

MIRZO ULUG'BEK NOMIDAGI
O'ZBEKISTON MILLIY UNIVERSITETI



D.T.Muxamedieva, O.N.Abduraxmonov, M.H.Primova

BIG DATA VA MA'LUMOTLAR TAHLILI

O'ZBEKIKISTON RESPUBLIKASI OLIY TA'LIM, FAN VA
INNOVATSIYALAR VAZIRLIGI

MIRZO ULUG'BEK NOMIDAGI
O'ZBEKISTON MILLIY UNIVERSITETI

D.T.Muxamedieva, O.N.Abduraxmonov,
M.H.Primova

BIG DATA VA MA'LUMOTLAR TAHLILI

5330100 – Axborot tizimlarining matematik va
dasturiy ta'minoti ta'lim yo'nalishlari uchun

O'QUV QO'LLAMMA

Toshkent
"Fidokor Yosh Avlod"
2023

.....	4
eti va8
.....	24
ari.....	39
.....	52
.....	58
ozori. 64
.....	89
.....	103
.....	107
....	116
... ..	119
....	122
... ..	148
... ..	151
.. ..	158
. ..	166
. ..	173
.....	177
.....	182
.....	187
.....	196
.....	98
.....	10
.....	15
.....	15
.....	2
.....	0
.....	5

UDK: 004.43(075.8)
BBK: 32.973-018ya73
M 93

Muxamediyeva D.T., Big data va ma'lumotlar tahlili: o'quv qo'llanma / D.T. Muxamediyeva O.N. Abduraxmonov M.H. Primova
Surxondaryo: Fidokor Yosh Avlod, 2023.-248 b.

Ushbu o'quv qo'llanma 5330100 – Axborot tizimlarining matematik va dasturiy ta'minoti ta'lim yo'nalishlari talabalariga Big data va ma'lumotlar tahlili, dasturiy injiniring va berilganlar bazasini boshqarish tizimi fanlaridan foydalanish maqsadida yozilgan bo'lib, konseptual obyektli modeldan foydalangan holda relyasion ma'lumotlar bazasidagi ma'lumotlar ustida turli amallar bajarishni o'rgatadi. Big data va ma'lumotlar tahlili ma'lumotlar bazasi ustida katta hajmli ma'lumotlarni yig'ish, saqlash, arxivlash, ma'lumotlarni foydalanuvchilarga tezlik bilan topib berish, katta hajmli ma'lumotlarni tahlil qilish nazarda tutiladi.

UDK: 004.43(075.8)
BBK: 32.973-018ya73
M 93

Mualliflar:

D.T.Muxamediyeva O'zMU, Hisoblash matematikasi va axborot tizimlari kafedrasini professori
O.N.Abduraxmonov O'zMU, Hisoblash matematikasi va axborot tizimlari kafedrasini o'qituvchisi
M.H.Primova O'zMU, Hisoblash matematikasi va axborot tizimlari kafedrasini doktoranti

Taqrizchilar:

T.F.Bekmuratov TATU huzuridagi "Axborot-kommunikatsiya texnologiyalari ilmiy –innovatsion" markazi akademigini
R.D.Aloev O'zMU, Hisoblash matematikasi va axborot tizimlari kafedrasini professori

ISBN: 978-9943-9119-6-3

©"Fidokor Yosh Avlod" nashriyoti, Toshkent 2023 y

MUNDARIJA

KIRISH	4
1-Mavzu: "Big data va ma'lumotlar tahlili" fanining mazmuni, predmeti va metodi.....	8
2- Mavzu: Ma'lumotlar	24
3- Mavzu: Data Mining metodlari va boskichlari. Data Mining masalalari.....	39
4- Mavzu: Sinflashtirish va klasterlash	52
5- Mavzu: Data Mining qo'llash sohasi. Prognozlash va vizualizatsiya.	58
6- Mavzu: Bank ishi. Sug'urta. Telekommunikatsiya. Marketing. Fond bozori. Bioinformatika. Meditsina. Farmatsevtika	64
7- Mavzu: Klasterli tahlil. OLAP va boshqa ma'lumot saqlagichlar	89
8- Mavzu: Berilganlarni intellektual tahlilining asosiy usullari.....	103
9- Mavzu: Tasniflash masalasi	107
10- Mavzu: Tasniflash jarayoni.....	116
11- Mavzu: Tasniflash masalalarini yechish uchun qo'llaniluvchi usullar	119
12- Mavzu: Noravshan to'plamlar usuli	122
13- Mavzu: Neyron to'rlar usuli. Perseptron.....	148
14- Mavzu: Neyron to'rlar usuli. Xopfild va Xemming neyron to'rlari.....	151
15- Mavzu: Yechimlar daraxtini qurish jarayoni.....	158
16- Mavzu: T ayanch vektorlar tushunchasi	166
17- Mavzu: "Eng yaqin qo'shni" usuli.....	173
18- Mavzu: Bayes taqsimoti	177
19- Mavzu: Bashoratlash masalasi.....	182
20- Mavzu: Tasniflash va bashoratlash usullari.....	187
21- Mavzu: Bashoratlash va vaqt qatorlari.....	196
22- Mavzu: Bashoratlash masalasini yechish.....	198
23- Mavzu: Klasterlash masalasi.....	210
24- Mavzu: Klaster tahlilini tadbqiq etish.....	215
25- Mavzu: Assotsiativ qoidalarni izlash usuli	225
26- Mavzu: Evolyutsion algoritmlar	232
27- Mavzu: Zamonaviy evolyutsion algoritmlar	240
Adabiyotlar ro'yxati.....	246

DENOV TADBIRKORLIK
VA PEDAGOGIKA
INSTITUTI ARM
№ 34283

Kirish

Ko'plab amaliy masalalarni yechish va amaliyotda qo'llaniladigan tizimlarini yaratish bo'yicha so'nggi o'n yildagi tajribalar aynan intellektual texnologiyalar avtomatlashgan boshqaruv tizimlarini yaratishda nisbatan asosli hamda iqtisodiy samarador ekanligini ko'rsatdi.

Ma'lumotlarni intellektual tahlil qilishning asosiy masalalari quyidagilar hisoblanadi:

- klassifikatsiya – kiruvchi vektor (obyekt, hodisa va jarayonlar) ni avvaldan ma'lum sinflardan biriga tegishiligidini aniqlash;
- klasterizatsiya — kiruvchi vektorlar to'plamini bir-biriga "o'xshashlik" darajasi bo'yicha guruh (klaster) larga ajratish;
- regressiya, xususan bashoratlash masalasi – uzluksiz kiruvchi va chiquvchi o'zgaruvchilarni o'zaro bog'liqliklarini o'rnatish;
- ketma-ket shablonlar – o'zaro bog'langan vaqtli hodisalar qonuniyatlarini o'rnatish, ya'ni "Agar X hodisa ro'y bersa, u holda ko'rsatilgan vaqt o'tgandan so'ng Y hodisa sodir bo'ladi" ko'rinishidagi bog'lanishlarni aniqlash;
- og'ishlarni tahlil qilish – sezilarli xususiyatga ega bo'lmagan shablonlarni aniqlash;
- bashoratlash – tahlil qilinayotgan jarayonlarning rivojlanish xususiyatlarini aniqlash;
- vizuallashtirish – tahlil natijalarini grafik, diagramma va hokazolar ko'rinishida aks ettirish.

Global optimallashtirishning xulq-atvorli usullari, ayniqsa gibrud usullari ko'plab erkin parametrlarga ega va ularning samaradorligi esa ko'pincha mazkur parametrlarning qiymatlariga bog'liq bo'ladi. Odatda ushbu parametrlar qiymatlarini aniqlash qoidalari bo'lmaydi. Shuning uchun xulq-atvorli usullar samaradorligini oshirishning asosiy yo'li bu ularning erkin parametrlari qiymatlarini aniq bir optimallashtirish masalasi yoki optimallashtirish masalalarining ma'lum bir sinfi xususiyatlariga moslashuvchi algoritmlarni ishlab chiqish, ya'ni xulq-atvorli usullarini metaoptimallashtirish hisoblanadi.

Metaoptimallashtirish algoritmlari tasnifi [1] ishda batafsil bayon etilgan. Masalan, Iben (*A.E.Eiben*) tasnifiga ko'ra metaoptimallashtirish algoritmlari parametrlarni sozlash (*parameter tuning*) va parametrlarni boshqarish

(*parameter control*) algoritmlariga ajratiladi [1]. Parametrlarni sozlash algoritmlari bir marta (*one-time parameter tuning*) va uzluksiz sozlanuvchi algoritmlarga ajratiladi. Tanlangan optimallashtirish usulini amalga oshiruvchi dasturni bir yoki boshqa sinf optimallashtirish masalalari uchun erkin parametrlarning turli qiymatlarida ko'p marta bajarish orqali ushbu parametrlarning eng yaxshi natija beruvchi qiymatlarini aniqlash bir marta sozlanuvchi algoritmlarning g'oyasi hisoblanadi. Mazkur g'oya gibrud usullar parametrlari qiymatlarini sozlashda keng qo'llaniladi.

Quyida gibrudlashning keng tarqalgan yondoshuvlari keltirilgan.

PSO (*Particle Swarm Optimization* – zarralar galasi bo'yicha optimallashtirish) usuli samarador va yuqori darajada ishlab chiqilganligi uchun ko'plab gibrud usullarni qurishda tayanch usul sifatida keng qo'llanilmoqda.

PSO-EO usuli [2] ekstremal optimallashtirish (*Extremely Optimization, EO*) va kanonik PSO usullarining gibrud hisoblanadi [3].

PSO va zaryadlangan zaryadlar usullarini gibrudlash [4,5] ishlarda keltirilgan.

Kombinirlashgan *IBPSO* usuli [6] ishda taklif etilgan bo'lib, u binar PSO va sun'iy immun tizimi (*Artificial Immune System, AIS*) usullarini birlashmasidir.

Qidiruvni diversifikatsiyalash va ko'p ekstremalli murakkab funksiyaning global ekstremumini lokallashtirish ehtimoligini oshirish uchun [5] ishda *PSO* va ancha mashhur bo'lgan differensial evolyutsiya (*Differential Evolution, DE*) usullarining *PSO-DE* gibrud usuli taklif etilgan. *PSO* usuli va sun'iy immun tizimi gibrudlashdan hosil qilingan *HPSI* usuli [4,5] ishda keltirilgan. *PSO* va tillaqo'ng'iz (*Glow-worm Swarm Optimization, GSO*) usullarini gibrudlashdan hosil bo'lgan *BSO* gibrud usuli [6] ishda batafsil bayon etilgan. *PSACO* gibrud usuli esa *PSO* va chumolilar kolonasi (*Ant Colony Optimization, ACO*) usullarini gibrudlashdan hosil qilingan [7]. *PSO* va begona o'tlar (*Invasive Weed Optimization, IWO*) usullarining gibrud [8] ishda qarab chiqilgan.

Noaniqlik sharoitida yechiladigan gibrud intellektual tizim masalalari tahlili ushbu masalalar yechimlarini sust shakllangan yechimlarni qabul qilish asosida olinishini ko'rsatdi.

Sust shakllangan yechimlar deb noaniq xususiyatga ega, to'liqmas, ziddiyatli, noravshan va qisman rost bo'lgan noaniq sharoitlarda qabul

qilinadigan yechimlarga aytiladi.

Tadqiq etilayotgan muammoli sohaning o'ziga xos xususiyati nazariy asoslarning cheklanganligi va hisoblash modellarining haqiqiy ma'lumotlar bilan mutanosibligi hisoblanadi. Shuning uchun amaliy masalalar ko'pincha tajribaviy ma'lumotlar asosida shakllantirilgan jadvallar orqali qurilgan modellar yordamida yechiladi. Bunda muayyan holatlarda vaziyatga ta'sir etuvchi barcha omillarni inobatga olish va ma'lumotlarni to'plash kabi muammolar paydo bo'ladi. Shuning uchun ko'p hollarda bir-birini takrorlovchi ma'lumotli yoki kam informativ belgili to'liq bo'lmagan axborotlarni qayta ishlashga to'g'ri keladi.

Qarorlarni qabul qilishning gibrid intellektual tizimlari, jumladan, sust shakllangan va murakkab jarayonlarni boshqarish masalalarini yechish usullari keng qamrovli masalalar sinfidan biri bo'lib, ularga qaror qabul qiluvchi har bir shaxs duch keladi [8, 9]. Ularning yechimlari esa qaror qabul qilish jarayonida an'anaviy matematikaning qayta baholanmagan va o'rmini pasaytirmagan holda zamonaviy matematik vositalar hamda intellektual axborot texnologiyalari imkoniyatlaridan to'g'ri foydalanishni talab qiladi. Murakkab tizimlarda qarorlarni qabul qilish jarayoni katta hajmli turli xil axborotni qayta ishlash va tahlil qilish orqali amalga oshiriladi. Insonning axborotni idrok etish va uni qayta ishlashda cheklangan imkoniyati qabul qilinadigan qarorlarni yetarli darajada samaradorligini ta'minlay olmaydi. Shuning uchun uni intellektual imkoniyatlarini kuchaytirish qarorlarni qabul qilishni qo'llab-quvvatlash intellektual tizimlari yordamida amalga oshiriladi.

Sust shakllangan jarayonlar holatini baholash modellarini qurishda quyidagi algoritmlardan foydalaniladi:

- o'z-o'zini o'qituvchi algoritmlar va mashinali o'qitish. Modellarini qurishning adaptiv algoritmlari: qaror daraxtlari, neyron tarmoqlari, o'z-o'zini tashkil etuvchi xaritalar, assotsiativ qoidalar. Bulardan foydalanish juda qulay bo'lib, tahlil qiluvchi faqat bog'liqliklar mavjudligi haqidagi farazni shakllantiradi, tizim esa mavjud ma'lumotlar asosida modellarini avtomatik quradi;

- vaqtli qatorlar modellarini qurish va tahlil qilish. Odatda mazkur tipdagi algoritmlar ko'pincha vaqtli qatorlarni bashoratlash masalasini yechishda qo'llaniladi;

- og'ishlarni tahlil qilish. Tahlil qilinadigan jarayonlar modellari

qurilgandan so'ng og'ishlar aniqlanadi va kuchli harakterga ega bo'lmagan belgilar aniqlanadi. Og'ishlarni tahlil qilish mexanizmiga o'xshash mexanizmlar diqqatni qaratish zarur bo'lgan hodisalarni avtomatik aniqlash va umumiy qonuniyatga tushmaydigan andozalarni topish imkonini beradi;

- modellar sifatini baholash. Mazkur tipdagi algoritmlar modellashtirish natijalarini taqqoslashda qo'llaniladi. Bunda taqqoslash sifatning rasmiy mezonlari va ekspert bilimlari asosida amalga oshiriladi;

- vizuallashtirish. Bu qulay aks ettirish mexanizmi bo'lib, natijalarni izohlash jarayonini yetarli darajada soddalashtirishga va ekspertlarni tahlil natijalariga bo'lgan ishonchini oshirishga xizmat qiladi.

Tasniflash masalalari uchun "o'qituvchi bilan o'qitish" o'ziga xos bo'lib, unda kiruvchi va chiquvchi vektorlarni o'zida mujassamlashtirgan tanlanma asosida (o'qitish) model quriladi.

Sinflashtirish va birlashtirish masalasi uchun "o'qituvchisiz o'qitish" qo'llanilib, unda chiquvchi parametrga ega bo'lmagan tanlanma asosida model quriladi.

Ma'lumotlarni intellektual tahlil qilish masalasini yechish quyidagi bosqichlarda amalga oshiriladi:

- 1- farazlarni shakllantirish;
- 2- ma'lumotlarni yig'ish (to'plash);
- 3- ma'lumotlarni tayyorlash (filtrlash);
- 4- modelni tanlash;
- 5- o'qitish algoritmi va model parametrlarini tanlash;
- 6- modelni o'qitish (modelning qolgan parametrlarini avtomatik topish);
- 7- o'qitish sifatini tahlil qilish, bunda agar tahlil qoniqarli bo'lmasa, u holda 5- yoki 4-bosqichga o'tish amalga oshiriladi;
- 8- hisoblangan qonuniyatlarni tahlil qilish, bunda agar tahlil qoniqarli bo'lmasa, u holda 1-, 4- yoki 5-bosqichga o'tish amalga oshiriladi.

Ma'lumotlarni intellektual tahlil qilish jarayoni yuqoridagi kabi ketma-ket bosqichlar orqali ifodalanishi mumkin: erkin qidiruv (jumladan, validatsiya) → bashoratli modellashtirish → istisnolar tahlili.

1. "BIG DATA VA MA'LUMOTLAR TAHLILI" FANINING MAZMUNI, PREDMETI VA METODI

Kalit so'zlar: Ta'lim, katta hajmli ma'lumotlar, Data Mining, raqamli savodxonlik, bilimni baholash, konfidentsial ma'lumotlar, bulutli saqlash, Blockchain, sun'iy intellekt va Deep Learning, Statistika.

Katta hajmli ma'lumot (inglizcha: Big Data) — katta hajmdagi ma'lumotlarga nisbatan qo'llanadigan termin hisoblanadi (odatda terabayt, ekzabayt va petabayt darajasida). Ushbu ma'lumotlarni amaldagi va yangi yaratilayotgan texnologiyalar yordamida to'plash, taqsimlash, o'zgartirishlar kiritish, tahlil qilish, saqlash va vizual shaklda tasvirlash imkoni bo'lishi kerak. Ma'lumotlar miqdorining tez o'sishiga so'nggi yillarda kuzatilayotgan texnologik taraqqiyot va buning natijasi hisoblanmish mashina tomonidan yaratilayotgan ma'lumotlarning ko'payib borishi sababchidir. RFID-texnologiyasining keng tarqalishi, elektron tranzaksiyalarning moliya sohasida keng qo'llanilishi, ilmiy-tadqiqot muassasalarining to'plagan natijalari, web-texnologiyalarining keng tarqalishi, telekommunikatsion tizimlardagi protokollar shular jumlasidandir. Big Data aslida ilmiy-tadqiqotga aloqador sohalarida avvaldan ham mavjud bo'lgan bo'lsa-da, lekin so'nggi yillardagina ushbu fenomen ko'proq tilga olinadigan bo'ldi. Buning sababi shundaki, katta miqdordagi ma'lumotlarni tahlil qilish endi iqtisodiy tashkilotlar tomonidan ham keng qo'llanila boshlanishi va buning oqibatida ma'lumotlar tahlilining raqobatbardoshlikni oshirish va samaradorlikni kuchaytirish kabi muhim masalalarda asosiy o'rinni egallaganidir. Big Data odatiy Bosh Informatson Mutaxassis (Chief Information Officer - CIO) uchun qo'shimcha vazifalar va talablar yaratishdan tashqari, yangi Ma'lumot Boshqaruvchisi (Data Steward) va Ma'lumot bo'yicha mutaxassis (Data Scientists) kabi kasblarning paydo bo'lishiga olib keldi.

Katta hajmli ma'lumotlar (Big Data) davri



Yangi texnologiyalar, qurilmalar va aloqa vositalarining tezkor rivojlanishi, ijtimoiy tarmoqlarning keng miqyosda yoyilishi ta'sirida insoniyat tarafidan ishlab chiqilayotgan ma'lumotlar miqdori keskin tarzda oshib bormoqda. Insoniyat yaralgan vaqtdan 2003-yilgacha to'plangan ma'lumotlar 5 milliard gigabayt miqdoriga teng deb hisoblanadi va bu ma'lumotlar diskka yozilgan holatda butun bir futbol maydonini egallashi mumkin bo'ladi.

