Research Assessment Exercise 2020
Impact Case Study

University: The Chinese University of Hong Kong

Unit of Assessment (UoA): 13 CSIT - Computer Studies/Science (incl. information Technology)

Title of case study: Automated **D**ata-driven s**O**ftware **R**eliability **E**ngineering (ADORE) Impact on IT Industry

(1) Summary of the impact (indicative maximum 100 words)

Reliability issues are a major concern for software systems, which could affect millions of users worldwide with great revenue loss and user dissatisfaction. The case study summarizes ADORE, a general framework for intelligent software reliability engineering with Quality of Service (QoS), log, and review data. The tools and datasets provided in ADORE have been downloaded by 370+ research organizations, cited 2000+ times, and applied to the reliability engineering of several IT giants. For example, the log-based toolset was successfully evaluated on real-world datasets at Microsoft and reached 1000x speed up. ADORE is also successfully deployed in Tencent and Huawei respectively.

(2) Underpinning research (indicative maximum 500 words)

ADORE aims to provide an intelligent solution to software reliability engineering, which consists of three components, QoS prediction, log management, and review mining. These components are widely adopted in various applications that dominate modern software nowadays, e.g., Web services, software systems, and mobile applications. In this framework, we designed and integrated machine learning techniques into the data-driven software reliability engineering. Our research outputs are widely recognized by both peer researchers and giant IT companies. Strong evidences also show that these research work have resulted in significant advancements in both research community and industrial deployment.

QoS Prediction: Reliability of the service-oriented systems heavily depends on the remote Web services as well as the unpredictable Internet. Designing effective and accurate reliability prediction approaches for the Web service systems has become an important research issue. To predict the quality of Web services, we designed a framework, WS-DREAM [3-1, 3-2, 3-3, 5-6], by integrating and improving advanced machine learning techniques, e.g., collaborative filtering. These techniques were implemented and deployed to the Internet for conducting real-world experiments, where we collected 1.5 million Web services invocation results from 150 service users in 24 countries. These research work are heavily cited in the research community from their publication date to present.

Log Management: In log-based reliability engineering, we use the log data generated by software systems to detect and identify severe reliability issues. To benefit both academia and industry, we collect a huge dataset from 16 different software systems, spanning distributed systems, super computers to operating systems, etc. In total, 440 million log messages that amounts to 70.8 GB are collected, which is the largest log dataset so far. Besides, we provide an intelligent log management toolset, LogPAI [3-4, 3-5], which is very comprehensive and covers the life cycle of software development, including the log compression, log parsing, anomaly detection, and problem identification. The toolset is published as an open-source toolset on GitHub for research reuse and industrial deployment.

Review Mining: Another line of work enhances the reliability with user reviews of software applications. User feedbacks can reflect apps' errors, bugs, and users concerns, which may not be easily detected before the deployment. In this framework, we crawled tens of millions of user reviews from multiple app stores, e.g., Google Play and App Store, for thousands of apps. We designed a review mining toolset, Armor [3-6], which contains a series of tools that automatically summarize helpful user reviews, extract emerging app issues, detect GUI- and functionality- related errors, and generate reasonable review response, etc. Our tools have been applied to 100+ industrial apps of Tencent, serving billions of users worldwide.

In the ADORE project, the research can be traced back to 2010, while all the impacts span from 2012 to present. The work included in this framework are mainly done by Dr. Zibin Zheng (CUHK), Dr. Jieming Zhu (CUHK), Dr. Pinjia He (CUHK), Dr. Cuiyun Gao (CUHK), and Mr. Shilin He (Ph.D. at CUHK), under the direction of Prof. Michael R. Lyu (CUHK).

(3) References to the research (indicative maximum of 6 references)

[3-1] Zibin Zheng, Michael R. Lyu, "Collaborative Reliability Prediction of Service-Oriented Systems," ACM/IEEE International Conference on Software Engineering (ICSE), 2010. [ACM SIGSOFT Distinguished Paper Award] (DOI: 10.1145/1806799.1806809)

[3-2] Zibin Zheng, Yilei Zhang, and Michael R. Lyu, "Investigating QoS of Real-world Web services," IEEE Transactions on Services Computing (TSC), Vol 7, No. 1, pp.32-39, 2014 (DOI: 10.1109/TSC.2012.34)

[3-3] Jieming Zhu, Pinjia He, Zibin Zheng, Michael R. Lyu, "Online QoS Prediction for Runtime Service Adaptation via Adaptive Matrix Factorization," IEEE Transactions on Parallel and Distributed Systems (TPDS), Vol 28, No. 10, pp. 2911-2924, 2017 (DOI: 10.1109/TPDS.2017.2700796)

[3-4] Jieming Zhu, Pinjia He, Qiang Fu, Hongyu Zhang, Michael R. Lyu, and Dongmei Zhang, "Learning to Log: Helping Developers Make Informed Logging Decisions," in Proceedings of the ACM/IEEE 32nd International Conference on Software Engineering (ICSE 2015), Florence, Italy, May 16-24, 2015. (DOI: DOI 10.1109/ICSE.2015.60)

