University of California
Santa Barbara

Geographic Knowledge Graph Summarization

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Geography

by

Bo Yan

Committee in charge:

Professor Krzysztof Janowicz, Chair
Professor Werner Kuhn
Professor Konstadinos Goulias

June 2019

The Dissertation of Bo Yan is approved.

Professor Werner Kuhn

Professor Konstadinos Goulias

Professor Krzysztof Janowicz, Committee Chair

May 2019

Geographic Knowledge Graph Summarization

by

Bo Yan

To my parents for their love and support

Acknowledgements

There are many people I am grateful for during the course of my life as a Ph.D. student at UC Santa Barbara. This incredible journey would not have been possible without them. To be honest, I am a bit emotional while writing this acknowledgement.

First and foremost, I would like to thank my advisor, Krzysztof Janowicz, who for the past six years has spent a great amount of time and effort chiseling me to who I am today as a competent researcher. This dissertation is the result of a close collaboration with him, who has given me invaluable feedback, has been constantly inspiring me to pursue new ideas, and has gone through each and every paper I wrote word by word to make sure I am able to convey the research clearly. There are a lot I learned from Jano. His meticulousness towards details has inspired me to dive deep into research questions and ideas. Working with him has made me realize the importance of communication and thinking outside the box. His Java class is one of the best I have taken. I still remember the first time I met him on Skype. I could feel his passion for research and I greatly appreciated his interest and patience when I talked about my immature research study through Skype at that time. Not knowing what life had in store for me, I took a leap of faith, came to this country that I barely knew, and joined his STKO lab. As I reminisce on the ups and downs over my past few years, I realize it was one of the best decisions I have made and I am very grateful for him for offering me this opportunity. For the past six years, I have witnessed the rapid growth of our lab and the amazing achievements of our lab members under his leadership.

Jano is not only a great mentor academically, but also a role model and a life coach. He has always told me that I should be a good human being first and a good researcher second. He is kind, caring, and supportive of every decision I made. He always has a positive outlook and believes that every cloud has a silver lining. During the dark times

of my Ph.D. life, he always believed in me and made me feel confident. His inquisitive mind has inspired me to learn to see the world in a grain of sand. Although twenty-two years of my school life have officially come to an end, the school of life continues. As I open my new chapter and commence an unknown journey, I will always remember what I learned from him and remain a scholar and a researcher at heart.

I have been very fortunate to have Werner Kuhn and Konstadinos Goulias in my committee. They have been giving me a lot of useful advice for my Ph.D. exams and this dissertation. Discussing with them helps me understand my research ideas and methods from different perspectives. They challenged my proposals and helped me distill the essence of my research. I really appreciate their comments and advice.

I would like to thank Gengchen Mai. We have been working together very closely since he joined the lab. I really enjoyed all the thought-provoking discussions and the binge-writing sessions before deadlines with him. His relentless effort in figuring out every detail in a paper, unbelievable productivity in research, and amazing ability to maintain a balanced life are beyond me. He is smart but humble. I am very fortunate to have the opportunity to work with him. Rui Zhu's passion for research and pursuit for raising the awareness of the fundamental research questions are impressive. I am glad that I can learn from his perseverance about his research topic and I am confident that he will see the light at the end of the tunnel soon. Thank you, Blake Regalia, for being my lunch buddy on campus although we have quite different tastes in food. I would like to thank Yingjie Hu who gave me a lot of comments and suggestions on my master's thesis and helped me on a number of research papers. I submitted my first paper with Yingjie's help. Song Gao (Super Song) was an energy source in our lab. He is always very active and energetic. Song helped me with my first co-authored paper. While my trip to Europe was a disaster during my first summer at UC Santa Barbara, I had a lot of fun with him and Tiange in Greece. Thank you, Song. In addition, I would like to thank other past and present STKO members who made my graduate school life more colorful: Yiting Ju, Grant McKenzie, Jiue-An (Jay) Yang, Ling Cai, Meilin Shi, Anagha Uppal, Zachary Canter, Kang Liu, Li Gong, Dara Seidl, Dingwen Zhang, Su Yeon Han, Amin Abdalla, and Marcela Suarez.

During my Ph.D. study, I squeezed some time for three summer internships at LinkedIn, HERE technologies, and ESRI. These internships are a great learning experience and had a lot of impact on my research trajectory. I would like to thank Matthew Walker, Shipeng Yu, Antonio Haro, Craig Barnes, and Sathya Prasad for their support and mentorship during my internships.

Finally and most importantly, I am grateful for my parents for their unconditional love and support. They have supported me through many challenges and genuinely cared for me. I feel forever in debt to them. I hope I can at least make them proud of what I have accomplished so far. Thank you, Mom and Dad.

Curriculum Vitæ
Bo Yan

Education

2019	Ph.D. in Geography, University of California, Santa Barbara.
2016	M.A. in Geography, University of California, Santa Barbara.
2013	B.S. in Geographic Information Science, Wuhan University.

Professional Experience

09/2013 – 06/2019	Research Assistant, University of California, Santa Barbara.
06/2018 – 09/2018	Machine Learning Research Intern, LinkedIn.
07/2017 – 09/2017	Research Engineer Intern, HERE Technologies.
06/2016 – 09/2016	Research Intern, ESRI.

Publications

Refereed Journal Articles, Conference Proceedings & Book Chapters

Yan, B., Janowicz, K., Mai, G., and Zhu, R. (2019): A Spatially-Explicit Reinforcement Learning Model for Geographic Knowledge Graph Summarization. *Transactions in GIS.*

Mai, G., Janowicz, K., and **Yan, B.** (2019): Deeply Integrating Linked Data with Geographic Information Systems. *Transactions in GIS.*

Mai, G., **Yan, B.**, Janowicz, K., and Zhu, R. (2019): Relaxing Unanswerable Geographic Questions Using A Spatially Explicit Knowledge Graph Embedding Model. *The 22nd AGILE International Conference on Geographic Information Science.* Jun. 17-10, Limassol, Cyprus.

Yan, B., Mai, G., Hu, Y., and Janowicz, K. (2018): Harnessing Heterogeneous Big Geospatial Data. In: M. Werner and Y. Chiang (Eds). *Big Geospatial Data (Under review).*

Mai, G., Janowicz, K., and **Yan, B.** (2018): Support and Centrality: Learning Weights for Knowledge Graph Embedding Models. *The 21st International Conference on Knowledge Engineering and Knowledge Management.* Nov. 12-16, Nancy, France.

Yan, B., Janowicz, K., Mai, G., and Zhu, R. (2018): xNet+SC: Classifying Places Based on Images by Incorporating Spatial Contexts. *The 10th International Conference on Geographic Information Science.* Aug. 28-31, Melbourne, Australia.

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), 321.

Yan, B., Janowicz, K., Mai, G., and Gao, S. (2017): From ITDL to Place2Vec – Reasoning About Place Type Similarity and Relatedness by Learning Embeddings From Augmented Spatial Contexts. *ACM SIGSPATIAL 2017*.

Gao, S., Janowicz, K., Montello, D., Hu, Y., Yang, J., McKenzie, G., Ju, Y., Gong, L., Adams, B., and **Yan, B.** (2017): A data-synthesis-driven method for detecting and extracting vague cognitive regions. *International Journal of Geographical Information Science*. 31 (6), 1245-1271, Taylor & Francis.

Gao, S., **Yan, B.**, Gong, L., Regalia, B., Ju, Y., and Hu. Y (2017): Uncovering the digital divide and the physical divide in Senegal using mobile phone data. *Advances in geocomputation*. 143-151.

Ju, Y., Adams, B., Janowicz, K., Hu, Y., **Yan, B.**, and McKenzie, G. (2016): Things and strings: improving place name disambiguation from short texts by combining entity co-occurrence with topic modeling. *International Conference on Knowledge Engineering and Knowledge Management*. Nov. 19-23, Bologna, Italy.

Refereed Short Papers & Workshop Proceedings

Janowicz, K., **Yan, B.**, Regalia, B., Zhu, R., and Mai, G. (2018): Debiasing Knowledge Graphs: Why Female Presidents are not like Female Popes. *International Semantic Web Conference (Posters & Demonstrations, Industry and Blue Sky Ideas Tracks)*. Oct. 8-12, Asilomar conference grounds, monterey, California, USA.

Mai, G., Janowicz, K., and **Yan, B.** (2018): Combining text embedding and knowledge graph embedding techniques for academic search engines. *The 4th Workshop on Semantic Deep Learning (SemDeep-4) at the International Semantic Web Conference*. Oct. 8-12, Asilomar conference grounds, monterey, California, USA.

Mai, G., Janowicz, K., Hu, Y., Gao, S., Zhu, R., **Yan, B.**, McKenzie, G., Uppal, A., and Regalia, B. (2018): Collections of Points of Interest: How to Name Them and Why it Matters. *Spatial big data and machine learning Workshop at GIScience 2018*. Aug. 28-31, Melbourne, Australia.

Mai, G., Janowicz, K., Prasad, S., and **Yan, B.** (2018): Visualizing The Semantic Similarity of Geographic Features. *The 21th AGILE International Conference on Geographic Information Science (short papers, posters and poster abstracts)*. Jun. 17-10, Limassol, Cyprus.

Gao, S., and **Yan, B.** (2018): Place2Vec: Visualizing and Reasoning About Place Type Similarity and Relatedness by Learning Context Embeddings. *The 14th International Conference on Location Based Services (Short Paper)*.

Zhu, R., Janowicz, K., **Yan, B.**, and Hu, Y. (2016): Which kobani? a case study on the role of spatial statistics and semantics for coreference resolution across gazetteers. *International Conference on GIScience (Short Paper)*. Sept. 2016, Montreal, Canada.

Yan, B., Janowicz, K., and Hu, Y. (2016): A Data-Driven Approach for Detecting and Quantifying Modeling Biases in Geo-Ontologies by Using a Discrepancy Index. *International Conference on GIScience (Short Paper).* Sept. 2016, Montreal, Canada.

Yan, B., Hu, Y., Kuczenski, B., Janowicz, K., Ballatore, A., Krisnadhi, A., Ju, Y., Hitzler, P., Suh, S., and Ingwersen, W. (2015): An Ontology For Specifying Spatiotemporal Scopes in Life Cycle Assessment. *Diversity++ Workshop at the International Semantic Web Conference.* Oct. 11-15, Bethlehem, Pennsylvania, USA.

Gao, S., Yang, J., **Yan, B.**, Hu, Y., Janowicz, K., and McKenzie, G. (2014): Detecting Origin-Destination Mobility Flows From Geotagged Tweets in Greater Los Angeles Area. *The 8th International Conference on Geographic Information Science (GIScience '14) Short Paper.* Vienna, Austria.

Gao, S., Yang, J., Janowicz, K., Hu, Y., and **Yan, B.** (2014): TrajAnalyst: Matching Data to Trajectory Analysis Modules via a Conceptual Framework. *The 8th International Conference on Geographic Information Science (GIScience '14) Short Paper.* Vienna, Austria.

Awards

2018	Jack & Laura Dangermond Graduate Research Fellowship.
2017	First Place in AAG GI Science and Systems Specialty Group Student Paper Competition.
2016	First Place in ESRI Intern Hackathon.

Abstract

Geographic Knowledge Graph Summarization

by

Bo Yan

Geographic knowledge graphs play a significant role in the geospatial semantics paradigm for fulfilling the interoperability, the accessibility, and the conceptualization demands in geographic information science. However, due to the immense quantity of information accompanying and the enormous diversity of geographic knowledge graphs, there are many challenges that hinder the applicability and mass adoption of such useful structured knowledge. In order to tackle these challenges, this dissertation focuses on devising ways in which geographic knowledge graphs can be digested and summarized. Such a summarization task, on the one hand lifts the burden of information overload for end users, on the other hand facilitates the reduction of data storage, speeds up queries, and helps eliminate noise. The main contribution of this dissertation is that it introduces the general concept of geospatial inductive bias and explains different ways this idea can be used in the geographic knowledge graph summarization task. By decomposing the task into separate but related components, this dissertation is based upon three peer-reviewed articles (Chapter 3, Chapter 4, and Chapter 5) which focus on the hierarchical place type structure, multimedia leaf nodes, and general relation and entity components respectively. Chapter 6 presents a spatial knowledge map interface to illustrate the effectiveness of summarizing geographic knowledge graphs. Throughout the dissertation, top-down knowledge engineering and bottom-up knowledge learning methods are integrated. We hope this dissertation would promote the awareness of this fascinating area and motivate researchers to investigate related questions.

Contents

Chapter 1

Introduction

This chapter provides a general introduction to the dissertation. It starts with the background and motivation for research in summarizing geographic knowledge graphs. It first introduces the interoperability demand, the accessibility demand, and the conceptualization demand that have given rise to the geospatial semantics paradigm in GIScience research. Then it explains the practical need for summarizing geographic knowledge graphs in this paradigm. Three concrete research questions are raised accordingly to constructively tackle the big question step by step. A dissertation synopsis is provided in the end to briefly outline the structure of this dissertation.

1.1 Background

Recent years have witnessed an increasing number of research endeavors in geospatial semantics [1, 2] as Geographic Information Science (GIScience) has entered the new paradigm that demands efficient processing of a large amount of heterogeneous geographic data, more accessible interfaces for the general audience, and a better conceptualization model to mitigate the inherent vagueness in geographic phenomenon. Such demands give rise to research studies that focus on geospatial ontologies and geographic knowledge graphs as they are the embodiment of the broader idea of geospatial semantics and semantic interoperability. In order to understand their relationship and provide a background of this research study, let us analyze these three major demands.

Interoperability Demand As a scientific discipline to develop and utilize theories, methods, technology, and data for understanding geographic processes, relationships, and patterns [3], GIScience has always been at the forefront of adopting and studying different kinds of data structures and data models. Traditionally, well-structured relational data models are the first choice for research as well as applications in the field because such data structures and models are well-studied and support efficient geospatial operations (such as spatial range queries or topological queries). As the *science of where* permeates almost every aspect of our daily life — from getting around the neighborhood to traveling in distant destinations, challenges arise because the ubiquity of geographic information demands a more flexible way of handling geographic data and the restriction of relational data models sets up a barrier to better solve real-world geographic problems. In addition to the well-structured data formats that researchers and practitioners alike have been favoring for a long time, geographic data comes in a variety of flavors as different data sources follow different protocols for capturing, storing, and transmitting data. For example, satellite imageries contain a lot of information about land use and land cover. However, different remote sensing platforms (such as the Landsat series, ASTER, and SPOT) have different spatial/temporal resolution and bands. In order to properly use satellite imageries, preprocessing steps are essential. Another example would be the large number of online documents that contain geographic locations, such as events in news articles. These geographic locations may appear in these documents in different surface forms (such as LA vs Los Angeles) and may need disambiguation steps [4] (the same surface form may refer to different geographic entities in different contexts) in order to find the correct association between the unstructured texts and geographic entities. In order to analyze these live events and discover the hidden geospatial patterns as a means to gain insights about the socioeconomic trend in our society, new methods are needed to close the gap between the demand of efficiently consuming the unstructured noisy geographic data and the supply of the lagging data handling ability of current Geographic Information System (GIS) tools and infrastructures. A good candidate is geospatial semantics because it can improve the interoperability of different geospatial data sources and operations [1]. For instance, Kemp et al. [5] used knowledge bases (a.k.a. knowledge graphs) as the middleware framework to accommodate semantic heterogeneity and provide analysis services for environmental information systems.

Accessibility Demand Moreover, the general trend in research and technology is that innovations and ideas are constantly being ported to a larger audience that are not necessarily experts in the field. This trend benefits both the research community and the general society in that it opens the dialog that facilitates the communication between them to share research progress as well as societal needs. By making research innovations more accessible, the general audience are able to appreciate the endeavors that have been made by numerous scholars and the research society in turn is able to collect feedback from the general audience about potential improvements. The recent effort in democratizing Artificial Intelligence (AI) research is a good example of this trend. By democratizing AI research, millions of people are able to be more efficient with everyday tasks. Realizing the potential of AI in our society, researchers have been extending the technology in various industries, including the financial sector and the health industry. Furthermore, new issues and challenges emerge as people start to deploy such technologies in a large scale, such as the ethical issues and the challenge of interpretability in AI. As an interdisciplinary field, GIScience has always been on the train towards a more democratized research agenda, i.e. a research practice that makes it more accessible to a wider audience. While traditional GIS tools such as ArcGIS[1] and Quantum GIS[2] require advanced knowledge in geoprocessing, more accessible tools such as Google Earth[3], CARTO[4], and Mapbox[5] have facilitated common users to create web maps and conduct basic geospatial analysis by importing data from spreadsheets. After realizing the power of GIS tools, people have started to embrace a variety of geospatial technologies, such as using Google Maps to find the shortest route, using Yelp to locate new restaurants, and geotagging photos on social media. People now have even higher expectations when they interact with geospatial tools. For example, they want the navigation systems to understand natural language commands and they want location search to be more intelligent in order to handle complex queries in addition to the address search or Points-Of-Interest (POI) search. In order to bridge the gap and make the interaction smoother, geospatial semantics comes into play as a means to tackle this challenge. Studies have known that semantics organized in (geographic) knowledge graphs can act as nexus between natural language and GIS systems in order to facilitate the question answering process in the

[1] https://www.arcgis.com/

[2] https://www.qgis.org

[3] https://www.google.com/earth/

[4] https://carto.com/

[5] https://www.mapbox.com/

geospatial domain [6, 7, 8].

Conceptualization Demand The inherent vagueness in geographic concepts has imposed a lot of challenges in processing, analyzing and understanding geographic phenomenon. In addition to mathematical and computational challenges, Montello et al. [9] discussed behavioral-science methods for determining the referents of vague geographic regions. Gao et al. [10] used a data-synthesis-driven approach to detect and extract vague cognitive regions. These attempts, though successful, only provide a temporary solution to the challenge. The root cause lies in the fundamental conceptualization of different geographic features or entities. Geo-ontology, as a sub-field of ontology which is a branch of philosophy, studies the constituents of reality and their relations within the geography domain in a systematic way [11]. In this sense, designing the ontology for various geographic features is an ideal approach to solving the conceptualization issue in GIScience. As a matter of fact, researchers have been working in this area from a theoretical as well as a pragmatical perspective. By examining the bona fide (i.e. natural) and fiat (i.e. artificial) characteristics of mountains, Smith and Mark [11] explained the implications for modeling landforms in geographic ontology to support environmental modeling and other GIS applications. In order to define vague geographic features, Bennett et al. [12] utilized standpoint semantics (a refinement of supervaluation semantics) that can be grounded in actual data by geometric analysis and segmentation of the data set. In a more applied manner, Hu et al. [13] designed an ontology for trajectory data. Grenon and Smith [14] proposed a modular ontology of the dynamic features of reality.

These three demands have given rise to the study of geo-ontologies and geographic knowledge graphs in GIScience. While they continue to help mitigate the demands, these areas are not without their own challenges. This dissertation is dedicated to addressing one of these challenges, namely geographic knowledge graph summarization, amid the big data era and the paradigm shift brought by geospatial semantics in GIScience. By first pointing out and clarifying the three demands, we hope to establish the background as well as our philosophy in dealing with the challenge of summarizing geographic knowledge graphs.

1.2 Motivation

While geospatial semantics and the semantic technology in general have been widely adopted in GIScience [15, 16, 13, 17], the amount of information accompanying this semantic lift is immense in terms of diversity as well as quantity. As a nexus component in geospatial semantics, geographic knowledge graph plays an important role in improving the interoperability, accessibility, and conceptualization in geographic data. However, the diversity of geographic knowledge graph has imposed a lot of challenges for researchers and general users. This diversity can be analyzed from two perspectives. Geographic knowledge graphs are linked with a diverse set of cross-domain knowledge graphs. Because of the interconnected nature of knowledge graphs, geographic knowledge graphs usually appear as subgraphs of cross-domain knowledge graphs, such as DBpedia[6], Wikidata[7], and Freebase[8]. These knowledge graphs interlink geographic entities with entities from life sciences, linguistic domain, media, social networks, and various user-generated contents. Such diversity, as a result of the linkage, has introduced a lot of possibilities as well as challenges. This diversity on the one hand provides geographic knowledge graphs with the ability to help solve cross-domain problems, such as question answering; on the other hand imposes challenge on organizing and digesting such heterogeneous information because usually different domains have different requirements and focus. For example, a

[6] https://wiki.dbpedia.org/

[7] https://www.wikidata.org

[8] https://developers.google.com/freebase/

geographic knowledge graph might include entities regarding human settlements which have a lot of human-centric information (e.g., demography) as well as entities regarding biogeography which focuses on the biological aspects. These differences entail different approaches in organizing, managing, and processing the structured data.

Another perspective for the diversity in geographic knowledge graph is about the large number of heterogeneous types of information for each entity. Take the DBpedia geographic entity *dbr:Los_Angeles*[9] as an example. This entity is connected to other entities through various relationship types, such as *(dbr:Los_Angeles, rdf:type, dbo:City)* and *(dbr:Leonardo_DiCaprio, dbo:birthPlace, dbr:Los_Angeles)* which link *dbr:Los_Angeles* with the class *city* and an entity (a person) *Leonardo DiCaprio* via a *type* relation and a *death place* relation respectively. In addition, entities can also be linked to literals. In the example of *dbr:Los_Angeles*, information such as population and elevation is expressed through literals (e.g., numbers). In principle, entities in knowledge graphs can be anything, tangible or intangible. In geographic knowledge graphs, such as the subgraph of DBpedia or Wikidata, multimedia (e.g., images) are usually part of the graph. These images, though represented as Uniform Resource Identifiers (URIs), encode information that is hard to extract otherwise. For example, the image might include different objects depicting different visual signals that can complement the graph itself.

In terms of quantity, the Linked Data Cloud[10] has been growing constantly. In the 2016-04 data dump, DBpedia contain more than 6 million entities of which 1.53 million are geographic entities. The DBpedia ontology has 754 classes and 2,711 relations/properties (including object and datatype properties). This sheer amount of information combined with the diversity of this information has introduced challenges for GIScience researchers to analyze and consume the powerful geographic knowledge graphs. More-

[9] `http://dbpedia.org/resource/Los_Angeles`

[10] `https://lod-cloud.net/`

over, although storage is not a major concern, the ability of end users to process such immense data is limited [18]. From a psychology perspective, the philosophy of less is more has been studied in human decision making under the notion of the paradox of choice [19]. Viewing the knowledge graph as an exploratory tool that users can interact with by iteratively choosing nodes to expand the knowledge, too many nodes on the graph lead to too many choices and would demotivate users from using the tool in the first place [20]. As a result, a novel research area, namely knowledge graph summarization has emerged. Analogous to text summarization where the summary provides a synopsis of the original text, knowledge graph summarization aims to identify the underlying structure and meaning of the graph using a digest graph.

The data mining community has a strong interest in (knowledge) graph summarization because graph structure is ubiquitous, such as the communication patterns in social networks and the molecular interactions in biochemistry, and the summarization process facilitates the reduction of data volume and storage, the speedup of graph algorithms and queries, the interaction and analysis of graph patterns, and the elimination of noise [21]. In the semantic web community, the summarization task is mostly focusing on the entity level [22, 23, 24] by considering diversity [25, 26], uniqueness [26], and popularity [26] with the optional assistance of human intervention [27, 28, 29] for knowledge graph exploration [30].

However, researchers in GIScience have yet started to explore the questions involving geographic knowledge graph summarization despite the fact that GIScientists are among the early adopters of knowledge graphs and semantic web related technologies due to the three demands mentioned in Section 1.1. This dissertation is motivated by such drastic contrast between the necessity of devising better ways to summarize geographic knowledge graphs and a dearth of research effort in the area. Such a research area is distinct from its siblings that spark interest in the data mining community and semantic web community in that spatial component is special [31] and by summarizing geographic knowledge graphs the original three demands that brought geographic knowledge graphs to GIScience in the first place should not be ignored.

1.3 Research Questions

In a broad sense, one could ask the research question *how to summarize geographic knowledge graphs?* Such a question is too general and could be decomposed into different aspects. Numerous methods for graph summarization have been proposed. Some of them are based on handcrafted features paired with machine learning and data mining algorithms [32, 33]. Others are based on top-down information such as the graph structure [34, 22]. In this dissertation, we would like to explore a hybrid approach, namely combining the bottom-up and top-down approaches in order to reap their complementary strengths. In this case, we would like to ask the question *How can we leverage both top-down knowledge engineering and bottom-up knowledge learning approaches to help summarize geographic knowledge graphs?*

This dissertation proposes to focus on both top-down and bottom-up approaches for three major reasons. First, the geographic system is a complex system which involves humans, environment, and the intricate interplay between them. Geographic knowledge graphs, acting as proxies to connect different components in this complex system in a semantically-enriched manner, are thus also complex. Similar to other domains such as vision, language, control, and decision-making where the dichotomy between hand-engineering (top-down knowledge) and end-to-end training (bottom-up learning) is stalling the progress of developing models and methods that can generalize, research in geographic knowledge graph summarization should take into account the challenge of combinatorial explosion due to the complexity of the system. Second, the combination of top-down and bottom-up approaches has proven to be effective for developing more generalized models. The principle of combinatorial generalization is constructing new inferences, predictions, and behaviors from known building blocks [35]. Such an idea has been explored in vision where the building blocks take advantage of the spatial translation invariance [36, 37] in images, in language where the building blocks are informed by the temporal translation invariance [38] in sentences, and in network analysis where the building blocks are from the node and edge permutation invariance [39] in graphs. To give a concrete example in the natural language domain, humans are able to utilize a few sets of elements (words) and combine them in limitless ways (sentences). This ability to make infinite use of finite means [40] marks the key component of human intelligence. Inspired by the wisdom in biology where nature and nurture complement each other, this dissertation acknowledges the fact that top-down and bottom-up methods are compatible with each other and they can work together to find a solution to the geographic knowledge graph summarization problem. Third, the spatial[11] component in geographic knowledge graphs is a natural source of top-down knowledge. Unlike domain-agnostic knowledge graphs, geographic knowledge graphs are accompanied by a number of hidden patterns informed by the geographic components, such as spatial correlation and spatial dependency. Leveraging such *geospatial inductive bias* (i.e. top-down knowledge) is helpful in the context of geographic knowledge graph summarization. Analogous to the spatial translation invariance, temporal translation invariance, and permutation invariance in vision, language, and network analysis, geospatial contextual invariance (i.e. the common ways in which geospatial context can inform nearby spatial/non-spatial attributes) is essential in dealing with the combinatorial generalization challenge in summarizing geographic knowledge graphs.

After establishing the general research question, let us decompose it into several

[11] *Spatial*, *geographic*, and *geospatial* are used interchangeably in this dissertation.

related small questions to obtain a more tangible idea of the scope and challenge of this research. Considering the composition and diversity of geographic knowledge graphs explained in Section 1.2, the question of summarization can be treated from three aspects, namely the hierarchical components, the multimedia leaf node components, and general relation and entity components. Thus, this dissertation is aiming to tackle three research questions regarding geographic knowledge graph summarization.

Research Question 1: How do we summarize the hierarchical place type information in geographic knowledge graphs?

This research question focuses on the hierarchical component, namely the place types. This is an important component in geographic knowledge graphs because the hierarchical place type structure reflects the way humans conceptualize the relationship of different geographic categories. In a sense, by isolating and emphasizing this hierarchical place type component, we also aim to preserve the conceptualization power of the ontology accompanied by the geographic knowledge graph and make sure that the summarization process would still satisfy the conceptualization demand of semantically lifting geographic data. While place type conceptualization and their hierarchical relationships (i.e. superclass and subclass) are merely part of the geo-ontology, they pose more challenge than the axioms in the summarization process because these conceptualizations are related to philosophical human construct — an area where machine intelligence still struggles — whereas the axioms are related to first-order logic — an area where machine traditionally excels.

To give a concrete example, consider the place type hierarchy in Yelp data[12] (shown in Figure 1.1). On the top level there are 22 root place types (e.g., *Restaurants*, *Shopping*, *Health & Medical*, etc.) and on the bottom level there are 1,030 place types. A naive approach would be to choose a particular level and use the place types in that level as the

[12] `https://www.yelp.com/dataset`

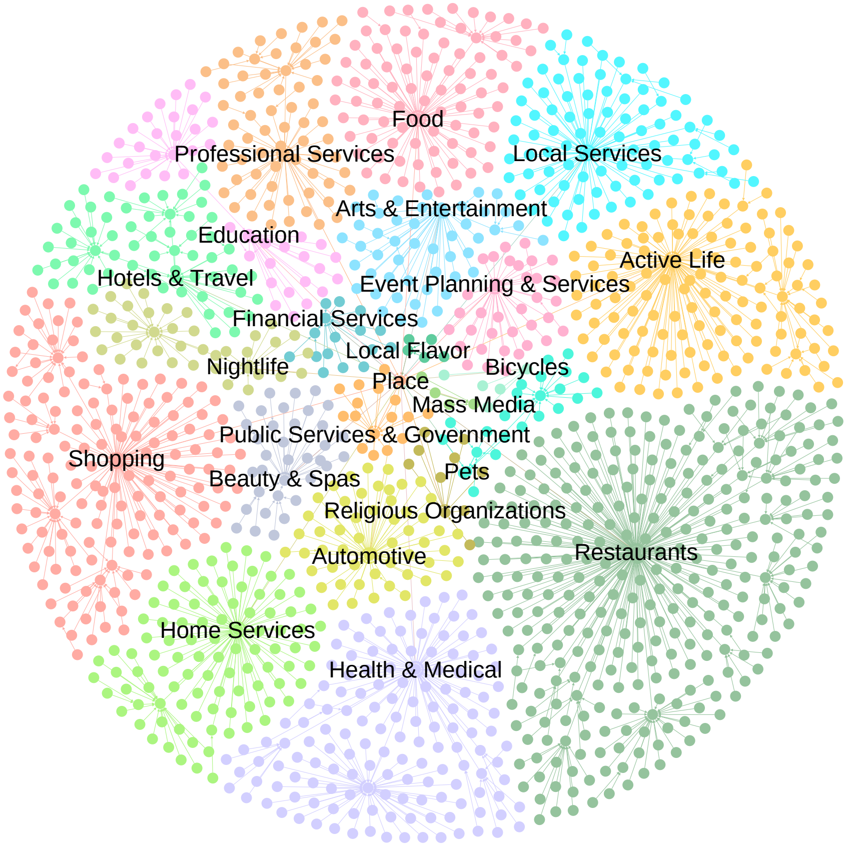


Figure 1.1: Place type hierarchy visualization for Yelp data. These place types are commonly used in geographic knowledge graphs and are important part of the conceptualization in the ontology.

summarized place types. However, it's hard to decide on the cutoff level because such a method is dependent on the particular place type hierarchy. In addition, uniformly choosing place types from the same level ignores the fact that hierarchical conceptualizations of different place types are not balanced. Certain place type conceptualizations are more expressive and informative than others. For instance, as shown in Figure 1.1, the branch for *Restaurants* has much more place types than the branch for *Education*. As a result, by uniformly cutting off at a particular level, a lot more information is lost for the *Restaurants* branch than for the *Education* branch. In this case, a bottom-up approach appears to be more generalizable for place type summarization across different geographic knowledge graph datasets. Instead of focusing on the hierarchical structure, the data-driven method could essentially take advantage of the linguistic aspects of the place type by examining the meaning of each concept in the hierarchy (ontology).

However, the words associated with these place types are merely proxies for geographic feature types in the knowledge graph. In order to reveal the underlying geospatial semantics [16], the model also need to be ware of the geospatial context. The main challenge of this research question then comes down to developing a model that marries the data-driven method and the geospatial contextual knowledge. The objective would be to use the result of the model to guide the place type summarization process, such as ranking and selecting relevant place types for the use case.

Research Question 2: How do we summarize multimedia leaf node information, such as images, in geographic knowledge graphs?

This research question focuses on the multimedia leaf node components. As mentioned in Section 1.2, geographic knowledge graphs are versatile because it can carry multimedia information, such as images, in the leaf nodes. This special ability also corresponds to the accessibility demand in the geospatial semantics paradigm because this visual information is a catalyst for a better human-machine interaction. Because of this, in research question 2, we are dedicated to developing approaches to summarizing these multimedia nodes using images as examples.

Images in geographic knowledge graphs only exist in leaf nodes as Uniform Resource Locators (URLs) and they do not have labels (e.g., mountains, rivers, etc.). In order to select a subset of relevant images and summarize the whole graph by striking the balance between commonality and variability, we need to develop robust algorithms to help label the numerous leaf image nodes first. Although existing image classification models can assist in labeling the images, they usually suffer from a tendency towards biases [41] because there is a discrepancy between the training data distribution and the distribution of more complex real-world systems [42]. Since our task domain is in GIScience, we are interested in devising ways in which such bias can be mitigated for labeling images in geographic knowledge graphs. One potential solution to such a challenge is to incorporate geospatial signals in addition to the visual stimuli in current state-of-the-art models. This idea corresponds to the notion of *geospatial inductive bias* (which will be explained in detail in Chapter 2) and is in line with the idea of integrating geospatial components in Research Question 1.

As there are different ways to consider geospatial components, such as geographic distance and topological relationship, under different granularities, such as neighborhood level, city level, and country level, it would be worthwhile to explore different integration strategies with the visual components on images. This exploration gives us guidance on the extent to which different geospatial components and different integration strategies can benefit the image labeling process and subsequently the summarization process of geographic knowledge graphs. Because both the image classification model and the *geospatial inductive bias* are generic methods and ideas, the resulting hybrid models can be applied to a variety of geographic knowledge graphs.

Research Question 3: How do we summarize relations and entities in geographic knowledge graphs in general?

While Research Question 1 and Research Question 2 aim to tackle the hierarchical place type structure in the geo-ontology and the multimedia leaf nodes in geographic knowledge graphs that correspond to the conceptualization demand and the accessibility demand respectively, Research Question 3 is focusing on finding the general solution to the summarization problem. After dealing with the place type information in the ontology and the leaf image nodes which are important components in geographic knowledge graphs, we would like to explore summarization approaches that can be applied to general relations and entities (besides place type relations and leaf image node entities).

For text summarization, there are usually two major types, namely extraction-based and abstraction-based text summarization. Extraction-based summarization extracts words and sentences directly from the original text based upon relevance and importance while the abstraction-based one involves paraphrasing and provides a more condensed summary. Similar to text summarization, knowledge graph summarization methods can also be categorized into extraction-based and abstraction-based approaches. More specifically, according to the core techniques employed, popular ones include grouping or aggregation based (extraction or abstraction), bit compression based (extraction or abstraction), simplification or sparsification based (extraction), and influence based methods (extraction or abstraction) [21]. Since we would like to maintain the interoperability that geographic knowledge graphs created, extracting a subset of relations and entities from the original graph subsequently making existing connections and conceptualizations intact is preferred. In this case, it is more desirable to employ the simplification or sparsification based approaches.

There are two major challenges in tackling this question. The first one is the subjectivity issue in summarizing geographic knowledge graphs. Since the relative significance of a relation or entity in identifying the graph is subjective to the application field, it is hard to universally define any metrics or evaluation schemes to justify the choice. Fortunately, studies [43, 44] have shown that Wikipedia articles are relatively unbiased. So leveraging curated, neutral summaries in Wikipedia articles would be a good start. Such an idea is more obvious considering that there is a clear correspondence between Wikipedia articles and many major knowledge graph repositories [45], such as DBpedia. The second challenge is related to the geospatial component in geographic knowledge graphs. While the semantics of a knowledge graph is well-established, evidence [46, 47] has shown that geospatial semantics (in a geographic knowledge graph) needs special care. In order to better summarize geographic knowledge graphs, the proposed method should explicitly model geographic components and make them the first-class citizens.

1.4 Dissertation Synopsis

This introduction chapter describes the background and motivation behind the rising interest in geographic knowledge graphs and the geospatial semantics domain in general. It then explains in detail the challenges and discusses three open questions that this dissertation is aiming to answer. The core content of this dissertation is based upon three individual yet related articles. These three articles (Chapter 3, Chapter 4, and Chapter 5) provide answers to three research questions raised in Section 1.3. These three chapters first appeared as various publications shown below:

- Chapter 3: Yan, Janowicz, Mai, and Gao [48].
- Chapter 4: Yan, Janowicz, Mai, and Zhu [49].
- Chapter 5: Yan, Janowicz, Mai, and Zhu [45].

The remainder of this dissertation is organized as follows.

Chapter 2 presents background knowledge and foundational concepts that have been used frequently in this dissertation. Specifically, we introduce and define the concept of geospatial context, geospatial inductive bias, knowledge graph, and geographic knowledge graph. By pointing out the fact that GIScience research has been implicitly using the idea of geospatial context and geospatial inductive bias, this chapter establishes the unification of these frequently used ideas as a foundation for the methods used in later chapters. It then introduces the summarization task for (knowledge) graphs, classifies existing work based on their core techniques, and reviews related methods and algorithms. Finally, it points out the need for methods that consider geospatial inductive bias to better summarize geographic knowledge graphs.

Chapter 3 presents a latent representation learning method for place types. Place types are typically represented in a hierarchical structure and are widely-used as conceptualizations in the ontology for geographic knowledge graphs. Traditionally, latent representations of place types are learned via their distributional semantics based on their occurrences in linguistic contexts. In this chapter, we ground these place types into the geospatial context and learn their embeddings based on geospatial distributional semantics. Geospatial inductive bias is applied using an information-theoretic distance lagged approach on both local and global neighborhoods. The final embeddings carry geospatial semantics content that is otherwise ignored by other approaches. These embeddings can be used to determine the similarity and relatedness of different place types and help decide the relevance scores in geographic knowledge graph summarization tasks.

Chapter 4 presents a method that utilizes geospatial contextual information as a Bayesian prior to help improve the classification of images for different place types. The major challenge in image classification is that the bias in training samples is likely to affect the classification result of unseen patterns. In order to facilitate the classification process, the model considers both the visual stimuli and the geospatial context in which the image is located. Geospatial inductive bias is applied by using the latent representations (Chapter 3), spatial co-location patterns, and spatial sequence patterns. This classification method can be used in geographic knowledge graph summarization to help select different types of images in a neighborhood as image labels are typically not specified in geographic knowledge graphs.

