

## МОНИТОРИНГ ДИНАМИКИ НЕПРЕРЫВНЫХ ПРОЦЕССОВ ТЕСТИРОВАНИЯ ПРИ ОЦЕНКЕ ЗНАНИЙ

Якубовский К.И.

*«Московский государственный университет печати имени Ивана Федорова», Москва, Россия (127550, Москва, ул. Прянишникова, 2А), e-mail: yakk2005@mail.ru*

В работе рассматривается задача статистической обработки результатов тестирования. Для анализа используется теория IRT (Item Response Theory). Рассматриваются значения исследуемой случайной величины, взвешенные средние значения и весовые коэффициенты. Используется специфика хранения и обработки знаний в системах автоматизации обучения. Значения оценок латентных переменных обладают свойствами нормального распределения и отражают взаимное расположение уровней подготовленности испытуемых и мер трудностей тестовых заданий на единой интервальной прямой. С помощью функций, задающих плотности распределения вероятностей, по найденным значениям рассчитываются вероятности решения испытуемыми тестовых заданий. Для отслеживания динамики некоторого показателя используются диаграммы статистического контроля. Предложенный алгоритм анализа данных результатов тестирований и отслеживания динамики с помощью накопленных сумм может применяться в системах по контролю и оценке знаний и компетенций в процессе обучения.

Ключевые слова: система тестирования, ItemResponseTheory, тестирование, алгоритм анализа данных результатов тестирований, дистанционные образовательные технологии

## DYNAMIC MONITORING OF UNINTERRUPTED TESTING PROCESSES BY KNOWLEDGE EVALUATION

Yakubovsky K.I.

*“Moscow State University of printing arts by Ivan Fedorov”. Moscow, Russia, (127550, Moscow, Pryanishnikova st., 2A), email: yakk2005@mail.ru*

In the article is described a task of result statistical processing. For analyse is used IRT (Item Response Theory) theory. Are described also value of observable figure, suspended middle values and suspended coefficients. Is used specific of knowledge holding and processing in automatized education systems. Values of latent valuation have normal distribution properties and reflect relative positioning of students skills and measures of tests difficulties on the one interval straight line. Probability density functions according to the founded values helps to students calculate solution probability. To monitor the dynamic of some value are used diagram of static control. The offered analysis algorithm of these results of testings and tracing of dynamics by means of the stored amounts can be applied in systems on monitoring and an assessment of knowledge and competences of training activity.

Keywords: testing system, Item Response Theory, testing, analyse algorithm of testing results, e-learning technologies.

Для создания развитой образовательной среды необходимо значительное количество разнообразных тестов. Однако разработка тестов, в полной мере отвечающих требованиям теории IRT (Item Response Theory) [5], является весьма трудоемким процессом. Их разработка требует от учебного заведения больших временных и финансовых затрат. Вновь созданный тест, прежде чем быть допущенным к контрольным тестированиям, должен пройти стадию апробации. Только после достаточного количества пробных тестирований, проводимых на значительном количестве испытуемых, тест переводится в разряд корректных. Количество заданий в тесте также должно быть достаточно большим.

В то же время в процессе обучения постоянно требуется проведение текущего контроля небольших групп учащихся [2, 7], разработка тестов для которого вполне по силам небольшому преподавательскому составу. Особенно актуально это для интенсивно развивающегося в последнее время дистанционного образования, где количество тестируемых может быть равно одному [1, 8].

### Постановка задачи

Рассмотрим кратко задачу статистической обработки результатов тестирования. Построим согласование оценок на основе одного часто используемого алгоритма обработки эмпирических данных современной теории тестирования с учетом специфики хранения и обработки знаний в системах автоматизации обучения. Будем также использовать методику и терминологию современной теории обработки тестирований IRT и достаточно известную модель Раша [6, 9].

Пусть в группе из  $N$  обучаемых проводится проверка успеваемости тестом, содержащим  $n$  заданий. Символ  $x_{ij}$  обозначает результат выполнения  $i$ -м испытуемым  $j$ -го задания.

$$x_{ij} = \begin{cases} 1, & \text{если ответ } i\text{-го испытуемого на } j\text{-ое задание верный} \\ 0, & \text{если ответ } i\text{-го испытуемого на } j\text{-ое задание неверный} \end{cases}$$

Вычисляются индивидуальные баллы каждого испытуемого  $X_i$  и каждого задания  $Y_j$ , как суммы правильно данных ответов

$$X_i = \sum_{j=1}^n x_{ij} \quad Y_j = \sum_{i=1}^N x_{ij}$$

Предварительная оценка значений параметра  $\theta_i$ , характеризующая уровень подготовки  $i$ -го ученика, вычисляется по формуле

$$\theta_i^0 = \ln \frac{p_i}{1 - p_i}, \text{ где } p_i = \frac{X_i}{n} \quad (1)$$

