

平成 30 年 6 月 5 日現在

機関番号：32675

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K01208

研究課題名(和文)品質指向ソフトウェア開発のためのデータマイニング型プロセス支援技術の開発と改善

研究課題名(英文) Development of assessment methods for aiding quality-oriented software development based on data-mining approach and their improvement

研究代表者

木村 光宏 (KIMURA, Mitsuhiro)

法政大学・理工学部・教授

研究者番号：20263486

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)：実際のソフトウェア開発プロジェクトにおいて採取された、プロジェクトごとの属性データを分析する方法を開発し、その改善に取り組んだ。具体的には、属性データからの情報を用いて、当該プロジェクトの成果物であるソフトウェアプロダクトの出荷後一定期間における不具合数の予測手法について研究した。特に、ソフトウェア開発の初期段階において観測される属性データを積極的に活用することにより、ソフトウェア開発プロジェクトが早期の段階で、出荷後のソフトウェアプロダクトの品質を予測することを目指した。また欠測や汚損のある属性データの扱いについて研究するとともに、変数間の従属性をうまく扱う方法についても研究した。

研究成果の概要(英文)：We have studied on a new methodology which can precisely forecast the software quality by using software development characteristics data sets which were provided from the actual software development companies. The data sets consist of a mixture of both quality and quantity data. In particular, our approach is unique in the sense that our method utilizes the data sets which can be obtained in the early stage of software development process. This means that our new method can predict the software quality in the earlier stage, and can provide some chances to improve the software development process on-the-fly. As a result, a framework of neural networks generally yields good forecast performance, whereas we have newly faced the issues of data contamination including missing data.

研究分野：信頼性工学

キーワード：ソフトウェア品質 定性的データ 多変量解析

1. 研究開始当初の背景

現代社会を支えるウェブシステム、クラウドコンピューティング、組込系などの動作を司る、中規模・大規模なサイズをもつソフトウェアの品質と信頼性（簡潔に言えばソフトウェアがユーザの期待通りに動作する度合い）が、開発工程の改善や開発作業者の能力の向上をもってしてもなかなか改善されない、あるいはそれらの定量的評価が困難であることは、今から45年ほど前、すなわち「ソフトウェア危機 (software crisis)」が叫ばれ始めた頃から知られている。

コンピュータの構成要素を大別すれば、ハードウェアとソフトウェアに分けることができるが、ハードウェアについては、その製造における技術の進歩と成熟により、ハードウェアに要求された品質を作り込むことは、許容範囲のコストの下でほぼ達成されていると言ってよい。これは各開発・製造者がそれぞれ培ってきた製造技術の成熟によるところが大きいが、本質的には、開発・製造したものが物体として目に見えること、つまり成果物の品質を確認しやすいという点が、その品質評価と改善に寄与するためである。ところがこれに対して、ソフトウェアはユーザの要求通りに動作するようその仕様が決定され、それを実現するアルゴリズムが構築された上で特定のプログラム言語により実装されることから、ソフトウェアは論理の集合体であって物理的実体をもつことがなく、実装されたアルゴリズムそのものは不可視である（ソースプログラムが可視であることとは本質が異なることに注意されたい）。したがって、不具合の原因（バグ、プログラムの論理的な誤り）をプログラム中から検出し、修正することは、ハードウェアに比して本質的に困難であることが知られている。

このような、ソフトウェアの特性に起因する品質の作り込みの困難さに対して、従来より各開発者・研究者らはこぞって様々なアプローチを展開し、改善を試みて来ている。たとえば、開発プロセス（つまり、ソフトウェアの作り方）を改善することにより、アルゴリズム実装上の論理的誤りの不用意な混入を避ける、あるいはプログラム言語仕様自体に制限を加え、バグの混入を防ぐ、またソフトウェアのテストの効率化のための方法論と実践法の開発などが挙げられる。

このような中、本代表者はこれまで、ソフトウェア開発において必ず実施される、開発の最終段階であるテスト工程から採取可能な時系列データに基づいて、テストの進捗と信頼性の度合い、つまり「あとどれくらいテストを続ければ、どの程度信頼度が改善されるか」あるいは「残りバグ数はいくらと見なせるか」などの定量的な評価法について、既存の研究成果（自らのもの以外も含む）を改良することを行ってきた。特に、データを分析してソフトウェアに潜在するバグ数やソフトウェア信頼度などを統計的に推定する

