Key:

Blue text == Needs revision/expansion

Title:

Spatial and temporal considerations with trip purpose inference using anonymous data.

Abstract:

The identity of *why* people travel within urban areas is something that has not received a lot of attention within broader mobility studies relative to the identity of *how* people travel. Often such studies focusing on the purpose of travel are structured around using personal information about the individual respondents such as age, gender, home and/or work locations as predictor variables. This type of volunteered geographic information can be difficult to collect in large quantities and is not widely accessible owing to sensitive information. In this study, we build a random forest classification model to predict the purposes of anonymous trips from the *2017 MTL Trajet Travel Survey*. To improve the model we combine POI information from the Google Places API and weather information from the ERA-5 climate reanalysis dataset. Our results show evidence of spatial and temporal dependencies in the models error terms [...] and we argue that a consideration of these dependencies within given cities is needed for any future classification models.

Introduction:

Movement can be thought of as an interaction between an origin and destination (Murray *et al.*, 2012). People move across space at a given time to go from where they are to where they want to be. Individuals have a particular motivation (or purpose) for doing this and so understanding these motivations at a city-wide scale underpins our comprehension of behavioural patterns within that city (Kwan & Neutens, 2012).

In recent years, improvements to GPS within smart-phones has provided researchers a new opportunity to study and record the large scale geo-spatial movement of people (Zhao *et al.*, 2019). Travel survey applications created for smart phones require much less effort from their participants than traditional travel surveys (i.e. where a separate GPS device is required to record movement) (Li et al., 2016). Therefore, it has become increasingly easy to collect qualitative information about movement within a city – including labels which detail the mode of transport used by the respondent and the purpose of their travel.

Despite this, research approaching studying *why* an urban population travels on large transport networks is less common when compared to the amount of literature into which modes of transport are used (Yazdizadeh et al.,2019). From our review of literature, one reason for this is that the key predictors for estimating travel purpose is relatively unknown.

The prediction of *how* people travel mode has been more extensively studied and can be effectively (70-80%; refs) modelled using a few variables including information such as time of day, day of week, distance and duration of trip – which, in combination, are usually sufficient to predict how a trip was taken (e.g. car, train, walk, etc.) Conversely, trips which are taken for a given activity (e.g. leisure, education, health, etc.) have been much harder to find a given set of predictor variables. And furthermore, the dynamics which control *why* people travel can vary across space and time e.g. with traffic jams, weather, annual festivals, etc. (Alsger et al., 2018).

About the data:

This report makes use of the most recent available dataset of the MTL Trajet travel survey: The *2017 MTL Trajet* travel survey project (Ville de Montréal) The *MTL Trajet* project was carried out between 18th September 2017 and 18th October 2017.

Data collection for this survey was carried out through a mobile app (available on both iOS and Android platforms) which automatically recorded a location trace using GPS provided from a user’s phone (**Figure 1.1A**; Patterson & Fitzsimmons, 2017a). When users were stopped in a given location for more than intervals of 120 seconds the app would prompt the user to ‘complete’ that trip and would be asked:

-  ‘Which travel modes did you use for this trip?’

-  ‘Why did you make this trip?’ (see similar example in **Figure 1.1B**).

Research Questions:

Can we effectively classify the purpose of trips without qualitative information from respondents?

Notes:

- [On why use anonymous data] GDPR problems with non-anonymous data

- [On problem with existing transport purpose research and VGI] This type of volunteered geographic information can be difficult to collect in large quantities is inherently tied to a specific regions at a specific time that it was recorded in.