Sitao Cheng

(86)18750166790 | sitaotonycheng@foxmail.com

EDUCATION

Nanjing University 2021.09 - 2024.06

Computer Science Master - NLP, LLM, Knowledge Graph - Websoft Lab

- Avg Score 92.35/100 (Top 5% of department)
- Honors: 1st Class Scholarship, 2nd Class Scholarship *2

University of Electronic Science and Technology of China 2017.09 - 2021.06

Software Engineering Bachelor

- GPA **3.99 / 4.00** Avg Score 90.74 (**top 3** of department)
- Honors: 2020 MCM/ICM H Prize, Outstanding Student of Sichuan Province, Outstanding Student of UESTC, MCM/ICM Campus Competition 2nd Prize, WeChat mini-Program Campus Competition 2nd Prize, 1st Class/Enterprise Scholarship*7

RESEARCH EXPERIENCE

Call me when necessary: LLMs can Efficiently and Faithfully Reason over Structured Environments 2023.09 - 2024.02

ACL24 1st Auther Microsof DKI Lab - Research Intern

- Innovation: Propose **Readi**, an efficient and faithful framework to call LLMs. LLMs initially **generate** a reasoning path, which is then **instantiatd** on structured environments. LLMs are called to **edit** the path if the instantiation goes wrong
  - Previous LLM-methods: Step-by-step iterative interaction with the structural environments; Consuming limited information at each step; Error propagation
  - Previous Finetuned-methods: Relying on annotations; Not ensuring faithfulness; Costy for beam search
- Error Message Design: Error reasons; Current instantiation progress; Possible candidate schemas
- Experiments: On three KBQA (Hit@1) and 2 TableQA (EM) tasks, Readi significantly outperforms other LLM-based methods and vanilla LLMs, and are competitive with sota finetuned methods. Analysis shows that Readi's initial reasoning path already surpasses finetuned methods in some extensive aspects, with editing further boosting the performance

QueryAgent: a Reliable and Efficient Reasoning Framework with Environmental Feedback based Self-Correction 2023.09 - 2024.02

ACL24 shared 1st Author Nanjing University with Microsoft DKI Lab

- Innovation: Propose **QueryAgent**, an Agent-based framework to build query by **step-by-step** invoking function tools over KB and conduct self-correction at each step for reliability.
  - Previous in-context learning methods: End-to-End query generation inducing many candidates; Hallucination of LLMs
  - Previous LLM agents: Environmental feedback including only part of entities and relations on KB; Function tools not step-wise executable; Hallucination of LLMs; Error propagation
- Innovation: Propose **ERASER** to detect error for executions, provide tailored guidelines and add directly to observations
  - Previous correction methods: Relying on LLM to identify the error; Mimicing cases in few shot examples to correct
  - Feedback source: KB engine, python interpreter and reasoning memory
  - Guidelines: Error types; Error reasons; Possible solutions. Directly adding to observation to generate new actions
- Experiments: On 4 KBQA(F1) tasks, QueryAgent significantly surpasses other LLM-methods; ERASER substantially boosts another agent-based method; QueryAgent also outperformances other LLM-methods on one TableQA(SP-based) task; QueryAgent imposses a lot fewer running time and less query engine calls.

MarkQA: A large scale KBQA dataset with numerical reasoning 2022.11 - 2023.06

EMNLP23 2nd Author - Nanjing University

- Innovation: Propose **NR-KBQA** to challenge both **multi-hop reasoning** and **numerical reasoning** ability over KB
  - Previous KBQA task: Only consider **complexity** of graph pattern(multi-hop reasoning), not computation **structure**
- Construct **MarkQA** (on Wikidata), scaling automatically to 32k from a small number of seeds
  - Provide both natural language and symbolic language forms of **reasoning steps**
- Design **PyQL** query, which can be converted into SPARQL, as the symbolic reasoning steps, alleviating labeling burden
- Experiments: MarkQA is challenging (especially in zero-shot), and reasoning steps significantly improves the results

Question Decomposition Tree for Answering Complex Questions over Knowledge Bases 2022.02 - 2022.11

AAAI23 2nd Author - Nanjing University

- Innovation: Propose a serializable Question Decomposition Tree **(QDT)** structure to represent natural language questions
  - Previous decomposition methods: Decompose the question insufficiently (just split into 2 parts)
- Innovation: Propose **Clue-Decipher**, a 2-staged method to ease the uncontrollable nature of LM to obrain QDT

- Experiments: Clue-Decipher outperforms other decomposition methods in two types of metrics. And QDT helps two types of QA systems to achieve **sota** results
  - Decomposition experiments: Compare the decomposition results (on QDTrees dataset) in the sequence-based (EM, BLEU, ROUGE) and tree-based (TDA, GED) metrics
  - Seq2Seq QA experiments: With the help of QDT, a T5-base model achieves SOTA on CWQ dataset. Results drop significantly when replacing QDT with other decompositions

PROFESSIONAL EXPERIENCE

Microsoft 2023.10 - Present

Research Intern DKI(data, knowledge and intelligence) Lab Beijing

- Research: Leading a research paper about reasoning over structured environments by LLMs submitted to ACL24
- Research: Cooperate with Nanjing University for a research paper about LLM-agents for KBQA submitted to ACL24
- Research: Cooperate with Nanjing University for a research paper about RAG by LLMs (on progress)
- Engineer: Adopting the idea of Readi (ACL24 paper) to Education and Medicine scenarios

Ant Group (Alipay) - Intelligent Office Assistant Project - QA System (Multi-hop Reasoning Module) 2023.06 - 2023.09

NLP Intern - Digitization Management - Department of AI Application Hangzhou

- Framework: Based on **MCR** (extension of langchain), LLM **iteratively** decomposes and answers the quest*ion*
- Goal: Introduce Knowledge Graph (KG - structured representation) to impove MCR. "**DENOISE**" and "**EXPAND**" text information, **guiding** LLM to decompose and answer the question
- Motivation: MCR relies heavily on LLM (prompt) and the retrieval model, leading to error propagation
- Details: Construct KG "offline and online" according to texts. Retrieve KG subgraphs according to the questions as context
- Results: Question decomposition and multi-hop use cases acc improve significantly. Achieve SOTA on HotpotQA subset.

OTHERS

- **Skills:** Common NLP models (LLM application, Transformer, attention mechanism, etc.), Pytorch, C++, Python
- **Languages:** Good English speaking and listening skills (TOEFL 106, CET-4 CET-6 excellent)
- **Interests:** Body building (over 6x body weight in the Big 3) , Basketball (member of department basketball team)