

ピアレビュー網羅率を用いた品質評価技法の提案

久野 倫義^{1,a)} 中島 毅¹

受付日 2011年5月26日, 採録日 2011年11月7日

概要: 本論文では, ソフトウェア開発におけるピアレビュー実施結果の品質評価において, 対象成果物に対してピアレビューが均質に実施できていない場合に欠陥を多く見逃してしまう問題を解決するための品質評価技法を提案する. 提案技法では, 従来のゾーン分析法の品質評価を補強し, 新しい指標であるピアレビュー網羅率を導入することでレビュー実施の均質さを評価する. 実プロジェクトに提案技法を適用し, テスト段階への流出欠陥数が減少する効果を確認した.

キーワード: ピアレビュー, ソフトウェアインスペクション, 定量的品質管理

A Method for Software Quality Evaluation Using the Peer-review-coverage Measure

NORIYOSHI KUNO^{1,a)} TSUYOSHI NAKAJIMA¹

Received: May 26, 2011, Accepted: November 7, 2011

Abstract: In this paper, we propose a quality evaluation method to solve the problem that many defects remain in some peer-reviewed products in case that the peer-reviews are not conducted uniformly over the target documents in the software development. Our method enhances the zone analysis method by using a new metrics of peer-review-coverage, which measures the uniformity of peer-reviews to target documents. An application of our method to real projects shows a good result of reducing defects that should be detected by peer-reviews and flowed into the test phases.

Keywords: peer-review, software inspection, quantitatively quality control

1. はじめに

ソフトウェアの欠陥が引き起こすシステム障害の社会的な影響が増大するなか, 高品質なソフトウェアを開発するための検証手法およびそれらを使った品質管理方法を確立することが求められている. 高品質なソフトウェアを最終検査工程だけで達成することは困難であり, 開発各工程で確実に欠陥を除去していき後工程に流出させないことが必要である [1], [2].

要求分析からコーディングに至る設計工程では, 設計文書やプログラムを対象としたピアレビューが主たる検証手段である [3], [4], [5]. ピアレビューは人手により実施する

ため参加者個人の技量に大きく依存し, その効果にバラツキが現れやすい [6]. 検証手法として, このバラツキを軽減することを目的に, 観点やチェックリストを用いる方法 [7] やプロセスを重視し組織力を活用する方法 [8], [9] などが提案・評価されている.

品質管理手法は, ピアレビューの結果を評価し成果物に残存する欠陥を計画目標内に抑えるために定量的なデータを用いて行う管理方法であり [1], [10], その管理精度向上のために種々の提案がなされている [11], [12], [13]. しかし, 従来の品質管理手法はレビュー効果のバラツキを扱う方法を十分に確立できていない. 実際, テスト段階で検出される欠陥の中でレビュー時に検出可能なものが多く, その原因はレビューの実施結果を適切に評価できていないことにある [8].

一般に, 各開発工程の品質管理者は, 成果物に対するレビュー実施結果を見て, 成果物が要求品質を達成している

¹ 三菱電機株式会社設計システム技術センター
Design Systems Engineering Center, Mitsubishi Electric Corporation, Kamakura, Kanagawa 247-8501, Japan

a) Kuno.Noriyoshi@cb.MitsubishiElectric.co.jp

か、再レビューなどの対策が必要であるかを判断することが多い。この判断は、主に成果物の成果物品質（およびそれによって想定される後工程のリスク）に基づいて行われる。

レビュー結果に基づく成果物品質の判断に用いる方法として、検出欠陥数に加えてレビュー密度の実績から品質を判断するゾーン分析法 [1], [10] や検出欠陥数の実績から残存欠陥数を推定する欠陥数推定法 [14], [15], [16], [17], [18] がある。

ゾーン分析法や欠陥数推定法は、ともにレビュー結果に基づく定量的な成果物の品質評価技法であり、レビューが成果物に対してムラなく実施されていることを前提としている。しかし、現実には、成果物に対して、その前半部分でレビュー時間を費やし後半までレビューできない場合が多い。この状況は潜在欠陥が多い成果物を対象としたときに生じやすい。こうした場合に上記の品質評価技法を単純に適用すると、収集されたデータが平均値で評価されてしまうため、残存欠陥を多く含む部分があるにもかかわらず、品質上問題なしと判断を誤ることがある。これが、設計工程から後工程に欠陥が流出する大きな要因となっている。

この問題を解決するために、我々は、実施したピアレビューが対象成果物をどの程度網羅しているかを評価するためにピアレビュー網羅率という新しい指標を導入することで、従来の品質評価技法の弱点を補完する品質評価技法を開発した。実プロジェクトに本技法を適用し、テスト段階への流出欠陥数が減少する効果を確認した。

本論文では、2章において、従来研究としてゾーン分析について特長と課題を述べ、その適用上の問題を明確にすることで、本論文で解決すべき課題を明確にする。次に、3章においては、レビュー実施の網羅性を考慮する手法を提案し、4章では提案手法の実プロジェクトへの適用結果を示しその有効性を示す。5章では関連研究について述べ、本提案技法との関係を明確にする。

2. 従来研究と解決すべき課題

レビュー結果に基づく成果物品質の判断を行う方法の1つであるゾーン分析法を、本研究の従来研究として述べ、本研究で解決すべき課題を明確にする。

2.1 ゾーン分析法

ある開発工程で、検出欠陥数の実績に基づき残存欠陥数を推定しようとする時、レビューによる検出欠陥密度（欠陥数を対象規模で割り正規化したもの）が過去の実績と比較して高い場合に、2つの相反する解釈が成り立つ。1つは、レビュー対象に元々含まれる欠陥数が多い、つまり残存欠陥数は多いという解釈、もう1つは、レビューが効果的であり欠陥をとりきった、つまり残存欠陥数は少ない、という解釈である [10]。

