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

Johan Vik Mathisen

June 4, 2024

0.1 Conclusion

Key takeaways: We are able to simultaneously reconstruct well and significantly improve downstream classification accuracy, which is very interesting from a representation learning perspective. 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. Additionally we see some differences in Barlow Twins and VIbCReg when it comes to sample diversity.

To summarize NC-VQVAE is able to capture the conditional distribution of the data better than naive VQVAE for a wide variety of datasets. For datasets with few training samples, our model can be prone to overfitting. We see our model as a step in the right direction, but further development is needed to ensure better intraclass diversity, possibly through a more refined sampling procedure.

NC-VQVAE is better able to mimic the training data. When data is abundant, then our model better captures the entire distribution, while covering

Some of the issues of TimeVQVAE are still highly relevant, such as the difficulty in modelling data with sharp differences in modularity, such as TwoPatterns and ElectricDevices.

Even though there are issues, we believe our model is a step in the right direction. The representations learned by NC-VQVAE are more expressive than naive VQVAE, demonstrating that we can optimize more than one objective, without sacrificing the reconstruction capability.

The representations enable easier learning of the semantics of the conditional distributions, to such a degree that one has to take measures not to overfit.

The representations makes it easier to capture the gobal consistency of the samples, which tha naive VQVAE has large issues with. This without the HF-LF split.

The added flexibility of NC-VQVAE, with possibility of choosing dataset specific augmentations, can in some applications be beneficial.

0.2 Further work

[**morningstar2024augmentations**] suggest 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.

HF-LF split - augmentations tailored for HF and LF, as they often have quite different characteristics.

Wavelet transform to improve HF-LF split.

Further optimize the relationship between aug recon loss and choice of augmentations.

Improving on the stage 2 learning to better handle the expressive representations, and be able to create more diverse samples. Higher masking ratio during training, lower value for T etc.

The differences in Barlow and VIbCReg indicate that further optimization of the SSL method/pretext task for generative performance is possible and could be an interesting extension of this project.

Investigation of the attention maps. Could one find if there is a HF direction, translation direction etc. similarly to how one in NLP can find gender direction, nationality direction etc?