

間接的メトリクスを用いて欠陥予測を行うレビュー方法の提案

- 欠陥の位置と種類の特定によりレビューの効率と効果を向上 -

The suggestion of review method of forecasting defect by using indirect metrics

- The efficiency and the effect of the review are improved by specifying the position and the kind of the defect -

2010 年度 SQiP 研究会 第 3 分科会

主査 細川 宣啓¹⁾ 副主査 永田 敦²⁾ アドバイザー 森崎 修司³⁾
 研究員 ○中谷 一樹⁴⁾ 諏訪 博紀⁵⁾ 田邊 哉好⁶⁾ 森崎 一邦⁷⁾ 末次 努⁸⁾
 小田部 健⁹⁾ 山本 浩之¹⁰⁾ 牧野 将治¹¹⁾ 小原 美帆⁴⁾ 奥山 剛¹²⁾

The 3rd Section Meeting Team of SQiP Study Group in 2010

CHIEF INVESTIGATOR: Nobuhiro Hosokawa¹⁾ SUB-CHIEF INVESTIGATOR: Atsushi Nagata²⁾

ADVISER: Shuji Morisaki³⁾

RESEARCHER:

○Kazuki Nakatani⁴⁾ Hiroki Suwa⁵⁾ Chikayoshi Tanabe⁶⁾ Kazukuni Morisaki⁷⁾ Tsutomu Suetsugu⁸⁾
 Ken Otabe⁹⁾ Hiroyuki Yamamoto¹⁰⁾ Shoji Makino¹¹⁾ Miho Ohara⁴⁾ Tsuyoshi Okuyama¹²⁾

研究概要

ソフトウェアの欠陥は仕様の複雑さ・人の行動・開発現場の環境など様々な要因により成果物に混入する。欠陥を上流工程で摘出するにはレビューが有効であるが、欠陥摘出効率の低さや重大な欠陥が残存する問題がある。本研究チーム（2010 年度 SQiP 研究会 第 3 分科会）では、レビュー実施前にとどの位置にとどのような欠陥が多く存在しているか特定できれば、それらの問題が改善できると考えた。

TIS 株式会社 技術本部 生産技術部

TIS Inc., Production Technology Dept, Technologies Div.

東京都港区海岸 1-14-5 Tel:03-5402-2625

1-14-5, Kaigan, Minato-ku, Tokyo Japan Tel:03-5402-2625

- | | |
|-------------------------|---|
| 1) 日本アイ・ビー・エム株式会社 | IBM Japan, Ltd. |
| 2) 株式会社ソニー | Sony Co., Ltd. |
| 3) 静岡大学 | Shizuoka University |
| 4) T I S 株式会社 | TIS Inc. |
| 5) 三菱UFJ トラストシステム株式会社 | Mitsubishi UFJ Trust Systems Co., Ltd |
| 6) 株式会社日立製作所 | Hitachi, Ltd. |
| 7) アルパイン株式会社 | Alpine Electronics Inc. |
| 8) 東芝電波システムエンジニアリング株式会社 | Toshiba Electronics Engineering Co., Ltd. |
| 9) 株式会社小野測器 | Ono Sokki Co., Ltd. |
| 10) 株式会社システムフロンティア | System Frontier Co., Ltd. |
| 11) 日本電子株式会社 | JEOL Ltd. |
| 12) 株式会社山武 | Yamatake Co., Ltd. |

レビュー実施前にプロジェクトやドキュメントの状態を表す間接的なメトリクスを用いて、欠陥の位置と種類を予測する手法について研究を行い、この手法の実現可能性を示唆するに至った。

間接的なメトリクスとは、人の行動や開発現場の環境や体制など、ドキュメントの欠陥とは直接関係なさそうなメトリクスである。レビュー経験が豊富な人は、一見無関係に思えるような物も含め、レビューの前や最中、意識的または無意識の内に総合的に判断しながら重大な欠陥を上手く見つけ出している。それを定量的に評価し、誰でも扱えるようにした物が間接的なメトリクスである。

本論文では、レビューの効率と効果の向上に貢献するであろう「間接的なメトリクスを用いて欠陥予測を行うレビュー方法」について提案する。

Abstract In software development, it is influenced from various factors, and the defect enters the product. The factors are, for instance, complexity of specification, person's behavior, and environment of development. In the upper process, the review of specification is effective to remove the defect. However, there are two problems. It is inefficient of the defect removal, and the serious defect remains. In this research team, it was thought that specifying the position and the kind of the defect before the review was executed improved this problem. We had researched the technique for forecasting the position and the kind of the defect before executing the review by using various metrics that showed the state of the project and the product. And, we were able to suggest the realizability of this technique.

