

潜在特性モデルによるプロジェクトリスク予測技術

Predicting project risks using latent trait model

井手 剛

1 はじめに

自動車開発、工場の立ち上げ、橋梁の建築、など、大規模開発業務ではプロジェクトマネジメントの成否が収益性を左右する。日本では、労働者の高い教育水準と、言語的な多様性の低さに由来する質の高いコミュニケーション能力を背景にして、プロジェクトマネジメント上の諸問題を人力のすり合わせで解決する傾向が強いように思われる。しかし最近のオフショア開発の増大などの環境変化にともない、労働者の自発的な改善・すり合わせのみに頼ることなく、客観的なデータに基づいてプロジェクトの状態を把握し、将来のリスクを予測することが望まれる。とりわけ IT (Information Technologies) 業界においては、通常システム構築の技術ではコモディティ化が進んでおり、人間の勘と経験に頼らぬ系統的なアプローチにより「筋の悪い」プロジェクトを早期に察知し、適切なアクションを取ることが市場競争力の源になっている。本稿では、大規模 IT システム開発におけるプロジェクトリスク予測の最近の試みを紹介する。

いでつよし, IBM T. J. Watson Research Center.
tide@us.ibm.com.

2 問題設定: IT システム構築プロジェクトの品質管理

大規模 IT システムの構築は、数ヶ月にわたる計画期間を経て、契約締結後に数ヶ月かけて徐々に安定稼動の状態に持ってゆく。図 1 に典型的なプロジェクト計画 (project design) のプロセスを要約する。営業部門とも開発部門とも独立な QA (quality assurance) と呼ばれる社内の第三者機関が、計画時において定期的に計画の監査を行い、プロジェクトマネージャーに改善の指示を出す。これは反復的なプロセスであり、そのプロセスが適切に行われている限りおおむねプロジェクトのリスク管理はうまくできるのだが、問題は、レビュープロセスが意図したとおりに動かない場合がままあるということである。これはレビューを有名無実化して、プロジェクトの計画が「うまくいっていることにする」という力が常に働くためである。

品質保証のプロセスにおいては多様なデータが取得されるが、現在のところ最も信頼性が高いと考えられているのは QA レビューにおける質問表データである [5]。レビューの属人性をできるだけ取り除くために、質問表は、たとえば

- 顧客との良好な関係が築けているか

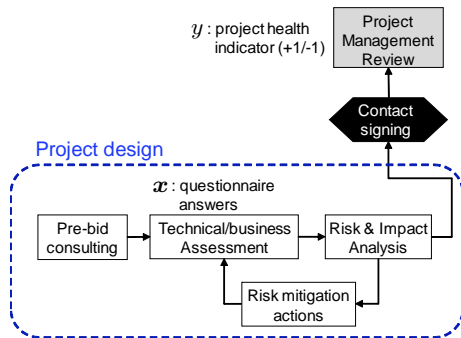


図 1 情報システム構築プロジェクトのリスク管理プロセスの例。

- 計画されているソリューションに経験はあるか
- コストの見積もりは完全と言えるか
- スケジュールには無理がないか

というようなプロジェクトの状態についての項目があり、それぞれに対して QA の専門家は自分の経験に基づき、リスクレベルを数値で付してゆく。問題を簡単にするために、契約締結直前の最後のレビューに注目し、質問表の入力はリスクあり (1) かなし (0) かの 2 値とする。採点結果 x とするとこれは多次元の 2 値ベクトルとなる。

形式的にいえば、IT プロジェクトの品質保証プロセスとは、計画時の特徴量としての x が与えられたときに、契約締結後、システム構築フェーズにおけるプロジェクトの成功不成功をあらわす指標 y を予測ないし制御するという問題である。 y としては、構築フェーズにおける財務指標を指数化したものをとることができる。たとえば予算超過の危険に応じて A、B、C、D のような指標を考慮することができるが、問題を単純化するために、問題あり ($y = +1$) か良好 ($y = -1$) かという 2 値を考慮ことにする。

3 プロジェクトの潜在リスクモデル

素朴に考えるとこれは数値ベクトル x を入力として y を予測する 2 値分類問題のように見える。実際、これまではそういう扱いがなされてきた [4, 5]。しかしデータを詳しく見るとそのような単純な定式

化では (作為的なデータを使わない限り) 実用的な精度で予測を行うのが難しいことがわかっている。

5 段階評価でリスクを判定する、のような質問表の場合、契約前の最終段階のレビューにおける質問表のデータは、非常に強く低リスク側に偏っており、たとえば正規分布に従う実数データとはまったく振る舞いが異なる。その上、レビューを行うのは IT システム構築の専門家であり、ランダムな誤差を考えるようなモデルはそもそも自然とは言えない。さらに重要なことは、どのような予測を行うにせよ、予測結果は非常に高いレベルでの説明可能性が必要であるということである。

