

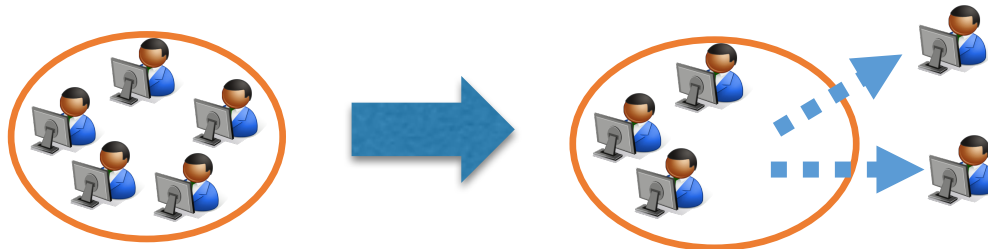
メンターとPoliteness分析

60181052

宮崎 智己

背景

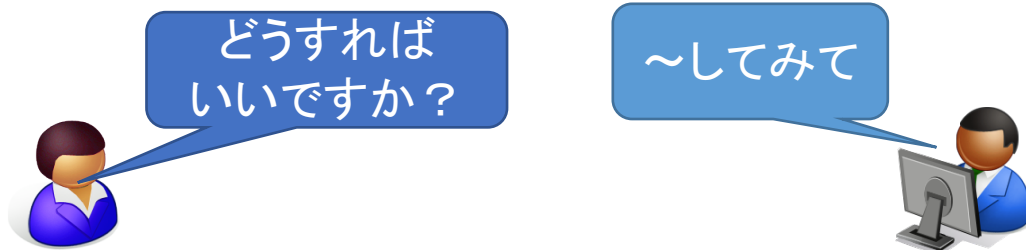
- OSSは、主にボランティアの開発者によって開発されている
- 多くの開発者が開発に参加できる一方で、開発コミュニティから離脱してしまう開発者も多く存在する



- 離脱する開発者の多くは短期間で活動をやめてしまう開発者のため、新たな人員を確保することは難しい

新人開発者の支援

- 新人開発者がコミュニティに参加を妨げる参入障壁を緩和するポータルの開発[1]
 - 主な参入障壁の例
 - 技術的障壁
 - コミュニケーション障壁
- メンターによるサポート
 - 新人開発者とコミュニケーションを行いサポート

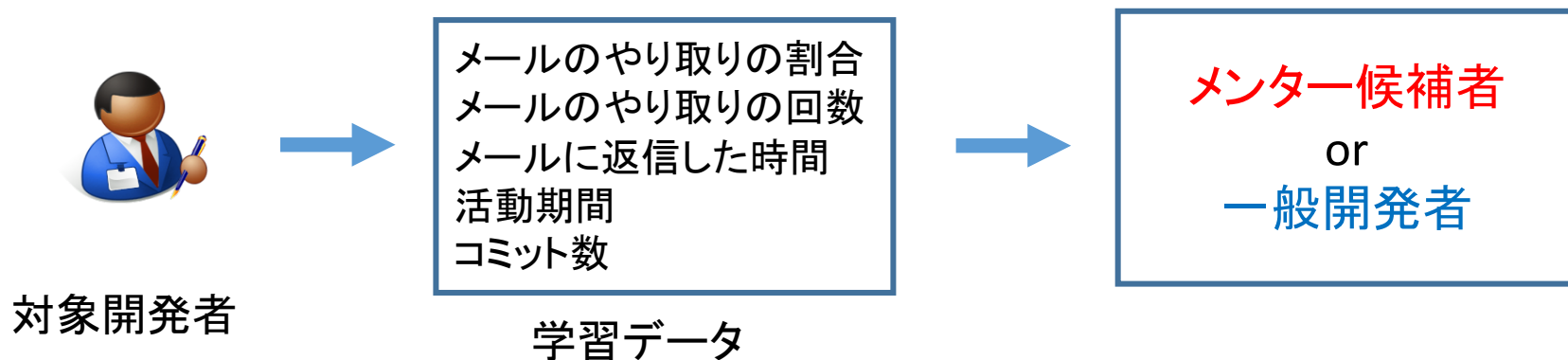


メンター

- 新人開発者, 経験の浅い開発者を育成
- 長期活動しているスキルのある開発者
- コア開発者であればいいというわけではない
 - コア開発者の粗末なコミュニケーションは新人開発者のやる気を削ぐため[2]