Chapter 5 presents a spatially-explicit reinforcement learning model for geographic knowledge graph summarization. The model formulates the summarization task as a sequential decision making process through trial and error. The learning process is powered by the theories in reinforcement learning. The geospatial inductive bias is applied by introducing an explicit spatial action for the reinforcement learning agent. While Chapter 2 and Chapter 3 present two special cases of summarization geographic knowledge graphs based on two important components (place types and image leaf nodes), this chapter aims to provide a more generic approach.

Chapter 6 presents a web map interface that facilitates the exploration of geographic knowledge as a case study for summarizing geographic knowledge. The interface takes advantage of the proposed geocoding enrichment process as well as an entropy-based geographic knowledge graph summarization approach. In order to provide a scalable system, it adopts the serverless and scalable framework using Amazon S3, AWS Lambda, and Amazon API Gateway provided by the Amazon Web Services. Linkage discovery and spatial pattern discovery are presented as examples to illustrate the usefulness of the web map interface as a means of discovering spatial knowledge.

Finally, Chapter 7 concludes the dissertation by providing a summary and discussion of previous chapters. In addition, research contributions including theoretical and practical implications are discussed for this research. Several limitations are listed and possible future research directions are proposed for further investigation.

Chapter 2

Foundational Concepts and Theories

This chapter provides two pieces of information — the background knowledge to understand the ideas and methods used in the dissertation and related work for (knowledge) graph summarization. The background knowledge includes the concept of geospatial context and geospatial inductive bias as well as the definition of knowledge graphs and geographic knowledge graphs. We introduce this background knowledge by means of providing intuitive examples, comparing with other related fields, and explaining based on existing work.

2.1 Geospatial Context and Geospatial Inductive Bias

This section first explains the concept of geospatial context which is usually implicitly used in a variety of GIScience research and reviews several research studies that used this idea. Then it brings up the idea of geospatial inductive bias that is central to the methods used in later chapters. While similar concepts have been used in research that incorporates geospatial components, we believe it is necessary to explicitly define these terms and clarify their implications.

2.1.1 Geospatial Context

There is a consensus among GIScientists that locations are not just two extra columns (one for latitude and one for longitude) in your spreadsheets. Such an understanding implies that location data or geospatial data should be treated differently compared with data with no spatial attributes. Indeed, spatial is special. By studying spatial data organization, analysis and interpretation, spatial statistics has pointed out the importance of many geospatial components, such as distance [50], direction [51], spatial autocorrelation [52], spatial nonstationarity [53], and spatial interaction [54]. While these interrelated concepts and statistics have their own emphasis on different aspects of geospatial data, they have some commonalities. First of all, they all have a reference area. While one can still calculate these statistics for an arbitrary number of geographic data in an arbitrary sized area, the scale and the context [55, 56] are crucial for interpreting these statistics meaningfully. Regardless of the scale (local or global), it is important to maintain the consistency of the reference area in measuring these values. Second, while these statistics are grounded in geographic locations, their values are usually from other attribute information, such as temperature and humidity.

In the hope of unifying these different but related terminologies as well as generalizing the ideas behind these research efforts, we propose the concept of *geospatial context*. As the name suggests, instead of defining the exact type of statistics or aspects in which geographic component is used, geospatial context is a broad concept and focuses more on the fundamental idea that the context in which geospatial data and their relationships are examined is an important consideration in GIScience research.

Definition 1 (Geospatial Context) *Given a geographic dataset, each entity e has a geospatial context $GC_e = (R_e, A_e, f_e)$ where R_e is the reference area, A_e is the set of values associated with e, and f_e is the function to encode the geospatial contextual information for e.*

The reference area R is a generic term that defines the scale of the geospatial context. It can be measured by the absolute area surrounding the entity, such as the buffer area or determined by the total area covered by a fixed number of nearby entities. This flexibility allows for customizable scales in which geographic data is aggregated. In addition, the reference area R is not uniformly defined across all entities in the dataset and can even change depending on the choice of A and f. This dynamic nature of R reflects the spatial homogeneity and heterogeneity of geographic data. The value set A determines the types of attributes used in the study. It is usually a subset of all the attributes associated with the dataset. This subset of attributes can be in the form of nominal, ordinal, interval, or ratio data. For example, activity categories (nominal), temporal bands (ordinal and interval), and check-in counts (ratio) are all used in a study to encode behavior-driven temporal signatures of different place types to improve reverse geocoding [57]. Typically, the same set of A is used across the whole dataset. There can be exception if certain attributes are missing for individual entities or certain attributes are related. The encoding function f takes the A values of current entity and all geographic entities within the reference area R as inputs and outputs the embedded *geospatial contextual information*. As mentioned before, this function f can be commonly-used spatial statistics, such as distance decay factors, Moran's I [58], and Ripley's K [59]. It can also be analytical patterns [60] or approximation functions learned from optimizing the parameterizations to satisfy particular geographic patterns [61, 62, 63].

Numerous research studies have either implicitly or explicitly adopted the concept of geospatial context and taken advantage of the patterns in the geospatial contextual information to improve model performance and help with the decision-making process. In criminology, distance-decay function has been used for profiling [64]. In data management, spatial signatures [65] have been used for ontology matching across different geospatial datasets. In information retrieval, geospatial context has been used for music recommendation and achieved better results compared with non-spatial models [66]. In computer vision, location context has been used as Baysian priors [61] as well as features in neural networks [62] to improve image classification accuracy.

2.1.2 Geospatial Inductive Bias

As this dissertation aims to tackle the geographic knowledge graph summarization problem by means of a combination of top-down and bottom-up approaches, one of the concerns is to develop an ideal learning process. Typically, the learning process gains knowledge by observing available data and finds the solution that better explains the underlying patterns or achieves high rewards. Since in any realistic learning process the instance space (all possible data points) would be too large to be covered by the training dataset, assumptions or preferences have to be made on the hypothesis space (e.g., rules that determine the decision boundaries for a classification problem) for inductive methods to achieve on average better results than random guessing [67]. This type of a priori assumption or preference is called the *inductive bias* [67, 68, 69] in the machine learning and AI community in general.

The inductive bias allows the learning algorithm to prioritize one solution over the other [67]. This bias is manifested in different forms. To overcome overfitting, regularization might be incorporated in the model as an inductive bias. For Bayesian models, an inductive bias might be a specific parameterization or the choice of prior distribution. In graphical models, the push towards the greatest randomness and the memorylessness assumptions are inductive biases in Maximum Entropy Markov Models. In neural networks, the inductive bias is represented as a preference towards a particular network architecture, such as convolutional vs. recurrent blocks.

Many approaches in GIScience as well as other application domains which utilize geospatial data use the idea of *geospatial inductive bias*. Instead of giving a formal definition, we use this term to generally refer to inductive biases which impose constraints or assumptions based upon geospatial contextual information obtained from the geospatial context. While linear regression has the inductive bias that the data-generating process can be explained simply as a line process corrupted by additive Gaussian noise [35], geographically weighted regression [53] has the geospatial inductive bias that separates the region into local subregions (geospatial contexts) to model spatially varying relationships (spatial nonstationarity). Geospatial inductive bias can also be the choice of spatial indexing, tessellation, and aggregation scale. Geospatial semantics is a source of geospatial inductive bias and it can be used as a means to inform people's sense of place to an extent comparable to that of pure cognitive approaches [10]. In Chapter 3, we will introduce an information-theoretic distance lagged approach as a means of geospatial inductive bias to adjust the distribution of POIs and learn latent representation for place types in local as well as global geospatial contexts. In Chapter 4, the geospatial inductive bias is based on the sequentially-dependent geospatial contextual information and is modeled as spatially-explicity Bayesian priors to facilitate the classification of images of different place types (e.g., restaurants, hotels, etc.). In Chapter 5, we exploit the geospatial inductive bias by explicitly introducing a spatial action in the policy agent in order to account for the spatial dependencies in the geographic knowledge graph summarization process.

2.2 Geographic Knowledge Graphs

In this section, we explain the components of a knowledge graph and the commonly-used data model to represent knowledge graphs. Then we provide the general definition of a geographic knowledge graph and point out its distinctions from other knowledge graphs.

While the idea behind knowledge graphs is nothing new [70, 71, 72, 73], it is experiencing a renaissance among a wide range of research communities (including GIScience) after its successful debut in large scale data management systems (including Google search engines) and becoming the common support for browsing, searching, and discovering knowledge. Knowledge graphs are structured datasets that describe entities and their relationships. The entities can be anything and the relationships can have any type. This flexibility has enabled knowledge graphs to become an important component of the semantic web [71]. Although knowledge graphs can be implemented using different underlying data structures and conform to different standards, most knowledge graphs, such as DBpedia, Wikidata, and Freebase, support the Resource Description Framework (RDF) model.

The atomic data entity in RDF is the *triple*, which is composed of three parts in the form of subject-predicate-object (s-p-o) expressions[1] (e.g. Santa_Barbara isPartOf California). Subjects and objects can be entities (objects can also be literal strings)

[1] The alternative head-relation-tail or h-r-t expressions are also used in some literature and this dissertation uses both.

while predicates define the relationships between subjects and objects. Every entity or relationship is represented by an URI which uniquely identifies it. Conceptually and equivalently, the entities in the RDF model are nodes and the predicates are edges (links) in the graph. Here we give the definition of a knowledge graph considering the duality that it is both a graph model and represented by the RDF triples.

Definition 2 (Knowledge Graph) *For a set of RDF triples T where each triple $t_i = (s_i, p_i, o_i) \in T$, a Knowledge Graph is a multi-relational graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ where $\mathcal{V} = \{s_i | (s_i, p_i, o_i) \in T\} \cup \{o_i | (s_i, p_i, o_i) \in T\}$ and $\mathcal{E} = \{p_i | (s_i, p_i, o_i) \in T\}$.*

From the definition we can see that a knowledge graph is conceptually a graph but represented as RDF triples for computational and reasoning purposes. Unlike homogeneous graphs, a knowledge graph usually contains hundreds of thousands of different types of relations (such as isPartOf, headquarterOf, locatedIn relations). Homogeneous graphs, such as the friendship network, the coauthorship network, or the molecular interaction network, only contain one type of relation, such as the friendship relation, collaboration relation, or molecular interaction. Because it models different relationships in the graph, a knowledge graph is considered a multi-relational graph (a.k.a. heterogeneous information network in some literature).

Because of the duality, research efforts have long been focusing on two parts, the RDF part, such as reasoning [74], SPARQL[2] queries [75, 76], and triple pattern fragment [77, 78, 79], and the network part, such as entity resolution [80, 32, 81]. There is a trend to unify these two separate yet related parts in the knowledge graph and semantic web community. Mika [82] introduced a tripartite model that extended the bipartite model of ontologies with the social dimension and showed how community-based semantics emerges from this model through a process of graph transformation. In order

[2] The query language for RDF data.

to provide support for fine-grained latent coherence between entities and predicates in graph-based authority ranking, Franz et al. [34] presented the TripleRank model to capture the additional latent semanticsby means of statistical methods in order to produce richer descriptions of the available data. Schlichtkrull et al. [83] adopted the idea of latent representation learning in Graph Convolutional Networks (GCN) and introduced Relational Graph Convolutional Networks (R-GCN) to handle the highly multi-relational data characteristic of realistic knowledge graphs.

Geographic knowledge graphs are domain specific knowledge graphs. In a broader sense, any knowledge graph that contains geographic entities is a geographic knowledge graph. Here we give the definition of geographic knowledge graphs.

Definition 3 (Geographic Knowledge Graph) *Given a Knowledge Graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$, for a propositional function $P(x)$ denoting 'g(x) is a meaningful geographic identifier' where function g maps an input x to a geographic identifier, if $\exists\, v \in \mathcal{V}\ P(v)$ is true then $\mathcal{G}$ is a Geographic Knowledge Graph.*

The mapping function g would try to map an entity/node to a geographic identifier. For example, if the entity s is *University of California Santa Barbara*, $g(s)$ would be, for instance, the centroid point coordinates or the polygon representing the university in Well-Known Text (WKT). Such geographic identifiers can be any geometry, such as points, lines, and polygons, in any representation, such as WKT, geojson, and shapefile. The most common geographic identifiers used in geographic knowledge graphs are points represented by WKT. This definition does not restrict geographic knowledge graphs to be knowledge graphs with only geographic entities because geographic entities are usually associated with nonspatial attributes such as names and geographic types. Ideally, a geographic knowledge graph should contain predicates that represents topological relationships of geographic entities. With the assistance of GeoSPARQL[3], geospatial queries can be conducted directly on RDF data and this capability accommodates systems based on qualitative spatial reasoning and systems based on quantitative spatial computations.

As the idea of geographic knowledge graph is still in its infancy, there are only a handful of research about it. Mai et al. [84] visualized the distribution of geographic features in a semantic space using the DBpedia geographic knowledge graph based on the idea of semantic enhancement in spatial visualization [85]. Janowicz et al. [86] comprehensively analyzed the systematic errors in geographic knowledge graphs and their potential causes and discussed lessons learned and means to avoid some of the introduced pitfalls in the future. Regalia et al. [87] used a geographic knowledge graph to showcase a proxy that can transparently run on top of arbitrary SPARQL endpoints to enable the on-demand computation. Kejriwal and Szekely [88] presented a dataset that embeds populated place in the Geonames[4] knowledge graph using neural embeddings methods.

2.3 Graph Summarization

In this section, we talk about two related concepts, namely graph summarization and knowledge graph summarization. Graph summarization is an area that has attracted a lot of research efforts in the data mining and data management communities. We explain the objectives of graph summarization, discuss its challenges, and review existing methods. Strictly speaking, for the knowledge graph summarization, we are in fact explaining the idea of semantic graph summarization which is a superset of knowledge graph summarization. Nonetheless, they have the same objective and share the same methods. The idea of knowledge graph summarization will be discussed and existing methods reviewed. Finally, we point out that methods specifically tailored towards geographic knowledge

[3] https://en.wikipedia.org/wiki/GeoSPARQL

[4] https://www.geonames.org/

graphs are needed in order to capture the geospatial inductive bias in geospatial contexts.

The idea of graph summarization has becoming popular in the context in which an enormous amount of graph structured data has been produced on a daily basis and human's ability to process such a large amount of information and identify hidden patterns in the data is limited. The objective of the graph summarization task is therefore to facilitate the identification of structure and meaning in data [21] and assist the discovery of hidden patterns. It helps to reduce data volume and storage, speed up graph algorithms and queries, support interactive analysis, and eliminate noise. However, there are also many challenges for this task. First of all, although graph summarization can reduce the data volume, the summarization algorithm itself is faced with the challenging of processing large data inputs. Efficient algorithms are thus critical in the summarization task. Second, graph data is complex. The graph structure is versatile because of the underlying structure (e.g. homomorphism and isomorphism) and this versatility has imposed challenges in dealing with graph data. The heterogeneity of nodes and edges in real-world graphs makes analyzing graph patterns a complicated task. The graph can be dynamic and evolve over time. In addition, the noise and missing information in graph data add another layer of complexity. Third, graph summarization is subjective. Depending on the domain knowledge and user preference, the consideration of the trade-off among space, time, and information preservation in the graph summary, the complexity of mapping the solutions to recover the original graph from the summary, and the optimization formulation, the summarization could be quite different. Decision for each of the components during the process is subjective. As a result, the evaluation is also subjective and challenging. The evaluation of a graph summarization result is heavily dependent on the application domain. For a social network, the evaluation will be based on whether the summary is able to preserve community information. For a visualization task, the evaluation will be based on user studies and qualitative criteria.

Depending on the core techniques employed, graph summarization methods include grouping or aggregation based, bit compression based, simplification or sparsification based, and influence based [21]. The grouping or aggregation based approaches aggregate nodes into supernodes based on application-dependent optimization functions. Grouping-based methods are among the most popular approaches for graph summarization. For example, by clustering nodes and discovering communities [89, 90], algorithms can help obtain a summary view of the graph. LeFevre and Terzi [91] proposed a hierarchical clustering-based node grouping algorithm (GraSS) for graph summarization that can target accurate query handling. Toivonen et al. [92] developed a method that merges structurally equivalent nodes in a way that minimizes approximation error and maximizes compression. The goal of bit compression-based methods is to minimize the number of bits needed to describe the original graph. Works in this category typically formulate summarization as a model selection task and employ the two-part Minimum Description Length (MDL) principle [21]. Simplification-based methods summarize the graph by removing unimportant components (nodes and edges) to produce a sparsified graph. For instance, OntoVis [93] provides a visualization tool to help filter nodes in order to facilitate the understanding of large heterogeneous networks. Influence-based approaches summarize the graph by formulating the tasks as an optimization process and understanding the patterns of influence propagation in the network. For example, Mehmood et al. [94] adopted community-level social influence information propagation analysis to summarize social networks.

In the context of knowledge graphs, the summarization task has a different scope. Knowledge graph summarization is specifically focusing on labeled, directed, heterogeneous graphs with semantics and type information. Based upon the algorithmic ideas, Čebirić et al. [95] classified knowledge graph summarization methods into four categories, namely structural methods, pattern-mining methods, statistical methods, and hybrid methods. Zhang et al. [96] proposed to summarize the ontology by extracting a set of salient RDF sentences according to a re-ranking strategy. Khatchadourian and Consens [97] developed a summarization method by combining text labels and bisimulation contractions. Zneika et al. [98] summarized large RDF graphs using top-K approximate graph patterns. Song et al. [99] utilized approximate graph pattern matching to summarize entities in terms of their neighborhood similarity up to a certain hop. Hose and Schenkel [100] proposed a sketch-based query routing strategy that takes into account source overlap in order to select sources from the knowledge graph cloud.

For geographic knowledge graphs, the geographic components add another layer of complexity. As mentioned in Section 2.1, the hidden pattern in geospatial contexts contain a variety of information that could potentially assist the summarization process for geographic knowledge graphs as a large number of entities have their corresponding geographic location. Existing methods fail to consider the rich geospatial contextual information and cannot take into account the inherent geospatial inductive bias in the summarization algorithms.

2.4 Summary

In this chapter, we laid out the foundational concepts and theories used in this dissertation. The first concept is geospatial context which we provided the formal definition to help unify existing work using the idea of geospatial contextual information to develop better models for GIScience problems. Based on the notion of geospatial context and the idea of inductive bias in AI research, we introduced the concept of geospatial inductive bias. We gave examples showing how this fundamental idea of imposing constraints or assumption based on geospatial contextual information is widely-adopted in GIScience research. After establishing these foundational concepts and ideas in solving GIScience questions, we provided the definition of knowledge graphs, acknowledging the its duality. The definition of geographic knowledge graphs was introduced. In the end, we presented challenges of summarizing (knowledge) graphs, reviewed existing algorithms, and pointed out the need for new methods that are tailored towards geographic knowledge graph summarization tasks.

Chapter 3

Reasoning About Place Type Similarity and Relatedness

In this chapter, a spatially-augmented latent representation learning method is proposed to embed place types. Such a method explicitly considers geospatial contextual information. By comparing the result with existing word embedding result using one hierarchy-based evaluation scheme and two human judgment-based evaluation schemes, we show that by applying the spatial context augmentation and geospatial inductive bias the embeddings are able to capture important geospatial semantics. Such embeddings are important for geographic knowledge graph summarization systems to select and rank the relevance of different place types.

Peer Reviewed Publication	
Title	From ITDL to Place2Vec — Reasoning About Place Type Similarity and Relatedness by Learning Embeddings From Augmented Spatial Contexts
Authors	Bo Yan, Krzysztof Janowicz, Gengchen Mai, and Song Gao
Venue	ACM Sigspatial Conference (SIGSPATIAL'17)
Editors	Erik Hoel, Shawn Newsam, Siva Ravada, Roberto Tamassia, Goce Trajcevski
Publisher	ACM
Pages	35:1 – 35:10
Submission Date	June 23, 2017
Acceptance Date	August 24, 2017
Publication Date	November 11, 2017
Copyright	Reprinted with permission from ACM

Abstract Understanding, representing, and reasoning about Points Of Interest (POI) types such as `Auto Repair`, `Body Shop`, `Gas Stations`, or `Planetarium`, is a key aspect of geographic information retrieval, recommender systems, geographic knowledge graphs, as well as studying urban spaces in general, e.g., for extracting functional or vague cognitive regions from user-generated content. One prerequisite to these tasks is the ability to capture the similarity and relatedness between POI types. Intuitively, a spatial search that returns body shops or even gas stations in the absence of auto repair places is still likely to satisfy some user needs while returning planetariums will not. Place hierarchies are frequently used for query expansion, but most of the existing hierarchies are relatively shallow and structured from a single perspective, thereby putting POI types that may be closely related regarding some characteristics far apart from another. This leads to the question of how to learn POI type representations from data. Models such as Word2Vec that produces word embeddings from linguistic contexts are a novel and promising approach as they come with an intuitive notion of similarity. However, the structure of geographic space, e.g., the interactions between POI types, differs substantially from

linguistics. In this work, we present a novel method to augment the spatial contexts of POI types using a distance-binned, information-theoretic approach to generate embeddings. We demonstrate that our work outperforms Word2Vec and other models using three different evaluation tasks and strongly correlates with human assessments of POI type similarity. We published the resulting embeddings for 570 place types as well as a collection of human similarity assessments online for others to use.

3.1 Introduction and Motivation

Semantic similarity and relatedness measures are prominent components of a variety of methods in geographic information retrieval, recommender systems, ontology engineering, and so forth; see [101] for a recent overview.[1] Given the importance of categorization for human cognition [103], place types are one of the three components (location and name being the other two) published by all major gazetteers and POI databases.[2] Place types act as a proxy for functions that a particular place of a given type affords. Intuitively, the presence of a nightclub (irrespective of its name or location) implies a certain exposure to noise during nights, the presence of a younger demographic, singles, a higher potential for drug related crimes, the possibility of getting a drink or snack late at night, and so forth. While each nightclub may differ to some degree, nightclubs share many of their characteristics with bars and the broader category of music venues, while they can neither act as substitute for bakeries nor barbers. Consequently, in the absence of POIs of a certain type, e.g., `Nightclub`, within a search radius, a system should return a place

[1] Similarity and relatedness are related concepts, in fact similarity is a subproperty of relatedness but not the other way around. To give an intuitive example, the Griffith Observatory is *related* to Griffith Jenkins Griffith via a `donorOf` relation but the observatory and the person are not *similar*. Many techniques, especially those based on linguistic aspects (including Word2Vec [102]) instead of formal semantics, cannot effectively distinguish between similarity and relatedness. Consequently, we approach them here together. Two of our three evaluation schemata, however, will explicitly focus on (human) assessments of similarity.

[2] In the following, we will use Point of Interest (POI) and place as synonyms.

of a similar type, e.g., `Bar`. This implies that semantic similarity measures should reflect human assessments of similarity, be it about place types or another topic.

To measure similarity, one may syntactically compare type labels, compute the distance in a place type hierarchy, count common place in their extensions, and so forth. New methods rely on comparing their linguistic meaning by learning word embeddings for all types and then computing their Cosine Similarity. However, such approaches do not consider any spatial information that is implicitly embedded in these place types, such as their co-occurrence patterns. This idea resembles the distributional semantics in linguistics and can be further summarized as: *place can be categorized by their neighbors.* The original counterpart in the linguistics is: *You shall know a word by the company it keeps* [104].

In this work, we embrace the idea of distributional semantics in geographic space and explore the similarity and relatedness of place types using different latent representations with augmented spatial contexts. Spatial contexts are augmented both intrinsically and extrinsically. In order to consider distance in our approach, distance decay and distance lags are used as intrinsic adjustments to augment the spatial contexts. We realize that there is a notable difference between place and space, namely *place is space infused with human meaning* [105], so we take check-in counts, i.e., popularity, as a proxy for human activities into consideration as well. Finally, and to adjust for the fact that place types follow a power law distribution, we also take the uniqueness of types at a certain distance into account. We approach both aspects from an information theoretic perspective, i.e., by measuring *information content.*

The contributions of this paper are as follows:

- We illustrate that the commonly used linguistic models alone cannot adequately capture the structure of geographic space such as the distinctive patterns in which places of different types co-occur. Instead, we propose a novel model based on *augmented spatial contexts* that make geographic distance a first-class citizen and adjust these contexts by an information theoretic perspective on the uniqueness of place types within a certain distance as well as their popularity as a proxy for human activities.

- We provide a comprehensive evaluation of different place type embeddings with respect to the top-down Yelp POI category hierarchy. This evaluation essentially brings inductive (bottom-up place type embeddings) and deductive (top-down place hierarchy structure) approaches together.

- We establish two baselines using Amazon's Mechanical Turk Human Intelligence Tasks (HIT) for measuring the similarity and relatedness of place types. Our evaluation result shows that our method has better accuracy than purely linguistically based embeddings, which confirms the importance of explicit spatial contexts. In fact, we demonstrate the remarkable fact that similarity assessments derived from embeddings created exclusively via our augmented spatial contexts, i.e., by merely studying spatial patterns of place types and their relative popularity, correlate strongly with human similarity judgments despite the fact that humans can rely on their rich cultural experience, the meaning of type labels, their background knowledge, and so forth.

- While the resulting place type embeddings can be used for a wide range of tasks that rely on similarity assessments such as commonly used in geographic information retrieval, co-reference resolution and ontology-alignment, as well as recommender system, we introduce a novel perspective, namely *compression*, as an interesting future area of study that deals with the question of whether place types can be substituted or act as proxies for other POI types, e.g., to summarize neighborhoods by a minimal number of place types.

- Finally, we make the embeddings as well as thousands of human similarity assessments from Mechanical Turk available online at `http://stko.geog.ucsb.edu/place2vec` for future use.

The remainder of this paper is organized as follows. Section 3.2 summarizes existing work on embeddings and geospatial semantics. Section 3.3 presents the dataset and provides basic concepts used throughout our work. Section 3.4 explains in detail how we model the augmented spatial contexts. Section 3.5 presents three evaluation schemes and Section 3.6 is evaluation. Finally, Section 3.7 summarizes the research and points to future directions.

3.2 Related Work

Most research on POI embeddings originates from word embedding techniques using neural network language models [106]. One of the most successful models in this class is Word2Vec, which is composed of Skip-Gram and Continuous-Bag-of-Words, proposed by Mikolov et al. [107, 102]. It uses neural networks that take advantage of the distributional semantics of natural languages. Skip-Gram learns the embeddings by predicting context words given center words whereas Continuous-Bag-of-Words does it the other way around.

Previous works on embeddings related to geographic information can be grouped into two categories. The first category considers the influence of geographic context on word embeddings. In a first attempt to investigate the extent to which geographic context affects the semantics of words, Cocos and Callison-Burch [46] trained word embeddings in geolocated tweets using geographic contexts derived from Google Places and OpenStreetMap (OSM). Their work is similar to ours in a sense that they also realize the importance of geospatial contexts, but the scope of their work remains limited to the linguistic domain. In addition, their result shows that geographic context is not as semantically rich as textual context. In contrast, we will demonstrate that *augmented spatial contexts* are indeed rich in semantic information. Zhang et al. [108] also acknowledges the variation in the semantics of words depending on the geographic space. They propose a vector space transformation under different topic distributions in order to generate a mapping between different geographic contexts. Yet again their approach is focusing on linguistic aspects whereas geographic aspects are not directly considered in their model.

The second category is more similar to our work which models geographic entities directly. Yao et al. [109] and Zhang et al. [110] have a very different focus compared to our study as they utilize embedding techniques in order to detect the spatial distribution of urban land use and uncover urban dynamics. We are focusing on exploring the extent to which different adjustment to the spatial context influences the embedding results. Feng et al. [47] and Zhao et al. [111] learn embedding in order to predict future POI visits or recommend POIs. This is a byproduct of the original prediction-based Word2Vec models. Our work has a different focus and therefore does not require temporally sequential data, such as check-in sequences of users. Instead, we are interested in the *semantics* of place types and utilize embeddings as a means to construct representations, share them, and to measure (semantic) similarity across types, e.g., in the context of query expansion [101] and extraction [112].

This relates our work to research on geographic information retrieval and geospatial semantics, and here more specifically to the social sensing framework of *semantic signatures* [16] which characterizes place types based on thematic, temporal, and spatial perspectives called *bands* in analogy to spectral signatures. For example, thematic bands for Points Of Interest have been studied by Adams and Janowicz [113] using Latent Dirichlet Allocation to extract topics from unstructured texts about place types. Quercini and Samet [114] proposes a set of graph-based similarity measures to determine the relatedness of a concept to a location in the Wikipedia link structure. These location-related concepts, which are referred to as *local lexicon* in their work, can be seen as signatures to differentiate geographic entities as well. Research on the temporal perspective has also shown promising results. Ye et al. [115] studied the temporal dimensions of places in the context of location-based social networks. McKenzie and Janowicz [57] applied temporal signature to reverse geocoding to adjust rankings returned by a spatial range search based on a temporal distortion model. So far, the spatial perspective, i.e., the question whether one can learn place (type) representations exclusively from spatial patterns, has received less attention. Mülligann et al. [60] used a measure based on combining point pattern analysis with semantic similarity, while Zhu et al. [65] proposes 27 spatial statistical features to characterize different aspects of place types in digital gazetteers. Our work can be seen as a continuation of this line of research and a contribution to the semantic signatures framework by using novel methods such as augmented spatial contexts to overcome the limitations of previous work. In fact, we will show that these contexts (even when taken on their own) are able to reproduce human similarity judgments, i.e., yield strong correlations between human assessments and our model.

3.3 Preliminaries

The individual Points of Interest and their categories used in this research are from the Yelp Dataset Challenge[3]. This dataset covers venues from 11 different cities from four countries (United Kingdom, Germany, Canada, and the United States). We selected Las Vegas as study region, but our methods can be generalized to different cities and place

[3] https://www.yelp.com/dataset_challenge

type schema; see [116] for a discussion about regional effects. The Yelp dataset groups their 1030 POI types into 22 root categories, such as `Restaurants`, `Shopping`, `Arts & Entertainment`, `Professional Services`, `Health & Medical`, and so forth. Each POI l_i in the POI set L is composed of three parts, a POI name $n \in N$, a geographic identifier (here, latitude and longitude of a place location modeled as centroid) $g \in G$, and a set of associated POI types $\{t_1, t_2, t_3, ..., t_k\} \subseteq T$.

After analyzing the 1030 place types and their frequencies in Las Vegas, we see a long tail in the rank-frequency distribution (Figure 3.1). The log-log plot also shows a linear trend. Fitting $log(frequency)$ and $log(rank)$ using linear regression, yields a value of 0.8543 for R-squared which indicates that the model fits strongly to the data and a p-value of 2.2e$-$16 which indicates that such a scaling effect is highly significant. Simply put, these statistics show that the rank-frequency indeed follows a power law distribution by which a few POI types dominate the data. This is an important motivation for the proposed information content-based frequency adjustment in our augmented spatial contexts discussed in the following section.

3.4 Methods

In this section, we describe the latent representation method and the augmented spatial contexts. The latent representation originates from natural language processing and has been used successfully in many domains. By acknowledging the difference in context formation between geographic space and linguistic expressions, we introduce three approaches to model the geographic influence in determining latent representations. These methods include, naive spatial context, simple augmented spatial context, and Information Theoretic, Distance Lagged (ITDL) augmented spatial context.

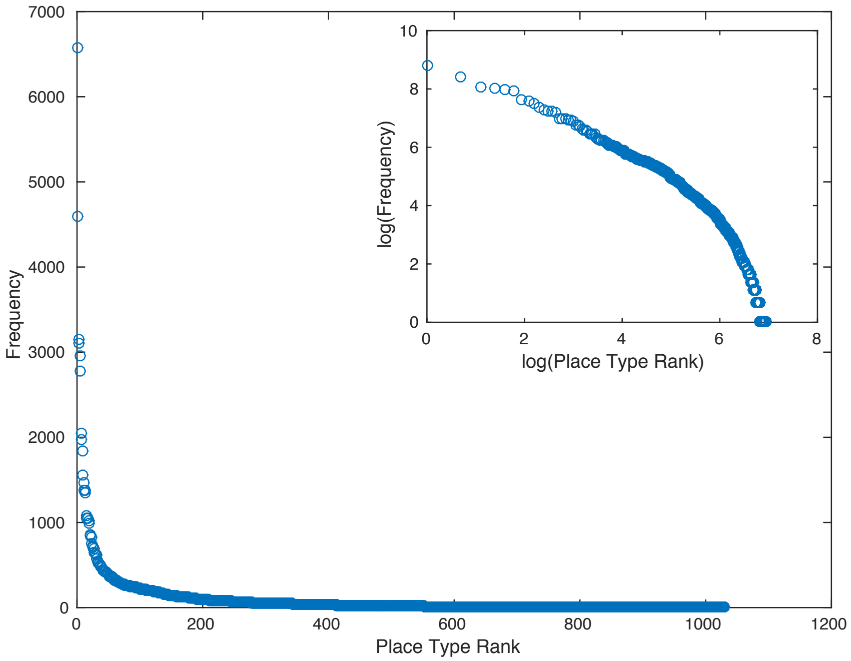


Figure 3.1: POI type rank-frequency and log-log plot.

3.4.1 Latent Representation Method

Recent work has shown that the latent representation model Word2Vec can effectively capture the semantic relationships in word spaces based on the distributional semantics assumption [107, 102]. From analyzing the POI type distribution, we know that, similarly to the word frequency distribution [117], it follows a power law distribution. This leads us to taking advantage of the Word2Vec model and its underlying distributional semantics assumption for the study of POI types in geographic space.

We selected the Skip-Gram model, which predicts context POI types given *center* types. Our objective is to approximate the true place type probability distribution from our training data. A typical approach is to use cross entropy to measure the difference between the learned probability and the true probability. Since our data is discrete and we only care about the center place type, the cross entropy can be simplified as:

$$D(\hat{y}, y) = -y_c log(\hat{y}_c) \tag{3.1}$$

where $\hat{y}$ and y are the learned probability distribution and true probability distribution, respectively. $\hat{y}_c$ is the predicted probability of the context POI types given the center place type (denoted by the index c), and y_c is the true probability of the context POI types given the center place type. $\hat{y}_c$ can be further defined as:

$$\hat{y}_c = P(t_1, t_2, t_3, ..., t_m | t_c) \tag{3.2}$$

where $t_1, t_2, t_3, ..., t_m$ are the context place types and t_c is the center place type. In order to calculate the probability, we apply the Naive Bayes assumption. Note that y_c will always be 1. Finally, we use the softmax function to turn the scores into probabilities and substitute the POI types with vector representations. The objective function is defined as:

$$\text{minimize } J = -log \prod_{t=1}^{m} \frac{exp(u_t^T v_c)}{\sum_{k=1}^{|T|} exp(u_k^T v_c)} \tag{3.3}$$

where u_t and v_c are the context place type vectors and center place type vectors, respectively; $|T|$ is the cardinality of a POI type, i.e., its *extension*. We implement the model in TensorFlow using Mini-Batch Gradient Descent and Noise-Contrastive Estimation [118].

3.4.2 Naive Spatial Context

An intuitive approach to utilize the structure of geographic space is to naively model the spatial context based on the center place type and context place type co-occurrences. We denote the context place type as $t_{context}$ and center place type as t_{center}. This naive method is faithful to the original Word2Vec model and captures the spatial contextual information using a nearest neighbor approach. Unlike natural languages which are sequential in nature, Points of Interest in Yelp are distributed in a 2D geographic space. As a result, instead of using a fixed-size sliding window to construct $(t_{center}, t_{context})$ pairs, we create spatial buffers around each center POI to detect the k-nearest neighbor POIs and record their respective place types as our training pairs. Since each center POI l_i and each context POI l_j can have a set of place types T_{li} and T_{lj} respectively, we use the Cartesian product $T_{li} \times T_{lj} = \{(t_{center}, t_{context}) | t_{center} \in T_{li} \wedge t_{context} \in T_{lj}\}$ to obtain the training pairs for each center POI and candidate context POI. We append these training pairs to the final list of training data SC_{naive}[4] as we iterate through all center and context POIs.

3.4.3 Simple Augmented Spatial Context

Within the naive spatial context the geographic component, namely the distance, is merely used as a criteria to search the neighborhoods and not modeled directly. In this second approach, we augment the naive spatial context by incorporating distance decay and/or aggregated check-in counts (as proxy for the relative popularity or dominance). The rationale behind this approach is that we acknowledge both distance and human activity as essential components in modeling the latent representations of POI types, and, hence, want to study how they can contribute to the final result by modeling them both individually and in combination. Here we define popularity P_{li} of a POI l_i as the number of total check-ins associated with l_i. By augmenting the spatial context, we increase the number of times a $(t_{center}, t_{context})$ tuple appears in our training dataset with a factor of β, where $\beta \in \{n | n \in \mathbb{Z}, n \geqslant 1\}$.

[4] We use SC as an abbreviation for Spatial Context and use different subscripts to denote different types of Spatial Contexts.

For incorporating activity alone, the factor β is defined as:

$$\beta^{lj}_{checkin} = \lceil 1 + ln(1 + P_{lj}) \rceil \tag{3.4}$$

where $\beta^{lj}_{checkin}$ is the augmenting factor for the training tuple $(t_{center}, t_{context})$ when the context POI is l_j. This is an extrinsic augmentation approach.

For incorporating distance decay alone, we define the augmenting factor as:

$$\beta^{lj}_{distance} = \left\lceil \frac{1 + \frac{\sum_{k=1}^{|L|} P_{lk}}{|L|}}{1 + d^{\alpha}(l_i, l_j)} \right\rceil \tag{3.5}$$

where $|L|$ is the total number of POIs, $d(l_i, l_j)$ is the distance between center POI l_i and context POI l_j, and α is an inverse distance factor, set to 1 in our case. The numerator is a smoothing constant for a given POI dataset. This is an intrinsic augmentation approach.

For combining both distance decay and human activities in the spatial context, the augmenting factor, which combines both intrinsic and extrinsic approaches, is defined as:

$$\beta^{lj}_{combined} = \left\lceil \frac{1 + ln(1 + P_{lj})}{1 + d^{\alpha}(l_i, l_j)} \right\rceil \tag{3.6}$$

As one can see, the proposed augmenting factors are based on the check-ins of the context POI as well as the distance from the center POI to the context POI, thus incorporating more geographic information in the spatial context. In fact, the naive spatial context is a special case of the augmented spatial context where the factor β equals to 1. For the simple augmented spatial contexts, our hypothesis is that the *popularity* of a POI as a context has a positive effect on the center POI whereas the influence of a context POI on a center POI decreases as the distance between them increases. By setting an augmenting factor β based on these geographic components, we are stretching the original distribution of POI types in a manner that reveals more latent information in geographic space. To give an intuitive example for our rationale, a single place of the type `Stadiums & Arenas` may dominate a neighborhood while many individual parking spaces and bars only play a supportive function despite their higher frequencies.

