THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

Department of Computing
電子計算學系

COMP RESEARCH STUDENT SEMINAR

Date : 8 January 2025 (Wed)
Time : 3:00 pm - 4:00 pm
Venue : PQ303 (Face-to-face)

On Leveraging Large Language Models for Ehancing Entity Resolution - A Cost-Efficient Approach

Abstract

Entity resolution, the process of identifying and merging records that refer to the same real-world entity, is essential in sectors such as e-commerce, healthcare, and law enforcement. With the advancement of Large Language Models (LLMs), they present a promising method to address this task through their linguistic capabilities. The ``pay-as-you-go'' nature of usage is particularly advantageous for non-experts in data science. However, most LLMs operate on a billing model that charges per API request, based on the number of tokens used, which can become costly in large-scale scenarios when all possible matching pairs are submitted for evaluation. In this paper, we explore how to utilize LLMs to enhance entity resolution results in a cost-effective manner. We introduce a novel uncertainty reduction framework that dynamically selects the most valuable matching questions for LLM verification, and subsequently adjusts the probability distribution of possible matches based on the responses from LLMs. Additionally, we design an error-tolerant technique to handle the potential mistake in LLM outputs. The experimental results demonstrate that our methods are both cost-efficient and effective, offering promising applications in real-world scenarios.

Mr Huahang LI
PhD student
Department of Computing

About the Speaker

Mr Huahang LI received his bachelor's degree in Computer Science from South China Normal University in 2023. He is now a PhD student at the Department of Computing, The Hong Kong Polytechnic University, under the supervision of Chen Zhang. His research interest boradly include Data Preparation, Data-centric AI, Generative AI, Machine Learning, and so on.

Interpretation-Empowered Neural Cleanse for Backdoor Attacks

Mr Liangbo NING
PhD student
Department of Computing

About the Speaker

Mr Liangbo NING is currently a PhD student of the Department of Computing, Hong Kong Polytechnic University, under the supervision of Dr Wenqi FAN and Prof. Qing LI. He received his Bachelor's and Master's degrees from Northwestern Polytechnical University (NPU), Xi'an, China, in 2020 and 2023, respectively. His research interest mainly focuses on Adversarial Attacks, Domain Adaptation, Transfer Learning, and Recommendation Systems. Personal Website: https://biglemon-ning.github.io/

Abstract

Backdoor attacks have posed a significant threat to deep neural networks, highlighting the need for robust defense strategies. Previous research has demonstrated that attribution maps change substantially when exposed to attacks, suggesting the potential of interpreters in detecting adversarial examples. However, most existing defense methods against backdoor attacks overlook the untapped capabilities of interpreters, failing to fully leverage their potential. In this paper, we propose a novel approach called interpretation-empowered neural cleanse (IENC) for defending backdoor attacks. Specifically, integrated gradient (IG) is adopted to bridge the interpreters and classifiers to reverse and reconstruct the high-quality backdoor trigger. Then, an interpretation-empowered adaptative pruning strategy (IEAPS) is proposed to cleanse the backdoor-related neurons without the pre-defined threshold. Additionally, a hybrid model patching approach is employed to integrate the IEAPS and preprocessing techniques to enhance the defense performance. Comprehensive experiments are constructed on various datasets, demonstrating the potential of interpretations in defending backdoor attacks and the superiority of the proposed method.