

呼び出されるメソッドのコメントに基づく メソッド名推薦手法

2019SE073 吉川圭彦

指導教員：名倉正剛

1 はじめに

プログラムにおけるメソッドには、そのプログラムの処理内容を反映した名前を付けることが一般的である [1]. メソッドの名前に使われる単語が処理内容に関連しない場合、プログラムの理解を妨げてしまう. このような問題から、メソッド呼び出し関係に基づき呼び出されるメソッドの名前からメソッド名を予測する手法 [2] や、メソッドの名前に含まれる振舞いを表す単語とメソッドのプログラム構造との整合性を評価する手法 [3] が提案されている.

そこで本研究では、メソッドのコメントがそのメソッドの振舞いを表現しているという前提のもとで、呼び出されるメソッド群のコメントを利用してメソッド名を推薦する手法を提案する. 提案手法では、メソッドの呼び出し関係を解析し、メソッドの名前とそのメソッドが呼び出すメソッド群のコメントの特徴語との関係を学習する. そしてその関係を利用し、新たにメソッドを作成した際にそのメソッドが呼び出す他のメソッド群のコメントを利用して、適切な名前を推薦する.

2 対象とする課題

米内ら [2] はメソッド名の予測において呼び出されるメソッドに振舞いを表す名前が付けられていることを前提としており、峯久ら [3] はメソッドの名前とそのプログラム構造との整合性を評価するにあたって対象のメソッドの名前と振舞いが一致することを前提としている. しかし前述のようにメソッドの名前に使われる単語が処理内容に関連しない場合があることを考慮すると、このような手法を用いたメソッドの名前づけが適切に行われているという前提が成り立たない場合も考えられる.

3 提案手法

3.1 概要

そこで本研究では、メソッドのコメントがそのメソッドの振舞いを表現しているという前提のもとで、呼び出されるメソッド群のコメントを利用してメソッド名を推薦する手法を提案する.

提案手法は、学習用ソースコード群から各メソッドの呼び出し関係を解析する. そして呼び出しているメソッドのコメントを取得し、そのコメントの特徴語を抽出する. 抽出した各メソッドの特徴語群を合成してベクトル化を行う. 名前をつける対象のメソッドに対しても同様に実施し、メソッドのコメントに対する特徴語をベクトル化する. これらのベクトルをコサイン類似度で計算し、類似した特

徴語のベクトルに対応する既存のメソッドのベクトルをメソッド名を候補として出力する.

3.2 提案手法の流れ

図 1 に提案手法の概要図を示す. 図 1 の提案手法の概要図から、各手順を次に示す.

• 学習処理

手順 1 既存ソースコードの各メソッドに対して、メソッド呼び出し関係を解析し、あるメソッド名と呼び出しているメソッドに付けられた Javadoc コメントの説明文を抽出する.

手順 2 呼び出している各メソッドに記述されているコメントの特徴語群を抽出する.

手順 3 抽出した各メソッドのコメントの特徴語群を合成してベクトル化する.

• 推薦処理

推薦処理では、推薦したいメソッドにも学習処理と同様に 1 から 3 の手順を実施し、学習処理の学習データに推薦したいメソッド名とその特徴語群を追加する. その後、学習データからベクトル化を行う. そのあとで次の手順を実施する.

手順 4 既存のソースコード群から特徴語に対するベクトルと名前を付けるメソッドから特徴語に対するベクトルをコサイン類似度で比較し、類似する特徴語ベクトルをメソッド名の候補として推薦する.

4 評価・検証

本実験の目的として、推薦したいメソッドに対してその呼び出されるメソッドのコメント文の特徴的な単語から推薦したい呼び出し元のメソッドの名前を推薦出来ているかを検証する. 本実験では対象とするメソッドを 3.2 節で示した手順に沿って実施し、推薦したいメソッドから推薦結果のメソッドに同じような単語が使われているメソッド名を推薦できるかを検証する. 評価基準は次のようにして評価対象のメソッド名が推薦出来ているかを評価する.

