

УДК 004.021

DOI: 10.33310/2524-0978-2019-1-7-33-40

Катерина КУЗЬМА

katushke2017@gmail.com

ORCID: 0000-0002-0937-7299

м. Миколаїв

АНАЛІЗ МЕТОДІВ ТА МОДЕЛЕЙ ОЦІНКИ ЗНАНЬ В СИСТЕМАХ ТЕСТУВАННЯ

В роботі розглянуто моделі та методи оцінки рівня знань в системах тестування. Виконано їх систематизацію на дві категорії: математичні та класифікаційні, наведено обчислювальні алгоритми, виявлено переваги та недоліки.

Визначено, що основним недоліком «чітких» тестів (використовується двійкова система оцінки правильності відповіді на кожне питання (правильно або неправильно)) є оцінка неповних або неточних відповідей як неправильних, тобто наближене, неповне знання відповіді на питання кваліфікується як незнання відповіді, а це не завжди виправдано. Застосування ідеї «нечітких» тестів дозволяє підвищити ефективність процесу ідентифікації знань. Аргументовано практичну значимість використання статистичних методів та ЕОМ при розробці та апробації тестів.

Ключові слова: методи оцінки знань, моделі оцінки знань, системи тестування, класифікація.

Постановка проблеми

Вирішення задач оцінки і контролю знань пов'язано з визначенням рівня та якості навчальних досягнень здобувачів вищої освіти з метою одержання об'єктивної інформації про рівень засвоєння навчального матеріалу, своєчасного виявлення недоліків та прогалин у знаннях.

Вирішення питання вимірювання результатів контролю пов'язано з поняттям «результативності», що розглядається як оцінка рівня знань студентів. Визначення кількості набраних балів може проводитися:

- за визначеним експериментальним методом шляхом знаходження відсотку виконання завдань тесту;
- за визначеним відсотком з урахуванням складності тесту;
- за визначеним відсотком з урахуванням диференціації завдань тесту за рівнем складності.

Аналіз останніх досліджень і публікацій

У наукових працях вітчизняних та зарубіжних авторів для об'єктивної оцінки використовуються наступні моделі [1-3]:

1. Математичні моделі:

- 1.1. Моделі, які враховують параметри завдань.

- 1.2. Моделі на основі імовірнісних критеріїв.

- 1.2.1. Модель Раша.

- 1.2.2. Метод уточнюючих питань

- 1.3. Моделі на основі нечітких множин.

2. Класифікаційні моделі:

- 2.1. Моделі на основі АОО.

Постановка завдання

Метою роботи є аналіз та систематизація існуючих підходів до оцінювання рівня знань для виявлення їх переваг та недоліків, можливості практичного застосування в системах тестування.

Виклад основного матеріалу

Вирішення проблеми оцінювання в автоматизованих системах складається з трьох етапів (рис. 1).

У репозиторії розміщуються метадані (складність, значимість) та параметри (число питань, час відповіді).

Найпростішою моделлю оцінки знань є модель, яка враховує кількість накопичених балів:

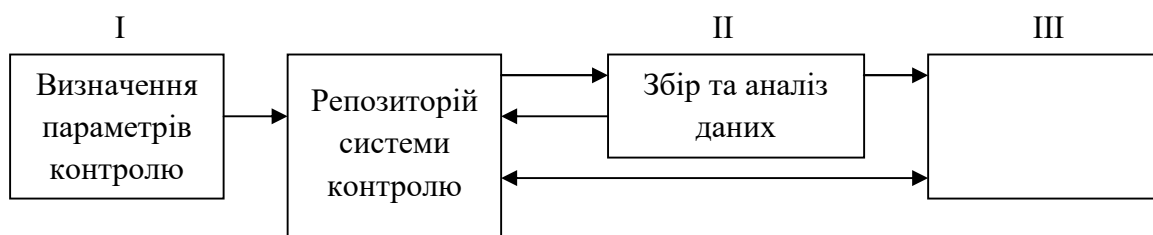


Рис. 1. Модель оцінки знань

$$S = \frac{\sum_{j=1}^k B_j}{n}, \quad (1)$$

де B_j – вірна відповідь на j питання; k – кількість вірних відповідей з n запропонованих ($k \leq n$).

