

# コスト予測における対数変換すべき変数の自動判別手法の提案

三好 健太<sup>†</sup> 生方 克馬<sup>‡</sup> 柿元 健<sup>†</sup> 楠本 真二<sup>‡</sup>

香川高等専門学校電気情報工学科<sup>†</sup> 大阪大学大学院情報科学研究科<sup>‡</sup>

## 1. はじめに

近年、ソフトウェア開発は大規模化・複雑化に伴い人員や時間といった多くのコストが必要となってきた。ソフトウェアが複雑になるにつれて、バグの発生や人間的な遅延により「品質」、「コスト」、「納期」の面で開発当初の予定通りに開発することは難しくなる。また、納期遅れやユーザの求める品質を満たしていない等のソフトウェア開発の失敗が結果として依頼元の企業経営に影響を及ぼすことになる場合もある[1]。そのため、ソフトウェア開発において、コスト超過や納期遅れなどのリスクを抑えつつ、ユーザから要求された品質や信頼性を確保することが重要である。

ソフトウェア開発において、スケジュール管理等の開発管理のために開発コストの予測が行われる。高い精度でコスト予測することでコスト超過・納期遅れ・品質の低下といったプロジェクトの失敗を未然に防ぐことができる。

コスト予測の精度を高める方法として、コスト予測に用いるメトリクスを対数変換することで値の分布を線形に近づける方法が従来から行われている[2]。しかし、どのメトリクスでも対数変換することで予測精度が向上するわけではなく、例えば、値の分布が線形に近いメトリクスでは対数変換してしまうと値の分布が線形から離れてしまい、予測精度を下げることになる。一方、総当たりで判別すると計算量が大きいため大きなデータセットを用いる場合には現実的な時間で処理が終わらない可能性がある。そこで、本研究では対数変換すべきメトリクスを少ない計算量で自動判別する手法を提案する。また、評価実験により単純変換との精度比較を行い、提案手法により精度が向上することを確認する。

## 2. 線形重回帰分析

線形重回帰分析は多変量解析の一手法であり、ソフトウェア開発のコスト予測に広く用いられている。

線形重回帰分析では、予測対象のメトリクス(目的変数)と目的変数に影響を与えるメトリクス(説明変

数)との関係を一次式で表す分析法である。メトリクス間の関係を数式モデル化することで既知の説明変数から未知の目的変数の値を予測することが可能となる。

一般的なモデル式は式(1)の形で表わされ、実測値とモデル式の残差が最も小さくなるように係数  $a$  と  $b$  を決める。 $\hat{Y}$  は目的変数の予測値、 $X_i$  は説明変数を表わしている。

$$\hat{Y} = a_1 X_1 + a_2 X_2 + \dots + a_n X_n + b \quad (1)$$

## 3. 提案手法

### 3.1 概要

少ない計算量で対数変換すべきメトリクスを判別するために、本論文ではメトリクスをいくつか分割し、分割した区間ごとの傾きの変移から値の分布を判断することにより対数変換すべきメトリクスであるか判断する手法を提案する。

### 3.2 手順

提案手法の判別手順は以下の通りである。

- ① 各メトリクスを昇順にソートする。
- ② ソートしたメトリクスを一定のプロジェクト数で分割し、分割した各区間の傾きを求める。
- ③ 前後の区間の傾きを比較することで、連続して傾きが増加しているか判断する。ある一定区間以上連続して傾きが増加していれば、そのメトリクスは対数変換すべきと判別する。
- ④ 対数変換する際にメトリクスの値が 0 であれば対数変換できないため、メトリクスの値をすべて +1 してから常用対数に変換する。
- ⑤ ②～④を全メトリクスで適用し、全メトリクスにおいて対数変換するべきかを判別する。

本論文の評価実験では 7 プロジェクトごとに分割し、連続数を 2 とした。

## 4. 評価実験

提案手法の性能を確かめるために評価実験を行った。提案手法で対数変換するメトリクスを判別し工数予測を行った場合と、全ての説明変数を対数変換し工数予測を行った場合の予測精度を比較する。評価実験にはカナダのソフトウェア開発企業の実績データである Desharnais データセット[3]を用いた。データセットにはプロジェクト 81 件、メトリクス 12 個が記録されているが、欠損値を含むプロジェクトを除いたプロジェクト 77 件を用いた。

Automatic identification of variables in the cost estimates to be logarithmic  
Kenta Misyoshi<sup>†</sup> Katsuma Ubukata<sup>‡</sup> Takeshi Kakimoto<sup>†</sup> Shinji Kusumoto<sup>‡</sup>  
Department of Electrical and Computer Engineering, Kagawa National College of Technology<sup>†</sup>  
Graduate School of Information Science and Technology, Osaka University<sup>‡</sup>

