レビュー記録のテキスト分析結果による多角的な品質管理方法の提案

Suggestion of a way of quality control by a multilateral viewpoint using the result of text analysis of a review record

佐藤　孝司
t-sato@ap.jp.nec.com
日本電気株式会社

発表要旨：

ウォーターフォール開発プロセスは，上流工程における成果物の完成度をできる限り高めることにより，当該工程以降に検出されたバグの修正による後戻り作業を低減するモデルといえる．そこで，品質管理部門では，上流工程の成果物の完成度の程度を測るための各種メトリクスを定義・運用して，品質上のリスクを見極める判断をしている．

筆者の組織では，プロセスメトリクスの一つであるレビュー工数の基準値を定め，レビュープロセスを確実に実行するように，組織的なプロセス標準を規程し品質管理を運用してきた．その結果，レビュー工数の増加と供に，出荷後のバグ数は減少し，一定の成果が得られたと考えられる．しかし，レビュー工数がある程度確保され，レビュープロセスが安定して実行されるようにプロセスが成熟した組織に対して，レビュー工数を軸にした品質管理だけでは品質リスクの見極めが難しくなってきた．

そこで，レビュー記録に着目し，レビュー記録のテキスト情報から得られるメトリクスについて設計品質に影響を与える要因を分析した．レビュー記録のテキスト情報から抽出したメトリクスには，バグの指摘率，レビュー指摘文の長さ，および，レビュー指摘文の述語の種類別頻出度などである．

本発表では，これらの分析結果をもとにして，レビューにおける多角的な品質管理方法について提案する．

キーワード：

品質管理，メトリクス，レビュー記録，テキスト分析

想定している聴衆

品質保証部門，プロジェクトリーダー

発表者の紹介（全角１００文字）：

日本電気株式会社入社以来、ＩＴ系ミドルソフトウェア/基本ソフトウェアなど汎用ソフトウェア製品の品質保証および CS 向上に従事。

Orchestrating a brighter world NEC

SQiPシンポジウム2017

レビュー記録のテキスト分析結果による多角的な品質管理方法の提案

2017年9月14日

佐藤　孝司　（NEC)

