

ソフトウェア欠陥数予測におけるトービットモデルの適用

村上 優佳紗
近畿大学理工学部情報学科
m.yukasa@gmail.com

角田 雅照
近畿大学理工学部情報学科
tsunoda@info.kindai.ac.jp

戸田 航史
福岡工業大学情報工学科
toda@fit.ac.jp

要旨

本研究では、ソフトウェア欠陥数予測に対しトービットモデルを適用し、その効果を確認した。実際のソフトウェア開発プロジェクトにおいて収集されたデータを用いてソフトウェア欠陥数を予測した。最小二乗法とトービットモデル、およびそれぞれに対数変換を適用した4つのモデルを評価した。その結果、対数変換を適用したトービットモデルは検討すべきモデルであるといえる。

1. はじめに

近年、プロジェクトは大規模化しており、納期の遅れ、品質の低下、コストの超過などを防止する重要性が高まっている。それらを防止するためには、定量的なプロジェクト管理は欠くことができない要素の一つである。例えば、開発言語やソフトウェアの規模に基づき、ソフトウェアリリース後に発見される欠陥数を予測し、欠陥数の予測値に基づき、テスト期間やテストケース数、テスト人員が適切となるように計画を立案する。

予測するために用いられる数学的モデルとして重回帰モデルなどがあり、過去に得られたデータを基に予測モデルが構築される。しかし、ソフトウェアプロジェクトのデータをそのまま用いた場合、ソフトウェアの欠陥数など、最小値が0となる変数がある。このため、最小二乗法に基づく重回帰モデルでは適切なモデルが構築されず、欠陥数や開発工数の予測値が負の値となる可能性がある。例えば、開発規模を説明変数とし、開発規模が100FP(ファンクションポイント)で欠陥数が0というデータが含まれるデータセットを用いて(対数変換を行わずに)モデルを構築したとする。この場合、予測対象プロジェクトの開発規模が100FP未満、例えば50FPの時、欠陥数の予測値が負の値となる可能性がある。この問題を解決するために、本研究ではトービットモデルをソフトウェア欠陥数予測に適用することを提案する。

2. トービットモデル

トービットモデルとは、データの分布の偏りを考慮したモデルである。データの偏りの分布は以下の3つに分類される[4]。

- 打ち切り(censored)
- 切断(truncated)
- 付随的切断(incidental truncation)

打ち切りとは、欠陥数など、目的変数の最小値が0以上など、値域があらかじめ決まっている場合である。切断とは、欠陥数が0のデータを除外するような場合である。付随的切断とは、レビュー指摘数の省略などが理由で本来0でない数値が0となる場合である。

ソフトウェア欠陥数は打ち切りに該当するため、本研究ではタイプIのトービットモデルを適用する。タイプIトービットモデルの打ち切り回帰モデルは以下ようになる[4]。

$$y^* = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (1)$$

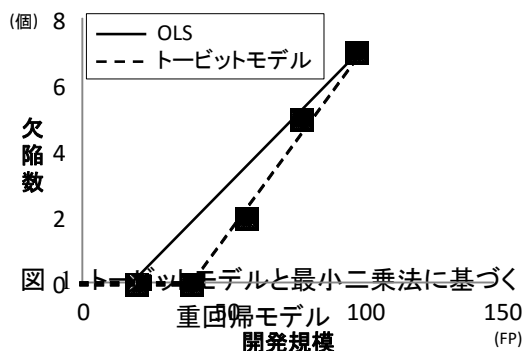
$$y = \begin{cases} y^* & y^* > 0 \\ 0 & y^* \leq 0 \end{cases} \quad (2)$$

ここで β_0 は回帰定数、 $\beta_1, \beta_2, \dots, \beta_k$ は偏回帰係数、 ε は誤差項である。最小二乗法では y が0となっているケースについても1つのモデル(式(1)のみ)で扱うが、トービットモデルではそれらを別の式(式(2))で扱うことにより、モデルが適切に構築される。トービットモデルはかなり以前に提案されたものであるが、我々の知る限り、これまでソフトウェア欠陥数の予測モデル構築に適用されていない。

図1に、説明変数を開発規模、目的変数をソフトウェア欠陥数として予測モデルを構築した場合のトービットモデルと最小二乗法に基づく重回帰モデル(OLS)のイメージを示す。グラフ内にプロットされている点は各プロジェクトのケースを表す。最小二乗法に基づく重回帰モデルと比べて、トービットモデルは欠陥数が0のケースを適切に扱っていることがわかる。

ソフトウェア欠陥数の予測は、ソフトウェアのテスト計画を立案する際に、ソフトウェアの品質を予測するために

われる。例えば品質が低い、すなわち欠陥数が多いと予測された場合、テストケース数の計画数を増加させるなどし、欠陥の発生を抑える。



なお、ソフトウェアの欠陥数が0となっているケースが多く存在するという、データの偏りが存在する理由は、データの収集方法などに問題があるためではない。一般的なソフトウェア開発において、リリース後の欠陥数は0となる場合が多い。このために上述の偏りが発生する。従来の最小二乗法に基づく重回帰モデルでは誤差には偏りが無いという前提のため、この偏りが考慮されないが、トービットモデルでは偏りを考慮したモデルが構築される。そのため、予測精度の向上が期待される。

3. 実験

最小二乗法に基づく重回帰モデル(OLS)、トービットモデル(Tobit)、対数変換を適用した最小二乗法に基づく重回帰モデル(OLS_{ln})、対数変換を適用したトービットモデル(Tobit_{ln})の計4つを用いてソフトウェア開発プロジェクトにおける欠陥数の予測を行い、精度を比較した。0は対数変換できないため、各変数にあらかじめ1を加えた上で対数変換した。欠陥数は0以上の整数であり、その最大値は149であったため、1を加えることとした。なお、これは対数変換をする場合のみの処理であり、対数変換後の値は0となる。

