

УДК 378.147.044.4'24(477)

Кузьма К. Т.

Канд. техн. наук, старший викладач кафедри комп'ютерної інженерії, Миколаївський національний університет ім. В. О. Сухомлинського, Миколаїв, Україна

ОБЧИСЛЮВАЛЬНА ТЕХНОЛОГІЯ ПЕРЕВІРКИ РІВНЯ ЗНАНЬ НА ОСНОВІ МЕТОДУ ПОСЛІДОВНОГО АНАЛІЗУ

Актуальність. Вирішено актуальну задачу підвищення ефективності процесу підтримки прийняття рішень під час статистичного контролю знань.

Мета роботи – розробка обчислювальної процедури вирішення чотирьохальтернативної задачі класифікації тестованих за рівнем навченості, що дозволяє здійснювати контроль знань диференційовано, мінімізує об'єм завдань, необхідний для виконання.

Метод. Запропоновано обчислювальну процедуру класифікації тестованих на чотири класи, які відповідають рівням навченості: «початковий», «середній», «достатній», «високий», що базується на використанні двохальтернативного критерію послідовного аналізу в декілька етапів та забезпечує здійснення контролю знань в процесі виконання завдань, мінімізуючи таким чином час перевірки знань, що дозволяє автоматизувати процес перевірки статистичних гіпотез у системах тестування та навчання з метою диференційної оцінки знань учасників навчального процесу. Для вирішення задачі оцінки придатності тесту запропоновано метод, що базується на побудові функції оперативної характеристики послідовного критерію, яка дозволяє визначити обсяг завдань достатній для досягнення бажаного рівня якості тесту за рахунок встановлення зв'язку між очікуваною ймовірністю прийняття гіпотези та випадковим значенням параметра ймовірності появи у виборці з $1, 2, \dots, n$ питань приймального числа невірно виконаних завдань.

Результати. Розроблено програмне забезпечення, яке реалізує запропоновану обчислювальну процедуру, що використано при проведенні обчислювальних експериментів тестового контролю знань.

Висновки. Проведені експерименти підтвердили працездатність запропонованої процедури і програмного забезпечення, що її реалізує, а також дозволяють рекомендувати їх для застосування на практиці для рішення задач автоматизованої перевірки рівня знань.

Ключові слова: послідовний аналіз, перевірка рівня знань, класифікація тестованих за рівнем знань, перевірка гіпотез, критеріально-орієнтований тест.

НОМЕНКЛАТУРА

ОТ – обчислювальна технологія;

СППР – система підтримки прийняття рішень;

d_h – число невірно виконаних завдань серед h перевірених;

h – кількості питань у виборці;

H_i – гіпотеза щодо віднесення об'єкта навчання до i -го класу навченості;

$L(p)$ – оперативна характеристика статистичного критерію;

n – загальна кількість тестових питань;

p_i – ймовірність невиконання завдань для i рівня навчальних досягнень, $i = \overline{1,4}$;

$f(d)$ – функція щільності розподілу неправильних відповідей;

$f(Q)$ – функція оцінювання знань;

μ_1 – математичне очікування числа правильно виконаних завдань для неатестованих об'єктів навчання, що відповідає гіпотезі A_1 – об'єкт навчання не володіє достатніми знаннями та не атестується;

μ_2 – математичне очікування числа правильно виконаних завдань для атестованих об'єктів навчання, що відповідає гіпотезі A_2 – об'єкт навчання володіє достатніми знаннями та атестується позитивно;

σ^2 – дисперсія випадкової величини X коректно виконаних завдань для атестованих та неатестованих об'єктів навчання;

X_0 – допустиме (прийнятне) число коректно виконаних завдань;

Pa_i – приймальна доля невірно виконаних завдань для i -ї класифікації;

Pr_i – неприймальна доля невірно виконаних завдань для i -ї класифікації;

g_i – i -та група із h завдань;

a_h – прийнятне число невірно виконаних завдань для певного класу;

r_h – неприйнятне число невірно виконаних завдань для певного класу;

$a_i(h)$ – пороги прийняття гіпотези щодо зарахування об'єкта навчання до одного з чотирьох класів навченості, $i = \overline{1,3}$;

$r_i(h)$ – пороги неприйняття гіпотези щодо зарахування об'єкта навчання до одного з чотирьох класів навченості, $i = \overline{1,3}$;

m – номер питання;

L – вага відповіді: $L = 0$ – вірна відповідь, $L = 1$ – невірна відповідь;

q – стандарт оцінювання;

$n_{\text{ср}_{\text{max}}}$ – максимальний обсяг завдань, відповіді на які необхідно перевірити;

$P_n(d)$ – ймовірність появи d невірних відповідей з перевірених n -завдань;

α – ймовірність помилки 1-го роду, ймовірність неприйняття головної гіпотези, якщо доля невірних відповідей менша за критеріальну;

β – ймовірність помилки 2-го роду, ймовірність прийняття головної гіпотези, якщо доля невірних відповідей більша за критеріальну;

C – прийнятне число невірних відповідей для заданої класифікації;

C_{ij} – ваги прийняття рішень, $i = \overline{1,2}$; $j = \overline{1,2}$.

ВСТУП

Тест, як інструмент стандартизованої процедури проведення і задалегідь спроектованої технології обробки та аналізу результатів, є зручним засобом вимірювання навчальних досягнень. Тестування як форма контролю навчальних досягнень студентів в останні роки все частіше стала застосовуватися в системі вищої, професійної освіти, оскільки є технологією, що дозволяє об'єктивно і швидко оцінити рівень досягнень кожного студента.

За цілями застосування педагогічні тести поділяються на два класи – нормативно-орієнтовані і критеріально-орієнтовані. Нормативно-орієнтований тест (norm-referenced test) дозволяє ранжувати випробовуваних за рівнем знань. Критеріально-орієнтований тест (criterion-referenced test) дозволяє виявити рівень засвоєння випробовуваним певного розділу в заданій предметній галузі. Зазвичай тестовий бал відображає частку правильних виконаних завдань і виражається у відсотках. При використанні критеріально-орієнтованого підходу особлива увага приділяється методиці оптимального вибору критеріального балу (або балів). Саме при критеріально-орієнтованому тестуванні задача перевірки знань розглядається як задача підтримки прийняття рішень та класифікації тестованих за рівнем їх підготовки.

