Contrastive self-supervised learning for climate spatio-temporal data

Alexander Marusov [1]

Abstract

The aim of the representation learning problem is to obtain *universal* and *high-quality* representations of the data. One of the ways to solve a such problem is to use Self-Supervised Learning (SSL). SSL is a fundamental paradigm originating from computer vision that now underlies numerous methods in different domains, including time series processing. In particular *contrastive-based methods* have become widespread because of their practical effectiveness and simple working idea — to make similar (*positive*) objects *closer* and distancing different (*negative*) ones. Such an approach naturally allows to prevent mode *complete collapse* when all representations of different objects are the same. Despite the wide usage of the self-supervised learning paradigm for different domains application of SSL for climate spatio-temporal data is scarce. We aim to fill this gap by proposing a new contrastive-based SSL approach for climate spatio-temporal data. The code is available in the Github. [1]

1. Introduction

Nowadays, more and more Deep Learning (DL) algorithms are being used to solve different kinds of tasks. Typically, it is about supervised DL methods, which rely on labelled data. However, many real applied objectives lack a sufficient amount of annotated data, resulting in poor quality of the supervised model. Self-supervised learning (SSL) methods show great potential to reduce this quality gap using a massive amount of unlabeled data available for many scenarios (Ericsson et al., 2022; Lin et al., 2023). The primary purpose of SSL methods is to obtain an encoder that produces *universal* representation (embedding) of an input signal that generalizes across various *downstream* tasks (Gálvez et al., 2023).

Generally, SSL approaches can be divided into generative and discriminative (Grill et al., 2020). The last one consists of contrastive and non-contrastive algorithms. Contrastive methods make *closer* representations of *similar* objects and *further* for *different* ones (Joshi & Mirzasoleiman, 2023). Such an approach naturally avoids the *complete collapse* problem, i.e., when representations of different objects are the same (Jing et al., 2022).

Many different algorithms such as SimCLR (Chen et al., 2020), MoCo (Chen et al., 2021), BYOL (Grill et al., 2020), Barlow Twins (Zbontar et al., 2021), VICReg (Bardes et al., 2021), DINO (Caron et al., 2021) were originally developed for **computer vision (CV)** domain. Moreover, current theoretical fundamental papers are also devoted to the SSL methods from the CV area (Balestriero & LeCun, 2022; Jing et al., 2022; Tian et al., 2021; Ji et al., 2023). Some of those methods were adapted to the time series domain. For example, BYOL (Marusov & Zaytsev, 2023; Niizumi et al., 2021) and Barlow Twins (Anton et al., 2023) provide quality representations for audio; a variant of the Triplet model (Romanenkova et al., 2022) shows promising results for the industrial sensors data analysis. However, a naive adaptation of those methods does not account for the peculiarities of dependent data.

Classic SSL approach for **time series** is Contrastive Predictive Coding (CPC) (Oord et al., 2018). CPC uses compressing and autoregressive techniques to create a low-dimensional embedding space and make predictions for several steps in the future. More recent TS2Vec (Yue et al., 2022) positions itself as a universal approach for any time series (Foumani et al., 2024). They use hierarchical contrastive loss to learn representations at different time scales. CoST (Woo et al., 2022) framework decomposes the original signal into season and trend components with a transition to the frequency domain.

Creating high-quality embeddings for spatio-temporal data—data that combines both spatial and temporal components—using self-supervised learning (SSL) is both challenging and essential. In particular climate data has its own peculiarities that should be accounted in SSL models. Available research efforts (Wang et al., 2022; 2023; Lee et al., 2024) present different approaches to handle climate spatio-temporal data. However most of those methods were straightforwardly adapted from computer vision domain thus being ineffective to process temporal dependency in the data.

[1] https://github.com/Astralex98/Contrastive-self-supervised-learning-for-climate-spatio-temporal-data

Novelty. Given the existing body of work, we aim to propose a new spatio-temporal SSL method that should properly handle both spatial and temporal correlation in the data. In particular we propose to generalize TS2Vec model from time series to spatio-temporal case. Since TS2Vec was originally developed for time series this method should properly handle *temporal correlations*. By applying different spatial augmentations that have already shown their effectiveness in (Wang et al., 2022) and using spatio-temporal encoder ConvLSTM (Shi et al., 2015) we handle *spatial correlations*.

