

平成 30 年 6 月 5 日現在

機関番号：14701

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K00101

研究課題名(和文) 派生開発における異常検知のためのプロアクティブマイニング手法の構築

研究課題名(英文) Proactive Data Mining for Detecting Anomalies in Derivational Development

研究代表者

大平 雅雄(OHIRA, MASAO)

和歌山大学・システム工学部・准教授

研究者番号：70379600

交付決定額(研究期間全体)：(直接経費) 3,500,000円

研究成果の概要(和文)：本研究課題の目的は、派生開発のような大規模な変更を伴う長期的な保守開発を対象とし、保守担当者がHigh Impact Defect(本研究における異常)の発生をいち早く察知できるようにするためのプロアクティブマイニング手法を構築することである。High Impact Defectを誘発するソフトウェアの変更をリアルタイムに監視する環境を構築するとともに、ソフトウェアの変更が一定期間後にHigh Impact Defectsとして顕在化する可能性を予測するプロアクティブマイニング手法を構築した。

研究成果の概要(英文)：The goal of this study is to construct a proactive mining method so that software maintainers can detect high impact defects (so called "anomalies" in this study) as soon as possible in software maintenance activities where large changes to existing software are required. We built a environment to monitor software changes which may induce high impact defects and construct a a proactive mining method to predict the possibility of being diagnosed with high impact defects in the future.

研究分野：ソフトウェア工学

キーワード：派生開発 プロアクティブマイニング High Impact Defects 異常検知 オープンソースソフトウェア

1. 研究開始当初の背景

ソフトウェアシステムの大規模化・短納期化に伴う開発リスクの高まりを受け、近年のソフトウェア開発では、新規開発よりも、品質が保証された既存システムの一部を流用し、システムを継続的に改良・拡張する「派生開発」[1]が主流になりつつある。

従来の「保守開発」では主に、製品出荷後に発見された欠陥の除去や、システム環境の変化に順応するためのソフトウェア変更が必要であった(図1上部)。一方、「派生開発」では、顧客ビジネスの急激な変化に対応するために、既存システムやオープンソースソフトウェア(OSS)を積極的に流用・移植しつつ短期間のうちに大幅な機能追加等を実現する必要がある(図1下部)。従来の保守開発においても、保守担当者の変更やドキュメント整備の不足等により、保守開発の生産性は必ずしも高くはなかったが、派生開発では、新規・流用・移植部分の担当者間で開発コンテキストの共有が困難になるため、予期しない箇所やタイミングで重大な欠陥が発見される場合がある[2]。

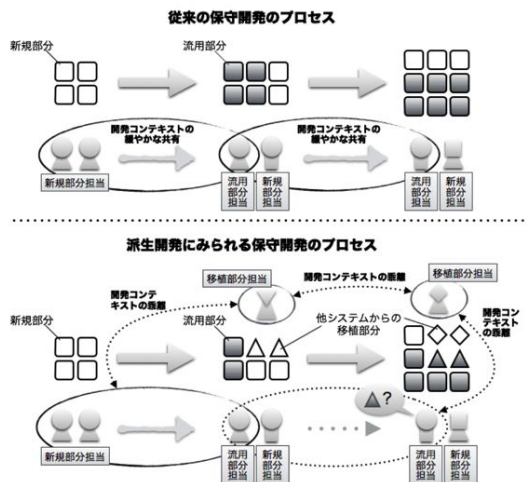


図1 保守開発プロセスの違い

ソフトウェア保守に関するトップカンファレンス ICSM は開催 30 周年を迎え、ICSME (International Conference on Software Maintenance and Evolution) と名称を一部変更し、派生開発による継続的かつ大幅な変更をソフトウェアの正統な進化と位置付けている。特に昨今の派生開発では、部品単体では高品質であるが故に大規模な流用や移植の過程で見逃される重大な欠陥 (High Impact Defects) が顧客のビジネスに大きな影響を与えることが多いため、High Impact Defects の早期発見や分析・修正を支援する技術の確立が喫緊の課題となっている。

これまでのソフトウェア進化に関する研究[3]やリポジトリマイニングに関する研究[4]は、従来の保守開発を対象としたものがほとんどであり、派生開発に着目した支援技術が研究されてこなかったという背景がある。

2. 研究の目的

そこで本研究課題では、派生開発のような大規模な変更を伴う長期的な保守開発を対象とし、保守担当者が異常の発生をいち早く察知できるようにするためのプロアクティブマイニング手法を構築することを目的とした。本手法は、プロダクトデータ(ソースコード)のみならず、保守担当者間のコミュニケーションや保守開発のプロセスデータを用いて、High Impact Defects の早期発見と原因特定を支援する点に特徴がある。

