Large-scale indoor movement analysis: the data, context and analytical challenges

Tomko, Martin ; Ren, Yongli ; Ong, Kevin ; Salim, Flora ; Sanderson, Mark

Abstract: The analysis of movement data can reveal rich information about animal or human behaviour in different environments. Human population is increasingly urbanised1 and in cities, people spend large portions of their time indoors. Human indoor movement patterns and indoor environments display characteristics different to the movement of other animals and natural environments, respectively, hence the need for tailored analytical approaches and methods. The specific characteristics of the movement are further emphasized by differences in data collection techniques in indoor environments, and consequently the data collected. In this paper, we introduce some of the major challenges in real-life indoor movement data collection impacting on the analysis of patterns mined from such data. Our examples are drawn from a large dataset of Wi-Fi-based tracking of visitors to large retail spaces (shopping malls) in two major Australian cities. The overall focus of our project (currently in its early stages) is the mining of contextualised indoor behavioural patterns that would enable improved product recommendation (Kantor et al. 2011) to visitors.

Large-Scale Indoor Movement Analysis: The data, context and analytical challenges

M. Tomko[1], Y. Ren[2], K. Ong[2], F. Salim[2], M. Sanderson[2]

[1]Department of Geography, University of Zurich, Switzerland
Email: martin.tomko@geo.uzh.ch

[2]School of Computer Science and Information Technology, RMIT University, Melbourne, Australia
Email: yongli.ren|kevin.ong|flora.salim|mark.sanderson@rmit.edu.au

1. Introduction

The analysis of movement data can reveal rich information about animal or human behaviour in different environments. Human population is increasingly urbanised[1] and in cities, people spend large portions of their time indoors. Human indoor movement patterns and indoor environments display characteristics different to the movement of other animals and natural environments, respectively, hence the need for tailored analytical approaches and methods. The specific characteristics of the movement are further emphasized by differences in data collection techniques in indoor environments, and consequently the data collected.

In this paper, we introduce some of the major challenges in real-life indoor movement data collection impacting on the analysis of patterns mined from such data. Our examples are drawn from a large dataset of Wi-Fi-based tracking of visitors to large retail spaces (shopping malls) in two major Australian cities. The overall focus of our project (currently in its early stages) is the mining of contextualised indoor behavioural patterns that would enable improved product recommendation (Kantor et al. 2011) to visitors.

2. Environment and Data

In collaboration with an industry partner – the operator of a network of large shopping malls – we have been granted access to large datasets covering two shopping malls. We believe that this is currently one of the largest datasets of this kind available to academics worldwide. The two shopping malls have different characteristics: an inner city shopping mall in Sydney, hosting 250 stores including luxury good stores; and a large suburban mall with over 500 stores supporting the more routine needs of a number of suburbs on the outskirts of Melbourne, hosting, amongst others, a number of supermarkets. Both malls have food courts.

The data cover over a year (September 2012 – October 2013) of indoor Wi-Fi-based tracking of mall visitors – registered users of the Wi-Fi network (positioning by Access Point (AP) ID), and associated Web access logs (duration and destination of Web activity, such as Web search and browsing). The Sydney mall is covered around 70 APs and the Melbourne mall by around 100 APs, installed in a configuration for improved positioning performance, prospectively allowing for coordinate-based positioning. Note that the Wi-Fi coverage is restricted to the *common* spaces servicing the malls, not to the shops themselves (other than natural signal spill-over). The datasets cover approximately 120,000 and 90,000 individual registered users (users

[1] http://www.worldbank.org/en/topic/urbandevelopment

register to use the Wi-Fi). The dataset has been anonymised for privacy protection. Just the AP log contains more than 3 million events together for both malls.

In addition to the activity logs, the locations of APs have been mapped on floor plans of the malls and combined with the categorisation of shops by dominant type of sold item (e.g., shoes, groceries, food court).