Недоліком цієї моделі є її залежність від єдиного параметру (кількості правильних відповідей), тобто вона не враховує неповністю правильні відповіді та характеристики завдань. Ця модель має найнижчу надійність, бо не дозволяє об'єктивно оцінити знання студента.

1. Математичні моделі, які враховують параметри завдань, при виставленні оцінки використовують характеристики контрольних питань.

Існують різні модифікації даного типу моделей:

1.1.1 Модель, яка враховує час виконання завдання та/або загальний час контрольної роботи. Для правильних відповідей значення B_j розраховується за формулою:

$$B_j = \begin{cases} 1, & t \leq t_{\max} \\ 0, & t > t_{\max} \end{cases}, \quad (2)$$

де t – час виконання завдання; t_{\max} – час, який виділяється на виконання завдання.

Далі підсумкова оцінка розраховується аналогічно «Простої моделі» за формулою (1).

1.1.2 Моделі на основі рівня засвоєння [4]. У цій моделі характеристикою завдання є рівень його засвоєння. Завдання поділяються на п'ять груп, які відповідають рівням засвоєння: розуміння, пізнання, відтворення, використання, творча діяльність. Для кожного завдання формується набір вагомих операцій. Під вагомими ро-

зуміють операції, що виконуються на рівні, який перевіряється. Операції, які належать до більш низького рівня, в число вагомих не входять.

Для виставлення оцінки використовується коефіцієнт K_α :

$$K_\alpha = \frac{P_1}{P_2}, \quad (3)$$

де P_1 – кількість правильно виконаних вагомих операцій у тесті; P_2 – загальна кількість вагомих операцій; $\alpha = 0, 1, 2, 3, 4$ – коефіцієнти, які відповідають рівням засвоєння.

Оцінка формується на основі заданих граничних значень за співвідношенням:

$K_\alpha < 0,7$ – незадовільно;

$0,7 \leq K_\alpha < 0,8$ – задовільно;

$0,8 \leq K_\alpha < 0,9$ – добре;

$K_\alpha \geq 0,9$ – відмінно.

1.1.3 Метод лінійно – кускової апроксимації [3]. Алгоритм оцінювання заснований на класифікації завдань (питань) за їх дидактичними характеристиками (значимість (z), складність (d), специфікація (s)). Число балів, отриманих студентом за виконання n завдань, визначається за формулою:

$$y = \sum_{i=1}^n w_i x_i, \quad (4)$$

де x_i – оцінка за виконання i -го завдання; n – число завдань; $W = \{w_1, w_2, \dots, w_n\}$ – вектор вагомих коефіцієнтів завдань, який залежить від їх дидактичних характеристик.

Після завершення контролю визначається середній бал A , отриманий студентом за виконання n завдань ($A = y / k_n$, де k_n –

кількість спроб виконання n завдань, $k_n \geq n$) і уточнений середній бал A' :

$$A' = A + \alpha_1 r + \alpha_2 \frac{k - n}{n} + \alpha_3 \frac{k_c}{n} + \alpha_4 \frac{k_b}{n}, \quad (5)$$

де r – ранг об'єкта навчання (1, 2, або 3); k_n – кількість спроб виконання n завдань; k_c – кількість звернень до довідкової інформації; k_b – кількість завдань, виконаних з перевищенням відведеного часу ($k_b \leq n$); $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ – коефіцієнти.

Далі оцінка формується за формулою:

$$I = \begin{cases} 1, & Q \leq c_1; \\ 2, & c_1 < Q \leq c_2; \\ \dots & \\ M, & Q > c_{M-1}, \end{cases} \quad (6)$$

де I – оцінка за тест; $\{c_1, c_2, \dots, c_M\}$ – вектор граничних значень; M – максимально можлива оцінка.

Аналогічно визначається і рівень засвоєння знань (ранг) студента. Перевага даної моделі: використання як чотирьох дидактичних характеристик завдань, так і рівня підготовленості (рангу) студента, що дозволяє підвищити надійність результатів контролю.

1.2 Моделі на основі ймовірнісних критеріїв. Головним у даних математичних моделях контролю знань є твердження про залежність ймовірності правильної відповіді студента від рівня його підготовки і від параметрів завдання. Суть цих моделей полягає в тому, що на основі відомих апіорних ймовірностей розраховуються апостеріорні ймовірності $P(H_i)$ гіпотези H_i , що студент заслуговує оцінку i . При обчисленні ймовірності $P(H_i)$ враховуються: складність і час виконання завдань; число запропонованих завдань; число неправильно виконаних завдань та ін. Розраховані ймовірності аналізуються й/або порівнюються із граничними значеннями, з огляду на ризики недооцінки й переоцінки формування оцінки i . Якщо отримані результати однозначно дозволя-

ють виставити оцінку, то контроль, як правило, завершується. У протилежному випадку студентові видається чергове завдання.

