

¹,王 勇¹,房丙午¹

1. 211106
2. 830054

L 2 ó

10 2005 ~2015

2 p N ó

Đ n t ì } ó TP311.5 Ô E T N “ ó A Ô d J } ó 0372-2112 2017 04-0982-07

É 1 Ñ URL ó [http //www. ejournal. org. cn](http://www.ejournal.org.cn) DOI 10.3969/j. issn. 0372-2112. 2017. 04. 030

Survey on Data Driven Software Defects Prediction

LI Yong^{1,2} HUANG Zhi-qiu¹ WANG Yong¹ FANG Bing-wu¹

1. *College of Computer Science and Technology Nanjing University of Aeronautics and Astronautics Nanjing Jiangsu 211106 China*

2. *Key Laboratory of Network Information Security and Public Opinion Analysis Xinjiang Normal University Urumqi Xinjiang 830054 China*

Abstract Data driven software defects prediction is considered as an effective means for the optimization of quality assurance activities. This paper tries to survey the studies of data driven software defects prediction from 2005 to 2015. First it briefly introduces the research background in this field. Then it summarizes and analyzes the software modules attribute metrics in detail and thoroughly surveys the state-of-the-art methods of data driven software defects prediction. The discussed topics include within-release cross-release and cross-project defects prediction. Finally the potential research directions in the field are discussed.

Key words software defects prediction data driven software metrics machine learning

1 ...p

1

- ①
- ②
- ③
- ④

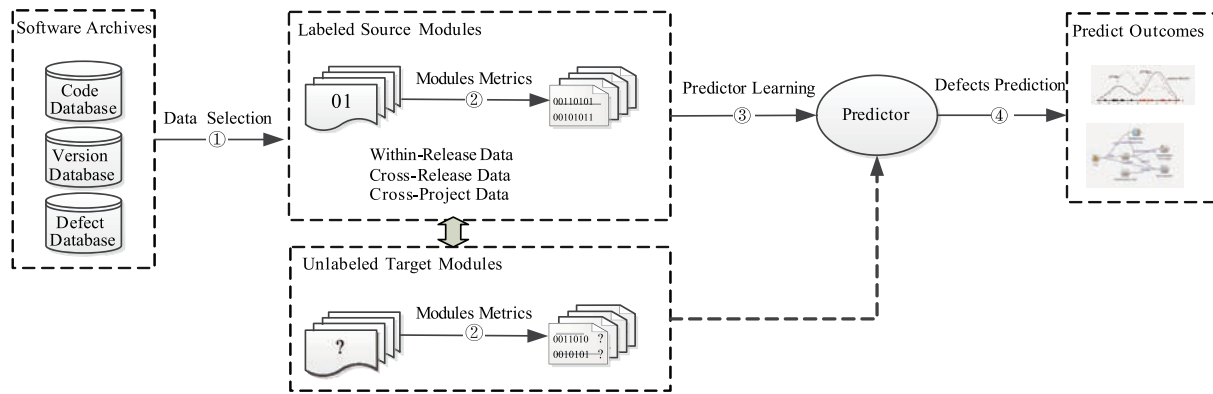


图1 数据驱动的软件缺陷预测流程图

3 É r • K • Œ [F

1

W1 É r • K • Œ [F

Product Metrics	Traditional Metrics	LOC		7 8
		Halstead		5 7 8
		McCabe		6 7 8
	OO Metrics	CK MOOD QMOOD Martin		8 9 10 11
Process Metrics	Change Metrics	Code Churn		12 13 14
		Revisions Count		15
		Pre-version Defects Count		16 17
	Developer Metrics	Code Developers		18 19

3.1 É r Å x Œ

7

1

8

LOC Halstead⁵

McCabe⁶

2 4 • { i | É r • K è a MD [F

CK MOOD 4.1 ò (> • { i | • K è a [F

QMOOD Martin 8 .

9 CK

10 11 .

CK MOOD QMOOD Martin 1

CK

Martin

3.2 É r Wò Œ 20 .