Рішення цієї задачі забезпечує необхідну об'єктивність при оцінці та контролі знань, можливість мінімізації об'єму завдань, який повинен вирішити опитуваний. Метою роботи є розробка обчислювальної процедури вирішення чотирьохальтернативної задачі класифікації тестованих за рівнем навченості, що дозволяє здійснювати контроль знань диференційовано, мінімізує об'єм завдань, необхідний для виконання.

1 ПОСТАНОВКА ЗАДАЧІ

На основі проведених досліджень, результати яких наведено в роботі [1], для вирішення задачі прийняття рішень щодо класифікації об'єктів навчання під час тестування обрано метод послідовного аналізу, оскільки його застосування дозволяє підвищити ефективність процесу перевірки знань за рахунок мінімізації часових витрат на його проведення.

На сьогодні для процесу ідентифікації рівня знань учасників навчального процесу з використанням методу послідовного аналізу розроблено обчислювальні процедури перевірки двохальтернативних випадків, наприклад, класифікації об'єктів навчання на дві групи: атестовані та неатестовані. Постає задача розробки алгоритму класифікації тестованих відносно декількох гіпотез. Правильним прийняття рішень є формування границь, за допомогою яких здійснюється класифікація тестованих на чотири групи (I група – «високий» рівень навченості, II – «достатній», III – «середній», IV – «початковий»).

Дано відповіді (x_1, \dots, x_h) на h запитань із загальної кількості завдань n . Припустимо, що $x_i = 0$, якщо завдання виконано вірно, $x_i = 1$, якщо завдання виконано невірно. Нехай p_i визначає відносне число невиконаних завдань. Для процесу перевірки відповідей на тестові питання під час контролю знань постає задача перевірки гіпотези про те, що p_i на деякому діапазоні h виконаних завдань із загальної кількості завдань не перевищує деякої заданої величини p'_i . Таким чином, задачу оцінки навчальних досягнень необхідно вирішити шляхом послідовної перевірки гіпотез $p_i \leq p'_i$ проти гіпотез $p_i > p'_i$, де $i = \overline{1,4}$ – номер класу навченості.

2 ОГЛЯД ЛІТЕРАТУРИ

Задача обробки даних педагогічного тестування та контролю знань за критеріально-орієнтованою методикою в роботах [2–5] розглядається як задача підтримки прийняття рішень при класифікації об'єктів навчання (тестованих) за рівнем підготовки, для вирішення якої використовуються методи теорії статистичних рішень (критерій Байєса, Неймана-Пірсона, мінімакса, Вальда). Застосування даних методів спрямоване на створення систем статистичного контролю знань за альтернативною ознакою.

При тестуванні за альтернативною ознакою використовується замкнена форма тесту, характеристиками якої є: функція щільності розподілу неправильних відповідей – $f(d)$, прийнятний рівень неправильних відповідей – p_0 , неприйнятний рівень неправильних відповідей – p_1 , ризик «заниженої» оцінки знань – α , ризик «завищеної» оцінки знань – β , функція оцінювання знань – $f(Q)$, обсяг освітньої інформації – N , обсяг вибірки завдань тесту – n та критерій прийняття розв'язків у вигляді граничного числа неправильних відповідей – X .

У роботі [2] запропоновано правило прийняття рішень, яке визначає допустиме (прийнятне) число коректно виконаних завдань X_0 (границя прийняття рішень) на основі критерію Байєса та поділяє всіх тестованих на дві групи: атестовані та неатестовані. Умовні щільності ймовірності випадкової величини X коректно виконаних завдань для атестованих та неатестованих об'єктів навчання подаються у вигляді гаусовських законів розподілення:

$$f(X/\mu_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(X - \mu_i)^2}{2\sigma^2}\right\}, i = 1, 2.$$

Стратегія прийняття рішень відносно границі X_0 має чотири результати: вірна гіпотеза A_1 та приймається рішення про A_1 ; вірна гіпотеза A_1 , а приймається рішення про A_2 ; вірна гіпотеза A_2 та приймається рішення про A_2 ; вірна гіпотеза A_2 , а приймається рішення про A_1 .

Перший та третій результат є правильними рішеннями, другий та четвертий – помилковими.

Байєсовський критерій базується на двох припущеннях: апіорні ймовірності гіпотез P_1 та P_2 задані та $P_1 + P_2 = 1$; задані ваги чотирьох зазначених дій: C_{11} , C_{12} , C_{22} , C_{21} . Слідуючи правилам прийняття рішень для вибору або A_1 , або A_2 , простір спостережень N ділиться на дві

частини N_1 та N_2 . Якщо результат тестування потрапляє в N_1 , то приймається гіпотеза A_1 , якщо в $N_2 - A_2$.

Байєсовський критерій будує рішення так, щоб у середньому втрати та ризику були мінімальними. Значення величини очікуваного ризику подається виразом:

$$R = P_1 C_{21} + P_2 C_{22} + \int_{N_1} \{ [P_2 (C_{12} - C_{22}) f_2(X/\mu_2)] - [P_1 (C_{21} - C_{11}) f_1(X/\mu_1)] \} dx.$$

Щоб мінімізувати ризик R за умови, що другий член підінтегрального виразу більший, ніж перший (значення вірних відповідей менше значення невірних), необхідно всі значення X включати до N_1 та навпаки. Таким чином, області рішень N_1 та N_2 , які відповідають гіпотезам A_1 та A_2 визначаються за наступних умов:

$$\text{якщо } P_2 (C_{12} - C_{22}) f_2\left(\frac{X}{\mu_2}\right) \geq P_1 (C_{21} - C_{11}) f_1\left(\frac{X}{\mu_1}\right), \quad (1)$$

то X відноситься до N_2 - приймається рішення щодо істинності гіпотези A_2 , та навпаки.

Вираз 1 подається у вигляді нерівності двох відношень:

$$\frac{f_2(X/\mu_2)^{A_2}}{f_1(X/\mu_1)} > \frac{P_1 (C_{21} - C_{11})}{A_1 P_2 (C_{12} - C_{22})}.$$

Ліва частина називається відношенням правдоподібності і визначається $\Lambda(x)$. Права частина називається границею прийняття рішення й визначається η та $\epsilon \text{ const}$, оскільки залежить від апіорних параметрів задачі. Таким чином, Байєсовський критерій зводиться до критерію відношення правдоподібності, записується у вигляді нерівності $\Lambda(x)^{A_2} > A_1 \eta$ та трактується наступним чином: якщо величина відношення правдоподібності для двох гіпотез більше границі прийняття рішень, то приймається гіпотеза A_2 , якщо менше, то - A_1 .

