

ピアレビュー網羅率を用いた 品質予測技法の提案

久野倫義[†] 中島毅[†]

本論文は、従来のピアレビュー指摘数やレビュー工数を用いた品質予測技法（残存欠陥数予測）における課題を明らかにし、ピアレビュー網羅率を用いたピアレビュー漏れを防止する品質予測技法の提案を行い、その適用結果を示したものである。

A Method For Software Quality Prediction Using the Peer-Review-Coverage Measure

Noriyoshi Kuno[†] and Tsuyoshi Nakajima[†]

In this paper, we propose a new method for software quality prediction using the peer-review-coverage measure after addressing the problem of existing quality prediction methods using the number of defects detected in the peer review and the effort for the review. We also shows the evaluation result for applying the method to real projects.

1. はじめに

ソフトウェア開発を行う上でのピアレビューは、品質向上の重要な鍵であり、多くの手法が提案されてきている¹⁾²⁾³⁾⁴⁾。ピアレビューの手法であるインスペクションは、不具合検出に最も効果があるといわれているが、その導入と実践において様々な課題がある。導入においては、導入前と比べて開発者の設計段階での負荷が高いことが挙げられる⁵⁾。実践においては、①レビュー実施の不徹底、②レビュー時間の浪費等が挙げられる¹⁾。これらの課題を解決すべく、我々は社内におけるピアレビュー会議時間に着目し、ムリムダムラを防止するピアレビューの改善を行ってき効果を上げてきている⁵⁾。

上記の改善活動を通じて、テスト段階で発見される不具合の要因として、レビュー漏れのものが多いことがわかった。この原因は、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。従来のレビュー実施結果の評価方法は、作業成果物全体に対してレビュー指摘密度での評価を行っているが⁶⁾⁷⁾、この評価方法ではソフトウェアの残存欠陥を正しく推定できないためである。その結果、流出した欠陥が、テスト段階で多発しリワークにより納期通りに出荷できないことや、テスト段階でも検出されずに市場での不具合発生が生じるというリスクがあった。

我々は、レビュー実施結果に関する従来の評価方法がもつ問題点を解決するため、レビューが作業成果物全体にわたって均一に行われたかを定量化する新たな指標を導入し、ソフトウェアをレビュー段階で適切に評価できるようにすることで、開発の早い段階で品質に関するリスクを軽減する取り組みを行った。本論文では、従来のピアレビュー評価方法がもつ問題点を解決する評価方法を提案し、その適用結果を示す。

2. 従来のピアレビュー評価方法の問題点

2.1 従来の評価方法

ピアレビュー結果に対して、従来の評価方法は、作業成果物全体に対するレビュー結果を指標として使い、レビュー後の成果物の残存欠陥数を予測している⁶⁾⁷⁾。従来の評価方法の代表的なものには、以下の3つがある。

① 目標レビュー指摘件数の達成の評価

レビューで検出する指摘数に目標を与えることで、それを達成するために必要な、一定以上のレビュー工数投資及びレビューの効率向上に対する動機付け

*[†] 三菱電機株式会社
Mitsubishi Electric Corporation

となる。残存欠陥数の予測の精度は低い。

② 指摘内容の分類別指摘件数による有効性の評価

指摘内容を分類し、誤記や軽微な誤りの比率から間接的に成果物の残存欠陥数を予測する。セルフレビューで誤記や軽微な誤りは見つけやすく、セルフレビューの実施度合いによって残存欠陥数の予測にバラツキがでる。

③ ゾーン分析法：レビュー工数密度とレビュー指摘密度の組合せによる評価

レビュー工数密度（工数／ページ）すなわちレビューをどれだけやったかと、レビュー指摘密度（件／ページ）すなわちそれによってどれだけ欠陥を検出できたかの2軸をとり、それぞれ目標値を置き、目標値からのずれで成果物の残存欠陥数の評価を行う。①よりレビュー実施度の要素を加味している点、②のように評価がセルフレビュー実施度の影響も受けにくい点で優れた方法である。

2.2 ゾーン分析法の問題点

本節では、ゾーン分析法について説明し、その適用上の問題点を挙げる。

図 2-1 は、4つの機能仕様書に関して、レビュー指摘密度（件／ページ）とレビュー工数密度（工数／ページ）を散布図にしたものである。図 2-1 では、機能仕様書 A に対しては、他機能に比べてレビュー工数密度が低いため、追加レビューを行うことという施策を実施する。また仕様書 B に関しては、レビュー工数密度が低いにも関わらず他に比べ高いレビュー指摘密度になっており、成果物の作成方法などの点検を行うという施策を行う。つまり、短時間で多くの指摘件数を抽出した場合やレビュー工数が少ない場合には、成果物の品質が悪いと予測し、多くの時間をかけるほど、成果物の品質向上ができるという評価方法である。

