

встановлення відповідностей, виявляється недостатньо для об'єктивного оцінювання правильності виконання завдань з математичних дисциплін.

У запропонованому підході модифікація редактора формул передбачає створення шаблонів для введення формул та використання алгоритму індексації елементів виразу, що дає можливість уникнути неоднозначності при введенні відповіді студентами. Завдання з комп'ютерною перевіркою аналітичного символічного виразу можуть бути застосовані для автоматизації перевірки вмінь і знань з природничих дисциплін шляхом відповідної модифікації шаблону.

Література

1. Аванесов В. С. *Научные проблемы тестового контроля знаний / В. С. Аванесов.* – М. : Исслед. центр, 1994. – 135 с.
2. Жарких Ю.С, Лисоченко С.В, Сусь Б.Б, Третьяк О.В. *Комп'ютерні технології в освіті. Навчальний посібник.* К.: ВПЦ "Київський університет", 2012. – 239 с. <http://iht.univ.kiev.ua/books-ih/comp-tech-osvita.pdf>
3. *Firemath - The MathML Editor. [Electronic Resource]. – Mode of access : URL : <http://www.firemath.info/> – Title from the screen.*

УДК 681.3

Роман Дубан

Криворізький національний університет

E-mail: romaduban@gmail.com

ПОБУДОВА ПРОФІЛІВ IRT НА ОСНОВІ ІНТЕГРАЛІВ ВІД БАЗИСНИХ СПЛАЙНІВ

© Роман Дубан, 2013

Основна увага приділяється практичним засобам дослідження властивостей завдань тесту та учасників тестування на основі профілів IRT. Для побудови профілів запропонована нова модель більш адекватна IRT профілям ніж відомі, що складається з інтегралів від базисних сплайнів. Запропоновано новий інструментарій для розрахунку та аналізу результатів тестування на основі представленої моделі.

Ключові слова: інформаційна система, профілі IRT, сплайн-модель.

The focus is on practical means to learn the properties of the test objectives and test participants based on profiles IRT. To construct a new model profiles more than adequate IRT profiles are known, which consists of integrals of basis splines. A new tool for calculating and analyzing test results based on the presented model.

Keywords: *information system, IRT profiles, spline model.*

Вступ. Тест, незважаючи на ряд його недоліків, на сьогодні, є найбільш об'єктивним технологічним інструментом вимірювання результатів педагогічного процесу. Він набув поширення в застосуванні завдяки стрімкому розвитку технологій та появі засобів, які дозволяють забезпечити процес тестування. Без використання таких засобів для підготовки, проведення й аналізу результатів тестування, процес є досить складним і дорогим.

Разом із широким застосуванням тестового контролю знань в освіті, більше уваги приділяють і характеристикам тесту, які забезпечують точність отриманих результатів вимірювання. На жаль, програмне забезпечення тестового контролю знань, якщо й містить модуль аналізу результатів, то дуже примітивний, заснований на базових статистичних показниках. Існує значний обсяг спеціалізованого програмного забезпечення, що виконує розрахунки одно-, двох- та трьох-параметричних функцій, а також непараметричних моделей IRT [1].

Програмне забезпечення для аналізу результатів тестування. Такі програми розрізняються за різними критеріями: обсягами вхідної інформації, кількістю математико-статистичних методів, обсягами (кількістю) результатних таблиць тощо. Існують як безплатні, так і пропріетарні програми цього класу. Вони дозволяють опрацювати великі обсяги даних, розрахувати потрібні статистики. У результаті користувач одержує інформацію про формальні властивості кожного тестового завдання й тесту в цілому, про рівень і структуру підготовленості осіб, які проходять тестування, про надійність і валідність результатів тестування, про міру відповідності результатів випробовуваних рівню складності завдань тощо. Програми цього класу здійснюють також шкалювання рівнів підготовки учасників тестування, і рівнів складності тестових завдань.

Наприклад, статистичний пакет QUEST призначений для аналізу як тестових завдань, так і питань соціологічних анкет. Обробка даних у ньому може проводитися згідно з класичною (статистичною) теорією педагогічних вимірів, так і відповідно до IRT. Програма WINMIRA має широкий набір методів обробки даних на основі Item Response Theory. Особливо слід відмітити

повну інтеграцію цього засобу із статистичним пакетом SPSS. Програмний засіб CONQUEST надає змогу проведення не лише одновимірних, але і багатовимірних вимірів. Пакет RASCAL дозволяє шкалювати завдання і учасників тестування на основі однопараметричної моделі Раша, якщо дані подані в дихотомічній шкалі (0 або 1).

