



И.А. Андрианов, А.В. Ржеуцкий, С.Ю. Ржеуцкая
Вологодский государственный университет

ПРОГНОЗИРОВАНИЕ ТРУДНОСТИ УЧЕБНЫХ ЗАДАНИЙ В АДАПТИВНОЙ СРЕДЕ ЭЛЕКТРОННОГО ОБУЧЕНИЯ

В статье решается задача автоматического индивидуального подбора учебных заданий в системе электронного обучения на основе прогнозирования их трудности с учетом подготовленности обучающегося. Исследуются возможности обучаемых моделей нейронных сетей и деревьев решений, приводятся результаты эксперимента на реальных данных.

Адаптивное обучение, прогнозирование трудности учебных заданий, автоматический подбор учебных заданий.

Применение интеллектуальных технологий Educational Data Mining (EDM) в средствах электронного обучения позволяет существенно повысить их адаптивность. В данной статье представлен один из возможных подходов к решению задачи автоматического индивидуального подбора учебных заданий, выполнение которых будет иметь наилучший обучающий эффект.

Обучение любой дисциплине, как правило, включает лекции, практические занятия, а также самостоятельную работу студентов. В рамках аудиторного лекционного курса все обучающиеся получают одинаковый образовательный контент. Для того чтобы каждый студент получил наилучший результат обучения, необходимо разумно организовать практические занятия и самостоятельную работу, используя интеллектуальные функции системы электронного обучения. По мнению авторов, одной из таких функций является рациональный индивидуальный подбор учебных заданий, предназначенных для самостоятельного выполнения студентами.

Для большинства заданий, размещенных в системе электронного обучения, имеется возможность автоматической проверки их решений с фиксацией результатов в базе данных. В отличие от контрольных заданий, для учебных заданий количество попыток их отправки на автоматическую проверку не ограничено, как и время их выполнения. Однако каждая попытка сдачи решения студентом отдельно фиксируется в системе, данные, по которым можно оценить время выполнения задания, также собираются.

Количество имеющихся в системе заданий, как правило, избыточно для студента. Например, в образовательной среде Вологодского государственного университета, предназначенной для подготовки ИТ-специалистов, имеется около 2000 задач по программированию, решения которых проверяются ав-

томатически, и еще больше тестовых заданий по различным дисциплинам [6]. При таком количестве учебных заданий задача их рационального подбора с учетом индивидуальных особенностей студентов требует компьютерной поддержки. Конечно, результаты автоматического подбора имеют характер рекомендаций, а окончательное решение по выбору заданий принимается преподавателем и студентом.

Постановка задачи представлена в виде схемы на рисунке 1.

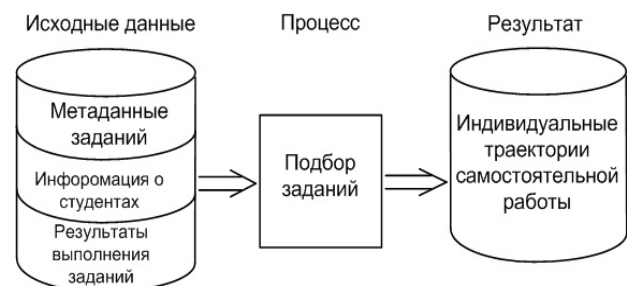


Рис. 1. Постановка задачи автоматического подбора заданий

Исходными данными для решения этой задачи являются множество данных, имеющихся в среде электронного обучения, которые представляют интерес для анализа в процессе подбора заданий. Некоторые данные фиксируются в системе при занесении нового задания или регистрации нового обучающегося, но большая их часть накапливается в процессе обучения. Результатом подбора являются последовательности учебных заданий по каждой теме, подобранные индивидуально для каждого студента. Назовем их индивидуальными траекториями самостоятельной работы.

Индивидуальные траектории формируются в соответствии с принципом разумного возрастания трудности заданий для студента. При этом слишком легкие и слишком трудные для конкретного студента задания должны отсутствовать в его индивидуальной траектории. Опыт авторов показывает, что начинать освоение навыков по каждой теме следует с 2–3 относительно нетрудных для студента заданий, затем тренироваться на заданиях средней сложности. После этого следует предложить студенту несколько трудных заданий, с которыми он в состоянии справиться, приложив усилия.

