

令和 2 年 5 月 15 日現在

機関番号：12601

研究種目：若手研究

研究期間：2018～2019

課題番号：18K18088

研究課題名（和文）コードの理解と教育におけるGitHubのプルリクエスト情報の有効性の検証

研究課題名（英文）Investigations on applications of GitHub pull requests for code understanding support and programming education

研究代表者

矢谷 浩司 (Yatani, Koji)

東京大学・大学院工学系研究科（工学部）・准教授

研究者番号：60741353

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：現在のソフトウェア開発の主流プラットフォームであるGitHubでは、コードの変更に
関する情報を一元的に管理する機能を利用した開発が活発に行われている。本研究ではGitHub上の情報を利活用
した2つの新たなシステム、CodeGlassとRealCodeを構築し、CodeGlassにより可視化されるプルリクエストがコ
ードの理解を支援する可能性があること、また公開されているコード変更を利用することで実世界でのプログラ
ミング体験を反映したプログラミング演習課題をRealCodeにより生成することが出来ることが確認された。

研究成果の学術的意義や社会的意義

本研究ではソフトウェア開発管理で広く用いられているGitHub上の情報を活用し、プログラミングにおける活動
を支援する技術の開発、その効果の評価を行いました。この研究により、GitHub上での情報を活用することで、
プログラマのコード理解を支援出来るだけでなく、教科書などでは見られない実践的なプログラミングスキルを
養える演習課題を生成できる可能性があることを示すことができました。

研究成果の概要（英文）：Modern software development actively utilizes Github for project management
including maintaining changes on code. In this research, we explore how interactive systems can
exploit various information on Github for code understanding and programming education through two
systems, CodeGlass and RealCode. We found that pull requests visualized through CodeGlass can
support developers' understanding on code. RealCode can repurpose code changes on Github to offer
programming exercises that otherwise would not be seen in standard textbooks and teaching materials,
complementing programming education at school.

研究分野：ヒューマンコンピュータインタラクション

キーワード：ヒューマンコンピュータインタラクション プログラミング教育 プログラミング支援

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

現在のソフトウェア開発の主流プラットフォームである GitHub では、プルリクエストという、コードの変更に関する情報を一元的に管理する機能を利用したコード開発が活発に行われている。申請者の研究グループが行った開発者とのインタビューでは、開発者がプルリクエストをコードの理解する際に利用している、またプルリクエスト内のコード変更内容を理解することで開発能力の向上につなげていることが示唆された。しかし、プルリクエストがコード開発や教育にどの程度有効であるかは客観的に検証されていない。そこで、申請者の支援システムの設計に関する経験を生かし、プルリクエストを活用した 2 つの新たなシステムの構築を通し、プルリクエストがコード開発や教育に対してどのように有効であるかを客観的な指標を用いて実証することを目指す。

2. 研究の目的

GitHub におけるプルリクエストは今までも様々な研究が行われている。しかしこれらの多くは、どのようなプルリクエストが投稿されているのかの調査と、プルリクエストの記述を支援するシステムの研究、に大別される。プルリクエストの重要性は、変更が行われたコードが正当性を持つものであることを判断することにおいて確認されている。一方、その変更が承認された後ではあまり積極的に利用されない。

本研究は特にコード理解とプログラミング教育におけるプルリクエストの効用を客観的な指標(例えば、コードに関する理解内容の項目数やそれにかかる時間など)によって検証する。また、既存のプルリクエストは上記のような利用を想定されていないため、そのまま開発するシステムへと流用することはできない。特に、1) ユーザが理解を深めたいと思うコード部分に関連する過去のプルリクエストのみを抽出するアルゴリズム、2) プログラミングの課題として適切なプルリクエストを抽出するアルゴリズム、が必要となる。これらの技術的ブレークスルーと、プルリクエストのコード開発や教育に対する客観的有効性の検証が、本研究の学術的独自性をなすものである。

3. 研究の方法

GitHub におけるプルリクエストは、コードの変更情報のみならず、その変更に関する記述、また他の開発者からのコメントやレビューを含んでおり、コードの変更を管理しやすくするための機能である。本研究では 2 つのプロジェクトを通して新たなインタフェースを構築することで、プルリクエストの客観的有効性を検証する。

