氏名	MAI Trong Khang		
学位の種類	博士（情報科学）		
学位記番号	博情第 558 号		
学位授与年月日	令和 7 年 9 月 24 日		
論文題目	Applications of Natural Language Processing and Large Language Models to Cybersecurity		
論文審査委員	BEURAN Razvan Florin	JAIST	准教授
	丹 康雄	同	教授
	リム 勇仁	同	教授
	井之上 直也	同	准教授
	柳生 智彦	日本電気株式会社	主任研究員

論文の内容の要旨

Like other domains, cybersecurity knowledge can be stored in textual data to facilitate cybersecurity practitioners' analysis, interpretation, and communication. For example, Cyber Threat Intelligence (CTI) reports, network design documents, system logs, and security guidelines are cybersecurity assets in a textual format. This leads to the crucial role of Natural Language Processing (NLP) in cybersecurity. However, state-of-the-art NLP models based on Deep Learning (DL) architecture (e.g., transformer) necessitate extensive training with significant labeled data. This issue demands the strong involvement of experts to create sufficient labeled data to ensure the model's performance. Additionally, the fast-changing nature of the cybersecurity field causes training data and the developed models to be quickly outdated, diminishing the practical applicability of these approaches. As a result, addressing the labeled data insufficiency and developing models with greater generalizability have become emerging trends in NLP and DL.

There are significant NLP and DL trends that can address the insufficiency of annotated data in developing and maintaining cybersecurity applications:

1. The Weak Supervision approach harnesses existing labeled data sources to create a new weak dataset suitable for model development. Weak Supervision-based approaches significantly reduce the need for extensive data labeling when developing new models for emerging problems.
2. The use of published open-source frameworks for NLP, like SpaCy, simplifies the development of NLP applications for those with limited linguistic knowledge. These frameworks provide tools to analyze sentences, paragraphs, and documents to gain linguistic insights. Then, the developers can take advantage of the analyzed results to solve their target problems. For example,

the obtained grammatical relationships can be used to replace the actual cybersecurity relationships among entities.

3. The recent appearance of Large Language Models (LLMs) introduces promising approaches to resolving data insufficiency problems in cybersecurity. With billions of parameters pre-trained on vast datasets, LLMs demonstrate strong generalizability, enabling them to tackle unseen tasks with very few labeled examples.

In this work, we aim to address the insufficiency problems of annotated data in developing cybersecurity applications. To achieve this goal, we studied advancements in NLP, including Weak Supervision and LLMs, and their feasibility to tackle practical downstream tasks. Additionally, we sought to create applications that can support cybersecurity experts. We applied these advancements to specific downstream tasks to develop practical applications, such as report analysis and policy generation.

Our first application was a framework called RAF-AG, which supports the information-sharing process in cybersecurity. RAF-AG can transform CTI reports into simplified versions, such as attack paths. In developing RAF-AG, we utilized Weak Supervision and open-source NLP tools to utilize already annotated data from similar problems to solve the target problem. For evaluating RAF-AG, we collected 30 CTI reports from various sources and compared its results with those generated by a similar report analysis framework, AttacKG. It was shown that RAF-AG can outperform AttacKG in precision, recall, and F1 scores, recording values of 0.717, 0.722, and 0.708 compared to 0.337, 0.535, and 0.393, respectively.

We recognized the limitations of RAF-AG and aimed to study new models that demonstrate high generalizability, eliminating the need for text normalization. The emergence and popularity of LLMs brought up new potential for this thesis. We utilized commercial LLMs to develop a framework for policy generation. This application aimed to assist experts in creating fine-grained access control policies tailored to a specific IT environment. We employed a typical ICS network as a case study to create 181 fine-grained ABAC policies. To enhance the access control performance of generated policies, we implemented priority optimization for policy conflict resolution. Our tests with various optimization algorithms showed that optimized priority values can significantly improve the effectiveness of the generated policies, resulting in an F1 score of 0.994.

We examined the benefits and drawbacks of previous applications. This turned the focus to open-source LLMs to develop CyLLM-DAP, a framework designed to

support the specialization of LLMs in cybersecurity. This effort promotes an effective DL technique for data scarcity, namely transfer learning, where we inject cybersecurity knowledge into open-source LLMs so that the models can be reused to better solve cybersecurity downstream tasks. The aim of this effort is to create cybersecurity-specific LLMs (CyLLMs). Our experiment showed that cybersecurity-specific LLMs can lead to significant performance enhancements (up to 4.75%) in downstream tasks such as text classification and Q&A when compared to the general base and instruct counterparts. Additionally, using insights from previously developed applications such as RAF-AG, CyLLM-DAP, and CyLLMs, we developed a methodology to work with cybersecurity problems where annotated data insufficiency is present. We also included a report analysis approach based on the proposed methodology.

For each of the mentioned tasks, we began by conducting a survey to identify the advantages and disadvantages of current approaches. Next, we developed a novel methodology to tackle the current issues. Based on the availability of existing approaches published in other research, our experiments successfully (1) demonstrated the effectiveness of the proposed techniques and (2) identified the best methodology among those available. Ultimately, the methodologies, models, and data in this work were published to assist in addressing similar downstream tasks in cybersecurity.
Keywords: Weak Supervision, Large Language Models, data insufficiency, CTI report analysis, policy generation, cybersecurity

論文審査の結果の要旨

The dissertation of Mr. Mai Trong Khang deals with the application of modern natural language processing (NLP) techniques and large language models (LLMs) to the cybersecurity domain. This is an important topic because a significant amount of knowledge about cybersecurity, such as Cyber Threat Intelligence (CTI) reports, network design documents, and security guidelines, is currently represented in text format.

Mr. Mai outlined in his dissertation several NLP and Deep Learning (DL) techniques that can be used to address the issue of annotated data insufficiency, which is critical in order to make it possible to leverage the full potential of NLP and LLMs in the cybersecurity domain. He then proceeded to extend and apply these techniques to several important applications related to cybersecurity.

One of his contributions, the RAF-AG framework, utilizes weak supervision and open-source NLP tools to extract essential information from CTI reports, such as attack graphs and attack paths. Another application is related to the use of commercial LLMs for the generation of fine-grained attribute-based access control (ABAC) policies that are tailored to specific IT environments. The generation was conducted by leveraging security guidelines documents relevant to those IT environments that are available in natural language form.

Another important contribution of Mr. Mai is the development of the CyLLM-DAP framework that supports the specialization of LLMs to cybersecurity via domain-adaptive pretraining techniques. This effort made it possible to create several cybersecurity-specific LLMs (CyLLMs) that were further used in various applications, such as the automated analysis of CTI reports.

The experimental evaluation of the frameworks that Mr. Mai developed has demonstrated the effectiveness of the proposed techniques with respect to equivalent approaches. For example, an F1 score improvement of more than 0.3 was obtained for RAF-AG compared to a previous work, AttacKG. Moreover, a 0.994 F1 score was achieved for ABAC policy generation after applying a novel optimization approach that the author developed.

The work of Mr. Mai has been presented in a top-level journal in the field of security, Elsevier Computers & Security (CiteScore 13.3, IF 5.4), as well as in two peer-reviewed international conference papers. Moreover, he received the best student paper award for one of those papers, which is related to CyLLM-DAP. This demonstrates both the validity of his results, and the high level of English language skills that he has.

Based on the above considerations, we conclude that this is an excellent dissertation, and we approve awarding the doctoral degree to Mr. Mai Trong Khang.