Monitoring Neighbourhood Health

Christopher Smith-Clarke
ICRI: Cities
London, UK
chris.smith@ucl.ac.uk

Licia Capra
University College London
London, UK
l.capra@ucl.ac.uk

ABSTRACT

Recent research has demonstrated that data collected from ubiquitous sources can be exploited to estimate socioeconomic factors. Here we discuss the kind of novel application that can be built using this new understanding and the challenges presented.

Author Keywords

HCI; urban computing; well being; data mining

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

General Terms

Human Factors; Design; Measurement.

INTRODUCTION

The planet is undergoing a rapid population shift towards urban environments, with an estimated growth of 5 million new city dwellers each month in developing countries [13]. With rapid growth comes an ever increasing need for effective planning and management of urban infrastructure. To efficiently allocate limited resources, policymakers and agencies first need to identify which areas are in most need of intervention in order to alleviate deprivation, where deprivation is often a multifaceted concept such as the English Index of Multiple Deprivation (IMD), which takes into account factors such as income, education and crime [8]. Currently, measures of socioeconomic factors are derived from survey data, or summary statistics of data such as that pertaining to social benefit claims and health, for example. A well known shortcoming of this approach is that measures can quickly become out of date. For example, the latest version of the IMD was published in 2010 yet pertains to data mostly from 2008 and even partly from as early as 2001. Indeed, in many cases, full censuses are undertaken only every ten years, and the larger the assessment window, the more likely that problem areas will deteriorate. Thus, developing new methods of identifying urban deprivation swiftly, continuously and at low cost would offer significant social and economic benefits.

CHI 2013 Workshop GeoHCI, April 2728, 2013, Paris, France.

The advent of large and rich datasets describing peoples mobility and communication patterns, such as from telephone call detail records (CDR), public transport usage and social media, has instigated a plethora of research analysing human behaviour and the social networks reflected in these data. Much of this work aims to discover features of human behaviour and relationships which relate to, and can thus be used to identify, socioeconomic deprivation. Examples of work in this vein include that by Kramer, who found that the difference between the number of positive and negative words used in Facebook status updates covaries with self-reported satisfaction with life [6] and Quercia, *et al.*, who found that features of geolocated tweets [11] in London correlate with IMD scores of neighbourhoods. The drawback to relying on social media for determining community well being is of course the bias towards certain demographic groups among the user base [9]. CDR offer a more robust, but still somewhat biased, source of data from which to derive features of communication and mobility patterns, such as in the work of Frias-Martinez *et al.* who have investigated in detail the use of CDR for producing estimated census maps [3].

A key element of the urbanisation process is the development of public transport systems, many of which are adopting automated fare collection systems. These systems record the travel history of passengers, making available a fine grained and much more representative dataset depicting the movement of people about the city. From this data we derived a number of features with which we trained a classifier which could identify areas of high deprivation. For example, by modelling urban flow as gravity, we were able to quantify flow restriction which provided a clue to the socio-economic health of neighbourhoods [12]. The results of this research are promising and suggest that before long we will be able to provide up to date and continuous estimates of socioeconomic factors, and indeed, provide forecasts based on the data models. It is therefore important to consider what will be the nature of the applications developed using these kind of data models, who will be the users of these applications and the ways in which users will be able to interact with the data. Here we focus on a particular vision of a tool designed for urban policymakers.

A MONITORING AND DECISION SUPPORT SYSTEM

For a reliable system to be developed which is able to highlight areas likely to be suffering high deprivation, policymakers would be concerned with having the most up to date estimate of an area's socioeconomic level, as well being able assess the effects of intervention and regeneration in different areas. Creating a system that meets these requirements would present a number of challenges. There is firstly the data mining challenge of discovering features in datasets and determining the best performing models in terms of predictive accuracy. Additionally, in order to be able to explore the effects of intervention, it needs to be understood how different areas affect one another. That is, how do the effects of intervention (or indeed degradation) diffuse through different interaction networks, for example through transport and communication networks, or through simple adjacency (spatial auto-correlation) or land use similarity.

