Zhong, Wei

Chinese name: 钟 威

w32zhong@uwaterloo.ca
(CAN) +01 226-600-8357
github.com/w32zhong
w32zhong.github.io

Summary

My interests are at the intersections of information retrieval and deep learning. I have created state-of-the-art structure search engine and neural retrievers for math language. I also wrote most backend, frontend, and ops&infra code for my online math search engine demo: `https://approach0.xyz`.

I plan to graduate at Fall 2023, and I am currently actively looking for internship or full-time job opportunities.

Education

University of Waterloo 2021 Apr. – Present
Ph.D. Candidate in Computer Science (Cum GPA: 89.25 / 100) *Waterloo, ON, CA*
(Advisor: Jimmy Lin)

Rochester Institute of Technology 2017 Aug. – 2019 Jul. (Transferred Out)
Ph.D. Candidate in Computer Science (Cum GPA: 3.780 / 4.0) *Rochester, NY, USA*
(Advisor: Richard Zanibbi)

University of Delaware 2013 Aug. – 2015 Aug.
M.S. in Electrical and Computer Engineering (Cum GPA: 3.867 / 4.0) *Newark, DE, USA*
(Advisor: Hui Fang)

China Ji Liang University (CJLU) 2009 Aug. – 2013 Jun.
B.S. in Information and Computation Science *Hangzhou, P.R. China*

Selected Publications

2022 CLEF (Best Paper) **Wei Zhong**, Yuqing Xie, Jimmy Lin *Applying Structural and Dense Semantic Matching for the ARQMath Lab 2022, CLEF.*

2022 EMNLP Findings (1st Math NLP workshop) **Wei Zhong**, Jheng-Hong Yang, Yuqing Xie and Jimmy Lin *Evaluating token-level and passage-level dense retrieval models for math information retrieval.*

2021 CLEF (#1 Formula Search System) **Wei Zhong**, Xinyu Zhang, Ji Xin, Richard Zanibbi, Jimmy Lin *Approach zero and anserini at the CLEF-2021 arqmath track: Applying substructure search and BM25 on operator tree path tokens.*

2021 SIGIR **Wei Zhong**, Jimmy Lin *PyAO: A Python Toolkit for Accessible Math-Aware Search.*

2020 ECIR **Wei Zhong**, Shaurya Rohatgi, Jian Wu, C. Lee Giles and Richard Zanibbi. *Accelerating Substructure Similarity Search for Formula Retrieval.*

2020 ECIR Gavin Nishizawa, Jennifer Liu, Yancarlos Diaz, Abishai Dmello, **Wei Zhong** and Richard Zanibbi. *MathSeer: A Math-Aware Search Interface with Intuitive Formula Editing, Reuse, and Lookup.*

2019 ECIR (Best Application Paper) **Wei Zhong** and Richard Zanibbi. *Structural Similarity Search for Formulas using Leaf-Root Paths in Operator Subtrees.*

Experience

University of Waterloo Apr. 2021 – Present
Research Assistant and Teaching Assistant *Waterloo, ON, Canada*

- Proposed new pretraining scheme and datasets for Math IR neural retrievers.
- Created a neural math-aware search engine, currently the state-of-the-art in two recent math IR datasets (CLEF ARQMath-2 and ARQMath-3). Won one best paper award at CLEF which is a major IR evaluation forum.
- Developed Domain-adapted DPR, ColBERT, CoCondenser, MAE models for Math IR.

DMAI Apr. 2020 – Sep. 2020
NLP Researcher (internship) *Guangzhou, P.R. China*

- Math expression simplifier using lock-free MCTS and CRF.
- Tag prediction using BERT, LDA, Mutual Information and SVM.
- Agent decision making using GBDT+LR and Reinforcement Learning.

Rochester Institute of Technology Aug. 2017 – Aug. 2019
Research Assistant *Rochester, NY, USA*

- Created a state-of-the-art structure search engine for math, won one best paper award at ECIR (European top IR conference).
- Discontinued and transfered to the University of Waterloo due to U.S. visa issue during the COVID-19 pandemic.

Huawei Technologies Sep. 2016 – July. 2017
Software developer *Shenzhen, P.R. China*

- STB (TV Box) Hardware Abstraction Layer code maintenance.
- Participated Peach Fuzzing testing for Android-based system interface

SevOne (2015 Glassdoor best places to work) Jun. 2015 – Aug. 2015
WEB backend intern (S.M.A.R.T.S program) *Wilmington DE, US*

- PHP, C/C++, MySQL code maintenance
- Search engine back-end rewriting using CLucene

Selected Projects

Approach Zero Math-aware search engine 2015 – Present

- Ranked top one in the year 2020 and 2021 Community Promotion Ads of Math StackExchange, one of the largest math Q&A community.
- Obtained the state-of-the-art scores in the NTCIR-12, CLEF-2021 and CLEF-2022 Tasks.
- Online version available, searching millions of posts in real-time using 5 low-end Linode instances.

PyAo Evaluation toolkit for various math IR systems and neural retrievers 2021 – Present

- Created effective training data for math IR from scratch.
- Implemented DPR, ColBERT, CoCondenser, MAE models in one code framework.
- Developed easy-to-use pipelines to cover pretraining, domain-adaptation, fine-tuning, inference and evaluation of math IR models (using BERT backbone).

TinyNN and MNN Lightweight (educational) deep learning libraries Sep. – Oct. 2019

- My contributions include denoising autoencoder, Restricted Boltzmann Machine (w/ CD-k training) and activation maximization (AM) visualization.
- Refactored its CNN convolutional layer (the *im2col* function).
- Hand-written gradient/Jacobian matrix derivation and GPU acceleration using CuPy.

Mathsteps-v2 A step-by-step math solver June. 2020

- I designed a declarative macro language for representing math transformations.
- Efficient lock-free MCTS math solver in C language, search space is reduced by a policy network.

Skill Set

Software Linux/Shell, Git, Web technologies, C/C++, Python

Hardware Embedded system design, VHDL (See my 8bits TTL CPU in Multisim and Flappy Bird game in VHDL!)

Teaching Skill Primer teaching skills (Instructed one algorithm colloquium on Dantzig's simplex algorithm, attended credited Teaching Skills Workshop in Rochester Institute of Technology)