Экспериментальная психология
2017. Т. 10. № 3. С. 33—45
doi:10.17759/exppsy.2017100303
ISSN: 2072-7593
ISSN: 2311-7036 (online)

Experimental Psychology (Russia)
2017, vol. 10, no. 3, pp. 33—45
doi:10.17759/exppsy.2017100303
ISSN: 2072-7593
ISSN: 2311-7036 (online)

НОВЫЙ ПОДХОД К КОМПЬЮТЕРИЗИРОВАННОМУ АДАПТИВНОМУ ТЕСТИРОВАНИЮ

КУРАВСКИЙ Л.С.*, *ФГБОУ ВО МГППУ, Москва, Россия*
e-mail: l.s.kuravsky@gmail.com
АРТЕМЕНКОВ С.Л. **, *ФГБОУ ВО МГППУ, Москва, Россия,*
e-mail: slart@inbox.ru
ЮРЬЕВ Г.А. ***, *ФГБОУ ВО МГППУ, Москва, Россия,*
e-mail: g.a.yuryev@gmail.com
ГРИГОРЕНКО Е.Л. *,*** *МГУ имени М.В. Ломоносова, Москва, Россия;*
Хьюстонский Университет, Хьюстон, США,
e-mail: elena.grigorenko@yale.edu

Представлен новый подход к компьютеризированному адаптивному тестированию, модель предъявления заданий которого описывается с помощью марковских процессов с дискретными состояниями и дискретным временем. Этот подход опирается на обобщение модели Г. Раша и имеет существенные преимущества перед адаптивным тестированием на базе современной теории тестирования (IRT). Преимуществами разработанного подхода являются: учёт особенностей процесса выполнения предъявленных заданий, включая затраченное на них время; прогнозирование поведения испытуемых; возможность самообучения и улучшения характеристик модели в процессе тестирования; простая процедура идентификации модели, использующая доступные результаты наблюдений. Разработанная модель адаптивного тестирования обобщается на случаи политомических заданий и нескольких шкал измерений.

Ключевые слова: Марковские цепи, адаптивное тестирование, IRT, компьютеризированное адаптивное тестирование.

Введение

Процедуры тестирования все чаще используются во многих современных приложениях, требующих оценки поведения людей или машин. Согласно традиционным моделям, ос-

Для цитаты:
Куравский Л.С., Артеменков С.Л., Юрьев Г.А., Григоренко Е.Л. Новый подход к компьютеризированному адаптивному тестированию // Экспериментальная психология. 2017. Т. 10. №. 3. С. 33—45. doi:10.17759/exppsy.2017100303
Англоязычный вариант статьи см. по ссылке http://psyjournals.ru/exp/

* *Куравский Л.С.* Доктор технических наук, профессор, декан факультета информационных технологий ФГБОУ ВО МГППУ. E-mail: l.s.kuravsky@gmail.com
** *Артеменков С.Л.* Кандидат технических наук, профессор кафедры прикладной информатики факультета информационных технологий ФГБОУ ВО МГППУ. E-mail: slart@inbox.ru
*** *Юрьев Г.А.* кандидат физико-математических наук, зам. декана, доцент, факультет информационных технологий, ФГБОУ ВО МГППУ. E-mail: g.a.yuryev@gmail.com
**** *Григоренко Е.Л.* Доктор психологических наук, профессор, доцент кафедры психологии образования и педагогики факультета психологии, Московский государственный университет имени М.В. Ломоносова; профессор Центра изучения ребенка, Йельский университет, Университет Хьюстона (США). E-mail: elena.grigorenko@yale.edu

нованным на классической теории тестирования, для измерения уровня определенного умения или способности испытуемого настолько точно, насколько это возможно, эти процедуры должны включать большое количество заданий, затрудняющих использование тестов. Выход из этой ситуации обеспечивает компьютеризированное адаптивное тестирование (CAT), развитию которого в значительной мере способствовали современные технологии и вычислительные возможности, появившиеся наряду с разработкой современной теории тестирования (IRT). Использование IRT или других подходов CAT — это метод администрирования тестов и измерения латентных конструктов с помощью небольшого количества тестовых заданий и как можно более точно (Thompson, Weiss, 2011). Эти конструкты могут включать в себя: способности, отношения, знания, навыки, черты и другие соответствующие категории. Ниже термин «конструкт» будет использоваться в качестве замены любой из этих категорий.

Во время компьютеризированной процедуры адаптивного тестирования используется адаптивный принцип выбора заданий, в соответствии с которым трудность заданий, предлагаемых для выполнения, должна соответствовать оценкам уровней достижений испытуемых. Согласно IRT, этот подход дает лучшую дифференциацию испытуемых по уровню их достижений. Другое преимущество представляет многомерное компьютеризированное адаптивное тестирование (MCAT), также включающее более короткое время тестирования и более точную и эффективную оценку конструкта. MCAT объединяет теоретические и практические достижения, полученные в CAT, позволяет получить доступ к большему количеству интересуемых конструктов, не увеличивая нагрузку на инструмент в виде дополнительных пула заданий, и, в конечном итоге, имеет большую точность (т. е. низкую стандартную ошибку измерения) (Torre de la, Patz, 2005). Несмотря на все эти преимущества, доступные методы CAT и MCAT, основанные на IRT, довольно сложны для реализации и не учитывают ряд специальных параметров таких, как время, историю тестирования и т. д.

Однако основная проблема, связанная с CAT и MCAT на основе IRT (которые обычно сочетаются с оценками максимального правдоподобия), состоит в приблизительном равенстве вероятностей для неправильных и правильных решений, поскольку трудность выбранного задания должна соответствовать оценке уровня достижений испытуемого. Этот факт делает результаты тестирования зависимыми в основном от посторонних случайных факторов, которые не связаны с изучаемыми конструктами, тем самым девальвируя полученные выводы, которые становятся фиктивными. Поэтому разработка новых подходов к CAT и MCAT является актуальной проблемой.