Для реализации этих разумных принципов формирования индивидуальной траектории требуется решить задачу прогнозирования трудности задания для студента с учетом его подготовленности, основываясь на данных, накопленных в системе электронного обучения. Задача педагогического прогнозирования относится к направлению EDM. Рассмотрим основные подходы к ее решению.

При занесении учебного задания в систему электронного обучения обычно задаются некоторые метаданные задания, одним из них является трудность задания. Как правило, трудность оценивает преподаватель-эксперт, который разработал это задание, в соответствии с принятой системой баллов, поэтому такая оценка трудности называется экспертной.

Преподаватель-эксперт оценивает трудность задания, ориентируясь на свой алгоритм выполнения этого задания и «среднего» студента. Однако реальную трудность учебного задания можно оценить только после того, как это задание выполнила группа обучающихся. Назовем оценки трудности, полученные по результатам выполнения задания, эмпирическими (статистическими) оценками.

Наиболее распространенной является эмпирическая оценка трудности как доли обучающихся, выполнивших (или не выполнивших) задание по отношению к общему количеству выполнявших это задание. Обычно такая оценка выражается в процентах и может считаться объективной характеристикой задания при условии, что количество выполнявших задание студентов достаточно велико для получения достоверной оценки.

С учетом специфики учебных заданий можно предложить дополнительные эмпирические оценки трудности заданий. По мнению авторов, объективной характеристикой трудности учебного задания, которую можно точно вычислить на основании имеющихся данных системы автоматической проверки, является среднее количество попыток сдать задание, прежде чем система приняла его как верное. Если учебное задание представляет собой обучающий тест, то для него задается пороговый балл, при котором система засчитывает тест как пройденный, – количество попыток пройти тест также легко вычислить.

Параметры, позволяющие судить о времени выполнения задания (интервал времени от момента открытия задания до его принятия системой автоматической проверки, возможно, интервалы времени между попытками сдать задание) также характеризу-

ют трудность этого задания, но их не всегда можно вычислить точно, поскольку в процессе выполнения задания студент может отвлекаться. Тем не менее в большинстве случаев эти параметры хорошо коррелируются с реальной трудностью задания.

Все перечисленные эмпирические оценки трудности заданий могут автоматически пересчитываться по мере накопления статистики выполнения заданий, при этом точность оценивания повышается.

Наш опыт показал, что экспертная оценка трудности учебных заданий далеко не идеально совпадает с эмпирическими данными, хотя и резкое различие между ними встречается довольно редко. Часто эмпирические оценки позволяют немного откорректировать выставленную ранее экспертную оценку трудности. В момент появления задания в системе при отсутствии статистических данных о выполнении задания в качестве меры трудности может быть использована только экспертная оценка.

Вернемся к вопросу прогнозирования трудности задания для конкретного студента в процессе подбора заданий для адаптивного обучения. Один из хорошо проработанных способов решения задачи прогнозирования трудности заданий представлен в теории тестирования, известной под названием Item Response Theory – IRT [5]. Согласно теории IRT, прогнозируемая трудность задания для студента оценивается как вероятность успешного выполнения студентом этого задания и является действительным числом в интервале (0–1). Такая интерпретация позволяет создавать качественные диагностические тесты, в которых задания упорядочены по возрастанию трудности.

В некоторых публикациях [1, 2] прогнозная модель на основе подсчета вероятности правильного выполнения задания развивается применительно к подбору учебных заданий в процессе адаптивного обучения. Однако в процессе практической реализации этих моделей в реальной среде электронного обучения выявились их недостатки, которые не позволили получить приемлемое качество прогноза трудности заданий.

Модель IRT была разработана в свое время для тестовых заданий с выбором одного правильного ответа из нескольких вариантов, поэтому в современных условиях она представляется чересчур упрощенной. Трудность задания в ней оценивается по доле ошибочных результатов при выполнении задания, а подготовленность студента – по доле правильно выполненных заданий от общего количества заданий. Упрощенный двухпараметрический подход не учитывает такие важные показатели трудности задания, как время его выполнения и количество попыток сдачи задания. Например, в дистанционном практикуме по программированию ВоГУ зафиксированы случаи, когда задачи сдавались упорными студентами с тридцатой и даже сороковой попытки, при этом решались много дней. С точки зрения моделей, основанных на IRT, такие задачи равноценны по трудности задачам, сданным с первой попытки, – в обоих случаях фиксируется успешное выполнение задания. Аналогично обстоят дела и с подготовленностью студентов – некоторые из них могут намеренно решать много легких для себя

задач, чтобы повысить уровень подготовленности, вычисляемый системой по упрощенной формуле. Подчеркнем, что в системе электронного обучения имеется огромное количество статистических данных, которые могли бы уточнить результат прогноза, но выходят за рамки IRT.

