Who are my visitors and where do they come from?

An analysis based on Foursquare check-ins and place- based semantics

P. Hallot[1], K. Stewart[2] and R. Billen[1]

Geomatics Unit, University of Liège, Belgium

Department of Geographical and Sustainability Science, The University of Iowa, Iowa - USA

ABSTRACT

Activity recommendation systems aims at providing relevant information depending on targeted users' groups. For instance in a city, it makes sense to differentiate local residents from tourists. This research investigates to what extent the anonymized data collected from social networks can be used as a basis for making activity recommendations associated with local residents versus tourists when visiting a public place, such as a museum or gallery. Using rules based on the spatial, temporal and semantics of visited places, we are able to infer if a user is likely to be local or a tourist, based on anonymous sample Foursquare data and place-based semantics retrieved using Google Places API. Using semantics of visited places, it becomes possible to infer additional information about a user based on their movements over space and time. Depending on the kind and frequency of visited places, inferences about the aim of a visit to a location are possible. This analysis could provide information to users in the form of recommendations based on their movements while travelling around an area. This study has been performed using Foursquare check-ins for visitors to the Art Institute of Chicago between March 2010 and January 2011.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Modeling Methodologies, Visualization theory, concepts and paradigms

1. Introduction

Exploring a city requires different information depending on the background of an individual. A person living in a city has the opportunity to visit a local attraction as frequently as desired, and may also have knowledge of additional nearby attractions. On the other hand, out of town visitors or tourists require different information in order to discover attractions, information that matches with the time of their visit or the kind of activities in which they are interested e.g. a general visit to a city or a more focused trip to see specific attractions such as art galleries, etc. Tailoring recommendations that would appeal to local residents *versus* tourists offers these two different groups opportunities to target information that is more useful for them based on their expected availability and interests with respect to a location, and is a key motivation for activity recommendation systems (Bereuter et al. 2012, Buchin et al. 2012, Zheng et al. 2010).

In this research, we are interested in knowing to what extent the anonymized data collected from Foursquare's social networking site (https://foursquare.com/) can be used as a basis for making activity recommendations associated with local residents *versus* tourists when visiting a public place, such as a museum or gallery. Semantic information associated with the locations where users have checked-in will be used to give clues as to whether visitors are local or not. In this study, *local* refers to a person performing most of their social and professional activities within a city where the attraction is located, while *tourist* refers to individuals whose movements originate from more than 50 miles of the tourist attraction. These definitions can be modified as desired for different domains. Using Foursquare data, a set of rules is developed to determine if a visitor can be specified as a *local* or a *tourist*. As a case study, Foursquare check-ins for visitors to the Art Institute of Chicago between March 2010 and January 2011 are analyzed. With this work, the utility of social media data for activity recommendation purposes is investigated where a distinction is made between visitors' origins. Knowing

whether visitors are local or are tourists from another region is relevant to targeting information and advertising, for example about the Art Institute's ongoing activities, such as weekly lunchtime talks that might be attended mostly by locals *vs.* blockbuster exhibitions that could be a particular draw for tourists.

Activity recommendation systems have been developed for different domains such as activity-based learning where learning processes are created and recommended based on former students activities and knowledge (Florian et al. 2011), business management where matching is between provided services and customers (Wang et al. 2010), and itinerary planning (Liu and Lu 2013). The growth of social network systems such as Foursquare, Twitter, and Flickr provides more frequent and spatially localized data on individuals' movement histories or activity paths that could be useful for designing automated recommenders (Chen et al. 2011). However, the analysis of big data associated with social media data collection efforts has some challenges including a possible lack of data completeness due to ensuring personal privacy (Damiani and Cuijpers 2012), and the management of temporally non-continuous information since the data are collected only when a user interacts with the network.

2. Modeling movement with Foursquare check-in data

Data provided by location-based social networks (LBSN) such as Foursquare, have been studied to determine the kinds of activities that users perform within an urban environment. This is useful, for example, for urban planning, content delivery, as well as activity recommendations (Anastasios 2013, Noulas et al. 2011). LBSN can be used to infer knowledge about individuals checking-in. For example, researchers have been able to determine the living area of users from Foursquare check-ins and to give recommendations about places users have visited (Pontes et al. 2012), and have analyzed how users manage the recommendations provided by the network (Silva et al. 2014). On the other hand, the semantic annotation of visited places remains an issue, for example, when a user checks-in, information is collected at different levels of granularity. To ensure that all these granularities are captured, algorithms automatically proposing to annotate places at different levels are being developed (Ye et al. 2011). It is notheworty to mention that pricacy policies change continuously in favor of a greater protection of social networks' users. In this research, Foursquare check-in data are combined with semantic information about the kind of places where check-ins occur using Google Places (https://developers.google.com/places/), to determine whether the visitors using check-ins are tourists or locals, and to derive the area of influence of a location (the Art Institute of Chicago).

