An Exploratory Study on Mobility Behaviours Divergence across Population Segments During Post Covid-19 Epidemics Mining App-derived Location Data

Lu Xia[1], Stephen Law[1], Miguel Trigo[2]
[1]University College London, [2]Movement Strategies

Background and Motivation

New waves of public policies such as nationwide lockdown, social distance rule and work from home have been emerging since the breakout of Covid-19 epidemics, which has inevitably made a huge impact on people lifestyles and movement patterns. This study aims to investigate into people's mobility patterns under current situation from a bottom-up analytical scheme, and also reveal the behaviour divergence across different demographic groups.

The objectives of this research can be summarised as follows:
1. Allocate demographic category to users based on their historical locations and Acorn population segments.
2. Discover crowd dynamic patterns and reveal mobility behaviour divergence at a macro level.
3. Apply machine learning approach for personal attribute inferences based on Individual mobility features.

Data and Methods

This research made use of approximately 500 million mobile app-derived location records from 80,000 London residents during one-month research period from May 27th 2021 to June 28th 2021, together with Acorn demographic classification developed by CACI to better understand users' social and economic status, which segments UK-wide population into 6 categories, 18 groups and 62 types.

The pipeline of processing location data contains the following steps: data cleaning, stop detection, spatial clustering, home location detection, map matching and feature engineering. And there are three vital techniques utilised in this research, firstly we developed stay point detection criterion which provide an effective segmentation of trajectories dependent on spatial proximity and time span. Secondly, we applied a DBSCAN based spatial clustering method that extracts venues from stay points. Lastly, we chose the XGBoost algorithm (Chen et al 2016) for model training, which is a scalable end-to-end decision tree boosting system for supervised learning, with an additive regularization term in its objective function that penalizes the complexity of the model to avoid over-fitting.

Key Findings

The findings of this research could be divided into two chapters. As for the crowd dynamic part, firstly, we plotted the daily temporal rhythm of mobility activeness and found that on weekend the high level of activeness lasts from 13 pm to 16 pm (BST), and there is a slight delay in activeness climax specifically for age attribute, where younger population's most active hour is 15pm to 16pm, and senior people's most active hour puts ahead between 13pm and 14pm. And there is multi-peak pattern on weekday which are 9am-10am and 16pm-17pm (BST), indicating a decay in evening rush hour, we think this phenomenon is caused by the more flexible office mode under current situation.

Secondly, we plotted the destination heatmaps for different travel purposes, and found different population segment has special regional preferences when conducting shopping and outdoor leisure activities. Generally speaking, higher household income groups prefer visiting west London, whereas people with lower incomes tend to go to eastern regions, and people living in suburban areas are more likely to stick around their neighbourhood areas.

Thirdly, we performed text analysis of historical destinations to reveal the potential interest and consumer behaviours of certain demographic groups. For example, Category 1 (Affluent Achievers) has a fancy for outdoor leisure, such as golf club and football club, whereas Category 2 (Rising Prosperity) prefers large shopping malls in central areas, such as Selfridges and John Lewis, and pays much attention to their spiritual enrichment, as they often visit church, museum and library.

In second chapter of study, we attempted to build a machine learning classifier using the XGBoost algorithm for personal attributes inferences. We used five mobility features mined from historical trajectories including radius of gyration, number of visited venues, uncorrelated entropy, maximum travel distance and home – 'work' distance. And we found that the accuracy for age, living area and household income classification

task reaches 67.83%, 70.44%, 57.03% respectively. Our model achieved the highest accuracy for living area inference, as those live in suburban regions empirically have longer travel distances and less variety in visited venues. And according to its confusion matrix, our model is especially effective for identifying high-income groups, due to their significant lower spatial coverage in daily routes. And we believe there would be great improvement of model performance after the remission of travel restriction policies, as people's mobility behaviours still yet to fully recover under current situation and the feature engineering might not be representative enough.

