

実践的ソフトウェア品質測定評価のための4つの「落とし穴」と7つの「コツ」：ゴール指向、不確実性、機械学習、実態調査ほか

鷲崎 弘宜

早稲田大学グローバルソフトウェアエンジニアリング研究所
国立情報学研究所、(株)システム情報

1. はじめに

ソフトウェアの品質を把握し、管理および改善するためには、測定を通じた評価が欠かせない。しかし、思慮なく測定評価を導入すると、誤った結論を導くなど、かえって悪影響をもたらしかねない。

本稿では、筆者が所長を務める早稲田大学グローバルソフトウェアエンジニアリング研究所が産学官連携により進めているソフトウェア品質測定評価の研究および実践結果に基づき、得られた知見を4つの「落とし穴」および7つの「コツ」として整理し解説する。

落とし穴とコツの対応を表・1に示す。本稿の以降において、それぞれの概要を紹介する。

表・1 ソフトウェア品質測定評価の落とし穴とコツ

落とし穴	コツ
負のホーン効果	<ul style="list-style-type: none">・ゴール指向・多面的な測定
組織上の不整合	<ul style="list-style-type: none">・組織構造と戦略に対する関係の可視化
不確実な未来	<ul style="list-style-type: none">・不確実性を組み入れた予測・機械学習に基づく測定システム改善PDCA
オレオレ品質	<ul style="list-style-type: none">・国際規格に基づく第三者評価と業界実態把握・各種パターン

2. ゴール指向に基づく多面的測定評価

負のホーン効果に陥ってはならない。これは、注目して測定した特徴は見かけ上は改善するものの、意図していない他の重要な側面に悪影響を及ぼすことを指す。ホーン工場における生産性向上の実験に由来する。当該実験では、いかなる要因を変更しても、さらに戻しても、生産性が改善した。注目して測定したために生産性は改善したと考えられる [1]。

その防止のために、測ることありきではなく、品質要求を目標（ゴール）として明確とし、その達成可否判定のために測定を導くべきである。これにより正のホーン効果が得られることが期待できる。そのようなゴール指向の枠組みとして Goal-Question-Metric (GQM) 法がある。GQM 法では、目標（Goal）を据えたうえで、目標達成を評価する質問（Question）およびその回答に必要なデータを得るための測定法（Metric）を決定する [1]。特に、抽象的で本来不可視なソフトウェアの品質を捉えるうえで重要である。

加えて、対象を多面的に捉えることが望ましい。これにより、ある側面の改善の過程で発生する思わぬ副作用を識別できる。

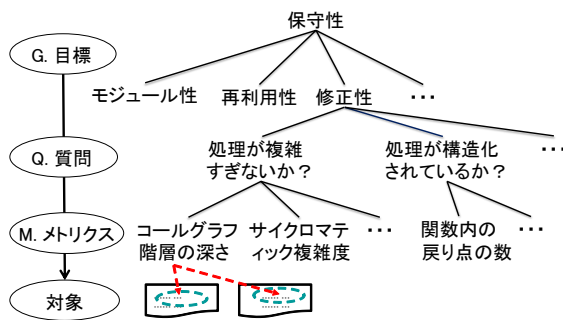
多面的な測定を経てプログラムソースコードの品質を評価する GQM モデルの例を図・1に示す。筆者らは類似モデルをオージス総研およびヤマハと共同で構築し、実サービスとして展開した実績を持つ [2][3][4]。

富士通コネクテッドテクノロジーズとは、OSS や実ソフトウェア製品の複雑さや欠陥、開発組織の変遷や開発者経験を考慮した評価の GQM モデルを構築し、活用を進めている [5][6]。ヤフーとは、様々な多面的な測定結果を可視化し総合的な意思決定に役立てるダッシュボード（例：図・2）の実装を進めている [7]。

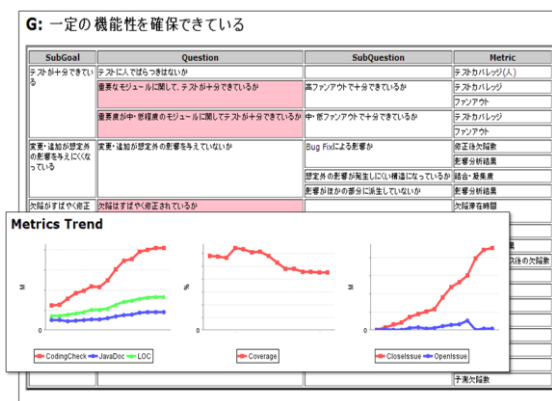
多面的な測定評価は、ソフトウェアプロダクトのみならず、ソフトウェア開発・運用プロジェクトについて

