

Методы и модели анализа качества тестовых заданий и моделирование компьютерного адаптивного тестирования в системах дистанционного обучения

*Национальный аэрокосмический университет им. Н. Е. Жуковского
«Харьковский авиационный институт»*

Статья посвящена обзору современных моделей и методов, используемых для анализа тестов в системах дистанционного обучения. Представлено описание теоретических подходов к оцениванию результатов тестирования. Рассмотрены математические модели современной теории тестирования (IRT), описывающие взаимосвязь латентных параметров обучаемых и заданий тестов. Приведено описание общих принципов компьютерного адаптивного тестирования с использованием дихотомических и полиномических заданий. Рассмотрен пример работы алгоритма компьютерного адаптивного тестирования с учетом различных начальных условий, условий выбора тестовых заданий и прекращения тестирования. В качестве программного инструментария предложено использовать пакеты программного инструментария для обработки статистических данных R.

Ключевые слова: современная теория тестирования, дистанционное обучение, компьютерное адаптивное тестирование, массовые открытые онлайн-курсы, качество теста.

Введение

В течение последних трех лет наблюдается широкое развитие систем дистанционного обучения и, в частности, стремительный рост рынка массовых открытых онлайн-курсов (МООК). МООК как системы дистанционного образования являются моделями для передачи образовательного контента онлайн к любому человеку, который хочет пройти обучение дистанционно, без ограничения по посещаемости. К основным новшествам таких систем можно отнести формат подачи теоретического материала (короткие видео), способы оценивания усвоенного материала (автоматический либо с помощью партнера, путем самостоятельного оценивания); постоянный доступ к контенту из ведущих высших учебных заведений мира, а также возможность общения через форумы, чаты и многие другие средства современных информационных технологий.

Количество зарегистрированных и текущих участников на платформах МООК огромно. Тем не менее, число пользователей, проходящих курс обучения до конца, значительно меньше. Исследованию факторов, которые влияют на процент студентов, завершивших курс, посвящено ряд работ [1, 2]. Одной из причин такой тенденции был назван некачественный механизм оценивания результатов. Предполагается, что разработчики курса должны тщательно рассмотреть вопрос о том, чтобы использовать в качестве оценивания полученных знаний автоматизированное тестирование с элементами интеллектуального оценивания, так как пользователями курсов являются люди с разными контекстуальными характеристиками, уровнем образования и культурными традициями. Тестовые задания, направленные на контроль полученных знаний, умений и навыков, должны не только соответствовать

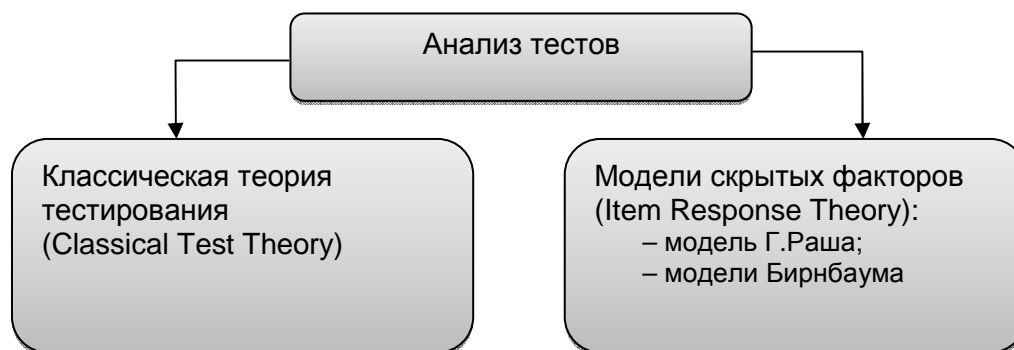
требованиям и целям обучения, но и учитывать темпы усвоения материала обучаемым и его индивидуальные способности. В противном случае обучаемый не может справиться с заданиями и у него пропадает интерес и мотивация к обучению.

Целью данного исследования является обзор методов, моделей и компьютерных средств, которые используются при оценивании уровня подготовки обучаемых в системах дистанционного обучения, в частности, методов адаптивного тестирования.

1. Анализ современных методов оценивания результатов тестирования

Контроль знаний и уровня подготовки обучаемого в дистанционных системах обучения чаще всего осуществляется с использованием технологий педагогического тестирования. Объективность, надежность и устойчивость оценок, получаемых в ходе тестирования, зависит, прежде всего, от качества разработанных тестов, так как именно тест и тестовые задания выступают инструментом оценивания латентных (скрытых) характеристик обучаемого.

В настоящее время оценивание результатов тестирования проводится с помощью математического аппарата классической (Classical Test Theory) или



современной теории тестирования (Item Response Theory - IRT) (рис. 1)

Рис. 1. Теоретические подходы к анализу тестов

Относительно достоинств и недостатков каждого подхода мнения исследователей различаются [3, 4]. Основным отличием классической теории от современной является подход к оцениванию обучаемого: результат – это итоговый балл по конкретному тесту с поправкой на ошибку. По сути в классической теории тестовые баллы испытуемых зависят от трудности заданий в тесте, а трудность задания – от данной выборки испытуемых. С другой стороны, стоит отметить прозрачность и ясность выводов при использовании классической теории. Кроме того, для получения достоверных результатов необходима малая выборка (100 или меньше) испытуемых [5].

В случае методов IRT итоговый балл рассматривается как результат совокупного взаимодействия скрытых параметров, а именно истинного уровня подготовки обучаемых и сложности рассмотренных заданий [3]. Современная теория тестов IRT основана на теории латентно-структурного анализа – одного из основных методов, применяя который можно определить латентный признак, т.е. оценить уровень способностей обучаемого на основе разработанных тестовых заданий, различных по уровню сложности, что определяется эмпирическим путем в ходе тестирования. Оцениваемый скрытый параметр должен быть одномерным

(один тест – один фактор – тест на интеллект). Существует два класса латентных параметров, от которых зависит вероятность правильного ответа на тестовое задание, – уровень подготовки испытуемого и уровень сложности задания.

