

Predicting developers' negative feelings about code review

Egelman, C. D., Murphy-Hill, E., Kammer, E., Hodges, M. M., Green, C., Jaspán, C., & Lin, J. (2020, June).

In Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering (pp. 174-185).

背景 | コードレビュー

- メリット
 - 欠陥の発見, 知識の共有, メンテナンスの改善
- 職場における結果的なストレス [21]
 - Linuxカーネルのベテラン開発者を失ってしまう [22]

コードレビューをストレスに感じる開発者がいる

[2][7][16]

問題点

コードレビュープロセスにおけるストレスを
定量化できていない

- 倫理的観点から公正で包括的なものである必要がある
 - 組織は才能のある開発者を維持しなければならない
 - 幸福度の高い開発者ほど生産性が向上している [9]
- コードレビューにおけるストレスを軽減する必要あり

研究目的

コードレビューにおける否定的な経験 (Pushback)
を理解し、測定すること

Pushback

コードレビューにおける不必要な対人対立の認識

- Pushbackに関連する5つの主な感情を明確化
- Pushbackを予測するログベースのメトリクスを作成し、検証

研究の流れ

RQ1,2

Googleのコードレビューについてインタビュー

- 否定的な経験と解決策,
Pushbackの定義や対処法を理解

RQ3

Pushbackを予測するためのメトリクス作成

- 会話数, レビューアのレビュー時間, 作成者の対応時間

RQ3

メトリクスの検証

- Pushbackの予測に有効なのか

RQ1：コードレビューに否定的な経験はどの程度の頻度であるか

コードプロセスに中程度満足	87% ($\pm 2\%$)
コードの品質が中程度改善された	96% ($\pm 1\%$)
1回/3ヶ月 で否定的な経験をした	56% ($\pm 3\%$)
1回/1ヶ月で否定的な経験をした	26% ($\pm 3\%$)

否定的な経験

過剰なレビュー遅れ, 過剰な指摘,
レビュー開始までの長い待ち時間, 対立的なコメント

RQ1 : Pushbackに関する5つの感情

Pushbackの感情

メトリクスの検証に用いる

- ・ 対人関係の対立
- ・ 受け入れがあまりにも長い間保留されたと感じること
- ・ レビューアーが過剰な変更を要求した
- ・ 今後のコードレビューについて否定的に感じる
- ・ 不満 (イライラ)

※ 5つの感情は排他的ではない
pushbackの潜在的な感情を網羅しているとは考えていない

RQ2 : Pushbackが発生する要因は？

定性分析の結果

- コードスタイルに関するコメント
- フィードバックが伝達されるトーン
- コードベースに対する馴染みのなさ
- 提案や要求に対する根拠の欠如



レビュアーの個人的な意見が多い

レビュアーがコードベースに
精通していない

ガイドラインに書かれていることを
遥かに超えた批判をする

正当な理由を示さずに些細な変更を
何度も指摘する

RQ3: Pushbackを予測するためのメトリクスを作成

- 既存のレビューアー推奨ツールのログから計算できる
3つの潜在的なメトリクスを作成

		48分	112分	9回
		レビュー 時間	対応時間	会話数
0 Flags	250			
	200	✓		
	200		✓	
1 Flag	200			✓
	100	✓	✓	
	100	✓		✓
2 Flags	100		✓	✓
	100	✓		
	100		✓	
3 Flags	100	✓	✓	✓

