様式Ｃ－１９

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成２４年６月８日現在

機関番号：１４４０１
研究種目：若手研究（B）
研究期間：２０１０年度～２０１１年度
課題番号：２２７００００３１
研究課題名（和文）：プログラム依存グラフを用いたコードクローン検出法の実用化に関する研究

研究課題名（英文）：Study on Code Clone Detection Using Program Dependency Graph For Practical Realization

研究代表者：肥後　芳樹（HIGO YOSHIKI）
大阪大学・大学院情報科学研究科・助教
研究者番号：７０４５２４１４

研究成果の概要（和文）：プログラム依存グラフを用いた検出法の高速化および高精度化手法を提案した．提案手法をツールとして実装し，オープンソースソフトウェアに対して実験を行った．実験により，提案手法の有効性を確認した．

研究成果の概要（英文）：In this research, we proposed methods to realize scalable and accuracy PDG-based code clone detection. The proposed methods were implemented as a software tool, and we applied it to several open source software systems. As a result, we confirmed that the proposed methods are effective.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
２０１０年度	1,100,000	330,000	1,430,000
２０１１年度	1,100,000	330,000	1,430,000
年度			
年度			
年度			
総　計	2,200,000	660,000	2,860,000

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：コードクローン，プログラム解析，ソフトウェア保守

１．研究開始当初の背景

コードクローンとは，ソースコード中に存在する同一または類似したコード片を表す．コードクローンはコピーアンドペーストや定型処理などのさまざまな理由によりソースコード中に作りこまれる．コードクローンの存在はソフトウェア保守を困難にするといわれている．例えばあるコード片に対して修正を加える場合，もしその部分がコードクローンであれば，対応する全てのコードクローンに対しても同様の修正の是非を検討しなければならない．また，修正すべきコード片を見落としてしまう危険性も含んでいる．

ソースコード中のコードクローンを検出するためのさまざまな手法が提案されている．既存の検出手法は用いている技術により，行単位での検出，字句単位での検出，抽象構文

木を用いた検出，プログラム依存グラフを用いた検出，メトリクスを用いた検出に大別される．既存の各手法は一長一短であり，全ての面において他の技術よりも優れているものはない．プログラム依存グラフ（Program Dependency Graph，以降 PDG）を用いた検出の長所と短所を示す．

長所： 非連続コードクローンを検出することができる．非連続コードクローンとは，1 つのコードクローンを構成する要素（プログラムの文や式など）が必ずしもソースコード上で隣接していないコードクローンを指す．コピーアンドペーストしたコード片に対して修正漏れがおこるという報告があり，そのような部分は非連続コードクローンとなるため，非連続コードクローンを検出することはソフトウェア保守の観点から重要である．

短所： 行単位，字句単位，および抽象構文木を用いた検出に比べると連続コードクローンの検出能力が劣る．また，検出に必要な計算コストが高いため，実規模ソフトウェアに対しては適用が難しい．

２．研究の目的

プログラム依存グラフを用いたコードクローン検出技術の問題点（検出に長い時間を必要とする・検出できない種類のコードクローンが存在する）を解決し，実規模ソフトウェアから短時間かつ十分な精度でコードクローン検出を行えるようにする．

３．研究の方法

目的を達成するためのキーアイディアは以下のとおりである．

(1)PDG 頂点間への実行依存関係の導入と双方向スライスの利用：

連続コードクローンの検出能力を高めることを目的とした提案である．従来手法に比べて，プログラムスライスでたどる頂点の範囲を拡大することができる．

(2)スライス基点とするPDG 頂点数の削減：

計算コストの削減を目的とした提案である．PDGを用いた手法の計算コストが高いのは2つの原因がある．1 つめの原因は，コードクローンを検出するための（同形部分グラフを特定するための）基点となるPDG 頂点の数が非常に多いことである．2 つめの原因は，同形部分グラフを検出すること自体がNP完全な，難しい問題であるからである．本アイデアは，両面から計算コストを削減するためのものである．

(3)距離の遠い頂点間の依存関係を無視：

非連続コードクローンの誤検出を削減するために，ソースコード上で遠く離れた頂点間には依存関係を引かない．本アイデアにより，検出する価値のない重複コードの検出を抑え，検出の精度をあげることができる．また副次的な効果として，プログラムスライスの範囲が狭まるため，計算コストの削減による検出時間の短縮も見込まれる．

(4)増分的な検出：

増分的なコードクローン検出とは，検出結果とその中間生成物をデータベースなどに保管し，次回以降の検出時に利用する検出手法である．検出対象ファイルが前回の検出時から更新されていない場合は，そのファイル自身は解析されず，ファイルに関する必要な情報はデータベースから取得される．このような枠組みを用いることによって，同じファイル集合から何度も検出を行う場合は，2 回目以降の検出時間を大幅に短縮することができる．

４．研究成果

4つの提案手法が，コードクローン検出能力の向上，および計算コストの削減に対して有効であることを実験により確かめた．また，提案内容を他の検出ツールと比較した．その結果，提案手法は，必ずしも他の手法に比べて多くのコードクローンを検出できるとは限らないが，他の手法では検出できていないコードクローンを検出できてきていることがわかった．

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６件）

[1]堀田圭佑，肥後芳樹，楠本真二，″プログラム依存グラフを用いたコードクローンに対するテンプレートメソッドパターン適用支援手法″，電子情報通信学会論文誌 D，2012 年．（to appear）

[2] Keisuke Hotta, Yui Sasaki, Yukiko Sano, Yoshiki Higo and Shinji Kusumoto, ″An Empirical Study on the Impact of Duplicate Code″, Advances in Software Engineering, 2012. (to appear)

