

УДК 001.891.572

А.А. Кравченко, Н.А. Яремчук

Національний технічний університет України «КПІ», Київ

СПОСІБ РОЗМНОЖЕННЯ ДАНИХ ТЕСТУВАННЯ НА ОСНОВІ РАНДОМІЗАЦІЇ

В роботі запропоновано спосіб розмноження даних тестування з метою зменшення невизначеності вимірювання трудності тестових завдань за цими даними. Спосіб розмноження вибірки даних тестування моделює вплив екзогенних та ендогенних факторів з застосуванням біноміального розподілу та моделі Раша в області нестабільності латентного параметру. Наведено результати, що демонструють ефективність впровадження даного способу.

Ключові слова: *трудність тестових завдань, рівень навченості, невизначеність вимірювання труднощі, розмноження даних, рандомізація.*

Вступ

При проектуванні та експлуатації систем тестування, систем контролю знань учнів, систем дистанційного навчання вирішується задача вимірювання трудності тестів, що використовуються при тестуванні. Так як кількість правильних відповідей на тестові завдання (ТЗ) використовується при вимірюванні рівня знань учнів, тестові завдання називають контрольно-вимірювальними матеріалами. Тому результати вимірювання трудності супроводжуються невизначеністю вимірювання. Невизначеність вимірювання трудності ТЗ зменшується зі збільшенням кількості осіб, що задіяні при тестуванні. При проведенні багатомасштабних тестувань (наприклад, при проведенні єдиного державного екзамену) питання зменшення невизначеності вимірювання трудності ТЗ вирішується просто. Але при проведенні вузькопрофільних тестувань або при проведенні тестування в малочисельних групах зі спеціальних (особливо технічних) дисциплін зменшення невизначеності вимірювання трудності ТЗ є проблемою. Її можна вирішити або залученням даних додаткових тестувань при проведенні моніторингу систем тестування, або з використанням методів штучного розмноження вибірки даних малого об'єму.

Результати тестування на основі яких судять про рівень навченості опитуваних θ_i та трудність тестових завдань δ_j , отримують у вигляді двомірної матриці відповідей A_{ij} обмеженого об'єму (матриця з певною кількістю стовпців k та обмеженою кількістю рядків n , де k – кількість запитань тесту, n – кількість опитуваних), яка дуже часто являється єдиною. Як показано в роботах [1], [2] невизначеність рівня навченості та трудності зменшується зі збільшенням n і k .

На даний час вже відомі способи, що дозволяють провести статистичний аналіз за даними єдиної вибірки, які базуються на розмноженні даних перві-

сної вибірки: «метод складного ножа» та «бутстреп-метод». Застосування цих методів для зменшення невизначеності вимірювання трудності ТЗ δ_j були досліджені в роботах Орлова А.І. [3] та Єлісеєва І.М. [4]. В роботі [3] показано, що застосування бутстреп-методу для розмноження матриці результатів тестування не забезпечило потрібного зменшення невизначеності вимірювання трудності ТЗ з причини генерації майже подібних псевдовибірок. А в роботі [4] продемонстровано, що використання бутстреп-методу та методу складного ножа для підвищення точності вимірювання трудності ТЗ за єдиною матрицею відповідей не має сенсу через випадковість формування псевдовибірок.

В зв'язку з цим, авторами даної роботи були проведені додаткові дослідження з метою розробки способу розмноження матриці тестування, результати яких наведено нижче.

Постановка задачі. Метою роботи є використання способу розмноження даних тестування (у вигляді матриці тестування) для зменшення невизначеності вимірювання трудності. Спосіб розмноження рядків матриці тестування заснований на рандомізації матриці у відповідності з біноміальним розподілом, параметри якого визначаються за моделлю Раша.

