Data Census: Charting the diverse geographies of User-Generated Content

Highlights

- A novel 'data census' of user-generated content
- Cross-platform analysis of representativeness and salience
- Multivariate comparison with socio-demographic variables

Level	PhD
First Supervisor	Dr Stefano De Sabbata
Second Supervisor	Dr Nick Tate
Application Closing Date	21st January 2019
Subject Areas	Data Analysis Geography Information Science

Overview

The past ten years have seen a momentous change in the way data is produced and used, driven by "web 2.0" and social media platforms (such as Wikipedia, Twitter and Instagram), and intertwined with the wide-spread adoption of smartphones and the open data movement. These phenomena have led to a deluge of digital data, including geospatial data, and the subsequent boom in big data analytics, data science, and machine learning (Kitchin, 2013). The use of these information sources is now commonplace, considered enabling of knowledge production (Smith, 2018) and a vital component of the recent advances in artificial intelligence.
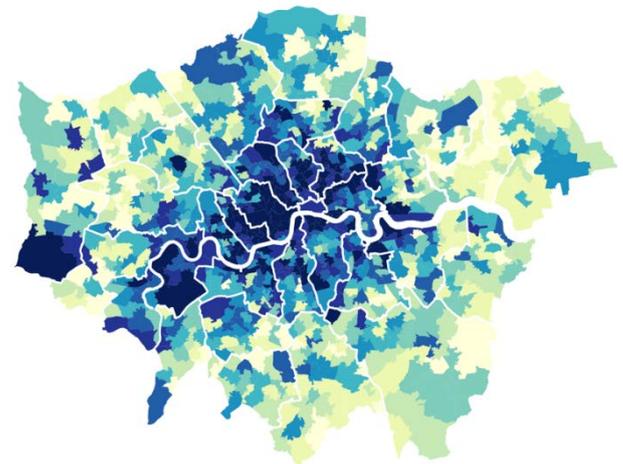
Distribution of geo-located tweets per workday population in London

However, the relationships between the geographies of user-generated content – commonly referred to as volunteered geographic information (VGI) in geographic information science – and the underlying socio-demographics at global (Graham et al., 2015), national (Bright et al., 2015) and urban (Ballatore and De Sabbata, 2018; Shelton et al., 2015) scales is complex and still largely unexplored. These studies indicate that user-generated content is highly concentrated in few locations, typically in urban areas with high socio-economic profiles, but only about half of the variation is accounted for by variation in socio-demographic variables, leaving the rest unexplained. Salient questions need to be asked about what part of our cities and regions are represented and how, and what impact these uneven geographies have on the algorithms and social science research that rely on those data.

The project aims to move forward from isolated analyses of single platforms to a contextualised, integrated understanding of the complex representations created on multiple platforms, and how user-generated content varies through time and space.

The project will:

- identify specific ontological dimensions of user-generated content (Kitchin and McArdle, 2016), including its geographic representativeness and relevance;
- conduct a cross-platform analysis of user-generated content and its relationship to socio-demographic data, including multiple platforms;
- establish the core of a novel data census, i.e. an open spatial dataset that charts the representativeness and salience of user-generated content, which will become a point of reference for analysts and developers on the advantages, biases and limitations of popular sources of big data.

Methodology

The first part of the project will encompass both a reflective analysis of the ontological dimensions to be covered, and the set-up of a series of data collectors, data scrapers, and mirror datasets covering the platforms selected for the study.

The second part of the project will start with a statistical and spatiotemporal analysis of the collected data, including cross-platform comparisons, and the analysis of its relationship with socio-demographic data. These analyses will lead to sophisticated modelling of these relationships through machine learning approaches and forecasting of future trajectories for each platform.

The third part of the project will focus on the establishment of a data census. This repository of aggregated information on each of the platforms will foster both a better understanding of the data for analyst and developers and reproducible research in the academic field.

Further Reading

1. Ballatore, A. and De Sabbata, S. (2018). Charting the geographies of crowdsourced information in Greater London. In *Geospatial Technologies for All. AGILE 2018*. Mansourian A., Pilesjö P., Harrie L., van Lammeren R. (eds).Lecture Notes in Geoinformation and Cartography. Springer, Cham.
2. Crampton, J.W., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M.W. and Zook, M. (2013). Beyond the geotag: situating 'big data'and leveraging the potential of the geoweb. *Cartography and geographic information science*, 40(2), pp.130-139.
3. Graham, M., De Sabbata, S. and Zook, M.A. (2015). Towards a study of information geographies:(im) mutable augmentations and a mapping of the geographies of information. *Geo: Geography and environment*, 2(1), pp.88-105.
4. Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in human geography*, 3(3), 262-267.

Funding

This research project is one of a number of projects in the Department. It is in competition for funding with one or more of these projects. Usually the project which receives the best applicant will be awarded the funding.

Home/EU Applicants

This project is eligible for a fully funded College of Science and Engineering studentship which includes :

- A full UK/EU fee waiver for 3.5 years
- An annual tax free stipend of £14,777 (2018/19)
- Research Training Support Grant (RTSG)

International Applicants

This project is eligible for a fully funded College of Science and Engineering studentship which includes :

- A full international fee waiver for 3.5 years
- Research Training Support Grant (RTSG)

Application Instructions

The online application and supporting documents are due by **Monday 21st January 2019**.

Any applications submitted after the deadline will not be accepted for the studentship scheme.

References should arrive no later than **Monday 28th January 2019.**

Applicants are advised to apply well in advance of the deadline, so that we can let you know if anything is missing from your application.

Required Materials

1. Online application form
2. Two academic references
3. Transcripts
4. Degree certificate/s (if awarded)
5. Curriculum Vitae
6. CSE Studentship Form
7. English language qualification

Applications which are not complete by the deadline will not be considered for the studentship scheme. It is the responsibility of the applicant to ensure the application form and documents are received by the relevant deadlines.

All applications must be submitted online, along with the supporting documents as per the instructions on the website.

Please ensure that all email addresses, for yourself and your referees, are correct on the application form.

For more information, please visit our website at :
https://www2.le.ac.uk/colleges/scieng/research/postgraduate-opportunities/cse-2019/instructions