

令和 5 年 6 月 27 日現在

機関番号：14303

研究種目：若手研究

研究期間：2019～2022

課題番号：19K20240

研究課題名（和文）類似ソースコード検索を用いたテストコード自動生成環境の構築

研究課題名（英文）Building an environment for automated test code generation using a similar source code search technique

研究代表者

崔 恩瀨（Choi, Eunjong）

京都工芸繊維大学・情報工学・人間科学系・助教

研究者番号：90755943

交付決定額（研究期間全体）：（直接経費） 2,100,000円

研究成果の概要（和文）：本研究では、任意のテストメソッドを実行可能性を保ちつつ自動的に移植できる手法および条件を考察した。さらに、その考察に基づいてJavaで開発されたオープンソースソフトウェアを調査し、テストコードの再利用元として使用できるリポジトリが1,862件存在することやほとんどのテストコードはテスト対象のコードにたかだか2つの依存関係しか持たないことを明らかにした。また、調査結果に基づいて類似ソースコード検索を用いたテストコード自動生成環境の構築し、本環境の有効性を評価した結果、本環境を利用した結果、本環境を用いて作成したテストコードはテストするの数が少なくして品質が高いことがわかった。

研究成果の学術的意義や社会的意義

本環境を用いることで、開発者はテストコードを再利用し、テストコードを生成することで、開発者が理解しやすいかつ高いカバレッジを持つテストコードを生成できると期待される。また、本環境を使用することで開発者がより迅速にテストを実施できる。最後に、テストコードの再利用と自動生成により、ソフトウェアの品質向上が期待できる。

研究成果の概要（英文）：This study presents a method and condition for automatically transplanting any given test method while preserving its executability. Furthermore, this study investigates open-source software projects developed in Java. It reveals several relevant findings, including 1,862 projects that can serve as sources for reusing test code, and the majority of test code in the investigated projects have at most two dependencies on the code under test. Based on these findings, this study developed an environment for automated test code generation using a similar source code search technique. Finally, the effectiveness of this environment was evaluated, and its results show that the test code generated using this environment exhibited a smaller number of test failures and higher overall quality.

研究分野：ソフトウェア工学

キーワード：類似ソースコード テストコード自動生成 ソフトウェア品質

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

ユーザのニーズの変化やハードウェアの進化に応じるために、ソフトウェアを短期間で開発させ、一定の品質を確保しつつリリースしていく必要がある。従って、ソフトウェアの不具合がユーザの元へリリースされることを防ぐためにソフトウェアのリリースにあたって、ソフトウェアが仕様書通り実装されているかを確認するためのソフトウェアテストが不可欠である。しかし、開発者がテストコードを記述するためには、ソフトウェアの仕様書およびテスト対象のソースコードを正しく理解する必要があるため、開発者が手作業でテストコードを記述して作成することはソフトウェア開発の全体の時間およびコストの増加を招く可能性がある。既存手法により生成されたテストコードは現実なシナリオに従っていないので、そのソースコードを開発者が理解することは難しい。そのため、開発者は既存手法によって生成されたテストコードに対する保守変更作業を行うことは困難であるという問題が指摘されている[1]。

