

# 工程別の欠陥埋め込み数の予測 —ソフトウェア詳細設計，コード作成を対象として—<sup>1</sup>

衣簾 宏和<sup>2</sup> 野中 誠<sup>3</sup> 阿萬 裕久<sup>4</sup>三菱電機マイコン機器ソフトウェア<sup>2</sup> 東洋大学 経営学部<sup>3</sup> 愛媛大学大学院 理工学研究科<sup>4</sup>

## 1. はじめに

ソフトウェア品質管理において，工程別の品質状況を評価したいというニーズがある．その際，レビューでの欠陥検出件数に加えて，当該工程で欠陥がどの程度埋め込まれているかを併せて検討することが，精度の高い品質管理のために必要である．

欠陥の埋め込み数は，開発規模に加えて，プロジェクトの定性的要因を考慮して推定される．しかし，実務に置いては経験的判断に基づいて推定される事が多く，データに基づいて導出されたモデルが適用される事は少ない．一方，従来研究では，開発ライフサイクル全体の欠陥埋め込み数を開発規模と定性的要因により予測する研究などが示されている．しかし，工程別の欠陥埋め込み数を実データに基づいて予測する研究は，十分に行われていると言いがたい．

筆者らは，ソフトウェア開発の工程のうち既にソフトウェア方式設計についてモデル構築，報告を行っている<sup>5</sup>．本稿では，ソフトウェア詳細設計工程とコード作成工程を対象に，開発規模に加えてプロジェクト環境などに関する定性的要因を考慮して，実データに基づいて欠陥埋め込み数の予測モデルの構築と評価を行った経験を報告する<sup>6</sup>．

## 2. 分析対象のデータ

分析データには，筆者の所属部門で 2008 年から 2010 年に開発されたソフトウェアの品質データを用いた．表 1 に分析対象データの概要を示す．表 2 に開発タスク単位で収集している定量的メトリクス，表 3 に定性的メトリクスを示す(いずれも抜粋)．なお，分析にあたっては開発タスク単位のデータを用いる．

表 1 分析対象データの概要

項目	値
データ収集期間	3年間(2008年から2010年まで)
プロジェクト数	5件
開発タスク数	46件
総開発規模(新規・変更)	88.92 KLOC

表 2 タスク単位で収集している定量的メトリクス(抜粋)

メトリクス名	意味	単位
欠陥埋め込み数	該当の工程で埋め込まれた欠陥件数	件
開発規模	新規開発と変更開発の規模	KLOC
流用規模	完全流用の規模	KLOC
欠陥検出数	該当工程のレビューで検出した欠陥数	件
レビュー工数	該当工程で実施したレビュー工数	人時

表 3 タスク単位で収集している定性的メトリクス(抜粋)

記号	メトリクス名	水準数
$M_1$	開発対象に対する事前学習の有無	2
$M_2$	性能要求の難易度	3
$M_3$	システム構造の複雑度	3
$M_4$	既存システムからの流用	2
$M_5$	仕様変更の可能性	3
$M_6$	未解決課題の数	3
$M_7$	直近の仕様変更の頻度	3

## 3. 欠陥埋め込み数の予測モデル

以下に，分析手順とソフトウェア詳細設計とコード作成のそれぞれの結果を示す．

**Step1:**それぞれのメトリクスについて分布の確認をした．その結果，「欠陥埋め込み数」と「開発規模」は右に偏った分布であったため，対数変換を行った(以降，対数変換した変数には LN を付す)．**Step2:**「LN 欠陥埋め込み数」と，これ以外の定量的および定性的メトリクスとの関連を分析した．散布図や相関係数，筆者らの経験的知識を考慮した上で，説明変数の候補をそれぞれ表 4，

表 5 以下のように絞り込んだ．この過程で，欠陥値のあるデータ，欠陥埋め込み密度の大きく異なるデータその他の変数について傾きの傾向の大きく異なるデータを除外し，ソフトウェア詳細設計の分析対象は 26 件，コード作成の分析対象は 24 件とした．

表 4 ソフトウェア詳細設計の説明変数の候補

記号	メトリクス名
$M_2$	性能要求の難易度
$M_3$	システム構造の複雑度
$M_4$	既存システムからの流用

1 Predicting Injected Defects by Phases: A Study for Software Detail Design and Coding

2 HIROKAZU KINUHATA

Mitsubishi Electric Micro-Computer Application Software Co., Ltd.

3 MAKOTO NONAKA

Faculty of Business Administration, Toyo University

4 HIROHISA AMAN

Graduate School of Science Engineering, Ehime University

5 「ウィンターワークショップ 2012 イン 琵琶湖」にて報告．論文番号 m01.

