Mining Human-Place Interaction Patterns from Location-Based Social Networks to Enrich Place Categorization Systems

Yingjie Hu, Grant McKenzie, Krzysztof Janowicz, Song Gao

STKO Lab, Department of Geography, University of California Santa Barbara, USA

Abstract. Place categorization plays an important role in location-based services as well as more recently in place-based geographic information systems. Traditionally, such categorization systems are often designed following a top-down approach in which a group of experts or users assign a *place type*, e.g. Restaurant, to a *place instance*, e.g., Bob's BBQ shack. While the output of such a process generally satisfies the requirements of a particular application, it often fails to incorporate the perception of the general public towards places. In today's online landscape, some parts of this perception are captured by location-based social network platforms. Contributions to these platforms, such as check-ins and reviews, enable a bottom-up approach to place categorization based on the actual interaction between humans and places. In this short paper, we outline selected advantages of a hybrid approach, which combines top-down and bottom-up methods to enhance place type hierarchies.

1 Introduction

Place is a key concept that has been widely discussed in various research communities [4, 1]. Geographers consider *place* as *space* filled with human experience [6]. Place categorization systems are generally designed to associate place instances with categories, and play an important role in location-based services, geographic information retrieval, and place-based Geographic Information Systems (GIS). For example, hierarchical place categorization systems allow users to easily retrieve all places associated with a certain branch of said hierarchy, e.g., all instances belonging to the subtypes of the *sports facility* super type.

Many existing place categorization systems have been developed through a top-down approach in which a group of experts or users collaboratively determine the classification, labels, and hierarchal position. For example, both *Yelp*[1] and *Foursquare*[2] have developed their own categorization systems to classify Points Of Interest (POI). Schema.org was developed through a joint effort from multiple companies, such as Google, Microsoft, and Yahoo. While these categorization systems generally satisfy the requirements of particular software applications or communities, they do not necessarily mirror how humans behave towards places, how they interact with them, describe them, which activities they perform there, and so forth.

[1] https://www.yelp.com/developers/documentation/v2/all_category_list
[2] https://developer.foursquare.com/categorytree

CESIP 2015: Cognitive engineering for spatial information processes: From user interfaces to model-driven design. Workshop at COSIT 2015, October 12, 2015. Santa Fe, NM, USA

Location-based social networks (LBSN) capture some of the interactions between people and places. Two main categories of interaction are recorded by LBSN: *check-ins* and *unstructured text-based comments*. Check-in data provide information about people's behavior related to places, such as when and where a user is. Text-based comments contain information about the feelings and perceptions of people towards places. With their large user bases and millions of check-ins and comments added every months, LBSN data have become increasingly valuable. A bottom-up approach, in which LBSN data are mined, could help develop place categorization systems that better reflect how people conceptualize place types.

In this work, we discuss a hybrid approach which mines human-place interaction patterns from LBSN to enrich place categorization systems. The strength of a top-down classification lies in its declarative inclusion of theoretical, historical, or political considerations in forming place hierarchies, which would be difficult to discover by purely mining human behavior. On the other side, while the bottom-up approach is grounded in observed human behavior and thus groups place types that have similar patterns, it is challenging to clearly label the categories in a way that is easy to interpret. Thus, we combine the two approaches for their merits while mitigating their limitations.

2 Semantic Signatures of Place

From LBSN data, we can extract information to describe the interaction patterns between people and places. We refer to such multi-dimensional quantitative information as *semantic signatures*. The concept of *semantic signatures* originally grew out of data-driven semantics research; see [3] for details and references. In the following paragraphs, we explain three types of *bands* (i.e., *temporal*, *thematic*, and *geospatial*) that jointly form such signatures.

Temporal bands and signatures represent check-in behavior of LBSN users from the perspective of time. Many place types show clear temporal patterns in terms of people's check-ins [2]. Intuitively, from an hourly perspective, people tend to visit bakeries in the morning and pubs in the evenings. Similarly, and considering the days of the week, people are more likely to visit colleges during weekdays and nightclubs during the weekend. Studying millions of check-ins to hundreds of POI types reveals interesting hourly and weekly bands that jointly form signatures for place types. Thus, given a user check-in at a given time, it is possible to estimate the visited place type.

Thematic bands and signatures characterize how people perceive places and are constructed through observing natural language descriptions. With different environments and services, places may be described by people in different ways. Intuitively, people are more likely to describe a state park by terms such as hike, camping, waterfall, nature, and so forth, than by music, chips, or traffic. Consequently, topics extracted from user comments are indicative of place types. To capture such thematic signatures, topic modeling methods (such as Latent Dirichlet allocation (LDA)) can be used.

Geospatial bands and signatures represent the geographic distribution of place types. Based on a variety of factors, different place types show different forms of clustering in an environment. For example, restaurants tend to cluster together to attract customers, while hospitals are distributed evenly to provide better coverage. To describe these geospatial distribution patterns, measures such as Ripley's K can be used.

3 Hybrid Place Categorization

We leverage semantic signatures mined from the LBSN data and integrate them into an existing place schema. The result is a hybrid place categorization system which combines a top-down schema with bottom-up knowledge that partially reflects the actual interaction between human and places. Figure 1 shows an overview of our approach.

