

卒 業 論 文

プログラミング演習における模範解答を用いた
テストケース評価基準の自動生成

English Title

指 導 教 員 中村 正樹 准教授

富山県立大学工学部 電子・情報工学科

学 籍 番 号 : 1515015

氏 名 尾崎 裕樹

提 出 年 月 2018 年 2 月

目次

| | | |
|---------|---------------------------|----|
| 第 1 章 | はじめに | 1 |
| 1.1 | 背景 | 1 |
| 1.2 | 目的 | 2 |
| 1.3 | 論文の構成 | 3 |
| 第 2 章 | プログラミング演習におけるテストケース評価システム | 4 |
| 2.1 | テストケース評価システム | 4 |
| 2.2 | 入力データの構造の定義 | 5 |
| 2.3 | 関数定義 | 6 |
| 2.4 | 評価基準の具体例 | 6 |
| 2.5 | システムの評価 | 7 |
| 2.6 | システムの課題 | 7 |
| 第 3 章 | テストケース評価基準の自動生成 | 8 |
| 3.1 | 背景 | 8 |
| 第 4 章 | 検証 | 9 |
| 4.1 | 背景 | 9 |
| 第 5 章 | おわりに | 10 |
| 謝 辞 | | 11 |
| 参 考 文 献 | | 12 |

第1章 はじめに

本研究では、プログラミング教育において、教員が学生の作成したプログラムに対して適切な評価を行う方法を考え、その際に利用できるシステムを提案する。

1.1 背景

プログラミング演習において、学生は最初に教科書のソースコードを書き写すことから始める。教科書のコードが正確に書き写しているかの確認をするには、教員が正しいソースコードを用意し学生の書き写したソースコードと比較すればよい。しかし、ただ教科書のソースコードを書き写すだけではプログラムを作成できるようにはならないので、プログラミング演習では提示された仕様を満たすプログラムを作成することを学ぶ必要がある。仕様を見てプログラムを作成した場合に教員と学生が同じ実装をするとは限らないため、教員のソースコードと学生のソースコードを単純に比較しただけでは正確な評価はできない。そこで、テストケースを設計してプログラムのテストを行うことによって学生が作成したプログラムを評価する。このテストケースを学生自身が適切に設計できるようになるためには、設計したテストケースを評価できなければならない。しかし、テストケースは境界値以外の値は答えが単一ではない。したがって、学生が設計したテストケースと教員が設計したテストケースが合致しなかったとしても、学生のテストケースが誤りであるとは限らない。つまり、テストケースを評価する場合も、教員のテストケースと学生のテストケースを単純に比較しただけでは正確な評価ができないため、テストケースの評価方法を考える必要がある。

例えば文献 [1] では、テスト駆動開発に基づきプログラミング演習を行うことによってコーディングだけでなく、ソフトウェアテストの学習が行われる。評価対象は学生のプログラムとテストケースで、教員の作成したプログラムとテストケースを用いて図 1.1 のようにトリプルチェックを行うことによって評価される。

