



## УЧЕБНО-МЕТОДИЧЕСКИЕ ВОПРОСЫ ИБ

БИТ

*О. Н. Густун, Н. М. Леонова, А. Д. Модяев, С. Е. Попов*

### ПОВЫШЕНИЕ ТОЧНОСТИ ОЦЕНОК ЛАТЕНТНЫХ ПАРАМЕТРОВ В СИСТЕМАХ УПРАВЛЕНИЯ ПЕРСОНАЛОМ

Основными задачами управления персоналом являются подбор и наем сотрудников, а также аттестация, повышение квалификации, подготовка, переподготовка и контроль персонала. Реализация перечисленных функций приводит к необходимости создания различных специализированных систем, обеспечивающих выявление скрытых, латентных параметров, при этом сотрудники подвергаются тем или иным видам испытаний. Набор средств, с помощью которых проводятся испытания, весьма широк — от тренажеров реального времени до тестирующих сред.

В частности, в специализированных организациях функция контроля персонала включает в себя аспект оценки лояльности и благонадежности сотрудников, которая может быть реализована посредством компьютерной системы тестирования.

В последнее время в социальной деятельности все большее применение получают адаптивные методы управления, например, в сфере образования адаптивные системы тестирования позволяют повысить точность оценки параметров испытуемого за счет предъявления тестовых заданий, в большей степени соответствующих его уровню знаний [1, 2]. Для этого необходимо знать оценки уровней сложности каждого из тестовых заданий в базе данных системы адаптивного тестирования. Очевидно, чем более точными будут эти оценки, тем более успешным будет и само адаптивное тестирование. Поэтому при построении адаптивной системы тестирования применяются методики получения оценок сложности тестовых заданий и уровней знаний испытуемых на базе калибровочного тестирования.

Такой подход может применяться в широкой области приложений, когда необходимо выявить и оценить уровень того или иного латентного признака (параметра) с использованием процедуры тестирования и соответствующего банка тестовых заданий. Далее будем рассматривать метод адаптивного тестирования на примере оценки значения контролируемого параметра испытуемых с помощью тестовых заданий различного уровня сложности.

В основу методики оценивания положен метод максимального правдоподобия, который позволяет получить статистически состоятельные, асимптотически несмещенные и асимптотически эффективные оценки латентных переменных — в нашем случае значений контролируемых параметров испытуемых и уровней сложности тестовых заданий [3].

Метод состоит в том, что для оцениваемых латентных параметров строится так называемая функция правдоподобия, значение которой тем выше, чем более соответствуют значения этих параметров модели измерения. Тогда максимизация функции правдоподобия и приведет к получению искомых оценок.

В приложении к задаче тестирования функция правдоподобия имеет следующий вид:

$$L(a_{ij}, \Theta_i, \delta_j) = \prod_{i,j} p\{a_{ij} | \Theta_i, \delta_j\}, \quad (1)$$

где  $a_{ij} = 1$ , если  $i$ -й испытуемый правильно отвечает на  $j$ -е задание, и  $a_{ij} = 0$ , если ответ неверный.  $p\{a_{ij} | \Theta_i, \delta_j\}$  — вероятность исхода  $a_{ij}$  при соответствующих значениях  $\Theta_i, \delta_j$ . Произведение вычисляется по всем фактически выданным во время тестирования заданиям.

В качестве модели описания результатов тестирования используем однопараметрическую логистическую модель Раша [4]. В ней функция успеха

$$p = \frac{\exp(\Theta)}{\exp(\Theta) + \exp(\delta)} = \frac{1}{1 + \exp(-(\Theta - \delta))} \quad (2)$$

определяет вероятность  $p$  того, что участник тестирования со значением контролируемого параметра  $\Theta$  верно ответит на задание сложности  $\delta$ .

Тогда функция правдоподобия имеет вид:

$$L(a_{ij}, \Theta_i, \delta_j) = \prod_{i,j} \frac{\exp(a_{ij} \cdot (\Theta_i - \delta_j))}{1 + \exp(\Theta_i - \delta_j)}. \quad (3)$$

Значения  $\hat{\Theta}_i$  и  $\hat{\delta}_j$ , при которых функция правдоподобия достигает своего максимума, являются точечными оценками искомых латентных параметров уровней сложности тестовых заданий  $\delta_j$  и значений параметров испытуемых  $\Theta_i$ .