Ushbu miqdordagi ma'lumotlar 2011-yilda har ikki kunda yaratilgan bo'lsa, 2013-yilga kelib, insoniyat tomonidan 5 milliard gigabayt ma'lumot har 10 daqiqada yaratilmoqda. Bu o'z navbatida, fanda yangi atama — katta hajmli ma'lumotlar (Big Data) tushunchasining yaralishiga sabab bo'ldi. Bu turdagi katta hajmli ma'lumotlar bilan ishlashda, hozirda yangidan-yangi algoritm va texnologiyalar ishlab chiqilmoqda va amaliyotda keng qo'llanilmoqda.

Atrofimizdagi milliardlab to'planib qolgan ma'lumotlarni yig'ish, qayta ishlash va ulardan samarali foydalanish orqali ma'lum holat yoki obyekt to'g'risida tarixiy, joriy va kelajak vaqt uchun muhim bo'lgan bilimlarga ega bo'lishimiz mumkin. Buning uchun dastlab, ushbu ma'lumotlarning manbasini aniqlash muhim amallardan biri hisoblanadi. «Katta hajmli ma'lumotlar» tushunchasi katta o'lchamdagi ishlarni bajara oladigan operatorlarga nisbatan ishlatiladi. Bu asrlar davomida o'rnatilgan tartibni o'zgartirib yuboradi va bizning fundamental bilimlarimizni, qaror qabul qilishni shubha ostiga qo'yadi. Haqiqiy evolyutsiya bu ma'lumotlarni chiqarib beradigan kompyuterlarda emas, balki, ma'lumotlarning o'zlarida va biz ularni to'g'ri qo'llay olishimizdadir. Google bir kunda bir petabaytdan ortiq ma'lumotlarni qayta ishlaydi. Ushbu ko'rsatkich AQSh kongressi kutubxonasida nashr qilinadigan materiallardan 100 barobar ko'proq. Atiga 10 yil avval mavjud bo'lmagan kompaniya—Facebook, soatiga 10 milliondan ortiq yangi rasmlar joylanishi bilan faxrlanishi mumkin. 800 millionta Googlening Youtube xizmati foydalanuvchilari oyiga har soniyada bir soatdan ko'p videolar joylashtiradi.

Bizni o'rab turgan ma'lumotlar haqiqiy hajmining jadalligini, o'sishini hisoblashni ko'p odamlar baholashga harakat qilishgan. Har xil narsalarni hisoblashgani uchun, ular har xil muvaffaqqiyatlarga erishishdi. Eng to'liq tadqiqotni Annenberg nomidagi Janubiy Kaliforniya universiteti qoshidagi kommunikatsiyaga ixtisoslashgan maktabdan Martin Gilbert olib bordi. U

hamma ishlab chiqilgan, saqlangan, yuborilgan narsalarni hisoblashga harakat qildi. Bu faqat kitoblar, rasmlar, elektron xatlar, raqamli musiqa va videolar emas, balki video o'yinlar, qo'ng'iroqlar va avtomobil navigatsion tizimlari hamdir. Undan tashqari u ommaviy axborot vositalari, televideniya va radioning auditoriya kattaliklarini ham hisobga olgan. Uning hisoblariga ko'ra, 2007- yilda qariyb, 2,25 zettabayt ma'lumot yuborilgan va saqlangan. Bu ko'rsatkich 20 yil avvalgisidan (435 ekzabayt) 5 barobar ko'proq.



Agar ommaviy axborot vositalaridan tashqari saqlanayotgan ma'lumotlar ko'rib chiqilsa, qiziq tendensiyalar paydo bo'ladi. 2007- yilda 300 ekzabayt saqlangan ma'lumotlarning 7% analog formatda (qog'oz hujjatlar, kitoblar, fotosuratlar), qolganlari esa raqamli formatda bo'lgan. 2000- yilda bo'lsa raqamli formatda saqlanadigan ma'lumotlar soni dunyodagi saqlanadigan ma'lumotlarning atiga bir chorak qismini tashkil etar edi. Qolgan 3 chorak ma'lumotlar esa qog'oz hujjatlarda, plyonkalarda, plastinkalarda, magnit plastinkalarda va boshqa tashuvchilarda bo'lgan. 2013- yilga kelib, butun dunyoda saqlanayotgan ma'lumotlar 1.2 zettabaytni tashkil etdi va ulardan sonli bo'lmagan ma'lumotlar atiga 2 foizni tashkil etdi. Dunyoda katta ma'lumotlardan katta hajmdagi ma'lumotlarni tahlil qilish, ba'zi hollarda esa hammasini qayta ishlash mumkin. Hamma ma'lumotlardan foydalangan holda biz aniq natijaga ega bo'lishimiz mumkin va kichik hajmdagi ma'lumotlar chegaralangandagi kamchiliklarini ko'rishimiz mumkin. Tanlash yo'li bilan baholab bo'lmaydigan katta ma'lumotlar detallarning segmentlarini va toifalari haqida aniq tasavvurga ega bo'lish mumkin. Katta hajmdagi ma'lumotlarga e'tibor qilgan holda aniqlikka bo'lgan qarashlarimizni kamaytirishimiz mumkin. Yaqin gacha, barcha sonli qurilmalar aniqlikka asoslangan edi. Bizda barcha elektron jadval ustunlari ma'lumotlar tizimi bazasi sonlari singari

bizning talablarga mos ko'rinishda yozilgan degan tushuncha shakllangan edi. Bunday fikr usuli «kichik ma'lumotlar» ustida qo'l keladi. Iloji boricha, hamma yozilgan ma'lumotlarni aniqroq hisoblash uchun kam ko'rsatkichlar o'lchanar edi.

Aniqlik ma'lumotlarning qattiq tekshiruvini talab qiladi. U kichkina hajmdagi ma'lumotlar uchun mos keladi va ba'zi hollarda u juda kerak (masalan, chek yozish uchun bank hisobida etarli miqdorda pul bor yoki yo'qligini tekshirish uchun). Lekin ma'lumotlar olamida 100% aniqlik bo'lishi mumkin emas va kerak ham emas. Agar biz ma'lumotlarning ko'p qismi doim o'zgaradigan ma'lumotlar bilan ishlasak, aniqlik ikkinchi o'ringa tushib qoladi. Katta hajmli ma'lumotlar — inson hayoti uchun dunyoni qurshab olish va hisoblashlardagi harakatlar borasidagi katta olg'a qadamdir. Katta hajmli ma'lumotlar haqida yangicha tushunchalarni shakllantirish uchun 3 ta asosiy holatni sanab o'tsak bo'ladi, bu holatlar bir-biriga bog'langan va bir-birini to'ldirib turadi. 1- barcha ma'lumotlardan qismlab yoki statik tanlanmadan ularni analiz qilish. 2- aniqlikka putur etkazgan holda betartiblik bilan ishlashga tayyor bo'lish. 3- fikrlashni o'zgartirish: qiyin sababiyatdan ko'ra korrelyatsiyalarga ishonish.

Tanlanma tushunchasi minimum materialdan maksimum foydalisini tanlab olish, eng kam berilganlardan katta hajmli tahlil qilishni ko'zlashdir. Endi esa tanlanma eski tushuncha deb qabul qilamiz va katta berilgan ma'lumotlar bilan ishlaymiz. Betartiblik — katta ma'lumotlarning ichki tavsifi emas, balki ishimiz bog'langan obyektiv reallikdir. Material obyektlar singari, ma'lumotlar bahosining iste'mol qilinishi kichiklashmaydi. Ma'lumotlarni qayta va qayta ishlash mumkin. Ular, iqtisodchilar tili bilan aytganda «taqqoslanmaydigan» tovarlar qatoriga kiradi. Shuningdek, moddiy egaliklardan farqli o'laroq, ma'lumot ishlatilishdan to'xtamaydi. Ma'lumotlarning asl bahosi — xuddi okeandagi aysberglardek. Bir qarashda faqat kichik qism ko'rinadi, ayni paytda qolgan qismi suv tagida yotadi, yani ma'lumotlar qimmatligini faqat yuzaki emas, balki har tomonlama qarash kerak. Katta hajmli ma'lumotlar atrofimiz haqida ancha aniqroq prognozlar bera olishni ta'minlaydi. Biz bunday prognozlarga tayyor bo'lmasligimiz mumkin. Chunki bizning dunyoqarashimiz va tizim muassasalarimiz bunday ko'p ma'lumotlarga emas, balki kam ma'lumotlarga moslashtirilgan.

Katta hajmli ma'lumotlar ularni tahlil qilgan, saqlagan va ularni ikkinchi

bor ishlatganlarga yangiliklarni ochib beradi. Chunki saqlanishlar qiymati ham shu darajada tushib ketadi, tahlil vositalari esa, kuchliroq, o'licham va masshtab yig'indisi kunma-kun, soatma-soat o'saveradi. Agar internet xavfsizlik tizimini xavf ostida qoldirgan bo'lsa, katta ma'lumotlar uchun bu masalani yechish imkoni bormi? Bu ularning qorong'u tomoni emasmi? Yana shunday bir xavf borki, biz axborot boshqaruvchilar qurbonlari bo'lishga tavakkal qilayapmiz. Bunda biz ma'lumotlarga va ular to'g'risida chiqayotgan tahlillarga sig'inishni boshlaymiz va oxir-oqibatda, ularni haddan tashqari ko'p ishlatishni boshlaymiz. Agar ulardan oqilona foydalanilsa, katta ma'lumotlar ratsional masalalar yechimi uchun yaxshi vosita hisoblanadi. Agarda ulardan aql bilan foydalanilmasa, u insoniyatning retsesiyasiga hissa qo'shishi, ya'ni mijozlar va ishchilarga noqulayliklar yaratishi yoki odamlarga ziyon etkazishi mumkin.

Dunyo katta ma'lumotlar asri tomon suzayotgani uchun, aholi o'sha «teknik» silkinishga yaqin yurmoqda. Katta ma'lumotlar bizning fundamental tasavvurlarimizni, o'zgartirishga majbur qilmoqda. Ular bizning yashash tarzimizga kirib borayotgani sababli ularning qanday o'sishini va potensial ziyonini boshqarishni o'rganishimizga to'g'ri kelayapti. Vaholangki, bunga asrlar emas, bir necha yillar qolgan bo'lishi ham mumkin. Shaxsiy hayot va ma'lumotlar xavfsizligini ta'minlash barchamizning o'z zimmamizda. Bizga o'z majburiyatlarimiz va huquqlarimizni bilib yurish zimmamizga yuklanadi. Boshqalarning qiziqishlarini nazorat qilish, bulardan ziyon ko'rmaslik maqsadida qiyin algoritmlarni tuzish uchun yangi muassasalar va ekspertlar kerak bo'ladi. Gap o'zgarishlarga ko'nikishda emas, balki yangi o'zgartirishlar kiritilishidir. Katta ma'lumotlar aqlbovar qilmaydigan masalalarni yecha oladigan texnologiya deb tushuniladi, lekin yangilari undan ham hayron qoldiradi. Katta ma'lumotlar hayotimizni, ishimizni, fikrlashimizni o'zgartirish imkoniga ega.



«Big Data» istiqbollari

Salmoqli ma'lumotlar aslida ilmiy-tadqiqotga aloqador sohalarda avvaldan ham mavjud bo'lgan bo'lsa-da, lekin so'nggi yillardagina ushbu fenomen ko'proq tilga olinadigan bo'ldi. Bunga sabab bugungi kunda yangi texnologiyalar, qurilmalar va aloqa vositalarining tezkor rivojlanishi, ijtimoiy tarmoqlarning keng miqyosda yoyilishi ta'sirida insoniyat tomonidan ishlab chiqilayotgan ma'lumotlar miqdori keskin tarzda ortib borishidir.

Hozirgi kunda insoniyat tomonidan ma'lumot har 10 daqiqada yaratilmoqda. Bu turdagi katta hajmli ma'lumotlar bilan ishlashda yangidan-yangi algoritmlar va texnologiyalar ishlab chiqilmoqda va amaliyotda keng qo'llanilyapti. Atrofimizdagi milliardlab to'planib qolgan ma'lumotlarni yig'ish, qayta ishlash va ulardan samarali foydalanish orqali ma'lum holat yoki obyekt to'g'risida tarixiy, joriy va kelajak vaqt uchun muhim bo'lgan bilimlarga ega bo'lishimiz mumkin. Buning uchun dastlab ushbu ma'lumotlar manbaini aniqlash muhim amallardan biri hisoblanadi.

Salmoqli ma'lumot tushunchasi katta o'lchamdagi ishlarni bajara oladigan operatorlarga nisbatan ishlatiladi. Misol uchun, «Facebook» kompaniyasi soatiga 10 milliondan ortiq yangi rasmlar joylashi bilan faxrlanishi mumkin. «Google» kompaniyasining 800 million nafar «Youtube» xizmati foydalanuvchilari oyiga har soniyada bir soatdan ko'p videolar joylashtiradi. Bundan ko'rinib turibdiki, hozirgi davrda salmoqli ma'lumot vazifasi o'ta muhim ahamiyatga ega.

Katta miqdordagi ma'lumotlarni tahlil qilish inson his etish imkoniyatidan tashqarida bo'lgan qonuniyatlarni aniqlashda yordam beradi. Bu esa kundalik hayotimizdagi barcha sohalar – hukumatni boshqarish, tibbiyot, telekommunikatsiya, moliya, transport, ishlab chiqarish va boshqa sohalarni yanada yaxshilash, ularning imkoniyatlarini oshirish, muammolarga muqobil yechimlar izlab topish imkonini yaratadi. Bugungi kunda salmoqli ma'lumotlarni moliyaviy tashkilotlar ham keng qo'llay boshlashi natijasida ma'lumotlar tahlilining raqobatbardoshligini oshirish va samaradorlikni kuchaytirish kabi muhim masalalar o'z yechimini topmoqda.

Salmoqli ma'lumotlardan foydalanish axborot mutaxassisi uchun qo'shimcha vazifalar va talablar yaratishdan tashqari, yangi ma'lumotlarni topish (Data mining) va ma'lumot bo'yicha mutaxassis (Data scientists) kabi kasblarning paydo bo'lishiga olib keladi.

«Data mining» – biron qonuniyatni topish maqsadida ma'lumotlarni tahlil

qilishga aytiladi.

«Data science» tushunchasiga ma'lumotlar omborini loyihalash va raqamlangan ma'lumotlarni qayta ishlashning barcha metodlari kiradi. «Data science» salmoqli ma'lumotning biznes nuqtai nazaridan hozirgi zamonaviy o'rindoshi hisoblanadi.

Bugungi kunda elektron hukumat tizimi butun dunyoda, shu jumladan, O'zbekistonda keng rivojlanayotgan tizim hisoblanadi. Bu tizim fuqarolar, tadbirkorlik subyektlari va davlat hokimiyati organlariga avvaldan shakllangan davlat xizmatlarini ko'rsatishda zamonaviy axborot-kommunikatsiya texnologiyalaridan foydalanishni ko'zda tutadi.

Mamlakatimizda interaktiv davlat xizmatlarini jismoniy va yuridik shaxslarga davlat organlarining veb-saytlari va Hukumat portali orqali taqdim etish uchun axborot kommunikatsiya texnologiyalarini joriy etish keng yo'lga qo'yilgan. Hozirgi kunga qadar elektron hukumat tizimining bir qancha loyihalari samarali amalga oshirildi.

Yuqorida ta'kidlaganidek, hukumatni boshqarishda salmoqli ma'lumotlar keng imkoniyatlar ochib beradi. Elektron hukumatda xalqaro standartlar talablariga mos keladigan «Big Data», «Data mining» va «Artificial intelligence» kabi zamonaviy texnologiyalarga asoslangan sifatini boshqarishning zamonaviy avtomatlashtirilgan tizimini yaratish sifat darajasi o'sishini amalga oshirib, bir qancha muammolarni yechishga imkon beradi.

Ayni damda «Elektron hukumat» tizimining axborot-resurslarini rivojlantirish departamentida aynan salmoqli ma'lumotlar bilan ishlash, unga butun tashkilotlarni jalb qilish rejalari ishlab chiqilmoqda. Salmoqli ma'lumotlarni qayta ishlash natijasida muqobil yechim izlash, ya'ni «Business intelligence» (biznes-intellekt) orqali kelajakda kelib chiqadigan natijalardan ham xabardor bo'lishimiz mumkin.

Ma'lumot o'rinda: IBS ma'lumotiga ko'ra, 2003- yili dunyoda 5 eksabayt (1 EB = 1 mlrd. gigabayt) ma'lumot to'plangan. 2008- yilga kelib bu hajm 0,18 zettabayt (1 ZB = 1024 eksabayt)ga, 2011- yili – 1,76 zettabaytgacha, 2013- yili esa 4,4 zettabaytgacha o'sgan. 2015- yilning mayida ma'lumotlarning global miqdori 6,5 zettabaytga ko'tarilgan. Tahminlarga ko'ra, 2020- yilga borib, insoniyat 40-44 zettabayt axborot ishlab chiqadi. «The Data Age» ning IDC kompaniyasi tahlilchilari tomonidan tayyorlangan hisobotida aytilishicha esa 2025- yilgacha bu miqdor 10 baravar o'sadi.

Salmoqli ma'lumot uchun Meta Group kompaniyasi 2001- yildayoq ishlab chiqqan "Uch V" an'anaviy xususiyatlar ahamiyatlidir:

1. Volume – ma'lumotlar hajmining kattaligi.
2. Velocity – ma'lumotlar o'sish tezligi va natijaga erishish uchun ularni tez qayta ishlash.
3. Variety – turli ma'lumotlarni bir vaqtning o'zida qayta ishlash imkoniyati.

Katta ma'lumotlarni tahlil qilish usullari. Big Data - Katta ma'lumotlar tizimlari nima?

Katta ma'lumotlar texnologiyasini ishlab chiqish

Katta ma'lumotlar (yoki Big Data) bu juda katta hajmdagi tuzilgan yoki tuzilmaviy ma'lumot bilan ishlash usullari to'plamidir. Katta ma'lumotlar bilan ishlash bo'yicha mutaxassislar vizual, inson tomonidan o'qiladigan natijalarni olish uchun uni qayta ishlash va tahlil qilish bilan shug'ullanadilar.

Katta ma'lumotlar bilan ishlashni istagan odamlar uchun ikkita asosiy ixtisoslashuv mavjud: katta ma'lumotlar bilan ishlash texnologiyalarini yaratadigan tahlilchilar va IT-maslahatchilar. Bundan tashqari, Big Data Analysts, ya'ni to'g'ridan-to'g'ri ma'lumotlar bilan ishlaydigan, mijozning IT-platformasi bilan shug'ullanadigan odamlar haqida ham gapirish mumkin. Ilgari bu oddiy matematik tahlilchilar edi, ular statistika va matematikani biladilar va statistik dasturlardan foydalanib ma'lumotlarni tahlil qilish bilan bir qatorda, ma'lumotlar va ma'lumotlar hayotining siklini tushunish ham zarur. Bu zamonaviy Data Analysts va ilgari bo'lgan tahlilchilar o'rtasidagi farq.

