Proceedings of the "Geographical and Cultural Aspects of Geo-Information: Issues and Solutions" AGILE 2019 Workshop, June 17th 2019, Limassol, Cyprus

Are Streets Indicative of Place Types?

Rui Zhu[1], ruizhu@geog.ucsb.edu

Grant McKenzie[2], grant.mckenzie@mcgill.ca

Krzysztof Janowicz[1], jano@geog.ucsb.edu

[1] STKO Lab, Department of Geography, UC Santa Barbara,
Ellison Hall, Isla Vista, CA 93117, USA

[2] Platial Analysis Lab, Department of Geography, McGill University,
805 Rue Sherbrooke Ouest, Montréal, QC H3A 0B9, Canada

Abstract

Places, and here more specifically Points of Interest (POI), can be described by characteristics such as their location, names, capacity, atmosphere, accessibility, reviews, opening hours, prices of services or products they offer, and so forth. Most importantly, however, places can be categorized into place types, e.g., *museum* or *fire station*. These types are best understood as proxies for a wide range of latent characteristics that we do not typically model explicitly in an information system. For example, we would expect to hear sirens nearby fire stations, find parking restrictions nearby, etc. Nonetheless, many modern (geographic) information retrieval systems treat place types as labels, i.e., atomic tokens. The same can be said about names of places and their locations, e.g., addresses. With regards to place (type) embedding, for instance, such a view ignores the cultural structure of these types, names, and addresses, thereby loosing important information. In this work we will show that addresses, here street types, are more than just atomic tokens. They are indicative of the types of places we can expect to encounter. Both the proximity to and suffix of streets are investigated to model the interaction between place types and streets, which are

1 Introduction

In contrast to typical spatial analysis, place-based (or platial) analysis focuses on characteristics that go beyond metric information about locations or geometries (Couclelis, 1992; Goodchild and Li, 2011; Merschdorf and Blaschke, 2018). Work towards place-based GIS and analysis is currently attracting significant attention in the GIScience community (Gao et al., 2013; Merschdorf and Blaschke, 2018; Blaschke et al., 2018; Westerholt et al., 2018), with multiple techniques being developed to analyze places from the perspective of the place hierarchies they form and what they afford to citizens. One family of these approaches focuses on crowdsourced textual descriptions of places, e.g., Adams and McKenzie (2013); Steiger et al. (2015); Siragusa and Leone (2018). These approaches are prevalent nowadays because they are capable of capturing moods, opinions, and experiences towards a place as well as many other latent characteristics such as atmosphere. Many place-based operations use these characteristics to derive a notion of place similarity (Medin et al., 1993) as an analogue to distance in space.

Places, specifically Points of Interest (POIs) in this work, and their types can be studied from a behavioural perspective by considering the thematic, temporal, and spatial patterns in which humans tend to interact with places of specific types. These patterns jointly form *semantic signatures*, i.e., the set of thematic, temporal, and spatial bands that uniquely characterize place types (Janowicz et al., 2019). Intuitively, places of type *museum* may be clustered in a specific district while *fire station* has to maximize coverage. Similarly, we would expect minimal activity around museums at night and early in the morning, but a more uniform distribution of temporal activity patterns at fire stations. Finally, news or reviews about museums are more likely to be about art, exhibitions, tickets, and so on than about rescues, emergencies, fires, and floods. Zhu et al. (2016), for instance, specifically investigated the role of spatial signature in modelling the semantics of place types through applying spatial statistics that quantify the spatial structures and interactions of places of given types.

Our work follows the aforementioned argumentation and further delves into one specific aspect, namely the spatial interaction between place types and addresses, here the street types (suffixes) associated with a place type. Put differently, street suffixes such as Avenue or Boulevard are not just atomic tokens, they carry meaning and reflect the types of places we can expect to encounter at a location. For example, airports are frequently located by main avenues that are close to highways while bookstores would be found on quieter and smaller streets. This paper introduces the proximity to and suffix of the closest street as two forms of spatial signature that describe the spatial interaction between places (and their types) and streets.