In this thesis, it proposes "Review method of forecasting the defect by using indirect metrics" that improves the efficiency and the effect of the review.

1. はじめに

1.1 研究の背景とレビューに関する問題点

筆者らの開発現場では、基本設計書や機能設計書のレビューで欠陥を抽出できなかったために、後工程で欠陥が多発し納期遅延や開発コストの増加を招いているという同じ悩みを抱えている。

筆者らが解決したい問題点は次の2点である。1点目は、レビューに費やした時間に対し欠陥抽出率が低いという効率の問題である。2点目は、誤字などの品質に大きく影響しない軽微な欠陥が多く抽出され、重大な欠陥が残存する効果の問題である。これらの問題点の解決策として「欠陥を予測する手法」「レビューのガイド」「チェックリストの活用」の3案を考えた。その中で筆者らが経験したことは無いが、直感的に最も有効で画期的な解決策に思えた「欠陥を予測する手法」について研究することにした。

1.2 研究の狙い

一般的なレビューは、「すべての位置」に存在する「すべての種類」の欠陥を探索・特定しようと試みる作業である。どんなに教育を受けて経験を積んだ担当者であっても、すべての欠陥を抽出することは、非現実的な労力を要することになる。換言すれば、レビュー実施前に抽出すべき重大欠陥の「位置」と「種類」が特定できれば、飛躍的にレビューの効率と効果を高めることができると考えた。

レビュー手法は強力な武器であるが、レビューに関する市販書籍は少なく、どのようにレビューするかという「レビューの種類や手順」いわゆる「How」の部分の解説が多い。「What」つまり何をレビューすべきか（欠陥の種類）、どこをレビューすべきか（欠陥の位置）を提示したものはほとんどない。「What」は、プロジェクトの特性や状況に合わせて決める必要があり、簡単には決められないからである。筆者らは、現状のままでは重大欠陥の効率的・効果的な抽出の実現は困難であると考えた。

重大欠陥の効率的・効果的な摘出を実現するために、レビュー実施の前後にできる工夫はないかを検討し、本研究では、「What」つまり『何を』に着目した以下の2点に焦点を当てた。

①『欠陥の多いドキュメントやチームを予測する』 → 重大欠陥の効率的な摘出施策

欠陥の多いドキュメントやチーム（＝欠陥の位置）を予測しレビューを実施することで、時間あたりの欠陥摘出件数は向上すると考える。さらにレビューを実施したドキュメントの欠陥情報を横展開し、他ドキュメントの見直しを実施する。見直し後のドキュメントをレビューする際は、その他の欠陥摘出に注力できるので、従来よりも短い時間でレビューが完了すると考える。

②『混入している欠陥種別を予測する』 → 重大欠陥の効果的な摘出施策

混入している欠陥種別が予測できれば、レビュー観点を絞ることができる。摘出する欠陥種別に絞ったレビューを実施することで、重大欠陥の欠陥摘出率は向上すると考える。

上記①②の探索的な2つのアプローチを導入し、「欠陥の位置と種類」を特定することにより、最も効果が高いが最も工数がかかる手作業のレビューの効率（レビュー時間の短縮）と効果（重大欠陥の摘出）を飛躍的に向上させる手法の開発と提案を研究の狙いとした。

2. 提案する欠陥予測の手法

2.1 欠陥予測につながるメトリクス

本研究では、これまでの研究で取り上げられているような複雑度・カバレッジなど直接状況を表すメトリクスだけでなく、喫煙者比率、机上のペットボトル空き本数、設計書の最終保存日時、特定キーワードの出現数など、レビュー経験豊富な人が意識的または無意識的に見ているであろう、プロジェクトや成果物の状態を間接的に表すメトリクスにも着目した。