Ниже представлен один из подходов к ее решению. Используемая модель представлена марковскими процессами с дискретными состояниями и дискретным временем (*цепями Маркова*). Особенностью данного подхода является определение трудностей заданий с использованием предельных распределений вероятностей пребывания в состояниях, полученных с помощью матриц вероятностей перехода. Преимущества представленного подхода по сравнению с адаптивным тестированием на основе теории IRT, использующей модель Г. Раша, заключаются в следующем:

1) оценка не выводится из локальных сопоставлений текущих измерений оценок и трудностей с использованием модели Г. Раша, но учитывает:

— всю наблюдаемую историю выполнения тестовых заданий, которая включает в себя распределение успешных и неуспешных решений заданий и их порядок;

— время, затраченное на выполнение тестовых заданий;

2) оценки основаны на прогнозируемых результатах в будущем, при условии, что время тестирования не ограничено, и они не используют локальные (то есть для определенного задания) сравнения, основанные на модели Г. Раша, которые могут быть неустойчивыми[1];

3) количество заданий, которые необходимо выполнить, существенно меньше;

4) выбранные трудности заданий связаны с историей выполнения тестовых заданий и не зависят напрямую от текущих приближенных оценок уровней достижений испытуемых;

5) существует возможность учитывать изменение трудности конструктов для испытуемого во время процедуры тестирования из-за усталости и других причин;

6) имеется возможность самообучения, которая приводит к улучшению характеристик модели адаптивного тестирования во время ее эксплуатации;

7) существует процедура идентификации модели, основанная на простых и доступных результатах наблюдений.

В то же время представленный подход может рассматриваться как расширение IRT, поскольку модель Г. Раша используется в качестве ее компонента.

Преимущества представленного подхода по сравнению с адаптивным тестированием на основе марковских моделей с непрерывным временем состоят в следующем:

1) вместо марковских процессов с непрерывным временем используются марковские процессы с дискретным временем (Куравский и др., 2013; Куравский и др., 2011; Kuravsky et al., 2016; Kuravsky et al., 2015), однако в рассматриваемой адаптивной модели время, затрачиваемое на выполнение заданий, учитывается с помощью ограничений на время пребывания в состояниях и переходов в состояния-«ловушки»;

2) допустимы политомические задания.

Модель адаптивного тестирования

Общее описание модели

Используются *марковские процессы с дискретными состояниями и дискретным временем (цепи Маркова)*, при этом *вероятности перехода* между их состояниями являются параметрами модели. Типичная структура, показанная на рисунке 1, представляет собой конечную цепочку из $2n + 2$ состояний, в которой переходы из состояния x_i ($i \neq 0$, $i \neq n$) возможны только в следующее состояние x_{i+1} или состояния x_{i*}. Доступными из состояний x_0 и x_n являются только состояния x_1, x_{0*} и x_{n*} соответственно. Будучи в состоянии x_{i*} ($i=0,...,n$), можно перейти только к состоянию x_i.

Состояния x_i и x_{i*} соответствуют *i-му уровню трудности заданий*. Для каждого i определен конкретный набор заданий соответствующей трудности. Как показано на рисунке 1, состояния с большим номером определяют задания, соответствующие большему уровню трудности, причём задания, соответствующие к самому высокому уровню трудности, соотносятся с состоянием, находящимся в крайнем положении справа.

Предполагается, что каждый испытуемый имеет один из указанных уровней достижений с индексами $l \in \{0, ..., z\}$, где $(z + 1)$ — число этих уровней, причём каждому из этих уровней соотносится набор заданий определённой трудности и $z < n$. На каждом уровне достижений задания соответствуют определённому уровню знаний, способностей или навыков.

[1] Когда в классической шкале логитов способности примерно равны трудностям, изменение в 0,5 логита в способности дает сдвиг 0,2 в вероятностях. Напротив, в неизменных условиях сдвиг 0,1 в вероятностной мере приводит к изменению в 0,25 логита в измерении способности. Таким образом, оценки весьма чувствительны к ошибкам.

Каждый уровень достижений соответствует определенному интервалу трудностей заданий, содержащему не менее одного состояния (см. рис. 1). Поэтому количество уровней трудностей равно или больше, чем количество уровней достижений.

Вероятности пребывания в состояниях модели как функции времени определяются следующим матричным уравнением:

$$\boldsymbol{p}(t+1) = \mathbf{M}(\boldsymbol{\lambda}_l)\mathbf{p}(t),$$

где t — дискретное время; $0 \le t \le T; t, T \in N$; T — конечный момент времени; N — множество натуральных чисел; $\mathbf{p}(t) = (p_0(t), \ldots, p_n(t), p_{0*}(t), \ldots, p_{n*}(t))^T$ — представляет вероятности пребывания в состояниях модели в момент времени t; $\mathbf{M}(\boldsymbol{\lambda}_l) = \|m_{ij}(\boldsymbol{\lambda}_l)\|$ — стохастическая квадратная матрица вероятностей перехода между состояниями цепи Маркова, в которой $m_{ij}(\boldsymbol{\lambda}_l)$ — вероятность перехода из состояния j в состояние i; $\boldsymbol{\lambda}_l = (p^+_{0,l}, \ldots, p^+_{n-1,l}, q^+_{0,l}, \ldots, q^+_{n,l}, q^-_{0,l}, \ldots, q^-_{n,l}, r_{0,l} \ldots, r_{n,l}, r_{0*,l}, \ldots, r_{n*,l})^T$ — упорядоченное множество рассматриваемых вероятностей перехода для уровня достижений l. Квадратные матрицы $\mathbf{M}(\boldsymbol{\lambda}_l)$ имеют порядок $2n + 2$.

