TURLI XILDAGI MA'LUMOTLARNI TAHLIL QILISH USULLARINING ZAMONAVIY HOLATI

Nishanov Axram Xasanovich[1], Tursunov Alisher Tulkunovich[2], Narzullayev Davronbek Zikrullayevich[2], Rahmonov Erkin Davlatovich[2]

[1]Muxammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti Tizimli va amaliy dasturlashtirish kafedrasi,
[2]Toshkent farmasevtika instituti Fizika, matematika va axborot texnologiyalari kafedrasi
E-mail:tursunovr484za@gmail.com

K E Y W O R D S

Ma'lumotlarni tahlil qilish, katta hajmdagi ma'lumotlar, klasterli tahlil, omilli tahlil, neyron tarmoqlar, funksiyalarni yaqinlashtirish, bashoratlash, optimallashtirish, qarorlar daraxti

A B S T R A C T

Ushbu maqolada turli xildagi ma'lumotlarni tahlil qilish masalalarida tarqoq ma'lumotlarni tizimlashtirish, eng muhim ma'lumotlarni tanlash, ularni ma'lum bir tartibda joylashtirish to'g'risidagi tushunchalar keltirilgan. Tahlil jarayonida ma'lum bir qonuniyatlar aniqlanadi, xulosalar chiqariladi vas shu asnosida bashoratlar qilinadi. Axborotni o'rganish katta hajmdagi ma'lumotlarni (Big data) tadqiq qilish, vaqt bo'yicha uzoq bo'lgan hodisalar natijalarini taqqoslash imkonini beruvchi turli zamonaviy texnologiyalardan foydalangan holda amalga oshiriladi. Hozirgi kunda katta hajmdagi ma'lumotlarni tahlil qilish usullarining zamonaviy holati haqida to'xtalib o'tilgan.

I.Kirish.

Ma'lumotlarni tahlil qilish har qanday faoliyat sohasining kalitidir, bu yangi korxonani shga tushirish, marketing qarorlarini qabul qilish, inson resurslarini boshqarish yoki tibbiyot sohasi vakillarining bemorlarni davolash bo'yicha to'g'ri qarorlar qabul qilishlarida muhim ro'y o'ynaydi. Ma'lumotlarni tahlil qilish usullarini tanlashdan oldin bajarilishi kerak bo'lgan ish hajmi, maqsadga muvofiqligi, maqsad va vazifa kabi omillarni hisobga olish muhimdir. Maqsadlarni tahlil qilish asosida aniq qo'yilgan vazifaga mos keladigan optimal yechimlarni tanlash yoki umumiy maqsadga erishish uchun ma'lumotlarni qayta ishlash, tahlil qilish, qidirish, vizualizatsiya qilishning turli usullaridan birgalikda foydalaniladigan murakkab tuzilmani ishlab chiqish mumkin. Tahlil usullarini tanlash o'rganilayotgan axborotning xarakteriga bog'liq.

Tahlil qilinayotgan ma'lumotlar quyidagi turlarga bo'lish qabul qilingan:

- sonli (miqdoriy - masalan, tovar qiymati ko'rsatkichlari);
- oraliqli (ma'lum bir oraliqda o'rganiladigan tanlanma ma'lumotlar);
- rangli (o'sib borish tartibida joylashgan qiyosiy ko'rsatkichlar);
- nominal (taqqoslab bo'lmaydigan sifat xususiyatlari - masalan, ismlar, kasblar, raqamlar va tibbiy ma'lumotlar).

II. Asosiy qism

Ma'lumotlarni tahlil qilish va qayta ishlash usullari

Klasterli tahlil. "Klasterli tahlil" atamasi (birinchi marta Tryon tomonidan 1939-yilda kiritilgan) aslida turli xil tasniflash algoritmlari to'plamini o'z ichiga oladi. Ko'plab sohalardagi tadqiqotchilar tomonidan qo'yiladigan umumiy savol shundan iboratki, ma'lumotlarni parametrlarning qiymatlari yaqin bo'lgan guruhlarga qanday ajratish kerak.