[3] 肥後芳樹，植田泰士，西野稔，楠本真二，″プログラム依存グラフを用いた増分的なコードクローン検出″，情報処理学会論文誌，Vol.53，No.2，pp.601-611，2012 年 2 月．

[4] 堀田圭佑，佐野由紀子，肥後芳樹，楠本真二，″修正頻度の比較に基づくソフトウェア修正作業量に対する重複コードの影響に関する調査″，情報処理学会論文誌，Vol.52，No.9，pp.2788-2798，2011 年 9 月．

[5] 肥後芳樹，楠本真二，″プログラム依存グラフを用いたコードクローン検出法の改良と評価″，情報処理学会論文誌，Vol.51，No.12，pp.2149-2168，2010 年 12 月．

[6] 肥後芳樹，宮崎宏海，楠本真二，井上克郎，″グラフマイニングアルゴリズムを用いたギャップを含むコードクローン情報の生成″，電子情報通信学会論文誌 D，Vol.J93-D，No.9，pp.1727-1735，2010 年 9 月．

〔学会発表〕（計１４件）

[1] Keisuke Hotta, Yoshiki Higo, and Shinji Kusumoto, ″Identifying, Tailoring, and Suggesting Form Template Method Refactoring Opportunities with Program Dependence Graph″, Proc. of the 16th European Conference on Software Maintenance and Reengineering (CSMR2012), pp.53-62, Szeged, Hungary, March 27-30, 2012.

[2] 石原知也，堀田圭佑，肥後芳樹，井垣宏，楠本真二，″大規模ソフトウェア群に対するメソッド単位のコードクローン検出，″電子情報通信学会技術研究報告 ss2011-xx，Vol.111，No.481，pp.31-36，てんぷす那覇，2012 年 3 月 13-14 日

[3] 村上寛明，堀田圭佑，肥後芳樹，井垣宏，楠本真二，″ソースコード中の繰り返し部分に着目したコードクローン検出手法の提案″，電子情報通信学会技術研究報告 ss2011-xx，vol.111，No.481，pp.25-30，てんぷす那覇，2012 年 3 月 13-14 日

[4] Yui Sasaki, Keisuke Hotta, Yoshiki Higo, and Shinji Kusumoto, ″Is Duplicate

Code Good or Bad? An Empirical Study with Multiple Investigation Methods and Multiple Detection Tools", Proc. of the 22nd International Symposium on Software Reliability Engineering (ISSRE2011), Hiroshima, Japan, November 29 - December 2, 2011.

[5] 堀田圭佑, 肥後芳樹, 楠本真二, "プログラム依存グラフを用いたテンプレートメソッドパターン適用によるリファクタリング支援手法の提案", ソフトウェア信頼性研究会 第7回ワークショップ論文集, 広島県休暇村大久野島,2011年11月27-28日

[6] Yoshiki Higo, Yasushi Ueda, Minoru Nishino, and Shinji Kusumoto, "Incremental Code Clone Detection: A PDG-based Approach", Proc. of the 18th Working Conference on Reverse Engineering (WCRE2011), pp.3-12, Limerick, Ireland, Octobar 17-20, 2011.

[7] 肥後芳樹, 植田泰士, 西野稔, 楠本真二, "プログラム依存グラフを用いた増分的なコードクローン検出", ソフトウェアエンジニアリングシンポジウム2011,東京女子大学, 2011年9月12-14日

[8] 佐々木唯, 堀田圭佑, 肥後芳樹, 楠本真二, "ソフトウェア保守におけるコードクローンの影響に関する調査方法の比較", 電子情報通信学会技術研究報告 SS2011-17, Vol.111, No.168, pp.25-30,北海道情報大学, 2011年7月28-30日

[9] 堀田圭佑, 肥後芳樹, 楠本真二, "プログラム依存グラフを用いたTemplate Methodパターン適用によるコードクローン集約支援", 情報処理学会研究報告 2011-SE-171, No.14, pp.1-8, 科学会館,2011年3月14-15日

[10] Yoshiki Higo, and Shinji Kusumoto, "Code Clone Detection on Specialized PDGs with Heuristics", Proc. of the 15th European Conference on Software Maintenance and Reengineering (CSMR2011), pp.75-84, Oldenburg, Germany, March 1-4, 2011.

[11] 肥後芳樹, 楠本真二, "複数のメソッドにまたがって存在するコードクローンの検出に向けて", 電子情報通信学会技術研究報告 SS2010-50, Vol.110, No.336, pp.67-72, 伊香保温泉ホテル天坊,2010年12月14-15日

[12] 兼光智子, 肥後芳樹, 楠本真二, "プログラム依存グラフを用いたリファクタリング候補の特定と可視化", 電子情報通信学会技術研究報告 SS2010-49, Vol.110, No.336, pp.61-66,伊香保温泉ホテル天坊, 2010年12月14-15日

[13] Keisuke Hotta, Yukiko Sano, Yoshiki Higo, and Shinji Kusumoto, "Is Duplicate Code More Frequently Modified Than Non-duplicate Code in Software Evolution?: An Empirical Study on Open Source Software", Proc. of the 11th International Workshop on Principles of Software Evolution (IWPSE-EVOL2010), pp.73-82, Antwerp, Belgium, September 20-21, 2010.

[14] 肥後芳樹, 楠本真二, "コードクローン検出に必要な計算コストの削減を目的としたプログラム依存グラフ頂点集約法の提案", ソフトウェアエンジニアリング最前線2010(ソフトウェアエンジニアリングシンポ

ジウム 2010 予稿集), pp. 127-134, 東洋大学白山キャンパス, 2010 年 8 月 31 日-9 月 1 日

６．研究組織

(1)研究代表者
肥後　芳樹（HIGO YOSHIKI）
大阪大学・大学院情報科学研究科・助教
研究者番号：７０４５２４１４