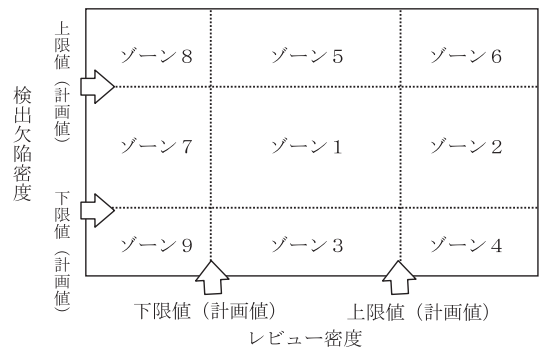


図 1 ゾーン分析法

Fig. 1 Zone analysis method.

ゾーン分析法 [1], [10] は、この相反する解釈の問題を解決するために、品質データとして検出欠陥数に加えてレビューにかけた工数を用いる。欠陥数推定法のように直接残存欠陥数を推定するのではなく、図 1 に示すようなレビュー密度と検出欠陥密度の 2 軸平面上で、9つのゾーンに分け、レビュー対象ごとに、計画値に対する実績値を、品質状態がどのゾーン入るかで評価する [10]。

ここで、計画値は、過去の実績から導出したレビュー密度と検出欠陥密度における上限値と下限値である。レビュー密度 Dr と検出欠陥密度 Df は、検出欠陥数 Nf 、レビュー工数 Tr 、および成果物規模 Nw から、以下の式を使って計算する。

$$Dr = Tr/Nw \quad (1)$$

$$Df = Nf/Nw \quad (2)$$

表 1 に、各ゾーンにおける品質評価、および判断と対策を示す。

ゾーン分析法は、レビュー結果に基づく成果物品質の判断を行う方法として、以下の特長を持つ。

- ① レビュー会議で通常収集されるデータ（レビューにかけた延べ時間と、それによって検出された欠陥数）だけから評価を行うことができる。
- ② 同一平面上で複数のレビュー対象を比較して見ること、同一プロジェクト内の均質な条件下で、同一作成担当者での複数対象および作成担当者間の対象の相对比较により、特異点を見出すことができる。
- ③ 同一レビュー対象に対する複数回のレビューを必須としない。
- ④ 複数のレビュー結果を統合して評価することや、累積で経時変化をプロットして評価することもできる。

2.2 ゾーン分析法の課題

ゾーン分析法には、次の 2 つの課題がある。

- ① ゾーンの決定の仕方：ゾーンを決定するのは、レビュー密度と検出欠陥密度の計画値である。計画値として、組織実績値（あるいは過去の類似プロジェクトの実績

表 1 ゾーンの意味

Table 1 Description of the zones.

ゾーン	品質評価	判断と対策
ゾーン1	良い	問題なし
ゾーン2	良い	レビュー方法(指摘内容及びレビュー人数や回数)を確認する
ゾーン3	良い	レビュー方法を確認後,追加レビューを検討する
ゾーン4	良い	レビュー方法の妥当性を確認後,追加レビューを検討する
ゾーン5	悪い	指摘内容を確認し,追加レビューを検討する
ゾーン6	悪い	記載内容を確認し,追加レビューを行う
ゾーン7	良い	指摘内容を確認し,追加レビューを検討する
ゾーン8	悪い	記載内容の見直しを行う
ゾーン9	判断保留(追加レビュー後に判断)	追加レビューを行う

値)を使う場合が多い。そのため、過去のデータの集積が少ない場合には、あるいは対象プロジェクトの特殊性(たとえば、開発体制・要員や流用率の変化)がある場合には、品質判断の信頼性が低くなってしまふ。

- ② 個々のレビュー実施に対する考慮：ゾーン上にプロットされる点は、個々のレビューの実績値を表す。実績値は規模で正規化されるため、対象に対する均一なレビューの実施が前提とされている。しかし、現実には、大量の成果物を対象にしていること、人とチームに依存して実施の仕方が異なることにより、実施方法と欠陥にバラツキが生じており、それに対する考慮がない。

課題①に対しては、過去の実績値が蓄積することで徐々に軽減されてくる。また、過去の実績がなくても、特長②を用いて実施の特異点を見つけることや定性的なデータを併用することで判断を補強し利用することができる。しかし、課題②に関しては、レビュー実施のバラツキを防止するために、品質管理者が直接レビューに参加し状況を把握することが対策として考えられるが、現実にはリソース制約上網羅的に実施することは困難である。そのため、ゾーン分析による品質評価の有効性を損なうという意味で、課題②がより重要な問題となっている。

2.3 本研究が扱う課題

本研究は、ゾーン分析の課題②に関して、検出効率のバラツキ以外の原因を明らかにし、その解決方法を提案するものである。本研究で扱う課題について以下に記述する。課題②が生じる例を以下で詳しく述べる。図2は、ゾーン上に4つの機能仕様書のレビュー密度(工数/ページ)と検出欠陥密度(件/ページ)の実績値をプロットしたものである。

図2では、機能仕様書Aに対しては、ゾーン9(レビュー密度および検出欠陥密度が低い)に入るため、追加レビューを行うという対策を実施する。仕様書Bに関しては、ゾ

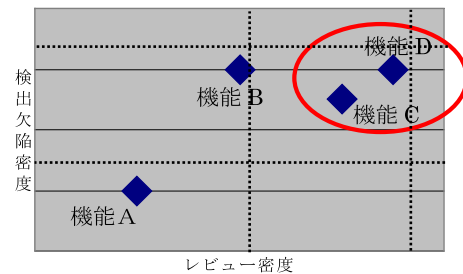


図 2 従来のゾーン分析手法の適用実例
Fig. 2 Example for zone analysis.

