

2-TOM, 11-SON

**BERILGANLARDAGI O'TKAZIB YUBORILGAN QIYMATLAR
MAVJUDLIGIDA SINFLAR OBYEKTлари UCHUN USTUNLIK
INTERVALLARINI QURISH**

Saydullayev Xumoyiddin Shuxratovich

Maktab, Magitr, O'qituvchi

Humoyiddin6737196@gmail.com

Tel: +998919077196

Annotatsiya: Sinf ob'ektlarini tavsiflashda ma'lumotlarda bo'shliqlar mavjudligini hisobga olgan holda har xil turdagi xususiyatlarning axborot to'plamlarini tanlash ko'rib chiqiladi. Tanlash sharti - bu ularning o'lchovlari miqyosiga mos kelmaydigan ma'lumotlarning xususiyatidir. O'zgaruvchanlikka xususiyat qiymatlarini ajratilgan intervallarga bo'lish usullarini qo'llash orqali erishiladi. Intervallarga bo'linish ma'lumotlarni qayta ishlashning ikki usulida qo'llaniladi. Oldindan ishlov berish natijalariga ko'ra ketma-ketliklar xususiyatlarning barqarorligi bilan bog'liq holda va juft xususiyatlar qiymatlarining sinflararo farqlariga nisbatan shakllanadi.

Kalit so'zlar: Intellektual, klassifikatsiya, miqdoriy, nominal

Kirish:

Zaif tuzilmaga ega bo'lgan fan sohalarida berilganlarni yig'ish, ularga ishlov berish va ulardan yashirin qonuniyatlarni izlashdagi eng muhim muamollardan biri bu – obyekt tasnifidagi berilganlarning to'liq emasligidir. Hamda, alomatlarni sintez qilish yangi alomatlar fazosini dastlabki fazodan olish NP to'liq murakkablikdagi masala hisoblanadi. Bu esa sun'iy intellekt usullarda evristik usullarni qo'llashni taqozo etadi. Evristik usullarni qo'llashdan maqsad bu obyekt tasnifidagi berilganlarda bo'sh qiymatlarning mavjudligini hisobga olgan holda sun'iy intellektning algoritmlarini ishlatishdir. Berilganlarni tavsiflashing eng keng tarqalgan usullaridan biri bu “ob'ekt - alomat” ko'rinishidagi jadvallardir. Ushbu dissertatsiyada tadqiqot mavzusi sifatida bo'shliqlarni o'z ichiga olgan “ob'ekt - alomat” jadvallarini ko'rib chiqamiz. Informativ alomatlarni tanlash sinflar ichidagi ob'ektlar va sinflar orasidagi munosabatlarning o'zgarishi bilan bog'liq. Hisoblash algoritmini amalga oshirishda ob'ektlar orasidagi munosabatning noaniqligi



2-TOM, 11-SON

alamatlar soni ko'pligiga bog'liq va bu muammo "o'lchov-la'nati" deb ataladi. Munosabatlarning turli tuzilmalarini tahlil qilish vositasi sifatida sinflarning kompaktligi o'lchovidan va umumiy namunadan foydalanish taklif qilingan¹. Sinflarning ixchamligi o'lchovlari qiymatlari yashirin xususiyatni shakllantirishning bir qator mezonlari uchun asos bo'lib xizmat qiladi

Asosiy qism:

Deterministik mezonlar asosida sonli alomatlarning (berilgan, latent) o'zaro kesishmaydigan intervallarga ajratishning ikki usuli ma'lum [4,5]. Mazkur usullar algoritmlari o'lchov masshtablariga invariant va quyidagi hollar uchun ishlatiladi:

- intuitiv yechimlar qabul qilish jarayonini modellashtirishda berilganlar bazasidan latent (oshkor, bevosita o'lchash mumkin bo'lmagan) alomatlarni qidirishda;
- sonli alomatlardan nominal alomatlarni shakllantirishda yo'qotiladigan axborotning minimal bo'lishini ta'minlashda;
- turli toifadagi alomatlardan informativ to'plamlarni tanlashda.

