

プロジェクトマネージャへのアンケートに対する 因子分析に基づいたソフトウェア開発コスト推定モデルの提案

足立 卓也[†] 水野 修[†] 菊野 亨[†] 高木 徳生^{††}

[†] 大阪大学大学院基礎工学研究科 情報数理系専攻

〒 562-8531 大阪府豊中市待兼山町 1-3

^{††} オムロン株式会社 ソーシャルシステムズカンパニー プロセス革新部

〒 525-0035 滋賀県草津市西草津 2-2-1

E-mail: [†]{adati,o-mizuno,kikuno}@ics.es.osaka-u.ac.jp, ^{††}taka@biwa.kusatsu.omron.co.jp

あらまし 本研究ではリスク調査アンケート結果に対して因子分析を行い、そこで抽出された因子に基づく開発コストを推定する手法について述べる。ソフトウェア開発プロジェクトに対するリスク調査アンケートの項目には、直接的には開発コストに対して統計的に有意な関係はみられないが、他の項目に対して影響を与えているものが存在する。従来の回帰モデルによるコスト推定手法ではこうした項目を除外してしまうため、その影響を考慮に入れることができなかった。本研究で新しく提案する手法では、因子分析によってアンケート項目の再編成を行い、23 個の質問項目をいくつかの因子に集約する。更に、得られる因子スコアを利用してコストに対する回帰式を作成する。適用実験の結果、多くのリスク要因を含むコスト予測が行えることを確認した。

キーワード ソフトウェア開発プロジェクト, リスク管理, コスト見積, 因子分析

A New Cost Estimation Model Refined by Factor Analysis based on Risk Questionnaire to Project Managers

Takuya ADACHI[†], Osamu MIZUNO[†], Tohru KIKUNO[†], and Yasunari TAKAGI^{††}

[†] Graduate School of Engineering Science, Osaka University
Machikaneyama 1-3, Toyonaka-shi, Osaka, 560-8531 Japan

^{††} Social Systems Company, OMRON corporation
Nishi-kusatsu 2-2-1, Kusatsu-shi, Shiga 525-0035 Japan

E-mail: [†]{adati,o-mizuno,kikuno}@ics.es.osaka-u.ac.jp, ^{††}taka@biwa.kusatsu.omron.co.jp

Abstract This paper proposes a cost estimation method using factor analysis based on a risk questionnaire for project managers. We have tried to develop a method to estimate costs and durations by the risk questionnaire and multiple regression analysis. However, some risk factors in the questionnaire, that have indirect effects to software development cost and duration, are sometimes excluded from estimating expressions. In the proposed method, we summarize risk factors into some representative factors by the factor analysis. We then perform multiple regression analysis to estimate development costs and durations using summarized factors. The result of experimental evaluation shows effectiveness of the proposed method.

Key words Software development project, Risk management, Cost estimation, Factor analysis

1. ま え が き

ソフトウェア開発プロジェクトには多くのリスク要因が複雑に関係して発生する問題が多いことが知られている [3],[6] . 特に, コストや開発期間の増加に伴う問題は其中でも重要なものである .

また, 実際の開発現場ではプロジェクトの状況が把握できなくなってしまい, プロジェクトが混乱してしまうことが少なくない [14] . 多くの企業ではプロジェクトマネージャがプロジェクトを監視し, 混乱を未然に防ぐ試みがなされている [4],[13] .

我々は, ある企業の協力の下でソフトウェア開発プロジェクトの混乱状態をロジスティック回帰分析により統計的に予測する研究を行ってきた [8] . また, この研究の過程で開発現場からのニーズとして, プロジェクトが混乱するという事象の予測よりも, コストや期間が見積りに比べ増加するかどうかのほう重要な問題であるとの認識を得た . そこでリスクアンケート調査の回答とコスト・期間の初期見積りからの誤差との関連を重回帰分析を用いて調べてきた [2],[9] .

これまでの研究で用いてきた手法では, リスク調査アンケートの回答項目を独立変数の候補と見なし, ステップワイズ法によって独立変数の絞り込みを行い, 回帰分析を行ってきた . しかし, この方法では関連が薄いと判断された項目は除外されてしまうので, 結果的にごく一部の特徴的な項目に基づいた回帰式による予測となっていた . またアンケートの回答値が 4 段階の値であったため推定されたコスト誤差, 期間誤差も順序関係しかもたないものであった .

