МОДЕЛИ ПЕДАГОГИЧЕСКОГО ИЗМЕРЕНИЯ С УЧЁТОМ ГРАДАЦИИ СТЕПЕНИ ПРАВИЛЬНОСТИ ОТВЕТА

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

Рассмотрены основные модели измерения с учётом градации степени правильности ответа. Дана оценка пригодности этих моделей для педагогического измерения, обозначены области возможного применения моделей. Описан подход к обработке результатов измерения на основе метода максимального (наибольшего) правдоподобия.

Ключевые слова: *тест, модель измерения, градация степени правильности ответа, критерий подбора параметров модели.*

Постановка проблемы

Математическое моделирование, теория вероятностей и математическая статистика создают математическую основу современного инструментария педагогических измерений. А это позволяет перейти на качественно новый уровень объективности измерения учебных достижений.

Распространённая сейчас двоичная (или дихотомическая) система оценки правильности ответа на тестовое задание (правильно 1, или неправильно 0) неполные и неточные ответы квалифицирует как незнание, что не всегда оправдано. Учёт неполного или частично правильного решения повышает информативность результатов тестирования и потенциально может способствовать достижению большей точности контроля знаний.

Известен целый ряд математических моделей, в которые заложена возможность градации степени правильности ответов. В данной статье рассмотрены возможности этих моделей в системе педагогических измерений.

Виды математических моделей, учитывающих частично правильные ответы

Основу многих математических моделей педагогических измерений составляет модель Г. Раша (G. Rasch)[1], связывающая вероятность правильного ответа P, уровень подготовленности испытуемого q и уровень трудности задания β:

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

где $e \approx 2{,}72$ — основание натурального логарифма.

В модели Раша принято, что результат выполнения задания x может принимать только одно из двух значений: 0 или 1. В моделях с градацией степени правильности ответов результат выполнения задания может принимать одно из значений из ряда 0 (неправильно), 1 (частично правильно) ... x (более правильно) ... $x_{\max}$ (абсолютно правильно). Теоретически количество градаций $x_{\max}$ не ограничено; однако в практических расчётах обычно не превышает одиннадцати (т.е. $x_{\max} \leq 11$).

Если процесс решения задания рассматривать поэтапно, то ошибка на каком-либо шаге решения лишает тестируемого возможности получить баллы на последующих стадиях. Например, получить 2 балла можно только в том случае, если правильно выполнен первый шаг, за что даётся 1 балл, а затем и второй шаг, за что добавляется ещё один балл.

С учётом градации степени правильности ответов модель Раша можно записать в виде:

$$P_x = \frac{e^{\theta-\beta_x}}{1+e^{\theta-\beta_x}} = \frac{1}{1+e^{-(\theta-\beta_x)}}, \quad (2)$$

где P_x — вероятность правильного выполнения шага x, β_x — уровень трудности шага x;

$x = 1, 2 \ldots x_{\max}$.

Существует несколько вариантов задания уровня трудности шага, каждый из которых соответствует определённой математической модели[2].

Модель с произвольными промежуточными категориями выполнения тестового задания, известная в англоязычной литературе как Partial Credit Model (PCM). Модель предполагает однократное выполнение тестового задания, уровни трудности

---

1 *Rasch G.* Probabilistic Models for Some Intelligence and Attainment Tests. Copengagen, 1960, Danish Institute of Educational Research (Expanded edition, Chicago, 1980, Mesa Press, 199 p.).

2 *Wright B.D., Masters G.N.* Rating Scale Analysis: Rasch Measurement. Chicago: Mesa Press, 1982. 206 p.

отдельных шагов не зависят друг от друга и не связаны с параметрами других тестовых заданий.

Модель с произвольными промежуточными категориями выполнения тестового задания ограничена только однократностью выполнения тестового задания, что вполне соответствует обычному сценарию тестирования. Благодаря возможности оценивать как пошаговое выполнение задания, так и частичную правильность ответа без учёта последовательности решения, эта модель наиболее пригодна для контроля знаний. Именно эта модель лежит в основе оценивания единого государственного экзамена, проводимого для выпускников образовательных учреждений, освоивших основные общеобразовательные программы среднего (полного) общего образования.

В **модель с фиксированными промежуточными категориями выполнения заданий** (Rating Scale Model) заложена однократность выполнения задания. Кроме того, для всех заданий одинаково количество градаций правильности ответа $x_{\max}$.

