

定量的プロジェクト管理におけるソフトウェア品質予測モデルの構築と適用

Construction of a Prediction Model for Software Quality and Its Application in Quantitative Project Management

岡野 麻子* 岡田 政之** 土屋 義兼*

Asako Okano, Masayuki Okada, Yoshikane Tsuchiya

鎌倉事業部では、ソフトウェア開発における定量的プロジェクト管理の一環として、品質予測モデルの構築を行っている。本研究ではソフトウェア開発における誤り累積数の予測モデルとしてゴンベルツ曲線を採用した。その場合、パラメータを決定するためには最低4つのフェーズの実績データが必要である。そのため、4フェーズに満たない初期段階ではゴンベルツ曲線が決定できない。それを補うため、終了していないフェーズのデータは計画値を用いて外挿し、その外挿した値と実績値を用いて予測モデルを構築する。本稿では、構築した品質予測モデルを定量的プロジェクト管理の中で使用した事例について述べる。

We have been trying to construct models to predict the software quality, as part of the quantitative project management in our division. We adopt a gompertzian curve as a prediction model for the cumulative number of bugs in the software development process. Four phases are required to determine the parameters of the gompertzian growth curve model; the curve cannot be determined at the early process less than four phases. The planned values, based on the results of former projects, are used to extrapolate the unfinished phase data. The prediction model is constructed using those finished and extrapolated data. This report mentions the construction of the model to predict software quality and its application to the quantitative project management of a pilot project.

1. まえがき

鎌倉事業部では、昨年度から引き続きCMMI® (Capability Maturity Model Integration: 能力成熟度モデル統合、米国Carnegie Mellon大学の登録商標⁽¹⁾)のレベル4と5を指標とした改善を進めている。成熟度レベル4以上は、統計的手法を用いて予測を行う定量的プロジェクト管理が求められる。つまり、プロセス実績ベースライン (Process Performance Baseline, PPB)⁽¹⁾を導出し、プロセス実績モデル (Process Performance Model, PPM)⁽¹⁾と呼ばれる予測モデルを構築して利用する予測管理・制御が必要になる。このモデルを利用することにより、下記の利点が得られるようになる。

- ・品質や生産性の向上
- ・プロジェクト管理制御の精度向上

昨年度に述べた「生産性と品質データの解析手法についての提案」⁽²⁾での試みに引き続き、この定量的プロジェクト管理を実施すべく、予測モデルを構築した。本稿では、改善活動における品質予測モデル構築への取り組み

み、品質予測モデルを利用した定量的プロジェクト管理のパイロットプロジェクトへの適用、および適用結果と今後の課題を述べる。

2. 品質予測モデル

品質予測モデルを構築するためには、下記項目についての検討が必要である。

- ①プロジェクトゴールの優先順位を決める。
- ②①の結果に基づいて予測対象を選択する。
- ③②を予測するための管理尺度を決める。
- ④過去実績を使用して③のPPB、PPMを作成する。

今回のパイロットプロジェクトでは、プロジェクトゴールに組織目標であるロスコスト抑制、品質向上等へつながら具体的なQCD (Quality, Cost, Delivery)を設定した。その中でも、出荷後に誤りを出さないことを最重要ゴールと考え、その管理尺度を出荷時残存誤り密度とした。この管理尺度を達成するための品質予測モデルを構築した。

品質予測モデル構築に必要なPPBの導出を2.1節、PPMの概要を2.2節、品質予測モデルの構築手順を2.3節に述べる。

2.1 プロセス実績ベースライン

CMMIのレベル4では、PPBを基に各工程内で目標値を定めることが求められる。そこで、対象組織/プロジェクトの過去実績を基に、データの中心傾向と幅を算出したもの⁽¹⁾から、箱ひげ図を利用して各フェーズの累積誤り検出率の目標値を導出した(図1参照)。計画値を第2四分位数⁽³⁾とし、許容範囲を累積誤り検出率の第1四分位数と第3四分位数の範囲とした。この値を用いて各フェーズの品質状況を定量的に把握した。

2.2 プロセス実績モデル

CMMIのレベル4では、PPMを利用し開発途中で測定可能な数値から将来の実績値を予測し制御することが求められる。そこで、2.1節で導出した各フェーズにおける累積誤り検出率の計画値を利用し、累積誤り検出率のPPM構築を試みた。これを品質予測モデルとする。

ソフトウェア開発において検出される誤りの累積数は、工程が進むにつれ一般にS字を描く傾向があり、成長曲線で近似される。その代表的な成長曲線には、ゴンベルツ曲線とロジスティック曲線がある⁽⁴⁾(図2参照)。

