

モジュールごとの適性を考慮した コミッター候補者予測

M3 山谷 陽亮

2016/06/14

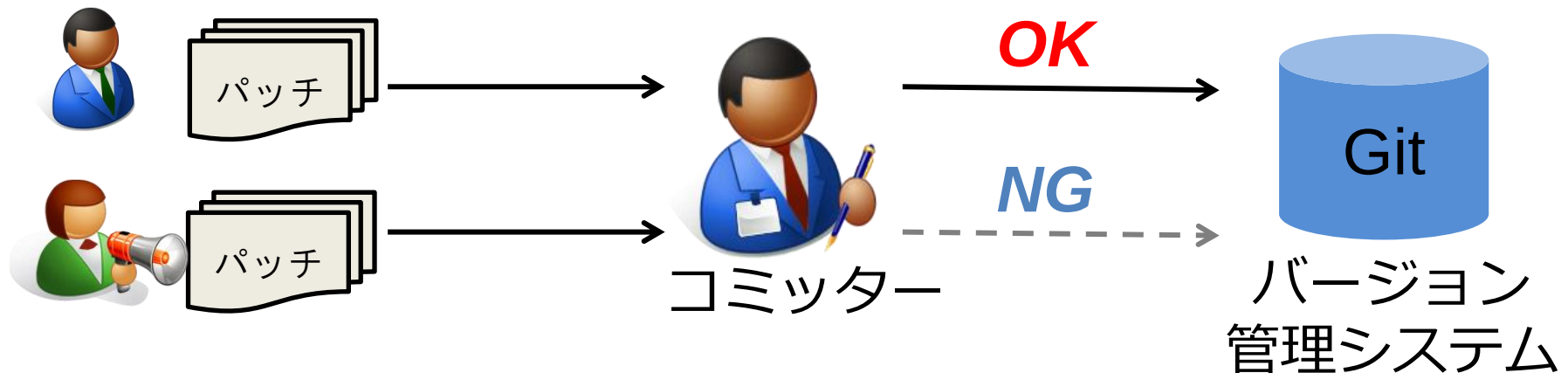
背景

- オープンソースソフトウェア(OSS)の普及によりOSSのユーザーが年々増加している
- 大規模OSSプロジェクトでは日々大量の不具合が報告されている[1]
- 報告された不具合はボランティアの開発者が中心となり修正が行われている

[1] John Anvik, Lyndon Hiew, and Gail C. Murphy. Who should x this bug?
In Proceedings of the 28th International Conference on Software Engineering (ICSE'06), pp. 361--370, 2006.

不具合修正プロセス

- 不具合を修正するためにパッチが投稿される
 - パッチ：コードの修正前後の差分を示すファイル
- コミッターによってパッチが検証される
 - コミッター：コードを変更する権限を持つ開発者