1 Code Churn

12 21 AdaBoost. NC

13 14 "

"

15 Change bursts 2

16 17

22 X-means

2 23

18 24

19 25 CoForest

ACoForest

4.2 u ò (• { i | • K è a [F



4.3 $u \setminus \# \bullet \{ \ddot{I} \mid \bullet K \dot{e}^a [F$

5 Ä ú ï ð U[-

1

2

3

39

4

41

40

6 3 Õ

- J Ô E

- 1 Lyu M R. Software reliability engineering a roadmap A . 2007 Future of Software Engineering C . Washington IEEE Computer Society 2007. 153 – 170.

- 2 Caglayan B Tosun Misirli A Bener A B et al. Predicting defective modules in different test phases J . Software Quality Journal 2015 23 2 205 – 227.
- 3 Shepperd M Bowes D Hall T. Researcher bias the use of machine learning in software defect prediction J . IEEE Transactions on Software Engineering 2014 40 6 603 – 616.
- 4 Hall T Beecham S Bowes D et al. A systematic literature review on fault prediction performance in software engineering J . IEEE Transactions on Software Engineering 2012 38 6 1276 – 1304.
- 5 Halstead M H. Elements of Software Science Operating and Programming Systems Series M . New York Elsevier Science Inc 1977. 1 – 128.
- 6 McCabe T J. A complexity measure J . IEEE Transactions on Software Engineering 1976 SE – 2 SE – 2 308 – 320.
- 7 Menzies T Greenwald J Frank A. Data mining static code attributes to learn defect predictors J . IEEE Transactions on Software Engineering 2007 33 1 2 – 13.
- 8 Radjenovi D Heri ko M Torkar R et al. Software fault prediction metrics a systematic literature review J . Information and Software Technology 2013 55 8 1397 – 1418.
- 9 Basili V R Briand L C Melo W L. A validation of object-oriented design metrics as quality indicators J . IEEE Transactions on Software Engineering 1996 22 10 751 – 761.
- 10 Elish M O Al-Yafei A H Al-Mulhem M. Empirical comparison of three metrics suites for fault prediction in packages of object-oriented systems a case study of eclipse J . Advances in Engineering Software 2011 42 10 852 – 859.
- 11 Olague H M Etzkorn L H Gholston S et al. Empirical validation of three software metrics suites to predict fault-proneness of object-oriented classes developed using highly iterative or agile software development processes J . IEEE Transactions on Software Engineering 2007 33 6 402 – 419.
- 12 Bell R M Ostrand T J Weyuker E J. Does measuring code change improve fault prediction A . Proceedings of the 7th International Conference on Predictive Models in Software Engineering C . New York ACM 2011. 1 – 8.
- 13 Munson J C Elbaum S G. Code churn a measure for estimating the impact of code change A . Proceedings of the International Conference on Software Maintenance C . Washington IEEE Computer Society 1998. 24 – 31.
- 14 Kpodjedo S Ricca F Galinier P et al. Design evolution

