連合転移学習におけるモデル凍結層の選択手法

A Layer Freezing Strategy Selection Method in Federated Transfer Learning

吉村 優 † 水谷 后宏 ‡,§
Yoshimura Yu Mizutani Kimihiro

1. はじめに

近年，IoT の普及などによってビッグデータの活用が進む中，複数の学習モデルを統合して高度な学習モデルを生成する連合学習という手法が提案されている．連合学習では，複数のクライアントがそれぞれのローカルデータを活用して個別にモデルを学習し，その後，学習済みのモデルパラメータを共有・統合することで，全体の性能を向上させる．この手法は，データのプライバシー保護や通信コストの削減といった利点を持ち，さまざまな分野での応用が進められている．

また，モバイルエッジコンピューティング環境においては，セキュリティリスクが大きな課題となっており，侵入検知システムの精度向上が求められている．深層学習を用いた侵入検知は高い精度を達成しているが，従来手法では中央サーバにデータを集約する必要があり，プライバシーの観点から問題がある．連合学習はこの問題を解決できるが，膨大な通信量が発生するという課題がある．

こうした課題に対し，通信量削減のため転移学習と組み合わせた連合転移学習が注目されている．具体的には，あるドメイン（ネットワーク）にて不正なトラフィックを検知するニューラルネットワークを連合学習で生成し，そのモデルを別のドメインに転移して再学習することで，学習の収束速度が向上し，通信量の削減と学習時間の短縮が可能となる．

しかし，モデルの転移においては，転送元の環境で得られたモデルの特性を保ちつつ，宛先環境特有の特性を新たに学習させる必要がある．このため，学習モデルの特定の層を凍結（学習パラメータの固定）し，他の層のみを再学習する手法が一般的である．層の凍結は，転送元の知識を保持し，不要な学習の重複を防ぐために重要な操作である．

先行研究 [1] では連合転移学習システムにおける層の凍結設定について検討が行われたが，比較方法の網羅性に欠けている．

本研究では，凍結させない層の組み合わせを網羅的に選択し，各層の凍結の有無による精度の変化を検証する．さらに，先行研究 [2] で提案された非独立同分布データに対応した最適化手法を組み合わせることで，連合転移学習における層凍結戦略の包括的な評価を行う．得られた結果から，どの層がモデルの性能にどのような効果をもたらすかを考察し，最適な層の凍結方法を特定することを目的とする．

2. 研究内容

2.1 既存研究の課題と本研究の位置づけ

先行研究 [3] では連合転移学習において，fc2 層のみを凍結させず，それ以外の層は凍結させる固定的なアプローチを採用していた．しかし，どの層を凍結するべきかという層選択の最適化については十分に検討されていない．また，連合学習においてよく用いられる最適化手法である FedAvg や FedSGD は，非独立同分布なデータに対してうまく学習できない場合があることが先行研究 [2] にて示されている．非独立同分布とは，独立同分布がランダムなデータ点が互いに独立であり，かつ同じ確率分布に従うという特性とは異なり，データ点が独立していない，または同じ確率分布に従わない特性を持つ．連合学習では，各クライアントが異なる環境やカテゴリのデータを保持することが多く，この問題はデータ分布が各クライアント間で大きく異なる状況において顕著に現れる．さらに，先行研究 [1] でも同様に層の凍結の組み合わせを検討しているが，比較方法の網羅性に欠けており，非独立同分布データに対する最適化手法については検討されていない．

本研究では，これらの課題を解決するため，層の凍結設定を網羅的に評価し，さらに非独立同分布データに対応した最適化手法を組み合わせることで，より効果的な連合転移学習手法の確立を目指す．

2.2 深層学習モデルの凍結層の最適化

本研究では，上記の背景を踏まえて，深層学習モデルの各層の凍結設定がモデルの性能に与える影響について網羅的な検証を行うことを目的とする．凍結させない層の組み合わせをあらゆるパターンで選択し，それぞれの組み合わせに対してモデルの検知精度の変化を測定する

† 近畿大学大学院総合理工学研究科, Graduate School of Science and Engineering, Kindai University
‡ 近畿大学情報学部, Faculty of Informatics (KDIX), Kindai University
§ 近畿大学情報学研究所, Cyber Informatics Research Institute, Kindai University

ことで，どの層がモデルの性能にどの程度寄与しているのかを明らかにする．さらに，単なる層の凍結設定の探索にとどまらず，モデルの性能向上を図るため，先行研究 [3] で用いられている最適化手法 FedSGD を基準として，追加の改善策も検討する．特に，非独立同分布データへの適応能力を高めるため，先行研究 [2] で提案された非独立同分布なデータにも耐性のある最適化手法を導入し，従来手法との比較を行うことで，より高精度なモデルを目指す．

