Vol.2021-ARC-244 No.24
Vol.2021-SLDM-194 No.24
Vol.2021-EMB-56 No.24
2021/3/25

特徴量選択アプローチと連合学習によるネットワーク侵入検知手法の検討

秦 洋[1]　近藤 正章[1]

概要：ネットワーク攻撃の増加と多様化により，それを検知するための技術として機械学習が活用されている．連合学習（Federated Learning）は機器学習を使用する際に，ユーザを特定できるデータを集約せず，ローカルモデルから協調的に予測モデルを訓練する分散型の学習手法である．しかしながら，ネットワーク攻撃は高度化・標的特化型化が進むとともに，攻撃種類の特性に応じて検知モデルを構築する必要性が高まっている．さらに，ネットワークトラフィックから有効な特徴量を選択することは，通信データ前処理の最も重要な課題の一つと考えられる．上記のような問題に取り組むため，本稿では，特定の攻撃種類に対してより優れた検知精度を達成可能な特徴量を選択するための貪欲アルゴリズムを提案する.また，各エッジデバイスで決定された特徴量に応じ，連合学習によりサーバ側で複数の共通モデルを構築する．提案手法の有効性を評価するために，異常検知のために提案されたオンデバイスニューラルネットワークを利用し，NSL-KDD データセットにおけるシミュレーション実験を行った．評価結果より，本提案手法が単純なモデル構築手法に比べて検知精度が大幅に向上することがわかった．

キーワード：ネットワーク侵入検知，特徴量選択，攻撃種類，連合学習，NSL-KDD データセット

Federated Learning-Based Network Intrusion Detection with a Feature Selection Approach

Yang Qin[1]　Masaaki Kondo[1]

Abstract: In the past few years, with the increase and diversity of network attacks, machine learning has shown its efficiency in realizing intrusion detection. Federated Learning (FL) has been proposed as a distributed machine learning approach, which collaboratively trains a prediction model by aggregating local models of users without sharing their privacy-sensitive data. However, since the attacks are becoming more sophisticated and targeted, there is a growing need to enhance detectin models according to the characteristics of attack type. Moreover, choosing effective feature sets from the network traffic characteristics is considered one of the most important tasks of data preprocessing. To tackle the problems above, we first suggest a greedy algorithm to help select features that achieve better intrusion detection accuracy regarding different attack categories. Afterward, multiple global models are generated by the server in federated learning, according to the decided features of edge devices. For evaluating the effectiveness of the proposed approach, simulation experiments based on the latest on-device neural network for anomaly detection are conducted over the NSL-KDD dataset. Experimental results demonstrate greatly improved accuracy of our method.

Keywords: Network intrusion detection, Feature selection, Attack type, Federated learning, NSL-KDD dataset

1. はじめに

デジタルトランスフォーメーション（DX）や Society 5.0 の推進により，Internet of Things（IoT）機器は家電，交通，医療など様々な分野で急速に普及している．IoT 機器は低価格ながら高度な処理能力を持った装置であるため，セキュリティ攻撃の拠点や踏み台として悪用されるケースが急増している．侵入検知システム（Intrusion Detection System: IDS）はネットワーク上のイベントを監視し，何らかの不正侵入の兆候を検知すると，管理者に通知するシステムである．不正の検出方法で分類すれば，既存 IDS は「シグネチャ型」と「異常検出型」の 2 種類がある．前者は，セキュリティ攻撃データベースに登録されているパターンに一致するイベント等を検出するもので，頻繁に攻撃データベースが更新されることが必要である．また，登録されていないパターンの不正な通信は検知することができない．一方，「異常検出型」は IoT 機器の正常通信状態を監視し，正常通信状態に一致していない場合に不正通信と判断する．そのため，「異常検出型」の IDS はデータベースの更新を必要がなく，効率的に未知の攻撃を検出することができる．IoT 環境におけるセキュリティ攻撃の多様性を考えると，「シグネチャ型」と比較して，「異常検出型」の方法が IoT 機器に適していると考えられる．

IoT 機器における攻撃検知を実現するためには，各 IoT 機器固有の制約や通信の特徴を考慮する必要がある．特に，①ネットワーク攻撃の多様化と未知性，②性能・電力の制約，そして③高次元となる時系列データへの対応，を考慮した手法が必要である．近年，深層学習技術が「異常検出

