



研究ノート

ソフトウェア開発における混乱プロジェクトの発生確率を推定するモデルの開発とその評価

菊野 亨*

On Predicting the Probability of Undesirable Events Occurring in Software Development Projects : Construction of Software Risk Model and Its Evaluation

Key Words : Software Development, Risky Project, Logistic Regression Model, Empirical Evaluation

1. ま え が き

最近のIT(情報技術)関連の製品開発においては開発期間と製品の品質についての要求がますます厳しくなっている。ソフトウェアの開発プロジェクトでもそうした傾向は顕著に表れており、その開発期間は短くなると共に、より高い信頼性が求められてきている。こうした要求を満たす1つの典型的な解決策は、十分な人と時間をソフトウェアテストにかけることであった。しかし、テストに注意を払うだけでは十分でないことは自明であり、ソフトウェア開発の全プロセスを対象とした分析とそれに基づく改善を行うことがソフトウェア工学の1つの重要な分野となってきた。

我々の研究室では、この10年間、ある企業のソフトウェア開発プロセスの改善に協力しつつ、研究活動を展開してきている。そうした中で、稀にはあるが、プロジェクトマネージャに把握できない事態が発生していることが分かってきた。こうした混乱した状況が繰り返されている間に、プロジェクトの開発期間、品質が見込みから大きくずれを生じてしまう。このようなプロジェクトは通常、失敗プロジェクトと呼ばれ、その発生をいかにして減らすかが協力企業での関心事であった。

そこで我々はプロジェクトマネージャが状況を把握できなくなってしまうプロジェクト(これを混乱プロジェクトと呼ぶことにする)をなるべく早期に発見し、適切な対応策を取ることに挑戦することにした。しかし、混乱プロジェクトを発見する技術はどこにも無かったので、まずその技術開発を行う必要があった。

一方、企業との共同研究を通じて、「開発現場のプロジェクトマネージャは直感的にはあるが、プロジェクトが混乱しているかどうかは判断できる。しかもその主な原因についてもうすうすは感づいている」ことが分かっていた。但し、それを断定するだけの理論的根拠を持ち合わせていなかった。

本研究はこのような背景の下に、我々がこの2年間で行った成果をまとめている。本研究ではプロジェクトマネージャへのアンケート調査によって、開発現場でのノウハウを抽出することを試みた。その調査結果に基づいて、プロジェクトが混乱するリスク要因をロジスティック回帰分析により明らかにした。更に、その分析によって構築されるロジスティック回帰モデルを利用すると、与えられたプロジェクトが混乱する確率を計算できる。この分析と評価には協力企業で過去3年間に行われた40個の実際のプロジェクトデータを利用している。

2. 混乱プロジェクト

我々の研究で混乱プロジェクトを導入したねらいは、失敗プロジェクトになる前の時点でその兆候を見つけることである。直感的には、現場のプロジェクトマネージャにとって制御不能になっているプロジェクトを混乱プロジェクトと呼ぶ。

ここで少しだけ形式的な話をすると、開発コスト



* Tohru KIKUNO
1947年9月11日生
1975年大阪大学大学院基礎工学研究科・博士課程(物理系専攻)修了
現在、大阪大学大学院・基礎工学研究科・情報数理系専攻・基礎工学部・情報科学、教授、工学博士、情報工学
TEL 06-6850-6565
FAX 06-6850-6569
E-Mail kikuno@ics.cs.osaka-u.ac.jp

と開発期間の2つに注目して混乱プロジェクトを定めることができる。これらについて、開発計画作成時に想定した値と実績値(あるいは、終了時点で予想される値)の間のずれを計算し、それがある基準値(例えば20%)を越えていると混乱プロジェクトとする。

但し、この形式的な定義だけに基づいて判断すると開発プロジェクトごとの事情を見落としてしまう恐れがある。そこで本研究では形式的な定義だけではなく、プロジェクトマネージャと品質管理部門のインタビューによる同意も混乱プロジェクトの判定に利用する。これによって、開発コストや開発期間については何の問題も無いが実際に開発現場が混乱したプロジェクトを混乱プロジェクトとして識別することができた。一方、開発コストや開発期間については問題のあるプロジェクトでもその原因のほとんどが技術的でないものについては、今回の議論の対象から除いた。

