
上流工程比率を用いたソフトウェア開発工数見積もり精度の定量的評価

Evaluation of software development effort estimation based on upper phase ratio

角田 雅照^{*} 戸田 航史[†] 伏田 享平[‡]

あらまし 本研究では要件定義などの上流工程の工数とその工程比率に基づいた工数見積もりに着目し、その見積り精度を評価する。分析ではISBSG データセットを用いて上流工程工数を用いたモデル、開発規模を用いたモデル、上流工程工数と開発規模の両方を用いたモデルを構築し、見積もり精度を比較した（工程比率はモデル構築時にパラメータとして推定される）。分析の結果、計画工数と開発規模の両方を用いることにより BRE (Balanced Relative Error) 平均値が 46.8% 改善することなどがわかった。

1 はじめに

近年、ソフトウェアはより大規模化しており、それにともないソフトウェア開発にはより多くの工数（コスト）が必要となっている。このような大規模なソフトウェア開発においては、スケジュールやコストに関するプロジェクト管理は必要不可欠であり、ソフトウェア開発工数見積もりは、その基礎となるものである。よって、工数見積もりには高い見積もり精度（見積もり工数と実績工数との差が小さいこと）が求められる。主要な工数見積もり方法の一つとして、数学的モデルによる工数見積もりがあり、過去に蓄積された開発データに基づき、工数見積もりモデルを構築する。これまで、高精度な見積もりを実現するために多数のモデルが提案されている。

多くの工数見積もりモデルの研究では、ファンクションポイントなどのソフトウェアの規模（開発規模）を主要な説明変数（見積もり時に既知の変数）として開発工数を見積もっている。一方開発現場では、上流工程の工数と、その工程が開発工程全体に占める割合に基づいて、工数を簡易的に見積もる場合がある[4][14]。例えば、要件定義の開発工数全体に占める割合が過去のプロジェクトの平均で 20% であり、ある開発中のプロジェクトの要件定義の工数が 2 人月である場合、総開発工数を 10 人月と見積もる。

本研究では、これまでの研究においてあまり重視されていなかった、要件定義などの上流工程の工数とその工程比率（当該工程にかけられた工数が開発工数全体に占める割合）に基づいた工数見積もりに着目し、それらに基づく工数見積もりモデルの精度を評価する。Yang ら[17]は開発工程の分布に関する分析を行った研究は非常に少ないことを

^{*}Masateru Tsunoda, 東洋大学 総合情報学部

[†]Koji Toda, 福岡工業大学 情報工学科

[‡]Kyohei Fushida, 株式会社 NTT データ 技術開発本部

指摘するとともに、上流工程（計画・要件定義、設計）の工程比率の分散はあまり大きくないことを示している。このことから、上流工程にかけられた工数（以下、上流工程工数と呼ぶ）を説明変数として重回帰分析に基づく工数見積もりモデルを構築した場合（工程比率はモデル構築時にパラメータとして推定されるため、説明変数に含まれない）、比較的高い見積もり精度となることが期待される（理由は3章で詳説する）。

我々の知る限り、開発規模を用いたモデルと上流工程工数を用いたモデルの見積もり精度を比較した研究は存在せず、従って、開発現場において見積もり精度の高いモデルを構築するためには、どちらの変数を使うべきかというガイドラインは存在しない。精度の差異を明らかにすることにより、そのガイドラインを示すことが本研究の目的である。のために、以下の3つのリサーチ・クエスチョンを設定した。

- RQ1: 開発規模のみ、もしくは上流工程工数のみで工数を見積もった場合、どちらの見積もり精度が高いのか。
- RQ2: 開発規模、上流工程工数に加え、その他のデータ項目（開発言語など）を説明変数に加えた場合、開発規模を含むモデルと上流工程工数を含むモデルでは、どちらの見積もり精度が高いのか。
- RQ3: 開発規模と上流工程工数、両方を用いてモデルを構築した場合、見積もり精度はより高まるのか。そのときに多重共線性が発生しないのか（発生する場合、不安定なモデルとなるため開発規模と上流工程工数の両方を用いるべきではない）。

