

The ABLoTS Approach for Bug Localization: is it replicable and generalizable?

Feifei Niu*, Christoph Mayr-Dorn[†], Wesley K. G. Assunção[‡], LiGuo Huang[‡],
Jidong Ge*, Bin Luo*, Alexander Egyed[†]

*State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China
Email: niufeifei@smail.nju.edu.cn, {gjd, luobin}@nju.edu.cn

[†]Institute for Software Systems Engineering, Johannes Kepler University, Linz, Austria
Email: {christoph.mayrdorn, wesley.assuncao, alexander.egyed}@jku.at

[‡]Department of Computer Science and Engineering, Southern Methodist University, Dallas, Texas, USA
Email: lghuang@lyle.smu.edu

背景

Bug Localization

- バグの原因をソースコードから特定
- デバッグコストを削減

IRBL(Information Retrieval-based Bug Localization)

- バグレポートを利用するアプローチ

先行研究

AmaLgam(Wangら)

- コード構造, 類似バグレポート, バージョン履歴を組み合わせる

ABLoTS (Rathら)

- AmaLgamの類似バグレポートをTraceScoreに置き換える
- AmaLgamを大きく上回る評価

研究目的

- ABLoTSの性能を示す他の論文はない
- より大きなデータセットでも性能を維持できるか不明

目的

- TraceScore及びABLoTSアプローチの再現性と一般化可能性を調査する

TraceScore

- ① アーティファクトの選択
 - 新しいバグ報告 b^* が提出された日から1年以内に修正されたバグ報告 b
 - 修正されたjavaファイル数が10以下のバグ報告 b
 - 修正されたjavaファイル数が20以下のバグ以外の報告 r
- ② 選別したファイルの前処理
- ③ b^* と各アーティファクト間のコサイン類似度を計算
- ④ b^* を根としたトレースグラフの作成
- ⑤ 各ファイルと b^* 間のTraceScoreを計算
- ⑥ ランク付けされたリストとして出力

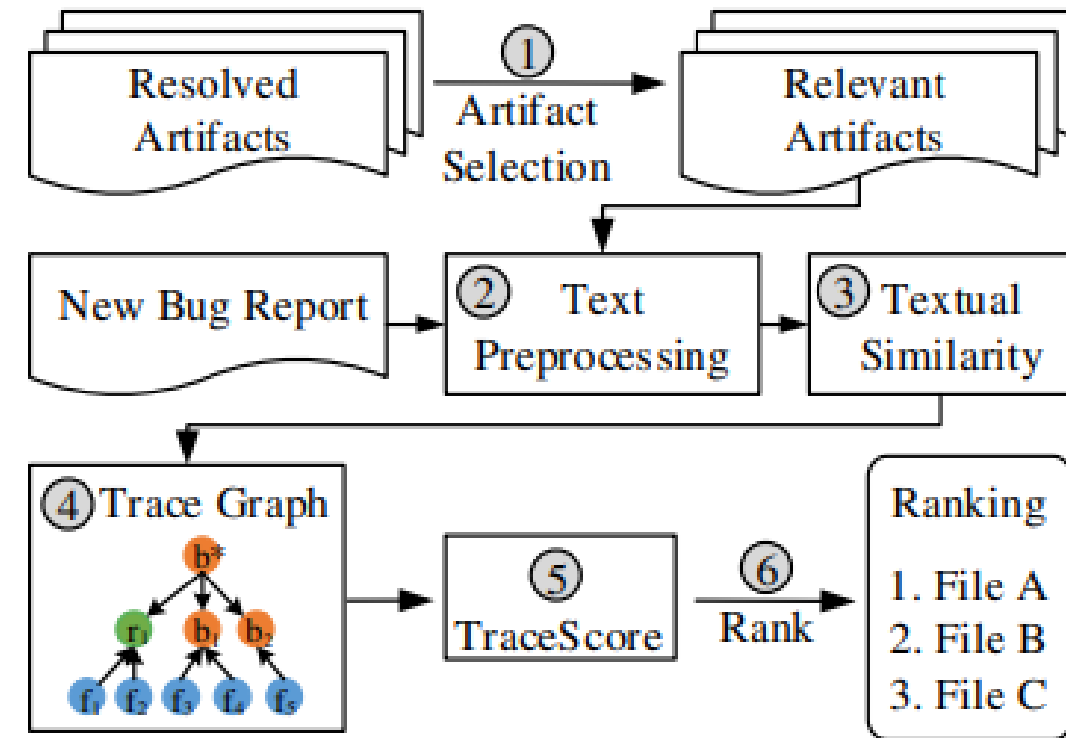


Fig. 2. TraceScore Component.