方法について取り組んできており、これを進めたいと考えた。これらが本研究の背景と動機となっている。

2. 研究の目的

前項に述べたソフトウェア開発管理における問題、特にソフトウェアの品質や信頼性に関する事項については、ソフトウェア製品が「仕様通りに動作して当たり前」であるべきという立場から、ソフトウェア品質・信頼性評価手法の更なる精緻化を目指すことを目的とした。特にソフトウェア開発の上流工程から採取し得るソフトウェア開発の属性データを、品質評価手法に組み込むことにより、従来のソフトウェア品質・信頼性評価手法をより正確な評価結果をもたらすよう改善することを目指した。同時に、そのようなソフトウェア開発作業に関する属性データには定量的なもの定性的なものが混在することや、それらにはソフトウェア開発の初期に値が確定し易いもの、あるいは末期に観測されるものなどの特性があることを踏まえ、当該ソフトウェアの出荷後の不具合発生数をできるだけ早い時点で精密に予測して、ソフトウェア開発作業の修正を on-the-fly にて実現させるため、早期に観測できる属性データを用いることに注力することとした。

3. 研究の方法

(1)平成27年度における取り組み

IPA/SEC（情報処理推進機構／ソフトウェアエンジニアリングセンター）から提供を受けた、ソフトウェア開発企業体から収集したソフトウェア開発プロセスに関するアンケートデータを利用して、その項目の一つである、「当該開発ソフトウェアの出荷後、1か月以内のバグの発生有無」をソフトウェアプロセスの良さの指標として着目し、それをゼロとできた企業体・ソフトウェアプロセスはどのような特徴があるのかについて、多変量解析的視点に基づいて調べた。具体的な特徴として、データ収集の際、それに協力した開発現場があまり労力を要しない、つまり定量的な数字を求めない形式のデータを説明変数として採用することとした。また、それにより質的変数が説明変数にも表れることから、いわゆる数量化理論 類を採用することが考えられたが、すでに得ている知見から、線形なモデルではうまくいかないことが予想されたため、新規に、機械学習の一つである Random Forests（ランダムフォレスト）法を採用して、精度のよいモデルが構築できるかについて調べた。また、派生的成果もあり、別の機械学習の手法として知られる、RBF（放射基底関数）ネットワークの基本性能を高めるため、コピュラ関数を導入することを試みた。このコピュラ関数の扱いについては更にいくつかの新しい知見が得られた。

(2)平成28年度における取り組み

前年度に引き続き、IPA/SEC との契約に基

づいて、当該研究代表者に貸与されているソフトウェアプロジェクトに関するデータセットを利用することとした。データの詳細を述べておくと、これは2005年から2012年までの間に、国内24社の企業が、実際に開発を行ったソフトウェア開発プロジェクトについて、開発形態がどのようなものであったか、出荷するソフトウェアが対象とする業務の種類、使用言語の種類などのプロジェクトの外形的性質、また定量化されたデータである、出荷後1か月以内の不具合の数などが一覧表の形式となっているものである。その規模は3089件のプロジェクトに対して、611項目について各社の担当者から得られた回答からなるものである。

本研究の主題である、ソフトウェアプロジェクトの効率性を高め、ソフトウェア製品の品質を精密に評価するための科学的方策の立案について、この年度の検討においては特に、データマイニングの観点から進め、興味のある研究項目としては、「どのようなソフトウェアプロジェクトが成功しにくい」かを判別したい、というものであった。これについては昨年度あるいはそれ以前の研究によって、上記のデータセットが、非常に欠測が多いことや定性的に答えるよう指示されたものがかなり多いことなどの特徴を持つことから、準備的研究である、欠測値の補完方法や定性的データに基づくニューラルネットワークの利用による出荷後のソフトウェア品質評価法の精緻化などを行った。

(3)平成29年度における取り組み

ソフトウェア開発プロジェクトの成果物である出荷されたソフトウェア製品の運用信頼性を、その出荷前のできるだけ早い時期に推定する問題において、構築した推定モデルに与えるデータセットの質が悪いことに、本年度は引き続き着目した。過去の(他の研究者によるものも含めた)研究では、多変量解析の要領で、ソフトウェアプロジェクトにおいて計測・記録される情報の中から説明変数を選び、目的変数を出荷後からある期間内に発生したソフトウェア故障数として、モデルの推定を行うといった簡易な方法によってそのモデルの適合性等を評価した。結果としては単純な線形回帰ではなく、非線形であることがモデルには求められること、また機械学習による予測モデルの方が総じて予測性能が高いことが示された。しかしながら、これらのモデルの同定に用いるデータセットが、実のところ信用が置けないものも含まれていることが看過できない問題として顕れてきた。利用したデータセットは、数百社を超える一般の企業の協力を得て採取された膨大なものであるが、ソフトウェア開発に携わる実務家が多忙な業務の合間に時間を割いて採取する形式のものについては、非常にノイズが多いと見做さざるを得ないものもあり、そのためモデルの精度が上がらないのではと推察された。そこで、データ(情報)

