Supplementary Material for Dynamic-Aware GANs: Time-Series Generation with Handy Self-Supervision

Anonymous authors
Paper under double-blind review

The source code can be found at `https://anonymous.4open.science/r/DAGAN-ICLR4119/README.md`. This supplementary material provides additional model descriptions, visualization and test results and ablation study, not included in the main paper.

1 More training details

Differencing:

In experiments, for DAGAN, we construct the stepwise conditional variable discussed in Section 4 of our main paper as the rolling mean or simple moving average of $\mathbf{d}_t$ with span k as a hyper-parameter, where $\mathbf{d}_t = \mathbf{x_{t+1}} - \mathbf{x_t}$. In other words, we obtain $\bar{\mathbf{d}}_t = \frac{1}{k} \sum_{i=0}^{k-1} \mathbf{d_{t-i}}$ and concatenate it as the condition for DAGAN. The simple moving average was considered as taking the average 1). smoothens the dynamics of $\mathbf{d}_t$, and potentially aids in its prediction; 2). accounts for previous time-steps; and 3). reduces the weight of the difference at the current time-step. From our experiments, we note that DAGAN benefits from greater values of k when data is noisy, such as in the stocks dataset when $k = 20$ yields the most optimal evaluation scores. The subsequent results of our DAGAN in the paper are presented with the optimized k value. The other hyperparameters, such as batch size and the number of iterations, require tuning for each dataset.

2 Visualization

PCA plots for DAGAN, TimeGAN, and ExtraMAE on the three main datasets (stocks, sine, and energy) are provided for both the complete and limited data settings.
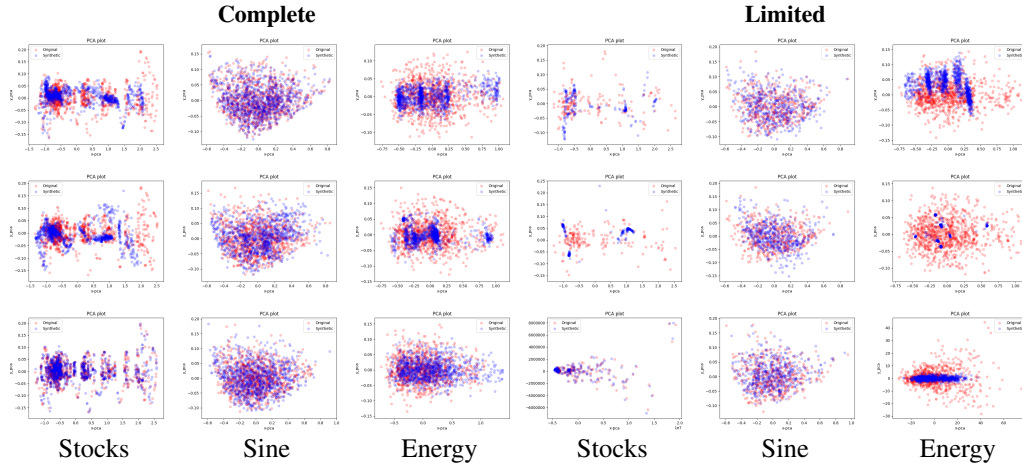

Figure 1: PCA visualization by DAGAN (1st row), TimeGAN (2nd row) and ExtraMAE (3rd row) for the complete (left) and limited (right) datasets. Each column is the results for Stocks, Sine, and Energy.

We note from Figure 1 that generally DAGAN has good overlap, particularly in the limited data case. Although ExtraMAE's visualization suggests better performance, such as in the complete

energy dataset, we argue that because of the nature of the model, it is more prone to memorizing the training data.

3 ADDITIONAL TESTS

The calculated MMD scores between X (synthetic data), Y (training data), and Z (test data) on the limited stock dataset are shown in Table 1. These scores give an indication of whether the model has simply memorized the training data.

Method	MMD(X, Y)	MMD(X, Z)
DAGAN	.002689	.002503
TimeGAN	.001302	.006861
ExtraMAE	.000021	.002467

Table 1: MMD scores of (limited) stock dataset between X, Y, and Z.

Ideally, both MMD scores (Gretton et al., 2006) between X and Y, and X and Z, should be low and comparable. This would mean that the model generates high quality and diverse data that is representative of the true data distribution. From Table 1, it is likely that both the TimeGAN and ExtraMAE has memorized the training data to some extent because the MMD between X and Z is much higher than the score between X and Y, despite the relatively high quality of synthetic data.

4 ABLATION STUDY

We conduct an ablation study to analyze the importance of joint distribution matching via the Discriminator. Hence, we modify DAGAN to exclude the joint distribution matching and compare its performance to the initial design. Table 2 suggests that joint distribution matching plays a crucial role in improving the quality of the synthetic data.

Table 2: Ablation analysis on DAGAN for both complete and limited data settings. The discriminative and predictive scores are computed for evaluation. Lower scores indicate superior performance.

Setting	Metric	Method	Sine	Stocks	Energy
Complete Data	Discriminative	DAGAN	$.006 \pm .004$	$.085 \pm .044$	$.439 \pm .013$
		w/o Joint Matching	$.016 \pm .029$	$.164 \pm .054$	$.493 \pm .003$
	Predictive	DAGAN	$.093 \pm .000$	$.038 \pm .000$	$.328 \pm .006$
		w/o Joint Matching	$.093 \pm .000$	$.038 \pm .000$	$.376 \pm .005$
		Original	$.094 \pm .001$	$.036 \pm .001$	$.250 \pm .003$
Limited Data	Discriminative	DAGAN	$.036 \pm .029$	$.119 \pm .059$	$.479 \pm .015$
		w/o Joint Matching	$.065 \pm .063$	$.246 \pm .092$	$.493 \pm .003$
	Predictive	DAGAN	$.096 \pm .001$	$.044 \pm .000$	$.291 \pm .005$
		w/o Joint Matching	$.094 \pm .000$	$.043 \pm .000$	$.380 \pm .004$
		Original	$.094 \pm .000$	$.037 \pm .001$	$.253 \pm .001$

REFERENCES

Arthur Gretton, Karsten Borgwardt, Malte Rasch, Bernhard Schölkopf, and Alex Smola. A kernel method for the two-sample-problem. *Advances in neural information processing systems*, 19, 2006.