3. 研究の方法

High Impact Defects の混入を早期検知する手法の構築へ向け、本研究課題では、次の段階に分けて研究を実施した。

- (1) High Impact Defects の実証的調査を実施し分布や特徴を分析する。
- (2) プロダクトおよびプロセスデータから High Impact Defects の混入原因を特定する。
- (3) 説明変数の変化が目的変数の一定期間後の変化に影響を与えるかどうかを分析できる遅延相関分析を応用することで、High Impact Defects を早期に検知するためのプロアクティブマイニング手法を構築する。
- (4) OSS プロジェクトでの検証実験およびソフトウェア開発企業での実証実験を通じて構築した手法の有用性を明らかにする。

4. 研究成果

(1) High Impact Defects の実証的調査

先行研究の調査結果から、High Impact Defects は主に、顧客(ユーザ)のシステム利用に直接的に大きな影響を与える欠陥と、保守担当者らに開発スケジュールの大幅な変更をもたらす欠陥とに大別できる。また、それぞれに3種類(計6種類)の欠陥が存在する。先行研究では1プロジェクトあたり100件分(合計400件)の不具合票を調査したに過ぎなかったが、6種類の High Impact Defects の特徴を一般化するために本研究では、調査範囲を1プロジェクトあたり1,000件(計4,000件)に拡張し調査を実施した。

(2) High Impact Defects の混入原因の特定

High Impact Defects の混入原因を、プロダクトおよびプロセスデータから多面的に分析した。

特に、モジュール間の依存関係を動的・静的に分析し、High Impact Defects に繋がる関係が存在するかどうか、流用・移植されたモジュールが他のモジュールの正常な動作を妨げていないかどうかに着目した分析を行った。

また、モジュールを作成した開発者とモジュールを利用(流用・移植)する保守担当者とのコミュニケーションの質を分析し、流用・

移植が円滑に行えるよう十分な意思疎通が図れているかどうかを分析した。

さらに、報告された1件の不具合の修正には複数のファイルにまたがる欠陥を除去する必要がある場合もあれば、1件の不具合を修正するために他の複数の不具合を先に修正する必要がある場合もある。複雑に絡み合ったこのような不具合報告と欠陥の関係や不具合同土の関係が、High Impact Defectsを誘発していないかどうかを確認した。

#### (3)プロアクティブマイニング手法の構築

まず、High Impact Defectsの種類と発生原因をマッピングし因果関係を抽出し学習データとして利用するためのデータセットを作成した。次に、High Impact Defectを誘発するソフトウェアの変更をリアルタイムに監視する環境を構築した。さらに、ソフトウェアの変更が一定期間後に High Impact Defectsとして顕在化する可能性を予測する手法を構築した。ソフトウェアに対する変更を様々なメトリクスで時系列に計測し、研究成果(2)で得られた High Impact Defectsとの相関を求めることで、ソフトウェア変更による異常を素早く検知できるかどうかを確認した。

#### (4)手法の評価

研究成果(3)で構築したプロアクティブマイニング手法の評価を行った。具体的には、複数のOSSプロジェクト(Eclipse Platformプロジェクト、Mozillaプロジェクト、Apacheプロジェクト)から取得したデータ(バージョン管理履歴および不具合管理履歴)を用いて、手法が異常を早期に検知できるかどうかの検証実験を行った。また、協力企業との共同研究を通じて小規模な実証実験を行い、構築した予測手法が実務においても有用かどうかを評価した。

#### 引用文献

- [1] 清水吉男,「派生開発」を成功させるプロセス改善の技術と極意,技術評論社 2007
- [2] E. Shihab, A. Mockus, Y. Kamei, B. Adams and A.E. Hassan, High-Impact Defects: A Study of Breakage and Surprise Defects, Proc. of the 19th ACM SIGSOFT Symposium and the 13th European Conference on Foundations of Software Engineering (ESEC/FSE '11), pp.300-310, 2011
- [3] 大森隆行,丸山勝久,林晋平,沢田篤史,ソフトウェア進化研究の分類と動向,コンピュータソフトウェア, Vol.29, No.3, pp.3-28, 2012
- [4] 門田暁人,伊原彰紀,松本健一,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, pp.52-65, 2013

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

##### [雑誌論文](計4件)

宮崎智己,伊原彰紀,太平雅雄,東裕之輔,山谷陽亮, OSS コミュニティにおける開発者の活動継続性を理解するための Politeness 分析, 情報処理学会論文誌, 査読有, Vol.59, No.1, pp.2-11, 2018  
鶴田雅明,太平雅雄,門田暁人,松本健一,ソフトウェアテストにおける静的解析ツールの段階的適用による不具合修正工数の更なる低減 民生用音響・映像機器向け組み込みソフトウェア開発への QAC の段階的適用とその実証評価, 情報社会学会誌, 査読有, Vol.11, No.1, pp.5-16, 2016

伊原彰紀,太平雅雄,オープンソースソフトウェア工学,コンピュータソフトウェア, 査読有, Vol. 33, No.1, pp.28-40, 2016

吉行勇人,太平雅雄,戸田航史, OSS 開発における管理者と修正者の社会的関係を考慮した不具合修正時間予測, コンピュータソフトウェア, 査読有, Vol.32, No.2, pp.128-134, 2015

