

発表の前に・・・

- ・ 来月締め切りのSANER2021のERAトラック（初期段階の研究を発表する場）に投稿予定の内容をまとめた
- ・ 客観的な意見や疑問など是非
- ・ 批判的な意見もウェルカム
- ・ 論文投稿に向けてブラッシュアップしていきたい！

Preliminary analysis of good first issues in GitHub to recommend issues for newcomers

新人向けIssue推薦のための GitHub上のgood first issueの予備分析

M1 堀口 日向

OSS (Open Source Software)



貢献 ⇒

プロジェクト存続のために重要^[1]

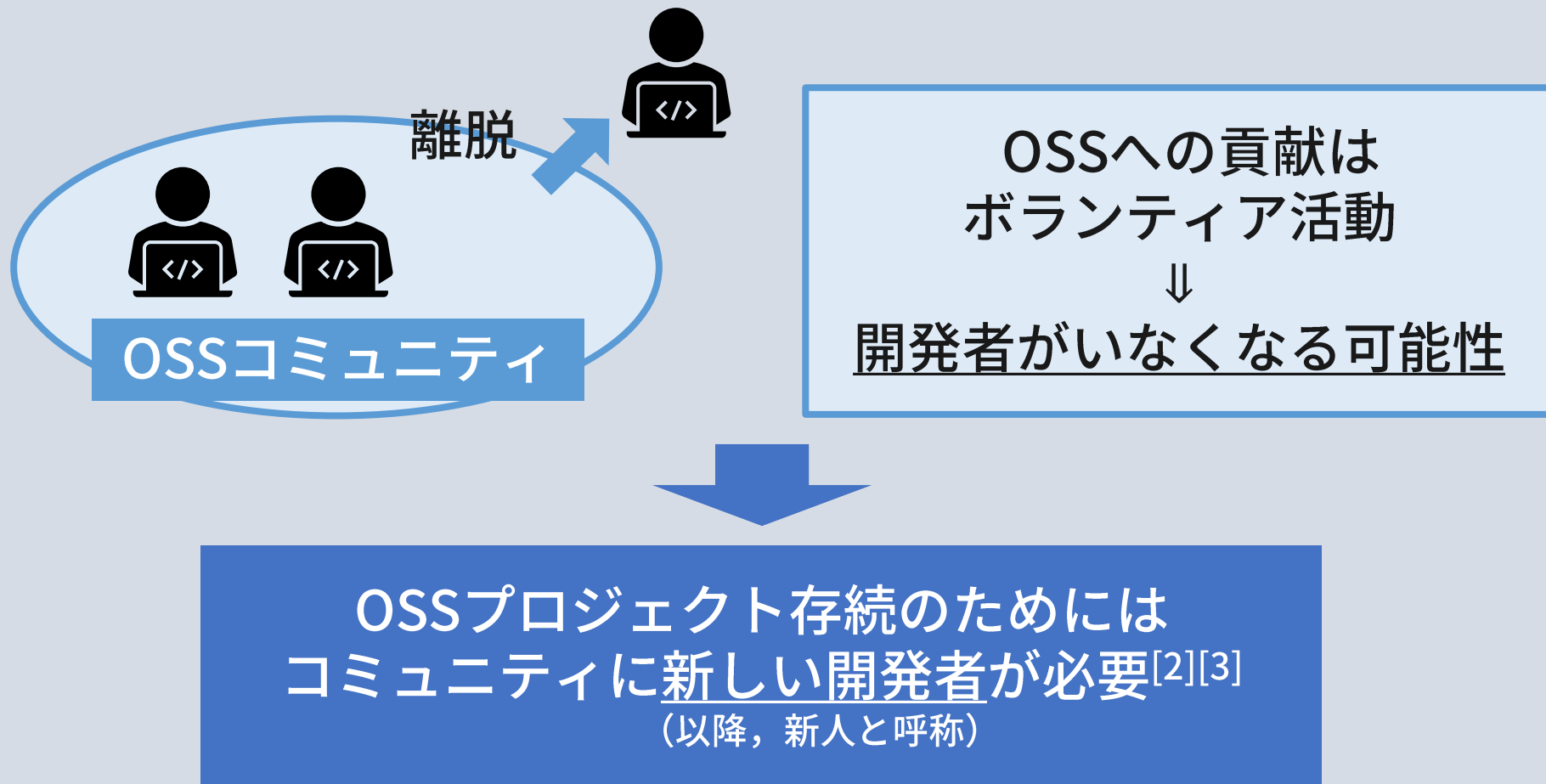


OSSコミュニティ

貢献

- バグの報告や修正
- 機能拡張

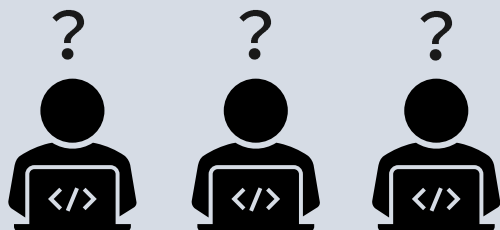
[1] Kevin Crowston, Hala Annabi, and James Howison. Defining open source software project success. In ICIS, pp. 327–340, 2003.



