Mining Collective Local Knowledge from Google MyMaps

Shaomei Wu[*] Shenwei Liu, Dan Cosley
Department of Information Science
Cornell University
{sw475,sl2277,drc44}@cornell.edu

Michael W. Macy
Department of Sociology
Cornell University
mwm14@cornell.edu

ABSTRACT

The emerging popularity of location-aware devices and location-based services has generated a growing archive of digital traces of people's activities and opinions in physical space. In this study, we leverage geo-referenced user-generated content from Google MyMaps to discover collective local knowledge and understand the differing perceptions of urban space. Working with the large collection of annotation-rich, publicly available MyMaps data, we propose a highly parallelizable approach in order to merge identical places, discover landmarks, and recommend places. Additionally, we conduct interviews with New York City residents/visitors to validate the quantitative findings.

Categories and Subject Descriptors

H2.8 [**Database Management**]: Database Applications-Data Mining; H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Algorithms, Human Factors

Keywords

Geo-tagged data, user-generated content, place recommendation

1. INTRODUCTION

With the growing popularity of location-based services such as Foursquare and Google Latitude, the increasing amount of user-generated geo-referenced data provides us a way to access knowledge about the physical environment that would otherwise be limited to locals. While most existing work focuses on data classification and visualization (e.g. [1, 2, 3]), the characteristics of places and the people who interact with the places are often overlooked. In this paper, we propose to mine community-specific insights about the urban space from the publicly available Google MyMaps data.

Google MyMaps is a Web service that enables Google users to create personalized annotated maps. This dataset features three unique properties: First, the annotation of places in Google MyMaps is centralized around the physical space per se, whereas in other services (e.g., Flickr), geotags are largely a collateral attribute of the data. Second, in addition to commercial/popular places that most consumer review sites usually cover, MyMaps also contains more heterogeneous content including many personally meaningful places. Third, as every MyMap is a manually-curated collection of places that people put together because of underlying association or specific purpose, the co-occurrence of places in maps potentially offers rich insights into the connections among them.

We make two main contributions. First, we use MyMaps data to show that locals' perception of important places is quite different from the kinds of globally-identified landmarks found in [2]: locals tend to identify cultural sites such as restaurants more frequently than non-locals, who tend to map tourist landmarks. We then show that how these data might be useful in a taste-sensitive recommender system that uses these local insights and latent connections from the maps people create online.

2. METHODS AND RESULTS

The first step when working with user-generated geographical data is to consolidate duplicated annotations of a single place. As reported in [2], textual information is crucial in identifying multiple representations of the same place. With this in mind, we developed an adaptive heuristic method that duplicates placemarks by merging nearby placemarks with similar names. By tuning the size of a *nearby* area and the threshold for the *similarity* between place names, our method outperforms previous approaches [1, 2] in two aspects: (1) it is a parallelizable and efficient algorithm; (2) it is fuzzy enough to recognize similar annotations of the same place (e.g., "Apple Store 5th ave" and "Apple Store") while differentiating geographically clustered places (e.g.,"Apple Store 5th ave" and "FAO Schwarz") with high resolution.

2.1 Discovering Landmarks

Similar to [2]'s use of photo density in space to identify landmarks, we use frequency of annotation as a signal of a location's salience. Table 1 shows the seven most salient places in New York City from our data and from [2]. Although there is significant overlap, three of the seven most salient places found in the MyMaps dataset are museums, whereas none of the most geo-tagged places in the Flickr dataset are museums. This result suggests the distinctive nature of places highlighted in these two datasets: Flickr is biased towards places that lend themselves to photography,

[*]Wu is supported by NSF grants IIS-0910664 and 08-35451.

WWW 2011, March 28–April 1, 2011, Hyderabad, India.
ACM 978-1-4503-0637-9/11/03.

Table 1: Top Landmarks of New York City

	MyMap	**Flickr** (Crandall et. al 2009)
1	metropolitan museum of art	empirestatebuilding
2	solomon r guggenheim museum	timesquare
3	museum of modern art	rockefeller
4	grand central station	grandcentralstation
5	times square	applestore
6	world trade center	columbuscircle
7	empire state building	libertyisland

Table 2: NYC Landmarks: Local vs. Non-Local

Local	**Non-local**
chelsea market	empire state building
momofuku noodle bar	museum of modern art
spotted pig	metropolitan museum of art
magnolia bakery	solomon r guggenheim museum
metropolitan museum of art	apple store fifth avenue
freemans	times square
museum of modern art	american museum of natural history
katz's delicatessen inc	grand central terminal
corner Bistro	katz's delicatessen inc
dinosaur barbecue	century 21 department stores

especially photography using an iPhone, whereas MyMaps accommodates landmarks where photography is prohibited.

