PI_3_2008.qxd 25.12.2008 18:29 Page 64

РЕАЛИЗАЦИЯ АЛГОРИТМА PROX В MICROSOFT EXCEL

Олег Деменчёнок
Восточно-Сибирский институт МВД России
AskSystem@yandex.ru

Рассмотрены основные подходы к нахождению уровней подготовленности тестируемых и трудности заданий в соответствии с моделью Раша. Описан алгоритм упрощённого решения PROX, а также технология его реализации средствами Microsoft Excel. Показано, что вычисления по алгоритму PROX в электронной таблице позволяют сделать процедуру оценивания более понятной и прозрачной.

Ключевые слова: *тест, Item Response Theory (IRT)*[1], *математическая модель Раша, вероятность правильного ответа, уровень подготовленности, уровень трудности задания, оценивание параметров модели.*

Читатель, интересующийся педагогическими измерениями, безусловно, знаком с публикациями, посвящёнными тем или иным аспектам Item Response Theory. Таких публикаций появилось немало в последние годы. Как известно, нет ничего практичнее хорошей теории. Однако до широкого применения Item Response Theory в отечественной педагогической практике ещё далеко. На взгляд автора, причина этого заключается не столько в сложности теории, сколько в новизне и непривычности идей IRT для педагогического сообщества.

1 На русский язык IRT В.С.Аванесов переводит как математическую теорию измерений (МТИ). См.: Педагогические измерения, № 3, 2007. с. 3.

Любое нововведение встречается скептически. И чем меньше новая теория согласуется со сложившимся комплексом взглядов, тем сильнее противодействие. Предлагаемый в IRT подход к измерению подготовленности обучаемых принципиально отличается от других теорий педагогических измерений, что существенно тормозит её распространение. Для раскрытия потенциала IRT педагогу нужно понять и принять эту теорию.

Основные подходы к нахождению уровней подготовленности тестируемых

Согласно IRT ответы испытуемых на тестовые задания — это достаточные исходные данные для определения уровня трудности заданий и измерения уровней подготовленности обучаемых. В табл. 1 приведена простая матрица данных, содержащая ответы восьми человек на тест из пяти заданий.

Таблица 1

		Номер задания					
		1	2	3	4	5	Количество правильно выполненных заданий
Обозначение испытуемого	A	1	0	0	0	0	1
	B	0	1	1	0	0	2
	C	1	1	0	0	0	2
	D	1	0	1	1	0	3
	E	1	1	0	1	0	3
	F	1	1	0	1	0	3
	G	1	1	1	1	0	4
	H	1	1	1	0	1	4
Количество правильных ответов		7	6	4	4	1	

Правильные ответы обозначены единицами, неправильные — нулями.

В IRT полагают, что наблюдаемые результаты выполнения теста обусловлены, в основном, разностью между уровнем подготовленности тестируемого q_i и уровнем трудности задания b_j. Другие факторы (угадывание правильных ответов; ошибки ввода данных; ошибки, вызванные неверным истолкованием условия задания; фрагментарность знаний испытуемого и т.п.) считаются случайными, а их влияние при анализе большого количества данных — взаимно компенсирующимся. Влияние уровня подготовленности тестируемого и уровня трудности задания на результат выполнения отдельного тестового задания задаётся аналитически в виде математической модели. Базовая модель IRT, предложенная Георгом Рашем[2], записывается в виде:

$$P = \frac{e^{\theta-\beta}}{1+e^{\theta-\beta}}, \qquad (1)$$

где P — вероятность правильного ответа; $e \approx 2{,}72$ — основание натурального логарифма.

Есть два основных подхода к нахождению уровней подготовленности тестируемых и трудности заданий в соответствии с выбранной математической моделью.

Первый подход заключается в непосредственном использовании уравнения математической модели. Этот вариант требует очень большого объёма вычислений, поэтому может быть реализован только с помощью компьютера. При этом достигается максимально возможная точность оценки параметров модели. Поэтому Б.Д. Райт и М.Х. Стоун[3] назвали такой путь идеальным. Определённым недостатком данного подхода является непрозрачность процесса оценивания. Программа, выполняющая обработку результатов тестирования, в глазах педагога представляет собой «чёрный ящик»[4]:

- известны входные данные (матрица ответов),
- на выходе получаем значения уровней подготовленности тестируемых и меры трудности заданий,
- как именно производится оценка — непонятно (перепроверить результаты вручную невозможно).

Непрозрачность процесса оценивания может вызвать недоверие как педагога, так и обучаемых к этому методу оценивания.

