

Understanding the Archived Projects on GitHub

Xiaoya Xia

East China Normal University

Shengyu Zhao

Tongji University

Xinran Zhang

East China Normal University

Zehua Lou

East China Normal University

Wei Wang

East China Normal University

Fenglin Bi

East China Normal University

Proceedings of the 30th International Conference on Software Analysis, Evolution and Reengineering
(SANER '23), pp13-24

OSSとアーカイブ

OSS (Open Source Software)

- GitHubなどにソースが公開されているソフトウェア

特徴

- 有志の開発者によるボランティア開発
- 厳格な開発工程を持たない（特に要求や設計など上流工程）
- トレンドやニーズに応じて開発目的が変化

開発者が少なくなるなど
さまざまな要因からメンテナンスされなくなる



プロジェクトのアーカイブ化

研究目的とResearch Question

目的

- OSSプロジェクトのメンテナンスにおける課題などの理解
 - プロジェクト運営や進化過程, ライフサイクルに関する知識の獲得

Research Question

RQ1 : なぜアーカイブにされるのか？

RQ2 : どのようにアーカイブにされるのか？

RQ3 : アーカイブプロジェクトの活動パターンは？

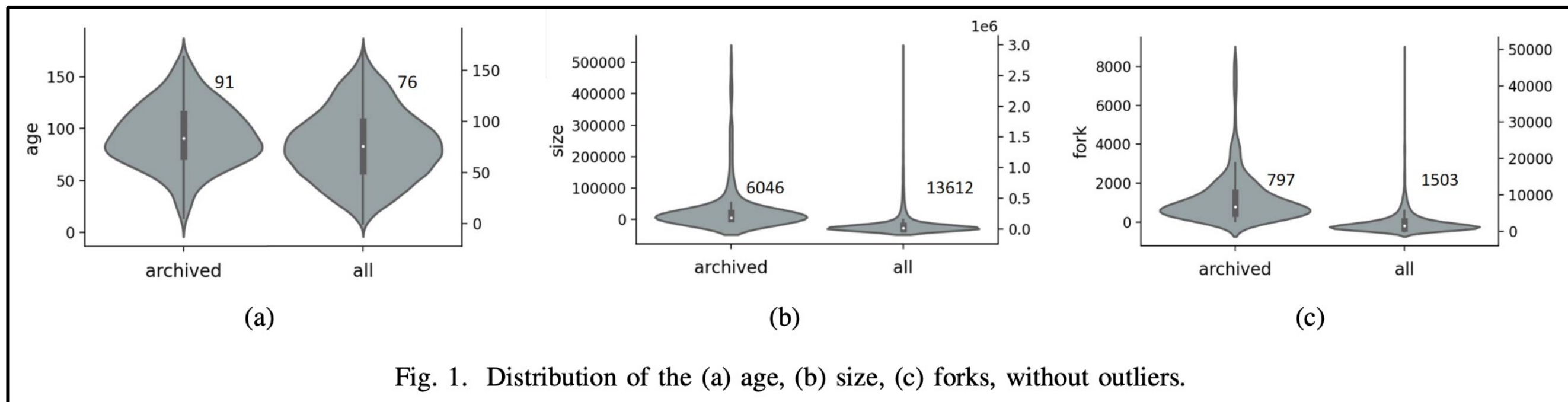
RQ4 : プロジェクトの持続性に関連しそうな要因は？

データセット

GitHubのアーカイブされた361リポジトリを対象

- スター数上位10,000リポジトリを収集
- アーカイブフラグが立っているプロジェクトを抽出
 - Awesome collectionや非英語圏プロジェクトは除外（45リポジトリ）

※ アーカイブプロジェクトとそれ以外の追加情報



方法

| Grounded Theory^[29]を用いた定性的分析

- README.mdの手動分析
 - アーカイブにした理由が書かれていた151プロジェクトを対象
- 開発者へのメールインタビュー
 - ユーザ191人のうち返信があった34件を対象（+14プロジェクト）

