

コードクローンの利用過程における 不具合混入原因を分析するための ツールの提案と評価

オープンソースソフトウェア工学研究室

60191008 大石 聖也

コードクローン

- ソースコード中に存在する, 互いに「一致」または「類似」するコード片
- ソフトウェア品質を低下させる要因
 - (例) 変更する際, コードクローン全体を変更しないといけない
 - コストがかかる
 - 変更忘れから, 新たな不具合が発生
- ソフトウェア品質を低下させる要因として, 様々な研究がされてきた^[1]

[1] 肥後 芳樹, 楠本 真二, 井上 克郎. コードクローン検出とその関連技術. 電子情報通信学会論文誌D, Vol.J91-D, No.6, pp.1465-1481, 2008年6月.

既存研究

- コードクローン分析に関する研究
 - バグ伝播に関与しやすいコードクロンの存在を発見[2]
 - ソースコードの理解が再利用に大きく関与する可能性を報告[3]
- コードクローン追跡に関する研究
 - 生存期間が長く、複数回修正されたコードクローンを対処すべきと主張[4]

[2] Manishankar Mondal, Chanchal K. Roy, Kevin A. Schneider, Bug Propagation through Code Cloning: An Empirical Study, IEEE International Conference on Software Maintenance and Evolution, pp. 227-237, 2017.

[3] 森脇匠哉, 堀田圭佑, 井垣宏, 楠本真二, 井上克郎, 複数のプロジェクトを対象としたクロンの系譜にもとづくソースコード再利用分析手法の提案, 電子情報通信学会技術研究報告, 第114巻, pp. 61-66, 2015.

[4] 堀田 圭佑, 肥後 芳樹, 楠本 真二. CRDを用いたコードクロンの生存期間と修正回数に関する調査. 情報処理学会論文誌 Vol.55 No.2 947-958 (Feb. 2014).

既存研究の問題点

- 既存研究で得られた成果
 - 不具合に関わるコードクローンの特性
 - コードクローンを利用する開発者の特徴
- 不具合が混入する要因として開発者の特徴を考慮した分析を行った事例はない

不具合を混入する開発者の特徴が分かっていない

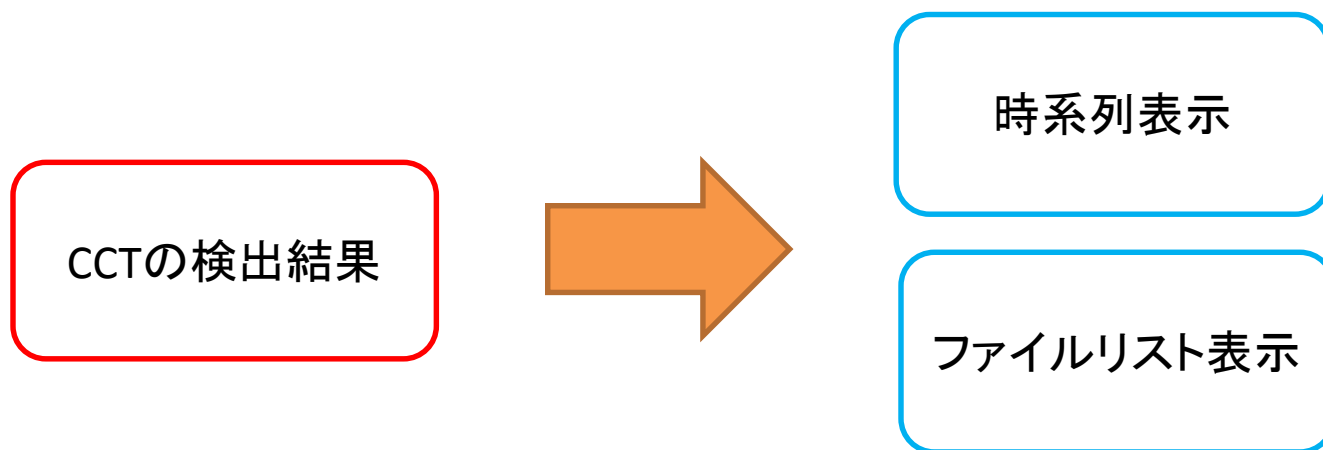
CCT (Code Clone Tracer) [5]

