ЕЩЕ РАЗ О КОРРЕКЦИИ ТЕСТОВОГО БАЛЛА

Виктор Кромер

kromer@newmail.ru

Опубликовано в ж. «Педагогические Измерения», №1, 2007 г. С. 89-94.

Аннотация

Анализируются недостатки системы коррекции тестовых баллов, проводимой без различения отказа испытуемого от ответа и неверного ответа. Анализируются вероятности угадывания верного ответа в заданиях различных форм и формулируются требования к заданиям, исходя из необходимого ограничения вероятности угадывания.

Ключевые слова: тесты, скорректированная матрица, угадывание.

В данное время, актуальна, как никогда ранее, проблема угадывания испытуемыми верных ответов к тестовым заданиям. Под угадыванием в широком смысле понятия понимается не только выбор испытуемым ответа на основе случайного выбора, как угадывается число, выпадающее при бросании игральной кости, но и выбор ответа с привлечением наличного неполного знания из тестируемой области знания.

Естественно начать с определения понятий, которыми мы будем пользоваться в дальнейшем. Истинной назовем матрицу тестирования, в которой угадывание не имеет места. Истинная матрица – гипотетическое понятие, при обработке результатов приходится сталкиваться в основном с искаженными матрицами, но по определению, лишь характеристики испытуемых и заданий, извлеченные из истинной матрицы, являются несмещенными оценками истинных характеристик. Искаженная матрица – матрица при реализации в процессе заполнения и обработки подхода, при котором не различаются (оцениваются нулем) пропуски задания и неправильный ответ.

Скорректированной назовем матрицу с внесенными в нее отрицательными корректирующими элементами при реализации описанного ранее[1] подхода. Предлагавшийся подход заключался во внесении в матрицу тестовых результатов (традиционно

[1] Кромер В.В. О некоторых вопросах тестовых технологий // Тез. докл. Второй Всеросс. конфер. "Развитие системы тестирования в России", г. Москва 23–24 ноября 2000 г. – Ч. 4. – М: Прометей, 2000. – С. 59–61.

заполняемую лишь нулями и единицами, в зависимости от исхода "противоборства" испытуемого с заданием), отрицательных дробных величин $\left(-\frac{1}{m_j - 1}\right)$ – корректирующих элементов за ошибочные ответы, где m_j – количество вариантов ответа на j-е задание, при этом m_j может меняться для различных заданий.

Заполнение матрицы единицами, нулями и корректирующими элементами и последующий подсчет сумм (с учетом знаков) элементов матрицы по строкам и столбцам дает несмещенные значения, математическими ожиданиями которых являются гипотетические истинные баллы испытуемого и задания, что позволяет объективно оценить уровни подготовленности испытуемых и уровни трудности заданий.

Необходимо также определить, хотя бы примерно, вероятность угадывания верного ответа, которой можно пренебречь вследствие ее малости. Все необходимые ориентировочные расчеты можно произвести в рамках классической статистической теории измерений. Имеем сбалансированный (по трудности заданий) тест длиной в k заданий. Оптимальное стандартное отклонение тестовых баллов на согласованной с тестом выборке составит около $\sigma_{опт} = \frac{k}{6}$ (известное правило «трех сигм»).[1] Выборка считается согласованной с тестом при нормальности распределения тестовых результатов. Известно, что кривая распределения индивидуальных баллов, получаемых на репрезентативной выборке, является следствием распределения трудности заданий теста. Таким образом, согласованная со стандартизированным нормативно-ориентированным тестом выборка – это случайная выборка из генеральной совокупности испытуемых, по размеру соответствующая мере трудности крайних тестовых заданий. Известно также, что размах (разница между наибольшим и наименьшим значениями в наборе наблюдений) зависит от размера набора наблюдений. Зависимости размаха от размера выборки изучены теоретически и табулированы. При наиболее часто встречающихся выборках в несколько сотен испытуемых значение размаха близко к шести стандартным отклонениям (плюс/минус три "сигмы" от среднего значения).

[1] Принято, что тестовые задания оцениваются дихотомично (0 или 1). При наличии в тесте оцениваемых политомично (0, 1, 2 …) заданий достаточно полагать, что k – максимально достижимый тестовый балл.

При надежности теста порядка $r = 0{,}9$ стандартная ошибка тестового результата определится как $s_e = s_y\sqrt{1-r} = \frac{k}{6}\sqrt{1-0{,}9} \approx 0{,}05k$. Ясно, что искажение результата на основе угадывания, меньшее в несколько раз, не повлияет на общую погрешность измерения, откуда вытекает требование ограничить вероятность угадывания верного ответа на одно задание одним процентом.

