Targeted Out-door Advertising Using Geo-Social Media Data

Juntao Lai*, Tao Cheng†

SpaceTimeLab for Big Data Analytics, Department of Civil, Environmental &Geomatic Engineering, University College London

Jan 11, 2017

Summary

Exploring new techniques of targeted advertising across public spaces has attracted considerable attention. This research demonstrates a novel way of identifying suitable location and time for targeted advertising using geo-social media, by using London Underground stations as case study. The number of target users are estimated by integrating the passenger counts data and topic of interests from geo-social media data, which is more effective and efficient than traditional methods.

KEYWORDS: social media data, spatial-temporal analysis, targeted advertising

1. Introduction

Exploring new techniques to improve the efficiency of targeting the selection of advertisements across public spaces has attracted considerable attention (Cronin, 2008). Efficient marketing is to deliver information that is relevant to consumers in the present situation, based on context such as location, time and social environments (Lasinger & Bauer, 2013). Digital out-door advertisements such as digital billboards have unlocked the capacity for advertisement companies to allocate their adverts in real-time, corresponding to consumers' interests. The position of out-door advertisement billboards are usually fixed. However, the population of passers-by, which are the potential audiences of their adverts, are dynamic. Whilst it is possible to ascribe local interests from surveys, these are costly, of low temporal resolution and have undoubtable sampling issues. Thanks to its spatial, temporal and semantic information, geo-social media data has been widely explored as a means of understanding public interests across an urban environment (Cranshaw et al., 2012; Hasan et al., 2013; Jenkins et al., 2016).

With as many as 4 million passenger journeys handled by London Underground system every day (Transport for London, 2015), the advertisements across the stations accumulate a considerable revenue. Advertising is already provided through digital screens in many of the more popular stations so practitioners already have the capacity to distribute their adverts across time and space. However, the planning of specific advertisements across time and space is a difficult process, as little is known about the passengers. Our previous research has explained how to harvest the topic of interests around London Underground stations from geotagged Tweets (Lai et al. 2015). Based on that, we are going to introduce an efficient and flexible way of identifying suitable location and time for advertising in different topics.

2. Identifying target stations for specific topic

From the advertising perspective, we would like to identify the place that has the biggest proportion as well as number of target user. If one place has a large number of people visiting, but who are not relevant to the particular product, it should not be considered as an ideal place for advertising as it will increase the budget but may not be as effective as we would expect. Therefore, we need to identify the number of passengers passing by Underground stations that are interested in specific advertising topics (e.g. Sports, Music & shows or Fashion & shopping).

* Juntao.lai.13@ucl.ac.uk

† Tao.cheng@ucl.ac.uk

We first identify interest of passengers based upon geo-tagged Twitter data around stations (Lai et al., 2015). The data were acquired from Twitter Streaming Application Programming Interface (API) service (https://dev.Twitter.com/). Latent Dirichlet Allocation (Blei et al, 2003) was used to extract topics from the corpus built by the words of the Tweets after text cleaning. After topic modelling, each Tweet had a distribution of probabilities of being assigned to 20 topics. Some of these generated topics did not have any actual meaning, for instance, the topics consisting of swear words or non-English words. Those topics could not help to explain the interests of the users, and hence, eight topics out of the 20 were removed. To ease the representation of the topics, each topic was manually labelled according to the top words. The Tweets were separated into eight time groups according to the associated time stamps (four 6-hour time periods at the weekend and on weekdays).

Then we estimate the number of target users of each topic. We used a sample of London Underground stations passenger counts data from October 11th to 24th in 2015, which is publically available from the TfL website (http://tap.data.tfl.gov.uk/). We aggregated the passenger counts to the same 6-hour time period groups, but did not include the time periods of 00:00-06:00 on both weekdays and weekends since most stations are closed during the night. With the number of passengers and topic probability distribution at each station in each time period, we can roughly estimate the number of our target group for each topic with the multiplication of both. It should be made aware that, this is based on an assumption that the topic distribution can perfectly reflect the composition of interests of the passenger group. However, this estimated number is supposed to reflect the relative differences between different stations, especially when two stations have similar passenger number but very different topic distributions.

