水野　修 教授

お問合せ先
corc@kit.ac.jp
産学公連携推進センター

不具合を含むソフトウェアモジュールの新しい検出法

迷惑メールフィルタのソフトウェア開発への応用

■キーワード
ソフトウェア　不具合検出　テキスト分類　データマイニング　実証ソフトウェア工学
モジュール　リポジトリ

■研究の概要

不具合を含みそうなソフトウェアモジュール(Fault-proneモジュール)の検出はソフトウェア工学における重要な課題の1つで、これまでにも多くの研究が行われてきました。それらの研究の多くはソフトウェアの複雑度を表した指標(メトリクス)を利用したモデルによる予測です。しかし、モデルの構築にはメトリクスの収集環境が必要となります。メトリクスの収集には意外と手間がかかるために、そもそものメトリクス収集が進まないという事態にもなり得ます。

そこで、本研究室ではソースコードさえ準備すれば簡単に利用可能なFault-proneモジュールの検出手法を提案します。この手法を私たちは「Fault-proneフィルタリング」と呼んでおり、迷惑メールフィルタで用いられるテキスト分類技術をソフトウェアのソースコードに適用することで、不具合のあるモジュールを発見する試みとなっています。本手法に必要なものはソースコードのみであり、全く事前の知識がない状態からでも開発プロジェクトに適用できるという特徴を持っています。

■研究・技術のプロセス／研究事例

提案する手法では、ソースコードの開発履歴、および、バグトラッキングシステムに蓄積された不具合の情報から、まず不具合を含むモジュールの特徴を表す辞書を作成します。そして、新規に作成されたモジュールに対して、この辞書に基づいたフィルタリングを行い、そのモジュールがfault-proneである確率を計算します。

システムの有効性を評価するために、オープンソースソフトウェアEclipseを用いた実験を行いました。実験では実際の開発を模した手順で知識ゼロの状態から学習・分類を繰り返し、最終的な予測精度を求めました。その結果、実験開始から通算50000モジュール程度の学習で精度が85%を超えることを確認できました。また、最終的な予測精度は90%前後になりました。さらに実験から得られた特徴として、「不具合が有るものを誤って不具合が無いと予測する」率が他手法に比べて低いことを確認しました。

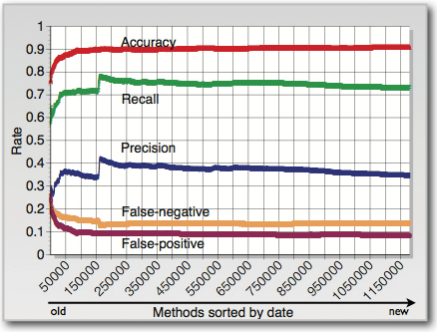

図1：Fault-prone　フィルタリング法の概要

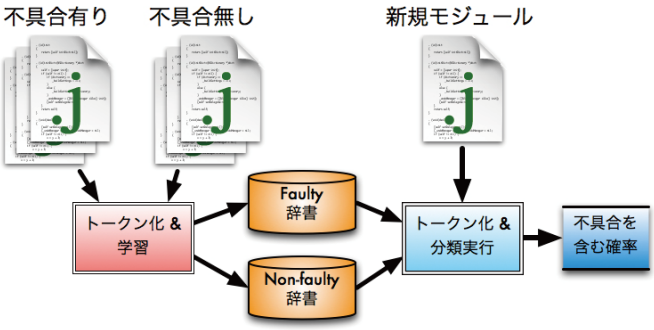

図2：Eclipseの100万モジュールに対して予測を行った結果
予測開始時は精度が低いが、次第に精度が上がっていることが分かる。

■セールスポイント

既存の版管理システムやバグトラッキングシステムのデータから学習させることができるため、現状利用しているシステムとの統合も簡単だと考えます。実証実験のためのデータを募集しています。