Learning Transportation Insecurity from Mobile Phone Data

Marco Nie

2025 CCAT Global Symposium on Mobility Innovation

A joint CCAT project with Alex Murphy, Ying Chen, Tianxing Dai, Peeter Kivestu
Previous collaborations with Gretchen Bella, Amanda Stathopoulos

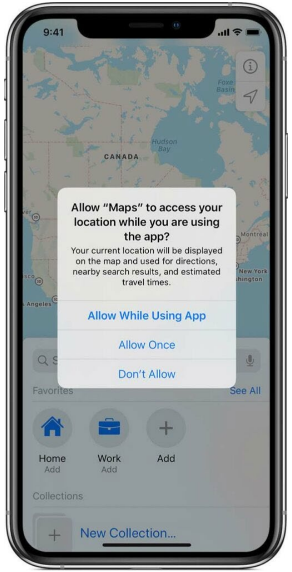
Northwestern

Outline

- Introduction
- Case study
- Learning mobility insecurity

Mobile phone data

- **Location-based service (LBS) data**
- **Sources**
  - GPS, Wi-Fi, cell tower
- **Commercially available**



- **What’s in the data?**
  - Timestamps
  - Location of the device at each timestamp
- **What’s not in the data?**
  - Identifiable personal information

Mobile phone data

- **Our dataset (PickWell Sp. z.o.o)**
  - February 2021 and May 2022
  - City of Chicago
  - 100,000+ devices in each period
  - About 4% of the Chicago population

Preprocessing

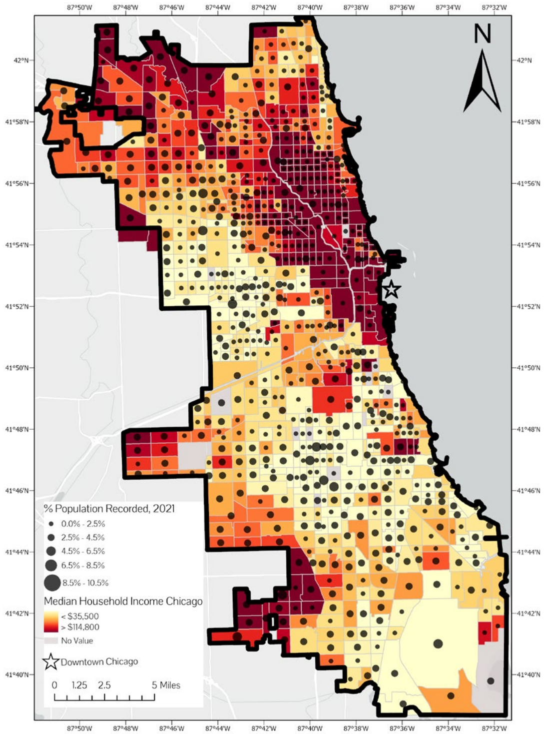
- **Activity**
  - A group of data points that stay roughly at the same location within a given interval
  - An activity is located to a census tract, and labeled by duration as
    - Home (a device spent most time outside the typical work time window)
    - Primary (a device spent most time other than Home)
    - Secondary (a device spent the second most time other than Home).
- **Trip**
  - A group of points that connects two consecutive activities.

CASE STUDY

Is mobile phone data representative?

- **Over-representation of the minority and low income groups**
  - Confirmed by statistical analysis
- **Impact of Apple privacy policy:**
  - User “opt-in” since Nov 2021
  - iOS users under-represented after that

iOS devices	2021 (“before”)	2022 (“after”)
Actual Market Share	56%	54%
Mobile phone data market share	51%	37%



2021data

Takeaways

- Mobile phone data can have significant representation biases.
- Opportunity to study minority and disadvantaged groups since they appear to be over-represented.
- Potential to complement and enrich traditional surveys.

Bella, Gretchen, Tianxing Dai, Peeter Kivestu, Marco Nie, and Amanda Stathopoulos. 2024. "Device Tracking Privacy Regulations Lead to Unexpected Data Bias in Smartphone Trace Data." *Available at SSRN*.