質問

Q1：なぜアーカイブにしたのか？

Q2：どのくらい円滑にアーカイブにできたか？

165プロジェクトを対象にアーカイブになった理由を分類

アーカイブにされる理由

Move and Transfer

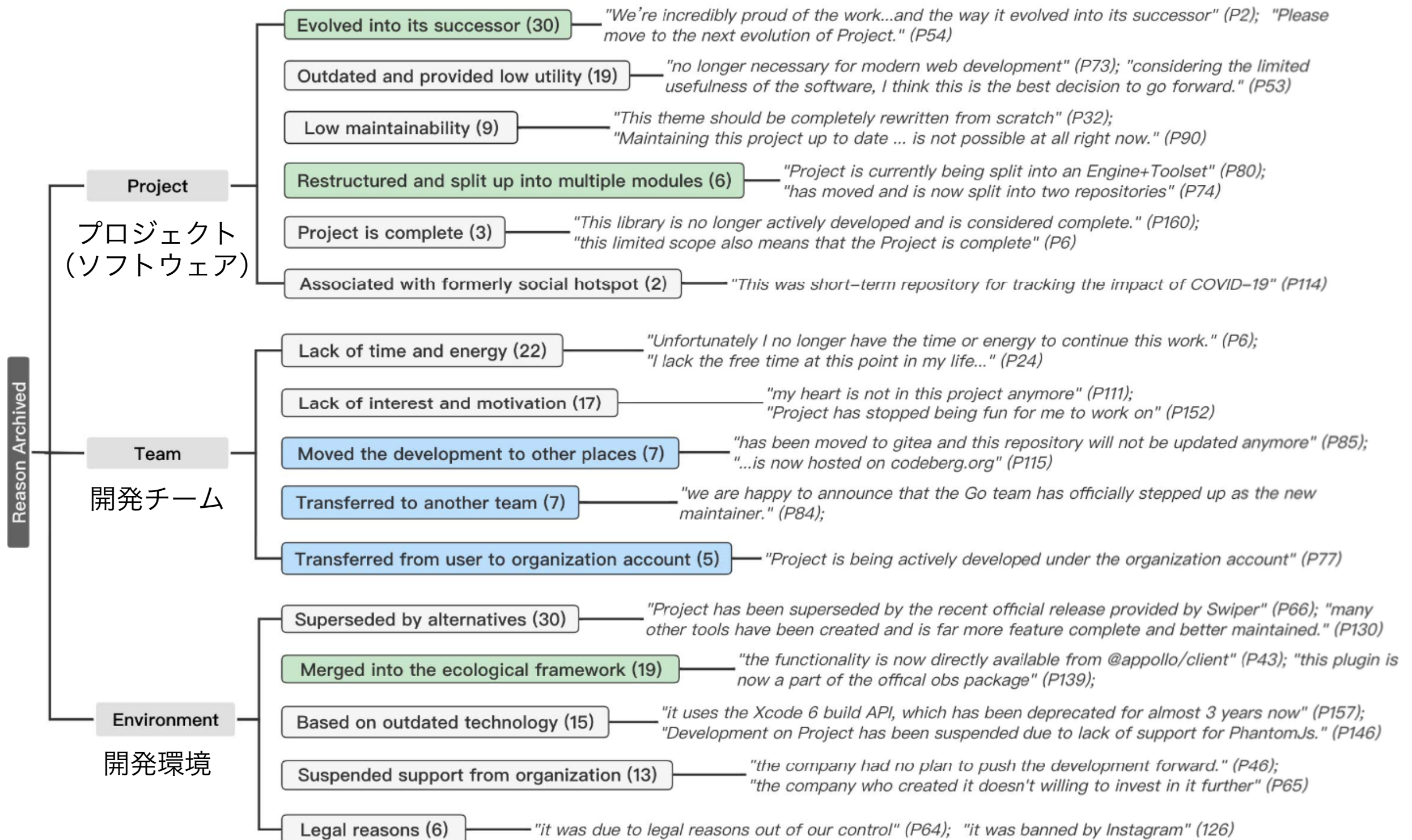
移動

Evolve and Upgrade

アップグレード

Unmaintained and Dead

保守していない



アーカイブにされる理由

Move and Transfer

- 異なるプラットフォームにリポジトリを移す
- ユーザ所有から組織所有に変更する
 - 例) CodebergからGitHubに移転させる

Evolve and Upgrade

- 後継のプロジェクトに切り替える
- より人気のあるプロジェクトにマージ
