Liet. mat. rink. LMD darbai, **50**, 2009, 202–207

Diagnostinio testo matematinio modelio tyrimas

Natalja KOSAREVA, Aleksandras KRYLOVAS
Vilniaus Gedimino technikos universitetas
Saulėtekio al. 11, LT-10223 Vilnius
el. paštas: natalja.kosareva@fm.vgtu.lt; akr@fm.vgtu.lt

Santrauka. Siūloma dichotominių diagnostinių operatorių kaip universalios testavimo priemonės kūrimo technologija pagal esamus empirinius duomenis. Vienintelis reikalavimas diagnostiniam operatoriui – tai turi būti nemažėjančioji funkcija. Pagal siūlomą metodiką skaičiuojamas testo rezultato tikimybinis skirstinys ir testo teikiamos informacijos kiekis.

Raktiniai žodžiai: užduoties sprendimo teorija, diagnostinis operatorius, matematinis modeliavimas.

Įvadas

Diagnostiniai uždaviniai atsiranda daugelyje mokslo sričių. Vieną iš jų sprendimo priemonių – testus nagrinėja klasikinė testų teorija ir vėliau susiformavusi užduoties sprendimo teorija (IRT – Item Response Theory). Šios teorijos atsirado, visų pirma, kaip matematiniai instrumentai žinioms bei psichologinėms žmogaus savybėms matuoti. Tokių matavimų specifika – matuojamas objektas yra latentinis, t.y. tiesiogiai nestebimas dydis, todėl matavimo priemonės taip pat yra netiesioginės – tai specialiai sudaryti testai arba klausimynai.

Klasikinėje testų teorijoje diagnostiniai operatoriai (testo klausimai) turi būti parinkti taip, kad stebimas latentinis kintamasis ir surinktų testo balų skaičius turėtų normalųjį skirstinį [1]. Tai gali būti pasiekta tik esant labai dideliam testuojamųjų skaičiui konstruojant standartizuotus testus. IRT buvo atsisakyta latentinio kintamojo normaliojo pasiskirstymo, tačiau reikalaujama specialaus pavidalo (logistinių) diagnostinių operatorių [6].

Siūlomame modelyje atsisakyta apribojimų matuojamo dydžio skirstiniui ir diagnostinio operatoriaus formai. Esminis reikalavimas diagnostiniam operatoriui – tai turi būti nemažėjančioji funkcija.

1. Diagnostinis operatorius

Tarkime, kad $\mathcal{A}$ yra tyrimo objektų (dalykinė) aibė, kurioje apibrėžtas tvarkos sąryšis $\prec$: $(\forall a, b, c \in \mathcal{A})$

1) $a \prec b \ \&\ b \prec a \Rightarrow a = b$ (antisimetriškumas);
2) $a \prec b \ \&\ b \prec c \Rightarrow a \prec c$ (tranzityvumas).

APIBRĖŽIMAI. Funkciją $p: \mathcal{A} \to [0, 1] \subset \mathcal{R}$, $p \neq \mathrm{const}$ vadinsime tyrimo objekto *diagnozuojamuoju požymiu*, jei $a \prec b \Rightarrow p(a) \leqslant p(b)$.

Dvejetą $(\mathcal{A}, \prec)$ vadinsime *empirine sistema*, o trejetą $(\mathcal{A}, \prec, p)$ – *tyrinių sistema*, t.y. tyrimui paruoštų objektų aibę.

Funkcijos p reikšmės $p(a)$ nustatymas *metrologijoje* vadinamas *matavimu* ir grindžiamas fizikiniais matavimo vienetais. Psichologijoje, medicinoje, sociologijoje ir kituose moksluose dažnai apsiribojama tik tam tikrais sąlyginiais vienetais. Pavyzdžiui, nustatoma kraujo apytakos nepakankamumo I, II arba III stadija; žemės drebėjimų intensyvumas vertinamas *seisminės skalės* balais; mineralų kietumas nustatomas Moso (Mohs) skale: 1 – talkas, 10 – deimantas; nespalvintų (naftos) produktų skaitiniam įvertinimui taikoma sutartinė spalvų skalė (*colour scale for undyed products*) nuo 0 iki 500 (maksimalus tamsumas).