ABLoTS

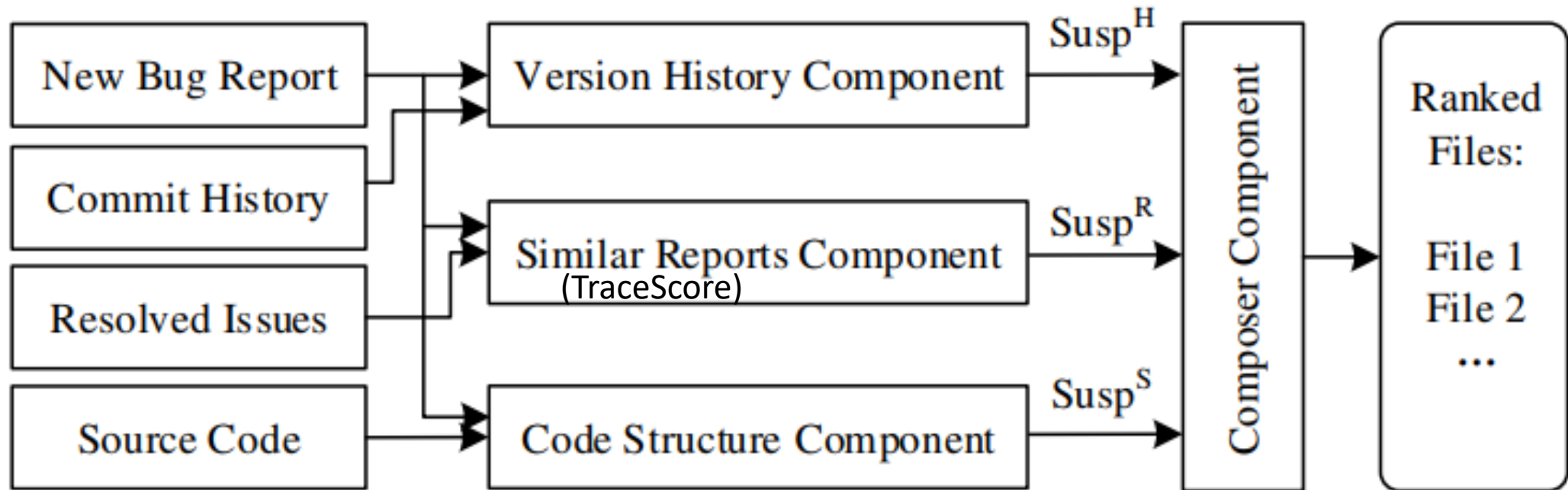


Fig. 1. Components of ABLoTS.

RQ1

緩い条件: 新しいバグ報告が提出された日から1年以内にバグ報告

厳しい条件: 緩い条件 + 新しいバグ報告が提出される前に修正が完了

RQ1.1 TraceScore コンポーネントの本来の性能を再現できるか？

緩い条件の場合再現可能

Metrics	Pairs		Deviation	P value
	Original	Replication		
MAP	0.191	0.145	0.05	0.000**
MRR	0.248	0.242	0.01	0.522
Top 1	0.163	0.156	0.01	0.407
Top 5	0.336	0.335	0.00	0.924
Top 10	0.419	0.419	0.00	0.99

*p < 0.05 **p < 0.01

RQ1

RQ1.2 TraceScoreコンポーネントを他のデータに適用しても同様のパフォーマンスが得られるか？

Metrics	TraceScore	
	K-S test	P value
MAP	0.438	0.124
MRR	0.409	0.175
Top 1	0.409	0.175
Top 5	0.409	0.175
Top 10	0.415	0.159

