

OSSプロジェクトにおける 開発者の特性を考慮した コミッター候補者予測

M2 松本 明

背景

- プロジェクトの保守・進化のためにコミッターは必要不可欠な存在
- コミッター不足が問題
 - ⇒少数のコミッターにタスクが集中
 - ⇒レビュー時間の長期化

コミッターへの昇格

- コミッター選出は, プロジェクト管理者や既存のコミッターによって行われる[1]
- プロジェクトへの**貢献**が認められればコミッターとして選出される



①パッチの作成・投稿



②投稿されたパッチのレビュー&フィードバック



③不具合に対する議論(コメントの投稿)

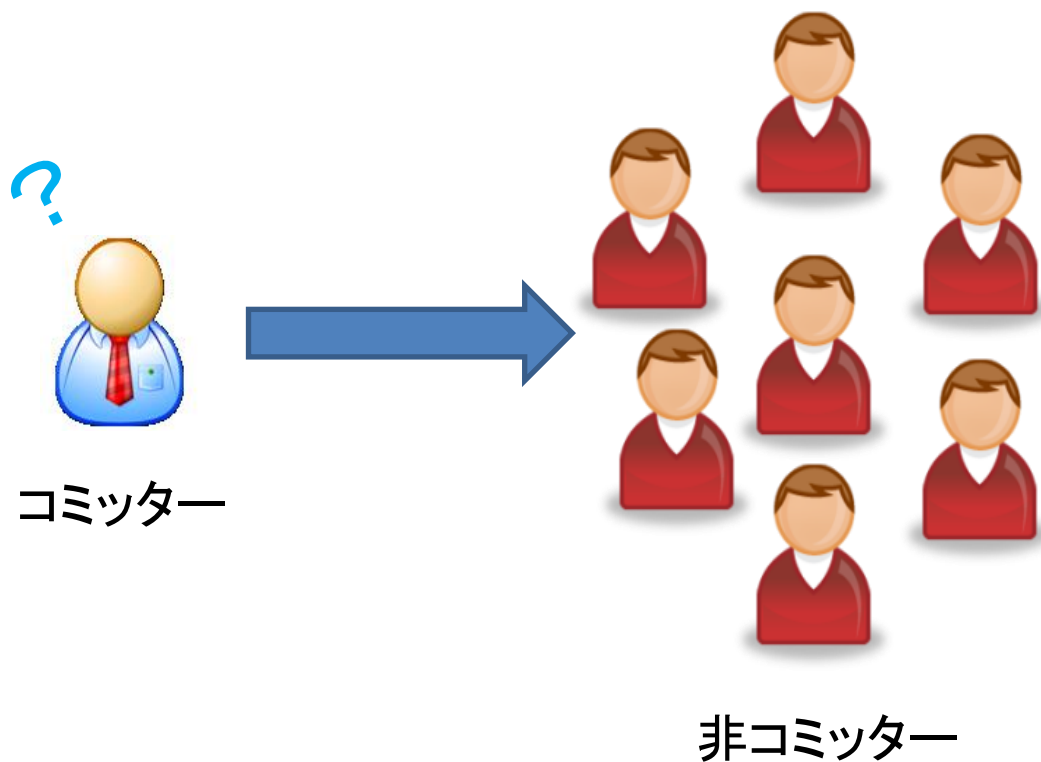


④パッチの妥当性確認後, 構成管理ツールへコミット

- コミッター選出のための客観的指標は存在しない
 - プロジェクト管理者は経験的にコミッター選出を行っている
 - 継続的参加に対する開発者(非コミッター)への動機付けが弱い