3.4.4 ITDL Augmented Spatial Context

While the simple augmented spatial context approach models distance and human activities directly, the augmenting factor only applies to the original spatial context using the k-nearest neighbor method. In this sense, the context POIs are limited to the k nearest neighbors regardless of how far or how close they are from the center POI. However, different place types are likely to follow different spatial distributions and form distinct spatial clusters. For example, places of type `Restaurants` may be located closely to many other places of types such as `Hotels`, `Bars`, and `Department Stores`, generating a dense spatial cluster, while POI of type `Police Departments` and other area-serving places will show very different patterns when compared to nearby places (via their types). This spatial variation means that different spatial context information can be captured within different distances. In addition, the distance we are focusing on rapidly increases for such types, so naively setting a single threshold for the search buffer or the number of nearest neighbors will result in homogeneous spatial contexts for many different place types, thus sacrificing spatial heterogeneity and numerous distinguishing geospatial semantic characteristics. In light of this, we suggest having multiple different spatial contexts for each POI. Inspired by the use of semi-variograms in spatial statistics such as Kriging, we make use of *distance lags*, i.e., discrete bins, for constructing our spatial contexts. Such binning by a given lag also adjusts for the uncertainty (also called *tolerance*) of place centroids. In fact, previous work shows that the median distance of a POI between different database providers, such as Yelp and Foursquare, is 63 meters [57].

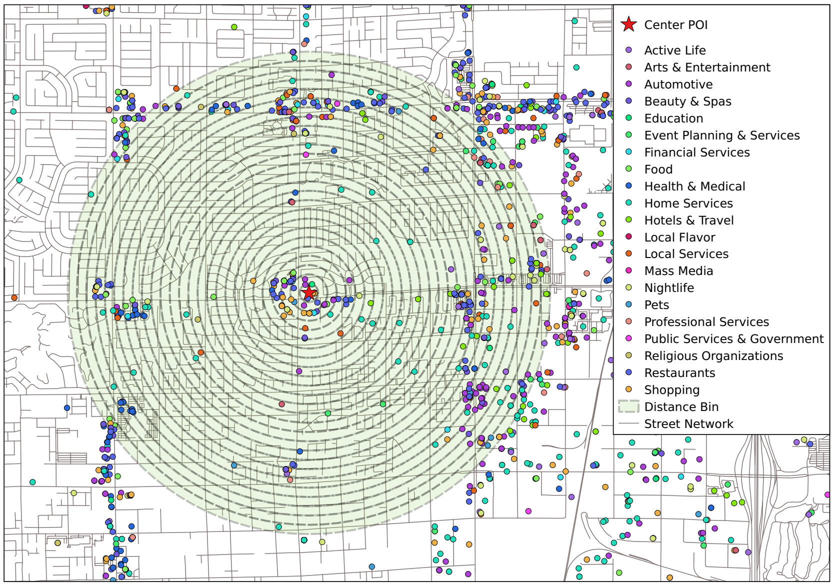


Figure 3.2: ITDL augmented spatial context example.

In the following, we will use a lag distance of $h = 100m$.

We use a default distance bin width for each distance lag, thus generating multiple spatial contexts for the same POI. Each spatial context can be used to learn a latent representation that encodes the distributional semantics between the center POI type and the context POI types within said distance bin. Our rationale behind this approach is that due to the nature (and function) of places and their interaction with other places and regions, an all-encompassing spatial context, even augmented with distance decay and human activities, is not sufficient for understanding the overall variation in the geographic patterns. Instead, we propose to first capture the local context by dividing the continuous geographic space, namely the distance, into discrete lags and then combine the semantic information from these different lags to obtain a more holistic global view of each place type; see Figure 3.2.

Since we aim to capture the spatial interaction between different place types, we want to set the maximum threshold of our spatial context based on this. We define D_{ti} as the set of pair-wise POI distances of the same type t_i. For each POI type t_i, we calculate the minimum intra-class distance $min(D_{ti})$ and use the maximum of these intra-class distances as our threshold TS for the spatial contexts (here the supremum of the per-type infimums):

$$TS = max(min(D_{t1}), min(D_{t2}), min(D_{t3}), ..., min(D_{tn})) \tag{3.7}$$

which is the maximum distance value, for at least one type among all place types, to search for context POIs that will not encounter the same type as the center. This TS value helps to capture as much inter-class spatial interaction as possible. Hence, for each center POI, there are $s = \lfloor \frac{TS}{h} \rfloor$ spatial contexts.

For each spatial context, we propose a novel information theoretic, distance lagged augmentation method. The simple augmented spatial context takes into consideration distance decay and human activities, in the ITDL augmented spatial context, however, we focus on the human activities within the local context as well as the uniqueness of each place types per distance bin. The first component that incorporates human activities is defined as:

$$A = -log_2\left(1 - \frac{P_{tj}}{1 + \sum_{k=1}^{|M|} P_{tk}^h}\right) \tag{3.8}$$

where P_{tj} is the *popularity* (check-in counts) of a place type t_j and $\sum_{k=1}^{|M|} P_{tk}^h$ is the total number of check-in counts of all place types within a distance bin with width h. This is a monotonically increasing function with respect to $\frac{P_{tj}}{1+\sum_{k=1}^{|M|} P_{tk}^h}$, which means that if a place type has high *popularity* among all place types within the bin, this component value will be very high. The second component adopts the idea of information content (here, *surprisal*) from information theory to model the uniqueness of a place type given a distance bin:

$$U = -log_2(F^h_{tj}) \tag{3.9}$$

where F^h_{tj} is the probability of encountering place type t_j in a distance bin. U essentially represents the information content of a place type t_j within a distance bin. Larger F^h_{tj} values will result in reduced information content. Finally, we integrate these two components using a convex combination and our ITDL augmentation is defined as:

$$\beta^{lj}_{ITDL} = \lceil \omega A + (1 - \omega)U \rceil \tag{3.10}$$

where ω and $1 - \omega$ are the weights for the components. Intuitively, this allows us to distinguish unique places (of a certain type) that are highly popular from places that are popular in virtue of their type. Algorithm 1 shows the detailed procedures to construct the ITDL augmented spatial context SC_{ITDL}. In order to improve the efficiency of this algorithm, we split the whole task into s tasks that can run in parallel, thus each worker only constructs a spatial context for one distance bin. In short, for the ITDL augmentation method, we use individual context settings to capture extrinsic components such as the popularity and the uniqueness of place types and use multiple spatial context bins combined to capture the intrinsic components such as distance and spatial variation.

3.5 Evaluation Schemes

In this section, we introduce three different ground truths that we establish to evaluate our proposed methods. These ground truth results can also be used to evaluate other tasks involving place type similarity and relatedness. The first ground truth is

Algorithm 1: Constructing ITDL-based Augmented Spatial Contexts SC_{ITDL}

```
Input  : $L = (N, G, T)$, $s$, $h$, $\omega$
Output: $SC_{ITDL}$
$SC_{ITDL}$ := initialize list
foreach $l_i \in L$ do
    $T_{li}$ := a set of place types associated with $l_i$
    for $n = 0; n < s; n$++ do
        $sc$ := check-in total of all place types in bin $n$
        $sp$ := POI total of all place types in bin $n$
        foreach $l_j \in L$ do
            $T_{lj}$ := a set of place types associated with $l_j$
            if $nh \leqslant d(l_i, l_j) < (n+1)h$ then
                foreach $t_{ki} \in T_{li}$ do
                    foreach $t_{kj} \in T_{lj}$ do
                        $cc$ := check-in of $t_{kj}$
                        $cp$ := count of $t_{kj}$
                        $A := -log_2(1 - cc/sc)$
                        $U := -log_2(cp/sp)$
                        $aug$ := ceil$(\omega A + (1-\omega)U)$
                        append tuple $(t_{ki}, t_{kj})$ to $SC^n_{ITDL}$ $aug$ times
                    end
                end
            end
        end
    end
end
```

built from the original Yelp place type hierarchy.[5] We take advantage of this *top-down* hierarchy and evaluate to what degree our *bottom-up* approaches can approximate Yelp's hierarchy. The second ground truth is obtained using Human Intelligence Tasks (HIT) via Amazon Mechanical Turk which is a binary test. The third one is obtained from another HIT which provides similarity and relatedness rankings for different POI types. These three ground truth results, one using top-down information from Yelp and the other two provided by human judges, provide a comprehensive evaluation for our work.

[5] https://www.yelp.com/developers/documentation/v3/all_category_list/categories.json

3.5.1 Hierarchy-based Evaluation Scheme

The original Yelp categories provide us with a natural way to calculate the similarity and relatedness of different POI types based on their hierarchical structure. There are two major ways to measure (semantic) similarity and relatedness for our tasks: distribution-based measures and knowledge-based measures [119]. While our proposed methods aims to capture the distributional semantics, the evaluation scheme derived from Yelp categories falls into the knowledge-based measures group. Numerous models have been proposed for such measures. In summary, edge-based measures and information content-based measures are two widely-used subgroups. In our study, we choose two measures from each subgroup to form our evaluation scheme. In addition, since the information content-based measures depend on the definition of information content, we also select two different definitions of information content in order to provide a more holistic evaluation scheme. In the end, we have 6 different measurements based on the Yelp hierarchy.

The first edge-based measurement is proposed by Wu & Palmer [120], which is defined as:

$$SIM_{WP}(t_1, t_2) = \frac{2N_3}{N_1 + N_2 + 2N_3} \tag{3.11}$$

t_{lcs} is defined as the least common superclass of place types t_1 and t_2. N_1 is the shortest path from t_1 to t_{lcs}. $N2$ is the shortest path from t_2 to t_{lcs}. N_3 is the shortest path from t_{lcs} to root. The second edge-based measurement is proposed by Leakcock & Chodorow [121]:

$$SIM_{LC}(t_1, t_2) = -log\Big(\frac{N}{2D}\Big) \tag{3.12}$$

where D is the maximum depth of the taxonomy and N is the shortest path between place types t_1 and t_2.

For the information content-based measurements, we use the models proposed by

Lin [122] and Jiang & Conrath [123]. Their definitions are shown in Eq. 3.13 and Eq. 3.14, respectively. IC is the information content of each place type and t_{lcs} is the least common superclass of place types t_1 and t_2 within the Yelp hierarchy. Jiang & Conrath's method calculates the distance between t_1 and t_2, so the similarity is equal to $SIM_{JC}(t_1, t_2) = 1/DIS_{JC}(t_1, t_2)$.

$$SIM_{Lin}(t_1, t_2) = \frac{2IC(t_{lcs})}{IC(t_1) + IC(t_2)} \tag{3.13}$$

$$DIS_{JC}(t_1, t_2) = IC(t_1) + IC(t_2) - 2IC(t_{lcs}) \tag{3.14}$$

Both models proposed by Lin and Jiang & Conrath depend on the definition of information content, so we also include two different definitions of information content that can be calculated from the place type hierarchy. The information content proposed by Sánchez et al. [124] is defined as:

$$IC_{Sanchez} = -log\left(\frac{\frac{|leaves(t_i)|}{|subsumers(t_i)|} + 1}{max\_leaves + 1}\right) \tag{3.15}$$

where $|leaves(t_i)|$ is the number of leaves of place type t_i in the hierarchy, $|subsumers(t_i)|$ is the number of place types that are more general than t_i in the hierarchy and max_leaves is the number of leaves for the root place type. The information content proposed by Seco et al. [125] is defined as:

$$IC_{Seco} = 1 - \frac{log(|hypo(t_i)| + 1)}{log(max\_types)} \tag{3.16}$$

where $|hypo(t_i)|$ is the number of POI types that are more specific than t_i and max_types is the maximum number of types in the hierarchy. Combining these definitions of information content with the methods by Lin and Jiang & Conrath, leads to four measures.

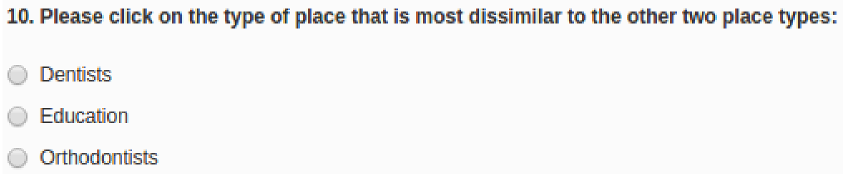
By using these semantic similarity measures, we calculate the pair-wise similarity of Yelp place types. Because these six measures differ in terms of what they measure, the resulting scores are also slightly different. Based on the similarity scores, for each place type, we generate a ranking of similar place types from the most similar to the least similar. We obtain six different groups of rankings for each of the POI types in Yelp. To confirm the validity of this evaluation scheme, we use Kendall's coefficient of concordance W to assess the agreement among these six groups of rankings. The average Kendall's W of all (1030) place types [6] among the six measurements is **0.981**, indicating a nearly perfect agreement among measures. Moreover, in our experiment, we use a subset of 93 place types (see Section 3.6) and the concordance remains stable at **0.979**. This result implies that our evaluation scheme based on the place type hierarchy is valid. To evaluate the result, we mimic the task of geographic information retrieval, e.g. finding the most similar place type based on a given place type. By choosing the first place type in each of the 1030 rankings, we can obtain the result for all six measurements. To evaluate our latent representations, we generate our own rankings of each place type based on the augmented spatial contexts using pair-wise similarity [7] and use Mean Reciprocal Rank (MRR) to test the performance of our methods.

3.5.2 Binary HIT Evaluation Scheme

The hierarchy-based evaluation scheme has some potential drawbacks. First, the hierarchy is created by a small set of people which may lead to a bias. Moreover, in this hierarchy of more than 1000 place types (nodes), the average path length is only 1.73 which indicates that the taxonomy is very shallow. This will result in ties in the rankings generated using the hierarchical structure. Finally, a hierarchy always encodes

[6] We only consider 570 place types, namely those that have at least 14 instances in our dataset and use various subsets of these 570 types in our experiments.

[7] All similarity scores for our place type embeddings are calculated using Cosine Similarity.



10. Please click on the type of place that is most dissimilar to the other two place types:

○ Dentists

○ Education

○ Orthodontists

Figure 3.3: Binary HIT example.

some underlying ontological commitments, e.g., grouping arts and entertainment in a common class. Hence, in addition to the hierarchy-based evaluation, we utilize Amazon's Mechanical Turk for a binary HIT evaluation scheme.

For the HIT task, we generate 80 triplets with each element in the triplet being a place type. For example, one of the triplets is `(Dentists, Education, Orthodontists)`. [8] The task is to choose the place type from each triplet that is most dissimilar from the other two. For each place type in the triplet, a human judge will make a binary decision; see Figure 3.3. We published the HIT task on Amazon Mechanical Turk and each of these 80 tests was done by 25 human workers. The final result of each test is determined by the mode answer of the 25 human workers. For instance, the final answer for the test `(Dentists, Education, Orthodontists)` is `Education` as this is the most often excluded type.

To evaluate the latent representations generated by augmented spatial contexts, for each triplet, we calculate the pair-wise similarity score using 2-combination. For example, for the above mentioned triplet, we calculate the similarity scores of three pairs `(Dentists, Education)`, `(Dentists, Orthodontists)` and `(Education, Orthodontists)`. We pick the one with the highest score and return the other place type as the result for this test using our methods. For instance, if `(Dentists, Orthodontists)` has the highest score, then `Education` is the result from our methods. We evaluate the accuracy of different methods on all triplets.

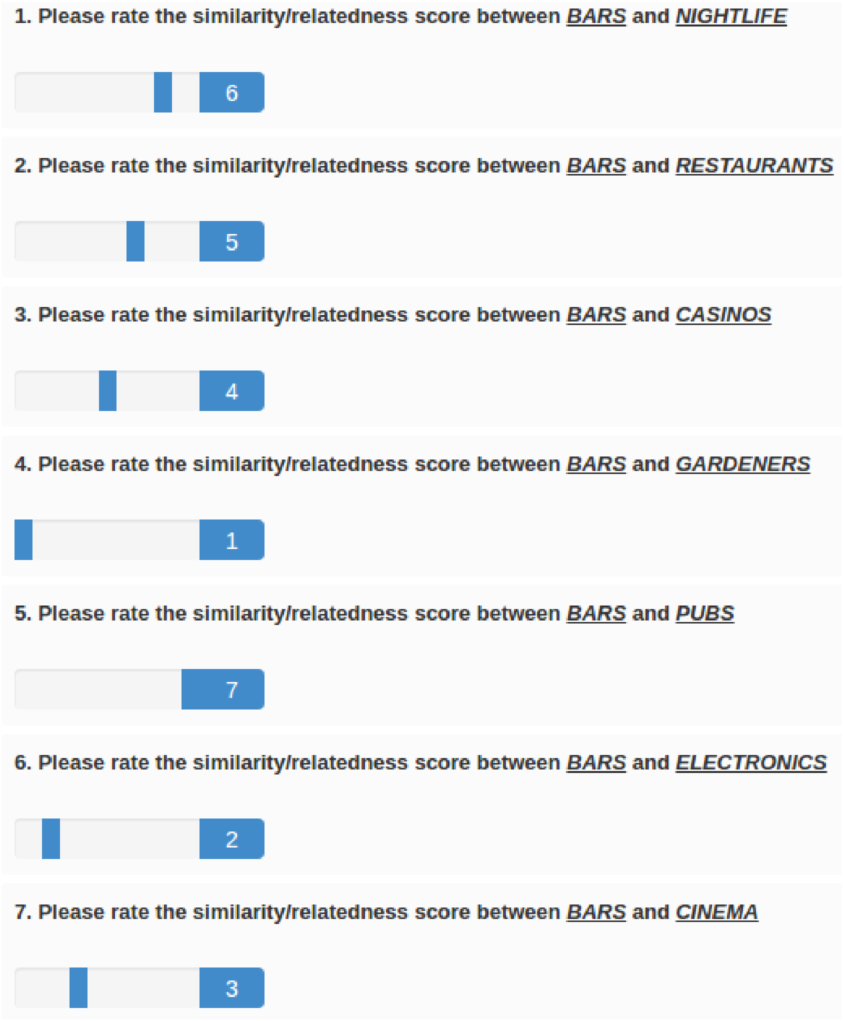
[8] See Goodman's deliberation on similarity for a rationale about using triples [126].

3.5.3 Ranking-based HIT Evaluation Scheme

While the binary-based HIT evaluation can complement the Yelp hierarchy task by relying on human judges, the task is relatively easy. Hence, for the ranking-based HIT evaluation scheme, we want to use human judges to generate a ranking result for each place type. We selected 10 place types and for each place type we selected 7 candidate place types for ranking, so altogether we have 70 POI type pairs. We ask 25 human judges on Amazon Mechanical Turk to rate on a scale of 1-7 the similarity of each of these pairs. Such task can be considered as very challenging in the context of studying semantic similarity [101] and requires more attention to user interface design (Fig. 3.4) to adjust for some well-known characteristics of human similarity judgments, notably that such judgments are known to be non-symmetric. In addition, we selected a slider-based design to ease visual comparison between pairs; see [126].

After receiving the results, we have rankings of each place type from 25 human judges. In order to check if the rankings are consistent, and, thus, whether the task is meaningful, we use Kendall's coefficient of concordance W to evaluate the agreement score among the judges. The average Kendall's W score over all place types in the test is **0.79** which indicates very high agreement.

In order to evaluate our place embeddings using the proposed augmented spatial contexts, we generate a ranking for each place type based on the pair-wise similarity score. We then calculate the average Spearman's rank correlation coefficient between our rankings and the rankings from the HIT task as the criteria to evaluate the performance of our models.



1. Please rate the similarity/relatedness score between *BARS* and *NIGHTLIFE*

6

2. Please rate the similarity/relatedness score between *BARS* and *RESTAURANTS*

5

3. Please rate the similarity/relatedness score between *BARS* and *CASINOS*

4

4. Please rate the similarity/relatedness score between *BARS* and *GARDENERS*

1

5. Please rate the similarity/relatedness score between *BARS* and *PUBS*

7

6. Please rate the similarity/relatedness score between *BARS* and *ELECTRONICS*

2

7. Please rate the similarity/relatedness score between *BARS* and *CINEMA*

3

Figure 3.4: Ranking-based HIT, showing one MTurk result.

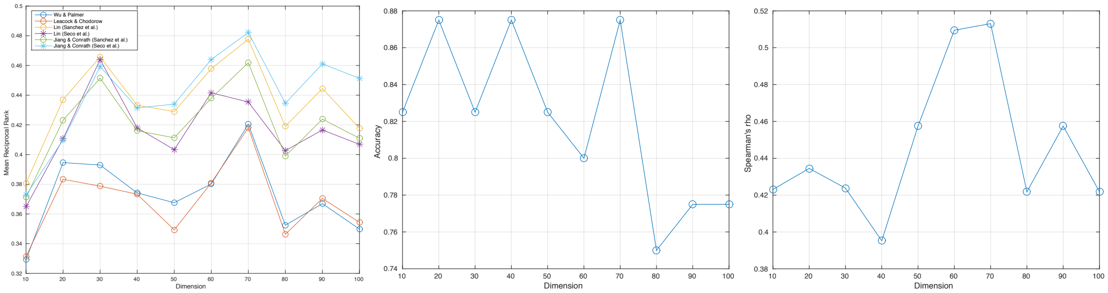


Figure 3.5: Left to right, Mean Reciprocal Rank (MRR) for the hierarchy-based evaluation, accuracy for the binary HIT evaluation, and Spearman's ρ for the ranking-based HIT evaluation.

3.6 Experiment and Result

In this section, we discuss the experiments to evaluate our work and their results. We also point to an interesting research question that arises from our work. First, we have to define the number of dimensions for the POI type embeddings. Next, we compare our embeddings with the state-of-the-art word embeddings trained from the Google News corpus as a baseline using the proposed evaluation schemes in order to reiterate the necessity of augmenting spatial contexts to obtain richer semantic information from geographic space. In addition, we visualize and analyze different embeddings spaces from different augmented spatial contexts using dimension reduction techniques and present *place type profile* as a visual assistance tool for understanding place type similarity and relatedness. Finally, we briefly look at a very interesting research question that arises from our work, namely whether there is potential for compression by merely using a *subset* of POI types to learn *all* POI types. From an urban planning perspective, this question can also be framed from a summarization perspective, by asking whether there are certain place types that are *indicative* of a neighborhood (when modeled as a set of POI) .

3.6.1 Selecting Dimensions

An important parameter for latent representation models is the number of dimensions for the embedding vectors. As the total number of place types is relatively small compared with the vocabulary size of natural language, we selected dimensions ranging from 10 to 100 with a step interval of 10 to determine the number of optimal dimensions for our model. Since we want to combine both intrinsic and extrinsic information in our spatial context, we focus on using the augmenting factor $\beta^{lj}_{combined}$ in this task, which takes into consideration the influence of geographic distance and POI *popularity.* Figure 3.5 shows the dimension test result using the Yelp hierarchy-based evaluation scheme, the binary HIT test, and the ranking-based HIT. Although there is a variation in the absolute values of the six measurements, the overall trend is very similar. It shows that using **70** dimensions yields the best overall results and we will use this number for the experiments described below.

3.6.2 Comparison

By introducing the augmented spatial contexts, we want to demonstrate the richness of semantic information latently encoded in geographic patterns. First, to justify the need for POI type embeddings, we compare the evaluation results of the word embeddings trained from the Google News corpus with the place type embeddings trained from Yelp POIs and our augmented spatial contexts. Word embeddings have been used in a variety of information retrieval tasks and have been frequently used as proxies for geographic information retrieval. Many of the word embeddings techniques, however, only consider unigrams, such as the pre-trained Word2Vec embeddings from Google, which means that they are not suitable for many place type names, such as `Auto Repair`. In addition, and as argued above, geographic space is inherently different from word space, and,

Table 3.1: Mean Reciprocal Rank for the hierarchy-based evaluation.

Model	SIM_{WP}	SIM_{LC}	SIM_{Lin} ($IC_{Sanchez}$)	SIM_{Lin} (IC_{Seco})	SIM_{JC} ($IC_{Sanchez}$)	SIM_{JC} (IC_{Seco})
Word2Vec	0.288	0.321	0.354	0.334	0.349	0.333
SC_{naive}	0.412	0.398	0.474	0.442	0.455	0.478
$SC_{checkin}$	0.385	0.387	0.448	0.428	0.452	0.474
$SC_{distance}$	0.381	0.396	0.458	0.426	0.443	0.458
$SC_{combined}$	0.420	0.418	0.478	0.435	0.462	0.482
SC_{ITDL}	**0.447**	**0.431**	**0.498**	**0.479**	**0.487**	**0.483**

thus, word embeddings lack the ability to capture spatial interaction among different geographic entities and distance (decay) effects which is a significant factor in measuring place type similarity and relatedness.

In order to support our argument, we compared the word embeddings with the proposed place type embeddings using different spatial contexts, namely one with the naive spatial context and four with the augmented spatial contexts. Recall that there is a weight parameter ω in the ITDL augmented spatial contexts, to adjust the relative importance of A (activity) and U (uniqueness). We tested our model with ω values ranging from 0.1 to 1 with 0.1 as step interval. Our TS value is 2644.5 meters, so the total number of spatial contexts for each ω value for the ITDL approach and a lag of 100m is $s = \lfloor 2644.5/100 \rfloor = 26$. In the end, we can obtain 234 different augmented spatial contexts and learn place type embeddings from each of these contexts using parallel threads. In order to compare the evaluation results, for each ω value, we test the performance of each of the 26 bins and concatenate the embedding vectors of the top five bins to generate the final place type embedding of 350 dimensions. We use the best ω values as our final result of the ITDL augmented spatial contexts.

We compared the pre-trained Google Word2Vec result with our place type embeddings using both the hierarchy-based evaluation scheme and the binary HIT evaluation scheme. SC_{naive} is the spatial context without augmentation. $SC_{checkin}$, $SC_{distance}$, $SC_{combined}$ and SC_{ITDL} are the methods detailed in Section 3.4. Table 3.1 shows the result of the hierarchy-based evaluation. As mentioned earlier, word embeddings trained using Google News corpus only contain unigrams, so we select a subset (93 place types) as our testing data. All methods are tested using the six measures described in Section 3.5. Table 3.2 shows the binary and ranking-based HIT results. The hierarchy and binary evaluations show that the results obtained by using spatial contexts, even without any augmentation, are substantially better than the one purely based on a linguistic perspective, thereby also showing the benefits of our approach over previous work outlined in Section 3.2. This confirms our hypothesis that geographic space carries rich latent semantic information that cannot be captured by the word space alone. For the ranking-based evaluation scheme, we dropped the Google Word2Vec embeddings to be able to use bigrams and because using a merely linguistic context already did not perform well for the two simpler tasks. In all three evaluations the ITDL augmented spatial contexts is able to model more semantic information, and, thus, yields better results for the place type similarity tests. With a ρ of **0.7**, i.e., a strong correlation with human judgments, and an accuracy of **0.95** this becomes most apparent for the more difficult HITs. This is a remarkable result as humans utilize substantially richer information to reason about similarity, e.g., the meaning (and similarity) of the type labels, background knowledge, e.g., about cultural and historic reasons why Asian foods are alike, and so forth. Financially, it is worth mentioning that short as well as long-distance bins contribute to these results, e.g., the highest ρ is obtained by a concatenation of bins 4-17-1-5-24 ($\omega = 0.1$), where 24 represents the 100m distance lag at 2400 meters from the center POI.

Table 3.2: Accuracy for binary HIT evaluation and Spearman's ρ for ranking-based HIT.

Model	Accuracy
Word2Vec	0.750
SC_{naive}	0.850
$SC_{checkin}$	0.700
$SC_{distance}$	0.875
$SC_{combined}$	0.875
SC_{ITDL}	**0.950**

Model	ρ
SC_{naive}	0.56
$SC_{checkin}$	0.56
$SC_{distance}$	0.57
$SC_{combined}$	0.51
SC_{ITDL}	**0.70**

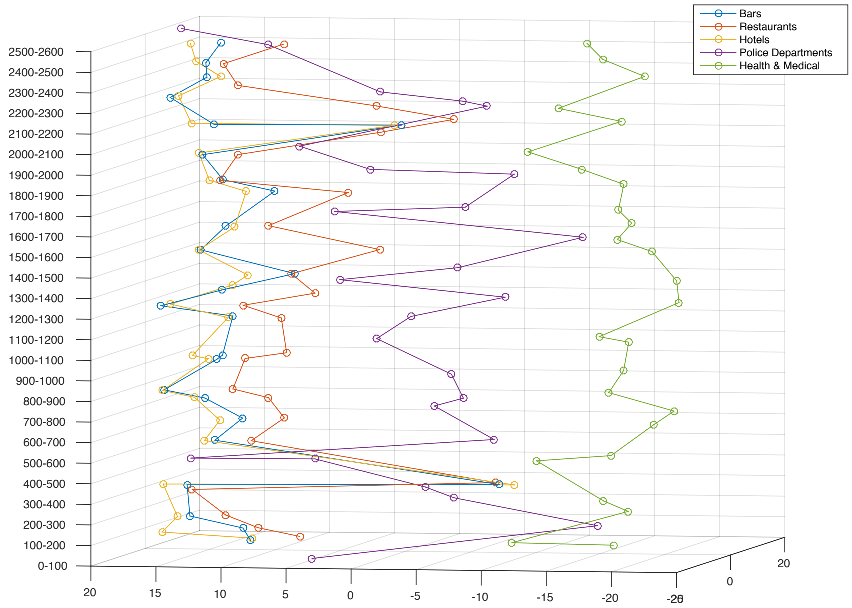


Figure 3.6: Place Type Profile with $\omega = 0.5$.

3.6.3 Place Type Profiles

Although we use the concatenated place type embeddings in our evaluation, individual augmented spatial context can be used separately for analyzing the characteristics of different place types. Here we propose a 3D visualization, namely *place type profile* as a tool to compare different POI types and their semantic relationships. We use t-Distributed Stochastic Neighbor Embedding (t-SNE) [127] to reduce our place type embeddings in each distance bin into two dimensions, then stack each of these 2D space together to build a 3D profile. Figure 3.6 shows the profiles of selected types generated with $\omega = 0.5$, the x-axis and y-axis are the two components after dimension reduction using t-SNE and the z-axis is the distance bin.

One can see that `Bars`, `Restaurants` and `Hotels` always cluster together no matter which distance bin they are in. `Police Departments` are a certain distance apart in each bin. `Health & Medical` remains far away from all other POI types. This pattern shows that `Bars`, `Restaurants`, and `Hotels` have very similar contexts in each distance bin,

Table 3.3: Place type compression result.

Model	Accuracy	ρ
All Place Types	0.950	0.70
W/O Restaurants	0.925	0.70
W/O Nightlife	0.925	0.70
W/O Professional Services	0.925	0.68
W/O Health & Medical	0.900	0.68
W/O 18 Place Types	0.875	0.59

which implies that they interact in similar ways with other POI type. We will return to this argument when discussing compression potential next.

3.6.4 Place Type Compression

So far, our experiments are all based on all POI types, which means that we generate our training data for each augmented spatial context using all types and run the latent representation model to retrieve place type embeddings. However, this approach is time-consuming as the number of $(t_{center}, t_{context})$ pairs increases in later distance bins and may also lead to overfitting. In order to obtain more condensed results, we proposed the novel idea of place type compression. Our intuition is that many place types such as `Restaurants` and `Nightlife` are co-located with other types (via their POI) following similar patterns. Hence, our hypothesis is that these types can serve as proxies in the sense that we can omit, for instance, all nightlife places (and places of their 17 subtypes) and still learn good embeddings for *all* types including `Nightlife`. Some place types such as `Professional Services` have weaker interaction patterns with other place types, thus making it harder to represent them by other POI types.

In order to test our hypothesis, we select four different root place types: `Restaurants`, `Nightlife`, `Professional Services`, and `Health & Medical`. We remove each of these place types and their subtypes from the context POI types in our training and run our models using the ITDL augmented spatial contexts. In addition, we run our model by removing all 18 place types aside of those four (there are 22 root place types). The accuracy result of the binary HIT evaluation and the Spearman's ρ result of the ranking-based HIT are shown in Table 3.3. The result shows that dropping either `Restaurants` or `Nightlife` does not have much effect on the final embeddings while dropping either `Professional Services` or `Health & Medical` will result in a (small) decrease in performance. Consequently, given the 570 studied types, removing even 69 from them, e.g., by removing the `Restaurants` supertype, leaves us with enough proxy types, i.e., types that interact with other types in similar ways. Dropping 18 place supertypes, however, and trying to generate embeddings merely on the 4 remaining supertypes will result in a substantial decrease. This confirms our hypothesis that we can compress our model and still obtain high-quality latent representations of place types.

3.7 Conclusion and Future Work

In this research we proposed a novel approach, namely augmented spatial contexts, to capture the semantics of place types by learning vector embeddings and using them to reason about place type similarity and relatedness, a common prerequisite for geographic information retrieval. By comparing the place type embeddings generated using the proposed methods with state-of-the-art word embeddings, we were able to show that our information-theoretic, distance lagged augmented spatial contexts substantially outperform the baseline and better capture the latent semantic information. We also established three different evaluation schemes to systematically evaluate the resulting POI embeddings. We published the embeddings as well as the HIT results online to foster reproducibility and in the hope that they will be reusable by others working on vector representations of place types. We used place type profiles as a way to visualize the semantic relationship among different place types. Finally, we outlined the idea of indicative POI types and their usage in compression as a novel research avenue.

In the future, we will explore place type compression in more detail to determine how different combinations of POI types can affect the quality of the overall place type embeddings and will follow up on the idea of using them to summarize neighborhoods. Finally, we focused on geodesic distance here but our methods can be generalized, e.g., using L1 distance (taxicab), in future work.

Chapter 4

xNet+SC: Classifying Places Based on Images by Incorporating Spatial Contexts

This chapter focuses on the multimedia leaf nodes (e.g. images) in geographic knowledge graphs. Since the labels for these images are typically not provided, we take advantage of the place type labels of surrounding images and utilize the hidden pattern in the geospatial contextual information to help classify images. Using re-ranking and Bayesian methods, we explore different ways in which geospatial contextual information can be incorporated. By combining visual stimuli with the help of the state-of-the-art convolutional neural networks (such as AlexNet, ResNet, and DenseNet) and spatial contexts (spatial relatedness, spatial co-location, and spatial sequence patterns), our model is able to improve classification accuracy. The model can be applied to label the leaf nodes considering geographic entities in the neighborhood for geographic knowledge graphs. Such information can then be used to inform different ways to select different multimedia leaf nodes in the summarization process.

Peer Reviewed Publication	
Title	xNet+SC: Classifying Places Based on Images by Incorporating Spatial Contexts
Authors	Bo Yan, Krzysztof Janowicz, Gengchen Mai, and Rui Zhu
Venue	GIScience 2018
Editors	Stephan Winter, Monika Sester, Amy L. Griffin
Publisher	Schloss Dagstuhl – Leibniz-Zentrum für Informatik
Pages	17:1 – 17:15
Submission Date	February 16, 2018
Acceptance Date	April 20, 2018
Publication Date	August, 2018
Copyright	Reprinted with permission from Schloss Dagstuhl – Leibniz-Zentrum für Informatik

Abstract With recent advancements in deep convolutional neural networks, researchers in geographic information science gained access to powerful models to address challenging problems such as extracting objects from satellite imagery. However, as the underlying techniques are essentially borrowed from other research fields, e.g., computer vision or machine translation, they are often not spatially explicit. In this paper, we demonstrate how utilizing the rich information embedded in spatial contexts (SC) can substantially improve the classification of place types from images of their facades and interiors. By experimenting with different types of spatial contexts, namely spatial relatedness, spatial co-location, and spatial sequence pattern, we improve the accuracy of state-of-the-art models such as ResNet – which are known to outperform humans on the ImageNet dataset – by over 40%. Our study raises awareness for leveraging spatial contexts and domain knowledge in general in advancing deep learning models, thereby also demonstrating that theory-driven and data-driven approaches are mutually beneficial.

4.1 Introduction

Recent advancements in computer vision models and algorithms have quickly permeated many research domains including GIScience. In remote sensing, computer vision methods facilitate researchers to utilize satellite images to detect geographic features and classify land use [128, 129]. In urban planning, researchers collect Google Street View images and apply computer vision algorithms to study urban change [130]. In cartography, pixel-wise segmentation has been adopted to extract lane boundary from satellite imagery [131] and deep convolutional neural network (CNN) has been utilized to recognize multi-digit house numbers from Google Street View images [132]. These recent breakthroughs in computer vision are achieved, in equal parts, due to advances in deep neural networks as well as the ever-increasing availability of extensive training datasets. For example, the classification error in the latest image classification challenge using the ImageNet dataset is down to about 0.023.[1]

However, such impressive results do not imply that these models have reached a level in which no further improvement is necessary or meaningful. On the contrary, such deep learning models which primarily depend on visual signals are susceptible to error. In fact, studies have shown that deep (convolutional) neural networks suffer from a lack of robustness to adversarial examples and a tendency towards biases [41]. Researchers have discovered that, by incorporating adversarial perturbations of inputs that are indistinguishable by humans, the most advanced deep learning models which have achieved high accuracy on test sets can be easily fooled [133, 134, 135]. In addition, deep learning models are also vulnerable to biased patterns learned from the available data and these biases usually resemble many unpleasant human behaviors in our society. For instance, modern neural information processing systems such as neural network language models and deep convolutional neural networks have been criticized for amplifying racial and gender biases [136, 137, 41, 138]. Such biases, which can be attributed to a discrepancy between the distribution of prototypical examples and the distribution of more complex real world systems [42], have already caused some public debates. To give a provocative example, almost three years after users revealed that Google erroneously labeled photos of black people as "gorillas", no robust solutions have been established besides simply removing such labels for now. [2]

The above-mentioned drawbacks are being addressed by improvements to the available training data as well as the used methods [139, 136]. In our work, we follow this line of thought to help improve image classification. In our case, these images depict the facades or interiors of different types of places, such as restaurants, hotels, and libraries. Classifying images by place types is a hard problem in that more often than not the training image data is inadequate to provide a full visual representation of different place types. Solely relying on visual signals, as most deep convolutional neural networks do, falls short in modeling the feature space as a result. To give an intuitive example, facades of restaurants may vary substantially based on the type of restaurant, the target customers, and the surrounding. Their facade may be partially occluded by trees or cars, may be photographed from different angles and at different times of the day, and the image may contain parts of other buildings. Put differently, the principle of spatial heterogeneity implies that there is considerable variation between places of the same type.

