

令和 5 年 6 月 26 日現在

機関番号：34315

研究種目：基盤研究(C) (一般)

研究期間：2019～2022

課題番号：19K12164

研究課題名(和文) 深層生成モデルによる潜在空間を用いたデータ拡張や転移学習手法の構築

研究課題名(英文) Data augmentation and domain adaptation using the latent space of the deep generative model

研究代表者

西川 郁子(Nishikawa, Ikuko)

立命館大学・情報理工学部・教授

研究者番号：90212117

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：深層生成モデルの潜在空間の利用として、主にドメイン適応と教師なし異常検知に取り組んだ。潜在空間の獲得には、符号化器と生成器を同時に獲得する自己符号化器や変分自己符号化器のほか、敵対的学習で生成器を獲得後に得られた潜在空間への符号化器を獲得する手法、複数の符号化器、生成器、分類器を組み合わせることでタスク固有空間とドメイン固有空間に分割して獲得する手法などを新たに提案した。それらを用いて、分類タスクにおける一般化ドメイン適応手法の提案、教師なし異常検知による生物データの逆遺伝学的アプローチなどを行った。

研究成果の学術的意義や社会的意義

生成モデルが、画像や音声、言語など多様なデータ形式に対して、精度の高いデータを出力できるのは、データをその実世界での姿に依存しない形で抽出し表現できているからです。その表現を、実世界での姿に依らないデータの本質だと期待して、潜在表現と呼んでいます。対象とするデータに対して良い潜在表現を獲得する方法を提案し、潜在空間上でのデータ間の相互関係やデータ分布を使うことで、あるデータセットでの学習結果を他のデータセットに転用して利用できるドメイン適応や、異常データ検知器の構築に正常データだけを用いる学習などを、高精度で実現しました。

研究成果の概要(英文)：The conducted research on the effective utilization of the latent space obtained by deep generative model has mainly focused on the universal domain adaptation and unsupervised anomaly detection. The methods of acquire the latent space include not only the simple simultaneous acquisition of the encoder and the generator by auto-encoder or variational auto-encoder, but also originally proposed networks: a combination of an generator obtained by the adversarial training followed by the encoder to its input space for the unsupervised anomaly detection, and a combination of multiple encoders, decoders, and a classifier to obtain a pair of task specific latent space and domain specific latent space for the universal domain adaptation. Then, a novel method using class-wise discriminators is proposed for universal domain adaptation for the classification task. Moreover, unsupervised anomaly detection is applied to some biological dynamical data as an approach of inverse genomics.

研究分野：機械学習・最適化

キーワード：深層生成モデル 潜在空間 異常検知 ドメイン適応 データ拡張 エントロピー

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

認識器の特徴量空間を用いて、認識の逆関数としてのデータ生成器が提案され始め、研究代表者らは制約付きボルツマンマシンや敵対的学習による生成器で獲得した潜在空間を用いて、正常データ多様体からの逸脱に基づいて、データの異常検知と正常データへの最小修正を実施し、さらに、生成データを操作可能な潜在空間として表現の纏れ解消を試みていた。それらを一般化し、有効な潜在表現の設計方法と、それを用いた課題解決を目指して、ドメイン適応法、異常検知法の開発と評価を計画した。

2. 研究の目的

(1) 分類などの学習タスクに必要なかつ十分な潜在表現の獲得を目指す。もしも、データを再構築できる潜在表現が、タスク特徴量と、タスクとは独立したドメインなどの特徴量の、2つの分割された潜在空間として獲得可能ならば、前者のタスク特徴空間を用いてドメイン不変な認識器が得られる。しかし、個別ドメインのデータのみからは、完全な分割表現を得ることは困難なためにドメインシフトが生じ、そのシフトを解消するドメイン適応の過程で、より高精度なタスク特徴空間へと修正される。そこで、タスク特徴とそれ以外の特徴に分割された潜在表現の獲得法と、ドメイン数やクラス数に依らないドメイン適応手法を開発する。異なるドメインのクラス集合の包括関係に応じて、クローズドセットからオープンセットへと拡張し、最終的にはユニバーサルなドメイン適応法を構築する。

(2) 潜在表現を用いた教師なし異常検知の実現と事例適用。正常と異常の2クラス分類タスクにおいて、正常データのみからその潜在表現とそこでの分布を獲得することで、そこからの逸脱量によってデータ異常を定量化し、逸脱の方向によって異常の種類を特徴付ける。

3. 研究の方法

(1) まず、分類タスクにおいて、タスク特徴とドメイン特徴に分割された潜在表現の獲得法を開発する。つぎにドメイン適応法として、クローズドセット条件下では、最小化目標である潜在空間上での分布間距離が重要であることから、距離尺度の設定を検討する。ユニバーサル条件下では、ドメイン固有クラスの判別が重要であることから、潜在空間各点での事後確率、あるいは、データ密度比を計測する。そこから、各データのエントロピーも算出し、分布を一致させる対象を選択することで、頑健なドメイン適応法を構築する。

