

卒業研究

OSSコミュニティへの参加支援を 目的とした新人開発者のGitHubと StackOverflowでの活動分析

オープンソースソフトウェア工学研究室

学籍番号 60210105

氏名 堀口 日向



1. 背景 | OSSについて

OSS (Open Source Software)



オペレーティングシステム



OpenJDK



ソフトウェア開発環境



Webアプリケーション実行環境

ソフトウェア開発にOSSは必要不可欠



1. 背景 | OSSプロジェクトの持続可能性 (1/2)

OSS (Open Source Software)



貢献 ⇒

プロジェクト存続のために重要^[1]



OSSコミュニティ

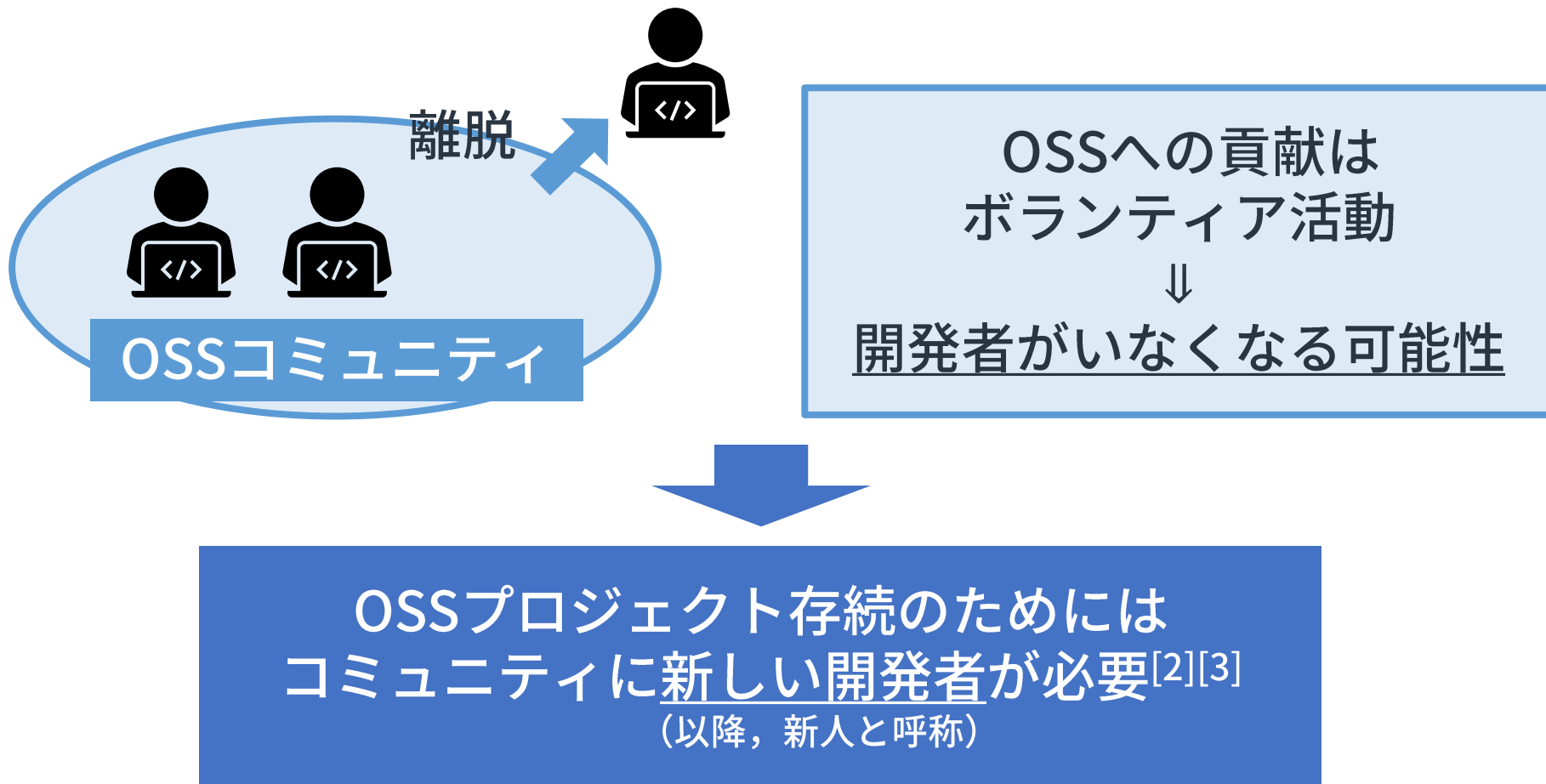
貢献

- バグの報告や修正
- 機能拡張

[1] Kevin Crowston, Hala Annabi, and James Howison. Defining open source software project success. In ICIS, pp. 327-340, 2003.