Проблема возникла и с интерпретацией результатов прогноза. Как связана вероятность выполнения задания студентом с обучающим эффектом этого задания? Как адекватно оценить качество прогноза? Путь решения проблемы предложен в работах М.Б. Челышковой и В.С. Клопченко. Суть его заключается в переходе от непрерывной шкалы оценки трудности задания к дискретной. Для этого предлагается разбить весь интервал значений вероятности p на несколько подинтервалов, каждому из которых может быть поставлен в соответствие определенный уровень трудности. Например, можно выделить пять уровней, соответствующих общепринятым понятиям о трудности заданий:

1. $(0,8 \leq p < 1)$ – задание слишком легкое для обучающегося.
2. $(0,6 \leq p < 0,8)$ – задание невысокой трудности.
3. $(0,4 \leq p < 0,6)$ – задание средней трудности.
4. $(0,2 \leq p < 0,4)$ – задание повышенной трудности.
5. $(0 < p < 0,2)$ – задание слишком трудное.

Количество уровней трудности (шаг дискретизации значений вероятности) не имеет принципиального значения, важна сама идея представления результата прогноза в виде интерпретируемого значения, которое может быть напрямую использовано в процессе подбора заданий. В развитие этой идеи в данном исследовании предлагается альтернативный подход к прогнозированию трудности учебных заданий.

Предлагаемый авторами способ состоит в том, чтобы решать задачу прогнозирования трудности задания для студента как задачу классификации пар «студент – задание» [5]. Результатом прогноза будет отнесение каждой пары к определенному классу трудности задания для студента на основе множества признаков, которые можно получить из системы электронного обучения. Множество классов может содержать, например, пять элементов, значения которых соответствуют приведенным выше пяти уровням трудности.

Задача автоматической классификации относится к задачам машинного обучения «с учителем», т.е. с использованием обучающей выборки [3]. Методы классификации используются и для анализа образовательных данных, часто они преследуют цель прогнозирования результатов обучения. Например, в [4] представлено решение задачи прогноза успеваемости на основе классификации студентов по множеству их признаков. Принципиальным отличием подхода авторов является расширение признакового пространства, которое должно включать характеристики и студента, и задания, а решение задачи должно выявить, насколько трудным является задание для студента.

Тем не менее представленная постановка задачи позволяет применить стандартный сценарий решения задачи машинного обучения (рис. 2).

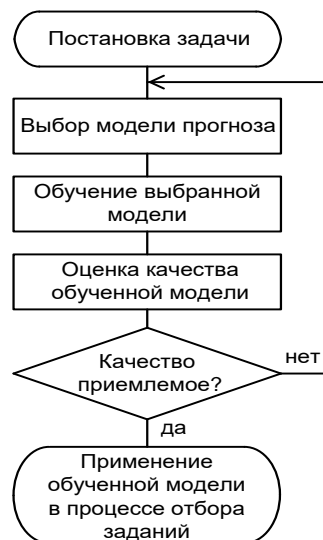


Рис. 2. Этапы решения задачи прогнозирования трудности заданий

Кратко поясним основные этапы.

Выбор модели прогноза на основе классификации требует внимательного анализа известных моделей, обзор которых представлен в [2, 4]. В данном исследовании эксперименты проводились с двумя существенно различными моделями – нейронными сетями и деревьями решений. Каждая из этих моделей имеет свои достоинства и недостатки применительно к решаемой задаче.

Нейросети – широко используемый, довольно точный, постоянно развивающийся способ классификации. Тем не менее, для задачи подбора учебных заданий представляет интерес и модель деревьев решений. Дерево решений представляет собой иерархически организованную систему правил «если..., то...», определяющую алгоритм принятия решения. Такая модель обладает рядом преимуществ:

- деревья решений в наглядной форме представляют правила прогноза на основе классификации, которые интересны сами по себе;
- модель деревьев решений обладает способностью отбора признаков, наиболее значимых в процессе классификации. Таким образом, предварительно можно отобрать избыточное множество признаков, а в процессе построения дерева решений малозначимые признаки будут отброшены.

