

**TAQSIMOT ZICHLIGI VA OBYEKTAR BOG'LANGANLIGI ASOSIDA
YASHIRIN QONUNIYATLARNI IZLASH**

Kuchkarova Marjonabonu Karimjon qizi

2-kurs magistranti.

Madraximov Sh.

Ilmiy rahbar. Texnika fanlari doktori, professor.

<https://doi.org/10.5281/zenodo.23303921>

Annotatsiya. Mazkur ishda ma'lumotlar to'plamida mavjud bo'lgan, bevosita kuzatish orqali aniqlash qiyin bo'lgan yashirin qonuniyatlarni taqsimot zichligi hamda obyektlar o'rtasidagi bog'langanlik munosabatlari asosida aniqlash masalasi ko'rib chiqiladi.

Obyektlarning ma'lumotlar fazosidagi joylashuvi, lokal zichlik ko'rsatkichlari va o'zaro yaqinlik munosabatlaridan foydalanish orqali tabiiy guruhlanishlar, anomal obyektlar va yashirin tuzilmalarni aniqlash yondashuvi muhokama qilinadi. Ushbu yondashuv klasterlash, anomalialarni aniqlash va qaror qabul qilish masalalarida qo'llanishi mumkin.

Kalit so'zlar: taqsimot zichligi, obyektlar bog'langanligi, yashirin qonuniyat, ma'lumotlar tahlili, klasterlash, anomal obyekt, yaqinlik, klassifikatsiya, mashinali o'qitish.

Kirish

Hozirgi vaqtda turli sohalarda katta hajmdagi ma'lumotlarning shakllanishi ularni avtomatik tahlil qilish va ma'lumotlarda mavjud yashirin qonuniyatlarni aniqlash masalasini dolzarb vazifalardan biriga aylantirmoqda. Ma'lumotlar to'plamida ayrim qonuniyatlar bevosita ko'rinmasligi, obyektlarning turli zichlikdagi hududlarda joylashishi yoki ular orasida murakkab munosabatlar mavjudligi sababli an'anaviy statistik va klassifikatsiya usullari har doim ham yetarli natija bermasligi mumkin.

Shu sababli obyektlarning ma'lumotlar fazosidagi joylashuvi, ularning lokal taqsimot zichligi va o'zaro bog'langanlik xususiyatlarini birgalikda tahlil qilish yashirin strukturaviy qonuniyatlarni aniqlash uchun istiqbolli yo'nalish hisoblanadi.

Asosiy qism

Taqsimot zichligi ma'lumotlar fazosining ma'lum bir sohasida obyektlarning qanchalik konsentratsiyalanganligini ifodalaydi. Agar ma'lum bir hududda obyektlar soni yuqori bo'lsa, ushbu hudud yuqori zichlikli soha sifatida qaraladi. Har bir obyekt uchun lokal zichlikni baholash orqali uning atrofidagi ma'lumotlar taqsimoti haqida qo'shimcha axborot olish mumkin.

$$\rho(x_i) = k / V_i$$

Bu yerda k — ko'rib chiqilayotgan qo'shnilar soni, V_i esa x_i obyektning k ta yaqin qo'shnisini qamrab oluvchi fazoviy soha hajmini bildiradi. Mazkur ifoda zichlikni tushuntirish uchun soddalashtirilgan modeldir; amaliy tadqiqotda o'lchamlar soni va masofa metrikasiga mos zichlik baholagichi tanlanishi lozim.

Biroq faqat zichlik qiymatidan foydalanish obyektlar o'rtasidagi munosabatlarni to'liq ifodalab bermaydi. Shu sababli obyektlarning bog'langanligi ham hisobga olinadi. Bog'langanlik ularning masofasi, umumiy qo'shnilari, bir xil zichlikli sohaga mansubligi yoki boshqa munosabat mezonlari orqali aniqlanishi mumkin.

Obyektlar orasidagi bog‘langanlikni grafik yordamida ifodalash mumkin. Bunda har bir obyekt grafikning tuguni, obyektlar o‘rtasidagi munosabat esa qirra sifatida qaraladi. Natijada ma‘lumotlar to‘plami bog‘langan grafik ko‘rinishiga keltiriladi.

Taqsimot zichligi va bog‘langanlikni birgalikda tahlil qilish orqali yuqori zichlikka ega tabiiy guruhlar, turli guruhlar orasidagi chegaraviy obyektlar, past zichlikda joylashgan anomal obyektlar hamda obyektlarning yashirin o‘zaro munosabatlarini aniqlash mumkin.

