УДК 378.14

І.В. Алєксєєва, В.О. Гайдей, О.О. Диховичний, Н.Р. Коновалова, Л.Б. Федорова
Національний технічний університет України “Київський політехнічний інститут”

СТАТИСТИЧНИЙ АНАЛІЗ ТЕСТОВИХ ЗАВДАНЬ ІЗ ЗАСТОСУВАННЯМ СУЧАСНИХ МАТЕМАТИЧНИХ МОДЕЛЕЙ

Проінформовано про особливості застосування сучасних математичних моделей тестів Мастерса та Андерсена до аналізу якості тестових завдань комплекту дистанційної освіти “Вища математика”, розробленого в НТУУ “КПІ”. Наведено методики оцінювання латентних параметрів за відповідними моделями.

Ключові слова: тестові завдання, латентні параметри, моделі Мастерса, Андерсена.

The present paper informs about features of application Masters and Andersen models to the Item Analysis for outlined web-based courses “Higher mathematics” made in NTUU “KPI”. Methodics of estimating latent parameters are given.

Keywords: Item Analysis, IRT, Masters and Andersen models.

Постановка проблеми

Сучасний період розвитку освіти, зокрема вищої технічної, характеризується всебічним поширенням тестового підходу до контролю знань і потребує застосування сучасних статистичних математичних методів до аналізу якості як загалом тестів, так і окремих тестових завдань.

Кафедра математичного аналізу та теорії ймовірностей НТУУ “КПІ” створила комплект дистанційної освіти “Вища математика” [1]. Комплект містить розгалужену базу тестових завдань за всіма розділами, на основі якої формуються електронні контрольні роботи та завдання. В зв’язку з цим постає проблема статистичного аналізу якості тестових завдань комплекту із застосуванням як класичних методів статистичного аналізу, так і сучасної математичної теорії параметризації тестових завдань, яка має назву Item Response Theory (IRT) [2]. Про експериментальне впровадження політомічної моделі було інформовано у [1], але її практичне застосування виявило низку проблем, шляхи подолання яких і пропонуються у цій статті.

Аналіз останніх досліджень і публікацій

Питанням статистичного аналізу тестів присвячено багато праць [3], але у доступних авторам літературних джерелах містились або загальні, або суто теоретичні поради щодо практичного застосування методів IRT.

Постановка завдання

Мета статті – дослідити особливості практичного застосування математичних моделей політомічних тестових завдань; виявити проблеми, що призводять до розбіжності ітераційного процесу оцінювання латентних параметрів для моделі Мастерса та запропонувати шляхи їх подолання; запровадити модель Андерсена як узагальнення моделі Мастерса; провести порівняльний аналіз моделей Мастерса та Андерсена; розробити алгоритми оцінювання латентних параметрів на підставі цих моделей та застосувати алгоритми до аналізу тестових завдань комплекту “Вища математика”.

Статистичний аналіз тестів

У політомічній моделі Partial Credit Scoring застосовано базову ідею моделі Г. Раша [2], яку детально вивчено у роботі Дж. Мастерса [4]. Надалі називатимемо її моделлю Раша–Мастерса.

Нехай N іспитників виконують тест, що складається з L завдань, причому кожне i-те завдання має m_i підрівнів. Розглянемо дві множини латентних параметрів:

θ_n – параметри підготованості n-го іспитника, $n = 1, 2, \ldots, N$, N – кількість іспитників;

β_{ij} – параметри складності j-го рівня i-го завдання, $i = 1, 2, \ldots, L$, $j = 0, 1, 2, \ldots, m_i$, m_i – максимальний рівень i-го завдання.

Кожен іспитник отримує за i-те завдання $j = 0, 1, \ldots, m_i$ балів. Тоді ймовірність отримання n-м іспитником j балів за i-те завдання означується як

$$P_{ij}(\theta_n) = \frac{\exp \sum_{l=0}^{j} (\theta_n - \beta_{il})}{\sum_{k=0}^{m_i} \exp \sum_{l=0}^{k} (\theta_n - \beta_{il})}.$$

