LEVERAGE CROWDSOURCED DATASETS FOR TRAVEL DEMAND ANALYSIS

Piergiorgio Vitello
Mobilab Transport Research Group, Faculty of Sciences, Technology and Communication, University of Luxembourg

SCOPE

For decades, transportation planners and researchers have relied on traditional methodologies and experience-based surveys to analyze transit demand.
[1] Problems of this approach:

- Mobility Service providers not willing to share their data
- Users misbehavior on travel surveys

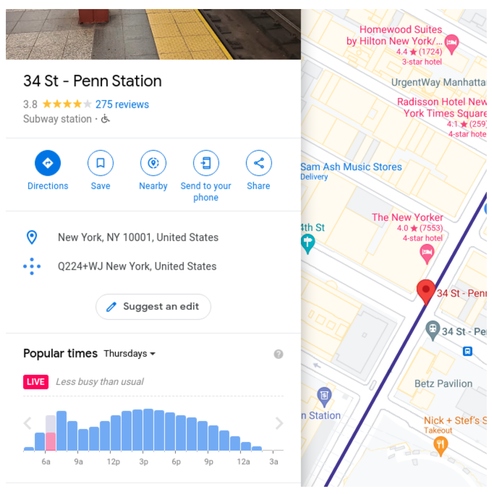
Crowdsourced data-driven approaches may provide novel solutions in this direction.[2]

PREDICTION OF LOCAL BUSINESSES ATTRACTIVENESS

We leverage crowdsensed data to enforce highly-accurate estimation of passengers demand at subway stations, this information can reach low level of aggregation not easily accessible for traditional tranist dataset.

DATASET USED

Google Popular Times(GPT)

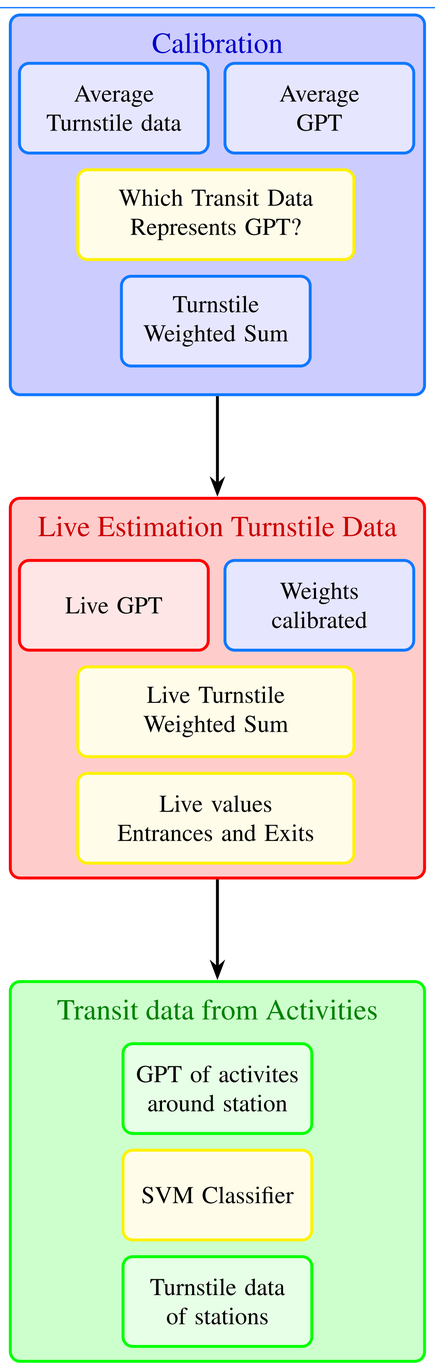


- How busy is a certain Business
- Between 0-100
- Live Value of Popularity

Turnstile Data

- Manhattan metro stations
- Entrances and Exits
- Granularity every 4 hours

METHODOLOGY



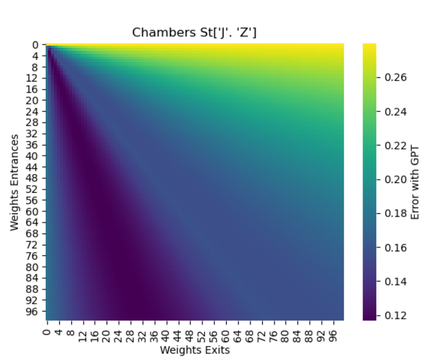
- We compare GPT of Stations and Turnstile data of previous month
- We compute a weighted sum $wsum = a \cdot \text{Entrances} + b \cdot \text{Exits}$
- We find the optimal a and b that minimize the error between $wsum$ and GPT

