Yaqing Wang

CONTACT INFORMATION

School of Electrical and Computer Engineering
Purdue University
E-mail: wang5075@purdue.edu
Mobile: +1-858-766-0845

*Homepage:*https://yaqingwang.github.io/
*Google Scholar:*Link
*LinkdedIn:*https://tinyurl.com/3bv26p22
*Github:*https://github.com/yaqingwang

RESEARCH INTERESTS

My research interests lie at the intersection of data mining, natural language processing, and machine learning with focus on ***minimally-supervised learning***. My research goal is to develop systems with human learning abilities to grasp new concepts from only a few examples and quickly adapt to unforeseen circumstances. Toward this goal, my research projects are formed as ***few-shot learning*** and ***domain adaption*** on various research tasks.

- **Few-shot Learning:** Prompt Learning, Self-training, Weakly-supervised Learning, Data Augmentation, Self-supervised Learning
- **Domain Adaption:** Adversarial Learning, Meta-learning, Transfer Learning
- **Research Tasks:** Natural Language Understanding, Knowledge Extraction and Validation, Fake News Detection, Disease and Risk Prediction, Activity Recognition
- **Research Domains:** Data Science, Natural Language Processing, Social Science, Healthcare, Bioinformatics, Education, Sensing Systems and E-commerce

EDUCATION

Purdue University, IN, USA — May 2022 (Expected)
Ph.D. in Electrical and Computer Engineering
Thesis: Be More with Less: Scaling Deep Learning with Minimal Supervision
Advisor: Jing Gao

University of California San Diego, CA, USA — May 2016
M.Sc. in Statistics

Shandong University, Shandong, China — May 2013
B.Sc. in Mathematics

PROFESSIONAL EXPERIENCE (INDUSTRY)

Part-time Researcher at Microsoft Research (MSR), Redmond, WA — Oct 2021 - Present
- **Project I: Augmented Self-training with Few Labels**
  - Working on few-shot learning with data augmentation upon large-scale language models.
- **Project II: Self-supervised Learning with Multimodal Data**
  - Working on large-scale model multi-task pre-training with multi-modal data.

Research Intern at Microsoft Research (MSR), Redmond, WA — May 2021 - Aug 2021
- **Mentor:** Subho Mukherjee, Xiaodong Liu, Ahmed H. Awadallah, Jianfeng Gao
- **Project: Lite Self-training for Few-shot Learning**
  - Worked on the project of natural language understanding with limited labels.
  - **Publication:** One preprint [1] under review

Research Intern at Microsoft Research (MSR), Redmond, WA — May 2020 - Aug 2020
- **Mentor:** Subho Mukherjee, Ahmed H. Awadallah.
- **Project: Few-shot Sequence Labeling**
  - Worked on few-shot sequence labeling task which aims to identify and categorize spans of text into a pre-defined set of classes with limited labeled data.
  - **Publication:** One paper [5] published at *KDD 2021* and shipped in Outlook service.

Applied Scientist Intern at Amazon, Seattle, WA — May 2019 - Jan 2020
- **Mentor:** Xin Luna Dong, Xian Li and Yifan Ethan Xu.
- **Project: Knowledge Validation in E-commerce**

- Worked on identifying correctness of product textual attribute phrase from millions of product types.
- **Publication:** Two papers [16, 17] published at *KDD 2020* and deployed in Product Graph.

Professional Experience (Academia)

Research Assistant at University at Buffalo — Aug 2016 - Dec 2020
Research Assistant at Purdue University — Dec 2020 - Present

- **Advisor:** Jing Gao
- **Project I: Few-shot Learning for Natural Language Understanding**
  - **Why important:** One of major bottleneck for deep learning is its heavy reliance on a large-scale labeled data.
  - **Approaches:** Improving supervision signals [9], incorporating unlabeled data [5] and developing a lightweight self-training method [1] with a small set of task-specific parameters for easy deployment.
- **Project II: Rapid Domain Adaptation to Changing World**
  - **Why important:** Significant performance drop of deep learning models happens when meeting domain shift, which refers to changes in the data distribution between training dataset and dataset models encounter when deployed.
  - **Approaches:** Evolving explorations in adversarial transfer learning [29], learning with weak supervision [18] and quick adaption with few-labels [6] as well as unlabeled data [16].

