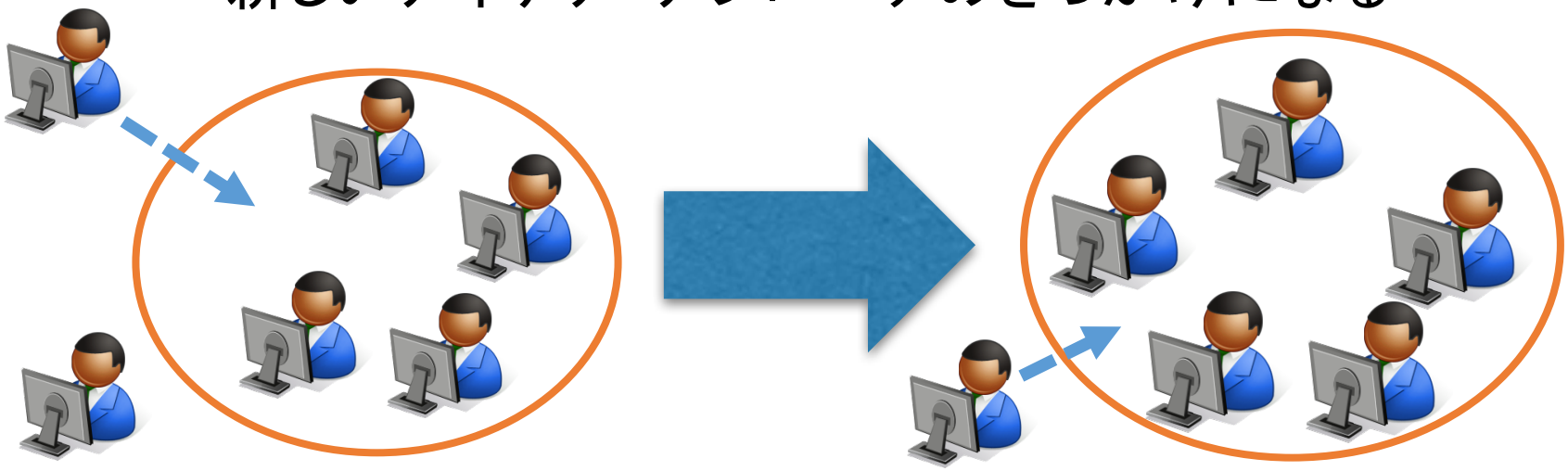
Politeness分析を用いたOSS開発コミュニティのコミュニケーション分析

宮崎　智己

背景

- OSSは企業におけるソフトウェア開発においても利用されている
- OSSは継続的に発展・保守されている必要がある

- 新人開発者の獲得はOSSの発展に重要な要因の一つ
  - 新しいアイデア・アプローチのきっかけになる



現状の問題点

- 新人開発者のプロジェクト参加を妨げる参入障壁が存在すると言われている[1]

- コミュニケーション障害
  - 新人開発者はメーリングリストにメールを送りにくい
  - 新人開発者が送ったメールに対して返信がない/無愛想な返事が返ってくる

実際にコミュニケーション障害が
新人開発者の参加を妨げているかは明らかでない

[1] Igor Steinmacher, Tayana Uchoa,Christoph Treude and Marco Aurélio Gerosa:Overcoming open source project entry barriers with a portal for newcomers Proceeding ICSE '16 Proceedings of the 38th International Conference on Software Engineering Pages 273-284

本分析の目的とアプローチ

- 目的
  - コミュニケーション障害は実際に新人開発者のプロジェクト参加を妨げているかどうかを明らかにする
- アプローチ
  - コア開発者が他の開発者とどのような様式でコミュニケーションを行っているかをPoliteness分析を用いて分析する
  - 分析結果からコア開発者が新人開発者とコミュニケーションを行っている場合と，一般/コア開発者とコミュニケーションを行っている場合とを比較

リサーチクエスチョン

- RQ１：1年未満でプロジェクトを離脱してしまった開発者と一年以上活動している開発者の間で，最初に送ったメールのPoliteness値に違いはあるのか？

- RQ２：1年未満でプロジェクトを離脱してしまった開発者と1年以上活動している開発者の間で，最初にコア開発者から受け取ったメールのPoliteness値に違いがあるのか？
  - RQ２−１：コア開発者全体で違いはあるのか？
  - RQ２−２：コア開発者ごとに違いはあるか？

