ZEKUN ZHAO

Department of Computer Science and Engineering
University of California Santa Cruz
zzhao99@ucsc.edu
zekunzhao.github.io

EDUCATION

UNIVERSITY OF CALIFORNIA, SANTA CRUZ Sep 2021 - Present
Ph.D. in Natural Language Processing
Advisors: Jeffrey Flanigan

UNIVERSITY OF CALIFORNIA, SANTA CRUZ Sep 2018 – Mar 2021
Master of Science in Computer Science

UNIVERSITY OF CALIFORNIA, BERKELEY Dec 2017 – Jun 2018
Exchange Student in the Department of Electrical Engineering and Computer Sciences

NANKAI UNIVERSITY, TIANJIN Sep 2014 – Jun 2018
Bachelor of Engineering in Intelligent Science and Technology

UNDER-REVIEW PAPERS

DuoCAMRA: Duo Copilots for Document Level AMR Annotation (Submitted)
Fast LLM Inference with Parallel Prompting (Submitted)

PROJECT

FAST LLM INFERENCE WITH PARALLEL PROMPTING Jun 2024 - Nov 2024
- Developing a novel inference method for Transformer Large Language Models(LLMs)
- Improving the inference efficiency without compromising generation quality
- Optimizing generation latency and throughput with fast parallel generation
- Reducing inference time on various popular datasets (SQUAD, QuAC, DROP) by over 70%

IMPLICIT ROLE RECOGNITION IN DOCUMENT Sep 2023 - Jun 2024
- Designed a novel prompt method for document-level implicit role recognition
- Constructed the QA Paris with predicate-argument relations extracted from PropBank
- Implemented full-document AMR parsing by incorporating concept coreference and implicit role recognition
- Improving over existing state-of-the-art methods in AMR graph representation

ABSTRACT MEANING REPRESENTATION (AMR) PARSING Sep 2022 – Jun 2023
- Designed a novel method for generating out-of-domain AMR data pairs
- Implemented a data pipeline with Automatic Keyword Extraction, Back Generation, and Pseudo-AMR Parsing
- Developed quality estimation method based on the semantic similarity score
- Fine-turned a language model for boosting the performance of AMR Paring in the out-of-domain scenario

SKILLS

Languages & Tools: Python, C++/C, Kubernetes, Git

EXPERIENCES

- Teaching Assistant for "Natural Language Processing" Fall 2022, Spring 2023, Winter 2024
- Teaching Assistant for "Applied Discrete Mathematics" Fall 2023