6V-02

機械学習による CAPEC の攻撃パターン間の関係の特定

宮田　陸歩†　鷲崎　弘宜†　須本　賢介†　深澤　良彰†
早稲田大学†

1. はじめに

情報セキュリティにおける脆弱性は多様化してきており，脆弱性のデータベースである Common Vulnerabilities and Exposures (CVE) [1] には, 19 万以上の脆弱性が報告されている. それらの脆弱性に対する脅威とその対策を把握するためには既知の脆弱性に対する攻撃パターンをカタログ化した Common Attack Pattern Enumeration and Classification (CAPEC) [2] が有効である. 攻撃パターンに記載されている情報として, ID や Name, Description, Relationships, Mechanisms などがある. 特に Relationships は, 関係する攻撃パターンとその性質が定義されており, この関係が特定されることで, 攻撃者が取りうる具体的な攻撃手段の洗い出しや分析が可能になり, 有効な対策を講じることが可能になる. しかし，攻撃パターンの関連付けは MITRE 社によって人手で行われているためこの関係に漏れが存在する可能性がある.

そこで本研究では, 深層学習モデルであるトランスフォーマー [3]及びグラフ構造を用いて ID が i, j である攻撃パターン p_i, p_jのペア$(p_i,\ p_j)$の関係を Name と Description から半自動的に特定することで漏れている関係を特定する.

2. 提案手法

トランスフォーマーベースのモデルである BERT と Longformer を使用し, これらを 2.2 で説明するデータによってファインチューニングすることで 6 クラス分類タスクに適応させる. さらに, 不適正・不適切なグラフ構造を定義し,予測結果の絞り込みをすることでより正確に$(p_i,\ p_j)$の関係を特定する.

2.1. 不適正・不適切なグラフ構造の定義

最初に次の命題を定義する.

$parent(p_i,\ p_j)$： p_iはp_jの親.
$precede(p_i,\ p_j)$： p_iはp_jに先行する.
$peer(p_i,\ p_j)$： p_iはp_jに類似する.
また, 推移的に次の関係が成り立つとする.

$$\exists p_i \exists p_j \exists p_k \left(parent(p_i,\ p_j) \wedge parent(p_j,\ p_k)\right) \rightarrow parent(p_i,\ p_k)$$
$$\exists p_i \exists p_j \exists p_k \left(peer(p_i,\ p_j) \wedge peer(p_j,\ p_k)\right) \rightarrow peer(p_i,\ p_k)$$

次に「不適正な構造」と「不適切な構造」を以下のような述語論理で定義する. また, このグラフ構造を図 1 に示す.

・ 不適正なグラフ構造

$$\exists p_i \exists p_j \exists p_k (parent(p_i,\ p_j) \wedge parent(p_j,\ p_k) \wedge (precede(p_i,\ p_k) \vee precede(p_k,\ p_i) \vee peer(p_k,\ p_i) \vee parent(p_k,\ p_i))) \quad (1)$$

・ 不適切なグラフ構造

$$\exists p_i \exists p_j \exists p_k (parent(p_i,\ p_j) \wedge peer(p_j,\ p_k) \wedge (precede(p_i,\ p_k) \vee precede(p_k,\ p_i))) \quad (2)$$

$$\exists p_i \exists p_j \exists p_k (peer(p_i,\ p_j) \wedge parent(p_j,\ p_k) \wedge (precede(p_i,\ p_k) \vee precede(p_k,\ p_i))) \quad (3)$$

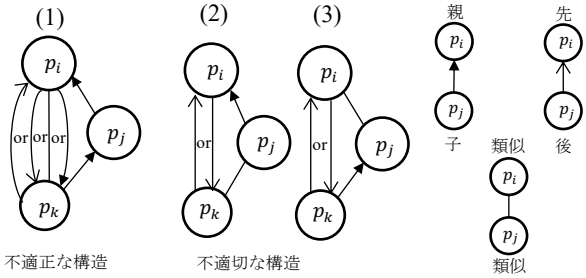

図 1 不適正・不適切なグラフ構造

2.2. 教師データの作成 / 不適正・不適切な構造の除去

現在 CAPEC では, 親子関係を示す *ParentOf, ChildOf*, 連鎖関係を示す *CanPrecede, CanFollow*, 類似関係を示す *PeerOf* という関係が定義されているため, 攻撃パターンのペア$(p_i,\ p_j)$に対してこの 5 つの関係をラベルとして付与する. さらに, 現状の関係において不適正・不適切な構造に当てはまる攻撃パターンを除去する. また, 関係がないペアに対しては *None* というラベル付与する.

