

2020/10/20 IBISML研究会

集合データの共通特徴抽出を利用したFew-shot異常検知

○佐藤一輝¹・中田智史²・松原 崇³・上原邦昭⁴

¹ 神戸大学 大学院システム情報学研究科, ² 株式会社地球快適化インスティテュート, ³ 大阪大学 大学院基礎工学研究科, ⁴ 大阪学院大学 経営学部

発表内容

- 背景
- 関連研究
- 提案手法
- 評価実験と結果
- まとめ

背景 | 異常検知

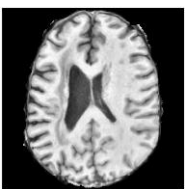
- 異常検知とは
 - 大多数の正常データとは異なる特徴を持つ異常データを検出するタスク
- 画像データにおける実用例

- 工業製品の欠陥検出



ビン底のヒビ割れ検出
[Bergmann+, 2019]

- 医用画像(MRI,CT等)を利用した医療診断の補助



ヒトの頭部MRIを用いた脳梗塞の発見
[Sato+, 2019]

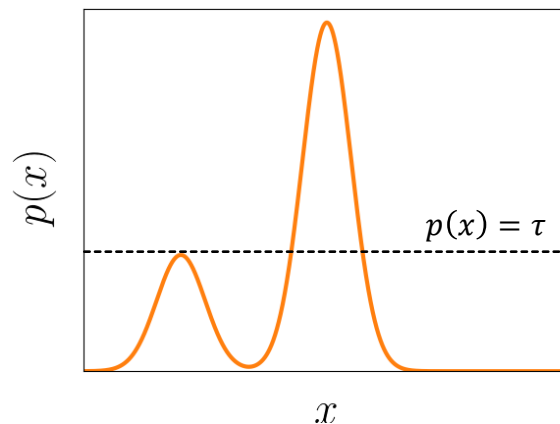
背景 | 異常検知の学習方法

- 異常検知の学習方法
 - 教師あり学習
 - データと正常/異常ラベルの対応関係を学習する方法
 - 未知のタイプの異常は基本的に検出不可
 - 教師データの収集とラベリングに手間が掛かる
 - 教師なし学習
 - 正常な画像の傾向や特徴を学習する方法
 - 未知のタイプの異常に対応可能
 - ラベル付けの手間が不要
- 異常検知の問題設定に適している

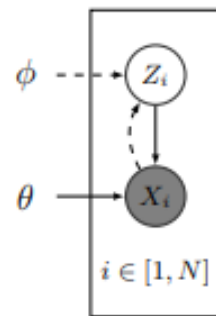
背景 | 異常検知と生成モデル

- 異常検知と生成モデル
 - 正常データの特徴表現を学習し, そこから離れた特徴を持つデータを異常とみなす
 - 正常データを用いて生成モデル $p_{\theta}(x)$ を学習する
 - 基本的には, 学習用データに対する尤度を最大化するようにパラメータを学習
 - 画像のような高次元のデータに対しては, ニューラルネットを利用した深層生成モデルが用いられる
 - 与えられたデータ x_i の「異常度」は, しばしば確率密度 $p_{\hat{\theta}}(x_i)$ で計られる

x_i is classified as $\begin{cases} \text{正常} & \text{if } p_{\hat{\theta}}(x_i) \geq \tau; \\ \text{異常} & \text{otherwise.} \end{cases}$



関連研究 | VAE



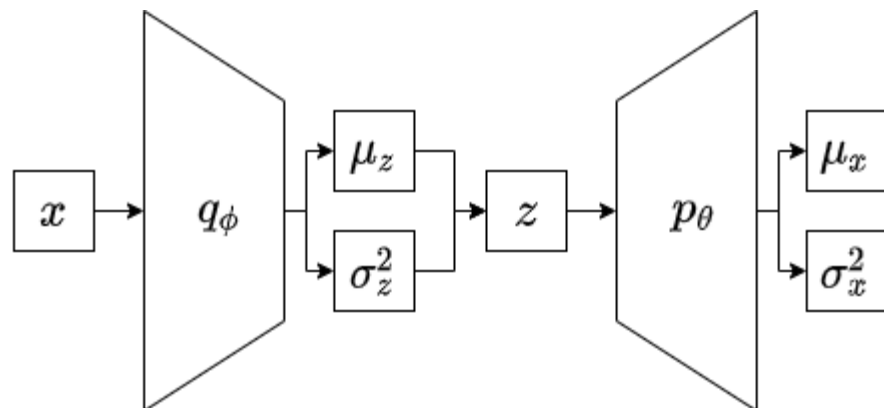
[Kingma+, 2014]

