1-2015

Improving Software Quality and Productivity Leveraging Mining Techniques: [Summary of the Second Workshop on Software Mining, at ASE 2013]

Ming LI
Nanjing University

Hongyu ZHANG
Tsinghua University

David LO
Singapore Management University, davidlo@smu.edu.sg

Lucia Lucia
Singapore Management University, lucia.2009@phdis.smu.edu.sg

DOI: https://doi.org/10.1145/2693208.2693219

Improving Software Quality and Productivity Leveraging Mining Techniques

[Summary of the Second Workshop on Software Mining, at ASE 2013]

Ming Li[1], Hongyu Zhang[2], David Lo[3], Lucia[3]
[1]Nanjing University, China
[2]Tsinghua University, China
[3]Singapore Management University, Singapore
lim@lamda.nju.edu.cn, hongyu@tsinghua.edu.cn, {davidlo,lucia.2009}@smu.edu.sg

ABSTRACT

The second International Workshop on Software Mining (Softmine) was held on the 11^{th} of November 2013. The workshop was held in conjunction with the 28^{th} IEEE/ACM International Conference on Automated Software Engineering (ASE) in Silicon Valley, California, USA. The workshop has facilitated researchers who are interested in mining various types of software-related data and in applying data mining techniques to support software engineering tasks. During the workshop, seven papers on software mining and behavior models, execution trace mining, and bug localization and fixing were presented. One of the papers received the best paper award. Furthermore, there were two invited talk sessions presented by two active researchers from software engineering and data mining community.

Keywords

Software quality, developer productivity, data mining, workshop report

1. INTRODUCTION

Software systems have been playing important roles in business, scientific research, and our everyday lives. It is critical to improve both software quality and developer productivity, which are major challenges to software engineering researchers and practitioners. In recent years, software mining has emerged as a promising means to address these challenges. It has been successfully applied to discover knowledge from software artifacts (e.g., specifications, source code, documentations, execution logs, and bug reports) to improve software quality and development process (e.g., to obtain the insights for the causes leading to poor software quality, to help software engineers locate and identify problems quickly, and to help the managers optimize the resources for better productivity) [6, 10, 3, 9, 1, 4, 7, 5, 8, 2]. Software mining has attracted much attention in both software engineering and data mining communities.

The Second International Workshop on Software Mining (SoftwareMining 2013) was held at Palo Alto, USA, on November 11, 2013, in conjunction with the 28th IEEE/ACM International conference on Automated Software Engineering (ASE 2013). The workshop aims to bridge research in the data mining community and software engineering community by providing an open and interactive forum for researchers who are interested in software mining to discuss the methodologies and technical foundations of software mining, approaches and techniques for mining various types of software-related data, and applications of data mining to facilitate specialized tasks in software engineering. Participants of diverse background in either data mining or software engineering can benefit from this workshop by sharing their expertise, exchanging ideas and discussing new research results. This series of workshop on software mining was expected to be held in data mining community and software engineering community in turns. The first workshop of this series was held in in conjunction with KDD 2012, a leading conference in data mining community.

Among 17 submissions that we received, 53% of them were from Asia, 23% from Europe, 12% from North America and 12% from Oceania. The submissions went through a rigorous reviewing process. Most submissions received three reviews. Reviews were done with the goal of providing constructive suggestions to the authors. The SoftwareMining-2013 chairs examined the reviews to further guarantee the reliability and integrity of the reviewing process. Finally, seven papers were accepted, one of which was granted the best paper award. These papers can be grouped into three categories: software mining and behavior models, execution trace mining, and bug localization and fixing. The best paper award was given to Du et al. for their paper entitled: "Interaction Traces Mining for Efficient System Responses Generation".

To further realize the goal of this workshop, we have two invited talk sessions, where two active researchers from software engineering and data mining community, i.e., Dr. Tao Xie from University of Illinois at Urbana-Champaign and Dr. Ling Huang from Intel Labs introduced their recent works on software mining. Besides, a panel session on "Software Mining, What Next?" was held for inspiring discussions. Among the topics discussed, there was a discussion on the best venue to organize this workshop in the future. The participants agree that the workshop should be continued in the future. Some participants prefer that the workshop is permanently co-located with a software engineering conference (e.g., future ASE conferences); others prefer the workshop to rotate between software engineering and data mining conferences.

2. WORKSHOP SESSIONS

The workshop was organized into two keynote sessions, three paper presentation sessions, and two discussions. Before the first paper presentation session, Dr. Tao Xie from University of Illinois at Urbana-Champaign gave a keynote on software analytics i.e., Software Analytics: Towards Software Mining that Matters. Also, before the third paper presentation session, Dr. Ling Huang from Intel Labs gave a keynote on their recent work on software mining.

