

Investigating the Impact of Bug Dependencies on Bug-fixing Time Prediction

Chenglin Li, Yangyang Zhao* Yibiao Yang, Yuming Zhou
Liming Nie Zuohua Ding

2023 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)

バグの依存関係について

追跡システムのIssue Links
フィールドを元にできる

たいてい多対多の関係

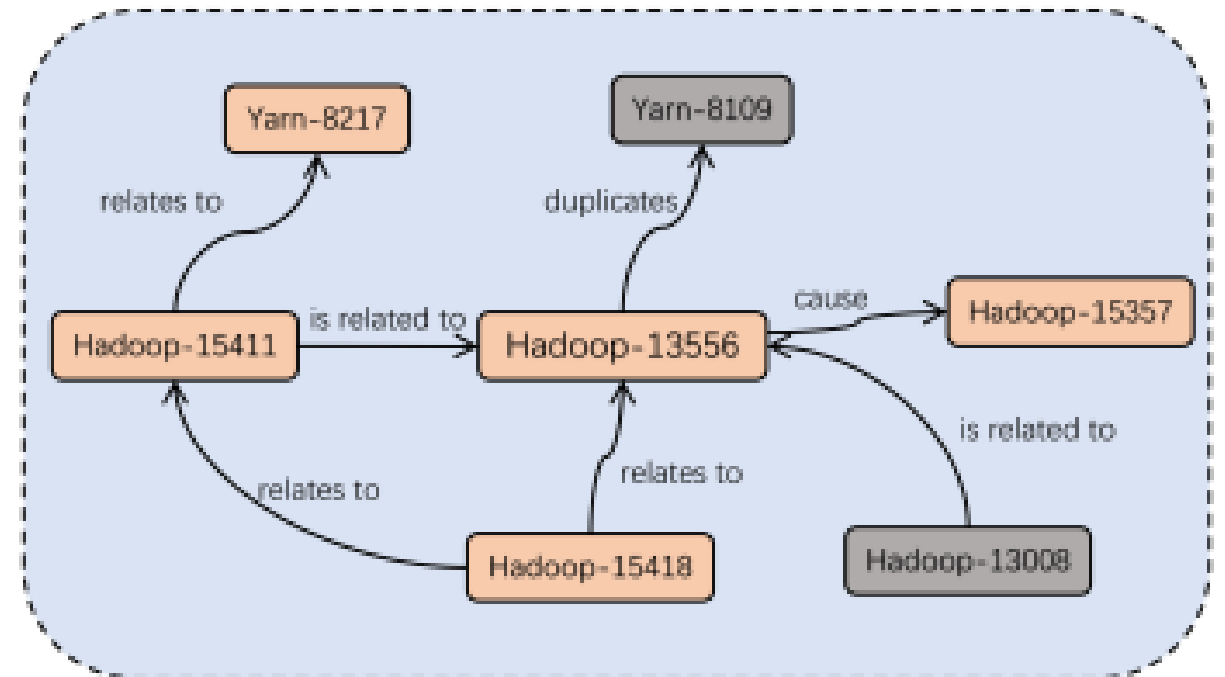


Fig. 2. An example of bug dependency graph

オレンジ：バグ グレー：非バグ

複雑な依存関係を持つバグは修正に時間がかかると
予想できる

目的・RQs

目的

- ・ 依存度合と修正時間が相関関係にあるかどうか調べる
- ・ 依存度合は予測の精度を向上させるか調べる

RQs

- 1, 修正時間と依存度合は相関関係であるか
- 2, 従来のメトリクスに対して冗長ではないか
- 3, バグ依存関係は予測の精度を向上させるのに有用か

依存関係メトリクス ネットワーク分析メトリクス

エゴネットワーク

隣接するノードに関するメトリクス

グローバルネットワーク

依存関係全体における重要性に注目する

- ・ 構造ホールメトリクス

依存関係の密接さ

- ・ 中心性メトリクス

ネットワーク内のノードの位置を説明

TABLE I
DESCRIPTION OF THE STUDIED NETWORK ANALYSIS METRICS.

