

MAHALLALAR MA'LUMOTLAR BAZASINI MODELLARINI ISHLAB CHIQUISHDA VA ISHINI YENGILLASHTIRISH MAQSADIDA SINFLARNI BIRLASHTIRISH ALGORITMLARIDAN FOYDALANISH

Raximboev Hikmat Jumanazarovich ⁽¹⁾ - Muhammad al-Xorazmiy nomidagi
Toshkent axborot texnologiyalari universiteti Urganch filiali "Axborot
texnologiyalari" kafedrasida dotsenti

hikmatbek75@yandex.com

Jumaniyozov Boburbek Umidbek o'g'li ⁽¹⁾ – Muhammad al-Xorazmiy
nomidagi Toshkent axborot texnologiyalari universiteti Urganch filiali
"Kompyuter injiniringi" kafedrasida M911-21 guruhi magistri

jumaniyozovboburbek170@gmail.com

Sapayev Sanjar G'ayrat o'g'li ⁽¹⁾ – Muhammad al-Xorazmiy nomidagi
Toshkent axborot texnologiyalari universiteti Urganch filiali "Kompyuter
injiniringi" kafedrasida M911-21 guruhi magistri

sapayevsanjar1994@gmail.com

Annotatsiya: Ushbu maqolada mutaxassislar mahallalar faoliyatini mahallaning bandlik ma'lumotlari asosida baholashda qulaylik yaratish maqsadida mahallalarni bandlik ko'rsatkichlari bo'yicha sinflarga birlashtirish masalasi muhokama qilinadi. Sinflarga integratsiya qilish uchun sinflarni integratsiyalash muammosini hal qilishda foydalaniladigan Fuzzy k-Means, K-means kabi mashinani o'rganish algoritmlari eslatib o'tiladi. Bu algoritmlar yordamida mahallalarni uchga (A'lo, Yaxshi va qoniqarli) guruhlash masalasi hal qilinadi. KNIME Analytics platformasida eksperimental ish olib borildi. Olingan natijalar ushbu algoritmlar uchun taqqoslandi va xulosalar chiqarildi.

Kalit so'zlar: ekspert, model, baholash, yumshoq k-O'rtacha, k-O'rtacha, xato ob'ekti.

Abstract: This article discusses the issue of uniting neighborhoods into classes based on employment indicators in order to make it easier for experts to evaluate the activity of neighborhoods based on the employment data of the neighborhood. For class integration, machine learning algorithms such as Fuzzy k-Means, K-means, which are used to solve the class integration problem, are mentioned. With the help of these algorithms, the problem of grouping neighborhoods into three (Excellent, Good and Satisfactory) is solved. Experimental work was carried out on the KNIME Analytics platform. The obtained results were compared for these algorithms and conclusions were drawn.

Keywords: expert, model, estimation, smooth k-Mean, k-Mean, error object

Kirish: Mahallaga oid ma'lumotlarni tahlil qilganda , asosan, aholining turmush sharoiti, moddiy va ma'naviy holatini belgilovchi ko'plab ma'lumotlar mavjud. Aholining turmush darajasi deganda odamlarning zarur moddiy va ma'naviy ne'matlar bilan ta'minlanganlik darajasi, ularning iste'moli va ehtiyojlarini qondirish darajasi tushuniladi. [1][2]

Aholining turmush darajasini ularning moddiy, ma'naviy va ijtimoiy ehtiyojlarini qondirish darajasi deb hisoblash mumkin . Bu boradagi muhim ko'rsatkichlardan biri tuman aholisining bandlik darajasidir. Tumanlar bo'yicha ma'lumotlarni yig'ishda bandlik haqidagi bo'sh ma'lumotlar ko'pincha topiladi, ammo dastlabki ishlov berish paytida ma'lumotlar o'chirib tashlansa, bu axborot ma'lumotlarining yo'qolishiga olib kelishi mumkin. [3][4] bu modelni yaratish jarayonida modelning muvofiqligini tekshirishda xatolar sonining ko'payishiga olib keladi.

mahalla faoliyatini mavjud bandlik ko'rsatkichlaridan kelib chiqib baholashda bir qator qiyinchiliklarga duch kelmoqda. Tumanlarni bandlik ko'rsatkichi bo'yicha sinflarga birlashtirish orqali guruhlashni amalga oshirish ekspertga baholash jarayonini osonlashtirishga yordam beradi. Mavjud ma'lumotlarga asoslangan sinflarga integratsiyalash uchun mashinani o'rganish usullari va algoritmlari muammoni hal qilishga yordam beradi. [5][6]

Savolning bayonoti. Xorazm viloyatidagi mahallalar fuqarolarining bandligi to'g'risidagi ma'lumotlardan foydalangan holda (1-jadval) har bir mahallani obyekt sifatida hisobga olgan holda ma'lumotlar to'plami ishlab chiqiladi .

