

平成 21 年 5 月 15 日現在

研究種目：基盤研究 (C)  
 研究期間：2006～2008  
 課題番号：18500021  
 研究課題名（和文） ソフトウェアプロジェクトに伴うリスク予知へのデータマイニング手法を利用した挑戦  
 研究課題名（英文） A Challenge to Predict Risks in Software Projects Using Data Mining Techniques  
 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：

本研究ではソフトウェア開発プロジェクト内で発生する危険因子間の関係を明らかにし、それらの関係の利用によって潜在的なリスクを予知する新しい手法を提案した。提案した予測手法は以下の3点から実践的なものとなっている。（1）実際の開発現場で収集されるデータに基づいて危険因子間の関係を明らかにしている。（2）実際の適用場面での危険度の高い因子と（1）で求めた関係を用いて、リスク予知を可能としている。（3）予知された危険因子とそれへの対処法の自動収集までを目指している。提案手法を現実の開発プロジェクトから得られたデータに対して適用した結果、危険度の高い因子の抽出とその改善案が抽出可能であることを確認した。

## 交付額

(金額単位：円)

	直接経費	間接経費	合計
2006 年度	2,000,000	0	2,000,000
2007 年度	1,000,000	300,000	1,300,000
2008 年度	600,000	180,000	780,000
年度			
年度			
総計	3,600,000	480,000	4,080,000

研究分野：総合領域

科研費の分科・細目：情報学・ソフトウェア

キーワード：ソフトウェア工学，データマイニング，ソフトウェア開発プロジェクト

## 1. 研究開始当初の背景

ソフトウェアが必要とされる応用分野は格段に広くなり、ソフトウェアによって解決が望まれる問題もますます複雑化・大規模化してきている。また、応用分野の広がりとともにその重要性も一層増してきており、品質への要求レベルもますます高まっ

てきている。こうしたソフトウェアの複雑化と高信頼性への要求に加え、ソフトウェアを取り巻くシステム環境が激しく変化していることがソフトウェア開発をより困難なものとしている。

このような環境下において大規模かつ高品質なソフトウェアを開発するためには、ソフトウェア開発を混乱させる危険因子の

状況（危険性の度合いや他への影響）を十分に考慮し、タイムリーに（あるいは事前予防的に）危険因子に対処するリスク管理活動を展開することが不可欠である。リスク管理活動の不備はソフトウェアプロジェクトを確実に混乱状態に陥れる。ここでプロジェクトの混乱とは、プロジェクトの進行中に誰が何の作業を行っているのか、製品の完成度がどの程度の状況にあるのかといったプロジェクトの進捗が開発メンバの誰にも分からなくなっていることを指す。こうした状況が存在することは、実際の開発現場で作業に従事する人々の間では広く知られている

## 2. 研究の目的

従来のリスク管理手法では定期的に個々の危険因子についてその評価を行い、その結果に基づいて対応策を実施するか否かの判断を行っている。しかし、実際には個々の危険因子は互いに影響を及ぼしあう関係にあると考えられるため、ある時点において危険度が高いと判断された因子のみに着目するだけではプロジェクトの混乱回避には不十分である。危険因子間の関係に基づき、潜在的なリスクについても着目して対処していくことの必要性が指摘されている。

本研究では危険因子間の様々な関係を明らかにし、それらの関係の利用によって潜在的なリスクを予知する新しい手法を提案する。ここでは予測手法を実践的なものとするために以下の点を重視する。まず、(1) 実際の開発現場で収集されるデータに基づいて危険因子間の関係を明らかにしていく。次に、(2) 実際の適用場面での危険度の高い因子と(1)で求めた関係を用いて、リスク予知（危険度が高くなる恐れのある因子を求めること）を可能とする。最終的には、(3) 予知された危険因子とそれへの対処法との関係についても明らかにしていく。

## 3. 研究の方法

本研究は次のようなフェーズから構成されるっており、以下に示す各フェーズに従って実施された。

フェーズ1：危険因子間に存在する関係の明確化

フェーズ2：危険因子間の関係に基づくリスク予知手法の構築

フェーズ3：実際の開発現場での利用による評価実験

本研究課題では平成18年度にフェーズ1を、平成19年度にフェーズ2を、そして、平成20年度にフェーズ3を実施する計画を作成した。実際の研究も当初計画に基づいて実行され、おおむね計画通りの実施状況となった。