そこで本研究ではリスク調査アンケート結果に対する因子分析を行い, そこで抽出された因子に基づく開発コストと開発期間の誤差を推定する試みについて述べる . 因子分析により, できる限り多くの項目から得られた情報を使った因子を抽出する . それにより得られた因子スコアを使用して重回帰式を作成することにより順序関係に依存しない予測式の作成をする . また, 実際のプロジェクトに対して行ったリスク調査アンケートから得られたデータに対して適用実験を行う .

2. 研究の背景と目的

2.1 これまでの成果

我々の研究グループでは, プロジェクトマネージャに対するリスク調査アンケートからプロジェクトの最終的な状態 (混乱したか否か) を予測する研究 [1],[8]

やコストと期間の見積りと実測の間の誤差を予測する研究 [2],[9] を行ってきた . ここではそれらについて簡単に述べる .

(a) リスク調査アンケート表の作成

まず, 開発現場やプロジェクトマネージャが意識しているリスク要因を調べるためのアンケート表を作成した (表 1) . これは実際に SEPG においてリスクのチェックリストとして使用されていたものを再編成したものである .

表 1 アンケート調査表

1.要求仕様
1.1要件が不明確なままでの要求
1.2要件の引出し不足
1.3要件の理解力不足
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見積り項目不足
2.2見積りの重要さの認識不足
2.3非技術的圧力に妥協
2.4技術的課題に対する楽観的見積り
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計画に対する関係者全員のコミットメントなし
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.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要件, 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

(b) ロジスティック回帰分析による統計的混乱予測モデルの作成 [1],[8]

プロジェクトの最終状態を 2 値の事象「混乱有り」「混乱無し」と位置づけ, 先に収集したリスク調査アンケートの回答から混乱の有無への回帰モデルを作成した . ここではロジスティック回帰分析を利用した . また, 1996, 1997, 1998 年に実施された合計 40 件のプロジェクトに対してアンケート調査を行った . 実験では 1996, 1997 年のアンケートデータから作成したモデルを 1998 年のプロジェクトの混乱予測に使用した . その結果, 非常に高い精度で混乱を予測できることが確かめられた . また, 上で述べたアンケートの配布・収集, そして回帰モデルの作成を Web ブラウザ上から行うシステムを実装した .

(c) 重回帰分析によるコスト・期間の見積と実績の誤差予測 [2],[9]

プロジェクトの最終状態でのコスト・期間の実測

値とそれらの見積りの誤差(コスト誤差, 期間誤差と呼ぶ)を目的変数とし, リスク調査アンケートの回答からコスト誤差, 期間誤差を推定する重回帰モデルを作成した. (b)と同様の予測実験を行ったところ, 提案するモデルがある程度の推定精度を持つことが確認できた.

2.2 本研究の目的

これらの研究を進めてきた中で, 次のような課題 P1, P2 が開発現場から指摘されてきた.

課題 P1 アンケート項目の評価が順序尺度であるため, 回帰分析の結果得られる推定値も順序関係しか保存していなかった. そのため, データのある全プロジェクト中での順位という形での評価しかできなかった. つまりコストの値そのものの評価はできなかった.

課題 P2 また, 現場へのフィードバックを考えた場合, 少数の項目に対する回答だけで判断が確定すると, ある特定のリスク要因のみが重要であり, その他の項目は重要でないという錯覚を招きやすい.

本研究では, これらの課題 P1, P2 に対処することを目的とする.

具体的には, まず, プロジェクトマネージャから得られたリスク調査アンケートの回答に対して, 因子分析を行う. 因子分析の結果, 回答の傾向が似たアンケート項目は集約され, 幾つかの因子にまとめられる. そして得られた因子に対して各プロジェクトが持つ因子スコアを算出し, この因子スコアに基づいて, コスト誤差に対する重回帰分析を行う.

この結果, 従来のように幾つかのアンケート項目だけを用いてコスト誤差に対する重回帰を行った場合に比べ, より多くのアンケート項目が因子の構成要素として考慮に入れられることになる. これにより, 一部の要因に依存した推定モデルではなく, より多くの情報を反映した推定モデルを構築できると期待できる. また, 因子分析により算出される因子スコアを重回帰分析に用いることにより, 推定値を比率尺度として扱うことができる.