type	metrics	description
Ego network	Size	# alters that ego is directly connected to
	Ties	# ties in the ego network
	Pairs	# pairs of alters in the ego network
	Density	% possible ties that are actually present
	nWeakComp	# weak components in the ego network
	pWeakComp	# weak components normalized by size
	2StepReach	# nodes ego can reach within two steps
	ReachEffic	2StepReach normalized by sum of alters' size
	Broker	# pairs not directly connected to each other
	nBroker	Broker normalized by the number of pairs
Global network	EgoBetw	% all shortest paths across ego
	nEgoBetw	normalized EgoBetween (by ego size)
	EffSize	# bugs minus the average degree of bugs
	Efficiency	effsize divided by number of bugs.
	Constraint	The extent to which ego is constrained.
	Hierarchy	measures how the constraint measure is distributed across neighbors.
	Degree	# issues adjacent to a given bug
	Degree centrality	Centrality score according to the degrees
	Closeness	How close to other vertices.
	Eigenvector centrality	Centrality score according to eigenvector.
	Betweenness centrality	Centrality score according to betweenness.

依存関係メトリクス

依存カウントメトリクス

カウント方法で収集されるメトリクス

依存関係グラフに基づいてバグの依存関係の数 方向 タイプ 範囲
に関する情報を捉える

ネットワークメトリクスと併用して、バグの依存関係を
より包括的に説明する

TABLE II
THE DESCRIPTION OF DEPENDENCY COUNTING METRICS.

dependency counting metrics	description
Bug count(BC)	# nodes in the bug dependency graph.
Entire dependencies count (EDC)	# edges in the bug dependency graph.
Dependencies density (DD)	the number of nodes divided by the maximum number of edges. $EDC/(BC*(BC-1))$
Inward dependencies (ID)	# dependencies from other issues.
Outward dependencies (OD)	# dependencies on other issues.
Total dependencies(TD)	Sum of ID and OD.
In-dependencies type (IDT)	# types of Inward dependencies.
Out-dependencies type(ODT)	# types of Outward dependencies.
Total type (TT)	Sum of IDT and ODT.
In-cross dependencies (ICD)	# dependencies from issues in other project.
Out-cross dependencies (OCD)	# dependencies on issues in other projects.
Total cross dependencies (TCD)	Sum of ICD and OCD.

データセット

Vieiraらによって作成されたデータセット

2009年から2018年までの55の大規模なオープンソースプロジェクトのバグレポートが含まれる

フィルタリング

他の問題と関係があるバグの数が占める割合が多いプロジェクトを選択

依存関係メトリクスを収集

NetworkX Pythonライブラリを使用し依存関係メトリクスを収集

RQ 1 結果

ピアソン相関係数を使用して相関関係を調査

・ 依存関係メトリクスの大部分が
有意な正の相関がある

・ グレーのメトリクスはバグの
複雑さを反映できていない
可能性がある

TABLE V
PEARSON CORRELATION BETWEEN THE BUG-FIXING TIME AND BUG DEPENDENCY METRICS.