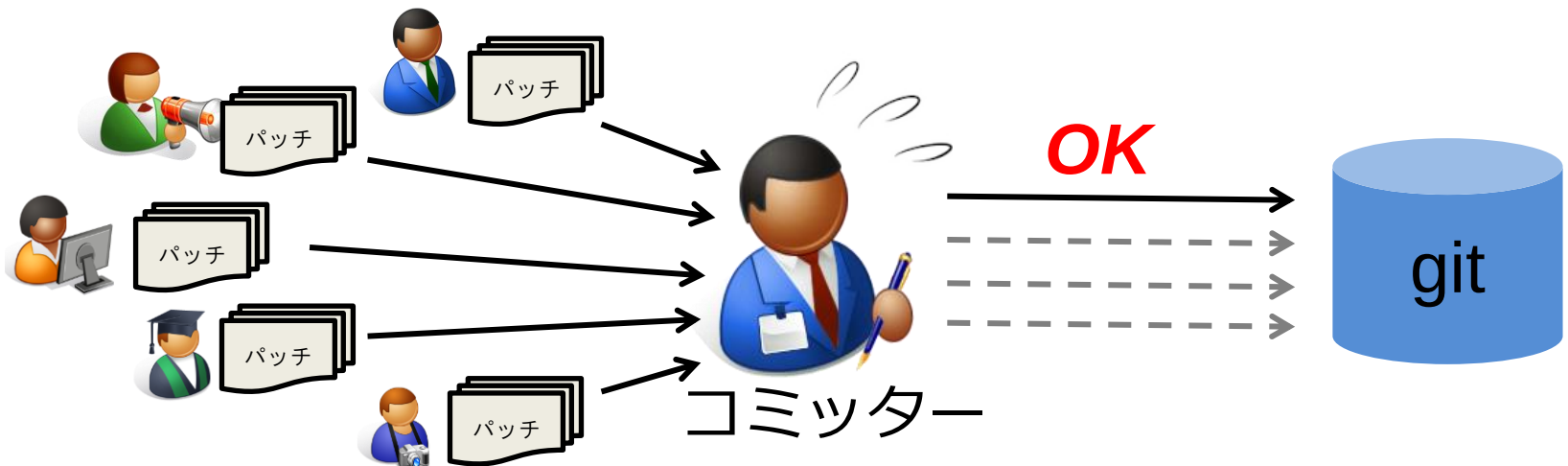
コミッター昇格

- コミッターはソフトウェアの品質管理の面で重要な役割を担っている
- 開発者にコミット権限を付与するかどうかの判断は慎重に行われ、既存のコミッターからの推薦や承認によってコミット権限が付与される [2]
- コミット権限を持たない開発者に比べコミッターは少人数で構成されている

[2] C. Jensen and W. Scacchi. Role Migration and Advancement Processes in OSSD Projects: A Comparative Case Study. In Proceedings of the 29th International conference on software engineering (ICSE'07), pp.364-374, 2007

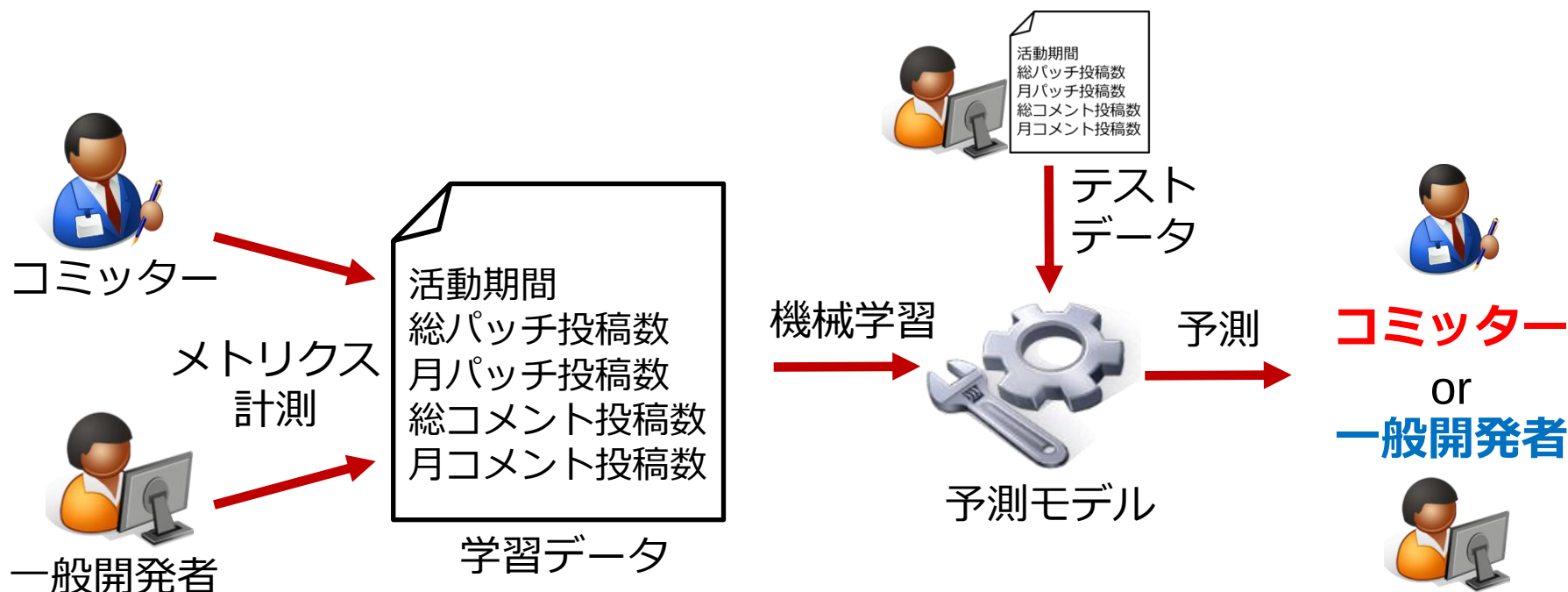
動機：コミッター不足

- 大規模OSSプロジェクトでは少人数で構成されるコミッターに過度な負担がかかっている
- コミッターに過度な負担がかかるとパッチの検証を効率的に行うことができなくなる