本研究チームのメンバーの経験や憶測を基にプロジェクトおよびプロダクトの状態を表すメトリクスを考案した。プロジェクトのメトリクスは、プロジェクト内のコミュニケーション・体制・作業環境などに関するものである。プロダクトのメトリクスは設計書のファイル属性情報・図表の数・特定キーワードの出現数などに関するものである。

2.2 欠陥予測の手順

筆者らが提案し実践した欠陥予測の手順を以下に示す。

(1)メトリクスの選定とメトリクス値の収集

プロジェクトの特性に合ったメトリクスを選定し、メトリクス値を収集する。収集データは予測する単位（チーム別、機能別、ファイル別、ページ別など）に分類する。

(2)データの分析と予測

収集したデータを分析し欠陥の位置と種別を予測する。1つのメトリクスだけでなく複数のメトリクスを組み合わせることでデータの分析を行うことで予測の精度を高める。データの分析方法については、3.3章、3.4章にて研究での具体例を後述する。

分析した結果から、欠陥が多いドキュメントやチーム、混入している欠陥種別を予測し、レビューする対象・順序とレビュー観点を設定する。

①レビュー対象・順序の設定

レビューの対象はチームや機能別に欠陥が多いドキュメントを1つずつレビューするか、全体で欠陥が多い順番にレビューするなどプロジェクトの特性に合わせて設定する。

②レビュー観点の設定

混入している欠陥種別の予測結果から、レビューの観点を設定する。予測した欠陥種別のうち除去したい種別を選択してレビュー観点を設定する。

3. 研究結果

3.1 研究対象としたプロジェクトの情報

本研究では時間的な制約により既に完了しているプロジェクトの開発実績データを使用し、予測結果との比較を行った。開発実績データは図 1 に機能仕様書のレビュー、図 2 にテスト工程での欠陥摘出件数を示す。

機能設計書のレビューで摘出した欠陥は、欠陥種別を分類するための情報が得られなかったため重大な欠陥と誤字レベルの軽微な欠陥の 2 つに大きく分類した。テスト工程で摘出した欠陥は、図 2 に示す欠陥種別で分類した。

レビューで摘出した欠陥種別毎の欠陥件数	A チーム	B チーム	C チーム	D チーム
重大な欠陥	500	450	100	40
軽微な欠陥	20	10	30	10
Ksあたりの欠陥密度	3.5	4.5	1.0	0.8

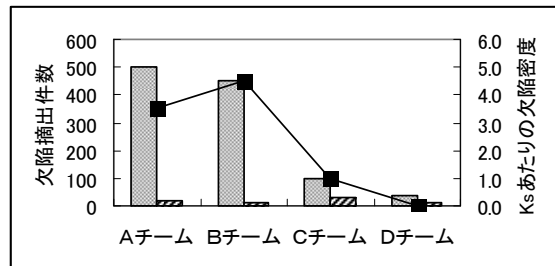


図 1 チーム別のレビュー欠陥摘出件数

テスト工程で摘出した欠陥種別毎の欠陥件数	A チーム	B チーム	C チーム	D チーム
基本設計の誤り	35	5	12	14
処理方式の誤り	35	3	1	5
異常処理の誤り	3	4	0	2
領域設計の誤り	2	2	1	0
内部変数使用の誤り	11	8	19	4
インタフェース設計の誤り	17	16	10	9
その他	14	11	6	4
合計	117	49	49	38
Ksあたりの欠陥密度	1.51	1.54	1.16	1.11

