Lietuvos matematikos rinkinys. LMD darbai ISSN 0132-2818
52 tomas, 2011, 195–199 www.mii.lt/LMR/

informacijos matavimo principus. Nekilnojamojo turto kainos modeliavimo pavyzdys

Aleksandras Krylovas[1], Natalja Kosareva[2], Laura Gudelytė[1]

[1] *Mykolo Romerio universitetas, Socialinės informatikos fakultetas*
Ateities 20, LT-08303 Vilnius

[2] *Vilniaus Gedimino technikos universitetas, Fundamentinių mokslų fakultetas*
Saulėtekio 11, LT-10223 Vilnius

E. paštas: krylovas@mruni.eu; natalja.kosareva@vgtu.lt; l.gudelyte@mruni.eu

Santrauka. Šiame darbe pritaikyta dichotominė testavimo metodika, kuria siekiama atlikti gyvenamojo nekilnojamojo turto norminį vertinimą pagal šio turto kainą. Ankstesniuose autorių darbuose buvo pasiūlytas metodas, kuris šiame darbe pritaikomas socialinių indikatorių konstravimui, nagrinėjant konkretų atvejį. Dichotominio testo metodika pagrįstam modeliui kuriama anketa. Pagal tam tikrą kriterijų parenkami anketos klausimai, iš kurių konstruojamas indikatorius. Jis leidžia atlikti Vilniaus miesto mikrorajonuose esančio gyvenamojo nekilnojamojo turto norminį vertinimą pagal kainą atsižvelgiant į informaciją, tiesiogiai susijusią tik su konkrečia vieta, kurioje yra nekilnojamasis turtas.

Raktiniai žodžiai: matematinis modeliavimas, indikatorių konstravimas, testas.

1 Įvadas

Nagrinėjama problematika yra glaudžiai susijusi su dichotominių kintamųjų užduočių sprendimo teorija (Item Response Theory) [7]. Autorių darbuose (žr. [3, 4, 5, 6]) pasiūlytas ir ištirtas indikatorių (testų) konstravimo modelis, kai turint apriorinę informaciją apie tam tikrų testo klausimų funkcijas ir apie populiacijos tiriamojo požymio skirstinį, konstruojamas testas (indikatorius) maksimizuojantis gaunamą informaciją. Modelis yra pagrįstas prielaida, kad indikatorius matuoja tam tikrą latentinį kintamąjį $p \in [0,1]$, o klausimų charakteristinės funkcijos[1] $k(p) : [0,1] \to [0,1]$ yra nemažėjančios. Tačiau socialiniams indikatoriams ši prielaida gali negalioti, nes kintamasis p dažnai būna tik tam tikro susitarimo apie objektų ranginę tvarką rezultatas. Pavyzdžiui, tai gali būti universitetų reitingai, šalies demokratizacijos lygmuo, įmonės vadybos kokybės laipsnis ir pan. Autorių darbe (žr. [6]) buvo išdėstyta bendra šios metodikos pritaikymo(žr. [4, 5, 6]) socialiniams indikatoriams konstruoti schema. Šiame darbe siūlomas indikatorių konstravimo algoritmas, kurį dėstysime nagrinėdami vieną modeliavimo pavyzdį: Vilniaus miesto mikrorajonų pasiskirstymą pagal vidutinę būsto kainą.

[1] T. y. tikimybės teisingai atsakyti į atitinkamą testo klausimą.

2 Nekilnojamojo turto kainos modeliavimo pavyzdys ir pradiniai duomenys

Tarkime, kad baigtinė aibė $A = \{a_1, a_2, \dots, a_n\}$ suskaidyta į nesikertančius blokus (klasterius) (žr. [2]) $K_0, K_1, \dots, K_m \subset A$, $K_i \neq \varnothing$, $K_i \cap K_j = \varnothing$, $\bigcup_{j=0}^{m} K_j = A$. Klasteriai K_j abibrėžia aibėje A dalinę tvarką (žr. [2]): $x < y \Leftrightarrow x \in K_i,\ y \in K_j, i < j$. Žemiau pateiksime Vilniaus miesto mikrorajonų klasterius, sunumeruotus pagal būsto vidutinę ketverių metų (2007–2010) kvadratinio metro kainą. Sudarykime keturis klasterius:
$K_0 = \{$Trakų Vokė (1), Pavilnys (2), Ž. Paneriai (3), Naujininkai (4), Vilkpėdė (5), Dvarčionys (6), A. Paneriai (7), Rasos (8)$\}$;
$K_1 = \{$N. Vilnia (9), Viršuliškės (10), Lazdynai (11), Karoliniškės (12), Pilaitė I (13), Šeškinė (14), Justiniškės (15)$\}$;
$K_2 = \{$Pašilaičiai (16), Fabijoniškės (17), Žirmūnai I (18), Žirmūnai II (19), Baltupiai (20), Santariškės (21), Pilaitė II (22)$\}$;
$K_3 = \{$Šnipiškės (23), Verkiai (24), Naujamiestis (25), Antakalnis (26), Žvėrynas (27), Centras I (28), Centras II (29), Senamiestis (30)$\}$.