プロジェクトのデータ収集はコストが掛かるため、組織によってはデータを詳細に収集、蓄積していない場合がある。そこで、開発規模または上流工程工数のみを説明変数として用いることを想定した、RQ1を設定した。RQ1にその他の説明変数（開発言語など）を加えた場合、結論が変化するのかを確かめるため、RQ2を設定した。RQ3はRQ1とRQ2両方の場合について確かめる。

2 分析に用いたデータセット

分析に用いたデータセットは、ISBSG（International Software Benchmarking Standards Group）が20ヶ国の中から収集したものである[3]。用いたデータセットはRelease 9と呼ばれるバージョンであり、1989年から2004年に実施されたソフトウェア開発プロジェクトが3026件、変数が99個含まれている。データには欠損値（値が記録されていないこと）が含まれている。

分析対象のプロジェクトの条件を整えるため、Locanら[9]が示した条件（ファンクションポイント（FP）がIFPUG法で計測されているなど）を満たしているプロジェクトを抽出した。さらにそれから、計画工数、要件定義工数、設計工数、コーディング工数、テスト工数、インプリメント工数（ソフトウェアのリリース、インストールなどの工数）の合計と、総工数の誤差が10%以内のプロジェクト173件を抽出し、分析対象とした。誤差は、各工程工数合計 < 総工数の場合、 $1 - \text{各工程工数合計} / \text{総工数}$ 、それ以外の場合、 $1 - \text{総工数} / \text{各工程工数合計}$ とした。完全に誤差がないデータだけを抽出するとデータ数が減少するため、10%以内の誤差があるプロジェクトを含めて分析を行った。

分析に用いた変数を表1に示す。上流工程工数、工程比率、工期、開発速度以外の変数についてはLocanら[9]と同じ変数を用いた。なお、設計工数が記録されたプロジェク

Evaluation of software development effort estimation based on upper phase ratio

表 1 分析に用いた変数

変数名	尺度	詳細
FP	比例	未調整 FP（調整係数の導出には問題点があり、かつ用いても精度が改善されない場合が多いため、未調整 FP の利用が好まれる[9]）
総工数	比例	総開発工数
生産性	比例	総工数 / FP
計画工数	比例	予備調査、プロジェクト計画作成などにかかった工数
要件定義工数	比例	要求分析、要件定義、アーキテクチャ定義などにかかった工数
計画比率	比例	計画工数 / 総工数
要件定義比率	比例	要件定義工数 / 総工数
工期	比例	全開発期間 - プロジェクト停止期間
開発速度	比例	FP/工期
開発種別	名義	新規開発、改修・保守開発、再開発
開発プラットフォーム	名義	MF（メインフレーム）、PC（パソコン）、MR（ミッドレンジ）
開発言語	名義	2GL, 3GL, 4GL (n GL は第 n 世代言語を示す)、自動生成コード

トは 173 件中 4 件のみであった。ISBSG の定義では、設計工程は外部設計、内部設計などの工数であるが、データ提供組織がこれらの工数を要件定義工数または次工程（コーディング）の工数に含めた可能性がある。すなわち、要件定義工数は基本設計にかけられた工数を含んでいる可能性が高く、場合によっては詳細設計にかけられた工数まで含んでいる可能性もある。

3 予備分析

数学的モデルによる工数見積もりでは、自社内や企業横断的に収集されたデータを用いて見積もりモデルを構築する[9]。モデル構築方法として対数（重）回帰分析がよく用いられ[13]、その場合の説明変数を 1 つだけ用いた最も単純なモデルは以下の式となる。

$$y = a x^b \quad (1)$$

ここで y は見積もり工数を表す。 b は定数であり、一般に 1 に近い値となる。開発規模を説明変数として用いる場合、 x は開発規模であるため、 a は生産性の逆数と見なせる。上流工程工数を用いる場合、 x は上流工程工数であるため、 a は上流工程の工程比率の逆数と見なせる。説明変数が 1 つだけの場合、 a は定数となるため、 a の分散（プロジェクトごとの差異）が大きいと、モデルの当てはまりが悪くなり、見積もり精度が低下する。