2. Related work

2.1. Climate spatio-temporal SSL

Climate-related tasks require special attention among spatio-temporal problems because of their critical importance to human well-being. The authors (Wang et al., 2022) adapted SimCLR to classify weather systems in East China. To create a positive pair for each sample x_i within the randomly sampled batch, they take a sample t_i hours later than x_i from the original dataset. The gap t_i is taken from the distribution that monotonically decreases with time. After the sample is taken, spatial augmentations (like resizing, cropping, and Gaussian blur) are applied, resulting in a positive pair and treating all other samples as negatives. However, such an approach has several issues. Firstly, batches consist of random samples. Hence, there is a possible situation where many truly positive examples can be considered as negatives. Secondly, contrasting occurs in a mixed way, without separately dividing the penalties for each dependency among themselves. Another approach (Lee et al., 2024) is to apply generative modeling to reconstruct masked spatio-temporal data and test their representations for rainfall probability estimation downstream task. The authors (Wang et al., 2023) adapted different approaches from computer vision (MoCo, DINO, MAE) to a large corpus of the satellite imagery to solve standard computer vision downstream tasks — scene classification and semantic segmentation. However, those methods mostly concentrate on handling only spatial dependency rather then temporal one.

2.2. Benchmarks

According to our research we found the following benchmarks for spatio-temporal data:

Drought dataset. The first benchmark is from (Marusov et al., 2024). This dataset collects the recordings of the drought index (Palmer Drought Severity Index, PDSI (Alley, 1984)) for different regions from various climate zones — Madhya Pradesh (India), Missouri(USA), Northern Kazakhstan, and Goias (Brazil) considering a time span from 1958 to 2022 with a monthly resolution.

WeatherBench. The authors (Tan et al., 2023) proposed a benchmark for different spatio-temporal tasks including climate domain. For weather forecasting tasks researchers used *WeatherBench* dataset containing history of different climatic variables (temperature, precipitation and etc.). The researchers considered different spatial resolutions (32×64 and 128×256) and different setups — single variable (*WeatherBench-S*) and multi-variable (*WeatherBench-M*). For the first setup authors used data from 2010 to 2018 with one hour temporal resolution while for the second one time range encompassed time period from 1979 to 2015 with six hour resolution.

SSL4EO-S12. The authors (Wang et al., 2023) shared an unlabeled dataset **SSL4EO-S12** corpus of various satellite images from different sources including Sentinel-1, Sentinel-2. The overall size of the dataset is $251,079$ images. The downstream tasks are mostly related to computer vision - scene classification, semantic segmentation.

2.3. Research gap

According to our review, we see that most of SSL-approaches for climate data are more concentrated on spatial dependencies without proper handling of temporal dependency.

3. Experiments

3.1. Downstream problems

3.1.1. Drought prediction

Problem importance. Among different climate-related problems, the drought prediction task is in particular demand these days since it is crucial in the global climate-changing context (Xiujia et al., 2022). Given the available history of the drought index for a particular region, our aim is to predict the spatial distribution of drought over the area up to twelve months ahead. This task belongs to long-term drought forecasting.

Target variable. The most suitable drought indicator target for our task is the Palmer Drought Severity Index, PDSI (Alley, 1984). Indeed, (McPherson et al., 2022) use PDSI as a suitable drought index for long-term drought forecasting. Moreover, PDSI is freely available in Google Earth Engine (Gorelick et al., 2017). We consider a time span from 1958 to 2022 with a monthly resolution. To make the forecast more interpretable, the regression task (predicting drought index directly) transforms into a binary classification problem by dividing PDSI values into two classes: drought or its absence. In particular, the case is considered as a drought if the PDSI value is less than -2 aligning with (Marusov et al., 2024; McPherson et al., 2022).

3.1.2. TEMPERATURE FORECASTING

Problem importance. Temperature is one of the key climate variables, since it is important for everyday life on the one hand, and is included in the calculation of other various indicators (for example, the drought index) on the other. The task of the temperature forecasting is to predict it for a week in advance.

Target variable. The target variable is the temperature at a height of two meters above the Earth's surface. The data is taken from the WeatherBench reference dataset (Rasp et al., 2020), which collects historical data on various climate characteristics. The temperature data ranges from 1979 to 2018 year inclusive with hour resolution.

