

# 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

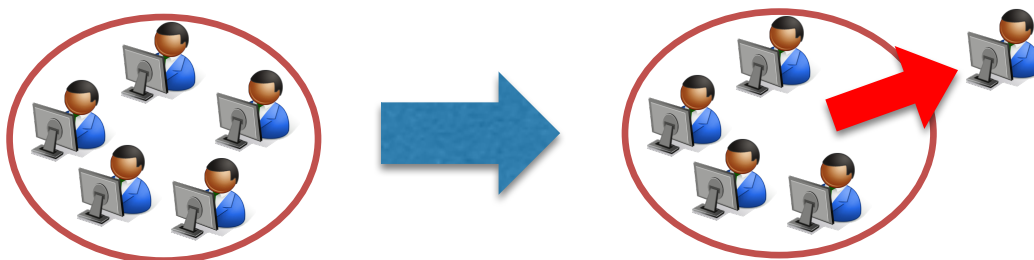
60181052 宮崎智己

# 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OSSはボランティア開発者によって開発・保守されているため
- どのOSSを再利用すべきかがわからない

# OSS開発コミュニティの継続性の研究

- 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



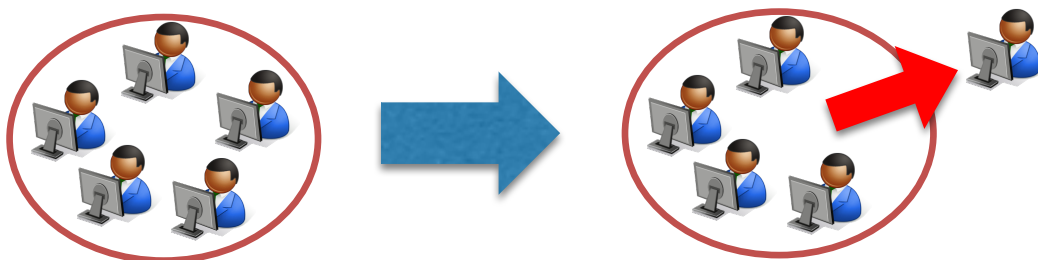
- 新人開発者に対する58個の参入障壁[2]
  - コミュニケーション障害：他の開発者からの**配慮のないメール (impoliteなメール)** がやる気を削ぐ

[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)

[2] 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

# OSS開発コミュニティの継続性の研究

- 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



開発者の配慮とOSS開発コミュニティの継続性  
に着目した研究はされていない

International Workshop on Mining Software Repositories (MSR'07), p. 6 (2007)

[2] 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

# 本研究の目的とアプローチ

- 目的

- 開発者の配慮がOSS開発コミュニティの継続性に影響をあたえるかを明らかにする

- アプローチ

- 開発者の配慮をPolitenessによって計測
- 長期間存続しているプロジェクトで活発に議論を行っている開発者のPolitenessを分析

# 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<br>高めるStrategy  |
| deference                    | great, good, nice          |                             |
| Greeting                     | hi, hey, hello             |                             |
| hedges                       | think, assume, feel        |                             |
| Please start                 | Please ~,                  | Politenessを<br>低くするStrategy |
| Factuality                   | really, actually, honestly |                             |
| Direct question              | What ~?, Why ~?. Where ~?  |                             |
| 2 <sup>nd</sup> person start | You ~.                     |                             |

# ケーススタディ

- 対象プロジェクト

| プロジェクト             | 対象期間                  | 総メール件数   |
|--------------------|-----------------------|----------|
| Apache HTTP Server | 1995/09/11～2016/03/30 | 133,272件 |
| Python             | 1999/04/02～2016/10/13 | 141,745件 |