そこで Yang らの分析[17]と同様に、計画比率と要件定義比率の基本統計量を算出するとともに、生産性の基本統計量を算出し、それぞれを比較する。上流工程の工程比率の分散が小さければ、上流工程工数のみで工数を見積もった場合の精度が高いと期待される。RQ1 に答えるための分析（開発規模のみ、もしくは上流工程工数のみで工数を見積もり、精度を比較すること）の妥当性を高めるためには、見積もり精度の差異が発生する原因を把握する必要があり、これはそのための分析である。

また、上流工程の工程比率と関連の強い変数は、説明変数の候補として採用する（モデル構築のガイドラインに、それらを必要性の高い説明変数として記述する）必要がある。単变量解析（分散分析の寄与率と Spearman の順位相関係数）を適用し、工程比率と他の変数との関連を明らかにする。さらに説明変数には含めない生産性などの関連も確かめた。なお、開発プラットフォームや開発言語が上流工程の「工数」と強い関

表 2 計画比率, 要件定義比率, 生産性の基本統計量

	平均値	最小値	Q_1	中央値	Q_3	最大値	Q_3 / Q_1	最大値 / 最小値
計画比率	0.090	0.005	0.039	0.071	0.124	0.434	3.2	86.8
要件定義比率	0.136	0.004	0.065	0.121	0.193	0.434	3.0	108.5
生産性	0.197	0.003	0.057	0.127	0.288	1.044	5.1	348.0

連がなくとも、上流工程の「工程比率」とは関連が見られる場合がある。例えば、同規模の開発を行う際、ある開発言語はその他の開発言語よりも下流工程工数が多くなりやすい場合、上流工程工数が変化しなければ、上流工程の工程比率は低下する。同様の理由で、開発種別と上流工程の「工数」との関連が強い場合でも、上流工程の「工程比率」との関連が弱いこともありうる。

3.1 計画比率と要件定義比率の基本統計量

表 2に計画比率、要件定義比率、生産性の基本統計量を示す（Yang らの分析[17]では生産性と工程比率は比較していない）。計画比率と要件定義比率の第 1 四分位数 (Q_1) と第 3 四分位数 (Q_3) の比、及び最大値と最小値の比（表中太字）は、計画比率と要件定義比率に比べて生産性のほうが大きかった。すなわち、計画比率、要件定義比率よりも、生産性のほうがプロジェクト間の差異が大きいといえる。これは、RQ1 の答えが「上流工程工数のみで工数を見積もった場合、開発規模のみで工数を見積もった場合よりも精度が高くなる」となることを示唆している。

3.2 工程比率と他変数との関連

まず、名義尺度の変数と工程比率との関係について、分散分析の調整済み寄与率 ω^2 [16]を用いて分析した。 ω^2 は名義尺度の変数と比例尺度の変数の関連の強さを分析する際に用いられ、おおむね 0 から 1 の値をとり、値が大きいほど関連が強いことを示す。工程比率と名義尺度変数との寄与率と、分散分析の p 値を表 3 に示す。*は有意水準 5%で関連があったものを表す。開発プラットフォームは要件定義比率と有意な関連があったが、あまり強い関連ではなかった。その他の変数と工程比率は関連が見られなかった。

次に、比例尺度の変数と工程比率との関係について、Spearman の順位相関係数を用いて分析した。ソフトウェア開発のデータセットでは値の大きいケース（開発規模の大きなケースなど）が含まれるなど、正規分布を仮定できないことが多く、Pearson の積率相関係数では正しく関係を分析できない場合があるため用いていない。工程比率と名義尺度変数との相関係数と p 値を表 4 に示す。*は有意水準 5%で関連があったものを表す。工期と計画比率、FP と要件定義比率に有意な関連が見られたが、それぞれあまり強い関連ではなかった。その他の変数と工程比率には関連が見られなかった。

データセットに含まれる変数に対して単変量解析を適用した限りにおいては、工程比

表 3 工程比率と名義尺度変数との
関連

変数		開発種別	開発プラットフォーム	開発言語
計画比率	寄与率	0.01	0.03	-0.01
	p 值	0.18	0.10	0.52
要件定義比率	寄与率	0.00	0.15	-0.01
	p 値	0.50	0.00*	0.82