3.1.3. DIVIDING DATA INTO SSL-PRETRAINING AND SUPERVISED DOWNSTREAM EVALUATION.

We used temperature and precipitation data from 1979 to 2016 to pretrain the encoder using SSL-regime. To evaluate performance of the embeddings we used temperature data from 2016 to 2018 and drought prediction task for Northern Kazakhstan and Goias (Brazil).

3.2. General description of the supervised model based on SSL embeddings

The general formalized description of the model for spatio-temporal task is the following:

- Input: History of the climatic variable — 3D-tensor: time, longitude, and latitude.
- Output: Predicted distribution of the climatic variable over the given area — a 2D tensor for a selected forecasting horizon.

To solve the target task using the SSL paradigm, we divide our model into two parts: a **SSL-pretrained encoder** and a **downstream classifier**. Using the contrastive SSL architecture described below, we pretrain the encoder. After SSL pretraining, we freeze the weights and provide feature representations of the input tensors for downstream problems.

3.3. Model architecture

The architecture of original TS2Vec is depicted in Figure 1. We generalized TS2Vec for the spatio-temporal case in the following way:

- **Augmentations.** We use standard spatial augmentations from computer vision following (Wang et al., 2022): Resizing (all regions reduce to one spatial size — 60×60), Gaussian blur, average pooling.
- **Encoder.** To account for both spatial and temporal dependencies, we use ConvLSTM architecture (Shi et al., 2015). The output dimension from ConvLSTM is $H \times W \times C_{cell}$ where H, W are spatial sizes of the region and C_{cell} is the embedding size of a single cell. We use one-dimensional convolutions and a fully connected layer to transform the tensor from $H \times W \times C_{cell}$ to $B \times T \times C$ dimension.

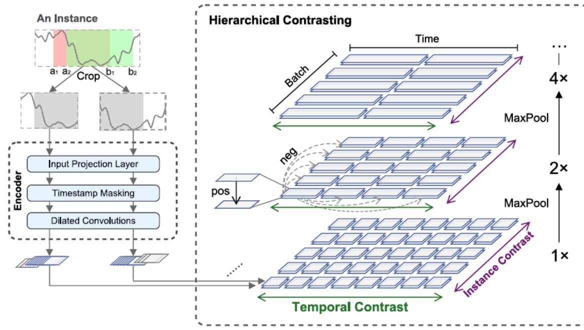

Figure 1: Original TS2Vec

3.4. Implementation details

General setting. We mostly follow the training protocol from the original paper (Yue et al., 2022). The maximum number of epochs is 100. The training stops if, during ten consequent epochs, the loss function on the validation set doesn't decrease. Learning rate is 0.001 following TS2Vec.

Encoder. The hidden dimension and embedding size for each cell are set to 8. The embedding size of the region is 512. Kernel size is $(3, 3)$. The number of layers is 3. The optimal history length is 60.

Augmentations. The optimum resize shape for all regions is 60×60. The kernel size for Gaussian blur and average pooling is 5.

3.5. Results

In the following, we examine the quality of the embeddings in two different downstream problems — *drought prediction* and *temperature forecasting*.

3.5.1. DROUGHT PREDICTION

We studied the quality of the embeddings via comparison of two classifiers:

1. **XGBoost.** XGBoost solved drought classification task using only the raw data.
2. **XGBoost with SSL.** XGBoost used embeddings from SSL-pretraining.

In the Fig. 2 you can find comparison between the XGBoost fitted on the raw data and using SSL-pretrained embeddings.

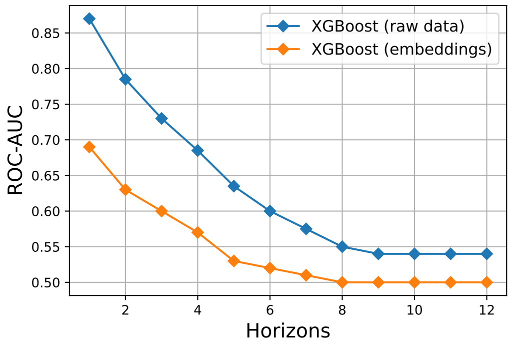

Figure 2: Performance averaged over regions

On the one hand we can see that currently SSL-pretrained embeddings are worse than original data. But on the other — representations contain some useful information since ROC-AUC of the first horizon is near 0.7.

3.5.2. Temperature forecasting

Besides drought prediction we also made a temperature forecasting for a week further. We used logistic regression as a classifier since it showed better results. The results are presented in the Fig. 3.

