

類似性に基づく工数予測における 適切でない予測プロジェクトの判別の試み

中村 哲彬[†] 柿元 健[†] 楠本 真二[†]

[†] 大阪大学大学院情報科学研究科

〒 565-0871 大阪府吹田市山田丘 1 番 5 号

E-mail: [†]{t-nakamr,kakimoto,kusumoto}@ist.osaka-u.ac.jp

あらまし ソフトウェア開発において、プロジェクトを円滑に進めるために開発工数の予測が行われる。適切な開発工数を予測することで、コスト超過や納期遅れ等のリスクを抑えることが可能になる。これまでに多数の工数予測手法が提案されているが、近年、過去に実施されたプロジェクトとの類似性に基づく工数予測 (Estimation by Analogy) 手法 (EbA 手法) が注目を集めている。しかし、EbA 手法は過去プロジェクト中に、予測対象のプロジェクト (以下、予測プロジェクト) と類似したプロジェクトが存在しない場合、適切な工数予測が期待できない。本稿では、EbA 手法において、高い精度での工数予測が期待できない (予測に適さない) プロジェクトを判別する方法を提案する。ソフトウェア開発企業で収集された 77 件のソフトウェア開発データを用いた評価実験の結果、提案手法で判別された予測に適さないプロジェクトを除いて工数予測を行った場合、提案手法を適用しない場合と比べて予測精度が改善した。

キーワード 工数予測, Estimation by Analogy, 類似度, ソフトウェア開発データ

Discrimination of Unsuitable Estimation Project in Estimation by Analogy

Tetsuaki NAKAMURA[†], Takeshi KAKIMOTO[†], and Shinji KUSUMOTO[†]

[†] Graduate School of Information Science and Technology, Osaka University

Yamadaoka1-5, Suita City, Osaka, 565-0871 Japan

E-mail: [†]{t-nakamr,kakimoto,kusumoto}@ist.osaka-u.ac.jp

Abstract Effort estimation is important to software development to reducing risk of project failure. Recently, Estimation by Analogy (EbA) method which is based on similar past projects has been attracting attention in various effort estimation methods. EbA method requires projects which are similar to estimation project for high estimation accuracy. In this paper, we propose the method which discriminates against unsuitable (high estimation accuracy is not expected) estimation project in EbA. We evaluated the proposed method by using a dataset of 77 projects. The result showed that estimation accuracy is improving by using the proposed method.

Key words Effort Estimation, Estimation by Analogy, Similarity, Software development data

1. はじめに

ソフトウェア開発プロジェクトにおいて開発を円滑に進めるために、開発工数の予測が行われる。工数の予測を行うことで、開発に必要な期間や要員数が決定され、コスト超過や納期遅れ等のリスクを抑えることが可能になるため、多くの工数予測手法が提案されている [1][2][3]。過去プロジェクトのデータセット全体に対してただ一つの予測式を

構築する重回帰分析などに対して、予測対象のプロジェクト (以下、予測プロジェクト) ごとに個別に予測式を構築する類似性に基づく工数予測 (Estimation by Analogy) 手法 (EbA 手法) [9][10] が、近年注目されている。

EbA 手法は、過去プロジェクト中に予測プロジェクトと類似したプロジェクトが含まれていることを前提としている。そのため、EbA 手法では、予測したいプロジェクトに類似したプロジェクトが過去プロジェクト中に存在しない

表 1 EbA 手法で用いるデータセット

	M_1	M_2	$\dots$	M_j	$\dots$	M_y
P_1	$V_{1,1}$	$V_{1,2}$	$\dots$	$V_{1,j}$	$\dots$	$V_{1,y}$
P_2	$V_{2,1}$	$V_{2,2}$	$\dots$	$V_{2,j}$	$\dots$	$V_{2,y}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
P_i	$V_{i,1}$	$V_{i,2}$	$\dots$	$V_{i,j}$	$\dots$	$V_{i,y}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
P_x	$V_{x,1}$	$V_{x,2}$	$\dots$	$V_{x,j}$	$\dots$	$V_{x,y}$