Here we present two topics as examples (Sports and Music & shows), which are displayed in figure 1. In these plots, each dot represents a station. The x value indicates the total number of target users of that station in one week, while the y value indicates the proportion of the topic over all the Tweets around the station. The perpendicular lines indicate the criteria, and dots located in the first quadrant are the ideal stations for advertising in that topic, which are highlighted in blue. The selecting criteria can be flexible and subjective, we chose x=50000 and y=0.1 for both topics. Stations that only meet either criteria are not considered as an ideal place for advertising. For example, King's Cross St Pancras is a major transport hub in central London, hence a high flow of people. However, only a little proportion of them are interested in these two topics, advertising in this station is not efficient but costs a lot. While stations, such as Goldhawk Road, has a high percentage but not a big number of people interested in Music, which should not be considered as well. In general, the stations identified are corresponding to our knowledge about the city in terms of each topic. For example, the stations that are popular in topic "Music" are close to the well-known music avenues, and stations in topic "Sports" are near football clubs or sport stadiums.

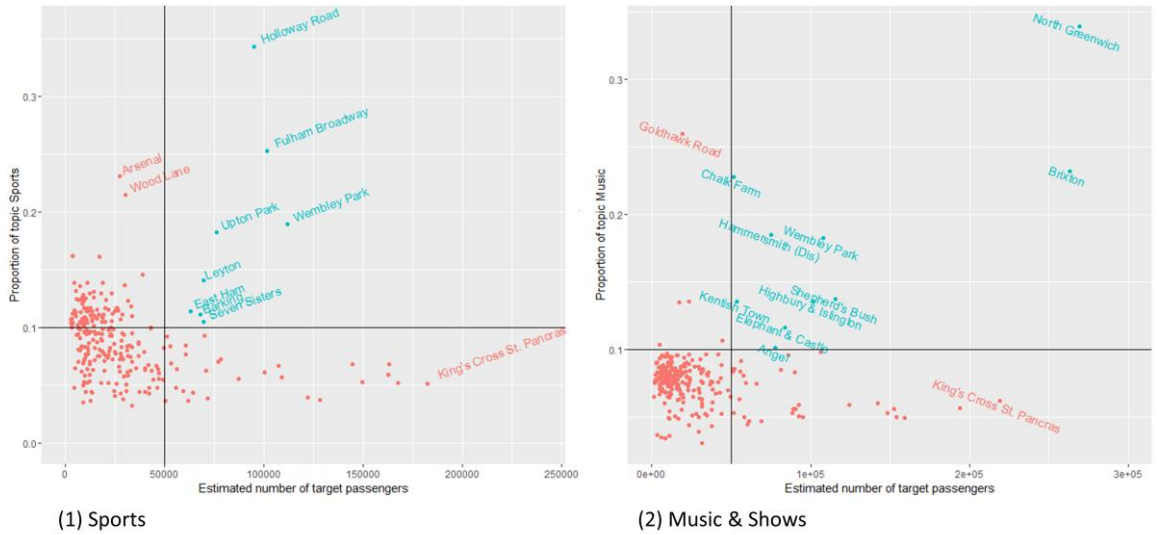
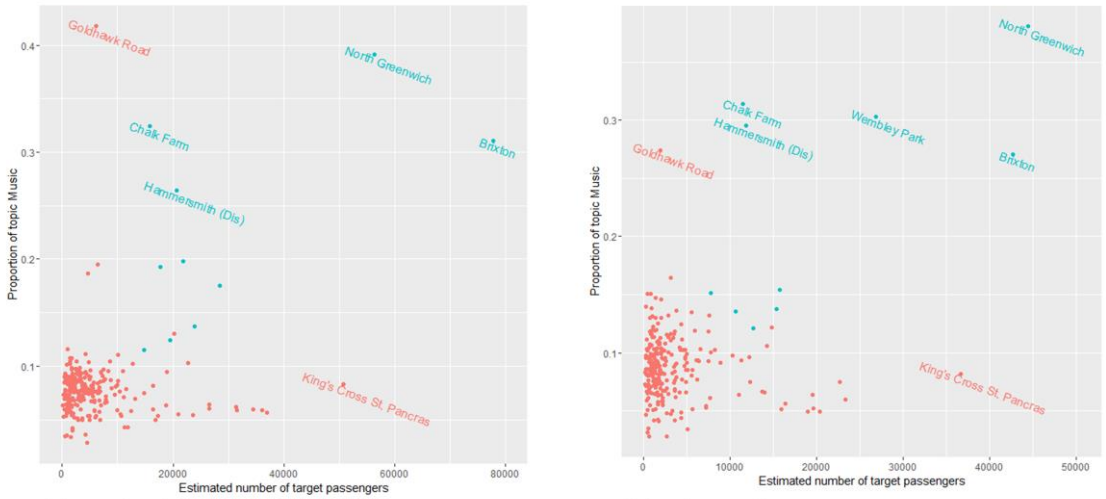