3. Semantic of checks-ins

The anonymous sample dataset covers worldwide check-ins for Foursquare for the period March 2010 to January 2011. The dataset is composed of 2,073,740 tuples representing each check-in with a unique id, a time, coordinates (latitude and longitude), and a location id.. Precise place names are anonymized by a number and this information cannot be obtained due to pricacy reasons. However, to avoid this reseach limitation and retrieve semantic information on each place checked-in, the reverse-geocoding capability of Google Place API was used (https://developers.google.com/places/). This allows a location to be associated with a place name, a type describing the kind of place, and the distance between closest retrieved location and the coded check-in place. In this case, we are interested in the places caterogies rather than the place itself by respect to the anonimity. Supported types provided by Google Place API include for example: airport, park, art gallery, café, church, library, health, and food. Often, a check-in place may fit more than one of these semantic categories, i.e., the classification is not a disjoint taxonomy. For example, type "church" is also a member of the type "place_of_worship". To ensure a better representation of the proposed types, a categorization is developed where each location is attributed with a unique type. If the location submitted to the Google API does not return a categorized place within a fixed radius (20m in this study, that represents the mean positioning error), the address and the neighborhood name is recorded. Figure 1 shows the frequency of types of places where users checked-in within the downtown Chicago area from March 2010 to January 2011.
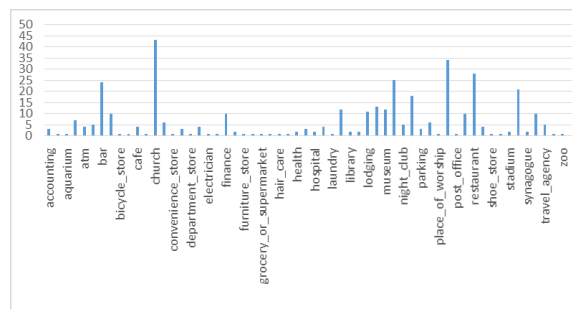

Figure 1: *Frequency of types of places checked-in within downtown Chicago, IL from March 2010 to January 2011.*

4. Determination of whether a visitor is local or tourist

All users who checked-in at least once within 20m of the Art Institute of Chicago are selected for this study, and all the types of places where these same users have checked-in during the previous nine months are also retrieved. The locations of the check-ins, frequency, distance, and the semantics of places checked-in are analyzed to determine if the user is estimated to live within Chicago or if they come from another location, in order to determine if the user is a *local* or a *tourist*. In this work, three rules are applied to determine if a user can be qualified as local or not. To analyze the spatial and temporal location of the check-ins, QGIS is used to compute spatial clusters of the check-ins that are within a maximal distance 50 miles of each other

(The presented results have been computed with a 50 miles value, others simulations have been tested. We consider that traveling more than 50 miles lead to define a journey). Three parameters are studied with respect to the clusters relating to the Art Institute:

1. *The interval of time during which the check-ins occur within a cluster.*

A short time period (i.e., a few days or a week) for check-ins within a region may suggests a tourist's visit, while a longer period of sustained check-ins (e.g, 9 months in this study) could suggest a local.

2. *The semantics of places retrieved within a cluster.*

A wider range of semantics could suggest a local user, for example, check-ins that correspond to a health clinic, hairdresser, and lawyer's office, while check-ins to a hotel or rental car dealer may suggest a tourist.

### 3. *The frequency of check-ins within the same places.*

Frequent visits of similar places suggests a local user. A tourist will tend to visit places less frequently (one or two times). A local may check-in frequently at her favorite restaurant, gym, or at a health clinic.

When examining the check-in history for two different individuals that have visited the Art Institute of Chicago, we see that, within the cluster located at the Art Institute of Chicago, one user checks-in 688 times in the cluster, 37 times over a period of three month at a furniture store (that suggests it is their work-place) (Figure 2a), while the other person's check-ins within the cluster cover a period of only 7 days (Figure 2b). With regard to the Art Institute, the local user has checked-in two times over two months, while the tourist user has only checked-in once (Figure 2b).

