CENTER FOR GEOSPATIAL INFORMATION SCIENCE

Travel behavior pattern analysis using passively collected location-aware mobile device data

Kathleen Stewart

Department of Geographical Sciences

Center for Geospatial Information Science

University of Maryland, College Park, MD USA

Human mobility analytics in the Big Data Era

The 3rd ISDE International Lecture Series, March 21, 2023

Mobility and human-vehicle interactions

- **Driving behaviors** are very important to understand in order to e.g., **reduce** the **number** and **types of accidents**, **improve efficiency** and **reduce fuel consumption**, and **improve** the **driving experience** overall
- They also contribute to an **understanding** of other **application domains**, e.g., health, economics, environmental topics where mobility patterns are relevant for domain-related research questions
- How do we get information about driver behaviors?
  - **Survey** travelers
  - **Simulate** driving behaviors
  - Install **onboard sensors** and directly measure behaviors
  - Collect and analyze **passive mobility data**
- Especially interested in **naturalistic driving**, i.e., driving behaviors experienced on roads during daily travel, i.e., patterns arising from everyday activities

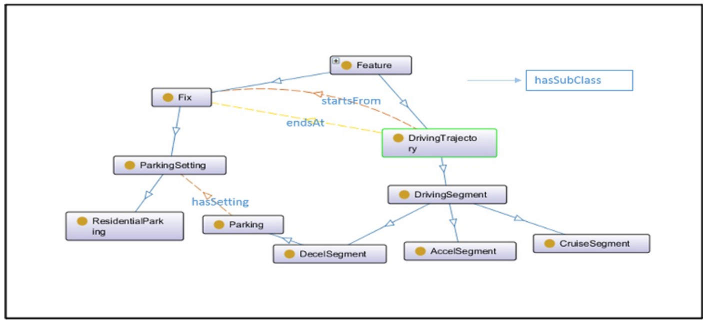
- One of our early investigations into driver behaviors involved a study of **naturalistic driving behaviors** using **onboard sensors** (not passive…)
- This study was designed to investigate how to possibly **decrease common driver errors** including **pedal misapplications** that can result in serious crashes and personal injuries.
  - Young et al. 2011 found that drivers engaged in **parking** generally produced more significant **pedal misapplication errors** than while driving on the road (e.g., missing the brake entirely).

- We designed a **geospatial model** of **vehicle movements** and **driver behaviors** associated with **parking**

This involved:

- Consideration of the **basic elements** of parking
- Identifying the key elements for a **formal geospatial model of parking**
- Designing an algorithm for **extracting parking segments** from big data driving trajectories (acceleration thresholds, parking distance, time)

- What does **analysis of vehicle movements** teach us about the **behaviors of individuals while driving**

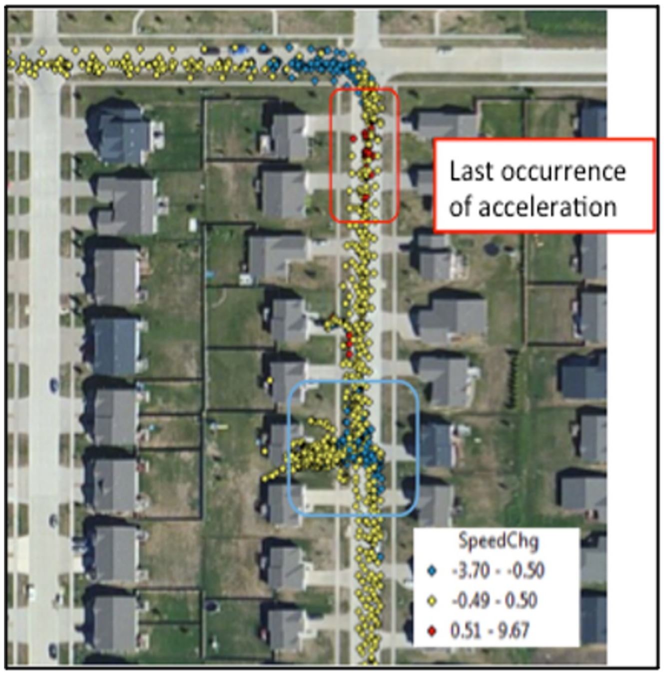


w/ J. Fan

Driving trajectory ending with residential parking

Parking distance is defined as the driving distance between start and end points.