先行研究

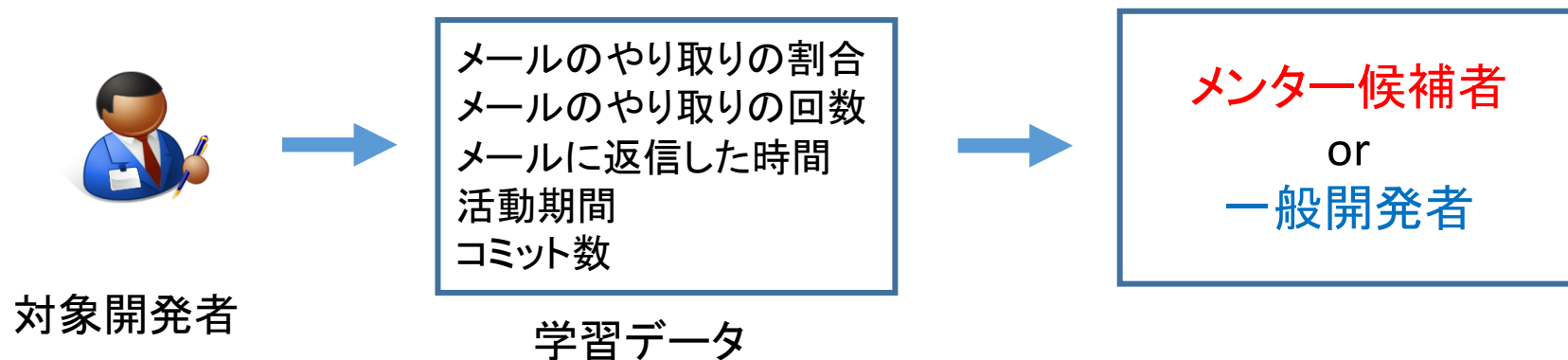
- メンター候補者の推薦[3]



[3] Gerardo Canfora, Massimiliano Di Penta, Rocco Oliveto, Sebastiano Panichella. Who is going to mentor newcomers in open source projects? Proceeding [FSE '12](#) Proceedings of the ACM SIGSOFT 20th International Symposium on the Foundations of Software Engineering

先行研究

- メンター候補者の推薦[3]



コミュニケーションの様式に着目しておらず、
新人開発者のやる気を下げるコミュニケーションをしていないとは言えない

コミュニケーション分析

- Sentiment Analysis
 - Sentiment(感情)を読み取る分析
 - 肯定, 中立, 否定の3段階に分類
 - ある議題に対しての意見の分析に適している
- Politeness Analysis
 - 相手との関係を円滑にするためのコミュニケーションをしているかどうかを分析
 - 質問や要求, 返答など様々な場面でのコミュニケーションの分析に利用可能

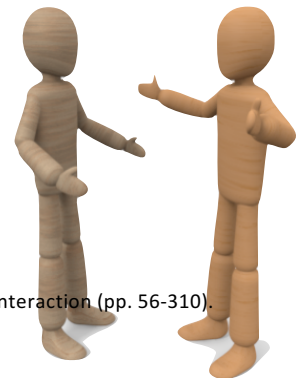
コミュニケーション分析

ある事象に対して肯定・否定を判断するSentiment Analysisよりも, Politeness Analysisの方が適している

- Politeness Analysis
 - 相手との関係を円滑にするためのコミュニケーションをしているかどうかを分析
 - 質問や要求, 返答など様々な場面でのコミュニケーションの分析に利用可能

Politeness [4]

- 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
- 単純な「丁寧さ」ではなく、相手との関係性により表現方法が異なる
 - 友人のような親密な関係:「よお！」
 - 上司と部下のような一定の距離を保った関係:「こんにちは」



Politeness Strategy

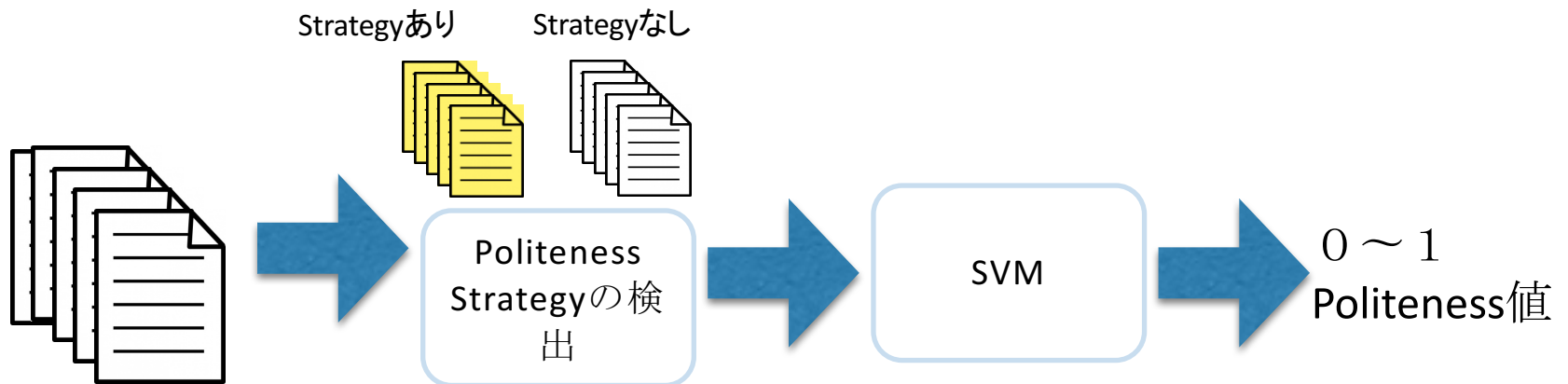
- Politenessを満たす20の要素
 - 単語の意味や位置によって分類

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 ~.	