- 対象開発者

  - メール総送信数トップ10名の開発者

- 手順

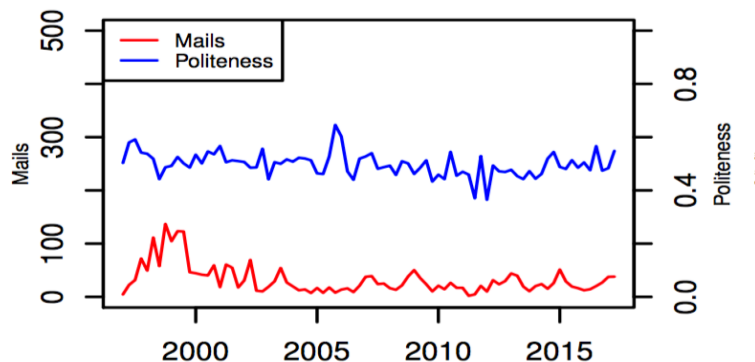


# リサーチクエスチョン

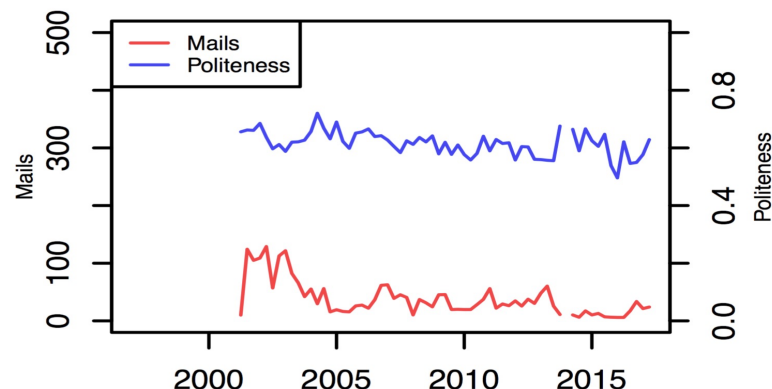
- RQ1
  - トップ開発者間でPolitenessに違いはあるのか？
- RQ2
  - トップ開発者と一般開発者のPolitenessに違いはあるのか？
- RQ3
  - 自ら発信する場合と他者へ返信する場合でPolitenessに違いはあるのか？
- RQ4
  - Politenessは時期により変動するのか？

# 分析結果

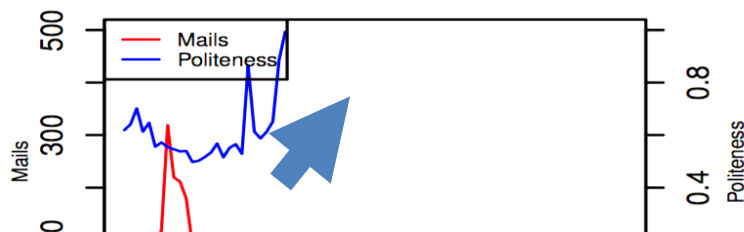
Dev. #1



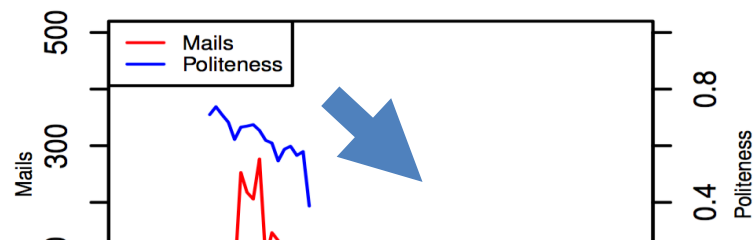
Dev. #2



Dev. #4

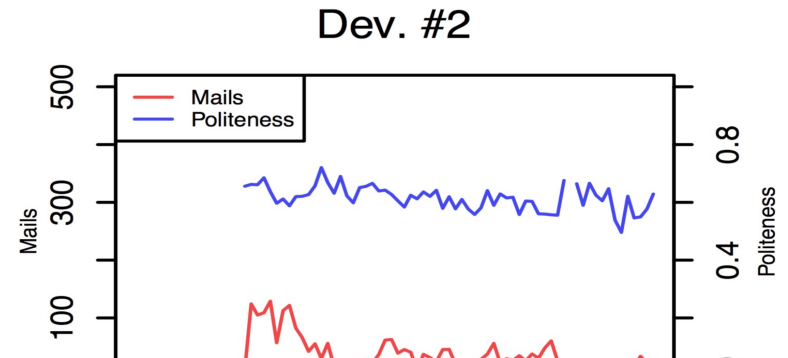
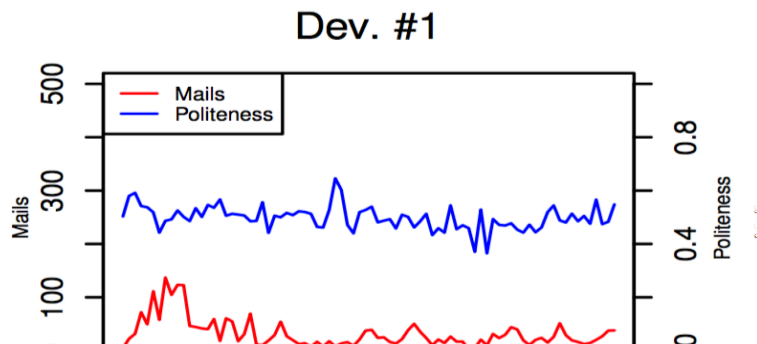