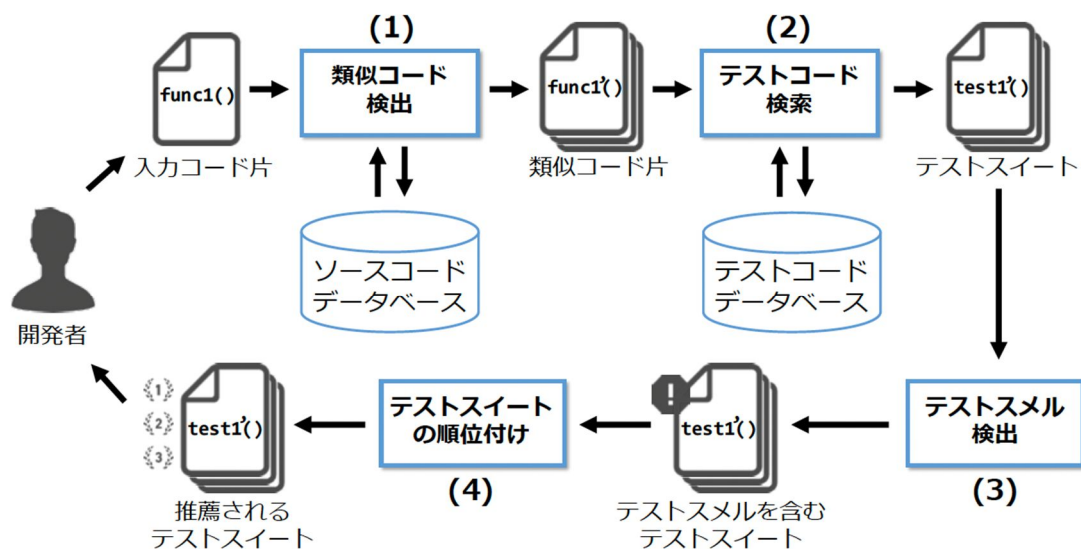
2. 研究の目的

本研究の主な目的は既存のソースコードを再利用することによって、より理解やすいかつ高いカバレッジ(テスト対象のコードを網羅する程度)を持つテストコードを自動生成する環境を開発することである。また、本環境によって類似ソースコードを再利用することで生成されたテストコードは開発者が読みやすくかつ品質が高いものである必要がある。従って、本研究の有効性を確認することも目的である。

3. 研究の方法

本研究では、まず、任意のテストメソッドを実行可能性を保ちつつ自動的に移植できる手法および条件を構想した。さらに、その考案に基づいてGitHub上のJavaで開発されたオープンソースソフトウェアにおける既存テストコードの移植可能性について調査した。

また、調査結果に基づいて類似ソースコード検索を用いたテストコード自動生成環境の構築した。本研究では、まず、既存の類似コード検出ツールを用いて開発者から与えられた入力コードに対する類似コードを検出する。また、検出された複数の類似コードに対するテストスイートをテストコードデータベース内から検索する。次に、検出された複数のテストスイートをテストスマイル検出ツールにかけ、各々テストスイートに含まれるテストスマイルを検出する。その後、類似コード検出ツールから取得した類似コードと入力コードの類似度に基づいてテストスイートを提示する。さらに、品質の高いテストコードを推薦できるようにしてテストスマイル情報に基づいて推薦優先順位を並び替える。最後に、被験者実験を通して本環境の有効性を評価した。



4. 研究成果

本研究では、まず、任意のテストメソッドを実行可能性を保ちつつ自動的に移植できる手法および条件を構想した。その考案に基づいた調査の結果、テストコードの再利用元として使用できるリポジトリが 1,862 件存在すること、ほとんどのテストコードはテスト対象のコードにたかだか 2 つの依存関係しか持たないことや、この依存関係を保つ移植によって本来必要なテストメソッドの平均 83% が生成できる可能性があることを明らかにした。この調査より、既存のテストコードを再利用する自動生成手法が現実的かつ有益であることが示された。

調査結果に基づいて類似ソースコード検索を用いたテストコード自動生成環境の構築した。また、本環境の有効性を評価した結果、本環境を利用した結果、条件分岐が多いプログラムのテストコードを作成する際にコードのカバレッジが向上したことや、本環境を用いて作成したテストコードはテストスメル数が少なく、品質が高いことがわかった。

