様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　６ 年　 ６ 月　 ６ 日現在

機関番号：１６３０１
研究種目：基盤研究(C)（一般）
研究期間：2021～2023
課題番号：２１Ｋ１１８３１

研究課題名（和文）開発者に依存しやすい品質特性の定量的分析と自動評価法の開発

研究課題名（英文）Quantitative Analysis and Automated Evaluation of Software Quality Characteristics Susceptible to Developers

研究代表者

阿萬　裕久（Aman, Hirohisa）

愛媛大学・総合情報メディアセンター・教授

研究者番号：５０３３３５１３

交付決定額（研究期間全体）：（直接経費）　　3,200,000円

研究成果の概要（和文）：本研究では，開発者に依存しやすい品質特性として考えられる次の観点について定量的な分析とツールの開発を行った：（1）変数・関数の名前，（2）ソースコードの書き方．
特に，変数や関数に対する命名には開発者の好みや経験が影響しやすく，一般的なコーディング規約では規定や評価が難しい．それゆえ，ソースコード中の名前は適切であるのか？可読性を損ねていないか？という点について分析を行ったところ，プログラムの文脈にそぐわない名前になってしまったり，名前そのものは適切でも他と酷似していて紛らわしい名前集合を形成してしまっていたりする実例が見られた．そして，それらの検出を自動的に行うためのツールを開発して公開した．

研究成果の学術的意義や社会的意義
本研究では，変数・関数の名前やプログラムの書き方といった開発者に依存しやすい特性注目している．これらの特性は属人性として排除されがちであったが，重要な観点であると研究代表者らは考え，これらに注目したデータ収集と分析を行った．
特に，変数の名前を分かりやすいものにすべきというのは一般的なプラクティスとしてコンセンサスが得られているものの，個々の名前を分かりやすいものに仕上げようとして長い説明的な名前になり，結果的に他と区別しにくい紛らわしい変数ペアが作られてしまいかねない．これまでそのような他の変数との名前の類似性については十分に研究されておらず，本研究ではその先駆けとなる学術論文を発表できた．

研究成果の概要（英文）：This study focused on the following quality characteristics that are likely to depend on the individual developer: (1) names of variables and functions, and (2) way of writing source code. Especially, naming variables and functions tend to depend on the developers' preference and experience, and it is hard to assess their names using the general coding conventions. Thus, we analyzed a lot of source programs from the perspectives of the names' appropriateness and readability. As a result, we encounters many instances in which variable/function names do not correspond to their program context. Moreover, we found instances that variable names are highly similar to each other and form a set of confusing variable pairs even though individual names are easy to understand. Hence, we examined those characteristics and developed automated tools for detecting problematic names in a program.

研究分野：ソフトウェア工学

キーワード：変数名　関数名　名前の類似性　人的要素　可読性　プログラムの書き方　ソフトウェア品質

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

(1) ソフトウェアは日常のありとあらゆる場面で活用され，安心・安全で便利な社会生活の基盤を成している．それだけに，ソフトウェアに不具合があった場合にはそれが社会へ及ぼす影響は極めて大きい．不具合の無いソフトウェアを開発して保守し続けることが理想ではあるが，実際のところソフトウェアの開発及び保守は人間による知的作業であり，人為的な誤りによるバグ（不具合要因）混入の可能性をゼロにするのは難しい．一般に，高品質なソフトウェアを開発し，その品質を維持し続けるには，適切な品質管理活動が必要不可欠である．具体的には，明確な評価基準と指標（目標値，基準値）を定めたり，それに合わせたチェック体制を整えたりといった活動が必要である．実際，ソフトウェアメトリクスを用いてプログラムの特徴を定量的に測定し，測定データを品質管理へ役立てるという研究や活動は従来から行われてきている．例えば，プログラムの規模や複雑さを数値化し，所定の閾値を超えるようであれば改善を促すといったことが代表例として挙げられる．これに関連して，所定のコーディング規約やガイドラインに従ってプログラムの書き方を確認するツールも研究・開発されており，品質管理活動の後押しとなっている．

(2) 上述した活動はいずれも有益なものではあるが，それらにおける視点は主としてプログラムの構造に対する一般的な傾向やルールに重きが置かれている．換言すれば，意味的な観点を含め，作業者に依存する特徴については十分に議論されていない．例えば，プログラムの中で使われる変数に対して「変数の数が多すぎる」という指摘は自動的にできても「変数の名前が不適切である」という指摘を自動的に行うのは難しい．後者の指摘については，漠然としたガイドラインはあるものの，定量的な基準や指標は見当たらない．

