氏　　名	Luu Thanh Son		
学位の種類	博士（情報科学）		
学位記番号	博情第 568 号		
学位授与年月日	令和 8 年 3 月 25 日		
論文題目	END-TO-END MULTIMODAL FACT-CHECKING SYSTEM FOR SOCIAL NEWS		
論文審査委員	Nguyen Le Minh	JAIST	Professor
	SHIRAI Kiyoaki	JAIST	Professor
	Satoshi Tojo	Asia University	Professor
	Quan Thanh Tho	HCMUT, VNU-HCMC	Associate Professor
	Naoya Inoue	JAIST	Associate Professor

論文の内容の要旨

Fact-checking is the process by which journalists verify whether a claim is true. With the rapid spread of misinformation and the rapid growth of social media, it is crucial to automate fact-checking to help network moderators reduce its impact on social news platforms. Moreover, the output of the fact-checking task requires not only the stance for the claim but also the explanation for the stance of the claim. In addition, current trends combine data from multiple modalities, such as text and images; thus, the fact-checking system must be able to process multimodal data.

This thesis aims to construct an end-to-end Fact-checking system on a multimodal dataset, comprising three main stages: retrieving evidence, verifying the claim based on the evidence, and justifying the claim's truthfulness. There are several appropriate approaches for an automatic end-to-end Fact-checking system, such as deep learning with multimodal fusion techniques and a sequence-to-sequence model for generating results in natural language. However, the main challenge for the deep learning approach is the lack of large-scale human-annotated data and potential bias in the dataset. Large language models (LLMs) trained on vast amounts of data can exploit valuable latent patterns even with limited or no training data, enabling efficient instruction prompting for fact-checking tasks. The flexibility of LLMs can then be leveraged to construct an end-to-end Fact-checking system.

We proposed MCVE - a transformer model with fusion techniques that efficiently combine text and image feature vectors within a classification network to determine the truthfulness label of the claim. Besides, we introduce a simple idea that leverages the image caption as a feature to combine with textual evidence, forming the input sequence to a sequence-to-sequence model for generating a justification of the claim's veracity. The experimental results show the efficiency of our deep neural network-based fusion method for the Multimodal fact-checking task, as evaluated on two multimodal fact-checking datasets. Next, we propose three instruction prompting frameworks — ZeFAV, TabV4FC, and M-RAV — to leverage the robustness of LLMs for verifying claims across diverse evidence modalities: text, images, and tables. Then, we evaluate our frameworks on different fact-checking datasets with multimodal evidence, including text-only, text-image, and

text-table, and the empirical results show the efficiency of our proposed frameworks in leveraging the LLMs for the Fact-checking task. Finally, we introduce FALCON - an end-to-end system that implements LLMs with a UI to demonstrate the fact-checking task in practice. The system can be hosted entirely locally with a single GPU, without any external API dependencies, demonstrating the potential for scalability and maintenance in a real system.

In the future, we would like to improve the performance of the end-to-end fact-checking system by optimizing inference runtime and integrating temporal information into LLMs to better understand time-based events in the data and verify the correctness of information, especially time-sensitive data such as financial, legal, and political.

Keywords: fact-checking, deep neural network, large language models, multimodal, end-to-end systems

論文審査の結果の要旨

This dissertation addresses the increasingly important problem of automated multimodal fact-checking, motivated by the rapid spread of misinformation on social media platforms. The work is well-grounded in real-world needs, emphasizing not only claim verification but also the generation of human-understandable justifications, which is essential for practical deployment. The thesis makes a solid contribution by developing an end-to-end multimodal fact-checking system that integrates evidence retrieval, claim verification, and justification generation.

The proposed model (MCVE) demonstrates effective multimodal fusion of textual and visual features for claim verification, and the idea of leveraging image captions as an intermediate representation for justification generation is simple yet effective. The experimental results on multiple multimodal fact-checking benchmarks validate the effectiveness of the proposed deep learning–based approaches. A key strength of the work lies in its systematic exploration of **LLM-based instruction prompting frameworks**—**ZeFAV**, **TabV4FC**, and **M-RAV**—for fact-checking across heterogeneous evidence modalities, including text, images, and tables. These frameworks demonstrate how large language models can be leveraged effectively in low-annotation settings, addressing data scarcity and bias issues commonly encountered in fact-checking datasets. The work is well-structured, experimentally grounded, and makes meaningful contributions at the intersection of multimodal learning, large language models, and fact-checking. The quality and scope of the contributions are suitable for publication at the leading journals, and the thesis provides a solid foundation for future research on scalable, trustworthy fact-checking systems.

In conclusion, the thesis is technically sound, well-executed, and experimentally rigorous. This is an excellent dissertation and we approve awarding a doctoral degree to Luu Thanh Son.