本研究で用いる侵入検知モデルの構成である 1D-CNN の構成を以下に示す．このモデル構成は先行研究 [3] で用いられているものを基準とし，ネットワークトラフィックの時系列的特徴を効果的に捉えるため 1 次元畳み込み層を採用している．conv1 と conv2，conv2 と pool1，conv3 と pool2 の間に sigmoid 関数を配置する．各層の設定は以下の通りである．

畳み込み層の設定:

- conv1: 入力データ数 1，出力データ数 32，カーネルサイズ 5，パディング 2，ストライド 1
- conv2: 入力データ数 32，出力データ数 32，カーネルサイズ 5，パディング 2，ストライド 1
- conv3: 入力データ数 32，出力データ数 32，カーネルサイズ 5，パディング 2，ストライド 1

プーリング層の設定:

- pool1: カーネルサイズ 2，ストライド 2
- pool2: カーネルサイズ 2，ストライド 2

その他の層の設定:

- dropout: ドロップアウト率 0.5
- fc1: 入力ニューロン数 320，出力ニューロン数 128
- fc2: 入力ニューロン数 128，出力ニューロン数 128

このモデルに対して層の凍結設定を適用し，各設定がモデル性能に与える影響を確認する．凍結する層の候補は畳み込み層（conv1，conv2，conv3）と全結合層（fc1，fc2）であり，これらの組み合わせについてすべてのパターンを評価することで，最適な凍結層の組み合わせを見つけることを目指す．

3. 実験

3.1 実装と実験環境

先行研究 [3] を参考にして，PyTorch を用いて実装を行う．本研究で提案する連合転移学習システムの概要を図 1 に示す．転送元ドメインにてモデルを学習させる時と宛先ドメインにてモデルを学習させる時に用いるデータセットやデータの分布が異なる．

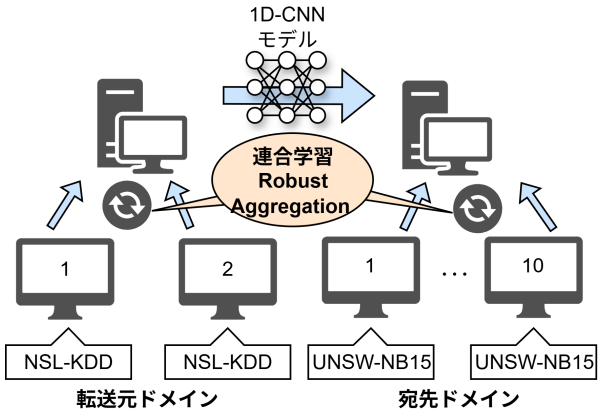

図 1: 連合転移学習システムの概要

3.2 データセットと前処理

NSL-KDD と UNSW-NB15[4, 5, 6, 7, 8] は侵入検知をタスクとした機械学習モデルの学習において非常に幅広く用いられているデータセットである．データセットに含まれる各レコードには，送信元 IP アドレス，宛先 IP アドレス，ポート番号，プロトコル種別，通信時間，サービス種別といったネットワーク通信に関する情報が含まれており，さらに DoS やファジングなど攻撃手法を示す特徴量も含まれる．これらの情報に対して，正常・異常の二値のラベルが設定されており，学習モデルは入力されたデータからそのトラフィックが正常なものか異常なものかを推定する．

これらのデータセットの特徴量の中には，一部同じ内容を示す特徴量がある．ただし，どの特徴量とも対応のとれない特徴量も多くある．2 つのデータセット同士の特徴量について，同じ内容を示すものは配列内で同じインデックスに配置するかどうかで結果に違いが生まれる可能性がある．そのため，データセットの特徴量の並び順を決める方法として以下の 4 つの方法で比較を行った．

1. 特徴量の対応が取れるもののみを利用し，配列内で同じインデックスになるように並び替える
2. 特徴量の対応が取れるものは配列内で同じインデックスになるように並び替え，対応の取れないものは適当に並べる
3. 並び替えは行わない
4. Recursive Feature Elimination を用いた特徴量選択

結果として，特徴量の対応が取れるものは配列内で同じインデックスになるように並び替え，対応の取れない

ものは適当に並べたものが最もモデルの検知精度が高くなったのでこれを採用する．配列の前から順に対応の取れているものを配置していき，後ろの方には対応の取れていないものを並べる．