ン7(レビュー密度が低いにもかかわらず欠陥検出密度が高い)に入るため、指摘内容の点検を行うという対策をとる。

問題は、機能仕様書CとDである。これらは、ゾーン1(レビュー密度と欠陥検出密度が目標値を満たしている)に入るため問題なしと判断されるが、実際にはテスト段階で検出された欠陥が多く、その中に明らかにレビュー検出漏れであると判断されるものが多く含まれているケースがあった[8]。

レビュー検出漏れが生じる原因の1つとしては、レビューの検出効率のバラツキがある。これは、レビュー活動中に成果物の説明や設計修正自体を実施し、欠陥検出の時間をかけていないことが原因であった。この問題を解決するため、レビュー会議の実施方法を見直すことで解決を図ってきたが、改善後においても、レビュー検出漏れを要因とするテスト段階で検出される欠陥数は減少しない場合があった[8]。

3. ピアレビュー評価手法を用いた品質評価技法の提案

前章で述べた問題点を解決することを目的に、成果物のピアレビューにおいて、レビュー実施漏れを防止し、テスト段階への流出欠陥数を削減するためのピアレビュー評価技法を提案する。

提案する品質評価技法は、ピアレビュー網羅率と呼ぶ指標を導入することで、ゾーン分析を実施する際に、実施したピアレビューが対象成果物をどの程度網羅しているかを定量的に評価できるようにするものである。以下、提案する新しい指標とそれを利用した品質評価技法について述べる。

3.1 事前調査

3.1.1 レビューでの欠陥検出パターン

図2のゾーン分析の結果において、品質良好と判断された機能のうち、テスト段階で欠陥が多発したものと、そうでなかったものを、レビュー記録を確認することで、比較分析を行った。レビュー記録は、指摘ごとに、① ページ番号、② 指摘内容、③ 指摘者、④ 重要度(大/中/小)、⑤ 回答、⑥ 回答者、⑦ 確認者を記載している。

図3と図4は、単一の成果物に対して、複数レビュー者がピアレビューを行った結果から、ページごとの指摘件数を集計して示したグラフである。図3はテスト段階でレビュー実施漏れの欠陥が多かった機能部分、図4は欠陥が少なかった機能部分に関するものである。同一指摘は欠陥数を1件としてカウントした。図3は、成果物の前半部

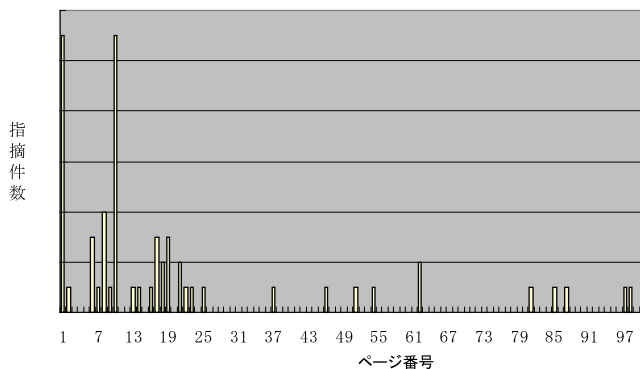


図3 ページ番号と指摘件数の関係(流出欠陥数の多かった機能)

Fig. 3 Number of defects in each page (in case of many defects detected in testing).

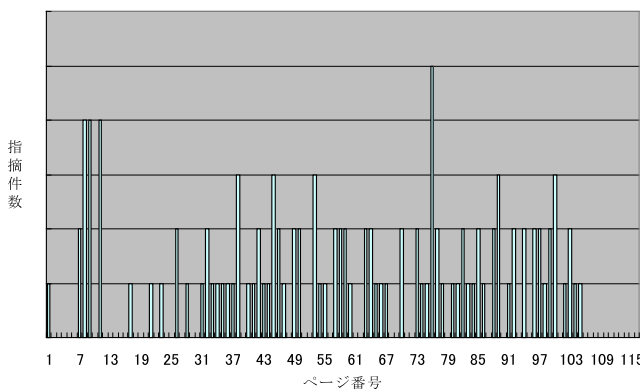


図4 ページ番号と指摘件数の関係(流出欠陥数の少なかった機能)

Fig. 4 Number of defects in each page (in case of few defects detected in testing).

分に多くの指摘があるが、後半部分ではレビュー指摘件数が0件のページが多いことを示している。これに対して、図4は前半から後半まで満遍なく指摘が上がっていることを示している。

この分析結果から、欠陥流出の原因は、レビュー対象の後半部分に欠陥検出数が少ない場合に多く、この場合後半部分に欠陥が多く残存している可能性が高いことが分かった。対象成果物に混入した欠陥は成果物全体に均質に存在すると仮定すると、レビュー対象の後半部分に欠陥検出数が少ない原因は、レビュー実施が不足しているか、検出効率が落ちているかのどちらかである。しかし、前半と後半部分に対するレビューは読み進む順序以外に差異が見当たらない。以上のことから、我々は、成果物の後半部分が前半部分に比して検出件数が少ない原因が、前半で多くのレビュー時間を費やしたため後半のレビューにかかる時間が不足したことにあるのではないかと推定した。