(3) プログラミングは人間の知的作業であるが故に，開発者の好みや感覚に依存する要素の影響は大きい．これを単なる“個人差”として切り捨てるのではなく，品質管理を行う上で適切に把握・活用していく必要があるのではないかと考えられる．ソフトウェアの開発・保守ではリポジトリの利用が進んでおり，それによりプログラムの１行１行について“誰が，いつ，何を”記述・修正したのかを追跡可能になっている．それゆえ，作業者情報の自動収集も可能であり，プログラムの特徴を作業者と紐付けできる．よって，開発者の違いが及ぼす影響を定量的に分析できるといえる．さらに，自然言語処理や機械学習の研究成果をツールやライブラリとして容易に利用可能になってきていた．従来のソフトウェア工学研究では，プログラムの構造に着目したアプローチが主流であり意味論的なアプローチは難しかったが，自然言語処理技術・機械学習技術を応用することで“人間の感覚”に近い評価も可能になりつつあった．

２．研究の目的

(1) 本研究では，プログラミングにおいてその作業者が影響すると思われる特徴，特に経験や好みによって違いが出やすい特徴に着目する．そして，そのような特徴が品質に及ぼす影響を定量データ分析の立場から明らかにしつつ，自然言語処理技術や機械学習技術を活用することで人手による評価に近い，意味的な側面に踏み込んだ評価を自動化することを目指す．

(2) さらに本研究では，変数に対する名前の付け方の評価（名前の特異性や紛らわしさ）について深堀りする．これまで，より良い変数名についての研究はいくつか行われてきているが，それらは個々の変数名のあり方についての研究であり，他の変数名との関係性については着目していなかった．しかしながら，個々の変数名が説明的で分かりやすいものであったとしても，他の変数と名前が酷似していた場合にはその紛らわしさ故に変数の取り違えといったミスを誘発してしまう．この点について名前の類似性を定量的に評価する仕組みを提案し，ツールのかたちで自動検出・警告できるようにする．

３．研究の方法

(1) さまざまな分野の開発プロジェクト（数百以上）から開発データを収集し，変数・関数の名前の付け方やプログラムの書き方，コメントの書き方といった，開発者の特色が出やすい特徴についてデータベースを構築する．特徴量の抽出には，プログラム解析のみならず自然言語処理技術も応用していく．そして，統計モデルを使ったデータ分析を駆使してどういった特徴が（バグの予測等に）どれほど有用であるかを明らかにし，さらには機械学習モデルを活用することで自動化可能な評価法の提案を行う．その上で，積極的に論文発表を行い，他の研究者との議論を通

じて評価法の洗練化を図る．

(2) 上述の活動で見出された特徴とその評価法について，その妥当性の評価とツール化（知の還元）に取り組む．例えば「変数名の紛らわしさ」について，実際に変数にさまざまな名前の付け方を行ったプログラムを用意し，その内容の理解や修正のしやすさに関するアンケート調査を行って，提案手法による評価の妥当性を確認する．そして，妥当性を確認できた手法について，その評価を自動的に行えるツールを開発して一般に公開していく．

４．研究成果

(1) GitHub で公開されているオープンソースの Java ソフトウェア 1000 件について，そこで使われている変数名とその特徴量（変数の名前の長さ，型，スコープの長さ等）を調べ，そのデータに対する定量的な分析を行った．その結果，特にスコープが広い変数の場合にはどういった単語が変数名には使われやすいのか，どういったスタイルの名前が多いのか，どういった長さの長さの名前になりやすいのかを明らかにした．さらには，スコープが広い場合にはプログラマたちはフルスペルの英単語や複合語を好む傾向にある一方，長過ぎる名前は敬遠しがちであることもわかった．実際，好んで使われる 1 つの単語や略語の長さの上限は 7-8 文字程度であり，これよりも長い名前は別のより短い省略形へ短縮されやすい傾向が見て取れた．この研究で得られた成果（データセット）は大学の Web サーバ上で公開しているが，実際に海外の研究者がこれを使った論文をいくつか発表している．