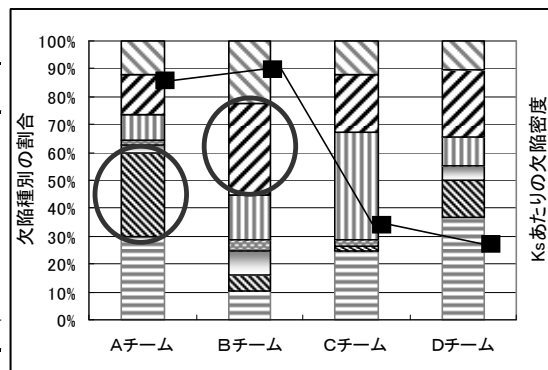


図 2 チーム別のテスト工程における欠陥摘出件数

3.2 評価方法

(1) 研究対象としたメトリクス

研究期間内にデータ収集可能であり、筆者らの経験から予測の精度が高いと考えて計測したメトリクスを表 1 に示す。

表 1 計測したメトリクス

#	分類	メトリクス
1	プロジェクト	コミュニケーション
2		環境
3		体制
4	プロダクト	語句
5		ファイル情報

(2) 各メトリクスのデータ収集

① プロジェクトの情報

プロジェクトの情報は、メトリクスごとに選択肢を 3 つ用意しアンケート形式でデータを収集した。

② プロダクトの情報

プロダクトの情報は、機能設計書の電子ファイルのプロパティ情報や、検索機能でメトリクスとした文字の出現回数を計測した。

(3) データ分析による欠陥の位置と種別を予測

収集したプロジェクト情報とプロダクト情報を、チーム単位に集計し傾向を分析した。その結果をもとに、欠陥が多いチームと混入している欠陥の種別を予測した。なお、収集したメトリクスのうちデータの信憑性が高く、チーム間で差異が見られるものだけを分析に使用した。それ以外のメトリクスは収集したが分析には使用しなかった。

(4) レビューおよびテストで抽出した欠陥データとの比較

上記の予測結果とレビューおよびテストで抽出した欠陥データを比較し、予測の妥当性を評価した。

3.3 欠陥が多いチームの特定

(1) プロジェクトのメトリクスから欠陥が多いチームを予測

開発担当者に、プロジェクトのメトリクスをアンケート形式にしてヒアリングを実施した。各チームの主な特徴・傾向は表 2 に示す。

表 2 主なプロジェクトメトリクスおよびその特徴・傾向

プロジェクトメトリクス	Aチーム	Bチーム	Cチーム	Dチーム	特徴・傾向
残業時間	毎日残業	毎日残業	半分残業	毎日定時	Aチーム・Bチームは、毎日残業なので、自己レビューを行う時間が確保できない、スケジュール優先で品質が無視されている、ミスが発生しやすい深夜に作業している可能性がある
机上のペットボトルの数	多い	1~2本	なし	なし	Aチーム・Bチームは、机上にペットボトルが散乱しているので、残業が多い状態と同じで、時間に追われ品質が疎かになっている可能性がある
喫煙者比率	約半数	少ない	少ない	約半数	Bチーム・Cチームは、喫煙者比率が低いので、喫煙所などリラックスした状態での会話が少なく、喫煙者比率が高いチームに比べるとコミュニケーションが取れていない可能性がある コミュニケーション不足が起因となるインタフェース設計の誤りが含まれる可能性がある
喫煙所での仕様決め	多い	たまに	ない	たまに	Aチームは、喫煙所での仕様決めが多いため、一部のメンバー間でのコミュニケーションが取れている状態で、仕様決めの背景などがメンバー全員に伝わっていない可能性がある

プロジェクトの状態が悪いチームは欠陥が多く含まれていると考え、どのチームのドキュメントをレビューするかという視点で予測を行った。

① 残業時間と机上のペットボトルに着目した予測

Aチーム・Bチームは、残業時間が多いので、自己レビューを行う時間が確保できない、スケジュール優先で品質が無視されている、ミスが発生しやすい深夜に作業しているなど、プロジェクトの状態が悪い可能性が高い。また、机上のペットボトル空き本数も、残業時間と同じ傾向にあり、図 3 に示す通り Aチーム、Bチームの余裕のなさが伺える。

