

生産性の比に基づいた類似プロジェクト選択手法

天 嵩 聡 介^{†1} 高 原 洋 平^{†1}

工数見積もり手法の一つに類推法がある。類推法では多次元の特性上のユークリッド距離を用いて類似プロジェクトを選択する。しかし、この方法では使用する特性の選択・重み付けにかかる計算量が大きい。本論文では生産性の比と RVM を用いた類似プロジェクト選択法を提案する。ワークショップでは本手法の有用性について議論したい。

Selecting Projects Based on Ratio of Productivity for Effort Estimation

SOUSUKE AMASAKI^{†1} and YOHEI TAKAHARA^{†1}

Analogy-based effort estimation (ABE) is one of the most popular methods. ABE selects similar projects based on Euclidean distance on multiple features. However, feature selection and weighting for ABE optimization is computationally intensive. This paper demonstrated a selection method using Relevance Vector Machine (RVM) and a ratio of productivity. We would like to discuss its validity at the workshop.

1. はじめに

工数見積もり手法の研究分野において、類推法¹⁾に関する研究が盛んである。類推法はプロジェクト特性を多次元ベクトルと見なし、ユークリッド距離によって類似したプロジェクトを選択する。そして選択されたプロジェクトを見積もりの根拠とする手法である。しかし、工数見積もりに用いる類似プロジェクト数や個々の特性の重み付け、見積もり工数の計算方法など、多くの組み合わせを試行して高い見積もり精度を示す組み合わせを発見しなければならない²⁾。

本論文では、個々のプロジェクト間の直接的な類似尺度に基づいて類似プロジェクトを選択する手法を提案する。提案法では Relevance Vector Machine (RVM) を利用することで効率的かつ効果的な類似プロジェクトの選択を行う。

2. 提案手法

提案手法ではプロジェクト間の類似尺度を生産性の比で定義する。この定義に従い過去のプロジェクトデータの組み合わせに対して類似度を計算する。そして、組み合わせたプロジェクトの特性をから類似度を予測するモデルを RVM で構築する。以下に詳細を

示す。

2.1 生産性の比による類似度の定義

本論文では生産性を以下に定義する。

$$\text{生産性} = \frac{\text{工数}}{\text{規模}}$$

そして、プロジェクト A, B の類似尺度として生産性の比の対数を定義する。

$$\text{非類似度}_{A,B} = \log \left(\frac{\text{生産性 A}}{\text{生産性 B}} \right)$$

生産性の比は常に正の値を取り分布に偏りがあるため対数化した。

2.2 RVM の学習

任意の 2 つの過去プロジェクトデータの組み合わせに対して生産性の比を計算して学習モデルの目的変数とする。また、双方のプロジェクトの特性を説明変数とする。この際、一方のプロジェクト特性には見積もり時に不明な特性 (工数など) を含めないようにする。このデータを RVM に適用して予測モデルを得る。今回は RVM にラプラスカーネルを使用した。

2.3 類似プロジェクトの選択

予測モデルに過去のプロジェクトと見積もり対象プロジェクトの特性を組み合わせたものを与えて類似度を予測する。そして類似度が高い過去のプロジェクト上位 n 件を類似プロジェクトとする。

3. 実 験

文献 3) と同じ条件で提案法の見積もり精度を評価

^{†1} 岡山県立大学

Okayama Prefectural University

表 1 実験結果

Models	MMRE	MdMRE	PRED(25)	MMAE	MdMAE	SA
X	0.29	0.34	0.33	1350	580	81
Y	0.32	0.16	0.58	961	702	86
提案法 ($n = 10$)	0.21	0.08	0.83	470	226	93