3. Approach

As the aim of our project is to provide improved shopping recommendation services to the customers, the analysis of the behavioral patterns of shoppers in these two environments is based on a combined approach mining the Web access logs (separated into Web Query log and Web browsing log (Ren et al. 2014) in combination with the spatio-temporal patterns captured by the AP logs.

We note that trajectory pattern analysis method from movement ecology have been found to not be applicable with the type of datasets and environments at hand. The the trajectories of shoppers are very coarse (an AP service area can be up to 20m in diameter), highly incomplete, and short (Figure 1). The Wi-Fi positioning is only working when a user is actively browsing. Within a few seconds of the mobile device entering sleep mode, the Wi-Fi disconnects for power management. Second, the average length of a trajectory (as number of APs recorded within a visit) is just over 2

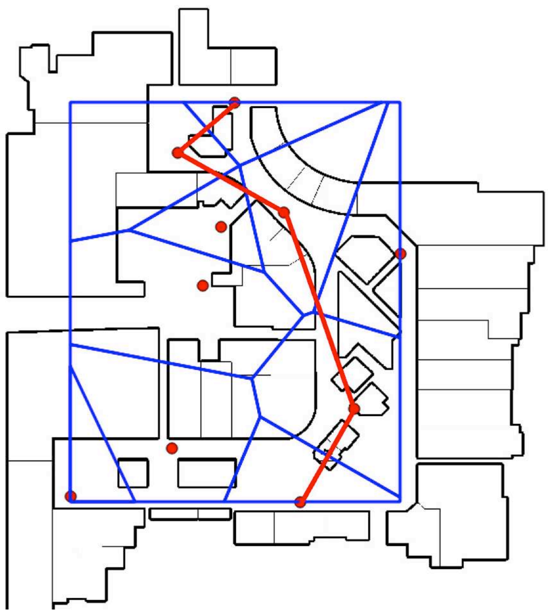


Figure 1 Part of the floorplan of the mall, with AP locations (red dots), their Voronoi regions (blue lines) and a recorded trajectory linking APs (red line).

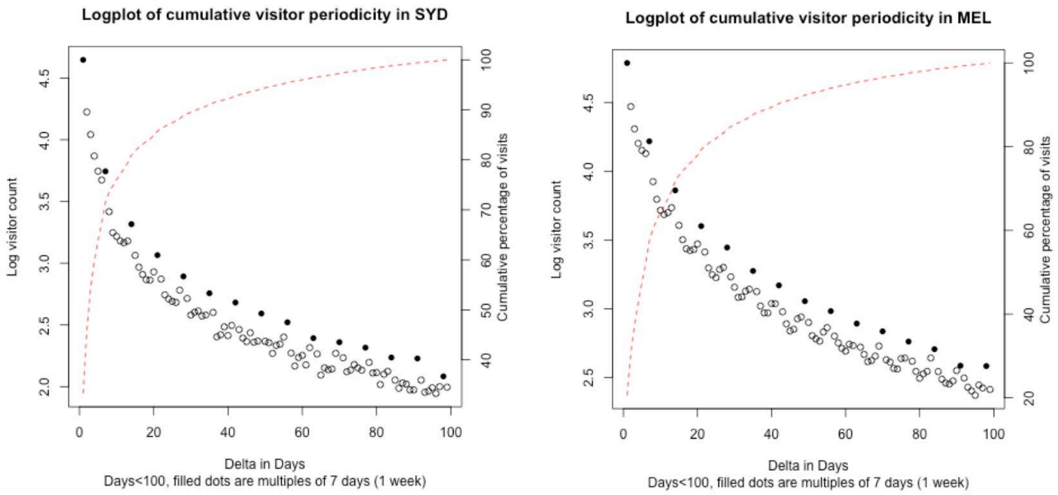
in our data.

In contrast to other approaches to movement data analysis (e.g., Buchin et al. 2012, Gschwend and Laube 2012), the contextual information is here provided by the location as proxy to ambient attributes (shops), and not by the activity of the user. The approximate catchment areas of each AP (constructed as Voronoi regions and refined to identify correct shop frontage) have been used to link the types of near shops to user's Web activity.

