

SUN'IY INTELLEKT MASALALARIDA BERILGANLAR BAZASINI TAYYORLASH VA ULARGA DASTLABKI ISHLOV BERISH TEXNOLOGIYALARI

Murodov Azizbek Muhammadqul o'g'li

O'zMU, Amaliy matematika va intellektual

texnologiyalar fakulteti 3-kurs talabasi

El.pochta: murodov_a@nuu.uz

[Tel: +998912112143](tel:+998912112143)

Annotatsiya. Berilganlarga dastlabki ishlov berish jarayoni ular ustida berilganlarning intellektual tahlili usullarini amalga oshirishga tayyorlashda bajariladigan amallardan iborat. Hozirda tahlil qilish uchun turli manbaalardan olinishi mumkin bo'lgan ma'lumotlar hajmi sezilarli darajada oshgan bo'lsa-da, bu har doim ham yaxshi natijalarga olib kelmaydi. Bu mavjud ma'lumotlarning axborot mazmuni va sifati bilan bog'liq bo'lib, ushbu tadqiqot mashinali o'rgatish algoritmlari uchun berilganlarni tayyorlash usullarini qo'llashga qaratilgan.

Kalit so'zlar. Berilganlarning intellektual tahlili, o'rgatuvchi tanlanma, dastlabki ishlov berish, tushirib qoldirilgan qiymatlar, mashinali o'rgatish (machine learning)

Sifatli berilganlarga qo'yiladigan talablar.

Berilganlarni sifatli tahlil qilish va mashinali o'rganish jarayonida foydalanish uchun ular quyidagi talablarni qondirishi lozim:

1. **To'liqlik** – berilgan ma'lumotlarda tushirib qoldirilgan qiymatlar minimal bo'lishi yoki ular mos usullar bilan to'ldirilgan bo'lishi lozim.
2. **Aniqlik va ishonchlilik** – ma'lumotlar aniq va ishonchli bo'lishi, noto'g'ri yoki noaniq qiymatlar bartaraf etilgan bo'lishi kerak.
3. **Muvofiqlik** – har bir atribut (xususiyat) aniq belgilangan diapazonda bo'lishi va mantiqiy bog'liqlikni saqlashi lozim.
4. **Normallashtirish va shkala bo'yicha moslashtirish** – ma'lumotlar turli o'lchov birliklaridan tozalanib, umumiy normalizatsiya qoidalariga keltirilishi lozim.
5. **Shovqinni kamaytirish** – keraksiz yoki noto'g'ri ma'lumotlar ajratib olinib, statistik usullar bilan shovqin kamaytirilishi lozim.
6. **Toifalash va kodlash** – nominativ atributlar mos kodlash usullari (one-hot encoding, label encoding) bilan raqamlashtirilishi lozim.
7. **Muvozanatli o'rgatuvchi tanlanma** – o'quv ma'lumotlari muvozanatli bo'lib, yetarli darajada aniqlikga ega bo'lishi lozim.

Tushirib qoldirilgan qiymatlarni aniqlash va ularni bartaraf qilish.

O'rgatuvchi tanlanmani nazorat qilish. Berilgan $A=\{a_{uv}\}_{m \times n}$ jadvalida m ta $S_1, \dots, S_m$ obyektlar to'plamini $X(n)=(x_1, \dots, x_n)$ miqdoriy alomatlar bilan tavsifi berilgan. Har bir $(x_i, x_j) \in X(n)$ alomatlar juftligi bo'yicha aniqlash talab etiladi:

– mumkin bo'lgan intervallar $[z_1, z_2]$ chegaralari ko'rinishida hisoblansin, bu yerda

$$R(i, j) = \{a_{ki}/P_i - a_{kj}/P_j\}_{k \in \{1, \dots, m\}} \quad (1)$$

bo'yicha hisoblansin, bu yerda $S_k=(a_{k1}, \dots, a_{kn})$, $P_i, P_j - (x_i, x_j)$ alomatlarining E_0 to'plamdagi qiymatlarining medianasi,

$$z_1 = \min_{E_0} R(i, j)$$

$$z_2 = \max_{E_0} R(i, j)$$

– E_0 to'plamdagi qiymatlari z_1 va z_2 intervallar chegaralari bo'lgan ob'ektlar nomerlari [2].

Berilgan $A=\{a_{uv}\}_{m \times n}$ jadvalida m ta $S_1, \dots, S_m$ obyektlar to'plami mavjud bo'lib, ushbu obyektlar n ta miqdoriy alomatlar $X(n)=(x_1, \dots, x_n)$ yordamida tavsiflanadi. Ushbu jadvalda har bir alomat juftligi $(x_i, x_j) \in X(n)$ bo'yicha quyidagi masalani hal qilish talab etiladi:

1. Har bir alomat juftligi (x_i, x_j) uchun mumkin bo'lgan qiymatlarning interval chegaralari $[z_1, z_2]$ ko'rinishida aniqlansin. Bu intervallar quyidagi qoidaga asoslanadi: bu yerda:

○ $S_k=(a_{k1}, \dots, a_{kn})$ obyektning alomatlar bo'yicha qiymatlarini ifodalaydi;

○ P_i va $P_j - E_0$ to'plamida x_i va x_j alomatlarining median qiymatlari.

2. E_0 to'plami har bir alomat x_i uchun qiymatlar to'plamini bildiradi. Alomat juftligi (x_i, x_j) uchun hisoblangan qiymatlar asosida z_1 va z_2 intervallar chegaralari aniqlanadi.