採取にコスト・時間が掛からず実務家に負担を強くないで採取される外形的なデータや、正確な定量的な値を要求せず、大・中・小のような荒い定性的なデータを用いて、信頼性予測が行えるかについて、モデルを開発し検討した。

4. 研究成果

研究期間初期においては、研究方法に述べた事項に基づき、ソフトウェアプロジェクトが多くは定性的・定量的データで表されるとき、失敗プロジェクトあるいは成功プロジェクトがどのような変数によって特徴づけられるかについて検討を行った。特に多変量データで、周辺度数が扱いにくい分布形状となっているようなものの処理法について検討した。また具体的なデータマイニング手法として、Random Forestsを実装するとともに、欠測値の補完方法や定性的データによる多変量解析に対する予測性能の向上を狙った。また本研究の目的達成に寄与する基礎技術について進捗が見られた。最終的には、機械学習(ニューラルネットワーク)に基づく手法がかなりよい予測を行えることが示された。この成果は一般論文誌での公表等には至らなかったが、一部は本学理工学研究科の修士論文「ソフトウェアプロジェクトデータに基づく運用信頼性の予測に関する研究」(著者:荻原佑実,研究協力者)としてまとめられた。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計4件)

Shuhei Ota and Mitsuhiro Kimura, A statistical dependent failure detection method for n-component parallel systems, Reliability Engineering and System Safety, 167, 2017, 376-382 (査読あり)
DOI: 10.1016/j.ress.2017.06.022

Shuhei Ota and Mitsuhiro Kimura, A Study on Regression Analysis by Expanded RBF Network Based on Copula with Linear Correlation and Rank Correlation, International Journal of Reliability, Quality and Safety Engineering, 22, 2015, 1550022 (査読あり)
DOI:10.1142/S0218539315500229

Shuhei Ota, Takao Kageyama and Mitsuhiro Kimura, Improvement of Reliability Evaluation for 2-Unit Parallel System with Cascading Failures by Using Maximal Copula, IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, E98-A, 2015, 2096-2100 (査読あり)
DOI: 10.1587/transfun.E98.A.2096

池田貴彦, 大原衛, 福本聡, 新井雅之, 岩

崎一彦, 木村光宏, ファイルバージョン管理機能を備えた分散データレプリケーションプロトコルの提案, 電子情報通信学会論文誌, J98-D, 2015, 684-699 (査読あり)
DOI: 10.14923/transinfj.2014JDP7061

[学会発表](計18件)

Mitsuhiro Kimura, Shunsaku Fujisawa and Shinji Inoue, A unified expression of Gompertz and logistic curves and its discretization, 2nd International Conference on System Reliability and Safety (ICRSRS2017), 2017年12月20日~2017年12月22日, Milan, Italy

齋藤仁貴, 太田修平, 木村光宏, 標本j次モーメントを用いた適合度検定について, 日本オペレーションズ・リサーチ学会中国・四国支部, 2017年09月07日~2017年09月08日, 中央森林公園研修室(広島県・三原市)

安部彰悟, 木村光宏, Box-Cox変換と不完全ガンマ関数で与えられる曲線群について, 日本オペレーションズ・リサーチ学会中国・四国支部, 2017年09月07日~2017年09月08日, 中央森林公園研修室(広島県・三原市)

今井純哉, 森村駿, 木村光宏, 相関を考慮したRBFネットワークに対するFireflyアルゴリズムの導入について, 日本オペレーションズ・リサーチ学会中国・四国支部, 2017年09月07日~2017年09月08日, 中央森林公園研修室(広島県・三原市)

太田修平, 木村光宏, n素子並列システムにおけるカスケード故障の尤度比検定を用いた発生検知に関する一考察, 電子情報通信学会信頼性研究会, 2017年07月28日~2017年07月28日, 稚内サンホテル(北海道・稚内市)

Shuhei Ota and Mitsuhiro Kimura, A study on reliability deterioration and improvement of coherent systems under dependent failure-occurrence environment, 10th International Conference on Mathematical Methods in Reliability (MMR2017), 2017年07月03日~2017年07月06日, Grenoble, France

