

Investigating Code Review Quality: Do People and Participation Matter?

Oleksii Kononenko* , Olga Baysal† , Latifa Guerrouj‡ ,
Yaxin Cao† , and Michael W. Godfrey

ICSME, Sept. 29 2015-Oct. 1 2015, Bremen, German

60181052 宮崎 智己

コードレビュー (code review)

- ソースコードの記述者とは別の人が詳細に調べ、開発者にフィードバックすること。
- コードが仕様や要求を満たすかどうか、誤りがないか等を検査する。
- 欠点は時間や労力がかかる。

背景

- コードレビューはソフトウェア開発プロジェクトにおいて、欠かせないものになっている
- しかし、コードレビューを行っているにもかかわらず、予期せぬバグがコードベースに入り込んでしまうことがある

→レビューアーの検査(コードレビュー)と社会的、個人的な要因の間に関係があるかどうかを調べる

リサーチクエスト

- RQ1
 - レビューアーは多くのバグを見逃しているのか？
- RQ2
 - 個人の要因はコードレビューの質に影響を与えているのか？
- RQ3
 - コードレビューの関与はコードレビューの質に影響を与えているのか？

アプローチ

- Mozilla all(full set of commits),Core,Firefox,Firefox for androidの4つをモジュールを対象
- BugzillaのITSから抽出したコードレビューのデータとリンクしたバグを誘発したコード変更をSZZアルゴリズムによって見つける
 - 35668件のコミットをリンク付けした

続き

- パッチとコミットのリンク付け
 - 外れ値の除去
 - バグ誘発したコード変更かどうかの確認
 - 要因の決定
-
- 多重線形回帰モデルを構築、分析

Mozillaのレビュー

- 2段階のレビュー方法

- Review・・・パッチ投稿に対して、そのモジュールに責任あるレビューワーがレビューする
- Super review・・・通常のレビューワーでは理解しづらい場合は経歴が長くドメイン知識があるレビューワーがレビューする

要因

Type	Metric	Description	Rationale
Technical	Size (LOC)	The total number of added and removed lines of code.	Large commits are more likely to be bug-prone [15]; thus the intuition is it is easier for reviewers to miss problems in large code changes.
	Chunks	The total number of isolated places (as defined by <code>diff</code>) inside the file(s) where the changes were made.	We hypothesize that reviewers are more likely to miss bugs if the change is divided into multiple isolated places in a file.
	Number of files	The number of modified files.	Similar, reviews are more likely to be prone to bugs if the change spread across multiple files.
	Module	The name of the Mozilla module (e.g., Firefox).	Reviews of changes within certain modules are more likely to be prone to bugs.
	Priority	Assigned urgency of resolving a bug.	Our intuition is that patches with higher priority are more likely to be rushed in and thus be more bug-prone than patches with lower priority levels.
	Severity	Assigned extend to which a bug may affect the system.	We think that changes with higher levels of severity introduce less bugs because they are often reviewed by more experienced developers or by multiple reviewers.
	Super review	Indicator of whether the change required super review or not	Super review is required when changes affect core infrastructure of the code and, thus, more likely to be bug-prone.
	Number of previous patches	The number of patches submitted before the current one on a bug.	Developers can collaborate on resolving bugs by submitting improved versions of previously rejected patches.
	Number of writer's previous patches	The number of previous patches submitted by the current patch owner on a bug.	A developer can continue working on a bug resolution and submit several versions of the patch, or so called resubmits of the same patch, to address reviewers concerns.
Personal	Review queue	The number of pending review requests.	While our previous research [14] demonstrated that review loads are weakly correlated with review time and outcome; we were interested to find out whether reviewer work loads affect code review quality.
	Reviewer experience	The overall number of completed reviews.	We expect that reviewers with high overall expertise are less likely to miss a bug.
	Reviewer experience for module	The number of completed reviews by a developer for a module.	Reviewers with high reviewing experience in a certain module are less likely to miss defects; and on the contrary, reviewers with no past experience in performing code reviews for some modules are more likely to fail to catch bugs.
	Writer experience	The overall number of submitted patches.	Developers who contribute a lot to the project — have high expertise — are less likely to submit buggy changes.
	Writer experience for module	The number of submitted patches for a module.	Developers who make few changes to a module are more likely to submit buggy patches.
Participation	Number of developers on CC	The number of developers on the CC list at the moment of review decision.	Linus's law states that "given enough eyeballs, all bugs are shallow" [16].
	Number of comments	The number of comments on a bug.	The more discussion happens on a bug, the better the quality of the code changes [5].
	Number of commenting developers	The number of developers participating in the discussion around code changes.	The more people are involved in discussing bugs, the higher software quality [5].
	Average number of comments per developer	The ratio of the comment count over the developer count.	Does the number of comments per developer has an impact on review quality?
	Number of reviewer comments	The number of comments made by a reviewer.	Does reviewer participation in the bug discussion influence the quality of reviews?
	Number of writer comments	The number of comments made by a patch writer.	Does patch writer involvement in the bug discussion affect review quality?

RQ1

TABLE IV: Number of code reviews that missed bugs.

System	# Reviews	# Buggy Reviews	% Buggy Reviews
Mozilla-all	28,127	15,188	54.0 %
Core	18,759	10,184	54.3 %
Firefox	2,668	1,447	54.2 %
Firefox4Android	2,160	1,210	56.0 %

全体のコードレビューの約54%がバグを見逃していた。
一番多いものは、Firefox for Androidの56%である。

RQ2

	Mozilla	Core	Firefox	FF4A
Adjusted R^2	0.128	0.123	0.173	0.138
Size (LOC)	0.102***	0.098 ***	0.108***	0.115***
Chunks	†	†	†	†
Number of files	0.058***	0.059 ***	0.109***	0.062*
Module	★	n/a	n/a	n/a
Priority	★	★	†	·
Severity	†	†	·	†
Super review	-0.139**	-0.177***	·	n/a
Review queue	0.017***	0.0204***	0.038**	0.045**
Reviewer exp.	-0.013***	-0.012***	-0.029***	-0.041***
Reviewer exp. (mod.)	†	†	†	0.018*
Writer exp.	·	-0.004*	†	†
Writer exp. (module)	†	†	†	·
# prev patches	†	†	†	-0.045***
# writer patches	-0.012***	·	·	†

4つのモジュールに対しては、パッチのサイズ、ファイルの数、保留中のレビューリクエストの数、レビュアーの経験の4つの要因が関係していた。

RQ3

	Mozilla	Core	Firefox	FF4A
Adjusted R^2	0.134	0.128	0.173	0.147
Size (LOC)	0.105***	0.103***	0.105***	0.117***
Chunks	†	†	†	†
Number of files	0.060***	0.059***	0.090***	0.067***
Module	★	n/a	n/a	n/a
Priority	‡	★	‡	★
Severity	★	‡	★	‡
Super review	-0.124***	-0.160***	‡	n/a
# of devs on CC	0.053***	0.056***	‡	0.049*
# comments	†	†	†	†
# commenting devs	-0.124***	-0.102***	-0.075***	-0.176***
# comments/ # dev	-0.039***	-0.029**	‡	‡
# reviewer comments	0.010**	‡	0.026*	‡
# writer comments	.	.	.	-0.047**

開発者のコメントの数がコードレビューの品質に関わっている。
多くのレビューコメントがあると、バグを見落としてしまう傾向もある。

まとめ

- コードレビューは半数以上バグを見逃してしまっていた
- レビューアーの個人的な要因がレビューの品質に関係していた