

The Review Linkage Graph for Code Review Analytics

Toshiki Hirao, Shane McIntosh
Akinori Ihara, Kenichi Matsumoto

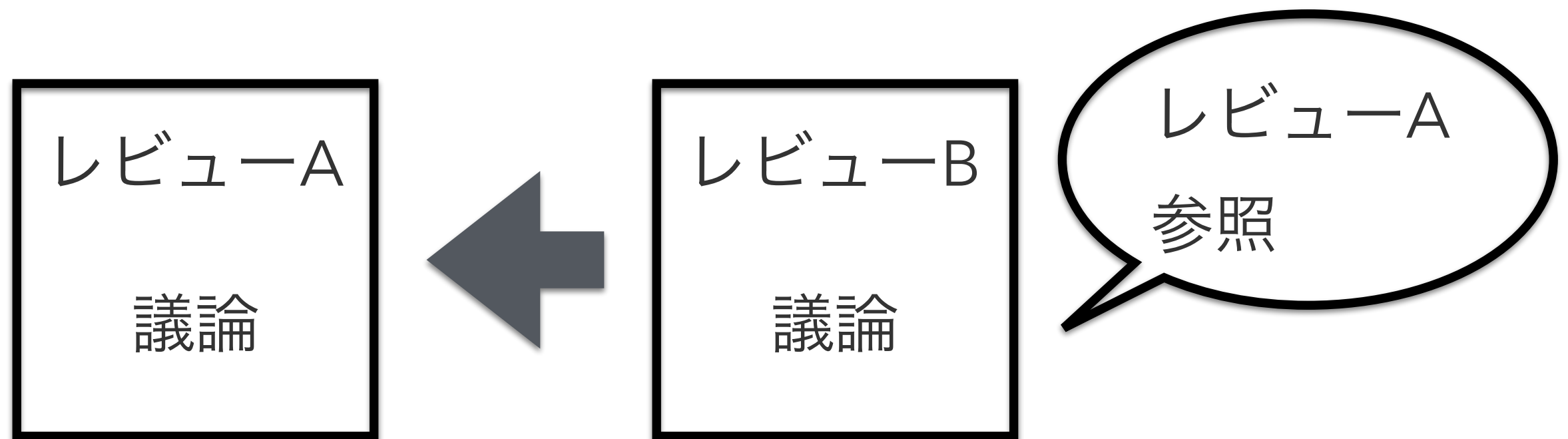
T. Hirao, S. McIntosh, A. Ihara, and K. Matsumoto, "The review linkage graph for code review analytics: a recovery approach and empirical study," in Proceedings of the 2019 27th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering - ESEC/FSE 2019, Tallinn, Estonia, 2019, pp. 578-589.

Modern Code Reviewとは

- オンライン上で行う，軽量版コード検査のこと
- 多くのコミュニティで活用されており，レビューデータが蓄積されている
- MCRのデータを用いてレビュアーの推薦をすることでコードレビュー改善の役に立つ

これまでのコードレビュー分析

- 各レビューは独立したものと扱われていた
- 実際はレビュー間に関係が存在する場合がある



本研究の目的

- 目的
 - レビュー間の関係（レビューリンク）がコードレビュー分析に与える影響の理解
- アプローチ
 - 6コミュニティからレビューリンクのグラフを作成し、分析する

- RQ1 レビューはどの程度リンクされているか
- RQ2 なぜレビューがリンクされているか
- RQ3 カテゴリがどの程度自動化でリンクできるか
- RQ4 リンクレビューはコードレビュー分析にどの程度影響するか

対象コミュニティ

- Gerritでコードレビューされているコミュニティを対象に実験

Table 1: An overview of our subject communities.

Product	Language	Studied Period	#Reviews	#Revs	#Projects
OPENSTACK	Python	09/2011–01/2018	533,050	12,359	1,804
CHROMIUM	JavaScript	04/2011–01/2018	364,079	7,442	410
ANDROID	Java	10/2008–01/2018	229,210	7,614	1,049
QT	C++	07/2011–01/2018	188,981	2,377	170
ECLIPSE	Java	04/2012–01/2018	106,515	2,191	380
LIBREOFFICE	Python	04/2012–01/2018	44,867	822	34
Total		31.8 years	1,466,702	32,805	3,847

リンク方法

- ランダムに410件のレビューを抽出し目視し， 2パターンのリンク方法を発見
 - Change ID
 - URL
- 正規表現を用いてコードの説明やコメントから自動的にリンクする

RQ1 レビューはどの程度リンクされているか

- OpenStack, Chromiumでのリンク率が高かった
- コードレビューに注力しているコミュニティほどリンクされやすい？

Table 2: Review graph characteristics in the subject communities.

