

プログラミング演習支援システム CAPES における受講者答案提出履歴を用いた要指導受講者推定システムの実現と評価

奥田 公将[†] 立岩佑一郎[†] 山本 大介[†] 高橋 直久[†]

[†] 名古屋工業大学大学院 〒466-8555 愛知県名古屋市昭和区御器所町

E-mail: [†]okuda@moss.elcom.nitech.ac.jp, ^{††}{tateiwa,yamamoto.daisuke,naohisa}@nitech.ac.jp

あらまし 我々は、プログラミング演習支援システム CAPES を提案し、本学の演習に適用している。CAPES を使用したプログラミング演習では、指導者は、CAPES の受講者の履歴データなどを用いて指導の必要な受講者を予測し、その受講者に対して指導を行う。しかし受講者数が多い場合や、指導者が演習問題に関する予備知識が少ない場合などでは、適切な指導の必要な受講者を見つけるのが難しい。この問題の解決のため、本稿では演習履歴を用いた要指導受講者推定システムを提案する。また、提案システムのプロトタイプを開発し、実際の演習のデータを用いて行った予備実験と評価実験の結果について述べる。

キーワード e-ラーニング, プログラミング演習, 演習履歴, 指導者補助

Implementation and Evaluation of a System for Inference of Students in Need of Educational Assistance with History of the Answer on CAPES

Hiroyuki OKUDA[†], Yuichiro TATEIWA[†], Daisuke YAMAMOTO[†], and Naohisa TAKAHASHI[†]

[†] Department of Computer Science and Engineering, Graduate School of Engineering, Nagoya Institute of Technology Gokiso-cho, Showa-ku, Nagoya-shi, Aichi, 466-8555 Japan

E-mail: [†]okuda@moss.elcom.nitech.ac.jp, ^{††}{tateiwa,yamamoto.daisuke,naohisa}@nitech.ac.jp

Abstract We proposed Computer-Aided Programming Exercise System(CAPES), and CAPES is used for exercise of Nagoya Institute of Technology. In the exercise using CAPES, instructor infer students in need of educational assistance by using history of the answer, and instruct them. However, it is difficult to find them appropriately when there are many students, and there is little prior knowledge of the instructor. To solve this problem, we propose a system for inference of students in needs of educational assistance with history of the answer on CAPES in this paper. And we develop prototype system, and describe a result of preliminary experiment and evaluation experiment by using exercise data.

Key words e-learning, programming exercise, exercise history, support instructor

1. はじめに

近年、パソコンやコンピュータネットワークなどを利用して教育を行う e-learning が発達した。e-learning により、遠隔地への教育の提供、コンピュータならではの教材の利用が可能となった。また、企業の社内研修で用いられているほか、英会話学校などがインターネットを通じて教育サービスを提供している例などがある。

我々は、プログラミング演習支援システム CAPES [1] を開発し、本校のプログラミング演習に適用している。図 1 は CAPES を使用した演習の流れの図である。まず始めに、演習の開始以前に指導者が CAPES に対して出題するプログラミング課題

を設定する。また、課題の集まりを単位として、単位ごとに答案提出開始時刻、答案提出締切時刻を設定する。受講者は、単位ごとに行われるプログラミング演習の授業時に、講義室で CAPES から出題される単元の課題を解き、答案プログラムを作成し、CAPES に対して答案プログラムを提出する。CAPES は提出された答案プログラムに対して自動で正誤判定を行い、結果を受講者に対して提示する。受講者は単元の答案提出締切時刻までに課題を全て正解するまで答案プログラムの提出を繰り返す。指導者は、演習時には、CAPES の受講者の演習履歴データを用いて、指導の必要な受講者に対して指導する。しかし指導者が受講者の演習履歴データを用いて指導の必要な受講者を予測することは困難である。この問題の原因として、以下