ゾーン分析法での大きな問題点は、レビュー工数密度とレビュー指摘密度が目標値を満たしている場合でも(図 2-1 では機能仕様書 C と D)、テスト段階で検出された欠陥の中には、明らかにレビュー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ることである。

レビュー漏れが生じる原因の一つとしては、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、欠陥検出のための時間が取れていないことがある⁵⁾。このため、レビュー工数とレビュー指摘密度の関係では、成果物品質を適切に評価できないという課題がある。これに関しては、レビュー会議の方法を見直すことで解決を図ってきているが⁶⁾、改善後においても、レビュー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。

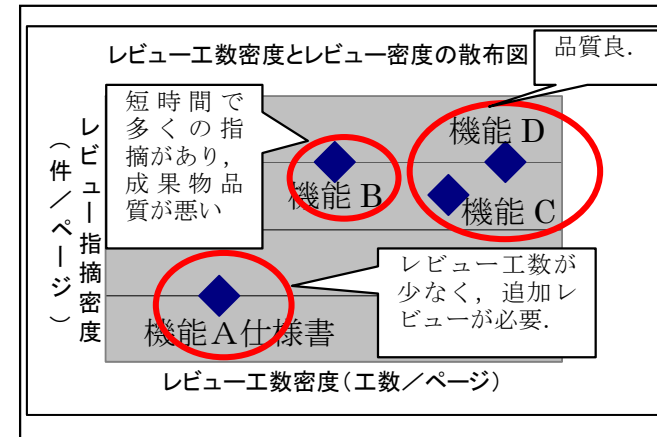


図 2-1 レビュー工数とレビュー指摘件数の関係イメージ

2.3 ゾーン分析法におけるレビュー漏れの原因の分析

本節では、前節のレビュー漏れの原因を分析する。

図 2-1 のような分析において品質良好と判断された機能のうち、テスト段階で不具合が多発したものと、そうでなかったものの比較を行った。比較方法は、全てのレビュー記録の確認とした。レビュー記録には、①頁番号、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 2-2 と図 2-2 は、単一の作業成果物に対して、複数レビューアがピアレビューを行った結果から、ページ毎の指摘件数を集計して示したグラフである。図 2-2 は、テスト段階でレビュー漏れの欠陥が多かった機能部分、図 2-2 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 1 件としてカウントした。

レビュー漏れの欠陥が多かった機能部分(図 2-2)では、作業成果物の前半部分に多くの指摘があるが、後半部分において、レビュー指摘数が 0 件のページが多くなっている。これに対して、レビュー漏れの欠陥が多かった機能部分(図 2-2)では、前半から後半まで満遍なく指摘が上がっている。この 2 つのデータより、図 2-2 のケースでの欠陥流出の原因は、成果物の後半部分でレビューが十分に実施されていないか、検出効率が落ちているかのどちらかである。前半と後半部分の差異は読み進む順序外に見当たらないので、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のが原因と考えられる。

