Title	A study on graph neural networks and pretrained models for analyzing cybersecurity texts
Author(s)	Nguyen, Chau Minh
Citation	
Issue Date	2021-09
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/17546
Rights	
Description	Supervisor:NGUYEN, Minh Le, 先端科学技術研究科, 修士（情報科学）

A study on graph neural networks and
pretrained models for analyzing cybersecurity texts

1910409 NGUYEN, Chau Minh

Analyzing cybersecurity texts is the task of identifying malware actions and determining their characteristics in the documents about cybersecurity threats, utilizing natural language processing (NLP) techniques. This task consists of four subtasks: (1) identifying malware-related sentences (i.e., sentences which describe malware actions) from cybersecurity texts, (2) identifying token labels (i.e., *Malware Action*, *Subject of Action*, *Objectof Action*, and *Modifier of Action*) in malware-related sentences, (3) identifying relation labels (i.e., *Subject-Action*, *Action-Object*, *Action-Modifier*, and *Modifier-Object*) between tokens, and (4) classifying malware actions into attribute labels. The attribute labels are defined and enumerated in the Malware AttributeEnumeration and Characterization (MAEC). Specifically, based on malware's behaviors and attack patterns, MAEC classifies malware into four categories, including *ActionName*, *Capability*, *StrategicObjectives* and *TacticalObjectives*; each category includes multiple malware attribute labels: 211 *ActionName* labels, 20 *Capability* labels, 65 *StrategicObjectives* labels, and 148 *TacticalObjectives* labels, results in a total of 444 attribute labels.

Recently, researchers in several disciplines have acknowledged the superior performance of graph neural networks (GNNs). Many NLP researchers also employed GNNs in multiple NLP tasks and achieved promising performance. Besides, pretrained language models are nowadays widely employed because of their robustness in language understanding. In this research, we aim to study on how GNN models and pretrained models can be employed for the task of analyzing cybersecurity texts. Specifically, in this research, we address all four subtasks. The experiment results demonstrate that our proposed models for the subtask 1 and subtask 2 achieve state-of-the-art performances on the *MalwareTextDBv2.0* dataset, which is the largest dataset for malware characteristic analysis.

For subtask 1, we propose two methods to exploit knowledge from an external document, which is the Attribute Reference Guide, to enrich the representation of the sentences. The first method utilizes GATs (Graph ATtention networks) for producing a *weak label* for each sentence. The second method considers the likelihood a sentence belongs to each of 444 malware attribute labels. We use those features to enrich the base features for representing sentences, and achieve the state-of-the-art performance for subtask 1 on the *MalwareTextDBv2.0* dataset.

For subtask 2, we propose an BERT-CRF model (a conditional random field layer on top of a BERT model) for the task of token labelling. With the post-processing phase, which utilizes the our predictions from subtask 1, we achieve the state-of-the-art performance for subtask 2 on the *MalwareTextDBv2.0* dataset.

In subtask 3, we employ the handcrafted rules to generate relation labels. After that, we address the coreference issue to construct a graph and visualize it.

For subtask 4, we propose to employ a GNN model for the task of malware action classification.

The experiments demonstrate the promising results of neural networks (in general) and graph neural networks (in particular) for the task of analyzing cybersecurity texts.

Keywords: cybersecurity texts, analyzing cybersecurity texts, cybersecurity text analysis, graph neural networks, pretrained language models, deep learning