

令和 7 年 6 月 16 日現在

機関番号：34406

研究種目：若手研究

研究期間：2019～2024

課題番号：19K20242

研究課題名（和文）大規模欠陥データを利用したソフトウェア開発におけるプロジェクト比較の枠組みの研究

研究課題名（英文）A Study on a Framework for Project Comparison in Software Development Using Large-Scale Defect Data

研究代表者

本田 澄（Honda, Kiyoshi）

大阪工業大学・情報科学部・講師

研究者番号：40732938

交付決定額（研究期間全体）：（直接経費） 2,800,000円

研究成果の概要（和文）：本研究では、ソフトウェア開発におけるドメインや開発手法の違いが欠陥の発生に与える影響を明らかにするため、様々なプロジェクトにおける欠陥情報を収集するウェブアプリケーションを開発し、大規模な欠陥データを収集した。収集したデータをもとにソフトウェア信頼度成長モデルを検証・拡張し、異なる組織やドメイン間で比較可能な評価枠組みを構築した。さらに、知見の共有化を図り、プロジェクトを比較する手法を提案し、その有効性を示し、開発したウェブアプリケーションに機能を追加し、利用者がこの手法を利用できるようにした。

研究成果の学術的意義や社会的意義

本研究は、異なる企業やドメインに属するソフトウェア開発プロジェクト間での比較を可能とする評価の枠組みを構築し、ソフトウェア信頼度成長モデルの汎用的な評価に貢献した点で学術的意義がある。特に、従来は比較が困難であった異種プロジェクト間の欠陥発生傾向の比較を可能とした点に新規性がある。また、これまで公開事例の少なかった欠陥データセットを整備・公開することで、実データに基づく研究や実践的応用の促進に寄与した。さらに、定量的評価に基づく開発プロセス改善を支援し、ソフトウェアの品質向上と開発効率の向上に資する点で社会的意義も有している。

研究成果の概要（英文）：This study aims to clarify the impact of differences in domains and development practices on defect occurrence in software projects. To this end, a web application was developed to collect defect data from various projects, enabling large-scale data collection. Based on the collected data, we validated and extended software reliability growth models, and constructed a comparative evaluation framework applicable across different organizations and domains. Furthermore, we proposed a method for project comparison, demonstrated its effectiveness, and enhanced the web application to allow users to apply this method.

研究分野：ソフトウェア工学

キーワード：ソフトウェア信頼度成長モデル プロジェクト比較 ソフトウェア品質

1. 研究開始当初の背景

ソフトウェア開発においては、ソフトウェアに不具合を与える様々な欠陥が作りこまれ、テストによって検出され管理される。開発中に欠陥がどれだけ存在するかを予測する方法であるソフトウェア信頼度成長モデル(SRGM)[Goel1985]は1980年代あたりから研究が盛んに行われてきた。数多くのSRGMが提案されているが、それぞれの方法は特定の状況下でのみ有効であり、条件が異なる場合に利用できないことが重大な問題である。

本研究では、様々なドメインや開発スタイルに属するソフトウェア開発に対して有効なSRGMを構築し活用方法を広く普及することでソフトウェア開発をより効果的で制御可能とすること目的とする。そのためには多くの企業の開発データの収集方法および普及方法としてウェブアプリケーションの開発が必要である。また企業の開発データのみならずオープンソースソフトウェア(OSS)における開発データも対象とする。

2. 研究の目的

本研究の目的は、多様な開発スタイルやドメインに対応した汎用的なSRGMを構築し、それを活用可能な評価枠組みを提供することである。そのために、企業やOSSにおける開発データを収集して蓄積し分析できるウェブアプリケーションを開発し、これを通じてデータ活用の普及を図る。また、得られたデータを用いたプロジェクト間の比較や、定量的なスケジュール決定の支援も目的とする。本研究を行うことで現在困難とされている開発スケジュールの定量的な決定と、異なるドメインや開発スタイルにおける欠陥データを比較することに役立つと考えられる。

3. 研究の方法

本研究では、SRGMの汎用化と評価手法の確立を目的として、以下の手順で研究を進めた。図1に研究の方法の概要を示す。①大規模なデータ収集を実現するために、企業や開発者が自らの開発プロジェクト情報を登録・提供できるウェブアプリケーションを構築する。このアプリケーションにより、開発ドメインや開発手法の異なる多数の企業・プロジェクトから、開発プロジェクトに関する定量的なデータ(例:リリース時期、欠陥数、開発期間、修正履歴など)を収集する。

