

УДК 517.9

Івохін Є.В.<sup>1</sup>, д.ф.-м.н, доцент,  
Махно М.Ф.<sup>2</sup>, аспірант

### Розробка засобів адаптивного тестування та автоматичного оцінювання знань

<sup>1,2</sup> Київський національний університет  
імені Тараса Шевченка, м.Київ, пр-т Глушкова,  
4д, e-mail: ivohin@univ.kiev.ua

Ivohin E.V.<sup>1</sup>, Dr.Sci.,  
Makhno M.F.<sup>2</sup>, postgraduate

### Adaptive testing and automatic grading methods development

<sup>1,2</sup> Taras Shevchenko National University of Kyiv,  
Kyiv, Glushkova st., 4d,  
e-mail: ivohin@univ.kiev.ua

*У даній статті розглянуто основні підходи, що використовуються для проведення адаптивного тестування. Розглянуто модель параметричної оцінки рівнів знань осіб, що проходять тестування, та рівнів складності тестових завдань за допомогою методики теорії тестових завдань Item Response Theory. Запропоновано реалізацію процесу автоматизованого проведення адаптивного тестування та інтерпретації його результатів.*

*Ключові слова: адаптивне тестування, теорія тестових завдань, параметрична оцінка.*

*This article analyzes the main approaches used for the process of adaptive testing. There are considered the concept of adaptive test control and methods effective implementation of adaptive testing. The model parametric estimation of levels of knowledge of persons tested and difficulty levels of test problems using methods of the theory test items Item Response Theory are considered. It was chosen model of one-parameter logistic Rush function, called the characteristic curve. It is illustrated calculation of latent characteristics. The formulas for calculating the values of the characteristics are discussed. The problem of adequate assessment of the complexity of test items and initial level of person tested knowledge is considered. The formula for estimating the initial level of tested persons is considered. The estimation of task complexity based on the fuzzy approach is proposed. The adaptive estimation scheme is given. The approach to the implementation of the automated adaptive testing process and interpretation of the results is proposed. The tasks of adaptive testing practical using are defined.*

*Key words: adaptive testing, item response theory, parametric estimation.*

Статтю представив д.т.н., проф. Волошин О.Ф.

Комп'ютерні системи контролю професійної підготовки є одним із різновидів автоматизованих систем навчання та контролю знань, які, у свою чергу, відносяться до спеціалізованих експертних систем. Визначальними характеристиками систем перевірки профпідготовки є багатокритеріальність процесу оцінювання знань і навичок осіб, що навчаються, та одержання висновків про рівень їх професійної підготовки і професійної спрямованості.

Комп'ютерні системи реалізують певні інформаційні технології, які за визначенням є сукупністю методів та засобів збору, збереження та перетворення вихідної інформації (апріорні дані) у кінцевий продукт (апостеріорні знання). Технологічність використання комп'ютерних систем для навчання та контролю знань визначає високий професійний рівень їх реалізації та забезпечує мінімальні часові витрати осіб, що проходять тестування (ОПТ), і максимально об'єктивний процес оцінювання. Одним з підходів для створення систем контролю профпідготовки є реалізація схем адаптивного тестування

та ефективної інтерпретації його результатів.

#### Поняття адаптивного тестування

Під адаптивним тестовим контролем [1] розуміють комп'ютеризовану систему науково обгрунтованої перевірки та оцінки результатів навчання, що відзначається високою ефективністю за рахунок оптимізації процедур генерації, пред'явлення та оцінювання результатів виконання адаптивних тестів. Ефективність контрольно-оціночних процедур підвищується при використанні багатокрокової стратегії відбору та пред'явлення завдань, що базується на алгоритмах з повною контекстною залежністю, в яких черговий крок здійснюється лише після оцінки результатів виконання попереднього кроку. Після виконання чергового завдання щоразу виникає потреба у прийнятті рішення про підбір складності наступного завдання в залежності від того, вірною або невірною була попередня відповідь. Алгоритм формулювання завдань будується за принципом зворотного зв'язку, коли при правильній відповіді ОПТ чергове завдання вибирається більш важким, а невірна відповідь

тягне за собою пред'явлення більш легкого завдання, ніж те, на яке було дано невірну відповідь. Існує можливість задання додаткових питань з інших тем, які тестований знає не дуже добре. Це дозволяє більш точно визначити рівень знань за обраними тестами. Таким чином, можна сказати, що адаптивна модель нагадує викладача на іспиті - якщо відповіді на поставлені питання здійснюються впевнено і правильно, викладач досить швидко ставить ОПТ позитивну оцінку. Якщо тестований починає «плавати», то викладач задає йому додаткові або спрямовуючі питання того ж рівня складності або з тієї ж теми. І, нарешті, якщо від ОПТ із самого початку викладач не отримує або отримує неправильні відповіді на поставлені питання, викладач теж досить швидко виставляє оцінку, але негативну.