2 Related Work

Semantic signatures have been discussed considerably in the literature (Adams and Janowicz, 2015; McKenzie et al., 2015; Zhu et al., 2016; Miller et al., 2019). From a spatial perspective, Zhu et al. (2016) introduced 41 spatial statistics to describe the spatial structure of places and their interactions with other geographic features such as population, climate zones, and street networks. Though a preliminary street interaction analysis was included in this work, street networks were examined in combination with a number of other approaches and not explicitly investigated themselves. In addition, these previous studies focused on aligning feature types across different gazetteers in which most of the features are natural resources such as mountains, rivers, and valleys. In contrast, this work focuses on places in urban areas, where the street networks play a larger role in place and place type identity.

Rather than characterizing the semantics of place types, street networks have also been investigated to model urban functional zones (Yuan et al., 2015), to measure the complexity of urban forms (Boeing, 2018), to predict the traffic interactions of streets (Liu et al., 2017), and so on. However, these techniques only model the interaction of street within a street network, without the association with places being taken into account.

3 Data

Two Point of Interest (POI) datasets were accessed in Maryland, USA, namely Google Places [1] and Foursquare Venues[2]. The data were accessed in January of 2018 using the respective companies' application programming interfaces (API). While both datasets offer similar spatial coverage, each employs a different *place type* schema. These different schemata reflect the underlying purpose for which these datasets were generated. Google Places puts an emphasis on navigation and local business search while Foursquare focuses on local venue recommendations, ratings, and reviews. Given this difference in purpose, Foursquare venues are classified at a finer thematic resolution than Google and include place types such as *Mexican restaurant* and *Japanese restaurant.* In contrast, Google provides only one *restaurant* place type. In total, 383,545 Google places were accessed and categorized into 99 different place types and 132,429 Foursquare venues were accessed and grouped into 403 place types. We selected the Maryland Road Centerlines dataset[3] for the street network, which contains about 4,816 street centerlines for all public roadways in Maryland.

4 Methods

With our goal of differentiating and characterizing place types, we explore two forms of interactions between places and streets, (a) *Proximity* to the closest streets and (b) The *suffix* of the closest street. The closest street of a place in this work is defined as the centerline that contains the point having the smallest geographic distance to the target place.

4.1 Proximity to the Closest Street

The geographic distance between a place and the closest street plays a significant role in identifying the *type* of the place. Such a theory comes from the observation that *nature features,* for instance, are often isolated and further from streets than *cafés* and *restaurants*, place types that must be close to streets in order to attract business. Put differently, the type of a place is implicitly embedded in its interaction with a street network given that the relationship between places and streets differs based on the properties and affordances of the place type. For example, people interact with restaurants on a daily basis as they provide necessary sustenance and social interactions, whereas natural features such as forests, lakes, and parks do not necessarily serve a human-centric purpose.

Considering this, we identify “distance to closest street” as one measure on which to differentiate place types. A set of statistics can be extracted from the distribution of this measure. For example, Equation 1 quantifies the mean distance between a place type and its closest streets, where d_j represents the distance of a place j to its closest street, and N is the total number of places associated with the target place type. Additional distance statistics such as minimum (*min*), maximum (*max*), and standard deviation (*std*) are computed as well to aid in describing the interaction between places and streets.

$$s^{proximity} = \frac{\sum_{j=1}^{N} d_j}{N} \qquad (1)$$

Three Google Places types are shown in Table 1 along with the “distance to closest street” values that distinguish them from one another. As expected, the place type *restaurant* reports a relatively small mean distance to the closest street, while *natural feature* shows a relatively larger distance. These values align with our aforementioned street interaction notion. With the inclusion of additional measures, i.e., *min* and *max*, we can further characterize place types such that *stadium* in Maryland has a much greater minimum but smaller maximum distance to their closest (major) streets when compared to *restaurants*, even though their means are relatively similar. Note that distances are computed based on centroids as places in Google Places and Foursquare Venues are represented as points and that our dataset contains only public streets. This effects the distance between large scale features and streets, particularly in more rural areas.