Мінімакний критерій використовується при відсутності інформації щодо значення апіорних ймовірностей гіпотез A_1 та $A_2 - P_1$ та P_2 відповідно. Якщо не визначені апіорні ймовірності гіпотез P_1 та P_2 і не задані ваги прийняття рішень C_{ij} , то для рішення задачі використовують критерій Неймана-Пірсона та значення ймовірностей помилок першого та другого роду.

Алгоритми прийняття рішень на основі критеріїв Байєса, Неймана-Пірсона та мінімаксу щодо класифікації об'єктів навчання передбачають, що N завдань, які видаються для об'єктів навчання, та X - число вірно виконаних завдань фіксовані. Якщо кількість вірно виконаних завдань не фіксовано, то прийняття рішень щодо класифікації тестованих, не перевіряючи всіх завдань, здійснюється на основі послідовного аналізу відповідей з використанням критерію Вальда.

3 МАТЕРІАЛИ ТА МЕТОДИ

На основі методів, поданих у роботах [2, 6, 7], розроблено алгоритм для чотирьохальтернативної задачі класифікації учасників навчального процесу. Пропонується класифікувати об'єкти навчання за рівнем знань: I клас - об'єкти навчання, які володіють «високим» рівнем знань (A, від 90 до 100 балів); II клас - «достатнім»

(BC від 65 до 89 балів); III клас - «середнім» (DE від 50 до 64 балів); IV клас - «початковим» (F, FX від 1 до 49 балів).

Сутність класифікації полягає в розбитті інтервалу тестових завдань N на чотири області: N_1, N_2, N_3, N_4 . Якщо число невірних виконаних завдань d_h потрапило в область N_1 , то приймається гіпотеза H_1 - об'єкт навчання має «високий» рівень знань; якщо в N_2 , то приймається гіпотеза H_2 (відповідає «достатньому» рівню знань); якщо в N_3 - гіпотеза H_3 (відповідає «середньому» рівню знань); інакше - приймається гіпотеза H_4 (відповідає «початковому» рівню знань).

Розроблений алгоритм базується на використанні дво-хальтернативного критерію послідовного аналізу в декілька етапів. Якщо p_i визначає відносне число невиконаних завдань на деякому діапазоні h , то ймовірність отримання вибірки (x_1, \dots, x_h) обчислюється за формулою:

$$p_h = p^{d_h} (1-p)^{h-d_h}. \quad (3)$$

На першому етапі рішення задачі визначимо гіпотези оцінки рівня навчальних досягнень тестованих. Оскільки ймовірність невиконання завдань p_i є випадковою величиною, то приймаємо наступні гіпотези: H_1 : якщо ймовірність невиконання завдання дорівнює p_1 ($p = p_1$) і тестований зараховується до класу I; H_2 : якщо ймовірність невиконання завдання дорівнює p_2 ($p = p_2$) і тестований зараховується до класу II; H_3 : якщо ймовірність невиконання завдання дорівнює p_3 ($p = p_3$) і тестований зараховується до класу III; H_4 : якщо ймовірність невиконання завдання дорівнює p_4 ($p = p_3$) і тестований зараховується до класу IV.

Допустима доля невірних відповідей для кожного класу p_i встановлюється на основі попередніх досліджень перевірки знань у контрольній групі або виходячи з ефективності контролю. При цьому перевірка здійснюється наступним чином. Із загальної кількості n завдань вибирається група g_1 , яка містить h завдань.

Перевірка закінчується прийняттям гіпотези про ате-стацію тестованого, якщо в групі g_1 для відповідного класу виконується умова: $a_h \geq d_h$. Перевірка закінчується прийняттям гіпотези про неатестацію тестованого, якщо в групі g_1 для відповідного класу виконується умова: $r_h \leq d_h$. Якщо $a_h < d_h < r_h$, обирається друга група g_2 , яка містить наступні h завдань. Знову тестований атестується позитивно, якщо загальне число невиконаних завдань у двох групах d_{2h} менше або дорівнює a_{2h} ; тестований неатестується, якщо $d_{2h} \geq r_{2h}$ та береться третя група g_3 із h запитань, якщо $a_{2h} < d_{2h} < r_{2h}$. Процес продовжується, поки тестований буде атестований або неатестований. Таким чином, коли спостереження проводяться над групами по h завдань, число визначених невірних виконаних завдань d_m порівнюється з відповідним приймальним числом a_m , або неприймальним числом r_m тільки при $m = h, 2h, 3h, \dots, n$.

Блок-схему ОТ перевірки рівня знань на основі методу послідовного аналізу наведено на рис. 1.

