オープンソースソフトウェア開発における役割分担の実態調査

プロジェクトマネジメントコース　矢吹研究室　1042067　関口　元基

1. 研究背景

オープンソースのソフトウェア開発のためには，バージョン管理システムやWiki，バグ追跡システム，メーリングシステムがよく利用される．これらのツールをまとめて提供するのがホスティングサイトであり，その代表的なものがGitHubである．現在GitHubでは多くのプロジェクトが行われている． オープンソース・ソフトウェア（以下，OSS）の最大の特徴は「フリー」であるということであり，ソースコードを公開し，著作権など企業によって独占されたクローズドソフトウェアを開放する，まさに自由な環境である．しかし，第三者はプロジェクト内部の役割を知るすべはない．

そこでプロジェクトの活動ログをAPIで取得することで，短時間で数多くのプロジェクトメンバの活動ログを収集することができる．GitHubのAPIは以下のようなものがある．

* ForkEvent : フォークを行ったイベント．自分のアカウント内に既存のリポジトリの複製をつくったという活動のログ．
* DeleteEvent : デリートを行ったイベント．プロジェクトで行われていたイベントを削除したという活動のログ．
* IssuesEvent : Issuesを行ったイベント．プロジェクトメンバに限らず，第三者もIssuesを発行したという活動のログ．
* IssueCommentEvent : Issuesにコメントを行ったイベント．プロジェクトメンバに限らず，第三者もIssuesにコメントしたという活動のログ．
* PushEvent : プッシュを行ったイベント．変更履歴をアップロードした活動のログ．
* PullRequestEvent : プルリクエストを行ったイベント．管理者に更新を依頼する活動のログ．

このようなGituHubのAPIのログを調べることによって，今まで明らかになっていなかったOSSプロジェクトにおける各メンバの役割の分担状況を明らかにすることが期待できる．

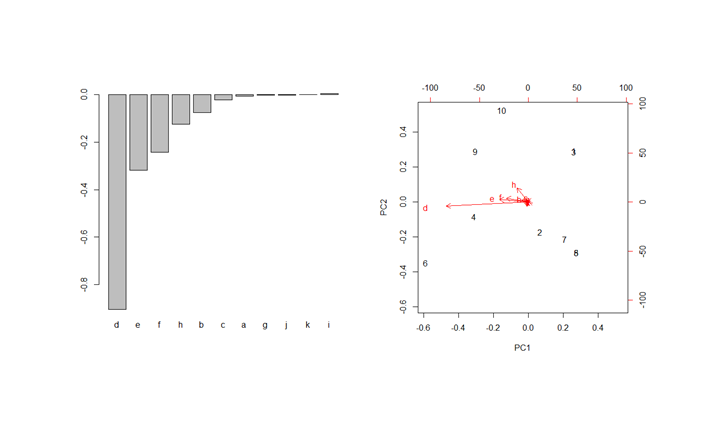
2．研究目的

　Adobe SystemsやLearnBoostなどのGitHub上で公開されているOSSをAPIやRを使用することで，今日のソフトウェア開発を学ぶとともに，プロジェクトメンバの役割分担の実態を明らかにさせる．GitHubのAPIから得られた各メンバの活動ログを統計計算ソフトRで主成分分析などの分析を行うことで，細かな統計ができ，より深く，様々なプロジェクトを学ぶことができる．

3．研究手法

　以下の手法で研究する．

* 1. 調査のためのツールを開発する．
  2. APIを使用し，GitHub上で行われているプロジェクトメンバのデータを収集する．
  3. 各イベントがそれぞれ何回行っているのかを，エクセルのピポットテーブルで一覧表にまとめる．
  4. Rを使用し，得られたデータを主成分分析を用い，活動ログの結果をを解釈する．

****図 1　活動ログの主成分分布図

4．考察・結果

主成分は次のような傾向が多く見られた．PushEventとIssuesCommentEventの絶対値が大きく，正負が逆であることが分かった．つまり,

PushとIssuesCommentは別々のメンバが行っていることが多い，ということが考察できる．

本研究のためのツールは開発できたので，このツールを利用することで，OSSプロジェクトの役割を把握することができ，第三者がより深くOSSでのプロジェクトを知ることができ，今後のOSS開発の発展やプログラマの技術の向上に役立てることができるのではないかと考える．

参考文献

[1] 経済産業省編OSSの利用状況調査導入検討ガイドライン「OSSの現状と今後の課題について<http://www.meti.go.jp/kohosys/press/0004397/1/030815opensoft.pdf>