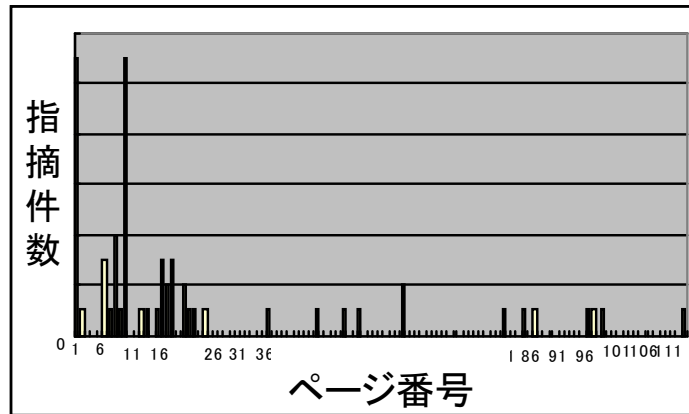


図 2-2 ページ番号と指摘件数の関係（欠陥の多かった機能）

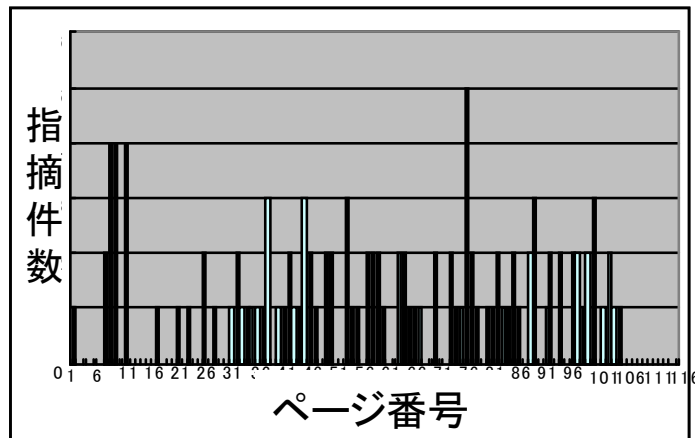


図 2-3 ページ番号と指摘件数の関係（欠陥の少なかった機能）

上記図 2-2 図 2-3 で示されるような問題点は、ゾーン分析法による評価では、顕在化できていない。上記 2つの機能でのレビュー指摘密度は、以下であった。

欠陥の多かった機能のレビュー指摘密度 = 0.51 件/ページ

欠陥の少なかった機能のレビュー指摘密度 = 1.02 件/ページ

本作業成果物に対するレビュー指摘密度の目標値は、0.5 件/ページであり、両機能とも目標を達成していたため、問題なしを判断された。

これは、ゾーン分析に代表される従来のレビュー実施結果の評価方法が、作業成果物全体に対するレビュー指摘密度でのマクロな評価を行っていることから発生する。従って、例えば特定ページで多くの指摘があった場合、その他のページでレビューがされていなくても、全体的にはレビュー指摘密度が目標通りという評価となってしまうからである。

3. ピアレビュー評価手法を用いた品質予測技法の提案

3.1 目的

2 で述べたように、仕様書のレビューにおいて、レビュー漏れを防止し、テスト段階への流出欠陥数を削減するためのレビュー評価手法を提案する。レビュー漏れを防止する方法は、レビュー指摘内容を確認し、レビューの有効性を判断するという方法がある。しかし、全ての指摘内容を確認することは困難であり、そこで、マトリクスを用いて、ピアレビューの有効性を判断することで品質を予測する方法を検討した。

3.2 ピアレビュー網羅率

ページ毎の指摘件数の有無による評価のため、ピアレビュー網羅率という指標を導入する。ピアレビュー網羅率の定義を次式に示す。

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

ここで、 D_i : ページ毎の指摘有無（有は 1 か無は 0 で示す）

i : ページ番号

N : 総ページ数

上式で示すように、ピアレビュー網羅率は、ピアレビュー対象作業成果物のページ毎の指摘の有無により、レビューされたかどうかを判断するものである。例えば、10 ページに 3 件の指摘があった場合には、10 ページの網羅率は 1 となり、11 ページの指摘が 0 件であれば、網羅率は 0 となる。これをピアレビュー対象の全ページで加算し、総ページ数で割ったものを、作業成果物のレビュー網羅率とする。過去に実施した成果物に対するピアレビュー網羅率を測定した結果を表 3-1 に示す。

表 3-1 レビュー網羅率結果

成果物 番号	総ページ数	指摘件数	網羅率
1	1	4	1.00
2	4	9	0.75
3	1	0	0.00
4	24	39	0.96
5	1	0	1.00
6	1	15	1.00
7	106	91	0.53
8	40	51	0.58
9	61	76	0.46
10	106	139	0.50
11	2	1	0.50
12	1	1	1.00
13	1	0	0.00
14	1	1	1.00
15	1	4	1.00
16	1	1	1.00
17	1	0	0.00