先行研究 [2] を参考に，各データセットに前処理を加える．非数値型の特徴量はそのままではモデルに読み込ませることができないため，数値に変換する必要がある．今回は，各値をバイナリベクトルに変換して扱う．バイナリベクトルエンコーディングは，カテゴリ型のデータを数値ベースの特徴量に変換する手法の一つであり，機械学習モデルでの処理が容易になる利点がある．

バイナリベクトルエンコーディングでは，カテゴリごとに一意なビット列が割り当てられる．例えば，ある特徴量が 4 種類のカテゴリ値を持つ場合，それぞれのカテゴリは次のようにエンコードされる：カテゴリ 1 → 0，カテゴリ 2 → 1，カテゴリ 3 → 10，カテゴリ 4 → 11．このようなエンコーディングは，カテゴリ間の関係をビット表現で捉えることで，数値型の特徴量としてモデルに入力できるようにする．

さらに，特定の数値型の特徴量には対数スケーリングを行った後，最小値 0，最大値 1 として正規化を施す．それ以外の特徴量には標準化を施す．特定の数値型の特徴量に施す前処理を式 1 と式 2 に示す．ここで，x' は元の特徴量 x に対数スケーリングを適用した値，x'' は x' を 0-1 正規化した最終的な特徴量値である．

$$x' = \log(x+1) \quad (式\ 1)$$

$$x'' = \frac{x' - \min(x')}{\max(x') - \min(x')} \quad (式\ 2)$$

各データセットの特徴量に対する前処理は以下のように行う．NSL-KDD の protocol_type，service，flag と UNSW-NB15 の proto，service，state に対してはバイナリベクトル変換を適用する．NSL-KDD の duration，src_bytes，dst_bytes と UNSW-NB15 の dur，sbytes，dbytes に対しては対数スケーリングを行った後に 0-1 正規化を適用する．それ以外のすべての特徴量に対しては標準化を適用する．

3.3　実験設定

本実験では，クライアント数を 10，統合回数（epoch）を 200，バッチサイズを 8，学習率を 0.00005 に設定した．損失関数にはクロスエントロピー誤差を用いた．改良型重み付け手法については，先行研究 [2] で提案された汎化境界推定に基づく手法を用いる．この手法では，各クライアントのモデル性能を汎化境界の推定値に基づいて評価し，その評価値を重みとしてモデル統合時に動的に調整する．

3.4　実験手順

凍結層の選択肢毎の検知精度の比較は以下のように行った．

1. 凍結しない層の数が 1 つの場合での比較
2. 凍結しない層の数が 2 つの場合での比較
3. 凍結しない層の数が 3 つの場合での比較

凍結しない層の数が 4 つの場合については，転移学習による恩恵である通信量の削減が得られないと考えて行わない．次に，その中で最も検知精度が高くなった凍結層の組み合わせを用いて，最適化手法の違いによる精度の変化を評価する．比較対象は以下の通りである．

1. 非独立同分布下での従来型重み付け（FedSGD）を採用した場合の検知精度
2. 非独立同分布下での改良型重み付けを採用した場合の検知精度
3. 独立同分布下での従来型重み付け（FedSGD）を採用した場合の検知精度
4. 独立同分布下での改良型重み付けを採用した場合の検知精度

3.5　実験結果

凍結層の組み合わせによる検知精度の変化を図 2 に示す．全ての組み合わせを表示すると見づらくなるため，違いが分かる程度に代表的な組み合わせを抜粋して表示している．グラフのラベルの表記については，conv1 を c1，fc1 を f1 のように省略して表示している．

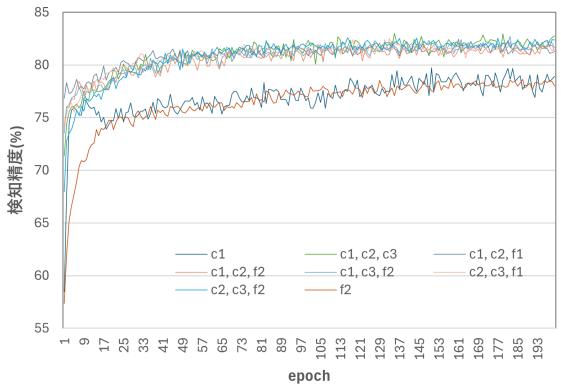

図 2: 凍結層の組み合わせによる検知精度の変化

図 2 に示すように，ほとんどの組み合わせで検知精度 82%程度を達成しているが，conv1 のみを凍結しなかった場合（c1），もしくは fc2 のみを凍結しなかった場合