- メソッドの名前に使われる単語が一部でも合っている：推薦可能
- 使用される単語が完全不一致、またはメソッドの役割として適していない単語：推薦不可能

また、今回の実験は推薦したいメソッドの推薦結果の過半数より上位のメソッドの特徴語ベクトルの内いずれかに推薦対象のメソッドを表す単語が含まれているものを判断基

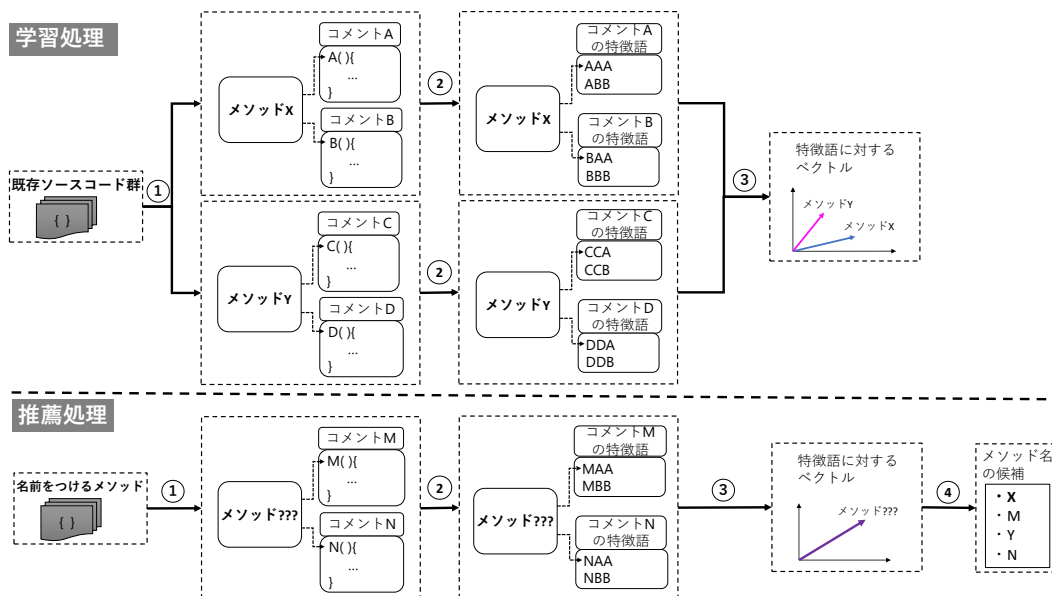


図 1: 提案手法の概要図

準とした。

評価に使用する推薦したいソースコード 1 は、外部クラスの四則演算を行うメソッドを利用して特定の計算を行うプログラムを用意した。

ソースコード 1: 推薦するソースコード例

```

1 public class Circumference {
2     public static void main(String[] args)
3     {
4         ...
5     }
6     public static double ??? (double width,
7         double height) {
8         return Calc.xyz(width, height);
9     }
10 }

```

提案手法によってメソッド名を推薦した結果として、出力結果 2 に示す。

出力結果 2: 推薦したメソッド例

```

1 [???]
2     around : 0.01950538530945778
3     area : 0.0025483479257673025
4     volume : -0.015642911195755005

```

出力結果 2 には推薦したいメソッド名 "???" が記述しており、以下には学習したデータの中に存在するメソッド名前とコサイン類似度で計算した数値が高い順に表示されている。出力結果 2 より、推薦結果としては around という名前が推薦された結果となった。実際のソースコード 1 を確認すると、呼び出しているメソッドは割り算を使

用するメソッドを利用しており、学習したソースコードの around メソッドを見てみると同じく割り算を使用するメソッドを呼び出していることから、評価基準より提案手法を用いたメソッド名推薦は可能だと言える。

5 おわりに

本研究では、メソッドに記述されているコメントがメソッドの振舞いに関する意味があると想定し、名前を付けるメソッドが呼び出しているメソッドのコメントに着目し、そのメソッドのコメントの特徴語を用いてメソッドの名前の推薦を行う手法を提案した。その結果、利用しているメソッドの識別子が不適切な場合でもそのメソッドに記述されているコメントを利用することで、そのメソッドの振舞いを表すメソッド名を推薦することができることを示した。今後の課題として、推薦処理で学習モデルに追加して学習を行う実装において、現在は学習処理で生成した学習データに対して追加して学習モデルを作成していたため、学習モデルに対して追加して学習する機能の改善が必要だと考えられる。

参考文献

- [1] D.Boswell, T.Foucher, 角 征典: リーダブルコード - より良いコードを書くためのシンプルで実践的なテクニック, オライリー・ジャパン, 東京, 2012.
- [2] 米内 裕史 1,a) 早瀬 康裕 2,b) 北川 博之 3,c): メソッド呼び出し関係に基づくメソッド名の予測. ソフトウェアエンジニアリングシンポジウム, pp.34-43, 2018.
- [3] 峯久 朋也, 阿萬 裕久, 川原 稔: 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. ソフトウェア工学の基礎 28, pp.81-90, 2021.