© О.М. Алексєєв, О.М. Король, 2010
ISSN 2076-8184. Інформаційні технології і засоби навчання. 2010. №6 (20).
Режим доступу до журналу: http://www.ime.edu-ua.net/em.html

УДК 378.146

Алексєєв Олександр Миколайович, кандидат технічних наук, доцент Сумського державного університету, м. Суми

Король Олена Миколаївна, завідувач кабінетом інформатики Сумського державного педагогічного університету ім. А. С.Макаренка, м. Суми

ФОРМУВАННЯ ТЕСТІВ В ІМІТАЦІЙНІЙ МОДЕЛІ КОНТРОЛЮ УСПІШНОСТІ НАВЧАННЯ

Анотація

У статті розвинені положення імітаційної моделі тестового контролю успішності навчання в частині вдосконалення процедури формування тесту. Пропонується для рішення про включення тестового завдання враховувати складність завдання і використовувати ітераційні обчислення з тим, щоб за допомогою алгоритмів оптимізації, які керують процесом випадкового перебору створити тест, у якому сума індексів складності завдань відповідає критеріям сукупної значущості або складності. Дається розгорнутий виклад математичного апарату, який використовується для обґрунтування схвалюваних рішень щодо формування тесту. Наводиться приклад, що демонструє основні етапи ітераційних обчислень і механізми досягнення оптимальності за критеріями сукупної значущості і складності.

Ключові слова: тестування, тестовий контроль успішності навчання, імітаційна модель, тестове завдання, тест, формування тесту, критерії оптимальності тесту, сукупна значущість і складність тесту.

Постановка проблеми. За неухильного застосування тестових форм контролю є наполеглива необхідність впровадити в практику тестування модель, яка, використовуючи усі переваги тестового контролю, водночас максимально спиралася б на вітчизняні традиції активної участі викладача в діагностуванні успішності навчання студентів. У роботі [1] описана імітаційна модель тестування, у якій поєднуються переваги комп'ютеризованого тестування з математичним обґрунтуванням процедур ухвалення рішень екзаменатором. Модель базується на сучасних програмних засобах, передбачаючи можливість регулювання кількості

запитань, що ставляться, залежно від результативності відповідей студентів, наявність розширеного набору прототипів типових завдань, з яких можна створити тест, максимально відповідний змісту контрольованого матеріалу, встановлення об'єктивних критеріїв переведення набраної суми балів в показники успішності навчання та ін.

Для забезпечення зіставлення результатів тестового контролю, що проводиться під час діагностування успішності засвоєння навчального матеріалу з різних тем і дисциплін, імітаційною моделлю пропонується формувати тести відповідно до критеріїв сукупної значущості (в основній сесії контролю) і сукупної складності (у додаткових сесіях контролю) [1, 2].

Встановлення критеріїв сукупної значущості і складності ставить перед необхідністю спроектувати тест так, щоб сума індексів складності тестових завдань, які включаються в нього, була б максимально близька до заданих критеріями однозначних числових значень. Зробити це простим підбором часто стає неможливим, враховуючи, що кожне з тестових завдань може мати будь-яке зі значень індексу складності – від мінімальних значень, прийнятих для найпростіших завдань, до максимальних значень, які призначені для найбільш складних із них, а завдання повинні включатися в тест випадково. Тобто, виходячи з опису ситуації, числові значення індексів складності завдань, що включаються в тест, вибираються випадково, а під час їх підсумовування вимагається отримати постійну величину.

