Vol.2021-DPS-188 No.18
Vol.2021-EIP-93 No.18
2021/9/10

勾配ブースティング決定木を用いたネットワーク侵入検知システムの提案

都留 悠哉[1,a)]　川上 朋也[1]

概要：情報セキュリティにおいて標的型攻撃の脅威が深刻化しており，対抗手段の一つとしてネットワーク侵入検知システム（Network-based Intrusion Detection System, NIDS）があげられる．NIDS はネットワークを監視し，不正な通信を検知する．NIDS に関するさまざまな研究が現在行われ，特に機械学習による高性能化が注目されている．そこで，本論文では勾配ブースティング決定木（Gradient Boosting Decision Tree, GBDT）を用いた NIDS を提案する．GBDT は教師あり学習の一つであり，高い汎化能力によって，未知のデータに対しても高い精度で判別できる．また，機械学習を用いた NIDS では学習や評価のために膨大な量のデータが必要であり，複数の NIDS 用データセットも公開されている．本論文では Kyoto 2016 Dataset を用いて提案システムの学習と評価を行い，既存システムより検知精度が向上することを確認した．

1. はじめに

独立行政法人情報処理推進機構の「情報セキュリティ 10 大脅威 2021」では組織を狙った標的型攻撃による被害が 2 位となっており，その脅威が深刻化している [1]．それら標的型攻撃から情報資産を守る方法の一つとして，ネットワーク侵入検知システム（Network-based Intrusion Detection System, NIDS）があげられる．NIDS はネットワークを監視し，不正な通信を検知する．一方，標的型攻撃の手段は多岐にわたり，新しい攻撃手法が日々出現している．そのため，機械学習を用いた侵入検知や異常検知に関する多くの研究が行われている．

機械学習を用いた NIDS では，学習や評価のために膨大な量のデータセットが必要である．複数の NIDS 用データセットが現在公開されており，例えば多田らは NIDS の高性能化のため，評価用データセットとして Kyoto 2016 Dataset を作成している [2]．文献 [2] では，ランダムフォレスト（Random Forest, RF），決定木（Decision Tree, DT），ナイーブベイズ（Naive Bayes, NB），サポートベクタマシン（Support Vector Machine, SVM），k 近傍法（k-Nearest Neighbor: k-NN），One Class SVM（OCSVM）の 6 種類の基本的な機械学習手法による分類精度の結果を示している．一方，機械学習手法には文献 [2] で用いられたもの以外にも多くの手法があり，特に教師あり学習の一つである勾配ブースティング決定木（Gradient Boosting Decision Tree, GBDT）[3] は汎化能力が高く，未知のデータに対しても高い精度で判別できる．

そこで，本論文では GBDT を用いた NIDS を提案し，Kyoto 2016 Dataset を用いて学習と評価を行う．提案システムの評価では，文献 [2] で示されている 6 手法と提案システムによる分類精度の結果を比較する．

以降，機械学習を用いた既存の NIDS と，学習や評価のためのデータセットについて 2 章で述べる．3 章では提案する NIDS で用いる GBDT と設計モデルについて述べ，4 章では Kyoto 2016 Dataset を用いた評価について述べる．最後に，5 章で本論文をまとめる．

2. 関連研究

2.1 機械学習を用いた NIDS

情報セキュリティにおいて標的型攻撃の手段は多岐にわたり，新しい攻撃手法が日々出現している．そのため，機械学習を用いた侵入検知や異常検知に関する研究が注目されている．

Ambusaidi らは，IDS に有効な特徴選択手法を考案し，その特徴選択手法と Least Square Support Vector Machine（LSSVM）を組み合わせた IDS を提案した [4]．また，Om らは，k-Means と k 近傍法，Naive Bayes を組み合わせたハイブリッド IDS を提案した [5]．Hosseini らは蟻コロニー最適化（Ant Colony Optimization, ACO）とホタルアルゴリズム（Firefly Algorithm）を組み合わせ，最適化

