

An Analysis of Fault Information Sharing in an OSS Project

伊原 彰紀† 亀井 靖高† 大平 雅雄† 松本 真佑† 松本 健一†

Akinori Ihara Kamei Yasutaka Masao Ohira Shinsuke Matsumoto Ken-ichi Matsumoto

1. はじめに

近年、中国やインドにおけるオフショア開発を代表として、ソフトウェアの分散開発が盛んに行われている。オープンソースソフトウェア (OSS) 開発もその一つである。分散開発は文化的背景の違いから開発者同士の意思疎通が困難と言われているが、OSS は商用のソフトウェアに劣らない品質と機能を備え、社会的に広く普及しつつある。そのため、分散開発を成功させるための要因や知見を得るために、OSS プロジェクトにおける情報共有の方法についての研究が現在盛んに行われている[1][3]。

本稿では、信頼性の高い OSS を実現する上で、障害に関する情報共有を円滑に行うためには、どのような人物の活動が重要なのか知るところとする。障害の発見に貢献しているバグ報告者の中で、その障害を修正している開発者との関わりをもつ人物に着目し、OSS プロジェクトにおけるバグ報告者と開発者がどのように障害情報を共有しているのか、その様子を分析する。

本稿の構成は以下の通りである。続く 2 章では関連研究を述べ本研究の立場を明らかにする。3 章では、本稿で扱う OSS プロジェクトにおける参加者から形成するネットワークについて説明する。そして、4 章においてケーススタディで行う分析方法について説明し、5 章のケーススタディで分析の対象とする PostgreSQL について述べる。6 章でケーススタディの結果と考察を行い、最後に 7 章においてまとめと今後の課題を示す。

2. 関連研究

OSS プロジェクトの開発形態を分析した事例が、数多く報告されている[2][3]。

Mockus らは、これまで OSS コミュニティでの開発について指摘されていた事柄の正しさについて、Apache コミュニティの CVS データ、バグ報告データから調査した[2]。その結果、Apache プロジェクトの 4% の開発者が 88% のソースコードを追加し、66% の不具合修正を行っていることが明らかとなった。Mockus らの研究は CVS データなどの開発データを利用して、どういった開発者が実装に寄与しているかを分析している。

一方で、Raymond は自身が運営する OSS プロジェク

トの分析を通じて、OSS 開発の成功には開発者だけでなくユーザの存在が重要であると指摘している[1]。ユーザは、OSS を利用している際に発見した問題を開発者に報告することがある。Raymond の指摘は、ユーザから開発者へのこのようなフィードバックが OSS の品質向上に繋がることを示唆するものであり、OSS 開発における開発者とユーザの協調作業の重要性を説くものである。

我々の先行研究では、Raymond の経験的な指摘を明らかにするために開発者とユーザの協調作業の分析を行ってきた[3]。OSS プロジェクトにおける開発者とユーザのコミュニティそれぞれが利用するメーリングリストを対象に、参加者間のメッセージの送受信関係をネットワークモデル化し、開発者とユーザの協調作業を分析した。その結果、両コミュニティに参加する人物が、コミュニケーションネットワークの中心に位置する傾向が強いことがわかった。

これまでの研究ではコミュニティ間の協調作業の様子を確認してきたが、信頼性の高い OSS を実現するための情報共有についての分析は行われていない。そこで我々は障害に関する議論を行っているバグ報告者コミュニティ参加者の中で、多くの障害を修正している開発者コミュニティとの情報共有をどのように行っているかを分析する。

3. OSS プロジェクトにおける参加者のネットワーク

3.1 コミュニティの参加者

OSS プロジェクトには開発者、ユーザといった人々のコミュニティが複数存在する。開発者コミュニティでは開発に関する議論が行われ、ユーザコミュニティでは利用する際の質問などの議論が行われている。本稿では、そのコミュニティの議論に参加した人物、つまりメッセージを投稿した人物をコミュニティの参加者とした。

3.2 障害に関する情報共有に寄与する人物

開発中の OSS に対する障害への対応を行うのは開発者であることが多いため[2]、その障害を報告するバグ報告者と開発者との情報共有は必要不可欠であると考えられる。