Методом ефективною реалізації систем адаптивного тестування повинні стати системи навчання, які з розвитком мультимедійних технологій вивели на новий рівень викладання навчальних дисциплін. Інтеграція систем навчання і контролю дозволяє реалізувати замкнений цикл навчання. Жорсткість такої структури стала причиною впровадження в автоматизовані системи (АС) адаптаційних механізмів, що базуються на різних принципах і моделях.

Одним з прикладів таких моделей є методика проведення оцінювання знань на основі тестових завдань (IRT). IRT - це проведення високоякісних педагогічних вимірювань рівня підготовленості тестованих і рівня складності завдань, куди входить пошук відповідних прогностичних моделей та проведення розрахунків придатності моделі для наявних даних.

Завдання IRT зводяться до розробки методів вимірювання, які дозволяють отримати найкращі (оптимальні) параметричні оцінки рівня підготовленості ОПТ і оцінки складності завдань на основі вибірових статистик та інших емпіричних даних [2].

У технологіях, що базуються на використанні IRT, вводиться основне припущення про існування деякого взаємозв'язку між результатами тестування, що спостерігаються, і латентними якостями осіб, що проходять тестування. Передбачається, що кожній особі ставиться у відповідність тільки одне значення латентного параметра. Елементи першої множини  $\theta_i$  - це рівні знань групи з  $N$  осіб, де  $i = \overline{1, N}$ . Другу множину утворюють значення латентного параметра  $\delta_j$ ,  $j = \overline{1, K}$ , що дорівнюють складності  $j$  завдання тесту. На практиці завдання полягає в оцінці значення латентних

параметрів за відповідями тестованих осіб на завдання тесту.

Функціональна залежність ймовірності правильної відповіді на завдання від латентної характеристики визначається у вигляді характеристичної кривої завдання (Item characteristic curve - ICC). Найчастіше в теорії IRT передбачається, що ICC мають S-подібну форму (рис. 1). З графіка випливає, що із зростанням оцінки латентної характеристики збільшується ймовірність правильної відповіді на завдання тесту. На відміну від класичної теорії з статистиками, що показують складність і відмінність завдань [3], у цьому випадку можна перевіряти, як ймовірність правильної відповіді залежить від латентної характеристики.



Рис.1. Приклад ICC.

Важливим моментом в моделях латентних характеристик є те, що вони дозволяють провести порівняння тестованих навіть у тих випадках, коли вони виконують різні завдання тесту. Цей ефект іноді називають вимірюванням, вільним (незалежним) від тесту. Для ілюстрації припустимо, що особа, яка проводить тестування, знає, що чотири завдання поділяються за складністю на два рівні. Екзаменатор пред'являє два найбільш легких завдання одній особі, а два найважчих - другій. Перший тестований відповідає на завдання №1 правильно, а на завдання №2 - неправильно. Таким чином, робиться висновок, що оцінка його латентної здатності знаходиться в інтервалі  $\theta_1 \leq \theta < \theta_2$ . Другий тестований відповідає на завдання №3 правильно, а на завдання №4 - неправильно, тому екзаменатор робить висновок про те, що оцінка латентної здатності випробовуваного  $\theta_3 \leq \theta < \theta_4$ . На підставі цього екзаменатор може зробити висновок, що перший відповідаючий має меншу латентну здатність, ніж другий.

Цей приклад показує, що завдяки можливостям теорії латентних характеристик результати ОПТ можуть бути розміщені на одній шкалі навіть у випадках, коли виконуються тести з різних завдань, за умови, що вимірюється одна

латентна характеристика.

Найбільш поширеною математичною моделлю з використання IRT є модель з одно-параметричною логістичною функцією Раша, що називається *характеристичною кривою* і має вигляд [3,4]:

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

де  $P_j$  - імовірність правильної відповіді особи будь-якого рівня підготовки на завдання певного рівня складності з номером  $j, j = \overline{1, K}$ ;  $\delta_j$  - рівень складності конкретного  $j$ -го завдання проєктованого тесту;  $\theta$  - рівень знань (латентна змінна).