This varied among drivers



Start of parking behavior

End of parking behavior

Study participant 009F

With 0.5m/s^2 as acceleration threshold

This threshold is used to define the start of parking

J. Fan

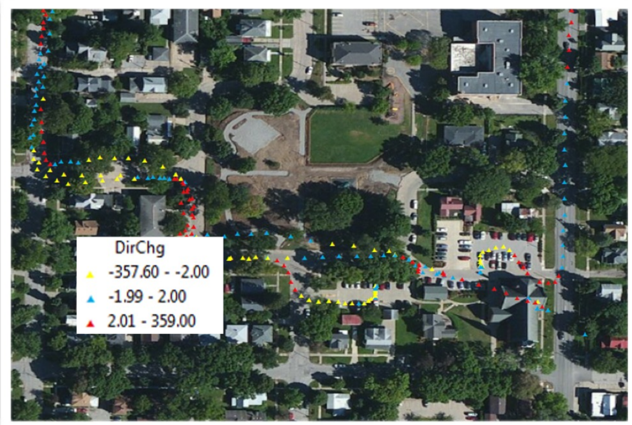
009F

Speed changes on approaches to parking

001M

Direction changes while driving to parking lot

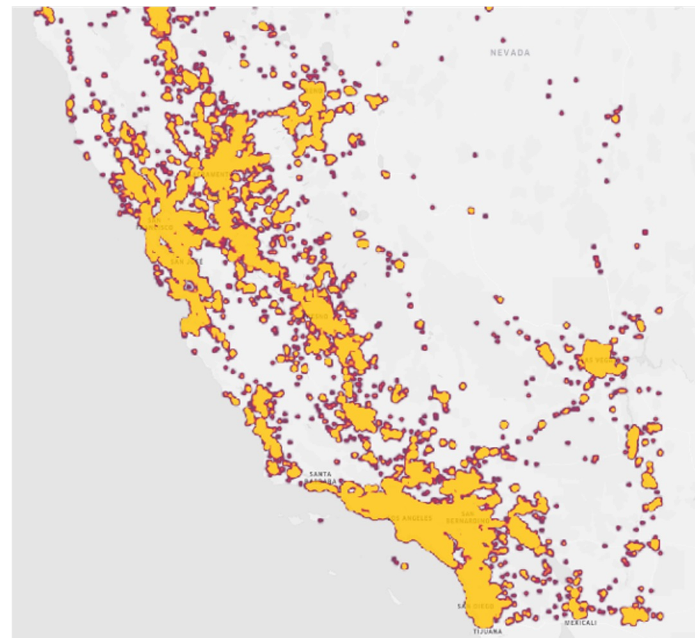
009F



001M

Passively collected mobility data

- With the rapid rise and prevalence of mobile technologies, **passively collected mobile data** has become more available
- Different from the traditional actively collected data, these data have the advantage of **larger data volumes**, broad **temporal** and **spatial distribution** and **coverage**, and **high collection frequency**
- Provides new opportunities for **large-scale** and **fine-grained** human **mobility analyses** and **travel behavior analyses**
- **Opens new doors** for analyzing, modeling and predicting travel behaviors
    - While **passive mobility data** brings **opportunities**, it also brings **new challenges** that we will discuss.



