прийнятним ризиком? Питання про рівень прийнятного ризику є найбільш важливим у прийнятті рішень

ЛІТЕРАТУРА

1. Егоров А. Ф., Савицкая Т. В. Управление безопасностью химических производств на основе новых информационных технологий. М : Химия. 2004. — 416 с.
2. Качинський А. Б. Безпека, загрози і ризик: наукові концепції та математичні методи. К : 2004. — 469 с.
3. Методики оценки последствий химических аварий на опасных производственных объектах. Методика «Токси-2.2.» Сборник документов. Изд. 2-е, испр. и доп. М.: НТЦ по безопасности в промышленности и Госгортехнадзора России. 2002. — 206 с.

пост.11.03.15

Побудова профілів IRT за допомогою сплайнів, які зберігають середнє значення

О. О. ШУМЕЙКО, В. М. КНУРЕНКО

Дніпродзержинський державний технічний університет

Побудова профілів IRT за допомогою сплайнів, які зберігають середнє значення. Стаття присвячена побудові агрегованих характеристик якості тестових завдань за допомогою сплайнів, які інтерполюють у середньому. Встановлено, що використання сплайнів з вільним вузлом дозволяє побудувати інтегральну характеристику якості складання тестових завдань.

Построение профилей IRT с помощью сплайнов, сохраняющих среднее значение. Статья посвящена построению агрегатных характеристик качества тестовых заданий с использованием сплайнов, интерполирующих в среднем. Установлено, что использование сплайнов со свободным узлом позволяет построить интегральную характеристику качества составления тестовых заданий.

IRT profiles scheme using average interpolating splines. The article deals with the construction characteristics of the aggregate quality of tests using average interpolating splines. It was found that the use of splines with free node allows to build an integral characteristic quality of compilation of tests task.

Контроль знань та вмінь студентів є одним з головних елементів навчального процесу. Від правильної організації проведення контролю залежить ефективність управління навчальною роботою та якість підготовки фахівців. Завдяки проведенню контролю, між викладачем та студентом встановлюється «зворотній зв’язок», який дозволяє проводити оцінку динаміки засвоювання навчального матеріалу, дійсний рівень знань, вмінь та навиків, та на їх основі робить відповідні зміни в організацію навчального процесу. Важливою складовою методів контролю знань є тестування. Система тестування - це універсальний інструмент для виявлення рівня знань студентів на всіх етапах навчального процесу. В сучасних умовах, знання методики тестування та створення баз тестових завдань є необхідною складовою роботи викладача.

Використання тестів у якості інструменту вимірювання знань, передбачає наявність певних характеристик якості, які випливають з теорії тестового контролю [3,6]. Теоретичним підґрунтям тестового контролю є класична теорія тестів і сучасна теорія Item Response Theory (IRT). Ці теорії почали формуватись у дослідженнях кінця XIX – початку XX сторіч у наукових працях Ф. Гальтона (F. Galton), Д. Кеттелла (J. Cattell), А. Біне (A. Binet), Т. Симона (T. Simon), Е. Торндайка (E. Thorndike), Ч. Спірмена (C. Spearman), Г. Гулліксена (H. Gulliksen), Л. Гуттмана (L. Guttman), Л. Крокера (L. Crocker), Д. Алгіна (J. Algina), Г. Раша (G. Rasch), А. Бірнбаума (A. Birnbaum) та інших. Постійне зростання кількості публікацій з пошуку й удосконалення моделей IRT свідчить про актуальність вибору цих моделей та їх широке застосування. Класичною моделлю для профілів запитань (ймовірність респондента з рівнем знань θ_i правильної відповіді на запитання зі складністю не вище за β_j) вважають двопараметричну модель Бірнбаума:

$$P(\theta_i,\beta_j)=\frac{e^{D\cdot a_j(\theta_i-\beta_j)}}{1+e^{D\cdot a_j(\theta_i-\beta_j)}},$$

де D=1.7 константа, а - коефіцієнт дискримінації (item discrimination parameter), що визначає нахил характеристичної кривої. Недоліком моделі в практичному використанні є її нелінійна залежність від параметрів, та обмежена "гнучкість". IRT базується на математичних моделях, які відрізняються виглядом функції $P(\theta_i, \beta_j)$. На основі цих моделей будуються профілі складності питань та рівня підготовленості студентів – характеристичні криві. Характеристичні криві завдань тесту є основним джерелом інформації в IRT, оскільки всі інші показники тесту отримуються з них. Характеристичним кривим завдань властиві:

1. Складність завдання, яка визначається за шкалою підготовленості учня на рівні вірогідності правильної відповіді $P(\theta) = 0.5$. Отже складність завдання є медіаною розподілу ймовірності правильної відповіді.