 - 例) KerasプロジェクトをTensorFlowプロジェクトにマージ

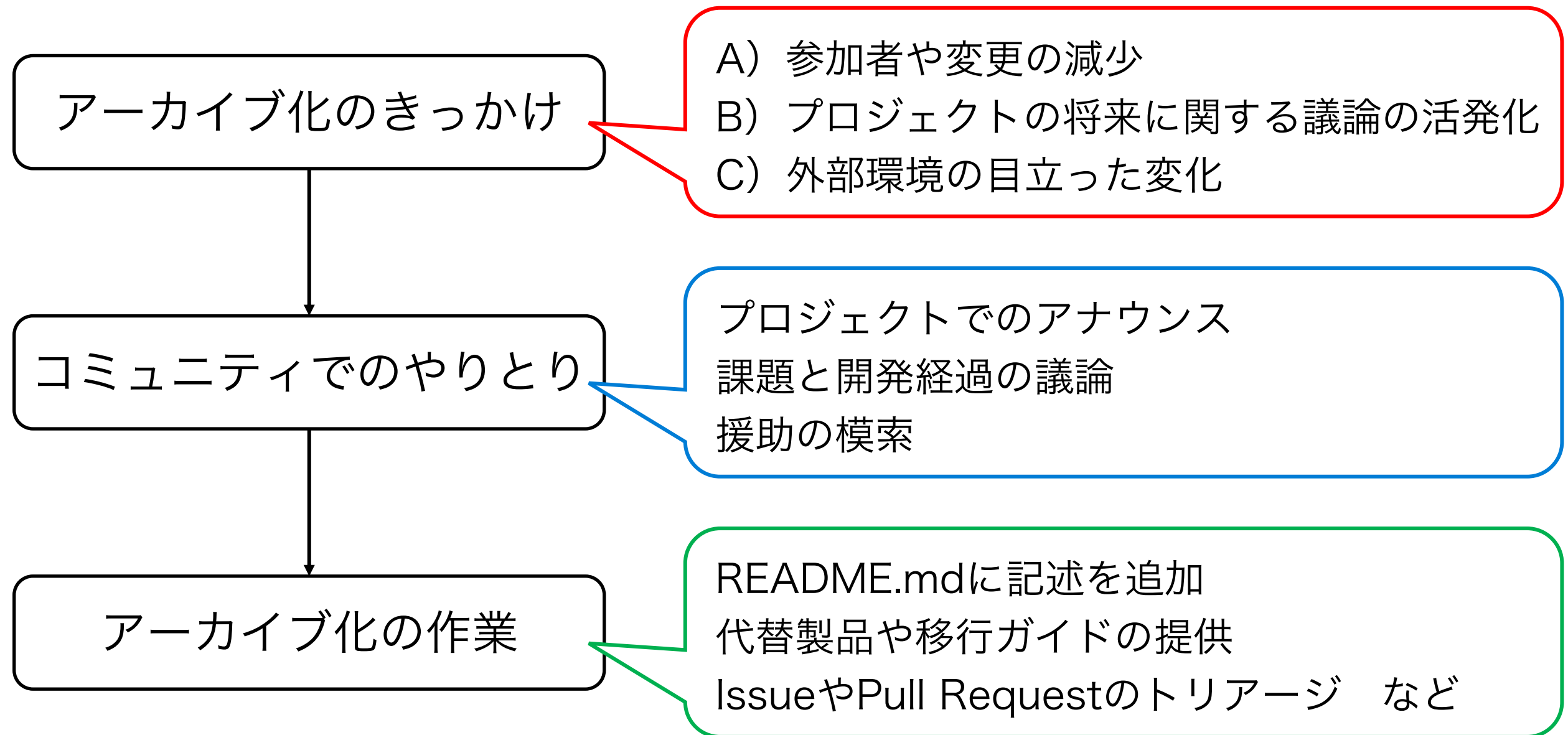
Unmaintained and Dead

- プロジェクトの人气がなくなったり開発意欲の低下

アーカイブにされるまで

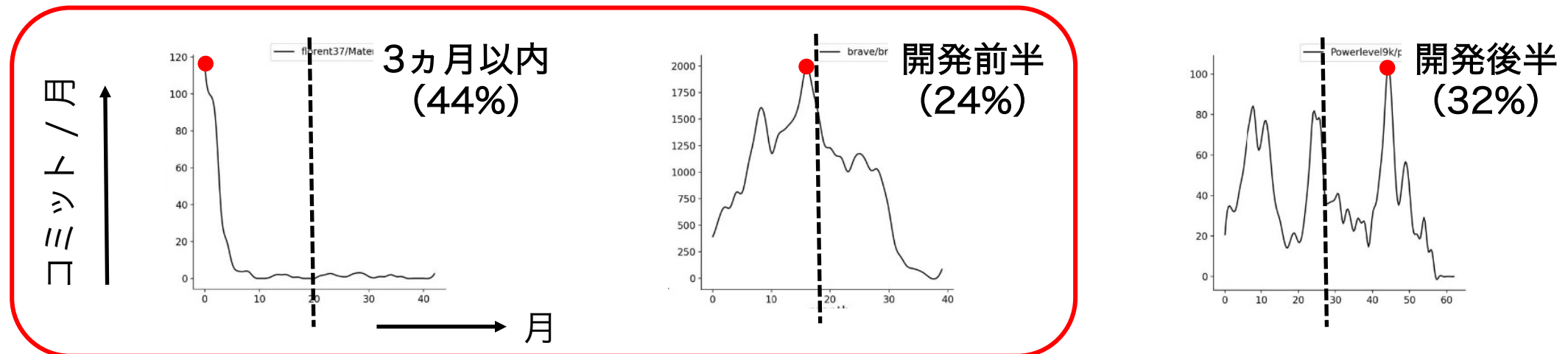
最新のIssueやPull Requestを分析してパターン化

- 最後の活動はアーカイブ化に関連していると想定



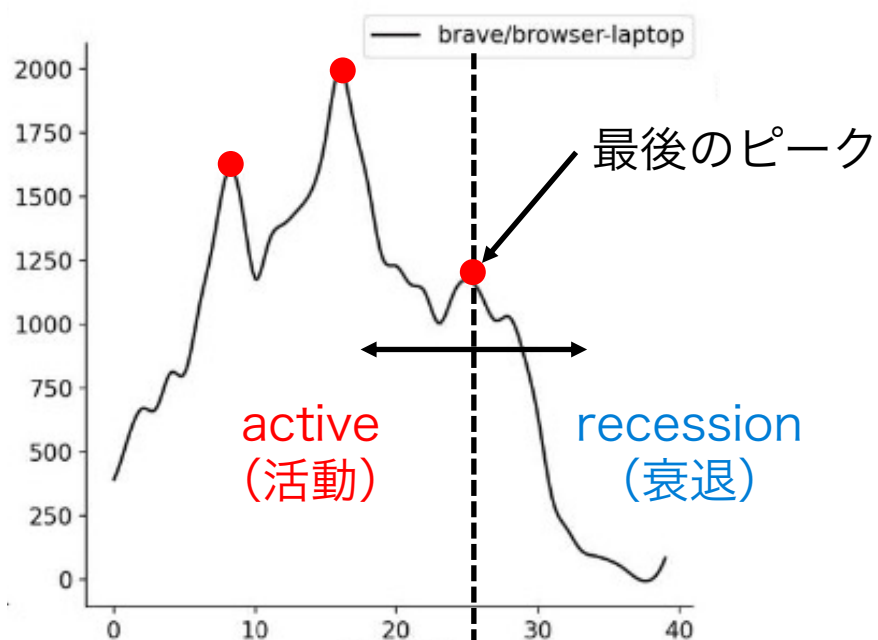
開発のピークと衰退

最も活動が活発な時期は？



開発の序盤もしくは前半でピークを迎える

活動期と衰退期の長さ



Active period

最後のピークが訪れるまでの期間
→ 中央値：21ヵ月

Recession period

最後のピークから最後のコミットが行われるまでの期間
→ 中央値：16ヵ月 (全体の約44%)

調査に利用する要因

CHAOSS^[9]で定義されているメトリクスを利用

- bus_factor_contributor
 - コミット数が多い開発者から順に足し合わせたときに、全コミット数の50%に到達するのに必要な開発者数
- issue_response_time
 - Issueが報告されてから、報告者以外の開発者にコメントされるまでの時間の中央値
- issue_resolution_duration
 - openからclosedになるまでの時間の中央値
- pr_merged_duration
 - Pull Requestがマージされるまでの時間の中央値
- pr_accepted_ratio
 - Pull Requestがマージされる割合

