5ZA-05

機械学習を用いた攻撃検知に関する学習手法の精度評価

平野　誠†　　八槇　博史†

東京電機大学†

1. はじめに

近年，コンピュータ及びネットワークを利用した犯罪による被害件数が増加している．この犯罪にはID やパスワードを盗み取る，サービスの利用を妨害するなど様々な種類が存在し，その中で最も悪質なものの１つに「標的型攻撃」がある．また，電子メールを使った手口が多く見受けられる．この手口で，攻撃者は標的となる組織のユーザにマルウェアを添付した電子メールを件名・内容を偽って送信する．電子メールを受け取ったユーザがそれをうっかり開いてしまうとマルウェアが同時に実行され，情報収集，改竄などの活動が行われる．近年，この標的型攻撃に用いられるマルウェアは攻撃通信を大量には発しなくなった．これによりほとんどすべてが正常な通信である中から一部の異常を発見する必要が生じた．また，標的となる組織に合わせた専用の攻撃プログラムが用いられるため，過去にあった攻撃を発見する手法（パターンマッチング）では攻撃を検知することができない．

そこで，本研究では機械学習を用いた攻撃検知を行う．機械学習の手法にはニューラルネットワークを使用する多層パーセプトロンを用い，学習・推定用データセットには京都大学に設置されたハニーポットのトラフィックデータである Kyoto 2016 Dataset を使用した．本稿では，学習手法と評価実験の結果，及びその考察を行う．

2. 使用した学習アルゴリズム

本実験では，モデルの学習アルゴリズムに多層パーセプトロン（Multilayer perceptron）を用いた．多層パーセプトロンとは，複数の単純パーセプトロンを繋いで多層構造にしたニューラルニューラルネットワークである．ここで，単純パーセプトロンは，入力データxと重みyをかけた結果がある閾値θを上回った場合は１，そうでない場合は 0 を出力する．また，出力結果と欲しいデータ（教師データ）から学習し重みを更新する．多層パーセプトロンは，単純パーセプトロンでは不可能であった線形分離できない問題に対して適用することができるため，より様々な課題に対して用いられている[1]．

Accuracy Evaluation of a Machine-Learning method on Attack Detection
†Tokyo Denki University

3. 使用したデータセット

本実験では，Kyoto 2016 Dataset を学習及び推定用データとして使用した．これは，京都大学内に設置されたハニーポットのトラフィックデータであり，2006 年 11 月から 2015 年 12 月にかけて収集されたデータが存在する．本実験では，2015 年 1 月から 12 月の１日ごとのデータをすべて使用した．

各々のデータには，セッションの長さからエラーが起こった割合まで 24 の特徴量が存在する．ここでは，表１に示す 12 の特徴量を使用した．クラスラベルは，1，-1，-2，の３種類存在し，1 が正常，-1 が既知攻撃，-2 が未知攻撃を表す[2]．

表1: 使用した特徴量

属性名	概要
Duration	セッションの長さ（s）
Source Bytes	送信バイト数
Destination Bytes	受信バイト数
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
Same_srv_rate	Count 特徴のセッションで，現在のセッションとサービスの種類が同じ割合
Serror_rate	Count 特徴のセッションで“SYN”エラーが起こった回数
Srv_serror_rate	“SYN”エラーが起こった割合
Dst_host_count	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
Dst_host_srv_count	現在のセッションと宛先 IP アドレスとサービス種類が同じ数
Dst_host_same_src_port_rate	Dst_host_count 特徴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

Dst_host_serror_rate	Dst_host_count 特徴で該当したセッションのうち，“SYN”エラーが起こった割合
Dst_host_srv_serror_rate	Dst_host_srv_count 特徴で該当したセッションのうち“SYN”エラーが起こった割合