3. ロジスティック回帰分析

ここではロジスティック回帰モデルを構築して、混乱プロジェクトの発生確率を推定する。そのモデルの構築に当たっては、開発現場のプロジェクトマネージャのノウハウを利用する。つまり、プロジェクトが混乱する原因とプロジェクトマネージャが考えるものをロジスティック回帰モデルの説明変数として採用する。

3.1 分析モデルの概要

ロジスティック回帰モデルは一般に次式

$$E(Y | x_1, x_2, \dots, x_n) = \frac{e^{b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n}}{1 + e^{b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n}}$$

で表現される。ここでYは目的変数で2値の事象に対応する。本研究ではYを「プロジェクトが混乱する/しない」という2値の事象に対応させる。一方、 x_1, x_2, \dots, x_n はこのモデルにおける説明変数である。 b_0, b_1, \dots, b_n は係数であって、ロジスティック回帰分析によって決定される。今回の場合、 $E(Y | x_1, x_2, \dots, x_n)$ はプロジェクトが混乱プロジェクトである確率を表す。

ロジスティックモデルを統計的に評価する基準として、p値とデビアンスが利用されている。p値は仮説「 $H_0: b_0 = b_1 = \dots = b_n = 0$ 」が成り立つ確率を示しており、モデルの有意性を判定する。通常、

$p < 0.05$ であれば仮説 H_0 は棄却される。一方、デビアンスはモデルの適合度を示す値であり、デビアンスがモデルの自由度に近い値をとると適合度が良いとされる。

3.2 混乱プロジェクトのリスク予測

プロジェクトが混乱する原因についてはこれまでもリスク管理という研究分野の中で多くの研究報告がなされている。本研究ではそれらの報告と協力企業で長年にわたって利用されている内部規約を調査した。その結果、混乱プロジェクトのリスク要因を次の5つの主要な要因に整理した。(従って、今回は3.1で述べた説明変数は5個($n=5$)となり、 x_1, x_2, \dots, x_5 が主要な要因に対応する。)

- (1) 要求仕様の定義と理解に関する問題点
- (2) 実現すべきプロダクトの規模や機能の見積りに関する問題点
- (3) 開発チームの編成と人材(能力)に関する問題点
- (4) 開発計画の作成方法とその内容に関する問題点
- (5) 技術的な事項や外的事項に対するプロジェクト管理に関する問題点

表1 リスク要因の分類

項目		評価
1.	要件	
1.1	要件が不明確なままでの要求(要求側の問題)	
1.2	要件の引出し力不足(実現側の問題)	
1.3	要件の理解力不足(実現側の問題)	
1.4	要件に対する顧客側、実現側相互の合意不足	
2.	見積り	
2.1	見積項目不足	
2.2	見積の重要さの認識不足	
2.3	非技術的圧力に妥協	
2.4	技術的課題を過小評価してしまった	
2.5	見積りの根拠不足	
3.	プロジェクト体制	
3.1	要員のスキル不足	
3.2	プロジェクト体制整備の必要性認識不足	
4.	工程計画	
4.1	作業に対する責任分担不明確	
4.2	作業成果物定義不十分	
4.3	マイルストーンやレビュー時期の設定不足	
4.4	計画に対する設計者全員のコミットメントがない	
4.5	計画に対する上級マネージャのレビュー不足	
4.6	進捗管理方法不明確	
5.	プロジェクト進捗管理	
5.1	技術的側面でのリスク管理不足	
5.2	モラル不足	
5.3	工数不足	
5.4	要求・仕様変更管理不足	
5.5	進捗状況報告不足	
5.6	進捗管理データ不足	