[1] 福井大学
3-9-1 Bunkyo, Fukui 910-8507, Japan
[a)] hb179073@g.u-fukui.ac.jp

したサポートベクター回帰（Support Vector Regression, SVR）を用いて，Denial of Service（DoS）侵入攻撃の検知手法を提案した [6]．Eskin らは，教師なし学習手法であるクラスタベース分類法と k 近傍法，One Class SVM を使用して，アノマリ検知による侵入検知手法を提案した [7]．RT らは，Software-defined Networking（SDN）技術によって制御されるネットワークが Distributed Denial of Service（DDoS）攻撃に対して脆弱であることをあげ，Support Vector Machine（SVM）を使用した DDoS 攻撃の検知手法を提案した [8]．Mukkamala らは，ニューラルネットワークと SVM を用いた IDS を構築して比較を行った [9]．Masarat らは，IDS のための改良型ランダムフォレストアルゴリズムを提案した [10]．Stein らは，遺伝的アルゴリズム（Genetic Algorithm, GA）を用いた特徴選択と決定木を組み合わせた侵入検知手法を提案した [11]．Amor らは，IDS における決定木とナイーブベイズの分類性能の比較を行った [12]．

2.2 NIDS のためのデータセット

機械学習を用いた NIDS では学習や評価のために膨大な量のデータが必要であり，現在，NIDS の学習，評価用データセットとして，複数のデータセットが公開されている．ここでは，一般公開されており，使用に当たっての許諾が不要なものについて述べる．

2.2.1 DARPA Intrusion Detection Data Sets

DARPA Intrusion Detection Data Sets は，MIT Lincoln Laboratory が作成し，1998 年から公開しているデータセットである．1998 年，1999 年，2000 年の 3 年分が公開されている．このデータセットは，実験用に作成されたネットワーク環境で正常な通信の中に意図的に悪性の通信を混入させて作られたものである．含まれる攻撃の種類としては，DoS 攻撃やバッファオーバーフロー攻撃などがある．通信データは TCPDUMP 形式で公開されているため，非常に柔軟に利用できるという特徴がある．

2.2.2 KDD Cup 1999 Data

KDD Cup 1999 Data は，University of California Irvine, Machine Learning Repository で公開されているデータセットである．このデータセットは，1998 DARPA Intrusion Detection Data Set の通信データをもとに作成されたものである．DARPA Intrusion Detection Data Sets とは異なり，通信データをセッション単位で扱い，各セッションについてセッションの前後関係をもとに算出した特徴量を追加して 42 次元のベクトルの形式で記録されている．セッションの前後関係から算出される特徴量には，過去 2 秒間のセッションという単位や同一ホスト間の過去 100 セッションという単位でセッションをひとまとめにし，集計して得られるものが含まれる．ペイロードはもちろん，IP アドレスやポート番号も含まれない．すべてのデータに正常か攻撃かを示すラベルが付けられ，攻撃の場合には攻撃の種類も示される．学習用と評価用にそれぞれデータが用意されており，評価用のデータには学習用に含まれない攻撃データが含まれるという特徴がある．

2.2.3 Kyoto 2006+ Dataset

Kyoto 2006+ Dataset は，当時 NIDS の評価用データセットとして広く用いられていた KDD Cup 1999 Data が古くなったことを受け，新たな NIDS の評価用データセットとして作成されたものである．通信データとして，京都大学に設置されているハニーポットのデータを使用して作成されたため，Traffic Data from Kyoto University’s Honeypots という名前で公開されている．公開されているのは，2006 年 11 月から 2009 年 8 月までの通信データから作成されたものである．データの作成方法は KDD Cup 1999 Data に準拠しており，通信データをセッション単位で扱いその前後関係から特徴量を算出して作成された．Kyoto 2006+ Dataset の特徴量は，KDD Cup 1999 Data の特徴量の一部である 14 種類と，独自に追加した 10 種類の特徴量を合わせた 24 種類である．

2.2.4 Kyoto 2016 Dataset

上記 3 つのデータセットは作成から期間が経過して最新の攻撃傾向を反映できていないことや，データの収集期間が短いといった問題がある．その問題を解決するため，Kyoto 2006+ Dataset の作成に使用されたハニーポットのデータを使用して新たに Kyoto 2016 Dataset [2] を作成された．

2.2.5 NSL-KDD Data set

前述の KDD Cup 1999 Data はデータが古い，冗長的である，データサイズが大きいなどの欠点が指摘されている．このため，上記の欠点を修正したのが，NSL-KDD Data set である．University of New Brunswick の Canadian Institute for Cybersecurity から提供されている．

2.2.6 CSE-CIC-IDS2018

Communications Security Establishment と Canadian Institute for Cybersecurity が提供するデータセット．Brute-force, Heartbleed, ボットネット, DoS, DDoS, Web 攻撃，および内部からのネットワークへの侵入という 7 つの異なる攻撃シナリオが含まれている．攻撃インフラには 50 台のマシンが含まれ，被害者の組織には 5 つの部門があり，420 台のマシンと 30 台のサーバーが含まれている．また，データセットにはキャプチャされたトラフィックから抽出された 80 の特徴とともに，各マシンのキャプチャされたシステムログが含まれている．

