

UDK: 01.00.00

KATTA MA'LUMOTLAR(BIG DATA) BILAN ISHLASHDA MATEMATIK USULLARNING QO'LLANILISHI

Yusupov Dilmurod Tasbaltayevich, TATU Nurafshon filiali axborot
texnologiyalari kafedrası katta o'qituvchisi

e-mail: cazinoroyal1987@gmail.com

Malikova Ozodaxon Ravshanbek qizi, TATU Nurafshon filiali kompyuter
injiniringi fakulteti talabasi

Annotatsiya: Bu maqolada katta ma'lumotlarni qayta ishlash matematik usul va algoritmlarni qo'llashning aniq misollar bilan keltirilgan. Katta ma'lumotlar bilan ishlashda matematik usullar ya'ni graflar nazariyasi, extimollik va statistika, chiziqli algebra, diskret matematika usullarini qo'llash ko'rib chiqilgan.

Kalit so'zlar: big data, matematik usul, timsollarni tanish, katta ma'lumotlar, algoritm.

Kirish

Bugungi dunyo raqamli ma'lumotlarga asoslangan dunyo bo'lib, hayotimizning barcha jabhalariga kirib bormoqda texnologiyalarning tez rivojlanishi natijasida ma'lumotlar katta hajmda ishlab chiqarilmoqda. Buning natijasida axborot texnologiyalar sohasida yangi "katta ma'lumotlar" (big data) tushunchasi paydo bo'ldi. "Big Data" – bu juda tez sur'atlarda o'sib borayotgan katta hajmdagi tizimli va tizimsiz raqamli ma'lumotlar to'plamidir.[1] «Big Data» biznes jarayonlarini optimallashtirish va avtomatlashtirish, moliya, hukumat va sog'liqni saqlash kabi sohalarda o'zgarishlar yaratish uchun ajoyib imkoniyatlarni taqdim etadi, shuningdek, firibgarlikning oldini olish, tabiiy ofat yuz berganda resurslarni taqsimlash yoki sog'liqni saqlash sohasida, to'plangan ma'lumotlarga asoslangan eng samarali qarorlarni qabul qilinishini ta'minlash maqsadida axborotni saqlash va qayta ishlashning innovatsion usullarini o'z ichiga oladi. Katta ma'lumotlar bilan ishlashda matematik usullarni qo'llash zamonaviy dunyoda hal qiluvchi ahamiyatga ega. Masalan, mashinani o'rganish va sun'iy intellekt sohalarida aniqroq va samarali modellarni ishlab chiqish uchun chiziqli algebra, statistika va ehtimollik kabi matematik tushunchalar qo'llaniladi. Bundan tashqari, matematik usullar ishonchli kriptografiya algoritmlarini yaratish va tarmoqlar xavfsizligini ta'minlash imkonini beradi. Shaharsozlik sohasida transport muammolarini hal qilish va infratuzilmani optimallashtirish uchun matematik usullardan foydalaniladi.[2] Shunday qilib, matematik usullarni tushunish va

qo'llash axborot texnologiyalar va katta ma'lumotlarni qayta ishlash sohasida muvaffaqiyatli ishlash uchun zaruriy bilim va ko'nikmadir.

Adabiyotlar tahlili

Katta ma'lumotlarni qayta ishlash va tahlil qilishda matematik usullar juda ko'p. Ulardan keng tarqalganlari:

1.Graflar nazariyasi: ijtimoiy tarmoqlar, transport tarmoqlari va boshqa juda ko'p murakkab tarmoqlardagi ma'lumotlar o'rtasidagi munosabatlarni tahlil qilish uchun ishlatiladi.

2.Ehtimollik va statistika: Katta hajmdagi ma'lumotlarni tahlil qilish, qonuniyatlarni aniqlash, ma'lumotlarni tasniflash, matnni tahlil qilish, timsollarni tanib olish va boshqa mashinani o'rganish vazifalari uchun ishlatiladi.

3.Chiziqli algebra: tasvirni qayta ishlash texnikasi, kompyuter grafikasi, ma'lumotlarni tahlil qilish va mashinani o'rganishda qo'llaniladi.

4.Diskret matematika va algebra: algoritmlarni ishlab chiqish va ma'lumotlarni qayta ishlash uchun ishlatiladi.

5.Modellar nazariyasi: so'zlashuv tillari va ularning talqinlari yoki modellari o'rtasidagi munosabatlar bilan shug'ullanadigan matematik mantiqning bir bo'limi.