上記の問題点は単純なゾーン分析法による評価では顕在化できない。実際、図3と図4の2つのケースにおいて、欠陥の多かった機能および欠陥の少なかった機能のレビュー指摘密度、検出欠陥密度はそれぞれ下限値上限値の範囲内であり、両機能とも問題なしと判断されている。

このことは、ゾーン分析法が成果物全体に対するレビュー密度という平均化された情報でのみ評価を行っていることに原因がある。たとえば、特定ページで多く時間を費やし多くの指摘を行い、その他のページでレビューがされなかった場合、全体的には欠陥検出密度が目標どおりという評価となってしまうからである。そして、これは非熟練者により作成された成果物に多くみられるパターンであり、品質管理上見逃してはいけないケースである。本研究はこの問題点を解決する技法を提案する。

3.1.2 作業成果物への混入欠陥の分布

図5は、テスト段階への欠陥流出の少なかった上位13の機能を選択し、その設計成果物を対象に、設計段階のレビューにおいて、成果物上のどの位置で欠陥が検出されたかを示すグラフである。横軸は、成果物全体のページを100%とした際の欠陥のあるページの相対位置である。縦軸は、相対ページ10%刻みごとに成果物の検出欠陥数を足

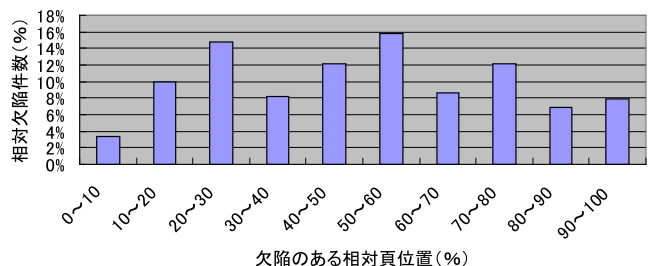


図5 成果物の相対頁位置における欠陥分布

Fig. 5 Distribution of defects over relative page position of artifacts.

し合わせたものを、相対件数で表している。

図5に示すように、テスト段階への欠陥流出の少ない場合において、検出欠陥が成果物の前半と後半に偏在せず成果物全体にわたっている。このことから、対象とするプロジェクト組織においては、設計成果物の混入欠陥が、レビュー対象のページ全体にわたり存在し、かつ前後半に偏在せず分布する可能性が高いことが分かる。

3.2 提案技法

3.2.1 ピアレビュー網羅率

一般には、ピアレビューが対象物に対して満遍なく実施されたかどうかを測定することは難しい。レビュー対象部分（たとえばページ）に対して直接レビューが行われた時間を測定できれば直観的であるが、実際のレビューでは対象を頻繁に変更させることが多く、どこを対象としているかを明確に区別できないため、その時間計測は困難である。

そこで、我々は、(ピアレビュー結果である) 個々の指摘情報から取得可能なデータを使って計算するピアレビュー網羅率と呼ぶ指標を、レビュー実施の網羅性を表す代替特性として採用した。ピアレビュー網羅率は、ページごとの指摘件数の有無から、対象成果物に対するピアレビュー実施のバラツキを評価する。定義を式(3)に示す。

$$D = \sum(D_i)/N \quad (3)$$

ここで、 D_i ：ページ*i*の指摘有無

(有は1、無は0で示す)

i：ページ番号、*N*：総ページ数

式(3)で示すように、ピアレビュー対象成果物のページごとの指摘が有なら1、無なら0を割り当てる。これによりそのページがレビューされたかどうかを判断している。この値を対象文書の全ページにわたり総和をとり、総ページ数で割ったものを、成果物のピアレビュー網羅率とする。たとえば、10ページに3件の指摘があった場合において、あるページに3件の指摘が集中していれば網羅率は0.1となり、指摘が3ページに分散していれば網羅率は0.3となる。

3.2.2 ピアレビュー網羅率(前半)およびピアレビュー網羅率(後半)

ピアレビュー網羅率は、ピアレビューが対象成果物の全体に対してレビュー実施の網羅性を表す代替特性として定義した。さらに、ピアレビュー網羅率を成果物の前半部分のページに対するものと後半部分のページに対するもので分離することで作業成果物後半におけるレビューの網羅性を評価するために2つの指標、ピアレビュー網羅率(前半) Df およびピアレビュー網羅率(後半) Db を以下のように定義する。

$$Df = \sum(D_i)/(N) \quad i = 1, \dots, N/2 \quad (4)$$

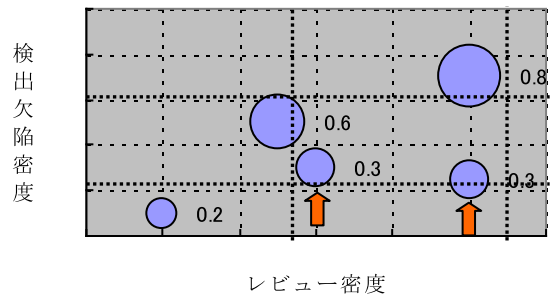


図6 ピアレビュー網羅率を用いたゾーン分析

Fig. 6 Example of zone analysis with the peer-review-coverage.

$$Db = \sum(D_i)/(N) \quad i = N/2 + 1, \dots, N \quad (5)$$

ここで、*N*：総ページ数

3.2.3 ゾーン分析とピアレビュー網羅率(後半)を組み合わせた品質評価技法

我々が提案する品質評価技法は、ピアレビュー網羅率(後半)の情報をゾーン分析法のグラフに付加することで、対象の網羅度に対する品質判断の観点を補うものである。図6は、提案技法におけるレビュー結果の評価に用いるグラフの表示例である。図6において、レビュー対象は(従来のゾーン分析法と同様に) レビュー密度と検出欠陥密度のゾーン上にプロットされ、各プロットはその大きさがピアレビュー網羅率(後半)を表すように表示される。