拡張データセットにも一般化可能

RQ2

RQ2.1 ABLoTSアプローチの結果を再現できるか？

RQ2.2 ABLoTS手法を他のデータに適用しても同様の結果が得られるか？

TABLE I
ORIGINAL, REPORTED PERFORMANCES [23]

Algorithm	MAP	MRR	Top 1	Top 5	Top 10
TraceScore	0.202	0.260	0.174	0.350	0.436
ABLoTS	0.488	0.545	0.487	0.610	0.649

TABLE VII
ABLoTS PERFORMANCE ON ORIGINAL DATA SET.

PROJECTS	MAP	MRR	TOP 1	Top 5	Top 10
DERBY	0.076	0.111	0.02	0.171	0.326
DROOLS	0.049	0.06	0.023	0.054	0.097
HORNETQ	0.057	0.067	0	0.056	0.185
IZPACK	0.086	0.11	0.016	0.172	0.375
KEYCLOAK	0.029	0.05	0.006	0.044	0.101
LOG4J2	0.065	0.072	0.011	0.067	0.146
RAILO	0.06	0.077	0	0.1	0.283
SEAM2	0.08	0.105	0.019	0.179	0.333
TEIID	0.056	0.079	0.015	0.104	0.231
WELD	0.02	0.024	0	0.018	0.027
WILDFLY	0.03	0.04	0.007	0.036	0.087
Average	0.055	0.072	0.011	0.091	0.199

- オリジナルの結果に遠く及ばなかった

RQ2

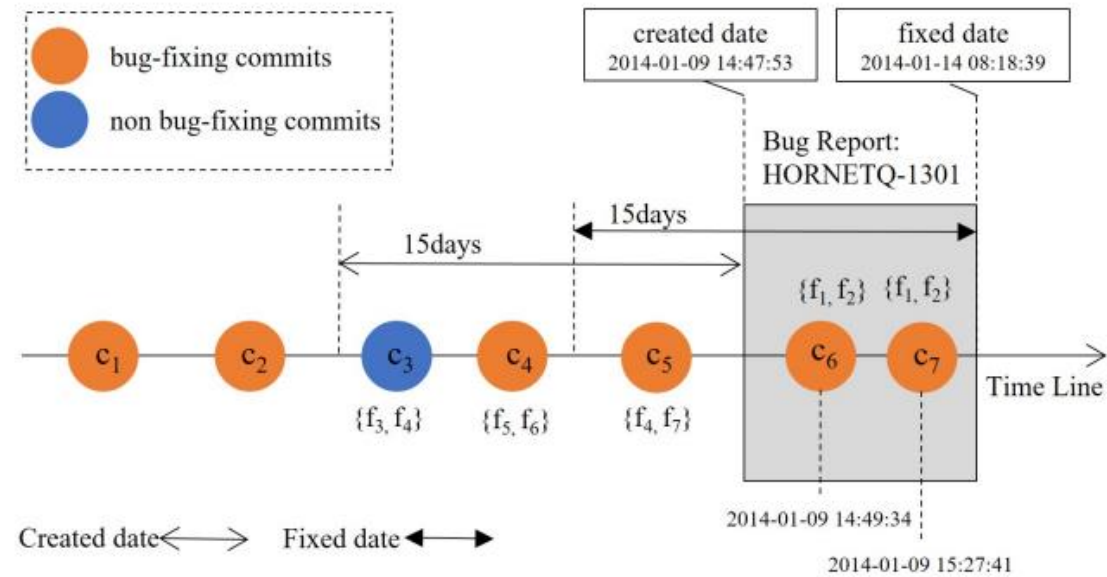


Fig. 4. BugCache using created date vs using fixed date.

バージョン履歴コンポーネントの設定ミスのため再現及び一般化不可

まとめ

- ABLoTSによるバグ位置特定手法の再現研究を行った
- TraceScoreは、緩い条件の下では複製可能で一般化可能
- ABLoTSはバージョン履歴コンポーネントの設定ミスのため再現及び一般化不可