主な例

Stanford Politeness

- Politenessを定量化するツール
- 処理の手順
 1. stackoverflowとwikipediaでの会話を人手により評価し, SVMにより学習させモデル構築
 2. テキストの構文解析
 3. Politeness Strategyの検出
 4. 学習モデルに基づいてPoliteness値を算出



本分析の目的とアプローチ

- 目的

- 新人開発者をサポートするメンターの適正にコミュニケーションのPolitenessが寄与するかを明らかにすること

- アプローチ

- Politenessを用いたメンター候補者推薦モデルの構築

今後やること

- 既存手法とPolitenessを用いた手法を比較
 1. 先行研究と同じデータセットの取得
 2. Politenessをメトリクスに用いたモデルを構築
 - Politeness値
 - Politeness Strategy
 - 両方
 3. 既存手法と比較
- 卒論の目次の修正

予備調査 ストラテジーの割合

全メールのストラテジーの割合

	Words	頻度	割合
1	1st_person	58137	0.213220031
2	1st_person_start	37729	0.138372784
3	1st_person_pl	35626	0.130659938
4	2nd_person	33588	0.123185482
5	hedges	33386	0.122444638
6	direct_start	20770	0.076174898
7	factuality	14345	0.052610925
8	direct_question	11929	0.043750138
9	gratitude	6894	0.025284051
10	2nd_person_start	5222	0.019151917
11	please	4058	0.014882895
12	indirect_	1996	0.007320419
13	greeting	1990	0.007298413
14	apologizing	1883	0.006905986
15	please_start	1797	0.006590577
16	indicative	1364	0.005002531
17	subjunctive	1036	0.003799576
18	deference	906	0.003322795
19	btw	6	

コア開発者の全メール(送信・受信含む)のストラテジーの割合

	Words	頻度	割合
1	1st_person	33759	0.212669854
2	1st_person_pl_	24233	0.152659397
3	1st_person_start	21989	0.138522984
4	hedges	19654	0.123813304
5	2nd_person	19164	0.120726476
6	direct_start	12383	0.078008555
7	factuality	8998	0.056684243
8	direct_question	7030	0.044286533
9	2nd_person_start	3117	0.019636006
10	gratitude	2480	0.01562313
11	please	2015	0.012693793
12	apologizing	847	0.005335803
13	please_start	834	0.005253907
14	indicative	602	0.003792389
15	subjunctive	571	0.0035971
16	deference	555	0.003496305
17	indirect_	254	0.001600111
18	greeting	252	0.001587512
19	btw	2	

コア開発者の送信メールの ストラテジーの割合

	Words	頻度	割合
1	1st_person	19350	0.205276726
2	1st_person_pl_	15151	0.160731146
3	1st_person_start	12552	0.133159352
4	hedges	11487	0.121861176
5	2nd_person	11475	0.121733872
6	direct_start	7291	0.077347422
7	factuality	5681	0.060267549
8	direct_question	4188	0.044428885
9	2nd_person_start	1912	0.020283674
10	please	1508	0.015997793
11	gratitude	1435	0.015223364
12	please_start	570	0.006046911
13	apologizing	438	0.004646574
14	deference	379	0.004020666
15	indicative	360	0.003819102
16	subjunctive	348	0.003691798
17	indirect_	69	0.000731995
18	greeting	68	0.000721386
19	btw	1	

コア開発者の受信メールの ストラテジーの割合

	Words	頻度	割合
1	1st_person	14409	0.223478504
2	1st_person_start	9437	0.146364539
3	1st_person_pl	9082	0.140858614
4	hedges	8167	0.126667287
5	2nd_person	7689	0.119253676
6	direct_start	5092	0.078975123
7	factuality	3317	0.051445499
8	direct_question	2842	0.044078417
9	2nd_person_start	1205	0.018689125
10	gratitude	1045	0.016207581
11	please	507	0.007863391
12	apologizing	409	0.006343446
13	please_start	264	0.004094547
14	indicative	242	0.003753335
15	subjunctive	223	0.003458651
16	indirect_	185	0.002869285
17	greeting	184	0.002853775
18	deference	176	0.002729698
19	btw	1	

離脱した開発者の離脱直前の 送信メールと受信メールのストラテジー

送信メール(その1)