[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.

OSS (Open Source Software)



新人

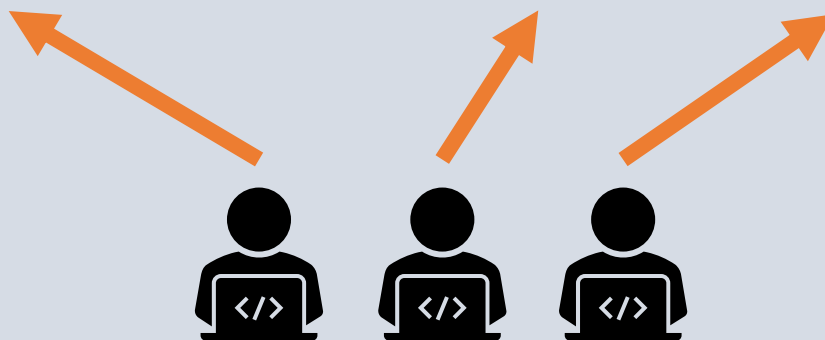
自分の興味関心やスキルアップのため
OSSプロジェクトに貢献したい^{[4],[5]}

どのプロジェクトに貢献すべきかわからない^[4]

[4]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

[5] Choi, B. et al, Socialization Tactics in Wikipedia and Their Effects. Proceedings of the 2010 ACM Conference on ComputerSupported Cooperative Work, pp. 107-116, 2010

OSS (Open Source Software)

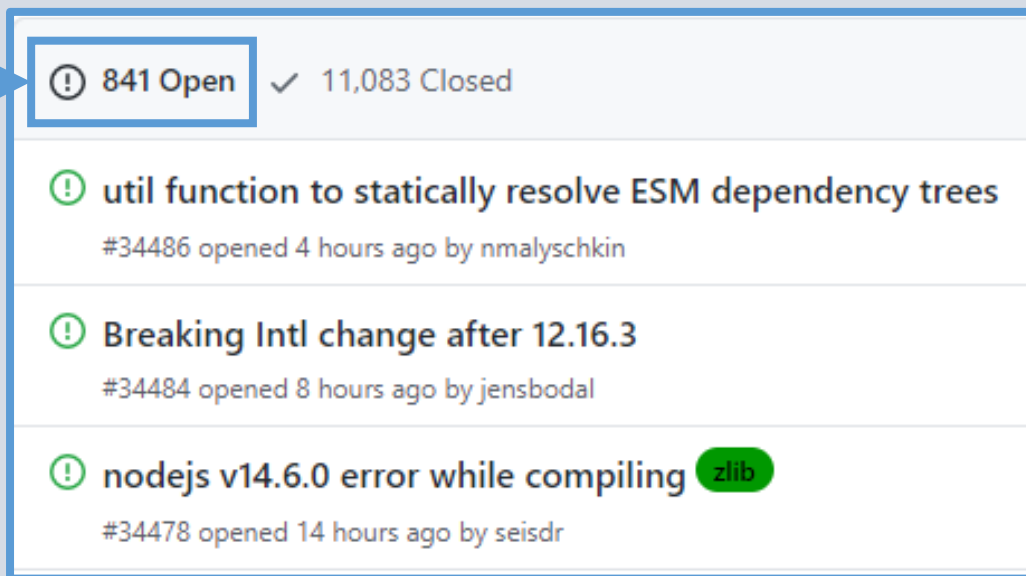


先行研究では
新人の過去の貢献履歴に基づき
OSSプロジェクトを推薦^[11]

[6] 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

プロジェクト内のタスクの例 (GitHubのnodejs/nodeリポジトリ)

タスクの数



841 Open ✓ 11,083 Closed

- util function to statically resolve ESM dependency trees
#34486 opened 4 hours ago by nmalyschkin
- Breaking Intl change after 12.16.3
#34484 opened 8 hours ago by jensbodal
- nodejs v14.6.0 error while compiling **zlib**
#34478 opened 14 hours ago by seisdr

最初にどのタスクに取り組め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^[7]

プロジェクトを推薦するだけでは不十分

[7] 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

背景 | good first issue

Display process warning stack traces by default #33349

Open szmarczak opened this issue on 11 May · 8 comments

投稿者 (szmarczak):

Is your feature request related to a problem? Please describe.

