

УДК 651.3:518.5

ТЕХНОЛОГІЯ АВТОМАТИЧНОЇ ОЦІНКИ СКЛАДНОСТІ ЗАВДАНЬ В АДАПТИВНІЙ СИСТЕМІ ДИСТАНЦІЙНОГО НАВЧАННЯ ТА КОНТРОЛЮ ЗНАНЬ

П.І. Федорук

Прикарпатський національний університет ім.Василя Стефаника

e-mail: pavlo@pu.if.ua

1. Вступ

Дистанційне навчання сьогодні набуло широкого впровадження у освітніх системах багатьох країн. Можливості такого навчання дозволяють як піднести на якісно новий рівень звичні форми навчання, так і організувати нову форму навчання, при якому викладач і студент можуть бути розділені в часі і просторі. Розвиток систем дистанційного навчання і контролю знань вийшов на рівень, коли вже недостатньо просто мати статичний набір електронних підручників і тестів, як це ми можемо бачити у більшості існуючих систем. Для використання в повній мірі можливостей дистанційного навчання такі системи повинні вміти враховувати індивідуальні особливості тих, хто навчається, тобто бути адаптивними відповідно до цих особливостей [1].

Однією із складових частин таких адаптивних систем є модуль контролю знань, який відповідає за адаптивне тестування. Адаптивний тест являє собою варіант автоматизованої системи тестування із заздалегідь відомими параметрами складності і здатності кожного завдання. В такому випадку говорять про тести з критеріально-орієнтованою інтерпретацією результатів. Критеріально-орієнтований підхід дозволяє оцінити, в якій мірі випробуваний оволодів необхідним для професійної діяльності навчальним матеріалом. Критеріально-орієнтовані тести дають змогу оцінити рівень ефективності програми навчання. При конструюванні критеріально-орієнтованого тесту важливо точно знати складність кожного завдання із тестової бази, для того, щоб в майбутньому об'єктивно оцінити рівень знань студента.

2. Оцінка рівня складності завдань

На практиці при тестуванні зазвичай формується та використовується кілька варіантів одного і того ж тесту. В основному, це пов'язано із необхідністю захисту бази тестових завдань від тиражування правильних відповідей серед учасників тестування. Однак виникає необхідність співставлення результатів, отриманих за паралельними формами тестів, що є непростим завданням. Процес оцінювання складності кожного завдання можна назвати калібруванням [2]. Якщо всі результати тестування обробляти окремо один від одного, то одержимо стільки метричних шкал, скільки було варіантів тесту – для кожного варіанту своя шкала [3].

Процедура створення єдиної метричної шкали на множині всіх тестових завдань різних варіантів тесту називається **вирівнюванням**. Для надійного виконання вирівнювання тести зазвичай складають таким чином, щоб різні варіанти тесту мали невелике перекриття, наприклад, три завдання, які повинні ділити всю шкалу рівнів складності завдань (тобто множину значень δ_j , j – номер завдання) приблизно на три рівні інтервали. Іншими словами, ці три завдання належать до трьох різних рівнів складностей завдань відповідно.

Спільні завдання різних варіантів одного і того ж тесту називають **вузловими**.

Даний підхід **вирівнювання за вузловими завданнями** базується на одночасній обробці всіх варіантів тесту методом найменших квадратів.

Отже, нехай обробці підлягають l , $l = \overline{1, M}$, різних варіантів тесту.

N_l – кількість учасників l^{ozo} варіанту тесту. K – кількість завдань в кожному варіанті тесту.

Таким чином, в результаті тестування буде отримано l різних матриць відповідей, кожна з яких має розмірність $N_l \times K$. Результати по кожній із матриць обробляються за моделлю Раша [4].

Тоді після математичної обробки відповідей одержуємо оцінки латентних параметрів складності завдання δ_{jl} , $j = \overline{1, K}$, $l = \overline{1, M}$, а також оцінки відповідних середньоквадратичних похибок $\sigma(\delta_{jl})$.

Латентні параметри складності завдань δ_{jl} , одержані для кожного із варіантів, належать до метричних, а не до нормованих шкал (можна виміряти відстань між параметрами в логітах, але не можна виміряти відстань від початку відліку). Всі обрахунки за такими шкалами можна змішувати без втрати інформації. Для зведення всіх результатів до єдиної метричної шкали необхідне перекриття завдань (одне і те ж завдання виконують різні учасники) в різних варіантах тесту.

Нехай у нас є три однакових для всіх варіантів тесту завдання із рівнями складності $\delta_{jl}^{(1)}$, $\delta_{jl}^{(2)}$, $\delta_{jl}^{(3)}$. Верхній індекс в круглих дужках показує взаємне розміщення складностей завдань (ці три завдання мають три різні рівні складності).