2. Диференційна здатність, яка показує наскільки добре завдання може розрізняти учнів з різним рівнем знань. Диференційна здатність оцінюється за значеннями нижньої та верхньої меж. Межі визначаються за профілем: нижня на рівні $P(\theta) = 0.25$ та верхня на рівні $P(\theta) = 0.75$. Ця властивість є рівнем нахилу характеристичної кривої завдання в середній частині. Відтак, чим більше рівень нахилу, тим краще завдання тесту зможе розрізнити рівні знань учнів

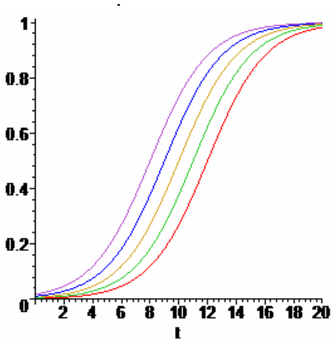

Рис.1. Характеристичні криві завдання з однаковою диференційною здатністю, але з різним рівнем складності

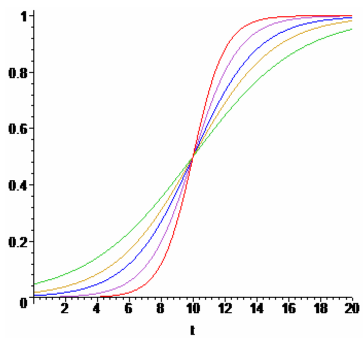

Рис.2. Характеристичні криві завдань з однаковим рівнем складності, але з різною диференційною здатністю

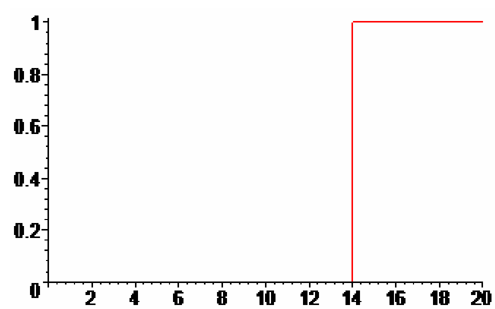

Рис.3. Характеристика завдання з ідеальною диференційною здатністю

Кількість математичних моделей в IRT постійно збільшується, їхній огляд з'являється в наукових періодичних виданнях. Причиною цього є, насамперед, значний інтерес до питань оцінки якості та надійності тестів в освіті, а також необхідність в отриманні найбільш точної, надійної та простої для використання моделі. Обґрунтовуючи недоліки параметричних моделей Дж. Рамсей, М. Абрахамовіч, С. Вінсберг, Д. Тіссен і Г. Вайнер (J. O. Ramsay, M. Abrahamowicz, S. Winsberg, D. Thissen, H. Wainer) запропонували методи оцінки характеристичних кривих, які базуються на застосуванні сплайн-моделей [**Ошибка! Источник ссылки не найден.**,2]. Треба зазначити, що використання інтерполяційних сплайнів не завжди коректно відображає реальні характеристики, тому доцільно було б розглянути сплайн-моделі на основі інших апроксимаційних моделей, так у роботах І.В.Шелевицького було запропоновано використовувати сплайн-регресійні моделі [4,5]. У даній роботі, у якості IRT моделі запропоновано використовувати сплайни, які зберігають середнє значення функцій, тобто інтерполюють у середньому.

Розглянемо допоміжні результати, необхідні у подальшому.

