

「ソフトウェア工学の実証的アプローチ」

ソフトウェアリポジトリマイニング

コンピュータソフトウェア, Vol.30, No.2(2013), pp.52-65.

門田 暁人

伊原 彰紀

松本 健一

発表：松尾 優太

Contents

- ❖ MSRとは
- ❖ 研究の魅力
- ❖ 主な研究テーマ
- ❖ マイニングの具体例
- ❖ 今後の発展
- ❖ まとめ



MSRとは

❖ 概要

Mining Software Repositoriesの略

- 何をマイニングするか
「新しい研究目的を見つけ出すこと」
- どのようにマイニングするか
 - マイニングの自由度が高い
 - 他分野の技術の取り入れ

MSRの研究テーマ

- ❖ ソフトウェア品質
- ❖ 開発プロセスの分析
- ❖ 協調作業の分析
- ❖ OSSを利用する開発者の支援
- ❖ MSR研究者の支援



MSRの研究テーマ①

❖ ソフトウェア品質

- バグモジュール予測
- Bug localization
バグ報告をもとにバグ位置を推定
- エラーパターンの検出
- バグ修正時間の予測
Firefoxプロジェクトの84.4%のバグは修正されない
- バグ修正誤りの分析

MSRの研究テーマ②

❖ 開発プロセスの分析

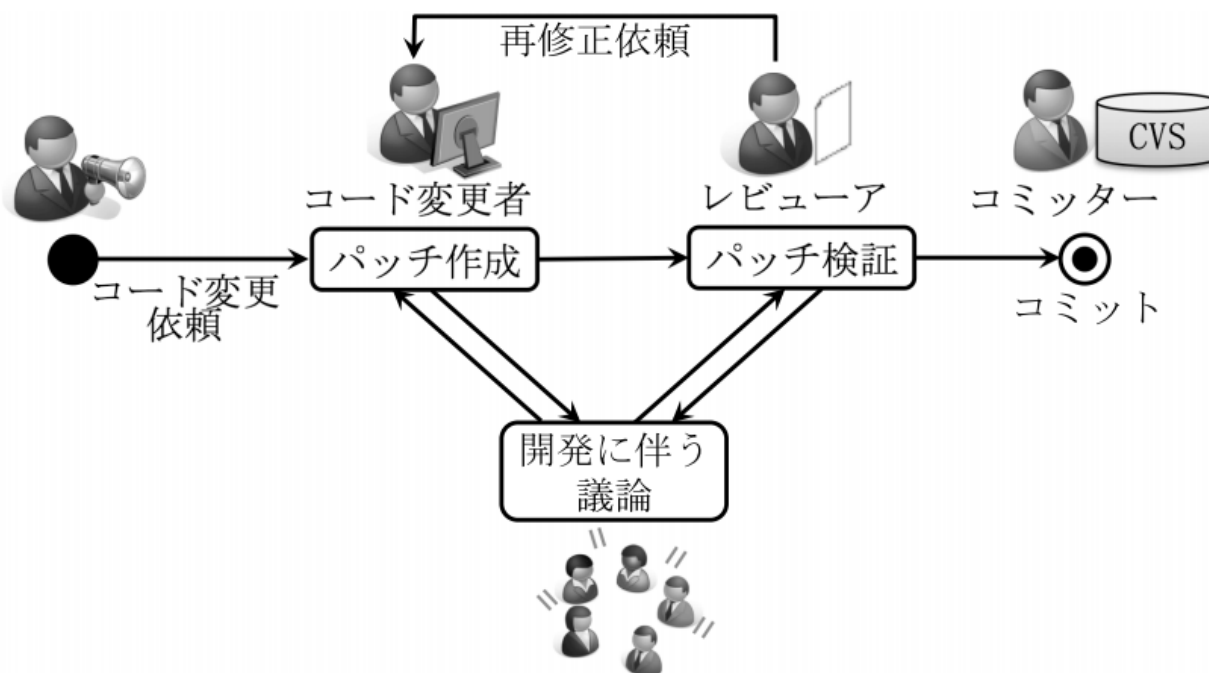


図 1 OSS 開発におけるソースコード変更プロセス

MSRの研究テーマ③

❖ 協調作業の分析

- エキスパート推薦
⇒特定のバグに適切な開発者は？
- コミッター推薦
- コミュニケーション分析
⇒SNA (Social Network Analysis)
- Openness(開放性)の分析

