ニューラルネットワークを用いた異常検知手法のアンサンブル学習に関する研究

1X19C132-1　山本陽一郎
指導教員　後藤正幸

1. 研究背景と目的

分析や活用のために蓄積されたデータには異常データが含まれていることがある．異常データとは例えば不正や故障によって生じたデータであり，これらを正常データと区別する必要があるが，一般に大多数は正常データであり，異常データはまれにしか発生しないため，単純な二値分類器による識別は多くの場合困難である．そのため，正常異常のラベル付けが行われていない既存のデータセットから正常データの特徴を学習することによって，未知のデータに対して異常を検知する教師なし異常検知技術が極めて重要な役割を担っている．

近年の論文では複雑な高次元のデータセットにおける教師なし異常検知において，深層学習を用いることが非常に有効であると報告されている．代表的な深層学習を用いた異常検知手法の 1 つに One Class Neural Network(以下，OC-NN)[1] がある．この手法ではまず，オートエンコーダと呼ばれる深層学習手法を用いてデータの特徴が反映された潜在表現を抽出する．次にこれを 3 層ニューラルネットワークに入力し，データが異常であるか否かを推定する異常検知指標を出力する．OC-NN のようにニューラルネットワークを用いた手法では目的関数が非凸になるため，勾配法を用いて学習を行う．そのため，局所的最適解に陥る可能性があり，安定して高い性能を出すことが容易ではない．さらに OC-NN の精度はニューラルネットワークの構造や重みの初期値の他，分類失敗を許容する度合いを表すハイパーパラーメータ ν の値によっても変化する．一般に，ν の値が大きいほど正常データを異常と判定する失敗を許容してより積極的に異常データを検出するモデルになり，異常検知精度も大きく左右されるため，ν を適切な値に設定することは OC-NN の重要な課題の 1 つである．

本研究では上記の課題を解決するため，同じネットワーク構造で重みの初期値が異なる複数の OC-NN を組み合わせてアンサンブル学習することにより，高い精度で安定した異常検知が可能な手法を提案する．提案手法では，複数の OC-NN で得られた異常検知指標を統合することで，新たな異常検知指標を作成する．最後に実際の手書き画像データセットに提案手法を適用し，その有効性を示す．

2. 準備

2.1. 問題設定

異常検知手法は教師あり型と教師なし型の 2 つのタイプに大別される．教師あり型は正常パターンと異常パターンをモデルに学習させる必要があり，正常・異常データそれぞれが学習時に十分に獲得されている場合においては有効である．一方，異常データの獲得が困難な場合は教師なし型の異常検知手法を適用する必要がある．教師なし型では学習データの大半を正常データと仮定して正常データの特徴を学習し，異常データの検出を行う．本研究では，異常データの獲得が困難な場合を想定し，教師なし型の異常検知を対象とする．

2.2. OC-NN

OC-NN はまず，オートエンコーダを学習する．次に，学習したオートエンコーダのエンコーダにより得られるデータの潜在表現を 3 層のニューラルネットワークに入力し，異常検知指標を出力する．ここでオートエンコーダとはエンコーダとデコーダによって構成されるモデルであり，エンコーダで入力を圧縮し，デコーダで入力を再構成するように学習される．OC-NN のイメージを図 1 に示す．

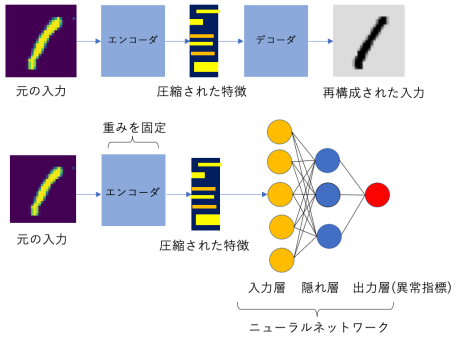

図 1: One Class Neural Network

OC-NN の学習で用いられる目的関数は式 (1) で表される．

$$\min_{\boldsymbol{w},\boldsymbol{V},r}\Big\{\frac{1}{2}\|\boldsymbol{w}\|_2^2+\frac{1}{2}\|\boldsymbol{V}\|_F^2+ \frac{1}{\nu}\cdot\frac{1}{N}\sum_{n=1}^{N}\max\left(0,r-\boldsymbol{w}^T g\left(\boldsymbol{V}\boldsymbol{x}_n\right)\right)-r\Big\} \tag{1}$$