先行研究

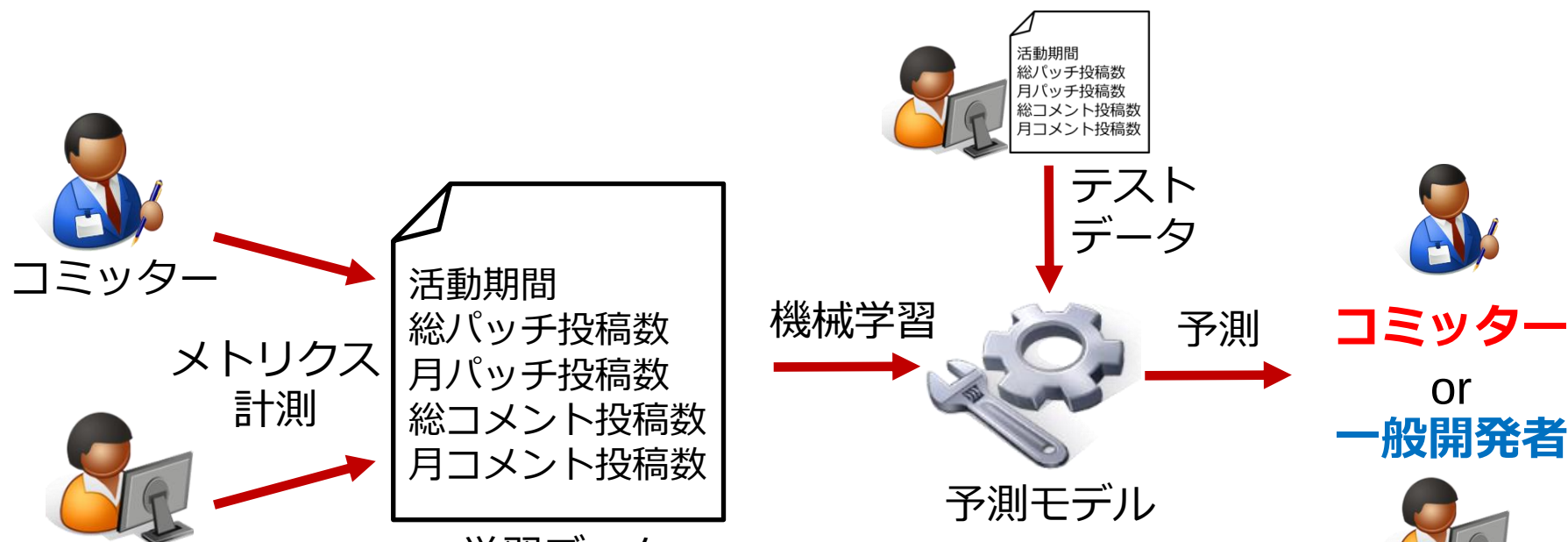
- コミッター不足を解消するためにコミッターに昇格可能な開発者を予測する研究が行われる[3]



[3] Akinori Ihara, Yasutaka Kamei, Masao Ohira, Ahmed E. Hassan, Naoyasu Ubayashi, and Kenichi Matsumoto. Early identification of future committers in open sourcesoftware projects. In Proceedings of the 14th International Conference on Quality Software (QSIC'14), pp. 47--56, 2014.

先行研究

- コミッター不足を解消するためにコミッターに昇格可能な開発者を予測する研究が行われる[3]