На практике удобным оказывается использование логарифмической функции правдоподобия  $\ln(L)$ , которая достигает своего максимума при тех же значениях  $\hat{\Theta}_i, \hat{\delta}_j$ , что и  $L$ .

Для рассматриваемой модели логарифмическая функция правдоподобия будет иметь вид:

$$\ln(L) = \sum_{i,j} a_{ij} \cdot (\Theta_i - \delta_j) - \sum_{i,j} \ln(1 + \exp(\Theta_i - \delta_j)). \quad (4)$$

Важно отметить, что функция имеет только один глобальный максимум, поэтому для решения задачи нахождения экспериментальных оценок  $\hat{\Theta}_i, \hat{\delta}_j$  можно успешно использовать методы оптимизации.

Для нахождения оптимальных значений  $\hat{\Theta}_i, \hat{\delta}_j$  предлагается использовать метод Хука—Дживса [5]. Этот метод демонстрирует очень хорошие результаты при оптимизации функций, имеющих большое число переменных, что важно для его применения при обработке результатов тестирования. Здесь количество латентных переменных может быть очень велико (оно складывается из количества испытуемых и количества тестовых заданий). На рис. 1 и рис. 2 показаны зависимости числа итераций метода Хука—Дживса от числа латентных переменных (количества тестовых заданий и испытуемых), а также от различных значений критерия останова алгоритма. Данные получены на основе моделирования процесса тестирования для различного числа испытуемых (от 15 до 75) и тестовых заданий (от 15 до 75), тест состоял из заданий, одинаковых для всех испытуемых. Результаты тестирования моделировались согласно однопараметрической логистической модели Раша.

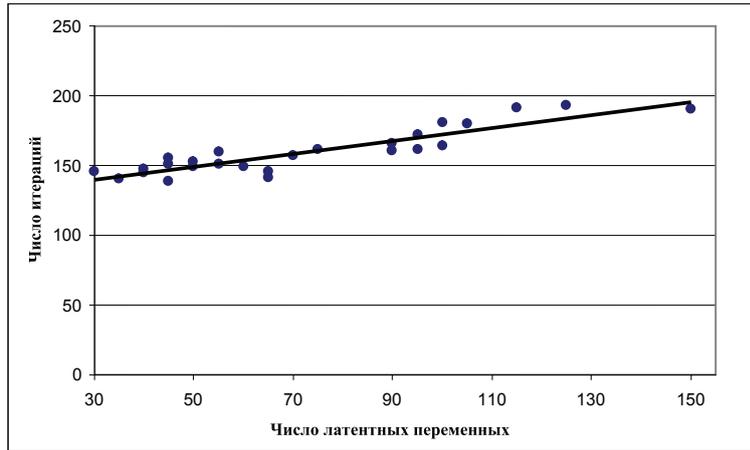


Рис. 1. Зависимость числа итераций от количества латентных переменных

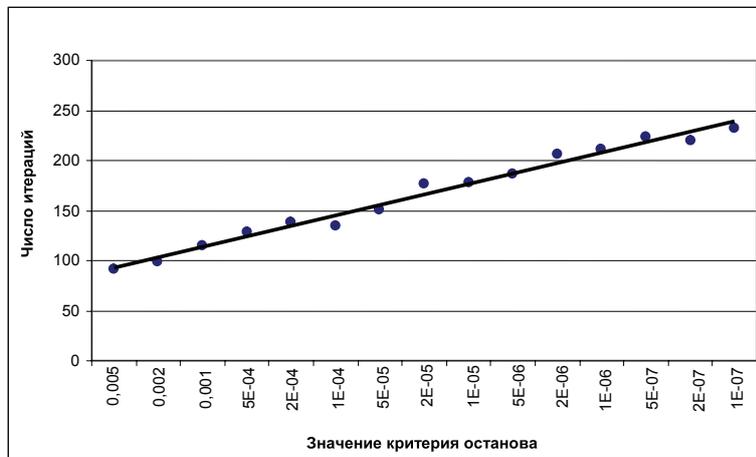


Рис. 2. Зависимость числа итераций алгоритма от значения критерия останова

Из рис. 3 следует, что точность оценки значений латентных параметров пропорциональна значению критерия останова, при этом уже для значения критерия останова  $5 \cdot 10^{-3}$  среднее отклонение полученных оценок от асимптотических значений составляет всего  $5 \cdot 10^{-3}$ .