Dev. #3

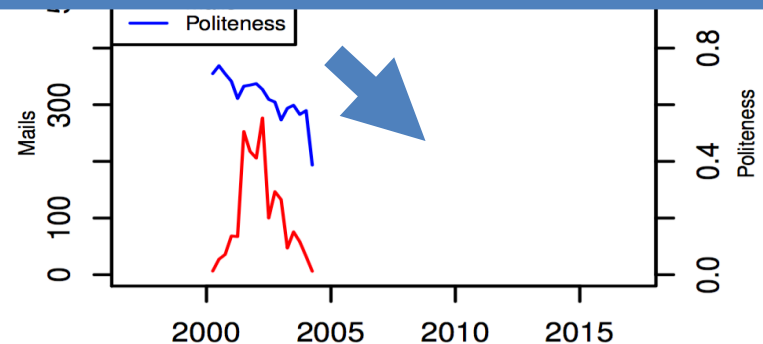
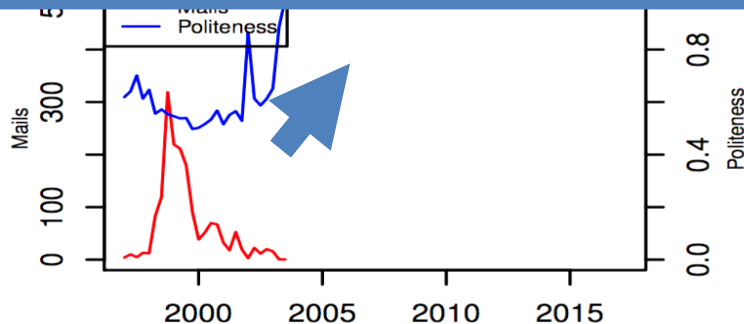


開発者毎にPoliteness値に個人差があること、  
時間による変動があることが確認できた

# Politenessは時期により変動するの か？



長期的に活動している開発者からは  
Politenessの急激な変動は見られなかった

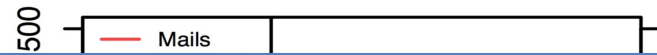


# Politenessは時期により変動するの か？

Dev. #1

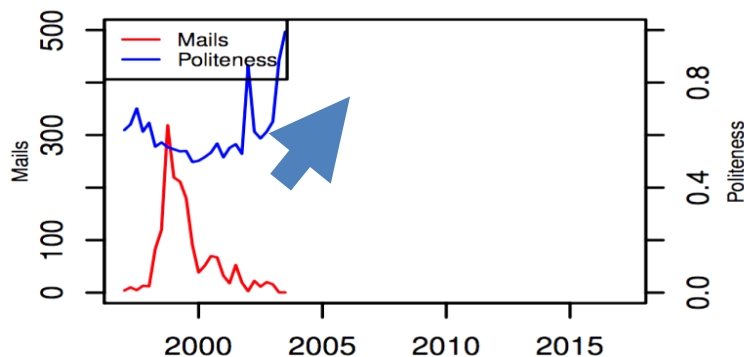


Dev. #2

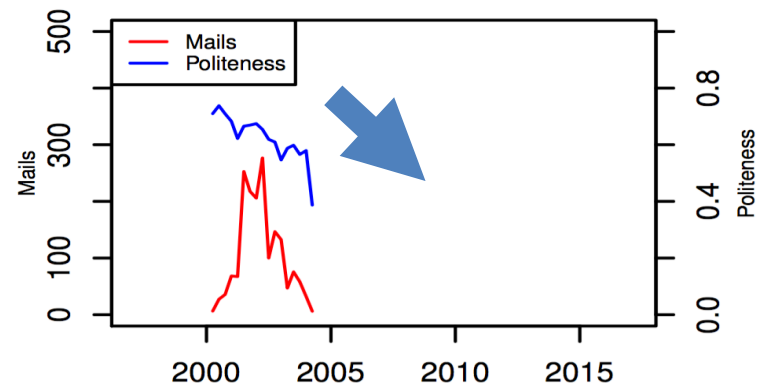


離脱した開発者は離脱前に急激な変動が見られた

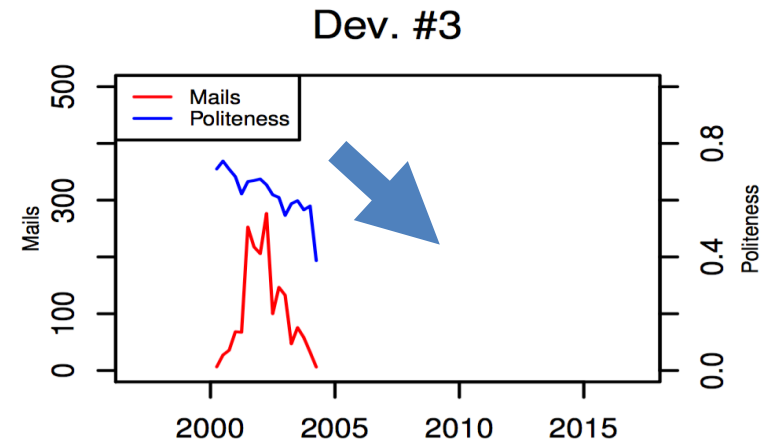
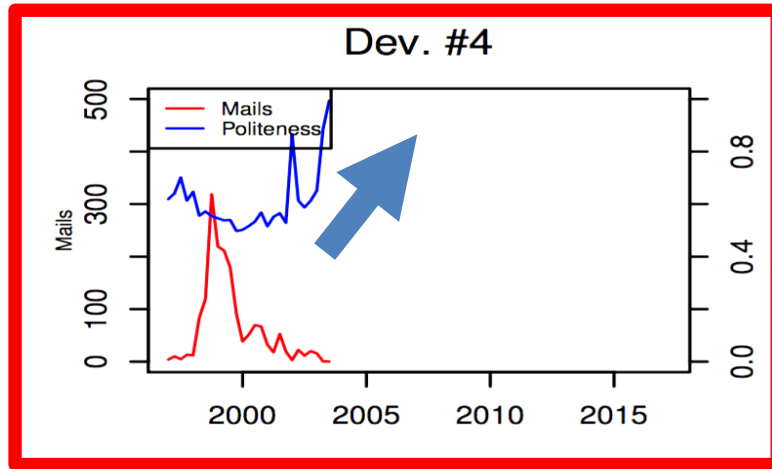
Dev. #4



