Computers, Environment and Urban Systems 47 (2014) 5–15

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/compenvurbsys

Studying commuting behaviours using collaborative visual analytics

Roger Beecham [a,*], Jo Wood [a], Audrey Bowerman [b]

[a] giCentre, Information Sciences, City University London, United Kingdom
[b] Delivery Planning - Cycling, Transport for London, United Kingdom

ARTICLE INFO

Article history:
Available online 12 November 2013

Keywords:
Collaborative visual analytics
Bicycle share schemes
Commuting behaviour

ABSTRACT

Mining a large origin–destination dataset of journeys made through London's Cycle Hire Scheme (LCHS), we develop a technique for automatically classifying commuting behaviour that involves a spatial analysis of cyclists' journeys. We identify a subset of potential commuting cyclists, and for each individual define a plausible geographic area representing their workplace. All peak-time journeys terminating within the vicinity of this derived workplace in the morning, and originating from this derived workplace in the evening, we label commutes. Three techniques for creating these workplace areas are compared using visual analytics: a *weighted mean-centres calculation, spatial k-means clustering* and a *kernel density-estimation method*. Evaluating these techniques at the individual cyclist level, we find that commuters' peak-time journeys are more spatially diverse than might be expected, and that for a significant portion of commuters there appears to be more than one plausible spatial workplace area. Evaluating the three techniques visually, we select the *density-estimation* as our preferred method. Two distinct types of commuting activity are identified: those taken by LCHS customers living outside of London, who make highly regular commuting journeys at London's major rail hubs; and more varied commuting behaviours by those living very close to a bike-share docking station. We find evidence of many interpeak journeys around London's universities apparently being taken as part of cyclists' working day. Imbalances in the number of morning commutes to, and evening commutes from, derived workplaces are also found, which might relate to local availability of bikes. Significant decisions around our workplace analysis, and particularly these broader insights into commuting behaviours, are discovered through exploring this analysis visually. The visual analysis approach described in the paper is effective in enabling a research team with varying levels of analysis experience to participate in this research. We suggest that such an approach is of relevance to many applied research contexts.

1. Introduction

Since its introduction in July 2010 over 20 million journeys have been made through the London Cycle Hire Scheme (LCHS). Recent analyses of LCHS usage data have found daily tidal flows of bikes into and out of central London, which coincide with commuting peaks (Lathia, Ahmed, & Capra, 2012; Wood, Slingsby, & Dykes, 2011). These flows disproportionately redistribute bikes to particular parts of the city, making many docking stations unusable – either rendered entirely full or empty of bikes. This is a problem common to most urban bike share schemes (OBIS, 2011). To keep the system as balanced as possible, bikes are manually transported across the city at peak times, and in priority areas docking stations are continually replenished with bikes or bikes continually removed from docking stations. Since such load rebalancing is expensive, Transport for London (TfL), the organisation responsible for the scheme's operation, wish to better understand commuting LCHS users and their journeys.

Working with a diverse team of colleagues at TfL, three questions motivate this research:

1. What are the characteristics of people who take part in commuting based activities?
2. Where do commuting events happen?
3. Under what circumstances are journeys made during the working day?

Before these three questions can be investigated, there is a broader question:

4. How can commuting journeys and commuting LCHS cyclists be reasonably detected?

The task of identifying commuting behaviour might initially seem like a straightforward data mining exercise. For example,

* Corresponding author. Address: giCentre, Information Sciences, City University London, London EC1V0HB, United Kingdom. Tel.: +44 020 7040 3914.
E-mail addresses: roger.beecham.1@city.ac.uk (R. Beecham), j.d.wood@city.ac.uk (J. Wood), audreybowerman@tfl.gov.uk (A. Bowerman).

0198-9715/$ - see front matter
http://dx.doi.org/10.1016/j.compenvurbsys.2013.10.007

one means of identifying commuting journeys might be to find all instances where a LCHS cyclist completes a closed peak-time 'loop', where their last journey of the day happens during the evening commute and is the inverse of their first journey of the day. Recent analysis of usage data from London's underground system has found such assumptions about commuting behaviour often do not hold (Lathia, Smith, Froehlich, & Capra, 2013). This might be especially true of LCHS behaviour. Usage of the scheme is perhaps more ad hoc and subject to a wider set of environmental and other variables than use of a large metro system. Moreover, whilst London underground users can generally *expect* access to a train at their most convenient station, competition for bikes during peak times means that LCHS cyclists may have more modest expectations: individuals may not be able to consistently collect or return bikes at their preferred station, and therefore LCHS commuting may not consist only of journeys between a single pair of stations.

