

OSSプロジェクトにおける 開発者の特性を考慮した コミッター候補者予測

大平研究室

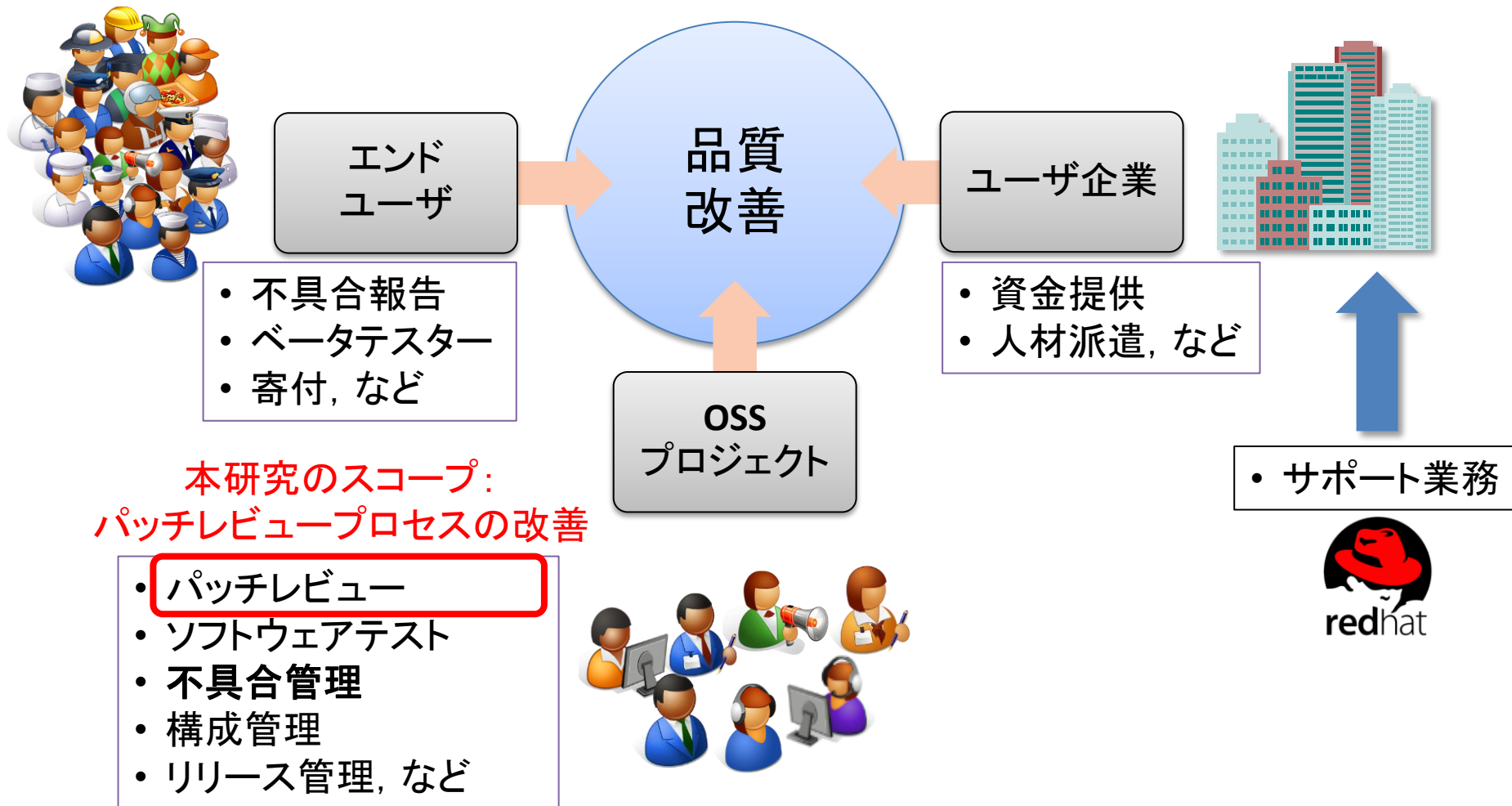
M2 松本 明

背景：OSSの品質に対する社会的懸念

- OSSを活用した製品開発のニーズが高まっている
 - ソフトウェアライフサイクルの短期化
 - グローバル競争の激化（コスト削減要請）
- OSSの品質に対する懸念も深刻化しつつある
 - 不具合報告の増加
 - 不具合修正の長期化（修正版リリースの遅れ）