表 4 工程比率と比例尺度変数との関連

		FP	総工数	生産性	工期	開発速度
計画比率	相関係数	-0.10	-0.17	0.12	-0.28	0.01
	p 値	0.25	0.05	0.18	0.00*	0.88
要件定義比率	相関係数	-0.18	-0.15	0.02	-0.05	-0.13
	p 値	0.04*	0.08	0.83	0.64	0.19

率との関連が強い変数ではなく、上流工程工数を用いた見積りモデルの説明変数の候補に必ず含めるべき（モデル構築のガイドラインに、必要性の高い説明変数として記述すべき）変数は見つからなかった。これらの変数は上流工程工数を用いた見積りモデルの精度向上に大きく寄与しないと考えられるが、精度改善に全く役立たないとも限らない。そこで次章の分析では、これらの変数を説明変数の候補として用いることとする。

4 上流工程工数を用いた工数見積り

4.1 実験概要

上流工程工数を用いた工数見積もりの精度を評価し、RQ1, RQ3 に答えるために、以下の変数を説明変数として用いた工数見積もりモデルを構築した。

- 計画工数
- 計画・要件定義工数
- FP
- 計画工数, FP
- 計画・要件定義工数, FP

ここで計画・要件定義工数とは計画工数と要件定義工数の和であり、新たに定義した変数である。FP, 計画工数, 要件定義工数が欠損していない 118 件のプロジェクトを用い、各変数を対数変換してからモデルを構築した。

さらに RQ2, RQ3 に答るために、上記の変数に開発種別、開発プラットフォーム、開発言語をダミー変数化して説明変数の候補として加え（変数名は表 1 の各カテゴリ名を用いた）、各変数が欠損していない 70 件のプロジェクトを用いてモデルを構築した。なお、開発種別が再開発、開発言語が 2GL または自動生成コードについては、それぞれプロジェクトが 1 件だったため、ダミー変数化時に除外した。モデル構築時には、AIC（赤池情報量規準）に基づくステップワイズ変数選択を行った。

モデル構築時には、5-fold cross validation を用いてモデル構築用のフィットデータとモデル評価用のテストデータにデータセットを分割した。試行回数が少ないと統計的にモデルの見積もり精度差を確かめることが難しいため、5-fold cross validation を 4 回繰り返して試行回数を増やした（合計 20 回の試行となる）。

4.2 評価尺度

工数見積もり精度の評価指標として、AE (Absolute Error), MRE (Magnitude of Relative Error) [2], MER (Magnitude of Error Relative to the estimate) [7], BRE (Balanced Relative Error) [11] の 4 つの指標の平均値（以降、例えば AE の平均値を MAE などと表す）と中央値（以降、例えば AE の中央値を MdAE などと表す）を用いた。各指標の値が小さいほどモデルの精度が高いことを示す。ただし、MRE と MER は過大見積もりと過少見積もりに対し、アンバランスな評価になるという問題がある[1][8]。そこで本研究では、それらをバランスよく評価する指標[12]である BRE を特に重視してモデルの精度を評価する。

4.3 開発規模、上流工程工数のみを用いたモデル

計画工数、計画・要件定義工数、FP を用いたモデルの標準化偏回帰係数と VIF（分散拡大要因）を表 5 に示す（20 回の試行の平均値。単回帰の場合の偏回帰係数は、紙面の都合上省略する）。VIF はモデルに多重共線性の発生の有無を確かめるための指標であり、各変数の VIF が 10 を超える場合、多重共線性が発生しているとされる[15]。FP と上流工程工数両方を用いたモデルでは、どの試行においても VIF が 10 を超えることはな

表 5 FP と上流工程工数を用いたモデルの偏回帰係数

(a) FP, 計画工数			(b) FP, 計画・要件定義工数		
変数	標準化偏回帰係数	VIF	変数	標準化偏回帰係数	VIF
FP	0.33	1.34	FP	0.25	1.37
計画工数	0.61	1.34	計画・要件定義工数	0.75	1.37

かった。これより、FPと上流工程工数両方を説明変数として用いても、多重共線性は発生しないといえる。これらのモデルでは、上流工程工数の偏回帰係数のほうがFPよりも大きかった。すなわち、上流工程工数の見積り工数に対する影響が、FPよりも大きいことを示している。

見積もり精度を表 6に示す. 表の 2行目は FP のみを用いたモデルの見積もり精度であり, 3行目以降は FP のみを用いたモデル (2 行目) との見積もり精度の差分 (20 回の試行の平均値) を示している. 値が大きいほど精度が改善したことを示し, 負となっている場合, FP のみを用いたモデルよりも精度が悪化したことを示す. ノンパラメトリックな検定である Wilcoxon の符号付順位和検定を適用し, *は有意水準 5%で有意差があつた場合を示している. 見積もり精度に関して以下の結果が得られた.