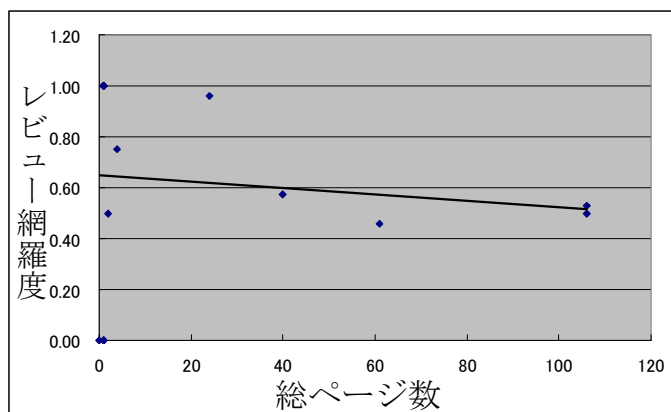


図 3-1 総ページ数とレビュー網羅率(D)の関係

表 3-1 から、ピアレビュー網羅率に関して、以下の考察ができる。

- 総ページ数が少ない場合、時間切れによるレビュー漏れの発生は少ない。
- 作業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漏れとの相関を見るために、作業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 3-1 に示す。サンプル数は少ないが、

図 3-1 より、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、レビュー漏れになり易い傾向があることがわかる。

またピアレビュー網羅率を作業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した 2 つの指標を以下のように定義する。

$$\text{ピアレビュー網羅率 (前半)} \quad Df = \sum (Di) / (N/2) \quad i=1 \dots N/2$$

$$\text{ピアレビュー網羅率 (後半)} \quad Db = \sum (Di) / (N/2) \quad i=N/2+1 \dots N$$

ここで、N: 総ページ数

上記定義に基づき、表 3-1 の作業成果物 7,8,9,10 に対して、計算した結果を表 3-2 に示す。

表 3-2 作業成果物 7,8,9,10 に対するピアレビュー網羅率

成果物 番号	総ページ数	指摘件数	網羅率	前半網羅率 (Df)	後半網羅率 (Db)
7	106	91	0.53	0.32	0.21
8	40	51	0.58	0.43	0.15
9	61	76	0.46	0.33	0.13
10	106	139	0.50	0.35	0.15

表 3-2 から、成果物全体でのピアレビュー網羅率 D より、ピアレビュー網羅率 (前半) : Df とレビュー網羅率 (後半) : Db という指標を導入することで、 Df と Db の関係より、作業成果物の後半部分でのピアレビュー網羅率が低いことがわかり、後半での漏れが発生し易い状況を示すことができる。

また、図 3-2 は、作業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。本グラフから作業成果物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することがわかる。原因としては、前半にレビュー時間をかけたため、後半のレビューが十分時間をかけなかったことが推定できる。

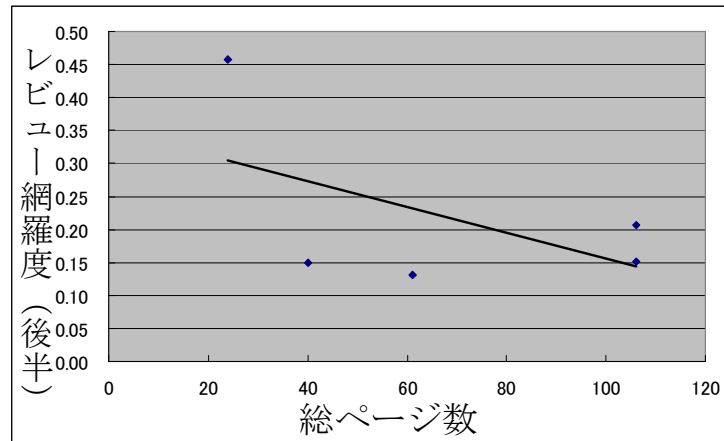


図 3-2 総ページ数と後半レビュー網羅率 (Db) の関係

4. 適用と評価

提案技法を実プロジェクトに適用し評価を実施した。

図 4-1 は、あるプロジェクトでのゾーン分析の結果である。図中の実線丸印で囲んだ機能部分は、ピアレビュー網羅率の評価により、後半のページ部分でレビュー漏れが発生していると判断された。その結果をプロジェクトリーダーへ通知し、該当機能部分に対して追加レビューが実施された。その結果、新たに欠陥が発見され、ゾーン上で点線の丸印で囲んだ位置へ移動した。このように、本技法を用いることによって、テスト段階へのレビュー漏れによる欠陥の流出を防止することが可能となった。

実際の適用にあたっては、以下のようにピアレビュー網羅率を利用した品質判断を行った。

- レビュー工数が目標に満たない場合、レビュー網羅率を確認する。
- 前半網羅率と後半網羅率の差が 2 倍程度であれば、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。

