

プロセスデータを用いたリスク解析と 東証障害の原因究明法に関する考察

長崎大学大学院生産科学研究科
木田 成美, 中村 剛

内容



- リスク解析の研究紹介
- 障害発生予防のための統計学的方法

リスク解析研究

- データ提供：株式会社SRA

ご協力いただいた方々

- 小島 勉 氏(株式会社SRA)
- 長谷川 享 氏(元 株式会社SRA、現 株式会社バイトルヒクマ)

- 研究目的・内容

ソフトウェア開発の早期段階で、プロジェクト失敗のリスクを予測する方法の提案。

受託ソフトウェア開発では、**開発コスト**が問題に

→そこで、「**利益**」に焦点を当て、プロジェクトの失敗を考えた。



本研究の仮説

要求分析／定義

設計

プログラミング

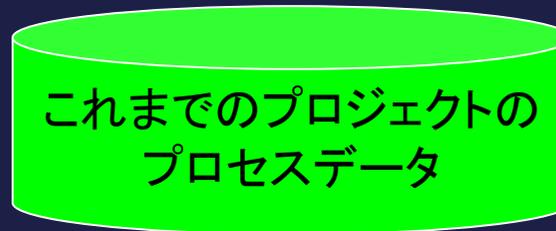
テスト

保守・運用

ソフトウェア開発の流れ

- ソフトウェア開発の早期段階で既に潜んでいる問題が、後の開発工程へ影響を及ぼし、結果としてプロジェクトの失敗を招く。

リスク解析のアウトライン



統計解析

プロセスデータとプロジェクトの
失敗との関係が数式で表される。

リスク推定モデル

- プロジェクト開始初期のプロセスデータから、そのプロジェクトが失敗するリスクを予測する統計モデル構築

統計モデルを構築できれば

- ソフトウェア開発の初期段階で、失敗のリスクを予測することができ、早期に対策を立て、修正することが可能。

失敗の予防！



方法 1-a 利益率について

$$\text{利益率} = \frac{\text{利益}}{\text{収入}}$$



- **最小利益率**

通常25~35%

- **予定利益率**

契約が成立したときに、プロジェクトマネージャーによって決定される

- **実利益率**

プロジェクト終了時に計算される

方法 1-b ソフトウェア失敗の定義

実利益率 < 予定利益率

かつ

実利益率 < 最小利益率



プロジェクト失敗！

- $Yore = 1$: 失敗
- $Yore = 0$: 成功

方法 2-a 統計学的手法

- モデル: ロジスティック回帰モデル
- サンプルサイズ: 58 (失敗例15を含む)
- 目的変数: ソフトウェアの失敗
(利益率に基づく)
- 説明変数: 失敗要因と考えられる変数12個

方法 2-b 要因と考えられる変数

- UTD ratio: プロジェクト開始1ヶ月のユニットテストの遅れ / 受注工数
- QAR ratio: プロジェクト開始1ヶ月の未解決のQAの数 / 受注工数
- CRR ratio: プロジェクト開始1ヶ月の顧客レビューリターン数 / 受注工数
- IRR ratio: プロジェクト開始1ヶ月の内部レビューリターン数 / 受注工数
- SPD ratio: プロジェクト開始1ヶ月の仕様書の遅れ数 / 受注工数
- BOC ratio: プロジェクト開始1ヶ月のバグの発生数 / 受注工数
- SPC ratio: プロジェクト開始1ヶ月の仕様書の変更数 / 受注工数
- SPA ratio: プロジェクト開始1ヶ月の仕様書の追加数 / 受注工数

方法 2-c 要因と考えられる変数(続)

- Gene Total:プロジェクト開始時にプロジェクトマネージャーによって評価される失敗のリスク
- TMD ratio: $CRR + SPC1 / \text{受注工数}$
- S ratio: $SPD1 + SPC1 + SPA1 / \text{受注工数}$
- R ratio: $QAR1 + CRR1 + IRR1 / \text{受注工数}$

方法 2-d ロジスティック回帰分析とは

- 疾病の発症、機械の故障・・・

ある事象が「起こる・起こらない」、「失敗・成功」

- ロジスティック回帰モデル

$$P(\text{失敗} \mid x_1, x_2, \dots, x_k) = \frac{\exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}{1 + \exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}$$

失敗の要因(集めたデータ)

要因 x_1, x_2, \dots, x_k があるときに失敗する確率

方法 2-e 重み付け

58のプロジェクトは、開発するソフトウェアの規模(工数)が小さいものから大きいものまでである

- 規模が大きいほど重要で、開発は複雑



- 重み: $\sqrt{\text{受注工数}}$

〈Pocock S. J. *et al.* (1981) for Environmental Risk〉

結果 1 ステップワイズ変数選択の結果

Parameter	Estimate	Chi Square	P-value
Intercept[1]	-4.606285	0	1.0000
TMD ratio	1.6480676	3.470365	0.0625
Gene Total	0.0399925	4.716253	0.0299
UTD ratio	0	0.004197	0.9483
QAR ratio	0	0.16731	0.6825
CRR ratio	0	0.359211	0.5489
IRR ratio	0	0.280249	0.5965
SPD ratio	0	0.862554	0.3530
BOC ratio	0	0.397161	0.5286
SPC ratio	0	0.359211	0.5489
SPA ratio	0	0.995466	0.3184
S ratio	0	1.079431	0.2988
R ratio	0	0.001737	0.9668