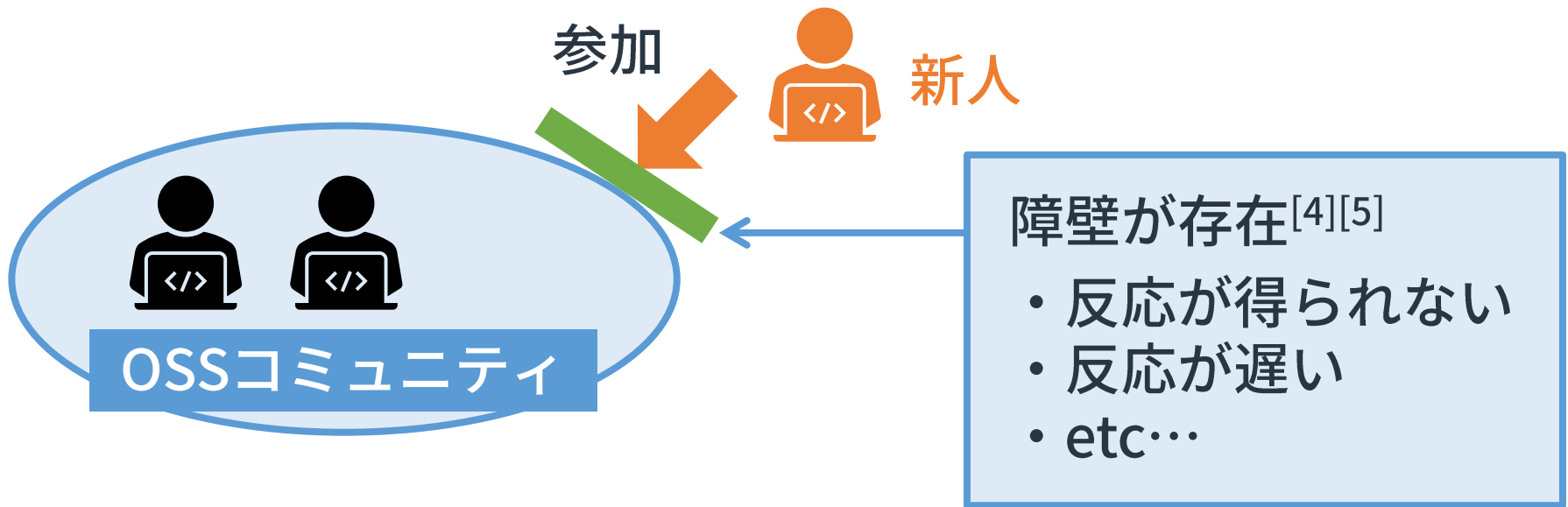
1. 背景 | OSSプロジェクトの持続可能性 (2/2)



[2] Israr Qureshi and Yulin Fang. Socialization in Open Source Software Projects: A Growth Mixture Modeling Approach. Organizational Research Methods, pp. 208–238, 2011.

[3] Forte, A. and Lampe, C. Defining, Understanding, and Supporting Open Collaboration: Lessons From the Literature. American Behavioral Scientist 57, 5 (2013), 535–547.

1. 背景 | OSSコミュニティへの新人の参加障壁



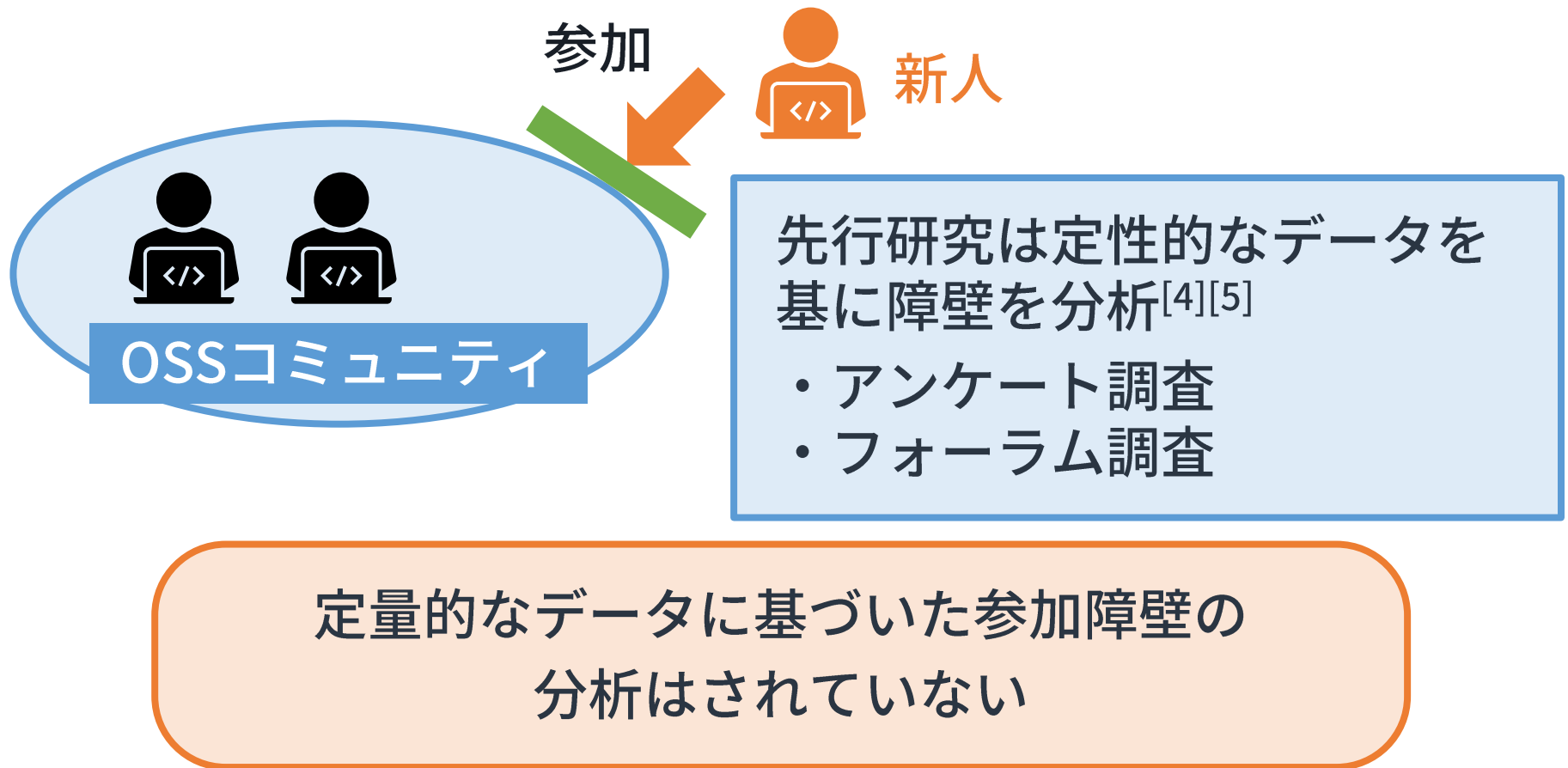
特にOSS開発経験のない初心者は障壁に直面しやすい^[4]

[4] Igor Steinmacher, Tayana Conte, Marco Aur´elio Gerosa, and David Redmiles. Social barriers faced by newcomers placing their first contribution in open source software projects. In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, pp. 1379–1392, 2015.