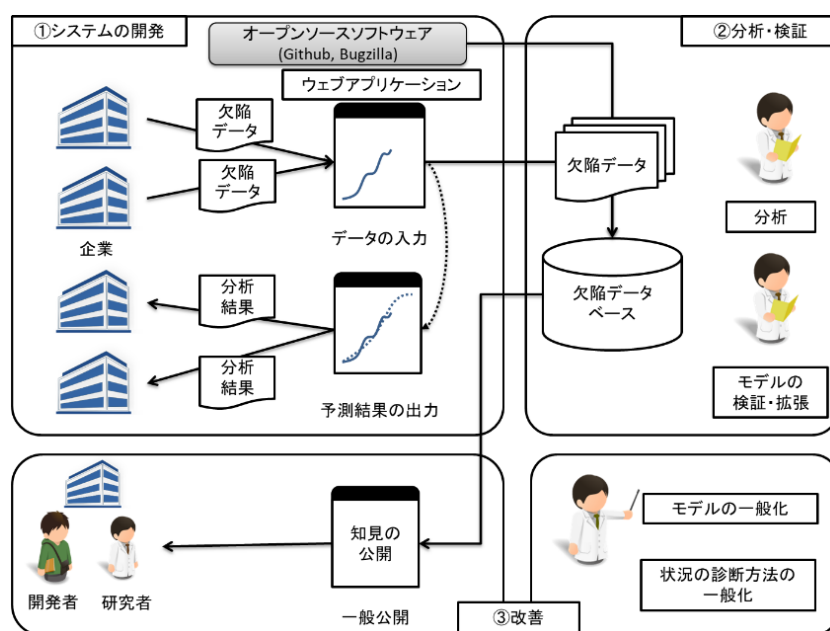


図1 研究の方法の概要

②収集したデータを用いて、既存のSRGMに適用し、モデルの適合性や予測精度を検証する。また、特定のドメインや開発手法などの特徴量に基づくモデル選択手法の開発・評価も行う。モデルの拡張にあたっては、異なる企業・ドメインにおける実データに対する適合度を比較し、共通する信頼度成長モデルの出力パターン抽出を試みる。

③開発状況を定量的に診断可能な指標や可視化手法を提案し、実務においてモデルを利用しやすくすることを目指す。

4. 研究成果

(1) 欠陥データの収集と分析基盤の整備

本研究では、欠陥データの収集、管理、および分析を支援するウェブベースのシステムを開発・運用している。このシステムは、ユーザごとにアクセス制御を設けることで、企業との連携においてセキュリティを確保しつつ、欠陥データの登録・取得・可視化を可能とするプラットフォームである。

現在、このプラットフォームには、8つの企業が公開している77件のOSSプロジェクトに関

する欠陥データが蓄積されており、SRGMの適用とその妥当性の検証に活用されている。[Honda2024-1]では、本プラットフォームを利用して収集された複数のプロジェクトの欠陥データに対し、各種SRGMを適用した結果とその分析を報告している。図2には、ある1つのプロジェクトに対してLogisticモデル、Gompertzモデル、Exponentialモデルなど複数のSRGMを、異なるアルゴリズムで適用した結果を示している。図の横軸は時間、縦軸は累積欠陥数を表しており、これらの結果から最も精度の高いモデルとアルゴリズムの組み合わせを選定し、そのモデルによって導出された最大予測欠陥数を提示することが可能である。

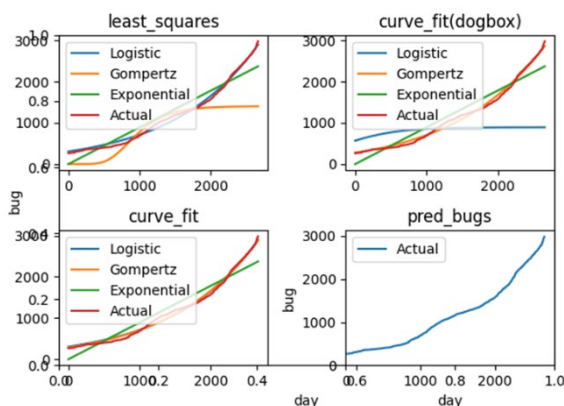


図2 ウェブアプリケーションの出力の例

さらに、[Honda2023]では、SRGMを逐次的に適用することで得られる最大予測欠陥数の推移に基づき、プロジェクトの特性を評価し、開発組織との関連性を分析している。プロジェクトの傾向は、(a) 最大予測欠陥数が安定している「安定タイプ」、(b) 徐々に増加する「漸増タイプ」、(c) 急激に増加する「爆発タイプ」の3つに分類される。77件のOSSプロジェクトを対象とした分析の結果、8つの組織のうち3つは安定タイプが多く、別の3つは漸増タイプが多く、残りの2つは複数のタイプが混在している傾向を明らかにした。