Через $S_r(\Delta_n)$ позначимо множину всіх поліноміальних сплайнів порядку r мінімального дефекту за розбиттям Δ_n, тобто множину всіх функцій з неперервною $r-1$-ю похідною, які співпадають на кожному з інтервалів (t_{i-1}, t_i) з алгебраїчним поліномом ступеню не вище r. Якщо для неперервної функції $x(t)$ сплайн $s(\Delta_n, t) \in S_r(\Delta_n)$ такий, що

$$s(\Delta_n, t_i) = x(t_i),$$

то кажуть - сплайн інтерполює функцію $x(t)$ у вузлах розбиття Δ_n, якщо виконується рівність

$$\frac{1}{h_i}\int_{t_{i-1}}^{t_i} s(\Delta_n, t)dt = \frac{1}{h_i}\int_{t_{i-1}}^{t_i} x(t)dt,$$

такий сплайн інтерполює у середньому або є таким, що зберігає середнє значення функції $x(t)$.

Теорема 1. Нехай $x(t)$ функція, яка інтегрується на R^1 та $X(t)$ первісна $x(t)$ така, що $X(0) = 0$. Окрім того, нехай $s(X, t)$ сплайн, який інтерполює $X(t)$ у вузлах t_i $0 \le t_0 < t_1 < ... < t_{n-1} < t_n$, тобто

$$X(t_i) = s(X, t_i)(i = 0,1,...,n). \quad (1)$$

Тоді сплайн $s'(X, t)$ зберігає середнє значення функції $x(t)$ на проміжку $[t_i, t_{i+1}]$.

Доведення. Розглянемо співвідношення

$$\int_{t_i}^{t_{i+1}} x(t)dt - \int_{t_i}^{t_{i+1}} s'(X,t)dt = \int_{t_i}^{t_{i+1}} \left(x(t) - s'(X,t)\right)dt = = \int_{-\infty}^{\infty} \left(x(t) - s'(X,t)\right)\chi\left(t, t_i, t_{i+1}\right)dt, \quad (2)$$

де

$$\chi(t,a,b) = \begin{cases} 1, & t \in [a,b], \\ 0, & t \notin [a,b], \end{cases}$$

функція Хевісайда. Застосуємо до (1) інтегрування по частинам

$$\int_a^b u dv = uv\Big|_a^b - \int_a^b v du,$$

нехай

$$u(t) = \chi\left(t, t_i, t_{i+1}\right) \text{ і } dv(t) = \left(x(t) - s'(X,t)\right)dt,$$

тоді

$$du(t) = \left(\delta\left(t - t_i\right) - \delta\left(t - t_{i+1}\right)\right)dt,$$

де $\delta(t)$ дельта-функція Дірака та

$$v(t) = \int_0^t \left(x(\tau) - s'(X,\tau)\right)d\tau = \int_0^t x(\tau)d\tau - \int_0^t s'(X,\tau)d\tau = X(t) - s(X,t).$$

Тоді

$$\int_{t_i}^{t_{i+1}} x(t)dt - \int_{t_i}^{t_{i+1}} s'(X,t)dt = \left(X(t) - s(X,t)\right)\chi\left(t, t_i, t_{i+1}\right)\Big|_{t_i}^{t_{i+1}} - - \int_{t_i}^{t_{i+1}} \left(X(t) - s(X,t)\right)\left(\delta\left(t - t_i\right) - \delta\left(t - t_{i+1}\right)\right)dt.$$

Зважаючи на те, що

$$\left(X(t) - s(X,t)\right)\chi\left(t, t_i, t_{i+1}\right)\Big|_{t_i}^{t_{i+1}} = \left(X(t_{i+1}) - s(X, t_{i+1})\right) - - \left(X(t_i) - s(X, t_i)\right)$$

то з умови (1) маємо

$$\left(X(t) - s(X,t)\right)\chi\left(t, t_i, t_{i+1}\right)\Big|_{t_i}^{t_{i+1}} = 0.$$

Окрім того, тому що

$$\int_{t_i}^{t_{i+1}} \left(X(t) - s(X,t)\right)\delta\left(t - t_i\right)dt = X(t_i) - s(X, t_i)$$

та

$$\int_{t_i}^{t_{i+1}} \left(X(t) - s(X,t)\right)\delta\left(t - t_{i+1}\right)dt = X(t_{i+1}) - s(X, t_{i+1}),$$

то з умови інтерполяції (1) маємо

$$\int_{t_i}^{t_{i+1}} \left(X(t) - s(X,t)\right)\left(\delta\left(t - t_i\right) - \delta\left(t - t_{i+1}\right)\right)dt = 0.$$

Таким чином,

$$\int_{t_i}^{t_{i+1}} x(t)dt = \int_{t_i}^{t_{i+1}} s'(X,t)dt,$$

тобто сплайн $s'(X,t)$ зберігає середнє значення функції $x(t)$ на проміжку $[t_i, t_{i+1}]$, іншими словами, є інтерполяційним у середньому.