LEARNING TRANSPORTATION INSECURITY

Transportation security index (TSI)

- The condition in which a person is **unable to regularly get from place to place in a safe or timely manner due to an absence of resources necessary for transportation**.
- TSI is created from a questionnaire designed by Prof. Murphy's team, which includes a range of questions:

In the past 30 days, how often

1. Did you have to **reschedule appointments** because of a problem with transportation
2. Did you **skip going somewhere** because of a problem with transportation
3. Were you **not able to leave your house** because of a problem with transportation
4. Did you **feel bad** because you did not have the transportation you needed
5. Did you **worry about inconveniencing your friends, family or neighbors** because you needed help with transportation
6. Did problems with transportation **affect your relationships with others**

Transportation security index

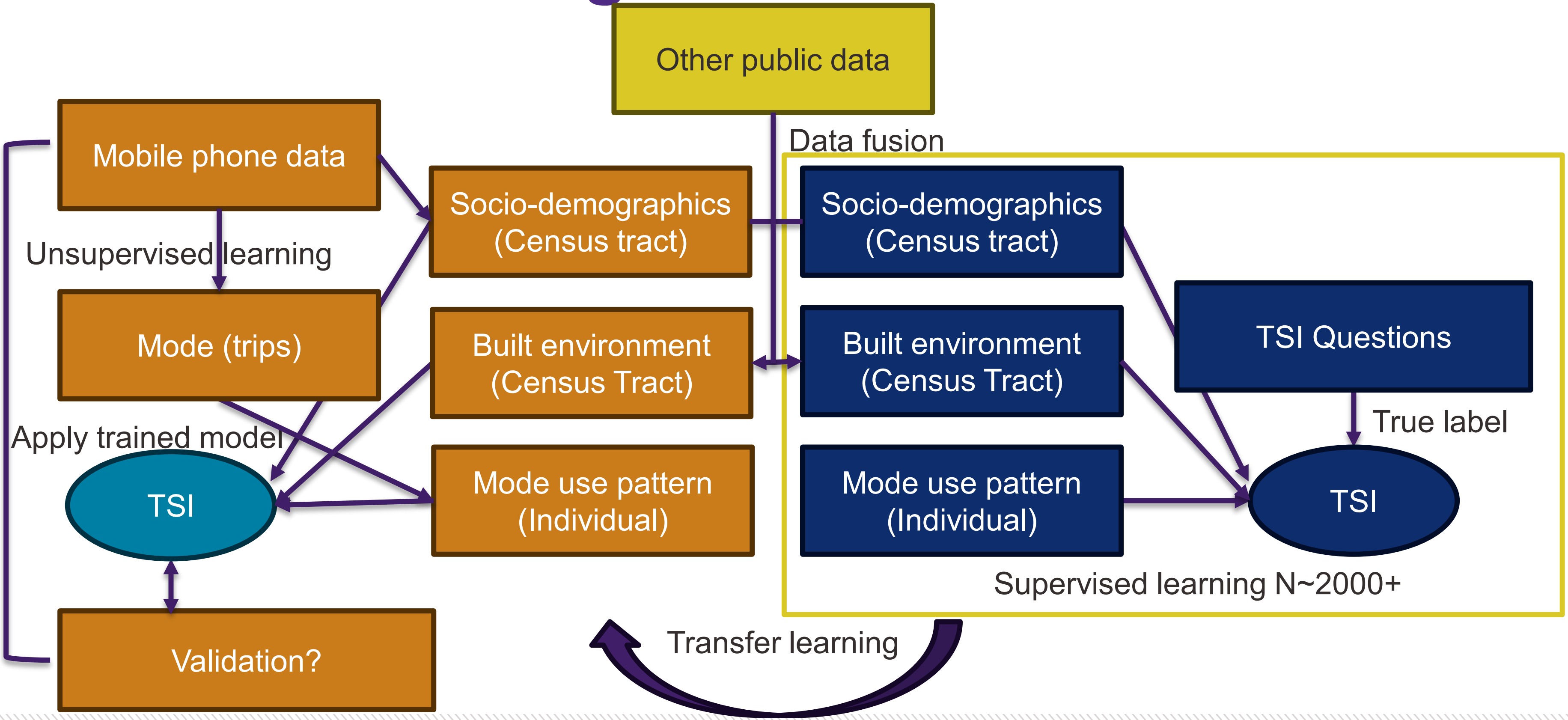
- **TSI was implemented in DMACS wave 18 (Nov 2023)**
  - 2000+ respondents
- **Minority, low-income and female respondents** are more likely to experience transportation insecurity
- **Transportation insecure people tend to:**
  - Have no car
  - Use more than three modes for daily travel

