Computers, Environment and Urban Systems 54 (2015) 336–346

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

How where is when? On the regional variability and resolution of geosocial temporal signatures for points of interest

Grant McKenzie [a,*], Krzysztof Janowicz [a], Song Gao [a], Li Gong [a,b]

[a] STKO Lab, Department of Geography, University of California, Santa Barbara, CA, USA
[b] Geosoft Lab, Institute of RS and GIS, Peking University, China

ARTICLE INFO

Article history:
Received 17 April 2015
Received in revised form 28 September 2015
Accepted 3 October 2015
Available online 23 October 2015

Keywords:
Semantic signature
Point of interest
Check-in
Geosocial
Time
Resolution

ABSTRACT

The temporal characteristics of human behavior with respect to points of interest (POI) differ significantly with place type. Intuitively, we are more likely to visit a restaurant during typical lunch and dinner times than at midnight. Aggregating geosocial check-ins of millions of users to the place type level leads to powerful temporal bands and signatures. In previous work these signatures have been used to estimate the place being visited based purely on the check-in time, to label uncategorized places based on their individual signature's similarity to a type signature, and to mine POI categories and their hierarchical structure from the bottom up. However, not all hours of the day and days of the week are equally indicative of the place type, i.e., the information gain between temporal bands that jointly form a place type signature differs. To give a concrete example, places can be more easily categorized into weekend and weekday places than into Monday and Tuesday places. Nonetheless, research on the regional variability of temporal signatures is lacking. Intuitively, one would assume that certain types of places are more prone to regional differences with respect to the temporal check-in behavior than others. This variability will impact the predictive power of the signatures and reduce the number of POI types that can be distinguished. In this work, we address the regional variability hypothesis by trying to prove that all place types are created equal with respect to their temporal signatures, i.e., temporal check-in behavior does not change across space. We reject this hypothesis by comparing the inter-signature similarity of 321 place types in three major cities in the USA (Los Angeles, New York, and Chicago). Next, we identify a common core of least varying place types and compare it against signatures extracted from the city of Shanghai, China for cross-culture comparison. Finally, we discuss the impact of our findings on POI categorization and the reliability of temporal signatures for check-in behavior in general.

1. Introduction

Points of interest (POI)[1] are inextricably linked to modern (mobile) search, recommender systems, location-based social networks, transportation studies, navigation and tourism systems, urban planning, predictive geo-analytics such as crime forecasting, and so forth. In terms of their computational representation, POI can be described and categorized in many different ways. Typical approaches are either based on features or functionality. The former describe POI based on attributes/properties such as price range, Wi-Fi availability, wheelchair access, ambience, noise level, room size, customer satisfaction, and so forth. Leaving pre-defined types such as restaurant, hotel, or national park, aside, POI can be grouped into ad-hoc categories (Barsalou, 1983) based on their common features such as "expensive places" or "attractions that offer wheelchair access". A functionality-centric view describes and categorizes POI based on what they afford, e.g., dining, travel, trade, or shelter (Jordan, Raubal, Gartrell, & Egenhofer, 1998; Winter & Freksa, 2014). While both approaches can be combined to account for their distinct strengths and weaknesses, they are typically realized in a schema-first manner in which features or functionalities are defined top-down and then populated with data (Glushko, 2014). An example of such a schema is shown in Fig. 1 which depicts properties defined for museum as well as the higher-level types from which these properties were inherited.

Alternatively, and assuming that meaning emerges from social structure (Gärdenfors, 1993), POI types can be described and categorized by aggregating how people behave towards places, e.g., when they visit them, what they say/write about them, and so forth. In addition to top-down schemata, such an approach reveals meaningful patterns suitable for a bottom-up, observations-first characterization

Abbreviations: API, application program interface; CHI, Chicago; DGC, difference in Gini coefficient; EMD, earth mover's distance; GINI, Gini coefficient; GPS, Global Positioning System; JSD, Jensen-Shannon divergence; KLD, Kullback–Leibler divergence; LA, Los Angeles; LBSN, location-based social network; MRR, mean reciprocal rank; NOLA, New Orleans; NYC, New York City; POI, point(s) of interest; RQ, research questions.

* Corresponding author.
E-mail address: grant.mckenzie@geog.ucsb.edu (G. McKenzie).

[1] We use the term point of interest here to keep in line with related work in research and industry and because these places are typically represented by point geometries. On the long term and due to the increase in richer geometric representations, place of interest seems to be the more appropriate name.

http://dx.doi.org/10.1016/j.compenvurbsys.2015.10.002
0198-9715/

Thing > Place > CivicStructure > Museum

A museum.

