

平成21年 4月30日現在

研究種目：若手研究(B)  
 研究期間：2007～2008  
 課題番号：19700023  
 研究課題名（和文） ソフトウェア再利用支援のためのコードスニペットマイ  
 ニング技術に関する研究  
 研究課題名（英文） A Technique for Mining Reusable Code Snippets

研究代表者  
 小林 隆志（Takashi Kobayashi）  
 名古屋大学・大学院情報科学研究科・准教授  
 研究者番号：50345386

## 研究成果の概要：

ソフトウェア部品を再利用する工程を支援するため、再利用部品の過去の利用方法に関する様々な情報を収集・抽象化することで、部品を利用するために必要なソースコード片の抽出方法およびその応用手法を提案した。提案手法では、構造と文字列双方の類似度を利用することで、制御構造を含む頻出コード片を効率良く発見可能であり、また、開発活動履歴を利用することで効果的なソースコード片提示を可能としている。

## 交付額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	1,900,000	0	1,900,000
2008年度	1,300,000	390,000	1,690,000
総計	3,200,000	390,000	3,590,000

研究分野：ソフトウェア工学

科研費の分科・細目：情報学ソフトウェア

キーワード：ソフトウェア再利用、プログラム解析

## 1. 研究開始当初の背景

近年、様々なソフトウェアがオープンソースで開発され、ソフトウェアモジュール、クラスライブラリ、フレームワークなど再利用可能な部品群も、Jakarta プロジェクトのように、非常に完成度の高いものが WWW 上に公開されるようになってきた。

これらの再利用部品を使用する際には、付属のドキュメント等の指示に従い、再利用部品が提供する機能や機構などを再利用するための実装や設定を行う必要がある。

単一の機能を提供する再利用部品であれば利用するための実装・設定は単純であるが、フレームワークのように汎用性の高い複雑な機能・機構を提供する再利用部品では、再利用のための実装や設定は多岐にわたるため、開発者が膨大なドキュメントやサンプルコードを熟読する必要があり、初期導入コストが高い、適切な部品を選択することが困難という問題があった。

そのため、設定項目の一覧とそれらの依存関係を解析・提供することで利用を支援する様々な手法が提案されており、近年では、設

定箇所を独立させ簡略化する手法なども提案されてきた。しかしながら、これらの支援手法は事前に想定できる一般的な利用方法に対してのみ有効であり、各開発者は部品の作成者が想定した一般的な利用方法をベースに、各自の目的に応じてサンプルコードを書き換えながら開発を行う必要があった。

## 2. 研究の目的

本研究では、支援のための情報を部品作者が想定する一般的なものや、部品自身を解析し取得するだけではなく、ソフトウェアリポジトリ等内にある実際の利用方法や、過去の開発でのデバッグ履歴からの実行履歴情報などからも取得することに着目する。

対象部品の実際の利用方法を活用することで、作成者が想定した一般的な利用方法以外の支援も可能になるほか、過去の開発でのデバッグ履歴による実行履歴情報を利用することで精度の高い依存関係解析を行い、再利用部品の利用のための実装・設定に対し、より高品質な支援が可能になると考える。

本研究では、上記の観点から、ソフトウェアリポジトリやデバッグ履歴から再利用部品の過去の利用方法を抽出し利用支援を行う手法の開発を目指す。具体的には、以下の点を明らかにすることを目的とする。

- (1) デバッグ履歴などの実行履歴情報とソースコードの情報から、ソースコードの構成要素間の依存関係、振る舞いに関する情報などを表現するグラフを作成し、再利用部品に関連するグラフ群に対してデータマイニングやリンクマイニングの手法を応用して、再利用部品の利用に必要な情報を発見する手法
- (2) 上記の情報から抽象化された利用方法モデルを構築する手法
- (3) 利用方法モデルを用い、開発者に対し利用方法を実装コードに近いコード断片である、コードスニペットとして提示する手法

