侵入検知システムのためのGNNとLLMの高次相互情報量に基づいた損失関数の設計と実装

渡部　柚[1,a)]　菅沼　拓夫[2]　和泉　諭[1]

概要：近年，サイバー攻撃の高度化に伴い，深層学習を用いたネットワーク侵入検知システムの研究が進展している．中でもグラフニューラルネットワークや大規模言語モデルは，複雑なネットワーク構造を効率的に捉え，従来手法では検出困難な攻撃にも対応可能である．一方，これらのモデルは（半）教師あり学習に依存しており，膨大なデータと学習時間を要するため，攻撃の変化に迅速に対応するのが困難である．そこで本研究では，少量のデータで効率的に調整可能な学習アーキテクチャと，高次相互情報量に基づく損失関数を提案する．実験の結果，提案手法は89%の分類精度を達成し，従来手法を上回る性能を達成した．さらに，100 エポックの学習時間は約 30%削減され，効率的な学習能力も確認された．

キーワード：侵入検知システム，GNN，LLM，高次相互情報量

Design and implementation of loss function based on high-order mutual information between GNN and LLM for intrusion detection systems

Yuzu Watanabe[1,a)]　Takuo Sugauma[2]　Satoru Izumi[1]

Abstract: In recent years, research on network intrusion detection systems based on deep learning has progressed as cyber-attacks become more sophisticated. Graph Neural Network（GNN）and Large Language Model（LLM）, in particular, can efficiently capture complex network structures and address attacks that are difficult to detect using conventional methods. On the other hand, these models rely on (semi-)supervised learning, which requires large amounts of data and training time, making it difficult to respond quickly to changes in attacks. In this study, we propose a learning architecture that can be efficiently tuned with a small amount of data and a loss function based on higher-order mutual information. Experimental results show that the proposed method achieves 89% classification accuracy, outperforming conventional methods. Furthermore, the learning time for 100 epochs was reduced by about 30%, confirming the efficient learning capability of the proposed method.

Keywords: Network Intrusion Detection System，GNN，LLM，High-order mutual information

1. はじめに

近年，IoT 技術の進展とネットワークの利用拡大に伴い，個人情報を含むデータのやり取りが急増している．それに伴い，企業や社会インフラを標的としたサイバー攻撃の脅威も深刻化している．実際，国内における攻撃関連の通信は，2018 年から 2022 年の 5 年間で約 2.8 倍に増加し，6197 億パケットに達していると報告されている [1]．中でも，持続型標的攻撃は検知の回避を目的として巧妙に設計されており，医療機器の停止や交通インフラの機能不全など，社会的に重大な被害を引き起こすリスクがある．

このような高度化・巧妙化するサイバー攻撃への対策として，ネットワーク侵入検知システム（Network Intrusion Detection System: NIDS）の研究が進められている．NIDS

[1] 仙台高等専門学校
National Institute of Technology, Sendai Collage
[2] 東北大学サイバーサイエンスセンター
Cyberscience Center, Tohoku University
[a)] a2411529@sendai-nct.jp

は，ネットワークを通過するパケットデータをリアルタイムで監視し，不正アクセスや異常な通信を検出することで，攻撃の初期段階での対応を可能にする重要なセキュリティ技術である [2]．従来の NIDS は，専門家が通信パターンを手動で解析し，異常を検知するためのルールやシグネチャを設定するルールベース方式が主流であった．この方式は即時的な対応が可能な一方で，ルールの更新に大きな労力を要するうえ，日々変化する攻撃手法に柔軟に対応することが難しい課題がある [3]．特に近年の攻撃は，正常な通信に酷似した挙動を装うことで既存ルールを回避する傾向が強まっており，静的な手法では十分な対応が困難になっている．そのため，より高い即応性と柔軟性を備えた自動化技術の導入が急務となっている．

こうした背景から，機械学習を用いたアノマリベースの NIDS が注目されている [2]．これにより，通信のパターンを自動で学習し，未知の攻撃にも対応可能となる．中でも深層学習は，多様で複雑な攻撃パターンを高精度で検出できる手法として注目を集めており，さまざまなネットワークセキュリティ分野に応用が進んでいる．

さらに最近では，グラフニューラルネットワーク（Graph Neural Network: GNN）と大規模言語モデル（Large Language Model: LLM）の特性を組み合わせた手法が提案されている [4][5]．GNN はネットワーク内の通信構造やパケット間の関係性を捉えることに長けており，LLM はパケット単位の数値的な特徴を抽出するのに優れている．特に文献 [5] では，両者を並列に用いた NIDS によって 99% の高精度な検出が可能であることが示されている．