Побудова сплайн-профілів IRT

Перейдемо до головних результатів даного дослідження. Розглянемо процес формування відповідей респондента на тестові запитання. У даному випадку ми маємо дві апріорно невідомі величини, які характеризують тестове запитання й респондента, а саме - рівень складності запитання і рівень знань респондента. При умові фіксації одного з параметрів θ або β, задача оцінювання $P(\theta,\beta)$ зводиться до визначення залежності від одного з параметрів: $P(\theta)$ – профіль запитання, або $P(\beta)$ – профіль респондента. Будемо вважати, що θ і β в процесі експерименту (тестування) залишаються незмінними, тоді можливо знайти оцінки ймовірностей, які пов'язані з θ і β. Припустимо, що результат відповіді i-го респондента на j-е завдання дорівнює $r_{i,j}$, де $r_{i,j}=1$, якщо відповідь вірна (але можна використовувати і зважену оцінку $r_{i,j}>0$), у зворотному випадку, дорівнює 0. Тоді оцінка рівня знань респондента дорівнює

$$\hat{\theta}_i = \frac{1}{M}\sum_{j=1}^{M} r_{i,j},$$

а оцінка рівня складності завдання тесту дорівнює

$$\hat{\beta}_i = 1 - \frac{1}{N}\sum_{i=1}^{N} r_{i,j},$$

де M — кількість завдань тесту, N — кількість респондентів.

Зважаючи на те, що респондент з більш високим рівнем знань відповідає вірно на завдання з ймовірністю не менше ніж респондент з нижчим рівнем знань, маємо

$$P(\theta_i) \geq P(\theta_k), \text{ якщо } \theta_i > \theta_k.$$

З цього випливає неспадаючий характер залежності $P(\theta_i)$ для фіксованого рівня складності запитання β. Тобто, $P(\theta_i,\beta)$ є характеристичним профілем завдання. Виходячи з цього, $P(\theta_i,\beta)$ є кумулятивною кривою ймовірностей, кожна точка якої відповідає ймовірності того, що респондент із рівнем знань не більшим за θ дає правильну відповідь на запитання з рівнем складності β. Таким чином, оцінки характеристичної кривої запитання визначаються як

$$P_j(\theta) = \frac{1}{N}\sum\left\{r_{i,j} \middle| \theta_i \leq \theta\right\}.$$

Оскільки ми маємо лише множину з $\hat{N}$ оцінок θ, то отримуємо $\hat{N}$ емпіричних точок

$\hat{P}_j(\theta_k) = \frac{1}{\hat{N}}\sum\left\{r_{i,j} \middle| \theta_i \leq \theta_k\right\}, k = 1,\ldots,\hat{N}$. Наведемо алгоритм побудови сплайн-профілю IRT, який зберігає середнє значення.

Нехай є множина $\hat{P}_j(\theta_k), k = 1,\ldots,\hat{N}$. Через

$$I\left(\hat{P}_j(\theta_k)\right)=\frac{1}{\hat{N}}\sum_{\nu=0}^{k}\hat{P}_j(\theta_\nu),\ k=1,\dots,\hat{N}$$

позначимо дискретний аналог первісної від $\hat{P}_j(\theta_k)$.

Через $S\left(I\left(\hat{P}_j(\theta_k)\right),t\right)$ позначимо кубічний сплайн мінімального дефекту з одним вузлом $t_0\in(0,1)$, такий, що визначається умовами

