

令和 6 年 6 月 20 日現在

機関番号：53203

研究種目：基盤研究(C)（一般）

研究期間：2021～2023

課題番号：21K12022

研究課題名（和文）疑似データ生成による異常検知性能向上と個人認証への応用

研究課題名（英文）Improvement of Anomaly detection and personal authentication ability by pseudo-data generation

研究代表者

北村 拓也（Kitamura, Takuya）

富山高等専門学校・その他部局等・准教授

研究者番号：40611918

交付決定額（研究期間全体）：（直接経費） 2,800,000 円

研究成果の概要（和文）：本研究では、異常検知問題において、正常データに比べて異常データの収集が困難であることに着目した。そこで、正常データから疑似異常データを生成し、それらのデータを学習に用いることによる異常検知性能向上を目的としている。疑似異常データ生成には、代表的な生成モデルである敵対生成ネットワーク（GAN）と変分自己符号化器（VAE）を用いており、生成した疑似異常データを学習に用いることによって汎化能力の向上を確認できた。また、空中筆記動作を用いた個人認証システムにも導入し、なりすまし実験を実施し、その有効性を確認した。

研究成果の学術的意義や社会的意義

異常検知問題は工場の生産ラインや医療の診断現場や個人認証などのように、様々な分野にて存在しており、国内外問わず、広く研究なされている。また、異常検知問題に対して、疑似データを生成することによる異常検知性能向上の研究は、これまでも広く実施されている。疑似データ生成に広く用いられる手法として、VAEやGANが挙げられるが、本研究のようにこれらを同時に考慮するシステムは少ない。また、生成モデルのみでなく、部分空間法やSVMなどの識別モデルを適用することにより、より高い精度の疑似データの生成と異常検知性能の向上を期待できると考えられ、新規性と有効性の高い研究であると考えられる。

研究成果の概要（英文）：In this study, I focus on the difficulty of collecting abnormal data compared to normal data in anomaly detection problems. Therefore, we aim to improve anomaly detection performance by generating pseudo-abnormal data from only normal data. I use adversarial generative networks (GAN) and variational autoencoders (VAE) for pseudo-abnormal data generation, and confirmed improvement in generalization capability by using the generated pseudo-abnormal data for training. Furthermore, we introduced this approach to a personal authentication system using aerial writing motion. I conducted impersonation experiments, and confirmed its effectiveness.

研究分野：機械学習

キーワード：異常検知 疑似異常データ 敵対生成ネットワーク 変分自己符号化器 サポートベクトルマシン

1. 研究開始当初の背景

通常、異常検知問題において、異常となる割合は正常となる割合に比べて、低くなる傾向にある。そのため、異常検知問題について学習を行わせる際、学習データは不均衡データとなりやすく、異常データが全くない場合も多く存在する。不均衡データであれば、分類精度（認識率）を指標として識別基準を決定したとしても、正常データをより重要視した基準となることは明らかである。そのため、AUCなどの再現率と適合率を考慮した指標から決定するべきである。しかしながら、学習段階では異常データが少ないため、AUCの予測は難しく、異常データを増加させることが重要となる。正常データのみでの問題では、1クラス識別器または自己符号化器（AE）などの生成モデルを用いて学習を行い、異常検知が可能である。しかしながら、少数であったとしても、異常データを考慮した識別モデルに比べれば、精度は低くなる傾向にある。また、個人認証問題では、対象とする人物であるか否かの二値問題であり、異常検知と同様に、学習段階では正常データのみ得られるケースが多い。さらに、システムへの登録の際、その正常データでさえ少数である場合が大半である。空中筆記動作などのように動的特徴量を対象とする問題において、データにばらつきが多く、少数の正常データのみでの学習では分類が困難となりやすい。これらのことより、異常検知問題においては、学習に用いる異常データ、前述のような個人認証問題においては、学習に用いる正常データを増加させることが重要となる。

2. 研究の目的

本研究では、疑似データを生成することによる異常検知性能向上と個人認証システムへの応用を目的とした。本研究では、以下の3パターンで、疑似データを生成する。

少数の異常データと多数の正常データを基にした疑似異常データの生成

多数の正常データのみを基にした疑似異常データの生成

少数の正常データを基にした疑似正常データの生成

ここで、では、正常データは多く取得できることが前提とした異常検知問題、では、異常データが得られない、かつ、正常データも少数しか得られない個人認証問題（本人かそうでないかの二値問題）を想定している。

3. 研究の方法

本研究では、大きく分けて、以下のような3種類の実験を実施した。

多数の正常データのみを基にした疑似異常データの生成

多数の正常データのみを学習データとして有する異常検知問題に対して、敵対生成ネットワーク（GAN）または変分AE（VAE）を用いて異常データを生成し、異常検知実験を実施し、従来手法との比較・評価を行う。このとき、AUCを評価基準として用いる。

大規模不均衡異常検知データにおける異常検知実験

少数の異常データと多数の正常データを有する異常検知問題に対して、一部の異常データを基にGANやVAEを用いて疑似異常データを生成し、異常検知実験を実施し、従来手法との比較・評価を行う。このとき、AUCを評価基準として用いる。

