

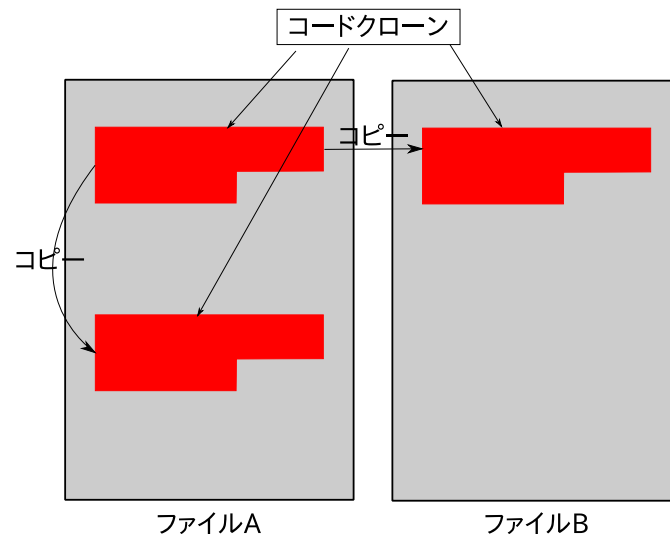
コードクローンへの欠陥混入防止に向けた欠陥混入クローンの特徴分析

オープンソースソフトウェア工学研究室

60206194 野口耕二郎

コードクローン

- ソースコード中に存在する, 「一致」 または「類似」するソースコードの断片 (コード片)
- ソースコードの品質を下げる要因となる



既存研究

- 複製された欠陥を自動的に特定する手法を構築した研究[1]
- 欠陥混入クローンから**33%**の割合で欠陥が複製されることを明らかにした研究[2]

[1]Li, J. and Ernst, M. D.: CBCD: Cloned Buggy Code Detector, ICSE '12, pp. 310–320 (2012).

[2]Mondal, M., Roy, C. K. and Schneider, K. A.: Bug Propagation through Code Cloning: An Empirical Study, 2017 IEEE International Conference on Software Maintenance and Evolution (ICSME), pp. 227–237 (2017).

問題点

- 欠陥が複製されると修正する手間が増える
- クローンへの欠陥混入を防止すれば修正する手間が省かれる
- しかし、クローンへの欠陥混入の防止に関する研究は十分に行われていない

目的・アプローチ

- 目的
 - コードクローンに欠陥が混入する原因を明らかにする
- アプローチ
 - 欠陥混入クロンのソースコードの特徴を分析する
 - クローンに混入した欠陥の特徴を分析する

リサーチクエスチョン

- RQ1：欠陥混入クローンのソースコードにどのような特徴があるか
 - RQ1-1：どれだけ複雑であるか
 - 条件分岐構造の経路数などを計測
 - RQ1-2：どれほどの規模があるか
 - 行数やステートメント数などを計測

リサーチクエスチョン

- RQ2：クローンに混入した欠陥の特徴はどのようなものか
 - RQ2-1：どのような種類の欠陥が多いか
 - RQ2-2：欠陥の影響範囲はどのようなものか

RQ1の分析手順

- 欠陥混入クローンと非欠陥混入クローンのメトリクスを計測し，比較する
- 比較結果からどのようなクローンに欠陥が混入しやすいかを分析する

RQ2の分析手順

- クローンに混入した欠陥とクローンでない部分に混入した欠陥の種類・修正範囲を比較する
- 比較結果からどのような欠陥が混入しやすいかを分析する

データセット

- プロジェクト：RxJava
 - リアクティブプログラミングをJavaで行うためのライブラリ
- 対象期間：2014年8月30日～2017年3月29 日
- リビジョン数：1703
- クローン数：616259

RQ1-1の分析結果

- メソッドに対する入力が多く、出力が小さいクローンに欠陥が混入しやすいという結果が得られた

メトリクス	p-value	cohen'sd
メソッドへの入力数	0.000	0.312
メソッドへの出力数	0.000	-0.399
サイクロマティック数	0.000	-0.070
本質的複雑度	0.000	-0.245
制御フローの交差数	0.000	-0.255
経路数	0.000	-0.165
経路数の対数	0.000	-0.298
最大ネスト数	0.000	0.269

RQ1-1の分析結果

- 条件分岐構造の最大ネスト数が大きく、経路数が小さいクローンに欠陥が混入しやすいという結果が得られた

メトリクス	p-value	cohen'sd
メソッドへの入力数	0.000	0.312
メソッドへの出力数	0.000	-0.399
サイクロマティック数	0.000	-0.070
本質的複雑度	0.000	-0.245
制御フローの交差数	0.000	-0.255
経路数	0.000	-0.165
経路数の対数	0.000	-0.298
最大ネスト数	0.000	0.269