しかし一方で，このような構成は計算資源と学習時間を大量に消費するため，攻撃の変化に迅速に対応する必要がある NIDS の実運用環境では適用が困難である．この課題に対し，異なる情報の関係性に基づいた損失関数と自己教師あり学習を導入し，学習コストを削減するアプローチが提案されている [6]．しかし，GNN と LLM の情報を効果的に統合できない課題がある．

以上の課題を踏まえ，本研究では，NIDS における再学習時の高コストおよび自己教師あり学習の非効率性という問題に対処することを目的とする．GNN と LLM がそれぞれ捉える広域な構造情報と局所的な数値情報の特性を活かし，これらの高次相互情報に基づいた Barlow Twins 型の損失関数を新たに設計する．これにより，ラベル付きデータの使用を最小限に抑えつつ，大規模な事前学習と軽量なファインチューニングを両立させ，高精度かつ即応性の高い NIDS の実現を目指す．

2. 関連研究

NIDS による悪意のあるトラフィックの検出と分類に関する研究が進められている．本章では本研究に密接に関連するこれらの手法をレビューし，2 つのカテゴリーに大別する．

2.1 機械学習に基づく手法

既存のトラフィック分類の研究は，ネットワークトラフィックが持つパケット長，パケット時間等の統計的特徴の解析を目的としている．例えば，Shekhawat[7] らは，暗号化トラフィックの解析のために，IP アドレス・ポート番号・サーバ名・暗号などの統計情報を利用した．これらの情報は XGBoost や SVM の学習で活用され，XGBoost では 98%の分類精度を達成した．Taylor ら [8] はパケット長の統計的特徴を利用し，Shen ら [9] は累積パケット長の統計的特徴を利用して，トラフィック分類のためのランダムフォレスト分類器を学習させている．

これらの機械学習に基づく手法の利点は，その迅速な学習にある．機械学習は，特徴選択を事前に実施した上で学習するため，少量の学習回数で高精度な検知が可能である．そのため，日々変化するサイバー攻撃に対しても，迅速な学習を持って対応することができる．しかし，特徴選択は人間による手動解析に依存しているため，大規模なトラフィックの解析には膨大な時間的コストと人的コストを必要とする．また，特徴選択による特徴量は，時間経過に伴うトラフィックパターンの変化により効果的な機能を失う．このように，トラフィックが変化する度に特徴選択を行う必要性があるため，非効率的な手法となっている．

2.2 深層学習に基づく手法

手動による特徴抽出の限界を受けて，近年では深層学習を活用したトラフィック検出手法が一般的なアプローチとなっている．Wang ら [10] は，生のトラフィックデータを IDX3 形式に前処理した上で，畳み込みニューラルネットワーク（Convolutional Neural Network: CNN）を用いて特徴を自動抽出する手法を提案した．さらに Lin ら [11] は，CNN と長短期記憶（Long Short-Term Memory: LSTM）を組み合わせ，トラフィックを時系列的に解析する手法を開発した．加えて，Lin ら [12] は，LLM の一種である BERT（Bidirectional Encoder Representations from Transformers）を応用し，トラフィックの文脈的特徴を捉えるモデル「ETBERT」を提案した．LLM はパケット単位の値の関係性を学習し，数値的な異常の検出には有効であるものの，ネットワーク全体の構造的な挙動を把握するのは難しい課題がある．

この問題に対処するために，Shen ら [4] は，サーバ・クライアント間のインタラクションからグラフ構造を構築し，それを GNN で解析するフレームワークを提案した．GNN は，通信間の構造的・時系列的関係を捉え，ネットワーク全体のパターンを把握するのに有効だが，一方で細かな数値変化の検出には不向きである．これら深層学習モデルそれぞれの利点と限界を踏まえ，Jin ら [5] は，GNN と LLM

を並列に組み合わせた学習アーキテクチャ「MTSecurity」を提案し，両者の特性を相補的に活用することで精度向上を図った．MTSecurity は，GNN が補足した大域的な構造特徴と LLM が学習した局所的な数値特徴を組み合わせた推論を行うことで 99%の検出精度を達成した．しかし，このような融合手法は確率モデルが複雑であるため，学習に大量のラベル付きデータと計算時間が必要であるという課題が存在する．