Аналіз останніх досліджень і публікацій. Велику роль у становленні тестового контролю успішності навчання як найточнішого засобу педагогічних вимірів відіграли праці таких учених як В. С. Аванесов, І. Є. Булах, Н. О. Гулюкина, Л. М. Давидова, Н. Ф. Єфремова, В. М. Кадневський, В. С. Ким, Є. Г. Кириленко, К. Г. Кречетников, А. М. Майоров, Є. А. Михайличев, О. В. Михєєв, І. А. Морев, В. І. Нардюжев, Ю. М. Нейман, В. Ю. Переверзев, Д. І. Попов, Б. У. Родіонов, І. Д. Рудинський, М. Б. Челишкова, Т. Ш. Шихнабієва, P. W. Airasian, F. B. Baker, B.S. Bloom, N. E. Gronlund, J. T. Hastings, R. L. Linn, G.F. Madaus, T. D. TenBrink та ін. Завдяки виявленню в їхніх дослідженнях нових дидактичних можливостей і технологічних переваг тестовий контроль останніми роками став широко застосовуватися для різних форм і напрямів навчання. Проте застосування тестування для контролю навчальних досягнень не завжди дає найкращі результати, оскільки в

проведених дослідженнях частково залишилася за рамками розгляду і не врахована повністю обов'язкова вимога розширити базу типових форм тестових завдань для того, щоб мати додаткові інструменти для контролю професійних навичок і вмінь. Крім того, до теперішнього часу не повною мірою забезпечується об'єктивність результатів контролю, оскільки за існуючих підходів залишається можливим значно впливати на його результати, наприклад, довільно змінивши критерії оцінювання. Не підвищує об'єктивності результатів контролю і те, що під час складання тестів кожен з проектувальників може по-своєму визначити вимоги до складності тестових завдань і тоді сукупна складність таких тестів буде непорівнянною. Очевидно, що і результати тестування в цих випадках будуть різні – за одного і того ж рівня підготовки студенти, що виконують завдання меншої складності, отримають вищі оцінки, ніж ті з них, хто виконував усі завдання підвищеної складності.

Проблема достовірності результатів тестування посилюється ще й тим, що більшість сучасних методик тестового контролю розробляються на базі теоретичних положень, постульованих зарубіжними авторами, які багато в чому не враховують традиції і переваги вітчизняної системи освіти, пов'язані з домінуючою роллю, яку в організації навчального процесу в першу чергу прийнято відводити викладачеві.

Постановка завдання. Мета цієї статті – розробити й описати стосовно імітаційної моделі тестування процедуру формування тестів, у якій при виборі випадковим способом тестових завдань непередбачуваної складності можна забезпечити формування тестів з рівнем сукупної складності, відповідним однозначно заданим детермінованим критерійним значенням.

Виклад основного матеріалу. Для розробки процедури формування тестів у рамках імітаційної моделі тестування використаний математичний апарат генетичних алгоритмів [3], відповідно до якого процес формування тесту пропонується здійснювати за допомогою такої послідовності наближення, яка сходиться до критерійного значення значущості (складності) контрольованого навчального матеріалу і будується рекурентно. Перехід від ітерації X_0 до X_1 і потім послідовно від X_i до X_{i+1} виробляється шляхом поетапного застосування правил включення завдань в тест, формування тестів і порівняння сумарної складності новостворюваних тестів з верхньою і нижньою межами довіри характеристик значущості або складності навчального матеріалу.

Створення початкового набору допустимих варіацій побудови тестів з тестових завдань $Nq_{r,i}$ на початку проектування припускає формування великої кількості векторів x_r^o

$$x_r^o = \{Nq_{r,1}, Nq_{r,2}, ...Nq_{r,i}, ...Nq_{r,n}\}, \quad (1)$$

кожен з яких створюється випадково і відображує один із можливих варіантів сформованого тесту.

Тестові завдання відбираються випадково, але з урахуванням того, що сумарне число завдань, які входять в будь-який з варіантів тесту, не повинне перевищувати заздалегідь встановленої величини $Nq_{\sum}$:

$$Nq_{r,i} \in 1, Nq_{\sum} - \sum_{i=1}^{i=n-1} Nq_{r,i}, \quad (2)$$

$$Nq_{r,i} = Z_{r,i},\ 1 \le Z_i \le Nq_{max_i},\ Nq_{r,i}, Z_{r,i} - \text{ціле } Z = rnd(X),$$

де $Z_{r,i}$ – випадкова величина, що задовольняє задані умови;

$Nq_{r,i}$ – кількість тестових завдань в *i*-ій темі;

Nq_{max_i} – максимальна кількість тестових завдань в *i*-ій темі.