Классификация испытуемых выполняется так, как представлено в разделе 2.

Если шкала измерения непрерывная, то весь диапазон ее значений следует разделить на несколько интервалов, каждый из которых интерпретируется как некоторый уровень, определяемый состоянием. Это и есть интервал значений шкалы, который должен быть выбран в результате процедуры тестирования. Чем больше число состояний, тем точнее оценка. Чем точнее оценка, тем большее количество эмпирических данных требуется для настройки модели.

Процедура тестирования определяется администрированием заданий, успешное выполнение которых требует наличия необходимого уровня достижений. Трудность задания, направленного испытуемому, соответствует состоянию модели, занятому им в текущее время.

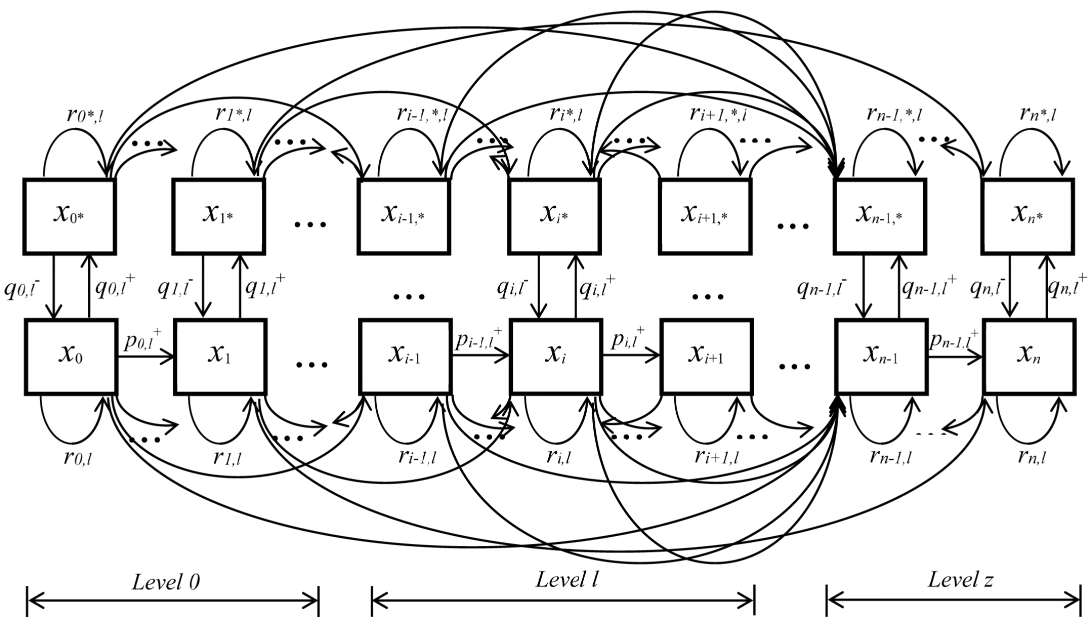

Рис. 1. Цепь Маркова, представляющая собой процедуру тестирования, состоящую из заданий: $\{x_i\}_{i=0,\ldots,n}$ *и* $\{x_{i*}\}_{i=0,\ldots,n}$ являются состояниями, $\boldsymbol{\lambda}_l = (p^+_{0,l}, \ldots, p^+_{n-1,l}, q^+_{0,l}, \ldots, q^+_{n,l}, q^-_{0,l}, \ldots, q^-_{n,l}, r_{0,l} \ldots, r_{n,l}, r_{0*,l}, \ldots, r_{n*,l})^T$— множество вероятностей перехода между этими состояниями, $l \in \{0, \ldots, z\}$— индекс уровня достижений

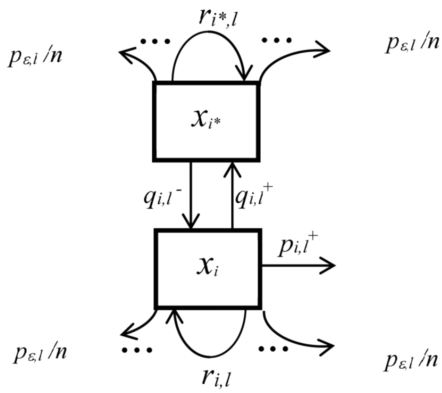

Рис. 2. Элемент цепи Маркова, представляющей процедуру тестирования: «ловушка»

Вероятности перехода вычисляются следующим образом ($i = 0, \ldots, n$):

$$q^-_{i,l} = p(i,l)k(t^*_i, i, l)(1 - p_{\varepsilon,l}),$$

$$r_{i,l} = (1 - p(i,l))\,k(t^*_i, i, l)(1 - p_{\varepsilon,l}),$$

$$r_{i^*,l} = ((1 - p(i,l))\,k(t^*_i, i, l) + (1 - k(t^*_i, i, l)))\,(1 - p_{\varepsilon,l}),$$

$$q^+_i = (1 - k(t^*_i, i, l)(1 - p_{\varepsilon,l}),$$

$$p(i,l) = \frac{e^{a(f(l)-i)}}{1 + e^{a(f(l)-i)}},$$

с $p^+_{i,l}$ определяемым для $i = 0, \ldots, n - 1$:

$$p^+_{i,l} = p(i,l)k(t^*_i, i, l)\,(1 - p_{\varepsilon,l}),$$

где функция $p(i, l)$ выражает зависимость Г. Раша для вероятности успешного выполнения тестового задания на уровне достижений l и уровне трудности задания i; параметры $k(t^*_i, i, l)$ представляют вероятности непревышения соответствующих временных ограничений t^*_i в случае заданных уровней трудности i и уровней достижений l; $p_{\varepsilon,l} << 1$ ($l \in \{0, \ldots, z\}$) — вероятность «ошибки» для каждого состояния, принадлежащего уровню достижений l; функция $f(l)$ и величина a являются параметрами модели Г. Раша. Малые вероятности $p_{\varepsilon,l}$, соответствующие уровням достижений, необходимы для поддержки адаптивности, поскольку эти значения отвечают за переходы, которые выполняются в случае изменения текущих оценок уровней достижений. Предполагается, что они равномерно распределены по соответствующим переходам из каждого состояния и, следовательно, равны $p_{\varepsilon,l} / n$. Включение небольшой вероятности «ошибки» в модель необходимо для формальных переходов, которые маловероятны для определенного уровня достижений, но характерны для какого-либо другого уровня достижений, в противном случае эти переходы были бы невозможны. Этот элемент модели полезен при классификации испытуемых. Малые вероятности перехода усредняются, поскольку их значения имеют порядок ошибки выборки (несмотря на то, что их можно оценить с использованием экспериментальных данных).

Появление в модели «ловушек», показанное на рисунке 2, обусловлено как необходимостью учитывать временную динамику процедуры тестирования (время учитывается в модели неявно), так и возможностью дифференцировать две важные группы испытуемых: тех, кто быстро генерируют ряд неправильных решений, и тех, кто находит правильное решение в течение длительного периода времени. Это важно для психологической диагностики, поскольку данная функция позволяет определять тех, кто не критичен в отношении получаемых ими результатов. Когда $t_i^* \to \infty$, соответствующие «ловушки» трансформируются из структуры с двумя состояниями в структуру с 1-м состоянием.

Правила переходов и классификация

В начальный момент процедуры тестирования предполагается, что испытуемый находится в состоянии модели x_0, т.е. получает самое простое задание (отвечающее самому нижнему уровню конструкта).

Когда испытуемый находится в состоянии x_k, то предъявляемое задание выбирается случайным образом из набора заданий, соответствующих данному состоянию, с временными ограничениями t_i^*, применяемыми для каждой пары состояний цепи Маркова (x_i, x_{i^*}).

Переходы испытуемого между состояниями определяются следующими правилами:

— если в состоянии x_i испытуемый правильно выполняет назначенное задание, и время выполнения не превышает заданное значение t_i^* (т. е. завершает назначенное задание в течение заданного периода времени), то он переходит в состояние x_{i+1};

— если в состоянии x_i испытуемый неправильно выполняет назначенное задание, а время выполнения не превышает заданное значение t_i^*, то он остается в состоянии x_i;

— если испытуемый находится в состоянии x_i, а время выполнения назначенного задания превышает заданное значение t_i^*, то испытуемый переходит в состояние x_{i^*};

— если испытуемый находится в состоянии x_{i^*} и время выполнения назначенного задания либо превышает заданное значение t_i^*, либо испытуемый неправильно выполняет назначенное задание, а время тестирования не превышает заданного значения t_i^*, то испытуемый остается в состоянии x_{i^*};

— если, находясь в состоянии x_{i^*}, испытуемый правильно выполняет назначенное задание, а время выполнения не превышает заданное значение t_i^*, то испытуемый возвращается в состояние x_i.

После каждого из указанных переходов выполняется коррекция состояний посредством классификации. Классификация осуществляется для определения уровня достижений испытуемого. Применяется подход, основанный на вычислении предельного стационарного распределения $\mathbf{p}^{\infty,l} = (p_0^{\infty,l}, \ldots, p_n^{\infty,l}, p_{0^*}^{\infty,l}, \ldots, p_{n^*}^{\infty,l})^T$ рассмотренных вероятностей состояний цепи Маркова, которые удовлетворяют следующему уравнению:

$$\mathbf{p}^{\infty,l} = \mathbf{M}(\boldsymbol{\lambda}_l)\mathbf{p}^{\infty,l}.$$

Доказано, что распределение $\mathbf{p}^{\infty,l}$ является собственным вектором выбранной в данный момент стохастической матрицы $\mathbf{M}(\boldsymbol{\lambda}_l)$, которая всегда существует и соответствует единичному собственному значению. В случае рассматриваемых матриц $\mathbf{M}(\boldsymbol{\lambda}_l)$ было также обнаружено, что процесс $\mathbf{p}(t)$ эргодичен, другие собственные значения данной матрицы находятся строго внутри единичной окружности в комплексной плоскости, $\mathbf{p}^{\infty,l} = \lim_{t\to\infty} \mathbf{p}^l(t)$. Численные методы вычисления $\mathbf{p}^{\infty,l}$ довольно просты и могут быть найдены в соответствующей литературе (Burden, Faires, 2001; Wilkinson, 1988).

Чтобы учесть траектории движения испытуемого между состояниями рассматриваемых цепей Маркова и сделать классификацию на основе распределения $\mathbf{p}^{\infty,l}$ более надежной и более подходящей реальной ситуации, столбцы матриц $\mathbf{M}(\boldsymbol{\lambda}_l)$, где $l \in \{0, ..., z\}$, корректируются после каждой попытки выполнить задание, а именно: в случае перехода между состояниями модели, соответствующим индексам j и i этих матриц (из состояния j в состояние i) элемент m_{ij} заменяется на 1, а другие элементы столбца j заменяются на 0, причем эти изменения остаются действительными только для процедуры тестирования данного испытуемого. Матрица $\mathbf{M}(\boldsymbol{\lambda}_l)$, в которой такие изменения были осуществлены, называется *матрицей пройденной траектории*. После каждого шага процедуры тестирования матрицы пройденных траекторий вычисляются для каждого исследуемого уровня достижений. Таким образом, они сохраняют информацию о динамике тестирования.

