

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 24 年 6 月 11 日現在

機関番号：14401

研究種目：基盤研究（C）

研究期間：2009 ～ 2011

課題番号：21500035

研究課題名（和文）大規模に収集された開発データからのプロジェクト管理のための知見の導出

研究課題名（英文） Knowledge extraction for software project management from a large set of project data

研究代表者

菊野 亨 (KIKUNO TOHRU)

大阪大学・大学院情報科学研究科・教授

研究者番号：50093745

研究成果の概要（和文）：

ソフトウェア開発のプロジェクトにおける混乱状態回避を目的として、プロジェクトに関するメトリクスから混乱するかどうかの診断手法を開発した。実際のソフトウェア開発プロジェクトにおいて収集されたメトリクスデータを利用して、その有効性を示した。

研究成果の概要（英文）：

This research developed a diagnosis approach for software development projects that is based on software project metrics. Using real-life examples of project metrics, the usefulness of the approach was demonstrated.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009年度	1,600,000	480,000	2,080,000
2010年度	1,100,000	330,000	1,430,000
2011年度	700,000	210,000	910,000
年度			
年度			
総計	3,400,000	1,020,000	4,420,000

研究分野：総合領域

科研費の分科・細目：ソフトウェア

キーワード：ソフトウェア工学、リスク管理、データマイニング

1. 研究開始当初の背景

社会の情報機器への依存が大きくなるにつれて、ソフトウェア開発現場において、プロジェクト管理の必要性が高まってきている。そのような背景から、ソフトウェア開発プロジェクトから収集したデータを利用して、ソフトウェアの品質や開発におけるコストなどを予測する研究が多く行われている。しかし、現実のソフトウェア開発現場に理論的には優れた既存の管理手法を適用することは、様々な点で困難であった。

このような困難さの原因としては、実際のソフトウェア開発現場においては、プロジェ

クトのデータ（メトリクス）を収集するコストが無視できないこと、そして、そのために十分な時間、手間をかけてデータを収集することができず、得られるデータが欠損を多く含むことが避けられないことなどが挙げられる。

2. 研究の目的

上記1の研究開始当初の背景で述べた、実際のソフトウェア開発プロジェクトにおける管理手法の適用の困難を克服し、開発現場で用いることの可能な先進的プロジェクト管理法を実現することが本研究の最終的な目

的である。期限内に達成すべき具体的な目的としては、このために必要な知見を得ることである。この知見は、実際のソフトウェア開発プロジェクトから収集された大規模のメトリクスデータを利用、分析することで、はじめて得ることが可能となる。

3. 研究の方法

まず、分析に用いる現実の開発プロジェクトから収集されたメトリックデータを確保した。具体的には、情報処理推進機構ソフトウェア・エンジニアリング・センター (IPA/SEC) からデータの提供を受けた。

次に、プロジェクトのデータ (メトリクス) の収集に要するコストの問題に取り組んだ。従来から、開発プロジェクトが失敗する (具体的には、フィールドバグが多く発生する) か成功するかの予測手法は存在したが、多くのチェック項目についてのデータを要するため、データ収集のためのコストが大きく、実プロジェクトにおける適用の大きな障害になっていた。

また、データ収集に時間がかかることから、開発後期 (例えば出荷直前) にならないと予測ができないという問題もあった。この場合、予測のタイミングが遅過ぎて、仮に予測できても、もはやプロジェクトを改善する手の打ち様がない。

そこで本研究では、開発プロジェクトを以下のような2状態に分けて考え、それぞれの状態に対して異なるデータ収集、分析方法を用いるアプローチを考案した。2状態は以下の通りである。

通簡易診断モード：少ないチェック項目についての監視を早い時期から定期的に繰り返して実行する

詳細診断モード：危ういと予測された場合であり、多くのチェック項目のデータ収集を行う。

通常状態でのプロジェクト管理では、どのようなメトリクスを用いるかをしぼり込むことで、データ収集のコストを下げる必要がある。同時に、選ばれたメトリクスから、プロジェクトの進捗予測が可能でなければならないため、どのメトリクスを選ぶかが非常に重要になる。そこで、ベイズ識別器による予測を前提として、定期的な監視の対象とするメトリクスの選別を行った。

詳細診断モードでのプロジェクト管理では、収集可能なデータをすべて利用することができる。しかし、予備実験により、従来の分析手法を適用すると、プロジェクトの成否

を50%程度の精度でしか予測できないことが分かった。この理由として、混乱プロジェクトではデータ収集がうまくできず未整理のままになっていること、および、そうした未整理のデータをそのまま利用していること、の二つを特定した。そして、この観察を元に、以下の二つの研究方法を実行した。

欠損データの削減：不完全なデータの混入する割合を極力抑えることを目的として、多くのプロジェクトで値が報告されていないメトリクスを削除する。さらに、データ収集が十分に出来ていないプロジェクトのデータを削除する。

メトリクス数の削減：相関ルールマイニングを利用して、利用するメトリクスを、予測に本当に必要なものだけに限定する。

最後に欠損を多く含むデータの取り扱い方法について検討した。用いたデータでは、実際に43.8%もの欠損が含まれており、このように欠損が多い場合でのデータの取捨選択が、精度の高い予測に重要な役割を果たすと考えられた。そこで、既存のデータを操作することで欠損率を人為的に変化させ、異なる欠損率の元での最適なメトリクスの選択と予測精度の向上法を検討した。具体的な手順は以下の(1)~(3)の通りである。