To address this problem and improve classification accuracy, we propose to go beyond visual stimuli by incorporating spatial contextual information to help offset the visual representational inadequacy. Although data availability is less of an issue nowadays, the biased pattern in the data poses a real challenge, especially as models such as deep convolutional neural networks take a very long time to train. Instead of fine-tuning the parameters (weights) by collecting and labeling more unbiased data, which are very

[2] https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/

resource-consuming, we take advantage of external information, namely spatial context. There are many different ways one can model such context; in this work, we focus on the types of nearby places. We explore and compare the value of three different kinds of spatial context, namely spatial relatedness, spatial co-location, and spatial sequence pattern.

We combine these context models with state-of-the-art deep convolutional neural network models using search re-ranking algorithms and Bayesian methods. The result shows that, by considering more complex spatial contexts, we can improve the classification accuracy for different place types. In fact, our results demonstrate that a *spatially explicit* model [140], i.e., taking nearby places into account when predicting the place type from an image, improves the accuracy of leading image classification models by at least 40%. Aside from this substantial increase in accuracy, we believe that our work also contributes to the broader and ongoing discussion about the role of and need for theory, i.e., domain knowledge, in machine learning. Finally, and as indicated in the title, our spatial context (SC) models, can be added to any of the popular CNN-based computer vision models such as AlexNet, ResNet, and DenseNet – abbreviated to *xNet* here.

The remainder of this paper is organized as follows. Section 4.2 provides an overview of existing work on spatial context and methods for incorporating spatial information into image classification models. Section 4.3 presents the image classification tasks and provides information about the convolutional neural network models used in our study. Section 4.4 explains in detail three different levels of spatial context and ways to combine them in image classification models. Section 4.5 presents the results. Finally, Section 4.6 concludes the research and points to future directions.

4.2 Related Work

There is a large body of work that utilizes spatial context to improve existing methods and provide deeper insights into the rich semantics of contextual information more broadly. For instance, spatial context has been recognized as a complementary source of information in computational linguistics. By training word embeddings for different place types derived from OpenStreetMap (OSM) and Google Places, Cocos and Callison-Burch [46] suggested that spatial context provides useful information about semantic relatedness. In Points of Interest (POI) recommendation, spatial context has been used to provide latent representations of POI, to facilitate the prediction of future visitors [47], and to recommend similar places [111]. By implementing an information theoretic and distance-lagged augmented spatial context, Yan et al. [48] demonstrated that high-dimensional place type embeddings learned using spatial contexts can reproduce human-level similarity judgments with high accuracy. The study showed that such a spatially explicit Place2Vec model substantially outperforms Word2Vec-based models that utilize a linguistic-style of context. Liu et al. [141] used spatial contexts to measure traffic interactions in urban area. In object detection, Heitz and Koller [142] leveraged spatial contexts in a probabilistic model to improve detection result. Likewise, by embracing the idea that spatial context provides valuable extrinsic signals, our work analyzes different kinds of spatial contexts and tests their ability to improve image classification of place types.

Existing work on image classification has realized the importance of including a geographic component. One direction of research focused on enriching images with geospatial data. Baatz et al. [143] took advantage of digital elevation models to help geo-localize images in mountainous terrain. Lin et al. [144] made use of land cover survey data and learned the complex translation relationship between ground level images and overhead imagery to extend the reach of image geo-localization. Instead of estimating a precise geo-tag, Lee et al. [145] trained deep convolutional neural networks to enrich a photo with geographic attributes such as elevation and population density. Another direction of research (which is more similar to our study) focused on utilizing geographic information to facilitate image classification. In order to better understand scenes and improve object region recognition, Yu and Luo [146] exploited information from seasons and location proximity of images using a probabilistic graphical model. Berg et al. [61] combined one-vs-most image classifiers with spatiotemporal class priors to address the problem of distinguishing images of highly similar bird species. Tang et al. [62] encoded geographic features extracted from GPS information of images into convolutional neural networks to improve classification results.

Our work differs from the existing work in that we explicitly exploit the distributional semantics found in spatial context [48] to improve image classification. Following the linguistic mantra that one *shall know a word by the company it keeps*, we argue that one can know a place type by its neighborhood's types. This raises the interesting question of how such a neighborhood should be defined. We will demonstrate different ways in which spatial contextual signals and visual signals can be combined. We will assess to what extent different kinds of spatial context, namely spatial relatedness, spatial co-location, and spatial sequence pattern, can provide such neighborhood information to benefit image classification.

4.3 Image Classification

In this section, we first describe the image classification task and the data we use. The task is similar to scene classification but we are specifically interested in classifying different business venues as opposed to natural environment. Then we explain four different deep convolutional neural networks that solely leverages the visual signals of images. These convolutional neural network models are later used as baselines for our experiment.

4.3.1 Classification Task

Our task is to classify images into one of the several candidate place types. Because we want to utilize the spatial context in which the image was taken, we need to make sure each image has a geographic identifier, e.g. geographic coordinates, so that we are able to determine its neighboring place and their types. In order to classify place types of images, we consider the scene categories provided by Zhou et al. [147] as they also provide pretrained models (Places365-CNN) that we can directly use. [3] Without losing generality, we select 15 place types as our candidate class labels. The full list of class labels and their alignment with the categories in Places365-CNN is shown in Table 4.1. For each candidate class, we selected 50 images taken in 8 states [4] within the US by using Google Maps, Google Street View, and Yelp. These images include both indoor and outdoor views of each place type. Please note that classifying place types from facade and interior images is a hard problem and even the most sophisticated models only distinguish a relatively small number of place types so far which is nowhere near the approximately 420 types provided by sources such as Foursquare. Places365, for instance, offers 365 classes but many of these are scenes or landscape features, such as waves, and not POI type, such as cinemas, in the classical sense.

[3] `https://github.com/CSAILVision/places365/blob/master/categories_places365.txt`

[4] Arizona, Illinois, Nevada, North Carolina, Ohio, Pennsylvania, South Carolina, and Wisconsin

Table 4.1: Class label alignment between Yelp and the Place365 model.

Class label	**Places365-CNN category**
Amusement Parks	amusement_park
Bakeries	bakery
Bookstores	bookstore
Churches	church
Cinema	movie_theater
Dance Clubs	discotheque
Drugstores	drugstore, pharmacy
Hospitals	hospital, hospital_room
Hotels	hotel, hotel_room
Jewelry	jewelry_shop
Libraries	library
Museums	museum, natural_history_museum, science_museum
Restaurants	fastfood_restaurant, restaurant, restaurant_kitchen, restaurant_patio
Shoe Stores	shoe_shop
Stadiums & Arenas	stadium

4.3.2 Convolutional Neural Network Models

To establish baselines for our study, we selected several state-of-the-art image classification models, namely deep convolutional neural networks. Unlike traditional image classification pipelines, CNNs extract features from images automatically based on the error messages that are backpropagated through the network, thus fewer heuristics and less manual labor are needed. Contrary to densely connected feedforward neural networks, CNN adopts parameter sharing to extract common patterns which help capture translation invariance and creates sparse connections which result in fewer parameters and being less prone to overfitting.

The architecture of CNNs has been revised numerous times and has become increasingly sophisticated since its first appearance about 30 years ago. These improvements in architecture have made CNN more powerful as can be seen in the ImageNet challenge. Some of the notable architectures include: LeNet [148], AlexNet [149], VGG [150], Inception [151], ResNet [152], and DenseNet [153]. We selected AlexNet, ResNet with 18 layers (ResNet18), ResNet with 50 layers (ResNet50), and DenseNet with 161 layers (DenseNet161). AlexNet is among the first deep neural networks that increased the classification accuracy on ImageNet by a significant amount compared with traditional classification approaches. By using skip connections to create residual blocks in the network, ResNet makes it easy to learn identity functions that help with the vanishing and exploding gradient problems when the network goes deeper. In DenseNet, a dense connectivity pattern is created by connecting every two layers so that the error signal can be directly propagated to earlier layers, parameter and computational efficiency can be increased, and low complexity features can be maintained [153]. These models were trained on 1.8 million images from the Places365-CNN dataset. We used the pretrained weights for these models.

4.4 Spatial Contextual Information

In this section, we introduce three different kinds of spatial contexts and explore ways in which we can combine them with the CNN models in order to improve image classification. The first type of spatial context is spatial relatedness, which measures the extend to which different place types relate with each other. The second type of spatial context is spatial co-location, which considers what place types tend to co-occur in space and the frequency they cluster with each other. The third type of spatial context is spatial sequence pattern which considers both spatial relatedness and spatial co-location. In addition, spatial sequence pattern considers the interaction between context place types and the inverse relationship between distance and contextual influence. We use POIs provided by Yelp as dataset. [5]

4.4.1 Spatial Relatedness

Since the output of CNN is the probability score for each class label, it is possible to interpret our task as a ranking problem: given an image, rank the candidate class labels

[5] https://www.yelp.com/dataset

based upon the visual signal and spatial context signal. For the visual signal, we can obtain the ranking scores (probability scores) from the CNN architectures mentioned in Section 4.3. Since the original CNN models has 365 labels, we renormalize the probability scores for each candidate place type by the sum of the 15 candidate ranking scores so that they sum up to 1. This renormalization procedure is also applied to the other two spatial context methods explained in Section 4.4.2 and Section 4.4.3. We will refer to the renormalized scores as CNN scores in this study. For the spatial context signal, the ranking scores are calculated using the place type embeddings proposed in [48]. These embeddings capture the semantics of different place types and can be used to measure their similarity and relatedness. In this regard, the task is equivalent to a re-ranking problem, which adjusts the initial ranking provided by the visual signal using auxiliary knowledge, namely the spatial context signal. Intuitively, the extent to which the visual signals from the images match with different place types and the level of relevance of the surrounding place types with respect to candidate place types jointly determine the final result.

Inspired by search re-ranking algorithms in information retrieval, we use a *Linear Bimodal Fusion* (LBF) method (here essentially a 2-component convex combination), which linearly combines the ranking scores provided by the CNN model and the spatial relatedness scores, as shown in Equation 4.1.

$$s_i = \omega^v s_i^v + \omega^r s_i^r \tag{4.1}$$

where s_i, s_i^v, and s_i^r are the LBF score, CNN score, and spatial relatedness score for place type i respectively, ω^v and ω^r are the weights for the CNN component and spatial relatedness component, and $\omega^v + \omega^r = 1$. The weights here are decided based on the relative performance of individual components. Specifically, the weight is determined using Equation 4.2.

$$\omega^v = \frac{acc^v}{acc^v + acc^r} \tag{4.2}$$

where acc^v and acc^r are the accuracies for CNN and spatial relatedness measurements for the image classification task. Intuitively, this means that we have higher confidence if the component performs better on its own and want to reflect such confidence using the weight in the LBF score.

In order to calculate the spatial relatedness scores, we use cosine similarity to measure the extend to which each candidate class embedding is related with the spatial context embedding of an image in a high dimensional geospatial semantic feature space. Following the suggestions in [48], we use a concatenated vector of 350 dimensions (i.e., 70D vectors for each of 5 distance bins) as the place type embeddings. The candidate class embeddings can be retrieved directly. Then we search for the nearest n POIs based on the image location, determine the place types of these n POIs, and calculate the average of these place type embeddings as the final spatial context embeddings for images. The cosine similarity score sm_i is calculated between the spatial context embedding of an image and the embedding of each candidate place type class i. Because sm_i ranges from -1 to 1, we use min-max normalization to scale the values to $[0, 1]$. Finally, we apply the same renormalization as for the CNN score to turn the normalized score sm'_i into probability score, i.e. spatial relatedness score s^r_i.

Combining these normalizations together with Equation 4.1 and Equation 4.2, we are able to derive that $0 \leq s_i \leq 1$ and $\sum_{i=1}^{N} s_i = 1$ where $N = 15$ in our case. This means that the LBF score s_i can be considered a probability score.

4.4.2 Spatial Co-location

The spatial relatedness approach follows the assumption that relatedness implies likelihood which is reasonable in cases where similar place types cluster together, such as restaurant, bar, and hotel. However, in cases of high spatial heterogeneity, this assumption will fall short of correctly capturing the true likelihood. An example would be places of dissimilar types that co-occur, e.g., grocery stores and gas stations. Moreover, the LBF method can only capture a linear relationship between the two signals.

Following Berg et al. [61], we also test a Bayesian approach in which we assume there is a complex latent distribution of the data that facilitates our classification task. Intuitively, the CNN score gives us the probability of each candidate class t given the image I, i.e., $P(t|I)$, and the spatial context informs us of the probability of each candidate class given its neighbors $c_1, c_2, c_3, ..., c_n$, denoted as C, around the image location, i.e., $P(t|C)$. We would like to obtain the posterior probability of each candidate class given both the image and its spatial context, i.e., $P(t|I, C)$. Using Bayes' theorem, the posterior probability can be written as:

$$P(t|I, C) = \frac{P(I, C|t)P(t)}{P(I, C)} \tag{4.3}$$

For variables I, C, and t, we construct their dependencies using a simple probabilistic graphical model, i.e., Bayesian network, which assumes that both the image I and the spatial context C are dependent on the place type t, which intuitively makes sense in that different place types will result in different images and different place types of their neighbors. We know that given information about the image I we are able to update our beliefs, i.e., the probability distributions, about the place type t. In addition, the changes in our beliefs about the place type t can influence the probability distributions of the spatial context C. However, if place type t is observed, the influence cannot flow between I and C, thus we are able to derive the conditional independence of I and C given t. So Equation 4.3 can be rewritten as:

$$\begin{aligned} P(t|I,C) &= \frac{P(I|t)P(C|t)P(t)}{P(I,C)} \\ &= \frac{P(t|I)P(I)}{P(t)}\frac{P(t|C)P(C)}{P(t)}\frac{P(t)}{P(I,C)} \\ &\propto \frac{P(t|I)}{P(t)}P(t|C) \end{aligned} \tag{4.4}$$

in which we have dropped all the factors that are not dependent on t as they can be considered as normalizing constants for our probabilities. It follows that the posterior probability $P(t|I,C)$ can be computed using the CNN probability score $P(t|I)$, the spatial context prior $P(t|C)$, and the candidate class prior $P(t)$. Instead of estimating the distribution of spatial context priors, we take advantage of the spatial co-location patterns and calculate the prior probabilities using the Yelp POI data directly. As mentioned earlier, the spatial context C is composed of multiple individual context neighbors $c_1, c_2, c_3, ..., c_n$; hence, we need to calculate $P(t|c_1, c_2, c_3, ..., c_n)$. In order to simplify our calculation, we impose a bag-of-words assumption as well as a Naive Bayes assumption in the spatial co-location patterns. The bag-of-words assumption simplifies the model by assuming that the position (or the order) in which different context POIs occur does not play a role. The Naive Bayes assumption implies that the only relationship is the pair-wise interaction between the candidate place type t and an individual neighbor's place type c_i and there is no interaction between neighboring places wrt. their types, i.e. $(c_i \perp\!\!\!\perp c_j|t)$ for all c_i, c_j. Using spatial co-location, we are able to calculate the conditional probability using place type co-location counts $P(c_i|t) = \frac{count(c_i,t)}{count(t)}$ where $count(c_i, t)$ is the frequency that neighbor type c_i and candidate type t co-locate within a certain distance limit and $count(t)$ is the frequency of candidate type t in the study area. Combining all these components, we can derive:

$$
\begin{aligned}
P(t|C) &= P(t|c_1, c_2, ..., c_n) \\
&= \frac{P(t)\prod_{i=1}^{n} P(c_i|t)}{P(c_1, c_2, c_3, ..., c_n)} \\
&= \frac{P(t)}{P(c_1, c_2, c_3, ..., c_n)} \frac{\prod_{i=1}^{n} count(c_i, t)}{count(t)^n}
\end{aligned}
\tag{4.5}
$$

Using Equation 4.4 and Equation 4.5, we can derive the final formula for calculating $P(t|I, C)$ shown in Equation 4.6. For the sake of numerical stability, we calculate the log probability $logP(t|I, C)$ using the natural logarithm. Since the natural logarithm is a monotonically increasing function, it will not affect the final ranking of the classification results.

$$
\begin{aligned}
logP(t|I, C) &\propto log\left(\frac{P(t|I)}{P(t)} P(t|C)\right) \\
&= log\left(\frac{P(t|I)}{P(c_1, c_2, c_3, ..., c_n)} \frac{\prod_{i=1}^{n} count(c_i, t)}{count(t)^n}\right) \\
&\propto logP(t|I) + \sum_{i=1}^{n} log(count(c_i, t)) - nlog(count(t))
\end{aligned}
\tag{4.6}
$$

where we also drop $P(c_1, c_2, c_3, ..., c_n)$ as it does not depend on t, so it will not affect the result ranking. The log posterior probability is then used to generate the final ranking of candidate place types and produce the classification results.

4.4.3 Spatial Sequence Pattern

The spatial co-location approach follows the bag-of-words assumption that the position of spatial context POIs does not matter and the Naive Bayes assumption that the context neighbors are independent of each other. However, in many cases this assumption is too strong. In fact, numerous methods, such as Kriging and multiple-point geostatistics, have been devised to model geospatial proximity patterns and complex spatial interaction patterns. However, incorporating these complex spatial patterns in a multidimensional space would adversely affect the model complexity and make the distribution in Section 4.4.2 intractable. In order to strike the right balance between the complexity of model and the integrity of spatial context pattern, we propose to capture the spatial sequence pattern in our model by collapsing the 2D geographic space into a 1D sequence.

Specifically, we use the Long Short-Term Memory (LSTM) network model, a variant of recurrent neural network (RNN), in our study. Recurrent neural networks are frequently used models to capture the patterns in sequence or time series data. In theory, the naive recurrent neural networks can capture long term dependencies in the sequence, however, due to the vanishing and exploding gradient problem, they fail to do so in practice. LSTM is explicitly designed to solve the problem by maintaining a cell state and controlling the input and output flow using forget gate, input gate, and output gate [154]. We use LSTM as a generative model in order to capture the latent distribution of place types using the spatial sequence pattern. In the training stage, the input is a sequence of context place types $c_1, c_2, c_3, ..., c_n$ and the output is the place type t of the POI from which the context is created. The input sequence is ordered in a way so that the previous one is further away from the output than the next one in the collapsed 1D space. Image one would drive around a neighborhood before reaching a destination. For each of the POIs encountered during the route, one would update the beliefs about the neighborhood by considering the current POI and all previously seen POIs. Upon arriving at the destination, one would have a reasonable chance of guessing this final POI's type. The structure of the LSTM model is shown in Figure 4.1. We apply a dropout after the LSTM layer to avoid overfitting. After training the LSTM model on Yelp's POI dataset, we are able to obtain the spatial context prior $P(t|c_1, c_2, c_3, ..., c_n)$ based on the spatial sequence pattern around

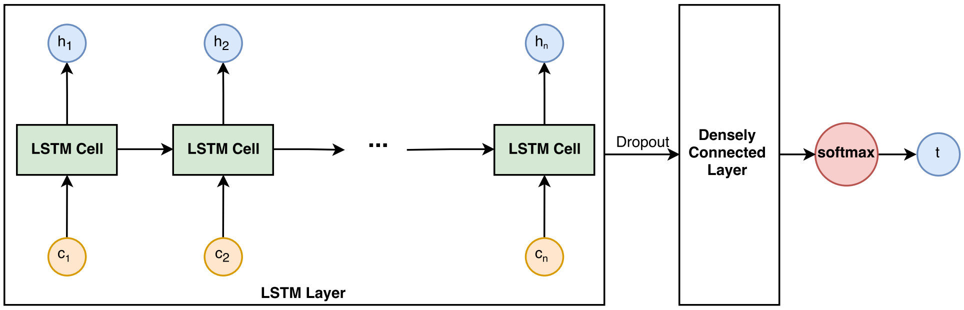


Figure 4.1: Structure of the LSTM.

the image locations in our test data. We specifically removed the image locations and their context in the training data. Similar to the spatial co-location approach, we use Bayesian inference and log probability to calculate the final result:

$$\begin{aligned} logP(t|I,C) &\propto log\left(\frac{P(t|I)}{P(t)}P(t|C)\right) \\ &= logP(t|I) + logP(t|c_1, c_2, c_3, ..., c_n) - logP(t) \end{aligned} \tag{4.7}$$

where the candidate class prior $P(t)$ can be computed using the Yelp data. Since we use LSTM as a generative model, in the prediction phase, sampling strategies, such as greedy search, beam search, and random sampling, can be applied based on the distribution provided by the output of the LSTM prediction. However, we only generate the next prediction instead of a sequence, so we do not apply these sampling strategies. Instead, we make use of the hyperparameter *temperature* τ to adjust the probability scores returned by the LSTM model before combining them with the CNN model in a Bayesian manner. Including the hyperparameter τ, the softmax function in the LSTM model can be written as:

$$P(t_i|C) = \frac{exp(\frac{logit_i}{\tau})}{\sum_{j=1}^{N} exp(\frac{logit_j}{\tau})} \tag{4.8}$$

where $logit_i$ is the logit output provided by LSTM before applying the softmax function and $N = 15$ in our case. Intuitively, when the temperature τ is high, i.e., $\tau \rightarrow \infty$, the probability distribution will become diffuse and $P(t_i|C)$ will have almost the same value for different t_i; when τ is low, i.e., $\tau \rightarrow 0^+$, the distribution becomes peaky and the largest $logit_i$ stands out to have a probability close to 1. This idea is closely related to the exploration and exploitation trade-off in many machine learning problems. The value of τ will affect the probability scores $P(t_i|C)$ but not the ranking of these probabilities.

In this study, we propose two ways to model the 2D geographic space as a 1D sequence. The first one is a distance-based ordering approach. For any given POI, we search for nearby POIs within a certain distance from it, choose the closest n POIs, and rearrange them by distance with descending order, thereby forming a 1D array. This distance-based method is isotropic in that it does not differentiate between directions while creating the sequence. The second method is a space filling curve-based approach. We utilize *Morton order* here which is also used in geohashing to encode coordinates into an indexing string that can preserve the locality of spatial locations. We use Morton order to encode the geographic locations of every POI and order them in a sequence based upon their encodings, i.e., indexing sequence. After obtaining the sequence, for each POI, we use the previous n POI in the sequence as the context sequence. Other space filling curves could be used in future work.

Because each POI can have multiple place types associated with it, e.g., restaurant and beer garden, the sequence of place types is usually not unique for the same sequence of POIs. As our LSTM input is a sequence of place *types*, we compute the Cartesian product of all POI type sets in the sequence of nearby places:

$$T_{c_1} \times T_{c_2} \times T_{c_3} \times ... \times T_{c_n} = \{(t_{c_1}, t_{c_2}, t_{c_3}, ..., t_{c_n})|\forall i = 1, 2, 3, ..., n,\ t_{c_i} \in T_{c_i}\} \tag{4.9}$$

where T_{c_i} is the set of place types associated with POI c_i in the context sequence. In practice, however, we randomly sample a fixed number of place type sequences from each of the Cartesian product for the POI context sequence as the potential combinations grow exponentially with increasing context size.

4.5 Experiment and Result

In this section, we explain our experimental setup for the models described above, describe the metrics used to compare the model performance for place type image classification, and present the results and findings.

4.5.1 Implementation Details

For all three types of spatial context, we use 10 as the maximum number of context POIs and a distance limit of 1000m for the context POI search. For the spatial sequence pattern approach, we use a fixed sample size of 50 to sample from the Cartesian product of all POI type sets in the sequence. [6] We use a one-layer LSTM with 64 hidden units. We train our LSTM model using the recommended Root Mean Square Propagation (RMSProp) optimizer with a learning rate of 0.005. A dropout ratio of 0.2 is applied in the LSTM and we run 100 epochs. The same settings are used for all LSTM trainings in our experiment. The total number of POI in the dataset is 115,532, yielding more than 5 million unique training sequences.

For evaluation, we use three different metrics, namely Mean Reciprocal Rank (MRR), Accuracy@1, and Accuracy@5. Another common metric for image classification would also be Mean Average Precision (MAP), but since there is only one true label per type in our task, we use MMR instead.

[6] The median for types per place in Yelp is 3.

4.5.2 Results

We run the 750 test images we collected, i.e., 50 images per each of 15 types, on the four CNN baseline models (AlexNet, ResNet18, ResNet50, and DenseNet161) as well as the combined models using our three different types of spatial context. [7] In addition to the two methods for converting geographic space into 1D sequences in the spatial sequence pattern approach, we also test one model using random sequences with the same context count and distance limits. We did so to study whether results obtained using the LSTM would benefit from distance-based spatial contexts. A higher result for the spatial sequence based LSTM over the random LSTM would indicate that the network indeed picked up on the distance signal.

The hyperparameter τ can be adjusted; a value of 0.5 has been proposed as a good choice before. In order to test this and find the optimal temperature value, we run the combined model using spatial sequence patterns with three types of sequencing approaches, namely random sequence, distance-based sequence, and Morton order-based sequence.

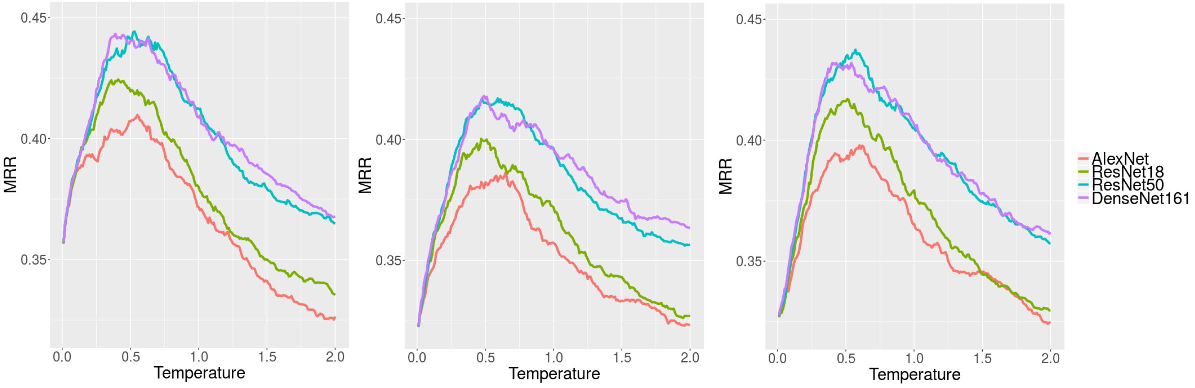


Figure 4.2: From left to right, MRR result using distance-based sequence, random sequence, and Morton code-based sequence with varying temperatures

[7] Transfer learning could be applied to fine tune the CNN models first, but we only have limited images and our hypothesis is that spatial context can be used as a powerful complement or alternative to the visual component for image classification.

We test temperature values ranging from 0.01 to 2 with a step of 0.01. We combine the spatial sequence pattern models with all CNN models. The MRR result with respect to temperature are shown in Figure 4.2. Although there are a slight variations, the MRR curves all reach their peaks around a τ value of 0.5. This confirms the suggestion from the literature. Figure 4.3 shows selected example predictions. The results for MRR, Accuracy@1, and Accuracy@5 using the baseline models as well as our proposed, spatially explicit models are shown in Table 4.2, Table 4.3, and Table 4.4. [8]

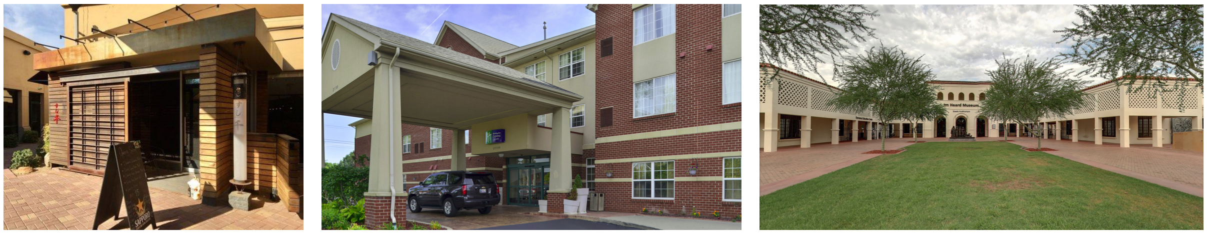


Figure 4.3: From left to right, images of a restaurant, a hotel, and a museum from Yelp, Google Street View, and Google Maps respectively. The first image is incorrectly classified as library using all 4 CNN models and it is correctly classified as restaurant using the spatial sequence pattern (distance) models. The second image is classified as hospital and library by the original CNN models and is classified as hotel by the spatial sequence pattern (distance) models. For the third image the correct label museum is in the third position in the label rankings of all 4 CNN models while, using the spatial sequence pattern (distance) models, ResNet18 and ResNet50 can correctly label it and in the label rankings of AlexNet and DenseNet161 museum is in the second position.

Table 4.2: MRR result using baseline models and proposed combination models using different types of spatial context and sequences

MRR	AlexNet	ResNet18	ResNet50	DenseNet161
Baseline	0.27	0.28	0.31	0.31
Relatedness	0.27	0.28	0.31	0.32
Co-location	0.30	0.31	0.31	0.32
Sequence Pattern (Random)	0.38	0.40	0.42	0.42
Sequence Pattern (Distance)	**0.41**	**0.42**	**0.44**	**0.44**
Sequence Pattern (Morton order)	0.39	**0.42**	0.43	0.43

[8] The baseline models are not comparable with a random classifier which would yield an expected accuracy of 1/15 in this case, because the baseline CNN models have 365 unique labels and we choose 15 labels in our experiment.

As we can see, by incorporating spatial context in the image classification model, we are able to improve the classification result in general. However, integrating spatial relatedness using the LBF method does not seem to affect the result. This essentially confirms our aforementioned assumption that relatedness does not always imply likelihood. The benefit of incorporating spatial relatedness in cases of spatial homogeneity are likely to be offset by cases of hight spatial heterogeneity in which spatial relatedness may have an negative effect as dissimilar places co-occur.

The Accuracy@1 measurement is improved by incorporating spatial co-location component in the models. This confirms our previous reasoning that considering the external signal, namely spatial contexts, and assuming a complex latent distribution of the data in a Bayesian manner improve image classification. However, for MRR the improvement is marginal and for Accuracy@5 there even is a decrease after incorporating the spatial co-location component because this type of spatial context falls short of taking into account the intricate *interactions* of different context neighbors. This shortcoming is not clear when only looking at the first few results in the ranking returned by the combined models, but it becomes clearer in later results in the ranking output, thus resulting in a decrease for Accuracy@5 and only a slight increase in the MRR measurement.

Table 4.3: Accuracy@1 result using baseline models and proposed combination models using different types of spatial context and sequences

Accuracy@1	AlexNet	ResNet18	ResNet50	DenseNet161
Baseline	0.07	0.07	0.09	0.09
Relatedness	0.07	0.07	0.09	0.09
Co-location	0.15	0.17	0.17	0.17
Sequence Pattern (Random)	0.18	0.18	0.19	0.20
Sequence Pattern (Distance)	**0.20**	**0.20**	**0.22**	**0.22**
Sequence Pattern (Morton order)	0.19	**0.20**	**0.22**	**0.22**

The Bayesian combination model using spatial sequence patterns shows better overall results compared with the baseline models, the spatial relatedness model, and the

Table 4.4: Accuracy@5 result using baseline models and proposed combination models using different types of spatial context and sequences

Accuracy@5	AlexNet	ResNet18	ResNet50	DenseNet161
Baseline	0.50	0.56	0.59	0.60
Relatedness	0.52	0.56	0.58	0.59
Co-location	0.42	0.44	0.45	0.44
Sequence Pattern (Random)	0.65	0.69	**0.73**	0.73
Sequence Pattern (Distance)	**0.67**	**0.70**	**0.73**	**0.75**
Sequence Pattern (Morton order)	0.65	**0.70**	0.72	0.71

spatial co-location model. This is because the spatial sequence patterns capture spatial interactions between the neighboring POIs that are neglected by the other models. From the result we can see that using a distance-based sequence is better than using a random sequence. To prevent confusion and to understand why the random model still performs relatively well, it is important to remember that this model utilizes spatial context. However, it does not utilize the distance signal within this context but merely the presence of neighboring POI. The results show that a richer spatially explicit context, one that comes with a notion of *distance decay*, indeed improves classification results. Interestingly, the sequence using Morton order, which is widely used in geohashing techniques, does not further improve the result compared to the distance-based sequence. There may be multiple reasons for this. First, we may have reached a ceiling of possible improvements by incorporating spatial contexts. Second, our Morton order implementation takes the 10 places that precede the target place in the index. This may result in directional effects. Finally, all space filling curves essentially introduce different ways to preserve local neighborhoods; utilizing another technique such as Hilbert curves may yield different results. Given that the Morton order-based sequence in many cases yield results of equal quality to the distance-based sequences, further work is needed to test the aforementioned ideas.

Summing up, the results demonstrate that incorporating a (distance-based) spatial context improves the MRR of state-of-the-art image classification systems by over **40%**.

The results for Accuracy@1 are more than **doubled** which is of particular importance for humans as this measure only considers the first ranked result.

4.6 Conclusion and Future Work

In this work, we demonstrated that utilizing spatial contexts for classifying places based on images of their facades and interiors leads to substantial improvements, e.g., increasing MRR by over 40% and doubling Accuracy@1, compared to applying state-of-the-art computer vision models such as ResNet50 and DenseNet161 alone. These advances are especially significant as the classification of places based on their images remains a hard problem. One could argue that our proposal requires additional information, namely about the types of nearby places. However, such data are readily available for POI, and only a few nearby places are needed. Secondly, and as a task for future work, one could also modify our methods to work in a *drive-by-typing* mode in which previously seen places are classified, and these classification results together with their associated classification uncertainty are used to improve estimation of the currently seen place, thereby relaxing the need for POI datasets. In the future, we would like to apply transfer learning and experiment with other ways to encode spatial contexts, e.g., by testing different space-filling curves. We plan to develop models to directly capture 2D spatial patterns rather than using a 1D sequence as a proxy and test whether spatial contexts also aid in recognizing objects beyond places and their facades.

Chapter 5

A Spatially-Explicit Reinforcement Learning Model for Geographic Knowledge Graph Summarization

This chapter presents a generic method for geographic knowledge graph summarization based on reinforcement learning by considering spatial relations explicitly. The proposed method tackles three major challenges in summarizing geographic knowledge graphs, namely the complexity of graph structure, the subjectivity of summarization criteria, and the richness of geospatial semantics. The complexity of graph structure is handled by formulating the summarization tasks as a Markov Decision Process and applying the Monte Carlo Policy Gradient method. Wikipedia summaries are used to provide a relatively objective summarization baseline. In order to capture the richness of geospatial semantics, we apply the geospatial inductive bias by introducing an explicity spatial relation in addition to existing relations in the graph. Results from Chapter 3 and Chapter 4 can be potentially incorporated in this method. The evaluation has shown promising results for the proposed spatially-explicit method.

Peer Reviewed Publication	
Title	A Spatially-Explicit Reinforcement Learning Model for Geographic Knowledge Graph Summarization
Authors	Bo Yan, Krzysztof Janowicz, Gengchen Mai, and Rui Zhu
Venue	Transactions in GIS (Esri UC special issue)
Editors	John P. Wilson
Publisher	Wiley
Pages	In press
Submission Date	February 3, 2019
Acceptance Date	April 3, 2019
Publication Date	In press
Copyright	Reprinted with permission from Wiley

Abstract Web-scale knowledge graphs such as the global Linked Data cloud consist of billions of individual statements about millions of entities. Unsurprisingly, many of the most densely connected entities are places and regions, often characterized by thousands of incoming and outgoing relationship to other places, actors, events, and objects. In recent years, this has fueled the interest in knowledge graph summarization techniques that compute representative subgraphs for a given collection of nodes. In this paper, we propose a novel summarization method that incorporates spatially-explicit components into a reinforcement learning framework in order to help summarize *geographic* knowledge graphs, a topic that has not been considered in previous work. Our model considers the intrinsic graph structure as well as the extrinsic information to gain a more comprehensive and holistic view of the summarization task. By collecting a standard dataset and evaluating our proposed models, we demonstrate that the spatially-explicit model yields better results than non-spatial models, thereby demonstrating that spatial is indeed special as far as summarization is concerned.

5.1 Introduction

Knowledge graphs were originally introduced to promote the creation, retrieval, and reuse of human and machine readable structured data. Recent advances in related technology stacks such as knowledge graph-based question answering systems as well as the adoption by commercial companies have highlighted the success of knowledge graphs in both academia and industry. Users and providers of geographic information have always been at the forefront of research in knowledge graphs and their representations as they address many key challenges in the areas of semantic interoperability and spatial data infrastructures that have plagued GIScience for years [155]. In fact, a large number of entities in Wikidata[1] — a sister initiative of Wikipedia to create a repository for structured information — are spatial and dedicated geospatial knowledge graph hubs such as LinkedGeoData[2] contain billions of statements about geographic entities.

In theory, today's abundance of geographic data facilitates new research and more powerful question answering systems. From a more practical perspective, however, sifting through the data deluge becomes increasingly challenging. Ramscar et al. [18] have shown that too much information may adversely influence our cognitive information-processing capacities and unavoidably result in lags and retrieval errors. As a result, researchers are working on ways to better present data for humans, such as interfaces and visualization tools to make knowledge graphs more user-friendly and more accessible for non-technical audiences. One novel area of study is *knowledge graph summarization*, namely selecting and identifying the property-value pairs that best represent the underlying entity from a large and convoluted graph [22].

The idea of summarizing a knowledge graph in a way such that the subgraph retains the significant substructures and meaning, here prominent nodes and edges, of the original

[1] https://www.wikidata.org

[2] http://linkedgeodata.org

graph is intriguing. However, this task is entangled with a lot of challenges, especially in the geospatial domain. One challenge is related to the inherently complex structure of graph data. Unlike other commonly-used structures such as the 1D sequence of natural languages and the 2D grids of images, graph structures are peculiar in their own ways. For example, on the global level, two graphs can be isomorphic, i.e., have the same structure, while they have distinct representations (e.g., labeling and visual representations). On the local level, substructures such as homophily and structural equivalence [156] coexist in the graph as proxies to encode the underlying patterns. In addition, since most knowledge graphs follow the so-called Open World Assumption (OWA) – which implies that there are possibly missing statements/triples in the knowledge graph without having to assume that those missing statements do not hold true in reality – the original structure of the graph might not represent the complete information. This adds another layer of complexity.