加えて、[Honda2019]では、ソフトウェア開発における各種メトリクスに対して時系列解析を行い、特に値の変化が顕著な時期を検出する手法を提案している。分析対象としたメトリクスには欠陥数も含まれており、欠陥数が大きく変動する時期の特定が可能となった。この手法は、欠陥データを分割して、SRGMを適用する際の有効な補助手段となる。実際の適用例として、同一組織のプロジェクトを対象に、2016年から2018年にかけて収集されたメトリクスに基づいて分析を行った結果、欠陥数、工数、およびテスト工数に基づく欠陥発見数において、同時期に顕著な変化が生じていることが確認された。さらに、欠陥数と他のメトリクスとの相関を調査することで、SRGMの予測精度を向上させる指標の特定につながる可能性が示唆された。

(2) SRGMの提案と比較評価

本研究では、年度ごとに収集された複数のソフトウェアプロジェクト間で、欠陥数の時系列的变化を比較可能とするため、SRGMの改良と比較評価のための分析枠組みの構築を進めてきた。特に、欠陥数とリリース間隔などの時間軸上の関係性、すなわち成長曲線の「形状」に着目し、開発期間や規模が異なるプロジェクト間においても比較が可能な手法を提案した。この手法により、従来は難しかったプロジェクト横断的な比較評価が可能となり、その有効性を国内会議にて報告している。

[Honda2024-1]では、複数プロジェクト間でのSRGMによる比較を実現するため、欠陥データを正規化して比較可能な形に整えた。これにより、プロジェクトごとの成長曲線の「形状」に基づいた分類が可能であることを示した。図3は、3つの異なる組織（組織A、組織B、組織C）に属する開発プロジェクトの欠陥情報を正規化し、時系列で比較した結果を示している。このような正規化により、プロジェクトごとの特徴を可視化し比較することが可能となった。

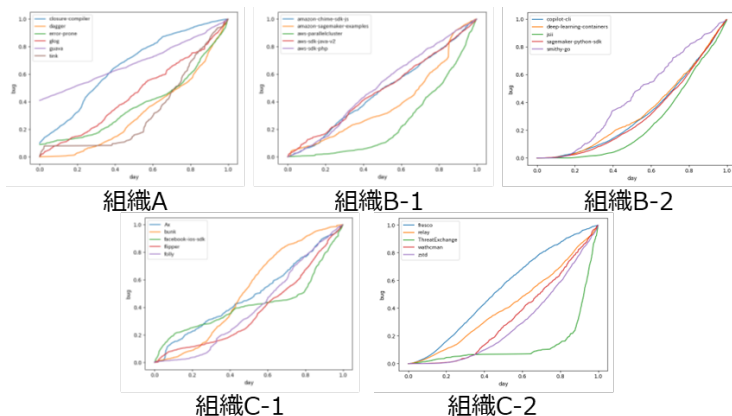


図3 複数プロジェクトの比較

さらに、組織をまたいで得られたプロジェクトデータに対し、成長曲線の形状に基づいて分類を行った結果、図4に示すように以下の5つのパターン（S字型、凸型、凹型、逆S字型、直線型）に整理可能であることを確認した。この分類により、開発のスタイルや欠陥の発見傾向に関する特徴を捉えることができる。

加えて、[Honda2024-2]では、欠陥予測の精度向上を目的として、ベイジアンネットワーク(BN)および動的ベイジアンネットワーク(DBN)を用いた分析手法を導入した。特にDBNを活用することで、あるリリース時点の開発要因（例：工数、バグ数など）が次期リリースの欠陥数にどのような影響を及ぼすかという、時系列での依存関係をモデル化することが可能となった。これにより、静的なモデルでは把握が困難であった中長期的な品質への影響を評価できるようになった。