Ushbu usullar ma'lumotlarni tahlil qilish, ularni qayta ishlash uchun optimal algoritmlarni ishlab chiqish, shuningdek, ularni turli sohalarida qo'llash imkonini beradi.[3]

Muhokama

Endi katta ma'lumotlarni qayta ishlashda matematik usullarni qo'llashni aniq misollarda ko'rib chiqamiz. Hozirgi kunda yurtimizda keng ommalashgan Яндекс.Такси xizmatini olsak bir manzildan boshqa manzilga borish yo'lini ko'rsatib beradi. Ya'ni borish yo'lini svetafor, qurilish, tirbandliklarni hisobga olgan holda topish kerak. Bu masalani matematik usullardan graflar nazariyasi orqali yechimini topish mumkin. Яндекс kompaniyasi ham aynan shu usuldan foydalanadi. Eng mashhur qidiruv tizimlari Google, Bing, Yahoo bir qancha matematik qidiruv algoritmlaridan foydalanadi. Masalan: TF-IDF(Term Frequency-Inverse Document Frequency) algoritmi. Bu algoritm berilgan so'rov bo'yicha hujjatning dolzarbligini aniqlash uchun ishlatiladi. TF-IDF statistik o'lchovi qanday formula bilan ishlashini ko'rib chiqaylik. Dastlab ba'zi belgilarni aniqlaymiz:

N - bizning ma'lumotlar to'plamimizdagi hujjatlar soni

d - bizning ma'lumotlar to'plamimizdan berilgan hujjat

D - barcha hujjatlar to'plami

w - hujjatdagi berilgan so'z

Birinchi qadamda hujjat chastotasi $tf(w,d)$ hisoblanadi.

$$tf(w,d) = \log(1 + f(w,d)) \quad (1)$$

Bu yerda $f(w,d)$ – d hujjatdagi w so'zining chastotasi.

Ikkinchi qadamda teskari hujjat chastotasi hisoblanadi.

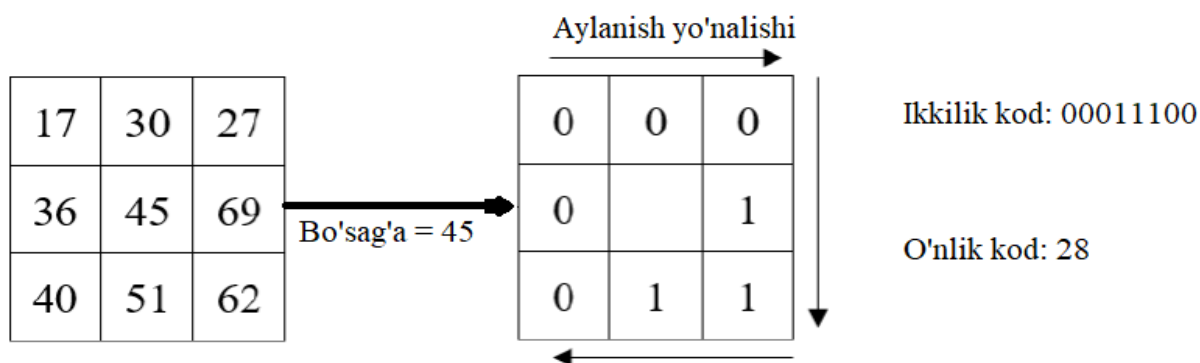
$$idf(w, D) = \log \left(\frac{N}{f(w, D)} \right) \quad (2)$$

Yakuniy qadam TF-IDF dolzarblik ballini quyidagi formula bo'yicha hisoblanadi:

$$tfidf(w, d, D) = tf(w, d) * idf(w, D) \quad (3)$$

Yana bir keng qo'llanilayotgan kamera, telefon, turniket orqali yuzni tanib olish texnologiyasi ham matematik usul va algoritmlar asosida amalga oshiriladi. Yuz tasviridagi lokal sohalar, ya'ni uning asosiy komponentlari (masalan, ko'zlar, burun, og'iz) joylashgan sohalar orqali ko'rsatilgan yuz komponentlarini tavsiflovchi belgilarni aniqlash masalasi yechiladi. Bunda, har bir komponent o'ziga xos belgilar to'plami bilan tavsiflanadi, ular turli xususiyatlarga ega bo'lishi mumkin. Belgilar soni me'yordan ortiqcha bo'lsa, tanib olish xatosi ehtimolini sezilarli darajada oshirmasdan, ularning sonini minimallashtirish maqsadga muvofiq. Bunday holatlar yuzaga kelganda, informativ belgilar to'plamini olish uchun turli usullardan, xususan, evristik usullardan foydalanish mumkin. Shuni ta'kidlash kerakki, dastlabki belgilar to'plamini shakllantirish muammosi aniq bir yechimga ega emas, shuning uchun u empirik tarzda hal qilinadi. Bunday holda, ob'ektning vizual tasviridagi o'zgarishlarga invariant bo'lgan bunday belgilarni tanlash maqsadga muvofiqdir. Invariantlik deganda tasvirda keltirilgan yuzning ko'rib chiqilayotgan yuz komponentlari tasvirlarini olish sharoitlariga (rakurs, yoritilganlik va h.k.) bog'liq bo'lmaslik tushuniladi.[5]