Pastaba. Matavimų teorijoje [5] apibrėžiama *ranginė* skalė reikalautų, kad sąryšis $\prec$ būtų pilnasis, o funkcija p – didėjančioji. Taigi mes formuluojame silpnesnius reikalavimus tyrinių sistemai: $\prec$ yra dalinės tvarkos sąryšis, o funkcija p – nemažėjančioji.

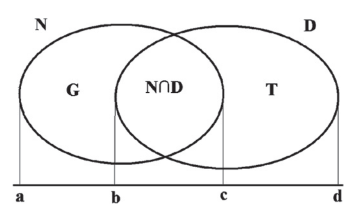
Parodykime, kad pasiūlytas modelis leidžia apibūdinti, pavyzdžiui, inžinerinio objekto *techninę būseną*. Pažymėkime aibes [8]: T – tvarkinga techninė būsena (atitinka visus specifikacijos reikalavimus), N – netvarkinga būsena (bent viena charakteristika neatitinka specifikacijos), D – darbingumas (veikumas – objektas gali atlikti funkcijas, tačiau nepagrindinės charakteristikos gali neatitikti reikalavimų), G – nedarbingumas (neveikumas, gedimas – būsena, kai objektas negali atlikti bent vienos funkcijos). Parodykime šias aibes diagramoje 1 pav.

Matome, kad $G = N \setminus D$, $T = D \setminus N$. Funkciją p galima apibrėžti įvairiais būdais. Pavyzdžiui galima sukonstruoti laiptinę funkciją: pasirinkti skaičius $0 \leqslant a < b < c < d \leqslant 1$ ir priskirti reikšmes: $p(x) = \frac{a+b}{2}$, kai $x \in G$, $p(x) = \frac{b+c}{2}$, kai $x \in N \cap D$ ir $p(x) = \frac{c+d}{2}$, kai $x \in T$. Tačiau, galime, pavyzdžiui, išsamiau nagrinėti aibę G ir suskaidyti ją į *taisomųjų* ir *netaisomųjų* objektų aibes.

Tarkime, kad $D: (\mathcal{A}, \prec, p) \to \{0, 1\}$ yra diskretusis atsitiktinis dydis: $\mathbf{P}\{D = 0\} = 1 - k_D(p)$, $\mathbf{P}\{D = 1\} = k_D(p)$.

APIBRĖŽIMAS. Nemažėjančiąją funkciją $k_D(p): [0, 1] \to [0, 1]$ vadinsime tyrinių sistemos $(\mathcal{A}, \prec, p)$ *diagnostiniu operatoriumi*. Diagnostinis operatorius yra tikimybė, kad į testo klausimą bus atsakyta teisingai (arba *taip*), nagrinėjama kaip latentinio kintamojo p funkcija.

APIBRĖŽIMAS. Diagnostinių operatorių rinkinį $T = \{D_1, D_2, \ldots, D_n\}$ vadinsime *diagnostiniu testu*.



1 pav. Diagnostinio operatoriaus konstravimas.

Funkcijos $k_D(p)$ monotoniškumas atitinka *validumo* reikalavimą diagnostiniam operatoriui [1]. Validumas reiškia savybę matuoti (vertinti) būtent tai, kas numatyta.

Diagnostikos uždavinys, t.y. reikšmės $p(a)$ nustatymas iš aibės $T(a)$ yra dalykinio tyrimo objektas ir šiame darbe nenagrinėjamas. Mūsų tikslas – ištirti, kiek informacijos apie funkciją p teikia *indeksas* – surinktų testo balų skaičius (arba jo vidurkis):

$$S(a) = \sum_{j=1}^{n} D_j(a), \quad s(a) = \frac{1}{n} S(a), \quad a \in \mathcal{A}.$$

Išnagrinėkime ribinius diagnostinių operatorių atvejus: laiptelio ir tiesinės funkcijos.

Absoliučiai tikslios diagnostikos modelis. Tarkime, kad aibė $\mathcal{A}$ suskaidyta į nesikertančius blokus $\mathcal{A} = \bigcup_{j=1}^{n} A_j$, $A_i \cap A_j = \varnothing$ ir p yra laiptinė funkcija $p(a) = p_j$, kai $a \in A_j$, $p_1 < p_2 < \ldots < p_n$. Jei egzistuoja tokie diagnostiniai operatoriai $k_j(p) = \begin{cases} 0, & p < p_j, \\ 1, & p \geqslant p_j, \end{cases}$ tai šie operatoriai yra absoliučiai tikslūs aibės $\mathcal{A}$ elementų A_j indikatoriai (2 pav.). Indekso reikšmė $s(a) = \frac{j}{n}$, kai $a \in A_j$ ir esant reikšmėms $p_j = \frac{j}{n}$, gausime $s(a) = p(a)$.

