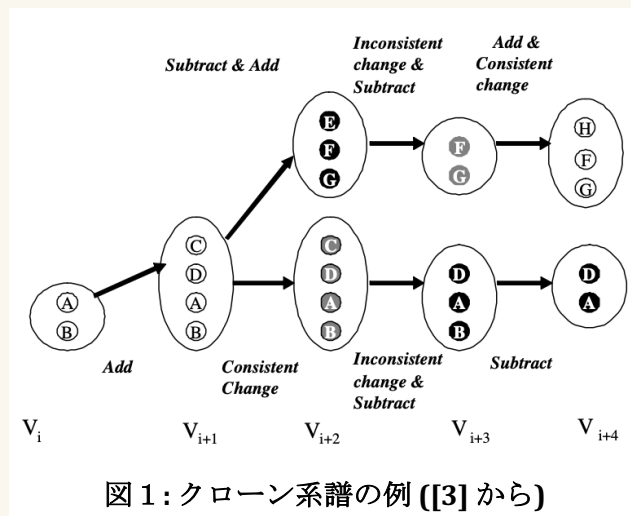


Gitリポジトリに対応した コードクローン系譜抽出 ツールの評価

オープンソースソフトウェア工学研究室 B4

60226192 NUR ATHIRAH RAMLI

クローン系譜とは



- クローン系譜は, コードクローンの変更の履歴のことである [1]
- クローン系譜から, クローンの色々な情報 (いつ発生したか, コードクローンがどのように変更されたかなど) が得られる [2]

[1] M. Zibran. Management Aspects of Software Clone Detection and Analysis. 2014.

[2] T. Moriwaki, K. Hotta, H. Igaki, K. Inoue, S. Kusumoto. Suggestion of the source code reuse analysis technique in multiple projects based on the clone genealogy. Technical report of IEICE, vol. 114 number 510, pages 61-66, 2015.

背景

- コードをコピーする時に, バグ伝搬を引き起こす可能性がある
- コードクローンを検出方法の一つはクローン系譜を使用する [3]
- 既存のクローン系譜抽出ツールはCVSやSVNからデータを受け取る [1]

[3] M. Kim, V. Sazawal, D. Notkin, and G. Murphy. An empirical study of code clone genealogies. SIGSOFT Softw. Eng. Notes 30, 5 (September 2005), 187–196.

背景

- 現代, 多数のソフトウェアはGitをリポジトリとして使用している [4]
- GitはCVSとSVNの構造が異なり[5], 既存のクローン系譜抽出ツールは対応していない可能性がある
- そのため, Gitに対応した新たなクローン系譜抽出ツールが作成された

[4] I. Tepavac, K. Valjevac, S. Kliba, M. Mijač. Version Control Systems, Tools and Best Practices: Case Git. CASE 27, Zagreb, 2015.

[5] S. Otte. Why use version control ? 2009.

目的

- Gitに対応したクローン系譜抽出ツールを評価する
- Gitに対応したツールと既存のクローン系譜抽出ツールを分析する
- 適合率と再現率を測定し, どちらの方が効率的であるかを明らかにする

関連研究

- CGE [3]
 - コードクローンの進化を理解するために, クローン系譜の概念を提案した
- gCad [6][7]
 - 最新のクローン系譜抽出ツール

[6] R. K. Saha, C. K. Roy and K. A. Schneider. An automatic framework for extracting and classifying near-miss clone genealogies. 2011 27th IEEE International Conference on Software Maintenance (ICSM), Williamsburg, VI, 2011, pp. 293-302.

[7] R. K. Saha, C. K. Roy and K. A. Schneider. gCad: A Near-Miss Clone Genealogy Extractor to Support Clone Evolution Analysis. 2013 IEEE International Conference on Software Maintenance, Eindhoven, 2013, pp. 488-491.

アプローチ

1. 複数のプロジェクトを使い, Gitに対応したツールと前のバージョンコントロールシステムに対応しているツールの適合率と再現率を測定する
2. 測定結果を用いて, 分析を実施する

今後の予定

- 対象プロジェクトを決める
- 他のクローン系譜抽出ツールの動かす方法を調べる
- 論文執筆の引き続きをやる