メール	Dev3	Dev4
1		Hedges/ Factuality/ 1st_person/ 1st_person_start/ 2nd_person/ Direct_start
2	Hedges/ 1st_person_pl./ 1st_person/ 1st_person_start/ 2nd_person/ 2nd_person_start	1st_person_start
3	1st_person/ 1st_person_start	Hedges/ 1st_person_pl./ 1st_person/ 1st_person_start/ 2nd_person/ Direct_start
4	1st_person/ 1st_person_start/ 2nd_person/ 2nd_person_start	Hedges/ 1st_person/ Indirect_(greeting)
5	1st_person/ 2nd_person/ 2nd_person_start/ Direct_question	
6	1st_person/ 1st_person_start/ 2nd_person	1st_person_pl./ 1st_person/ 2nd_person
7	1st_person_pl./ 1st_person/ 1st_person_start/ 2nd_person/ 2nd_person_start/ Direct_start	Hedges/ 1st_person/ 2nd_person
8	1st_person_pl./ 1st_person/ 1st_person_start	Hedges/ Factuality/ 1st_person_pl./ 1st_person/ 2nd_person/ Direct_question
9	Please_start/ 1st_person_pl./ 1st_person/ 1st_person_start	Hedges/ Factuality/ 1st_person/ 2nd_person/ Direct_start
10	1st_person_pl./ 1st_person/ 1st_person_start/ 2nd_person/ 2nd_person_start/ Direct_start	Please/ 2nd_person/ Direct_start
11	Factuality	Hedges/ 1st_person/ 2nd_person_start
12		Hedges/ 1st_person_start
13	Please/ Hedges/ Factuality/ 1st_person_pl./ 1st_person/ 2nd_person/ 2nd_person_start	2nd_person
14	Please_start/ Hedges/ 1st_person_pl./ 2nd_person	
15		Hedges/ Factuality/ 1st_person_pl./ 1st_person/ 1st_person_start/ Direct_start

送信メール(その2)

メール	Dev8	Dev9
1	1st_person	1st_person/ 1st_person_start
2	1st_person/ 1st_person_start/ Direct_question/ Direct_start	Hedges/ 1st_person/ 1st_person_start
3	Hedges/ 1st_person/ 1st_person_start/ Direct_start	
4	Hedges/ Factuality/ 1st_person_pl./ 1st_person/ 2nd_person/ Direct_question/ Direct_start	Direct_question
5	Direct_start	
6	1st_person_pl./ 2nd_person/ Direct_start	
7	1st_person_pl.	2nd_person/ Direct_question
8	1st_person_pl.	
9	Hedges/ 2nd_person	
10	1st_person_pl./ 1st_person_start/ 2nd_person/ 2nd_person_start/ Direct_start	
11	1st_person/ 1st_person_start	
12	2nd_person	Hedges/ 1st_person/ 1st_person_start
13		
14	Hedges/ 1st_person_pl./ 1st_person/ 1st_person_start/ Direct_question/ Direct_start	2nd_person
15	Hedges/ Factuality/ 2nd_person/ 2nd_person_start	1st_person

受信メール(その1)

メール	Dev3	Dev4
1	Hedges/ 1st_person_pl./ 1st_person/ 1st_person_start/ 2nd_person	1st_person_start
2	Direct_start	
3	1st_person_pl./ 2nd_person/ 2nd_person_start	Hedges/ Deference/ 1st_person
4	1st_person_pl./ 1st_person/ 1st_person_start/ Direct_start	Hedges/ 1st_person_start/ 2nd_person
5	1st_person_pl./ 2nd_person/ Direct_start	Hedges/ 1st_person
6	Hedges/ 1st_person_pl./ 1st_person/ 1st_person_start/ Direct_start	1st_person
7	2nd_person/ Direct_start	1st_person_pl./ 1st_person/ 1st_person_start
8	1st_person/ 1st_person_start	Factuality/ 1st_person/ 1st_person_start/ 2nd_person/ Direct_question
9		
10	Hedges/ 1st_person_pl./ Direct_start	1st_person
11	1st_person_pl./ 1st_person	Hedges/ 1st_person_pl./ 1st_person/ 1st_person_start
12		Hedges/ 1st_person_pl./ 1st_person_start
13		1st_person_pl./ 1st_person_start
14	Hedges/ 1st_person_pl./ 1st_person_start/ Direct_start	
15	Hedges/ 1st_person/ 2nd_person	Hedges/ 1st_person_pl./ 1st_person_start

受信メール(その2)

メール	Dev8	Dev9
1	Hedges/ 1st_person_pl./ 1st_person/ 1st_person_start/ 2nd_person_start/ Direct_question	Hedges/ 1st_person/ 1st_person_start
2	Direct_question	1st_person_start
3	1st_person_pl./ 1st_person/ 1st_person_start	
4	Hedges/ 1st_person_pl./ 1st_person/ 1st_person_start/ Direct_question	Hedges/ 1st_person_start
5	1st_person	
6		
7	Please/ Hedges/ Factuality/ 1st_person_pl./ 1st_person/ 2nd_person/ Direct_question	
8	1st_person_pl.	
9	2nd_person/ Direct_question	1st_person_pl.
10		Hedges/ 1st_person/ 1st_person_start
11	1st_person_pl./ 1st_person	1st_person_pl.
12	Hedges/ 1st_person_pl./ 1st_person_start/ Direct_start	1st_person
13	2nd_person	2nd_person
14	Gratitude	1st_person/ Direct_question
15		1st_person_start/ 2nd_person/ SUBJUNCTIVE