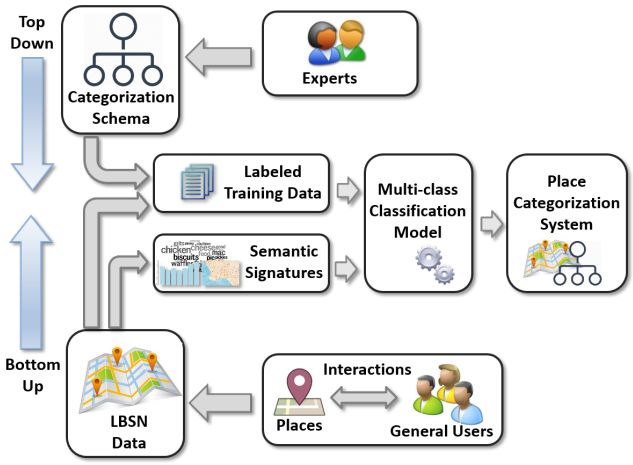

Fig. 1: An overview of the hybrid approach to place categorization.

There are three parts in this approach: 1) the top-down approach, 2) bottom-up approach, and 3) combining the two. The top-down approach employs a formally-defined place categorization schema, and reuses its structures and category labels. The bottom-up portion is based on the LBSN data generated through people's interactions with places. These two are then combined in several steps. First, the temporal, thematic, and geospatial bands are extracted from LBSN data, which quantifies human experiences towards places and which will become valuable *input features* for the later machine learning step. Meanwhile, a labeled training dataset is developed based on the LBSN data and the categorization schema. Such a training dataset trains a machine learning model with the LBSN places to which the top-down categories should belong. Manual labeling or semi-automatic methods can be used to develop this training

dataset. With the extracted semantic signatures (i.e., the combination of bands that uniquely identifies a place type) as well as the labeled training dataset, we then train a multi-class classification model for place categorization, since one place could be associated with multiple categories. Finally, we apply the trained classification model to all places in the LBSN dataset, and derive a place categorization system.

The proposed approach has been applied to an experiment in which a top-down POI classification schema from the Ordnance Survey[3] and a bottom-up LBSN dataset from Foursquare were combined to derive a hybrid place categorization system. A Web-based user interface[4] has been developed to visualize the categorization result. Details about this experiment can be found in our previous work [3]. The experimental result from our hybrid approach raises several interesting questions which are worthy of further investigation. In the following, we outline some of these questions.

- **Should single-class or multi-class categorization be used?** Many places serve multiple functions, and therefore could be classified into more than one category. For example, a *park* can be considered as an attraction for tourists, but can also be considered as a place for recreational activities such as boating. Similarly, a *soccer stadium* may be considered as a *sports facility* by the soccer players, but can also be considered as an *attraction* by the audience (which is, by far, the larger group). In essence, people can interact with and perceive the same places in different manners. In our hybrid approach, we adopted a multi-class categorization in order to capture such multiple perspectives. Figure 2 shows the categorization result derived from our experiment for two place types, *Park* and *Soccer Field*, respectively based on the Foursquare place type schema. In addition, our approach assigns a score to each category based on the probabilities returned from a SVM classifier trained using the previously described semantic signatures. These scores represent the quantities by which a specific place can be categorized. Compared with the crispy single-class categorization, the multi-class approach nevertheless carries the potential risk of confusing the end users from a cognitive perspective. In our case we used geographic scale to address this issue, i.e., on small scale, only 10 higher-order categories are shown and the multiple types only become visible when a user changes to a larger scale by zooming in.

- **Can category prototypes be modeled using a hybrid approach?** Prototypes have been suggested as one driver of categorization. For example, people can generally agree that "*a robin is a bird*", but are less likely to say "*a penguin is a bird*" (although people may say "technically, a penguin is a bird") [5]. In this example, *robin* has been considered as a prototypical case for the category of *bird*. In place categorization, such prototypical views

[3] http://www.ordnancesurvey.co.uk/business-and-government/products/points-of-interest.html

[4] Accessible at http://poipulse.com

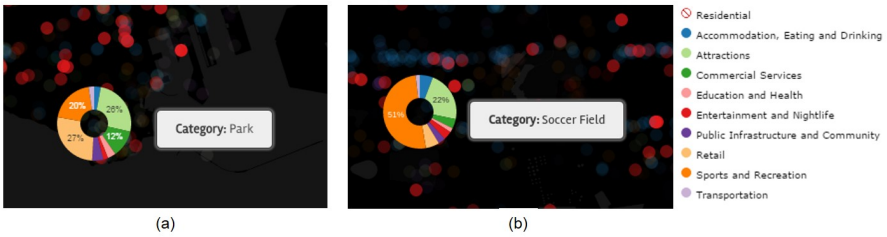

Fig. 2: The multi-class categorization: (a) Park; (b) Soccer Field.

may also exist, and a place categorization system that captures these views might better fit the perception of end users. We believe that such effects may also hold between different hierarchical levels of a place schema. With this idea in mind, we have examined the categorizations of two place types from our experiment: *American Restaurant* and *Burger Joint* (see Figure 3), and observed that *American Restaurant* shows a higher percentage of belonging to the category of *Accommodation, Eating and Drinking* than *Burger Joint*. Does this mean that the *American Restaurant* type is considered more representative for eating places than *Burger Joint*? Answering this questions requires further research on how to capture prototypical views into a place categorization system and how to validate the result.