場合，その予測プロジェクトは高い精度で工数予測することができないという問題がある（以降，このような EbA 手法では高い精度が期待できない予測プロジェクトを「予測に適さないプロジェクト」と呼ぶ）。従来 EbA 手法を改善するために様々な研究が行われてきているが，予測手法そのものや，過去プロジェクトに焦点を当てた研究が大半であり，予測プロジェクトに焦点を当てた研究はほとんどない。

そこで本稿では，予測に適さないプロジェクトを判別する方法を提案する．提案手法では，各過去プロジェクトが予測プロジェクトを類似プロジェクトとして選択するかに基づいて，予測プロジェクトが高い精度での工数予測が期待できるか（適切かどうか）を判別する．本稿では，提案手法の有効性を調べるために，ソフトウェア開発企業で収集されたソフトウェア開発データを用いた評価実験についても報告する．評価実験では，提案手法による判別後に EbA 手法で予測した場合と提案手法で判別せずに EbA 手法で予測した場合を比較し，予測精度の改善の程度を観察した．

以降，2. では類似性に基づく工数予測手法について述べ，3. では提案手法について説明する．4. で実験方法，評価指標について述べた後，5. で実験の結果，結果に対する考察を述べる．最後に 6. で本稿のまとめを述べる．

2. Estimation by Analogy 手法

EbA 手法は，プロジェクトの特性を表すメトリクス同士が類似しているプロジェクトは工数も互いに似た値をとるであろうという仮定に基づいて予測を行う手法である．例えば，予測プロジェクトの開発規模が小さい場合には，過去プロジェクトの中から開発規模が小さいプロジェクトを選出し，その過去プロジェクトの実績工数を基に予測プロジェクトの工数を予測する．EbA 手法では表 1 の $x \times y$ 行列のように表されるデータセットを入力として用いる． P_i はプロジェクト i を， M_j はメトリクス j を， $V_{i,j}$ はプロジェクト i のメトリクス j の値を表す．

EbA 手法は 3 つの手順（正規化，類似度計算，予測値計算）から構成される．以下で各手順の詳細について述べる．

手順 1 メトリクスの正規化

メトリクスの正規化は，データセットに記録されている値はメトリクスによって値域が大きく異なるので，予測に与える影響を各メトリクスで均等にするために行う．メトリクス j の最大値を $\max(V_j)$ ，最小値を $\min(V_j)$ としたと

き，正規化したプロジェクト i のメトリクス j の値 $V'_{i,j}$ は式 (1) によって求められる．

$$V'_{i,j} = \frac{V_{i,j} - \min(V_j)}{\max(V_j) - \min(V_j)} \quad (1)$$

手順 2 類似度計算

類似度計算では，予測プロジェクトと各過去プロジェクト間の類似度を計算する．類似度計算には EbA 手法でよく用いられるユークリッド距離に基づいた類似度 [6] を採用した．ユークリッド距離の逆数を類似度として用いる．つまりユークリッド距離の値が小さいほど，類似度は大きくなる．プロジェクト P_a とプロジェクト P_i の類似度 $\text{sim}(P_a, P_i)$ は式 (2) で表される．

$$\text{sim}(P_a, P_i) = \frac{1}{\text{dist}(P_a, P_i)} \quad (2)$$

ここで， $\text{dist}(P_a, P_i)$ はプロジェクト P_a とプロジェクト P_i のユークリッド距離であり，式 (3) で求められる．

$$\text{dist}(P_a, P_i) = \sqrt{\sum_{j=1}^y (V'_{a,j} - V'_{i,j})^2} \quad (3)$$

手順 3 予測値計算

予測値計算では，類似度が高い過去プロジェクトの工数の実測値から，予測プロジェクトの工数の予測値を計算する．予測値計算には EbA 手法でよく用いられる，Weighted Sum [8] を用いた．工数のメトリクスを M_y とすると，予測プロジェクト P_a の工数の予測値 $V_{a,y}$ は式 (4) で求められる．

$$V_{a,y} = \frac{\sum_{i \in k\text{-nearestProjects}} M_{i,y} \times \text{sim}(P_a, P_i)}{\sum_{i \in k\text{-nearestProjects}} \text{sim}(P_a, P_i)} \quad (4)$$

