

УДК 004.5

РАЗРАБОТКА АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ ТЕСТИРОВАНИЯ ЗНАНИЙ СТУДЕНТОВ В ОБЛАСТИ КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ С ИСПОЛЬЗОВАНИЕМ МЕХАНИЗМОВ АДАПТАЦИИ

Трофименко Е.С., Скобцов Ю.А.

Донецкий национальный технический университет

e-mail: trofimenko@mls-automation.com

В статье рассматриваются вопросы построения систем тестирования знаний, а так же возможность использования в данных системах механизмов адаптации.

Общая постановка задачи

Важной частью процесса обучения является контроль знаний, который позволяет получить всестороннюю оценку уровня знаний обучаемых. Одной из хорошо зарекомендовавших себя форм контроля знаний является тестирование. Тестирование, обладает множеством определенных недостатков, однако, является единственным по-настоящему технологическим средством для измерения уровня знаний и незаменимо как инструмент, позволяющий должным образом организовать управление учебным процессом и обеспечить эффективный педагогический контроль.

Существует определенное многообразие типов тестовых заданий: закрытые (многоальтернативные и одноальтернативные), открытые, на установление соответствия между элементами, на установление правильной последовательности, ситуационные тестовые задания [3]. Одним из сложных и противоречивых вопросов при проведении тестирования является проблема оценивания знаний. Для эффективного решения данной проблемы необходимо использовать политомическую систему оценивания, в которой допускается несколько категорий ответа на задание, каждая из которых оценивается по-разному [5].

Исследования в области построения системы контроля знаний показали необходимость разделения заданий на уровни сложности. Отсутствие разделения заданий на уровни сложности приводит к недостаточной объективности оценивания знаний и часто не коррелируется с истинным уровнем знаний обучаемых [5]. Распределение заданий по уровням сложности преподавателем делает субъективным процесс оценивания знаний обучаемых, потому что не всякое легкое задание для преподавателя является легким и для студентов. Таким образом, актуальным является разработка технологии распределения тестовых заданий по уровням сложности [2].

4

Постановка задачи исследования

Цель исследования: разработать систему тестирования знаний студентов в сфере компьютерных технологий; внедрить, экспериментально проверить и программно оптимизировать технологию адаптивного компьютерного тестирования.

Объект исследования: профессиональная подготовка студентов в сфере компьютерных технологий; система создания адаптивных тестовых заданий.

Предмет исследования: создание и использование технологии адаптивного компьютерного тестирования в профессиональной подготовке студентов в сфере компьютерных технологий.

Решение задачи и результаты исследований

Для создания тестовой последовательности необходимо включить в нее задания разных типов. Для расчета объективной оценки тестовых заданий разных типов для каждого из них используется свой специализированный подход. Для этого вводится коэффициент оценивания γ_i . Итоговая оценка выполнения всего теста R , состоящего из набора тестовых заданий, содержащего Z уровней сложности, определяется по формуле (1).

$$R = \frac{B}{\dot{Z}} \times \sum_{i=1}^N r_i \times z_i, \quad (1)$$

где B – балльность системы, N – количество тестовых заданий, включенных в тест, z_i – уровень сложности i -го задания, \dot{Z} – суммарная сложность теста, которая определяется по формуле (2):

$$\dot{Z} = \sum_{i=1}^N z_i \quad (2)$$

Для построения тестовых последовательностей применяются различные методики. Однако одними из самых эффективных являются алгоритмы, построенные на принципах ИРТ и генетических алгоритмах (далее ГА)[1].

Здесь, ИРТ (Item Response Theory)[4] – математическая теория оценки качества тестовых заданий и вычисления уровня знаний обучаемых. В фокусе изучения ИРТ находятся не тесты, как системы заданий, а отдельные тестовые задания.

