PINJIA HE

Computer Science & Engineering Department at The Chinese University of Hong Kong
(+852) 93486230 ✉ pjhe@cse.cuhk.edu.hk www.cse.cuhk.edu.hk/∼pjhe

BRIEF INTRODUCTION

Pinjia He is supervised by Prof. Michael R. Lyu at CUHK. His research focuses on software engineering and data mining. He has published 15 papers, including papers at top conferences/journals (e.g., TDSC, ICSE, DSN, TPDS). His research has been transformed into the product lines of Microsoft and Huawei.

RESEARCH INTERESTS

- **Software Engineering**, especially **log mining, logging practice, code analysis, failure prediction**
- **Data Mining**, especially **classification, nlp, machine learning application testing**

EDUCATION

The Chinese University of Hong Kong (CUHK) *Shatin, N. T., Hong Kong*
Ph.D. Candidate, Computer Science & Engineering *Aug. 2013–Feb. 2018*
Advisor: Prof. Michael R. Lyu (ACM/IEEE/AAAS Fellow, H-index: 81)

University of Illinois at Urbana-Champaign (UIUC) *Urbana, United States*
Visiting Researcher, Computer Science *July 2016–Sep. 2016*
Advisor: Prof. Tao Xie (IEEE Fellow)

South China University of Technology (SCUT) *Guangzhou, China*
B.Eng., Computer Science and Technology, (Ranked 2nd) *Sept. 2010–July 2013*
∗ Selected by SCUT for a special program for most promising students and finished the undergraduate study in 3 years

SELECTED PUBLICATIONS

Part 1: Log Analysis Papers:

[1] **Pinjia He**, Jieming Zhu, Shilin He, Jian Li, Michael R. Lyu. *Towards Automated Log Parsing for Large-Scale Log Data Analysis*. IEEE Transactions on Dependable and Secure Computing (TDSC), accepted.

[2] **Pinjia He**. *An End-to-end Log Management Framework for Distributed Systems*. The 36th International Symposium on Reliable Distributed Systems (SRDS 2017), PhD Forum, Hong Kong, Sept. 26-29, 2017.

[3] **Pinjia He**, Jieming Zhu, Zibin Zheng, Michael R. Lyu. *Drain: An Online Log Parsing Approach with Fixed Depth Tree*. The 24th International Conference on Web Service (ICWS 2017), Honolulu, Hawaii, June 25-30 2017.

[4] **Pinjia He**, Jieming Zhu, Shilin He, Jian Li, Michael R. Lyu. *An Evaluation Study on Log Parsing and Its Use in Log Mining*. The 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2016), Toulouse, France, June 28-July 1, 2016.

[5] Shilin He, Jieming Zhu, **Pinjia He** Michael R. Lyu. *Experience Report: System Log Analysis for Anomaly Detection Factorization*. The 27th International Symposium on Software Reliability Engineering (ISSRE 2016), Ottawa, Canada, Oct. 23-27 2016.

[6] Jieming Zhu, **Pinjia He**, Qiang Fu, Hongyu Zhang, Michael R. Lyu, Dongmei Zhang. *Learning to Log: Helping Developers Make Informed Logging Decisions*. The 37th International Conference on Software Engineering (ICSE 2015), Firenze, Italy, May 16-24, 2015.

Part 2: Software Reliability Prediction and Machine Learning Papers:

[1] **Pinjia He**, Jieming Zhu, Zibin Zheng, Jianlong Xu, Michael R. Lyu. *Location-Based Hierarchical Matrix Factorization for Web Service Recommendation*. The 21st International Conference on Web Service (ICWS 2014), Alaska, USA, June 27-July 2, 2014.

[2] **Pinjia He**, Jieming Zhu, Jianlong Xu, Michael R. Lyu. *A Hierarchical Matrix Factorization Approach for Location-Based Web Service QoS Prediction*. The International Workshop on Internet-based Virtual Computing Environment (iVCE 2014), Oxford, UK, Apr. 7-11, 2014.

[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), Volume 28, Issue 10.

[4] Jieming Zhu, **Pinjia He**, Zibin Zheng, Michael R. Lyu. *A Privacy-Preserving QoS Prediction Framework for Web Service Recommendation*. The 22nd International Conference on Web Service (ICWS 2015), New York, USA, June 27-July 2, 2015.

[5] Cuiyun Gao, Baoxiang Wang, **Pinjia He**, Jieming Zhu, Yangfan Zhou, Michael R. Lyu. *PAID: Prioritizing App Issues for Developers by Tracking User Reviews Over Versions*. The 26th International Symposium on Software Reliability Engineering (ISSRE 2015), Washington DC, USA, Nov. 2-5, 2015.

[6] Jieming Zhu, **Pinjia He**, Zibin Zheng, Michael R. Lyu. *Towards Online, Accurate, and Scalable QoS Prediction for Runtime Service Adaptation*. The 34th International Conference on Distributed Computing Systems (ICDCS 2014), Madrid, Spain, June 30-July 3, 2014.