Heatmap of origin and destination points of passively collected mobile device data for 02/01/2020 in CA, estimated to be 5% sample. *From P. Zhang*

Speeding behaviors (P. Zhang)

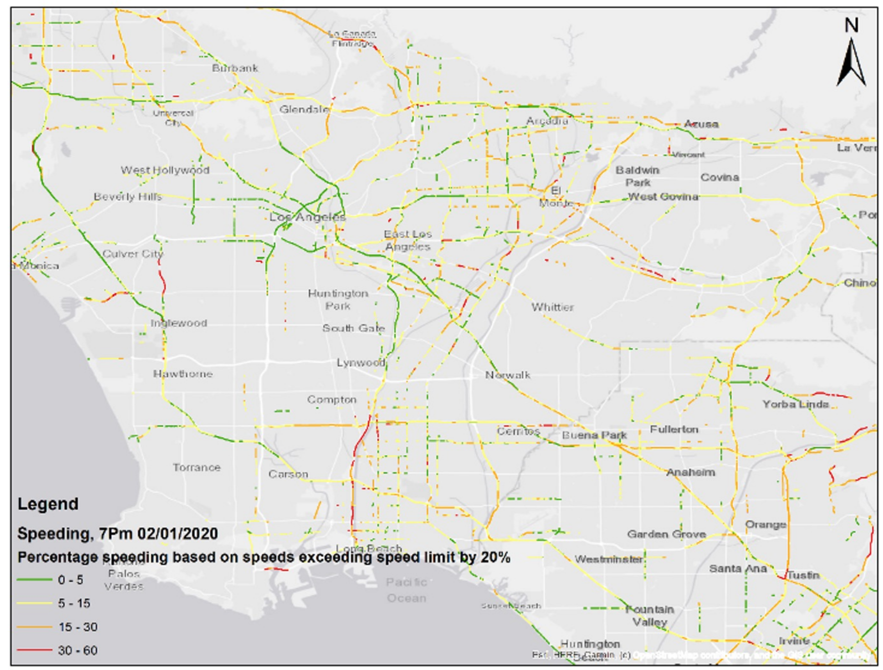
- A key behavior experienced during driving is **speeding**, where a vehicle is driven at speeds that **exceed the legal speed limit** or that may be too fast for road conditions or allow a driver to maintain control
- We have been investigating methods to **estimate vehicle speeds** from passively collected mobile data and **identify speeding** on roads

Speeding behavior analysis using passively collected data

- For this research, we measured speeding using the **proportion of waypoints that exceeded the speed limit** by a certain percentage as compared to all waypoints in a certain time interval.
- A **speeding index** is defined as:

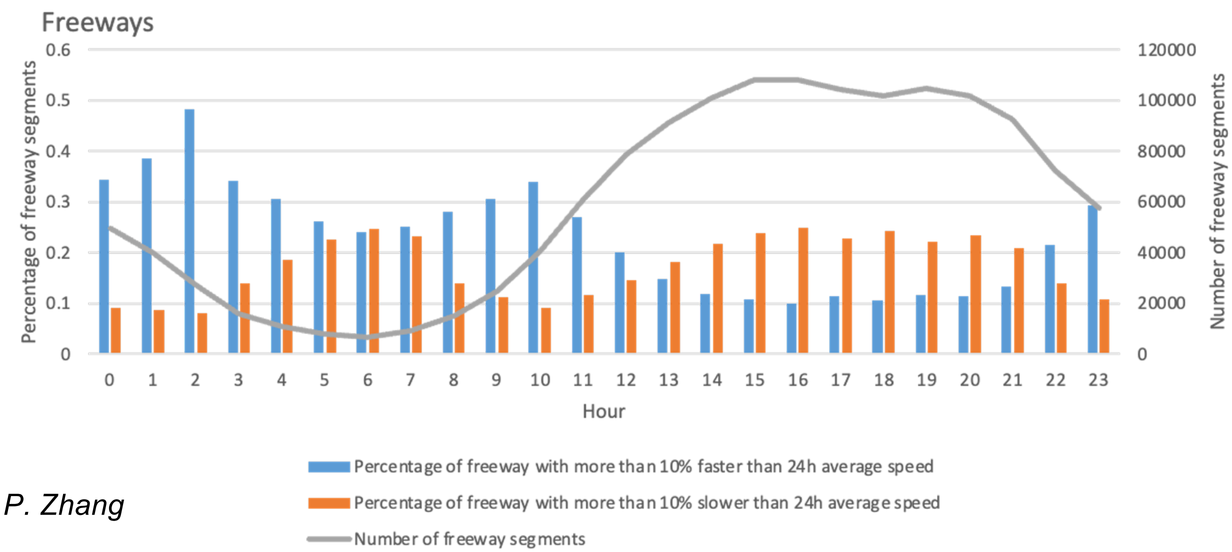
$$I_{threshold} = \frac{N_{r(speed > speed\ limit * (1+ threshold))}}{N_{r(all)}}$$