## 4. 研究成果

まず、初年度には実際の開発現場から収集された危険因子に関するアンケートデータに対して、データマイニング手法を適用して、危険因子間の関係を明確化する作業に着手した。使用したデータマイニング手法は相関ルールマイニング法である。これにより、相関ルールと呼ばれる危険因子間の強い関係を知識として発見することが可能となる。また、得られた危険因子間の関係について妥当性の検証を行った。具体的には、抽出した危険因子間のルールが開発現場での経験に照らし合わせて有効かどうかの判定、得られた危険因子ルールを利用したプロジェクト混乱の予測実験、の2つを実施した。

研究の2年目は危険因子間の関係に基づくリスク予知手法の構築を中心に研究を行った。すなわち、既に得られている危険因子間の関係を利用したリスク予知手法の構築を試みた。この手法では、相関ルールマイニング法によって得られた危険因子間の関係に基づいて危険度の高い因子と関係のある危険因子を導出する。具体的には、まず、ソフトウェア開発プロジェクトの進行中にリスク要因アンケートを実施する。この結果はリスク予知システムへ入力され、相関ルールマイニング法で抽出された危険因子間の関係に基づいて危険性が高いと思われる因子が最終的に利用者に提示される。また、リスク予知手法は、計算機上にリスク予知システムとして構築した。

さらに、実際のプロジェクトから得られたデータを用いた実験によって、本手法の有効性を確認することを試みた。実験の結果、得られる相関ルール数は莫大な数になることが判明し、その絞り込みを行う技法の確立が必要となることが分かった。そのため、クラスタリング手法を応用することでルールの絞り込みを行い、膨大なルールを分かり易くクラスタ化することに成功した。

最終年度には、(独)情報処理推進機構ソフトウェア・エンジニアリング・センターの協力を得て、前年度までに構築したリスク予知システムの評価実験を行った。実験の過程では、各因子に対する対処策とその効果についても情報を収集することで、潜在的な危険因子に対する有効な対処法についての知識を

収集した。

実際の開発現場での適用が最も望ましい形であったが、そうした実環境の準備は困難であったため、実験は(独)情報処理推進機構ソフトウェア・エンジニアリング・センターが収集した企業横断的データを利用して行った。本手法の適用可能性と有効性を調べるために、ソフトウェアプロセスの改善案(例えば、プロジェクトが新規の業種・業務である場合、品質目標を明確にして実行可能性を検討し、テストでは開発者のスキルと人数について十分な体制を構築することで、不具合工数密度が低くなる、と言ったもの)を自動的に抽出し、その改善案がどの程度妥当かをソフトウェア技術者との議論によって調査した。

提案手法により抽出されたルールの中から得られた改善案の候補を、IPA/SEC側の研究グループが有用であるかを検討し、次に示す10個の改善案が有用であると判断された。

- (1) 言語、ツール利用に関するトレーニングをプロジェクト初期に実施する、または、メンバに経験者を多少追加したり、コーチングを行うことなどで、開発言語やツールを使いこなせるレベルに早急に上げていくことが必要である。
- (2) ドキュメント作成ツールを利用しない場合に、構成管理ツールを利用する。
- (3) ドキュメント作成ツールを利用しない場合に、デバッグ・テストツールを利用する。
- (4) プロジェクトマネージャのスキルが高い人を配置する。または、スキル不足のPMの場合は組織としてPMサポート機能で支援する。
- (5) プロジェクト内での責任所在を明確にする。
- (6) 計画時に、品質目標を明確にし、実行可能性を検討しておく。
- (7) ドキュメント作成ツールを利用しない場合に、プロジェクト管理ツールを利用する。
- (8) 要求仕様を明確にする
- (9) スキルのあるテスト要員を、十分な人数配置する
- (10) 計画時にコストの根拠と計画工数の妥当性を確認する

実験の結果、(1) 相関ルールの形で示される改善案は非常に理解しやすいこと、(2) 得られた改善案は納得のいくものであることが確認できた。このことから、得られた相関ルールをデータベース化することで、ソフトウェア開発におけるリスク予知が十分可能であると判断できた。