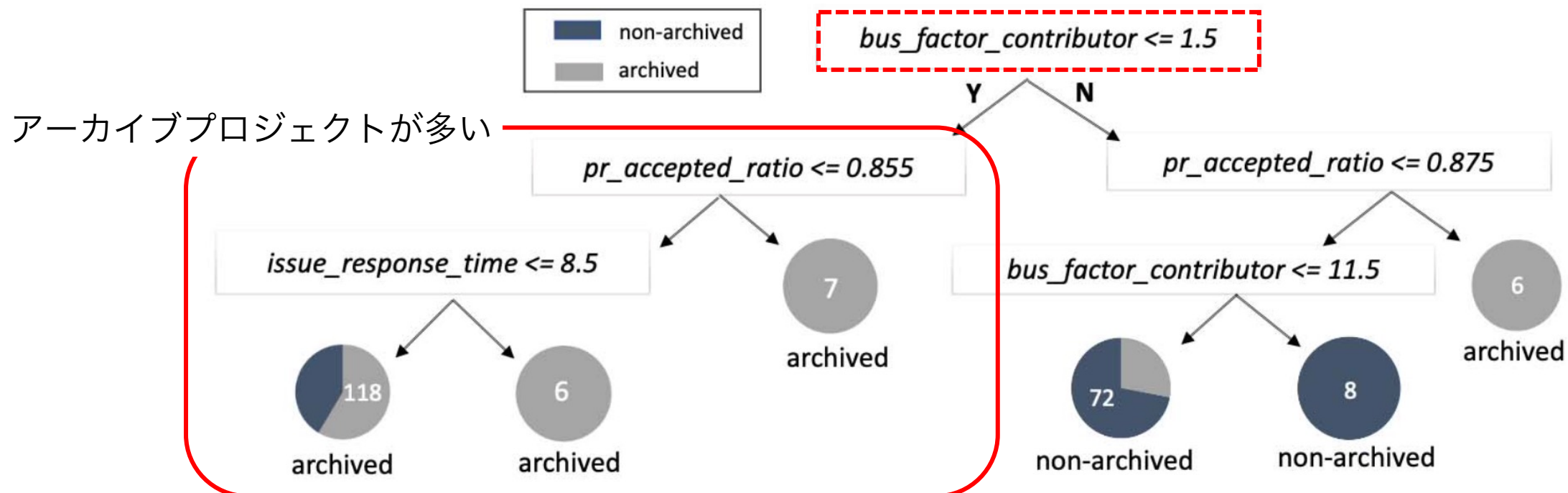
分析方法と結果

分析方法

- 前述の5つの要因と決定木アルゴリズムを用いてアーカイブプロジェクトとそれ以外に分類

結果

- コア開発者が少ないプロジェクトはアーカイブにされやすい
 - $\text{bus_factor_contributor} \leq 1.5$ なので、1人の開発者が50%以上貢献



まとめ

プロジェクトのアーカイブ化について調査を実施

組織管理者に向けて

プロジェクトを適切に管理するためにアーカイブプロセスの標準を設計し円滑に行えるようにすることが重要

- ・ 組織の利用だけでなくエンドユーザの利用にも関係

メンテナに向けて

コミュニケーションをしっかりと取りましょう

- ・ 貢献者に対するサポートなどを怠るとプロジェクトの非活性化に繋がるので注意組織の利用だけでなくエンドユーザの利用にも関係

アーカイブプロジェクトの分布

スター数とアーカイブプロジェクトの分布

stars	10k	5k	2k	1k	500
Archived	66	216	813	1,771	3,611
All	2,347	5,919	16,991	34,306	66,207
Ratio	2.81%	3.65%	4.78%	5.16%	5.45%

スター数：多い



スター数：少ない

プロジェクト数：少ない



プロジェクト数：多い

- スター数が少なくなるほど、アーカイブプロジェクトの割合が増加する
 - 人気があるプロジェクトほど貢献者が多くメンテナンスされやすい