Mezonlar talqini. Ikki o'zaro kesishmaydigan K_1, K_2 sinflarga ajratilgan obyektlarning mumkin bo'lgan to'plami $E_0 = \{S_1, \dots, S_m\}$ berilgan bo'lsin. Har bir obyekt n ta turli toifadagi alomatlar $X(n) = (x_1, \dots, x_n)$ asosida ifodalanadi, δ ($\delta > 0$) tasi interval shkalada, qolgan, $n - \delta$ tasi nominal shkalalarda o'lchanadi. $X(n)$ dan olingan alomatlar $Y(\mu) = (y_1, \dots, y_\mu)$ sonli alomatlarga akslantiruvchi operator mavjud bo'lsin va uning elementlari ichida $X(n)$ dan olingan δ sondagi latent alomatlar bor bo'lsin. Latent (sonli) alomatlarga misol tariqasida $x_i x_j, x_i x_j^{-1}$ kombinatsiyalar, hamda sonli va nominal alomatlardan olingan umumlashgan ko'rsatkichlarni ko'rsatib o'tish mumkin [5].

E_0 tanlanmadagi $Y(\mu)$ to'plam ostisidan olingan alomatlar qiymatlarini kesishmaydigan intervallarga ajratishning ikki mezoni aniqlangan bo'lsin. Birinchi mezon sinflar soniga teng intervallar soniga amal qilishga asoslangan. Biz qarayotgan holda bu son ikkiga teng.

Har bir $y_j \in Y(\mu)$ alomatning mezonga mos optimal ajratish quyidagicha amalga oshiriladi. Alomatning tartiblangan qiymatlar to'plami ikkita $[c_0, c_1][c_1, c_2]$ intervalga ajratiladi, bu yerda $c_0 = \min_{S_v \in E_0} y_{vj}$ va $c_2 = \max_{S_v \in E_0} y_{vj}$, ($S_v = (y_{v1}, \dots, y_{v\mu})$). Intervalning

¹ Ignatyev N A 2018 Structure Choice for Relations between Objects in Metric Classification Algorithms Pattern Recognition and Image Analysis vol 28 pp 590–597.



2-TOM, 11-SON

chegarasining qiymatlarini hisoblash quyidagi gipoteza, ya'ni har bir interval obyektlar alomatlari qiymatlarining K_t yoki K_{3-t} ($t=1,2$) sinfdan olinganiga asoslanadi.

Faraz qilaylik, $u_1^1, u_1^2(u_2^1, u_2^2) - y_j \in Y(\mu)$ alomatning $K_1(K_2)$ sinflarga $[c_0, c_1]$ va $(c_1, c_2]$ intervallarga tegishli qiymatlari soni bo'lsin. $A = (a_0, a_1, a_2), a_0 = 1, a_2 = m, a_1 = E_0$ tanlanmadan olingan $y_j \in Y(\mu)$ alomat qiymatlarining o'sib borish tartibida tartiblangan va interval chegarasini $c_1 = r_{a_1}, m_t = |K_t \cap E_0|, t = 1, 2$ aniqlovchi ketma-ketlik.

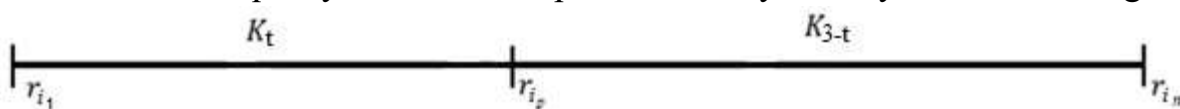
Quyidagi

$$\left(\frac{\sum_{p=1}^2 u_1^p (m - m_t - u_2^p) + u_2^p (m_t - u_1^p)}{2m_1 m_2} \right) \left(\frac{\sum_{p=1}^2 \sum_{i=1}^2 u_i^p (u_i^p - 1)}{m_1 (m_1 - 1) + m_2 (m_2 - 1)} \right) \rightarrow \max_{\{A\}} \quad (1)$$

mezonni intervalning c_1 chegarasining optimal qiymatini hisoblash va uning (mezonning) qiymatidan E_0 to'plam obyektlarini sinflarga ajratishda sonli alomatning kompaktilik ko'rsatkichi sifatida foydalanish mumkin. Agar ikkita obyektning xar birining chegaralarida faqat K_t yoki K_{3-t} olingan obyektlar a'loomatlari qiymatlari $y_j \in Y(\mu)$ joylashgan bo'lsa, u holda (1) mezon qiymati birga teng bo'ladi (1-rasm).

