データマイニング手法を用いたOSS開発プロジェクトの分類調査

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

　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試みがあった．（久保孝樹（2013）「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（未公刊））．この研究では分類の解釈を人間が主観的に行っており，客観性に欠けていた．そこで，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の特徴として，Issueを利用してのチケット駆動開発の運用が挙げられる．チケット駆動開発とは，チケット管理から生まれたプロジェクト管理技法の一つである．ソフトウェア開発中に発生した課題や変更の内容をチケットという形で発行してから開発するという手法である．どのような課題が発生し，どのような変更を行ったのかを可視化できるため，進捗管理が容易となる．

　チケット駆動開発の運用を中心としたプロジェクトは，未消化チケット数と消化済チケット数の時系列変化から進捗状況を判断できる．未消化チケット数の増加は何らかの課題が発生した事を示し，消化済チケット数の増加は課題の解決を示している．つまり，この2種類のチケット数の差が小さければ進捗状況は順調であり，差が大きければ進捗状況は不順であると判断できる．

2.　研究目的

　チケット駆動開発を運用して成果物を作成している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定量的に行う．

3.　研究方法

　Issue（チケットと同義）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（未消化チケットと同義）とCloseIssue（消化済チケットと同義）の2種類である．この2種類のIssue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」を用いて定量的に行う．

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，4つ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．変数には標準化した多項式の次式を指定した．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．

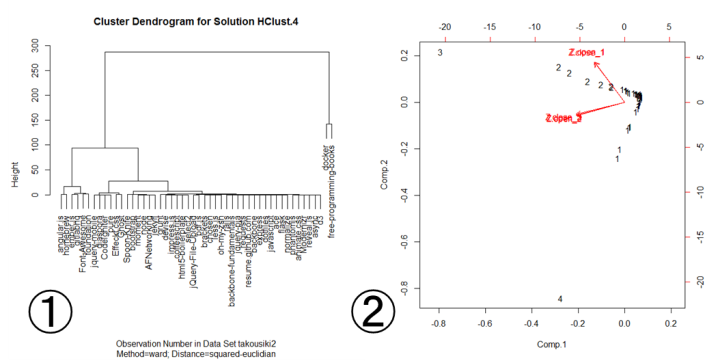


図 1. OSS開発プロジェクトの分類化

　解析した50件のプロジェクトを4つのパターンに分類することができた．①は階層的クラスター分析，②は非階層的クラスター分析の結果である．

* ラベル1はcomp.1とcomp.2の数値が小さい．
* ラベル2はcomp.1とcomp.2の数値が大きい．
* ラベル3はcomp.1とcomp.2の数値が非常に大きい．
* ラベル4はcomp.1が非常に大きく，comp.2が小さい．

　comp.1とcomp.2の数値が小さいとIssue数の増加率は小さい．逆に数値が大きいとIssue数の増加率は大きい．

6．今後の計画

　プロジェクトを4つのパターンに分類することはできたが，どのような定義で分類されているのか完全に把握できていない．今後は分類の定義を明確にしていく．

　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

**「平成26年度PM学科卒研中間審査用研究概要作成要領」**

・指定枚数：1枚／1テーマ

・用紙サイズ：A4

・余白：

上22mm

下22mm

左19mm

右19mm

・段組：タイトル，氏名の部分は1段，

本文は2段組み

・本文の文字数：22字

・本文の行数：50行（ページ設定で指定する値であり，タイトル等があるため実際には50行に満たない）

・日本語フォント：MS明朝

・英語フォント：Times New Roman

・フォントサイズ：

タイトル 　　　　14pt

サブタイトル 10.5pt

所属・氏名 　　　　10.5pt

本文 10.5pt

・句読点：カンマ「，」およびピリオド「．」を使用する

注）

１．原則として，このテンプレートを用いて作成する．

２．英文タイトルは記載しない．

３． 右段下の空行は2行以下にする．最低限，右段下から3行目まで本文を書く．

４．本文は，研究の背景，目的，研究方法，成果物のイメージ，進捗状況（これまでの成果），今後の計画などを簡潔に記述する．

５．その他，不明な点は，事前に指導教員へ確認すること

最低限ここまで（下から3行目）本文を書くこと