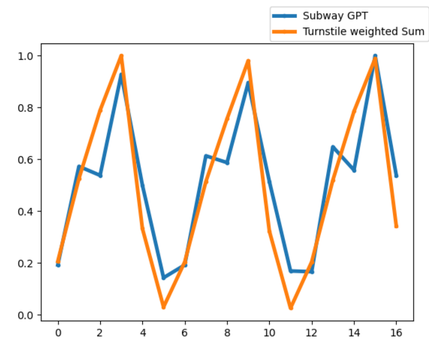
- We apply a and b calibrated before on Live GPT of the station
- We compare such data with the turnstile data of the same week
- Using the proportion computed on calibration phase, we reconstruct the real amounts of Entrances and Exits

- We create a profile of the GPT around the station (Fingerprint)
- We Use a SVM classifier with:
  - **Inputs**: Intervals of the fingerprint of the area of the station
  - **Outputs**: The transit demand at the station on the corresponding interval

RESULTS

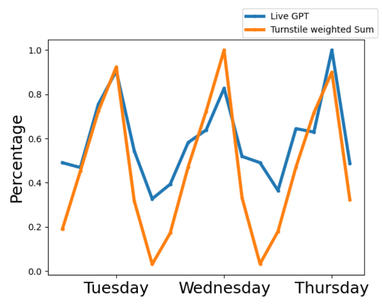
Calibration

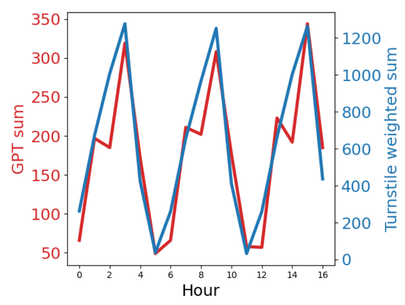




- We calibrates the weights of wsum for each station
- The majority of stations have higher weights for entrances than exits (passengers entering the station have to wait the subway)
- Wsum is a good representation of GPT of stations

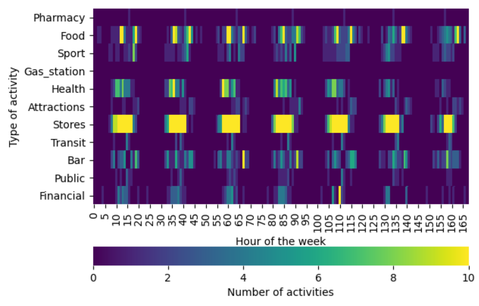
Live Estimation Turnstile data through Live GPT (Transit)





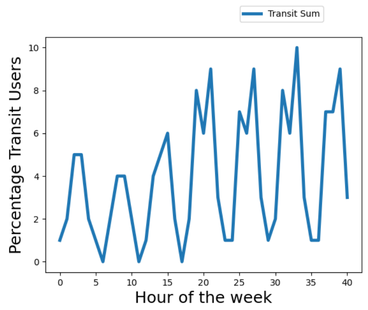
- First we try to replicate Wsum on Live GPT of a specific week
- We obtain a similar profile but with a lower accuracy than calibration phase
- Then using the live GPT we reconstruct the real number of passengers using the proportion between GPT and Turnstile computed during calibration

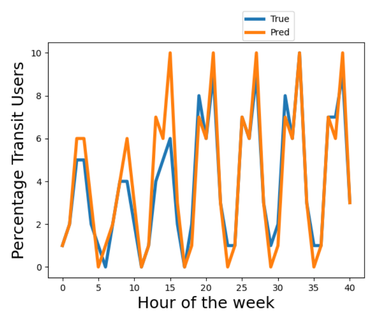
Estimation Turnstile data through GPT of Local businesses



- The heatmap represents the behavior of activities around the station
- Lighter colors mean higher popularity for a specific category
- We use slices of this heatmap as Input for a SVM Classifier

- We extract the tunrstile behavor on the interval of the heatmap
- We use slices of this data as output class for SVM





- The SVM estimate the Turnstile data with a high accuracy

ACKNOWLEDGEMENT

The Doctoral Training Unit **Data-driven computational modelling and applications** (DRIVEN) is funded by the Luxembourg National Research Fund under the PRIDE programme (PRIDE17/12252781). https://driven.uni.lu

REFERENCES

[1] Marie-Pier Pelletier, Martin Trépanier, and Catherine Morency. “Smart card data use in public transit: A literature review”. In: *Transportation Research Part C: Emerging Technologies* 19.4 (2011), pp. 557–568.

[2] Andrea Capponi et al. “Crowdsensed Data Learning-Driven Prediction of Local Businesses Attractiveness in Smart Cities”. In: *IEEE Symposium on Computers and Communications (ISCC), Barcelona, Spain, 2019*. 2019.