

5X-2 CPUとアセンブラ授業のための事例に基づくプログラム 評価支援システム – 事例ベース構築法の改良とその評価 –

荒井 正之

渡辺 博芳

武井 恵雄

帝京大学理工学部

1. はじめに

我々は、CASLを対象として、提示した課題に対して学生が作成したプログラムの評価を支援するシステムを実現し、1999年度の授業で実際に使用することで、教員の評価作業の軽減に大きな効果があることを明らかにした[1]。また、このシステムのための事例ベース構築に関して、プログラムリストの一般化は有効であるが、評価事例の事例ベース化においては、CASL文法のまま収納したほうが良いことを指摘し、それらを効果的に管理するための階層的なインデックス構成法を提案した[2]。

本稿では、提案した事例ベース構築法の概要を述べた後、授業での使用に基づく評価結果を報告する。

2. 事例ベース構築法の改良点

(1) 以前の手法の問題点：本システムは、プログラムの動作の評価と、事例に基づく実現方法の評価を行うことで、教員の評価作業を支援する[1]。授業での実用の結果、非常に有効であったが、事例ベース構築において、以下のような問題が顕在化した。

・事例がフラットに配置されており、事例照合処理において、評価対象のプログラムを事例ベース中の全ての事例と照合するので、効率が悪い。

・事例中のプログラムリストを一般化表現[1]で記述することで、1つの事例で複数のバリエーションに対処できるようにしたが、一方で、本システムのユーザとしての教員が一般化表現を習得する必要がある。

(2) 改良版事例ベース構築法の概要：事例中のプログラムリストは、一般化表現とせずに、CASLの文法に従った表現をとる。事例に含まれるプログラムリストを一般化しない場合、そのままでは1つの事例の適用範囲が狭くなってしまいが、プログラムリストの一般化レベルに基づいて、事例へのインデックスを

表 1: 問題の概要

問題名	問題の説明	対象
P1	2つの値の大きい方を求める	A, B
P2	整数データの符号を調べる	A
P3	N個の整数データの加算	A, B
P4	N個の整数データのうち、0以上と0未満のデータ数を求める	A
P5	2つのデータの論理和をとった結果において1のビット数を求める	A
P6	文字列データから英大文字のみを取り出し、スタックを使って逆順	A
P7	整数データの符号を反転	B
P8	N個の文字データのうち、指定した文字の数を求める	B

階層化することで対応する。事例ベースは次の4つの階層から成る[2]。

第1階層 インデックスのルート。

第2階層 プログラムリストを可能な限り一般化したノード(最大一般化)。

第3階層 (a)命令の種類、(b)命令の順序、(c)冗長命令の有無の3つの観点ごとに、3種類のノードのグループを形成。

第4階層 典型的な事例を配置。ここでは、プログラム中のラベル名、レジスタ番号が異なる場合も同一プログラムと見なす。

3. 授業での使用状況

実現したシステムは、本学2年生を対象とした演習授業の2つのクラス(仮にA組、B組と呼ぶ)で、実際に使用した。授業を履修した学生の人数は、A組が83人、B組が78人である。本システムを使用した授業で、これまでに出题した問題を表1に示す。

これらの問題に対する総提出数は、1906プログラムで、うち、991プログラムが動作が正しくないため、不受理、544プログラムが確信度Surelyでシステムが評価、135プログラムが確信度Probablyでシステムが評価、236プログラムはシステムでは評価不可であっ

A Case-Based Evaluating Assistant System for the Students' Programs of the CPU and Assembler Course – Improvement of Case-base Construction and Its Evaluation

School of Science and Engineering, Teikyo University
Masayuki Arai, Hiroyoshi Watanabe and Shigeo Takei

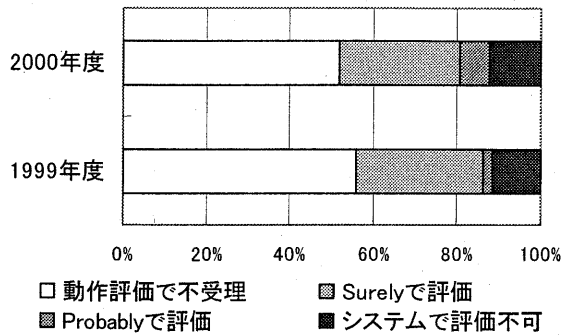


図 1: 本システム使用によるプログラム評価結果

た。1999年度は、7問[1]について本システムを使用した。総提出数1349、動作評価で不受理757、Surelyで評価406、Probablyで評価31、システムでは評価不可155であった。これらの結果を、提出されたプログラムに対する割合で表したのが、図1である。