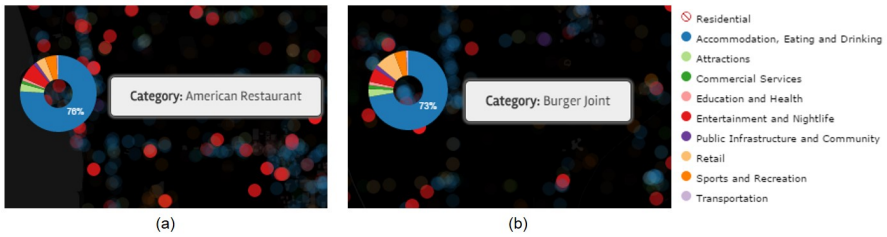

Fig. 3: Prototypical categories: (a) American Restaurant; (b) Burger Joint.

– **Can place schema reflect type similarities perceived by people?** A categorization system that takes into account human cognition should, at least, partially reflect the general understanding of people towards places. From a perspective of similarity, the places which have been considered as similar by people may also show similar result in terms of categorization. We examined the categorization of four place types: *Doctor's Office*, *Dentist's Office*, *Wine Bar*, and *Whiskey Bar* (Figure 4). Intuitively, *Doctor's Office* should be more similar to *Dentist's Office*, whereas *Wine Bar* should be more similar to *Whiskey Bar*. While such similarity has been observed from the categorization result of the four place types, confirming that a hybrid categorization system can reflect human cognition on place similarities requires capturing the actual perception of people. Crowdsouring approaches, e.g., Amazon Mechanical Turk, could be employed for this purpose.

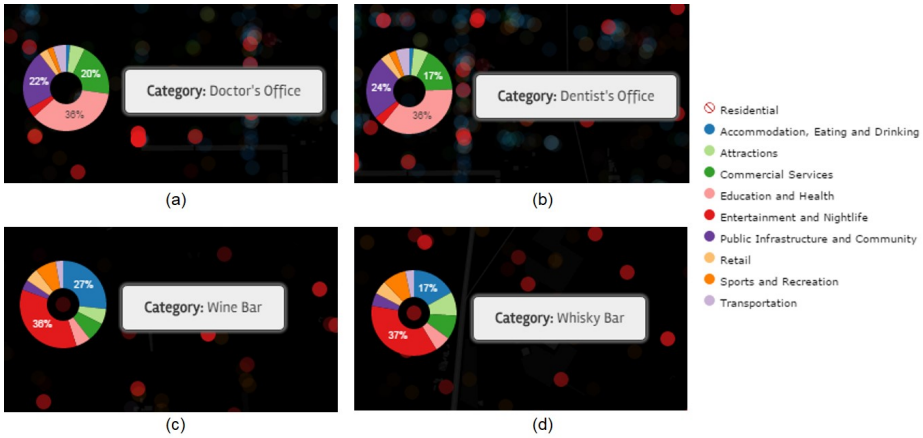

Fig. 4: The categorization similarity: (a) Doctor's Office; (b) Dentist's Office; (c) Wine Bar; (d) Whiskey Bar.

4 Conclusions

In this short paper, we outlined a workflow that mines human-place interaction patterns from LBSN data to enrich place categorization systems. Such a workflow combines top-down and bottom-up methods, and generates a hybrid system that partially incorporates human perception into place categorization. We have described temporal, thematic, and geospatial signatures, and we discussed some interesting questions that arise from the derived categorization result. The proposed hybrid approach also has its limitations that should be addressed in future work. For example, due to credibility issue, LBSN data might introduce noise and various biases into the categorization result.

References

1. Jordan, T., Raubal, M., Gartrell, B., Egenhofer, M.: An affordance-based model of place in GIS. In: Symposium on Spatial Data Handling. vol. 98, pp. 98–109 (1998)
2. McKenzie, G., Janowicz, K.: Where is also about time: A location-distortion model to improve reverse geocoding using behavior-driven temporal semantic signatures. Computers, Environment and Urban Systems 54, 1–13 (2015)
3. McKenzie, G., Janowicz, K., Gao, S., Yang, J.A., Hu, Y.: POI pulse: A multi-granular, semantic signatures-based approach for the interactive visualization of big geosocial data. Cartographica: The International Journal for Geographic Information and Geovisualization 50, 71–85 (2015)
4. Pred, A.: Place as historically contingent process: Structuration and the time-geography of becoming places. Annals of the Association of American Geographers 74(2), 279–297 (1984)
5. Roach, E., Lloyd, B.B.: Cognition and categorization. Hillsdale, New Jersey (1978)
6. Tuan, Y.F.: Space and place: The perspective of experience. U of Minnesota Press (1977)