離脱開発者の送信メールラスト15件の ストラテジーの割合

	Words	頻度	割合
1	1st_person	29	0.188311688
2	2nd_person	24	0.155844156
3	1st_person_start	21	0.136363636
4	hedges	19	0.123376623
5	1st_person_pl	17	0.11038961
6	direct_start	14	0.090909091
7	2nd_person_start	9	0.058441558
8	factuality	8	0.051948052
9	direct_question	7	0.045454545
10	please	2	0.012987013
10	please_start	2	0.012987013
12	greeting	1	0.006493506
12	indirect_	1	0.006493506

離脱開発者の送信メールラスト15件の ストラテジーの割合

	Words	頻度	割合
1	1st_person	23	0.191666667
2	1st_person_pl	22	0.183333333
2	1st_person_start	22	0.183333333
4	hedges	18	0.15
5	2nd_person	12	0.1
6	direct_start	8	0.066666667
7	direct_question	7	0.058333333
8	2nd_person_start	2	0.016666667
8	factuality	2	0.016666667
10	deference	1	0.008333333
10	gratitude	1	0.008333333
10	please	1	0.008333333
10	subjunctive	1	0.008333333

リサーチクエスト

- 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

リサーチクエスト

- 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%

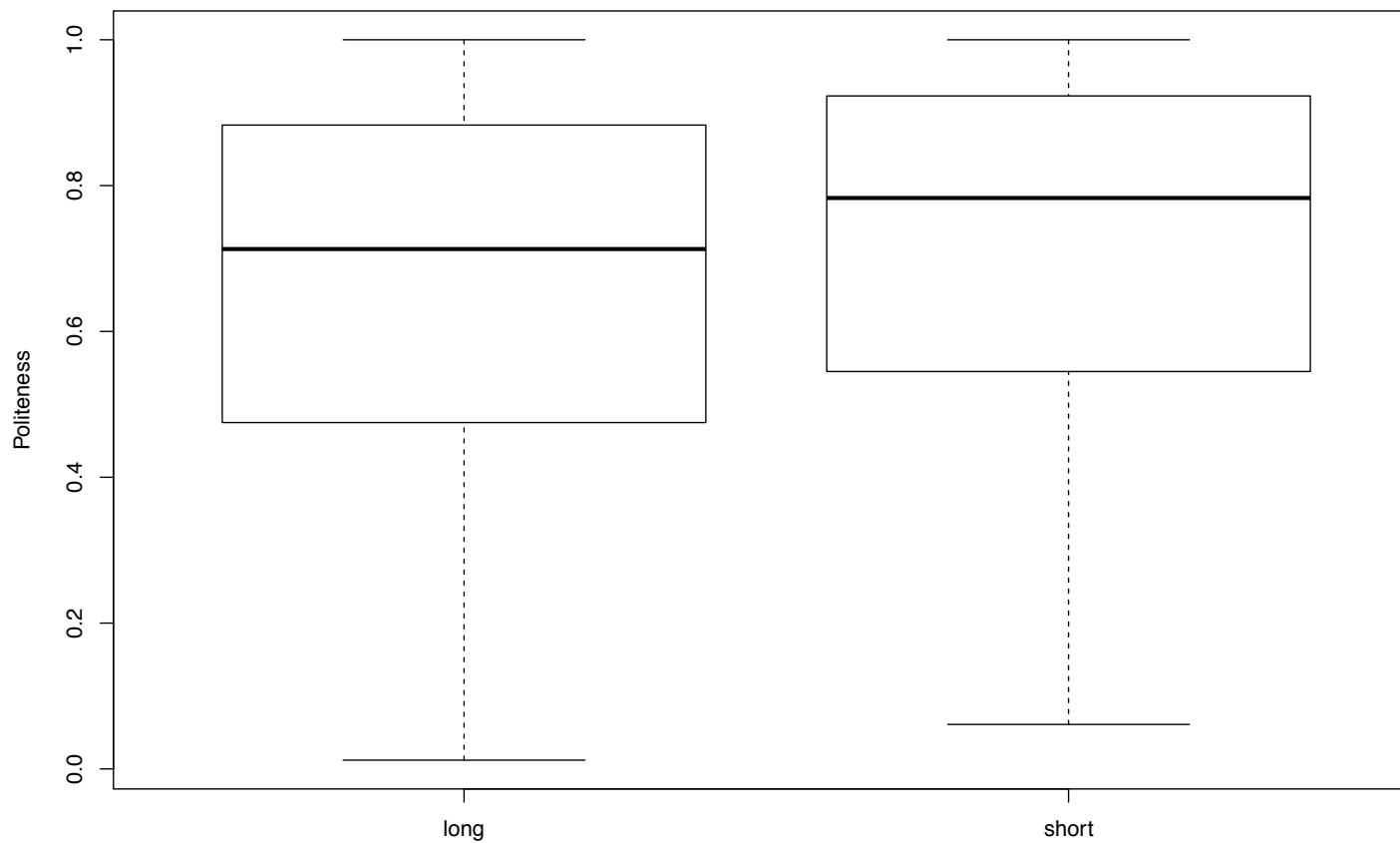
RQ1の結果