ゴンベルツ曲線：

$$y = Ka^{bx} \quad (K > 0, 0 < a < 1, 0 < b < 1) \quad (1)$$

ロジスティック曲線：

$$y = \frac{K}{1 + ae^{-bx}} \quad (K, a, b > 0) \quad (2)$$

見かけの上では、式(1)と(2)は類似したS字型の曲線で、両者ともxが増加するにつれてyは一定値Kに近づく。ロジスティック曲線は変曲点⁽⁵⁾に対して点対称になるが、ゴンベルツ曲線はこの対称性がない。変曲点に対して点対称性がないゴンベルツ曲線はロジスティック

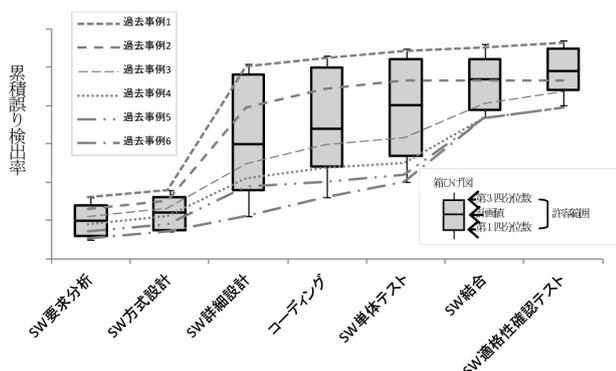


図1 累積誤り検出率の許容範囲

曲線より汎用性があるので、誤りの累積数の近似曲線としてより適合していると考えられる。このことから、本稿では、実際にソフトウェア品質管理の数学モデルとして多く使われているゴンベルツ曲線⁽¹⁾をモデルとして採用した。ゴンベルツ曲線の変曲点の導出概要を付録Aにまとめる。

説明変数xを各フェーズの番号(例：SW要求分析フェーズ：x=1)、目的変数yを2.1節で算出した各フェーズの累積誤り検出率として、式(1)のパラメータを決定する。ゴンベルツ曲線⁽¹⁾のパラメータK,a,bのうちKとbは最小二乗法を使用して決定する。その際、yの差分Δyを用いているために最低4点のデータが必要となる。つまり、少なくとも実施済のフェーズが4つ以上あることが必要である。パラメータaは基準点x=0に対する $y = Ka \equiv y^*$ と最小二乗法で得られたKより決定する(付録B参照)⁽⁶⁾。

2.3 品質予測モデルの構築

方式設計フェーズなどのプロジェクト初期段階では、実測値が4点に満たないため、そのままではゴンベルツ曲線を描くことができない。

この問題を解決し、設計の早期から品質予測を行うための施策を以下に述べる。

2.3.1 設計・製造フェーズにおける品質予測モデル

本試行では、終了していないフェーズのデータは計画値を用いて外挿する方法を採用した。これは、今回予測するプロジェクトと同一組織、メンバーなど類似かつ成功プロジェクトだけから誤り検出傾向を実績値として採用したため、プロジェクトゴール達成を予測するデータとして妥当と考えたからである。

上記方法を用いた初期フェーズからの予測の手順を以下に述べる。

- ①各フェーズにおける累積誤り検出率の計画値を用いてゴンベルツ曲線(計画)を決定する。

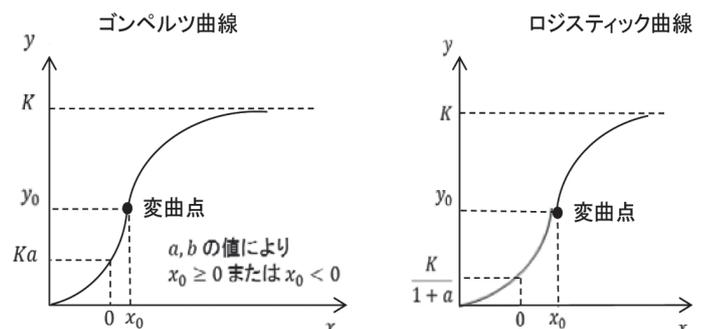


図2 代表的な成長曲線

- ②①の曲線から各フェーズ間の累積誤り検出率の増分を算出する。
- ③終了したフェーズの実績値と②で算出した増分を用いて、終了していないフェーズの累積誤り検出率を外挿して全フェーズのデータを準備する。
- ④準備されたデータを用いて再度ゴンベルツ曲線（予測）を決定する。