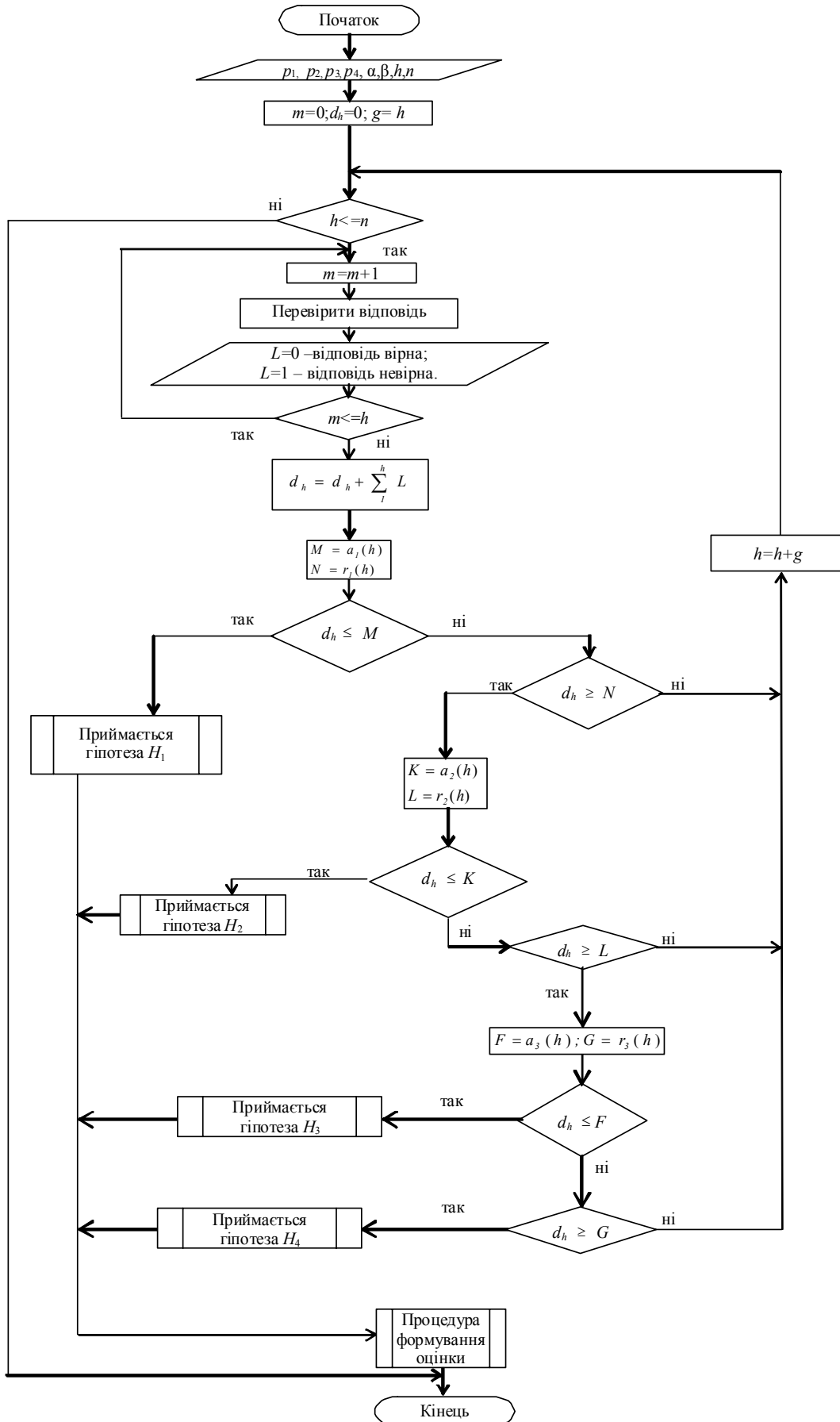


Рисунок 1 – Блок-схема алгоритму ОТ перевірки рівня знань на основі методу послідовного аналізу

Алгоритм ОТ передбачає виконання наступних етапів:

Етап 1. Уведення меж розбивки вихідного параметра – ймовірностей $p_i, i = \overline{1,4}$ невиконання завдань для кожного рівня навчальних досягнень.

Етап 2. Введення α, β, n, h .

Етап 3. Цикл за всіма діапазонами. Визначається на поточному діапазоні виконаних завдань h значення вихідного параметру d_h . Послідовно в поточному діапазоні обчислюються пороги прийняття $a_i(h), i = \overline{1,3}$ або неприйняття $r_i(h), i = \overline{1,3}$ гіпотези щодо зарахування об'єкта навчання до одного з чотирьох класів навченості за формулами:

$$a_i(h) = \frac{\ln \frac{\beta}{1-\alpha}}{\ln \frac{p_{r_i}}{p_{a_i}} - \ln \frac{1-p_{r_i}}{1-p_{a_i}}} + \frac{h \cdot \ln \frac{1-p_{a_i}}{1-p_{r_i}}}{\ln \frac{p_{r_i}}{p_{a_i}} - \ln \frac{1-p_{r_i}}{1-p_{a_i}}};$$

$$r_i(h) = \frac{\ln \frac{1-\beta}{\alpha}}{\ln \frac{p_{r_i}}{p_{a_i}} - \ln \frac{1-p_{r_i}}{1-p_{a_i}}} + \frac{h \cdot \ln \frac{1-p_{a_i}}{1-p_{r_i}}}{\ln \frac{p_{r_i}}{p_{a_i}} - \ln \frac{1-p_{r_i}}{1-p_{a_i}}}, i = \overline{1,3}.$$

Етап 4. Перевіряється, в яку область попадає значення d_h . Якщо значення d_h не попадає в границі відповідної області, здійснюється перехід до наступного діапазону питань $g_i(x_{h+1}, \dots, x_{ih})$ де $i = 2, \dots$. Якщо $h = n$, а рішення ще не прийнято, відбувається усічення послідовного критерію з використанням наступного правила: у випадку коли $d_h \geq (a_i(h) + r_i(h))/2$ приймається рішення щодо перевірки гіпотез H_{i+1}, \dots, H_4 , а при $d_h < (a_i(h) + r_i(h))/2$, приймається гіпотеза H_i , тестування завершується.

Для визначення ймовірності невиконання завдань $p_i, i = \overline{1,4}$ необхідно здійснити вибір стандарту оцінювання q_i . Методи вибору тестового стандарту засновані на експертних оцінках змісту тестового завдання. На основі дослідження даних методів, описаних у роботах [8–10], обрано метод Ангофф, який заснований на послідовних експертних оцінках змісту тестових завдань. Спочатку виконується вибір стандарту оцінювання для IV класу. Експерт-викладач для кожного завдання тесту встановлює ймовірність того, що мінімально компетентний студент дасть на нього вірну відповідь. Для однозначності експерту (або групі експертів) пропонується обрати значення ймовірності P_i зі значень 0,9, 0,8, ..., 0,1. Визначивши суму значень даних ймовірностей отримаємо критеріальний бал: $K = \sum_{i=1}^n P_i$. Стандарт оцінювання виз-

начається за формулою: $q = \frac{K}{n}$.

Тоді ймовірність невиконання завдань $p = 1 - q$. Після вибору критеріального балу для «початкового» рівня навченості, експерт проводить оцінку кожного тестового завдання вже на більш високий стандарт «середній», «достатній» та «високий».

Процедура контролю знань потребує вирішення задачі оцінки придатності тесту, що полягає у виборі таких критеріїв класифікації ($\alpha, \beta, p_i, i = \overline{1,4}$), які б зробили помилки першого та другого роду мало ймовірними та забезпечили визначення обсягу завдань достатнього для досягнення бажаного рівня якості тесту (ймовірність віднесення об'єктів навчання, які мають індивідуальний бал вищий або нижчий критеріального на величину не більше 10%, до певного класу навченості не повинна бути меншою за 0,8). Для розв'язку даної задачі використовується метод, що базується на побудові функції оперативної характеристики послідовного критерію відношення ймовірностей, яка дозволяє встановити зв'язок між очікуваною ймовірністю прийняття гіпотези та випадковим значенням параметра ймовірності появи у виборці з 1, 2, ... питань приймального числа невірно виконаних завдань:

$$L(p) = P_0 + P_1 + \dots + P_C = \sum_{d=0}^C P_n(d). \quad (4)$$

Якість обраних критеріїв класифікації визначається рівняннями: $L(p) \geq 1 - \alpha$, якщо $p = p_{\alpha_i}$; $L(p) \leq \beta$, якщо $p = p_{r_i}$. Значення оперативної характеристики статистичного критерію для кожного класу обчислюється за формулою:

$$L(p) \approx ([(1-\beta)/\alpha]^l - 1) / ([(1-\beta)/\alpha]^l - [\beta/(1-\alpha)]^l),$$