Різні модифікації моделі успішно застосовуються і в цей час.

1.2.1 Біноміальний розподіл випадкових подій.

В роботах Коджи Т.І., Нарожного О.В. [5,6] використовується біноміальний розподіл випадкових подій угадування як правильної, так і невірної відповіді.

$$R_n = \sum_{r=m}^n P_{nr} = \sum_{r=m}^n \frac{n!}{r!(n-r)!} P^r (1-P)^{n-r}, \quad (7)$$

де n – число питань в тесті; r – число правильних відповідей, при якому виставляється позитивна оцінка; m – мінімальне число правильних відповідей для одержання задовільної оцінки; P – ймовірність випадкового угадування правильної відповіді на одне питання, $P = \frac{1}{C}$; C – число

можливих варіантів введення результативної відповіді; P_{nr} – ймовірність одержання r_i правильних відповідей.

1.2.2 Оцінка на основі нормального закону [7].

Теоретичним обґрунтуванням того, що розподіл рівня навченості відповідає нормальному (гаусовському закону), є одна з центральних граничних теорем, згідно якої розподіл середнього значення m незалежних результатів перевірки знань об'єктів навчання з рівнями навченості, розподіленими за різними законами, але з кінцевими математичними очікуванням та дисперсією, при збільшенні числа спостережень у вибірці (тобто при $n \rightarrow \infty$) наближається до нормального.

Не дивлячись на те, що застосування центральної граничної теореми обумовлене великою кількістю завдань протягом сеансу тестування, розподіл середнього вибіркового прагне до нормального навіть при відносно невеликих значеннях n , якщо величина дисперсії досягнень якого-небудь індивіда з даного класу навчання не є переважаючою і розподіл досягнень об'єктів не дуже відхиляється від нормального.

Тоді щільність розподілу ймовірностей навчальних досягнень описується функціями вигляду:

$$f_1(x) = \frac{1}{\delta\sqrt{2\pi}} \exp\left[-\frac{(x-\mu_1)^2}{2\delta^2}\right];$$
$$f_2(x) = \frac{1}{\delta\sqrt{2\pi}} \exp\left[-\frac{(x-\mu_2)^2}{2\delta^2}\right], \quad (8)$$

де μ_1 , μ_2 та δ – відповідно параметри, що характеризують центри розподілів і їх масштаб. Величина μ_1 – середнє значення тестових завдань, відповіді на які не відповідають вимогам, величина μ_2 – середнє значення тестових завдань, відповіді на які відповідають вимогам. Вважаємо, що дисперсії нормальних розподілів однакові і рівні δ^2 .

1.2.3 Оцінка на основі перевірки статистичних гіпотез.

Під статистичною гіпотезою розуміють будь-який вислів про генеральну сукупність (випадкову величину), який перевіряється за наслідками спостережень. Для завдання оцінки рівня навчальних досягнень тестованих, під яким розуміється ймовірність правильного виконання m завдань тесту, статистичною гіпотезою є наступні твердження:

– якщо процес тестування представлений схемою Бернуллі (7) та ймовірність p правильного виконання завдання невідома, приймається гіпотеза про те, що $p = p_0$; ця гіпотеза перевіряється на основі отриманої вибірки, тобто на основі результатів перевірки якості виконання респондентом завдань тесту;

– якщо вважаємо, що ймовірність правильного виконання завдань m є випадковою величиною, розподіленою за нормальним законом, причому респонденти атестовані позитивно, в середньому вирішують μ_2 завдань, а атестовані негативно – μ_1 завдань, то висувається гіпотеза у вигляді висловлювання: математичне очікування m рівне – μ_1 (або μ_2).

Отже, не маючи в своєму розпорядженні відомостей про всю генеральну су-

купність, за певними правилами на основі вибірових даних приймають або не приймають висловлену гіпотезу. Процедура зіставлення висловленої гіпотези з вибіровими даними називається перевіркою гіпотези.