As a result of the versatility and peculiarity of graph data, traditional methods that rely heavily on handcrafted features/rules (such as clustering coefficients and other graph summary statistics) for knowledge graph summarization are not sufficient enough because they do not generalize well. Another challenge is the subjectivity of the summarization criteria. The relative importance of a node (entity) or an edge (relation/property/predicate) in the knowledge graph is not universally defined and different application fields can interpret it differently. For instance, a knowledge graph that primarily models friendship relation among individuals may take advantage of the connectivity information (such as degrees, betweenness, closeness, or eigenvector centrality) to determine important nodes in the summarization process. On the contrary, to summarize the DBpedia[3] knowledge graph — a crowd-sourced community effort to extract structured information from various Wikimedia projects — where there are a large number of distinct relation types

[3] `https://wiki.dbpedia.org/`

and the whole graph is densely connected, latent information embedded in the labels and abstracts of each entity and relation is required to determine the extent to which each component of the graph is related with one another in order to rank the relative importance. Besides the aforementioned challenges, *geographic* knowledge graph summarization has its distinct challenges. Given the inherent richness of geospatial semantics [48, 49], geospatial components such as spatial contexts play a significant role in understanding spatial entities and their dependencies. However, existing (knowledge) graph summarization methods [21] are not tailored towards the geospatial domain thus neglecting such special components. For instance, a summary about Santa Barbara, CA is also always a partial summary of California. As humans we give special weight to the places where important historic figures were born even if they spent their entire life somewhere else. Hence, every summary of the city of Ulm, Germany, e.g., the first paragraph of its Wikipedia article, lists Albert Einstein as notable resident despite his family moving to Munich a year after his birth. For Munich in turn, his name is not prominently featured in the city's description. This may be related to the broader phenomenon of duration neglect [157].

In light of this, we propose to adopt a reinforcement learning-based approach to explicitly incorporate spatial contextual information. Our method combines both intrinsic structure and extrinsic information to help summarize *geographic* knowledge graphs as most domain-agnostic work [22, 158, 159, 160, 33, 99] fails to consider the inherent richness of geospatial semantics. In fact, we believe that there is no prior work about geographic knowledge graph summarization at all – despite places such as Vienna, Austria being represented by tens of thousands triples in modern knowledge graphs, and, hence, in desperate need for graph summarization. In order to strike the balance between diversity and uniformity in summarizing geographic knowledge graphs, our model utilizes the idea of distance decay and information entropy to determine the relatedness of different spatial/non-spatial entities.

By intrinsic structure, we mean the graph structure where each entity is connected by properties. We embrace the current trend of utilizing vector representations, namely translation-based embedding models [161], to embed the structural information of knowledge graphs. The semantic information – by which we mean latent information encoded in natural language, and, hence, not directly available to structural analysis – of the knowledge graph is captured by the embeddings of entity and relation labels. For extrinsic information, we take advantage of the Wikipedia abstracts of different places (geographic entities) to guide our summarization process since these abstracts are exemplary summaries of each geographic entity produced by human ingenuity, and there is a clear tractable correspondence between Wikipedia articles and knowledge graphs [162, 163, 164]. By combining reinforcement learning with knowledge graph embeddings, word embeddings, information theory, and spatial contexts, we aim to tackle the challenges mentioned above. Knowledge graph embeddings efficiently encode the hidden structure of the graph. Word embeddings facilitate the transmission of semantic information in the knowledge graph to the summarization process. Information theory together with the reinforcement learning framework (guided by Wikipedia summaries) is employed to partially mitigate the subjectivity issue that impacts knowledge graph summarization tasks. After all, Wikipedia abstracts provide relatively neutral [43, 44], curated, concise, and generic digests that highlight the distinctive and significant aspects of different places. Spatial contexts are used to help recover missing links in the geographic knowledge graph and uncover the hidden geospatial patterns.

The research contributions of this paper are as follows:

- We utilize Wikipedia summaries to guide the geographic knowledge graph summarization process using reinforcement learning. Instead of mostly relying on intrinsic information, such as node groups in grouping and aggregation-based approaches and the number of bits needed to describe the graph in bit compression-based approaches, our approach reaps the complementary strengths of intrinsic information from the graph structure and extrinsic knowledge using Wikipedia summaries by framing the task as a sequential decision making process that can be optimized using reinforcement learning.

- We account for the richness of geospatial semantics in geographic knowledge graphs and incorporate such information in the summarization process in order to better capture the relatedness of geographic entities and provide better results. We do so by following established GIScience methods, namely modeling distance decay, as well as from an information theoretic perspective.

- We create a dataset DBP369[4] that includes 369 place summaries from Wikipedia and a subgraph of DBpedia for geographic knowledge graph summarization tasks and make it openly available. A lack of standard datasets has been one of the obstacles that hinder research development in the area of geographic knowledge graph summarization and to some degree geographic information retrieval in general. By taking the initiative to collect this dataset, we hope it will foster further research in this area.

- We establish different baselines for the geographic knowledge graph summarization task for the DBP369 dataset. Our result shows that by considering spatial contextual components the summarized graph better resembles the Wikipedia summary.

- Finally, to the best of our knowledge this is the first research to consider the problem of geographic knowledge graph summarization. This is remarkable as Web-scale

[4] `http://stko.geog.ucsb.edu/gkg/`

knowledge graphs such as Linked Data store tens of millions of places and often thousands of statements (subject-predicate-object triples) about them.

The remainder of this paper is organized as follows. Section 5.2 summarizes existing work on knowledge graph summarization, spatially-explicit models, and utilizing reinforcement learning in the context of knowledge graphs. Section 5.3 describes the basic procedure of our data collection and provides detailed information about the DBP369 dataset. Section 5.4 explains the proposed method for geographic knowledge graph summarization. Section 5.5 applies the model to the DBP369 dataset and evaluates the results. Section 5.6 concludes the research and points to directions for future work.

5.2 Related Work

Most graph summarization techniques fall into one of the four categories [21] namely: grouping or aggregation-based approaches, bit compression-based approaches, simplification or sparsification-based approaches, and influence-based approaches. Knowledge graph summarization usually adopts the simplification or sparsification-based approach for the reason that the prime motivation for summarizing knowledge graphs is to provide a subgraph that highlights the important entities and relations of the original graph. Cheng et al. [22] and Thalhammer and Rettinger [159] proposed to utilize the graph structure and performed PageRank to identify relevant entities and summarize the graph. Pirrò [160] formalized the notion of relatedness in knowledge graphs to better harness the large variety of information. While these papers primarily take advantage of the intrinsic information of knowledge graphs, some work is geared towards extrinsic knowledge. For instance, Bast et al. [33] utilized textual information from Wikipedia to build logistic regression and generative models to calculate relevance scores for relations in knowledge graph triples. Our work takes the best of both worlds by considering intrinsic knowledge graph structure and extrinsic information simultaneously.

In addition, all the work mentioned above aims at retrieving/ranking entities/relations based on certain criteria such as relevance scores with respect to a user's queries rather than providing a subgraph that captures the essence of the original graph. Our work provides a subgraph that summarizes the relations and connected entities for each geographic entity based on corresponding Wikipedia abstracts. With the recent trend towards learning latent representations of graphs [165], methods based on matrix factorization strategies (such as Singular Value Decomposition (SVD), CUR [166], and Compact Matrix Decomposition (CMD) [167]) have been used in which low-rank approximations of adjacency matrices are viewed as sparse approximation summaries of the original graphs. Our work embraces the idea of adopting a more scalable neural network-based approach, namely the TransE [161] model, to learn low-dimensional latent knowledge graph representations and applying these embeddings within our summarization pipeline.

In order to study the influence of geospatial contexts on identifying different types of places, Yan et al. [48] proposed a latent representation learning method based on augmented spatial contexts. Similarly, Yan et al. [49] used spatial sequence patterns of neighborhoods as Bayesian priors and combined them with state-of-the-art convolutional neural network models to help improve image classification for different place types using data collected from Yelp and Google Street View. Mai et al. [8] incorporated geographic weights into the latent representation learning process in order to provide better knowledge graph embeddings for geographic question answering tasks. Our work, follows the same line of reasoning, namely that *spatially-explicit models* substantially outperform more general models when applied to geographic data. Kejriwal and Szekely [88] presented a geospatial data source generated using weighted neural embeddings methods on Geonames[5] data. The resulting embeddings encode geographic contextual information.

[5] `https://www.geonames.org/`

Researchers working on knowledge graphs have been exploring different ways in which reinforcement learning can be used. For example, Xiong et al. [168] adopted the REINFORCE (Monte Carlo Policy Gradient) algorithm [169] to make a policy-based agent learn multi-hop relational paths for knowledge graph reasoning tasks by considering accuracy, diversity, and efficiency in their reward function. Das et al. [170] framed the knowledge graph reasoning task as a finite horizon, deterministic partially observable Markov Decision Process (MDP) and designed a randomized non-stationary history-dependent policy parameterized by a long short-term memory network (LSTM) [154]. Shen et al. [171] developed the M-Walk graph-walking agent using recurrent neural network (RNN) to encode the history of the walked path and Monte Carlo Tree Search (MCTS) with a neural policy to generate trajectories yielding more positive rewards to overcome the challenge of sparse rewards under the off-policy Q-learning framework for knowledge graph completion. However, none of these approaches used a geographic dataset. Moreover, our work is based on the novel idea of treating the geographic knowledge graph summarization task as an MDP and the decision at each summarization step is made by the reinforcement learning agent.

5.3 Dataset

Given the lack of existing work on geographic knowledge summarization and related benchmarks, we collected the dataset DBP369 for our research and hope it can be adopted in similar research studies in the future. We initially picked 500 places from different areas of the world, as shown in Fig. 5.1. In this work, we would like to explore the possibility of guiding the summarization process of geographic knowledge graphs by means of unstructured human knowledge. There are two parallel parts of our dataset: 1) Wikipedia summaries of each of these places, 2) A geographic knowledge graph containing each of these places and their related entities. These places include well-known metropolitan areas such as New York City and Los Angeles as well as areas with archaeological and historic significance such as Olympia, Greece. We used the MediaWiki API[6] to find the corresponding Wikipedia pages for these places, from which summary texts were extracted. These summaries provide a human-generated guidance for summarizing geographic knowledge graphs.

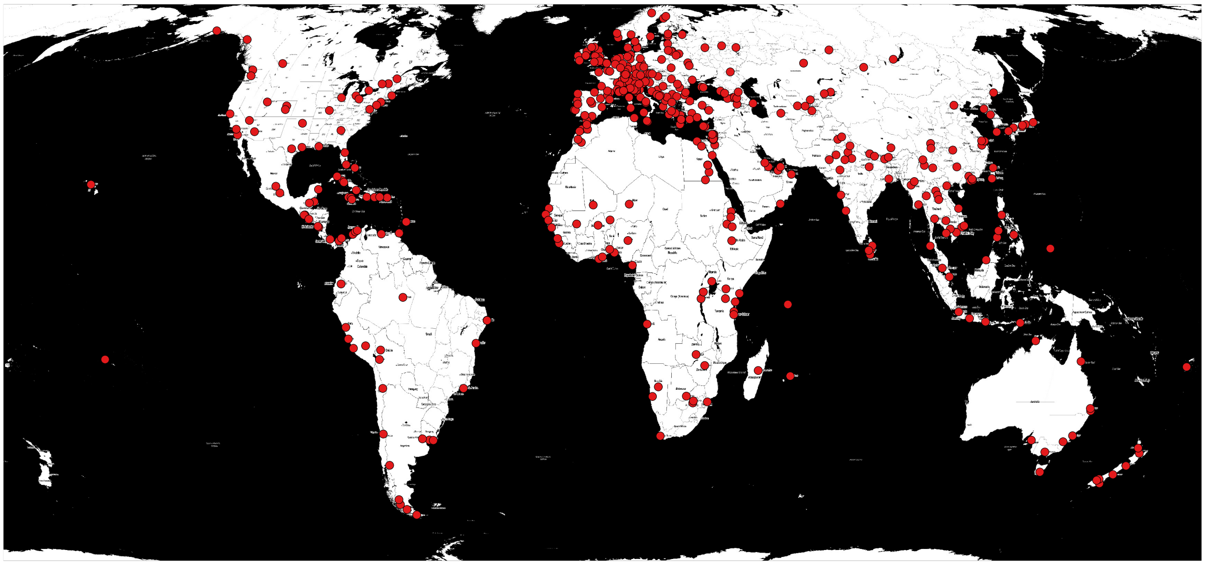


Figure 5.1: Place distribution map (Eckert IV).

For the geographic knowledge graph part, we selected DBpedia as our data source, as it has numerous geographic entities, is being actively maintained and updated, has a clear one-to-one correspondence for each Wikipedia article, and provides a diversified and comprehensive coverage of properties. In order to construct our geographic knowledge graph from DBpedia, we prepared these 500 places from Wikipedia and retrieved all links that appeared in the summaries of these 500 articles. We generated mappings to find the corresponding entities for these places as well as the links. After obtaining these

[6] https://en.wikipedia.org/w/api.php

seed entities, we generated SPARQL[7] queries to retrieve 1-degree and 2-degree neighbors iteratively in order to form subgraphs surrounding these seed nodes. In DBpedia all statements are organized as *(head, relation, tail)* or *(subject, predicate, object)* triples. Query 5.1 shows an example query that uses a basic graph pattern to obtain 1-degree (both incoming and outgoing) neighboring nodes of DBpedia entity *dbr:Los_Angeles*.

```
PREFIX  dbr: <http://dbpedia.org/resource/>
SELECT DISTINCT * WHERE {{
dbr:Los_Angeles ?p1 ?o .
FILTER(CONTAINS(str(?p1),’http://dbpedia.org/ontology/’) && !isLiteral
   ↪ (?o))}
UNION {
?s ?p2 dbr:Los_Angeles .
FILTER(CONTAINS(str(?p2),’http://dbpedia.org/ontology/’) && !isLiteral
   ↪ (?s))}}
```

Listing 5.1: An example SPARQL query for retrieving the 1-degree neighbors for *dbr:Los_Angeles*, using it as both the head (subject) and the tail (object) entity.

We only considered relations with prefix *http://dbpedia.org/ontology/* since these mapping-based relations have a much higher quality. For the purpose of our modeling strategy, we further removed duplicate triples/statements and filtered out entities that appeared less than 10 times. In the end we obtained a dataset that contains 369 Wikipedia place summaries and a DBpedia subgraph that connects these 369 place entities with various other spatial and non-spatial entities, e.g., historical figures, via different relations, thus forming our geographic knowledge graph.

For the 369 places, the average length for the Wikipedia summary is 299 words and each summary on average contains 28 links. For the geographic knowledge graph, there are all together 419,579 entities, 534 unique relations, and 3,248,715 triples/statements.

[7] https://www.w3.org/TR/rdf-sparql-query/

The data is split into a training set of 334 places and a test set including 35 places. Fig. 5.2 shows a slice of our dataset. The text in the middle is part of the summary for *Los Angeles (dbr:Los_Angeles)* and the graph surrounding the text illustrates the way in which different entities are connected with each other. We highlight the correspondence between the links in the summary and DBpedia entities.

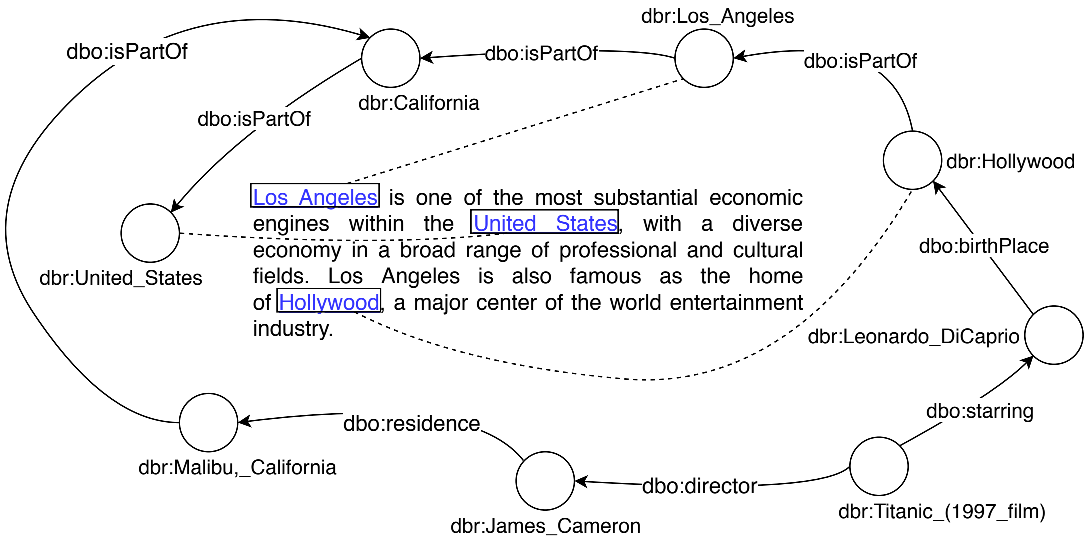

dbo:isPartOf
dbr:Los_Angeles
dbo:isPartOf
dbo:isPartOf
dbr:California
dbo:isPartOf
dbr:Hollywood
dbr:United_States
Los Angeles is one of the most substantial economic engines within the United States, with a diverse economy in a broad range of professional and cultural fields. Los Angeles is also famous as the home of Hollywood, a major center of the world entertainment industry.
dbo:birthPlace
dbr:Leonardo_DiCaprio
dbo:starring
dbo:residence
dbr:Malibu,_California
dbo:director
dbr:Titanic_(1997_film)
dbr:James_Cameron

Figure 5.2: Three links *Los Angeles*, *United States*, and *Hollywood* in this text are mapped to three entities *dbr:Los_Angeles*, *dbr:United_States*, and *dbr:Hollywood* respectively. By retrieving the 1-degree and 2-degree neighbors of these entities, we are able to find their connections as well as information about other related entities.

5.4 Methods

In this section, we introduce our spatially-explicit reinforcement learning method. Instead of pruning the graph as explored by previous studies [99], we decide to tackle the problem in a reverse manner. We formulate the task as a sequential decision making problem where we start from the simplest graph, namely a single node (the geographic entity in question), and iteratively propose to make the graph more complex and expressive by sequentially adding new relations (edges) and entities (nodes) through trial and error until the graph representation closely resembles Wikipedia's textual summary. We first introduce the reinforcement learning model by explaining the basic components such as the environment, agent, actions, states, and rewards. Our policy-based agent learns to pick meaningful relations by interacting with the geographic knowledge graph environment. Then we describe the training pipeline where the model is first trained on a supervised policy followed by being retrained using the reward function.

5.4.1 Reinforcement Learning Framework

The geographic knowledge graph summarization task is formalized as a Markov Decision Process (S, A, P_a, R_a) where two components, namely the environment and the agent, interact with each other, as shown in Fig. 5.3. $S = \{s_1, s_2, ..., s_n\}$ is a set of states that contains useful information from the history of the MDP. $A = \{a_1, a_2, ..., a_n\}$ is a set of actions that the agent can take for the state provided by the environment. Because of the memorylessness of the MDP, the state transition probability matrix $P_a(s, s') = \Pr(s_{t+1} = s'|s_t = s, a_t = a)$ represents the probability of reaching state s' at time $t+1$ after the agent takes action a in state s at time t. $R_a(s, s')$ is the immediate reward after taking action a and transitioning from state s to state s'.

To intuitively understand the process, let us suppose the MDP starts with a graph that is composed of the place entity itself and the Wikipedia summary of the place. At each step, the agent analyzes the current state (by considering information about the graph as well as the Wikipedia summary) of the process and decides to add one of the possible relations to the graph to grow it in the hope of more closely resembling the Wikipedia abstract. The agent gets a certain amount of reward depending on the extent to which this step was successful in reaching this goal. When the process terminates, i.e., an episode of MDP has been conducted, the graph is expected to be a good summary

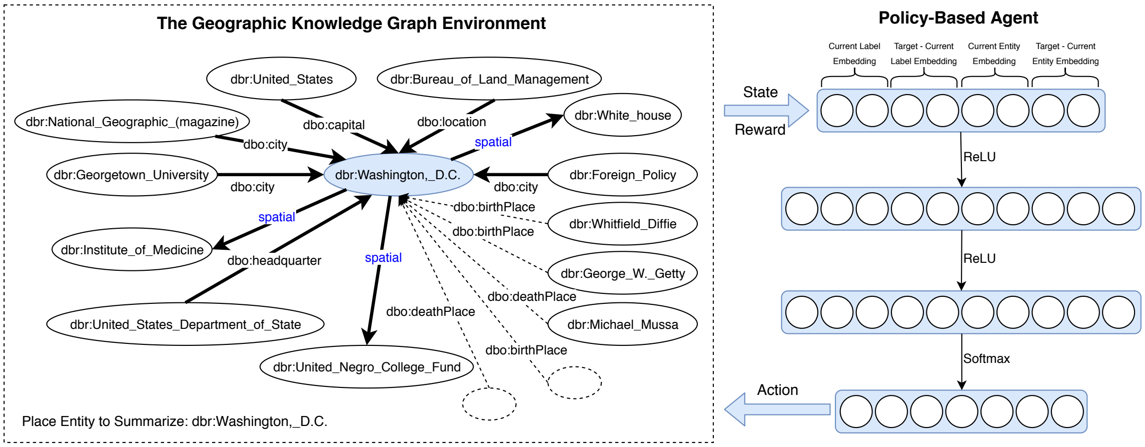


Figure 5.3: The geographic knowledge graph environment and the policy-based agent interact with each other in the reinforcement learning model. The graph environment on the left shows how the place entity *dbr:Washington,_D.C.* is connected with other spatial/non-spatial entities via various relations. The agent on the right interacts with the environment in the MDP and learns to pick relations to help summarize the graph.

of the original geographic knowledge graph for this place. The goal of the agent is to maximize the amount of reward it receives. During this process, the agent is learning to discover the sweet spot on the spectrum between information deficit (a graph with a single node for the place entity itself) and information overload (the whole geographic knowledge graph containing 419,579 nodes) by considering the textual summarization counterpart, namely the Wikipedia abstract. In order to balance the trade-off between exploration and exploitation, the behavior of the agent is defined by the stochastic policy $\pi(a|s) = \Pr(a_t = a|s_t = s)$ which is a probability distribution that determines the likelihood of the agent taking action a in state s at time step t.

In our model, the policy network (shown in Fig. 5.3) is used to learn an approximation function that captures the dynamics of the interaction and to parameterize the policy $\pi_\theta(a|s)$ of the agent. It is a fully-connected neural network with two hidden layers. Rectified Linear Units (ReLU) are used as activation functions in the hidden layers and the softmax function is used in the output layer to generate probabilities for each possible action. Before diving into the training pipeline, we further explain each concept in the context of our summarization task.

5.4.2 States

The states capture the information in the MDP. Since our model aims to capture both intrinsic and extrinsic information, we utilize the geographic knowledge graph structure as well as the semantic information from the Wikipedia summaries.

Since there are more than 400,000 entities in our geographic knowledge graph, modeling them as discrete atomic symbols using one-hot vectors in the states is not feasible. In order to provide a condensed representation of the entities, we use the translation-based knowledge graph embedding approach (TransE) [161]. The TransE model provides a scalable and generic way to embed nodes and edges in a heterogeneous graph into the same vector space. By considering the relations in the graph as translations in the embedding space, the model extracts local and global connectivity patterns between entities. The intrinsic structures of the graph are, thus, embedded in these latent representations of entities and relations. The states in the MDP are supposed to help the agent understand the current environment in order to make decisions about the next step. In this case, the entity embeddings can help capture the progress in the summarization process with respect to the Wikipedia summary. We use the sum of the entity embeddings $\mathbf{z}_t = \sum_{i \in E_t} \mathbf{e}_i$ in the current summarization graph at step t to capture the intrinsic structural information where $\mathbf{e}_i$ is the embedding for entity i in a set of entities E_t.

As these entities also appear as links in Wikipedia summaries, we denote the sum of the embeddings of entities from a target Wikipedia place summary as $\mathbf{z}_{target} = \sum_{i \in E_{target}} \mathbf{e}_i$ where E_{target} is a set of entities that appear in the target Wikipedia place summary. The intrinsic component of the state representation is defined as $\mathbf{s}_t^{intrinsic} =$

$(\mathbf{z}_t, \mathbf{z}_{target} - \mathbf{z}_t)$ where the first component (left) encodes the structure of the summarization graph at step t and the second component (right) encodes the gap between the current graph structure $\mathbf{z}_t$ and the desired structure $\mathbf{z}_{target}$.

For the extrinsic component of the state representation $\mathbf{s}_t^{extrinsic}$, we consider the labels of the entities and relations in the graph as well as the Wikipedia text summary. Neural word embeddings have proven to be an efficient and effective way of encoding meaning of words in our natural languages [102, 107]. We adopt the fastText word embedding model proposed by Bojanowski et al. [172] as it handles out-of-vocabulary words and considers the morphology of words by viewing each word as a bag of character n-grams.

After obtaining the word embeddings using the fastText model, we use the sum of the entity label and relation label embeddings $\mathbf{h}_t = \sum_{l \in L_t} \mathbf{v}_l$ to help capture the semantic information of the graph at step t. In order to obtain the latent representation of the Wikipedia textual summary, we utilize the Smooth Inverse Frequency (SIF) embedding approach to generate paragraph embeddings $\mathbf{h}_{target}$ using the word embeddings. The theoretical justification of this method is provided by Arora et al. [173]. The idea is to multiply each word vector $\mathbf{v}_w$ by the inverse of its probability of occurrence $p(w)$. Here α is a smoothing constant and is set to 0.001 by default. We then obtain $\mathbf{h}_{target}^{'}$ by summing these normalized and smoothed word vectors and dividing them by the number of words $|W|$:

$$\mathbf{h}_{target}^{'} = \frac{1}{|W|} \sum_{w \in W} \frac{\alpha}{\alpha + p(w)} \mathbf{v}_w \tag{5.1}$$

As suggested by Arora et al. [173], we obtain the matrix representation of all 369 Wikipedia summaries and remove the first principal component from this matrix to generate the final embeddings $\mathbf{h}_{target}$ for each Wikipedia place summary because the top singular vector tends to contain syntactic information and removing it cleans up the embeddings' ability to better express semantic information.

Similar to the intrinsic component, the extrinsic component of the state is represented as $\mathbf{s}_t^{extrinsic} = (\mathbf{h}_t, \mathbf{h}_{target} - \mathbf{h}_t)$ and the state representation is a concatenation of these two components:

$$\mathbf{s}_t = (\mathbf{s}_t^{intrinsic}, \mathbf{s}_t^{extrinsic}) = (\mathbf{z}_t, \mathbf{z}_{target} - \mathbf{z}_t, \mathbf{h}_t, \mathbf{h}_{target} - \mathbf{h}_t) \tag{5.2}$$

After calculating state representations, the cosine distance is calculated between the current graph and the target Wikipedia summary for both entity embeddings and label embeddings, denoted as $dist_{\mathbf{z}_t} = 1 - \cos(\mathbf{z}_t, \mathbf{z}_{target})$ and $dist_{\mathbf{h}_t} = 1 - \cos(\mathbf{h}_t, \mathbf{h}_{target})$ respectively. The termination of the process is decided by:

$$\mathcal{T} = \begin{cases} 1, & \text{if } dist_{\mathbf{z}_t} \leqslant \frac{dist_{\mathbf{z}_1}}{2} \text{ or } dist_{\mathbf{h}_t} \leqslant \frac{dist_{\mathbf{h}_1}}{2} \\ 0, & \text{otherwise} \end{cases} \tag{5.3}$$

where $dist_{\mathbf{z}_1}$ and $dist_{\mathbf{h}_1}$ denotes the initial cosine distance between the subgraph and the Wikipedia summary for entities and labels respectively. The process terminates if $\mathcal{T} = 1$. This means that if either the cosine distance for entity embeddings or label embeddings is at most half of the initial cosine distance the process will terminate.

5.4.3 Actions

Given the place entity and Wikipedia summary, the agent aims to choose actions that iteratively leads to a better summary of the geographic knowledge graph for the place in question. Starting from the initial state s_0, the policy network (shown in Fig. 5.3) outputs the probability of choosing each action a. Since there are 534 unique relations in our geographic knowledge graph, the normal action space is of size 534.

After the agent takes an action and decides to add a relation to the current subgraph, the environment checks possible ways of connecting the entities on the current subgraph with potential new entities via the chosen relation. Let us suppose (by checking the graph) that there are n potential triples to be added to the current subgraph. Each triple contains an entity that is already in the graph, the chosen relation, and a new entity (either a head or a tail entity) to be added. We use the index i to denote the new entity where $1 \leqslant i \leqslant n$ and $triple_i$ to denote the corresponding triple for entity i. Our model picks the triple (and the new entity) among all candidate triples from a distribution where the probability for each triple $p(triple_i)$ is proportional to the information content of the new entity:

$$p(triple_i) = \frac{-\log(p(i))}{\sum_{j=1}^{n} -\log(p(j))} \tag{5.4}$$

where $p(i)$ is the probability of encountering entity i in the whole geographic knowledge graph and $-\log(p(i))$ is its information content. The rationale behind this approach is that entities that are rich in information content carry latent information that can more efficiently enrich our knowledge about the place we wish to summarize.

In addition to the normal 534 actions, we also propose a novel step by including a dedicated *spatial* action to make the model *spatially-explicit*. This idea stems from the data-driven approach that exploits the hidden patterns of geographic data [16] and is inspired by previous work on spatially-explicit models where spatial contextual information facilitates place type embeddings [48], image classification for places [49], and geographic question answering [8]. Following a similar school of thought, we aim to utilize spatial context to help improve geographic knowledge graph summarization. Another reason to incorporate this special *spatial* action is that, as mentioned in Section 5.1, it helps in discovering missing links in the geographic knowledge graph by connecting spatially related entities together. Simply put, a human (textual) summary of San Diego will include the adjacent border with Mexico. However, such adjacency relation does not exist in DBpedia, and, hence, Tijuana (and Mexico in general) would not be reachable within the graph for the agent.

The *spatial* action itself is modeled as an extra action that the agent can take at any step t. However, if the agent decides to take a *spatial* action, our model only gathers candidates that are geographic entities and are not connected with any entities in the current subgraph directly. We execute a spatial query retrieving all geographic entities within k-meter radius of the place we want to summarize. Our spatially-explicit model selects one geographic entity among these candidate geographic entities from a distribution where the probability for each candidate $p(i)$ is proportional to the inverse of the distance between the candidate and the place q in question:

$$p(i) = \frac{d(i,q)^{-1}}{\sum_{j=1}^{n} d(j,q)^{-1}} \tag{5.5}$$

where $d(i,q)$ denotes the geodesic distance between candidate i and place q. This inverse distance strategy favors nearby geographic entities over distant ones. While the spatial radius buffer gives a local geographic view around the center place entity, we also propose to incorporate a global view that is modeled by the PageRank score of each entity in the whole geographic knowledge graph [174]. Intuitively, some places, e.g., landscape features, are characteristic for an entity to be summarized despite their distance due to their overall importance. Mount Fuji is such an example despite its distance of over 100 km from Tokyo. Each entity is assigned a score pr_i after running the PageRank algorithm. This score represents the relative importance of each entity in the graph by examining the incoming and outgoing link connections. By combining the global graph view and the local geographic view, we propose to use a weighted inverse distance in the probability calculation:

$$p(i) = \frac{pr_i d(i,q)^{-1}}{\sum_{j=1}^{n} pr_j d(j,q)^{-1}} \tag{5.6}$$

After deciding on the relations and entities to add into the subgraph through either spatial or non-spatial actions, new state representations are generated using the methods explained in Section 5.4.2 and the new state is then presented to the agent to help it decide on the next action.

5.4.4 Rewards

The reward function plays an important role in guiding the agent to summarize the geographic knowledge graph as the goal of our reinforcement learning model is to find an optimal behavior strategy for the agent to obtain optimal rewards. In our model, there are three components in the reward function, namely similarity score, diversity score, and connection score.

In order to help the agent select the actions (relations) that make the subgraph representation resembles the Wikipedia summary representation from such a large action space, an intuitive way is to incorporate such mechanism in the immediate reward. In addition to cosine distance calculated after the agent takes an action as described in Section 5.4.2, the cosine similarity is also calculated. The normal similarity score is then defined as the sum of the cosine similarities:

$$r_{similarity}^{normal} = \cos(\mathbf{z}_t, \mathbf{z}_{target}) + \cos(\mathbf{h}_t, \mathbf{h}_{target}) \tag{5.7}$$

where larger cosine similarity values will result in higher scores for the reward component $r_{similarity}^{normal}$. Moreover, considering the fact that sometimes the TransE model does not handle one-to-many and many-to-many relationships well [161] and summing or averaging the entity embeddings may exacerbate such issues because the connectivity information of individual nodes/entities may be dwarfed by the crude aggregation of other nodes/entities, we propose to substitute the entity similarity score $\cos(\mathbf{z}_t, \mathbf{z}_{target})$

by another measurement to highlight the difference of the intrinsic structure between the subgraph and the Wikipedia summary. Such a measurement is inspired by the Hausdorff distance commonly-used to measure the difference between two geometries. Instead of using a metric such as Euclidean distance as in Hausdorff distance, we use cosine distance because it is insusceptible to the change of magnitude of embedding vectors. This measurement is defined as:

$$sim_{maxmin}(E_t, E_{target}) = 1 - \max_{i \in E_t} \min_{j \in E_{target}} (1 - \cos(\mathbf{e}_i, \mathbf{e}_j)) \tag{5.8}$$

where E_t is a set of entities on the subgraph at time step t, E_{target} is a set of entities in the Wikipedia summary, and $\mathbf{e}_i$ and $\mathbf{e}_j$ are entity embeddings for entity i and j respectively. The max-min similarity score is then defined as:

$$r^{maxmin}_{similarity} = sim_{maxmin}(E_t, E_{target}) + \cos(\mathbf{h}_t, \mathbf{h}_{target}) \tag{5.9}$$

While there are 535 possible relations/actions (including the *spatial* action), these relations follow a long-tail distribution, which might lead the agent to pick the most possible relations in order to avoid penalties. In addition, a good graph summary should exhibit a balance between diversity and uniformity. In light of this, we propose to incorporate a diversity score into the reward function:

$$r_{diversity} = \begin{cases} +0.5, & \text{if relation is not already on the subgraph} \\ -0.5, & \text{otherwise} \end{cases} \tag{5.10}$$

Since it is possible that the model might pick relations and entities that are not directly connected to the place entity in question, we would like to discourage such behavior. For example, to summarize *dbr:Los_Angeles*, the model might add new triples regarding *dbr:California* (because *dbr:California* became part of the subgraph for *dbr:Los_Angeles* at some point) instead of *dbr:Los_Angeles*. This behavior is the result of the data bias in knowledge graphs [175] as prominent entities are safer for the model to target and would mislead the model to summarize the wrong place. In order to alleviate this potential issue, we propose to include the connection score:

$$r_{connection} = \begin{cases} +0.5, & \text{if entity is directly connected to the place} \\ -0.5, & \text{otherwise} \end{cases} \tag{5.11}$$

The reward function is then defined as the combination of the three components:

$$R = r_{similarity} + r_{diversity} + r_{connection} \tag{5.12}$$

It is worth noting that simply reducing relations to be selected from 1-degree queries relative to the entity to be summarized would not be a suitable solution. This would restrict the summary subgraph to a star-shape.

5.4.5 Training Procedure

As mentioned in Section 5.4.1, we use a policy-based method to train our spatially-explicit reinforcement learning model. The advantage of policy-based methods over value-based methods such as Q-learning [176] and SARSA [177] is that they solve an easier problem by optimizing the policy π directly, can provide a stochastic policy, and can be applied to a wider range of problems where the state space is large or even continuous. The objective of the policy-based method is to maximize the total future expected rewards J:

$$J(\theta) = \mathbb{E}_{s \sim \Pr(s), a \sim \pi_\theta(a|s)} R(s, a) \tag{5.13}$$

Following the REINFORCE (Monte Carlo Policy Gradient) method [169], the policy network is updated using the gradient:

$$\begin{aligned}\nabla_\theta J(\theta) &= \mathbb{E}_{s\sim \Pr(s), a\sim\pi_\theta(a|s)} Q(s,a)\nabla_\theta \log \pi_\theta(a|s) \\ &\approx \frac{1}{N}\sum_{i=0}^{N}\sum_{s,a\in eps_i} Q(s,a)\nabla_\theta \log \pi_\theta(a|s)\end{aligned} \tag{5.14}$$

where N episodes eps are sampled from the process, $Q(s_t = s, a_t = a) = \mathbb{E}[G_t|s_t = s, a_t = a]$ is the expected return starting from state s after taking action a, and the return $G_t = \sum_{k=0}^{\infty}\gamma^k R_{t+k+1}$ is the total discounted reward from time step t with discount factor $\gamma \in [0,1]$. A low γ value implies that the agent is myopic in evaluating the situation and values immediate reward over delayed future reward. In addition, similar to the idea of diversity reward in Section 5.4.4, we include the entropy of the policy as a regularization term in the optimization where we encourage a more diversified set of actions. The entropy is defined as:

$$H(\theta) = -\sum_{a\in A}\pi_\theta(a|s)\log\pi_\theta(a|s) \tag{5.15}$$

In order to maximize the total future expected rewards J and the entropy H, the loss function is formulated as:

$$\mathcal{L}_{REINFORCE} = -(J + \alpha H) \tag{5.16}$$

where α is the regularization factor.

Due to the size of the action space, it would be challenging for the policy agent to learn to pick actions purely based on trial and error. In order to solve this problem and inspired by imitation learning [178] and the training pipeline proposed by Silver et al. [179], we first train our model with supervised learning and then retrain the supervised policy with the proposed reward function to learn summarizing the geographic knowledge graph.