この課題に対するアプローチとして，ラベルを必要としない学習手法である自己教師あり学習が存在する [13]．この手法では，事前学習の段階でデータから自動的に教師信号を生成することで，ラベルなしでも高い汎化能力を獲得できる．さらに，自己教師あり学習の利点として，ファインチューニングによるモデルの迅速な適応が挙げられる．少量のラベル付きデータでモデルを微調整できるため，攻撃パターンが変化した場合にも，短時間かつ高効率に再学習が可能となる．Jure ら [14] らは学習時間の削減のために，負例を使用せず正例の相互情報量のみを最大化する Graph Barlow Twins 損失を提案した．Jing ら [6] は，GNN が持つ 3 種の特徴に対する高次相互情報量を定義し，教師信号の補強による効率的な学習を行う損失関数を提案した．このように学習コストの削減を目的とし，GNN や LLM において多様な自己教師あり学習手法が提案されているが，両モデル間の情報連携が不十分で補完効果に限界がある．具体的には，GNN と LLM で特徴捕捉の情報共有ができないため，学習の冗長性による学習コストが課題となっている．

そこで本研究では，NIDS における自己教師あり学習の非効率性という課題に対処することを目的とし，GNN と LLM の高次相互情報量を活用した自己教師あり学習フレームワークを提案する．これにより，ラベル付きデータの使用を最小限に抑えることで再学習時の学習時間と使用データ量を削減する．本手法により，高精度かつ少量の学習データ・学習時間で運用可能な侵入検知システムを構築する．

3. 提案手法

本節では，高次相互情報量に基づく提案手法の学習手順を説明する．はじめに，全体的な学習フレームワークの構成を概観する．次に，学習に用いるトラフィックデータの前処理方法を示す．続いて，特徴抽出を行う各モジュールの構造と処理内容を説明する．最後に，抽出された特徴間の関係性を考慮した新たな損失関数と，最適化手法について述べる．

3.1 学習フレームワーク

3.1.1 手法の全体像

本手法の全体的なフレームワークを**図 1** に示す．事前処理の段階では，データは複数のフローから構成されるキャプチャパケット（PCAP）ファイルであるため，まず双方向フロー（bi-flow）に分割する．その後，グラフの構築とトラフィックのデータ整形（3.1.2）を実施して，各双方向フローのグラフと行列データを生成する．学習モジュールでは，グラフデータを GNN モジュールに入力し，行列データを LLM モジュールに入力する．この操作により，フローの構造的特徴とパケットの時系列的な文脈特徴を自動的に抽出する．トラフィック分類モジュールでは，生成されたパケットレベルの文脈特徴とフローレベルの構造特徴を結合してトラフィック全体の特徴を強化し，マルチクラス分類のためのソフトマックス層を通過させる．損失関数では，各々のモジュールの特徴学習の冗長性を削減するために，GNN の構造特徴量と LLM の文脈的特徴量の相互情報量に基づく損失値を Barlow Twins により計算する．

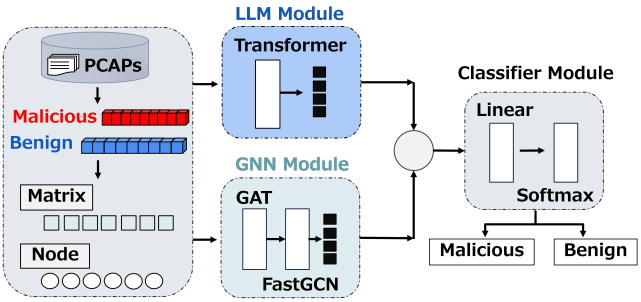

図 1: 学習フレームワークの全体像

Fig. 1 Overall picture of the learning framework

3.1.2 データの前処理

データの前処理の全体像を**図 2** に示す．データの前処理部では，パケットデータの解析とグラフの構築を実施する．パケットデータの解析では，データの精緻性を向上させるために，ネットワーク解析において不要となるリンク層ヘッダを削除する．また，ユーザープライバシーの保護やモデルの一般化可能性の維持のために，IP アドレスやポート番号も削除する．具体的には，送信元/送信先 IP アドレスと送信元/送信先ポート番号があるバイトデータを削除する．最後に，解析済みデータに対して整形処理を適用することで一貫した行列形式のパケットデータを作成する．具体的には，フロー内のパケット数と個々のパケットの長さは常に異なるため，学習には一貫した（$\sqrt{M} \times \sqrt{M}$）形状の行列データに変形する必要がある．そこで，フロー全体の最初の N 個のパケットを選択し，その長さを L で固定する（$N \times L = M$）．不足分に対してはゼロパディングを適用し，超過したデータは削除する．整形したデータは $\sqrt{M} \times \sqrt{M}$ となるように行列形式に変形する．