Етапи перевірки гіпотези щодо оцінки рівня навчальних досягнень:

Етап 1. Маючи у своєму розпорядженні вибірові дані x_1, x_2, \dots, x_N та конкретні умови поставленого завдання, формулюють гіпотезу H_0 , яку називають основною або нульовою, а гіпотезу H_1 , яка конкурує з гіпотезою H_0 , називають альтернативною.

За спостережуваною вибіркою випадкової величини приймається рішення щодо справедливості для генеральної сукупності гіпотези H_0 або гіпотези H_1 .

Етап 2. Задаються величиною α – ймовірністю помилки першого роду, причому $\alpha = P(H_1 / H_0)$,

де $P(H_1 / H_0)$ – ймовірністю того, що буде прийнята гіпотеза H_1 тоді як для генеральної сукупності справедлива гіпотеза H_0 . Ймовірність α задається заздалегідь, зазвичай у вигляді стандартних значень: 0,05; 0,01; 0,005; 0,0001. Ймовірність помилки другого роду β визначається як:

$$\beta = P(H_0 / H_1),$$

де $P(H_0 / H_1)$, ймовірність того, що буде прийнята гіпотеза H_0 , тоді як для генеральної сукупності справедлива гіпотеза H_1 .

У результаті перевірки гіпотез H_0 і H_1 може бути прийнято і правильне рішення щодо стану генеральної сукупності. Тоді:

$$1 - \alpha = P(H_0 / H_0),$$

де $P(H_0 / H_0)$ – ймовірність прийняття рішення про гіпотезу H_0 , якщо вона насправді має місце.

$$1 - \beta = P(H_1 / H_1),$$

де $P(H_1 / H_1)$ – ймовірність прийняття гіпотези H_1 , коли вона вірна для даної генеральної сукупності.

Етап 3. Формують статистику φ , що є функцією вибірових даних, значення яких дозволяють судити про «розбіжність» вибірки з гіпотезою H_0 :

$$\varphi = \varphi(X_1, X_2, \dots, X_N). \quad (9)$$

При справедливості гіпотези H_0 значення статистики відповідають відомому закону розподілу. У результаті, величину φ називають критерієм.

Етап 4. З області допустимих значень критерію φ виділяють підобласть ω таких значень, які свідчать про істотну розбіжність вибірки з гіпотезою H_0 . Тоді, при попаданні значень φ у цю область, гіпотеза H_0 відкидається, а приймається гіпотеза H_1 . Підобласть ω називають критичною.

Необхідно мати на увазі, що таке рішення може бути помилковим: насправді для генеральної сукупності справедлива гіпотеза H_0 . Тобто, орієнтуючись на критичну ω , можна зробити помилку першого роду, ймовірність якої задана заздалегідь і дорівнює α . З цього слідує вимога до критичної підобласті ω : ймовірність того, що критерій φ прийме значення з критичної підобласті ω повинна дорівнюватися заданому числу α , тобто

$$P(\varphi \in \omega) = \alpha,$$

і критична величина ω повинна бути визначена так, щоб при заданій ймовірності помилки першого роду α ймовірність помилки другого роду β була б мінімальною з критичними значеннями ω .

Щодо функції розподілу значень статистики можливі три види розташування критичних значень ω :

– правобічна критична підобласть, що складається з інтервалу $(x_{кр, \alpha}^{пр}, +\infty)$, де

$x_{кр, \alpha}^{пр}$ визначається з умови:

$$P(\varphi > x_{кр, \alpha}^{пр}) = \alpha$$

– лівобічна критична область, що складається з інтервалу $(-\infty, x_{кр, \alpha}^{лів})$, де

$x_{кр, \alpha}^{лів}$ визначається з умови:

$$P(\varphi < x_{кр, \alpha}^{лів}) = \alpha$$

– двобічна критична область, що складається з наступних двох інтервалів:

$(-\infty, x_{кр, \alpha/2}^{лів})$ і $(x_{кр, \alpha/2}^{пр}, +\infty)$, де точки

$x_{кр, \alpha/2}^{лів}$ і $x_{кр, \alpha/2}^{пр}$ визначаються з умов:

$$P(\varphi < x_{кр, \alpha/2}^{лів}) = \frac{\alpha}{2} \text{ і } P(\varphi > x_{кр, \alpha/2}^{пр}) = \frac{\alpha}{2}.$$

Етап 5. На підставі спостережень за вибірковими даними у формулу критерію (9) замість X_1, X_2, \dots, X_N підставляють числа та обчислюють значення критерію φ . Якщо це значення потрапляє в критичну підобласть ω , то гіпотеза H_0 відкидається і приймається гіпотеза H_1 . Якщо не потрапляє, то гіпотеза H_0 не відкидається. Це не означає, що H_0 єдина гіпотеза. Тоді слід вважати, що розбіжність між вибірковими даними та гіпотезою H_0 не суперечить результатам спостережень.

