Tianqing Fang

Email : tfangaa@cse.ust.hk

http://www.fangtq.com

Research Interest

- **Knowledge Acquisition:** Acquisition of commonsense knowledge, knowledge conflicts, and complex knowledge with information extraction, crowdsourcing, and LLM distillation.
- **Knowledge Injection and Reasoning on LLMs** (Lightweight) injection of knowledge, including constrained decoding, retrieval-augmented, and information-theoretic injections. Elicit complex reasoning ability of LLMs using internal/external knowledge, particularly on complex, counterfactual, long-tail, and long-context knowledge.
- **Machine Learning for NLP:** Data denoising; training dynamics; ...

Education

- **Hong Kong University of Science and Technology (QS 2024 world rank 60)** — Hong Kong SAR
  Ph.D. of Computer Science; Supervisor: Yangqiu Song — *Aug. 2019 – June 2024*
- **Zhejiang University (QS 2024 world rank 44, top 4 in China)** — Hangzhou, China
  B.E. of Automation; GPA: 3.94/4.00, top 5%; Supervisor: Yang Yang — *Aug. 2015 – June 2019*
  Minor Advanced Class of Engineering Education in Chu Ko Chen Honors College
  Special Scholarship Awardee, 2018 (top awards for undergraduate students, one per year)

Experience

- **École Polytechnique Fédérale de Lausanne (EPFL)** — Lausanne, Switzerland
  Visiting Doctoral Student; Supervisor: Antoine Bosselut — *July 2023 - Present*
  Topic: Complex Commonsense Reasoning
- **University of Southern California** — Los Angeles, CA
  Visiting Research Scholar; Supervisor: Muhao Chen — *July 2022 - June 2023*
  Topic: Event-level knowledge conflicts in LLMs; Denoising Data Augmentation
- **NVIDIA** — Hong Kong SAR
  Research Intern; Supervisor: Ginny Y. Wong and Simon See — *Feb. 2022 - June 2022*
  Topic: Semi-supervised learning on commonsense reasoning

Publications

Preprints:

[1] **On-the-fly Denoising for Data Augmentation in Natural Language Understanding**
***Tianqing Fang**, Wenxuan Zhou, Fangyu Liu, Hongming Zhang, Yangqiu Song, and Muhao Chen*
arxiv.2212.10558, 2023

- A self-regularization + distillation enabled denoising framework for data augmentation in NLU.

[2] **Getting Sick After Seeing a Doctor? Diagnosing and Mitigating Knowledge Conflicts in Event Temporal Reasoning**
***Tianqing Fang**, Zhaowei Wang, Wenxuan Zhou, Hongming Zhang, Yangqiu Song, Muhao Chen.*
arxiv.2212.10558, 2023

- Define and study knowledge conflicts in event temporal reasoning, and propose data-augmentation driven mitigation methods for both Pre-trained Language Models (PLMs) and Large Language Models (LLMs).

[3] **Acquiring and Modelling Abstract Commonsense Knowledge via Conceptualization**
*Mutian He, **Tianqing Fang**, Weiqi Wang, and Yangqiu Song.*
Submitted to Journal of Artificial Intelligence. arxiv.2206.01532, 2022

- First work to study the role of conceptualization in event-based commonsense knowledge graphs. Identify the problem of "proper level of conceptualization" for commonsense understanding.

- Construct a contextualized and conceptualized ATOMIC knowledge graph with WordNet, Probase, and human annotation. For example, "(PersonX drinks milk, `xAttr`, thirsty)" can be conceptualized as "(PersonX drinks beverage, `xAttr`, thirsty)". But in "(PersonX drinks milk, `xWant`, to grow taller)", the head event cannot be conceptualized as "PersonX drinks beverage" but should be "PersonX drinks diary-product" in view of the context that PersonX wants to grow taller.

Conference:

[4] **KCTS: Knowledge-Constrained Tree Search Decoding with Token-Level Hallucination Detection**

- *Sehyun Choi, **Tianqing Fang**, Zhaowei Wang, Yangqiu Song*
- EMNLP 2023
  - An inference-time algorithm based on Monte-Carlo Tree Search for future step estimation and a token-level hallucination detector, applicable for general LLMs.

[5] **CAR: Conceptualization-Augmented Reasoner for Zero-Shot Commonsense Question Answering**

- *Weiqi Wang*, **Tianqing Fang***, Wenxuan Ding, Baixuan Xu, Xin Liu, Yangqiu Song, Antoine Bosselut*
- Findings of EMNLP 2023
  - Use one-hop conceptualization to 1) augment ATOMIC, a commonsense knowledge graph. 2) to guide zero-shot question-answering pair generation to reduce false-negative options.
  - Zero-shot models outperform ChatGPT in several commonsense question answering benchmarks and achieve state-of-the-art.

