

Математические модели оценки знаний

Александров М.А.

Научный руководитель: старший преподаватель кафедры ЭМИС Матолыгин А.А.

Томский государственный университет систем управления и радиоэлектроники

Кафедра экономической математики, информатики и статистики.

Аннотация: в работе представлено описание стандартной модели математической оценки знаний и модель современной теории тестов (IRT), с помощью которой проводится оценка знаний с учетом множества факторов.

Ключевые слова: IRT, математическая оценка, логит.

Введение

В наше время существует множество разных математических моделей и подходов, служащие для оценки знаний. Оцениваемые учатся обманывать простые системы, чтобы добиться желаемого результата. Из-за этого модели усложняются, меняются.

Для этого используется теория вероятностей и математическая статистика, теория графов, теория нечетких множеств и популярная ныне нечеткая логика.

Можно разделить модели оценки знаний на три больших класса:

- модель оценки уровня знаний;
- модель диагностики знаний;
- модель классификации.

Целью данной работы состоит в том, чтобы улучшить программу тестирования студентов оптимизации его модуля математической модели оценки знаний [1].

Простая модель оценки уровня знаний

Данная модель является самой простой и самой распространённой. Ответ студента на каждое задание оценивается по двухбалльной (правильно или неправильно) или много балльной (чаще всего пятибалльной) шкале. Оценка выставляется путем вычисления значения R .

$$R = \frac{\sum_{i=1}^k R_i}{n},$$

где R_i -правильный ответ обучаемого на i -е задание; k -количество правильных ответов из n предложенных ($k \leq n$), которое затем обычно округляется по правилам математики.

Окончательная оценка, как правило, определяется по формуле:

$$I = \begin{cases} 1, R \leq c_1 \\ \dots \\ M, R > c_{M-1} \end{cases},$$

где I – окончательная оценка, $\{c_1, \dots, c_M\}$ – вектор граничных значений, M -максимально возможная оценка (например $M=5$).

К достоинствам данной модели следует отнести простоту ее реализации. Но недостаток заключается в том что она зависит от количества правильных ответов. Из-за этого она не позволяет объективно оценить знания студента.

Метод линейно-кусочной аппроксимации

Для улучшения оценки знаний был разработан алгоритм основанный на классификации заданий по их дидактическим характеристикам (значимость (z), трудность (d), спецификация (s)), который используется в современных онлайн тестированиях. Число баллов, полученных студентом за выполнение n заданий, определяется по формуле:

$$y = \sum_{i=1}^n w_i x_i,$$

где x_i – оценка за выполнение i -го задания; n – число заданий; $W = \{w_i\}$ – вектор весовых коэффициентов заданий, зависящий от их дидактических характеристик.

По завершению контроля определяется средний балл A , полученный студентом за выполнение n заданий ($A = y / k_n$, где k_n – количество попыток выполнения n заданий, $k_n \geq n$) и уточненный средний балл A' :

$$A' = A + a_1 r + a_2 \frac{k_n}{n} + a_3 \frac{k_c}{n} + a_4 \frac{k_b}{n},$$

где r – ранг обучаемого (1, 2, или 3); k_n – количество попыток выполнения n заданий; k_c – количество обращений к справочной информации; k_b – количество заданий, выполненных с превышением отведенного времени ($k_b \leq n$); a_1, a_2, a_3, a_4 – коэффициенты.

Далее значения уточненного среднего балла с помощью вектора граничных значений переводятся в обычную пятибалльную шкалу.

Параметры контроля, а именно значения весовых коэффициентов w_i , коэффициентов a_i и значения элементов вектора граничных значений определяются на этапе обучения по результатам контрольного эксперимента[2].

Модели современной теории тестов

В данной работе в качестве оптимизации оценки знаний студентов была выбрана модель современной теории тестов IRT.

В отличие от классической теории, где индивидуальный балл тестируемого рассматривается как постоянное число, в IRT латентный параметр трактуется как некоторая переменная. Начальное значение параметра получается непосредственно из эмпирических данных тестирования. Переменный характер измеряемой величины указывает на возможность последовательного приближения к объективным оценкам параметра с помощью тех или иных итерационных методов.

В рамках основного предположения IRT устанавливается связь между латентными параметрами испытуемых и наблюдаемыми результатами выполнения теста. При установлении связи важно понимать, что первопричиной являются латентные параметры. Если говорить точно, то взаимодействие двух множеств значений латентных параметров порождает наблюдаемые результаты выполнения теста.

Элементы первого множества – это значения латентного параметра, определяющего уровень подготовки N испытуемых θ_i ($i = 1, 2, \dots, N$). Второе множество образуют значения латентного параметра β_j ($j = 1, 2, \dots, n$), равные трудностям n заданий теста.

Однако на практике всегда ставится обратная задача: по ответам испытуемых на задания теста оценить значения латентных параметров θ и β . Для ее решения нужно ответить по меньшей мере на два вопроса.

Первый связан с выбором вида соотношения между латентными параметрами θ и β . Идея установления соотношения принадлежит датскому математику Г. Рашу, который предложил ввести его в виде разности $(\theta - \beta)$, предполагая, что параметры θ и β оцениваются в одной и той же шкале.

Ответ на второй вопрос, который является центральным в IRT, связан с выбором математической модели для описания рассматриваемой связи между латентными параметрами и наблюдаемыми результатами выполнения теста.