Agar biror kishi Big Data sohasida ma'lumotlarni tahlil qilish yoki IT-konsalting bilan shug'ullanmoqchi bo'lsa, unda birinchi navbatda yaxshi matematik ma'lumotga ega bo'lgan matematik yoki texnikaviy ma'lumot muhimdir. Shuningdek, ma'lumot texnologiyalarini, masalan, SAS, Nadoop, R tili yoki IBM yechimlarini o'zlashtirish foydali bo'ladi. Bundan tashqari, siz Big Data uchun qo'llaniladigan vazifalar bilan faol qiziqishingiz kerak, masalan, ularni bankda skoring yaxshilanishi yoki mijozning hayotiy siklini boshqarish uchun qanday foydalanish. Ushbu va boshqa bilimlarni mavjud manbalardan olish mumkin: masalan, Coursera va Big Data Universiti.

Pensilvaniya shtatining Uarton universitetida haridorlarni tahlil qilish tashabbusi ham mavjud bo'lib, u juda ko'p qiziqarli materiallarni nashr etgan.

Bu sohada ishlashni istaganlar uchun jiddiy muammo bu Big Data haqida aniq ma'lumot etishmasligi. Siz kitob do'koniga yoki biron-bir vebsaytga bora olmaysiz va, masalan, banklarda Big Data texnologiyalarining barcha ilovalarida ishlarning to'liq to'plamini olmaysiz. Bunday kataloglar yo'q. Ma'lumotlarning bir qismi kitoblarda, boshqa qismi konferensiyalarda to'planadi va siz o'zingiz nimadir bilan tanishishingiz kerak.

Yana bir muammo shundaki, tahlilchilar raqamlar dunyosida o'zlarini yaxshi his qilishadi, ammo ular biznesda har doim ham qulay emas. Bunday odamlar ko'pincha introvert qilinadi, ular bilan aloqa qilish juda qiyin va shuning uchun tadqiqot natijalari to'g'risidagi ma'lumotlarni mijozlarga ishonchli tarzda yetkazish juda qiyin. Ushbu ko'nikmalarni rivojlantirish uchun "Piramida printsiipi", "Grafik tilida gapiring" kabi kitoblarni tavsiya qilinadi. Ular taqdimot ko'nikmalarini rivojlantirishga, o'z fikrlarini aniq va aniq ifoda etishga yordam beradi.

Jahonda yirik kompaniyalar - banklar, uyali aloqa operatorlari va boshqalar katta ma'lumotlarni tahlil qilishlari kerak, shuning uchun ushbu sohada ishlashni istaganlar uchun istiqbollar mavjud. To'g'ri, hozirda ko'plab loyihalar integratsiya, ya'ni xorijiy ishlanmalar yoki ochiq manbali texnologiyalar asosida amalga oshirilmoqda. Bunday loyihalarda tubdan yangi yondashuvlar va texnologiyalar yaratilmaydi, aksincha, mavjud amaliyotlar moslashtiriladi.

Katta ma'lumotlarni qayta ishlash sohasiga ta'sir qiluvchi tendensiyalar va iqtisodiy kuchlarni tushunish muhimdir. Buning uchun ko'p o'qish kerak, IT sohasidagi nufuzli mutaxassislarning chiqishlarini tinglash, tematik konferensiyalarda qatnashish kerak. Endi deyarli har bir konferensiyada Big Data haqida bo'lim mavjud, ammo ularning barchasi bu haqda boshqa nuqtai nazardan gaplashadi: texnologiya, biznes yoki marketing nuqtai nazaridan. Siz ushbu mavzu bo'yicha loyihalarni amalga oshiradigan kompaniyada dizaynerlik yoki stajirovkaga borishingiz mumkin. Agar siz o'zingizning qobiliyatingizga ishonchingiz komil bo'lsa, unda Big Data sohasida startapni tashkil qilish juda kech emas.

Bozor bilan doimiy aloqada bo'lmasdan rivojlanishning yangi xatarlari talab qilinmaydi.

Biroq, siz yangi mahsulot uchun javobgar bo'lganingizda, bozorni tahlil qilish va potensial mijozlar, sheriklar, mijozlar va ularning ehtiyojlari haqida ko'p narsalarni biladigan professional tahlilchilar bilan muloqot qilish uchun ko'p vaqt sarflanadi. Bozor bilan doimiy aloqada bo'lmasdan, yangi rivojlanish talab qilinmasligi xavfini tug'diradi. Har doim noaniqliklar ko'p: siz birinchi foydalanuvchi kim bo'lishini (erta asrab oluvchilar), siz ular uchun nimalarga ega ekanligingizni va keyin qanday qilib ommaviy auditoriyani jalb qilish kerakligini tushunishingiz kerak. Ikkinchi eng muhim vazifa - ishlab chiqaruvchilarga yakuniy mahsulot to'g'risida aniq va yaxlit tasavvurni shakllantirish, ba'zi talablar o'zgarishi mumkin bo'lgan muhitda ishlashga undash uchun va ustuvorliklar birinchi mijozlarning fikr-mulohazalariga bog'liq. Shu sababli, bir tomondan mijozlar va boshqa tomondan ishlab chiquvchilarning talablarini boshqarish muhim vazifadir. Shunday qilib, na biri, na boshqasi qiziqishni yo'qotmaydi va loyihani yakunlaydi. Birinchi muvaffaqiyatli loyihadan so'ng, bu osonlashadi va asosiy vazifa yangi biznes uchun to'g'ri o'sish modelini topish bo'ladi.

Katta ma'lumotlar - bu nafaqat ma'lumotlarning o'zi, balki ularni qayta ishlash va ishlatish texnologiyalari, kerakli ma'lumotlarni katta massivlarda topish usullari. Katta ma'lumotlar muammosi hali ham ochiq va har qanday tizim uchun o'nlab yillar davomida turli xil ma'lumotlarni to'plagan.

Bir ibora ushbu atama bilan bog'liq "Hajm, tezlik, turli-tumanlik" - katta ma'lumotlar bilan ishlash tamoyillari. Bu to'g'ridan-to'g'ri ma'lumotlar miqdori, ishlov berish tezligi va ma'lumotlarning xilma-xilligi bir qator saqlanadi. Yaqinda ular uchta asosiy printsipga yana birini qo'shishni boshladilar - qiymati degan ma'noni anglatadigan ma'lumotlarning qiymati. Ya'ni, uni saqlash va qayta ishlash xarajatlarini oqlaydigan nazariy yoki amaliy jihatdan foydali va zarur bo'lishi kerak.

Katta ma'lumotlarning odatiy manbasiga misol ijtimoiy tarmoqlardir - har bir profil yoki jamoat sahifasi tuzilmagan ma'lumot okeanidagi bir tomchi. Bundan tashqari, ma'lum bir profilda saqlanadigan ma'lumotlarning hajmidan qat'i nazar, foydalanuvchilarning har biri bilan o'zaro aloqa imkon qadar tez bo'lishi kerak.

Katta ma'lumotlar doimiy ravishda inson hayotining har qanday sohasida to'planadi. Bunga insonning o'zaro aloqasi yoki hisoblash bilan bog'liq har qanday soha kiradi. Bular ijtimoiy media, tibbiyot va bank ishi, shuningdek,

kunlik hisob-kitoblarning ko'plab natijalarini beradigan qurilmalar tizimlari. Masalan, astronomik kuzatishlar, meteorologik ma'lumotlar va yerni ovoz chiqaradigan qurilmalardan olingan ma'lumotlar.

Turli kuzatuv tizimlaridan olingan ma'lumotlar real vaqt rejimida ham kompaniya serverlariga keladi. Televizion va radio eshittirishlar, uyali aloqa operatorlarining qo'ng'iroq bazalari - har bir shaxsning ular bilan o'zaro ta'siri minimal, ammo jami bu barcha ma'lumotlar katta ma'lumotlarga aylanadi.

Katta ma'lumotlar texnologiyalari tadqiqot va tijoratning ajralmas qismiga aylandi. Bundan tashqari, ular davlat boshqaruvi sohasini qamrab olishni boshlamoqdalar va hamma joyda axborotni saqlash va boshqarish uchun tobora samarali tizimlarni joriy etish talab etiladi.

"Katta ma'lumotlar" atamasi birinchi marta 2008- yilda "Nature" jurnalining muharriri Clifford Lynch katta hajmdagi ma'lumotlar bilan ishlash texnologiyalaridan foydalangan holda ilm-fan kelajagini rivojlantirish to'g'risida maqola chop etganda paydo bo'ldi. 2009- yilga qadar bu atama faqat ilmiy tahlil nuqtai nazaridan ko'rib chiqilgan, ammo yana bir nechta maqolalar nashr etilgandan so'ng, matbuot Big Data tushunchasini keng ishlatishni boshladi va hozirgi kunda ham foydalanishda davom etmoqda.

2010- yilda o'sib borayotgan katta ma'lumotlar muammosini hal qilish uchun birinchi urinishlar paydo bo'ldi. Dasturiy ta'minot mahsulotlari katta hajmli massivlardan foydalanishda xatarlarni minimallashtirishga qaratilgan.

2011- yilga kelib, Microsoft, Oracle, EMC va IBM kabi yirik kompaniyalar katta ma'lumotlarga qiziqish bildirdilar - ular o'zlarining rivojlanish strategiyalarida Big Data ishlanmalaridan birinchi bo'lib foydalanganlar va juda muvaffaqiyatli bo'lishgan.

2013- yilda universitetlar katta ma'lumotlarni alohida fan sifatida o'rganishni boshladilar - hozirda nafaqat ma'lumotshunoslik, balki muhandislik va hisoblash fanlari bilan bir qatorda ushbu sohadagi muammolar bilan shug'ullanishmoqda.

Tahlil qilish va ma'lumotlarni qayta ishlashning asosiy usullari quyidagilarni o'z ichiga oladi.

Sinf usullari yoki chuqur tahlil qilish (Data Mining)

Ushbu usullar juda ko'p, ammo ularni bitta narsa birlashtiradi: axborot texnologiyalari sohasidagi yutuqlar bilan birgalikda ishlatiladigan matematik

vositalar.

Crowd sourcing

Ushbu texnik usul sizga bir vaqtning o'zida bir nechta manbalardan ma'lumot olish imkonini beradi va ularning soni deyarli cheksizdir.

A / B sinov

Ma'lumotlarning umumiy hajmidan elementlarning biriga tegishli elementlar to'plami tanlangan, u boshqa o'xshash to'plamlar bilan taqqoslanadi, bu yerda elementlardan biri o'zgartirilgan. Bunday sinovlarni o'tkazish parametrlarning qaysi biri tanlanmalarini nazorat populyasiyasiga eng katta ta'sir ko'rsatishini aniqlashga yordam beradi. Katta hajmdagi ma'lumotlar tufayli ko'p sonli iteratsiyalarni amalga oshirish mumkin, ularning har biri eng ishonchli natijaga yaqinlashmoqda.

Bashoratli tahlil

Ushbu sohadagi mutaxassislar ushbu vaziyatda eng maqbul qaror qabul qilish uchun boshqariladigan obyekt qanday ishlashini oldindan aytib berishga va rejalashtirishga harakat qilishadi.

Mashinani o'rganish (sun'iy aql)

Bu ma'lumotlarning empirik tahliliga va tizimlarning o'z-o'zidan o'rganish algoritmlarini tuzishga asoslangan.

Tarmoq tahlili

Ijtimoiy tarmoqlarni tadqiq qilishning eng keng tarqalgan usuli shundan iboratki, statistik ma'lumotlarni olgandan so'ng, panjara ichida joylashgan tugunlar tahlil qilinadi, ya'ni individual foydalanuvchilar va ularning jamoalari o'rtasidagi o'zaro munosabatlar.

2017- yilda katta ma'lumotlar yangi va noma'lum bo'lib qolishni to'xtatganida, ularning ahamiyati nafaqat kamaymadi, balki yanada oshdi. Endilikda mutaxassislar katta hajmdagi ma'lumotlarni tahlil qilish nafaqat yirik tashkilotlar, balki kichik va o'rta biznes vakillari uchun ham mavjud bo'lishiga ishonishadi. Ushbu yondashuvni quyidagi tarkibiy qismlardan foydalangan holda amalga oshirish rejalashtirilgan:

Bulutli saqlash

Ma'lumotni saqlash va ishlov berish tezroq va tejamkor bo'lmoqda - o'z ma'lumotlar markazini saqlash va xodimlarning kengayishi bilan taqqoslaganda, bulutli ijaraga olish arzonroq alternativaga o'xshaydi.

Qorong'i ma'lumotlardan foydalanish

“Qorong'u ma'lumotlar” deb ataladigan narsa bu to'g'ridan-to'g'ri foydalanishda muhim rol o'ynamaydigan kompaniya haqidagi barcha raqamli bo'lmagan ma'lumotlar, ammo ma'lumotni saqlashning yangi formatiga o'tish uchun sabab bo'lishi mumkin.

Sun'iy intellekt va chuqur o'rganish

Inson miyasining tuzilishi va ishiga taqlid qiluvchi mashina zakovatini o'rganish texnologiyasi doimiy ravishda o'zgarib turadigan ma'lumotlarning katta hajmini qayta ishlash uchun eng mosdir. Bunday holda, mashina odam bajarishi kerak bo'lgan narsani bajaradi, ammo xato ehtimoli juda kamayadi.

Blockchain

Ushbu texnologiya ko'plab onlayn, shu jumladan xalqaro tranzaksiyalarni tezlashtirish va soddalashtirishga imkon beradi. Blockchainning yana bir tomoni shundaki, uning yordamida tranzaksion harajatlar kamayadi.

O'z-o'ziga xizmat ko'rsatish va narxlarni pasaytirish

2017 yilda “o'ziga o'zi xizmat ko'rsatish platformalari” ni joriy etish rejalashtirilgan - bu bepul platformalar bo'lib, unda kichik va o'rta biznes vakillari o'zlari saqlagan ma'lumotlarni mustaqil ravishda baholaydilar va tashkil etadilar.

Barcha marketing strategiyalari qandaydir tarzda ma'lumotni boshqarish va mavjud ma'lumotlarni tahlil qilishga asoslangan. Shuning uchun katta ma'lumotlardan foydalanish kompaniyaning kelajakdagi rivojlanishini bashorat qilishi va amalga oshirishi mumkin.

Masalan, katta ma'lumotlar asosida yaratilgan RTBkim oshdi savdosi sizga reklamadan samaraliroq foydalanish imkonini beradi - ma'lum bir mahsulot faqat uni sotib olishga qiziqqan foydalanuvchilar guruhiga namoyish qilinadi.

Marketing va biznesda katta ma'lumotlar texnologiyalaridan foydalanish nima?

Ularning yordami bilan siz xaridorlarning talabiga aylanishi mumkin bo'lgan yangi loyihalarni ancha tez yaratishingiz mumkin.

Ular mijozlarning talablarini mavjud yoki rejalashtirilgan xizmat bilan o'zaro bog'liqlashtirishga yordam beradi va shu bilan ularni tuzatadi.

Ma'lumotlarning katta usullari barcha foydalanuvchilar va har bir

kishining hozirgi qoniqish darajasini baholashga imkon beradi.

Mijozlarning sodiqligini oshirish katta ma'lumotlarni qayta ishlash usullari bilan ta'minlanadi.

Internetda maqsadli auditoriyani jalb qilish katta hajmdagi ma'lumotlarni boshqarish qobiliyati tufayli osonlashadi.

Masalan, mahsulotning ehtimoliy ommaviylikini bashorat qilish uchun eng mashhur xizmatlardan biri bu Google.trends. U marketologlar va tahlilchilar tomonidan keng qo'llaniladi, bu ularga ushbu mahsulotning o'tmishda ishlatilishi va keyingi mavsum uchun prognoz ma'lumotlarini olishga imkon beradi. Bu kompaniya menejerlariga reklama byudjetini yanada samarali taqsimlashga, pulni qaysi sohaga investitsiya qilish yaxshiroq ekanligini aniqlashga imkon beradi.

Katta ma'lumotlarga misollar

Big Data texnologiyalarini bozorga va zamonaviy hayotga faol joriy etish, dunyoning taniqli kompaniyalari dunyoning deyarli har bir burchagida o'z mijozlariga ega bo'lganlaridan foydalanishni boshlaganidan keyin boshlandi.

Bular Facebook va Google, IBM., Matser Card, VISA va Bank of America kabi moliyaviy institutlar kabi ijtimoiy gigantlardir.

Masalan, IBM uzluksiz pul o'tkazmalariga katta ma'lumotlar usullarini qo'llaydi. Ularning yordami bilan 15 foizga ko'proq soxta bitimlar aniqlandi, bu esa himoyalangan mablag'lar miqdorini 60 foizga oshirish imkonini berdi. Shuningdek, tizimning noto'g'ri signallari bilan bog'liq muammolar hal qilindi - ularning soni yarmidan ko'piga kamaytirildi.

VISA shu kabi yoki boshqa operatsiyani amalga oshirish uchun qilingan firibgarliklarni kuzatib, Big Data -dan foydalangan. Buning evaziga ular har yili 2 milliard dollardan ortiq mablag'ni qochqindan asraydi.

Germaniya Mehnat vazirligi ishsizlik bo'yicha nafaqa berish bo'yicha katta ma'lumot tizimini joriy etish orqali xarajatlarni 10 milliard evroga kamaytirishga muvaffaq bo'ldi. Shu bilan birga, fuqarolarning beshdan biri bu imtiyozlarni hech qanday sababsiz olayotganligi aniqlandi.

Katta ma'lumotlar o'yin sohasi tomonidan o'tib ketmadi. Shunday qilib, World of Tanksning ishlab chiquvchilari barcha o'yinchilar haqida ma'lumotni o'rganishdi va ularning faoliyatining mavjud ko'rsatkichlarini solishtirishdi. Bu kelajakda o'yinchilarning ketishini taxmin qilishda yordam berdi - qilingan

taxminlarga asosanib, tashkilot vakillari foydalanuvchilar bilan yanada samarali aloqada bo'lishdi.

Taniqli yirik ma'lumotlar tashkilotlariga QSBC, Nasdaq, Coca-Cola, Starbucks va AT&T kiradi.

Katta ma'lumotlarning eng katta muammosi - uni qayta ishlashning narxi. Bunga qimmatbaho uskunalar va katta miqdordagi ma'lumotlarga xizmat ko'rsatadigan malakali mutaxassislarning ish haqi xarajatlari kiradi. Shubhasiz, uskunalar doimiy ravishda yangilanib turilishi kerak, shunda u ma'lumotlarning ko'payishi bilan minimal ish qobiliyatini yo'qotmaydi.

Ikkinchi muammo yana qayta ishlanishi kerak bo'lgan ma'lumotlarning katta miqdori bilan bog'liq. Agar, masalan, tadqiqot 2-3 emas, balki ko'p sonli natijalarni beradigan bo'lsa, obyektiv bo'lib qolish va umumiy ma'lumot oqimidan faqat hodisaning holatiga real ta'sir ko'rsatadiganlarni tanlash juda qiyin.

Katta ma'lumotlarning maxfiyligi muammosi

Ko'pgina mijozlarning xizmatlari onlayn ma'lumotlardan foydalanishga o'tishlari sababli, kiber jinoyatchilar uchun boshqa maqsadga aylanish juda oson. Hatto biron bir onlayn tranzaksiya qilmasdan shaxsiy ma'lumotlarni saqlash ham bulutli saqlash mijozlari uchun nomaqbul oqibatlariga olib kelishi mumkin.