[5] Singh, Vandana. Newcomer Integration and Learning in Technical Support Communities for Open Source Software. In Proceedings of the 17th ACM International Conference on Supporting Group Work, pp. 65–74, 2012.



2. 本研究の着眼点（1/3）



[4] Igor Steinmacher, Tayana Conte, Marco Aur´elio Gerosa, and David Redmiles. Social barriers faced by newcomers placing their first contribution in open source software projects. In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, pp. 1379–1392, 2015.

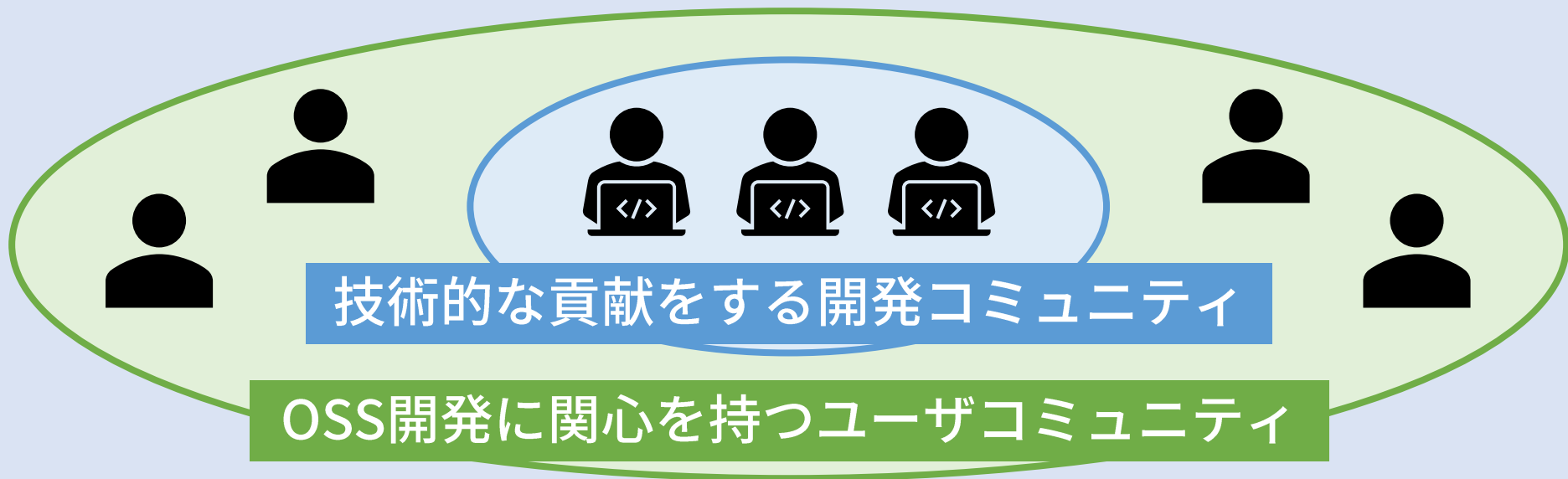
[5] Singh, Vandana. Newcomer Integration and Learning in Technical Support Communities for Open Source Software. In Proceedings of the 17th ACM International Conference on Supporting Group Work, pp. 65–74, 2012.

2. 本研究の着眼点 (2/3)

狭義のOSSコミュニティ

OSSプロジェクトに対し技術的な貢献をする開発者ら

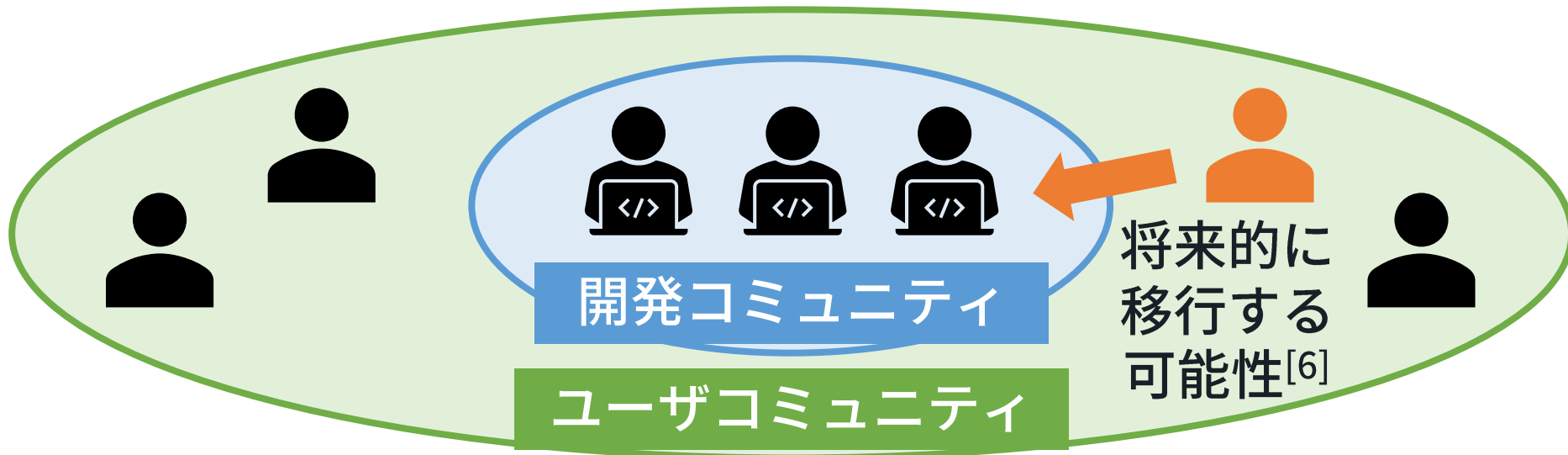
広義のOSSコミュニティ^[6]