OSS開発における品質保証活動の支援が求められている

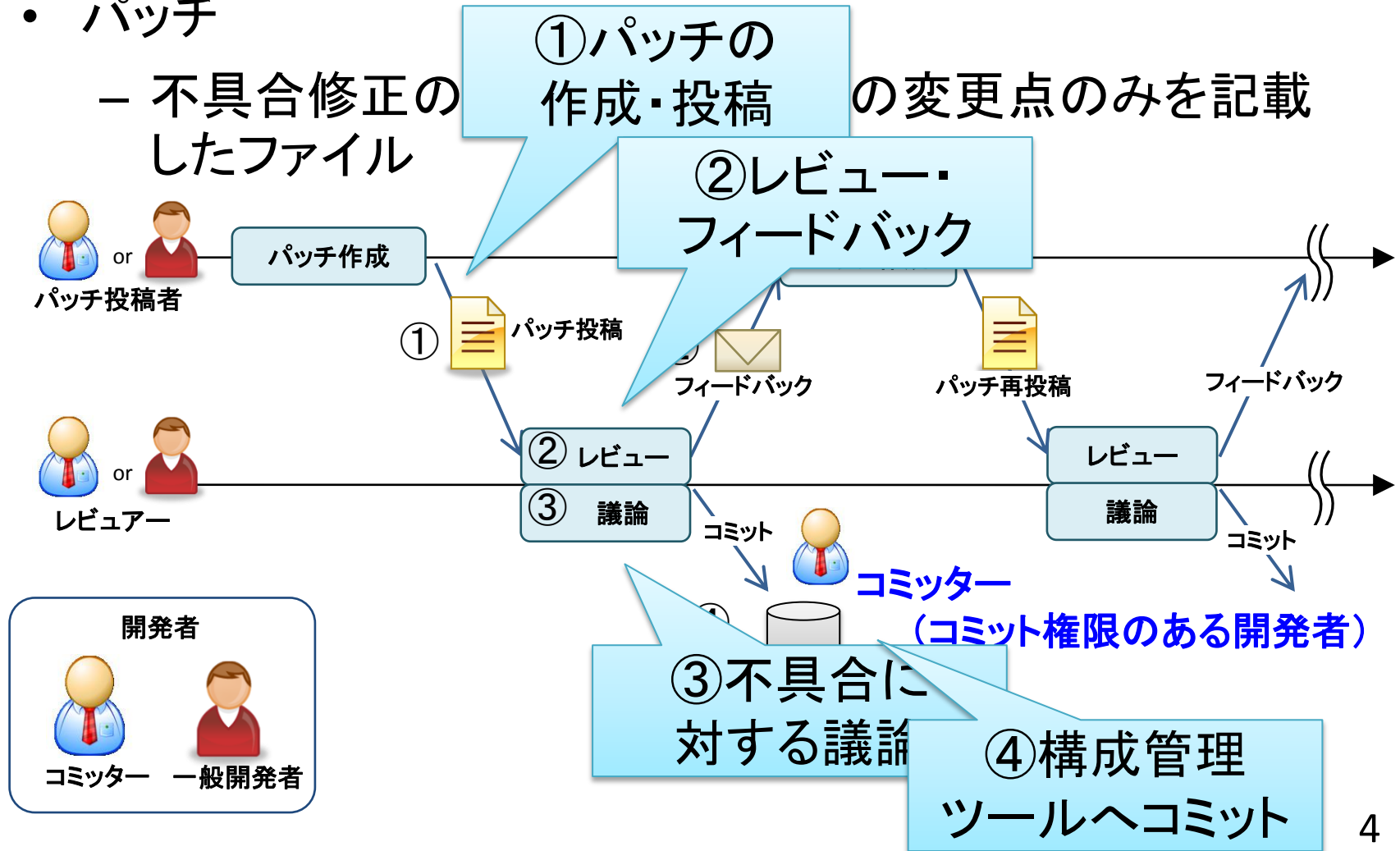
本研究の位置付け: OSS開発における品質保証活動



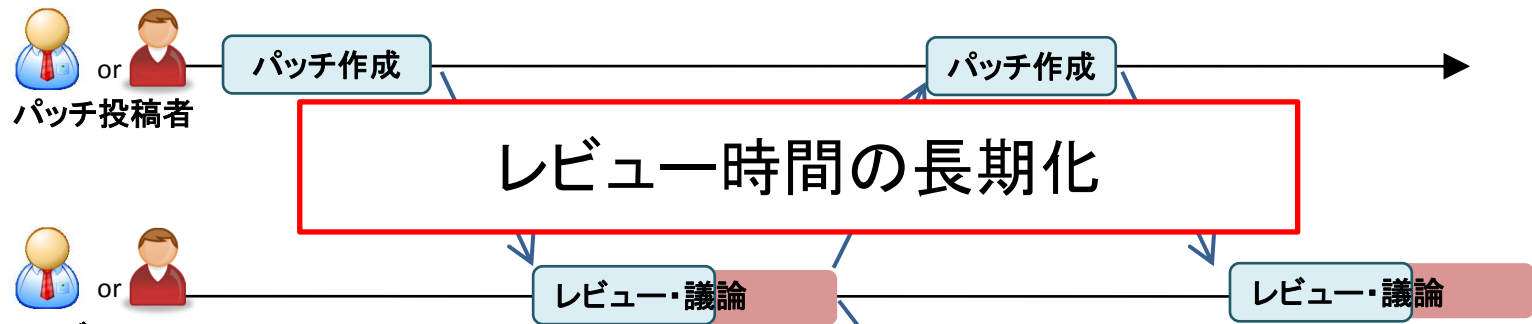
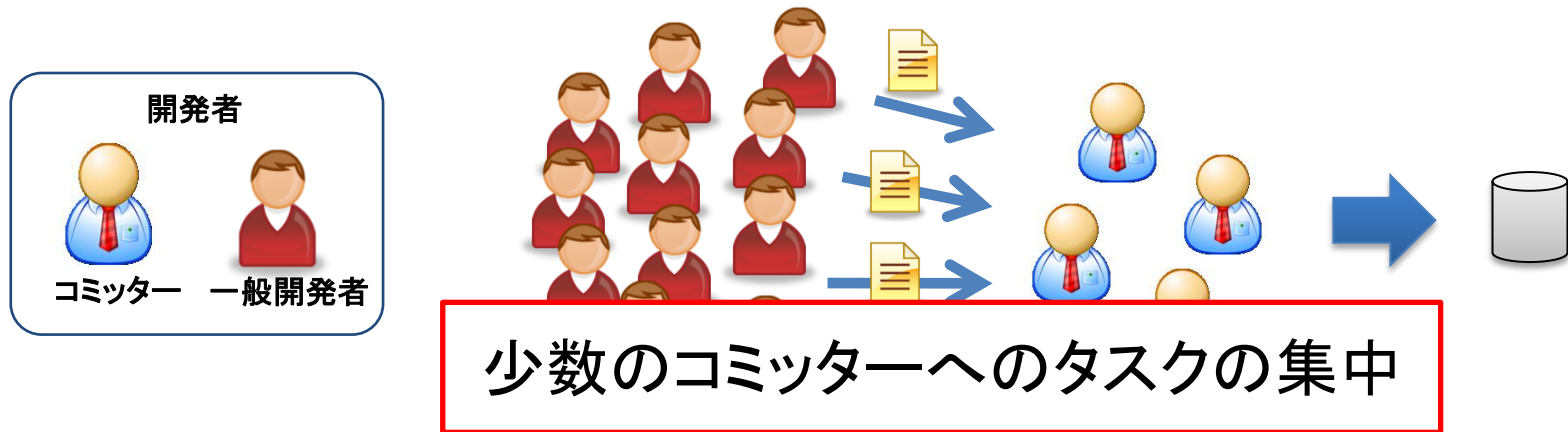
パッチレビュープロセス

- パッチ

- 不具合修正の
したファイル
- ①パッチの
作成・投稿
- ②レビュー・
フィードバック
- の変更点のみを記載



パッチレビュープロセスの問題点



コミッターに昇格すべき開発者を見つけ出すことが重要

コミッターへの昇格

- コミッター選出はプロジェクト管理者や既存のコミッターによって行われる[1]
- プロジェクトへの**貢献**が認められればコミッターとして選出される



①パッチの作成・投稿

②投稿されたパッチのレビュー&フィードバック

③不具合に対する議論(コメントの投稿)

