

## Predicting Population Flow Using Mobile Phone and Point of Interest Data in Glasgow

Jiawei Zhang<sup>1</sup>, Duncan Hay<sup>1</sup>, Franck Lenoir<sup>2</sup>, Irem Ozen<sup>2</sup>

<sup>1</sup>University College London, <sup>2</sup>Tamoco

### Background and Motivation

The mobility pattern reveals humans' travel routine. The study of mobility has great significance to the real world, including transportation networks, business sites selection, etc. The application of location-based technology generates a great amount of data with high accuracy in real-time, enabling a more accurate and comprehensive population flow prediction. This paper investigates the relationship between population flow and local activities in Glasgow city based on the Least Angle Regression model. POI (Point of Interest) and mobile phone location data are employed to construct an activity-based spatial-temporal population prediction framework. It seeks to provide some insight on estimating crowd flow based on local activities.

### Data and Methods

The anonymous mobile phone dataset provided by Tamoco is collected from January 1st to February 29th in 2020. It describes the location and time of an anonymous device user engaging with the mobile phone. The mobile phone data is spatially aggregated to the 300×300 metres cells given the privacy concern. To generate a cleaned dataset, hotspots, inactive users and inaccurate locations are filtered out. Then mobile phone data is converted into human trajectories, where stop detection and home location identification are applied to generate a cleaned trajectory dataset. Finally, the human trajectory dataset is transformed into a weighted network in order to calculate population flow by time segments and regions.

POI data and POI opening hour data are provided by Tamoco to represent local activities. This study applied a novel method in POI data cleaning, it removes POI without a chain name, in order to provide a clear picture of local activities and improve the model performance. Given the fact the different types of POI have various opening and closing hour during a week, the POI

matrix is constructed to describe the number of POIs in each category, time segment and region.

Least Angle Regression is employed to predict population flow based on local activities by assuming a linear relationship between the POI matrix and the population flow. The regression model generates a series of coefficients for each type of POI in different time segments and day-of-week, which reflects the time-varying contribution or importance of each type of activity to population flow. Mean Square Error (MSE), Rooted Mean Squared Error (RMSE), R-Square ( $R^2$ ), Mean Absolute Percentage Error (MAPE) are used to evaluate the performance of the model.

### Key Findings

The results show: City centre is the most active area compared with the surrounding areas. The distribution of city highways has an obvious influence on population flow, the regions close to city highways have larger population flow. (Figure 1)

Distinct mobility patterns are observed on weekends and weekdays morning. On weekends, most people go out after 10 am, while on weekdays, most people go out at 7 am (Figure 2). In terms of travel distance, it is found that a majority of journeys are short-distance trips that are within 500 meters, meanwhile, people tend to make long-distance journeys during the daytime (Figure 3).

The results of the regression model show that eating and drinking, as well as transport, are major activities that influence population flow. Other types of POI have less contribution to crowd flow, but still, show a regular changing pattern over time. Contribution of activity is different from weekends and weekdays, where eating and drinking and transport facility is high on weekdays, while at weekends, the contribution of transport drops significantly. (Figure 4)

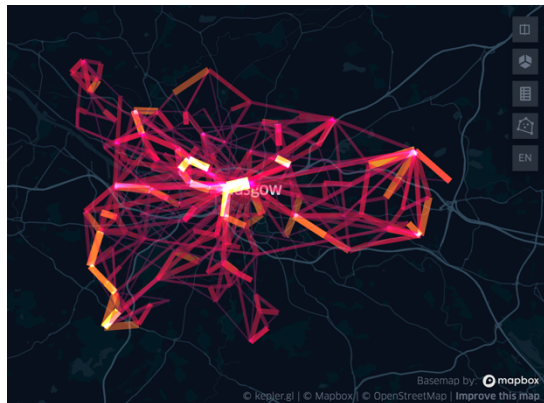
The model generates a promising prediction result using cleaned POI data generated with a novel data cleaning method. The model shows

that over 40% of the variance of population flow can be explained by local activities (Figure 5). The above results indicate that mobile phone data is able to reflect human mobility patterns, the identified mobility pattern is consistent with ground truth and can be verified by previous studies. Besides, the results suggest the effectiveness of the POI data cleaning method, which provide some sights into effectively pre-process the POI data before analysis.

### Value of the research

Overall, this study has successfully developed an activity-based spatial-temporal population prediction framework to predict population flow from local activities. Meanwhile, it confirms the effectiveness of the POI data cleaning method with regard to mobility prediction. The study provides some insights into estimating the population flow for a newly planned region.

Figure 1:



Available at: <https://jiawei997.github.io/>  
 (click PLAY ► to show the animated network)

Figure 2:

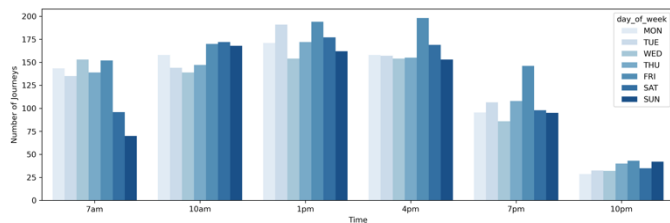


Figure 3:

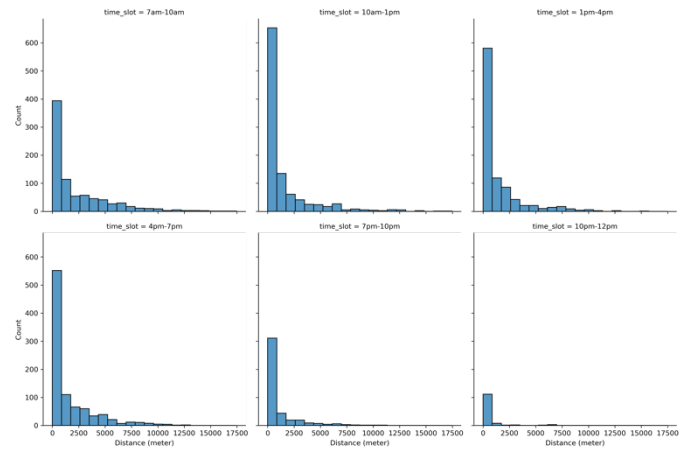


Figure 4:

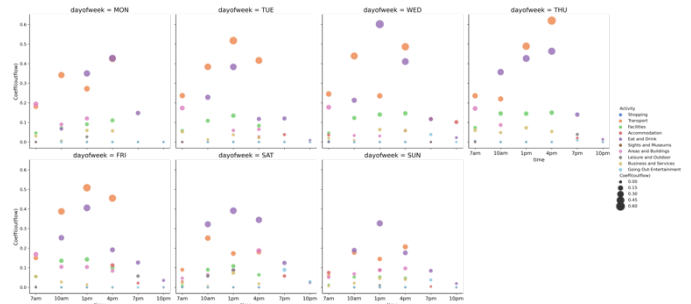


Figure 5:

Evaluation	Cleaned POI		Uncleaned POI	
	Inflow Model	Outflow Model	Inflow Model	Outflow Model
MSE	0.13	0.13	0.19	0.19
RMSE	0.36	0.36	0.44	0.43
R-square	0.45	0.41	0.11	0.21