For the supervised learning stage, we use the links in Wikipedia summaries to help gather positive training samples. We query the whole graph to check if the links in the Wikipedia place summary are directly connected to the place entity itself and keep track of these connections. In addition, in order to learn about the *spatial* action as well, we randomly incorporate nearby geographic entities via the special *spatial* relation. This procedure is applied to every place in the training place set in order to get our positive training samples for the supervised learning. A reward of +1 is used for each step in these positive training samples. After the supervised training stage, we retrain the model using the reward function described in Section 5.4.4 to help the agent pick up desired relations to better summarize the graph. Summarizing one place is considered an episode *eps*. The model starts with a single node (the place entity itself) for the graph and follows the stochastic policy $\pi(a|s)$ to iteratively add relations. We limit the maximum length of the episode with an upper bound max_eps_len to improve the training efficiency.

5.5 Experiment and Results

In this section, we explain our experiment setup for the model, describe the evaluation metrics used to test the model performance, and present our results and findings.

5.5.1 Implementation Details

Since we use 50-dimensional vectors for both entity and label embeddings, the resulting state representations are 200-dimensional vectors. For *spatial* action, we use a search radius of $k = 100,000$ meters in our geopatial query. The discount factor γ for the cumulative reward we use in the model is 0.99. In the policy network, the first hidden layer has 512 units and the second hidden layer has 1024 units. The Adam Optimizer [180] is used to update the parameters in the policy network. The upper bound for the episode length is set to $max\_eps\_len = 20$.

Different alternative settings are proposed for actions and rewards in Section 5.4.3 and Section 5.4.4 respectively. The alternatives in the actions component are non-spatial actions vs. spatial actions and unweighted inverse distance (Eq. 5.5) vs. PageRank-weighted inverse distance (Eq. 5.6). The alternatives in the reward component are $r_{similarity}^{normal}$ vs. $r_{similarity}^{maxmin}$. In order to better understand the contribution of different component alternatives and to testify our assumption that spatially-explicit models are superior in modeling geographic data, we examine our method with different combinations of these alternatives, resulting in 5 models, namely $RL_{nonspatial-normal}$ (model without spatial actions using $r_{similarity}^{normal}$ score), $RL_{spatial-normal}$ (model with spatial actions using $r_{similarity}^{normal}$ score), $RL_{nonspatial-maxmin}$ (model without spatial actions using $r_{similarity}^{maxmin}$ score), $RL_{spatial-maxmin}$ (model with spatial actions using $r_{similarity}^{maxmin}$ score), and $RL_{spatial-maxmin-pr}$ (model with spatial actions and PageRank-weighted inverse distance using $r_{similarity}^{maxmin}$ score).

5.5.2 Results

To evaluate the models, we consider the intrinsic and extrinsic components separately. For the summarization results, we would like to see the improvements of using our summarization approach compared with the initial information, i.e., we compute the difference between the cosine similarity of the summarized graph and the Wikipedia summary and the cosine similarity of the initial place entity/label and the Wikipedia summary:

$$\text{diff}_{entity} = \cos(\mathbf{z}_T, \mathbf{z}_{target}) - \cos(\mathbf{z}_1, \mathbf{z}_{target}) \tag{5.17}$$

$$\text{diff}_{label} = \cos(\mathbf{h}_T, \mathbf{h}_{target}) - \cos(\mathbf{h}_1, \mathbf{h}_{target}) \tag{5.18}$$

where $\text{diff}_{entity} \in [-2, 2]$ and $\text{diff}_{label} \in [-2, 2]$ are the difference of cosine similarities between entity and label embeddings and $\mathbf{z}_T$ and $\mathbf{h}_T$ are the final entity and label embeddings for the summarized graph. Higher diff scores show better summarization results. In addition to this evaluation metrics, we also calculate the Mean Reciprocal Rank (MRR) score for these 5 models. We calculate the cosine similarity scores between the summarized graph of the place with all 35 candidate places in our test set and then rank them. We record the rank position of the corresponding Wikipedia place summary for each place entity, take the reciprocal of the rank, and then calculate the mean of these reciprocal ranks for all 35 places in the test set. Higher MRR scores correspond to better model performance.

Table 5.2 and Table 5.3 show the diff_{entity} and diff_{label} scores for all 35 test places. As we can see, on average all 5 models show positive diff_{entity} and diff_{label} scores, implying that these models are effective in creating subgraphs that resemble the Wikipedia summary, thus facilitating the summarization of these places. In general, the scores for the intrinsic component diff_{entity} are lower than the ones for the extrinsic component diff_{label} for the same place and on average. One reason might be that the TransE model takes into account the local and global connectivity information of entities and since the place entity itself is usually closely connected with the Wikipedia links for this place entity the initial single-node graph $\mathbf{z}_0$ tends to be quite similar to $\mathbf{z}_{target}$, making further improvements less prominent. On average, incorporating the *spatial* action or using the $r^{maxmin}_{similarity}$ component in the reward function helps improve the performance and including both further improves the result. The best model is $RL_{spatial-maxmin-pr}$ for both the intrinsic diff_{entity} and extrinsic diff_{label} components. On average it has a 147% and 90% increase compared with the $RL_{nonspatial-normal}$ model for the intrinsic and extrinsic components respectively.

Table 5.1: MRR result for 5 models.

	Entity	Label
$RL_{nonspatial-normal}$	0.9190	0.6975
$RL_{spatial-normal}$	0.9380	0.7183
$RL_{nonspatial-maxmin}$	0.9428	0.7095
$RL_{spatial-maxmin}$	0.9571	0.7396
$RL_{spatial-maxmin-pr}$	0.9523	0.7742

By examining the results in Table 5.2 and Table 5.3 for $RL_{spatial-normal}$ and $RL_{nonspatial-maxmin}$, we can see that adding the spatial action is beneficial for the model to capture more semantic information and using the $r^{maxmin}_{similarity}$ reward component facilitates the model to capture intrinsic structural information as the diff$_{label}$ result is better for $RL_{spatial-normal}$ than for $RL_{nonspatial-maxmin}$ and vice versa in the case of diff$_{entity}$. The MRR result in Table 5.1 aligns with our findings.

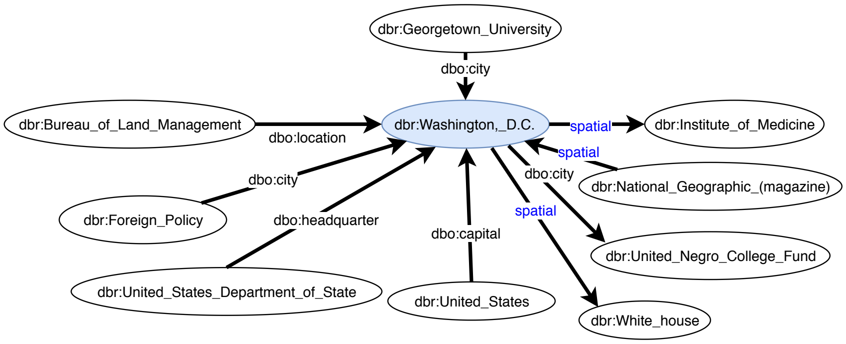

dbr:Georgetown_University
dbo:city
dbr:Bureau_of_Land_Management
dbo:location
dbr:Washington,_D.C.
spatial
dbr:Institute_of_Medicine
spatial
dbr:National_Geographic_(magazine)
dbo:city
dbr:Foreign_Policy
dbo:headquarter
dbo:city
spatial
dbo:capital
dbr:United_Negro_College_Fund
dbr:United_States_Department_of_State
dbr:United_States
dbr:White_house

Figure 5.4: Summarization result for *dbr:Washington,_D.C.*.

Fig. 5.4 and Fig. 5.5 show the summarization results for *dbr:Washington,_D.C.* and *dbr:Guangzhou* using the $RL_{spatial-maxmin-pr}$ model. The model learns to pick different relations such as *dbo:capital*, *dbo:city*, *dbo:headquarter*, *dbo:location*, *dbo:isPartOf*, and the *spatial* relation. In the case of *dbr:Washington,_D.C.*, the relationship between

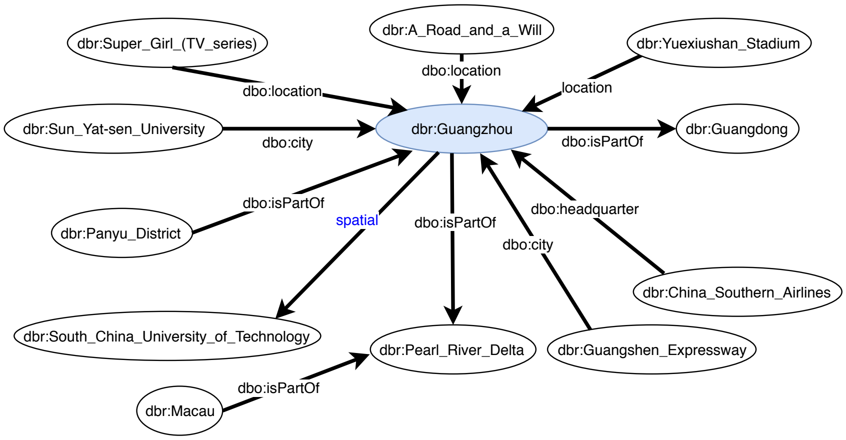


Figure 5.5: Summarization result for *dbr:Guangzhou*.

dbr:White_House and *dbr:Washington,_D.C.* is missing in the original geographic knowledge graph. Without the *spatial* relation, such certainly important information would have been lost. Our spatially-explicit model shows advantage over non-spatial models. In the case of *dbr:Guangzhou*, as we incorporate the connection reward $r_{connection}$ into the model, it refrains from summarizing other entities even though *dbr:Macau* is included in the subgraph at some point.

5.6 Conclusions and Future Work

In this research, we introduced and motivated the need for geographic knowledge graph summarizations and proposed a spatially-explicit reinforcement learning framework to learn such graph summaries. Due to the lack of benchmark and standard datasets, we collected a dataset that contains Wikipedia place summaries as well as a geographic knowledge graph for 369 places as seed. In order to explore different possibilities of modeling the summarization process, we suggested different alternatives for the actions and rewards formulation in the model. By testing 5 models using different combinations of the alternative components, we conclude that a spatially-explicit model yields superior summarization results compared to non-spatial models, thereby confirming that spatial is indeed special as far as knowledge graph summarization is concerned.

In the future, we would like to test if reducing the variance in the Monte Carlo Policy Gradient method by using an advantage function or the Actor-Critic framework would help improve the performance. Finally, our and other approaches do not consider datatype properties which is an important goal for future research.

Table 5.2: Result for diff$_{entity}$ scores for 35 test places for 5 models.

	$RL_{nonspatial-normal}$	$RL_{spatial-normal}$	$RL_{nonspatial-maxmin}$	$RL_{spatial-maxmin}$	$RL_{spatial-maxmin-pr}$
dbr:New_Orleans	0.0287	0.0721	0.0757	0.0866	0.0946
dbr:Boston	0.0008	0.0091	0.0101	0.0114	-0.0012
dbr:Canberra	0.0863	0.1038	0.1020	0.1084	0.1206
dbr:Osaka	0.0759	0.1015	0.0953	0.1135	0.1057
dbr:Lyon	0.0529	0.0676	0.0617	0.0754	0.0764
dbr:Heidelberg	0.0614	0.0858	0.0901	0.1062	0.1086
dbr:Krakw	0.0065	0.0166	0.0409	0.0449	0.0491
dbr:Johannesburg	0.0097	0.0176	0.0131	0.0325	0.0233
dbr:Oxford	0.0231	0.0257	0.0344	0.0602	0.0941
dbr:Milan	0.0043	0.0561	0.0445	0.0746	0.1092
dbr:Montreal	0.0050	0.0192	0.0250	0.0342	0.0451
dbr:Braslia	0.0265	0.0699	0.0651	0.0951	0.1140
dbr:Tel_Aviv	0.0122	0.0238	0.0234	0.0366	0.0391
dbr:Frankfurt	0.0777	0.1073	0.1125	0.1212	0.1235
dbr:Philadelphia	0.0066	0.0168	0.0451	0.0573	0.0682
dbr:Washington,_D.C.	0.0403	0.0545	0.0578	0.0701	0.0949
dbr:Shanghai	0.0323	0.0626	0.0543	0.1002	0.0925
dbr:Saint_Petersburg	0.0279	0.0562	0.0497	0.0699	0.0787
dbr:Seattle	0.0194	0.0310	0.0228	0.0689	0.0406
dbr:San_Diego	0.0350	0.0484	0.0508	0.1267	0.0868
dbr:Seoul	0.0215	0.0291	0.0273	0.0558	0.0599
dbr:Las_Vegas	0.0107	0.0202	0.0332	0.0648	0.0818
dbr:Athens	0.0159	0.0390	0.0499	0.0612	0.0675
dbr:Guangzhou	0.0090	0.0176	0.0378	0.0935	0.0802
dbr:Hangzhou	0.0240	0.0464	0.0500	0.0713	0.0680
dbr:Madrid	0.0380	0.0566	0.0594	0.0670	0.0782
dbr:Edinburgh	0.0335	0.0767	0.0771	0.0931	0.1053
dbr:Barcelona	0.0130	0.0239	0.0383	0.0577	0.0813
dbr:Denver	0.0239	0.0412	0.0498	0.0631	0.0641
dbr:Mexico_City	0.0044	0.0149	0.0177	0.0411	0.0397
dbr:Manila	0.0606	0.0756	0.0859	0.0891	0.0953
dbr:Amsterdam	0.0913	0.1046	0.0966	0.1112	0.1129
dbr:Ho_Chi_Minh_City	0.0495	0.0614	0.0591	0.0848	0.0684
dbr:Kyoto	0.0377	0.0651	0.0561	0.0728	0.0732
dbr:Prague	0.0123	0.0192	0.0183	0.0417	0.0212
Average	0.0307	0.0496	0.0523	0.0732	0.0760

Table 5.3: Result for diff_{label} scores for 35 test places for 5 models.

	$RL_{nonspatial-normal}$	$RL_{spatial-normal}$	$RL_{nonspatial-maxmin}$	$RL_{spatial-maxmin}$	$RL_{spatial-maxmin-pr}$
dbr:New_Orleans	0.2520	0.3804	0.3725	0.3883	0.3959
dbr:Boston	0.1476	0.3025	0.3027	0.3765	0.4243
dbr:Canberra	0.1033	0.2775	0.2532	0.3138	0.3829
dbr:Osaka	0.0747	0.1172	0.1078	0.1479	0.1971
dbr:Lyon	0.3296	0.4490	0.4396	0.5298	0.5111
dbr:Heidelberg	0.1653	0.2079	0.2140	0.2321	0.2592
dbr:Krakw	0.1041	0.1534	0.1158	0.2258	0.2181
dbr:Johannesburg	0.1593	0.2436	0.2461	0.2931	0.3002
dbr:Oxford	0.1656	0.3358	0.3206	0.3899	0.4136
dbr:Milan	0.2647	0.3249	0.3217	0.3579	0.3823
dbr:Montreal	0.2049	0.2320	0.2753	0.3004	0.3074
dbr:Braslia	0.0676	0.1682	0.0694	0.2071	0.2148
dbr:Tel_Aviv	0.1588	0.2143	0.2069	0.2288	0.2431
dbr:Frankfurt	0.1867	0.3025	0.2900	0.3347	0.3386
dbr:Philadelphia	0.0762	0.1274	0.1214	0.1484	0.1618
dbr:Washington,_D.C.	0.1509	0.3166	0.3290	0.3655	0.3889
dbr:Shanghai	0.1655	0.1840	0.1810	0.2689	0.3629
dbr:Saint_Petersburg	0.1622	0.1981	0.1911	0.2506	0.2381
dbr:Seattle	0.2090	0.2609	0.2634	0.2881	0.2944
dbr:San_Diego	0.2412	0.3575	0.2752	0.3962	0.3986
dbr:Seoul	0.1295	0.1616	0.2061	0.3086	0.2893
dbr:Las_Vegas	0.1652	0.2300	0.2197	0.3613	0.3678
dbr:Athens	0.1770	0.1999	0.2258	0.2466	0.2390
dbr:Guangzhou	0.1122	0.1711	0.1693	0.2193	0.2334
dbr:Hangzhou	0.1045	0.2032	0.1864	0.2151	0.2397
dbr:Madrid	0.1624	0.2214	0.2232	0.2373	0.2364
dbr:Edinburgh	0.1938	0.2737	0.2695	0.3708	0.3944
dbr:Barcelona	0.0697	0.2311	0.2140	0.2528	0.2375
dbr:Denver	0.5028	0.6273	0.6034	0.6688	0.6421
dbr:Mexico_City	0.1383	0.1610	0.1698	0.1869	0.2187
dbr:Manila	0.1013	0.2114	0.1852	0.2595	0.2407
dbr:Amsterdam	0.0745	0.1418	0.1420	0.2233	0.2186
dbr:Ho_Chi_Minh_City	0.2000	0.2974	0.2857	0.3321	0.3603
dbr:Kyoto	0.1350	0.1801	0.1718	0.2304	0.2456
dbr:Prague	0.1537	0.3808	0.3868	0.4317	0.4620
Average	0.1659	0.2527	0.2444	0.3025	0.3159

Chapter 6

Summarizing Geographic Knowledge: A Case Study of Enriching Geocoding Services

In this chapter, we develop an enriched geocoding service using knowledge graphs and illustrate the importance of summarization to help users understand the context knowledge of different geographic entities and gain insights about spatial patterns. the chapter explains the motivation of developing this web map interface, presents related work on georeferencing and geospatial knowledge graphs, discusses data source selection, formalize the geocoding enrichment process, illustrates the entropy-based summarization method, introduces the serverless and scalable framework, and demonstrate the utilities of the geographic knowledge map using two example functions.

6.1 Introduction

Since its inception dating back as many as 55,000 years [181], map has become an essential element of our life – from various atlases of the world to numerous applications embedded with maps on mobile devices. With the rapid development of GIScience and computer science, the making of maps has transitioned from a cartographer's privilege to a layman's trivia. The advances in web technologies, especially the rise of Web 2.0, have lead to the renaissance of maps [182]. By hybridizing different data sources, map mashup has extended the original meaning of map, making it an enriched source of obtaining geographic knowledge.

One essential component of all these web map applications is geocoding. Geocoding is a computational process that bridges the gap between humans recognition of places and machine-readable place representations. Geocoding services have become so ubiquitous that people have taken them for granted in daily life. For example, in routing and navigation, geocoding acts as a proxy to encode addresses, intersections, or even places of interests (POI) into coordinates so that more complicated route optimization algorithms can proceed using these coordinates.

Unlike traditional gazetteers, which normally include spatial footprints, place types and toponyms [183], places in human mind are more vivid than simple coordinates. For instance, when someone talks about a place, he will most likely also talk about the demographics of the place such as the population distribution of a country or terrains of a place such as whether an area is mountainous, because people tend to use such associated information rather than coordinates to relate to places. We call these pieces of information pertaining to places as examples of geographic knowledge. Such geographic knowledge can be grouped into two categories based on whether or not the knowledge is obtained from reasoning: primary geographic knowledge and derived geographic knowledge. Primary geographic knowledge does not require any extent of reasoning. The population of a city is primary geographic knowledge because it is a piece of factual information that does not include any insight or external experience whereas the knowledge that China is the most populous country in the world is derived geographic knowledge for the simple reason that this information is obtained from a comparison. Because of the heterogeneity of geographic information, geographic knowledge can be obtained from a variety of sources and in different formats. Conventional approaches to utilize this geographic knowledge would entail complex techniques such as natural language processing especially if the knowledge was originated from unstructured text data, which is the major data format on the Web. The geospatial semantics paradigm [155] has provided the GIScience community with an alternative and a more promising source of geographic knowledge.

By incorporating geospatial semantics, more specifically geographic knowledge graphs, with the traditional geocoding mechanism, we present a system that can help users discover geographic knowledge of an area of interest. While the enriched geocoding system is better at conveying useful information than traditional geocoding systems, the challenge of selecting, presenting, and summarizing the large amount of information to end users arises. In order to tackle this challenge, we employ a method based upon information theories to prune the part of the geographic knowledge graph that is associated with the matching entities to prevent flooding users with loads of information. The raison d'être of this approach is four-fold. First and foremost, by incorporating one or two knowledge graphs, the entire Linked Data Cloud, in essence, is exposed to us due to the interconnection among numerous datasets. As of January 2019[1], the Linked Data Cloud consists of about 1,234 individual datasets, providing not only domain-agnostic

[1] Linking Open Data cloud diagram 2019, by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch, Richard Cyganiak, and Vladimir Andryushechkin. http://lod-cloud.net/

knowledge but also domain-specific content which includes geography, government, life sciences, linguistics, media, publications, social networking and user-generated information. All this data can be potentially used with our approach, drastically enriching our knowledge about places. Second, since all of them share the same data model — RDF — and are highly structured, machines can easily interpret them without sophisticated data massaging such as natural language processing. Third, the ontologies that come with these datasets are the proxies of higher level human knowledge which can further enrich our geographic knowledge. The knowledge from ontologies can be subclass-superclass relationships, transitivity of certain properties, topological relationships, etc. Fourth, by introducing an information theory-based summarization approach, the system is able to strike the right balance between information overload and information deficit in order to help users navigate and assimilate important geographic knowledge empowered by the enriched geocoding service.

The remainder of this paper is organized as follows. In Section 6.2, we will review some of the related work on georeferencing and geospatial linked open data and emphasize the key difference between their work and our work. In Section 6.3, we will talk about the implementation details and a serverless framework we adopt. Section 4 will be focusing on demonstrating our geographic knowledge map interface and Section 5 will conclude our research as well as point out some future directions.

6.2 Related Work

In this section, we summarize two areas of related work. The first area is georeferencing, which is a broader concept that includes geocoding. While geocoding technologies are relatively mature, we also review existing work on geotagging and geoparsing and point out that our research has a different perspective comparing to existing work in georeferencing. The second area is geospatial Linked Open Data (a.k.a. geographic knowledge graphs). We review existing work combining the spatial component in GIScience with knowledge graphs from both the data and methodology perspectives.

6.2.1 Georeferencing

Georeferencing is a broad term that describes the process of associating data with spatial coordinate systems. It usually includes geotagging, geocoding and geoparsing. Geotagging means manually assigning geographic identifications to an object. Geocoding in most cases requires gazetteers to facilitate the mapping between textual geographic information such as place name or address and spatial coordinate systems. Geoparsing, however, goes beyond geocoding in that it deals with ambiguous geographic references in unstructured natural language texts.

Research on geocoding has been focusing on two perspectives: geocoding methods and geocoding quality. Hutchinson and Veenendaal [184] developed an agent-based method that utilized the belief, desire, intention (BDI) model for intelligent geocoding. Ratcliffe [185] examined the accuracy of geocoding based on TIGER-typed files in relation to cadastral and census areal units. Zandbergen [186] proposed a comprehensive framework to evaluating the quality of geocoding with different address data models, namely address points, parcels and street networks. In order to increase the accuracy, most geocoding systems provide multiple candidates for each query and utilize a hierarchy-based criterion to select the best result. Goldberg and Cockburn [187] formalized the candidate selection criteria in geocoding and presented three alternative strategies, namely uncertainty-based, gravitationally-based, and topologically-based strategies. While many geocoding methods are particularly designed for spatial data, Murray et al. [188] argued that it is especially challenging to geocode spatio-temporal data and designed a geocoding method for spatio-temporal data, taking advantage of the use of supplementary land use information, aerial photographs and local knowledge. For the task of georeferencing ambiguous or sometimes cryptic place names in social networks, Davis Jr et al. [189] recursively expanded the network of locatable users using social relationships to enrich the location information of tweets. Georeferencing has also been used in Search and Rescue (SAR) in a context that lacks explicitly spatial data [190]. Our work, however, differs from the above-mentioned research in that it does not focus on either the method or quality of geocoding or georeferencing in general but pays more attention on enriching geocoding results instead.

6.2.2 Geospatial Linked Open Data

From a data perspective, geospatial Linked Open Data represents the marriage of Linked Data and GIS. Both the Linked Data community and GIS community are working actively together to refine the definition of geospatial Linked Open Data. Their mutual influence has incubated a variety of research studies and technical advances in both fields. Many Linked Data sources such as DBpedia, Freebase and Wikidata have hosted a large amount of geographic data. The GIS community also started to provide dedicated geospatial Linked Dataset such as GeoNames Linked Data and LinkedGeoData. Abdelmoty et al. [191] summarized the limitations of Semantic Web language for the representation of geographic place and proposed approaches that combine rules and Semantic Web language to alleviate such limitations. Lopez-Pellicer et al. [192] proposed to publish resources alongside their metadata in RDF to help identify real world entities with geospatial Web resources. In an attempt to facilitate geographic information retrieval, Lopez-Pellicer et al. [193] presented a geographic knowledge vocabulary and a tool for building large knowledge bases of geographic places. In order to provide spatial reasoning support for geographic knowledge graphs, as contributors to the GeoSPARQL OGC standard, Battle and Kolas [194] examined the overall state of geospatial data in the Semantic Web and implementated GeoSPARQL in the Parliament triple store. From a data quality perspective, Janowicz et al. [86] presented a comprehensive study of systematic errors and their potential causes in geographic knowledge graphs.

From a methodology perspective, spatially-explicit models have been using the idea of geospatial inductive bias to handle the geospatial semantics in geographic knowledge graphs. For example, Chapter 3 presented a latent representation learning method that utilized biased sampling strategy to uncover hidden geographic patterns in order to embed different place types. These place types are essential components in the ontology for various geographic knowledge graphs. Chapter 5 developed a spatially-explicit reinforcement learning model to help summarize geographic knowledge graphs. The proposed model introduced an explicit spatial action for the agent to capture the geospatial semantics in geographic knowledge graphs. Mai et al. [8] incorporated the weighting scheme based on geographic distance between spatial entities in the knowledge graph in order to learn better embeddings for question answering tasks.

6.3 Framework & Implementation

In this section, we first describe our data source selection process by comparing different geocoding services as well as geographic knowledge graphs. We propose 4 criteria on which our selection is based. Then we discuss our geocoding enrichment mechanism by giving formal definition and implementation details. In order to make our system more scalable, we take advantage of the serverless architecture provided by Amazon Web Services (AWS) and explain the deployment of our system on AWS. Finally, we present an entropy-based summarization approach for selecting a subset of informative nodes from the original geographic knowledge graph in order to help users navigate the geographic knowledge map that we develop.

6.3.1 Data Source

For the geocoding service, we choose both OpenStreetMap (OSM) Nominatim[2] and GeoNames[3] as they are freely available robust geocoding services. We make use of OSM data because it provides a comprehensive set of APIs and tools for researchers to take advantage of the crowd sourcing data. GeoNames is a geographic database that covers toponyms around the world. GeoNames has both a traditional version and a Linked Data version. OSM and GeoNames are used as access points into the Linked Data Cloud as they are partially linked to Wikipedia from which many Linked Datasets are originated primarily. It is not a trivial task to choose our knowledge graph data sources from a selection of 1,234 datasets. We narrow it down by first filtering out datasets that are not directly associated with geographic information. In this case, we only focus on datasets that are either cross-domain or in geography domain. Furthermore, the quality and coverage of these datasets vary a lot. We review the use cases and research [162, 163, 164, 195, 196] based on these datasets and select 4 candidate knowledge graph datasets. These 4 candidates are: Freebase, DBpedia, Wikidata, LinkedGeoData. We pick our final datasets based on 4 major criteria: 1) whether the data source contains a sufficient amount of geographic entities, 2) whether the data source is actively maintained and up-to-date, 3) whether there is a clear correspondence between the data source and OSM Nominatim or GeoNames, 4) whether the data source has a comprehensive coverage of different properties/predicates for each entity.

Knowledge graph data in LinkedGeoData can be traced back to OpenStreetMap

[2] `http://nominatim.openstreetmap.org/`

[3] `http://api.geonames.org/findNearbyJSON`

Table 6.1: A comparison of different candidate data sources with respect to 4 criteria.

	LinkedGeoData	Freebase	DBpedia	Wikidata
Geographic Entities	Yes	Yes	Yes	Yes
Up-to-date	Yes	No	Yes	Yes
Clear Correspondence	Yes	Partially	Yes	Yes
Comprehensive Coverage	No	Yes	Yes	Yes

data. Every entity in this case is a geographic entity and it is being regularly maintained and updated based on the OpenStreetMap data. Since LinkedGeoData contains mostly geographic entities, it lacks the non-spatial entities that are needed in providing a general spatial knowledge for map users. The relations/predicates in the LinkedGeoData graph are typically about geometries, thus failing to provide a comprehensive view of geographic entities. Freebase data has been used frequently in knowledge graph related research as benchmark and standard dataset and it contains geographic entities. However, this dataset is not currently being maintained and only a snapshot of the data is available. As a consequence, it does not correspond to the evolving OSM or GeoNames data very well.

In the end, we select 2 datasets — DBpedia and Wikidata — which satisfy the 4 criteria. All these data sources are freely accessible. DBpedia aims to extract structured information from Wikipedia and provides numerous data for semantic queries on the Web. Unlike DBpedia, Wikidata creates structured information from scratch and is constantly maintained and curated by a community similar to the one for Wikipedia. DBpedia and Wikidata represent two of the most interconnected nodes in the Linked Data Cloud and are used as the sources of our geographic knowledge. While there is no direct concept of entity in GeoNames and OSM, in GeoName unique features are treated as entities and in OSM each node, way or relation is considered as an entity. As dedicated geographic datasets, GeoNames and OSM have a much larger coverage of spatial entities. It's also interesting to note that, although DBpedia predates Wikidata, its coverage for spatial data is much less than Wikidata. However, the richness of the information in DBpedia is the highest among all these datasets.

6.3.2 Geocoding Enrichment

The main challenge of incorporating geographic knowledge into geocoding services is to correctly correspond entities from geocoding services with entities from our geographic knowledge bases, namely DBpedia and Wikidata. To tackle this challenge, we take advantage of OSM and GeoNames dataset and harness the power of geospatial SPARQL queries provided by DBpedia and Wikidata endpoints.[4][5] We call this process *Geocoding Enrichment*, illustrated in Figure 6.1. SPARQL, made standard by the World Wide Web Consortium (W3C), is a semantic query language for knowledge bases. SPARQL by itself does not natively have a strong support for geospatial queries. As a geospatial extension for SPARQL, GeoSPARQL defines a vocabulary for representing geospatial data in RDF and supports a variety of complex geospatial query functions. However, none of the DBpedia and Wikidata endpoints supports GeoSPARQL. In the case of DBpedia, we can only use the keyword FILTER to apply spatial queries. In the case of Wikidata, we can use a set of predefined spatial services for the same purposes.

Most geocoding services provide both forward geocoding and reverse geocoding functions. In either scenario, two variables will be determined by the function, namely the toponym or address and coordinates. In order to formalize the operations, we define some of the basic terms we will be using in this research. We use the general term *label* for toponym or address and denote it by L, and we denote *coordinates* by C. The process of geocoding can be formalized as $C = f_g(L)$ (given the label L, the service returns the coordinates C) and the process of reverse geocoding can be formalized as $L = f_r(C)$ (given

[4] `http://dbpedia.org/sparql/`

[5] `https://query.wikidata.org/`

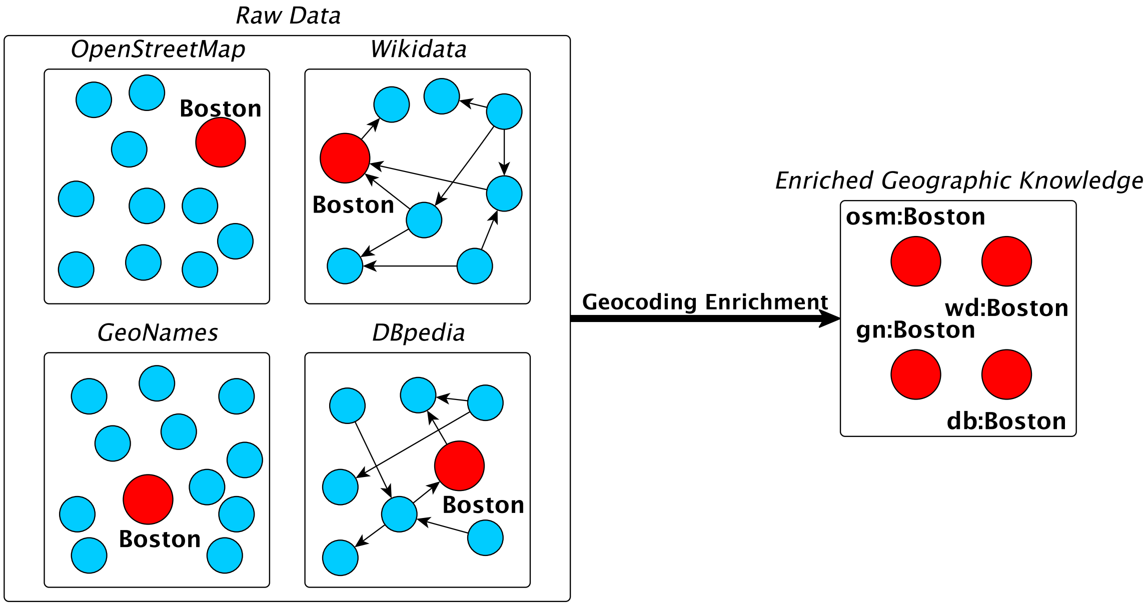


Figure 6.1: Geocoding enrichment illustration. This process combines information from two geographic databases, namely OpenStreetMap and GeoNames, with knowledge from two knowledge graphs, namely Wikidata and DBpedia to help enrich the geocoding results.

the coordinates C, the service returns the label L), where f_g and f_r are different functions of the geocoding service. Based on Definition 2 for knowledge graphs and Definition 3 for geographic knowledge graphs, we define the task of *Geocoding Enrichment*.

Definition 4 (Geocoding Enrichment) *Given a Geographic Knowledge Graph* $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ *and a geocoding service with functions* $F = \{f_g, f_r\}$ *which return the (label, coordinates) pair* (l, c) *where* $l \in L$ *and* $c \in C$*, the task of Geocoding Enrichment is to find the entity set* $\{s | (s, \textit{rdfs:label}, o_1) \in T \wedge (s, \textit{:coordinates}, o_2) \in T \wedge o_1 \leftrightarrow l \wedge o_2 \leftrightarrow c\} \subseteq \mathcal{V}$.

In this formal definition, *rdfs:label* is a common predicate in knowledge graphs that connects an entity with its label and *:coordinates* is a predicate variable that links an entity with its geographic identifications depending on the specific knowledge graphs. For example, in DBpedia *:coordinates* would be *geo:geometry*[6] and in Wikidata it would be

[6] http://www.w3.org/2003/01/geo/wgs84_pos#geometry

wdt:P625[7]. The notation $o_1 \leftrightarrow l$ indicates that o_1 and l are *comparable*. Since in many cases, toponyms of the same spatial entity with difference provenance do not match with each other strictly, e.g. *Los Angeles* in GeoNames and *LA* in DBpedia. Making o_1 and l comparable relaxes the strict string matching conditions. However, due to the nature of geocoding services, a high precision is preferred over a high recall, so in our implementation we still use strict string matching instead of Levenshtein distance or even more sophisticated fuzzy string matching methods. For the spatial footprint part, due to the precision of different datasets and potential errors, we also use the relaxed form of spatial matching $o_2 \leftrightarrow c$. In practice, in order to increase precision, we only use a small radius buffer to search spatial entities.

While it is possible to solely follow the *Geocoding Enrichment* definition for the task, it is not efficient enough and can sometimes be redundant. Figure 6.2 shows the detailed workflow of our implementation. Since OSM are partially linked to Wikipedia, the original source of DBpedia, it is an ideal starting point to access the Geospatial Linked Open Data Cloud. Likewise, GeoNames is partially linked by Wikidata, which makes it our secondary access point. In the cases in which no link can be found by using both OSM and GeoNames, we follow the formal definition of the task and generate SPARQL queries to match entities from geocoding services with entities from the geographic knowledge base directly. If no matching entities are found even via SPARQL queries, the system will reduce the zoom level of the web map and try to match a coarser level toponym in the database. For example, if the label l returned by the geocoding service is in the form "street address, neighborhood, city, county, state, country" and the system fails to match on the state address level, it will try to match the neighborhood level, and so forth. In realization, the formal *Geocoding Enrichment* task is transformed into the following SPARQL query 6.1 in Wikidata. *$wktLiteral$* and *$label$* correspond to two variables c

[7] `https://www.wikidata.org/wiki/Property:P625`

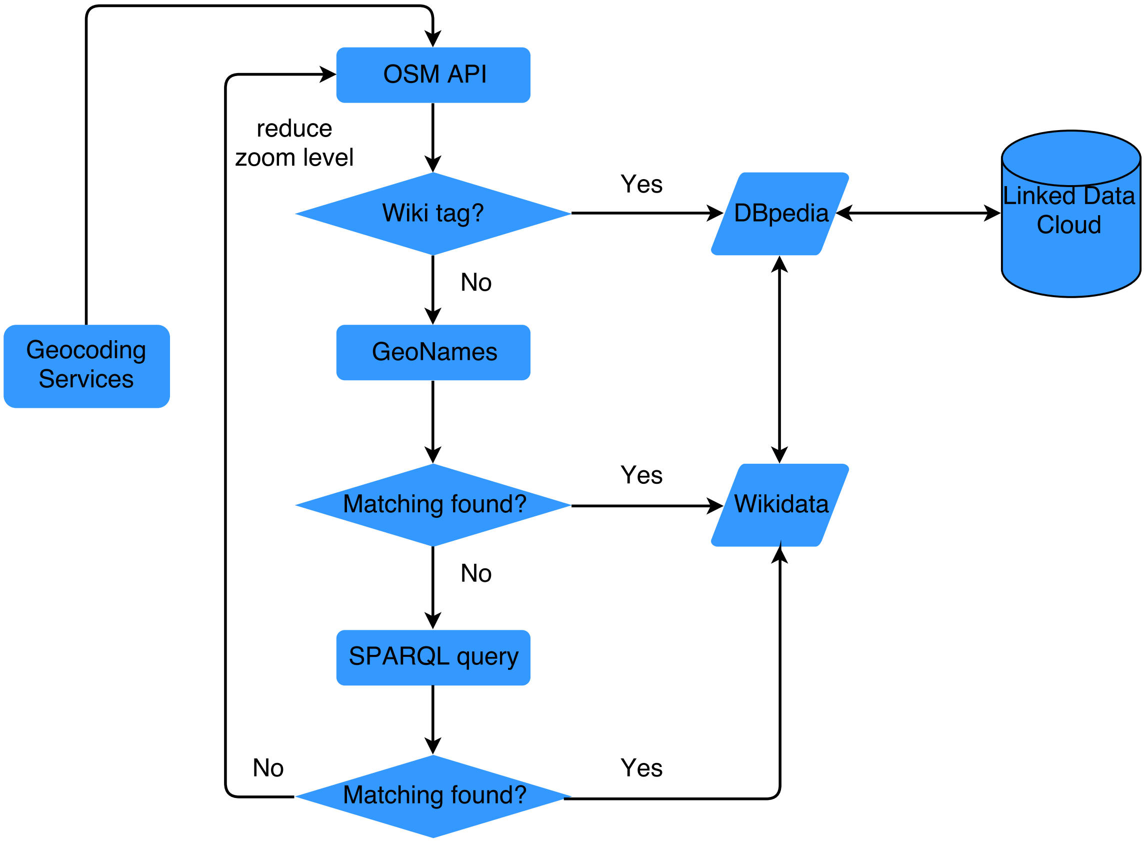


Figure 6.2: Geocoding enrichment workflow

and l respectively.

```
# Geocoding Enrichment
SELECT ?place ?placeLabel ?distance
WHERE {
        # match spatial footprint
        SERVICE wikibase:around {
                ?place wdt:P625 ?location
                # create a buffer
                bd:serviceParam wikibase:center $wktLiteral$ .
                        # radius 2km
                        bd:serviceParam wikibase:radius "2" .
                        bd:serviceParam wikibase:distance ?distance .
        }
        # match toponym
        SERVICE wikibase:label {
                bd:serviceParam wikibase:language "en" .
                ?place rdfs:label ?placeLabel .
        }
        FILTER(lcase(str(?placeLabel)) = $label$)
} ORDER BY ?distance
```

Listing 6.1: An example SPARQL query.