[6] **QADYNAMICS: Training Dynamics-Driven Synthetic QA Diagnostic for Zero-Shot Commonsense Question Answering**

- *Haochen SHI, Weiqi Wang, **Tianqing Fang**, Baixuan Xu, Wenxuan Ding, Xin Liu, Yangqiu Song*
- Findings of EMNLP 2023
  - A novel option-level training dynamics algorithms for selecting informative negative options. Achieving SOTA on zero-shot commonsense QA on five datasets.

[7] **Doolittle: Benchmarks and Corpora for Academic Writing Formalization**

- *Shizhe Diao, Yongyu Lei, Liangming Pan, **Tianqing Fang**, Wangchunshu Zhou, Sedrick Scott Keh, Min-Yen Kan, Tong Zhang*
- EMNLP 2023
  - A new benchmark for paragraph-level academic writing formalization. A metric-based RL method for improve LLMs' writing refinement ability.

[8] **STORYANALOGY: Deriving Story-level Analogies from Large Language Models to Unlock Analogical Understanding**

- *Jiayang Cheng, Lin Qiu, Tsz Ho CHAN, **Tianqing Fang**, Weiqi Wang, Chunkit Chan, Qipeng Guo, Hongming Zhang, Yangqiu Song, Yue Zhang, Zheng Zhang*
- EMNLP 2023
  - A new benchmark for story-level analogy reasoning.

[9] **CAT: A Contextualized Conceptualization and Instantiation Framework for Commonsense Reasoning**

- *Weiqi Wang*, **Tianqing Fang***, Baixuan Xu, Chun Yi Louis Bo, Yangqiu Song, Lei Chen.*
- ACL 2023.
  - Semi-supervised conceptualization-instantiation framework for commonsense knowledge bases.

[10] **COLA: Contextualized Commonsense Causality Reasoning from the Causal Inference Perspective**

- *Zhaowei Wang, Quyet V. Do, Hongming Zhang, Jiayao Zhang, Weiqi Wang, **Tianqing Fang**, Yangqiu Song, Ginny Y. Wong, Simon See*
- ACL 2023.

[11] **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, achieving state-of-the-art.
  - Propose a filtering strategy for pseudo labels using influence function and self distillation (the student model's own predictions).

[12] **MICO: A Multi-alternative Contrastive Learning Framework for Commonsense Knowledge Representation**
- *Ying Su, Zihao Wang,* ***Tianqing Fang****, Hongming Zhang, Yangqiu Song and Tong Zhang*
- Findings of EMNLP 2022.
  - A novel commonsense knowledge embedding pipeline, well used for CSKB completion and zero-shot CSQAs.

[13] **SubeventWriter: Iterative Sub-event Sequence Generation with Coherence Controller**
- *Zhaowei Wang, Hongming Zhang,* ***Tianqing Fang****, Yangqiu Song, Ginny Y. Wong and Simon See*
- EMNLP 2022.
  - An iterative neural text generation framework to generate multi-step instructions.

[14] **Weakly Supervised Text Classification using Supervision Signals from a Language Model**
- *Ziqian Zeng, Weimin Ni,* ***Tianqing Fang****, Xiang Li, Xinran Zhao, and Yangqiu Song.*
- Findings of Annual Conference of the North American Chapter of the Association for Computational Linguistics (Findings of NAACL). 2022.

[15] **Benchmarking Commonsense Knowledge Base Population with an Effective Evaluation Dataset**
- ***Tianqing Fang****, Weiqi Wang*, Sehyun Choi, Shibo Hao, Hongming Zhang, Yangqiu Song, Bin He*
- Conference on Empirical Methods in Natural Language Processing (**EMNLP**), 2021 (Main Conference).
  - Commonsense Knowledge Base (CSKB) Population is different from Completion as it requires reasoning over unseen assertions in external resources, while Completion only fills missing links within the CSKB.
  - Propose a dataset aligning four popular CSKBs, ConceptNet, ATOMIC, ATOMIC$^{20}_{20}$, and GLUCOSE with a large-scale eventuality graph, ASER, to populate commonsense knowledge. ˜31K triples are annotated as the evaluation set to check neural models' reasoning ability.
  - Developed KG-BertSAGE to better incorporate graph structures in the commonsense reasoning task.

[16] **DISCOS: Bridging the Gap between Discourse Knowledge and Commonsense Knowledge**
- ***Tianqing Fang****, Hongming Zhang, Weiqi Wang, Yangqiu Song, and Bin He.*
- The Web Conference (**WWW**), 2021.
  - Align the Commonsense Knowledge Base ATOMIC with a large-scale eventuality graph ASER. Use the knowledge in ATOMIC as ground-truth to train a reasoning model. Populate ATOMIC with novel edges in ASER .
  - Such commonsense knowledge acquisition method can alleviate selection bias and produce more diverse commonsense knowledge.

