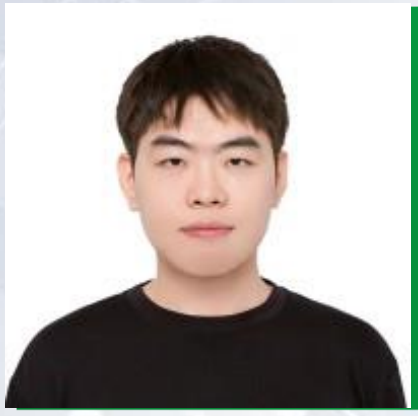
香港浸會大學
HONG KONG BAPTIST UNIVERSITY

DEPARTMENT OF COMPUTER SCIENCE
計算機科學系

TMLR Young Scientist SEMINAR

2024 SERIES

Trustworthy Machine Learning and Reasoning Group



Mr. Wei Xiong

PhD student,
Computer Science,
University of Illinois Urbana-Champaign.

Date: 13 Nov 2024 (Wednesday)

Time: 10:00 – 11:00 (HKT)

Meeting: https://meeting.tencent.com/dm/yIrmabVB5ScV

Building math agent with iterative multi-turn direct preference learning

ABSTRACT

Recent studies have shown that large language models' (LLMs) mathematical problem-solving capabilities can be enhanced by integrating external tools, such as code interpreters, and employing multi-turn Chain-of-Thought (CoT) reasoning. While current methods focus on synthetic data generation and Supervised Fine-Tuning (SFT), this paper studies the complementary direct preference learning approach to further improve model performance. However, existing direct preference learning algorithms are originally designed for the single-turn chat task, and do not fully address the complexities of multi-turn reasoning and external tool integration required for tool-integrated mathematical reasoning tasks. To fill in this gap, we introduce a multi-turn direct preference learning framework, tailored for this context, that leverages feedback from code interpreters and optimizes trajectory-level preferences. This framework includes multi-turn DPO and multi-turn KTO as specific implementations. The effectiveness of our framework is validated through training of various language models using an augmented prompt set from the GSM8K and MATH datasets. Our results demonstrate substantial improvements: a supervised fine-tuned Gemma-1.1-it-7B model's performance increased from 77.5% to 83.9% on GSM8K and from 46.1% to 51.2% on MATH. Similarly, a Gemma-2-it-9B model improved from 84.1% to 86.3% on GSM8K and from 51.0% to 54.5% on MATH.

BIOGRAPHY

Wei Xiong is a second-year Ph.D. student in computer science at UIUC, working with Tong Zhang and Nan Jiang. He also concurrently works with Google Deepmind as a full-time or part-time student researcher. His research interests focus on the theoretical understanding of decision-making problems and the practical algorithm designs inspired by mathematical insights.

ENQUIRY

Email: bhanml@comp.hkbu.edu.hk