1 東京大学
The University of Tokyo

Vol.2021-ARC-244 No.24
Vol.2021-SLDM-194 No.24
Vol.2021-EMB-56 No.24
2021/3/25

型」の IDS に実装され, 侵入攻撃の検知に活用されている[1-3]. 一方で, 深層学習を利用した手法では, 検知モデル構築のために, 高い計算能力および高品質の正常・異常アクセスデータが教師データとして必要である. IoT 機器の性能・電力制約を考えると, 通信内容の分析処理やアクセス情報の保存は通常クラウド側で行う必要がある. この場合, 通信コストや検知までの遅延時間, またクラウドへの通信によるセキュリティリスク増大などが問題となる. これらの背景より, 計算資源や電力資源に制約がある IoT 機器において, IoT 機器自身が高次元の通信データを高速に分析しつつ, 不正アクセルをリアルタイムで検知する技術の研究開発が重要となっている. 本研究は, 限りある計算・電力資源の中で IoT 機器が自身の通信パターンを学習により, セキュリティ攻撃を検知する技術を開発することを目的とする.

リアルタイムで不正なネットワークを検知するために, 本稿では軽量なニューラルネットワーク ONLAD [4]を利用し, IoT 機器自身がトラフィックを監視しつつ, ネットワーク攻撃を迅速かつ自動的に検知するシステムを検討する. ONLAD のニューラルネットワークの構築手法は第 2 章で紹介する. 一般に, ネットワーク攻撃は高度化・標的特化型化が進んでおり, すべてのネットワーク攻撃を高精度に検出できるモデルの構築は困難になってきている. 本研究は, 特定の攻撃種類を監視対象として, 異常検知モデルを学習することを提案する. 例えば, 主に Dos 攻撃を受ける IoT 機器では, Dos 攻撃に対する機械学習モデルを構築することで, 検出精度の向上が期待される.

一方, 高次元の通信データに対応するために, ネットワークトラフィックからの特徴量選択手法が注目を集めている. 最近では, 機械学習における検出精度の向上や計算時間の短縮を目的として特徴量を選択する手法も研究されている[5, 6]. 通常, 最適な特徴量セットを算出するのは時間がかかる処理であり, 一般的に使用されているn個の特徴量から生成可能な特徴量セット数は$2^n - 1$に達する. また, 攻撃に応じて特定の特徴量が変化するため, 攻撃の種類によって最適な特徴量セットの選択が異なる. 例えば, Network Security Laboratory-Knowledge Discovery and Data Mining （NSL-KDD）データセット[7]では, Dos 攻撃に対する検出精度が最も高い特徴量セットは, Probe 攻撃にはあまり適切でない. そのため, 攻撃の種類に応じて適切な特徴量を選択することが重要である.

連合学習（Federated Learning）[8]はユーザを特定できるデータを集約せず, ローカルに構築されたモデルから協調的に予測モデルを学習する手法であり, 近年機械学習の分野で注目されている. 一般に, IoT 機器は少量のトラフィックデータを生成するため, ネットワーク侵入検知の分野で連合学習のフレームワークを利用することも検討されている[9, 10]. しかしながら, 機器の種類と動作環境の多様性により, 機器の特性に応じて通信データは異なる. 加えて, 上記に述べたように, 多種類の攻撃を検出することを目的とした検知モデルでは, 単一の攻撃を対象にするのに比べて検出精度が低下する. そのため, 連合学習を用いてすべてのローカルモデルを集約すると, 検出精度が低下してしまう可能性がある. 本稿では, 従来の研究とは異なり, ターゲットとなる攻撃の種類に応じてエッジデバイスをグループ化して連合学習を行うことで, 侵入検知精度を向上させる手法を提案する. 最初に検出目的とする攻撃種類に応じてデバイスごとに特徴量選択を行い, その後各デバイスで決められた特徴量に基づいてローカルモデルを集約する. 同じ特徴量を選択したデバイスが連合学習される. この際に, サーバでは複数のグローバルモデルが構築されることになる.

本研究の貢献を下記に示す.

- **オンデバイス侵入検知手法 :** 最新の逐次学習および教師無し異常検知手法を統合したニューラルネットワークを用い, オンデバイス学習による攻撃検知システムを構築する.
- **攻撃種類に応じた特徴量選択手法** : 攻撃の種類に応じた学習モデルの構築手法を検討する. さらに, 検知精度を向上させるため, 特徴量選択における貪欲アルゴリズムを提案する.
- **連合学習に基づく侵入検知システム :** 連合学習フレームワークを用いて, モデル集約による高精度な検出モデルを構築する. 集約アルゴリズムは, 各 IoT 機器で決められる特徴量の選択に基づいて設計される.

本論文の構成は以下に示す通りである. 第 2 章では, 本研究で使用する検知モデルと連合学習の基本知識を紹介する. 第 3 章では, 特徴量選択アルゴリズムと連合学習ベースの検知システムを提案する. 第 4 章では, NSL-KDD データセットを用いて提案手法の有効性を評価する. 最後に, 第 5 章で論文のまとめと今後の課題を述べる.