ここで， $k\text{-nearestProjects}$ は，予測プロジェクト P_a と類似度の高い上位 k 個の過去プロジェクトの集合を表す． k の値は実験的に別途求める必要がある．

3. 提案手法

本稿では，各過去プロジェクトが予測プロジェクトを類似プロジェクトとして選択するかという観点により，EbA 手法では高い精度が期待できない予測プロジェクト（予測に適さないプロジェクト）を判別する方法を提案する．提案手法では，予測に適さないプロジェクトの判別基準として，各過去プロジェクトが予測プロジェクトを類似プロジェクトとして選択するかを用いる．具体的には，予測プロジェクトとある過去プロジェクトとの類似度が，その過去プロジェクトとそれ以外の各過去プロジェクトとの類似度と比較し，設定された閾値よりも高い順位の場合，その過去プロジェクトは予測プロジェクトを類似プロジェクトとして選択したとする．そして，全ての過去プロジェクトから類似プロジェクトとして選択されなかった予測プロジェクトを予測に適さないプロジェクトと判別する．

提案手法は以下の手順で行われる．

表2 実験に用いたデータセットの概要

メトリクス	平均値	中央値
開発期間(月)	11.3	10.0
開発チームの経験年数(年)	2.3	2.0
PMの経験年数(年)	2.6	3.0
トランザクション数	177.5	134.0
調整済みファンクションポイント	282.4	247.0
エンティティ数	120.5	96.0
開発総工数(人時)	4833.9	3542.0

手順1 予測プロジェクト P_a と過去プロジェクト $P_i (1 \leq i \leq n)$ の類似度 $\text{sim}(P_a, P_i)$ をユークリッド距離に基づく類似度により式(2)で算出する。

手順2 過去プロジェクト P_i と P_i 以外の過去プロジェクトとの類似度 $\text{sim}(P_i, P_1), \text{sim}(P_i, P_2), \dots, \text{sim}(P_i, P_n)$ を同様に式(2)で算出する。

手順3 手順1で求めた類似度 $\text{sim}(P_a, P_i)$ を、手順2で求めた各類似度 $\text{sim}(P_i, P_1), \text{sim}(P_i, P_2), \dots, \text{sim}(P_i, P_n)$ に含めて昇順でソートし、 $\text{sim}(P_a, P_i)$ の順位 R_i を算出する。

手順4 すべての過去プロジェクトに対して手順1～手順3を行い、 $R_1 \sim R_n$ を求める。

手順5 $R_1 \sim R_n$ 中に、設定した閾値 t 以上の順位が含まれている場合、予測プロジェクト P_a は適切な予測プロジェクトと判別する。一方、 $R_1 \sim R_n$ 中に閾値 t 以上の順位が含まれていない場合、予測プロジェクト P_a は予測に適さないプロジェクトと判別する。

4. 評価実験

4.1 実験目的

提案手法において、予測に適さないプロジェクトの判別の効果を調べるために、評価実験を行った。評価実験では、提案手法で予測に適切なプロジェクトと判別されたプロジェクトをEbA手法で予測した場合と、提案手法で判別せずにEbA手法で予測した場合を比較する。また、閾値 t の値を変化させ、閾値 t の予測精度への影響の変化も確かめた。

4.2 実験に用いたデータセット

実験に用いたデータセットは、Desharnais[4][7]によって収集されたカナダのソフトウェア開発企業における1980年代のデータである。元のデータセットは81件のプロジェクト、12種類のメトリクスが記録されており、5個の欠損値(記録されていない値)が含まれている。このデータから、EbA手法で扱うことができない欠損値や、 unnecessaryメトリクス(“調整済みファンクションポイント”に対する“未調整ファンクションポイント”と“調整係数”)を除去した、プロジェクト77件、メトリクス8種類のデータセットを実験に用いた。実験に用いたデータセットの各メトリクスの名称、平均値、中央値を表2に示す。

4.3 評価指標

本稿では、判別の効果を評価する指標として絶対誤差平

均値(MMAE)、相対誤差平均値(MMRE)、MER平均値(MMER)の3種類を用いた。各評価指標の概要と式を以下に示す。予測プロジェクト数を N 件、工数の実測値と予測値をそれぞれ $X_i, \hat{X}_i (i=1 \sim N)$ としている。