(2) 提案法の有効性の検証には、公開データで類似手法との定量比較を行った上で、実データによる応用事例で成果を得る。ドメイン適応では、地図画像のセグメンテーション課題を、異常検知では、逆遺伝学に用いる生物動態データに対して形態異常と時間発展異常を検知する。

4. 研究成果

(1) 分類タスクにおけるタスク特徴空間とドメイン特徴空間の獲得法を構築し、それを用いた

疑似データを生成した。具体的な構成は、両者を統合すると元データが再構築されること、タスク特徴からはデータのドメインが推定できず、逆にドメイン特徴からはクラスが推定できないことを条件として損失関数で与え、両空間への符号化器と生成器、分類器の逐次学習により実現される。手書き数字の複数の公開データセットを用いて検証し、学習で得られた潜在空間を用いて、所望のクラスとドメインの組み合わせに対して、疑似的なデータが生成できることも確認した。ただ、複数の損失関数に対する係数が、生成データの質に影響すること、係数の調整はドメインに依存することも見られた。さらに、疑似データを用いたドメイン適応を引き続き検討している。

(2) ドメイン適応に用いる分布間距離尺度として、クローズドセット条件で、特に画像データに対する Wasserstein 距離の有効性を数値的に検証した。事例適用として、図形による定量的情報を担う画像として地図画像を対象に、適応元をベクトルデータから生成されたデジタル地図、適応先を複数スタイルの地図画像としてドメイン適応を行い、提案法の有効性を確認した。学習で得られた潜在空間を用いて、手書きイラスト地図に対する道路セグメンテーションを実施した。

(3) ドメイン固有クラスが存在するオープンセット条件下では、潜在空間の各点における各ドメイン分布の密度比を推定することで、ドメイン固有クラスの判別を実現した。分布密度の推定には、符号化器との敵対的学習においてドメイン判別器の損失関数に用いる距離尺度に K-L ダイバージェンスを用いると、判別器の出力がドメイン密度比に対応することに着目し、事後確率を高精度に算出し、ドメイン共通クラスのみで分布を一致させることを可能にした。ここで導入した密度比算出用の判別器は、適応に用いる判別器とは独立に設定する。

(4) 前項のドメイン判別器と同様に、クラス分布の密度比を推定するクラス判別器を設定することで、潜在空間の各点におけるクラス確率を算出し、それに基づくエントロピーを用いて、ユニバーサル条件下でのドメイン適応法を提案した。従来から用いられてきた分類器に基づくエントロピーでは、最大事後確率クラスになった後は分類器が学習しないのに対して、判別器を用いることでより高精度で事後確率を得られる。逆に、ドメイン適応の進行につれて判別器の過適合が生じることには留意を要する。共通クラスと固有クラスで、エントロピー分布に顕著な差が見られ、従来法と比較して高精度での適応が実験的にも確認できた。クラスごとに計測用と適応用の判別器を設定するため、クラス数に応じて計算量が増大する点、データ数が少ないと計測用判別器が早期に過適合する点を、引き続き検討している。

(5) 教師なし異常検知における指標に、データの再構築誤差のみでなく、潜在空間上の位置を用いた。適用事例として、正常データ群が一定の分布を示す生物データを対象とし、表現型異常の検知により遺伝子機能を推定する逆遺伝学の設定に着目して、個別遺伝子をロックアウトした個体に対する表現型の3次元動態データの異常を定量的に検知した。再構築誤差から形態異常を、潜在空間における時間発展のズレから時間異常を、それぞれ定量化した。遺伝子機能に対する既知の知見や、同じ公開データに対する生物学的異常検知結果とも比較し、遺伝子機能を新たに示唆する可能性を示した。

5. 主な発表論文等

〔雑誌論文〕 計2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 谷村亮介、東優貴、高野諒、西川郁子	4. 巻 36-7
2. 論文標題 階層的探索による二次元配置配線最適化	5. 発行年 2023年
3. 雑誌名 システム制御情報学会論文誌	6. 最初と最後の頁 212-219
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 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 甲野晴太、植田考哉、高野諒、西川郁子	4. 巻 35-5
2. 論文標題 敵対的学習によるクラス推定を利用した部分ドメイン適応	5. 発行年 2022年
3. 雑誌名 システム制御情報学会論文誌	6. 最初と最後の頁 101-108
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 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16件（うち招待講演 0件 / うち国際学会 9件）

1. 発表者名 Eisuke Ito, Takaya Ueda, Ryo Takano, Yukako Tohsato, Koji Kyoda, Shuichi Onami, and Ikuko Nishikawa
2. 発表標題 Phenotype anomaly detection for biological dynamics data using a deep generative model
3. 学会等名 31st International Conference on Artificial Neural Networks (ICANN2022) (国際学会)
4. 発表年 2022年