type	metrics	+	-	/	Derby	Hadoop	Hbase	Hdfs	Hive	Solr	Spark	Ww	Yarn
Network analysis metrics	Size	9	0	0	0.253***	0.139***	0.120***	0.128***	0.160***	0.178***	0.217***	0.128***	0.287***
	Ties	9	0	0	0.251***	0.139***	0.120***	0.128***	0.159***	0.177***	0.216***	0.128***	0.285***
	Pairs	9	0	0	0.141**	0.133***	0.115***	0.119***	0.122***	0.147***	0.214***	0.110***	0.281***
	Density	7	1	1	0.061*	/	0.065***	0.041**	0.099***	0.114***	0.120***	0.104***	-0.217***
	WeakComp	9	0	0	0.254***	0.139***	0.120***	0.130***	0.163***	0.178***	0.217***	0.124***	0.295***
	nWeakComp	9	0	0	0.219***	0.100***	0.099***	0.100***	0.147***	0.167***	0.181***	0.123***	0.225***
	2StepReach	3	0	6	/	/	0.003**	/	/	0.040**	0.025**	/	/
	ReachEffic	2	0	7	/	/	0.039***	/	/	0.028**	/	/	/
	Broker	0	0	9	/	/	/	/	/	/	/	/	/
	nBroker	1	0	8	/	/	/	/	/	/	0.028**	/	/
	EgoBetw	8	0	1	0.084***	0.112***	0.105***	0.069***	0.101***	0.088***	0.161***	/	0.268***
	nEgoBetw	9	0	0	0.186***	0.122***	0.107***	0.091***	0.112***	0.095***	0.162***	0.067*	0.267***
	Dgree	9	0	0	0.253***	0.139***	0.120***	0.128***	0.160***	0.178***	0.217***	0.128***	0.287***
	EffecSize	9	0	0	0.205***	0.126***	0.107***	0.135***	0.136***	0.118***	0.17***	0.108***	0.302***
	Constraint	5	0	4	/	/	0.041***	0.061***	0.056***	0.070***	0.054***	/	/
	Efficiency	9	0	0	0.146***	0.08***	0.072***	0.106**	0.101***	0.098**	0.102***	0.081**	0.131***
	Hierarhcy	8	0	1	0.159***	0.066***	0.086***	0.076***	0.13***	0.142***	0.154***	0.112***	/
	DCentrality	8	0	1	0.166***	0.070***	0.086***	0.074***	0.126***	0.146***	0.160***	0.128***	0.088***
	Closeness	7	0	2	0.107***	0.037*	0.063***	/	0.097***	0.108***	0.142***	0.089**	/
	Eigenvector	5	0	4	/	/	0.041***	/	0.060***	0.093***	0.109***	0.074**	/
	BetwCentrality	8	0	1	0.159***	0.100***	0.066***	0.097***	0.094***	0.09***	0.116***	/	0.181***
Dependency counting metrics	BC	9	0	0	0.154***	0.101***	0.108***	0.108***	0.125***	0.143***	0.196***	0.118***	0.183***
	EDC	9	0	0	0.155***	0.101***	0.108***	0.107***	0.123***	0.144***	0.197***	0.121***	0.181***
	DD	9	0	0	0.143***	0.097***	0.105***	0.103***	0.111***	0.128***	0.191***	0.114***	0.186***
	ID	9	0	0	0.203***	0.110***	0.093***	0.068***	0.130***	0.156***	0.186***	0.065*	0.128***
	OD	9	0	0	0.194***	0.102***	0.091***	0.127***	0.114***	0.106***	0.139***	0.123**	0.207***
	IDT	9	0	0	0.189***	0.104***	0.092***	0.050***	0.119***	0.159***	0.178***	0.079***	0.103***
	ODT	9	0	0	0.146***	0.101***	0.084***	0.115***	0.116***	0.117**	0.121***	0.121***	0.178***
	ICD	7	0	2	/	0.047***	0.052***	0.038**	0.033***	/	0.052***	0.038***	0.089***
	OCD	7	0	2	/	0.073***	0.075***	0.054***	0.053***	0.048**	0.037***	/	0.121***
	TD	9	0	0	0.256***	0.139***	0.119***	0.128***	0.161***	0.176***	0.216***	0.128***	0.214***
	TT	9	0	0	0.229***	0.136***	0.116***	0.109***	0.157***	0.188***	0.199***	0.134***	0.181***
	TCD	7	0	2	/	0.080***	0.083***	0.060***	0.058***	0.050***	0.040***	/	0.134***

