欠陥モデリングの重要性と欠陥情報の移転と共有

- ソフトウェア欠陥を売買する IT 社会の実現に向けて -

†細川 宣啓

ソフトウェア欠陥に関わる研究は，定性的側面や属性に関する研究が少ない．欠陥の類似性や分類・粒度に関する研究がない．一方，様々な欠陥の定性的分析手法も属人性や恣意性等様々な課題が残存する．本発表では，人は環境因子によって「誤らされる」のであり，欠陥を「誘発する」環境因子と人の「過失」との関係を示した欠陥標本（モデル）が必要であることを述べる．はじめにソフトウェア欠陥標本作成の方法（欠陥モデリングという）と，標本を作成・共有・移転・流通させる技術について紹介する．その後，欠陥標本の利点を示し，最後に標本の商品化について議論する．

The study of Defect Modeling and their transferring and sharing

†Nobuhiro Hosokawa

There are few researches about qualitative aspects of software defects. On the other hand, there are several issues that the dependency on personal experience and arbitrariness in qualitative analysis techniques will still remain.

In this presentation, the Human negligence tends to be caused by environmental induction factors. Moreover, it is important that the defect specimen should be expressed with the relationship between human "negligence" factors and “induction” factors. Introduction how to create the software defect specimens (referred to as defect modeling), and we introduce the technology to be creating a sample, share, and transfer and distribution. At last, we discuss about the commercialization of software defect specimens.

1. はじめに

1.1. 背景：ソフトウェア欠陥の既存研究

一般的なソフトウェア欠陥に関わる既存研究は「欠陥の定量的分析手法([1]Bagchi 2009), (Henderson 2008)」「欠陥の混入予測に関するモデル・推定手法(Jeet, Bhatia et al. 2011)」に関する研究が主流である．品質関連する研究の中でもソフトウェア欠陥に関する研究の多くは，「時系列の変化」「欠陥の粒度・重み」を考慮したものが少ない．ソフトウェア欠陥の既存研究では，欠陥の定性的側面＝属性に関する研究が少ないのが実情である．

そもそも個々のソフトウェア欠陥に関する国際的な定義や規格が存在しないため，欠陥の個数・数え方にも課題が残存する．例として，「Defect」や「Fault」という用語定義は存在するが，「メモリーリーク」「キャストエラー」「例外未実装」等は厳密な定義が存在しない．欠陥の類似性や分類・比較し難い理由の一つである．

また Industrial Paper・経験事例紹介に於いても，個々のソフトウェア欠陥に関する経済面からの調査研究が不足している．理由はソフトウェア欠陥という自社のセンシティブデータ（＝多くは機密条項）を市場に情報公開し難いためであろうと推測される．

1.2. ソフトウェア欠陥分析の恣意性問題

一方，欠陥や障害の定性的分析の観点では「なぜなぜ分析（5Why）」「根本原因分析（RCA: Root Cause Analysis）(Dalal and Chhillar 2013)」といった手法が主流であるが，手法そのものが分析者の属人的経験に左右されるの「分析者の恣意性問題」と，欠陥そのものを特定個人の能力不足・愚行に帰結させようとする「原因の恣意性問題」が存在する．前者の属人的恣意性はいわゆるパターン・フレーミング・バイアス（既に経験済の事例に現象面から当てはめようとする認知偏向）の問題と，欠陥を過度に単純化して考える問題に分解される．また後者は実プロジェクトに散見される「対策アクションの取りやすい原因に帰結させせよう」とする思考バイア

†細川 宣啓, Quality Engineer, Researcher
Service Software Engineering, Industrial & Solution,
IBM Research-Tokyo.IBM Japan Ltd.
email: CARVIN@jp.ibm.com

© 2016 Information Processing Society of Japan

スの問題である.

