

Investigating Code Review Quality: Do People and Participation Matter?

個人的要因, レビューへの参加数の観点から
“コードレビューの品質”を調査

Oleksii Kononenko¹

Latifa Guerrouj³

Michael W. Godfrey¹

Olga Baysal²

Yaxin Cao²

¹David R. Cheriton School of Computer Science,
University of Waterloo, Canada
Email: okononen@uwaterloo.ca,
migod@uwaterloo.ca

²Department of Computer Science
and Operations Research,
Universit  de Montr al, Canada
Email: yaxin.cao@umontreal.ca,
olga.baysal@umontreal.ca

³Department of Computer Science
and Software Engineering,
Concordia University, Canada
Email: latifa.guerrouj@polymtl.ca

背景

- ◎ コードレビューはソフトウェア開発にとって重要
 - ▶ コストはかかるが、未然に欠陥を防ぐための有効な方法
 - ▶ レビューワはソフトウェアの品質を保つためのゲートキーパー
- ◎ しかし、コードレビューを行なった後も欠陥が含まれている
- ◎ 先行研究では、コードレビューとソフトウェアの品質との関係などが調査されている
- ◎ 本研究では、“**コードレビューの質**”に着目して調査

RQ

◎ RQ1

- ▶ レビューワーはどれくらいバグを見逃すのか

◎ RQ2

- ▶ **個人的な要因**はコードレビューの品質に影響するのか
 - オーナーシップ（モジュールでの変更割合）とコード変更後の欠陥に強い相関がある

◎ RQ3

- ▶ **コードレビューへの参加数**はコードレビューの品質に影響するのか
 - レビューへの参加数が少ないと欠陥が出やすい

対象プロジェクト

TABLE I: Overview of the studied systems.

System	Commits	Reviews	Writers	Reviewers
Mozilla-all	27,270	28,127	784	469
Core	18,203	18,759	544	362
Firefox	2,601	2,668	214	110
Firefox for Android	2,106	2,160	108	72

- Mozilla

- ▶ 2段階レビューがある
- ▶ Mozillaの主要なプロジェクト3つ

- 期間

- ▶ 2013/1/1 - 2014/1/1

データの取得

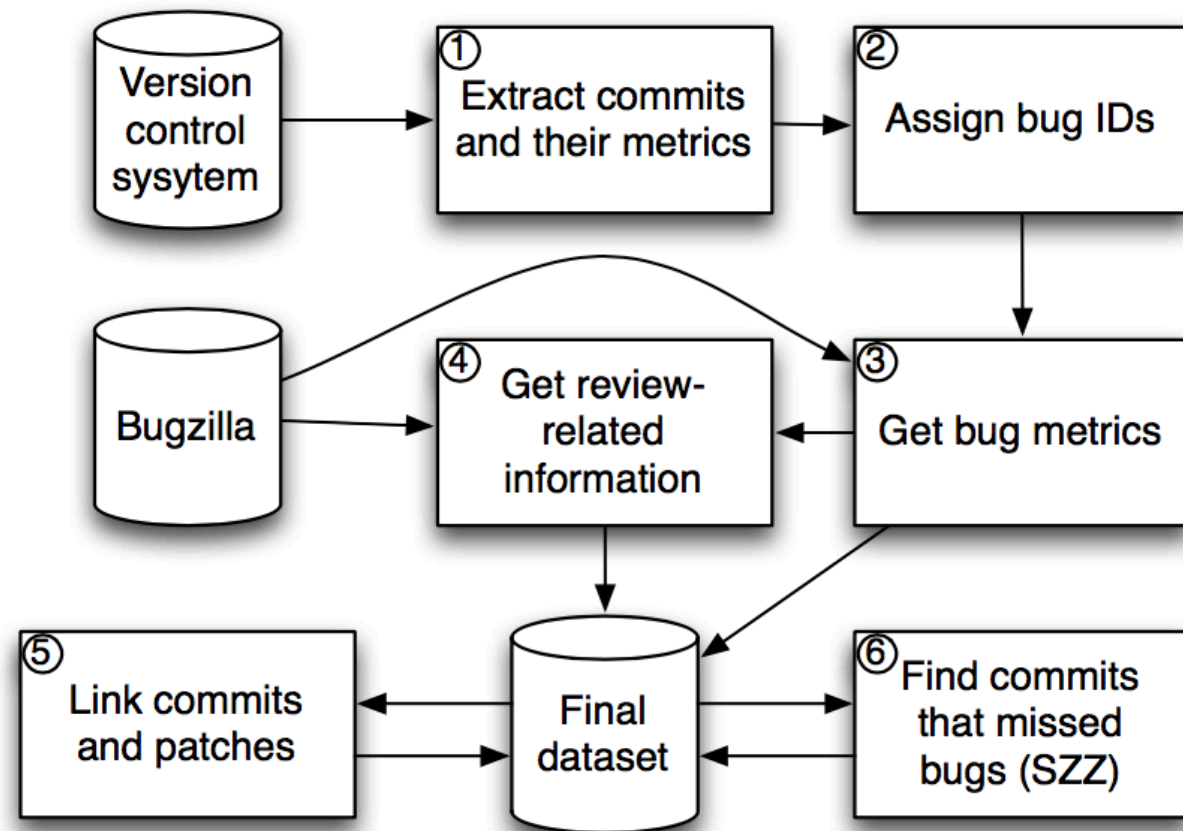


Fig. 1: Process overview.