評価実験では精度評価によく用いられる Leave-one-out 方を用いた。Leave-one-out 法の説明を以下に示す。

- ① 全プロジェクト  $n$  件から 1 件抽出する
- ② 抽出した 1 件を予測対象プロジェクト、残り  $n-1$  件を過去プロジェクトとし、過去プロジェクトをもとに予測対象プロジェクトの工数を予測する
- ③ 予測対象プロジェクトの工数の実測値は既知であるので工数の実測値と予測値から誤差を計算する。
- ④ 予測対象プロジェクトとして抽出するプロジェクトを変更し、①～③を  $n$  回繰り返す。
- ⑤  $n$  回繰り返された③により求められた  $n$  個の誤差をもとに予測精度を算出する。

評価指標には MMRE, MdMRE, Pred(25), MMER, MBRE を用いた。MMRE は実測値からみた予測の相対誤差(MRE)の平均値, MdMRE は実測値からみた予測の相対誤差(MRE)の中央値, Pred(25)は MRE が 0.25(25%)以下の予測結果が全体に占める割合である。MMER は予測値からみた実測値の相対誤差(MER)の平均値である。MRE は過大見積を MER は過小見積で性能を発揮するため、MRE と MER の良いところを取った評価指標である BRE の平均値である MBRE も評価指標に用いた。

MRE, MER, BRE はそれぞれ式(2)-(4)で表わされる。ここで、real は実測値、est は予測値を表す。

$$\text{MRE} = \frac{|\text{real} - \text{est}|}{\text{real}} \quad (2)$$

$$\text{MER} = \frac{|\text{real} - \text{est}|}{\text{est}} \quad (3)$$

$$\text{BRE} = \frac{|\text{real} - \text{est}|}{\max(\text{real}, \text{est})} \quad (4)$$

MMRE, MdMRE, MMER, MBRE は値が小さいほど、Pred(25)は値が大きいくほど予測精度が高いことを表わす。

## 5. 実験結果

表 1 に提案手法と全メトリクスを対数変換した場合、および、対数変換しなかった場合の各評価指標の値を示す。対数変換しない場合と比較すると、提案手法、全変換した場合ともに予測精度は著しく向上している。そして、対数変換した場合でも、提案手法は、MMER, MBRE で 0.025 をはじめ、全ての評価指標において、全変換した場合よりも予測精度が向上している。

## 6. 考察

提案手法では対数変換することにより値の分布が線形に近づく場合のみ対数変換し、対数変換することにより値の分布が線形から離れていく場合は対数

表 1 各手法の評価指標の値

	提案手法	全変換	変換なし
MMRE	0.456	0.466	0.891
MdMRE	0.392	0.396	0.459
Pred(25)	0.377	0.364	0.254
MMER	0.417	0.442	0.934
MBRE	0.564	0.589	1.407

変換をしない。つまり、対数変換により精度向上が期待されるメトリクスのみを対数変換するため対数変換をしない場合よりも予測精度が良くなったと考えられる。

対数変換するメトリクスの判別において最も精度を向上させる方法は総当たりで行うことである。しかし、計算量で比較すると総当たりでの計算量は  $O(n^3)$ 、提案手法の計算量は  $O(n)$ 、対数変換しない場合の計算量は  $O(1)$  であり、総当たりでは莫大な計算量が必要となる。そのため、メトリクス数が多い場合には現実的な時間で処理が終わらない可能性もあり、対数変換するメトリクスの判別をコスト予測ツールに組み込む場合等に問題となることが考えられる。

## 7. まとめ

本論文では対数変換すべきメトリクスを自動判別する方法を提案した。評価実験の結果、全メトリクスを対数変換した場合や対数変換しなかった場合よりも予測精度が向上することが確認された。

提案手法のデータを分割する個数や判別する閾値である連続数を変化させ、予測精度が最良になる値について検討することが今後の課題である。

## 謝辞

本研究は一部、日本学術振興会科学研究費補助金基盤研究(A)(課題番号:21240002)、基盤研究(C)(課題番号:20500033)の助成を得た。また一部、文部科学省「次世代IT基盤構築のための研究開発」の委託に基づいて行われた。

## 参考文献

- [1]国立大学法人奈良先端科学技術大学院大学:ソフトウェア開発におけるエンピリカルアプローチ, 株式会社アスキー, 2008.
- [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, 2010.
- [3]Desharnais, J. M. :Analyse statistique de la productive des projets informatique a partie de la technique des point des fonction, Master's thesis, Univ. of Montreal, 1989.