

令和元年度 情報工学コース卒業研究報告要旨

山本 研究室	氏 名	山 内 泉 水
卒業研究題目	ドキュメントとバグ報告における出現単語数の比較によるバグ集中箇所の予測	
<p>ソフトウェアテストや保守において、バグの集中している部分を特定することは、バグの修正はもちろんテスト工数割り当ての優先順位の決定に役立つ。そうした手法のひとつとして、fault-prone モジュール（欠陥を含む確率の高いモジュール）を特定するモデルの研究が行われている。しかし、モジュール毎の欠陥データが収集できていない場合、予測精度の高い fault-prone モジュール予測モデルの構築は難しい。</p> <p>本研究では、fault-prone モジュール予測とは別のデータであるソフトウェアにおけるバグ報告とドキュメントの出現単語の違いを利用することで、バグの集中している部分を特定する手法を提案する。提案手法は、次のような手順で行う。はじめにバグ報告とドキュメントにおける出現頻度が独立であり、かつ出現頻度がバグ報告のほうが多い語を取り出す。次に、取出した語の出現するバグ報告を取り出す。取り出したバグ報告を、バグの集中しているものとする。</p> <p>ドキュメントが公開されているオープンソースのソフトウェアを対象として、提案手法によって取り出した単語から、実際にバグが集中している部分が特定できるかどうかを調べた。調べた結果、取り出した単語の中の 30 語のうち、実際にバグが集中している部分の特定ができた単語が存在した。次に、実際にバグが集中していると特定できた単語、特定できなかった語はどれだけ存在するのかを調べた。調べた結果、取り出した単語の中の 30 語のうち、実際にバグが集中している部分の特定ができた単語は 13 語であり、43.3%であった。また、取り出した単語を「設計書ドキュメントと比べてバグ報告に多い単語」「マニュアルと比べてバグ報告に多い単語」「設計書ドキュメントとマニュアル両方と比べてバグ報告に多い単語」に分けると、「設計書ドキュメントと比べてバグ報告に多い単語」は 10 語中 9 語で 90%、「マニュアルと比べてバグ報告に多い単語」は 10 語中 1 語で 10%、「設計書ドキュメントとマニュアル両方と比べてバグ報告に多い単語」は 10 語中 3 語で 30%であった。</p> <p>さらに、本研究の対象ソースコードのモジュールのメトリクス値を求め、メトリクス値を用いて基本的な fault-prone モジュール予測を行い、結果を比較する事で、提案手法の評価を行った。バグが集中していると判断したバグ報告の修正モジュール 15 のうち、10 モジュールが、メトリクス値の 1 つであり、値が大きいほど複雑で欠陥のある可能性のあるモジュールとされる Cyclomatic Complexity の値の上位 15 モジュールに含まれていた。Cyclomatic Complexity の値が低いモジュールは、値が 3 のモジュールが 1 つと、最低値である 1 のモジュールが 1 つ、本研究の提案手法において取り出したモジュールに含まれていた。Cyclomatic Complexity の値の上位 15 モジュールの中で本研究の提案手法において取り出したモジュールを除く残りの 5 モジュールのうち、バグによる修正が 1 つしか存在しないモジュールが 1 つ含まれていた。</p>		