処理内容の類似性から見たコードの類似性判定

西　田　皓　司†

回答者は博士課程前期の学生であり,“ バグを引き起こしたクローンとそうでないクローンとの違い ”という点を主眼に置き，開発の履歴情報に基づいた分析を行っている．クローンの“ 良い ”“ 悪い ”を見分けることができれば，コードクローンのメンテナンスにおける効率を改善することができると考えている．

1. はじめに

回答者は博士課程前期の学生であり，OSS 開発の履歴情報に着目し,“ バグを引き起こしたコードクローン（以下，クローン）とそうでないクローンとの違い ”について分析を行っている．コードクローンの“ 良い ”“ 悪い ”を見分けることができれば，コードクローンのメンテナンスにおける効率を改善することができると考えている．

プログラミングに関しては，研究活動において小規模なコードは書いてきたが，規模の大きいものやチームでの開発経験はない．Java に関しては，基礎を一通り学んだ程度で知識は深くないが，他のオブジェクト指向プログラミング言語の経験はある．

2. アンケート回答

2.2 ソース No.2 についての回答理由

Protected と public の違いしかないが，protected と限定的な使い方をしているため，場合によってはまとめられないのではないかと考え，観点Aは no にした．

2.7 ソース No.7 についての回答理由

多少の違いはあるが，処理の内容的には同じだと思ったのすべての観点で Yes にした．

2.8 ソース No.8 についての回答理由

コード自体は類似していないが，ソートするという意味的な部分では一致しているので，観点Aと観点Cについては迷ったが，No にした．

2.9 ソース No.9 についての回答理由

明らかに類似してはいるが，型や数字の関連性がありそうなのでまとめにくいと考え，観点Aを No にした．

表 1 設問への回答

ソース No.	A	B	C	D	X
1	yes	yes	yes	yes	-
2	no	yes	yes	yes	-
3	yes	yes	yes	yes	-
4	yes	yes	yes	yes	-
5	yes	yes	yes	yes	-
6	yes	yes	yes	yes	-
7	yes	yes	yes	yes	-
8	no	no	no	no	-
9	no	yes	no	yes	-
10	yes	yes	yes	yes	-
11	no	no	no	no	-

2.10 ソース No.10 についての回答理由

観点 A については，具体的にどうすればいいのか思いつかないが，350 行以上一致しているという点で，まとめられるならまとめるべきだと考えた．

2.11 ソース No.11 についての回答理由

単純な作業であり，int と Boolean では使う目的も全然違うと考え，全て no にした．

3. 議　　論

将来的にはなんらかの形でクローンの良し悪しを判断したいと考えているが，クローンとするかどうかの判断や，クローンを生成した人の意図等，客観的な判断の難しい部分が多いため,「このクローンは良い（もしくは、悪い）」と明確な線引きをすることの難しさを痛感している．

過去の文献からも，同じような問題を抱えている印象を受けたので，やはり「曖昧さ」をいかに対処するかというところが今後の課題であると考えている．

† 奈良先端科学技術大学院大学 情報科学研究科
Graduate School of Information Science, NARA INSTITUTE of SCIENCE and TECHNOLOGY