3. 因子分析を用いたコスト誤差, 期間誤差の予測手法

3.1 メトリクスの定義

プロジェクトマネージャに対して表 1 に示すアンケート調査表を配布し, 彼らの手がけたプロジェクトに対する回答を回収した. プロジェクトマネージャはこれらの項目に対して「はい」「いいえ」「どちら

でもない」「分からない」の中から回答を選択する. それらの回答に対して「はい」なら 4 点を「いいえ」なら 1 点を「どちらでもない」なら 3 点を「分からない」なら 2 点を割り当てた. それらの値は順序関係しか持たないので, 順序尺度のメトリクスと見なされる.

また, それぞれのプロジェクトのコストが, 見積りコストに対してどれくらいの割合で増減したのかを測るコスト誤差 RE_{cost} を評価した. 更に見積り期間に対しても開発期間がどれくらい増減したのかを測る期間誤差 $RE_{duration}$ を評価する. これら 2 つのメトリクスの定義は次に示す通りである.

$$RE_{cost} = \frac{actCOST}{estCOST}$$

RE_{cost} : コスト誤差
 $actCOST$: コスト実測値 (人月)
 $estCOST$: コスト見積り値 (人月)

$$RE_{duration} = \frac{actDUR}{estDUR}$$

$RE_{duration}$: 期間誤差
 $actDUR$: 期間実測値 (月)
 $estDUR$: 期間見積り値 (月)

3.2 因子分析

因子分析とは, 多変量データをより低次元のデータに集約し, 現象の理解を容易にすることを目的としている. 本来, 様々なデータは互いに完全に独立ではなく, 相関のあるデータも存在している. 因子分析により, それらのデータから, より少数の独立した因子を探り出すことが可能である. 従来の我々の研究では, 観測データをそのまま重回帰分析の独立変数の候補と考えて回帰式の作成を行ってきた. しかしその場合採用されるデータにおいても完全に独立ではなく, またデータのうち一部しか結果に反映されない回帰式となっていた. 先に因子分析をデータに適用することにより, よりデータを有効に用いたモデルを作成することが可能になる.

ここでは観測対象数が N , 観測項目数が n , 抽出された因子数が m の場合の因子分析の基本モデルを示す.

$$Z = FA' + UD$$

Z は観測値を標準化したデータの $N \times m$ の行列である. 本報告では Z がアンケートに対する回答データにあたる. F は因子スコア行列と呼ばれる. 行に観測対象 (特に本報告ではプロジェクトとなる), 列に因子をとった $N \times m$ の行列である. つまり抽出さ

れた因子とプロジェクトの関係を示したものになる。また A は因子負荷行列と呼ばれる。これは、行に変量（本報告ではアンケート項目）、列に因子をとった $n \times m$ の行列である。それぞれの因子がアンケート項目とどのような関係にあるのかを $-1 \sim 1$ の数値で示したものとなる。 UD は独自因子で因子分析では説明できない部分となり、考慮には入れない。

この基本式を基に、 F と A の値を推定するのが因子分析の主な役割である。まず観測データから因子数を推定する。そしてその因子に対する因子負荷行列、因子スコア行列を計算していく。因子分析の手法には多くのものが知られている [12]。本報告での因子負荷行列の推定には最も一般的に使われる主因子法、因子数の推定には 75% 分散法を利用している。

3.3 因子スコア

因子分析の結果を外的要因と関連付けて考える場合に利用されるのが因子スコアである。前述の F がそれにあたる。これは抽出されたそれぞれの因子に対して観測対象がどれだけの影響を持っているのかを表したものである。つまり本報告の場合、プロジェクトがそれぞれの因子に対してどれだけ高いリスクをもっているのかを表していることになる。この値を基に重回帰分析を行い、コスト誤差、期間誤差の推定式を作成する。なお因子スコアの値は $-1 \sim 1$ の間の連続値をとる。

3.4 重回帰分析

重回帰分析では 2 つ以上の独立変数から従属変数の値を予測する。具体的には次に示す重回帰式を作成する。

$$Y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n$$

ここで Y は従属変数、 $x_i (1 \leq i \leq n)$ は独立変数であり、収集されたデータを用いて係数 $a_i (0 \leq i \leq n)$ の値を推定する。運用時には、こうして作成した予測式に x_i の値を代入して予測値 Y を得る。