- コードクローンの利用過程における人的影響の調査を目的とした, コードクローン追跡ツール
 - 開発者・不具合・コードクローンの紐づけが行える
- CCTに不足していること
 - 検出結果が膨大になるため, 複数あるファイルから必要なデータを取得・整理が必要
 - 利用過程の把握には, 図示化が必要

[5] 久木田 雄亮, 大平 雅雄. コードクローン変更過程における開発者のインタラクションとソフトウェア品質の関係. ソフトウェア・シンポジウム2017 in 宮崎. pp.120-129, 2017年6月.

目的

- CCTの検出結果を元に, コードクローンの利用過程において不具合が混入する原因の分析を補助するツールを作成



ツール概要

1. 時系列表示

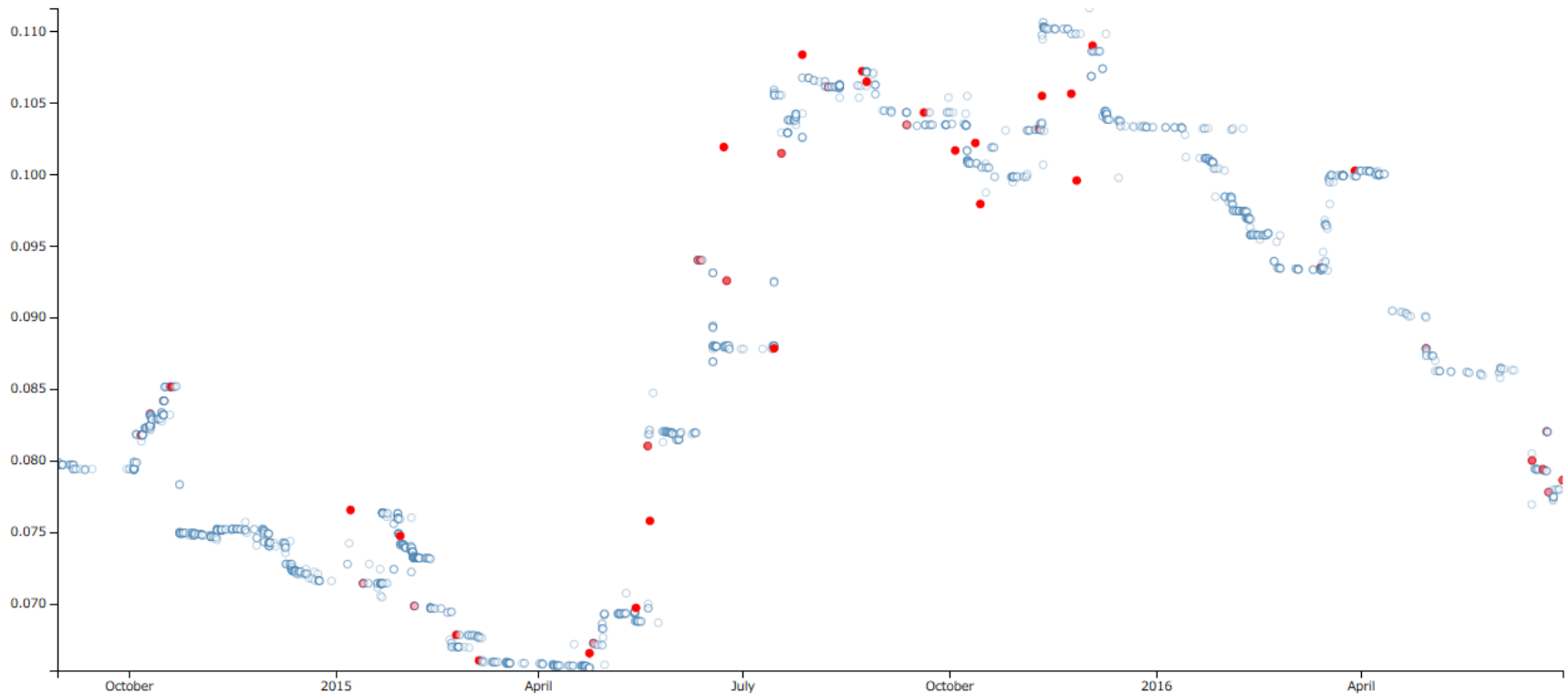
- コードクローンの利用過程を俯瞰的に認識
- コミットが含むクローン密度の時系列的変化を表示
 - クローン密度 = $\frac{\text{全クローン行数}}{\text{全コード行数}}$ [issue/kloc]