6.3.3 Serverless & Scalable Framework

Geocoding enrichment is an extension to our common geocoding services. Our geocoding enrichment workflow connects geocoding services to geographic knowledge graphs on the fly and it is capable of providing stand-alone services for any geocoding queries. In order to demonstrate and showcase the power of our geocoding enrichment extension, we use it as a server side for a web map that can help users discover geographic knowledge. To deal with the potential challenge of high request volume, we take advantage of the serverless architecture of AWS. The whole geocoding enrichment workflow is converted into and deployed as several microservices in the serverless framework.

A detailed framework outline is illustrated in Figure 6.3. The core part of the framework is the middle part, which includes AWS Lambda and Amazon API Gateway. AWS Lambda hosts functions we implement for the geocoding enrichment so that geographic knowledge could be retrieved as a result. Amazon API Gateway communicates and exchanges requests and responses between our functions and the Geospatial Linked Data Cloud. It also acts as a messenger to listen to the requests sent by the web map interface and respond with the data returned by the AWS Lambda functions.

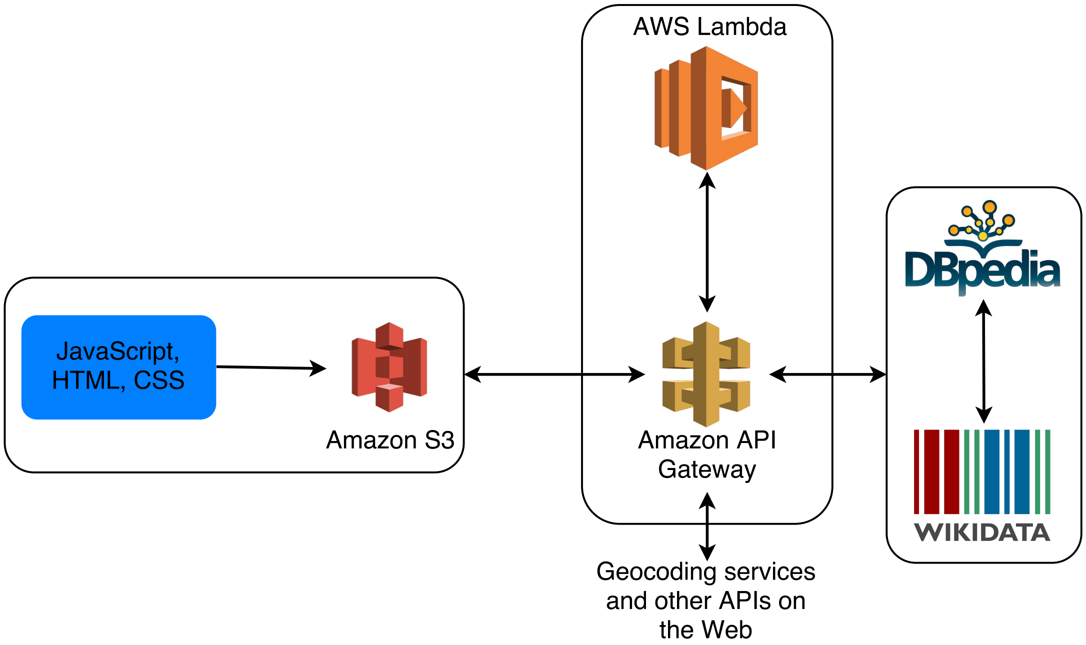


Figure 6.3: A serverless framework for geographic knowledge map

6.3.4 Geographic Knowledge Summarization

As mentioned in Chapter 1 and Chapter 2, in most cases, the amount of information provided by geographic knowledge graphs tends to overwhelm end users and prevent them from analyzing and digesting useful information. The geocoding enrichment process in our system connects the traditional geocoding services with the world of knowledge provide by knowledge graphs and at the same time exposes users to an enormous amount of information, including redundancy, noise, and error associated with it. For example, in DBpedia the geographic entity *dbr:San_Francisco* is directly connected to 248 objects. It is unrealistic to present all these objects to users and expect them to sift through these entities and gain useful knowledge. The whole process would become counterproductive without selecting and summarizing such large amount of information.

Since the main goal of incorporating knowledge graphs into geocoding services is to help users gain more geographic knowledge, we follow the philosophy of maximizing knowledge gain while minimizing redundancy as well as striking the balance between information overload and information deficit. In light of this, we choose to adopt idea in information theory and use entropy to help select and summarize our geographic knowledge returned by the geocoding enrichment process. Entropy is utilized for three reasons. 1) Entropy measures the information contained in an entity as opposed to the portion that is determined or predictable. In this sense, entropy is an ideal proxy for measuring the knowledge gain or loss by adding or removing different entities. 2) Entropy can measure the diversity. For a uniform distribution where each piece of information is equally likely to be obtained (high diversity), the entropy will be high. For a skewed distribution or a distribution where one piece of information is much more likely to be obtained than others (low diversity), the entropy will be low. 3) Entropy effectively sets the bound of the performance of the strongest lossless compression possible. In this sense, entropy is a good measurement for data compression which is closely related to our geographic knowledge summarization task in this context.

Once the geocoding enrichment process is done, the focus is switched to the geographic knowledge graph component to bring more geographic knowledge on the map. The anchoring point on the geographic knowledge graph now becomes the access point towards the world of knowledge. Our prototype system is designed to gather information from geographic knowledge graphs from two perspectives, as shown in Figure 6.4. The first one is the additional attribute information (including a short description, population information for populated places, images if available, etc.) for each geocoded and anchored entity. The second one is grounding connected entities on the map. By retrieving related geographic entities, the spatial knowledge map is able to present the users with linkage information in the geographic space. In addition, our spatial knowledge map also shows spatial patterns for different types of geographic features. More detailed examples will be described in Section 6.4.

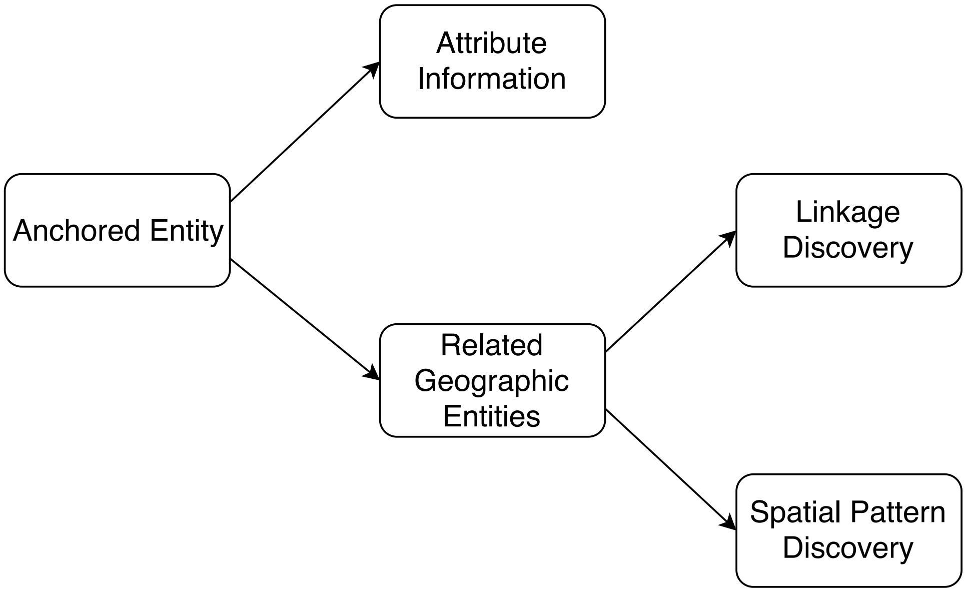


Figure 6.4: Spatial knowledge map components.

For the attribute information part, we use heuristics to select a subset of predicates that is more relevant for facilitating the understanding of geographic entities and remove irrelevant ones, such as *topic's main Wikimedia portal*[8], *topic's main template*[9], and *topic's main category*[10]. For the related geographic entities part, we adopt an entropy-based strategy to summarize the resulting graph. By default, we retrieve all geographic entities directly connected to the anchored entity using Query 6.2. However, this would result in a lot of geographic entities (denoted as ?*o* in the query) on the map. The situation would aggravate if users are interested in higher degrees of connected geographic entities (such as 2nd-degree or 3rd-degree geographic entities) as the result size would grow exponentially.

```
SELECT DISTINCT ?wdPLabel ?o ?oLabel ?coords
WHERE {
```

[8] https://www.wikidata.org/wiki/Property:P1151

[9] https://www.wikidata.org/wiki/Property:P1424

[10] https://www.wikidata.org/wiki/Property:P910

```
        VALUES ?propType {wikibase:WikibaseItem wikibase:Url wikibase:
            ↪ String wikibase:Monolingualtext}
        $anchoredEntity$ ?wdtP ?o .
        ?o wdt:P625 ?coords .
        ?wdP wikibase:directClaim ?wdtP .
        ?wdP wikibase:propertyType ?propType .
        SERVICE wikibase:label {
                bd:serviceParam wikibase:language "en" .
        }
}
```

Listing 6.2: A SPARQL query to retrieve all geographic entities directly connected to $anchoredEntity$.

In order to tackle this issue, we propose a summarization mechanism for the retrieved geographic knowledge. Instead of returning all connected geographic entities, our system caches these entities and retrieves predicate and objects for the connected geographic entities. This is done by using Query 6.3. Suppose an entity e_i has n predicates $p_i^1, p_i^2, ..., p_i^n$. We denote the number of objects for a predicate p_i^j as $freq_{ij}$. The probability for predicate p_i^j is defined as

$$\Pr(p_i^j) = \frac{freq_{ij}}{\sum_{k=1}^{n} freq_{ik}} \tag{6.1}$$

The entropy for the geographic entity e_i is then defined as

$$H(e_i) = -\sum_{k=1}^{n} \Pr(p_i^k) \log(\Pr(p_i^k)) \tag{6.2}$$

```
SELECT ?sub ?wdtP ?oLabel
WHERE {
        VALUES ?sub { $connectedGeographicEntityList$ }
        ?sub ?wdtP ?o .
        ?wdP wikibase:directClaim ?wdtP .
        SERVICE wikibase:label {
                bd:serviceParam wikibase:language "en" .
        }
}
```

Listing 6.3: A SPARQL query used for calculating entropy.

After calculating the entropy for each connected geographic entity e_i, the system ranks them based on entropy values and only retains a certain percentage of entities that have entropy values above the threshold. This percentage parameter is adjustable. A higher percentage value would entail that fewer nodes are removed and a smaller percentage value would result in a graph that is summarized to a greater extent. Figure 6.5 shows the original graph for retrieving all 1st-degree neighboring nodes and part of the 2nd-degree nodes for *Los Angeles*. Figure 6.6 shows the result of retaining only 40% of the nodes from Figure 6.5 based on the entropy of each node calculated using Equation 6.2.

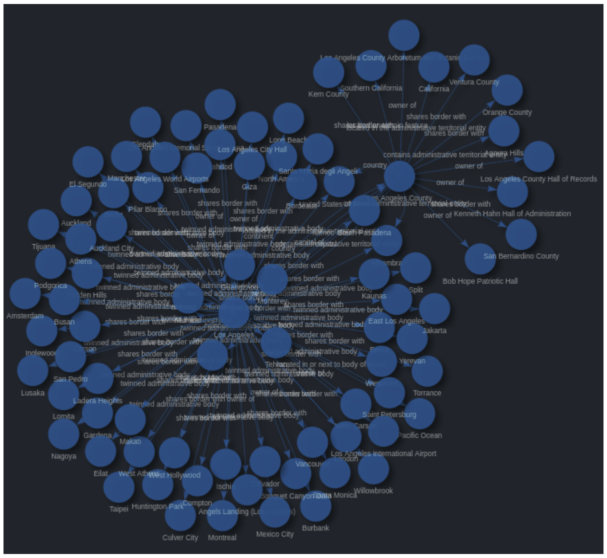


Figure 6.5: The original graph containing all 1st-degree nodes and part of the 2nd-degree nodes for *Los Angeles*.

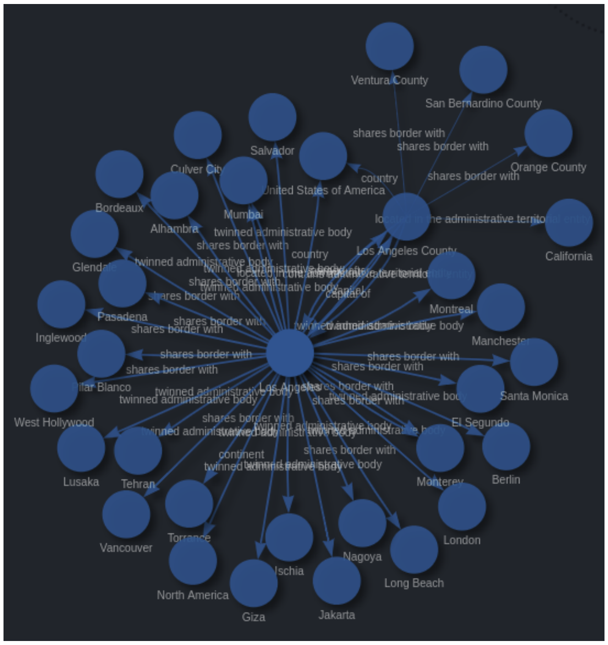


Figure 6.6: Summarized graph for *Los Angeles* for the same graph shown in Figure 6.5.

6.4 Spatial Knowledge Discovery

By introducing geocoding enrichment, we expose ourselves to a sea of possibilities. Geocoding is one of the key components in many common applications such as navigation and web maps. In many cases, geocoding is also the nexus between spatial and non-spatial information. Spatial information such as the geographic identification of an entity can be combined with non-spatial information such as population to generate geographic knowledge. In this section, we showcase some of the potentials of incorporating geographic knowledge into geocoding services with two examples on a web map application. In this application, we show that, by enriching the geocoding result, we are able to not only explore the linkage between distinct spatial entities in non-spatial space but also compute and analyze some basic geospatial statistics such as kernel density estimation (as shown in Figure 6.4).

6.4.1 Linkage Discovery

Although the primary building blocks in a knowledge graph are RDF triples, a knowledge graph as a whole can be viewed as a gigantic directed graph with each node being an entity and each edge being a property connecting a pair of entities. Exploring the network structures of the graph or even simply the links between different entities can reveal some hidden information. For instance, Janowicz et al. [86] coined the term *spatial degree* by counting the number of property paths between a geographic identification and an entity of an arbitrary type. Using this concept, they were able to identify some of the potential modeling errors in geographic knowledge graphs. In contrast, map distances are Euclidean distances. Combining graph distance with map distance is an ideal way to explore and discover spatial entity linkage and association.

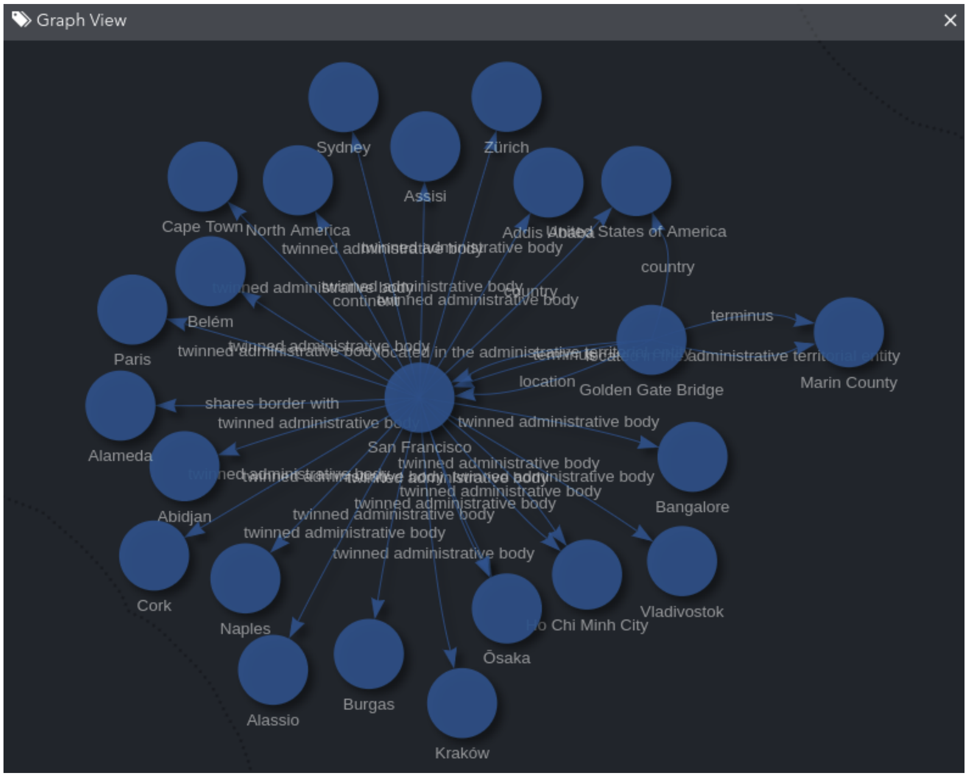


Figure 6.7: 2nd-degree places of Golden Gate Bridge (graph view).

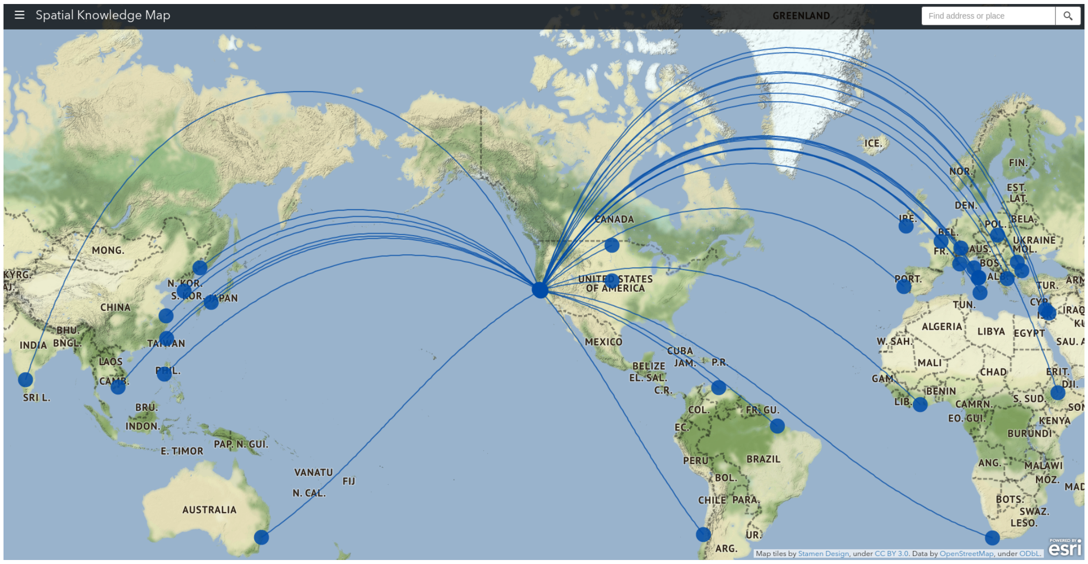
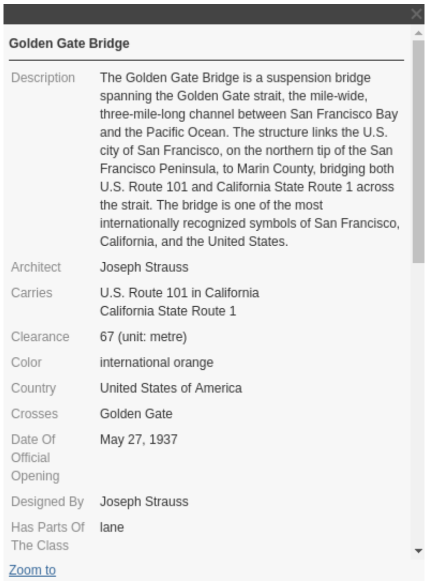


Figure 6.8: 2nd-degree places of Golden Gate Bridge (map view).

However, naively plotting the entire knowledge graph on a map is not very useful because there are too many spatial entities and it will create an information overflow. A more intuitive approach is to adopt the follow-your-nose pattern of Linked Data. Users can either use the forward geocoding or reverse geocoding function to locate a place on the map. Since we are using the enriched geocoding service, a corresponding spatial entity from the knowledge graph will be automatically retrieved on the fly. Meanwhile, we avail ourselves of the power of SPARQL to essentially perform a breadth-first search of the knowledge graph in order to find all the directly connected spatial entities with respect to the original query point. Then by following any of these 1st-degree places (with respect to the original query point), users are able to discover more connections. For example, when a user query about Golden Gate Bridge, all the 1st-degree places including Marin County, Golden Gate, San Francisco and United States will be shown on the map. If the user is interested in the connected place San Francisco, he can further retrieve all 1st-degree places with respect to San Francisco (2nd-degree places with respect to Golden Gate Bridge).In other words, the user can follow his nose and explore the network structure of these places while being able to compare it with the Euclidean counterpart on the map at the same time. This is because we have both a graph view (see Figure 6.7) and a map view (see Figure 6.8). Although the number of paths the user can explore between different places is infinite, it has been shown that the average graph distance between entities is less than or equal to 5 [197]. One observation from the graph view is that, for the 2nd-degree graph of Golden Gate Bridge, we can tell that San Francisco has already started to create its own *community* and it has way more connections than the original place — Golden Gate Bridge. Likewise, from the map view, we can notice that San Francisco is connected to many spatially distant places and the links between Golden Gate Bridge and its 1st-degree places have been shadowed by those of San Francisco's because they are too close to each other spatially. In addition to the linkage discovery, for each individual place, the user can also view its attributes retrieved from the Linked Data Cloud (see Figure 6.9).



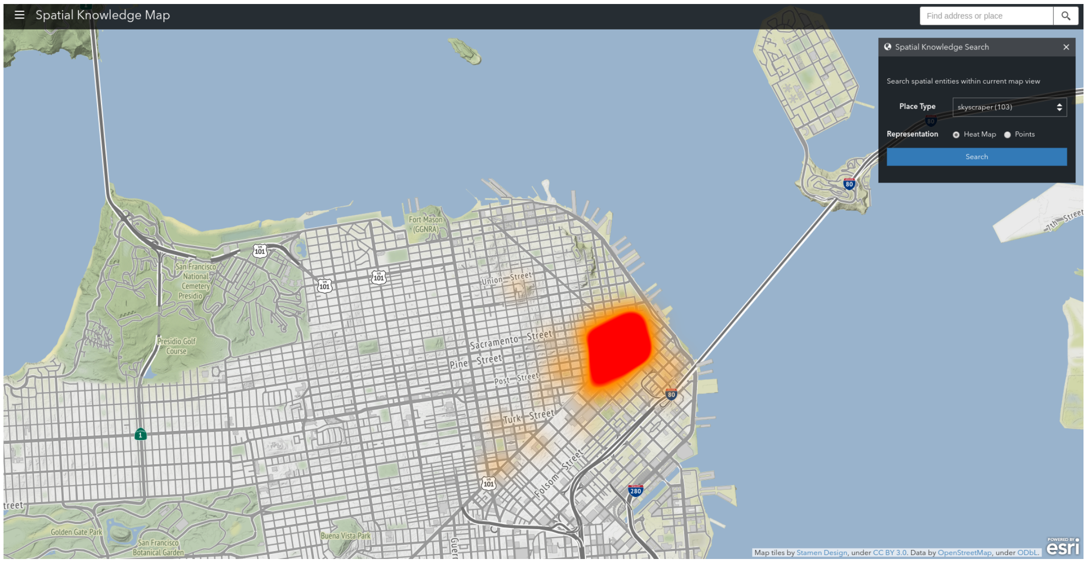
Golden Gate Bridge

Description	The Golden Gate Bridge is a suspension bridge spanning the Golden Gate strait, the mile-wide, three-mile-long channel between San Francisco Bay and the Pacific Ocean. The structure links the U.S. city of San Francisco, on the northern tip of the San Francisco Peninsula, to Marin County, bridging both U.S. Route 101 and California State Route 1 across the strait. The bridge is one of the most internationally recognized symbols of San Francisco, California, and the United States.
Architect	Joseph Strauss
Carries	U.S. Route 101 in California California State Route 1
Clearance	67 (unit: metre)
Color	international orange
Country	United States of America
Crosses	Golden Gate
Date Of Official Opening	May 27, 1937
Designed By	Joseph Strauss
Has Parts Of The Class	lane

Figure 6.9: Attribute information from the knowledge graph.

6.4.2 Spatial Pattern Discovery

Another use case of the geocoding enrichment extension is that people can retrieve essential information to derived spatial patterns for an area of interest. We take advantage of one of the most well-represented properties in the Linked Data Cloud — *rdf:type* to retrieve information about place types. *rdf:type* in our case is a general group of properties that can help us identify the place type associations and each knowledge graph has its own version, for instance, Wikidata uses *wd:P31* (instance of). Several research studies have demonstrated the validity of using place types to discover spatial patterns and spatial signatures [198, 65].



Search spatial entities within current map view

Place Type: skyscraper (103)

Representation: ● Heat Map ○ Points

Search

Figure 6.10: Heatmap showing the distribution of skyscrapers in San Francisco

We implement a minimal spatial pattern discovery function using the place type information retrieved from Wikidata. To activate this function, users can choose *Spatial Knowledge Search* from the drop down menu. A control panel will then appear. Users can zoom in, zoom out or pan to desired spatial extent. After clicking on the *search* button, a list of place types will be shown based on the query result of the geocoding enrichment extension. For each of the place type, users can either view the kernel density map (heatmap) or directly visualize them as points. Figure 6.10 shows an example of the spatial pattern. It shows that there are 103 skyscrapers in downtown San Francisco in the Wikidata knowledge base. From this kernel density map we can tell that most skyscrapers are located in the Financial District and some of them are in nearby neighborhoods such as Rincon Hill and South of Market. Users will also be able to see the attributes of individual place as shown in Figure 6.9.

6.5 Conclusions

In this research, we propose to incorporate geographic knowledge into the current geocoding services using a systematic workflow. We develop the workflow by first formalizing our *Geocoding Enrichment* task and then we relax some of our constraints for pragmatic purposes. In order to scale our geocoding enrichment extension, we take advantage of AWS serverless framework and deploy it as a web service so that any applications with the same interface can access our service. In addition, we adopt an entropy-based approach to help summarize the geographic knowledge associated with the anchored entity. Finally, to demonstrate the power of geographic knowledge enriched geocoding service, we build an interactive web map application with two major functionalities, namely linkage discovery and spatial pattern discovery. Both functions have shown that the enriched geocoding service can help users gain much more geographic knowledge with a few simple operations.

For future work, in terms of the current geographic knowledge map, we can modify the way the linkage is presented by considering the property types as well. In this way, we are able to distinguish between topological properties and non-topological properties. For the spatial pattern discovery part, semantic signatures can be incorporated to help users identify spatially functional regions. Another direction is to apply the enriched geocoding service to other scenarios to discover more potentials.

Chapter 7

Conclusions

In this dissertation, the topic of geographic knowledge graph summarization has been dissected into different but related parts. This chapter concludes the dissertation by synthesizing these different parts and provide insights about the theoretical as well as the practical contributions. In addition, several limitations are listed as directions for improvement. Areas for further investigation are also discussed in the end.

7.1 Summary and Discussions

This dissertation explores the question of leveraging both top-down knowledge engineering and bottom-up knowledge learning approaches to help summarize geographic knowledge graphs under the context of the geospatial semantics paradigm. Such a paradigm shift is propelled by three major demands in the course of the development of GIScience, namely the interoperability demand, the accessibility demand, and the conceptualization demand. While bringing geospatial semantics (geographic knowledge graphs in particular) to the GIScience community helps dealing with these demands, it also creates new challenges. It creates diversity from two perspective, namely the diversity in terms of cross-domain interconnection for various knowledge graphs and the diversity in terms of heterogeneous types of information for each entity in the same knowledge graph. The sheer amount of information (the number of entities, predicates, and datasets) provided by geographic knowledge graphs are overwhelming.

The idea of *less is more* is the central philosophy of geographic knowledge graph summarization. From a human-centric perspective, the overwhelmingly large amount of information imposes burden on our cognitive load, hinders our cognitive information-processing capacities, and results in lags and retrieval errors. Too much information from geographic knowledge graphs also leads to too many choices for us to decide in terms of selecting and filtering useful information. This creates the undesired paradox of choice and demotivates users from using geographic knowledge graphs in the first place. From a data-centric perspective, geographic knowledge graph summarization helps reduce the data volume, speeds up queries, supports large scale interactive analysis, and facilitates the elimination of noise [21].

While there are different ways to tackle the problem, this dissertation aims to treat the summarization task from three aspects considering the composition of geographic knowledge graphs, namely the hierarchical components, the multimedia leaf node components, and the general relation and entity components. The advantage of such a strategy is that by decomposing geographic knowledge graphs into different parts the major demands that gave rise to geospatial semantics in the first place are taken into consideration. For instance, by isolating the hierarchical structure of the place type ontology in geographic knowledge graphs and learning their latent representations, our approach not only preserves the original conceptualization in the ontology but also provides a robust embedding mechanism that is accessible and interoperable across different machine learning models. By focusing on the images and their labels for multimedia leaf nodes in geographic knowledge graphs, our approach promotes the accessibility of geographic knowledge graphs among common users as people tend to utilize multimedia to understand data.

Specifically, Chapter 3 focuses on place types in the ontology hierarchical structure and presents a latent representation learning method to embed these place types. This latent representation learning strategy takes advantage of geospatial contextual information by considering local as well as global contexts. Instead of using language as proxies, the proposed information-theoretic and distance lagged method direct uses geographic entities to model place types which results in an improvement in performance compared with existing models using a hierarchy-based evaluation scheme, a binary-based human judgment evaluation scheme, and a ranking-based human judgment evaluation scheme. These place type embeddings can be used for calculating semantic similarities for different place types and eventually be used for selecting and summarizing type information in geographic knowledge graphs. Chapter 4 focuses on leaf image nodes and presents an image classification model that combines visual signals with spatial contextual signals to help improve classification results for different place types. It explores the possibility of incorporating different types of spatial contexts, namely spatial relatedness, spatial colocation, and spatial sequence patterns. The evaluation shows that spatial sequence patterns modeled as Bayesian priors are able to substantially improve the classification result. Such a model can be use to help label unknown leaf image nodes in the geographic knowledge graph and facilitate the selection and summarization process while preserving the accessibility. Chapter 5 proposes a spatially-explicit reinforcement learning model for geographic knowledge graph summarization. Instead of focusing on a particular component, this model provides a generic approach. In order to tackle the challenge of the inherent geospatial semantics, the model adopts the idea of geospatial inductive bias and introduces a special spatial action for the reinforcement learning agent in an attempt to provide a robust algorithm. Chapter 6 presents a spatial knowledge map interface that illustrate the functionality and effectiveness of geographic knowledge graph summarization in the context of enriching geocoding services. The interface is empowered by a serverless and scalable framework using AWS. It can be used to facilitate the discovery of spatial knowledge such as linkage discovery and spatial pattern discovery.

In summary, this dissertation provides the background and motivation for summarizing geograhic knowledge graphs, explains foundational concepts, decomposes the summarization task into separate but related parts, answers these research questions separately, and illustrates the usefulness of geographic knowledge graph summarization via a spatial knowledge map interface.

7.2 Research Contributions

The main contribution of this dissertation is the hybrid approach of summarizing geographic knowledge graphs by decomposing them into related components. The reason for the decomposition and the resulting hybrid approach is that three major demands (interoperability demand, accessibility demand, and conceptualization demand) in GIScience can be preserved. The dissertation points out the fact that, while graph summarization and knowledge graph summarization have been intriguing research topics for the data mining community and the semantic web community, research on geographic knowledge graphs is still scarce. Although existing methods and algorithms for graph summarization can be adopted, these models are not capable of explicitly taking into account the inherent geospatial semantics embedded in geographic knowledge graphs. To elaborate on this point, this dissertation summarizes foundational concepts and ideas behind various research in GIScience and argues that such ideas are important for the geographic knowledge graph summarization task as well.

In the following, we discuss the specific theoretical contributions and practical implications of this research.

7.2.1 Theoretical Contributions

This dissertation has made a number of theoretical contributions to the GIScience field. In this subsection, we summarize them as follows.

Major demands in the current GIScience development. Although listed as the background for the research explored in this dissertation, three major demands have been the driving force for embracing the geospatial semantics paradigm. These three demands, namely the interoperability demand, the accessibility demand, and the conceptualization demand, have been entangled with a lot of research questions in GIScience in recent years. By explaining them as means of improving efficiency in processing the large amount of heterogeneous geographic data, providing more accessible interfaces for the general audience, and establishing a better conceptualization model to mitigate the inherent vagueness in geographic phenomenon, we hope this dissertation can stimulate further studies in understanding the implications of the geospatial semantics paradigm.

The idea of geospatial inductive bias. By formalizing the concept of geospatial context, this dissertation brings up the idea of geospatial inductive bias based upon the idea of inductive bias in machine learning models. The hidden patterns in geospatial context are frequently exploited by a variety of models in geography (such as geographically weighted regression), but such an idea was not formalized as a general approach towards developing more robust and suitable models for geographic data. This dissertation provides a formalization of such an idea and proposes that geospatial inductive bias should be adopted in order to develop generalizable models for geographic data.

A hybrid approach for geographic knowledge graph summarization. This dissertation provides the definition of knowledge graphs by considering its duality in conceptual representation and implementation. The definition of geographic knowledge graphs is subsequently clarified by extending existing definition of knowledge graphs. While there are many ways to summarize geographic knowledge graphs, in the dissertation we present a hybrid approach based on the decomposition of the graph. Such an approach allows us to target specific areas, such as the place type ontology, the multimedia leaf nodes, and the generic properties and entities separately.

Spatially-explicit models. While many data mining and machine learning models provide a general strategy to solving problems, naively applying these models in the GIScience domain is not ideal because of the heterogeneous nature and the hidden semantics of geographic data. The idea of *spatial is special* demands that models dealing with geographic data should explicitly consider spatial patterns by exploring the geospatial inductive bias. Chapter 3, Chapter 4, and Chapter 5 have shown that by developing spatially-explicit models for geographic data, the performance can be substantially improved for a variety of machine learning models, including representation learning, classification, and reinforcement learning.

7.2.2 Practical Implications

This research also has practical values and can be applied in pragmatic settings. In the following, we outline several applications that could benefit from this research.

Semantic search for locations. While Chapter 3 is primarily focusing on place types, the proposed methods can be used for a variety of geographic entity or entity types. The latent representation learning approach provides low dimensional vector representations for geographic entities/entity types which can be used as inputs for the search engines. Because these embeddings contain semantic relatedness and similarity, search engines empowered by these embeddings are able to understand the meaning of these entities/-types in the geospatial context. As a result, these search engines can help provide a semantic search (as opposed to string-based matching) for locations.

Location recommendation. The embeddings for different place types can also be used for location recommendation applications. Because these embeddings encode information about the spatial distribution as well as popularity (check-in counts) among users, they can imply the preference of different users under different spatial-temporal contexts. By combining trajectory information of a user, these embeddings can be used as inputs to help identify hidden patterns and predict future locations. Such a model is essential for location recommendation systems.

Image classification systems. As explained in Chapter 4, the bias in training data has imposed a lot of challenges for correctly classifying images. The proposed idea incorporate spatial contextual signals into image classification to help improve the performance. This idea is particularly useful for classifying images that are less common and can potentially help improve sample efficiency. This has important implications for practical image classification systems as efficiency and accuracy are two major considerations.

Geographic knowledge graph visualization and exploration systems. The ultimate goal of this dissertation is to find ways to help better summarize geographic knowledge graphs. In Chapter 5, a generic approach has been proposed and sample summary graphs have been provided. In practice, this approach can be integrated in geographic knowledge graph visualization and exploration interfaces as demonstrated in Chapter 6. Such a system is able to benefit from the concise representation and insightful digest of the geographic knowledge graph and provide an improved user experience.

7.3 Limitations and Future Work

In the following, we discuss limitations in this research as well as potential areas that could be integrated in the future work.

Spatial modeling in high dimensions. While real-world geographic entities are usually in 3D or 4D (if time is considered an extra dimension) space, the spatially-explicit models developed in this dissertation are mostly 1D or 2D. For instance, in Chapter 3, we calculate the neighborhood distribution in a 2D space. In Chapter 4, the neighborhood is collapsed into a 1D sequence in the proposed model. While these strategies are a result of the trade-off between model complexity and performance, it is important to note that by collapsing the dimensions of geographic entities the loss of information is inevitable. Future improvements could be made by considering model architectures that are specifically designed for high dimensional structures.