де параметр l змінюється від $-\infty$ до $+\infty$ та визначається із рівняння:

$$p_i = \left[1 - \left((1-p_{r_i}) / (1-p_{a_i}) \right)^l \right] / \left[\left(p_{r_i} / p_{a_i} \right)^{l-1} - \left((1-p_{r_i}) / (1-p_{a_i}) \right)^l \right], i = \overline{1,3}.$$

Максимальний обсяг завдань, відповіді на які необхідно перевірити для прийняття рішень щодо наявності відповідної гіпотези H_1, H_2, H_3 , пропонується обчислювати за формулою:

$$n_{\text{срmax}} = \sum_{i=1}^3 \left[- \left(\ln \frac{\beta}{1-\alpha} \right) \cdot \left(\ln \frac{1-\beta}{\alpha} \right) \right] / \left[3 \cdot \ln \frac{p_{r_i}}{p_{a_i}} \cdot \ln \frac{1-p_{a_i}}{1-p_{r_i}} \right].$$

4 ЕКСПЕРИМЕНТИ

Для проведення експерименту необхідно виділити контрольну групу студентів та з використанням СППР «ManageEdu» [11] або власноруч розробленого програмного продукту, який реалізовуватиме запропонований алгоритм класифікації тестованих, провести підсумковий контроль знань з будь-якої дисципліни за традиційною методикою та з використанням запропонованої обчислювальної технології перевірки рівня знань на ос-

нові методу послідовного аналізу. Програмний додаток повинен реалізовувати наступні функції: розраховувати ймовірнісні характеристики можливого віднесення тестованого до того чи іншого класу навченості ($p_1, p_2, p_3, p_4, \alpha, \beta, n, h$), формувати границі прийняття рішень для чотирьох гіпотез, для перевірки якості обраних параметрів будувати оперативні характеристики послідовного критерію для гіпотез H_1, H_2, H_3 , проведення контролю знань з використанням запропонованої ОТ та стандартним методом.

5 РЕЗУЛЬТАТИ

Для проведення експерименту було задіяно контрольну групу студентів (23 чоловіка) ВНЗ «Міжнародний технологічний університет «Миколаївська політехніка» III курсу економічного факультету з дисципліни «Фінансовий облік».

З використанням модуля «Последовательный анализ» СППР «ManageEdu» були сформовані ймовірнісні характеристики можливого віднесення тестованого до того чи іншого класу навченості: $p_1=0,1; p_2=0,35; p_3=0,6; p_4=0,75; \alpha = 0,05; \beta = 0,01; n = 100; h = 10$.

Значення параметрів границь прийняття рішень, сформовані на основі запропонованого алгоритму, наведено на рис. 2. в табличному та графічному вигляді.

Для перевірки якості обраних параметрів побудовано оперативні характеристики послідовного критерію для гіпотез H_1, H_2, H_3 . Аналіз графіка оперативної характеристики $L(p)$ для гіпотези H_1 підтверджує оптимальність обраних α, β , приймального числа p_1 та неприймального числа p_2 : очікуваний відсоток студентів, які будуть атестовані «на відмінно» при 10–16% невірних відповідей з масиву тестових завдань складає 80%, $L(p) > 80\%$. Оптимальна кількість завдань для даної класифікації – 24.

Для даного тесту середній обсяг завдань, який є достатнім для досягнення бажаного рівня якості процедури тестування, складає $n_{cp} = \frac{(n_{cpI} + n_{cpII} + n_{cpIII})}{3} = 45$.

Якщо при постійній кількості завдань збільшувати приймальне число відповідей, то контроль буде менш «жорстким» – збільшуються ймовірності атестації об'єктів навчання за відповідними рівнями навченості. Та навпаки: якщо при постійному значенні приймального числа збільшувати кількість завдань, – контроль буде більш «жорстким».

Після збереження отриманих значень вихідних параметрів процедури послідовного аналізу в діалоговому вікні налаштування параметрів теми на стороні клієнтської частини програмного комплексу відбувається процес контролю знань.