Используя матрицы пройденных траекторий, для прогнозирования уровня достижений l испытуемого, после каждой попытки выполнить задание рассчитывается математическое ожидание индекса конечной пары состояний (x_e, x_{e^*}) цепи Маркова:

$$e_l = \sum_{i=0}^{n} i\,(p_i^{\infty,l} + p_{i^*}^{\infty,l}).$$

Для дальнейших оценок перед процедурой тестирования для каждого рассматриваемого уровня достижений l и матрицы $\mathbf{M}(\boldsymbol{\lambda}_l)$, используемой в начальный момент времени процедуры тестирования, должен быть определен стационарный индекс уровня трудности $e_{\infty,l}$. Чтобы выбрать уровень достижений, который лучше всего подходит для исследуемого процесса тестирования, следует рассчитать абсолютные разности между полученным индексом e_l и стационарными индексами уровня трудности $e_{\infty,l}$ для каждого рассматриваемого уровня достижений l (см. рис. 1) и затем выбрать минимальное значение e_{min}:

$$e_{min} = \min_{l \in \{0, \ldots, z\}} \left| e_l - e_{\infty,l} \right|.$$

Уровень l_{min}, соответствующий этому значению e_{min}, дает наилучшую оценку уровня достижений. Очевидно, что чем больше испытуемый соответствует данному уровню достижений, тем меньше значения e_{min} и наоборот.

Если уровень l_{min} отличается от текущего, *уточнение состояния* реализуется с помощью перехода к состоянию, которое соответствует уровню l_{min} и имеет целый индекс, который ближе всего соответствует математическому ожиданию e_l. Уточнение состояния не производится, если либо оно требует уменьшения номера состояния после правильного решения задания, либо номер состояния увеличивается после неправильного решения задания или превышения установленного для решения лимита времени. Также целесообразно выполнять эти сдвиги номе́ра состояний, только если они не превышают заданного порога сдвига. Если какая-либо матрица пройденной траектории повторяется на предыдущих этапах процедуры тестирования, то должно выполняться уменьшение номера состояния. Используемое значение сдвига может меняться со временем.

Матрица $\mathbf{M}(\boldsymbol{\lambda}_l)$ для данного уровня достижений l называется *согласованной*, если упомянутое выше математическое ожидание попадает в набор состояний модели, соответствующих этому уровню (см. рис. 1).

При выполнении описанной здесь процедуры испытуемый «захватывается» в одном из состояний, которое наилучшим образом соответствует оценке его уровня конструкта.

В качестве альтернативного метода для классификации могут быть использованы *байесовские оценки*. Зная состояние модели, в котором испытуемый оказывается после решения последнего задания в определенный момент времени, и вероятности нахождения в этом состоянии в указанное время для каждого уровня достижений, которая может быть рассчитана с использованием ранее данного матричного уравнения, можно оценить последующие вероятности уровней достижений с помощью *формулы Байеса*:

$$P(C_l \mid S) = \frac{P(C_l)P(S \mid C_l)}{\sum_{k=0}^{z} P(C_k)\,P(S \mid C_k)},$$

где C_l — событие, указывающее, что испытуемый достиг уровня достижений l ($l \in \{0, ..., z\}$), S — событие, указывающее, что испытуемый находится в указанном состоянии модели, соответствующем заданному уровню трудности задания в заданное время, $P(C_l)$ — является априорной вероятностью достижения испытуемым уровня достижений l, $P(S \mid C_l)$ — вероятность находиться в указанном состоянии модели в указанное время, при условии, что испытуемый имеет уровень достижений l, а $P(C_l \mid S)$— это вероятность попадания в уровень достижений l при условии, что испытуемый находится в заданном состоянии модели в указанное время.

Уровень достижений, при котором достигается максимальная условная вероятность $P(C_{max} \mid S) = \max_l \{P(C_l \mid S)\}_{l=0, \dots, z}$, дает требуемую классификацию. В результате выполнения последовательности заданий получается распределение вероятностей $\{P(C_l \mid S)\}_{l=0, \dots, z}$, что позволяет оценить надежность производной классификации.

Идентификация модели

Идентификация рассмотренных марковских моделей выполняется, используя результаты тестирования. Каждый уровень достижений $l \in \{0, ..., z\}$ обрабатывается отдельно и имеет свою идентифицированную матрицу $\mathbf{M}(\boldsymbol{\lambda}_l)$, с которой связан уникальный набор оценок параметров модели $\boldsymbol{\lambda}_l$. Это позволяет проводить дальнейшую классификацию, вычисляя уровень достижений, который наилучшим образом подходит для проверяемой процедуры тестирования.

В отличие от подхода, представленного в работах (Куравский и др., 2013; Куравский и др., 2011; Kuravsky et al., 2016; Kuravsky et al., 2015, no. 21), идентификация исследуемой модели не требует решения сложных задач оптимизации. Если имеется соответствующая база данных с результатами наблюдений, настройка модели сводится к довольно простой оценке следующих характеристик: функции $f(i)$, величины a, коэффициента $k(t_i^*, i, d)$ и предельного времени $\{t_i^*\}_{i=0, \dots, n}$. База данных должна позволять оценивать частоты правильных и неправильных решений для каждой комбинации уровней достижений испытуемого и трудности задания с учетом превышений предельного времени выполнения задания.

Характеристики $f(i)$ and a могут быть оценены методом максимального правдоподобия с использованием эмпирических данных, представляющих результаты тестирования для испытуемых с заранее определенными уровнями достижений и заданий с заранее известными уровнями трудности.

Параметры $k(t_i^*, i, d)$ определяются непосредственно с помощью эмпирических данных через отношения количеств испытуемых, которые превышают и не превышают соответствующие временные пределы t_i^* в случае заданного уровня достижений.