表 2 測定データ

Projects in '96 and '97	要求仕様	見積り	プロジェクト体制	工程計画	プロジェクト管理	混乱 / 成功
PJ1	1.50	1.00	0.50	0.83	1.00	混乱
PJ2	1.50	1.00	0.50	1.17	1.00	混乱
PJ3	0.75	1.40	1.00	0.50	0.17	混乱
PJ4	1.50	0.40	1.00	0.83	0.83	混乱
PJ5	1.00	1.00	0.50	0.50	0.83	混乱
PJ6	0.75	0.00	0.00	1.00	0.33	混乱
PJ7	1.50	1.40	2.00	1.67	1.17	混乱
PJ8	1.75	1.80	2.00	1.83	1.33	混乱
PJ9	0.75	0.60	0.00	1.00	0.83	混乱
PJ10	0.00	0.00	2.00	2.00	1.00	混乱
PJ11	1.75	1.20	1.50	0.67	1.00	成功
PJ12	0.50	0.00	0.00	0.00	0.00	成功
PJ13	1.00	0.00	0.50	0.67	1.17	成功
PJ14	0.00	0.00	0.00	0.00	0.33	成功
PJ15	0.00	0.00	0.00	0.33	0.17	成功
PJ16	0.00	0.20	0.00	1.17	0.33	成功
PJ17	0.00	0.80	1.50	1.17	0.67	成功
PJ18	0.75	0.40	1.50	0.00	0.67	成功
PJ19	0.00	0.00	0.00	0.00	0.50	成功
PJ20	0.50	0.00	1.00	1.17	0.50	成功
PJ21	0.75	0.20	0.00	0.33	0.17	成功
PJ22	1.25	0.80	0.00	0.00	0.50	成功
PJ23	0.00	0.40	0.00	0.33	0.50	成功
PJ24	0.50	0.00	0.00	0.67	0.50	成功
PJ25	1.25	0.40	0.00	0.67	0.83	成功
PJ26	0.75	0.00	0.00	0.50	0.17	成功
PJ27	0.75	0.80	0.00	0.17	0.17	成功
PJ28	0.00	0.00	0.00	0.50	0.33	成功
PJ29	1.50	0.40	0.50	0.50	0.50	成功
PJ30	0.00	0.60	0.00	0.17	0.33	成功
PJ31	0.00	0.00	0.00	0.17	0.00	成功
PJ32	0.00	1.20	0.50	0.17	0.33	成功

5つの主要な要因はそれぞれ、より詳細なレベルの項目に展開される。例えば、(1)の要求仕様に関しては、(1.1)カスタマからの要求が不合理なものである、(1.2)要求仕様が実現上必要な項目を含んでいない、(1.3)開発者が要求仕様の中味を誤解した、(1.4)要求仕様がカスタマと開発者の間で合意されたものになっていない、と詳細化されている。その他の要因も含めたリスク要因の記述を表1に示す。

4. 適用実験 1 (モデル式の作成)

協力企業において1996年と1997年に行われた32個のプロジェクトPJ1~PJ32のマネージャに対し、表1に示すリスク要因の詳細な各項目毎に回答を求めた。回答は「該当する」「恐らく該当する」「分からない」「該当しない」の4種類のものを許した。得られた各回答にそれぞれ評価値2, 1, 1, 0を割り当てた。続いて各リスク要因ごとに評価値の平均を取った。得られた測定データを表2に示す。

なお、プロジェクトPJ1~PJ32についてはそれが混乱プロジェクトであったかどうかの判定が既になされている。その判定結果も表2に示している。32個のプロジェクトの中で10個が混乱プロジェクトであった。今の場合の混乱プロジェクトの比(10/32 ≃ 0.3)を次節では混乱プロジェクトを識別するときのしきい値として採用する。

表2のデータに対してロジスティック回帰分析を適用した。その結果、次のモデル式が求まった。

$$E(Y|x_1, x_2) = \frac{e^{b_0 + b_1x_1 + b_2x_2}}{1 + e^{b_0 + b_1x_1 + b_2x_2}}$$

このモデルに採用されたのは見積り(x₁)と工程計画(x₂)の2つである。表3に係数b₀, b₁, b₂の値を示す。表3でオッズ比はその係数が目的変数(今の場合、プロジェクトが混乱する確率)に対して与える影響の大きさを表す。p値はその説明変数をモデルに含めることが統計的に有意であるかどうかを検定した結果である。一般的に言って、0.2以下であ

表3 ロジスティックモデルの係数

	係 数	オッズ比	p-value
定 数	-5.251(b_0)	0.005	0.0041
見 積 り(x_1)	2.727(b_1)	15.288	0.0005
工程計画(x_2)	3.984(b_2)	53.749	0.0175

ればその変数をモデルに含めることが妥当である。

次に、作成したモデルに対するデビアンスの値と尤度比検定の結果を表4に示す。モデルのp値は $p < 0.0001$ であり、統計的に有意であることが示された。また、デビアンスの値が20.6であるのに対し、自由度が25なのでモデルの適合度も悪くないと判断できる。