Для створення єдиної метричної шкали за всіма варіантами тесту необхідно в першу чергу задати умовний початок (нуль) метричної шкали для всіх варіантів. Для цього від всіх оцінок латентних параметрів δ_{jl} віднімаємо значення $\delta_{jl}^{(2)}$.

Обчислюємо середньоквадратичні відхилення $\sigma^2(\delta_{jl}^{(1)})$ та $\sigma^2(\delta_{jl}^{(3)})$:

$$\sigma^2(\delta_{jl}^{(1)}) = \frac{\sum_{k=1}^3 (\delta_{kl} - \delta_{jl}^{(1)})^2}{3}, \quad \sigma^2(\delta_{jl}^{(3)}) = \frac{\sum_{k=1}^3 (\delta_{kl} - \delta_{jl}^{(3)})^2}{3}.$$

Наприклад,

$$\begin{aligned} \sigma^2(\delta_{11}^{(1)}) &= \frac{\sum_{j=1}^3 (\delta_{j1} - \delta_{11}^{(1)})^2}{3}, & \sigma^2(\delta_{31}^{(3)}) &= \frac{\sum_{j=1}^3 (\delta_{j1} - \delta_{31}^{(3)})^2}{3}, \\ \sigma^2(\delta_{12}^{(1)}) &= \frac{\sum_{j=1}^3 (\delta_{j2} - \delta_{12}^{(1)})^2}{3}, & \sigma^2(\delta_{32}^{(3)}) &= \frac{\sum_{j=1}^3 (\delta_{j2} - \delta_{32}^{(3)})^2}{3}. \end{aligned}$$

Тут j – номер вузлових завдань в варіантах тесту, l – номер варіанта тесту, $l = \overline{1, M}$.

Обчислюємо ваги оцінок $\delta_{jl}^{(1)}$, $\delta_{jl}^{(3)}$:

$$W_{jl}^{(1)} = \frac{1}{\sigma^2(\delta_{jl}^{(1)})}, \quad W_{jl}^{(3)} = \frac{1}{\sigma^2(\delta_{jl}^{(3)})}.$$

Обчислюємо середні вагові значення $\bar{\delta}^{(1)}$, $\bar{\delta}^{(3)}$ ($\bar{\delta}^{(2)} = 0$).

$$\bar{\delta}^{(1)} = \frac{\sum_{l=1}^M W_{jl}^{(1)} \delta_{jl}^{(1)}}{\sum_{l=1}^M W_{jl}^{(1)}}, \quad \bar{\delta}^{(3)} = \frac{\sum_{l=1}^M W_{jl}^{(3)} \delta_{jl}^{(3)}}{\sum_{l=1}^M W_{jl}^{(3)}}.$$

В результаті складностей вузлових завдань присвоюються наступні значення:

$$\bar{\delta}^{(1)} < 0, \quad \bar{\delta}^{(2)} = 0, \quad \bar{\delta}^{(3)} > 0$$

Обчислюємо $V_l^{(1)}$, $V_l^{(3)}$ — відхилення δ_{jl} :

$$V_l^{(1)} = \bar{\delta}^{(1)} - \delta_{jl}^{(1)}, \quad V_l^{(3)} = \bar{\delta}^{(3)} - \delta_{jl}^{(3)}, \quad l = \overline{1, M}.$$

На наступному етапі накладаємо результати (початкові складності завдань, отримані на основі моделі Раша [4]) на єдину метричну шкалу, використовуючи формулу :

$$\begin{cases} \delta_{jl} < 0, & \text{то} \quad \delta_{jl}^* = \delta_{jl} + \frac{V_l^{(1)}}{\delta_{jl}^{(1)}} \delta_{jl}, \\ \delta_{jl} \geq 0, & \text{то} \quad \delta_{jl}^* = \delta_{jl} + \frac{V_l^{(3)}}{\delta_{jl}^{(3)}} \delta_{jl}. \end{cases}$$

Але оскільки

$$\delta_{jl}^* = \delta_{jl} + \frac{V_l^{(1)}}{\delta_{jl}^{(1)}} \delta_{jl} = \delta_{jl} \left(1 + \frac{\bar{\delta}^{(1)} - \delta_{jl}^{(1)}}{\delta_{jl}^{(1)}} \right) = \delta_{jl} \left(1 + \frac{\bar{\delta}^{(1)}}{\delta_{jl}^{(1)}} - 1 \right) = \delta_{jl} \frac{\bar{\delta}^{(1)}}{\delta_{jl}^{(1)}}, \quad (\delta_{jl} < 0)$$

і аналогічно $\delta_{jl}^* = \delta_{jl} \frac{\bar{\delta}^{(3)}}{\delta_{jl}^{(3)}}, \quad (\delta_{jl} \geq 0).$

Отримаємо остаточний результат.

