

コードレビューでの議論内容が バギーレビューに与える影響の調 査

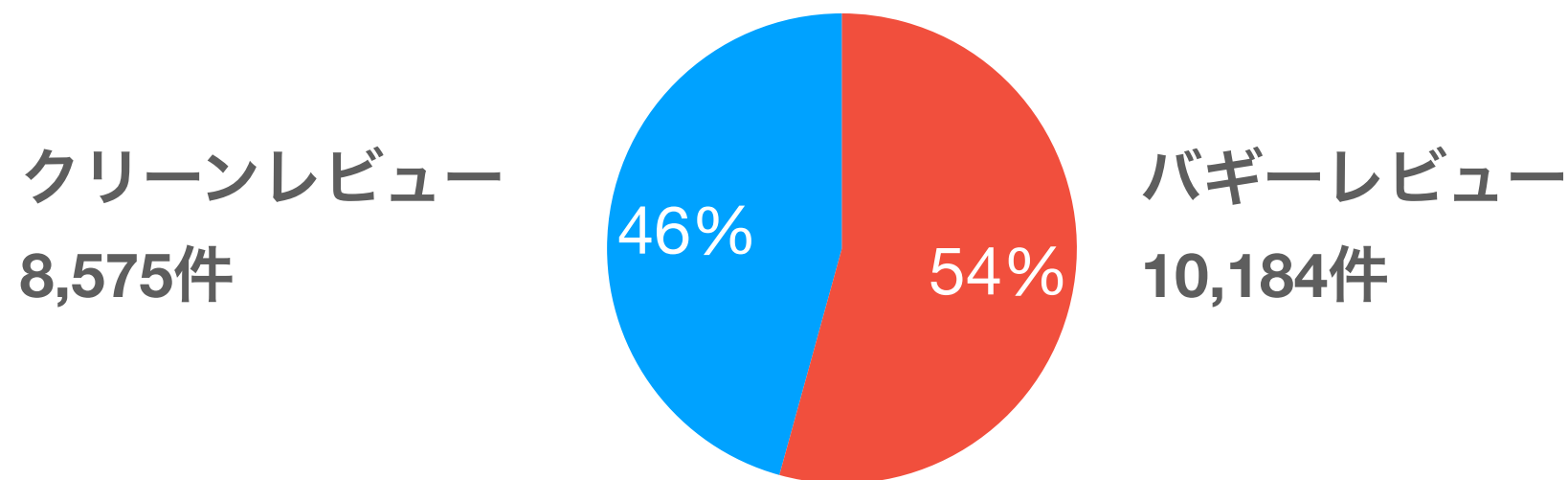
和歌山大学大学院 システム工学研究科
オープンソースソフトウェア工学研究室
60190128 山崎 大輝

コードレビューとは

- ・ ソフトウェア開発においてテスト前に不具合を発見する方法の1つ^[1]
- ・ コードを目視して以下の不具合を検証
 - ・ コーディング規約に沿っているか
 - ・ コードの可読性が高いか
 - ・ ロジックが間違っていないか

コードレビューでの問題点

- ・ バグーレビューとは
 - ・ 不具合を発見できなかったコードレビュー
- ・ **約半数**のコードレビューがバグーレビュー^[2]



- ・ バグーレビューを見つけることで保守コストを減らしたりシステムの故障を防ぐことができる

[2] O. Kononenko, O. Baysal, L. Guerrouj, Y. Cao, and M. W. Godfrey, "Investigating code review quality: Do people and participation matter?", IEEE International Conference on Software Maintenance and Evolution, pp. 111–120, 2015

バギーレビューの既存研究

- ・ バギーレビューの予測モデルを構築するため、以下の特徴量を用いている^{[2][3][4]}
 - ・ コード
 - ・ 変更行数
 - ・ 開発者
 - ・ レビュー者の人数, 経験年数
 - ・ プロセス
 - ・ コメント数, レビューにかかった時間