Value of the research
Firstly, the heatmaps of regional preferences provide constructive insights for retail marketing and urban management targeted at various demographic groups. Secondly, locations data is a rich resource for inferring individual point of interest and consumer behaviours, which could be applied in customised advertising. Moreover, the five numeric mobility features are proven to be effective for geo-demographic profiling, making it possible for predicting sensitive personal attributes such as income without conducting high-cost surveys. Lastly, as an exploratory study, the pipelines of mining GPS location data from social and economic perspectives utilised in this research hopes to be inspirational and reproducible for follow-up studies.
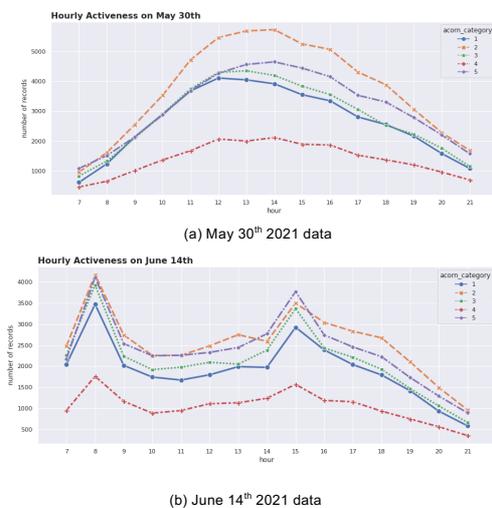
Figure 1:

(a) May 30th 2021 data

(b) June 14th 2021 data

Figure 1. Temporal Rhythm of Hourly Activeness Status between 7:00 - 22:00 (GMT)
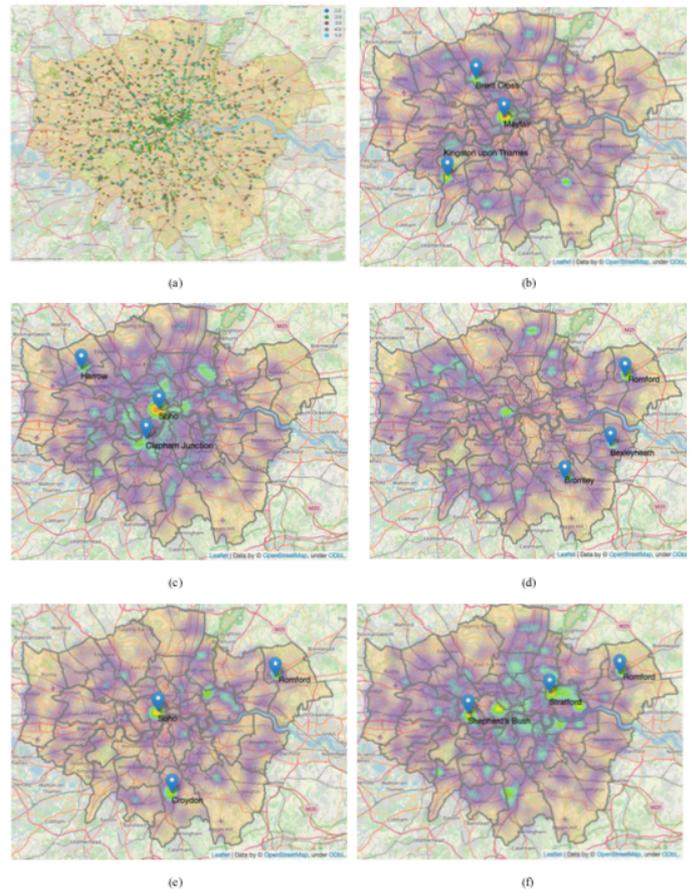
Figure 2:

Figure 2. Regional Distribution of Shopping Activities. (a): Scatterplot of all shopping activities. (b) – (f): Destination heatmaps of Acorn category 1 – 5.

Figure 3:

(a) Category 1 (Affluent Achievers)

(b) Category 2 (Rising Prosperity)

Figure 3. Word cloud of venue visitations