④パッチの妥当性確認後, 構成管理ツールへコミット

- コミッター選出のための客観的指標は存在しない
 - プロジェクト管理者は経験的にコミッター選出を行っている
 - コミッター選出にはコストがかかる

目的とアプローチ

- 目的

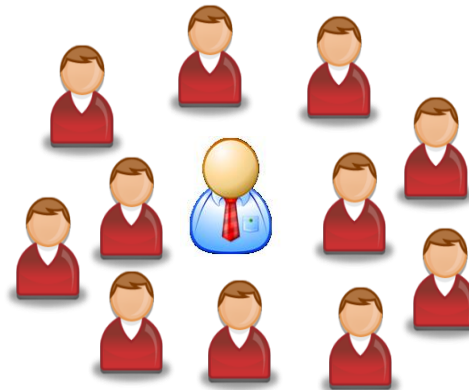
- コミッター候補者の選出による, コミッターのリソースの拡大およびコミッター1人当たりの負荷軽減

- アプローチ

- 既存コミッターの過去の活動とその活動量を分析し, **コミッター候補者予測モデルを構築する**



予測モデル

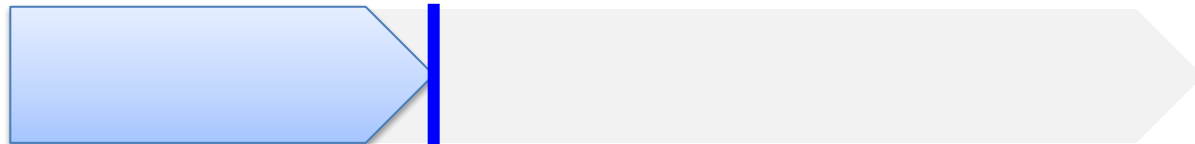


分析

- 不具合管理データを用いて、開発者の活動内容を比較分析する

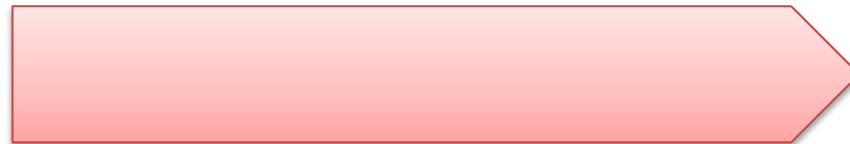
RQ1: コミッター候補者と一般開発者の活動量に差はあるか？

- コミッター昇格前と一般開発者を比較



コミッター候補者

コミッター昇格



一般開発者



分析：活動内容メトリクス

- 活動期間(月)
- 活動量
 - 総パッチ投稿数
 - 総パッチレビュー数
 - 総コメント投稿数
 - 月パッチ投稿数
 - 月パッチレビュー数
 - 月コメント投稿数

一時期のみ積極的に活動した開発者と継続して活動している開発者を区別

ケーススタディ

- パッチの投稿・レビューに不具合管理システムJiraを利用しているOSSプロジェクト, Apache Ambariを分析



Apache Ambari
<http://incubator.apache.org/ambari>

分析期間	2011/9 ~ 2014/12
全コミッター数	47
コミッター候補者数	43
一般開発者数	248

不具合管理システムJira



The Apache Software Foundation
<http://www.apache.org/>

Dashboards ▾ Projects ▾ Issues ▾ Agile ▾

Search 

 ▾ Log In

 Ambari / AMBARI-2412
Disable Hadoop 2.0 stack on trunk

Agile Board 

Details

Type:	 Task	Status:	RESOLVED
Priority:	 Major	Resolution:	Fixed
Affects Version/s:	1.2.5	Fix Version/s:	1.2.5
Component/s:	ambari-web		
Labels:	None		

Description

Until we merge the Hadoop 2.0 work being done on branch-1.4.0, we need to make Hadoop 2.0 stack inactive on trunk.

> Attachments

Activity

All **Comments** Work Log History Activity Transitions 

▼  Mahadev konar added a comment - 19/Jun/13 05:16

+1 for the patch.

