1E-01

大規模言語モデルを用いた可視化データの推論手法

篠原 正紀†

日本電信電話株式会社　社会情報研究所†

1.はじめに

装置やシステム等に関する情報を可視化したデータ（以下，「可視化データ」と呼ぶ）を用いて，構成検査，ライセンス検査，脆弱性検査などを行う技術が存在する．特に，ソフトウェア構成情報を可視化し共有する方法として，SBOM[1] の利用が活発化している．

しかし，可視化データの提供については，提供者にとっての機密情報が漏洩する懸念から，全ての可視化データを提供することが難しいという課題がある．また，提供者が意図しない間違いが含まれる可能性もある．そのため，可視化データが提供されていても，検査に必要な情報が足りない場合や間違いがある場合には，それを用いた検査の精度が低下してしまう．

よって本稿では，検査の精度を向上させるため，大規模言語モデル（LLM）を用いて，不足する可視化データを推定する手法を提案する．

2.提案方式

本章では，大規模言語モデルを用いて SBOM 等の可視化データを推定する手法について説明する．本手法で用いる情報推定装置の概要を図 1に示す．情報推定装置は，学習フェーズと推論フェーズの２つの動作フェーズを実施する．

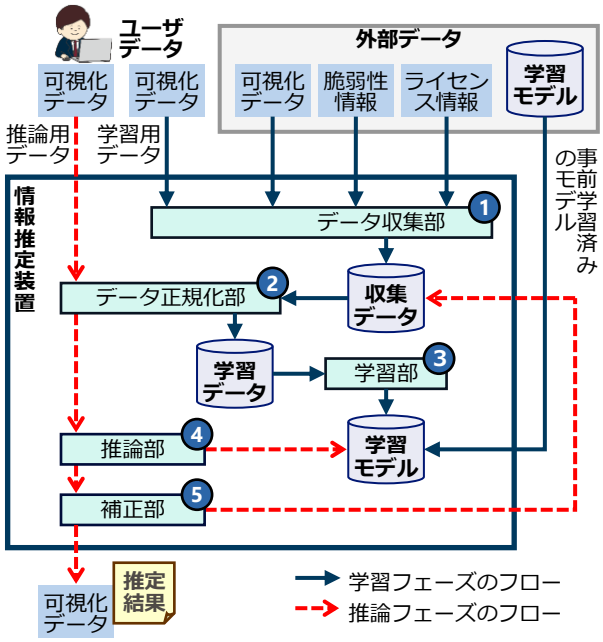

図 1　可視化データの推定動作概要

Inference Method for Visualization Data Using LLM
†Masanori Shinohara
†NTT Social Informatics Laboratories

①データ収集

事前学習，およびファインチューニングに用いるデータを収集する．収集データの一例を表 1に示す．

表 1　収集データの例

情報源	情報取得方法
公開情報（OSINT 等）	クローリング
構成情報（SBOM 等）	製品の提供者から取得
仕様書 / マニュアル	電子化されたデータとして取得
ソースコード	外部サイトもしくは製造者から取得
外形的仕様	外観画像等を実機から取得
通信ログ	実機を動作させ、通信経路上で取得
バイナリ	実機から取得
動作ログ	実機に蓄積されたログを取得
詳細な構成情報	実機で取得プログラムを動作させる

②データ正規化

データ収集部が収集した収集データに対して，推論に役立つ部分を選別，若しくは推論に役立つデータを付加し，大規模言語モデルの学習に適した形式に変換して学習データを生成する．

③学習

データ正規化部の生成した学習データを用いて，事前学習モデルの生成，もしくは，外部から取得してきた事前学習モデルに対するファインチューニングを行う．

④推論

学習部により生成されたモデルを用いて，推定対象の可視化データの推論を行う．推論の入力データも，学習データと同様に正規化されたデータを用いる．

⑤補正

推論部により推論された結果に対して，辞書データ等との類似性を計算することで，結果の補正を行い，最終的な推定結果を出力する．

3.評価実験

本章では，提案方式の評価として，可視化データの一つである CPE（Common Platform Enumeration）の CPE ネームを推定した実験結果について説明する．

3.1. 学習

学習では，T5（Text-to-Text Transfer Transformer）と GPT-2（Generative Pre-trained Transformer 2）の事前学習モデルを外部サイト（Hugging Face）より取得し，CPE に関する 261 万件の学習データを用いてファインチューニングを行った（エポック数 3 回）．T5 の学習データの例を，図 2に示す．データ本体には，NVD から取得した CPE タイトルや，CPE に対応するソフトウェアが持つ脆弱性（CVE）の説明文等を用いた．