Recently I've come to an issue, where the author provided only `TimeoutOverflowWarning` without any stack traces v therefore I'm unable to locate the source.

Describe the solution you'd like

I'd propose to enable process warning stack traces by default, so people don't have to `node --trace-warnings index` would be another option to disable that, `--no-trace-warnings`.

3 likes

メンテナ (mmarchini):

Sounds good to me. This looks like a good first issue, unless I'm wrong? (I added the label, if anyone thinks that's not an issue feel free to remove it).

2 likes

mmarchini added the **good first issue** label on 20 May

メンテナがラベル付与

GitHubでは新人が最初に取り組みべきタスクにメンテナが「good first issue」ラベルを付与する慣習

運用していないリポジトリも多い

ラベル付与はメンテナの負担

good first issueの例 (GitHubのnodejs/nodeリポジトリ)

目的，リサーチクエスト

目的

将来的に新人向けのIssue推薦手法を構築するため，
good first issueが新人の最初に取り組むべき
タスクとして有用かを確かめる

- RQ1：good first issueを解決した開発者のOSS開発経験はどの程度か？
- RQ2：good first issueは解決しやすいのか？
- RQ3：good first issue解決後にPRを投稿した開発者数はどの程度か？

データセット作成

1. good first issueの収集

GitHub上の「good first issue」ラベルが付与されたすべてのIssueを，WebAPI経由で収集

2. 分析対象プロジェクトのフィルタリング

多くの新人が関わっていることが望まれるため，good first issueの数が500件以上の上位11プロジェクトを選択

3. プロジェクトからデータ収集

各RQの分析に利用するため，Issue，Pull Request等のデータをWebAPI経由で収集

データセット概要

上位11プロジェクトのリポジトリ一覧

repository
osmlab/name-suggestion-index
radareorg/radare2
zulip/zulip
mui-org/material-ui
pandas-dev/pandas
CleverRaven/Cataclysm-DDA
elastic/elasticsearch
rust-lang/rust-clippy
tgstation/tgstation
WordPress/gutenberg
elastic/kibana

上位11プロジェクトから収集したデータ数

	total number
issues	140,421
good first issues	9,475
pull requests	200,838
resolved issues	43,373
resolved good first issues	4,970
developers	4,595
developers who resolved good first issue	1,986

動機

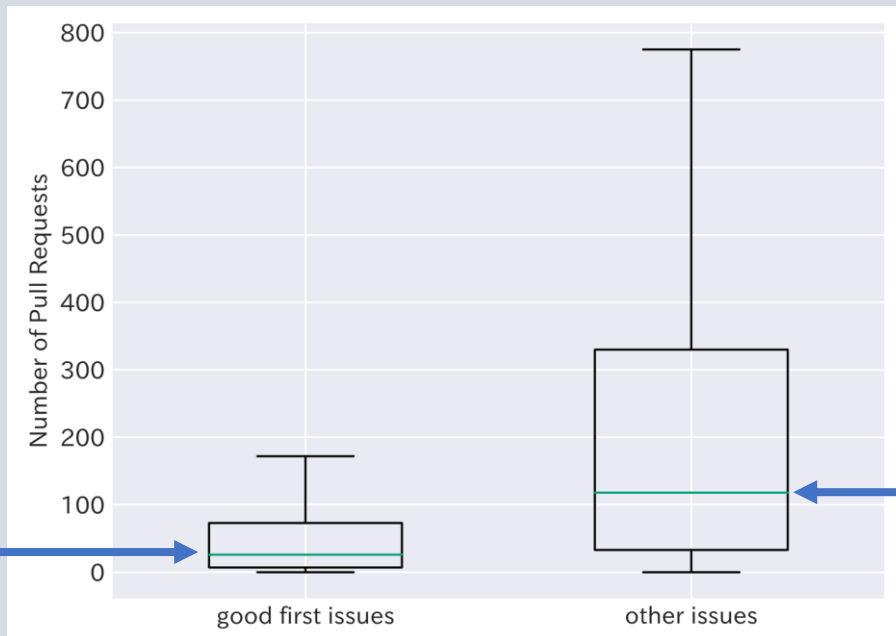
good first issueはOSS開発経験が少ない開発者でも取り組めるのか確かめるため

