Master Thesis Proposal: Self-Supervised Extreme Event Prediction

Joel Oskarsson, joel.oskarsson@liu.se

November 2, 2021

1 Project

In extreme event prediction tasks there is sometimes plenty of sensor data available, but very few events to act as positive examples [3]. We would like to make robust predictions for these events, but the few positive examples are not enough to learn useful representations for the sensor data.

The idea is to instead perform explicit representation learning as an initial step and later use the learned representation to predict the extreme events (the downstream task). For this we want to use contrastive self-supervised learning [4, 2, 1] to learn a representation of the sensor data. This allows us to learn a lower-dimensional useful representation in an unsupervised manner. Using this representation, we then want to use the few positive examples to learn to predict the extreme events. We can then rely on some less data-hungry method for the actual event prediction. Finding the exact method (and evaluating different ones) will be part of the project. Bayesian methods are of interest.

This method will be applied specifically to earthquake prediction. Initially it will be enough to predict if an event happens on each day, creating a binary classification task. Even more interesting would be to also predict properties of the event, such as location and scale. This also brings an additional spatial aspect into the problem, since the sensors are spatially distributed and will contain contextual information as to the position of the event. Also the uncertainty in predictions is of great interest.

2 Data description

The dataset to be used is earthquake data from the AETA competition[1]. This dataset contains data on just 200 earthquakes, but relevant sensor information from 159 sensor stations over 3 years. The sensors measure electromagnetic and acoustic signals with a period of 10 minutes. A total of 95 features are provided for each such measurement.

3 Research objectives

- Apply self-supervised representation learning to the AETA earthquake data. Verify the usefulness of the learned representations for the task of earthquake prediction.
- Find and compare prediction models making use of the learned representation for the downstream task.
- Investigate how the spatio-temporal structure of the data can be used in the representation learning and/or downstream task.

[1] https://aeta.io/

4 Required student background

The project will require wide knowledge of different machine learning models and a solid understanding of the theory underpinning them. As self-supervised learning is done using deep neural networks, experience implementing and working with deep learning models and frameworks is a must.

5 Application

Send grades and a short description of why you are interested in the project to `joel.oskarsson@liu.se`. Get in touch if you have any questions.

References

[1] Kumar Ayush, Burak Uzkent, Chenlin Meng, Kumar Tanmay, Marshall Burke, David Lobell, and Stefano Ermon. Geography-aware self-supervised learning. pages 10181–10190. URL `https://openaccess.thecvf.com/content/ICCV2021/html/Ayush_Geography-Aware_Self-Supervised_Learning_ICCV_2021_paper.html`.

[2] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. URL `http://arxiv.org/abs/2002.05709`.

[3] A. Galkina and N. Grafeeva. Machine learning methods for earthquake prediction: A survey.

[4] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. URL `http://arxiv.org/abs/1807.03748`.