Dev. #3

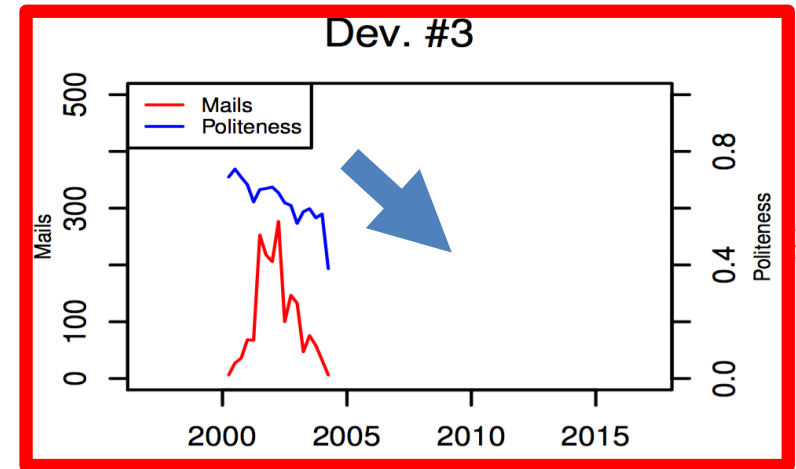
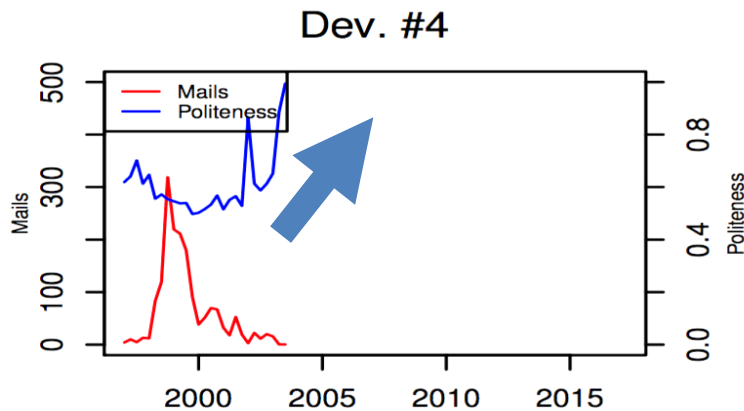


# 離脱前の急激な変動についての考察



コミュニティに対して嫌気がさしてきて  
距離を置こうとしているためではないか

# 離脱前の急激な変動についての考察



離脱前に複数の開発者と口論していた  
Politeness値の急激な減少は  
コミュニケーションの異常の目安となる

# まとめ

- 開発者のPolitenessがOSS開発コミュニティの継続性に影響を与えるかを明らかにするために、トップ開発者のPolitenessを分析した
- 開発者のPolitenessとOSS開発コミュニティの継続性の関係は見られなかった
- 開発者ごとにPoliteness値が違ふこと、また時期によりPoliteness値が違ふことがわかった

# 今後の課題

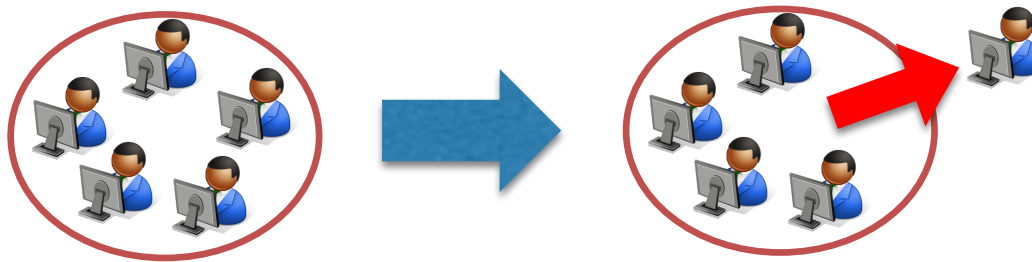
- プロジェクトの時期を考慮
  - 開発が活発に行なわれている時期とそうでない時期
- 開発者間の関係を考慮
  - コア開発者間
  - コア開発者・一般開発者間