Research questions

- Can we extract mobility features from mobile phone data that are related to transportation insecurity?

- Can we infer transportation insecurity using those features?

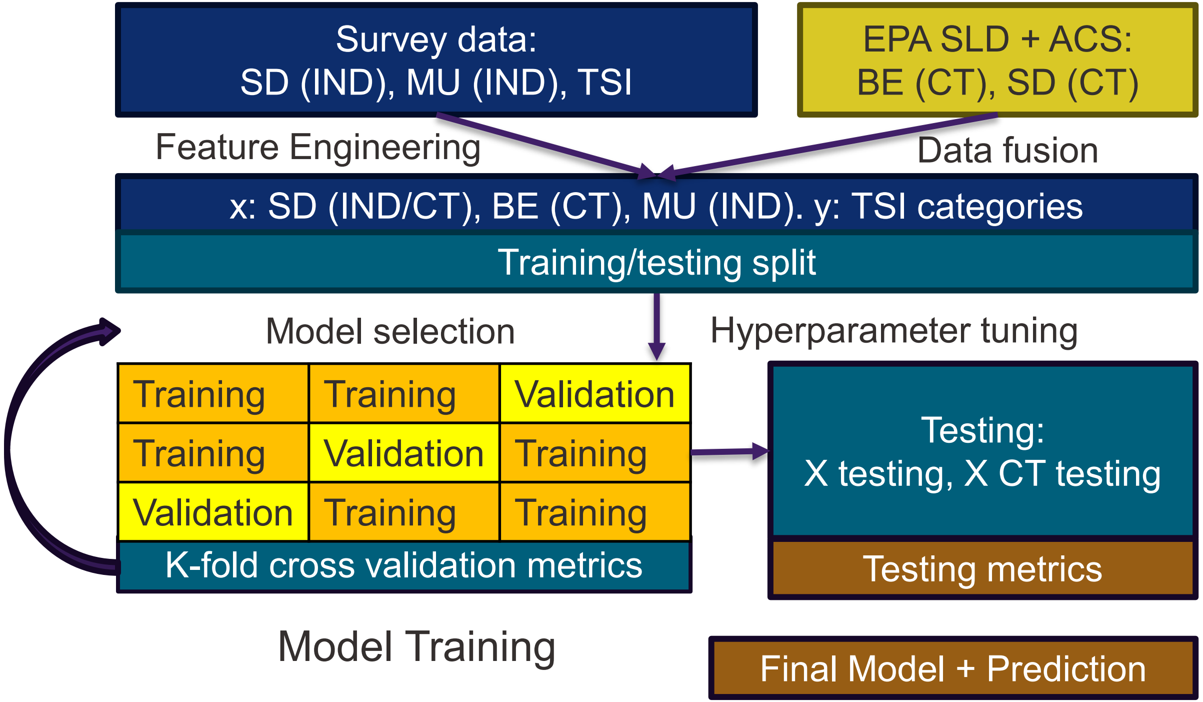
A transfer learning method



Classification model setup

- **Dependent variable: secure (1) or insecure (0) (labels derived from the 6-item TSI questionnaire)**
- **Predictors:**
  - Built environment (e.g. road density, transit frequency, etc.) – **BE**
  - Socio-demographics (e.g., income) – **SD (CT), SD (IND)**
  - Mode use (e.g. use frequency of different modes) – **MU**