[6] Yunwen Ye and Kouichi Kishida. Toward an understanding of the motivation open source software developers. In Proceedings of the 25th International Conference on Software Engineering, pp. 419–429, 2003.



2. 本研究の着眼点 (3/3)



ユーザコミュニティに新人が参加しやすいかどうかは
開発コミュニティを維持するためにも重要

ユーザコミュニティへの参加障壁については
分析されていない

[6] Yunwen Ye and Kouichi Kishida. Toward an understanding of the motivation open source software developers. In Proceedings of the 25th International Conference on Software Engineering, pp. 419–429, 2003.



3. 目的とアプローチ

目的

OSSの開発コミュニティとユーザコミュニティの
新人の参加障壁を定量的に明らかにする

アプローチ

新人と新人以外の投稿に対する反応の違いを
コミュニティごとに仮説検定で確かめる

使用するデータ

開発コミュニティ

 GitHubの
IssueとPull Request

ユーザコミュニティ

 StackOverflowの
質問



4. GitHubについて | Issue

- Issueとは

バグ報告や新機能の
提案(Issue)を投稿



コメント欄で
コミュニティ
メンバーと議論

Bug with namespaced ActiveRecord models on 4.1.x (PostgreSQL) #15811

① Open

thelucid opened this issue on 19 Jun 2014 · 25 comments



thelucid commented on 19 Jun 2014

Contributor + 😊 ...

I've come up against a bug with ActiveRecord when using namespaces models and polymorphic associations, it's a common use-case so should probably be fixed before a 4.2 release.

Error:

```
NoMethodError: undefined method `relation_delegate_class' for Comment:Module
```



rafaelfranca commented on 20 Jun 2014

Member + 😊 ...

Which version are you using? Have you tried the latest 4.1.2.rc2? Could you create a reproduction script with this template
https://github.com/rails/rails/blob/master/guides/bug_report_templates/active_record_master.rb?



rafaelfranca added the **activerecord** label on 20 Jun 2014



taish commented on 23 Jun 2014

+ 😊 ...

I think it is problem and should be fixed

I create an reproduction script [here](#)

This test should be passed but actually failed

I'll try to solve



4. GitHubについて | Pull Request

- Pull Requestとは

開発者がバグ修正や機能拡張のため、ソースコードの変更を投稿(Pull Request)



プロジェクトの中心メンバーがレビュー追加の修正，議論を行う



最終的に変更を統合するか判断

```
▼ 5 ■■■■■ activerecord/lib/active_record/core.rb
@@ -552,6 +552,11 @@ def slice(*methods)
552 552      Hash[methods.flatten.map! { |method| [met
553 553      end
554 554
555 +      # Returns an array of the given methods wit
556 +      def values_at(*methods)
557 +      methods.flatten.map! { |method| public_se
```

This conversation was marked as resolved



rafaelfranca on 28 Jul Member

Does not `flat_map` work here?



guillaumebriday on 29 Jul Author

it does not. Because:

```
a.flat_map(&b) works exactly like a.map(&b).flatten!(1).
```

So if you send an array as parameter, it will try to `public_send` on an array

I added test for array as params if needed

中心メンバー

投稿者



5．分析対象データ | プロジェクトの選定

1.  GitHubと  StackOverflowの両方でコミュニティを持つOSSプロジェクトを目視で抽出
2. 分析に十分なデータを確保するため、  StackOverflowでの質問数が多い順に6つのプロジェクトを選択

プロジェクト名	StackOverflowでの質問数
Ruby on Rails	311,356
Node.js	297,653
Angular.js	259,240
Swift	247,299
Django	212,227
Angular	191,240



5. 分析対象データ | データ収集

3.  GitHubのWebAPIから、対象プロジェクトのリポジトリが保有するIssueとPull Requestを収集
4.  StackOverflowのWebAPIから、対象プロジェクトを示すタグが付与されている質問を収集

収集できたデータ件数

データ名	件数
 GitHubのIssue	158,011
 GitHubのPull Request	106,082
 StackOverflowの質問	1,519,015



6. 分析 | 開発コミュニティの新人の参加障壁

分析方法

1. 参加障壁に関する定量的なデータを取得

先行研究で指摘された以下の2つの参加障壁

- コミュニティからの反応が得られない
 - Issue：コメントの有無（名義尺度）
 - Pull Request：統合されたかどうか（名義尺度）
- コミュニティからの反応が遅い
 - Issue：最初のコメントが来るまでの時間（比率尺度）
 - Pull Request：統合されるまでの時間（比率尺度）



6. 分析 | 開発コミュニティの新人の参加障壁

分析方法

2. 仮説検定

新人と新人以外の間で  開発コミュニティからの反応に違いがあるかどうかを仮説検定

- 名義尺度

カイ二乗検定（有意水準：1%），効果量はCohen's w ^[7]

- 比率尺度

マンホイットニーのU検定（有意水準：1%），効果量はCliff's d ^[8]

[7] Jacob Cohen. Statistical Power Analysis for the Behavioral Sciences. 2 edition, 1988.