目次

1．品質プロセス概要

2．品質管理の効果・課題

3．新たなメトリクス探究

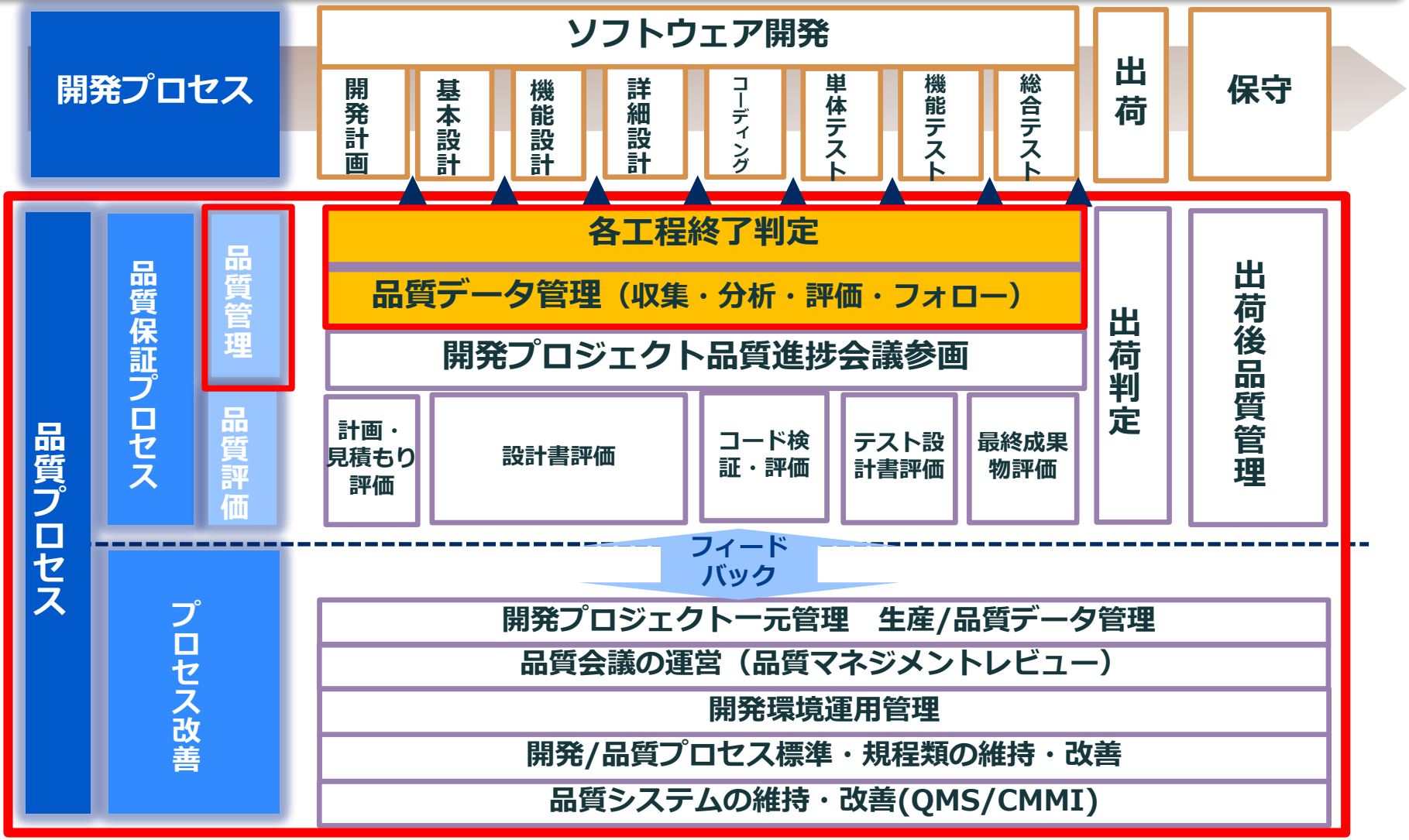
4．レビュー記録のメトリクス抽出・分析

5．分析結果の品質管理への応用

6．まとめ

品質プロセスの概要

① 個々の開発プロジェクトに密着した品質管理・品質評価（品質保証プロセス）
② 組織横断的な標準化・仕組みの改善（プロセス改善）



© NEC Corporation 2017

品質管理の特徴

- 定期的（基本は毎週）なメトリクスデータの収集・分析
- 各工程のメトリクスの実績値が基準値に達しているかを確認
- 品質保証部門が各工程の「工程終了判定」を実施