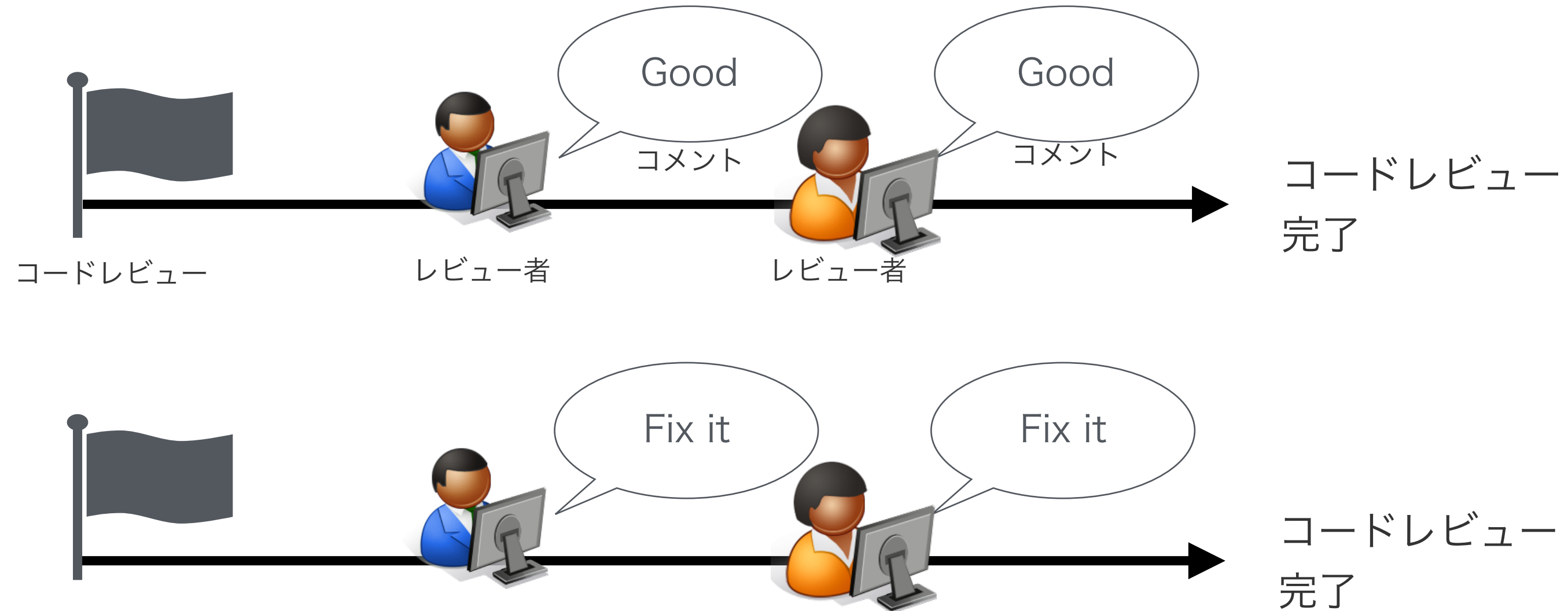
[3]B. Soltanifar, A. Erdem, and A. Bener, “Predicting Defectiveness of Software Patches,” in Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement - ESEM ’16, Ciudad Real, Spain, 2016, pp. 1–10.

バギーレビューの既存研究

- ・ 多くのモデルではソースコードに関するメトリクスに注目している
- ・ Touraniら[4]はコードレビューでの議論がバギーレビューに与える影響を調査している
- ・ e.g. コードレビューでのインラインコメント数, コメントの評判