Второй подход заключается в упрощении решения за счёт введения ряда допущений. Целесообразность этого Б.Д. Райт и М.Х. Стоун в 1979 году мотивировали непрактичностью использования математической модели ввиду больших затрат машинного времени. В настоящее время такой довод неактуален: матрица ответов десятков испытуемых на десятки тестовых заданий в соответствии с моделью Раша на обычном персональном компьютере обрабатывается около секунды. Читатель может лично убедиться в этом, скачав бесплатную компьютерную программу Estimate2PL (сайт www.asksystem.narod.ru). Кроме того, упрощения неизбеж-

2 *Rasch G.* Probabilistic models for some intelligence and attainment tests. — Copenhagen, Denmark: Danish Institute for Educational Research, 1960.

3 *Wright B.D., Stone M.H.* Best Test Design. Chicago: MESA PRESS. 1979.

4 Чёрный ящик — объект изучения, внутреннее устройство которого либо неизвестно, либо слишком сложно для того, чтобы можно было по свойствам его составных частей (элементов) и структуре связей между ними делать выводы о поведении объекта; метод исследования таких объектов. (Большая советская энциклопедия, электронная версия. М.: Большая Российская энциклопедия, 2002)

но снижают точность педагогического измерения. Однако упрощённый вариант вычислений ценен тем, что может быть выполнен без специализированных программ. Тем самым процесс оценивания становится более понятным и обретает прозрачность — любое заинтересованное лицо получает возможность проверки результатов. Таким образом, упрощённый вариант вычислений помогает педагогу понять и принять IRT.

Простой алгоритм оценивания — PROX

Этот алгоритм для модели Раша предложен Лесли Коэном (Leslie Cohen)[5]. PROX базируется на предположении о нормальности распределения уровня подготовленности (среднее значение $\bar{\theta}$, стандартное отклонение σ_θ) и уровня трудности (среднее значение $\bar{\beta}$, стандартное отклонение σ_β).

Оценивание проводится в следующем порядке. Сначала определяются начальные значения уровней подготовки:

$$\theta_i^0 = \ln\frac{p_i}{q_i} = \ln\frac{R_i}{m-R_i}, \qquad (2)$$

где p_i и q_i — доля правильных и неправильных ответов i-го тестируемого; R_i — количество правильно выполненных заданий i-м тестируемым; m — число тестовых заданий.

Начальные значения трудности заданий:

$$\beta_j^0 = \ln\frac{q_j}{p_j} = \ln\frac{n-s_j}{s_j}, \qquad (3)$$

где p_j и q_j — доля правильных и неправильных ответов на j-е задание теста; s_j — количество правильных ответов на j-е задание; n — число испытуемых.

Примечание: при делении на ноль получается неопределенность, а логарифм нуля не существует. Следовательно, и числитель, и знаменатель выражений (2) и (3) должен быть ненулевым. Поэтому в рамках IRT не могут быть достаточно точно оценены испытуемые, все ответы которых правильны или неправильны, а также задания, на которые даны только правильные или только неправильные ответы. В таких случаях может быть принята следующая интерпретация:

- если все ответы испытуемого правильны, то уровень его подготовленности превышает уровень трудности любого из выполненных им заданий;
- если все ответы испытуемого неправильны, то уровень его подготовленности ниже уровня трудности любого из выполненных им заданий;
- если все ответы на тестовое задание правильны, то уровень его трудности меньше уровня подготовленности любого из испытуемых;
- если все ответы на тестовое задание неправильны, то

5 *Cohen L.* A modified logistic response model for item analysis. Manuscript, 1976.

уровень его трудности больше уровня подготовленности любого из испытуемых.

Далее находим значения поправочных коэффициентов X и Y:

$$X = \sqrt{\frac{1+\frac{U}{2,89}}{1-\frac{UV}{8,35}}}, \qquad (4)$$

$$Y = \sqrt{\frac{1+\frac{V}{2,89}}{1-\frac{UV}{8,35}}}, \qquad (5)$$

где V и U — дисперсии уровней подготовленности тестируемых и уровней сложности заданий:

$$V = \frac{\sum_{i=1}^{n}(\theta_i^0)^2 - n\bar{\theta}}{n-1}, \qquad (6)$$

$$U = \frac{\sum_{j=1}^{m}(\beta_j^0)^2 - m\bar{\beta}}{m-1}. \qquad (7)$$

Производим перерасчет полученных ранее значений уровней подготовленности и уровней трудности заданий:

$$\theta_i = \overline{\beta^0} + X\theta_i^0, \qquad (8)$$

$$\beta_j = \overline{\theta^0} + Y\beta_j^0. \qquad (9)$$

Для оценки стандартных ошибок предлагаются формулы:

