

OSS開発コミュニティの継続性と 開発者のPolitenessの関係分析

オープンソースソフトウェア工学研究室

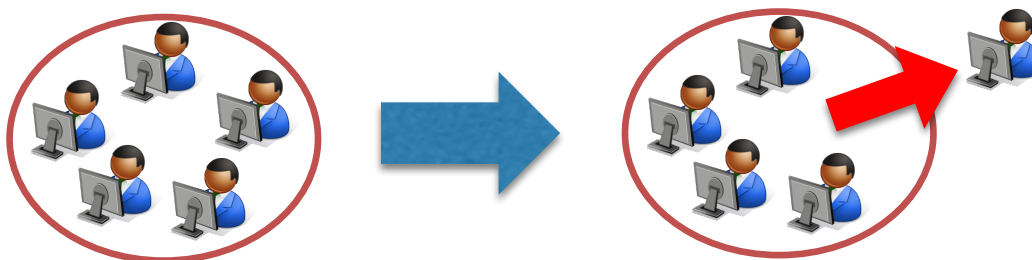
60181052 宮崎智己

背景

- 生産性向上のために，OSSを再利用する企業が増えてきている
- OSSは継続的に開発・保守されるかわからない
 - 多くの開発者はボランティア開発者
 - ごく一部の開発者が継続的に活動
- どのOSSを再利用すべきかわからない

開発者の活動の継続性の研究

- 開発者の平均活動期間は約1.5年[1]
 - 多くの開発者は参加してすぐに離脱してしまう
 - 長くプロジェクトに貢献してもらうためには、1.5年以内に重要な役割を与えることが必要



なぜ開発者が離脱してしまうかは
明らかにされていない

[1] Bird, C., Gourley, A., Devanbu, P., Swaminathan, A. and Hsu, G.: Open Borders? Immigration in Open Source Projects, Proceedings of the 4th International Workshop on Mining Software Repositories (MSR'07), p. 6 (2007)

考えられる開発者の主な離脱要因

- 開発スキルの不足
- 開発スタイル・方向性の違い
- 人間関係での問題
 - コミュニティに馴染めない
 - 他の開発者と口論

本研究のモチベーション

- OSS開発は協調作業なため，継続的に活動をおこなうためにはコミュニティ内の人間関係が重要
- 良好な人間関係を確立するためには，他の開発者と配慮のあるやり取りが必要
- 継続的に活動を行なっている開発者は他の開発者への配慮を行えているからではないか

本研究の目的とアプローチ

- 目的
 - 開発者の配慮と開発者の活動継続性が関係あるかどうかを明らかにする
- アプローチ
 - 開発者の配慮をPolitenessで定義
 - 開発者のPolitenessと活動期間の関係を分析

リサーチクエスチョン

- RQ1: 長期開発者の受信メールのPolitenessと短期開発者の受信メールのPolitenessに違いはあるか？
- RQ2: 長期開発者の送信メールのPolitenessと短期開発者の送信メールのPolitenessに違いはあるか？

Politeness [3]

- 相手との円滑な人間関係を確立・維持しつつコミュニケーションを行うための配慮
- Politenessの算出手順
 - 入力テキストからPoliteness Strategyを正規表現を用いて抽出
 - 抽出したPoliteness StrategyをSVMに入力してPoliteness値を0~1の値で算出

[3]Brown, P., & Levinson, S. (1978). Universals in language usage::Politeness phenomena. In E. Goody (Ed.), Questions and politeness: Strategies in social interaction (pp. 56-310). Cambridge: Cambridge University Press.

Politeness Strategy

- Politenessを満たす20の要素
 - 単語の意味や位置によって分類

Strategy	主な単語,文	
Gratitude	thank, appreciate	Politenessを 高めるStrategy
deference	great, good, nice	
Greeting	hi, hey, hello	
hedges	think, assume, feel	
Please start	Please ~,	Politenessを 低くするStrategy
Factuality	really, actually, honestly	
Direct question	What ~?, Why ~?. Where ~?	
2 nd person start	You ~.	