+:有意な正の相関 -:有意な負の相関 /: 非有意な相関

RQ 2 従来のメトリクスに対して冗長ではないか

RQ2結果

主成分分析を用いて冗長でないか判断

- Hadoopデータセットでは
主成分の半分が依存関係メトリクス

- すべてのプロジェクトで
第一主成分が依存関係メトリクス

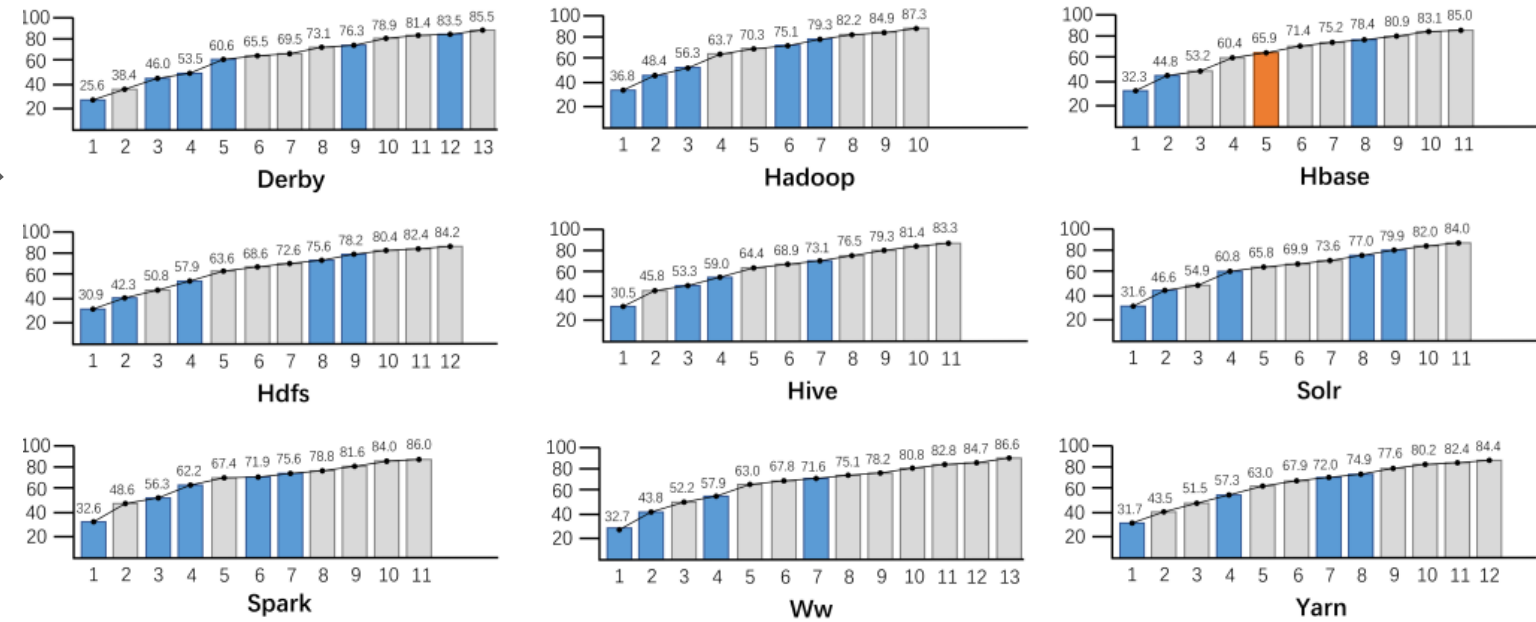


Fig. 6. The rotated components on nine datasets from PCA.