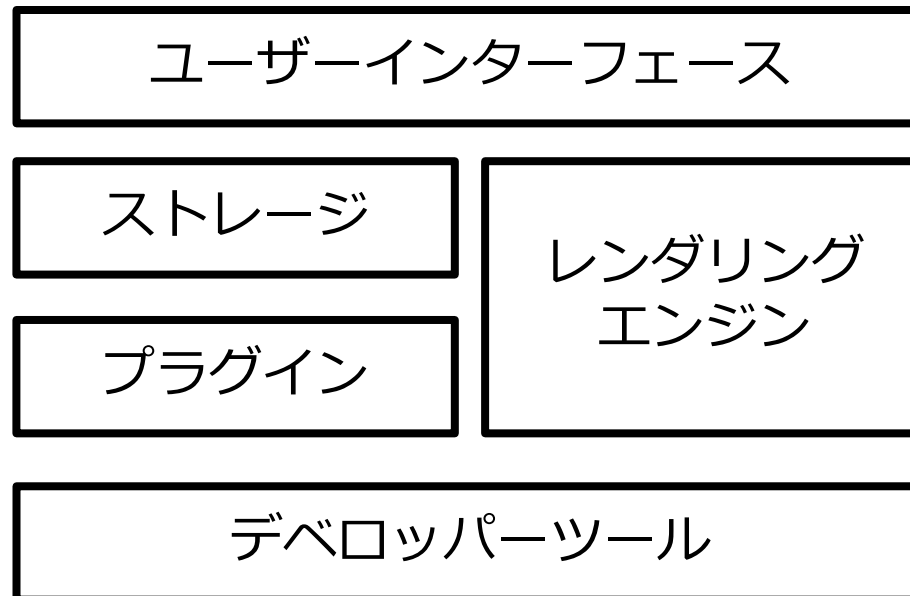


不具合修正の効率化のためには
モジュールごとにコミッターを予測する必要がある

ソフトウェアモジュール

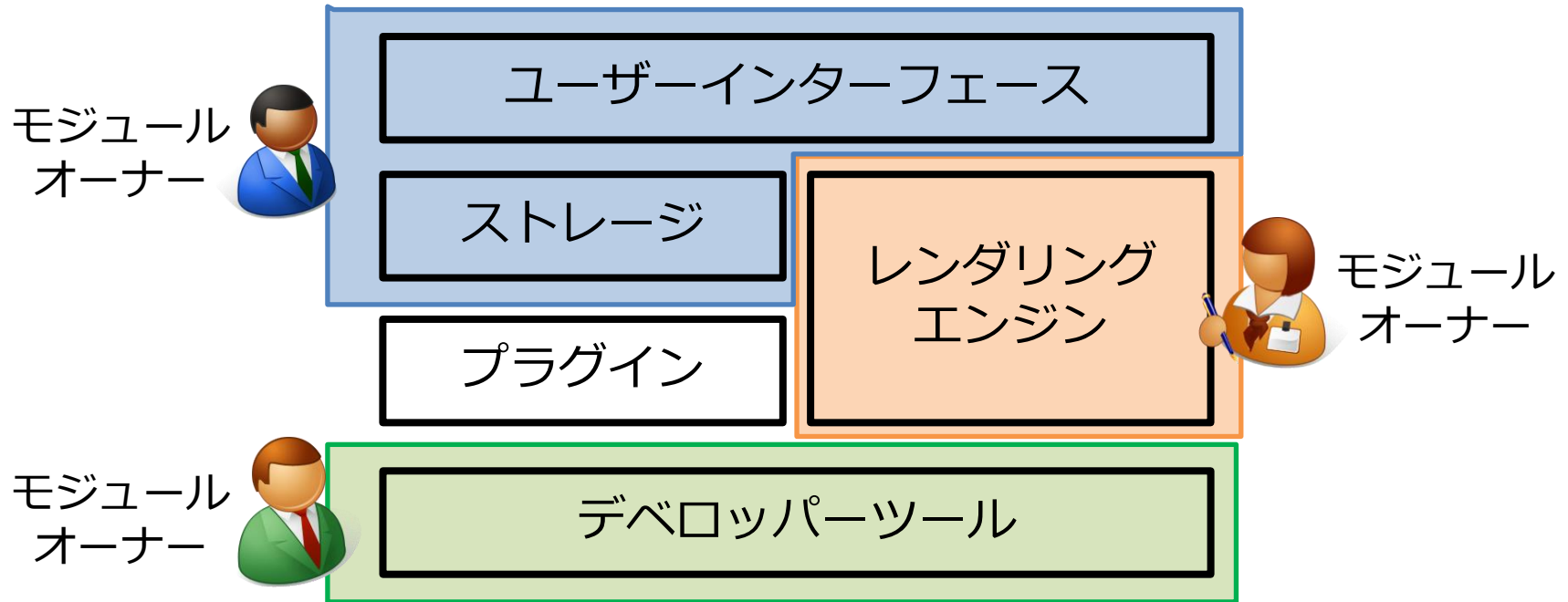
- ソフトウェアは保守性や再利用性を高めるために複数のモジュールが組合わさり構成される
 - モジュール：ひとまとまりの機能をもつ部品