ケーススタディ

- データセット

プロジェクト	長期開発者	短期開発者
Apache HTTP Server	183人	173人
Python	227人	183人

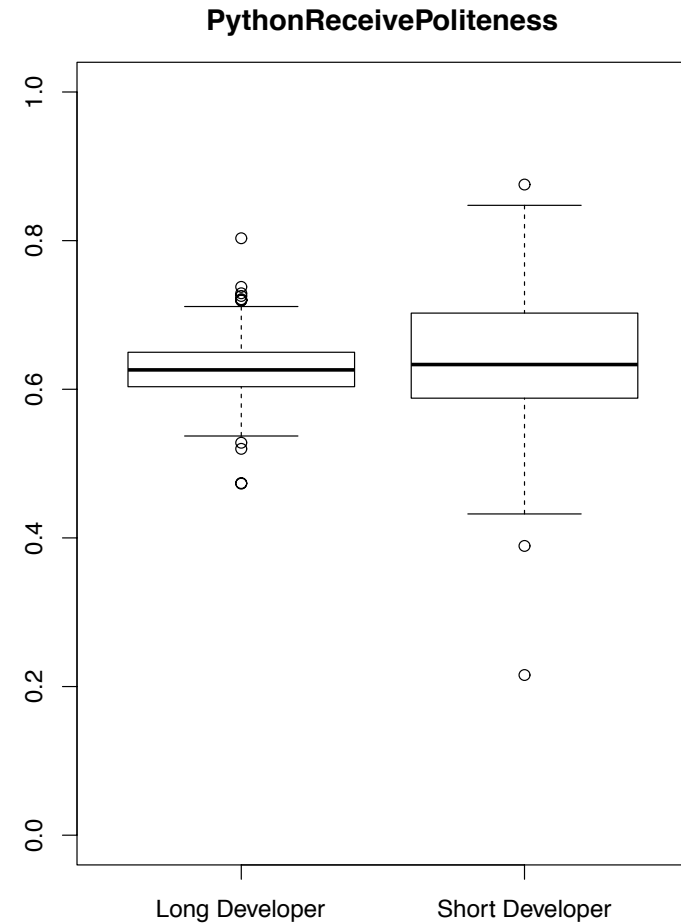
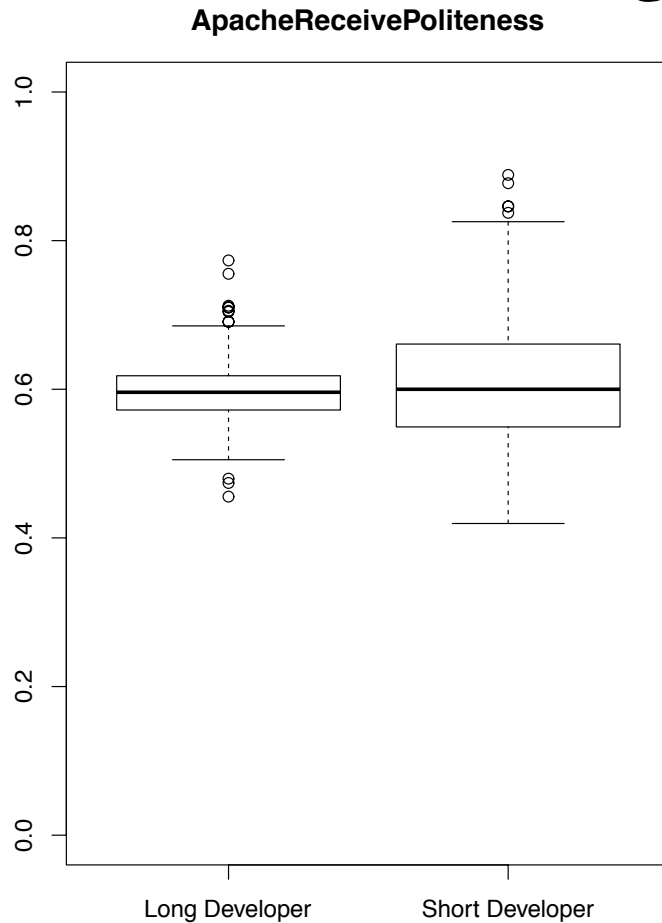
長期開発者：活動期間270日以上開発者

