

偏差値メトリクスを用いた欠陥モジュール予測モデルの提案

藤野 啓輔 坂口 英司 Papon Yongpisanpop
伊原 彰紀 松本 健一

本論文は、ソフトウェアモジュールの欠陥の有無によってソフトウェアメトリクスの分布の違いを示すことにより、新たに開発したモジュールが第三者評価の観点から欠陥を評価することができるソフトウェアメトリクス標準値域を提案する。そして、新たな開発したモジュールの欠陥が混入している可能性（偏差値メトリクス）を標準値域に基づき算出し、fault-prone モジュール予測モデルの構築を試みた。

Toward Building Fault-Prone Module Prediction Using Class Curve

KEISUKE FUJINO HIDESHI SAKAGUCHI PAPON YONGPISANPOP AKINORI IHARA
and KEN-ICHI MATSUMOTO

In this paper, by showing difference in distribution software metrics depending on existence of fault in software module, we propose software standard range which possible to fault evaluation for new product from a third party's point of view. Further, by calculating possibility of fault in new developed module based on software standard range, we tried to model construction of fault-prone prediction.

1. はじめに

ソフトウェアの大規模化に伴い、網羅的なソフトウェアテストの実施が、予算、納期の両面から困難となっている。そこで、欠陥が混入している可能性の高いモジュールに対し、優先的にテスト工数を割り当てるための欠陥モジュール予測モデルが提案されているが導入されている¹⁾²⁾。従来研究では、欠陥モジュール予測モデルは、機械学習により構築され、モジュールから測定したソフトウェアメトリクス（コード行数、サイクロマティック数、分岐数等）を入力として与えることにより行われていた³⁾⁴⁾。しかし、従来研究において提案された欠陥モジュール予測モデルは、各モジュールの欠陥の有無、欠陥が混入していると判定された要因となるメトリクス等を出力するが、そのように判定された要因については示されない。

本論文では欠陥が混入された理由を定量的に示すための指標として、偏差値メトリクスを提案する。ソフトウェアモジュールの欠陥の有無によってソフトウェアメトリクスの分布の違いを示すことにより、新たに開発したモジュールが第三者評価の観点から欠陥を評

価することができる。そして、新たな開発したモジュールの欠陥が混入している可能性（偏差値メトリクス）を標準値域に基づき算出するアプローチを示す。標準値域は2種類あり、一つは、あるプロジェクトが開発したモジュールから標準値域を算出するプロダクト標準値域。もう一つは、複数のプロジェクトでそれぞれ開発されたモジュールから標準値域を算出する一般標準値。一般標準値域は、他プロジェクトが開発したプロダクトも混在されているため、第三者から見たソフトウェアの品質評価を示す。標準値域を用いることにより、評価対象モジュールが、高品質モジュールの標準値域から逸脱している程度を確認することができる。

2. 第三者評価のための欠陥モジュール予測

ソフトウェアの品質評価には、テスト等による開発者自身の評価に加え、検証結果の妥当性を向上させるために開発企業と独立した第三者評価機関による評価が求められている。しかし、第三者評価の依頼のためにコストを要してしまうという問題点も存在している。

本論文では、二種類の標準値域（プロダクト内標準値域、一般標準値域）を提案する。プロダクト内標準値域は、プロジェクト特有の性質を考慮するために、あるプロジェクトが開発した過去のデータからモジュール

†1 奈良先端科学技術大学院大学
Nara Institute of Science and Technology

ルの品質を計測し、メトリクスの標準値域を作成する一般標準値域は、第三者から見たソフトウェアの品質を評価するために、数十の OSS プロジェクトの開発データから、モジュールのメトリクス値を計測することにより、標準値域を作成する。二種類の標準値域を用いることにより、低コストで、自動的に開発したモジュールの第三者評価を低コスト、且つ、自動化することができる。