$x_{\max 1} = x_{\max 2} = ... = x_{\max m} = \text{const}.$

Также одинаковы относительные трудности шагов для всех заданий:

$$\beta_x = \beta + \tau_x, \qquad (3)$$

где β — уровень трудности тестового задания; τ_x — относительная трудность шага x, одинаковая для всех заданий.

Таким образом, фиксируются расстояния между уровнями трудности соседних шагов, т.е. для всех заданий справедливо:

$$\beta_2 - \beta_1 = \beta + \tau_2 - (\beta + \tau_1) = \\ = \tau_2 - \tau_1 = \text{const}, \\ \beta_3 - \beta_2 = \tau_3 - \tau_2 = \text{const и т.д.}$$

Модель с фиксированными промежуточными категориями часто используется при проведении опросов, анкетировании, диагностировании различных проблем. Пример:

ВАШЕ ОТНОШЕНИЕ К РЕФОРМЕ ОБРАЗОВАНИЯ:

–3 — крайне отрицательное;
–2 — отрицательное;
–1 — скорее отрицательное, чем положительное;
0 — нейтральное;
1 — скорее положительное, чем отрицательное;
2 — положительное;
3 — полностью одобряю.

Для целей педагогического анкетирования можно использовать вопросы с фиксацией степени уверенности испытуемого, например:

МОЖЕТ ЛИ ШАРОВАЯ МОЛНИЯ СУЩЕСТВОВАТЬ БОЛЬШЕ ЧАСА?

Меру уверенности выберите по шкале:

–2 — нет, не может;
–1 — наверное, не может;
0 — затрудняюсь ответить;
1 — наверное, может;
2 — да, может.

Биномиальная модель (Binomial Trials Model) описывает ситуацию, в которой тестируемому дано $x_{\max}$ независимых попыток выполнить тестовое задание. Каждая попытка оценивается дихотомически (правильно/неправильно), причём вероятность успеха в любой из попыток одинакова. Первым шагом считается получение хотя бы одного балла, вторым — получение, как минимум, двух баллов и так далее. Уровень трудности шага x определяется выражением:

$$\beta_x = \beta + \ln \frac{x}{x_{\max} - x + 1}. \qquad (4)$$

Биномиальная модель находит применение для измерения быстроты реакции, точности движений, рациональности распределения усилий и т.д. Тестовое задание может состоять из серии попыток щёлкнуть мышкой по маленькому объекту, кратковременно появляющемуся в произвольной точке экрана. Для контроля знаний тестовое задание можно составить из серии близких по уровню трудности заданий, каждое из которых оценивается одним баллом. Например, серия заданий на проверку знания таблицы умножения.

Модель Пуассона (Poisson Counts Model) можно рассматривать как частный случай биноминальной модели. Если испытуемому предоставляется достаточно большое количество независимых попыток выполнения задания, а вероятность успеха в каждой попытке достаточно мала, то биномиальное распределение может быть аппроксимировано распределением Пуассона. Тогда уровень трудности шага x можно найти по формуле:

$$\beta_x = \beta + \ln x. \qquad (5)$$

Модель Пуассона может использоваться для определения надёжности, эффективности, напряжённости и других параметров человеко-машинных систем. В педагогическом тестировании модель Пуассона подходит для оценки уровня подготовленности по количеству ошибок, допущенных испытуемым при чтении (переводе) теста, или количеству успешно выполненных заданий в тестах с жёстким ограничением времени.

Из проведённого анализа следует, что каждая из названных выше моделей имеет определённый потенциал для педагогических измерений. Ввиду специфичности заданий педагогических измерений модели с фиксированными промежуточными категориями, биноминальная и Пуассона пригодны с ограничениями.

Вероятность получения *x* баллов

Для любой из описанных моделей справедливо уравнение (2), т.е. все эти модели основаны на модели Раша. Однако непосредственное использование уравнения (2) затруднено тем, что вероятность правильного выполнения шага x не равна вероятности получения x баллов за решение тестового задания.

