

プロジェクト失敗リスク予測モデル

Project Failure Risk Prediction Model for Software Development Projects

森 俊樹 覚井 真吾 田村 朱麗

■ MORI Toshiki ■ KAKUI Shingo ■ TAMURA Shurei

ソフトウェア開発プロジェクトの大規模化と複雑化に伴い、定量的なプロジェクト管理の重要性が増している。しかし、一般的にデータのばらつきが大きく、欠損値も多く含まれるソフトウェア開発データに対しては、重回帰分析やレイリーモデルなど従来の技術の有効性は限られていた。

東芝は、プロジェクトが目標未達となる確率を定量的に管理するため、ナイーベイズ識別器に基づくプロジェクト失敗リスク予測モデルを開発した。この技術は、データのばらつきや欠損に強く、プロジェクトの途中段階で得られたデータに基づいて、逐次的な予測の更新が可能である。このモデルを実際の社会インフラシステムのソフトウェア開発プロジェクトに試行適用し、リスクの早期検知と対策立案に役だつことを確認した。

With the expanding scale and complexity of software development projects, quantitative project management has become increasingly important. However, conventional approaches based on a statistical model including multiple regression analysis and the Rayleigh model generally have limited applicability to actual software development data containing low-accuracy data, outliers, and missing values.

Toshiba has been constructing a new approach using a project failure risk prediction model based on the naive Bayes classifier technique in order to manage the probability of nonachievement of objectives in a project. Trial application of this approach to actual software development projects in the social infrastructure field has confirmed that it achieves successful performance without being affected by low-accuracy data and missing values in software development data, and can sequentially renew predictions in accordance with data collected in each project phase.

1 まえがき

ソフトウェアには、目に見えない（不可視性）、容易に変更できる（変更容易性）などの特徴があり、その開発管理は従来から課題とされてきた。更に、近年、ソフトウェア開発の大規模化と複雑化が進んでおり、国内や海外の関連会社などへの業務委託が一般的になるとともに、ひとりのプロジェクトリーダーが複数のプロジェクトを兼務するというような状況も生じている。このような環境の変化に伴い、プロジェクトの状況を正確に把握できず、リスクへの対応が後手に回るなど、プロジェクトリーダーの勘と経験に頼ったプロジェクト管理は限界に近づいている。こうした変化に対応していくためには、定量的なデータに基づく客観的なプロジェクトの状況把握、及び将来のリスク予測に基づく先回りの管理が重要になる。

東芝は、ナイーベイズ識別器に基づいて、プロジェクトが目標未達になる確率を定量的に予測するモデルを開発した。ここでは、その仕組みと構築手順、及び適用事例について述べる。

2 プロジェクト管理における定量データ活用の課題

従来、プロジェクト管理における定量データ活用の主な技術としては、メトリクスの可視化によるプロジェクトの状況把握

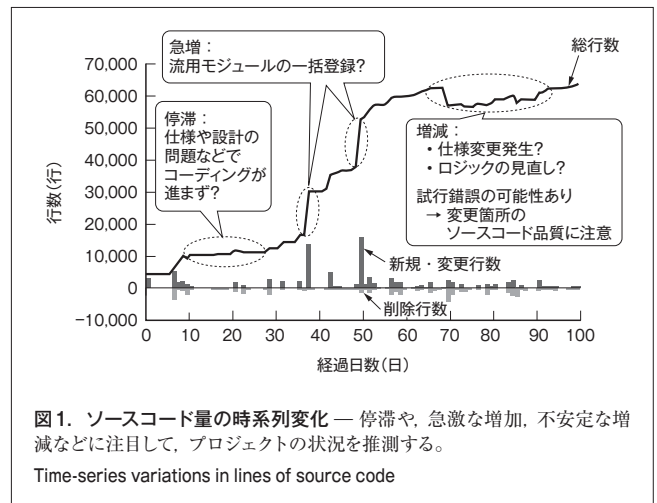


図1. ソースコード量の時系列変化 — 停滞や、急激な増加、不安定な増減などに注目して、プロジェクトの状況を推測する。

Time-series variations in lines of source code

や、重回帰分析などによる統計モデルの構築とそれに基づく将来の状況予測があり、それぞれが様々な開発部門で適用され、一定の効果を上げている⁽¹⁾⁻⁽³⁾。

マトリクスの可視化によるプロジェクトの状況把握は、定量データ活用の基礎的な技術の一つである。まず最初に、開発部門におけるソフトウェア開発の課題に沿ったメトリクスの定義と収集を行い、次に、収集したメトリクスを様々な種類のグラフ（折れ線グラフや、棒グラフ、散布図など）で可視化して、