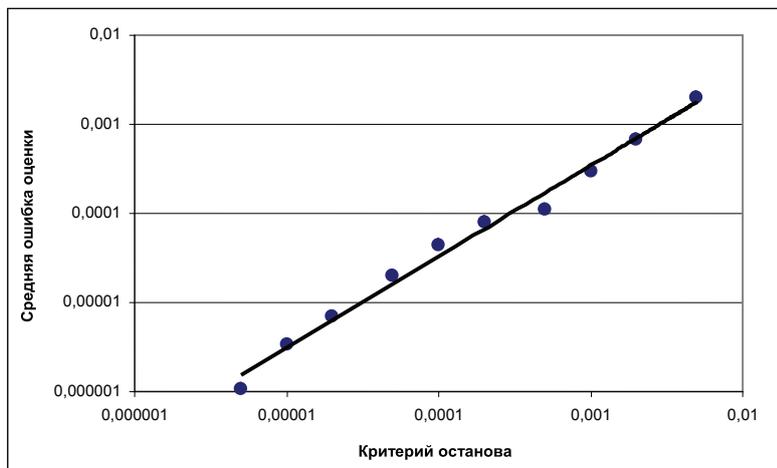


Рис. 3. Зависимость точности оценки латентных параметров от значения критерия останова (логарифмические шкалы)



Погрешность в оценке значений латентных параметров, обусловленная ограниченной информацией о результатах тестирования, значительно превосходит в данных условиях погрешность алгоритма оптимизации.

На рис. 4 приведена зависимость точности оценок контролируемых параметров испытуемых от количества предъявленных тестовых заданий. График приведен для случая произвольного выбора тестовых заданий при моделировании тестирования 500 испытуемых. Сложность всех тестовых заданий полагалась известной, оценка контролируемого параметра проводилась при помощи метода наибольшего правдоподобия и алгоритма Хука–Дживса. Значения контролируемых параметров испытуемых изменялись в диапазоне от  $-5$  до  $5$  логит. Зависимость точности оценки уровня сложности тестового задания от количества испытуемых в силу симметричности однопараметрической модели Раша имеет аналогичный вид.

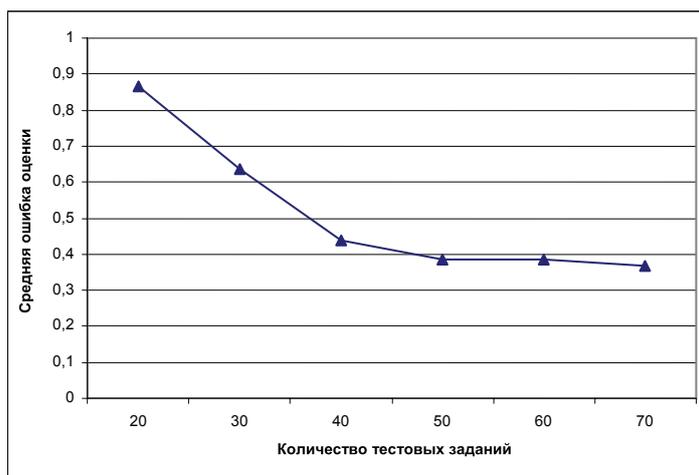


Рис. 4. Зависимость точности оценок контролируемых параметров испытуемых от количества предъявленных тестовых заданий

Таким образом, решающим фактором при оценке сложности тестовых заданий является количество предъявлений каждого из заданий испытуемым.