- 計画工数を用いたモデル、計画・要件定義工数を用いたモデルでは、全ての評価指標において FP を用いたモデルよりも改善しており、特に *MBRE*, *MdBRE* については有意水準 5%で有意差があった。
 - FP と上流工程工数の両方を用いたモデルの見積もり精度は、FP を用いたモデルよりも大きく改善しており、全ての評価指標で有意差があった。
 - 計画工数を用いたモデルと計画・要件定義工数を用いたモデルを比較すると、後者のほうがより精度が改善しており、FP と計画工数を用いたモデルよりも計画・要件定義工数を用いたモデルのほうが、改善度合いが大きかった。

これより、RQ1に対する答えは「開発規模のみ、もしくは上流工程工数のみで工数を見積もった場合、後者の見積もり精度が高い」となる。

4.4 開発規模、上流工程工数、その他変数を用いたモデル

FP, 上流工程工数とその他の変数を用いたモデルの標準化偏回帰係数と VIF を表 7に示す（紙面の都合上, FP を用いたモデルの偏回帰係数は省略する）．採用回数とは, 20回の試行のうち説明変数として採用された回数を示す．構築されたモデルにおいて, 以下の結果が得られた．

- どの試行においても VIF が 10 を超えることはなかった。FP と上流工程工数に加え、

表 6 FP, 上流工程工数のみを用いたモデルの見積もり精度

Evaluation of software development effort estimation based on upper phase ratio

表 7 FP と上流工程工数を用いたモデルの偏回帰係数

(a) 計画工数

変数	標準化 偏回帰係数	VIF	採用回数
計画工数	0.74	1.08	20
3GL	0.17	1.05	12
MF	-0.18	1.20	8

(b) 計画・要件定義工数

変数	標準化 偏回帰係数	VIF	採用回数
計画・要件定義工数	0.87	1.11	20
3GL	0.13	1.11	8
MF	-0.16	1.17	16
MR	-0.12	1.52	1

(c) FP, 計画工数

変数	標準化 偏回帰係数	VIF	採用回数
FP	0.44	1.33	20
計画工数	0.50	1.40	20
3GL	0.19	1.06	20
MF	0.18	1.52	3

(d) FP, 計画・要件定義工数

変数	標準化 偏回帰係数	VIF	採用回数
計画・要件定義工数	0.66	1.60	20
FP	0.30	1.52	20
3GL	0.14	1.11	19
MF	0.14	1.75	1
新規開発	-0.09	1.25	2

その他の変数を説明変数に加えた場合でも、多重共線性は発生しなかったといえる。

- FP と上流工程工数の両方を説明変数としたモデルにおいて、偏回帰係数の大きさを比較すると、計画・要件定義工数のほうが FP よりも係数が大きかったが(表 7(d))、計画工数と FP では差が小さかった(表 7(c))。すなわち、計画工数と FP では、見積り工数に対する影響の差が小さいことを示している。
- FP と上流工程工数以外では、3GL と MF が説明変数として採用される回数が多かったが、FP と上流工程工数の両方を用いたモデルでは、MF の採用回数は低かった。
- 開発種別はどのモデルでもほとんど採用されなかった。すなわち、上流工程工数を用いた工数見積もりにおいて、開発種別は重要ではないことを示している。

各モデルの見積もり精度を表 8に示す。見積もり精度に関して以下の結果が得られた。

- 計画工数を用いたモデルの評価指標は、8 個中 6 個が FP を用いたモデルより悪化していた。特に *MdBRE* については有意水準 5%で有意に悪化していた。
- 計画・要件定義工数を用いたモデルでは、ほとんどの評価指標が FP を用いたモデルよりも改善しており、*MBRE* については有意水準 5%で有意に改善していた。
- FP と上流工程工数の両方を用いたモデルは、FP を用いたモデルよりも見積もり精度が改善しており、*MBRE* については有意水準 5%で有意に改善していた。
- 計画工数を用いたモデルと計画・要件定義工数を用いたモデルを比較すると、他の変数を説明変数の候補としなかったモデルと同様に、後者のほうが改善の度合いが大きかった。ただし、FP と計画工数を用いたモデルと、計画・要件定義工数を