個々のレビュー結果は、従来のゾーン分析法の評価にレビュー実施の網羅性の観点からの評価を加えて品質判断される。図6の例において、矢印の成果物は、ゾーン分析法ではゾーン1に入るため品質が良いと判断されるが、ピアレビュー網羅率(後半)が0.3であることから、後半部分でのレビューが不十分であることが分かり、追加レビューが必要であるとの品質判断を下すことができる。なおピアレビュー網羅率の閾値0.3は、検出欠陥密度の組織実績値を参考に算出した。検出欠陥密度は、1ページあたりの平均欠陥数であり、フェーズごと(ソフトウェア要求分析など)に組織閾値を定めている。1ページあたりの平均欠陥数の閾値が0.3であることから、少なくともレビュー網羅率としても、0.3以上であるとした。

4. 適用と評価

提案技法を実サブプロジェクトに適用し評価を実施した。レビュー対象成果物はソフトウェア要求仕様書(総ページ数500超)、レビュー者は10名であった。

4.1 ピアレビュー網羅率の評価

過去のレビュー結果から、レビュー対象成果物別に算出したピアレビュー網羅率を表2に示す。なお、表2では総ページ数が2ページ以下のものは含めていない。

表2から、ピアレビュー網羅率に関して以下のことが分かる。

表 2 過去の成果物に対するレビュー網羅率

Table 2 Peer-review-coverage for work products.

成果物番号	総ページ数 (N)	指摘件数	網羅率 (D)
1	106	91	0.53
2	106	139	0.50
3	61	76	0.46
4	40	51	0.58
5	24	39	0.96
6	4	9	0.75

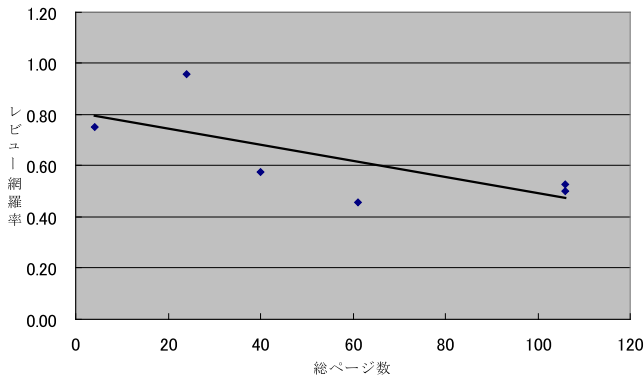


図 7 総ページ数とレビュー網羅率 (D) の関係

Fig. 7 Relation between the number of pages and the peer-review-coverage (D).

表 3 成果物 1, 2, 3, 4, 5 に対するピアレビュー網羅率

Table 3 Peer-review-coverage for work products 1, 2, 3, 4, 5.

成果物番号	総ページ数 (N)	指摘件数	網羅率 (D)	網羅率 (前半) (Df)	網羅率 (後半) (Db)
1	106	91	0.53	0.32	0.21
2	106	139	0.50	0.35	0.15
3	61	76	0.46	0.33	0.13
4	40	51	0.58	0.43	0.15
5	24	39	0.96	0.50	0.46

- 総ページ数が少ない場合、時間切れによるレビュー実施漏れの発生は少ない。
- 成果物の総ページ数が多くなると、網羅率に差異が表れ、レビュー実施漏れとの相関を検討できる。

ピアレビュー網羅率とレビュー実施漏れとの相関を見るために、成果物の総ページ数とレビュー網羅率の相関を分析した結果を図 7 に示す。サンプル数は少ないが、図 7 から、総ページ数が多くなるほどピアレビュー網羅率は低下し、レビュー実施漏れになりやすい傾向があることが分かる。

4.2 ピアレビュー網羅率 (前半) およびピアレビュー網羅率 (後半) の評価

表 3 に、総ページ数の多い成果物 1, 2, 3, 4, 5 に対して、計算した結果を示す。

表 3 から、成果物全体でのピアレビュー網羅率 D は、 Df と Db の関係より、成果物の後半部分でのピアレビュー

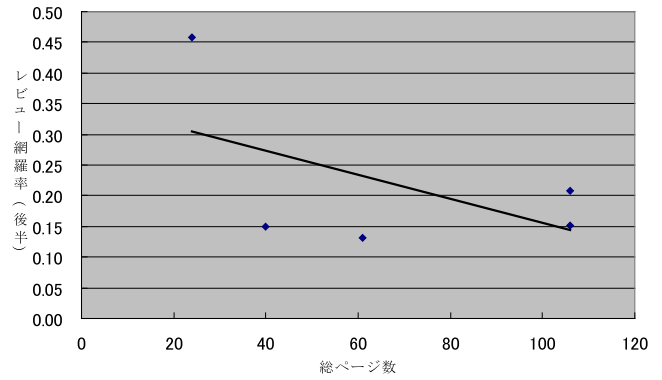


図 8 総ページ数と後半レビュー網羅率 (Db) の関係

Fig. 8 Relation between the number of pages and the last half peer-review-coverage (Db).