In Ren et al. (2011), we have demonstrated that the spatial location in the mall significantly influences the Web activity of users, as characterized by the categories of the Websites visited. In this paper, we focus on other coarse spatio-temporal patterns of the mall users' behaviour: namely the temporal patterns in their visits. We have analysed the aggregate periodicity of shoppers' visits and subsequently analysed the stability of individual shoppers temporal return patterns. It is assumed that this information is a strong indicator of the type of shopper.

4. Preliminary Results

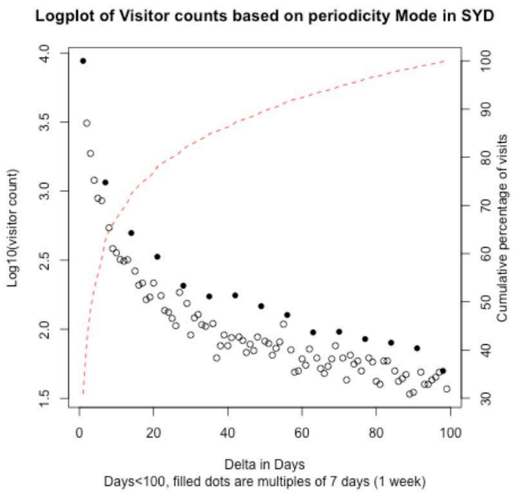
Based on the hypothesis that the frequency with which people visit an environment is linked to the kind of needs the environment satisfies, we have started exploring the frequency of return patterns to the shopping malls. The inner city shopping mall has a large proportion of one-time-only visitors (assumed to be dominantly tourists) in the period covered. As it sells also luxury goods, some of these one-time only visits may also be associated with shoppers purchasing goods that are seldom acquired (e.g., jewelry). The aggregate return patterns of visitors with more than a single visit are shown in Figure 1. The displayed periodicities are cut of at 100 days, as return periods are increasingly influenced by the dataset size (just over a year) and are therefore noisy.

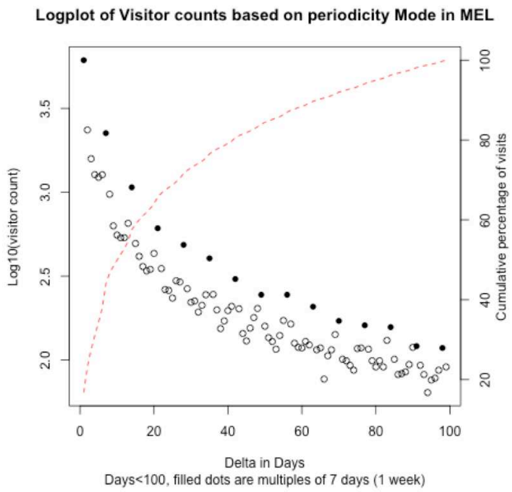


(a) Sydney – inner city mall (b) Melbourne – suburban mall

Figure 2 Cumulative differences between consecutive visits of all visitors, in days. The systematically higher occurrence of visits differences of 7 day multiples is clear in both environments. Red line shows the cumulative percentage of visits covered (excluding visits of users that visited only once).

A pattern of increased likelihood of a return visit to a mall in a multiple of 7 days is visible in both centers. These Figures 1(a) and (b), however, do not show if such patterns also occur on individual levels and how stable they are (Figure 2).





(a) Sydney – inner city mall (b) Melbourne – suburban mall

Figure 3 Individual visitor periodicity (based on most common (mode) of differences in days per user).