重回帰分析とは、複数の独立変数が従属変数に及ぼす線形の効果を探るものである。本研究では、因子スコアと開発コスト誤差、開発期間誤差の間に線型関係の仮定をおき、その関係の説明を試みる。また推定された係数をみることで、どの因子がコスト、期間に大きく影響を及ぼしているのかということが分かる。

3.5 提案する手法

提案する手法は次のステップから構成される。

Step1 アンケートにより得られたアンケート回答値に因子分析を適用することにより、アンケート項目

を統合し簡略化をする。その際に得られる値は、因子数、因子負荷量、因子スコアである。

Step2 得られた因子スコアを独立変数、開発コスト誤差、開発期間誤差を従属変数として重回帰分析を適用する。それにより、開発コスト誤差、開発期間誤差にそれぞれ対する重回帰式が得られる。

4. 適用実験

ここでは、提案する手法に基づいて実際のプロジェクトデータに対して行った適用実験について述べる。

4.1 対象プロジェクト

本報告で対象にしたプロジェクトは、ある企業において 1996 年から 1998 年に実施された 40 個の組み込みソフトウェアの開発プロジェクトである。これらに対してリスク調査アンケートを配布し、それらに対する回答を得た。さらに、この企業で従来から実施されているソフトウェアメトリクス収集活動の結果 $actCOST, estCOST, actDUR, estDUR$ を得た。しかしながら、3 プロジェクトにおいて期間に関するデータが不正確であったため、これらを分析から除外することにした。その結果、今回の分析では 37 個のプロジェクトデータを用いる。

4.2 因子分析結果

まず各プロジェクトに対するアンケートの回答値に因子分析を適用した。因子数の決定法として 75% 分散法、因子の抽出法として主因子法、また因子解の回転法としてバリマックス変換を利用した。

因子分析を実行した結果、アンケート項目は 7 つの因子に集約された。表 2 に各アンケート項目に対する因子負荷量行列を、また、表 3 に各プロジェクトに対する因子スコア行列を示す。

表 2 に示す因子負荷量を元に、それぞれの因子の意味を考えることができる。ここでは特に因子 2 と因子 3 について例をあげる。

表 2 における因子 2 に対する因子負荷量から図 1 に示す。図 1 からアンケート項目の 1.3, 1.2, 1.4 が因子 2 に特に強く関わっていることが分かる。表 1 から、項目 1.3 は「要件の理解力不足」、1.2 は「要件の引出し力不足」、1.4 は「要件に対する顧客側、実現側相互の合意不足」、となっている。これらのことから、因子 2 は要求仕様段階での危険性を表している。

同様に因子 3 に対する因子負荷量 (図 2) から、アンケート項目の 2.5, 2.3, 2.2 が因子 3 に特に強く関わっていることが分かる。表 1 から、項目 2.5 は「見積りの根拠不足」、2.3 は「非技術的圧力に妥協」、2.2 は

表2 各アンケート項目に対する因子負荷量

	factor1	factor2	factor3	factor4	factor5	factor6	factor7
1.1	-0.096	0.235	0.054	0.195	0.173	-0.067	0.759
1.2	0.072	0.911	-0.005	-0.024	0.112	-0.001	-0.148
1.3	-0.008	0.958	0.030	0.017	0.006	-0.025	-0.084
1.4	-0.039	0.793	-0.085	0.239	0.273	0.118	0.030
2.1	-0.064	0.547	0.292	-0.350	-0.050	0.116	0.353
2.2	0.213	0.460	0.541	-0.055	-0.404	-0.198	0.066
2.3	0.035	0.071	0.592	0.072	0.387	0.282	-0.001
2.4	0.082	0.040	0.414	-0.391	0.190	0.058	0.518
2.5	-0.149	-0.095	0.808	0.137	-0.113	0.087	0.018
3.1	0.078	0.011	0.142	0.144	0.027	0.504	0.348
3.2	0.102	-0.097	0.435	0.135	0.070	0.591	-0.052
4.1	-0.092	0.032	0.214	0.766	-0.048	0.247	-0.120
4.2	0.076	0.014	-0.025	0.447	0.000	0.343	0.447
4.3	0.370	0.001	-0.035	0.686	-0.057	-0.248	0.171
4.4	0.294	0.071	-0.008	0.162	-0.031	0.636	-0.019
4.5	0.611	0.000	-0.179	0.015	0.183	0.480	-0.077
4.6	0.863	-0.024	-0.057	0.010	0.009	0.161	-0.010
5.1	-0.041	0.358	-0.170	-0.176	-0.263	0.554	0.382
5.2	-0.022	0.705	-0.115	0.014	-0.222	0.002	0.349
5.3	-0.084	0.043	0.098	-0.128	-0.074	0.822	-0.058
5.4	0.060	0.073	-0.015	-0.094	0.864	-0.115	0.104
5.5	0.602	-0.162	-0.018	0.180	-0.007	0.188	0.423
5.6	0.898	0.110	0.110	-0.004	-0.042	-0.220	-0.128