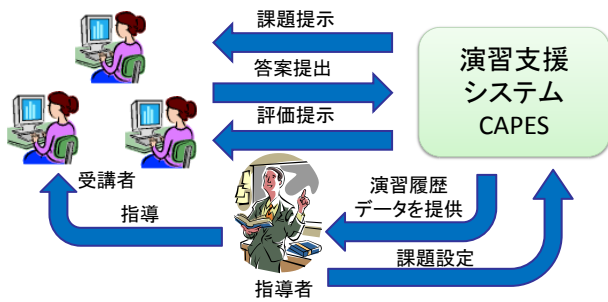


図1 CAPESを使用した演習の流れ

の2つがある。

- 受講者の人数が多く、さらにCAPESの受講者の演習履歴データは、各受講者の課題の答案プログラムの正誤判定、答案プログラムの提出時刻、答案提出回数など様々なデータが存在する。このため指導者が全受講者の演習履歴データを用いて予測することは困難となっている。

- CAPESが提示する受講者の演習履歴データの中からどのデータを用いれば指導の必要な受講者の予測を適切に行うことが出来るのかわからない。

本稿では、この問題を解決するために、受講者答案提出履歴を用いた要指導受講者推定システムを提案する。提案システムは以下の特徴を持つ。

- プログラミング演習支援システムCAPESのデータベースから、答案提出、正誤判定結果などの受講者の答案作成状況を表すデータ(答案作成状況データ)を抽出し、前年度以前の演習履歴(過去データ)を用いた教師あり学習により、答案作成状況データに従って受講者を分類し、指導の必要な受講者と判定する決定木を作成する機能(要指導受講者決定木作成機能)を実現する。これにより、演習実施時に、過去データに基づいて、受講者の答案作成状況から、指導の必要な受講者を推定して選択的に指導できるようになる。

- 過去データに決定木を適用して、決定木の葉ごとに、その葉に分類された受講者が指導の必要な受講者となる割合(要指導率)を求める機能(要指導率設定機能)を実現する。これにより、演習実施時に、要指導率を用いて、指導の必要な受講者になる可能性の高い順に受講者を並べ、指導の順番を定めることが可能になる。

2. 関連研究

CAPESを用いた演習で要指導受講者を見つける関連研究として、文献[2]がある。文献[2]では、CAPESを使用した演習において指導の必要な受講者を見つけるために、受講者の着席情報と受講者の進捗状況を提示する指導者補助システムのための進捗状況提示機能を提案している。この研究では、受講者を進捗状況を基に分類して、その結果を用いて受講者に(グループ名、度合)からなる属性を付与する。グループとしては「進捗が遅れている」、「課題に行き詰っている」、「頻繁に答案を提出している」があり、指導者により自由にカスタマイズできる。

度合は、課題の提出回数、課題の正解数、答案提出からの経過時間などを用いて数値で表される。グループと度合を用いた進捗状況提示機能により、要指導受講者をグループの度合から色分けして表示することで、指導者が指導の必要な受講者を見つける補助を行なっている。

本研究は、発見する指導の必要な受講者が違う。文献[2]では、現在の進捗状況が悪い受講者を指導の必要な受講者としている。一方、本研究では、要指導受講者を将来的に進捗状況が悪くなると推定される受講者を指導の必要な受講者としている。このようにすることで、文献[2]で発見される指導の必要な受講者よりも指導が必要とされる受講者の発見を目指す。

3. 提案システムの概要

本稿で用いる用語の定義を表1、用語に関する図を図2に示す。

提案システムは、現在時刻である推定要求時刻 t_1 までのデータを用いて、答案提出締め切り近くに設定した認定時刻 t_2 での進捗状況が要指導受講者判定割合 r 未満である要指導受講者を推定する。そのために、まず、過去データの各単元の演習授業開始時刻から $|t_2|$ 経過時の進捗状況のデータより、 $|t_2|$ 経過時の要指導受講者であることを示すデータ(以下要指導受講者判定属性)を求める。続いて、過去データの各単元の演習授業開始時刻から $|t_1|$ 経過時のデータと、要指導受講者判定属性に、決定木アルゴリズムを利用して、推定要求時刻 t_1 までのデータより、要指導受講者判定属性のデータを推定する決定木(以下要指導受講者決定木)を作成する。そして、要指導受講者決定木から、決定木の葉ごとに、その葉に分類された受講者が要指導受講者となる割合である要指導率を求める。

