Research Statement

Tianqing Fang

Knowledge, including but not limited to skills, factual and commonsense knowledge, is fundamental to human cognition and of vital importance in developing human-level language models. Despite being knowledgeable on many aspects by scaling up to billions of parameters, Large Language Models (LLMs) are still criticized for *hallucination* and a limited capacity of *reasoning* with knowledge. To this end, there is an urgent need for developing both knowledge-enhanced LLMs and robust reasoning paradigms. My research vision is thus centered on building **robust, trustworthy, and knowledge-grounded Natural Language Processing (NLP) systems beyond scaling up**.

My current research direction can be roughly divided into two scopes:

1. **(Complex) Knowledge Acquisition**: Acquisition of commonsense knowledge, knowledge conflicts, and complex logical queries with information extraction, crowdsourcing, and LLMs.

2. **Reasoning with Knowledge**: (Lightweight) injection of knowledge, including constrained decoding, retrieval-augmented, and information-theoretic injections. Elicit the complex reasoning ability of LLMs using internal and external knowledge, particularly on complex structured, counterfactual, long-tail, and long-context knowledge.

The acquisition of knowledge in different forms serves as a foundational basis for language models. On top of the curated knowledge resources, my research also focuses on developing efficient and effective knowledge-grounded reasoning systems.

1 Complex Knowledge Acquisition

I have studied the acquisition of commonsense knowledge through the form of relational knowledge bases, abstraction, and complex logical queries. I also study knowledge conflicts, where the actual context contradicts the parametric knowledge in the language models.

Commonsense. I mainly studied Commonsense Knowledge mining and am the creator of the paradigm **Commonsense Knowledge Base Population** [1, 2, 3], which aims at automatically populating commonsense knowledge defined in a certain knowledge base via relation extraction and a commonsense discriminator. Based on the commonsense paradigm of inferential knowledge on events and situations (e.g., *PersonX repels PersonY's attack*, then PersonX is seen as *brave*, defined in ATOMIC and GLUCOSE), and entities (e.g., *eat* results in *full*, defined in ConceptNet), I used graph-enhanced BERT-based models to automatically convert information-extracted discourse relations to commonsense knowledge [1]. On top of this, I built several follow-up works including benchmarking such a commonsense knowledge base population process [2], using semi-supervised learning for knowledge acquisition [4], and constrained prompt-based reasoning for Large Language Models [5].

Abstraction. Abstraction indicates the acquisition of a higher level of knowledge (e.g., conceptualizing *watching football games* to *relaxing activity*). We first crowdsourced abstraction knowledge on a proportion of triples in ATOMIC [6]. Since it's hard to acquire conceptualization knowledge manually due to the inherent difficulty of such *IsA* relations, which require deep linguistic knowledge, I further built a semi-supervised abstraction and instantiation acquisition system [7] to scalably acquire million-scale abstraction-related knowledge by populating the small-scale annotated knowledge on large commonsense knowledge bases. To further improve the annotation quality and embrace the power of LLMs, I also contribute to conduct knowledge distillation from LLMs (ChatGPT and Llama2) [8] in a dual-process of both abstraction and instantiation, thus acquiring even more abstraction instances with better quality. By synthesizing the abstract knowledge into question-answering pairs, we trained a zero-shot commonsense question answering model, showing state-of-the-art zero-shot reasoning performance, even better than ChatGPT [9].

Complex Reasoning. However, despite being able to understand those one-hop inferences, LLMs still struggle to reason about complicated structures, such as logical queries on knowledge graphs [10]. I sampled complex first-order logic queries from ATOMIC and verbalized them to narratives to derive both a harder commonsense evaluation set and better reasoning supervision signals for LLMs. Experiments show that complex queries equip language models with better reasoning ability on both complex reasoning and original one-hop reasoning tasks.

Knowledge Conflict. Knowledge conflicts refer to the cases where the parametric knowledge from the language model contradicts with the actual context. I used the idea of *reporting bias* to calculate knowledge conflict statistics to mine temporal knowledge conflicts of various types [11]. I developed counterfactual data augmentation that can be used for both fine-tuning and in-context learning to mitigate such knowledge conflicts.

2 Knowledge Injection and Reasoning

I study injecting knowledge into LLMs without scaling them up from three perspectives. First, lightweight **Knowledge Constrained Decoding**. In the framework KCTS [12], a Monte-Carlo Tree Search module is applied to estimate the future groundness to the intended knowledge, and a novel token-level hallucination detection method is used by constructing synthetic supervision signals by setting a hallucination inflection point at a token level. KCTS is plug-and-play to LLMs and only requires fine-tuning on 0.21% of additional training weights while significantly improving factuality. Second, I study knowledge fusion in a **Data Augmentation** perspective to denoise the knowledge [13] with training dynamics, i.e., a clean distiller and a self-regularization module, and to use counterfactual data augmentation for both fine-tuning and in-context learning to mitigate knowledge conflicts [11]. Last but not least, I used **Graph Reasoning Networks** to fuse knowledge subgraphs to language models. I studied integrating supporting sub-graphs of knowledge to reasoning modules, including using GraphSAGE to aggregate the knowledge in ATOMIC to perform commonsense knowledge base population [2] and fusing embeddings of knowledge graphs to each layer of transformers to perform better dialogue generation [14].