リサーチクエスチョン

- RQ1：1年未満でプロジェクトを離脱してしまった開発者と一年以上活動している開発者の間で，最初に送ったメールのPoliteness値に違いはあるのか？

- 動機
  - コミュニティへの最初のメールは送りにくいという障害が本当かどうかを明らかにするため
  - 障害があるのであれば，使われている言葉のニュアンスに変化がある

リサーチクエスチョン

- RQ2：1年未満でプロジェクトを離脱してしまった開発者と1年以上活動している開発者の間で，最初にコア開発者から受け取ったメールのPoliteness値に違いがあるのか？
  - RQ2-1：コア開発者全体で違いはあるか？
  - RQ2-2：コア開発者ごとに違いはあるのか？
- 動機
  - 古参開発者からの配慮のないメールは新人開発者のやる気を削ぐという参入障壁が本当にあるかを明らかにするため
  - 古参開発者の中には新人開発者に対して丁寧に扱っている者/扱っていない者もいるのではないか

コア開発者・新人開発者の定義

- コア開発者
  - 1995年3月～2016年3月の間の送信メール数TOP20の開発者
  - プロジェクト全体のメール数の50%を占めている
- 新人開発者
  - 上記の期間で活動期間が1年未満の者
  - １０日未満の者は除く
- 一般開発者
  - 上記の分類以外の開発者

分析手順

1. メールアドレスから開発者を特定

2. 開発者ごとに，一番最初に送ったメールと一番最初にコア開発者から受け取ったメールのPoliteness値を取得

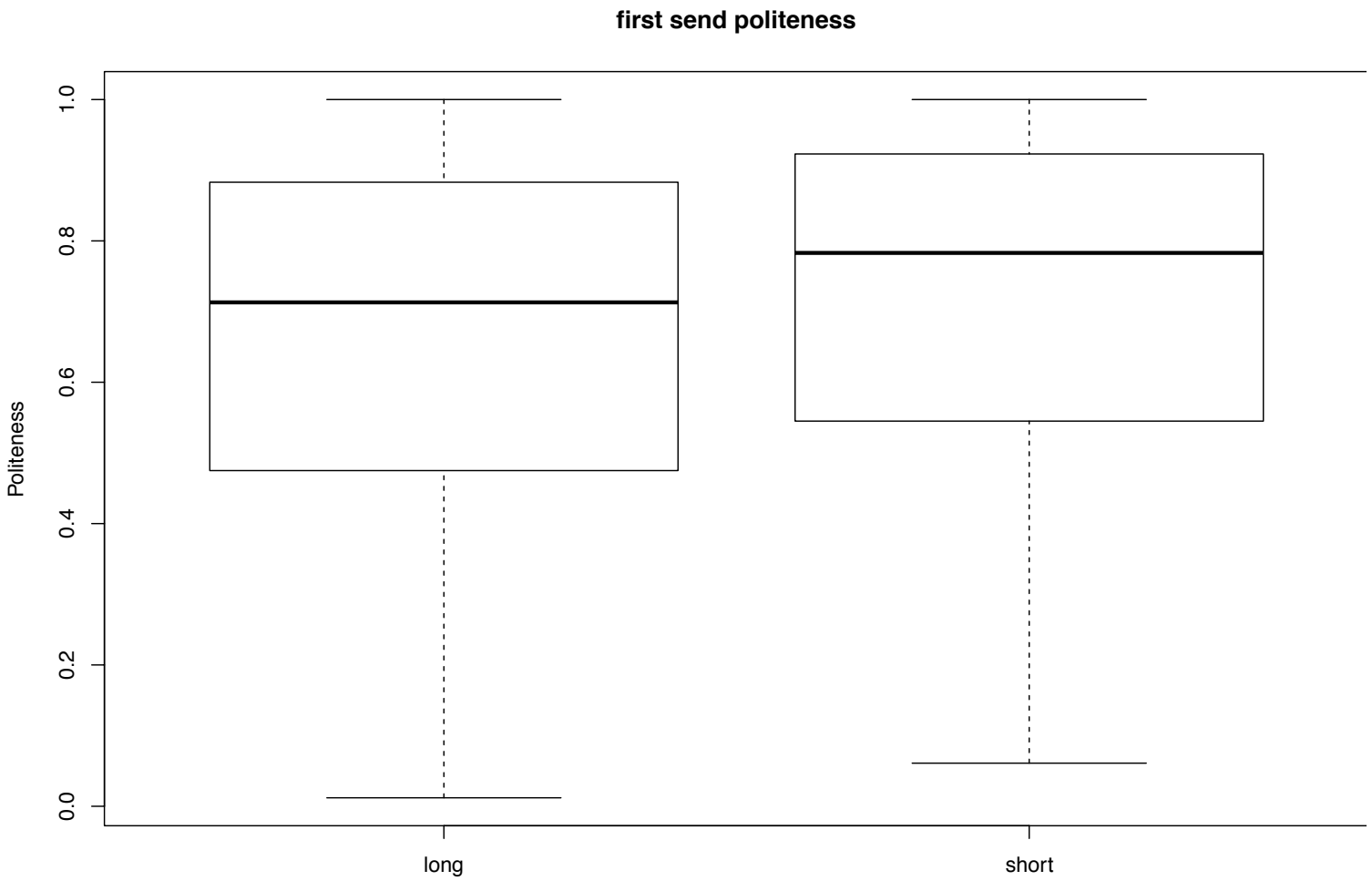
3. 一般・コア開発者と新人開発者で違いがあるのかをウィルコクソンの順位和検定で検定
   - 有意水準５％