Table 1: Example statistics for proximity to closest street. Values are based on a sample of > 50 POI per place type

Place Types	Distance to Closest Street (in meters)			
	Min	Max	Mean	Std
restaurant	0.01	15084.88	503.29	785.35
natural feature	8.90	14881.89	1423.70	2172.93
stadium	15.20	1870.40	468.42	387.72

4.2 Closest Street Suffix

In addition to street proximity, place types can also be characterized through other properties such as *street width*. This rational lies on the notion that place types such as *café* or *bakery* are more likely to be close to local, narrower single lane streets as opposed to place types such as *car dealerships*. Fortunately, thanks to the historical and cultural conventions, many properties of a street are implicitly encoded in its suffix[4]. For instance, one expects to find a short and narrow street categorized by the suffix *lane* in a local neighborhood. In contrast, the *parkway* suffix implies a wide, multi-lane street. Based on this, we propose to utilize the distribution of closest street suffix to identify and characterize place types.

Using the Maryland Street Centerlines dataset, we find that streets are categorized into 14 suffix types including streets (*RD*), turnpikes *(PIKE*), avenues (*AVE*), boulevards (*BLVD*), streets (*ST*), parkways (*PKWY*), connectors (*CONNECTOR*), circles (*CIR*), lanes (*LA*), ramps (*RAMP*), drives (*DR*), express ways (*EXPWY*), and no names (*NO NAME*). For each place type, we build a suffix distribution based on each place’s closest

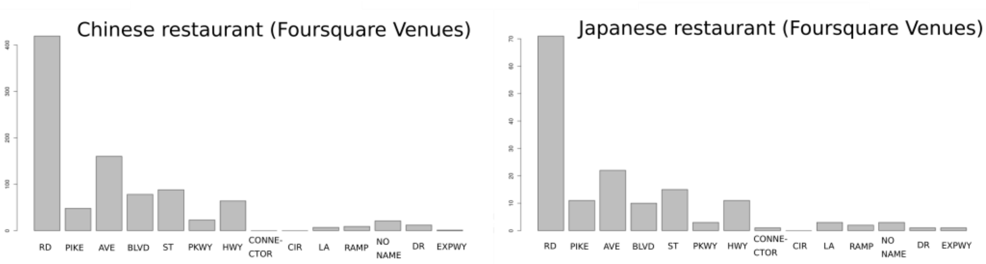
[1] https://cloud.google.com/maps-platform/places/

[2] https://developer.foursquare.com/

[3] http://data.imap.maryland.gov/datasets/

[4] https://pe.usps.com/text/pub28/28apc_002.htm

Figure 1: The distribution of street suffix for *Chinese restaurant* and *Japanese restaurant* from Foursquare Venues.



street and compare it with those produced from other place types. Figure 1 illustrates the distribution of *Chinese restaurant* and *Japanese restaurant* from Foursquare Venues. As expected, they share relatively similar patterns with the type *RD* occurring the most in both, with *ST* and *AVE* second and third, respectively. Moreover, we observe that these two types are barely located close to streets that belong to *CONNECTOR* or *CIR*.

In addition to characterizing *similar* place types, Figure 2 demonstrates how street suffix distribution is capable of distinguishing different place types. For example, the three types, *football stadium, department store*, and *bakery*, illustrate different patterns, despite the common domination of *RD* in their distributions. Specifically, *RAMP* has a prominent contribution in the pattern of *football stadium*, which we barely observe in other place types. Bakeries in general are located more close to *AVE* and *ST*, while department stores have a relatively equal likelihood of being near a *PIKE*, *AVE, BLVD, ST* or *HWY* .

In order to extract representative statistics from the distribution, Equation 2 is introduced, which measures the entropy of closest street suffix for each place type. In Equation 2, p_k represents the probability of observing the suffix k in a distribution of *M* different street suffixes (*M* equals 14 in this work). The larger the value, the more balanced (i.e., uncertain) the distribution. For example, *department store* shows a relatively larger entropy value (2.63) as compared to *aquarium* (1.78). This is due to the fact that department stores can be found near a wide range of street suffixes, while this is not the case for aquariums.

$$S^{suffix} = -\sum_{k=1}^{M} p_k \log p_k \qquad (2)$$

In summary, we propose five descriptive statistics to quantitatively describe the interaction between places and their closest streets. These five statistics are: the *mean, minimum, maximum* and *standard deviation of distance to closest streets,* and *the entropy of closest street suffix.*

5 Experiments and Discussions

Next we discuss exploratory experiments to verify the feasibility of the proposed street-based signatures on characterizing and differentiating place types. First, we used the street signatures to explore the relation of place types within one dataset (i.e., Google Places). Second, we use these measures to assess the similarity of place types across different datasets.