2.2.7 MWS Datasets

MWS Datasets [13, 14] は，マルウェア対策研究人材育成ワークショップ（Anti-Malware Engineering Workshop, MWS）で使用する研究用データセットである．以下の複数のデータセットから構成されている．

- Augma Dataset 2020 2021
  Web クライアントハニーポットで収集した攻撃通信データ
- FFRI Dataset 2013〜2021
  株式会社 FFRI セキュリティで収集したマルウェアの動的解析ログ，表層解析ログ
- Soliton Dataset 2018〜2021
  エンタープライズ向け EDR 製品 InfoTrace Mark II で収集したマルウェア動作ログ
- NICTER Dataset 2013〜2021
  サイバー攻撃観測・分析・対策システム NICTER で収集したダークネットトラフィックデータ，メールサーバに届いたダブルバウンスメールのデータ
- MWS Cup Dataset 2015〜2020
  MWS Cup 2015〜2020 参加チームにより作成されたデータセットや発表スライド，課題を解くにあたって作成したスクリプト
- BOS 2014〜2019
  総務省実証事業「サイバー攻撃解析・防御モデル実践演習の実証実験の請負」にて実施し，研究者コミュニティから提供された組織内ネットワークへの侵害活動を観測したデータ
- D3M (Drive-by-Download Data by Marionette) 2010〜2015
  研究者コミュニティから提供された Web 感染型マルウェアデータ
- PRACTICE (AmpPot) Dataset 2015
  インターネット上のオープンなサーバ（DNS，NTP 等）を踏み台にして通信を増幅させることでサービス妨害を行う分散反射型サービス妨害攻撃（DRDoS 攻撃）を観測したデータセット
- NCD in MWS Cup 2014
  MWS Cup 2014 会期中に収集したホワイトデータセット
- PRACTICE Dataset 2013
  総務省「国際連携によるサイバー攻撃予知・即応に関する実証実験」（略称： PRACTICE）の挙動観察システムで，マルウェアを長期観測した際の通信トラフィック（マルウェア感染後の通信挙動）を示すデータ
- CCC DATAset 2008〜2013
  マルウェア検体を収録したボット観測データ群であり，CCC 運営連絡会が運用するサイバークリーンセンターハニーポットで収集したマルウェア検体とウイルス対策ソフト 6 製品での検知名をリスト化したデータ

3. 提案システム

3.1 概要

本論文では，勾配ブースティング決定木（Gradient Boosting Decision Tree, GBDT）を用いた NIDS を提案する．GBDT は教師あり学習の一つであり，汎化能力が高く，未知のデータに対しても高い精度で判別できる．データ分析コンペティションの Kaggle [15] でも高いパフォーマンスを出す事例が多く，さまざまなアプリケーションで多用される．

3.2 勾配ブースティング決定木

GBDT とは Gradient（勾配降下法）と Boosting（アンサンブル学習），Decision Tree（決定木）を組み合わせた手法である．GBDT の基本となる 3 つの概念について述べる．

3.2.1 勾配降下法

機械学習の主な目的は，正確な予測を行うことである．予測が正確かどうかは誤差の大きさで判断することができる．つまり，誤差が小さいと予測が正確と置き換えることができる．そこで誤差を小さくする方法の 1 つに勾配降下法がある．勾配降下法は勾配を降下する最適化の手法で，精度の誤差を最適化する役割をもっている．誤差関数の勾配を下れば下るほど誤差が小さくなり，予測が正確になる．

3.2.2 アンサンブル学習

アンサンブル学習とは，精度の低い学習器を複数組み合わせて，精度を高くする手法である．アンサンブル学習の手法には，さまざまな手法があるが，その中でも GBDT に利用されているのが Boosting である．Boosting は，弱学習器を複数重ねてそれぞれがデータセットを学習する学習器を形成する．また，弱学習器同士は連なっていて，前の弱学習器の誤分類情報を優先的に学習し，弱学習器の数だけその弱学習器の誤分類情報を次の弱学習器に学習させる操作を行う．

3.2.3 決定木

木構造を用いて分類や回帰を行う機械学習の手法の一つである．決定木は，構築のために学習データが全てルートノードに集められる．そこで，そのデータの持つ特徴の中で集められたデータを一番よく分割する特徴と閾値の組を選ぶ．その特徴と閾値で分割後，それぞれのノードで分割を繰り返し行っていく．この手順で決定木は構築されていく．