[R1] プルリクエストを用いたコードの理解を支援するシステム

コードの理解は様々なプログラミング活動において重要である。既存のシステムではドキュメンテーションと呼ばれる、コードに関する情報を記載したドキュメントを利用することで、コードの理解を支援するものがある。しかしドキュメンテーションの問題点として、開発者が必ずしもその記録を残さない、コードの更新が行われてもドキュメンテーションへの反映が行われないうちがある。またドキュメンテーションは比較的大きな単位(ファイルやクラス、関数など)で行われることが多い一方、現実的な開発環境においては、コードの一部(例えば、if 節の内部)に関する情報を得たいことが多い。

そこで本研究においてはユーザが興味を持つコード行に関するプルリクエストを抽出し、その情報を提示することによりコードの理解を支援するシステムを構築する。本システムにおいてはユーザが選択したコード行に関するプルリクエストを取得する必要がある。GitHub においてコードの変更は commit によりその差分が管理されている。しかし、この情報だけではリファクタリングなどにより、コードが移動・変更された場合や、commit が複数の変更を含む場合には、対応関係が明らかでないために、過去のバージョンに遡ることが不可能となる。そこで本研究ではそのような過去のコード変更を遡ることのできるアルゴリズムを開発する。さらに構築したシステムをユーザ実験により評価し、どのように開発者のコード理解を支援できるかを検証する。

[R2] プルリクエストの情報から実世界のコード例を用いたプログラミング教育システム

プログラミング教育において、学習者に適切な課題を提示することは教育効果を高める点において重要である。しかし、既存の教科書や教育システムにおいては、現実のプロジェクトにおいて現れるようなバグの修正や機能の実装などを体験することは難しい。GitHub におけるプルリクエストには変更に関する理由や背景だけでなく、実際にどのような変更が行われたか、の情報が含まれている。従って、この情報を用いることでプログラミング課題の問題と解答とできる可能性がある。

一方、必ずしもすべてのプルリクエストが学習教材として適切であるとは限らない。例えば、関数や変数の単純な名前変更、コメントの追記なども中には存在するが、それらは学習教材としては不適切である。またリファクタリングのようにソースコードの構造を大きく変更するものも変更の自由が大きいため、学習教材として利用するのは難しい。

本研究ではコードの変更量やその複雑さに加えて、上記のようなプルリクエストに付属す

る情報を利用することにより、プログラミング課題に適切なプルリクエストを抽出する手法の構築を目指す。また抽出されたプルリクエストにおいても、含まれている変更をそのまま利用するだけでは、プログラミング課題に不向きであることがある。これらの処理を経た上で、ユーザに課題を提示し、その課題の適切さや有用性をユーザ実験により検証する。その後、生成されたプログラミング課題を用いた教育プログラムを設計し、中級者のプログラマーなどに受講してもらい、その教育的効果をテスト等により検証する。

