

クローンメトリクスの差異が fault-prone モジュール判別に与える影響の分析

梶村 和輝 角田 雅照

近畿大学理工学部情報学科

1はじめに

ソフトウェア開発において、ソフトウェアの品質を高めるために、テスト実施前にfault-proneモジュールを判別することが行われる。いくつかの研究では[1]、判別モデルの説明変数としてコードクローンメトリクスを用いることにより、予測精度が向上することが示されている。

コードクローンとはモジュール内、モジュール間における類似したコード片のことを指し、コードクローンメトリクスとは各モジュールのコードクローン含有率などを指す。コードクローンメトリクスはモジュールの保守性を示す特性の一つで、faultと関連を持つため、予測精度が向上すると考えられる。

2 クローンメトリクスの差異

コードクローン検出方法は様々なものが提案されており、例えばトークンの並びやAST (Abstract Syntax Tree)に基づく検出方法などがある。前者に基づくツールとしてCC Finder X[2]、後者に基づくツールとしてClone DR[3]がある。CC Finder Xは非常にスケーラブルという特長がある。それぞれクローンの定義が異なるため、同一のソースコードを計測対象とした場合でも、クローンメトリクスに差異が生じる。

本稿では、検出手法の違いから生じるメトリクスの差異に着目する。もし検出手法の差異が影響を与える場合、予測精度とスケーラビリティとのバランスを考慮してコードクローン検出方法を選択する必要がある。そこで以下のリサーチクエスチョンを設定した。

- RQ1: クローン検出（メトリクス算出）手法の違いにより判別モデルの精度に違いが生じるのか？
- RQ2: どのクローン検出手法を用いると、最もモデルの精度が高まるのか？
- RQ3: 複数のクローン検出手法を適用して、それぞれから得られたメトリクスを同時に用いた場合、判別モデルの精度は高まるのか？

3 実験

リサーチクエスチョンに答えるため、2種類のfault-proneモジュール判別モデルを構築し、それぞれの精度を比較した。

- モデル1: 単一のコードクローン検出手法に基づくメトリクスを用いた判別モデル
- モデル2: 2つのコードクローン検出手法に基づくメトリクス両方を用いた判別モデル

表1 各モデルの予測精度

モデル	CC Finder X	Clone DR	CC Finder X, Clone DR
分類木	0.586	0.587	0.601
線形判別	0.651	0.688	0.672
ロジスティック回帰	0.664	0.701	0.673
平均	0.634	0.659	0.649

このモデルを線形判別分析、ロジスティック回帰分析、分類木により構築し、それぞれの精度を比較した。モデルの予測精度の評価指標として、AUC (Area Under the Curve) を用いた。AUCの値域は0から1であり、値が大きいほど予測精度が高いことを示す。

オープンソースプロジェクトから収集したソースコードメトリクスとバグデータが公開されており[4]、これを用いてモデルを構築した。用いたデータはLucene 2.4.0から収集されたものである。クローンメトリクスは、ソースコードをダウンロードし、クローン検出ツールを適用することにより算出した。モデル構築時には10-fold cross validationを適用した。

実験結果を表1に示す。平均は、3つの判別モデル構築方法の平均AUCを示す。Clone DRを用いた場合の精度が高く、両方のツールを用いた場合の精度は低かった。よって、RQ1に対する答えは「Yes」、RQ2に対する答えは「Clone DR」、RQ3に対する答えは「No」となる。

4 おわりに

クローンメトリクスの差異が、fault-proneモジュール判別予測モデルの精度に与える影響について分析した。実験に用いるデータを増やし、結果の信頼性を高めることが今後の課題である。

謝辞 本研究の一部は文部科学省科学研究補助費（基盤C：課題番号 25330090）による助成を受けた。

参考文献

- [1] 馬場慎太郎、吉田則裕、楠本真二、井上克郎、 “Fault-Proneモジュール予測へのコードクローン情報の適用,” 電子情報通信学会論文誌D, J91-D, no. 10, pp. 2559-2561, 2008.
- [2] <http://www.ccfinder.net/ccfinderxos-j.html>
- [3] <http://www.semdesigns.com/Products/Clone/>
- [4] <http://bug.inf.usi.ch/download.php>