

OSS の障害修正における開発者ネットワークの分析

伊原彰紀¹ 大平雅雄¹ 小山貴和子¹ 松本健一¹

1. はじめに

近年、オープンソースソフトウェア (OSS) における障害修正に関する研究が盛んに行われている[1]。一般的な OSS プロジェクトは、地理的に分散した環境で開発を行っている。障害に関する情報は、障害管理システムを用いて共有・管理される。しかし、障害管理システムを利用しているにも関わらず、多くの OSS プロジェクトでは障害の修正作業が円滑に行われないことが多くある[2]。分散環境下で開発を行う OSS プロジェクトでは、開発者間の情報共有が特に重要であるが[3]、障害修正のための情報共有を円滑に行うための方法は明らかではない。

本研究は、効率的に障害修正を行うために開発者らに必要となる活動を明らかにすることを目的とする。本稿では、障害修正が円滑に行われている時期と停滞している時期の開発者らの議論の様子を、社会ネットワーク分析指標を用いて比較する。

2. 準備

本章では、本稿で用いる用語の定義と説明を行う。

2.1 障害管理システムにおける議論

障害管理システムは、ソフトウェア開発中に発生した障害を登録し、修正状況などを一元管理するシステムである。障害に関する情報共有を支援するために、各障害に対して掲示板(BBS)を設ける機能を利用することができる。障害管理システムの BBS では、修正に関するアドバイスや修正方法の相談が開発者らの間で行われる。

2.2 開発者間の議論ネットワーク

開発者は、複数の障害を修正するためにそれぞれの BBS において議論を行っているが、議論の様子を俯瞰的に捉えるのは困難である。本稿では、障害に関して開発者が行う議論の様子をネットワーク化し、その特徴を分析する。具体的には、図 1 に示すように、1 つのノードを掲示板に書き込んだ 1 人の開発者とする。開発者が掲示板にコメントを書き込む際、それまでに書き込んだ開発者全てに情報を発信したと考える。コメントを投稿した開発者から、これまでに投稿した開発者に向けてエッジを張り、開発者間で行う議論の様子を表現する。例えば、図 1 において、2 回目に掲示板に投稿した C は B と A にそれぞれ 1 本のエッジを張る。

3. 分析方法

本稿では、障害修正が円滑に行われている時期と停滞している時期の開発者らの議論の様子を、社会ネットワーク分析指標を用いて比較分析し、効率的に障害を修正するための方法を明らかにする。

3.1 障害修正が円滑に行われている時期の特定

本稿では、開発者らが継続的に多くの障害を修正している時期を、障害修正が円滑に行われている時期と考え、障害管理システムに登録された障害のうち、修正された障害の割合(障害修正率)が継続して高かった時期を特定する。ただし、多くの開発者が存在するプロジェクトでは、修正される障害の数も増加するため、開発者一人当たりの障害修正率を求め、開発者一人当たりの障害修正率が数か月連続で増加もしくは減少している時期で、かつ変動幅がそれぞれ最大の時期と 2 番目に大きい時期を選択する。

3.2 社会ネットワーク分析

開発者一人当たりの障害修正率の高い時期と低い時期それぞれについて、2.2 節で説明した開発者間のネットワークを構築する。次に、「べき指数」「クラスタリング係数」「平均経路長」を用いて開発者の議論ネットワークを分析し、開発者一人当たりの障害修正率の高い時期と低い時期の議論ネットワークにどのような違いがあるかを確認する。

4. ケーススタディ

Firefox プロジェクトを対象にケーススタディを行った。2001 年 1 月 1 日から 2008 年 12 月 10 日の間に障害管理システムに投稿された障害の内、修正が完了した 2376 件の障害を対象とした。

4.1 分析結果

障害が報告されてから修正完了するまでの時間(滞留時間)の結果を報告する。本稿で対象とした障害の内、60%は滞留時間が約 1 か月以内、80%は滞留時間が約 3 か月以内、90%は滞留時間が約 7 か月以内であった。開発者一人当たりの障害修正率の推移が 7 か月連続で増加(減少)することはないため、開発者一人当たりの障害修正率の推移が 3 か月間連続して増加(減少)する時期を分析対象とした。

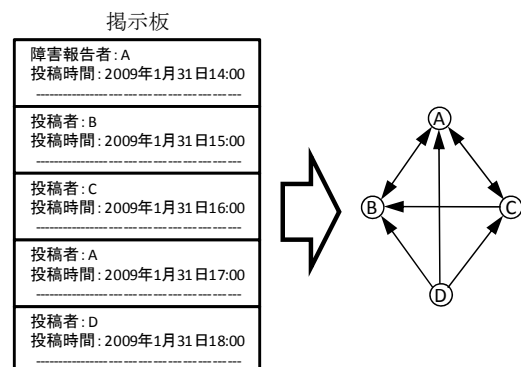


図 1 開発者間の議論ネットワーク

¹奈良先端科学技術大学院大学 情報科学研究科
¹Graduate School of Information Science,
Nara Institute of Science and Technology

円滑な障害修正活動を行う時期の特定

開発者一人当たりの障害修正率の推移を図2に示す。横軸は期間、縦軸は障害修正率を表す。図2より、3か月続けて開発者一人当たりの障害修正率が増加し、かつ変動幅が最大の時期は2003年8月(03.08)と2008年10月(08.10)であった。一方、3か月続けて減少し、かつ変動幅が最大の時期は2002年12月(02.12)と2008年6月(08.06)であった。

