

コミッター候補者予測における OSSプロジェクトの開放性に関する考察

金城 清史^{†1} 大平 雅雄^{†1}

従来のコミッター候補者予測の研究では、OSSプロジェクトに参加する開発者の様々な活動量（例えば、パッチ投稿数など）を基にコミッターとして昇格させるべき候補者を推薦するための予測モデルを構築している。本研究では、開発者の活動量だけではなく、OSSプロジェクトの状態（例えば、成長期なのか成熟期なのか）や昇格方針（例えば、指名方式なのか投票方式なのか）などが、コミッター予測の精度を向上するために考慮すべき要因であると考えられる。本ワークショップでは、主にOSSプロジェクトの特性に起因し、コミッター予測の精度に影響を与え得る要因を、「コミッター昇格プロセスにおけるOSSプロジェクトの開放性」と定義し、OSSプロジェクトの開放性を表現するためのメトリクスについて議論したい。

On Openness in Open Source Software Projects for Predicting Committer Candidates

KIYOFUMI KINJO^{†1} and MASAO OHIRA^{†1}

Existing approach to predicting committer candidates in an open source software (OSS) project have built a prediction model using a variety of developers' activities such as the number of created patches and the number of comments. Our study focuses not only on developers' activities but also various attributes of each OSS project such as the maturity of the project and the promotion policy in the project. In this workshop, we would like to discuss what kinds of metrics we should select to improve the accuracy of the prediction model for committer candidates.

1. はじめに

大規模オープンソースソフトウェア（OSS）プロジェクトには、日々数百件の不具合報告があると言われて³⁾。それら大量の不具合を修正するために開発者により作成されプロジェクトに投稿される不具合修正パッチは、コミッターと呼ばれる特別な権限を与えられた開発者（以降、一般開発者との区別のためにコミッターと呼ぶ）によって、その妥当性が検証されプロダクトに反映される（コミットされる）。しかしながら、パッチの品質を保証するために検証作業を行うコミッターは、優れた開発能力を示した開発者の中から選出されるため、OSSプロジェクトの全開発者に占めるコミッターの数は圧倒的に少数であることが一般的である⁴⁾。そのため、近年の大規模OSSプロジェクトでは、少数のコミッターに過度な負担がかかるのと同時に、結果的に不具合修正の長期化を招くことが問題と

されている³⁾。

コミッターの負担を軽減させることで不具合修正プロセスの改善を図る方法として、多数の開発者の中からコミッターとなり得る有能な開発者を発見し、コミット権限を付与することが考えられる。特に、一般の開発者がOSSプロジェクトに長期間参加する事は稀である¹⁾ため、出来るだけ早い段階でコミッター候補者を特定することが重要となる。

先行研究⁶⁾では、「OSS開発者の特性（例えば、パッチ投稿数などの活動量）」に基づいてコミッターとして昇格させるべき候補者を推薦するための予測モデルを構築している。本研究では、開発者の活動量だけではなく、OSSプロジェクトの状態（例えば、成長期なのか成熟期なのか）や昇格方針²⁾（例えば、指名方式なのか投票方式なのか）などが、コミッター予測の精度向上を図る上で考慮すべき要因であると考えられる。本研究では、「OSSプロジェクトの特性」に起因し、コミッターの昇格（のしやすさ）に影響を与える要因を「OSSプロジェクトの開放性」として捉える。

