

An Empirical Study of Inadequate and Adequate Test Suite Reduction Approaches

一言で言うと……

テストケース削減手法を相互検証し、欠陥検出力とテストケース数とのトレードオフを明らかにした。

出典：

Carmen Coviello, Simone Romano, and Giuseppe Scanniello. 2018. An empirical study of inadequate and adequate test suite reduction approaches. In Proceedings of the 12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM '18). ACM, New York, NY, USA, Article 12, 10 pages. DOI: <https://doi.org/10.1145/3239235.3240497>

紹介担当：伏田 享平（NTTデータ）

前提

- 回帰試験：
システムに対する修正が、振る舞いに変更を及ぼさないかを確認するための技法
- Test Suite Reduction (TSR)：
テストケースの数を削減する手法
 - Adequate : テストに対する要求を保持する
 - Inadequate : テストに対する要求が落ちることを許容する

※この論文では要求 = テストカバレッジとする

目的

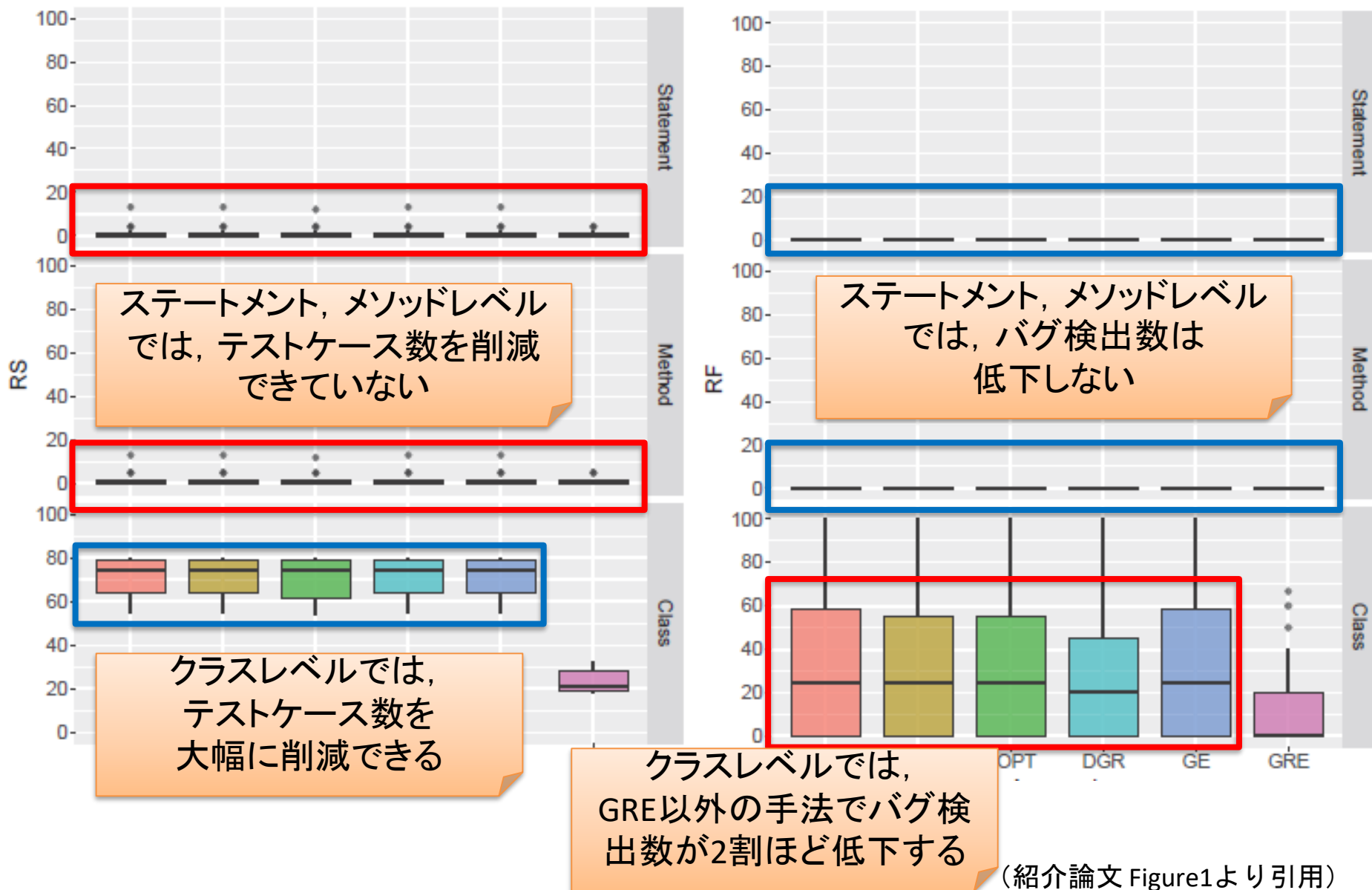
TSR手法により回帰テストの効率，欠陥検出力がどの程度変化するのか検証する

- RQ1：カバレッジの取り方ごとに，テストケース数および欠陥検出力が最適となるAdequateな手法はどれか？
- RQ2：カバレッジの取り方ごとに，テストケース数および欠陥検出力が最適となるInadequateな手法はどれか？
- RQ3：Inadequateな手法はAdequateな手法と比較してテストケース数をより多く削減しつつ同程度の欠陥検出力を持つのか？

