Positional Encoder Graph Neural Networks for Geographic Data

Konstantin Klemmer
Microsoft Research

Nathan Safir
University of Georgia

Daniel B. Neill
New York University

Abstract

Modeling spatial dependencies in geographic data is of crucial importance for the modeling of our planet. Graph neural networks (GNNs) provide a powerful and scalable solution for modeling continuous spatial data. However, in the absence of further context on the geometric structure of the data, they often rely on Euclidean distances to construct the input graphs. This assumption can be improbable in many real-world settings, where the spatial structure is more complex and explicitly non-Euclidean (e.g., road networks). In this paper, we propose PE-GNN, a new framework that incorporates spatial context and correlation explicitly into the models. Building on recent advances in geospatial auxiliary task learning and semantic spatial embeddings, our proposed method (1) learns a context-aware vector encoding of the geographic coordinates and (2) predicts spatial autocorrelation in the data in parallel with the main task. We show the effectiveness of our approach on two climate-relevant regression tasks: 3d spatial interpolation and air temperature prediction. The code for this study can be accessed via: `https://github.com/konstantinklemmer/pe-gnn`.

1 Introduction

Geographic data is characterized by a natural geometric structure, which often strongly affects the observed spatial pattern. While traditional neural network approaches do not have an intuition to account for spatial dynamics, graph neural networks (GNNs) can represent spatial structures graphically. The recent years have seen many applications leveraging GNNs for climate-related modeling tasks in the geographic domain, such as precipitation forecasting [2] or traffic modeling [3]. Nonetheless, as we show in this study, GNNs are not necessarily sufficient for modeling complex spatial effects: spatial context can be different at each location, which may be reflected in the relationship with its spatial neighborhood. The study of spatial context and dependencies has attracted increasing attention in the machine learning community, with studies on spatial context embeddings [17, 22] and spatially explicit auxiliary task learning [14].

In this study, we seek to merge these streams of research. We propose the positional encoder graph neural network (**PE-GNN**), a flexible approach for better encoding spatial context into GNN-based predictive models. **PE-GNN** is highly modular and can work with any GNN backbone. **PE-GNN** contains a positional encoder (PE) [20, 17], which learns a contextual embedding for point coordinates throughout training. **PE-GNN** also generalizes the spatial autocorrelation auxiliary learning approach proposed by [14] to continuous spatial data.

Lastly, we train **PE-GNN** by constructing a novel training graph, based on k-nearest-neighborhood, from a randomly sampled batch of points at each training step. This forces PE to learn generalizable features, as the same point coordinate might have different spatial context (neighbors) at different training steps.

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2022.

2 Related work

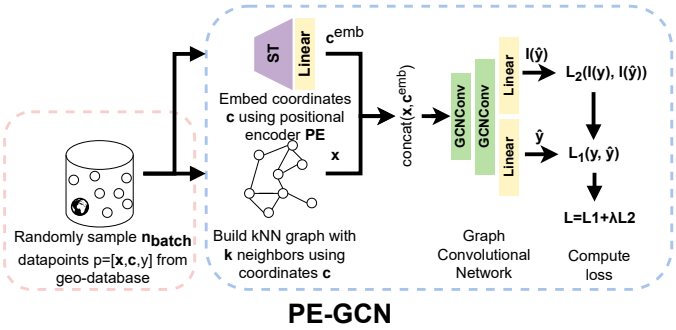

Figure 1: **PE-GCN** contains a (1) positional encoder network, learning a spatial context embedding throughout training which is concatenated with node-level features and (2) an auxiliary learner, predicting the spatial autocorrelation of the outcome variable simultaneously to the main regression task.

