

2023年度 独創的研究助成費 実績報告書

2024年 3月29日

報告者	学科名	情報システム工学	職名	准教授	氏名	天壽 聡介
研究課題	埋め込み表現技術を用いたリリース前後のソフトウェア品質向上に関する調査					
研究組織	氏名	所属・職	専門分野	役割分担		
	代表	天壽 聡介	情報システム工学科・准教授	ソフトウェア工学	全体を担当	
研究実績の概要	<p>本研究では、ソフトウェアの不具合を効率的に除去するために、リリース前で不具合が含まれそうなソースコードを特定する技術の改善に取り組む。</p> <p>ソフトウェアから継続的に収益を得るためには、ユーザの品質に対する期待に応えることが重要である。不具合を含む可能性が高いモジュール（ソフトウェアの部品）は優先的にチェックできることが品質確保の観点から望ましい。従来、リリース前に不具合を含む可能性が高いモジュールを予測する研究、リリース後の不具合報告から原因となるモジュールを特定する研究がそれぞれ行われてきた。</p> <p>従来研究では、メトリクス（変更行数や識別子の数）を予測・特定に用いていたが、近年では、不具合予測、不具合箇所特定に埋め込み表現技術を適用する研究が行われ始めている。埋め込み表現技術は単語やデータの構造などを機械学習モデルで利用しやすい多次元の数値ベクトルに変換する手法の総称である。</p> <p>本研究では、従来メトリクスに加えて埋め込み表現技術を併用した場合に、不具合予測の性能にどのような変化が生じるか評価した。ソフトウェア工学の分野では、実際の開発プロジェクトで収集されたデータを用いて提案手法を実証的に検証することが一般的になっている。本研究もこの方針に従い、実際の開発プロジェクトで収集されたソースコードとその更新履歴から収集できる埋め込み表現とメトリクスを組み合わせた。ソースコードの更新履歴に関する情報を細粒度で収集したデータが必要であるが、先行研究で収集されたデータを用いた。</p> <p>本課題では埋め込み表現技術の有効性が主眼である。特に、埋め込み表現を得るために行われる事前学習で用いられるデータの違いが大きく影響する。そこでソフトウェア開発に関する質問サイトである Stack Overflow で学習した埋め込み表現と、Google News で学習された埋め込み表現を比較評価した。なお、ここでは自然言語に関する埋め込み表現技術をソースコードに適用している。具体的には、ソースコードに出現する識別子を自然言語とみなす。</p> <p>不具合の中でも近年特に重要視されている脆弱性に関する不具合に着目して、これらの埋め込み表現と、従来型のメトリクスと組み合わせた場合の予測性能の変化について評価した。その結果、(1) 一般的な語彙で学習された埋め込み表現の方が予測に役立っていること、(2) 従来型のメトリクスと組み合わせることで一定の性能向上は見られたが、統計的優位な変化でないこと、が明らかとなった。</p>					

※ 次ページに続く

研究実績 の概要	今後の研究としては、利用するメトリクスの種類を増やすこと、自然言語の埋め込みに加えて、ソースコードをグラフ表現とみなす埋め込み表現などを用いた実験・評価が考えられる。
成果資料目録	[1] S. Amasaki, H. Aman, T. Yokogawa, “An Evaluation of Word Embeddings on Vulnerability Prediction with Software Metrics,” vol.9, no.2, pp.26–29, Work in Progress in Embedded Computing Journal, September, 2023.