Empirical Observation of User Activities: Check-ins, Venue Photos and Tips in Foursquare

Yi Yu[1], Suhua Tang[2], Roger Zimmermann[1], Kiyoharu Aizawa[3]
[1]School of Computing, National University of Singapore, Singapore
[2]Graduate School of Informatics and Engineering, The University of Electro-Communications, Japan
[3]Department of Information and Communication Engineering, The University of Tokyo, Japan
{yuy,rogerz}@comp.nus.edu.sg, shtang@uec.ac.jp, aizawa@hal.t.u-tokyo.ac.jp

ABSTRACT

Location-based social networking platform (e.g., Foursquare), as a popular scenario of participatory sensing system that collects heterogeneous information (such as tips and photos) of venues from users, has attracted much attention recently. In this paper, we study the distribution of these information and their relationship, based on a large dataset crawled from Foursquare, which consists of 2,728,411 photos, 1,212,136 tips and 148,924,749 check-ins of 190,649 venues, contributed by 508,467 users. We analyze the distribution of user-generated check-ins, venue photos and venue tips, and show interesting category patterns and correlation among these information. In addition, we make the following observations: i) Venue photos in Foursquare are able to significantly make venues more social and popular. ii) Users share venue photos highly related to food category. iii) Category dynamics of venue photo sharing have similar patterns as that of venue tips and user check-ins at the venues. iv) Users tend to share photos rather than tips. We distribute our data and source codes under the request of research purposes (email: yi.yu.yy@gmail.com).

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behavioral Sciences; G.3 [**Mathematics of Computing**]: Probability and Statistics; H.4 [**Information Systems Applications**]: Miscellaneous

Keywords

Social media, participatory sensing, Foursquare, check-ins, venue photos and tips, venue category dynamics, venue analytics, online sharing activities

1. INTRODUCTION

Widespread of mobile devices has enabled people to easily engage with the Internet anywhere and anytime. Meanwhile, explicitly checking in at venues and instantly sharing photos and comments on location-based social networking platforms (e.g., Foursquare, Facebook, Places, Google + Local) have been popular user activities. Accordingly, vast volumes of check-ins, venue photos, and venue comments are aggregated over time [1], which are potentially valuable knowledge source about user physical activities and online sharing behaviors in the era of big data.

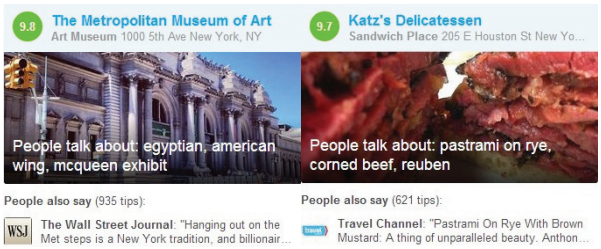

Figure 1: Foursquare helps you find the perfect places in New York to go with your friends.

Foursquare is one of the most popular location-based social networks (LBSNs). As an intersection of virtual social networks and physical world, it helps connect people with friends and venues around the world. People check in at various venues to tell their friends where they love to go and leave tips and photos at their favorite places to share experiences with friends, as shown in Fig. 1. As of December 2013, it is reported that Foursquare has 45 million registered users (with 5 billion check-ins) who are spreading the world about their favorite spots[1]. Such location sharing service also is a popular type of participatory sensing systems [2][3] of venues. The heterogeneous information of a particular venue could be check-ins, tips, and photos. A check-in indicates a physical visit to a venue, while a photo or a tip reflects a user's personal comments on this venue. As a venue is associated with a geographic category such as bar and beach, check-ins, tips and photos of a venue all inherit this property.

Sensed data in Foursquare contains user activities in physical world and sharing activities on the Internet, and denotes whereabouts of their social behaviors. It not only contains personalized information of each individual user, but also reflects social behaviors related to the society through aggregating more users' data streams. Such heterogeneous data source of people activities is able to play a very important

[1]http://en.wikipedia.org/wiki/Foursquare

MM'14, November 07 2014, Orlando, FL, USA

http://dx.doi.org/10.1145/2661714.2661724.