上流工程で品質確保（レビュープロセス定着化）のために，
レビュー工数密度とレビュー摘出バグ密度を重点的にフォロー

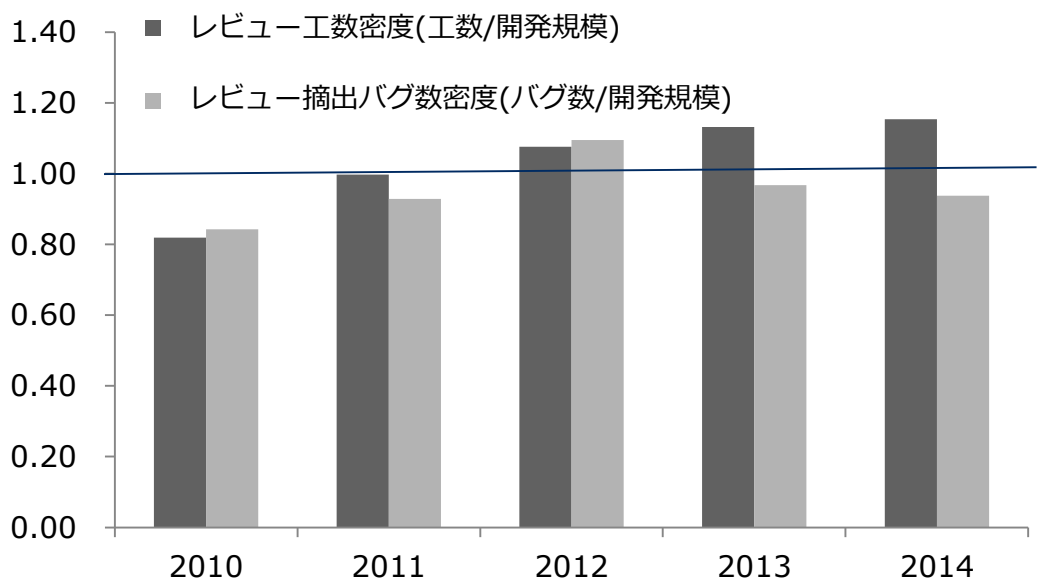
主なメトリクス一覧

No	メトリクス	算出式
1	レビュー工数密度	上流工程（設計，コーディング）のレビュー（設計レビュー，コードレビュー）工数（人・時間） / 開発規模 (KLOC)
2	レビュー摘出バグ密度	上流工程（設計，コーディング）のレビューで摘出したバグ数 / 開発規模(KLOC)
3	上流工程バグ摘出率	上流工程のレビューで摘出したバグ数/開発工程で摘出した総バグ数*100
4	上流工程工数密度	上流工程（設計，コーディング，レビュー）の工数（人・時間） / 開発規模 (KLOC)
5	テスト工程摘出バグ密度	テスト工程で摘出したバグ数 / 開発規模 (KLOC)
6	テスト項目数密度	テスト項目数 / 開発規模 (KLOC)
7	テスト工数密度	テスト工程の工数（人・時間） / 開発規模 (KLOC)