RQ1の結果

開発者	中央値	P値
コア・一般開発者	0.713	8.551e-05
新人開発者	0.783	

1年以上活動している開発者のほうが
新人開発者よりPoliteness値が低い傾向がある

箱ひげ図

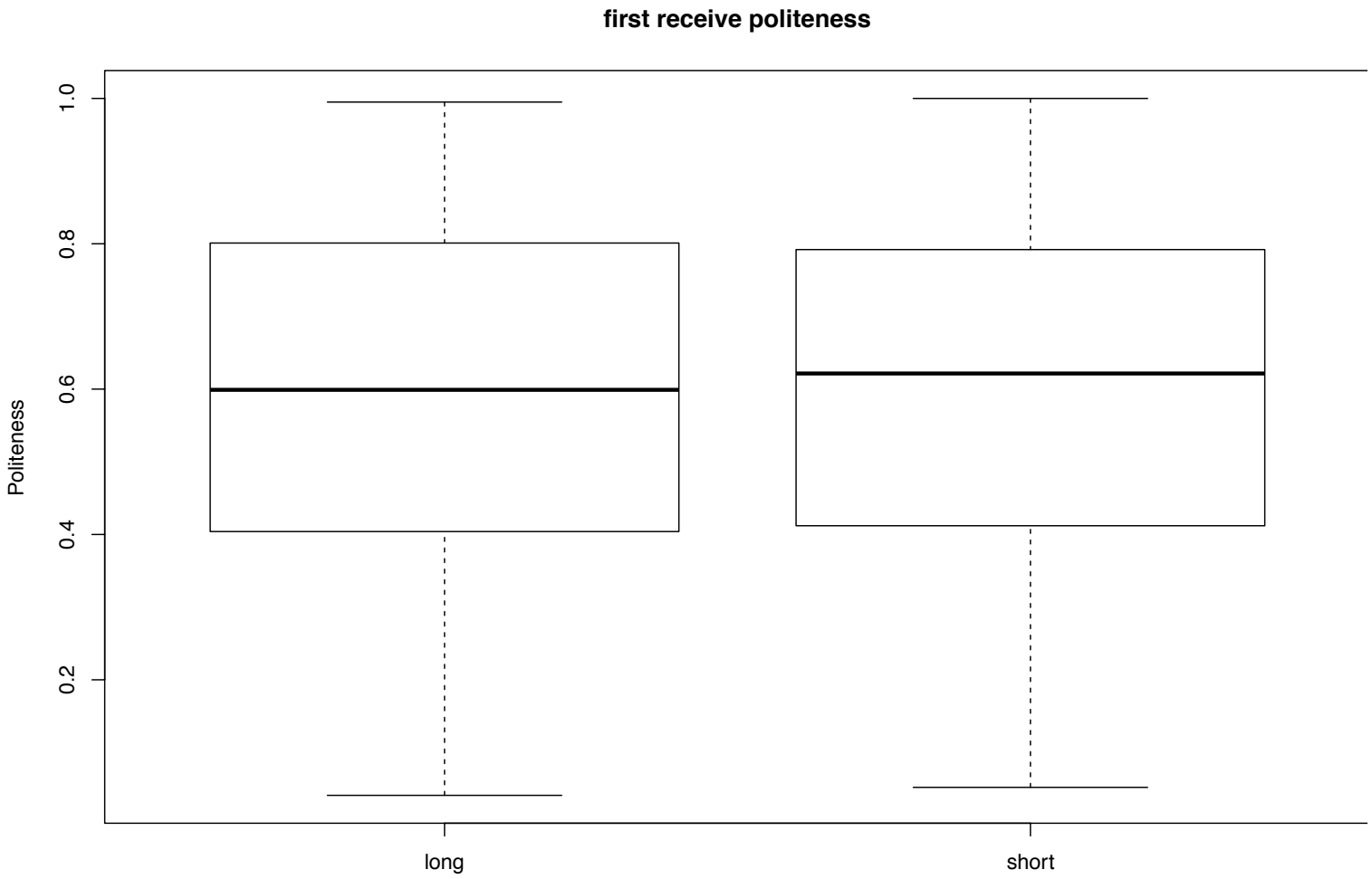

first send politeness
Politeness
0.0
0.2
0.4
0.6
0.8
1.0
long
short

RQ2-1の結果

開発者	中央値	P値
コア・一般開発者	0.599	0.9012
新人開発者	0.6215	

新人開発者が受け取ったメールの値の方が僅かに高かったが，有意差は見られなかった

RQ2-1の箱ひげ図


first receive politeness
Politeness
1.0
0.8
0.6
0.4
0.2
long
short

RQ2-1の結果

コア開発者	開発者の分類	中央値	P値	コア開発者	開発者の分類	中央値	P値
Dev1	一般・コア	0.4435	0.05634	Dev11	一般・コア	0.601	0.7679
	新人	0.473			新人	0.5945	
Dev2	一般・コア	0.632	0.0004698	Dev12	一般・コア	0.653	0.3561
	新人	0.724			新人	0.645	
Dev3	一般・コア	0.6265	0.6084	Dev13	一般・コア	0.599	0.1291
	新人	0.663			新人	0.571	
Dev4	一般・コア	0.562	0.7991	Dev14	一般・コア	0.47	0.1199
	新人	0.558			新人	0.558	
Dev5	一般・コア	0.562	0.6939	Dev15	一般・コア	0.602	0.3755
	新人	0.555			新人	0.646	
Dev6	一般・コア	0.551	0.7921	Dev16	一般・コア	0.83	0.175
	新人	0.544			新人	0.862	
