Vol.2019-MPS-122 No.14
2019/3/1

勾配ブースティング木を用いた Kyoto 2016 Dataset に対する攻撃検知精度の向上に関する研究

寺戸綾佑†，林田守広†

概要：サイバー攻撃に対する防御法として，ネットワークベース侵入検知システム（Network-based Intrusion Detection System : NIDS）による攻撃の検知が挙げられる．本研究では，NIDS 評価用データセット Kyoto 2016 Dataset に対して勾配ブースティング木を用いることで，サイバー攻撃の検知精度向上を目指す．

Improvement of Intrusion Detection Accuracy on Kyoto 2016 Dataset Using Gradient Boosting Tree

RYOSUKE TERADO†，MORIHIRO HAYASHIDA†

1. はじめに

近年，技術の進歩によって IT 化が促進されている一方で，サイバー攻撃件数の増加が問題視されている．特に，発電所や大規模工場，医療，交通，政府関連施設といった最重要インフラへのサイバー攻撃は国民のライフラインを途絶えさせる可能性があるため，これらを支えるシステムにおけるサイバーセキュリティの強化が進められている．

サイバー攻撃に対する防御法として，ネットワーク型侵入検知システム（Network-based Intrusion Detection System : NIDS）による検知があげられる．多田ら[1]は，機械学習による NIDS の高性能化のために NIDS 評価用データセットとして Kyoto 2016 Dataset を作成し，ランダムフォレスト（Random Forest : RF）[2]，決定木，ナイーブベイズ（Naïve Bayes），サポートベクトルマシン（Support Vector Machine : SVM），k 近傍法，One Class SVM の 6 種類の機械学習手法を適用した場合の分類精度に関する結果を示した．

本研究では，以上 6 手法に比べてより高い汎化能力を持つとされる勾配ブースティング木（Gradient Boosting Tree : GBT）[3]を用いることで，サイバー攻撃の検知精度向上を目指す．

2. 学習及び検証用データセット

2.1 Kyoto 2016 Dataset

本研究で用いる Kyoto 2016 Dataset は，先代の Kyoto 2006+ Dataset が古くなったことや，より巧妙になっている攻撃に対応するため以下の特徴を考慮して作成された[1].

(1) 最新の攻撃傾向を含む
(2) 実際のトラフィックから作成（ハニーポットデータ）
(3) データ収集期間が長い（2006 年 11 月から 2015 年 12 月までのおよそ 10 年間）

10 年間で観測されたセッション数は約 8 億セッションで，そのうち正常なセッションが約 1.6 億セッション，悪性なセッションが約 6.5 億セッションである．

2.2 特徴量

本研究で考慮する特徴量は次に示す 12 特徴量である．

(1) セッションの長さ[秒]
(2) 送信バイト数
(3) 受信バイト数
(4) 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) (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) (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) 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) 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) 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) (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) (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) (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

3. 実験方法

図 1 に GBT モデルを示す．図 1 は，弱学習器となる決定

† 松江工業高等専門学校
National Institute of Technology, Matsue College

ⓒ 2019 Information Processing Society of Japan

木の数が N 本の GBT を表す．本研究では弱学習器となる決定木の数は 400，深さは 3 とし，実装には Python の機械学習ライブラリである scikit-learn を使用した．

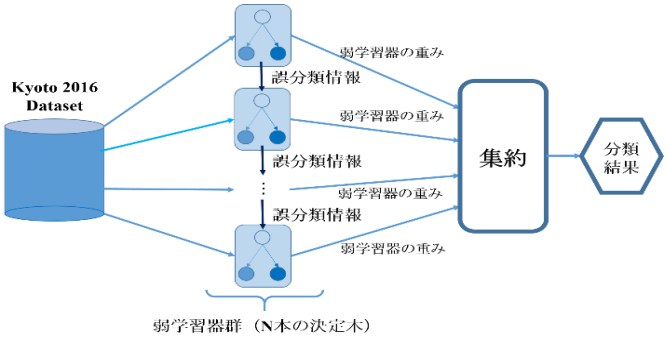

図 1　GBT モデル

多田ら[1]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，2.2 節に示した 12 特徴量を，各期間からサイバー攻撃（正例）と正常なアクセス（負例）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 GBT 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．また，分類にかかる時間を計測した．

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（正常アクセス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．

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$TPR = \frac{TP}{TP + FN} \quad (3)$$

$$FPR = \frac{FP}{TN + FP} \quad (4)$$

ここで，TP は攻撃と予測して実際に攻撃だった件数，FP は攻撃と予測したが実際は正常だった件数（第 1 種の誤り），TN は正常と予測して実際に正常だった件数，FN は正常と予測したが実際は攻撃であった件数（第 2 種の誤り）を意味する．

4. 実験結果

表 1 に各期間における GBT の平均分類精度と実際に先行研究[1]で示された RF の平均分類精度を参考として示す．GBT は RF 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．期間ごとにみると I-J 間の誤検知率は 36.02%から 26.90%に減少しており，平均値を比べても GBT では誤検知率を減少させることができた．しかし，GBT と RF の正解率に対して有意水準 5%の片側 t 検定を行った結果，0.2431 と有意差は認められなかった．また，テストデータの分類には，GBT では 0.0671±0.0106 秒，RF では 0.113±0.0201 秒の時間を要した．

表 1　各期間における GBT の平均分類精度と RF の平均分類精度（参考）

期間	正解率	適合率	検知率	誤検知率
A-B	95.21%	92.64%	92.32%	1.77%
B-C	96.51%	99.07%	99.12%	6.09%
C-D	98.52%	98.62%	98.52%	1.58%
D-E	93.94%	93.35%	93.23%	5.26%
E-F	97.55%	98.59%	98.62%	3.52%
F-G	98.68%	99.13%	99.24%	1.78%
G-H	91.87%	99.16%	99.29%	15.55%
H-I	97.72%	96.97%	96.82%	1.48%
I-J	86.03%	86.58%	98.96%	26.90%
J-K	98.86%	98.95%	98.80%	1.23%
K-L	98.96%	98.86%	99.01%	0.94%
L-M	97.79%	97.40%	97.38%	1.79%
GBT 平均	95.97%	96.61%	97.61%	5.66%
RF 平均[1]	94.68%	94.46%	95.79%	6.43%

5. まとめ

GBT による分類を行い，先行研究で最も分類精度の高かった RF との比較を行った．各結果から，GBT を適用した場合の正解率，適合率，検知率は向上し，誤検知率は低下したため，Kyoto 2016 Dataset に対して攻撃検知精度向上を達成した．また，期間 G-H，I-J では，学習データには含まれない正常サンプルを検証データ内に多く含んでいることが先行研究[1]で確認されており，本研究でも同様の傾向が見られ，誤検知率が高くなった．

本研究では，正解率における有意水準 5%の t 検定による有意差を認めるには至らなかったが，RF に比べて分類にかかる時間を約半分に抑えながら全ての指標において RF を超える精度のモデルを構築できた．今後の課題としては，過去に受けた攻撃と同様の攻撃に対する検知精度の向上のために新たな特徴量を検討すること，および，過去に経験していない攻撃への対処方法の検討が挙げられる．

参考文献

[1] 多田竜之介ら: 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).
[2] Breiman, L.: Random Forests, *Machine Learning*, Vol. 45, pp. 5-32 (2001).
[3] Friedman, J. H.: Greedy Function Approximation: A Gradient Boosting Machine, *The Annuals of Statistics*, Vol. 29, No. 5, pp. 1189-1232 (2001).

ⓒ 2019 Information Processing Society of Japan