$$\sigma_{\theta_i} = X\sqrt{\frac{m}{R_i(m-R_i)}}, \qquad (10)$$

$$\sigma_{\beta_j} = Y\sqrt{\frac{n}{s_j(n-s_j)}}. \qquad (11)$$

Ориентировочно можно полагать, что истинные значения уровней подготовленности испытуемых и уровней трудности заданий находятся в диапазонах:

$$\theta_i - \sigma_{\theta_i} \ldots \theta_i + \sigma_{\theta_i}, \qquad (12)$$

$$\beta_j - \sigma_{\beta_j} \ldots \beta_j + \sigma_{\beta_j}. \qquad (13)$$

Чем меньше стандартные ошибки, тем точнее результат измерения.

Имеются модификации алгоритма PROX для других моделей IRT, например, для модели с произвольными промежуточными категориями выполнения тестовых заданий[6].

Очевидно, что результат оценивания по алгоритму PROX тем точнее, чем ближе к истине предположения о нормальности распределения уровня подготовленности испытуемых и уровня трудности заданий.

Реализация PROX в электронной таблице Microsoft Excel

Хотя оценивание по алгоритму PROX можно выполнить вручную, это слишком трудоёмко и ненадёжно. Гораздо проще выполнить расчёты в электронной таблице Microsoft Excel. Кроме того, при изменении матрицы ответов новые значения уровней подготовленности тестируемых и трудности заданий

6 *Wright B.D., Masters G.N.* Rating Scale Analysis: Rasch Measurement. Chicago: Mesa Press, 1982. 204 p.

будут получены автоматически. Реализация PROX в Microsoft Excel позволяет многократно снизить трудоёмкость работы, перенести акцент с выполнения вычислений на анализ результатов.

Для примера взята задача подбора параметров модели Г. Раша из работы Б.Д. Райта и М.Х. Стоуна[7] (табл. 2). Исходные данные — это результаты выполнения 14 тестовых заданий 34 испытуемыми. Результаты по некоторым заданиям и испытуемым были удалены для того, чтобы обеспечить отсутствие строк и столбцов, состоящих только из нулей или только из единиц, поэтому нумерация оказалась не сплошной.

7 *Wright B.D., Stone M.H.* Best Test Design. Chicago: MESA PRESS. 1979.

Таблица 2. **Результаты выполнения теста**

		Задания													
		4	5	7	6	9	8	10	11	13	12	14	15	16	17
Испытуемые	25	0	1	0	1	0	0	0	0	0	0	0	0	0	0
	4	1	0	1	0	1	0	0	0	0	0	0	0	0	0
	33	1	0	1	0	0	0	1	0	0	0	0	0	0	0
	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
	27	1	1	1	1	0	0	0	0	0	0	0	0	0	0
	11	0	1	1	1	1	1	0	0	0	0	0	0	0	0
	12	1	1	1	0	1	0	1	0	0	0	0	0	0	0
	17	1	0	1	1	1	1	1	0	0	0	0	0	0	0
	19	1	1	1	1	1	1	0	0	0	0	0	0	0	0
	30	1	1	1	1	1	1	0	0	0	0	0	0	0	0
	2	1	1	1	1	1	1	1	0	0	0	0	0	0	0
	3	1	1	1	1	1	1	0	0	0	1	0	0	0	0
	5	1	1	1	1	1	1	1	0	0	0	0	0	0	0
	6	1	1	1	1	1	1	1	0	0	0	0	0	0	0
	8	1	1	1	1	1	1	1	0	0	0	0	0	0	0
	9	1	1	1	1	1	1	1	0	0	0	0	0	0	0
	13	1	1	0	0	1	1	1	1	0	1	0	0	0	0
	16	1	1	1	1	1	1	0	1	0	0	0	0	0	0
	26	1	1	1	1	1	1	1	0	0	0	0	0	0	0
	28	1	1	1	1	1	1	0	1	0	0	0	0	0	0
	29	1	1	0	1	1	0	1	1	0	0	1	0	0	0
	31	1	1	1	1	1	1	1	0	0	0	0	0	0	0
	10	1	1	1	1	1	1	1	1	0	0	0	0	0	0

PI_3_2008.qxd 25.12.2008 18:29 Page 70

	18	1	1	1	1	1	1	1	0	1	0	0	0	0	0
	14	1	1	1	1	1	1	1	1	0	0	0	0	0	0
	32	1	1	1	1	1	1	1	1	0	0	0	0	0	0
	20	1	1	1	1	1	1	1	0	1	0	0	0	0	0
	21	1	1	1	1	1	1	1	1	1	0	0	0	0	0
	22	1	1	1	1	1	1	1	1	0	1	0	0	0	0
	23	1	1	1	1	1	1	1	0	1	0	1	0	0	0
	34	1	1	1	1	1	1	1	0	0	1	1	0	0	0
	15	1	1	1	1	1	1	1	1	1	1	0	0	0	0
	7	1	1	1	1	1	1	1	1	1	1	0	1	0	0
	24	1	1	1	1	1	1	1	1	1	0	0	0	1	1

Технология реализации PROX в электронной таблице Microsoft Excel.