Параметры $\{t_i^*\}_{i=0, \dots, n}$, а также вероятности «ошибок» $p_{\varepsilon,l}$ могут быть определены путем решения задачи оптимизации с критерием $C = \sum_{l=0}^{z} (e_l - e_{mean,l})^2$. Таким образом, параметры временного ограничения выбираются так, чтобы сделать ожидаемую конечную пару состояний цепи Маркова ближе к центру рассматриваемого диапазона уровней достижений модели. Так как фактические диапазоны значений для параметров $\{t_i^*\}_{i=0, \dots, n}$ известны заранее, можно использовать численный метод решения задачи оптимизации (Kuravsky et al, 2015, No. 8).

Цепь Маркова (см. рис. 1) идентифицируется отдельно для каждого уровня достижений.

Самообучение модели

После определения уровня достижений с заданной точностью C_*, а также после каждой попытки выполнить задание, происходит корректировка значений вероятностей элементов m_{ij}, соответствующих уже осуществлённым переходам между состояниями модели этого уровня. Их оценки заменяются на увеличенные малые значения $m_{ij}(1 + \delta)$, где $\delta << 1$, в то время как остальные элементы столбца j суммарно уменьшаются на значения $m_{ij}\delta / (2n + 1)$, чтобы сохранить общую сумму элементов столбца равной единице (т.е. сохранить матрицу $\mathbf{M}(\boldsymbol{\lambda}_l)$ стохастической). Эта коррекция выполняется для всех испытуемых, проходящих тестирование в течение определённого периода времени.

Представленная серия малых поправок для матричных элементов фактически реализует метод самообучения Кохонена (Kohonen, 2001).

Свойства процедуры тестирования

Процедура тестирования прекращается, когда происходит одно из следующих событий:

— значение e_{min} становится меньше заданного *порогового значения* (этот случай обычно сокращает время тестирования, так как данное условие может быть выполнено после предъявления нескольких заданий);

— превышено заданное предельное время тестирования;

— задание, соответствующее состоянию x_n, выполняется успешно за время, не превышающее t_n^*.

В случае непрерывной шкалы измерений представленный подход может быть использован в режиме «микроскопа», где мы на первом этапе получаем приблизительную оценку с использованием довольно грубых интервалов. Затем, когда испытуемый окажется в конце процедуры тестирования, мы разделяем интервал этапа на несколько подынтервалов меньшего размера, затем на следующем этапе повторяем процедуру тестирования, используя новую цепь Маркова, приспособленную для более точной оценки с этими подынтервалами, и так далее. Чем больше число таких этапов, тем точнее оценка.

Как цепи Маркова, так и указанные выше адаптивные переходы остаются скрытыми для испытуемых, которые взаимодействуют с реализованной математической моделью только на уровне получения новых заданий.

Политомические задания

Рассмотренная модель может быть обобщена на случай политомических заданий. Рассмотрим w возможных вариантов оценки результатов работы для задания, предъявленного в состоянии x_i. В этом случае соответствующий переход модели к состоянию x_{i+1}, показанный на рис. 1, заменяется многовариантным переходом, представленным на рис. 3 (каждое состояние x_{i+1} имеет свою собственную «ловушку»). Предполагается, что вероятности политомических переходов $\{p_0^+, \dots, p_{w-1}^+\}$ пропорциональны эмпирическим частотам переходов, доступным посредством наблюдения.

Трудность заданий

Для оценки трудности заданий разработана «теория двойственности», в которой конструкт испытуемого (например, способность) и шкала трудности заданий рассматриваются как двойственные концепты, заменяющие друг друга. Цепь Маркова, используемая для оценки уровней достижений, может быть применена для вычисления уровней трудности, если рассматриваемые состояния представляют собой уровни достижений, а не уровни трудности заданий, как в предыдущей модели. Задание, которое будет оцениваться, «идёт»

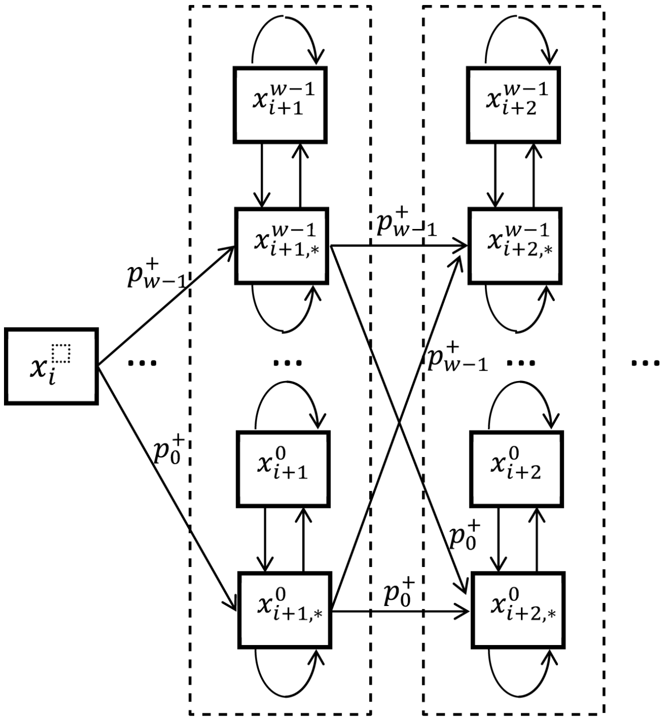

Рис. 3. Переходы в случае политомических заданий

по цепи Маркова и даётся испытуемым, уровень достижений которых соответствует состоянию, в котором исследуемое задание находится в данный момент. Все элементы теории, включая структуру модели, правила перехода, идентификацию модели, а также правила классификации, остаются неизменными.