5.1 Experiments Within One Dataset

As a first step, we applied multidimensional scaling (MDS) to our place type dataset using the five-dimensional (i.e., *min, max, mean, std distance* to street and *entropy* of street suffix), street-based, spatial signatures computed from the interaction with closest streets. MDS transforms the relation of place types in high dimensional space into a lower one, by which we can visualize in a 2D map the perceived similarity between place types as reported by our new street-based spatial signatures. Using this method, the relationship between place types of

Figure 2: The distribution of street suffix for *football stadium, department store*, and *bakery* from Foursquare Venues.

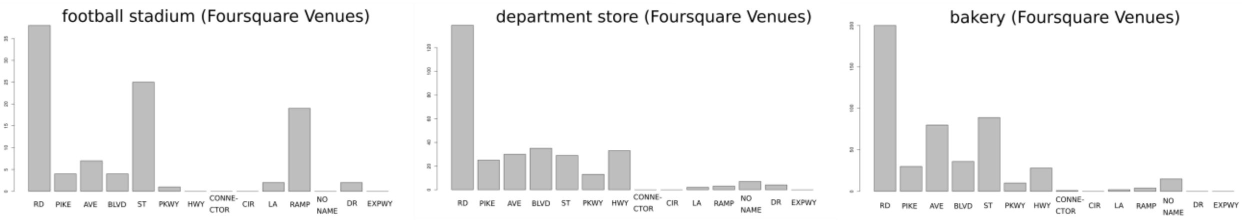
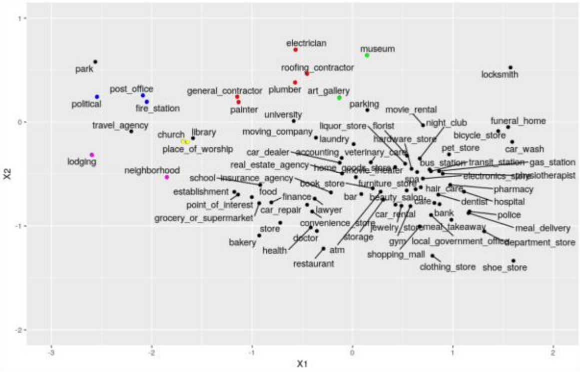


Figure 3: Multi-dimensional scaling map for place types of Google Places.



Google Places were visualized as a two-dimensional chart shown in Figure 3, with the scaling stress achieved at 6.46%. Note that the x1 and x2 axes of Figure 3 are transformed dimensions implying the greatest variation of the signatures without any practical interpretations.

From this initial experiment, we observe that the proposed signatures are capable of revealing similarities between place types. First, place types such as *electrician, roofing contractor*, *plumber, general contractor*, and *painter* form a noticeable group in this map (highlighted in red). Interestingly, they are all related to the construction trade. Second, *post office, political* and *fire station* cluster together providing public services (in blue). In addition, we observe that *museum* and *art gallery* are in close proximity in the figure (in green), both relevant to arts. Finally, the religion-related place types, *church* and *place of worship*, are near to each other (in yellow), indicating a high degree of similarity. Many other types of places exhibit similarity to one another, as can be seen in the figure.

In summary, statistics designed by leveraging the interaction with closest streets have the ability to uniquely characterize and cluster place types (in the Google Places dataset), similar to what most humans would intuitively perceive. Specifically, we demonstrate here that street-based signatures are capable of quantitatively characterizing place types with respect to religions, art, housing modeling and public services.

5.2 Experiments across Different Dataset

In addition to understanding place types within one dataset, this section concentrates on employing the proposed measures to compare place types across different datasets. We particularly investigated the distribution of closest street suffix with the goal of aligning place typing schemata between Google Places and Foursquare Venues. We applied Jensen-Shannon divergence (JSD) to compare the suffix distribution of place types between two datasets. Specifically, the pairwise JSD are computed and ranked, based on which of the top places are selected as candidate matches for a target place type.