青：依存関係 灰色：従来 オレンジ：両方のタイプ

→依存関係メトリクスはバグ他の側面をとらえることができる

RQ3バグ依存関係は予測の精度を向上させるのに有用か

RQ 3 結果

重回帰分析を行う

従来のメトリクスのみを用いた場合と両方を含めた場合とで比較する

MAE : 平均絶対誤差

MSE: 平均二乗誤差

RMSE: 平方根平均二乗誤差

ほとんどの場合で両方を含めた方が
良い結果を示している

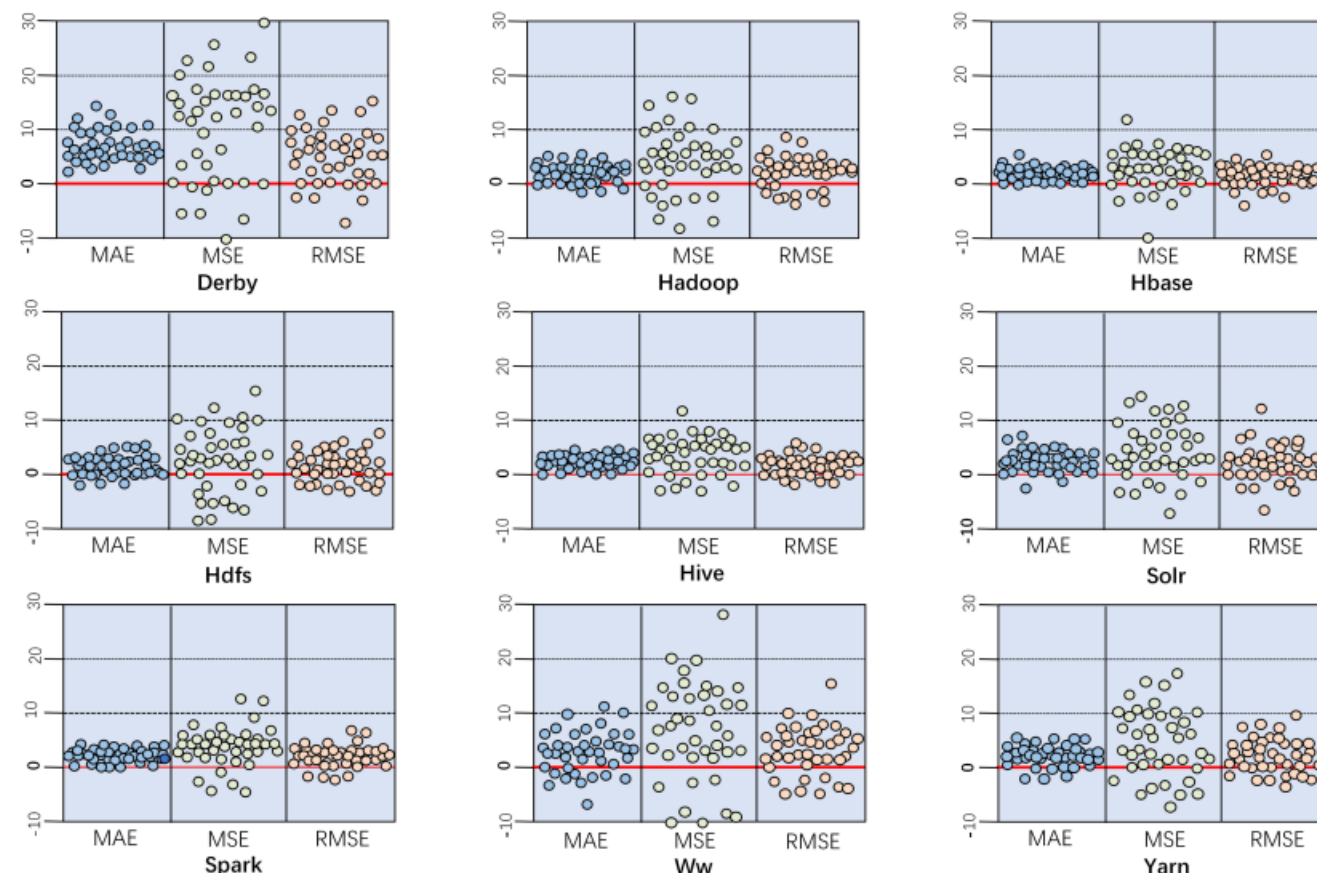


Fig. 7. The result of comparison in MAE, MSE and RMSE

→依存関係を考慮することが従来のメトリックのみを使用するよりも
優れたパフォーマンスを示した

総括

依存関係が修正時間の予測に与える影響を調べた

ほとんどの依存関係メトリクスが修正時間と正の相関があることがわかった

依存関係メトリクスは従来のメトリクスと補完関係にある

依存関係を考慮することで修正時間予測の有効性を向上させることが分かった