Для вузлових завдань:
$$\begin{cases} \delta_{jl} < 0, & \text{то} \quad \delta_{jl}^* = \delta_{jl} \frac{\bar{\delta}^{(1)}}{\delta_{jl}^{(1)}}, \\ \delta_{jl} \geq 0, & \text{то} \quad \delta_{jl}^* = \delta_{jl} \frac{\bar{\delta}^{(3)}}{\delta_{jl}^{(3)}}. \end{cases}$$

Для не вузлових завдань :
$$\begin{cases} \delta_{jl} < 0, & \text{то} \quad \delta_{jl}^* = \delta_{jl} \frac{\bar{\delta}^{(1)}}{\delta_{1l}^{(1)}}, \\ \delta_{jl} \geq 0, & \text{то} \quad \delta_{jl}^* = \delta_{jl} \frac{\bar{\delta}^{(3)}}{\delta_{3l}^{(3)}}. \end{cases}$$

Розглянемо даний метод вирівнювання за вузловими завданнями на прикладі двох варіантів одного і того ж тесту.

Обидва тести складаються і з шести завдань, три з яких спільні. Використавши однопараметричну модель Раша [4], знаходимо значення латентних параметрів складності для всіх тестових завдання. Складність кожного завдання позначаємо δ_{jl} , де j – номер завдання в тесті,

l – номер варіанту тесту. Одержимо :

$\delta_{11}, \delta_{21}, \delta_{31}, \delta_{41}, \delta_{51}, \delta_{61}$ – множина складностей завдань першого тесту;

$\delta_{12}, \delta_{22}, \delta_{32}, \delta_{42}, \delta_{52}, \delta_{62}$ – множина складностей завдань другого тесту.

Табл.1. Перший варіант тесту.

δ_{11}	δ_{21}	δ_{31}	δ_{41}	δ_{51}	δ_{61}
-1,71	0,31	1,2	-0,99	-0,47	0,5
-2,02	0	0,89	-1,3	-0,78	-0,19

Табл.2. Другий варіант тесту.

δ_{12}	δ_{22}	δ_{32}	δ_{42}	δ_{52}	δ_{62}
-1,66	0,39	1,5	0,47	0,99	-1,71
-2,05	0	1,11	0,08	0,6	-2,1

Позначимо $\delta_{11}^{(1)} = -2,02$, $\delta_{12}^{(1)} = -2,05$,

$$\delta_{31}^{(3)} = 0,89, \quad \delta_{32}^{(3)} = 1,11,$$

$$\sigma^2(\delta_{11}^{(1)}) = \frac{(-2,02 + 2,02)^2 + (0 + 2,02)^2 + (0,89 + 2,02)^2}{3} = 4,18,$$

$$\sigma^2(\delta_{31}^{(3)}) = \frac{(-2,02 - 0,89)^2 + (0 - 0,89)^2 + (0,89 - 0,89)^2}{3} = 3,09,$$

$$\sigma^2(\delta_{12}^{(1)}) = \frac{(-2,05 + 2,05)^2 + (0 + 2,05)^2 + (1,11 + 2,05)^2}{3} = 4,73,$$

$$\sigma^2(\delta_{32}^{(3)}) = \frac{(-2,05 - 1,11)^2 + (0 - 1,11)^2 + (1,11 - 1,11)^2}{3} = 3,74.$$

Ваги оцінок складностей вузлових завдань:

$$W_{11}^{(1)} = \frac{1}{4,18} = 0,24,$$

$$W_{12}^{(1)} = \frac{1}{4,73} = 0,21,$$

$$W_{31}^{(3)} = \frac{1}{3,09} = 0,32,$$

$$W_{32}^{(3)} = \frac{1}{3,74} = 0,27.$$

Середні вагові значення складностей вагових завдань:

$$\bar{\delta}^{(2)} = 0, \quad \bar{\delta}^{(1)} = \frac{W_{11}^{(1)}\delta_{11}^{(1)} + W_{12}^{(1)}\delta_{12}^{(1)}}{W_{11}^{(1)} + W_{12}^{(1)}}, \quad \bar{\delta}^{(3)} = \frac{W_{31}^{(3)}\delta_{31}^{(3)} + W_{32}^{(3)}\delta_{32}^{(3)}}{W_{31}^{(3)} + W_{32}^{(3)}},$$

$$\bar{\delta}^{(1)} = \frac{0,24(-2,02) + 0,21(-2,05)}{0,24 + 0,21} = -2,03,$$

$$\bar{\delta}^{(3)} = \frac{0,32 \cdot 0,89 + 0,27 \cdot 1,11}{0,32 + 0,27} = 0,99.$$