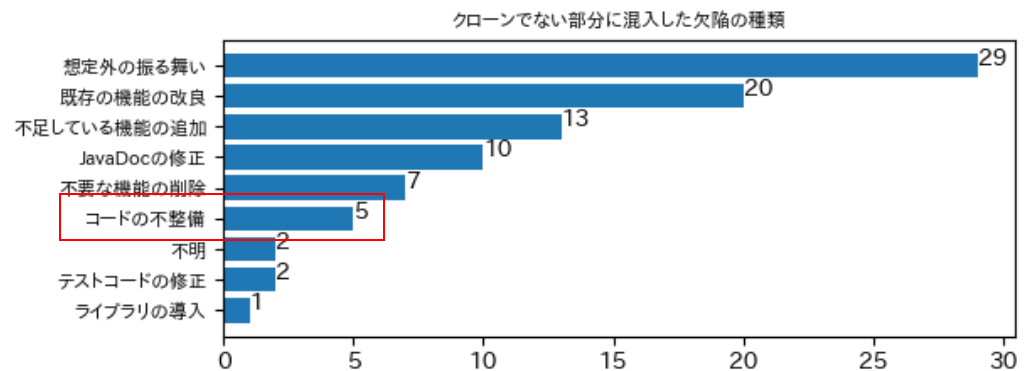
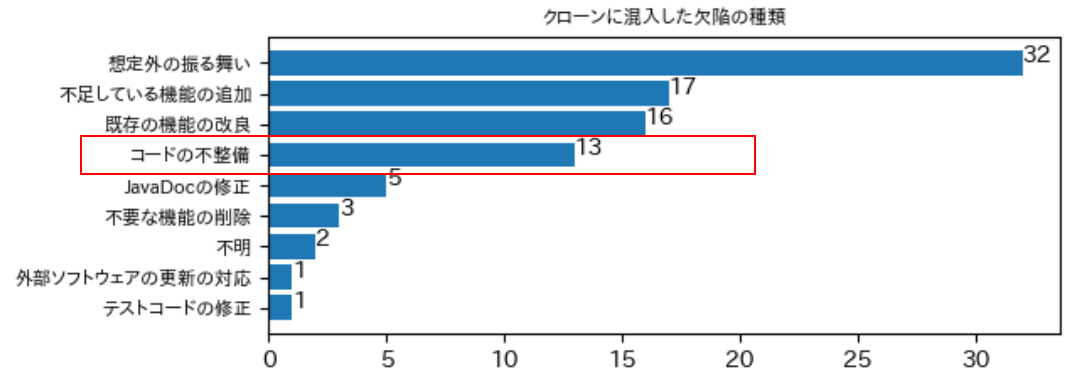
RQ1-2の分析結果

- 行数やステートメント数など、規模の小さいクローンに欠陥が混入しやすいという結果が得られた

メトリクス	p-value	cohen'sd
行数	0.000	-0.251
空白行数	0.000	0.033
有効な行数	0.000	-0.220
宣言文の行数	0.000	-0.065
実行可能な行数	0.000	-0.235
コメント行数	0.000	-0.164
セミコロンの数	0.000	-0.263
ステートメント数	0.000	-0.223
宣言ステートメント数	0.000	-0.127
実行可能なステートメント数	0.000	-0.226
行数に対するコメントの比	0.000	-0.191

RQ2-1の分析結果

- 「コードの不整備」が多いという結果が得られた



RQ2-2の分析結果

- クローンに混入した欠陥の影響範囲が大きいことが明らかになった

メトリクス	p-value	cohen's d
行数	0.000	0.321
ファイル数	0.000	0.418

RQ1の結果に関する考察

- 入力数が多いクローンに欠陥が混入しやすい
 - 他のコードとの関係が多いため
- ネスト数が多いクローンに欠陥が混入しやすい
 - 可読性が低下し、コードの理解が困難になるため
- 規模が小さいクローンに欠陥が混入しやすい
 - 変更の余地があるため

RQ2の結果に関する考察

- 「コードの不整備」が多い
 - クローンが保守性を低下させる原因となるため
- 影響範囲が大きい
 - 影響範囲が大きい箇所を変更する際、クローンに影響が及んでいる可能性がある

結論

- 以下の特徴を示すクローンに欠陥が混入しやすいことが明らかになった
 - 他のコードとの関係が多い
 - ネスト数が多いため可読性が低い
 - 変更の余地がある
- クローンによる保守性の低下が見られた場合に修正される
- 影響範囲が大きい箇所を変更する際は、クローンにも影響が及んでいる可能性がある