

プログラミング演習支援システムCAPESにおける 演習履歴を用いた要指導受講者予測システムの提案

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 工学研究科 情報工学専攻[‡]

1. はじめに

我々は、プログラミング演習支援システム CAPES[1]を提案し、学部の演習に適用している。CAPES を適用した演習では、受講者は各演習ごとに設定された複数の課題に回答を行う。また、CAPES は、答案プログラムを自動評価し、答案の提出時刻や正誤判定結果などの演習履歴のデータを詳細に収集し、指導者に提供している。指導者はこのようなデータを元に、課題に躓いている受講者を予測する。しかし、以下のような問題がある。

問題 1 指導者が、各受講生の過去の情報をすべて記憶することは難しい。そのため、指導者が各受講生の過去の情報を用いて指導に必要な受講者を予測することは難しい。

問題 2 指導者が、各受講生の過去の情報を用いて指導に必要な受講者を予測することは手間がかかる。

問題 3 指導者が、各受講生の過去の情報を用いて指導に必要な受講者を予測した場合、予測精度は指導者の経験や直感に依存する。

本稿では、これらの問題を解決するために、指導に必要な受講者(以下要指導受講者とする)の機械的予測を行うシステムの提案、さらにプロトタイプシステムによる評価実験の結果について考察を行う。

2. 要指導受講者の機械的予測システムの課題

要指導受講者の機械的予測システムの実現にあたって、以下のような課題がある。

課題 1 指導者によって、答案提出期限内に全問正解できない受講者、授業時間内に正解数が少ない受講者、現在解いている課題に躓いている受講者、授業中であるが課題を解く意欲のない受講者など要指導受講者の定義は異なる。そのため指導者によって異なった要指導受講者の予測、予測結果の提示を行うようにする。

課題 2 要指導受講者を予測するために必要なデータを明らかにする。

これらの課題の解決案として、以下のような特徴を持つ要指導受講者の機械的予測システムを提案する。

特徴 1 提案システムは要指導受講者の定義を複数もつ。指導者はそれらの要指導受講者の定義から、自身の要指導受講者の定義と合致するものに優先度をつけて複数選択する。提案システムは指導者が選択した複数の要指導受講者の定義のパターンによって異なった要指導受講者の予測を行う。これにより、問題 1 を解決する。

特徴 2 数年間に渡る多数のクラスのプログラミング演習における答案提出開始時刻からの経過時間、答案提出開始時刻、答案提出時刻、答案正誤判定結果などの詳細な演習履歴から判別分析可能なデータベース(以下 DADB とする)を作成する。これにより、各種判別式を作成して妥当性の検査を容易にすることで、問題 2 を解決する。

3. 提案システムの実現方法

提案システムは DADB 作成プログラム、DADB、要指導受講者データ作成プログラム、統計分析プログラムで構成される。提案システムの構成図は図 1 のようになる。提案システムは以下のような手順で DADB の構成、要指導受講者の予測に使用する属性選択を行う。

手順 1 CAPES の履歴データベースから、学籍番号、課題の答案提出開始時刻、課題の答案提出締め切り時刻、受講者の答案提出時刻、答案の正誤判定結果などのデータを取得する。この手順は、CAPES の履歴データベースで取得する一次データの変更、追加が生じた場合に実行される。

手順 2 手順 1 で取得したデータから、各演習の各受講者ごとに課題正解数などの複数の説明変数のデータ、答案提出期限内に全問正解したかどうかなどの要指導であるかどうかを示す複数の目的変数のデータを求め、DADB を作成する。

手順 3 DADB から複数の目的データの集合のすべての部分集合に対し、それらの論理積に対して要指導受講者の予測精度が最も高い属性集合をそれぞれ求め、その属性集合を保存する。

また、提案システムは、以下のような手順で指導者に要指導受講者の予測結果を提供する。

手順 1 指導者は提案システムに設定された要指導受講者の複数の定義に対し、指導者の要指導受講者の定義と合致するものに優先度をそれぞれ設定する。その後要指導受講者の予測を提案システムに要求する。

手順 2 提案システムは、指導者が優先度を設定した要

Suggestion of the system which predicts students who should be instructed

Hiroyuki Okuda [†], Yuichiro Tateiwa [†], Daisuke Yamamoto [†], Naohisa Takahashi [†]

[†] Dept. of Computer Science and Engineering, Graduate School of Engineering, Nagoya Institute of Technology

指導受講者の定義の集合のすべての部分集合に対し、保存された要指導受講者の予測精度の最も高い属性集合を用いて要指導受講者の予測式を作成する。そして要指導受講者の予測結果を取得し、結果を指導者に提供する。

