

## 平成26年度 情報工学コース卒業研究報告要旨

高田 研究室	氏 名	古橋拓人
卒業研究題目	パターンマイニング技術を用いた 組込みプログラムからのイディオムの検出	
<p>ソフトウェア開発では、プログラムにイディオムが用いられることが知られている。イディオムとは、プログラミング時に発生するコーディングパターンのことである。イディオムは横断的関心事になることがある。横断的関心事とは、ロギングや同期処理といったモジュール化が困難なソフトウェアの機能のことである。これらの機能は複数のモジュールにまたがるプログラム群として存在する。こうしたプログラム群は、元となるプログラムを複製し、配置された場所に応じて適宜変更されるなどして実装される。こうした横断的関心事では、もしそのプログラム群の1つに不具合が発見された場合、別の場所に配置された同じプログラム群で同様の不具合が発生する可能性が考えられる。しかし一方で、横断的関心事は散在しており、また類似していても個々に異なる記述を持つ。そのため、横断的関心事の検出は困難である。以上の経緯から、イディオムなどの横断的関心事は開発や保守を困難にする要因になっている。</p> <p>このようなイディオムの自動検出を行うモチベーションは主に2つある。1つ目は、ソフトウェア開発メンバの間でイディオムについての認識を共有することである。メンバ全員がイディオムについて認識を共有することは、開発や保守において有用である。しかしメンバ間での認識の共有は容易ではない。多人数で開発が行われる中で、開発メンバが途中で入れ替わるようなプロジェクトも存在する。また、ソフトウェアに関するドキュメントも全てのプロジェクトで整備されているわけではない。そのような場合にイディオムが自動的に検出できれば、イディオムについての認識の共有が容易になる。2つ目は、プログラムのチェックを支援することである。イディオムを検出し、同じイディオムのプログラム群と比較することで、あるプログラム片に必要な記述が抜けていたり、不必要な記述がされていたりすることが判別しやすくなる。</p> <p>本研究では、C言語で記述された組込みソフトウェアのプログラムを対象とし、系列パターンマイニング手法を用いてイディオムの検出を試みた。組込みソフトウェアは汎用ソフトウェアと異なり、リアルタイム性などの要求がある。それらの要求に対応するため割込み機構といった特殊な機能を持つ。このため、プログラムの記述も汎用ソフトウェアと異なってくる。従来の汎用ソフトウェアを対象としたイディオム検出に対し、組込みソフトウェアの特徴を踏まえた検出を行った。評価実験では、組込みOSの1つであるTOPPERS/ATK2を対象とした。まず、イディオムの構成要素を抽出するツールを作成し、プログラムからパターンマイニングに必要な情報として、関数呼出しと制御構造を抽出した。次に抽出結果に対し、統計解析環境Rを用いて系列パターンマイニングを行った。最後に、パターンマイニングの結果をフィルタリングし、パターンマイニングの結果から包含関係を整理するなどして冗長な系列パターンを取り除いた。フィルタリングを行うことで、系列パターンマイニングで検出された数千から数十万の膨大な量のパターンを、たかだか200個までのパターンに絞り込んだ。以上の検出実験をTOPPERS/ATK2の4つのプロダクトに実施し、検出されたイディオムを分析した。検出対象である4つのプロダクトの中で、マルチコア対応またはメモリコントロール機能が付加されたプロダクトではパターンの数やパターンの長さが大きくなっていることが確認された。</p>		