2. ファイルリスト表示

- 問題と考えられる箇所の変更の差分を確認
- 不具合と結びついたコードクローンについて、変更が加えられたコミットをリスト化
 - コミットした開発者, 変更内容, ファイルパスなど

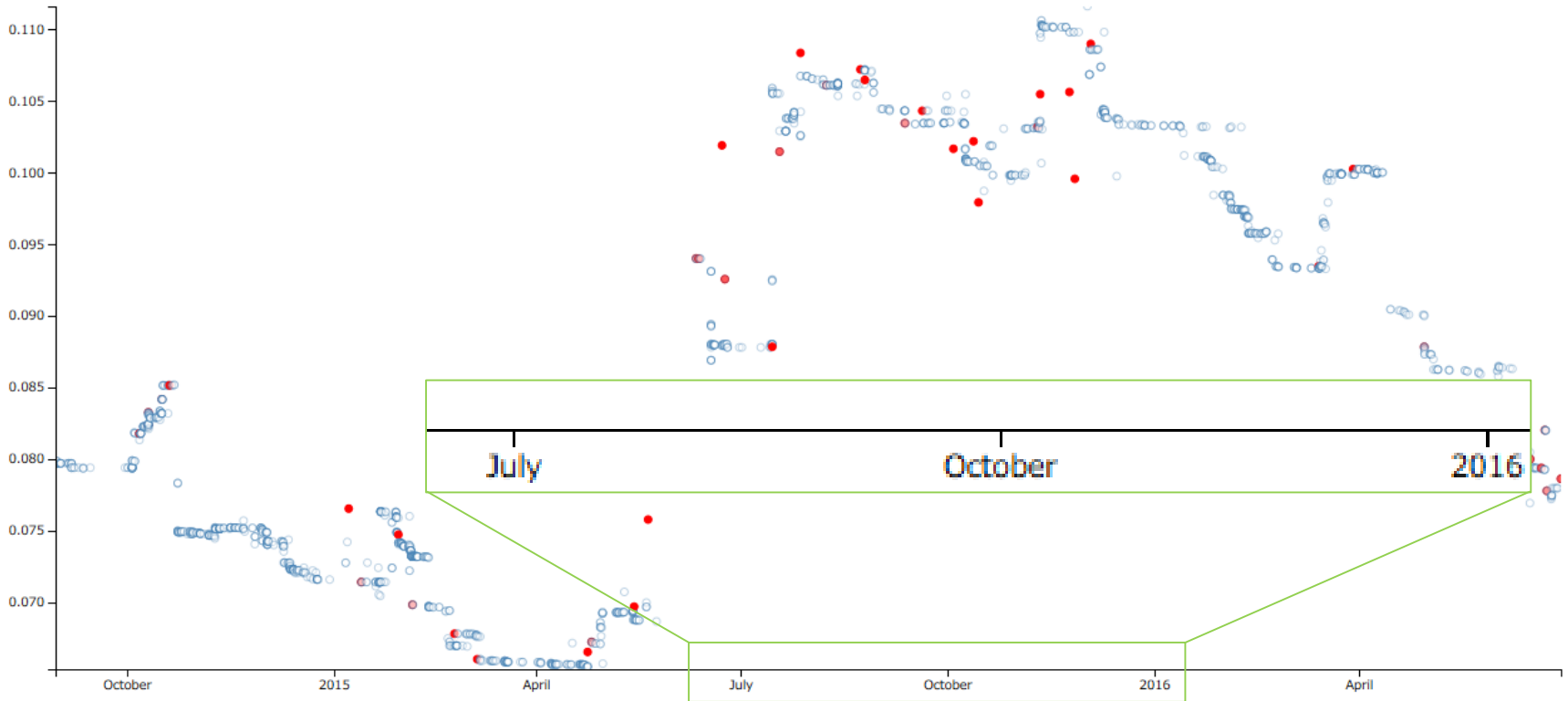
時系列表示

- x軸: 時間 [月]