	NYC		LA	
Cat.	#users	#venues	#users	#venues
1	3,914,458	6,065	1,421,047	3,301
2	515,256	2,910	236,508	1,499
3	13,484	55	1,674	19
4	16,987,902	31,641	6,415,261	15,941
5	6,219,712	7,865	1,214,428	2,748
6	3,407,437	6,729	916,367	3,400
7	2,400,257	21,684	777,010	10,548
8	161,609	4,427	38,756	1,338
9	6,779,043	31,620	3,025,047	17,992
10	4,860,249	8,256	1,675,119	2,819

Table 1: Numbers of users (user number may be duplicate when a user visits venues with different categories) and venues per-category in NYC and LA.

role in location-aware recommendation [4] (e.g., provide media advertisement and travel plan [5]) and urban computing [6] (e.g., provide potential sustainability and outlook of urban environment, people life quality [7], city planning [8], and social sciences [9]).

Taking a large dataset crawled from Foursquare as an example, we present a case study of heterogeneous data sensed by users. We have collected from Foursquare information of 190,649 venues located in New York City and Los Angeles, which contains 2,728,411 venue photos, 1,212,136 tips, and 148,924,749 check-ins contributed by 508,467 users. We present an analysis of spatial and temporal dynamics of heterogeneous data sensed by users and show how venue photos, check-ins, and tips provide a means to discover people lifestyle patterns. Our results indicate that heterogeneous data sensed by users in location-based social networks provide us more opportunities to predict users' physical activities and online sharing behaviours. User activities are firstly investigated over heterogeneous information. The potential applications of this initial study is able to complementarily serve the smart online services spanning from location recommendation, trip planning, media advertising to urban environment improvement and city operation. We distribute our data and source codes related to this study under the request of research purposes (email: yi.yu.yy@gmail.com & URL: http://www.comp.nus.edu.sg/∼yuy/).

2. FOURSQUARE DATA DESCRIPTION

A Foursquare venue can be created with Foursquare API. It is associated with a physical location in the real world, e.g., John F. Kennedy International Airport, Yankee Stadium, Union Square Park. Users in Foursquare also can leave tips and venue photos to share their experiences with people when they check in at certain venues. Through making use of Foursquare API, we are able to acquire the following information about a venue: venue name, venue category, geographic coordinates, number of check-ins, number of visited users, number of tips and details, number of photos and details.

In our initial investigation, from February 21 to February 27 2014, we collected information of venues in two geographic regions: New York City (NYC) and Los Angeles (LA). We utilize 10 primary categories provided by Foursquare to analyze statistics of our data. As shown in Table 1, category indices from 1 to 10 respectively correspond to: Arts & Entertainment, College & University, Event, Food, Nightlife Spot, Outdoor & Recreation, Professional & other places, Residence, Shop & Service, and Travel & Transport.

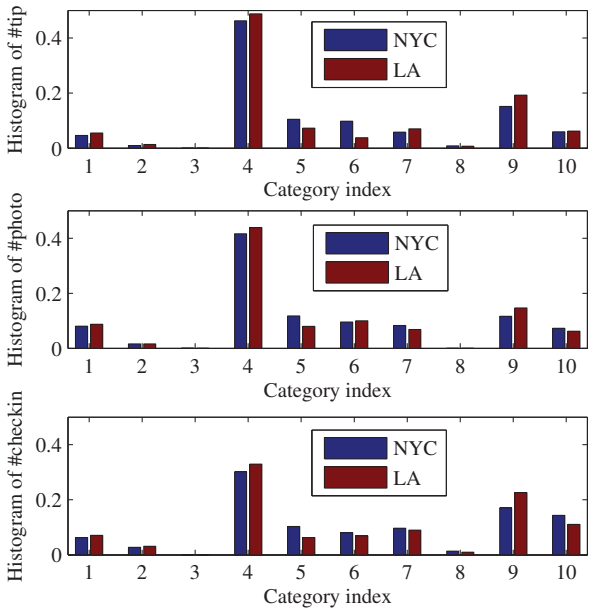

Figure 2: Distributions of #tips, #photos, and #check-ins in NYC and LA, in terms of the 10 primary categories defined in Foursquare.

