Jingfeng Yang

1000 NORTHSIDE DR NW, Atlanta, GA, 30318 +1 (404)513-4445 jyang690@gatech.edu https://jingfengyang.github.io/

EDUCATION

Master of Science: Computer Science Expected to Graduate in May 2021
Georgia Institute of Technology – Atlanta, GA **GPA**: 4.00/4.00
Courses: Machine Learning (4.0)/Deep Learning (4.0)/Deep Learning for Text Data (4.0)/Design and Analysis of Algorithms (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)
Courses: Empirical Method in Natural Language Processing(95.5)/ Discrete Mathematics(94)/ JavaScript Web Programming(94)/ Software Engineering(93)/ Java Programming(92)/ Data Structure and Algorithm(90.5)/ Introduction to Database(90)/ Compiler Design(89)/ Introduction to Parallel & Distributed Computing(89)/ Linear Algebra(89)/ Probability Theory & Statistics(89)/ …

RESEARCH EXPERIENCE

Semantic Parsing | Georgia Institute of Technology | Research Intern | Atlanta | August 2020 to now
Advisor: Diyi Yang, assistant professor in the School of Interactive Computing, Georgia Tech
● Used two LSTM and Transformer-based semantic parsers with encoders optionally replaced by XLM-R, and found that Universal Dependency relation features significantly and consistently improves zero-shot cross-lingual semantic parsers.
● Proposed to recompose grammar rules to augment training data, where CCG or CFG grammars were first induced.
Disfluency Detection and Generation | Georgia Institute of Technology | Research Intern | Atlanta | Aug 2019 to May 2020
Advisor: Diyi Yang, assistant professor in the School of Interactive Computing, Georgia Tech
● Used PyTorch to build state-of-the-art LSTM, Transformer and BERT sequence labeling models in disfluency detection.
● Adapted OpenNMT to design a model (Planner and Generator) for generating disfluent sentences, which were used as augmented data to train the disfluency detection model, achieving SOTA in both disfluency generation and disfluency detection.
Stylistic Text Generation | Georgia Institute of Technology | Research Intern | Atlanta | Jan 2020 to May 2020
Advisor: Diyi Yang and Yangfeng Ji, assistant professor in the Department of Computer Science, University of Virginia
● Built a style transfer benchmark dataset, which included formality transfer, sentiment transfer, neutralizing subjective biases etc.
● Experimented with several models (e.g. Seq2Seq, CopyNet, BART) and achieved SOTA in the style transfer benchmark dataset.
Noun-Verb disambiguation in POS tagging | Microsoft Research Asia | Research Intern | Beijing, China | Dec 2018 to Mar 2019
Advisor: Jinge Yao and Chin-Yew Lin
● Used C# to develop a Structured Perceptron Part-of-Speech (POS) tagger for query understanding in Microsoft Forms, where the POS tagger can make predictions for hundreds of sentences within one second and have more than 92% accuracy.
● Used Meta Learning and ELMo to achieve state-of-the-art POS tagging results in PTDB (98%) and Noun-verb (89%) Dataset.
Cross-lingual Semantic Parsing | University of Edinburgh | Research Intern | Edinburgh, UK | 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 for semantic parsing in English Parallel Meaning Bank dataset.
● Leveraged Universal Dependency and cross-lingual word embeddings to transfer the model to Italian, German, and Dutch.
Elementary Discourse Units segmentation | Peking University | Research Assistant | Beijing, China | June 2017 to June 2019
Advisor: Sujian Li, associate professor in Institute of Computational Linguistics, Peking University
● Designed an adversarial multi-task neural network for cross-lingual Elementary Discourse Units segmentation with TensorFlow. The model was composed of a private feature extractor, a common feature extractor and a language discriminator to extract language-independent representations, and a discourse segmenter. The model achieved SOTA in cross-lingual EDU segmentation.
● Used SVM, Logistic Regression, BiLSTM-CRF, self-attention and ELMo to achieve SOTA in English EDU segmentation.

PUBLICATIONS

● **Jingfeng Yang**, Diyi Yang, Zhaoran Ma. 2020. Planning and Generating Natural and Diverse Disfluent Texts as Augmentation for Disfluency Detection. In *EMNLP 2020.*
● **Jingfeng Yang**, Federico Fancellu, Bonnie Webber, Diyi Yang. 2020. Frustratingly Simple but Surprisingly Strong: Using Language-Independent Features for Zero-shot Cross-lingual Semantic Parsing. Submitted to *NAACL 2021.*

● Yang Zhong, **Jingfeng Yang**, Wei Xu, Diyi Yang. 2020. MsBC: On Automatically Detecting Multi-Span Subjective Bias. Submitted to *NAACL 2021.*

● Stephanie Schoch, Wanyu Du, **Jingfeng Yang**, Diyi Yang, Yangfeng Ji. 2020. It Takes Two - Tasks and Datasets! Linguistically -Informed Analysis of Style Transfer. Submitted to *TACL.*

● **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 *EMNLP 2018.*

● 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*).

WORK EXPERIENCE

CS 4650/7650 Natural Language Processing | Georgia Tech | Teaching Assistant | Atlanta, GA Fall 2020

● Instructor: Diyi Yang

Unified Repair Portal | Amazon | Software Development Engineer intern | San Francisco, CA May 2020 to July 2020

● Used Spring (Java), Amazon/AWS tools (Horizonte, DynamoDB, S3, AWS Lambda, DJS, EDX, ETL Manager etc.), Elasticsearch and Weblab to build a unified portal for repair vendors to easily exchange information with Amazon.

CS 4650/7650 Natural Language Processing | Georgia Tech | Teaching Assistant | Atlanta, GA Spring 2020

● Instructor: Diyi Yang

SKILLS

● C++ / C# ● Java / Spring ● Python/PyTorch/TensorFlow/Numpy/Scikit-Learn/OpenNMT/fairseq

● AWS tools. ● SQL / NoSQL / Spark / Hadoop ● JavaScript/Jquery/Express/HTML/CSS/XML

SELECTED PROJECT EXPERIENCE

Predictive Text Embeddings (PTE) in Text Classification | Python Oct 2019 to Nov 2019

● Used PTE-based CNN, logistic regression and SVM for semi-supervised and supervised text classification in three datasets.

Humor Classification | Python / PyTorch Oct 2019 to Nov 2019

● Implemented CNN and BERT for humor classification., where active learning and multi-task learning were used to improve binary humor classification with fine-grained humor classification. Also, gradient ascent were used in the word embedding space to create adversarial examples, making the model more robust.

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

● Implemented 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 (e.g. constituency paths) and SVM in semantic role labeling, achieving 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, where Express (Javascript) framework and MySQL were used in backend, Android (Java) in frontend, and JIGUANG used for real-time server push.

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