- 1.VCSからコミットを抽出
- 2.BTSからコミットと対応する不具合を紐付け
- 3.バグに関するメトリクスを取得
- 4.レビューに関連するメトリクスを取得
- 5.コミットとパッチを紐付け
- 6.バグが混入したコミット
の特定のためにSZZを使用

メトリクス

- ◎ 3つの種類に分類
 - ▶ 技術 (Technical) : レビュー対象の難易度
 - ▶ 個人 (Personal) : プロジェクトに対する経験
 - ▶ 参加 (Participation) : レビューへの参加数
- ◎ 以下, 取得するメトリクスと取得する理由を表にした
 - ▶ 論文中表 II

メトリクス：技術

メトリクス	説明	根拠
Size(LOC)	追加/削除した コード行数の合計	サイズが大きいコミットは バグが含まれやすい□
Chunks	バグのひと塊り	仮説：複数のファイルにまた がる変更はミスしやすい モジュールによって欠陥が 発生する可能性が異なる 直感的に、優先度が高いバグ は欠陥が含まれやすそう
ファイルの数	変更したファイルの数	
モジュール	Mozillaのモジュールの名前 (例：Firefox)	
優先度	修正したバグの優先度	直感的に、優先度が高いバグ は欠陥が含まれやすそう
Severity（難易度）	割り当てられたバグがシステ ムに与えたであろう影響	難易度が高いパッチは関与 する人が多いのでバグの発 生が少なそう
スーパーレビュー	バグが含まれるモジュールの オーナーにスーパーレビュー を指示された	スーパーレビューはシステ ムのコアな部分に影響する ので、バグが出やすい
以前のパッチ数	投稿されたパッチ数	？
Writer's パッチの数	特定の人が投稿したパッチ数	？

メトリクス：個人

メトリクス	説明	根拠
キュー	レビュー中のレビューの数	レビューキューが長いと時間と労力がかけれないのでレビューの質に影響する[]
レビュワーの経験	レビューをした総数	経験が多いとミスが少なそう
モジュールでのレビュワーの経験	特定のモジュールでレビューをした総数	
パッチを書いた経験	パッチが承認された数	
モジュールでパッチを書いた経験	特定のモジュールでパッチが承認された数	

- データセットの期間が短いためツールを使用
 - ▶ Elastic Search
 - その地点の個人のメトリクスを計算できる
- レビューキューに関しては, 途中で放棄されたものは除外

メトリクス：参加

メトリクス	説明	根拠
CCの開発者の数	レビュー決定時にCCリストに載っている開発者の数	『目の数があれば、バグも問題ではない』 Linusの法則
コメントの数	バグに関するコメントの数	コメントがたくさんあるとバグが出にくい
コメントした開発者の数	コード変更時に議論に参加した開発者の数	たくさん人がいるとソフトウェアの質が上がる
開発者ごとのコメントの数の平均	コメント数を開発者で割った数	各メトリクスはコードレビューの品質に影響するのか
レビュワーコメントの数	レビュワーからのコメント数	
パッチ作成者のコメントの数	パッチ作成者からのコメント数	

モデルの構築

- ◎ MLR (Multiple Linear Regression) モデル

- ▶ 重回帰分析

- $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$

- ◎ y : 目的変数

- ▶ バグが出たかどうか

- ◎ x : 説明変数

- ▶ 上記のメトリクス

結果：RQ1

- ◎ Q. レビューワーはどれくらいバグを見逃すのか

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 %

- ◎ 半分以上
- ◎ これは, 先行研究の知見とも一致する

結果：RQ2

◎ Q. 個人的な要因はコードレビューの品質に影響するのか

(a) Technical and personal factors.

	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***	·	·	†

パッチサイズは
バグが出る確率が上がる

スーパーレビューは
バグが出る確率が下がる
> 有効

†Disregarded during VIF analysis (VIF coefficient > 5).

★ “It’s complicated”: for categorical variables see explanation of the results in Section IV.

FF4A = Firefox for Android.

正：不具合が出やすい

負：不具合が出にくい

結果：RQ2

◎ Q. 個人的な要因はコードレビューの品質に影響するのか

(a) Technical and personal factors.

	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***	·	·	†

†Disregarded during VIF analysis (VIF coefficient > 5).

★ “It’s complicated”: for categorical variables see explanation of the results in Section IV.

FF4A = Firefox for Android.

レビュー負荷が重いと低品質になる可能性が高い
「主要レビュワーへの負荷を分散する」

レビュー担当者の経験が多いとバグが出にくい

正：不具合が出やすい

負：不具合が出にくい

結果：RQ3

- Q. コードレビューへの参加数はコードレビューの品質に影響するのか

(b) Technical and participation metrics.

	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**

†Disregarded during stepwise selection.

Statistical significance: '***' $p < 0.001$; '**' $p < 0.01$; '*' $p < 0.05$; '.' $p \geq 0.05$.

CCリストへの登録者が多い
とバグが出る確率が上がる
> CCリストに登録している
からといって必ず全員がバグ
に関与するわけではない

コメントが多いと
バグが出にくい

レビュワーが多いバグは
レビュワーからの関心が高い
???

開発者はバグ修正に関与
する

正：不具合が出やすい
負：不具合が出にくい

まとめ

- ◎ 何がレビューの品質に影響を与えるのかを調査した
 - ▶ 個人的要因とレビューへの参加についてレビューの品質への影響を定量的に調査した
- ◎ 技術的要因
 - ▶ コミットサイズが大きいとバグが出やすい
 - コミットサイズを少なくするように勧める
- ◎ 個人的要因
 - ▶ 負担が多いとミスをしやすくなる
 - ▶ 知識を積むといい
- ◎ 参加数の要因
 - ▶ 予測モデルの有効な因子になる