Consumer Data Research Centre

An ESRC Data Investment

GEOLYTIX

Predicting Micro-Scale Employment Geographies with Points of Interest

Gavin Rolls[1], Chris Vardy and Ilya Ilyankou[2],
[1]University College London, [2]Geolytix

Background and Motivation.

Knowledge of urban employment geography, particularly at the local scale, is a precondition for predicting a range of phenomena, including estimated demand for nearby businesses and location planners.

This study uses POI data from the Overture Maps Foundation, amongst other open data products, to train a machine learning model that predicts employment density in London and Birmingham, UK with a particular emphasis on identifying the spatial distribution of office workers.

Data and Methods

To train a machine learning model capable of accurately predicting local employment, this study used data on over 350,000 POIs across London and Birmingham, in addition to building footprints and type classifications from OSM. Models were generated using an iterative, ensemble method which incorporated the use of Random Forest, and gradient boosting learning approaches. A 4-fold cross validation strategy was used to evaluate predictive performance.

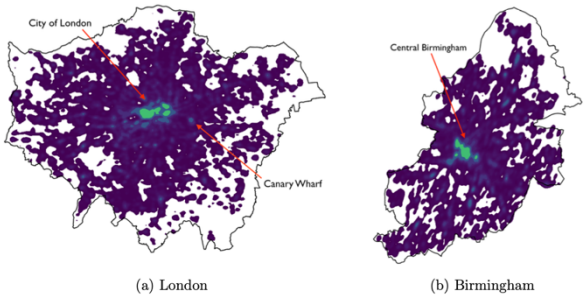

Figure 1: Kernel Density Estimation (σ = 200m) of POIs in London (left) and Birmingham (right)

Predictions were generated for each of London's ~5,000 and Birmingham's ~600, LSOAs, the most spatially granular UK census geographies. Spatially explicit models were also generated by incorporating information from spatially adjacent LSOAs into the model feature space. Overall employment and office employment predictions were made separately and validated against the UK Business Register and Employment Survey (BRES), which collects ground-truth employment data at the same spatial aggregation.

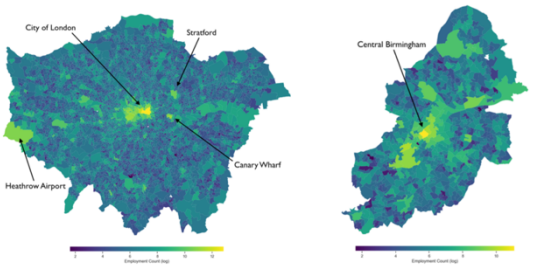

Figure 2: BRES Employment Counts per LSOA in London (left) and Birmingham (right)

Key Findings

The best performing models were able to predict overall employment density in London and Birmingham with reasonably high accuracy (R^2 = 0.76 for London, R^2 = 0.64 for Birmingham). Models trained exclusively on Overture POI data and without use of building footprints or spatially lagged features also proved effective (R^2 = 0.71 for London, R^2 = 0.62 for Birmingham). These results also show that POI category data improves predictive power beyond simple POI counts.

Models were far less accurate when predicting office-employment density. This could be due to methodological shortcomings in how employment was classified as 'in-office' or 'white-collar' but could also suggest that POIs are a less effective proxy for office density.

A detailed analysis of model residuals, specifically the construction of regression models against local demographic indicators, demonstrate that local area deprivation is a possible confounding variable in the relationship between local POI composition and corresponding employment.

Value of the Research

This research demonstrates that POIs are a strong proxy for employment and that the Overture Places dataset is a useful input for such work. It also represents a first step toward predicting specific subsets of employment, although the weak performance presented shows that further work is needed to do so accurately. Additionally, it highlights the potential role of deprivation in employment geography, an incorporation of which could further improve predictive power.