品質, Vol.46, No.3, 品質管理学会, 2016

でも重要である。筆者が日本支部の代表を務める SEMAT の提案する枠組み [8]は、「アルファ」と呼ばれるプロジェクトの進捗・健全性測定評価のための7つの視点を与えており、筆者らはその活用をシステム情報やインフォメーション・テクノロジー・アライアンス (ITA) の研究会と進めている [9]。



図・1 コード品質評価のGQMモデル例



図・2 品質測定評価ダッシュボードの例

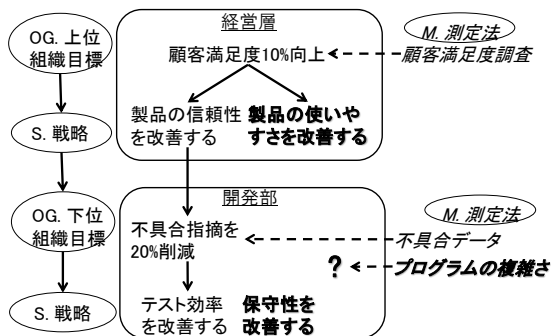
3. 組織構造と見える化: 不整合の防止

品質の測定や評価が、組織構造および組織上の目標 (Organizational Goal)、戦略 (Strategy) と整合していなければならない。それらの整合関係を可視化および改善する枠組みとして、GQM 法を拡張した GQM+Strategies [10] が有用である。

可視化の例 [11] を図・3 に示す。この例では、経営層において決定した戦略を下位部門が実現していないことや、逆に開発部の戦略が組織に貢献していないこと、および、測定しているプログラムの複雑さが組織目標とも結びついていないことを示している。不整合の特定により、真に組織上の目標と戦略へと寄与する測定を実施するように改善可能となる。

筆者はゴール指向経営研究会を主宰し、

GQM+Strategies の普及展開および応用研究を進めている。例えば、利害関係者間の関係から目標や戦略を支える様々な根拠を識別する手法を提案し、リクルート住まいカンパニーと共同で同社における目標・戦略へと適用し有効性を確認した [11]。またヤフーとは共同で、ユーザビリティの定量化と改善を一つの主要な目的として、ビジネス要求とユーザ要求にしばしば起きる対立を特定し解消へと導くゴール指向の手法 GO-MUC (Goal-oriented Measurement for Usability and Conflict) を提案し、同社への実サービスへの適用を通じて有効性を確認している [13]。



図・3 GQM+Strategiesによる可視化の例 (出典[11])

4. 信頼性予測と機械学習: 不確実な未来の扱い

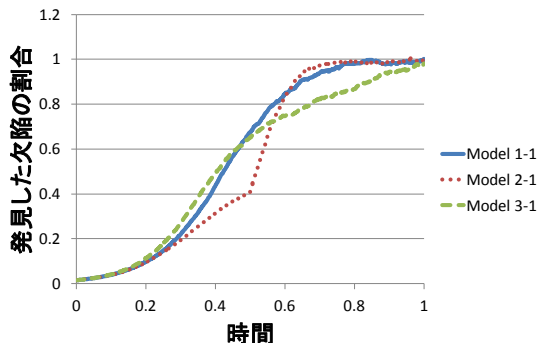
未来は不確実であり、ある時点における品質の予測や基準が、将来にわたって正しい保証はない。そこで、不確実性を加味した幅を持たせた予測や、ある時点の測定結果に基づく次の測定に向けた PDCA サイクルが欠かせない。

筆者らは品質予測の一種として、不確実性を加味して不具合数や欠陥数を幅をもって予測可能なソフトウェア信頼性モデルを実現した [14][15] (図・4)。同モデルは、従来からあるロジスティックモデルについて、不確実性を確率過程として追加拡張したものである。

筆者らは富士通研究所と共同で、同モデルによる信頼性の観点によるプロジェクト進行中のリスク検知支援を進めている [16]。また住友電工とは共同で、多数の過去プロジェクトへの適用を通じたベースラインの導出と比較評価への応用を進めている [17]。