(f2) についてのみ，目立って検知精度が低下していることがわかった．結果について述べると，3 つ全ての比較どれにおいても conv1，もしくは fc2 を凍結しない層の組み合わせの一つに含めると他の組み合わせと比べて精度の低下を招いていることが分かった．この結果から，conv1 と fc2 を含まない組み合わせが最も高い検知精度を示すことが確認された．

次に，今回用いた非独立同分布に適した最適化手法の効果について述べる．最適化手法の比較結果を図 3 に示す．図 3 では，データ分布条件（独立同分布・非独立同分布）と重み付け手法（従来型・改良型）の 4 つの組み合わせについて，学習過程における検知精度の変化を示している．凡例の表記について，独立同分布を IID，非独立同分布を Non-IID，非独立同分布に強い最適化手法を S，弱い最適化手法を W とし，IID_S，Non-IID_S，IID_W，Non-IID_W のように表示している．

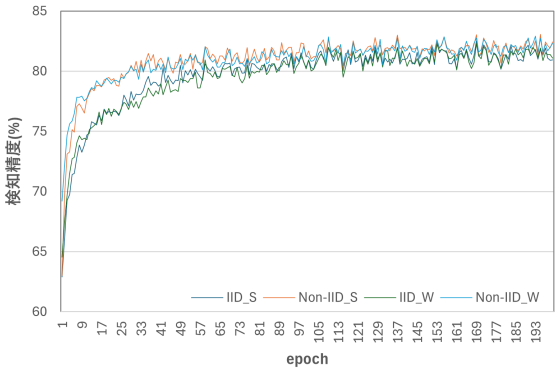

図 3: 最適化手法毎の検知精度の比較結果

図 3 に示すように，4 つの手法すべてにおいて最終的な検知精度は約 82%に収束しており，最適化手法の違いによる精度への明確な差は見られなかった．収束過程においても，わずかな違いはあるものの，有意な差があるとは言えない結果となった．

この結果から，今回の実験設定においては，非独立同分布に対応した改良型重み付けの効果は限定的であることが示された．これは使用したモデルの特性や実験設定に起因する可能性があり，より詳細な検証が必要である．

4. 結論・今後の課題

本研究では連合転移学習における凍結層の組み合わせ毎の精度の調査と非独立同分布に適した最適化手法の適用を行った．実験の結果，conv1 と fc2 を含まない凍結層の組み合わせが最も高い検知精度を示すことが確認された．一方，改良型重み付け手法については，今回の実験設定では従来手法との明確な差は見られなかった．

今後の課題として，モデル構成の変化に対する手法の汎用性の検証や，層数増加に対応した効率的な探索手法の開発が挙げられる．また，先行研究 [9] で示されているバッチ正規化層の問題への対応や，より多様なデータセットでの検証により，連合転移学習の実用性向上が期待される．

謝辞

本研究は JSPS 科研費 JP23K16876 の助成を受けたものである．

参考文献

[1] Yu Yoshimura and Kimihiro Mizutani. Adjusting neural network layer transfer in a federated learning environment. In *International Conference on Emerging Technologies for Communications*, 2024.

[2] Mingwei Xu, et al. Aggregation weighting of federated learning via generalization bound estimation. *arXiv preprint arXiv:2311.05936*, pp. 1–17, 2023.

[3] Yanyu Cheng, et al. Federated transfer learning with client selection. *IEEE Communications Letters*, Vol. 26, No. 3, pp. 552–556, March 2022.

[4] Nour Moustafa, et al. UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In *Proc. MilCIS*, 2015.

[5] Nour Moustafa, et al. The evaluation of network anomaly detection systems: Statistical analysis of the UNSW-NB15 dataset and the comparison with the KDD99 dataset. *Information Security Journal: A Global Perspective*, pp. 1–14, 2016.

[6] Nour Moustafa, et al. Novel geometric area analysis technique for anomaly detection using trapezoidal area estimation on large-scale networks. *IEEE Transactions on Big Data*, 2017.

[7] Nour Moustafa, et al. Big data analytics for intrusion detection system: Statistical decision-making using finite dirichlet mixture models. In *Data Analytics and Decision Support for Cybersecurity*, pp. 127–156. Springer, Cham, 2017.

[8] Mohanad Sarhan, et al. NetFlow datasets for machine learning-based network intrusion detection systems. In *Proc. WiCON*. Springer Nature, 2020.

[9] Yanmeng Wang, et al. Batch normalization damages federated learning on NON-IID data: Analysis and remedy. In *IEEE ICASSP*, pp. 1–5, 2023.