- y軸: クローン密度 [issue/kloc]
- 各点: コミット
 - 赤: コードクローンと不具合を紐づけしたコミット



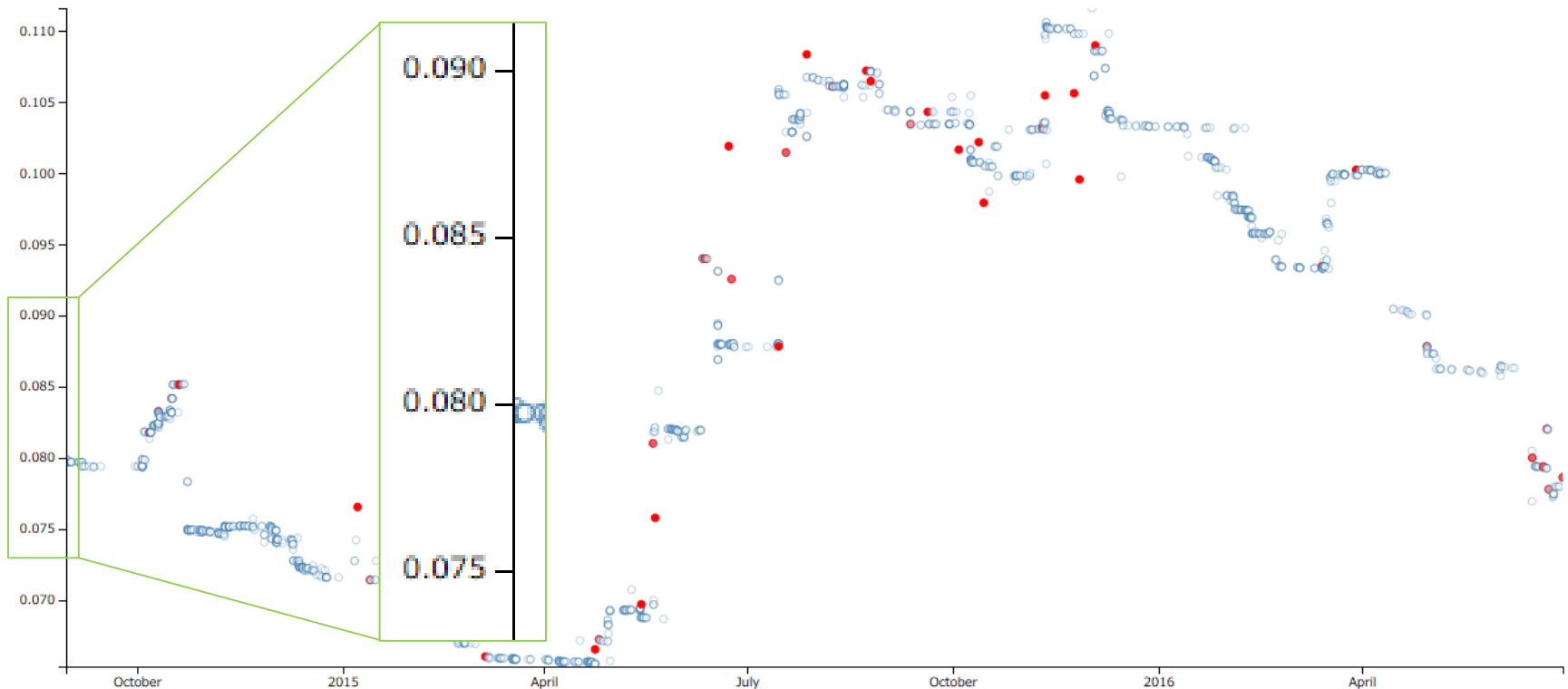
時系列表示

- x軸: 時間 [月]
- y軸: クローン密度 [issue/kloc]
- 各点: コミット