# 先行研究

- OSS開発コミュニティの継続性に関する研究
  - 開発者の平均活動期間は約1.5年[1]



- OSS開発のコミュニケーション分析の研究
  - 修正した不具合数と開発者の感情に相関がある[2]
  - 不具合の修正時間と開発者の感情に相関がある[3]

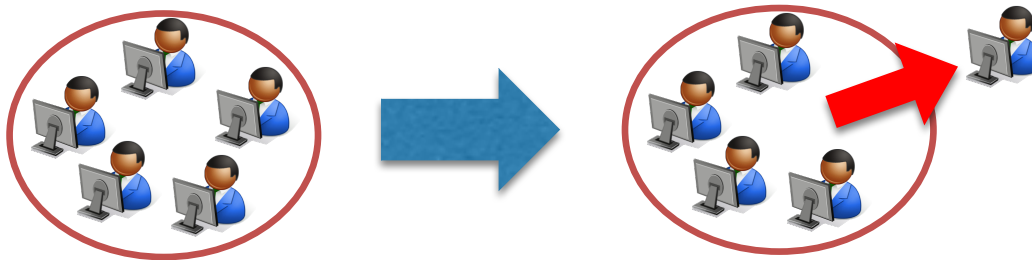
[1] Bird, C., Gourley, A., Devanbu, P., Swaminathan, A. and Hsu, G.: Open Borders? Immigration in Open Source Projects, Proceedings of the 4th International Workshop on Mining Software Repositories (MSR'07), p. 6 (2007)

[2] Parastou Tourani, Yajuan Jiang, and Bram Adams. Monitoring sentiment in open source mailing lists: Exploratory study on the apache ecosystem. In Proceedings of 24th Annual International Conference on Computer Science and Software Engineering (CASCON'14), pp. 34–44, 2014.

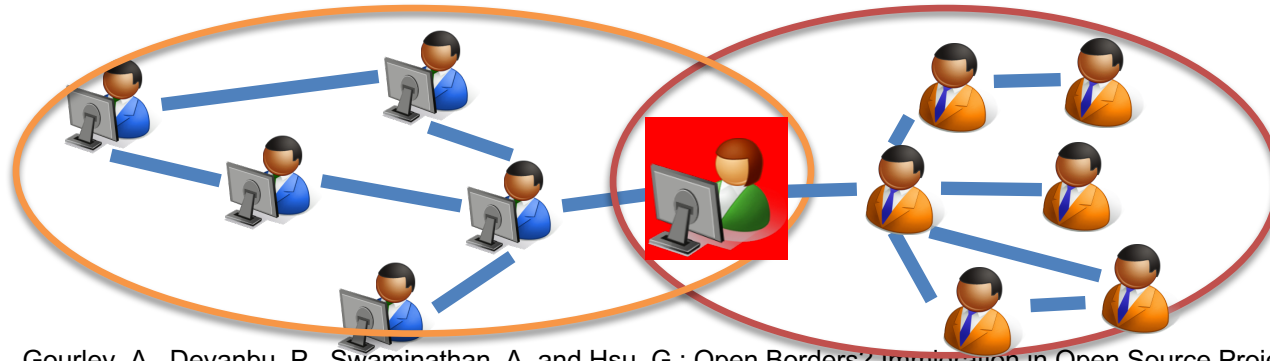
[3] Marco Ortu, Bram Adams, Giuseppe Destefanis, Parastou Tourani, Michele Marchesi, and Roberto Tonelli. Are bullies more productive?: Empirical study of affectiveness vs. issue fixing time. In Proceedings of the 12th Working Conference on Mining Software Repositories (MSR '15), pp. 303–313, 2015.

# OSSコミュニティの継続性の研究 ：コーディネータの研究

- 開発者の平均活動期間は約1.5年[1]
  - 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



- コミュニティの維持にコーディネータが貢献[2]



[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)

[2] Matsumoto, S., Kamei, Y., Ohira, M. and Matsumoto, K.: A Comparison Study on the Coordination between Developers and Users in Foss Communities, Socio- Technical Congruence (STC'08), pp. CD-ROM-NO.8- 1-9 (2008).

# OSSコミュニティの継続性の研究 ：コーディネータの研究

- 開発者の平均活動期間は約1.5年[1]
  - 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

コーディネータが送っているメールの内容は対象とされておらず、コーディネータがどのようにコミュニティを維持しているか不明



[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)

[2] Matsumoto, S., Kamei, Y., Ohira, M. and Matsumoto, K.: A Comparison Study on the Coordination between Developers and Users in Foss Communities, Socio- Technical Congruence (STC'08), pp. CD-ROM-NO.8- 1-9 (2008).

