Towards Learning of Spatial Triad from Online Text

Jina Kim
kim01479@umn.edu
University of Minnesota

Yao-Yi Chiang
yaoyi@umn.edu
University of Minnesota

ABSTRACT

The Spatial Triad model provides a framework for studying human interactions and experiences with the environment, which helps to improve human well-being and quality of life. Typical studies that use this framework require time-consuming and expensive surveys. This paper presents a simple yet effective approach to learning what humans feel, think, and see about their surroundings from easily accessible online text descriptions (e.g., descriptions of listings on real estate or travel blogs). The proposed technologies learn meaningful document and locality representations in a unified representation space, capturing important concepts shared among documents within the same locality. The proposed approach outperforms the existing method in finding associated localities in online text and shows exciting insights into locality similarity using the learned representations.

CCS CONCEPTS

• **Information systems** → *Document representation* .

KEYWORDS

spatial triad, online text, representation learning

ACM Reference Format:
Jina Kim and Yao-Yi Chiang. 2023. Towards Learning of Spatial Triad from Online Text. In *The 31st ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL '23), November 13–16, 2023, Hamburg, Germany* , 2 pages. https://doi.org/10.1145/3589132.3628372

1 INTRODUCTION

Understanding how humans feel, think, and see their surroundings (i.e., Spatial Triad [10]) is important, for example, for policymakers and urban planners to design a better space to improve human well-being and quality of life [3]. Traditional methods rely on extensive surveys or interviews over several years to capture the required information (e.g., life quality measures) [1, 7]. Recent work leverages distinct elements from satellite imagery and street-level photos to capture aspects of spaces (e.g., [6]). However, satellite imagery cannot always capture diverse physical environments (e.g., building facets), and street view images are unavailable for many places (e.g., India). Furthermore, as human perceptions encompass more than just the visual sense, image data may lack perceptions of culture, activities, and interactions with the surroundings, to name a few (i.e., the conceived and lived spaces in the Spatial Triad).

SIGSPATIAL '23, November 13–16, 2023, Hamburg, Germany

ACM ISBN 979-8-4007-0168-9/23/11…$15.00
https://doi.org/10.1145/3589132.3628372

In contrast, many types of online text data describe the physical space and how humans interact with the space. They are easily accessible and have excellent spatial coverage (e.g., Redfin listings, Figure 1). A recent study demonstrates an approach to generate urban-characteristic scores (e.g., transportation, safety) from Airbnb guest reviews [8]. They first detect a sentiment score for each review in a neighborhood. Then, they extract important words, such as nominal objects, from each review and assign a review to an indicator using rule-based methods based on the extracted words. Finally, they calculate the indicator score for a neighborhood by averaging the sentiment scores of all reviews assigned to an indicator. They show that the indicator scores generated are highly correlated with several objective measures of urban characteristics in a neighborhood. However, the proposed approach requires human-defined rule-based methods and could introduce subjectivity. Also, existing language-model-based approaches to identifying shared concepts from documents, such as supervised topic modeling (e.g., BERTopic [9]), often work on a few numbers of topics (compared to large numbers of localities) and deal with easily distinguishable topics (e.g., news topics - sports, business) in contrast to subtle differences in locality characteristics.

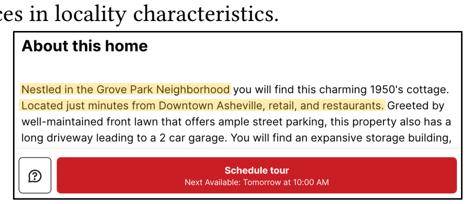

Figure 1: Example listing description on Redfin.

This paper proposes an automatic method for learning locality characteristics directly from online text. The input includes online text data and their geocoordinates (e.g., Redfin listing descriptions and property locations). The goal is to learn shared concepts of documents associated with the same predefined locality (e.g., cities, neighborhoods), allowing the observation of interesting relationships among learned locality representations with similar or dissimilar concepts. In addition, such locality representations will enable downstream tasks that require understanding locality characteristics, such as predicting greenness indices of localities. The challenges include how to encode text data in the same locality to distill relevant locality characteristics and leverage their geocoordinates to guide the generation of meaningful locality representations. The following sections present our proposed approach to addressing these challenges and experiment results.

2 LEARNING LOCALITY CHARACTERISTICS FROM ONLINE TEXT

The proposed method starts by constructing locality representations and then learns a joint space that binds document representations of the same locality to distal shared information using contrastive learning techniques.

Constructing Locality Representation. Each input data resource consists of a set of documents and their corresponding geocoordinates. We augment the geocoordinates of each document with a full place phrase of the corresponding locality. For example, for city-level localities, if a document is within the boundary of Wailuku, Hawaii, the place phrase associated with the document is 'Wailuku, Hawaii, US'. Using place phrases instead of raw geocoordinates enables learning similar locality representations within the same geographic hierarchies (e.g., same states/countries), which are highly likely to share similar concepts.