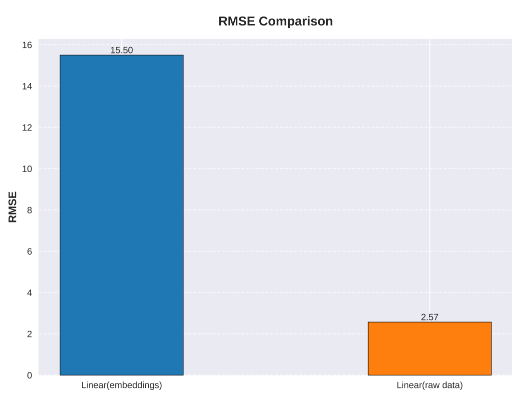

Figure 3: RMSE metric for a week forecasting

Unfortunately we can see that the quality of the SSL-embeddings here is worse than original data.

4. Conclusions

In this work the TS2Vec architecture was generalized from time series to the spatio-temporal data. The model was pre-trained on a huge corpus of data which contains temperature and precipitation indicators from 1979 to 2016 year. Those embeddings were tested on the two downstream problems — *drought prediction* and *temperature forecasting*.

The experiments showed that classifiers show better performance using only raw data rather than SSL-embeddings. I see several steps to improve the quality of SSL-embeddings:

- **More appropriate accounting of temporal information.** Despite that TS2Vec positions itself as a universal SSL-framework for time series authors claim that two nearby observations should be treated as a negative pair. However this claim is wrong for specific type of data (e.g. climate), where nearby observations for sure can be treated as a positive pair. For example, temperature indicators for 2 p.m. and 3 p.m. are definitely correlate with each other. Hence we need a specific treatment for such type of data.
- **Data periodicity.** The potential periodicity of the data (like in the climate case) should be mentioned and processed separately due to its importance.

References

Alley, W. M. The palmer drought severity index: limitations and assumptions. *Journal of climate and applied meteorology*, 1984.

Anton, J., Coppock, H., Shukla, P., and Schuller, B. W. Audio barlow twins: Self-supervised audio representation learning. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5. IEEE, 2023.

Balestriero, R. and LeCun, Y. Contrastive and non-contrastive self-supervised learning recover global and local spectral embedding methods. In *Advances in Neural Information Processing Systems (NeurIPS 2022)*, pp. 26671–26685, 2022.

Bardes, A., Ponce, J., and LeCun, Y. Vicreg: Variance-invariance-covariance regularization for self-supervised learning. In *arXiv preprint arXiv:2105.04906.*, 2021.

Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., and Joulin, A. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9650–9660, 2021.

Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. A simple framework for contrastive learning of visual representations. In *International conference on machine learning (ICML 2020)*, pp. 1597–1607, 2020.

Chen, X., Xie, S., and He, K. An empirical study of training self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision (IEEE/CVF 2021)*, pp. 9640–9649, 2021.

Ericsson, L., Gouk, H., Loy, C. C., and Hospedales, T. M. Self-supervised representation learning: Introduction, advances, and challenges. *IEEE Signal Processing Magazine*, 39(3):42–62, 2022.

Foumani, N. M., Tan, C. W., Webb, G. I., Rezatofighi, H., and Salehi, M. Series2vec: similarity-based self-supervised representation learning for time series classification. *Data Mining and Knowledge Discovery*, 2024.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, 202:18–27, 2017.

Grill, J. B., Strub, F., Altché, F., Tallec, C., Richemond, P., Buchatskaya, E., and Valko, M. Bootstrap your own latent-a new approach to self-supervised learning. In *Advances in neural information processing systems (NeurIPS 2020)*, pp. 21271–21284, 2020.

Gálvez, B. R., Blaas, A., Rodríguez, P., Golinski, A., Suau, X., Ramapuram, J., and Zappella, L. The role of entropy and reconstruction in multi-view self-supervised learning. In *International Conference on Machine Learning (ICML 2023)*, pp. 29143–29160, 2023.

Ji, W., Deng, Z., Nakada, R., Zou, J., and Zhang, L. The power of contrast for feature learning: A theoretical analysis. *Journal of Machine Learning Research*, 24(330): 1–78, 2023.

Jing, L., Vincent, P., LeCun, Y., and Tian, Y. Understanding dimensional collapse in contrastive self-supervised learning. In *International conference on learning representations (ICLR 2022)*, 2022.

