

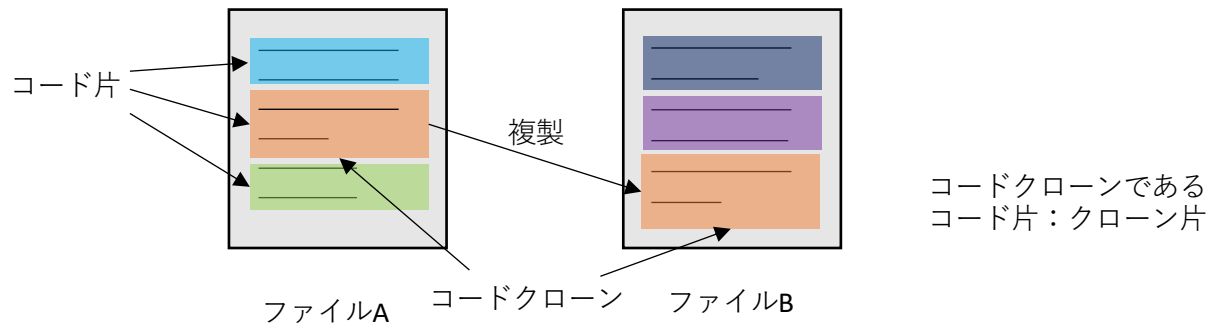
コードクローンに特化した 欠陥混入予測手法の構築

オープンソースソフトウェア工学研究室 M2

60200089 野口 耕二郎

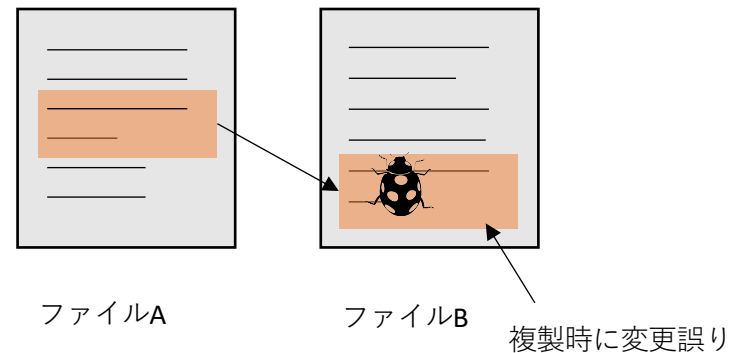
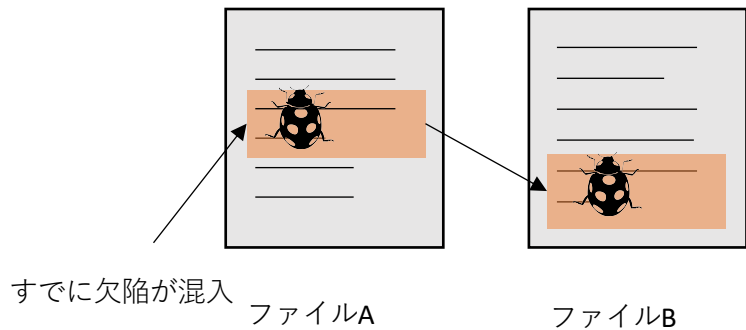
コードクローン

- 「一致」または「類似」するソースコードの断片
- 生産性を向上させるためにコードクローンの作成が頻繁に行われる
- コードクローン作成時に欠陥が複製されたり新たに混入する場合がある



コードクローン生成時に混入する欠陥

- 欠陥が混入しているコードを欠陥が混入したまま複製（左）
- 複製時に適切な変更を加えなかったことで複製先のコードに欠陥が混入（右）



問題点

- 欠陥を複製しやすいクローンはソースコードの品質に悪影響を与える
- コードクローンに混入した欠陥予測手法は構築されていない

目的

1. 既存の欠陥予測手法で欠陥混入クロンを予測できるかを明らかにする
2. コードクロンに特化した欠陥混入予測手法の構築

先行研究

- Just-In-Time Defect Prediction[1][2]
 - コミット時に欠陥が混入しているかどうかを予測
 - 従来手法に比べて欠陥混入箇所の特定範囲が細かい