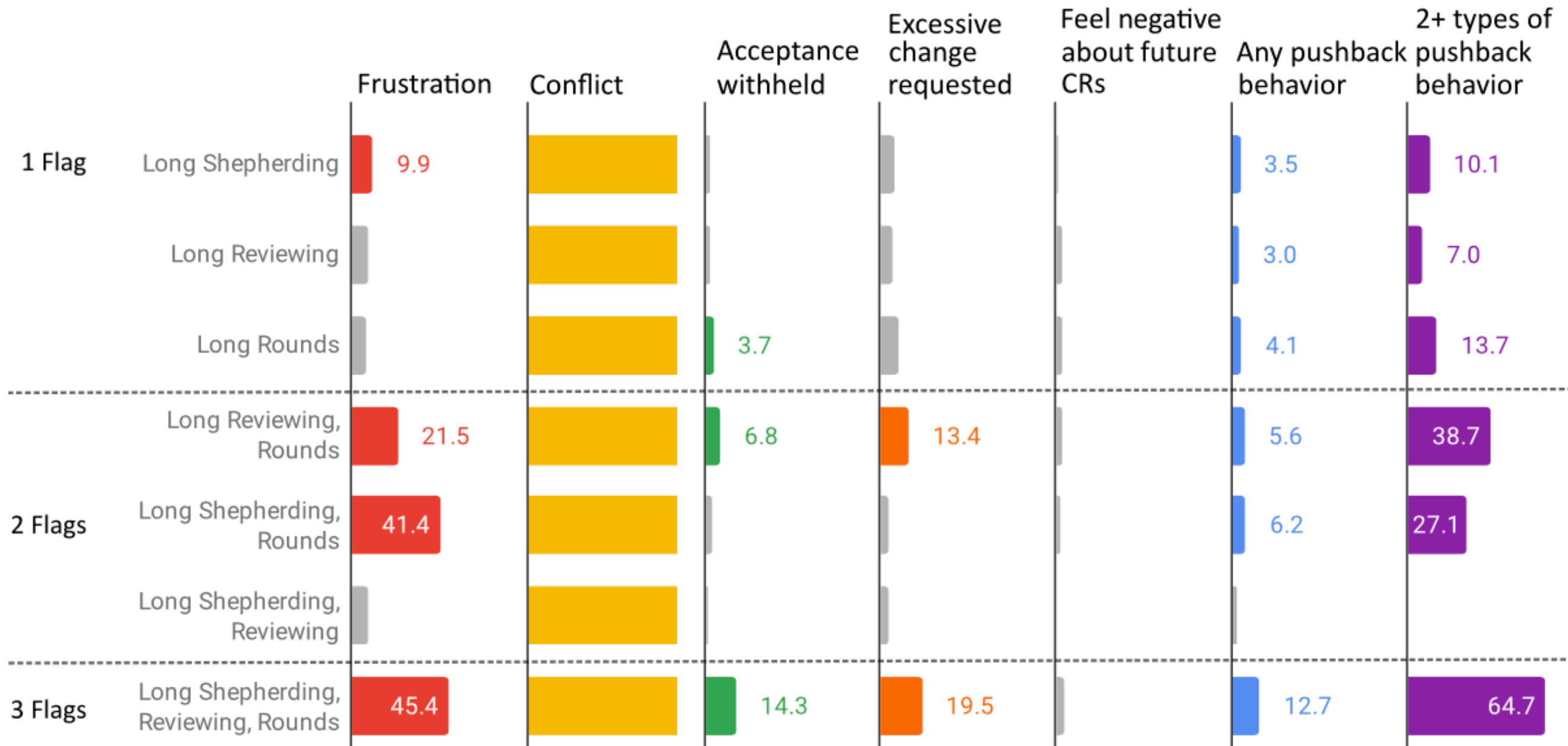
RQ3: Pushbackを検出するメトリクスの検証

pushback発生率を比較

Feeling of Pushback	0 Flags	3 Flags
frustration	1%	27%
interpersonal conflict	0%	22%
acceptance was withheld for too long	4%	27%
reviewer asked for excessive changes	1%	16%
feeling negatively about future CRs	4%	8%
any pushback behavior	7%	49%
2+ types of pushback behavior	1%	27%