短期開発者：活動期間90日以上270日未満の開発者

- 分析方法

- 長期開発者のPoliteness値と短期開発者のPoliteness値の平均値を比較
- マンホイットニーのU検定で検定

RQ1 : 結果

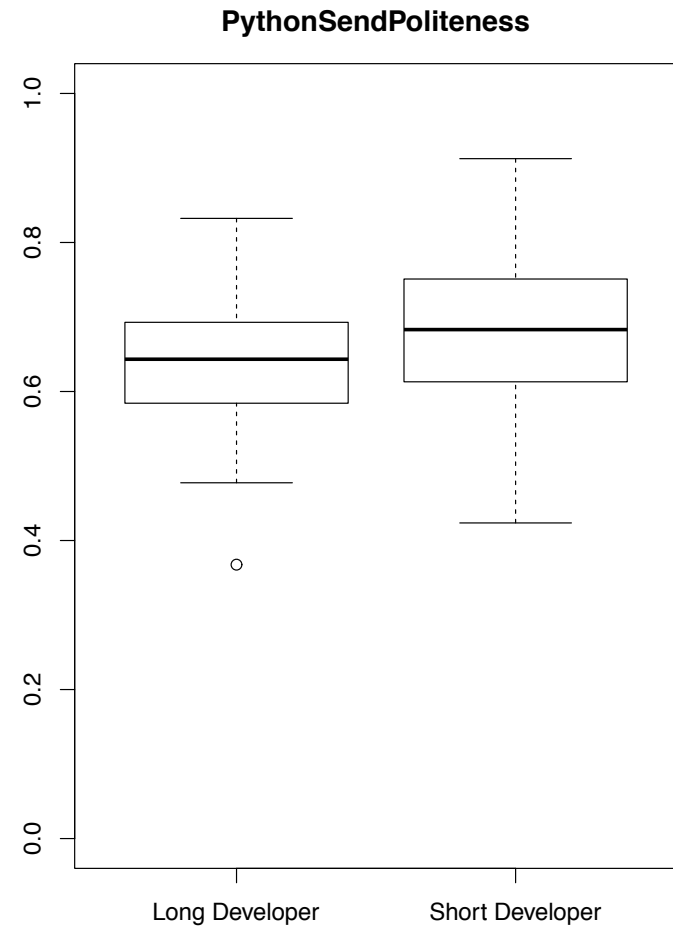
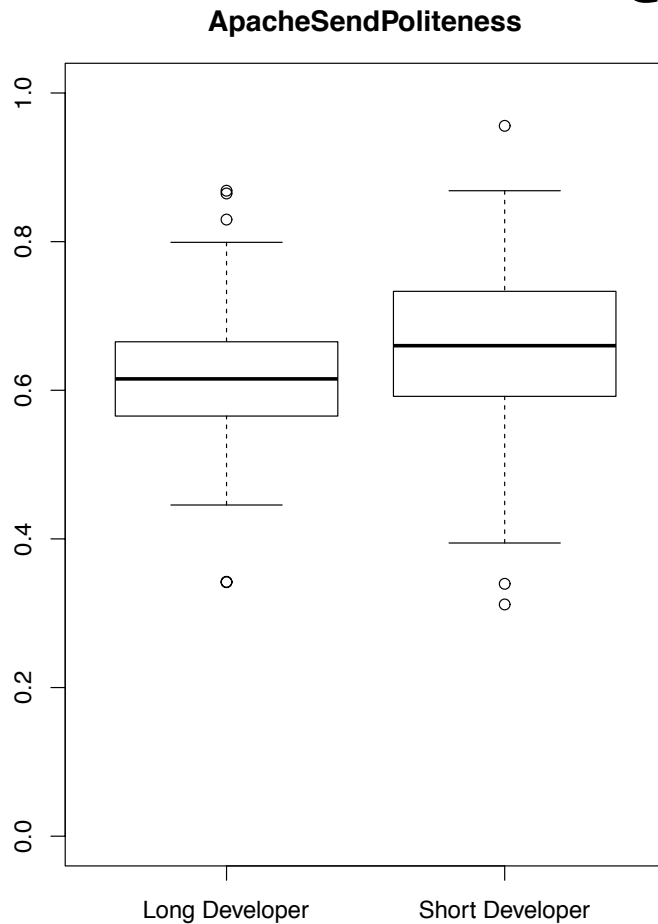


RQ1 : 結果

	長期開発者			短期開発者			p値
	中央値	最大値	最小値	中央値	最大値	最小値	
Apache	0.60	0.77	0.46	0.60	0.89	0.42	0.44
Python	0.63	0.80	0.47	0.63	0.88	0.22	0.06

