Rongzhi Zhang

Email: rongzhi.zhang@gatech.edu | Tel: (+1) 857-999-7305

Address: CODA 1312, 756 W Peachtree St NW, Atlanta, GA 30308, USA

EDUCATION BACKGROUND:

Georgia Institute of Technology *M.S. in Electrical and Computer Engineering* 08/2019 – Present
Overall GPA: 3.87/4.00

Harvard Medical School *Senior Thesis Research* 09/2018 – 05/2019

Zhejiang University *B.E in Measurement Control Technology and Instruments* 09/2015 – 06/2019
Major GPA: 3.93/4.00 Overall GPA: 3.83/4.00 Ranking: Top 3%

PUBLICATIONS:

1. **Rongzhi Zhang**, Yue Yu and Chao Zhang, "SeqMix: Augmenting Active Sequence Labeling via Sequence Mixup", In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP'20),* 2020.

2. **Rongzhi Zhang**, Yang Luo, Yue Yu and Chao Zhang, "Few-shot event extraction via modularized network", In *2021 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL'21),* 2021. (Submitted)

RESEARCH PROJECTS:

Few-shot Event Extraction *Prof. Chao Zhang*
Special Problem, Georgia Tech 07/2020 – 11/2020

- Propose a novel few-shot task for event extraction regarding the pretraining on base event types and the finetuning on unseen event types.
- Design a module-based argument detection model to enhance the arguments composition of the unseen events.

Augmenting Active Sequence Labeling *Prof. Chao Zhang*
Special Problem, Georgia Tech 02/2020 – 06/2020

- Propose an augmentation method for active sequence labeling by generating extra labeled sequences in each active learning round.
- Perform linear interpolation for both sequences and token-level labels using the queried samples.
- Employ a discriminator to judge whether the generated sequences are plausible or not.
- The proposed framework improve the standard active sequence labeling method by 2.27%-3.75% in terms of F1 scores.

Role Identification in Online Discussion of Political Events *Prof. Srijan Kumar*
Web Search & Text Mining, Georgia Tech 08/2019 – 12/2019

- Construct a comprehensive representation of the Reddit users who participated in political discussion from multiple perspectives like topics, discourse acts, linguistic features and social network analysis.
- Propose a framework to utilize Gaussian Mixture Model to identify roles in online discussion of political events.
- Interpret and evaluate the 11 derived roles and access the stability and dynamics of roles.

New Words Detection in Medical Field
AI Team, Beijing GoodWill Corp. 05/2019 – 07/2019

- The electronic health records documented in Chinese were processed by a model of CRF on top of bi-LSTM network to detect new words by tagging them as named-entities.
- Use the Chinese characters and their radicals as the input features, considering Chinese corpus involving diseases and organs reflect the same radical characteristics more closely.
- The dataset consists of 680k BIEO-labeled data and dropout training was implemented, reaching 92.0% precision and 88.4% recall.

Noninvasive and Continuous Blood Pressure Measurement *Prof. Quan Zhang*

Senior Thesis Research, Harvard Medical School 09/2018 – 05/2019

- Proposed a novel multi-element blood pressure measurement method and designed a sensor array that reduce the sensing area by 64%, overcame the difficulty of precise positioning and frequent calibration.
- Configured multiple sensors in C++ and explored multi-channel physiological signal fusion to increase the frequency band and enhance the monitoring accuracy
- Trained LSTM to reduce the mean absolute deviation of blood pressure estimation lower than 7, meeting the IEEE 1708 standard for wearable BP instruments.

Data Mining and NLP for Analysis of Electronic Medical Record *Prof. Zhengxing Huang*

Medical Informatics Lab, ZJU 04/2017 – 06/2018

- Analyzed medical record structure and proposed 10 rules based on regular expression to process complex syntactic structures in clinical electronic health records (EHRs).
- Applied CRF-based model in word segmentation and part-of-speech tagging.
- Applied rule-based model to extract key clinical information documented in EHRs and output structural summary for the EHRs which originally in the form of natural language.
- The extracted results of 6000 medical records give a 97% precision and 80% recall.

ACTIVITIES EXPERIENCE:

Leading Player of University Basketball Team, Zhejiang University 09/2016 – 06/2018

- Enrolled in the finals of Chinese University Basketball Association, the top series in which undergraduates are eligible to participate
- Led the team to win the championship in the University Basketball Association of Zhejiang Province

Leading Player of University Chinese Basketball Team, Harvard University 09/2018 – 05/2019

- Enrolled in the Northeastern Chinese Basketball League on behalf the Harvard Chinese Students & Schoolers Association (HCSSA).

AWARDS:

- The Outstanding Graduate Student 06/2019
- The Outstanding Undergrad Thesis 05/2019
- The First Prize of Excellent Undergraduate Scholarship (top 3%) 2017 – 2018 Academic Year
- The First Prize of Academic Scholarship (top 3%) 2017 – 2018 Academic Year
- Zhejiang Provincial Government Scholarship 2017 – 2018 Academic Year
- Student Innovation Research Funding in Zhejiang Province 05/2016
- The First Grade Award in the 31st Chinese Physics Olympiad (CPhO) 10/2014

MISC:

- Teaching Assistant for CS 7641: Machine Learning 08/2020 - Present
- Reviewer for TKDE