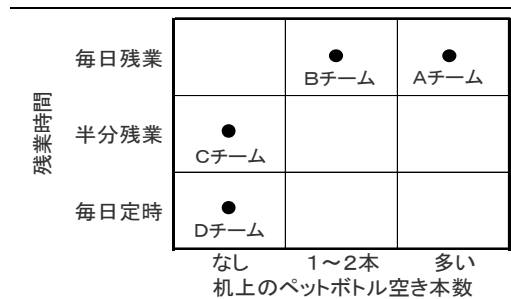


図 3 残業時間と机上のペットボトルの空き本数の散布図

② 喫煙者比率と喫煙所での仕様決めに着目した予測

喫煙所は作業フロアよりも気軽に話しやすい雰囲気があり、設計書に取替えて記載していない既知の仕様や不安を抱えている箇所などの確認が行い易いと考えた。

Bチーム・Cチームは、喫煙者比率が低いので、設計書を読むだけでは理解できない事柄について、他のチームに比べると、コミュニケーションが不足する可能性がある。

一方 Aチームは、喫煙所での仕様決めが多いため、一部のメンバー間でのコミュニケーションが取れて

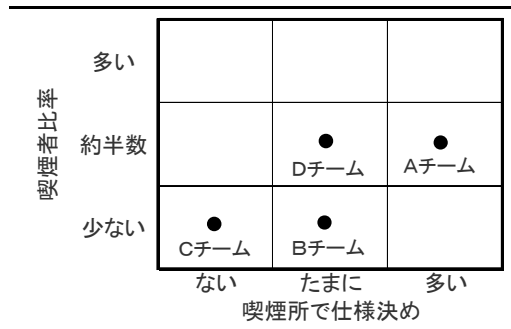


図 4 喫煙者比率と喫煙所での仕様決めの散布図

いる反面、仕様決定の背景などがメンバー全員に伝わっていない可能性がある。よって、図 4 に示す通り喫煙者比率が約半数で、喫煙所でたまに仕様決定している D チームが、最も適度にコミュニケーションが図られていると予測した。

このように、忙しさ・情報共有・意志疎通の度合いを測るためのいくつかのメトリクス値を単独あるいは組み合わせることで、B チームの欠陥混入率は高く、D チームは低いと予測した。

(2) プロダクトのメトリクスから欠陥が多いチームを予測

ファイルのプロパティ情報や、ワードの検索機能または目視により、各種ファイル情報や各キーワードの数を取得した。

プロダクトのメトリクスから読み取れる主な特徴・傾向を表 3 に示す。

表 3 主なプロダクトのメトリクスおよびその特徴・傾向

プロダクトメトリクス	Aチーム	Bチーム	Cチーム	Dチーム	特徴・傾向
「。」「,」「」	大	大	中	小	Aチーム、Bチームは、文章量が多い 文章量が単純に多いことから、日本語の表現が起因となる処理方式の誤りが含まれる可能性がある
「。」「,」「」 /表の数	小	大	中	中	Bチームは、表の数に対して文章量が多い Aチームは、表の数に対して文章量が少ない
「場合」 /「。」「,」「」	中	中	大	小	Cチームは、文章量に対して条件が多い Dチームは、文章量に対して条件が少ない
「および」	大	小	中	小	Aチームは、条件が複雑 条件が複雑なことから、処理の複雑さが起因となる処理方式の誤りが含まれる可能性がある
「且つ」「かつ」	大	中	中	小	Aチームは、条件が複雑 条件が複雑なことから、処理の複雑さが起因となる処理方式の誤りが含まれる可能性がある

ファイルや頁の単位で欠陥が多い位置を予測することも可能だが、今回はどのチームのドキュメントをレビューするかという視点で予測を行った。

① 句読点の数と表の数に、着目した予測

句読点の数から、A チーム・B チームは、文章量が多いことが分かる。日本語には多義語や二義文など、作者の意図と違う意味で解釈される表現がある。その為、日本語の表現に起因する欠陥が含まれている可能性が高い。一方、図 5 の横軸（表の数に対する句読点の数）を見ると、A チームは句読点の数の比率が低いので、なるべく日本語の平文で記述せず、表を用いて条件や分類毎に整理して記述している可能性が高い。

