Liet. mat. rink. LMD darbai, **48/49**, 2008, 217–221

Žinių tikrinimo matematinis modelis

Aleksandras KRYLOVAS, Natalja KOSAREVA (VGTU)
el. paštas: akr@fm.vgtu.lt

Reziumė. Pasiūlytas matematinis modelis testo rezultatų tikimybiniam skirstiniui gauti. Tai leidžia formuluoti rekomendacijas parenkant testo užduotis įvairioms testuojamųjų grupėms, priklausomai nuo jų žinių lygio.

Raktiniai žodžiai: testų užduočių sprendimo teorija, žinių matavimas, entropija, tikimybinis skirstinys.

1. Darbo tikslas

Matematiniai modeliai sudaro testų teorijos mokslinį pagrindą ir yra įprastas testų taikymo praktikos įrankis. Pagrindinis klasikinės testų teorijos modelis remiasi matuojamų savybių skirstinio normališkumu, o vienas iš svarbiausių uždavinių – testu išmatuotų dydžių nuokrypių nuo tam tikrų standartinių dydžių skaičiavimas (žr. [1]). Žinių matavimo testuose, ypač kai testuojamųjų nėra labai daug ir kai taikomi nestandartizuoti testai, sunku tikėtis gauti konstrukto (šiuo atveju žinių lygio) normalųjį skirstinį ir tokie atvejai reikalauja sudėtingesnių tikimybinių modelių. Testų užduočių sprendimo matematinės teorijos (IRT – Item Response Theory [2]; dar vadinamos tikimybine testų teorija [3]) pradininku laikomas danų matematikas G. Rašas (G. Rasch) [4]. Jo modelio esmę sudaro prielaida: tikimybė, kad testuojamasis teisingai išspręs testo užduotį priklauso nuo užduoties sunkumo ir nuo testuojamojo asmens tiriamosios savybės – konstrukto. IRT tikslas yra išmatuoti šią savybę – latentinį modelio parametrą.

Tarkime, kad S yra testuojamojo savybė, o D – klausimo sunkumas. Rašo modelyje tikimybė, kad į klausimą bus atsakyta teisingai, išreiškiama formule

$$P(S, D) = \frac{S}{S + D} = \frac{1}{1 + \frac{D}{S}}. \tag{1}$$

Taigi, esant vienodoms abiejų parametų S ir D reikšmėms, teisingo atsakymo tikimybė lygi $0,5$. Apibrėžęs naujus kintamuosius $\theta = \ln S$ ir $\delta = \ln D$ (jų matavimo vienetas vadinamas *logitu*), Rašas gavo *vienparametrinį logistinį modelį*

$$P(\theta, \delta) = \frac{1}{1 + e^{\delta - \theta}}. \tag{2}$$

Galimi įvairūs (1), (2) modelio apibendrinimai, kai papildomi parametrai leidžia atsižvelgti į klausimo skiriamąją gebą, į atsakymo atspėjimo tikimybę ir pan. [3].

1 pastaba. Visais atvejais klausimo charakteristinė funkcija $P(\theta, \delta)$ yra nemažėjanti parametro θ – testuojamųjų savybės lygio – atžvilgiu.

Tarkime, kad $p \in [0, 1]$ yra testuojamojo žinių lygis. Jį galima suprasti kaip studento įsisavintą dalyko programos dalį, t.y. tam tikrą kriterinį žinių įvertį [5]. Atkreipkime dėmesį, kad p ne tik sunkiai išmatuojamas, bet ir nėra vienareikšmiškai apibrėžiamas dydis. Pažymėkime $k_j(p)$ tikimybę, kad studentas, turintis žinių lygį p, teisingai atsakys į j-ąjį testo klausimą. Taigi, atsižvelgdami į 1 pastabą, galime pasirinkti bet kurią nemažėjančią funkciją k: $[0, 1] \to [0, 1]$. Sudarykime n klausimų rinkinį (testą):