Axborotni yo'qotish muammosi. Ehtiyotkorlik bilan oddiy bir martalik ma'lumotni zaxiralash bilan cheklanmaslik kerak, lekin kamida 2-3 ta zaxira nusxasi kerak. Biroq, hajmning o'sishi bilan, zaxira bilan bog'liq qiyinchiliklar kuchaymoqda - va IT-mutaxassislari ushbu muammoning maqbul yechimini topishga harakat qilmoqdalar.

Dunyodagi katta ma'lumotlar texnologiyalari bozori 2014 yil ma'lumotlariga ko'ra, katta ma'lumotlar bozorining 40% xizmatlardan iborat. Katta ma'lumotni kompyuter uskunalarida ishlatishdan olingan daromadlar ushbu ko'rsatkichdan biroz pastroqdir (38%). Qolgan 22% dasturiy ta'minot.

Statistikaga ko'ra, Big Data muammolarini hal qilish uchun jahon segmentidagi eng foydali mahsulotlar bu In-memory va NoSQL analitik platformalaridir. Mos ravishda bozorning 15 va 12 foizini Log-fayl analitik dasturlari va Columnar platformalari egallaydi. Ammo Nadoop/ MapReduce amalda katta ma'lumotlar bilan bog'liq muammolarni samarali hal qila

olmaydi.

Yirik ma'lumotlar texnologiyalarini joriy etish natijalari:

- mijozlarga xizmat ko'rsatish sifatini oshirish;
- ta'minot zanjiriga integratsiyani optimallashtirish;
- tashkilotni rejalashtirishni optimallashtirish;
- mijozlar bilan o'zaro munosabatlarni tezlashtirish;
- mijozlar so'rovlarini ko'rib chiqish samaradorligini oshirish;
- xizmat narxining pasayishi;
- mijoz buyurtmalarini qayta ishlashni optimallashtirish.

Katta ma'lumotlarni qayta ishlash texnologiyalarini dastlabki o'rganish uchun javob beradi. Bu ma'lumotlarning ko'pligi kundalik hayotga va uning barcha sohalariga: ilm-fan, biznes, tibbiyot va boshqalarga qanday ta'sir qilganligini aniq ko'rsatib turibdi. Unda ko'plab rasmlar mavjud, shuning uchun ular ko'p harakat qilmasdan qabul qilinadi.

2. MA'LUMOTLAR

Kalit so'zlar: *ma'lumotlar, jadval, dasturiy ta'minot, atribut, obyekt, yozuv, gipoteza, joy, o'zgaruvchi, o'lchov, navbat, qiymat, aniqlik, balandlik, vazn, uzunlik, ma'lumotlar to'plami, ta'rif, hisoblash, fayl, matn, CSV, SAS, excel, bilim, qidirish, kirish, ma'lumotlar bazasi menejeri, mantiqiy dizayn, kaskadni o'chirish, OLAP, konstantalar, koordinatalar, foyda, so'rov.*

Keng ma'noda ma'lumotlar alomatlar, matn, grafiklar, rasmlar, tovushlar, analogli yoki raqamli video-segmentni ifodalaydi. Ma'lumotlar o'lchovlar, tajribalar, arifmetik hamda mantiqiy amallar natijasida olinishi mumkin.

Ma'lumotlar saqlash, uzatish va qayta ishlashga mos shaklda taqdim etilishi lozim. Boshqa so'z bilan aytganda ma'lumotlar bu yetkazib beruvchilar tomonidan taqdim etiladigan qayta ishlanmagan hamda itse'molchilar tomonidan ma'lumotlar asosida ma'lumotlarni shakllantirish uchun qo'llaniladigan materialdir.

Ma'lumotlar to'plamini ifodalashda ikki o'lchovli jadvaldan foydalaniladi. Jadvalning gorizontal bo'ylab obyekt atributlari yoki uning alomatlari joylashadi. Jadvalning vertikal bo'yicha obyektlar.

Obyekt - atributlar to'plami sifatida ta'riflanadi. Obyekt jadval satri sifatida ifodalanishi mumkin.

Atribut - obyektning xarakterlovchi xossa. Atribut shuningdek o'zgaruvchi, o'lcham, tavsif deb ham ataladi. U jadval maydoni sifatida taqdim etilishi mumkin.

Umumiy kategoriyalardan aniq kattaliklarga o'tish natijasida o'rganiluvchi tushunchaning o'zgaruvchilar to'plami hosil bo'ladi.

O'zgaruvchi (variable) - o'rganilayotgan Obyektlar uchun umumiy bo'lgan xossa yoki xarakteristika, uning namoyon bo'lishi obyektidan obyektga qarab o'zgarishi mumkin.

Qiymat (value) alomatning ifodalanishidir.

Ma'lumotlarni tahlil qilishda obyektlar to'plamini qarab chiqishning iloji yo'q. Katta hajmdagi ma'lumotlarni o'rganish juda ham qimmat jarayon hisoblanadi, u katta vaqt harajatlarini talab qiladi hamda inson omili bilan bog'liq xatoliklarga olib keladi.

Butun to'plamning ma'lum bir qismini, ya'ni tanlanmani qarash va uning asosida kerakli ma'lumotni olish etarlidir.

Ammo tanlanma hajmi bosh to'plamda keltirilgan obyektlar xilma-

xilligiga bog'liq bo'lishi kerak. Tanlanmada bosh to'plamning turli xil kombinatsiyalari va elementlari taqdim etilishi lozim.

Bosh to'plam (population) - tadqiqotchini qiziqtirayotgan obyektlarning butun to'plami.

Tanlanma (sample) - bosh to'plamning xossalari va tavsiflari to'g'risida xulosalar chiqarish va o'rganish maqsadida saralab olingan bosh to'plamning bir qismi.

Parametrlar - bosh to'plamning sonli tavsiflari.

Statistikalar - tanlanmaning sonli tavsiflari.

Ko'pincha tadqiqotlar farazlarga asoslanadi. Farazlar ma'lumotlar asosida tekshiriladi. Gipoteza bu o'z qismida tekshirilishi kerak bo'lgan obyektlar to'plami to'g'risidagi faraz.

Gipoteza - turli xil empirik alomatlar o'rtasida bog'liqlik o'rnatish, yoki alomatni izohlash uchun xizmat qiladigan qisman asoslab berilgan qonuniyatdir.

Aytaylik, erksiz o'zgaruvchi (umr ko'rish davomiyligi) erkli o'zgaruvchilar bo'lgan ba'zi sabablarga (oziq-ovqat sifati, turmush tarzi, yashash joyi va boshqalar) qarab o'zgaradi.

Ammo, o'zgaruvchi dastavval erkli yoki erksiz bo'lmaydi.

U ma'lum bir gipoteza aniqlangandan keyin shunday bo'ladi. Bir gipotezada erksiz bo'lgan o'zgaruvchi boshqasida erkli bo'lishi mumkin.

O'lchov - ma'lum bir qoidaga muvofiq o'rganilayotgan obyektlarning xususiyatlarni raqamlash jarayoni.

Ma'lumotni tayyorlash jarayonida obyektning o'zi emas, balki uning xususiyatlari o'lchanadi. O'lchov ma'lum bir shkala bo'yicha amalga oshiriladi.

Shkala - obyektlarga raqam beradigan qoida.

O'zgaruvchilar sonli yoki belgilar to'g'risidagi ma'lumotlar bo'lishi mumkin.

O'z navbatida sonli ma'lumotlar diskret va doimiy bo'lishi mumkin.

Diskret ma'lumotlar xarakterli qiymatlar bo'lib, ularning umumiy soni hisobga olinmaydi.

Uzluksiz ma'lumotlar - qiymatlari ma'lum vaqt oralig'ida bo'lgan ma'lumotlardir.

Shkalalar, odatda, quyidagi o'lchov shkalalari qaraladi: nominal, tartibli,

oraliq, nisbiy va dixotomik.

Nominal shkala - faqat toifalarni o'z ichiga olgan shkala; undagi ma'lumotlarni tartiblab bo'lmaydi, ular bilan arifmetik amallarni bajarish mumkin emas. Ushbu masshtab uchun faqat shunday operatsiyalar qo'llaniladi: teng/ teng emas.

Ordinal shkala - obyektlarning nisbiy pozitsiyasini ko'rsatadigan raqamlar berilgan, ammo ular orasidagi farqning kattaligini anglatmaydigan masshtab.

O'lchov shkalasi o'zgaruvchilar qiymatlarini tartiblash imkonini beradi.

Tarkibiy shkaladagi o'lchovlar faqat miqdorlarning ketma-ketligi to'g'risida ma'lumotni o'z ichiga oladi, ammo "bitta miqdor boshqasidan kattaroq" yoki "boshqasidan qancha kam" deb aytishga imkon bermaydi.

Bu shkala uchun faqat shunday operatsiyalar qo'llaniladi: teng / teng emas, ko'proq/kamroq.

Interval shkalasi (interval scale) - qiymatlari orasidagi farqni hisoblash mumkin bo'lgan shkala, ammo ularning munosabati mantiqiy emas.

Ushbu shkala sizga ikki qiymat o'rtasidagi farqni topishga imkon beradi, nominal va tartibli o'lchovlarning xususiyatlariga ega, shuningdek belgining miqdoriy o'zgarishini aniqlashga imkon beradi.

Bu shkala uchun faqat shunday operatsiyalar qo'llaniladi: teng / teng emas, kattaroq / kamroq, qo'shish / ayirish.

Nisbiy shkala (ratio scale) - bu ma'lum bir yo'naltiruvchi va shkala qiymatlari o'rtasida mumkin bo'lgan munosabatlar mavjud bo'lgan shkala.

Ushbu shkala uchun quyidagi amallar qo'llaniladi: teng / teng emas, kattaroq / kamroq, qo'shish / ayirish, ko'paytirish / bo'lish.

Dixotomoz shkala (dichotomous scale) - faqat ikkita toifani o'z ichiga olgan shkala, odatda "ha-yo'q" (nominal shkalaning alohida holati).

Ma'lumotlar to'plami turlari

Yozuvlardan iborat ma'lumotlar

Eng keng tarqalgan ma'lumotlar bu yozuvlardan iborat ma'lumotlar (record data). Bu jadvalli ma'lumotlar, matritsali ma'lumotlar, hujjatli ma'lumotlar, tranzaksion yoki operatsion.

Jadval ma'lumotlari yozuvlardan iborat bo'lib, ularning har biri belgilangan atributlar to'plamidan iborat.

Tranzaksion ma'lumotlar bu maxsus ma'lumotlar turi bo'lib, unda tranzaksiya hisoblangan har bir yozuv bir qator qiymatlarni o'z ichiga oladi.

Do'kon mijozlarining xaridlari ro'yxatini o'z ichiga olgan tranzaksion ma'lumotlar bazasiga misol 1-rasmda keltirilgan.

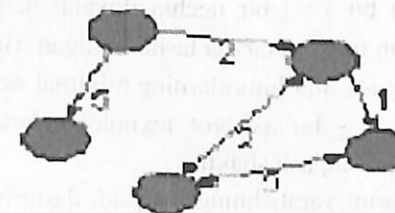
Tranzaksiya- ma'noga ega hamda faqat to'liq bajarilishi mumkin bo'lgan minimal mantiqan ma'no berilgan amaldir.

ID	Items
1	Bread, Coke, Milk
2	Beer, Bread
3	Beer, Coke, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Coke, Diaper, Milk

1-rasm. Tranzaksiya ma'lumotlariga misollar

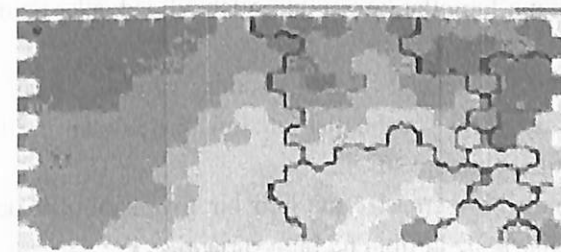
Grafik ma'lumotlarga misollar:

WWW-ma'lumotlar; molekulyar tuzilmalar; graflar (2-rasm); xaritalar.



2-rasm. Grafga misol

Masalan, xaritalardan foydalanib, obyektlardagi o'zgarishlarni vaqt va fazoda kuzatish, ularning tekislikda yoki kosmosda tarqalish xususiyatini aniqlash mumkin.



3-rasm. "Koxonen xaritasi" turidagi ma'lumotlarga misol

Ma'lumotni grafik taqdim etishning afzalligi, masalan, jadval jadvaliga qaraganda, uni idrok qilishning osonligi. Koxonen xaritasining namunasi (neyron tarmoqlari modeli) 3- rasmda keltirilgan.

Ma'lumotni saqlash formatlari.

Zamonaviy dunyo ma'lumotlarining asosiy xususiyatlaridan biri shundaki, ular juda ko'p.

Ma'lumotlar bilan ishlashning to'rtta yo'nalishi mumkin: ma'lumotlarni aniqlash, hisoblash (o'lchash), manipulyasiya (formatni o'zgartirish) va ishlov berish (to'plash, uzatish va boshqalar).

Ma'lumotlarni manipulyasiya qilishda, tipdagi ma'lumotlar tuzilishi "fayl". Fayllar har xil formatlarga ega bo'lishi mumkin.

Data Mining uchun eng keng tarqalgan ma'lumotlarni saqlash shakli ma'lumotlar bazasi.

Ma'lumotlar bazalari (asosiy qoidalar).

Ma'lumotlar bazasi (Database) - maxsus tashkil etilgan va elektron shaklda saqlanadigan ma'lumotlar.

Ma'lumotlar ularni bir yoki bir nechta ilovalar uchun qidirish va unga kirishni osonlashtiradigan maxsus tarzda tashkil etilgan. Bunday ma'lumotlarni tashkillashtirish, shuningdek, ma'lumotlarning minimal zaxirasini ta'minlaydi.

Ma'lumotlar bazalari - bu axborot texnologiyalarining turlaridan biri, shuningdek ma'lumotlarni saqlash shakli.

Ma'lumotlar bazalarini yaratishning maqsadi dasturiy ta'minotga, amaliy dasturiy ta'minotga va kompyuterda ma'lumotlarning jismoniy joylashishiga bog'liq bo'lmaydigan ma'lumotlar tizimini yaratishdir. Bunday ma'lumotlar tizimini qurish izchil va yaxlit ma'lumotlarni taqdim etishi kerak. Ma'lumotlar bazasini loyihalashda undan ko'p maqsadli foydalanish qabul qilinadi. Eng oddiy holatda ma'lumotlar bazasi ikki o'lchovli jadvallar tizimi sifatida taqdim etiladi.

Ma'lumotlar sxemasi - ma'lumotlar tavsifi tilida ko'rsatilgan va ma'lumotlar bazasini boshqarish tizimi tomonidan ishlangan mantiqiy ma'lumotlar tuzilishini tavsifi.

Foydalanuvchi sxemasi - ma'lum bir foydalanuvchi uchun belgilangan jadval maydonlarining tartibining bitta varianti.

Ma'lumotlar bazasini boshqarish tizimlari.

Ma'lumotlar bazasini boshqarish tizimi - bu ma'lumotlar bazasida ma'lumotlarni tashkillashtirish, saqlash, yaxlitlash, o'zgartirish, o'zgartirish va o'qishni boshqaruvchi dastur.

BBBT (Database Management System, DBMS) - bu qobiq bo'lib, uning yordamida jadvallarning tuzilishini tashkil etish va ularni ma'lumotlar bilan to'ldirishda u yoki bu ma'lumotlar bazasi olinadi.

Relational Database Management System - bu relyasion ma'lumotlar modeliga asoslangan ma'lumotlar bazasi.

Relyasion ma'lumotlar modelida har qanday ma'lumotlar namoyishi relyasion jadvallar to'plamiga qisqartiriladi (maxsus tipdagi ikki o'lchovli jadvallar).

Ma'lumotlar omborini (domenga tegishli ma'lumotlar bazasi) qurish uchun ma'lumotlar bazasini boshqarish tizimlari qo'llaniladi.

Ma'lumotlar bazasi dasturiy, texnik va tashkiliy tuzilishga ega.

Dasturiy ta'minot vositalari ma'lumotlarni kiritish-chiqarish va qayta ishlash, ma'lumotlar bazasini yaratish, o'zgartirish va sinovdan o'tkazishni ta'minlovchi boshqaruv tizimini o'z ichiga oladi.

Ichki BBBT dasturlash tillari to'rtinchi avlod tillari (C, C ++, S#, Java, Python). Ma'lumotlar bazasi tillari, dasturlar, ma'lumotlar bazalari va foydalanuvchi interfeysi, shu jumladan ekrandagi shakllar, menyu, hisobotlar yaratiladi.

Ma'lumotlar bazalari, shuningdek, BBBT uchun quyidagi talablar qo'yiladi:

1. Yuqori tezlik;
2. Ma'lumotlarni yangilash qulayligi;
3. Ma'lumotlarning erkinligi;
4. Ma'lumotlardan ko'p foydalanuvchi tomonidan foydalanish imkoniyati;
5. Ma'lumotlar xavfsizligi;
6. Ma'lumotlar bazasini qurish va ishlashini standartlashtirish (deyarli BBBT);
7. Tegishli mavzu doirasidagi ma'lumotlarni ko'rsatishning yetarliligi;
8. Do'stona interfeysi.

BBBT ma'lumotlar bazasiga so'rovlarni qayta ishlash va javob olish uchun javobgardir. Ma'lumotlarni saqlash usullari har xil bo'lishi mumkin:

ma'lumotlar modeli relatsion yoki ko'p o'lchovli, tarmoqli yoki ierarxik bo'lishi mumkin.

- **Ma'lumot turlarining tasnifi.**

- **Relyasion ma'lumotlar** bu relyasion ma'lumotlar bazalari - jadvallar (mantiqiy ma'lumotlar modeli) dan olingan ma'lumotlar.

- **Ko'p o'lchovli ma'lumotlar** bu OLAP kublarida berilgan ma'lumotlar (real vaqtda analitik ishlov berish).

O'lchov yoki o'q - ko'p o'lchovli ma'lumotlarda - bu bir xil turdagi ma'lumotlar to'plamidir, bu sizga ko'p o'lchovli ma'lumotlar bazasini tuzishga imkon beradi.

- **Vazifani yechish jarayonida qiymatlari barqarorligi mezoni bo'yicha ma'lumotlar quyidagicha bo'lishi mumkin:**

O'zgaruvchi - masalani yechish davomida qiymatlarini o'zgartiradigan ma'lumotlar muammoning yechimlari;

O'zgarmas - masalani yechishda o'z qiymatlarini saqlaydigan ma'lumotlar (matematik konsantalar, qo'zg'almas jismlarning koordinatalari) va tashqi omillarga bog'liq emas;

Shartli doimiy - ba'zan qiymatlarini o'zgartirishi mumkin bo'lgan ma'lumotlar, ammo bu o'zgarishlar muammoni hal qilish jarayoniga bog'liq bo'lmasdan tashqi omillar bilan belgilanadi.

Ma'lumotlar, ularning bajaradigan funksiyalariga qarab: xabar beruvchi, tezkor, arxiv bo'lishi mumkin.

Davriy ma'lumotlar hamda nuqtaviy ma'lumotlarini farqlashingiz kerak. Ushbu farqlar ma'lumot to'plash tizimini loyihalashda, shuningdek o'lchash jarayonida muhimdir.

Bir davr uchun olingan ma'lumotlar ma'lum vaqt oralig'ini tavsiflaydi.