各フェーズの実績データが取得されるごとに、上記手順②～④をくり返す。

2.3.2 テストフェーズにおける品質予測モデル

この段階では、既に4フェーズ以上が実施済であるので、2.3.1節で述べたような外挿処理をする必要がない。つまり、実績値のみでゴンベルツ曲線（予測）を決定することができる。

3. 品質予測モデルの適用

本節では、2節で述べた品質予測モデルをパイロットプロジェクトの定量的プロジェクト管理に適用し、得られた効果を①～③に述べる。また、①については3.1節、②については3.2節、③については3.3節に詳細を述べる。

- ①過去の類似かつ成功プロジェクトのデータを用いて導出したPPB⁽¹⁾を利用することで、定量的な状況把握、および的確な対処の確認を行うことができた。
- ②各フェーズ実績の累積誤り検出率を利用して構築したゴンベルツ曲線による品質予測によって、プロジェクトゴール（出荷時残存誤り密度）への到達を制御した。
- ③目標値からの逸脱やゴール未到達と予測された場合の対策として、予測値の許容範囲に対する実績値の分布から、過去情報から想定される原因、確認方法、及び対応方法を示したガイドラインを事前に準備し、フェーズに内在する問題（及びリスク）の是正の手掛かりとして活用した。

3.1 PPBに基づく実績値・許容範囲の監視・管理

図3は、PPB⁽¹⁾の各フェーズの許容範囲に対しての実績値の状態を示している。実績値が、2.1節で導出した許容範囲内にあるか否かを確認し、許容範囲外の場合には、原因を分析しガイドラインを参考としてその原因に応じた対処を行った。コーディングフェーズでの範囲外の対策を実施することでソフトウェア単体テストフェーズにおいては実績値が許容範囲内に収まり、当該フェーズでは「問題無し」と判断することができた。

3.2 PPMによるプロジェクトゴール達成の判断

図4は、ソフトウェア方式設計フェーズ終了時点における、後フェーズのトレンド予測を示している。2.3.1節で述べた外挿方法を採用し予測を行った結果である。方式設計フェーズ時点での誤り検出率をみると許容範囲を超えてしまっているが、以降のフェーズで計画通り誤り検出をしていくことでプロジェクト最終フェーズであるソフトウェア適格性確認テスト終了地点ではプロジェクト目標とする許容範囲内に収まり（目標値におよそ10%以内の誤差で到達）、プロジェクトゴール達成となる予測が得られた。

3.3 「コントロールノブ」による品質向上施策の試行

図4で示したソフトウェア方式設計フェーズでは、最終フェーズでのゴール未達の危険度合いは低いものの、実績値が許容範囲（=指標範囲）を逸脱し上限方向にある状態であり、このまま後続フェーズで指標範囲の上限を辿るような誤り指摘件数で進む場合、プロジェクトゴールを達成できないことも予測できた。このため、各フェーズにおいて実績値の許容範囲の逸脱に対する対策を過去事例より抽出し、ガイドライン（図5に示す「コン

