L-011

MLP-Mixer を用いたマルウェア分類

ダオ・ヴァン・トゥアン† 佐藤　浩† 久保　正男†
Tuan Dao Van Hiroshi Sato Masao Kubo

1. はじめに

近年、マルウェアはサイバー空間のセキュリティを脅かす重要な存在となっている。マルウェアの新しい亜種は常に出現し、ウィルス対策ソフト及び侵入検知システムは挑戦を受けている。さらに危険なことに、複数の攻撃者や組織がマルウェアを悪用して、企業、政府機関、金融機関、医療機関などを攻撃している。AV-TEST Institute [1] によると、1 日あたり 45 万以上の新しい悪意のあるプログラムや潜在的に望ましくないアプリケーションが登録されている。今年 5 月までに発見されたマルウェアの数は 13 億 6,000 万個以上である。この数は 10 年前の 7 倍にもなっている。

従来のマルウェア検知作業は、署名と振る舞いによるアプローチに基づいている。しかし，複数のマルウェア自動作成ツールの開発が急速に進んでおり，マルウェアの生成速度にこれらの手法が追いついていないのが現状である。

近年の機械学習は、その高度に発達したアルゴリズムによって、ほぼすべての分野で遭遇する問題のほとんどを解決できるほど強力なものとなっている。それらは、悪意のあるファイルを実行することなく、マルウェアの特徴をとらえることができる。特に、マルウェアを画像化し、画像処理を用いて分類を行うアプローチの精度が向上し、注目を浴びている。理由としては、攻撃者は難読化技術を使用してなりすましを実行するが、同じファミリのマルウェアバリアントは依然として同様のコードとデータの順序を維持しており、同じ場所に表示されないものの同じパターンが保持されている可能性があるためである。

画像処理において、畳み込みニューラル（CNN）は多層パーセプトロン（MLP）より優れているが、IoT などに使われる低パワーの機械では GPU などがなく、並列処理ができないため、CNN の適用は困難である。

その後、CNN とは異なるネットワーク構造を持つ、Attention 機構を用いた Vision Transformer [2]が提案されたが、複雑な Attention 機構を導入することによって、よい精度が得られるとは限らない。

このような状況に対し、MLP-mixer[3]が提案された。MLP-mixer は畳み込みニューラルネットワーク及び Attention を持たず、簡略な構造で高い精度が得られる。

本研究では MLP-Mixer をより効率的に学習向上させるため、Sharpness-Aware Minimization（SAM）[4]を導入し、マルウェア分類を行い、提案手法が、少数のパラメータで CNN に匹敵するほど高い分類性能を発揮することを実験により確認した。

2. MLP-mixer

本節では MLP-mixer のアーキテクチャー（図 1）について紹介する。ネットワークへの入力は 3 次元の画像パッチ(S,P,P,3)の形式である。ここでは、S はパッチ数、P はパッチのサイズである。これらのパッチは、H 次元の潜在表現に線形に投影され、Mixer 層に渡される。投影されたパッチは，X と呼ばれるテーブルのような形式に変換される。最初のブロック（Token-mixing MLP ブロック）は，X の転置に対して作用する。このブロックは，各パッチの同じ位置の情報をすべてのチャンネルに渡って集約する。2 番目のブロック（Channel-mixing MLP ブロック）は，X の転置後の行に作用する。ここでは，各パッチが取り出され，そのパッチの各チャンネルについて計算が適用される。Token-mixing ブロックがパッチを跨いで特徴を探すのに対し，Channel-mixing ブロックはそのパッチでのみ特徴を探し，それをチャンネルに関連付ける。

3. SAM　(Sharpness-Aware Minimization)

従来のニューラルネットでは損失が最小になるパラメータを学習させていたが、SAM では損失が最小かつその周りも平坦となっているパラメータを探索する。SAM のアルゴリズムを図 2 に示す。まず、パラメータ w の周辺で最大の損失をとる w+ϵ^(w)を求める。その後、w+ϵ^(w)における損失を求め、損失に対する勾配でパラメータ w を更新する。つまり、単にトレーニング損失値 LS（w）が低いパラメーター値 w を探すのではなく、近隣全体のトレーニング損失値が均一に低いパラメーター値を探すことが SAM の特徴である。

