Xiao Liu

No. 5 South Zhongguancun Street, Haidian District, Beijing, China, 100081
+86 18610789560 • xiaoliubit@hotmail.com
https://xiaoliubit.github.io • *Google Scholar* • *GitHub*

RESEARCH INTERESTS

Information Extraction; Text Mining; Event Extraction & Schema Induction; Knowledge-boosted Application; Application of NLP & Machine Learning

EDUCATION

Beijing Institute of Technology, Beijing, China

- **Ph.D. Candidate**, Computer Science — Sep 2017 –
  - Adviser: *Prof. Heyan Huang*
  - Topic: *Event extraction on low-resources domains, including automatically inducting event schemas using unsupervised approachs and jointly extracting multiple events in texts*
- **M.S.**, Computer Science (Rank 5/192) — Sep 2016 – Jun 2017
  - Adviser: *Prof. Heyan Huang*
  - *Successive postgraduate and doctoral program*
- **B.S.**, Computer Science (Rank 4/220) — Sep 2012 – Jun 2016
  - ***Outstanding Graduates at BIT and in Beijing***

Singapore University of Technology and Design, Singapore

- **Joint Ph.D. student**, Information Science Technology and Design — Aug 2018 – Mar 2019
  - Host: *Prof. Yue Zhang* and *Prof. Jun Sun*
  - Topic: *Information extraction using neural latent variable models and variational inferences*

Nanyang Technological University, Singapore

- **Joint Ph.D. student**, School of Computer Science and Engineer — Mar 2019 – Sep 2019
  - Host: *Prof. Erik Cambria* and *Prof. Yue Zhang*
  - Topic: *Event applications, knowledge-based sentiment analysis and stock market prediction*

EXPERIENCE

Microsoft Research Asia, Beijing, China

- **Research Intern**, Data Mining and Enterprise Intelligence Group — Dec 2016 – Nov 2017
  - Worked with Dr Jun Yan and Yaobo Liang on semantic similarity computation and generation.
  - Participated in the FAQ Retrieval project for Pfizer that provides every input question with an answer by retrieving similar standard questions with well-written answers.
  - Built a high-recall filter for top-n similar question-question pairs.
  - Built pair-wise CNNs to classify the question intents for boosting question matching.
  - Built paraphrase generation models by synonyms replacement with language models and beam search.
  - Built a deep learning paraphrase generation model based on an NMT seq2seq framework.

Netease Youdao, Beijing, China

- **Research and Development Intern**, Machine Translation Group — Nov 2015 – Dec 2016

- Improved the result of the Japanese-to-Chinese machine translation service by 4% point increment in BLEU.
- Cleaned corpus from different sources, optimized word segmentation and ordering strategies, enlarged bilingual and monolingual corpus, trained larger language models, and deployed them into the online production environment.
- Designed the framework and built the first version of parallel sentence pairs crawler with switchable alignment methods on Spark clusters.

PROFESSIONAL SERVICES

Program Committee Member / Reviewer

- **Conferences:** *ACL 2021, ACL 2020, EMNLP 2021, EMNLP 2020, EACL 2021, NAACL 2021, AAAI 2021, NLPCC 2021*
- **Journals:** *TASLP, TKDE, TALLIP, PLoS One and JCCE*

Secondary Reviewer

- **Conferences:** *ACL 2019, ACL 2018, EMNLP 2019, EMNLP 2018, COLING 2018, CCL 2017, NLPCC 2016*

PUBLICATION

[1] **Xiao Liu**, Heyan Huang, Yue Zhang. Open Domain Event Extraction Using Neural Latent Variable Models. *In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (**ACL**), 2019. (**Long Oral**)

[2] **Xiao Liu**, Zhunchen Luo, Heyan Huang. Jointly Multiple Events Extraction via Attention-based Graph Information Aggregation. *In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (**EMNLP**), 2018. (**Long Oral**)

[3] **Xiao Liu**, Heyan Huang, Yue Zhang. End-to-End Event Factuality Prediction Using Directional Labeled Graph Recurrent Network. *Information Processing & Management* (**IPM**), 2022, 59(2): 102836.

[4] Heyan Huang, **Xiao Liu**, Yue Zhang, Chong Feng. News-Driven Stock Prediction via Noisy Equity StateRepresentation. *Neurocomputing* (**NEUCOM**), 2021, 470: 66-75.

[5] Heyan Huang, **Xiao Liu**, Qian Liu. Knowledge-Enhanced Graph Encoding Method for Metaphor Detection in Text. 基于知识增强的文本隐喻识别图编码方法. *Journal of Computer Research and Development,* 计算机研究与发展, 2022. (**in Chinese, to appear**)

[6] Heyan Huang, **Xiao Liu**. A Survey on Event Extraction in New Domains. 面向新领域的事件抽取研究综述. *CAAI Transactions on Intelligent Systems*, 智能系统学报, 2022, 17(1): 201-212. (**in Chinese**)

[7] **Xiao Liu**, Ge Shi, Bo Wang, Changsen Yuan, Heyan Huang, Chong Feng, Lifang Wu. BIT-Event at NLPCC-2021 Task 3: Subevent Identification via Adversarial Training. *In Proceedings of the 10th CCF International Conference on Natural Language Processing and Chinese Computing* (**NLPCC**), 2021. (**Rank 1/15**)

[8] Heyan Huang, **Xiao Liu**, Ge Shi. Information Extraction and Applications. 文本信息抽取及应用研究（专著）. *Publishing House of Electronics Industry,* 电子工业出版社, 2022. (**in Chinese, before publishing**)

[9] Heyan Huang, **Xiao Liu**. An Event Factuality Prediction Method Based on Graph Recurrent Neural Network (Patent). 一种基于图循环神经网络的事件事实性检测方法（专利）. CN202011626720, 2020. (**in Chinese, under examination**)

[10] Qingkai Min, Libo Qin, Zhiyang Teng, **Xiao Liu**, Yue Zhang. Dialog State Tracking Using Neural Latent Variable Models. *In Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence* (**IJCAI**), 2020.