Для оцінювання відповідних латентних параметрів було застосовано методи роботи [4], згідно з якими оцінки параметрів можна отримати шляхом розв'язання нелінійної системи рівнянь:

$$\begin{cases} r_n - \sum_{i=1}^{L} \sum_{k=1}^{m_i} k P_{ik}(\theta_n) = 0, & n = 1, 2, \ldots, N, \\ -S_{ij} + \sum_{n=1}^{N} \sum_{k=j}^{m_i} P_{ik}(\theta_n) = 0, & i = 1, 2, \ldots, L,\ j = 1, 2, \ldots, m_i, \end{cases}$$

де $r_n = \sum_{i=1}^{L} x_{ni}, n = 1, 2, \ldots, N$ – кількість балів n-го іспитника за тестову роботу; S_{ij} – кількість іспитників, які отримали за i-те завдання не менш як j балів.

Цю систему можна розв'язати, наприклад, методом Ньютона–Рафсона, який дає такі ітераційні формули:

$$\theta_n^{(t+1)} = \theta_n^{(t)} - \frac{r_n - \sum_{i=1}^{L} \sum_{k=1}^{m_i} k P_{ik}(\theta_n)}{-\sum_{i=1}^{L} \left[\sum_{k=1}^{m_i} k^2 P_{ik}(\theta_n) - \left(\sum_{k=1}^{m_i} k P_{ik}(\theta_n) \right)^2 \right]}, \quad n = 1, 2, \ldots, N,$$

$$\beta_{ij}^{(t+1)} = \beta_{ij}^{(t)} - \frac{-S_{ij} + \sum_{n=1}^{N} \sum_{k=j}^{m_i} P_{ik}(\theta_n)}{-\sum_{n=1}^{N} \left[\sum_{k=j}^{m_i} P_{ik}(\theta_n) - \left(\sum_{k=j}^{m_i} P_{ik}(\theta_n) \right)^2 \right]}, \quad \begin{array}{l} i = 1, 2, \ldots, L, \\ j = 1, 2, \ldots, m_i \end{array}$$

за такої умови завершення ітераційного процесу:

$$\sqrt{\sum_{n=1}^{N} (\theta_n^{(t)} - \theta_n^{(t-1)})^2 + \sum_{i=1}^{L} \sum_{j=1}^{m_i} (\beta_{ij}^{(t)} - \beta_{ij}^{(t-1)})^2} < \varepsilon,$$

де ε – заздалегідь задана точність обчислення.

Але з широким впровадженням у практику такого підходу з'явилися проблеми спотворення результатів оцінювання та розбіжності ітераційного процесу, зокрема через:

1) вибір початкових значень;

2) зсув розв'язків;

3) появу тривіальних рівнів.

1. Наведені у роботі [1] рекомендації щодо вибору початкових значень містили неточності. Існує узагальнення відповідних початкових значень параметрів дихотомічної моделі на випадок політомічної:

$$\theta_n^{(0)} = \ln \frac{r_n}{M - r_n}, n = 1, 2, \ldots, N,$$

$$\beta_{ij}^{(0)} = \ln \frac{N - S_{ij}}{S_{ij}}, i = 1, 2, \ldots, L, \; j = 1, 2, \ldots, m_i,$$

де M – максимальний бал за тестову роботу. Такі значення у більшості випадків під час виконання інших додаткових умов забезпечують збіжність ітераційного процесу.

2. Специфіка системи рівнянь призводить до появи такого ефекту: якщо

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_N), \;\; \boldsymbol{\beta} = \left(\beta_{11}, \ldots, \beta_{1m_1}, \ldots, \beta_{L1}, \ldots, \beta_{Lm_L}\right)$$

є розв'язками системи рівнянь, то

$$\boldsymbol{\theta}^* = (\theta_1 + C, \theta_2 + C, \ldots, \theta_N + C),$$

$$\boldsymbol{\beta}^* = \left(\beta_{11} + C, \ldots, \beta_{1m_1} + C, \ldots, \beta_{L1} + C, \ldots, \beta_{Lm_L} + C\right),$$

де C – довільна стала, також є розв'язками системи рівнянь, тобто зсунуті параметри є розв'язками системи. Проведенням процедури усереднення [3]:

$$\frac{\sum_{i=1}^{L} \sum_{j=1}^{m_i} \beta_{ij}}{\sum_{i=1}^{L} m_i} = 0$$

вибирають правильне значення оцінених параметрів.