太田修平, 木村光宏, A note on a statistical detection method of cascading failure for parallel systems, 日本信頼性学会, 第25回春季信頼性シンポジウム, 2017年05月31日, 一般財団法人日本科学技術連盟本部(東京都・新宿区)

木村光宏, 藤澤峻作, 井上真二, ゴンペルツ曲線とロジスティック曲線の一般化とその離散化について, 電子情報通信学会信頼性研究会, 2017年05月26日~2017年05月26日, ピュアリティまきび(岡山県・岡山市)

太田修平, 木村光宏, 依存故障するn素子直並列システムの信頼性解析~EFGMコンピュータを用いて~, 日本オペレーションズ・リサーチ学会第18回信頼性研究会(招待講演) 2016年12月17日~2016年12月17日, マ

ホロバマインズ三浦(神奈川県・三浦市)

Shuhei Ota and Mitsuhiro Kimura, A study on the MTTF of parallel systems under dependent failure-occurrence environment, 7th Asia-Pacific International Symposium on Advanced Reliability and Maintenance Modeling (APARM2016), 2016年08月24日~2016年08月26日, Seoul, Korea

太田修平, 木村光宏, 2ユニット並列システムにおける故障時間データに基づく従属故障の統計的検知について, 電子情報通信学会信頼性研究会, 2016年07月29日~2016年07月29日, 小樽経済センター(北海道・小樽市)

Shuhei Ota and Mitsuhiro Kimura, A method of dependent failure-occurrence detection for parallel systems, 9th IMA International Conference on Modelling in Industrial Maintenance and Reliability (MIMAR2016), 2016年07月12日~2016年07月14日, London, UK

Mitsuhiro Kimura, One-step-ahead MTTF Prediction for Online Software Reliability Monitoring with Bootstrap Scheme, 2nd East Asian Workshop on Industrial Engineering (EAWIE2015), 2015年11月06日~2015年11月07日, Seoul, Korea

Shuhei Ota and Mitsuhiro Kimura, A Dependent-Multivariate Data Analysis Method by Skewed-RBF Network Based on FGM Copula, 21st ISSAT International Conference on Reliability & Quality in Design (RQD2015), 2015年08月06日~2015年08月08日, Philadelphia, U.S.A.

Yuichiro Arai and Mitsuhiro Kimura, Operational Software Reliability Prediction by Random Forest Based on Development Project Data with Qualitative Variables, 21st ISSAT International Conference on Reliability & Quality in Design (RQD2015), 2015年08月06日~2015年08月08日, Philadelphia, U.S.A.

影山孝夫, 太田修平, 木村光宏, コピュラによるネットワークシステムの従属故障に関する考察, 電子情報通信学会信頼性研究会, 2015年07月31日~2015年07月31日, ホテルニューグリーン(青森県・むつ市)

太田修平, 木村光宏, FGMコンピュータを用いたRBFネットワークの一拡張について, 電子情報通信学会信頼性研究会, 2015年06月19日~2015年06月19日, 機械振興会館(東京都・港区)

Mitsuhiro Kimura, Naomichi Hata, and Takaji Fujwara, Hidden Markov Analysis for Software Testing Performance Evaluation Based on Two-stage Testing by Two Teams, 9th International Conference on Mathematical Methods in Reliability 2015 (MMR2015), 2015年06月01日~2015年06

月 04 日 , Tokyo, Japan

〔図書〕(計 2 件)

Syouji Nakamura, Cun Hua Qian, Toshio Nakagawa (Eds.), Mitsuhiro Kimura (chapter 8 を分担執筆), Reliability Modeling with Computer and Maintenance Applications, World Scientific, 2017, 396 (155-174)

K. M. Jung, M. Kimura, L.-R. Cui (Eds.), Advanced Reliability and Maintenance Modeling VII: Recent Developments on Reliability, Maintainability and Dependability, McGraw-Hill International Enterprises, LLC, Taiwan Branch, 2016, 634 (共同編集のため執筆ページは特定できない)

6 . 研究組織

(1)研究代表者

木村 光宏 (KIMURA, Mitsuhiro)

法政大学・理工学部・教授

研究者番号 : 20263486

(2)研究協力者

新井 雄一郎 (ARAI, Yuichiro)

太田 修平 (OTA, Shuhei)

荻原 佑実 (OGIWARA, Yumi)