## 3. 研究の方法

- (1) 本研究では、まずマイニング対象を表現するためのモデル化を検討した。ソースコードの抽象構文木に静的解析によって解析できる各種依存関係を加えたソースコードに関する基本モデルを作成し、このモデルに対して、利用方法と考

えることのできる部分を探査し、コードスニペットの候補を抽出する方法を検討した。

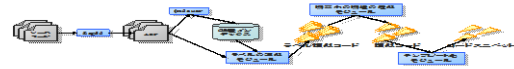


図 1：提案手法の概要

- (2) 次に分析基盤を整備した。複数のコード検索サイトを統合検索し、検索結果が依存するコード群を解析しそれらを収集するツールを作成し、Web 上で利用できるコード検索サイトを活用し、部品の利用方法を発見するための基盤を作成した。さらに、プログラム内の依存関係等を視覚的に表現できるように、モデルの可視化方法及びそのツールの開発を行った。

- (3) これらの基盤を利用し、制御構造を含む再利用性の高いコードスニペット抽出技術を開発した。提案手法では既存のソースコード群から、頻出するコードパターンを効率よく抽出するために、図に示すように構文要素の出現数を特徴とした構造特徴ベクトルと、出現する文字列の類似度を合わせた識別子類似度を計算する。本研究ではこれらの類似度を利用して効果的にスニペットを抽出する手法を開発した。

この手法を用いることで、再利用の際に必要なが従来方式では検出できない制御構造を含むコードスニペットの効果的な抽出を可能とした。

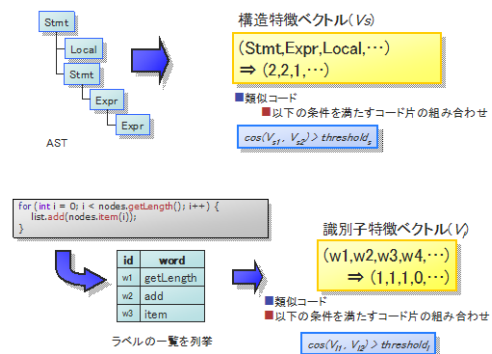


図 2：提案手法における特徴ベクトル

- (4) また、適切なコードスニペットの検索手法として、過去の開発履歴における同様の問題とその解法を効果的に検索する手法を検討し、スタックトレース情報を

利用する検索方法を提案した。提案手法では、開発活動を記録するツールによって取得した、開発活動中のテスト実行における例外発生スタクトレースを記録し、それらの類似度を利用する検索手法を提案した。

- (5) 上記と並行して、コードスニペット抽出の際にデザインパターンの情報を利用できるように、デザインパターンが利用されている箇所を静的情報と実行時の動的情報を合わせて効率よく検出する手法を提案した。さらに、再利用対象としてフレームワークに着目し、フレームワーク利用コードに関する分析精度を向上させるため、要求仕様およびフレームワークに実装されている機能をモデル化し、それらを比較することで、要求仕様に合致する複数のフレームワークから適切なものを選択し、その使用方法を導出するための手法を開発した

#### 4. 研究成果

解析基盤に関しては、依存関係や振る舞いの可視化ツールおよびコンポーネント検索システムを実装し、可視化ツールを用いたコードナビゲーション手法、コンポーネント検索手法をそれぞれ発表した。

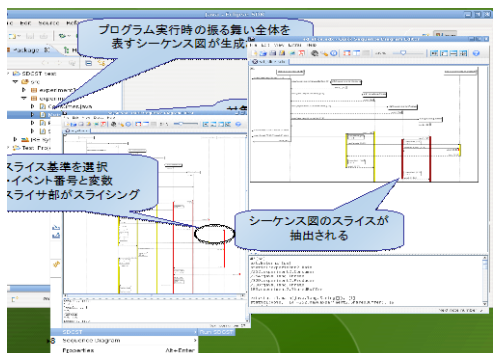
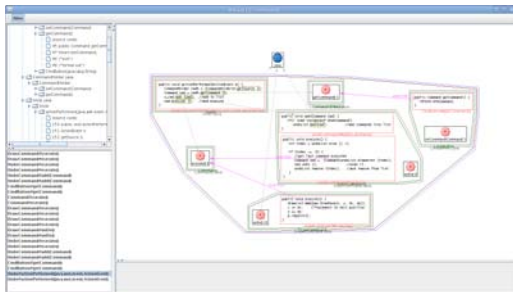