MSRの研究テーマ④

❖ OSSを利用する開発者の支援

- ライブラリの使用状況の分析
⇒機能性、信頼性の指標に.
- ソフトウェア部品検索
- ソフトウェアの分類
⇒類似ソフトウェアの発見
- ライセンスの分析
⇒ライセンスを自動で同定
- コーディングパターン検出

MSRの研究テーマ⑤

❖ MSR研究者の支援

- リポジトリ間のリンク付け
⇒ バグ管理システム＋構成管理システム
- 名寄せ(後述)
- MSRのサービス化
⇒ MSRをWebサービス化
⇒ 誰でも簡単にMSRが行える

マイニングの具体例①

❖ データ取得方法

- 構成管理システム
 - ⇒Webを介して取得
 - ⇒正規表現で必要な情報を抽出
- メーリングリストアーカイブ
 - ⇒添付ファイル

マイニングの具体例②

❖ データ成形方法

➤ 名寄せアルゴリズム

類似するアカウント, emailアドレスを同一人物かどうか判断するためのアルゴリズム.

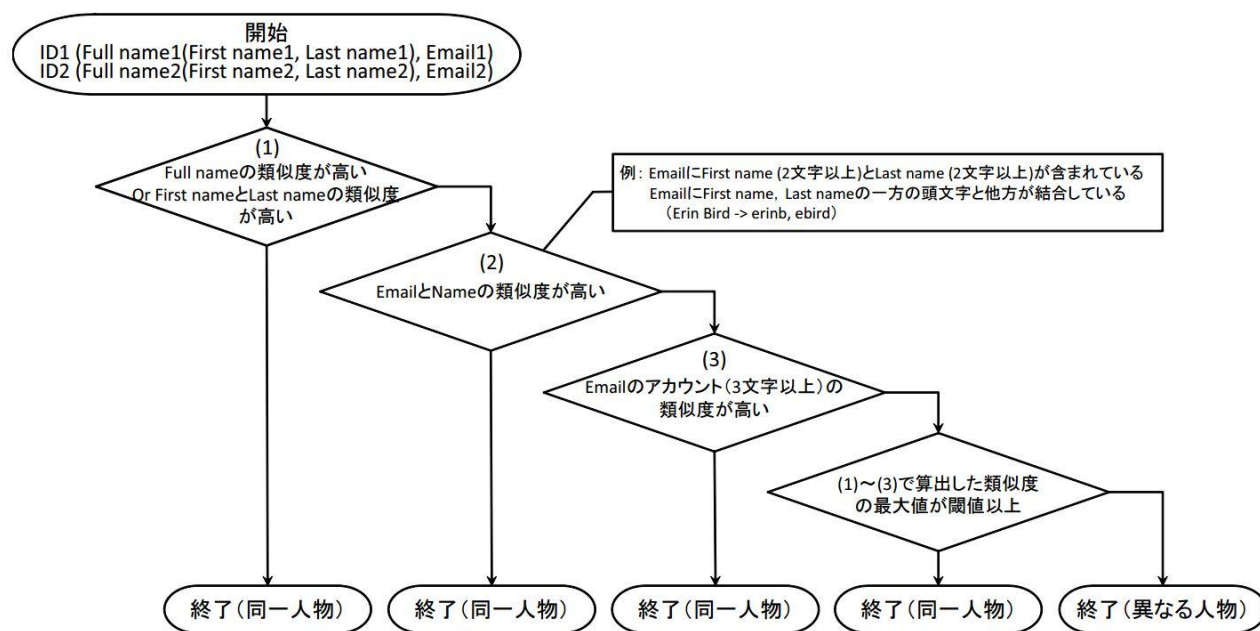


図3 “名寄せ” アルゴリズム

今後の発展

MSR vision 2020(2012.8)の議題より

❖ Large/Small-Scale mining

- 小規模プロジェクトも対象に！
- 課題: 十分なデータが存在するか

❖ IR

- ビッグデータの技術をMSRにも！
- 課題: ノイズがどれだけ影響するか

❖ Github

- 分散リポジトリが開発をどう変えるか

まとめ

❖ MSRの魅力

- 利用可能なデータが多い
⇒これまでの研究をより詳細に.
- 対象の拡大
⇒企業の開発支援から, 開発者, 利用者, 研究者へ.

❖ MSRに注目が集まっている