Задания, оцениваемые с использованием нескольких шкал измерений

Когда процедура тестирования содержит задания, оцениваемые с использованием нескольких шкал измерений, рассмотренная оценка для каждой шкалы может быть выполнена независимо с использованием модели, представленной на рис. 1. Полученные результаты могут быть представлены с помощью многомерных структур, составленных из «ловушек», показанных на рис. 2. В качестве примера, трёхмерная структура этого типа приведена на рис. 4 (ее составные элементы представляют собой упомянутые выше «ловушки»).

Демонстрационная программа

Программа, демонстрирующая особенности рассмотренной модели адаптивного тестирования, доступна по адресу: http://it.mgppu.ru/files/model.zip.

Результаты и выводы

Разработан новый подход к адаптивному тестированию, модель предъявления заданий которого описывается с помощью марковских процессов с дискретными состояниями

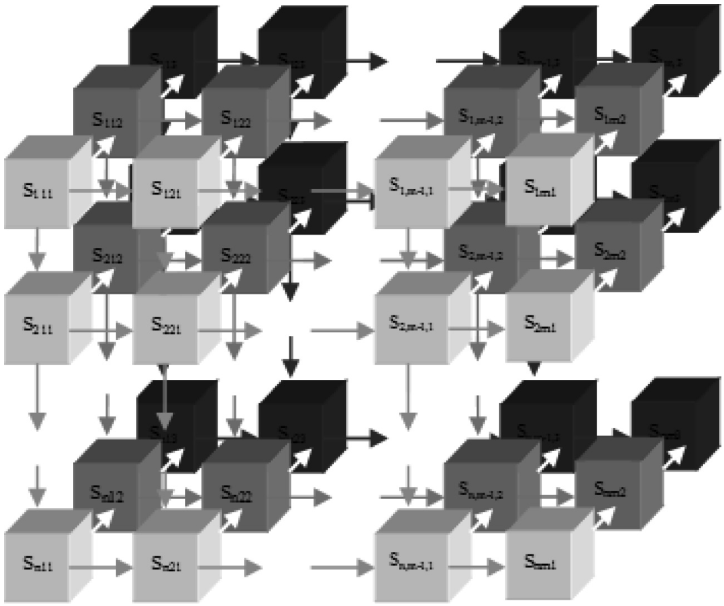

Рис. 4. Трёхмерная структура, составленная из «ловушек», представленных на рис. 2

и дискретным временем. Этот подход опирается на обобщение модели Г. Раша и имеет существенные преимущества перед адаптивным тестированием на базе современной теории тестирования (IRT). Его особенностями являются:

1) учёт в модели адаптивного тестирования всей наблюдаемой история выполнения тестовых заданий, включая порядок успешных и неуспешных результатов;

2) учёт времени, затраченного на выполнение тестовых заданий, с помощью временных ограничений в элементах модели, названных «ловушками»;

3) оценки, основанные на прогнозировании будущего поведения испытуемых, которые допускают изменение значений конструктов испытуемого во время процедуры тестирования;

4) возможность самообучения модели, что приводит к улучшению её характеристик в период эксплуатации;

5) использование марковских процессов с дискретным временем вместо аналогичных процессов с непрерывным временем, что облегчает практическое применение рассмотренного подхода;

6) простая процедура идентификация параметров модели на основе доступных результатов наблюдений;

7) связь трудностей предъявляемых заданий с историей прохождения тестов, а не текущими оценками уровней достижений испытуемых;

8) возможности обобщения разработанного подхода на случай политомических заданий и нескольких шкал измерений.

Финансирование

Работа выполнена при поддержке Российского фонда фундаментальных исследований (проект № 17-06-00277).

Литература
1. *Burden R.L., Faires J.D.* Numerical Analysis. Brooks/Cole, Cengage Learning, 9th Ed., 2011. 895 pp.
2. *Kohonen T.* Self-Organizing Maps. Springer, 3th Ed., 2001. 501 pp.
3. *Kuravsky L.S., Marmalyuk P.A., Yuryev G.A., Dumin P.N.* A Numerical Technique for the Identification of Discrete-State Continuous-Time Markov Models// Applied Mathematical Sciences. 2015. Vol. 9. № 8. P. 379—391. URL: http://dx.doi.org/10.12988/ams. 2015.410882.
4. *Kuravsky L.S., Margolis A.A., Marmalyuk P.A., Panfilova A.S., Yuryev G.A., Dumin P.N.* A Probabilistic Model of Adaptive Training // Applied Mathematical Sciences. 2016. Vol. 10. № 48. P. 2369—2380. http://dx.doi.org/10.12988/ams.2016.65168.
5. *Куравский Л.С., Марголис А.А., Мармалюк П.А., Юрьев Г.А., Думин П.Н.* Обучаемые марковские модели в задачах оптимизации порядка предъявления психологических тестов // Нейрокомпьютеры: разработка и применение. 2013. № 4. С. 28—38.
6. *Kuravsky L.S., Marmalyuk P.A., Baranov S.N., Alkhimov V.I., Yuryev G.A., Artyukhina S.V.* A New Technique for Testing Professional Skills and Competencies and Examples of its Practical Applications // Applied Mathematical Sciences. 2015. Vol. 9. № 21. P. 1003—1026. http://dx.doi.org/10.12988/ams.2015.411899.
7. *Куравский Л.С., Юрьев Г.А.* Использование марковских моделей при обработке результатов тестирования // Вопросы психологии. 2011. № 2. С. 98—107.
8. *Thompson N.A., Weiss D.J.* A framework for the development of computerized adaptive tests // Practical Assessment, Research & Evaluation. 2011. Vol. 16(1). P. 1—9.
9. *Torre J. de la, Patz R.J.* Making the Most of What We Have: A Practical Application of Multidimensional Item Response Theory in Test Scoring // Journal of Educational and Behavioral Statistics. 2005. Vol. 30(3). P. 295—311. doi:10.3102/10769986030003295.
10. *Wilkinson J.H.* The Algebraic Eigenvalue Problem, Oxford, Clarendon Press, 1988. 662 pp.

