

# Zhenghui Wang

CODA E1540L – Georgia Institute of Technology – Atlanta, USA

✉ zhwang@gatech.edu • 🌐 zhenghuiwang.net

## Education

---

### Georgia Institute of Technology

Atlanta, GA

Aug. 2019 - Now

- PhD Student, School of Computational Science and Engineering
- Advisors: Prof. [Jimeng Sun](#)

### Shanghai Jiao Tong University

Shanghai, China

Sep. 2015 - Jun. 2019

- B.Eng. in Computer Science and Technology
- Zhiyuan Honors Program in Engineering (an elite program for top 5% talented students)
- Advisors: Prof. [Weinan Zhang](#) and Prof. [Yong Yu](#)

## Internship

---

### ByteDance AI Lab

Shanghai, China

Feb. 2019 - Jun. 2019

- Research Intern
- Advisor: Dr. [Hao Zhou](#)

### Carnegie Mellon University

Pittsburgh, USA

Jul. 2018 - Jan. 2019

- Research Assistant, Computational Biology Department, School of Computer Science
- Advisor: Prof. [Jian Ma](#)

## Publications

---

### Label-aware Double Transfer Learning for Cross-specialty Medical Named Entity Recognition

- [Z. Wang](#), [Y. Qu](#), [L. Chen](#), [J. Shen](#), [W. Zhang](#), [S. Zhang](#), [Y. Gao](#), [G. Gu](#), [K. Chen](#), and [Y. Yu](#).
- In *Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Techniques. NAACL HLT 2018* (oral, 6.73%).

### Sampled in Pairs and Driven by Text: A New Graph Embedding Framework

- [L. Chen](#), [Y. Qu](#), [Z. Wang](#), [L. Qiu](#), [W. Zhang](#), [K. Chen](#), [S. Zhang](#), and [Y. Yu](#).
- In *Proceedings of the 30th Web Conference. WWW 2019*.

## Honors and Awards

---

- **Zhiyuan Honor Degree of B.Eng in Computer Science and Technology** (Top 1% in SJTU) 2019
- **Outstanding Graduate of Shanghai** 2019
- **A-class Scholarship for Excellent Academic Performance** (Top 1% in SJTU) 2018
- **Rong Chang Innovation Scholarship** (Top 1% in SJTU) 2018
- **Scholarship of Excellent Undergraduates** (Top 2% in CS Department) 2018
- **Wish Scholarship** (Top 3% in CS Department) 2017
- **Finalist Winner Prize, Mathematical Contest in Modeling** (Top 1% among global participants) 2017
- **China National Scholarship** (Top 2% in CS Department) 2016
- **Zhiyuan Honor Scholarship** (Top 5% in SJTU) 2016 & 2017
- **B-class Scholarship for Excellent Academic Performance** (Top 10% in SJTU) 2016 & 2017
- **Second Prize in China Undergraduate Mathematical Contest in Modeling, Shanghai Division** 2016
- **SJTU Merit Student** (Top 5% in SJTU) 2016

## Research Experiences

---

### Question Answering for Information Extraction

ByteDance AI Lab, Shanghai

Guide: Dr. Hao Zhou, ByteDance AI Lab

Feb. 2019 - Jul. 2019

- o Constructed a huge Chinese reading comprehension dataset, which contains more than 800K articles and 3M QA pairs (5% of which contain multiple answers).
- o The implemented QA4IE system achieves 0.4364 F1 on title triple extraction and 0.2273 F1 on full triple extraction.

### Interpretable Prediction of Gene Sequence

CMU, Pittsburgh

Guide: Prof. Jian Ma, SCS, CMU

Jul. 2018 - present

- o Incorporated advanced NLP techniques (e.g., deep contextualized word representations) into the modeling of the gene sequence.
- o Improved *learning to explain* framework with attention mechanism to do interpretable prediction of the gene sequence.

### Transfer Learning for Named Entity Recognition

SJTU, Shanghai

Guide: Prof. Weinan Zhang, Prof. Yong Yu, CSD, SJTU

Jun. 2017 - Dec. 2017

- o Proposed a label-aware double transfer learning framework for named entity recognition and applied in the medical domain.
- o Introduced a label-aware assumption which is critical in real-world named entity recognition systems.
- o Proved the equivalence of the L2 distance in parameter space and the KL-divergence in model output distributions.
- o Accepted as a long oral paper at **NAACL HLT 2018**.

### Multi-Task Learning in Gradient Boosting Decision Tree

SJTU, Shanghai

Guide: Prof. Weinan Zhang, Prof. Yong Yu, CSD, SJTU

May. 2017 - Sep. 2018

- o Proposed and analyzed the *negative task gain* problem of gradient boosting decision tree (GBDT) in a multi-task learning scenario.
- o Proposed a multi-task GBDT model and improved the performance of diabetes prediction in data collected from 21 different centers in China.
- o In submission.

### Text-driven Graph Embedding with Pairs Sampling

SJTU, Shanghai

Guide: Prof. Weinan Zhang, Prof. Yong Yu, CSD, SJTU

Nov. 2017 - Sept. 2018

- o Proposed a new explanation of Random Walk from the perspective of neighborhood joint probability.
- o Proposed an efficient sampling policy which reduces more than 99.9% training pairs compared with Random Walk.
- o Proposed an inductive graph embedding model to make full use of textual information on graphs.
- o Accepted to **WWW 2019**.

### Automated ICD Coding

Synyi LLC, Shanghai

Guide: Prof. Weinan Zhang, CSD, SJTU and Dr. Shaodian Zhang, Synyi LLC

Nov. 2017 - May. 2018

- o Worked as a research intern in Synyi LLC and focused on NLP techniques for automated ICD coding.
- o Cleaned massive raw medical records data from hospital.
- o Implemented a state-of-the-art multi-label document classification model, hierarchical attention networks, for automated ICD coding.