2. 関連研究

本章では, ネットワーク攻撃の検知に利用するオンデバイスニューラルネットワークを詳細に述べる. 次に, このニューラルネットワークに適用する連合学習アルゴリズムを紹介する.

2.1 オンデバイス学習モデル

本研究では, On-line Sequential Extream Learning Machine（OS-ELM）[11]とオートエンコーダ（Autoencoder）[12]の組み合わせを利用する ONLAD [4]と呼ばれる逐次学習型のモデルを利用する. 各機械学習モデルの詳細は以下に示す通りである.

1) OS-ELM : OS-ELM は, 単層フィードフォワードニューラルネットワーク（SLFN）の形状を取り, バッチ学習アルゴリズム ELM[13]の利点を維持しつつ逐次学習を実装し

Vol.2021-ARC-244 No.24
Vol.2021-SLDM-194 No.24
Vol.2021-EMB-56 No.24
2021/3/25

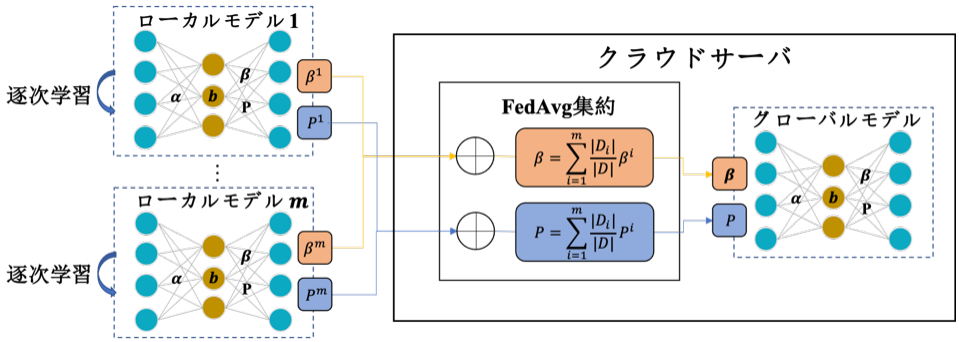

図 1　逐次学習型のニューラルネットワーク ONLAD における FedAvg を用いた連合学習の概要

たモデルである．バッチサイズkのn次元の入力データ$x \in R^{k\times n}$に対し，m次元の出力データ$y \in R^{k\times m}$は以下の数式で表現される．

$$y = G(x \cdot a + b)\beta \quad (1)$$

$G(\cdot)$は活性化関数，aは入力層-隠れ層の重み，b は隠れノードに関するバイアス，βは隠れ層-出力層の重みである．さて，m次元の教師データ$t \in R^{k\times m}$は損失 0 で推論できるとした場合，次の等式を満たす$\tilde{\beta}$が計算できる．

$$G(x \cdot a + b)\tilde{\beta} = t \quad (2)$$

$H \in R^{k\times m}$をネットワークの隠れ層の各ノードの値を表現する行列$H = G(x \cdot a + b)$と定義する．式 (2) を満たす最適な重み$\tilde{\beta}$は，$\tilde{\beta} = H^{\dagger}t$として定式化される．ここで，$H^{\dagger}$はムーア-ペンローズの一般逆行列である．

逐次学習に合わせるために，i番目の学習データ$\{x_i \in R^{k_i\times n},\ t_i \in R^{k_i\times m}\}$が到来した場合，出力重みの更新式は以下のように求められる．

$$P_i = P_{i-1} - P_{i-1}H_i{}^T\left(I + H_iP_{i-1}H_i{}^T\right)^{-1}H_iP_{i-1}$$
$$\beta_i = \beta_{i-1} + P_iH_i{}^T(t_i - H_i\beta_{i-1}) \quad (3)$$

特に，0 番目の訓練データに対し，出力層の重みを次の式で計算する．

$$P_0 = \left(H_0{}^TH_0\right)^{-1}$$
$$\beta_0 = P_0H_0{}^TH_0 \quad (4)$$

2) Autoencoder：Autoencoder は，ニューラルネットワークを利用した教師無し機械学習の 1 つであり，高次元データに対し，次元削減や特徴抽出のために用いられることが多く，異常検知においても注目を集めている．Autoencoder のネットワークは，エンコーダとデコーダの 2 つの部分から構成される．エンコーダの部分は入力データ次元削減や特徴抽出の機能を獲得する．一方，デコーダの部分は圧縮した低次元の情報を入力データに復元する．このネットワークの目的は，元の入力データにできるだけ近い形で出力を再構成する．学習の過程では，入出力が一致するように各エッジの重みを調整する．学習データと同様な特徴を持つデータを入力した場合，損失が小さくなる．逆に学習データに持ちられなかった特徴を持つ入力データに対しては損失値が大きくなる．そこで，Autoencoder により入力データに対する損失を評価することにより，入力データがこれまでに頻繁に入力されたデータであるか，つまり正常値であるかどうかを判断できる．

