

## 異常検知のための Few-shot 表現学習

山本綾華<sup>†</sup>東京理科大学大学院経営学専攻<sup>†</sup>安藤晋<sup>‡</sup>東京理科大学大学院経営学専攻<sup>‡</sup>

## 1 はじめに

実用的な異常検知問題は正常例が複数のクラスから構成される場合があり、クラス間の相違が検知をより困難な問題とする。本研究では、少数のラベル付正常例を用い、事前学習したモデルを再訓練することでこれに対処する。これは半教師付きの深層異常検知 [3] 手法に類する設定だが、異常例が与えられない点で異なる。

われわれは各クラスから少数の代表例のみが与えられる条件下で、クラス内類似性とクラス間距離に関する損失を情報ボトルネック原理に基づくレート歪み損失 [4, 5] として定義し、さらにマージン最大化損失の形式で定式化する。

## 2 関連研究

## 2.1 情報ボトルネック原理

情報ボトルネック [4] は、入力変数  $X$  の符号表現  $Z$  を学習する原理であり、 $Y$  の予測に寄与しない情報を除いて圧縮するためにレート歪み損失  $\mathcal{L}$  を最小化する問題を解く。

$$\mathcal{L} = I(X; Z) - \beta I(Z; Y) \quad (1)$$

$\mathcal{L}$  の第 1 項は  $Z$  の持つ情報量であるが、 $P(z|x)$  を  $x$  の射影  $f(x)$  を中心とする等方正規分布とすると、定数項を除いて以下のような全ての例間の深層空間での正規化二乗距離の和

の形式で表される [5]。

$$I(X; Z) = \frac{1}{N^2} \sum \sum \left( \frac{\|z - f(x_i)\|^2}{2\sigma_k^2} + \log \sigma_k^2 \right)$$

一方、第 2 項は  $Z$  が  $Y$  について持つ情報量であり、 $P(z|y)$  を正規分布に従う生成確率とすると深層空間のクラス中心からの正規化二乗距離の和の形式で表される。

2つの項を同時に最小化することでクラス内で類似度が増大しつつ、異なるクラス間では距離を増大させる埋め込み  $f$  の学習が促される。本研究ではこれを深層空間で異常例が既知のクラスと重複を避けるために用いる。

## 3 提案手法

提案手法では GAN を教師無しで事前学習し、その判別器の埋め込み層を少数のラベル付き正常例を用い再学習する。 $\mathcal{X}_t$  を  $N$  クラス各  $K$  例からなるラベル付き例集合とする。 $\mathcal{X}_t = \{p_j, y_j\}$ 、ただし、 $y_j \in \{1, \dots, N\}$  とする。判別器の埋め込み関数を  $f$  とする。

2.1 節で示した情報ボトルネック損失は埋め込み空間で正規分布を仮定するが、異常検知ではクラス境界付近の例が精度に強く影響するため、本研究では以下のようにマージン最大化形式の損失を定義する。まず、異なるクラスのペア  $\{j, k : y_j \neq y_k\}$  について  $R_{dis}$  未満の距離にあるものに以下のペナルティを課し、

$$\max \{0, R_{dis}^2 - \|f(p_j; W) - f(p_k; W)\|\}$$

閾値  $R_{dis}$  は、ミニバッチのクラス内距離の平均と 2 倍の標準偏差の合計の最大値とする。

また、同クラスのペア  $\{j, k : y_j = y_k\}$  についてミニバッチのクラス内距離の中央値  $R_{sim}$  を

Anomaly Detection with Few-shot Learning over Normal Classes

<sup>†</sup> Ayaka Yamamoto, Tokyo University of Science

<sup>‡</sup> Shin Ando, Tokyo University of Science

超えたペアの損失を下式の通り定める。

$$\max \{0, \|f(p_j; W) - f(p_k; W)\| - R_{\text{sim}}^2 \}$$

## 4 実験

### 4.1 設定

数値実験は Fashion-MNIST, CIFAR10 の画像データセットを用い, 1つのクラスを異常, 他9クラスを正常例とした検出問題を扱う. EGBAD[6] および GANomaly[3] をベースラインとして提案手法と合わせて AUROC および AUPRC を評価する. 提案手法は事前学習で全ての正常クラス例をラベル無しで用い, 再学習では  $K = 20$  のラベル付き正常クラス代表例を用いる. ベースラインの2手法はそれぞれ教師無し, 半教師付きの学習を行う.

### 4.2 結果

図1, 2にベースラインと提案手法の比較を示す. 提案手法は XDD と表記する. 横軸は異常クラスの番号を示す. 縦軸はそれぞれ AUROC, AUPRC を示す. これらの結果は, 提案手法が教師無し深層異常検知手法を上回り, 半教師付手法とも同等以上の性能を持つことを示す.

## 5 おわりに

本研究では少数のラベル付き正常例を用いた深層異常検出モデルの学習手法を提案した. 数値実験では画像データにおいて訓練時に与えられていない異常例を検知する問題で, 従来の教師無し手法および異常例を用いる半教師付き学習を上回る検出指標を示した. また, 学習した深層表現空間の視覚的分析でも正常クラス間の分離を向上することを確認した.

## 参考文献

- [1] Ruff et Al. Deep one-class classification. *ICML2017, Proceedings of Machine Learning Research*, vol. 80, pp. 4393–4402, 2018.

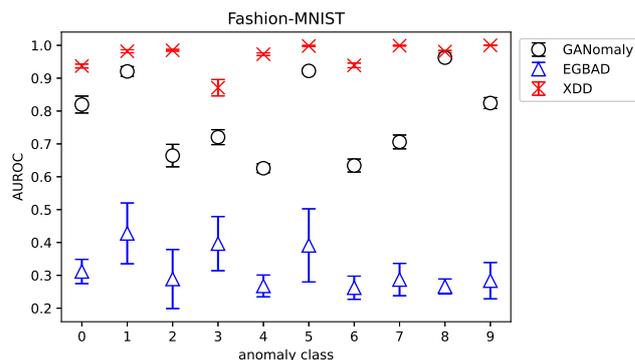


図1 AUROC の比較 (Fashion MNIST)

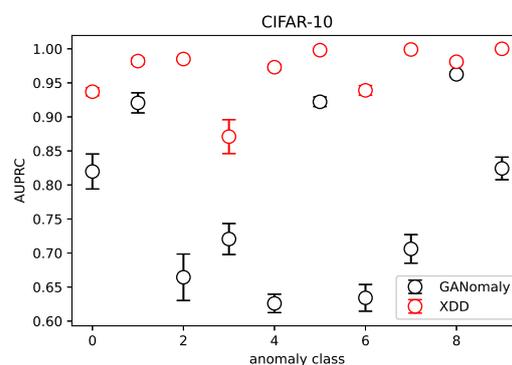


図2 AUPRC の比較 (CIFAR10)

- [2] Wang et al. Generalizing from a few examples: A survey on few-shot learning. *ACM Comput. Surv.*, 53(3), June 2020.
- [3] Ruff et Al. Deep semi-supervised anomaly detection. In *8th International Conference on Learning Representations, ICLR 2020*.
- [4] Tishby et al. The information bottleneck method, 2000. cite arxiv:physics/0004057.
- [5] Ando, Shin. Deep representation learning with information-theoretic Loss. *CoRR*, abs/2111.12950, 2021.
- [6] Zenati et al. Efficient gan-based anomaly detection. *CoRR*, abs/1802.06222, 2018.