学位報告４

報告番号	※甲　　　　第　　　　号

主論文の要旨

論文題目　Effective Application of Natural Language Processing Techniques in Automated Cyber Threat Intelligence (サイバー脅威情報自動処理における自然言語処理適用に関する研究)

氏　　名　MENDSAIKHAN Otgonpurev

論文内容の要旨

The digital age has presented various opportunities to society and to business in general. However, these opportunities also bring with them different kinds of risks such as cyber-attacks, data breaches, loss of intellectual property, financial fraud, etc. To mitigate and minimize these risks the field of cybersecurity has been developing various defense approaches. One such approach is called Cyber Threat Intelligence (CTI) which utilizes existing knowledge about cyber threats to proactively mitigate a cyber-attack before it happens, or resolve the intrusion with minimal damage using the systematic knowledge. Through the CTI, the organizations are able to systematically identify the threat actors, prioritize the defense, share the threat indicators with each other, and mitigate them effectively.

On the other hand, with the growth of the processing power and accumulation of digital data, various forms of machine intelligence are forming that would further complicate the cyber threat landscape. Once a theoretical only concept such as techniques of Machine Learning (ML), Natural Language Processing (NLP), etc have been applied in the day to day life and it is a matter of time for cyber adversaries to utilize those approaches efficiently. In this arms race, defenders have to embrace the technological shift, so that the latest researches of machine intelligence have to be embedded in the defensive tools. There have been various approaches to utilize ML and other Artificial Intelligence (AI) fields in cyber defense, especially in network security through various Intrusion Detection Systems (IDS), Security Information and Event Management (SIEM) systems, etc. However, I believe that cyber defense could be further improved by utilizing NLP and other AI techniques in the CTI process. To do t hat, I propose to apply various NLP techniques in the prototype system that could be used for collecting, analyzing, and enriching cyber threat information in text format. Natural Language Processing is a subfield of Artificial Intelligence that processes and analyzes the natural language into machine-understandable form. Hence, the proposed prototype system identifies and extracts the cyber threat-related information from massive textual documents, analyzes their significance and relevance to the user, and finally enriches the document with the more systematic threat information.

The goal of this thesis is to prove the applicability of the various NLP techniques in the CTI process. Therefore, the proposed system uses embedding, transfer learning, text classification, Named Entity Recognition (NER), Knowledge Graph, various algorithms of the single label and multi-label classification methods to demonstrate the effectiveness of such techniques in the CTI process.

The proposed system consists of the following modules.

- Natural Language Filter module classifies and filters the cybersecurity-related text documents from any information source. It has been implemented using Doc2Vec and BERT language models to identify and filter the security-related text documents. The evaluation results show that the BERT-based Natural Language Filter outperformed Doc2Vec-based Natural Language Filter by 28 points in the F1 Score. From the experimental results, it can be seen that a language model pre-trained with generic knowledge (Wikipedia and books) performs better when fine-tuned with few domain-specific data as compared to a language model that has been trained on a large amount of domain-specific data. Hence, it has been concluded that BERT would be the most suitable language model to implement the Natural Language Filter module for the proposed system.
- Analyzer module determines the significance and relevance of the threat information to the user. It has been implemented using a novel approach of engineering the features of the text through Knowledge Graph and Named Entity Recognition methods. I believe this approach could potentially address the problem of processing a massive amount of unstructured text for cybersecurity situational awareness. I propose to do it, through textual similarity with pre-defined important documents that the significance of the text can be determined and that by utilizing the existing Cybersecurity Knowledge Graph to correlate the named entities, the subjective relevance of the cybersecurity text could be found. For that, I trained a custom Named Entity Recognition model using over 17 million words and constructed a Cybersecurity Knowledge Graph with 221,202 semantic tuples in order to generate features that would represent their significance and subjective relevance. Combining these features, the significance and relevance of the text document could be represented in quantitative numbers. Due to the

constraints such as a probable lack of identifiable cybersecurity named entity in test data and the uncertainty of identified Mentioned Entities to exist in CKG the effectiveness of the proposed architecture could not be proven directly on the raw test documents; however, by simulating the controlled environment by manipulating the test document achieved a classification accuracy of 88% using the logistic regression classifier. Since it is impossible to expect the controlled environment in a real-life situation, the experiment has to be improved to reconcile the simulated dataset with real-life data. I believe by improving the NER performance and extending the scope of CKG the experiment would come closer to producing production-grade results.

- Mapper module enriches the threat information with the adversarial tactics and techniques. It uses an approach to automatically map the vulnerability information to adversary techniques in the cybersecurity context. Vulnerability descriptions have been converted into vector space and experimented with various multi-label classification methods to identify the most suitable method to map the vulnerability into MITRE ATT&CK adversarial techniques. 8,077 examples from open datasets prepared by ENISA have been used for this experiment, of which 7,877 have been used to train and test 7 multi-label classification methods in 9 evaluation measures. A comprehensive analysis of the remaining 200 examples as a prediction only task has been also performed using the best performing neural LabelPowerset model. Due to the partial nature of the experimental dataset, the experimental result could not be fully tested in real-life scenarios. However, in the given dataset, the chosen methods show good performance, indicating a comprehensive dataset may yield a production-ready system that could be used to automate and prioritize the cyber defense operations.

The individual results of the independent experiments support the utilization of particular method for that problem. Essentially, it could be inferred that by utilizing various Natural Language Processing techniques in the Cyber Threat Intelligence process the cyber defense could be improved, specifically in situational awareness and security automation operations.