Для некоторых форм тестовых заданий это требование выполняется без особых на то усилий, для некоторых форм оно невыполнимо в принципе. Рассмотрим известные формы заданий и рассчитаем вероятность угадывания верного ответа.

а) Задания открытой формы.

В заданиях этой формы вероятность угадать правильный ответ практически равна нулю. В самом деле, вписать верный ответ в задании «Коня Александра Македонского звали __________» без знания соответствующего фактического материала невозможно.

б) Задания на установление соответствия.

Левый столбец имеет a_1 элементов, правый a_2 элементов. Угадать правильное соответствие первому элементу левого столбца можно с вероятностью $\frac{1}{a_2}$, второму – с вероятностью $\frac{1}{a_2-1}$, третьему – с вероятностью $\frac{1}{a_2-2}$ и т.д. до исчерпания всех элементов левого столбца. Вероятность выполнить правильно задание (т.е. угадать все соответствия), равна $p = \prod_{i=1}^{a_1}\frac{1}{a_2+1-i} = \frac{(a_2-a_1)!}{a_2!}$. При $a_2 = a_1$ вероятность угадать соответствие последнему элементу левого столбца равна 1 (нежелательный случай), откуда вытекает известное условие $a_2 > a_1$. При минимально допустимом $a_2 - a_1 = 1$ условию $p < 0{,}01$ соответствует $a_2 \geq 5$. Таким образом, задание с 4 элементами в левом столбце и 5 элементами в правом столбце уже вполне удовлетворительно с позиций устранения угадывания.

в) Задания на установление правильной последовательности.

Всего в последовательности a элементов. Первый элемент последовательности можно угадать с вероятностью $\frac{1}{a}$, второй – с вероятностью $\frac{1}{a-1}$ и т.д. Общая вероятность угадывания правильного ответа составит $p=\frac{1}{a!}$. Условие $p<0{,}01$ выполняется при $a \geq 5$. Таким образом, при числе элементов в последовательности, большей или равной 5, влияние угадывания на результат практически исключается.

г) Задания с выбором нескольких правильных ответов.

К заданиям предусмотрено *а* ответов. Поскольку на каждый из приведенных вариантов ответа можно наугад дать как правильный, так и неправильный ответ, вероятность правильного угадывания всех ответов составляет $p=\frac{1}{2^a}$. При наложении ограничения (хотя-бы один неверный ответ должен быть[1]) вероятность правильного выбора верных ответов составит $p=\frac{1}{2^a-1}$ (исключается вариант ответа с подтверждением всех приведенных ответов). Наложив еще одно ограничение (наличие хотя бы одного правильного ответа), получаем $p=\frac{1}{2^a-2}$. Условию $p<0{,}01$ соответствует $a \geq 7$.

д) Задания с выбором одного правильного ответа.

Количество вариантов ответа равно *а*. Задания подобной формы являются самыми распространенными, и вероятность угадывания равна $p=\frac{1}{a}$. Вероятность угадывания недопустимо велика, поскольку $a \leq 6$[2], и не снижается ниже 0,16, что и порождает известные проблемы. Ввиду большой вероятности угадывания в заданиях подобной формы прямое использование тестовых результатов невозможно. Необходим либо отказ от заданий подобной формы, либо коррекция результатов. В рамках классической статистической теории измерений коррекция производится либо вычитанием из индивидуальных тестовых результатов величины коррекции[3], либо внесением в матрицу, наряду с нулями (за отказ от ответа) и единиц (за верный ответ), также отрицательных корректирующих элементов за неверные ответы[4].

---

[1] Аванесов В.С. Форма тестовых заданий. – М.: Центр тестирования, 2005. – С. 64.

[2] Аванесов В.С. Форма тестовых заданий. – М.: Центр тестирования, 2005. – С. 46–47.

[3] Suen H.K. Principles of Test Theories. – Hillsdale, NJ: Erlbaum, 1990. – С. 73.

[4] Кромер В.В. О некоторых вопросах тестовых технологий // Тез. докл. Второй Всерос. конф. "Развитие системы тестирования в России", г. Москва 23–24 ноября 2000 г. – Ч. 4. – М: Прометей, 2000. – С. 59–61.

В рамках Item Response Theory коррекция проводится введением в формулу функции успеха соответствующего параметра, характеризующего вероятность угадывания для данного задания[1], либо двух параметров, характеризующих раздельно вероятность угадывания для задания и склонность испытуемого к угадыванию[2].