我々はこの「説明可能性」について次の 3 つの基準を定義した [1]。

- ある質問における選択肢同士の相対比較可能性
- 質問表の異なる質問同士の相対比較可能性
- プロジェクト同士の比較可能性

まず最初の点を考える。先に述べたとおり、プロジェクトの計画段階の最終レビューにおいてはほとんどの項目はリスクなしとなっており、たとえばリスクあり項目の総計を求めても予測の役に立たない [2]。この設定では、第 1 の説明可能性とは、ある質問 i について $x_i = 0$ と 1 の選択についての確率モデルを明示的に立てるということである。 $x_i = 1$ が選択される確率として、我々は計量心理学に由来する次のモデルを考える (図 2)。

$$P(\theta, a_i, b_i, c_i) \equiv c_i + \frac{1 - c_i}{1 + e^{-a_i(\theta - b_i)}} \quad (1)$$

ここで、 θ は未知の潜在変数で、各プロジェクトの真のリスクを表す。重要な仮定は、真のリスクと $x_i = 1$ が選択される確率は比例しないということである。図 2 を直感的に解釈すれば、真のリスクが低いときは過度に楽観的、真のリスクがある値を超えると過度に悲観的、さらに、時折確信なくあてずっぽうに答えることもあるということである。

このモデルを質問数 M だけ積み重ねて x の確

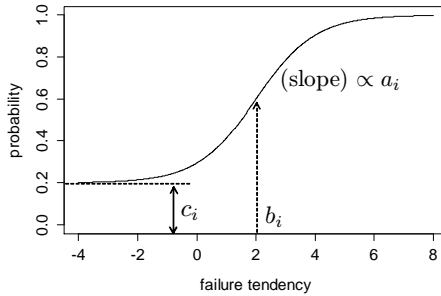


図 2 3 パラメータ項目応答曲線。

率モデルを得る。

$$p(\mathbf{x}|\theta, \mathbf{a}, \mathbf{b}, \mathbf{c}) = \prod_{i=1}^M P(\theta, a_i, b_i, c_i)^{\delta(x_i,1)} [1 - P(\theta, a_i, b_i, c_i)]^{\delta(x_i,0)} \quad (2)$$

ただし M は質問表における質問数、 $\delta(x_i, 1)$ はクロネッカーのデルタで、 $x_i = 1$ のときに 1、それ以外では 0 となる。 $\delta(x_i, 0)$ も同様に定義される。

計量心理学の分野では、このモデルを教師なし学習の枠組みで扱うが、我々は y の情報を取り込むために、潜在リスク変数 θ に対して次の事前分布を設定する。

$$f(\theta|y) = \begin{cases} \frac{\gamma}{\sqrt{2\pi}} \exp(-\frac{\gamma}{2}\theta^2) & \text{for } y = -1, \\ \frac{\gamma}{\sqrt{2\pi}} \exp(-\frac{\gamma}{2}(\theta - \omega)^2) & \text{for } y = +1, \end{cases}$$

以上のモデルから、 N プロジェクトに対する $(\mathbf{x}^{(n)}, y^{(n)} | n = 1, \dots, N)$ のデータをもとに、次のような周辺化対数尤度を考えることができる。

$$L(\mathbf{a}, \mathbf{b}, \mathbf{c} | \mathcal{D}) = \sum_{n=1}^N \ln [\pi(y^{(n)}) p(\mathbf{x}^{(n)} | \mathbf{a}, \mathbf{b}, \mathbf{c}, y^{(n)})]$$

$$p(\mathbf{x}^{(n)} | \mathbf{a}, \mathbf{b}, \mathbf{c}, y^{(n)}) \equiv \int_{-\infty}^{\infty} d\theta^{(n)} p(\mathbf{x}^{(n)} | \theta^{(n)}, \mathbf{a}, \mathbf{b}, \mathbf{c}) f(\theta^{(n)} | y^{(n)})$$

ただし $\pi(y)$ は $y = \pm 1$ に対応する事前分布であり、通常、 ± 1 のクラスの標本数（プロジェクト数）をそれぞれ全標本数で割ったものとしてよい。これを最大化することで、各質問を特徴付けるパラメータ $\{a_i, b_i, c_i\}$ が求められる。

このモデルは共役事前分布を使わないベイズ推論のひとつの例になっている。指数型分布族の共役分布とは異なり、期待値-最大化法は簡易な反復公式を与えない。しかし、幸い、 $f(\theta|y)$ のような正規分布を重み因子とする周辺化積分には、Gauss-Hermite 求積法という巧妙な手法が直接適用でき、周辺化尤度およびその勾配の数値計算は実は容易である。計算の詳細は論文 [1] を参照されたい。

4 プロジェクトのリスク予測モデルの構築

以上のモデルは本質的には \mathbf{x} と y の同時分布を与えるので、原理的には y の予測分布を導くことができる。しかし説明可能性についての実用上の強い要請、とりわけ第 3 のサンプル間比較の容易さを考慮して、上記のモデルを計量学習の枠組みに統合し、 k 近傍法により予測を行うものとする。