Рассмотрим, как изменяется оценка контролируемого параметра испытуемого в процессе тестирования при последовательной выдаче тестовых заданий с увеличивающейся сложностью, что представлено на рис. 5:

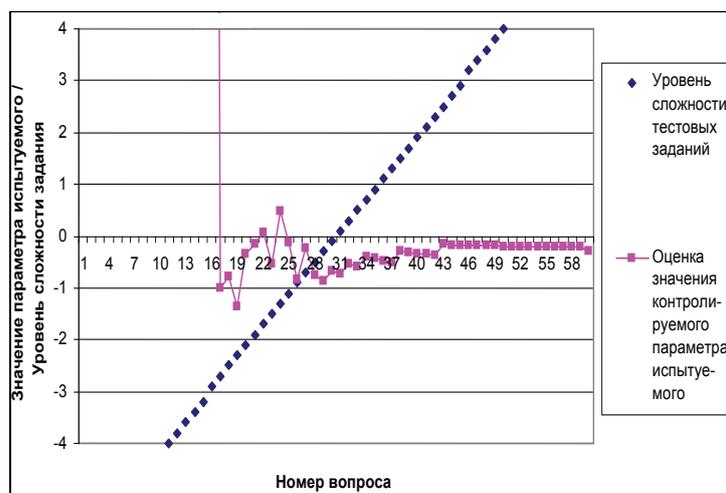


Рис. 5. Изменение оценки контролируемого параметра испытуемого в процессе тестирования



Можно заметить, что при использовании классического подхода, когда тестовые задания покрывают широкий диапазон возможных значений контролируемого параметра испытуемого, лишь часть тестовых заданий дает полезную информацию для его оценки. Задания с номерами, меньшими 15, оказываются слишком легкими для испытуемого, и на их основе невозможно получить оценку. Задания с номерами, большими 45, оказываются слишком сложными, поэтому их предъявление испытуемому также почти не дает дополнительной информации о контролируемом параметре испытуемого и имеющаяся оценка не изменяется. Таким образом, задания, не соответствующие по сложности значению параметра испытуемого, будут малоинформативны с точки зрения оценки этого значения.

Разберем теперь более детально адаптивный механизм предъявления тестовых заданий и рассмотрим, насколько он позволяет получить более точную оценку контролируемого параметра испытуемого по сравнению с классическими подходами. Адаптивный механизм, предлагаемый для использования, состоит в следующем:

- после каждого выполненного испытуемым тестового задания производится новая оценка значения контролируемого параметра испытуемого с учетом всех выполненных им к этому моменту заданий, включая последнее. При этом используются уже имеющиеся в системе (полученные по итогам калибровки) оценки сложности тестовых заданий;
- исходя из полученной оценки из базы тестовых заданий выбирается следующее задание для предъявления, наиболее близкое по уровню сложности к текущей оценке контролируемого параметра испытуемого.

Динамика изменения оценки значения параметра испытуемого в процессе тестирования с адаптивным механизмом выдачи тестовых заданий показана на рис. 6:

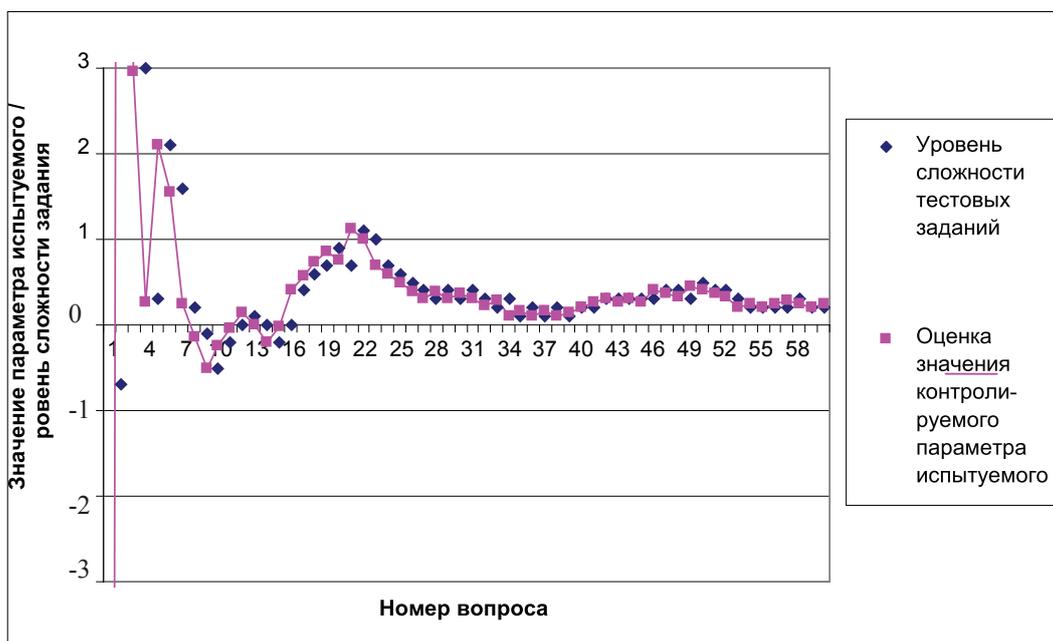


Рис. 6. Изменение оценки контролируемого параметра испытуемого в процессе адаптивного тестирования

В этом случае затухание колебаний и стабилизация оценки контролируемого параметра происходят вследствие накопления достаточной информации для получения точной оценки испытуемого. При этом с максимальной эффективностью используется вся протяженность теста.

На рис. 7 приведены кривые, показывающие точность оценки значения параметра испытуемого при использовании различных подходов к предъявлению тестовых заданий.



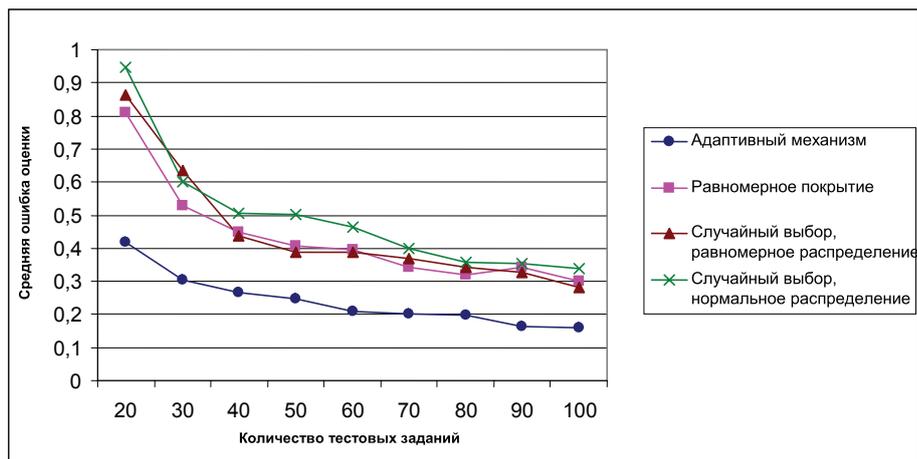


Рис. 7. Точности оценок контролируемых параметров испытуемых для различных методов предъявления тестовых заданий

На графиках отчетливо видно, что адаптивный механизм обеспечивает более точную оценку значения контролируемого параметра испытуемого при том же количестве тестовых заданий по сравнению с классическими схемами, которые не демонстрируют существенной разницы между собой. Предложенный адаптивный механизм позволяет на тесте, состоящем лишь из 30 тестовых заданий, достичь такой же точности оценки, какая при классических подходах достигается при длине теста в 100 заданий.

Таким образом, в адаптивных системах тестирования возможно добиваться большей точности измерения латентных параметров испытуемых при существенном сокращении тестового испытания. Использование данного преимущества адаптивных систем позволяет повысить надежность при принятии решений в специализированных компьютерных системах контроля персонала.

## СПИСОК ЛИТЕРАТУРЫ:

1. Леонова Н. М. Параметрически адаптивное управление образовательной деятельностью: монография / Под ред. А. Д. Модяева. М.: МИФИ, 2006.
2. Леонова Н. М., Марковский М. В. Имитационные математические модели процессов адаптивного управления образовательной деятельностью: монография / Под ред. А. Д. Модяева. М.: МИФИ, 2006.
3. Бендат Дж., Пирсол А. Прикладной анализ случайных данных: Пер. с англ. М.: Мир, 1989.
4. Baker F. B. The Basics of Item Response Theory. 2nd ed. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD, 2001.
5. Трифонов А. Г. Постановка задачи оптимизации и численные методы ее решения. М.: Наука, 2003.