- Variational Auto-Encoder (VAE) [Kingma+, ICLR2014]

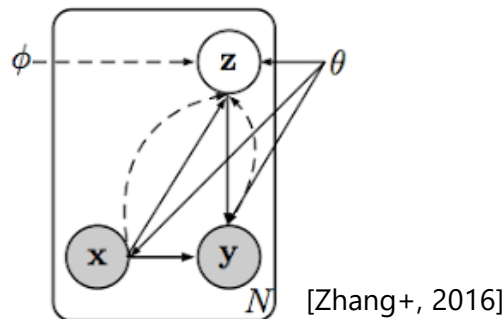
- 対数尤度 $\log p_\theta(x)$ の変分下界 :

$$\log p_\theta(x) \geq \underbrace{-D_{KL}(q_\phi(z|x) \parallel p(z))}_{\text{正則化項}} + \underbrace{\mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)]}_{\text{再構成誤差}} \triangleq -\mathcal{L}(x)$$

- $-\mathcal{L}(x)$ を最大化するよう $p_\theta(x|z)$, $q_\phi(z|x)$ のパラメータをニューラルネットで学習する
 - 事前分布 $p(z)$ を例えば多変量標準正規分布 $\mathcal{N}(\mathbf{0}, I)$ とすると, 正則化項は容易に計算可能
 - 再構成誤差はモンテカルロ近似により計算可能
- 異常度として, 近似された確率密度の代わりに再構成誤差がよく用いられる



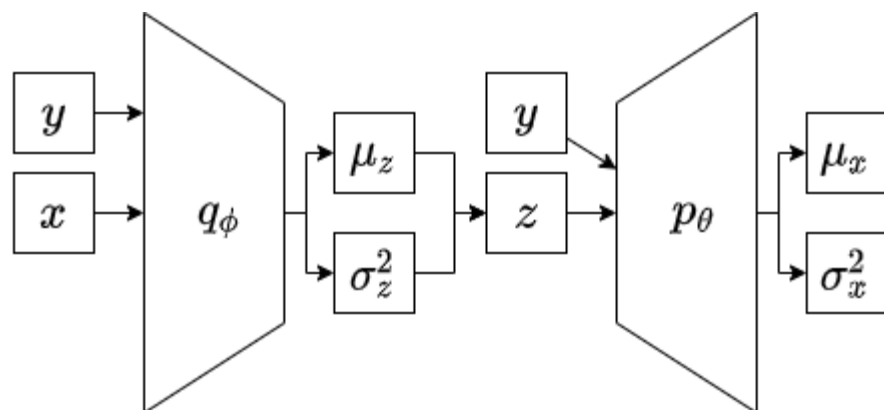
関連研究 | Conditional VAE



- Conditional VAE [Kingma+, ICLR2014]
 - ラベル y を与えて条件付き生成を行うVAEの拡張

$$\log p_{\theta}(x|y) \geq \underbrace{-D_{KL}(q_{\phi}(z|x, y) \parallel p(z))}_{\text{正則化項}} + \underbrace{\mathbb{E}_{q_{\phi}(z|x, y)}[\log p_{\theta}(x|z, y)]}_{\text{再構成誤差}} \triangleq \mathcal{L}(x, y)$$