さらに、本論文では標準値域から欠陥が混入している可能性（偏差値メトリクス）を算出し、欠陥モジュール予測モデルを開発する。従来の欠陥モジュール予測モデルは、目的変数に最も寄与する説明変数を明らかにするアプローチが取られていたが、各メトリクス値が他のモジュールと比べて逸脱する程度を判断することは容易ではなかった。しかし、メトリクス偏差値の利用により、対象モジュールの各メトリクスにおける、標準値域内での分布を把握することができ、過去のデータや他のソフトウェアと比較して、どの程度逸脱しているのかを定量的に判断できるようになる。従って、モジュールから定量的に計測した値に比べて、欠陥が混入する可能性を考慮した値となることから、偏差値メトリクスを用いることにより欠陥モジュール予測モデルの精度向上が期待される。

3. メトリクス偏差値を用いた欠陥予測

3.1 メトリクス偏差値の標準値域作成

本論文では、欠陥モジュール予測のために、従来研究で使用されているプロダクトメトリクスとプロセスメトリクスを使用する。複数プロジェクトのモジュールから各メトリクスを抽出し、欠陥有り、欠陥無し of モジュール群において、それぞれメトリクス偏差値を求める。そして、各プロジェクトで求めたメトリクス偏差値を用いて、欠陥有り、欠陥無しのモジュール群におけるメトリクス偏差値の標準値域を作成する。新たに開発したモジュールを標準値域から評価する事例を図 1 に示す。任意のモジュールを入力として、当該モジュールから計測したソフトウェアメトリクスが、欠陥ありのモジュールから開発した標準値域（図 1 左側）、欠陥なしのモジュールから開発した標準値域（図 1 右側）から評価する。例えば、高品質モジュールの標準値域内であれば、入力モジュールは欠陥を含まない可能性が高く、高品質モジュールの標準値域から離れていた場合、欠陥混入可能性が高くなるといった利用が考えられる。メトリクス偏差値は、標準値域の中央値を 0.5 とし、中央値からの逸脱している程度を偏差値で算出する。

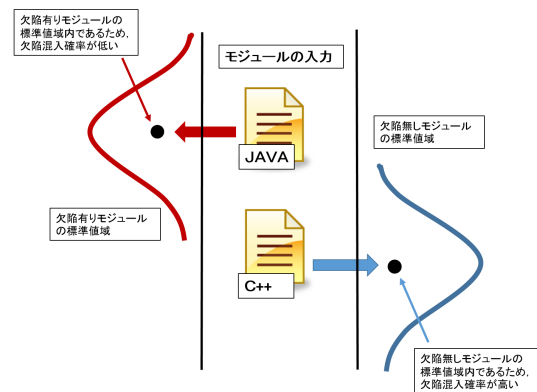


図 1 提案手法のイメージ図

4. おわりに

本論文では、欠陥モジュール予測モデル構築に向けた偏差値メトリクスを提案した。ワークショップにおいては、各メトリクスの欠陥モジュール予測における有用性、偏差値メトリクスの妥当性について議論したいと考えている。

謝辞 本研究の一部は、独立行政法人情報処理推進機構（IPA）「2013 年度ソフトウェア工学分野の先導的研究支援事業」の委託に基づいて行われた。また、公益財団法人中島記念国際交流財団による助成を受けた。

参考文献

- 1) N Nagappan, T Ball, and A Zeller. Mining metrics to predict component failures. In Proceedings of the 28th International Conference on Software Engineering (ICSE'06), pp. 452-461, 2006.
- 2) E. J. Weyuker and T. J. Ostrand. An automated fault prediction system. In Making Software, pp. 145-160. O'Reilly Media, 2010.
- 3) 額綱伸子, 川村真弥, 野村准一, 野中誠, “プロセスおよびプロダクトメトリクスを用いた fault-prone クラス予測の適用事例,” 情報処理学会研究報告, ソフトウェア工学研究会報告, no.6, pp.1-8, 2010.
- 4) まつ本真佑, 亀井靖高, 門田暁人, 松本健一, “Fault-prone モジュール判別に対する外れ値除去法の適用効果,” 情報処理学会研究報告, ソフトウェア工学研究会報告, no.33, pp.49-56, 2007.