Klaster tahlilini o'tkazish uchun ma'lumotlarni to'plashdan tashqari, ikkita narsani aniqlash kerak: ma'lumotlarni nechta klasterga bo'lish kerak va ma'lumotlardagi o'xshashlik o'lchovini qanday aniqlash kerak. Masalan, dori vositalari yoki tibbiy ma'lumotlarni bir nechta klasterga birlashtirish mumkin. Shunda o'xshashlik o'lchovi tibbiy ma'lumotlarning bir-

© A.X. Nishanov, A.T. Tursunov, D.Z. Narzullayev, E.D. Rahmonov

biriga kommunikatsion yaqinligi bilan belgilanadi. Murakkabroq hollarda klaster tahlili bo‘yicha adabiyotlarda batafsil tavsiflangan boshqa o‘xshashlik o‘lchovlarini qo‘llash mumkin.

Omilli tahlil. Agar ko‘p sonli parametrlar (100 dan ortiq) mavjud bo‘lsa, bu parametrlarni guruhlash va endi har bir parametrni alohida emas, balki parametrlar guruhlarini yagona kompleks parametr (omil) sifatida tahlil qilish zarur bo‘ladi.

Omilli tahlil asosida yaqqol berilgan belgilarning murakkab o‘zaro aloqalari ortida o‘rganilayotgan hodisaning eng muhim xususiyatlarini aks ettiruvchi nisbatan oddiyroq tuzilma yotadi, “tashqi” belgilar esa bu tuzilmani belgilovchi yashirin umumiy omillarning funksiyalari hisoblanadi degan g‘oya yotadi.

Omilli tahlilni qo‘llash, birinchidan, ko‘rib chiqilayotgan parametrlar sonini kamaytirish (reduksiyalash), ikkinchidan, har biri bitta mustaqil parametr bo‘ladigan parametrlarning mazmunli guruhlarini topish imkonini beradi.

Ushbu usulning o‘ziga xosligi shundaki, parametrlarni omillarga birlashtirishda har bir omil barcha parametrlardagi umumiy qonuniyatlarni o‘zida jamlaydi, har bir parametrning o‘ziga xos xususiyatlarini alohida tashlab yuboradi.

Neyron tarmoqlar. Ma’lumotlarni tahlil qilish vositasi sifatida neyron tarmoqlarining paydo bo‘lishi 40-yillarning boshlarida MakKallok va Pitsning ishlarida boshlangan edi. Ushbu tadqiqqot ishlarida ishda sun’iy neyron modeli taklif qilingan.

Miyaning neyron tuzilishini modellashtirish orqali sun’iy intellektga yaqinlashish mumkinligi taxmin qilingan. O‘sha paytga kelib, inson miyasi maxsus biologik hujayralar - neyronlardan tashkil topganligi ma’lum edi va neyronlardan tarmoqlar qurish inson miyasi har kuni hal qiladigan murakkab vazifalarni hal qilish imkonini beradigandek tuyulardi. Hozirda neyron tarmoqlari ma’lumotlarni tahlil qilishning yetarlicha mashhur vositalaridan biri hisoblanadi.

Neyron tarmoqlari deyarli har qanday sohaga qo‘llanilishi mumkin. Michigan davlat universiteti xodimi Anil K. Jain va IBM Jianchang tadqiqot markazi mutaxassislarining fikricha, Mao va K.M.Mohiuddin, neyron tarmoqlari uchun vazifalar ro‘yxatini quyidagicha tasniflash mumkin [8].

Timsollarni tasniflash. Ma’lum ilovalarga harflarni tanib olish, nutqni tanib olish, elektrokardiogramma signalini tasniflash, qon hujayralarini tasniflash, biometrik skanerlar faoliyatini ta’minlash va boshqalar kiradi.

Klasterlash/toifalash. Klasterlash bilimlarni ajratib olish, ma’lumotlar o‘lchamini kamaytirish va ma’lumtlarning xususiyatlarini o‘rganish uchun qo‘llaniladi.

Funksiyalarni yaqinlashtirish. Tipik misol sifatida uzatilayotgan axborotga bog‘liq bo‘lmagan holda turli tabiatli signalni qabul qilishda shovqinni so‘ndirishni keltirish mumkin.

Bashoratlash/prognoz. Bunga misol tariqasida fond birjasidagi narxlar prognozi va ob-havo prognozini keltirish mumkin.