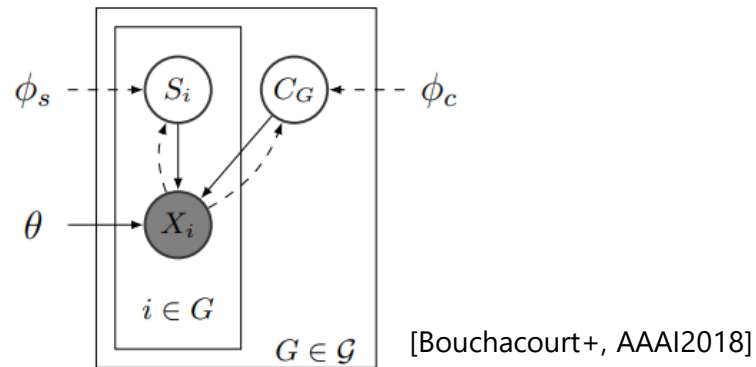
- 異常検知に既知のクラスに関する情報を活用したい場合には用いることができる [Pol+, 2018]



提案手法 | Few-shot異常検知

- Few-shot 異常検知
 - 想定する状況
 - 工業製品の欠陥検出で、新型が出るたびにモデルの再学習が必要になることを避けたい場合
 - モデルの学習時と実運用時でデータの取得環境に差異がある場合 ex. 医療データ
 - 問題設定
 - 学習時とは異なる未知のクラス(i.e. 新型, 別の環境)のデータから異常を検出する
 - 同一のクラスに属する N_G 個のデータの集合 $G = \{x_i\}_{i=1}^{N_G}$ を入力として補助的に用いても良い
 - 実験では $N_G = 10$ とした
 - 未知クラスに対して分類を行うFew-shot分類タスクの異常検知バージョンという意味で、この問題設定を**Few-shot 異常検知**と呼ぶことにする
 - 以降ではFew-shot異常検知で使える生成モデルについて考える

提案手法 | MLVAE

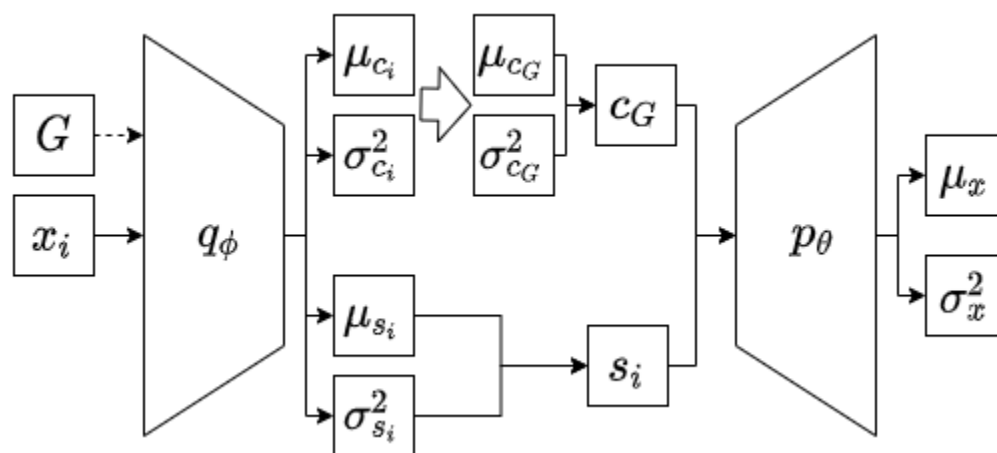


- Multi-Level VAE (MLVAE) [Bouchacourt+, AAAI2018]
 - 集合 $G = \{x_i\}_{i=1}^{N_G}$ についての生成モデル $p_\theta(G)$ を考える
 - 潜在変数として, **集合で共通**の c_G と **データ固有**の s_i を持つものとする
 - VAE同様, 以下の下界 $\mathcal{L}(G)$ を最大化する

$$\log p_\theta(G) \geq \underbrace{-D_{KL}(q_\phi(c_G|G) \parallel p(c_G))}_{c_G \text{の正則化項}} - \underbrace{\sum_{x_i \in G} D_{KL}(q_\phi(s_i|x_i) \parallel p(s_i))}_{s_i \text{の正則化項}} + \underbrace{\sum_{x_i \in G} \mathbb{E}_{q_\phi(c_G, s_i|x)} [\log p_\theta(x|c_G, s_i)]}_{\text{再構成誤差}} \triangleq \mathcal{L}(G)$$

