ConcordAI Hackathon – Itinerum
4 May 2019
8AM – 6:30PM
Concordia University
9th Floor of John Molson School of Business
1450 Guy St., Metro Guy-Concordia
Project Description

Project Description: Mode Detection from Smartphone Travel Survey Data (MTL Trajet 2017)
The purpose of this project is to develop machine learning classifier models that will use location data on respondent trips from the 2017 MTL Trajet study. MTL Trajet is an Itinerum-based smartphone travel survey app that has been used for annual smartphone studies in the fall of 2016, 2017 and 2018.

The two main data sources for the project are the open versions of this data from the Ville de Montréal Open Data portal. The data can be accessed here:

http://donnees.ville.montreal.qc.ca/dataset/mtl-trajet/resource/64278608-6bf4-4a30-83c0-2bc6e3a459e1

The two primary datasets are the “trips” (trajets) data and the “points” data.

The trips data can be downloaded here:
http://donnees.ville.montreal.qc.ca/dataset/mtl-trajet/resource/71d43835-1eef-4ae0-a0c4-578d5c8c605d?inner_span=True

Metadata for trips

Column	Definition
id_trip	Unique trip identifier
mode	Trip mode (multiple options possible). Not all trips are associated with a mode.
purpose	Trip purpose (not all trips are associated with a purpose)
starttime	Start time of the trip
endtime	End time of the trip

Possible Mode Designations (multiple answers possible)

Mode (FR)	Mode (EN)
À pied	On foot
Vélo	Bicycle
Transport collectif	Public Transportation
Voiture / Moto	Car / Motorbike
Taxi	Taxi
Autopartage	Automobile ride share

Autre	Other

The points data can be downloaded here: http://donnees.ville.montreal.qc.ca/dataset/mtl-trajet/resource/cb661499-3fa9-4793-a5fd-1ba320342fcd?inner_span=True

Metadata for points

Column	**Definition**
latitude	Latitude of location
longitude	Longitude of location
speed	Estimate (by phone) of speed
altitude	Estimate (by phone) of altitude
h_accuracy	Estimate of horizontal accuracy of point
v_accuracy	Estimate of vertical accuracy of point
timestamp	Time of when point was recorded
id_trip	An identifier of the trip to which a point is associated (see metadata above)

Proposed methodology

To get started with this project the software and libraries listed below will likely be used and would be good to have installed. It would be best to work only with single mode trips (e.g. Bicycle instead of Bicyle and Walk). It would also be best to work only with a subset of these trips; a few hundred should be fine to facilitate processing and estimation.

Mode identification for the points of a trip (the classification) needs to be obtained from the trips (trajets) data. Possible factors/variables to be used in the mode classifier models could include average travel speed, 85th percentile speed, overall trip distance, acceleration, etc. It might also be possible to include other information such as the location of metro stations to help in such models.

GIS and Database Software

QGIS (an open-source cross-platform desktop geographic information system application that supports viewing, editing, and analysis of geospatial data)

PostGIS (an open source software program that adds support for geographic objects to the PostgreSQL object-relational database)

Postgres/PostgreSQL (open-source relational database management system)

ogr2ogr (a general purpose command-line GIS tool for transforming data and loading to PostGIS)

Python Libraries

NumPy (Library for Python programming language)

scikit-learn (machine learning in Python)

- https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html (a helpful resource for finding an appropriate classifier)

Pandas (data manipulation and processing,and interacting with PostgreSQL databases)

Deep learning libraries (Tensorflow, PyTorch, Keras)

Loading GeoJSON Data to PostgreSQL/PostGIS

1. Have PostgreSQL installed with PostGIS extension enabled
    a. Create a new database for MTL Trajet data
2. Download the open MTL Trajet points and trip .geojson files (above) to a local directory
3. Load .geojson files to PostgreSQL using ogr2ogr (working directory from step 2; can take a while)

```
$ ogr2ogr -f "PostgreSQL" PG:"dbname=mtltrajet user=postgres" "./points_mtl_trajet_2017-1.geojson" -nln points
```

```
$ ogr2ogr -f "PostgreSQL" PG:"dbname=mtltrajet user=postgres" "./trajets_mtl_trajet_2017-1.geojson" -nln trips
```