УДК 519.257

Н.С. Нифонтов, А.Е. Архипов

ОБ ИСПОЛЬЗОВАНИИ ПАРАМЕТРИЧЕСКИХ МОДЕЛЕЙ В ТЕСТАХ ЗАКРЫТОГО ТИПА

Введение

Современный мир развивается очень динамично. Появляются новые технологии, прикладные науки и предметные области. Разнообразность современных технологий, их тесная взаимосвязь и довольно высокая сложность требует от специалистов наличия глубоких знаний в различных областях современной науки для успешного решения поставленных задач.

Последние 50 лет большое внимание уделяют разнообразным информационным технологиям, которые дают возможность применить разнообразные методики и системы тестового контроля и диагностики для определения уровня квалификации конкретных специалистов, а также определения качества их знаний, и возможности решения поставленных перед ними задач.

В области дистанционного образования, система тестового контроля, аттестации, а также система оценивания знаний и умений в конкретных предметных областях является важным базовым элементом учебного процесса.

Постановка задачи

Одним из существенных препятствий на пути массового внедрения систем тестового контроля (СТК) в современных условиях является отсутствие объективных показателей и критериев качества полученных результатов.

Отсутствие эффективных методов интеллектуальной обработки и анализа полученных данных, большой багаж эмпирических методов и приемов построения и последующей обработки результатов тестового контроля не гарантирует достоверных результатов в современных условиях.

Поэтому сегодня остро стоит проблема разработки эффективной методологии построения систем тестового контроля, разработки моделей и методов объективной обработки результатов теста.

Главным этапом на пути решения поставленной задачи является формализация процедуры работы систем тестового контроля [2].

Рассмотрим СТК на основе тестов закрытого типа. Задача тестируемого в тестах закрытого типа это выбор правильного ответа из множества предложенных вариантов. Процедуру проведения теста можно описать как процесс измерения некой физической величины L (уровень знаний тестируемого).

Модель системы тестового контроля можно представить в виде системы:

$$\textit{СТК} =< I, L, F, V, Q >, \qquad (1)$$

где $Q = \{q_i\}, i = \overline{1, N}$ – множество вопросов, а $V = \{\nu_{11}, \ldots, \nu_{1m}\}, \ldots, \{\nu_{n1}, \ldots, \nu_{nm}\}$ – множество m вариантов возможных ответов на задание (содержащих лишь один правильный). Множество знаний $I = \{I_k\}, k = \overline{1, N}$ тестируемого характеризуется интегральным показателем L (уровень знаний).

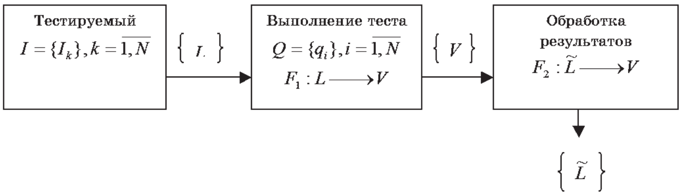

Рис. 1 – Система тестового контроля

Количественное значение уровня знаний L переводит переменную V процедура выполнения теста F_1, а процедура обработка результатов F_2 выполняет восстановление величины уровня знаний $F_2 : \tilde{L} \to V$.

Для идеального случая в тестах закрытого типа выполняется отношение вида $F_1 * F_2 = 1$, что позволяет на основе величины $\tilde{L}$ точно восстановить значение измеряемого параметра уровня знаний L. К сожалению, задание алгоритма обратного отображения $F_2 = F_1^{-1}$, является на практике невозможным, так как не известна природа механизма отображения $F_1 : L \to V$.

Поэтому на практике получаем соотношение вида $\tilde{L} = L + \Delta_{LF_2}$, где Δ_{LF_2} – ошибка тестирования, определяющая принципиальную возможность измерения с помощью построенной СТК нужной характеристики тестируемого.

Использование параметрических моделей

Как упоминалось выше, отсутствие эффективных методов обработки и анализа данных тестового опроса привело к появлению моделей совместного оценивания параметров тестируемого и параметров задания.

Теоретической основой моделей такого семейства является IRT (Item Response Theory). Оценивание уровня знаний L тестируемого и уровня сложности задания в IRT происходит на единой интегральной шкале.