Publication Summary

As of October 2021: I have published **30+ Papers**, among which I first-authored **12 Papers** (**4 KDD, 1 AAAI, 1 EMNLP finding, 2 ICDM, 1 CIKM, 1 SDM, 2 Preprint**). These papers have received **730+ Citations**. More up-to-date information could be found in Google Scholar Profile.
Several research works [5, 16, 17, 18] have been successfully shipped and deployed in a wide range of applications to make real-life impact.

Preprints And Submissions (* equal contribution)

[1] *LiST: Lite Self-training Makes Efficient Few-shot Learners*
Yaqing Wang, Subhabrata Mukherjee, Xiaodong Liu, Jing Gao, Ahmed Hassan Awadallah and Jianfeng Gao.
https://arxiv.org/abs/2110.06274

[2] *Decomposed adversarial learned inference*.
Hanbo Li*, **Yaqing Wang***, Changyou Chen, Jing Gao. (* Equal Contribution)
https://arxiv.org/abs/2004.10267

[3] *FedSemi: An Adaptive Federated Semi-Supervised Learning Framework*
Zewei Long, Liwei Che, **Yaqing Wang**, Muchao Ye, Junyu Luo, Jinze Wu, Houping Xiao and Fenglong Ma.
https://arxiv.org/abs/2012.03292

[4] *A Benchmark Dataset for Understandable Medical Language Translation*
Junyu Luo, Zifei Zheng, Hanzhong Ye, Muchao Ye, **Yaqing Wang**, Quanzeng You, Cao Xiao and Fenglong Ma.
https://arxiv.org/abs/2012.02420

Peer-Reviewed Conference papers (* equal contribution)

[5] *Meta Self-training for Few-shot Neural Seuqence Labeling*
Yaqing Wang, Subhabrata Mukherjee, Haoda Chu, Yuancheng Tu, Ming Wu, Jing Gao, Ahmed Hassan Awadallah.
Proceedings of 2021 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (**KDD 2021**), August 2021
Shipped in Outlook Natural Language Email Search, Microsoft

[6] *Multimodal Emergent Fake News Detection via Meta Neural Process Networks*
Yaqing Wang, Fenglong Ma, Haoyu Wang, Kishlay Jha and Jing Gao.
Proceedings of 2021 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (**KDD 2021**), August 2021

[7] *MedRetriever: Target-Driven Health Risk Prediction via Retrieving Unstructured Medical Text*
Muchao Ye, Suhan Cui, **Yaqing Wang**, Junyu Luo, Cao Xiao and Fenglong Ma.
Proceedings of the 30th ACM International Conference on Information and Knowledge Management (**CIKM 2021**), November 2021

[8] *A Lightweight Knowledge Graph Embedding Framework for Efficient Inference and Storage*
Haoyu Wang, **Yaqing Wang**, Defu Lian and Jing Gao.
Proceedings of the 30th ACM International Conference on Information and Knowledge Management (**CIKM 2021**), November 2021

[9] *Learning from Language: Low-shot Named Entity Recognition via Decomposed Framework*
Yaqing Wang, Haoda Chu, Chao Zhang, Jing Gao.
Findings of the 2021 Conference on Empirical Methods in Natural Language Processing, (**EMNLP Findings**), November 2021.

[10] *Knowledge-Guided Paraphrase Identification*
Haoyu Wang, Fenglong Ma, **Yaqing Wang** and Jing Gao.
Findings of the 2021 Conference on Empirical Methods in Natural Language Processing, (**EMNLP Findings**), November 2021.

[11] *MedPath: Augmenting Health Risk Prediction via Medical Knowledge Paths.*
Muchao Ye, Suhan Cui, **Yaqing Wang**, Junyu Luo, Cao Xiao and Fenglong Ma.
Proceedings of the 30th The Web Conference (**WWW 2021**), April, 2021.

[12] *Fair Classification Under Strict Unawareness.*
Haoyu Wang, Hengtong Zhang, **Yaqing Wang** and Jing Gao.
Proceedings of the SIAM International Conference on Data Mining (**SDM 2021**), March, 2021.

[13] *Towards Learning Outcome Prediction via Modeling Question Explanations and Student Responses.*
Tianqi Wang, Fenglong Ma, **Yaqing Wang**, Tang Tang, Longfei Zhang, and Jing Gao.
Proceedings of the SIAM International Conference on Data Mining (**SDM 2021**), March 2021.

