Map Data

Grebu et al 2017. Using Deep Learning and Google Street View to Estimate the Demographic Makeup of the US

Use convolutional neural networks and google street view to identify make, model and year of vehicles, and link these vehicles to the neighbourhoods in which they are seen. They then use this information and logistic regressions (for demographics) or ridge regressions (for income and voting patterns) to determine the sociodemographic characteristics and voting patterns of the neighbourhood (precinct). They get correlation coefficints in the .56-.8 level for most outcomes.

Baylis notes: Interesting that they are using a simple logistic regression analysis for demographics, whereas one would think that given that there are joint outcomes (at a minimum, multiple categories), other more sophisticated models might perform better. Second, they don’t seem to compare their prediction outcomes relative to naïve models (such as predicting voting patterns from past voting patterns. E.g. caspar Wyoming voting republican in all districts is something you don’t need machine learning methods to predict).

Dong et al 2017. Measuring Economic Activity in China with Mobile Data

Use a combination of cell phone data, GPS locations from smartphone searches, which record location, point of interest (i.e. location of search query subject), query keywords and a timestamp. They sample users who consistently registered location points at least once a month during a 13-month rolling window period to address potential biases generated by changes in online services. They calculate the work location of each user by looking for location during working hours, then use a density-based clustering algorighm to filter these data into a likely ‘point’. For consumers, they track the point of interests, and the search volume to estimate potential customers. As far as I can tell, they aren’t using ‘ground-truthing’ data. They then validate their employment and consumer results against box-office revenues and apple store revenues, and several firm-level employment case studies.

Baylis notes: interesting and clever combination of data. The ‘proof of concept’ cases are not perhaps the most overwhelming, but it’s still an impressive paper.