4. 研究成果

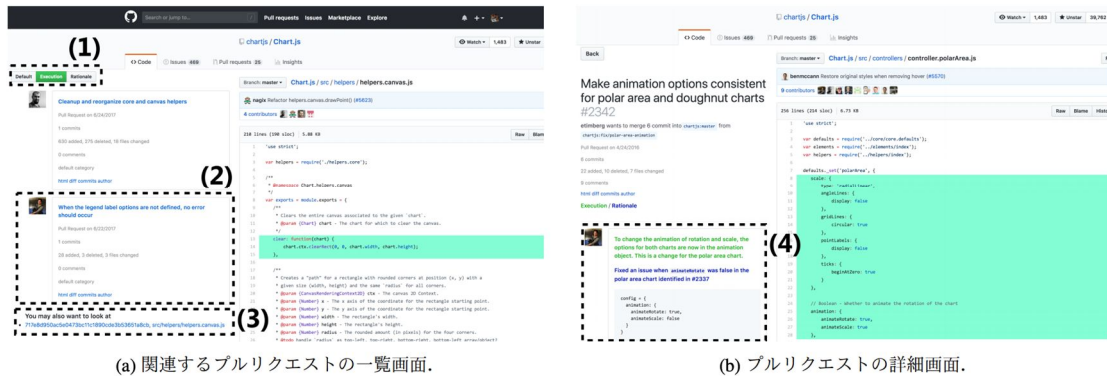
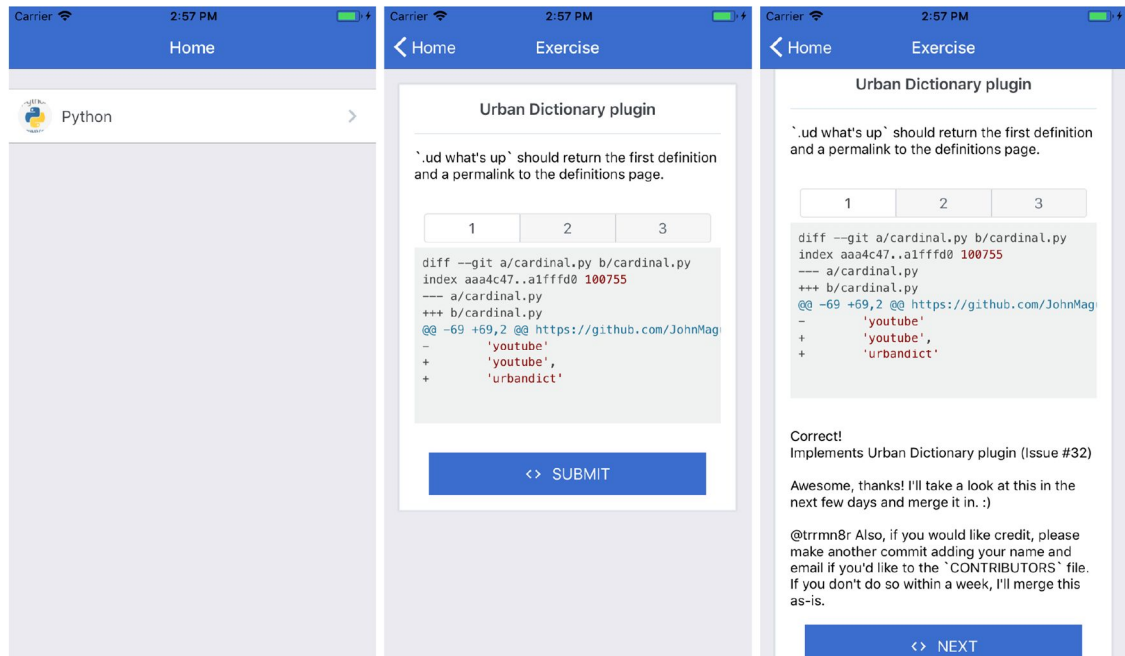


図 1. CodeGlass のインターフェースは、ユーザが選択したコード断片（緑にハイライトされたコード）に関連する過去のプルリクエストの一覧を表示する。抽出されたプルリクエストの表示順は、時系列順、実装内容に関する情報が多い順、開発背景に関する情報が多い順に並べることができる (1)。各プルリクエストにはタイトル、コミット数、コードの追加および削除行数、コメント数、そして関連する各ウェブページへのリンクが表示される (2)。さらに、選択されたコード断片と一致する可能性があるコード断片が過去のバージョンの 1 つで複数ある場合、そのバージョンにおけるソースコードへのリンクが表示される (3)。プルリクエストのタイトルをクリックすると、その詳細情報を確認できる。詳細画面では、プルリクエストの情報に加えて、レビューコメントも表示される。さらに、実装内容に分類される文章は緑色で、開発背景に分類される文章は青色で表示される (4)。

[R1]に対応する研究成果として、プルリクエストを用いたコードの理解を支援するシステムとして、ユーザが選択したコード断片に関連する過去のプルリクエストを抽出し、ユーザに提供する CodeGlass を開発した。CodeGlass はユーザが選択したコード断片に関連する過去のプルリクエストを抽出し、ユーザに提供する。さらに、プルリクエストの説明文を解析し、実装内容や開発背景に関する文章をインターフェース上で強調して表示することが可能となっている。

本システムにおいては既存のアルゴリズムにさらなる改良を加えたことで、選択されたコード断片が過去のバージョンにおいて分裂していた場合にも、CodeGlass は関連する過去のプルリクエストを提供することができるような新しい手法を実現した。さらには実装内容や開発背景に関する情報は特に開発者にとっては有用であると考えられるため、それらを抽出し強調して表示できるようなインタフェースを実現した。

ユーザ実験では、学生による CodeGlass の定量的評価により、コード断片の理解のための情報収集において、CodeGlass が有用であることが分かった。特に「なぜこのような実装になっているのか」、「どのような経緯でこのような実装になっているか」ということが CodeGlass ではより明確に支援できることが確認された。さらにプログラマーを対象とした定性的評価を行った結果、専門的用途における CodeGlass の有用性が明らかとなった。



(a) プログラミング言語の選択画面.

(b) 演習問題の画面.

(c) 演習問題の解答後の画面.