Основной математической моделью ИРТ является однопараметрическая логистическая функция Раша [5]. В данной модели вводится понятие вероятности P правильного выполнения i -м испытуемым j -го тестового задания, зависящей от параметров испытуемого и задания. Таким образом функцией успеха называется зависимость P от уровня измеряемого свойства испытуемого и уровня трудности задания [1]. В однопараметрической логистической модели Раша данная вероятность определяется разностью уровней развития свойства испытуемого θ_i и уровня трудности задания β_j и определяется по формуле (3):

$$P_j = \frac{1}{1 + e^{(\beta_j - \theta_i)}} \quad (3)$$

Задания обладают различной дифференцирующей способностью. Тогда, в соответствии с логистической моделью Бирнбаума, [5] вероятность успеха испытуемого составляет:

$$P_j = \frac{1}{1 + e^{d_j(\beta_i - \theta_i)}}, \quad (4)$$

где d_j – дифференцирующая способность задания. Понятие о дифференцирующей способности было введено с целью исключения из тестовой последовательности заданий с низкой дифференцирующей способностью.

Возможна также и логистическая трехпараметрическая модель с функцией успеха (трехпараметрическая модель Бирнбаума):

$$P = \frac{1}{1 + e^{\frac{d_i d_j}{\sqrt{(d_i^2 + d_j^2)}(\beta_j - \theta_i)}}} \quad (5)$$

4
 Параметры трехпараметрической функции находятся применением метода максимального правдоподобия путем проведения ряда итераций, при этом в качестве начального приближения целесообразно использовать значения θ_i и β_j , полученные исходя из решения системы уравнений для однопараметрической модели, а в качестве d_i и d_j – значения $\sqrt{2} \approx 1.41$, что дает

$$d_s = \frac{d_i d_j}{\sqrt{d_i^2 + d_j^2}} = 1, \quad (6)$$

т.е. значение, при котором модель Бирнбаума вырождается в модель Раша.

Задача распределения тестовых заданий по уровням сложности сводится к определению сложности тестовых заданий, исходя из экспериментальных данных предварительного тестирования.

В процессе создания тестов с помощью IRT, используется стратегия выбора элементов, основанная на целях разработчиков. Стратегия выбора элемента относится к задаче комбинаторной оптимизации. Когда увеличивается банк вопросов, время для расчетов увеличивается в несколько раз. Данная задача является NP-трудной, что предполагает многошаговый расчет в формировании лучшего

решения. ГА с успехом используются в данной ситуации [1].

ВданномГА, каждая хромосома представляет собой множество тестов k , которое содержит n бит (где, n – число тестов в банке), среди которого m бит (где, m – количество тестовых вопросов) равно 1, а остальные равны 0. x_i указывает на то, присутствует ли вопрос в тестовой последовательности или нет. Для каждой хромосомы, можно вычислить информационную емкость результируемого теста и отклонение от только что созданных тестов.

Квадрат суммы отклонений определяется как часть фитнес функции эволюционного процесса:

$$d_s = \frac{d_i d_j}{\sqrt{d_i^2 + d_j^2}} = 1, \quad (7)$$

где d_j – значение целевой информации о сложности уровня j , $j(=1 \sim s)$ – сложность уровня, s – количество уровней.

$$E(X^k) = \sum_{j=1}^S (d_j - o_j^k)^2, \quad (8)$$

где o_j^k ссылается на k -ю хромосому и сложность теста на j -м уровне. $i(=1 \sim n)$ i -й элемент в банке вопросов, n – количество вопросов в банке. w_{ij} – сложность i -го вопроса на j -м уровне. $x_i^k \in \{0,1\}$ – выбор i -го условия в k -й хромосоме.

С помощью генетических операторов, таких как репродукция, скрещивание и мутация, производится потомство особей с лучшими характеристиками.

Установка начальной популяции и параметров эволюционного процесса происходит следующим образом: случайно генерируется n бинарных строк. Вероятности кроссинговера, мутации и репродукции представляют переменные P_c , P_m , и P_r соответственно. В эволюционной алгебре T , начальное значение числа поколений равно 0, а максимальное значение – `gener_no`.

