

# Analogy 手法における Analogy 数決定法の比較

矢野 博暉<sup>†1</sup> 柿元 健<sup>†2</sup>

Analogy 手法における類似プロジェクト (Analogy) 数の決定法は明らかにされていない。そこで、本稿では Analogy 数の決定法を比較する。

## Comparison of analogy decision methods in estimation by analogy

HIROKI YANO<sup>†1</sup> TAKESHI KAKIMOTO<sup>†2</sup>

In this paper, we comparison of the decision method of the number of analogies in estimation by analogy.

### 1. はじめに

ソフトウェア開発において予算の決定はプロジェクトを開始するための必須事項である。ソフトウェア開発上における予算は、プロジェクトの作業量 (工数) を用いて決定することが多い<sup>2)</sup>。従って、プロジェクトの予算決定には、工数予測が重要となる。工数を予測する際、予測対象のプロジェクト (予測プロジェクト) に過去に行われた類似しているプロジェクトの工数の実測値に基づいて予測する手法である類似性に基づく予測手法 (Analogy 手法)<sup>3)</sup> が広く用いられている。

従来研究において Analogy 手法を用いる際、予測プロジェクトに関わらず類似プロジェクト (Analogy) 数は固定して予測している。しかし、Analogy 数の明確な決定方法は定まっていない<sup>1)</sup>。また、予測プロジェクトによって、プロジェクトの特性は異なるため、予測プロジェクトごとに動的に Analogy 数を決定することが望ましいが、Analogy 数を動的に決定する方法も明確とはなっていない。

そこで、本稿では、Analogy 数を固定した場合といくつかの方法により Analogy 数を動的に決定した場合を比較することにより、最適な Analogy 数決定法の検討を行う。

### 2. Analogy 手法

Analogy 手法は、予測プロジェクトに各メトリクスの値が類似している過去プロジェクトを選択し、選択された類似プロジェクトの値に基づいて予測プロジェクトの予測値を算出する手法である。以下に本稿で用いた Analogy 手法の手順を示す。

- ① データセットに含まれるカテゴリー変数に対してダミー変数化を行う。
- ② メトリクスごとの類似度への影響を均等にするため、

メトリクスごとに正規化を行う。

- ③ 予測プロジェクトと各過去プロジェクトの間のユークリッド距離に基づいた類似度を算出する。
- ④ 決定された Analogy 数の類似プロジェクトの実測値を用いて、類似度による加重平均により予測値を算出する。

### 3. Analogy 数決定法

#### 3.1 Analogy 数固定

予測プロジェクトに関わらず、決められた Analogy 数で予測値を算出する方法である。本稿では Analogy 数を 1~3 で固定した場合を用いる。

#### 3.2 変動係数を用いた方法

変動係数は標準偏差を平均値で除した値であり、単位の無い指標である。類似プロジェクト  $P_i$  の工数の実測値を  $c_i$ 、類似プロジェクトのセットを  $A$ 、類似プロジェクトセット  $A$  の工数の実測値の平均を  $A_{AV}$ 、Analogy 数を  $k$  個としたとき、変動係数  $CV_A$  は式(1)で定義される。

$$CV_A = \frac{1}{A_{AV}} \sqrt{\frac{1}{k-1} \sum_{i \in A} (c_i - A_{AV})^2} \quad (1)$$

#### 3.3 四分位散布係数を用いた方法

四分位散布係数は四分位偏差を第1四分位と第3四分位の和で除した値である。類似プロジェクトのセットを  $A$ 、類似プロジェクトセット  $A$  の第1四分位を  $Q_{A1}$ 、第3四分位を  $Q_{A3}$  としたとき、四分位散布係数  $QC_A$  は式(2)で定義される。

$$QC_A = \frac{Q_{A3} - Q_{A1}}{Q_{A3} + Q_{A1}} \quad (2)$$

<sup>†1</sup> 国立高専機構香川高等専門学校専攻科創造工学専攻  
Advanced Course in Industrial and Systems Engineering, National Institute of Technology, Kagawa Campus

<sup>†2</sup> 国立高専機構香川高等専門学校電気情報工学科  
Department of Electrical and Computer Engineering, National Institute of Technology, Kagawa Campus

### 3.3 メトリクスの値の差に基づいた方法

メトリクスの値の差に基づいた方法は、予測プロジェクトの各メトリクスの値と類似プロジェクトの各メトリクスの値の差を算出し、正負の偏りの差を用いる方法である。以下に、メトリクスの値の差に基づいた手法の手順を示す。