- Where $I_{threshold}$ is the speeding indicator for a certain threshold, i.e., speeds 20% and 50% faster than the speed limit.
- Certain roads in CA, e.g., in central Los Angeles showed a higher proportion of drivers that were speeding. (more than 45% of segments on these roads had more than half of the drivers speeding at more than 50% of the speed limit)



Speeding on California freeways at 7PM, 02/01/2020 where driving speeds were 20% over the speed limit, *P. Zhang*

Temporal patterns of traffic speed



P. Zhang

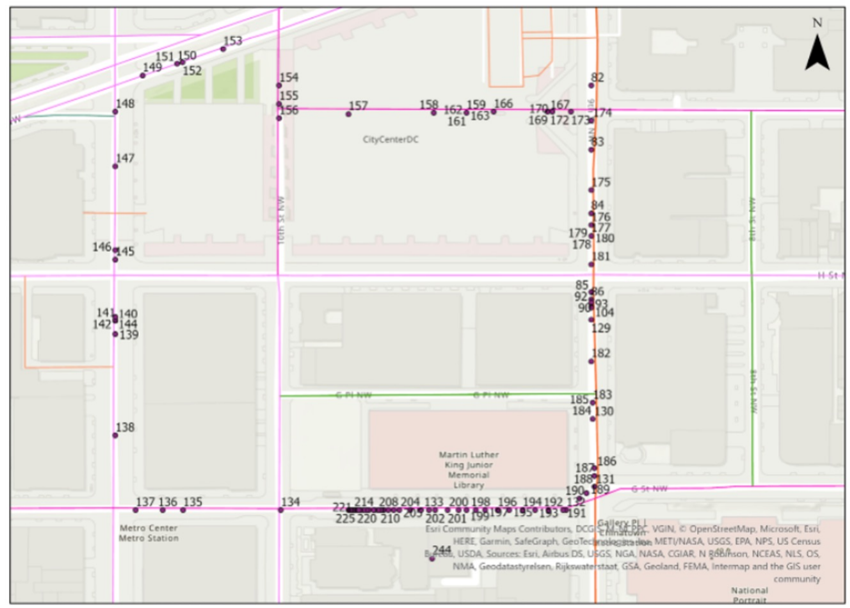
Using passively collected data we could estimate **driving speeds** for a **large percentage of road segments** including especially freeway road segments (motorways, trunk roads, and primary and secondary roads)

- For this day of data, and compared with the **all-day average speed** of each segment, the speeds of drivers on **freeway segments** showed that **more road segments had faster speeds at night**, decreasing steeply from 3am to 6 am.

- Since the mobility data was collected for a **Saturday**, our results do not reflect commuting patterns with morning and evening peaks corresponding to home-work trips but rather what **weekend trips** look like.

