

プログラム構造比較システム

5X-3

福田 勇一
芝浦工業大学

1. はじめに

本システムはプログラミング言語教育において、学生から提出される、多数のプログラムを分類し、詳細に比較することを目的としている。この報告では、標準化して分類したプログラム・グループ間のアルゴリズムの相違を明確にするための「プログラム構造表の自動作成」および、「個人間およびグループ間の類似度計算」を中心に記述している。

2. 処理の概要

処理の概要を以下に示しておく。

- (1) 学生は「レポート提出プログラム」を使用して、教員に課題プログラムをファイル転送する。もしも、提出済のプログラムを訂正したい場合は、再度、ファイル転送を行えば、上書きが行われる。また、提出期限を過ぎると、ファイル転送は行えなくなる。
- (2) エラープログラムの検出
提出された多数のプログラムを、一括実行して、その結果によりエラープログラムを検出し、削除する。
- (3) プログラムの文を、対応する数値に変換し以後はこの数値列に対して処理を行う。
- (4) 数値列を標準化(後述)し、数値列の大きさ(=プログラムの行数)の等しいものについて比較を行う。この結果、等しいものについては、プログラム構造が似ているものとして、グループ化し、グループ数と人数を出力する。
- (5) 本システムの使用時には、目標グループ数を指定することができる。この場合標準化を段階的に行なって目的グループ数に最も近くなった時、処理を終り、結果を出力する。
- (6) 分類された各グループ内で、個人間の類似度を計算して出力する。これにより、「学生のコピー状況」を見ることができる。
- (7) 類別されたグループについて、プログラム

構造表を自動作成し、また、グループ間の類似度を計算して出力する。これにより、「アルゴリズムの多様性」を見ることができる。また、「本システムの欠点・改良すべき点」を容易に把握できる。

3. プログラムの標準化

プログラムの構造を比較するため、以下の順序により、標準化を行っている。この標準化の順序と組合せは、自由に容易に変更することができる。

- (1) コメント行を削除する。
- (2) CONTINUE文を削除する。
- (3) END文の前のSTOP文とRETURN文を削除する。
- (4) DATA文とFORMAT文を定位置へ移動する。
- (5) 宣言文の順序を整理する。
- (6) 連続する同じ文を1つの文とみなす。
- (7) INTEGER文、REAL文を削除する。
- (8) CHARACTER文を削除する。
- (9) FORMAT文を削除する。
- (10) WRITE文を削除する。
- (11) 連続する同じ文を1つの文とみなす。
- (12) 宣言文を削除する。(ただし、PARAMETER文を除く)
- (13) PARAMETER文を削除する。
- (14) READ文を削除する。
- (15) DATA文を削除する。
- (16) 代入文を削除する。
- (17) 連続する同じ文を1つの文とみなす。
- (18) 連続する同じ数値列を1つとみなす。

4. プログラム構造表の自動作成

標準化を行い、グループ分けを行い、各グループのプログラム構造に、どのような相違があるかを考察するために、次ページに示すプログラム構造表が有用である。この構造表を手作業で作成するのは、大変手間がかかるので自動作成のプログラムを開発した。

これは、プログラムの文を数値で置き換えた複数の数値列において、同じ数値が、同じ列に

PG01	15	16	41	42	24	42	24	42	22	52	
PG02	15	16	41	42	24	42			22	52	
PG03	15	16	41		24	42			22	52	
PG04	15	16	41		24	42				52	
PG05	15	16		24	42	24	42	24	42	22	52
PG06	15	16		24	42	24	42			22	52

プログラム構造表

並ぶように、並べ替えればよい。このプログラムの概要を以下に示す。

- (1) 数値列の最も長いものをコピーして、代表数値列とする。
- (2) 代表文字列と対象文字列を、先頭から順次比較し、同じ場合はそのまま、異なる場合は、①または②を行う。
 - ① 対象文字列の b を右にシフトして、0 を入れる。

代表文字列	..	a
PG01	..	a
PG02	..	0	b.....
.			
.			
pgxx	..	a

- ② 代表文字列の a を右にシフトして、b を入れる。

代表文字列	..	b	a.....
PG01	..	0	a.....
PG02	..	b
.			
.			
pgxx	..	a

①、②のどちらを行うかは、残された文字列を調べて、一致する数値の多い方を選ぶ。

5. プログラムの類似度計算

プログラムの類似度については、各グループ間の類似度と個人間の類似度について、計算し表示している。

類似度の計算対象とするプログラムの標準化レベルは任意に選べば良いが、現在は以下のように行うことが多い。

- (1) グループ間の類似度については、アルゴリズムの種類と特色を見るため、標準化後のプログラムについて計算している。
- (2) 個人間の類似度については、学生のプログラムコピーの度合いを見るため標準化後の構造が同じ学生について、コメント行を除いた段階で計算している。また、プログラムを点検する場合は、標準化以前のものを表示するよう注意している。

つぎに、類似度の計算例を示しておく。この例は、レベル(6)まで標準化した、簡単なプログラムのグループ間の類似度計算である。

プログラム 1	15-16-41-42-24-42-22-52
プログラム 2	15-16-24-42-24-42-22-52

同じ値が同じ列に並ぶように整える。
(プログラム構造比較)

プログラム 1	15-16-41- 0-42-24-42-22-52
プログラム 2	15-16- 0-24-42-24-42-22-52

ここで、数値が等しい列と、等しくない列をカウントし、それぞれ、E、Nとする。ここでは、E=7、N=2となる。

類似度：C(P1、P2)は、つぎの式で計算している。

$$C(P1, P2) = E / (E + N) \times 100$$

この例では、類似度は77.8%となる。
 $7 / (7 + 2) \times 100 = 77.8$

6. おわりに

1989年より、本システムを試作・使用してきた。この間、教育・研究用システムのメーカーが変わり、また、実習用の機器が、スーパーコンピュータから、エンジニアリング・ワークステーションに変更になるなど、大きな変化があった。このように教育環境の急変する中で開発が遅れた部分もあったが、漸く本格的に使用できる時期がきたと考えている。今後については、提出プログラムの傾向を見ながら、標準化機能の追加を行って完成度を高めたい。

[参考文献]

- ①福田「プログラム比較システムⅡ」、1991年電子情報通信学会春季全国大会
- ②福田「プログラム比較システムⅢ」、情報処理学会第43回(平成3年後期)全国大会