[8] Jeanine Romano, Jeffrey D. Kromrey, Jesse Coraggio, and Jeff Skowronek. Appropriate Statistics for Ordinal Level Data: Should We Really be Using t-test and Cohen's d for Evaluating Group Differences on the NSSE and other Surveys? In annual meeting of the Florida Association of Institutional Research, 2006.



6. 分析 | 開発コミュニティの新人の参加障壁

分析方法

2. 仮説検定

新人と新人以外の間で  開発コミュニティからの反応に違いがあるかどうかを仮説検定

- 新人と新人以外の判別

過去に1度もPull Requestを投稿したことがない開発者

⇒ 新人



6. 分析 | 開発コミュニティの新人の参加障壁

分析結果

新人と新人以外の投稿したIssueの違い

指標	p値	効果量
コメントの有無	0.000 (有意差あり)	0.05 (無視できる程小さい)
最初のコメントが寄せられるまでの時間	0.000 (有意差あり)	0.27 (小程度)

小さいが障壁がある可能性

新人と新人以外の投稿したPull Requestの違い

指標	p値	効果量
統合されたかどうか	0.000 (有意差あり)	0.05 (無視できる程小さい)
統合されるまでの時間	0.058 (有意差なし)	0.04 (無視できる程小さい)

有意差はあるものの効果量は小さい

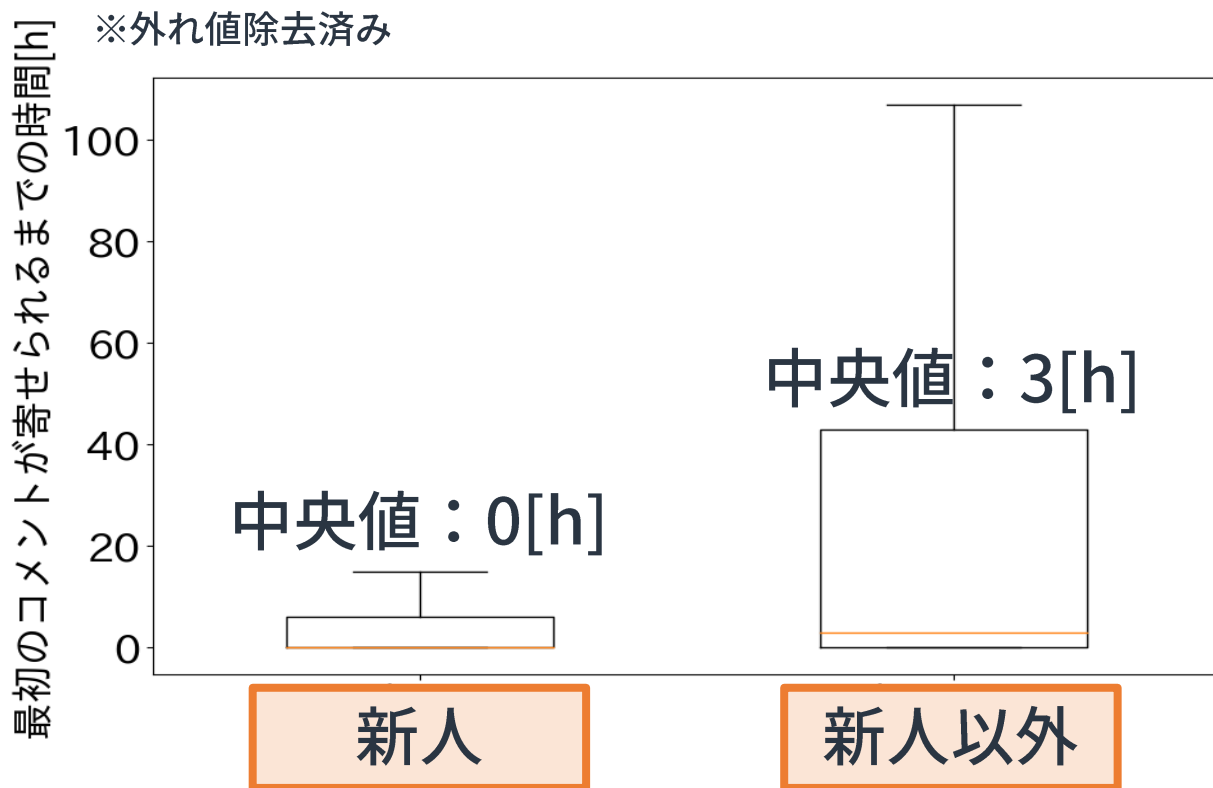
⇒  開発コミュニティに参加障壁があるとはいえない



6. 分析 | 開発コミュニティの新人の参加障壁

分析結果

指標	p値	効果量
最初のコメントが寄せられるまでの時間	0.000 (有意差あり)	0.27 (小程度)



新人への
反応の方が早い
↓
参加障壁はなく
むしろ新人が
ケアされている？
考察で追加調査



6. 分析 | ユーザコミュニティの新人の参加障壁

分析方法

1. 参加障壁に関する定量的なデータを取得

分析する参加障壁については🕒開発コミュニティと同様

- ・コミュニティからの反応が得られない
質問への回答の有無（名義尺度）
- ・コミュニティからの反応が遅い
質問への最初の回答が来るまでの時間（比率尺度）

2. 仮説検定

🕒開発コミュニティの分析と同様



6. 分析 | ユーザコミュニティの新人の参加障壁

分析結果

新人と新人以外の投稿した質問の違い

指標	p値	効果量
回答の有無	0.000 (有意差あり)	0.03 (無視できる程小さい)
最初の回答が寄せられるまでの時間	0.005 (有意差あり)	0.01 (無視できる程小さい)

