Yian Zhang

+86 136 7192 6669 | yian.zhang@nyu.edu | 1555 Century Ave, Shanghai, China, 200122 | Personal website: yianzhang.github.io

EDUCATION

New York University Shanghai Expected graduation time: May 2021

Bachelor of Science in Computer Science, Minor in Mathematics

Cumulative GPA: 3.91/4.00 | Major GPA: 3.95/4.00

Selected Coursework: Machine Learning for Language Understanding (A), Artificial Intelligence (A), Parallel Computing (A) Natural Language Processing (A), Operating Systems (A), Basic Algorithms (A), Data Structures (A)

Scholarship: NYUSH 2018 Dean's Undergraduate Research Funding, NYUSH 2019 Recognition Award, NYUAD 2019 Visiting Undergraduate Research Scholarship, NYUSH 2020 Recognition Award

PUBLICATIONS & PREPRINTS

[1] **Yian Zhang***, Alex Warstadt*, Haau-Sing Li, and Samuel R. Bowman. When Do You Need Billions of Words of Pretraining Data? *arXiv:2011.04946 preprint*, 2020. [URL]

[2] **Yian Zhang**. Latent Tree Learning with Ordered Neurons: What Parses Does It Produce? *The 2020 EMNLP Workshop on Analyzing and interpreting neural networks for NLP (BlackboxNLP)*, 2020. [URL]

[3] Alex Warstadt, **Yian Zhang**, Haau-Sing Li, Haokun Liu, and Samuel R. Bowman. Learning Which Features Matter: RoBERTa Acquires a Preference for Linguistic Generalizations (Eventually). *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020. [URL]

[4] Daniel Chin, **Yian Zhang**, Tianyu Zhang, Jake Zhao, and Gus Xia. Interactive Rainbow Score: A Visual-centered Multimodal Flute Tutoring System. *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, 2020. [URL]

[5] **Yian Zhang**, Yinmiao Li, Daniel Chin, and Gus Xia. Adaptive Multimodal Music Learning via Interactive-haptic Instrument. *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, 2019. [URL]

RESEARCH

Investigating the Amount of Pretraining Data Required to Learn Different NLU Skills

Research Assistant advised by Professor Sam Bowman ML^2, CILVR, NYU, May 2020 – Current

- Probe a group of 16 RoBERTa models pretrained on different amounts of data using classifier probing, information-theoretic probing, and unsupervised grammaticality judgement probing, and test them on downstream NLU tasks.
- Find that 90% of the attainable learning of the linguistic knowledge we test can be made with 0.3% of RoBERTa's original pretraining data, while commonsense knowledge and practical NLU skills take much more data to acquire.
- Co-author a paper (as the **1st author**) which will be submitted to the conference of **ACL 2021.**

Investigating the Impact of Pretraining on RoBERTa's Inductive Bias

Research Assistant advised by Professor Sam Bowman ML^2, CILVR, NYU, January 2020 – Current

- Using a downward hyperparameter and architecture searching algorithm, pretrain a total of 102 RoBERTa models on 1M, 10M, 100M, 1B words.
- Find that increasing pretraining data does improve RoBERTa's inductive bias towards making linguistic rather than trivial generalizations, but teaching inductive biases using pretraining is extremely data inefficient.
- Publish a paper (as the **2nd author)** at the conference of **EMNLP 2020**; a follow-up work that focuses on spoken language and structural biases will be submitted to the journal *Language*.

An Analysis of the Unsupervised Parsing Behavior of ON-LSTM

Course Project inspired by Professor Sam Bowman NYU, March 2020 – October 2020

- Reproduce ON-LSTM five times and compute the average F1 score between each pairing of the five parses (self F1 score) and the model's constituency parsing accuracy on different constituent types.
- Find that the model is fairly consistent (self F1 41 higher than the random baseline), but it often makes mistakes in parsing the internal structures of complex noun phrases and has a tendency to branch before verbs too early.
- Propose solving both issues by using more syntactically demanding training tasks that support bidirectional training.
- Publish a paper (as the **1st author**) at the **EMNLP 2020 workshop BlackboxNLP.**

Haptodont: Haptic-based Dental Simulation

Advised by Professor Mohamad Eid AIM Lab, NYU Abu Dhabi, June 2019 – August 2019

- The goal of the project is to use virtual reality technologies to simulate an oral environment, where dental students can interactively practice dental probing under haptic and visual guidance.
- Design and implement the 3D recording and playback modes which support recording instructor demonstration and using the recording to give real-time haptic and visual guidance to the learner according to his/her hand movement.
- Resolve the problems of abrupt force variation and oscillation, by developing a force transition smoothing function and a basic PID controller.

Interactive Multimodal Music Learning System

Advised by Professor Gus Xia Music X Lab, NYU Shanghai, April 2018 – Current

- The project aims to build an interactive learning environment that teaches flute playing by giving real-time haptic, audio, and visual feedbacks.
- Design the innovative "Clutch" mechanism that allows instant adjustment of the haptic guidance level and a dynamic learning algorithm that boosts the learning rate by 45.3% and shrinks the forgetting chance by 86%.
- Build both the hardware and software of the system, and design and conduct user studies to test the learning effect.
- Publish a paper at **NIME 2019 (1st author)** and another paper at **NIME 2020 (2nd author).**

PROJECTS

Eticket Booking System September 2020 – Current

- Using MySQL, Express, and Node.js, build an airline ticket booking system where clients can book tickets and airline employees can manage flight information.

Boring Blog January 2020 – May 2020

- Using Mongodb, Express, React, and Node.js, build a blog platform where users can sign up and post articles.
- Use the TF-IDF algorithm to efficiently recommend to the users the articles they are least interested in.

Sentiment Transfer September 2019 – December 2019

- Train a GRU-based seq2seq model with a disentangled latent space to transfer negative yelp reviews to positive ones without changing the content.

Chat App September 2018 – December 2018

- Build an encrypted chat system over TCP with Tkinter powering its GUI.

INTERNSHIPS

Summer Intern

Gopher Asset July 2018 – August 2018

- Perform statistical analysis and industry research at the fund of funds team to assist decision making.

Summer Associate

AlphaSights June 2018 – July 2018

- Connect strategic consultants from McKinsey & Company to professional experts that provide industry insights.

SKILLS & INTERESTS

Languages: Native in Mandarin, Working proficiency in English

Programming: Python, C++, Java, Javascript

Hobbies: Piano, Kung Fu, Soccer, NBA