Расчет значений фитнес-функции для всех хромосом в популяции и определение оптимизированной фитнес хромосомы `chromosome_opt` определяется по формуле (9):

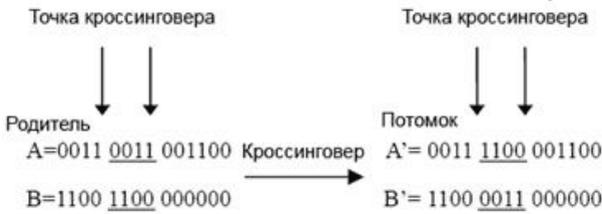


Рисунок 1 – Двухточечный кроссинговер

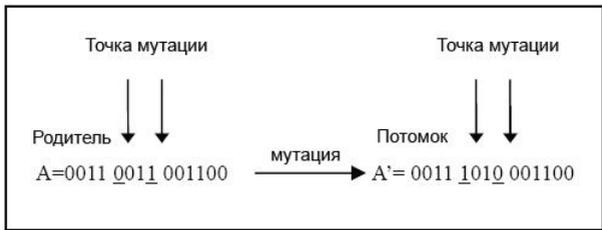


Рисунок 2 – Двухточечная мутация

$$fitness(X^k) = E(X^k) + r^k = \sum_{j=1}^s (d_j - o_j^k)^2 + r^k, \quad (9)$$

где r^k – число, удовлетворяющее условию $\sum_{i \in C_q} x_i > k_q$ или $\sum_{i \in C'_q} x_i > k'_q$,

где $E(X^k)$ – значение целевой функции, $q(=1 \sim p)$ – число q констант, p – общее число констант, C_q – количество q констант, включая множественные вопросы. k_q, k'_q – параметры для значения q -го ограничителя, r^k – число ограничений [6].

Для каждой хромосомы из популяции, применяются следующие генетические операции для генерации P потомков. Применяется двухточечный кроссинговер для каждой пары родителей с вероятностью P_c .

Так же используется двухточечная мутация с вероятностью P_m . Случайно выбирается хромосома из родительской популяции и в ней выбирается 2 позиции.

Оператор мутации производится, если x_i хромосома равна 0,

а x_j хромосома равна 1.

Лучшие хромосомы, найденные в родительской популяции, становятся потомками в новой популяции.

Если лучшая хромосома в дочерней популяции удовлетворяет требованиям создателя тестов, или количество генераций достигает максимума, тогда эволюционный процесс останавливается. Иначе, число итераций увеличивается на 1, и цикл продолжается, начиная с расчета фитнес функций особей в новой популяции.

В генетическом алгоритме Деба используется штраф-функция, которая не требует дополнительного штрафного параметра. Этот метод основан на простом ГА, в комбинации с эволюционной методикой Деба, предназначенной для стратегии выбора элементов с ограничительными условиями. Он отличается от простого ГА двумя аспектами.

Во-первых, методом расчета фитнес функции (10):

$$fitness(X_k) = \begin{cases} [E(X)]^k; \\ f_{max} + r_k. \end{cases} \quad (10)$$

Первое значение выбирается, если оно приемлемо. Иначе - второе.

Во-вторых, оператором кроссинговера:

Применяется двухточечный кроссинговер, использующий конкурентный выбор двух родителей. Далее применяются следующие действия:

1. Когда встречаются 2 приемлемых решения, выбор зависит от правила
 ЕСЛИ (E_d пороговая) ТО выбирается решение с наименьшим $E(X_k)$,
 ИНАЧЕ предыдущее решение комбинируется с приемлемым для получения E_d
 ЕСЛИ (E_d пороговая) ТО выбирается комбинированное решение,
 ИНАЧЕ выбирается предыдущее решение.
2. При выборе между приемлемым и неприемлемым решениями, выбирается приемлемое.