Recently, there has been a rise of research on applications of neural network models for spatial modeling tasks. More specifically, graph neural networks (GNNs) are often used for these tasks with the spatial data represented graphically. GNNs offer flexibility and scalability advantages over traditional spatial modeling approaches such as Gaussian processes [4]. Specific GNN architectures including Graph Convolutional Networks [13], Graph Attention Networks [21] and GraphSAGE [8] are powerful methods for inference and representation learning with spatial data. Recently, GNN approaches tailored to the specific complexities of geospatial data have been developed. The authors of Kriging Convolutional Networks [2] propose using GNNs to perform a modified kriging task. [8] apply GNNs for a spatio-temporal Kriging task, recovering data from unsampled nodes on an input graph. We look to extend this line of research by providing stronger, explicit capacities for GNNs to learn spatial structures. Additionally, our proposed method is highly modular and can be combined with any GNN backbone.

Through many decades of research and applications–from ecology to epidemiology–a myriad of measures, metrics, and statistics have been developed to cover a broad range of spatial interactions. Measures of spatial autocorrelation such as Moran's I [1] are particularly popular. Specifically, they have already been shown to be useful for improving neural network models through auxiliary task learning [14], model selection [15], embedding losses [16] and localized representation learning [7]. Recent years have also seen the emergence of neural network based embeddings for geographic information. Often trained in an unsupervised fashion, many of these embeddings are learnt from spatial context such as points-of-interest (POIs) or local social media data [7, 22, 17] and maybe be deployed for different downstream tasks.

3 Method

Graph Neural Networks with Geographic Data We elaborate on our method using Graph Convolutional Networks (GCNs) as an example backbone for our novel PE-GNN approach. We define a datapoint $p_i = \{y_i, \mathbf{x_i}, \mathbf{c_i}\}$, where y_i is a continuous target variable (scalar), $\mathbf{x_i}$ is a vector of predictive features and $\mathbf{c_i}$ is a vector of point coordinates, mapping the datapoint into geographic space (e.g., latitude and longitude values). Using a k-nearest-neighbor approach we create a graph $G = (V, E)$, consisting of a set of vertices (or nodes) $V = \{v_1, \ldots, v_n\}$ and a set of edges $E = \{e_1, \ldots, e_m\}$, assigned by the adjacency matrix $\mathbf{A}$. Each vertex $i \in V$ has respective node features $\mathbf{x_i}$ and target variable y_i. As proposed by [13], a GCN layer can now be defined as

$$\mathbf{H}^{(l)} = \sigma(\bar{\mathbf{A}}\mathbf{H}^{(l-1)}\mathbf{W}^{(l)}), l = 1, \ldots, L \tag{1}$$

where σ describes an activation function (e.g., ReLU), $\bar{\mathbf{A}}$ the normalized adjacency matrix and $\mathbf{W}^{(l)}$ is a weight matrix parametrizing GCN layer l. The input for the first GCN layer $\mathbf{H^{(0)}}$ is given by the feature matrix $\mathbf{X}$ containing all node feature vectors $\mathbf{x_1}, \ldots, \mathbf{x_n}$. The assembled GCN predicts the output $\hat{\mathbf{Y}} = GCN(\mathbf{X}, \Theta_{GCN})$ parametrized by Θ_{GCN}.

Context-aware spatial coordinate embeddings GCNs struggle with tasks that explicitly require learning of complex spatial dependencies. Their performance is highly susceptible to the graph definition (e.g., the chosen distance metric, number of neighbors). For example, we show

in our experiments that simple GCNs are not able to solve simple spatial interpolation tasks, i.e., predicting a continuous outcome variable from the point coordinates only. We propose a novel approach to overcome these difficulties, by devising a new positional encoder module, and learning a flexible spatial context encoding for each geographic coordinate, motivated by recent advances in transformers and spatial representation learning. Specifically, we define a positional encoder $PE(\mathbf{C}, \sigma_{min}, \sigma_{max}, \Theta_{PE}) = NN(ST(\mathbf{C}, \sigma_{min}, \sigma_{max}), \Theta_{PE})$, consisting of a sinusoidal transform $ST(\sigma_{min}, \sigma_{max})$ and a fully-connected neural network $NN(\Theta_{PE})$, parametrized by Θ_{PE}. Following the intuition of transformers [20] for geographic coordinates [17], the sinusoidal transform ST is a concatenation of scale-sensitive sinusoidal functions at different frequencies. The output from ST is then fed through the fully connected neural network $NN(\Theta_{PE})$ to transform it into the desired vector space shape, creating the coordinate embedding matrix $\mathbf{C}_{emb} = PE(\mathbf{C}, \sigma_{min}, \sigma_{max}, \Theta_{PE})$.