少数の正常データを基にした擬似正常データの生成

空中筆記動作を入力とした個人認証問題に対して、EfficientNetや準ハードマージンSVM（SH-SVM）を用いて擬似正常データを生成し、認証実験を実施する．ここで、学習で用いる空中筆記動作データ、1人につき数個データと少数の正常データのみとする．これまで、申請者の先行研究にて用いられてきたDPマッチングやSVMや部分空間法や部分空間法に基づくSVM（SS-SVM）などの手法との比較・評価を行う．このとき、AUCと認識率を評価基準として用いる．

4．研究成果

前述の - に関する実験における研究成果を以下に述べる．

多数の正常データのみを基にした擬似異常データの生成

異常検知問題において広く用いられるMVTecデータセットを用いてGANとVAEを用いて図1に示すような擬似異常データを生成した．このとき、異常データは部分的な欠損による異常が多いことから、あらかじめ正常データにマスクし、そのマスク画像を基にGANとVAEにより正常データに近づけるように学習することによって、図のような小さい欠損のあるデータとすることを可能とした．計算機実験では、従来手法に比べてAUCが15種類のデータセットで平均6%以上向上し、50%以上向上したデータセットも見受けられ、全体としても安定して高い検知性能を示した．

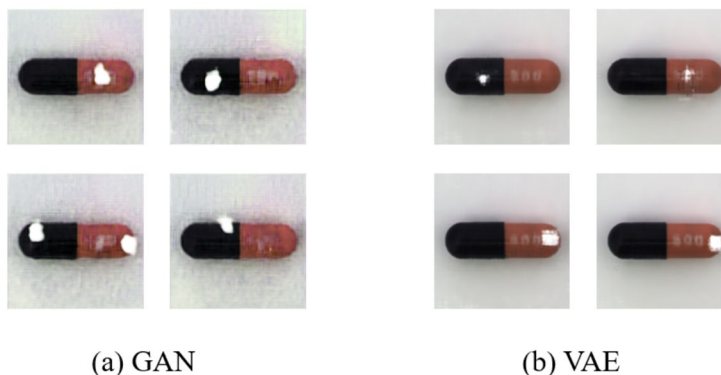


図1．生成モデルを利用した擬似異常データの生成

大規模不均衡異常検知データにおける異常検知実験

の実験に対して正常データによる学習の代わりに少数の異常データを用いて学習させ、擬似異常データを生成した．その結果、で示した擬似異常データとほぼ同様の擬似異常データとなり、少数の異常データを入れることによる性能向上が見られず、異常データが少数である場合においても正常データのみを用いた擬似異常データ生成手法が有効であることが確認できた．

少数の正常データを基にした擬似正常データの生成

計算機実験では、図1に示すようにLeapMotionと呼ばれるモーションキャプチャデバイスを用いて空中筆記動作を検知し、得られた少数のデータを正常データとして学習に用いた．このとき、被験者には自分の名前を30回書いてもらった．擬似異常データは畳み込みニューラルネットワークモデルのひとつである EfficientNet と SH-SVM を用いた．得られた擬似異常データの例を図 2 に示す．これらの擬似異常データと正常データを用いて学習を行うことにより、図 2 に示すように、正常データ（赤）と異常データ（青）を分類するような分類超平面が生成できたことが確認できた．