さらに筆者らはコマツと共同で、機械学習を組み入れた品質測定評価システムの PDCA 改善プロセスの研究を進めている [18]。具体的には、GQM 法により初期セットとして測定法を設定の上、ソフトウェアモジュール群の一部に対して人手により品質評価を与え、

機械学習により有用な測定法の取捨選択および評価のための評定水準（基準値）を求める。



図・4 不確実性を考慮した信頼性モデル適用の例
(不確実性パターンの異なる3種の予測)

5. 第三者評価と業界実態およびパターン

品質要求と測定評価方法はソフトウェアそれぞれに定められるとしても、外部や業界を参考とせずに関係組織内で純粋に閉じて進めることは、自らが自らの品質を一方向的に主張する「オレオレ品質」とでも呼ぶべき状況である。そこで、開発主体から一定の独立性を持つ部門や組織による独立的、第三者的な品質評価の実施、さらには、組織外の業界傾向の把握が望ましい。

筆者らは情報処理推進機構の委託を受けて、2015年度ソフトウェア工学分野の先導的研究支援事業「測定評価と分析を通じたソフトウェア製品品質の実態定量化および総合的品質評価枠組みの確立」を実施している [19]。これはソフトウェア品質、測定および評価の国際規格ISO/IEC 25000シリーズ (SQuaREシリーズ) [20] を用いて客観的に、20のソフトウェアパッケージ製品やSaaS、ソフトウェアサービスの内部、外部、利用時品質を測定評価し、個別に評価結果と改善の助言を返却するとともに、品質間の関係を含めて広く業界傾向を明らかとし、組織を超えた世界初の網羅的品質ベンチマークとして公開するものである。

この品質実態調査を2016年度いっぱいまで進める予定である。さらに、適合の場合は、ISO/IEC 25051:2014に基づく第三者品質認証制度であるコンピュータソフトウェア協会実施のPSQ認証 [21] が得られる。客観的かつ国際的に通用する品質評価が得られるばかりか、以降の各組織におけるグローバルスタンダードに基づく品質評価・改善の基礎が得られるものであり、開発したソフトウェアやサービス製品をお持ちの場合はぜひ調査への参画を検討されたい。

また上記の品質実態調査に限らず、しばしば、特定の品質に焦点を当てた品質の組み入れや保証の方法が業界において一定の抽象度においてパターン化されている。それらのパターンの参照および活用もまた、組織外の傾向の把握および効率的・効果的な品質の測定評価に役立つ。

例えば筆者らは国内外の様々な機関と連携して、セキュリティをソフトウェアシステムの要求定義や設計、実装において組み入れるセキュリティパターン [22] の適用検証 [23][24]や、適用によって得られるセキュリティレベルの測定 [25]を進めている。

6. おわりに

本稿では、筆者らの実績と経験に基づき、ソフトウェア品質測定評価を実践する上で陥りやすい「落とし穴」、および、その回避あるいは軽減策としての「コツ」を解説した。本稿の内容および筆者らの取り組みが、読者、さらには業界ならびに研究コミュニティにおける真に価値あるソフトウェア品質測定評価の実践と研究発展の基盤となれば幸いである。

研究コミュニティに関連して、2017年3月13-18日に、筆者を共同プログラム委員長として、ソフトウェアテスト、検証、妥当性確認に関する代表的国際会議10th IEEE International Conference on Software Testing, Verification and Validation (ICST 2017) [26]を早稲田大学にて開催する。日本から多くの論文投稿ならびに参加をいただければ幸いである。

謝辞

研究の一部は、独立行政法人情報処理推進機構 技術本部 ソフトウェア高信頼化センターSEC: Software Reliability Enhancement Center) が実施した「2015年度ソフトウェア工学分野の先導的研究支援事業」の支援を受けたものです。研究の一部は、JSPS科研費JP25330091, JP16H02804およびIISF SSRフォーラム2015, 2016の助成を受けたものです。また共同研究、受託研究機会をいただいた各社に御礼申し上げます。

参考文献

- [1] L. M. Laird and M. C. Brennan 著, 野中誠, 鷲崎弘宜 訳 (2009): “演習で学ぶソフトウェアメトリクスの基礎”, 日経 BP 社
- [2] 鷲崎弘宜ほか (2007): “プログラムソースコード