Але найбільш популярними і розповсюдженими програмними продуктами вважаються австралійська розробка RUMM 2030 та американський програмний засіб WINSTEPS. Програма WINSTEPS є одним з різновидів програм, призначених для аналізу результатів тестування, побудована на технології за теорією Г.Раша. Програма є комерційною, але існує також безкоштовна версія, яка називається MINISTER, що дозволяє використовувати усі можливості комерційного варіанту, але має обмежений обсяг кількості питань у тесті до 25 і кількості осіб до 75. Елементи таких розробок, також, входять в потужні статистичні системи такі, як MatLab, R та інші [2].

Для застосування згаданих вище програм викладачі повинні мати неабиякі знання із статистики. Метою існуючих програм є, здебільшого, доведення теоретичних положень IRT і вони майже не пристосовані для використання в площині емпіричних даних у зв'язку із складністю інтерфейсів та введеними обмеженнями на кількість вхідних даних. Але ці програми дозволяють досліджувати поведінку моделей IRT, які в них підтримуються.

Особливості застосування IRT. IRT потенційно дозволяє створювати дихотомічні тести із визначеними характеристиками якості. Серед досягнень теорії слід зазначити побудову профілів тестових запитань та респондентів, які відповідають найбільш повним їх характеристикам. Існує велике різноманіття моделей профілів. Найбільш відомі серед них однопараметрична функція Г. Раша, двох- і трьохпараметричні функції А. Бірнбаума. Однак, існує проблема апроксимації цими функціями емпіричних даних, отриманих з результатів тестування контрольних груп та в процесі використання тестів. Функції нелінійно залежать від параметрів. У зв'язку з цим необхідно застосовувати методи нелінійної оптимізації та оцінювання, які не гарантують отримання глобального мінімуму критерію наближення. Тому, процес оцінювання параметрів існуючими методами відбувається індивідуально для кожного тестового запитання й респондента під контролем досвідченого оператора.

Поряд із вказаними функціями IRT розроблено й досліджуються багато інших різноманітних функцій. Саме аналізу цих функцій і процедурам оцінювання їх параметрів присвячена велика кількість публікацій у сфері теорії тестового контролю [3]. Це також ілюструє відсутність функції, яка б достатньо

повно задовольняла існуючі теоретичні й практичні потреби тестового контролю. Особливі труднощі виникають, коли емпіричні дані виявляються далекі від очікуваних теоретичних. У таких випадках деякі автори рекомендують вважати питання невдало сформульованими. Однак, залишається проблема, що робити з “неправильними” респондентами.

Серед великої кількості функцій, що потенційно придатні для моделювання профілів особливу увагу привертають сплайни. Є ряд спроб застосовувати сплайни для опису профілів. Слід зауважити, що ідеальною моделлю профілю з найвищою диференційною здатністю є ступінчата функція, що є найпростішим сплайном. Безперечною перевагою сплайн-функцій перед класичними є їх хороші апроксимуючі властивості у поєднанні з простотою розрахунків. Завдяки цим властивостям вдалося вирішити ряд апроксимаційних задач, які неможливо було вирішити класичними параметричними функціями IRT. Хоча сплайн-моделі IRT завдяки своїм властивостям мають кращі показники, вони не враховують деякі особливості IRT, а саме, функція профілів тестових завдань має неспадаючий характер та її значення знаходяться на проміжку від 0 до 1.

Сплайн з інтегральних базисних сплайнів в якості моделі IRT. В результаті пошуку найбільш адекватної сплайн-моделі для розрахунків профілів IRT з врахуванням апріорних умов, в роботі [4] була запропонована сплайн-модель, що складається з інтегралів базисних сплайнів (рис. 1.).

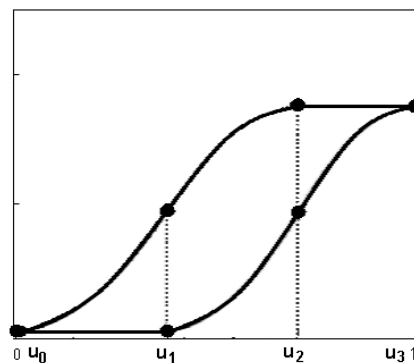


Рис. 1. Базиси сплайну, інтеграли базисних сплайнів.

Формула для сплайна, зображеного на рисунку 1 має вигляд:

$$SI(x) = A_0 Iq_0(x) + A_1 Iq_1(x), \quad (1)$$

де $SI(x)$ – сплайн-модель з базисом інтеграл від лінійного B-сплайна, A – числові коефіцієнти (параметри) сплайна, а базис $Iq(x)$ розраховується за формулою:

$$Iq(x) = \begin{cases} 0, & \text{якщо } x \in (-\infty, u_0) \\ -\frac{(x-u_0)^2}{2(u_0-u_1)}, & \text{якщо } x \in [u_0, u_1) \\ \frac{u_2-u_0}{2} + \frac{(x-u_2)^2}{2(u_1-u_2)}, & \text{якщо } x \in [u_1, u_2) \\ \frac{u_2-u_0}{2}, & \text{якщо } x \in (u_2, \infty) \end{cases} \quad (2)$$

де u_i – вузлові значення сплайна.