# OSSコミュニティの継続性の研究 ：新人開発者の参入障壁

- 開発者の多くはプロジェクトに参加してすぐにコミュニティから離脱している
- 新人開発者に対する58個の参入障壁[3]
  - コミュニケーション障害：コーディネータからの配慮のないメール（impoliteなメール）がやる気を削ぐ

どのようなメールが配慮がないのか、  
新人開発者の活動に影響を与えているかを理解  
する必要がある

[3] 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

# 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
- 相手との関係性により表現方法が異なる
  - 相手との関係に適した言葉遣い
    - 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」
  - 単純な「丁寧さ」ではない

[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.

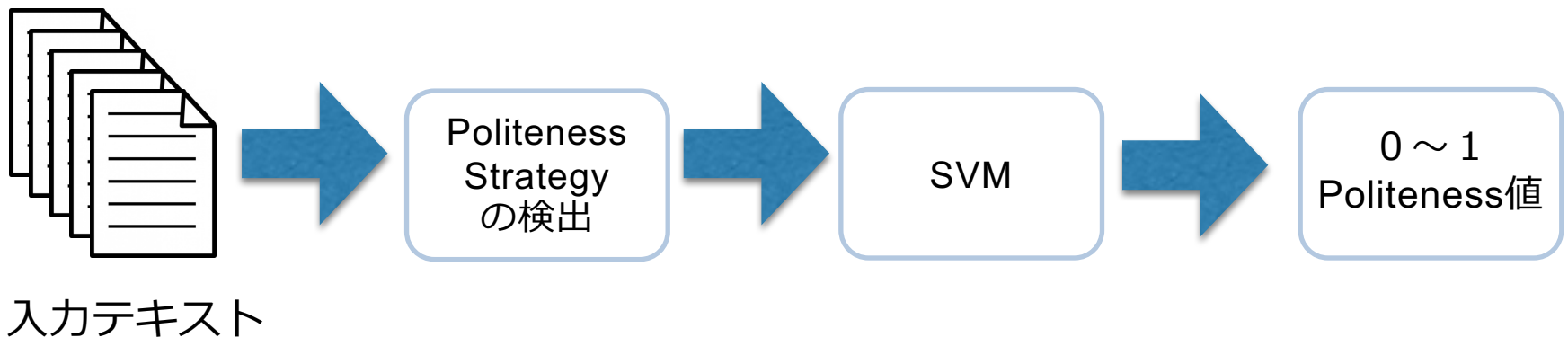
# Politeness Strategy

- 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<br>高めるStrategy  |
| deference                    | great, good, nice          |                             |
| Greeting                     | hi, hey, hello             |                             |
| hedges                       | think, assume, feel        |                             |
| Please start                 | Please ~,                  | Politenessを<br>低くするStrategy |
| Factuality                   | really, actually, honestly |                             |
| Direct question              | What ~?, Why ~?. Where ~?  |                             |
| 2 <sup>nd</sup> person start | You ~.                     |                             |

# Stanford Politeness[4]

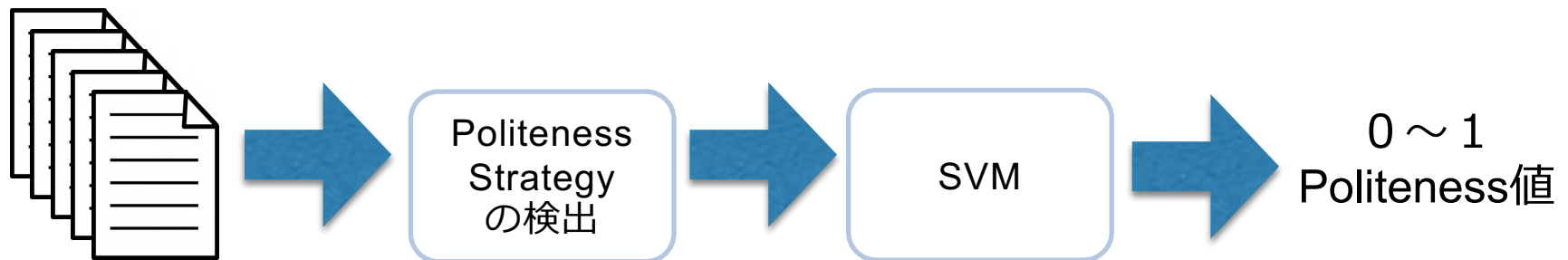
- Politenessを定量化するツール
  - 入力テキストからPoliteness Strategyを正規表現を用いて抽出
  - 抽出したPoliteness StrategyをSVMに入力しPoliteness値を算出