それらのグラフを定常的に監視することで、プロジェクトの状況把握や異常検知につなげる(図1)。

マトリクスの可視化によるプロジェクトの状況把握は、直感的でわかりやすく、ソフトウェア開発の進捗把握など多くの場面において有効であるが、いくつかの課題も残る。まず、各グラフが表している情報は、プロジェクト全体のごく一部を表す断片的な情報であり、プロジェクト全体の監視を行うには、定常的に多数のグラフを生成して、求める情報をそこから読み取る必要がある。グラフの解釈や適切な対策の実施は、プロジェクトリーダーの能力や経験に大きく依存しており、一般化は難しい。

もう一方の定量データ活用の技術として、重回帰分析や、レイリーモデル、モンテカルロシミュレーションなどによる統計モデルの構築と、それに基づく将来の状況予測がある。これらは、開発部門の特徴を反映した精度の高いモデルが構築できれば有効であるが、実際に適用してみると、実用的なモデルの構築が困難である場合も多い。その理由として、重回帰分析やレイリーモデルは外れ値の影響を受けやすく、また、一部に欠損値が含まれているとサンプル数が大幅に減少するなどの問題がある。更に、モデルの構造自体はシンプルであるが、モデルが満たすべき制約条件が強く、正しいモデルを構築するには十分な注意が必要である。

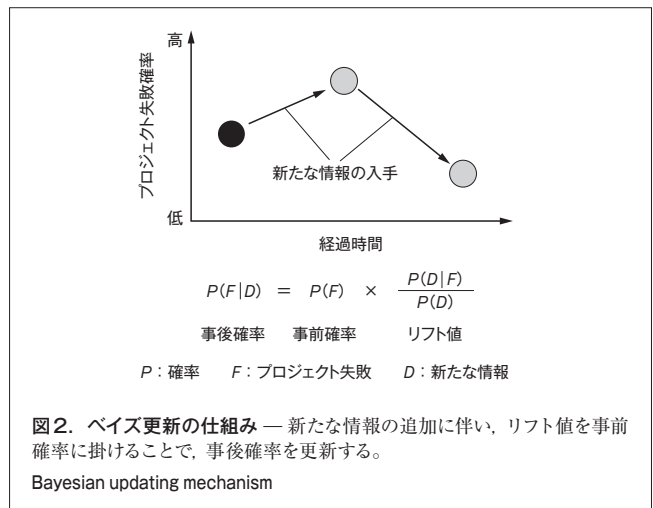
3 プロジェクト失敗リスクの予測

当社は、2章で述べた課題に対応した新しい定量的管理のアプローチとして、ベイズ統計の応用の一種であるナীবベイズ識別器に基づく予測モデルを開発した^{(4), (5)}。これは、2章で述べた二つのアプローチ(可視化による状況把握と統計的予測モデルの構築)の利点を兼ね備えた技術であり、重回帰分析などによる予測モデルが、例えば工数や欠陥数などマトリクスの値そのものを予測しようとしているのに対して、この技術は、マトリクスがあるしきい値を上回る(又は、下回る)確率を予測する。

3.1 予測の仕組み

プロジェクト失敗リスクとは、“組織があらかじめ設定した目標に対して、対象プロジェクトが目標未達となる確率”である。組織目標は、例えば“工数が計画値以下”や“総合テスト以降の不具合が0件”など、品質、コスト、及び納期のマトリクスで定義されるのが一般的であり、ここでの失敗プロジェクトは、組織目標に対して改善余地のあるプロジェクトと捉えることができる。

ベイズ統計では、ある事象の確率の初期値として、今まで得られた情報から推定された事前確率が与えられ、ある時点で新たな情報が得られると、確率は得られた情報の価値に応じて更新される。このように、新たな情報の追加に伴って、次々に確率の更新を行う仕組みがベイズ更新であり、その際、新たな情報の影響度として事前確率に掛けられる数値はリフ



