Comparison of Self-Distilled and Contrastive Time-Series Representations for Noisy Classification

Assignee Research

June 13, 2026

Abstract

Self-supervised learning for time-series data holds potential similar to that recently unleashed in Natural Language Processing and Computer Vision. While most existing works in this area focus on contrastive learning, we propose a conceptually simple yet powerful non-contrastive approach, based on the data2vec self-distillation framework. The core of our method is a student-teacher scheme that predicts the latent representation of an input time series from masked views of the same time series. This strategy avoids strong modality-specific assumptions and biases typically introduced by the des

1 Introduction

This paper examines: Self-Distilled Representation Learning for Time Series. Research question: How do self-distilled time-series representation methods compare to contrastive learning approaches in terms of classification accuracy under varying levels of Gaussian noise on the UCR archive?.

2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.2/10.

3 Results

13 papers retrieved. 14 claims extracted; 11 independently verified. Quality review score: 7.2/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
The experimental setup follows a two-step procedure: (1) learning the encoder in a self-supervised fashion without any l	✓	0.35
The protocol is closely aligned with the one of TS2Vec [32], which serves as the primary reference point for comparison	✓	0.22
Self-distillation is prone to representation collapse, which is why hyperparameter optimization (HPO) was performed on a	✓	0.30
For time-series classification, instance-level representations are obtained by performing a max-aggregation over all tim	✓	0.25
The approach is benchmarked on the UCR archive [10] and UEA archive [3], which consist of 128 (univariate) and 30 (multi	✓	0.23
The data2vec scheme is highly competitive with existing SSL methods, as shown in Table 1.	✓	0.16
The reported scores for all comparison methods are taken from [32].	✓	0.20
The self-distillation training objective follows the SSL approach of data2vec [2], where the teacher model provides a ta	✓	0.37
The target representation is computed by averaging the hidden activations over the last K layers of the teacher model.	✓	0.23
The teacher's weights follow the student model according to an Exponential Moving Average (EMA) mechanism during trainin	✓	0.28
The data2vec approach is well-suited for time-series representation learning due to its simplicity and generalizability.	×	0.12
The timestamp masking strategy bypasses the limitations and unintentional biases that typically occur when handcrafting	✓	0.27
The average accuracy scores over all datasets of each archive are reported in Table 1: 0.832 for UCR and 0.738 for UEA.	×	0.13
The average MSE and MAE scores for time-series forecasting are reported in Table 2, with averages of 0.1688 and 0.2971 f	×	0.14

References

- http://arxiv.org/abs/2306.10125v4
- http://arxiv.org/abs/2210.16611v2
- http://arxiv.org/abs/2311.11335v1