

新しい迷惑メールフィルタリング検出法

ソフトウェアのソースコードで不具合も発見！！

Fault-prone フィルタリングの概要

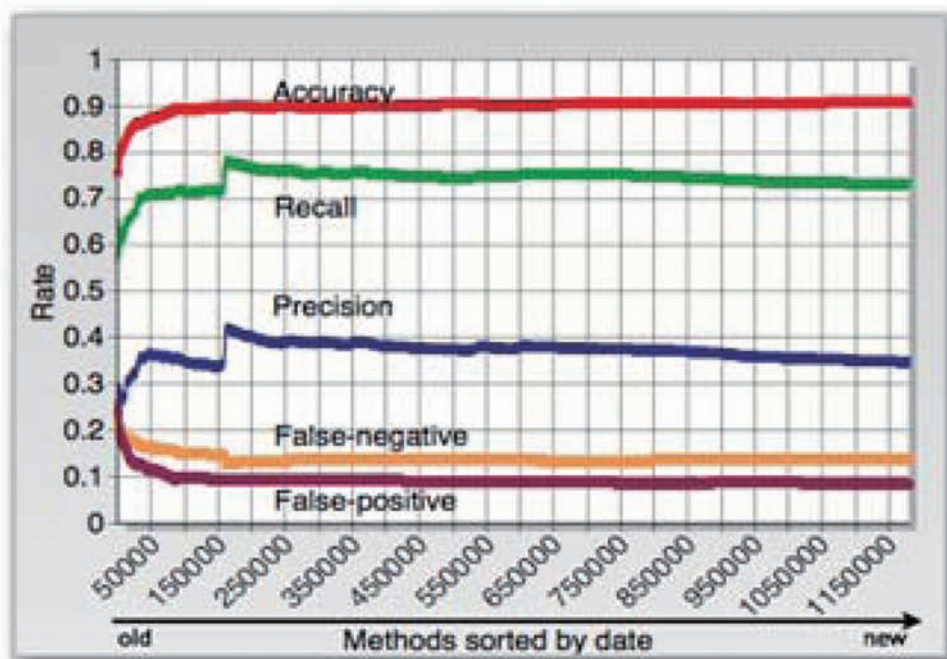
着想の背景

Fault-prone フィルタリングはスパムメール（迷惑メール）の判定を行うスパムフィルタで用いられるテキスト分類フィルタ技術を利用しています。スパムフィルタは、過去に受信した電子メール内の単語群を利用して、スパムメールと通常のメールを判別するための辞書を作成します。そして、新たに受信した電子メールについては、ベイズ識別等の技術により、スパムか否かを判定します。

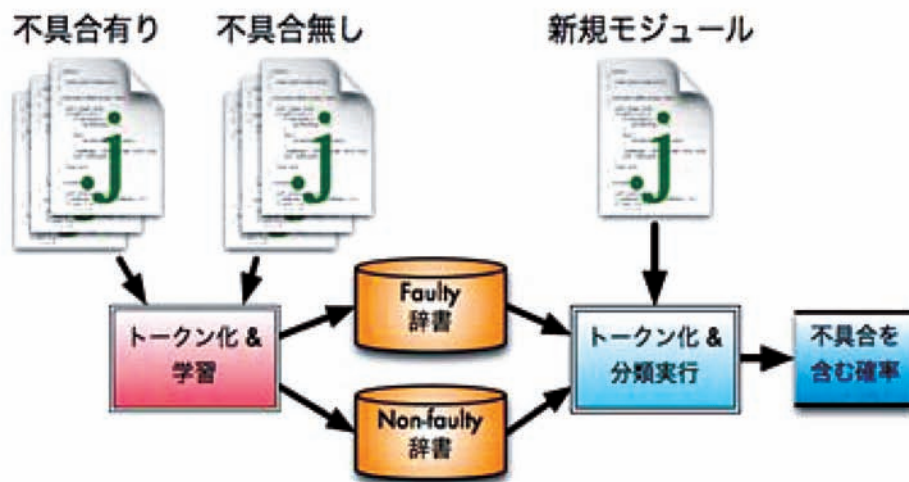
Fault-prone フィルタリングの適用

従来技術と開発手法

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により、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。



Fault-prone フィルタリング法の概要



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、次第に精度が上がっていることが分かる。)

研究のプロセスと今後の課題

ここで提案する手法では、ソースコードの開発履歴、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、新規モジュールに対して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。

スパムフィルタを利用した Fault-prone モジュールの予測手法、Fault-prone フィルタリングとその実用的な適用手法、について述べました。

提案する Fault-prone フィルタリングはソースコードのみを入力して実施することが出来るため、実プロジェクトへの適用も容易であると考えます。

今後は企業で実施された開発への適用が挑戦すべき課題の1つです。

研究者

京都工芸繊維大学
大学院工芸科学研究科
情報工学部門

准教授 水野 修

研究テーマ

実証的ソフトウェア工学研究
不具合ソフトウェアモジュール開発研究
情報学基礎研究
ソフトウェア開発