グラフの構築では，パケットをノード，パケットの時系列関係をエッジとして保有するグラフを構築する．各ノードにはパケットの固有特性（パケット長，パケット方向，プロトコル等）を埋め込み，パケットの表現を強化する．ま

た，各ノードをパケットのバースト情報に基づいてグループ化し，グループ内の隣接ノードを無向エッジで接続する．グループ間では，グループ i の先頭ノードと末尾ノードに，他のグループ $i+1$ の先頭ノードと末尾ノードをそれぞれ接続する．

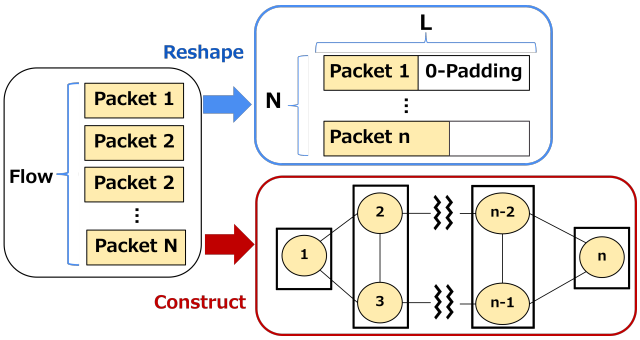

図 2: データの前処理

Fig. 2 Data preprocessing

3.1.3 学習モジュール

学習モジュールの全体像を**図 3** に示す．提案手法では，グラフ形式のデータと行列形式のデータに対し，それぞれに最適な深層学習モジュールを適用することで，埋め込み表現を作成する．

まず，グラフ形式のデータには GNN モジュールを適用する．このモジュールは，各パケット間の接続関係やフロー全体の構造的特徴に基づいて学習を行い，グラフ全体の意味を内包するグラフ埋め込みを生成する．GNN モジュールは，3 層の Graph Attention Network（GAT）と 3 層の FastGCN の並列接続で構成され，GAT は全てのノードに対して周囲ノードの重要度を動的に学習することで，文脈に応じた情報集約が可能である．FastGCN は各層で固定されたサンプリング確率に基づいて計算を行うため，大規模グラフに対しても効率的な学習を実現できる．このモデルは既存研究 [5] の GraphSAGE のようにノードごとの確率計算を必要としないため，計算量を指数時間から線形時間に削減し，モデル全体の学習時間を大幅に短縮する．一方，行列形式のデータには LLM モジュールを適用する．このモジュールは，前後パケット間の時間的・意味的な相関関係に基づいて特徴抽出を行い，パケット列に対する高次元のバイト埋め込み表現を生成する．LLM には 12 層の Transformer を採用しており，マスク付き言語モデルを用いた自己教師あり学習により，未知の通信パターンにも柔軟に対応できる表現獲得を可能としている．

これらの GNN および LLM モジュールを並列的に接続することにより，学習領域が異なるモジュールが相互補完的に特徴表現を強化する．さらに，既存研究 [5] においては GraphSAGE を活用していたが，上述の通り GraphSAGE は層数に対して指数的な計算コストが発生する．これに対して本研究では，同等以上の性能を維持しつつ，より効率的な FastGCN へ置き換えることにより学習時間の大幅な短縮を可能にする．これにより，実運用における学習コストの削減と効率的な再調整を実現する．

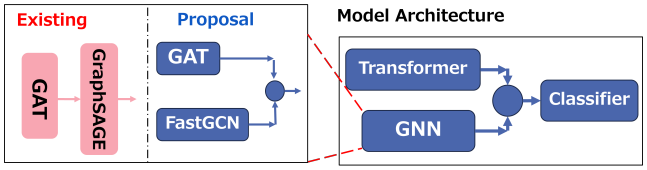

図 3: 学習モジュールの全体像

Fig. 3 Overall picture learning module

最後に，分類モジュールでは 2 つのモジュールが持つ埋め込み表現を線形層に入力し，softmax 関数に通して予測値を獲得する．

3.2 損失関数

本研究では，対照学習に基づく事前学習を採用し，GNN によるグラフ埋め込みと LLM によるバイト埋め込みの相互依存性を考慮する損失関数を提案する．特に，負例を使用しない Barlow Twin 損失を適用することで，計算効率を向上させる．