Nuqtaviy ma'lumotlar ma'lum bir vaqtning o'zida o'zgaruvchining qiymatini anglatadi. Ma'lumotlar birlamchi va ikkilamchi bo'ladi.

Ikkilamchi ma'lumotlar - bu dastlabki ma'lumotlarga nisbatan qo'llanilgan ma'lum hisob-kitoblarni natijasida olingan ma'lumotlar. Ikkilamchi ma'lumotlar, qoida tariqasida, saqlanadigan ma'lumot miqdorini ko'paytirish orqali foydalanuvchi so'roviga tezkor javob olishga olib keladi. Birlamchi ma'lumotlar to'g'ridan-to'g'ri o'lchash (kuzatish) orqali olinadi.

Metadata

Metadata - ma'lumotlar haqidagi ma'lumotlar.

Metadata tarkibiga quyidagilar kiradi: kataloglar, ma'lumotnomalar, registrlar. Metadata ma'lumotlarning tarkibi, mazmuni, holati, kelib chiqishi, joylashuvi, sifati, taqdimot shakllari va formatlari, murojaat, olish va ulardan foydalanish shartlari, mualliflik huquqi, ma'lumotlarga nisbatan mulk huquqi va boshqalar to'g'risidagi ma'lumotlarni o'z ichiga oladi.

Saqlashni boshqarishda ishlatiladigan metadata uning konfiguratsiyasi va ishlatilishi uchun zarur bo'lgan ma'lumotlarni o'z ichiga oladi.

Biznes meta-ma'lumotlari va operatsion meta-ma'lumotlar farqlanadi.

Biznes metama'lumotlarida biznes atamalari va ta'riflari, ma'lumotlarga egalik va saqlash xizmatlarini to'lash qoidalari mavjud.

Operativ metadata - bu ma'lumotlar ombori ishlagan davr mobaynida to'plangan ma'lumotlar, ular quyidagilarni o'z ichiga oladi:

- 1) ko'chirilgan va o'zgartirilgan ma'lumotlarning kelib chiqishi;

- 2) ma'lumotlardan foydalanish holati (faol, arxivlangan yoki o'chirilgan);

- 3) monitoring ma'lumotlari, masalan, foydalanish statistikasi, xatolik to'g'risidagi xabarlar va boshqalar.

Ombor metama'lumotlari odatda repozitoriyada joylashadi (odatda omborda ma'lumotlar tarmoq orqali tarqatish uchun mavjud bo'lgan fayllar sifatida saqlanadi). Bu meta-ma'lumotlardan omborni loyihalash, o'rnatish, ishlatish va boshqarishdagi turli xil vositalar va jarayonlar bilan birgalikda foydalanishga imkon beradi.

Data Mining jarayonining boshlang'ich bosqichi

Data Mining jarayoni muayyan bosqichlardan iborat bo'lib, taqqoslash, tiplashtirish, tasniflash, umumlashtirish, mavhumlashtirish, takrorlash elementlarini o'z ichiga oladi. Data Mining jarayoni qarorlarni qabul qilish jarayoni bilan uzviy bog'liqdir. Data Mining jarayoni modelni yaratadi va qarorlarni qabul qilish jarayonida ushbu modeldan foydalaniladi.

An'anaviy Data Mining jarayoni quyidagi bosqichlarni o'z ichiga oladi:

- fan sohasining tahlili;

- masalaning qo'yilishi;

- ma'lumotlarni tayyorlash;

- modellarni qurish;

- modellarni tekshirish va baholash;
- modelni tanlash;
- modelni tadbiiq etish;
- modelni to'g'rilash va yangilash.

1-bosqich. Fan sohasini tahlil etish.

Tadqiqot - bu ma'lum bir maqsad doirasiga ega bo'lgan ma'lum bir sohani, obyekt yoki hodisani anglash jarayoni.

Tadqiqot jarayoni tadqiqotchilar nuqtai nazaridan muhimligini aniqlash va baholash uchun obyektlarning xususiyatlarini kuzatishdan iborat bo'lib, bu xususiyatlar ko'rsatkichlari o'rtasidagi doimiy aloqalarni o'z ichiga oladi.

Fan sohasi - bu tasvirlash yoki modellashtirish va tadqiq qilish kerak bo'lgan aqliy jihatdan cheklangan borliq sohasi.

Fan sohasi xususiyatlar bilan farqlanuvchi va ular o'rtasida ma'lum munosabatlarda bo'lgan yoki biron-bir tarzda o'zaro ta'sir qiladigan obyektlardan iborat.

Fan sohasi haqiqiy dunyoning bir qismidir, unda olib borilayotgan tadqiqotlar nuqtai nazaridan muhim va ahamiyatsiz ma'lumotlar mavjud. Ma'lumotlarning ahamiyati tanlangan mavzu doirasiga bog'liq.

Mavzu sohasini o'rganish jarayonida uning modelini yaratish kerak. Turli manbalardan olingan ma'lumot har qanday vositalar yordamida rasmiylashtirilishi kerak. Bu mavzu doirasining matnli tavsifi yoki ixtisoslashtirilgan grafik yozuvlar bo'lishi mumkin. Domenlarni tavsiflash usullari juda ko'p: masalan, SADT tarkibiy tahlil uslubi va unga asoslangan IDEF0, Xayn-Sarson ma'lumotlar oqimining diagrammasi, obyektga yo'naltirilgan UML tahlil uslubi va boshqalar.

Fan sohasining modeli unda ro'y beradigan jarayonlar va ushbu jarayonlarda ishlatiladigan ma'lumotlarni tavsiflaydi. Data Mining dasturini yanada rivojlantirishning muvaffaqiyati fan sohasining naqadar to'g'ri modellashtirilishiga bog'liq.

2-bosqich. Masalaning qo'yilishi.

Data Mining masalasining qo'yilishi quyidagi bosqichlarni o'z ichiga oladi:

- muammoni bayon qilish;
- muammoni rasmiylashtirish.

Masalaning qo'yilishi o'rganilayotgan obyektlarning statik va dinamik

xatti-harakatini ham o'z ichiga oladi.

Statikaning tavsifi obyektlar va ularning xususiyatlarini tavsiflashni anglatadi. Dinamikani tavsiflashda obyektlarning harakati va ularning harakatlariga ta'sir etadigan sabablar tavsiflanadi. Obyektlar harakati dinamikasi ko'pincha statika bilan bir qatorda tavsiflanadi.

Ba'zan mavzu sohasini tahlil qilish bosqichlari va masalaning qo'yilishi bitta bosqichga birlashtiriladi.

3-bosqich. Ma'lumotlarni tayyorlash.

Bosqichning maqsadi: Data Mining uchun ma'lumotlar bazasini yaratish.

Ma'lumotni tayyorlash eng muhim bosqich bo'lib, uning sifati Data Mining butun jarayonining yuqori sifatli natijalarini olish imkoniyatini belgilaydi. Ushbu bosqichdagi qadamlarni ko'rib chiqamiz.

1. Ma'lumotlarga qo'yilgan talablarni aniqlash va tahlil qilish.

Ushbu bosqichda ma'lumotlarni modellashtirish deb ataladigan, ya'ni Data Miningni amalga oshirish uchun zarur bo'lgan ma'lumotlar talablarini aniqlash va tahlil qilish amalga oshiriladi.

Shu bilan birga, ma'lumotlarni taqsimlash (geografik, tashkiliy, funksional); tahlil qilish uchun zarur bo'lgan ma'lumotlarga kirish, tashqi va / yoki ichki manbalarga bo'lgan ehtiyoj; shuningdek tizimning analitik tavsiflari (ma'lumotlarni o'lchash, chiqish hujjatlarining asosiy turlari, ma'lumotni o'zgartirish ketma-ketligi va boshqalar) o'rganiladi.

2. Ma'lumotlarni to'plash.

Tashkilotda ma'lumotlar omborining mavjudligi tahlilni osonroq va samaraliroq qiladi, undan foydalanish investitsiyalar nuqtai nazaridan alohida ma'lumotlar bazalari yoki ma'lumotlar marshrutlaridan foydalanishga qaraganda ancha arzonroq tushadi. Biroq ma'lumotlar omborlari har doim ham mavjud bo'lavermaydi. Bunday holda dastlabki ma'lumotlarning manbai operatsion, ma'lumotnoma va arxiv ma'lumotlar bazalari, ya'ni mavjud axborot tizimlaridan olingan ma'lumotlar hisoblanadi.

Bundan tashqari Data Mining uchun ma'lumotlar tashqi manbalar, qog'ozli axborot vositalarining axborot tizimlaridan, shuningdek ekspertlarning bilimlari yoki so'rov natijalaridan kerak bo'lishi mumkin.

Ushbu qadamda ba'zi ma'lumotlar kodlanadi.

Kerakli ma'lumot miqdorini aniqlashda ma'lumotlarning tartiblangan yoki yo'qligini hisobga olishingiz kerak.

Agar ma'lumotlar tartiblangan bo'lsa va vaqt ketma-ketligi bilan shug'ullanadigan bo'lsa, bunday ma'lumotlar to'plami mavsumiy /siklik tarkibiy qismlar kiritilganligini bilish tavsiya etiladi. Agar ma'lumotlar to'plamida mavsumiy / davriy tashkil etuvchi mavjud bo'lsa, kamida bitta fasl /siki uchun ma'lumotlar bo'lishi kerak.

Agar ma'lumotlar tartiblanmagan bo'lsa, ya'ni ma'lumotlar to'plamidagi voqealar vaqtga bog'liq bo'lmasa, ma'lumotlarni yig'ishda quyidagi qoidalarga rioya qilish kerak.

- Ma'lumotlar to'plamidagi yozuvlarning yetarli emasligi noto'g'ri modelning qurilishiga olib kelishi mumkin. Statistik nuqtai nazardan, model aniqligi o'rganilayotgan ma'lumotlarning ko'payishi bilan ortadi.

- Ba'zi ma'lumotlar eskirgan bo'lishi mumkin yoki noaniq vaziyatni tavsiflaydi va ularni ma'lumotlar bazasidan chiqarib tashlanishi kerak.

- Juda katta ma'lumotlar bazalarida modellarni yaratish uchun ishlatiladigan algoritmlar kengaytirilishi kerak.

- Ko'p algoritmlardan foydalanganda kirish o'zgaruvchilarining ma'lum bir (istalgan) nisbati va kuzatishlar soni kerak. Ma'lumotlar to'plamidagi yozuvlar (misollar) soni omillar (o'zgaruvchilar) sonidan ancha ko'p bo'lishi kerak.

- Ma'lumotlar to'plami iloji boricha ko'p taqdimotli va iloji boricha ko'proq vaziyatni namoyish etishi kerak. Ma'lumotlar to'plamida turli xil misollar taqdim etish nisbati haqiqiy vaziyatga mos kelishi kerak.

3. Ma'lumotlarni oldindan qayta ishlash.

Iloji boricha yuqori sifatli, ham past sifatli ma'lumotlarni tahlil qilish.

Ikkala holatda ham natijaga erishiladi. Sifatli tahlilni ta'minlash uchun ma'lumotlarni datslabki qayta ishlashni amalga oshirish kerak, bu **Data Mining** jarayonining zaruriy bosqichidir.

To'plash natijasida olingan ma'lumotlar ma'lum sifat mezonlariga javob berishi kerak. Shunday qilib, biz **Data Mining** jarayonining muhim qism bosqichi - ma'lumotlar sifatini baholashni ajrata olamiz.

Ma'lumotlar sifati (Data quality) - bu ma'lumotlar to'liqligi, aniqligi, o'z vaqtida va ma'lumotlarni talqin qilish qobiliyatini belgilaydigan mezon. Ma'lumotlar yuqori sifatli va pats sifatli bo'lishi mumkin, ikkinchisi - iflos yoki "yomon" deb yuritiluvchi ma'lumotlar.

Yuqori sifatli ma'lumotlar talqin qilinishi mumkin bo'lgan to'liq, aniq va

o'z vaqtida berilgan ma'lumotlar. Bunday ma'lumotlar sifatli natijani beradi: qaror qabul qilish jarayonini qo'llab-quvvatlaydigan bilimlarni ta'minlaydi.

Amaliy nuqtai nazardan pats sifatli yoki iflos ma'lumotlar mavjud bo'lmagan, noto'g'ri yoki foydasiz ma'lumotlar (masalan, standartga mos kelmaydigan va noto'g'ri formatda keltirilgan).

Iflos ma'lumotlarning eng keng tarqalgan turlari:

- yetishmayotgan qiymatlar;
- ma'lumotlarning nusxasi;
- shovqin va chiqindilar.

Tushirib qoldirilgan qiymatlar (Missing Values).

Ma'lumotlarning ayrim qiymatlari quyidagilarga bog'liq holda tushirib qoldirilishi mumkin:

- ma'lumotlar umuman to'planmagan bo'lsa (masalan, yoshni so'roq qilishda);

- ayrim atributlar ayrim obyektlar uchun qo'llanilmasligi mumkin bo'lsa (masalan, "yillik daromad" atributini bolaga qo'llamaymiz).

Tushirib qoldirilgan ma'lumotlarni qayta ishlash.

1. Tahlilda tushirib qoldirilgan qiymatlarga ega obyektlarni chiqarish.
2. Tushirib qoldirilgan ma'lumotlar uchun yangi qiymatlarni hisoblash.
3. Tahlil jarayonida yo'qotilgan qiymatlarni e'tiborsiz qoldirish.
4. Tushirib qoldirilgan qiymatlarni mumkin bo'lgan qiymatlar bilan almashtirish.

Ma'lumotlarni nusxasi (Duplicate Data).

Ma'lumotlar to'plamida takroriy ma'lumotlar bo'lishi mumkin, ya'ni dublikatlar. **Dublikatlar** bu barcha atributlar uchun bir xil qiymatlarga ega yozuvlardir. Ma'lumotlar to'plamida dublikatlarni mavjud bo'lishi ba'zi yozuvlarni ahamiyatini oshirish usuli bo'lib hisoblanadi. Bunday zaruriyat ba'zan ma'lumotlar to'plamini alohida yozuvlarini ajratib ko'rsatish uchun tug'iladi.

Biroq ko'p hollarda, takroriy ma'lumotlar ma'lumotlarni tayyorlashda xatolar natijasi bo'lib hisoblanadi.

Dublikatlarni qayta ishlashni ikkita varianti mavjud. Birinchi variantda dublikatlardan iborat barcha yozuvlar guruhi yo'qotiladi. Ushbu variant shunday holda qo'llaniladiki, agar dublikatlarni mavjudligi axborot ishonchsizligiga imkon bersa, uni butunlay qadrsiz qilib qo'yadi. Ikkinchi

variant dublikatlar guruhini birgina kam uchraydigan yozuvga almashishidan iborat.

Shovqin va chiqindilar.

Chiqindilar – ma'lumotlarni to'plashda qat'iy farq qiluvchi obyektlar yoki kuzatuvlardir. Shovqin va chiqindilar ma'lumotlarni tahlil qilishda yetarlicha umumiy ma'lumot bo'lib hisoblanadi. Chiqindilar alohida kuzatuvlar bo'lishi mumkin, yoki ma'lum guruhlariga birlashtirilishi mumkin. Tahlil qiluvchining vazifasi – nafaqat ularni aniqlash, balki ularning keyingi tahlil natijalariga ta'sir darajasini baholashdir. Agar chiqindilar tahlil qilinuvchi ma'lumotlar to'plamini informativ qismi bo'lib hisoblansa, ishonchli usullar va protseduralar qo'llaniladi.

Ikki bosqichli tahlilni o'tkazish amaliyoti etarlicha keng tarqalgan – chiqindilar va ularning yo'qligi bilan – va olingan natijalarni taqqoslash.

Data Miningning turlicha usullari chiqindilarga turlicha sezgirliklarga ega, ma'lumotlarni tahlil qilish usulini tanlashda ushbu faktni hisobga olish zarur. Shuningdek Data Miningning ba'zi instrumentlari shovqin va chiqindilarni tozalash uchun o'rnatilgan protseduralarga ega.

Ma'lumotlarni vizuallashtirish grafik ko'rinishidagi chiqindilarni ham ma'lumotlarini taqdim etishga imkon beradi.

Ayonki, iflos ma'lumotlarga asoslangan Data Miningning natijalari ishonchli va foydali hisoblanmaydi, chunki ma'lumotlarni tozalash juda zarur.

Ma'lumotlarni tozalash (data cleaning, data cleaning yoki scrubbing) ma'lumotlarni sifatini yaxshilash maqsadida ma'lumotlardagi xatolar va nomuvofiqliklarni aniqlash va olib tashlash uchun ishlatiladi.

Tozalash uchun maxsus vositalar odatda muayyan sohalarga - asosan ismlari va manzillariga ega, yoki ularning nusxalari bundan mustasno. O'zgarishlar qoidalar kutubxonasi shaklida yoki foydalanuvchi tomonidan interaktiv rejimda taqdim etiladi. Ma'lumotlarning o'zgarishi sxemaning moslash vositalaridan foydalangan holda avtomatik ravishda olinishi mumkin.

Ma'lumotlarni tozalash usuli bir qator mezonlarni qoniqtirishi mumkin.

1. Usul barcha xatolarni va nomuvofiqliklarni alohida ma'lumot manbalari sifatida ham va bir nechta manbalarni birlashtirish sifatida ham aniqlashi va olib tashlashi kerak.

2. Usul qo'lda tekshirish va dasturlash hajmini kamaytirish uchun belgilangan instrumentlar bilan qo'llab-quvvatlashi zarur va qo'shimcha

manbalar bilan ishlashda moslashuvchan bo'lishi lozim.

3. Ma'lumotlarni tozalash murakkab metama'lumotlar asosida bajariluvchi ma'lumotlarni o'zgartirish sxemasi bilan bog'liq ma'lumotlardan ajratilgan holda amalga oshirilmasligi zarur.

4. Tozalash va boshqa ma'lumotlarni o'zgartirish uchun xaritalash funksiyasi deklarativ ravishda belgilanishi va boshqa ma'lumot manbalarida va so'rovlarni qayta ishlashda foydalanish uchun mos bo'lishi kerak

5. Texnologik jarayon infratuzilmasi ma'lumotlar ombori uchun ayniqsa faol ravishda qo'llab-quvvatlanishi kerak, bu esa manbalar to'plami va katta hajmdagi ma'lumotlar to'plami uchun barcha o'zgartirish bosqichlarini samarali va ishonchli amalga oshirishni ta'minlaydi.

Ma'lumotlarni tozalash bosqichlari

Umuman olganda, ma'lumotlarni tozalash quyidagi bosqichlarni o'z ichiga oladi.

1- bosqich. Ma'lumotlarni tahlil qilish.

O'chirilishi kerak bo'lgan xatolar va nomuvofiqliklar turlarini aniqlash uchun ma'lumotlarni batafsil tahlil qilish zarur. Bu yerda ma'lumotlar xossalari va sifat muammolarini aniqlash to'g'risidagi metama'lumotlarni olish uchun ma'lumotlarni va ularning shablonlarini qo'lda tekshirishdan ham, maxsus dasturlardan ham foydalanish mumkin.

2-bosqich. Ma'lumotlar tartibini aniqlash va o'zgartirish qoidalari.

