AnoShift: A Distribution Shift Benchmark for Unsupervised Anomaly Detection

Marius Dragoi [1] Elena Burceanu [1,2] Emanuela Haller [1,3] Andrei Manolache [1] Florin Brad [1]

[1]Bitdefender, Romania [2]University of Bucharest [3]Politehnica University of Bucharest

Introduction

We approach the distribution shift for network intrusion detection and introduce **AnoShift**, an unsupervised anomaly detection benchmark built over Kyoto-2006+.

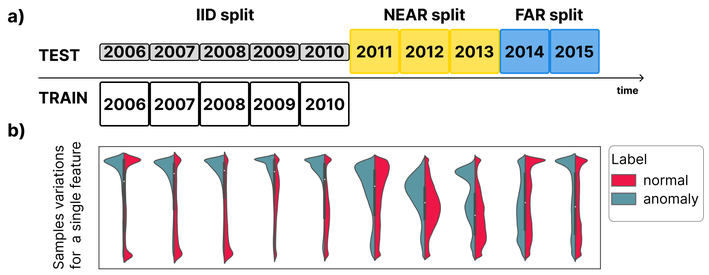

Figure 1. The proposed AnoShift splits over Kyoto-2006+ dataset.

In **AnoShift**, we split the data in **IID**, **NEAR**, and **FAR** testing splits.

Contributions

- We analyzed a large dataset for the unsupervised anomaly detection in network traffic (Kyoto-2006+) and prove that it is affected by distribution shifts.
- We propose a chronology-based benchmark, which focuses on splitting the test data based on its temporal distance to the training set: `IID`, `NEAR`, `FAR` (Fig. 2).
- We prove that acknowledging the distribution shift may improve anomaly detection, with a distillation method impacting the performance by 3% on average.

Per-feature distribution shift

We extract the normalized histogram per year for each feature and compute the Jeffreys divergence between those histograms.

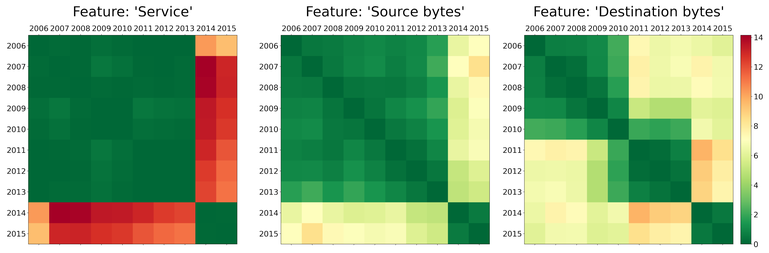

Figure 2. Jeffreys divergence between Kyoto years for 3 features

References

[1] Mohiuddin Ahmed, Abdun Naser Mahmood, and Jiankun Hu. A survey of network anomaly detection techniques. *J. Netw. Comput. Appl.*, 2016.

[2] Zhipeng Cai, Ozan Sener, and Vladlen Koltun. Online continual learning with natural distribution shifts: An empirical study with visual data. In *IEEE/CVF International Conference on Computer Vision, ICCV*, 2021.

Inherent non-stationarity

We generate a t-SNE representation in 2D for each split and observe that the discrepancy between point clouds increases with the temporal distance between splits, with clouds becoming more separated over time.

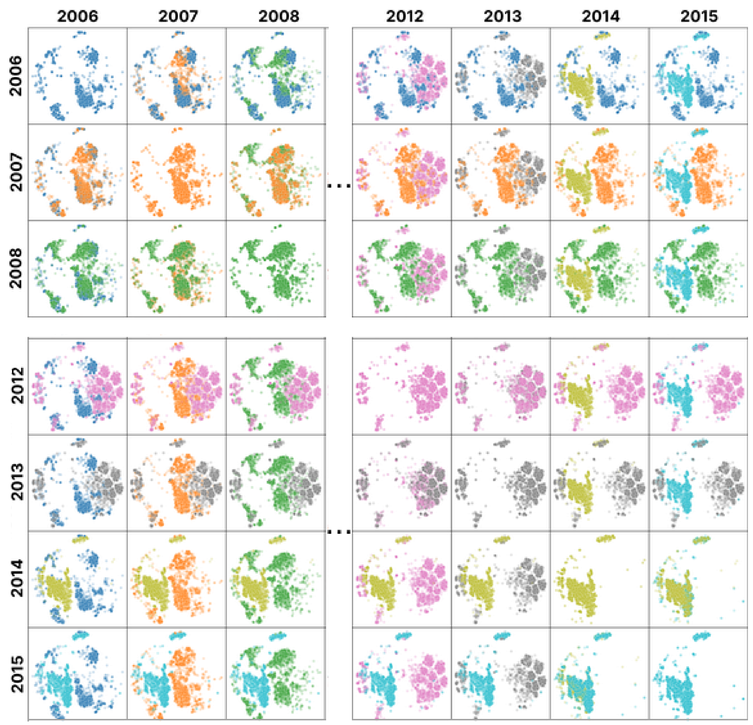

Figure 3. Comparison between yearly splits using t-SNE visualization.