1-jadval. Tumanlar bo'yicha demografik ko'rsatkichlar va tarmoqlar bo'yicha bandlik ko'rsatkichlari

1-jadval

Mahalla	Aholi soni	Oilalar soni	Pensionerlar soni	18 yoshgacha bo'lganlar soni	Bolalar bog'chasi xodimlari soni	To'liq vaqtda ishlaydiganlar soni	Tadbirkorlik bilan shug'ullanuvchilar	Ishsizlar soni
Uslar	2371	649	194	726	182	486	149	189
Yuksalish	3314	667	314	981	248	617	164	208
Namuna	2857	797	356	1208	245	64	129	138
Gulzor	4012	915	283	1374	150	423	917	270
Bog'zor	3723	745	389	1344	205	0	144	243
O'zbekiston	4861	1203	124	1754	386	811	202	215

Navbaxor	6347	1945 yil	389	2443	432	281	328	234
----------	------	-------------	-----	------	-----	-----	-----	-----

O'zgaruvchan tanlama uchun hosilalarni sinflarga birlashtirish (klasterlash) muammosi hal qilindi . Klasterlash muammosining matematik tavsifi quyidagicha ko'rinadi:

Menga ma'lumotlar ob'ektlari to'plami beriladi , ularning har biri atributlar to'plami bilan ifodalanadi.

$$i_j = \{x_1, x_2, \dots, x_h, \dots, x_m\} \quad (1.1)$$

qayerda i_j - o'rganilayotgan ob'ekt

Evklid masofasi. Ba'zan bir-biridan uzoqda joylashgan ob'ektlarga ko'proq og'irlik berish uchun standart Evklid masofasini kvadratga solish vasvasasi bo'lishi mumkin. Bu masofa quyidagicha hisoblanadi va to'plamlarga bo'linish yo'li bilan Ekspertning chiqishini olish uchun ishlatiladi.

$$d_2(x_i, x_j) = \sqrt{\sum_{t=1}^m (x_{it} - x_{jt})^2} \quad (1.2)$$

Oldindan belgilangan klasterlar soniga asoslangan mavjud metrikaga asoslangan ob'ektlarni guruhlariga guruhlash klasterlash muammosi deb ataladi.

Muammoni hal qilishning taklif qilingan usuli. Mashinani o'rganishda sinflarga birlashtirish (klasterlash) uchun bir nechta algoritmlar mavjud. Ular orasida hozirda quyidagilar juda keng qo'llaniladi

Fuzzy K-means algoritmi

Klasterlar endi loyqa to'plamlardir va har bir nuqta sinchkovlik bilan turli darajadagi turli klasterlarga tegishli. Nuqta ma'lum bir klasterga tegishli bo'lgan maksimal mezon bo'yicha ma'lum bir klasterga tegishli. Bu holatda asosiy tushunchalar:

Trening to'plami $M = \{m_j\}_{j=1}^d$, d ma'lumotlar nuqtalari soni (vektorlar);

• metrik masofa (2)

• Klaster markazlarining vektori $C = \{c^{(i)}\}_{i=1}^c$, keyin

$$c^{(i)} = \frac{\sum_{j=1}^d (u_{ij})^w m_j}{\sum_{j=1}^d (u_{ij})^w}, \quad 1 \leq i \leq c \quad (1.3)$$

Matritsani ajratish $U = \{u_{ij}\}$, keyin :

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_A^2(m_j, c^{(i)})}{d_A^2(m_j, c^{(k)})} \right)^{\frac{1}{w-1}}} \quad (1.4)$$

Maqsad funktsiyasi:

$$J(M, U, C) = \sum_{i=1}^c \sum_{j=1}^d u_{ij}^w d_A^2(m_j, c^{(i)}) \quad (1.5)$$