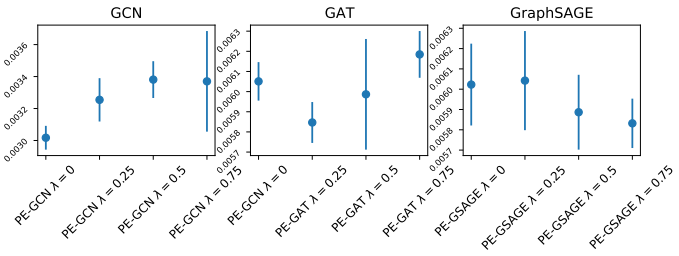

Figure 2: 3d road: MSE bar plots of mean performance and 2σ confidence intervals obtained from 10 different training checkpoints.

Auxiliary learning of spatial autocorrelation A further particularity of geographic data is that it often exhibits spatial autocorrelation: observations are related, in some shape or form, to their geographic neighbors. Measures of spatial autocorrelation have recently been successfully integrated into neural networks for discrete spatial data (images) [14, 16], however no such approach exists for continuous spatial coordinates. Spatial autocorrelation can be measured using the Moran's I metric of local spatial autocorrelation [1]. For our outcome variable y_i, it is defined as:

$$I_i = (n-1)\frac{(y_i - \bar{y}_i)}{\sum_{j=1}^{n}(y_j - \bar{y}_j)^2} \sum_{j=1, j\neq i}^{n} a_{i,j}(y_j - \bar{y}_j), \tag{2}$$

where $a_{i,j} \in \mathbf{A}$ denotes adjacency of observations i and j.

As proposed by [14], predicting the Moran's I metric of the output can be used as auxiliary task during training. Auxiliary task learning [19] is a special case of multi-task learning, where one learning algorithm tackles two or more tasks at once. The approach is commonly used, for example in reinforcement learning [6] or computer vision [10, 11]. Translated to our GCN setting, we seek to predict the outcome $\mathbf{Y}$ and its local Moran's I metric $I(\mathbf{Y})$ using the same network, so that $[\hat{\mathbf{Y}}, I(\hat{\mathbf{Y}})] = GCN(\mathbf{X})$. The local Moran's I metric is highly scale-sensitive which can limit its power [14, 5, 18]. To overcome this issue, we propose to train our model on a new, randomly sampled batch of geographic coordinates at each training step. The Moran's I for point i can thus change throughout iterations, reflecting a differing set of more distant or closer neighbors. Altogether, we refer to this altered Moran's I as "shuffled Moran's I".

Positional Encoder Graph Neural Network (PE-GNN) We now assemble the different modules of our method and introduce the Positional Encoder Graph Neural Network (**PE-GNN**). Assuming a batch B of randomly sampled points $p_1, \ldots, p_{n_{batch}} \in B$, a spatial graph is constructed from point coordinates $\mathbf{c}_1, \ldots, \mathbf{c}_{n_{batch}}$ using k-nearest-neighborhood, resulting in adjacency matrix $\mathbf{A}_B$. The point coordinates are then subsequently fed through the positional encoder $PE(\Theta_{PE})$, consisting of the sinusoidal transform ST and a single fully-connected layer with sigmoid activation, embedding the $2d$ coordinates in a customizable latent space, returning vector embeddings $\mathbf{c}_1^{emb}, \ldots, \mathbf{c}_{n_{batch}}^{emb} = \mathbf{C}_B^{emb}$. We then concatenate the positional encoder output with the node features, to create the input for the first layer of our GCN. To integrate the Moran's I auxiliary task, we compute the metric $I(\mathbf{Y}_B)$ for our outcome variable $\mathbf{Y}_B$ at the beginning of each training step. Prediction is then facilitated by creating two prediction heads, here linear layers, while the graph operation layers (e.g. GCN layers)