絶対誤差平均値(MMAE) 実測値と予測値の絶対誤差の平均値である。数値が小さいほど予測精度が高いことを示す。

$$MMAE = \frac{1}{N} \sum_{i=1}^N (|X_i - \hat{X}_i|) \quad (5)$$

相対誤差平均値(MMRE) 実測値と絶対誤差の比の平均値である。実測値と比べて誤差がどの程度であるかを表す。数値が小さいほど予測精度が高いことを示す。

$$MMRE = \frac{1}{N} \sum_{i=1}^N \left(\frac{|X_i - \hat{X}_i|}{X_i} \right) \quad (6)$$

MER平均値(MMER) 予測値と絶対誤差の比を表す指標であるMER[5]の平均値をとった値である。MERは予測値と比べて誤差がどの程度の大きさであるかを評価する指標であり、MMREでは値が1以下にしかない、過小予測時に、値が大きくなる。

$$MMER = \frac{1}{N} \sum_{i=1}^N \left(\frac{|X_i - \hat{X}_i|}{\hat{X}_i} \right) \quad (7)$$

4.4 実験手順

提案手法の予測に適さないプロジェクトの判別の効果を調べるために、以下の手順で実験を行った。

(1) データセットに含まれる77件のプロジェクトを、10回無作為に二分割し、38件の過去プロジェクト(フィットデータ)と、39件の予測プロジェクト(テストデータ)を作成する。

(2) (1)で作成した各フィットデータとテストデータのペアにおいて、テストデータの各プロジェクトについて、フィットデータを使って提案手法により判別する。このとき、閾値 t を1～5まで変化させて判別する。

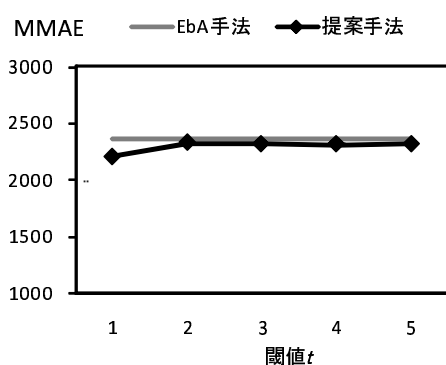
(3) (2)で予測に適切なプロジェクトと判別されたプロジェクトについて、EbA手法を用いて工数予測を行い、予測値と実測値の誤差を計算し、各評価指標を求める。

(4) (1)で作成したフィットデータとテストデータのペアに対して、提案手法で判別せず全予測プロジェクトについてEbA手法を用いて工数予測を行い、各評価指標を求め、提案手法で判別した場合との比較を行う。

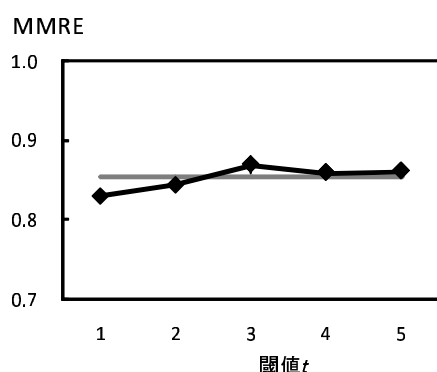
なお、両者ともEbA手法の k の値は、予備実験の結果 $k=4$ を採用した。

5. 結果と考察

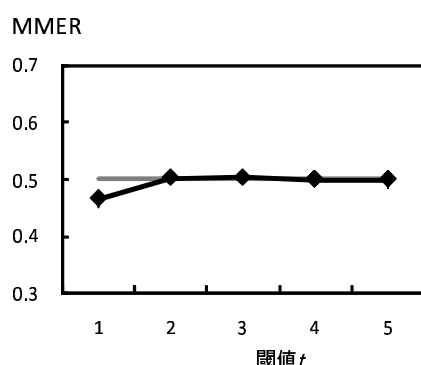
提案手法で判別された予測に適切なプロジェクトについてEbA手法で工数予測した場合と、判別せずにEbA手法で予測した場合の、各評価指標の値を図1(a)、図1(b)、図1(c)に示す。各グラフの横軸は閾値 t 、縦軸は各評価指標の値を表す。