Рассмотрим случай, когда за выполнение задания можно получить от 0 до 2 баллов. Очевидно, что набрать 2 балла можно только при условии, что первый шаг выполнен правильно. Согласно теореме умножения вероятностей двух независимых событий, вероятность одновременного появления двух событий (первое событие — успешное выполнение первого шага, второе — второго шага) равна произведению их вероятностей:

$$\pi_2 = P_1 \cdot P_2, \qquad (6)$$

где P_1 и P_2 — вероятность успешного выполнения соответственно первого и второго шагов; π_2 — вероятность получить 2 балла (во избежание путаницы, вероятность успешного выполнения i-го шага будем обозначать P_i, а вероятность получить i баллов — π_i).

Очевидно, что вероятность правильного выполнения первого шага равна сумме вероятностей получить один и два балла:

$$P_1 = \pi_1 + \pi_2. \qquad (7)$$

Подставим (7) в выражение (6) и зададим P_2 по формуле (2):

$$\pi_2 = (\pi_1 + \pi_2) \cdot P_2 = \pi_1 \cdot P_2 : (1 - P_2),$$

$$\pi_2 = \frac{\pi_1}{1+e^{-(\theta-\beta_2)}} : \frac{e^{-(\theta-\beta_2)}}{1+e^{-(\theta-\beta_2)}} = (8)$$

$$= \frac{\pi_1}{e^{-(\theta-\beta_2)}} = \pi_1 \cdot e^{\theta-\beta_2},$$

где β_2 — уровень трудности второго шага.

Аналогично можно получить вероятность результата выполнения задания с одним баллом:

$$\pi_1 = \pi_0 \cdot e^{\theta-\beta_1}, \qquad (9)$$

где β_1 — уровень трудности первого шага.

Так как получение 0, 1 и 2 баллов составляют полную группу возможных событий, то сумма их вероятностей равна единице:

$$\pi_0 + \pi_1 + \pi_2 = 1,$$

$$\pi_0 = \frac{1}{1+e^{\theta-\beta_1}+e^{\theta-\beta_1}\cdot e^{\theta-\beta_2}}. \qquad (10)$$

Подставив (10) в формулы (8) и (9), получим:

$$\pi_1 = \frac{e^{\theta-\beta_1}}{1+e^{\theta-\beta_1}+e^{\theta-\beta_1}\cdot e^{\theta-\beta_2}},$$

$$\pi_2 = \frac{e^{\theta-\beta_1}\cdot e^{\theta-\beta_2}}{1+e^{\theta-\beta_1}+e^{\theta-\beta_1}\cdot e^{\theta-\beta_2}}.$$

Обобщая полученное, найдем вероятность достижения тестируемым результата x_{ij} (т.е. того, что тестируемый i выполнит ровно x шагов в задании j):

$$\pi_{ijx} = \frac{\prod_{k=0}^{x_{ij}} e^{\theta_i-\beta_{jk}}}{\sum_{l=0}^{x_j^{\max}} \prod_{k=0}^{l} e^{\theta_i-\beta_{jk}}}$$

или

$$\pi_{ijx} = \frac{e^{\sum\limits_{k=0}^{x_{ij}} (\theta_i - \beta_{jk})}}{\sum\limits_{l=0}^{x_j^{\max}} e^{\sum\limits_{k=0}^{l} (\theta_i - \beta_{jk})}}, \qquad (11)$$

где $k = 0, 1 \ldots x_{ij} \ldots x_{\max}$ — количество шагов; $x_j^{\max}$ — максимально возможное количество баллов за задание j;

Уравнение (11) представляет собой модификацию модели Раша для заданий с градацией степени правильности ответа. Подставляя в уравнение (11) значения соответствующие значения β_{jk} (уровней трудности k-го шага в задании j), можно рассчитать вероятность получения i-м испытуемым x баллов в j-том задании.

Подбор параметров модели измерения

Суть измерения на основе математической модели заключается в следующем: полагая, что результаты выполнения тестовых заданий соответствуют выбранной математической модели, нужно найти такие значения параметров модели, при которых результаты конкретного тестирования и математическая модель совпадают наилучшим образом. Математическую основу поиска решения может составить метод максимального (наибольшего) правдоподобия или метод наименьших квадратов.