表 8 FP, 上流工程工数とその他の変数を用いたモデルの見積もり精度

説明変数		MAE	MdAE	MMRE	MdMRE	MMER	MdMER	MBRE	MdBRE
FP		2966.6	1202.3	76.5%	48.7%	112.8%	48.8%	148.4%	72.5%
計画工数	差分	-200.2	-92.3	-3.3%	-11.6%	13.0%	-1.8%	13.2%	-13.9%
	p 値	0.23	0.47	0.45	0.00*	0.43	0.60	0.41	0.03*
計画・要件定義工数	差分	107.6	240.9	22.1%	-1.5%	38.3%	9.1%	56.7%	11.6%
	p 値	0.99	0.05*	0.00*	0.70	0.20	0.00*	0.00*	0.05
計画工数, FP	差分	22.0	231.8	15.6%	2.6%	34.3%	5.9%	46.8%	11.4%
	p 値	0.22	0.03*	0.00*	0.52	0.01*	0.06	0.00*	0.11
計画・要件定義工数, FP	差分	373.6	403.9	29.2%	10.3%	50.6%	14.5%	71.4%	25.9%
	p 値	0.00*	0.00*	0.00*	0.01*	0.00*	0.00*	0.00*	0.00*

用いたモデルの見積もり精度の差は小さかった。

これより、RQ2に対する答えは「開発規模を用いたモデルと比較して、計画工数を用いたモデルは見積り精度が低く、計画・要件定義工数を用いたモデルは見積り精度が高い」となる。また、前節の分析結果と合わせ、RQ3に対する答えは「開発規模と上流工程工数、両方を用いてモデルを構築した場合、見積り精度はより高まる。多重共線性は発生しない」となる。

なお、2章で述べたように、プロジェクトによって計画工数と要件定義工数の定義が異なる可能性があるが、目的変数である総工数はそれらの差異に影響を受けていない。従って、それらの差異は精度に影響を与えても、精度評価の妥当性には影響を与えない。現場への適用時にこれらの差異の発生を抑えるためには、ある時点まで、例えば委託契約から請負契約への変更時までを上流工程と定義するとよいと考えられる。

4.5 工数見積もりモデル構築のガイドライン

RQ1～RQ3への答えに基づいた、工数見積もりモデル構築の新たなガイドラインを以下に示す。収集しているデータ項目が非常に少ない組織では、ある程度の見積り精度が期待できる、上流工程工数のみで総工数を見積もるモデルを作成するとよい。また、項目を多数収集している組織においては、精度向上が期待できるため、上流工程工数に加えてそれらの項目（開発言語など）を用いてモデルを構築するとよい。多重共線性が起きる可能性が低く、また精度向上が期待できることから、開発規模を計測している組織では、開発規模に加えて上流工程工数を工数見積もりモデルに用いるとよい。

5 考察

ISBSG データにおける上流工程比率と生産性の基本統計量を、他のデータセットと比較することにより分析結果の妥当性について考察する。比較対象は、経済調査会と情報処理推進機構・ソフトウェアエンジニアリングセンター（IPA/SEC）それぞれが収集したデータセットである。Yang ら[17]が用いている CSBSG データセットは、一部の統計量が不明であるため比較対象としなかった。経済調査会のデータセット[6]は、2001 年から 2008 年にかけて、日本の情報処理サービス企業 268 社から収集されたものであり、中小企業から大企業まで幅広い層から収集していることに特色がある。IPA/SEC のデータセット[5]は、大手 IT ベンダー 23 社から 2000 年代のプロジェクトを収集したものである。

経済調査会データにおける基本設計比率、生産性の基本統計量を表 9 に、IPA/SEC データのものを表 10 に示す。表 9 のデータでの FP の計測手法は主に IFPUG 法（一部例外あり）、表 10 は IFPUG 法のみである。どちらも計画・要件定義工程比率は（直接）示されていなかった。また、新規開発と改造開発に分けて統計量が示されていた。生産性の単位は人月である。