3 Future Works

In the future, I will pursue my lifelong research goal to enable machines with the ability of human cognition and reasoning that leads to AGI, especially in the following directions:

Grounded Complex Reasoning. Though current LLMs possess knowledge about numerous one-hop scenarios, they are not robust in handling complex graph-structured reasoning tasks, even with the chain-of-thought. In my future research, I aim to identify and integrate complex knowledge and reasoning capabilities into language models, enabling them to effectively address complex reasoning challenges prevalent in real-world applications, such as planning. I will start with constructing knowledge resources and formulate them as the format of logical queries, where variables can be entities, propositions, and even formulas. LLMs then synthesize and compose knowledge into relevant reasoning contexts, generating synthetic training data. This data is utilized to train or self-reward LLMs, facilitating their self-evolution. The grounded synthetic data can train a retriever that connects real-world context with intricate knowledge, enabling a bidirectional process of reasoning and grounding. This advancement further promotes trustworthy NLP research.

Effective and Efficient Knowledge Injection. I aim to focus on two main research problems regarding knowledge injection and reasoning. First, inject knowledge into LLMs without catastrophic forgetting and high-cost fine-tuning. To accomplish this, I will focus on inference-time algorithms, including guided decoding [12], retrieval-augmentation, and information-theoretic integration of in-context examples. Second, reason with entailment knowledge inspired human cognition (*k-line* theory by Marvin Minsky). Certain long-tail scenarios, particularly in commonsense reasoning, can be abstracted/entailed to a higher level for reasoning. For instance, if an LLM lacks knowledge about *ghijak*, conceptualizing it as an *instrument* enables effective resolution of reasoning scenarios related to instruments without requiring scaling up or fine-tuning the LLM.

Real-world Application of Knowledge Reasoning. Most existing (commonsense) reasoning benchmarks are limited to toy datasets designed to challenge language models rather than addressing practical tasks. To bridge this gap, I aim to integrate various forms of knowledge, including scientific knowledge, moral and cultural knowledge, and long-context knowledge from textbooks, into language models as curriculums, guiding their utilization in real-world applications involving interactions with human agents, procedural planning, and science discovery. My objective is to establish a comprehensive pipeline that encompasses essential information extraction or direct knowledge generation from (long) contexts. This pipeline involves grounding to intricate knowledge and formulas in logical forms, employing tools to ensure deterministic computation, and employing LLM reasoning based on intermediate computation and planning outcomes. Data synthesis, as explained in the ”Grounded Complex Reasoning” paragraph, is a crucial step in providing supervision signals, which harnesses the existing ability of LLMs to enhance weak-to-strong generalization.

References

[1] T. Fang, H. Zhang, W. Wang, Y. Song, and B. He, "DISCOS: bridging the gap between discourse knowledge and commonsense knowledge," in *WWW '21: The Web Conference*, 2021.

[2] T. Fang, W. Wang, S. Choi, S. Hao, H. Zhang, Y. Song, and B. He, "Benchmarking commonsense knowledge base population with an effective evaluation dataset," in *EMNLP*, 2021.

[3] T. Fang, Q. V. Do, S. Choi, W. Wang, and Y. Song, "Ckbp v2: An expert-annotated evaluation set for commonsense knowledge base population," *arXiv*, 2023.

[4] T. Fang, Q. V. Do, H. Zhang, Y. Song, G. Y. Wong, and S. See, "Pseudoreasoner: Leveraging pseudo labels for commonsense knowledge base population," *Findings of EMNLP*, 2022.

[5] Q. V. Do, T. Fang, S. Diao, Z. Wang, and Y. Song, "Constraintchecker: A plugin for large language models to reason on commonsense knowledge bases," *EACL*, 2024.

[6] M. He, T. Fang, W. Wang, and Y. Song, "Acquiring and modelling abstract commonsense knowledge via conceptualization," 2022.

[7] W. Wang*, T. Fang*, B. Xu, C. Y. L. Bo, Y. Song, and L. Chen, "CAT: A contextualized conceptualization and instantiation framework for commonsense reasoning," *ACL*, 2023.

[8] W. Wang, T. Fang, C. Li, H. Shi, W. Ding, B. Xu, Z. Wang, J. Bai, X. Liu, J. Cheng, *et al.*, "Candle: Iterative conceptualization and instantiation distillation from large language models for commonsense reasoning," *arxiv*, 2024.

[9] W. Wang*, T. Fang*, ..., Y. Song, and A. Bosselut, "CAR: conceptualization-augmented reasoner for zero-shot commonsense question answering," *Findings of EMNLP*, 2023.

[10] T. Fang and A. Bosselut, "Complex commonsense reasoning on commonsense knowledge graphs," *ongoing work (to be submitteed to ACL 2024)*, 2024.

[11] T. Fang, Z. Wang, H. Zhang, Y. Song, and M. Chen, "Getting sick after seeing a doctor? diagnosing and mitigating knowledge conflicts in event temporal reasoning," *arxiv*, 2023.

[12] S. Choi, T. Fang, Z. Wang, and Y. Song, "KCTS: knowledge-constrained tree search decoding with token-level hallucination detection," *EMNLP*, 2023.

[13] T. Fang, W. Zhou, H. Zhang, F. Liu, M. Chen, and Y. Song, "On-the-fly denosing for data augmentation in natural language understanding," *Fingdings of EACL*, 2024.

[14] T. Fang, H. Pan, H. Zhang, Y. Song, K. Xu, and D. Yu, "Do boat and ocean suggest beach? dialogue summarization with external knowledge," in *AKBC*, 2021.

* indicates equal contribution.