[4] <https://github.com/sudhof/politeness>

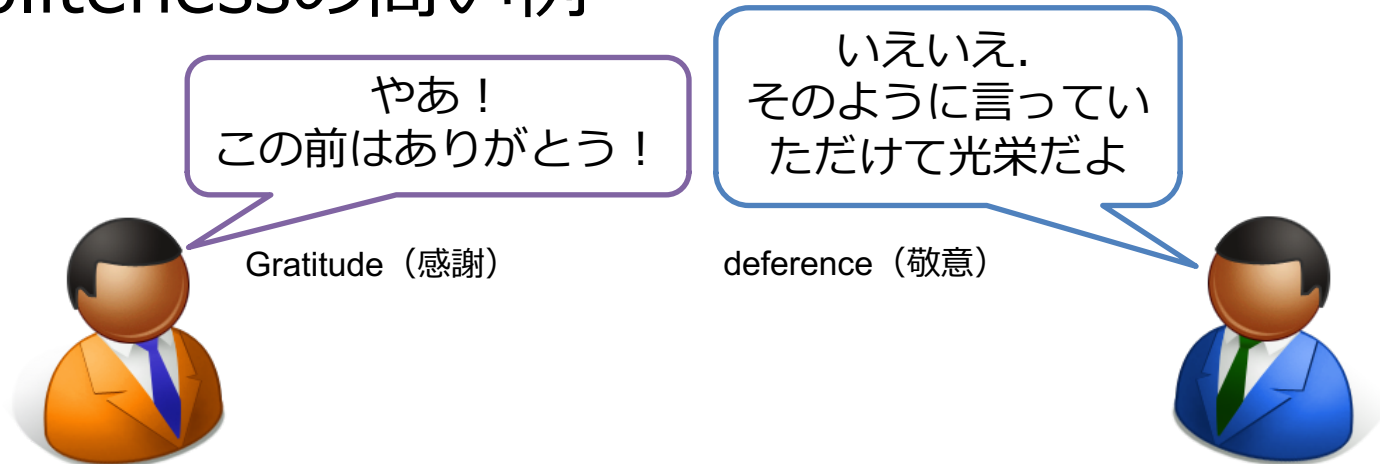
# Stanford Politenessの算出手順

1. テキストの構文解析
2. Politeness Strategyの検出
3. 学習モデルに基づいてPoliteness値を算出

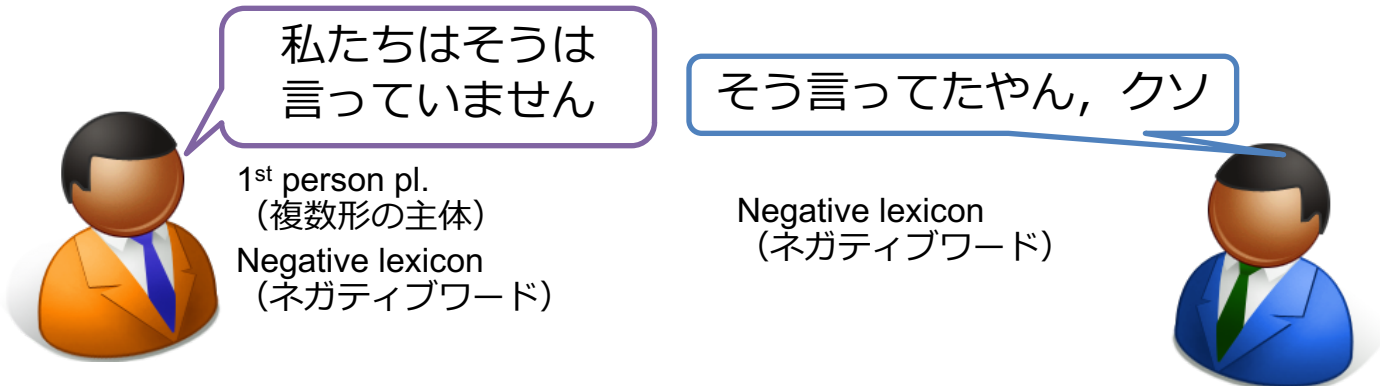


# 会話の例

- Politenessの高い例



- Politenessの低い例



# RQ1: トップ開発者間で Politenessに違いはあるか？

| HTTPD<br>開発者 | メール送信の<br>Politeness値 |      |
|--------------|-----------------------|------|
|              | 平均値                   | 中央値  |
| Dev#1        | 0.51                  | 0.48 |
| Dev#2        | 0.63                  | 0.65 |
| Dev#3        | 0.64                  | 0.66 |
| Dev#4        | 0.55                  | 0.55 |
| Dev#5        | 0.57                  | 0.57 |
| Dev#6        | 0.69                  | 0.71 |
| Dev#7        | 0.55                  | 0.53 |
| Dev#8        | 0.57                  | 0.56 |
| Dev#9        | 0.55                  | 0.54 |
| Dev#10       | 0.59                  | 0.58 |