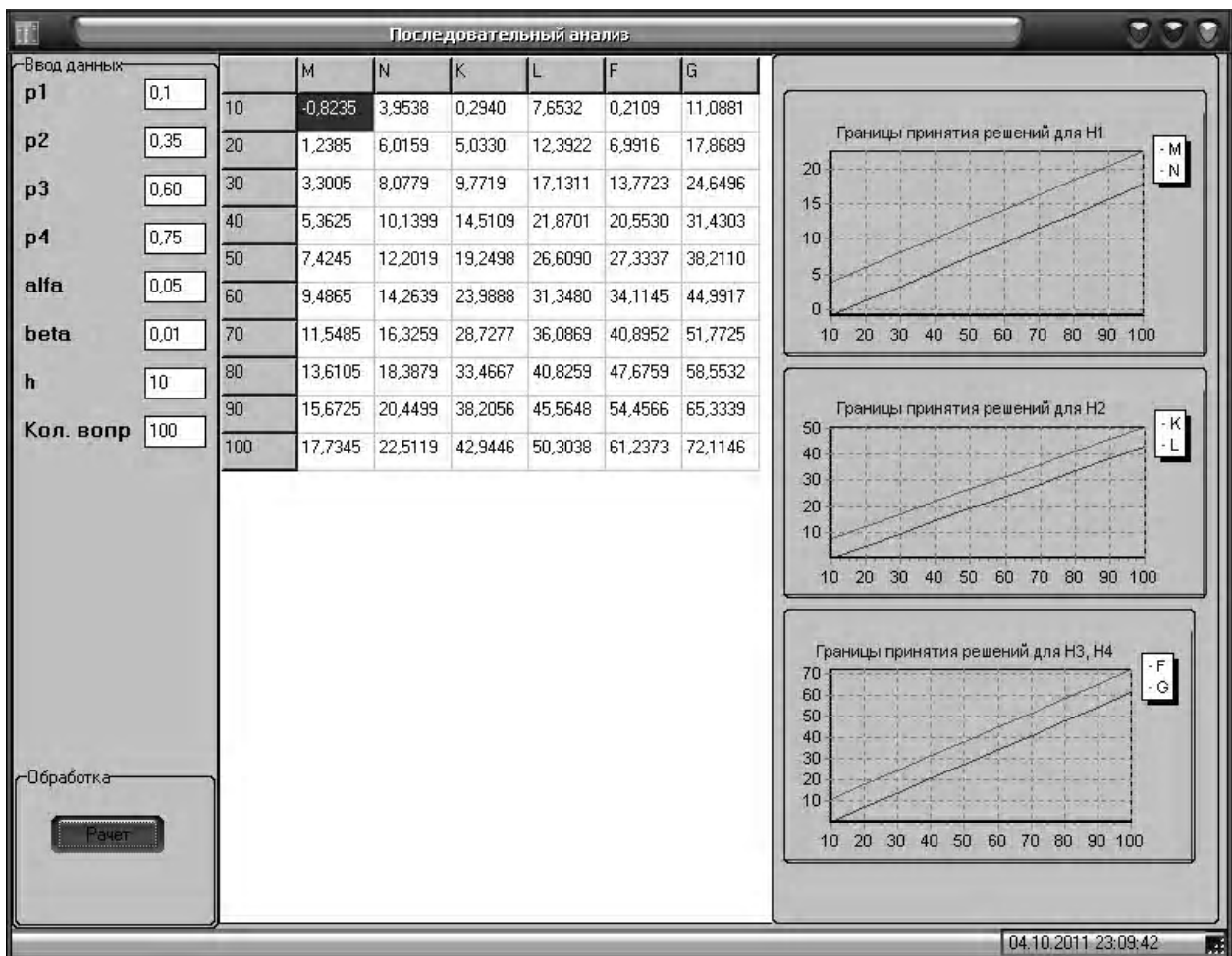


Рисунок 2 – Діалогове вікно формування границь прийняття рішень

На рис. 3. наведено приклад вирішення задачі тестування на основі процедури послідовного аналізу для одного об'єкта навчання.

Таким чином, якщо тестований з десяти завдань отриманих на першому етапі тестування виконав вірно підряд вісім, то згідно з правилами прийняття рішень йому буде видана наступна вибірка з десяти питань та продовжена перевірка гіпотези H_1 . Після перевірки гіпотези H_1 , буде прийняте рішення про перевірку гіпотези H_2 , відповідно до якої студент отримає наступні десять завдань, оскільки $K < d_h < L$. Після виконання 30 завдань приймається рішення щодо прийняття гіпотези H_2 при $d_h = 7$ та зарахування студента до класу II, оскільки $d_h \leq K$. Тестований отримає оцінку «BC» за системою ECTS або «4» за 5-ти бальною системою.

6 ОБГОВОРЕННЯ

В порівнянні з двохальтернативною класифікацією тестованих на дві групи: «атестовані» та «неатестовані», представленою в роботі [1], запропонована чотирьохальтернативна класифікація дозволяє здійснювати оцінку навчальних досягнень диференційовано, є адаптованою для автоматизованого контролю знань у вищих закладах освіти.

На рис. 4. зображено розподіл студентів за класами навченості за результатами перевірки знань з використанням ОТ послідовного аналізу та стандартного режимів тестування.

Оцінка якості класифікації виконується шляхом визначення наскільки тісно розташовані об'єкти в класах у порівнянні з розташуванням об'єктів у всій групі для стандартного та адаптивного режимів перевірки. Для IV класу

(«початковий» рівень навченості) відхилення результатів «адаптивного» режиму тестування від стандартного складає +4%, для III – –4%, для II та I –0%. Отже, класифікації рівнозначні, при цьому для стандартного режиму тестування середня кількість завдань, яку виконує кожний студент дорівнює «50», а для адаптивного – «30».

Запропонована обчислювальна технологія дає можливість системі тестування приймати рішення щодо класифікації поведінки студента, не перевіряючи всі n завдань, що скорочує число завдань, які необхідно перевірити, та забезпечує індивідуальну мінімізацію часу навчання.

Додатковими умовами забезпечення високої ефективності застосування запропонованої технології контролю знань є:

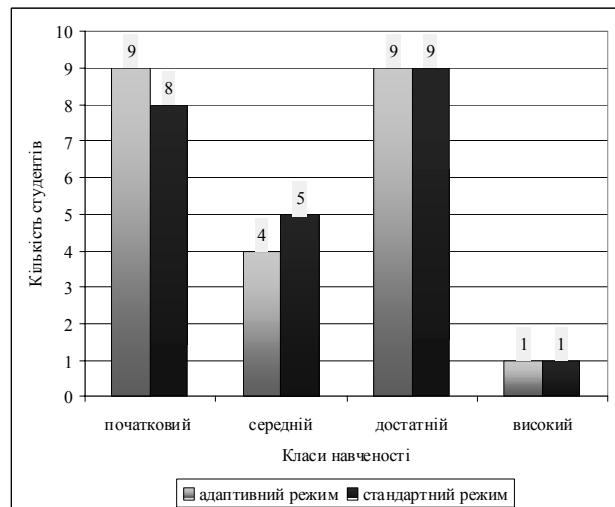
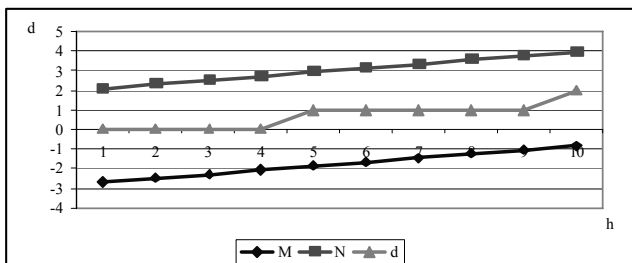
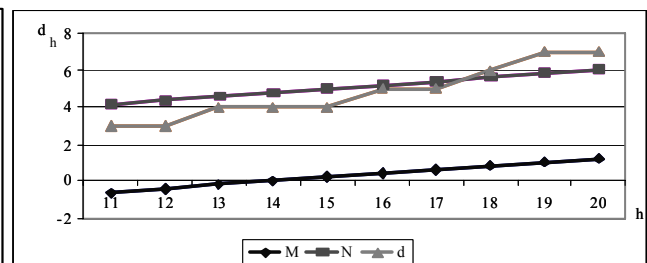