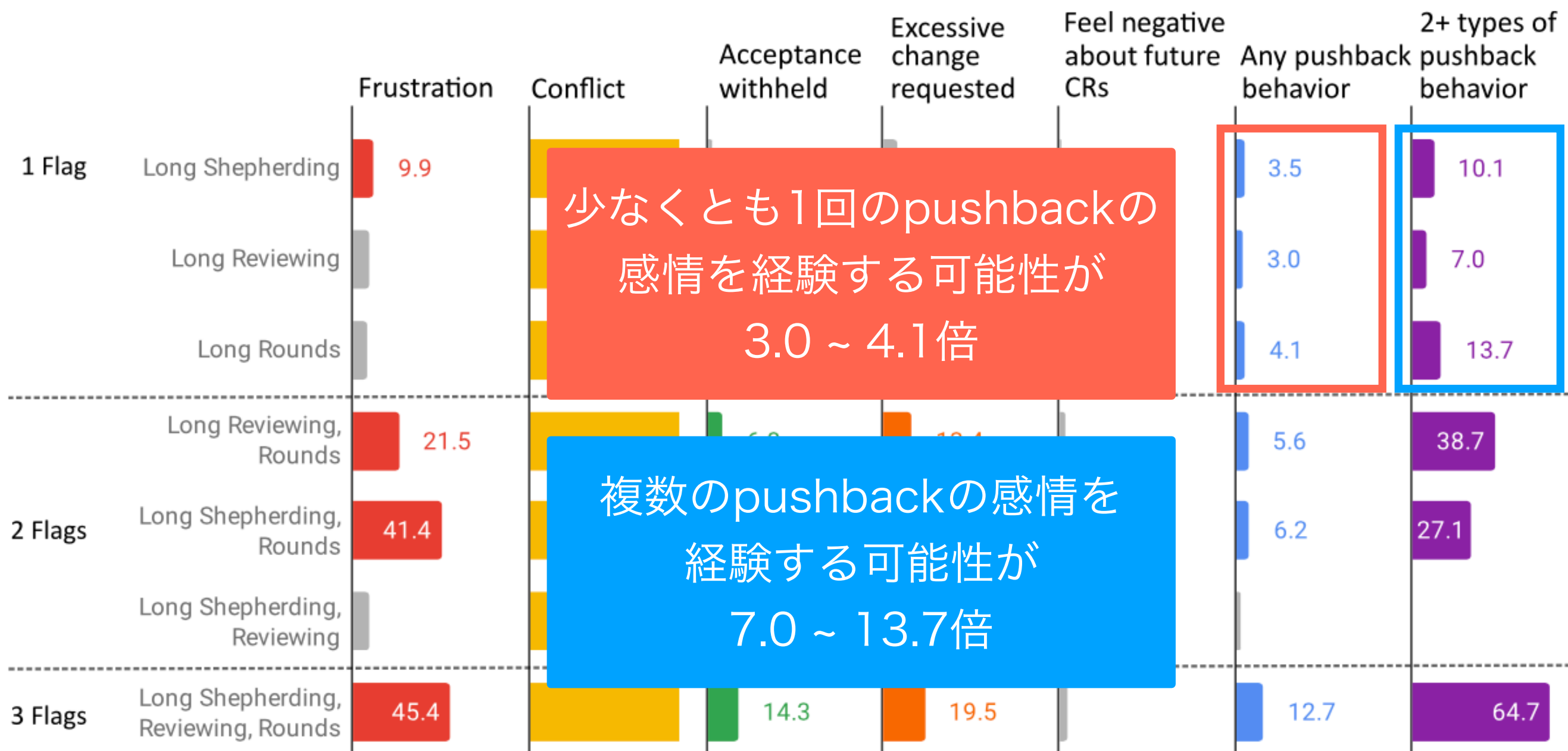
メトリクスが効果的であることを示唆

RQ3: Pushbackを検出するメトリクスの検証



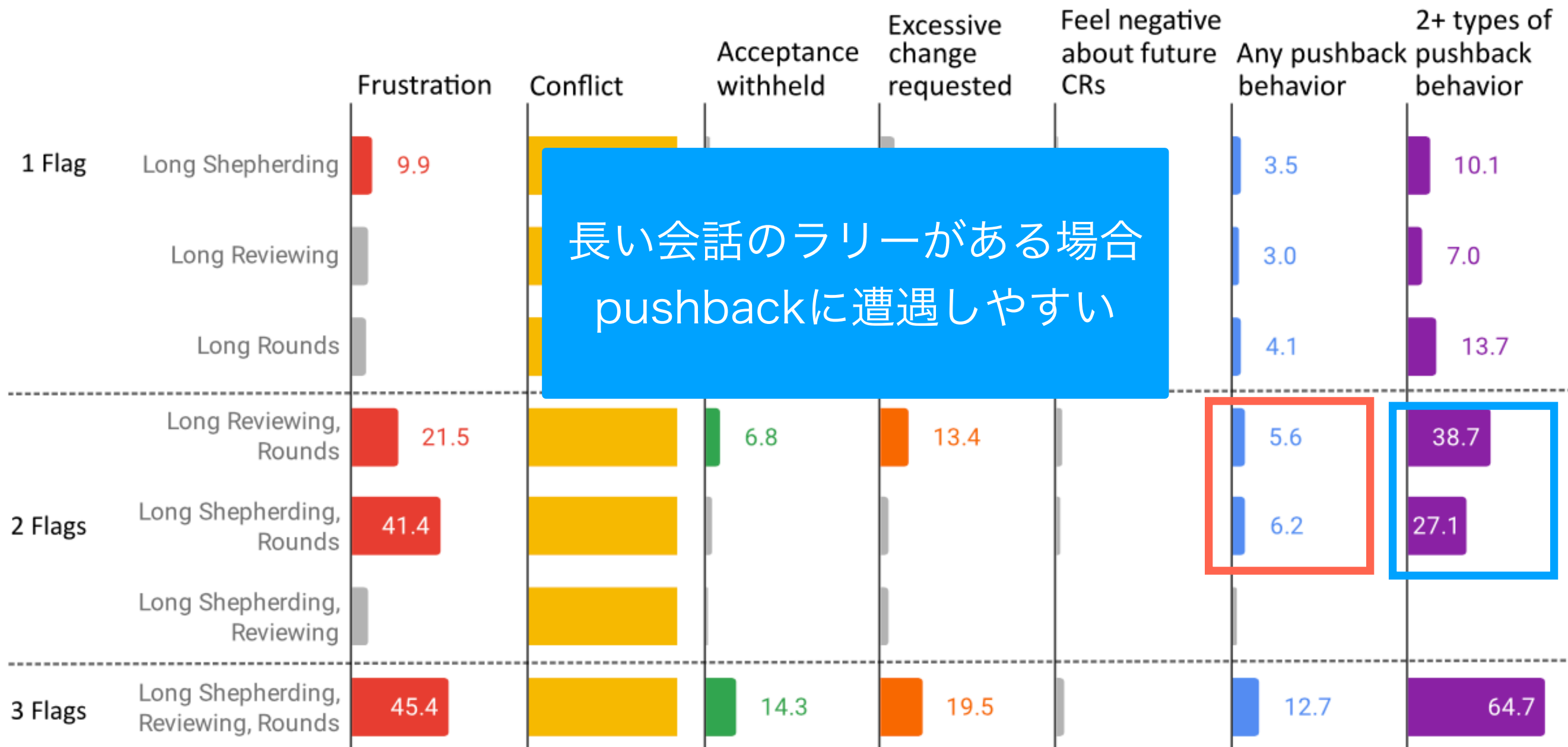
ベースラインカテゴリがフラグなしである場合の
適合されたロジスティック回帰モデルのオッズ比

RQ3 | 0 flag vs 1 flag



ベースラインカテゴリがフラグなしである場合の
適合されたロジスティック回帰モデルのオッズ比

RQ3 | 0 flag vs 長い会話数を含む 2 flag



レビュー時間や対応時間が長くともフィードバックが簡潔であればpushbackに遭遇しない可能性がある

まとめ

- コードレビューにおけるネガティブな経験は頻度は少ないものの存在する一定数存在する
- ネガティブな感情(pushback)を予測するために有効なメトリクスの提案
 - 会話数, レビュー時間, 対応時間

※付録にアンケート内容や, 回帰分析結果等有

所感

- 論文内の3つのメトリクスのみでは不十分であると述べられていたが、他にどのようなメトリクスであれば有効なのか気になった.
- 複数コミュニティに属している開発者がいるとしたら、コードスタイルをコミュニティごとに合わせるのは大変だろうなと思った.
- 追加資料の値を理解できなかった. . .

Pushback Feeling	Precision	Recall
frustration	0.11	0.98
interpersonal conflict	0.10	1.00
acceptance was withheld for too long	0.10	0.93
reviewer asked for excessive changes	0.07	0.97
feeling negatively about future CRs	0.06	0.87