

KATTA MA'LUMOTLAR OMBORIDA MA'LUMOTLARINI QAZIB OLISH UCHUN
YANGI MA'LUMOTLARNI KLASTERLASH YONDASHUVI

Xojalbibi To'xtaovunova
Andijon Davlat Universiteti
Kompyuter ilmlari va Dasturlash
Texnologiyalari yo'nalishi
2-bosqich magistranti

Katta ma'lumotlar (big data) juda xilma-xil ma'lumotlar bilan ishlaydi, ularning aksariyati qat'iy tuzilishga ega emas, shuning uchun bu ma'lumotni tahlil qilish vazifasi ancha murakkab va qiziqarli. Kerakli ma'lumotlar namunasini tahlil qilish natijasida siz ob'ekt yoki ob'ektlar guruhi haqida batafsil ma'lumot olishingiz mumkin.

Zamonaviy dunyoda Big Data IT, Internet, modellashtirish, biznes va inson faoliyatining boshqa sohalarida keng qo'llaniladi. Katta ma'lumotlarning odatiy misoli - bu turli xil jismoniy eksperimental qurilmalardan, masalan, doimiy ravishda juda katta hajmdagi ma'lumotlarni ishlab chiqaradigan Katta adron kollayderidan keladigan ma'lumotlar. O'rnatish doimiy ravishda katta hajmdagi ma'lumotlarni ishlab chiqaradi, ular yordamida olimlar turli xil muammolarni hal qilishga harakat qilishadi [6]. Katta ma'lumotlardan foydalanishning yana bir misoli - bu "berilgan" algoritmlardan o'rganishga asoslangan barcha qidiruv tizimlari. Ular foydalanuvchining qidiruv so'rovlari tarixini, u tashrif buyurgan sahifalarini, foydalanuvchini qiziqtiradigan so'rovlarni qidirish yoki ularni to'ldirish, shuningdek, avvalo ular eng foydali deb hisoblagan natijalarni berish, uning afzalliklarini eslab qoladi va olingan ma'lumotlar yordamida ular taklif qilishlari mumkin. Ammo Internet faoliyati tahlili natijalaridan foydalanish shu bilan tugamaydi. Jinoyatchilarni qidirish, teraktlarning oldini olish, shuningdek, o'z joniga qasd qilish tendentsiyalarini faqat Internet-syorfing tarixi bilan aniqlash allaqachon qisman sinovdan o'tkazilmoqda.

Biznesda Big Data Internet va ITdagidan kam emas. Xaridorning afzalliklari kabi kerakli ma'lumotlarni hisobga olgan holda, do'konlar ularga yoqadigan kiyimlarni yoki ular aniq qiziqadigan hikoyasi bo'lgan kitoblarni taklif qilishlari mumkin.

Bularning barchasini, shuningdek, boshqa vazifalarni turli xil algoritmlar bilan hal qilish mumkin, ulardan biri "Klaster tahlili" bo'lib, u ma'lum bir tuzilishga ega ob'ektlar to'plamini ba'zi birlashtirilgan belgilarga ko'ra kichik to'plamlarga bo'lish uchun xizmat qiladi. Uning asosiy xususiyati - qismlarga ajratish uchun qat'iy belgilangan parametrlar to'plamining yo'qligi. Bo'linish xususiyatlar to'plamiga ko'ra sodir bo'ladi, shuning uchun bir xil to'plamning ob'ektlari taxminan bir xil xususiyatlarga ega.

Yuqoridagilarni umumlashtirib, biz katta ma'lumotlarni klaster tahlili bugungi kunning dolzarb vazifasidir, degan xulosaga kelishimiz mumkin.

Foydalanuvchilarga Internetga kirishni ta'minlaydigan proksi-serverning ishlashi davomida har bir foydalanuvchining Internetdagi faoliyati to'g'risidagi ma'lumotlar xavfsizlik va tizim monitoringi uchun qayd etiladi.

Proksi-server - bu boshqa serverda joylashgan mijozlar va resurslar o'rtasida oraliq qatlam yoki vositachi vazifasini bajaradigan server. Global tarmoqdan mavjud bo'lgan har qanday ma'lumot resurs bo'lishi mumkin, chunki ularning barchasi turli serverlarda saqlanadi.

Internetdagi faoliyat ma'lumotlariga asoslanib, turli vazifalarni hal qilish mumkin: tizimni optimallashtirish, tarqatish orqali resurslar yukni kamaytirish, uning himoyasini yaxshilash, tizimga zarar etkazishi mumkin bo'lgan harakatlarni yoki shubhali faoliyatga ega foydalanuvchilarni kuzatish.

Har bir so'rov jurnali yozuvida foydalanuvchi va keyingi tahlil uchun so'rov haqida etarli ma'lumot mavjud, xususan:

- 1) Foydalanuvchi ma'lumotlari: login va IP-manzil.
- 2) So'rov haqidagi ma'lumotlar: bajarilish vaqti, hajmi, usuli, URL manzili va kontent turi.

Olingan ma'lumotlar to'plamini tahlil qilish orqali proksi-server foydalanuvchilari to'plamini, masalan, afzal ko'rgan sayt mavzulariga ko'ra, ba'zi kichik guruhlariga bo'lish va uzoq vaqt davomida bir-biriga yopishib oladigan va shunga o'xshash xususiyatlarga ega bo'lgan foydalanuvchilar guruhlarini aniqlash mumkin.

Klasterlash (yoki Klaster tahlili) — bu ob'ektlar to'plamini klasterlar deb ataladigan guruhlariga bo'lish vazifasi. Har bir guruh ichida "o'xshash" ob'ektlar bo'lishi kerak va har xil ob'ektlar guruhlar iloji boricha boshqacha bo'lishi kerak. Klasterlash va tasniflash o'rtasidagi asosiy farq shundaki, guruhlar ro'yxati aniq belgilanmagan va algoritm jarayonida aniqlanadi.