Результат вирівнювання для вузлових завдань :

$$\delta_{11} < 0, \quad \text{modi} \quad \delta_{11}^* = \delta_{11} \frac{\bar{\delta}^{(1)}}{\delta_{11}^{(1)}} = (-2,02) \frac{-2,03}{-2,02} = -2,03;$$

$$\delta_{12} < 0, \quad \text{modi} \quad \delta_{12}^* = \delta_{12} \frac{\bar{\delta}^{(1)}}{\delta_{12}^{(1)}} = (-2,05) \frac{-2,03}{-2,05} = -2,03;$$

$$\delta_{21} = \delta_{22} = 0, \quad \text{modi} \quad \delta_{21}^* = \delta_{22}^* = \delta_{21} \frac{\bar{\delta}^{(2)}}{\delta_{21}^{(2)}} = 0;$$

$$\delta_{31} > 0, \quad \text{modi} \quad \delta_{31}^* = \delta_{31} \frac{\bar{\delta}^{(3)}}{\delta_{31}^{(3)}} = 0,89 \frac{0,99}{0,89} = 0,99;$$

$$\delta_{32} > 0, \text{ тоді } \delta_{32}^* = \delta_{32} \frac{\bar{\delta}^{(3)}}{\delta_{32}^{(3)}} = 1,11 \frac{0,99}{1,11} = 0,99.$$

Результат вирівнювання для решти шести невузлових завдань :

$$\delta_{41} < 0, \text{ тоді } \delta_{41}^* = \delta_{41} \frac{\bar{\delta}^{(1)}}{\delta_{41}^{(1)}} = (-1,3) \frac{-2,03}{-2,02} = -1,306;$$

$$\delta_{51} < 0, \text{ тоді } \delta_{51}^* = \delta_{51} \frac{\bar{\delta}^{(1)}}{\delta_{51}^{(1)}} = (-0,78) \frac{-2,03}{-2,02} = -0,7839;$$

$$\delta_{61} > 0, \text{ тоді } \delta_{61}^* = \delta_{61} \frac{\bar{\delta}^{(3)}}{\delta_{61}^{(3)}} = 0,19 \frac{0,99}{0,89} = 0,21;$$

$$\delta_{42} > 0, \text{ тоді } \delta_{42}^* = \delta_{42} \frac{\bar{\delta}^{(3)}}{\delta_{42}^{(3)}} = 0,08 \frac{0,99}{1,11} = 0,07;$$

$$\delta_{52} > 0, \text{ тоді } \delta_{52}^* = \delta_{52} \frac{\bar{\delta}^{(3)}}{\delta_{52}^{(3)}} = 0,6 \frac{0,99}{1,11} = 0,5351;$$

$$\delta_{62} < 0, \text{ тоді } \delta_{62}^* = \delta_{62} \frac{\bar{\delta}^{(1)}}{\delta_{62}^{(1)}} = (-2,1) \frac{-2,03}{-2,05} = -2,0795.$$

Звідси отримуємо остаточний результат (Табл.3).

Табл.3. Остаточний результат.

№ завд.	1	2	3	4 (№41)	5 (№51)	6 (№61)	7 (№42)	8 (№52)	9 (№62)
δ_j	-2,03	0	0,99	-1,306	-0,78	0,21	0,07	0,54	-2,08

3. Висновок

Таким чином, ми **однозначно** отримали числове значення складності кожного завдання із тестової бази. На наступному етапі отриману числову шкалу розбивають на інтервали складності (як правило, три). Тепер кожне завдання належить до одного рівня складності (низького, середнього чи важкого).

Вище описані дії можна назвати первинним етапом тестування, на якому ми оцінюємо не самих студентів, а завдання. В майбутньому, після першого тестування конкретного учасника і визначення його початкового рівня знань, система повинна підбирати тестові завдання з тим рівнем складності, що відповідає його рівню знань.

Використання даної технології автоматичної оцінки рівня складності тестових завдань дозволить забезпечити об'єктивну оцінку рівня знань тих, хто навчається і організувати навчальний процес згідно до індивідуальної навчальної траєкторії кожного студента.

Література

1. Федорук П.І. Графо-автоматна модель адаптивної системи дистанційного навчання та контролю знань // Математичні машини і системи. – 2006. – №4. – С. 144-154.
2. Frank B. Baker. The Basics of Item Response Theory. University of Wisconsin, ERIC Clearinghouse on Assessment and Evaluation, 2001, pp. 133-142.
3. Stocking M.L., Lord F.M. Developing a common metric in Item Response Theory. Applied Psychological Measurement, vol. 7, No. 2, 1983, pp. 201-210.
4. Rash G. Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen, Denmark: Danish Institute for Educational Reserver, 1960. – 199 p.