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

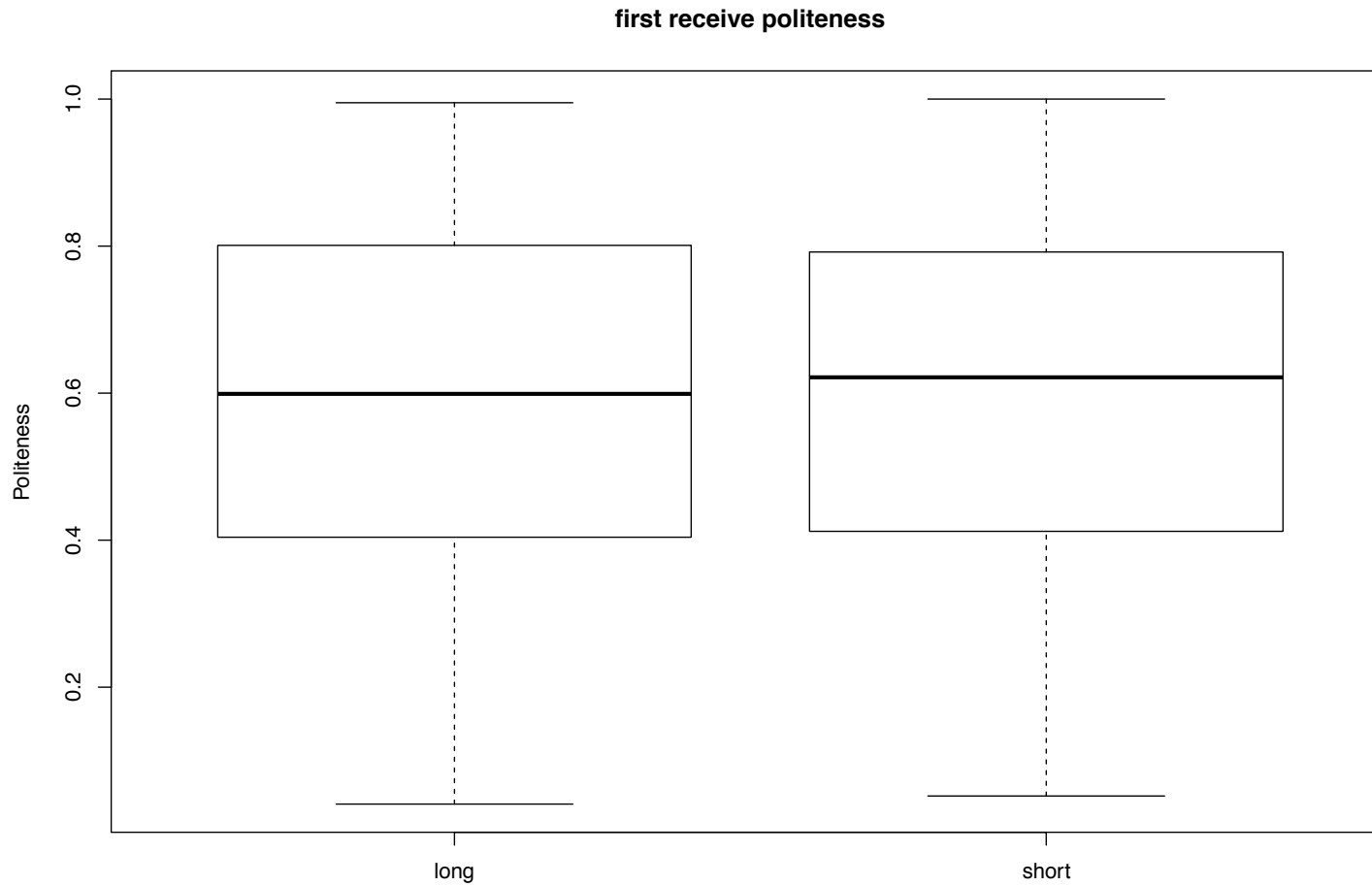


RQ2-1の結果

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



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	

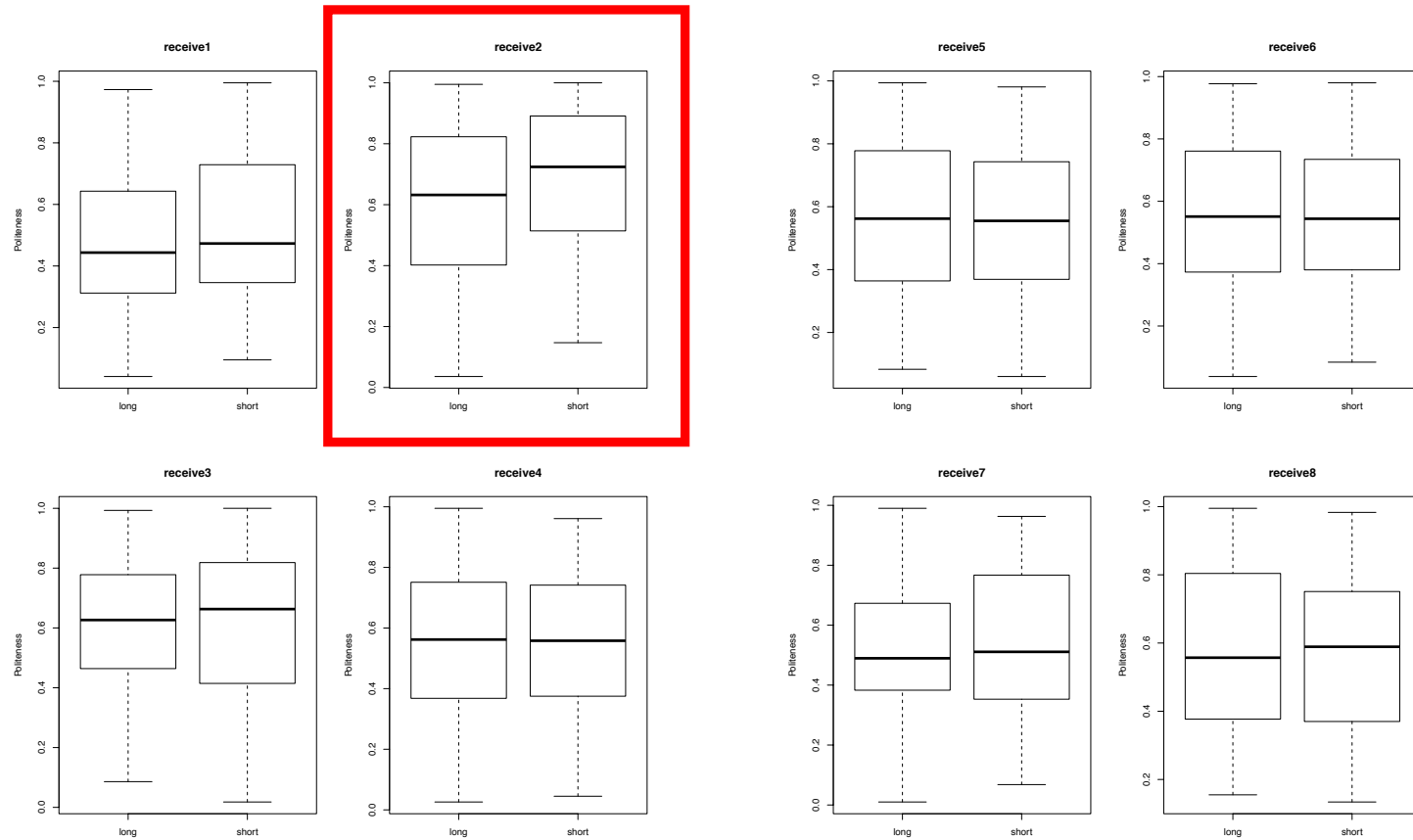
RQ2-1の結果

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
	新人	0.668			新人	0.574	

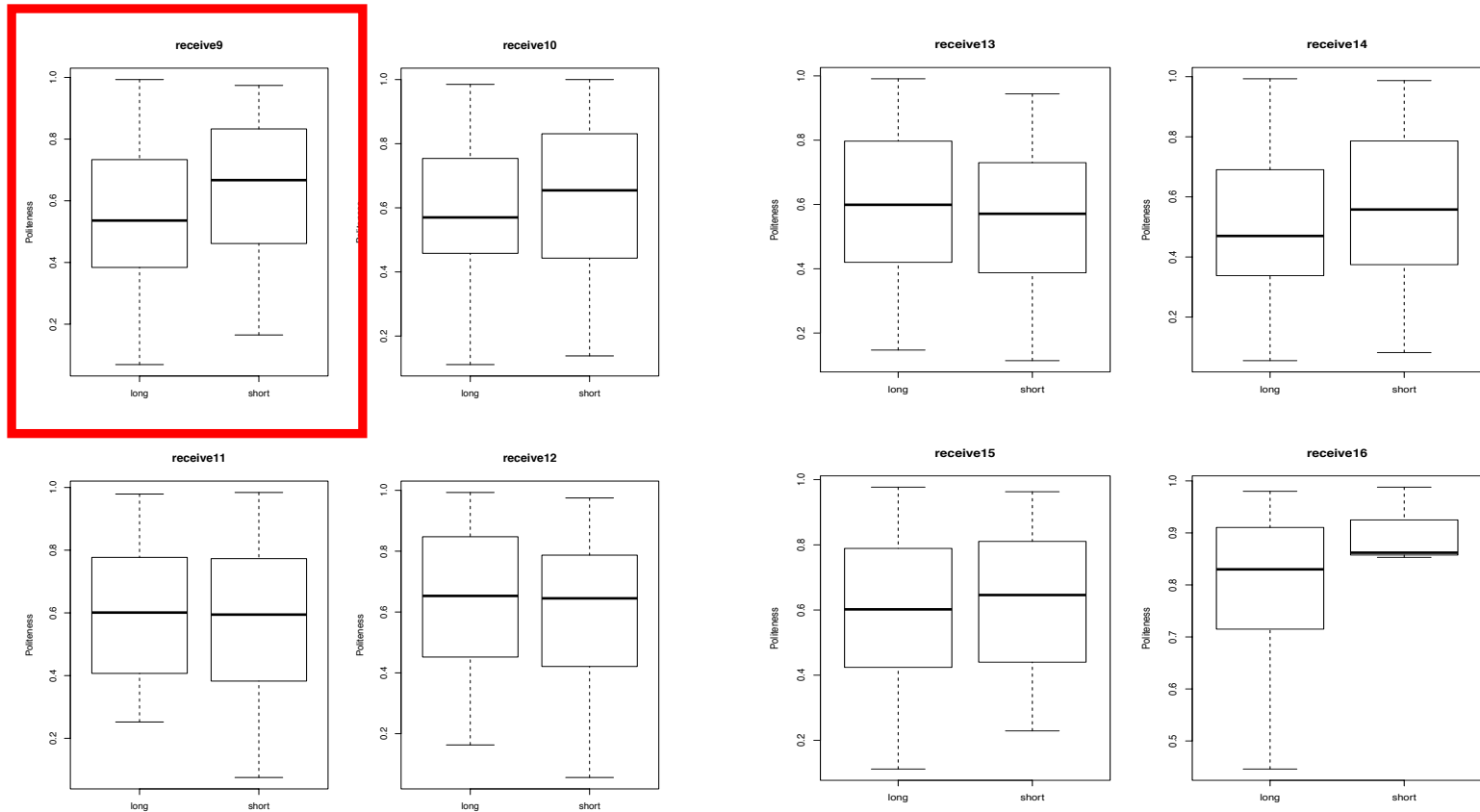
Dev2,9においては有意差が見られた

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	

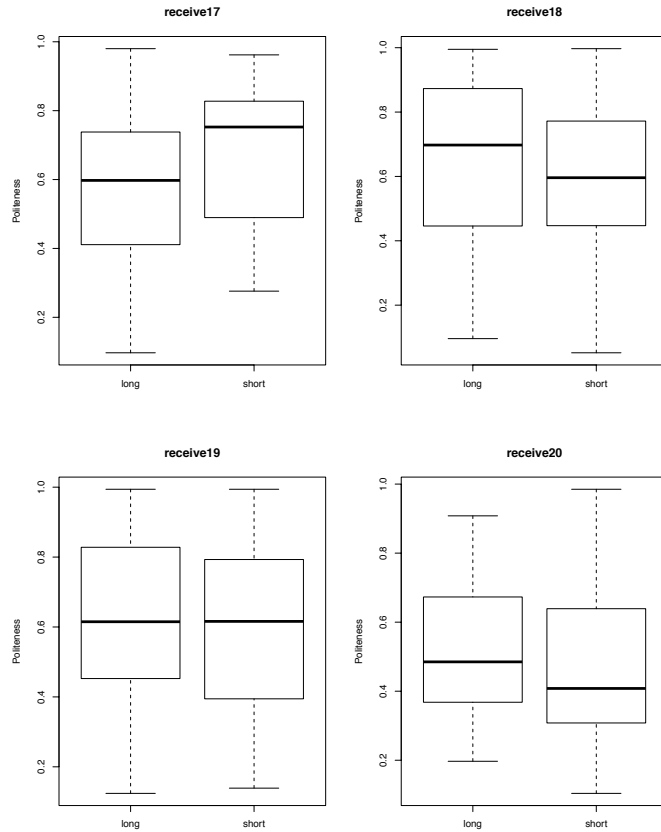
RQ2-2の箱ひげ図(1~8)



RQ2-2の箱ひげ図(9~16)



RQ2-2の箱ひげ図(17~20)



コアから受け取ったメールの Politeness値と一般から受け取った メールのPoliteness値の比較(1)

	Politeness値	P値
一年以上活動	0.6035	0.824
一年未満活動	0.6055	

コアから受け取ったメールの Politeness値と一般から受け取った メールのPoliteness値の比較(2)

	core	noncore	P値
一年以上	0.57	0.6035	0.2342
一年未満	0.604	0.6055	0.952

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もので、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

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

Dev2,9においては有意差が見られた

	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	