В якості методу перевірки гіпотези може використовуватися послідовний аналіз [7].

1.2.4 Модель Раша. Ця модель є моделлю адаптивного тестування. Теоретичною основою адаптивного контролю є теорія IRT у поєднанні з дидактичним принципом індивідуалізації навчання. Адаптивним тестуванням знань називається спосіб екзаменаційного контролю рівня підготовки тестуючого, при якому процедура вибору та пред'явлення йому чергового тестового завдання на $(t + 1)$ -му кроці тестування визначається відповідями на попередніх t кроках тесту [8]. Математичну основу такого методу складає модель об'єднання тестових завдань в тематичній послідовності із зваженим ранжуванням як окремих завдань, так і цілих послідовностей та виведенням підсумкової оцінки за тест з урахуванням нормованої суми балів, що накопичується за вибраними варіантами відповідей.

Модель Раша є однопараметричною логістичною функцією Раша, яка відображає ймовірність правильної відповіді на питання j :

$$P_j \{x_{ij} = 1 | \beta_j\} = \frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}}, \quad (10)$$

де P_j – ймовірність вірної відповіді i -го студента на j -те завдання; θ_i – рівень знань i -го студента (в логітах); β_j – рівень складності j -го завдання.

Г. Раша запровадив дві міри рівня: «логіт рівня знань», який визначається як $\ln \frac{P_i}{q_i}$ та «логіт рівня складності завдання» $\ln \frac{P_j}{q_j}$. Першу він визначив як натуральний логарифм відношення частки правильних відповідей i -го випробовуваного на всі завдання тесту, до частки неправильних відповідей, а другу – як натуральний логарифм іншого відношення – доли неправильних відповідей на j -те завдання тесту до частки правильних відповідей на теж саме завдання для безлічі випробовуваних. Єдина логарифмічна шкала дозволяє встановити необхідну відповідність між рівнем засвоєння та складністю завдання і, більш того, провести корекцію результатів тестування при тестах різної складності.

Статистичні оцінки рівня підготовленості θ та рівня складності β будуть незміщеними оцінками, якщо їх математичне очікування при будь-якому об'ємі вибірки випробовуваних буде дорівнюватися параметру, який оцінюється. На практиці зазвичай використовують асимптотичну незміщену оцінку, математичне очікування якої прагне до дійсного значення параметра, який оцінюється, при необмеженому збільшенні об'єму вибірки. Для отримання оцінки параметрів β та θ застосовуються методи моментів або найбільшої правдоподібності.

1.2.5 Моделі уточнюючих питань

Концепція базується на автоматизації методики уточнюючих питань, яка широко використовується в педагогічній практиці для виявлення «глибини» знань. Відносна важливість питань визначається їх вагови-

ми коефіцієнтами, що враховуються при підведенні результатів тестування.

Кожен ланцюжок є послідовністю близьких за тематикою питань, що формулюються для уточнення знань. Чергове питання в ланцюжку задається тільки після отримання відповіді на попереднє питання. Залежно від стратегії тестування, що обирається організатором контролю знань, чергове питання в ланцюжку може пред'являтися до першої помилки («строгий» викладач), або надається можливість демонструвати максимум знань, відповідаючи на всі питання даної тематичної послідовності.

У ланцюжок може об'єднуватися необмежена кількість тематично близьких питань. Процедура кількісного оцінювання знань складається з трьох етапів. На першому розраховуються бали, набрані за правильні відповіді в рамках кожної окремої тематичної послідовності. На другому етапі розраховується сумарний бал S за відповіді на всі питання тесту з урахуванням кількості ланцюжків питань, на які встиг відповісти тестуючий за відведений час. На третьому етапі визначається підсумкова оцінка знань. Для цього набраний ним сумарний бал S проєктується на шкалу оцінювання, що має вигляд $[0; I_1; I_2; I_3; 1]$, де $0 < I_1 < I_2 < I_3 < 1$ – межі інтервальних діапазонів оцінок, що задаються викладачем при організації тестування.