した。この実験では、Maxwell データセットのうち、1991 年以前の 50 件を過去プロジェクトとし、残り 12 件を見積もり対象としている。そして、以下の工数見積もりモデル X, Y の見積もり精度を評価している。

$$X: \log(\text{Effort}) = \log(\text{Size}) + T09 + \text{Time} + \text{Ifc}$$

$$Y: \log(\text{Effort}) = \log(\text{Size}) + T08 + T09 + T14 + \text{Ifc}$$

本研究ではモデル X, Y の見積もりと類似プロジェクト n 件を使った見積もりの精度を比較した。以下の方法で見積もりを行った。

$$\text{見積もり工数} = \frac{1}{n} \sum_{i=1}^n \text{生産性}_i \times \text{対象の規模}$$

今回のモデルでは予測値を類似プロジェクトの生産性の補正に利用できる。そこでこの結果とも比較した。

過去プロジェクトは 50 件であるので、単純に組み合わせを作成すると $50 \times 49/2 = 1225$ となる。しかし、今回は 2 つのプロジェクトの開始日の前後関係が守られるようプロジェクトを組み合わせため、合計 1218 件のデータを用意した。

評価基準としては、MMRE, MdMRE, PRED(25), MMAE, MdMAE, そして SA⁴⁾ を用いた。

4. 結 果

表 1 に実験結果を示す。PRED(25) 及び SA は値が大きいくほど性能が良い。他は値が小さいほど性能が良い。まず、モデル X と Y には大きな性能差がある。文献 3) では全てのプロジェクトデータを用いて変数を選択しているが、プロジェクト特性の選択が見積もり精度に影響を及ぼすことがわかる。

提案手法は $n = 10$ のときモデル X, Y より高い性能を示した。変数選択を明示的に行わずに十分な性能の工数見積もりモデルが構築できたと言える。

図 1 に選択プロジェクト数と性能の関係を示す。白丸が見積もり工数である。選択プロジェクト数が $n = 43$ までモデル Y と同程度以上の性能を保持している。選択プロジェクト数が増えると性能は低下しているが、性能の変化の傾向はほぼ連続的であり、類似プロジェクトの選択が適切に行われていると考えられる。

黒丸は補正見積もり工数である。生産性の比を正確に予測しているため、補正を行うと選択プロジェクト数に関わらず高い精度を維持できることがわかる。

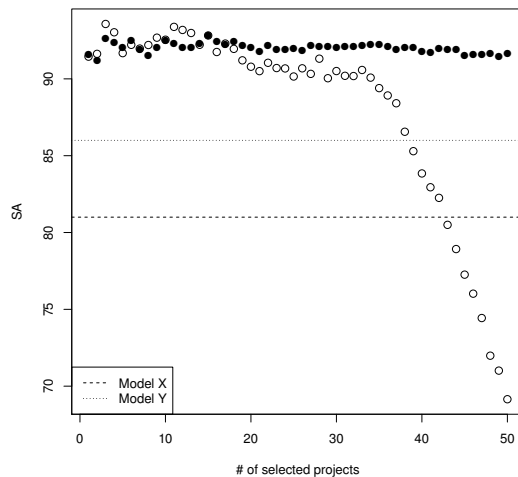


図 1 選択したプロジェクト数と性能の関係

5. おわりに

実験により、提案手法の有効性が一部確認できた。今後は類推法との比較を行い、提案手法の有効性を検証していきたい。また、本手法の有効性を他のデータセットにおいても検証していきたい。

参 考 文 献

- 1) Shepperd, M. and Schofield, C.: Estimating Software Project Effort Using Analogies, *Software Engineering, IEEE Transactions on*, Vol.23, No.11, pp.736-743 (1997).
- 2) Kocaguneli, E., Menzies, T., Bener, A. and Keung, J. W.: Exploiting the Essential Assumptions of Analogy-Based Effort Estimation, *Software Engineering, IEEE Transactions on*, Vol.38, No.2, pp.425-438 (2012).
- 3) Maxwell, K.D.: *Applied Statistics for Software Managers*, Prentice Hall (2002).
- 4) Shepperd, M. and MacDonell, S.: Evaluating prediction systems in software project estimation, *Information and Software Technology*, Vol.54, No.8, pp.820-827 (2012).