Figure 1: Identifying stations for advertising in different topics.

The results shown in figure 1 are simply based on data for the whole time period. However, more detail insights of the targeted stations can gained by analysing the results in different time periods. Figure 2 displays two plots of the popularity of topic "Music" in different time periods, nights on weekday and weekend. As shown on the map, one station could have different popularity of the same topic as time changes. For example, Wembley Park is one of the stations with a big number of target passengers on weekend evening. However, there are not that many people visiting Wembley Park who are interested in "Music" on weekdays. As contrast, stations such as North Greenwich and Brixton are always popular for "Music".



(1) Monday-Thursday 18:00-24:00 (2) Friday-Sunday 18:00-24:00

Figure 2: Identifying stations for advertising in topic "Music" at different time periods.

3. Conclusion

This research presents a novel approach of estimating potential users of specific topic in marketing using Geo-social media data, taking London Underground stations as our case study. Although there are many assumptions and limits in the research, such as the reliability and representativeness of using Twitter user to represent tube users, and whether the topic discussed on geotagged Tweets really reflect people's interests in that situation. The approach proposed in this research shows its potential of extracting target stations of different topics in a purely data-driven way, and the results match our understanding of these places. Moreover, it offers opportunity of gaining insights to the variation of topics in different time periods in a statistical perspective, which are beyond our subjective and static understanding of the places. Therefore, this research is useful in targeted advertising, by making use of spatial, temporal and semantical information in the urban environment.

4. Biography

Juntao Lai is a PhD student in department of Civil, Environmental and Geomatic Engineering at University College London. His research interest includes semantic analysis of social media data and spatial-temporal analytics.

Tao Cheng is a Professor in GeoInformatics, and Director of SpaceTimeLab for Big Data Analytics (http://www.ucl.ac.uk/spacetimelab), at University College London. Her research interests span network complexity, Geocomputation, integrated spatio-temporal analytics and big data mining (modelling, prediction, clustering, visualisation and simulation), with applications in transport, crime, health, social media, and environmental monitoring.

References

Blei DM, Ng AY, Jordan MI (2003). Latent dirichlet allocation." *the Journal of Machine Learning Research* 3: 993-1022.

Cranshaw J, Hong JI, Sadeh N (2012). The Livehoods Project : Utilizing Social Media to Understand the Dynamics of a City. Icwsm, pp.58–65.

Cronin AM (2008). Mobility and Market Research: Outdoor Advertising and the Commercial Ontology of the City. Mobilities, 3(1), pp.95–115.

Hasan S, Zhan X, Ukkusuri SV (2013). Understanding Urban Human Activity and Mobility Patterns Using Large-scale Location-based Data from Online Social Media. *In Proceedings of the 2Nd ACM SIGKDD International Workshop on Urban Computing.* p. 6:1--6:8.

Lai J, Cheng T, Lansley G (2015). Spatio-Temporal Patterns of Passengers ' Interests at London Tube Stations. *In Proceedings of the 23rd GIS Research UK (GISRUK) conference*. Available: http://leeds.gisruk.org/abstracts/GISRUK2015_submission_26.pdf

Lasinger P and Bauer C (2013). Situationalization, the New Road to Adaptive Digital-out-of-Home Advertising. *Proceedings of IADIS International Conference e-Society*, (October), pp.162–169.

Jenkins A, Croitoru A, Crooks AT, Stefanidis A (2016). Crowdsourcing a Collective Sense of Place. *PloS one*. 11(4):e0152932.

Transport for London. London Underground [Internet]. (2015) [cited 29 March 2016]. Available: https://tfl.gov.uk/corporate/about-tfl/what-we-do/london-underground