網羅率が低く、後半での漏れが発生しやすい状況であることが分かる。

また、図 8 は、成果物総ページ数と後半網羅率 Db の関係を示したものである。

図 8 から、成果物の総ページ数が大きくなると後半のページにおいてレビュー指摘のあるページが減少していることが分かる。このことから、前半にレビュー時間を多くかけ、後半のレビューに十分に時間をかけなかったことが推定できる。

4.3 ゾーン分析とピアレビュー網羅率 (後半) を組み合わせた品質評価技法の評価

レビュー結果の品質評価は、ゾーン分析を基本として、本技法により品質判断を補うことで実施した。具体的には、以下のようにピアレビュー網羅率を利用した品質判断と対策を行った。

- レビュー工数が目標に満たない場合、ピアレビュー網羅率を確認する。
- ピアレビュー網羅率の前半と後半で差があれば、成果物の後半部分でレビューが不足していると判断する。
- 後半部分に対して、短時間の追加レビューを行う。
- 追加の指摘件数を加えてレビュー網羅率を評価する。

(1) 適用結果と評価

図 9 は、1つのプロジェクトにおける成果物のレビュー実施結果に対するゾーン分析結果である。対象としたプロジェクトは、14個のサブプロジェクトで構成され、1つのサブプロジェクトは、2名から5名程度の独立した開発者により開発を実施している。1つの丸が1つのサブプロジェクトの作業成果物、丸の中の数値は、ピアレビュー網羅率 (後半) を示す。図中の数値の付いている部分は、従来のゾーン分析では品質が良いと判断される領域にあり問題なしと判断される。しかしピアレビュー網羅率 (後半) が0.27と低い場合、後半のページ部分でレビュー実施漏れが発生していると判断され、当該機能部分に対して追加レ

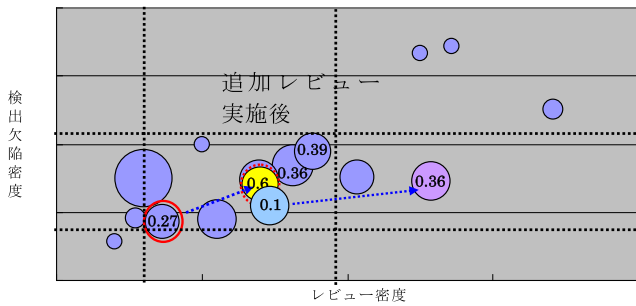


図9 ピアレビュー網羅率（後半）付ゾーン分析結果

Fig. 9 Result of zone analysis with the peer-review-coverage.

レビューを実施した。その結果、新たに欠陥が発見された。結果として図9の矢印の位置へ移動し、ピアレビュー網羅率が0.27（レビュー密度：0.36，欠陥検出密度：0.43）からピアレビュー密度0.60（レビュー密度：0.65，欠陥検出密度：0.71）へ向上した。また別成果物において、レビュー網羅率（後半）が、0.1（レビュー密度：0.67，欠陥検出密度：0.53）から0.36（レビュー密度：1.29，欠陥検出密度：0.73）へ向上した。

このように、提案する品質評価技法の適用により、ピアレビュー網羅率（後半）の低い作業成果物を検出することにより、後半部分に対するレビュー実施漏れのある成果物を見逃さず、追加レビューを実施させる管理が可能となった。

結果として、適用サブプロジェクトのテスト段階での欠陥検出数を過去実績比で34%削減できた。ここで、比較対象の過去プロジェクトは、本技法の適用を除き、同様な開発体制、スタイル、および品質基準で開発を実施している。このため、コーディング差異などによる欠陥の防止効果などの影響は小さいと考えている。

また、開発者へのヒアリング結果から、本技法を用いることで、ピアレビュー網羅率を用いて追加レビューを行った場合、後半部分に重点的なピアレビューを行うことができ、欠陥検出を効果的に実施できる点が効果としてあげられた。さらにスケジュール制約が厳しい3件の工事に対する適用において、適用実施したプロジェクト管理者から、プロジェクト残存欠陥の可能性の高い箇所に集中的対策を行うことができるとの評価を受けた。

(2) 見い出された課題

適用プロジェクトにおいて、ケース数は少ないが、ピアレビュー網羅率が低いため追加レビューを実施したのに新たに欠陥が検出できない場合があった。これは、ピアレビュー網羅率が、成果物の潜在欠陥数が元々少ない場合でも、低い値をとることが原因であった。

ピアレビュー網羅率を用いて追加レビューの要否を判断する際には、その絶対値が基準値を満たすか満たさないかで単純に評価せず、ピアレビュー網羅率の前半対後半比が高いかどうかなどを判断材料として加えることが必要であると考える。

(3) 技法の適用範囲と限界

本技法は、1回のレビューにおいて対象成果物を前から後ろへ読み進むことを前提としている。そのため、部分に分けて均等に時間配分を行うやり方や、レビュー対象をサンプリングするやり方のレビューには適用できない。

なお、本技法は、上記前提を満たせば、レビューを成果物の途中から読み始める場合にも適用可能である。ピアレビュー網羅率の計算では、開始ページ番号を m 、レビュー終了ページ番号を n として、式(3)から式(5)の N 以下を当てはめればよい。

$$N = n - m \tag{6}$$

また、本技法は、レビュー対象成果物に対して均質に欠陥が分布していることを前提としている。このため、たとえば、前半部分に欠陥が偏重する場合においてもピアレビュー網羅率が前半でより高くなってしまい、追加レビューの判断を誤る可能性がある。

この前提については、3.1.2項において、この開発現場において過去の実績が欠陥の前後半での均質さを裏付けていることを示した。しかし、レビュー対象成果物の部分ごとに、成果物の複雑度が異なる場合、作成担当者あるいはかけた設計工数が異なる場合などは有意差が現れる可能性があるため、技法適用時に考慮する必要がある。

