УДК 519-257

Н.С. Нифонтов, А.Е. Архипов

АНАЛИЗ РЕЗУЛЬТАТОВ ЗАКРЫТОГО ТЕСТА НА ОСНОВЕ МНОГОПАРАМЕТРИЧЕСКИХ МОДЕЛЕЙ IRT

Введение

Широкое внедрение систем тестового контроля в современных условиях часто приводит к ситуациям, когда необходимо быстро и качественно проанализировать полученные данные. Возникает острая необходимость в использование объективных показателей и критериев качества для оценки полученных результатов. Применение классических методик не всегда приводит к ожидаемому результату. К примеру, известны только данные тестового опроса, но абсолютно ничего неизвестно про характер теста - неизвестны веса заданий и другие характеристики, заложенные экспертами в процессе разработки теста. Использование классических статистических методик может быть не эффективно в результате недостаточного объема данных или неизвестного уровня репрезентативной выборки.

В таких случаях обработку данных теста можно производить, используя модель совместного оценивания параметров тестируемого и параметров теста. Теоретической основой такого семейства моделей являются модели IRT (Item Response Theory). Использование параметрических моделей позволяет построить эффективный адаптивный механизм обработки данных путем выявления связи, между скрытыми параметрами тестируемого и теста.

В зависимости от поставленной задачи строится параметрическая модель с необходимыми группами характеристик тестируемого и задания, и производиться оценка параметров модели на основе репрезентативной выборки данных.

Постановка задачи

Совокупность входных данных представлена 2-мя тестами закрытого типа (входной и выходной контроль), длинной 44 и 53 вопроса и тестируемыми лицами в количестве 964 человека. Используется 3-х параметрическая логистическая модель IRT. В данной модели вероятность правильного ответа P_{ij} при решении i-м тестируемым j-ого задания определяется отношением вида:

$$P_{ij}(b_{ij}=1|L_i) = P_j(L_i) = c_j + \frac{1-c_j}{1+e^{D\alpha_j(\beta_j - L_i)}}, \tag{1}$$

где β_j – уровень сложности задания, L_i – уровень знаний тестируемого, α_j – дифференцирующая сила задания, D – константа равная 1.7, а c_j

характеризует вероятность правильного ответа на задание j в том случае, если тестируемый применял стратегию угадывания при ответе на вопрос.

Оценки параметров модели можно найти, воспользовавшись методом максимального правдоподобия. Уровень знаний L_i дискретная величина, которая может принимать $l_k, k = \overline{1,K}$значений с неизвестными вероятностями $\pi_k, k = \overline{1,K}$. Таким образом, уровень знаний L_i обладает мультиномиальным распределением с вероятностями $\pi = (\pi_1, \ldots, \pi_k)$.

Результаты проведения тестирования можно представить в виде бинарной матрицы ответов B размерностью NxM.

Вероятность того, что случайному i-му тестируемому из выборки с распределением уровня знаний, задающимся вероятностями $\pi = (\pi_1, \ldots, \pi_k)$, будет соответствовать i-ый вектор матрицы ответов $B_i = (b_{i1}, \ldots, b_{iM})$, $i = \overline{1,N}$ можно представить виде:

$$\begin{aligned} f(b_{ij}|\Delta,\pi) &= \sum_{k=1}^{K} f(b_{ij}, l_k|\pi_k, \Delta) = \sum_{k=1}^{K} f(b_{ij}|l_k, \Delta)\pi_k \\ &= \sum_{k=1}^{K} \pi_k \prod_{j=1}^{M} P(l_k|\delta_j)^{b_{ij}} [1 - P(l_k|\delta_j)]^{1-b_{ij}} \end{aligned}, \tag{2}$$

где $\Delta = (\delta_1, \ldots, \delta_M)$, $\delta_j = (\alpha_j, \beta_j, c_j)$ – набор параметров j-ого вопроса.

Таким образом, функция правдоподобия для набора ответов всей группы тестируемых будет иметь вид:

$$\varphi(b_{ij}|\Delta,\pi) = \prod_{i=1}^{N} \left(\sum_{k=1}^{K} \pi_k \prod_{j=1}^{M} P(l_k|\delta_j)^{b_{ij}} [1 - P(l_k|\delta_j)]^{1-b_{ij}} \right). \tag{3}$$

Соответствующая полная функция правдоподобия будет иметь вид:

$$\begin{aligned} \varphi(b_{ij}, L_i|\Delta,\pi) &= \prod_{i=1}^{N} \prod_{j=1}^{M} P(L_i|\delta_j)^{b_{ij}} [1 - P(L_i|\delta_j)]^{1-b_{ij}} f(L_i|\pi) \\ &= \prod_{j=1}^{M} \prod_{k=1}^{K} P(l_k|\delta_j)^{x_{jk}} [1 - P(l_k|\delta_j)]^{y_k - x_{jk}} \pi_k^{y_k} \end{aligned}, \tag{4}$$

где y_k – количество тестируемых с уровнем знаний $L_i = l_k$, $f(L_i|\pi) = \pi_k$ при $L_i = l_k$, x_{jk} это количество тестируемых с $L_i = l_k$, которые ответили правильно на j-й вопрос.

