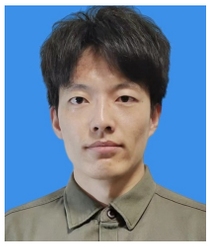
Qiang Gao



24 years old | Male | https://github.com/cooper12121 |
https://scholar.google.com/citations?user=eoUnS60AAAAJ&hl=en&authuser=1
gaoqiang.nlp@gmail.com

Education

Sep 2018 - Jun 2022

East China University of Science and Technology (ECUST) **Energy and Power Engineering (Bachelor)**

Computer Science and Technology(second major).

Sep 2022 - Jun 2025

WuHan University (WHU) **Computer Science and Technology (Master)**

I am affiliated with the Language and Cognition Computing Laboratory at Wuhan University, under the guidance of Professor Fei Li. The laboratory focuses on natural language processing, primarily engaging in research areas such as information extraction, multimodal content recognition, and analysis of large models. My current work mainly involves research in information extraction, fine-tuning of large models, and multimodal content recognition.

Publications

- Enhancing Cross-Document Event Coreference Resolution by Discourse Structure and Semantic Information COLING2024, First author

Existing cross-document event coreference resolution models either compute the similarity between event mentions directly or enhance mention representation by extracting event arguments, lacking the capability to utilize document-level information. This makes it challenging to capture long-distance dependencies, leading to poor performance in coreference decisions for highly similar events, events with different expressions but identical meanings, or events whose argument information depends on distant relations. For the first time, we propose to enhance document-level representation by using discourse information. By constructing a document-level Rhetorical Structure Theory (RST) tree and cross-document lexical chains, we model the structural and semantic information of documents, respectively, and build graphs for each. A Graph Attention Network (GAT) is used to learn from these structural and semantic graphs, and the results from the Encoder are fused with those from the GAT for coreference resolution and clustering of coreferent events. Additionally, we have built a large-scale, event-type agnostic Chinese cross-document event coreference resolution dataset. Experimental results demonstrate significant improvements achieved by our model, with further analyses indicating that these benefits are derived from the rich discourse relations captured by our RST and lexical chains. Our method offers a new approach to discourse-level tasks, efficiently enhancing information representation capabilities for various discourse tasks.

- Harvesting Events from Multiple Sources: Towards a Cross-Document Event Extraction Paradigm ACL2024 (Finding), First author

In this paper, we introduce a pioneering approach to cross-document event extraction (CDEE), significantly enhancing the extraction and integration of event information from multiple documents. We address the limitations of traditional document-level event extraction by utilizing a novel dataset and a multifaceted pipeline. The dataset, CLES, comprises over 37,688 mention-level events and 3,633 concept-level events, enabling comprehensive studies across different document sources. Our proposed CDEE pipeline consists of event extraction, coreference resolution, entity normalization, role normalization, and entity-role resolution, achieving approximately 72% F1 score in end-to-end performance.

- MMLSCU: A Dataset for Multimodal Multi-domain Live Streaming Comment Understanding WWW2024 (Oral), Second author

Interactive audience participation in live-streaming scenarios provides constructive feedback for both streamers and platforms. Analyzing these live comments to uncover underlying intentions is crucial for enhancing the quality of broadcasts and promoting the healthy development of the live-streaming ecosystem. We have introduced a multimodal dataset specific to the live streaming domain, which includes video, audio, and comment text from live sessions. We propose four tasks: audience comment intent detection, intent cause mining, audience comment explanation, and streamer strategy recommendation. Utilizing Chain of Thought (CoT) technology, we have developed an end-to-end multimodal model to tackle these tasks. Our model includes an Audio Branch and a Visual Branch, processing auditory and visual information through visual and audio encoders, respectively. These are then integrated with textual content, and a chain of reasoning pathways is designed. The tasks are sequentially inferred by inputting them into a Large Language Model (LLM) for reasoning.

Internship experience

Feb 2024 - Present **Tencent AI Lab**

since February 2024, I have been engaged as a research intern at Tencent AI Lab, supervised by researcher Jian Li, where I have delved into large language models, including multimodal variations. My work has primarily revolved around enhancing the performance of Mixture of Experts (MoE) models on business data. I have contributed to initiatives aimed at improving the training efficiency and stability of warm-starting MoE models and have conducted detailed analyses of expert distribution strategies.

- **Project 1 - Warm-starting MoE**

Here, we explore the performance of warm-started MoE models on business data. We constructed warm-started Yi-8x6b and Yi-4x6b models based on the Yi-6b model, including both base and Instruct versions (by copying the MLP parameters as experts, and creating randomly initialized routing and load balancing losses). Our goals are:

1. For the Instruct version, fine-tune the constructed MoE model to leverage the strong performance of the original model and achieve better performance on gaming business data with minimal data,
2. For the base version, perform post-pretraining to evaluate the routing distribution strategy and training stability.
3. To determine under what conditions stable routing can be trained.

