Prologue

Abstract

Measuring the distribution and dynamics of population at granular level both spatially and temporally is crucial for understanding the structure and function of built environment. In this era of big data, there have been numerous attempts at approaching this pursuit using the dearth of digital, unstructured, passive and incidental data which are generated from day to day human activities. In such attempts to collect, analyse and link these widely available datasets at massive scale, it is easy to put privacy of the study subjects at risk.

This research looks at one such data source - Wi-Fi probe requests generated by mobile devices, in detail and processes it into a granular, long term footfall information at retail high streets of United Kingdom (UK). Though this is not the first study to use this data source, it specifically targets and tackles the uncertainties introduced in recent years caused by implementation of features that protect the privacy of the users. This research starts with the design and implementation of multiple experiments to examine Wi-Fi probe requests in detail then later instruments a data collection methodology to collect multiple sets of probe requests at locations across London. It also uses these datasets along with the massive dataset generated by the 'Smart Street Sensor' project to devise novel data cleaning and processing methodologies which resulted in the generation of a dataset of sufficient quality which describes the volume footfall at retail high streets of UK with a granularity of 5 minute intervals since August 2015.

In parallel, this research also compiles a bespoke *'Medium data toolkit'* for processing Wi-Fi probe requests or any data with similar size and complexity. Finally the research demonstrates the value and possible applications of such footfall information through a series of case studies. By successfully avoiding the use of any personal information, this research also demonstrates that it is feasible to prioritise the privacy of the users while deriving detailed and meaningful insights from the data generated by them.

Impact Statement

We live in the age of data deluge where the data is generated at a pace that far exceeds our capacity to digest and analyse them. Putting these vast amount of data to use within the constraints of available resources and time is one of the biggest challenges faced by researchers today. The primary impact of this research is in solving this issue. This research picked one such vast data - Wi-Fi signals, generated by millions of mobile phones all around the year and available to anyone with a Wi-Fi receiver, then cleaned and processed them into highly granular and longitudinal information volume of footfall at retail high streets across UK.

The research, to convert the unstructured data into useful information, developed two novel methods - one for filtering Wi-Fi signals based on their strength and other for grouping them based on their source mobile device, without actually revealing the identity of the users. These techniques enables researchers to deal with datasets exhibiting similar challenges such as Bluetooth signals, records of people's clicking as they navigate through websites etc. These methodologies and their results have been published under a peer reviewed journal *International Journal for Geographic Information Science* for the benefit of the wider community. They were also presented back to the data partner who collaborated with the research unit - Consumer Data Research Center (CDRC), for the Smart Street Sensor project and were considered for inclusion in the data partner's commercial project.

When dealing with the large and complex Wi-Fi dataset, the research designed and implemented a bespoke toolkit and a data processing pipeline comprising of open source and free software which could be used by other researchers for use with similar datasets. The work on this 'Medium-data toolkit' was presented at the conference - *Geographic Information Science Research UK*. Moreover the research directly led to the creation and maintenance of the *aggregated footfall data* product disseminated by CDRC [2] and has served as the data source for multiple research projects within and outside CDRC and UCL.

[2] https://data.cdrc.ac.uk/dataset/local.data.company.ucl.smartstreetsensor.footfall.data.research.aggregated.data

Apart from the technical impact, the primary output of the research - footfall volumes on retail locations, has commercial and policy impact for all the stakeholders involved with the retail industry in UK. From this information, the retailers can derive insights on the patterns of customer movement around their shops; the landlords can find a reliable way to value their properties; local authorities gain a way to quantify and track the vibrancy of their retail centers over long periods of time; and the customers get information on which areas might be crowded at a given time. Finally, in the past 3 years, the outputs from this research have been disseminated to the broader academic community and industry through a series of paper presentations at conferences such as *GIS Research UK* and *Conference of Complex Systems*, talks at *Data natives*, *Geo+Data London* and *Smart Urban Policy Futures Workshop*, industry events such as *Oxford Retail Futures Conference* and public engagement events such as the *Big Data Here* exhibition.