Optimallashtirish. Bir qator ko‘nikmalar va omillar bo‘yicha xodimlar shtatini tayinlash optimallashtirish masalalarining klassik misollari hisoblanadi.

Neyron tarmoqlari yordamida ma’lumotlarni tahlil qilishning umumiy sxemasi quyidagilardan iborat:

Tarmoq tipologiyasini tanlash. Tarmoqlarning 9 ta turi mavjud bo‘lib, bu bosqichda vazifaga eng mos keladigan tarmoq turi tanlanadi.

Tarmoq xarakteristikalarini eksperimental tanlash. Turni tanlagandan so‘ng, tarmoq tuzilishini tanlash kerak (neyronlar soni, ularning og‘irligi, o‘zaro bog‘liqligi va boshqalar).

O‘qitish parametrlarini eksperimental tanlash. Keyinchalik o‘qitish parametrlarini tajriba yo‘li bilan aniqlash zarur: o‘qitishning maksimal vaqti, ma’lumotlar soni, yo‘l qo‘yilishi mumkin bo‘lgan maksimal xato va boshqalar.

Tarmoqni o‘qitish. O‘rgatuvchi tanlanma bo‘yicha tarmoqni o‘qitish amalga oshiriladi. O‘quv tanlanmasi ma’lumotlarni umuman tavsiflovchi ma’lumotlarni o‘z ichiga oladi deb taxmin qilinadi.

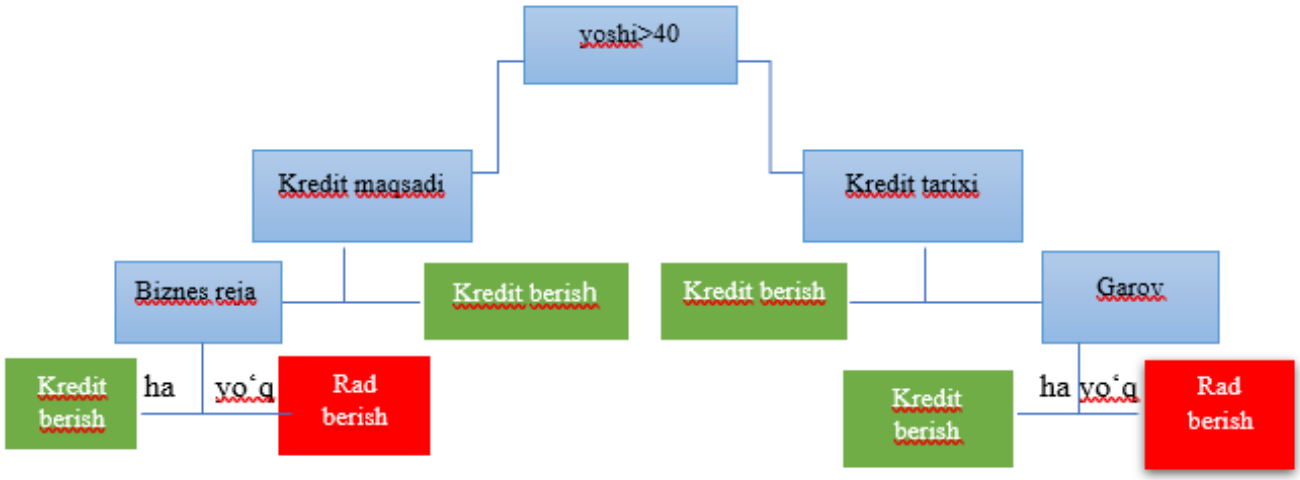
© A.X. Nishanov, A.T. Tursunov, D.Z. Narzullayev, E.D. Rahmonov

O‘qitishning adekvatligini tekshirish. O‘quv tanlanmasiga kirmagan ma’lumotlar bo‘yicha olingan natijalar tahlil qilinadi. Neyron tarmog‘ining ish natijalari qo‘lda nazorat qilinadi.