[11] Changsen Yuan, Heyan Huang, Chong Feng, **Xiao Liu**, Xiaochi Wei. Distant Supervision for Relation Extraction with Linear Attenuation Simulation and Non-IID Relevance Embedding. *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence* (**AAAI**), 2019.

[12] Zhunchen Luo, **Xiao Liu**. Real-time Scholarly Retweeting Prediction System. *In Proceedings of the 27th International Conference on Computational Linguistics* (**COLING**), 2018.

[13] Wenhan Chao, Ping Wei, Zhunchen Luo, **Xiao Liu**, Guobin Sui. Selective Expression for Event Coreference Resolution on Twitter *In Proceedings of the 2019 International Joint Conference on Neural Networks* (**IJCNN**), 2019.

[14] Zhunchen Luo, Jun Chen, **Xiao Liu**. Real-Time Scientific Impact Prediction in Twitter. *In Proceedings of the 2018 CCF Conference on Big Data. Part of the Communications in Computer and Information Science, vol 945. Springer, Singapore.* (**BigData**), 2018.

SELECTED OPEN-SOURCE PROJECTS

★ *represents the number of github stars.*

- **JMEE** ★195 *https://github.com/lx865712528/EMNLP2018-JMEE*
  A neural sequence labeling framework with graph convolution network layer to jointly extract multiple events in a sentence. It gives the best performance on the event extraction task of the ACE 2005 dataset in 2018.
- **ODEE** ★94 *https://github.com/lx865712528/ACL2019-ODEE*
  A neural-extended unsupervised schema induction and open domain event extraction framework which supports incrementally learning. A business report corpus is published along with this framework.

TEACHING

Beijing Institute of Technology, Beijing, China

- **Teaching Assistant** Sep 2014 – Jan 2015
  - Algorithm and Data Structure
- **Teaching Assistant** Sep 2017 – Nov 2017
  - Selected Topics In New Technology for Computer

HONORS AND AWARDS

High-level Joint Doctoral Students Scholarship, China Aug 2018 – Sep 2019

- China Scholarship Council *No. 201806030142*
- Fully funded high-level Ph.D. students for jointly studying abroad officially.

Beijing Outstanding Graduates, Beijing Jun 2016

- This award program recognizes undergraduate students in Beijing universities who show outstanding grades and potential in their studies.

Outstanding Students Award, BIT 2012 – 2019

- Six times.
- Beijing Institute of Technology Annual Award and Honor. For rewarding excellent (top 5%) students in BIT in the academic year.

China Century Group Scholarship 2014

- Sponsor and award 4 undergraduate students and 2 postgraduate students who major in computer science every year in BIT by China Century Group.

First-class Ph.D. Academic Scholarship, BIT 2017 – 2019

- Two times.
- This award program recognizes top 5 Ph.D. students majored in computer science every year.

First-class M.S. Academic Scholarship, BIT 2016

- This award program recognizes top 30 master students majored in computer science every year.

First-class Undergraduate Academic Scholarship, BIT 2013 – 2016

- Four times.
- As for undergraduate student, select top 5% by last semester's GPA.

COMPETITIONS AND HACKATHONS

Rank 1st in the NLPCC 2021 Sharedtask 3 Jun 2021

- The sharedtask 3 in NLPCC 2021 is about Sub-event Identification, whose goal is to build an Information Extraction system that can quickly adapt to a new occurring sub-event.
- The team I lead won the championship, outperforming the second place 1.88% in accuracy.

Gold Medal in CCPC Regional Oct 2016

- The 2nd China Collegiate Programming Contest Regional Hefei Onsite
- The rules and difficulties are the same to the ACM/ICPC Regional contests.

Bronze Medal in ACM/ICPC Regional Oct 2014

- The 39th ACM/ICPC Regional Beijing Onsite

Bronze Medal in ACM/ICPC Regional Dec 2014

- The 39th ACM/ICPC Regional Shanghai Onsite

ACTIVITIES AND ORGANIZATIONS

Association of Programming and Algorithms, BIT

- **Co-founder** Sep 2013 – Jun 2014
  - Organize and invite programming contest experts to deliver lectures.
  - Host Freshman Enrollment Programming Contest in the end of the first semester every year.

Annual Programming Contest, BIT

- **Problem Provider and Onsite Contest Judge** May 2015 – May 2017
  - Provided several problems for the contests and verify all the selected problems.
  - Judged program solutions and did the real time Q&A about the problem descriptions during the contests.

PROGRAMMING LANGUAGES

- Python, C/C++, LaTeX, Java, Scala, Bash under Linux/Unix/Windows.
- PyTorch, Tensorflow, Keras.