Tadqiqotning umumiy jarayoni quyidagi bosqichlardan iborat: ma‘lumotlarni tayyorlash; obyektlar orasidagi masofani aniqlash; lokal zichlikni hisoblash; obyektlarning bog‘langanligini aniqlash; yashirin guruh va strukturalarni ajratish; anomal va chegaraviy obyektlarni aniqlash; natijalarni klassifikatsiya yoki prognozlashda qo‘llash.

Ushbu jarayonda Evklid masofasi, Manhattan masofasi yoki masalaga mos boshqa metrikalardan foydalanish mumkin. Yuqori o‘lchamli ma‘lumotlarda masofa funksiyasini tanlash alohida ahamiyatga ega.

Tadqiqotning ilmiy ahamiyati

Tadqiqotning ilmiy ahamiyati ma‘lumotlar obyektlarining faqat individual xususiyatlarini emas, balki ularning lokal taqsimot zichligi va o‘zaro bog‘langanlik munosabatlarini ham hisobga olish g‘oyasi bilan belgilanadi. Bunday yondashuv obyektning ma‘lumotlar fazosidagi kontekstual holatini tahlil qilish imkonini beradi. Zichlik va bog‘langanlikni birgalikda ishlatish klassifikatsiya hamda klasterlash jarayonlarida shovqinli, anomal yoki chegaraviy obyektlarni aniqlashga yordam berishi mumkin.

Ushbu tezisda bayon qilingan yondashuv tadqiqot yo‘nalishi sifatida taklif etilmoqda.

Uning ilmiy yangiligi va samaradorligi alohida matematik model, mavjud usullar bilan taqqoslash hamda tajribaviy natijalar orqali asoslanishi kerak.

Amaliy ahamiyati

Taklif etilayotgan yondashuvni tibbiy diagnostikada bemorlar guruhlarini tahlil qilish, moliyaviy ma‘lumotlarda noodatiy holatlarni aniqlash, mijozlarni segmentatsiya qilish, sanoat tizimlaridagi anomal holatlarni aniqlash, ijtimoiy-iqtisodiy ma‘lumotlarni tahlil qilish va mashinali o‘qitish uchun ma‘lumotlar bazasini optimallashtirish kabi masalalarda o‘rganish mumkin.

Xususan, obyektlarning zichligi va bog‘langanligini hisobga olish etalon obyektlarni tanlash, shovqinli obyektlarning ta‘sirini baholash va klassifikatsiya sifatini oshirish masalalarida qo‘shimcha imkoniyatlar yaratishi mumkin. Biroq bu imkoniyatlar tajriba orqali tekshirilishi lozim.

Xulosa

Taqsimot zichligi va obyektlar bog‘langanligini birgalikda tahlil qilish ma‘lumotlar to‘plamidagi yashirin qonuniyatlarni aniqlashning istiqbolli yo‘nalishlaridan biridir. Ushbu yondashuv obyektlarning lokal holatini baholash, tabiiy guruhlarini aniqlash, anomal va chegaraviy obyektlarni ajratish hamda obyektlar orasidagi munosabatlarni o‘rganish imkonini beradi.

Keyingi bosqichda zichlik va bog‘langanlik ko‘rsatkichlarini birlashtiruvchi mezonning matematik modelini ishlab chiqish, uning parametrlarini asoslash va ochiq ma‘lumotlar to‘plamlarida tajribaviy sinovdan o‘tkazish maqsadga muvofiq. Olingan natijalar klassifikatsiya va etalon obyektlarni tanlash algoritmlarini takomillashtirishga xizmat qilishi mumkin.

Foydalanilgan adabiyotlar

1. Ester M., Kriegel H.-P., Sander J., Xu X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD), 1996, pp. 226–231.
2. Breunig M. M., Kriegel H.-P., Ng R. T., Sander J. LOF: Identifying Density-Based Local Outliers. Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, 2000, pp. 93–104.
3. Schubert E., Sander J., Ester M., Kriegel H.-P., Xu X. DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN. ACM Transactions on Database Systems, 2017, 42(3), Article 19.
4. Игнатьев Н.А., Згуральская Е.Н. Кластерный анализ с применением обучения на основе отношений связанности и плотности распределения // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2024. № 68. С. 66–74. DOI: 10.17223/19988605/68/7.