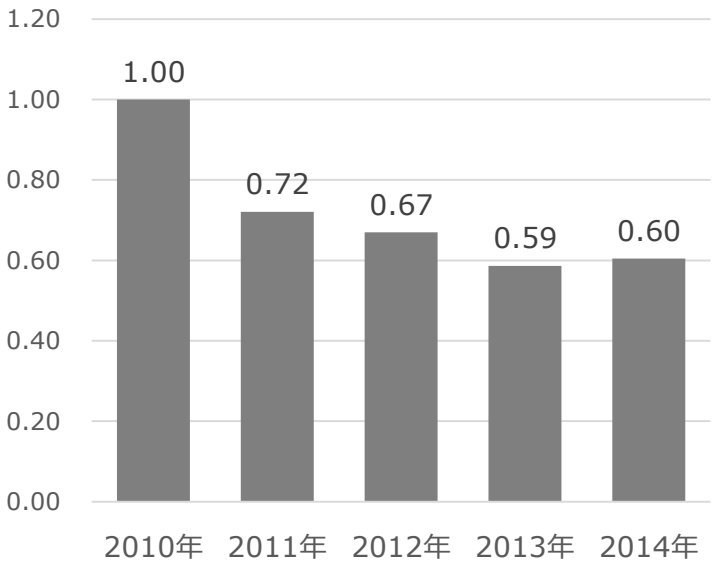
© NEC Corporation 2017

品質管理の効果

- プロセス改善活動の継続により，2012年にレビュー工数密度とレビュー摘出バグ密度は組織の目標値に到達
- 出荷後バグ数も低減し，フィールド品質が向上



レビュー工数密度とレビュー摘出バグ密度推移
（各メトリクス目標値に対する実績値比）



出荷後バグ数推移
（2010年実績値に対する比率）

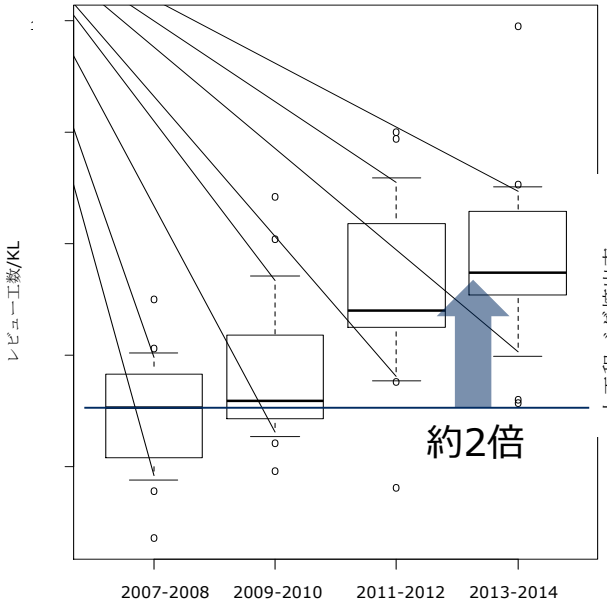
© NEC Corporation 2017

品質管理の課題①

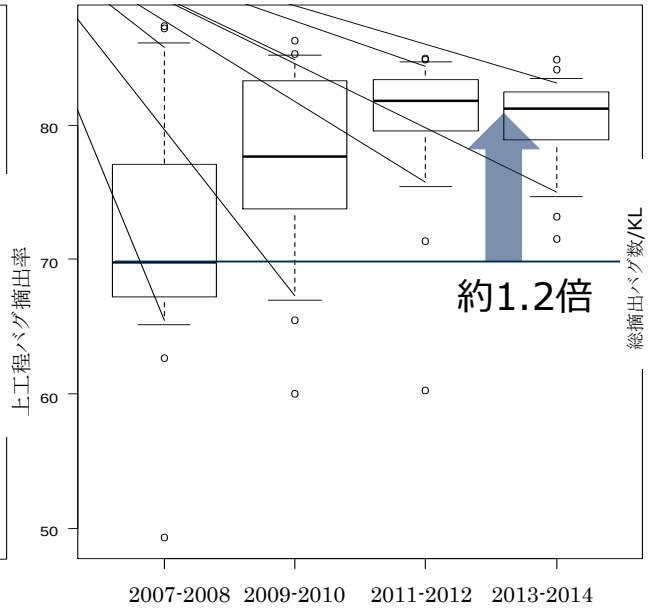
レビュープロセスの成熟に伴い，レビュー工数強化による効果に限界？

- レビュー工数密度の中央値が７年間で約2倍に増加
- 上工程バグ摘出率の中央値は７年間で約1.2倍に増加
- 開発総摘出バグ数の中央値も７年間で約1.2倍に増加

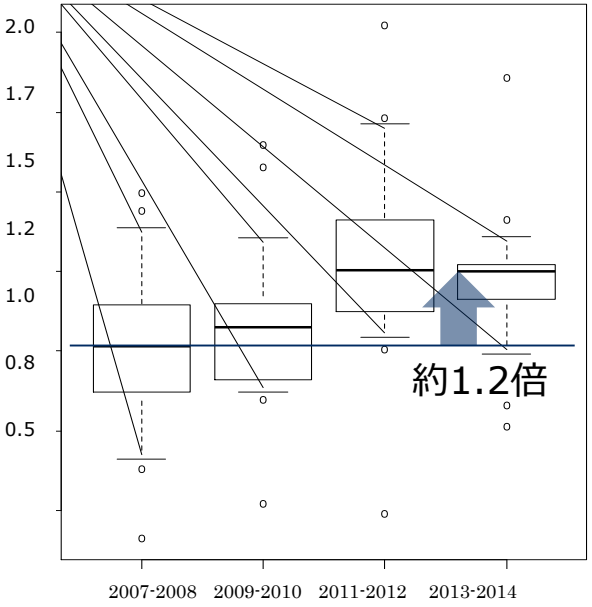
⇒レビュー工数密度の増加の割合に対して，上工程バグ摘出率および開発時の総摘出バグ数の増加の割合は小さい（レビュー成熟に従い効果が鈍化？）



レビュー工数密度の推移



上工程バグ摘出率の推移



開発時の総摘出バグ密度の推移
(“2007-2008”の値を1.0とした相対値)

© NEC Corporation 2017

品質管理の課題②

メトリクスの組み合わせの中で，品質判断が難しいケースがある

レビュー工数密度と摘出バグ密度の組み合わせ

	レビューが少ない	レビューが多い
摘出バグが少ない	A:レビュー不十分	C:品質が良い？
摘出バグが多い	B:品質が悪い	D:品質が悪い？

A	レビュー工数密度が少なく，かつ，摘出バグ密度も少ない場合	レビューが十分に実施されていないため，バグも検出されていないと考えて，まだ品質を判定できる状態ではなく，レビュープロセスを十分に実施すべきと判断する．
B	レビュー工数密度が少なく，かつ，摘出バグ密度が多い場合	品質が悪い状態であると判断して，レビューの上積みと，検出されたバグの傾向から，レビュー対象となる設計書の品質見直しを要求する．
C	レビュー工数密度が多く，かつ，摘出バグ密度が少ない場合	レビュープロセスは十分に実施されたと判断するが，十分なレビューを実施したにも関わらずバグが少ないのは設計品質が良いと判断すべきか，レビュー工数は多いけれどもレビューの質が低いため，十分にバグを摘出できていないと判断すべきか難しい．
D	レビュー工数密度が多く，かつ，摘出バグ密度も多い場合	レビュープロセスは十分に実施されたと判断するが，十分なレビューを実施した成果としてバグも十分に摘出できたと判断すべきか，それとも，設計品質がかなり悪いと推測して，まだ多くのバグが潜在していると判断すべきか，どちらの場合も想定できるため，品質の判定が難しい．