5. 主な発表論文等

〔雑誌論文〕 計10件（うち査読付論文 8件 / うち国際共著 0件 / うちオープンアクセス 2件）

1. 著者名 西浦 生成、水野 修、崔 恩滯	4. 巻 62
2. 論文標題 Javaテストコードの再利用による自動生成に向けた移植可能なテストメソッドの調査	5. 発行年 2021年
3. 雑誌名 情報処理学会論文誌	6. 最初と最後の頁 1019 ~ 1028
掲載論文のDOI (デジタルオブジェクト識別子) 10.20729/00210552	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 西浦 生成、渡辺 大輝、水野 修、崔 恩滯	4. 巻 62
2. 論文標題 組合せテストにおける実行順序に起因する非決定的不具合誘発要因特定法の提案	5. 発行年 2021年
3. 雑誌名 情報処理学会論文誌	6. 最初と最後の頁 1008 ~ 1018
掲載論文のDOI (デジタルオブジェクト識別子) 10.20729/00210551	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 藤原 裕士、崔 恩滯、吉田 則裕、井上 克郎	4. 巻 J104-D
2. 論文標題 深層学習を用いたソースコード分類のための動的な学習用データセット改善手法の提案	5. 発行年 2021年
3. 雑誌名 電子情報通信学会論文誌D 情報・システム	6. 最初と最後の頁 275 ~ 284
掲載論文のDOI (デジタルオブジェクト識別子) 10.14923/transinfj.2020PDP0005	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 藤原 裕士、崔 恩滯、吉田 則裕、井上 克郎	4. 巻 J104-D
2. 論文標題 深層学習を用いたソースコード分類手法の比較調査	5. 発行年 2021年
3. 雑誌名 電子情報通信学会論文誌D 情報・システム	6. 最初と最後の頁 622 ~ 635
掲載論文のDOI (デジタルオブジェクト識別子) 10.14923/transinfj.2020JDP7068	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 徳井 翔梧、吉田 則裕、崔 恩瀨、井上 克郎	4. 巻 38
2. 論文標題 Cross-Polytope LSH を用いた コードクローン検出のためのパラメータ決定手法	5. 発行年 2021年
3. 雑誌名 コンピュータ ソフトウェア	6. 最初と最後の頁 4_60~4_82
掲載論文のDOI (デジタルオブジェクト識別子) 10.11309/jssst.38.4_60	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 瀬村 雄一、吉田 則裕、崔 恩瀨、井上 克郎	4. 巻 103-D
2. 論文標題 多様なプログラミング言語に対応可能なコードクローン検出ツールCCFinderSW	5. 発行年 2021年
3. 雑誌名 電子情報通信学会論文誌D	6. 最初と最後の頁 215-227
掲載論文のDOI (デジタルオブジェクト識別子) 10.14923/transinfj.2019PDP0025	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Karim Md. Rejaul, Ihara Akinori, Choi Eunjong, Iida Hajimu	4. 巻 31
2. 論文標題 Identifying and predicting key features to support bug reporting	5. 発行年 2019年
3. 雑誌名 Journal of Software: Evolution and Process	6. 最初と最後の頁 1-26
掲載論文のDOI (デジタルオブジェクト識別子) 10.1002/smr.2184	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 石津 卓也、吉田 則裕、崔 恩瀨、井上 克郎	4. 巻 60
2. 論文標題 コードクローンのリファクタリング可能性に基づいた削減可能ソースコード量の分析	5. 発行年 2019年
3. 雑誌名 情報処理学会論文誌	6. 最初と最後の頁 1051-1062
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 MON Khine Yin、KONDO Masanari、CHOI Eunjong、MIZUNO Osamu	4. 巻 E106.D
2. 論文標題 Commit-Based Class-Level Defect Prediction for Python Projects	5. 発行年 2023年
3. 雑誌名 IEICE Transactions on Information and Systems	6. 最初と最後の頁 157 ~ 165
掲載論文のDOI (デジタルオブジェクト識別子) 10.1587/transinf.2022MPP0003	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 崔 恩瀨、藤原 裕士、吉田 則裕、水野 修	4. 巻 39
2. 論文標題 コードクローン検索手法の調査	5. 発行年 2022年
3. 雑誌名 コンピュータ ソフトウェア	6. 最初と最後の頁 3_47 ~ 3_59
掲載論文のDOI (デジタルオブジェクト識別子) 10.11309/jssst.39.3_47	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計16件 (うち招待講演 2件 / うち国際学会 5件)

1. 発表者名 中森 陸斗
2. 発表標題 リモートワークにおけるソフトウェア開発者間のコミュニケーション方法の調査
3. 学会等名 日本ソフトウェア科学会 第28回ソフトウェア工学の基礎ワークショップ (FOSE 2021)
4. 発表年 2021年

1. 発表者名 福家 範浩
2. 発表標題 深層学習を用いたコードクローン検出器のベンチマーク間精度調査
3. 学会等名 第210回ソフトウェア工学研究発表会
4. 発表年 2022年

1. 発表者名 近藤 将成
2. 発表標題 リモートワークにおけるソフトウェア開発者間のコミュニケーション方法の調査
3. 学会等名 ソフトウェアエンジニアリングシンポジウム2020
4. 発表年 2020年