本稿では、バグ報告者コミュニティの中でも開発者コミュニティにも参加する人物が OSS の障害に関する情報共有を行う重要な人物であり、コミュニティ間で情報の橋渡し役を行っていると考え、開発者、バグ報

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

告者コミュニティの両方に参加する人物に着目して分析を行う。

3.3 コミュニケーションのネットワーク化

OSS プロジェクトは参加人数やメッセージ数の規模が大きく、参加者間のコミュニケーションの様子を俯瞰的に捉えることが困難である。そこで本稿では、参加者間でやり取りされるメッセージの送受信関係に基づいて、参加者間のコミュニケーションをネットワーク化し、その特徴を分析する。コミュニケーションから形成したネットワークを用いて分析対象を単純化することで、参加者間やコミュニティ間のコミュニケーションについて巨視的な分析が可能となる。

具体的には図1のように1つのノードをOSSコミュニティに参加している1人の人物とし、1本のエッジを2人の参加者間でやり取りされるメッセージの送受信関係として表す。Aのメッセージに対し、Bが返信している場合、BからAにエッジが引かれる。

特に、本稿では障害に関する情報共有に寄与するであろう人物(バグ報告者コミュニティに参加し、かつ、開発者コミュニティに参加する人物)に着目して分析を行う。バグ報告者コミュニティと開発者コミュニティに参加する人物の集合を1つのグループ、バグ報告者コミュニティのみ参加する人物の集合を1つのグループとみなして、障害を報告するグループとそれを開発者に伝えるグループでの情報のやり取りを分析する。

OSS プロジェクトにおける参加者のコミュニケーションをグループ単位で巨視的に捉えるために、個人間のネットワークとは別に一つのグループを一つのノードで表し、異なるグループ間でやり取りされるメッセージの送信エッジ、受信エッジをそれぞれまとめて1本のエッジで表すことでグループ間のネットワークを形成して分析を行う。

4. 分析方法

OSS プロジェクトにおける障害に関する情報共有を理解するためには、開発者コミュニティ、バグ報告者コミュニティ間の情報伝達の様子を詳細に分析する必要がある。そこで、グループ間のネットワークを用いOSS プロジェクトにおけるバグ報告者コミュニティの参加者で、開発者コミュニティにも参加しているグループと参加していないグループのグループ内/間でどのようなコミュニケーション関係を持っているのか分析する。次に、障害に関する議論でより多く発言している参加者の活動傾向を分析する。最後にバグ報告者コミュニティの参加者で多くの発言をしている人物ら

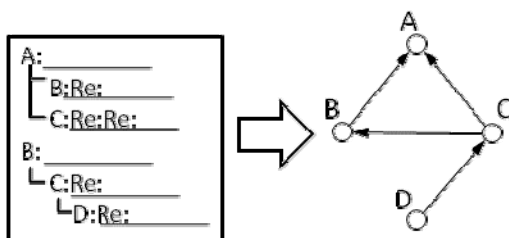


図1 個人間のコミュニケーションネットワーク

は継続して活動を行っているのかを分析する。

4.1 グループ間における情報共有

バグ報告者コミュニティの中でも開発者コミュニティに参加しているグループと参加していないグループの間でメッセージの送受信関係がどのように異なるのかを調べるために、ある一時期のグループ内/間のネットワークを分析する。分析方法として、グループ内/間のエッジ数を調べる。分析を通して、バグ報告者と開発者の仲介をする人物らが障害に関する情報共有にどれだけ寄与しているのかを知ることができる。

4.2 障害に関する情報共有を行う重要な人物

障害に関する議論を主として行うバグ報告者コミュニティで、多くの議論に参加し、発言を行っている人物は開発者コミュニティにも参加している人物なのかを分析するために、ある一時期の参加者の次数を比較する。参加者の次数を比較することで、バグ報告者コミュニティと開発者コミュニティの両方に参加する人物が障害に関する議論についてどれだけ発言しているのかを知ることができる。

