令和 4 年度卒業研究発表会（日本大学工学部情報工学科）
C－2－10

# 学習識別再構成を用いたネットワークトラフィックの異常検知

Network Traffic Anomaly Detection Using Learning Discriminative Reconstructions

316093　　設楽　夏希　　［和泉研究室］

## 1.　まえがき

近年，サイバー攻撃の急激な増加が問題となっており，セキュリティ対策として悪性のネットワークトラフィックを検出することが重要である．

機械学習を用いたネットワークトラフィックの異常検知では，良性通信と悪性通信を明確に区別したデータセットが必要である．しかし，良性通信と悪性通信の判別，ラベル付けには専門的な知識や膨大な量を処理しなければならずコストが問題となるため，十分なデータセットが存在しないのが現実である．

そこで本研究ではオートエンコーダ，学習識別再構成による異常データ除去，動的な閾値を用いたラベル無し学習データによる異常検知の手法を提案する．

## 2.　オートエンコーダ

オートエンコーダ(以下 AE と)とは，ニューラルネットワークの１つであり，入力した次元を圧縮(エンコード)し，重要な特徴量だけを抽出した後，再度元の次元に復元処理(デコード)をするアルゴリズムである．

この復元されたデータと AE に入力前のデータの平均二乗誤差を再構成誤差としてこれを最小化するように学習を行うことで異常検知を実現する．また，正常データのみを学習させることで未知の異常にも対応できる．

## 3.　学習識別再構成による異常データ除去[1]

このアルゴリズムはミニバッチごとに AE で算出された再構成誤差の分散を元に正常データか異常データかを推定し判別する判別ラベリング処理とそれにより正常データと判別されたデータの再構成誤差を減らすことで誤差分布の分離性能を向上させる再構成学習の 2 つのステップを繰り返すことにより異常データを取り除きながら学習していくモデルである．

### 3.1.　判別ラベリング

このステップでは，データラベルの推定，つまりミニバッチの各データのラベルを正常データラベル$yi=1$と異常データラベル$yi=0$のいずれかに分類することを目的とする．以下の関数を最適化することにより実現する．

$$\min_{y} h = \frac{\sum_{yi=1}(\epsilon_i - c^+)^2 + \sum_{yi=0}(\epsilon_i - c^-)^2}{\sum(\epsilon_i - c)^2} \qquad (1)$$

ここで$\epsilon_i$とはミニバッチの各データの再構成誤差であり，$c^+$，$c^-$，$c$とはそれぞれミニバッチの正常データの平均，異常データの平均，全データの平均である．

$h$ は誤差分布の分離性能が高いほど小さい値を示す．

### 3.2.　再構成学習

このステップでは判別ラベリング処理によってラベリングされたデータの正常データのみを用いて再構成誤差を最小化するように学習させる．これにより再構成誤差の分布を分離しやすくする．

### 3.3.　アルゴリズム

以下に，学習識別再構成の具体的なアルゴリズムについて述べる．

1. AE に学習データを入力し，ミニバッチの各データの再構成誤差$\epsilon_i$ を求め，ラベルを異常とする．
2. 判別ラベリングのため，求めた再構成誤差をソートし，再構成誤差の小さいものから正常データとしてラベリングを行う，その度に式（1）の$h$を求め，h が最小となるところで終了する．
3. 判別ラベリングにより正常データと判別されたデータと誤差分布の分離性能を表す$h$ を用いて再構成学習を行う．これは以下の目的関数，式（2）の損失$L$を最小化するよう学習することで実現される．

$$L = \frac{1}{n^+}\sum_{yi=1}\epsilon_i + \lambda h \qquad (2)$$

ここで，右辺の$\sum_{yi=1}\epsilon_i$ は正常データの再構成誤差の総和であり，$n^+$ はその数，$\lambda$ はトレードオフを制御する変数である．

4. 1. ～3. を損失$L$ が収束するまで繰り返す．

## 4.　提案手法

次々に新たなサイバー攻撃が生まれる状況において未知の異常を検知することが重要になっている．そこで正常データのみを AE に学習させることが有効であると考えられる．これを学習識別再構成と動的な閾値を用いて実現する手法を提案する．

従来は異常画像除去に用いられていた学習識別再構成であるが学習データを正常と異常トラフィックの混在するラベル無しデータとすることでトラフィックデータの再構成誤差の分布を分離させるように学習させる．また，異常検知を行う際は判別ラベリングを用いて入力データの再構成誤差の分布から動的に閾値を決定し，閾値未満のデータを正常，閾値以上のデータを異常と推定する．

以下に，閾値決定のアルゴリズムを示す．

1. トラフィックデータを入力し再構成誤差を求める．
2. 判別ラベリングを用いて再構成誤差の大きい$(yi=0)$クラスタと小さい$(yi=1)$クラスタに分ける．
3. 再構成誤差の大きいクラスタの最小値，小さいクラスタの最大値の平均値を閾値とする．

この手法によりトラフィックデータの異常検知を実現する．

## 5.　実験

### 5.1　実験方法

本実験では実験データとしてNSL-KDDデータセット[2]を使用する．このデータセットは通信データをセッション単位で加工したものであり 41 個の特徴[3]がある．また，使用した AE モデルを図１に示す．

NSL-KDD データセットの学習に用いる正常データ，異常データの割合を変化させたデータを用いて，提案手法を使用した際と提案手法を使用せず AE のみを用いた際，また，正常データのみを AE に学習させた際との比較を行う．

