Zihan Wang

PH.D. STUDENT @ CS, UC, SAN DIEGO

☎ (+1) 949-299-8265 | ✉ ziw224@ucsd.edu | ⌂ zihanwangki.github.io

RESEARCH INTERESTS

Natural Language Processing • Weak Supervision • Language Modeling

My research aims to address NLP tasks with (extremely) weak supervision. I strive to answer the following questions.

- What knowledge do pre-trained neural language models encode? Is the current fine-tuning or prompting paradigm the best to distill the right part of these knowledge to domain-specific tasks?

- What is the best forms of getting supervision from human? How to inform machines about the NLP task goals with minimum human effort?

- Can machines go beyond extracting information to discovering novel information to expand human knowledge?

EDUCATION

University of California, San Diego

PH.D. IN COMPUTER SCIENCE   *Oct 2020 - Present*

Advised by Prof. Jingbo Shang; GPA 4.0

University of Illinois, Urbana-Champaign

B.S. IN COMPUTER SCIENCE   *Sep 2016 - May 2020*

Graduated with Highest Honor & Bronze Tablet; Major GPA 4.0, Overall GPA 3.92

PROFESSIONAL EXPERIENCE

Microsoft Research, Language and Information Technologies   *Remote*

RESEARCH INTERN   *June 2022 – Sep 2022*

Working with Dr. Guoqing Zheng, Dr. Yu Cheng and Dr. Ahmed H. Awadallah on efficient fine-tuning for large language models.

Adobe, Document Intelligence Lab   *Remote*

RESEARCH INTERN   *June 2021 – Sep 2021*

Worked with Dr. Tong Sun and Dr. Jiuxiang Gu on efficient transformers and proposed an adaptive axis attention pattern.

University of Pennsylvania   *Philadelphia, PA, USA*

RESEARCH ASSISTANT   *May 2019 – Aug 2019*

Worked with Prof. Dan Roth on low-resource, cross-lingual named entity recognition. We also analyzed the effectiveness of Multilingual BERT.

University of Illinois at Urbana-Champaign   *Urbana, IL, USA*

RESEARCH ASSISTANT   *Dec 2018 – May 2019*

Worked with Prof. Jiawei Han on TextCubes, Topic Extraction, and Representation Learning.

Facebook   *Melon Park, CA, USA*

SOFTWARE ENGINEER INTERN   *May 2018 – Aug 2018*

Worked on front-end and back-end for new features for Facebook Groups in the Facebook App.

INVITED TALKS

[2] University of Colorado Boulder, Online, Jan 2021

[1] Google N2Formal, Online, Jan 2022

SELECTED HONORS & AWARDS

2021	Bronze Medal, ICPC World Finals at Moscow, 2nd in North America, 11/118 worldwide
2021	Bronze Medal, ICPC North American Championship
2021	**Champion**, ICPC North American Division Championship (West) all schools in the west coast, including Stanford, UC Berkeley, University of Washington
2021 – 2022	UCSD Jacob School of Engineering Fellowship
2020 – 2023	UCSD Halıcıoğlu Data Science Fellowship
2019	**World Champion**, IEEE Xtreme 13.0, among 2781 teams worldwide
2019	Champion, ICPC Mid-Central Regional
2019	Champion, C1 Terminal@UIUC
2019	Champion, UIUC Undergrad Math Contest
2016 – 2019	UIUC Dean's List
2018	Runner-up, Illinois Tech Association (ITA) Tech Challenge
2018	UIUC Jeffry P. Blahut Memorial Scholarship

Papers & Pre-prints

[13] Waveformer: Linear-Time Attention with Forward and Backward Wavelet Transform
Yufan Zhuang, Zihan Wang, Fangbo Tao, Jingbo Shang
arXiv preprint arXiv:2210.01989, 2022.