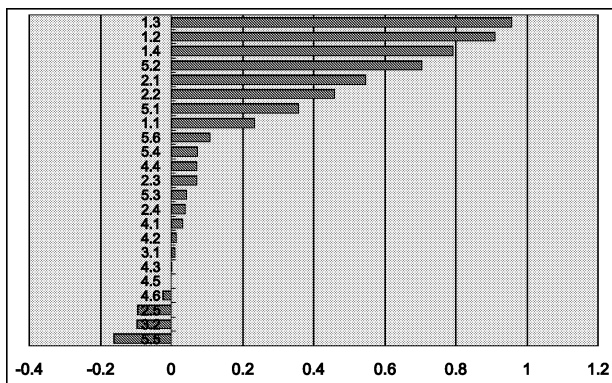


図1 因子2の因子負荷量

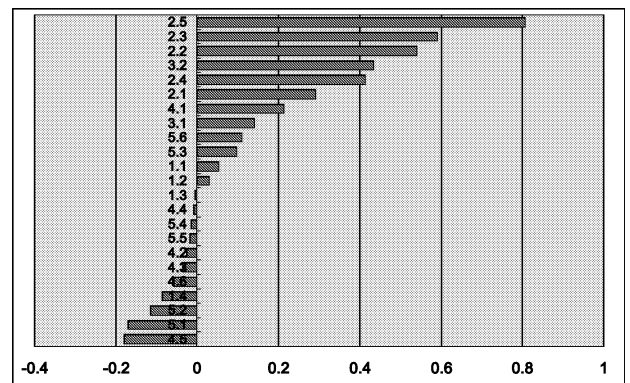


図2 因子3の因子負荷量

「見積りの重要さの認識不足」, となっている。これらのことから, 因子3は見積もり段階での危険性を表していると考えられる。

同様に考えると残りの因子は以下のように名前を付けることができる。

- 因子1: 進捗管理に関する危険性
- 因子4: 工程計画に関する危険性
- 因子5: 要件や仕様の変更の管理に関する危険性
- 因子6: プロジェクトの体制に関する危険性
- 因子7: プロジェクトの目標設定に関する危険性

このように, それぞれの因子に対して特に影響の強

いアンケート項目をそれぞれの因子の特徴と考える。

4.3 重回帰モデル

ここでは表3に示す因子スコアを利用してコスト誤差, 期間誤差の回帰式を求める。

先の因子分析によって得られた7つの因子に対する因子スコアを独立変数の候補とし, 開発コスト誤差, 開発期間誤差を従属変数として重回帰式を作成する。予測式作成についてはステップワイズ法を用いた。その結果得られた開発コスト誤差の回帰式は次のように求まった。(予測式の左辺の re_{cost} は実測値として求まる RE_{cost} と区別する目的で小文字の表記を用いる。)