3. Якщо всі іспитники не досягли певного рівня або всі подолали певний рівень, то виникає явище так званих тривіальних рівнів. Формально це визначається так.

Нехай R_{ij} — кількість іспитників, які в i-му завданні здобули результат j. Тоді j-й рівень i-го завдання називається тривіальним, якщо $R_{ij} = 0.$ Відповідні ймовірності дорівнюють нулеві:

$$\sum_{n=1}^{N} P_{ij}(\theta_n) = 0.$$

Це спричиняє розбіжність ітераційного процесу:

$$\theta_n^{(t)} - \beta_{ij}^{(t)} \xrightarrow[t \to \infty]{} -\infty, \forall n = 1, 2, \ldots, N,$$

де t – номер кроку ітерацій.

Найпростіший спосіб усунення цього явища – об'єднання тривіальних стовпців із сусідніми. Але, згідно з [4], це спотворює результати оцінювання.

Альтернативним шляхом подолання вказаних проблем є впровадження моделі Е. Андерсена [5], згідно з якою ймовірність отримання n-м іспитником j-го рівня за i-те завдання визначається як

$$P_{ij}(\theta_n) = \frac{\exp(a_{ix}\theta_n - \eta_{ix})}{\sum_{k=0}^{m_i} \exp(a_{ik}\theta_n - \eta_{ik})},$$

де $\eta_{i0} \equiv 0$; $\theta_n, n = 1, 2, \ldots, N,$ – параметр підготованості n-го іспитника;

$\eta_{ij}, i = 1, 2, \ldots, L, \; j = 1, 2, \ldots, m_i,$ – параметр складності j-го рівня i-го завдання;

$a_{ij}, i = 1, 2, \ldots, L, \; j = 1, 2, \ldots, m_i,$ – бал за досягнення j-го рівня i-го завдання;

$d = a_{ij} - a_{i(j-1)}, i = 1, 2, \ldots, L, \; j = 2, 3, \ldots, m_i.$

Якщо запровадити

$$a_{ij} = j, i = 1, 2, \ldots, L, \; j = 0, 1, \ldots, m_i$$

та

$$\eta_{ij} = \sum_{k=1}^{j} \beta_{ik}, i = 1, 2, \ldots, L, \; j = 1, 2, \ldots, m_i,$$

то модель Андерсена збігається з моделлю Раша–Мастерса.

Для тривіального рівня

$$\exp(-\eta_{ij}) \equiv 0, a_{ij} \equiv 0.$$

Окрім того,

$$a_{ib_i} \equiv 0, \eta_{ib_i} \equiv 0,$$

де b_i – порядковий номер найнижчого нетривіального рівня i-го завдання. Отже, модель Андерсена охоплює випадок виникнення появи тривіальних стовпців у моделі Раша–Мастерса.

Значення параметрів знаходяться за системою

$$\begin{cases} \sum_{i=1}^{L} a_{ix_{ni}} - \sum_{i=1}^{L} \sum_{k=0}^{m_i} a_{ik} P_{ik}(\theta_n) = 0, & n = 1, 2, \ldots, N, \\ -R_{ij} + \sum_{n=1}^{N} P_{ij}(\theta_n) = 0, & i = 1, 2, \ldots, L, \; j = 1, 2, \ldots, m_i, \end{cases}$$

де x_{ni} – кількість балів n-го іспитника за i-те завдання.

Відповідні ітераційні формули оцінювання параметрів набувають вигляду:

$$\theta_n^{(t+1)} = \theta_n^{(t)} - \frac{\sum_{i=1}^{L} a_{ix_{ni}} - \sum_{i=1}^{L} \sum_{k=0}^{m_i} a_{ik} P_{ik}(\theta_n)}{-\sum_{i=1}^{L} \left[\sum_{k=0}^{m_i} a_{ik}^2 P_{ik}(\theta_n) - \left(\sum_{k=0}^{m_i} a_{ik} P_{ik}(\theta_n) \right)^2 \right]}, \quad n = 1, 2, \ldots, N,$$

$$\eta_{ij}^{(t+1)} = \eta_{ij}^{(t)} - \frac{-R_{ij} + \sum_{n=1}^{N} P_{ij}(\theta_n),}{-\sum_{n=1}^{N} \left(P_{ij}(\theta_n) - (P_{ij}(\theta_n))^2 \right)}, \quad \begin{matrix} i = 1, 2, \ldots, L, \\ j = 1, 2, \ldots, m_i. \end{matrix}$$

