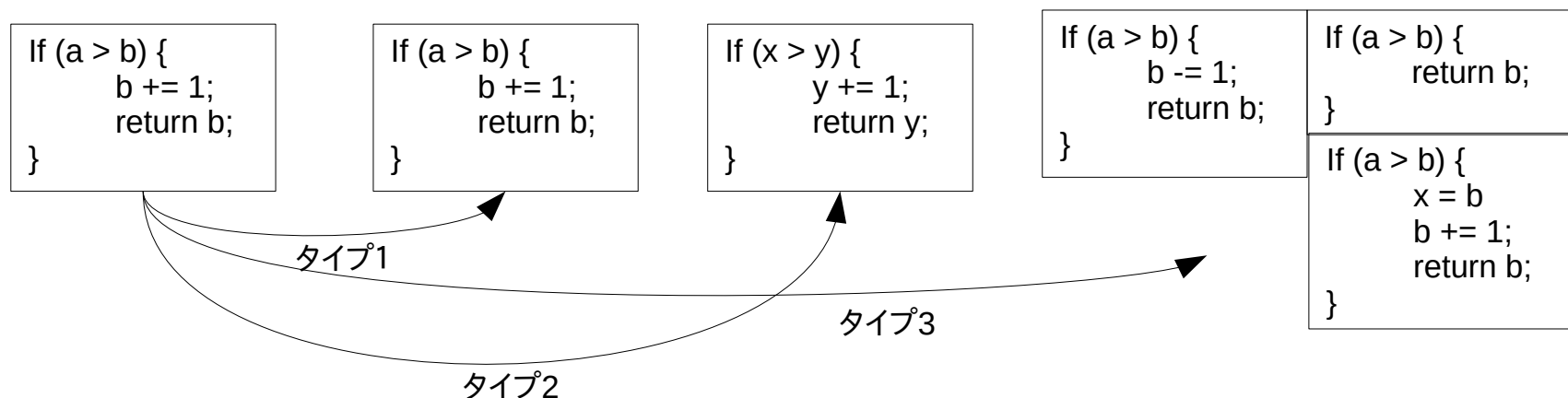


研究テーマ案

野口耕二郎

コードクローンの大まかな説明

- コードクローンとは、ソースコード中に含まれる一致または類似したコード片の集合
- 3つのタイプが存在する
 - タイプ1: スペースなどの有無を除き、完全に一致している
 - タイプ2: 変数名や関数名のみ異なる
 - タイプ3: 一部の行で追加、削除、変更が行われている



コードクローンの主な問題

- クローンの中にバグが見つかった、同じクローンセットのすべてのコード片を修正する必要がある
- 多くのクローンは信頼性の高い部分から複製されており、特別な処置を必要としないクローンがかなり多い
- バグが混入しているクローンの特徴を知り、どのようなクローンを除去すべきかを知りたい

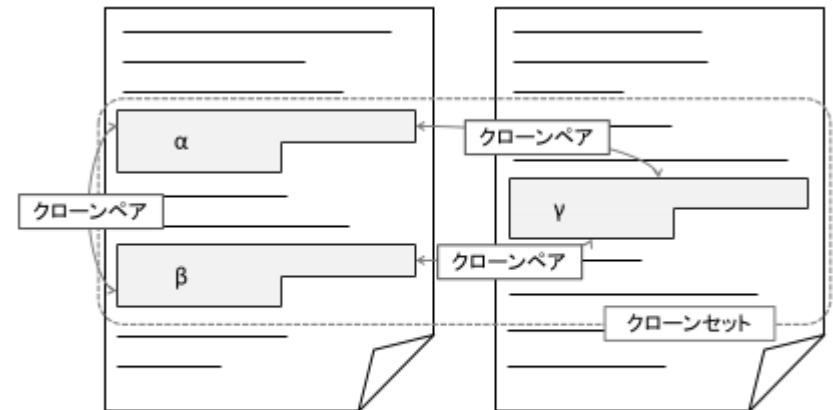


図 1 クローンペアとクローンセット

片桐さんの研究の元となった研究

- A Comparative Study of Bug Patterns in Java Cloned and Non-cloned Code
- Bug Propagation through Code Cloning: An Empirical Study

