

保守工程に悪影響を与える類似コード片とは

福島 義彦[†]

回答者は博士前期課程の学生であり、類似コード片が存在するファイルに基づいた類似コード片の可視化手法の研究をしている。可視化によってシステム内に横断的に存在する類似コード片の発見を支援する。プログラミング言語の知識はあるがソフトウェア開発の経験はほとんどない。Java に関しては、研究を通してある程度コードを読んだことはあるが、書いたことはなく、知識も深くはない。

1. はじめに

回答者は博士前期課程の学生であり、類似コード片が存在するファイルに基づいた類似コード片の可視化手法の研究をしている。可視化によってシステム内に横断的に存在する類似コード片の発見を支援する。

プログラミング言語の知識として C, C++, C#, Python 等を習得している。しかしそれらを用いたソフトウェア開発の経験はほとんどない。Java に関しては、研究を通してある程度コードを読んだことはあるが、書いたことはなく、知識も深くはない。今回の回答は他のプログラミング言語の知識をもとにして行った。

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 には大きな差はなく、他の冗長的なコードを残すリスクのほうが大きいので、まとめたほうが良いと考えた。

2.4 ソース No.4 についての回答理由

これも差分に比べて他の冗長的なコードの方が圧倒的に多いので、可読性を考えまとめたほうが良いと考えた。

2.5 ソース No.5 についての回答理由

コピー＆ペーストしているのは明らかだと思う。しかし、全体が短い上に扱っている変数のクラスが違うので、同じようなコードに見えても実際は違う処理をしている可能性があると考え、観点 A だけ no にした。

2.8 ソース No.8 についての回答理由

同一コード部分は短い上に定型的なので類似とは言

表 1 設問への回答

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

えないと考えた。

2.9 ソース No.9 についての回答理由

観点 A について迷いましたが、うまく 1 つのコードにまとめる方法が思いつかなかったので no にした。

2.10 ソース No.10 についての回答理由

No.9 と同じく観点 A について迷い、まとめる方法が思いつかなかった。しかし No.9 とは違い同一コード部分が膨大なのでまとめるべきだと考えた。このままだと冗長なコードが多くなり可読性が低くなると考えた。

2.11 ソース No.11 についての回答理由

コード片の構造はよく似ているが、よくある構造であり、長さも短いので類似しているとは言えないと考えた。

3. 議論

回答者は現在、類似コード片が存在するファイルに基づいて類似コード片の可視化を行い、優先的にリファクタリングすべきコード片の発見を支援する研究をしている。具体的には、コードクローンが存在するファイル群に着目して、同じファイル群に存在する

[†] 奈良先端科学技術大学院大学

Graduate School of Information Science, Nara Institute of Science and Technology

コードクローン群をグループ化することで、コードクローン同士の関係を可視化する手法を提案している。本手法では、保守を困難にするコードクローンの一つとして、全く別の機能を実現している複数のファイルにまたがって存在するコードクローンの抽出を目的とする。

上記の研究を進める上で、CCFinder 等の類似コード片の検出ツールを用いる場合、どのようなパラメータを使用すればいいのかで迷っている。何トークン以上一致すればコピー＆ペーストしたものだといえるのかの基準が欲しい。

また、グループ化された同じファイル群に存在する類似コード片の中で、各々のコードの類似性を判定したいと考えている。それ故、コードの類似性を構文で判断するのではなく、コードの機能で判断した類似性の指標等があれば有用であると考えている。