図 2. RealCode のインタフェース (モバイル版). RealCode では, プログラミング演習問題とその解答の説明文を, イシューとプルリクエストの説明文から生成する. さらに解答となるソースコードをプルリクエストのコード変更から生成する. この図では 3 択問題 (問題文に該当するコード変更として 3 つ提示 (b ではその内 1 つが提示されている) し, ユーザは正しいと思うものを選択する.

また, [R2] に対応する実世界のコード例を用いたプログラミング教育システムとして, プルリクエストに存在するイシューとコード変更を用いることで, プログラミング課題を自動で生成する RealCode というシステムを構築した.

このシステムにおいてはまず, GitHub の膨大なイシューのデータから, プログラミング演習問題への転用に適さないプルリクエストに付随するイシューを分類するために, 我々はヒューリスティックにより事前処理と決定木によるイシューの分類を行った. 決定木の精度を評価した結果, F 値の平均が 0.79 の精度で転用に適さないイシューを取り除くことができた. また, 決定木による分類においては, 変更行数が重要な特徴量であることが明らかとなった. さらに, 機械学習を用いて課題内容を分類する手法を実装し検証した所, 例えば UX (ユーザ体験) などというキーワードで表されるようなプログラミング課題を生成することが可能であることが確認された.

さらにはユーザ実験により RealCode により出題される演習課題を評価してもらった結果, 教科書等ではカバーしきれない問題を RealCode により提供できる可能性があることが確認できた.

当初予定では, 研究期間終了直前に RealCode を公開し, より多くのユーザに試用していただくとともに, さらなるデータ拡充を行う予定であったが, コロナウイルスの影響により, 公開に耐えうる十分な準備ができないまま研究期間終了をせざるを得ないこととなった. しかしながら, 研究期間終了後も準備を進める予定であり, 最終的には RealCode の社会実装を目指していきたい.

以上の研究成果は, 国内の会議, 研究会等で発表され, すべての年度を合わせて計 6 篇の論文として発表された.

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計7件（うち招待講演 1件 / うち国際学会 0件）

1. 発表者名 松井 秀憲, 矢谷 浩司.
2. 発表標題 GitHub上のコード変更を利用したプログラミング演習問題の実用性の評価
3. 学会等名 情報処理学会HCI研究会
4. 発表年 2020年

1. 発表者名 柴藤大介, 有園拓也, 宮崎章太, 矢谷浩司.
2. 発表標題 CodeGlass: GitHubのプルリクエストを活用したコード断片のインタラクティブな調査支援システム
3. 学会等名 インタラクシオン2019
4. 発表年 2019年

1. 発表者名 柴藤大介, 有園拓也, 宮崎章太, 矢谷浩司.
2. 発表標題 CodeGlass: GitHubのプルリクエストを活用したコード断片のインタラクティブな調査支援システム
3. 学会等名 インタラクシオン2019
4. 発表年 2019年

1. 発表者名 柴藤大介, 矢谷浩司.
2. 発表標題 GitHubのプルリクエストを用いたプログラミング課題自動生成システムの実現可能性に関する検討
3. 学会等名 第80回情報処理学会全国大会
4. 発表年 2018年

1. 発表者名 柴藤大介, 三島潤平, 矢谷浩司.
2. 発表標題 GitHub のプルリクエストを用いたプログラミング課題自動生成システム
3. 学会等名 第4回SIGPX勉強会
4. 発表年 2018年

1. 発表者名 三島潤平, 柴藤大介, 矢谷浩司.
2. 発表標題 DiDA: GitHub上のコード変更と開発履歴に関するデータセット
3. 学会等名 第4回SIGPX勉強会
4. 発表年 2018年

1. 発表者名 Antony Chacon and Daisuke Shibato
2. 発表標題 Interactive Sensing, Wearable, and Tangible User Interfaces and Support for Programming Activities
3. 学会等名 Invited talk at National Taiwan University (招待講演)
4. 発表年 2018年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

<p>GitHub上のコード変更を利用したプログラミング演習問題の実用性の評価 https://iis-lab.org/research/programming-exercises/ CodeGlass https://iis-lab.org/research/codeglass/ GitHub のプルリクエストを用いたプログラミング課題自動生成システムの実現可能性に関する検討 https://iis-lab.org/research/codequiz/ DiDA: GitHub 上のコード更と開発履歴に関するデータセット https://iis-lab.org/research/dida/</p>

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
--	---------------------------	-----------------------	----