2. 提案:

2.1. 欠陥を標本化し共有・移転・流通させる

一般的に摩耗・腐食といった物理法則に支配されるハードウェア故障や欠陥と異なり，ソフトウェア欠陥は100%が「人間の思考の産物」である．従って，認知心理学や神経科学の手法を以て「人の誤り方」を分析するべきである．特に，開発者が意図的に欠陥・バグを混入させないという前提に立脚すれば，人は環境因子によって「誤らされる」のであり，欠陥を「誘発する」環境因子と，人の犯す「過失」との関係を示した標本に基づいて定性分析すべきである．(図1)

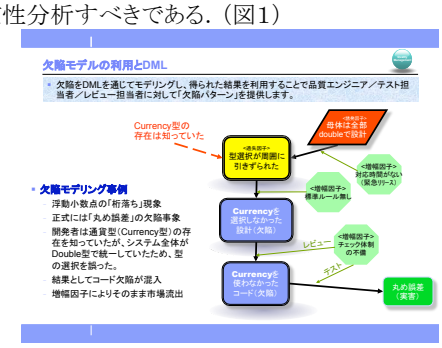

図1:欠陥標本の事例

本発表では，標本事例を交えながら数千件の実例データに基づいた自動標本抽出の事例を紹介し，前述の1.1および1.2から，欠陥標本の作成手法を提案する．欠陥標本作成については，生物学での生物標本作成，および医学分野に於ける病理学(＝病気そのものを研究する学問)研究と同様，ソフトウェア欠陥を「無毒化・弱毒化」する作業手順を述べる．この作業を欠陥モデリングといい，検出・抽象化・固定化・分類等のプロセスについても言及する．

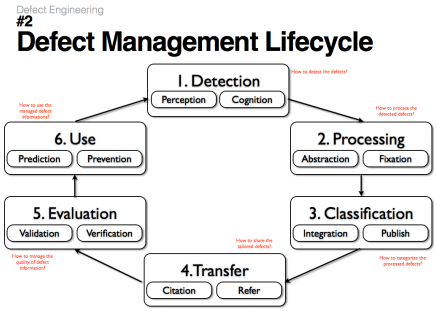

図2:欠陥モデリングプロセス

欠陥モデリングを経て得られた標本は，同一組織内や同一業界内での共有，あるいは現在実施中のプロジェクト間でトレンド情報として流通させる等，各種の用途に用いる事が可能となる．また前述1)で述べた「センシティブデータ」の課題も，抽象化の際にセキュリティー情報とプライバシー情報を削除する事で共有・移転・流通しやすくなる．

さらに前述の課題に対して，欠陥の重み・粒度・複雑度の定量化が行ないやすくなり，欠陥の難易度(＝検出難易度・除去難易度・再発予防難易度)のスコアリングや経済性の算出が従来よりも簡便に行なう事が可能となる．

3. 議論:

近い将来，ビジネス分野として，ソフトウェア欠陥をプロジェクト間または同業種異企業間，異業種間で移転させる仕組みを作り，最終的には売買する商品として扱いたい．また技術分野として，欠陥はソフトウェア・エンジニアリング研究の中核となる一分野として確立する．このためには，今後何が必要であろうか？

この仕組みの前提として，欠陥共有のシステム基盤としてのデータベースや以下項目も議論が必要である．

- 欠陥を交換・売買するための共通プロトコル定義
- 欠陥分類方法の確立
- 欠陥の経済性検討(価格決定原理と ROI 計算)

既存のソフトウェア・エンジニアリング研究の多くが「手法＝How」の研究であったが，特に品質エンジニアリングの分野では「対象ソフトウェア欠陥＝What」の研究が黎明期を迎えている．このようなソフトウェア欠陥研究の分野確立は社会的意義があると考える．

参考文献

[1] Bagchi, T. P. (2009). "Models for software defects and testing strategies." ACM SIGSOFT Software Engineering Notes **34**(2): 1.

[2] Dalal, S. and R. S. Chhillar (2013). "Empirical study of root cause analysis of software failure." ACM SIGSOFT Software Engineering Notes **38**(4): 1.

[3] Henderson, C. (2008). "Managing software defects." ACM SIGSOFT Software Engineering Notes **33**(4): 1.

[4] Jeet, K., et al. (2011). "A bayesian network based approach for software defects prediction." ACM SIGSOFT Software Engineering Notes **36**(4): 1.

© 2016 Information Processing Society of Japan