Agar bo'lsa, mezon qiymati 0 ga teng bo'ladi. Boshqa barcha hollarda (1) mezon qiymatlari (0,1) intervalga tegishli qiymatlarni qabul qiladi, (1) mezon sinflar soni $l > 2$ bo'lgan hollarda ham ishlatilishi mumkin.

1-rasm. Miqdoriy alomatni kompaktilik kriteriyasi bo'yicha intervallarga bo'lish.



Hisoblashni soddalashtirish uchun berilganlarga dastlabki ishlov berish tavsiya qilinadi. Berilganlarni dastlabki ishlov berish deganda tartiblangan ketma-ketlik asosida quyidagi butun sonli matritsani shakllantirish tushuniladi:

$$D = \begin{pmatrix} d_{10} & d_{11} & \dots & d_{1m} \\ \dots & \dots & \dots & \dots \\ d_{l0} & d_{l1} & \dots & d_{lm} \end{pmatrix}, \quad (2)$$

Bu yerda $d_{pi}, p = 1, \dots, l, i = 1, \dots, m$ ustun elementi alomat qiymati r_{ij} bo'lgan $S \in E_0$ obyektga tegishli.

D matritsa elementlari quyidagicha hisoblanadi:



2-TOM, 11-SON

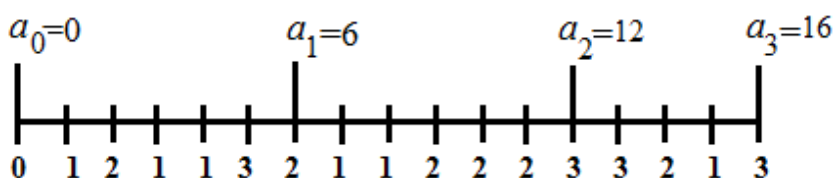
$$d_{pi} = \begin{cases} 0, & i = 0, \\ d_{p,i-1} + g(p,i), & i > 0, \end{cases} \text{ bu yerda } g(p,i) = \begin{cases} 1, S \in K_p, \\ 0, S \notin K_p. \end{cases}$$

$K_p, p = 1, \dots, l, i = 1, \dots, l$ sinfning $[c_1, c_2]$ intervaldagi, o'ng va chap chegaralari quyidagi indekslarga mos keluvchi $\eta = a_{t-1}, v = a_t, c_{2t-1} = r_{j\eta}, c_{2t} = r_{jv}, t = 1$ va $t > 1$ da $(c_{2t-1}, c_{2t}]$ vakillari soni quyidagicha topiladi:

$$u_t^p = d_{pv} - d_{p\eta} \tag{3}$$

Sonli alomat qiymatlarini (1) mezon asosida o'zaro kesishmaydigan intervallarga ajratish algoritmini vizual namoyishi 2-rasmda keltirilgan. Bunda $m=16$, sinflar soni $l=3$ va sinflar elementlari soni . Bunda (3) dastlabki ishlov berish natijalaridan foydalanilgan.

2-rasm. Algoritmning vizual namoyishi



$$D = \begin{pmatrix} 0 & 1 & 1 & 2 & 3 & 3 & 3 & 4 & 5 & 5 & 5 & 5 & 5 & 5 & 6 & 6 \\ 0 & 0 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 3 & 4 & 5 & 5 & 5 & 6 & 6 & 6 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 3 & 3 & 3 & 4 \end{pmatrix}$$

Intervallarga ajratish variantlari 1-jadvalda keltirilgan.