Использование параметрических моделей позволяет измерить уровень знаний тестируемого независимо от уровня сложности задания, путем выявления связи между латентными параметрами тестируемого и результатами выполнения теста.

Для использования интервальной шкалы требуется введение классификации заданий по уровню их сложно (“силы”, ”веса”) β_j в соответствии с поставленными требованиями к тесту. В классической теории тестов характерным является определение уровня сложности задания β_j как отношение количества не справившихся с заданием к общему количеству испытуемых, т.е $\beta_j = \frac{N_j}{N}$. Таким образом уровень сложности задания β_j зависит от объема и качества репрезентативной выборки.

В современных условиях мы все чаще сталкиваемся со случаями, когда тестирование носит локальный характер. Тестирование проводится по определенному курсу или пройденной теме среди студентов одного курса и одной специальности. В результате репрезентативной выборкой будут все тестируемые. В таком случае, для качественной "настройки" системы тестового контроля нужно накапливать статистические данные. Процесс сбора статистики может растянуться во времени, за которое обычно требуется внесение корректив в тест (к примеру, изменения в учебной программе или устаревание теста).

Эти ограничения можно преодолеть с помощью параметрических моделей семейства IRT. Базовой для семейства IRT является модель Раша. В параметрических моделях используется понятие вероятности правильного ответа P_{ij} при решении i-м тестируемым j-го задания. Вероятность P_{ij} зависит от уровня знаний тестируемого L_i и уровня сложности задания β_j.

В однопараметрической (так как единственным параметром тут есть разница $(L_i - \beta_j)$) логистической модели Раша данная вероятность P_{ij} определяется лишь разностью уровней знаний тестируемого L и уровня сложности задания β_j и имеет вид:

$$P_{ij} = \frac{1}{1 + e^{-(L_i - \beta_j)}}. \qquad (2)$$

Также ее удобно представить в форме:

$$P_{ij} = \frac{e^{(L_i - \beta_j)}}{1 + e^{(L_i - \beta_j)}}. \qquad (3)$$

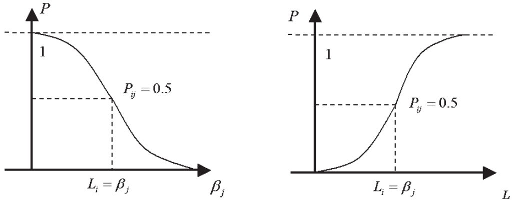

Рис. 2 – Характеристические кривые j-го задания и i-го тестируемого

Функция P_{ij} является монотонно возрастающей. С ростом уровня знаний тестируемого вероятность получить правильный ответ возрастает. Точкой перегиба характеристической кривой j-го задания будет точка где уровень знаний тестируемого равен уровню трудности задания, т.е будет выполняться отношение $L_i - \beta_j = 0$. Вероятность решения такого задания будет $P_{ij} = 0.5$.

Характеристическая кривая i-го тестируемого будет монотонно убывающей, так как вероятность успешно ответить на задание будет убывать с увеличением сложности задания β_j. Точкой перегиба для кривой будет $L_i - \beta_j = 0$. С увеличением уровня знаний тестируемого происходит сдвиг данной кривой вправо.

“Крутизна” характеристической кривой определяет различающую способность задания. Задания с более “крутыми” характеристическими кривыми позволяют эффективней различать тестируемых, особенно в среднем диапазоне, чем задания с “пологими” кривыми.

“Крутизну” характеристической кривой называют дифференцирующей силой задания.

В модели Раша все задания считаются одинакового различающими тестируемых. Для учета фактора “крутизны” характеристической кривой задания нужно дополнить модель Раша еще одним параметром – дифференцирующей силой задания.

Модель такого типа называется двухпараметрической моделью Бирнбаума. Вероятность правильного ответа на j-е задание i-м тестируемым определяется отношением вида:

$$P_{ij} = \frac{1}{1 + e^{-\alpha_j (L_i - \beta_j)}}, \tag{4}$$

где P_{ij}, L_i, β_j имеют такой же смысл, как и в модели Раша, а параметр α_j это и есть дифференцирующая сила задания.