長期開発者の受信メールのPoliteness値と
短期開発者のPoliteness値の違いは確認できなかった

RQ2 : 結果

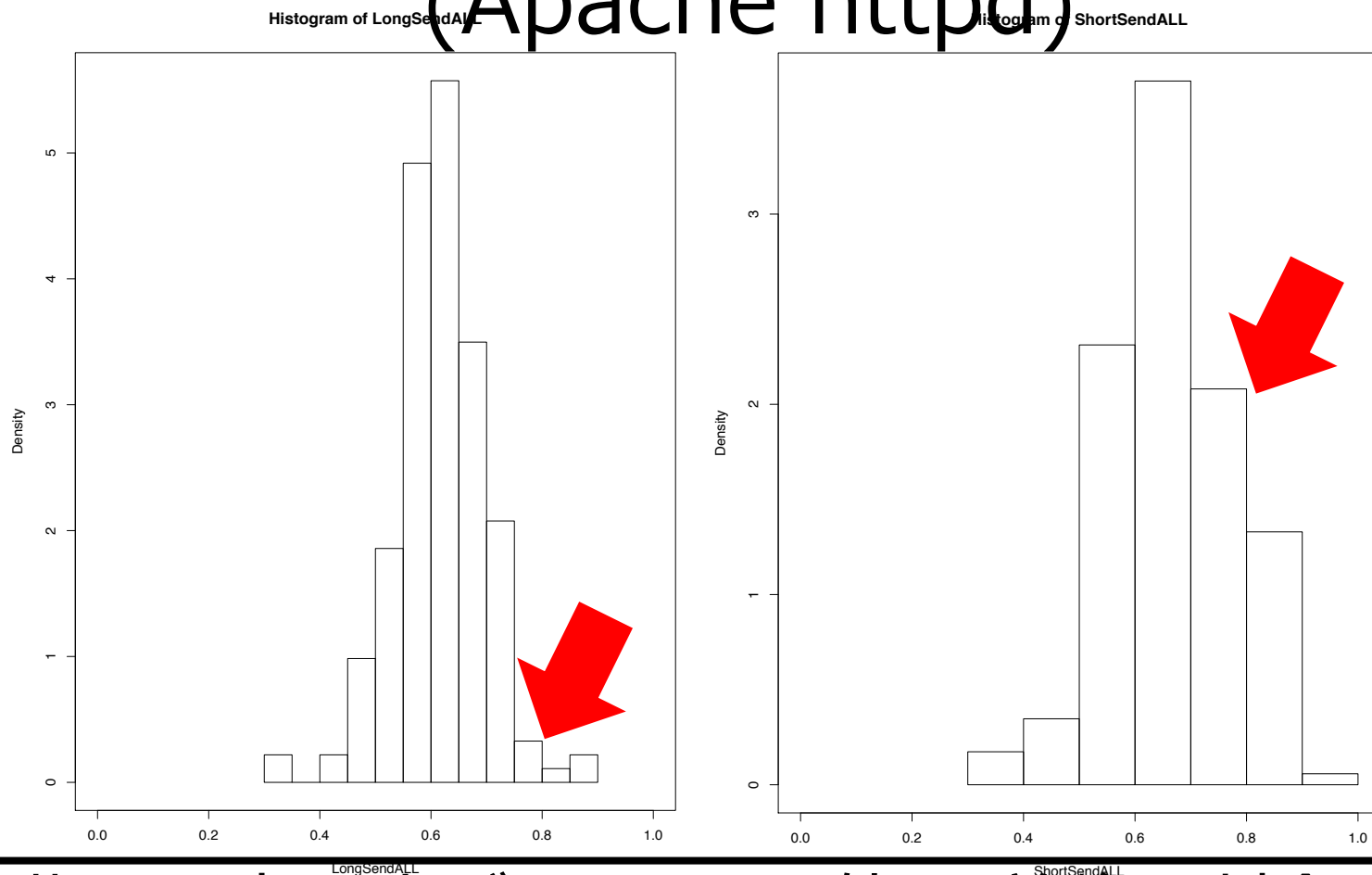


RQ2 : 結果

	長期開発者			短期開発者			p値
	中央値	最大値	最小値	中央値	最大値	最小値	
Apache	0.62	0.87	0.34	0.66	0.96	0.31	0.01
Python	0.64	0.83	0.37	0.68	0.91	0.42	0.00