3.3 設計モデル

提案する NIDS で使う GBDT モデルを図 1 に示す．図 1 は，弱学習器となる決定木の数が N 本の場合の GBDT を表す．実装には機械学習ライブラリである scikit-learn と，GBDT を実装するために XGBoost [16] を使用した．各パラメータは，XGBoost のデフォルト値として，弱学習器となる決定木の数は 400，深さは 6 とした．

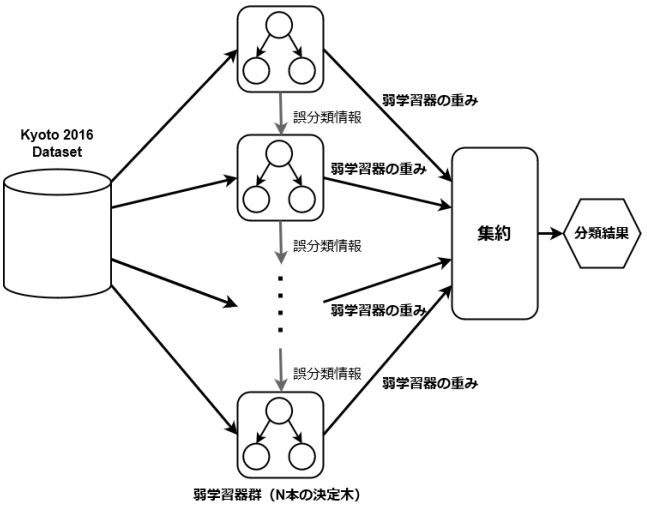

図 1 GBDT モデル

4. 評価

4.1 実験環境

多田ら [2] と同様に，Kyoto 2016 Dataset の 2006 年 11 月から 2008 年 12 月を 2 か月ごとに区切り，2006 年 11–12 月を期間 A，2007 年 1–2 月を期間 B というようにして期間 M まで作成した．そして，多田ら [2] と同様に基本 14 特徴量のうち，カテゴリ値特徴である “Service” と “Flag” を除いた 12 特徴量を，各期間からサイバー攻撃と正常な通信に分けて 1 万件ずつ無作為抽出した．抽出した期間 A のデータを GBDT に入力して学習させた後，期間 B のデータを分類した．同様に，期間 B のデータを学習させて期間 C を分類し，同様の操作を期間 L で学習して期間 M を分類するまで繰り返した．それぞれの期間の分類結果から，正解率（予測の正確さ，Accuracy），適合率（攻撃と予測し正解した割合，Precision），検知率（攻撃を攻撃として検知した割合，True Positive Rate, TPR），誤検知率（正常な通信を攻撃と誤検知した割合，False Positive Rate, FPR）の 4 種類の指標を算出した．なお，無作為抽出による結果のばらつきを防ぐため 10 回試行時の平均値を結果として使用した．各指標の定義式は以下の通りである．

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$TPR = \frac{TP}{TP + FN} \tag{3}$$

$$FPR = \frac{FP}{FP + TN} \tag{4}$$

ここで，TP（True Positive）は攻撃と予測して実際に攻撃だった件数，FP（False Positive）は攻撃と予測したが実際は正常だった件数，TN（True Negative）は正常と予測して実際に正常だった件数，FN（False Negative）は正常と予測したが実際は攻撃であった件数を意味する．

4.2 実験結果

各期間における GBDT の平均分類精度を表 2 に示す．また，GBDT と文献 [2] で示された 6 手法の全期間における平均分類精度を表 3 に示す．表 3 より，GBDT は 6 手法に比べて全ての指標においてより精度の高いモデルであることがわかる．また，全期間を通して誤検知率も減少させることができた．期間（G-H），（I-J）では他の期間と比べ平均誤検知率が高い．これは，評価データに学習データには含まれない正常サンプルが含まれるためである．学習データに含まれる評価データの正常サンプルは，当然正しく正常サンプルであると分類できる．しかしながら，学習データに含まれない評価データの正常サンプルは分類器の作成時には未知の存在であり，境界を決定する際に考慮されない．このため，分類器は学習データに含まれない評価データの正常サンプルを正しく分類することができず，誤検知が増加したと考えられる．