4. 実験手法及び結果

本研究は，攻撃検知システムの構築及び評価を目的とする．また，実験では多層パーセプトロンを用いた学習・各種精度の算出を行った．

本手法は学習フェーズと推定フェーズに分かれる．学習フェーズでは，はじめに入力ファイルより学習に使用する箇所の抜き出し・データ整形をし，入力データを作成する．次に，抽出したデータを多層パーセプトロンによって学習させる．ここで，多層パーセプトロンの入力層は 12，中間層は 16，出力層は 2 とした．また，batch_size は 100，epochs は 50 とし，callbacks には，Early Stopping を使用した．batch_size は，1 回に学習するデータのまとまり（データサイズ），epochs は，同じデータを何回繰り返して学習させるかを表す．Early Stopping は，学習が進んで精度の向上がこれ以上見込めないと判断されたとき，その時点で epochs の数を無視して学習を打ち切る Keras の関数である．推定フェーズでは，学習済の学習器を用いて各種精度を算出する．学習・推定手法図を図 1 に示す．

本実験では，推定フェーズにおいて Accuracy（正解率），Precision（適合率），Recall（再現率），F_measure（F 値）の 4 つを各々算出した．Accuracy は，正や負と予測したデータのうち，実際にそうであるものの割合であり，Precision は，正と予測したデータのうち，実際に正であるものの割合である．また，Recall は実際に正であるもののうち，正であると予測されたものの割合であり，F_measure は，Precision と Recall の調和平均を表す．真陽性，偽陽性，偽陰性，真陰性をそれぞれ，TP，FP，FN，TN とすると，各種精度は以下の式で表される．

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

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

$$Recall = \frac{TP}{TP + FN}$$

$$F\_measure = \frac{2Recall \cdot Precision}{Recall + Precision}$$

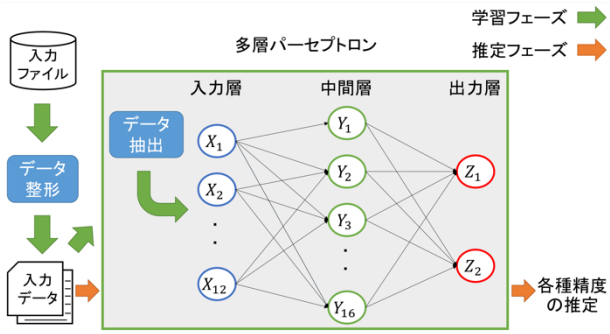

図 1： MLP による学習・推定

実験の結果，1 月から 12 月において，月ごとの各種平均精度が 99%となった．

5. まとめ及び今後の課題

本研究では，機械学習を用いた攻撃検知を行う．また，多層パーセプトロンを用いた各種精度の算出を行った．結果としては，月ごとの各種平均精度が 99%となった．しかし，この結果には疑問が残る．この精度が本当に正確なものであれば，ほぼすべてを正しく推定できることを意味する．しかし，一方で過学習の可能性も存在する．過学習とは，学習時のデータに偏りが生じることで，未学習のデータを正しく判別することのできない状態である．本実験では，ある 1 日のデータを学習・推定用に分割して精度の算出を行った．しかし，同時系列のデータでは特徴が似通っているため，推定用のデータが学習させているデータとほぼ同じものとなり，精度が極端に上がってしまったと考えられる．この問題の対策としては，違う時系列（別の日の）データを推定用に用いるということが挙げられる．また，プログラムの見直しも必要だと考えられる．

今後は，多層パーセプトロンによる精度の算出を別の時系列データを用いた上で再度行い，その後，アンサンブル学習に用いる学習法を選定する．それが終わり次第，アンサンブル学習による学習・推定及び異常攻撃検知システムの構築・評価を行う．

参考文献

[1] Rumelhart, David E., Geoffrey E. Hinton, and R. J. Williams. “Learning Internal Representations by Error Propagation”. Parallel distributed processing: Explorations in the microstructure of cognition, Volume 1: Foundation. MIT Press. 1986.

[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, Sep. 2017.