Данное отношение удобно представить в форме:

$$P_{ij} = \frac{e^{\alpha_j (L_i - \beta_j)}}{1 + e^{\alpha_j (L_i - \beta_j)}}. \tag{5}$$

Применение двухпараметрической модели позволяет уменьшить фактор влияния низкоэффективных заданий с невысокой различающей способностью при определении уровня знаний тестируемого. Тем самым снимается одна из проблем при шкалировании заданий и теста – проблема неадекватных заданий.

Геометрический смысл параметра α_j сводиться к тангенсу угла наклона касательной в точке перегиба $L_i = \beta_j$ характеристической кривой задания. Соответственно интервал изменения параметра α_j может быть любым.

Характеристическая кривая задания обладает свойствами монотонности. С уменьшением уровня знаний тестируемого, вероятность правильного ответа на j-й вопрос падает, а при увеличении уровня знаний она подымается до уровня $P_{ij} = 1$.

С увеличением уровня сложности задания β_j вероятность ответа для одного и того уровня знаний уменьшается.

При небольших значениях параметра “крутизны” α_j характеристическая кривая принимает пологий вид и в результате для тестируемых с

 ISSN 1562-9945

уровнем знаний $L_i < \beta_j$, а также тестируемых с уровнем знания $L_i > \beta_j$, вероятности их успешного ответа на поставленный вопрос будут примерно равны.

С увеличением "крутизны" происходят существенные различия в вероятностях успеха. Что позволяет эффективно проводить тестирование.

Использование заданий с "крутизной" $\alpha_j < 0$ приводит к тому что, с увеличением уровня знаний тестируемого вероятность успешного ответа на данное задание уменьшается.

При построении СТК рекомендуется использовать задания с "крутизной" α_j характеристической кривой в диапазоне $[0.5 \ldots 3]$.

Одним из преимуществ использования однопараметрической модели Раша является ее наглядность, а именно наличие у нее монотонной зависимости сырого балла и конечной оценки уровня знаний тестируемого (зависимость нелинейная, связана с тем что достаточной статистикой для оценки параметров системы является сырой бал), что очень важно в условиях массового тестирования (государственный экзамен), так как является подтверждающим фактором валидности теста.

В двухпараметрической модели достаточной статистикой для оценки уровня знаний L_i тестируемого является сумма $\sum\limits_{j=1}^{m} b_{ij}\alpha_j$, где m – количество заданий в тесте а b_{ij} – сырой бал ответа i-го тестируемого на j-й вопрос. Следствием этого, является не наглядность модели (тестируемые набравшие одинаковый сырой бал могут получить разный окончательный бал).

Этап "настройки" (оценивания параметров) системы тестового контроля носит итерационный характер. Построение характеристических кривых позволяет исследователю сделать выводы и рекомендации касательно чистки заданий, корректировки заданий (нехватки заданий определенной трудности).

Оценивание параметров модели

Одним из важных этапов в "настройке" СТК является процедура оценки параметров.

Результаты проведения теста из M вопрос среди N тестируемых можно представить в виде бинарной матрицы ответов A, где $a_{ij} = 1$ соответствует правильному ответу на вопрос а $a_{ij} = 0$ не правильному:

$$A = \begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1M} \\ a_{21} & a_{22} & \ldots & a_{2M} \\ \ldots & \ldots & \ldots & \ldots \\ a_{N1} & a_{N2} & \ldots & a_{NM} \end{vmatrix}. \tag{6}$$

Вероятность того, что $a_{ij} = 1$ будет $P_{a_{ij}=1} = P_{ij}$, а вероятность принятие "0" – $P_{a_{ij}=0} = 1 - P_{ij} = Q_{ij}$.

Использую бинарную матрицу ответов A и метод наибольшего правдоподобия можно произвести оценки латентных параметров системы тестового контроля [3,4].