図3：作成したナビゲーションツールと可視化ツール

(3) のコードスニペット抽出手法についても学会発表を行った。さらに、システムを実

装し、The Java Tutorialsに含まれる767のJavaソースファイル及びWeb Tools Platformのorg.eclipse.jstパッケージ以下のクラスである5457のJavaファイルから、それぞれのソフトウェアに対してコードスニペットを抽出する実験を行った。その結果、The Java Tutorialsからは106個のJavaの標準ライブラリの、Web Tools Platformからは1080個のEclipseプラグインのコードスニペットを抽出できることを確認した。それらのコードスニペットのうち、制御構造を含むものはそれぞれ、35個と754個であった。以下に抽出されたコードスニペットの例を示す。

```
JFrame f = new JFrame("ImageOps");
f.addWindowListener(new WindowAdapter() {
    public void windowClosing(WindowEvent e) {System.exit(0);}
});
JApplet applet = new ImageOps();
f.getContentPane().add("Center", applet);
applet.init();
f.pack();
f.setSize(new Dimension(550, 550));
f.show();
```

図4：抽出されたコードスニペット例

提案手法では、構造特徴を利用して抽出を行うため、既存手法では取得できない制御構造をコードスニペットに含めることができる。部品を再利用し実際に使用する際には、既存手法の単なるメソッド列の提示だけで不十分であり、入出力の条件設定や、ループ処理など典型的な処理を同時に提示することが重要である。提案手法では、制御構造を含むテンプレート型のスニペットを効率よく抽出することが可能であり、再利用支援に有効な手法であるといえる

(4) のコードスニペット検索に関する手法に関しても提案手法を実現するツールを実装し、ツールを用いた評価結果を学会発表した。

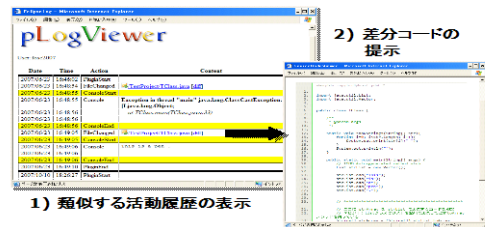


図5：作成した検索システム

実験では、複数の被験者の開発作業履歴を1年間記録し、その情報をもとに、1) 一般的な部品、2) 組織内のプロジェクトに固有の部品、それぞれに対して誤った利用とその訂正に関するスニペットを検索する実験を行

った。

実験の結果、提案手法を用いることで、問題解決に必要なコードスニペットを効果的に発見できることを確認した。Web 検索や従来のコード検索では、組織内の部品に対する再利用コードに関しては情報が少ないため、特に困難であり、一般 API に関しても問題解決に必要な情報は一般的な利用法に隠れてしまい発見が困難である。提案手法では例外発生時のスタックトレースにより作業状況の類似度を判定できるため、効果的に必要な情報が提示できるという利点がある。

部品を再利用する際に必要となるコードスニペットをソースコードの構造情報を考慮した類似性を利用したマイニングである(3)の手法は、トークン単位での類似計算やマイニングが一般的な従来手法と比べ、構造情報を利用することの大きな利点が存在する。今後、より高度な構造情報を利用するように拡張をさせることで、適用範囲の拡大および、精度向上が期待できると考える。