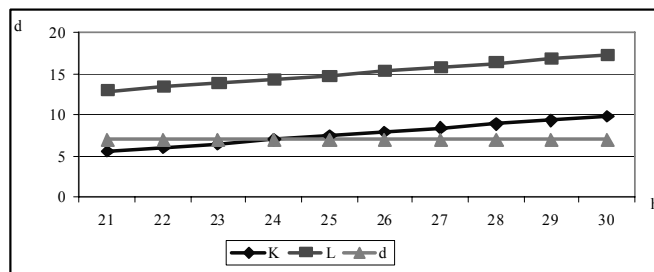
Рисунок 4 – Розподіл студентів за класами навченості на основі ОТ послідовного аналізу та стандартного режиму тестування



а



б



в

$$d_h = (000011111233444556777777777777)^T$$

$$h = (1234567891011121314151617181920212223242526272829,30)^T$$

Рисунок 3 – Графіки границь прийняття рішень:

а – відносно гіпотези H_1 при $h = 10$; б – відносно гіпотези H_1 при $h = 20$; в – відносно гіпотези H_2 при $h = 30$

1. Вимога підвищення якості тестових завдань, основними показниками якої є валідність і надійність тестів. Дані показники характеризують, відповідно, стабільність отриманих результатів тестування і їх здатність правильно відображати рівень підготовки.

2. Можливість у процесі тестування не тільки визначення рівня навченості слухачів, але й швидкого відновлення знань шляхом перегляду правильної відповіді і отримання (при необхідності) іншої додаткової дидактичної інформації з предмета.

ВИСНОВКИ

В роботі розглянуто задачу оцінки знань тестованих як задачу їх класифікації за рівнем знань, умінь та навичок, для рішення якої використано послідовну процедуру перевірки гіпотез і правила прийняття рішень на основі критерію Вальда. Розроблено обчислювальну процедуру класифікації тестованих на чотири класи, які відповідають рівням навченості: «початковий», «середній», «достатній», «високий».

Проаналізовані обчислювальні технології підтримки прийняття рішень під час контролю знань об'єктів навчання, які базуються на методах теорії статистичних рішень. Обґрунтовано вибір методу послідовного аналізу для оцінки рівня знань у комп'ютерній системі підтримки навчальної діяльності.

Досягнуто підвищення ефективності процесу підтримки прийняття рішень під час статистичного контролю знань за рахунок удосконалення методу послідовного аналізу відповідей шляхом класифікації учасників навчального процесу на чотири групи, які відповідають рівням навченості, що забезпечило індивідуальну мінімізацію часу перевірки знань та дозволило виконати диференційовану оцінку їхнього обсягу в задачах контролю за альтернативною ознакою.

Здійснено апробацію процедури підтримки прийняття рішень на основі методу послідовного аналізу з використанням СППР «ManageEdu», в результаті чого сформовані результуючі оцінки навчальних досягнень студентів III курсу ВНЗ МТУ «Миколаївська політехніка» економічного факультету з дисципліни «Фінансовий облік» та здійснено їх класифікацію за групами, які відповідають рівням навченості. За підсумками оцінки відхилення результатів адаптивного режиму тестування від стандартного підтверджено достатню якість запропонованої класифікації.

Перспективи подальших досліджень спрямовані на розробку методик встановлення стандарту оцінювання кваліфікації (ймовірності вірно виконаних завдань для кожного класу), методика оцінки рівня помилок цього критерію, застосування запропонованої обчислюваль-

ної технології для вирішення практичних задач контролю рівня знань в системах тестування.

ПОДЯКИ

Роботу виконано в рамках спільних наукових досліджень кафедри комп'ютерної інженерії й кафедри комп'ютерних наук та прикладної математики Миколаївського національного університету ім. В. О. Сухомлинського. Результати досліджень здійснювались у рамках науко-дослідної роботи за темою: «Моделі та методи інтелектуального аналізу даних в предметно-орієнтованій інформаційній системі» (номер реєстрації 0115U001249).

СПИСОК ЛІТЕРАТУРИ

1. Кузьма К. Т. Інформаційні технології контролю та оцінки знань / К. Т. Кузьма // Труды IX Міжнародної наук.-практ. конференції студентів та молодих учених «Політ». – К. : Вид-во Нац. авіац. ун-ту «НАУ-друку», 2009. – С. 221.
2. Васильев В. И. Основы культуры адаптивного тестирования / В. И. Васильев, Т. Н. Тягунова. – М. : Издательство ИКАР, 2003. – 584 с.
3. Galeev I. A Learning Model in MONAP / I. Galeev, V. Ivanov, M. Akhmadullin // Human-Computer Interaction. The 6th International Conference. EWHCI'96. Moscow, 1996. – P. 320–323.
4. Автоматизация контроля обученности в процессе подготовки специалистов для систем безопасности / [А. Н. Членов, И. Г. Дровникова, Т. А. Буцынская, П. А. Орлов] // Научный информационный сборник «Проблемы безопасности и чрезвычайных ситуаций». – М. : Винити, 2009. – № 4. – С. 107–116.
5. Переверзев В. Ю. Критериально-ориентированные педагогические тесты для итоговой аттестации студентов / В. Ю. Переверзев. – М. : НМЦ СПО Минобразования РФ, 1999. – 152 с.
6. Вальд А. Последовательный анализ / А. Вальд. – М. : Физматгиз, 1960. – 328 с.
7. Левин Б. Р. Теоретические основы статистической радиотехники. В трех книгах. Книга вторая / Б. Р. Левин. – М. : Сов. радио, 1975. – 392 с.
8. Люсин Д. В. Основы разработки и применения критериально-ориентированных педагогических тестов / Д. В. Люсин. – М. : Исследовательский центр, 1993. – 51 с.
9. Angoff W. H. Scales, norms, and equivalent scores / W. H. Angoff. – Princeton, NJ: ETS, 1984. – P. 153. – URL: <http://www.ets.org/Media/Research/pdf/Angoff.Scales.Norms.Equiv.Scores.pdf>.
10. Kaftandjiev F. Methods for Setting Cut Scores in Criterion-referenced Achievement Tests / F. Kaftandjiev. – Cito, Arnhem: EALTA, 2010. – P. 170. – URL: http://www.ealta.eu.org/documents/resources/FK_second_doctorate.pdf
11. А. с. 39905 України Комп'ютерна програма підтримки навчальної діяльності студентів «ManageEdu» / Кузьма К. Т., Байбуз О. Г. – № 40241; заявл. 30.06.2011; опуб. 01.09.2011.

Стаття надійшла до редакції 16.12.2017.

Після доробки 22.02.2017.

Кузьма К. Т.