Критерии подбора параметров модели по методу максимума правдоподобия

Метод максимума правдоподобия можно сформулировать так: наилучшее описание явления то, которое даёт наибольшую вероятность получить в результате измерений именно те значения, которые и были фактически получены[3]. В формальной записи этот метод может быть представлен в виде максимума произведения вероятностей всех наблюдаемых независимых событий[4]:

$$p_1 \cdot p_2 \cdot \ldots \cdot p_N = \prod_{i=1}^{N} p_i \to \\ \to \max. \qquad (12)$$

По методу максимального правдоподобия наилучшим будет признан тот набор значений θ_i и β_{jk}, при котором произведение расчётных вероятностей фактически полученных результатов π_{ijx} максимально:

$$F = \prod_{i=1}^{n} \prod_{j=1}^{m} \pi_{ijx} = \\ = \prod_{i=1}^{n} \prod_{j=1}^{m} \frac{e^{\sum\limits_{k=0}^{x_{ij}} (\theta_i - \beta_{jk})}}{\sum\limits_{l=0}^{x_j^{\max}} e^{\sum\limits_{k=0}^{l} (\theta_i - \beta_{jk})}} \to \max, \qquad (13)$$

где n — количество испытуемых; m — число тестовых заданий.

Математические модели, построенные на основе модели

3 *Львовский Б.Н.* Статистические методы построения эмпирических формул. М.: Высшая школа, 1988.

4 *Айвазян С.А., Мхитарян В.С.* Прикладная статистика и основы эконометрики. М.: ЮНИТИ, 1998.

Г. Раша, оперирует разностью параметров θ и β_x. Поэтому возможно бесконечное количество равноценных решений, отличающихся на произвольную постоянную величину. Для определённости целесообразно ввести ограничение, например, в виде равенства нулю среднего значения уровня подготовленности всех испытуемых:

$$\bar{\theta} = 0. \qquad (14)$$

В этом случае наилучшее решение будет единственным. Для реализации в вычислительных алгоритмах желательно преобразовать ограничение (14) в неравенство:

$$\left|\bar{\theta}\right| \le \varepsilon, \qquad (15)$$

где ε — некоторая допустимо малая величина, например равная 0,001.

Выражений (13) и (15) достаточно для непосредственного численного решения задачи подбора параметров модели.

Решение можно упростить путём логарифмирования. Так как логарифм — монотонно возрастающая функция, то максимум функции правдоподобия совпадает с максимумом её логарифма:

$$F_l = \ln F =$$

$$= \ln\left(\prod_{i=1}^{n}\prod_{j=1}^{m}\frac{e^{\sum_{k=0}^{x_{ij}}(\theta_i-\beta_{jk})}}{\sum_{l=0}^{x_j^{\max}} e^{\sum_{k=0}^{l}(\theta_i-\beta_{jk})}}\right) \to \max,$$

$$F_l = \sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij}\theta_i - \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{x_{ij}}\beta_{jk} -$$

$$-\sum_{i=1}^{n}\sum_{j=1}^{m}\ln\left[\sum_{l=0}^{x_j^{\max}} e^{\sum_{k=0}^{l}(\theta_i-\beta_{jk})}\right] \to \max, \qquad (16)$$

где $\sum_{k=0}^{x_{ij}}\beta_{jk} = \sum_{k=1}^{x_{ij}}\beta_{jk}$, поскольку $\beta_{j0}=0$.

Нетрудно заметить, что $\sum_{k=1}^{x_{ij}}\beta_{jk}$ — сумма уровней трудности шагов, выполненных i-м тестируемым в задании j. Соответственно, $\sum_{i=1}^{n}\sum_{k=1}^{x_{ij}}\beta_{jk}$ — сумма уровней трудности шагов, выполненных всеми испытуемыми в задании j. Введём обозначение: s_{jk} — число тестируемых, закончивших шаг k в задании j. Тогда

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{x_{ij}}\beta_{jk} = \sum_{j=1}^{m}\sum_{k=1}^{x_{ij}} s_{jk}\beta_{jk}.$$

Также обозначим R_i — сумма баллов, набранных i-м тестируемым (R_i также называют исходным баллом) $R_i = \sum_{j=1}^{m} x_{ij}$.