A NEW APPROACH TO COMPUTERIZED ADAPTIVE TESTING

KURAVSKY L.S.*, *Moscow State University of Psychology & Education, Moscow, Russia,*
e-mail: l.s.kuravsky@gmail.com

ARTEMENKOV S.L.**, *Moscow State University of Psychology & Education, Moscow, Russia,*
e-mail: slart@inbox.ru

YURYEV G.A.,*** *Moscow State University of Psychology & Education, Moscow, Russia,*
e-mail: g.a.yuryev@gmail.com

GRIGORENKO E.L.*,*** *Moscow State University, Moscow, Russia; University of Houston, USA,*
e-mail: elena.grigorenko@yale.edu

For citation:
Kuravsky L.S., Artemenkov S.L., Yuryev G.A., Grigorenko E.L. A New approach to computerized adaptive testing. *Eksperimental'naya psikhologiya = Experimental psychology (Russia)*, 2017, vol. 10, no. 3, pp. 33—45. doi:10.17759/exppsy.2017100303

* *Kuravsky L.S.* Doctor in Technical Sciences, Professor, Dean of the Department of Information Technologies, Moscow State University of Psychology & Education. E-mail: l.s.kuravsky@gmail.com
** *Artemenkov S.L.* PhD, Professor, Moscow State University of Psychology & Education. E-mail: slart@inbox.ru
*** *Yuryev G.A.* Ph.D. in Physics and Matematics, associate professor, Deputy Dean of the Department of Information Technologies, Moscow State University of Psychology & Education. E-mail: g.a.yuryev@gmail.com
**** *Grigorenko E.L.* PhD, Professor, Head of Laboratory of Behavoir Genetics, Chair of Psychology of Education and Pedagogics, Lomonosov Moscow State University, Professor, Yale University, University of Houston. E-mail: elena.grigorenko@yale.edu

A new approach to computerized adaptive testing is presented on the basis of discrete-state discrete-time Markov processes. This approach is based on an extension of the G. Rasch model used in the Item Response Theory (IRT) and has decisive advantages over the adaptive IRT testing. This approach has a number of competitive advantages: takes into account all the observed history of performing test items that includes the distribution of successful and unsuccessful item solutions; incorporates time spent on performing test items; forecasts results in the future behavior of the subjects; allows for self-learning and changing subject abilities during a testing procedure; contains easily available model identification procedure based on simply accessible observation data. Markov processes and the adaptive transitions between the items remain hidden for the subjects who have access to the items only and do not know all the intrinsic mathematical details of a testing procedure. The developed model of adaptive testing is easily generalized for the case of polytomous items and multidimensional items and model structures.

Keywords: Markov processes, adaptive testing, IRT, computerized adaptive testing.

Funding
This work is supported by Grant 17-06-00277 from the Russian Foundation for Basic Research.

References

1. Burden R.L., Faires J.D. *Numerical Analysis*. Brooks Cole, 2011. 895 p.
2. Kohonen T. *Self-Organizing Maps*. Springer, 2001. 501 p.
3. Kuravsky L. S., Marmalyuk P. A., Yuryev G. A., Dumin P. N. A Numerical Technique for the Identification of Discrete-State Continuous-Time Markov Models. *Applied Mathematical Sciences*, 2015, vol. 9, no. 8, pp. 379–391. URL: http://dx.doi.org/10.12988/ams. 2015.410882.
4. Kuravsky L.S., Margolis A.A., Marmalyuk P.A., Panfilova A.S., Yuryev G.A., Dumin P.N. A Probabilistic Model of Adaptive Training. *Applied Mathematical Sciences*, 2016, vol. 10, no. 48, pp. 2369—2380. http://dx.doi.org/10.12988/ams.2016.65168.
5. Kuravsky L.S., Margolis A.A., Marmalyuk P.A., Yuryev G.A., Dumin P.N. Trained Markov Models to Optimize the Order of Tasks in Psychological Testing. *Neirokomp'yutery: Razrabotka, Primenenie [Neurocomputers: Development, Application],* 2013, no. 4, pp. 28–38 (In Rus.).
6. Kuravsky L.S., Marmalyuk P.A., Baranov S.N., Alkhimov V.I., Yuryev G.A., Artyukhina S.V. A New Technique for Testing Professional Skills and Competencies and Examples of its Practical Applications. *Applied Mathematical Sciences*, 2015, vol. 9, no. 21, pp. 1003—1026. http://dx.doi.org/10.12988/ams.2015.411899.
7. Kuravsky L.S., Yuryev G.A. Primenenie markovskikh modeley dlya testirovania resultatov [Applying Markov Models to Processing Testing Results]. *Voprosy Psychologii,* 2011, no. 2, pp. 112—121, (In Rus.).
8. Thompson N.A., Weiss D.J. A framework for the development of computerized adaptive tests. Practical Assessment. *Research & Evaluation*, 2011, vol. 16, no. 1, pp. 1—9.
9. Torre J. de la, Patz R.J. Making the Most of What We Have: A Practical Application of Multidimensional Item Response Theory in Test Scoring. *Journal of Educational and Behavioral Statistics*, 2005, vol. 30, no. 3, pp. 295—311. doi:10.3102/10769986030003295.
10. Wilkinson J.H. The *Algebraic Eigenvalue Problem*. Oxford, Clarendon Press, 1988, 662 p.