▼  Yusaku Sako added a comment - 19/Jun/13 05:20

Committed to trunk

People

Assignee:	 Yusaku Sako
Reporter:	 Yusaku Sako
Votes:	 0 Vote for this issue
Watchers:	 2 Start watching this issue

Dates

Created:	18/Jun/13 08:00
Updated:	19/Jun/13 05:20
Resolved:	19/Jun/13 05:20

Agile

[View on Board](#)

11

不具合管理システムJira

・パッチ投稿

 Jaimin D Jetly made changes - 07/Feb/14 19:49

Field	Original Value	New Value
Attachment		AMBARI-4560.patch [12627684]

・パッチレビュー

▼  Mahadev konar added a comment - 19/Jun/13 05:16

+1 for the patch.

・コメント投稿

▼  Mahadev konar added a comment - 18/Jun/13 06:31

This can be run post ambari-server upgrade to update the ambari server metainfo.

活動期間(月)の比較

	中央値	平均値	分散
コミッター候補者	4	4.77	9.52
一般開発者	1	1.90	2.69

t検定の結果, $P = 0.00 < 0.05$

コミッター昇格前と一般開発者で有意差がある

⇒コミッター選出の指標として有効

パッチ投稿数の比較

総パッチ投稿数

	中央値	平均値	分散
コミッター候補者	30	32.82	555.30
一般開発者	0	4.21	335.87

月パッチ投稿数

	中央値	平均値	分散
コミッター候補者	5.25	7.57	57.89
一般開発者	0	1.08	11.24

t検定の結果, $P = 0.00 < 0.05$ (総パッチ投稿数)
 $P = 0.00 < 0.05$ (月パッチ投稿数)

コミッター昇格前と一般開発者で有意差がある
⇒コミッター選出の指標として有効

パッチレビュー数の比較

総パッチレビュー数

	中央値	平均値	分散
コミッター候補者	1.5	10.27	409.26
一般開発者	0	0.26	2.10

月パッチレビュー数

	中央値	平均値	分散
コミッター候補者	0	2.68	49.44
一般開発者	0	0.08	0.26

t検定の結果, $P = 0.00 < 0.05$ (総パッチレビュー数)
 $P = 0.00 < 0.05$ (月パッチレビュー数)

コミッター昇格前と一般開発者で有意差がある
⇒コミッター選出の指標として有効

コメント投稿数の比較

総コメント投稿数

	中央値	平均値	分散
コミッター候補者	12	15.91	205.61
一般開発者	1	3.20	24.27

月コメント投稿数

	中央値	平均値	分散
コミッター候補者	2	2.91	17.75
一般開発者	1	1.49	5.20

t検定の結果, $P = 0.00 < 0.05$ (総コメント投稿数)
 $P = 0.02 < 0.05$ (月コメント投稿数)

コミッター昇格前と一般開発者で有意差がある
⇒コミッター選出の指標として有効

予測モデルの構築

- RQ2: 開発者の活動量からコミッター候補者をどの程度の精度で予測できるか？
- ロジスティック回帰分析によって予測
 - 説明変数
 - 7種類の活動内容メトリクス
 - 目的変数
 - コミッターに昇格するか否か

予測モデルの評価

	適合率	再現率	F1値
本実験モデルによる予測	0.64	0.76	0.70
ランダム予測	0.14	0.50	0.22
向上率	4.57	1.52	3.18

ランダムに予測するよりも高い精度で予測可能

今後の予定

- アプローチ①
 - コミッターをタイプ分類(検討中)
 - 長期間活躍するコミッター
 - パッチ投稿得意コミッター
 - パッチレビュー得意コミッター
 - コミュニケーション得意コミッター etc
 - タイプに応じてメトリクスを追加し、予測モデルを作成(ランダムフォレストetc)
- アプローチ②
 - 3ヶ月ほどの短期間のデータからコミッター候補者を見つける

伊原先生のコメント

- 予測よりClassificationの方がいいのでは？
- メーリングリストとかプルリクエストも見たら？
- HCEつかえば？
- とりあえず10プロジェクトみろ！
- To do
 - 10プロジェクトの分析(Doing, 来週まで)
 - 開発者分類の論文サーベイ
 - アプローチの決定