Ma'lumotlar manbalari soniga, birjinsli emasligi va ifloslik darajasiga qarab, ma'lumotlar etarlicha katta o'zgartirishni va tozalashni talab qiladi. Ba'zida ma'lumotlarni umumiy modelini manbalarini ko'rsatish uchun o'g'irish sxemasidan foydalaniladi; odatda ma'lumotlar ombori uchun relyasion ko'rinishdan foydalanadilar. Tozalashning birinchi bosqichlari ma'lumotlarning alohida manbalari muammolari tavsiflarini aniqlashi yoki o'zgartirishi mumkin. Keyingi qadamlar sxemani ma'lumotlar integratsiyasiga va ko'p elementlar, masalan dublikatlar muammolarini bartaraf etishga qaratilgan bo'lishi kerak. Saqlash uchun aniqlanishi bo'yicha amalga oshirish ETL (*Extract, Transform, Load*) ni aniqlash jarayonida o'zgartirish va tozalashga tegishli nazorat qilish usullarini ma'lumotlar oqimini aniqlash zarur bo'ladi.

O'chirish bilan bog'liq ma'lumotlarning o'zgarishi, shuningdek tozalash

bosqichlari imkon qadar deklarativ so'rov va xaritalash tili yordamida aniqlanishi kerak va shu bilan o'zgartirishning avtomatik generatsiya kodini ta'minlaydi. Bundan tashqari o'zgartirish jarayonida foydalanuvchi tomonidan yozilgan tozalash kodi va maxsus vositalarni ishga tushirish kerak.

O'zgartirish bosqichlari foydalanuvchi bilan tozalash mantig'iga ega bo'lmagan ma'lumotlar elementlari bo'yicha teskari bog'lanishni talab qilishi mumkin.

3-bosqich. Tasdiq.

Ushbu bosqichda jarayonning to'g'riligi va samaradorligi hamda o'zgartirishning ta'riflari aniqlanadi. Buni testlashtirish va baholash yo'li bilan amalga oshiriladi, masalan, berilgan manba misolida yoki nusxasida, - ushbu ta'riflarni qandaydir yaxshilash zarur yoki yo'qligini oydinlashtirish maqsadida. Loyihalashtirish va tasdiqlashni tahlil qilishda faqat ma'lum o'zgarishlar amalga oshirilgandan keyingina ayrim xatolar sezilarli bo'lib qolishiga bog'liq iteratsiyalar to'plami talab qilinishi mumkin.

4-bosqich. O'zgarishlar.

Ushbu bosqichda yuklash va ma'lumotlar omborini yangilash uchun ETL jarayonida yoki manbalar to'plamlari bo'yicha so'rovlarga javoblarda o'zgarishlarni bajarish amalga oshiriladi.

5-bosqich. Tozalangan ma'lumotlarning qarshi oqimi.

Alohida manbadagi xatolar bartaraf etilgandan so'ng, yaxshilangan ma'lumotlar eski ilovalarga tushishi va keyinchalik olish jarayonida qo'shimcha tozalashni talab qilmasligi uchun dastlabki manbalardagi ifloslangan ma'lumotlar tozalanganiga almashtiriladi. Tozalangan ma'lumotlar omborlar uchun ma'lumotlarni saqlash sohasida joylashgan.

O'zgarishning bunday jarayoni metama'lumotlarning katta hajmini talab qiladi (sxemalar, sxema darajasidagi ma'lumotlarning xususiyatlari va boshqalar). Boshqa holatlarda moslashuvchanlik, egiluvchanlik va foydalanish qulayligi uchun ushbu meta ma'lumotlar MBBT asosida omborda saqlanishi kerak. Ma'lumotlar sifatini saqlab turish uchun o'zgarish jarayoni to'g'risidagi batafsil axborot ham omborda, ham o'zgartirilgan ma'lumot elementlarida, xususan dastlabki ma'lumotlarning to'liqligi va yangiligi to'g'risidagi axborotni hamda o'zgartirilgan manbalarning asl manbai va ularga kiritilgan o'zgarishlar to'g'risidagi ma'lumotlarni yozilishi kerak.

3. DATA MINING METODLARI VA BOSQICHLARI. DATA MINING MASALALARI.

3.1. Ma'lumot va bilim

Data Miningning muhim jihati – bu keng matematik instrumentlar (klassik statistik tahlildan toki yangi kibernetik usullargacha) va axborot texnologiyalari sohasida so'nggi yutuqlar birlashmasidir. DataMining texnologiyasida tahlilning qat'iy shakllantirilgan usullari va noformal usullari mos holda birlashtirildi, ya'ni ma'lumotlarni miqdor va sifat jihatdan tahlil qilishlar.

Data Mining usullari va algoritmlariga quyidagilar kiradi: sun'iy neyron to'rlari, daraxt yechimlari, belgili qoidalar, yaqin qo'shni va k-yaqin qo'shni usullari, tayanch vektorlar usuli, bayes to'rlari, chiziqli regressiya, korrelyasion-regression tahlil, klaster tahlillashning ierarxik va noierarxik usullari, assotsiativ qoidalarni qidirish usullari, cheklangan perebor usuli, evolyutsion dasturlashtirish va genetik algoritmlar, ma'lumotlarni vizuallashtirishning turlicha usullari.

Data Mining texnologiyasida foydalaniluvchi analitik usullarning aksariyat qismi – bu ma'lum matematik algoritmlar va usullardir. Ularning qo'llanilishida yangi narsa – texnik va dasturiy vositalarning paydo bo'lishi imkoniyatlari tufayli muayyan muammolarni hal qilishda ulardan foydalanish imkoniyatidir.

Ta'kidlash kerakki, Data Mining usullarining aksariyati sun'iy tafakkur doirasida ishlab chiqilgan.

Usul (method) - o'zi bilan norma yoki qoidani, ma'lum yo'lni, usulni, nazariy, amaliy, bilishga oid, boshqaruv tabiatli masala yechimlarni ifodalaydi.

Algoritm (algorithm) - izlanayotgan natijani dastlabki ma'lumotlarini qayta o'zgartiruvchi qadamlar ketma-ketligiga nisbatan qo'yilgan aniq buyruqdir.

Algoritm tushunchasi elektron hisoblash mashinalaridan ancha ilgari paydo bo'lgan. Hozir algoritmlar inson faoliyatining turli sohalarida ko'pgina amaliy va nazariy masalalarni hal qilish uchun asos bo'lib hisoblanadi, ko'pchilikda ushbu masalalar kompyuterdan foydalangan holda ta'minlanuvchi yechimlardir.

Data Mining bosqichlari klassifikatsiyasi

Data Mining ikki yoki uch bosqichdan iborat bo'lishi mumkin.

1-bosqich. Erkin qidiruv (Discovery)

Erkin qidiruv bosqichida yashirin qonuniyatlar topish maqsadida ma'lumotlar to'plamini tadqiq qilish amalga oshiriladi. Bu yerda qonuniyatlar turlari xususida taxminiy farazlar aniqlanmaydi.

Qonuniyat (law) – turlicha hodisa va jarayonlarni shakllanish, rivojlanish jarayonining bosqichlari va shakllarini aniqlovchi mavjud va doimiy takrorlanuvchi o'zaro bog'lanish.

Ushbu bosqich Data Mining tizimi OLAP tizimida olinadigan shablonlarni aniqlaydi, masalan, tahlil uchun savollar to'plamini o'ylab topish va tuzish zarur. Bu yerda tahlil qiluvchi shunday ishdan ozod qilinadiki – shablonlar uning uchun tizimni qidiradi. Ayniqsa ushbu yondashuvni juda katta ma'lumotlar bazasida qo'llash foydalidir, bu yerda so'rovlarni tuzish orqali qonuniyatni topish juda qiyin, buning uchun turlicha variantlar to'plamini sinab ko'rish talab etiladi.

Erkin qidiruv quyidagi amallar bilan berilgan:

- shartli mantiq qonuniyatlarini aniqlash (conditionallogic);
- assotsiativ mantiq qonuniyatlarini aniqlash (associations and affinities);
- trendlar va tebranishlarni aniqlash (trends and variations).

Qonuniyatlarni erkin qidirish natijasida tizim mantiqiy qoidalar to'plami "agar ..., u holda ..." ni shakllantiradi. Bunda maqsad o'zgaruvchini berish talab etiladi.

Erkin qidiruv bosqichi doirasida tizim amallari quyidagilar yordamida amalga oshiriladi:

- shartli mantiq qoidalari induksiyalari (sinflashtirish va klasterlash masalalari, obyektlarni yaqin va o'xshash ixcham shaklida tavsiflash);
- assotsiativ mantiq qoidalari induksiyalari (assotsiatsiya masalalari va ketma-ketlik masalalari va ularning yordamida axborotlar olgan);
- trendlar va tebranishlarni aniqlash (bashoratlash masalasining dastlabki bosqichi).

Erkin qidiruv bosqichida tekshirish qonuniyatlarini ham amalga oshirish, ya'ni ularni qonuniyatlarni shakllantirishda qatnashmagan ma'lumotlar qismiga bo'lgan ishonchligini tekshirish. Ma'lumotlarni o'qitish va tekshirishga bo'lishni bunday qabuli tez-tez neyron to'rlar va daraxtga oid

yechish usullarida foydalaniladi.

2-bosqich. Bashoratli modellashtirish (Predictive Modeling).

Berilgan bosqichda birinchi bosqich ishlari natijalaridan foydalaniladi. Bu erda topilgan qonuniyatlar bashoratlash uchun bevosita qo'llaniladi.

Bashoratli modellashtirish quyidagi amallarni o'z ichiga oladi:

- noma'lum qiymatlarni bashorat qilish (outcomeprediction);
- jarayonlarni rivojlantirishni bashoratlash (forecasting).

Bashoratlashga oid modellashtirish jarayonida sinflashtirish va bashoratlash masalalari hal qilinadi.

Sinflashtirish masalalarini hal qilishda birinchi bosqich ishlari natijalari (qoidalar induksiyalari) ma'lum bir qiymatlarga asoslangan oldindan belgilangan ma'lum sinflardan biriga aniq ishonch (ehtimollik) bilan yangi obyektning tayinlash uchun foydalaniladi.

Bashoratlash masalalarini yechishda birinchi bosqich natijalari (trend va tebranishlarni aniqlash) maqsad o'zgaruvchisi (o'zgaruvchilari) noma'lum (tushirib qoldirilgan yoki so'nggi)bo'lgan qiymatlari uchun foydalaniladi.

Ta'kidlaymizki, erkin qidiruv (1-bosqich) umumiy qonuniyatlarni ochib beradi. U o'z tabiati bo'yicha induktivdir. Ushbu bosqichda olingan qonuniyatlar qismli holatdan umumiy holatiga shakllanadi. Natijada ushbu sinfning alohida vakillarini tadqiq qilish asosida ma'lum bir obyekt sinfi to'g'risida ba'zi umumiy bilimlar hosil bo'ladi.

Bashoratga oid modellashtirish (2-bosqich), aksincha, deduktivdir.

Ushbu bosqichda olingan qonuniyatlar umumiy holatidan qismli va yakkali holatiga shakllantiriladi. Bu yerda quyidagilar asosida obyektlarni ba'zi obyektlari yoki guruhlari to'g'risidagi yangi bilimlar olinadi:

- tadqiq etilayotgan obyektga tegishli sinf bilimlari;
- obyektlarni berilgan sinfi chegaralarida amal qiluvchi umumiy qoidalar bilimi.

Ta'kidlash kerakki, olingan qonuniyatlar, aniqrog'i ularning dizayni tahlilchi tomonidan sharhlanishi mumkin bo'lgan (yuqorida ko'rib chiqilgan qoidalar) shaffof bo'lishi mumkin va "qora qutilar" deb nomlanuvchi noshaffof bo'lishi mumkin. Oxirgi kostruksiyanı tipik misoli – neyron to'ri.

3-bosqich. Istisnolar tahlili (forensic analysis)

DataMiningning uchinchi bosqichida topilgan qonuniyatlarda aniqlangan istisnolar va anomalıyalar tahlil qilinadi.

Og'ishlarni aniqlash uchun (deviation detection) erkin qidiruv bosqichida hisoblanuvchi normani aniqlash zarur.

Normadan og'ish uchun ikkita bo'lishi mumkin. Ulardan birinchisi – og'ishlarning bir qancha qoida ko'rinishida shakllantirish mumkin bo'lgan mantiqiy izohlari mavjud. Ikkinchi variant – bu dastlabki ma'lumotlar xatolari. Ushbu holatda istisnolar tahlilining bosqichi ma'lumotlarni tozalash sifatida ishlatilishi mumkin bo'ladi.

Ma'lumotlarni intellektual tahlil qilish usullari orqali aniqlanadigan qonuniyatlar turi [10] ishda batafsil yoritilgan. V.A.Dyuku fikriga ko'ra beshta standart qonuniyat turlari mavjud. Bular tasniflash, sinflashtirish (qonuniyat, guruhlar berilmaganligi bilan sinflashtirishdan farqlanuvchi sinflashtirish (vaqtli qonuniyatlar)).

Monitoring va sust shakllangan qarorlarni qabul qilish intellektual tizimlarda ma'lumotlarni tadqiq qilish usullari [11] ishlarda keng yoritilgan bo'lib, ular quyidagilar:

- regressjon, dispersion va korrelyasion tahlil;
- empirik modellarga asoslangan aniq bir predmet sohasini tahlil qilish usullari;
- neyrotarmoq algoritmlari. Ushbu algoritmlarning g'oyasi kiruvchi parametrlar signal sifatida qaralib, ular "neyronlar" orasidagi bog'lanishlar asosida o'zgartirib va javob sifatida dastlabki ma'lumotlardagi butun tarmoqning tahlil natijasi o'lishidan iborat. Bunda dastlabki ma'lumotlar va javoblarni o'zida mujassamlashtirgan tanlanma orqali o'qitish amalga oshiriladi;

• "yaqin qo'shnilar" usuli. Bunda tarixga ega bo'lgan ma'lumotlardan dastlabki ma'lumotlarga o'xshashligi yaqin bo'lgan ma'lumotlar ajratiladi;

• yechimlar daraxti ierarxik tuzilmaga ega bo'lib, u "ha" yoki "yo'q" javoblarni nazarda tutuvchi savollar majmuasiga asoslanadi; ma'lumotlarga ishlov berishning ushbu usuli qonuniyatlarni har doim ham yuqori darajada aniqlamasada, biroq u olinadigan javoblar aniqligi tufayli ko'plab bashoratlash tizimlarida qo'llaniladi;

• klasterli modellar, ba'zida ular segmentatsiya modellari deb ham yuritiladi. Mazkur tipdagi modellar ma'lumotlar to'plamidagi bir nechta maydonlarni o'xshash qiymatlari asosida o'xshash hodisalarni yagona guruhga birlashtiradi;

• chekli tanlov algoritmlari. Ushbu algoritmlar ma'lumotlar qismlaridagi soddada mantiqiy hodisalarning kombinatsiyalari chastotalarini hisoblovchi algoritmlar hisoblanadi;

• evolyusion dasturlash. Mazkur usul g'oyasi qidiruv jarayonida boshlang'ich algoritmi rivojlantirish orqali ma'lumotlardagi o'zaro bog'liqliklarni ifodalovchi algoritmlarni generatsiyalash va topishdan iborat. Bunda ba'zi hollarda o'zaro bog'liqliklarni topish biror-bir aniqlangan funksiyalar asosida amalga oshiriladi. Masalan, polinomlar ko'rinishida.

Sust shakllangan yechimlarni qabul qilishni qo'llab-quvvatlash va intellektual monitoring tizimlarining asosiy xususiyati bu ularning ko'plab matematik apparatlarni birlashtirishi (statistikaning klassik usullaridan tortib to yangi kibernetik usullargacha) va axborot texnologiyalari sohasining so'nggi yutuqlaridan foydalanishidir. Sust shakllangan yechimlarni qabul qilishni qo'llab-quvvatlash va intellektual monitoring tizimlari texnologiyalarida qat'iy shakllangan va rasmiy bo'lmagan tahlil usullari, ya'ni ma'lumotlarni miqdoriy va sifat tahlil usullari o'zaro muvofiqlashtirilgan [12].

Intellektual monitoring tizimlarining barcha texnologik usullari tasnifi [11,12] ishda batafsil yoritilgan. Unda mazkur usullar boshlang'ich o'qituvchi ma'lumotlar bilan ishlash tamoyiliga ko'ra ikkita katta qismga ajratilgan. Ushbu tasniflashda yuqori pog'ona ma'lumotlarni intellektual tahlildan so'ng saqlanishi yoki ularni keyinchalik foydalanish uchun tozalanishi asosida aniqlanadi.

Ma'lumotlardan bevosita foydalanish yoki ularni saqlash. Mazkur holda boshlang'ich ma'lumotlar aniq detallashtirish holda saqlanadi va bashoratli modellashtirish bosqichlarida yoki istisnolar tahlilida to'g'ridan-to'g'ri foydalaniladi. O'ta katta hajmli ma'lumotlar ombori tahlilida qiyinchiliklarni foydalanish uchun ushbu turdagi usullarning muammosi hisoblanadi. Ushbu turdagi usullarga klasterli tahlil, yaqin qo'shnilar usuli, k -yaqin qo'shnilar usuli va o'xshashlik bo'yicha mulohazalash usullari kiradi.

Shakllantirilgan qonuniyatlarni aniqlash va ulardan foydalanish yoki tozalov shablonlari. Ushbu texnologiya yordamida dastlabki ma'lumotlardan ko'rinishi ma'lumotlarni tahlil qilishda foydalaniladigan usulga bog'liq bo'lgan axborotlar ajratib olinadi va biror bir formal tuzilmaga almashtiriladi. Birinchi guruh usullarida erkin qidiruv bosqichi mavjud bo'lmaganda mazkur jarayon bajariladi. Bashoratli modellashtirish va istisnolarni tahlil qilish

bosqichida erkin qidiruv bosqichi natijalaridan foydalaniladi va ular ma'lumotlar omboriga nisbatan ancha ixcham bo'ladi. Ushbu modellar tuzilmasi tahlilchilar uchun yechimlar olish yo'li bo'lishi ham bo'lmasligi ham mumkin [13].

Ushbu guruh usullariga mantiqiy usullar, vizuallashtirish usullari, kross-tabulyatsiya va tenglamalarga asoslangan usullar misol bo'ladi.

Mantiqiy yoki mantiqiy induksiya usullari noravshan so'rovlar va tahlillarni, belgili qoidalarni, yechimlar daraxti va genetik algoritmlarni qamrab oladi. Mazkur guruh usullari boshqa usullarga nisbatan tushunarli bo'lib, ular aniqlangan qonuniyatlarni rasmiylashtirib boradi. Ko'p hollarda esa rasmiylashtirish jarayoni foydalanuvchilar nuqtai-nazaridan yetarli darajada shaffof amalga oshiriladi. Bunda olingan qoidalar uzluksiz yoki diskret o'zgaruvchilarga ega bo'lishi mumkin. Yechimlar daraxti belgili qoidalar to'plamiga bir qoidani daraxt ildizidan boshlab to terminal uchigacha generatsiyalash orqali oson almashtiriladi. Yechimlar daraxti va qoida bir masalani yechish turli yo'llari bo'lib, ular bir-biridan faqat imkoniyatlari bilan farq qiladi. Bundan tashqari, qoidalarni joriy etish yechimlar daraxti induksiyasiga nisbatan tezkor bo'lmagan algoritmlar asosida amalga oshiriladi.