$$re_{cost} = 1.091 + 0.112x_2 + 0.124x_3 + 0.075x_5$$

表3 37プロジェクトに対する因子スコア

PJ.NO.	factor1	factor2	factor3	factor4	factor5	factor6	factor7
1	1.619	0.927	1.080	0.500	-2.112	-2.109	2.035
2	0.961	1.070	-1.085	-1.590	0.568	1.518	1.099
3	-0.989	0.105	2.205	0.258	1.475	0.327	-0.725
4	0.671	1.587	0.992	-1.018	-0.735	0.125	-1.248
5	1.180	0.728	1.643	0.760	1.130	1.031	0.016
6	-0.969	1.855	-0.404	1.781	-0.095	-0.062	0.976
7	-1.110	-1.196	1.316	-0.190	0.947	1.174	1.962
8	0.557	1.054	1.555	0.719	0.304	1.727	0.549
9	2.124	-0.280	-0.490	-0.617	1.315	-1.166	1.264
10	-0.090	1.104	-1.275	1.048	0.640	0.027	1.156
11	2.060	-1.411	-0.943	1.496	-1.393	1.306	0.204
12	-0.783	-0.843	-0.405	0.117	-0.325	-0.476	-0.514
13	-0.370	-0.957	0.249	-0.370	1.035	-0.284	-0.444
14	-1.208	0.336	0.032	1.090	1.277	-0.885	-0.191
15	-1.359	-1.303	0.042	0.525	-1.138	1.035	1.070
16	-0.548	0.061	-0.599	-0.622	-0.646	-0.561	-0.885
17	2.125	0.486	-1.259	-1.044	0.938	-0.204	-0.672
18	-0.538	-0.866	-0.564	-0.743	0.702	-0.700	-0.183
19	-0.874	-0.941	-0.568	0.475	-0.349	-0.376	-0.126
20	0.467	-1.315	-0.308	1.262	-1.153	0.415	-0.223
21	1.619	-1.483	1.450	0.099	0.940	0.894	-0.784
22	-1.165	0.388	0.178	-1.003	0.156	1.278	-0.928
23	-0.629	-0.805	-0.885	-0.967	0.378	-0.345	0.086
24	0.247	-0.352	-1.172	0.840	-0.691	2.034	-0.812
25	-0.092	0.489	0.473	1.132	-0.562	-1.230	-1.442
26	-0.733	1.469	-0.772	-1.793	-0.973	-0.341	0.447
27	-0.801	1.181	0.242	0.568	-0.338	-0.028	1.972
28	-0.574	-0.750	0.182	-0.676	-0.477	0.332	-0.401
29	0.923	0.158	-0.615	0.653	1.229	-0.919	-1.288
30	-0.252	1.523	-1.260	-0.598	-1.217	1.636	-1.628
31	-0.708	0.247	-0.638	1.531	0.656	-1.053	-0.738
32	0.126	0.741	2.061	-0.199	-0.980	-1.304	-1.673
33	0.301	-1.080	-0.531	0.657	0.018	-0.956	0.079
34	-0.757	0.767	-0.819	-0.308	1.486	-0.017	0.642
35	0.141	-0.953	0.049	-1.842	0.802	-0.924	0.697
36	-0.381	-0.939	-0.581	-0.028	-0.911	-0.888	-0.041
37	-0.194	-0.805	1.421	-1.902	-1.900	-0.032	0.692

x_2 : 因子2

x_3 : 因子3

x_5 : 因子5

式の有意性を示す p 値^(注1)は0.0001より小さく、係数の p 値と共に0.05以下であったので統計的に有意である。またモデルの適合度を表す R^2 値^(注2)は0.482であった。

また、開発期間誤差の回帰式は次のように求まった。

$$re_{duration} = 1.157 + 0.099x_2 + 0.127x_3 + 0.093x_7$$

x_2 : 因子2

x_3 : 因子3

x_7 : 因子7

式の有意性を示す p 値は0.002であり、係数の p 値

(注1): ここで、 p 値とは「式中の係数が本当はすべて0であり上記のような値になったのはただの偶然だった」という仮説が成立する確率を示す。

(注2): ソフトウェア工学の文献では R^2 値は0.3程度の値があれば良いとされていることが多い[7]。

と共に0.05以下であったので統計的に有意である。またモデルの適合度を表す R^2 値は0.446であった。

またこれらの式から得られる推定値は比率尺度であり、推定値そのものの評価が可能である。

ここでコスト誤差に関して従来法と本手法の比較を考える。従来法ではコスト誤差に強い関わりを持つとされたアンケート項目は「要件の理解力不足」、「技術的課題に対する楽観の見積り」、「モラル不足」の3つであった。コスト誤差に影響を及ぼしているリスク要因がこの3つだけであるとは現実的には考えにくい。これに比べ今回の提案法においてコスト誤差の増減に関連しているとみられるアンケート項目は、それぞれの因子において影響の強い項目だけを考えても因子2では「要件の理解力不足」、「要件の引出し力不足」、「要件に対する顧客側、実現側相互の合意不足」などが挙げられ、また因子3では「見積りの根拠不足」、「非技術的圧力に妥協」、「見積りの重要さの認識不足」、因子5では「要件、仕様