Оценки параметров системы по своему характеру являются случайными величинами (связано с тем, что значения матрицы ответов сами являются случайными величинами). Важным свойством оценки является ее несмещенность. У несмещенных оценок математическое ожидание оценки параметров системы совпадает с точными значениями (не зависимо от объема репрезентативной выборки):

$$M\{\hat{L}\} = L \text{ и } M\{\hat{\beta}\} = \beta. \tag{7}$$

Что позволяет избежать сдвига относительно истинных значений параметров системы. Эффективность оценки характеризуется дисперсией, если оценка обладает минимальной дисперсией, то она называется эффективной. Поэтому главной задачей стоит нахождение несмещенных оценок с минимальной дисперсией (на практике часто ищут асимптотически несмещенные и асимптотически эффективные оценки).

Если несмещенную оценку не удается найти, то можно использовать смещенные оценки с достаточно большой дисперсией.

Если в результате оценивания параметров системы методом максимального правдоподобия, будут найдены оценки, то они будут подчиняться нормальному распределению, будут несмещенными и эффективными (если эффективные оценки существуют, иначе будут получены асимптотически эффективные).

За требуемые оценки параметров системы нужно выбирать такие значения, при которых функция правдоподобия принимает наибольшее свое значение.

В однопараметрической модели Раша функция правдоподобия для элемента матрицы a_{ij} (дискретная случайная величина) будет зависеть от параметров L_i, β_j. Ее можно представить виде произведения вероятностей:

$$L\{a_{ij}|L_i,\beta_j\} = \prod_{i=1}^{N}\prod_{j=1}^{M} P\{a_{ij}|L_i,\beta_j\} = \frac{e^{\sum\limits_{i=1}^{N}\sum\limits_{j=1}^{M} a_{ij}(L_i-\beta_j)}}{\prod\limits_{i=1}^{N}\prod\limits_{j=1}^{M}(1+e^{(L_i-\beta_j)})}. \tag{8}$$

Использование $Ln(L\{a_{ij}|L_i,\beta_j\})$ позволяет упростить вид исходного выражения ($Ln(L\{a_{ij}|L_i,\beta_j\})$ и $L\{a_{ij}|L_i,\beta_j\}$ достигают максимального значения при одинаковых значениях аргументов):

$$\ln\left(L\{a_{ij}|L_i,\beta_j\}\right) = \sum_{i=1}^{N}\sum_{j=1}^{M} a_{ij}(L_i-\beta_j) - \sum_{i=1}^{N}\sum_{j=1}^{M} \ln(1+e^{(L_i-\beta_j)}). \tag{9}$$

Методом максимального правдоподобия составляем систему из $(N + M)$ уравнений частных производных $Ln(L\{a_{ij}|L_i,\beta_j\})$:

 ISSN 1562-9945

$$\begin{cases} \frac{\partial \ln(L\{a_{ij}|L_i,\beta_j\})}{\partial L_i} = \sum\limits_{j=1}^{M} a_{ij} - \sum\limits_{j=1}^{M} \frac{e^{(L_i-\beta_j)}}{1+e^{(L_i-\beta_j)}} = \sum\limits_{j=1}^{M} a_{ij} - \sum\limits_{j=1}^{M} P_{ij} = 0 \\ \frac{\partial \ln(L\{a_{ij}|L_i,\beta_j\})}{\partial \beta_j} = -\sum\limits_{i=1}^{N} a_{ij} - \sum\limits_{i=1}^{N} \frac{e^{(L_i-\beta_j)}}{1+e^{(L_i-\beta_j)}} = -\sum\limits_{i=1}^{N} a_{ij} - \sum\limits_{i=1}^{N} P_{ij} = 0 \end{cases}. \quad (10)$$

Решение данной системы уравнений не вызывает затруднений и позволяет найти требуемые оценки если они существуют.

По аналогичной схеме происходит оценивание параметров системы для двухпараметрической модели Бирнбаума. Функция правдоподобия для элемента матрицы a_{ij} зависит от параметров L_i, β_j, α_j и будет носить смысл плотности совместного распределения наблюдаемых случайных величин.