欠陥混入クロンの予測に特化した
手法は構築されていない

[1] Kamei, Y.; Shihab, E.; Adams, B.; Hassan, A. E.; Mockus, A.; Sinha, A. & Ubayashi, N. A large-scale empirical study of just-in-time quality assurance
IEEE Transactions on Software Engineering, 2013, 39, 757-773

[2] McIntosh, S. & Kamei, Y. Are Fix-Inducing Changes a Moving Target? A Longitudinal Case Study of Just-In-Time Defect Prediction IEEE
Transactions on Software Engineering, 2018, 44, 412-428

先行研究で用いられている予測モデル

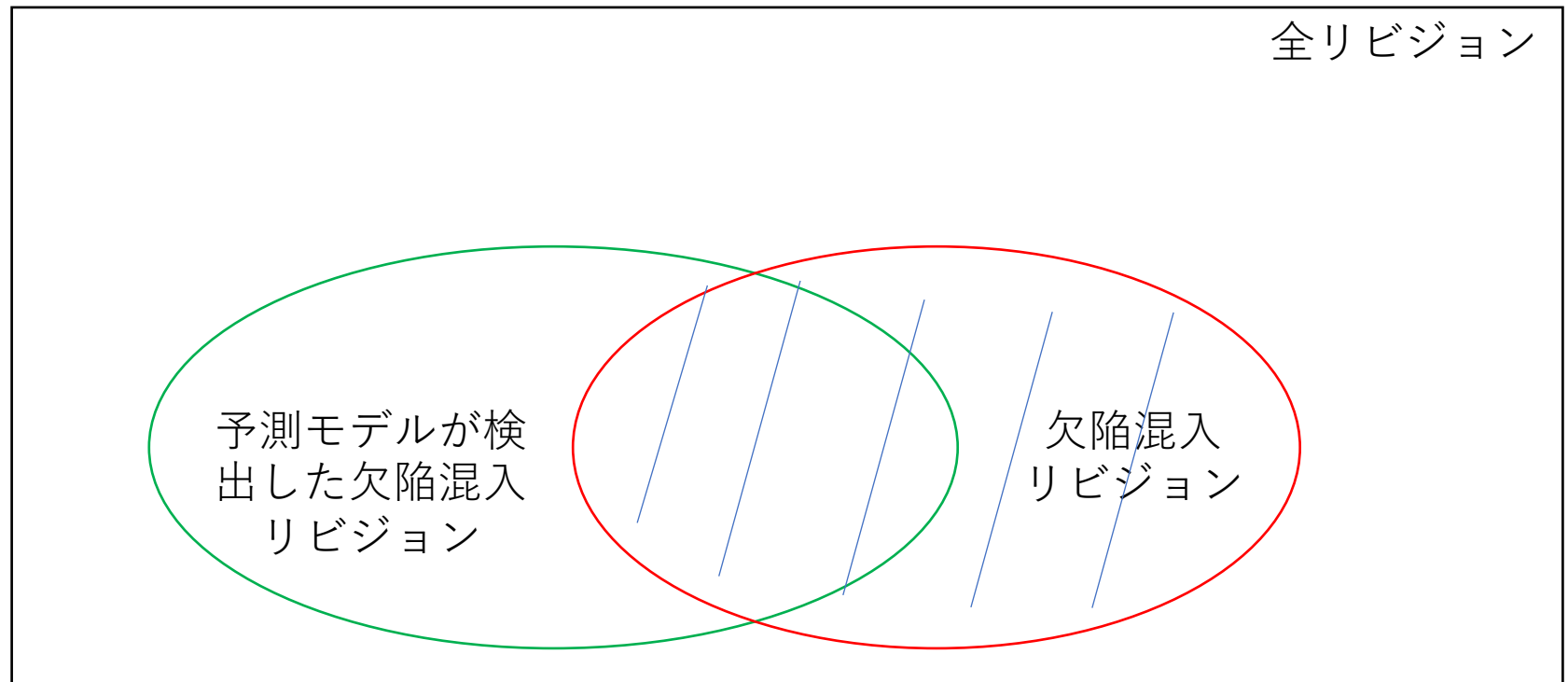
- 使用するアルゴリズムはロジスティック回帰

- $$isBuggy = \frac{1}{1 + e^{-(a_1 \cdot x_1 + a_2 \cdot x_2 + \dots + a_n \cdot x_n)}}$$