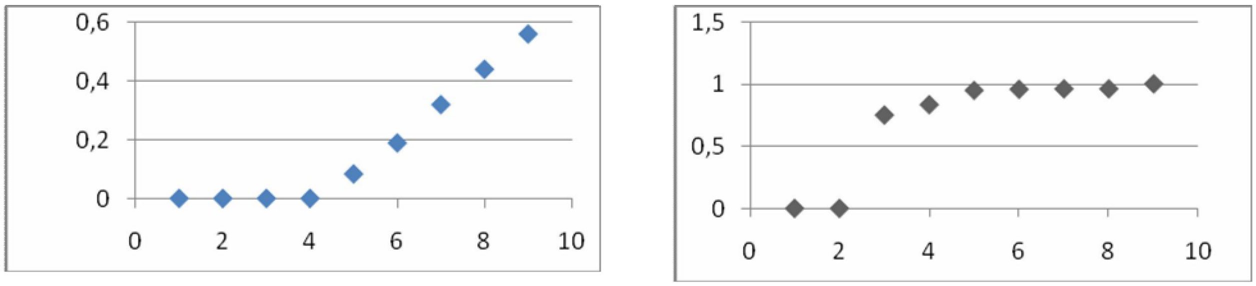

Рис.4. Поведінка $\hat{P}_j(\theta_k)$ для різних тестових завдань I та II.

$$S\left(I\left(\hat{P}_j(\theta_k)\right),0\right)=0,\ S'\left(I\left(\hat{P}_j(\theta_k)\right),0\right)=0,\ S''\left(I\left(\hat{P}_j(\theta_k)\right),0\right)=0,$$
$$S\left(I\left(\hat{P}_j(\theta_k)\right),1\right)=I\left(\hat{P}_j(\theta_{\hat{N}})\right),\ S'\left(I\left(\hat{P}_j(\theta_k)\right),1\right)=\hat{P}_j(\theta_{\hat{N}}),$$
$$S''\left(I\left(\hat{P}_j(\theta_k)\right),1\right)=0.$$

Нескладно отримати цей сплайн в явному вигляді: для $t\in[0,t_0]$

$$S\left(I\left(\hat{P}_j(\theta_k)\right),t\right)=\frac{1}{3t_0}\left(2(t_0-1)\hat{P}_j(\theta_{\hat{N}})+3I\left(\hat{P}_j(\theta_{\hat{N}})\right)\right)t^3,$$

а для $t\in[t_0,1]$

$$S\left(I\left(\hat{P}_j(\theta_k)\right),t\right)=\frac{1}{3(t_0-1)^2}\left((2t_0-3)\hat{P}_j(\theta_{\hat{N}})+3I\left(\hat{P}_j(\theta_{\hat{N}})\right)\right)t^3-$$
$$-\frac{1}{(t_0-1)^2}\left((2t_0-3)\hat{P}_j(\theta_{\hat{N}})+3I\left(\hat{P}_j(\theta_{\hat{N}})\right)\right)t^2+$$
$$+\frac{1}{(t_0-1)^2}\left((t_0^2-2)\hat{P}_j(\theta_{\hat{N}})+3I\left(\hat{P}_j(\theta_{\hat{N}})\right)\right)t+$$
$$-\frac{t_0}{(t_0-1)^2}\left((3t_0-4)\hat{P}_j(\theta_{\hat{N}})+(6-3t_0)I\left(\hat{P}_j(\theta_{\hat{N}})\right)\right)$$

Згідно теореми 1 похідна $S'\left(I\left(\hat{P}_j(\theta_k)\right),t\right)$ є параболічним сплайном, який зберігає середнє значення для будь-якого значення $t_0\in(0,1)$.

Так для $\hat{P}_j(\theta_k), k=1,\dots,\hat{N}$ тестового завдання I (див. *рис.4-I*) маємо