Large scale study of the summarization task. Because of the lack of standard dataset for research studies in geographic knowledge graph summarization, we take the initiative to collect a subgraph from DBpedia to form our geographic knowledge graph in this research. While this dataset is useful for establishing baselines and comparison in model performance, a large scale study of the summarization task is still needed. Such a large scale study would require a substantially larger geographic knowledge graph dataset. While the same method could be potentially applied, the noise and error in the real-world dataset might pose additional challenges. Additional improvements are also needed to provide an efficient strategy for summarizing large scale geographic knowledge graphs. Future work can potentially focus on improving the space and time complexity of existing models and incorporate mechanism to handle noise to make the model more robust.

Data type properties and literals in geographic knowledge graphs. Although this dissertation specifically considers place type information as well as multimedia leaf nodes, it does not consider data type properties and literals which contain a lot of information in geographic knowledge graphs. In Chapter 5, the knowledge graph embedding model is not able to handle data type properties and literals. These literals, although are not typically encoded in knowledge graph embedding models, contain important semantics in the context of geographic knowledge graphs, such as the population information of a city and the elevation of a mountain, etc. In future work, data type properties and literals can be incorporated in the embedding learning process and should be considered in the summarization process.

Bibliography

[1] W. Kuhn, *Geospatial semantics: why, of what, and how?*, in *Journal on data semantics III*, pp. 1–24. Springer, 2005.

[2] K. Janowicz, S. Scheider, T. Pehle, and G. Hart, *Geospatial semantics and linked spatiotemporal data–past, present, and future*, *Semantic Web* **3** (2012), no. 4 321–332.

[3] D. M. Mark, *Geographic information science: Defining the field*, *Foundations of geographic information science* **1** (2003) 3–18.

[4] Y. Ju, B. Adams, K. Janowicz, Y. Hu, B. Yan, and G. McKenzie, *Things and strings: improving place name disambiguation from short texts by combining entity co-occurrence with topic modeling*, in *European Knowledge Acquisition Workshop*, pp. 353–367, Springer, 2016.

[5] Z. Kemp, L. Tan, and J. Whalley, *Interoperability for geospatial analysis: a semantics and ontology-based approach*, in *Proceedings of the eighteenth conference on Australasian database-Volume 63*, pp. 83–92, Australian Computer Society, Inc., 2007.

[6] M. Wang, R. Wang, J. Liu, Y. Chen, L. Zhang, and G. Qi, *Towards empty answers in sparql: Approximating querying with rdf embedding*, in *International Semantic Web Conference*, pp. 513–529, Springer, 2018.

[7] W. Chen, E. Fosler-Lussier, N. Xiao, S. Raje, R. Ramnath, and D. Sui, *A synergistic framework for geographic question answering*, in *Semantic Computing (ICSC), 2013 IEEE Seventh International Conference on*, pp. 94–99, IEEE, 2013.

[8] G. Mai, B. Yan, K. Janowicz, and R. Zhu, *Relaxing unanswerable geographic questions using a spatially explicit knowledge graph embedding model*, in *The Annual International Conference on Geographic Information Science*, 2019.

[9] D. R. Montello, M. F. Goodchild, J. Gottsegen, and P. Fohl, *Where's downtown?: Behavioral methods for determining referents of vague spatial queries*, *Spatial Cognition & Computation* **3** (2003), no. 2-3 185–204.

[10] S. Gao, K. Janowicz, D. R. Montello, Y. Hu, J.-A. Yang, G. McKenzie, Y. Ju, L. Gong, B. Adams, and B. Yan, *A data-synthesis-driven method for detecting and extracting vague cognitive regions*, *International Journal of Geographical Information Science* **31** (2017), no. 6 1245–1271.

[11] B. Smith and D. M. Mark, *Do mountains exist? towards an ontology of landforms*, *Environment and Planning B: Planning and Design* **30** (2003), no. 3 411–427.

[12] B. Bennett, D. Mallenby, and A. Third, *An ontology for grounding vague geographic terms.*, in *FOIS*, vol. 183, pp. 280–293, 2008.

[13] Y. Hu, K. Janowicz, D. Carral, S. Scheider, W. Kuhn, G. Berg-Cross, P. Hitzler, M. Dean, and D. Kolas, *A geo-ontology design pattern for semantic trajectories*, in *International Conference on Spatial Information Theory*, pp. 438–456, Springer, 2013.

[14] P. Grenon and B. Smith, *Snap and span: Towards dynamic spatial ontology*, *Spatial cognition and computation* **4** (2004), no. 1 69–104.

[15] M. J. Egenhofer, *Toward the semantic geospatial web*, in *Proceedings of the 10th ACM international symposium on Advances in geographic information systems*, pp. 1–4, ACM, 2002.

[16] K. Janowicz, *Observation-driven geo-ontology engineering*, *Transactions in GIS* **16** (2012), no. 3 351–374.

[17] B. Yan, K. Janowicz, and Y. Hu, *A data-driven approach for detecting and quantifying modeling biases in geo-ontologies by using a discrepancy index*, in *International Conference on GIScience Short Paper Proceedings*, vol. 1, 2016.

[18] M. Ramscar, P. Hendrix, C. Shaoul, P. Milin, and H. Baayen, *The myth of cognitive decline: Non-linear dynamics of lifelong learning*, *Topics in cognitive science* **6** (2014), no. 1 5–42.

[19] B. Schwartz, *The paradox of choice: Why more is less*, vol. 6. HarperCollins New York, 2004.

[20] S. S. Iyengar and M. R. Lepper, *When choice is demotivating: Can one desire too much of a good thing?*, *Journal of personality and social psychology* **79** (2000), no. 6 995.

[21] Y. Liu, T. Safavi, A. Dighe, and D. Koutra, *Graph summarization methods and applications: A survey*, *ACM Computing Surveys (CSUR)* **51** (2018), no. 3 62.

[22] G. Cheng, T. Tran, and Y. Qu, *Relin: relatedness and informativeness-based centrality for entity summarization*, in *International Semantic Web Conference*, pp. 114–129, Springer, 2011.

[23] K. Gunaratna, K. Thirunarayan, A. Sheth, and G. Cheng, *Gleaning types for literals in rdf triples with application to entity summarization*, in *International Semantic Web Conference*, pp. 85–100, Springer, 2016.

[24] F. Hasibi, K. Balog, and S. E. Bratsberg, *Dynamic factual summaries for entity cards*, in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 773–782, ACM, 2017.

[25] M. Sydow, M. Pikuła, and R. Schenkel, *The notion of diversity in graphical entity summarisation on semantic knowledge graphs*, *Journal of Intelligent Information Systems* **41** (2013), no. 2 109–149.

[26] K. Gunaratna, K. Thirunarayan, and A. P. Sheth, *Faces: Diversity-aware entity summarization using incremental hierarchical conceptual clustering.*, 2015.

[27] G. Cheng, D. Xu, and Y. Qu, *C3d+ p: A summarization method for interactive entity resolution*, *Web Semantics: Science, Services and Agents on the World Wide Web* **35** (2015) 203–213.

[28] D. Xu, G. Cheng, and Y. Qu, *Facilitating human intervention in coreference resolution with comparative entity summaries*, in *European Semantic Web Conference*, pp. 535–549, Springer, 2014.

[29] D. Xu, G. Cheng, and Y. Qu, *Preferences in wikipedia abstracts: Empirical findings and implications for automatic entity summarization*, *Information Processing & Management* **50** (2014), no. 2 284–296.

[30] A. G. Nuzzolese, V. Presutti, A. Gangemi, S. Peroni, and P. Ciancarini, *Aemoo: Linked data exploration based on knowledge patterns*, *Semantic Web* **8** (2017), no. 1 87–112.

[31] L. Anselin, *What is special about spatial data? alternative perspectives on spatial data analysis (89-4)*, .

[32] L. Zhu, M. Ghasemi-Gol, P. Szekely, A. Galstyan, and C. A. Knoblock, *Unsupervised entity resolution on multi-type graphs*, in *International Semantic Web Conference*, pp. 649–667, Springer, 2016.

[33] H. Bast, B. Buchhold, and E. Haussmann, *Relevance scores for triples from type-like relations*, in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 243–252, ACM, 2015.

[34] T. Franz, A. Schultz, S. Sizov, and S. Staab, *Triplerank: Ranking semantic web data by tensor decomposition*, in *International semantic web conference*, pp. 213–228, Springer, 2009.

[35] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner, *et. al.*, *Relational inductive biases, deep learning, and graph networks*, *arXiv preprint arXiv:1806.01261* (2018).

[36] K. Fukushima, *Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position*, *Biological cybernetics* **36** (1980), no. 4 193–202.

[37] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, *Backpropagation applied to handwritten zip code recognition*, *Neural computation* **1** (1989), no. 4 541–551.

[38] J. L. Elman, *Finding structure in time*, *Cognitive science* **14** (1990), no. 2 179–211.

[39] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. R. Salakhutdinov, and A. J. Smola, *Deep sets*, in *Advances in neural information processing systems*, pp. 3391–3401, 2017.

[40] W. Humboldt, *On language: On the diversity of human language construction and its influence on the mental development of the human species*, .

[41] P. Stock and M. Cisse, *Convnets and imagenet beyond accuracy: Explanations, bias detection, adversarial examples and model criticism*, *arXiv preprint arXiv:1711.11443* (2017).

[42] B. Kim, R. Khanna, and O. O. Koyejo, *Examples are not enough, learn to criticize! criticism for interpretability*, in *Advances in Neural Information Processing Systems*, pp. 2280–2288, 2016.

[43] F. Å. Nielsen, *Scientific citations in wikipedia*, *arXiv preprint arXiv:0705.2106* (2007).

[44] S. Greenstein and F. Zhu, *Is wikipedia biased?*, *American Economic Review* **102** (2012), no. 3 343–48.

[45] B. Yan, K. Janowicz, G. Mai, and R. Zhu, *A spatially-explicit reinforcement learning model for geographic knowledge graph summarization*, *Transactions in GIS* (2019).

[46] A. Cocos and C. Callison-Burch, *The language of place: Semantic value from geospatial context*, in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, vol. 2, pp. 99–104, 2017.

[47] S. Feng, G. Cong, B. An, and Y. M. Chee, *Poi2vec: Geographical latent representation for predicting future visitors*, in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[48] B. Yan, K. Janowicz, G. Mai, and S. Gao, *From itdl to place2vec: Reasoning about place type similarity and relatedness by learning embeddings from augmented spatial contexts*, in *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, SIGSPATIAL '17, (New York, NY, USA), pp. 35:1–35:10, ACM, 2017.

[49] B. Yan, K. Janowicz, G. Mai, and R. Zhu, *xnet+sc: Classifying places based on images by incorporating spatial contexts*, in *10th International Conference on Geographic Information Science (GIScience 2018)*, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

[50] W. R. Tobler, *A computer movie simulating urban growth in the detroit region*, *Economic geography* **46** (1970), no. sup1 234–240.

[51] M. L. Stein, *Interpolation of spatial data: some theory for kriging.* Springer Science & Business Media, 2012.

[52] J. K. Ord, *Spatial autocorrelation: A statisticians reflections*, in *Perspectives on Spatial Data Analysis*, pp. 165–180. Springer, 2010.

[53] C. Brunsdon, A. S. Fotheringham, and M. E. Charlton, *Geographically weighted regression: a method for exploring spatial nonstationarity*, *Geographical analysis* **28** (1996), no. 4 281–298.

[54] G. Mariethoz, P. Renard, and J. Straubhaar, *The direct sampling method to perform multiple-point geostatistical simulations*, *Water Resources Research* **46** (2010), no. 11.

[55] S. Openshow, *A million or so correlation coefficients, three experiments on the modifiable areal unit problem*, *Statistical applications in the spatial science* (1979) 127–144.

[56] D. Wong, *The modifiable areal unit problem (maup)*, *The SAGE handbook of spatial analysis* **105** (2009) 23.

[57] G. McKenzie and K. Janowicz, *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** (2015) 1–13.

[58] P. A. Moran, *Notes on continuous stochastic phenomena*, *Biometrika* **37** (1950), no. 1/2 17–23.

[59] B. D. Ripley, *The second-order analysis of stationary point processes*, *Journal of applied probability* **13** (1976), no. 2 255–266.

[60] C. Mülligann, K. Janowicz, M. Ye, and W.-C. Lee, *Analyzing the spatial-semantic interaction of points of interest in volunteered geographic information*, in *International Conference on Spatial Information Theory*, pp. 350–370, Springer, 2011.

[61] T. Berg, J. Liu, S. Woo Lee, M. L. Alexander, D. W. Jacobs, and P. N. Belhumeur, *Birdsnap: Large-scale fine-grained visual categorization of birds*, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2011–2018, 2014.

[62] K. Tang, M. Paluri, L. Fei-Fei, R. Fergus, and L. Bourdev, *Improving image classification with location context*, in *Proceedings of the IEEE international conference on computer vision*, pp. 1008–1016, 2015.

[63] B.-H. Lee, H.-N. Kim, J.-G. Jung, and G.-S. Jo, *Location-based service with context data for a restaurant recommendation*, in *International Conference on Database and Expert Systems Applications*, pp. 430–438, Springer, 2006.

[64] G. F. Rengert, A. R. Piquero, and P. R. Jones, *Distance decay reexamined*, *Criminology* **37** (1999), no. 2 427–446.

[65] R. Zhu, Y. Hu, K. Janowicz, and G. McKenzie, *Spatial signatures for geographic feature types: Examining gazetteer ontologies using spatial statistics*, *Transactions in GIS* **20** (2016), no. 3 333–355.

[66] M. Schedl, A. Vall, and K. Farrahi, *User geospatial context for music recommendation in microblogs*, in *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pp. 987–990, ACM, 2014.

[67] T. M. Mitchell, *The need for biases in learning generalizations*. Department of Computer Science, Laboratory for Computer Science Research, Rutgers Univ., 1980.

[68] R. S. Michalski, *A theory and methodology of inductive learning*, in *Machine learning*, pp. 83–134. Springer, 1983.

[69] D. Haussler, *Quantifying inductive bias: Ai learning algorithms and valiant's learning framework*, *Artificial intelligence* **36** (1988), no. 2 177–221.

[70] O. Lassila and R. R. Swick, *Resource description framework (rdf) model and syntax specification*, .

[71] T. Berners-Lee, J. Hendler, O. Lassila, *et. al.*, *The semantic web*, *Scientific american* **284** (2001), no. 5 28–37.

[72] G. Klyne and J. J. Carroll, *Resource description framework (rdf): Concepts and abstract syntax*, .

[73] C. Bizer, T. Heath, and T. Berners-Lee, *Linked data: The story so far*, in *Semantic services, interoperability and web applications: emerging concepts*, pp. 205–227. IGI Global, 2011.

[74] B. Parsia, N. Matentzoglu, R. S. Gonçalves, B. Glimm, and A. Steigmiller, *The owl reasoner evaluation (ore) 2015 resources*, in *International Semantic Web Conference*, pp. 159–167, Springer, 2016.

[75] J. Pérez, M. Arenas, and C. Gutierrez, *Semantics and complexity of sparql*, *ACM Transactions on Database Systems (TODS)* **34** (2009), no. 3 16.

[76] R. Angles and C. Gutierrez, *The multiset semantics of sparql patterns*, in *International semantic web conference*, pp. 20–36, Springer, 2016.

[77] R. Verborgh, M. Vander Sande, O. Hartig, J. Van Herwegen, L. De Vocht, B. De Meester, G. Haesendonck, and P. Colpaert, *Triple pattern fragments: a low-cost knowledge graph interface for the web*, *Journal of Web Semantics* **37** (2016) 184–206.

[78] O. Hartig and C. Buil-Aranda, *Bindings-restricted triple pattern fragments*, in *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, pp. 762–779, Springer, 2016.

[79] O. Hartig, I. Letter, and J. Pérez, *A formal framework for comparing linked data fragments*, in *International semantic web conference*, pp. 364–382, Springer, 2017.

[80] R. Usbeck, A.-C. N. Ngomo, M. Röder, D. Gerber, S. A. Coelho, S. Auer, and A. Both, *Agdistis-graph-based disambiguation of named entities using linked data*, in *International semantic web conference*, pp. 457–471, Springer, 2014.

[81] J. Pujara, H. Miao, L. Getoor, and W. Cohen, *Knowledge graph identification*, in *International Semantic Web Conference*, pp. 542–557, Springer, 2013.

[82] P. Mika, *Ontologies are us: A unified model of social networks and semantics*, *Web semantics: science, services and agents on the World Wide Web* **5** (2007), no. 1 5–15.

[83] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. Van Den Berg, I. Titov, and M. Welling, *Modeling relational data with graph convolutional networks*, in *European Semantic Web Conference*, pp. 593–607, Springer, 2018.

[84] G. Mai, , K. Janowicz, S. Prasad, and B. Yan, *Visualizing the semantic similarity of geographic features*, in *The Annual International Conference on Geographic Information Science Short Paper Proceedings*, Springer, 2018.

[85] S. Sen, A. B. Swoap, Q. Li, B. Boatman, I. Dippenaar, R. Gold, M. Ngo, S. Pujol, B. Jackson, and B. Hecht, *Cartograph: unlocking spatial visualization through semantic enhancement*, in *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, pp. 179–190, ACM, 2017.

[86] K. Janowicz, Y. Hu, G. McKenzie, S. Gao, B. Regalia, G. Mai, R. Zhu, B. Adams, and K. Taylor, *Moon landing or safari? a study of systematic errors and their causes in geographic linked data*, in *The Annual International Conference on Geographic Information Science*, pp. 275–290, Springer, 2016.

[87] B. Regalia, K. Janowicz, and S. Gao, *Volt: a provenance-producing, transparent sparql proxy for the on-demand computation of linked data and its application to spatiotemporally dependent data*, in *European Semantic Web Conference*, pp. 523–538, Springer, 2016.

[88] M. Kejriwal and P. Szekely, *Neural embeddings for populated geonames locations*, in *International Semantic Web Conference*, pp. 139–146, Springer, 2017.

[89] M. E. Newman and M. Girvan, *Finding and evaluating community structure in networks*, *Physical review E* **69** (2004), no. 2 026113.

[90] J. Yang, J. McAuley, and J. Leskovec, *Community detection in networks with node attributes*, in *2013 IEEE 13th International Conference on Data Mining*, pp. 1151–1156, IEEE, 2013.

[91] K. LeFevre and E. Terzi, *Grass: Graph structure summarization*, in *Proceedings of the 2010 SIAM International Conference on Data Mining*, pp. 454–465, SIAM, 2010.

[92] H. Toivonen, F. Zhou, A. Hartikainen, and A. Hinkka, *Compression of weighted graphs*, in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 965–973, ACM, 2011.

[93] Z. Shen, K.-L. Ma, and T. Eliassi-Rad, *Visual analysis of large heterogeneous social networks by semantic and structural abstraction*, *IEEE transactions on visualization and computer graphics* **12** (2006), no. 6 1427–1439.

[94] Y. Mehmood, N. Barbieri, F. Bonchi, and A. Ukkonen, *Csi: Community-level social influence analysis*, in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 48–63, Springer, 2013.

[95] Š. Čebirić, F. Goasdoué, H. Kondylakis, D. Kotzinos, I. Manolescu, G. Troullinou, and M. Zneika, *Summarizing semantic graphs: A survey*, *The VLDB Journal* (2017).

[96] X. Zhang, G. Cheng, and Y. Qu, *Ontology summarization based on rdf sentence graph*, in *Proceedings of the 16th international conference on World Wide Web*, pp. 707–716, ACM, 2007.

[97] S. Khatchadourian and M. P. Consens, *Explod: Summary-based exploration of interlinking and rdf usage in the linked open data cloud*, in *Extended Semantic Web Conference*, pp. 272–287, Springer, 2010.

[98] M. Zneika, C. Lucchese, D. Vodislav, and D. Kotzinos, *Rdf graph summarization based on approximate patterns*, in *International Workshop on Information Search, Integration, and Personalization*, pp. 69–87, Springer, 2015.

[99] Q. Song, Y. Wu, P. Lin, L. X. Dong, and H. Sun, *Mining summaries for knowledge graph search*, *IEEE Transactions on Knowledge and Data Engineering* **30** (2018), no. 10 1887–1900.

[100] K. Hose and R. Schenkel, *Towards benefit-based rdf source selection for sparql queries*, in *Proceedings of the 4th International Workshop on Semantic Web Information Management*, p. 2, ACM, 2012.

[101] K. Janowicz, M. Raubal, and W. Kuhn, *The semantics of similarity in geographic information retrieval*, *Journal of Spatial Information Science* **2011** (2011), no. 2 29–57.

[102] T. Mikolov, K. Chen, G. Corrado, and J. Dean, *Efficient estimation of word representations in vector space*, *arXiv:1301.3781* (2013).

[103] S. Harnad, *To cognize is to categorize: Cognition is categorization*, *Handbook of categorization in cognitive science* (2005) 20–45.

[104] J. R. Firth, *A synopsis of linguistic theory, 1930-1955*, .

[105] Y.-F. Tuan, *Space and place: The perspective of experience.* Uni. of Minnesota, 1977.

[106] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, *A neural probabilistic language model*, *Journal of machine learning research* **3** (2003), no. Feb 1137–1155.

[107] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, *Distributed representations of words and phrases and their compositionality*, in *Advances in neural information processing systems*, pp. 3111–3119, 2013.

[108] Y. Zhang, A. Jatowt, and K. Tanaka, *Is tofu the cheese of asia?: Searching for corresponding objects across geographical areas*, in *Proceedings of the 26th International Conference on World Wide Web Companion*, pp. 1033–1042, International World Wide Web Conferences Steering Committee, 2017.

[109] Y. Yao, X. Li, X. Liu, P. Liu, Z. Liang, J. Zhang, and K. Mai, *Sensing spatial distribution of urban land use by integrating points-of-interest and google word2vec model*, *International Journal of Geographical Information Science* **31** (2017), no. 4 825–848.

[110] C. Zhang, K. Zhang, Q. Yuan, H. Peng, Y. Zheng, T. Hanratty, S. Wang, and J. Han, *Regions, periods, activities: Uncovering urban dynamics via cross-modal representation learning*, in *Proceedings of the 26th International Conference on World Wide Web*, pp. 361–370, International World Wide Web Conferences Steering Committee, 2017.

[111] S. Zhao, T. Zhao, I. King, and M. R. Lyu, *Geo-teaser: Geo-temporal sequential embedding rank for point-of-interest recommendation*, in *Proceedings of the 26th International Conference on World Wide Web Companion*, pp. 153–162, International World Wide Web Conferences Steering Committee, 2017.

[112] J. Kim, M. Vasardani, and S. Winter, *Similarity matching for integrating spatial information extracted from place descriptions*, *International Journal of Geographical Information Science* **31** (2017), no. 1 56–80.

[113] B. Adams and K. Janowicz, *Thematic signatures for cleansing and enriching place-related linked data*, *International Journal of Geographical Information Science* **29** (2015), no. 4 556–579.

[114] G. Quercini and H. Samet, *Uncovering the spatial relatedness in wikipedia*, in *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 153–162, ACM, 2014.

[115] M. Ye, K. Janowicz, C. Mülligann, and W.-C. Lee, *What you are is when you are: the temporal dimension of feature types in location-based social networks*, in *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 102–111, ACM, 2011.

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

[117] W. Li, *Random texts exhibit zipf's-law-like word frequency distribution*, *IEEE Transactions on information theory* **38** (1992), no. 6 1842–1845.

[118] A. Mnih and K. Kavukcuoglu, *Learning word embeddings efficiently with noise-contrastive estimation*, in *Advances in neural information processing systems*, pp. 2265–2273, 2013.

[119] S. Harispe, S. Ranwez, S. Janaqi, and J. Montmain, *Semantic similarity from natural language and ontology analysis*, *Synthesis Lectures on Human Language Technologies* **8** (2015), no. 1 1–254.

[120] Z. Wu and M. Palmer, *Verbs semantics and lexical selection*, in *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*, pp. 133–138, Association for Computational Linguistics, 1994.

[121] C. Leacock and M. Chodorow, *Combining local context and wordnet similarity for word sense identification*, *WordNet: An electronic lexical database* **49** (1998), no. 2 265–283.

[122] D. Lin *et. al.*, *An information-theoretic definition of similarity.*, in *Icml*, vol. 98, pp. 296–304, 1998.

[123] J. J. Jiang and D. W. Conrath, *Semantic similarity based on corpus statistics and lexical taxonomy*, *arXiv preprint cmp-lg/9709008* (1997).

[124] D. Sánchez, M. Batet, and D. Isern, *Ontology-based information content computation*, *Knowledge-Based Systems* **24** (2011), no. 2 297–303.

[125] N. Seco, T. Veale, and J. Hayes, *An intrinsic information content metric for semantic similarity in wordnet*, in *Proceedings of the 16th European conference on artificial intelligence*, pp. 1089–1090, IOS Press, 2004.

[126] N. Goodman, *Problems and projects*, .

[127] L. v. d. Maaten and G. Hinton, *Visualizing data using t-sne*, *Journal of Machine Learning Research* **9** (2008), no. Nov 2579–2605.

[128] M. Castelluccio, G. Poggi, C. Sansone, and L. Verdoliva, *Land use classification in remote sensing images by convolutional neural networks*, *arXiv preprint arXiv:1508.00092* (2015).

[129] W. Sun, V. Heidt, P. Gong, and G. Xu, *Information fusion for rural land-use classification with high-resolution satellite imagery*, *IEEE Transactions on Geoscience and Remote Sensing* **41** (2003), no. 4 883–890.

[130] N. Naik, S. D. Kominers, R. Raskar, E. L. Glaeser, and C. A. Hidalgo, *Computer vision uncovers predictors of physical urban change*, *Proceedings of the National Academy of Sciences* **114** (2017), no. 29 7571–7576.

[131] A. Zang, R. Xu, Z. Li, and D. Doria, *Lane boundary extraction from satellite imagery*, in *Proceedings of the 1st ACM SIGSPATIAL Workshop on High-Precision Maps and Intelligent Applications for Autonomous Vehicles*, p. 1, ACM, 2017.

[132] I. J. Goodfellow, Y. Bulatov, J. Ibarz, S. Arnoud, and V. Shet, *Multi-digit number recognition from street view imagery using deep convolutional neural networks*, *arXiv preprint arXiv:1312.6082* (2013).

[133] M. Cisse, Y. Adi, N. Neverova, and J. Keshet, *Houdini: Fooling deep structured prediction models*, *arXiv preprint arXiv:1707.05373* (2017).

[134] I. J. Goodfellow, J. Shlens, and C. Szegedy, *Explaining and harnessing adversarial examples*, *arXiv preprint arXiv:1412.6572* (2014).

[135] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, *Intriguing properties of neural networks*, *arXiv preprint arXiv:1312.6199* (2013).

[136] T. Bolukbasi, K.-W. Chang, J. Y. Zou, V. Saligrama, and A. T. Kalai, *Man is to computer programmer as woman is to homemaker? debiasing word embeddings*, in *Advances in Neural Information Processing Systems*, pp. 4349–4357, 2016.

[137] A. Caliskan, J. J. Bryson, and A. Narayanan, *Semantics derived automatically from language corpora contain human-like biases*, *Science* **356** (2017), no. 6334 183–186.

[138] J. Zhao, T. Wang, M. Yatskar, V. Ordonez, and K.-W. Chang, *Men also like shopping: Reducing gender bias amplification using corpus-level constraints*, *arXiv preprint arXiv:1707.09457* (2017).

[139] N. Papernot, P. McDaniel, X. Wu, S. Jha, and A. Swami, *Distillation as a defense to adversarial perturbations against deep neural networks*, in *Security and Privacy (SP), 2016 IEEE Symposium on*, pp. 582–597, IEEE, 2016.

[140] M. F. Goodchild and D. G. Janelle, *Thinking spatially in the social sciences*, *Spatially integrated social science* (2004) 3–22.

[141] K. Liu, S. Gao, P. Qiu, X. Liu, B. Yan, and F. Lu, *Road2vec: Measuring traffic interactions in urban road system from massive travel routes*, *ISPRS International Journal of Geo-Information* **6** (2017), no. 11 321.

[142] G. Heitz and D. Koller, *Learning spatial context: Using stuff to find things*, in *European conference on computer vision*, pp. 30–43, Springer, 2008.

[143] G. Baatz, O. Saurer, K. Köser, and M. Pollefeys, *Large scale visual geo-localization of images in mountainous terrain*, in *Computer Vision–ECCV 2012*, pp. 517–530. Springer, 2012.

[144] T.-Y. Lin, S. Belongie, and J. Hays, *Cross-view image geolocalization*, in *Computer Vision and Pattern Recognition*, pp. 891–898, IEEE, 2013.

[145] S. Lee, H. Zhang, and D. J. Crandall, *Predicting geo-informative attributes in large-scale image collections using convolutional neural networks*, in *Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on*, pp. 550–557, IEEE, 2015.

[146] J. Yu and J. Luo, *Leveraging probabilistic season and location context models for scene understanding*, in *Proceedings of the 2008 international conference on Content-based image and video retrieval*, pp. 169–178, ACM, 2008.

[147] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, *Places: A 10 million image database for scene recognition*, *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2017).

[148] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, *Gradient-based learning applied to document recognition*, *Proceedings of the IEEE* **86** (1998), no. 11 2278–2324.

[149] A. Krizhevsky, I. Sutskever, and G. E. Hinton, *Imagenet classification with deep convolutional neural networks*, in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[150] K. Simonyan and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*, *arXiv preprint arXiv:1409.1556* (2014).

[151] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, *et. al.*, *Going deeper with convolutions*, IEEE Conference on Computer Vision and Pattern Recognition, 2015.

[152] K. He, X. Zhang, S. Ren, and J. Sun, *Deep residual learning for image recognition*, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[153] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten, *Densely connected convolutional networks*, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, vol. 1, p. 3, 2017.

[154] S. Hochreiter and J. Schmidhuber, *Long short-term memory*, *Neural computation* **9** (1997), no. 8 1735–1780.

[155] W. Kuhn, T. Kauppinen, and K. Janowicz, *Linked data-a paradigm shift for geographic information science*, in *International Conference on Geographic Information Science*, pp. 173–186, Springer, 2014.

[156] A. Grover and J. Leskovec, *node2vec: Scalable feature learning for networks*, in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 855–864, ACM, 2016.

[157] B. L. Fredrickson and D. Kahneman, *Duration neglect in retrospective evaluations of affective episodes.*, *Journal of personality and social psychology* **65** (1993), no. 1 45.

[158] A. Thalhammer, I. Toma, A. Roa-Valverde, and D. Fensel, *Leveraging usage data for linked data movie entity summarization*, *arXiv preprint arXiv:1204.2718* (2012).

[159] A. Thalhammer and A. Rettinger, *Browsing dbpedia entities with summaries*, in *European Semantic Web Conference*, pp. 511–515, Springer, 2014.

[160] G. Pirrò, *Explaining and suggesting relatedness in knowledge graphs*, in *International Semantic Web Conference*, pp. 622–639, Springer, 2015.

[161] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, *Translating embeddings for modeling multi-relational data*, in *Advances in neural information processing systems*, pp. 2787–2795, 2013.

[162] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, *Dbpedia: A nucleus for a web of open data*, in *The semantic web*, pp. 722–735. Springer, 2007.

[163] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, *Freebase: a collaboratively created graph database for structuring human knowledge*, in *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pp. 1247–1250, ACM, 2008.

[164] D. Vrandečić and M. Krötzsch, *Wikidata: a free collaborative knowledgebase*, *Communications of the ACM* **57** (2014), no. 10 78–85.

[165] W. L. Hamilton, R. Ying, and J. Leskovec, *Representation learning on graphs: Methods and applications*, *arXiv preprint arXiv:1709.05584* (2017).

[166] P. Drineas, R. Kannan, and M. W. Mahoney, *Fast monte carlo algorithms for matrices iii: Computing a compressed approximate matrix decomposition*, *SIAM Journal on Computing* **36** (2006), no. 1 184–206.

[167] J. Sun, Y. Xie, H. Zhang, and C. Faloutsos, *Less is more: Compact matrix decomposition for large sparse graphs*, in *Proceedings of the 2007 SIAM International Conference on Data Mining*, pp. 366–377, SIAM, 2007.

[168] W. Xiong, T. Hoang, and W. Y. Wang, *Deeppath: A reinforcement learning method for knowledge graph reasoning*, in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 2017)*, (Copenhagen, Denmark), ACL, September, 2017.

[169] R. J. Williams, *Simple statistical gradient-following algorithms for connectionist reinforcement learning*, *Machine learning* **8** (1992), no. 3-4 229–256.

[170] R. Das, S. Dhuliawala, M. Zaheer, L. Vilnis, I. Durugkar, A. Krishnamurthy, A. Smola, and A. McCallum, *Go for a walk and arrive at the answer: Reasoning over paths in knowledge bases using reinforcement learning*, *arXiv preprint arXiv:1711.05851* (2017).

[171] Y. Shen, J. Chen, P.-S. Huang, Y. Guo, and J. Gao, *M-walk: Learning to walk over graphs using monte carlo tree search*, in *Advances in Neural Information Processing Systems*, pp. 6787–6798, 2018.

[172] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, *Enriching word vectors with subword information*, *Transactions of the Association for Computational Linguistics* **5** (2017) 135–146.

[173] S. Arora, Y. Liang, and T. Ma, *A simple but tough-to-beat baseline for sentence embeddings*, in *International Conference on Learning Representations*, 2017.

[174] G. Mai, K. Janowicz, and B. Yan, *Support and centrality: Learning weights for knowledge graph embedding models*, in *International Conference on Knowledge Engineering and Knowledge Management*, pp. 212–227, Springer, 2018.

[175] K. Janowicz, B. Yan, B. Regalia, R. Zhu, and G. Mai, *Debiasing knowledge graphs: Why female presidents are not like female popes*, in *International Semantic Web Conference*, 2018.

[176] C. J. Watkins and P. Dayan, *Q-learning*, *Machine learning* **8** (1992), no. 3-4 279–292.

[177] G. A. Rummery and M. Niranjan, *On-line Q-learning using connectionist systems*, vol. 37. University of Cambridge, Department of Engineering Cambridge, England, 1994.

[178] A. Hussein, M. M. Gaber, E. Elyan, and C. Jayne, *Imitation learning: A survey of learning methods*, *ACM Computing Surveys (CSUR)* **50** (2017), no. 2 21.

[179] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, *et. al.*, *Mastering the game of go with deep neural networks and tree search*, *nature* **529** (2016), no. 7587 484.

[180] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, *arXiv preprint arXiv:1412.6980* (2014).

[181] K. C. Clarke, *What is the worlds oldest map?*, *The Cartographic Journal* **50** (2013), no. 2 136–143.

[182] M. Batty, A. Hudson-Smith, R. Milton, and A. Crooks, *Map mashups, web 2.0 and the gis revolution*, *Annals of GIS* **16** (2010), no. 1 1–13.

[183] L. L. Hill, *Georeferencing: The geographic associations of information.* Mit Press, 2009.

[184] M. J. Hutchinson and B. Veenendaal, *An agent-based framework for intelligent geocoding*, *Applied Geomatics* **5** (2013), no. 1 33–44.

[185] J. H. Ratcliffe, *On the accuracy of tiger-type geocoded address data in relation to cadastral and census areal units*, *International Journal of Geographical Information Science* **15** (2001), no. 5 473–485.

[186] P. A. Zandbergen, *A comparison of address point, parcel and street geocoding techniques*, *Computers, Environment and Urban Systems* **32** (2008), no. 3 214–232.

[187] D. W. Goldberg and M. G. Cockburn, *Improving geocode accuracy with candidate selection criteria*, *Transactions in GIS* **14** (2010), no. s1 149–176.

[188] A. T. Murray, T. H. Grubesic, R. Wei, and E. A. Mack, *A hybrid geocoding methodology for spatio-temporal data*, *Transactions in GIS* **15** (2011), no. 6 795–809.

[189] C. A. Davis Jr, G. L. Pappa, D. R. R. de Oliveira, and F. de L Arcanjo, *Inferring the location of twitter messages based on user relationships*, *Transactions in GIS* **15** (2011), no. 6 735–751.

[190] P. Doherty, Q. Guo, Y. Liu, J. Wieczorek, and J. Doke, *Georeferencing incidents from locality descriptions and its applications: a case study from yosemite national park search and rescue*, *Transactions in GIS* **15** (2011), no. 6 775–793.

[191] A. I. Abdelmoty, P. Smart, and C. B. Jones, *Building place ontologies for the semantic web:: issues and approaches*, in *Proceedings of the 4th ACM workshop on Geographical information retrieval*, pp. 7–12, ACM, 2007.

[192] F. J. Lopez-Pellicer, M. J. Silva, M. Chaves, F. J. Zarazaga-Soria, and P. R. Muro-Medrano, *Geo linked data*, in *International Conference on Database and Expert Systems Applications*, pp. 495–502, Springer, 2010.

[193] F. J. Lopez-Pellicer, M. J. Silva, and M. Chaves, *Linkable geographic ontologies*, in *Proceedings of the 6th Workshop on Geographic Information Retrieval*, p. 1, ACM, 2010.

[194] R. Battle and D. Kolas, *Enabling the geospatial semantic web with parliament and geosparql*, *Semantic Web* **3** (2012), no. 4 355–370.

[195] S. Auer, J. Lehmann, and S. Hellmann, *Linkedgeodata: Adding a spatial dimension to the web of data*, in *International Semantic Web Conference*, pp. 731–746, Springer, 2009.

[196] S. Hahmann and D. Burghardt, *Connecting linkedgeodata and geonames in the spatial semantic web*, in *The 6th International Conference on Geographic Information Science*, 2010.

[197] J. Ugander, B. Karrer, L. Backstrom, and C. Marlow, *The anatomy of the facebook social graph*, *arXiv preprint arXiv:1111.4503* (2011).

[198] G. McKenzie, K. Janowicz, S. Gao, J.-A. Yang, and Y. Hu, *Poi pulse: A multi-granular, semantic signature–based information observatory for the interactive visualization of big geosocial data*, *Cartographica: The International Journal for Geographic Information and Geovisualization* **50** (2015), no. 2 71–85.