Our approach:

1. For the Instruct version, fine-tune the MoE model with context-rich gaming business data and context-poor advertising business data.
2. For the base version: post-pretrain and sft are performed using a mixed Chinese-English corpus, utilizing wudao and firefly data.

Experimental results:

1. We found that the Instruct version of the MoE model, with less data, could perform better than the original 6b model.
2. In the initial phase of the base version post-pretrain, the 8x6b routing distribution was extremely unstable, while the 4x6b showed better stability. Tests on general metrics showed a significant drop in performance compared to the original 6b model before 0.3 epochs, followed by gradual improvement. During the sft phase, the MoE model needed more data to understand the instructions, as evidenced by lengthy and repetitive text responses before 0.02 epochs. However, as training progressed, the performance of the MoE gradually improved, while the base model experienced a drastic decline at 0.5 epochs, with severe forgetting of general knowledge.

Our challenges include:

1. How to determine when the routing distribution strategy is sufficiently trained.
2. Whether the parameter advantages of the 8x6b can compensate for its unstable routing distribution: simple sft routing showed no clear pattern, whereas the 4x6b routing distribution exhibited a distinct pattern.
3. Whether the 8x6b is necessary, and if the increased number of expert parameters can significantly enhance performance. These are subjects for ongoing exploration.

- **Project 2 - More anthropomorphic NPCs**

Our goal is to make the responses of NPCs in games more anthropomorphic. This means that NPCs should not only answer players' questions but also be able to resonate emotionally with players. This is primarily manifested in the following ways:

1. NPCs should be able to recall content shared by players. This requires NPCs to proactively mention information previously brought up by players, such as if a player says, "I've been wanting to eat hotpot recently," the NPC's response should not merely mention hotpot, but could proactively mention, "Didn’t you say you have a sensitive stomach? Eating hotpot might not be healthy for you," making the player feel like the NPC cares and listens like a human.
2. NPCs should provide emotional support to players.
3. NPCs should have the ability to refuse answers outside of their knowledge domain.

Approach:

1. Memory capability: Construct an event summarization model to summarize historical conversations between players and NPCs and extract eight types of events.
2. Dialogue response model: Based on the NPC's persona and historical conversations, use the RAG module to retrieve relevant events, integrate prompts using GPT-4 to construct dialogue data that utilizes event information for more anthropomorphic responses, and train using the constructed data. This data construction includes four steps to ensure the data meets the requirements.

Currently, the main focus is on constructing data and combining RAG to achieve the interestingness and personification of NPCs. This requires high-quality data as a guarantee. We are exploring how to integrate the personification of multiple NPC tasks (different NPCs have different personification directions).

Research interest

Over the past two years, the rise of large models has spawned various new research fields. As a master’s student, I have been actively exploring new areas and knowledge. My research over the past two years has primarily covered traditional natural language processing tasks, Mixture of Experts (MoE) models, and multimodal large models. Recently, I have started to focus on the field of model integration and plan to continue my research on multimodal LLM-MoE models in the future. I hope to further expand my knowledge base in the final year of my master’s program, deeply exploring various aspects of LLMs to enrich my options for PhD research.

Looking ahead to my PhD, I hope to develop tools that facilitate the LLM community and explore more interesting AI applications。

The current directions in LLM research can broadly be classified into “useful AI” and “fun AI.” This blog post, “Should AI Agents Be More Useful or More Interesting?” (The content is in Chinese and requires translation for browsing, sorry), has greatly inspired me, inclining me to focus on the latter during my PhD. Since AI applications often require the integration of multiple modalities, mastering multimodal LLM knowledge during my PhD is crucial. Moreover, to sustain the socio-economic value of LLMs, application is key. Therefore, as a PhD student, I should consider the practical applications of LLMs more thoroughly. This is a brief overview of my future research directions.

Important Aspect: Large Model Safety. As large models become increasingly powerful, their safety will become the foremost consideration in any AI development. Therefore, ensuring the safety of any type of large language model (LLM) is also a very important aspect.

Self-evaluation

- Throughout my academic journey, I have maintained a profound interest in research and have always been eager to explore new technologies and directions. This passion has driven me to venture into uncharted territories

- I approach my research with persistence and seriousness, and I possess a high enthusiasm for coding and experimentation. This attitude not only aids in my progress but also allows me to remain resolute and focused in the face of challenges.
- I have strong programming skills and have open-sourced several projects on GitHub.
- I have extensive experience with LLM code and am familiar with various common frameworks currently in use.
- I love academia and maintain an optimistic life attitude, skilled at balancing work and life.