3.2.1 異なる埋め込みの高次相互情報量

二つの変数間の相互依存性を示す指標として相互情報量があり，これは次の式で表される．

$$I(X_1; X_2) = H(X_1) - H(X_1|X_2) \tag{1}$$

$I(X_1; X_2)$ は変数 X_1 と X_2 の相互情報量であり，$H(X_1)$ と $H(X_1|X_2)$ はそれぞれエントロピーと条件付きエントロピーを表す．直感的にこの数式は，事前情報として変数 X_1 が与えられた際の情報 X_2 の不確実性を示す指標であり，この相互情報量が小さいほど情報間の依存性が高いと評価される．この相互情報量を $N \geq 3$ 個の確率変数に対して一般化したものが高次相互情報量である．n 個の確率変数の集合 $X_1, X_2, ..., X_N$ が与えられたとき，高次相互情報量は相互情報量 式 (2) に類似して定義される．

$$I(X_1; X_2; ...; X_N) = \sum_{n=1}^{N} (-1)^{n+1} \sum_{i_1 < ... < i_n} H(X_{i_1}, ..., X_{X_{i_n}}) \tag{2}$$

ここで，$H(X_{i_1}, ..., X_{X_{i_n}})$ は $X_{i_1}, ..., X_{X_{i_n}}$ の結合エントロピーを表し，総和 $\sum_{i_1 < ... < i_n} H(X_{i_1}, ..., X_{i_n})$ は確率変数の組み合わせ $(i_1, ..., i_n) \in [1, ..., N]$ の全てに対して実施される．この定義によれば，N=3 のとき以下の式が成り立つ．

$$\begin{aligned} I(X_1; X_2; X_3) = {} & (X_1) + H(X_2) + H(X_3) \\ & - H(X_1, X_2) - H(X_1, X_3) - H(X_2, X_3) \\ & + H(X_1, X_2, X_3) \end{aligned} \tag{3}$$

式（3）は，さらに次の式に書き換えることができる．

$$\begin{aligned} I(X_1;X_2;X_3) &= (X_1) + H(X_2) - H(X_1,X_2) \\ &\quad + H(X_1) + H(X_3) - H(X1,X_3) \\ &\quad - H(X_1) - H(X_2,X_3) + H(X_1,X_2,X_3) \\ &= I(X_1;X_2) + I(X_1;X_3) - I(X_1;X_2,X_3) \end{aligned} \tag{4}$$

式（4）において，$I(X;Y,Z)$ は X_1 の分布と X_2,X_3 の同時分布との相互情報を表す．ここで，式 (4) の各変数をグラフ埋め込み s，ノード特徴量 h_n^g，バイト埋め込み h_n^t に置き換えると，グラフ埋め込みとバイト埋め込みの相互情報量は次のように表される．

$$I(s;h_n^g;h_n^t) = I(s;h_n^g) + I(s;h_n^t) - I(s;h_n^g,h_n^t) \tag{5}$$

式（5）において $I(s;h_n^g)$ は，グラフ埋め込みとノード埋め込み間の相互依存性というグラフの内在的な監視信号であり，$I(s;h_n^t)$ はグラフ埋め込みとバイト埋め込みの外在的な監視信号を捉える．異なる相互情報量に対して係数をかけることにより調整し，最終的な損失関数を提案する．

$$L = \lambda_E I(s;h_n^g) + \lambda_I I(s;h_n^t) + \lambda_j I(s;h_n^g,h_n^t) \tag{6}$$

3.2.2 相互情報量の算出

各相互情報量 I の算出には Barlow Twins 損失を使用する．Barlow Twins 損失は，2 つの埋め込み表現のコサイン類似度により作成した相関行列に対して，対角要素を最大化，非対角要素を最小化するように最適化を行う．

$$\sum_{i=1}^{n}(1-C_{i,i})^2 + \sum_{i=1}^{n}(C_{i,j})^2$$

ここで，$C_{i,i}$ は相関行列の対角要素であり，$C_{i,j}$ は非対角要素を表す．

4. 実験

初期実験として，提案する学習モジュールの実装と分類精度および学習時間の計測に関する実験を行った．本章では，実験の内容や結果について説明する．

4.1 データセット

本研究では，マルウェアを含むトラフィックデータセットとして代表される 3 つのデータセット：UNSW-NB15, MCFP, USTC-TFC2016 を用いて実験を行う．MCFP データセットには，Yakes や HTBot などの複数のマルウェアが多数含まれており，USTC-TFC2016 データセットには，攻撃関連の悪性トラフィックと正常通信の良性トラフィックが 20 カテゴリ含まれている．

