

Haoyue Shi (a.k.a., Freda Shi)

Address: 6045 S. Kenwood Ave., Chicago, IL, USA, 60637

Email: freda@ttic.edu

Homepage: <https://home.ttic.edu/~freda>

Last updated: Wednesday 28th April, 2021

Research Interests

Computational linguistics, natural language processing and machine learning: compositional semantics, grounded language acquisition, unsupervised and semi-supervised representation learning, structured prediction, narrative understanding, and information theory for natural language processing.

Education

Toyota Technological Institute at Chicago, Chicago, IL, USA 2018-
Ph.D. student in Computer Science (Master's degree awarded Sept. 2020)
Advisors: Karen Livescu and Kevin Gimpel

Peking University, Beijing, China 2013-2018
B.S. in Intelligence Science and Technology (Computer Science Track), *Summa Cum Laude*
Minor in Sociology
Thesis title: On Multi-Sense Word Embeddings via Matrix Factorization and Matrix Transformation
Advisor: Junfeng Hu

Referred Conference Publications

* indicates equal contribution.

1. **Haoyue Shi**, Karen Livescu and Kevin Gimpel. 2020. On the Role of Supervision in Unsupervised Constituency Parsing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
2. **Haoyue Shi***, Jiayuan Mao*, Kevin Gimpel and Karen Livescu. 2019. Visually Grounded Neural Syntax Acquisition. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL; Best Paper Nominee)*.
3. **Haoyue Shi**, Hao Zhou, Jiaze Chen and Lei Li. 2018. On Tree-Based Neural Sentence Modeling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
4. **Haoyue Shi***, Jiayuan Mao*, Tete Xiao*, Yuning Jiang and Jian Sun. 2018. Learning Visually-Grounded Semantics from Contrastive Adversarial Samples. In *Proceedings of the 27th International Conference on Computational Linguistics (COLING)*.
5. **Haoyue Shi**, Jia Chen and Alexander G. Hauptmann. 2017. Joint Saliency Estimation and Matching using Image Regions for Geo-Localization of Online Video. In *Proceedings of the 2017 ACM International Conference on Multimedia Retrieval (ICMR)*.

Conference Publications Referred by Abstract

1. **Haoyue Shi**, Xihao Wang, Yuqi Sun and Junfeng Hu. 2018. Constructing High Quality Sense-specific Corpus and Word Embedding via Unsupervised Elimination of Pseudo Multi-sense. In *Proceedings of the 11th Language Resources and Evaluation Conference (LREC)*.
2. Shan Xu, **Haoyue Shi**, Xiaohui Duan, Tiangang Zhu, Peihua Wu and Dongyue Liu. 2016. Cardiovascular Risk Prediction Method Based on Test Analysis and Data Mining Ensemble System. In *Proceedings of the 2016 IEEE International Conference on Big Data Analysis*.

Referred Workshop Publications

1. Shubham Toshniwal, **Haoyue Shi**, Bowen Shi, Lingyu Gao, Karen Livescu and Kevin Gimpel. 2020. A Cross-Task Analysis of Text Span Representations. In *Proceedings 4th Workshop on Representation Learning for NLP*.
2. Yuqi Sun, **Haoyue Shi** and Junfeng Hu. 2018. Implicit Subjective and Sentimental Usages in Multi-sense Word Embeddings. In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*.
3. **Haoyue Shi**, Caihua Li and Junfeng Hu. 2016. Real Multi-Sense or Pseudo Multi-Sense: An Approach to Improve Word Representation. In *Proceedings of the 1st Workshop on Computational Linguistics for Linguistic Complexity*.

Conference and Workshop Presentations without Proceedings

1. **Haoyue Shi**, Jiayuan Mao, Kevin Gimpel and Karen Livescu. 2019. Visually Grounded Neural Syntax Acquisition. Talk, Midwest Speech and Language Days, Chicago, IL, USA, May 2-3.

Invited Presentations

1. *Naturally Supervised Parsing: Assumptions, Methods and Evaluation*. Yahoo! Remote Research Seminar NYC, April 15th, 2021.
2. *Visually grounded neural syntax acquisition*. Remote Computer Music Seminar, Carnegie Mellon University, August 19th, 2020.
3. *Structures in Natural Language: How to learn it and how to use it?* Remote NLP Seminar, University of Alberta, May 4th, 2020.
4. *Visually grounded neural syntax acquisition*. NLP Seminar, Peking University, Decemember 24th, 2019.