Итерационный метод оценивания параметров модели

Максимизацию функции правдоподобия удобней проводить в $\log(\varphi)$ форме:

$$\log(\varphi(b_{ij}, L_i|\Delta,\pi)) = \sum_{j=1}^{M} \begin{array}{l} \sum_{k=1}^{K} x_{jk} \log(P(l_k|\delta_j)) \\ +(y_k - x_{jk}) \log[1 - P(l_k|\delta_j)] + y_k \log(\pi_k) \end{array}. \tag{5}$$

Таким образом, нам нужно найти оценки параметров Δ и π, которые будут максимизировать полную функцию правдоподобия. Данная про-

 ISSN 1562-9945

цедура носит итерационный характер и состоит из двух основных вычислительных блоков.

Первый этап на итерации $s = (0, 1, \ldots)$ заключается в вычислении ожидаемого значения $\log$-функции правдоподобия (на итерации s вычисляем новые значения $x_{jk}^{(s)}$и $y_k^{(s)}$). Значения $x_{jk}^{(s)}$ и $y_k^{(s)}$ вычисляем на основе условного распределения уровня знаний L_i (полученного по входным данным из бинарной матрицы ответов B) и значениям параметров $\Delta^{(s)}$и $\pi^{(s)}$ которые вычислили на втором этапе оценки на итерации $s-1$ (где $\Delta^{(0)}$ и $\pi^{(0)}$ – начальные значения параметров).

Условная вероятность уровня знаний $L_i = l_k$ для i-ого тестируемого ответ которого представлен вектором $B_i = (b_{i1}, \ldots, b_{iM}), i = \overline{1, N}$ с параметрами $\Delta^{(0)}$ и $\pi^{(0)}$ будет иметь вид:

$$\begin{aligned} f(l_k|b_{ij}, \Delta^{(s)}, \pi^{(s)}) &= \frac{f(b_{ij}|l_k, \Delta^{(s)})\pi_k^{(s)}}{\sum\limits_{k'=1}^{K} f(b_{ij}|l_{k'}, \Delta^{(s)}, \pi_{k'}^{(s)})} \\ &= \frac{\pi_k^{(s)} \prod\limits_{j=1}^{M} P(l_k|\delta_j^{(s)})^{b_{ij}} [1-P(l_k|\delta_j^{(s)})]^{1-b_{ij}}}{\sum\limits_{k'=1}^{K} \pi_{k'}^{(s)} \prod\limits_{j=1}^{M} P(l_{k'}|\delta_j^{(s)})^{b_{ij}} [1-P(l_{k'}|\delta_j^{(s)})]^{1-b_{ij}}}. \end{aligned} \tag{6}$$

Значение $y_k^{(s)}$ можно получить, просуммировав все условные вероятности (6) уровня знаний для всех тестируемых с $L_i = l_k$:

$$\begin{aligned} y_k^{(s)} &= \sum_{i=1}^{N} f(l_k|b_{ij}, \Delta^{(s)}, \pi^{(s)}) = \sum_{i=1}^{N} \frac{f(b_{ij}|l_k, \Delta^{(s)})\pi_k^{(s)}}{\sum\limits_{k'=1}^{K} f(b_{ij}|l_{k'}, \Delta^{(s)}, \pi_{k'}^{(s)})} \\ &= \sum_{i=1}^{N} \frac{\pi_k^{(s)} \prod\limits_{j=1}^{M} P(l_k|\delta_j^{(s)})^{b_{ij}} [1-P(l_k|\delta_j^{(s)})]^{1-b_{ij}}}{\sum\limits_{k'=1}^{K} \pi_{k'}^{(s)} \prod\limits_{j=1}^{M} P(l_{k'}|\delta_j^{(s)})^{b_{ij}} [1-P(l_{k'}|\delta_j^{(s)})]^{1-b_{ij}}}. \end{aligned} \tag{7}$$