 - 赤: コードクローンと不具合を紐づけしたコミット



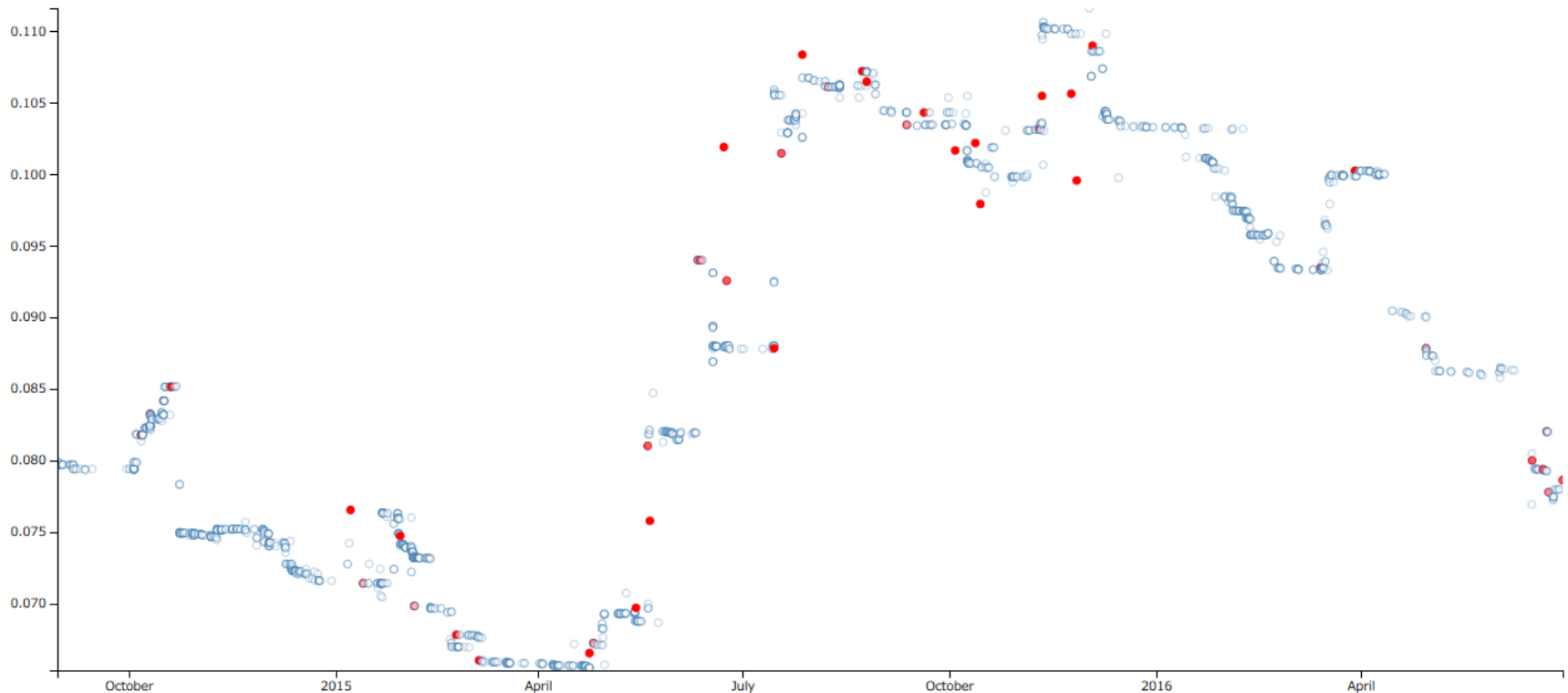
時系列表示

- x軸: 時間 [月]
- y軸: クローン密度 [issue/kloc]
- 各点: コミット
 - 赤: コードクローンと不具合を紐づけしたコミット