Bu erda $w \in (1, \infty)$ noaniqlik ko'rsatkichi (so'nish koeffitsienti), parchalanish loyqaligi. Odatda $w = 2$ ishlatiladi, cheklovlar to'plami:

$$u_{ij} \in [0,1]; \sum_{i=1}^c u_{ij} = 1; 0 < \sum_{j=1}^d u_{ij} < d, \quad (1.6)$$

har bir ma'lumotlar vektorining turli darajadagi alohida klasterlarga tegishli ekanligi klasterlar tegishli bo'lgan barcha ma'lumotlar elementi bo'shliqlarining yig'indisi bitta ekanligini aniqlaydi. Strukturaviy ravishda, algoritm quyidagi puflash shaklining iterativ protsedurasidir.

1-qadam. $2 \leq c \leq d$ klasterlar sonini tanlang

2-qadam. Ma'lumotlar vektorlarini ifodalash uchun skalar metrik.

3-qadam. To'xtatish variantini d tanlang.

Qadam 4. Loyqalik omilini tanlang $w \in (1, \infty)$, masalan, $w = 2$.

Qadam 5. Bo'linish matritsasini ishga tushiring (masalan, tasodifiy qiymatlar).

6-qadam. Formula yordamida klasterlarning prototiplarini (markazlarini) hisoblang:

$$c_l^{(i)} = \frac{\sum_{j=1}^d (u_{ij}^{(l-1)})^w m_j}{\sum_{j=1}^d (u_{ij}^{(l-1)})^w}, \quad 1 \leq i \leq c \quad (1.7)$$

7-qadam. Barcha ma'lumotlar elementlari uchun formuladan foydalanib, barcha (markaziy) klasterlargacha bo'lgan masofalarning kvadratlarini hisoblang:

$$d_A^2(m_j, c_l^{(i)}) = (c_l^{(i)} - m_j)^t A((c_l^{(i)} - m_j)). \quad (1.8)$$

8-qadam. Quyidagi formuladan foydalanib, bo'linish matritsasini yangilang:

$$u_{ij}^{(l)} = \frac{1}{\sum_{k=1}^c \left(\frac{d_A^2(m_j, c^{(k)})}{d_A^2(m_j, c^{(i)})} \right)^{\frac{1}{w-1}}}, \quad 1 \leq i \leq c, 1 \leq j \leq d, \quad (1.9)$$

k -Means algoritmi

yanada murakkab algoritmlarga kiritilgan prototipni yaxshiroq tushunishga yordam beradi. Ushbu algoritm doirasidagi asosiy ta'riflar va tushunchalar quyidagi shaklga ega:

- Trening paketi $M = \{m_j\}_{j=1}^d$, d ma'lumotlar nuqtalari soni (vektorlar);
- Formula bo'yicha hisoblangan masofa ko'rsatkichi;
- Klaster markazlarining vektori $C = \{c^{(i)}\}_{i=1}^c$,

$$C^{(i)} = \sum_{j=1}^d u_{ij} m_j, \quad 1 \leq i \leq c; \quad (2.1)$$

Matritsalarini ajratish $U = \{u_{ij}\}$, keyin

$$u_{ij}^{(l)} \left\{ \begin{array}{l} 1 \text{ } d(m_j, c_i^{(l)}) = \min_{1 < k < c} d(m_j, c_k^{(l)}) \\ 0 \text{ qolgan hollarda} \end{array} \right\}$$

(2.2)

Maqsad funktsiyasi:

$$J(M, U, C) = \sum_{i=1}^c \sum_{j=1}^d u_{ij} d_A^2(m_j, c^{(i)})$$

(2.3)

• cheklovlar to'plami

$$u_{ij} \in \{0, 1\}; \quad \sum_{i=1}^c u_{ij} = 1; \quad 0 < \sum_{j=1}^d u_{ij} < d$$

(2.4)

bu har bir ma'lumot vektori boshqalarga emas, faqat bitta klasterga tegishli bo'lishi mumkinligini aniqlaydi. Har bir klasterda kamida bitta nuqta bor, lekin umumiy nuqtalar kamroq. Strukturaviy ravishda, algoritm quyidagi puflash shaklining iterativ protsedurasidir.