- のための実用的な品質評価枠組み”, 情報処理学会論文誌, 48(8)
- [3] 鷲崎弘宜ほか (2010): “ソースコード解析による品質評価の仕組み”, 日経エレクトロニクス
- [4] H. Washizaki, et al. (2007): “A Framework for Measuring and Evaluating Program Source Code Quality,” Proc. PROFES, LNCS 4589
- [5] S. Sato, H. Washizaki, et al. (2013): “Effects of Organizational Changes on Product Metrics and Defects,” Proc. APSEC
- [6] R. Ando, S. Sato, C. Uchida, H. Washizaki, et al. (2015): “How Does Defect Removal Activity of Developer Vary with Development Experience?,” Proc. SEKE
- [7] H. Nakai, K. Honda, H. Washizaki, et al. (2014): “Initial Industrial Experience of GQM-based Product-Focused Project Monitoring with Trend Patterns,” Proc. APSEC
- [8] I. Jacobson ほか著, 鷲崎弘宜ほか訳 (2013): “ソフトウェアエンジニアリングのエッセンス: SEMAT カーネル”, 日経BP ITPro
- [9] 鷲崎弘宜, 小林浩ほか (2016): “マルチディメンショナル・レトロスペクティブ - SEMAT の7つの視点で気づきの多い多面的な振り返り -”, XP 祭り 2016, 2016年9月24日, <http://xpjug.com/>
- [10] V. Basili, et al. 著, 鷲崎弘宜ほか 監訳, “ゴール&ストラテジ入門: 残念なシステムの無くし方 (GQM+Strategies)”, オーム社, 2015
- [11] J. Heidrich and A. Trendowicz (2013): “測定を基にした、ソフトウェア戦略とビジネス目標の整合”, Fraunhofer IESE プランニングエキスパート用チュートリアル (IPA/SEC WG 訳)
- [12] T. Kobori, H. Washizaki, et al. (2016): “Exhaustive and efficient identification of rationales using GQM+Strategies with stakeholder relationship analysis,” IEICE Trans. Information and Systems, Vol. E99-D
- [13] C. Uchida, K. Honda, H. Washizaki, et al. (2016): “GO-MUC: A Strategy Design Method Considering Requirements of User and Business by Goal-Oriented Measurement,” Proc. CHASE
- [14] K. Honda, H. Nakai, H. Washizaki, et al. (2014): “Predicting Time Range of Development Based on Generalized Software Reliability Model,” Proc. APSEC
- [15] H. Washizaki, K. Honda, Y. Fukazawa (2015): “Predicting Release Time for Open Source Software based on the Generalized Software Reliability Model,” Proc. Agile
- [16] K. Honda, H. Washizaki, et al. (2015): “Detection of Unexpected Situations by Applying Software Reliability Growth Models to Test Phases,” Proc. ISSRE
- [17] K. Honda, N. Nakamura, H. Washizaki, et al. (2016): “Case Study: Project Management Using Cross Project Software Reliability Growth Model,” IEEE Workshop Trustworthy Computing
- [18] N. Tsuda, M. Takada, H. Washizaki, et al. (2015): “Iterative Process to Improve GQM Models with Metrics Thresholds to Detect High-risk Files,” SANER
- [19] H. Nakai, N. Tsuda, K. Honda, H. Washizaki, et al. (2016): “Initial Framework for a Software Quality Evaluation based on ISO/IEC 25022 and ISO/IEC 25023,” Proc. QRS
- [20] ISO/IEC 25000:2005 Software engineering-Software product Quality Requirements and Evaluation (SQuaRE)-Guide to SQuaRE
- [21] コンピュータソフトウェア協会, “PSQ 認証制度”, <http://www.psq-japan.com/>
- [22] A. Nhlabatsi, et al. (2010): “Security Patterns: Comparing Modeling Approaches”, in “Software Engineering for Secure Systems”, pp75-111, IGI Global
- [23] T. Kobashi, N. Yoshioka, H. Kaiya, H. Washizaki, et al. (2014): “Validating Security Design Pattern Applications by Testing Design Models,” IJSSE, 5(4)
- [24] M. Yoshizawa, H. Washizaki, et al. (2016): “Implementation Support of Security Design Patterns Using Test Templates,” Information, 7(2)
- [25] E. B. Fernandez, N. Yoshioka, H. Washizaki, et al. (2010): “Measuring the level of security introduced by security patterns,” Proc. SPattern
- [26] 10th IEEE International Conference on Software Testing, Verification and Validation, <http://aster.or.jp/conference/icst2017/>