Neyron tarmoqlarining asosiy kamchiligi shundaki, neyron tarmog‘ini o‘qitish jarayoni va qaror qabul qilish jarayoni mutlaqo nazorat qilinmaydi. Boshqacha aytganda, neyron tarmog‘i “qora quti” bo‘lib, uning kirishida ma’lumotlar beriladi, chiqishida esa natija olinadi. Neyron tarmog‘i o‘z ichida nima qilayotganini tushunishning iloji yo‘q, chunki test ma’lumotlari tahlil qilinadi (neyron tarmog‘ini o‘qitish amalga oshiriladi), shu bilan birga tizim ichki parametrlarni (vaznni) avtomatik ravishda o‘zgartirib, xatolikni minimallashtirishga harakat qiladi. O‘qitilgan tarmoqda vazn qiymatlarini olish mumkin, ammo bu vaznlar berishi mumkin bo‘lgan yagona javob quyidagicha: test ma’lumotlaridagi qaysi parametr qanday rol o‘ynaydi, u qanday muhimlik darajasiga ega. Bu rollarning ma’nosi haqida bu usul bilan hech qanday tushuntirish olib bo‘lmaydi.

Qarorlar daraxti. Qarorlar daraxtlarini yaratishning dastlabki g‘oyalari Xovlendning ishlariga borib taqaladi. (Hoveland) va Xanta (Hunt) XX asrning 50-yillari ohirida ushbu yo‘nalishning rivojlanishiga turtki bergan.

Qaror daraxtlari qoidalarni ierarxik ketma-ket mantiqiy tuzilmada taqdim etish usuli bo‘lib, u kirishdagi obyekt yoki vaziyatni bir yoki bir nechta chiqish (terminal) tugunlari bilan bog‘lash imkonini beradi. Qoida deganda “agar... bo‘lsa” ko‘rinishidagi mantiqiy konstruksiya tushuniladi. Quyidagi masalani ko‘rib chiqamiz: jismoniy shaxslarga kredit berish bo‘yicha hal qiluvchi qoidani tuzish kerak. Bu holda qarorlar daraxti quyidagi ko‘rinishga ega bo‘lishi mumkin (1-rasm):



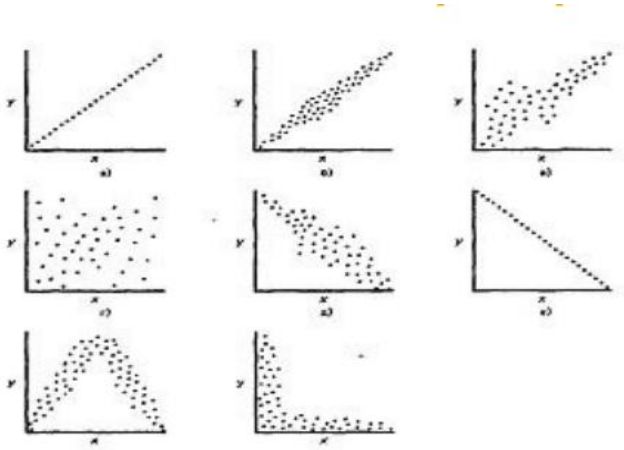
1-rasm. Aniq qarorlar daraxti

Regression tahlil. Korrelyatsion tahlil o‘zgaruvchilar o‘rtasidagi bog‘liqlikni o‘lchashda, regression tahlil esa bir yoki bir nechta mustaqil o‘zgaruvchilardan foydalangan holda bog‘liq o‘zgaruvchini prognoz qilishda qo‘llaniladi. Korrelyatsiya o‘zgaruvchilar orasidagi o‘zaro bog‘liqlikning kuchi va yo‘nalishini aniqlashga yordam beradi, regression esa ma’lumotlarni prognozlash uchun samarali vosita hisoblanadi. Shuningdek, korreylatsion va regression tahlilning iqtisodiyot, tibbiyot va ijtimoiy sohalarda qanday qo‘llanilishi misollar bilan ko‘rsatilgan. Bu metodlar ilmiy tadqiqotlar va amaliy sohalarda ma’lumotlarni tahlil qilishda keng foydalanilmoqda. Korrelyatsion tahlil qanchalik o‘xshashligini aniqlash imkonini beradi. Umumiy ko‘rinishda gipotezani qabul qilish korrelyatsiya mavjudligi A o‘zgaruvchi qiymatining o‘zgarishi B qiymatining mutanosib o‘zgarishi bilan bir vaqtda sodir bo‘lishini anglatadi: agar ikkala o‘zgaruvchi ham o‘sib borsa, u holda korrelyatsiya musbat, agar bir o‘zgaruvchi o‘sib, ikkinchisi kamaysa - korrelyatsiya manfiy bo‘ladi. Korrelyatsiyalarni o‘rganishda bitta tanlanmadagi ikkita ko‘rsatkich o‘rtasida qandaydir bog‘liqlik bor-yo‘qligini aniqlashga harakat qilinadi