提案システムの構成図を図3に示す。提案システムは、要指導受講者決定木作成機能、要指導率設定機能からなる。指導者が、推定要求時刻 t_1 、認定時刻 t_2 、要指導受講者判定割合 r 、指導者所属年度 Y_1 、クラス C_1 、学習年度 $Y_2(Y_1 > Y_2)$ 、クラス C_2 、を設定すると、提案システムが要指導受講者を推定する。設定について簡単な具体例を挙げると、指導者が演習授業開始から90分後に演習中の課題が半分未満しか解けていない受講者を要指導受講者として推定をする場合、 $t_2=90$ 分、 $r=1/2$ 、と設定する。提案システムは、指導者から取得したデータを用いて、学習用答案履歴 $SAH(Y_2, C_2, t_1, P_1) \cdots SAH(Y_2, C_2, t_1, P_n(n \text{ は } Y_2, C_2 \text{ の単元数}))$ 、 $SAH(Y_2, C_2, t_2, P_1) \cdots SAH(Y_2, C_2, t_2, P_n)$ 、テスト用答案履歴 $TAH(Y_1, C_1, t_1, P_m(m \text{ は現在行なっている演習の単元番号}))$ をCAPESのデータベースより取得し、要指導受講者決定木作成機能、要指導率設定機能により指導者の所属コースの各受講者の要指導率を求め指導者に提示する。

4. 提案システムの実現法

4.1 答案作成状況データの構成

答案作成状況データは、要指導受講者の推定を行うための式を作成するために用いる学習用答案作成状況データ、受講者が要指導であるかの推定を行うために用いるテスト用答案作成状況

表 1 用語定義

用語	記号	意味
単元	P	学習内容の区分としてのまとまり，単元ごとに演習を行う．
演習時刻	t	プログラミング演習の授業を開始する時刻を 0 とした相対時刻．
答案提出締切時刻	td	プログラミング演習の答案提出締切の演習時刻．
演習時間	t	プログラミング演習の授業を開始する時刻を 0 とした相対時間．
演習授業前時間	ts	プログラミング演習の開始時刻からプログラミング演習の授業を開始する時刻までの時間．
演習授業時間	[0,te]	時間割に指定された演習のための授業時間．通常，受講者は演習室に集合して対面指導を受ける． $te=90(\text{分})$ ．
答案イベント	e	受講者が演習時間に行う答案提出，および，演習システムが行う答案評価結果の通知．
答案イベント記録	R(e)	答案イベント e の内容と e が発生した時刻の対．演習システムが e を観測して記録．
推定要求時刻	t1	指導者が提案システムに対して要指導受講者の推定を要求した時刻． $0 \leq t1 \leq td$ ．
認定時刻	t2	提案システムが要指導受講者を認定する演習時刻．提案システムでは， $0 \leq t1 \leq t2 \leq td$ ．
指導者所属年度	Y1	指導者の所属する年度．
指導者所属クラス	C1	指導者の所属するクラス．
学習年度	Y2	要指導受講者を推定するための学習データとして使用する年度． $Y1 > Y2$
学習クラス	C2	要指導受講者を推定するための学習データとして使用するクラス．
学習用答案履歴	SAH(Y2,C2,t,P _n)	学習年度 Y2，学習クラス C2 の単元番号 n の演習の演習授業開始時刻から， t 経過時までの R(e) より，学習用に選択した R(e) の集合．
テスト用答案履歴	TAH(Y1,C1,t1,P _m)	指導者所属年度 Y1，指導者所属クラス C1 の演習中である単元番号 m の演習の演習授業開始時刻から， t1 経過時までの R(e) より，テスト用に選択した R(e) の集合．
コース	CO	年度とクラスの 2 つのデータの組み合わせのデータ．
要指導受講者判定割合	r	ある受講者が要指導受講者と判定されるか否かを判定する値． $0 < r \leq 1$ ．
要指導受講者	IS(Y,C,P _m ,r,t2)	年度 Y，クラス C の単元番号 m の演習の課題が t2 時に解けている割合が r 未満の受講者．