$$T = \{k_1(p), k_2(p), \dots, k_n(p)\}. \tag{3}$$

Pažymėkime $K(p)$ testuojamojo, turinčio žinių lygį p, teisingų atsakymų skaičių, t.y. diskretųjį atsitiktinį dydį

$$P\{K(p) = m\} = t_m(p), \quad m = 0, 1, \dots, n. \tag{4}$$

Tarkime, kad $f(p) = \begin{cases} f_0(p) \geqslant 0, & \text{kai } p \in [0,1], \\ 0, & \text{kai } p \notin [0,1] \end{cases}$ $\left(\int_0^1 f_0(x)\,\mathrm{d}x = 1\right)$ – testuojamųjų populiacijos žinių lygio tikimybinio tankio funkcija. Tada kito diskrečiojo atsitiktinio dydžio $K = \int_0^1 K(p) f_0(p)\,\mathrm{d}p$ prasmė – testo rezultatų pasiskirstymas populiacijoje:

$$P\{K = m\} = p_m = \int_0^1 t_m(p) f_0(p)\,\mathrm{d}p, \quad m = 0, 1, \dots, n. \tag{5}$$

Taigi, m testo taškų gaus vidutiniškai $p_m \cdot 100\%$ testuojamųjų.

Mūsų testavimo modelio tikslas – sudaryti testą T, kuris leistų efektyviai atlikti testuojamųjų *norminį* vertinimą [5], t.y. kad testo rezultatai leistų atskirti silpnesnius studentus nuo stipresnių, tačiau mes **neturime tikslo** išmatuoti latentinį parametrą p. Kitaip tariant, testavimo rezultatai turi būti matuojami *ranginėje skalėje* [6].

Suformuluokime bendruosius modelio sudarymo principus.

- (3) *testo T klausimų skaičius n, klausimų charakteristinės funkcijos $k_j(p)$ ir testuojamųjų populiacijos tankis $f(p)$ laikomi žinomais uždavinio duomenimis.* Taigi šiame darbe mes ne tik turime *a priori* apibrėžtą funkcijų pavidalą, bet ir neturime tikslo įvertinti latentinius parametrus ir tai esminis skirtumas lyginant su IRT teorijos modeliais.
- *Funkcijos $k_j(p)$ ir $f_0(p)$ turi būti pritaikytos neryškioms sąvokoms „lengvas“ arba „sunkus“ klausimas, „stipri“ arba „silpna“ testuojamųjų grupė.* Atkreipkime dėmesį, kad IRT teorijos sudėtingumas gerokai komplikuoja jos taikymus praktikoje [7]. Mes bandome sukurti modelį, kuris ateityje galėtų tapti patogiu praktiniu įrankiu.
- *Modelis pritaikytas testo teikiamos informacijos kiekio maksimizavimui ir orientuotas į norminius vertinimus.*

2. Matematinis modelis

Apskaičiuokime (4) tikimybes $t_m(p)$. Jos lygios generuojančiosios funkcijos daugianario [8] koeficientams:

$$\Psi_n(p,x) = \prod_{j=1}^{n} \bigl(1 - k_j(p) + k_j(p)x\bigr) = t_0(p) + t_1(p)x + \cdots + t_n(p)x^n. \quad (6)$$

Testo teikiamos informacijos kiekį gauname apskaičiavę entropijos funkcijos [9] reikšmę