本実験は初期段階での評価を目的としており，簡易的な検証を行うために，USTC-TFC2016 データセットの約 20%にあたる 60,000 件のデータを使用した．このデータ選定による特徴分布の偏りを抑えるため，対象データはガウス分布に従ってランダムに抽出する．また，データ選択から学習・評価までの実験を 5 回繰り返し，その平均を評価することにより結果の安定性と再現性を確保する．

4.2 評価指標

提案モデルの性能を評価・比較するために，NIDS の研究で一般的に利用される Accuracy（分類精度），Precision（適合率），Recall（再現率），F1-Score を使用した．その式を（7）-（10）に示す．これらの式においては，以下の 4 つの分類結果が用いられる．

- True Positive（TP）：悪性と判断すべきデータを正しく悪性と分類した件数
- True Negative（TN）：正常と判断すべきデータを正しく正常と分類した件数
- False Positive（FP）：正常なデータを誤って悪性と分類した件数
- False Negative（FN）：悪性のデータを誤って正常と分類した件数

これら全ての指標に対し，式（11）で表される各カテゴリの平均値を計算し，データの不均衡による結果の偏りを回避する．ここで，n はカテゴリの総数であり，X は特定の評価指標を表す．

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{7}$$

$$Precision = \frac{TP}{TP+FP} \tag{8}$$

$$Recall = \frac{TP}{TP+FP} \tag{9}$$

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{10}$$

$$Macro-Average = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{11}$$

また，事前学習および微調整時の全エポックにおける学習時間を計測し，環境変化への適応効率を評価する．

4.3 データの前処理

本実験では，USTC-TFC2016 データセットに対して，教師あり学習とテストを行う．実験に先立ち，Split Cap ツールを使用してトラフィックデータを五分位に従って双方向フローに分割する．その後，各フローを 3 章で説明した手法を適用し，行列形式のデータとグラフ形式のデータに変換する．元データセットにはカテゴリごとの偏りが存在するため，アップサンプリング [15] により少数カテゴリのサンプルを増やし，ダウンサンプリング [15] で多数カテゴリのサンプルを削減する．これにより，カテゴリ間のバランスの改善を図る．最終的に，データセットは 6:2:2 の割合で学習用・検証用・テスト用に分割して使用する．

4.4 実験結果

本研究では，マルウェアトラフィックの分類を目的としたマルチクラス分類タスクに焦点を当て，提案手法の性能と実用性を評価した．まず，USTC-TFC データセットを用いて提案モデルの学習とテストを行うことにより，モジュールの分類性能を確認した．続いて，提案手法の有効性を示すため，既存手法である MTSecurity[5] の追試実験を行い，両手法の分類結果を比較した．

各モデルの分類精度と 100 エポックにおける学習時間を表 1 と表 2 にそれぞれ示す．表 1 より，提案モデルは全指標で 0.85 以上のスコアを記録し，特に Recall と F1-Score で顕著な向上を確認した．Recall は MTSecurity に比べて約 0.3 向上しており，攻撃通信の見逃しを大幅に抑えられることが示された．

表 2 に示すように，MTSecurity の学習時間（30 分 2 秒）に対し，提案手法は 21 分 42 秒と約 30%の時間短縮を実現した．これは FastGCN の導入などによる処理効率の向上によるものであり，計算時間や計算コストに限りのある実環境においても有利であると考えられる．一方で，推論時間は 3 ミリ秒増加したが，推論計算の並列化により容易に対応可能 [16] であるため，NIDS のリアルタイム性に大きな影響はないと考えられる．

以上より，提案手法は分類精度と計算効率の両方で既存手法を上回る有効なアプローチであることが示された．

表 1: 各モデルの分類精度
Table 1 Classification accuracy of each model

Method	Acc(%)	Precision	Recall	F1-Score
MTSecurity	88	0.84	0.86	0.84
Ours	89	0.87	0.87	0.86

