

## 工数見積もりモデルの予測精度を低下させるプロジェクトの特徴分析

戸田航史†

ソフトウェア開発プロジェクトの計画立案のためには、開発工数の予測が必要となる。そのために、重回帰モデルなどの見積もりモデルが多数提案されている。本研究では、予測が外れやすい、または、ばらつきやすいプロジェクトの特徴を実験的に明らかにする。

## Characterizing Software Projects that Degrade Effort Estimation Performance

Koji Toda†

Effort prediction is necessary for project managers to make and control a software development plan. To predict development effort, various estimation models such as multivariate regression models have been proposed. This paper analyzes project characteristics whose prediction accuracy tends to become worse.

### 1. はじめに

ソフトウェア開発プロジェクトを成功に導くためには、開発工数の予測を定期的に行い、作業済み工数(実工数)との乖離を随時把握することが重要となる。そのために、従来、重回帰モデル、ニューラルネット、事例ベース推論 (CBR) などの数多くの工数見積もりモデルが提案されてきた。

しかし、それらのモデルによる工数の予測精度には、プロジェクトごとに大きなばらつきがある。一つの実例として、ステップワイズ対数重回帰モデルを用いて 68 件のプロジェクトの試験工数の予測を試みた。その結果を図 1 に示す。工数予測の相対誤差が 30%以内の「予測がほぼ的中した」プロジェクトが多数存在する一方で 100%を超える「予測に失敗した」プロジェクトも少なからず存在した[2]。このように、たとえ確率が低くとも予測が大きく外れる可能性がある以上、予測誤差の平均値がいくら小さくとも、プロジェクト開発の現場では予測値を採用することは難しい。

本稿の主張は、予測モデルを現場で活用するためには、平均的に予測精度を向上させるのではなく、予測が外れるやすい(もしくは、外れにくい)プロジェクトの特徴を明らかにすることが必要ではないか、ということである。予測が外れやすい(特異点となるような)プロジェクトは、モデル構築に最初から用いないことが望ましいと考えられる。

我々はこれまでに、予測が外れやすい、もしくは、予測精度がばらつきやすいプロジェクトの特徴を実験的に明らかにすることを目的として研究を行ってきた[2]。実験では、多数のプロジェクトについて工数予測を行い、各プロジェクトの予測誤差を求めた。そして、様々なプロジェクト特性値に基づいてプロジェクトをグループ分けし、t 検定、及び、アソシエーション分析を用いて誤差が平均的に大きいグループを調査した。

### 2. 検証実験について

#### 2.1. データセット概要

本研究で用いたデータセットは International Software Benchmark Standard Group (ISBSG)が収集した 20 カ国のソフトウェア開発企業の実績データである[1]。データセットには 1989 年から 2004 年までの 3026 件のプロジェクトについて、それぞれ 99 種類のプロジェクト特性が含まれている。ただしこのデータセットには数多くの欠損値が含まれているため、実験ではそのサブ

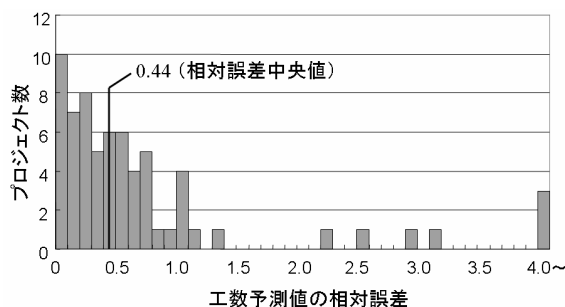


図 1. 工数予測値の相対誤差の分布の例

†奈良先端科学技術大学院大学 情報科学研究科  
Graduate School of Information Science, Nara Institute of  
Science and Technology

表 1. 実験に用いたプロジェクト特性の一部

特性値名	説明
Development Type	New Development or Enhancement
Adjusted Function Points	数量データ
Architecture	Client Server, Stand Alone or Multi-tier
Primary Programming Language	COBOL, Java, PL/I, TELON or Others
Recording Method	Productive Time Only, Stuff Hours or Others
Resource Level	Level 1~4
Effort Plan	数量データ
Effort Specify	数量データ
Effort Build	数量データ
Effort Test	数量データ

セットを用いた。実験に用いたデータセットに含まれるプロジェクト特性の一部を表 1 に示す。

## 2.2. 実験手順

実験用データセットを用いて工数予測を行い、各プロジェクトの予測値の相対誤差を算出した。

プロジェクト特性を表す変数のうち、数値変数(比尺度, 間隔尺度)については、中央値を閾値とした2つのカテゴリを持つ名義尺度へと変換した。そして、カテゴリ間で相対誤差の平均値に有意差のある変数を、t 検定を用いて特定した。

アソシエーション分析では、「相対誤差が大」、もしくは、「相対誤差が小」というカテゴリを結論部を持つアソシエーションルールの抽出を試みた。

## 2.3. 実験結果

t 検定を用いた分析では Primary Programming Language の COBOL と Others, Recording Method の Productive Time Only と Others, Stuff Hours Only と Others, Resource Level の Level 2 と Level 4 の間で相対誤差平均値に有意な差が見られた。

一方、アソシエーション分析では十分に支持度の高いルールは得られなかった。

## 2.4. 考察

t 検定の結果からは、COBOL 言語を用いたプロジェクトは予測が外れにくい傾向が分かった。COBOL 以外のプロジェクトをモデル構築から予め除外しておくことが望ましい可能性がある(ただし、COBOL を用いたプ

ロジェクトを予測したい場合に限る)。

また、Recording Method と Resource Level によっては予測が外れやすことが分かった。前者はプロジェクトに含める作業時間の記録方法の定義であり、後者はプロジェクトに含める作業の種類定義である。予測精度を向上させるためには、データ記録方法(定義)を統一することが望ましいことが示唆された。

アソシエーション分析において有用なルールが得られなかったこと理由として、数値変数からカテゴリ変数への変換方法に問題があった可能性がある。実験では数値変数の中央値を閾値として大小の二群に分けることでカテゴリ変数へと変換したが、この閾値の設定やカテゴリ数が適切でない、すなわち大小の二群ではなく、より適切な群に分割し、カテゴリ変数化できれば、有用な(支持度の高い)ルールが得られた可能性がある。しかしどのようにしてプロジェクトを適切な群に分割すべきか、という点については今後の課題である。

## 3. 議論したい問題

これまで、工数や信頼性の予測を目的とした研究が多数行われてきたが、予測精度を従来モデルより向上させることばかりが目され、実際に開発現場へ適用する際の問題については触れられないことが多かった。たとえ予測精度が従来モデルよりも向上したとしても、平均相対誤差が50%を超えるようでは現場で用いることはできないし、平均相対誤差が10%だったとしても、数回に1回の割合で予測が大きくはずれる(例えば、相対誤差が100%を超える)場合には、やはり現場で用いるのは危険である。

ワークショップでは、ソフトウェア工学分野における従来の研究成果を現場で活用するための様々なアプローチについて議論したい。また、アソシエーション分析のように、他分野で用いられている技術をソフトウェア工学分野に取り入れる際の問題についても議論したい。

## 参考文献

- [1] International Software Benchmarking Standards Group (ISBSG), "ISBSG Estimating, Benchmarking and Research Suite Release 9," ISBRG, 2004.
- [2] 戸田 航史, 角田 雅照, 門田 暁人, 松本健一, "工数見積もりモデルで予測できないソフトウェアプロジェクトの特徴分析, SIGSS 2005-74, pp.67-72, December 2005.