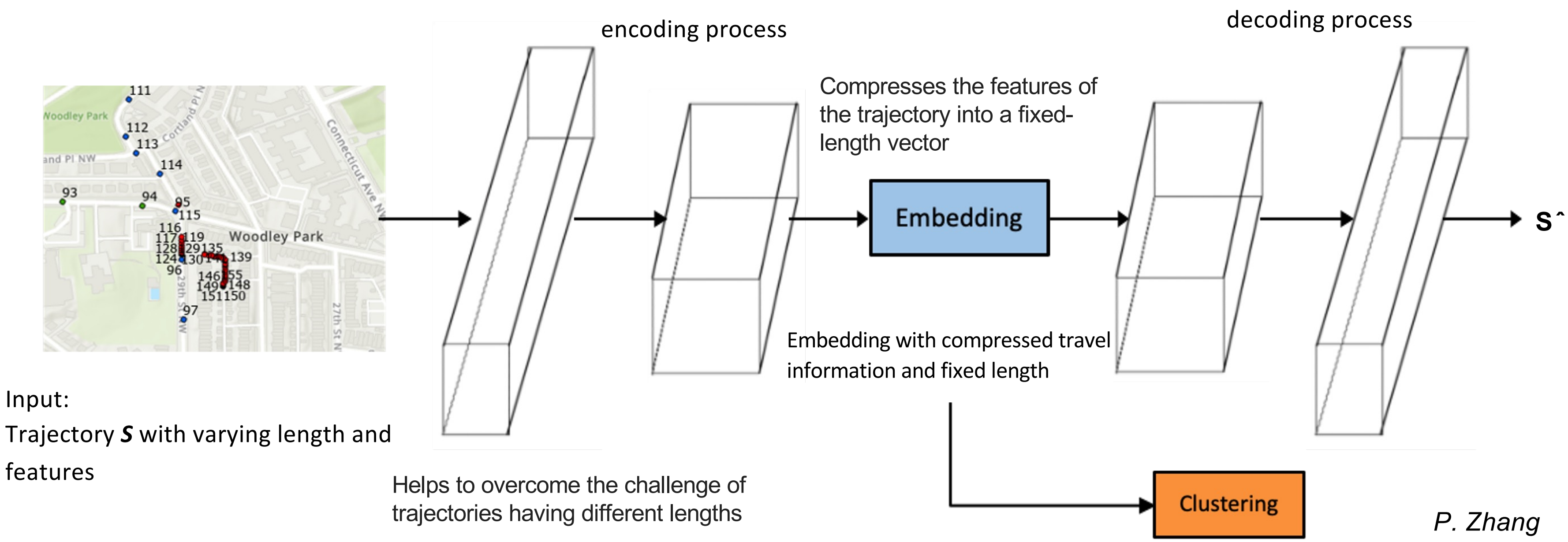
Parking behavior analysis using passive mobility data (P. Zhang)

- In addition to **speeding,** we have been using the **passively collected mobility data** to extract **parking behaviors**
- We are working on approaches to **identify** and **classify different parking behaviors**
- **Deep learning-based unsupervised classification** is being tested to overcome the limitations of no direct labels for parking in the data, and identify parking behaviors.