3. DATA ANALYSIS

We take Foursquare as a participatory sensing platform to collect check-ins, tips, and photos of the venues. Participatory sensing [2] is the concept of communities (or other groups of people) contributing sensory information to collectively form knowledge. It also is an approach to data collection and interpretation in which individuals, acting alone or in groups, use their personal mobile devices and web services to systematically explore interesting aspects of their worlds. Here, we investigate user activity patterns behind heterogeneous data of venues.

The distributions of tips, photos, check-ins in terms of venue categories, shown in Fig. 2, reflect two similarities: i). These distributions are similar in different regions, as can be seen by comparing two curves in each plot. This is because both NYC and LA are American cities. Users there have the same culture and their activities exhibit similarities. Another potential factor might be the preferences of user activities in physical locations are correlated with online sharing activities, e.g., most people love to check in at various restaurants to share photos and leave tips online. ii). Tips, photos, and check-ins show similar trends in different categories. This infers that there is a consistency in user preferences on venue categories, which may be in different forms though. Since Category 3 is newly added in Foursquare, we cannot find many tips, photos, and check-ins in this category.

In Figs. 3 and 4, we plot the histograms of photos and tips across the seven most popular categories within the course of a week. Although the numbers of tips and photos are different for each category, they show similar trend for some

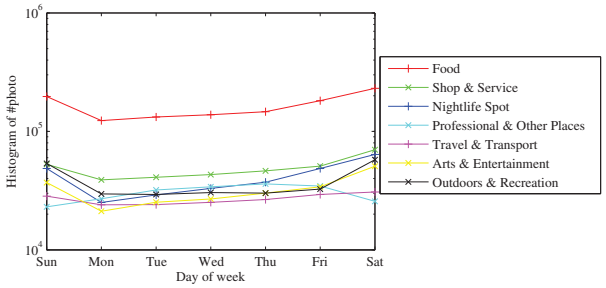

Figure 3: Number of photos per weekday.

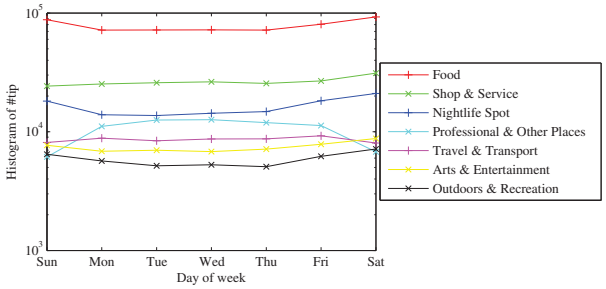

Figure 4: Number of tips per weekday.

categories. Food, nightlife spot, arts & entertainment, outdoors & recreation are more popular at weekends (in the free time) than weekdays. In comparison, more professional activities are on weekdays. There are some differences in the category of travel & transport. This category has more tips on weekdays but more photos on weekend.

Venues have different numbers of tips, photos, and check-ins. The CCDF (Complementary cumulative distribution functions) results of tips, photos, and check-ins are respectively shown in the three plots of Fig. 5. A common trend is that only a few venues have a large number of events. 50% venues have only one tip while 0.83% (NYC) or 0.25% (LA) venues have more than 100 tips, and the average is 7.04 in NYC and 5.03 in LA. A similar result can be obtained for photos and check-ins.

Users post tips or photos at irregular intervals and different places. CCDF results of inter-visit time (in terms of tip or photo) and inter-visit distance (the distance between two venues being successively visited) are shown in Fig. 6. 50% inter-visit time is more than 7.3 days for tips, and 1.83 days for photos, and the average is 50.0 and 17.3 days, respectively. As for inter-visit distance, 50% is more than 3.72 km for tips and 4.03 km for photos, and the average is 6.10 and 6.67km, respectively.