IPA/SEC データの新規開発での最大値と最小値の比を除き、第 1 四分位数 (Q_1) と第

表 9 経済調査会データにおける基本設計比率、生産性の基本統計量[6]

		平均値	最小値	Q_1	中央値	Q_3	最大値	Q_3 / Q_1	最大値 / 最小値
新規	基本設計比率	0.157	0.030	0.100	0.150	0.200	0.660	2.0	22.0
	生産性	20.6	0.4	8.2	15.7	23.2	189.5	2.8	473.8
改造	設計比率	0.141	0.029	0.100	0.138	0.200	0.280	2.0	9.7
	生産性	23.6	3.3	7.7	15.2	23.8	118.5	3.1	35.9

表 10 IPA/SEC データにおける基本設計比率、生産性の基本統計量[5]

		平均値	最小値	Q_1	中央値	Q_3	最大値	Q_3 / Q_1	最大値 / 最小値
新規	基本設計比率	0.161	0.001	0.095	0.143	0.205	0.589	2.2	589.0
	生産性	18.0	0.8	7.7	11.9	20.1	118.2	2.6	140.7
改造	設計比率	0.147	0.002	0.095	0.137	0.189	0.557	2.0	278.5
	生産性	26.0	0.3	7.8	14.8	34.9	235.8	4.5	873.2

3 四分位数 (Q_3) の比及び最大値と最小値の比（表中太字）は、設計比率に比べて生産性のほうが差異が大きかった。このことから、他のデータセットにおいても上流工程比率、または開発規模以外の説明変数を使わない場合、上流工程比率を用いたモデルの見積もり精度が高くなる、すなわち、少なくとも RQ1 に対する答えは他のデータセットにおいても正しい、と考えられる。

2 章で述べたように、ISBSG データでは要件定義工数に基本設計工数や詳細設計工数が含まれている可能性がある。ただし、4 章の実験で用いたプロジェクトでは全て計画工数と要件定義工数が記録されており、これらの計画工数には詳細設計工数は含まれていないと考えられる。よって、少なくとも計画工数を用いたモデルについては、上流工程工数を用いた場合の見積もり精度を示しているとみなしてよい。

全ての開発において工程の区切りが明確ではないため、工程比率は事後的に推定され記録されている、すなわち数値が正確でない可能性もある。ただし、ISBSG データ、経済調査会データ、IPA/SEC データは、全て企業横断的に収集されたデータである。これより、仮にデータの記録者の違いによる推定誤差が存在するとしても、誤差は吸収されていると考えられる。よって、これらのデータセットの分析結果から得られた「プロジェクトによる工程比率の差は小さい」という結果は信頼してよいと考える。

6 関連研究

上流工程の工数と、その工数が全工数に占める割合に基づいて工数を簡易的に見積もる方法は、いくつかの文献で紹介されている[4][14]。文献[4]では、プロジェクトの進捗を確認するために、あるフェーズ終了後の実績工数と、その工数が全工程に占める割合の平均値を用いて工数を再見積もりする方法が紹介されている。文献[14]では、要件定義工数を見積もり、その見積り工数と要件定義が全工程に占める割合から、工数を見積もる方法を紹介している。ただし、開発規模を用いた場合とどちらの見積もり精度が高いのか、また、開発規模と上流工程工数の両方を用いた場合、更に精度が向上するのかについては、我々の知る限りこれまで明らかにされていなかった。Yang ら[17]は、開発工程の分布は工数見積もりに重要であるにもかかわらず、これまで見過ごされがちであったことを指摘している。本研究の主要な貢献は、新たな見積もり方法の提案ではなく、見積もりモデル構築における新たなガイドラインを示したことである。

丸地ら[10]は、各工程の工数と、各工程における手戻り工数と不具合数を用いて、不具合削減施策を行って不具合数が減った場合、どの程度工数を減るのかを見積もる（評価する）モデルを提案している。ただし、このモデルの目的は不具合削減施策の効果を評価することであり、各工程の工数は既知であることが前提となっているため、一般的な（プロジェクト計画立案のための）工数見積もりに利用することはできない。

