You're your own best teacher: A Self-Supervised Learning Approach For Expressive Representations in Time Series Generation

Johan Vik Mathisen

June 13, 2024

We are able to simultaneously reconstruct well and significantly improve downstream classification accuracy. Additionally, from the t-SNE plots **??**, **??** and **??**, we observe that NC-VQVAE can efficiently perform clustering as well. We improve both IS, FID and CAS for most datasets, indicating that the conditional distribution is better captured, as well as the synthetic data being closer to the ground truth.

NC-VQVAE is more adept at mimicking training data and less likely to overlook underrepresented classes. However, for datasets with limited training samples, our model can be prone to overfitting. Some challenges associated with TimeVQVAE remain, particularly in modeling data with sharp differences in modularity.

While issues persist, we believe our model represents a step forward. The representations learned by NC-VQVAE are more expressive than those from the naive VQVAE, showing that multiple objectives can be optimized without sacrificing reconstruction capability. These expressive representations facilitate capturing the semantics of the training data and producing synthetic samples with better global consistency.

In summary, NC-VQVAE demonstrate the ability to classify accurately and perform effective clustering while maintaining reconstruction capabilities, as well as enhancing generative quality.

0.1   Further work

There are many exiting directions this work can be taken further. In [**morningstar2024augmentations**] they underscore that focus on augmentations is of great importance. The hunt for good augmentations in the time series domain is ongoing and should probably get more attention. An easy to use library of time series augmentations would be very beneficial to the community.

As we considered a simplified version of TimeVQVAE, it is natural to investigate effective ways of incorporating the HF-LF split. A possible approach could be to concatenate HF and LF representations and pass them through the projector. This would also open the possibility of applying augmentations tailored for HF and LF. In the same direction of sort of obvious extension, would be to further optimize the relationship between the reconstruction loss on the augmented branch and choice of augmentations, and understanding their interplay. Undergoing a more thorough analysis on the effect of the different losses, would be interesting.

It is of interest to better understand the geometry and topology of the discrete latent representations. A possible direction of future research could be to further investigate the effect of different augmentations and SSL methods by applying

topological data analysis techniques.

The discrepancy between validation prior loss and generative quality observed with NC-VQVAE is not understood. NC-VQVAE consistently produces samples with higher fidelity and improved generative scores, but it fails to model the latent space in such a way that the validation prior loss is minimized. It would be very interesting to investigate further, and one possibility would be to consider the attention maps and try to deduce relationships between the tokens.