Dev7	一般・コア	0.4895	0.3866	Dev17	一般・コア	0.5975	0.05489
	新人	0.511			新人	0.7525	
Dev8	一般・コア	0.557	0.8633	Dev18	一般・コア	0.6975	0.1492
	新人	0.589			新人	0.596	
Dev9	一般・コア	0.536	0.0315	Dev19	一般・コア	0.615	0.6441
	新人	0.667			新人	0.616	
Dev10	一般・コア	0.57	0.07742	Dev20	一般・コア	0.485	0.09419
	新人	0.6545			新人	0.408	

RQ2-1の結果

コア開発者	開発者の分類	中央値	P値	コア開発者	開発者の分類	中央値	P値
Dev1	一般・コア	0.4435	0.05634	Dev11	一般・コア	0.601	0.7679
	新人	0.473			新人	0.5945	
Dev2	一般・コア	0.632	0.0004698	Dev12	一般・コア	0.653	0.3561
	新人	0.724			新人	0.645	
Dev3	一般・コア	0.6265	0.6084	Dev13	一般・コア	0.599	0.1291
	[illegible]	[illegible]			[illegible]	[illegible]	
Dev7	一般・コア	0.4895	0.3866	Dev17	一般・コア	0.5975	0.05489
	新人	0.511			新人	0.7525	
Dev8	一般・コア	0.557	0.8633	Dev18	一般・コア	0.6975	0.1492
	新人	0.589			新人	0.596	
Dev9	一般・コア	0.536	0.0315	Dev19	一般・コア	0.615	0.6441
	新人	0.667			新人	0.616	
Dev10	一般・コア	0.57	0.07742	Dev20	一般・コア	0.485	0.09419
	新人	0.6545			新人	0.408	