3) ONLAD：論文[4]では，Autoencoder と OS-ELM の組み合わせにより，エッジデバイスにおいて異常検知を行うための手法を提案した．ONLAD では，Autoencoder において入力データ x に対するエンコーダの結果は $H = G(x \cdot a + b)$と表現し，デコーダによる復元結果は $y = H \cdot \beta$となる．その結果，逐次学習に対すi番目の学習データが到来した場合，ニューラルネットワークの入力と出力の損失値を用いることで異常検知を実現できる．

2.2　連合学習

連合学習は，個々のエッジデバイスのデータ自体を共有せずに，差分のパラメータ（ニューラルネットワークの重みやバイアス等）の情報のみを用いて，クラウドサーバ上で共通学習モデルを構築する機器学習手法である．本稿では，最もよく使用されている Federated Averaging（FedAvg）[14]アルゴリズムを用いて，各エッジデバイスにおけるニューラルネットワークの重みを集約することを想定する．m個の IoT 機器が連合学習に参加していると仮定する．FedAvg は共有モデルを構築するため，式 (3) に示したモデルの重み$\{\beta, P\}$の平均値をとる．したがって，連合学習後のニューラルネットワークの重みは，次のように更新される．

$$P = \sum_{i=1}^{m}\frac{|D_i|}{|D|}P_i,\ \beta = \sum_{i=1}^{m}\frac{|D_i|}{|D|}\beta_i \quad (5)$$

ここで，β_iとP_iはエッジデバイスiに対するローカルモデルのパラメータ，Pとβは連合学習された後の共通モデルのパラメータ，$|D_i|$はエッジデバイスiの学習データ数，$|D|$は連合学習に参加しているm個のデバイスのデータ数の合計である．逐次学習型のニューラルネットワーに ONLAD おける FedAvg を用いた連合学習アルゴリズムの構造を図 1 に示す．

Vol.2021-ARC-244 No.24
Vol.2021-SLDM-194 No.24
Vol.2021-EMB-56 No.24
2021/3/25

3. 提案手法

本章では，データの前処理を含む貪欲特徴量選択アルゴリズムと，連合学習に基づく侵入検知システムについて述べる．

3.1 徴量選択手法

NSL-KDD データセットはネットワークの通信に関する 41 個の特徴量と 1 個のラベルで構成されている．特徴量の内容は，シンボリック型の特徴量，数値型の特徴量とブーリアン型的特徴量がある．シンボリック型特徴量（例：*protocol*, *service*, *flag*）はニューラルネットワークで学習を行う際にデータの前処理が必要である．最もよく使用されているワンホットエンコーディングであるが，特徴数が増加するため，特徴量選択において最適な特徴量を統計的に解析する際に難しくなることがある．本稿では，シンボリック型の特徴量を値に応じたユニークな整数に変換して表現する．シンボリックの特徴量を非負の整数に変換した後，全ての学習データを[0,1]の範囲に正規化し，ニューラルネットワークで学習させる．

最適な特徴量のセットを導出するために全組み合わせパターンを試すことは難しいため，なるべく優れた検知精度を達成する特徴量セットを求めるために，本稿では貪欲特徴量選択アルゴリズムを提案する．概要をアルゴリズム 1 に示す．特徴量セットV_0が n 個の利用可能な特徴量を持ち，それらをすべて利用した場合の検出精度を$S(0)$として計算する（1 行目）．そして，各段階で局所的に最適な結果を採用する貪欲法のアルゴリズムを用いて特徴量の選択を行う．具体的には，デクリメンタル学習のように各段階で 1 つの特徴量を削除する（2 行目）．つまり，0 から特徴量数$Length(V_i)-1$までの各特徴インデックスjをテストし，削除すべき特徴量インデックスmを求める（3-7 行目）．mを見つけた後，F_mを特徴量セットV_iから削除し，検出精度を$S(i)$として記録する（8-9 行目）．削除された特徴量数iに対し，全ての算出した検出精度$S(i)$を評価し，最も高い検出精度を達成した特徴量セットVを選択する（11-12 行目）．加えて，選択された特徴量の数Nを$n-t$として記録する（13 行目）．

3.2 連合学習ベースの検知システム

本稿で提案する侵入検知システムは，ユーザデータのプライバシーを保護するために，局所的な異常検知モデルを集約する連合学習をベースにしている．連合学習は，以下のようなシナリオで有用と考えられる．