例：ウェブブラウザのモジュール



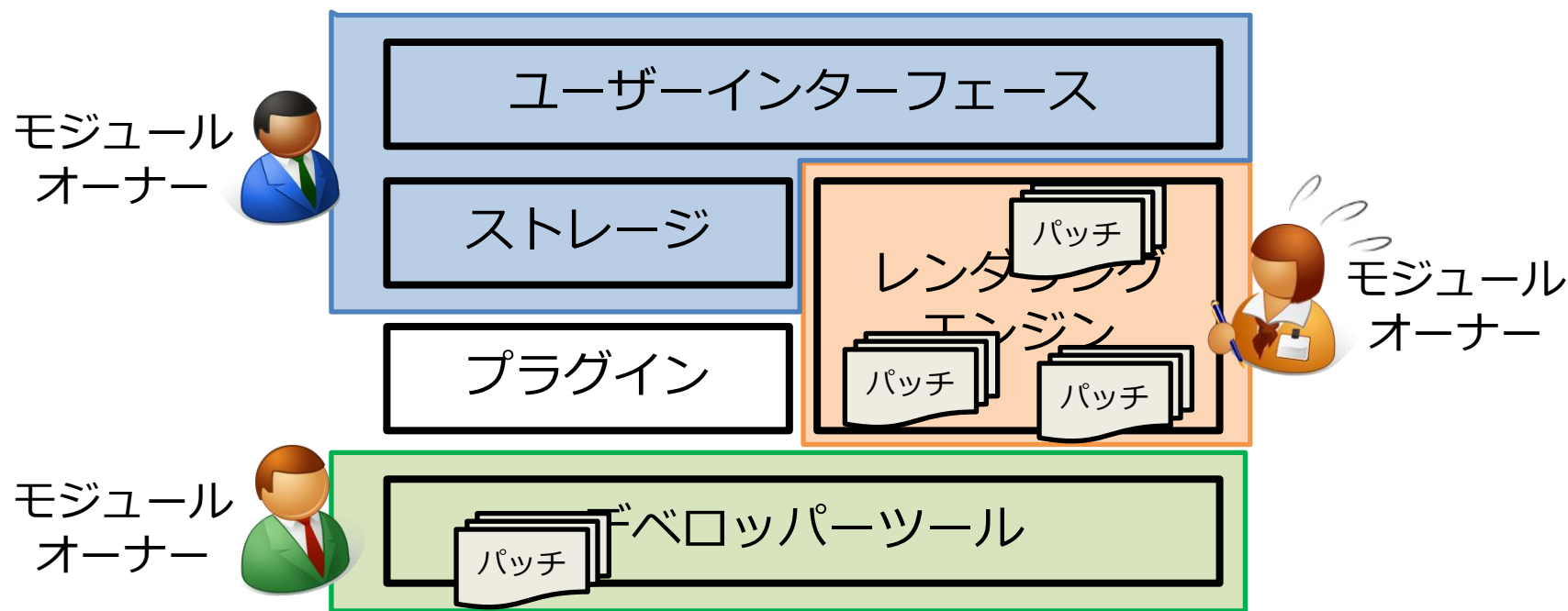
モジュールオーナー

- 大規模OSSプロジェクトでは各コミッターが担当すべきモジュールが決められている
 - モジュールオーナー：**
 - 特定のモジュールを担当するコミッター



モジュールオーナー候補者

- 不具合修正の効率化には人手が不足している**モジュールのオーナーに適した開発者（モジュールオーナー候補者）**を推薦することが必要となる



本研究の目的

- 目的
 - モジュールオーナーの不足を解消するために、モジュールオーナー候補者を推薦する
- アプローチ
 - モジュールオーナーの適性を示すメトリクスを定義
 - モジュールオーナー候補者を特定するための予測モデルを構築

リサーチクエスチョン

- RQ：モジュールごとの適性を考慮して，コミッター候補者を推薦することができるのか？
- RQ1：技術的貢献に差があるか？
- RQ2：社会的貢献に差があるか？
- RQ3：技術的・社会的貢献のある開発者は周りからの評判がよいのでは？
- RQ4：依存関係の強いモジュールでの評判が高い開発者は，特定のモジュールを担当するコミッターになりやすいのか？
- RQ5：技術的・社会的貢献と評判から特定のモジュールを担当するコミッターを予測できるのか？
- RQ6：開発者の活動初期の評判から特定のモジュールを担当するコミッターを予測することができるのか？