(2) オープンソースの Java ソフトウェア 1876 件と Python ソフトウェア 2427 件について，そこで使われている変数を収集し，同じスコープで利用可能な変数のペアについて変数名の類似性を定量的に評価する手法を提案した．提案手法では文字列としての類似度と（自然言語処理技術を活用した）意味的な類似度の 2 種類を考慮し，それぞれの観点から 2 つの変数名がどれほど似ているかを数値化した．そして，アンケート調査を通じて，どの程度類似度が高いと 2 つの変数を取り違えてしまう恐れがあるかを明らかにし，その閾値を超えるようであれば自動的に警告できるツールを開発して GitHub で公開した．この成果をまとめた論文は，ソフトウェア工学分野でのトップジャーナルの 1 つである Empirical Software Engineering に掲載された．そして，その功績から情報処理学会ソフトウェア工学研究会より卓越研究賞を受賞した．2024 年には本件に関して，招待講演の依頼を 2 件受けている．さらに，この研究を時系列解析へ発展させた論文が Core ランク A の国際会議である EASE (International Conference on Evaluation and Assessment in Software Engineering) の Research Track で採録され，2024 年 6 月に発表予定である．

(3) 変数の名前の自動評価を行うため，自然言語処理技術 Doc2Vec ならびに大規模言語モデル CodeBERT を活用した手法を提案した．本研究では評価対象の変数を使っているプログラム断片を一種の文書と見なし，Doc2Vec を使った研究ではその文書のベクトル表現を活用して，類似したプログラム断片でどのような変数名が使われているかを検索するという手法をとった．いわばそのプログラム断片は当該変数の使われ方を表した文書であるため，仮に当該変数が妥当な名前になっていれば，類似したプログラム断片でも類似した名前の変数が使われているはずである．この考え方のもとで変数名の自動評価を行う手法を提案した．さらに近年注目されている大規模言語モデルも活用し，CodeBERT というソースコードで事前学習されたモデルを使って変数名のマスク予測を行い，当該変数をマスクした場合に元の変数名が適切に復元できるかどうかでもって元の変数名の妥当性を自動評価するという手法も提案した．これらの提案の妥当性は他の研究者からも認められ，学術雑誌論文ならびに国際会議論文としてそれぞれ出版できた．

(4) 関数の名前の自動評価を行うため，Transformer アーキテクチャを活用した手法を提案した．Transformer は近年注目を集めているニューラルネットワークのアーキテクチャであり，本研究ではこれを使って関数の内容から当該関数の名前に該当するキーワード（関数名の先頭に出現すべき単語）を予測するというタスクにこれを利用した．これまでに Doc2Vec，Word2Vec 及び畳込みニューラルネットワークを使ったモデルが提案されていたが，本手法ではその先行手法を上回る性能でもって不適切な関数名を自動的に検出・警告できることを実験を通じて確認できた．そして，その成果を学術雑誌論文として出版できた．

５．主な発表論文等

〔雑誌論文〕　計6件（うち査読付論文　6件／うち国際共著　0件／うちオープンアクセス　4件）

１．著者名 Aman Hirohisa、Amasaki Sousuke、Yokogawa Tomoyuki、Kawahara Minoru	４．巻 28
２．論文標題 An automated detection of confusing variable pairs with highly similar compound names in Java and Python programs	５．発行年 2023年
３．雑誌名 Empirical Software Engineering	６．最初と最後の頁 108:1--108:32
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s10664-023-10339-2	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 OHARA Kosuke、AMAN Hirohisa、AMASAKI Sousuke、YOKOGAWA Tomoyuki、KAWAHARA Minoru	４．巻 E106.D
２．論文標題 A Comparative Study of Data Collection Periods for Just-In-Time Defect Prediction Using the Automatic Machine Learning Method	５．発行年 2023年
３．雑誌名 IEICE Transactions on Information and Systems	６．最初と最後の頁 166～169
掲載論文のDOI（デジタルオブジェクト識別子） 10.1587/transinf.2022MPL0002	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 峯久 朋也、阿萬 裕久、川原 稔	４．巻 39
２．論文標題 Transformerによるメソッド名推定を活用したネーミングバグの検出	５．発行年 2022年
３．雑誌名 コンピュータ ソフトウェア	６．最初と最後の頁 4_17～4_23
掲載論文のDOI（デジタルオブジェクト識別子） 10.11309/jssst.39.4_17	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 Amasaki Sousuke、Aman Hirohisa、Yokogawa Tomoyuki	４．巻 27
２．論文標題 An extended study on applicability and performance of homogeneous cross-project defect prediction approaches under homogeneous cross-company effort estimation situation	５．発行年 2022年
３．雑誌名 Empirical Software Engineering	６．最初と最後の頁 46:1～46:29
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s10664-021-10103-4	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 Amasaki Sousuke、Aman Hirohisa、Yokogawa Tomoyuki	4．巻 27
2．論文標題 An extended study on applicability and performance of homogeneous cross-project defect prediction approaches under homogeneous cross-company effort estimation situation	5．発行年 2022年
3．雑誌名 Empirical Software Engineering	6．最初と最後の頁 46:1--46:29
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s10664-021-10103-4	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 山中 啓太、阿萬 裕久、川原 稔	4．巻 38
2．論文標題 プログラムスライスとDoc2Vecを用いた変数名評価法の提案	5．発行年 2021年
3．雑誌名 コンピュータソフトウェア	6．最初と最後の頁 4_9--4_15
掲載論文のDOI（デジタルオブジェクト識別子） 10.11309/jssst.38.4_9	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