$$q_\phi(c_G|G) \propto \prod_{x_i \in G} q_\phi(c_{x_i}|x_i)$$

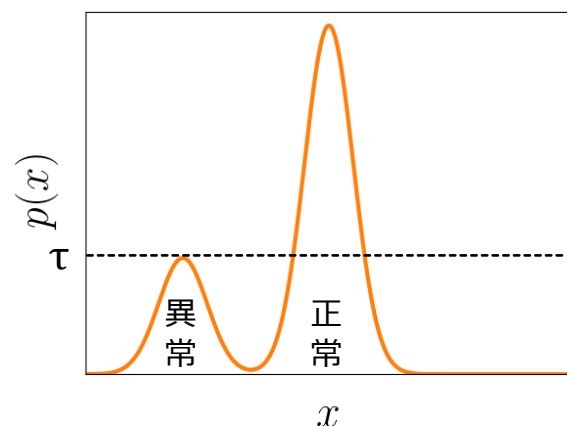
c_G の分布 $q_\phi(c_G|G)$ は,
集合に含まれるデータから得た $q_\phi(c_{x_i}|x_i)$ の積で表す



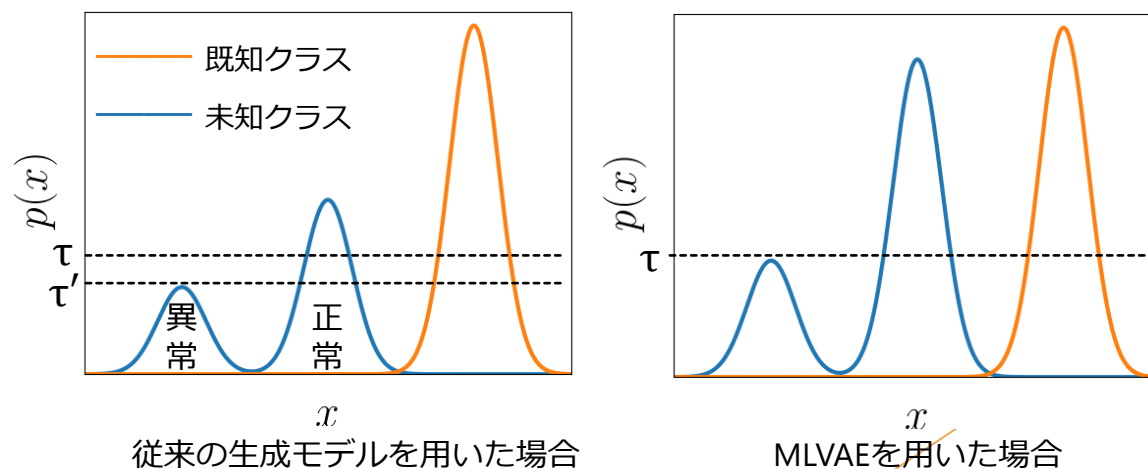
提案手法 | Few-shot異常検知への適用

- MLVAEを用いたFew-shot異常検知
 - VAEと同様に, MLVAEの再構成誤差を異常度として用いることにする
- 期待される効果
 - 未知のクラスに対する偽陽性を抑制することで, 性能が向上
 - 通常の異常検知の設定から閾値の流用が可能
 - つまり, 未知クラスのラベル付きデータがなくても, 既知クラスのデータを利用して閾値が決定できる

通常の異常検知



Few-shot異常検知



評価実験 | トイデータ

- 実験目的
 - MLVAEに期待される**偽陽性の抑制**による性能向上, および**閾値の流用可能性**について確かめる
- 使用したデータセット① (トイデータ)
 - データセット
 - SVHN : 家屋番号(0-9)の画像
 - CIFAR-10 : オブジェクト(乗り物や動物)の画像
 - それぞれ90度回転させた画像を異常とみなした

