

Politeness 分析に基づく OSS 開発者の活動継続性の理解

宮崎 智己^{1,a)} 伊原 彰紀^{2,b)} 大平 雅雄^{2,c)} 東 裕之輔^{1,d)} 山谷 陽亮^{1,e)}

概要：本研究では、開発者の Politeness（コミュニケーション上の配慮）と活動継続性に関係があるかどうか分析を行った。Apache HTTP Server 及び Python プロジェクトを対象とするケーススタディを行った結果、開発者自身の Politeness と活動継続性には一定の関係があることを確認した。

1. はじめに

オープンソースソフトウェア (OSS) は、自由に利用、拡張、再配布できる特徴から、多くのボランティア開発者が、新たな機能拡張を提案し、自ら開発に参加している。一方で、多数の開発者が 1 年未満で OSS プロジェクトを離脱しているため、長期的に参加する開発者の不足が OSS の継続的な開発保守を困難にしている。

従来研究では、開発者が継続的に活動を行う要因を明らかにする研究がされてきている。Zhou らは、長期的にプロジェクトに参加する開発者 (Long Term Contributor) を参加初期に予測するモデルの開発を行い、開発者のプロジェクト参加直後の活動意欲 (コメントや不具合報告など)、および、他の開発者からのサポートが継続的な活動に重要であることを示した [1]。

様々な地域に点在する OSS 開発者は、主にメールを用いてコミュニケーションを行なっている。メールのような非対面コミュニケーションでは、意思疎通に誤解が生じやすいため、配慮のあるコミュニケーションを行っていることが考えられる。しかしながら、継続的に活動する開発者を対象にコミュニケーション上の配慮について分析した研究はない。

本研究では、開発者のコミュニケーション上の配慮を分析することで、開発者が継続的に活動することを可能にしている要因を明らかにする。本研究では、開発者のコミュ

ニケーション上の配慮を Politeness[2] を用いて定義する。Apache HTTP Server プロジェクトと Python プロジェクトにおける開発者メーリングリストのコミュニケーションデータを用いたケーススタディを行った結果、開発者の自身の Politeness と活動継続性に一定の関係があることがわかった。

2. Politeness

Politeness とは、日本語の「礼儀正しさ」とはニュアンスが異なる概念であり、円滑な人間関係を確立・維持するための相手への配慮を指す。Politeness を適切に表現するためには、他者に認められたい・評価されたいという欲求 (ポジティブフェイス) と、他者から批判されたくない、行動を妨げられたくない欲求 (ネガティブフェイス) のどちらも満たすよう、言葉遣いを工夫する必要がある。例えば、“このエラーの原因は何?” のような質問は、聞き手に心理的負荷をかけてしまい聞き手のネガティブフェイスを侵害してしまう。一方で、“このエラーの原因を教えてくださいませんか?” のような質問は、回答することでポジティブな効果 (感謝されるなど) があり、たとえ回答できなくても問題ないことを暗黙的に示しているため聞き手への心理的負荷が低くなる。

本研究では、Stanford Politeness[3]*¹を用いて、開発者のコミュニケーションデータを対象にコミュニケーション上の配慮を数値化する。Stanford Politeness とは、言語表現に込められた Politeness を定量化するツールである。

¹ 和歌山大学大学院システム工学研究科
Graduate School of Systems Engineering, Sakaedani 930,
Wakayama-city 640-8510, JAPAN

² 和歌山大学システム工学部
Faculty of Systems Engineering, Sakaedani 930, Wakayama-city 640-8510, JAPAN

a) miyazaki.tomoki@g.wakayama-u.jp

b) ihara@sys.wakayama-u.ac.jp

c) masao@sys.wakayama-u.ac.jp

d) higashi.yunosuke@g.wakayama-u.jp

e) yamatani.yosuke@g.wakayama-u.jp



図 1: Stanford Politeness における Politeness の算出方法

*¹ <https://github.com/sudhof/politeness>

表 1: データセット

対象プロジェクト	対象開発者数		対象メール件数	
	長期開発者 (活動期間 90 日以上)	短期開発者 (活動期間 90 日未満)	送信メール	受信メール
Apache	183	173	109,685	82,190
Python	227	183	106,993	93,874

表 2: Politeness 値 (送信メール) の統計量と検定結果

プロジェクト	長期開発者			短期開発者			p 値
	中央値	最大値	最小値	中央値	最大値	最小値	
Apache	0.62	0.87	0.34	0.66	0.96	0.31	0.00
Python	0.64	0.83	0.37	0.68	0.91	0.42	0.00