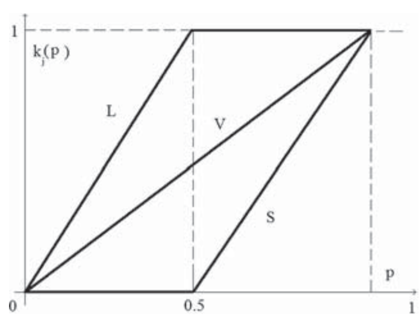
$$E(k_1,\ldots,k_n;\, f_0) = \sum_{m=0}^{n} p_m \ln \frac{1}{p_m}, \quad (7)$$

kai tikimybes p_m, gauname iš (5), (6) formulių. Pastebėkime, kad entropijos matavimo vienetas (šiuo atveju *natas*) priklauso nuo pasirinkto (7) formulėje logaritmo pagrindo, o didžiausią entropiją (informacijos maksimumą) gauname, kai $p_0 = p_1 = \ldots = p_n = \frac{1}{n+1}$. Tada testuojamųjų, gavusių testo rezultatus 0, 1, …, n bus vidutiniškai po lygiai ir testas turės maksimalią skiriamąją gebą.

Klausimų sunkumus apibūdinančias charakteristines funkcijas laikysime atkarpomis tiesinėmis:

$$k_j(p;\alpha,\beta) = \begin{cases} 0, & \text{kai } p < \alpha, \\ \frac{p-\alpha}{\beta-\alpha}, & \text{kai } \alpha \leqslant p \leqslant \beta, \\ 1, & \text{kai } p > \beta. \end{cases} \quad (8)$$

1 pav. parodyti funkcijų $k_L(p;0,0.5)$, $k_V(p;0,1)$ ir $k_S(p;0.5,1)$ grafikai. Šios funkcijos sąlyginai pavadintos atitinkamai lengvo, vidutinio ir sunkaus klausimo charakteristinėmis funkcijomis ir leidžia apibrėžti atitinkamas *neryškias* (*fuzzy*) [10] klausimų aibes. Aišku, kad konkrečios parametrų α ir β reikšmės turėtų būti tikslinamos eksperimentais.



1 pav. Klausimų charakteristinės funkcijos: (L) – lengvas, (V) – vidutinis ir (S) – sunkus.

Žinių lygio pasiskirstymo populiacijoje tankį $f_0(p)$ taip pat apibrėžkime atkarpomis tiesine funkcija

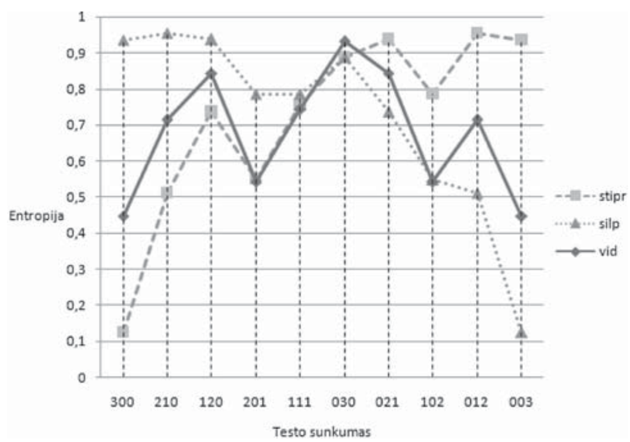
$$f_0(p; a, b, c, d) = \begin{cases} 0, & \text{kai } p \leqslant a, \\ M \cdot \frac{p-a}{b-a}, & \text{kai } a < p < b, \\ M, & \text{kai } b \leqslant p \leqslant c, \\ M \cdot \frac{d-p}{d-c}, & \text{kai } c < p < d, \\ 0, & \text{kai } p \geqslant d, \\ M = \frac{2}{d-a+c-b}. \end{cases} \tag{9}$$

3. Skaitiniai eksperimentai

Modeliui tirti buvo sukurta programa ir atlikti testo rezultatų tikimybinio skirstinio ir entropijos funkcijos reikšmių skaičiavimai trijų klausimų testo atveju. Skaičiavimai atlikti trim testuojamųjų populiacijoms – turintiems silpnas $f_0(p; 0, 0.33, 0.33, 0.67)$, vidutines $f_0(p; 0.25, 0.5, 0.5, 0.75)$ ir stiprias $f_0(p; 0.33, 0.67, 0.67, 1)$ žinias su visomis galimomis klausimų sunkumo funkcijų $k_L(p; 0, 0.5)$, $k_V(p; 0,1)$ ir $k_S(p; 0.5,1)$ kombinacijomis.