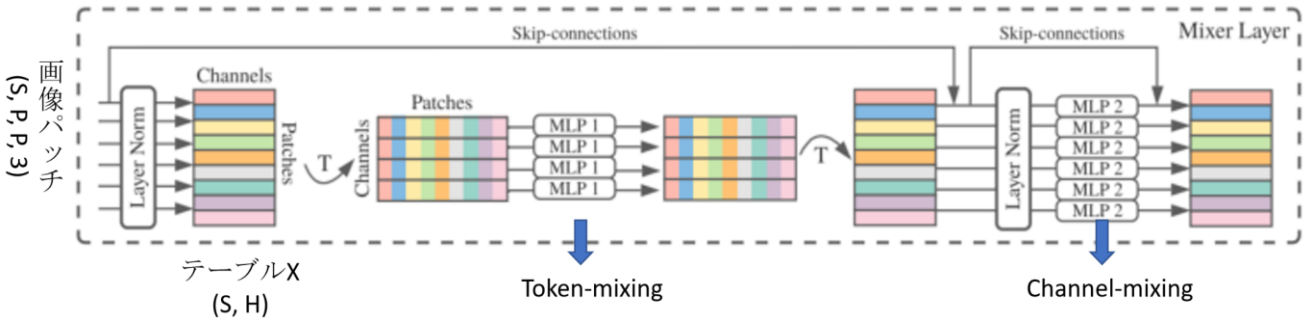
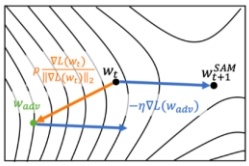


図1　MLP-mixerのアーキテクチャー[3]

† 防衛大学校情報工学科 Computer Science, National Defense Academy.

Input: Training set $\mathcal{S} \triangleq \cup_{i=1}^{n}\{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}$, Loss function $l : \mathcal{W} \times \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+$, Batch size b, Step size $\eta > 0$, Neighborhood size $\rho > 0$.
Output: Model trained with SAM
Initialize weights $\boldsymbol{w}_0$, $t = 0$;
while *not converged* **do**
  Sample batch $\mathcal{B} = \{(\boldsymbol{x}_1, \boldsymbol{y}_1), ...(\boldsymbol{x}_b, \boldsymbol{y}_b)\}$;
  Compute gradient $\nabla_{\boldsymbol{w}} L_{\mathcal{B}}(\boldsymbol{w})$ of the batch's training loss;
  Compute $\hat{\boldsymbol{\epsilon}}(\boldsymbol{w})$ per equation 2;
  Compute gradient approximation for the SAM objective (equation 3): $\boldsymbol{g} = \nabla_{\boldsymbol{w}} L_{\mathcal{B}}(\boldsymbol{w})|_{\boldsymbol{w}+\hat{\boldsymbol{\epsilon}}(\boldsymbol{w})}$;
  Update weights: $\boldsymbol{w}_{t+1} = \boldsymbol{w}_t - \eta \boldsymbol{g}$;
  $t = t + 1$;
end
return $\boldsymbol{w}_t$



Algorithm 1: SAM algorithm

図 2 SAM のアルゴリズム[4]

4. 実験

本研究では Malimg データセット[5]を用いて実験を行った。データセットは 25 の異なるマルウェアファミリーの 9,339 のサンプルで構成されている。

実験結果を表 1 に示す。SAM を適用することで、MLP-mixer の精度は 0.4%上がった。図 3 及び図 4 は SAM なしとありの学習過程であり、MLP-mixer に SAM を追加することによって、少し円滑に学習ができると見られる。

表 2 は提案手法を CNN モデルである ResNet50 と Attention 機構を持つモデルである Vision Transformer と比較した結果である。[3]と同じく、MLP-mixer は ResNet より 1.71%低い精度であるが、パラメータは 20 倍減少している。このことは、提案手法が幅広い機器に対し適用可能であることを意味する。さらに、図 5 より、提案手法は ResNet より安定した学習ができるため、コールバック EarlyStopping を利用することで、学習時間を短縮することもできる。提案手法は、ViT とほぼ同じパラメータ数であるが、精度は圧倒的に良い。本実験で用いたマルウェアの画像化データは写真のような通常の画像とは異なるため、Attention の効果が発揮できなかったと考えられる。

表 1 SAM の影響

	SAMなし	SAMあり
MLP-mixer	95.89	**96.29(+0.4)**

表 2 精度の比較

モデル	Accuracy (%)	パラメータ数
ResNet50	**98.00**	23.538.690
ViT	16.06	1.244.930
提案手法	96.29	**1.149.570**

