

# 尖度と歪度を考慮した予測モデルの検討

福居誠二<sup>†1</sup> 門田暁人<sup>†1</sup>

**概要:** 本稿では、尖度と歪度を考慮した予測モデル有効性について検討した結果について報告する。

## A Study on Prediction Models Considering Kurtosis and Skewness

SEIJI FUKUI<sup>†1</sup> AKITO MONDEN<sup>†1</sup>

**Abstract:** This paper discusses about the effectiveness of prediction model considering kurtosis and skewness.

### 1. はじめに

ソフトウェア開発プロジェクトにおいて、納期の遅れ、コストの超過といったプロジェクトの失敗を未然に防ぐためには、ソフトウェア工数の予測が必須であり、その予測精度と高めることが重要となる。このため、様々な工数予測方法が提案されてきた。過去の開発実績データに基づく工数予測方法として、線形重回帰分析がよく用いられており、特に、予測モデル構築に用いるデータを対数変換し、データを正規分布に近づけることで予測精度を向上させる log-log 重回帰分析が効果的である[2]。

本稿では、log-log 重回帰分析の予測精度をさらに高める試みとして、データの対数変換後に、尖度と歪度も変換することで、データの分布をより正規分布に近づける方式について検討する。

以降、2 章では、データの尖度と歪度について説明し、3 章では、本研究のアプローチを述べる。4 章では、本研究のアプローチのケーススタディについて述べる。5 章はまとめの今後の課題を述べる。

### 2. 尖度と歪度

分布の特徴を表す指標として、平均  $\mu$  や分散  $\sigma^2$  はよく知られているが、これらの尺度は、分布のピークの鋭さや非対称性については考慮されない。尖度と歪度はこれらを定量化する指標である。

尖度とは、分布のピークと裾が正規分布からどの程度異なるかを表す指標であり、正規分布に完全に従ったデータは、尖度の値が 0 になる。尖度の値が正の分布は、正規分布と比べて鋭いピークと重い裾を持ち、尖度の値が負の分布は、正規分布と比べて平坦なピークと軽い裾を持つことを表す。従って、尖度の値が 0 から大きく外れると、データが正規分布から遠くなる。

歪度とは、データの非対称性を示す指標であり、データが対

称的になるほど、歪度の値は 0 に近づき、正規分布に完全に従ったデータは、歪度の値が 0 になる。歪度の値が正の分布は、分布の裾が右側に向かい、歪度の値が負の分布は、分布の裾が左側に向かうことを表す。

### 3. 研究のアプローチ

#### 3.1 尖度と歪度の変換方法

対数変換したデータに対し、尖度と歪度の値を変更するために、(1)の式で定義される sinh-arcsinh transformation [1]を用いる。

$$H(x; \epsilon, \delta) = \sinh[\delta \sinh^{-1}(x) - \epsilon] \quad (1)$$

この関数は、標準正規分布化したデータを入力  $x$  とし、 $\epsilon$ ,  $\delta$  を実数の範囲で与えることで、尖度と歪度の値を変更したデータを得ることができる。

本研究では、与えられたデータ群に対し、尖度と歪度の値が共に 0 となるような  $\epsilon$ ,  $\delta$  を探索により求める。

#### 3.2 提案方法

本研究で提案する工数予測方法を図 1 のフローチャートに示し、以下で説明する。なお、図中のフローチャート左部分をモデル構築フェーズ、右部分を予測フェーズとする。

モデル構築フェーズでは、予測モデル構築に用いる全ての変数について、データ(変数値)を対数変換し、標準正規分布に近づけるために平均を 0、標準偏差を 1 に変換する。次に、(1)の関数を用いて、尖度、歪度を 0 に変更し、線形重回帰分析を行う予測モデルを構築する。

予測フェーズでは、まず、構築した予測モデルを用いて目的変数を算出する。この算出値は、対数変換、平均と標準偏差の変換、尖度と歪度の変換が行われていることを前提とした値であるため、それぞれ逆変換を行うことで、予測工数を得る。

<sup>†1</sup> 岡山大学  
Okayama University

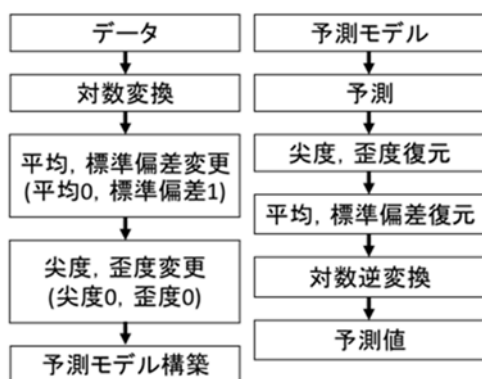


図 1 提案方法 (左:モデル構築フェーズ, 右:予測フェーズ)

Figure 1 Proposed method (left side: model construction phase, right side: estimation phase)

表 1 各方法による工数予測値の絶対誤差平均, 相対誤差平均

Table 1 MMRE and MMAE of effort estimation of each method.

	方法(i)	方法(ii)	方法(iii)
絶対誤差平均	5.826	6.200	5.667
相対誤差平均	0.814	0.382	0.307

## 4. 実験

### 4.1 実験方法

Albrecht データセット[3](パラメータ数 7, プロジェクト数 24) に対して, (i)線形重回帰分析, (ii)log-log 重回帰分析, および, (iii)提案方法をそれぞれ用いて予測モデル構築を行い, 各予測値を算出する. そして, 実測値と各方法での予測値との絶対誤差平均と相対誤差平均を比較する. なお, 本実験では, 交差検証は行っていないために, ここでの誤差というのは残差のことである.

### 4.2 結果

各方法で算出された予測値と実測値との絶対誤差平均と相対誤差平均を表 1 にまとめる.

表 1 から, 提案方法で算出された予測値は, 単純な線形重回帰分析より改善しているのみならず, log-log 重回帰分析と比較しても, 絶対誤差平均と相対誤差平均の両方の値が小さくなっている. つまり, 他の 2 つの方法に比べて, 提案方法により構築された予測モデルは, 実データに対する当てはまりが良いことを示している.

## 5. おわりに

本稿では, 対数変換したデータの尖度と歪度を変換させ, より正規分布に近づけることによる, 工数予測モデルの予測性能向上に関して検討した. 今後は, より多くのデータを対象とした提案方法の評価, および, 交差検証による未知のデータに対する提案方法の有効性の評価を行う予定である.

## 参考文献

- [1] Jones, M. C. and Pewsey A, "Sinh-arcsinh distributions," *Biometrika*, Vol.96, No.4, pp.761-780, Dec. 2009.
- [2] 門田暁人, 小林健一, "線形重回帰モデルを用いたソフトウェア開発工数予測における対数変換の効果," *コンピュータソフトウェア*, Vol.27, No.4, pp.234-239, Nov. 2010.
- [3] The tera-PROMISE Repository, <http://openscience.us/repo/>