(a) 絶対誤差平均



(b) 相対誤差平均



(c) MER 平均

図1 各評価指標

MMAE (図 1(a)) の結果を見ると、閾値 t が 1 の時、提案手法では EbA 手法よりも、MMAE が 157 小さくなっており、提案手法により予測精度が向上している。一方、閾値 t が 2~5 においては、ほとんど差はみられない。

MMRE (図 1(b)) の結果を見ると、閾値 t が 1 および 2 の時、提案手法では EbA 手法よりも MMRE がそれぞれ、0.025、および、0.014 小さくなっており、提案手法により

表 3 判別の予測精度への効果

	MMAE	MMRE	MMER
提案手法 (閾値 $t=1$)	2205.3	0.83	0.47
EbA 手法	2362.3	0.85	0.50
検証結果 (閾値 $t=1$)	2553.3	0.92	0.57

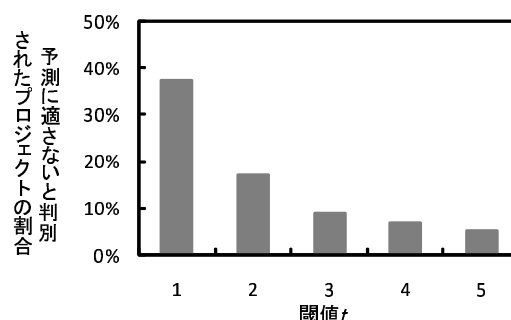


図2 予測に適さないと判別されたプロジェクトの割合

予測精度が向上している。一方、閾値 t が 3~5 においては、EbA 手法よりも提案手法のほうが予測精度が悪くなっている。

MMER (図 1(c)) の結果を見ると、閾値 t が 1 の時、提案手法では EbA 手法よりも MMER が 0.03 小さくなっており、提案手法により予測精度が向上している。一方、閾値 t が 2~5 においては、ほとんど差はみられない。

実験の結果、閾値 t が 1 の時、MMRE、MMER とともに予測精度が向上していることから、EbA 手法では過小予測してしまうプロジェクト、および、過大予測してしまうプロジェクトとも提案手法によって予測に適切でないプロジェクトとして判別されていると言える。一方で、閾値 t が 2 以上の時は提案手法で判別してもほとんど効果がなく、逆に MMRE では判別精度が悪化した場合もあった。従って、予測プロジェクトを類似プロジェクトとするような過去プロジェクトが 1 つでもある場合には、EbA 手法によって高い精度で予測できる可能性が高いと言える。

閾値 t が 1 の時に提案手法で予測に適切でないプロジェクトと判別されたプロジェクトが本当に予測に適していないプロジェクトであることを確かめるため、予測に適切でないと判別されたプロジェクトに対して EbA 手法を用いて工数予測を行った (検証結果)。その結果を表 3 に示す。比較対象として提案手法と EbA 手法の結果も示している。

表 3 より、予測に適切でないと判別したプロジェクトの予測精度は、提案手法により改善した精度以上に EbA 手法よりも悪化している。また、MMRE、MMER とともに悪化している。従って、提案手法では、実際に EbA では高い精度で予測できないプロジェクトを予測に適切でないプロジェクトと判別できていると言える。

図 2 に、閾値 t ごとの提案手法で予測に適切でないプロジェクトと判別されたプロジェクトの割合 (10 回の平均値) を示す。グラフの横軸は閾値 t を、縦軸は判別されたプロ

プロジェクトの割合（母数 39 件）を示す。

図 2 より、提案手法は、予測精度が向上した閾値 t が 1 の時、37.1%（39 件のうち約 14.5 プロジェクト）と非常に多くのプロジェクトを予測に適さないプロジェクトと判別している。また、閾値 t と予測に適さないと判別されたプロジェクトの割合はほぼ反比例している。提案手法で予測に適さないと判別したプロジェクトは、EbA 手法では、高い精度での予測が期待できないプロジェクトではあるが、EbA 手法以外の工数予測手法を用いて工数予測することで高い精度で工数予測できる可能性がある。従って、これらの予測に適さないと判別されたプロジェクトについても、予測しないのではなく、何らかの方法で高い精度で予測することを今後検討する必要がある。