なお，GPT-2 の学習でも，記載フォーマットが異なるが，T5 と同じ情報を含む学習データを利用した．

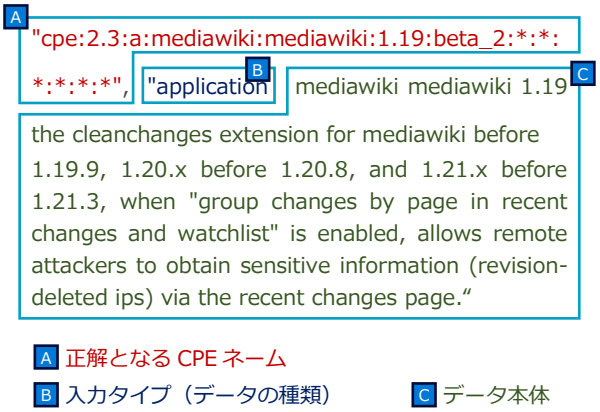

図 2　学習データの例

3.2. 推論

推論の入力データは，実際の学習には使用していない学習データから 1,000 件を抽出し，正解となる CPE ネームを削って作成した．推論により CPE ネーム部分を出力させる．T5 の推論に用いた入力データの例を図 3に示す．

なお，GPT-2 の推論でも，記載フォーマットは異なるが，T5 と同じ情報を含む入力データを利用した．

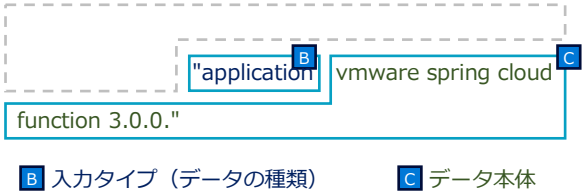

図 3　推論の入力データの例

3.3. 評価結果

CPE ネームを脆弱性検査に用いる場合，種別，ベンダ名，製品名，バージョンの４項目が正しく出力されなければ脆弱性の判断が困難となるため，ここでは，種別，ベンダ名，製品名，バージョンの４項目が全て正しい場合に，生成された CPE ネームが正解であると定義する．例えば，ある CPE ネームが生成された場合に，表 2 で示すように，種別，ベンダ名，製品名は正しいが，バージョンが誤っている場合には，不正解とする．

表 2　評価基準の適用例

正解	cpe:2.3:**a:php:php:4.0.6**:*:*:*:*:*:*:*		
推論結果	cpe:2.3:**a:php:php:4.0.7**:rc1:*:*:*:*:*:*		
判定	**NG**		
		Part（種別）	OK
		Vendor（ベンダ名）	OK
		Product（製品名）	OK
		Version（バージョン）	**NG**

評価結果を表 3に示す．完全一致は，推論部による推論出力がそのまま正解と判断された場合を表し，補正後一致は，補正部において辞書との照合による補正を行った結果，正解と判断された場合を表す．

表 3　評価結果

学習データ数	T5			GPT-2		
	合計正解数	完全一致	補正後一致	合計正解数	完全一致	補正後一致
2,000	612	508	104	686	547	139
5,000	748	648	100	716	588	128
10,000	789	699	90	723	607	116
20,000	**815**	734	81	753	649	104
50,000	741	651	90	761	661	100
100,000	644	503	141	762	667	95
200,000	727	616	111	757	664	93
500,000	366	200	166	772	688	84
1,000,000	487	317	170	773	682	91
2,537,000	749	551	198	**810**	695	115

T5 を用いた場合，ファインチューニングに使用するデータ数が 2 万件付近で完全一致および合計正解数ともにピークを示し，その後増減して，50 万件以上で精度改善の兆しが見られた．また，GPT-2 を用いた場合，完全一致および合計正解数ともに，ファインチューニングに使用するデータ数が増えるにしたがって，推論における正解数が向上する．いずれのモデルを用いた場合にも，最大の正解率は 80%を超えることを確認できた．

4.まとめと今後の課題

本稿では，大規模言語モデル（LLM）を用いて，SBOM 等の可視化データを推定することにより，検査の精度を向上させる手法を提案した．可視化データとして，CPE ネームを推定する場合について評価実験を行い，事前学習モデルとして T5 および GPT-2 を用いた場合に，最大の正解率は 80%を超えることを確認した．これにより，情報が不完全な場合でも，情報の推定を行って検査に利用したり，また取得可能な可視化データについても，別途推定した可視化データの正確性を確認することで，検査の精度をより向上させることができるようになる．

今後の課題としては，CPE 文字列の共起性や脆弱性の類似性に基づいた補正手法[2]と組み合わせて補正を行うことなどが考えられる．さらには，CPE ネームの推定だけではなく，対象をより多くの可視化データの項目に広げていく必要がある．そのためには，より多くの情報源から学習用のデータを収集すると共に，GPT-4 など新しい事前学習モデルの適用や，効果的な学習および推論方式などを検討し，精度を向上させていく必要がある．

参考文献

[1] The Minimum Elements For a Software Bill of Materials (SBOM)
https://www.ntia.doc.gov/files/ntia/publications/sbom_minimum_elements_report.pdf
[2] 佐々木満春，山嵜麿与：“脆弱性管理のための CPE マッチング手法の提案”，情報処理学会 コンピュータセキュリティシンポジウム 2021 論文集 484-491, 2021-10-19