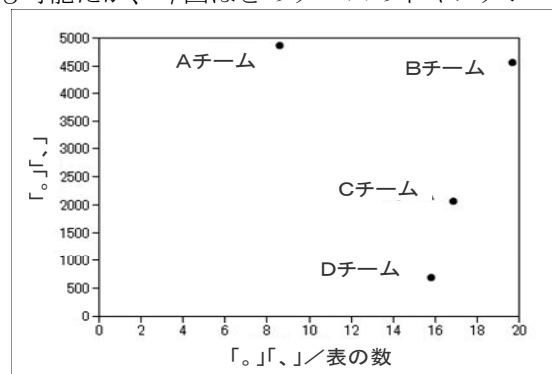


図 5 句読点と表の散布図

日本語の平文では分からなかったが、表を用いて整理して記述すると、ある条件での処理が記述されていなかった事に気付くこともある。よって、B チームの欠陥混入率が最も高いと予測した。

② 句読点の数と「場合」の数に、着目した予測

図 6 に示す句読点の数に対する「場合」の数から、D チームは、条件が少なく簡単な機能であり、欠陥が含まれる可能性が低いと予測した。

このように、規模・簡潔さ・複雑度の度合いを測るための情報となるいくつかのメトリクス値を単独あるいは組み合わせて見ることで、B チームの欠陥混入率は高く、D チームの欠陥混入率は低いと予測した。

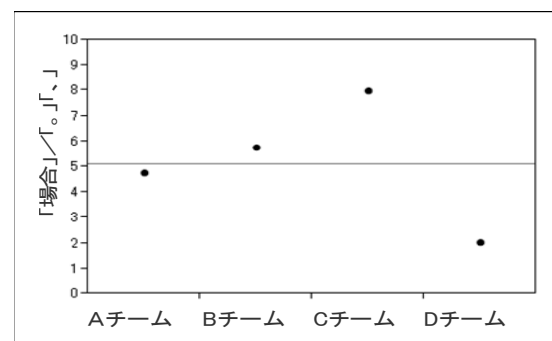


図 6 句読点と場合の散布図

(3) 予測結果の検証

レビューで抽出した欠陥件数と、テスト工程で抽出した欠陥件数は、図 1、図 2 に示した通りである。欠陥混入率が最も高かったチームはBチームで、欠陥混入率が最も低かったチームはDチームであった。この結果は、筆者らが行った「プロジェクトのメトリクス」と「プロダクトのメトリクス」からの欠陥予測と同じ結果であった。

3.4 欠陥種別の特定

(1) プロジェクトのメトリクスから多く発生する欠陥種別を予測

ヒアリング結果から得られた主な特徴・傾向は、前述の表 2 に示した通りである。その中から、欠陥種別の予測につながりそうなメトリクスとして、喫煙者の比率に着目した。

喫煙所は作業フロアよりも気軽に話しやすい雰囲気があり、設計書に取って記載していない既知の仕様や不安を抱えている箇所などの確認が行い易いと考えた。

Bチーム・Cチームは、喫煙者比率が低いので、設計書を読むだけでは理解できない事柄について、他のチームに比べると、コミュニケーションが不足する可能性がある。よって、コミュニケーション不足が起因となる「インタフェース設計の誤り」が多く含まれると予測した。

(2) プロダクトのメトリクスから多く発生する欠陥種別を予測

プロダクトのメトリクス値から読み取れる主な特徴・傾向は、前述の表 3 に示した通りである。その中から、欠陥種別の予測に使えるメトリクスとして、句読点・「および」・「且つ」・「かつ」・「場合」に着目した。

Aチームは、上記のメトリクス値が全て高いので、日本語の表現や処理の複雑さに起因する「処理方式の誤り」が多く含まれると予測した。

また、Aチーム・Bチーム・Cチームは、「場合」／句読点のメトリクス値が高いため、条件分岐の処理が多くなると考えられ「異常処理の誤り、領域設計の誤り」が多いと予測した。

(3) 予測結果の検証

テスト工程で抽出した欠陥の種別は、図 2 に示した通りである。「インタフェース設計の誤り」の抽出比率が最も高いのはBチームであり、「(1) プロジェクトのメトリクスから多く発生する欠陥種別を予測」で導いた結果と同じであった。また、「処理方式の誤り」の抽出比率が最も高いのはAチームであり、「(2) プロダクトメトリクスから多く発生する欠陥種別を予測」で導いた結果と同じであった。但し、「異常処理の誤り、領域設計の誤り」は、抽出数が少なく予測の妥当性は確認できなかった。