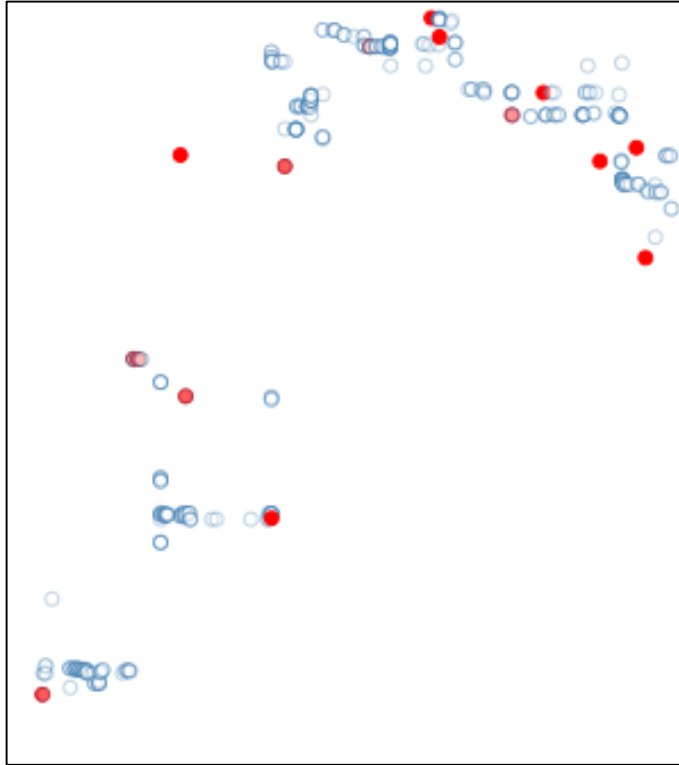


時系列表示

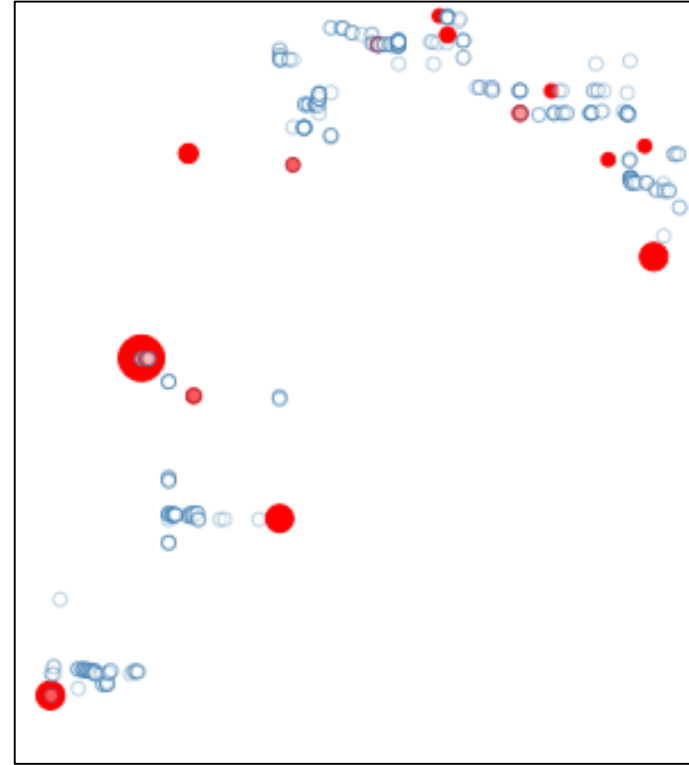
- x軸: 時間 [月]
- y軸: クローン密度 [issue/kloc]
- 各点: コミット
 - 赤: コードクローンと不具合を紐づけしたコミット



