Jingfeng Yang

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

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

WORK HISTORY

Unified Repair Portal | Amazon | Software Development Engineer intern | San Francisco 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.

● Enabled Repair vendors to view, filter and download repair reports that are automatically generated from Business Intelligence table, which helps business managers save 90% time generating and sending reports each week.

Microsoft Forms | Microsoft Research Asia | Research and Software Engineer Intern | Beijing, China Dec 2018 to March 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.

Disfluency Detection and Generation | Georgia Institute of Technology | Research Assistant | 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 Encoder-Decoder) in disfluency generation for dialogue and text-to-speech.

Stylistic Text Generation | Georgia Institute of Technology | Research Assistant | 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 includes formality transfer, sentiment transfer, neutralizing subjective bias etc.

● Experimented with several models (e.g. Seq2Seq, CopyNet, BART) and achieved SOTA in the style transfer benchmark dataset.

Semantic Parsing | University of Edinburgh | Research Assistant | 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 to conduct 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

● Applied an adversarial multi-task neural network to cross-lingual Elementary Discourse Units segmentation using Tensorflow.

● Used SVM, Logistic Regression, Random Forest, BiLSTM-CRF, self-attention and ELMo for English EDU segmentation.

PUBLICATIONS

● **Jingfeng Yang**, Diyi Yang. 2020. Planning and Generating Natural and Diverse Disfluent Texts as Augmentation for Disfluency Detection. In *2020 EMNLP.*

● **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 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*).

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) /Introduction to Parallel & Distributed Computing(89)/Linear Algebra(89)/Probability Theory & Statistics(89)/Compiler Design(89)/Numerical Methods(88)/Advanced Mathematics(87)/ Operating Systems/ Introduction to Computer Networks / C++ Programming Language

TEACHING EXPERIENCE

CS 4650/7650 Natural Language Processing | Teaching Assistant Fall 2020

Georgia Institute of Technology – Atlanta, GA Instructor: Diyi Yang

CS 4650/7650 Natural Language Processing | Teaching Assistant Spring 2020

Georgia Institute of Technology – Atlanta, GA Instructor: Diyi Yang

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 binary humor classification with 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