本稿の提案した手法は、まだ簡易な手法であり、評価実験において予測精度の改善はそれほど大きくは見られなかった。提案手法を改良することで更なる予測精度の改善が見込めると考えられる。例えば、互いに非常に類似したプロジェクトが固まっていると、それらのプロジェクトでは、ある程度類似している予測プロジェクトであっても、類似プロジェクトとは選択しないため、その予測プロジェクトは予測に適さないプロジェクトと判別される可能性が高い。一方、過去プロジェクトの中でも特異なプロジェクトで、他の過去プロジェクトとの類似度が低いプロジェクトは、あまり類似していない予測プロジェクトであっても類似プロジェクトとして選択してしまい、過去プロジェクトとあまり類似していない予測プロジェクトが予測に適切なプロジェクトと判別される可能性がある。このような問題は、閾値を順序だけでなく、類似度の値もある程度考慮することで軽減することが可能であると考えられるため、閾値について更なる改良が必要であると言える。

6. ま と め

本稿では、ソフトウェア開発プロジェクトにおける工数予測手法の 1 つである類似性に基づく工数予測（EbA）手法において、高い精度での工数予測が期待できない予測プロジェクト（予測に適さないプロジェクト）を判別する方法を提案した。ソフトウェア開発企業で収集されたデータセットを用いた評価実験の結果、判別手法を用いることで、判別手法を用いない場合よりも MMAE、MMRE、MMER の全ての評価指標において予測精度の向上がみられ、予測に適さないプロジェクトを判別できていることを確認した。

今後の課題としては、提案手法を改良することで更なる予測精度の改善を目指すこと、EbA 手法で予測に適さないと判別したプロジェクトを、EbA 手法以外の工数予測手法で予測することを検討することが挙げられる。

謝辞

本研究で利用した予測エンジンの作者である大杉直樹氏、角田雅照氏に、深く感謝致します。本研究の一部は、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。また、文部科学省科学研究費補助金基盤

研究 (C)(課題番号:20500033) の研究助成を受けて行われた。

文 献

- [1] A. Albrecht, and J. Gaffney, "Software function, source lines of code, and development effort prediction," IEEE Transactions on Software Engineering, vol.9, no.6, pp.83-92, 1979.
- [2] B.W. Boehm, Software Engineering Economics, Prentice Hall, New Jersey, 1981.
- [3] S.D.Conte, H. E.Dunsmore, and V.Y. Shen, "Software engineering metrics and models," The Benjamin/Cummings Publishing Company, Inc., 1986.
- [4] J.M. Desharnais, "Analyse statistique de la productivite des projets informatique a partie de la technique des point des fonction," Masters Thesis, University of Montreal, 1989.
- [5] T. Foss, E. Stensrud, B. Kitchenham, and I. Myrtveit, "A simulation study of the model evaluation criterion MMRE," IEEE Transactions on Software Engineering, vol.29, no.11, pp.985-995, 2003.
- [6] C. Kirsopp, E. Mendes, R. Premraj, and M.J. Shepperd, "An empirical analysis of linear adaptation techniques for case-based prediction," Proceedings of 5th International Conference on Case-Based Reasoning(ICCBR2003), pp.231-245, 2003.
- [7] C. Mair, G. Kadoda, M. Lefley, K. Phalp, C. Schofield, M. Shepperd, and S. Webster, "An investigation of machine learning based prediction systems," Journal of Systems and Software, Vol.53, Issue 1, pp.23-29, 2000.
- [8] B.M. Sarwar, G. Karypis, J.A. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," In Proceedings of 10th International World Wide Web Conference, pp.285-295, 2001.
- [9] M. Shepperd, and C. Schofield, "Estimating software project effort using analogies," IEEE Transactions on Software Engineering, vol.23, no.12, pp.736-743, 1997.
- [10] 角田雅照, 大杉直樹, 門田暁人, 松本健一, 佐藤慎一, "協調フィルタリングを用いたソフトウェア開発工数予測方法," 情報処理学会論文誌, vol.46, no.5, pp.1156-1164, 2005.