##### [学会発表](計16件)

福井克法,太平雅雄, E-mail データマイニングに基づく適任開発者の推薦手法の検討, 第 24 回ソフトウェア工学の基礎ワークショップ (FOSE2017), 査読有, pp.239-240, 2017

蘆田誠人,太平雅雄, 時系列モデルを用いた遅延相関分析の評価, 第 24 回ソフトウェア工学の基礎ワークショップ (FOSE2017), 査読有, pp.241-242, 2017  
柏祐太郎,山谷陽亮,太平雅雄, OSS 開発における不具合修正プロセスの改善に向けたモジュールオーナー候補者推薦, 第 24 回ソフトウェア工学の基礎ワークショップ (FOSE2017), 査読有, pp.165-170, 2017

久木田雄亮,太平雅雄, コードクローン変更過程における開発者のインタラクションとソフトウェア品質の関係, ソフトウェア・シンポジウム 2017 in 宮崎, 査読有, pp.120-129, 2017

宮崎智己,太平雅雄, OSS 開発者の離脱要因理解のための Politeness の質的調査, ソフトウェア・シンポジウム 2017 in 宮崎, 査読有, pp.75-84, 2017

福井克法,築山文香,太平雅雄, OSS 開発における長期貢献者の特徴理解のための注目度と活動期間の関係分析, 情報処理学会マルチメディア,分散,協調とモバイル(DICOMO2017)シンポジウム論文集, pp.1274-1284, 2017

Masao Ohira, Hayato Yoshiyuki, and

Yosuke Yamatani, A Case Study of the Misclassification of Performance Reports in an Issue Tracking System," In Proceedings of 15th IEEE/ACIS International Conference on Computer and Information Science (ICIS2016), 査読有, pp.855-860, 2016

松野祐介, 山谷陽亮, 大平雅雄, 重大不具合に対する OSS 開発者の認識調査, ソフトウェア・シンポジウム 2016 in 米子, 査読有, pp.11-20, 2016

大平雅雄, 久木田雄亮, コードクローンの作成・利用過程における人的影響を調査するための追跡ツールの試作, 情報処理学会 第 193 回ソフトウェア工学研究発表会, Vol.2016-SE-193, No.25, 5, 2016

宮崎智己, 山谷陽亮, 東裕之輔, 大平雅雄, OSS 開発コミュニティの進化の理解を目的としたコミュニケーション分析: Politeness 分析適用の試み, 情報処理学会マルチメディア, 分散, 協調とモバイル (DICOMO2016) シンポジウム論文集, pp.703-713, 2016

大平雅雄, OSS のメンテナンスに関わる最新の研究動向, ソフトウェア・メンテナンス・シンポジウム 2016, 招待講演, 2016

Masao Ohira, Yutaro Kashiwa, Yosuke Yamatani, Hayato Yoshiyuki, Yoshiya Maeda, Nachai Limsettho, Keisuke Fujino, Hideaki Hata, Akinori Ihara, and Kenichi Matsumoto, A Dataset of High Impact Bugs: Manually-Classified Issue Reports, In Proceedings of 12th Working Conference on Mining Software Repositories (MSR '15), 査読有, pp.518-521, 2015.

Yunosuke Higashi, Yuki Manabe, and Masao Ohira, Clustering OSS License Statements Toward Automatic Generation of License Rules, In Proceedings of the 7th IEEE International Workshop on Empirical Software Engineering in Practice (IWESEP2016), 査読有, pp.30-35, 2016.

Masao Ohira, Hayato Yoshiyuki, and Yosuke Yamatani, A Case Study of the Misclassification of Performance Reports in an Issue Tracking System, In Proceedings of 15th IEEE/ACIS International Conference on Computer and Information Science (ICIS2016), 査読有, pp.855-860, 2016.

大平雅雄, 開発データマイニングのススメ: 「見える化」の次のステップ, 独立行政法人情報処理推進機構 (IPA) 技術本部ソフトウェア高信頼化センター主催セミナー「定量的なプロジェクト管理・プログラム管理のススメ ~見える

化による生産性・信頼性の向上~」, 招待講演, 2015

大平雅雄, プロジェクト管理ツール EPM-X の機能拡張によるプロアクティブ型プロジェクトモニタリング環境の構築 ~次世代の定量的プロジェクト管理ツールとリポジトリマイニング研究基盤~, Embedded Technology West 2015 / 組込み総合技術展関西 IPA/SEC プレゼン, 招待講演, 2015

〔図書〕(計0件)

〔産業財産権〕

出願状況 (計0件)

名称:  
発明者:  
権利者:  
種類:  
番号:  
出願年月日:  
国内外の別:

取得状況 (計0件)

名称:  
発明者:  
権利者:  
種類:  
番号:  
取得年月日:  
国内外の別:

〔その他〕  
ホームページ等

## 6. 研究組織

(1) 研究代表者  
大平 雅雄 (OHIRA, Masao)  
和歌山大学・システム工学部・准教授  
研究者番号: 70379600

(2) 研究分担者 ( )

研究者番号:

(3) 連携研究者 ( )

研究者番号:

(4) 研究協力者 ( )