3. При выборе между двумя неприемлемыми решениями, выбирается то, которое нарушает ограничения в меньшей степени.

Расстояние и отклонение между двумя решениями после кроссинговера рассчитываются как стандартные.

Таким образом, алгоритм эффективно решает проблему «множественных ограничительных условий» в стратегии выбора элементов. Минимальное значение отклонения достигается для обеспечения более практичной методики, которая дает более высокое качество теста.

Выводы

Основными целями адаптивного тестирования являются:

1. Стремление к повышению эффективности тестовых измерений, что включает в себя уменьшение числа заданий, времени, стоимости тестирования
2. Повышение точности оценок, полученных испытуемыми по результатам выполнения теста
3. Повышение мотивации к тестированию у слабых и сильных студентов.

Именно это и легло в основу преимуществ описанных методов над тестами фиксированной длины. Существует возможность повышения эффективности адаптации тестов, трудность которых учитывает диапазон подготовленности тестируемых. При компьютерном адаптивном тестировании тестовые задания формируются индивидуально для каждого студента с учетом его ответов на предыдущие вопросы. Типы заданий, их количество и порядок следования – индивидуальны.

Таким образом, адаптивное тестирование:

- дает более объективную оценку знаний, умений и навыков обучаемых;
- позволяет выявлять, какие знания ошибочны или неполны;
- позволяет давать рекомендации для дальнейшего

построения образовательного процесса.

Благодаря развитию теории адаптивного тестирования стала возможной адаптация не только тестовых заданий, но и тестирующих систем направленная на:

- приспособление к предметной области, выбранной для тестирования;
- приспособление к текущим потребностям конкретного испытуемого;
- приспособление к текущему состоянию конкретного испытуемого.

Литература

- [1] Скобцов Ю.А. Основы эволюционных вычислений. – Донецк: ДонНТУ, 2008. - 20 с.
- [2] Бондаренко М.Ф., Семенец В.В., Белоус Н.В., Куцевич И.В., Белоус И.А. «Оценивание тестовых заданий разных типов и определение их уровня сложности» [Электронный ресурс] // «Искусственный интеллект», Национальная библиотека Украины им. В.И.Вернадского. – К., 2009г. – Электронный ресурс. Режим доступа: http://www.nbuv.gov.ua/portal/natural/ii/2009_4/7%5C00_Bondarenko_Semenets_Belous_Kutsevich_Belous.pdf – 10.11.2009г.
- [3] Кромер В.В., «Об одной трехпараметрической модели тестирования» [Электронный ресурс] // Cornell University Library – Электронный ресурс. Режим доступа: <http://arxiv1.library.cornell.edu/ftp/cs/papers/0506/0506057.pdf> – 03.09.2009г.
- [4] Использование Теории тестовых заданий (Item Response Theory) в адаптивном тестировании. [Электронный ресурс] // Викизнание – Режим доступа: [http://www.wikiznanie.ru/ru-wz/index.php/Использование_Теории_тестовых_заданий_\(Item_Response_Theory\)_в_](http://www.wikiznanie.ru/ru-wz/index.php/Использование_Теории_тестовых_заданий_(Item_Response_Theory)_в_)

адаптивном_тестировании – 19.12.2009г.

- [5] Попов А.П. «Критический анализ параметрических моделей Раша и Бирнбаума» [Электронный ресурс] // Ростовский государственный педагогический университет – Электронный ресурс. Режим доступа: <http://www.ast-centre.ru/books/favorites/273/> – 20.09.2009г.
- [6] Koun-Tem Sun*, Yi-Chun Lin, Yueh-Min Huang «An efficient genetic algorithm for item selection strategy» [Электронный ресурс] // Dep. of Information and Learning Technology, National University of Tainan*, Dep. of Engineering Science, National Cheng Kung University – Режим доступа: <http://dspace.lib.fcu.edu.tw/bitstream/2377/3728/1/ce07ics002006000254.pdf> – 25.08.2010г.