不具合が多いコミットを確認する機能

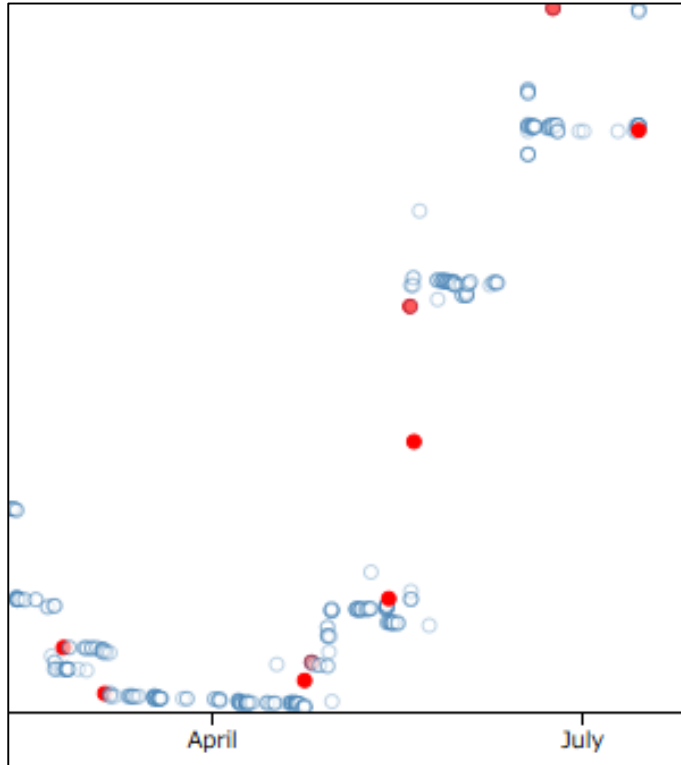


変更前

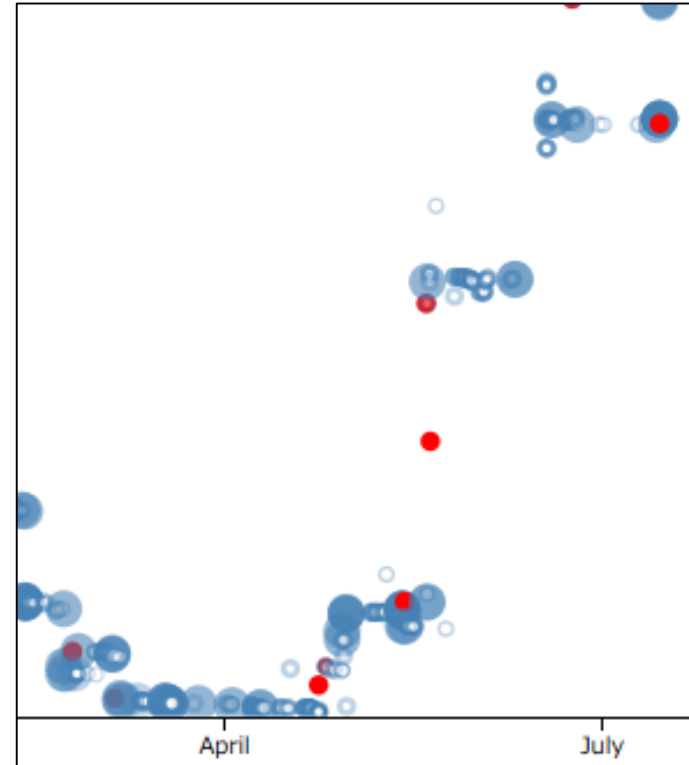


変更後

開発者を見分ける機能



変更前



変更後

評価実験

- 本ツールにより, コードクローンの利用過程における不具合混入原因の分析を補助できるかを示す
- 11の質問を設定
 - Q1, Q2, Q3, Q4
 - クローンの利用過程を把握できるかを確認
 - Q5, Q6, Q7, Q8, Q9, Q10, Q11
 - 不具合混入原因の分析を補助できるかを確認

質問 Q1～Q4

クローンの利用過程を把握できるかを確認

Q1	最もコミットがされていない期間はいつか.
Q2	クローン密度が減少している期間にどのような作業が行われているか.
Q3	コミット回数が最も多い開発者は どの開発者 か.
Q4	Q3の開発者が、実験期間ではじめて行ったコミットと最後に行ったコミットはいつか.

