疎構造学習および グラフ畳み込みニューラルネットワークによる異常検知

熊谷　将也 †　　松本　亮介 †
†さくらインターネット株式会社

1　はじめに

我々が普段利用するインターネットは，スマホや IoT デバイスの急速な普及に伴い，今や欠かすことができない社会基盤となっている [1]．一方，サイバー攻撃の脅威も年々急速に増加しており，サイバー攻撃対策の必要性が高まってきている [2]．特に最近では，サイバー攻撃対策の 1 つとして機械学習を利用した侵入検知システム (IDS: Intrusion Detection System) の研究が盛んに行われている．すでに多くの機械学習手法の適用が提案されており，高い検知率が得られている [3]．ところが，深層学習をはじめとする機械学習によって生成されるモデルは，ブラックボックスになっていることが多く，異常なトラフィックであると判断した理由の説明が困難である．もし「どのような異常」で「なぜ異常と判断したのか」を説明するができれば，運用上最終的に判断する人間の助けとなり，検知した異常に対する具体的な対応がしやすくなると予想される．

そこで本研究では，トラフィックデータをグラフ構造化することに着目し，異常となった要因を部分グラフの可視化によって説明する機械学習型 IDS を提案する．

2　提案手法

2.1　データセットと前処理

提案手法（図 1）では，ラベルが付与されたパケットキャプチャ型のトラフィックデータが必要であるため，DARPA1998 データセットを使用した [4]．DARPA1998 は，仮想的に構築した空軍基地ネットワークに対するサイバー攻撃をシミュレートし，Tcpdump でトラフィックデータを収集したデータセットである．トラフィックデータの中から説明変数となり得る情報（送受信 IP や TCP/UDP ポート，TCP フラグやパケットサイズなど）を複数選択し，各変数でフィルタした 1 分単位のパケット数を時系列特徴量として用意した．ここでは，得られた各時系列特徴量の変化率から異常を検知するため，それぞれを対数差分系列に変換し，標準化を施した．またトラフィックの変化は，輻輳や異常の影響がなければ，短い時間スケールにおいて平均や分散が大きく変化しないとされている [5]．今回使用する時系列特徴量も特定の確率分布に従うデータとして仮定した．

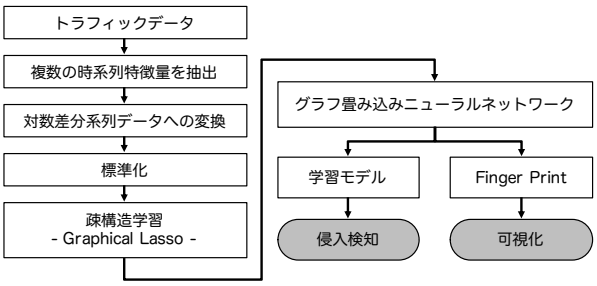

図 1: 提案手法の概略図

2.2　Graphical Lasso

各時系列特徴量間に存在する直接相関関係のグラフ構造を求める（構造学習）ため，多変量正規分布を想定したガウス型グラフィカルモデルを利用した．ガウス型グラフィカルモデルにおいて，精度行列 Λ は変数間の直接相関を表し，構造学習はこの Λ を推定する問題と考えることができる．ただし，データにはノイズが含まれていることが多く，そのノイズの影響を排除するため，疎な Λ を求める工夫が必要となる．そこで，疎なガウス型グラフィカルモデルの学習を可能にする手法である，Graphical Lasso に着目した．Graphical Lasso は，次の L_1 正則化項付きの最適化問題を解き，疎な Λ を推定する疎構造学習手法である [6, 7].

$$\arg\max_{\Lambda}(\ln\det\Lambda - \mathrm{tr}(S\Lambda) - \rho\|\Lambda\|_1) \tag{1}$$

ここで, S は標本共分散行列, tr は行列の対角和, det は行列式を表す．ρ は正則化パラメータであり，どの程度を Λ を疎な構造にするかを決定する．$\|\Lambda\|_1$ は $\sum_{i,j=1}^{M}|\Lambda_{i,j}|$ により定義される．

ここでは，各時系列特徴量を 30 分単位でグラフ化し，1 分毎に更新するグラフの時系列データを作成した．

2.3　グラフ畳み込みニューラルネットワーク

グラフ畳み込みニューラルネットワーク（GCNN: Graph Convolutional Neural Network）は，画像を入力とした問題に有効な CNN と同様の処理を，グラフ構造を入力とした問題にも対応させるために研究されてき

Anomaly detection by the method combined with sparse structure learning and graph convolutional neural network
†Masaya Kumagai　†Ryosuke Matsumoto
†SAKURA Research Center, SAKURA Internet Inc.