В IRT при учете явления угадывания видом функции успеха нет прямой необходимости в разделении между отказом от ответа и неверным ответом, поскольку в зависимости от профилей испытуемых (и заданий) будут подобраны значения наиболее вероятных параметров угадывания, при которых реализуются данные профили.

Сказанное неприемлемо при обработке результатов в рамках классической статистической теории измерений. Неразличение отказа от ответа и ошибки приводит к дискриминации добросовестных испытуемых, т.е. испытуемых, не прибегающих к угадыванию, откуда и проистекают известные рекомендации по заполнению наугад матрицы в сомнительных случаях[3].

Иногда приходится сталкиваться с рекомендациями по коррекции, предполагающими введение коррекции без учета степени добросовестности испытуемых. Так, в работе[4] рассматривается несколько вариантов коррекции с однозначной зависимостью между скорректированным и нескорректированным значениями. Вне зависимости от вида вводимой коррекции (с фиксированной поправкой, линейная/нелинейная модель) введение подобных поправок не имеет заметного смысла, поскольку шкала тестовых баллов классической модели (в общем случае) порядковая, и никакое монотонное преобразование тестового балла не изменит порядка следования испытуемых в ранжированном ряду.

Нежелание проводить различие между отказом от ответа и неверным ответом приводит к педагогическому просчету – часть реальных достижений слабых, но добросовестных испытуемых "списывается" на не имевшее место угадывание, а тем самым добросовестность наказывается. Подобная "коррекция" приносит больший педагогиче-

---

[1] Suen H.K. Principles of Test Theories. – Hillsdale, NJ: Erlbaum, 1990. – С. 92.

[2] Кромер В.В. О многопараметрической оценке уровней подготовленности испытуемых и трудности заданий // Педагогические измерения. – 2005. – № 3. – С. 65–72. – С. 72.

[3] Шаповал В.В. О нравственной оценке одной тактики "высшего результата" при сдаче ЕГЭ // Вопросы тестирования в образовании. – 2002. – № 2. – С. 252–260.

[4] Ким В.С. Коррекция тестовых баллов на угадывание // Педагогические измерения. – 2006. – № 4.

ский вред, чем отсутствие коррекции, где знающие "правила игры" испытуемые получают за счет случайного угадывания незаслуженные баллы.

Некорректность модели следует из сопоставления формул (7) и (8) в рассматриваемой статье[1]. Формула (8) не следует из формулы (7), поскольку применение формулы (8) предполагает раздельный учет неверных ответов и отказа от ответа, а в формуле (7), согласно позиции автора, p и q – всего лишь доли единиц и нулей в профиле испытуемого без упомянутого выше раздельного учета. Между тем, смысл и назначение коррекции заключается именно в изменении полученного без коррекции порядка ранжирования испытуемых. При правильно проведенной коррекции испытуемый, чаще прибегающий к угадыванию, чем некий среднестатистический испытуемый, отодвигается ближе к концу рангового списка, а прибегающий к угадыванию реже или совсем отказывающийся от него продвинется ближе к началу списка. Иными словами, добродетель будет вознаграждена, а порок наказан.

Сомнительна также педагогическая основа более жесткой реакции на угадывание для слабых испытуемых, в случае использования т.н. параболической и кубической моделей[2]. Но модель, в силу наложенного условия по подбору коэффициента μ ,обеспечивает одинаковую реакцию для самых слабых испытуемых, а тем самым дает преимущество с уменьшением необходимой коррекции всем без исключения испытуемым, и в тем большей степени, чем испытуемый сильнее. Это в случае, если путем особых ухищрений удастся добиться интервальной тестовой шкалы. В более общем случае порядковой шкалы, как уже говорилось, любое монотонное преобразование тестовых баллов бессмысленно.

Внедрение модели приведет лишь к снижению валидности тестовых результатов. В рассматриваемой модели заложено рациональное зерно – варьирование степени применения полной наиболее справедливой коррекции. Это имеет смысл делать в переходный период внедрения коррекции на угадывание, хотя эта модель, по нашему мнению, не отвечает своему назначению, да и степень коррекции необходимо варьировать не параметром n модели, а параметром μ при $n = 1$.

[1] Ким В.С. Коррекция тестовых баллов на угадывание // Педагогические измерения. – 2006. – № 4.
[2] Ким В.С. Коррекция тестовых баллов на угадывание // Педагогические измерения. – 2006. – № 4.