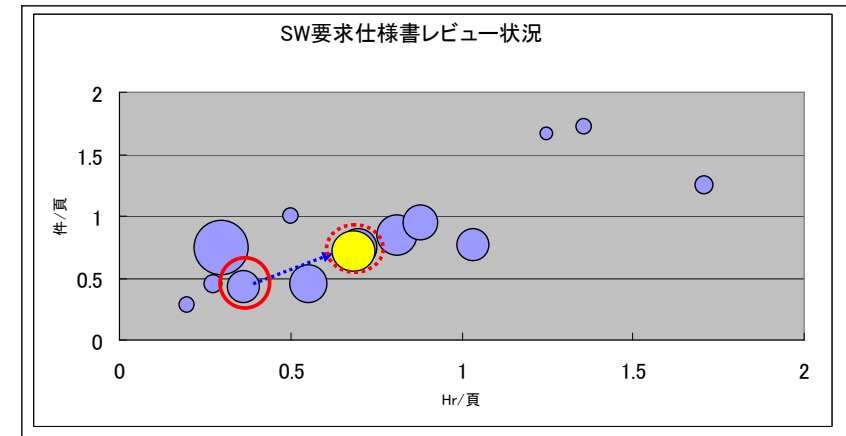


図 4-1 本技法の適用例 (有効であった場合)

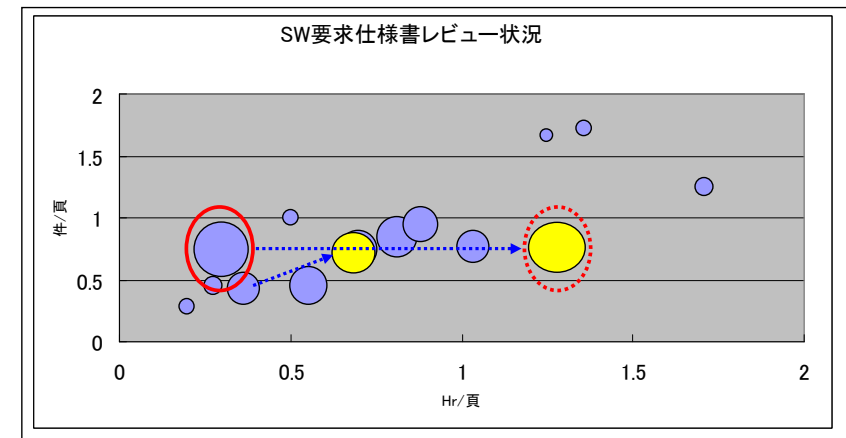


図 4-2 本技法の適用例 (有効でなかった場合)

図 4-2 に示すように、ピアレビュー網羅率による評価結果から追加レビューを実施しても、追加の指摘が検出できない場合もあった。ピアレビュー網羅率は、成果物品質そのものを測定してはいないため、ピアレビュー網羅率をどこまで向上させれば良いかという最適な値を見つけるには、実績データの蓄積が必要である。精度向上のためには、実績データの蓄積が必須であると考えられる。

5. おわりに

本論文では、ピアレビューの結果に基づく品質評価において用いる従来の予測手法、すなわち成果物全体に対するレビュー結果を指標として使いレビュー後の成果物の残存欠陥数を予測するやり方の問題点を明らかにし、成果物の部分へのレビュー実施程度を測定するための指標、ピアレビュー網羅率を定義し、その指標を用いた品質予測技法（残存欠陥数予測技法）を提案した。成果物の後半部分へ指標を適用し全体との比率をみることで、後半部分のレビュー漏れを推定することができる。

さらに、実プロジェクトへの適用結果から、ピアレビュー網羅率を用いてレビュー時間不足によるレビュー漏れを防止できることを示した。ピアレビュー網羅率のレビュー漏れを防止への利用方法例を示した。今後は、レビュー網羅率とテスト段階における不具合数のデータを蓄積し、両値の相関について調査することで、定量的なレビュー網羅率の最適値を検討していく予定である。

参考文献

- 1) 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）
- 2) T.Gilb and D.Graham(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）
- 3) D. フリードマン，G. ワインバーグ（岡田正志 監訳）：ソフトウェア技術レビューハンドブック（T B S 出版会，1987）
- 4) GLASS, R L. : Inspections. Some Surprising Findings. Communications of ACM Vol.42, No.4, Page17-19 (1999.04)
- 5) 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（日科技連，2006）
- 6) 中野裕也，他：コードレビューの密度と効率がコード品質に与える影響の分析（SEC journal 2008.8）
- 7) 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）。