Ответы индивидуума не зависят друг от друга. Вероятность того, что при прохождении теста экзаменуемый даст конкретную последовательность ответов, равна произведению вероятностей ответов на отдельные задания. Наиболее весомое отличие IRT от классической теории тестов заключается в следующем: в IRT пренебрегают решением проблемы эмпирической валидности и надежности теста. Тест является однофакторным, соотносится только с одним свойством и считается априорно валидным. В классической теории тестов уровень свойства (или же индивидуальный балл) является некоторым постоянным значением. В современной же теории тестов скрытый параметр рассматривается как непрерывная переменная. Кроме сложности задания и уровня подготовки испытуемого в модель IRT могут включаться и другие параметры. Все модели IRT различаются по количеству используемых в них переменных. Наиболее известные – это однопараметрическая модель Раша, двухпараметрическая и трехпараметрическая модели Бирнбаума.

Дихотомическая модель Раша представляет собой простую связь между тестируемым и заданиями [6]. Каждый тестируемый характеризуется уровнем способности, выраженным в виде числа вдоль бесконечной линейной шкалы соответствующей способности. Как и для физических измерений, для удобства выбирается локальное начало шкалы.

Связь между уровнем способности тестируемого и сложностью задания выражается дихотомической моделью Раша

$$\ln\left(\frac{P_1}{P_0}\right) = B - D, \quad (1)$$

где P_1 – вероятность того, что тестируемый с уровнем способности B ответит правильно на задание сложностью D ; P_0 – вероятность того, что он ответит неправильно.

Единица измерения интервальной шкалы, построенной по этой модели, называется логитом. Расстояние в логитах вдоль одномерной шкалы измерений между тестируемым, который имеет 50% шансов ответить верно на тестовое задание ($\ln(50\%/50\%)=0$), т.е. уровень способности находится в той же точке на шкале, что и уровень сложности задания, и тестируемым, который имеет 75% шансов ответить верно ($\ln(75\%/25\%)=1.1$), будет равняться 1.1 логит.

Выразив из уравнения (1) вероятность правильного ответа, получим

$$P_1 = \frac{e^{B-D}}{1 + e^{B-D}}. \quad (2)$$

Подставив коэффициент 1,7, который используют для совместимости модели Раша и модели Фергюсона [7, 8], где вероятность правильного ответа выражена интегралом нормального распределения, получим

$$P_{nil} = \frac{e^{1,7*(B_n - D_i)}}{1 + e^{1,7*(B_n - D_i)}}. \quad (3)$$

Индекс n соответствует идентификационному номеру тестируемого; i – идентификационному номеру задания.

Представленная зависимость (3) описывает вероятность правильного ответа как функцию одного параметра, еще её называют однопараметрической моделью IRT. Корректное использование модели Раша позволяет отделить оценки уровней подготовленности испытуемых от оценки уровней сложностей заданий.

Однако однопараметрическая модель Раша не может быть применена в том случае, если тест содержит задания с различной дифференцирующей способностью или дискриминативностью (способность заданий теста выявлять сильных и слабых учащихся, дифференцировать испытуемых по их подготовке). Решение данной проблемы предложил А. Бирнбаум, который ввел еще один параметр – a :

$$P_{ni1} = \frac{e^{1,7*a_i*(B_n-D_i)}}{1 + e^{1,7*a_i*(B_n-D_i)}} \quad (4)$$

Параметр a_i определяет дифференцирующую способность i - го задания.

Для еще более точного соответствия эмпирическим данным Бирнбаум ввел третий параметр – коэффициент угадывания c_i :

$$P_{ni1} = c_i + (1 - c_i) \frac{e^{1,7*a_i*(B_n-D_i)}}{1 + e^{1,7*a_i*(B_n-D_i)}} \quad (5)$$

Также существует ещё четырёх параметрическая модель, которая включает в себя дополнительно параметр «невнимательности», – d_i :

$$P_{ni1} = c_i + (1 - c_i - d_i) \frac{e^{1,7*a_i*(B_n-D_i)}}{1 + e^{1,7*a_i*(B_n-D_i)}} \quad (6)$$

Однако расчет всех параметров четырех параметрической модели является достаточно сложной задачей. Если для модели Раша получен последовательный алгоритм нахождения параметров модели [5, 9], то для остальных моделей для нахождения параметров используют численные методы на основе метода максимального правдоподобия, что при большом объеме данных требует значительных вычислительных ресурсов [10, 11]. Кроме того, в [12, 13] предложены сплайн-модели для оценки параметров модели IRT, которые дают более точные оценки, но разработаны только для двухпараметрической модели Бирнбаума.

На рис. 2 показана характеристическая кривая четырех параметрической модели IRT, где a – точка, в которой вероятность правильного ответа на задание, характеристическая кривая которого изображена на рис. 2, равна 0,5. Иными словами, сложность задания равняется уровню подготовки тестируемого; b – асимптота, угол наклона которой относительно оси x отображает дифференцирующую способность представленного вопроса; c – асимптота, характеризующая коэффициент угадывания; d – асимптота, характеризующая коэффициент невнимательности.