[17] **Do Boat and Ocean Suggest Beach? Dialogue Summarization with External Knowledge**
- ***Tianqing Fang****, Haojie Pan, Hongming Zhang, Yangqiu Song, Kun Xu, Dong Yu.*
- Conference on Automated Knowledge Base Construction (**AKBC**). 2021.
  - Address the situation where summarization may include something out of the dialogue context but can be implicitly inferred. Develop a knowledge-attention network to tackle this problem and achieves promising results.

[18] **Probing Toxic Content in Large Pre-Trained Language Models**
Nedjma Ousidhoum, Xinran Zhao, ***Tianqing Fang****, Yangqiu Song, and Dit-Yan Yeung*
Annual Meeting of the Association for Computational Linguistics (**ACL**). 2021.

Journal:

[19] **ASER: Towards Large-scale Commonsense Knowledge Acquisition via Higher-order Selectional Preference over Eventualities**
Hongming Zhang, Xin Liu*, Haojie Pan*, Haowen Ke, Jiefu Ou,* ***Tianqing Fang****, and Yangqiu Song.*
Artificial Intelligence. 2022

Academic Achievements

- HKUST Overseas Research Award (2023)
- HKUST RedBird Academic Excellence Award for Continuing PhD Students in 2022/23 (2023)
- HKUST RedBird Academic Excellence Award for Continuing PhD Students in 2021/22 (2022)
- Hong Kong Ph.D. Fellowship (2019-2023, top 100 Ph.D. students in all disciplines in Hong Kong, 41.5K USD per year)
- Special Scholarship for Undergraduate Students in Zhejiang University (One of the most prestigious awards for undergraduates in Zhejiang University) (2018)
- Provincial Scholarship (top 5% in 2017-2018, highest annual scholarship in Zhejiang Province) (2018)
- 1st Place and MATLAB Innovation Award (1st/36k+) in Contemporary Undergraduate Mathematical Contest in Modeling (The most authoritative mathematical modeling competition in China) (2017)
- Provincial Scholarship (top 5% in 2016-2017, highest annual scholarship in Zhejiang Province) (2017)
- National Scholarship (top 3% in 2015-2016, highest annual scholarship in China) (2016)

Skills

- **Programming skills**: C++, Python
- **Languages**: English (TOEFL 110, 26 in speaking), Mandarin Chinese (Native), Cantonese (Elementary).
- **Miscs**: I enjoy taking pictures. Some highlights are at https://unsplash.com/@tqfang.

Teaching

- MSBD 5018: Natural Language Processing (Spring 2022)
- COMP5222/MATH5471: Statistical Learning Models for Text and Graph Data. (Fall 2021)
- MSBD6000H: Natural Language Processing. (Spring 2021)
- COMP4901K/MATH4824B: Machine Learning for Natural Language Processing. (Fall 2020)
- COMP4332/RMBI4310: Big Data Mining. (Spring 2020, awarded best TA award in the CSE department.)

Mentoring

- **Haochen Shi**: HKUST UG. QADynamics (Findings of EMNLP 2023).
- **Weiqi Wang**: HKUST UG → HKUST Ph.D., *Hong Kong Ph.D. Fellowship* Awardee. Worked on Commonsense Knowledge Acquisition and Reasoning (WWW 2021, EMNLP 2021, ACL 2023, Findings of EMNLP 2023).
- **Van Quyet Do**: HKUST UG → HKUST M.Phil. Worked on semi-supervised learning for commonsense knowledge base population (EMNLP 2022 and ongoing works).
- **Sehyun Choi**: HKUST UG → HKUST M.Phil., *Asian Future Leaders Scholarship Program* Awardee. Worked on graph neural networks for commonsense reasoning and knowledge-constrained decoding (EMNLP 2021 and EMNLP 2023).
- **Shibo Hao**: Peking University UG → UCSD Ph.D., worked on commonsense acquisition (EMNLP 2021).

Reference

- **Dr. Yangqiu Song**:
  Associate Professor at HKUST. Ph.D. Supervisor. Email: yqsong [at] cse [dot] ust [dot] hk
- **Dr. Muhao Chen**:
  Assistant Professor at UC Davis. Supervisor of my visit in USC. Email: muhaoche [at] usc [dot] edu
- **Dr. Antoine Bosselut**:
  Assistant Professor at EPFL. Supervisor of my visit in EPFL. Email: antoine [dot] bosselut [at] epfl [dot] ch