4. 考察

研究の結果、筆者らが行った欠陥予測の結果と、検証対象プロジェクトのレビューおよびテストの結果はほぼ同じであった。これは、レビューやテストを実施する前に、プロジェクトやプロダクトのメトリクスを用いて、欠陥混入率の高いチーム、低いチームを予測することや、混入している欠陥の種別を予測することが原理的に可能であるということが示唆できたと言える。

但し、欠陥種別の予測で「異常処理の誤り、領域設計の誤り」は、テスト工程での抽出数が少なく予測の妥当性は評価できなかった。欠陥は混入していたがレビューで除去できていたのか、この種の欠陥が入り込まないような言語や開発手法などプロジェクトの特性があったのか、欠陥抽出数が少ない要因の特定には至らなかったためである。

5. 今後の課題

5.1 研究の狙いに至らなかった事項

(1) 考案したメトリクス全ての検証

今回は研究期間が十分でなかったため、考案したメトリクスのうちの一部しか計測することができなかった。また、計測したがデータに差異が見られないものやデータの信憑性が低いものがあり、分析に使用したメトリクスはその一部だけである。検証できなかった残りのメトリクスについても欠陥予測の妥当性について検証する必要がある。

(2) 欠陥予測手法適用による効果の測定

提案した手法を用いることである程度、予測した通りに欠陥が混入していることが実証できた。しかし、この手法を適用することで「レビューの効果と効率の向上」がどの程度図れるか、また、従来手法（複雑度やカバレッジ等を用いた欠陥予測）と比べてどの程度効果的なのかについては、今後、多くの開発中のプロジェクトに欠陥予測手法を適用し、データを分析する必要があると考える。

5.2 将来的な課題事項

(1) メトリクス収集の精度向上・効率化

メトリクスとして収集したワード文書のプロパティ情報（作成日時や編集時間など）は、有効な情報でない場合があり、手作業で収集したため、労力を要した。プロジェクト計画段階で①有効な情報が収集できる仕組み、②自動収集できる仕組みを検討する必要があると考える。

(2) 欠陥混入予防策へのフィードバック

今回の実験から、例えば、設計書の記述において日本語の平文で記述するよりも、できるだけ表を用いると欠陥混入率が下がるという予測の有効性を評価できた。設計書記述の基準として、表を多く使用することを推奨するなど、欠陥混入予防策へつながらせるものはフィードバックすべきと考える。

但し、間接的なメトリクスの多くは、欠陥混入の予兆や痕跡であるため、メトリクスの値そのものを操作することに意味が無いものが多い。例えば、フッタのフォントが違う頁に欠陥が多いと予測できるからといって、フッタのフォントを統一することを設計書作成時のチェック項目に入れてしまうと、欠陥混入の痕跡を逆に消してしまうことになる。

レビューアが欠陥予測のために密かに利用するメトリクスと、作成者へフィードバックする内容は上手く切り分けながら、有効なフィードバックを行う必要があると考える。

6. おわりに

本研究では、レビュー関連の書籍で解説されているレビューのやり方（技法）ではなく、何を・どういうところをレビューすべきかについて着目した。そして、レビューの前に、様々なプロジェクトやプロダクトのメトリクスから欠陥が多く存在するところや欠陥の種類を予測し、効率的にかつ効果的にレビューを行う考え方を提案した。実証作業を通し、ただ漠然とレビューするのではなく、レビューの前にプロジェクトのコンテキストに目を向けることが効率的なレビューにつながると確信した。早く、安いシステム開発が求められている昨今では、有効な施策の1つと言える。

参考文献

- [1] Karl E. Wiegers (著) 「ピアレビュー」 日経BPソフトプレス社 2004年
- [2] SQuBOK 策定部会 「ソフトウェア品質知識体系ガイド—SQuBOK Guide」 オーム社 2007年
- [3] 野中誠 「fault-prone モジュール予測技法の基礎と研究動向」 SES2010 2010年