are shared between tasks. Finally, we obtain predicted values $\hat{\mathbf{Y}}_B$ and $I(\hat{\mathbf{Y}}_B)$. The loss of **PE-GCN** can be computed with any regression criterion, for example mean squared error (MSE).

4 Experiments

Model	Air Temp.		3d Road	
	MSE	MAE	MSE	MAE
GCN [13]	0.0225	0.1175	0.0169	0.1029
PE-GCN $\lambda = 0$	0.0040	0.0432	**0.0031**	**0.0396**
PE-GCN $\lambda = 0.25$	0.0037	0.0417	0.0032	0.0416
PE-GCN $\lambda = 0.5$	**0.0036**	**0.0401**	0.0033	0.0421
PE-GCN $\lambda = 0.75$	0.0040	0.0429	0.0033	0.0424
GAT [21]	0.0226	0.1165	0.0178	0.0998
PE-GAT $\lambda = 0$	**0.0039**	0.0429	0.0060	0.0537
PE-GAT $\lambda = 0.25$	0.0040	**0.0417**	**0.0058**	**0.0530**
PE-GAT $\lambda = 0.5$	0.0045	0.0465	0.0061	0.0548
PE-GAT $\lambda = 0.75$	0.0041	0.0429	0.0062	0.0562
GraphSAGE [8]	0.0274	0.1326	0.0180	0.0998
PE-GraphSAGE $\lambda = 0$	0.0039	0.0428	0.0060	**0.0534**
PE-GraphSAGE $\lambda = 0.25$	0.0040	0.0418	0.0059	**0.0534**
PE-GraphSAGE $\lambda = 0.5$	0.0043	0.0461	0.0060	0.0536
PE-GraphSAGE $\lambda = 0.75$	**0.0036**	**0.0399**	**0.0058**	0.0541
KCN [2]	0.0143	0.0927	0.0081	0.0758
PE-KCN $\lambda = 0$	0.0648	0.2385	**0.0025**	**0.0310**
PE-KCN $\lambda = 0.25$	**0.0059**	**0.0593**	0.0037	0.0474
PE-KCN $\lambda = 0.5$	0.0077	0.0664	0.0077	0.0642
PE-KCN $\lambda = 0.75$	0.0122	0.0852	0.0110	0.0755
Approximate GP	0.0481	0.0498	0.0080	0.0657
Exact GP	0.0084	0.0458	-	-

Table 1: *Spatial Interpolation:* Test MSE and MAE scores from four different datasets, using four different GNN backbones with and without our proposed architecture.

Data We evaluate **PE-GNN** and baseline competitors on two climate relevant, real-world geographic datasets: The *air temperature* dataset [9] contains the coordinates of $3,000$ weather stations around the globe. Here, we seek to predict mean temperatures y from a single node feature x, mean precipitation, and locations c. Air temperature prediction is relevant to many-climate related applications, from forecasting crop growth under increasingly extreme weather conditions, to the modeling of animal movement and behaviour. The *3d road* dataset [12] provides over $430,000$ 3-dimensional spatial coordinates (latitude, longitude, and altitude) of the road network in Jutland, Denmark. Here, altitude y is predicted using latitude and longitude coordinates $\mathbf{c}$. Such digital elevation models (DEMs) also have many applications in climate relevant domains, from modeling flooding exposure to species distribution models.

Baselines We compare **PE-GNN** with four different backbones: The original GCN formulation, introduced by [13] and outlined in the Methods section, graph attention mechanisms (GAT) [21], GraphSAGE [8] and Kriging Convolutional Networks (KCN) [2]. We compare the naive version of all these approaches to the same four backbone architectures augmented with our **PE-GNN** modules. We also provide Gaussian Process baselines. For all approaches, we compare a range of different training settings and hyperparameters.

