

OSS 開発プロジェクトにおけるタスク処理過程の定量分析

プロジェクトマネジメントコース 矢吹研究室 1142009 安藤 勇樹

1. 研究背景

ソフトウェア開発のためのホスティングサービスである GitHub では様々なソフトウェアが開発されている。2013 年 12 月には GitHub 上に 1000 万件のリポジトリが作成され、ユーザ数は 400 万人を超えた。数多くのプロジェクトが公開されている GitHub を調査すれば、ソフトウェア開発プロジェクトの分類が可能であると考えられる。

過去に GitHub 上のプロジェクトのチケットを調査し、プロジェクトを分類するという研究があり、プロジェクトの分類が可能であるということが明らかにされていた[2]。しかし、この研究では分類の解釈を人間が主観的に行っており、客観性に欠けているという問題があった。そのため、本研究ではデータマイニング手法を用いて分類を客観的に行う。

GitHub には、リポジトリの人気指標の 1 つにスターが存在する。スターとは、気になるリポジトリをブックマークできる機能である。このスターの数が多いリポジトリは人気が高いことを示している。本研究では、スター数の多いプロジェクトを調査する。

本研究では、プロジェクトを分類するためにチケットを調査する。チケットとは、ソフトウェア開発中に発生した作業や変更履歴の内容を登録する進捗管理ツールである。チケットには未完了チケットと完了済チケットの 2 種類が存在する。未完了チケットは作業が完了されていないチケットを示し、完了済チケットは作業が完了されているチケットを示す。チケットによって作業の進捗状況を可視化できるため、進捗管理が容易になる。

このチケットを中心に開発する手法をチケット駆動開発という。これは作業を開始する前に必ずチケットを発行することを原則とした開発手法である。この開発手法を運用しているプロジェクトは、未完了チケット数と完了済チケット数の時系列変化から進捗状況を判断できる。

2. 研究目的

GitHub 上のプロジェクトを対象とする。チケット数の時系列変化に着目し、データマイニング手法を用いてプロジェクトを分類する。

3. 研究方法

Issue (GitHub 上でのチケット) を GitHub 内のスター数ランキング上位 50 件のプロジェクトから API を用いて収集する。収集する Issue は OpenIssue (GitHub 上での未完了チケット) と CloseIssue (GitHub 上での完了済チケット) の 2 種類である。この 2 種類のチケットの時系列変化を調査し、プロジェクトを分類する。

4. 成果物のイメージ

50 件のプロジェクトの Issue データを時系列解析し、幾つかのパターンに分類する。

5. 進捗状況

50 件のプロジェクトの Issue データを時系列解析した。利用した解析手法は階層クラスタ分析と非階層クラスタ分析である。その変数は、時系列データにフィットする多項式の係数を標準化したものである。以下の図は解析した結果である。

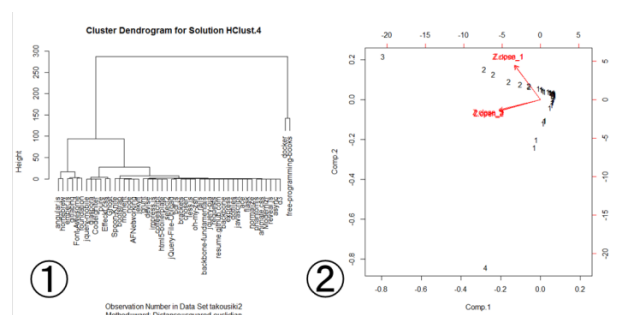


図 1. OSS 開発プロジェクトの分類化

①は階層クラスタ分析、②は非階層クラスタ分析の結果である。Issue 数の増加率の違いにより、幾つかのパターンに分類することができた。

6. 今後の計画

今後はプロジェクトのスター数や共同開発者数が分類パターンに関わっているのかを調査する。これらの関係性を見出せば分類パターンの新たな解釈ができるだろう。

参考文献

- [1] 小川明彦, 酒井誠. チケット駆動開発. 翔泳社, 2012-8-23.
- [2] 久保孝樹. チケットを活用するオープンソースソフトウェア開発の実態調査. 千葉工業大学, 2013, 卒業論文.