表4 モデルの統計量

	自由度	χ^2	p-value
デビアンス	25	20.596	0.7149
尤度比検定	2	19.154	<0.0001

5. 適用実験2(モデル式による混乱予測)

1998年に行われた8個のプロジェクトPJ33~PJ40を新規プロジェクトと見なして、4節で作成したモデル式に基づく混乱プロジェクトの予測が可能かどうかを調べる。

まず、モデル式に従って、8個のプロジェクトのマネージャに対し表1中の2つのリスク要因(見積りと工程計画)について(4節と同様の方法で)回答を求め、それから測定データを算出する。その結果を表5の x_1 と x_2 に示す。次に、4節で作成したモデル式にそれらの値を代入して $E(Y|x_1, x_2)$ の値を求める。 E の値も表5に示す。

プロジェクトが混乱する確率 E の値は $0 \leq E \leq 1$

表5 計算された混乱確率

Projects in 1998	見積り (x_1)	工程計画 (x_2)	$E(Y x_1, x_2)$	混乱/成功
PJ33	0.60	2.00	0.987	混乱
PJ34	1.20	1.33	0.966	混乱
PJ35	1.20	0.33	0.343	混乱
PJ36	0.00	0.17	0.010	成功
PJ37	0.40	0.33	0.056	成功
PJ38	0.20	0.33	0.033	成功
PJ39	0.00	0.17	0.010	成功
PJ40	0.40	0.67	0.182	成功

となっている。既に説明した様に、モデル式の導出に利用したデータの1つの特性値である0.3をしきい値として利用する。つまり、 $E \geq 0.3$ のプロジェクトを混乱(すると予測される)プロジェクトであると判定する。その結果、表5に示すように、3個のプロジェクトPJ33, PJ34, PJ35が混乱プロジェクトであると判定される。

上述したように、本実験で利用した8個のプロジェクトも既に開発を終了しており、それが混乱プロジェクトであったかどうかの判定も既になされている。その判定結果を表5の混乱/成功に示している。

更に、これら2つの判定結果の比較を表6に示す。表6より全てのプロジェクトについて判定結果が一致していることが分かる。

表6 98年のプロジェクトに対する予測結果

実 測	予 測 結 果	
	成 功	混 乱
成 功	5	0
混 乱	0	3

6. む す び

本研究ではソフトウェア開発プロジェクトのリスク要因をプロジェクトマネージャからのアンケート回答を利用して解明した。次に、その成果に基づいて、プロジェクトが混乱する確率を推定するモデル式を導出した。更に、実際の開発プロジェクトデータを利用した評価実験から、提案するモデル式が非常に高い精度で推定できることを確認した。

謝 辞

本研究の一部は平成12年度文部省科学研究費補助金(基盤研究C(2))(課題番号:12680347)、平成11年度HITOCC研究資金支援、オムロン(株)ソーシャル事業グループ開発・生産センタからの奨学寄付金からの援助によって行われたものであります。

また、本研究を行うに当たってオムロン株式会社(現在SPIコンサルタント)坂本啓司博士、オムロン株式会社 高木徳生氏からのご協力は極めて重要でありましたので深謝いたします。

最後に、本研究における詳細なデータ分析を担当し、非常に本質的な貢献をしていただいた大阪大学大学院基礎工学研究科 水野修氏に心より御礼申し上げます。

参 考 文 献

- 1) 足立, 水野, 菊野, 高木, 坂本, “アンケート調査に基づく開発中のプロジェクトの混乱予測とその予測作業支援システムの開発”, ソフトウェアシンポジウム論文集, pp.146-153, 2000.
- 2) O. Mizuno, T. Kikuno, K. Inagaki, Y. Takagi and K. Sakamoto, “Analyzing effects of cost estimation accuracy on quality and productivity”, Proc. of 20th International Conference on Software Engineering (ICSE'98), pp.410-419, 1998.
- 3) O. Mizuno, T. Kikuno, Y. Takagi and K. Sakamoto, “Characterization of Risky Projects based on Project Managers' Evaluation”, Proc. of 22nd International Conference on Software Engineering(ICSE2000), pp.387-395, 2000.
- 4) 丹後, 山岡, 高木, ロジスティック回帰分析—SASを利用した統計解析の実験—, 朝倉書店, 1996.