Joshi, S. and Mirzasoleiman, B. Data-efficient contrastive self-supervised learning: Most beneficial examples for supervised learning contribute the least. In *International Conference on Machine Learning (ICML 2023)*, pp. 15356–15370, 2023.

Lee, J., An, S., You, S., and Cho, N. Self-supervised learning with probabilistic density labeling for rainfall probability estimation. *arXiv preprint arXiv:2412.05825*, 2024.

Lin, W., He, C., Mak, M. W., and Tu, Y. Self-supervised neural factor analysis for disentangling utterance-level speech representations. In *International Conference on Machine Learning (ICML 2023)*, pp. 21065–21077, 2023.

Marusov, A. and Zaytsev, A. Noncontrastive representation learning for intervals from well logs. *IEEE Geoscience and Remote Sensing Letters*, 20, 2023.

Marusov, A., Grabar, V., Maximov, Y., Sotiriadi, N., Bulkin, A., and Zaytsev, A. Long-term drought prediction using deep neural networks based on geospatial weather data. *Environmental Modelling & Software*, 179, 2024.

McPherson, R. A., Corporal-Lodangco, I. L., and Richman, M. B. A place-based approach to drought forecasting in south-central oklahoma. *Earth and Space Science*, 2022.

Niizumi, D., Takeuchi, D., Ohishi, Y., Harada, N., and Kashino, K. Byol for audio: Self-supervised learning for general-purpose audio representation. In *International Joint Conference on Neural Networks (IJCN 2021)*, pp. 1–8, 2021.

Oord, A. V. D., Li, Y., and Vinyals, O. Representation learning with contrastive predictive coding. In *arXiv preprint arXiv:1807.03748*, 2018.

Rasp, S., Dueben, P. D., Scher, S., Weyn, J. A., Mouatadid, S., and Thuerey, N. Weatherbench: a benchmark data set for data-driven weather forecasting. *Journal of Advances in Modeling Earth Systems*, 12:e2020MS002203, 2020.

Romanenkova, E., Rogulina, A., Shakirov, A., Stulov, N., Zaytsev, A., Ismailova, L., and AlShehri, A. Similarity learning for wells based on logging data. *Journal of Petroleum Science and Engineering*, 215, 2022.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., and Woo, W.-c. Convolutional lstm network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems*, 28, 2015.

Tan, C., Li, S., Gao, Z., Guan, W., Wang, Z., Liu, Z., Wu, L., and Li, S. Z. Openstl: A comprehensive benchmark of spatio-temporal predictive learning. *Advances in Neural Information Processing Systems*, 36:69819–69831, 2023.

Tian, Y., Chen, X., and Ganguli, S. Understanding self-supervised learning dynamics without contrastive pairs. In *International Conference on Machine Learning (ICML, 2021)*, pp. 10268–10278, 2021.

Wang, L., Li, Q., and Lv, Q. Self-supervised classification of weather systems based on spatiotemporal contrastive learning. *Geophysical Research Letters*, 49(15), 2022.

Wang, Y., Braham, N. A. A., Xiong, Z., Liu, C., Albrecht, C. M., and Zhu, X. X. Ssl4eo-s12: A large-scale multimodal, multitemporal dataset for self-supervised learning in earth observation [software and data sets]. *IEEE Geoscience and Remote Sensing Magazine*, 11(3):98–106, 2023.

Woo, G., Liu, C., Sahoo, D., Kumar, A., and Hoi, S. Cost: Contrastive learning of disentangled seasonal-trend representations for time series forecasting. In *arXiv preprint arXiv:2202.01575*, 2022.

Xiujia, C., Guanghua, Y., Jian, G., Ningning, M., and Zihao, W. Application of wnn-pso model in drought prediction at crop growth stages: A case study of spring maize in semi-arid regions of northern china. *Computers and Electronics in Agriculture*, 2022.

Yue, Z., Wang, Y., Duan, J., Yang, T., Huang, C., Tong, Y., and Xu, B. Ts2vec: Towards universal representation of time series. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI, 2022)*, pp. 8980–8987, 2022.

Zbontar, J., Jing, L., Misra, I., LeCun, Y., and Deny, S. Barlow twins: Self-supervised learning via redundancy reduction. In *International conference on machine learning (ICML, 2021)*, pp. 12310–12320, 2021.