1. Спосіб розмноження даних тестування

В даній роботі пропонується та розглядається спосіб розмноження рядків матриці відповідей на тестові завдання на основі рандомізації у відповідності до моделі m_b випадкових незалежних експериментів, що мають розподіл Бернуллі з параметром p , тобто біноміальний розподіл (m_b, p) , де p – ймовірність моделі Раша:

$$P_{ij} = \frac{\exp(\theta_i - \delta_j)}{1 + \exp(\theta_i - \delta_j)}, \quad (1)$$

Причиною нестабільності реакції опитуваних в області значень $\tilde{p}_{ij}^{(1)} \approx 0,5$ являється вплив екзогенних (зовнішніх) та ендогенних (внутрішніх) факторів. До числа екзогенних факторів відносять стресову тактику педагогічного впливу, невідповідність методик і технологій тестування функціональним можливостям опитуваних. До числа ендогенних факторів відносять стан здоров'я, ступінь сформованості вищих психічних функцій. Проте, у ряді випадків впливають так звані змішані фактори, які поєднують вплив як зовнішніх, так і внутрішніх факторів. Тому, в залежності від стану опитуваного, він може відповісти, а може не відповісти на запитання, коли $\theta_i \approx \delta_j$. Таку область можна назвати областю нестабільності латентного параметру (дана область виділена в табл. 5).

В якості величини, що відповідає нестабільності латентного параметру, обрано саме ймовірність моделі Раша \tilde{p}_{ij} , яку будемо вважати емпіричною ймовірністю. Мірою відхилення емпіричної ймовірності \tilde{p}_{ij} від теоретичної, яку отримують за значеннями θ_i і δ_j , обрано різницю $\tilde{p}_{ij} - p_{ij}$, що асимптотично нормальна з нульовим середнім і дисперсією [5]:

$$\sigma^2[\tilde{p}_{ij}] = \frac{\tilde{p}_{ij}(1-\tilde{p}_{ij})}{n} \quad (2)$$

Емпірична ймовірність використовується для рандомізації даних тестування з використанням біноміального розподілу, за яким визначається ймовірність 0 або 1 в матриці тестування. Одиниця відповідає правильній відповіді, а 0 – неправильній.

В основу розробки алгоритму розмноження даних тестування на основі рандомізації покладено

факт можливої зміни 1 на 0 та 0 на 1 в матриці відповідей, коли ймовірність в моделі Раша p_{ij} знаходиться біля значення 0,5, що відповідає $\theta_i \approx \delta_j$.

Для аналізу маємо дані тестування групи опитуваних ($n=12, k=10$), які представлені в табл. 1 та підлягають подальшому розмноженню.

Перехід від ймовірності до дискретних значень c_j^M можна отримати за модою біноміального розподілу, виходячи з максимальної правдоподібності за розміщенням моди біноміального розподілу, що є цілим числом

$$c_j^M = M_0,$$

де

$$\left[(\tilde{p}_{ij}(m_b + 1) - 1) \leq M_0 \leq (\tilde{p}_{ij}(m_b + 1)) \right], \quad (3)$$

якщо $m_b = 1$, то $\left[(2\tilde{p}_{ij} - 1) \leq M_0 \leq (2\tilde{p}_{ij}) \right]$.

2. Обчислення невизначеності вимірювання труднощі після рандомізації матриці тестування

Зв'язок між середніми квадратичними відхиленнями труднощі $\sigma[\delta_j]$ та кількості правильних відповідей $\sigma[c_j]$ можна виразити наступним співвідношенням:

$$\sigma[\delta_j] = \frac{\partial \delta_j}{\partial c_j} \sigma[c_j], \quad \text{де } \delta_j = \frac{n - c_j}{c_j}.$$

$\sigma[c_j]$ визначається за СКВ біноміального розподілу та дорівнює

Таблиця 1

Результати тестування групи опитуваних ($n=12, k=10$)

n	Номер завдання										b_i	θ_i	$\hat{\theta}_i$
	1	2	3	4	5	6	7	8	9	10			
1	0	0	0	0	0	0	0	0	0	1	1	-2,197	-2,082
2	0	1	0	0	0	0	0	0	0	1	2	-1,386	-1,271
3	0	0	0	1	0	0	0	1	1	0	3	-0,847	-0,732
4	0	0	0	1	0	0	0	1	1	1	4	-0,405	-0,290
5	0	0	0	0	1	1	0	1	1	1	5	0,000	0,116
6	0	0	0	1	1	0	1	0	1	1	5	0,000	0,116
7	0	0	0	0	0	1	1	1	1	1	5	0,000	0,116
8	1	0	0	0	1	0	1	1	1	1	6	0,405	0,521
9	0	0	0	0	1	1	1	1	1	1	6	0,405	0,521
10	0	0	1	0	0	1	1	1	1	1	6	0,405	0,521
11	0	1	1	0	0	1	1	1	1	1	7	0,847	0,963
12	0	0	1	1	1	1	1	1	1	1	8	1,386	1,502
c_j	1	2	3	4	5	6	7	9	10	11		-0,116	0,000
δ_j	2,40	1,61	1,10	0,69	0,34	0,00	-0,34	-1,10	-1,61	-2,34		0,07	
$\hat{\delta}_j$	2,33	1,54	1,03	0,62	0,27	-0,07	-0,41	-1,17	-1,68	-2,47		0,00	