Property	Expected Type	Description
Properties from CivicStructure		
openingHours	Duration	The opening hours for a business. Opening hours can be specified as a weekly time range, starting with days, then times per day. Multiple days can be listed with commas ',' separating each day. Day or time ranges are specified using a hyphen '-'. - Days are specified using the following two-letter combinations: Mo, Tu, We, Th, Fr, Sa, Su. - Times are specified using 24:00 time. For example, 3pm is specified as 15:00. - Here is an example: <time itemprop="openingHours" datetime="Tu,Th 16:00-20:00">Tuesdays and Thursdays 4-8pm</time>. - If a business is open 7 days a week, then it can be specified as <time itemprop="openingHours" datetime="Mo-Su">Monday through Sunday, all day</time>.
Properties from Place		
address	PostalAddress	Physical address of the item.
aggregateRating	AggregateRating	The overall rating, based on a collection of reviews or ratings, of the item.
containedIn	Place	The basic containment relation between places.
event	Event	Upcoming or past event associated with this place, organization, or action. Supersedes events.
faxNumber	Text	The fax number.
geo	GeoCoordinates or GeoShape	The geo coordinates of the place.

Fig. 1. A fragment of the museum type from schema.org.

of POI (types). To give a few concrete examples, certain types of places are visited mostly during the weekends, while others are visited primarily during the workweek. Similarly, some types have their visitation peaks during the evenings while others peak during typical business hours from 9 am–5 pm. Even the lack of such distinct peaks is indicative (e.g. of major airports). Textual descriptions and other sources of observations can be used accordingly. For instance, mining latent topics from social media such as textual user reviews of places from Los Angeles reveals very characteristic Spanish-language topics (McKenzie, Janowicz, Gao, Yang, & Hu, 2015).

As an analogy to spectral signatures and bands in remote sensing, we have proposed semantic signatures that support the categorization of POI based on a multitude of spatial, temporal, and thematic bands (Janowicz, 2012). Simply put, in the domain of remote sensing, geographic entities on the surface of the Earth are classified via their unique reflection and absorption patterns in different wavelengths of electromagnetic energy called spectral bands (Schowengerdt, 2006). In some cases a particular band is sufficiently indicative to distinguish entity types (e.g., paved concrete from bare red brick), while in other cases a combination of multiple bands is required to form a unique spectral signature (e.g., deciduous and conifer trees cannot be distinguished via the visible light band alone).

Temporal signatures and bands are of particular interest as they are relatively easy to mine and at the same time are strongly indicative for a variety of POI types (Shaw, Shea, Sinha, & Hogue, 2013; Ye, Janowicz, Mülligann, & Lee, 2011). Consequently, they have been successfully used for the labeling of uncategorized places, for data cleansing and deduplication, for the construction of bottom-up POI hierarchies, for geolocation tasks such as estimating which place a user visited based on Global Positioning System (GPS) fixes, and further tasks that benefit from this kind of social sensing. Recognizing the role of time has also lead to new fields of study such as time-aware POI recommendation (Yuan, Cong, Ma, Sun, & Thalmann, 2013). Some POI types require additional (non-temporal) bands for their more fine-grained classifications (McKenzie et al., 2015). However, we will exclusively focus on temporal signatures in this work.

Interestingly, not all hours of the day and days of the week are equally indicative for the classification of POI types, i.e., the information gain of temporal bands differs. Intuitively, places can be more easily categorized into evening and morning place types (e.g., bars versus bakeries) than into early morning and late afternoon places. To further exploit the analogy to spectral signatures, it is interesting to note that the resolution of temporal bands is characterized and bound by human behavior. While hourly, daily, and seasonal bands have predictive power, second or minute-based bands do not (at least not for POI). This leads to the question of whether temporal signatures also have a platial, i.e., place-based, resolution.[2] Note that we use the term platial (or regional) instead of spatial here as the variation is expected to be non-linear. For example, San Diego, CA and Tijuana, Mexico are neighboring cities, yet we expect them to vary more with regards to the temporal signatures (due to cultural differences between Mexico and the United States) than San Diego, CA and San Francisco, CA which are over 700 km apart but within the same country. Conversely, non-spatial typically implies platial (regional) invariance.

Clearly, as temporal signatures are mined from human behavior, certain POI types will be affected by cultural differences. For instance, the peak dinner time for restaurants in Italy is around 8 pm while it is approximately 6 pm in the United States. We may even expect differences between the West and East Coasts of the U.S. In contrast, meaningful differences between the neighboring cities of New York, NY and Newark, NJ are less likely. Understanding such regional variations, their resolution, and magnitude, is important as they will effect the indicativeness of the signatures and thus their contribution to the

[2] We use the term platial here in reference to place, similar to how spatial refers to space.