Users have different preferences of posting tips and photos. To find whether the two factors are correlated, we investigated the numbers of tips and photos posted by each user, and the numbers of tips and photos posted to each venue. The histograms of users and venues in terms of the tuples <#tip, #photo> are shown in Fig. 7 and Fig. 8, respectively. The surface of user histogram is relatively flat. In contrast, the surface of venue histogram is much sharper. This can be explained as follows: The numbers of tips and photos contributed by a specific user might be greatly different; But popular venues attract many users; Users as a

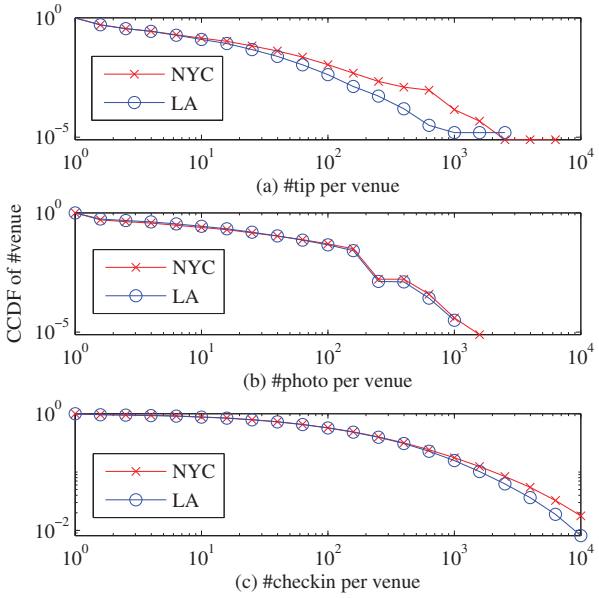

Figure 5: CCDF of number of venues in terms of tips, photos, and check-ins.

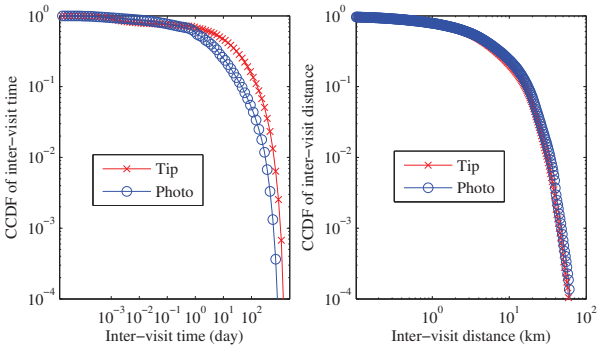

Figure 6: Distribution of inter-visit time and inter-visit distance by a same user, in terms of tips or photos.

population exhibit a trend in preferring to posting photos than tips.

4. DISCUSSION

Foursquare implements a location layer for the Internet, which can provide useful data about the context at any moment. This work analyzes the spatial and temporal patterns of heterogeneous data of venues sensed by Foursquare users and studies the relationships among venue check-ins, photos, and tips. Check-ins implicate users' activities in physical world. Photos and tips as interesting social contents of Foursquare can help users to easily make judgements about a venue as shown in Fig. 9: what is there? Then, users have opportunity to visually watch the venue environment by viewing the shared venue photos (e.g., venue photos mainly describe business storefronts and interiors and service contents) and enjoying the experience written in venue tips. To the best of our knowledge, this is the first attempt to char-

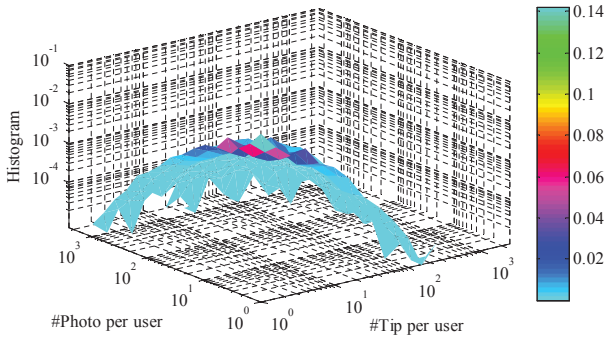

Figure 7: Histogram of users in terms of the tuples <#tip, #photo>.