In this paper we investigate approaches to identifying commuting journeys that involve spatial analysis of individual cyclists' peak-time journeys. Our general approach is to find a broad spatial area, or set of spatial areas, representing each commuting LCHS user's workplace, and identify all journeys that end (in the morning) or start (in the evening) from this workplace area. This is our first contribution:

Contribution 1. A new technique for deriving customers' workplace areas and labelling commuting journeys, based on a spatial analysis of travel behaviours.

We believe this technique is novel to the extent that, unlike similar studies by Lathia et al. (2013) and Agard, Morency, & Trepanier (2006) that identify individuals whose dominant temporal usage coincides with peak-times, it relies on a *spatial* as well as temporal evaluation of travel behaviours. Our technique might reasonably be applied to other large-scale bike share schemes.

The second contribution relates to our approach. We use visual analytics to evaluate various workplace identification techniques. This visual approach allows relatively abstract data transformations to be made intelligible to both data analysis specialists (ourselves) and domain experts (colleagues at TfL). The paper describes a process of *chauffeuring*, whereby colleagues at TfL articulate a research problem, we propose a set of solutions and, using tailored visual analytics, we collectively explore and evaluate this solution space. We believe that such an approach is particularly suited to research contexts where decisions are required from a diverse team containing analytic, policy-related and operational specialisms. Such a requirement may be common to many applied analysis contexts.

Contribution 2. A visual analytics approach that facilitates a data-driven discourse between a diverse set of individuals.

Finally, we describe several new insights into commuting LCHS behaviour that were generated through this analysis, and that were not previously known to colleagues at TfL. These may be relevant to others with an interest in studying usage of bike-share schemes.

Contribution 3. A set of empirically-generated findings relevant to those interested in studying bike-share cycling behaviour.

2. Related work

2.1. Inferring commuting behaviour from origin–destination datasets

Automatically collected usage datasets from shared transport systems have only recently become available to researchers, and the literature on approaches to inferring commuting behaviour from such data is not widespread. One relevant study was conducted by Lathia et al. (2013). With a month's usage data from the London underground, the authors identify the nature and extent of commuting behaviour using various data mining algorithms. First, they make reasonable assumptions about commuting travel behaviours: that commuters will make on average two journeys or more per day; that they will typically repeat the same origin–destination (OD) pair; and that commuters will have a closed loop whereby the first origin and last destination of the day should be the same. Lathia et al. (2013) subsequently find that many travellers do not fit these expected patterns. Sixty-six percent of users take less than one trip every two days and only 8% meet the expected two trips per weekday criteria. Half of all trips taken by users in the month-long study period are entirely unique OD pairs for those people. However, 50% of all users form a closed loop on Monday–Thursdays, 44% on Fridays and 37% on weekends (Lathia et al., 2013). The authors later propose clustering algorithms for automatically finding groups of travellers with similar temporal travel profiles – who typically travel at particular times of day and days of the week. This approach is also taken by Agard et al. (2006) when analysing bus usage data in Quebec. The authors find a large group of bus passengers whose usage almost exclusively coincides with peak commuting times.

Since we are interested in studying whether or not LCHS members make journeys within their working day, an important aspect of this study is to identify with a degree of certainty all journey *events* that we think might be commutes. Whilst the unsupervised clustering algorithms proposed by Agard et al. (2006) and Lathia et al. (2013) would enable those who apparently use the scheme almost exclusively for commuting to be distinguished from those with more varied usage characteristics, it would not enable a total, journey-level view of commuting. It is reasonable to assume that those who commute may also often use the LCHS for non-commuting, leisure-oriented or utilitarian weekend journeys. If commuting users were only defined as people whose dominant travel patterns coincide with commuting times, then we potentially miss the commuting behaviour of individuals who typically use the scheme for other purposes. In addition, that usage of the LCHS is likely to be more ad hoc than, for instance, the London underground, assumptions around commuting activity made by Lathia et al. (2013) might be even more problematic when applied to the LCHS dataset.

An alternative approach, taken in this study, is a spatial analysis of LCHS users' peak-time journeys. We attempt to identify a broad spatial area representing each cyclist's workplace and label all peak-time journeys arriving at this workplace area in the morning and departing from this workplace area in the evening as commutes. Clearly we have no *a priori* knowledge of such workplace areas and instead we derive them from exploring spatial patterns of LCHS cyclists' peak-time journeys.

2.2. Collaborative visual analytics and chauffeuring

For Tukey & Wilk (1966), the aim of any data analysis is to find insights that can be easily stated and are intelligible to the individuals conducting an analysis: '*at all stages of data analysis the nature and detail of output* [...] *need*[s] *to be matched to the capabilities of the people who use it and want it*' (Tukey, 1966:697). Visual analytics refers to the application of tools and techniques for synthesising information and discovering insights from generally large and complex datasets (Thomas, 2005). It can play an important role in making research outputs interpretable to individuals with a range of specialisms. For instance, Robinson (2008) notes that the use of interactive visual interfaces enables teams of analysts with different skill sets to collectively engage