リサーチクエスチョン

- RQ1：技術的貢献に差があるのでは？
- 動機
 - コミッターに求められる資質として技術的な知識が挙げられているため
- アプローチ
 - パッチ投稿に関するメトリクスを計測する
 - 投稿したパッチの数（量的な貢献）
 - 重大な不具合に対するパッチ投稿数（質的な貢献）

リサーチクエスチョン

- RQ2：社会的貢献に差があるのでは？
- 動機
 - コミッターに求められる資質として社会的な貢献が挙げられているため（マネジメント能力的な）
- アプローチ
 - コメント投稿に関するメトリクスを計測する
 - 投稿したコメントの数（量的な貢献）
 - コミッターとやり取りを行っているか？（質的な貢献）

リサーチクエスチョン

- RQ3：技術的・社会的貢献のある開発者は周りからの評判がよいのでは？
- 動機
 - コミッターになるには既存のコミッター等から信頼を得ている必要がある
- アプローチ
 - 周りからの評判を定量化する
 - Google Playストアのレビューデータから、レビューテキストの学習をし、Ratingの値を返す

リサーチクエストション

- RQ4：依存関係の強いモジュールでの評判が高い開発者は、特定のモジュールを担当するコミッターになりやすいのか？
- 動機
 - 依存関係にあるモジュールから人が移ることがある
- アプローチ
 - モジュール間の依存関係を定量化する

Major-Minor-Dependency Relationship

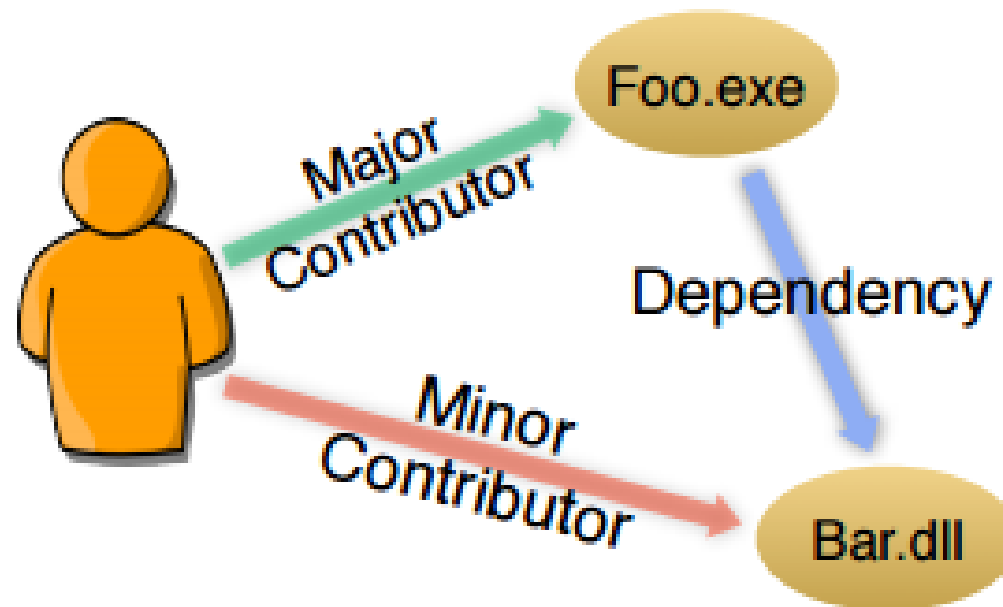


Figure 3: Illustration of the major-minor-dependency relationship commonly observed in Vista

リサーチクエスチョン

- RQ5：技術的・社会的貢献と評判から特定のモジュールを担当するコミッターを予測できるのか？
- 動機
 - 先行研究でも予測が行われており，比較するため
- アプローチ
 - 予測を行い，予測に寄与する変数を特定する

リサーチクエスチョン

- RQ6：開発者の活動初期の評判から特定のモジュールを担当するコミッターを予測することができるのか？
- 動機
 - OSS開発者は離脱する可能性もあるのでできるだけ早くコミッターになるか否かを知りたい
- アプローチ
 - 各開発者の活動初期のメトリクスのみを用いて予測する（1か月のみ，3か月のみ，など）

これまでやってきたこと

- モジュールの定義の再検討
- 評判の定量化

モジュールの定義の再検討

- 問題点

- 前は、コミットテキストに含まれるバグ番号とBTSのコンポーネント情報を基にひもづけを行っていたので、コミットテキストにバグ番号が含まれない場合、ひもづけが不可能であった