Чим вище рівень зростання функції  $P_j$  для довільного завдання  $j$ , тим вужче інтервал рівня складності, для якого ймовірна правильна відповідь. Пропонується поліпшити модель Раша за рахунок введення у вираз додаткового параметра  $a_j, j = \overline{1, K}$ , який дає інформацію про завдання з точки зору оцінки його диференціуючої здібності на заданому інтервалі. Геометрично значення параметра виражається рівнем зростання характеристичної кривої, аналітично - значенням похідної функції в точці перегину. Після введення в вираз параметра  $a_j$  отримаємо двопараметричну модель:

$$P_j = \frac{e^{a_j(\theta - \delta_j)}}{1 + e^{a_j(\theta - \delta_j)}}, \quad (2)$$

у якій емпіричні межі значень для параметра  $a_j, j = \overline{1, K}$ , задаються в межах від -2,80 до +2,80.

Наступною є проблема адекватної оцінки складності тестових завдань і початкового рівня знань тестованих. Початкова оцінка рівня знань ОПТ визначається за формулою  $\theta_i = \ln(p_i / q_i)$ , де  $p_i$  - частка правильних відповідей  $i$ -ї особи,  $q_i$  - частка неправильних відповідей,  $i = \overline{1, N}$ . Аналогічно визначається початкова оцінка рівня складності завдання тесту  $\delta_j = \ln(q_j / p_j)$ , де  $p_j, q_j$  - частки правильних та неправильних відповідей на  $j$ -е завдання тесту,  $j = \overline{1, K}$ .

З огляду на дії різних випадкових факторів оцінки параметрів  $\theta_i$  і  $\delta_j$  для будь-яких  $i = \overline{1, N}$  та  $j = \overline{1, K}$ , отримані на різних вибірках, будуть, звичайно, різними. Якщо обсяг вибірки досить великий, то можна ставити питання про обчи-

слення значень параметрів  $\theta_i$  і  $\delta_j$ , що будуть найбільш ефективними та об'єктивними оцінками параметрів. При довільному тестуванні результати обчислення  $\overline{\theta}_i (i = \overline{1, N})$  - статистичних оцінок  $\theta_i$  і  $\overline{\delta}_j (j = \overline{1, K})$  - статистичних оцінок  $\delta_j$  будуть відрізнятися від існуючих точних значень.

### Оцінка та інтерпретація результатів

При звичайному оцінюванні результатів тестів кожна тестова оцінка є сумою оцінок за завданнями для кожного відповідаючого, якому присуджується 1 бал за правильну відповідь і 0 за будь-який інший варіант відповіді. Це правило іноді називають оцінкою числа правильних відповідей. Такий результат оцінювання можна виразити формулою:

$$X_a = \sum_{i=1}^n x_{ai}, \quad (3)$$

де  $X_a$  - загальна (повна) тестова оцінка ОПТ  $a$ ;  $x_{ai}$  - оцінка виконання ОПТ  $a$  завдання  $i$ . Якщо завдання оцінюються дихотомічно, значення  $x_{ai}$  дорівнюють 0 або 1. Згідно з цим оціночним правилом всі завдання мають однакову вагу.

Концепція оцінюючої формули може бути врахована при отриманні наближених оцінок істинних балів для розроблених тестів. Необхідно згадати, що  $P_j(\theta)$  можна інтерпретувати як імовірність того, що ОПТ з рівнем здатності  $\theta$  виконає завдання  $j$  правильно. Тому справжній бал тестованого може бути оцінений шляхом врахування ймовірностей усіх правильних відповідей на всі завданням. Лордом Ф.М.[3] запропоновано модифікацію, за якою визначають справжню оцінку числа правильних відповідей ОПТ  $a$  за допомогою процедури:

- 1) встановити всі завдання, на які відповів  $a$ ;
- 2) для кожного з цих завдань отримати  $P_j(\theta)$  - ймовірність, з якою  $a$  з оцінкою здатності  $\theta$  відповість на це завдання вірно;
- 3) підсумувати ці ймовірності.