4.3 重要人物の活動の継続性

前節までは、ある一時期に限って、次数の高い人物が開発者コミュニティにも関わっているかを分析していたが、長期間において同じ結果が得られるとは限らない。そこで、長期間において、各月で次数の高い人物(どのグループに所属するかも含めて)の推移を知ること、OSS プロジェクトの障害に関する情報共有は同一人物によって支えられているのか否かを知ることができる。

5. ケーススタディ

本章では、PostgreSQL プロジェクトにおける開発者、バグ報告者の2者間で障害に関する情報共有の実態を2つのコミュニティに参加する人物の視点から分析したケーススタディについて述べる。

5.1 PostgreSQL

PostgreSQLはBSDライセンスに基づき配布されているフリーの関係データベース管理システム(RDBMS)である。商用/非商用を問わず無償で利用ことができ、企業システムへの採用も増えていることから、商用のRDBMSに劣らないソフトウェアとして広く認知されている。また、データベース管理システムであるという点から、特に高い信頼性が求められるソフトウェアであると言える。

PostgreSQLは、1986年にPOSTGRESプロジェクトとして発足し、開発されてきたものが、1995年にオープンソース版(Postgres95)としてWeb上でリリースされた後、1996年終わりに名称をPostgreSQLに変更して1997年1月にver.6.0としてリリースされ現在まで開発が続いている。2004年8月にver.8.0のベータ版となる

ver.8.0β がリリースされ、障害の修正が加えられた後、翌年の 2005 年 1 月に正式に ver.8.0 がリリースされた。

5.2 分析対象データ

分析対象データは各コミュニティの議論の場として主に用いられるメーリングリストとする。本稿では、各メーリングリストの利用者の役割に応じてコミュニティを定義した。以下では各コミュニティの参加者が利用しているメーリングリストで主に扱われる議論内容について述べる。

- 開発者コミュニティ

pgsql-hackers : PostgreSQL の開発に関する話題を取り扱うメーリングリストであり、開発中の障害や新規機能の追加についての議論が主として行われる。

pgsql-patches : PostgreSQL の開発に関する話題を取り扱うメーリングリストであり、リリース後の障害の修正や新規機能の追加についての議論が主として行われる。

- バグ報告者コミュニティ

pgsql-bug : PostgreSQL の障害に関する話題を取り扱うメーリングリストであり、障害の修正についての議論などが行われる。また、報告された障害の修正が行われた場合は pgsql-patches へメッセージを送ることになっている。

本稿では、障害に関する情報共有が必要と思われる、ベータ版 ver.8.0β をリリースしてから正式版 ver.8.0 をリリースするまでの期間に着目した。正式版がリリースされるまでは、OSS のベータ版をリリースし、開発者がバグ報告者の意見を取り入れる必要がある。これらの作業を達成するためには、開発者コミュニティとバグ報告者コミュニティの情報共有が必要であると考える。この期間のコミュニケーション関係を分析することで、OSS プロジェクトにおける障害に関する情報共有が円滑にできているかを明らかにすることが期待できる。

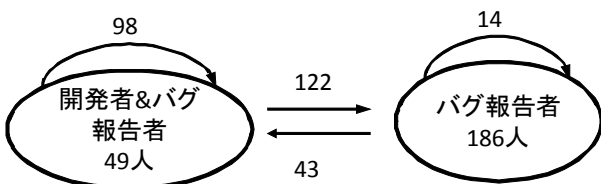


図2 バグ報告者コミュニティのグループ間のネットワーク (2004年9月)

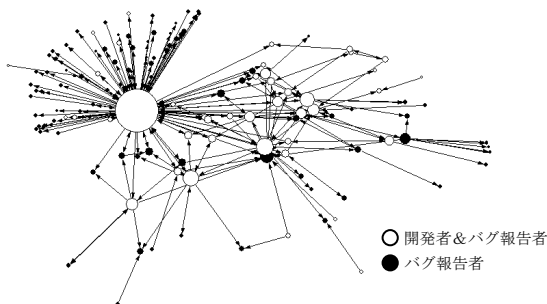


図3 バグ報告者コミュニティの個人間ネットワーク (2004年9月)