また、(4)で提案している、開発活動履歴におけるスタックトレース情報とソースコードの変化を紐付け、再利用部品を使用する際の問題解決に必要なコードスニペットを検索する手法は、発表の場において、企業等の実開発者などから、社内における再利用部品の開発、およびその利用の際に有用な手法であるとの評価を受けている。プロダクトライン開発が注目され、企業内の再利用可能な部品を資産としてとらえる考えが広まってきた昨今では、過去の開発者の活動履歴を有効に活用し開発を効率化できる本手法は非常に有用であると考え

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 1 件)

- (1) Shinpei Hayashi, Junya Katada, Ryota Sakamoto, Takashi Kobayashi, Motoshi Saeki: Design Pattern Detection by Using Meta Patterns, IEICE Transactions on Information and Systems, Vol. E91-D, No. 4, pp. 933-944 (2008) 査読有

[学会発表] (計 12 件)

- (1) 野田訓広, 小林隆志, 山本晋一郎, 阿草清滋: 高精度なデータ依存解析に基づくシーケンス図スライシング手法, 情報処理学会 第 163 回 ソフトウェア工

学研究発表会, 2009. 3. 17-18, 東京

- (2) 小林隆志, 林晋平: 多量のソフトウェア関連データを用いた開発支援, ウィンターワークショップ 2009・イン・宮崎, 2009. 1. 23-24, 宮崎
- (3) 林晋平, 小林隆志: マイニング結果の評価に関する一考察 -- デザインパターン検出を例に, ウィンターワークショップ 2009・イン・宮崎, 2009. 1. 23-24, 宮崎
- (4) 大須賀俊憲, 小林隆志, 山本晋一郎, 阿草清滋: 構文木の構造とラベルの類似性に基くコードスニペット抽出 ウィンターワークショップ 2009・イン・宮崎, 2009. 1. 23-24, 宮崎
- (5) Teruyoshi Zenmyo, Takashi Kobayashi, Motoshi Saeki: A Technique to Check the Implementability of Behavioral Specifications with Frameworks, 15th Asia-Pacific Software Engineering Conference, 2008. 12. 5, Beijing, China
- (6) 谷聡貴, 小林隆志, 山本晋一郎, 阿草清滋: スタックトレース情報を用いた問題解決経験の検索, ソフトウェア科学会 第 15 回ソフトウェア工学の基礎ワークショップ, 2008. 11. 14 淡路
- (7) 大須賀俊憲, 小林隆志, 山本晋一郎, 阿草清滋: コードドクローンをを用いたコードスニペット抽出の試み, ソフトウェア科学会 第 15 回ソフトウェア工学の基礎ワークショップ, 2008. 11. 14 淡路
- (8) 善明晃由, 小林隆志, 佐伯元司: ユースケース記述に対するフレームワーク利用法の導出, 情報処理学会 第 159 回ソフトウェア工学研究会 2008. 3. 17-18, 東京
- (9) 上原伸介, 大須賀俊憲, 小林隆志, 金子伸幸, 山本晋一郎, 阿草清滋: ソフトウェア理解支援のための多粒度ソフトウェアマップ, 情報処理学会 第 159 回 ソフトウェア工学研究会 2008. 3. 17-18, 東京
- (10) 小林隆志, 大須賀俊憲, 上原伸介, 金子伸幸, 阿草清滋: コードスニペットマイニング -- 再利用過程の再利用にむけて --, ウィンターワークショップ 2008・イン・道後 2008. 1. 24-25, 道後
- (11) 大須賀俊憲, 金子伸幸, 山本晋一郎, 小林隆志, 阿草清滋: ソフトウェア統合検索を利用した再利用支援システム, 日本ソフトウェア科学会 第 14 回ソフトウェア工学の基礎ワークショップ, 2007. 11. 8-10, 下関
- (12) 小林隆志: ソフトウェア再利用技術: 経験の記述と利用, 平成 19 年度電気関係学会東海支部連合大会シンポジウム,

2007. 9. 27, 信州大学

6. 研究組織

(1) 研究代表者

小林 隆志 ( Takashi Kobayashi )

名古屋大学・大学院情報科学研究科・特任  
准教授

研究者番号：50345386