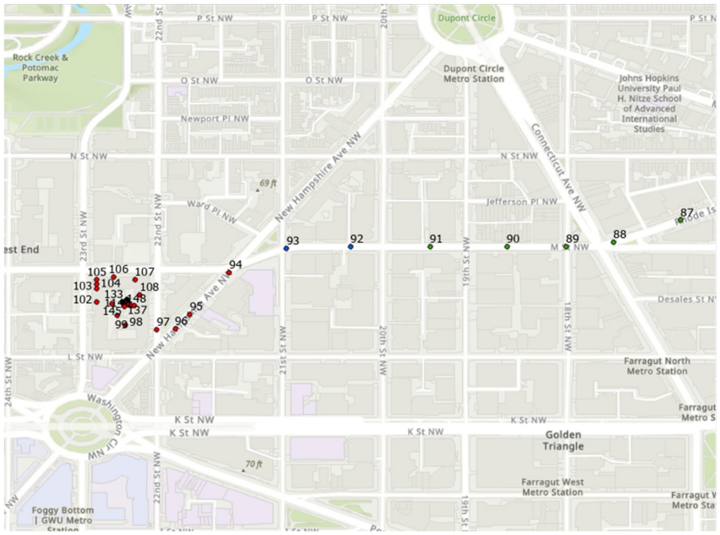


An example of cruising for parking in Washington, DC, *P. Zhang*

Parking behavior classification using autoencoder-decoder deep clustering



For **classifying trajectories into different types of parking**, we use an **LSTM autoencoder-decoder deep clustering model** to overcome the challenges brought by the **lack of labels** and **varying lengths** of parking trajectories.



Buffer radii	Avg speed	Avg angle
250 m	4.83	39.60
500 m	32	11.32
1000 m	34.5	3.65

P. Zhang

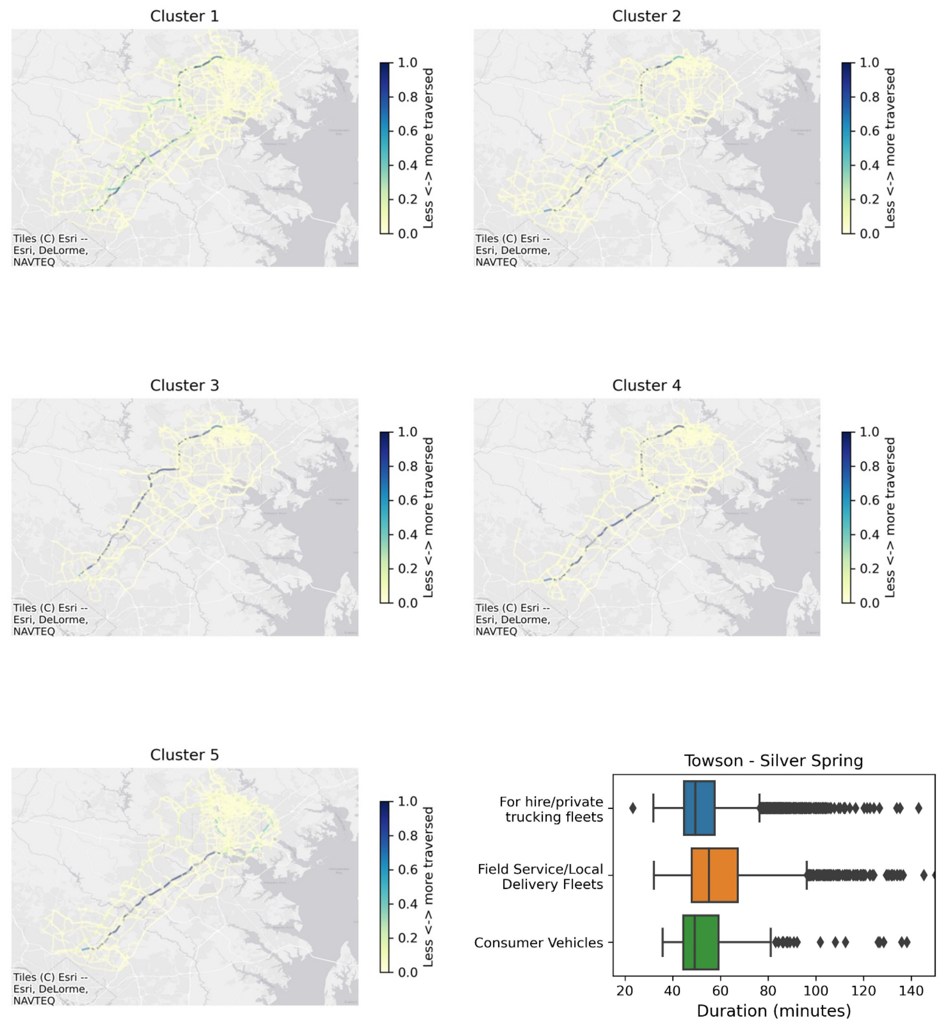
- For this research, we are combining definitions for identifying parking trajectories from previous studies (**acceleration thresholds**, **buffer distances**…) and are developing a working definition that includes, using **a buffer area** around the destination point and **certain travel speed thresholds**
- Classifying **cruising for street parking** and **garage parking**

Travel time estimation (G. Zhu)

- In a study on **travel time by drivers**, we analyzed over 9,000 trips from 2019 for travel from Towson, MD to Silver Spring, MD

- We applied **K-Means with DTW clustering** to identify different patterns of **route choices** generating 5 clusters.