5. 関連研究

ピアレビューの実施結果から品質を評価する技法として、2章で述べたゾーン分析法以外に、検出欠陥数の実績から残存欠陥数を推定する欠陥数推定法がある。以下に、欠陥数推定法のアプローチおよびレビューの品質評価への適用上の特徴と制約をまとめ、提案する技法との関係について考察する。

5.1 欠陥数推定法とレビュー評価への適用上の特徴と制約

(1) Capture-Recapture モデル[14], [15]

同一対象・同一観点での複数人による独立したレビューの結果から残存欠陥数を推定する技法である。ある地域での野生動物の生態数を求める統計手法を適用し、地域をレビュー対象に野生動物を欠陥に置き換えて、レビュー対象に含まれる総欠陥数を推定する。レビュー者1とレビュー者2により検出された欠陥をそれぞれ $n1$, $n2$ とする、そのうち共通の欠陥を m とすると総欠陥数は式(7)で計算できる。

$$N = n1 \cdot n2 / m \tag{7}$$

残存欠陥数は式(8)で計算する。

$$R = N - (n1 + n2 - m) \tag{8}$$

Capture-Recapture モデルには、レビュー者の検出能率を可変と考える推定法 (Mt)、欠陥の検出確率を可変と考

える推定法 (Mh) などのバリエーションがある。Vander Wiel ら [15] はシミュレーション結果から、Mt が Mh に比較して良い推定結果を得ること、Mt が欠陥クラスを分類することにより推定結果に改善がみられることを示した。Briand ら [14] は実適用のデータに基づき全バリエーションを評価し、以下の結果を報告している。

- ① 4 以上の独立レビューを実施しないと良い推定値を得られない。
- ② 推定値 (中央値) が低く出る傾向がある。これを調整するバリエーションは安定性が悪い (極端に高い推定値を出すことがある)。

現場での実運用を考慮すると、① をつねに前提にすることは適用上難しい。② は残存欠陥数を過少評価することになり、再レビューの必要性を判断する目的には合わない。

(2) 信頼度成長モデルによる外挿法

レビュー中に検出した欠陥のタイムスタンプを使い、レビュー時間に対する累積欠陥検出数をプロットし信頼度成長曲線にフィッティングすることで残存欠陥数を推定する技法である [16]。Goel らは、信頼度成長曲線として Goel-Okumoto のモデル [17] を採用している。Goel-Okumoto のモデルは、レビュー時間 t での累計検出欠陥数を $D(t)$ 、検出能力を k とすると、式 (9) となる。

$$D(t) = N(1 - e^{-kt}) \tag{9}$$

総欠陥数 N と検出能力 k は実データを使って推定する。Goel-Okumoto のモデルは、レビュー者グループの検出能力は一定と仮定し、レビュー時間を経るとレビュー対象成果物は残存欠陥数が減少するため検出効率 (件/h) は指数関数的に減少していくものとしている。

レビューにおいて、検出能力を時間にわたって一定とする前提は適切ではない。実際のレビューは、複数人が並行して進める事前準備段階と、レビュー者が集まり会議形式で行う段階からなることが多いが、事前準備段階は正確な実施時間を計画・把握することはできないため、タイムスタンプをとることで検出能力一定の前提をおくことができない。また、会議形式においてもレビュー参加者が変化することがあるためこの前提は崩れており、この技法をレビューの品質評価に用いる場合には、こうした実施状況に注意してデータを取り扱う必要がある。

なお、山田 [19] は、テストプロセスでの検出能力の時間的変化を考慮した信頼度成長モデルとして習熟 S 字形モデルを提案しているが、レビューへの適用は報告されていない。

(3) 欠陥プロファイル推定法 [18]

複数レビュー者が同一対象を独立してレビューする。検出された欠陥を、その欠陥を検出したレビュー者の数で値づけし、その値の多い順にソートする。図 10 に示すように、指数関数にフィッティングし、閾値 0.5 を設定してそれ以下になった欠陥番号が総欠陥数の推定値とする方法である。

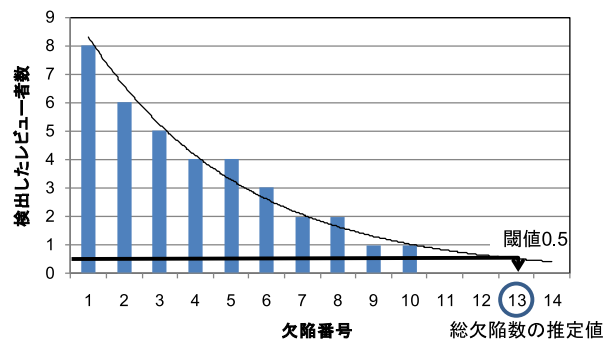


図 10 欠陥プロファイル推定法
Fig. 10 Defect profiling estimating method.

Biffi ら [18] は、欠陥プロファイル推定法が、リーディング法の影響も受けにくく、また平均誤差および標準偏差で安定した結果を得ることができたと報告している。しかし、この推定法では、指数関数へのフィッティングを用いているが、指数関数を選択する理論的根拠は示されていない。その点が信頼度成長曲線の場合と大きく異なる。

5.2 欠陥数推移法と提案技法との関係

5.1 節で述べたように、Capture-Recapture モデルは種々の制約から追加レビューの必要性を判断する目的には利用しにくい。これに対して、信頼度成長モデルによる外挿法および欠陥プロファイル推定法は、制約事項はあるものの、この目的に利用することが可能である。