Čia miesto rajonų klasteriai sunumeruoti kainos didėjimo tvarka. Pastebėsime, kad klasterizacija gali būti daroma ir pagal kitus kriterijus, kurie nėra tiesiogiai išmatuojami (pvz., gyvenimo kokybė, psichologinis komfortas ir pan. (žr. [8])).

3 Diagnostiniai operatoriai

Apibrėžkime diagnostinio operatoriaus (žr. [4, 5, 6]) diskretųjį analogą. Tarkime, kad turime funkciją $p : A \to \{0, 1\}$, t. y. kiekvienam aibės A elementui a priklausomai nuo testo (indikatoriaus) klausimo, priskiriama reikšmė $p(a)$ – nulis arba vienetas. Funkcija p suskaido aibę A į du klasterius: $A^0 = \{a \in A\colon p(a) = 0\}$ ir $A^1 = \{a \in A\colon p(a) = 1\}$. Pažymėkime C_i klasterį, sudarytą iš pirmųjų $i + 1$ klasterių K_i: $C_i = K_0 \cup K_1 \cup \dots \cup K_i$, $i = 0, 1, 2, \dots, m-1$. Pavyzdžiui, nagrinėjamų Vilniaus mikrorajonų atveju $C_0 = K_0 = \{(1)\text{–}(8)\}$, $C_1 = K_0 \cup K_1 = \{(1)\text{–}(15)\}$, $C_2 = K_0 \cup K_1 \cup K_2 = \{(1)\text{–}(22)\}$, $C_3 = K_0 \cup K_1 \cup K_2 \cup K_3 = \{(1)\text{–}(30)\}$.

Tarkime, kad A – bet kokia baigtinė aibė. Aibės A elementų skaičių žymėsime $|A|$. Sakysime, kad funkcija $p : A \to \{0, 1\}$ yra dichotominis diagnostinis operatorius, kai $\forall i \in \{0, 1, 2, \dots, m-1\}$

$$\left|A^0 \cap C_i\right| + \left|A^1 \cap (A \setminus C_i)\right| \geqslant \left|A^0 \cap (A \setminus C_i)\right| + \left|A^1 \cap C_i\right|. \tag{1}$$

(1) nelygybė turi tokią prasmę: diagnostinio operatoriaus teisingai priskirtų reikšmių 0 ir 1 skaičius turi būti nemažesnis, nei neteisingai priskirtų reikšmių skaičius. Kitaip tariant, elementų $x_i \in K_i$, $x_j \in K_j$, $i < j$, kuriems galioja nelygybė $p(x_i) \leqslant p(x_j)$ yra nemažiau negu elementų, kuriems ši nelygybė negalioja. Tai yra diskretusis analogas tikimybinių reikalavimų [7, 5, 4] diagnostiniams operatoriams, kuris turi validumo prasmę: sudarytas iš diagnostinių operatorių $p(a)$ testas (indikatorius) turi matuoti tai, kam jis sukurtas – mūsų atveju apibrėžti naują klasterizaciją $\widetilde{K}$, kiek įmanoma „artimą" klasterizacijai K. Panagrinėkime diagnostinio operatoriaus pavyzdį. Tarkime, kad x – mikrorajono atstumo nuo miesto centro standartizuota reikšmė. Tada

operatoriai

$$p_\alpha(a;x) = \begin{cases} 1, & \text{jei } x \leqslant \alpha, \\ 0, & \text{jei } x > \alpha, \end{cases} \tag{2}$$

turi atitikti (1) nelygybę. Pastebime, kad egzistuoja tik baigtinis skaičius $n+1$ galimų skirtingų diagnostinių operatorių $p_\alpha(a;x)$, t. y. diagnostinių operatorių, kurie apibrėžia skirtingas aibės A klasterizacijas. Vilniaus mikrorajonams nagrinėdami diagnostinį operatorių, kai standartizuota atstumo nuo centro reikšmė yra palyginama su $\alpha = 0.26$, gauname tokius klasterius[2]:

$K_0 = \{$Trakų Vokė (1), Pavilnys (0.44), Ž. Paneriai (0.36), Naujininkai (0.21), Vilkpėdė (0.16), Dvarčionys (0.54), A. Paneriai (0.73), Rasos (0.28)$\}$;

$K_1 = \{$N. Vilnia (0.64), Viršuliškės (0.26), Lazdynai (0.36), Karoliniškės (0.27), Pilaitė I (0.43), Šeškinė (0.21), Justiniškės (0.36)$\}$;

$K_2 = \{$Pašilaičiai (0.42), Fabijoniškės (0.36), Žirmūnai I (0.28), Žirmūnai II (0.16), Baltupiai (0.32), Santariškės (0.6), Pilaitė II (0.44)$\}$;

$K_3 = \{$Šnipiškės (0.13), Verkiai (0.88), Naujamiestis (0.10), Antakalnis (0.27), Žvėrynas (0.11), Centras I (0.04), Centras II (0.00), Senamiestis (0.10)$\}$.