動作評価で不受理のケースと、システムがSurelyで評価したケースは教員が評価する必要がないとすると、2000年度は約81%の軽減量があったと見ることができる。提出されたプログラムの約20%のみが教員の評価対象となることから、本システムがきわめて有効であることがわかる。

同様に考えると、1999年度は、約86%の軽減量があったと見ることができる。2000年度よりも1999年度の方が軽減量が多い理由の1つは、1999年度の方が動作の正しくないプログラムが提出された割合が多いことにある。本システムを使用した場合、教員に直接プログラムを提出するのに比較して、学生は“気軽に”提出するため、自分のプログラムの動作を十分に検証しないケースが少なくない。課題提示時に、プログラムの動作を十分に確認してから提出するように学生に注意すると、動作の正しくないプログラムの提出は減少した。ただし、プログラムの動作の要求仕様が厳しい問題では、動作の正しくないプログラムの提出は多くなる傾向があった。

4. 事例ベース構築法改良の評価

図2は、動作の正しくないプログラムを除き、事例ベース推論の対象となったプログラムについてまとめた結果である。事例ベース活用率(システムが評価を行えた割合)は、2000年度が74.2%、1999年度が73.8%であり、ほぼ同等である。すなわち、ユーザとしての教員が一般化表現を知らなくとも良いように事例ベース構築法を改良しても、事例ベース活用率が低下しなかった。従って、提案した事例ベース構築法が有効であることが明らかになった。

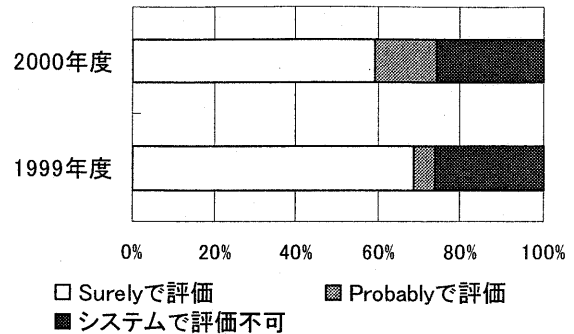


図 2: 事例ベース推論によるプログラム評価結果

図2において、システムがSurelyで評価したケースの割合が、2000年度よりも1999年度の方が大きいのは、1999年度は初期事例としてプログラムリストを一般化表現に変換した事例を用いていたためと思われる。一方、システムがProbablyで評価したケースの割合が2000年度の方が大きいのは、事例ベース構築法の改良により、事例ベースに付加した階層化インデックスが有効に機能したことを示している。つまり、一般化した事例を用いないことが、提案した事例ベース構築で補完されたことが分かる。

合否の評価について、システムの評価結果と教員の評価結果が一致した割合(判定制度)は、2000年度は98.1%、1999年度は99.4%であった。システムが確信度Surelyで評価した場合は、評価対象プログラムと事例のプログラムはほとんど同一のため、判定精度が100%になることが予想されるが、2000年度は99.2%、1999年度は100%であった。2000年度は、確信度Surelyの場合も、4ケースについてシステムと教員の判定が一致しなかったが、その原因は教員のケアレスミスにより、合否を誤って判定したプログラムが事例として登録されたためであった。これらから、システムの判定精度は十分に高いと言える。

5. おわりに

CASLを対象としたプログラム評価支援システムにおいて、事例ベース構築法を改良し、実際の授業で評価を行った結果、提案手法の有効性が確認された。今後、CASL IIに対応したプログラム評価支援システムを実現する予定である。

参考文献

- [1] 渡辺博芳, 荒井正之, 武井恵雄: 事例に基づく初等アセンブラプログラミング評価支援システム, 情処論, Vol.42, No.1, 2001.
- [2] 渡辺博芳, 荒井正之, 武井恵雄: 初等アセンブラプログラミングを対象とした事例に基づくプログラム評価のための事例ベース構築法, 情処学会第61回全大, 3P-6, 2000.