ここで，$\|.\|_2^2$ は L2 ノルム，$\|.\|_F^2$ はフロベニウスノルム，$\boldsymbol{x}$ は圧縮された学習データ，$\boldsymbol{V}$ は入力から隠れ層までの重み行列，$g(.)$ はベクトルを出力とする活性化関数，w は隠れ層から出力までの重みベクトルである．ニューラルネットワークの出力 $\hat{y}(\boldsymbol{w},\boldsymbol{V})=\boldsymbol{w}^T g(\boldsymbol{V}\boldsymbol{x})$ は異常検知指標であり，小さいほど異常の可能性が高いことを意味し，r よりも小さい場合は異常，大きい場合は正常と判断する．OC-NN はニューラルネットワークの隠れ層によってオートエンコーダで得られた潜在表現を異常検知に適した表現に変換することができ，構造が複雑なデータセットに対しても有効であるとされる．

3. 提案

3.1. 着想

OC-NN は複雑で高次元なデータセットに対しても高精度であるが，学習によって得られるニューラルネットワークの重み $\boldsymbol{w},\boldsymbol{V}$ の値が，その初期値やニューラルネットワークの構造，ハイパーパラメータ ν の値に依存するため，精度が安定しないという課題がある．そこでニューラルネットワークの構造，ハイパーパラメータ ν の値を固定し，重み $\boldsymbol{w},\boldsymbol{V}$ の初期値を変化させた複数の OC-NN を組み合わせた

アンサンブル学習により，安定して高い精度が出せる異常検知手法を提案する．

3.2. 提案手法

提案手法ではまず，同じ学習データを用いてニューラルネットワークの重み $\boldsymbol{w}, \boldsymbol{V}$ の初期値が異なる複数の OC-NN をアンサンブル学習の考え方に基づき，並列に学習する．学習後の OC-NN は最適化された $\boldsymbol{w}, \boldsymbol{V}$ の値がそれぞれ異なる．そして，これらの OC-NN から得られた出力の合計値を新たに異常検知指標 S_T として定義する．

$$S_T = \sum_{t=1}^{T} \hat{y}_{nt}(\boldsymbol{w}_t, \boldsymbol{V}_t) \tag{2}$$

ただし，T は OC-NN のアンサンブル数である．提案手法のイメージを図 2 に示す．

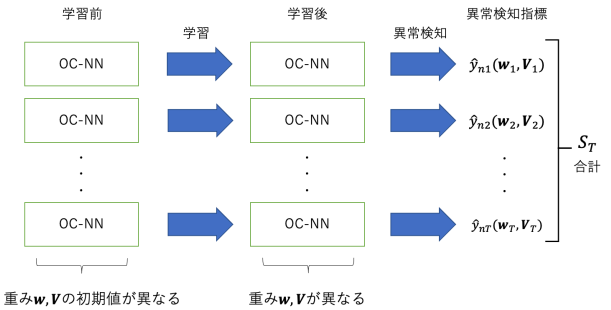

図 2: 提案手法

4. 実験

4.1. 実験条件

従来手法 OC-NN と提案手法の精度及びその安定性の比較を行うために，手書き数字データセットである MNIST[2] の 1 を正常，7 を異常データとして異常検知実験を行う．本実験では学習データを 6,000 件，テストデータを 1,000 件とし，異常割合をそれぞれ 10%とする．各データの特徴量数はオートエンコーダにより 784 から 32 に圧縮し，ニューラルネットワークの入力層および隠れ層のニューロン数は 32 とする．エポック数は 200，提案手法における OC-NN のアンサンブル数 T は 10，ハイパーパラメータ ν は 0.10 から 0.50 まで 0.10 間隔で変化させた 5 通りで実験を行った．評価指標には，各条件で実験を 10 回行った際の Area Under the ROC Curve(AUC) の平均と 95%信頼区間を用いる．

