



Title	Fault-prone Module Prediction Using Version Histories
Author(s)	Hata, Hideaki
Citation	大阪大学, 2012, 博士論文
Version Type	VoR
URL	<a href="https://hdl.handle.net/11094/2569">https://hdl.handle.net/11094/2569</a>
rights	
Note	

*The University of Osaka Institutional Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

氏名	畑 秀 明
博士の専攻分野の名称	博士 (情報科学)
学位記番号	第 25285 号
学位授与年月日	平成24年3月22日
学位授与の要件	学位規則第4条第1項該当 情報科学研究科情報システム工学専攻
学位論文名	Fault-prone Module Prediction Using Version Histories (履歴情報を用いたフォールトブローンモジュール予測に関する研究)
論文審査委員	(主査) 教授 菊野 亨  (副査) 教授 楠本 真二 教授 尾上 孝雄 准教授 土屋 達弘

## 論文内容の要旨

ソフトウェア開発やメンテナンスにおいて、テストやレビューといった品質確保の活動は、実用的な時間とコストで効果的に行うことが期待される。このため、重点的に品質確保の工数をかけるべき、フォールトブローン（不具合の潜在が疑われる）モジュールを早期に予測することが求められる。近年、ソフトウェアリポジトリからのデータマイニングによって得られる開発履歴メトリクスを用いたフォールトブローンモジュール予測が注目されている。本研究では、2つの手法を提案して現状の課題に取り組んだ。

多くの研究が行われているフォールトブローンモジュール予測であるが、適用が容易でないことが実用化の障壁として挙げられる。特に予測モデル構築に必要な開発履歴メトリクスの計測は、実際の開発に用いられている個別のソフトウェアリポジトリに対応したマイニングツールの開発が必要であり、容易ではない。我々はテキストマイニングに基づく予測手法を提案した。これはソースコードをテキストとして扱い、トークンの出現回数を計測するだけでよいため適用が容易である。オープンソースソフトウェアへの適用実験から、従来の複雑度メトリクスや開発履歴メトリクスに基づく予測モデルと比べても、テキストマイニングに基づく予測モデルは高い精度で予測ができることを確認した。

これまでの開発履歴メトリクスに基づいたフォールトブローンモジュール予測は、モジュールが主にファイルレベルであった。予測されたモジュールを調査する工数を考慮した場合、モジュールのサイズはより小さいことが望ましい。しかし、細粒度なモジュールである、メソッドに対するFault-proneモジュール予測では開発履歴の取得が難しい。この問題に対して、我々は細粒度版管理システム Historage を開発し、メソッドレベルの開発履歴取得を可能にした。このHistorageを用いて計測した開発履歴メトリクスから、メソッドレベルのフォールトブローンモジュール予測を行い、工数を考慮した評価を行った。オープンソースソフトウェアへの適用実験から、同程度の工数で、ファイルレベルと比べて多くのフォールトを調査できることを示した。

開発されるソフトウェアプロダクトの品質確保は極めて重要であり、テストやレビューといった活動によって実現が図られる。しかし現状では、許される開発時間、コスト等への制約が厳しいために十分な活動が展開できず、品質の確保は必ずしも達成されているとは言えない状況にある。このため、フォールトブローン（つまり、フォールトの潜在が疑われる）モジュールを予測して、そのモジュール群に重点的に工数をかけることで作業の効率化を図る開発方法に関する研究が注目されている。

具体的には、ソフトウェアリポジトリに蓄積されている開発データにデータマイニング手法を適用して開発履歴メトリクスを求め、それを用いてフォールトブローンモジュールを予測する多くの試みが報告されている。本論文では、新たに2つの予測手法を提案し、オープンソースソフトウェアへの適用実験を通じて、その有効性を確認している。

先ず、フォールトブローンモジュール予測手法そのものについては既に多くの研究があるが、どれも適用が容易でないために未だ実用化には至っていない。特に、予測モデル構築に必要な開発履歴メトリクスの獲得に困難がある。そこで、本論文ではテキストマイニングに基づく予測手法を提案した。ソースコードをテキストとして扱い、トークンの出現回数から予測モデルを構築する。そのため予測モデルの構築を高速に、かつ容易に実行できる。またオープンソースソフトウェアに対する適用実験から、従来の予測モデルと比べても、本手法は高い予測精度が達成できることを確認した。

次に、開発履歴メトリクスに基づいたフォールトブローンモジュール予測に関する研究はこれまでは主として、ファイルレベルでの予測であった。そのため、実際にテストやレビューを行うには、フォールトを含むメソッドをファイルの中から人手で探索する必要があった。予測結果を実際の開発で利用しようとするなら、予測するモジュールのサイズは小さいことが望ましい。本論文では、この問題に対処するために細粒度版管理システムHistorageを開発し、メソッドレベルでの開発履歴メトリクス取得とフォールトブローンモジュール予測を可能にした。オープンソースソフトウェアへの適用実験から、従来のファイルレベルでの方法と比べ、より多くのフォールトを検出できることを示した。しかも、そのためのオーバーヘッドは従来法とほぼ同程度であった。

以上のように、本論文はフォールトブローンモジュール予測の新たな手法を提案し、ソフトウェア工学研究に対して大きな貢献を与えるものであり、博士（情報科学）の学位論文として価値のあるものと認める。