| Python<br>開発者 | メール送信の<br>Politeness値 |      |
|---------------|-----------------------|------|
|               | 平均値                   | 中央値  |
| Dev#1         | 0.65                  | 0.67 |
| Dev#2         | 0.63                  | 0.65 |
| Dev#3         | 0.67                  | 0.71 |
| Dev#4         | 0.61                  | 0.62 |
| Dev#5         | 0.56                  | 0.55 |
| Dev#6         | 0.68                  | 0.72 |
| Dev#7         | 0.58                  | 0.58 |
| Dev#8         | 0.63                  | 0.65 |
| Dev#9         | 0.58                  | 0.57 |
| Dev#10        | 0.64                  | 0.65 |

# RQ1: トップ開発者間で Politenessに違いはあるか？

| HTTPD<br>開発者 | メール送信の<br>Politeness値 |      | Python<br>開発者 | メール送信の<br>Politeness値 |      |
|--------------|-----------------------|------|---------------|-----------------------|------|
|              | 平均値                   | 中央値  |               | 平均値                   | 中央値  |
| Dev#1        | 0.51                  | 0.48 | Dev#1         | 0.65                  | 0.67 |
| Dev#2        | 0.63                  | 0.65 | Dev#2         | 0.63                  | 0.65 |
| Dev#3        | 0.64                  | 0.66 | Dev#3         | 0.67                  | 0.71 |
| Dev#4        | 0.55                  | 0.55 | Dev#4         | 0.61                  | 0.62 |
| Dev#5        | 0.57                  | 0.57 | Dev#5         | 0.56                  | 0.55 |

中央値が最大で0.2以上の差が見られた

|        |      |      |        |      |      |
|--------|------|------|--------|------|------|
| Dev#10 | 0.59 | 0.58 | Dev#10 | 0.64 | 0.65 |
|--------|------|------|--------|------|------|

# RQ2: トップ開発者と一般開発者で Politeness に違いはあるか？

| HTTPD<br>開発者 | メール送信の<br>Politeness値 |      |
|--------------|-----------------------|------|
|              | 平均値                   | 中央値  |
| 一般開発者        | 0.63                  | 0.65 |
| Dev#1        | 0.51                  | 0.48 |
| Dev#2        | 0.63                  | 0.65 |
| Dev#3        | 0.64                  | 0.66 |
| Dev#4        | 0.55                  | 0.55 |
| Dev#5        | 0.57                  | 0.57 |
| Dev#6        | 0.69                  | 0.71 |
| Dev#7        | 0.55                  | 0.53 |
| Dev#8        | 0.57                  | 0.56 |
| Dev#9        | 0.55                  | 0.54 |
| Dev#10       | 0.59                  | 0.58 |

| Python<br>開発者 | メール送信の<br>Politeness値 |      |
|---------------|-----------------------|------|
|               | 平均値                   | 中央値  |
| 一般開発者         | 0.62                  | 0.62 |
| Dev#1         | 0.65                  | 0.67 |
| Dev#2         | 0.63                  | 0.65 |
| Dev#3         | 0.67                  | 0.71 |
| Dev#4         | 0.61                  | 0.62 |
| Dev#5         | 0.56                  | 0.55 |
| Dev#6         | 0.68                  | 0.72 |
| Dev#7         | 0.58                  | 0.58 |
| Dev#8         | 0.63                  | 0.65 |
| Dev#9         | 0.58                  | 0.57 |
| Dev#10        | 0.64                  | 0.65 |

# RQ3:メールの発信・返信で Politenessに違いはあるのか

| HTTPD<br>開発者 | 発信メール |      | 返信メール |      | Python<br>開発者 | 発信メール |      | 返信メール |      |
|--------------|-------|------|-------|------|---------------|-------|------|-------|------|
|              | 平均値   | 中央値  | 平均値   | 中央値  |               | 平均値   | 中央値  | 平均値   | 中央値  |
| Dev#1        | 0.57  | 0.58 | 0.50  | 0.47 | Dev#1         | 0.77  | 0.84 | 0.65  | 0.66 |
| Dev#2        | 0.69  | 0.73 | 0.62  | 0.64 | Dev#2         | 0.73  | 0.77 | 0.63  | 0.65 |
| Dev#6        | 0.77  | 0.84 | 0.57  | 0.57 | Dev#3         | 0.65  | 0.66 | 0.67  | 0.71 |
| Dev#9        | 0.50  | 0.43 | 0.61  | 0.61 | Dev#4         | 0.80  | 0.88 | 0.60  | 0.60 |

メールを発信する場合と返信する場合で  
Politenessの平均値と中央値に違いがあった