[14] *InterHG: an Interpretable and Accurate Model for Hypothesis Generation.*
Haoyu Wang, Xuan Wang, **Yaqing Wang**, Guangxu Xun, Kishlay Jha, and Jing Gao.
Proceedings of the 2021 IEEE International Conference on Bioinformatics and Biomedicine (**BIBM 2021**)., Dec. 2021

[15] *FedTriNet: A Pseudo Labeling Method with Three Players for Federated Semi-supervised Learning.*
Liwei Che, Zewei Long, Jiaqi Wang, **Yaqing Wang**, Houping Xiao, Fenglong Ma.
Proceedings of the 2021 IEEE International Conference on Big Data (**BigData 2021**), December 15-18 2021.

[16] *Automatic Validation of Textual Attribute Values in ECommerce Catalog by Learning with Limited Labeled Data*.
Yaqing Wang, Yifan Ethan Xu, Xian Li, Xin Luna Dong and Jing Gao.
ACM SIGKDD Conference on Knowledge Discovery and Data Mining (**KDD 2020**).
Deployed in Product Graph, Amazon

[17] *AutoKnow: Self-Driving Knowledge Collection for Products of Thousands of Types*.
Xin Luna Dong, Xiang He, Andrey Kan, Xian Li, Yan Liang, Jun Ma, Yifan Ethan Xu, Chenwei Zhang, Tong Zhao, Gabriel Blanco Saldana, Saurabh Deshpande, Alexandre Michetti Manduca, Jay Ren, Surender Pal Singh, Fan Xiao, Haw-Shiuan Chang, Giannis Karamanolakis, Yuning Mao, **Yaqing Wang**, Christos Faloutsos, Andrew McCallum, Jiawei Han.
ACM SIGKDD Conference on Knowledge Discovery and Data Mining (**KDD 2020**).
Deployed in Product Graph, Amazon

[18] *Weak Supervision for Fake News Detection via Reinforcement Learning*.
Yaqing Wang, Weifeng Yang, Fenglong Ma, Jin Xu, Bin Zhong, Qiang Deng, Jing Gao.
Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence (**AAAI 2020**), New York, Feb 2020.
Deployed in WeChat, Tencent

[19] *Efficient Knowledge Graph Validation via Cross-Graph Representation Learning*.
Yaqing Wang, Fenglong Ma, Jing Gao.
Proceedings of the 29th ACM International Conference on Information and Knowledge Management (**CIKM 2020**).

[20] *Rare Disease Prediction by Generating Quality-Assured Electronic Health Records*.
Fenglong Ma*, **Yaqing Wang***, Jing Gao, Houping Xiao, Jing Zhou.
Proceedings of the SIAM International Conference on Data Mining (**SDM 2020**), Cincinnati, Ohio, May 7-9, 2020.

[21] *LP-Explain: Local Pictorial Explanation for Outliers*.
Haoyu Liu, Fenglong Ma, **Yaqing Wang**, Shibo He, Jiming Chen, Jing Gao.
Proceedings of the IEEE International Conference on Data Mining (**ICDM 20**), Sorrento, Italy, November 2020.

[22] *Hypothesis Generation From Text Based On Co-Evolution Of Biomedical Concepts*.
Kishlay Jha, Guangxu Xun, **Yaqing Wang**, Aidong Zhang.
Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (**KDD 2019**), Alaska, USA, August 2019.

[23] *A General Framework for Diagnosis Prediction via Incorporating Medical Code Descriptions*.
Fenglong Ma, **Yaqing Wang**, Houping Xiao, Ye Yuan, Radha Chitta, Jing Zhou and Jing Gao.
Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (**BIBM 18**), Madrid, Spain, December, 2018.

[24] *Multivariate Sleep Stage Classification using Hybrid Self-Attentive Deep Learning Networks.*
Ye Yuan, Fenglong Ma, Guangxu Xun, **Yaqing Wang**, Kebin Jia, Lu Su and Aidong Zhang.
Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (**BIBM 18**), Madrid, Spain, December, 2018.

[25] *Leveraging The Power of Informative Users for Local Event Detection.*
Hengtong Zhang, Fenglong Ma, Yaliang Li, Chao Zhang, Tianqi Wang, **Yaqing Wang**, Jing Gao, Lu Su.
IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (**ASONAM 18**), Barcelona, Spain, August, 2018.

[26] *Interpretable Word Embeddings For Medical Domain*.
Kishlay Jha*, **Yaqing Wang***, Guangxu Xun, and Aidong Zhang.
Proceedings of the 18th IEEE International Conference on Data Mining (**ICDM 18**), Singapore, November 2018.