- 5種類, 17項目のメトリクスを説明変数とし, 欠陥の有無を目的変数とする
 - 変更規模
 - 変更の拡散度合い
 - 歴史
 - 開発者の経験
 - コードレビュー

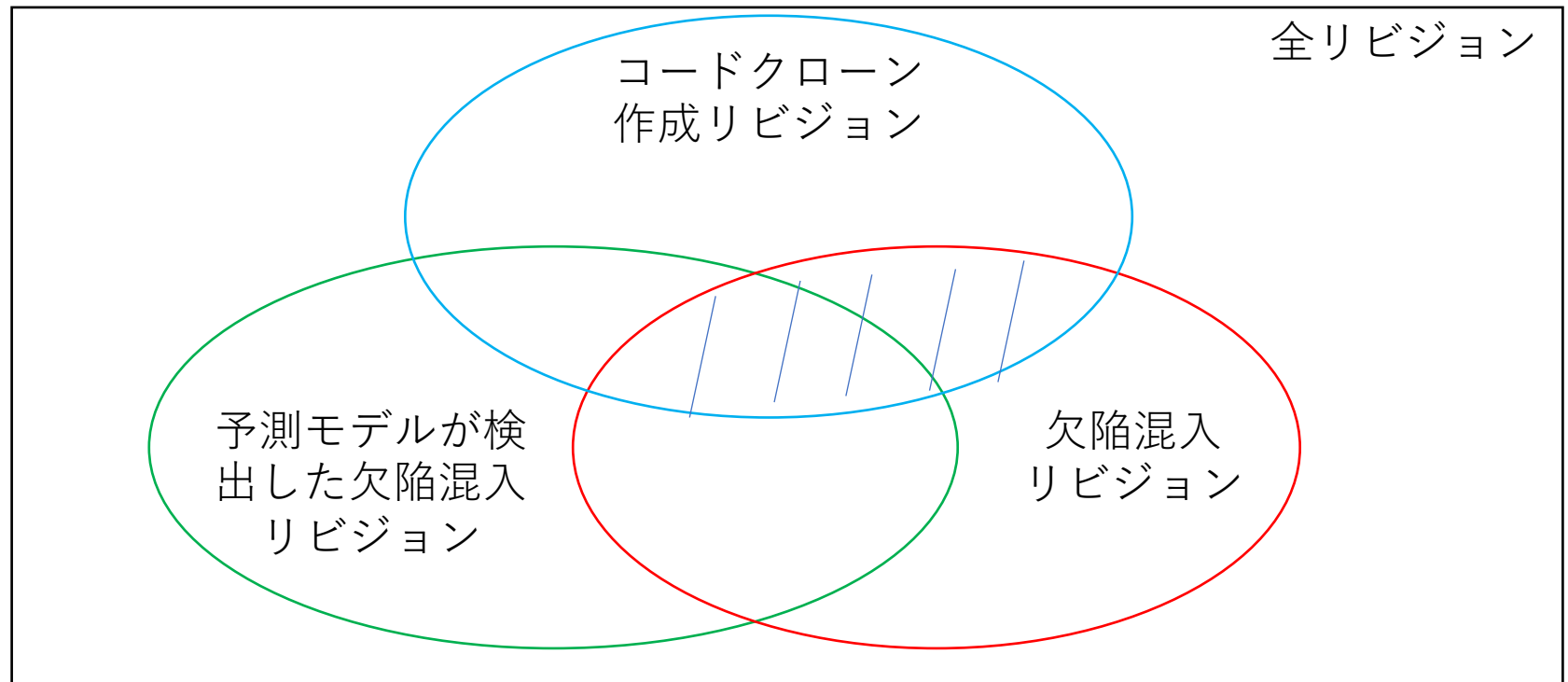
先行研究の予測対象

- 欠陥混入リビジョンを予測対象としている



本研究の予測対象

- 欠陥混入クローン作成リビジョンを予測対象としている



アプローチ

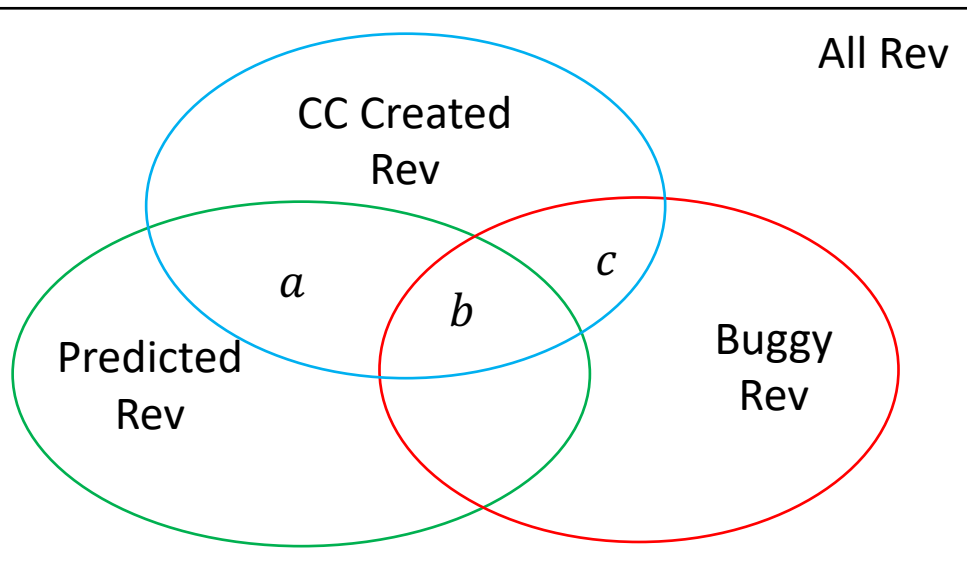
1. 既存手法で欠陥混入クローンを予測した際の予測精度を測る
2. 1で測定した予測精度に基づいて分析を実施する
3. 分析結果に基づいて予測手法を構築する

既存手法を用いた予測精度の算出

• a , b , c をもとに再現率, 適合率を算出する

- 再現率 : $\frac{b}{b+c}$

- 適合率 : $\frac{b}{b+a}$



$b + c$: 欠陥混入クローン作成リビジョン

b : 既存手法で予測できた欠陥混入クローン作成リビジョン

c : 既存手法で予測できなかった欠陥混入クローン作成リビジョン

a : 既存手法が誤って予測したクローン生成リビジョン

データセット

- プロジェクト : OpenStack
- リビジョン数 : 25150
- 期間 : 2011/6 – 2014/3
- 言語 : python

結果

- 今週中まで

予測精度に基づく分析

- 欠陥混入リビジョンのみ予測した場合に比べて,
 - 再現率が小さい：欠陥混入クローンの検出に特化した説明変数を調査
 - 適合率が小さい：誤検知の原因の特定
 - どちらも大きい：既存手法で用いている説明変数の中で欠陥混入クローンの検出に向いているものを調査

分析結果を用いて予測モデル構築

- 欠陥混入クローンの予測モデル構築に適したメトリクスをロジスティック回帰の説明変数に使用
- 予測精度が向上するかどうかを実験
- 影響のあるメトリクスを調査

今後の予定

- 既存手法を用いて欠陥混入クローンを予測した際の精度に基づいて分析を実施する
- 分析に基づいて欠陥混入クローンに特化した予測手法を構築する