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Starが存在する．Star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Star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，Star数を基準にプロジェクトを選択し，調査をする．

　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幾つかのパターンに分類する．分類後，OSS開発プロジェクトの成功パターンを発見する．

3.　研究方法

　Issues（GitHub上でのチケット）をGitHub内のStar数ランキング上位50件のプロジェクトとランダムに選択した50件のプロジェクトから，APIを用いて収集する．収集するIssuesはOpenIssues（GitHub上での未完了チケット）とClosedIssues（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，プロジェクトを幾つかのパターンに分類する．分類後，成功パターンを発見するため，Star数とプロジェクト成功の成否を調査する．

4.　研究結果

　100件のプロジェクトのIssuesデータを時系列解析し，成功パターンを発見した．使用した解析手法は，階層クラスター分析と非階層クラスター分析，自己組織化マップである．その変数は，時系列データにフィットする線形式の係数とIssuesの数，Issuesが発行されてから完了されるまでの時間の平均と標準偏差である．これらの変数は単位が異なるため，標準化する．解析結果から，プロジェクトを幾つかのパターンに分類する．

　本研究では，Star数がプロジェクトの成功要因であると仮定する．成功要因であるStar数にどのような要素が関係しているのかを調査するため，Star数を目的変数とし，重回帰分析をする．

　プロジェクト成功の成否はリリースされたソフトウェアのVerから判断する．Ver1.0以上をリリースしているプロジェクトを成功とし，Ver1.0未満でリリースが止まっているプロジェクトを失敗とする．

　結果として，成功しているプロジェクトにはStar数が多く，Issuesが発行されてから完了されるまでの時間が短いという共通点が見られた．また，失敗しているプロジェクトにはStar数が少なく，Issuesが発行されてから完了されるまでの時間が長いという共通点が見られた．

　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.