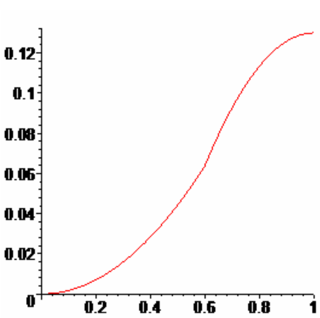

Рис.5. Сплайн для t_0=0.6

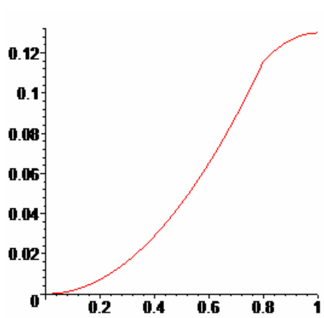

Рис.6. Сплайн для t_0=0.8

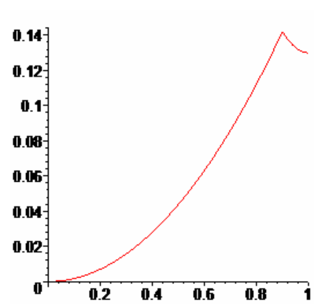

Рис.7. Сплайн для t_0=0.9

а для завдання II

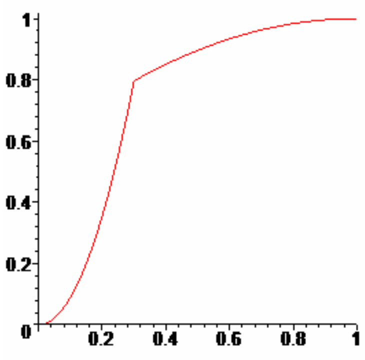

Рис.8. Сплайн для t_0=0.3

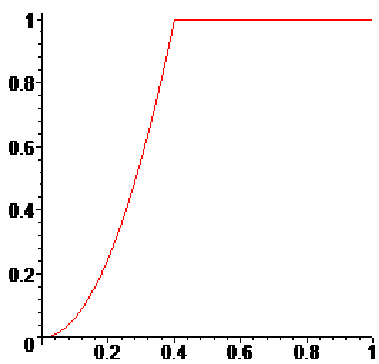

Рис.9. Сплайн для t_0=0.4

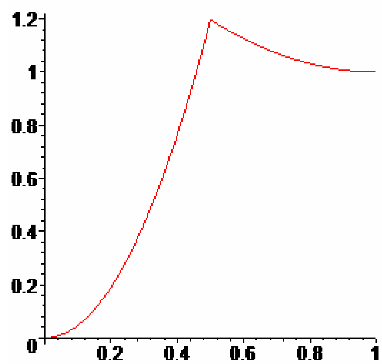

Рис.10. Сплайн для t_0=0.5

Як ми бачимо, на *рис. 7* та *10*, порушується властивість монотонності, тобто отримані сплайн-моделі некоректно описують профіль IRT. Максимальне значення параметру t_0, при якому отримана функція буде монотонною, можна використовувати в якості характеристики завдання тесту. Згідно характеристики ідеального завдання (див. *рис.3*), якщо t_0=0.75, то це завдання можна вважати оптимальним. Таким чином, отримано агреговану характеристику, за допомогою якої можна автоматизувати процес оцінки якості тестових завдань і проводити аналіз складності конкретного завдання тесту (так для приклада, наведеного на *рис.4*, завдання I складено досить якісно, а завдання II є занадто простим).

Висновок

Використання моделі IRT на основі сплайнів, які інтерполюють у середньому, дозволяє отримати агрегатну характеристику оцінки якості складання тестових завдань, що дозволяє реалізувати автоматичну діагностику якості тестів.

ЛІТЕРАТУРА

1. Abrahamowicz M., Multicategorical spline model for item response theory / M. Abrahamowicz, J. O. Ramsay // Psychometrika. — 1992. — vol. 57, no.1 — P. 5—27.
2. Winsberg S. Fitting item characteristic curves with spline functions / S. Winsberg, D. Thissen, H. Wainer // Educational Testing Service. Technical Report. — Princeton: NJ. 1984. — P. 84—52.
3. Алгина Дж. Введение в классическую и современную теорию тестов: учебник / Дж. Алгина, Л. Крокер. — М.: Логос. 2010. — 668 с.
4. Дубан Р. М. Сплайн-моделі в тестових вимірюваннях знань / Р. М. Дубан, І. В. Шелевицький // Рідна школа. — 2010. — №7—8 (967—968). — С.11—14.
5. Дубан Р. М. Сплайн-моделі профілів складності питань та знань респондентів в тестовому контролі знань / Р. М. Дубан, І. В. Шелевицький // Автоматизированные системы управления и приборы автоматики: всеукр. межведомств. научн.-техн. сб. — Х : — 2011. — вып. 156. — С. 71—77.
6. Челышкова М. Б. Теория и практика конструирования педагогических тестов: учеб. пособ. / М. Б. Челышкова. — М. : Логос. 2002. — 432 с.

пост.24.03.15