表 2: 各モデルの学習時間
Table 2 Training time for each model

Method	100 エポックの学習時間	1 バッチの推論時間
MTSecurity	30m3s	0.0066
Ours	21m42s	0.0096

さらに，提案手法のより包括的な評価を行うため，USTC-TFC データセットにおける各カテゴリごとの分類精度を分析した．その結果を**図 4** に示す．図 4 は混同行列であり，縦軸が実際のラベル，横軸がモデルの予測ラベルを表す．したがって，対角線上の値が各カテゴリにおける正解率を示しており，値が高いほどそのマルウェアを正しく識別できていることになる．

マルウェア分類の結果としては，Cridex が 98%と最も高い分類精度を示した一方，Neris は 57%と最も低い精度となった．また，正常通信に関しては 94%と高い精度を記録している．これらの結果から，マルウェアの種類によって分類精度に差が確認できる．この汎化性能のばらつきは，各クラスのデータ数の不均衡に起因していると考えられる．特に，Cridex に比べて Neris のデータ数が少ないことから，学習の偏りによって分類性能の差が生じた可能性がある．

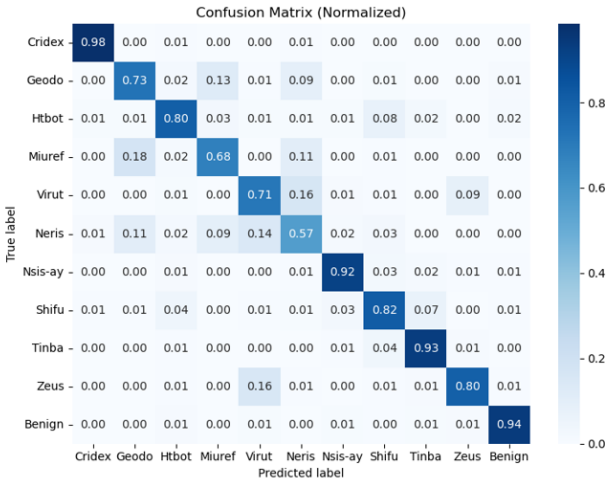

図 4: クラスごとの分類精度
Fig. 4 Classification accuracy by class

5. 考察

この章では，まず，比較検証の結果から明らかとなった提案手法の柔軟性と分類能力を説明する．その後，混同行列により明確化した各クラスの精度差について説明する．

5.1 提案手法の利点

サイバー攻撃は日々多様化・高度化しており，それに対応するにはモデルの迅速な再学習や調整が不可欠である．提案手法はこの要件を満たしており，実際に既存研究が要した学習時間を 30%削減した．さらに，分類精度の面でも提案手法は既存研究を上回っており，特に攻撃の見逃しを大幅に抑制する．

また，GNN と LLM の高次相互情報量を活用した損失関数により，正解ラベルを含まない少量のデータでも高い分類性能を実現できる．これにより，学習準備や再調整にかかる負担も軽減される．さらに，本手法はファインチューニングにも柔軟に対応でき，既存モデルの重みを保持したまま新たな攻撃パターンに迅速に適応可能である．これは，常時更新が求められる NIDS において実運用上の大きな利点となる．

本手法はこれらの取り組みにより，社会インフラへの被害を未然に防ぎ，信頼性の高いネットワーク環境の構築に貢献できると考えている．

5.2 データの不均衡性とサンプリング手法

本実験で用いた USTC-TFC2016 データセットは各クラスのデータ数の分布が大きく，不均衡性が確認されている．実際，表 3 に示すように，多数クラスである Zeus が 10000 件であるのに対して Neris は 6511 件であり，約 4000 件少ない．提案手法では，このようなデータの不均衡性を解消するために 4.3 章で説明した手法を用いてクラスの均等化を図った．具体的には，アップサンプリングでは少数クラスのサンプルを複製し，ダウンサンプリングでは多数クラスのデータを一部削除することで，サンプル数を平均化した．しかし，この手法はデータの類似性や特異性を考慮しないため，トラフィックの元の特徴分布を破壊する可能性がある．特にアップサンプリングにおいては，情報量が小さいサンプルを複製した際にデータの冗長性が増加し，新たな特徴量に対する学習が非効率的になると考えられる．

そのため，今後はデータ拡張を実施すると同時に，データ分布を考慮したアップサンプリング手法である GICaPS[17] の実装を行う必要があると考えている．

表 3: USTC-TFC2016 の各クラスのデータ数
Table 3 Number of data points for each class in USTC-TFC2016