Функцию правдоподобия удобно представить в виде:

$$L\{a_{ij}|L_i,\beta_j,\alpha_j\} = \prod_{i=1}^{N}\prod_{j=1}^{M} P\{a_{ij}|L_i,\beta_j,\alpha_j\} = \frac{e^{\sum\limits_{i=1}^{N}\sum\limits_{j=1}^{M} a_{ij}\alpha_j(L_i-\beta_j)}}{\prod\limits_{i=1}^{N}\prod\limits_{j=1}^{M}\left(1+e^{\alpha_j(L_i-\beta_j)}\right)}. \quad (11)$$

Оценками латентных параметров системы L_i, β_j, α_j будут такие значения $\hat{L}_i$, $\hat{\beta}_j$, $\hat{\alpha}_j$ (определены с точность до линейных преобразований вида $\hat{L}_i^{'} = a\hat{L}_i + b$, $\hat{\beta}_j^{'} = a\hat{\beta}_j + b$, $\hat{\alpha}_j^{'} = \frac{\hat{\alpha}_j}{a}$) при которых функция правдоподобия достигает значения $L\{a_{ij}|\hat{L}_i,\hat{\beta}_j,\hat{\alpha}_j\} = \mathrm{MAX}$.

Учитывая линейность преобразований нужно найти оценки параметров при условии, что среднее значение оценки уровня знаний будет равно “0” и несмещенная оценка дисперсии среднего значения будет равна “1”:

$$\bar{\hat{L}}_i = 0 \text{ и } D\{\hat{L}_i\} = \frac{\sum\limits_{i=1}^{N}(\hat{L}_i - \bar{\hat{L}}_i)}{N} = 1. \quad (12)$$

Методом максимального правдоподобия составляем систему из $(N + 2M)$ уравнений частных производных $Ln(L\{a_{ij}|L_i,\beta_j,\alpha_j\})$.

$$\begin{cases} \frac{\partial L\{a_{ij}|L_i,\beta_j,\alpha_j\}}{\partial L_i} = 0 \\ \frac{\partial L\{a_{ij}|L_i,\beta_j,\alpha_j\}}{\partial \beta_j} = 0 \\ \frac{\partial L\{a_{ij}|L_i,\beta_j,\alpha_j\}}{\partial \alpha_j} = 0 \end{cases}, \ i = \overline{1,N}, j = \overline{1,M}. \quad (13)$$

Решение системы уравнений даст искомые значения оценок. Для двухпараметрической модели эти уравнения являются нелинейными и их решение можно получить только при использовании численных методов и алгоритмов. С увеличением репрезентативной выборки происходит существенное увеличение количества уравнений, что вызывает трудности при использовании известных численных методов.

В СТК на основе тестов закрытого типа возможно применение стратегии угадывания при ответе на вопрос. Что негативно сказывается на оценке параметров системы (особенно в условиях тестирования с небольшой репрезентативной выборкой).

Одним из возможных способов решения данной проблемы является усложнение параметрических моделей путем введения дополнительных параметров характеризующих тестируемого и задания. Главными недостатками такого подхода к решению задачи является усложнение процедуры оценки параметров системы, не наглядность полученных результатов (сложная зависимость сырого балла от конечной оценки, не интуитивная валидность результатов) а также существенное увеличение объема репрезентативной выборки.

Другим подходом к решению данной проблемы является введение дополнительных методов и моделей, характеризующих поведение тестируемого и задания в условиях возможного применения стратегии угадывания (исследование стратегий поведения тестируемого – "*двоечники*", "*середнячки*", "*отличники*") и возможного наличия перекрестной информации в вопросах (исследование наличия и учет корреляции между уровнем знаний и перекрестной информацией).

Выводы

Применение параметрических моделей в тестах закрытого типа позволяет построить эффективные методы обработки и анализа данных тестового опроса. Характер тестов закрытого типа требует введение дополнительных методов и приемом адаптации параметрических моделей для построения эффективной СТК.

Литература

1. Аванесов В.С. Тесты в социологическом исследовании. –М.: Наука, 1982.-200 с.
2. Архипов А.Е. Формализация и моделирование тестовых систем закрытого типа. // Системні технології ті технічної діагностики: Збірн. Наук. Праць. – Вип.5. – Дніпропетровськ, 1999. – С. 29-33
3. Нейман Ю.М., Хлебников В.А. Введение в теорию моделирования и параметризации педагогических тестов. – М : Прометей, 2000. – 169 с.
4. Гмурман В.Е. Теория вероятностей и математическая статистика. - М.: Высшая школа, 1997, – 480 с.

Получено 08.11.2007

 ISSN 1562-9945