$$\sigma[c_j] = \sqrt{n \cdot p(1-p)} = \sqrt{\frac{1}{n} \cdot c_j(n-c_j)}$$

В результаті, невизначеність вимірювання трудності тестових завдань знаходять як:

$$\begin{aligned} \sigma[\delta_j] &= \frac{n}{c_j(n-c_j)} \cdot \sqrt{\frac{c_j(n-c_j)}{n}} = \\ &= \sqrt{\frac{n}{c_j(n-c_j)}} = \sqrt{\frac{1}{p_j q_j n}} \end{aligned} \quad (4)$$

Але співвідношення (4) може бути застосовано тільки в середній частині діапазону вимірювання трудності. Для крайніх точок діапазону, де нелінійність суттєва [6]:

$$\sigma[\delta_j] = \sqrt{D[\delta_j]}$$

$$D[\delta_j] = \left(\frac{\partial \delta_j}{\partial c_j}\right)^2 \sigma^2[\delta_j] + \frac{1}{4} \left(\frac{\partial^2 \delta_j}{\partial^2 c_j}\right)^2 \times$$

$$\times \left\{ \mu_4[c_j] - \sigma^4[\delta_j] \right\} + \left(\frac{\partial \delta_j}{\partial c_j}\right) \cdot \left(\frac{\partial^2 \delta_j}{\partial^2 c_j}\right) \cdot \mu_3[c_j],$$

$\mu_4[c_j]$ – четвертий момент біноміального розподілу:

$$\mu_4[c_j] = 3n^2 p^2 (1-p)^2 + n \cdot p(1-p)(1-6p(1-p));$$

$\mu_3[c_j]$ – третій момент біноміального розподілу:

$$\mu_3[c_j] = n \cdot p(1-p)(1-2p).$$

Тому значення стандартної невизначеності для крайніх точок діапазону вимірювання були отримані за співвідношенням (5).

Дані розрахунку невизначеності вимірювання трудності тестових завдань (табл. 1), приведені в табл. 2, де δ_j та $\sigma[\hat{\delta}_j]$ подані в логітах.

З таблиці видно, що стандартна невизначеність вимірювання трудності сумірна з трудностю, а в деяких випадках – перевищує її. Отже, очевидна необхідність підвищення точності вимірювання трудності, тобто в розмноженні даних матриці тестування.

Таблиця 2

Дані розрахунку невизначеності вимірювання трудності тестових завдань (табл. 1)

j	1	2	3	4	5
c_j	1	2	3	4	5
δ_j	2,33	1,54	1,03	0,62	0,27
$\sigma[\hat{\delta}_j]$	0,89	0,72	0,42	0,37	0,34
j	6	7	8	9	10
c_j	6	7	9	10	11
δ_j	-0,07	-0,41	-1,17	-1,68	-2,47
$\sigma[\hat{\delta}_j]$	0,33	0,34	0,42	0,72	0,89

3. Алгоритм розмноження даних тестування

Структурна схема, що ілюструє алгоритм розмноження даних тестування, наведена на рис. 1.