We then analyze how people from different groups pay attention to different parts of the city. We classify a user as "local" when his profile's self-reported location matches the city ("NYC", "New York city"), a borough, or a neighborhood name. Other users with location information are "non-local". About 20% of MyMaps creators had profile locations, 925 local and 1611 non-local. Table 2 shows that the top 10 landmarks for each group are quite different, with only two places in common (*Museum of Modern Art* and *Katz's Delicatessen Inc*). Local users mapped places related to daily life (restaurants, cafes, bars), while non-locals focused on tourist attractions and stores, suggesting different perspectives on the city: for local people it is the home where they live and socialize; for non-locals it is a metropolitan area to explore and consume.

These findings are supported by interview data. When asked about their favorite places, four NYC locals and long-time residents mentioned newly-blooming neighborhoods such as Chelsea and Williamsburg, while none of our non-local informants were familiar with these areas. Some locals also expressed the tendency to avoid areas popular with tourists, as "the sidewalks are so full, you basically can't walk there!"

2.2 Recommending Related Places

Based on the success of recommender systems in helping people find books, movies, and music to explore, we believe that a personalized recommender system equipped with collective local knowledge could be compelling. We apply collaborative filtering [4] to recommend places with similar characteristics based on the probability of them appearing in the same map. In our algorithm, the pairwise similarity between places is measured by the cosine correlation coefficient of their occurrences in all the maps. After a user picks a place he likes, our algorithm recommends the 10 places with the highest similarity scores.

Table 3 shows the recommendations given for *Katz's Delicatessen* by our algorithm and Google Maps' "nearby places you might like"[1]. Our results are more diverse: while most existing place recommender systems (e.g. Yelp, Google Maps) usually classify places into pre-fixed categories and make recommendations within a category, our system discovers latent connections and may better serve a user's need to explore. NYC local interviewees were impressed by our results and confirmed the informal but strong connections among the places our system recommends. For example, as they pointed out, places like *Katz's Delicatessen*, *corner bistro* and *shake shack* are "cheap", "bold", and "extremely satisfying"—"they are the true New York experience!"

Table 3: Recommendations Related to Katz's Delicatessen

Rank	**Our method**	**Google**
1	spotted pig	Carnegie Deli
2	momofuku noodle bar	Noah's Ark Original Deli
3	carnegie hall	Lombardi's Pizza
4	russ & daughters	Sarges Deli & Restaurant
5	lombardi's pizza	Italian Food Center
6	magnolia bakery	Bon Vivant Diner
7	museum of modern art	2nd Ave Deli
8	clinton street baking co	Stage Deli & Restaurant Inc
9	shake shack	Russ & Daughters
10	chelsea market	Lahore Deli

3. CONCLUSION

This paper presents our exploration of mining collective knowledge about places from user-generated, geo-referenced data at scale. Aiming at understanding the heterogeneous meaning of the physical space, we present the difference in the salience of places across communities (local/non-local) and develop a personalized place recommendation system that captures the taste-related latent qualities of places. We plan to extend our notion of social groups to spatial (particular neighborhoods) and socioeconomic (occupation and income) differences. Another interesting future direction is to bridge the online/offline space and predict possible social ties from the geo-reference dataset based on the hypothesis that "birds of a feather map together".

4. REFERENCES

[1] S. Ahern, M. Naaman, R. Nair, and J. Yang. World explorer: Visualizing aggregate data from unstructured text in geo-referenced collections. In *JCDL 2007*.

[2] D. J. Crandall, L. Backstrom, D. Huttenlocher, and J. Kleinberg. Mapping the world's photos. In *WWW 2009*.

[3] T. Rattenbury, N. Good, and M. Naaman. Towards automatic extraction of event and place semantics from flickr tags. In *SIGIR 2007*.

[4] B. Sarwar, G. Karypis, J. Konstan, and J. Reidl. Item-based collaborative filtering recommendation algorithms. In *WWW 2001*.

[1]The "nearby places you might like" section appears when a user clicks on "more info" link in the search result in Google Maps.