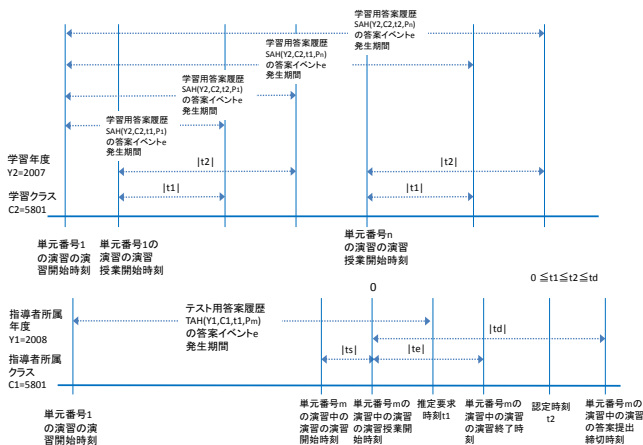


図 2 用語に関する図

況データの 2 種類がある．学習用答案作成状況データは，学習に使用するコースの受講者番号，単元番号ごとにサンプルが存在し，テスト用答案作成状況データはテストに使用するコースの受講者番号ごとにサンプルが存在する．答案作成状況データの説明変数の属性は，学習用答案作成状況データ，テスト用答案作成状況データともに共通であり，各単元のプログラミング演習を開始する時刻から | 推定要求時刻 t1 | 経過時のデータであり，表 2 に示す $a_1 \sim a_{14}$ の合計 14 種類の属性を持つ．また，目的変数の属性は学習用答案作成状況データ，テスト用答案作成状況データともに共通であり，要指導受講者判定属性の 1 種類である．要指導受講者判定属性のデータは，学習用答案作成

表 2 答案作成状況データの説明変数の属性

a_1	単元番号までの合計課題数
a_2	単元番号の課題数
a_3	単元番号の課題合計提出回数
a_4	単元番号までの課題の合計正解数
a_5	単元番号の課題の合計正解数
a_6	単元番号までの課題の解けた割合
a_7	単元番号の課題の解けた割合
a_8	課題合計提出回数
a_9	単元番号までのコースの課題の平均正解数
a_{10}	単元番号に対するコースの課題の平均正解数
a_{11}	単元番号までのコースの課題の解けた割合
a_{12}	単元番号に対するコースの課題の解けた割合
a_{13}	要指導受講者認定回数
a_{14}	要指導受講者認定回数の比率

状況データでは設定されるが，テスト用答案作成状況データでは空のデータとなる．

4.2 要指導受講者決定木作成機能

要指導受講者決定木作成機能の構成図を図 4 に示す．要指導受講者決定木作成機能では，推定要求時刻 t1，学習年度 Y2，クラス C2，CAPES のデータベースから取得した学習用答案履歴 SAH(Y2, C2, t1, P₁)…SAH(Y2, C2, t1, P_n(n は Y2, C2 の単元数)) より，P₁ から P_n までの学習用答案作成状況データを作成する．次に，作成した答案作成状況データと指導者から取得した認定時刻 t2 と要指導受講者判定割合 r，CAPES のデータベースから取得した SAH(Y2, C2, t2, P₁)…SAH(Y2,

する人数 $n=1, 3, 5, 10, 20$ の入力データを使用し, Y1, C1 の総ての単元それぞれにたいして, 表 2 の答案作成状況データの説明変数 $a_1 \sim a_{14}$ の 14 種類の属性の 2 の 14 乗通り総ての組み合わせをそれぞれ使用した際の Y1, C1 の各受講者の要指導率を求めた. そして, 要指導率の高い n 人を要指導受講者と推定し, 選択した n 人が実際に要指導である確率 (以下要指導的中率) を求めた. さらに, 属性の組み合わせに対して, 上記のように各単元の要指導的中率を求め, それらの平均を求めた. その結果, 表 2 の $a_1 \sim a_{14}$ の属性を用いた以下の組み合わせで要指導的中率が高くなった.