$p_i$  — доля правильных ответов  $i$ -го ученика. Аналогично предварительная оценка значений параметра  $\beta_j$ , характеризующая трудность  $j$ -го задания

$$\beta_j^0 = \ln \frac{1 - p_j}{p_j}, \text{ где } p_j = \frac{Y_j}{N}$$

$p_j$  — доля правильных ответов на  $j$ -ое задание теста. После этого применяется дальнейшая обработка статистического тестового материала с использованием метода наибольшего правдоподобия. Составляется функция правдоподобия для  $i$ -го испытуемого

$$L_i(\theta_i) = \prod_{j=1}^n P_{ij}^{x_{ij}} Q_{ij}^{1-x_{ij}} \quad (2)$$

Где  $P_{ij}$  — вероятность правильного выполнения  $i$ -м испытуемым  $j$ -го задания теста.  $Q_{ij}$  — вероятность неправильного выполнения  $i$ -м испытуемым  $j$ -го задания теста.  $x_{ij}$  — результат выполнения задания 1 или 0. В качестве вероятностной функции берется функция конкретной модели IRT, например логистическая для однопараметрической модели Раша

$$P_{ij} = \frac{e^{1.7(\theta_i - \beta_j)}}{1 + e^{1.7(\theta_i - \beta_j)}} = 1 - Q_{ij}$$

Далее ищется значение  $\theta_i$ , при котором функция правдоподобия достигает максимума. Это значение будет объективной оценкой искомого параметра. Его удобнее определять для логарифмической функции правдоподобия, решая следующее уравнение

$$\frac{\partial \ln L_i}{\partial \theta_i} = 0$$

Аналогично составляется функция правдоподобия для получения оценки  $\beta_j$  — уровня трудности  $j$ -го задания. И далее итерационным методом по очереди ищется решение этих систем, пока изменения оцениваемых параметров не станут меньше некоторого  $\varepsilon$

$$\left| (\theta_i)_{k+1} - (\theta_i)_k \right| < \varepsilon \quad \left| (\beta_j)_{k+1} - (\beta_j)_k \right| < \varepsilon$$

Полученные таким способом значения оценок латентных переменных обладают свойствами нормального распределения и отражают взаимное расположение уровней подготовленности испытуемых и мер трудностей тестовых заданий на единой интервальной прямой. С помощью функций задающих плотности распределения вероятностей, например указанной выше функции Раша, по найденным значениям рассчитываются вероятности решения испытуемыми тестовых заданий.

### **Модификация алгоритма анализа данных**

Использование цифровых технологий и автоматизированных систем в производстве позволяет усовершенствовать технологические процессы и приводит к их существенным качественным изменениям. В сфере образования значительный выигрыш приносит автоматизация контроля успеваемости в целом и результатов тестирований как его частной формы. При наличии достаточного количества вычислительной техники и соответствующего программного обеспечения проверка успеваемости проводится быстро и качественно. В современном образовании эти изменения в технологии образования оказываются весьма востребованными, так как позволяют быстро диагностировать появляющиеся проблемы и оперативно реагировать на них.

Регулярное проведение тестирований порождает процесс получения разнообразных данных. В дистанционном образовании он практически непрерывный. Для очной формы

обучения тестирование имеет более дискретный характер, но, учитывая масштаб всего учебного заведения и необходимость отслеживания текущего и промежуточного контроля, данный процесс также можно рассматривать как непрерывный.

Каждое отдельное тестирование формирует набор характеристик. Из них основными являются оценки за тест, дополнительными – уровни сложности заданий, качество ответов, время выполнения и др. Некоторые характеристики относятся к конкретному тестированию, другие оказываются общими для многих тестирований. Некоторые подлежат точному измерению, другие имеют приближенный или оценочный характер. Исследуя процесс в динамике, можно оценивать погрешности, выявлять грубые ошибки измерений и отслеживать тренды. Ниже описано получение указанных показателей для произвольного параметра тестирования.

Будем рассматривать три последовательности: значения исследуемой случайной величины, взвешенные средние значения и весовые коэффициенты. Обозначим их  $\{x_n\}$ ,  $\{s_n\}$ ,  $\{w_n\}$ . Начальные значения  $s_0 = 0, w_{s_0} = 0$ . На шаге  $n + 1$  элементы последовательностей будут вычисляться по рекуррентным формулам

$$s_{n+1} = w_{s_n} s_n + w_{x_{n+1}} x_{n+1}$$

$$w_{s_{n+1}} = w_{s_n} + w_{x_{n+1}}$$

Соотношение (1) можно представить в другом виде

$$s_{n+1} = s_n + \frac{w_{x_{n+1}}}{w_{s_n} + w_{x_{n+1}}} (x_{n+1} - s_n) = s_n + v(x_{n+1}) \quad (3)$$