Check-ins of a
"local" user
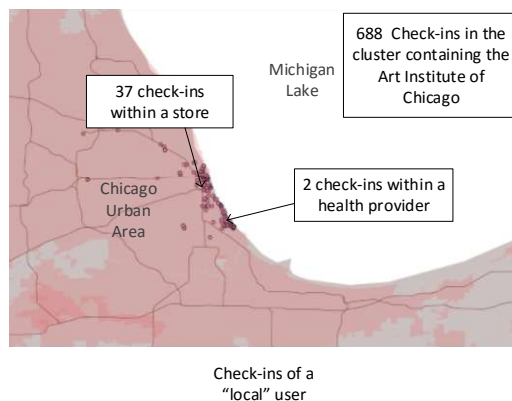
(a)

Figure 2a: *Check-ins of two different users within a cluster covering the Art Institute of Chicago. This shows a pattern of check-ins typical of a local user over a 9 month period.*

Check-ins of a
"tourist" user
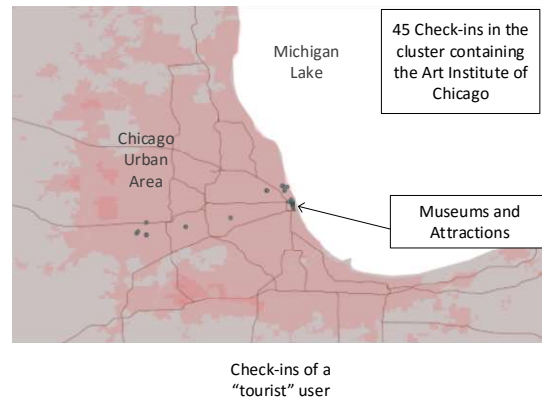
(b)

Figure 2b: *Check-ins of two different users within a cluster covering the Art Institute of Chicago. This shows tourist check-ins that all occur during one week.*

4. Conclusions

Using rules based on the spatial, temporal and semantics of visited places, we are able to infer if a user is likely to be local or a tourist, based on anonymous Foursquare data and place-based semantics retrieved using Google Places API. Using semantics of visited places, it becomes possible to infer additional information about a user based on their movements over space and time. Depending on the kind and frequency of visited places, inferences about the aim of a visit to a location are possible. This analysis could provide information to users in the form of recommendations based on their movements while travelling around an area. Furthermore, for an attraction such as the Art Institute of Chicago, learning whether specific events in their schedule brings in more local or tourists could be useful communications about events. The model is undergoing a full evaluation to discern the accuracy of the semantic annotations as well as the check-in data. Additional research is possible, for example, automatically deriving an attraction's area of influence over time. The model can also be extended to include other indicators such as the detection of a commuting path within a cluster (Buchin et al. 2011), or through daily activities detection. These extensions will be investigated in future work.

5. Acknowledgements

References

[BUC*11] BUCHIN, K., et al. 2011. Detecting commuting patterns by clustering subtrajectories. International Journal of Computational Geometry & Applications, 21(03), 253-282.

[BUC*12] BUCHIN, M., DODGE, S. AND SPECKMANN, B., 2012. Context-Aware Similarity of Trajectories. In: XIAO, N., et al. eds. Geographic Information Science. Springer Berlin Heidelberg, 43-56

[CHE*11] CHEN, J., et al. 2011. Exploratory data analysis of activity diary data: a space–time GIS approach. Journal of Transport Geography, 19(3), 394-404.

[DAM*12] DAMIANI, M. L. AND CUIJPERS, C., 2012 Privacy-aware geolocation interfaces for volunteered geography: a case study. ed. Proceedings of the 1st ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information, 83-90.

[FLO*11] FLORIAN, B., et al., 2011. Activity-Based Learner-Models for Learner Monitoring and Recommendations in Moodle. In: KLOOS, C., et al. eds. Towards Ubiquitous Learning. Springer Berlin Heidelberg, 111-124.

[LIU*13] LIU, J.-S. AND LU, W.-H., Location and Activity Recommendation by Using Consecutive Itinerary Matching Model. ed. ROCLING 2013 Kaohsiung, Taiwan.

[NOU*11] NOULAS, A., et al. 2011. An Empirical Study of Geographic User Activity Patterns in Foursquare. ICWSM, 11, 70-573.

[PON12] PONTES, T., et al., 2012. We know where you live: privacy characterization of foursquare behavior. Proceedings of the 2012 ACM Conference on Ubiquitous Computing. Pittsburgh, Pennsylvania: ACM, 898-905.

[SIL*14] SILVA, T. H., et al. 2014. You are What you Eat (and Drink): Identifying Cultural Boundaries by Analyzing Food & Drink Habits in Foursquare. arXiv:1404.1009.

[WAN*10] WANG, C.-Y., WU, Y.-H. AND CHOU, S.-C. 2010. Toward a ubiquitous personalized daily-life activity recommendation service with contextual information: a services science perspective. Information Systems and e-Business Management, 8(1), 13-32.

[YE*14] YE, M., et al., On the semantic annotation of places in location-based social networks. ed. Proceedings of the 17th ACM SIGKDD

[ZHE*10] ZHENG, V. W., et al., 2010. Collaborative location and activity recommendations with GPS history data. Proceedings of the 19th international conference on World wide web. Raleigh, North Carolina, USA: ACM, 1029-1038