2.3. ファインチューニング

BERT には[CLS]トークンと[SEP]トークンがあり, [SEP]トークンによって二つの文章を分ける事ができる. そのため, 図 2 のように$(p_i,\ p_j)$というペ

Identifying relationships between CAPEC attack patterns through machine learning
† Rikuho Miyata, Hironori Washizaki, Kensuke Sumoto, Yoshiaki Fukazawa, Waseda University

アを入力する際は[SEP]の前にp_iを[SEP]の後ろにp_jを入力することで二つの攻撃パターンの関係を学習する．Longformer については[CLS]トークンに替わり<s>トークン, [SEP]トークンに替わり</s>トークンとなるが, 入力方法に関しては BERT と同様である．さらに, 先頭のトークンである[CLS], <s>トークンと二つの攻撃の関係を区別するために用いた[SEP], </s>トークンの特徴量を結合して[SEP]</s>トークンの有意性を検証する．

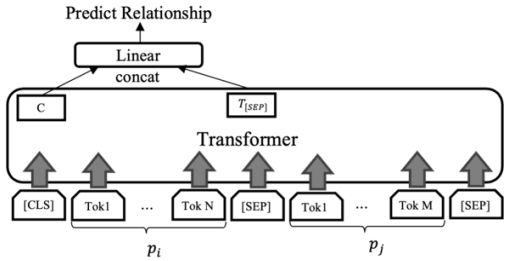

図 2. トランスフォーマーへの入力

2.4. ファインチューニング済みモデルによる関係の予測及びグラフ構造による関係の特定

推論を行うにあたり, Softmax 関数を用いることでモデルの分類における確信度を得る．BERT, Longformer の二つのモデルに, 攻撃パターンのペア(p_i, p_j)とその順序を入れ替えた(p_j, p_i)を入力し, この二つのペアの出力結果の組み合わせが対照的（親である関係と子である関係, 先行する関係と後行する関係など）でない場合, または二つのモデルの結果が異なる場合にその結果を棄却する. さらに, 予測された結果の組み合わせが親子関係の時, Mechanisms が異なる場合にも結果を棄却する．今回は二つのモデルの確信度の平均が 99%を超えた関係のみを抽出した．さらに抽出された関係において, 2.1 で示したグラフ構造が成り立つ場合は結果を棄却する．

3. 結果と考察

3.1. 不適正・不適切な構造の除去

現在では, 計 711 組の関係が定義されており, そのうち, $parent(p_{616}, p_{630})$, $precede(p_{616}, p_{691})$, $peer(p_{630}, p_{691})$ の 3 組のみが図 1 の(2)に当てはまる構造であったため, この 3 組を教師データから除いた．その結果, 今回扱うデータ数について *ParentOf, ChildOf* は 526, *CanPrecede, CanFollow* は 164, *PeerOf* は 32, *None* は 1537 になった．

3.2. BERT, Longformer の精度

表 1 にモデルの精度の結果を示す. <s>, </s>トークンの特徴量を結合した Longfomer$_{<s>+</s>}$ の Accuracy は 0.859, F1 値は 0.760 で最も高い結果を得た. 一方で, [CLS]トークンの特徴量だけを用いた BERT の値は最も低く Accuracy は 0.816, F1 値は 0.674 であった. この結果は BERT の最大入力トークン数が 512 であるのに対して, Longformer は 4096 であることに起因すると考えられる. また, BERT, Longformer 共に[SEP](</s>)トークンを結合した方が，精度が高いことから関係の分類において，[SEP] (</s>)トークンが有効に働いていると考えられる．

表 1 トランスフォーマーモデルの精度

トランスフォーマーモデル	Accuracy	F1
BERT$_{[CLS]}$	0.816	0.674
Longformer$_{<s>}$	0.853	0.772
BERT$_{[CLS]+[SEP]}$	0.838	0.735
Longformer$_{<s>+</s>}$	0.859	0.760

3.3. 提案手法によって特定された関係

今回の手法により 362 件の関係が抽出され，そのうち既存の関係からグラフ構造上, 推移的な関係が 91 件存在した. また, 第一著者と第二著者が残りの 271 件を確認し, 41 件が意味的に正しい関係と判断された. 特定された関係の一例として, *ParentOf* 関係と特定された(p_{114}, p_{509})を挙げると, p_{114}(Authentication Abuse)は認証機能の弱点を突いて不正アクセスを行う攻撃であり, p_{509}（Kerberoasting）はサービスアカウントの Kerberos 認証と SPN（Service Principal Names）の使用方法を突いて不正アクセスを行う攻撃である．そのため, (p_{114}, p_{509}) の予測結果である *ParentOf* は正しい関係だと判断した．

4. まとめ

本稿では, トランスフォーマーモデルをファインチューニングし, 攻撃パターンペアの 6 クラス分類タスクとしてモデル化し,「不適正なグラフ構造」,「不適切なグラフ構造」を定義することで現状の Relationships において漏れている関係を特定した. その結果，特定した候補群に対する著者によるレビューにより 41 件が漏れている可能性の高い関係であると特定できた．

参考文献

[1]. The MITRE Corporation. Common Vulnerability and Exposures (CVE). https://cve.mitre.org/, 2022.
[2]. The MITRE Corporation. Common Attack Pattern Enumeration and Classification (CAPEC). https://capec.mitre.org/, 2022.
[3]. Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).