Dev2,9においては有意差が見られた

RQ2-2の結果

コア開発者	開発者の分類	中央値	P値
Dev1	一般・コア	0.4435	0.05634
	新人	0.473	
Dev2	一般・コア	0.632	0.0004698
	新人	0.724	
Dev3	一般・コア	0.6265	0.6084
	新人	0.663	
Dev4	一般・コア	0.562	0.7991
	新人	0.558	
Dev5	一般・コア	0.562	0.6939
	新人	0.555	

RQ2-2の結果

コア開発者	開発者の分類	中央値	P値
Dev6	一般・コア	0.551	0.7921
	新人	0.544	
Dev7	一般・コア	0.4895	0.3866
	新人	0.511	
Dev8	一般・コア	0.557	0.8633
	新人	0.589	
Dev9	一般・コア	0.536	0.0315
	新人	0.667	
Dev10	一般・コア	0.57	0.07742
	新人	0.6545	

RQ2-2の結果

コア開発者	開発者の分類	中央値	P値
Dev11	一般・コア	0.601	0.7679
	新人	0.5945	
Dev12	一般・コア	0.653	0.3561
	新人	0.645	
Dev13	一般・コア	0.599	0.1291
	新人	0.571	
Dev14	一般・コア	0.47	0.1199
	新人	0.558	
Dev15	一般・コア	0.602	0.3755
	新人	0.646	

RQ2-2の結果

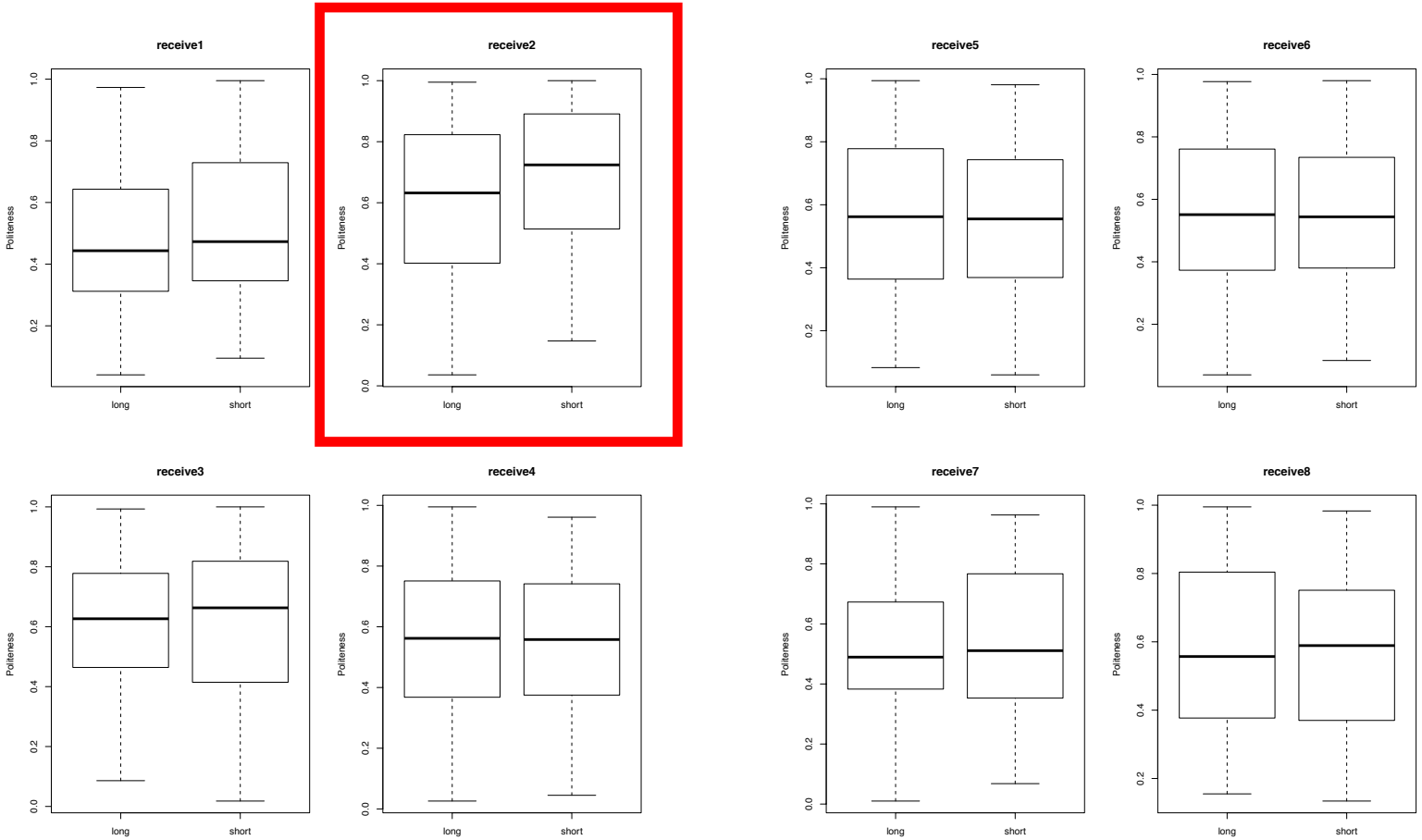
コア開発者	開発者の分類	中央値	P値
Dev16	一般・コア	0.83	0.175
	新人	0.862	
Dev17	一般・コア	0.5975	0.05489
	新人	0.7525	
Dev18	一般・コア	0.6975	0.1492
	新人	0.596	
Dev19	一般・コア	0.615	0.6441
	新人	0.616	
Dev20	一般・コア	0.485	0.09419
	新人	0.408	