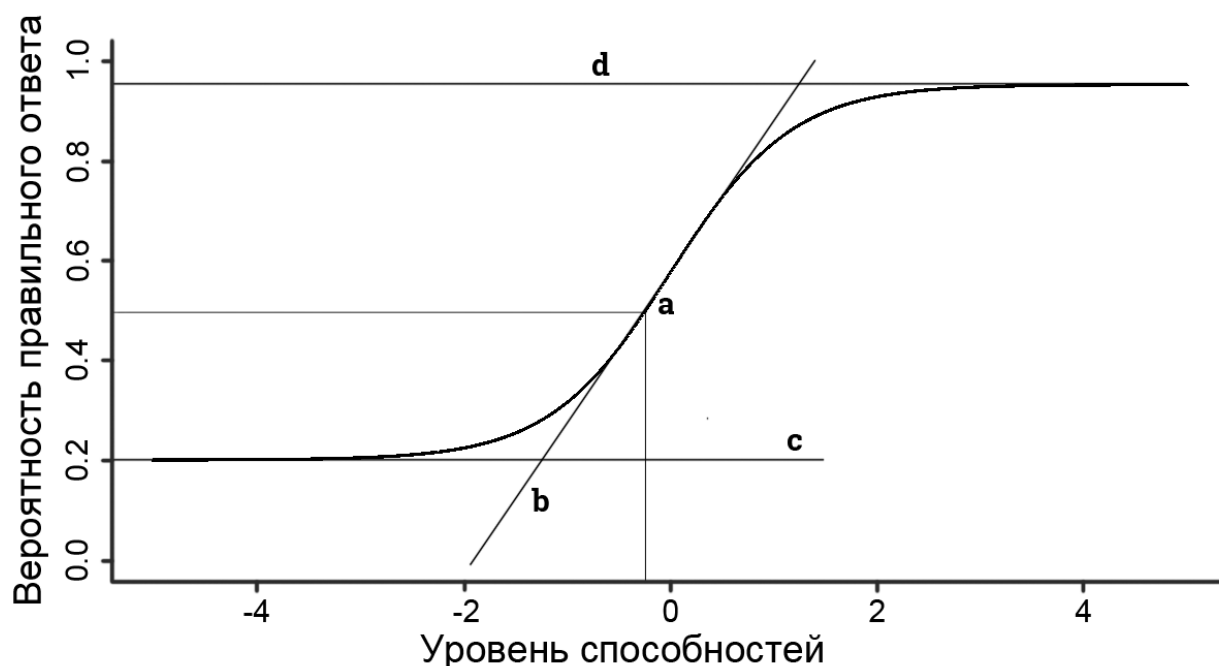


Рис. 2. Характеристическая кривая четырех параметрической модели

2. Общие принципы компьютерного адаптивного тестирования

Компьютерное адаптивное тестирование (КАТ) предполагает автоматизированный процесс оценивания уровня подготовки испытуемых в системах дистанционного обучения. КАТ максимально адаптировано уровню способностей обучаемого и предполагает выбор тестовых заданий определенной сложности в процессе проведения тестирования. Рассмотрим общий принцип реализации КАТ.

Прежде всего необходимо подготовить банк тестовых заданий, откалиброванных по уровню сложности. Наиболее распространённым вариантом является банк дихотомических заданий с несколькими вариантами ответа. Каждое задание имеет сложность, которая заранее рассчитана согласно моделям IRT. Рассмотрим в качестве примера тестовое задание по геометрии. В общем случае диапазон измерения уровня сложности тестового задания для любого теста принимается в пределах от -5 до 5 логитов. Предположим, что утверждение «квадрат гипотенузы равен сумме квадратов катетов» имеет сложность 1 логит. Тестируемые, для которых определение данного утверждения является простым, имеют уровень способности выше, чем 1 логит. Тестируемые, для которых определение является слишком сложным, имеют уровень способности ниже, чем 1 логит. Тестируемые, которые с вероятностью 50% правильно дадут определение, имеют предполагаемый уровень способности 1 логит, т.е. равный сложности задания. Это задание называется целевым для таких тестируемых.

На следующем этапе важно определить, как будут назначаться задания в процессе КАТ. Перед началом основного тестирования тестируемому предлагается несколько предварительных практических заданий в присутствии администратора теста (преподавателя), чтобы убедиться, что тестируемый понимает, как правильно работать с компьютерным тестом. Далее либо администратор, либо автоматически по заранее разработанному алгоритму

назначается предполагаемый начальный уровень способности тестируемого (зачастую принимается среднее значение уровня подготовки).

Выбор первого вопроса не критичен для измерения, но он может иметь решающее значение для психологического состояния испытуемого. Если предложить испытуемому слишком трудное задание, то он может даже не попытаться решить его и почувствовать неуверенность в ходе тестирования. Если же предложить на первом шаге слишком легкое задание, тестируемый может отнестись к его решению несерьезно и совершить ошибку по невнимательности. Поэтому рекомендуется, чтобы задания были немного ниже предполагаемого уровня способности испытуемого.

Если же заранее есть информация о предполагаемом уровне способности обучаемого, то стартовое задание должно иметь сложность, соответствующую уровню подготовки испытуемого. В этом случае задание будет скорее всего решено правильно и не будет допущена ошибка по невнимательности.

Общий процесс КАТ можно представить в виде диаграммы (рис. 3), где по оси абсцисс приведены номера заданий, а по оси ординат – уровень сложности (в логитах).

