

DO, Van Quyet

Email: vqdo@connect.ust.hk

Homepage: <https://dovanquyet.github.io/>

Research Interests

I currently work on Commonsense Reasoning and Large Language Models (LLM). Even though LLMs excel in many NLP tasks, they still fall short in problems that deal with implicit and unseen commonsense knowledge. Thus, my current research aims to find LLM prompting techniques to better deal with commonsense reasoning, as well as the design of LLMs to enable cognitive functionalities such as abstraction in the models.

Education

- **M.Phil, Computer Science and Engineering** Aug 2022 - Present
Hong Kong University of Science and Technology (HKUST)
Supervisor: Yangqiu Song
Specialization: Natural Language Processing, Data Mining
CGA: 4.0/4.3 as of Fall'22
- **B.Sc, Data Science and Pure Math (Advanced)** Aug 2018 - Jun 2022
Hong Kong University of Science and Technology (HKUST)
CGA: 4.0/4.3 as of Graduation
First Class Honors with Academic Achievement Medal

Experience

- **Software Engineer Assistant** Jun 2022 - Aug 2022
Eureka FinTech Limited, Hong Kong
- Work on the core (NLP) engine, including Data Crawling and Information Extraction
- **Research Assistant** Sep 2021 - Jun 2022
HKUST Knowledge Computation Group, led by Prof. Yangqiu SONG
- Help to train and test Knowledge Models
- Propose, experiment ideas to populate Commonsense Knowledge Graph
- **AI Internee** Mar 2021 - Sep 2021
Vietnam Technology International, Research and Development group
- Involved in projects in Computer Vision and Natural Language Processing (NLP), gained interest and specialty in NLP
- Key person of an internal project in Machine Translation, took part in almost all aspects of the project (in MLOps cycle)

Awards and Honors

- Academic Achievement Medalist of HKUST UG Class of 2022 [2022]
- HKUST Chern Class Achievement Scholarship with outstanding performance [2022]
- Awardees for the 16th, 17th HKUST Epsilon Fund Award [2021, 2022]
- Dean's List Student of HKUST [All semesters except Spring 2021]
- The Bronze Medal at the 58th International Mathematical Olympiad [2017]

Skills and Hobbies

- Technical: Python, PyTorch, familiar to TPU-training and Cloud Computing.
- Soft skills: Motivate others, work effectively under pressure and limited supervision.
- Organizational skills: Serve in the organizing committee of Vietnam Students' Association in Hong Kong.
- Language: Vietnamese (Native), English (proficient), Chinese (beginner).
- Hobbies: I like playing badminton, and especially love singing.

Publications

*Remark: * indicates equal contribution*

2023

- **COLA: Contextualized Commonsense Causal Reasoning from the Causal Inference Perspective**
Zhaowei Wang, **Quyet V. Do**, Hongming Zhang, Jiayao Zhang, Weiqi Wang, Tianqing Fang, Yangqiu Song, Ginny Y. Wong, S. See
ACL 2023
 - Introduce a new task of contextualized commonsense causal reasoning, with a new benchmark crowdsource-annotated dataset,
 - Propose a theoretical framework and an implementation, namely COLA, to solve the task from the causal inference perspective
- **CKBP v2: An Expert-Annotated Evaluation Set for Commonsense Knowledge Base Population**
Tianqing Fang*, **Quyet V. Do***, Sehyun Choi, Weiqi Wang, Yangqiu Song
Preprint, arXiv:2304.10392
 - Introduce a new expert-annotated evaluation set, that addresses the problems of its predecessor CKBP v1,
 - Conduct extensive experiments with state-of-the-art methods for CSKB Population and LLMs on the new evaluation set for future research comparisons
- **A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity**
Yejin Bang, Samuel Cahyawijaya, Nayeon Lee, Wenliang Dai, Dan Su, Bryan Wilie, Holy Lovenia, Ziwei Ji, Tiezheng Yu, Willy Chung, **Quyet V. Do**, Yan Xu, Pascale Fung
Preprint, arXiv:2302.04023
 - Extensive empirical evaluation of ChatGPT, focus on three topics: reasoning, hallucination, and interactivity,
 - Discuss the future direction to improve ChatGPT and other LLMs

2022

- **PseudoReasoner: Leveraging Pseudo Labels for Commonsense Knowledge Base Population**
Tianqing Fang, **Quyet V. Do**, Hongming Zhang, Yangqiu Song, Ginny Y. Wong and Simon See
Findings of EMNLP 2022
 - Use the idea of pseudo labels to perform semi-supervised learning on CSKB Population,
 - Propose a filtering strategy for pseudo labels using influence function and self distillation (the student model's own predictions)