

Zhenghui Wang

GHC 7413 – Carnegie Mellon University – Pittsburgh, USA

✉ felixwzh@outlook.com • 🌐 zhenghuiwang.net

Education

Carnegie Mellon University

Pittsburgh, USA

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

Jul. 2018 - Jan. 2019

Shanghai Jiao Tong University

Shanghai, China

- Senior undergraduate, Dept. of Computer Science
- Zhiyuan Honors Program in Engineering (an elite program for top 5% talented students)
- GPA: Overall: 90.84/100 (Ranking: 10/142) | Major: 92.70/100 | Final Year: 93.76/100
- Standard Test: TOEFL: 102 (R28, L27, S22, W25), GRE: V151, Q168, W4.0
- Advisors: Prof. [Weinan Zhang](#) and Prof. [Yong Yu](#)

Sep. 2015 - Jun. 2019 (Expected)

Research Interests

Machine Learning, Natural Language Processing, Data Mining.

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%).

Gradient Tree Boosting Based Multi-Task Learning for Diabetes Prediction

- [Z. Wang](#), Y. Qu, G. Sui, J. Shen, W. Zhang, Z. Zhao, G. Ning, and Y. Yu.
- In submission to *the 30th Web Conference. WWW 2019*.

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 submission to *the 30th Web Conference. WWW 2019*.

Honors and Awards

- **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
- **EIC Education Scholarship** (Top 2% in CS Department) 2017
- **Yitu Tech Scholarship** (Top 3% in CS Department) 2017
- **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
- **Member of SJTU RONG CHANG CHUCAI Program** (Top 2% in SJTU) since 2016
- **Member of SJTU Zhiyuan Honors Program in Engineering** (Top 5% in SJTU) since 2016

Research Experiences

Interpretable Prediction of Gene Sequence

CMU, Pittsburgh

Guide: Prof. Jian Ma, SCS, CMU

Jul. 2018 - present

- Incorporated advanced NLP techniques (e.g., deep contextualized word representations) into the modeling of the gene sequence.
- 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

- Proposed a label-aware double transfer learning framework for named entity recognition and applied in the medical domain.
- Introduced a label-aware assumption which is critical in real-world named entity recognition systems.
- Proved the equivalence of the L2 distance in parameter space and the KL-divergence in model output distributions.
- 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

- Proposed and analyzed the *negative task gain* problem of gradient boosting decision tree (GBDT) in a multi-task learning scenario.
- Proposed a multi-task GBDT model and improved the performance of diabetes prediction in data collected from 21 different centers in China.
- Applied in product and served more than 10 thousands people.
- Submitted to **WWW 2019**.

Text-driven Graph Embedding with Pairs Sampling

SJTU, Shanghai

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

Nov. 2017 - Sept. 2018

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

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

Power Consumer Profiling

SJTU, Shanghai

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

Nov. 2016 - Jan. 2017

- Cleaned and Inferred on a huge power consumer dataset with millions of records.
- Implemented state-of-the-art models to predict whether a power consumer is sensitive to the power price.

Selected Academic Presentations

- [Label-Aware Double Transfer Learning for NER](#)
- [A Survey on Automated ICD Coding](#)

NAACL HLT 2018, New Orleans
Apex Lab, SJTU

Open Source Projects

- [Implementation of Label-aware Double Transfer Learning for NER](#)
- [Implementation of Multi-Task Gradient Boosting Decision Tree](#)

External Reviewer

- The 11th ACM International Conference on Web Search and Data Mining WSDM 2018
- The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval SIGIR 2018
- The 17th International Semantic Web Conference ISWC 2018
- The 7th CCF International Conference on Natural Language Processing and Chinese Computing NLPC 2018