При відборі сформованих варіантів тестів для подальшого аналізу використовуються переваги, засновані на тому, що тест, якісніший за критерієм значущості (складності) контрольованого матеріалу, має найбільшу вірогідність для включення в новий набір. Під час формування тестів це означає, що чим ближче сумарна складність тестових завдань, що становлять аналізований варіант, до критерію оптимальності цільової функції

$$F_i = [Q] - \sum_{i=1}^{n} \sum_{j=1}^{m} Q_{i,j}, \quad (3)$$

де *[Q]* – критерійне значення допустимої значущості (складності) контрольованого матеріалу;

$Q_{i,j}$ – індекс складності тестового завдання (визначається відповідно до рекомендацій роботи [4]),

тим тест вважається якіснішим і вище вірогідність його відбору для створення нових варіантів побудови тестів.

Водночас приймається, що вірогідність включення тесту в пару для утворення нових варіантів (т. з. "батьківську" пару") має бути пропорційною якості вихідних тестів, що відбираються, і один і той же тест може використовуватися для утворення декількох "батьківських" пар. Виходячи з принципу трансформації найбільш пристосованих особин [3], умова записується як:

$$K_i^p = \frac{\sum_{i=1}^{N} F_i - F_i}{\sum_{i=1}^{N} F_i},\ G_i^L < Z \le G_i^R, \qquad (4)$$

де K_i^p – коефіцієнт, що враховує вірогідність участі тесту у формуванні нових варіантів;

G_i^L, G_i^R – межі інтервалів, які відповідають вірогідності включення тесту в пару

$$G_i^L = \sum_{i=1}^{h} K_i^p,\ G_i^R = \sum_{i=1}^{h} K_{i+1}^p,\ h=1,2\ldots n-1. \qquad (5)$$

Суть процедури формування нових варіантів тесту полягає в тому, що її застосування повинно забезпечити обмін тестовими завданнями між двома тестами, які утворюють "батьківську" пару. Для формування нових варіантів в процедурі використовується двоточковий оператор, відповідно до якого розігрують дві випадкові цілі величини, значення яких не можуть співпадати і повинні знаходитися в інтервалі від 2 до n-1. Розіграні випадкові величини ранжуються і приймається менша – як перша точка розриву t_1, а більша – як друга точка розриву t_2.

Створення нових варіантів виконується так, щоб в плані розміщення завдань в тесті обміняти завдання на ділянках від 1-го завдання до завдання з координатою t_1 і від завдання з координатою t_2 до кінцевого завдання, а завдання на ділянці від t_1 до t_2 залишалися б незмінними:

$$\begin{cases} x_{i,j}^{c1} = x_{i,j}^{1},\ x_{i,j}^{c2} = x_{i,j}^{2},\ \text{при}\ j = t_1 + 1 \ldots t_2 + 1 \\ x_{i,j}^{c1} = x_{i,j}^{2},\ x_{i,j}^{c2} = x_{i,j}^{1},\ \text{при}\ j = 1 \ldots t_1; t_2 \ldots n \end{cases}. \qquad (6)$$

У разі, якщо в результаті такого обміну тестовими завданнями утворився варіант тесту, що не реалізовувався, тобто не виконуються умови

$$Nq_i \le Nq_{max\,i} \text{ і } \sum_{i=1}^{n} Nq_{r,i} \le Nq_{\sum} \tag{7}$$

виконується коригування, яке можливе одним із двох способів.

Перший спосіб можливий за умов значної кількості завдань в "батьківських" тестах і при цьому в кожному з них кількість контрольованих тем понад десять. Для коригування в тестах випадково визначаються дві точки розриву і завдання обмінюються на початковій і крайній ділянках, а завдання середньої ділянки не змінюються. При цьому пошук точок розриву й обмін завданнями повторюється до тих пір, доки процес виникнення нових варіантів тестів не буде задовольняти умові (7).