Klaster tahlilini umumiy shaklda qo'llash quyidagi bosqichlarga qisqartiriladi:

Klasterlash uchun ob'ekt namunalarini tanlash.

Namunadagi ob'ektlar baholanadigan o'zgaruvchilar to'plamini aniqlash. Agar kerak bo'lsa, o'zgaruvchilar qiymatlarini normallashtirish.

Ob'ektlar orasidagi o'xshashlik o'lchovining qiymatlarini hisoblash.

Shunga o'xshash ob'ektlar (klasterlar) guruhlarini yaratish uchun Klaster tahlil usulini qo'llash.

Tahlil natijalarini taqdim etish.

Natijalarni olgandan va tahlil qilgandan so'ng, tanlangan metrikani va klasterlash usulini maqbul natijaga erishish uchun sozlash mumkin

nulll Klasterlash

1.1 Klasterlash nima

“Klasterlash – ma'lum bir to'plam o'xshashligiga qarab avtomatik ravishda guruhlariga bo'linishi. To'planning elementlari har qanday ob'ektlar bo'lishi mumkin, masalan, ma'lumotlar yoki xususiyat vektorlari. Guruhlarning o'zi ham odatda klasterlar deb ataladi” [4]. Klasterlash ko'pincha shunga o'xshash protsedura, tasniflash bilan taqqoslanadi. Ularning farqi shundan iboratki, tasniflashda natijaviy guruhlar to'plami aniq belgilanadi, klasterlashda esa algoritmning o'zi uning ishlashi jarayonida aniqlanadi.

Klasterlashning kompyuter fanida ham, boshqa sohalarida ham ko'plab amaliy qo'llanilishi mavjud. Klasterlashdan foydalanishga misollar keltiramiz [7]:

- 1) Ma'lumotlarni tahlil qilish
 - Axborot bilan ishlashni soddalashtirish
 - Ma'lumotlarni vizuallashtirish
- 2) Axborotni olish va qidirish
 - Qulay klassifikatorlarni qurish
- 3) Ob'ektlarni guruhlash va tanib olish
 - Shaklni aniqlash
 - Ob'ektlarni guruhlash

Klaster tahlilini quyidagi harakatlar ketma-ketligi sifatida ko'rsatish mumkin [3]:

- 1) Ob'ektlar to'plamini tanlash.
- 2) Ob'ektlarni baholash uchun o'zgaruvchilar to'plamini aniqlash va xarakteristikalar vektorlarini tuzish.
- 3) Mavjud usullardan biri bilan xarakterli vektorlarni normallashtirish.
- 4) Berilgan ko'rsatkich bo'yicha ob'ektlar orasidagi o'xshashlikni aniqlash.
- 5) Ob'ektlar to'plamini o'xshashlik darajasiga ko'ra klasterlarga bo'lish uchun klaster tahlilining tanlangan usulini qo'llash.
- 6) Tahlil natijalarini taqdim etish.

Klasterlash natijalarini tahlil qilgandan so'ng, natijalarni yaxshilash uchun tanlangan parametrlarni, metrikaning yoki klasterlash usulini sozlashingiz mumkin. Ushbu yaxshilanishlar orqali siz eng yaxshi natijaga erishishingiz mumkin.

Xususiyat vektorini ajratib olish

Birinchi qadam - klasterlash jarayonida foydalaniladigan ob'ektlarning xususiyatlarini tanlash. Bu miqdoriy (balandlik, vazn, koordinatalar, hisoblagichlar, ...) va sifat ko'rsatkichlari (rang, holat, kayfiyat, ...) bo'lishi mumkin. Ko'pincha ular miqdoriy xususiyatlar bilan ishlaydi, chunki ular uchun ko'p sonli ko'rsatkichlar qo'llaniladi. Katta xususiyat maydonida klasterlash jarayoni ancha sekin kechadi va uning natijalari har doim ham qabul qilinmaydi. Shuning uchun, xarakteristikalar maydonining katta o'lchami bilan siz ob'ektlarning eng muhim xususiyatlarini qoldirib, uni kamaytirishga harakat qilishingiz kerak. Yaxshiroq natijalarga erishish uchun har bir ob'ektning xossalari to'plamini normallashtirish kerak. Vektorni normallashtirish uni belgilangan o'lchamga keltirishni anglatadi. Har bir normallashtirilgan vektorning xarakteristikalari sobit segmentda yotadi

FOYDALANILGAN ADABIYOTLAR:

1. O.S. Radjabov, A.A. Temirov, M.R. Avezova Kriterii otsenki delovoy aktivnosti predpriyatiya // "Globalashuv jarayonida korporativ boshqaruv: muammo va yechimlar" ilmiy maqolalar to'plami (Respublika ilmiy-amaliy konferensiya materiallar to'plami). Toshkent-2019. -S. 248-250

2. Temirov A., Sohobiddinov A. "Mamlaktimiz ta'lim tizimida axborot kommunikatsiya va innovatsiyon texnologiyalardan foydalangan holda bilim olish" // "WORLD SOCIAL SCIENCE" // 15-16 bet.

3.Кенжабаев, А.Т., Х., Хуррамов А. (2015). КИЧИК БИЗНЕС ВА ХУСУСИЙ ТАДБИРКОРЛИК КОРХОНАЛАРИ РИВОЖЛАНИШИНИ МОДЕЛЛАШТИРИШ УСЛУБИЁТИ. XI Международная научно-практическая интернет-конференция, 1(1), 75-85.

4.Ro'ziyev, A.O., Xurramov, A.X. (2013). Axborottizimlari. 1, 172.