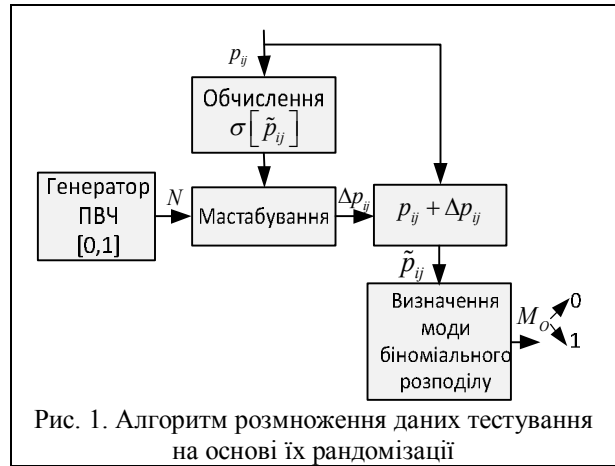


Рис. 1. Алгоритм розмноження даних тестування на основі їх рандомізації

На виході генератора псевдовипадкових чисел (ПВЧ) отримують послідовність чисел N з нормальним розподілом, математичне сподівання якого дорівнює нулю, а середнє квадратичне відхилення дорівнює 1. В залежності від теоретичної ймовірності (формула 1), яку отримують за даними матриці тестування θ_i і δ_j , проводиться масштабування послідовності ПВЧ за значенням $\sigma[\tilde{p}_{ij}]$ (формула 2). Після масштабування отримуємо послідовність $\Delta p_{ij} = \sigma[\tilde{p}_{ij}] \cdot N$, яка додається до значень теоретичної ймовірності $\tilde{p}_{ij} = p_{ij} + \Delta p_{ij}$ (табл. 5). З використанням емпіричної ймовірності \tilde{p}_{ij} за формулою (3) визначається мода біноміального розподілу.

В табл. 3 продемонстрована розмножена матриця тестування. На матриці позначені зміни моди з 0 на 1 та навпаки при двох циклах рандомізації.

$\sigma_{\text{ран}}[\hat{\delta}_j]$ – невизначеність вимірювання трудності ТЗ після рандомізації (табл. 4).

Можна виділити наступні етапи розмноження даних тестування на основі рандомізації.

1. Отриману матрицю відповідей опитуваних на тестові завдання A_{ij} , де $i = 1 \dots n, j = 1 \dots k$ упорядковуємо за зростанням кількості відповідей по рядках та стовпцях (табл. 1).

2. Знаходимо значення трудності ТЗ δ_j та рівня навченості опитуваних θ_i за алгоритмом, описаним

$$\text{в [2]: } \delta_j = \ln \frac{n - c_j}{c_j}, \theta_i = \ln \frac{b_i}{k - b_i}.$$

3. Перерахуємо отримані значення δ_j та θ_i з урахуванням умовного нуля. Нові значення трудності ТЗ $\hat{\delta}_j$ та рівня навченості опитуваних $\hat{\theta}_i$ заносимо до табл. 1.

Таблиця 3

Розмножена матриця тестування

<i>i</i>	<i>j</i>	1	2	3	4	5	6	7	8	9	10
1	A_{1j}	0	0	0	0	0	0	0	0	0	1
	A_{1j}^1	0	0	0	0	0	0	0	0	0	1
	A_{1j}^2	0	0	0	0	0	0	0	0	1	0
2		0	1	0	0	0	0	0	0	0	1
		0	0	0	0	0	0	0	0	1	1
		0	0	0	0	0	0	0	1	1	1
3		0	0	0	1	0	0	0	1	1	0
		0	0	0	0	0	0	0	1	1	1
		0	0	0	0	0	0	1	1	1	1
4		0	0	0	1	0	0	0	1	1	1
		0	0	0	0	0	0	0	1	1	1
		0	0	0	0	0	0	1	1	1	1
5		0	0	0	0	1	1	0	1	1	1
		0	0	0	0	0	1	1	1	1	1
		0	0	0	0	1	1	1	1	1	1
6	A_{6j}	0	0	0	1	1	0	1	0	1	1
	A_{6j}^1	0	0	0	1	0	0	1	1	1	1
	A_{6j}^2	0	0	0	0	0	1	1	1	1	1
7	A_{7j}	0	0	0	0	0	1	1	1	1	1
	A_{7j}^1	0	0	0	0	0	1	0	1	1	1
	A_{7j}^2	0	0	0	0	1	0	0	1	1	1
8		1	0	0	0	1	0	1	1	1	1
		0	0	1	0	1	1	1	1	1	1
		0	0	0	1	0	1	1	1	1	1
9		0	0	0	0	1	1	1	1	1	1
		0	0	0	1	1	1	1	1	1	1
		0	0	1	0	1	1	1	1	1	1
10		0	0	1	0	0	1	1	1	1	1
		0	0	0	0	1	1	1	1	1	1
		0	0	0	1	1	1	1	1	1	1
11		0	1	1	0	0	1	1	1	1	1
		0	0	1	1	1	1	1	1	1	1
		0	0	1	1	1	1	1	1	1	1
12	A_{12j}	0	0	1	1	1	1	1	1	1	1
	A_{12j}^1	0	0	1	1	1	1	1	1	1	1
	A_{12j}^2	0	1	1	1	1	1	1	1	1	1