Results The results of our experiments are shown in Table 1. Figure 2 shows MSE bar plots of the different methods on the 3d road dataset. Generally, **PE-GNN** substantially improves over naive baselines. Most of the improvement can be attributed to the positional encoder, however the auxiliary task learning also has substantial beneficial effects in some settings, especially for the KCN models. The best setting for the task weight hyperparameter λ seems to heavily depend on the data, which confirms findings by [14]. While being substantially more scalable, **PE-GNN** also performs well compared to Gaussian Processes.

5 Conclusion

With **PE-GNN**, we introduce a flexible, modular new GNN-based learning framework for geographic data. **PE-GNN** leverages recent findings in embedding spatial context into neural networks to improve predictive models. Our empirical findings confirm a strong performance. This study highlights how geographic domain expertise can help improve machine learning models for Earth observation data, a task relevant to many climate-related applications.

References

[1] L. Anselin. Local Indicators of Spatial Association—LISA. *Geographical Analysis*, 27 (2):93–115, sep 1995. ISSN 15384632. doi: 10.1111/j.1538-4632.1995.tb00338.x. URL `http://doi.wiley.com/10.1111/j.1538-4632.1995.tb00338.x`.

[2] G. Appleby, L. Liu, and L. P. Liu. Kriging convolutional networks. In *AAAI 2020 - 34th AAAI Conference on Artificial Intelligence*, volume 34, pages 3187–3194. AAAI press, apr 2020. ISBN 9781577358350. doi: 10.1609/aaai.v34i04.5716. URL www.aaai.org.

[3] C. Chen, K. Li, S. G. Teo, X. Zou, K. Wang, J. Wang, and Z. Zeng. Gated residual recurrent graph neural networks for traffic prediction. In *33rd AAAI Conference on Artificial Intelligence, AAAI 2019, 31st Innovative Applications of Artificial Intelligence Conference, IAAI 2019 and the 9th AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019*, volume 33, pages 485–492. AAAI Press, jul 2019. ISBN 9781577358091. doi: 10.1609/aaai.v33i01.3301485. URL www.aaai.org.

[4] A. Datta, S. Banerjee, A. O. Finley, and A. E. Gelfand. Hierarchical Nearest-Neighbor Gaussian Process Models for Large Geostatistical Datasets. *Journal of the American Statistical Association*, 111(514):800–812, apr 2016. ISSN 1537274X. doi: 10.1080/01621459.2015.1044091. URL https://www.tandfonline.com/doi/full/10.1080/01621459.2015.1044091.

[5] Y. Feng, L. Chen, and X. Chen. The impact of spatial scale on local Moran's I clustering of annual fishing effort for Dosidicus gigas offshore Peru. *Journal of Oceanology and Limnology*, 37(1):330–343, jan 2019. ISSN 25233521. doi: 10.1007/s00343-019-7316-9.

[6] Y. Flet-Berliac and P. Preux. MERL: Multi-Head Reinforcement Learning. In *NeurIPS 2019 - Deep Reinforcement Learning Workshop*, sep 2019. URL http://arxiv.org/abs/1909.11939.

[7] Y. Fu, P. Wang, J. Du, L. Wu, and X. Li. Efficient region embedding with multi-view spatial networks: A perspective of locality-constrained spatial autocorrelations. In *33rd AAAI Conference on Artificial Intelligence, AAAI 2019, 31st Innovative Applications of Artificial Intelligence Conference, IAAI 2019 and the 9th AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019*, volume 33, pages 906–913. AAAI Press, jul 2019. ISBN 9781577358091. doi: 10.1609/aaai.v33i01.3301906. URL www.aaai.org.