Furthermore, we must consider what is the best way to represent these connections to the user. The second challenge, or group of challenges, thus relates to the HCI aspect of the system. A number of sources of uncertainty exist that need to be appropriately conveyed to the user so that decisions are not made under false pretences. Within the Geography and Geographical Information Science literature there is a large amount of work tackling the problem of how to communicate the level of uncertainty in spatial data [4]. An example of relevant work is that of Kardos *et al.*, who identify three kinds of uncertainty in the spatial visualisation of census data: temporal, spatial and attribute uncertainty [5]. *Temporal* uncertainty arises from data becoming out of date and thus is largely avoided by the system under consideration. *Spatial* uncertainty of census data is a consequence of aggregation of individuals' data to the area level. The true distribution of an attribute within an area is unknown and the choice of area boundaries can also affect the attribute values (Modifiable Unit Area Problem [10])). In our case spatial uncertainty is introduced when defining how deprivation levels estimated from interaction between nodes, (e.g., transit stations, telecoms antenna) relate to the surrounding area. For example, it is not clear if the estimated value at a node should be assigned uniformly to the area in which it is located (in which case the choice of boundary will effect the results, or to a circular area around the node, or a population weighted overlay of administrative areas and tessellation of the nodes. Finally, *attribute* uncertainty in census mapping refers to the error due to sampling or methodology in the census data acquisition, but in the case of estimated values, the attribute uncertainty derives from the error in the statistical models themselves.

A possible way to decrease uncertainty would be to overlay data models derived from the different kinds of dataset mentioned above (social media, communication and mobility networks), thereby increasing coverage and accuracy. This would however increase complexity, thereby cementing the 'black box' nature of the system. The problem of complexity in dealing with spatial data has led to the development of Spatial Decision Support Systems (SDSS). These are GIS based software tools that combine spatial data and algorithmic intelligence, and which are designed to lubricate the spatial planning and problem solving process by enabling the exploration of various alternate strategies through modelling, simulation and prediction [1]. Examples of common use cases for SDSS are land use planning [7] and urban infrastructure planning [2]; problems which involve several conflicting criteria to be satisfied. The kind of system we envisage could form part of an SDSS that incorporates socioeconomic factors as new criteria to be met.

In this position paper we have outlined a vision, and the challenges in realising this vision, of a system which, by building on the data mining methods developed in previous work, can provide continuously updated estimates of socioeconomic factors, identify areas most in need of intervention and forecast the effects of such intervention.

REFERENCES

1. M. Batty and P. J. Densham. Decision support, gis, and urban planning, Jan. 1996.
2. J. a. Coutinho-Rodrigues, A. Simão, and C. H. Antunes. A GIS-based multicriteria spatial decision support system for planning urban infrastructures. *Decision Support Systems*, 51(3):720–726, 2011.
3. V. Frias-martinez, V. Soto, J. Virseda, and E. Frias-martinez. Computing Cost-Effective Census Maps From Cell Phone Traces. In *Proceedings of the 2nd Workshop on Pervasive Urban Applications @ Pervasive'12*, 2012.
4. G. J. HUNTER and M. F. GOODCHILD. Communicating uncertainty in spatial databases. *Transactions in GIS*, 1(1):13–24, 1996.
5. J. Kardos, G. Benwell, and A. Moore. The Visualisation of Uncertainty for Spatially Referenced Census Data Using Hierarchical Tessellations. *Transactions in GIS*, 9(1):19–34, Jan. 2005.
6. A. D. I. Kramer. An Unobtrusive Behavioral Model of Gross National Happiness. In *Proceedings of the 28th ACM CHI*, pages 287–290, 2010.
7. K. B. Matthews, A. R. Sibbald, and S. Craw. Implementation of a spatial decision support system for rural land use planning. *Computers and Electronics in Agriculture*, 23(1):9–26, 1999.
8. D. Mclennan, H. Barnes, M. Noble, J. Davies, and E. Garratt. The English Indices of Deprivation 2010. Technical report, 2011.
9. A. Mislove, S. Lehmann, Y. Ahn, and J. Onnela. Understanding the Demographics of Twitter Users. *Fifth International AAAI*, pages 554–557, 2011.
10. S. Openshaw. Ecological fallacies and the analysis of areal census data. *Environment & planning A*, 16(1):17–31, Jan. 1984.
11. D. Quercia, J. Ellis, L. Capra, and J. Crowcroft. Tracking Gross Community Happiness from Tweets. In *Proceedings of ACM CSCW 2012*, 2012.
12. C. Smith, D. Quercia, and L. Capra. Finger On The Pulse: Identifying Deprivation Using Transit Flow Analysis. *Proceedings of ACM CSCW 2013*, 2012.
13. United Nations Human Settlement Program. *State of the World'd Cities 2008/2009*. EarthScan, 2008.