既存研究の問題点

- 既存研究では議論内容を考慮していない



同じコメント数, 人数でもコードレビューの内容が異なる

本研究の目的

- ・ 目的
 - ・ バグーレビューの予測精度向上に向け,
コードレビューの議論内容がバグーレビューに与える
影響を調査する

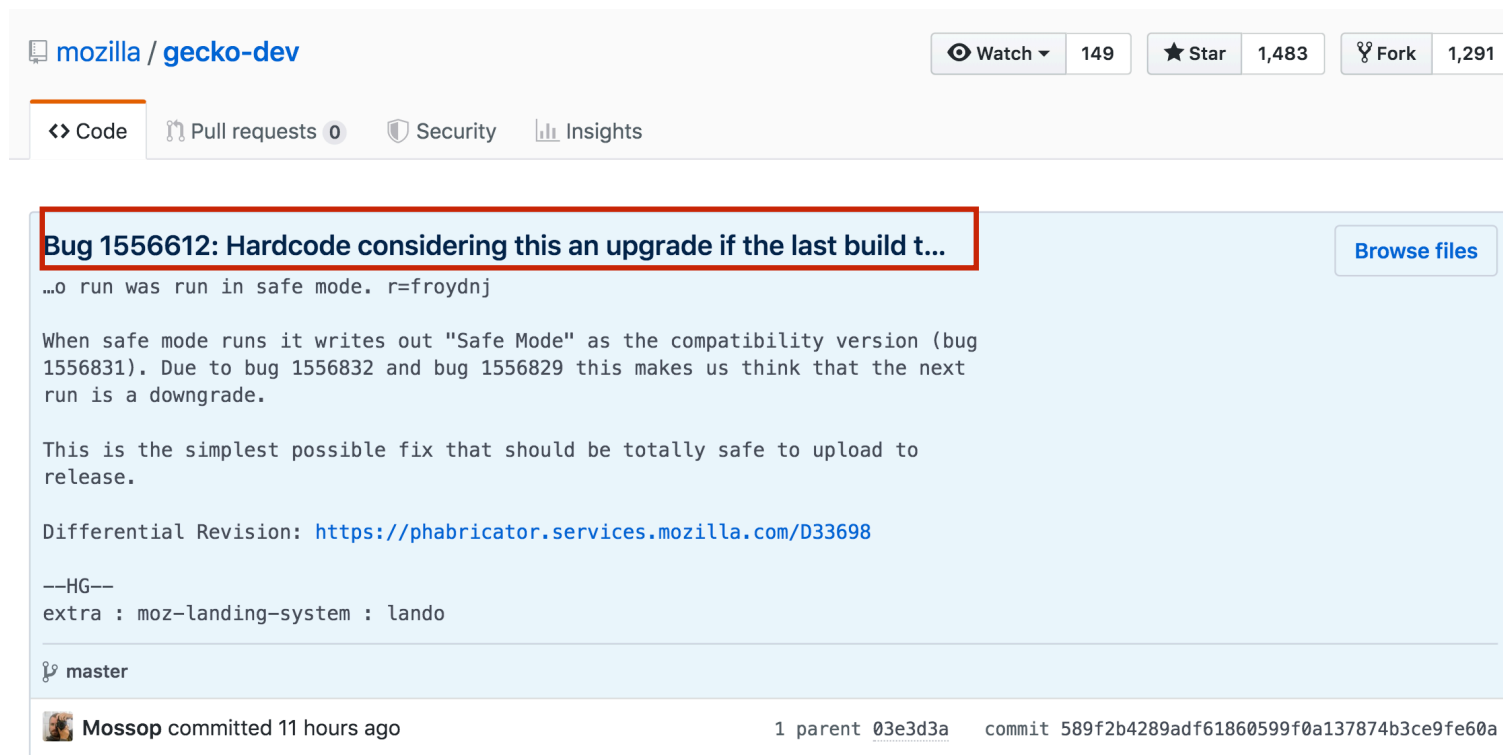
リサーチクエスション

- ・ RQ1：バギー／クリーンなコードレビューの議論内容に違いがあるか
 - ・ 議論内で使われている単語をもとに調査，メトリクス化
- ・ RQ2：RQ1で抽出したメトリクスは既存モデルの精度向上にどれだけ寄与するか
 - ・ 既存研究のメトリクス+本研究のメトリクスで予測

アプローチ

1. コードレビューがバギーかクリーンか特定する
 1. 不具合混入ファイルを特定
 2. 不具合混入コミットからバギーレビューを抽出
2. バギー/クリーンなコードレビューの議論内容を比較する
 1. コードレビュー内のコメントを数値変換する
 2. カウントした単語数をもとにモデル構築