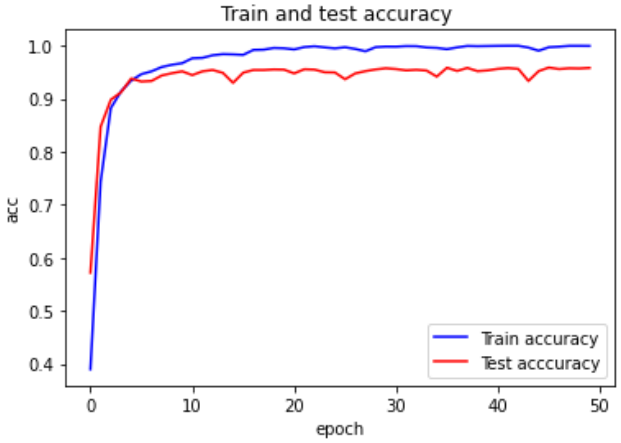


図 3 SAM なしの学習状況

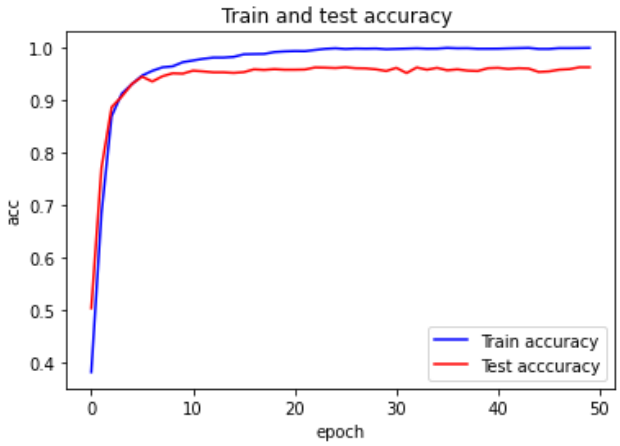


図 4 SAM ありの学習状況

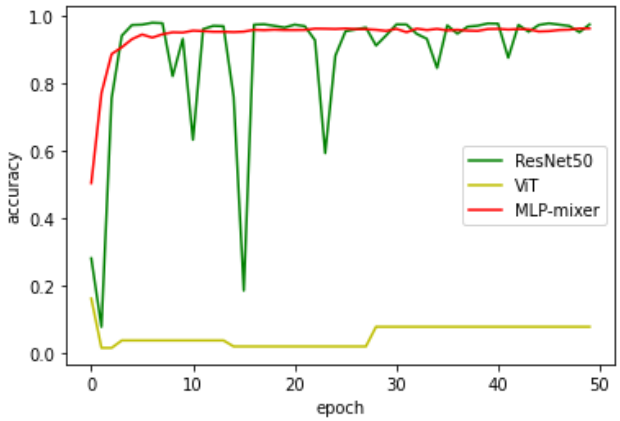


図 5 各手法の精度の比較

5. おわりに

CNN は、画像から空間情報を効率的に抽出することができるため、古くからコンピュータビジョンに君臨してきた。しかし、パラメータが多く存在するため CPU で処理することが困難であった。本研究では簡易な構造を持つ MLP-mixer に SAM を導入することによって、畳み込みや Attention を使用せずとも高い分類性能を発揮するできることをマルウェアの分類実験により確認した。

参考文献

[1] https://www.av-test.org/en/statistics/malware.
[2] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit and Neil Houlsby, “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale”, ICLR 2021.
[3] Ilya Tolstikhin, Neil Houlsby, Alexander Kolesnikov, Lucas Beyer, Xiaohua Zhai, Thomas Unterthiner, Jessica Yung, Andreas Steiner, Daniel Keysers, Jakob Uszkoreit, Mario Lucic and Alexey Dosovitskiy, “MLP-Mixer: An all-MLP Architecture for Vision”, NeurIPS 2021.
[4] Pierre Foret, Ariel Kleiner, Hossein Mobahi, Behnam Neyshabur, “Sharpness-Aware Minimization for Efficiently Improving Generalization”, ICLR 2021.
[5] L. Nataraj, S. Karthikeyan, G. Jacob and B.S. Manjunath, “Malware images: visualization and automatic classification”, Proceedings of the 8th International Symposium on Visualization for Cyber Security, 2011.