По сравнению с нейронными сетями деревья решений на многих задачах показывают более низкую точность прогноза. Однако на начальном этапе проводимого исследования преимущества деревьев решений более важны, чем некоторое снижение точности.

Этап обучения модели классификации состоит в настройке параметров модели с использованием обучающей выборки. В ходе вычислительного эксперимента для формирования обучающей выборки использовались реальные данные системы электронного обучения студентов ИТ-направлений, в том числе результаты выполнения заданий, необходимые для обучения моделей. Далее будут представлены примеры обучения и использования прогнозных моделей.

Перед использованием обученной модели прогноза требуется оценить ее точность на экзаменационной выборке, содержащей такие прецеденты выполнения заданий, которые не были использованы в процессе обучения модели. Точность прогнозной модели определяется как процент совпадения результатов прогноза с фактическими данными. Для рассматриваемой задачи целесообразно дополнить эту оценку введением функции потерь [3], которая определяет величину штрафа за ошибку в зависимости от класса, в который попал прогноз. Штраф будет минимальным, если результат прогноза отнесен к соседнему, наиболее близкому классу. По мере удаления результатов прогноза от фактических результатов величина штрафа возрастает.

Далее представим ход и результаты вычислительного эксперимента с использованием данных, накопленных в дистанционном практикуме по программированию с автоматической проверкой решений задач, который является частью среды обучения студентов ИТ-направлений ВоГУ и используется в учебном процессе более 15 лет. За это время накоплено огромное количество данных, которые можно использовать для анализа в процессе подбора задач и организации адаптивного обучения.

Предварительный этап эксперимента состоял в выделении множества классов и множества признаков с учетом специфики организации данных в практикуме. Для объективной проверки точности прогнозной модели требуется формально описать каждый из классов, чтобы полученное значение прогноза трудности заданий можно было однозначно сравнить с фактическими результатами выполнения заданий. В базе данных практикума фиксируется каждая попытка сдать задачу, при этом сохраняются данные о статусе решения (верно или неверно) и времени поступления решения задачи на проверку. Таким образом, при определении классов можно использовать такие критерии:

- 1) success – есть ли среди попыток сдать решение задачи хотя бы одна успешная;
- 2) cnt – количество попыток до принятия решения задачи системой автоматической проверки;

3) t – интервал времени в часах между первой и успешной попыткой.

Формальное определение выделенных нами классов представлено в таблице 1. Числовые параметры, определяющие количество попыток и время, легко настраиваются и всегда могут быть уточнены. Приведенные в таблице 1 значения установлены, исходя из опыта и здравого смысла.

Уточним исходные данные для классификации – признаки задачи и студента, которые мы выделили как значимые (табл. 2). Напомним, что при использовании модели деревьев решений из них будут автоматически отобраны наиболее важные.

Отметим, что динамические данные по задачам и результатам студентов в период активного обучения обновляются каждый день. Личностные качества определяются анкетированием и тестированием на добровольной основе – студенты охотно принимают участие в таких экспериментах.

Статистических данных о результатах решения задач в практикуме хватило и на обучающую, и на экзаменационную выборки по нескольким учебным курсам. Для обучения моделей использовался популярный пакет алгоритмов машинного обучения WEKA (Waikato Environment for Knowledge Analysis), который поддерживает целую линейку различных моделей классификации, в том числе нейронные сети и деревья решений. Кроме того, в рамках студенческих ВКР было реализовано несколько программных продуктов на языках C++ и Python.

В ходе эксперимента было получено несколько вариантов деревьев решений, представляющих вполне разумные, но довольно объемные правила классификации. Как и следовало ожидать, самым важным признаком трудности задачи (корень дерева решений) оказался процент студентов, сумевших решить задачу. Заслуженно высокий приоритет оказался у признаков студента «мотивация», «процент успешно решенных задач», «средняя трудность решенных задач».