Якщо ж число завдань в тесті невелике або в них невелика кількість контрольованих тем, то застосовують другий варіант. Відповідно до нього тест, що не відповідає вимогам умови (7), піддається коригуванню відповідно до такого правила:

- випадково визначаються точки розриву;
- виконується обмін завданнями на початкових ділянках тесту (вектора *N*);
- перевіряється виконання умови

$$Nq_{\sum} - \sum_{i=1}^{n} Nq_{r,i} = 0; \tag{8}$$

- якщо умова не виконується й отримана різниця позитивна, то кількість тестових завдань по останній контрольованій темі збільшується до рівня, що забезпечує дотримання цієї умови;
- якщо умова не виконується й отримана різниця негативна, то кількість тестових завдань по останній контрольованій темі зменшується до тих пір, поки умова не стане дотримуватися або ж останній темі відповідатиме тільки одне тестове завдання. У останньому випадку виконується коригування тестових завдань інших тем, для яких планувалося змінити кількість тестових завдань. Коригування планів здійснюється в зворотному порядку – від кінцевої теми тесту до першої.

Для забезпечення можливості формування більшої кількості варіантів побудови тестів і тим самим запобігання зацикленню завдання застосовується оператор випадкових змін. Вірогідність його застосування перевіряється з умови

$$P_{mut} > Z\ , \tag{9}$$

де P_{mut} – вірогідність застосування оператора випадкових змін, P_{mut} = 0,001 ... 0,01;

Z – випадкова величина в діапазоні від 0 до 1.

В операторі випадкових змін довільно визначаються дві точки розриву вектора $\mathbf{X_i}$, а потім виконується обмін між двома фрагментами тестів з координатами t_1 і t_2. У разі спрацьовування оператора випадкових змін новий план побудови тесту будується відповідно до правила:

$$\begin{cases} x^m_{i,t_1} = x_{i.t_2} \\ x^m_{i,t_2} = x_{i.t_1} \\ x^m_{i,t_1} = x_{i.j}\text{, при } j \neq t_1, t_2\,. \end{cases} \tag{10}$$

Послідовне застосування описаних перетворень до спочатку сформованого плану тестування призводить до створення нових варіантів побудови тестів, які разом з початковими складають розширену групу, що складається з удвічі більшого числа тестів. У зв'язку з цим виконується відбір тестів, у яких сумарна складність завдань найбільш близька до критерійних значень допустимої значущості або складності контрольованого навчального матеріалу. Врешті-решт чисельність нового набору можливих варіантів побудови тестів доводиться до спочатку встановленого рівня шляхом виключення з набору варіантів, найменшою мірою тих, що задовольняють умову (3). При цьому працює оператор "виживання", в основі якого лежить механізм визначення якості сформованого тесту за показниками критерію оптимальності. Тести найгіршої якості виключаються з популяції.

Роботою оператора "виживання" закінчується ітераційний розрахунок і у разі, якщо рішення не отримане, то обчислення повторюється відносно до знову сформованого плану побудови тестів відповідно до встановлених правил. Такі

циклічні обчислення повторюються до тих пір, поки не буде досягнутий оптимум цільової функції (*F=0*) або ж кількість ітераційних циклів не перевищить заздалегідь встановлене граничне значення *N*ц (при формування тестів з використанням описаних процедур рекомендується застосовувати *N*ц =10^4-10^6).

Розглянемо пропоновану процедуру формування тестів на прикладі відбору тестових завдань для рейтингового тестового контролю по дисципліні "Інформаційно-комунікативні технології".

Відповідно до встановленого критерію сукупної складності для рейтингового контролю з дисципліни вимагається сформувати тест із завдань різної складності (див. таблицю 1) так, щоб сумарна складність тесту склала 1296 одиниць.

Таблиця 1

Індекси складності завдань, включених в базу для формування тесту

	Тема 1	Тема 2	Тема 3	Тема 4	Тема 5
	Індекси складності				
Завдання 1	92	82	39	27	46
Завдання 2	68	64	19	54	76
Завдання 3	86	89	82	89	85
Завдання 4	34	94	65	66	63
Завдання 5	69	70	91	48	91
Завдання 6	42	34	93	90	-
Завдання 7	93	-	-	84	-
Завдання 8	-	-	-	32	-

Формуємо тест у наступній послідовності.

