

識別子の難読化と命名の収集による命名学習方法の提案

式見 遼†

松浦 佐江子†

†芝浦工業大学 理工学研究科 電気電子情報工学専攻

1 はじめに

現在のソフトウェア開発において、他人の書いたソースコードやライブラリを利用することは殆どの場合避けられない。また、自分で定義したプログラムを他人が利用する場合もある。このように、他人のプログラムを理解したり、他人が使うプログラムを自分が書いたりするとき、識別子に適切な命名を行えることは非常に重要である。こうした命名の重要性は広く知られているが、適切な命名の客観的な指標が無いため命名を学習する方法は確立されていない。

我々はこれまでにソースコードの識別子名を変更し、そのソースコードから元の識別子名を推測することで、識別子の命名を学習する手法を提案した。さらに、名前を変更する識別子の種類を変えることで目的別の学習が可能であることを示した[1]。しかし、[1]では適切な命名の客観的な評価結果が課題であった。そこで本論文では、適切な命名の客観的な指標として、複数人の学習者の各々が考える命名のうち、最も多くの学習者がつけた識別子の名前を適切な命名と評価する手法を提案する。そして先行研究による命名の学習手法と提案手法による命名の評価を組み合わせた学習環境の実装事例として、web ベースでの学習支援環境を示す。

2 学習手法

本手法における識別子の命名学習は、学習問題の出題、学習者の解答、解答の評価からなる。本節では、これらの学習過程を順番に説明する。

2.1 問題の出題 - 難読化

教授者は適当なソースコードについて、任意の識別子の名前を置換したソースコードを作成する。この操作を難読化と呼ぶ。難読化したソースコードはコンパイル・実行が可能で、その挙動はオリジナルと全く同じである。

生成した難読化コードは「適切な識別子名を考える問題」として、学習者に出題する。

2.2 学習者の解答

学習者は出題された難読化コードについて、

コードの構造や実行結果から、各識別子について適切だと思われる名前を推測し解答する。

2.3 解答の評価 - 命名の収集

識別子の適切な命名とは、他の人がソースコードを見たとき、その識別子の役割を理解できる名前である。逆に言えば、多くの人が定義した名前は、少数の人が定義する名前よりも命名として適切であることが多いと考えられる。

そこで、本手法では全学習者のソースコードから定義された識別子名の集計を行う。集計結果からわかる、多くの学習者が定義した識別子名は、多くの人がその識別子の役割を理解できるという意味で、適切だと考えられる。また、この収集結果を学習者に公表することで、学習者が一般的に適切である命名を直感的に理解する効果が期待できる。教授者は収集結果の識別子の一覧の何件かを適切な命名と評価し、指導することができる。

学習者の命名技量に対して難読化が難しすぎた場合や学習者の人数が少ない場合、良い命名が多数派にならないことがある。この時、教授者は収集結果を次に出題する難読化コードの難しさに反映させて再度学習を行うことで、学習者に合った命名学習ができる。

3 実装事例 - web 上での命名学習

本手法による学習は Web 教科書[5]に組み込み、web 上で行うことを想定している。Web 教科書とはブラウザ上で動作するプログラミング学習に特化した e ラーニングシステムである。解説と問題からなる教科書コースと、各種言語に対応した IDE を備えており、学習したコードを web 上で実行して確認することができる。

本節では、この Web 教科書上で動作する命名学習支援を学習過程に沿って示す。

3.1 問題の出題 - 難読化

ソースコードの難読化は専用のツールを用いてクライアント環境で行う。学習効果の観点から、難読化を行う際は次の 2 点に注意する。

「対応する識別子の名前を一括で置換し、置換後のソースコードに、コンパイルエラーが起きないようにする」、「難読化は過度に行わず、難読化コードから難読化された識別子の役割を理解できる程度にする」専用ツールはこれらの注意点を踏まえ、「同一識別子の一括難読化」、「ソース外で定義された識別子の難読化の禁

Collaborative Learning to Identifier Naming from Obfuscating

†Ryo Shikimi †Saeko Matsuura

†Graduate School of Engineering Division of Electrical Engineering and Computer Science, Shibaura Institute of Technology

止」, 「難読化ソースコードのプレビュー」などの機能を備えている。

難読化したソースコードは Web 教科書にアップロードして, 難読化された識別子の適切な名前を問う問題として学習者に出題する。

3.2 学習者の解答

学習者による識別子名の解答は図 1 の Web 教科書の IDE 上で行う。IDE はコンパイル, 実行, 識別子の置換機能を持ち, 学習者はソースコードの動作理解と識別子の名前の推測のみに集中することができる。

命名を終えた学習者が Web 教科書上で識別子の名前を確定し解答を送信すると, その解答は Web 教科書によって学習者ごとに収集される。

3.3 解答の評価 - 命名の収集

収集された命名は図 2 のように識別子ごとに, ある名前をつけた学習者が何人いるのかを集計して表示・公開される。

各学習者は多くの学習者が定義した識別子名を良い手本として, 少数の学習者が定義した識別子名を悪い手本として見ることで, 識別子に適切な名前をつける能力を養うことができる。

このように, web の命名学習環境では, 教授者の負担を極力減らした学習が可能である。

4 関連研究

ソースコードの可読性と理解性の向上には識別子の命名の適切さが重要であり, 識別子の命名に関する研究が行われている [2][3][4]。Deißenböck 等[3] は, コーディング時に開発システムの概念知識を蓄え, その知識から命名を指摘する手法を提案した。高精度な評価のために知識の蓄積が必要なこの手法に対し, 知識によらず人間の思考によって評価ができることに本稿手法の優位性があると考えられる。Butler 等[4] は, 不適切な命名がプログラムの品質に影響を与えていることを統計的に示しており, その中で良い命名の条件に「規定される命名規約を守ること」と, 「識別子自身の概念や機能を自然言語で簡潔に表していること」としている。これについて, 本研究の手法では多数決を取ることで自身の概念や機能を簡潔に表す識別子名を評価する事ができていると考えられる。

5 まとめ

本論文では, 初学者を対象とした識別子の命名学習方法として, 多数決による命名の評価や, 他者の解答を閲覧することによる良い命名と悪い命名への理解学習を提案した。

今後取り組むべき課題には, 多数決以外の命名の評価基準として, 命名規約に一致しているかの判定を可能にすることや, 長期的な評価実

験により, 手法の効果を実証することがある。



図 1 Web 教科書における学習画面例

クラス class1		メソッド method1_1	
命名	人数	命名	人数
Rectangle	42	getArea	29
Square	11	getSize	9
Rect	6	menseki	9
class1	1	area	7
		getarea	4
		mult	1
		method1_1	1

図 2 Web 教科書における命名収集結果画面例

6 参考文献

- [1] 式見遼, 松浦佐江子: 命名学習を目的としたソースコード難読化ツールの開発, 第 27 回ソフトウェア科学大会, 2010.
- [2] P.A. Relf, Tool Assisted Identifier Naming for Improved Software Readability: An Empirical Study, Proc of Int'l Symp. Empirical Software Eng., pp.53-62, 2005.
- [3] F. Deißenböck and M. Pizka, Concise and consistent naming, Proc of the 13th International Workshop on Program Comprehension (IWPC 2005), IEEE Computer Society, pp.1-10, 2005.
- [4] S. Butler, M. Wermelinger, Y. Yu, and H. Sharp, Relating identifier naming flaws and code quality: an empirical study, Proc of WCRE'09, pp.31-39, 2009.
- [5] Web Study, <http://www.sayo.se.shibaura-it.ac.jp/incusphere/webStudy.html>