Таблица 1

Возможное определение классов трудности задач

№	Обозначение	Определение
1	Слишком легкое	success=true \wedge cnt=1
2	Легкое стартовое	success=true \wedge cnt=2 \wedge t \leq 1
3	Средней трудности	success=true \wedge (cnt=2 \wedge t>1 \vee cnt=3 \wedge t \leq 3)
4	Трудное, но в меру	success=true \wedge (cnt=3 \wedge t>3 \vee cnt>3 \wedge cnt \leq 6)
5	Слишком трудное	success=false \vee (success=true \wedge cnt>6)

Таблица 2

Признаковое пространство для решения задачи прогнозирования

Признаки задачи	Признаки студента
Статические: раздел (тема), тренируемые навыки; экспертный уровень трудности; дата занесения задания в систему. Динамические: число студентов, решавших задачу; процент студентов, решивших задачу; среднее время с первой до успешной попытки; среднее число попыток до принятия решения	Результаты на данный момент: число задач, которые решались; процент принятых решений задач; среднее число попыток до принятия решения; средняя эмпирическая оценка трудности принятых решений. Личностные качества (при наличии данных): память, внимание, мотивация, ответственность

Точность прогноза, оцененная как процент точного совпадения прогнозируемого и фактического класса трудности, в первых экспериментах оказалась довольно низкой у обеих классификационных моделей, но при анализе неверных результатов прогноза было обнаружено, что их подавляющее большинство попало в соседние классы. Например, если задача реально оказалась для студента легкой (допустим, быстро сдал, но со второй попытки), то результатом прогноза могло быть «слишком легкое» или «средней трудности», но как трудное для студента такое задание классифицировалось крайне редко. Таким образом, среднее значение штрафа за неверный прогноз оказалось минимальным. Заметим, что в решаемой задаче нечастое попадание в соседний класс не принесет особого вреда.

Существенного повышения точности прогноза удалось добиться путем сокращения количества классов трудности задачи до трех (легкая, средней трудности, трудная) и изъятия из обучающей выборки всех прецедентов, у которых зафиксирован высокий уровень плагиата (таких прецедентов использования студентами чужих решений, к сожалению, немало). В этом случае в некоторых экспериментах удалось довести точность до 78 % при минимальном значении штрафа. По результатам эксперимента уже можно сделать вывод о том, что более высокой точности прогноза для такой предметной области, как обучение, добиться очень трудно, скорее всего, невозможно.

Представленный способ прогнозирования трудности заданий пока используется в дистанционном практикуме в экспериментальном порядке, поэтому его влияние на результаты обучения оценивать рано. Разумеется, автоматический индивидуальный подбор заданий – только одна из интеллектуальных функций,

которые можно реализовать в среде электронного обучения для эффективной поддержки учебного процесса.

Литература

1. Andersen, P.-A. Adaptive task assignment in online learning environments / P.-A. Andersen, C. Krakevik, M. Goodwin, A. Yazidi // Proceedings of the 6th international conference on web intelligence, mining and semantics. – New York : ACM, 2016. – pp. 1–10.
2. Yazidi, A. Balanced difficulty task finder: an adaptive recommendation method for learning tasks based on the concept of state of flow / A. Yazidi, A. Mofrad, M. Goodwin, H. Hammer, E. Arntzen // Cognitive Neurodynamics. – 2019, 14. – pp. 675–687.
3. Бринк, Х. Машинное обучение / Х. Бринк, Дж. Ричардс, М. Феверолф. – Санкт-Петербург : Питер, 2017. – 336 с.
4. Котова, Е. Е. Прогнозирование успешности обучения в интегрированной образовательной среде с применением инструментов онлайн-аналитики / Е. Е. Котова // Компьютерные инструменты в образовании. – 2019. – № 4. – С. 55–80.
5. Ржеуцкая, С. Ю. Способ автоматического подбора учебно-тренировочных заданий в информационной среде обучения студентов ИТ-направлений / С. Ю. Ржеуцкая, М. В. Харина // Открытое образование. – 2020. – Т. 24, № 2. – С. 17–28.
6. Архитектура интеллектуального агентно-ориентированного учебного комплекса для подготовки специалистов технического профиля / А. Н. Швецов, С. Ю. Ржеуцкая, А. П. Сергушичева, А. А. Суконщиков // Открытое образование. – 2018. – 22 (3). – С. 14–24.

I.A. Andrianov, A.V. Rzhetskiy, S.Yu. Rzhetskaya
Vologda State University

PREDICTING DIFFICULTY OF TRAINING TASKS IN ADAPTIVE E-LEARNING ENVIRONMENT

The article solves the problem of automatic individual selection of training tasks in the e-learning system based on predicting their difficulty, taking into account the readiness of a student. The possibilities of the trained models of neural networks and decision trees are investigated, the results of the experiment on real data are presented.

Adaptive learning, predicting the difficulty of training tasks, automatic selection of training tasks.