Qadam 1. Dastlabki bo'linishdan boshlang (masalan, tasodifiy), aniq d ni tanlang (Ma algoritmining tugallanish holatida ishlatiladi), $l=0$ takrorlash raqamini bajaring.

2-qadam. Formula yordamida klaster markazlarini aniqlang:

$$c_l^{(i)} = \frac{\sum_{j=1}^d u_{ij}^{(l-1)} m_j}{\sum_{j=1}^d u_{ij}^{(l-1)}}, \quad 1 \leq i \leq c. \quad (2.5)$$

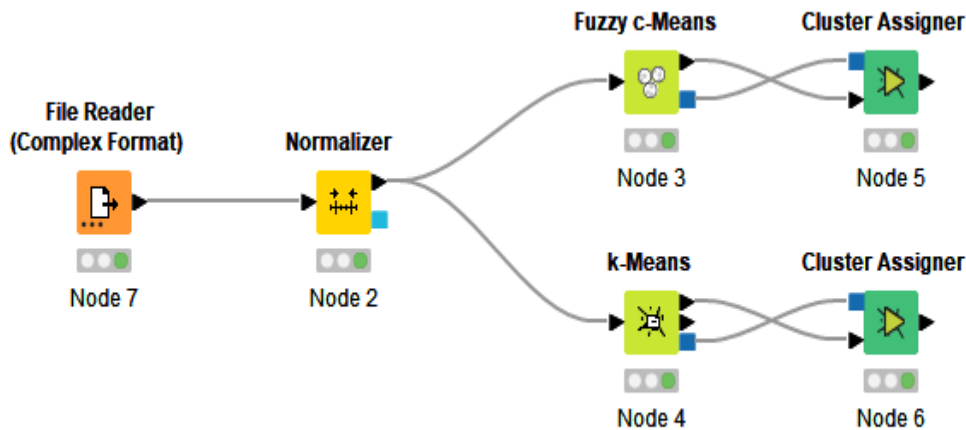
Qadam 3. Formuladan foydalanib xato kvadratlarini minimallashtirish uchun bo'linish matritsasini yangilang

$$u_{ij}^{(l)} \left\{ \begin{array}{l} 1 \text{ } d(m_j, c_i^{(l)}) = \min_{1 < k < c} d(m_j, c_k^{(l)}) \\ 0 \text{ qolgan hollarda} \end{array} \right\}$$

(2.6)

4-qadam. Shartni tekshiring $\|U^{(l)} - U^{(l-1)}\| < d$

Ushbu ma'lumotlar va algoritmlar yordamida ANIME Analytics Platform[6] dasturidan foydalangan holda Xorazm viloyatidagi besh yuzdan ortiq mahallalar klasterlashtirilib, uch guruhga birlashtiriladi. Buning uchun KNIME Analytics platformasi dasturiy ta'minotidan foydalanib, quyidagi ketma-ketliklar asosida hisob-kitoblar amalga oshiriladi :[9][10][12]



Shkl 1. KNIME Analytics platformasini qo'llashda joriy muammoni hal qilish uchun ish jarayoni ishlab chiqilgan.

1-rasmdagi Ish jarayoniga ko'ra, algoritmnning hisoblash jarayonlari quyidagicha amalga oshiriladi .

Finereader : Ma'lumotlar ushbu komponent orqali yuklanadi. Yuklangan ma'lumotlar shunday ko'rinadi.

Row ID	Col0	Col1	Col2	Col3	Col4	Col5	Col6	Col7
Row0	2371	649	194	726	182	486	149	312
Row1	3314	667	314	981	248	617	164	421
Row2	2857	797	356	1208	245	64	129	83
Row3	4012	915	283	1374	150	423	917	489
Row4	3723	745	389	1344	205	0	144	147
Row5	4861	1203	124	1754	386	811	202	316
Row6	6347	1945	389	2443	432	281	328	285
Row7	4114	1039	397	1402	29	786	365	342
Row8	3386	935	283	1010	115	481	414	280
Row9	3305	925	334	926	23	781	298	345
Row10	3468	874	523	1285	29	102	124	275
Row11	2730	830	204	874	61	468	621	315
Row12	1902	530	165	296	233	342	358	284
Row13	3045	848	397	453	276	321	218	327

