Group Presentation on TranAD: Deep Transformer Networks for Anomaly Detection in Multivariate Time Series Data

https://arxiv.org/pdf/2201.07284.pdf

Tobramycin[1], Shreshth Tuli[2], Giuliano Casale[2], Nicholas R. Jennings[2]

[1] Presenting author, [2] Paper author,

May 25, 2022

THE UNIVERSITY OF MEMPHIS

Table of Contents

Introduction

Problems & Contributions

Problems:

- Simple transformer-based encoder-decoder networks tend to miss anomalies

Contributions:

- Tuli propose a transformer-based anomaly detection model (TranAD)
- TranAD uses self-conditioning. robust, training stability, generalization.
- TranAD uses an adversarial training process. Adversarial training procedure that can amplify reconstruction errors

Methodology

Framework

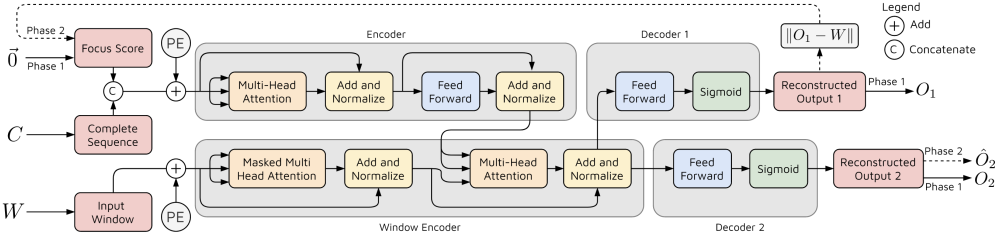
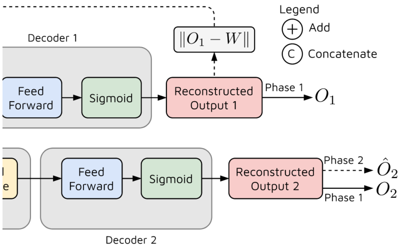


Figure 1: The TranAD Model.

$$\text{focus score } F = [0]_{K\times m}, \text{time slice } C \in \mathbb{R}^{K\times m}, \text{input window} W_t = \{x_{t-K+1}, \ldots, x_t\} \in \mathbb{R}^{K\times m} \quad (1)$$

$$O_1, O_2, \hat{O}_2 \in \mathbb{R}^{1\times m} \quad (2)$$

Loss



GAN loss:

$$\min_G \max_D V(D, G) \tag{3}$$

TranAD loss:

$$\min_{\text{Decoder1}} \max_{\text{Decoder2}} \left\| \hat{O}_2 - W \right\|_2 \tag{4}$$

Decoder 1 want to minimize $\left\| \hat{O}_2 - W \right\|_2$,

Decoder 2 want to maximize $\left\| \hat{O}_2 - W \right\|_2$

The loss of 2 phases:

$$L_1 = \epsilon^{-n} \|O_1 - W\|_2 + (1 - \epsilon^{-n}) \left\| \hat{O}_2 - W \right\|_2$$
$$L_2 = \epsilon^{-n} \|O_2 - W\|_2 - (1 - \epsilon^{-n}) \left\| \hat{O}_2 - W \right\|_2$$

Loss

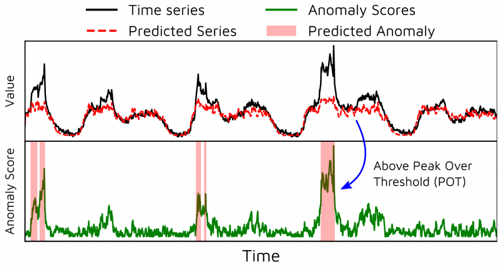


Figure 2: Visualization of anomaly prediction.

anomaly score:

$$s = \frac{1}{2}\left\|O_1 - \hat{W}\right\|_2 + \frac{1}{2}\left\|\hat{O}_2 - \hat{W}\right\|_2 \quad (5)$$