Kross-tabulyatsiya usullariga vakillar, bayes(ishonchli) tarmoq, kross-jadvalli vizualizatsiya usullari mansub. Kross-jadvalli vizualizatsiya usuli ma'lumotlarni intellektual tahlil qilishning biror-bir xossasiga ega emas. Biroq u analitik tizimlar qonuniyatlarini mustaqil izlaydi. Mazkur usulda axborotlarni taqdim etish ma'lumotlarni tahlil qilishning asosiy masalasini joriy etishni, ya'ni shablonlarni qidirishni ta'minlab beradi. Shuning uchun ushbu usul ma'lumotlarni intellektual tahlil qilish usullaridan biridir.

Tenglamalarga asoslangan usullar. Ushbu guruh usullarida qonuniyatlar matematik ifodalar, ya'ni tenglamalar ko'rinishida aniqlanadi. Ular sonli o'zgaruvchilar asosida ishlaydi. Bunda o'zgaruvchilarning boshqa tiplarini esa mos shaklda kodlashtirish talab etiladi. Bu esa ushbu guruh usullarini tadbqiq etish sohalarini cheklanishiga olib keladi. Biroq ular turli amaliy masalalarni yechishda keng qo'llaniladi, ayniqsa bashoratlash bilan bog'liq bo'lgan masalalarini yechishda.

Statistik usullar. Ushbu guruh usullarini asosini tashkil etadi va ular ko'plab bashoratlash masalasini yechishda qo'llaniladi. Hozirgi kunda ma'lumotlarni tahlil qilishning ko'plab statistik usullari mavjud. Masalan,

korrelyatsion-regression tahlil, dinamik qatorlar korrelyatsiyasi, garmonik tahlil va dinamik qatorlar intilishini aniqlash usullari shular jumlasidandir.

Data Mining masalalari

Data Mining texnologiyasi poydevoriga o'zida qonuniyatlarni ifodalovchi shablonlar konsepsiyasi qo'yilgan. Ushbu yashirin qonuniyatlarni aniqlash natijasida Data Mining masalalari hal etiladi. Odamga tushunarli shaklda ifodalanishi mumkin bo'lgan qonuniyatlarning turlicha tiplari ma'lum Data Mining masalalariga muvofiq keladi.

Data Mining masalalari (tasks) ba'zan qonuniyatlar (regularity) yoki texnikalar (techniques) deb ataladi.

Odatda quyidagilar ajratib ko'rsatiladi: sinflashtirish, klasterizatsiya, bashoratlash, assotsiatsiya, vizuallashtirish, og'ishlarni, baholashni, bog'lanishlarni tahlil qilishni va natijalarni xulosalashni tahlil qilish va aniqlash.

Sinflashtirish (Classification). Data Miningning oddiyroq va eng ko'p tarqalgan masalasi. Sinflashtirish masalasini yechish natijasida taqiq qilinayotgan ma'lumotlar to'plamining obyektlari guruhi tavsiflanuvchi belgilar – sinflar aniqlanadi; ushbu belgilar bo'yicha yangi obyekt u yoki boshqa sinfga o'tkazish mumkin. Sinflashtirish masalasini yechish uchun usullardan foydalanish mumkin: yaqin qo'shnilar (Nearest Neighbor); k- yaqin qo'shnilar (k-Nearest Neighbor); bayes to'rlari (Bayesian Networks); daraxt yechimlar induksiyasi; neyron to'rlar (neural networks).

Klasterlash (Clustering). Klasterlash sinflashtirish g'oyasini mantiqiy davomidir. Ushbu masala birmuncha murakkabroq, klasterlashning afzalligi shundan iboratki, obyektlar sinfi boshidan avvaldan aniqlanmagan. Klasterlash natijalari bo'lib obyektlarni guruhlariga bo'linishi hisoblanadi. Klasterlash masalasining yechish usuliga misol: neyron to'rlari ko'rinishiga asoslangan o'rgatish alohida turi "o'qituvchisiz" o'qitish - Koxonenning o'z-o'zini tartibga soluvchi xaritasi.

Assotsiatsiya (Associations). Assotsiativ qoidalarni qidirish masalasini yechishga kirishda ma'lumotlar to'plamida o'zaro bog'liq hodisalar o'rtasidagi qonuniyatlar izlanadi. Qonuniyatlarni izlashda Data Miningning avvalgi ikki masalasining assotsiatsiyasi farqi tahlil qilinayotgan obyekt xossalari asosida emas, balki bir vaqtda ro'y beradigan bir qancha hodisalar o'rtasida amalga oshiriladi. Assotsiativ qoidalarni izlash masalasini yechishning eng tanilgan

algoritmi – Apriori algoritmidir.

Ketma-ketlik (Sequence), yoki ketma-ket assotsiatsiya (sequential association). Ketma-ketlik tranzatsiyalar orasidagi davriy qonuniyatlarni topishga imkon beradi. Ketma-ketlik masalasi assotsiatsiyaga o'xshash, biroq uning maqsadi bir paytni o'zida yaqinlashayotgan hodisalar o'rtasida emas, balki vaqt bilan bog'liq hodisalar o'rtasidagi qonuniyatlarni o'rnatish bo'lib hisoblanadi (ya'ni ma'lum bir vaqt oralig'ida sodir bo'ladigan). Boshqacha aytganda ketma-ketlik hodisalarning vaqtga bog'liq zanjirlarini yuqori ehtimolligi bilan aniqlanadi. Haqiqatda assotsiatsiya nolga teng vaqtinchalik lag bilan ketma-ketlikning xususiy holatidir. Data Miningning ushbu masalasi shablonlarni ketma-ket topish masalasi deb ham ataladi (sequential pattern). Ketma-ketlik qoidasi: X hodisasidan keyin ma'lum vaqtdan keyin Y hodisa yuz beradi.

Bashoratlash (Forecasting). Bashoratlash masalasini yechish natijasida tarixiy ma'lumotlarni afzalliklari asosida butun sonli ko'rsatkichlarning tushirib qoldirilganlar yoki kelgusi qiymatlari baholanadi. Bunday masalalarni yechish uchun matematik statistika, neyron to'rlar va boshqa usullar qo'llaniladi.

Og'ishlarni yoki chiqindilarni aniqlash (Deviation Detection). Berilgan masalani yechishdan maqsad – ma'lumotlarni, o'ziga xos bo'lmagan shablonlarni aniqlash.

Baholash (Estimation). Baholash masalasi belgining uzluksiz qiymatlarini avvaldan aytishga keltiriladi.

Bog'lanishlarni tahlil qilish (Link Analysis) – ma'lumotlar to'plamida bog'liqliklarni topish masalasi.

Vizuallashtirish (Visualization, Graph Mining). Vizuallashtirish natijasida tahlil qilinuvchi ma'lumotlarni grafik tasviri yaratiladi. Vizuallashtirish masalasini yechish uchun ma'lumotlarda qonuniyat borligini ko'rsatuvchi grafik usullardan foydalaniladi. Vizuallashtirish usullariga misol - ma'lumotlarni 2-D va 3-D o'lchovlarda berish.

Xulosa chiqarish (Summarization) – maqsadi tahlil qilinuvchi ma'lumotlar to'plamidan obyektlarni aniq guruhlarini tavsiflash bo'lgan masaladir.

Data Mining masalalarini sinflashtirish

Strategiyalar bo'yicha sinflashtirishga muvofiq Data Mining masalalari quyidagi guruhlariga bo'linadi:

- o'qituvchilik o'qitish;
- o'qituvchisiz o'qitish;
- boshqalar.

O'qituvchilik o'qitish kategoriyasi Data Miningning quyidagi masalalari bilan beriladi: sinflashtirish, baholash, bashorat qilish. O'qituvchisiz o'qitish kategoriyasi klasterizatsiya masalalari bilan beriladi. «Boshqalar» kategoriyasiga avvalgi ikki bosqichni o'z ichiga olmagan masalalar kiradi.

Data Mining masalalari foydalaniluvchi modellarga bog'liq holda deskriptiv va bashoratli bo'lishi mumkin.

Ushbu sinflashtirishlarga muvofiq holda Data Mining masalalari tavsifli va bashoratli masalalar guruhlarini bilan beriladi.

Tavsifli (descriptive) masalalarni yechish natijasida tahlil qiluvchi talqin qilinuvchi ma'lumotlarni tavsiflovchi shablonlarni oladi. Ushbu masalalar tahlil qilinuvchi ma'lumotlarni umumiy tushunchasini tavsiflaydi ma'lumotlarning informativ, yakuniy, farqlovchi xususiyatlarini aniqlaydi. Tavsifli masalalar tushunchasi ma'lumotlar to'plamlarini tavsiflash va taqqoslashni o'z ichiga oladi. Ma'lumotlar to'plamlarini tavsiflash ayrim ma'lumotlar to'plamini qisqa va ixcham tavsifini ta'minlaydi. Ikki yoki undan ortiq ma'lumotlar to'plamini qiyosiy tavsifini ta'minlaydi.

Bashorat qilinuvchi masalalar (predictive) ma'lumotlarni tahlil qilishda, modelni yaratish, yangi yoki noma'lum ma'lumotlarning tendensiyalari yoki xususiyatlarini bashorat qilishga asoslangan.

Yuqorida eslatilgan sinflashtirishga yetarlicha yaqin bo'lib Data Mining masalalarini quyidagilarga bo'linishi hisoblanadi: tadqiqotlar va kashfiyotlar, bashoratlar va sinflashtirishlar, tushuntirishlar va tavsiflar.

Avtomatik tadqiqot va kashfiyot (erkin qidiruv).

Masala misoli: bozorning yangi segmentlarini kashf qilish. Masalaning berilgan sinfini yechish uchun klaster tahlil usulidan foydalaniladi.

Bashoratlash va sinflashtirish. Masala misoli: joriy qiymatlar asosida xaridlar hajmini avvaldan aytish. Usullar: regressiya, neyron to'rlar, genetik algoritmlar, daraxt yechimlar.

Sinflashtirish va bashoratlash masalalari induktiv modellashtirish deb nomlanadigan guruhni o'z ichiga oladi, natijada tahlil qilinuvchi obyekt yoki

tizimni o'rganish ta'minlanadi. Ma'lumotlar to'plami asosidagi ushbu masalalarni yechish jarayonida umumiy model va faraz ishlab chiqiladi.

Tushuntirish va tavsiflash. Masala misoli: demografik ma'lumotlar va xaridlar tarixi bo'yicha mijozlar xarakteristikasi. Usullar: daraxt yechim, qoidalar tizimi, assotsiatsiyalar qoidasi, bog'lanishlar tahlili. Umumlashgan modelni talqin qilishda tahlilchi yangi bilim oladi. Obyektlarni guruhlashtirish ularni o'xshashliklari asosida amalga oshadi.

Data Miningning asosiy qiymati – bu berilgan texnologiyaning amaliy yo'nalishi, xomaki ma'lumotlardan aniq bilimga yo'l, qaror qabul qilish mumkin bo'lgan qo'llab-quvvatlashda masalani qo'yilishidan tayyor dasturgacha bo'lgan yo'l. Sinflashtirish va klasterizatsiya masalalarini batafsil ko'rib chiqamiz.

3.2. Ma'lumotlar va bilimlar

Ma'lumotlar - biror predmet sohadagi obyektlar, jarayonlar va hodisalarni xarakterlovchi, ularning xarakterlarini bildiruvchi alohida olingan dalillardir.

EHMda qayta ishlanganda ma'lumotlar shartli ravishda quyidagi bosqichlardan o'tadi;

- ma'lumotlar, o'lchovlar va kuzatishlar natijasi sifatida;
- ma'lumotlar strukturasi (modelli) diagramma, grafik, funsiyalar kurinishida;
- ma'lumot tashuvchilar bo'lgan ma'lumotlar bazalarida.

Bilimlar esa ma'lumotlar bilan bog'langan bo'lib, ularga asoslanadi, lekin ular inson fikrlash qobiliyati natijasi bo'lib, biror amaliy faoliyatdagi olgan tajribalarni birlashtiradi. Ular empirik yul bilan hosil qilinadi.

Bilimlar bir sohadagi masalalarni yechishi mumkin bo'lgan, biror predmet oblastining aniqlangan qonuniyatidir (prinsiplar, bog'likliklar, konunlar).

EHMda qayta ishlashda bilimlar ma'lumotlar kabi ko'rinishlarni oladi:

- bilimlar inson xotirasida fikrlash natijasi sifatida;
- bilimlarni moddiy tashuvchilari (o'quv qo'llanma, uslubiy qo'llanmalar);
- bilimlar maydoni - biror predmet sohadagi asosiy obyektlarni, ularning atributlarini va bog'lovchi qonuniyatlarni shartli tavsifi;
- bilimlarni taqdim etish tillarida tavsiflangan bilimlar (produksion tillar, semantik tarmoklar, freymalar).

Ko'pincha bilimlarni quyidagi ta'rifi ishlatiladi:
Bilimlar - bu yaxshi strukturlashgan, yoki ma'lumotlar haqidagi ma'lumotlardir.

Har xil predmet bo'limlari uchun bir qancha bilimlarni tavsiflash modellari mavjud. Ularning ko'pchiligini quyidagi sinflarga ajratish mumkin:

- produksion;
- semantik tarmoqlar;
- freymalar;
- formal mantiqiy modellar.

3.2.1. Produksion model

Ushbu model qoidalarga asoslangan bo'lib, bilimlarni quyidagicha ko'rinishda tavsiflaydi:

Agar (shart), u holda (harakat).

Bu yerda shart deganda bilimlar bazasidan qidirish lozim bo'lgan ko'rinish asosida berilgan biror so'z yoki jumla, harakat deganda esa shartdagi jumla topilganda bajariladi. Ushbu shart va qadamlar boshqa biror qadam orasida, maqsadli, tizim ishini tugatuvchi bo'lishi mumkin.

Produksion modelni qo'llaganda bilimlar bazasi qoidalar to'plamidan iborat bo'ladi. Kerakli qoidani topuvchi dastur esa chiqarish mashinasi deyiladi. Ko'pincha chiqarish to'g'ri (ma'lumotlardan maqsadni qidirish) yoki teskari (maqsaddan uni tasdiqlash uchun ma'lumotlarga) bo'ladi. Ma'lumotlar bu - qoidalarni bazadan tanlovchi dasturni ishga tushirishga asos bo'luvchi boshlang'ich faktlardir.

3.2.2. Semantik tarmoqlar

Semantik termini - ma'noli degan ma'noni anglatadi, semantika esa belgilar ma'nosini aniqlovchi, ya'ni simvollar va obyektlar o'rtasidagi bog'liklikni aniqlovchi fandır.

Semantik tarmoq bu yo'naltirilgan graf bo'lib, uning uchlari tushuncha, yoylari esa tushunchalar o'rtasidagi munosabatlardir.

Tushunchalar bo'lib abstrakt va konkret obyektlar ishlatiladi, munosabatlar esa quyidagi bog'lanishlardir: bu ("is"), qismga ega ("hasrat"), ga tegishli (), "yaxshi ko'rish".

Semantik tarmoqlarning harakterli tomoni - ularda quyidagi 3 xil munosabatlar bo'lishidir:

- sinf - sinf elementi;

- xossa - qiymat;
- sinf elementi misoli.

Semantik tarmoqlarning bir necha klassifikatsiyasini keltirish mumkin.

Masalan munosabatlar turlari soni buyicha:

- bir jinsli (munosabatning bir tipiga esa);
- bir jinsli emas (har xil tipli munosabatga ega).

Munosabatlar turlariga ko'ra:

- binar (munosabat ikki obyektga bog'laydi);
- narli (ikkita dan ortiq tushunchani bog'lovchi maxsus munosabatlar bor).

Ko'pincha semantik tarmoqlarda quyidagi munosabatlar ishlatiladi:

- "qism - butun" tipli bog'lanish ("sinf - sinf qism", "element - to'plam" va hokazo);

- funksional bog'lanishlar (odatda "ishlab chiqarish", "ta'sir qiladi" kabi fe'llar bilan aniqlanadi).

- miqdoriy bog'lanishlar (ko'p, kam, teng...);
- fazoviy bog'lanishlar (...dan uzoq, ...ga yaqin, orqasida, ostida, ustida...);
- vaqtli bog'lanishlar (oldin, keyin, orasida...);
- atributli bog'lanishlar (xossaga ega, qiymatga ega...);
- mantiqiy bog'lanishlar (va, yoki, emas) va h.k.

Semantik tipdagi tarmoqda yechimni qidirish qo'yilgan savolga mos keluvchi, quyi tarmoqqa mos, tarmoq fragmentini qidirish masalasiga olib kelinadi.

3.2.3. Semantik tarmoq

Ushbu modelning asosiy afzalligi insonning uzoqqa saqlanuvchi xotirasi to'g'risidagi zamonaviy qarashlarga mosligidir.

Modelning kamchiligi - semantik tarmoqdan yechimning qidirilishini murakkabligidir.

Semantik tarmoqlarni qo'llash uchun NET deb nomlangan tilga o'xshash tarmoq tillari yaratilgan.

3.2.4. Freymlar

Freym (frame -karkas yoki ramka) so'zi 70- yillarda bilimlar strukturasi fazoviy ko'rinishda tasvirlash uchun taklif qilingan.

Freym asosida abstrakt obraz yoki holat yotadi.

Abstrakt obraz tushunchasi psixologiya va falsafadan yaxshi ma'lum. Masalan, "xona" so'zi eshitgan kishida xona obrazini qo'zg'atadi;

"6- 20 m.kv. maydonga ega, eshik va derazalarga ega, polli, to'rt devorli yashash xonasi" degandan bir narsani olsak, ma'no buziladi, masalan, derazalar so'zini olsak, yashash emas, yotish xonasi bo'lib qolishi mumkin, lekin unda teshiklar, ya'ni slotlar ayrim atributlarning to'ldirilmagan qiymatlari mavjud, masalan, oynalar soni, devor rangi va hokazo.

Freymlar nazariyasida bunday obraz freym deyiladi. Obrazni akslantiruvchi formallashgan model ham freym deyiladi. Freym strukturasi quyidagicha tasvirlash mumkin. **Freym nomi:**

- (1 - slot nomi: 1 slot qiymati),
- (2 - slot nomi: 2 slot qiymati),
- (n - slot nomi : n - slot qiymati).

Freym modeli yetarlicha universaldir, chunki u bilimlar ko'p xilligini quyidagilar yordamida akslantirishga imkon beradi:

- Freymlar - obyekt va tushunchalarni belgilash strukturalar uchun;
- Freymlar - rollar (menedjer, kassir, klient);
- Freym - ssenariyalar (bankrotlik, aksionerlik yigilishi);
- Freym holatlar (trevoga, avariya, qurilma ish rejimi).

Freymlar nazariyasining asosiy hossasi - semantik tarmoqlar nazariyasiga xos bo'lgan hossalar vorisligidir. Freymlarda ham, semantik tarmoqlarda ham vorislik AKO - bog'lanishlar orqali bo'ladi. AKO sloti o'xshash slotlar qiymatlari oshkormas holda o'tadigan ierarxiyaning yuqoriroq darajasini ko'rsatadi.

4. SINFLASHTIRISH VA KLASTERLASH

Sinflashtirish masalasi

Sinflashtirish eng oddiy va shu bilan birga eng tez-tez hal qilinadigan Data Mining masalasi bo'lib hisoblanadi. Tasniflash masalasi keng tarqalganligi tufayli ushbu tushunchaning mohiyatini aniq tushunish kerak. Bu yerda ayrim ta'riflar mavjud.