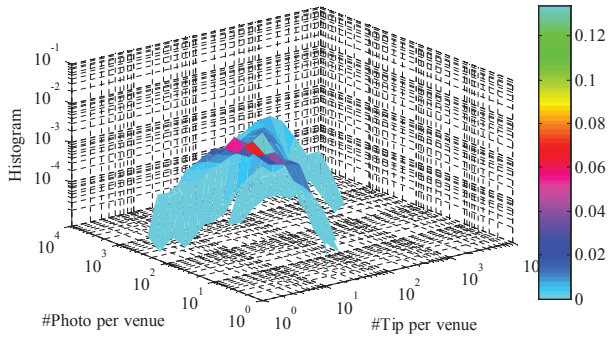

Figure 8: Histogram of venues in terms of the tuples <#tip, #photo>.

acterize heterogeneous data (e.g., check-ins, venue photos and tips) sensed by users in location-based social networking platform, e.g., Foursquare.

In our initial study, taking Foursquare as an example platform of participatory sensing system, we collected a dataset from two regions in USA. Our aim is to understand implications behind heterogeneous data from different regions with different users, with respect to the study of user activities (e.g., an individual level or an regional level). We summarize our observations here:

- Category dynamics among venue photo sharing, tips sharing, and check-ins have analogous geo-temporal rhythms.
- Shared venue photos are highly relevant to food.
- Users prefer to share photos rather than tips.
- Venue photos are more important in promoting venues.

We currently are collecting data from broader regions to investigate more detailed properties of sensed data. We plan to learn prediction models on user activity with regard to multimedia retrieval and recommendation, urban computing and social mobility so as to provide better services to match users' preferences.

Acknowledgments

This research has been supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office through the Centre of Social Media Innovations for Communities (COSMIC).

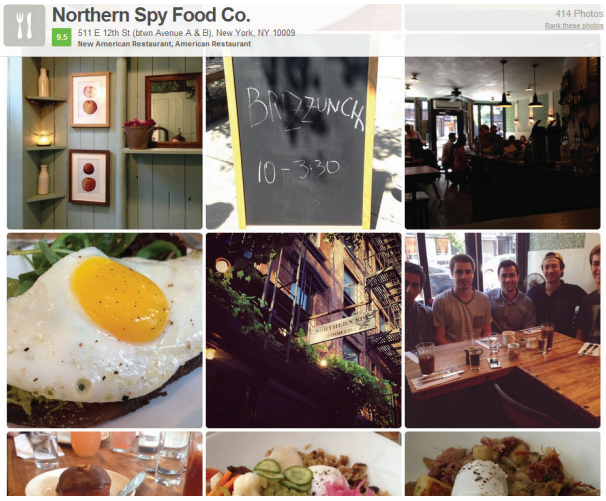

Figure 9: Venue photos mainly describe business storefronts and interiors and service contents.

5. REFERENCES

[1] Y. Yu, S. Tang, and R. Zimmermann. Edge-based locality sensitive hashing for efficient geo-fencing application. In Proc. of 21^{st} ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 576-579, 2013.

[2] D. Estrin. Participatory sensing: applications and architecture. Internet Computing, IEEE, vol.14, no.1, pp.12-14, 2010.

[3] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava. Participatory sensing. Workshop on World-Sensor-Web: Mobile Device Centric Sensor Networks and Appilitions, pp. 117-134, 2006.

[4] R. Zimmermann and Y. Yu. Social interactions over geographic-aware multimedia systems. In Proc. of ACM MM'13, pp. 1115-1116, 2013.

[5] Y. Yu, K. Aizawa, T. Yamasaki, and R. Zimmermann. Emerging topics on personalized and localized multimedia information systems. In Proc. of ACM MM'14, 2014.

[6] J. Cranshaw, R. Schwartz, J. I. Hong, and N. Sadeh. The livehoods project: utilizing social media to understand the dynamics of a city. In Proc. of the Sixth Int'l Conf. on Weblogs and Social Media, pp. 58-65, 2012.

[7] R. Jain, L. Jalali, and M. Fan. From health-persona to societal health. In Proc. of WWW'13 Companion, pp.1329-1334, 2013.

[8] T. H. Silva, P. O. S. Vaz De Melo, J. M. Almeida, and A. A. F. Loureiro. Large-scale study of city dynamics and urban social behavior using participatory sensing. Wireless Communications, IEEE, vol.21, no.1, pp.42-51, 2014.

[9] Y.-S. Cho, G. V. Steeg, and A. Galstyan. Where and why users "check in". In Proc. of AAAI'14, pp.269-275, 2014.