- 大量に分散されたデバイス: 多数のエッジデバイスが存在し，単一のエッジデバイスのトラフィックデータ数は少ない．
- 異種デバイス混在: 様々なデバイスが存在し，各デバイスのトラフィックデータは，デバイスの種類とネットワーク環境に依存する．

アルゴリズム 1: 貪欲特徴量選択

入力：トレーニングデータセット
出力：特徴量セットV

1. $S(0) = Accuracy(V_0)$
2. for $i = 1, 2, \dots, n-1$ //i: 削除される特徴量の数
3. for $j = 0, 1, \dots, Length(V_i) - 1$ //j:特徴量インデックス
4. $V_i' = Delete\_feature(V_i, F_j)$
5. $s(j) = Accuracy(V_i')$
6. end for
7. $m = argmax(s(j))$
8. $V_i = \text{Delete\_feature}(V_i, F_m)$
9. $S(i) = s(m)$
10. end for
11. $t = argmax(S(i))$
12. $V = V_t$
13. $N = n - t$ //N: 特徴量インデックス

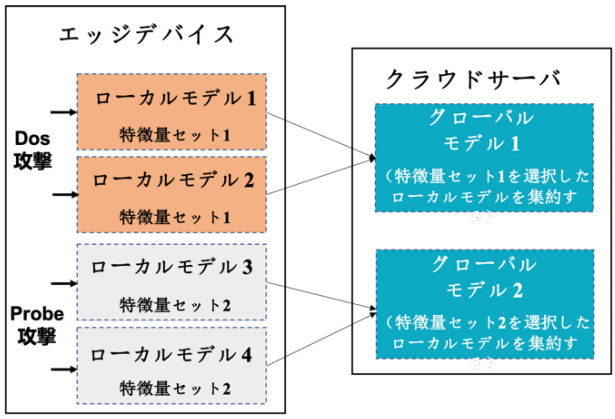

図 2：連合学習ベースの検知システムの概要

第 1 章で述べた「異常検出型」の検知システムでは，多種類の攻撃を高精度に検知するための学習モデルを構築することは困難である．さらに，種類が多いほど検出精度が低下する．そのため，各エッジデバイスで決定された特徴量セットに対し，連合学習を用いて複数の共通モデルを構築することを提案する．連合学習アルゴリズムとしては，共有のニューラルネットワークモデルを得るために，ONLAD におけるモデルの重みの平均を取る．この手法は，各 IoT 機器の限られた情報を利用しつつ，検出精度を向上させることに貢献すると考えられる．図 2 に，提案する連合学習ベースの検出システムの概要を示す．

4. 評価

提案手法の有効性を評価するために，逐次学習型のニューラルネットワーク ONLAD を用い，NSL-KDD データセッ

Vol.2021-ARC-244 No.24
Vol.2021-SLDM-194 No.24
Vol.2021-EMB-56 No.24
2021/3/25

トにける評価実験を行う．本章では，提案手法評価のための実験設定と実験結果を述べる．

4.1 実験設定

ニューラルネットワーク：ネットワーク攻撃の検出には，異常検出手法である ONLAD を利用する．このニューラルネットワークのハイパーパラメータ等は以下のように設定する．入力層と出力層の次元は学習された特徴量の数に応じて変化するが，隠れ層の次元は 64 のままで固定する．初期化パラメータは[-1,1]の範囲でランダムに生成され，逐次学習の初期学習データとして，バッチサイズ 64 で 1000 個の学習データを用意する．隠れ層と出力層の活性化関数は，それぞれ Sigmoid と Identity を使用する．

データセット： 本研究では，NSL-KDD データセット[7]により侵入検知の有効性を評価する．NSL-KDD データセットは，KDDCUP 1999 データセット[15]の更新版である．このデータセットは，セキュリティ攻撃の分野における様々な IDS 手法を評価するために広く利用されている．データセットには，表 1 に示すように Dos, Probe, R2L, U2R の 4 つの攻撃カテゴリを持つ．各データインスタンスには，同じ 41 個のネットワークトラフィック特徴量が含まれている．訓練データとテストデータの数を表 2 に示す．

連合学習フレームワーク：連合学習フレームワークをシミュレートするために，8 つのエッジデバイスを仮定する．正常データセットを 8 等分してそれぞれのデバイスに与え，また攻撃データセットは各攻撃カテゴリを 2 つずつのエッジデバイスに 2 等分して与えることとした．つまり，各エッジデバイスには 1/8 の個数の正常データが与えられ，またエッジデバイス 0, 1 に与える攻撃データは Dos 攻撃データ，デバイス 2, 3 には Probe 攻撃データ，デバイス 4, 5 には R2L 攻撃データ，デバイス 6, 7 には U2R 攻撃データが与えられると仮定する．