2 pav. pavaizduotos normuotos entropijos funkcijų reikšmės įvairiems trijų klausimų rinkiniams. Abscisių ašyje išdėstyti testai pagal jų sunkumo augimą nuo 300 (trys lengvi klausimai) iki 003 (trys sunkūs klausimai). Matome, kad entropijos funkcijos reikšmė, tuo pačiu ir testo teikiamos informacijos kiekis, silpnai populiacijai yra didžausia, kai testą sudaro 2 lengvi ir 1 vidutinio sunkumo klausimai (210), vidutiniai populiacijai, kai yra 3 vidutinio sunkumo klausimai (030), stipriai populiacijai, kai testą sudaro 1 vidutinio sunkumo ir 2 sunkūs klausimai (012).

Dėl apribojimų straipsnio apimčiai mes neteikiame analogiškų skaičiavimų 10 klausimų testui. Paminėsime tik, kad vidutinio stiprumo populiacijai (visų parametrų



2 pav. Trijų klausimų testo entropijos funkcijos silpnai, vidutinei ir stipriai testuojamųjų populiacijoms.

reikšmės nekeičiamos) didžiausią entropiją (nuo 73 iki 83 %) gavome, esant klausimų sunkumo kombinacijoms 334, 442, 253, 235,... ir tokių kombinacijų buvo pakankamai daug.

4. Išvados ir numatomų tyrimų tąsa

Pasiūlytas matematinis modelis leidžia formuluoti praktines rekomendacijas žinių tikrinimo testo klausimams parinkti skirtingoms testuojamųjų grupėms. Modelio patikimumui didinti svarbūs šie klausimai:

- Palyginti parinktas atkarpomis tiesines funkcijas su IRT teorijos logistinėmis funkcijomis.
- Ištirti rezultatų stabilumą funkcijų k_j ir f_0 parametrų α, β, a, b, c, d atžvilgiu.
- Pasiūlyti šių parametrų statistinių įverčių paieškos algoritmus.
- Atlikti eksperimentus su realiais duomenimis.

Literatūra

1. A. Anastasi, S. Urbina *Psychological Testing* (7th edition), Prentice Hall (1997).
2. R.K. Hambleton, H. Swaminathan, H.J. Rogers, *Fundamentals of Item Response Theory*, Sage Publications, Newburg Park (1991).
3. D. Kiseliova, A. Kiseliovas, *Matematinių gebėjimų diagnostika*, Pirmoji knyga. Antroji knyga. Šiaulių universiteto leidykla (2004).
4. G. Rasch, *Probabilistic Models for some intelligence and Attainment Tests*, Danish Institute for Educational Research, Copenhagen (1960). Expanded edition: The University of Chicago Press, Chicago (1980).
5. S. Girdzijauskas, *Studentų žinių kontrolė ir vertinimas*, VU, Vilnius (1999).
6. B. Bitinas, *Pedagoginės diagnostikos pagrindai*, VPU, Vilnius (2002).
7. L.F. Buralčiuk, S.M. Morozov, *Slovar'-spravočnik po psichodiagnostike*, PITER, Sankt-Peterburg (2002).
8. J. Kruopis, *Matematinė statistika*, Mokslo ir enciklopedijų leidykla, Vilnius (1993).
9. V. Stakėnas, *Informacijos kodavimas*, VU, Vilnius (1996).
10. L.A. Zadeh, Fuzzy sets, *Information and Control*, **8**(3), 338–353 (1965).

SUMMARY

A. Krylovas, N. Kosareva. Mathematical model for knowledge testing

There is a mathematical model for obtaining test responses probability distribution proposed. This enables to formulate recomendations for selection of test items for different groups of tested population according to their attainment level.

Keywords: item response theory, attainment testing, entropy, probability distribution.