Open-Sourced Projects

1. **Global statistics-based bilingual lexicon inducer and word aligner**. Implementation of the paper *Bilingual Lexicon Induction via Unsupervised Bitext Construction and Word Alignment* (Shi et al., 2020).
2. **Visually grounded neural syntax learner** (developed with Jiayuan Mao). Implementation of the paper *Visually Grounded Neural Syntax Acquisition* (Shi et al., 2019). <https://ttic.uchicago.edu/~freda/project/vgns1>
3. **Tree-based neural sentence encoders**. Implementation of the paper *On Tree-Based Neural Sentence Modeling* (Shi et al., 2018). <https://github.com/explorerfreda/TreeEnc>
4. **Contrastive adversarial caption generator and evaluation framework for visual semantic embeddings** (developed with Jiayuan Mao). Implementation of the paper *Learning Visually-Grounded Semantics from Contrastive Adversarial Samples* (Shi et al., 2018). <https://github.com/explorerfreda/vse-c>
5. **Multimodal concreteness score estimator** (developed with Victor Silva). Implementation of the paper *Quantifying the Visual Concreteness of Words and Topics in Multimodal Datasets* (Hessel et al., 2018). <https://github.com/victorssilva/concreteness>
6. **Structured self-attentive sentence embeddings**. Implementation of the paper *A Structured Self-Attentive Sentence Embedding* (Lin et al., 2017). <https://github.com/explorerfreda/structured-self-attentive-sentence-embedding>

Honors and Awards

Google Ph.D. Fellowship	2021
Finalist, Facebook Ph.D. Fellowship	2021
Nomination for the Best Paper Award, ACL	2019
Best Undergraduate Dissertation Award (top 10 out of 400), School of EECS, Peking University	2018
Excellent Graduate Student, Peking University	2018
Pacemaker to Merit Student, Peking University	2016
Robin Lee Scholarship (top 2 out of 400), Peking University	2016
Top-Notch Scholarship, Chinese Ministry of Education	2016
WeTech Qualcomm Global Scholarship	2016
Merit Student, Peking University	2015
Arawana Scholarship (top 4 out of 400), Peking University	2015
Honorable Mention Prize, Mathematical Contest in Modeling (MCM)	2015
Yitian Minsheng Scholarship, School of EECS, Peking University	2014
Gold Medal and the Best Female Team, ACM-ICPC Chengdu Site	2013

Teaching Experience

Teaching Assistant at Toyota Technological Institute at Chicago TTIC 31220 Unsupervised Learning and Data Analysis	Instructor: Karen Livescu	Winter 2021
Teaching Assistant at School of EECS, Peking University Practice of Programming in C&C++	Instructor: Wei Guo	Spring 2018
Programming & Algorithms (MOOC on Coursera)	Instructor: Wei Guo	Fall 2016
Practice of Programming in C&C++	Instructor: Jiaying Liu	Spring 2015
Volunteer Lecturer in Mathematics, Rongxian High School, Guangxi, China		Summer 2014

Industrial Research Internships

Facebook AI Research (Remote Internship), Seattle, WA, USA Mentors: Sida Wang and Luke Zettlemoyer Project: Bilingual lexicon induction.		Jun. 2020-Dec. 2020
ByteDance AI Lab , Beijing, China Mentors: Hao Zhou and Lei Li Project: Structural sentence modeling.		Mar. 2018-Aug. 2018
Megvii (Face++) Research , Beijing, China Mentors: Yuning Jiang and Jian Sun Project: Visually-grounded semantics learning.		Oct. 2017-Mar. 2018
Microsoft Research Asia , Beijing, China Mentors: Zhongyuan Wang and Jun Yan Project: Metaphor extraction and interpretation.		Sep. 2016-Feb. 2017

Industrial Engineering Internships

4th Paradigm Inc. , Beijing, China Mentors: Weiwei Tu and Yuqiang Chen Project: Machine learning methods for diabetes risk prediction.		Mar. 2017-Jun. 2017
Google Inc. , Beijing, China Mentors: Xiaoyi Ren and Jie Mao Project: Wikipedia HTML template monitor.		Jul. 2015-Dec. 2015

Skills

Programming Languages:

- Proficient: C/C++, Python(2/3), MATLAB, Pascal, C#, SCOPE, HTML/CSS
- Capable: JavaScript, Java, Scala, Mathematica, Bash

Natural Languages:

Mandarin (native), English (fluent), classical Chinese (advanced reading & writing), Cantonese (intermediate listening & speaking), Japanese (intermediate), German (beginner), Hebrew (beginner), Spanish (beginner)

Tools & Frameworks: Vim, Caffe, Torch, PyTorch, GDB, Git, L^AT_EX, CMake, Visual Studio, ssh

Service

Program committee member/Reviewer: ACL, COLING, EMNLP, LREC, NAACL, NLPCC, ICLR, ICML, IJCAI, ViGiL Workshop, AAAI (secondary), NeurIPS (secondary), CVPR (secondary).

Co-organizer of TTIC Student Workshop, 2020.

Student Representative at TTIC, 2020-2021.

Student co-chair of the Women at TTIC group and **coordinator** with UChicago Graduate Women in CS, 2019-2021.

Peer mentor for new students at TTIC, 2019-2020.

Chief of PKU Guqin Society, 2016-2017.