EGU24-17389, updated on 11 Mar 2024
https://doi.org/10.5194/egusphere-egu24-17389
EGU General Assembly 2024

Analyzing Spatio-Temporal Machine Learning Models through Input Perturbation

Claire Robin et al.

The biogeoscience community has increasingly embraced the application of machine learning models across various domains from fire prediction to vegetation forecasting. Yet, as these models become more widely used, there is sometimes a gap in understanding between what we assume the model learns and what the model actually learns. For example, Long-short Term Memory (LSTM) models are applied to long time series, hoping they benefit from access to more information, despite their tendency to rapidly forget information. This can lead to erroneous conclusions, misinterpretation of results, and an overestimation of the models, ultimately eroding trust in their reliability.

To address this issue, we employ an explainable artificial intelligence (XAI) post hoc perturbation technique that is task-agnostic and model-agnostic. We aim to examine the extent to which the model leverages information for its predictions, both in terms of time and space. In other words, we want to observe the actual receptive field utilized by the model. We introduce a methodology designed to quantify both the spatial impact of neighboring pixels on predicting a specific pixel and the temporal periods contributing to predictions in time series models. The experiments take place after training the model, during inference. In the spatial domain, we define ground-truth pixels to predict, then examine the increase in prediction error, caused by shuffling their neighboring pixels at various distances from the selection. In the temporal domain, we investigate how shuffling a sequence of frames within the context period at different intervals relative to the target period affects the increase in prediction loss. This method can be applied across a broad spectrum of spatio-temporal tasks. Importantly, the method is easy-to-implement, as it only relies on the inference of predictions at test time and the shuffling of the perturbation area.

For our experiments, we focus on the vegetation forecasting task, i.e., forecasting the evolution of the Vegetation Index (VI) based on Sentinel-2 imagery using previous Sentinel-2 sequences and weather information to guide the prediction. This task involves both spatial non-linear dependencies arising from the spatial context (e.g., the surrounding area, such as a river or a slope, directly influencing the VI) and non-linear temporal dependencies such as the gradual onset of drought conditions and the rapid influence of precipitation events. We compare several models for spatio-temporal tasks, including ConvLSTM and transformer-based architectures on their usage of neighboring pixels in space, and context period in time. We demonstrate that the ConvLSTM relies on a restricted spatial area in its predictions, indicating a limited utilization of the spatial context up to 50m (5 pixels). Furthermore, it utilizes the global order of the time series sequence to capture the seasonal cycle but loses sensitivity to the local order after 15 days (3 frames). The introduced XAI method allows us to quantify spatial and temporal behavior exhibited by machine learning methods.

How to cite: Robin, C., Benson, V., Requena-Mesa, C., Alonso, L., Poehls, J., Russwurm, M., Carvalhais, N., and Reichstein, M.: Analyzing Spatio-Temporal Machine Learning Models through Input Perturbation, EGU General Assembly 2024, Vienna, Austria, 14–19 Apr 2024, EGU24-17389, https://doi.org/10.5194/egusphere-egu24-17389, 2024.