За відповідної умови збіжності ітераційного процесу:

$$\sqrt{\sum_{n=1}^{N} (\theta_n^{(t)} - \theta_n^{(t-1)})^2 + \sum_{i=1}^{L} \sum_{j=1}^{m_i} (\eta_{ij}^{(t)} - \eta_{ij}^{(t-1)})^2} < \varepsilon,$$

де ε – заздалегідь задана точність обчислення.

Наведені алгоритми разом із графічними засобами інтерпретації результатів покладено в основу комплексу програм для пакета MATLAB.

Порівняння результатів обробки еталонних прикладів на підставі моделей Раша–Мастерса та Андерсена підтвердило з точністю $\varepsilon = 10^{-3}$ тотожність оцінок параметрів у випадку відсутності тривіальних стовпців та відмінність їх оцінок у випадку присутності.

На підставі розроблених методик було проаналізовано результати електронних контрольних робіт за різною тематикою, якими було охоплено понад 200 студентів Інституту телекомунікаційних систем і факультету авіаційной та космічної техніки НТУУ "КПІ".

Контрольна робота містила як дихотомічні, так і політомічні тестові завдання. Завдання з множинним вибором або завдання на відповідність розрізнялися за рівнем складності.

Висновки

1. Застосування IRT-методів дає змогу об’єктивніше поглянути на тестові завдання, які переважно складають на підставі досвіду та інтуїції викладача, і значно спрощує первинний аналіз результатів тестів.

2. Впровадження методики оцінювання параметрів за моделлю Андерсена підтверджує її зручність та ефективність порівняно з моделлю Раша–Мастерса.

3. Збільшення кількості студентів, охоплених тестуванням, підвищило вірогідність результатів.

4. Автори вбачають подальшу перспективність втілення відповідних методик для аналізу контролю знань студентів різних форм освіти.

1. Алєксєєва І.В., Гайдей В.О., Диховичний О.О., Коновалова Н.Р., Федорова Л.Б. Статистичний аналіз тестових завдань із застосуванням сучасних математичних моделей. – Інноваційні технологій у вищій школі / Матеріали III Наук.-практ. конф.– Львів, 18–20 жовтня 2011. – С. 119–123. 2. Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. – Chicago: Univ. of Chicago Press, 1980. 3. Linden W., Hambleton R. Handbook of Modern Item Response Theory. – NY: Springer-Verlag, 1997. – 510 p. 4. Masters, G. N. A Rasch model for partial credit scoring // Psychometrika. – 1982. – Vol. 47, №2, June 1982. – p. 149-174. 5. Andersen E.B. A general latent structure model for contingency table data / Wainer H., Messick S. Principles of modern psychological measurement. – New Jersey: Lawrence Erlbaum Associates, 1983.

УДК 709.4, 710.5

О.Б. Біленька
Національний університет “Львівська політехніка”

МОЖЛИВОСТІ АНАЛІЗУ ТА ПІДВИЩЕННЯ ЯКОСТІ ТЕСТОВИХ ЗАВДАНЬ З ВИКОРИСТАННЯМ СИСТЕМИ УПРАВЛІННЯ НАВЧАННЯМ MOODLE

Описано, як за допомогою вбудованих засобів для автоматизації обчислень статистичних показників за результатами виконання тестових завдань можливо контролювати та об’єктивно оцінювати якість тестів.

Ключові слова: тест, тестове завдання, курс, студент, віртуальне навчальне середовище.

Control and objective estimation of a test’ quality are possibly carry out wih help of build-in means for statistic indeces calculation by executive results of the test tasks.

Keywords: test, test task, e-course, student, virtual learning system.

Постановка проблеми

Широке використання електронних систем управління навчанням не в останню чергу обумовлене наявністю в них засобів, які дають змогу зробити технологічнішою найважливішу складову процесу навчання, яка пов’язана з оцінюванням рівня засвоєння студентами навчального матеріалу, що вивчається. Йдеться про використання автоматизованих тестових систем як засобу вимірювання рівня підготовки студентів.