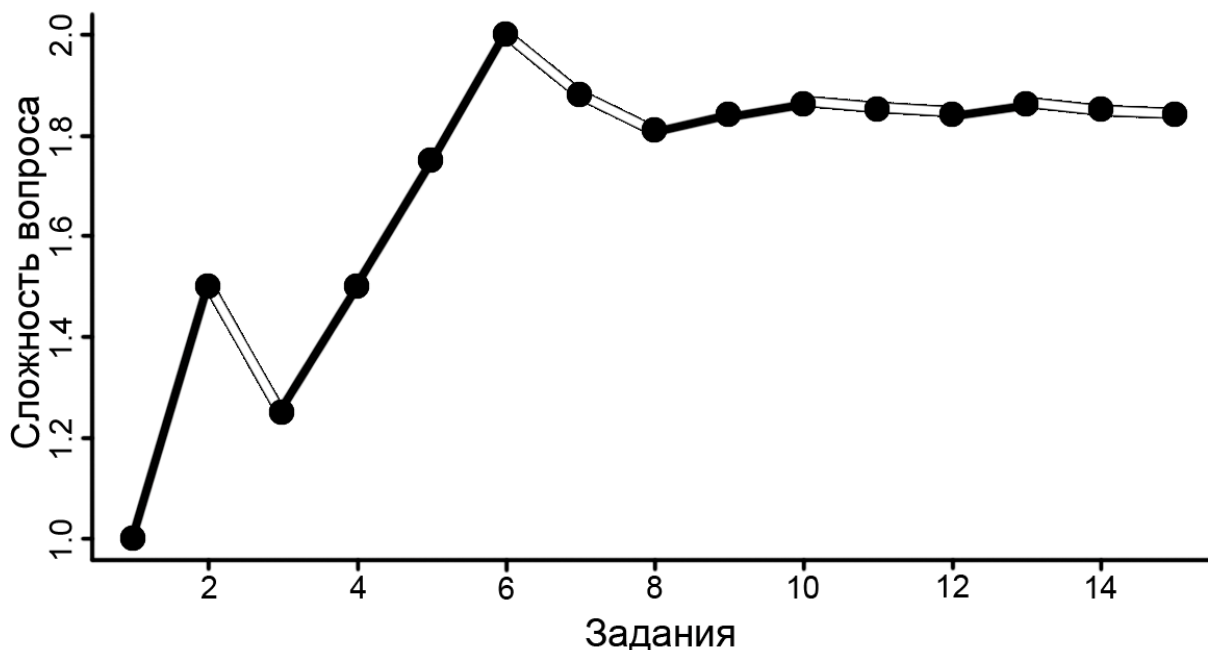


Рис. 3. КАТ-назначение дихотомических заданий:

- а) черный маркер – респондент дал верный ответ на задание;
- б) белый маркер – респондент дал неверный ответ на задание

В нашем примере (см. рис. 3) предположим, что первое задание, которое предлагается испытуемому, имеет уровень сложности 1 логит. Тестируемый скорее всего решит такое задание. Далее автоматически выбирается более сложное задание, например, с уровнем 1,5 логита. Тестируемый его не решает. Поэтому испытуемому предлагается задание с уровнем сложности примерно 1,25 логита. На этот раз тестируемый успешно справляется с заданием. Опять назначается задание с уровнем сложности 1,5 логита. Тестируемый решает его правильно. Как и на предыдущем шаге, испытуемому предлагается задание

сложнее (1,75 логита). Процесс продолжается до того момента, пока не будет получена оценка уровней способностей испытуемого с заданной точностью. Если же испытуемый все время то верно, то неверно решает задания, то точкой останова тестирования может быть максимальный предел задаваемых тестовых заданий. В общем случае чем больше заданий будет назначено, тем более точной будет оценка уровня способности. Однако тест может быть пройден и за минимальное количество шагов (количество определяет администратор теста), если испытуемый в процессе тестирования все время правильно решает тестовые задания.

Для того, чтобы использовать тот или иной тест в процессе КАТ, необходимо выполнение следующего условия: тест должен обладать свойством одномерности, т.е. за одно тестирование могут быть определены латентные параметры лишь в одной области знаний; банк тестовых заданий должен быть откалиброван по уровню сложности на основе моделей IRT; заданы все условия по выбору тестовых заданий, условий останова и уровню точности оценивания.

В случае, когда ответ на вопрос в тесте не может быть определён, как «верно» или «неверно», применяют политомические задания. В этом случае каждый вариант ответа на тестовое задание имеет свой определенный вес, который характеризует степень близости данного ответа к правильному ответу. Как правило, тестовые задания с множественными ответами, задания на соответствие, с открытым числовым решением являются политомическими заданиями.

Общий алгоритм КАТ показан на рис. 4. Процесс КАТ требует откалиброванного банка тестовых заданий (БТЗ) и может быть разделен на 4 основных этапа. Первый этап является начальным (initial step) и заключается в выборе одного или нескольких соответствующих заданий в качестве первых заданий в процессе тестирования. Вторым этапом является непосредственное тестирование (test step), который заключается в том, что задания последовательно выбираются из банка тестовых заданий и повторно оценивается уровень способности испытуемых после каждого ответа. Третий этап – это этап останова теста (stopping test), где определяются правила останова КАТ. На заключительном этапе (final step) дается окончательная оценка способностей испытуемого и, возможно, другая информация об обучаемом.

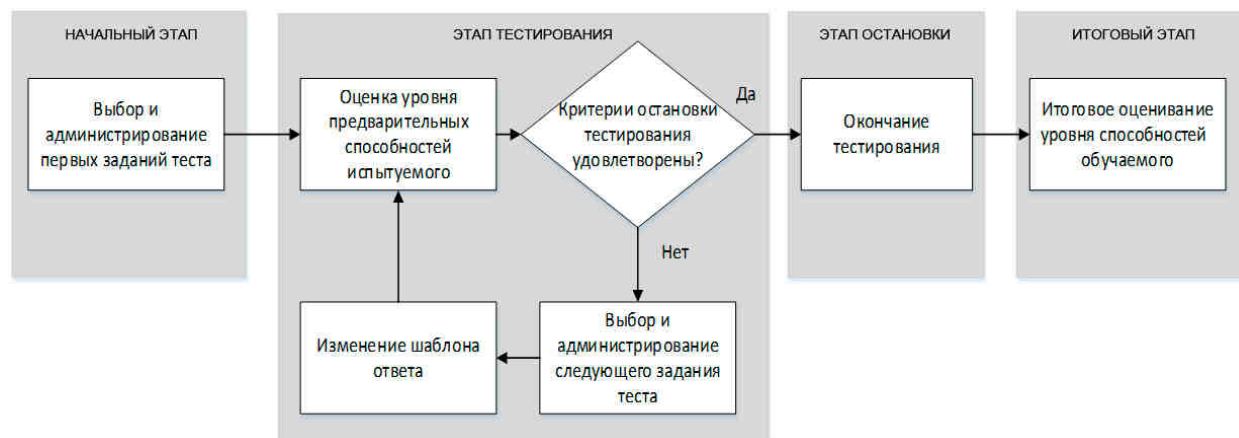


Рис. 4. Схема процесса КАТ

Непосредственно адаптивное тестирование включает в себя следующие этапы:

1. Оценивается уровень способности на основе имеющейся в настоящее время информации (предыдущего ответа на задание). Устанавливается предварительная оценка уровня способностей участника тестирования.