図5には、同一組織が開発した複数のサービスについて、それぞれの開発要因間の関係をDBNで表現したグラフを示している。この結果から、同じ組織であってもプロジェクトごとに開発要因間の影響構造が異なることが明らかとなり、DBNを用いた手法がプロジェクトごとの開発特性を把握する上で有効であることが示された。

(3) コードメトリクスおよびコードスメルの関係分析

本研究では、欠陥情報に加えて、ソースコードに関する静的解析指標（メトリクス）やコードスメルに関するデータも収集し、それらと欠陥発生傾向との関連性について分析を行ってきた。これにより、SRGMに組み込むことが可能なコード品質指標の選定に資する知見を得ることができた。

[Honda2024-3]では、特にコードクローン中に含まれるコードスメルの実態に着目した調査・分析を行った。ここで、コードクローンとはソースコード中に存在する一致あるいは類似するコード片を指し、コードスメルは将来的にソフトウェアの保守性や信頼性に悪影響を及ぼす可能性のある構造的問題を含むコード片を指す。本研究では、コードスメルを含むコード片がコードクローンによって複製されることが、ソフトウェア品質に与えるリスクであると仮定し、これを定量的に検証するための分析を行った。

具体的には、開発履歴を含めたコードクローンの分析が可能なツールであるCloneTrackerを用いて、対象とするOSSプロジェクトのコードクローンを検出した。また、静的解析ツールSonarQubeを用いて、対象コードに含まれるコードスメルを抽出した。さらに、CloneTrackerによって検出されたコードクローンと、SonarQubeにより検出されたコードスメルを統合し、コードクローン中に含まれるコードスメルを抽出・分類・分析した。その結果、クローン中のコードスメルに一定のパターンや特徴が存在することを明らかにし、特定の種類のコードスメルがクローンによって複製されやすい傾向があることが確認された。

この研究成果は、コードの保守性向上および欠陥予測精度向上の観点から、SRGMへの品質指

標統合の可能性を広げるものであるとともに、ソフトウェア開発実務におけるコード品質管理の重要性を示すものとなった。本研究の成果は、日本ソフトウェア科学会第41回大会において優秀発表賞を受賞しており、その学術的貢献が評価されている。

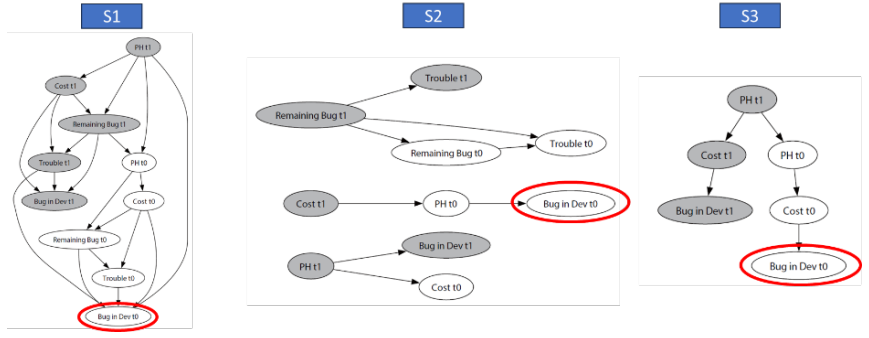
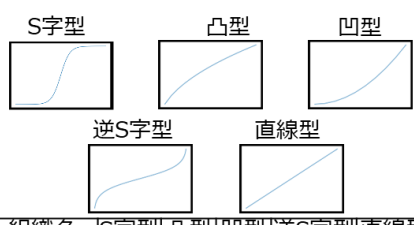