3. Har bir $[z_1, z_2]$ intervali uchun ushbu intervalga mos keladigan ob'ektlarning raqamlari aniqlansin. Ya'ni, interval chegaralari ichida yotuvchi qiymatlarni aniqlovchi ob'ektlarning indeksleri qayd qilinadi.

Natijada, har bir alomat juftligi uchun mumkin bo'lgan qiymatlarning chegaralari va ushbu qiymatlarga mos keluvchi obyektlarning tartib raqamlari

hisoblab chiqiladi. Bu jarayon alomatlar orasidagi munosabatlarni chuqurroq tahlil qilish va intervalni aniqlash orqali obyektlarni tavsiflash imkonini beradi.

Hisoblash eksperimenti

1-jadval. Alomatlar juftligi bo'yicha mumkin bo'lgan intervallar chegaralari.

Alomatlar juftligi	z_1 - (3.2.1) bo'yicha minimal qiymat	z_2 - (3.2.1) bo'yicha maksimal qiymat
Vitamin D, Tug'ilgan yili	-0, 99375	0, 107311
Vitamin D, Viloyat	-0, 99391	0, 100633
Vitamin D, Yoshi	-33, 3872	0, 498773
Vitamin D, Ma'lumoti	-210, 373	0, 890033
Vitamin D, Ish joyi	-1, 69172	0, 249929
Vitamin D, Homila muddati	-1, 27339	0, 778843
Vitamin D, Homiladorlik soni	-181, 969	0, 91738
Vitamin D, Tug'ish soni	-2, 51551	0, 59477
Vitamin D, Abort	-1, 02108	1, 099605
Vitamin D, Oldingi tug'ruq	-1, 01416	1, 099612
Vitamin D, 4kg dan ortiq vazn	-1, 14482	1, 099479
Vitamin D, Bo'yi, sm	-1, 08668	0, 346513
Vitamin D, Vazni, kg	-13, 1735	0, 500519
Vitamin D, Homiladorlikkacha TVI	-1, 29894	0, 414397
Vitamin D, TVI	-11, 4674	0, 439693
Vitamin D, ИМТ ранжировка	-2, 35117	1, 100633
Vitamin D, Yelka o'lchami	-1, 28859	0, 425689
Vitamin D, Homiladorlikkacha bo'lgan vazn	-1, 33662	0, 443708
Vitamin D, Ortiqcha vazn	-13, 313	14, 76325
Vitamin D, Yaqin qarindoshlarda QD	-1, 10789	1, 099518
Vitamin D, Farzand nobud bo'lishi	-1, 01416	1, 099612
Vitamin D, Ortiqcha vazn >15-20 kg	-1, 18735	1, 099438
Vitamin D, Yoshi >30 yosh	-1, 25717	1, 099368
Vitamin D, ortiqcha vazn TVI>25	-1, 11229	1, 099513
Vitamin D, АГ >140/90	-1, 01761	1, 099608
Vitamin D, Gepatit	-1, 10693	0, 729625
Vitamin D, Och qoringa glikemiya	-1, 04583	1, 095684
Vitamin D, Ovqatlanishdan 1 soatdan so'ng glikemiya	-22, 3197	1, 011547
Vitamin D, HBA1C	-1, 15034	1, 098318
Vitamin D, GDM	-1, 06673	0, 564221

Xulosa

Ushbu tadqiqot ishi natijasida berilganlarga dastlabki ishlov berish texnologiyalarining mashinali o'rgatish tizimlaridagi ahamiyati o'rganildi. Berilganlarning sifatli tahlili uchun ularning to'liqligi, aniqligi va ishonchliligi, muvofiqlik, normallashtirilganlik, shovqindan tozalanganlik, to'g'ri toifalanganlik va muvozanatli o'rgatuvchi tanlanma kabi muhim talablar aniqlab berildi. Olib borilgan hisoblash eksperimentlari natijasida alomatlar juftligi bo'yicha mumkin bo'lgan intervallar chegaralari (z_1 , z_2) aniqlandi va ushbu masalalar yechimida (1) formulaga asoslangan usul samarali ekanligi ko'rsatildi. Eksperimental natijalar turli tibbiy ko'rsatkichlar misolida tekshirilib, amaliy ahamiyatga ega bo'lgan dastlabki ishlov berish jarayonlari asoslab berildi.

Berilganlar bazasini tayyorlash va ularga dastlabki ishlov berish usullarini takomillashtirish orqali sun'iy intellekt tizimlarining aniqligi va ishonchliligini sezilarli darajada oshirish mumkinligi aniqlandi. Kelajakda ushbu metodlarni turli sohalarda qo'llash va ularni yanada takomillashtirish bo'yicha tadqiqotlarni davom ettirish maqsadga muvofiq.

Foydalanilgan adabiyotlar:

1. Bhaya, Wesam. (2017). Review of Data Preprocessing Techniques in Data Mining. Journal of Engineering and Applied Sciences. 12. 4102-4107. 10.3923/jeasci.2017.4102.4107.
2. N.A. Ignatev Sun'iy intellektual tizimlar bo'yicha ilmiy-metodik masalalar to'plami // Ilmiy-uslubiy qo'llanma. "Zamin nashr", 2024, 136-b. ISBN 978-9910-9296-0-1.