4.2. 結果

各手法の AUC の平均値と，95%信頼区間を表 1 および表 2 に示す．青字の部分は各手法で最も高い AUC の平均値であることを示す．表 1 より OC-NN は 95%信頼区間が大きく，精度にばらつきがあることがわかる．さらに，ハイパーパラメータ ν の値によっても精度に大きな差が出ていることがわかる．また表 2 より提案手法は 95%信頼区間が非常に小さく，ν の値にも精度がほとんど左右されておらず，OC-NN よりも安定したモデルであることがわかる．また $\nu = 0.10$ の場合を除いて提案手法の方が AUC の平均値が高く，$\nu = 0.20$ 及び 0.50 では 95%信頼区間において有意差があることがわかる．それぞれの手法における最も高い AUC の平均値を比較しても提案手法の方が高く，提案手法は安定性だけでなく，精度においてもオリジナルの OC-NN より優れていると言える．

表 1: OC-NN の実験結果

ν の値	0.10	0.20	0.30	0.40	0.50
AUC	**0.95652**	0.85192	0.93973	0.92547	0.72776
95%信頼区間(±)	0.00786	0.08862	0.03778	0.04052	0.13864

表 2: 提案手法の実験結果

ν の値	0.10	0.20	0.30	0.40	0.50
AUC	0.94561	**0.96104**	**0.96607**	**0.96379**	**0.96546**
95%信頼区間(±)	0.00200	0.00126	0.00037	0.00032	0.00023

5. 考察

実験より，本提案手法は OC-NN に比べて安定して高い精度で異常検知を行うことができることがわかった．このことは，モデルを安全に運用可能な点から有用である．また，提案手法はハイパーパラメータ ν の値に対してロバストであることもわかった．OC-NN では最適な ν の値を設定する必要があるが，提案手法は ν をアンサンブル学習に用いる個々の OC-NN にとって最適な値に設定しなくても高い精度で異常検知を行えるため，ν の値を探索する手間が省ける点で有用である．

さらに，OC-NN では ν の値が小さい場合で最も高い精度を示したのに対し，提案手法では ν の値が比較的大きい場合で最も高い精度を示した．つまり異常データが少ない場合において，分類失敗を許容して異常を積極的に検出する比較的精度の低い OC-NN を組み合わせる方が，精度の高い OC-NN を組み合わせるよりも高い精度で異常を検出できることがわかった．このことは，弱学習器を複数組み合わせることにより精度を向上させるというアンサンブル学習の基本的な考え方に沿っている．

6. まとめと今後の課題

本研究では，深層学習を用いた教師なし異常検知手法 OC-NN の安定性向上を目的として，同手法のアンサンブル学習を提案した．具体的には，複数の OC-NN から得られた異常検知指標を統合することで新たな異常検知指標を作成した．そして，MNIST を用いた異常検知精度の評価実験によって，提案手法は精度が高く，安定していることに加え，ハイパーパラメータ ν の値に対してロバストであることも明らかになった．

今後の課題としては，OC-NN のアンサンブル数 T の最適化が挙げられる．提案手法では，ν の値を設定する必要はないが，T の設定が必要となった．この値は精度だけでなく，学習時間や使用メモリ量にも大きな影響を及ぼすため適切な値に設定することは非常に重要である．

参考文献

[1] Raghavendra Chalapathy, Aditya Krishna Menon, and Sanjay Chawla, “Anomaly detection using one-class neural networks.” *arXiv preprint, arXiv*:1802.06360, 2018

[2] Yann LeCun, Corinna Cortes, and Christopher JC Burges. Mnist handwritten digit database. *AT&T Labs [Online]. Available:* http://yann.lecun.com/exdb/mnist, 2, 2010.