Table 2 depicts examples of top matches from Foursquare Venues to Google Places. These examples show the merits of using the proposed signature in aligning place types. First of all, many place types are labeled as different tokens in different data sets, hence using traditional string matching (e.g., Levenshtien distance) would fail to align them. However, the interaction between place type and street suffix helps to address this issue. For instance, *amusement park* and *theme park* have different string names while their similar distributions of street suffix correctly align them, as shown in Table 2. On the other hand, even though two place types from different data sources share the same string names, they are by no means guaranteed to have the same semantics. Take the *hospital* from Google Places as an example, its top 5 matching candidates do not include the *hospital* from Foursquare Venues despite their exactly the same string names. On the contrary, *medical center* is ranked semantically closest to *hospital* in Google Places (with respect to the interaction with streets). As Figure 4 illustrates, hospitals in Foursquare Venues have a high probability of being located near a ST suffix, while both medical centers in Foursquare Venues and hospitals in Google Places are more likely to be found close to a RD suffix. However, it is still worth noting that street-based signatures do not work for all cases. As the third row of Table 2 illustrates, only applying proposed street-based signatures fails to align *post office* in Google Places to its correspondence in Foursquare Venues.

In summary, this section demonstrates that a "suffix-based" spatial signature is of use when aligning two different place type vocabularies. Further work, outside of this short paper, will investigate the limits of this approach.

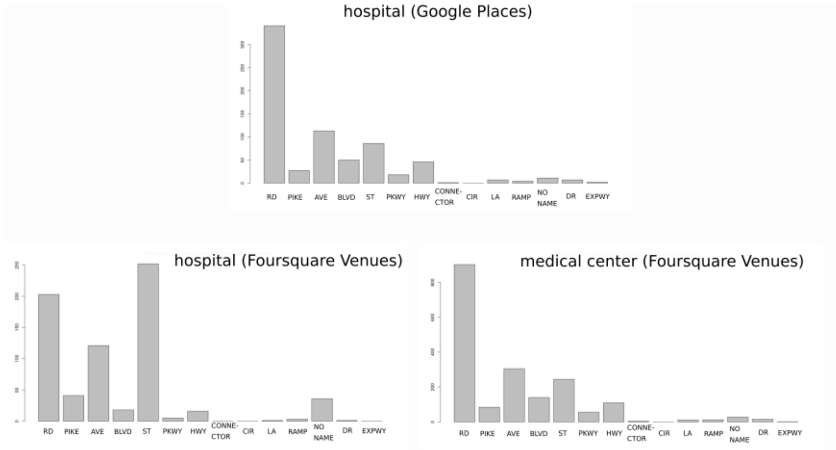
6 Conclusions and Future Work

This paper introduces a new aspect of spatial signature to quantify the semantics of place types based on the interaction with streets. Two types of statistics were proposed: the distance to the closest street, with the *mean, minimum, maximum* and *standard deviation* being selected as the specific statistics, and the distribution of the closest street suffix, with the *entropy* being extracted as the statistic. A series of experiments were conducted to illustrate the feasibility of proposed signatures in terms of understanding the semantics of place types both within one dataset and across different datasets. Thanks to the cultural implication behind both place types and street names, we discovered that the streets, specifically their geographic footprints and suffixes, are in fact indicative of place types. The interaction between places and streets is particularly beneficial

Table 2: Example of typing schema alignment from Foursquare Venues to Google Places. They are ranked by the Jensen-Shannon divergence on their street suffix distribution.

Place Type in Google Places	Top 5 Match in Foursquare Venues				
	1	2	3	4	5
amusement park	**theme park**	bike rental bike share	motel	lounge	market
hospital	**medical center**	salon barbershop	miscellaneous	drugstore pharmacy	laundry service
post office	fire station	city	bridge	flower shop	brewery

Figure 4: Street suffix distribution of hospital from Google places and hospital and medical center from Foursquare. Venues



to identify semantics that are relevant to public services, home improvement, art, health and so on.