Product	Linked Reviews		Per Review (Mean)		Per Reviewer Comments	Description		General Comment		Inline Comment		Cross-project	
	#Reviews	%Reviews	Comments	Reviewers		#Links	%Links	#Links	%Links	#Links	%Links	#Links	%Links
OPENSTACK	133,650	25%	17	5	3	77,400	28%	151,411	54%	50,110	18%	110,964	40%
CHROMIUM	62,065	17%	12	3	4	10,295	5%	182,929	93%	3,968	2%	44,869	23%
ANDROID	11,584	5%	8	2	3	6,082	22%	21,491	77%	466	2%	2,508	9%
QT	14,266	8%	8	2	3	574	3%	19,450	87%	2,309	10%	4,353	19%
ECLIPSE	8,227	8%	6	2	3	1,255	6%	19,995	90%	917	4%	1,359	6%
LIBREOFFICE	1,549	3%	4	2	2	211	10%	1,774	86%	88	4%	107	5%

RQ2 なぜレビューがリンクされているか

- オープンコーディング&ソーティングによってグループとしてまとめる

Table 3: The frequency of the discovered types of review linkage in OPENSTACK NOVA and NEUTRON.

Category	Frequency	
	NOVA	NEUTRON
<i>C1: Patch Dependency</i>	269 (55%)	148 (56%)
Patch Ordering	124 (26%)	62 (24%)
Root Cause	50 (11%)	28 (11%)
Shallow Fix	6 (2%)	4 (2%)
Follow-up	28 (6%)	29 (11%)
Merge Related Reviews	19 (4%)	13 (5%)
Multi-part	42 (9%)	12 (5%)
<i>C2: Broader Context</i>	96 (20%)	50 (19%)
Related Feedback	43 (9%)	14 (6%)
Demonstration	29 (6%)	26 (10%)
Additional Evidence	24 (5%)	10 (4%)
<i>C3: Alternative Solution</i>	69 (14%)	39 (15%)
Superseding	35 (8%)	17 (7%)
Duplicated	34 (7%)	22 (9%)
<i>C4: Version Control Issues</i>	27 (6%)	17 (6%)
Integration Concern	15 (4%)	13 (5%)
Gerrit Misuse	5 (2%)	2 (1%)
Revert	7 (2%)	2 (1%)
<i>C5: Feedback Related</i>	23 (5%)	10 (4%)
Fix Related Issues	11 (3%)	3 (2%)
Feedback Inspired Reviews	12 (3%)	7 (3%)

- パッチ依存のリンクが多く、レビューアー推薦に影響する可能性がある

RQ3 カテゴリがどの程度自動化でリンクできるか

- ブートストラップ法を使用(1000回)
- 5つの分類器を使用し, ZeroRとの精度を比較

Table 4: The performance of our five link category classifiers, all of which outperform the ZeroR and random guessing baselines in all cases.

Model	NOVA				NEUTRON			
	Prec.	Rec.	F1.	AUC	Prec.	Rec.	F1.	AUC
SVM	0.75	0.88	0.79	0.72	0.72	0.92	0.77	0.82
RF	0.77	0.72	0.71	0.57	0.76	0.84	0.72	0.89
MNB	0.71	0.80	0.74	0.72	0.71	0.82	0.75	0.76
MLP	0.74	0.81	0.76	0.61	0.75	0.84	0.74	0.68
MLR	0.75	0.88	0.79	0.65	0.72	0.92	0.77	0.79
ZeroR	0.26	0.50	0.34	–	0.27	0.50	0.34	–

RQ4 リンクレビューはコードレビュー分析にどの程度影響するか

- レビューアー推薦時にメトリクスを組み込むことで精度がどのように変動するか
- 予測時間のバッファを増やしてもそれほど変わらない

Table 5: The mean performance scores of cHRev [48] and our proposed link-aware reviewer recommenders at different prediction delays. The bulk of the performance improvement is achieved in the no delay (0 hour) setting.

Model	NOVA									NEUTRON								
	Top 1			Top 3			Top 5			Top 1			Top 3			Top 5		
	Pre.	Rec.	F1.	Pre.	Rec.	F1.	Pre.	Rec.	F1.	Pre.	Rec.	F1.	Pre.	Rec.	F1.	Pre.	Rec.	F1.
Baseline	0.25	0.06	0.09	0.28	0.18	0.22	0.30	0.32	0.30	0.24	0.05	0.08	0.26	0.17	0.20	0.28	0.31	0.29
0 hour	0.39	0.09	0.14	0.40	0.26	0.31	0.40	0.43	0.41	0.39	0.09	0.14	0.40	0.26	0.31	0.41	0.44	0.42
1 hour	0.39	0.09	0.14	0.40	0.26	0.31	0.40	0.43	0.41	0.41	0.09	0.14	0.41	0.26	0.32	0.42	0.44	0.43
3 hours	0.40	0.09	0.14	0.40	0.26	0.31	0.40	0.43	0.41	0.41	0.09	0.14	0.41	0.26	0.32	0.42	0.45	0.43
6 hours	0.40	0.09	0.14	0.40	0.26	0.31	0.40	0.43	0.41	0.41	0.09	0.15	0.41	0.27	0.32	0.42	0.45	0.43

まとめと所感

- まとめ
 - 6つのコミュニティにおいてレビューリンクの存在は珍しくない
 - リンクをレビュー分析に組み込むとパフォーマンスが向上する
- 所感
 -