Research Statement

Sun Jun
School of Computing and Information Systems, Singapore Management University
Tel: (65) 90269318; Email: junsun@smu.edu.sg
23 Dec 2023

Background

Our daily life is more and more dictated by computer software (hereafter 'code' or AI), being it control software in cyber physical systems (which provides water and electricity to us), business services (e.g., online banking), applications in the smart phone (e.g., Google map and Siri) or AI which is likely dominating the future. These codes or AIs are thus entrusted with powers to have enormous societal impact, which could be either good or bad or horrifying.

The critical question is whether these systems are built in a way which deserves our trust. Time and time again, it has been demonstrated such trust may be misplaced (e.g., discrimination by AI systems) or over-looked (e.g., vulnerabilities in control software systems). The goal of my research is to make the world a better place by developing theories, methods, and tools which help improve the quality of these systems so that they deserve our trust.

Research Areas

Towards the goal of developing methods and tools to improve code quality, my research has so far been focused on three main areas: formal methods, software engineering, and AI safety.

Formal methods

Formal methods are a family of methods which aim to provide formal guarantee on the safety and quality of a system. They are built on solid foundation of mathematics, e.g., first order logic, set theory, graph theory and category theory. With formal methods, programs are treated as mathematical objects, which are subject to analysis based on mathematical methods. We establish formal guarantees on a program's correctness (i.e., it will satisfy users' requirements all the time in all circumstances) in the same way mathematical theories are established (e.g., by proving lemmas and theories). Even though researchers (including many Turing award winners) have worked on formal methods for decades, formal methods are still limited to small systems/programs (and a few instances of large systems with years of dedicated effort). That is, there is still a gap between what formal methods set out to solve and what they can achieve now. One of my research goals is to close the gap (by connecting formal methods with research on software engineering, AI safety, among others; as well as by developing fully automated software toolkits).

Software engineering

Software engineering is a research area which covers many topics related to how to develop and maintain software systems. Compared to research on formal methods, research on software engineering is in general more applied. I have researched on many topics related to software engineering. What I truly enjoy working on are research problems related to program analysis and verification. The primary reason is that I love programming. My research goal is to develop systematic scalable approaches and tools which could be useful for ordinary programmers' daily job. It is my belief that experienced programmers can solve program analysis and verification problems which are impossible in theory because of their ability in learning what the program is supposed to do from various sources. Thus, to develop a tool which can solve complicated program analysis problems like the experts do, we would need state-of-art machine learning techniques.

AI safety

AI safety is a new area that I have been focused on recently. Firstly, AI safety is technically interesting! The war between connectivism (e.g., large language models) and constructivism (e.g., knowledge graphs and traditional programs) is simply inspiring. Secondly, AI safety is very much relevant these days, and will be more so in the future given the wide variety of AI applications that are being developed. Thirdly, AI safety is a fruitful and important application area of my research on formal methods as well. That is, it presents a large set of real-world systems with important safety properties (such as robustness, fairness, privacy and so on) which must be checked, and it would certainly prove the usefulness of my research if my research could solve these problems.

Selected Publications and Outputs

- Yedi Zhang, Fu Song, and Jun Sun: "QEBVerif: Quantization Error Bound Verification of Neural Networks", CAV 2023.
- Bozhi Wu, Shangqing Liu, Yang Xiao, Zhiming Li, Jun Sun, and Shang-Wei Lin: "Learning Program Semantics for Vulnerability Detection via Vulnerability-specific Inter-procedural Slicing", FSE 2023.
- Lida Zhao, Sen Chen, Zhengzi Xu, Lyuye Zhang, Jiahui Wu, Jun Sun and Yang Liu: "Software Composition Analysis for Vulnerability Detection: An Empirical Study on Java Projects", FSE 2023.
- Mengdi Zhang, Jun Sun, Jingyi Wang, and Bing Sun: "Interpretable Testing of Neural Networks Against Subtle Group Discrimination", TOSEM 2023.
- Yuan Zhou, Yang Sun, Yun Tang, Yuqi Chen, and Jun Sun: "Specification-based Autonomous Driving System Testing", TSE 2023.
- Richard Schumi, and Jun Sun: "Semantic-based Neural Network Repair", ISSTA 2023.
- Chen Yang, Junjie Chen, Xingyu Fan, Jiajun Jiang, and Jun Sun: "Silent Compiler Bug De-duplication via Three-Dimensional Analysis", ISSTA 2023.
- Xiaodong Zhang, Zhao Wei, Yang Sun, Jun Sun, Yulong Shen, Xuewen Dong, and Zijiang Yang: "Testing Automated Driving Systems by Breaking Many Laws Efficiently", ISSTA 2023.
- Christopher M. Poskitt, Yuqi Chen, Jun Sun, and Yu Jiang: "Finding Causally Different Tests for a Cyber-Physical System", ICSE 2023.

- Xiaoning Ren, Yun Lin, Yinxing Xue, Ruofan Liu, Jun Sun, Zhiyong Feng, and Jinsong Dong: “DeepArc: Modularizing Neural Networks for the Model Maintenance”, ICSE 2023.
- Shuzheng Gao, Cuiyun Gao, Chaozheng Wang, Jun Sun, David Lo, and Yue Yu: “Two Sides of the Same Coin: Exploiting the Impact of Identifiers in Neural Code Comprehension”, ICSE 2023.
- Yuhan Zhi, Xiaofei Xie, Chao Shen, Jun Sun, Xiaoyu Zhang, and Xiaohong Guan: “Seed Selection for Testing Deep Neural Networks”, TOSEM 2023.
- Jialuo Chen, Jingyi Wang, Xingjun Ma, Youcheng Sun, Jun Sun, Peixin Zhang, and Peng Cheng. “QuoTe: Quality-oriented Testing for Deep Learning Systems”, TOSEM 2023.