

VAEを用いた線画のパターン分類

石田 汐音^{1,a)} 数藤 恭子^{1,b)}

概要: 近年、情報技術の発展により、実世界の多様かつ大量なデータを得られるようになった。しかし、それらのデータには正解ラベルが伴っていないことが多い。また、大量のデータに対して手作業でラベル付けを行うことは困難である。そのため、ラベルを必要としない教師なし学習によって、データの持つ特徴を獲得し、パターン分類を行いたいという用途は多い。本研究では、生成モデルの一種である変分オートエンコーダ (VAE) を用いて、教師なし学習による潜在変数を用いたパターン分類を行う。シンプルな線画のお手本を真似て描くタスクの評価指標作りへの適用可能性を検討した。パーキンソン病の検査の一種である公開データの NewHandPD データセットを用いた実験を行い、VAE の潜在ベクトルを用いた場合に、オートエンコーダ (AE) の潜在ベクトルを用いた場合や主成分ベクトルを用いた場合と比較し、健常者クラスとそれ以外のクラスの分離度が良いことを確認した。

Sketch Classification using the Latent Variables of VAE

Abstract: The improvement of the information technology enables us to obtain various and enormous data of the real world. However, in many cases, their ground truth labels are not available, nor the ground truth labels are not easy to annotate manually. So, there are many purposes of unsupervised learning to obtain good features of data to classify the patterns. We adapt a variational auto encoder (VAE), which is one of the generation model, to classify the patterns using the latent variables obtained through unsupervised learning. We evaluated the applicability of the VAE latent space to detect anomaly of the sketch data. In the evaluation using the public dataset, NewHandPD dataset which is one of the assessment of the Parkinson's disease, it was confirmed that we can separate the class of patients and class of healthy using the VAE latent variables better than when using the AE latent variables or using the PCA vectors.

1. はじめに

変分オートエンコーダ (VAE)[1] は、入力を圧縮するエンコーダ部分と、圧縮した特徴を元に出力を入力と同じ次元数にまで復元するデコーダ部分で構成され、出力を入力により近づけるように学習する。入力は、潜在変数を用いた条件付き確率分布に従い生成されると仮定し、対数尤度最大化により確率分布を推定する。連続した潜在変数として有力な意味を持つ低次元の特徴を獲得できる性質から、データの属性抽出や、属性の制御を伴う画像生成、ラベルのないデータに対するパターン分類にも応用される。

本研究では、VAE を用いた教師なし学習による線画データからの異常データの検出を検討する。リハビリテーション

分野における診断等では、一般的に、患者のデータ数が少ないという問題や、データが多く集められても診断ラベルを付与するのが困難であるという問題が存在する。これらの問題に対し、教師なし学習による診断手法が有効であると考えられる。今回は、パーキンソン病の患者と健常者に関する公開データセット NewHandPD dataset[2] を用いた実験により、VAE を用いた手法の適用可能性を検討した。

2. 関連研究

教師なしの異常検出は、特徴空間におけるはずれ値検出として実現される。はずれ値の求め方は、a). クラスタリングに基づく方法、b). 確率モデルに基づく方法、c). 再構成モデルに基づく方法などがあり、Ruff らによる異常検出手法の分類 [3] では、1 クラスの SVM などの方法が a)、マハラノビス距離や GMM モデル近似による方法が b)、深層生成モデルである VAE や GAN を用いる手法は b) と c) の中間、主成分分析 (PCA) を用いる手法は c) に属すると

¹ 東邦大学 〒274-8510 千葉県船橋市三山 2-2-1
Toho University Miyama 2-2-1, Funabashi-shi, Chiba 274-8521, Japan

a) 5518009i@st.toho-u.jp

b) kyoko.sudo@sci.toho-u.ac.jp

して整理されている。

VAEを用いる異常検出は、Anら[4]が再構成誤差を用いる方法を提案し、再構成誤差を指標としてMNISTデータの手書き数字画像のはずれ値が検出できることを示した。近年の様々な実際のデータへの応用例としては、3次元点群の異常検知[5]、工業部品の異常検知[6]などがある。はずれ値の指標としては、再構成誤差のほか、損失関数に含まれる項のうち、一つの項を取り出して異常度とした非正規化異常度を用いる方法が提案されている。

本研究では、対象とするデータが手書きの線画であり、データの性質としては数字画像にも近いと考察し、同様の手法で効果が得られるのではないかと考えVAEの適用を試みる。また、AEやPCAを用いる方法との比較を行う。

3. 提案手法

3.1 提案手法の概要

パーキンソン病 (PD) の患者と健常者に関する公開データセット NewHandPD dataset[2] を対象に、VAEを用いた異常検知を行う。学習時は健常者のデータのみを入力とする。

3.2 データセット

NewHandPD dataset は、124名のPD患者と140名の健常者による手書き試験の画像とその信号をまとめたデータセットである。手書き試験は、一個人に対し、meander(四角の渦巻き)とspiral(丸の渦巻き)の2パターンで行われている。データデータセットに含まれる画像の例を図1に示す。