4.2 評価結果

特徴量選択の有効性評価：まず特徴量選択による検出精度に対しての有効性について評価・考察する．表 3 は全特徴量を用いてモデルを構築した場合，および特徴量選択手法により各攻撃向けに特徴量を絞り込みモデルを構築した場合の攻撃検出精度を比較したものである．表 3 に示す通り，特定の攻撃カテゴリに検出モデルを生成した方が，全特徴量を利用したモデルよりも高い検出精度が得られることがわかる．また，攻撃の種類によって，選択された特徴量の数や精度の向上率も異なることがわかる．表 4 に，単一の攻撃，および混合攻撃に対して選択された特徴量を示す．表では，“1”で示された部分の特徴量が各攻撃に対して選択されたことを示している．各攻撃の種類に対し，特徴量の選択が異なることがわかる．以上の結果より，攻撃の種類に応じて特徴量を適切に選択することで，検出精度を向上できることが明らかになった．

表 1: NSL-KDD データセットでの攻撃種類

攻撃種類	訓練データセット	テストデータセット
Dos	back, land, neptune, pod, smurf, teardrop	back, land, neptune, pod, smurf, teardrop, apache2, mailbomb, processtable, udpstorm, worm
Probe	ipsweep, nmap, portsweep, satan	ipsweep, nmap, portsweep, satan, mscan, saint
R2L	ftp_write, guess_passwd, imap, multihop, phf, warezclient, warezmaster, spy,	ftp_write, guess_passwd, imap, multihop, phf, warezmaster, sendmail, named, snmpgetattack, snmpguess, xsnoop, xlock, httptunnel
U2R	buffer_overflow, rootkit, perl, loadmodule	buffer_overflow, rootkit, perl, loadmodule, sqlattack, xterm, ps

表 2: 攻撃による訓練データ数とテストデータ数

	正常	Dos 攻撃	Probe 攻撃	R2L 攻撃	U2R 攻撃	混合攻撃
訓練データセット	67343	45927	11656	995	52	58630
テストデータセット	9711	7460	2421	2885	67	12833

表 3: 貪欲アルゴリズムを用いた特徴量選択による攻撃検出精度の比較

	Dos 攻撃	Probe 攻撃	R2L 攻撃	U2R 攻撃	混合攻撃
全特徴量を用いた場合の精度	63.4%	68.9%	78.3%	97.3%	49.4%
特徴量選択を用いた場合の精度（選択特徴量数）	89.1% (8)	94.5% (7)	84.8% (9)	99.5% (6)	78.7% (15)

各特定の攻撃，および混合攻撃の場合について，削除された特徴量数に対する検出精度の比較を図 3 に示す．図より特定の攻撃に特化して特徴量を絞り込むことで，多くの場合で検出精度が大幅に向上することがわかる．しかしながら，ある特徴量を削除すると精度が低下場合もあるため，特徴量は適切に選択する必要がある．提案する貪欲法により選択することで，適切な選択が行えるため，本手法は有効であると考えられる．

Vol.2021-ARC-244 No.24
Vol.2021-SLDM-194 No.24
Vol.2021-EMB-56 No.24
2021/3/25

表 4：単一・混合の攻撃における特徴量の選択

特徴量 No.	特徴量の名称	Dos 攻撃	Probe 攻撃	R2L 攻撃	U2R 攻撃	混合 攻撃
1	duration	0	0	0	0	0
2	protocol_type	1	0	0	0	0
3	service	0	0	1	0	0
4	flag	1	0	0	0	1
5	src_bytes	0	0	0	0	0
6	dst_bytes	0	0	0	0	0
7	land	0	0	0	0	1
8	wrong_fragment	8	0	0	0	0
9	urgent	0	0	1	0	0
10	hot	0	0	0	0	1
11	num_failed_logins	0	0	0	0	1
12	logged_in	0	1	0	0	1
13	num_compromised	1	0	0	1	1
14	root_shell	1	0	0	1	0
15	su_attempted	0	0	0	0	1
16	num_root	0	1	1	0	1
17	num_file_creations	0	0	0	1	1
18	num_shells	1	1	1	0	0
19	num_access_files	0	0	1	0	0
20	num_outbound_cmds	0	0	0	1	1
21	is_host_login	0	1	0	0	0
22	is_guest_login	0	0	1	0	1
23	count	0	0	0	0	0
24	srv_count	0	0	0	0	0
25	serror_rate	0	0	0	0	0
26	srv_serror_rate	0	1	1	0	0
27	rerror_rate	0	0	0	0	0
28	srv_rerror_rate	0	1	0	0	0
29	same_srv_rate	1	0	0	0	1
30	diff_srv_rate	0	0	0	0	0
31	srv_diff_host_rate	0	1	0	0	0
32	dst_host_count	0	0	0	0	0
33	dst_host_srv_count	0	0	0	1	1
34	dst_host_same_srv_rate	0	0	0	0	0
35	dst_host_diff_srv_rate	0	0	0	0	1
36	dst_host_same_src_port_rate	0	0	1	0	0
37	dst_host_srv_diff_host_rate	0	0	0	1	0
38	dst_host_serror_rate	1	0	0	0	1
39	dst_host_srv_serror_rate	0	0	0	0	0
40	dst_host_rerror_rate	0	0	1	0	0
41	dst_host_srv_rerror_rate	0	0	0	0	0