2. Выбирается следующее задание из БТЗ, которое еще не было отобрано для испытуемого в соответствии с методом отбора и параметрами задания (трудностью, дискриминативностью и т.д.).

3. Проводится администрирование выбора тестового задания для испытуемого, т.е. обновляется шаблон ответа (фиксируется, верно или не верно участник тестирования ответил на вопрос).

4. Шаги 1-3 повторяются.

Наиболее важным шагом данного алгоритма является определение методов оценивания уровня способностей испытуемого. Методы оценивания базируются на различных моделях и критериях оценивания. Среди методов можно выделить метод максимального правдоподобия (ML estimator), когда шаблон ответов предполагает наличие только двух событий – верного и неверного ответа на тестовое задание, и метод, базирующийся на предположении нормального распределения ответов (BM estimator) для остальных типов заданий. В процессе выбора следующего задания также могут использоваться другие эвристические алгоритмы, базирующиеся на теории байесовских сетей, нечеткой логики и др.

Существуют различные критерии выбора следующего тестового задания, например критерий максимальной информативности (maximum information criterion – MFI), минимальной ожидаемой последующей дисперсии (minimum expected posterior variance MEPPV), критерий максимального правдоподобия взвешенной информации (maximum likelihood weighted information – MLWI), максимальной последующей взвешенной информации (maximum posterior weighted information – MPWI), максимально ожидаемой информации (maximum expected information – MEI), комбинированного случайного выбора. Подробно эти критерии описаны в [14].

Этап завершения адаптивного тестирования предполагает использование одного из трех правил: достигнута максимальная длина теста (the length criterion), достигнута требуемая величина точности оценивания (precision criterion) – испытуемый устойчиво отвечает на задания данного уровня сложности или участник тестирования классифицирован (classification criterion).

Максимальную длину теста задает администратор, т.е. исходя из типа теста и целей тестирования, требований к точности оценивания определяется оптимальное количество тестовых заданий. Слишком малое количество не даст достоверной оценки уровня способностей испытуемого, тогда как и слишком большой тест может внести ошибки по невнимательности, так как участники тестирования просто могут устать.

Требуемая величина точности определяется на основе стандартной ошибки оценки уровня способностей. Чем ниже значение этой ошибки, тем лучше. Пороговое значение также задается администратором теста в зависимости от вида теста и целей тестирования.

Критерий классификации используется тогда, когда целью тестирования является классификация участников, а не оценивание их способностей, т.е. если

испытуемый достиг некоторого порогового значения способностей и нет необходимости продолжать тестирование.

На заключительном этапе тестирования проводится итоговое оценивание уровня способностей испытуемого на основе полученного шаблона ответов адаптированного теста. Рассчитывается стандартная ошибка и результаты оценивания могут отображаться в выбранной шкале оценивания, например, не в логитах, а в шкале 0...100 баллов. На заключительном этапе также могут использоваться различные методы оценивания параметров модели (ML estimator или BM estimator).

3. Моделирование КАТ с помощью пакетов ltm и catR программного инструментария обработки статистических данных R

Наряду с теоретическими исследованиями существует большое количество программных решений и аппаратных средств для реализации тестового контроля знаний и оценки качества тестов и тестовых заданий. Однако далеко не все включают в себя системы адаптивного компьютерного тестирования. В основном программные решения, такие, как ITEMAN, WINSTEPS, XCALIBRE, RASCAL, позволяют оценить качество теста и тестовых заданий согласно классической и/или адаптивной теорий тестирования [15]. Такие программные продукты, как OpenTEST2, TCExam, iTest, Moodle, успешно решают задачи компьютерного тестирования [16, 17]. Однако решения в области реализации адаптивных тестов на рынке Украины практически отсутствуют. Поэтому, в качестве программного инструментария для оценки качества и формирования банка тестовых заданий, а также моделирования процесса КАТ был выбран программный продукт R и набор библиотек ltm и catR. Такой выбор прежде всего обусловлен доступностью программного обеспечения, поскольку среда R является свободной программной средой вычислений с открытым исходным кодом. Особенностью R является также возможность интегрирования в другие программные среды (java, c++, c#)

Рассмотрим пример оценки качества дихотомических тестовых заданий и моделирования процесса КАТ в среде R. Ниже приведен скрипт программного кода для оценки параметров тестовых заданий и построения характеристических кривых в среде R, а также простой пример моделирования процесса КАТ на основе созданного банка тестовых заданий.

Оценка качества тестовых заданий на основе моделей IRT

```
#Настройка опций, подключение служебных библиотек и считывание и
отображение данных
options(digits=3,scipen = 10) # установка формата вывода данных
library(foreign)
Dataset1 <- read.spss("D:/data1.sav",use.value.labels=TRUE,
max.value.labels=Inf, to.data.frame=TRUE, reencode = TRUE)
View(Dataset1)
datasetdich<-Dataset1[60:101]
View(datasetdich)

# Подключения библиотеки для расчета параметров модели IRT
install.packages('ltm')
library("ltm", lib.loc="~/R/win-library/3.2")
descript(datasetdich)
```

```

#Задание исходных данных для модели Раша, расчет основных
коэффициентов
fit1<-rasch(datasetdich,constraint =
cbind(length(datasetdich)+1,1))
summary(fit1)
coef(fit1,prob=TRUE,order=TRUE)

#Проверка точности полученной модели
margins(fit1)
factor.scores(fit1)
residuals(fit1)

#Задание исходных данных для модели Раша (с поправкой 1,7),
расчет основных коэффициентов
fit2<-rasch(datasetdich)
summary(fit2)
coef(fit2,prob=TRUE,order=TRUE)

#Проверка точности полученной модели
margins(fit2)
factor.scores(fit2)
residuals(fit2)
anova(fit1,fit2)

#Задание исходных данных для модели Бирнаума, расчет основных
коэффициентов
fit3<-ltm(datasetdich ~ z1)
summary(fit3)
coef(fit3,prob=TRUE,order=TRUE)

#Проверка точности полученной модели
margins(fit3)
factor.scores(fit3)
anova(fit2,fit3)

#Оценка информационных кривых
information(fit1,c(-4,4))
information(fit2,c(-4,4))
information(fit3,c(-4,4))

#Построение графиков характеристических кривых
dev.new()
plot(fit1)
dev.new()
plot(fit2)
dev.new()
plot(fit3)
dev.new()