Для оцінювання параметрів сплайн-моделі обрано зважений метод найменших квадратів. Вибір обґрунтовано лінійною залежністю моделі від параметрів та наявністю похибки в емпіричних даних з різною дисперсією.

Окрему складність моделювання профілів становить обмеження області значень рамками $0 \div 1$ та неспадаючий (або незростаючий) характер профілів. Для відповідності цим умовам було застосовано комплексний критерій оптимізації, що складається з

Для врахування цих обмежень введемо трикомпонентний критерій оптимізації, який є сумою трьох складових помножених на вагові коефіцієнти: середньоквадратична похибка наближення до емпіричних даних; штрафне значення за вихід значень побудованого сплайна за межі інтервалу $[0; 1]$; штрафна функція за спадаючий характер сплайна. Наближення мінімуму значення критерію оптимізації здійснюється методом покоординатного спуску-підйому, із застосуванням алгоритму золотого прерізу.

Застосування запропонованої моделі для побудови та оцінки параметрів профілів IRT. Описана вище модель знайшла своє практичне втілення в інформаційній системі тестового контролю знань Logit, що доступна за адресою [<http://logit.kdpu.edu.ua/>]. Система має на меті забезпечення підтримки тесту на всіх етапах життєвого циклу, як інформаційного продукту [5]. В системі передбачена послідовна робота з тестом на кожному етапі, від вибору тематики тесту та призначення, формування розділів та наповнення їх завданнями, рецензування завдань, проведення випробувальних тестувань, та визначення характеристик тесту. Але, вона не дозволяє отримати розрахунки профілів IRT за результатами тестування отриманими з інших систем.

Аби не обмежувати користувачів інших систем тестування та надати зручний сервіс для розрахунку профілів завдань тесту та респондентів, на базі Криворізького педагогічного інституту ДВНЗ "Криворізький національний університет" планується впровадження веб-сервісу "IRT Calculator". Метою сервісу є розрахунок та оцінка профілів IRT за наданими результатами тестування в автоматичному режимі. Результати тестування повинні бути у

форма CSV з роздільним знаком – кома. Перший рядок містить список ідентифікаторів респондентів, окрім першого елемента. Перший елемент першого рядка має бути ключовим символом "#". Рядки з другого по останній містять ідентифікатор завдання тесту та відповіді на це завдання кожним з респондентів. Відповідь має бути позначена "1" – якщо відповідь респондента була правильна та "0" – якщо хибна. Також необхідно вказати назву та опис представлених результатів тестування і електронну пошту, на яку буде відправлено повідомлення по завершенню розрахунків. У зв'язку з тим, що в процесі побудови профілів відбувається оптимізація сплайн-моделі за комплексним критерієм спрогнозувати час розрахунків неможливо, але черга розрахунків просуватиметься щогодини. Веб-сервіс "IRT Calculator" незабаром буде розміщено за адресою [<http://irtcalc.kdpu.edu.ua>].

Висновки. Важливою умовою функціонування методів й алгоритмів побудови профілів має бути їх універсальність, у сенсі можливості отримання коректних результатів для усього різноманіття емпіричних даних без участі людини-оператора (в автоматичному режимі). Оцінки профілів завдань тесту дозволяють будувати адаптовані за рівнем складності тести, що скорочує час тестування без суттєвого погіршення точності оцінок рівня знань учасників тестування.

Література

1. Лісова Т. В. Програмні засоби для аналізу результатів тестування. / Т. В. Лісова // Наукові записки. Серія: Психолого-педагогічні науки. – 2011. – кн. 7. – С. 46-52.
2. Фетісов В. С. Програмні засоби для обґрунтування якості тестових завдань. // В. С. Фетісов, Е.О. Чернишова // Наукові записки. Серія: Психолого-педагогічні науки. – 2011. – кн. 10. – С. 106-111.
3. Wang, W. *Are All Item Response Functions Monotonically Increasing: Dissertation Ph.D.: 31.05.2012* / Wenhao Wang. – Kansas, 2012. – 166 p.
4. Дубан Р.М. Застосування інтегралу від лінійного В-сплайна в якості моделі IRT / Р. М. Дубан, І. Ф. Бойко // Електроніка та системи управління. – К. – 2012. – №1(31). – С. 131-138.
5. Дубан Р.М. Інформаційна система "LOGIT" – інструмент підтримки тестового контролю. / Р. М. Дубан, І. В. Шелевицький // Вісник Національного університету "Львівська політехніка": серія Інформатизація вищого навчального закладу. – Львів. – 2012. – № 731. – С. 23-28.