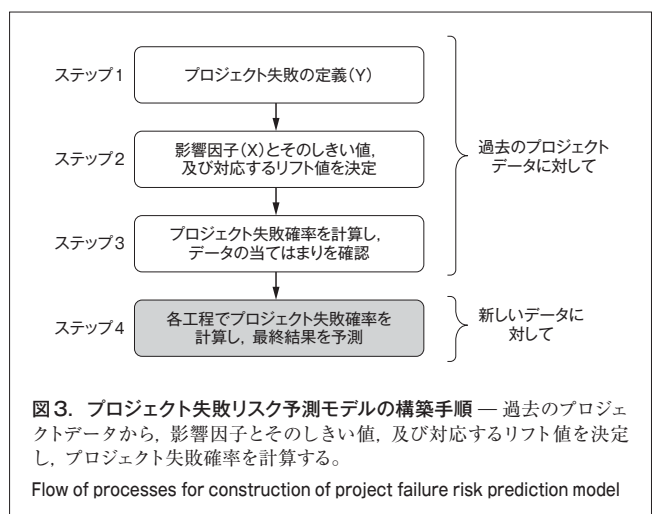
ト値と呼ばれる(図2)。

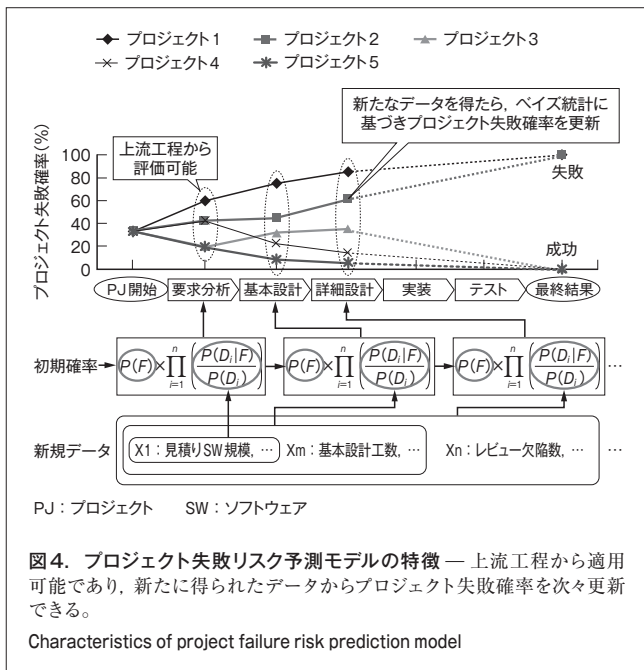
ベイズ更新の仕組みは、曖昧な状況下で人間が意思決定する際の思考プロセスと非常に近く、2章で述べた可視化による状況把握においても、熟練したプロジェクトリーダーは同様の思考や判断を行っていると思定される。この技術で採用したナীবベイズ識別器は、迷惑メールフィルタなど様々な分野で応用されており、ベイズ更新の仕組みの上に更に各事象の独立性を仮定することで、計算量を大幅に低減した。

3.2 予測モデル構築手順

プロジェクト失敗リスク予測モデルの構築手順を図3に示し、各ステップにおける作業内容を以下に述べる。

まずステップ1で、プロジェクト失敗の定義を明確化する。続くステップ2で、各種検定などの統計的手法を用いて、過去のプロジェクトデータから、プロジェクト失敗に影響の大きいマトリクス(以下、影響因子と呼ぶ)とそのしきい値、及び対応するリフト値を求める。それらが決まれば、予測モデルの構成要素が確定し、次のステップ3で、ベイズの定理から、各マトリクスの独立性を仮定したプロジェクト失敗確率の近似値(以下、





リスクスコアと呼ぶ)が計算できる。リスクスコア自体は厳密には確率の定義に当てはまらないため、更に、リスクスコアの対数のロジスティック回帰分析を実施することで、0から1までの確率値に正規化されたプロジェクト失敗リスクが得られる。最後のステップ4では、新たに得られたデータから、ベイズ更新の仕組みに基づいてプロジェクト失敗確率を計算し、プロジェクトの最終結果を予測する。

プロジェクト失敗リスク予測モデルは、上流工程から適用可能であり、更に、新たなデータを得たら、それに応じて予測モデルの更新が可能という特徴があり、異常プロジェクトの早期検知や監視強化などに活用できる(図4)。

4 適用事例

プロジェクト失敗リスク予測モデルの適用事例として、社会インフラシステムのソフトウェア開発プロジェクトに試行適用した2件の事例について述べる。

4.1 適用事例1

3章で述べたように、プロジェクト失敗リスク予測モデルを適用する際、まず組織の目標を明確化する必要がある。適用事例1の組織では、次のいずれかの状態であった場合、ソフトウェア開発プロジェクトが目標未達であると定義した。