変更管理不足」など多くのアンケート項目が関連していると言える。期間誤差についても同様である。

このように、コスト誤差、期間誤差について統計的に有意でモデルの適合度の高い回帰式を作成することができた。またここで二つの回帰式に採用された因子2と因子3については先に述べたように要求仕様、見積りに関わる因子であるので、これらの段階での影響がコスト誤差と期間誤差に強く反映していることが分かる。

4.4 予測実験

得られた重回帰式は開発コスト誤差、開発期間誤差について共に統計的に有意であり、モデルの適合度も悪くない。そこで、この手法の有効性を調べるために予測実験を行った。

予測実験の手順は次の通りである。

(1) 全37プロジェクトのうち、任意の1つを予測対象のプロジェクトとする。

(2) 残りの36プロジェクトのデータを用いて重回帰式を作成する。

(3) 得られた重回帰式を用いて(1)で選択した1プロジェクトのコスト誤差、期間誤差の予測を行う。

(4) (1)–(3)の手順を37プロジェクトの全てについて繰り返す。

この手順に基づいて37回の実験を行った結果を表4に示す。また、コスト誤差についての実測値と予測値の関係を表したグラフを図3に、期間誤差についての実測値と予測値の関係を図4に示す。

表4 予測実験結果

PJ.NO.	RE _{cost}	re _{cost}	RE _{duration}	re _{duration}
1	1.323	1.131	1.222	1.661
2	1.421	1.086	1.455	1.194
3	1.633	1.446	1.909	1.268
4	1.691	1.279	1.353	1.319
5	2.086	1.350	1.043	1.492
6	1.069	1.266	1.273	1.400
7	1.195	1.191	1.333	1.402
8	1.086	1.474	1.818	1.464
9	1.025	1.104	1.100	1.192
10	1.175	1.095	1.158	1.231
11	0.641	0.725	0.615	0.952
12	1.013	0.917	1.000	0.973
13	1.039	1.099	1.125	1.048
14	1.000	1.248	1.000	1.182
15	0.896	0.861	1.333	1.109
16	1.011	0.973	1.000	1.006
17	0.989	1.067	1.000	0.981
18	1.000	0.975	1.000	0.982
19	1.019	0.881	1.000	0.978
20	1.000	0.795	1.000	0.964
21	1.062	1.201	1.000	1.146
22	0.792	1.181	1.000	1.140
23	1.000	0.914	1.000	0.971
24	0.886	0.852	1.000	0.888
25	1.062	1.168	1.250	1.120
26	1.081	1.087	1.167	1.256
27	1.221	1.228	2.000	1.380
28	1.033	0.992	1.143	1.065
29	0.977	1.137	1.000	0.973
30	0.992	1.019	1.091	0.971
31	0.998	1.094	0.714	1.050
32	1.000	1.438	1.143	1.398
33	1.002	0.898	1.143	0.979
34	1.171	1.188	1.000	1.203
35	1.093	1.048	1.000	1.143
36	0.700	0.859	1.250	0.970
37	0.984	1.048	1.182	1.342