実験設定

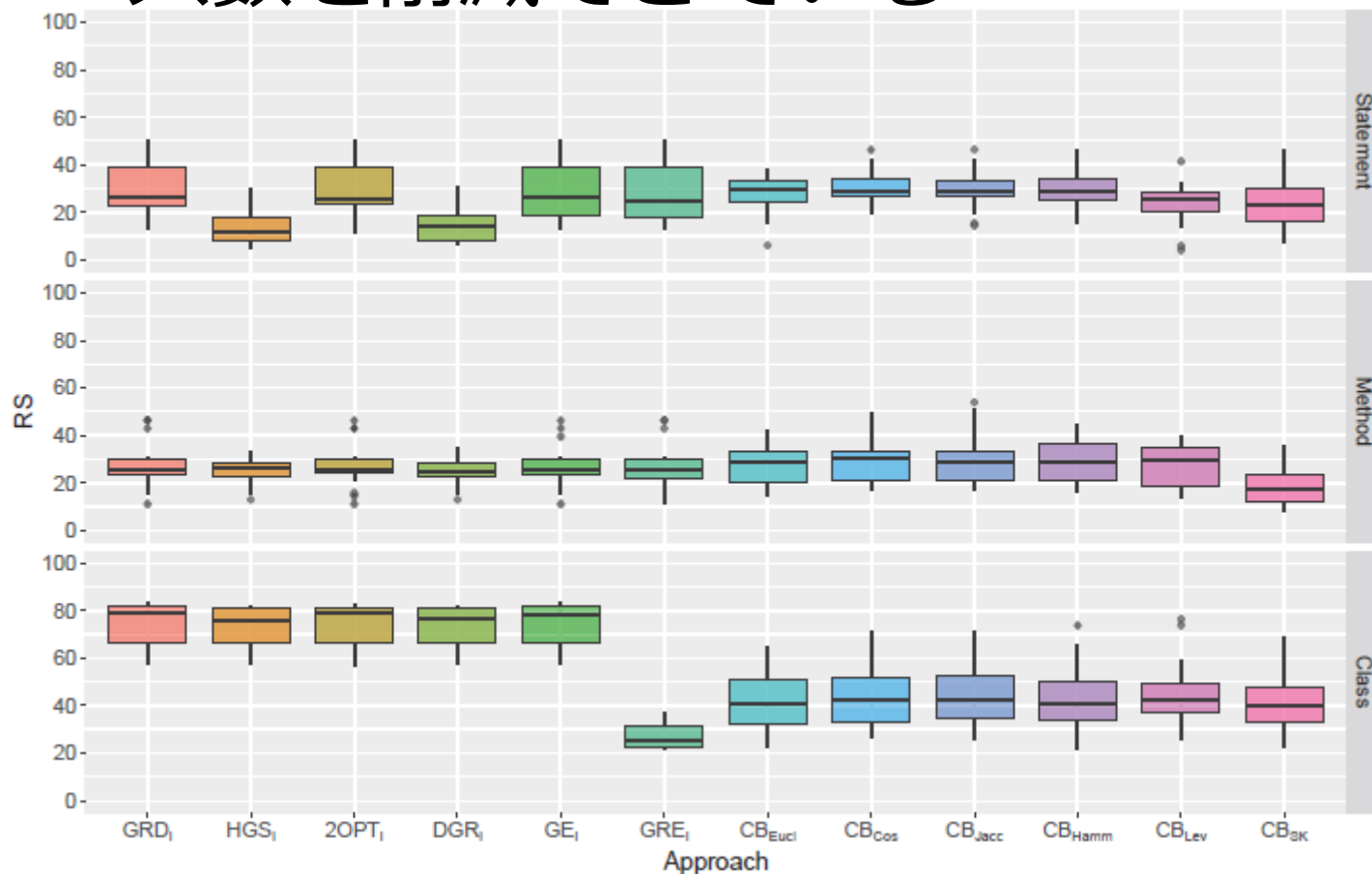
- 対象とするTSR手法（2章参照）
 - Adequate : GRD, HGS, 2OPT, DGR, GE, GRE
 - Inadequate:
GRD_I, HGS_I, 2OPT_I, DGR_I, GE_I, GRE_I,
CB_{EUCL}, CB_{COS}, CB_{JACC}, CB_{HAMM}, CB_{LEV}, CB_{SK}
- 実験対象：
Ant, JTopas, Jmeter, XMLSecurity
- 評価指標
 - RS : テストケースの削減率
 - RF : 検出できたバグ数の減少率
- カバレッジ：
ステートメント, メソッド, クラス単位で評価