7 おわりに

本研究では、上流工程の工数との工程比率に基づいた工数見積もりに着目し、上流工程工数を用いた工数見積もりモデルの精度を評価した（工程比率はモデル構築時にパラメータとして推定されるため、説明変数に含まれない）。上流工程の工数を用いた工数見積もりは開発現場で用いられる場合があったが、開発規模を用いたモデルと比較してどちらの見積もり精度が高いのかはこれまで明らかではなかった。分析の結果、上流工程工数を用いることにより見積もり精度の向上が期待できることがわかった。開発規模との間に多重共線性も発生しないことから、上流工程工数は工数見積もりモデルに積極的に採用すべきであるといえる。

8 参考文献

- [1] C. Burgess, and M. Lefley, “Can genetic programming improve software effort estimation? A comparative evaluation,” Journal of Information and Software Technology, Vol.43, No.14, pp.863-873, 2001.
- [2] S. Conte, H. Dunsmore, and V. Shen: Software Engineering, Metrics and Models, Benjamin/Cummings, 1986.
- [3] International Software Benchmarking Standards Group (ISBSG): ISBSG Estimating: Benchmarking and research suite, ISBSG, 2004.
- [4] International Software Benchmarking Standards Group (ISBSG): Practical ways to use the ISBSG data, [http://www.isbsg.org/isbsgnew.nsf/WebPages/EB33DF6DB82ED1C5CA2576DC0081F046/\\$file/Practical%20Ways%20To%20Use%20The%20ISBSG%20Data.pdf](http://www.isbsg.org/isbsgnew.nsf/WebPages/EB33DF6DB82ED1C5CA2576DC0081F046/$file/Practical%20Ways%20To%20Use%20The%20ISBSG%20Data.pdf)
- [5] 情報処理推進機構 ソフトウェア・エンジニアリング・センター, “ソフトウェア開発データ白書 2010-2011,” 情報処理推進機構, 2010.
- [6] 経済調査会 経済調査研究所, “ソフトウェア開発データリポジトリの分析,” 経済調査会, 2010.
- [7] B. Kitchenham, S. MacDonell, L. Pickard, and M. Shepperd, “What Accuracy Statistics Really Measure,” In Proc. of IEE Software, Vol.148, No.3, pp.81-85, 2001.
- [8] C. Lokan, “What Should You Optimize When Building an Estimation Model?,” In Proc. of International Software Metrics Symposium (METRICS), pp.34, Como, Italy, Sep. 2005.
- [9] C. Lokan, and E. Mendes, “Cross-company and single-company effort models using the ISBSG Database: a further replicated study,” In Proc. of the International Symposium on Empirical Software Engineering (ISESE), pp.75-84, Rio de Janeiro, Brazil, Sep. 2006.
- [10] 丸地康平, 今井健男, 岩政幹人, 一條泰男, “工数削減評価のための総工数見積もり方法の提案,” ソフトウェア品質シンポジウム (SQiP) , 2010.
- [11] Y. Miyazaki, M. Terakado, K. Ozaki, and H. Nozaki, “Robust Regression for Developing Software Estimation Models,” Journal of Systems and Software, Vol.27, No.1, pp.3-16, 1994.
- [12] K. Møllokkæn-Østvold, and M. Jørgensen, “A Comparison of Software Project Overruns-Flexible versus Sequential Development Models,” IEEE Transactions on Software Engineering, Vol.31, No.9, pp.754-766, Sep. 2005.
- [13] 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.
- [14] 尾崎憲和, “上流工程の問題解決 見積もり編 - 後編,” 日経システム構築, pp.49, 2005.
- [15] 田中豊, 垂水共之 (編) :Windows 版 統計解析ハンドブック 多変量解析, 共立出版, 東京, 1995.
- [16] B. Winer, D. Brown, and K. Michels: Statistical Principles in Experimental Design, McGraw-Hill, New York, 1991.
- [17] Y. Yang, M. He, M. Li, Q. Wang, and B. Boehm, “Phase Distribution of Software Development Effort,” In Proc. of the International Symposium on Empirical Software Engineering and Measurement (ESEM), pp.384-392, Kaiserslautern, Germany, Oct. 2008.