Тогда логарифмическая функция правдоподобия может быть записана в виде:

$$F_l = \sum_{i=1}^{n} R_i\theta_i - \sum_{j=1}^{m}\sum_{k=1}^{x_{ij}} s_{jk}\beta_{jk} -$$

$$-\sum_{i=1}^{n}\sum_{j=1}^{m}\ln\left[\sum_{l=0}^{x_j^{\max}} e^{\sum_{k=0}^{l}(\theta_i-\beta_{jk})}\right] \to \max. \qquad (17)$$

В точке максимума производные функции равны нулю. Для ускорения поиска экстремума найдём частные производные логарифмической функции правдоподобия по каждому из аргументов.

$$\frac{\partial F_l}{\partial \theta_i} = \frac{\partial\left[\sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij}\theta_i - \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{x_{ij}}\beta_{jk} - \sum_{i=1}^{n}\sum_{j=1}^{m}\ln\left(\sum_{l=0}^{x_j^{\max}} e^{\sum_{k=0}^{l}(\theta_i-\beta_{jk})}\right)\right]}{\partial\theta_i} = R_i -$$

$$-\sum_{j=1}^{m}\left[\frac{1}{\sum_{l=0}^{x_j^{\max}} e^{\sum_{k=0}^{l}(\theta_i-\beta_{jk})}}\cdot\frac{\partial\left(\sum_{l=0}^{x_j^{\max}} e^{\sum_{k=0}^{l}(\theta_i-\beta_{jk})}\right)}{\partial\theta_i}\right] = R_i - \sum_{j=1}^{m}\left[\frac{\sum_{l=0}^{x_j^{\max}}\left(e^{\sum_{k=0}^{l}(\theta_i-\beta_{jk})}\cdot\frac{\partial\left(\sum_{k=0}^{l}\theta_i-\beta_{ik}\right)}{\partial\theta_i}\right)}{\sum_{l=0}^{x_j^{\max}} e^{\sum_{k=0}^{l}(\theta_i-\beta_{jk})}}\right] =$$

$$= R_i - \sum_{j=1}^{m}\frac{\sum_{l=0}^{x_j^{\max}} l\cdot e^{\sum_{k=0}^{l}(\theta_i-\beta_{jk})}}{\sum_{l=0}^{x_j^{\max}} e^{\sum_{k=0}^{l}(\theta_i-\beta_{jk})}},$$

$$\frac{\partial F_l}{\partial\theta_i} = R_i - \sum_{j=1}^{m}\sum_{l=1}^{x_j^{\max}} l\cdot\pi_{ijk}. \tag{18}$$

Последняя часть уравнения (18) представляет собой сумму произведений количества баллов на вероятность их получения по всем заданиям для i-го тестируемого, т.е. математическое ожидание полученного им количества баллов. Следовательно, частная производная логарифмической функции по уровню подготовленности i-го тестируемого равна разности между фактическим и расчётным количеством баллов, набранных этим тестируемым.

$$\frac{\partial\left[\sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij}\theta_i - \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{x_{ij}}\beta_{jk} - \sum_{i=1}^{n}\sum_{j=1}^{m}\ln\left(\sum_{l=0}^{x_j^{\max}} e^{\sum_{k=0}^{l}(\theta_i-\beta_{jk})}\right)\right]}{\partial\beta_{jk}} = -s_{ij} -$$

$$-\sum_{j=1}^{m}\left[\frac{1}{\sum\limits_{l=0}^{x_j^{\max}} e^{\sum\limits_{k=0}^{l}(\theta_i-\beta_{jk})}}\cdot\frac{\partial\left(\sum\limits_{l=0}^{x_j^{\max}} e^{\sum\limits_{k=0}^{l}(\theta_i-\beta_{jk})}\right)}{\partial\beta_{jk}}\right]=-s_{ij}+\sum_{j=1}^{m}\frac{\sum\limits_{l=k}^{x_j^{\max}} e^{\sum\limits_{k=0}^{l}(\theta_i-\beta_{jk})}}{\sum\limits_{l=0}^{x_j^{\max}} e^{\sum\limits_{k=0}^{l}(\theta_i-\beta_{jk})}},$$

$$\frac{\partial F_l}{\partial\beta_{jk}}=-s_{ij}+\sum_{i=1}^{n}\sum_{l=k}^{x_j^{\max}}\pi_{ijk}. \qquad (19)$$

Последняя часть выражения (19) — это сумма вероятностей получения не менее k баллов всеми испытуемыми в j-м задании. Поэтому эта частная производная определяется разностью между расчётным и фактическим количеством тестируемых, выполнивших в j-том задании шаг k.

Уравнения (18) и (19) образуют совместно с ограничением (15) образуют математическую основу для нахождения параметров модели.