The first paper presentation session presented two papers on software mining and behavior models i.e., "Towards a Model-based

DOI:10.1145/2693208.2693219 http://doi.acm.org/2693208.2693219

Software Mining Infrastructure" and "Calling Network: A New Method for Modeling Software Runtime Behaviors". The second session presented three papers on mining execution traces i.e., "Interaction Traces Mining for Efficient System Responses Generation", "Achieving Effective Test Suites for Reactive Systems using Specification Mining and Test Suite Reduction Techniques", and "Automata-based Pattern Mining from Imperfect Traces". After the second session, we had a discussion about model and mining execution traces.

The third session presented two papers on bug localization and bug fixing i.e., "Comparing Incremental Latent Semantic Analysis Algorithms for Efficient Retrieval from Software Libraries for Bug Localization" and "On Incomplete Bug Fixes in Eclipse and Programmers' Intuition on These". This session was then followed by MALIR SE-2013 session which presented two papers. After that, we had a joint session with MALIR SE-2013 to discuss about bugs and anti-patterns as well as about future directions in the field of data mining, machine learning, and information retrieval for software engineering.

3. CONCLUSION

The second workshop on Software Mining has been successfully run in conjunction with the 28^{th} IEEE/ACM International Conference on Automated Software Engineering. Seven papers on software mining and behavior models, execution trace mining, and bug localization and fixing were presented. Also, two interesting talk sessions were held during the workshop.

4. ACKNOWLEDGEMENT

The workshop would not have been a success without the support of many people. We wish to take this opportunity to thank the program committee members for their efforts and engagements to help produce a rich and interesting scientific program as well as providing comments and suggestions to the authors. We wish to express our gratitude to the ASE 2013 Workshop Chairs for their invaluable supports. We are also grateful to the invited speakers for their insightful and inspirational talks, and we also want to thank all authors and all workshop participants for their contributions and supports.

5. REFERENCES

[1] R. Ding, Q. Fu, J. Lou, Q. Lin, D. Zhang, and T. Xie, "Mining historical issue repositories to heal large-scale online service systems," in *44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, DSN 2014, Atlanta, GA, USA, June 23-26, 2014*, 2014, pp. 311–322.

[2] K. Hossen, H. H. Kagdi, and D. Poshyvanyk, "Amalgamating source code authors, maintainers, and change proneness to triage change requests," in *22nd International Conference on Program Comprehension, ICPC 2014, Hyderabad, India, June 2-3, 2014*, 2014, pp. 130–141.

[3] M. Li, H. Zhang, R. Wu, and Z. Zhou, "Sample-based software defect prediction with active and semi-supervised learning," *Autom. Softw. Eng.*, vol. 19, no. 2, pp. 201–230, 2012.

[4] J. Lou, Q. Lin, R. Ding, Q. Fu, D. Zhang, and T. Xie, "Software analytics for incident management of online services: An experience report," in *2013 28th IEEE/ACM International Conference on Automated Software Engineering, ASE 2013, Silicon Valley, CA, USA, November 11-15, 2013*, 2013, pp. 475–485.

[5] R. K. Saha, M. Lease, S. Khurshid, and D. E. Perry, "Improving bug localization using structured information retrieval," in *28th IEEE/ACM International Conference on Automated Software Engineering, ASE 2013, Silicon Valley, CA, USA, November 11-15, 2013*, 2013, pp. 345–355.

[6] F. Thung, D. Lo, and L. Jiang, "Automatic defect categorization," in *19th Working Conference on Reverse Engineering, WCRE 2012, Kingston, ON, Canada, October 15-18, 2012*, 2012, pp. 205–214.

[7] S. Wang and D. Lo, "Version history, similar report, and structure: putting them together for improved bug localization," in *22nd International Conference on Program Comprehension, ICPC 2014, Hyderabad, India, June 2-3, 2014*, 2014, pp. 53–63.

[8] S. Wang, D. Lo, and J. Lawall, "Compositional vector space models for improved bug localization," in *30th IEEE International Conference on Software Maintenance and Evolution, ICSME 2014, Victoria, BC, Canada*, 2014.

[9] X. Xia, D. Lo, X. Wang, and B. Zhou, "Accurate developer recommendation for bug resolution," in *20th Working Conference on Reverse Engineering, WCRE 2013, Koblenz, Germany, October 14-17, 2013*, 2013, pp. 72–81.

[10] J. Zhou, H. Zhang, and D. Lo, "Where should the bugs be fixed? more accurate information retrieval-based bug localization based on bug reports," in *34th International Conference on Software Engineering, ICSE 2012, June 2-9, 2012, Zurich, Switzerland*, 2012, pp. 14–24.