コミュニケーションは時間の経過とともに変化するため、ある一定期間ごとにコミュニケーションの様子を見る必要がある。そこで本稿ではコミュニケーションのネットワークを求める期間を一定期間ずつスライドさせる Sliding Time Method を用いる[4]。本稿では3か月間のコミュニケーションデータを一つのネットワークで表現し、1か月ごとスライドさせてデータを算出する。例えば、2004年9月の分析データは2004年9月から11月のデータとなる。

6. 結果と考察

6.1 グループ間における情報共有

2つのグループ内/間で行われるコミュニケーションの送受信エッジ数について図2に示す。各グループの数字はグループの人数を表す。また各エッジの添え字は送信エッジ数と受信エッジ数を表す。

図2よりグループ間の送受信数には約3倍の差があることがわかった。また、グループ内の個人間で行われるメッセージの送受信数は、バグ報告者コミュニティのみに参加するグループでは参加者数に対して少なく、開発者コミュニティにも参加するグループでは参加者数に対して多いことがわかった。

これまでの研究ではコミュニティ間の協調作業を分析していたため、複数のコミュニティに参加している人物のグループ内で行われる議論はどのコミュニティで話された議論かを区別していなかった。本稿ではバグ報告者コミュニティのみを対象とし、その中で開発者コミュニティにも参加する人物を分類することで、複数のコミュニティに参加する人物間のコミュニケーションがどのコミュニティで行われた会話なのかを明確にした。この分析結果から障害の報告に対し、開発者コミュニティにも参加する人物が障害に関する議論の中心となって PostgreSQL の障害に関する情報共有を行っていると考えられる。

6.2 障害に関する情報共有を行う重要な人物

PostgreSQL プロジェクトのバグ報告者コミュニティの参加者が行う個人間のコミュニケーションをネット

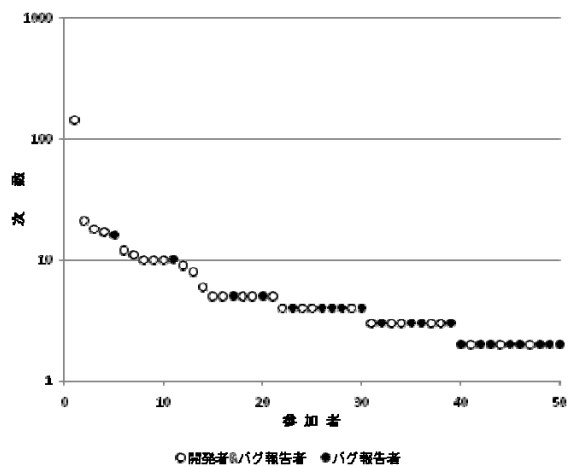


図4 バグ報告者コミュニティにおけるグループ別の参加者の度数 (2004年9月)

ワーク化したものを図 3 に示す。各ノードはバグ報告者コミュニティの参加者を示し、白いノードは開発者コミュニティにも参加する人物、黒いノードはバグ報告者コミュニティのみに参加する人物を表す。各ノードの大きさは次数の大きさを反映している。また、参加者間のコミュニケーションを表すエッジはメッセージの返信者から送信者に引いて表現している。

図 3 から開発者コミュニティにも参加する人物がバグ報告者コミュニティの中で開発者とバグ報告者の情報共有に大きく寄与していることがわかる。

さらに、図 3 で示したバグ報告者コミュニティに参加する人物のうち次数が高い上位 50 人を図 4 に示す。横軸はバグ報告者コミュニティにおいて次数の高い参加者を降順に並べ、各参加者に対する次数を片対数グラフにプロットした。

図 4 より次数の高い 50 人のうち 36 人は開発者コミュニティにも参加しており、その中でも開発者コミュニティにも参加している人物は比較的次数が高いことがわかる。よって、障害に関する議論は開発者コミュニティにも参加する人物が多く発言する傾向にあり、障害に関する情報共有に開発者コミュニティにも参加している人物が大きく影響していると考えられる。