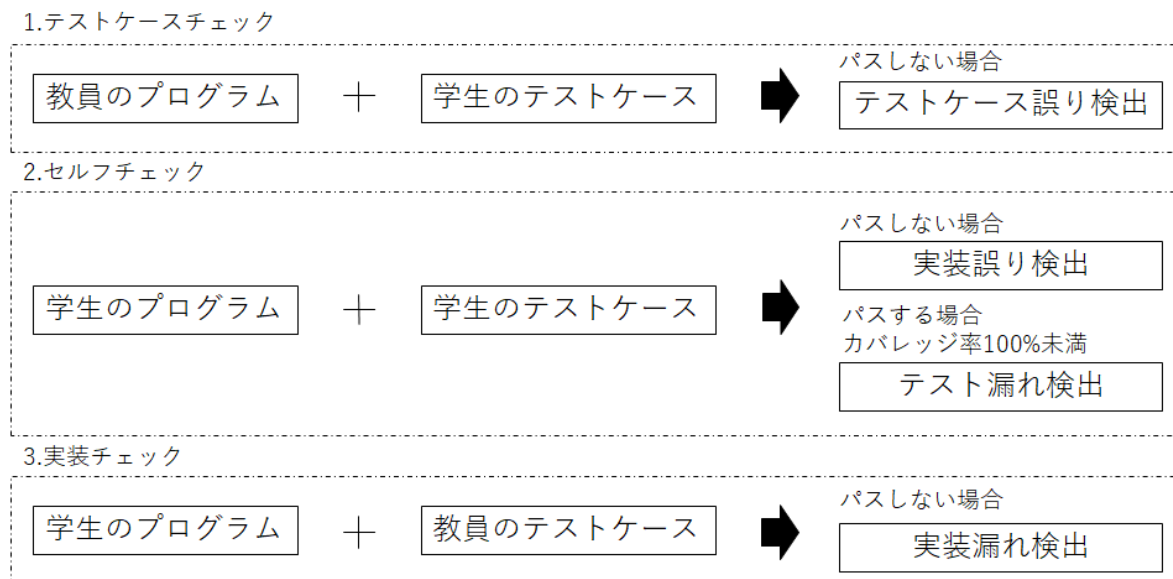


図 1.1 トリプルチェックによるコード判定原理

テストケースチェックによって、仕様要求を満たしていないテストケースが検出され、学生のテストケースの間違いが早期に指摘される。なお、仕様要求とは、プログラムにおける関数あるいはサブルーチンの動作が規定されたものであり、戻り値や引数などが定められたものである。教員のプログラムとテストケースはこの仕様要求が満たされるとする。次のセルフチェックでは、テストの実行だけでなくコードカバレッジの測定が行われる。テストにパスしない場合は、学生のプログラムが正しく実装されていないことが分かり、テストにパスしたが、カバレッジが 100% でない場合は、テスト漏れが検出され学生に指摘される。最後の実装チェックでは、学生のプログラムが仕様要求の要求をすべて満たしていないことが検知され、次のテスト駆動開発サイクルを始めるよう学生に通知される。実装チェックをパスすることによって、学生が作成したプログラムとテストケースが仕様要求の要求をすべて満たしたことが確認される。

しかし、この手法では学生が仕様通りのプログラムを作成した後に仕様通りのテストケースであるかの確認が行われるため、学生が自身のテストケースを用いてプログラムの確認を行うことはできない。

1.2 目的

本研究では、学生がテストケースを設計できる

1.3 論文の構成

．．．

第2章 プログラミング演習におけるテストケース評価システム

本章では文献 [2] で提案されているテストケース評価システムについて紹介する。テストケース評価システムでは、学生自身が適切なテストケースを設計できるようになるために、学生が作成したテストケースを評価してアドバイスを行うシステムが提案されている。このシステムでは、教員が演習問題毎にテストケースの評価基準とテストケースが不足していた場合に表示するアドバイスを与える必要がある。演習問題毎にテストしなければならない値や入力データの数が異なるため、評価基準は演習問題を分析した上で、入力データの構造を定義してから記述される。

2.1 テストケース評価システム

テストケース評価システムでは、教員があらかじめテストケース評価基準を作成し、学生が設計したテストケースが評価基準をどの程度パスできるかを判定することによって、テストケースの評価が行われる。評価基準をパスしない場合にアドバイスを表示することで、不足しているテストケースを学生に知らせる。テストケース評価システムにおいて、評価基準は次のようにタブ区切り形式でテキストとして記述される。

評価基準の記述形式

評価基準番号 TAB 判定条件 TAB アドバイス

「判定条件」はテストケースにおける入力データが満たすべき条件で、その条件を満たすテストケースがなかったときに、学習者に「アドバイス」が表示される。同値クラスを評価するための基準が複数ある場合などは、同じ「評価基準番号」を指定することで、評価基準がグループ化できる。例えば、年齢（整数値）を入力して、20 歳以上は成年、20 歳未満は未成年と表示する課題の評価基準は、入力データを変数 `age` として、成年、未成年、エラーの場合について次のように記述される。

評価基準の記述例 1

```
1 TAB age==20 TAB 成年の最低年齢
1 TAB age>=20 TAB 成年の場合
2 TAB age==19 TAB 未成年の最高年齢
2 TAB age==0 TAB 0 歳
2 TAB age>=0 && age<20 TAB 未成年の場合
3 TAB age<0 TAB 年齢がマイナスの場合
```

学生が作成したテストケースの網羅率を、次の式で定義される評価基準のテストカバレッジにより提示する。

評価基準のテストカバレッジ (%)

$$\frac{\text{学習者のテストケースがパスした評価基準グループ数}}{\text{教員が記述した評価基準グループ数}} \times 100$$

判定条件の記述において、課題毎に入力データの型や個数が異なるため、テストケースの入力のデータ構造を定義し、そこで定義された変数を用いて判定条件が記述される。

2.2 入力のデータ構造の定義

データ構造は型と変数と出現回数から成る。入力のデータ構造の記述法はバックス・ナウア記法 (BNF) に類した形式で設計された。入力のデータ構造の定義例を以下に示す。なお、* は 0 回以上の繰り返しを表し、pint は正の整数、uint は 0 と正の整数、udouble は 0 と正の実数の型である。