An initial evaluation of the stability of the mode statistic (the most common value) shows that the proportion of visits covered may be sufficient to characterize the type of mall visitor as a daily, weekly, bi-weekly, monthly or sporadic visitor (Figure 3). Such

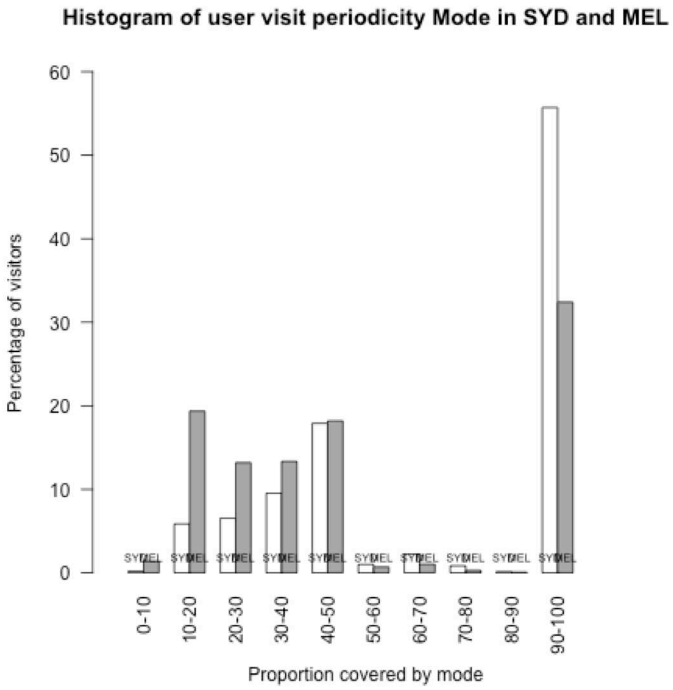


Figure 4 Mode of individual visit periodicity as an indicator of metric stability. The 90-100% bin mostly covers users with only two visits (hence the mode covers 100% of their visits).

coarse categories are, however, highly significant for the targeted provision of information to visitors by the mall operator.

4. Discussion and Conclusions

We provide an overview of a Eulerian movement dataset that is representative of those collected by operators of large indoor spaces, but rarely available to academics – a comprehensive dataset of users' indoor behaviour in retail spaces. The dataset covers both users' physical and information behaviours. We argue that this dataset the context information is provided by location itself, and the main information is the Web activity that has been shown to be associated with the shopping environment (Ren et al., 2014). We argue that user classification based on trajectory similarity analysis may not be applicable in realistic indoor tracking datasets due to coarseness and incompleteness of positioning. Finally, we show the results of the initial temporal analysis of the return visits of shoppers – their most basic movement pattern.

We are currently analysing the social networks inherent in the data and extracted based on spatio-temporal co-occurrence of users. Common co-occurrence may indicate e.g., family units, but it is assumed that highly frequent co-occurrence will only indicate accidental associations between staff or e.g., staff and a frequent customer to a coffee shop.

Future analysis will also focus on the Wi-Fi Received Signal Strength Indicator (RSSI) data from the access logs to improve the accuracy of position estimation within a Voronoi region. This can be used to refine the estimation of shoppers' trajectory patterns. The entropy of the patterns of individuals' visits will also be explored to improve shoppers' classification.

Acknowledgements

This research has been supported by a Linkage Project grant of the Australian Research Council (LP120200413).

References

Buchin M, Dodge S and Speckmann B 2012, Context-aware similarity of trajectories. *Geographic information science*, Springer Berlin Heidelberg, 43-56.

Gschwend C and Laube P 2012, Challenges of Context-Aware Movement Analysis - Lessons learned about Crucial Data Requirements and Pre-processing. In Whyatt D and Rowlingson B (eds.), *Proceedings of the 20th GIS Research UK Annual Conference GISRUK 2012*, April 11-13, 2012, Lancaster, UK, 241-246.

Kantor P B, Rokach L, Ricci F and Shapira B 2011, *Recommender systems handbook,* Springer.

Ren Y, Tomko M, Ong K, Yuntian B and Sanderson M 2014, The Influence of Indoor Spatial Context on User Information Behaviours. *Proceedings of the Workshop on Information Access in Smart Cities, held in conjunction with the 36th European Conference on Information Retrieval ECIR 2014*, Amsterdam, NL.