```

В результате работы данной программы были рассчитаны сложности тестовых заданий теста и оценены параметры моделей. На рис. 5 и 6 показаны характеристические кривые тестовых заданий по модели Раша и двухпараметрической модели Бирнбаума. Как видно, в тесте присутствуют вопросы различного уровня сложности с различной дифференцирующей способностью.

Например, по своим характеристикам вопросы 20 и 11 являются достаточно сложными с неудовлетворительной дискриминативностью. Их рекомендуется не включать в банк тестовых заданий, поскольку они не позволят получить достоверные оценки относительно уровня подготовки обучаемых.

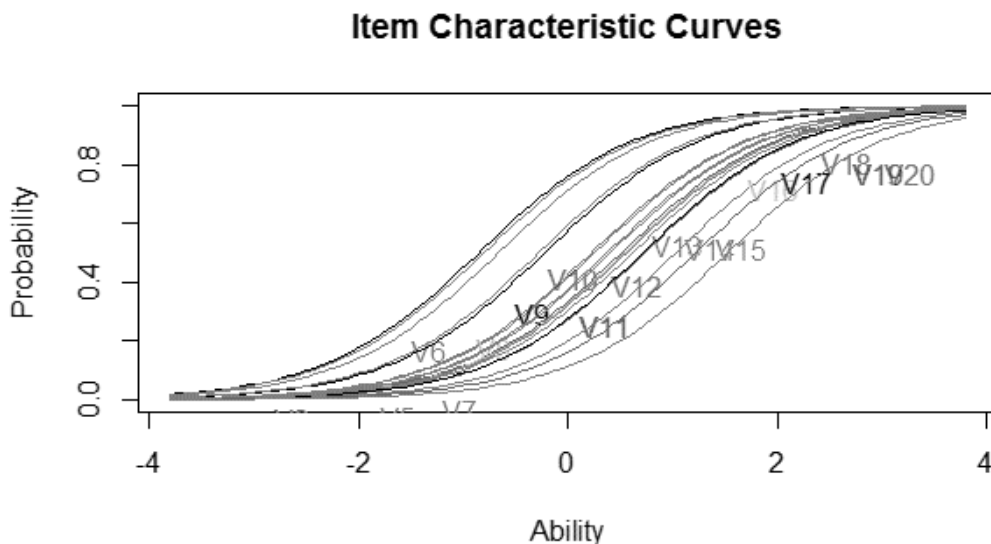


Рис. 5. Характеристические кривые заданий теста однопараметрической модели Раша

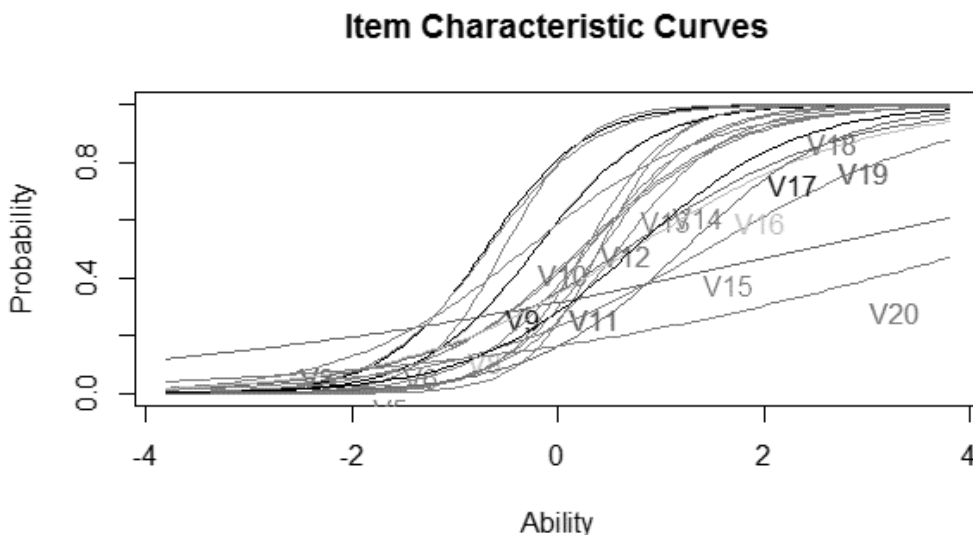


Рис. 6. Характеристические кривые заданий теста двухпараметрической модели Бирнбаума

Далее приведен пример программного кода работы с пакетом `catR`, который позволяет промоделировать процесс КАТ. Результаты моделирования показаны на рис. 7. Подробное описание настройки пакетов можно найти в [18].

Моделирование процесса КАТ

```
## Подключения библиотеки для моделирования КАТ
install.packages("catR")
library("catR")
require(catR)

# Настройка параметров КАТ
cbList <- list(names = c("Group1", "Group2", "Group3",
"Group4"), props = c(0.2,0.4,0.3,0.1))
mybank<-genDichoMatrix(items = 500, cbControl = cbList)
View(mybank)
mybank1<-genDichoMatrix(items = 500)
hist(mybank$b, breaks=30,main="Distribution of Item
Difficulties", xlab="b parameter")
bank <- genPolyMatrix(items = 200, nrCat = 5,model = "PCM")
start <- list(nrItems = 3, seed = 1)
test <- list(method = "WL", itemSelect = "MFI")
stop <- list(rule = "length", thr = 10)
final <- list(method = "ML")

# Моделирование процесса КАТ
res <- randomCAT(trueTheta = 0, itemBank = bank,
model = "PCM", start = start,
test = test, stop = stop,
final = final)

# Построение графика

plot(res, ci = TRUE, trueTh = TRUE, classThr = 2)
```

Из графика видно, что по мере выполнения тестовых заданий стандартная ошибка уменьшается, а уровень способностей стремится к истинному установленному значению, т.е. к единице. Когда верхнее значение доверительного интервала становится меньше двух, тестирование прекращается, т.е. выполняется правило останова.