Atrinkę narius, iš klasterio $C_0 = K_0$, kurių atstumo nuo centro standartizuota reikšmė yra didesnė, nei 0.26 gauname $|A^0 \cap C_0| = 6$, o visų likusių narių, netenkinančių šios sąlygos, K_0 klasteryje yra $|A^1 \cap C_0| = 2$. Tikriname likusius K_1, K_2, K_3 klasterius, sudarančius aibę $A \setminus C_0$ ir gauname, kad mikrorajonų, kurių atstumas nuo centro yra didesnis, nei 0.26, yra $|A^0 \cap (A \setminus C_0)| = 13$, o likusių yra $|A^1 \cap (A \setminus C_0)| = 9$. Mūsų atveju, kai $i = 0$, visai klasterizacijai pritaikę (1) formulę, gauname: $6 + 9 = 13 + 2$. Matome, kad pasirinkta funkcija $\mathbf{1}_{\{x \geqslant 0.26\}}$ tenkina (1) formulę. Likusiems C_1, C_2, C_3 formulė (1) tikrinama analogiškai.

4 Indikatoriaus skaitinės charakteristikos

Diagnostinių operatorių rinkinį vadiname indikatoriumi $I(A) = (p_1(a), p_2(a), \ldots, p_m(a))$. Indikatorius $I(A)$ apibrėžia naują aibės A klasterizaciją[3] $\tilde{K} = \{\tilde{K}_0, \tilde{K}_1, \ldots, \tilde{K}_m\}$:

$$\tilde{K}_j = \left\{ a \in A\colon \sum_{i=1}^{m} p_i(a) = j \right\}. \tag{3}$$

Pažymėkime aibės A poaibius $D_j = \{a \in A\colon a \in K_i \cap \tilde{K}_l,\ |i-l| = j\}$, $j = 0, 1, \ldots, m$. Skaičius $|D_0|$ rodo, kiek aibės A elementų priklauso klasteriams su tais pačiais numeriais, t. y. $0 \leqslant \frac{|D_0|}{|A|} \leqslant 1$ yra naujos klasterizacijos suderinamumo su sena matas. Kai $\frac{|D_0|}{|A|} < 1$, svarbu žinoti nesuderinamumo laipsnį. Autorių darbe (žr. [6]) pasiūlytas toks kriterijus

$$S\bigl(I(A), K\bigr) = \sum_{j=1}^{m} j \cdot |D_j|. \tag{4}$$

[2] Skliausteliuose nurodytas sunormuotas atstumas nuo centro.

[3] Nerašysime į rinkinį $\tilde{K}$ klasterių, kurie yra tuščios aibės.

Pavyzdžiui, paimkime tris diagnostinius operatorius $p_\alpha(a,x)$, $p_\beta(a,y)$, $p_\gamma(a,z)$, čia x – rajono atstumo nuo miesto centro, y – darbo vietų tankio, z – gatvių tankio standartizuotos reikšmės. Taikant iš šių operatorių sukonstruotą indikatorių sudaroma mikrorajonų (jie sunumeruoti) klasterizacija pagal tris kintamuosius, esant $\alpha = 0.26$, $\beta = 0.04$, $\gamma = 0.15$:
$\tilde{K}_0$ = {Pilaitė II, Pavilnys, A.Paneriai, Trakų Vokė, Lazdynai, Pilaitė I, N. Vilnia, Fabijoniškės, Santariškės};
$\tilde{K}_1$ = {Ž. Paneriai, Rasos, Karoliniškės, Šeškinė, Justiniškės, Antakalnis, Pašilaičiai, Baltupiai, Dvarčionys, Verkiai};
$\tilde{K}_2$ = {Viršuliškės, Žirmūnai I};
$\tilde{K}_3$ = {Centras I, Centras II, Naujininkai, Naujamiestis, Vilkpėdė, Senamiestis, Žvėrynas, Žirmūnai II, Šnipiškės}.