Learning Document and Locality Representations. Given input documents and their locality place phrases, we employ BERT [2] to encode them separately. The document encoder aims to generate contextualized representations, while the location encoder aims to generate similar representations if the localities are in the same geographic hierarchies. Note that other deep networks can be used as encoders if each network can address the encoding goals.

As we aim to learn shared concepts across documents within the same locality, we employ InfoNCE [5] as a pretraining loss. We bind document representations to their corresponding location representation, inspired by existing approaches using contrastive learning on pairs of related data (e.g., image-audio/image-text in ImageBind [4]). Such contrastive learning approaches have demonstrated superior performance in extracting relations between data without explicit supervision. Given document-locality pairs in a batch, we pull the associated locality representations (i.e., positives) to the corresponding document representations while pushing the non-associated locality representations (i.e., negatives) far apart. In this way, we can generate document representations that focus on shared contexts among other documents within the same localities and locality representations that distill shared concepts of the associated documents.

3 EXPERIMENTAL RESULTS AND DISCUSSION

The learned document representations contain concepts shared within a certain locality, disregarding other information. Thus, the learned document representations are close together if the documents are located in the same locality. We construct a Redfin data resource for the experiments. Redfin, a residential real estate brokerage, provides a detailed property description in the 'About this home' section with the geocoordinates of the property. After preprocessing text descriptions in English and constructing document-locality pairs, we used 37,355 online text descriptions from 27 cities and 103 LA neighborhoods. We evaluate the ability of the learned document representations by predicting their associated locality using the closest locality representation from each document representation. As a baseline approach, we employ BERTopic [9] to classify the localities in a supervised way. Table 1 shows the experimental results where our approach significantly outperforms BERTopic in average top 1 accuracy from a 10-fold cross-validation using the Redfin data resource.

In addition, we show that the proposed method can identify cities with similar characteristics by performing hierarchical clustering on the learned locality representations of a set of cities. An interesting finding is that the learned embeddings show that New York and Boston in the US share similar locality characteristics, which aligns with the expert-curated data from 'What are your city's twins? article. [1] Also, Figure 2 shows the experimental results at the neighborhood level in the city of Los Angeles, in which Silver Lake and Eco Park have similar locality characteristics, which are the neighbors known as 'Hipster L.A.' with unique characteristics. [2]

Locality	Model	Avg. Top 1 Accuracy
City	BERTopic [9]	63.14%
	Our Approach	83.93%

Table 1: Experimental results on identifying locality of document representations using Redfin data resource.

Figure 2: Sampled hierarchical clustering results using learned locality representations of LA neighborhoods.

We plan to evaluate the usefulness of locality representations by performing supervised tasks that require understanding localities (e.g., predicting the greenness indices). As we show the potentiality of capturing locality characteristics from online text descriptions in a simple but effective way, this method will help build a robust foundation model that comprehensively defines the Spatial Triad of localities by incorporating other modalities of data (e.g., street-level imagery for perceived space of the Spatial Triad).

ACKNOWLEDGMENTS

We thank Filip Biljecki for providing insight into the data resources and Zekun Li for the constructive discussion on contrastive learning. The work is supported in part by the LASI-DAD grant 5 U01 AG064948-05.

REFERENCES

[1] Bieling et al. 2014. Linkages between landscapes and human well-being: An empirical exploration with short interviews. *Ecological Economics* 105 (2014), 19–30.
[2] Devlin et al. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proc. of NAACL '19*. ACL, 4171–4186.
[3] Goonewardena et al. 2008. *Space, difference, everyday life: reading Henri Lefebvre.* Routledge.
[4] Girdhar et al. 2023. ImageBind: One Embedding Space To Bind Them All. In *Proc. of CVPR '23*. 15180–15190.
[5] Oord et al. 2018. Representation Learning with Contrastive Predictive Coding. arXiv:1807.03748 [cs.CL]
[6] Ramírez et al. 2021. Measuring heterogeneous perception of urban space with massive data and machine learning: An application to safety. *Landscape and Urban Planning* 208 (2021), 104002.
[7] Senlier et al. 2009. A perception survey for the evaluation of urban quality of life in Kocaeli and a comparison of the life satisfaction with the European cities. *Social indicators research* 94 (2009), 213–226.
[8] Wang et al. 2023. Insights in a city through the eyes of Airbnb reviews: Sensing urban characteristics from homestay guest experiences. *Cities* 140 (2023), 104399.
[9] Maarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv:2203.05794 [cs.CL]
[10] Eugene J McCann. 1999. Race, protest, and public space: Contextualizing Lefebvre in the US city. *Antipode* 31, 2 (1999), 163–184.

[1] nytimes.com/interactive/2018/04/03/upshot/what-is-your-citys-twin.html
[2] sandiegoreader.com/news/2017/dec/01/travel-hipster-los-feliz-echo-park-silver-lake