今のモデルでは、各質問項目は、潜在リスク度 θ が与えられたときに条件付き独立であるので、質問表の回答 \mathbf{x}, \mathbf{x}' との 2 乗距離を

$$d^2(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^M A_{i,i} (x_i - x'_i)^2 \quad (3)$$

のように、対角型の計量テンソル \mathbf{A} を使い自然に表現することができる。詳細は省くが、確率モデルの生成するエントロピーについての等式から、この計量テンソルを先に示した確率モデルから求めることが可能である [1]。計量テンソルの対角成分は定義から正であり、これは各質問項目の有用度 (informativeness) を表す便利な指標となる。

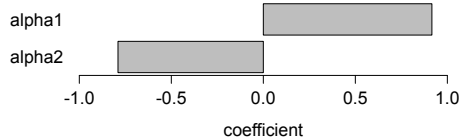
距離尺度が計算できれば、 k 近傍法を使い $y = \pm 1$ の予測結果と、「似たプロジェクトとして過去にどのようなものがあったか」という疑問に答えることが可能となる。ひとつ気をつけることは、 $y = \pm 1$ が著しく不均衡であるということである。この点についての一般的な注意は、最近出版した教科書 [3] を参照されたい。

5 実験結果

以上述べた「説明可能性」についての問題意識を説明するために、 $M = 2$ 個の質問のみからなる非

表 1 2次元の人工データ。

\mathbf{x}	$(y = +1)$	$(y = -1)$
(0,0)	8	9
(0,1)	6	16
(1,0)	20	20
(1,1)	16	16

図 3 L_1 正則化ロジスティック回帰の係数。

常に素朴な分類問題を考えてみよう。表 1 にデータ、図 3 に比較のため L_1 正則化ロジスティック回帰で学習した分類モデルの係数、図 4 に本モデルで学習した項目応答曲線と（実線が $x_i = 1$ の確率を表し、破線は 1 からその確率を引いたものである）、計量テンソルの対角成分 A_{11}, A_{22} を示す。

解釈可能性の観点でのロジスティック回帰の問題は、係数が負になりえるということである。そのため、「どの質問がどれだけ有用か」というような質問に答えるのは簡単ではない。我々のモデルでは、有用度において第 1 番目の質問が明確に高いことがわかる。このことは、項目応答曲線が、閾値関数としてきれいな形になっていることから確認できる。一方、第 2 番目の質問の項目応答曲線の傾きはなだらかで、閾値を定めにくい。したがって分類においても有用度は低いであろうことが読み取れる。勾配が負になっていることは、yes と答えたときに $y = -1$ となる可能性が高いということであり、これはロジスティック回帰の係数が負であることと一貫している。

以上は簡単は人工データの結果であるが、我々の論文 [1] ではまた、ある IT ベンダーの実データを使った評価結果も述べられている。1000 程度のプロジェクトについて数十項目の質問を含む質問表

を使い、一つ抜き交差確認で精度を計算した結果、F 値にして 7 割前後の予測精度を出せることがわかった。興味深いことは、説明可能性が低い通常の

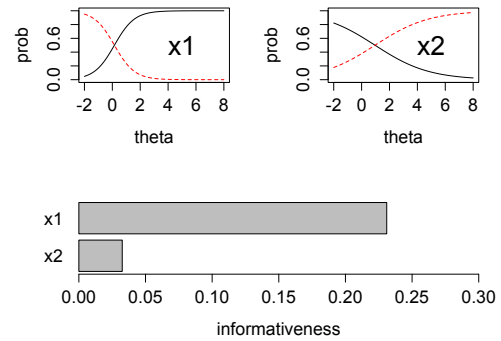


図 4 項目応答曲線と係数の有用度。

分類器よりもむしろ高い分類精度を与えていることである。これは、人間の評価のバイアスを計量心理学的に明示的にモデリングすることの意義を示している。

参考文献

- [1] Tsuyoshi Idé and Amit Dhurandhar. Informative prediction based on ordinal questionnaire data. In *Proc. 2015 IEEE Conference on Data Mining, ICDM 15*, 2015.
- [2] Tsuyoshi Idé, Sinem Güven, Ee-Ea Jan, Sergey Makogon, and Alejandro Venegas. Latent trait analysis for risk management of complex information technology projects. In *Proceedings of the 14th IFIP/IEEE International Symposium on Integrated Network Management, IM 15*, pages 305–312, 2015.
- [3] Tsuyoshi Idé and Masashi Sugiyama. *Anomaly detection and change detection*. Kodansha, 2015. (in Japanese).
- [4] Aleksandra Mojsilović, Bonnie Ray, Richard Lawrence, and Samer Takriti. A logistic regression framework for information technology outsourcing lifecycle management. *Computers & Operations Research*, 34(12):3609–3627, December 2007.
- [5] K. Ratakonda, R. Williams, J. Bisceglia, R.W. Taylor, and J. Graham. Identifying trouble patterns in complex it services engagements. *IBM Journal of Research and Development*, 54(2):5:1–5:9, March 2010.