$D_1:(a_6, a_9, a_{12}, a_{13}, a_{14})$

$D_2:(a_3, a_4, a_{14})$

$D_3:(a_3, a_{13})$

$D_4:(a_6, a_{11}, a_{14},)$

$D_5:(a_8, a_1, a_4, a_{14})$

$D_6:(a_8, a_4, a_5, a_6, a_{13})$

$D_7:(a_3, a_1, a_5, a_6, a_9, a_{11})$

$D_8:(a_6, a_{13})$

$D_9:(a_2, a_5, a_{11}, a_{13})$

$D_{10}:(a_4, a_5, a_{13})$

$D_{11}:(a_8)$

$D_{12}:(a_7, a_{10}, a_{11}, a_{14})$

$D_{13}:(a_4, a_5)$

7. 評価実験

評価実験は, 予備実験で要指導的中率が高い結果となった $D_1 \sim D_{13}$ を使用した提案システム (以下それぞれ $M_1 \sim M_{13}$) と, 進捗状況の悪い受講者を要指導受講者と推定する手法 (以下従来手法 M_{14}) を比較した. 入力データを指導者の所属する年度 Y1=2011, クラス C1=5607, 5608, 5609 とし, 他は予備実験と同様にして, 提案手法による推定と従来手法による推定によって要指導的中率を求め, 平均を求めると, 表 3 の結果を得た. M_4 は $n=1$ の場合, M_{12} は $n=3, n=5$ の場合, M_3 は $n=10, 20$ の場合に最も要指導的中率が高い結果となった. また, n が大きくなるに連れて, どの組み合わせの場合も従来手法とほぼ変わらない結果となっている. これらの結果より n が小さな場合に M_3, M_4, M_{12} は, 従来手法 M_{14} に比べ, 要指導的中率の高い推定を期待できると考えられる.

8. おわりに

本研究では, プログラミング演習支援システム CAPES における受講者答案提出履歴を用いた要指導受講者推定システムを提案した. また, プロトタイプシステムの作成, 実際の演習データを用いた予備実験の結果, 評価実験の結果について述べた.

今後の課題として, 学習用答案作成状況データの母集団や属性に関する検証, 要指導率の計算方法の改善が挙げられる.

表 3 要指導的中率の比較結果

手法	n=1	n=3	n=5	n=10	n=20
M_1	71.7	66.3	58.6	47.4	38.3
M_2	73.3	67.4	59.9	48.9	40
M_3	77	72.5	67.3	55.7	43.7
M_4	83.4	72	62.8	49.3	39.7
M_5	78.2	72.1	63.4	50.7	40.5
M_6	59.9	60.4	57.3	50.5	41.6
M_7	48.2	47.4	45.2	41.5	36.5
M_8	76.2	72.2	65.7	53.1	40.4
M_9	58.7	55.8	52.5	46.2	39
M_{10}	57.4	58.2	55.5	48.1	39.9
M_{11}	41.7	40.2	38.7	35.4	33.5
M_{12}	82	77	68	53.5	41.1
M_{13}	63.2	60.4	58	51.7	42.1
M_{14}	51.2	51.2	50.9	47.6	42.2

文 献

- [1] 中島秀樹, 宮地恵佑, 高橋直久, "プログラミング演習支援システム CAPES のための答案評価機構の実現", 情報処理学会研究報告, 2006-CE-83(18), pp127-134, 2006.
- [2] 片桐由裕, 山本大介, 立岩佑一郎, 高橋直久, "プログラミング演習指導者補助システムのための進捗状況提示機能の実現法", 電子情報通信学会総合大会講演論文集 2009 年情報・システム (1), 216, 2009-03-04
- [3] java.com <http://www.java.com/ja/>
- [4] MySQL 世界でもっとも普及している、オープンソース データベース <http://www-jp.mysql.com/>
- [5] Apache Software Foundation - Apache Tomcat <http://tomcat.apache.org/index.html>
- [6] Weka 3: Data Mining Software in Java <http://www.cs.waikato.ac.nz/ml/weka/>