図5 DBNを用いた分析

参考文献

- [Goell1985] Goel, Amrit L. Software reliability models: Assumptions, limitations, and applicability. IEEE Transactions on software engineering (12), 1985, 1411-1423.
- [Honda2019] Honda, K., Washizaki, H., Fukazawa, Y., Taga, M., Matsuzaki, A. (2019, December). Industrial case study on time series analysis of metrics changes based on GQM models. In 2019 10th International Workshop on Empirical Software Engineering in Practice (IWESPE), 2019, 13-18. IEEE.
- [Honda2023] 本田澄, 神山飛悠河, 鷺崎弘宜, 深澤良彰, 予測欠陥数の変化を用いた組織とプロジェクトの関係分析. 研究報告ソフトウェア工学 (SE), 2023 (23), 1-8.
- [Honda2024-1] 本田澄, 松田望, 鷺崎弘宜, 深澤良彰, 複数プロジェクトを利用したソフトウェア信頼度成長モデル構築のためのウェブアプリケーション. 2024年情報科学技術フォーラム(FIT), 2024, 145-146.
- [Honda2024-2] Honda, K., Washizaki, H., Fukazawa, Y., Taga, M., Matsuzaki, A., Nakagawa, K., Sakai, Y. An Empirical Study on Predicting Software Development Bugs Using Dynamic Bayesian Networks. In 2024 IEEE 35th International Symposium on Software Reliability Engineering Workshops (ISSREW), 307-310. IEEE.
- [Honda2024-3] 本田澄, 奥野真伍, 吉田則裕, コードクローン中のコードスメルに関する調査研究. 日本ソフトウェア科学会第41回大会(2024年度)講演論文集, 2024, 1-2.

グラフの形状分類



組織名	S字型	凸型	凹型	逆S字型	直線型
組織A	0	1	3	0	2
組織B	0	0	7	0	3
組織C	2	0	6	1	1
組織D	0	2	3	0	6
組織E	4	0	5	0	1
組織F	0	0	10	0	0

図4 形状のパターン

5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 4件/うち国際共著 0件/うちオープンアクセス 0件）

1. 著者名 Honda Kiyoshi, Hashimoto Naohiro, Washizaki Hironori, Fukazawa Yoshiaki, Taga Masahiro, Matsuzaki Akira, Nakagawa Kazuyuki, Sakai Yusuke	4. 巻 ISSRE 2023
2. 論文標題 A Bayesian Network-Based Analysis of Changing Impact of Organizational Goals on Software Development	5. 発行年 2023年
3. 雑誌名 2023 IEEE 34th International Symposium on Software Reliability Engineering Workshops (ISSREW)	6. 最初と最後の頁 154-159
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/ISSREW60843.2023.00080	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 本田 澄, 神山 飛悠河, 鷺崎 弘宜, 深澤 良彰	4. 巻 SE
2. 論文標題 予測欠陥数の変化を用いた組織とプロジェクトの関係分析	5. 発行年 2023年
3. 雑誌名 情報処理学会 ソフトウェア工学研究会 (IPSJ/SIGSE) 第213回SE研究発表会	6. 最初と最後の頁 1-8
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 西村 滋幸, 本田 澄, 山下 育男	4. 巻 29
2. 論文標題 Grad-CAMを用いた画像認識AIの特徴分析の試み	5. 発行年 2022年
3. 雑誌名 日本ソフトウェア科学会 ソフトウェア工学の基礎研究会第29回ソフトウェア工学の基礎ワークショップ (FOSE2022) in 松江しんじ湖温泉	6. 最初と最後の頁 1-2
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 本田 澄, 西尾 達哉, 鷺崎 弘宜, 深澤 良彰	4. 巻 2021
2. 論文標題 オープンソースソフトウェアにおけるCode Smellと対応するリファクタリングの特徴に関する調査	5. 発行年 2021年
3. 雑誌名 ソフトウェアエンジニアリングシンポジウム2021論文集	6. 最初と最後の頁 226--234
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

[学会発表] 計15件(うち招待講演 0件/うち国際学会 3件)

1. 発表者名 本田 澄, 吉永 圭佑, 鷲崎 弘宜, 深澤 良彰
2. 発表標題 オープンソースソフトウェアにおけるバージョンアップと品質の関係分析の試み
3. 学会等名 ソフトウェアエンジニアリングシンポジウム2023
4. 発表年 2023年

1. 発表者名 本田 澄, 里中 俊介, 鷲崎 弘宜, 深澤 良彰
2. 発表標題 複数プロジェクトを利用したソフトウェア信頼度成長モデル構築に向けた試み
3. 学会等名 第30回ソフトウェア工学の基礎ワークショップ
4. 発表年 2023年

1. 発表者名 Honda Kiyoshi, Hashimoto Naohiro, Washizaki Hironori, Fukazawa Yoshiaki, Taga Masahiro, Matsuzaki Akira, Nakagawa Kazuyuki, Sakai Yusuke
2. 発表標題 A Bayesian Network-Based Analysis of Changing Impact of Organizational Goals on Software Development
3. 学会等名 2023 IEEE 34th International Symposium on Software Reliability Engineering Workshops (ISSREW) (国際学会)
4. 発表年 2023年