Todėl pagal aprašytą metodiką [4] sudaromi klasterizacijų neatitikimai: $D_j = \{a \in A\colon a \in K_i \cap \tilde{K}_l,\ |i-l| = j\}$, $j = 1,\dots,k-1$; čia $A = \{1,2,\dots,30\}$ – mikrorajonų aibė, $\tilde{K}$ ir K yra dvi aukščiau aprašytos klasterizacijos. Idealiuoju atveju indikatorius $I(A)$ apibrėžia identiškus suskirstymus (t. y. tą pačią klasterizaciją) pagal tris testo klausimus ($\tilde{K}$) ir kainą (K), todėl gauname, kad $D_0 = A; D_1 = D_2 = \dots = D_{k-1} = \varnothing$.

Lyginame seną klasterizaciją pagal kainą ir naują klasterizaciją pagal trijų funkcijų reikšmių sumą (3). Dydis $|D_0|$ parodo, kiek yra mikrorajonų, patenkančių į tuos pačius klasterius, $|D_1|$ – kiek patenka į gretimus klasterius, $|D_2|$ – kiek į dar toliau vienas nuo kito esančius klasterius ir t. t. Klasterizacijų nesuderinamumo laipsnį išmatuojame funkcija (4). Nagrinėjamo pavyzdžio atveju gauname, kad $S(I(A),K) = 1\cdot 10 + 2\cdot 5 + 3\cdot 2 = 26$.

Taigi geriausią indikatorių $I_0(A)$ gausime, kai

$$S\bigl(I_0(A),K\bigr) = \min_{\{p_1,p_2,\dots,p_m\}} S\bigl(\{p_1,\dots,p_m\},K\bigr). \tag{5}$$

Idealiuoju atveju

$$S\bigl(I_0(A),K\bigr) = 0 \quad \text{ir} \quad \frac{|D_0|}{|A|} = 1.$$

Pastebime, kad operatoriai $p_j(a)$ yra (2) pavidalo. T. y., kai pasirenkamos išmatuojamų standartizuodų dydžių tam tikros reikšmės, turime (5) kombinatorinį uždavinį. Kol kas neaišku ar jis gali būti priskirtas NP uždavinių klasei (žr. [2]) ir ar galima ieškoti (5) uždavinio minimumo nuosekliai pagal kiekvieną p_j, t. y. ar tinka godžioji strategija (žr. [1]). Tai yra mūsų tolimesnių tyrimų objektas.

5 Išvados

Šioje publikacijoje išnagrinėtas dichotominio diagnostinio testo matematinis modelis naudojant duomenis, netiesiogiai nusakančius agreguotą gyvenamojo ploto kainą Vilniaus mieste. Pasiūlyta metodika leidžia atlikti gyvenamųjų rajonų norminį vertinimą latentinio kintamojo – nekilnojamojo turto kainos – atžvilgiu, kai vertinimui naudojami kiti tiesiogiai stebimi ir išmatuojami kintamieji. Šių kintamųjų parinkimo kriterijus yra funkcijos (5), kuri parodo dviejų klasterizacijų neatitikimą, minimizavimas. Šiame straipsnyje aptartas metodas gali būdas taikomas ir kitos rūšies nekilnojamojo turto norminiam vertinimui.

Literatūra

[1] R. Čiegis. *Duomenų struktūros, algoritmai ir jų analizė.* Technika, Vilnius, 2007.

[2] A. Krylovas. *Diskrečioji matematika.* Technika, Vilnius, 2009.

[3] A. Krylovas and N. Kosareva. Mathematical modelling of forecasting the results of knowledge testing. *Technological and Economic Development of Economy*, **14**(3):388–401, 2008.

[4] A. Krylovas and N. Kosareva. Mathematical modelling of diagnostic tests, knowledge-based technologies and or methodologies for strategic decisions of sustainable development. In (*KORSD*-2009): *5th International Conference*, pp. 120–125, 2009.

[5] A. Krylovas and N. Kosareva. Politominio diagnostinio testo matematinis modelis. *Liet. mat. rink. LMD darbai*, **51**:279–284, 2010.

[6] A. Krylovas and N. Kosareva. Socialinių indikatorių kosntravimas. In *Social Technologies'*10*: Challenges, Opportunities, Solutions. Conference Proceedings*, pp. 48–55. Vilnius, November 25–26, 2010.

[7] 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

[8] V. Rudzkienė ir M. Burinskienė. Darnaus vystymosi principai vilniaus miesto plėtroje. *Darnaus vystymose strategija ir praktika: mokslo darbai*, pp. 39–46, Mykolo Romerio universitetas, Vilnius, 2006.

SUMMARY

Construction of social indicators using information measuring principles. Case study of real estate prices simulation model

A. Krylovas, N. Kosareva, L. Gudelytė

The article analyses the data concerning the real estate market of last years in Vilnius. The approach of dichotomous diagnostic operators as general testing tool creation from the empirical data is used and deterministic clusterization of catchments is provided. Based on this analysis, questionnaire concerning the information about catchment is constructed for the evaluation of aggregated appartment price.

Keywords: mathematical modeling, creation of indicators, test, statistical methods.