〔学会発表〕　計25件（うち招待講演　1件／うち国際学会　13件）

1．発表者名 Yahiro Mori, Hirohisa Aman, Sousuke Amasaki, Tomoyuki Yokogawa, and Minoru Kawahara
2．発表標題 An Application of Program Slicing and CodeBERT to Distill Variables With Inappropriate Names
3．学会等名 The 22nd IEEE/ACIS International Conference on Software Engineering, Management and Applications（国際学会）
4．発表年 2024年

1．発表者名 森 哉尋, 阿萬 裕久, 川原 稔
2．発表標題 変数の型名と代入式に着目した命名パターンと大規模言語モデルを活用した変数名評価に関する考察
3．学会等名 電子情報通信学会ソフトウェアサイエンス研究会
4．発表年 2024年

１．発表者名
伏原 裕生, 阿萬 裕久, 川原 稔

２．発表標題
テストコードにおけるテストスメルの存在とバグ潜在性の関係に関する定量的調査

３．学会等名
電子情報通信学会ソフトウェアサイエンス研究会

４．発表年
2024年

１．発表者名
Shinnosuke Irie, Hirohisa Aman, Sousuke Amasaki, Tomoyuki Yokogawa and Minoru Kawahara

２．発表標題
A Comparative Study of Hybrid Fault-Prone Module Prediction Models Using Association Rule and Random Forest

３．学会等名
The 5th World Symposium on Software Engineering（国際学会）

４．発表年
2023年

１．発表者名
Yuki Fushihara, Hirohisa Aman, Sousuke Amasaki, Tomoyuki Yokogawa, and Minoru Kawahara

２．発表標題
A Trend Analysis of Test Smells in Python Test Code Over Commit History

３．学会等名
The 49th Euromicro Conference on Software Engineering and Advanced Applications（国際学会）

４．発表年
2023年

１．発表者名
大嶋 琉太, 阿萬 裕久, 川原 稔

２．発表標題
記号実行とミューテーションを活用したプログラム正誤判定の効率化

３．学会等名
第24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
高橋 佑介，阿萬 裕久，川原 稔

２．発表標題
スペクトル情報とソースコード行の新しさを組み合せたバグ限局手法

３．学会等名
第24回ソフトウェア工学の基礎ワークショップ

４．発表年
2024年

１．発表者名
阿萬裕久

２．発表標題
ソフトウェア工学におけるデータサイエンス

３．学会等名
電気学会 2023年1月19日-2023年1月20日通信研究会（招待講演）

４．発表年
2023年

１．発表者名
大嶋 琉太，阿萬 裕久，川原 稔

２．発表標題
プログラム正誤判定におけるプログラムのベクトル化と類似度評価の関係について

３．学会等名
情報処理学会ウィンターワークショップ2023

４．発表年
2023年

１．発表者名
高橋 佑介，阿萬 裕久，川原 稔

２．発表標題
SBFL手法における疑惑値の分布とバグ限局精度の関係について

３．学会等名
情報処理学会ウィンターワークショップ2023

４．発表年
2023年

１．発表者名
Sousuke Amasaki, Hirohisa Aman and Tomoyuki Yokogawa

２．発表標題
An Evaluation of Cross-Project Defect Prediction Approaches on Cross-Personalized Defect Prediction

３．学会等名
PROFES 2022: Product-Focused Software Process Improvement（国際学会）

４．発表年
2022年

１．発表者名
Kazuki Wayama, Tomoyuki Yokogawa, Sousuke Amasaki, Hirohisa Aman, Kazutami Arimoto

２．発表標題
Verifying Game Logic in Unreal Engine 5 Blueprint Visual Scripting System Using Model Checking

３．学会等名
The 37th IEEE/ACM International Conference on Automated Software Engineering: Workshop ASE4Games（国際学会）

４．発表年
2022年

１．発表者名
Tenma Kita, Hirohisa Aman, Sousuke Amasaki, Tomoyuki Yokogawa, and Minoru Kawahara

２．発表標題
Have Java Production Methods Co-Evolved With Test Methods Properly?: A Fine-Grained Repository-Based Co-Evolution Analysis