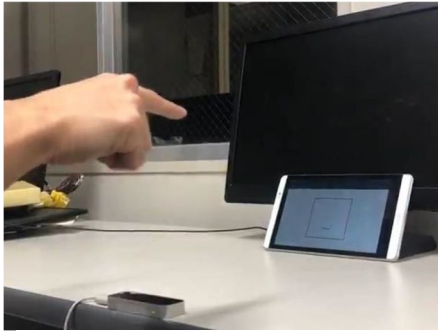


図 1 . 空中筆記動作の読み取り

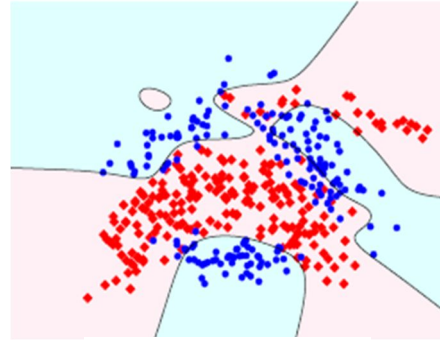


図 2 . 異常検知結果

5. 主な発表論文等

〔雑誌論文〕 計8件（うち査読付論文 2件／うち国際共著 0件／うちオープンアクセス 0件）

1. 著者名 川原啓翔, 北村拓也	4. 巻 123
2. 論文標題 ファインチューニング画像分類モデルによるEEGデータを用いたASD診断システム	5. 発行年 2024年
3. 雑誌名 電子情報通信学会技術研究報告	6. 最初と最後の頁 130-134
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 吉原沙江, 北村拓也	4. 巻 123
2. 論文標題 ファインチューニングモデルによるfMRIを用いたASD診断システムの評価	5. 発行年 2024年
3. 雑誌名 電子情報通信学会技術研究報告	6. 最初と最後の頁 135-140
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 大上雄大, 北村拓也	4. 巻 123
2. 論文標題 異常検知問題におけるマルチカーネルを用いた1クラス量み込みELM	5. 発行年 2023年
3. 雑誌名 電子情報通信学会技術研究報告	6. 最初と最後の頁 7-12
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 土尾崇太, 北村拓也	4. 巻 123
2. 論文標題 疑似異常データを用いたDSVDDによる異常検知システムの汎化能力向上	5. 発行年 2023年
3. 雑誌名 電子情報通信学会技術研究報告	6. 最初と最後の頁 25-30
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 田村明人, 北村拓也	4. 巻 123
2. 論文標題 HSI分類問題におけるPCAによる複合カーネルSVMの改良	5. 発行年 2023年
3. 雑誌名 電子情報通信学会技術研究報告	6. 最初と最後の頁 1-6
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 田村明人, 北村拓也	4. 巻 122
2. 論文標題 HSI分類問題における複合SVMの改良	5. 発行年 2023年
3. 雑誌名 電子情報通信学会技術研究報告	6. 最初と最後の頁 96-100
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Yoshida Takeshi, Kitamura Takuya	4. 巻 12894
2. 論文標題 Semi-Hard Margin Support Vector Machines for Personal Authentication with an Aerial Signature Motion	5. 発行年 2021年
3. 雑誌名 Lecture Notes in Computer Science (Proceedings of 30th International Conference on Artificial Neural Networks)	6. 最初と最後の頁 333 ~ 344
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-030-86380-7_27	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Takaoka Shogo, Kitamura Takuya, Suzuki Aiga, Murakawa Masahiro	4. 巻 -
2. 論文標題 Initialization Method of Batch Uniformization Auto Encoder by Principal Component Analysis	5. 発行年 2021年
3. 雑誌名 Proceengs of IEEE Symposium Series on Computational Intelligence (SSCI2021)	6. 最初と最後の頁 1-6
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/SSCI50451.2021.9660064	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計10件（うち招待講演 0件 / うち国際学会 2件）

1．発表者名 土尾崇太，北村拓也
2．発表標題 異常検知問題における深層SVDDのデータ選択による学習効率化
3．学会等名 第50回知能システムシンポジウム
4．発表年 2023年

1．発表者名 大上雄大，北村拓也
2．発表標題 1クラス畳み込みELMによる異常検知システム
3．学会等名 第50回知能システムシンポジウム
4．発表年 2023年

1．発表者名 大上雄大，北村拓也
2．発表標題 1クラス畳み込みELMによる異常検知
3．学会等名 第25回情報論的学習理論ワークショップ（IBIS2022）
4．発表年 2022年

1．発表者名 土尾崇太，北村拓也
2．発表標題 深層サポートベクトルデータ記述法（DSVDD）におけるファインチューニングの効率化と異常検知への適用
3．学会等名 第25回情報論的学習理論ワークショップ（IBIS2022）
4．発表年 2022年

1．発表者名 田村明人，北村拓也
2．発表標題 ハイパースペクトル画像分類における畳み込みSVM
3．学会等名 第25回情報論的学習理論ワークショップ (IBIS2022)
4．発表年 2022年

1．発表者名 石倉 優資，北村 拓也
2．発表標題 1クラス多層多重カーネル ELM の開発
3．学会等名 第34回自立分散シンポジウム
4．発表年 2022年

1．発表者名 大石 悠貴，北村 拓也
2．発表標題 部分空間法におけるマージン最大化を考慮した原点移動
3．学会等名 第49回知能システムシンポジウム
4．発表年 2022年

1．発表者名 本田 一期，北村 拓也
2．発表標題 センテンス間の連続性を考慮した SNS における感情分析
3．学会等名 第49回知能システムシンポジウム
4．発表年 2022年

1 . 発表者名 Yoshida Takeshi , Kitamura Takuya
2 . 発表標題 Semi-Hard Margin Support Vector Machines for Personal Authentication with an Aerial Signature Motion
3 . 学会等名 International Conference on Artificial Neural Networks(ICANN2021) (国際学会)
4 . 発表年 2021年

1 . 発表者名 Takaoka Shogo , Kitamura Takuya , Suzuki Aiga , Murakawa Masahiro
2 . 発表標題 Initialization Method of Batch Uniformization Auto Encoder by Principal Component Analysis
3 . 学会等名 IEEE Symposium Series on Computational Intelligence (SSCI2021) (国際学会)
4 . 発表年 2021年

〔 図書 〕 計1件

1 . 著者名 北村拓也 , 他69名	4 . 発行年 2022年
2 . 出版社 技術情報協会	5 . 総ページ数 560
3 . 書名 機械学習・ディープラーニングによる “ 異常検知 ” 技術と活用事例集 (担当範囲： 1 章第 5 節 サポートベクターマシンの基礎と異常検知 , 故障予測への活用)	

〔 産業財産権 〕

〔 その他 〕

-

6 . 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	-------------------------------	-------------------------	----

7 . 科研費を使用して開催した国際研究集会

〔 国際研究集会 〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------