Для взвешенного среднего значения ряда нормально распределенной случайной величины существует более 20 оценок. Предпочтительно использовать простое среднее, для которого

$$w_{x_n} = \frac{1}{\sigma_{x_n}^2}$$

Эта оценка является состоятельной, несмещенной, эффективной и достаточной, использует весь массив статистической информации и является оценкой максимального правдоподобия. Стандартная ошибка среднего в этом случае

$$\sigma_s = (\sqrt{w_{s_n}})^{-1/2}$$

Ввиду возможных случайно возникающих грубых ошибок измерений некоторые значения  $x$  лучше не учитывать в общей сумме. Простейшим критерием выявления ошибки может служить сравнение текущего значения со среднеквадратическим отклонением  $\sigma_x$ .

$$x_{n+1} - s_n < C \cdot \sigma_x(x)$$

$$v(x_{n+1}) < c \cdot \frac{\sigma_x(x)}{w_{s_n}}$$

Можно использовать либо теоретическое значение  $\sigma$  либо с увеличением статистической информации по случайной величине вычислять оценочное значение

$$\sigma_{x_n}(x) = \frac{s_n^{(2)} - w_{s_n} s_n^2}{w_{s_n} - 1}, \text{ где}$$

$$s_{n+1}^{(2)} = s_n^{(2)} + x_{n+1}^2$$

Оценка среднеквадратического отклонения будет приемлемой для  $n > n_0 = 30$ .

Для отслеживания динамики некоторого показателя обычно используют диаграммы статистического контроля. Воспользуемся этим подходом и построим последовательности нарастающих сумм:  $\{c_n^{inc}\}$  – для отслеживания тенденции роста,  $\{c_n^{desc}\}$  – для отслеживания тенденции снижения,  $c_0^{inc} = c_0^{desc} = 0$ . Элементы последовательностей определяются на каждом шаге:

$$c_{n+1}^{inc} = \max(c_n^{inc} + \psi(x_{n+1}), 0)$$

$$c_{n+1}^{desc} = \min(c_n^{desc} + \psi(x_{n+1}), 0)$$

Положим  $\psi(x) = x$ . При превышении величин некоторого заданного порогового значения  $h_0$  можно утверждать, что обнаруживается устойчивый тренд к изменению средневзвешенного значения ряда

$$\max(c_n^{inc}, -c_n^{desc}) > h_0$$

Другие полезные динамические характеристики: статистическое отклонение из (3),  $\psi(x) = \sigma(x)$  и сдвиг средневзвешенной суммы (2),  $\psi(x) = v(x)$ .

В последнем случае вместо константы  $h_0$  нужно использовать асимптотическую функцию вида  $\Psi(1/n)$ . Превышение порога укажет тренд и смещения, и разброса значений.

### Заключение

Предложенный алгоритм анализа данных результатов тестирований и отслеживания динамики с помощью накопленных сумм обладает преимуществами простоты компьютерной реализации и устойчивостью. Этот алгоритм может применяться в системах по контролю и оценке знаний и компетенций в процессе обучения [3]. Использование указанного алгоритма в автоматизированной системе поддержки образовательного процесса позволит осуществлять непрерывный контроль [4] и существенно повышать характеристики качества образовательного процесса.

### Список литературы

1. Попов Д.И. Проектирование интеллектуальных систем дистанционного образования; Известия Южного федерального университета. Технические науки. — 2001. — № 4 (22). — С. 325–332.
2. Попов Д.И., Демидов Д.Г. Адаптивная стратегия обучения персонала предприятий; В мире научных открытий. — 2011. — № 9. — С. 65–71.
3. Попов Д.И., Комолова Т.И., Попова Е.Д., Якубовский К.И. Особенности формализации компетентностного подхода при обучении в области полиграфии и издательского дела; Известия высших учебных заведений. Проблемы полиграфии и издательского дела. — 2013. — № 3. — С. 106–112.
4. Попов Д.И., Якубовский К.И., Демидов Д.Г. Нечеткая модель выбора тестовых заданий для аттестации персонала полиграфических предприятий; Известия высших учебных заведений. Проблемы полиграфии и издательского дела. — 2014. — № 3. — С. 3–9.
5. Hambleton, R. K. Application of Item Response Theory. Vancouver : Educ. Res. Inst. B. C., 1983.
6. Linacre, J.M. Many-faceted Rasch measurement. Chicago: MESA. 1989.
7. Popov D.I. Adaptive testing algorithm based on fuzzy logic; International Journal of Advanced Studies. — 2013. — Т. 3. № 4. — С. 23–27.
8. Popov D.I. Designing the intelligent distance learning systems. Proceedings of South federal university; Engineering in Life Sciences. — 2001. — Т. 22. № 4. — С. 325.
9. Wright, B.D., & Masters, G.N. Rating scale analysis. Chicago: MESA Press. 1982

**Рецензенты:**

Майков К.А., д.т.н., профессор кафедры программного обеспечения ЭВМ и информационных технологий МГТУ имени Н.Э.Баумана, г. Москва;

Николаев А.Б., д.т.н., профессор, декан факультета «Управление», заведующий кафедрой «Автоматизированные системы управления» ФГБОУ ВПО «Московский автомобильно-дорожный государственный технический университет (МАДИ)», г. Москва.