- metrics for defect prediction in object oriented systems J . Empirical Software Engineering 2011 16 1 141 – 175.
- 15 Nagappan N Zeller A Zimmermann T et al. Change bursts as defect predictors A . Proceedings of the 21st IEEE International Symposium on Software Reliability Engineering C . Washington IEEE Computer Society 2010. 309 – 318.
 - 16 Khoshgoftaar T M Allen E B Halstead R et al. Using process history to predict software quality J . Computer 1998 31 4 66 – 72.
 - 17 Illes-Seifert T Paech B. Exploring the relationship of a file's history and its fault-proneness an empirical method and its application to open source programs J . Information and Software Technology 2010 52 5 539 – 558.
 - 18 Okutan A Yildiz O T. Software defect prediction using Bayesian networks J . Empirical Software Engineering 2014 19 1 154 – 181.
 - 19 Ostrand T J Weyuker E J Bell R M. Programmer-based fault prediction A . Proceedings of the 6th International Conference on Predictive Models in Software Engineering C . New York ACM 2010. 1 – 10.
 - 20 Malhotra R. A systematic review of machine learning techniques for software fault prediction J . Applied Soft Computing 2015 27 0 504 – 518.
 - 21 Wang S Yao X. Using class imbalance learning for software defect prediction J . IEEE Transactions on Reliability 2013 62 2 434 – 443.
 - 22 Catal C Sevim U Diri B. Metrics-driven Software Quality Prediction Without Prior Fault Data M . Berlin Springer 2010. 189 – 199.
 - 23 Catal C Sevim U Diri B. Clustering and metrics thresholds based software fault prediction of unlabeled program modules A . Proceedings of the 2009 Sixth International Conference on Information Technology New Generations C . Washington IEEE Computer Society 2009. 199 – 204.
 - 24 Alan O Catal C. Thresholds based outlier detection approach for mining class outliers an empirical case study on software measurement datasets J . Expert Systems with Applications 2011 38 4 3440 – 3445.
 - 25 Li M Zhang H Wu R et al. Sample-based software defect prediction with active and semi-supervised learning J . Automated Software Engineering 2012 19 2 201 – 230.
 - 26 Rahman F Devanbu P. How and why process metrics are better A . Proceedings of the 2013 International Conference on Software Engineering C . USA IEEE Press 2013. 432 – 441.
 - 27 Kastro Y Ay Bener E B. A defect prediction method for software versioning J . Software Quality Control 2008 16 4 543 – 562.
 - 28 Weyuker E Ostrand T Bell R. Do too many cooks spoil the broth Using the number of developers to enhance defect prediction models J . Empirical Software Engineering 2008 13 5 539 – 559.
 - 29 Misirli A T Murphy B Zimmermann T et al. An explanatory analysis on eclipse beta-release bugs through in-process metrics A . Proceedings of the 8th International Workshop on Software Quality C . New York ACM 2011. 26 – 33.
 - 30 Turhan B. On the dataset shift problem in software engineering prediction models J . Empirical Software Engineering 2012 17 1 62 – 74.
 - 31 Watanabe S Kaiya H Kaijiri K. Adapting a fault prediction model to allow inter language reuse A . Proceedings of the 4th International Workshop on Predictor Models in Software Engineering C . New York ACM 2008. 19 – 24.
 - 32 Nam J Pan S J Kim S. Transfer defect learning A . Proceedings of the 2013 International Conference on Software Engineering C . USA IEEE Press 2013. 382 – 391.
 - 33 He P Li B Liu X et al. An empirical study on software defect prediction with a simplified metric set J . Information and Software Technology 2015 59 170 – 190.
 - 34 Turhan B Menzies T Bener A B et al. On the relative value of cross-company and within-company data for defect prediction J . Empirical Software Engineering 2009 14 5 540 – 578.
 - 35 Herbold S. Training data selection for cross-project defect prediction A . Proceedings of the 9th International Conference on Predictive Models in Software Engineering. Baltimore C . New York ACM 2013. 1 – 10.
 - 36 Ma Y Luo G Zeng X et al. Transfer learning for cross-company software defect prediction J . Information and Software Technology 2012 54 3 248 – 256.
 - 37 Menzies T Butcher A Cok D et al. Local versus global lessons for defect prediction and effort estimation J . IEEE Transactions on Software Engineering 2013 39 6 822 – 834.
 - 38 Bettenburg N Nagappan M Hassan A E. Think locally act globally Improving defect and effort prediction models A . Proceedings of the 9th IEEE Working Conference on Mining Software Repositories C . USA IEEE Press 2012. 60 – 69.
 - 39 Peters F Menzies T Gong L et al. Balancing privacy and utility in cross-company defect prediction J . IEEE Transactions on Software Engineering 2013 39 8 1054

- 1068.
- 40 Lewis C Lin Z Sadowski C et al. Does bug prediction support human developers findings from a google case study A . Proceedings of the 2013 International Conference on Software Engineering C . USA IEEE Press 2013. 372 – 381.
- 41 Czerwonka J Das R Nagappan N et al. CRANE failure prediction change analysis and test prioritization in practice—experiences from Windows A . Proceedings of the 2011 Fourth IEEE International Conference on Software Testing Verification and Validation C . Washington IEEE Computer Society 2011. 357 – 366.

U • f »



Li Yong 1983 12

E-mail liyong@live.com



Zhiyuan Zhang 1965

CCF

E-mail zhuang@nuaa.edu.cn