３．学会等名
The 48th Euromicro Conference on Software Engineering and Advanced Applications（国際学会）

４．発表年
2022年

１．発表者名
Sousuke Amasaki, Hirohisa Aman, and Tomoyuki Yokogawa

２．発表標題
An Evaluation of Effort-Aware Fine-Grained Just-in-Time Defect Prediction Methods

３．学会等名
The 48th Euromicro Conference on Software Engineering and Advanced Applications（国際学会）

４．発表年
2022年

１．発表者名
Kazuma Toyota, Tomoyuki Yokogawa, Sousuke Amasaki, Hirohisa Aman and Kazutami Arimoto

２．発表標題
A Visual Modeling Environment for the nuXmv Model Checker Intended for Novice Users

３．学会等名
The 7th International Conference on Enterprise Architecture and Information Systems（国際学会）

４．発表年
2022年

１．発表者名
大嶋 琉太, 阿萬 裕久, 川原 稔

２．発表標題
プログラムのベクトル化と記号実行を活用した正誤判定の効率化

３．学会等名
第23回ソフトウェア工学の基礎ワークショップ

４．発表年
2022年

１．発表者名
峯久 朋也, 阿萬 裕久, 川原 稔

２．発表標題
メソッド名の整合性評価のためのデータセット

３．学会等名
第23回ソフトウェア工学の基礎ワークショップ

４．発表年
2022年

１．発表者名
高橋 亮至, 阿萬 裕久, 川原 稔

２．発表標題
Pycodestyleによる警告とバグ修正の関係に関する定量分析

３．学会等名
情報処理学会第212回ソフトウェア工学研究会

４．発表年
2022年

１．発表者名
峯久 朋也, 阿萬 裕久, 川原 稔

２．発表標題
機械学習によるメソッド名推定を活用したネーミングバグの検出

３．学会等名
ソフトウェア信頼性研究会 第16回ワークショップ

４．発表年
2022年

１．発表者名
Sousuke Amasaki, Hirohisa Aman and Tomoyuki Yokogawa

２．発表標題
Searching for Bellwether Developers for Cross-Personalized Defect Prediction

３．学会等名
22nd International Conference on Product-Focused Software Process Improvement（国際学会）

４．発表年
2021年

１．発表者名
Hirohisa Aman, Sousuke Amasaki, Tomoyuki Yokogawa and Minoru Kawahara

２．発表標題
An Investigation of Compound Variable Names Toward Automated Detection of Confusing Variable Pairs

３．学会等名
1st Workshop on Automated Support to Improve code Readability（国際学会）

４．発表年
2021年

１．発表者名
Hirohisa Aman, Sousuke Amasaki, Tomoyuki Yokogawa and Minoru Kawahara

２．発表標題
A Large-Scale Investigation of Local Variable Names in Java Programs: Is Longer Name Better for Broader Scope Variable?

３．学会等名
14th International Conference on Quality of Information and Communications Technology（国際学会）

４．発表年
2021年

１．発表者名
Sousuke Amasaki, Hirohisa Aman and Tomoyuki Yokogawa

２．発表標題
A Preliminary Evaluation of CPDP Approaches on Just-in-Time Software Defect Prediction

３．学会等名
47th Euromicro Conference on Software Engineering and Advanced Applications（国際学会）

４．発表年
2021年

１．発表者名
Tomoya Minehisa, Hirohisa Aman, Tomoyuki Yokogawa and Minoru Kawahara

２．発表標題
A Comparative Study of Vectorization Approaches for Detecting Inconsistent Method Names

３．学会等名
18th IEEE/ACIS International Virtual Conference on Software Engineering, Management and Applications（国際学会）

４．発表年
2021年

１．発表者名
峯久 朋也, 阿萬 裕久, 川原 稔

２．発表標題
ソースコードの難読化解除手法を活用したメソッド名の整合性評価

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

４．発表年
2021年

〔図書〕　計0件

〔産業財産権〕

〔その他〕
発表論文 -- ソフトウェア工学研究室
https://se.cite.ehime-u.ac.jp/jp/research/paper/
Supplementary Materials for "An Automated ..." --- https://zenodo.org/record/7493554#.ZEyDKM7P24Q
cvpfinder -- https://github.com/amanhirohisa/cvpfinder

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
研究分担者	天嵜　聡介 (Amasaki Sousuke) (00434978)	岡山県立大学・情報工学部・准教授 (25301)	
研究分担者	横川　智教 (Yokogawa Tomoyuki) (50382362)	岡山県立大学・情報工学部・准教授 (25301)	

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