1. 発表者名 伊藤瑛南、植田考哉、高野諒、遠里由佳子、京田耕司、大浪修一、西川郁子
2. 発表標題 深層生成モデルの潜在空間を用いた生物動態データの異常検知
3. 学会等名 第66回システム制御情報学会 研究発表講演会(SCI'22)
4. 発表年 2022年

1. 発表者名 田淵裕貴, 双見京介, 高野諒, 西川郁子
2. 発表標題 アイウェア型赤外線距離センサレイを用いた視線ジェスチャ認識
3. 学会等名 第66回システム制御情報学会 研究発表講演会 (SCI '22)
4. 発表年 2022年

1. 発表者名 荒木章伍, 宮野尚哉, 西川郁子
2. 発表標題 線形モデルと非線形モデルを複合したモデルによる風向風速予測
3. 学会等名 第66回システム制御情報学会 研究発表講演会 (SCI '22)
4. 発表年 2022年

1. 発表者名 谷村亮介, 東優貴, 高野諒, 西川郁子
2. 発表標題 有望な部分解空間の絞り込みによる二段階最適化を用いた二次元配置最適化
3. 学会等名 第66回システム制御情報学会 研究発表講演会 (SCI '22)
4. 発表年 2022年

1. 発表者名 Takumi Oibayashi, Takaya Ueda, Yuki Kimura, Yukako Tohsato, and Ikuko Nishikawa
2. 発表標題 Phenotype Anomaly Detection in Early <i>C. elegans</i> Embryos by Variational Auto-Encoder
3. 学会等名 2021 IEEE 9th International Conference on Bioinformatics and Computational Biology (ICBCB) (国際学会)
4. 発表年 2021年

1. 発表者名 甲野晴太, 植田考哉, 西川郁子
2. 発表標題 敵対的学習による分布間差異の推定を利用した部分ドメイン適応
3. 学会等名 第65回システム制御情報学会研究発表講演会(SCI ' 21)
4. 発表年 2021年

1. 発表者名 追林拓光, 木村勇貴, 植田考哉, 西川郁子
2. 発表標題 画像の構造的類似度を用いた自己符号化器による異常領域検出
3. 学会等名 第65回システム制御情報学会研究発表講演会(SCI ' 21)
4. 発表年 2021年

1. 発表者名 東優貴, 利根大輝, 山根悠, 高野諒, 西川郁子
2. 発表標題 二階層のメタヒューリスティクスによる電子機器と配線経路の二次元最適化
3. 学会等名 第65回システム制御情報学会研究発表講演会(SCI ' 21)
4. 発表年 2021年

1. 発表者名 Takaya Ueda, Yukako Tohsato and Ikuko Nishikawa
2. 発表標題 Temporal Anomaly Detection by Deep Generative Models with Applications to Biological Data
3. 学会等名 29th International Conference on Artificial Neural Networks, ICANN2020 (国際学会)
4. 発表年 2020年

1. 発表者名 Seita Kono, Takaya Ueda, Enrique Arriaga-Varela and Ikuko Nishikawa
2. 発表標題 Wasserstein Distance-Based Domain Adaptation and Its Application to Road Segmentation
3. 学会等名 International Joint Conference on Neural Networks 2021, IJCNN2021 (国際学会)
4. 発表年 2021年

1. 発表者名 Takaya Ueda and Ikuko Nishikawa
2. 発表標題 Latent space decomposition into task-specific and domain-specific subspaces for domain adaptation
3. 学会等名 IEEE World Congress on Computational Intelligence 2020 (IEEE WCCI 2020) (国際学会)
4. 発表年 2020年

1. 発表者名 Takumi Oibayashi, Takaya Ueda, Yuki Kimura, Yukako Tohsato, and Ikuko Nishikawa
2. 発表標題 Phenotype Anomaly Detection in Early <i>C. elegans</i> Embryos by Variational Auto-Encoder
3. 学会等名 2020 IEEE 8th International Conference on Bioinformatics and Computational Biology (ICBCB 2020) (国際学会)
4. 発表年 2020年

1. 発表者名 Takaya Ueda, Masataka Seo, Yukako Tohsato, Ikuko Nishikawa
2. 発表標題 Analysis of Time Series Anomalies Using Causal InfoGAN and Its Application to Biological Data
3. 学会等名 The International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD) 2019 (国際学会)
4. 発表年 2019年

1. 発表者名 Yuki Kimura, Takaya Ueda, Seo Masataka, Yukako Tohsato, Ikuko Nishikawa
2. 発表標題 Characterizing Phenotype Abnormality by Variational Auto Encoder
3. 学会等名 The International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD) 2019 (国際学会)
4. 発表年 2019年

1. 発表者名 Yu Yamane, Masataka Seo, Ikuko Nishikawa
2. 発表標題 Diversity Preservation in Genetic Algorithm by Lifespan Control
3. 学会等名 The International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD) 2019 (国際学会)
4. 発表年 2019年

〔図書〕 計0件

〔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