アプローチ

good first issueを解決した開発者群と，普通のIssueを解決した開発者群で，Issue解決前に投稿したことがあるPull Requestの数の分布を，箱ひげ図で比較する

RQ1 : good first issueを解決した開発者のOSS開発経験はどの程度か？ | 結果

開発者が
Issue解決前に
投稿したことがある
Pull Request数[件]



中央値 : 26[件]

中央値 : 118[件]

good first issue
を解決した開発者

普通のIssueを
解決した開発者

good first issueを解決している開発者はPR投稿数が少ない



good first issueは、OSS開発経験の少ない開発者にとって取り組みやすい

RQ2. good first issueは解決しやすいのか？ | 方法

動機

good first issueはOSS開発経験が少ない開発者でも解決できるのかを確かめるため

アプローチ

それぞれのプロジェクトで、good first issueと普通のIssueが解決される割合を比較する

RQ2. good first issueは解決しやすいのか？ | 結果

リポジトリ	普通のIssueが 解決される割合[%]	good first issueが 解決される割合[%]
mui-org/material-ui	21.0	92.9
tgstation/tgstation	45.0	80.5
CleverRaven/ Cataclysm-DDA	37.6	71.6
pandas-dev/pandas	26.3	67.3
WordPress/gutenberg	24.1	61.7
elastic/kibana	28.2	60.2
rust-lang/rust-clippy	24.8	56.4
elastic/elasticsearch	25.1	55.2
osmlab/ name-suggestion-index	7.3	43.6
radareorg/radare2	9.0	13.5
zulip/zulip	7.2	10.6

全てのプロジェクトで
good first issueの方が
解決される割合が高い



OSS開発経験の少ない
開発者にとって
good first issueは
解決しやすい

ただし
プロジェクトごとに
解決される割合に
大きな差がある

なぜプロジェクトごとに大きな差があるのか

考えられる原因

- good first issueの内容がプロジェクトによって違う
bug, enhancement, documentation, etc…
- プロジェクトのその他特徴
 - ソフトウェアの種類
UI, ゲーム, チャットツール, etc…
 - 成熟度
 - 新人へのケアの程度
 - etc…

確かめるためには追加分析が必要

動機

good first issueを解決することによって，次以降の貢献につながるのかを確かめるため

アプローチ

good first issue解決後に，普通のIssueを解決するためのPR投稿をした開発者の割合をプロジェクトごとに集計

(複数のgood first issueを解決している開発者は，最後に解決したgood first issueから起算)

RQ3. good first issue解決後にPRを投稿した開発者数はどの程度か？ | 結果

リポジトリ	good first issueが 解決される割合[%]	PR投稿した 開発者の割合[%]
mui-org/material-ui	92.9	24.9
tgstation/tgstation	80.5	83.9
CleverRaven/ Cataclysm-DDA	71.6	70.6
pandas-dev/pandas	67.3	35.8
WordPress/gutenberg	61.7	62.3
elastic/kibana	60.2	80.4
rust-lang/rust-clippy	56.4	40.9
elastic/elasticsearch	55.2	60.3
osmlab/ name-suggestion-index	43.6	33.8
radareorg/radare2	13.5	63.5
zulip/zulip	10.6	80.4

赤：50%以下

青：50%以上

プロジェクトごとに
PR投稿する開発者の
割合に大きな差がある

good first issueが
解決しやすくても
good first issue解決後に
PR投稿する割合が高い
とは限らない

なぜgood first issueが解決される割合が高いのに、
解決後PR投稿する開発者が少ない？

考えられる原因

- ・ good first issueの内容と普通のIssueの内容に大きな違いがあり， good first issueが解決できたとしても，普通のIssueを解決することができない可能性

確かめるためには追加分析が必要

- good first issueは、OSS開発経験の少ない開発者にとって取り組みやすく、解決しやすい
 - ⇒ 新人向けIssue推薦の教師データとして、good first issueを利用できる
- 一方、good first issue解決後に次の貢献につながる割合は、プロジェクトごとに差がある
 - ⇒ good first issue解決後にも新人に貢献してもらうためには何らかの工夫が必要

- GitHubが推薦するIssueと good first issueの比較

GitHubが公開している新人向けIssueリストがある．これと，メンテナによってラベル付けされたgood first issueを比較
⇒ この研究のIssue推薦の方向性をはっきりさせ，有用性を示す

- プロジェクトごとのgood first issueの比較

RQ2では，good first issueが解決される割合，RQ3では，解決後に貢献してくれる開発者の割合が，プロジェクトごとに差があった
→ good first issueの内容を比較し，解決されやすい，解決後に貢献してくれるようなgood first issueの特徴を調べる