[12] WeDef: Weakly Supervised Backdoor Defense for Text Classification
Lesheng Jin*, Zihan Wang*, Jingbo Shang
In Proc. of *ACL SIGDAT Empirical Methods in Natural Language Processing (**EMNLP**), 2022.*

[11] Formulating Few-shot Fine-tuning Towards Language Model Pre-training: A Pilot Study on Named Entity Recognition
Zihan Wang, Kewen Zhao, Zilong Wang, Jingbo Shang
In Proc. of *ACL SIGDAT Empirical Methods in Natural Language Processing (**EMNLP**) (Findings), 2022.*

[10] Learning Adaptive Axis Attentions in Fine-tuning: Beyond Fixed Sparse Attention Patterns
Zihan Wang, Jiuxiang Gu, Jason Kuen, Handong Zhao, Vlad I Morariu, Ruiyi Zhang, Ani Nenkova, Tong Sun and Jingbo Shang
In Proc. of *60th Annual Meeting of the Association for Computational Linguistics (**ACL**) (Findings), 2022.*

[9] "Average" Approximates "First Principal Component"? An Empirical Analysis on Representations from Neural Language Models
Zihan Wang, Chengyu Dong, and Jingbo Shang
In Proc. of *ACL SIGDAT Empirical Methods in Natural Language Processing (**EMNLP**), 2021.*

[8] UCPhrase: Unsupervised Context-aware Quality Phrase Tagging
Xiaotao Gu*, Zihan Wang*, Zhenyu Bi, Yu Meng, Liyuan Liu, Jiawei Han, Jingbo Shang (* **Equal Contribution**)
In Proc. of *ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (**KDD**), 2021.*

[7] X-Class: Text Classification with Extremely Weak Supervision
Zihan Wang, Dheeraj Mekala, and Jingbo Shang
In Proc. of *2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (**NAACL**), 2021.*

[6] Extending Multilingual BERT to Low-Resource Languages
Zihan Wang*, Karthikeyan K*, Stephen Mayhew, Dan Roth
In Proc. of *ACL SIGDAT Empirical Methods in Natural Language Processing (**EMNLP**) (Findings), 2020.*

[5] CatE: Category-Name Guided Word Embedding
Yu Meng, Jiaxin Huang, Guangyuan Wang, Zihan Wang, Chao Zhang, Jiawei Han
In Proc. of *Int. World Wide Web Conference (**WWW**), 2020.*

[4] Cross-Lingual Ability of Multilingual BERT: An Empirical Study
Karthikeyan K*, <u>Zihan Wang</u>*, Stephen Mayhew, Dan Roth (∗ **Equal Contribution**)
In Proc. of *International Conference on Learning Representations (**ICLR**), 2020.*

[3] Raw-to-End Name Entity Recognition in Social Media
Liyuan Liu, <u>Zihan Wang</u>, Jingbo Shang, Dandong Yin, Heng Ji, Xiang Ren, Shaowen Wang, Jiawei Han
arXiv preprint arXiv:1908.05344, 2019.

[2] TopicMine: User-Guided Topic Mining by Category-Oriented Embedding
Yu Meng, Jiaxin Huang, <u>Zihan Wang</u>, Chenyu Fan, Guangyuan Wang, Chao Zhang, Jingbo Shang, Lance Kaplan, Jiawei Han
In Proc. of *ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (**KDD**) (demo), 2019.*

[1] CrossWeigh: Training Named Entity Tagger from Imperfect Annotations
<u>Zihan Wang</u>*, Jingbo Shang*, Liyuan Liu*, Lihao Lu, Jiacheng Liu and Jiawei Han (∗ **Equal Contribution**)
In Proc. of *ACL SIGDAT Empirical Methods in Natural Language Processing (**EMNLP**), 2019.*

LEADERSHIP & TEACHING

Co-Coach *UCSD*
UCSD ICPC CLUB *2020–Present*
Coaching and teaching students at UCSD for algorithmic programming contests.

Guest Lecturer *UCSD*
UCSD CSE 109 *Spring 2022, Spring 2021*
Giving invited lectures on algorithmic programming contests.

PROFESSIONAL SERVICE

Reviewer: EMNLP (2021), NeurIPS CtrlGen Workshop (2021), EMNLP (2022), COLING (2022), AAAI (2023)

Volunteer: KDD (2021)

Student Organizer: SoCal ML & NLP Symposium (2021)

TECHNICAL SKILLS

Skills: Machine Learning, Natural Language Processing, Language Modeling, Weakly Supervised Learning, Unsupervised Learning, Information Retrieval, Software engineering, Android Development, Full-stack Development

Programming Languages: Python, Java, C++, Javascript, C, PHP

Frameworks: Tensorflow, Jax/Flax, Pytorch, Kubernetes