Absoliučiai blogos diagnostikos modelis. Tarkime, kad diagnostiniai operatoriai $k_\delta(p)$, $\delta \ll 1$ yra tiesinės funkcijos: $k_\delta(p) = 2\delta p + \frac{1}{2} - \delta$. Jei operatoriai D_j yra nepriklausomi vienodai pasiskirstę atsitiktiniai dydžiai, tai $s(a) = 2\delta p + \frac{1}{2} - \delta$. Taigi, kai $\delta \to 0$, gauname $s(a) \approx 0,5$, nepriklausomai nuo $a \in \mathcal{A}$ (2 pav.)

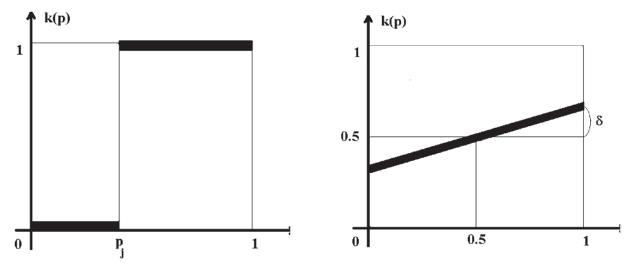
Tarp šių dviejų ribinių atvejų galima sukonstruoti įvairių formų diagnostinius operatorius. Tam tikslui naudojami empiriniu būdu surinkti testavimo duomenys.

2. Diagnostinio operatoriaus savybės

APIBRĖŽIMAS. Diagnostinio operatoriaus $k_D(p)$ *skiriamąja geba* intervale $(\alpha, \beta) \subset [0, 1]$ vadinamas dydis

$$d_{(\alpha,\beta)}(k_D) = \lim_{p \to \beta+0} k_D(p) - \lim_{p \to \alpha-0} k_D(p).$$

Taigi idealaus operatoriaus $k_j(p)$ skiriamoji geba taško p_j aplinkoje lygi 1, operatoriaus $k_\delta(p)$ skiriamoji geba bet kokiame intervale $(\alpha, \beta) \subset [0, 1]$ yra $d_{(\alpha,\beta)}(k_\delta) = 2\delta(\beta - \alpha) \to 0$, kai $\delta \to 0$.



2 pav. Absoliučiai tikslios ir absoliučiai blogos diagnostikos modeliai.

Kita diagnostinio operatoriaus (testo klausimo) charakteristika – jo *sunkumas*. Pavyzdžiui, klausimas, ar universitete dirba Nobelio premijos laureatas yra sunkus: tikimybė, kad į jį bus atsakyta teigiamai yra labai maža, jei tik universiteto reitingas nėra labai aukštas.

APIBRĖŽIMAS. Diagnostinių operatorių aibėje $\mathcal{D}$ apibrėžkime sunkumą apibūdinantį sąryšį $\prec$: jei $\forall p \in [0,1]$ $k_1(p) \leqslant k_2(p)$, rašome $D_1 \prec D_2$ (arba $k_1 \prec k_2$). Tokiu atveju sakysime, kad klausimas D_1 yra *sunkesnis* už D_2.

Pastebėkime, kad sąryšis $\prec$ nėra visiškasis (pilnasis). Pavyzdžiui, 3 pav. pavaizduotoms funkcijoms negalioja nei $k_1 \prec k_2$, nei $k_2 \prec k_1$.