- 改善点

- コミットテキストに含まれるバグ番号とBTSのコンポーネント情報、変更されたファイルパスから、ソースコードの構造からモジュールを分割する

Eclipse Platform

モジュール名	判別方法
UI	• bundles/org.eclipse.ui~
Workspace	• bundles/org.eclipse.team~ • bundles/org.eclipse.core.resources~ • bundles/org.eclipse.compare~
SWT	• bundles/org.eclipse.swt~
Releng	• bundles/org.eclipse.releng • ~.sh
Debug	• org.eclipse.debug~

Mozilla Firefox

モジュール名	判別方法
App	• browser/app
Tab	•
Preferences	• browser/themes
General	•
DeveloperTools	• browser/devtools
Bookmark	• browser/components/places
Bar	

GCC

モジュール名	判別方法
C	• gcc/c~ • libstdc~ • libcpp~
BackEnd	• gcc/config~
MiddleEnd	• gcc/tree~ • libgomp~
Fortran	• gcc/fortran~ • libfortran~

GIMP

モジュール名	判別方法
General	• app/core • app/base
UserInterface	• app/actions • app/display • app/gui • app/tools
Tools	• app/paint • app/widgets • app/text
plug-ins	• plug-ins/ • app/plug-in

http://wiki.gimp.org/wiki/Hacking:Source_Tree#The_app_directory

WebKit

モジュール名	判別方法
JavaScriptCore	• Source/JavaScriptCore
WebCore	• Source/WebCore/platform • Source/WebCore/dom • Source/WebCore/bindings • Source/WebCore/page • Source/WebCore/loader
WebKit2	• Source/WebKit2
Build	• WebKitTools/
API	消した
Inspector	• Source/WebCore/inspector
Layout	• Source/WebCore/rendering • Source/WebCore/css • Source/WebCore/svg

評判の定量化

- 概要
 - コミッターから肯定的な意見を得ている開発者を特定するためにセンチメンタル分析を行う
- 分析方法
 - Google Play StoreのレビューデータのRatingからテキストとRatingのデータを作る
 - メリット
 - 5段階でRatingがついているからわかりやすい
 - より一般的な結果を得ることができる

分析手順

1. データの取得

- ICSE'14で発表されているオープンデータを取得

2. Ratingごと、カテゴリごとのデータを作成

- すべてのカテゴリを含む場合
- Gameをはずした場合
- 対象のカテゴリに絞った場合

3. Bug of Wordsを生成

- can not や don't などを変換し抽出できるように

対象プロジェクトのカテゴリ

- 対象プロジェクトに合わせたカテゴリのレビューデータを取得する際の情報
 - Eclipse : TOOLS
 - Firefox : COMMUNICATION
 - GCC : TOOLS
 - GIMP : PHOTOGRAPHY
 - WebKit : COMMUNICATION

すべてのカテゴリ

review1		review2		review3		review4		review5	
term	time	term	time	term	time	term	time	term	time
game	234	game	251	app	224	app	205	game	222
app	225	app	229	game	187	game	192	app	160
dont	222	dont	174	good	147	good	184	love	154
cant	123	work	138	dont	112	great	107	good	123
work	122	fix	120	work	110	like	101	great	103
get	102	get	115	use	106	love	87	like	72
fix	97	use	111	updat	88	fun	82	nice	65
time	91	time	110	fix	87	nice	79	use	60
play	90	cant	108	get	84	work	72	best	57
updat	85	updat	108	like	83	use	70	awesom	54
even	81	play	95	time	74	time	58	fun	47
download	78	like	90	pleas	69	dont	54	just	45
just	73	good	86	cant	65	get	51	work	44
pleas	73	just	84	need	65	play	48	can	41
like	68	pleas	80	star	60	need	47	play	41
use	63	now	73	can	59	one	46	one	34
now	62	tri	69	great	57	cool	43	excel	33
phone	62	one	64	just	56	fix	43	get	33

Gameなし

review1		review2		review3		review4		review5	
term	time	term	time	term	time	term	time	term	time
app	348	app	323	app	278	app	314	app	279
dont	233	dont	170	good	138	good	183	good	145
work	198	work	158	use	120	use	113	love	122
updat	135	fix	138	dont	116	great	104	great	115
cant	128	updat	136	work	111	nice	91	use	98
use	125	use	128	like	96	like	85	work	82
fix	122	cant	114	fix	92	work	84	best	63
time	104	pleas	102	cant	86	love	75	nice	63
now	92	get	93	updat	80	dont	65	like	51
tri	80	time	84	get	69	get	62	awesom	47
get	77	now	82	need	68	star	57	excel	43
even	76	like	79	pleas	67	just	56	can	34
uninstal	75	need	79	can	65	one	53	one	33
pleas	70	phone	79	great	65	can	51	realli	33
say	61	good	78	time	62	updat	40	thank	32
want	61	just	77	just	57	better	39	easi	31
need	60	can	58	now	56	need	38	cool	25
phone	60	star	57	one	55	pleas	36	time	25

