Haoran Xu

443-453-8801 | hxu64@jhu.edu | haoranxu.com | github.com/haoranxu

Education

Johns Hopkins University (JHU) — Baltimore, USA
Incoming Ph.D. - Computer Science — Sep 2021
M.S.E - Computer Science; GPA: 3.93/4.0 — Aug 2019 – Dec 2020
Concentration on Human Language Technology (HLT)

University of Illinois at Chicago (UIC, 3+2 exchange program) — Chicago, USA
M.S. - Electrical and Computer Engineering; GPA: 3.93/4.0 — Aug 2017 – May 2019

East China University of Science and Technology (ECUST) — Shanghai, China
B.E. - Information Engineering; Major Ranking: 1/90 — Sep 2014 – July 2018

Main coursework: Nature Language Processing, Machine Learning, Human Language Technology, Algorithms, Pattern Recognition, Neural Networks, Computer Vision

Current Research areas: Multilingual Embeddings, Cross-Lingual Transfer Learning, Machine Translation

Publications

[1] **Haoran Xu**, Seth Ebner, Mahsa Yarmohammadi, Aaron Steven White, Benjamin Van Durme and Kenton Murray. **Gradual Fine-Tuning for Low-Resource Domain Adaptation**. **To appear at** Adapt-NLP, EACL, 2021.

[2] **Haoran Xu** and Philipp Koehn. **Zero-Shot Cross-Lingual Dependency Parsing through Contextual Embedding Transformation**. **To appear at** Adapt-NLP, EACL, 2021.

[3] **Haoran Xu** and Philipp Koehn. **Cross-Lingual BERT Contextual Embedding Space Mapping with Isotropic and Isometric Conditions**. **Under Review** at NAACL-HLT, 2021.

[4] The other paper in wireless communication: **Haoran Xu**, Shahin Khobahi and Mojtaba Soltanalian. **Efficient Quadratic Programming for Peak-to-Average Power Ratio Reduction in Communication Systems**.

Research Experience

Better Extraction from Text Towards Enhanced Retrieval Program [1] — Research Intern
Center for Language and Speech Processing, mentors: Benjamin Van Durme, Mark Dredze, JHU — May 2020 – Dec 2020

- Cooperated with the research group in cross-lingual transfer learning for the Information Extraction (IE) task.
- Proposed a novel data augmentation approach for domain adaptation, which improved the performance of the original IE system by 37%.
- Investigated the general applicability of the augmentation approach and achieved state-of-the-art performance in dialogue state tracking and event extraction tasks.

Zero-Shot Cross-Lingual Dependency Parsing [2] — Research Assistant
Center for Language and Speech Processing, advised by Philipp Koehn, JHU — May 2020 – Sep 2020

- Developed a zero-shot cross-lingual dependency parsing by building a multilingual shared semantic embedding space.
- Investigated the properties of embedding spaces across various languages.
- Surpassed state-of-the-art methods by 2.82% in 6 languages from 4 language families on average.

Cross-Lingual Contextual Embedding Space Mapping [3] — Research Assistant
Center for Language and Speech Processing, advised by Philipp Koehn, JHU — Sep 2019 – May 2020

- Developed a context-aware embedding space mapping method, which is sense-level and dictionary-free.
- Outperformed traditional static embedding alignment approaches by approximate 10% accuracy on the bilingual dictionary induction task by providing a higher degree of isomorphism.
- Revealed the tight relationship of isotropy, isometry and isomorphism in contextual embedding spaces, and explained their strong correlation with the quality of the mapping.

Efficient Quadratic Programming in Wireless Communication [4] Research Assistant

Waveform Optimization Lab, advised by Mojtaba Soltanalian, UIC Nov 2018 – Mar 2019

- Developed a new algorithm to reduce the peak energy of data transmission based on *Unimodular Quadratic Programming (UQP)*, and achieved higher stability and faster convergence rate than existing methods.
- Provided more mature and general techniques for solving the *UQP* problem in radar code design scenarios as well as other active sensing and communication applications.

Projects

Neural Machine Translation with Pre-trained Language Models Research Assistant

Center for Language and Speech Processing, JHU Dec 2020 – Present

- Developed neural machine translation models based on *fairseq* framework.
- Investigated an ensemble model composed of transformers and pre-trained language models.
- Achieved state-of-the-art BLEU scores between multiple language pairs.

Image Expansion with GANs Team Leader

Science and Engineering Laboratory, UIC Jan 2019 – May 2019

- Built a deep learning method based on *GAN* to naturally expand the boundaries of incomplete images.
- Designed an encoding-decoding hybrid *CNN* composed of dilated convolution and the normal one, rendering the extended image to acquire better realism and conform to the semantics of the whole image.
- Replaced traditional global discriminator with local discriminator to reduce the blur of the vertical part of the image and enhance authenticity.

Invisible Signature Security System Team Leader

Science and Engineering Laboratory, UIC Jan 2018 – May 2018

- Created *GUI* and local database for human-computer interaction, which allowed users to sign their names in the air to implement signature recognition.
- Developed high-dimension dynamic signature features extracted from the signature process to promise the uniqueness of the signature and the security of the system.
- Utilized *Fast-DTW (Fast Dynamic Time Warping)* and statistical information of high-dimension dynamic features to recognize the signature with high precision.
- Received First Place at the 2018 Engineering Expo at UIC and the Best Research Paper Award at ECUST.

Selected Awards

- Best Research Paper Award ECUST – 2018
- First Place at the 2018 Engineering Expo UIC – 2018
- Principal Scholarship (top 2%-ranked student) ECUST – 2015, 2016, 2018
- Social Work Award (top 5%-ranked student in community contribution) ECUST – 2015, 2016, 2018
- Excellent Student Award (4 selected among 1000 students) ECUST – 2015

Skills

Programming Languages: Python, MATLAB, C, bash, C++, Java, LaTeX, SQL

Toolkits and Libraries: PyTorch, Allennlp, Fairseq, Moses, NLTK, Sklearn, Keras, TensorFlow, PyQt, MySQLdb