質問 Q5～Q11

不具合混入原因の分析を補助できるかを確認	
Q5	クローンと不具合が紐づけされたコミットの集中する期間はあるか.
Q6	Q5の原因として, クローンの変更内容 から考えられることはなにか.
Q7	一度のコミットでクローンと紐づいた不具合の混入量が最も多い開発者はどの開発者か.
Q8	一度のコミットに不具合と紐づいたクローンが多く含まれる原因として考えられる, Q7の 開発者の行動 はなにか.
Q9	クローンと不具合が紐づいたコミットを多くしているのはどの開発者か.
Q10	クローンと不具合の結びつきがQ9の開発者のコミットの多くで発生した原因として考えられる, この 開発者の行動 はなにか.
Q11	Q6とQ8, Q10を踏まえて, 対象プロジェクトについて, クローンの利用過程における 不具合が混入した原因 として考えられることはなにか.

評価方法

- 本ツールとExcelで11問の正答率, 回答時間を比較
- 回答者(学部生3名, 大学院生2名)
 - 本ツール: 3名, Excel: 2名
- 対象プロジェクト: RxJava

	対象リビジョン数	対象期間
Excel	2,628	2012/03/18～2014/04/30
本ツール	2,630	2014/04/30～2018/01/19

実験結果

	本ツール	Excel
Q1	2	1
Q2	0	1
Q3	3	2
Q4	0	2
Q5	3	1
Q6	0	1
Q7	3	0
Q8	1	0
Q9	2	1
Q10	0	0
Q11	0	0
正答率	42.4%	40.9%

	本ツール	Excel	差分
Q1	02:42.3	04:21.4	-01:39.1
Q2	06:12.5	13:47.2	-07:34.7
Q3	03:54.8	05:05.9	-01:11.1
Q4	03:24.5	01:45.3	+01:39.2
Q5	00:08.0	07:20.1	-07:12.1
Q6	04:10.0	13:09.4	-08:59.4
Q7	00:51.7	19:25.1	-18:33.4
Q8	03:08.6	10:04.3	-06:55.7
Q9	03:09.7	12:37.4	-09:27.7
Q10	05:41.9	02:05.3	+03:36.6
Q11	01:26.9	02:29.9	-01:03.0
平均	03:10.1	08:22.9	-05:12.8

回答に関する考察

- Q1, Q3, Q5, Q7, Q9
 - 全体の中から特定の期間・開発者を発見できるか
- 利用過程を俯瞰的に認識し、問題を含む期間・開発者の発見に有効

Q1	最もコミットがされていない期間はいつか.
Q3	コミット回数が最も多い開発者はどの開発者か.
Q5	クローンと不具合が紐づけされたコミットの集中する期間はあるか.
Q7	一度のコミットでクローンと紐づいた不具合の混入量が最も多い開発者はどの開発者か.
Q9	クローンと不具合が紐づいたコミットを多くしているのはどの開発者か.

	本ツール	Excel	差分
Q1	02:42.3	04:21.4	-01:39.1
Q3	03:54.8	05:05.9	-01:11.1
Q5	00:08.0	07:20.1	-07:12.1
Q7	00:51.7	19:25.1	-18:33.4
Q9	03:09.7	12:37.4	-09:27.7

結論

- CCTの検出結果を元に, コードクローンの利用過程における不具合混入原因の分析を補助するツールを作成
- 評価結果
 - 利用過程を俯瞰的に認識し, 問題を含む期間・開発者の発見に有効
 - 分析の初期段階の補助に役立つ