RQ2-2の結果

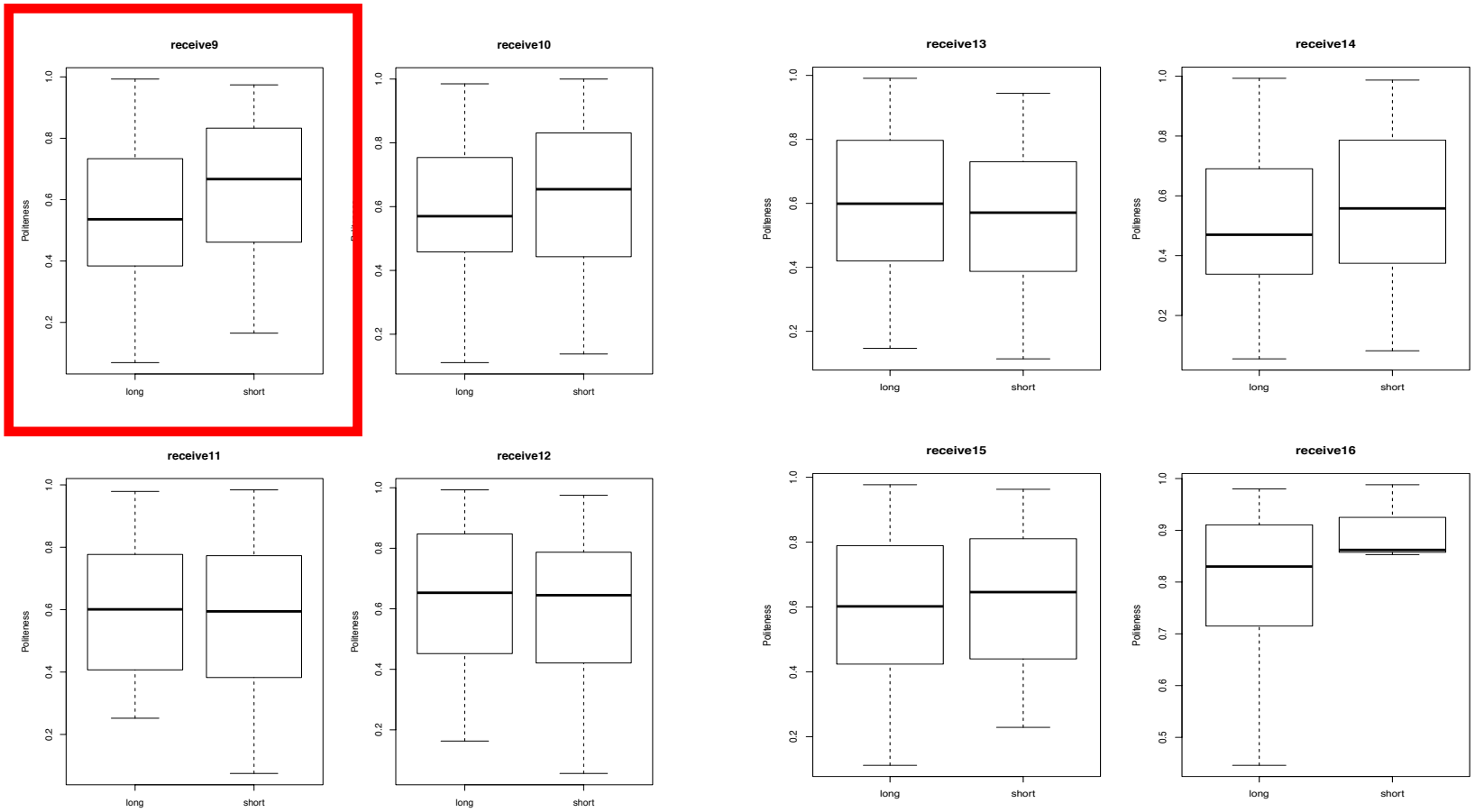
コア開発者	開発者の分類	中央値	P値
Dev16	一般・コア	0.83	0.175
	新人	0.596	
Dev19	一般・コア	0.615	0.6441
	新人	0.616	
Dev20	一般・コア	0.485	0.09419
	新人	0.408	

