA**)

广州地铁
Guangzhou Metro

ScienceDirect

Ca



- *MULTRA* provides a forum for high quality, cutting-edge research in transportation science and technology.
- Papers concerning multimodal integration or emerging transportation technologies are particularly welcome

Upcoming special issue:

Innovative applications of operations research and machine learning in traffic and transportation management