Це можна представити у вигляді формули

$$\xi_a = \sum^{(a)} P_j(\theta), \quad (4)$$

де  $\xi_a$  - істинний бал, отриманий за кількістю правильних відповідей  $a$ ;  $\sum^{(a)}$  - сумування результатів тільки за тими завданнями, на які відповідав  $a$ . Наближена оцінка істинного бала, отриманого за кількістю правильних відповідей  $a$ , коригується для врахування поправки на випадок вгадування за формулою:

$$\eta_a = \sum^{(a)} P_j(\theta) - \frac{\sum^{(a)} Q_j(\theta)}{k-1}, \quad (5)$$

де  $Q_j(\theta) = 1 - P_j(\theta)$ ;  $k$  - кількість варіантів відповіді на кожне завдання.

Припустимо, що заздалегідь відома оцінка параметра  $\delta_j$ ,  $j = \overline{1, K}$ . Тоді для реалізації адаптивного тестування може бути запропонована процедура опитування. Якщо відповідь на питання є правильною, то вважається, що рівень підготовки тестованого вище складності завдання, і він здатний вирішувати задачі заданої складності, інакше - нездатний. Надалі, якщо відповідаючий успішно впорався з наданим завданням, то у нього з'являється бажання отримати більш важке завдання. Якщо ні - то йому буде запропонована ще одна спроба розв'язування завдання тієї ж складності. У випадку, якщо воно також не розв'язане, то пропонується більш просте завдання. У випадку невиконання цього завдання, пропонується завдання ще меншою складності.

Аналогічно відбувається процес підвищення складності завдань. У результаті особа, що проходить тестування, вибере для себе певний рівень складності, відповідно до якого і буде формуватися складність завдань. Таким чином, функція «рівня знань» є перетворенням функції «складності» завдання через «здатність розв'язання задач» певної «складності». Ці терміни мають нечіткий характер. Тому для формалізації цих понять доцільно застосовувати теорію нечітких множин [5].

#### Список використаних джерел

1. Hambleton R.K., Jones R.W. Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development. University of Massachusetts at Amherst, 1993.
2. Нейман Ю.М. Введение в теорию моделирования и параметризации педагогических тестов / Ю.М.Нейман, В.А.Хлебников. - М.: Прометей, 2000. - 168 с.
3. Lord, F.M., Novick, M.R. Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley, 1968.
4. Croker L., Algina J. Introduction to Classical and Modern Test Theory. - SF: Cengage Learning, 2006. - 527 p.
5. Орловский С.А. Проблемы принятия решения при нечеткой исходной информации. - М.: Наука, 1981. - 206с.

Існують три варіанти схем адаптивного контролю, що відповідають сучасному рівню автоматизації процесів освіти:

- пірамідальне тестування (завдання однакового середнього рівня складності, що в залежності від відповідей доповнюються або більш важкими, або більш простими);
- flexilevel-тестування (контроль починається з будь-якого рівня складності завдань);
- stradaptive тестування (кожне наступне завдання відрізняється від попереднього за складністю на один рівень).

Залучення сучасних методик адаптивного тестування спрямоване на наближення процесу навчання до індивідуальних особливостей в умовах колективного навчання. Контроль за результатами навчання дозволяє оцінити стан навчання і необхідні дії для відповідної корекції навчального процесу. В цілому, адаптивне тестування є адекватним сучасним напрямком розвитку дистанційної освіти і відкриває нові можливості у підвищенні ефективності навчальних процесів.

**Висновок.** У даній статті розглянуто основні поняття адаптивного тестування. Для проведення адаптивних тестів пропонується застосування теорії тестових завдань IRT, яка дозволяє визначити рівні знань осіб, що проходять тестування за допомогою оцінки латентних характеристик  $\theta_i$  і  $\delta_j$ . Запропоновано підхід до проведення адаптивного тестування та інтерпретації його результатів.

#### References

1. HAMBLETON R. and JONES R. (1993) *Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development*, University of Massachusetts at Amherst.
2. NEUMANN J. and KHLEBNIKOV V. (2000). *Vvedenie v teoriju modelirovanija i parametrizacii pedagogicheskikh testov*. Moskva:Promethey.
3. LORD, F., and NOVICK, M. (1968) *Statistical theories of mental test scores*. Reading, Mass.: Addison-Wesley.
4. CROKER L. and ALGINA J. (2006) *Introduction to Classical and Modern Test Theory*. SF: Cengage Learning.
5. ORLOVSKY S. (1981) *Problemy prinjatija reshenija pri nechetkoj ishodnoj informacii*. Moskva: Nauka.

Надійшла до редколегії 22.12.2014