

ZHE YU

Department of Software Engineering, RIT, Rochester, NY, 14623

(919)-949-1382 ♦ <https://zhe-yu.github.io> ♦ zxyvse@rit.edu

I believe the future of AI is not replacing humans, but, rather, better supporting humans with automated intelligences. Hence, my research focuses on the creation of “human in the loop” machine learning environments.

RESEARCH INTEREST

Software Engineering, Machine Learning, Information Retrieval, Human-Computer Interaction

EDUCATION

PhD in Computer Science Aug 2015 - May 2020

North Carolina State University

MS in Control Science and Engineering Sep 2011 - Mar 2014

Shanghai Jiao Tong University

BS in Automation Sep 2007 - May 2011

Shanghai Jiao Tong University

EMPLOYMENT

Assistant Professor Aug 2020 - Present

Department of Software Engineering Rochester Institute of Technology, Rochester, NY

Graduate Research Assistant Aug 2015 - May 2020

Department of Computer Science North Carolina State University, Raleigh, NC

Summer Intern May 2019 - Aug 2019

Traffic Estimation for Ads Google, Los Angeles, CA

Summer Intern May 2018 - Aug 2018

Data Engine Machine Learning Google, Mountain View, CA

Summer Intern May 2017 - Aug 2017

LexisNexis, Raleigh, NC

Summer Intern May 2016 - Aug 2016

LexisNexis, Raleigh, NC

SERVICE

Conference Committee Member: MSR 2021, SCAM 2020 NIER track

Conference Review: CSCW 2020

Journal Review: IEEE Transactions on Software Engineering (TSE), ACM Transactions on Software Engineering and Methodology (TOSEM), Empirical Software Engineering (EMSE), Information and Software Technology (IST), IEEE Access

TEACHING

DSCI-633: Foundation of Data Science Aug 2020 - Present

RECENT RESEARCH PROJECTS

Few-shot Active Learning for Better Information Retrieval Aug 2020 - Present

Rochester Institute of Technology

- This project aims to improve efficiency of information retrieval with a combination of few-shot learning and active learning.
- When a deep neural network model learns to retrieve different relevant information (of the same type, e.g. literature reviews of different topics), it optimizes its inner structure so that it will learn faster and better to retrieve new, unseen information.
- Active learning is also important in retrieving new, unseen information by (1) continuously learning from human decisions on which is relevant and (2) suggesting what should be reviewed next based on the learned model.

Better self-admitted technical debt identification with adaptive CNN. Aug 2020 - Present

Rochester Institute of Technology

- Incorporate a CNN model in the framework of a human-in-the-loop self-admitted technical debt identification system.
- Explore ways to efficiently update the CNN model with incrementally obtained labels.
- Make the prediction model of CNN adaptive to the target software project.

Human Ethical Bias Detection with Machine Learning Software

Aug 2020 - Present

Rochester Institute of Technology

- Instead of trying to reduce bias in machine learning software, this project aims to reduce bias from its source—the human decisions.
- That is, this work proposes to utilize the machine learning bias inherited from the training data (human decisions) as an indicator for detecting human bias.
- While it is difficult to directly test whether a human has bias, with current research on machine learning bias, it is now easy to test, in large scale at low cost, whether a machine learning software has bias.

Trace Link Prediction with Dual-Encoders

Aug 2021

Rochester Institute of Technology

- Trace link prediction is a crucial problem to facilitate the maintenance of software projects.
- This work proposes to customize the dual-encoder architecture for trace link prediction, especially in the inference strategy and the loss function for training.
- The designed dual-encoders will be able to encode different software artifacts into a shared encoding space. In that shared encoding space, artifacts connected by trace links would be closer to each other than the unconnected ones.

SELECTED PUBLICATIONS

-
- [1] Zhe Yu, Fahmid Morshed Fahid, Huy Tu, and Tim Menzies. "Identifying Self-Admitted Technical Debts with Jitterbug: A Two-step Approach." *IEEE Transactions on Software Engineering*.
- [2] Yang, Xueqi, Zhe Yu, Junjie Wang, and Tim Menzies. "An Expert System for Learning Software Engineering Knowledge (with Case Studies in Understanding Static Code Warning)." *Expert Systems with Applications*.
- [3] Yang, Xueqi, Jianfeng Chen, Rahul Yedida, Zhe Yu, and Tim Menzies. "How to Recognize Actionable Static Code Warnings (Using Linear SVMs)." *Empirical Software Engineering*.
- [4] Chakraborty, Joymallya, Suvodeep Majumder, Zhe Yu, and Tim Menzies. "Fairway: A Way to Build Fair ML Software." In *Proceedings of ESEC/FSE 2020*.
- [5] Agrawal, Amritanshu, Tim Menzies, Leandro L. Minku, Markus Wagner, and Zhe Yu. "Better software analytics via DUO: Data mining algorithms using/used-by optimizers." *Empirical Software Engineering* 25, no. 3 (2020): 2099-2136.
- [6] Huy Tu, Zhe Yu, Tim Menzies. 2020. "Better Data Labelling with EMBLEM (and how that Impacts Defect Prediction)." *IEEE Transactions on Software Engineering*.
- [7] Zhe Yu, Christopher Theisen, Laurie Williams, Tim Menzies. 2019. "Improving Vulnerability Inspection Efficiency Using Active Learning." *IEEE Transactions on Software Engineering*.
- [8] Zhe Yu, Fahmid M. Fahid, Tim Menzies, Gregg Rothermel, Kyle Patrick, Snehith Cherian. 2019. "TERMINATOR: Better Automated UI Test Case Prioritization." In *Proceedings of ESEC/FSE'19, Software Engineering in Practice*, 883-894. <http://doi.acm.org/10.1145/3338906.3340448>
- [9] Zhe Yu and Tim Menzies. 2019. "FAST2: An intelligent assistant for finding relevant papers." *Expert Systems with Applications*. 120: 57-71. <https://www.sciencedirect.com/science/article/pii/S0957417418307413>
- [10] Zhe Yu, Nicholas A. Kraft, and Tim Menzies. 2018. "Finding Better Active Learners for Faster Literature Reviews." *Empirical Software Engineering*. 23(6): 3161-3186. <https://link.springer.com/article/10.1007/s10664-017-9587-0>
- [11] Vivek Nair, Zhe Yu, Tim Menzies, Norbert Siegmund, and Sven Apel. 2018. "Finding faster configurations using flash." *IEEE Transactions on Software Engineering*.
- [12] Agrawal, Amritanshu, Tim Menzies, Leandro L. Minku, Markus Wagner, and Zhe Yu. "Better software analytics via DUO": Data mining algorithms using/used-by optimizers." *Empirical Software Engineering* 25, no. 3 (2020): 2099-2136.