

OSS開発者のIssueラベル分析 によるオンボーディング支援

発表者：大森伊月

背景

- OSSではプロジェクトを存続していくために常に開発者を求めている[1]
- プロジェクトへの定着を促進するためのオンボーディングがメンテナによって実施されている
- オンボーディングはコミットやバグ報告を通じて行われ、それらには課題票(以下 Issue)が用いられている
- Issueにはその課題の特徴を示すためのラベルが付与されている

[1]Shaosong Ou Alexander Hars. 2002. Working for free? Motivations for participating in open-source projects. International journal of electronic commerce 6, 3 (2002), 25–39.

問題点

- ・ 貢献を行う際にどのIssueを選択すれば良いかわからない
- ・ オンボーディングはプロジェクトのメンテナにとって負担になっている

目的・アプローチ

目的

- ・ OSSプロジェクトのオンボーディングに有効なIssueの内容を理解する

アプローチ

- ・ 一般的な開発者が過去に携わったIssueのラベルを分析する

データセット

- プロジェクト
 - 新規開発者向けの「GoodFirstIssue」(以下 GFI)ラベルがIssueで運用されている
 - 整合性のあるデータが取得可能
- 開発者
 - プロジェクトにおける一般的な開発者を抽出
 - 一般的な開発者は平均1~1.5年で離脱^[2]
 - > 活動期間が1年以上
 - 継続的な貢献をしている開発者 -> 月に一回以上活動
 - 一部の熟練した開発者を除外するため -> 上位20%を除外

[2] C. Bird, A. Gourley, P. Devanbu, A. Swaminathan and G. Hsu, "Open Borders? Immigration in Open Source Projects," Fourth International Workshop on Mining Software Repositories (MSR'07:ICSE Workshops 2007), 2007, pp. 6-6, doi: 10.1109/MSR.2007.23.

パーセンタイル別Issueラベル分析

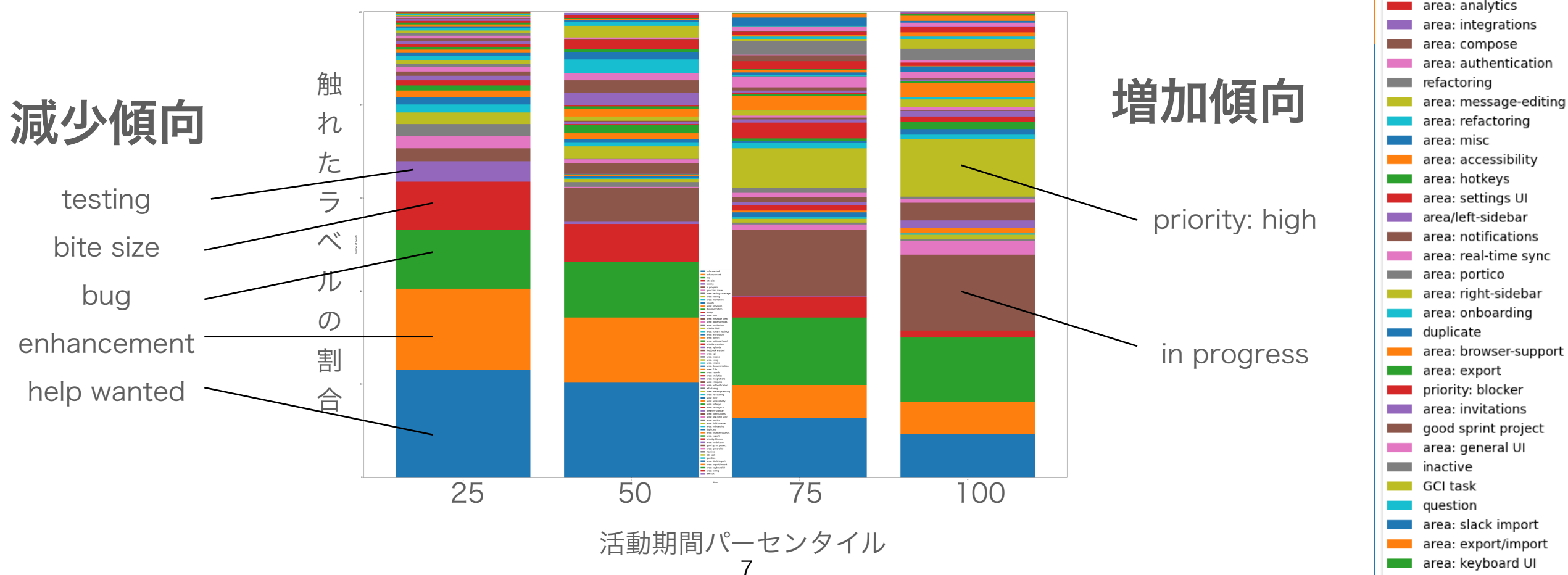
- 前述の条件を満たすプロジェクトの開発者がアクション(コメント, コミット, PRなど)をおこなったIssueを追跡
- 貢献開始(最初のIssueへのアクション)から貢献終了(最後のIssueへのアクション)までの期間を4分割し, それぞれの区間でそれぞれ開発者が携わったラベルを集計する

結果

zulip(チャットアプリ)を利用

- ・ 25パーセンタイル区間での上位5件は減少傾向
- ・ 「priority:high」や「in progress」が著しい増加傾向

開発者が活動を継続するにつれて「priority:high」や「in progress」ラベルが付与されているIssueに携わることが増え、「testing」「bite size」「bug」「enhancement」「help wanted」ラベルが付与されているIssueに携わることが減っている



考察

- 「bug」などからプロジェクトに定着していき、最終的には「priority:high」(優先度:高)や「in progress」(進行中)などのプロジェクトの根幹に関わるissueに携わっている
- 「documentation」は25パーセントから少なく、先行研究^[3]で示唆されていたドキュメントに関連するGFIはオンボーディングのための触媒として機能しないことと似た結果を得ることができた

[3] H. Horiguchi, I. Omori and M. Ohira, "Onboarding to Open Source Projects with Good First Issues: A Preliminary Analysis," 2021 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER), 2021, pp. 501-505, doi: 10.1109/SANER50967.2021.00054.

今後の予定

- ・ 得られた結果のより深い考察・解釈
- ・ 目的が達成できているかの確認
- ・ (達成できていない場合) どうすれば達成できるか