有意差はあるものの効果量は小さい

⇒  ユーザコミュニティに参加障壁があるとはいえない



6. 分析 | OSSコミュニティの新人の参加障壁

分析結果まとめ

 開発コミュニティ,  ユーザコミュニティの
新人の参加障壁があるとはいえない



先行研究の質的調査とは異なる結果

先行研究と異なる結果となった理由を考察するため
効果量が最も大きかった指標に着目し追加調査



7. 考察 | 開発コミュニティの追加調査

調査方法

- ① 「最初のコメントが寄せられるまでの時間」が中央値と一致するIssueの最初のコメントを抽出
- ② 新人，新人以外が投稿したIssueについて各10件ランダムに抽出して目視し，特徴を見つける



7. 考察 | 開発コミュニティの追加調査

新人のIssueへの
最初のコメントの内容

- 不完全な内容を補う
ための質問
- 解決策の提示

新人以外のIssueへの
最初のコメントの内容

- 投稿内容についての議論

新人は  開発コミュニティからケアされている



7. 考察 | 先行研究との比較

先行研究^{[4][5]}

新人はコミュニティ
メンバーに相手にされない
場合がある

本研究

新人はコミュニティ
メンバーによってケアされる
場合がある



分析対象コミュニティの新人に対する扱いが異なる

先行研究：独自にフォーラムを持つOSSプロジェクト

本研究：GitHub上のOSSプロジェクト

⇒ コミュニティの性質が異なることが扱いの違いの原因？

[4] Igor Steinmacher, Tayana Conte, Marco Aurélio Gerosa, and David Redmiles. Social barriers faced by newcomers placing their first contribution in open source software projects. In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, pp. 1379–1392, 2015.

[5] Singh, Vandana. Newcomer Integration and Learning in Technical Support Communities for Open Source Software. In Proceedings of the 17th ACM International Conference on Supporting Group Work, pp. 65–74, 2012.

8. 本研究のまとめ

定量的な分析では、

 開発コミュニティ， ユーザコミュニティ双方で
新人の参加障壁があるとはいえない

先行研究と異なる結果となった理由

追加調査の結果，先行研究と本研究では
分析対象コミュニティの新人に対する扱いが異なる

今後の展望

先行研究と本研究のコミュニティについて比較すれば
参加障壁の有無とコミュニティの特徴の関係性が判明する可能性

大学院での研究

初めてのOSSプロジェクトに 貢献する開発者のための Issue推薦手法の提案

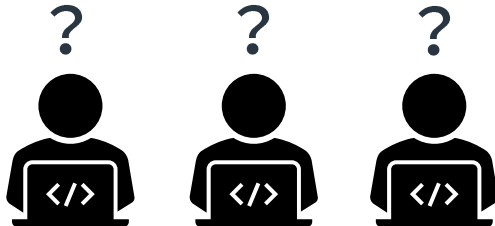
オープンソースソフトウェア工学研究室

学籍番号 60210105

氏名 堀口 日向

1. 背景

OSS (Open Source Software)



自分の興味関心やスキルアップのため
OSSプロジェクトに貢献したい^{[9],[10]}

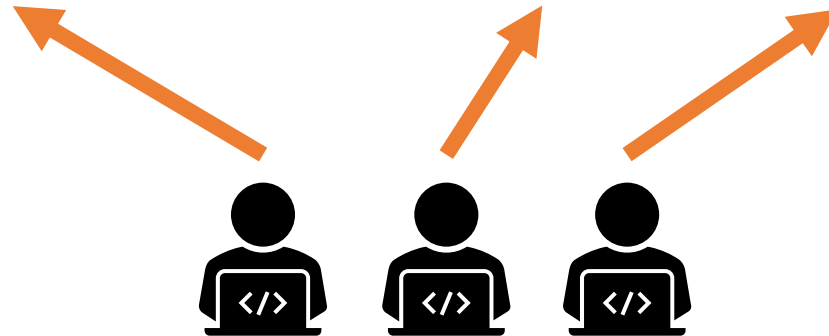
どのプロジェクトに貢献すべきかわからない^[9]

[9]Capiluppi, A. and Michlmayr, M. From the Cathedral to the Bazaar: An Empirical Study of the Lifecycle of Volunteer Community Projects, Open Source Development, Adoption and Innovation, pp. 31–44, 2007

[10] Choi, B. et al, Socialization Tactics in Wikipedia and Their Effects. Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work, pp. 107–116, 2010

1. 背景

OSS (Open Source Software)