結果：RQ1



結果:RQ2:テスト削減割合

全体的にAdequateな手法と比較して、テストケース数を削減できている

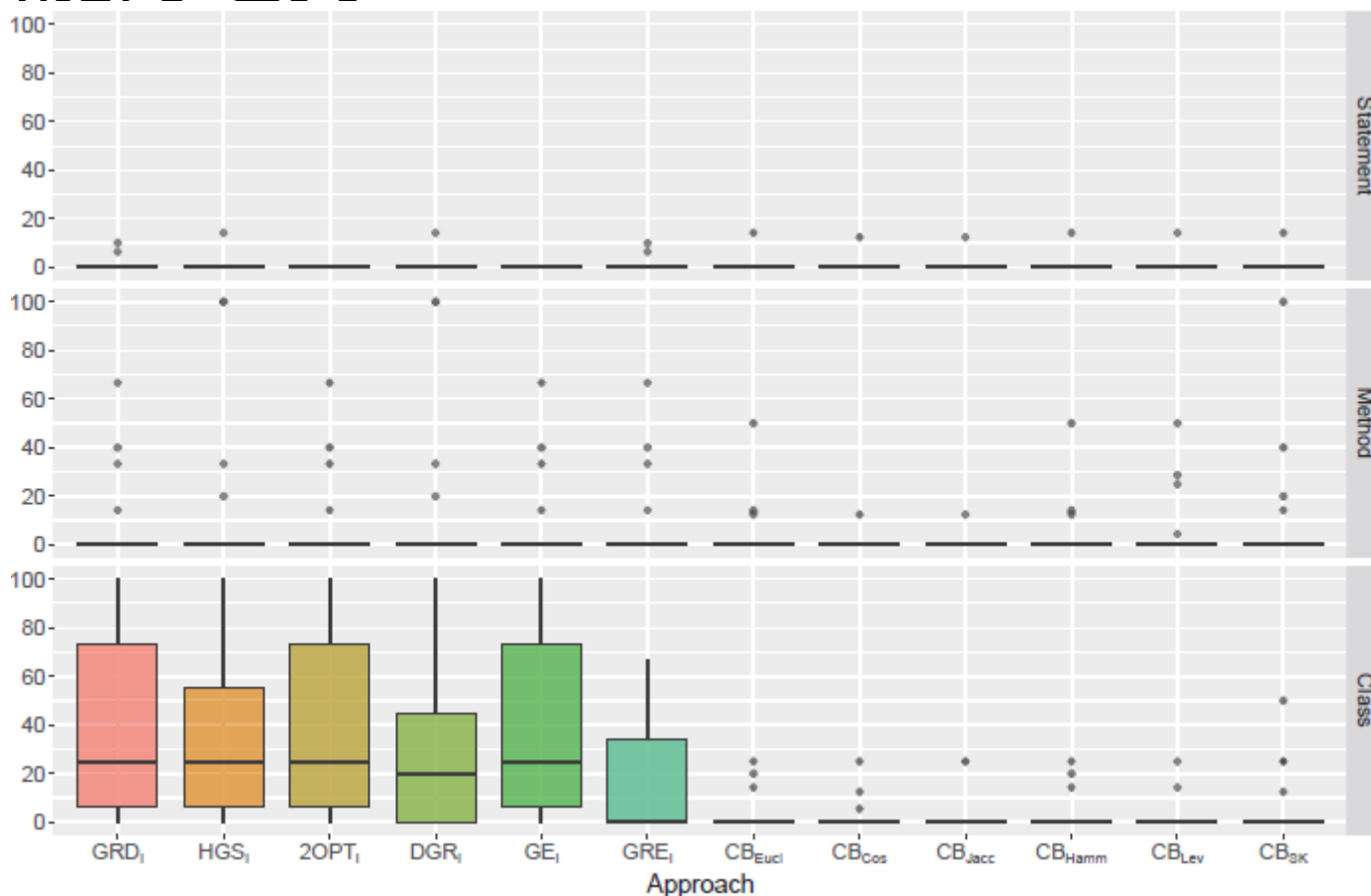


5%のカバレッジ低下を許容している

(紹介論文 Figure2aより引用)

結果:RQ2:バグ検出数減少率

CB手法はテストケース削減前と同様のバグ検出能力を持つ



5%のカバレッジ低下を許容している

(紹介論文 Figure2bより引用)

RQ3:結果 & 全体考察

- Inadequateな手法の方がよりAdequateな手法よりもテストケース数を削減しつつ、一定の欠陥検出力を発揮した
 - 特にCBについてはクラスレベルでもテストケース数を削減しつつ欠陥検出力は下がらない
- 今回の結果は、効率よくテストを行うための基準となる