1.2.6 Моделі на основі нечітких множин

Методи, що базуються на теорії нечітких множин, відносяться до методів оцінки і прийняття рішень в умовах невизначеності. Багато тестів є складними комп'ютерними системами, мають ієрархічну структуру, містять допоміжні або повчальні питання тощо. Такі тести можна назвати «чіткими», оскільки в них фактично використовується двійкова система оцінки правильності відповіді на кожне питання (правильно або неправильно).

Основним недоліком «чітких» тестів є оцінка неповних або неточних відповідей як

неправильних, тобто наближене, неповне знання відповіді на питання кваліфікується як незнання відповіді, а це не завжди виправдано. В результаті «чіткі» тести дуже жорстко, недостатньо справедливо оцінюють рівень знань фахівця, занижують оцінку.

Застосування ідеї «нечітких» тестів дозволяє підвищити ефективність процесу ідентифікації знань на основі припущення, що всі варіанти відповіді на питання є правильними, причому ступінь їх правильності визначається значенням так званої функції приналежності $F(X)$ варіанту відповіді X , що приймає значення від 0 до 1 включно. Чим правильніша відповідь, тим більше значення цієї функції. Такі тести умовно можна назвати «нечіткими», їх доцільно застосовувати, якщо метою тестування є визначення рівня ознайомлення тестованого з деякою предметною областю. «Нечіткі» тести не слід застосовувати, якщо від тестованого потрібні знання на рівні інструкцій, а також - в тих областях діяльності, де неточні або неповні знання можуть завдати істотного збитку. Для формування оцінки за результатами тестування в якій-небудь n -бальній рівномірній шкалі досить помножити розмірність шкали n на отримане значення. Складність «нечіткого» тесту залежить від близькості неточних варіантів відповіді до цілком правильного варіанту. Чим ця «відстань» менша, тим важче ефективно відповідати на питання.

2. Основна ідея класифікаційних моделей полягає у віднесенні студента до одного із стійких класів з урахуванням сукупності ознак, що визначають даного студента. При цьому використовується спеціальна процедура обчислення ступеня схожості (оцінки) рядка, що розпізнається (сукупності ознак об'єкта навчання), з рядками, приналежність яких до класів заздалегідь відома.

Алгоритм, заснований на обчисленні оцінок (АОО), був уперше запропонований Ю.І. Журавльовим [9] і пізніше використовувався для класифікації за рівнем підготовленості [3] і для оцінки знань як додатковий метод в навчальних системах. Дана мо-

дель передбачає побудову таблиці навчання T_{nm} , в якій кожен рядок є набором ознак, що характеризують роботу студента в процесі КЗ: кількість запропонованих завдань (n), середній бал (A), кількість спроб виконання завдань (k_n), кількість звернень до довідкової інформації (k_c), ранг (r). При виставленні оцінки обчислюється ступінь схожості сукупності ознак конкретного студента $I(S) = \{\beta_1, \beta_2, \dots, \beta_m\}$ рядкам, що входять в таблицю навчання T_{nm} на підставі чого відносять його до певного класу K_j . Для цього обчислюється число рядків кожного класу K_j , близьких за обраним критерієм класифікаційному об'єкту S . Рядок таблиці навчання $T_{nm}I(S_i) = \{\alpha_{i1}^j, \dots, \alpha_{im}^j\}$ і рядок, який розпізнається $I(S) = \{\beta_1, \beta_2, \dots, \beta_m\}$ вважаються схожими, якщо виконуються нерівності $|\alpha_{ik}^j - \beta_k| \leq \varepsilon_k$ де $\varepsilon_k (k=1, \dots, m)$ – точність порівняння. Студент відноситься до того класу K_j , який має максимальну оцінку $\max K_j(S, K_j), j = 1, m$.

Висновки і перспективи досліджень

Сучасні методи навчання та контролю знань повинні ґрунтуватися на широкому застосуванні засобів автоматизації, які не тільки підвищують ефективність використаних інформаційних технологій, а й створюють умови для подальшого їх удосконалення. Використання статистичних методів та ЕОМ при розробці та апробації тестів забезпечує можливість ліквідувати такі недоліки тестування як:

- можливість угадування студентами правильних відповідей;
- можливість помилкової оцінки;
- складність розробки науково обґрунтованих тестів, які містять високі та стійкі показники надійності та валідності.