表 5：連合学習ベースの侵入検知の有効性

	自己学習	連合学習	
		シングル 共通モデル	複数 共通モデル
全特徴量利用	49%	45.8%	50.5%
特徴量選択	68%	70.4%	

連合学習の有効性評価：次に，特徴量の選択結果に基づいて連合学習を行った場合のネットワーク侵入検知システムの精度を評価する．攻撃検出精度の評価結果を表 5 に示す．比較のため，従来のデバイス自身の収集データのみを用いて学習した場合のデバイス間の平均精度（自己学習と表記）も示している．表より，連合学習により精度が向上している場合があることがわかる．連合学習により，他のデバイスのデータの特徴も利用することで検出モデルを構築することができる．そのため，自己学習アプローチに比べて連合学習アプローチの検知精度を高められる可能性がある．一方，単一の共通モデルを作成した場合は，各デバイスからのデータの多様性を考慮しないため，逆に精度が

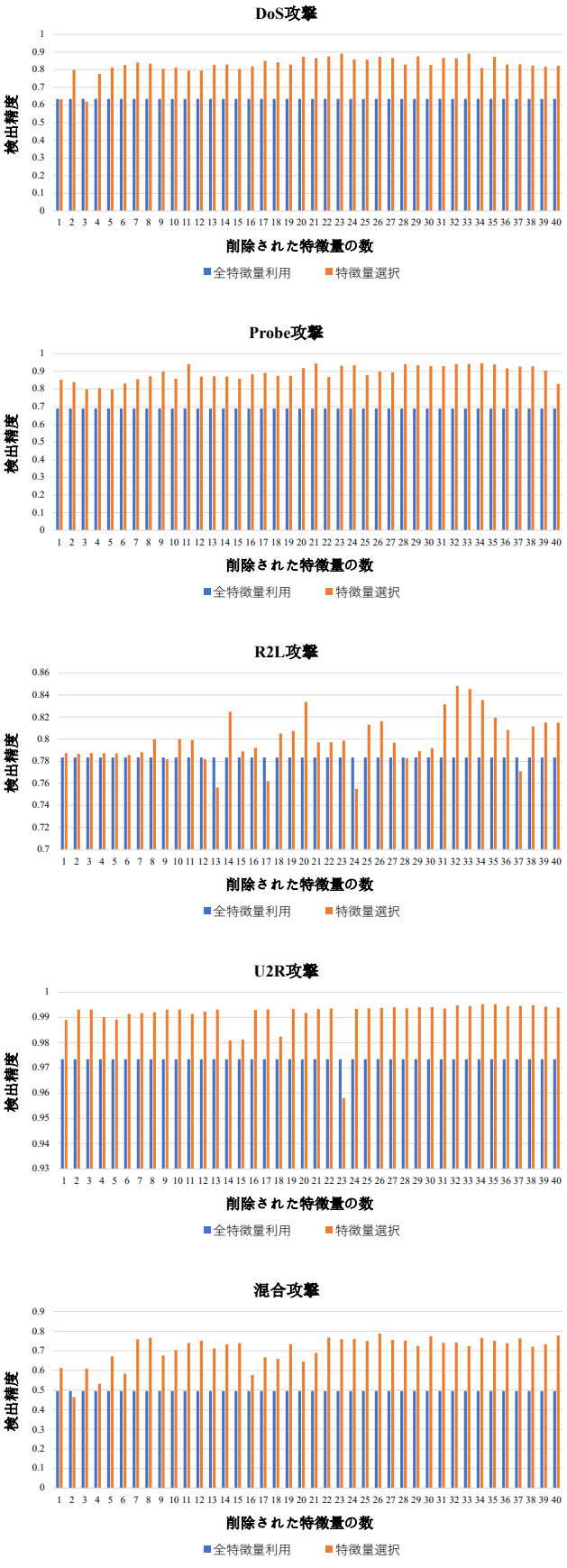

図 3: 単一・混合の攻撃の場合，削除された特徴量の数に対する検出精度

Vol.2021-ARC-244 No.24
Vol.2021-SLDM-194 No.24
Vol.2021-EMB-56 No.24
2021/3/25

悪くなっている．複数の共通モデルを構築することで，単一の共通モデルよりも高い検出精度が得られることが示された．本稿では特に，各エッジデバイスで特徴量の選択を行い，その結果をもとに複数の共通モデルを構築することが提案である．表より，その場合に 70.4%という最大の検知精度が得られている．なお，連合学習の評価では，データセットを各エッジに分割することで，エッジデバイス毎の学習データ数が減少するため，表 3 の結果に比べて検知精度の絶対値は低下しているが，データ数がより多くなれば問題とはならない．

5. おわりに

ネットワークに接続された IoT 機器の増加に伴い, IoT 機器で高次元の通信データを高速に分析し，不正アクセスをリアルタイム検知することが重要な課題になっている．本稿では，軽量な逐次学習型のニューラルネットワークにより, IoT 機器自体でトラフィックを監視しつつ，ネットワーク攻撃を検知するシステムを検討した．特に攻撃種類に応じて適切な特徴量を選択する貪欲アルゴリズムを提案した．NSL-KDD データセットを用いて検知精度の評価を行い，特定の攻撃に特化した学習モデルを，特徴量を適切に選択しつつ構築することで，検知精度が上がることが示された．また，特徴量に応じて連合学習を行うと検知精度がさらに向上することもわかった．

今後の課題としては，オンライン学習時の攻撃検知精度を評価することが挙げられる．本稿では，事前に用意したトレーニングデータを用いて，検知モデルを構築したが，IoT デバイスのトラフィックデータは通信環境に応じても変化する貯め，それに追随できるオンライン学習は有用であると考えられる．

謝辞 本研究の一部は JST CREST の研究プロジェクト JPMJCR20F2 の助成によるものである．

参考文献

[1] Drewek-Ossowicka, A., Pietrołaj, M., & Rumiński, J. (2020). A survey of neural networks usage for intrusion detection systems. *Journal of Ambient Intelligence and Humanized Computing*, 1-18.

[2] Gamage, S., & Samarabandu, J. (2020). Deep learning methods in network intrusion detection: A survey and an objective comparison. *Journal of Network and Computer Applications*, *169*, 102767.

[3] Ahmad, Z., Shahid Khan, A., Wai Shiang, C., Abdullah, J., & Ahmad, F. (2021). Network intrusion detection system: A systematic study of machine learning and deep learning approaches. *Transactions on Emerging Telecommunications Technologies*, *32*(1), e4150.