- Сортуємо завдання усередині тем випадково (таблиця 2).

Таблиця 2

Результати сортування тестових завдань

Тема 1	Тема 2	Тема 3	Тема 4	Тема 5
93	89	65	66	85
68	82	39	27	46
42	70	91	89	91
92	94	93	32	76

86	34	19	84	63
34	64	82	90	
69			54	
			48	

- Формуємо первинний план формування тесту (таблиця 3).

Таблиця 3

Початковий план з чотирьох варіантів формування тесту

	Тема 1		Тема 2		Тема 3		Тема 4		Тема 5	
	n	*Qsum*	*n*	*Qsum*	*n*	*Qsum*	*n*	*Qsum*	*n*	*Qsum*
Варіант 1	1	93	5	369	4	288	3	182	2	131
Варіант 2	3	203	4	335	2	104	2	93	5	361

Продовження табл. 3

Варіант 3	1	93	3	241	4	288	4	214	2	131
Варіант 4	3	203	5	369	2	104	5	490	3	222
n – кількість завдань з кожної теми для включення в тест (розігрується випадково), Qsum – сума індексів складності *n* завдань.										

- Розраховуємо значення функції *F* і відповідні їм коефіцієнти *Кр*.

Варіант 1: *F*=233, *Кр*=0,217413, Варіант 2: *F* = 200, *Кр* = 0,233831

Варіант 3: *F* = 329, *Кр* = 0,169652, Варіант 4: *F* = - 92, *Кр* = 0,379104

- Визначаємо межі інтервалів.

1) від 0 до 0,217413, 2) понад 0,217413 до 0,451244,

3) понад 0,451244 до 0,620896, 4) понад 0,620896 до 1

- Відбираємо варіанти тесту для включення в "батьківські" пари.

Таблиця 4

Формування "батьківських" пар

Розігрування	*z*=*rnd*(1)=0,612	*z*=*rnd*(1)=0,834	*z*=*rnd*(1)=0,312	*z*=*rnd*(1)=0,964
Відібрані варіанти	Перша пара		Друга пара	
	3	4	2	4

- Виконуємо обмін завданнями і складаємо новий план формування тесту.

Таблиця 5

Обмін завданнями (t_1=*rnd*(4), t_2=*rnd*(4), t_1<t_1)

t_1	t_2	Нові варіанти тесту

2	5	1	3	2	5	2
2	5	3	5	4	4	3
3	5	3	4	2	5	5
3	5	3	5	2	2	3

- Перевіряємо необхідність внесення випадкових змін до плану формування тесту.

Таблиця 6

Перевірка спрацьовування оператора випадкових змін

Варіант	1	2	3	4
Z= *rnd*(1)	0,632542	0,078325	0,009229	0,132163
Z > *Pmut* = 0,008 – оператор випадкових змін не застосовується				

- Складаємо загальний план формування тесту (для кожного нового варіанту розраховується значення функції *F*).

Таблиця 7

Загальний план формування тесту

Варіанти завдань						Сумарна складність	*F*
Варіант 1	93	369	288	182	131	1063	233
Варіант 2	203	335	104	93	361	1096	200
Варіант 3	93	241	288	214	131	967	329
Варіант 4	203	369	104	490	222	1204	-92
Варіант 5	93	241	104	298	131	867	429
Варіант 6	203	369	288	214	222	1296	0
Варіант 7	203	335	104	298	361	1291	-5
Варіант 8	203	369	104	93	222	991	305

- Відбираємо кращі варіанти для нового плану формування тесту.

Таблиця 8

Новий план формування тесту

	Нові варіанти тесту					Сумарна складність	*F*
Варіант 6	3	5	4	4	3	1296	0
Варіант 7	3	4	2	5	5	1291	-5

Варіант 4	3	5	2	5	3	1204	-92
Варіант 2	3	4	2	2	5	1096	200

- Аналізуємо необхідність продовження ітераційного процесу.