	既知クラス (学習用データ)									未知クラス	
										正常	異常
SVHN											
CIFAR-10	airplane	automobile	bird	cat	deer	dog	frog	ship	truck	horse	

10クラス中, 1クラスを除いて学習

結果 | トイデータ

- 性能比較

- 評価指標：ROC-AUC（高いほど良い）
- Few-shot異常検知に加え，既知クラスを扱う通常の異常検知の性能についても検証
- データ集合 G として，同一のクラスからではなくランダムなクラスから選ばれたデータを用いた場合と比較

SVHN			CIFAR-10		
Model	未知クラス	既知クラス	Model	未知クラス	既知クラス
AE	0.644	0.656	AE	0.601	0.606
VAE	0.700	0.717	VAE	0.605	0.616
MLVAE (ランダム)	0.707	0.717	MLVAE (ランダム)	0.633	0.614
MLVAE	0.723	0.745	MLVAE	0.640	0.642

結果 | トイデータ

- 閾値の流用可能性の検証
 - 性能評価指標：Youden's index（陽性率-偽陽性率）
 - 以下の2つの条件で， Few-shot異常検知の性能比較を行なった
 - (a) 未知クラスの正常/異常データを用いて閾値を決める場合
 - (b) 既知クラスの正常/異常データを用いて閾値を決める場合
 - 現実的な状況(b)において， どれだけ理想的な状況(a)に近い性能が得られるかを評価

SVHN

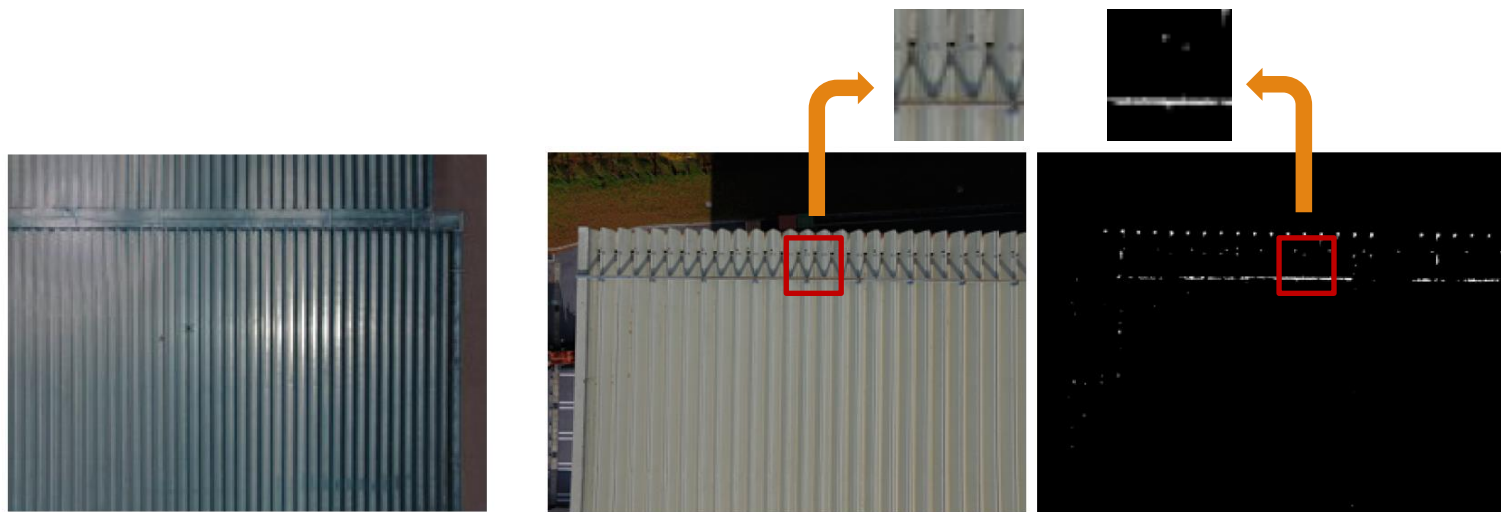
Model	(a) 未知	(b) 既知	(a)-(b)
AE	0.210	0.195	0.015
VAE	0.294	0.279	0.015
MLVAE	0.332	0.325	0.007