5. まとめ

本論文ではネットワーク侵入検知システム（Network-based Intrusion Detection System, NIDS）の高性能化として，勾配ブースティング決定木（Gradient Boosting Decision Tree, GBDT）を用いた NIDS を提案した．GBDT は教師あり学習の一つであり，高い汎化能力によって，未知のデータに対しても高い精度で判別できる．提案システムは XGBoost を用いて実装し，Kyoto 2016 Dataset を用いて学習と評価を行った．実験結果から，GBDT を用いた提案システムが先行研究 [2] の 6 手法より誤検知率を抑えつつ，検知精度も高いことを確認した．

今後の課題として，学習や評価のデータにプライバシ情報が含まれることを想定し，分散型 NIDS の検討があげられる．1 章や 2.2 節でも述べたとおり，機械学習には膨大な量のデータが必要となる．そのような大量のデータは大企業や研究機関等大きな組織が所有することが多いが，プライバシの保護や組織間の競合等の観点から外部に公開されることはほとんどない．しかし，組織が所有するデータにも量や特徴において限りがあり，機械学習で訓練したモデルを構築してもその汎化性能に限界がある．この課題の改善には，組織が所有するデータの他に個々の収集・所有するデータの使用が有望であるが，プライバシ保護やリソースの制限の観点で困難である．この課題の改善として，学習可能な程度にノイズを加えてデータ侵害を困難にすることでプライバシの保護を実現して公開する手法（差分プライバシ）があるが，ノイズを加えるためデータ量が増加する．一方，Google は Federated Learning という呼ぶ手法を提案している [17]．Federated Learning は，複数のデバイスが分散して持つデータをモデルの訓練に使用する手法である．この手法の利点として，ローカルにあるデータをデバイス上で訓練し，デバイスごとに訓練したモデルを外

表 1 Kyoto 2016 Dataset から用いた 12 の特徴量

名前	詳細
(1) Duration	セッションの長さ [秒]
(2) Source bytes	送信バイト数
(3) Duration bytes	受信バイト数
(4) Count	過去 2 秒間のセッションのうち，現在のセッションと宛先 IP アドレスが同じ数
(5) Same_srv_rate	(4) で該当したセッションのうち，現在のセッションとサービスの種類が同じ割合
(6) Serror_rate	(4) で該当したセッションのうち，“SYN” エラーが起こった割合
(7) Srv_serror_rate	過去 2 秒間のセッションで現在のセッションとサービスの種類が同じセッションのうち，“SYN” エラーが起こった割合
(8) Dst_host_count	宛先ポートが同じ過去 100 セッションのうち，現在のセッションと送信元 IP アドレスと宛先 IP アドレスが同じ数
(9) Dst_host_srv_count	宛先ポートが同じ過去 100 セッションのうち，現在のセッションと宛先 IP アドレスとサービスの種類が同じ数
(10) Dst_host_same_src_port_rate	(8) で該当したセッションのうち，現在のセッションと送信元ポートが同じ割合
(11) Dst_host_serror_rate	(8) で該当したセッションのうち，“SYN” エラーが起こった割合
(12) Dst_host_srv_serror_rate	(9) で該当したセッションのうち，“SYN” エラーが起こった割合

表 2 各期間における GBDT の平均分類精度

期間	正解率	適合率	検知率	誤検知率
A-B	94.55%	98.83%	90.18%	1.07%
B-C	96.14%	93.51%	99.17%	6.88%
C-D	98.75%	98.44%	99.07%	1.57%
D-E	92.96%	95.05%	90.62%	4.71%
E-F	97.53%	96.26%	98.91%	3.84%
F-G	98.67%	98.29%	99.07%	1.72%
G-H	91.81%	86.34%	99.38%	15.75%
H-I	97.82%	98.67%	96.94%	1.31%
I-J	89.00%	83.14%	98.92%	20.92%
J-K	98.64%	98.76%	98.53%	1.24%
K-L	98.92%	98.96%	98.87%	1.04%
L-M	97.69%	98.12%	97.23%	1.86%

表 3 GBDT と 6 手法の平均分類精度

GBDT 平均	96.04%	95.36%	97.24%	5.16%
RF 平均 [2]	94.68%	94.46%	95.79%	6.43%
DT 平均 [2]	94.33%	94.99%	94.07%	5.42%
NB 平均 [2]	84.02%	83.07%	86.88%	18.84%
SVM 平均 [2]	91.67%	92.58%	91.13%	7.78%
k-NN 平均 [2]	82.88%	81.48%	85.69%	19.93%
OCSVM 平均 [2]	81.44%	81.27%	82.81%	19.94%