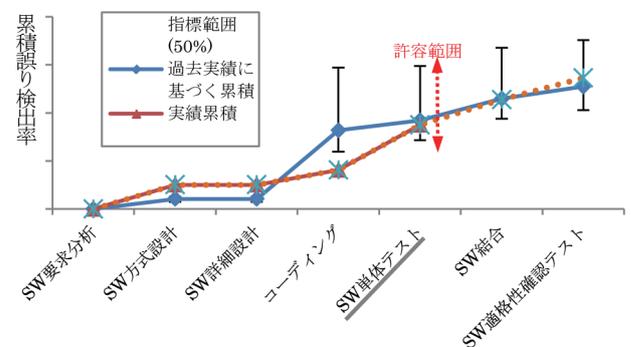


図3 許容範囲と実績状態

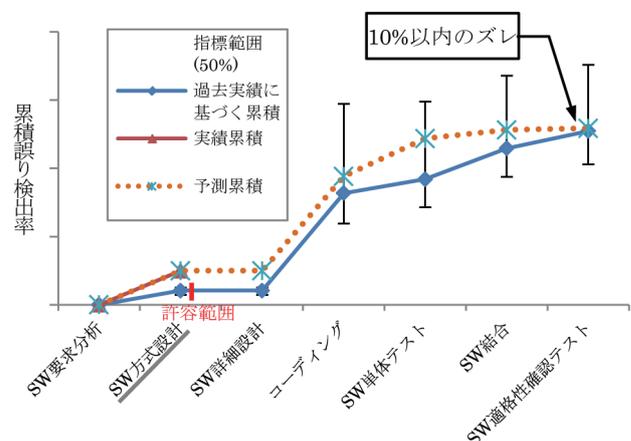


図4 PPMによる品質予測

フェーズ	パターン	想定される原因	確認方法	対応方法
フェーズA プロジェクト開始	想定以上の作業を要求されている。	見積り内容以上の詳細性を要求しているため。	見積り内容以上の詳細性を要求しているため。	顧客と作業チームの認識を調整する。
	作業以上の工程を要している。	作業以上の工程を要している。	作業以上の工程を要している。	作業の要否を再確認する。
	顧客の要求の意味である。	顧客仕様に対し、要求内容的結果が異なるため。	顧客仕様に対し、要求内容的結果が異なるため。	顧客の要求を再確認する。
作業の進捗が不足している。	複雑な顧客仕様に対し、要求分析の結果が異なるため。	複雑な顧客仕様に対し、要求分析の結果が異なるため。	要求仕様からは見えない、隠れた要求が明かを確認する。	
フェーズB プロジェクト進行	作業以上の工程を要している。	作業以上の工程を要している。	作業以上の工程を要している。	作業の要否を再確認する。
	顧客の要求の意味である。	顧客仕様に対し、要求内容的結果が異なるため。	顧客仕様に対し、要求内容的結果が異なるため。	顧客の要求を再確認する。
	作業の進捗が不足している。	複雑な顧客仕様に対し、要求分析の結果が異なるため。	複雑な顧客仕様に対し、要求分析の結果が異なるため。	要求仕様からは見えない、隠れた要求が明かを確認する。
フェーズC プロジェクト完了	作業以上の工程を要している。	作業以上の工程を要している。	作業以上の工程を要している。	作業の要否を再確認する。
	顧客の要求の意味である。	顧客仕様に対し、要求内容的結果が異なるため。	顧客仕様に対し、要求内容的結果が異なるため。	顧客の要求を再確認する。
	作業の進捗が不足している。	複雑な顧客仕様に対し、要求分析の結果が異なるため。	複雑な顧客仕様に対し、要求分析の結果が異なるため。	要求仕様からは見えない、隠れた要求が明かを確認する。

図5 コントロールノブの例

「コントロールノブ」と命名)を参考にプロジェクトの実情に合わせて対策を実施した。

実績値が許容範囲の上限を越えていたので、コントロールノブの対象フェーズとパターンから、想定される原因と対策案を特定しパイロットプロジェクトへ助言できた。結果として、対象プロジェクト内で本対策案を適用したことで、以降のフェーズで許容範囲に実績結果が収まり、最終フェーズでのプロジェクトゴールとして目標設定した出荷時残存誤り密度以内に達成できるという予測結果を導き出すことができた。

4. 結果

3節に述べたことより、プロジェクトが実施中の各フェーズ内においてプロジェクトゴールの目標とする許容範囲に予測値が収まるか否かを定量的に判断できた。これは、プロジェクトゴール逸脱の危険予知を早期に行え、問題(リスク)解決への取り組みを促す効果を得られたと考えられる。組織の出荷時の目標値でのプロジェクトのリリースもできた。

以上のことから、本品質予測モデルを使った定量的プロジェクト管理が有効であるという結果を得ることができたと考える。

5. むすび

予測モデルを利用した定量的プロジェクト管理を行う目的は、組織目標につながるプロジェクトゴールを達成することにある。よって、予測モデルを利用した定量的プロジェクト管理を組織内で継続して利用し改善していくことが重要であり、そのためには予測精度を高くすることが必要であると考えられる。

組織内にはさまざまなプロジェクトがあり、プロジェクトごとに特性が異なる。そのため、評価尺度(プロジェクトの成否を評価するものや、プロジェクトの中間状況を評価するもの)が同じでも、部門、プロジェクトの達成目標によっては予測モデルが適合しない場合がある。これは、プロジェクトの特性が異なると、評価尺度

ごとに予測モデルが異なるからである。それぞれの部門、プロジェクトごとに評価尺度に合った実績データを収集し分析することにより、予測を実施できると考えると考える。

本稿の改善活動による品質改善の実施を継続しながら、上記に挙げた改善検討事項の解決に向けた活動を並行して実施することが、更なる予測精度向上に繋がり、利用の促進を促すことができると考える。