General distribution shift

An Optimal Transport Dataset Distance for Kyoto shows the distances between the inliers (first), inliers and outliers (second), and outliers (third) of each set.

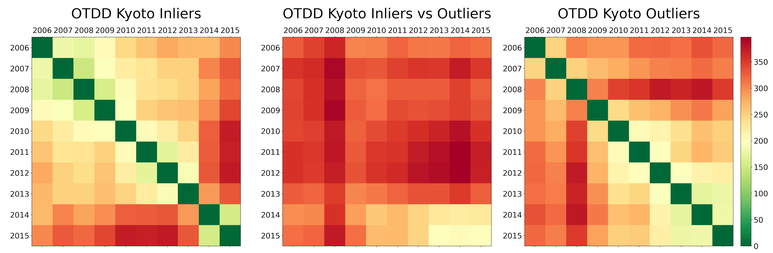

Figure 4. Optimal Transport Dataset Distance for Kyoto.

Impact on IID anomaly detection models

We evaluate the ROC-AUC evolution of several baselines over time: **IID** vs **NEAR** vs **FAR** and prove that the **ROC-AUC** is dropping over time in all cases and the variance for **FAR** is the highest.

	ROC-AUC ↑				
Split	IsoForest	OC-SVM	deepSVDD	LOF	BERT for anomalies
`IID`	78.73 ± 1.23	76.78 ± 0.43	77.87 ± 2.69	**86.58** ± 2.35	84.54 ± 0.07
`NEAR`	58.08 ± 6.53	72.73 ± 2.02	74.78 ± 6.29	74.45 ± 1.24	**86.05** ± 0.25
`FAR`	26.54 ± 2.65	**46.96** ± 3.07	44.33 ± 6.72	32.74 ± 12.55	28.15 ± 0.06

Monthly performance

A monthly evaluation shows that the performance for the inliers is slowly decreasing during **IID** and **NEAR** splits, dropping suddenly just before the **FAR** split, showing how the language model fails to recognize inliers once it moves further from the train data.

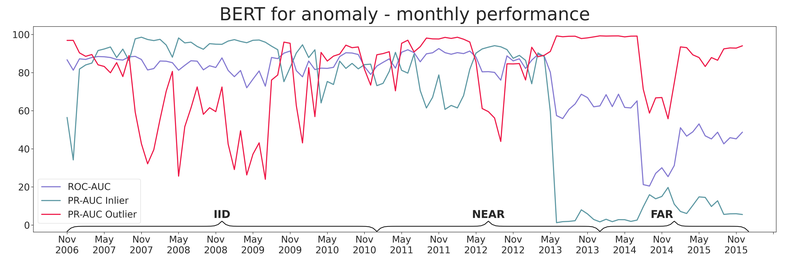

Figure 5. BERT for anomaly, evaluated on months: ROC-AUC, PR-AUC inliers, and PR-AUC outliers.

Addressing the shifted data

We next compare the performance of a BERT model in 3 training regimes: **iid**, **finetune**, and **knowledge distillation** and observe that the best ROC-AUC is achieved by the final **distilled** model, outperforming **iid** and **finetune** by over 3% on average.

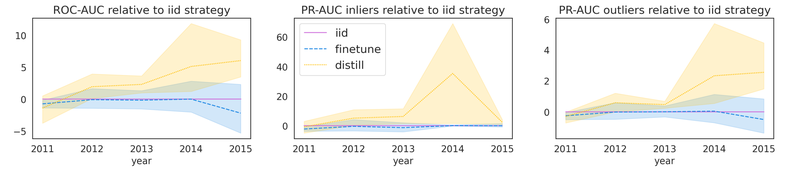

Figure 6. ROC-AUC, PR-AUC-in, PR-AUC-out for Finetune and Distill strategies, relative to the iid.