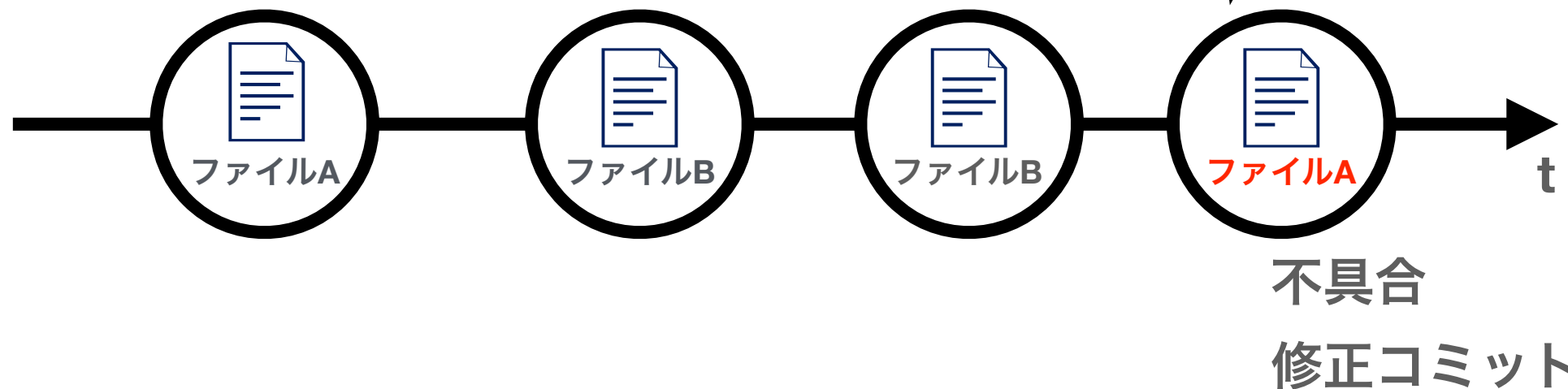
1.1 不具合混入ファイルを特定[4]