Сначала вводим данные табл. 2 в клетки B2:O35 (рис. 1).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		*4*	*5*	*7*	*6*	*9*	*8*	*10*	*11*	*13*	*12*	*14*	*15*	*16*	*17*
2	*25*	0	1	0	1	0	0	0	0	0	0	0	0	0	0
3	*4*	1	0	1	0	1	0	0	0	0	0	0	0	0	0
4	*33*	1	0	1	0	0	0	1	0	0	0	0	0	0	0
5	*1*	1	1	1	1	0	0	0	0	0	0	0	0	0	0

Рис. 1. Ввод исходных данных (курсивом выделены номера тестируемых и заданий)

Находим сумму баллов каждого тестируемого R: в клетку P2 вводим **=СУММ(B2:O2)** и копируем эту формулу в клетки P3:P35. Для суммы баллов по каждому заданию в B36 вводим **=СУММ(B2:B35)** и копируем в C36:O36.

По формулам (2) и (3) рассчитываем начальные значения параметров модели θ^0 и β^0:

• для уровней подготовленности испытуемых в Q2 вводим **=LN(P2/(14 – P2))** и копируем в клетки Q3:Q35 (14 — число тестовых заданий);

• для уровней трудности заданий в B37 вводим **=LN((34 – B36)/B36)** и копируем в C37:O37 (34 — количество испытуемых).

По формулам (6) и (7) определяем V и U — дисперсии уровней подготовленности тестируемых и уровней трудности заданий:

• для дисперсии уровней подготовленности в любую свободную клетку (например, B40) вводим **=ДИСП(Q2:Q35)**;

- для дисперсии трудности заданий в клетку B41 вводим **=ДИСП(B37:O37)**.

Далее по формулам (4) и (5) находим значения коэффициентов *X и Y*:

- для коэффициента *X в* клетку B42 вводим **=КОРЕНЬ((1+B41/2,89)/(1 – B40 *B41/8,35))**;
- для коэффициента *Y* в клетку B43 вводим **=КОРЕНЬ((1+B40/2,89)/(1 – B40 *B41/8,35))**.

По формулам (8) и (9) производим перерасчёт полученных ранее значений уровней подготовки и уровней сложности заданий:

- для уровней подготовленности испытуемых в R2 вводим формулу **=СРЗНАЧ($B$37:$O$37) +$B$42*Q2** и копируем в клетки R3:R35 (символом $ обозначается абсолютная ссылка, которая не меняется при копировании формулы);
- для уровней трудности заданий в B38 вводим **=СРЗНАЧ($Q$2:$Q$35)+$B $43*B37** и копируем в C38:O38.

Найдём стандартные ошибки значений уровней подготовки и уровней трудности заданий формулам (10) и (11):

- для ошибок уровней подготовленности испытуемых в S2 вводим формулу **=$B$42*КОРЕНЬ(14/(P2*(14-P2)))** и копируем в клетки S3:S35;
- для ошибок уровней трудности заданий в B39 вводим **= $ B $ 4 3 * К О Р Е Н Ь (34/(B36*(34 – B36)))** и копируем в C39:O39.

В качестве аргумента за реализацию алгоритма PROX в электронной таблице можно привести ещё один довод. Расчёт по алгоритму PROX является последовательным: находим суммы по столбцам и строкам; по этим суммам рассчитываем начальные значения уровней; по начальным значениям уровней определяем дисперсии; по дисперсиям — поправочные коэффициенты и так далее. Если при ручном счёте использовать округленные значения (не восемь значащих цифр, как на калькуляторе, а меньше), то незначительная вначале ошибка на каждом шаге будет увеличиваться. В итоге ошибка собственно вычислений может стать существенной.

Реализация алгоритма PROX в Microsoft Excel делает процесс оценивания более понятным и прозрачным. Однако следует учитывать, что результаты оценивания, полученные более точными методами, могут несколько отличаться от результатов, полученных по этому алгоритму.