Classification model setup



Data processing

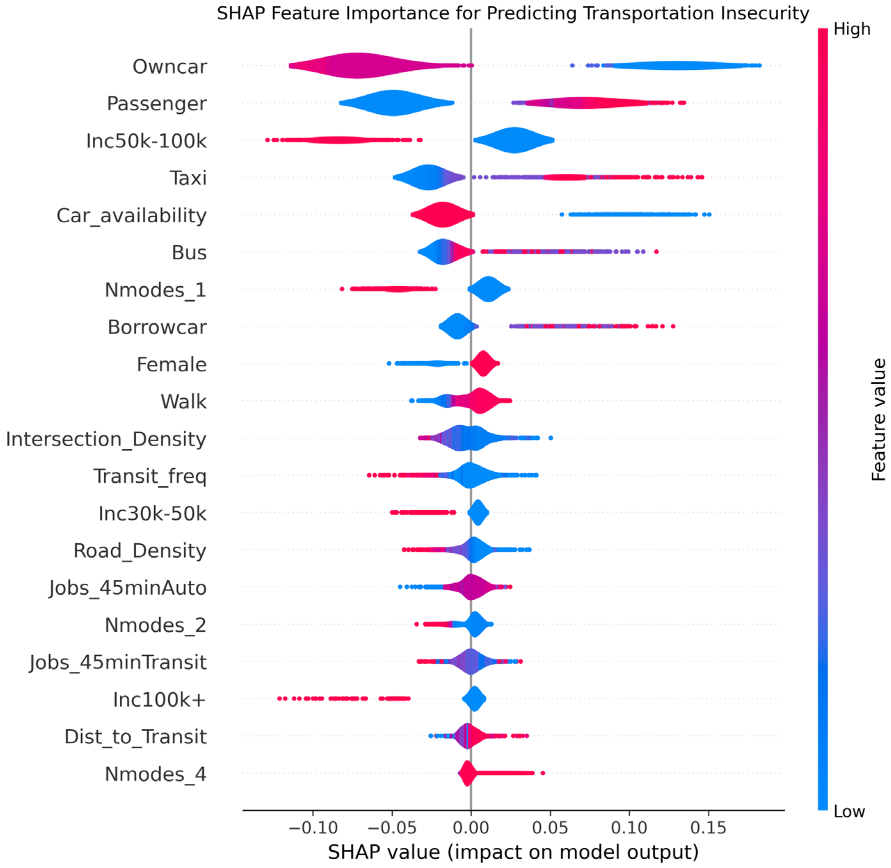
Model Performance

Results (Random forest classifier)

- **Key metrics for model evaluation**
  - Accuracy = correct predictions/all predictions
  - Specificity (identify insecure population) = true neg/(true neg + false pos)
- **Performance for testing data using SD (IND)**
  - Accuracy: **0.77**
  - Specificity: **0.81**
- **Model performance for testing data using SD (CT)**
  - Accuracy: **0.73**
  - Specificity: **0.81**

Results

- **Predictor of insecurity**
  - Owncar (-)
  - Passenger (+), taxi (+), bus (+), borrowcar (+), walk (+)
  - Higher income (-), female (+)
- **Mode use patterns are stronger predictor of transportation insecurity**



Next steps

- Obtain predictors from mobile phone data
  - Built-environment and social-demographics: from Home location
  - Mode use:
    - Unsupervised learning methods to infer modes for each trip
    - Assemble at device level to get mode use features
- **Adjust the prediction model to accommodate mobile phone data.**
- Develop validation methods.

Thank you!

Email: y-nie@northwestern.edu

Representation bias paper

Telework paper

Northwestern

Methodology