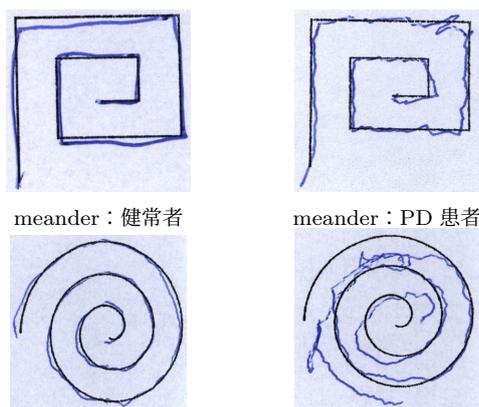


図1 NewHandPD dataset の一例

3.3 モデル

本研究では、Encoder部分とDecoder部分で構成される基本的なVAEモデルを用いる。ネットワーク構成として、Encoderは三層の全結合層で構成される。Encoderへの入力 は $28 \times 28 \times 3$ 次元であり、 $2352 \rightarrow 768 \rightarrow 128 \rightarrow N_z$ と

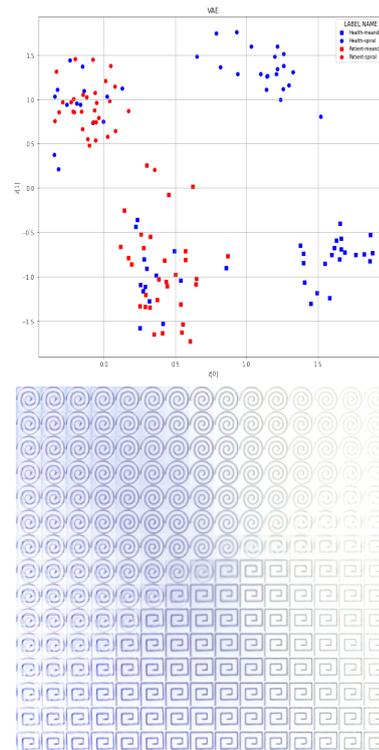


図2 学習したVAEの2次元の潜在空間におけるテストデータの分布(上)と、学習したVAEの潜在変数を連続的に変化させて自動生成したパターン(下)

Fig. 2 Data distribution of the 2-D latent variables of VAE(upper), and generated patterns from the latent variables.

表1 潜在変数の次元数と患者データの検出のROC-AUC
Table 1 Number of the dimension of the latent variables and ROC-AUC of the detection of the patients' data

n_z	ROC-AUC					
	2	5	10	20	50	100
VAE	0.83	0.85	0.84	0.86	0.82	0.81
AE	0.72	0.76	0.76	0.69	0.76	0.75
PCA	0.61	0.58	0.55	0.50	0.33	0.19

徐々に次元を圧縮していく。DecoderはEncoderと対になるように構成されている。中間層の活性化関数はReLU関数を使用し、出力層にはSigmoid関数を使用した。損失関数 L は、式(1)となっている。ここで、第一項は潜在変数 z を正規分布に近づけるための正則化項であり、式(2)で表される。第二項は入力と出力を近づけるための再構築誤差であり、式(3)とで表すことができる。

$$L = D_{KL}(N(\mu, \sigma^2) || N(0, I)) + L_{rec} \quad (1)$$

$$D_{KL} = -\frac{1}{2} \sum_{j=1}^{N_z} (1 + \log \sigma_j^2 - \mu_j^2 - \sigma_j^2) \quad (2)$$

$$L_{rec} = -\sum_{i=1}^{N_x} (x_i \log x'_i + (1 - x_i) \log(1 - x'_i)) \quad (3)$$