1-jadval. Intervallarga ajratish variantlari

№	a_1	a_2	(1) mezon qiymatlari
1	1	2	0.19444444444444445
2	1	3	0.21296296296296297
3	1	4	0.25
...			
17	2	6	0.2976190476190476
18	2	7	0.30357142857142855
19	2	8	0.34523809523809523(optimal)
20	2	9	0.25462962962962965



2-TOM, 11-SON

21	2	10	0.20634920634920634
22	2	11	0.25793650793650796
. . .			
85	10	14	0.19345238095238096
86	11	12	0.26388888888888889
87	11	13	0.23611111111111111
88	11	14	0.23611111111111111
89	12	13	0.23148148148148148
90	12	14	0.23809523809523808
91	13	14	0.19642857142857142

Optimal ajratishda ($a_1 = 2$ va $a_2 = 8$) bir sinfga tegishli obyektlar-ning barcha alomatlari qiymatlarini o'z ichiga oluvchi birorta intervalning yo'qligi ko'rinib turibdi.

(1) asosida optimal ajratish intervallarining nomerlarini sonli belgi qiymatining nominal alomat qiymatlariga o'tkazish gradatsiyalari sifatida qarash mumkin. Bunday shakl almashtirish turli toifadagi maksimal tarzda o'zaro bog'liq bo'lmagan informativ to'plamlarini qidirishda foydalanilgan [6].

(1) matritsa asosida (3) dan foydalanish intervallar va latent (oshkor tarzda o'lchash mumkin bo'lmagan) alomatlar salmog'ini (1) asosida hisoblash imkoniyati yaratildi. Alomatning vazni deganda (1) mezonning optimal qiymati tushuniladi. Amaliyotda latent alomatlar ko'pincha turli indekslar ko'rinishida ishlatiladi. Masalan, meditsinada tana massasi indeksi, Kerdo indeksi tushunchalari ishlatiladi. Latent alomatlarni salmog'ining birga yaqin yoki birga teng qiymatlari intuitiv yechimlar qabul qilish modellarini qurishga asos bo'la oladi.

Alomatning (1) mezon bo'yicha salmog'i o'z ichiga alomatning informativligi xaqidagi muhim ma'lumotni oladi. Biroq, alomatlar axborotliligi to'plamlarini ajratganda, ularning salmoqlari bo'yicha tartiblanishiga to'laligicha asoslanish maqsadga muvofiq emas, ya'ni *“alomatning salmog'i qancha katta bo'lsa, alomatlar to'plamida ana alomatning informativligi shuncha yuqori bo'ladi”* tamoyili hamma vaqt ham o'rinli emas. Bunda alomatlarning o'zaro korrelyatsion bog'liqligidan ham foydalaniladi. Turli toifadagi informativ alomatlar to'plamlarini ajratish va ularning sun'iy neyron to'rlarini qo'llash samaradorligiga ta'sirini tadqiq qilishga qo'llanilgan [4].

Ikkinchi mezon [5] da keltirilgan bo'lib, u mumkin bo'lgan kesishmaydigan intervallar soni $p \geq 2$ bo'lgan ikkita $K1, K2$ sinflar vakillari uchun mo'ljallangan. Intervallar sonini p



2-TOM, 11-SON

ni aniqlash, sonli alomatning $r_1, \dots, r_u, \dots, r_v, \dots, r_m$ tartiblangan qiymatlarida K_i sinf obyektlarining $\frac{d_i(u, v)}{|K_i|}$, $i = 1, 2, u \leq v$ munosabati bo'yicha hisoblanuvchi uchrashlar chastotalarini tahlil qilishi asosida amalga oshiriladi. $d_i(u, v)$ alomat qiymatlari r_u, \dots, r_v to'plam ostini hosil qiluvchi K_i sinf obyektlari sonini ifodalaydi. Kesishmaydigan intervallar chegaralari $[r_{c_u}, r_{c_v}]^i, i = 1, \dots, p$ quyidagi mezonni rekursiv optimizatsiyalash natijasida topiladi:

Xulosa:

Berilganlarni tahlil qilish uchun ma'lumotlarni shakllantirishda obyekt-alomat ko'rinishidan foydalanildi. Predmed sohalarda, masalan meditsinada katta hajmdagi berilganlar tarkibida o'tkazib yuborilgan qiymatlar ko'plab uchrab turadi. Buning sabablari turlicha bo'lishi mumkin, topshirilmagan tahlillar, qurilmalarning yetishmasligi, inson omili va hokazo. Bu berilganlar intellektual tahlilining an'anaviy usullarini ishlatishda katta muammolarni keltirib chiqaradi.