- (1) 開発に要した工数が計画値を超過し、更に、総合試験における欠陥の数が基準値を超過
 - (2) 開発に要した工数が計画値を超過し、更に、総合試験完了時の節目レビューの実施日が予定日から遅延
- 次に、このように定義した目標未達の条件に対して影響因子を探した。適用事例1の組織では過去のプロジェクトデータが

29件蓄積されており、その中から各開発工程で品質、コスト、及び納期に関わるメトリクスを影響因子の候補として選び出した。この際、基本測定量(単一の属性の測定量)だけではなく、複数の測定量から計算される導出測定量も含めて、100以上のメトリクスを候補として採用した。その後、過去のプロジェクトを目標達成だった群と未達成だった群に分け、2サンプルt検定を実施して影響因子を絞り込んだ。

検定の結果、p値が0.05以下となったデータは統計上明らかに有意差があるので、影響因子として採用した。また、p値が0.05より大きくなったデータについては、統計上有意差があるとは断定できないが、改善の対象となりうるデータについては、p値が0.1以下程度であれば、今回は影響因子として採用することにした。最終的に次の6種類に影響因子であると結論づけた。

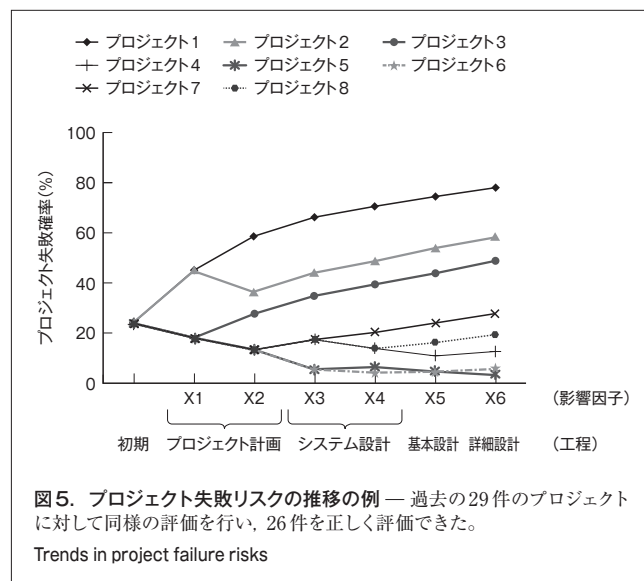
- X1: 新規ステップ数と改造ステップ数の合計値
- X2: ステップ生産性
- X3: システム設計工程におけるレビュー工数密度
- X4: システム設計工程におけるレビュー欠陥密度
- X5: 基本設計工程におけるレビュー工数密度
- X6: 詳細設計工程におけるレビュー工数密度

影響因子とそのしきい値が決定されれば、過去のプロジェクトデータから各工程におけるプロジェクト失敗リスクを計算でき、プロジェクトの途中段階で得られたデータに基づいて、逐次的に予測を更新することが可能になる(図5)。

プロジェクト失敗確率の予測値が50%以上を失敗、50%未満を成功として、過去の29件のプロジェクトに対し詳細設計完了時点での予測精度を評価したところ、26件を正しく予測できた。

4.2 適用事例2

二つ目の適用事例として、一つ目とは別の社会インフラシス



テムのソフトウェア開発部門において、プロジェクト失敗リスク予測モデルを構築し試行した事例を述べる。適用事例2の組織では、組織目標を“システムテスト以降の不具合件数が基準値以下”と定義した。

この目標に対して、まず、影響因子の候補を網羅的に洗い出し、各候補に対して、影響度や発生頻度などで優先順位付けを行い、詳細分析するデータを絞り込んでからカイ2乗検定を実施し、統計的な有意性を確認した。最終的には、次の8種類を影響因子とした。

- X1：要求仕様の確定遅れ
- X2：機能仕様の難易度
- X3：ソースコード規模
- X4：担当者の経験
- X5：各工程でのスケジュール遅れ
- X6：各工程でのコスト超過
- X7：設計レビュー速度
- X8：コードレビュー速度