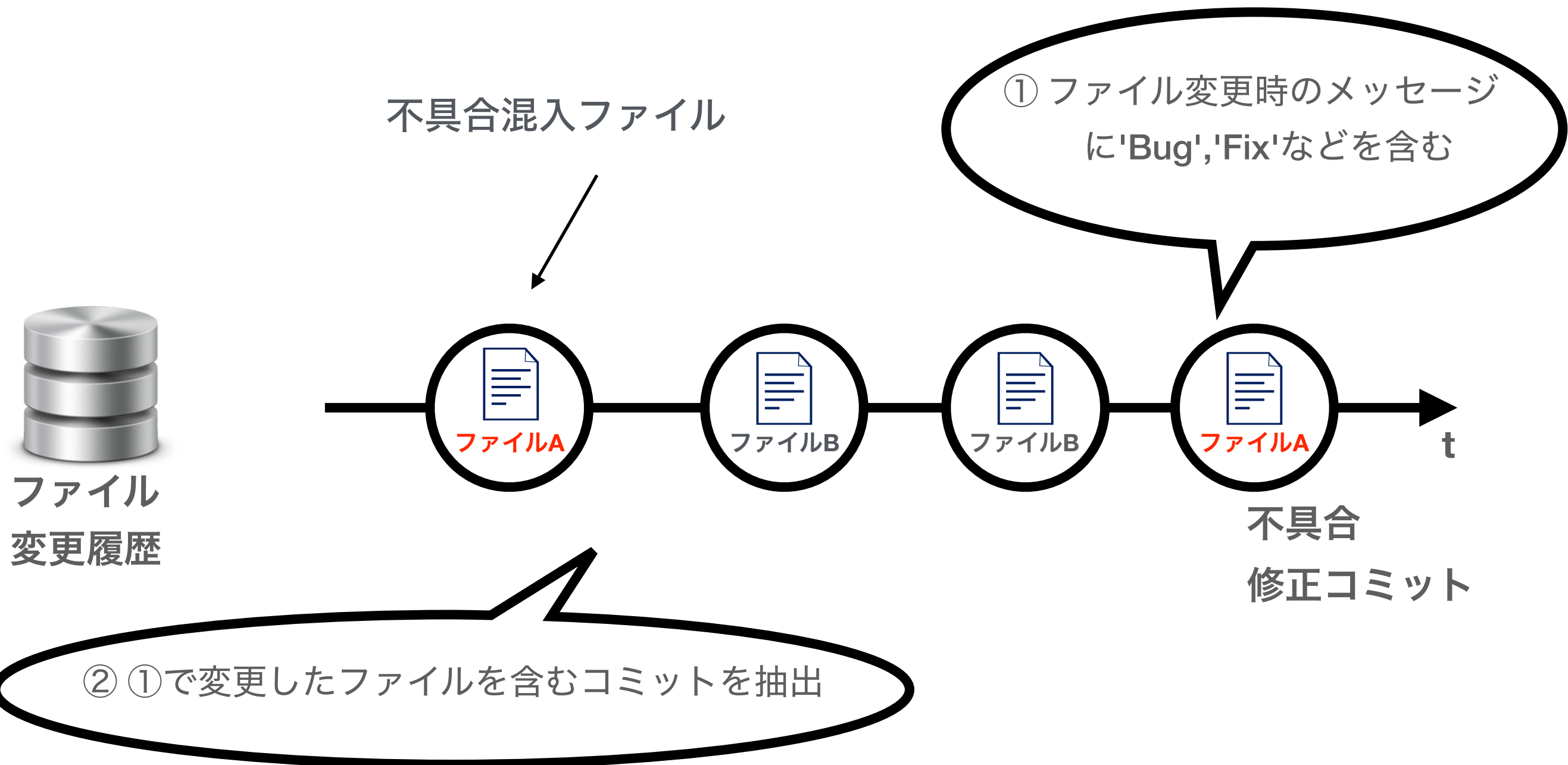
① ファイル変更時のメッセージ
に'Bug','Fix'などを含む



ファイル
変更履歴

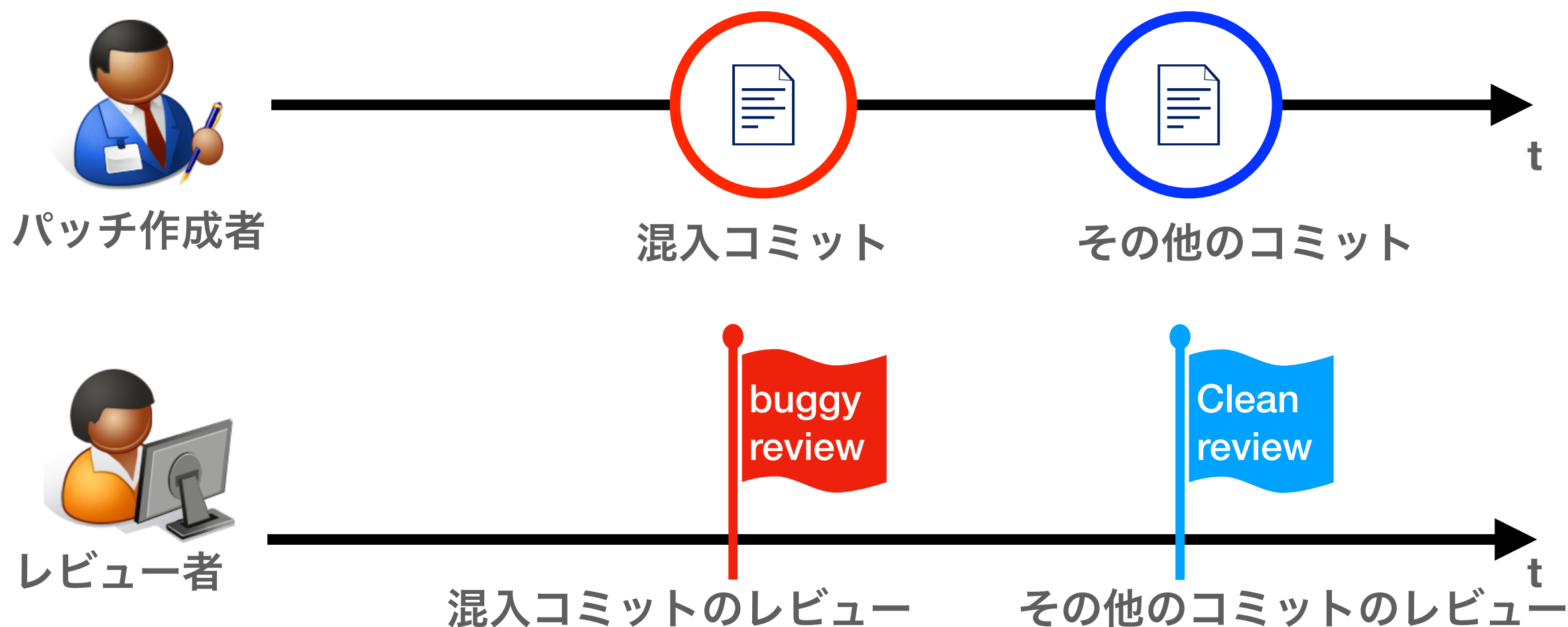


1.1 不具合混入ファイルを特定[4]