[3-5] Shilin He, Qingwei Lin, Jian-Guang Lou, Hongyu Zhang, Michael R. Lyu, Dongmei Zhang, "Identifying Impactful Service System Problems via Log Analysis," 26th ACM Symposium on the Foundations of Software Engineering (FSE 2018), Florida, United States, November 4-9, 2018. (DOI: 10.1145/3236024.3236083)

[3-6] C. Gao, J. Zeng, M. R. Lyu, and I. King, "Online App Review Analysis for Identifying Emerging Issues," In Proceedings of the 40th International Conference on Software Engineering (ICSE), 2018. (DOI: 10.1145/3180155.3180218)

(4) Details of the impact (indicative maximum 750 words)

Reliability is at the core of the usability of modern software, such as Web services, software systems and mobile applications. These software systems are widely utilized nowadays and provide a variety of services to billions of users globally. One tiny reliability issue could lead to serious problems, e.g., network interrupting and service outage, which further leads to revenue loss and user complaints. Thus, reliability engineering is very vital to modern software operation and maintenance. To resolve these problems, we designed a general framework, ADORE [5-1] from an intelligent and data-driven perspective. ADORE mainly analyses several typical software data from three components, i.e., QoS, log, and review. These components work in a closely integrated manner and together enable system reliability and service quality. Through about ten years devoting ourselves to this area, we have not only built certain reputations in the research community, but also showed our impacts in the industry. In the following, we will introduce more details.

QoS Prediction: As a core component of Web service reliability management, QoS prediction [5-6], aims to predict the quality of Web services. The area has been studied by many researchers all over the world for a long time. However, until the proposal of our datasets and tools, there was no comprehensive real-world Web service QoS datasets or tools for validating various QoS-based approaches. The collection of several real-world datasets and benchmark of 30+ QoS prediction approaches greatly facilitate the research achievements and benefit the entire researcher and practitioner community in the Web service area. Since the publication and open-source of datasets and tools, more than 370+ research institutes have requested for the datasets and tools from us, including large companies and top-tier universities, such as Microsoft, IBM, Amazon, Carnegie Mellon University, Imperial College London, etc. Our publications on QoS have received more than 2000 times citation so far [5-2, 5-3], indicating the impact of our QoS research outcomes, which inspires many following research studies. Together, the real-world datasets, tools and publication demonstrate the significant impacts of our QoS investigations and deployments.

Log Management: ADORE provides the automatic detection and identification of reliability issues through log management. We provide a set of log management tools powered by artificial intelligence to the research community and industry by a open-source package, namely, LogPAI [5-4], including LogAdvisor, Loglizer, Log3C, etc. These tools have been liked and recognized as useful by more than 1200 individuals from both industry and research institutes. We also provide a comprehensive dataset, which has been downloaded more than 4000 time by 200+ organizations globally such as ETH Zurich, Microsoft, HSBC, clearly demonstrating the impact of our log-based reliability management on both academia and industry. Besides, through our close collaboration with Huawei, our tools have been integrated deeply inside its product lines to trace the root cause of network reliability issues. We also work closely with Microsoft, where our methods clearly enhance the effectiveness and efficiency of reliability issue prediction. Our LogAdvisor provides high-accurate automatic suggestions to developers on where to insert a logging statement. Besides, our Log3C tool was successfully deployed on three real-world datasets at Microsoft and reached 1000x speed up than conventional methods on a huge amount of synthetic data. For example, Log3C successfully detected a serve problem occurred to a service system at Microsoft and reported it to the service team [5-7].

Review Mining: Review mining serves as a major component in software reliability management, which extracts actionable insights from user feedback at app stores, such as Google Store, App Store, etc. The extracted insights can help companies better upgrade their mobile applications, e.g.,

Instagram. Moreover, the extracted knowledge may provide informative business suggestions and guidance to companies. In this component, we designed a combination of several tools, i.e., Armor [5-5], by utilizing natural language processing techniques to automatically extract edifying knowledge from user reviews. Through our collaboration with Tencent, one of the largest IT companies in China, our tools can extract insights efficiently yet effectively. Our Armor toolset can process thousands of reviews in seconds, which may cost several hours or even days for human inspection, achieving 70% accuracy meanwhile. The toolset deployed at Tencent reduced their analysis cost by at least a factor of 500 [5-8]. These tools have been successfully validated in showing impacts in the practical environment at Tencent (e.g., deployed on 100+ industrial apps.)

To conclude, ADORE enables the software reliability engineering by providing a number of intelligent tools and various valuable datasets. It demonstrates significant impacts on both academic development and industrial deployment.

(5) Sources to corroborate the impact (indicative maximum of 10 references)

[5-1] ADORE. https://cuhk-adore.github.io/

[5-2] Papers that use ADORE. http://cuhk-adore.github.io/archive/biblist.html

[5-3] Citations of ADORE publications. http://cuhk-adore.github.io/archive/citation.html

[5-4] LogPAI project. https://github.com/logpai

[5-5] Armor. https://remine-lab.github.io

[5-6] WS-DREAM https://github.com/wsdream

[5-7] Microsoft Support Letter

[5-8] Tencent Support Letter