^{†1} 和歌山大学

Wakayama University

表 1 開放性メトリクス
Table 1 Openness Metrics

メトリクス	説明	仮説
新規開発者の数	プロジェクトに新規に参加した（例えば 1 年未満の）開発者の数.	新規開発者の数が多いプロジェクトは、コミッターの昇格数も多い.
新規開発者数の発言率	ML などでの新規開発者の発言数が全発言数に占める割合.	新規開発者が活発に議論できるプロジェクトは、コミッターの昇格数も多い.
新規開発者への返信率	ML などでの新規開発者の発言に対するコアメンバー（コミッター）の返信率.	新規開発者を大事に扱う（育てる）プロジェクトは、コミッターの昇格数も多い.
上位開発者の流動性	ML や BTS での発言やパッチ投稿数の上位（例えば上位 10%）の開発者の流動率	活躍する開発者が固定的ではなく流動的なプロジェクトは、コミッターの昇格数も多い.
パッチレビュー率	新規/一般開発者が投稿したパッチに対するレビュー率.	新規/一般開発者とコアメンバのパッチレビュー率が同等なプロジェクトは、コミッターの昇格数も多い.
機能要求充足率	BTS などで要求された機能（Enhancement など）が実現された割合.	要求された機能を積極的に開発するプロジェクトは、コミッターの昇格数も多い.

2. OSS プロジェクトの開放性

プロジェクトの状態や昇格方針などの違いにより、コミッター昇格のしやすさに影響を与える要因、すなわち、「OSS プロジェクトの開放性」は、プロジェクト毎に異なると考えられる。本研究ではまず、「OSS プロジェクトの開放性」を複数のメトリクスで定義する。表 1 に開放性メトリクスの一例を挙げる。

例えば、新規/一般開発者の「パッチレビュー率」は、コアメンバ以外の貢献を積極的に受け入れているかどうかの目安になり得る。成熟したプロジェクトでは、新規機能の追加などの必要がなくなるため、新規/一般開発者のパッチレビュー率は低くなる⁵⁾ことが知られており、コミッターを新たに昇格させる必要性も低くなると考えられる。

今後はいくつかの OSS プロジェクトを対象に、これらのメトリクスとコミッター昇格率との相関を時間経過（例えば、1 年毎に計測するなど）を考慮しながら詳細に調査する予定である。

3. おわりに

本稿では、コミッター推薦の予測精度の向上を目的として、先行研究で利用されている開発者の属性⁶⁾だけではなく、プロジェクトの特性（OSS プロジェクトの成熟度や昇格方針）を考慮する必要性について述べた。また、OSS プロジェクトの開放性を調査するための複数のメトリクスを挙げ、コミッター昇格のしやすさとの関係について、仮説を提示した。本ワークショップでは、列挙した開放性メトリクスの妥当性およびその他の有望なメトリクスについて議論したい。

謝辞 本研究の一部は、文部科学省科学研究補助金（基盤 (B):23300009）および（基盤 (C):24500041）による助成を受けた。

参 考 文 献

- 1) Bird, C., Gourley, A., Devanbu, P., Swaminathan, A. and Hsu, G.: Open Borders? Immigration in Open Source Projects, *Proceedings of the Fourth International Workshop on Mining Software Repositories (MSR'07)*, p.6 (2007).
- 2) Jensen, C. and Scacchi, W.: Role Migration and Advancement Processes in OSSD Projects: A Comparative Case Study, *Proceedings of the 29th International Conference on Software Engineering (ICSE'07)*, pp.364–374 (2007).
- 3) Jeong, G., Kim, S. and Zimmermann, T.: Improving bug triage with bug tossing graphs, *Proceedings of the 7th joint meeting of the European Software Engineering Conference and the ACM SIGSOFT symposium on The Foundations of Software Engineering (ESEC/FSE '09)*, pp.111–120 (2009).
- 4) Mockus, A., Fielding, R. T. and Herbsleb, J. D.: Two Case Studies of Open Source Software Development: Apache and Mozilla, *ACM Transactions on Software Engineering and Methodology (TOSEM)*, Vol.11, No.3, pp. 309–346 (2002).
- 5) Yamamoto, M., Ohira, M., Kamei, Y., Matsumoto, S. and Matsumoto, K.: Temporal Changes of the Openness of an Oss Community: a Case Study of the Apache Http Server Community, *The Fifth International Conference on Collaboration Technologies 2009 (CollabTech 2009)*, pp.64–65 (2009).
- 6) 伊原彰紀, 亀井靖高, 大平雅雄, 松本健一, 鶴林尚靖: OSS プロジェクトにおける開発者の活動量を用いたコミッター候補者予測, 電子情報通信学会論文誌, Vol.J95-D, No.2, pp.237–249 (2012).