今後は、PPB、PPMを活用し組織として製品品質やプロセスについて定量的なマネジメントが実施できることを目指す。

参考文献

- (1) Japanese Language Translation of CMMI for Development, <http://cmminstitute.com/resource/japanese-language-translation-of-cmmi-for-development-v1-3/>, CMMI Institute, 2012
- (2) 岡野麻子・矢田部学, 生産性と品質データの解析手法についての提案, MSS技報, Vol.25, 2014
- (3) 東大教養学部統計学教室編, 統計学入門, 東大出版会, 1991
- (4) 三觜武, ソフトウェアの品質評価法, 日科技連出版社, 1981
- (5) 石原繁・浅野重初, 微分積分, 裳華房, 1997
- (6) 石田秀人, シグモイドのゴンペルツ曲線の求め方, <http://home.a02.itscom.net/coffee/tako06.html>

付録

A. ゴンペルツ曲線の変曲点

ゴンペルツ曲線をxで微分するために、本文の式(1)の両辺の自然対数をとると

$$\ln y = \ln K + b^x \ln a$$

両辺をxで微分すると、以下の結果を得る。

$$y' = K \cdot \ln a \cdot \ln b \cdot a^{b^x} \cdot b^x$$

さらに両辺の対数をとって微分すると、以下の結果を得る。

$$y'' = K \cdot \ln a \cdot (\ln b)^2 \cdot a^{b^x} \cdot b^x \cdot (\ln a \cdot b^x + 1)$$

変曲点のx座標とy座標は、 $a^{b^x} > 0$ および $b^x > 0$ に注意すると、 $y'' = 0$ より以下のようになる。

$$x_0 = \frac{\ln(-1/\ln a)}{\ln b}, \quad y_0 = Ke^{-1}$$

B. ゴンペルツ曲線のパラメータの決定

本文の式(1)の両辺の常用対数をとると

$$\log y = \log K + b^x \log a \quad (B-1)$$

$x \rightarrow x+1$ のとき、 $y \rightarrow y + \Delta y$ とすると

$$\log(y + \Delta y) = \log K + b^{x+1} \log a \quad (\text{B-2})$$

(B-2) - (B-1) より

$$\log(y + \Delta y) - \log y = (b - 1)b^x \log a$$

(B-1) を代入して $b^x \log a$ を消去すると

$$\log(y + \Delta y) - \log y = (b - 1) \log y + (1 - b) \log K$$

これを以下のように表す。

$$Y = AX + B \quad (\text{B-3})$$

ただし

$$X \equiv \log y, \quad Y \equiv \log(y + \Delta y) - \log y$$

$$A \equiv b - 1, \quad B \equiv (1 - b) \log K,$$

(B-3) の係数AとBは最小二乗法で決定する。すなわち、実測データ (X_i, Y_i) に対して

$$L \equiv \sum_i [Y_i - (AX_i + B)]^2 \rightarrow \min$$

より

$$\frac{\partial L}{\partial A} = -2 \sum_i X_i [Y_i - (AX_i + B)] = 0$$

$$\Rightarrow A \sum_i X_i^2 + B \sum_i X_i = \sum_i X_i Y_i \quad (\text{B-4})$$

$$\frac{\partial L}{\partial B} = -2 \sum_i [Y_i - (AX_i + B)] = 0$$

$$\Rightarrow A \sum_i X_i + Bn = \sum_i Y_i \quad (\text{B-5})$$

(B-4) と (B-5) を解くと

$$A = \frac{n \sum_i X_i Y_i - \sum_i X_i \sum_i Y_i}{n \sum_i X_i^2 - (\sum_i X_i)^2}, \quad B = \frac{\sum_i Y_i - A \sum_i X_i}{n}$$

以上のようにAとBが具体的に算出されるとパラメータbとKは

$$b = A + 1, \quad K = 10^{\frac{B}{1-b}}$$

のように算出できる。

パラメータaについては、以下のように算出する。本文の式 (1) で $x=0$ とすると $y=Ka$ なので、 $x=0$ に対する $y=y^*$ を与えると、以下ようになる。

$$a = \frac{y^*}{K}$$

ただし、パラメータaの決定方法は、 y^* の与え方に依存するので、ゴンペルツ曲線が実測データにフィットするように調整が必要となる。

執筆者紹介

岡野 麻子

1997年入社。入社以降、鎌倉事業部で防衛分野に従事。2005年4月より品質保証に従事。2012年4月より生産技術部門としてプロセス改善に従事。

岡田 政之

1991年入社。入社以降、鎌倉事業部で防衛分野に従事。2012年1月より品質保証に従事。

土屋 義兼

2013年入社。入社以降、鎌倉事業部で生産技術部門としてプロセス改善に従事。