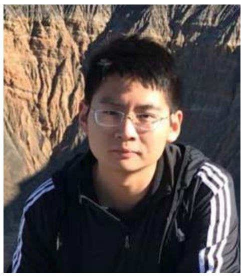
GENGCHEN MAI
STKO Lab
Department of Geography
University of California, Santa Barbara
Email: gengchen_mai@ucsb.edu



Gengchen Mai is a Ph.D. candidate at Space and Time for Knowledge Organization Lab, Department of Geography, University of California, Santa Barbara, advised by Prof. Krzysztof Janowicz. He is interested in Geographic Question Answering, NLP, Geographical Information Science (GIScience), Geographic Information Retrieval, Machine Learning/Deep Learning, and Semantic Web. Mai has done multiple machine learning/engineering internships at different companies such as Esri Inc., Mosaix.ai, and Apple Maps where he focused on cutting edge research projects about geospatial related machine learning/deep learning problems.

Spatially Explicit Machine Learning Model

Introduction

To my understanding, spatial data science, as a subfield of data science, refers to the discipline focusing on developing methods, algorithms, tools, and data sets for complex spatial problems. Although many efforts have been made by data scientists in general and many problems they tackle are inherently spatial problems (e.g., price prediction for housing and ride sharing), some key aspects of spatial information are neglected intentionally or unintentionally. In this position paper, we advocate the necessity of *spatial thinking* and *spatially-explicit models* [5,3,1] in contrast to rather general models.

Spatial information plays an important role in our daily life. Massive amount of location-based services enable us to plan beforehand which makes our life easier. For example, we may want to search for a nearby Japanese ramen restaurant with vegetarian options for dinner, or we may want to know which bus can take us to the destination and when it can arrive. If we have an important meeting planned for tomorrow during the rush hours, we may want to know how long it will take to drive there in advance to avoid being late. All these daily requests can be answered by different location-based services by using data-driven machine learning models in which spatial information is a non-negligible component.

Except for these problems where spatial information is an obvious component of the underlining data sets, there are other tasks where spatial information can help to improve the model performance significantly. Example tasks can be image classification[4,5] and fine grained image recognition[1]. When you go hiking and come across a bird of an unfamiliar species, you may want to take a picture, upload it to an AI image recognition app, and ask for its species. Instead of purely based on the visual information, when and where you took this photo can also help this recognition problem because different species have different habitats and different migration patterns[1].

Many machine learning tasks is geospatial related. However, it is not easy to process spatial information into a proper representation in order to be easily consumed by machine learning models due to its complexity. For example, Google Maps just launches a bus delay forecasting feature[1] in which how to prepare the bus routes and traffic information properly to be easily consumed by deep neural net architectures is an important problem. While we can divide the routes into road units and feed them into a sequential model, how can we encode the location and time information in each road unit such that it can be well handled by deep neural nets?

Spatially Explicit Model

As said by Michael Goodchild, "A model is said to be spatially explicit when it differentiates behaviors and predictions according to spatial locations"[2]. In the context of machine learning, we see a necessity of *spatially explicit machine learning models* which focus on better ways to design machine learning model architectures for consuming spatial information.

Many research has demonstrated the superiority of spatially-explicit models over more general models on several spatial related tasks. In the context of geographic question answering, by explicitly incorporating distance decay effect into the question relaxation model architecture, Mai et al.[3] show that the spatially-explicit knowledge graph embedding model outperform the traditional knowledge graph embedding models in geographic query relaxation task. As for geographic knowledge graph summarization, Yan et al.[6] demonstrate that the spatially-explicit model which adds spatial inductive bias in the reinforcement learning framework outperforms the traditional monte carlo policy model.

However, adding spatial inductive bias to make spatially-explicit models will yield more complex machine learning model architectures. Investigating the trade-off between designing a spatially-explicit machine learning architecture versus a more general setup that would have to learn to value space implicitly is an important research question for the spatial data science community.

References

1. **Grace Chu, Brian Potetz, Weijun Wang, Andrew Howard, Yang Song, FernandoBrucher, Thomas Leung, and Hartwig Adam**. Geo-aware networks for fine grained recognition. *arXiv preprint arXiv:1906.01737*, 2019.
2. **Michael F Goodchild and Donald G Janelle**. Thinking spatially in the social sciences.*Spatially integrated social science*, pages 3–17, 2004.
3. **Gengchen Mai, Bo Yan, Krzysztof Janowicz, and Rui Zhu**. Relaxing unanswerablegeographic questions using a spatially explicit knowledge graph embedding model. In *The Annual International Conference on Geographic Information Science*, pages 21–39. Springer, 2019.
4. **Kevin Tang, Manohar Paluri, Li Fei-Fei, Rob Fergus, and Lubomir Bourdev**. Improving image classification with location context. In *Proceedings of the IEEE international conference on computer vision*, pages 1008–1016, 2015.

[1] https://ai.googleblog.com/2019/06/predicting-bus-delays-with-machine.html

5. **Bo Yan, Krzysztof Janowicz, Gengchen Mai, and Rui Zhu**. xnet+ sc: Classifyingplaces based on images by incorporating spatial contexts. In *10th International Conference on Geographic Information Science (GIScience 2018)*. Schloss DagstuhlLeibniz-Zentrum fuer Informatik, 2018.
6. **Bo Yan, Krzysztof Janowicz, Gengchen Mai, and Rui Zhu**. A spatially explicit reinforcement learning model for geographic knowledge graph summarization. *Transactions in GIS* 23(3): 620–640, 20.