Stanford Politeness は、図 1 に示す通り、2 つの処理から構成される。

Politeness Strategy の検出。文章中の言語表現に込められた Politeness を定量化するために、相手を配慮するための具体的な表現を分類した Politeness Strategy[2] を用いる。Politeness Strategy は、相手のポジティブフェイスとネガティブフェイスを満たす語彙を分類しており、Politeness の値が高くなる語彙群には、ポジティブフェイスを満たす語彙 (Gratitude, Deference など) とネガティブフェイスを満たす語彙 (Please, Hedges など) が含まれる。また、Politeness の値が低くなる語彙群には、ネガティブフェイスを侵害する語彙 (Please start, Direct question など) が含まれる。Politeness Strategy では、このような Politeness に影響する語彙群を 18 種類に分類している。本論文では、入力テキストを Stanford Parser^{*2}により構文解析し、Politeness Strategy に登録される語彙を検出する。

SVM による Politeness 値の予測。Stanford Politeness は、機械学習アルゴリズムの 1 つであるサポートベクタマシン (SVM) を用いて、入力テキストの Politeness 値を算出する。SVM 構築には、Stanford Politeness Corpus[3] を学習データとして用いている。

Stanford Politeness Corpus を学習データとし、検出した語彙を用いて入力テキストの Politeness 値を 0~1 の値 (値が大きいほど相手を配慮する表現が取り入れられている) で予測する。

3. 分析結果

本研究では、開発者の Politeness と活動継続性に関係があるかどうかを明らかにするために、活動期間が 270 日以上の開発者を長期開発者、活動期間 270 日未満の開発者として、長期開発者と短期開発者の送信メールの Politeness 値と受信メールの Politeness 値をそれぞれ比較する。長期開発者と短期開発者の Politeness 値に統計的有意差があるか否かを確認するために、マンホイットニーの U 検定を用いる。対象プロジェクトは大規模かつ長期間開発・保守されている OSS プロジェクトである、Apache HTTP Server

プロジェクトと Python プロジェクトである。分析に用いたデータセットを表 1 に示す。本論文では、ページの都合上、開発者の送信メールの Politeness 値と活動継続性に関する分析の結果について述べる。

表 2 は長期開発者の Politeness と短期開発者の Politeness の統計量 (中央値, 最大値, 最小値) とマンホイットニーの U 検定の結果を示す。Apache HTTP Server プロジェクト、及び、Python プロジェクトにおいて、開発者が送信したメールの Politeness 値の中央値は、短期開発者の方が長期開発者に比べて高く、長期開発者と短期開発者の間に統計的有意差を確認した。この結果から、長期開発者と短期開発者が送信したメールの Politeness に違いがあり、開発者自身の Politeness と開発者の活動継続性に関係があることがわかった。

4. おわりに

本研究では、OSS 開発プロジェクトで交わされるコミュニケーションに着目し、開発者が継続的に活動を行うことを可能にしている要因を明らかにするために、開発者の送受信メールを対象に良好な人間関係を確立・維持するための配慮を表す Politeness を分析した。Apache HTTP Server と Python を対象に分析を行った結果、長期開発者の方が短期開発者よりも送信メールの Politeness 値が有意に低いことがわかった。しかしながら、開発者の活動継続性を Politeness 値のみで説明できるほどの大きな差ではないため、開発者の自身の Politeness は活動継続性に影響を与える要因の一つとして捉えるのが妥当である。

本研究では、開発者間のコミュニケーションに着目したが、今後、開発者間の人間関係、及び、パッチ投稿、コードレビューなどの技術的貢献を分析することで、Politeness 値が変動する要因を明らかにし、短期開発者が継続的に貢献するための方法論の確立を目指す。

謝辞 本研究の一部は、JSPS 科研費 JP17H00731, JP18K11243 による助成を受けた。

参考文献

- [1] Zhou, M. and Mockus, A.: What Make Long Term Contributors: Willingness and Opportunity in OSS Community, *Proc. of the 34th Int'l Conf. on Soft. Eng. (ICSE '12)*, pp. 518–528 (2012).
- [2] Brown, P. and Levinson, S. C.: *Politeness: some universals in language usage*, Cambridge University Press (1987).
- [3] Danescu-Niculescu-Mizil, C., Sudhof, M., Jurafsky, D., Leskovec, J. and Potts, C.: A computational approach to politeness with application to social factors, *Proc. of 51st Annual Meeting of the Association for Computational Linguistics (ACL '13)* (2013).

^{*2} <http://nlp.stanford.edu/software/lex-parser.shtml>