Anomaly diagnosis label:

$$y_i = 1\left(s_i \geq \mathrm{POT}\left(s_i\right)\right), \quad (6)$$

$$y = \vee_i y_i. \quad (7)$$

Experiments

Dataset

(1) Numenta Anomaly Benchmark (NAB): is a dataset of multiple real-world data traces, including readings from temperature sensors, CPU utilization of cloud machines, service request latencies and taxi demands in New York city [2].
(2) HexagonML (UCR) dataset: is a dataset of multiple univariate time series (included just for completeness) that was used in KDD 2021 cup [13, 26].
(3) MIT-BIH Supraventricular Arrhythmia Database (MBA): is a collection of electrocardiogram recordings from four patients, containing multiple instances of two different kinds of anomalies.
(4) Soil Moisture Active Passive (SMAP) dataset
(5) Mars Science Laboratory (MSL) dataset
(6) Secure Water Treatment (SWaT) dataset
(7) Water Distribution (WADI) dataset
(8) Server Machine Dataset (SMD)
(9) Multi-Source Distributed System (MSDS) Dataset

Result

Table 2: Performance comparison of TranAD with baseline methods on the complete dataset. P: Precision, R: Recall, AUC: Area under the ROC curve, F1: F1 score with complete training data. The best F1 and AUC scores are highlighted in bold.

Method	NAB				UCR				MBA			
	P	R	AUC	F1	P	R	AUC	F1	P	R	AUC	F1
MERLIN	0.8013	0.7262	0.8414	0.7619	0.7542	0.8018	0.8984	0.7773	0.9846	0.4913	0.7828	0.6555
LSTM-NDT	0.6400	0.6667	0.8322	0.6531	0.5231	0.8294	0.9781	0.6416	0.9207	0.9718	0.9780	0.9456
DAGMM	0.7622	0.7292	0.8572	0.7453	0.5337	0.9718	0.9916	0.6890	0.9475	0.9900	0.9858	0.9683
OmniAnomaly	0.8421	0.6667	0.8330	0.7442	0.8346	0.9999	0.9981	0.9098	0.8561	1.0000	0.9570	0.9225
MSCRED	0.8522	0.6700	0.8401	0.7502	0.5441	0.9718	0.9920	0.6976	0.9272	1.0000	0.9799	0.9623
MAD-GAN	0.8666	0.7012	0.8478	0.7752	0.8538	0.9891	0.9984	0.9165	0.9396	1.0000	0.9836	0.9689
USAD	0.8421	0.6667	0.8330	0.7442	0.8952	1.0000	0.9989	0.9447	0.8953	0.9989	0.9701	0.9443
MTAD-GAT	0.8421	0.7272	0.8221	0.7804	0.7812	0.9972	0.9978	0.8761	0.9018	1.0000	0.9721	0.9484
CAE-M	0.7918	0.8019	0.8019	0.7968	0.6981	1.0000	0.9957	0.8222	0.8442	0.9997	0.9661	0.9154
GDN	0.8129	0.7872	0.8542	0.7998	0.6894	0.9988	0.9959	0.8158	0.8832	0.9892	0.9528	0.9332
TranAD	0.8889	0.9892	**0.9541**	**0.9364**	0.9407	1.0000	**0.9994**	**0.9694**	0.9569	1.0000	**0.9885**	**0.9780**

Method	SMAP				MSL				SWaT			
	P	R	AUC	F1	P	R	AUC	F1	P	R	AUC	F1
MERLIN	0.1577	0.9999	0.7426	0.2725	0.2613	0.4645	0.6281	0.3345	0.6560	0.2547	0.6175	0.3669
LSTM-NDT	0.8523	0.7326	0.8602	0.7879	0.6288	1.0000	0.9532	0.7721	0.7778	0.5109	0.7140	0.6167

Acknowledgment

Thank you!