※“多い”や“少ない”と抽象的に表しているが，基準値比較や相対比較である等はここでは言及しない

© NEC Corporation 2017

新たなメトリクスの探究

品質管理の課題整理

課題①	レビュー工数効果の限界	プロセスが成熟した組織では，レビュープロセスを繰り返し安定的に実行できる状態から，さらにレビュー工数を増加させることで，品質をさらに向上させるという改善策には限界がある
課題②	メトリクスの組み合わせで，品質判断が難しいケースがある	レビュー工数密度と摘出バグ密度のメトリクスの組み合わせパターンの中で，レビュー工数が多いといっても品質が担保されたと判断することは難しい

新たなメトリクスを探し，品質管理のさらなる改良を図る

- **より品質に影響を与えるメトリクスを探す**
- **品質判断がより高度になるメトリクスを探す**
- **上流で品質を作り込むには，レビュー活動が重要**

→レビュー工数以外でレビューの質を測るメトリクスを探す

→レビュー記録に着目！

© NEC Corporation 2017
\Orchestrating a brighter world NEC

レビュー記録表からメトリクスを測定する意味

設計書ドラフト作成（レビュー前）において

- **完成度の高い設計書を作成**することは，設計者のスキルにかなり依存すると考えられる
- **完成度の低い設計書を作成**したとしても，多くの技術者がレビューすることで，最終的に成果物の品質や完成度を高めることができる

⇒レビューの質を上げることが最終的に設計品質を向上できる

レビュー記録表の情報からレビューの質を測る手法を検討

→レビュー指摘内容の文章の**述語に着目**

- 主語や目的語は固有名詞が多く，その開発に特化した技術用語が多く頻出する.
- しかし，レビュー指摘内容の文章の述語は，設計の誤りや設計の不足などの汎用的な表現が多く含まれるため，分類して計測が可能と考えた

PM学会2016年度秋季研修発表大会の発表資料を一部流用

© NEC Corporation 2017

レビュー記録表のテキスト文等から得られるメトリクス

- **レビュー流出バグ密度：**
  - **レビュー以降の開発中および出荷後に検出されたバグの中で，本来ならば当該レビューで検出すべきだったバグ**
- **レビュー指摘内容の述語の種類：**
  - **レビューの指摘文中に含まれる述語の意味から，期待する処置の種類により分類**

レビュー記録表から得られるメトリクス

メトリクス	意味と算出式
1レビュー流出バグ密度	レビューで見逃したバグ数 / 開発規模(KLOC)
2レビュー指摘密度	レビューで指摘した総数 / 開発規模(KLOC)
3バグ指摘率	レビューでバグを指摘した割合（バグ指摘数 / レビュー総指摘数）
4指摘文の長さ	指摘文の長さの平均；レビュー記録表に書かれた指摘文の総文字数 / レビュー総指摘数
5レビュー指摘内容	レビューで指摘した内容の分類；各分類の指摘数/レビュー総指摘数
(1) 述語の種類	レビューで指摘した文に含まれる述語の種類を以下の6種類の述語に分類する（重複可）
A) 間違いの指摘率	修正する,訂正する,改修する,改善する,変更する,更新する,反映する,対応する,対処する,処置する
B) 漏れの指摘率	漏れている，不足している,抜けている,加える,追加する,追記する
C) 要確認の指摘率	確認する，検討する，調整する
D) 削除の指摘率	削除する
E) 明確化の指摘率	明確にする
F) その他の指摘率	上記の5種類の分類に当てはまらない場合
(2) 質問率	レビューの指摘ではなく，質問の場合（アクションは必要ない）
(3) コメント率	レビューの指摘ではなく，単なるコメントの場合（アクションは不要）

PM学会2016年度秋季研修発表大会の発表資料を一部流用

© NEC Corporation 2017

レビュー記録表の質的要因の分析結果

7開発プロジェクト（下表A～G）のレビュー記録表を分析

- レビュー記録表の総指摘数3,439個を分析（測定対象の開発規模は約210KLOC）
- 「レビュー指摘内容の述語の種類」は，「漏れの指摘（設計不足）」が平均値0.315と一番多く頻出
- 二番目は，「間違いの指摘（設計誤り）」が平均値0.266