部で統合するため，プライバシ情報を含むデータを外部に提供する必要がないことがあげられる．さらに，差分プライバシと比べて低通信負荷で機械学習を実現できる．プライバシ情報の有無に依存しない機械学習が可能になれば，多様で膨大なデータに対応可能なモデルを獲得することが容易になり，汎化性能の高いモデルによって様々な分野での応用範囲が広がると期待されている．筆者らは今後，Federated Learning を用いて，NIDS がそれぞれ独立したデータを使って学習し，NIDS の連携により，全てのデータを使用した学習と同程度の性能を達成するシステムの提案に取り組む予定である．

参考文献

[1] 独立行政法人情報処理推進機構：情報セキュリティ 10 大脅威 2021. Available at https://www.ipa.go.jp/security/vuln/10threats2021.html (accessed Aug. 18, 2021).
[2] 多田竜之介，小林良太郎，嶋田 創，高倉弘喜：NIDS 評価用データセット：Kyoto 2016 Dataset の作成，情報処理学会論文誌， Vol. 58, No. 9, pp. 1450–1463 (2017).
[3] Friedman, J. H.: Greedy Function Approximation: A Gradient Boosting Machine, *The Annals of Statistics*, Vol. 29, No. 5, pp. 1189–1232 (2001).
[4] Ambusaidi, M. A., He, X., Nanda, P. and Tan, Z.: Building an Intrusion Detection System Using a Filter-Based Feature Selection Algorithm, *IEEE Transactions on Computers*, Vol. 65, No. 10, pp. 2986–2998.
[5] Om, H. and Kundu, A.: A Hybrid System for Reducing the False Alarm Rate of Anomaly Intrusion Detection System, *Proceedings of the 1st International Conference on Recent Advances in Information Technology (RAIT 2012)*, pp. 131–136 (2012).
[6] Hosseini, Z. S., Chabok, S. J. S. M. and Kamel, S. R.: DOS Intrusion Attack Detection by Using of Improved SVR, *Proceedings of the 2015 International Congress on Technology, Communication and Knowledge (ICTCK 2015)*, pp. 159–164 (2015).
[7] Eskin, E., Arnold, A., Prerau, M., Portnoy, L. and Stolfo, S.: A Geometric Framework for Unsupervised Anomaly Detection, *Applications of Data Mining in Computer Security*, Vol. 6, pp. 77–101 (2002).
[8] RT, K., Selvi, S. T. and Govindarajan, K.: DDoS Detection and Analysis in SDN-Based Environment Using Support Vector Machine Classifier, *Proceedings of 6th International Conference on Advanced Computing (ICoAC 2014)*, pp. 205–210 (2014).
[9] Mukkamala, S., Janoski, G. and Sung, A.: Intrusion Detection Using Neural Networks and Support Vector Machines, *Proceedings of the 2002 International Joint Conference on Neural Networks (IJCNN 2002)*, pp. 1702–1707 (2002).
[10] Masarat, S., Sharifian, S. and Taheri, H.: Modified Parallel Random Forest for Intrusion Detection Systems, *The Journal of Supercomputing*, Vol. 72, pp. 2235–2258 (2016).

[11] Stein, G., Chen, B., Wu, A. S. and Hua, K. A.: Decision Tree Classifier for Network Intrusion Detection with GA-Based Feature Selection, *Proceedings of the 43rd Annual Southeast Regional Conference (ACM-SE 43)*, Vol. 2, pp. 136–141 (2005).
[12] Amor, N. B., Benferhat, S. and Elouedi, Z.: Naive Bayes vs Decision Trees in Intrusion Detection Systems, *Proceedings of the 2004 ACM symposium on Applied computing (SAC 2004)*, pp. 420–424 (2004).
[13] マルウェア対策研究人材育成ワークショップ（MWS）：MWS Datasets. Available at http://www.iwsec.org/mws/datasets.html (accessed Aug. 18, 2021).
[14] Hatada, M., Akiyama, M., Matsuki, T. and Kasama, T.: Empowering Anti-malware Research in Japan by Sharing the MWS Datasets, 情報処理学会論文誌，Vol. 56, No. 9 (2015).
[15] Kaggle Inc.: Kaggle. Available at https://www.kaggle.com/ (accessed Aug. 18, 2021).
[16] Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2016)*, pp. 785–794 (2016).
[17] McMahan, H. B., Moore, E., Ramage, D., Hampson, S. and y Arcas, B. A.: Communication-Efficient Learning of Deep Networks from Decentralized Data, *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS 2017)* (2017).