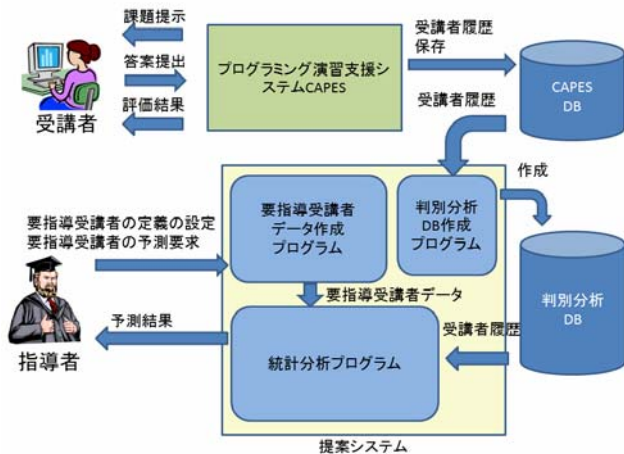


図1 提案システムの構成図

4. 評価実験と結果

提案システムによる要指導受講者の予測精度の検証のためプロトタイプシステムの開発を行い、従来手法による要指導受講者の予測精度との比較を行った。プロトタイプシステムでは、CAPESのデータベース上の2006年から2008年間の受講者履歴データを使用し、要指導受講者は提出期限内に全問正解できない受講者としている。プロトタイプシステムでは、指定時刻 t 以前の受講者データを使用して各受講者の演習開始日の0時から t' 経過した時点でのデータを取得し、各受講者が演習開始日の0時から t' 経過した時点で要指導受講者であるかどうかを判別する予測式を作成する。そしてその予測式に指定時刻 t 以降のデータから取得した各受講者の演習開始日の0時から t' 経過した時点でのデータを代入することで、 t 以降のデータの各受講者データに対し演習開始日の0時から t' 経過した時点で要指導であるかを予測する。

今回の評価実験では、 t を2008年1月7日0時、2008年12月7日の2通り、 t' を24時間の1通りとした。要指導受講者の予測式の作成、予測式に代入する各受講者のデータは、 t' の時点での正解数、課題の数、過年度生であるか、進捗状況、クラスの平均正解数、クラスの平均進捗状況、過去の演習で要指導受講者に判定された回数の7つである。要指導受講者の予測時に、7つとも使用する場合、進捗状況、クラスの平均進捗状況、過去の演習で要指導受講者に判定された回数の3つを使用する場合、過去の演習で要指導受講者に判定された回数のみを使用する場合をそれぞれ手法1, 2, 3とする。予測式の作成には、決定木のアルゴリズムであるJ48アルゴリズムを使用した。

また、従来手法による予測方法を演習ごとに演習開始日の0時から t' 経過した時点で最も正解数の少ない受講者を要指導受講者とする方法とした。

従来手法、プロトタイプシステムによる手法による

要指導受講者の予測の結果は、表1, 2のようになった。従来手法は要指導受講者の検出率が高いが、要指導受講者の予測的中率は低く実際に予測を使用した場合には効率が悪いと考えられる。

手法3では、他の手法に比べ要指導受講者の予測的中率が高く、要指導受講者の検出率も従来方法以外の手法と同等となった。過去の演習で要指導受講者に判定された回数のデータは演習開始日の0時から24時間後に要指導受講者を予測するためのデータとして妥当であると考えられる。

一方、手法1, 2では手法3と比較して要指導受講者の検出率が高い場合もあるものの、要指導受講者の予測的中率には差があった。これは、演習開始日の0時から24時間後の時点では受講者の進捗状況のデータでは要指導受講者の予測は難しいと考えられ、また過学習により要指導受講者の予測的中率が下がっていると考えられる。

表1 $t=2008$ 年1月7日0時
評価データ数2010の場合の
要指導受講者の予測結果

	従来手法	手法1	手法2	手法3
要指導受講者と予測した場合の的中率	12.19	46.15	36.36	78.26
要指導受講者の検出率	86.27	35.29	36.60	35.29

表2 $t=2008$ 年12月7日0時
評価データ数847の場合の
要指導受講者の予測結果

	従来手法	手法1	手法2	手法3
要指導受講者と予測した場合の的中率	14.00	55.32	57.78	82.35
要指導受講者の検出率	85.71	33.77	33.77	36.36

5. 今後の課題

授業時間内に要指導受講者を予測する場合の予測精度の検証と向上、要指導受講者の予測の人数指定、指導者ごとに異なった要指導受講者の予測を行う機能の実現が挙げられる。

参考文献

- [1] 中島秀樹, 高橋直久, 細川宜秀: “プログラミング学習のためのQAサイクル 受講者の習得度に応じた問題自動提示メカニズム”, 電子情報通信学会論文誌, VOL. J88-D-I, NO. 2, 2005
- [2] 片桐由裕, 高橋直久: “操作履歴分析機能を有する状況適応型プログラミング演習指導者支援システムに関する研究”, 平成21年度修士論文