[7] Tong Zhao, Junjie Hu, **Pinjia He**, Hang Fan, Irwin King, Michael R. Lyu. *Exploiting Homophily-based Implicit Social Network to Improve Recommendation Performance*, 2014 International Joint Conference on Neural Networks (IJCNN 2014), Beijing, China, July 6-11, 2014.

PROFESSIONAL ACTIVITIES

Reviewer

- Computer & Security (COSE) *2018*
- ACM Transactions on Software Engineering and Methodology (TOSEM) *2018*
- IEEE Transactions on Knowledge and Data Engineering (TKDE) *2017*

External Reviewer (selected)

- IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), Luxembourg *2018*
- International Joint Conference on Artificial Intelligence (IJCAI), Stockholm, Sweden *2018*
- IEEE International Conference on Distributed Computing Systems (ICDCS), Atlanta, U.S. *2017*
- International Conference on Software Engineering (ICSE), Austin, U.S. *2016*
- Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), Toulouse, France *2016*
- International World Wide Web Conference (WWW), Montreal, Canada *2016*
- ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), Sydney, Australia *2016*
- International Conference on Software Engineering (ICSE), Florence, Italy *2015*
- International World Wide Web Conference (WWW), Florence, Italy *2015*

TEACHING ASSISTANT

CSCI3100: Software Engineering *2017 Spring*
CSCI3100: Software Engineering (**Excellent Teaching Assistantship**) *2016 Spring*
ENGG1110: Problem Solving By Programming *2016/2015/2014/2013 Fall*
ENGG2600: Technology, Society and Engineering Practice *2015 Spring*
CSCI3250: Computer and Society *2014 Spring*

SELECTED AWARDS

CUHK Excellent Teaching Assistantship (CSCI3100 Software Engineering) *2015-2016*
CUHK Overseas Research Attachment Programme Scholarship *2015-2016*
CUHK Reaching Out Award *2015-2016*
DSN Student Travel Grant *2016*
CUHK Postgraduate Student Scholarship *2013–2017*
Honorable Mention in Mathematical Contest in Modeling *2013*
The Second Prize Scholarship awarded by SCUT *2011–2012*
Merit Student of SCUT in 2011-2012 *2011–2012*
The First Prize Scholarship awarded by SCUT *2010–2011*
Merit Student of SCUT in 2010-2011 *2010–2011*

PROJECTS

Mining Massive Logs, collaborate with MSRA *Mar. 2016–Present*

- Design end-to-end log collection and automated analysis platform for Microsoft cloud.
- Use large-scale data processing platform Spark in massive log analysis, including log parsing, anomaly detection, etc.
- Implement a parallel log parser that can parse logs by **1 GB/min** with more than **95%** accuracy. Our implementation of the parallel log parser and other representative log parsers have been released as a Github project Logparser. Part of the results has been published in DSN'2016 with **DSN Student Travel Grant**.

WDM Network Fault Diagnosis, collaborate with Huawei *July 2015–July 2016*

- Design an automated fault diagnosis framework for Huawei WDM network.
- Use machine learning techniques to conduct network data (including alarms, performance data and logs) mining.
- Develop a series of anomaly detection and failure diagnosis approaches based on various techniques, including clustering, association rule mining, linear invariants mining, etc. We obtain **99%** accuracy on labeled **Hadoop File System** (HDFS) data, and **91%** accuracy on unlabeled HDFS data. These approaches and evaluation results have been published in ISSRE'2016.

- The prototypes of these techniques have been released as a Github project Loglizer. Part of this project has been successfully applied to **Huawei fault diagnosis cases**.

Learning to Log, collaborate with MSRA *2013–2014*

- First use machine learning techniques to mine the logging practice of master developers from high-quality source codes.
- The effectiveness of this method has been determined on source codes of two **Microsoft projects**. It obtains **90%** accuracy on the "Whether to log or not" problem.
- This technique has been developed as a prototype for **Visual Studio**, and its effectiveness is confirmed by the Visual Studio product team. Besides, this technique has been published in ICSE'2015 and ICSE'2014.
- The implementation of this technique has been released as a Github project LogAdvisor.

Other Research Grants *2013–Present*

- Core Member, "Research on Efficient and Reliable Internet-based Virtual Computing Environments", General Research Project (a.k.a. **973** Project), the National Basic Research Program of China, 2011-2015.
- Core Member, "Towards Trustworthy Cloud Computing with Component-based Design, Online Evaluation, and Runtime Optimization Techniques", **NSFC/RGC** Joint Research Scheme, 2012-2014.
- Core Member, "Fault Tolerance and Personalized Reliability Prediction in Services Computing", General Research Fund for Young Scientists, **NSFC**, 2012-2014
- Core Member, "On the Development of Reliable Internet-of-Things Systems: Theory, Technique, and Experimentation", General Research Fund for Young Scientists, **NSFC**, 2012-2014.

TRANSFERABLE SKILLS

- **TOEFL**: 101 (Reading 29, Listening 24, Speaking 23, Writing 25) in 2012
- **2 Course Certifications on edX**: Introduction to Big Data with Apache Spark, Scalable Machine Learning.