© A.X. Nishanov, A.T. Tursunov, D.Z. Narzullayev, E.D. Rahmonov

(masalan, bolalarning bo‘yi va vazni yoki IQ darajasi va maktabdagi o‘zlashtirish o‘rtasida) yoki ikki xil namuna o‘rtasida (masalan, juftliklarni taqqoslashdaegizaklar) va agar bu bog‘liqlik mavjud bo‘lsa, o‘sish kuzatiladimi bir ko‘rsatkichning o‘sishi (musbat korrelyatsiya) yoki kamayishi bilan taqqoslanadi. Quyidagi korrelyatsion tahlilga misollar keltirilgan (2-rasm)

1. qat’iy musbat korrelyatsiya;
2. musbat korrelyatsiya;
3. kuchsiz musbat korrelyatsiya;
4. nol korrelyatsiya;
5. manfiy korrelyatsiya;
6. qat’iy manfiy korrelyatsiya;
7. nochiziqli korrelyatsiya;
8. nochiziqli korrelyatsiya.



2-rasm. Korrelyatsion tahlil

III. Xulosa

Bog‘liqliklarni tiklash masalalari 1794-yilda K. Gauss tomonidan eng kichik kvadratlar usuli ishlab chiqilganidan beri 200 yildan ortiq vaqt davomida faol o‘rganilmoqda. Matematik statistikada shu vaqtdan boshlab ma’lumotlarni tahlil qilishning juda ko‘p usullari va vositalari ishlab chiqildi. Ushbu maqolada hozirda vaqtda jahonda amaliy fanning barcha sohalarida - fizika, biologiya, iqtisodiyot, psixologiya va boshqalarda ma’lumotlarni tahlil qilish uchun eng keng qo‘llaniladigan usullar tavsiflab o‘tildi. Hozirgi vaqtda kompyuterlar matematik statistikada katta rol o‘ynaydi. Ular hisoblashlar uchun ham, imitatsion modellashtirish uchun ham (xususan, tanlanmalarni ko‘paytirish usullarida va asimptotik natijalarning yaroqliligini o‘rganishda) qo‘llaniladi.

IV. Foydalanilgan adabiyotlar

1. Камилов М.М.. Фазилов Ш.Х.. Мирзаев Н.. Раджабов С.С. Модели алгоритмов распознавания, основанных на оценке взаимосвязанности признаков. – Ташкент.: Изд–во “Фан ва технология”, 2020. – 148 с.
2. Турсунов А.Т., Нишанов А.Х. Применение методов оцифровки в задачах распознавания образов. ТАТУ хабарлари 3 (67) 2023 ISSN 2010–9857. 79–93–б.
3. Гнеденко Б.В. Курс теории вероятностей: Учебник. – Изд. 6–е изд, исправл– М.: Эдиториал. Гл. ред. физ.–мат. лит., 2011.– 448 с.
4. Клейн Ф. Лекции о развитии математики в XIX столетии. т. 2. – Москва–Ижевск, институт компътерных исследований 2003. – 240 с.
5. Плошко Б.Г., Елисеева И.И. История статистики: Учеб. пособие. – М.: Финансы и статистика. 1990. – 295 с.
6. Орлов А.И. Вероятностно–статистические модели данных – основа методов прикладной статистики / Заводская лаборатория. Диагностика материалов. 2020. Т.86. № 7. – с. 5–6.
7. Kevin P. Murphy. Machine Learning: A Probabilistic Perspective MIT Press, 2012 . – pp.1104.
8. Полуакова В.В., Шаброва Н.В.. Прикладнаya статистика: методы анализа эмпирической информатции. Учебно–методическое пособие. РФ, Екатеринбург : Изд–во Урал. ун–та, 2020. – 188 с
9. Tryon R.C. Cluster Analysis. NY.: McGraw-Hill, 1939.

© A.X. Nishanov, A.T. Tursunov, D.Z. Narzullayev, E.D. Rahmonov