

Jingfeng Yang

1000 NORTHSIDE DR NW APT 1319, Atlanta, Georgia (GA) 30318 +1 (404)513-4445 jyang690@gatech.edu

RESEARCH EXPERIENCE

Disfluency Detection | College of Computing, Georgia Institute of Technology | Research Intern Aug 2019 to present
Advisor: Diyi Yang, assistant professor in the School of Interactive Computing, Georgia Tech

- Used Pytorch to implement LSTM and BERT to improve the sequence labeling models in Disfluency Detection in Dialogue.
- Used Data Augmentation and Semi-supervised techniques, including generating auto-denoising fake data and using Cross View Training to achieve the state-of-the-art performance in Disfluency Detection.

Noun-Verb disambiguation in POS tagging | Microsoft Research Asia | Research Intern Dec 2018 to March 2019
Advisor: Jinge Yao and Chin-Yew Lin

- Used C# to develop a fast and accurate Structured Perceptron part-of-speech (POS) tagging tool as a part of Microsoft Forms.
- Used Meta Learning and ELMo to reproduce the state-of-the-art POS tagging results in PTDB dataset and Noun-verb Dataset.

Cross-Lingual and Mono-lingual Semantic Parsing to generate Discourse Relation Structure for Parallel Meaning Bank | University of Edinburgh | Research Assistant July 2018 to Sep 2018

Advisor: Bonnie Webber, professor in Institute for Language, Cognition and Computation

- Used Pytorch to implement a model composed of a tree-LSTM encoder and a three-stage coarse-to-fine decoder with attention mechanism and copying mechanism to conduct semantic parsing.
- Applied the model to generating Discourse Relation Structure Representations for Parallel Meaning Bank.
- Leveraged Universal Dependency structure information and cross-lingual word embeddings to conduct experiments in Italian, German, and Dutch while training the model using English data.

Cross-lingual and mono-lingual Elementary Discourse Units segmentation | Peking University | Research Assistant June 2017 to June 2019

Advisor: Sujian Li, associate professor in Institute of Computational Linguistics, Peking University

- Applied an adversarial multi-task neural network to cross-lingual Elementary Discourse Units segmentation using Tensorflow.
- Used SVM, Logistic Regression and Random Forest along with some useful features, including some specific constituency parsing tree features, to conduct English EDU segmentation.
- Developed a Fast and Accurate Elementary Discourse Units segmenter using BiLSTM-CRF, self-attention and ELMo.

PUBLICATIONS

- **Jingfeng Yang**, Federico Fancellu, Bonnie Webber. 2019. A survey of cross-lingual features for zero-shot semantic parsing. *arXiv:1908.10461*
- **Jingfeng Yang**, Sujian Li. 2018. Chinese Discourse Segmentation Using Bilingual Discourse Commonality. *arXiv:1809.01497*.
- Yizhong Wang, Sujian Li, **Jingfeng Yang**. 2018. Toward Fast and Accurate Neural Discourse Segmentation. In *2018 Empirical Methods in Natural Language Processing (EMNLP)*.
- Yizhong Wang, Sujian Li, **Jingfeng Yang**, Xu Sun, Houfeng Wang. 2017. Tag-enhanced tree-structured neural networks for implicit discourse relation classification. In *The 8th International Joint Conference on Natural Language Processing (IJCNLP)*.

TEACHING EXPERIENCE

CS 4650-7650 Natural Language Processing | Teaching Assistant Spring 2020
Georgia Institute of Technology – Atlanta, GA Instructor: Diyi Yang

EDUCATION

Master of Science: Computer Science Expected in May 2021
Georgia Institute of Technology – Atlanta, GA **GPA:** 4.00/4.00

Core Courses: Machine Learning (4.0) / Deep Learning (4.0) / Deep Learning for Text Data (4.0)

Bachelor of Science: Computer Science and Biological Science Sep 2015 to July 2019
Peking University – Beijing, China **GPA:** 3.70/4.00 (Computer Science) | 3.57/4.00 (Biological Science)

Core Courses: Empirical Method in Natural Language Processing (95.5) / Mathematical Logic(95) / Discrete Mathematics (94) / JavaScript Web Programming (94) / Software Engineering (93) / Java Programming (92) / Data structure and Algorithm (90.5) /

Introduction to The Database (90) / Introduction to Parallel and Distributed Computing (89) / Linear Algebra(89) / Probability Theory and Statistics (A) (89) / Compiler Design (89) / Principle of Microcomputer (89) / Numerical Methods(88) / Networking Technology and Practices (88) /Advanced Mathematics(87) / Operating Systems / Algorithm Design and Analysis / Introduction to Computer Networks / the C++ Programming Language /...

SKILLS

- | | | |
|-------|--------|--|
| ● C++ | ● Java | ● Python/PyTorch/TensorFlow/Numpy/Scipy/Scikit-Learn |
| ● C# | ● SQL | ● JavaScript/Jquery/Express/HTML/CSS/XML |

SELECTED PROJECT EXPERIENCE

Predictive Text Embeddings in Text Classification | Python Oct 2019 to Nov 2019

- Use Predictive Text Embeddings in the semi-supervised and supervised text classification task.
- Use CNN, logistic regression and SVM to test embeddings in three text classification datasets.

Humor Classification | Python Oct 2019 to Nov 2019

- Use Pytorch to implement CNN and BERT in humor classification.
- Use Active Learning and Multi-task learning to improve the task of binary humor classification with the task of fine-grained humor classification.
- Use gradient ascent in the word embedding space to create adversarial examples, which makes the model more robust.

Biomedical Question Answering | Python/PyTorch Feb 2019 to June 2019

- Used PyTorch to implement a question-based extractive summarization system to automatically answer questions in the biomedical domain and ranked the 5st in the BioASQ Task 7b Phase B.

Semantic Role Labeling | Python/ Scikit-Learn April 2018 to May 2018

- Used linguistic features, especially constituency tree path, and SVM in semantic role labeling and achieved competitive results in PropBank.

YouKnow App | Javascript/Express Feb 2018 to June 2018

- Designed an app where users with mobile phones could subscribe some websites and the server could push useful messages to users when the subscribed website is updated.
- Used Express (Javascript) framework and MySQL in backend and Android (Java) in frontend. Implemented real-time server push using JIGUANG.

Battle City | Java Feb 2018 to June 2018

- Implemented a fully functional game Battle City using Java, including GUI, two-player mode and human-machine mode.
- Generated maps of battlefields automatically according to any given pictures.

Parallel Implementation of Single-Source Shortest Path (SSSP) | C++/OpenMP Feb 2019 to June 2019

- Used C++ and OpenMP to implement both serial and parallel versions of Bellman-Ford algorithm, Dijkstra algorithm and Δ -stepping algorithm to solve SSSP problem. Improved the speed in the benchmark of the 9th DIMACS Implementation Challenge.

Value Range Analysis in Compiler | Python May 2018 to June 2018

- Transformed SSA file to eSSA code, constructed constraints graph and computed strongly connected components
- Conducted widening, future resolution and narrowing to finish range value analysis.

FELLOWSHIPS/AWARDS

- May 4th Scholarship, 2016-2017
- Kwang-Hua Scholarship, 2015-2016
- Merit Student of Peking University, 2015-2016, 2016-2017
- Silver medal in China Mathematics Olympiad (CMO), 2015