データ構造の定義例

```
(A) データ数 (n) を入力し、その個数分身体データを入力する場合
(pint n)
(pint id, string name, uint age, udouble height, udouble weight){n}
(B) ファイルの終端まで分身体データを入力する場合
(pint id, string name, uint age, udouble height, udouble weight){*}
```

2.3 関数定義

判定条件の記述に関数を用いることができる。複数の身体データを読み込んで、身長 of 最大値を表示する課題の場合、最大値が複数存在するテストの評価基準は、引数のリスト of 最大値の個数を返す関数 `countmax()` を定義して、次のように記述される。なお、実引数 `height` は全身長データのリストである。関数の評価基準で用いる場合、教員は関数定義も与える必要がある。

関数を用いた評価基準の記述例

```
1 TAB countmax(height)>=2 TAB 身長の最大値が複数の場合
```

Listing 2.1 `countmax` 関数

```
function countmax($list){  
  
    $lmax = max($list);  
  
    return count(array_keys($list, $lmax));  
  
}
```

2.4 評価基準の具体例

入力が比較的複雑になる課題として、ボウリングの得点計算が例示されている。

ボウリングの点の計算の課題

ボウリングは、1 フレーム目から 10 フレーム目までの倒れたピンの数によって得点が決まる。フレームごとに倒したピンの数を読み込み、各フレームごとの得点を計算して表示すること。なお、ガーターとミスは 0 と表示するものとし、ダブルがあった場合はメッセージを表示する。

ボウリングのデータ構造

```
(uint thr1, uint thr2){9}  
(uint last1, uint last2, uint last3)
```


ボウリングの評価基準

```
1 TAB thr1 == 10 TAB ストライクのスコア計算
2 TAB thr1 != 10 && thr1+thr2 == 10 TAB スペアのスコア計算
3 TAB last1 == 10 TAB ストライクの通常加算
3 TAB last1 != 10 && last1+last2 == 10 TAB スペアの通常加算
4 TAB double_strike(thr1) || all_strike([last1, last2]) TAB ダブル
4 TAB last_strike(thr1) == 1 && last1 == 10 TAB 最終 F をまたぐダブル
```

ストライクとスペアだった場合のスコア計算が正しいかをテストする必要がある（評価基準番号 1, 2）。最終フレームではストライク、スペアでも通常の加算になるので、それらのテストも必要である（評価基準番号 3）。評価基準番号 4 はダブルについてのテストである。ダブルがあれば true を返す関数 `double_strike()`、リストがすべて 10 ならば true を返す関数 `all_strike()`、リストの最後の要素から 10 が何回連続しているかを返す関数 `last_strike()` が定義されている。

2.5 システムの評価

参考書の例題のテストケースのデータ構造が評価システムに則した形式で記述できるか調査が行われ、実行時に入力データの構造が変化する共用体などの問題以外ではデータ構造を記述できることが確認された。

また、評価システムにより、学生がテストケースを作成できるようになるのか確認をするために、プログラミングを学習済みの学生 7 人が評価システムを使ってボウリングの課題のテストケースを作成した。その結果最終的に 6 人の学生がカバレッジ 100% を達成したことが確認された。

2.6 システムの課題

テストケース評価システムでは、教員が問題文を分析した上で、評価基準を記述しているので、システムを運用する上で教員の負担が大きくなってしまふことが考えられる。そこで、本研究は教員の負担を減らすために、評価基準を自動で生成する方法を提案する。

第3章 テストケース評価基準の自動生成

3.1 背景

第4章 検証

4.1 背景

第5章 おわりに

謝 辭

参 考 文 献

- (1) 吉田英輔，角川裕次： テスト駆動開発に基づくプログラミング学習支援システム : 初心者開発者のためのセルフトレーニングアーキテクチャ，信学技報. SS，ソフトウェアサイエンス，Vol. 105，No.331(2005)，pp.27-32
- (2) 蜂巢吉成，小林悟，吉田敦，阿草清磁： プログラミング演習におけるテストケース評価システム，コンピュータソフトウェア第 34 巻第 4 号，2017，pp.54-60