Просуммировав все условные вероятности для всех тестируемых с $L_i = l_k$ которые правильно ответили на j-й вопрос (если ответил правильно то $b_{ij} = 1$ иначе $b_{ij} = 0$) мы получаем новое значение$x_{jk}^{(s)}$:

$$x_{jk}^{(s)} = \sum_{i=1}^{N} b_{ij} f(l_k|b_{ij}, \Delta^{(s)}, \pi^{(s)}) = \sum_{i=1}^{N} \frac{b_{ij} f(b_{ij}|l_k, \Delta^{(s)})\pi_k^{(s)}}{\sum\limits_{k'=1}^{K} f(b_{ij}|l_{k'}, \Delta^{(s)}, \pi_{k'}^{(s)})}, \tag{8}$$

Второй этап на итерации s заключается в подстановке вычисленных на первом этапе на итерации s значений $x_{jk}^{(s)}$ и $y_k^{(s)}$ в $\log$-функции правдоподобия и нахождения значения параметров Δ и π ($\Delta^{(s+1)}$и $\pi^{(s+1)}$) которые максимизируют ее. В дальнейшем эти значение используем на следующей итерации в первом этапе.

Функцию правдоподобия в $\log$-виде можно представить в форме:

$$
\begin{aligned}
&\log(\varphi(b_{ij}, L_i|\Delta, \pi)) = \sum_{j=1}^{M} \Phi(\Delta) + \Psi(\pi), \\
&\Phi(\Delta) = \sum_{k=1}^{K} x_{jk}^{(s)} \log(P(l_k|\delta_j)) + (y_k^{(s)} - x_{jk}^{(s)}) \log[1 - P(l_k|\delta_j)], \\
&\Psi(\pi) = \sum_{k=1}^{K} y_k^{(s)} \log(\pi_k)
\end{aligned}
\tag{9}
$$

Оценивание параметров Δ и π на данном этапе можно производить отдельно, путем нахождения значений Δи π которые приводят к максимуму $\Phi(\Delta)$ и $\Psi(\pi)$.

$\pi_k^{(s+1)}$ вычисляется на втором этапе на итерации s по правилу:

$$
\pi_k^{(s+1)} = \frac{y_k^{(s)}}{N}. \tag{10}
$$

$\delta_j^{(s+1)} = (\alpha_j, \beta_j, c_j)$ вычисляется на втором этапе на итерации s путем решения системы уравнений:

$$
\left\{
\begin{array}{c}
\frac{\partial \Phi(\Delta)}{\partial \alpha_j} = \sum_{k=1}^{K} \frac{x_{jk}^{(s)} - y_k^{(s)} P(l_k|\delta_j)}{P(l_k|\delta_j)[1 - P(l_k|\delta_j)]} \cdot \frac{\partial P(l_k|\delta_j)}{\partial \alpha_j} = 0 \\
\frac{\partial \Phi(\Delta)}{\partial \beta_j} = 0 \\
\frac{\partial \Phi(\Delta)}{\partial c_j} = 0
\end{array}
\right.
\tag{11}
$$

Решение данной системы уравнений не вызывает затруднений. Классические итерационные методы, к примеру, метод Ньютона, позволяют получить необходимый результат.

Итерации выполняем до тех пор, пока не получим необходимую точность вычисления параметров. Критерием тут может служить разница изменений параметров между итерациями $s-1$ и s.

Особенностью входных данных (тесты из 44 и 53 вопросов и выборка из 941 тестируемого) было наличие дополнительной проверочной информации, а именно оценки выставленной преподавателем (экспертная оценка) по этим тестам. Фактически проверку эффективности применения заданной параметрической модели для данной выборки можно выполнить, сравнив конечные результаты с экспертной оценкой (вначале проведя шкалирование уровня знаний и оценки).

Оценивание параметров модели производилось на 2-х репрезентативных выборках из 300 и 500 тестируемых, которые были случайно выбраны из всей совокупности. Перед сравнением полученных результатов с экспертной оценкой из входных данных были исключены аномальные тестируемые (явно завышенный или заниженный бал по сравнению с сырым балом теста, скорее всего в таких случаях эксперт, выставляя оценку, руководствовался дополнительной информацией о тестируемом, накопленной им в учебном процессе). В результате отсева аномальных данных была сформирована выборка из 896 тестируемых. Таким образом, репрезентативные выборки (настроечные выборки) для оценивания параметров системы составили 300 (596 тестируемых это рабочие данные, по которым производилось сравнения с экспертной оценкой) и 500 (рабочая выборка – 396).