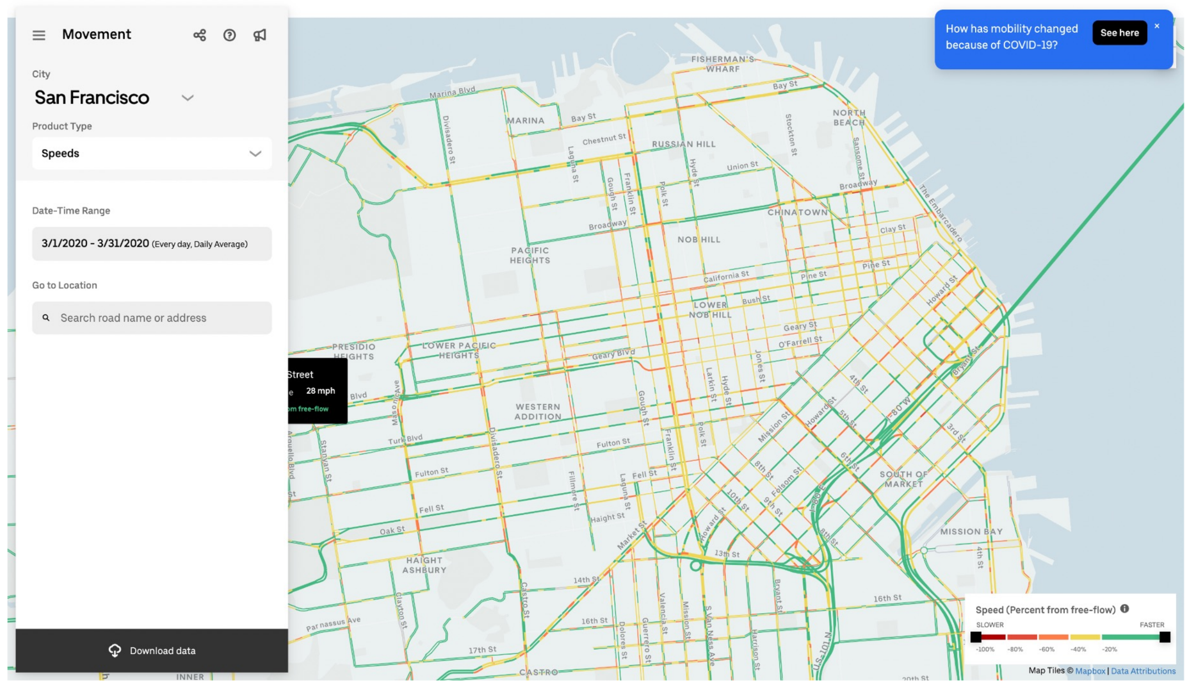
- **Random forest** models have been applied to examine how different factors impact travel time, including:
  - Departure time of day, day of week, vehicle profile (driver behavior)
  - Percent of motor/highway used, direction of route, changes of road classes (road network)
  - Incident duration, precipitation, snow (external factors)



G. Zhu

Mobility datasets are available e.g., Uber movement

- **Uber Movement** released travel time, aggregated travel speed, and mobility heatmap datasets for **certain large cities** around the world.
  - The **travel speed dataset** is aggregated for one-hour intervals for road segments, and the **travel time dataset** is aggregated for small zones.
  - Both datasets are based on GPS trajectories collected from **Uber drivers.**