使用した解析ソフト: JMP

結果 2 失敗する確率の計算

- 回帰モデル 切片の推定値 Gene Total の係数推定値

$$P(\text{失敗} | x_1, x_2) = \frac{\exp(\alpha + \beta_1 x_1 + \beta_2 x_2)}{1 + \exp(\alpha + \beta_1 x_1 + \beta_2 x_2)}$$

- TMD ratio と Gene Total TMD ratio の係数推定値



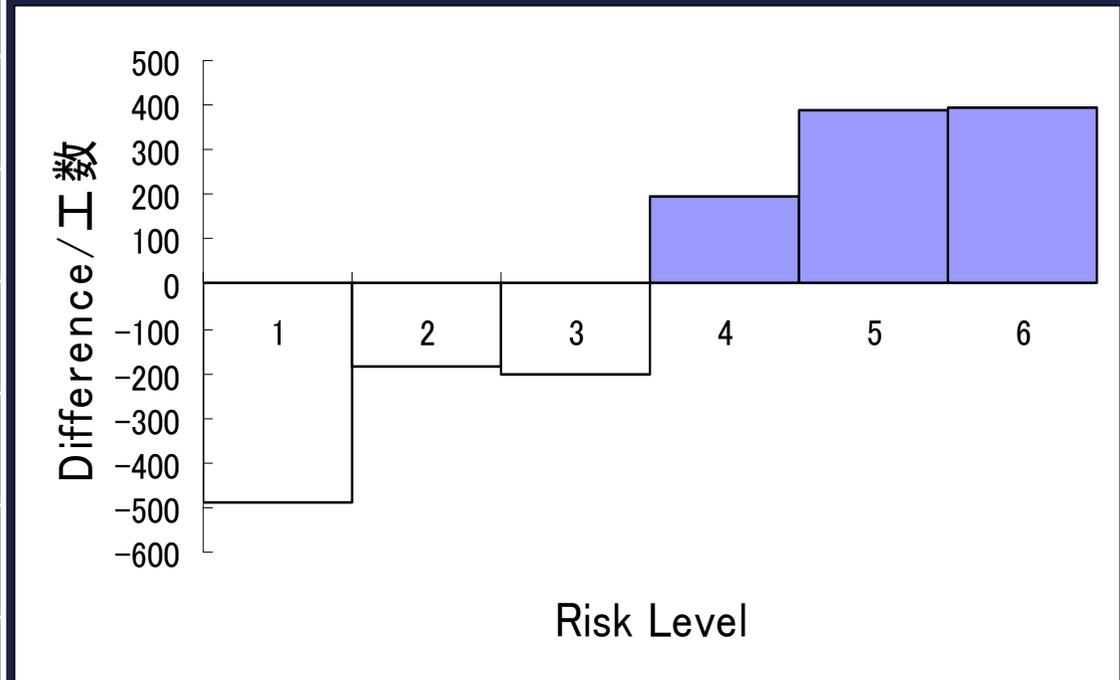
- あるプロジェクトにおいて、TMD ratio, Gene Total の値がそれぞれ a_1, b_2 のとき、プロジェクトが失敗する確率

$$P(\text{失敗} | a_1, b_2) = \frac{\exp(-4.61 + 1.65 \cdot \underline{a_1} + 0.04 \cdot \underline{b_2})}{1 + \exp(-4.61 + 1.65 \cdot \underline{a_1} + 0.04 \cdot \underline{b_2})}$$

そのプロジェクトのTMD ratio と Gene Totalのデータ

結果 3 失敗する確率の推定値と 利益・損失の関係

Risk Level	確率の 平均値	Difference/ 工数
1	0.65	-486
2	0.32	-186
3	0.20	-200
4	0.12	193
5	0.08	390
6	0.03	393



※ Difference = 実利益 - 予定利益

結論 1

- **TMD ratio** と **Gene Total** が説明変数に選ばれた

【TMD ratioについて】

TMD ratio =

$$\frac{\text{CRR1 (プロジェクト開始1ヶ月間の顧客レビューリターン数)} + \text{SPC1 (プロジェクト開始1ヶ月間の仕様書の変更数)}}{\text{受注工数}}$$

顧客と強い関係がある

- 高いTMD ratio値をもつ失敗プロジェクトの損失の原因は、幾分か顧客にあるかもしれない

結論 2

【Gene Totalについて】

- **Gene Total** はプロジェクト開始時にプロジェクトマネージャによって評価された、そのプロジェクトのもつリスクの合計



- **経験的な勘**は、**失敗のリスクを発見するために有効**である
- 計測できない要因も、**主観で構わないので数値的に記録しておく**と良い

まとめ 1

- ソフトウェア開発におけるリスク解析の一手法として提案する
- モデル式に新たなプロジェクトのプロセスデータを代入することにより、そのプロジェクトの失敗リスクを予測できる
- リスク推定モデルに組み込まれている変数（＝失敗の要因と考えられるプロセス）のコントロールにより、失敗のリスクを低減させる

まとめ 2

- 他にも「成功・失敗」「良・不良」などを定義できるならば、同様の手法を用いてリスク解析を行うことが可能
- 説明変数にはコントロールできるものを入れておくと、リスク制御に活かせる