しかしながら、これらの欠陥数推移法は、ゾーン分析法と同様に、成果物全体に対する平均化された情報でのみ評価を行っている。このため、レビュー実施の不均一さを見逃してしまう可能性がある。我々が提案するピアレビュー網羅率を用いることで、これらの欠陥数推移法の収束判断を補完することができる。特に、信頼度成長モデルによる外挿法は、レビューの進捗にともなう基準値付近での推定残存欠陥数の変化 (つまり収束具合) を判断に用いるが、これにピアレビュー網羅率 (および前半対後半比) の変化を見ることで、レビュー実施の不均一さがないことを見ることができると考える。

6. おわりに

本論文では、ピアレビューの結果に基づく従来の品質評価技法が、成果物全体に対する平均化された情報のみに基づいているため、レビューが均質に実施できていない場合に欠陥を多く見逃してしまうという問題を明確にした。この問題を解決するために、レビュー実施の均質さを測定するための指標として、ピアレビュー網羅率を定義し、その指標を用いて従来技法を補完する品質評価技法を提案した。本指標を、レビュー網羅率 (前半) とレビュー網羅率 (後半) に分離することで、後半部分のレビュー実施漏れを推定することができる。さらに、提案技法によるレビューの

品質管理を実プロジェクトへ適用し、レビュー実施漏れを防止しテスト段階への欠陥流出を減少する効果を確認し、提案技法の有効性を示した。

今後は、ピアレビュー網羅率とテスト段階における欠陥数のデータを蓄積し、両値の相関を分析することで、ピアレビュー網羅率を品質判断の基準値として用いるうえでの適値範囲を決定していく予定である。

参考文献

- [1] Kan, S.H.: *Metrics and Models in Software Quality Engineering* (2nd Edition), Addison-Wesley Professional (2002).
- [2] 中島 毅, 東 基衛: ソフトウェア開発における品質プロセスのコスト最適化のためのモデルとシミュレーションツール, 電子情報通信学会論文誌 D, Vol.J91-D, No.5, pp.1216-1230 (2008).
- [3] 織田 巖: ソフトウェア・レビュー技術, ソフトウェア・リサーチ・センター (2006).
- [4] Gilb, T. and Graham, D.: *Software Inspection*. 伊土誠一, 富野 壽 (監訳): ソフトウェアインスペクション, 構造計画研究所 (1999).
- [5] D. フリードマン, G. ワインバーグ: ソフトウェア技術レビューハンドブック, TBS 出版会 (1987).
- [6] 森崎修司: ソフトウェアインスペクションの動向, 情報処理, Vol.50, No.5, pp.377-384 (2009)
- [7] 野中 誠: 設計・ソースコードを対象とした個人レビュー手法の比較実験, 情報処理学会研究報告, Vol.2004, No.118, pp.25-32 (2004).
- [8] 久野倫義, 丹羽友光, 前川隆昭: デザインレビューの効果的な実施方法, 第 26 回ソフトウェア品質シンポジウム (2006).
- [9] 細川宣啓: 第三者インスペクションによる品質検査と欠陥予測, 情報処理, Vol.50, No.5, pp.405-411 (2009).
- [10] ソフトウェア・エンジニアリングセンター: 定量的品質予測のススメ (2008).
- [11] 野中 誠: ソフトウェアインスペクションの効果と効率, 情報処理, Vol.50, No.5, pp.385-390 (2009).
- [12] 小室 睦ほか: 開発現場の実態に基づいたピアレビュー手法の改善と改善効果の定量的分析, SEC Journal, No.4, pp.6-15 (2005).
- [13] 飯山俊介: 設計レビュー指標値の算出, 第 28 回ソフトウェア品質シンポジウム (2008).
- [14] Briand, L.C., Eman, K.E., Freimut, B. and Laitenberger, O.: Quantitative Evaluation of Capture-Recapture Models to Control Software Inspections, *Proc. 8th International Symposium on Software Reliability Engineering (ISSRE'97)*, pp.234-245 (1997).
- [15] Wiel, S.V. and Votta, L.: Assessing Software Designs using Capture-Recapture Methods, *IEEE Trans. Software Engineering*, Vol.19, No.11, pp.1045-1054 (1993).
- [16] Goel, A.L.: Software Reliability Models: Assumptions, Limitations, and Applicability, *IEEE Trans. Software Engineering*, Vol.11, No.12, pp.1411-1423 (1985).
- [17] Wohlin, C. and Runeson, P.: Defect Content Estimations from Review Data, *Proc. 20th International Conference on Software Engineering*, pp.400-409 (1998).
- [18] Biffi, S. and Halling, M.: Investigating Reinspection Decision Accuracy Reading Product-Quality and Cost-Benefit Estimates, *Proc. 25th Annual International Computer Software and Applications Conference (COMPSAC'01)*, pp.87-96 (2001).

[19] 山田 茂: ソフトウェア信頼性モデル, 日科技連 (1994).



久野 倫義 (正会員)

1989 年筑波大学大学院理工学研究科修士課程修了。同年三菱電機 (株) 入社。現在, 三菱電機 (株) 設計システム技術センターグループマネージャ。ソフトウェア生産技術に関する研究開発に従事。



中島 毅 (正会員)

1984 年早稲田大学大学院理工学研究科電気工学修士課程修了。同年三菱電機 (株) 入社。1991 年から 1 年間米国コロラド大学コロラドスプリングス校客員研究員。2008 年早稲田大学大学院博士課程修了, 博士 (工学)。現在, 三菱電機 (株) 設計システム技術センター主管技師長, ソフトウェア生産技術に関する研究開発に従事。著書に『IT Text ソフトウェア開発』(共著, オーム社)。技術士 (情報工学/総合技術監理)。IEEE Computer Society, 電子情報通信学会, 電気学会各会員。