

平成27年度 情報工学コース卒業研究報告要旨

高田・本田 研究室	氏 名	清水 隆也
卒業研究題目	データマイニング技術を用いたC言語プログラムからのパターン違反検出	
<p>自然言語におけるイディオムと同じく、プログラミング言語においてもコーディングパターンが存在する。例えば、C言語では、ファイルを開くために <code>fopen</code> 関数を呼び出した後、ファイルの読み書きを行う処理を記述し、最後にファイルを閉じるため <code>fclose</code> 関数を呼び出す。自然言語においてイディオムが誤用されるように、プログラミング言語においてもコーディングパターンの誤用が行われ、コーディングパターンの誤用を含むコード片はパターン違反と呼ばれる。パターン違反は、ソフトウェアの不具合や保守性の低下の原因となる。例えば、<code>fopen()</code> 関数の後に <code>fclose()</code> 関数を呼出し忘れると不具合の原因になりうる。また、パターン違反を含むプログラムは一貫性が欠如しているため、理解しづらく保守性が低いと言える。</p> <p>組込みプログラムには高い信頼性や保守性が求められることが多いため、パターン違反を除去することが強く望まれる。しかし、組込みプログラムではコーディングパターンが多く作りこまれる傾向にある。その理由は、組込みプログラムは、信頼性とともに対応性も求められるため、JavaやC#のようなモジュール性の高い言語ではなくC言語で記述されることが多いからである。コーディングパターンが多く作りこまれるということは、パターン違反が作りこまれる可能性が高まる。</p> <p>これまでに、C言語で記述されたプログラムを対象として、パターン違反を検出する手法がいくつか提案されてきた。これら手法は、まず関数呼出し系列に対してパターンマイニング技術を適用することでコーディングパターンを抽出し、その後コーディングパターンに違反するプログラムをパターン違反として検出する。しかし、これら手法は組込みプログラムを適用対象としていないため、組込みプログラムにおいて多用されるプリプロセッサ命令やジャンプ命令をコーディングパターンの構成要素として抽出しない。</p> <p>そこで本研究では、組込みプログラム向けパターン違反検出ツールの開発を行う。本ツールは、プリプロセッサ命令やジャンプ命令をコーディングパターンの構成要素として抽出し、抽出したコーディングパターンからパターン違反を検出する。コーディングパターンの検出には高速に系列パターンマイニングを行うことができる <code>PrefixSpan</code> アルゴリズムを採用した。既存研究が開発したコーディングパターン検出ツールにはスケーラビリティ上の問題があったため、系列パターンマイニングに用いるライブラリを変更することで、スケーラビリティの改善を行った。また、パターン違反に対するフィルタリング方法や順位付け方法を改善することで、不具合や保守性低下の原因になりやすいと考えられるパターン違反を優先的に提示できるようにした。</p> <p>適用実験では、本ツールをC言語で記述された組込みプログラムに適用することで、パターン違反の検出数や検出時間の評価を行った。また、本ツールが検出および優先度付けしたパターン違反に対して、不具合や保守性の低下の原因になりうるか検討を行うことで、フィルタリング方法や順位付け方法の有効性を評価した。</p>		