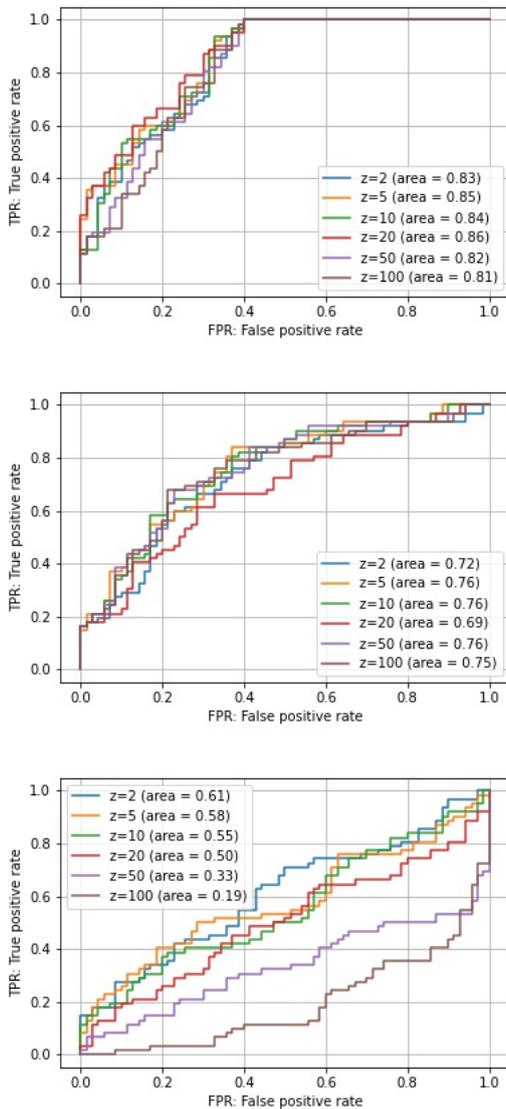


図 3 上から VAE, AE, PCA による患者データ検出精度の ROC 曲線

Fig. 3 ROC curves of the patients' data detection using VAE, AE and PCA(top to bottom)

4. 評価実験

本稿では、VAE と AutoEncoder(AE)[7], 主成分分析 (PCA) の異なる学習モデルを用いて異常検知性能を評価する。VAE については、潜在変数が 2 次元の場合について、潜在空間上の分布の可視化による確認も行なった。

4.1 実験条件

NewHandPD dataset の画像データのみを使用する。訓練データには健常者のデータ 210 枚を用い、テストデータには 132 枚 (健常者 70 枚, PD 患者 62 枚) の画像を用いた。画像は RGB 画像であり、 28×28 にリサイズした。最適化手法は学習率 $1e-3$ の Adam, バッチサイズを 15 とし, 50 エポック学習を行なった。AE は, VAE と同じネットワーク

構成とし, 損失関数には平均二乗誤差 (MSE) を使用した。また, 学習パラメータなども VAE と同様である。VAE, AE の潜在変数の次元数 N_z は, $N_z = [2, 5, 10, 20, 50, 100]$ の 6 パターンで学習を行なった。PCA のコンポーネント数は, N_z と同様の 6 パターンで実験した。評価指標は分離度と ROC 曲線の下部面積 (ROC-AUC) を用いた。ROC-AUC は値が 1.0 に近いほど異常検知性能が高いことを示す指標である。

4.2 実験結果

潜在変数の学習

学習した VAE の 2 次元の潜在空間におけるテストデータの分布は, 潜在変数の数 $n_z = 2$ の場合に図 2(上) のようになった。横軸では, 健常者 (青) の一部と患者 (赤) とが分離されており, 縦軸では四角と丸の形状が分離されるように学習されている。z の値を連続的に変化させて自動生成したパターンは図 2(下) のようになった。これらの結果から, 被験者のカテゴリと模様のパターンが主な潜在属性として学習されたと考えられる。

患者データの検出

再構成誤差を指標とした異常検出手法による患者データの検出について, VAE, AE, PCA による結果の ROC 曲線を求めた結果を図 3 に, AUC の値を表 1 に示す。VAE による結果が最も良く, 潜在変数の次元数は 5 ~ 20 が良いという結果が得られた。

5. まとめ

線画の再構成に基づく異常検知を試みた。教師なし学習により, 潜在空間においてパターンの形の違いや, 正常パターンや異常パターンの線の揺れやお手本からのズレを獲得できたと考えられる。その結果, AE や主成分分析などの手法を比べて異常検知性能で上回った。今回は単純な手書き試験画像を対象としたが, 今後はより複雑な線画への適用や, 異常の度合いを定量化できるかどうかの検討を行う。

参考文献

- [1] D. P. Kingma, M. Welling, "Auto-Encoding Variational Bayes", in Proc. 2nd International Conference on Learning Representations, (2014).
- [2] C. R. Pereira, S. A. T. Weber, C. Hook, G. H. Rosa and J. Papa, "Deep Learning-aided Parkinson's Disease Diagnosis from Handwritten Dynamics," 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI) (2016).
- [3] L. Ruff, J. R. Kauffmann, R. A. Vandermeulen, G. Montavon, W. Samek, M. Kloft, T. G. Dietterich and K. Muller, "A Unifying Review of Deep and Shallow Anomaly Detection," arXiv:2009.11732v3 (2021).
- [4] J. An and S. Cho, "Variational Autoencoder based Anomaly Detection using Reconstruction Probability,"

- Special Lecture on IE, vol.2, pp. 1-18 (2015).
- [5] M. Masuda, R. Hachiuma, R. Fujii, H. Saito and Y. Sekikawa, "Toward Unsupervised 3D Point Cloud Anomaly Detection using Variational Autoencoder," IEEE International Conference on Image Processing(ICIP) (2021).
 - [6] T. Matsubara, K. Sato, K. Hama, R.Tachibana, K. Uehara, " Deep Generative Model Using Unregularized Score for Anomaly Detection With Heterogeneous Complexity." IEEE Transactions on Cybernetics (2020).
 - [7] G. E. Hinton, R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks,"Science, vol.313, no.5786, pp. 504-507 (2006).