However, our current work, as an initial exploration, has several limitations. First, the proposed street-based signatures were represented equally in the multidimensional scaling (MDS) map illustrated in Section 5.1, but such an assumption is not preferable in practice and assigning different weights to different signatures will be explored in future studies. Second, the MDS exploration only focused on a small subset of place types and the analysis was rather subjective and qualitative. Future studies will extend the work to the whole set of place types, and new approaches, such as clustering algorithms, will be introduced to quantitatively investigate the semantic relevance of place types using street-based signatures. Furthermore, we only showed several examples of using proposed signatures to align place types across different data sources, more sophisticated models and systematic evaluations will be investigated in future studies. In practice, the proposed signature has the potential to address practical challenges such as co-reference resolution, open geospatial data cleaning, and place disambiguation, which are the future directions of this work as well. Last but not least, we plan to apply the approach across different cities and countries as a new means to compare and understand the culture implication on places.

References

Adams, B. and Janowicz, K. (2015) Thematic signatures for cleansing and enriching place-related linked data. *International Journal of Geographical Information Science* 29 (4), 556-579.

Adams, B. and McKenzie, G., (2013) Inferring thematic places from spatially referenced natural language descriptions. In *Crowdsourcing Geographic Knowledge* (pp. 201-221). Springer, Dordrecht.

Blaschke, T., Merschdorf, H., Cabrera-Barona, P., Gao, S., Papadakis, E. and Kovacs-Györi, A. (2018) Place versus Space: From Points, Lines and Polygons in GIS to Place-Based Representations Reflecting Language and Culture. *ISPRS International Journal of Geo-Information*, *7*(11), p.452.

Boeing, G. (2018) Measuring the complexity of urban form and design. *URBAN DESIGN International*, *23*(4), pp.281-292.

Couclelis, H. (1992) Location, place, region, and space. In: Abler, R. F., Marcus, M. G., Olson, J. M. (Eds.), *Geography's Inner Worlds*. Rutgers University Press, New Jersey, pp. 215-233.

Gao, S., Janowicz, K., McKenzie, G., Li, L. (2013) Towards platial joins and buffers in place-based gis. In: *Comp@ Sigspatial*. pp. 42-49.

Goodchild, M., Li, L. (2011) Formalizing space and place. In: *CIST2011-Fonder les sciences du territoire*. pp. 177-183.

Janowicz, K., McKenzie, G., Hu, Y., Zhu, R., Gao, S. (2019) Using semantic signatures for social sensing in urban points of interest with representation learning. *Computers, Environment and Urban Systems* 75, 146–16.

Liu, K., Gao, S., Qiu, P., Liu, X., Yan, B. and Lu, F., (2017) Road2vec: Measuring traffic interactions in urban road system from massive travel routes. *ISPRS International Journal of Geo-Information*, 6(11), p.321.

McKenzie, G., Janowicz, K., Gao, S., Gong, L. (2015) How where is when? on the regional variability and resolution of geosocial temporal signatures for points of interest. *Computers, Environment and Urban Systems* 54, 336–346.

Medin, D. L., Goldstone, R. L., Gentner, D. (1993) Respects for similarity. *Psychological Review* 100 (2), 254–278.

Merschdorf, H., Blaschke, T. (2018) Revisiting the role of place in geographic information science. I*SPRS International Journal of Geo-Information* 7 (9), 364.

Miller, H.J., Jaegal, Y. and Raubal, M., (2019) Measuring the Geometric and Semantic Similarity of Space–Time Prisms Using Temporal Signatures. *Annals of the American Association of Geographers*, pp.1-24.

Steiger, E., Westerholt, R., Resch, B., Zipf, A. (2015) Twitter as an indica- tor for whereabouts of people? correlating twitter with uk census data. *Computers, Environment and Urban Systems* 54, 255–265.

Westerholt, R., Gr¨obe, M., Zipf, A., Burghardt, D. (2018) Towards the statistical analysis and visualization of places (short paper). In: *10th International Conference on Geographic Information Science*, GIScience 2018, August 28-31, 2018, Melbourne, Australia.

Yuan, N. J., Zheng, Y., Xie, X., Wang, Y., Zheng, K., Xiong, H. (2015) Discovering urban functional zones using latent activity trajectories. *IEEE Transactions on Knowledge and Data Engineering* 27 (3), 712–725.

Zhu, R., Hu, Y., Janowicz, K., McKenzie, G., (2016), Spatial signatures for geographic feature types: Examining gazetteer ontologies using spatial statis- tics. *Transactions in GIS* 20 (3), 333–355.