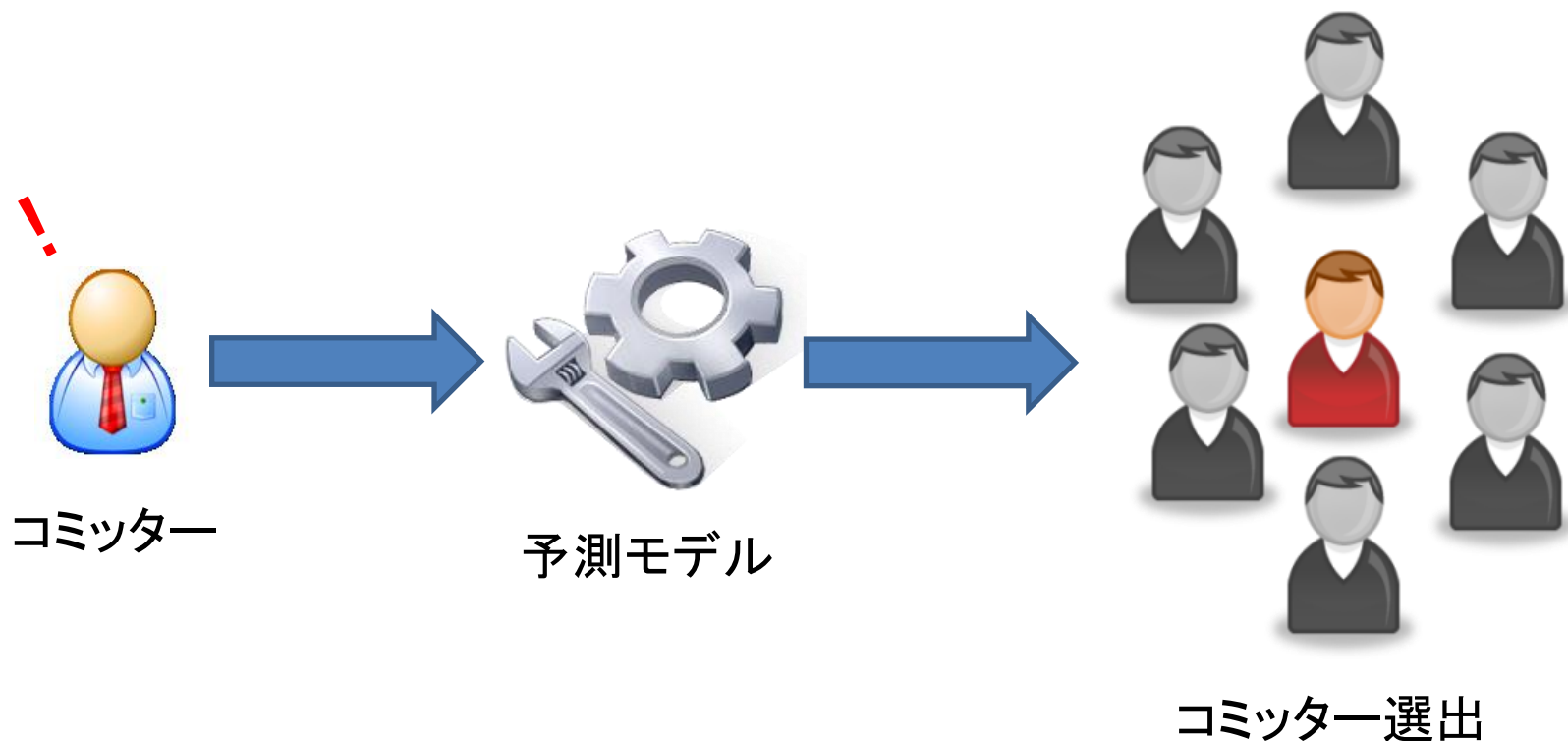
課題

- コミッターにふさわしい開発者を見つけ出すのにはコストがかかる



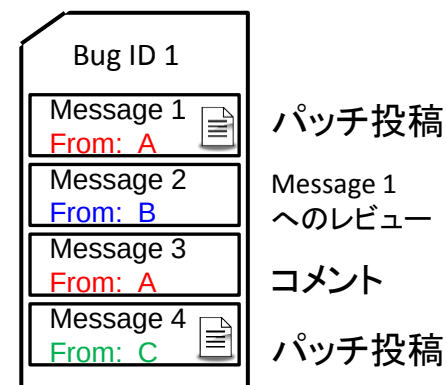
研究目的

- 開発者の特性を考慮したコミッター候補者予測モデルの構築

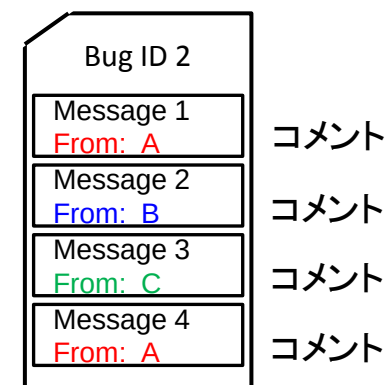


既存研究における活動内容の定義

- 活動期間(月)
 - メッセージ投稿を行っている月数
- 活動量
 - パッチ投稿数(月平均・総数)
 - レビュー数(月平均・総数)
 - コメント数(月平均・総数)



パッチを含む不具合票



パッチを含まない不具合票

コミッター候補者予測の結果

表 6 コミッター候補者の予測結果 (Mozilla Firefox)

Table 6 Committer prediction results in Mozilla Firefox project.

		適合率	再現率	F1 値	正確に予測できた コミッター候補者数
本実験モデル (全メトリックス) を用いた予測		0.02	0.83	0.05	20
単一メトリックス を用いた予測	総コメント投稿数	0.05	0.67	0.09	16
	月コメント投稿数	0.00	0.54	0.01	13
	総パッチ投稿数	0.05	0.54	0.10	13
	月パッチ投稿数	0.06	0.46	0.11	11
	総パッチ検証数	0.05	0.67	0.10	16
	月パッチ検証数	0.04	0.38	0.07	9
	活動期間	0.02	0.79	0.04	19
ランダム予測		0.00	0.50	0.01	13
向上率 ^(注4)		5.13	1.66	5.00	1.54

(注 4) 本論文のモデルを用いて予測する場合がランダムに予測した場合に比べて, どの程度精度が向上したかを意味する.

新たな開発者特性の定義(1/2)

- 仮説をたてて分析し、予測モデルのメトリクスとして用いるかを検討
- 貢献活動間隔
 - 貢献活動を一定におこなう開発者はプロジェクトを支えるコミッターになりうるのでは？
- 特定コンポーネントへの専門性
 - 特定コンポーネントへの貢献度が高い開発者は、プロフェッショナルであるとみなされ、コミッターになりうるのでは？

新たな開発者特性の定義(2/2)

- 他プロジェクトでのコミッター兼任・経験
 - 他プロジェクトでコミッターを兼任していたり, 経験がある開発者はスキルが十分である(また, コミッターとしてのノウハウがある)とみなされ, コミッターになりうるのでは?
- その他, 考え中

前回からの進捗

- ASFのプロジェクト10個のデータ(Jira,Git)を取得
 - 報告が1000件以上かつGitのコミットと関連付けられるプロジェクト
- データクレンジング, 開発者名の名寄せ

今後の予定

	4		5		6	
分析メトリクスの決定						
分析準備						
予測モデルの構築						
予測モデルの評価						
大ゼミ						
？						

どこで学外発表しよう...