1. 発表者名 本田澄
2. 発表標題 ソフトウェア信頼度成長モデルのウェブアプリケーション
3. 学会等名 ウィンターワークショップ2024・イン・鹿児島
4. 発表年 2024年

1. 発表者名 本田 澄, 神山 飛悠河, 鷺崎 弘宜, 深澤 良彰
2. 発表標題 予測欠陥数の変化を用いた組織とプロジェクトの関係分析
3. 学会等名 第213回SE研究発表会
4. 発表年 2023年

1. 発表者名 西村 滋幸, 本田 澄, 山下 育男
2. 発表標題 Grad-CAMを用いた画像認識AIの特徴分析の試み
3. 学会等名 日本ソフトウェア科学会 ソフトウェア工学の基礎研究会第29回ソフトウェア工学の基礎ワークショップ (FOSE2022) in 松江しんじ湖温泉
4. 発表年 2022年

1. 発表者名 本田 澄, 橋本 直宙, 世羅 俐恵, 鷺崎 弘宜, 深澤 良彰, 多賀 正博, 中川 和之, 酒井 優介
2. 発表標題 ベイジアンネットワークを利用した組織目標が与える影響の可視化の試み
3. 学会等名 情報処理学会/ソフトウェア工学研究会ウィンターワークショップ2023・イン・富山
4. 発表年 2023年

1. 発表者名 松田 宗一郎, 本田 澄, 鷺崎 弘宜, 深澤 良彰
2. 発表標題 Web APIを模倣するデバッグツールの開発
3. 学会等名 日本ソフトウェア科学会 ソフトウェア工学の基礎研究会第29回ソフトウェア工学の基礎ワークショップ (FOSE2022) in 松江しんじ湖温泉
4. 発表年 2022年

1. 発表者名 本田 澄、西尾 達哉、鷲崎 弘宜、深澤 良彰
2. 発表標題 オープンソースソフトウェアにおけるCode Smellと対応するリファクタリングの特徴に関する調査
3. 学会等名 ソフトウェアエンジニアリングシンポジウム2021
4. 発表年 2021年

1. 発表者名 本田 澄、鷲崎 弘宜、深澤 良彰、多賀 正博、松崎 明
2. 発表標題 DIDを利用したアジャイル化による効果の評価に向けた試み
3. 学会等名 第28回ソフトウェア工学の基礎ワークショップ (FOSE2021)
4. 発表年 2021年

1. 発表者名 西尾 達哉, 本田 澄
2. 発表標題 バージョン間の差分に着目したリファクタリング実施状況の調査
3. 学会等名 FOSE2020: 第27回 ソフトウェア工学の基礎ワークショップ
4. 発表年 2020年

1. 発表者名 Honda Kiyoshi、Washizaki Hironori、Fukazawa Yoshiaki、Taga Masahiro、Matsuzaki Akira
2. 発表標題 Industrial Case Study on Time Series Analysis of Metrics Changes Based on GQM Models
3. 学会等名 10th International Workshop on Empirical Software Engineering in Practice (IWESEP) (国際学会)
4. 発表年 2019年

1. 発表者名 San Kyawt Kyawt、Washizaki Hironori、Fukazawa Yoshiaki、Honda Kiyoshi、Taga Masahiro、Matsuzaki Akira
2. 発表標題 DC-SRGM: Deep Cross-Project Software Reliability Growth Model
3. 学会等名 2019 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW) (国際学会)
4. 発表年 2019年

1. 発表者名 本田 澄、鷲崎 弘宜、深澤 良彰、多賀 正博、松崎 明
2. 発表標題 GQMを利用した組織目標が与えるメトリクスへの影響についての時系列解析
3. 学会等名 IPSJ/SIGSE ソフトウェアエンジニアリングシンポジウム2019
4. 発表年 2019年

1. 発表者名 本田 澄、鷲崎 弘宜、深澤 良彰
2. 発表標題 欠陥データを利用したソフトウェアプロジェクト比較手法に関する研究に向けて
3. 学会等名 日本ソフトウェア科学会 第26回 ソフトウェア工学の基礎ワークショップ FOSE2019
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

ソフトウェア信頼度成長モデル利用ページ https://softwarereliabilitylab.github.io/srgm/
--

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------