A Comparative Study of Bug Patterns in Java Cloned and Non-cloned Code

- FindBugsを用いてバグとクローンの関係を調査、分析した研究
 - FindBugs:Java バイトコードを対象とした解析を行い、FindBugs が定義した欠陥パターンに一致するソースコードの箇所を出力する
- バグの密度がクローンでない部分のほうが大きいことが分かった
- クローンに含まれるバグのカテゴリーが危険性の低いものであることが分かった

Bug Propagation through Code Cloning: An Empirical Study

- クローン系譜に含まれるバグの拡散の様子を分析した研究
- バグ拡散のパターンを定義している
 - パターン1: 同時にクローンペアがコミットされた場合
 - パターン2: あるクローン片が先にコミットされ、そのクローン片のコピーが後でコミットされた場合
- クローンタイプ2,3のクローンが拡散されやすく、パターンについてはパターン1のほうが拡散されやすいことが分かった

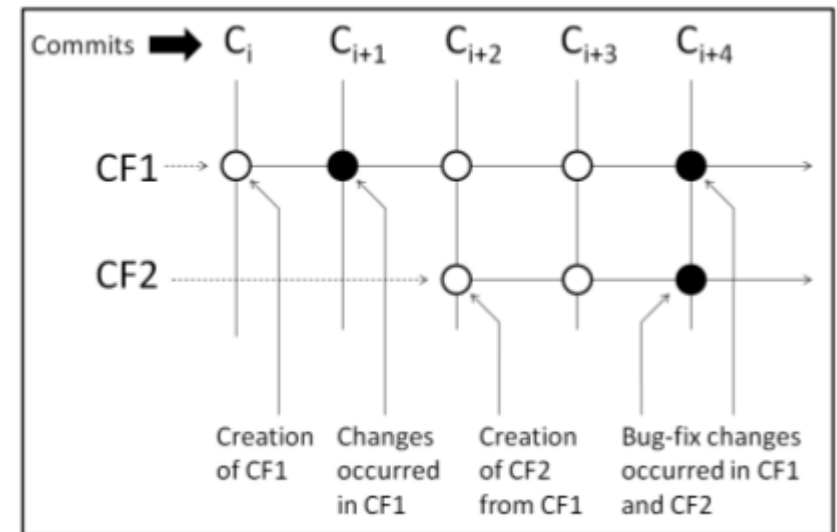


Fig. 1. A possible way of bug-propagation through code cloning

クローン系譜の中でバグが拡散している様子

片桐さんの研究

- FindBugsとCCTを用いてクローンに混入したバグの拡散の様子を分析

開発者のミスによって意図しない結果が返る欠陥を含むカテゴリ

- クローンに含まれるバグカテゴリが、ソフトウェアが成長するに連れて危険性が増していることを発見

悪意ある攻撃に対する脆弱性のある欠陥を含むカテゴリ

- 欠陥混入クローンは少数であることが分かった

欠陥カテゴリ	前期	後期
BAD_PRACTICE	0	614
CORRECTNESS	0	1,893
EXPERIMENTAL	0	0
INTERNATIONALIZATION	0	0
MALICIOUS_CODE_VULNERABILITY	0	1,628
MULTI_THREADED_CORRECTNESS	0	222
PERFORMANCE	2,174	25
SECURITY	0	0
DODGY_CODE	2,944	290

RxJavaのバグカテゴリの前期と後期の割合の変化. 赤枠で囲んだ部分が危険なカテゴリ

片桐研究ではできなかった部分

- 不具合表の内容を分析しておらず、より詳しい不具合混入クローンの特徴がわかっていない
- バグカテゴリの危険性の悪化についてあまり詳しく分析されていない
- クローンタイプやバグ拡散パターンについて考慮していない

自分の研究でやりたいこと

- 不具合票の内容を分析して、不具合混入クローンの更に詳しい特徴を知りたい
- バグカテゴリの悪化についてより詳しく分析したい
- 可能であれば、クローンタイプやバグ拡散パターンを考慮して分析を行いたい