Учет влияния времени t_{ij} потраченного при ответе на вопрос

При наличии во входных данных такого показателя как время t_{ij} потраченное i-м тестируемым при ответе на j-й вопрос имеет смысл воспользоваться модифицированной версией модели $(\alpha_j, \beta_j, c_j, L_i)$:

$$P_{ij}(b_{ij}=1|L_i) = P_j(L_i) = c_j + \frac{1-c_j}{1+e^{-D\alpha_j(L_i - \frac{\rho_i d_j}{t_{ij}} - \beta_j)}}, \qquad (12)$$

где параметры $(\alpha_j, \beta_j, c_j, L_i)$ имеют тот же смысл, а ρ_i параметр характеризует скорость тестируемого при ответе на вопрос (ρ_i дискретная величина, которая может принимать $r_l, l = \overline{1, L}$ значений), d_j – параметр характеризующий теоретическую оценку затрат времени, необходимую для подготовки решения j-ого вопроса.

Процедура оценки параметров для модели остается такой же и характеризуется лишь усложнением решения системы уравнений на втором этапе.

Формула (9) для модифицированной модели принимает вид:

$$f(l_k, r_l|b_{ij}, t_{ij}, \Delta^{(s)}, \pi^{(s)}) = \frac{f(b_{ij}|r_l, l_k, t_{ij}, \Delta^{(s)})\pi_{kl}^{(s)}}{\sum\limits_{k'=1}^{K}\sum\limits_{l'=1}^{L} f(b_{ij}|r_l, l_k, t_{ij}, \Delta^{(s)})\pi_{k'l'}^{(s)}} =$$

$$\frac{\pi_{kl}^{(s)}\prod\limits_{j=1}^{M} P(l_k, r_l|t_{ij}, \delta_j^{(s)})^{b_{ij}}[1-P(l_k, r_l|t_{ij}, \delta_j^{(s)})]^{1-b_{ij}}}{\sum\limits_{k'=1}^{K}\sum\limits_{l'=1}^{L}\pi_{k'l'}^{(s)}\prod\limits_{j=1}^{M} P(l_{k'}, r_{l'}|t_{ij}, \delta_j^{(s)})^{b_{ij}}[1-P(l_{k'}, r_{l'}|t_{ij}, \delta_j^{(s)})]^{1-b_{ij}}}. \qquad (13)$$

Таблица 1

Оценки параметров для модели (α_j, β_j, c_j) и модели с учетом t_{ij}

	3-х параметрическая модель			Модель с учетом времени t_{ij}			
j	α	β	c	α	β	c	d
1	0,8211	0,8627	0,2350	0,7762	0.7526	0,2231	1.8742
2	1,3225	0,6252	0,3694	1,2891	0.3711	0.3313	9,8127
3	1,1442	0,5817	0.3092	1,1943	0.3140	0.2956	9,7185
4	1.3661	0,6820	0.3319	1,3024	0,3069	0.3247	9,8251

Выводы

Для репрезентативной выборки из 500 тестируемых, результаты параметрической модели приблизились к показателям экспертной оценки, и в целом разброс значений не превысил 10-15%, что может служить показателем эффективности модели в данных условиях. Увеличение объема выборки позволяет увеличить эффективность оценивания.

Регулирую состав репрезентативной выборки можно адаптировать процедуру оценки под определенный уровень знаний тестируемых из общей совокупности.

Большой интерес представляет сравнение полученных оценок параметров (α_j, β_j, c_j) для двух моделей, а также возможность обоснования необходимости учета времени t_{ij}, затраченного тестируемым для ответа на вопрос (насколько важно учитывать время, затраченное тестируемым для ответа на вопрос, и каких образом этот фактор может отразиться на оценке уровня знаний тестируемого).

Анализирую оценки параметров для двух моделей можно сделать вывод, что введение в модель учета времени t_{ij} практически не оказывает влияния на параметр c_j. Аналогичную тенденцию можно наблюдать и для параметров (α_j, β_j).

Литература

1. Schnipke D.L., Scrams D.J. Modeling item response times with two-state mixture model: A new methods of measuring speededness. // Journal of Educational Measurement – 1997. - 34 – P.213-232

2. Bock R.D., Aitken M. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm // Psychometrika – 1981. – 46 – P. 443-459

3. Нейман Ю.М., Хлебников В.А. Введение в теорию моделирования и параметризации педагогических тестов. – М.: Прометей, 2000. - 169 с.

4. Нифонтов Н.С., Архипов А.Е. Об использовании параметрических моделей в тестах закрытого типа // Адаптивные системы автоматического управления. Межведомственный научно-технический сборник. - Днепропетровск: Системные технологии, 2007.-Вып. 11(31).- С.55-63

5. Гмурман В.Е. Теория вероятностей и математическая статистика. - М.: Высшая школа, 1997, -480 С.

Отримано 12.03.2009 р.