過去の100件のプロジェクトデータを用いて各工程におけるプロジェクト失敗リスクを計算し、プロジェクト失敗確率の予測値が50%以上を失敗、50%未満を成功として、コーディング完了時点での予測精度を評価したところ、86件を正しく予測できた。また、新規プロジェクトに対する予測性を評価するために、モデル構築時には含めなかった同システム群のプロジェクトデータに対して同じ予測モデルを適用したところ、予測正答率は約80%であった。

更に、進行中のプロジェクトに適用した。そこでは、定期的にプロジェクト失敗リスクを確認し、リスクが高い場合には、後工程のレビュー強化など、重点的な対策を実施した場合と実施しなかった場合のリスクの変化を確認して、対策要否の検討につなげた(図6)。

これらの試行適用を通して、開発したプロジェクト失敗リスク予測モデルが、リスクの早期検知と対策立案に役だつことを確認した。

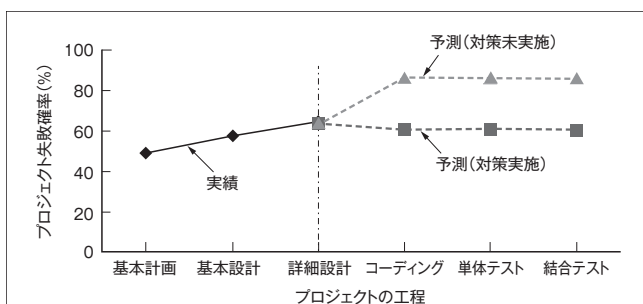


図6. 対策の実施がプロジェクト失敗リスクに及ぼす影響 — 進行中のプロジェクトにおいて、プロジェクト失敗リスクの予測に基づき、対策の要否を検討した。

Effects of control policies on project failure risks

5 あとがき

ナイーブベイズ識別器に基づくプロジェクト失敗リスク予測モデルを開発し、実プロジェクトデータによりその予測精度を評価した。その結果、この技術はデータのばらつきや欠損に強く、利用できるデータの数や種類に応じた実用的な予測モデルの構築が可能であることを確認した。

この技術は、プロジェクト状況の定量的モニタリングや、異常検知の仕組み作りなどに活用することが期待される。精度の高い予測を実現するには、ある程度のプロジェクトのサンプル数が必要であり、サンプル数が極端に少ない場合には、一つのプロジェクトを複数のサブプロジェクトに分割したり、機能モジュールごとの比較を行うなどの工夫が必要になる。

今後の課題としては、予測精度の向上やプロジェクト実行中のコントロールが挙げられる。予測精度の向上には、予測モデル自体を改良するアプローチと、入力データを加工(変換や補完など)するアプローチがあり、どちらも、まだおおいに改善の余地がある。プロジェクト実行中のコントロールについては、予測の背後に隠れた本質的なプロジェクト制御のメカニズムを明らかにすることが必要で、困難ではあるが非常に重要な課題であり、取組みを進める。

文献

- (1) 小室 睦 他. 開発現場の実態に基づいたピアレビュー手法改善と改善効果の定量的分析. SEC journal. 1, 4, 2005, p.6-15.
- (2) 竹本昇司. 実践型EVMを活用したプロジェクト管理の適用研究. SEC journal. 1, 4, 2005, p.16-23.
- (3) 相磯正司 他. ソフトウェア開発における統計的プロジェクト管理手法の導入と実践. 富士フイルム研究報告. 56, 2011, p.31-34.
- (4) Mori, T. et al. "Incremental Estimation of Project Failure Risk with Naïve Bayes Classifier". Proceedings of the ACM-IEEE International Symposium on Empirical Software Engineering and Measurement. Baltimore, MD, USA, 2013-10, ACM/IEEE. p.283-286.
- (5) 森 俊樹 他. プロジェクト失敗リスク予測モデルの構築. プロジェクトマネジメント学会誌. 15, 4, 2013, p.3-8.



森 俊樹 MORI Toshiki

ソフトウェア技術センター プロセス・品質技術開発担当参事。定量的プロジェクト管理技術の開発に従事。IEEE会員。Corporate Software Engineering Center



覚井 真吾 KAKUI Shingo

ソフトウェア技術センター プロセス・品質技術開発担当参事。定量的プロジェクト管理技術の開発に従事。情報処理学会会員。Corporate Software Engineering Center



田村 朱麗 TAMURA Shurei

ソフトウェア技術センター プロセス・品質技術開発担当主務。定量的プロジェクト管理技術の開発に従事。情報処理学会会員。Corporate Software Engineering Center