Таблиця 4

Невизначеність вимірювання трудності ТЗ після рандомізації

<i>j</i>	1	2	3	4	5
c_j	1	3	9	12	16
$\sigma_{\text{ран}}[\hat{\delta}_j]$	0,76	0,32	0,15	0,12	0,11

<i>j</i>	6	7	8	9	10
c_j	20	23	30	33	34
$\sigma_{\text{ран}}[\hat{\delta}_j]$	0,11	0,12	0,19	0,32	0,44

4. Проводимо калібрування матриці отриманих результатів з метою наближення даних до моделі Раша ітераційним методом, описаним в [6] та отримуємо нові значення трудності ТЗ $\hat{\delta}_j^{(1)}$ та рівня навченості опитуваних $\hat{\theta}_j^{(1)}$.

5. Робимо розрахунок ймовірностей по моделі Раша (формула 1) для $\hat{\delta}_j^{(1)}$ та $\hat{\theta}_j^{(1)}$. Отримані результати розрахунків заносимо до табл. 5.

6. Визначаємо моду відповідно до розробленого алгоритму розмноження даних тестування (рис.1).

7. Розраховуємо невизначеність вимірювання трудності тестових завдань (формула 4) після двох тактів рандомізації $\sigma_{\text{ран}}[\hat{\delta}_j]$.

Дані проведених розрахунків для двох циклів рандомізації наведено в таблиці 3.

З отриманих результатів видно, що точність визначення трудності тестових завдань підвищилась, але в середній частині діапазону залишилась низькою. Це пояснюється тим, що біля умовного нуля розрізнення значень трудності потребує більшого об'єму початкової вибірки.

Висновки

Відомо, що за результатами тестування в малих вибірках невизначеність вимірювання трудності тестових завдань практично сумірна зі значеннями вимірюваної трудності. Зменшення невизначеності можливе або з залученням даних додаткових тестувань при проведенні моніторингу систем тестування, або з використанням штучного розмноження вибірки даних малого об'єму.

Таблиця 5

Ймовірності $\hat{p}_{ij}^{(1)}$ вірного виконання *i*-м учасником *j*-го завдання

<i>i</i> <i>j</i>	1	2	3	4	5	6	7	8	9	10
1	0,01	0,01	0,02	0,03	0,05	0,07	0,10	0,22	0,33	0,55
2	0,01	0,03	0,05	0,08	0,11	0,16	0,22	0,41	0,56	0,75

i j	1	2	3	4	5	6	7	8	9	10
3	0,03	0,06	0,10	0,14	0,20	0,27	0,36	0,58	0,72	0,86
4	0,05	0,10	0,16	0,23	0,32	0,41	0,50	0,72	0,82	0,95
5	0,08	0,17	0,26	0,35	0,45	0,55	0,64	0,82	0,89	0,95
6	0,08	0,17	0,26	0,35	0,45	0,55	0,64	0,82	0,89	0,95
7	0,08	0,17	0,26	0,35	0,45	0,55	0,64	0,82	0,89	0,95
8	0,13	0,26	0,38	0,49	0,59	0,68	0,76	0,89	0,94	0,97
9	0,13	0,26	0,38	0,49	0,59	0,68	0,76	0,89	0,94	0,97
10	0,13	0,26	0,38	0,49	0,59	0,68	0,76	0,89	0,94	0,97
11	0,22	0,39	0,53	0,63	0,72	0,79	0,85	0,94	0,96	0,98
12	0,36	0,56	0,69	0,77	0,84	0,88	0,92	0,97	0,98	0,99

Як показав огляд літератури, застосування способів розмноження даних з використанням «складеного ножа» або «бутстреп-методу» не дали можливості зменшити невизначеність вимірювання трудності.