Для варіанту 6 оптимум досягнутий, оскільки F=0. Відповідно до нього остаточно до тесту включаємо 19 завдань: 3 з першої теми, 5 з другої, по 4 з третьої і четвертої і 3 з п'ятої теми. Їх сумарна складність складе 1296 одиниць, що відповідає встановленому критерію.

Висновки. Наведені результати дозволяють рекомендувати для включення в імітаційну модель тестового контролю процедуру, що пропонує розглядати сукупність операцій з формування тестів як ітераційний процес, у якому кожен новостворюваний варіант є випадковим набором тестових завдань різної складності. Закладені в ітераційному процесі механізми оптимізації забезпечують формування тесту найкращого за критеріями сукупної складності і значущості контрольованого навчального матеріалу. Подальше проведення робіт з формування імітаційної моделі дасть можливість розширити набір математично точних проектних рішень для більшості типових процедур розробки, проведення і підведення підсумків контролю, які традиційно покладалися на викладача.

Список використаних джерел

1. *Алексеев А. Н.* Дистанционное обучение инженерным специальностям. – Сумы: Университетская книга, 2005. – 333 с.

2. *Алексєєв О. М.*, *Алексєєва Г. В.* Імітаційна модель тестового контролю знань і умінь // Комп’ютерно-орієнтовані системи навчання: Збірник наукових праць. – К: НПУ ім. М. П. Драгоманова, 2009. – Випуск 7(14). – С. 65–71.

3. Базы данных. Интеллектуальная обработка информации / В. В. Корнеев, А. Ф. Гареев, С. В. Васютин, В. В Райх. – М.: Нолидж, 2000. – 452 с.

4. *Ефремова Н. Ф.* Тестовый контроль качества учебных достижений в образовании: дисс. ... д-ра пед. наук: 13.00.01 / Надежда Федоровна Ефремова. – Ростов-на-Дону, 2003. – 458 с.

ФОРМИРОВАНИЕ ТЕСТОВ В ИМИТАЦИОННОЙ МОДЕЛИ КОНТРОЛЯ УСПЕШНОСТИ ОБУЧЕНИЯ

Алексеев А. Н., Король Е. Н.

Аннотация

В статье развиты положения имитационной модели тестового контроля успешности обучения в части совершенствования процедуры формирования теста. Предлагается при принятии решения о включении тестового задания учитывать сложность задания и использовать итерационные вычисления с тем, чтобы с помощью алгоритмов оптимизации управляющих процессом случайного перебора создать тест, у которого сумма индексов сложности заданий соответствует критериям совокупной значимости или сложности. Дается развернутое изложение математического аппарата, используемого для обоснования принимаемых решений по формированию теста. Приводится пример, демонстрирующий основные этапы итерационных вычислений и механизмы достижения оптимальности по критериям совокупной значимости и сложности.

Ключевые слова: тестирование, тестовый контроль успешности обучения, имитационная модель, тестовое задание, тест, формирование теста, критерии оптимальности теста, совокупная значимость и сложность теста.

HOW TO BUILD TESTS IN THE IMITATION MODEL FOR TEST-BASED KNOWLEDGE CONTROL

Alexeyev A., Korol E.

Resume

The principles of imitation model for knowledge test control, specifically on the perfection of the procedure of constructing a test, are developed in this model. The authors suggest taking into account the difficulty of the question when one is making a decision about including a given test question into the test. They also suggest using iterational calculations in order to create a test with the help of optimization algorithms that are used for the process of random searching. The sum of task complexity indices for such test will meet the criteria of joint value and difficulty. Detailed explanation of mathematical apparatus that is used for decision-making during the test-building process is given in the article as well as an example that demonstrates main steps of iterational calculations and mechanisms for achieving optimal test structure based on the criteria of joint value and difficulty.

Keywords: testing, knowledge test control, imitation model, test question, test, building a test, criteria for optimal test structure, joint value and difficulty of a test.