Yuz tasviridan lokal teksturaviy belgilarni ajratib olish uchun lokal binar shablonlar (LBSh) usuli keng qo'llaniladi. Tadqiqotlar shuni ko'rsatdiki, LBSh yorug'lik sharoitlari va tasvir aylanishidagi o'zgarishlarga nisbatan qisman invariant hisoblanadi. LBSh – bu tasvir pikseli va uning atrofidagi axborotning ikkilik ko'rinishdagi tavsifi. Tasvirning ma'lum bir nuqtasida bazaviy LBShni hisoblash uchun uning sakkizta qo'shni pikselidan foydalaniladi va markaziy pikselning intensivlik qiymati bo'sag'aviy qiymat sifatida olinadi (1-rasm). Intensivlik qiymati bo'sag'aviy qiymatidan katta yoki unga teng bo'lgan piksellar birga teng qiymatlarni oladi, qolganlari esa nolga teng bo'lgan qiymatlarni oladi. Shunday qilib, operatsiya natijasi sifatida ushbu pikselning atrofini tavsiflovchi sakkiz bitli binar kod hisoblanadi.



1-rasm. Bazaviy LBShni hisoblashga misol.

LBShni hisoblash uchun [6] ishda mualliflar quyidagi ifodani taklif qilganlar:

$$LBSh(x, y) = \sum_{l=0}^{P-1} s(z_l) 2^l; \quad s(z_l) = \begin{cases} 1, & z_l \geq 0; \\ 0, & z_l < 0; \end{cases}$$

$$z_l = g_l - g_c,$$

2-rasm. LBSh ni hisoblash funksiyalari.

Bu yerda (x, y) – markaziy piksel koordinatas i , P – markaziy piksel atrofidagi qaralayotgan pikseller soni, l – qaralayotgan piksel raqami ($l = \overline{0, P-1}$), g_c – markaziy piksel yorqinlik qiymati, g_l – qaralayotgan piksel yorqinlik qiymati. Real vaqtda yuzni tanib olish tizimlarini qurishda belgilar fazosini shakllantirish masalasini hal etishda LBSh algoritmi takomillashtirilgan variantlaridan foydalanish tezkorlik va samaradorlik jihatdan maqsadga muvofiq hisoblanadi.

Xulosa

Shunday qilib katta ma'lumotlar bilan ishlashda matematik usullar, algoritmlar katta yordam beradi. Bundan buyon raqamli ma'lumotlar ko'payib boraveradi. Bu esa yanada tezroq, optimal yechim topish zaruriyatiga olib keladi. Buning uchun matematik usullarni chuqurroq, har bir soha uchun optimal yechimlarni topish, ko'plab matematik modellarni sinashga to'g'ri keladi. Katta ma'lumotlar barcha sohaga kirib bormoqda. Ko'plab sohalarida bu usul va algoritmlar muammolarni muvaffaqiyatli hal etmoqda. Lekin hali yechimi topilmagan muammolar ham talaygina. Bunga tibbiyot sohasi, zilzila, suv toshqini bashoratlash, xavf-xatarni aniqlash, zararni minimallashtirish, iqtisodiyot va qishloq xo'jaligini rivojlantirish misol bo'la oladi. Bu holat shuni anglatadiki har bir jarayonni alohida o'rganish, ko'plab matematik usullarni qo'llab eng optimal natija beradigan usulni tanlab olish zarur.

Foydalanilgan adabiyotlar

1. Tuyg'un Asatov, "Big Data texnologiyasi va undagi ma'lumotlarni tahlil etish usullari"// <https://ict.xabar.uz/uz/startap/big-data-texnologiyasi-va-undagi-malumotlarni-tahlil-etish> (2022).
2. Victor I. Stepanov, Сафронова Т.И. Применение математических методов в исследовании процессов природообустройства// http://www.researchgate.net/publication/330193490_PRIMENENIE_MATEMATICHESKIH_METODOV_V_ISSLEDOVANII_PROCESSOV_PRIRODOOBUSTROJSTVA
3. Сюй Юань. (2022). Взаимосвязь между математическими алгоритмами и большими данными. ЖУРНАЛ Инновации и инвестиции// <https://cyberleninka.ru/article/n/vzaimosvyaz-mezhdu-matematicheskimi-algoritmami-i-bolshimi-dannymi>
4. Narziyev Nosir Baxshilloyevich, Qidiruv tizimlari algoritmlarining tahlili// <https://zenodo.org/records/7993938> 2023 yil
5. F.N. Numonov Sh.S. Kaxarov, Yuzni tanib olish tizimlarida byelgilar fazosini shakllantirish masalasi, Qo'qon universiteti xabarnomasi// https://www.herald.kokanduni.uz/index.php/public_html/article/download/583/369 2023 yil.
6. Ojala T., Pietikainen M., Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). – 2002. – Vol. 24. – No. 7. – P. 971–987.