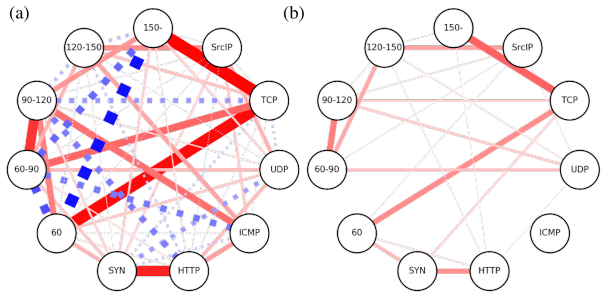

図 2: トラフィックデータから作成したグラフ構造：Graphical Lasso の (a) 適用前および (b) 適用後

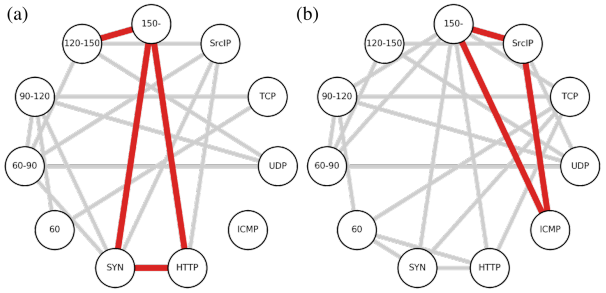

図 3: 異常の要因となった部分グラフの一例：(a) 異常 1 , (b) 異常 2

た手法である．ここでは特に，化学分野において報告された Neural Fingerprint（NFP）に着目した．NFP は，分子構造をグラフと捉えて学習することで，水溶度や毒性の高精度な予測を可能にしている [8]．また，NFP では Fingerprint（部分グラフの集合）を生成することができ，各部分グラフが予測結果にどの程度影響したかを表すことができる．そのため，毒性の学習であれば毒性に最も影響した部分グラフを特定，可視化することができる．そこで，元の DARPA1998 データセットの正常および異常のラベルを利用して，グラフの時系列データにもラベルを付与し，NFP で学習することによって異常状態の予測および異常の要因となる部分グラフの可視化を行なった．

3 実験結果

3.1 トラフィックデータのグラフ化

トラフィックデータから作成したグラフ構造を図 2 に示す．グラフの描画には，精度行列から算出した偏相関係数を利用した．

$$r^{i,j} \equiv -\frac{\Lambda_{i,j}}{\sqrt{\Lambda_{i,i}\Lambda_{j,j}}} \tag{2}$$

実線は $r^{i,j} > 0$，点線は $r^{i,j} < 0$ を表し，線の太さは $r^{i,j}$ の絶対値の大きさを表す．Graphical Lasso を適用する前のグラフ構造（図 2(a)）は，全ての特徴量間に相関が存在することを表す完全グラフとなった．一方，Graphical Lasso を適用した後のグラフ構造（図 2(b)）は，疎なグラフ構造となった．疎なグラフ構造は，完全グラフと比べてノイズの影響を排除できるため，ノイズに対する頑強性の観点から有効である．

3.2 侵入検知および可視化

グラフ化したトラフィックデータを NFP で学習した結果，98%の精度で異常を予測できた．この値は，先行研究と同程度に高い値である [3]．また図 3 は，異常を検知した際にその要因と判断された部分グラフである．特に，今回対象とした異常は，80 番ポートに対する Syn-flood 攻撃，および複数の送信 IP から 1 つの受信 IP に対して行われる Smurf 攻撃である．実際にそれぞれの部分グラフの結果から，異常 1 が HTTP と SYN に関わる Syn-flood 攻撃，異常 2 が送信 IP と ICMP に関わる Smurf 攻撃であると推測できる．

4 まとめ

本研究では，トラフィックデータを疎構造学習によって疎なグラフ構造で表せることを示した．そのグラフ構造を GCNN によって学習することで，98%の精度で異常を予測できることを示した．また異常の検知だけでなく，異常となった要因を部分グラフの可視化によって説明できることを示した．今後は，本手法における説明性の要である時系列特徴量の再検討，および最新のトラフィックデータでの有効性の調査を進める．本手法は，トラフィックデータのみならず，様々な時系列データへの応用も期待できる．

参考文献

[1] 総務省：平成 29 年版　情報通信白書 (online), 入手先 ⟨http://www.soumu.go.jp/johotsusintokei /whitepaper/ja/h30/index.html⟩ (2018.12.26).

[2] 情報通信研究機構：NICTER 観測レポート 2017 (online), 入手先 ⟨https://www.nict.go.jp/cyber/report /NICTER_report _2017.pdf⟩ (2018.12.26).

[3] Li, G., Yan, Z., Fu, Y. and Chen, H: Data Fusion for Network Intrusion Detection: A Review, *Security and Communication Networks* (2018).

[4] MIT: DARPA1998Dataset (online), 入手先 ⟨https://www.ll.mit.edu/r-d/datasets/1998-darpa-intrusion-detection-evaluation-dataset⟩ (2018.10.23).

[5] 原 聡：機械学習における解釈性, 人工知能 (2018).

[6] Friedman, J., Hastie, T., and Tibshirani, R.: Sparse inverse covariance estimation with the graphical lasso, *Biostatistics* (2008).

[7] 井出 剛，杉山 将：異常検知と変化検知 (2017).

[8] Duvenaud, D.K., Maclaurin, D., Iparraguirre, J., Bombarell, R., Hirzel, T., Aspuru-Guzik, A. and Adams, R.P.: Convolutional Networks on Graphs for Learning Molecular Fingerprints, *In Advances in neural information processing systems* (2015).