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[3]．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．

　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，進捗管理が容易となる．

　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の分類を行う．

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，プロジェクトの分類を行う．分類には統計解析ソフトRを用いる．

4.　成果物のイメージ

　50件のプロジェクトのIssueデータをRで時系列解析し，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Rで時系列解析した．利用した解析手法は階層クラスター分析と非階層クラスター分析である．変数には，時系列データにフィットする多項式の係数を標準化してから指定した．これはIssueデータを折れ線グラフ化した際に追加できる近似曲線である．以下の図は解析した結果である．

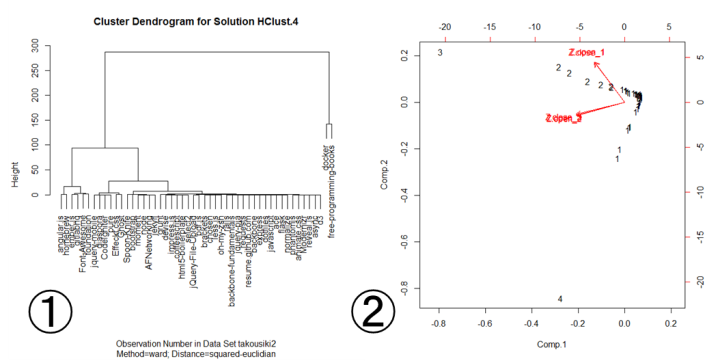


図 1. OSS開発プロジェクトの分類化

　①は階層クラスター分析，②は非階層クラスター分析の結果である．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

　プロジェクトを幾つかのパターンに分類できたが，どのような定義で分類されているのかは完全に把握できていない．今後は分類の定義を明確にする．

　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] 大塚弘記. GitHub実践入門~Pull Requestによる開発の変革~. 技術評論社, 2014-4-25

[3] 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.