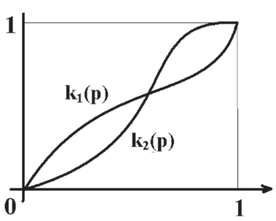
Jei turime kelių sunkumo tipų diagnostinius operatorius $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_k$, tokius, kad $\forall D_i \in \mathcal{D}_i$, $\forall D_j \in \mathcal{D}_j$ $(i < j)$ $D_i \prec D_j$, tai testų rinkinių aibėje irgi galime apibrėžti dalinę tvarką:

$$T^1 = \{D_1^1, \ldots, D_n^1\} \prec T^2 = \{D_1^2, \ldots, D_n^2\}.$$

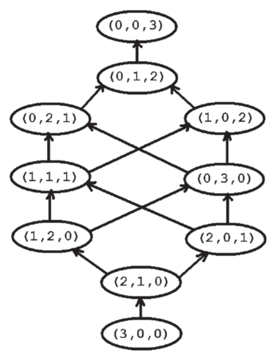
Tarkime, kad $n = k = 3$, t.y. testą sudaro 3 diagnostiniai operatoriai, kurių kiekvienas gali būti trijų sunkumo tipų. Jei visi operatoriai yra $\mathcal{D}_1$ tipo, pažymėkime testą $(3,0,0)$, kai visi trys operatoriai iš skirtingų tipų $\mathcal{D}_1$, $\mathcal{D}_2$, $\mathcal{D}_3$ – $(1,1,1)$ ir t.t. Testo $(3,0,0)$ sunkumas yra didžiausias, o testo $(0,0,3)$ – mažiausias. 4 pav. parodyta dešimties testų $T = \{(3,0,0), (2,1,0), \ldots, (0,0,3)\}$ sunkumo hierarchija. Pastebėję, kad tvarkos sąryšis yra tranzityvusis, paveiksle neparodėme savaime egzistuojančių ryšių, pavyzdžiui, $(3,0,0) \prec (2,0,1)$.

3. Skaitiniai eksperimentai

Tarkime, kad yra žinoma latentinio dydžio p tikimybinio tankio funkcija $f(p)$, $p \in [0,1]$. Sakykime, kad testą sudaro n klausimų. Nagrinėsime atsitiktinio dydžio $S(p) = ns(p)$ – surinktų testo balų skaičiaus skirstinį: $\mathbf{P}\{S = j\} = p_j$, $j = \overline{0,n}$. Jei testas sudarytas iš n nepriklausomų dichotominių diagnostinių operatorių $k_1(p), k_2(p), \ldots,$



3 pav. Diagnostinių operatorių palyginimas.



4 pav. Testų sunkumo hierarchija.

1 lentelė. Teisingų atsakymų į 10 testo klausimų skirstinys ir testo teikiamos informacijos kiekis (I) 100 testuojamųjų grupei silpnai $\mathcal{B}(2;5)$, vidutinio stiprumo $\mathcal{B}(5;5)$ ir stipriai $\mathcal{B}(5;2)$ populiacijoms

Populiacija	I	Teisingų atsakymų skirstinys										
		0	1	2	3	4	5	6	7	8	9	10
$\mathcal{B}(2;5)$	0,88	22	19	15	12	9	7	6	4	3	2	1
$\mathcal{B}(5;5)$	0,96	2	4	7	9	11	13	13	14	12	10	5
$\mathcal{B}(5;2)$	0,76	0	1	1	2	3	5	7	10	15	23	33

$k_n(p)$, tai bendrasis testo balas $S = S(p)$ turi apibendrintąjį binominį skirstinį [4], kurio generuojančioji funkcija lygi:

$$\Psi(p,x) = ES^x = \prod_{j=0}^{n}(1 - k_j(p) + k_j(p)x) = t_0(p) + t_1(p)x + \cdots + t_n(p)x^n.$$

Gautame daugianaryje esantys koeficientai prie skirtingų x laipsnių yra lygūs atitinkamoms sąlyginėms tikimybėms $t_j(p) = P(S = j|p)$, $j = \overline{0,n}$. Bendrojo testo balo skaičiaus S skirstinys visoje populiacijoje gaunamas integruojant tikimybes $t_j(p)$:

$$p_j = P(S = j) = \int_0^1 t_j(p)f(p)\,\mathrm{d}p, \quad j = \overline{0,n}.$$

Testuodami aibę $\tilde{\mathcal{A}} \subset \mathcal{A}$ testu $T = \{D_1, D_2, \ldots, D_n\}$, gauname testo teikiamos (aposteriorinės) informacijos kiekį [7]: $I(k_1, k_2, \ldots, k_n; f) = -\sum_{j=1}^{n} p_j \ln p_j$.