レビュー記録表メトリクスの測定結果

	開発規模 KLOC	レビュー総指摘数	レビュー流出バグ密度	レビュー指摘密度	バグ指摘率	指摘文の長さ	間違いの指摘	漏れの指摘	要確認の指摘	削除の指摘	明確化の指摘	その他の指摘	質問	コメント
A	6.06	137	0.825	22.607	0.350	83.503	0.569	0.386	0.073	0.007	0.131	0.000	0.065	0.175
B	23.36	284	0.941	12.157	0.235	69.225	0.331	0.415	0.133	0.035	0.098	0.007	0.056	0.112
C	76.34	1723	3.942	22.570	0.252	82.948	0.200	0.240	0.074	0.036	0.082	0.007	0.178	0.193
D	32.77	678	1.983	20.689	0.435	59.498	0.231	0.377	0.070	0.092	0.054	0.007	0.064	0.138
E	46.80	45	5.491	0.961	0.133	80.000	0.177	0.222	0.000	0.000	0.133	0.288	0.044	0.133
F	8.77	376	1.710	42.873	0.199	68.031	0.183	0.308	0.069	0.042	0.101	0.021	0.188	0.085
G	16.00	196	1.687	12.250	0.744	61.326	0.168	0.255	0.051	0.030	0.015	0.306	0.066	0.107
s	210.10	3439	-	-	-	-	-	-	-	-	-	-	-	-
m	30.01	491.28	2.368	19.158	0.335	72.076	0.266	0.315	0.067	0.034	0.088	0.088	0.095	0.135
sd	24.83	580.24	1.717	13.013	0.205	10.081	0.144	0.078	0.039	0.030	0.042	0.143	0.061	0.038