6 本研究(の一部)は，国立情報学研究所トップエスイーププロジェクトにおける修了製作として実施した．

表 5 コード作成の説明変数の候補

記号	メトリクス名
$M_2$	性能要求の難易度
$M_4$	既存システムからの流用

**Step3:** ステップワイズ変数選択法による重回帰分析を実施し、表 6の結果を得た。説明変数は、いずれも統計的には比較的有意であり、また、予測モデルの説明力も高い。得られた予測モデルを式で表すと、それぞれのようになる。尚、式(1)はソフトウェア詳細設計であり、式(2)はコード作成のモデルを示す。

$$\log_e D = 2.23 + 0.99 * \log_e S - 0.29 * M_2 \quad (1)$$

$$\log_e D = 1.71 + 0.82 * \log_e S \quad (2)$$

ここで、 $D$  は欠陥埋め込み数、 $S$  は開発規模である。

表 6 欠陥埋め込み数の重回帰分析結果

工程	説明変数	偏回帰係数 (P 値)	相関係数 (P 値)
ソフトウェア 詳細設計	定数項	2.23 (0.000)	0.92 (0.000)
	LN 開発規模	0.99 (0.000)	
	$M_2$	-0.29 (0.153)	
	調整済み $R^2$	0.847	
	$N$	26	
コード作成	定数項	1.71 (0.000)	0.81 (0.000)
	LN 開発規模	0.82 (0.000)	
	調整済み $R^2$	0.64	
	$N$	24	

#### 4. 予測モデルの評価

選択された説明変数を用いて、分析対象のデータ、ソフトウェア詳細設計 26 件、コード作成 24 件に対して leave-one-out 交差検証を行った。予測誤差の評価にあたっては、予測値と実績値の差を実績値で割った MRE (Magnitude of Relative Error) を用いた。表 7 に、各工程の MRE の平均値 MMRE, 中央値 MdMRE, 及び MRE の標準偏差を示す。また図 1 に MRE の分布を示す。尚、これらの数値は対数値から元の測定単位である欠陥数に戻した上で算出している。

図 1 から、四分位範囲に入ったデータについては十分な予測精度と考えられるが、それ以外のデータでは予測誤差が大きい事が分かる。特に外れ値となった 3 件のデータについては欠陥埋め込み件数を過大に予測している事が分かる。

表 7 leave-one-out 交差検証の結果

工程	項目	値
ソフトウェア 詳細設計	MMRE	1.301
	MdMRE	0.313
	MRE の標準偏差	3.306
コード作成	MMRE	0.650
	MdMRE	0.189
	MRE の標準偏差	1.631

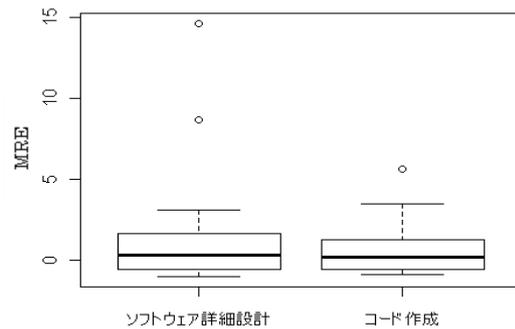


図 1 MRE の分布

#### 5. 考察

ステップワイズ変数選択法により、ソフトウェア詳細設計では「LN 開発規模」と「性能要求の難易度 ( $M_2$ )」が、コード作成では「LN 開発規模」のみが選ばれた。ソフトウェア詳細設計、コード作成ともに「LN 開発規模」が選ばれた事は開発規模の増減が欠陥の埋め込みに影響していることを考えると妥当と判断する。ソフトウェア詳細設計において「性能要求の難易度 ( $M_2$ )」が選択された事は要求されたソフトウェア仕様に対する実現の難易度が、不具合の埋め込みに影響することを考慮すれば妥当と判断する。コード作成の説明変数には定性的メトリクスが選択されていない。また、コード作成に関しては説明変数として開発規模のみが選択され、定性的メトリクスは何れも採用されない結果になった。コード作成の段階まで開発が進むと、プロジェクト管理的な視点よりもむしろソースコードそのものやコーディングに着目した分析が重要になって来るのではないかと思われる。

#### 6. おわりに

「LN 開発規模」、「性能要求の難易度」からそれぞれ「欠陥埋め込み数」のバラつきを説明出来る予測モデルが得られた。交差検証においては課題が残る結果となった。交差検証における外れ値については、外れ値となった理由を説明しうる要因の調査・分析を行う必要がある。また、コード作成についてはコードやコーディングに着目した分析を行う為にもデータ収集と分析が必要である。

#### 参考文献

- 1) Fenton, N., et. al., : Project Data Incorporating Qualitative Facts for Improved Software Defect Prediction, *Proc. Of Int' l Workshop on Predictor Models in Softw. Eng.*, pp.2 (2007)
- 2) 衣簾 宏和, 野中 誠, 阿萬 裕久: 工程別の欠陥埋め込み数の予測-ソフトウェア方式設計を対象として-, ウィンターワークショップ 2012 イン 琵琶湖