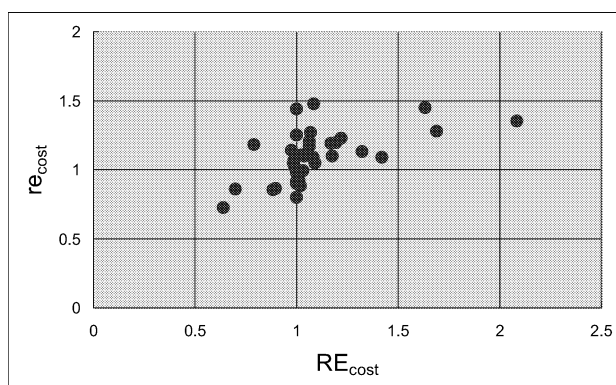


図3 コスト誤差の実測と予測の関係

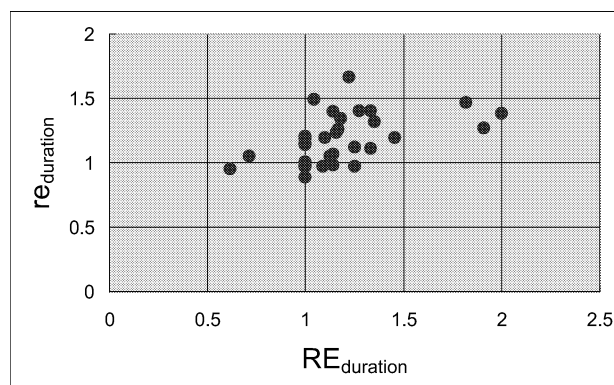


図4 期間誤差の実測と予測の関係

コスト誤差の予測値と実測値の相関係数は0.575、また期間誤差の予測値と実測値の誤差の相関係数は0.524となり、共にかなり相関があると言える[5]。この結果より、提案する因子分析を用いたコスト誤差、期間誤差の推定手法の有効性を統計的に確認できた。しかしながら、プロットしたグラフを見てみると、実際に非常に大きなコスト誤差、期間誤差を出してい

るプロジェクトに関してはあまり予測精度が良くないと言える。これは、回帰式の作成に使用したデータの中にそのようなプロジェクトの例が少ないことが主な原因と考えられ、またコスト誤差、期間誤差が非常に増大してしまうことにはアンケート項目以外の要因が関連しているとも考えられる。

5. ま と め

Developer's Guide to Surviving 'Mission Impossible' Projects, Prentice Hall Computer Books, 1997.

本報告ではソフトウェア開発プロジェクトのコスト誤差，期間誤差の推定手法として，因子分析と重回帰分析を用いる手法を提案した．提案する手法では，まずプロジェクトマネージャに対するリスク調査アンケートの結果に対する因子分析を行い，特徴的な因子の抽出を行った．次に，求めた因子を用いて重回帰分析を行い，コスト誤差，期間誤差の推定モデルを作成した．この結果より多くのアンケートの回答を反映し，従来法よりも現実的な推定モデルを作成できた．実際のプロジェクトデータに対する適用実験の結果，ある程度の精度でコスト誤差，期間誤差の予測が行えることを確認した．

今後の課題としては，より多くのプロジェクトデータに対する本手法の適用が考えられる．更には現場への適用により本手法の有効性の検証の必要がある．また，因子分析で求めた因子に対する名前付けを通じて，既存のアンケート項目をより統計的に意味のある項目へと再編成していくことも考えられる．

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 2000 論文集, pp.146-153, 2000.
- [2] 足立, 水野, 菊野, 高木, “リスク調査に基づくコストと期間の誤差予測とその統計的分析,” ソフトウェアシンポジウム 2001 論文集, pp.109-115, 2001.
- [3] B.W.Boehm, "Industrial software metrics top 10 list", IEEE Software, vol.4, no.5, pp.84-85, 1987.
- [4] E. H. Conrow and P. S. Shishido, "Implementing risk management on software intensive projects," IEEE Software, vol.14, no.3, pp.83-89, 1997.
- [5] 石村貞夫, 統計解析のはなし, 東京図書, 1989.
- [6] J. Jiang and G. Klein, "Software development risks to project effectiveness," Journal of Systems and Software, 52, pp.3-10, 2000.
- [7] M.S.Krishnan and M.I.Kellner, "Measuring process consistency: Implications for reducing software defects," IEEE Trans.Software Engineering.vol.25, no.6, pp.800-815, 1999.
- [8] 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, pp.387-395, 2000.
- [9] O. Mizuno, T. Adachi, T. Kikuno, and Y. Takagi, "On prediction of cost and duration for risky software projects based on risk questionnaire," Proc. of 2nd Asia-Pacific Conference on Quality Software, pp.120-128, 2001.
- [10] 水野欽司, 多変量解析講義, 朝倉書店, 1996.
- [11] 佐伯, 松原, 実践としての統計学, 東京大学出版会, 2000.
- [12] 芝祐順, 因子分析法, 東京大学出版会, 1986.
- [13] R. C. Williams, J. A. Walker and A. J. Dorofee, "Putting risk management into practice," IEEE Software, vol.14, no.3, pp.75-82, 1997.
- [14] E. Yourdon, *Death March : The Complete Software*