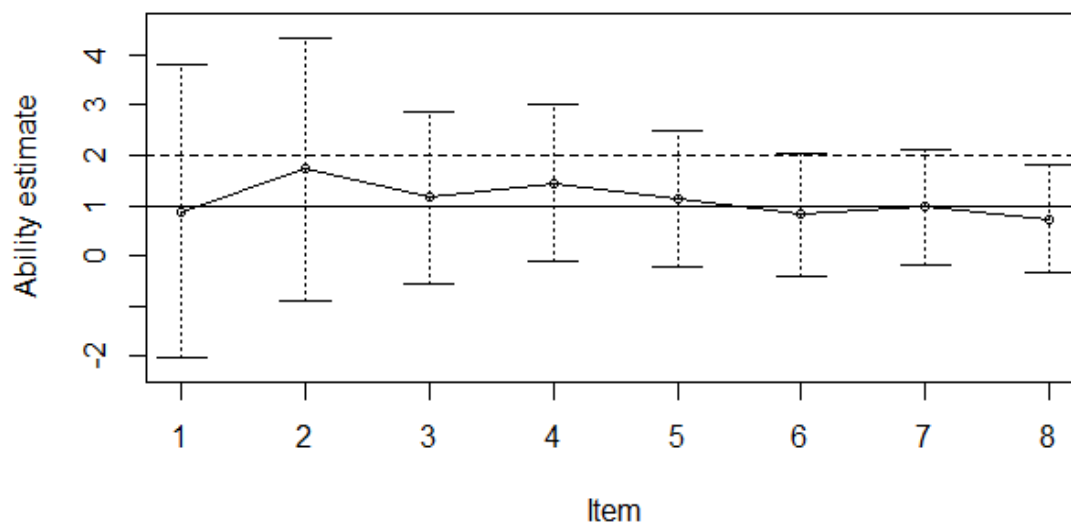


Рис. 7. График результатов моделирования КАТ: 1 – истинный уровень способностей испытуемых; 2 – порог классификации

Выводы

Широкое внедрение в систему образования компьютерного оценивания знаний (повсеместное использование дистанционного обучения) привело к необходимости качественного и быстрого мониторинга результатов подготовки обучаемых. С одной стороны, полученные данные должны быть легко интерпретируемыми, с другой – применяемые подходы должны отвечать всем необходимым критериям качества (задания должны быть откалиброваны, тесты – валидными и надежными, сложность задания – максимально отвечать уровню подготовки респондента и т.д.).

В данной работе рассмотрены математические модели современной теории тестирования, так как эти методы считаются более мощными, чем методы, основанные на классической теории тестирования. Описаны основные характеристики данных моделей. Важным преимуществом IRT является то, что оценки могут быть использованы для адаптации процесса тестирования с точки зрения способностей обучаемого. Такой подход используется в компьютеризированном адаптивном тестировании.

В работе приведен общий алгоритм работы КАТ, а также предложен вариант использования пакетов программной среды R для оценки качества тестовых заданий на основе моделей IRT и моделирования процесса catR.

Список литературы

1. Jordan, K. Massive Open Online Course Completion Rates Revisited: Assessment, Length and Attrition [Electronic resource] // The international review of research in open and distributed learning.– 2015. – Way of access: <http://www.irrodl.org/index.php/irrodl/article/view/2112/3340> June, 2015. – Title from the screen.
2. Harvard Magazine. What modularity means for MOOCs. [Electronic resource] // Harvard Magazine, – 2013. – Way of access: <http://harvardmagazine.com/2013/12/harvard-mit-online-education-views-changing>. –05.12.2015. – Title from the screen.
3. Wiberg, M. Classical test theory vs. item response theory [Electronic resource] // EM No 50, - 2004. Way of access: http://www.edusci.umu.se/digitalAssets/59/59529_em-no-50.pdf. – Title from the screen
4. Ronald, K. Comparison of classical test theory and item response theory and their applications to test development // Educational Measurement: issues and practice. – 1993. – Way of access: http://www.internationalgme.org/Resources/Pubs/ITEMS_Module_16.pdf – Title from the screen.
5. Loken, E. Estimation of a Four-Parameter Item Response Theory Model [Text] / E. Loken, K. L. Rulison // British Journal of Mathematical and Statistical Psychology. – 2010. – № 63. – P. 509–525.
6. Крокер, Л. Введение в классическую и современную теорию тестов [Текст] : учеб. / Л. Крокер, Дж. Алгина., пер. с англ. И. Н. Найденовой, В. Н. Симкина, М. Б. Челышковой / под общ. ред. В. И. Звонникова, М. Б. Челышковой. – М.: Логос, 2010. – 668 с.
7. Ким, В. С. Тестирование учебных достижений [Текст]: моногр. / В. С. Ким. – Уссурийск: Изд-во УГПИ, 2007. – 214 с.
8. Дубан, Р. Н. Построение профилей IRT в тестовом контроле знаний [Текст] / Р. Н. Дубан // Актуальные проблемы гуманитарных и естественных наук. – М. – 2014. – № 03(62). Ч. 1. – С. 52 – 56.