Challenges with passively collected mobility data

- Due to the **massive data volume** and **large spatio-temporal coverage**, many mobility analysis methods (e.g., clustering-based methods) can be **hard to use** and need to be **scaled up**
- Passively collected trajectories may have **varying sampling intervals**
  - trajectory waypoints may present **different sampling errors generated by the GPS sensors** in the mobile devices (7-13 m horizontal error on average with a maximum error within 100 m)
  - **Not all roads are covered**, roads that people travel on the most are more likely to be recorded in the passively collected datasets
- Trajectories representing similar mobility behaviors (e.g., parking) may exhibit **different spatial and/or temporal scales**.
- **Duplicated waypoints**
- It is not uncommon for incorrect trajectories to be generated due to **reversed GPS points** (i.e., the next GPS point is behind the previous GPS point on the road), which requires additional pre-processing work.
- Many (most?) behaviors (e.g., parking) are **not explicitly labeled**
  - **Driver profiles** (e.g., age, gender, travel purposes, etc.) that could provide rich information on travel behaviors, are usually **not included** in the dataset (to protect driver privacy).

Biases

- Behaviors are being inferred from s**mall sample sizes**
- **Differences in technology** (e.g., smartphones) use can lead to **how or whether individuals are captured** in data collections
  - not everyone has a smartphone
- Access to digital technologies are often closely linked to different **socioeconomic, geographic, race/ethnic, gender, age, ability, and class disparities**
  - Mobility patterns of certain groups may be captured more frequently (**selection biases**)
  - Commuting patterns, shift workers
- **Different OS settings** may affect **geolocation accuracy**

Summary and conclusions

- **Driving behaviors** are very important to understand in order to e.g., **reduce** the **number** and **types of accidents**, i**mprove efficiency** and **reduce fuel consumption**, and **improve** the **driving experience** overall
- Interested in **naturalistic driving** captured by **passively collected big** mobility data
- Examining methods for **behaviors** including **speeding**, **parking**, **travel time estimation**
  - **spatial methods (map matching** and **gap filling)**
  - **machine learning (random forest), deep learning (Autoencoder-decoder), clustering (K means, DBSCAN** and **Dynamic Time Warping)**
- **Validation** using **manual classification** and **actively collected datasets.**
- **Biases** need to be considered and analyzed

Selected related papers

Zhu, G., K. Stewart, D. Niemeier, J. Fan (2021) Understanding the Drivers of Mobility during the COVID-19 Pandemic in Florida, USA using a Machine Learning Approach, special issue on Geospatial Approaches for Understanding the Social, Economic and Environmental Impacts of COVID-19, *International Journal of Geo-Information.* 10(7), 440; https://doi.org/10.3390/ijgi10070440

Fan, J. and K. Stewart (2021) Understanding collective human movement dynamics during large-scale events using big geosocial data analytics, *Computers, Environment and Urban Systems,* 87, 101605.

Fan, J., C. Fu, K. Stewart, L. Zhang (2019) Using Big GPS Trajectory Data Analytics for Vehicle Miles Traveled Estimation, *Transportation Research Part C: Emerging Technologies,* 103:298-307.

Nasri, A., L. Zhang, J. Fan, K. Stewart, H. Younes, C. Fu, S. Jessberger (2019) Advanced vehicle miles traveled (VMT) estimation methods for non-federal-aid system roadways using GPS vehicle trajectory data and statistical power analysis, *Transportation Research Record.* https://doi.org/10.1177/0361198119850790

Stewart, K., J. Fan, C. Schwarz, and D. McGehee (2018) Geospatial analysis of residential parking behaviors using a semantic modeling approach, *Travel Behavior and Society*, 11: 9-20. https://doi.org/10.1016/j.tbs.2017.12.004

- Thank you!

UNIVERSITY OF MARYLAND 18 56

CENTER FOR GEOSPATIAL INFORMATION SCIENCE