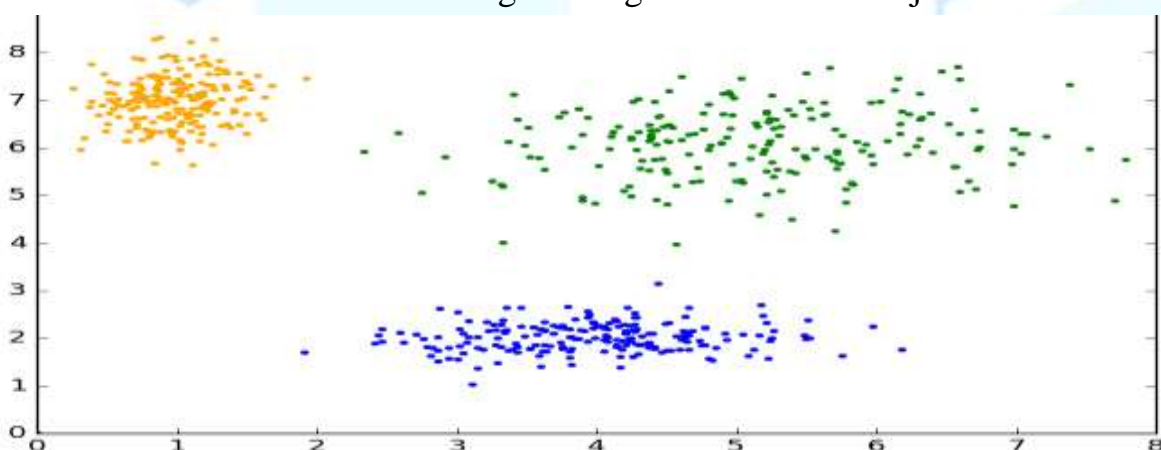
Mahalla ma'lumotlari asosida yaratilgan ma'lumotlarni yuklab olish .

2. Normalizator: bu komponent orqali bizda mavjud bo'lgan ma'lumotlar normallashtiriladi (min max normalizatsiya orqali [0 ..1] intervalga o'tkaziladi. Bu jarayon barcha belgilarga bir xil vazn berish uchun kerak.)

3. Fuzzy k – Means ushbu komponentda keltirilgan ma'lumotlar to'plami uchun yuqorida aytib o'tilgan Fuzzy k-Means algoritmi asosida guruhlashni amalga oshiradi. Ushbu ma'lumotlar to'plami uch guruhga bo'lingan : bandlik ko'rsatkichlari a'lo darajadagi mahallalar , yaxshi ko'rsatkichlarga ega bo'lgan mahallalar va qoniqarli ko'rsatkichlarga ega bo'lgan mahallalar . Algoritmdan foydalangandan keyin mahallalar Ular uchta sinfga birlashtirilgan , buni quyidagi rasmda ko'rish mumkin.

