Topic: Fake News Detection Using Large Language Models

Presenter: Kaixiang Zhao (kzhao5@nd.edu)

Overall Research Direction:

This paper studies how large language models (LLMs) can be used to improve fake news detection. It examines whether LLMs can detect fake news using their internal knowledge and, more importantly, whether the diverse, multi-perspective rationales generated by LLMs can help task-specific small language models (SLMs) such as BERT. The research investigates the potential of using LLMs as advisors rather than as stand-alone detectors.

Specific Research Questions:

• "Can LLMs help detect fake news with their internal knowledge and capability?"
• "What solution should we adopt to obtain better performance using LLMs?"

Challenges and Research Elements:

Problem: Although LLMs can produce rich, multi-perspective rationales, they are not effective at making direct veracity judgments because they struggle to select and integrate these insights properly.

Claims: The paper argues that LLMs should not replace fine-tuned SLMs. Instead, LLMs can serve as valuable advisors by providing instructive rationales that, when adaptively fused with SLMs, enhance overall fake news detection.

Evidence: Experiments were conducted on two real-world datasets—a Chinese dataset (Weibo21) and an English dataset (GossipCop). The results demonstrate that the combined approach using the adaptive rationale guidance (ARG) network, as well as its distilled version (ARG-D), outperforms methods based solely on LLMs or SLMs. These experiments show that the integration of LLM-generated rationales with SLMs leads to more accurate fake news detection.

Statistical Analysis: The study employs evaluation metrics such as macro F1 scores and overall accuracy to compare different methods. Ablation studies, which remove key components like the LLM judgment predictor and the rationale usefulness evaluator, reveal significant drops in performance.