[4] Tsukada, M., Kondo, M., & Matsutani, H. (2020). A neural network-based on-device learning anomaly detector for edge devices. IEEE Transactions on Computers, 69(7), 1027-1044.

[5] Stiawan, D., Idris, M. Y. B., Bamhdi, A. M., & Budiarto, R. (2020). CICIDS-2017 dataset feature analysis with information gain for anomaly detection. IEEE Access, 8, 132911-132921.

[6] Kang, S. H., & Kim, K. J. (2016). A feature selection approach to find optimal feature subsets for the network intrusion detection system. Cluster Computing, 19(1), 325-333.

[7] NSL_KDD dataset: http://nsl.cs.unb.ca/NSL-KDD/

[8] Nguyen, T. D., Marchal, S., Miettinen, M., Fereidooni, H., Asokan, N., & Sadeghi, A. R. (2019, July). DÏoT: A federated self-learning anomaly detection system for IoT. In 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS), 756-767.

[9] Preuveneers, D., Rimmer, V., Tsingenopoulos, I., Spooren, J., Joosen, W., & Ilie-Zudor, E. (2018). Chained anomaly detection models for federated learning: An intrusion detection case study. Applied Sciences, 8(12), 2663.

[10] McMahan, B., & Ramage, D. (2017). Federated learning: Collaborative machine learning without centralized training data. Google Research Blog, 3.

[11] Liang, N. Y., Huang, G. B., Saratchandran, P., & Sundararajan, N. (2006). A fast and accurate online sequential learning algorithm for feedforward networks. IEEE Transactions on neural networks, 17(6), 1411-1423.

[12] Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. science, 313(5786), 504-507.

[13] Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2004, July). Extreme learning machine: a new learning scheme of feedforward neural networks. In 2004 IEEE international joint conference on neural networks (IEEE Cat. No. 04CH37541), 2, 985-990.

[14] McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017, April). Communication-efficient learning of deep networks from decentralized data. In Artificial Intelligence and Statistics, 1273-1282.

[15] KDD Cup 1999 dataset: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html