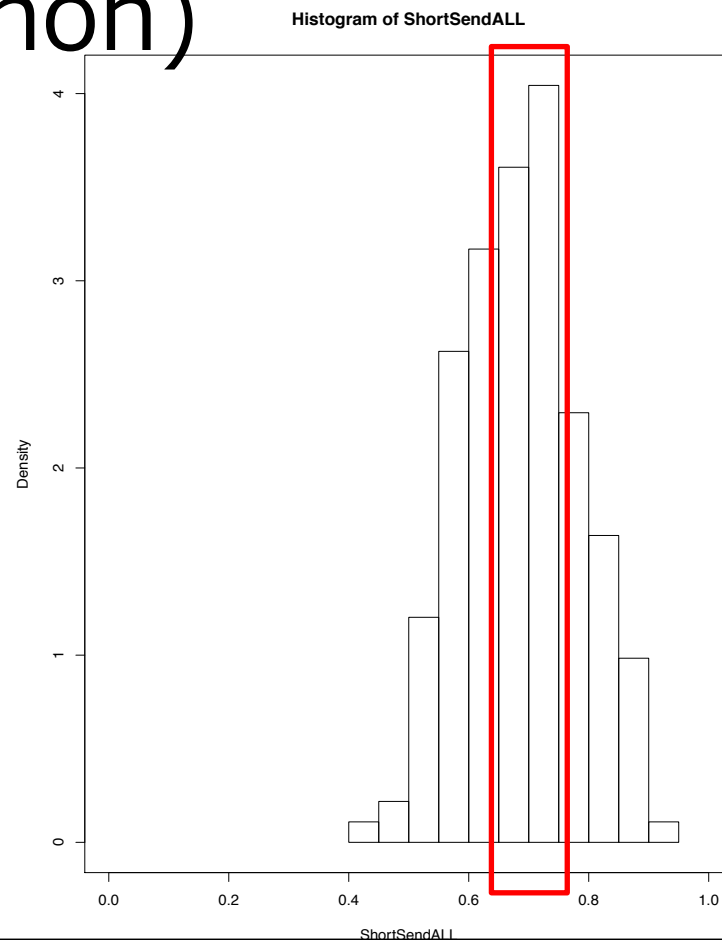
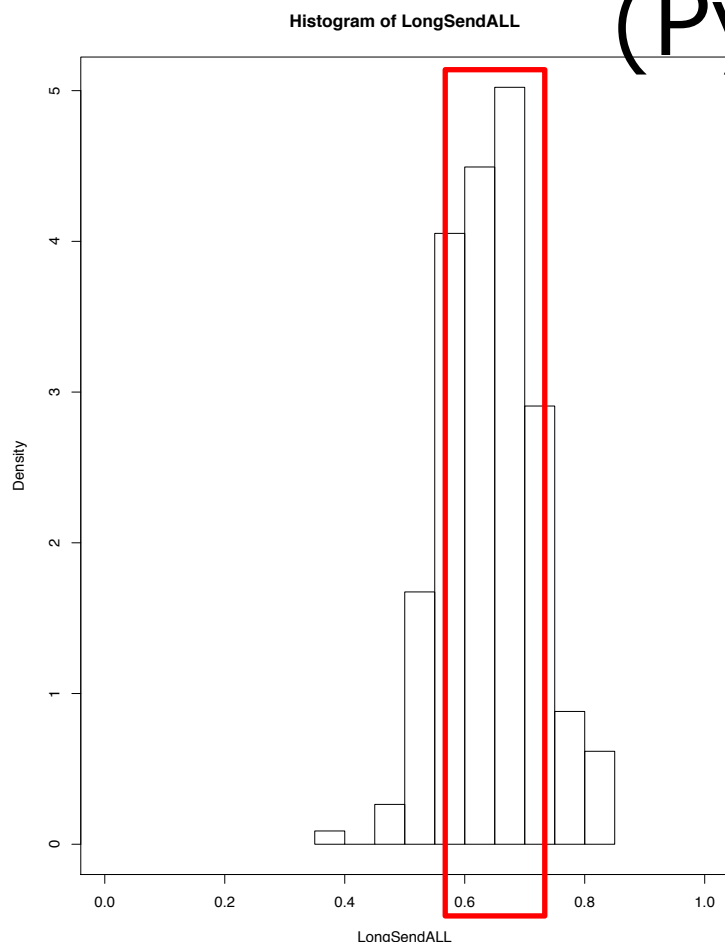
長期開発者の受信メールのPoliteness値と
短期開発者のPoliteness値に違いが確認できた

送信メールのPoliteness値の分布 (Apache httpd)



短期開発者の方がPoliteness値0.8前後の割合が多い

送信メールのPoliteness値の分布 (Python)



長期開発者のうち0.65前後の割合が最も多いが、
短期開発者では0.70前後の割合が最も多い

考察

- RQ1とRQ2の結果から、他者からの振る舞いよりも自分の振る舞いの方が重要だと考えられる
- Politenessは高くても低くてもダメ
 - 短期開発者群の方がPoliteness値の高い開発者の割合が多い
 - 送信メールのPoliteness値0.4未満の開発者はわずかしかない

まとめ

- 開発者のPolitenessが開発者の活動継続性との関係を明らかにするために、長期開発者と短期開発者のPolitenessを比較した
- 長期開発者と短期開発者の送信メールのPolitenessに違いがあることがわかった