社会ネットワーク分析

障害修正率が増加傾向もしくは減少傾向にある時期について、社会ネットワーク分析した結果を表1に示す。また、障害修正率が増加傾向にある2003年8月の議論ネットワークを図3に、障害修正率が減少傾向にある2008年6月の議論ネットワークを図4に示す。

表1より、障害修正率が増加傾向にある時期と減少傾向にある時期とでは、べき指数、平均経路長に違いは見られないが、クラスタリング係数に関しては障害修正率が減少傾向の期間に比べ、増加傾向の期間の方が高いことが分かる。つまり、多くの障害修正が行われている時期は、開発者間の議論が盛んに行われているということである。また、障害修正率が増加傾向の期間(図3)は減少傾向の期間(図4)に比べ、集中的に議論が行われていることが分かった。

4.2 考察

開発者一人当たりの障害修正率の高い時期はクラスタ係数が高く、議論ネットワークからも議論を集中的に行っている開発者の存在を確認することができた。このことから、障害管理において中心となって議論(情報共有)を行っている開発者が存在すると、円滑な障害修正活動が行われると考えられる。

また、障害の中には、別の障害の修正が完了しなければ修正が開始できないという依存関係のある障害が多く存在する。依存関係のある障害は、双方の障害について情報共有が必要で、橋渡しの役割を担う開発者が必要となる。橋渡し役となっている開発者が多いと、図3のようなネットワーク構造となり、円滑な障害修正に繋がると考えられる。

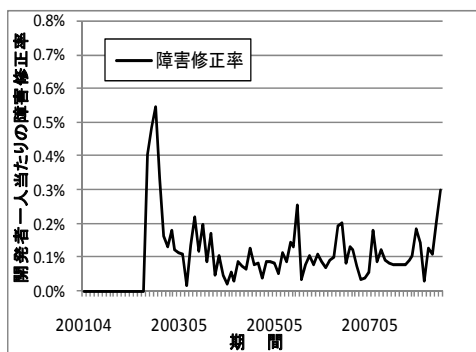


図2 障害修正率の推移

表1 開発者間の議論ネットワーク

	期間	ノード数	べき指数	クラスタリング係数	平均経路長
障害修正率増加傾向	03.08	439	0.85	0.51	2.55
障害修正率減少傾向	08.10	364	0.88	0.50	2.55
障害修正率増加傾向	02.12	151	0.94	0.38	2.82
障害修正率減少傾向	08.06	595	0.91	0.35	2.98

5. おわりに

本稿では、効率的な障害修正を行うために必要な要件を明らかにすることを目的として、Firefoxの障害修正活動を分析した。その結果、議論が集中的に行われている時期ほど多くの障害修正が行われていることが分かった。

本研究の今後の課題は、障害修正率をより妥当な指標とするために、それぞれの障害の内容(どういったモジュールに存在する障害なのか)を加味することや、優先度や重要度によって障害に重みをつけるなどの改良を加えた指標を構築することである。

謝辞

本研究の一部は、文部科学省「次世代IT基盤構築のための研究開発」の委託に基づいて行われた。また、本研究の一部は、文部科学省科学研究補助費(若手B:課題番号20700028)および公益信託マイクロソフト知的財産研究助成基金による助成を受けた。

参考文献

- [1] 伊原彰紀, 亀井靖高, 大平雅雄, 松本真佑, 松本健一, OSSプロジェクトにおける生涯に関する情報共有の分析, 情報処理学会関西支部 支部大会 講演論文集, pp.69-72, 2008.
- [2] Wang, Y., Guo, D., and Shi, H.: Measuring the evolution of open source software systems with their communities, SEGSOFT Software Engineering Notes, Vol.32, No.6, 2007.
- [3] Bird, C., Gourley, A., Devanbu, P., Gertz, M. and Swaminathan, A.: Mining Email Social Networks, In Proceedings of the 2006 International Workshop on Mining Software Repositories(MSR'06), pp.137-143, 2006.

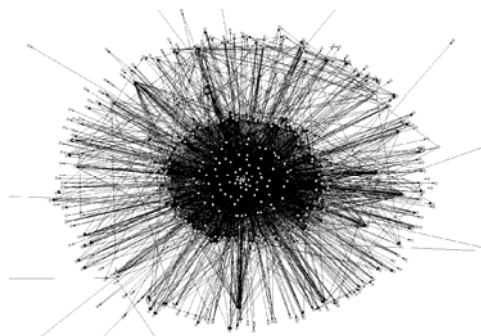


図3 2003年8月の議論ネットワーク

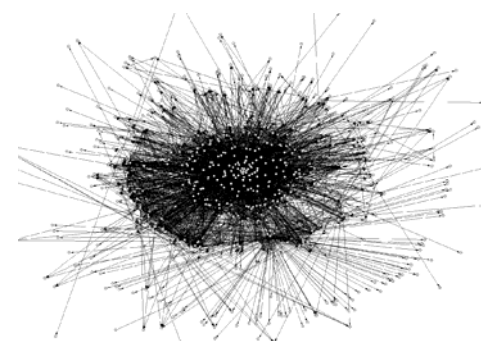


図4 2008年6月の議論ネットワーク