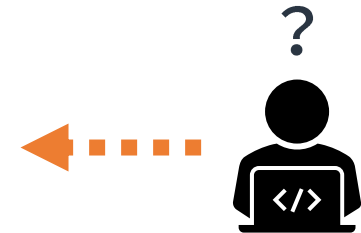
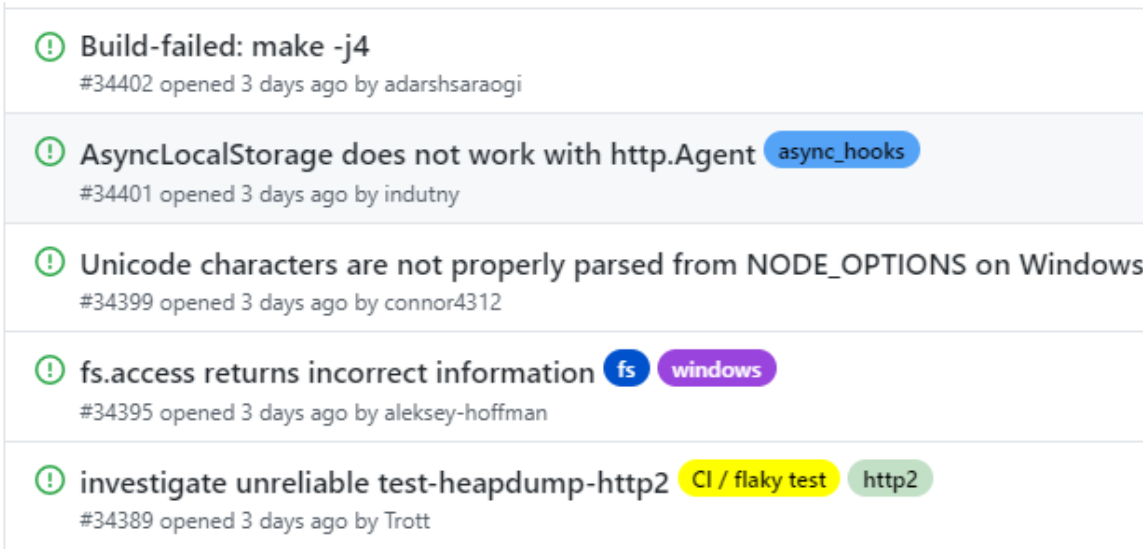


先行研究では
開発者の過去の貢献履歴に基づいた
OSSプロジェクトを推薦^[11]

[11] C. Liu, D. Yang, X. Zhang, B. Ray and M. M. Rahman, Recommending GitHub Projects for Developer Onboarding, in IEEE Access, vol. 6, pp. 52082-52094, 2018

1. 背景

プロジェクト内のタスク（例：Node.jsのIssue）



最初にどのタスクを選択すればいいかわからないため
コミュニティからのサポートが必要^[12]

プロジェクトを推薦するだけでは不十分

[12] I. Steinmacher, T. U. Conte and M. A. Gerosa, Understanding and Supporting the Choice of an Appropriate Task to Start with in Open Source Software Communities, 48th Hawaii International Conference on System Sciences, pp. 5299-5308, 2015

2. 目的とアプローチ

目的

初めてのOSSプロジェクトに貢献する開発者が
解決すべきIssueを推薦する手法を提案

アプローチ

プロジェクトによってラベル付けされた
初めて貢献する開発者向けのIssueを分析



普通のIssueと，初めて貢献する開発者向けのIssueの
違いを明らかにし，推薦手法を考案

3．今後の予定

- 修士1年

GitHub上のgood first issueラベルが付与されたIssueと付与されていないIssueの違い分析

- Issueの内容
- ラベル付与の際にかかるコスト（時間，人数）
- Issueの難易度
- etc…

- 修士2年

分析結果を基に，初めてのOSSプロジェクトに貢献する開発者向けのIssue推薦手法を考案

補足資料．StackOverflowについて



プログラミングに関する
質問・回答を行う
知識共有コミュニティ

開発者が質問



別の開発者が回答

Rails/Ruby - Format string into Date with locale

Asked today Active today Viewed 20 times

Is there a way in rails with localize or in plain ruby with Date.strptime to convert a string into a date with a specific months locales?

For example, I have the following string and I need it to convert it to a Date object

ruby-on-rails ruby date

share edit

add a comment

New contributor

質問に関連する技術名や
プロダクト名でタグを付ける

1 Answer

active oldest votes

If you have the date month in english, you can use Date.strptime with this format:

```
date_str = '5 apr 19'  
date = Date.strptime(date_str, '%d %b %y')
```

For other languages (excepts english I would just translate the month, only a few months changes:

```
{ 'abr' => 'apr', 'ago' => 'aug', 'dic' => 'dec' }  
gsub(/ene|abr|ago|dic/, months) # only replace the months that should  
e_str_aux, '%d %b %y')
```

share edit

answered 3 hours ago

arielljuod
10.5k ●2 ●15 ●27

補足資料． StackOverflowの選定理由

GitHub（開発コミュニティ）との関係を示す先行研究

- StackOverflowへの参加はGitHub開発者の生産性を向上させる^[9]
- GitHub開発者はStackOverflowのコードスニペットを頻繁に利用する^[10]

GitHubとの関係性があるユーザコミュニティを調べれば、ユーザコミュニティから開発コミュニティに移行した開発者が多く含まれる可能性がある

[9] Vasilescu, B., Filkov, V., and Serebrenik, A. StackOverflow and GitHub: Associations between software development and crowdsourced knowledge. International Conference on Social Computing , pp. 188-195, 2013.

[10] Manes, S. S., and Baysal, O. How often and what stackoverflow posts do developers reference in their GitHub projects?. In Proceedings of the 16th International Conference on Mining Software Repositories (MSR '19), pp 235–239, 2019.