9. Моделі та методи сучасної теорії тестів: [навч.-метод. посібник] / Т. В. Лісова. – Ніжин: Видавець ПП Лисенко М.М., 2012. — 112 с.
10. Тестирование как форма контроля знаний учащихся [Электронный ресурс] / Дистанционный курс «Электронные тесты». – Режим доступа: \www/ URL: <http://elektest.narod.ru/p7aa1.html> – 10.05.2016 г. – Загл. с экрана.
11. Зайцева, Л. В. Модели и методы адаптивного контроля знаний [Текст] / Л. В. Зайцева, Н. О. Прокофьева // Educational Technology & Society. – Nr. 7(4), 2004 ISSN 1436-4522 (Международный электронный журнал) [Электронный ресурс]. – Режим доступа: http://ifets.ieee.org/russian/depository/v7_i4/html/1.htm. – С. 265-277.
12. Дубан, Р. М. Слайн-модели профилей сложности вопросов и знаний респондентов в тестовом контроле знаний [Текст] / Р. М. Дубан, І. В. Шелевицький // Автоматизированные системы управления и приборы автоматики: всеукр. межведомств. науч.-техн. сб. – Х., – 2011. – Вып. 156. – С. 71-77.
13. Дубан Р. М. Застосування інтегралу від лінійного В-сплайна в якості моделі IRT [Текст] / Р. М. Дубан, І. Ф. Бойко // Електроніка та системи управління. – К. – 2012. – № 1(31). – С. 131-138.
14. Ying, C. Computerized adaptive testing / C. Ying // New developments and applications, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN – 2008. 92 p.
15. The Adaptest® computerized adaptive testing and item response theory (IRT) [Electronic resource] Adaptest®. – Way of access: \www/ URL: <https://adaptest.vpgcentral.com/> – 20.05.2016 г. – Title from the screen.
16. Мясникова, Т. С. Система дистанционного обучения MOODLE [Текст] / Т.С. Мясникова, С.А. Мясников – Х., – 232 с.
17. Гильмутдинов, А. Х. Электронное образование на платформе Moodle [Текст] / А.Х. Гильмутдинов, Р.А. Ибрагимов, И.В. Цивильский // Казанский Государственный Университет, 2008. – 169 с.
18. catR: Procedures to Generate Patterns under Computerized Adaptive Testing [Electronic resource] CAT. – Way of access: \www/ URL: <http://CRAN.R-project.org/package=catR>. – 21.05.2016 г. – Title from the screen.
19. Прокофьев, Н. О. Модели и методы компьютерной оценки знаний обучаемых [Текст] / Н. О. Прокофьев // Материалы Междунар. научно-практ. Конф. «Информационные технологии в многоуровневой системе образования». – Казань: ЗАО «Новое знание», 2005. – С. 139–143.
20. Шостак, И. В. Критический анализ подходов к организации контроля учебно-познавательной деятельности в системах дистанционного обучения [Текст] / И. В. Шостак, А. С. Носиков // Радиоэлектронные и компьютерный системы. – 2011.– № 10(87). – С. 229–235
21. Шкиль, А. С. Методика оценивания в компьютерной системе тестирования знаний [Текст] / А. С. Шкиль, С. В. Чумаченко, С. В. Напрасник // Образование и виртуальность, 2002 : сб. науч. тр. 5-й Междунар. конф. – Харьков – Ялта: УАДО, 2003. – С. 340–345.
22. Лаптев, В. В. Учет времени при оценивании результатов автоматизированного контроля [Текст] / В. В. Лаптев, В. И. Сербин // Изв. Волгоградск. гос. техн. ун-та: межвуз. сб. науч. ст. № 11 (71) / ВолгГТУ. – Волгоград: ИУНЛ ВолгГТУ, 2010. – (Сер. Актуальные проблемы управления, вычислительной техники и информатики в технических системах. Вып. 9). – С. 102–105.

23. Попов, Д. И. Способ оценки знаний в дистанционном обучении на основе нечетких отношений [Текст] / Д. И. Попов // Дистанционное образование. – 2000. – № 6. – С. 100–109.

Поступила в редакцию 20.09.2016

Методи і моделі аналізу якості тестових завдань і моделювання комп'ютерного адаптивного тестування у системах дистанційного навчання

Стаття присвячена огляду сучасних моделей і методів, які використовують для аналізу тестів у системах дистанційного навчання. Наведено опис теоретичних підходів до оцінювання результатів тестування. Розглянуто математичні моделі, що описують взаємозв'язок латентних параметрів учнів і завдань тестів. Описані загальні принципи комп'ютерного адаптивного тестування з використанням дихотомічних і політомічних завдань. Розглянуто приклад алгоритму КАТ з урахуванням початкових умов, умов вибору завдання і припинення тестування. На прикладі вхідного тесту з математики наведено спосіб оцінювання аналізу якості тесту. Як програмний інструментарій запропоновано використовувати програмний продукт RStudio.

Ключові слова: сучасна теорія тестування, дистанційна освіта, комп'ютерне адаптивне тестування, масові відкриті онлайн-курси, якість тесту.

Methods and Models of Analysis of the Quality of Tests and Modeling of the Computer Adaptive Testing in Distance Learning Systems

The article provides an overview of current models and methods that are used to analyze test in distance education systems. The description of the theoretical approaches to the evaluation of the test results. The mathematical model describing the relationship of latent parameters of learners and items. The description of the general principles of computerized adaptive testing using dichotomous and polytomous items. An example of the CAT algorithm taking into account the initial conditions, select the items conditions and stop the testing. For example, the first test in mathematics given way to test the quality assessment analysis. As a software tool is proposed to use the software RStudio.

Key words: item response theory, distance learning, computerized adaptive testing, massive open online course, test quality.

Сведения об авторах:

Мазорчук Мария Сергеевна – к.т.н., доцент кафедры информатики Национального аэрокосмического университета им. Н. Е. Жуковского «Харьковский авиационный институт», Украина.

Добряк Виктория Сергеевна – к.т.н., ст. преподаватель кафедры информатики Национального аэрокосмического университета им. Н. Е. Жуковского «Харьковский авиационный институт», Украина.

Емельянов Павел Сергеевич – аспирант кафедры информатики Национального аэрокосмического университета им. Н. Е. Жуковского «Харьковский авиационный институт», Украина.