

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 brought us a growing collection of digital traces of people’s activities and opinions in physical space. In this study, we leverage geo-referenced user-generated content from Google MyMaps¹ to discover collective local knowledge and understand the differing perceptions of urban space. Working with the large collection of publicly available, annotation-rich 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 findings from quantitative analysis.

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 Foursquare² and Google Latitude³, 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 focus 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. Comparing to most other location-aware services, this dataset features the following three

¹<http://maps.google.com/help/maps/mymaps/>

²<http://foursquare.com>

³<http://google.com/latitude>

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), geo-tags 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 us rich insights into the connections among them.

We make two main contributions in this work. 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 locally-frequented 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 cluster and 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 develop a more adaptive heuristic method that deduplicates 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 highly parallelizable and efficient automatic 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], where landmarks were defined as peaks of photo density in space, we use the frequency of a place being annotated in MyMaps as the signal of its salience. The 7 most salient places in New York City from our data are shown in Table 1, in comparison with the most geo-tagged places from Flickr photos as in [2]. It is noticeable that although there is a significant overlap between the top landmarks found in two datasets, the difference mainly comes from museums: 3 out of the 7 most salient places found by our algorithm 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 natures of places highlighted in these

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 ssam 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

two datasets: Flickr is biased towards places that lend themselves to photography, especially photography using an iPhone, whereas MyMaps accommodates worthy landmarks where photography is prohibited.

We then analyze how people from different groups pay attention to different parts of the city. Using the self-reported location information in the user's profile, we classify a user as NYC "local" when his location matches with the name of the city ("NYC", "New York city"), 5 boroughs, and neighborhoods. Users with location information that does not match any of above will be classified as non-locals. As a result, we are able to categorize 20% of all MyMaps creators, identifying 925 local users and 1611 non-local users. Accordingly, we show top 10 landmarks for different groups in Table 2. Not surprisingly, we see a substantial difference of the top places mapped by local and non-local users: with only two place in common (*Museum of Modern Art* and *Katz's Delicatessen Inc*), the places local users mapped are tightly related to day-to-day life (restaurants, cafes, bars), while the places mapped by non-locals are mostly tourist attractions and stores. Such difference suggests the different perspectives about New York City from the two communities: for local people the city is more of a home to live and socialize; for non-local people the city is rather a metropolitan area to explore and consume.

These findings are supported by our interviews. When asked about favorite places in NYC, all the 4 NYC locals and long-time residents we interviewed mentioned several newly-blooming neighborhoods such as Chelsea and Williamsburg, while none of our non-local informants were familiar with these areas. Additionally, some locals expressed the tendency to avoid passing 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 very beneficial and 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

Table 3: Recommendations of Places Related to the Katz's Delicatessen Restaurant

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

in all the MyMaps. After a user picks a place he likes, our algorithm will recommend the top 10 places with the highest similarity scores. Table 3 shows the recommendations our system gives for "Katz's Delicatessen" (a famous deli restaurant in NYC), in comparison with the "nearby places you might like" recommendations given by Google Maps⁴. As we can see, our results are promising in the sense that there is a better diversity in our recommendations: while most existing place recommender systems (e.g. Yelp, Google Maps) usually classify places into pre-fixed categories and make recommendations within category, our system does not rely on categories and can better serve a user's need to explore. When discussing the recommendations generated by our systems during the interviews, the NYC local interviewees were very impressed by our results and confirmed the informal but very 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!"

3. CONCLUSION

This paper synthesizes 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 is able to capture the taste-related latent qualities of places. We plan to extend our notion of social groups to more dimensions, such as spatial (particular neighborhoods) and socioeconomic (occupation and income). 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*.

⁴The "nearby places you might like" section shows up when a user clicks on "more info" link in the search result in Google Maps.