6.3 重要人物の活動の継続性

前節までは、ベータ版をリリースしてから、正式版をリリースする間 (3 か月間) に着目して分析、考察してきた。ただ、開発者コミュニティにも参加している人物が常に障害に関する議論を行っているかは定かではない。そこで、ベータ版、正式版のリリース時期を含む長期間について次数が高い人物の推移を表 1 に示す。各行(R1~R10)は各月の次数の高い 10 名を昇順に示しており、各列は 6 ケタの数字で年、月を示す。表中の p1~p10 は前節まで分析してきた 2004 年 9 月に次数が高かった 10 名を示し、空白のセルはその他の参加者であることを示す。また、灰色のセルは開発者コミュニティにも参加している人物を表し、白色のセルはバグ報告者コミュニティのみに参加している人物を示す。

表 1 より、分析期間中に次数が上位 10 位以内に含まれた人物は 110 人存在するが、そのうち 90 人(約 82%)は開発者コミュニティにも参加している人物であった。また、正式版のリリースが近くなるほど開発

表 1 各月における次数の高い 10 人の推移

	200402	200403	200404	200405	200406	200407	200408	200409	200410	200411	200412
R1	p1	p1	p1	p1	p1	p1	p1	p1	p1	p1	p1
R2	p5	p5	p2	p6		p2	p2	p2	p2	p3	p2
R3	p2	p2	p5		p2			p3	p3	p2	p3
R4		p6		p2	p6			p4	p4		
R5	p6			p5	p5	p8	p8	p5	p5	p10	p10
R6			p6		p10	p4	p4	p6	p6	p9	
R7		p4	p9	p9		p10	p10	p7	p9	p5	p9
R8	p4					p5	p5	p8		p6	p4
R9				p10	p8			p9	p10		
R10					p9	p3	p3	p10	p7	p4	

者コミュニティにも参加する人物の発言数が多くなっていることがわかる。

次に 2004 年 9 月に次数が高かった 10 人 (p1-p10) の変動についての結果を分析する。各月で次数が最も高い人物 p1 は変動せず、常に開発者コミュニティにも積極的に参加していることがわかる。また、p2 についても多少の変動があるが、どの期間でも次数の上位に位置し、p1 と同様に開発者コミュニティにも常に参加していることがわかる。

ベータ版がリリースされたことによって、障害に関する報告が積極的に行われ、障害の修正を行う開発者に障害情報を伝える開発者コミュニティにも参加する人物の発言が多くなったと考えられる。さらに p1, p2 のように常に次数が高く開発者コミュニティにも参加する人物が障害に関する情報共有において重要であると考えられる。

7. おわりに

本稿では信頼性の高い OSS を実現する上で、障害に関する情報共有を円滑に行うためには、どのような人物の活動が重要であるのかを知ることを目的として、PostgreSQL プロジェクトのメーリングリストを分析した。その結果、開発者コミュニティにも参加している人物が中心となって、バグ報告者コミュニティのみに参加する人物と議論を行っていることがわかった。特に議論の中心となっている人物は、常にコミュニティの中心であり続けた。

障害の修正数と、修正者が参加するコミュニティがどのように変化するのかを分析し、障害への対応状況を正確に把握することが今後の課題である。

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) による助成を受けた。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] Raymond, E. S.: The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary, O'Reilly and Associates (1999).
- [2] 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, Vol. 11, No. 3, pp. 309-346 (2002).
- [3] Matsumoto, S., Kamei, Y., Ohira, M. and Matsumoto, K.: A Comparison Study on the coordination between Developers and Users in Foss Communities, Socio-Technical Congruence (STC 2008), No.8, pp.1-9 (2008).
- [4] Kakimoto, T., Kamei, Y., Ohira, M. and Matsumoto, K.: Social Network Analysis on Communications for Knowledge Collaboration in OSS Communities, In Proc. the International Workshop on Supporting Knowledge Collaboration in Software Development (KCS'D'06), pp. 35-41 (2006).