[8] W. L. Hamilton, R. Ying, and J. Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*, volume 2017-Decem, pages 1025–1035. Neural information processing systems foundation, jun 2017. URL http://arxiv.org/abs/1706.02216.

[9] J. Hooker, G. Duveiller, and A. Cescatti. Data descriptor: A global dataset of air temperature derived from satellite remote sensing and weather stations. *Scientific Data*, 5(1):1–11, nov 2018. ISSN 20524463. doi: 10.1038/sdata.2018.246. URL https://www.nature.com/articles/sdata2018246.

[10] Y. Hou, Z. Ma, C. Liu, and C. C. Loy. Learning to Steer by Mimicking Features from Heterogeneous Auxiliary Networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33 (01):8433–8440, jul 2019. ISSN 2159-5399. doi: 10.1609/aaai.v33i01.33018433.

[11] M. Jaderberg, V. Mnih, W. M. Czarnecki, T. Schaul, J. Z. Leibo, D. Silver, and K. Kavukcuoglu. Reinforcement learning with unsupervised auxiliary tasks. In *International Conference on Learning Representations (ICLR)*, nov 2017. URL https://youtu.be/Uz-zGYrYEjA.

[12] M. Kaul, B. Yang, and C. S. Jensen. Building accurate 3D spatial networks to enable next generation intelligent transportation systems. In *Proceedings - IEEE International Conference on Mobile Data Management*, 2013. doi: 10.1109/MDM.2013.24.

[13] T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*. International Conference on Learning Representations, ICLR, sep 2017. URL http://arxiv.org/abs/1609.02907.

[14] K. Klemmer and D. B. Neill. Auxiliary-task learning for geographic data with autoregressive embeddings. In *SIGSPATIAL: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems*, 2021.

[15] K. Klemmer, A. Koshiyama, and S. Flennerhag. Augmenting correlation structures in spatial data using deep generative models. *arXiv:1905.09796*, 2019. URL http://arxiv.org/abs/1905.09796.

[16] K. Klemmer, T. Xu, B. Acciaio, and D. B. Neill. SPATE-GAN: Improved Generative Modeling of Dynamic Spatio-Temporal Patterns with an Autoregressive Embedding Loss. In *AAAI 2022 - 36th AAAI Conference on Artificial Intelligence*, 2022.

[17] G. Mai, K. Janowicz, B. Yan, R. Zhu, L. Cai, and N. Lao. Multi-Scale Representation Learning for Spatial Feature Distributions using Grid Cells. In *International Conference on Learning Representations (ICLR)*, feb 2020. URL http://arxiv.org/abs/2003.00824.

[18] Y. Meng, C. Lin, W. Cui, and J. Yao. Scale selection based on Moran's i for segmentation of high resolution remotely sensed images. In *International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 4895–4898. Institute of Electrical and Electronics Engineers Inc., nov 2014. ISBN 9781479957750. doi: 10.1109/IGARSS.2014.6947592.

[19] S. C. Suddarth and Y. L. Kergosien. Rule-injection hints as a means of improving network performance and learning time. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 412 LNCS, pages 120–129. Springer Verlag, 1990. ISBN 9783540522553. doi: 10.1007/3-540-52255-7_33.

[20] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 2017-Decem, pages 5999–6009, 2017. URL https://research.google/pubs/pub46201/.

[21] P. Veličković, A. Casanova, P. Liò, G. Cucurull, A. Romero, and Y. Bengio. Graph attention networks. In *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*. International Conference on Learning Representations, ICLR, oct 2018. URL https://arxiv.org/abs/1710.10903v3.

[22] Y. Yin, Z. Liu, Y. Zhang, S. Wang, R. R. Shah, and R. Zimmermann. GPS2Vec: Towards generating worldwide GPS embeddings. In *SIGSPATIAL: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems*, pages 416–419, New York, NY, USA, nov 2019. Association for Computing Machinery. ISBN 9781450369091. doi: 10.1145/3347146.3359067. URL https://dl.acm.org/doi/10.1145/3347146.3359067.