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

Address: 6045 S. Kenwood Ave., Chicago, IL, USA, 60637
Email: freda@ttic.edu
Homepage: https://ttic.uchicago.edu/~freda
Last updated: Tuesday 4th June, 2019

Research Interests

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

Education

Toyota Technological Institute at Chicago, Chicago, IL, USA 2018-
Ph.D. student in Computer Science
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

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*. (*: Equal contribution.)

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*.

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*. (*: Equal contribution.)

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*.

Conference Publications Referred by Abstract

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*.

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

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*.

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

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.

Open-Sourced Projects

- **Visually grounded neural syntax learner** (developed with Jiayuan Mao).
Implementation of the paper *Visually Grounded Neural Syntax Acquisition* (Shi et al., 2019).
`https://github.com/explorerfreda/vgnsl`

- **Tree-based neural sentence encoders**.
Implementation of the paper *On Tree-Based Neural Sentence Modeling* (Shi et al., 2018).
`https://github.com/explorerfreda/TreeEnc`

- **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`

- **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`

- **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

Excellent Graduate Student, Peking University	2018
Best Undergraduate Dissertation Award, School of EECS, Peking University	2018
Pacemaker to Merit Student, Peking University	2016
Robin Lee Scholarship, Peking University	2016
Top-Notch Scholarship, Chinese Ministry of Education	2016
WeTech Qualcomm Global Scholarship	2016
Merit Student, Peking University	2015
Arawana Scholarship, Peking University	2015
Honorable Mention Prize, Mathematical Contest in Modeling (MCM)	2015

Best Foreign Team, ACM-ICPC Daejeon Site 2014
Yitian Minsheng Scholarship of Peking University 2014
Gold Medal and the Best Female Team, ACM-ICPC Chengdu Site 2013

TEACHING EXPERIENCE

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

ByteDance (TouTiao) AI Lab, Beijing, China Mar. 2018-Aug. 2018
Mentors: Hao Zhou and Lei Li
Project: Structural sentence modeling.

Megvii (Face++) Research, Beijing, China Oct. 2017-Mar. 2018
Mentors: Yuning Jiang and Jian Sun
Project: Visually-grounded semantics learning.

Microsoft Research Asia, Beijing, China Sep. 2016-Feb. 2017
Mentors: Zhongyuan Wang and Jun Yan
Project: Metaphor extraction and interpretation.

INDUSTRIAL ENGINEERING INTERNSHIPS

4^{th} Paradigm Inc., Beijing, China Mar. 2017-Jun. 2017
Mentors: Weiwei Tu and Yuqiang Chen
Project: Machine learning methods for diabetes risk prediction.

Google Inc., Beijing, China Jul. 2015-Dec. 2015
Mentors: Xiaoyi Ren and Jie Mao
Project:Wikipedia HTML template monitor.

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 (intermidiate listening), Japanese (beginner), Spanish (beginner)

Tools & Frameworks: Vim, Caffe, Torch, PyTorch, GDB, Git, LaTeX, CMake, Visual Studio, ssh

SERVICE

Secondary Reviewer: EMNLP 2018, NLPCC 2018, AAAI 2019.