Канд. техн. наук, старший преподаватель кафедры компьютерной инженерии, Николаевский национальный университет им. В. А. Сухомлинского, Николаев, Украина

ВЫЧИСЛИТЕЛЬНАЯ ТЕХНОЛОГИЯ ПРОВЕРКИ УРОВНЯ ЗНАНИЙ НА ОСНОВЕ МЕТОДА ПОСЛЕДОВАТЕЛЬНОГО АНАЛИЗА

Актуальность. Решена актуальная задача повышения эффективности процесса поддержки принятия решений при статистическом контроле знаний.

Цель работы – разработка вычислительной процедуры решения четырехальтернативной задачи классификации тестируемых по уровню обученности, что позволяет осуществлять контроль знаний дифференцированно, минимизирует объем задач, необходимый для выполнения.

Метод. Предложено вычислительную процедуру классификации тестируемых на четыре класса, которые соответствуют уровням обученности: «начальный», «средний», «достаточный», «высокий», основанную на использовании двухальтернативного критерия последовательного анализа в несколько этапов и обеспечивающую выполнение контроля знаний в процессе выполнения заданий, минимизируя таким образом время проверки знаний, что позволяет автоматизировать процесс проверки статистических гипотез в системах тестирования и обучения с целью дифференциальной оценки знаний участников учебного процесса. Для решения задачи оценки пригодности теста предложен метод, основанный на построении функции оперативной характеристики последовательного критерия, которая позволяет определить объем задач достаточный для достижения желаемого уровня качества теста за счет установления связи между ожидаемой вероятностью принятия гипотезы и случайным значением параметра вероятности появления в выборке из $1,2 \dots n$ вопросов приемочного числа неверно выполненных заданий.

Результаты. Разработано программное обеспечение, которое реализует предложенную вычислительную процедуру, использованное при проведении вычислительных экспериментов тестового контроля знаний.

Выводы. Проведенные эксперименты подтвердили работоспособность предложенной процедуры и программного обеспечения, которое ее реализует, а также позволяют рекомендовать их для применения на практике для решения задач автоматизированной проверки уровня знаний.

Ключевые слова: последовательный анализ, проверка уровня знаний, классификация тестируемых по уровню знаний, проверка гипотез, критериально-ориентированный тест.

Kuzma K. T.

Ph.D., Senior Lecturer of Department of Computer Engineering, V. O. Sukhomlynsky Mykolaiv National University, Mykolaiv, Ukraine

COMPUTER TECHNOLOGIES OF VERIFICATION OF KNOWLEDGE BASED ON THE METHOD OF SEQUENTIAL ANALYSIS

Context. The actual task of increasing the effectiveness of decision support in the statistical control of knowledge has been solved.

Objective is a development of a computational procedure for solving the problem of classification the educational process participants into four groups, according to their trainability level that allows to control knowledge differently, minimizing the amount of tasks required to accomplished.

Method. The computer technology for classification the educational process participants into four groups, according to their trainability level: “initial”, “medium”, “sufficiently”, “high”, has been proposed, based on the use of sequential hypothesis testing procedure and allows to perform the controlling of the knowledge while accomplishing tasks, thus minimizing the time for testing and provides the automation of the process of verification the statistical hypotheses in a testing and learning systems with the purpose of differential assessment of knowledge the educational process participants. To solve the problem of definition the assessment standard the method based on creation the function of operational characteristics of sequential criteria is used. The function allows to establish a link between the expected probability of the hypothesis and random probability of the fact of presence in the sample with $1.2 \dots n$ questions appropriate number of incorrect answers.

Results. The software implementing proposed computational procedure have been developed and used in computational experiments of knowledge testing.

Conclusions. The experiments confirmed the efficiency of the proposed procedure and software. The experiments also allow to recommend them for use in practice to solve the problems of automated assessment of knowledge.

Keywords: sequential analysis, the assessment of knowledge, classification the educational process participants according to their trainability level, testing of hypotheses, criterion-referenced test.

REFERENCES

1. Kuz'ma K.T. Informacijni tehnologii kontrolju ta ocinki znan', *Trudi IX Mizhnarodnoï nauk.-prakt. konferencii studentiv ta molodih uchenih «Polit»*. Kyiv, Vid-vo Nac. aviac. un-tu «NAU-druk», 2009. – S. 221.
2. Vasil'ev V. I., Tjagunova T. N. *Osnovy kul'tury adaptivnogo testirovanija*. Moscow: Izdatel'stvo IKAR, 2003, 584 p.
3. Galeev I., Ivanov V., Akhmadullin M. A Learning Model in MONAP, *Human-Computer Interaction. The 6th International Conference, EWHCI'96*. Moscow, 1996, pp. 320–323.
4. Chlenov A.N., Drovnikova I.G., Bucynskaja T.A., Orlov P.A. Avtomatizacija kontrolja obuchennosti v processe podgotovki specialistov dlja sistem bezopasnosti, *Nauchnyj informacionnyj sbornik «Problemy bezopasnosti i chrezvyčajnyh situacij»*. Moscow, Viniti, 2009. No. 4, pp. 107–116.
5. Pereverzev V. Ju. Kriterial'no-orientirovannye pedagogicheskie testy dlja itogovoj attestacii studentov. Moscow, NMC SPO Minobrazovanija RF, 1999, 152 p.
6. Val'd A. *Posledovatel'nyj analiz*. Moscow, Fizmatgiz, 1960, 328 p.
7. Levin B. R. *Teoreticheskie osnovy statisticheskoj radiotekhniki. V treh knigah. Kniga vtoraja*. Moscow, Sov. radio, 1975, 392 p.
8. Ljusin D.V. *Osnovy razrabotki i primenenija kriterial'no-orientirovannyh pedagogicheskikh testov*. Moscow, Issledovatel'skij centr, 1993, 51 p.
9. Angoff W. H. *Scales, norms, and equivalent scores*. Princeton, NJ, ETS, 1984, P. 153. URL: <http://www.ets.org/Media/Research/pdf/Angoff.Scales.Norms.Equiv.Scores.pdf>.
10. Kaftandjjeva F. *Methods for Setting Cut Scores in Criterion-referenced Achievement Tests*. Cito, Arnhem: EALTA, 2010, P. 170. URL: http://www.ealta.eu.org/documents/resources/FK_second_doctorate.pdf
11. Kuz'ma K. T., Bajbuz O. G. A. s. 39905 Ukraini Komp'juterna programa pidtrimki navchal'noï dijtal'nosti studentiv «ManageEdu» / № 40241; zajavl. 30.06.2011; opub. 01.09.2011.