Актуальність проблеми діагностики процесу навчання не знижується і в теперішній час, особливо при зрості тенденції впровадження комп'ютерної техніки в навчальний процес та передачі технічним пристроям навчальних та контролюючих функцій викладача.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Волков Н.И., Алексеев А.Н., Алексеев Н.А. Тестовый контроль знаний: учебное пособие. Сумы: ИТД «Университетская книга», 2004. 109 с.
2. Ягодзінський А.Й., Муровцева А.О., Иванова Л.В. та ін. Оцінка знань студентів та якості підготовки фахівців: навч. посібник. Київ: ІЗМН, 1997. 216 с.
3. Зайцева Л.В., Прокофьева Н.О. Модели и методы адаптивного контроля знаний. Рига: Рижск. политехн. ин-т., 2004. URL: http://ifets.ieee.org/russian/depositary/v7_i4/pdf/1.pdf (дата звернення 31.03.2019).
4. Соловов А.В. Проектирование компьютерных систем ученого назначения: учебное пособие. Самара: СГАУ, 1995. 138с.
5. Коджа Т. І. Автоматизована система управління та контролю знань в процесі навчання: дис. ... канд. техн. наук: 05.13.06. Одеса, 2003. 233 с.
6. Нарожний О.В. Моделі інтелектуального управління процесом навчання за допомогою мультиагентних систем: автореф. дис. на здобуття наук. ступеня канд. техн. наук: 05.13.23 «Системи та засоби штучного інтелекту». Одеса, 2007. 19 с.
7. Васильев В.И., Тягунова Т.Н. Основы культуры адаптивного тестирования. Москва: Издательство ИКАР, 2003. 584 с.
8. Современные информационные технологии в учебном процессе: тезисы докладов уч.-мет. конференции 25-26 апреля 2000 г. Ростов. URL: <http://www.uic.rsu.ru/~nprohoro/DO> (дата звернення 31.03.2019).
9. Журавлев Ю.И. Об алгебраическом подходе к решению задач распознавания или классификации. *Проблемы кибернетики*. Москва: Наука, 1978, № 33. С. 5-68.

Kateryna KUZMA
Mykolayiv

ANALYSIS OF METHODS AND MODELS FOR ASSESSMENT KNOWLEDGE IN TESTING SYSTEMS

The paper considers models and methods for assessing the level of knowledge in the testing systems. Their classification into two categories was executed: mathematical and classification, computational algorithms were given, advantages and disadvantages were revealed.

It was determined that the main disadvantage of «clear» tests (using a binary system to assess the correctness of the answer to each question (right or wrong)) is a evaluation of incomplete or inaccurate responses as wrong, that is, approximate, incomplete knowledge of the answer to the questions qualifies as ignorance of the answer, and this is not always justified. Applying the idea of «fuzzy» tests can improve the efficiency of the process of identifying knowledge. The practical significance of the use of statistical methods and computers in the development and applying of tests was argued.

Keywords: *knowledge assessment methods, knowledge assessment model, testing system, classification.*

Екатерина КУЗЬМА
Николаев

АНАЛИЗ МЕТОДОВ И МОДЕЛЕЙ ОЦЕНКИ ЗНАНИЙ В СИСТЕМАХ ТЕСТИРОВАНИЯ

В работе рассмотрены модели и методы оценки уровня знаний в системах тестирования. Выполнено их систематизацию на две категории: математические и классификационные, приведены вычислительные алгоритмы, выявлены преимущества и недостатки.

Определено, что основным недостатком «четких» тестов (используется двоичная система оценки правильности ответа на каждый вопрос (правильно или неправильно)) является оценка неполных или неточных ответов как неправильных, то есть приближенное, неполное знание ответа на вопрос квалифицируется как незнание ответа, а это не всегда оправдано. Применение идеи «нечетких» тестов позволяет повысить эффективность процесса идентификации знаний. Аргументированно практическую значимость использования статистических методов и ЭВМ при разработке и апробации тестов.

Ключевые слова: *методы оценки знаний, модели оценки знаний, системы тестирования, классификация.*

Стаття надійшла до редколегії 31.03.2019