補足資料．分析対象のOSSプロジェクトの詳細

分析対象プロジェクトのリポジトリ名とタグ名の対応

プロジェクト名	リポジトリ名	タグ名	対象期間
Ruby on Rails	rails/rails	ruby-on-rails	2008/08/02 ~ 2019/11/01
Node.js	nodejs/node	node.js	2008/11/11 ~ 2019/11/07
Angular.js	angular/angular.js	angularjs	2009/10/19 ~ 2019/11/11
Swift	apple/swift	swift	2009/01/19 ~ 2019/11/13
Django	django/django	django	2008/08/03 ~ 2019/11/14
Angular	angular/angular	angular	2013/04/15 ~ 2019/11/18

分析結果

指標	p値	効果量
コメントの有無	0.000（有意差あり）	0.05（無視できる程小さい）

クロス集計表

		コメントがあった	コメントがなかった	合計
新人の Issue 投稿	度数	8915	24483	33398
	期待度数	9325.10	24072.90	33398.00
	調整残差	-10.86	10.86	
新人以外の Issue 投稿	度数	2923	6077	9000
	期待度数	2512.90	6487.10	9000.00
	調整残差	10.86	-10.86	
合計		11838	30560	42398

新人のIssue
に対して
コメントが
ない傾向

分析結果

指標	p値	効果量
統合されたかどうか	0.000（有意差あり）	0.05（無視できる程小さい）

クロス集計表

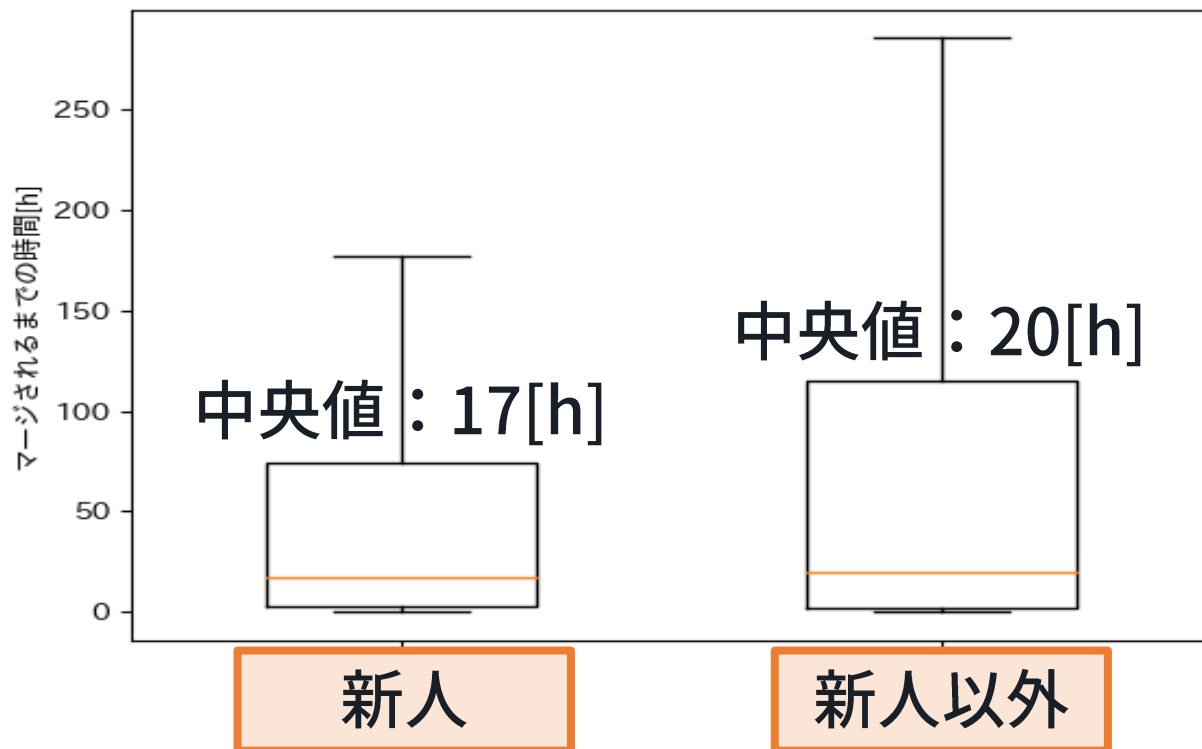
		マージされた	マージされなかった	合計
新人の PR 投稿	度数	1048	1548	2596
	期待度数	857.00	1739.00	2596.00
	調整残差	8.36	-8.36	
新人以外の PR 投稿	度数	8312	17445	25757
	期待度数	8503.00	17254.00	25757.00
	調整残差	-8.36	8.36	
合計		9360	18993	28353

新人のPR
の方が統合
されやすい
傾向

分析結果

指標	p値	効果量
統合されるまでの時間	0.058（有意差なし）	0.04（無視できる程小さい）

※外れ値除去済み



新人のPRの方が
統合されるまでの
時間が短い

補足資料．分析 | ユーザコミュニティの新人の参加障壁 | 回答の有無

分析結果

指標	p値	効果量
回答の有無	0.000（有意差あり）	0.03（無視できる程小さい）

クロス集計表

		回答があった	回答がなかった	合計
新人の質問	度数	4217	33642	37859
	期待度数	4391.70	33467.30	37859.00
	調整残差	-5.96	5.96	
新人以外の質問	度数	1418	9300	10718
	期待度数	1243.30	9474.70	10718.00
	調整残差	5.96	-5.96	
合計		5635	42942	48577

新人の質問
に対して
回答がない
傾向

分析結果

指標	p値	効果量
最初の回答が寄せられるまでの時間	0.005（有意差あり）	0.01（無視できる程小さい）

※外れ値除去済み



時間の
分布に
違いはない

補足資料．考察 | 開発コミュニティの追加調査

新人のIssueに対する最初のコメントの内訳

- 投稿内容に関する追加の説明を求めるコメント（例を求めるなど）：4件
- 投稿者とコミュニティの間に議論する余地がないような投稿内容で最初のコメントで投稿された問題に対する解決策が提示されている：4件
- 最初のコメントを含めて投稿内容について議論があったのが：2件

新人以外のIssueに対する最初のコメントの内訳

- Issueの投稿内容に関する追加の説明を求めるコメントが：1件
- 最初のコメントで投稿された問題に対する解決策が提示されている：2件
- 議論があったのが：5件
- 投稿内容の重複を示すコメント：2件