Tian Yun

[placeholder_personal_webpage] | tian_yun@brown.edu | (530) 551-9835

EDUCATION

Brown University Providence, RI
Master of Science in Computer Science. *GPA:4.00/4.00* *August 2020 - Expected May 2022*

Wake Forest University Winston-Salem, NC
Bachelor of Science in Mathematical Statistics *GPA:3.92/4.00* *August 2016 - May 2020*
Honors Bachelor of Science in Computer Science *GPA:3.96/4.00*

- **Honors**: Summa Cum Laude, Dean's list all semesters, Member of Phi Beta Kappa Honor Society

RESEARCH EXPERIENCE

LUNAR Lab, Brown University Providence, RI
Co-advised by Ellie Pavlick and Chen Sun *August 2020 – Present*

- Investigating whether vision-and-language (VL) models learn generalizable concepts.
- Conducted a set of probing experiments to investigate whether the presence of grounded pretraining in VL models yield better linguistic representations.
- Found that VL pretraining in its current form does not have benefits for NLP in general.

Humans to Robots Lab, Brown University Providence, RI
Advised by Stefanie Tellex *March 2021 – Present*

- Proposing a situation-aware coaching dialogue system that guides users to accomplish daily tasks by reasoning over uncertainties in the environment and asking clarification questions to the users.

BigScience Workshop, HuggingFace Providence, RI
Advised by Ellie Pavlick *May 2021 – Present*

- Co-leading the engineering team in Evaluation working group to build a zero-shot evaluation pipeline
- Constructing a suite of tests to evaluate how well language models understand word relations and word meanings on low-level concepts (e.g. colors, spatial directions).

DataMine Research Group, Wake Forest University Winston-Salem, NC
Advised by Natalia Khuri *August 2019 – August 2020*

- Built a pipeline with machine learning algorithms to extract information about safety and efficacy of drugs in pediatric populations from approved drug product labels.

PUBLICATIONS

[1] *Does Vision-and-Language Pretraining Improve Lexical Grounding?*
Tian Yun, Chen Sun, Ellie Pavlick.
Findings of EMNLP, 2021.

[2] *Mining Biomedical Texts for Pediatric Information.*
Tian Yun, Deepti Garg, Natalia Khuri
14th International Joint Conference on Biomedical Engineering Systems and Technologies, 2020.

THESIS

An Introduction to Topic Modeling and Sentiment Analysis Winston-Salem, NC
Advised by Natalia Khuri *April 2020 – August 2020*

- Re-implemented LDA with Gibbs sampling process to label documents with latent topics in R.
- Conducted sentiment analysis on literature works and Twitter comments dataset using bag-of-words, n-grams, word hashing, TF-IDF, etc.

Segmentation and Classification of Satellite Maps with Deep Learning Winston-Salem, NC
Advised by Nicole Dalzell *September 2019 – May 2020*

- Constructed a U-Net architecture to extract roads from Dstl satellite maps (Jaccard index of 0.12).
- Expanded and augmented limited number of satellite maps by cropping, sampling, and flipping.

PROFESSIONAL EXPERIENCE

Tencent Holdings Limited Shenzhen, China
Data Analyst Intern *May – August 2019*

- Extracted and processed billions of rows of data to monitor QQ Kandian and Kandian APP using Hive-SQL.
- Used Flask and Git Webhook to develop a back-end service to automate script uploading and task scheduling.
- Evaluated and predicted the loss of daily active users in the following week (AUC score of 0.9173)

China Minsheng Bank Beijing, China
Data Analyst Intern *May – July 2018*

- Aggregated and evaluated data for 8 million customers and obtained 500 thousand records with all the information of the customers (i.e. over 200 features) using SQL on DBM2.
- Fitted statistical models to predict the target customers (i.e. 48,000/8 million) who may have over 50,000 RMB savings next month.

TEACHING ASSISTANTSHIPS

CSCI-1470/2470: Deep Learning, Brown University *September 2021 – Present*
Teaching Assistant Instructor: Chen Sun

CSCI-1460: Computational Linguistics, Brown University *January 2021 – April 2021*
Teaching Assistant Instructor: Eugene Charniak

Computer Science Peer Tutor, Wake Forest University *September 2017 – May 2020*

AWARDS

ICPC Mid-Atlantic USA Regional Contest 2019, UNC – Chapel Hill Chapel Hill, NC
4th Place *November 2019*

- Designed and implemented algorithms in Python to solve problems

ASA DataFest 2019, Duke University Durham, NC
Honorable Mention *April 2019*

- Proposed Fatigue Composite Index to quantify the individual variation in self-reported data from athletes in Canadian National Women's Rugby Team

COMAP's Mathematical Contest in Modeling 2019 Winston-Salem, NC
Honorable Mention *January 2019*

- Designed Environmental Degradation Index to measure environmental degradation costs from human activities.

SERVICE

Conference Reviewer

- Computer Vision and Pattern Recognition (CVPR) *2022*