Muammoning shartlariga ko'ra quyidagilar talab qilindi:

- alomatlar majmui bo'yicha ketma-ketlik qonuniyatlarini aniqlash;
- bo'shliqlar mavjud bo'lganda har xil turdagi xususiyatlar uchun yaqinlik matritsalarini yaratish;
- hisoblash eksperimenti orqali o'lchovsiz ma'lumotlar qiymatlarining nisbatlariga qarab funktsiyalar ketma-ketligi tartibining barqarorligini aniqlash.

Interval usullar asosida dasturiy ta'minot ishlab chiqildi. Tahlil qilish uchun intervalli usullarni tanlash, ularning yordami bilan, ma'lumotlarni tahlil qilishda bunday muhim xususiyat belgilari o'lchovlari uchun o'zgaruvchanlik sifatida amalga oshirilganligi bilan izohlanadi. Ichki yashirin qonuniyatlarni izlash miqdoriy alomatning tartiblangan ketma-ketligini ustunlik intervallariga bo'lish, har bir interval tegishlilik funksiyasi qiymatlari va alomat qiymatlarini intervallarga bo'lish turg'unligini hisoblash orqali hal qilindi. Bo'shliqlar bo'lgan holatlarda ham olingan qonuniyatlarni tekshirish uchun intervallarga bo'lish turg'unligi qiymati olinadi. Bo'shliqlar qiymati ko'p bo'lganda alomatlarning turg'unligi bo'yicha ustunlik ketma-ketligi o'zgarib ketishi mumkin. Bunday holatlarda ham turg'unlik qiymati yuqori bo'lsa aniqlangan qonuniyatga ishonishish mumkinligi ehtimoli ham yuqori bo'ladi.



2-TOM, 11-SON

FOYDALANILGAN ADABIYOTLAR

1. O‘zbekiston respublikasi prezidentining “Innovatsion loyihalar va texnologiyalarni ishlab chiqarishga tatbiq etishni rag‘batlantirish borasidagi qo‘shimcha chora-tadbirlar to‘g‘risida”gi qarori Toshkent sh.,2008-yil 15-iyul, PQ-916-son
2. Вапник В.Н. Алгоритмы и программы восстановления зависимостей. М.: Наука, 1984.- 816 с.
3. Наследов А.Д. SPSS: Компьютерный анализ данных в психологии и социальных науках. - СПб.: Питер, 2005. 416 с.
4. Дюк В.А. Осколки знаний. Экспресс-Электроника, 2002, № 6, С. 60-65
5. Згуральская Е.Н. Алгоритм выбора оптимальных границ интервалов разбиения значений признаков при классификации // Известия Самарского научного центра Российской академии наук. Т. 14, №4(3), 2012.- С.826-829.
6. Игнатьев Н.А. Вычисление обобщённых показателей и интеллектуальный анализ данных // Автоматика и телемеханика. 2011. № 5. С.183-190.
7. Згуральская Е.Н. Выбор информативных признаков для решения задач классификации с помощью искусственных нейронных сетей // Нейрокомпьютеры: разработка, применение. 2012. № 2. С. 20 - 27.
8. ***М. П. Кривенко “Обучаемая классификация неполных клинических данных”, Информ. и её примен., 2017, том 11, выпуск 3, 27–3.***
9. N.A.Ignat’ev, R.N.Usmanov, Sh.F.Madrahimov. “Berilganlarning intellektual tahili”. // O‘quv qo‘llanma, Toshkent – 2018, 55 – 60s.
10. Игнатьев Н.А. “Обобщенные оценки и локальные метрики объектов в интеллектуальном анализе данных”. Монография.- Ташкент: Издательство “Университет”, 2014. – 72 с.
11. Python.org – Python dasturlash tilining rasmiy web sayti.
12. Library.ziyonet.uz – respublika ta’lim portali.
13. Doc.qt.io – Qt freymworkinging rasmiy web sayti.