Row ID	D_Col0	D_Col1	D_Col2	D_Col3	D_Col4	D_Col5	D_Col6	D_Col7	D_Col8	D_cluster_0	D_cluster_1	D_cluster_2	S_Winner ...
Row0	0.133	0.182	0.153	0.164	0.392	0.295	0.037	0.256	0.213				cluster_0
Row1	0.266	0.191	0.268	0.26	0.552	0.374	0.045	0.345	0.234				cluster_2
Row2	0.201	0.255	0.309	0.345	0.545	0.039	0.028	0.068	0.156				cluster_2
Row3	0.364	0.314	0.238	0.407	0.314	0.256	0.372	0.4	0.304				cluster_2
Row4	0.323	0.229	0.34	0.396	0.488	0	0.035	0.12	0.274				cluster_2
Row5	0.483	0.456	0.086	0.55	0.888	0.492	0.06	0.259	0.242				cluster_1
Row6	0.693	0.824	0.34	0.808	1	0.17	0.115	0.233	0.264				cluster_1
Row7	0.378	0.375	0.348	0.438	0.019	0.476	0.131	0.28	0.153				cluster_2
Row8	0.276	0.323	0.238	0.271	0.229	0.292	0.152	0.229	0.245				cluster_0
Row9	0.264	0.318	0.288	0.239	0.095	0.473	0.102	0.283	0.163				cluster_0
Row10	0.287	0.293	0.469	0.374	0.019	0.062	0.026	0.225	0.287				cluster_0
Row11	0.183	0.271	0.163	0.22	0.097	0.284	0.243	0.258	0.172				cluster_0
Row12	0.067	0.123	0.125	0.003	0.516	0.207	0.128	0.233	0.139				cluster_0
Row13	0.228	0.28	0.548	0.062	0.62	0.195	0.067	0.268	0.16				cluster_2
Row14	0.358	0.424	0.22	0.356	0.431	0.494	0.125	0.186	0.297				cluster_2
Row15	0.351	0.376	0.605	0.23	0.513	0.336	0.086	0.274	0.237				cluster_2
Row16	0.301	0.376	0.344	0.247	0.594	0.113	0.086	0.342	0.256				cluster_2
Row17	0.248	0.235	0.238	0.055	0.467	0.473	0.049	0.12	0.165				cluster_0
Row18	0.151	0.195	0.152	0.101	0.066	0.352	0.226	0.14	0.139				cluster_0
Row19	0.42	0.421	0.194	0.192	0.516	0.473	0.353	0.07	0.151				cluster_2
Row20	0.374	0.429	0.434	0.313	0.348	0.076	0.149	0.251	0.21				cluster_2
Row21	0.479	0.555	0.768	0.608	0.509	0.153	0.115	0.335	0.173				cluster_1
Row22	0.258	0.305	0.046	0.257	0.423	0.159	0.106	0.186	0.154				cluster_0
Row23	0.204	0.23	0.238	0.088	0.399	0.231	0.038	0.243	0.151				cluster_0
Row24	0.3	0.343	0.188	0.425	0.068	0.473	0.354	0.206	0.17				cluster_2

rasm.Mahallalarni guruhlarga birlashtirish natijalari



Shakl 5. Fuzzy k dan keyin ma'lumotlar to'plamining ko'rinishi - Bajarish vositalari.

4. k – vositalar. Ushbu komponent yuqorida aytib o'tilgan k-Means algoritmi yordamida

Xulosa

Xorazm viloyatidagi barcha mahallalarni ekspert bahosidan oldin olingan ma'lumotlar asosida aholi turmush darajasini baholash algoritmlaridan foydalangan holda guruhlarga birlashtirish ekspert baholash jarayonida vaqtni tejash va baholash sifatini oshirish imkonini beradi. Xususan, bu jarayonda Fuzzy k-Means algoritmidan foydalanish ancha samarali bo'lishi mumkin.

Ishlatilgan adabiyotlar

[1]. Belginskaya O. Mezonlar tizimi, munitsipalitetning ijtimoiy-iqtisodiy rivojlanish ko'rsatkichlari // Munitsipal hokimiyat. - 2007. Yo'q . 6. - p . 35-42
 [2]. Belkina TD Ko'rsatkichlar tizimi yordamida shahar rivojlanishining diagnostikasi // Prognozlash muammolari. - 2007. - No 2. - b. 77 - 89.
 [3]. “ Rahimboev H.J., Jabborov D.Z. Mahalliy ijro etuvchi hokimiyat organlari tomonidan axborotni boshqarish tizimini qurish xususiyatlari // Radiotexnika,

telekommunikatsiya va ahlat tahnologiasi : muammolar va rivojlanish istiqbollari, Xalqaro ilmiy-texnikaviy konferentsiya to'plami, »mi. II jild Toshkent, 2015 yil 25-22 may

[4]. Sergeev Aleksey Viktorovich . Aholining turmush darajasi ko'rsatkichlarini modellashtirish va prognozlash usullarini ishlab chiqish va tadqiq qilish: birinchi Samara viloyatida: iqtisod fanlari nomzodi dissertatsiyasi: 08.00.13 / Samara, 2011 .- 188 b.: kasal .

[5]. N.I.Rustamov . _ O'zbekistonda turmush darajasi va uni yaxshilash imkoniyati. "Iqtisodiyot va innovatsion texnologiyalar" ilmiy elektron jurnali. 2015 yil sentyabr-oktyabr, 5-son.

[6]. Golovina.AN 2018 Mahalliy boshqaruv organlarida boshqaruv qarorlarini qabul qilish va amalga oshirish texnologiyasi Kamchiliklarni bartaraf etish yo'llari Ilmiy va amaliy elektron jurnal Alley of Science 1 (17).