[27] *MuVAN: A Multi-view Attention Network for Clinical Temporal Data*.
Ye Yuan, Guangxu Xun, Fenglong Ma, **Yaqing Wang**, Nan Du, Kebin Jia, Lu Su and Aidong Zhang.
IEEE International Conference on Data Mining (**ICDM 18**), Singapore, November 2018.

[28] *Towards Environment Independent Device Free Human Activity Recognition*.
Wenjun Jiang, Chenglin Miao, Fenglong Ma, Shuochao Yao, **Yaqing Wang**, Ye Yuan, Hongfei Xue, Chen Song, Xin Ma, Dimitrios Koutsonikolas, Wenyao Xu, and Lu Su.
The 24th Annual International Conference on Mobile Computing and Networking (**MobiCom 2018**), New Delhi, India, October 29-November 2, 2018.

[29] *EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection*.
Yaqing Wang, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su and Jing Gao.
Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (**KDD 2018**), London, United Kingdom, August, 2018.
Over 300 citations since 2018, ranked as the 5th most influential paper at KDD 2018 by Paperdiges

[30] *Concepts-Bridges: Uncovering Conceptual Bridges Based on Biomedical Concept Evolution*.
Kishlay Jha, Guangxu Xun, **Yaqing Wang**, Vishrawas Gopalakrishnan, Aidong Zhang.
Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (**KDD 2018**), London, United Kingdom, August, 2018.

[31] *Discovering Truths from Distributed Data*.
Yaqing Wang, Fenglong Ma, Lu Su, Jing Gao.
IEEE International Conference on Data Mining (**ICDM 2017**), New Orleans, USA, November 2017.

PEER-REVIEWED JOURNAL PAPERS

[32] *MeSHProbeNet: A Self-attentive Probe Net for MeSH Indexing*.
Guangxu Xun, Kishlay Jha, Ye Yuan, **Yaqing Wang** and Aidong Zhang.
Bioinformatics, Oxford University Press, 2019

[33] *Incorporating Medical Code Descriptions for Diagnosis Prediction in Healthcare*.
Fenglong Ma, **Yaqing Wang**, Houping Xiao, Ye Yuan, Radha Chitta, Jing Zhou, Jing Gao.
BMC bioinformatics 19.

[34] *A Hybrid Self-attention Deep Learning Framework for Multivariate Sleep Stage Classification*.
Yuan Ye, Kishlay Jha, Fenglong Ma, Guangxu Xun,, **Yaqing Wang**, Lu Su, Aidong Zhang.
BMC bioinformatics 20.

PROFESSIONAL SERVICE

Program Committee

- SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) 2020, 2021
- International Conference on Learning Representations (ICLR) 2021, 2022
- International Conference on Machine Learning (ICML) 2020, 2021
- Neural Information Processing Systems (NeurIPS) 2021
- AAAI Conference on Artificial Intelligence (AAAI) 2021
- TheWebConf (WWW) 2022
- ACL Rolling Review (ARR) 2022
- Annual Meeting of the Association for Computational Linguistics (ACL) 2021, 2022
- North American Chapter of the Association for Computational Linguistics (NAACL) 2021
- Conference on Empirical Methods in Natural Language Processing (EMNLP) 2021
- SIAM International Conference on Data Mining (SDM) 2022

Journal Reviewer

- The International Journal on Very Large Data Bases (VLDB Journal)
- ACM Computing Surveys (CSUR)
- WIREs Data Mining and Knowledge Discovery
- ACM Transactions on Knowledge Discovery from Data (TKDD)
- IEEE Transactions on Multimedia
- World Wide Web Journal
- IEEE Access

Conference Volunteer

- The 26th SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2020
- The 24th SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2018
- IEEE International Conference on Data Mining(ICDM), 2017.

HONORS AND AWARDS

- Bilsland Dissertation Fellowship, *Purdue University* 2022
- Student Registration Award, *KDD, Virtual* 2021
- Student Registration Award, *CIKM, Virtual* 2020
- Student Registration Award, *KDD, Virtual* 2020
- Student Travel Award, *AAAI, New York, USA* 2020
- Student Travel Award, *KDD, London, UK* 2018
- Best Poster Award, *UB CSE Graduate Research Conference* 2017
- Student Travel Award, *ICDM, New Orleans, USA* 2017
- Presidential Fellowship, *SUNY Buffalo* 2016-2020

PROGRAMMING SKILLS

- Python (PyTorch, TensorFlow), Jave, C++, C, R, LaTeX, MATLAB, SQL