

研究発表

大石 聖也

コードクローン

- ソースコード中に存在する類似、または同一なコード片
- 一部でバグが発生すれば、コードクローン全体を修正しないといけない
 - 変更忘れがあるとそれが新たなバグになる
- ソフトウェアの品質を低下させる原因として、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

コードクローン追跡手法

- ソフトウェアの改良・保守の過程でコードクローンがどのように拡散していくのかを分析するための手法

CCT (Code Clone Tracer) [2]

- Gitなどの分散型バージョン管理システムを対象としたクローン追跡ツール
- クローンの特定・追跡を行う機能と、不具合管理システムと連携する機能を持つ
 - クローンの作成・利用過程を、開発者との関連性も含めて調査できる

[2] 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

CCTの問題点

- 検出結果をcsvファイル形式で出力するため、「ソフトウェアの改良・保守の過程」を直観的に認識することが困難
- Issueと関係があるコードクローンを検出できるが、Issueとどう関係があるかが分かりづらい
 - 不具合の混入に関わったコードクローンなのか
 - 不具合の修正に関わったコードクローンなのか

目的とアプローチ

- 目的
 - コードクローンと不具合の関係を示し、不具合が混入される原因を理解するための補助を行う
- アプローチ
 - 3段階の可視化を用いて、不具合の混入・修正に関係があるクローンを抽出する

準備

- CCTの検出結果から使う値の説明
 - issueIds: クローンと関係があるissueのID
 - SOURCE_COMMIT_ID: 不具合が混入したリビジョンのID
 - FIX_COMMIT_ID: 不具合を修正したリビジョンのID

樹形図表示

- ソフトウェアの改良・保守の過程を直観的に認識できるように、リビジョンをもとに樹形図を表示
- issueldsの数が多いクローン集合を追跡の起点とするクローンを含むリビジョンを選択できるようにする
 - issueldsが多い = 不具合と関わることが多い

	rootCloneId	cloneSize	cloneLineC	issue_coun	author_coun	maxCloneC	otherReuse	cloneAuth	issueIds				
1	44034	110	39	4	3	16903783	0	David Karn	3491;4117;4091;3183;				
2	31748	35	22	2	4	11717325	0	David Karn	3034;2335;				
3	76806	93	40	6	5	5355255	0	David Karn	3696;4007;4119;3678;3951;3487;				
4	43533	34	8	7	4	4102160	0	George Ca	3428;4007;3787;4091;3979;3931;3183;				
5	28174	127	43	2	3	20283204	0	David Karn	2896;2897;				
6													

ファイル表示・コード表示

- ファイル表示
 - クローンに変更のあったファイルだけを表示
 - Issueとどう関係しているかで、ファイル名の表示の仕方(色など)を変える
- コード表示
 - クローンの部分だけ色を変えて、コード全体を表示

今後の課題

- どういったクローンが不具合の混入・修正に関わるかを分かりやすくする
 - 不具合が混入される原因を理解するのに役立つ
- どんな分析につながってる？