s:総和, m:平均値, sd:標準偏差

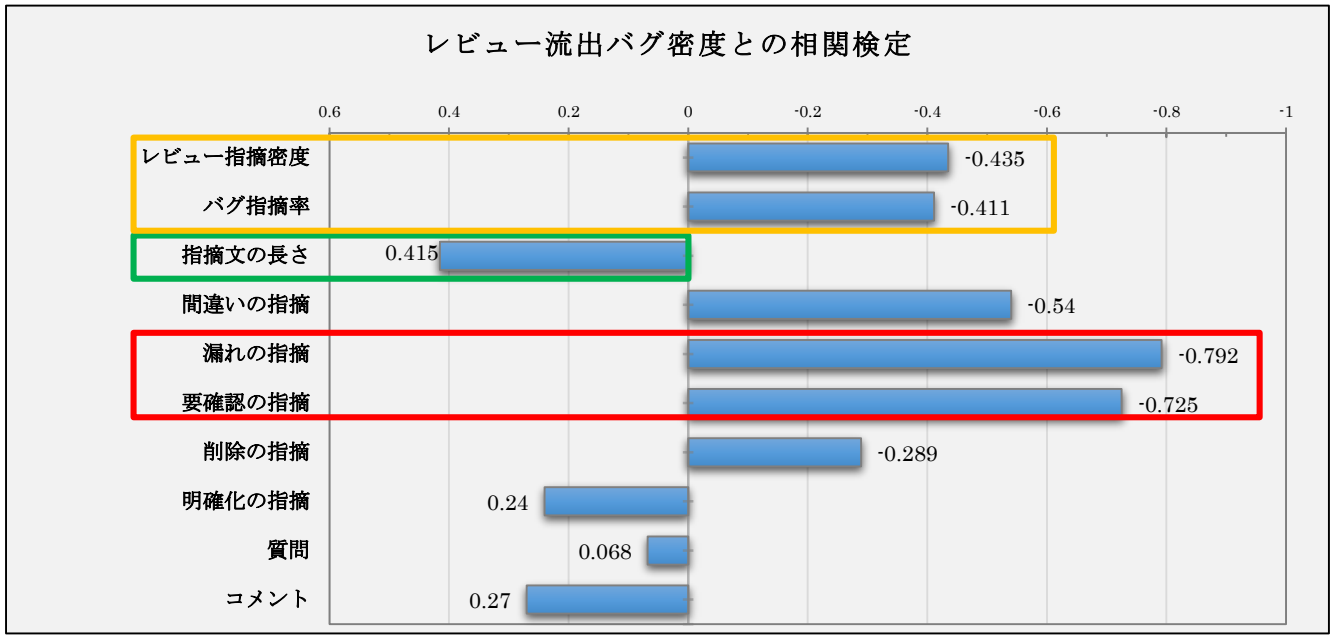
PM学会2016年度秋季研修発表大会の発表資料を一部流用

© NEC Corporation 2017

レビュー流出バグ密度との相関分析

- レビューの質を示す目的変数は，レビュー流出バグ密度
- レビューの質が高い（負の相関）変数は，漏れの指摘と要確認の指摘
- レビューの質が低い（正の相関）変数は，指摘文の長さ
- レビュー指摘密度やバグ指摘率も，レビューの質が良いことを表す（有意差なし）

レビュー流出バグとの相関係数	P値
-0.435	0.3298
-0.411	0.36
0.415	0.3544
-0.54	0.2108
-0.792*	0.0339
-0.725	0.0653
-0.289	0.5294
0.24	0.6038
0.068	0.8848
0.27	0.5575



レビュー記録表から得られたメトリクスとレビュー流出バグ密度との相関検定

PM学会2016年度秋季研修発表大会の発表資料を一部流用

© NEC Corporation 2017

レビュー記録表のテキスト分析結果の考察

「漏れの指摘」が多い場合はレビューの質が高い

- 設計書に**記述されていないことを指摘**することであり，記述されている事柄の誤りを指摘するより，**気づき難い高度な指摘**であるといえるので，質の高いレビューが実施されたものと推察

「要確認の指摘」が多い場合はレビューの質が高い

- 修正が必要であると断定したわけではなく，再検討や再調査などを要求して，**より設計の完成度を高める追求を期待する行為**であるため，質の高いレビューが実施されたものと推察

レビュー記録の指摘文が長い場合はレビューの質が低い

- 文章が冗長になり指摘内容の**要点を読み取ることが曖昧**になることから，レビューの質が低くなると推察

© NEC Corporation 2017

分析結果の品質管理への応用

- 必要時にレビュー記録を評価してレビューの質を判定することを応用すると，以下の品質判定を行うことができる

レビュー記録の評価結果と従来メトリクスの組み合わせ

レビュー工数もバグ数も多い場合	レビュー記録表を評価した結果，レビューの質が高い場合には，品質に問題なしと判断できる
	レビュー記録表を評価した結果，レビューの質が低い場合には，設計品質が悪い可能性があるので，バグの傾向分析等を実施して，再度，品質上の問題がないかを確認する必要があると判断する
レビュー工数は多いがバグ数が少ない場合	レビュー記録表を評価した結果，レビュー記録表を評価して，レビューの質が高い場合には，設計品質が高いので品質に問題なしと判断できる
	レビュー記録表を評価した結果，レビューの質が低い場合には，たとえレビュー工数が多かったとしてもレビューアのスキルが低いなどの要因により，質的に十分なレビューができていないと判断して，再度，レビューの参加者や進め方やレビュー対象範囲などには十分かを確認する必要があると判断する

© NEC Corporation 2017

まとめ

上流工程における品質判定方法を改善
品質判断が難しいケースを細分化し判断基準を明確化

レビュー工数密度と摘出バグ密度の組み合わせ

	レビューが少ない	レビューが多い
摘出バグが少ない	A:レビュー不十分	C:品質が良い？
摘出バグが多い	B:品質が悪い	D:品質が悪い？

改善

レビュー工数密度と摘出バグ密度とレビューの質の組み合わせ

	レビューが少ない	レビューが多い	
摘出バグが少ない	A:レビュー不十分	レビューの質が高い	C1:問題なし
		レビューの質が低い	C2:品質不十分
摘出バグが多い	B:品質が悪い	レビューの質が高い	D1:問題なし
		レビューの質が低い	D2:品質不十分

品質に影響を与える要因は多種多様で複雑に絡み合っている．
今後も，開発の早い段階（設計時）から品質を見極めるために，
多角的な分析を継続研究し品質管理のさらなる高度化を目指す．

© NEC Corporation 2017

Orchestrating a brighter world

NEC