提案手法の有用性を評価するために、ISBSG (International Software Benchmarking Standards Group) が収集したプロジェクトデータ (ISBSGデータ[1]) を用いて分析を行った。ISBSGデータは、世界20ヶ国のソフトウェア開発企業から収集されたものであり、予測手法の有

効性を評価する研究で広く用いられている[2][3][5]。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。データには欠損値が含まれている。欠損値とは値が記録されていないことを指す。

分析に用いた変数は総欠陥数、未調整FP、開発種別(新規開発など)、業種(製造業など)、開発プラットフォーム(メインフレームなど)であり、本研究では総欠陥数を目的変数とした。開発種別などのカテゴリ変数はダミー変数化した。予備分析において、OLSとAIC(Akaike Information Criterion)による変数選択を行い、モデルにおいて説明変数として用いる変数を絞り込んだ。具体的には、未調整FP、新規開発、金融業、メインフレーム、ミッドレンジを説明変数として採用した。

分析対象のプロジェクトの条件を整えるため、データ品質評価がAまたはB、FP計測法がIFPUG (International Function Point Users Group) のデータを抽出した。また、分析で用いた変数に欠損が含まれているデータを除去した。その結果、221件のプロジェクトが分析対象となった。欠損が含まれているデータを除去することはリストワイズ除去と呼ばれ、広く用いられており、予測モデルの構築時に適用することが妥当であることが指摘されているため[6]、これを採用した。

モデルの予測精度を評価するために、5-fold cross validationによる実験を行い、評価指標は、絶対誤差(AE)、Balanced Relative Error(BRE)それぞれの平均値、中央値とPred25を用いた。AE、BREは値が小さいほど精度が高いことを示す。Pred25は、誤差25%以内のデータの割合を指し、数値が高いほど精度が良いことを示す。

実験結果を表1に示す。評価値は、OLSの各評価指標から、Tobitの各評価指標を引いた値である。ただし、Pred25のみTobitからOLSを引いた値とする。これはそれぞれの評価指標の評価値をみた際に正の値ならばTobitモデルの方が優れている(負の値ならばOLSの方が優れている)ということを分かりやすくするためである。実験結果では以下の傾向が見られた。

- (1) OLSとTobitの比較: AE平均値, AE中央値, BRE中央値は改善し, BRE平均値, Pred25は悪化した。
- (2) OLS_{ln}とTobit_{ln}の比較: BRE平均値, Pred25は大

表1 各モデルの予測精度

	AE平均値	AE中央値	BRE平均値	BRE中央値	Pred25
OLS	10.74	4.11	284%	165%	12%
Tobit	9.66	2.04	378%	115%	8%
評価値	1.08	2.06	-94%	49%	-4%
OLS _{ln}	8.24	1.86	226%	107%	21%
Tobit _{ln}	8.45	1.53	260%	93%	28%
評価値	-0.21	0.33	-34%	14%	7%

幅に改善し、*AE*平均値、*AE*中央値は大きな差がなかった。

(1)については、特に*BRE*中央値を見るとOLSと比べ100%近く悪化していた。また、Pred25をみると、誤差の大きなデータ件数もOLSより多かった。(2)については、*BRE*平均の評価値には大幅な改善がみられた。Pred25に関してはOLS_Inよりも高い精度となった。

これらにより、対数変換を適用しない場合のトービットモデルは*BRE*平均値とPred25に問題を残すが、対数変換後はこれらの指標を改善することが可能となり、有効性を示すことができたと考える。

4. おわりに

本研究では、ソフトウェア欠陥数の予測にトービットモデルを用いることを提案した。実験結果では、対数変換を適用したトービットモデルは、対数変換を適用した最小二乗法に基づく重回帰モデルよりも5つある評価尺度のうち3つに対し予測精度が高いと評価された。このことから、対数変換を適用したトービットモデルは検討すべきモデルであるといえる。対数変換を適用したトービットモデルでは、最小二乗法に基づく重回帰モデルよりも*BRE*平均値が悪化する傾向があるが、その理由の分析と改善方法について、他のデータセットを用いて検討することは今後の課題である。

謝辞 本研究の一部は、文部科学省科学研究補助費(基盤 C:課題番号 25330090)による助成を受けた。

参考文献

- [1] International Software Benchmarking Standards Group, “ISBSG Estimating Benchmarking and Research Suite Release 9,” ISBSG (2004).
- [2] Jeffery, R., Ruhe, M. and Wiczorek, I., “Using Public Domain Metrics to Estimate Software Development Effort,” Proc. International Software Metrics Symposium (METRICS), pp.16-27(2001).
- [3] Mendes, E., Lokan, C., Harrison, R. and Triggs, C., “A Replicated Comparison of Cross-company and Within-company Effort Estimation Models using the ISBSG Database,” Proc. International Software Metrics Symposium (METRICS), p.36 (2005).
- [4] 水落正明, “打ち切り・切断データの分析”, 理論と方法(数理社会学会 機関誌), Vol. 24, No. 1, pp. 129-138, (2009)
- [5] Sentas, P., Angelis, L., Stamelos, I. and Bleris, G., “Software productivity and effort prediction with ordinal regression,” Information and Software

Technology, Vol.47, No.1, pp.17-29(2005).

- [6] Strike, K., El Eman, K., and Madhavji, N., “Software Cost Estimation with Incomplete Data,” IEEE Transactions on Software Engineering, Vol.27, No.10, pp.890-908 (2001).