В роботі запропоновано спосіб розмноження даних тестування, заснований на рандомізації матриці даних тестування у відповідності з біноміальним розподілом, параметри якого визначаються за моделлю Раша.

За область рандомізації обрано область, що на практиці відповідає нестабільності латентного параметру, а при моделюванні області, де ймовірність за моделлю Раша знаходиться біля значення 0,5 і відповідає факту зміни розташування моди біноміального розподілу та відповідно зміні рішення в матриці тестування з 0 на 1, і навпаки. Після опрацювання даних, визначених при розмноженні вибірки, отримуємо уточнене значення трудності тестових завдань зі зменшеною невизначеністю.

Список літератури

1. Нейман Ю.В. Введение в теорию моделирования и параметризации педагогических тестов / Ю.В. Нейман, В.А. Хлебников. – М.: Прометей, 2000. – 168 с. – ISBN 5-7042-1068-6.

2. Яремчук Н.А. Зменшення невизначеності вимірювання трудності тестових завдань / Н.А. Яремчук, А.А. Кравченко // Збірник IX Міжнародної науково-технічної конференції «Гіротехнології, навігація, керування рухом та конструювання авіаційно-космічної техніки. – К.: НТУУ «КПІ», 2012. – 86 с.

3. Орлов А.И. Эконометрика. Учебник / А.И. Орлов. – М.: Издательство "Экзамен", 2002. – 576 с.

4. Елисеев И.Н. Математические модели и комплексы программ для автоматизированной оценки результатов обучения с использованием латентных переменных: дис...к.т.н.: 05.13.18 / Елисеев Иван Николаевич. – Новочеркасск, 2013. – 371 с.

5. Смирнов Н.В. Курс теории вероятностей и математической статистики / Н.В. Смирнов, Дунин-Барковский Н.В. – Изд. 3-е. – М.: Наука, 1969. – 511 с.

6. Сборник задач по теории вероятностей, математической статистике и теории случайных функций / под ред. А.А. Свейникова. – М.: Наука, 1970. – 656 с.

7. Кравченко А.А. Порівняння методів калібрування результатів тестування / А.А. Кравченко, Н.А. Яремчук // Збірник X Всеукраїнської науково-методичної конференції «Вища освіта: Проблеми і шляхи забезпечення якості». – К.: НТУУ «КПІ». – 2012. – 86 с.

Поступила в редколлегию 28.03.2014

Рецензент: д-р техн. наук проф. И.В. Руженцев, Харьковский национальный университет радиоэлектроники, Харьков.

СПОСОБ РАЗМНОЖЕНИЯ ДАННЫХ ТЕСТИРОВАНИЯ НА ОСНОВЕ ИХ РАНДОМИЗАЦИИ

А.А. Кравченко, Н.А. Яремчук

В работе предложен способ размножения данных тестирования с целью уменьшения неопределенности измерения трудности тестовых заданий по этим данным. Способ размножения выборки данных тестирования моделирует влияние экзогенных и эндогенных факторов с применением биномиального распределения и модели Раша в области нестабильности латентного параметра. Приведены результаты, демонстрирующие эффективность внедрения данного способа.

Ключевые слова: трудность тестовых заданий, уровень обученности, неопределенность измерения трудности, размножение данных, рандомизация.

REPRODUCTION METHOD OF TEST DATA BASED ON THEIR RANDOMIZATION

A.A. Kravchenko, N.A. Jaremchuk

In this paper, we proposed a method for reproduction test data in order to reduce the vagueness of measurement the difficulty of test tasks on these data. Method of reproduction data sample of testing simulates the effect of exogenous and endogenous factors using the binomial distribution and the Rash model in the area of instability the latent parameter. We have also presented the results which demonstrate the effectiveness of the introduction of this method.

Keywords: the difficulty of test tasks, the level of knowledge, the vagueness of measurement the difficulty, reproduction of test data, randomization.