- (1) 目標とする欠損率を有するデータの作成：収集されたメトリクスデータを、欠損数の大きい順に並び替えた。引き続き、先頭から順番にメトリクスを削除して、データ全体の欠損率を目標値に近づけた。欠損率の妥当な値が事前には判明しないため、20%、15%、10%、5%、0%の5つを目標値を選んで以降の(2)、(3)を検討した。
- (2) メトリクスの選択：(1)で求めた5種類のメトリクスデータ毎に、相関ルールマイニングを適用して、プロジェクト状況に大きな変化を与えるメトリクスの選択・抽出を行った。
- (3) 予測精度の算出：(2)で選択したメトリクスに基づいて、(1)で求めたデータから切り出しを行った。引き続き、ベイズ識別器を適用して、予測精度を算出した。

4. 研究成果

通常状態でのプロジェクト管理におけるメトリクスの選別に関しては、プロジェクトデータへの適用実験を行って、候補メトリクスを求めることに成功した。

また、その検討の中で、使用するチェック

項目（メトリクス）の数と予測精度の間にトレードオフの関係が存在することを発見した。更に、メトリクス以外に、通常の見積もりに含まれるプロジェクトデータを利用すると予測精度の向上が見込まれることも明らかになった。

詳細診断モードでのプロジェクト管理に関しては、相関ルールマイニングの結果、利用可能な全 78 個のメトリクスを 6 個にまで削減しても、プロジェクトの成功、失敗に関する予測精度はむしろ向上することが分かった。具体的な予測精度の数値としては、6 個のメトリクスの利用でも 70%の精度が得られ、従来の約 50%という精度と比較すると、大幅な予測精度の向上に成功した。

最後の欠損を多く含むデータの取り扱い方法に関しては、考慮したそれぞれの欠損率に対して、選択されるメトリクスと予測精度を得ることができた。例えば、最も高い予測精度が得られたのは欠損率を 10%とした場合であった。このとき、7 個のメトリクスが選択され、82.8%の精度であった。この精度の値は、欠損の削除やメトリクスの選択を行わないで予測を行う場合に比べ、十分に高い予測精度を示すものであった。

以上のように、不完全なプロジェクトデータであっても、予測で用いるメトリクスを適切に選択することで、むしろ予測精度が向上することを、様々な条件下で示すことができた。メトリクスの絞り込みの結果、メトリクス収集にかかるコストが低減され、実際のソフトウェア開発でのデータ収集が容易になるといえる。同時に、データ収集の時間も低減されるため、早期の予測が可能となり、その結果をプロジェクト管理へフィードバックすることも可能となる。

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 3 件)

出張純也, 菊野亨, 菊地奈穂美, 平山雅之, 欠損率の高いプロジェクトデータを利用したプロジェクトの成否予測, 情報処理学会論文誌, 査読有, 53 巻, 2012, 662-671

尾形憲一, 出張純也, 菊野亨, 水野修, 菊地奈穂美, 平山雅之, ベイズ識別器による不具合予測のための相関ルールマイニングを用いたメトリクス絞り込み, 情報処理学会研究報告, 査読無, Vol. 2010-SE167, No. 4. 2010, 1-7

出張純也, 尾形憲一, 菊野亨, 水野修, 菊地奈穂美, 平山雅之, ソフトウェア開発データに対する相関ルールマイニングを利用した不具合増加要因の調査, 情報処理学会研究報告, 査読無, Vol. 2010-SE167, No. 3, 2010, 1-8

[学会発表] (計 6 件)

Junya Debari, Tohru Kikuno, Nahomi Kikuchi, and Masayuki Hirayama, Prediction of Successful Projects Using Public Project Data Set with Incomplete Data, 22nd International Symposium on Software Reliability Engineering (ISSRE2011), 2011 年 12 月 2 日, 広島

Junya Debari, Tohru Kikuno, Nahomi Kikuchi, and Masayuki Hirayama, An Improvement of Accuracy in Product Quality Prediction Using Imbalanced Project Data in Japan, 3rd International Workshop on Empirical Software Engineering in Practice (IWESEP2011), 2011 年 11 月 1 日, 奈良

Masayuki Hirayama, Junya Debari, Nahomi Kikuchi, and Tohru Kikuno, An Extraction Method of Project's Failure Factor in IT System Development, 2011 International Conference on Software and Intelligent Information, 2011 年 10 月 23, Puerto Rico (USA)

出張純也, 菊野亨, 菊地奈穂美, 平山雅之, 欠損率の高いプロジェクトデータを利用したプロジェクトの成否予測, ソフトウェアエンジニアリングシンポジウム 2011 (SES2011), 2011 年 9 月 13 日, 東京

尾形憲一, 出張純也, 菊野亨, 菊地奈穂美, 平山雅之, ソフトウェアプロジェクト予測に用いるメトリクスの削減, 情報処理学会ソフトウェア工学研究会, 2010 年 11 月 11 日, 大阪大学

出張純也, 尾形憲一, 菊野亨, 水野修, 菊地奈穂美, 平山雅之, 相関ルールマイニングを利用したソフトウェアプロジェクト混乱要因の関連性に関する調査, 情報処理学会創立 50 周年記念全国大会 (第 72 回全国大会), 2010 年 3 月 11 日, 東京大学 (東京都)

6. 研究組織

(1) 研究代表者

菊野 亨 (KIKUNO TOHRU)

大阪大学・大学院情報科学研究科・教授

研究者番号：50093745

(3) 連携研究者

水野 修 (MIZUNO OSAMU)

京都工芸繊維大学・大学院工芸科学研究科・准教授

研究者番号：60314407

(H21 まで分担者、H22 から連携研究者として参画)