- ① メトリクスごとに予測プロジェクトと各類似プロジェクトとのメトリクスの値の差を求める。
- ② Analogy 数ごとに各メトリクスの差が正の平均値、負の平均値を求め、それぞれの平均値の和を算出する。
- ③ 各メトリクスの正負の平均値の和の二乗平均を算出する

動的な Analogy 数決定法は、いずれも各指標が最小値をとるプロジェクト数を Analogy 数とする。

### 3. 評価実験

評価実験の目的は、Analogy 決定法を比較することで、最適な Analogy 決定法の検討を行うことである。評価実験には、tera-PROMISE リポジトリの Effort<sup>4)</sup>のうち、Albrecht, China, Desharnais, Kemerer, Miyazaki94 の5つの実績データを用いた。

評価実験の方法として Leave-one-out 法を採用し、得られた工数の予測値と記録されている実測値から工数予測精度を算出し、Analogy 決定法の比較を行った。

工数予測精度の評価尺度として、過大予測と過小予測とともに評価可能な指標である BRE の平均値である MBRE(Mean Balanced Relative Error)を用いた。予測プロジェクト数を  $n$  件、実測値を  $X_i$ 、予測値を  $E_i$  ( $i=1 \sim n$ ) とすると、MBRE は式(3)で表される。

$$MBRE = \frac{1}{n} \sum_{i=1}^n \left( \frac{|X_i - E_i|}{\min(X_i, E_i)} \right) \quad (3)$$

MBRE は、値が小さいほど予測精度が高いことを示す。

### 4. 実験結果

評価実験の結果を表 1 に示す。表中の  $k$  は Analogy を固定した場合における Analogy 数を示しており、CV は変動係数を用いた方法、QC は四分位散布係数を用いた方法、DV はメトリクスの値の差に基づいた方法を示す。

表 1 より desharnais のデータセットの CV, QC などや miyazaki94 のデータセットの DV など、データセットの種類によっては動的な決定手法の場合の精度が高いという結果となった。さらに Analogy 数固定同士での予測精度を比較すると、Analogy 数を 1 に固定した場合よりも 2,3 に固定した場合のほうが予測精度が高いという結果となった。

表 1 各 Analogy 決定法の MBRE

データセット		A	C	D	K	M
固定	k=1	1.11	0.69	1.18	1.12	1.35
	k=2	0.93	0.60	1.00	0.94	1.06
	k=3	0.98	0.61	0.95	0.86	1.01
動的	CV	0.96	0.66	0.65	0.89	1.09
	DV	1.00	0.64	0.90	0.92	0.90
	QC	0.92	0.66	0.65	0.94	1.04
プロジェクト数		24	498	81	15	48
メトリクス数		7	17	10	6	7

A は Albrecht, C は China, D は Desharnais, K は Kemerer, M は Miyazaki94 を示している。

### 5. 考察

変動係数、四分位散布係数を用いた手法では Desharnais のようにメトリクスの数が多いデータセットにおいて、高い精度が得られたことから変動係数、四分位散布係数を用いた手法はメトリクスの数が多いプロジェクトに対して適していると考えられる。メトリクスの値の差に基づいた手法では China のようにプロジェクト数が多いデータセットにおいて高い精度が得られたことからメトリクスの値に差に基づいた手法は過去プロジェクト数が多い予測プロジェクトに対して適していると考えられる。したがって、予測対象のデータによって各手法を使い分けた場合、更なる精度の向上が期待される。

### 6. おわりに

本稿では Analogy 手法の複数の Analogy 数決定法について工数予測精度を比較した。評価実験の結果、データセットの種類によって各動的な Analogy 数決定手法の精度の違いが見られた。

各 Analogy 決定手法に適したデータセットの条件を検討することが今後の課題である。

### 参考文献

- 1) Kocaguneli, E., Menzies, T., Bener, A., and Keung, J.W.: Exploiting the essential assumptions of analogy-based effort estimation, IEEE Transaction Software Engineering, Vol.38, No.2, pp.425-438 (2012).
- 2) Menzies, T., and Sayyad, J.: Effort, tera-PROMISE, <http://opensci-ence.us/repo/effort/>
- 3) Shepperd, M., and Schofield, C.: Estimating software project effort using analogies, IEEE Transaction Software Engineering, Vol.23, No.12, pp.736-743 (1997).
- 4) The Standish Group: Extreme Chaos, (2001) [http://www.Standishgroup.com/sample\\_reserc](http://www.Standishgroup.com/sample_reserc)