1. 発表者名 福家 範浩
2. 発表標題 深層学習を用いたコードクローン検出器の汎化性能に関する調査
3. 学会等名 情第207回ソフトウェア工学研究発表会
4. 発表年 2020年

1. 発表者名 倉地亮介、崔 恩潯、飯田元
2. 発表標題 類似コード検出ツールを用いたテストコード再利用に向けた調査
3. 学会等名 第26回 ソフトウェア工学の基礎ワークショップ (国際学会)
4. 発表年 2020年

1. 発表者名 Norihiko Yoshida, Seiya Numata, Eunjong Choi, Katsuro Inoue
2. 発表標題 Proactive Clone Recommendation System for Extract Method Refactoring
3. 学会等名 3rd IEEE/ACM International Workshop on Refactoring (IWor 2019) (国際学会)
4. 発表年 2019年

1. 発表者名 Shogo Tokui, Norihiro Yoshida, Eunjong Choi, Katsuro Inoue
2. 発表標題 Clone Notifier: Developing and Improving the System to Notify Changes of Code Clones
3. 学会等名 the 27th IEEE International Conference on Software Analysis, Evolution and Reengineering (国際学会)
4. 発表年 2020年

1. 発表者名 Md. Rejaul Karim, Akinori Ihara, Eunjong Choi, Hajimu Iida
2. 発表標題 Identifying and predicting key features to support bug reporting
3. 学会等名 35th IEEE International Conference on Software Maintenance and Evolution (ICSME) (国際学会)
4. 発表年 2019年

1. 発表者名 Hirotaka Honda, Shogo Tokui, Kazuki Yokoi, Eunjong Choi, Norihiro Yoshida, Katsuro Inoue
2. 発表標題 CCEvovis: A Clone Evolution Visualization System for Software Maintenance
3. 学会等名 the IEEE/ACM 27th International Conference on Program Comprehension (ICPC 2019) (招待講演)
4. 発表年 2020年

1. 発表者名 崔 恩滯
2. 発表標題 プログラム解析技術を用いたソフトウェア開発支援
3. 学会等名 令和元年 電気関係学会学会関西連合大会 (招待講演)
4. 発表年 2020年

1. 発表者名 石津 卓也, 吉田 則裕, 崔 恩瀨, 井上 克郎
2. 発表標題 プロジェクト間クローンに対する変更傾向の調査
3. 学会等名 情報処理学会研究報告
4. 発表年 2019年

1. 発表者名 鐘ヶ江由佳, 崔 恩瀨, 飯田元
2. 発表標題 OSSプロジェクトにおけるCIツールの変更の影響調査に向けて
3. 学会等名 第26回 ソフトウェア工学の基礎ワークショップ
4. 発表年 2019年

1. 発表者名 石津 卓也, 吉田 則裕, 崔 恩瀨, 井上 克郎
2. 発表標題 プロジェクト間クローンに対する変更傾向の調査
3. 学会等名 ソフトウェアエンジニアリングシンポジウム2019
4. 発表年 2019年

1. 発表者名 渡邊紘矢, 崔恩瀨, 水野修
2. 発表標題 ソースコードコメントに着目した不確かさとソフトウェア品質の関係調査
3. 学会等名 ソフトウェア・シンポジウム2022
4. 発表年 2022年

1. 発表者名 中森陸斗、崔恩瀨、吉田則裕、水野修
2. 発表標題 テストケース生成ツールを用いたバグ限局ツールAutoSBFLの提案
3. 学会等名 ソフトウェア工学の基礎 28 日本ソフトウェア科学会ソフトウェア工学の基礎研究会 F O S E 2 0 2 1
4. 発表年 2021年

1. 発表者名 Zhu, W, Yoshida, N, Kamiya, T, Choi, E, and Takada, H
2. 発表標題 MSCCD: Grammar Pluggable Clone Detection Based on ANTLR Parser Generation
3. 学会等名 International Conference on Program Comprehension (ICPC) 2022 (国際学会)
4. 発表年 2022年

〔図書〕 計1件

1. 著者名 Katsuro Inoue, Chanchal K. Roy 他	4. 発行年 2021年
2. 出版社 Springer	5. 総ページ数 246
3. 書名 Code Clone Analysis: Research, Tools, and Practices	

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関