1. NSL-KDDデータセットの学習データのうち正常データと異常データの割合がそれぞれ約 8:2，5:5，2:8 のラベルを隠したデータセットを作成する．
2. 各データセットに対し本手法を用いて学習させる．

また，比較のため各データセットと正常データのみのデータセットに対し本手法を使用せず AE のみを用いて学習させる．

3. 学習させたモデルに対しテストデータを入力し，動的に閾値を決定することで入力データが正常か異常かを推定しその性能を比較する．

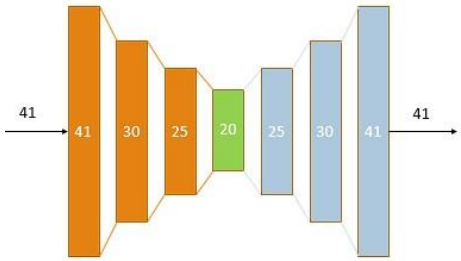


**図 1　A E モデル**

### 5.2　実験結果

実験によって得られた結果を表 1 に示す．

**表 1 実験 5.1 結果−混同行列**

| 処理 | | | 予測クラス | | |
|---|---|---|---|---|---|
| | | | 正常 | 異常 | 正答率 |
| 正常80% 異常20% 提案手法あり | 真のクラス | 正常 | 76.8% | 23.2% | 82.1% |
| | | 異常 | 12.7% | 87.3% | |
| 正常80% 異常20% 提案手法なし | | 正常 | 84.3% | 15.7% | 82.1% |
| | | 異常 | 20.1% | 79.9% | |
| 正常53% 異常47% 提案手法あり | | 正常 | 77.3% | 22.7% | 82.3% |
| | | 異常 | 12.7% | 87.3% | |
| 正常53% 異常47% 提案手法なし | | 正常 | 60.9% | 39.1% | 42.0% |
| | | 異常 | 76.9% | 23.1% | |
| 正常20% 異常80% 提案手法あり | | 正常 | 84.3% | 13.7% | 77.5% |
| | | 異常 | 29.4% | 70.6% | |
| 正常20% 異常80% 提案手法なし | | 正常 | 28.7% | 71.3% | 29.0% |
| | | 異常 | 70.7% | 29.3% | |
| 正常データのみ学習 | | 正常 | 83.7% | 16.3% | 83.0% |
| | | 異常 | 17.8% | 82.2% | |

## 6.　考察

実験結果から，AE に正常データのみを学習させた際の正答率が 83.0%であるのに対し，同じ正常データに異常データを追加しラベル無しで学習させた正常 53% 異常 47% 提案手法ありの正答率は 82.3%と，AE に正常データのみを学習させた際の 99.1%の結果を得られた．このことから本手法は正常データのみを学習させた際と同程度の性能があり，人力によるラベル付けを必要とせずラベルがある場合と同程度の性能を得られると考えられる．

また，正常データが少ない場合，偽陽性の値が大きくなっていることから異常データを正常データと誤検知してしまう傾向が強くなると考えられる．しかし，提案手法では誤差分布の反転が起きていない，ミニバッチごとの適切なラベリングにより誤差分布の反転を防ぐという効果もこの提案手法を用いる利点であると考えられる．

正常データ 80% 異常データ 20%の結果を見ると正答率は 82.1%と同じであるが提案手法ありの誤差分布 図 2 と提案手法なしの誤差分布 図 3 を比較すると提案手法を用いた際の方が明らかに誤差分布の分離性能が高いことが見て取れる．また，提案手法を用いた際は偽陽性の値が高く，偽陰性の値が小さくなっており，提案手法を用いなかった際は偽陰性の値が低く，偽陽性の値が高くなっている．異常検知を行う上では異常データを見逃さないために偽陰性の値が小さいことが重要であるため提案手法を用いた結果の方が良いと考えられる．

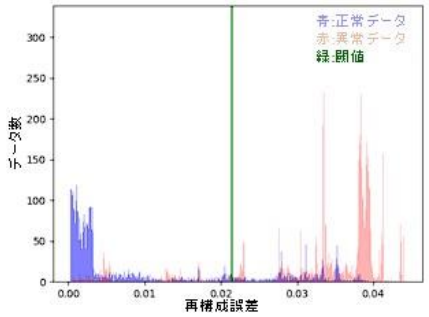


**図 2　正常 80% 異常 20% 提案手法あり−誤差分布**

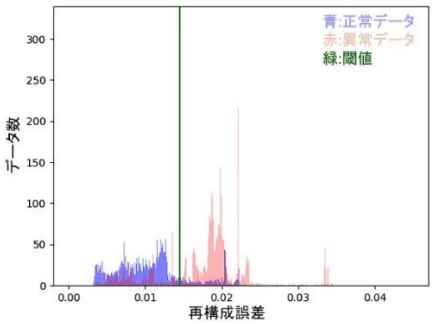


**図 3　正常 80% 異常 20% 提案手法なし−誤差分布**

## 7.　むすび

本研究では，AE と学習識別再構成による異常データ除去，動的な閾値を用いたラベル無しデータによる異常検知を行った．本手法は AE に正常データのみを学習させた際と同等な性能があるという結果を得ることが出来た．今後の課題として，精度の向上があげられるが，CNNAE やその他の AE モデルを用いて検証をしていく必要がある．

## 参考文献

[1] Y. Xia, X. Cao, F. Wen, G. Hua and J. Sun, ″Learning Discriminative Reconstructions for Unsupervised Outlier Removal″, 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015, pp. 1511-1519, doi: 10.1109/ICCV.2015.177.

[2] M. Tavallaee, E. Bagheri, W. Lu, and A. Ghorbani, “A Detailed Analysis of the KDD CUP 99 Data Set”, Submitted to Second IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA), 2009.

[3] Selecting Optimal Subset of Features for Intrusion Detection Systems - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/The-41-features-of-NSL-KDD-dataset_tbl1_287302551 [accessed 24 December, 2022]