Autorių straipsniuose [2,3] aprašyti skaitiniai eksperimentai, nagrinėjant *atkarpomis tiesines* funkcijas $f(p)$ ir $k_j(p)$ ir galiojant atsitiktinių dydžių D_j *nepriklausomumo* reikalavimui. Esant silpnai, vidutinio stiprumo bei stipriai populiacijoms parodyta, kaip galima iš esamo diagnostinių operatorių (testo klausimų) banko parinkti tokį diagnostinių operatorių rinkinį, kad testo teikiamos informacijos kiekis $I(k_1, k_2, \ldots, k_n; f)$ būtų didžiausias.

Visual C++ terpėje sukurta programa, kuri leidžia skaičiuoti bendrojo testo balo skirstinį ir testo informacijos kiekį esant 4 skirtingiem latentinio kintamojo skirstinių tipams ir 6 diagnostinių operatorių klasėms. Čia pateiksime skaičiavimo eksperimentų rezultatus, kai diagnostinius operatorius modeluojame priklausoma nuo vieno parametro $a \geqslant 0$ funkcija $k(p;a) = \operatorname{arccot}(\frac{a\ln(p)}{\ln(1-p)})$. Populiacija aprašoma Beta skirstiniu, testą sudarė 10 vidutinio sunkumo klausimų. Parametro a reikšmės buvo parinktos taip: 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9; 1; 1,5; 2. Rezultatai pateikti 1 lentelėje.

Pastaba. Lentelėje pateiktas testo teikiamos informacijos kiekis procentais nuo maksimalios reikšmės.

Matome, kad šis testas tinkamiausis vidutinio stiprumo populacijai. Testo informatyvumas ženkliai mažėja, kai jis pateikiamas stipriai populiacijai.

4. Išvados

Siūloma testų kūrimo technologija leidžia kiekvienai tiriamųjų grupei parinkti individualius diagnostinius operatorius ir įvertinti gaunamos informacijos kiekį apie tiriamųjų stebimą savybę. Ši metodika nereikalauja dirbtinių apribojimų tiriamo latentinio dydžio skirstiniui ir diagnostinių operatorių pavidalui. Diagnostiniai operatoriai gali būti labai įvairaus pavidalo funkcijos ir tai esminis mūsų metodikos skirtumas palyginti, pavyzdžiui, su IRT metodais. Vienintelis reikalavimas diagnostiniam operatoriui – tai turi būti nemažėjančioji funkcija, įgyjanti reikšmes intervale $[0, 1]$. Pasiūlyta metodika leidžia konstruoti orientuotus į populiacijos ypatumus diagnostinius testus. Šiems tikslams pasiekti reikia realizuoti skaitinius algoritmus diagnostinių operatorių parametrams bei tiriamojo latentinio dydžio p skirstiniui nustatyti.

Literatūra

1. A. Anastasi, S. Urbina. *Psychological Testing* (7th edition). Prentice Hall, 1997.
2. A. Krylovas, N. Kosareva. Žinių tikrinimo matematinis modelis. *Liet. mat. rink. LMD darbai* 48/49:217 – 221, 2008.
3. A. Krylovas, N. Kosareva. Mathematical modelling of forecasting the results of knowlegdge testing. *Technological and Economic Development of Economy*. *Baltic Journal on Sustainability.* 14(3):388–401, 2008.
4. J. Kruopis. *Matematinė statistika*. Mokslo ir enciklopedijų leidykla, Vilnius, 1993.
5. J. Pfanzagl (in cooperation with V. Baumann and H. Huber). *Theory of Measurement.* Physica-Verlag, Wurzburg–Wien, 1971.
6. G. Rasch. *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen, Danish Institute for Educational Research, 1960. Expanded edition: Chicago, The University of Chicago Press, 1980.
7. V. Stakėnas. *Informacijos kodavimas*. Vilnius, VU, 1996.
8. I. Višniakas, K. Slivinskas. *Patikimumo teorija*. Objektų kokybės vertinimo ir patikimumo skaičiavimų metodikos nurodymai. Vilnius, Technika, 2005.

SUMMARY

N. Kosareva, A. Krylovas. A research of mathematical model for diagnostical tests

The technology of creation dichotomous diagnostic operators as general testing tool using the empirical data is proposed. The only requirement to diagnostic operator – it must be nondecreasing function. Probability distribution of test result and test information are calculated according to the proposed technique.

Keywords: item response theory, diagnostic operators, mathematical modelling.