

Review of the manuscript:

GNNWR: An Open-Source Package of Spatiotemporal Intelligent Regression Methods for Modeling Spatial and Temporal Non-Stationarity

This work introduces a software package called GNNWR for spatiotemporal regression tasks using neural networks. The package is a following implementation of the works of Du et al. (2020a) and Wu et al. (2020). The work further validates the implementation on two works namely, Du et al. (2020b) and Qi et al. (2023).

General comments

In my opinion, the manuscript lacks novelty and scientific contribution since it describes a typical framework which is used in most papers/works using machine learning. In addition, there are many similar frameworks such as Pytorch Lightning by Falcon et al. (2020) and there was no discussion about other existing frameworks. The model also seems very similar to Graph Neural Network (GNN) i.e., see Scarselli et al. (2009).

How does the proposed software differ from other deep learning frameworks such as Pytorch Lightning by Falcon et al. (2020)?

Regarding the model:

How does GNNWR differ from the idea of Graph Neural Network (GNN)?

Was there a comparison between GNNWR and other baselines which are widely used as regression models such as Multi-Layer Perceptron (MLP), XGBoosts, and GNN?

I think a discussion about a critical limitation is missing in the manuscript i.e., the estimation of each point is dependent on the distances to all points in the dataset. How is this scaled with large-scale data or global domain? And why shouldn't the estimation be dependent on the neighborhood? Is it essential to have an estimation based on all points in the dataset?

Since the spatiotemporal weights w_{ip} are computed by a neural network (hidden layers), I suspect that β_p can be removed because the network can learn these weights when estimating w_{ip} . Can you please comment on this?

Specific and technical comments

Lines 5-7: Please rephrase this sentence to make it clearer.

Line 75: Please update the reference for PyTorch i.e.:

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

Equation (2): What are u_i and v_i ? Do you mean the coordinates, please introduce the terms.

Equation (4): What is \hat{y}_i and where is the bias term? If you denote the estimated value as \hat{y}_i , please indicate this when you introduce the term.

Equation (5): What is w_{ij} and what are i and j ?

Line 97: What happens if we use an inverse of the distance?

Figure 1: What are the hidden layers? Do you mean Multi-Layer Perceptron?

Line 116: What is this symbol \otimes ? Do you mean the Hadamard product?

Lines 125-126: “Just like the GTWR model, we can use these coefficients to explain the estimated relationship and make spatiotemporal inference.” Was there an example of using the coefficients to explain the estimation described in the manuscript? Can you please refer to it?

Equation (12): What is X and where is β ?

Line 130: Please modify the name of this section. It is about the package description rather than a usage example.

Lines 181-187: Is a random split used to generate the training, validation and test sets? That means both validation and test sets will have the same distribution and will be similar to the training set.

Lines 189-191: Note that dropout in most deep learning frameworks such as Pytorch scales the output during inference to preserve the distribution inside the neural network. Does dropout have a similar affect for regression as for classification tasks? Was dropout also effective for the regression tasks mentioned in the manuscript?

Line 239: What are AIC and AICs?

Lines 254-265: Are the reported metrics for the test set?

Lines 303-306: “This method computes the distance vectors between the features in the dataset to be predicted and the reference points, and applies the identical scaling transformation to the independent variables as in the training dataset, guaranteeing the reliability of the prediction outcomes.” This doesn’t guarantee reliability. It guarantees that the input for the model inference follows the same statistical distribution of the training data.

Figures: It would be better to provide more description in the captions.

Line 425: “... may be retrieved ...” What is meant by *may be*?

Lines 408-409: “Employing advanced neural network techniques, these models enhance the precision of discerning spatiotemporal non-stationary features.” What is exactly meant by advanced neural networks techniques? I do not think MLP is an advanced neural network. Advanced neural networks include i.e., ViT by Dosovitskiy et al. (2020) etc.

Lines 543-545: Please correct the date for the reference Wu et al. (2020).

References:

Dosovitskiy et al. (2020): <https://doi.org/10.48550/arXiv.2010.11929>.

Du et al. (2020a): <https://doi.org/10.1080/13658816.2019.1707834>.

Du et al. (2020b): <https://doi.org/10.12082/dqxxkx.2020.190533>.

Falcon et al. (2020): <https://zenodo.org/doi/10.5281/zenodo.3530844>.

Qi et al. (2023): <https://doi.org/10.1016/j.scitotenv.2023.163981>.

Scarselli et al. (2009): <https://doi.org/10.1109/TNN.2008.2005605>.

Wu et al. (2020): <https://doi.org/10.1080/13658816.2020.1775836>.