В частности, можно рассматривать условную вероятность правильного выполнения i -м испытуемым с уровнем подготовки θ_i различных по трудности заданий теста, считая θ_i параметром i -го ученика, а β – независимой переменной. В этом случае условная вероятность будет функцией латентной переменной β :

$$P_j(x_{ij} = 1 | \beta_j) = \varphi(\theta - \beta_j), j = 1, 2, \dots, n.$$

В теории IRT функции $f(\beta)$ и $\varphi(\theta)$ получили название Item Response Functions (IRF). Специальное название имеют и их графики. График функции P_j – это характеристическая кривая j -го задания (ICC), а график функции P_i – индивидуальная кривая i -го испытуемого (PCC).

При выборе вида функций P_i и P_j учитываются обстоятельства как эмпирического, так и математического характера. В предположении нормального распределения значений латентных переменных θ и β таких функций предлагаются две. Одна из них, обычно обозначаемая $\varphi(x)$, относится к семейству логистических кривых, другая $\Phi(x)$, является интегральной функцией нормированного нормального распределения. Поскольку для одних и тех же значений x ординаты точек графиков функций $\Phi(x)$ и $\varphi(1,7x)$ отличаются друг от друга достаточно мало, то в том, что их две, нет ни ошибки, ни противоречия. А именно, для всех x , принадлежащих области определения этих функций,

$$|\Phi(x) - \varphi(1,7x)| < 0,01.$$

Наиболее сильный аргумент в пользу логистической функции связан не с качеством измерений, а с относительной простотой ее аналитического задания, выгодной при оценивании параметров θ и β . Поэтому в практических приложениях предпочтение обычно отдают функции $\varphi(1,7x)$.

Число параметров, входящих в аналитическое задание функций, является основанием для подразделения семейства IRF на классы [3]. Среди логистических функций различают: однопараметрическую модель Г. Раша

$$P_j(\theta) = \frac{e^{1,7(\theta - \beta_j)}}{1 + e^{1,7(\theta - \beta_j)}}$$

$$P_i(\beta) = \frac{e^{1,7(\theta_i - \beta)}}{1 + e^{1,7(\theta_i - \beta)}}.$$

где θ и β – независимые переменные для первой и второй функций соответственно.

двухпараметрическую модель А. Бирнбаума

$$P_j(\theta) = \frac{e^{1,7a_j(\theta - \beta_j)}}{1 + e^{1,7a_j(\theta - \beta_j)}}$$

$$P_i(\beta) = \frac{e^{1,7a_i(\theta_i - \beta)}}{1 + e^{1,7a_i(\theta_i - \beta)}}.$$

Кроме прежних обозначений в этой модели появляются параметры a_i и a_j . Параметр a_j был введен А. Бирнбаумом для характеристики дифференцирующей способности задания при измерении различных значений θ ; параметр a_i указывает на меру структурированности знаний ученика.

трехпараметрическую модель А. Бирнбаума

$$P_j(x_{ij} = 1|\beta_j) = c_j + (1 - c_j) \frac{e^{1.7aj(\theta - \beta_j)}}{1 + e^{1.7aj(\theta - \beta_j)}}$$

где c_j является третьим параметром модели, характеризующим вероятность правильного ответа на задание j в том случае, если этот ответ угадан, а не основан на знаниях ученика.

В каждой из представленных моделей параметры θ и β выражаются как показатели единой для всех моделей шкалы логитов. При отношении двух величин, равном e , их различие составит 1 логит. Таким образом, получается шкала, в которой можно говорить, что знания двух испытуемых или трудности двух упражнений различаются *на столько-то логит* (а не во столько-то раз).

Начальные значения параметров θ и β находят по формулам:

$$\theta_i^0 = \ln \frac{p_i}{q_i},$$

$$\beta_j^0 = \ln \frac{q_j}{p_j},$$

где p_i и q_i – доли правильных и неправильных соответственно ответов i -го ученика на задания теста; p_j и q_j – доли правильных и неправильных ответов учеников на j -е задание теста.

Затем начальные значения параметров переводятся в единую интервальную шкалу стандартных оценок. Стандартизация достигается с помощью ряда специальных преобразований, в ходе которых вычисляются среднее значение для множества ($i = 1, 2, \dots, N$), среднее значение для множества ($j = 1, 2, \dots, n$), дисперсии по этим множествам и поправочные коэффициенты X и Y . Оценки параметров θ и β в единой интервальной шкале находятся по формулам:

$$\theta_i = \beta + X\theta_i^0,$$

$$\beta_j = \theta + X\beta_j^0.$$

Роль последних двух формул в развитии современной теории тестов трудно переоценить, хотя на первый взгляд они имеют узкую практическую направленность. Эти формулы позволяют преодолеть ряд существенных недостатков классической теории тестов, поскольку с их помощью можно получить объективные оценки параметров испытуемых и заданий, не зависящие друг от друга и выраженные в единой интервальной шкале[4].

СПИСОК ЛИТЕРАТУРЫ

1. Педагогический энциклопедический словарь / Под ред. Б.М. Бим-Бад. – М.: Большая рос. энцикл., 2002. – 527 с.
2. Зайцева Л.В., Прокофьева Н.О. Модели и методы адаптивного контроля знаний. // Educational Technology & Society 7(4) 2004.
3. Чельщикова М.Б. Теория и практика конструирования педагогических тестов: Учебное пособие. – М.: Логос, 2002. – 432 с.: ил.
4. <http://luna.cas.usf.edu/~mbrannic/files/pmet/irt.htm>