COMMUNICATION

review1		review2		review3		review4		review5	
term	time	term	time	term	time	term	time	term	time
app	293	app	280	app	203	app	237	app	250
dont	229	fix	160	good	139	good	192	good	125
work	169	dont	156	use	108	use	90	great	102
cant	155	work	148	cant	104	nice	84	love	102
updat	155	updat	147	dont	102	work	83	use	74
use	138	messag	144	work	102	great	73	best	73
bug	133	use	131	messag	101	like	69	nice	71
phone	111	cant	119	fix	97	dont	67	work	68
fix	103	pleas	99	like	93	love	60	excel	59
now	95	time	98	pleas	87	need	53	awesom	43
get	94	get	93	updat	79	best	48	like	43
messag	81	now	91	get	72	call	48	browser	42
call	80	phone	85	need	70	fix	48	dont	33
time	80	email	82	can	69	one	45	one	33
just	78	send	77	phone	60	pleas	45	realli	33
pleas	77	call	73	call	59	can	44	better	32
like	67	good	73	contact	55	browser	43	thank	32
even	66	can	71	time	55	just	42	can	30

TOOLS

review1		review2		review3		review4		review5	
term	time	term	time	term	time	term	time	term	time
app	306	app	295	app	317	app	286	app	300
work	238	work	241	work	185	good	208	good	133
dont	228	dont	202	good	177	use	126	work	122
phone	104	use	132	dont	143	work	126	great	121
uninstal	101	fix	102	use	131	great	100	use	117
use	90	updat	101	fix	82	nice	82	best	66
time	82	phone	94	updat	81	like	65	love	66
cant	77	cant	89	like	79	dont	64	like	59
just	77	pleas	89	phone	71	phone	61	nice	58
updat	72	like	80	will	70	need	44	excel	49
download	66	good	76	great	63	can	43	easi	48
get	64	time	69	cant	62	get	43	thank	46
now	62	get	68	get	62	time	42	just	42
fix	59	need	67	need	61	just	40	phone	42
even	56	now	63	can	60	thank	37	awesom	38
instal	56	can	62	pleas	60	star	35	perfect	36
say	54	just	59	star	60	awesom	34	realli	33
suck	47	will	59	nice	53	updat	34	one	31

GHOTOGRAPHY

review1		review2		review3		review4		review5	
term	time	term	time	term	time	term	time	term	time
app	308	app	345	app	295	app	288	app	354
dont	192	work	148	good	161	good	162	love	202
work	149	use	130	like	105	love	105	great	100
pictur	111	dont	129	photo	98	nice	105	good	96
cant	104	fix	129	fix	95	great	101	nice	85
use	100	pictur	113	use	92	use	101	use	83
time	92	photo	112	work	88	like	94	best	68
fix	91	cant	108	pictur	82	cool	59	like	64
photo	84	like	103	pleas	80	pictur	59	cool	63
download	72	pleas	95	dont	79	photo	58	photo	63
just	64	good	84	nice	77	realli	54	awesom	53
phone	59	time	75	love	74	easi	51	easi	43
wont	58	just	73	cant	70	make	51	pictur	42
tri	57	phone	68	great	63	work	50	work	41
like	56	get	67	just	63	camera	45	amaz	38
save	54	take	67	can	58	just	42	can	36
suck	53	updat	66	camera	57	dont	41	make	35
uninstal	53	pic	64	realli	57	pleas	40	realli	35

予備調査：評判の違い

- 昇格前のコミッターと一般開発者の周りからの評判（どんなコメントをもらってるか）に違いがあるのかを確かめる
- 調査方法

Dirk Pranke 2013-07-31 11:45:34 PDT

(In reply to [comment #10](#))

```
> It sounds like it need to update it to import into:  
> LayoutTests/csswg/contributors/adobe/submitted/shapes/shape-outside
```

Yup, that's exactly what I had in mind.

調査結果

- ぱっとみてわかるような違いはなかった
- キーワードをしぼると少し違いがみえなくもない
 - good, great などは若干Committerの方が多い