Label	Flows	Label	Flows
BitTor	4491	Nesis-ay	1862
Cridex	8189	Outlook	7370
Facetime	6000	Shifu	9631
FTP	10000	Skype	5679
Geodo	6033	SMB	10000
Gmail	5252	Tinba	7749
HtBot	4298	Virut	5788
Miuref	4900	Weibo	10000
MySQL	9992	Wow	7786
Neris	6511	Zeus	10000

6. まとめ

本研究では，NIDS における再学習時のコスト問題に対処することを目的とし，GNN と LLM で構成される新たな学習モジュールと高次相互情報に基づいた Barlow Twins 型の損失関数を提案した．従来手法との比較実験により，提案するモジュールは比較的短い計算時間で優れた学習能力を有することが明らかとなった．また，カテゴリごとの分類精度を検証した結果，多くのマルウェアを高確率で捕捉可能であることが判明した．

一方，Neris 等の一部のマルウェアに対して分類精度は低く，これはデータの不均衡性や単純なデータ拡張による特徴分布の拡散が原因であると考えられる．今後の課題として，MCFP データセットや UNSW-NB15 等のデータセットの追加，データ分布を考慮したアップサンプリング手法である GICaPS の実装を行う予定である．また，提案する損失関数を適用し，他手法との比較評価を実施する．

参考文献

[1] NICT: NICTER 観測レポート 2023 の公開，`https://www.nict.go.jp/press/2024/02/13-1.html` (2024). 参照: 2024-03-12.

[2] T.Alsmadi, et al.: A Survey on malware detection techniques, *International Conference on Information Technology (ICIT)*, pp. 371–376 (2021).

[3] T.Bilot., et al.: Graph Neural Networks for Intrusion Detection: A Survey, *IEEE Access*, Vol. 11, pp. 49114–49139 (2023).

[4] M.shen, et al.: Accurate decentralized datagram representation via encrypted traffic analysis using graph neural networks, *IEEE Transactions on Information Forensics and Security*, Vol. 16, pp. 2367–2380 (2021).

[5] Y.Jin, et al.: MTSecurity: Privacy-Preserving Malicious Traffic Classification Using Graph Neural Network and Transformer, *IEEE Transactions on Network and Service*, Vol. 21, No. 3, pp. 3583–3597 (2024).

[6] B.jing, et al.: HDMI: High-order Deep Multiplex Infomax, *Association for Computing Machinery*, p. 2414–2424 (2021).

[7] S.Shekhawat et al.: Feature analysis of encrypted malicious traffic[J], *Expert Systems with Applications*, Vol. 125, pp. 130–141 (2019).

[8] V.Taylor, et al.: Robust smartphone app identification via encrypted network traffic analysis, *IEEE Transactions on Information Forensics and Security*, Vol. 13, No. 1, pp. 63–78 (2017).

[9] M.Shen, et al.: Fine-grained webpage fingerprinting using only packet length information of encrypted traffic, *IEEE Transactions on Information Forensics and Security*, Vol. 16, pp. 2046–2059 (2020).

[10] W.Wang, et al.: End-to-end encrypted traffic classification with one-dimensional convolution neural networks, *IEEE international conference on intelligence and security informatics (ISI).*, pp. 43–48 (2017).

[11] W.Wang, et al.: TSCRNN: A novel classification scheme of encrypted traffic based on flow spatiotemporal features for efficient management of IIoT, *Computer Network*, Vol. 190, pp. 1389–1286 (2021).

[12] X.Lin, et al.: ET-bert: A contextualized datagram representation with pre-training transformers for encrypted traffic classification, *Proceedings of the ACM Web Conference 2022*, pp. 633–642 (2022).

[13] L.Yixin, et al.: Graph Self-Supervised Learning: A Survey, *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, Vol. 35, No. 6, pp. 5879–5900 (2023).

[14] Z.Jure et al.: Barlow Twins: Self-Supervised Learning via Redundancy Reduction, *arXiv* (2021).

[15] V.Dumoulin, et al.: A guide to convolution arithmetic for deep learning, *arxiv preprint* (2016).

[16] King, Isaiah J, et al.: Euler: Detecting Network Lateral Movement via Scalable Temporal Link Prediction, *ACM Trans. Priv. Secur.*, Vol. 26, No. 3, p. 36 (2023).

[17] M.Anima, et al.: A Method for Handling Multi-class Imbalanced Data by Geometry based Information Sampling and Class Prioritized Synthetic Data Generation (GICaPS), *arXiv* (2020).