CIFAR-10

Model	(a) 未知	(b) 既知	(a)-(b)
AE	0.163	0.121	0.042
VAE	0.181	0.150	0.031
MLVAE	0.042	0.031	0.023

評価実験 | 工場屋根の空撮画像

- 使用したデータセット② (実データ)
 - 工場屋根の空撮画像データセット
 - 学習用に正常な屋根の画像を759枚, テスト用に異常(錆び)を含む屋根の画像を25枚含む
 - 解像度3000×4000の画像を100×100のパッチに分割
 - 「そのパッチがどの工場の屋根から得られたものか」をクラスとして想定
→学習時とは異なる工場屋根から異常を検出する性能を評価した



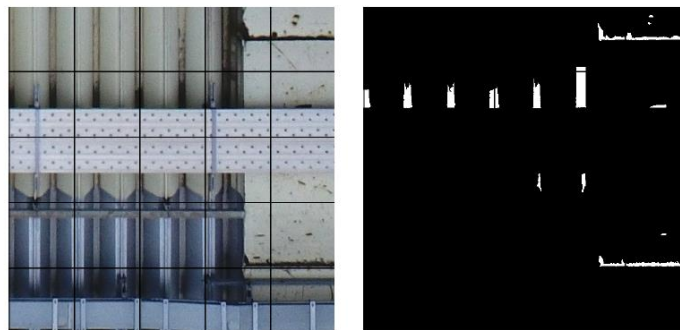
結果 | 工場屋根の空撮画像

- 性能の定量比較
 - 評価指標 : IoU@TPR{95,90,50}, ROC-AUC

Model	IoU@TPR95	IoU@TPR90	IoU@TPR50	ROC-AUC
AE	0.031	0.102	0.268	0.901
VAE	0.051	0.104	0.235	0.948
MLVAE	0.130	0.197	0.294	0.972

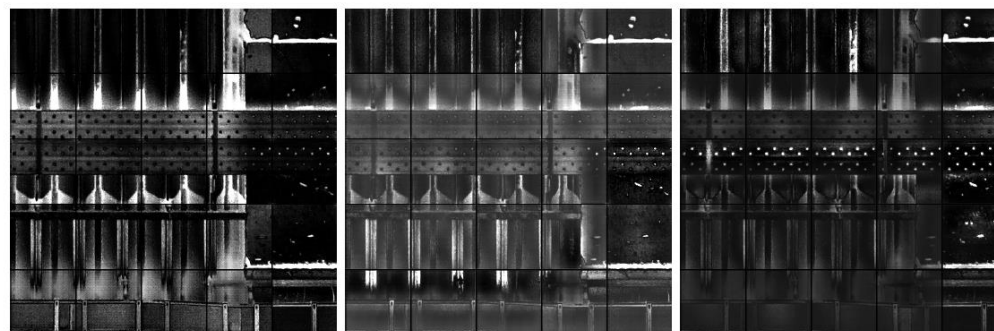
結果 | 工場屋根の空撮画像

- 異常度の可視化
 - [0,255]に正規化, 白色に近づくほど異常度が高いことを示す



sample

ground truth



AE, 二乗誤差

VAE, 再構成誤差

MLVAE, 再構成誤差

まとめ

- まとめ
 - 生成モデルを用いた異常検知では、学習時とは異なる未知クラスのデータが偽陽性になり得る
 - 学習時とは異なるクラスのデータに対する異常検知の問題を**Few-shot異常検知**と呼ぶ
 - この問題設定のもとで、同じクラスのデータの集合から共通の特徴を抽出する生成モデルのひとつであるMLVAEを用いることで、性能が向上することを確認した
- 今後の方針
 - 集合内で共通の特徴情報がデータ固有の潜在変数 s_i に流入することを防ぐよう、相互情報量最小化等を利用したモデルの拡張を検討