## 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)

- [1] 出張 純也, 水野 修, 菊野 亨, 菊地 奈穂美, 平山 雅之, "企業横断的データからのプロジェクト改善案の相関ルールマイニングによる抽出," 情報処理学会論文誌, 49(8), pp. 2791-2801, 2008年8月. 査読有.
- [2] 浜野 康裕, 天寄 聡介, 水野 修, 菊野 亨, "相関ルールマイニングによるソフトウェア開発プロジェクト中のリスク要因の分析," コンピュータソフトウェア, 24(2), pp. 79-87, 2007年. 査読有.
- [3] 中野 裕也, 水野 修, 菊野 亨, 阿南 佳之, 田中 又治, "コードレビューの密度と効率がコード品質に与える影響の分析," SEC journal, 2(4), pp. 10-17, 2006年11月. 査読有.

[学会発表] (計 9 件)

- [1] Junya Debari, Osamu Mizuno, Tohru Kikuno, Nahomi Kikuchi, and Masayuki Hirayama, "On Deriving Actions for Improving Cost Overrun by Applying Association Rule Mining to Industrial Project Repository," In Proc. of International Conference on Software Process 2008 (ICSP2008), LNCS 5006, pp. 51-62, May 2008. (Leipzig, Germany)
- [2] 出張 純也, 水野 修, 菊野 亨, 菊地 奈穂美, 平山 雅之, "相関ルールマイニングによる企業横断データにおける不具合工数密度の分析," 電子情報通信学会技術研究報告, 107(275, SS2007-36), pp. 35-40, 2007年10月. (仙台市)
- [3] 出張 純也, 浜野 康裕, 水野 修, 菊野 亨, "リスク要因間の特徴抽出のための相関ルールに対するクラスタ分析," ソフトウェア信頼性研究会第4回ワークショップ論文集, pp. 9-16, 2007年6月. (松山市)
- [4] 飯田 哲也, 安部 誠也, 水野 修, 菊野 亨, "ソフトウェアプロジェクト混乱予測に用いるベイズ識別器における順序尺度への対応による学習モデルの改良," ソフトウェア信頼性研究会第4回ワークショップ論文集, pp. 91-99, 2007年6月. (松山市)
- [5] 浜野 康裕, 水野 修, 菊野 亨, 菊地 奈穂美, 平山 雅之, "相関ルールマイニングの適用によるソフトウェア生産性の決定要因の分析," 情報処理学会第155回ソフトウェア工学研究会, 2007(33, 2007-SE-155), pp. 65-72, 2007年3月. (東京都)
- [6] 近堂 高広, 安部 誠也, 水野 修, 菊野 亨, "ベイズ識別器による混乱予測に基づくソフ

トウェアプロジェクト管理支援ツールの試作," 情報処理学会第155回ソフトウェア工学研究会, 2007(33, 2007-SE-155), pp. 57-64, 2007年3月.(東京都)

[7] 浜野 康裕, 天寄 聡介, 水野 修, 菊野 亨, "相関ルールマイニングを用いた混乱プロジェクトの特徴分析," 情報処理学会第154回ソフトウェア工学研究会, 2006(125, 2006-SE-154), pp. 1-8, 2006年11月.(吹田市)

[8] Sousuke Amasaki, Yasuhiro Hamano, Osamu Mizuno, and Tohru Kikuno, "Characterization of Runaway Software Projects Using Association Rule Mining," In Proc. of 7th International Conference on Product Focused Software Process Improvement (PROFES2006), LNCS 4034, pp. 402-407, June 2006. (Amsterdam, The Netherlands)

[9] Seiya Abe, Osamu Mizuno, Tohru Kikuno, Nahomi Kikuchi, and Masayuki Hirayama, "Estimation of Project Success Using Bayesian Classifier," In Proc. of 28th International Conference on Software Engineering (ICSE2006), pp. 600-603, May 2006. (Shanghai, China)

## 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

### (2)研究分担者

水野 修 (MIZUNO OSAMU)

大阪大学 大学院情報科学研究科 助教

研究者番号: 60314407