1.2 不具合混入コミットからバギーレビューを抽出

- 不具合混入コミットのコードレビューをバギーレビュー, それ以外をクリーンレビューとする



2.1 コードレビュー内のコメントを数値変換する

- ・ テキストのクレンジング後, tf-idfにより各単語をメトリクス化する
 - ・ $m \times n$ 行列 (m =コードレビュー数, n =各単語数)を作成する
- ・ クレンジング内容
 - ・ 大文字→小文字
 - ・ 数字→0に統一
 - ・ ストップワード除去
 - ・ URL除去
 - ・ ステミング

$$w_{i,j} = tf_{i,j} \times \log \left(\frac{N}{df_i} \right)$$

tf_{ij} = number of occurrences of i in j
 df_i = number of documents containing i
 N = total number of documents

2.2 カウントした単語数をもとにモデル構築

- ・ ロジスティック回帰でモデル構築
 - ・ 説明変数：各単語のtf-idf値
 - ・ 目的変数：コードレビューがバグー/クリーン
- ・ コードレビューの半分を訓練データ，残りをテストデータ
- ・ 目的変数に偏りがあるためAUCで評価
 - ・ ランダム分類器のAUC = 0.5

データセット

- 対象プロジェクト
 - OpenStack

プロジェクト名	コードレビュー数 (コメント付き)	バギーレビュー数	コメント数
cinder	5889	2459	696586
devstack	4190	447	87867
heat	2086	432	44517

RQ1:バギー／クリーンなコードレビューの議論内容に違いがあるか

- ・ $AUC > 0.5$ のためバギー／クリーンなコードレビューの議論内容に違いがある

プロジェクト名	単語の種類	AUC
cinder	1699	0.753
devstack	1038	0.687
heat	417	0.730

各単語のオッズ(バギーレビューになりやすい)

cinder

devstack

heat

feature_name feature_odds

feature_name feature_odds

feature_name feature_odds

259	comments	208.580419	155	comments	4.987800	54	comments	4.256308
133	bdd	6.215189	872	successful	3.450477	360	think	3.516758
845	logs	5.847531	429	inline	3.388022	53	comment	3.502362
700	identity	5.009644	125	change	2.469317	393	validate	2.815400
1586	v0	4.067528	279	etc	2.427034	350	template	2.814854
624	glusterfs	3.555204	72	bash0	2.416665	374	tripleo	2.384911
350	datacore	3.136092	818	sh	2.379613	318	seems	2.337819
1100	post_failure	2.892552	326	fine	2.151537	306	resource	2.323608
746	inline	2.839013	937	turn	2.148124	136	go	2.301442
926	ml0	2.715972	277	error	2.080383	13	another	2.246638

comments:インラインコメントをする時の定型表現

各単語のオッズ(クリーンレビューになりやすい)

cinder

	feature_name	feature_odds
1250	requirements	0.053468
1400	sphinx	0.242704
687	hpmsa	0.268123
1150	py0	0.293234
680	horcm	0.327676
1176	rc0	0.418510
354	db	0.421644
1698	zuul	0.447676
820	lgm	0.460718
1624	vsafte0	0.475083

devstack

	feature_name	feature_odds
794	score	0.196280
700	pxe_ssh	0.342052
1030	xenial	0.456356
212	default	0.496806
471	legacy	0.547936
593	odl	0.556019
476	lgm	0.584977
80	better	0.588196
902	thanks	0.611537
832	slow	0.617081

heat

	feature_name	feature_odds
298	releasenotes	0.273399
304	requirements	0.296703
243	openstack	0.372881
367	timeout	0.470468
332	sorry	0.524816
172	know	0.529039
159	install	0.534431
348	sure	0.567222
142	guide	0.636385
323	service	0.638344

今後の予定

- ・ RQ2の調査
- ・ インラインコメントの単語も分析する
- ・