Dev2,9においては有意差が見られた

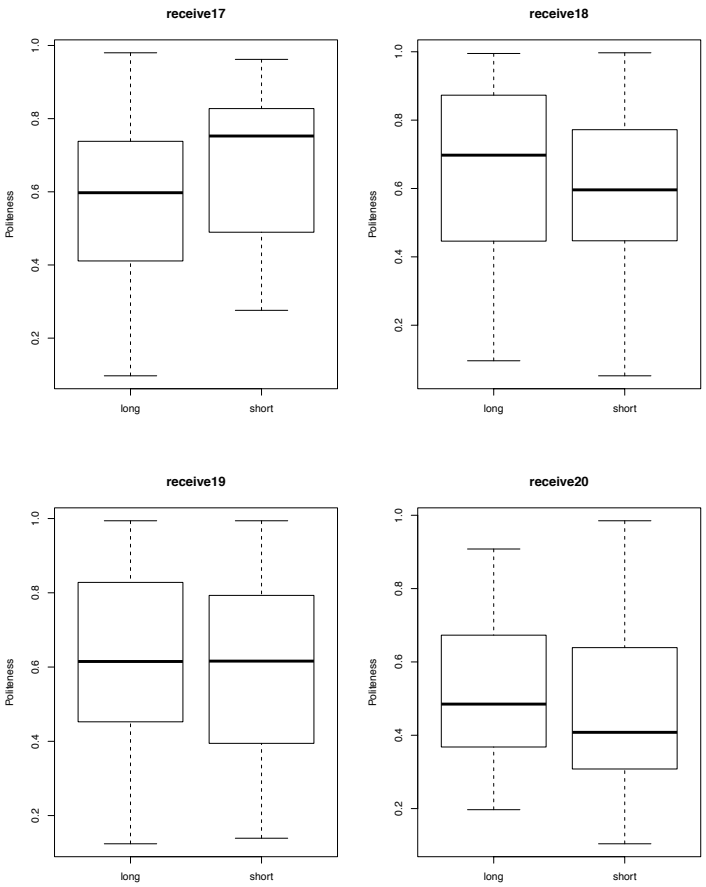
RQ2-2の箱ひげ図（1～8）



RQ2-2の箱ひげ図（9～16）



RQ2-2の箱ひげ図（１７～２０）



結果からの考察

- RQ1
  - 新人開発者が最初にメールを送る際には，砕けた表現のメールを送ればよいと考えられる
- RQ2
  - 有意差は見られなかったので，コア開発者による無礼なメールが新人開発者の継続的な貢献を妨げていないと考えられる

考察

- 新人開発者の中には単純に質問だけをしている者も含まれている可能性がある
  - 10日未満の者は除いているが・・・
  - 実際の活動量を考慮する

- RQ2では最初にコア開発者から受け取ったメールのPoliteness値を調べているが，受け取った時期を考慮していない
  - もしかしたら一年以上経過しているかも
  - 参加からの期限を定める