Tasniflash – o'rganilayotgan predmetlar, hodisalar, jarayonlarni jinsi, turi, tipi bo'yicha, ularni o'rganish qulayligi uchun har qanday muhim belgilar bo'yicha tizimli taqsimlash; ushbu o'xshashlik darajasini aks ettiruvchi dastlabki tushunchalarni guruhlash va ularni belgilangan tartibda joylashishi.

Tasniflash – o'xshash sinflashtirish belgilariga (bir yoki bir nechta belgilar) ega ba'zi bir tamoilga muvofiq obyektlar to'plami bo'lib, ushbu obyektlar o'rtasidagi o'xshashlik yoki farqni aniqlash uchun tanlangan.

Tasniflash quyidagi qoidalarni o'z ichiga oladi:

- bo'linmaning har bir aktida faqat bir asosni qo'llash zarur;
- bo'linma mutanosib bo'lishi kerak, ya'ni turga oid tushunchalarning hajmi bo'linuvchi turga oid tushunchaga teng bo'lishi kerak;
- bo'linma hadlari o'zaro bir-birini chiqarib tashlashi kerak, ularning hajmi bir-biriga zid bo'lmasligi kerak;
- bo'linma ketma-ket bo'lishi kerak. Shu bilan farqlanadiki:
- tashqi belgi bo'yicha amalga oshiriluvchi va predmetlar (jarayonlar, hodisalar) to'plamini kerakli tartibda berish uchun xizmat qiluvchi yordamchi (sun'iy) tasniflash;
- predmetlar va hodisalarni ichki umumiylikni tavsiflovchi muhim belgilar bo'yicha amalga oshiriluvchi tabiiy tasniflash.

Oxirgisi ilmiy tadqiqotni muhim vositasi va natijasi bo'lib hisoblanadi, shuningdek sinflashtiriluvchi obyektlar qonuniyatlarini o'rganish natijalarini taxmin qiladi va mustahkamlaydi.

Tanlangan belgilarga, ularning birlashishi va bo'lish tushunchalariga bog'liq holda tasniflash quyidagicha bo'lishi mumkin:

• oddiy – turga oid tushunchani faqatgina belgi bo'yicha va barcha turlarni ochishdan avval faqat bir marta bo'lish. Bunday tasniflashga misol bo'lib dixotomiya bo'lib, unda faqat ikkita tushuncha bo'linma hadlari bo'lib, har biri boshqasiga ziddir (ya'ni tamoilga rioya qilinadi: "A va A emas");

• murakkab – bir tushunchani turlicha asoslar bo'yicha bo'lish va bunday

bo'lishlarni bir butun holda sintez qilish qo'llaniladi. Bunday tasniflashga misol bo'lib, kimyoviy elementlarni davriy sistemasi bo'lib hisoblanadi.

Tasniflash deganda obyektlarni (kuzatishlarni, hodisalarni) ma'lum bir sinflardan biriga tayinlash tushuniladi.

Tasniflash – bu ma'lum bir guruhning xususiyatlarini aniqlash to'g'risidagi nisbiy xulosa chiqarishga imkon beruvchi qonuniyat. Shunday qilib, tasniflashni o'tkazish uchun u yoki boshqa hodisa yoki obyektga tegishli bo'lgan guruhni tavsiflovchi belgilar mavjud bo'lishi zarur (odatda, bunda allaqachon sinflashtirilgan hodisalarning tahlili natijasida ba'zi qoidalar tuziladi).

Tasniflash nazorat qilinuvchi yoki boshqariluvchi o'qitish deb ham nomlanuvchi o'qituvchili o'qitish (supervised learning) strategiyasini anglatadi.

Tasniflash masalalari odatda uzluksiz va/yoki kategoriyali o'zgaruvchilar asosida kategoriyali bog'liq o'zgaruvchini avvaldan aytish (ya'ni, kategoriya bo'lib hisoblangan bog'liq o'zgaruvchi) deb nomlanadi.

Masalan, oldindan aytish mumkinki, kim firmaning mijozlaridan bo'lsa ma'lum bir mahsulotning potensial xaridori bo'lib hisoblanadi, kimdir firmaning xizmatidan foydalanadi, kimdir esa yo'q va hokazo. Masalalarni ushbu turi binarli tasniflashga tegishli, ularda bog'liq o'zgaruvchi faqat ikkitagina qiymatni qabul qilishi mumkin (masalan, ha yoki yo'q, 0 yoki 1).

Tasniflashning boshqa turi, agar bog'liq o'zgaruvchi oldindan belgilangan sinflarni ba'zi to'plamlaridan qiymatni qabul qilishi mumkin bo'lsa, yuzaga kelishi mumkin. Masalan, mijoz avtomobilni qanday markalisini xohlashini oldindan aytib berish mumkin bo'lgan paytda. Bunday hollarda bog'liq o'zgaruvchi uchun sinflar to'plami ko'rib chiqiladi.

Tasniflash bir o'lchamli (bir belgisi bo'yicha) va ko'p o'lchamli (ikki yoki undan ko'proq belgilar bo'yicha).

Tasniflash jarayoni.

Tasniflash jarayonining maqsadi shundan iboratki, kirish parametrlari sifatida bashorat qilinuvchi atributlardan foydalaniladigan va bog'liq atribut qiymati olinadigan modelni qurish. Tasniflash jarayoni obyektlar to'plamini ma'lum mezon bo'yicha sinflarga bo'lishdan iborat. Tasniflagich belgilarning vektori bo'yicha obyektini qaysi oldindan belgilangan sinflarga tegishlilikini aniqlaydi.

Matematik usullar bilan tasniflashni o'tkazish uchun tasniflashning matematik apparatidan foydalangan holda bajarishi mumkin bo'lgan obyektning formal tavsifiga ega bo'lishi zarur. Bizni holatda bunday tavsif bilan ma'lumotlar bazasi chiqadi.

Har bir obyekt (ma'lumotlar bazasini yozib olish) obyektini ayrim xossalari to'g'risidagi axborotni tashiydi.

Dastlabki ma'lumotlar to'plami (yoki ma'lumotlar tanlanmasi) ikkita to'plamga bo'linishi mumkin: o'qitiluvchi va test qilinuvchi.

O'qitiluvchi to'plam (training set) modelni o'qitish (qurish) uchun foydalaniladigan ma'lumotlarni o'z ichiga olgan to'plam. Bunday to'plam misollarni kirish va chiqish qiymatlaridan iborat.

Chiqish qiymatlari modelni o'qitish uchun avvaldan belgilangan.

Testli (test set) to'plam misollarni kirish va chiqish qiymatlaridan iborat. Bu yerda chiqish qiymatlaridan modelni ishlash layoqatini tekshirish uchun foydalaniladi.

Tasniflash jarayoni ikki bosqichdan iborat: modelni qurish va undan foydalanish.

1. Modelni qurilishi: oldindan ma'lum sinflar to'plamini tavsiflash.

• Ma'lumotlar to'plamini har bir misoli oldindan belgilangan sinfga tegishli bo'ladi.

• Ushbu bosqichda o'qitiluvchi to'plamdan foydalaniladi, unda modelni qurish amalga oshiriladi.

• Olingan model daraxt yechim yoki matematik formulaning tasniflashga oid qoidalari bilan berilgan.

2. Modeldan foydalanish: yangi yoki noma'lum qiymatlarni tasniflash.

• Modelning to'g'riligini (aniqligini) baholash.

a) Olingan modeldan foydalanish natijalari bilan ma'lum testga oid misol qiymatlari taqqoslanadi.

b) Aniqlik darajasi – testli to'plamda to'g'ri tasniflangan misollarni foizi.

c) Testli to'plam, ya'ni qurilgan model test qilinuvchi to'plam, o'qitiluvchi to'plamga bog'liq bo'lmasligi zarur.

• Agar model aniqligi yo'l qo'yilgan bo'lsa, sinfi noma'lum bo'lgan yangi misollarni tasniflash uchun modeldan foydalanish mumkin.

Sinflashtirish jarayoni aynan modelni qurish va undan foydalanish 7-va 8-rasmlarda keltirilgan.

Tasniflash masalalarini yechish uchun qo'llaniluvchi usullar.

Tasniflash uchun turlicha usullardan foydalaniladi. Ulardan asosiylari:

- daraxt yechimlar yordamida tasniflash;
- bayes (sodda) tasniflash;
- sun'iy neyron to'rlari yordamida tasniflash;
- tayanch vektorlar usuli bilan tasniflash;
- statistik usullar, jumladan, chiziqli regressiya;
- yaqin qo'shnilar usuli tasniflash;
- CBR-usul bilan tasniflash;
- Genetik algoritmlar yordamida tasniflash.

Tasniflash aniqligi: xatolar darajasini baholash.

Tasniflash aniqligini baholash kross-tekshirish yordamida o'tkazilishi mumkin. **Kross-tekshirish** (Cross-validation) – bu kross-tekshirish to'plami deb ham nomlanuvchi testli to'plam ma'lumotlarida tasniflash aniqligini baholash jarayonidir. Testli to'plamni tasniflash aniqligi o'qitiluvchi to'plam tasniflash aniqligi bilan solishtiriladi. Agar testli to'plamni tasniflash o'qitiluvchi to'plamni tasniflash kabi taxminan aniqlik bo'yicha shunday natijalarni bersa, berilgan model kross-tekshiruvdan o'tgan bo'lib hisoblanadi.

O'qitiluvchi va testli to'plamga bo'lish ma'lum nisbatda tanlanmani bo'lish yo'li bilan amalga oshiriladi, masalan, o'qitiluvchi to'plam – uchdan ikki qism, testli to'plam – uchdan bir qism ma'lumotlardan iborat. Ushbu usuldan ko'p sonli misollarni tanlash uchun foydalaniladi. Agar tanlanma kichik hajmga ega bo'lsa, qisman kesishishi mumkin bo'lgan o'qitiluvchi va testli tanlanmadan foydalanishda maxsus usullarni qo'llash tavsiya qilinadi.

Klasterlash masalasi

Klasterlash masalasi tasniflash masalasiga o'xshash va uning mantiqiy davomi bo'lib hisoblanadi. Farqi shuki, o'rganilayotgan ma'lumotlar to'plami sinflari avvaldan belgilanmagan.

“Klasterlash” terminining sinonimlari “avtomatik tasniflash”, “o'qituvchisiz o'qitish” va “taksonomiya” bo'lib hisoblanadi.

Klasterlash obyektlar to'plamini bir turdagi guruhlariga ajratish uchun mo'ljallangan (klasterlar yoki sinflar). Agar tanlanma ma'lumotlar belgilar fazosida nuqta sifatida talqin etilsa, u holda klasterlash masalasi “nuqtalarni quyushuvi” ni aniqlashga keltiriladi.

Klasterlashning maqsadi – mavjud tuzilishni izlash.

Klasterizatsiya tavsifli protsedura bo'lib, u hech qanday statistik xulosalar qilmaydi, biroq izlanish tahlili va "ma'lumotlar tuzilishi" ni o'rganish imkoniyatini beradi.

"Klaster" tushunchasi noaniq tarzda aniqlanadi: har bir tadqiqotda o'zining "klasterlari" mavjud. Klasterni umumiy xususiyatlarga ega obyektlar guruhi sifatida tavsiflash mumkin.

Quyidagi ikki belgini klaster xususiyatlari deb atash mumkin:

- ichki bir jinslilik;
- tashqi izolyatsiya.

Ko'plab masalalarni hal qilishda tahlil qiluvchilar tomonidan beriladigan savol – bu yaqqol tuzilmalarda ma'lumotlar qay tarzda shakllantirish ya'ni taksonomiyani rivojlantirishdan iborat. Klasterizatsiyani dastlab biologiya, antropologiya, psixologiya kabi fanlar uchun eng ko'p qo'llanildi. Iqtisodiy masalalarni hal qilish uchun iqtisodiy ma'lumotlar va hodisalarning o'ziga xos xususiyatlari tufayli uzoq vaqt davomida kam foydalanilgan.

Klasterlar kesishmaydigan, yoki eksklyuziv (non-overlapping, exclusive), va kesishadigan (overlapping) ko'rinishda bo'ladi.

Ta'kidlash kerakki, klasterli tahlilning turlicha usullarini qo'llash natijasida turlicha shaklning klasteri hosil bo'ladi. Masalan, klasterning "zanjirli" turi bo'lishi mumkin, klasterlarning uzun "zanjirli" turi, klasterning uzaytirilgan shakli va h.k. bo'lganda, ayrim usullar esa klasterning erkin shaklini yaratishi mumkin.

Turlicha usullar ma'lum o'lchamdagi klasterlarni yaratishga intilishi (masalan, kichik va katta) yoki ma'lumotlar to'plamida turli o'lchamdagi klasterlarni mavjudligini taxmin qilishi mumkin. Klaster tahlilning ayrim usullari shovqinlarga yoki chiqindilarga ayniqsa ta'sirchan, boshqalarga kamroq. Klasterlashning turli usullarini qo'llash natijasida har xil natijalar olinishi mumkin, bu normal holat va u yoki boshqa algoritmnining ishining afzalligi bo'lib hisoblanadi. Klasterlash usulini tanlashda ushbu xususiyatlarni hisobga olinadi.

Hozirgi kunda klasterlashning yuzdan ortiq turlicha algoritmlarni ishlab chiqilgan.

Klasterlash yondashuvlariga qisqa tavsifnoma keltiramiz:

- Ma'lumotlarni ajratishga asoslangan algoritmlar, (Partitioning algorithms), shu jumladan iterativ;

- obyektlarni kklasterlarga ajratish;
- klasterlashni yaxshilash uchun obyektlarni iterativ qayta taqsimlash;
- ierarxik algoritmlar (hierarchy algorithms);
- aglomeratsiya: har bir obyekt dastlab klaster bo'lib hisoblanadi, bir-biri bilan bog'laydigan klasterlar kattaroq klaster hosil qiladi va hokazo;
- obyektlarni konsentratsiyasiga asoslangan usullar (density-based methods);
- obyektlarni birlashtirish imkoniyatiga asoslangan;
- shovqinlarga e'tibor bermaslik, erkin shakl klasterlarini topish;
- grid-usullar (grid-based methods);
- grid-tuzilishda obyektlarni kvantlash;
- model usullar (model-based);
- ma'lumotlarga mos keladigan klasterlarni topish uchun modeldan foydalanish;

Klasterlash sifatini baholash.

Klasterlash sifatini baholash quyidagi protseduralar asosida o'tkazilishi mumkin:

- qo'lda tekshirish;
 - nazorat nuqtalarini o'rnatish va olingan klasterlarni tekshirish;
 - o'zgaruvchilar modeliga qo'shish yo'li bilan klasterlash barqarorligini aniqlash;
 - turlicha usullar foydalanish bilan klasterlarni yaratish va taqqoslash;
- Klasterlashning turlicha usullari har xil klasterlarni yaratishi mumkin va bu normal hodisa bo'lib hisoblanadi. Biroq o'xshash klasterlarni turli usullar bilan yaratish klasterlashning to'g'riligini ko'rsatadi.

Klasterlash jarayoni.

Klasterlash jarayoni tanlangan usulga bog'liq va deyarli har doim iterativ bo'lib hisoblanadi. Bu qiziqarli jarayon bo'lishi mumkin va turli xil parametrlarni tanlash bo'yicha tajribalar to'plamini o'z ichiga oladi, masalan, masofa o'lchovlari, o'zgaruvchilarni standartlashtirish turi, klasterlar miqdori va boshqalar. Olingan natijalar shakllantirilgan klasterlarni aniq tavsiflash imkoniyati uchun o'ziga xos obyektlarni xususiyatlarini va xossalarni keyingi talqinini tadqiq qilish va o'rganishni talab qiladi.

5. DATA MINING QO'LLASH SOHASI. PROGNOZLASH VA VIZUALIZATSIYA.

Bashoratlash masalasi

Bashoratlash masalalari inson faoliyatining turlicha sohalarida, shuningdek fan, iqtisodiyot, ishlab chiqarish va boshqa ko'plab sohalarida hal qilinadi. Bashoratlash alohida xo'jalik subyektlarini ham, iqtisodiyotni ham boshqarishni tashkil qilishning muhim elementi bo'lib hisoblanadi. Bashoratlash usullarini rivojlanishi axborot texnologiyalarini rivojlanishi bilan bevosita bog'liq, jumladan Data Mining instrumentlarida amalga oshirilgan bashoratlash usullari va algoritmlarini murakkablashishi va saqlanuvchi ma'lumotlar hajmini o'sishi bilan.

Bashoratlash (forecasting) - keng ma'noda ushbu so'z kelajakdagi oldindan akslanish sifatida belgilanadi. Bashoratlash maqsadi kelgusidagi voqealarni oldindan aytish bo'lib hisoblanadi.

Bashorat qilish (prognostics) - bashoratlash nazariyasi va amaliyotidir.

Bashoratlash retrospektiv ma'lumotlar asosida aniq obyekt yoki hodisaning dinamikasini tendensiyalarini aniqlashga yo'naltirilgan, ya'ni uning keyingi va hozirgi holatlarini tahlil qilish. Shunday qilib, bashoratlash masalasini yechish ma'lumotlarni ayrim o'qitiluvchi tanlanmasini talab qiladi.

Bashoratlash - bog'liq va bog'liq bo'lmagan o'zgaruvchilar o'rtasida funksional bog'lanish o'rnatish. Marketingda odatiy masala bo'lib bozorni bashorat qilish (market forecasting) hisoblanadi. Berilgan masalani yechish natijasida belgilangan bozorni konyunkturasi istiqboli, kelgusi davrga bozor sharoitlarini o'zgarishi baholanadi, bozor tendensiyalari aniqlanadi (tarkibiy o'zgarishlar, mijozlar talablari, narxlarning o'zgarishi).

Bashoratlash masalalarini hal qilishda eng umumiy tomonlari quyidagi qism masalalarni yechishga keltiriladi:

- bashoratlash modelini tanlash;
- qurilgan bashoratning adekvatligi va aniqligini tahlil qilish.

Data Miningning ko'plab usullari tasniflash va bashoratlash masalalarini yechish uchun foydalaniladi: chiziqli regressiya, neyron to'rlari, daraxt yechimlar (ba'zan ularni bashoratlash va tasniflash daraxtlari deb ataladi).

Ikkala masalani yechishda o'qitiluvchi to'plam asosida modelni qurishning ikki bosqichli jarayonidan foydalaniladi va undan bog'liq

o'zgaruvchini noma'lum qiymatlarini oldindan aytish uchun foydalaniladi.

Tasniflash va bashoratlash masalasining farqi shundan iboratki, birinchi masalada bog'liq o'zgaruvchini sinfi oldindan aytiladi, ikkinchisida esa tushirib qoldirilgan yoki noma'lum (kelajakka tegishli) bog'liq o'zgaruvchining sonli qiymatlari.

Bashoratlash va davriy qatorlar

Bashoratlash uchun asos bo'lib davriy qatorlar ko'rinishidagi ma'lumotlar bazasida saqlanuvchi tarixiy axborot bo'lib xizmat qiladi.

Davriy qatorlar Data Mining tushunchasi mavjud (Time-Series Data Mining). **Davriy qator** - tasodifiy bo'lmagan vaqt onlarida tartiblangan qandaydir belgining kuzatiluvchi qiymatlari ketma-ketligi.

Davriy qatorning o'ziga xos misoli - birja savdolari ma'lumotlari.

Davriy qatorni tahlil qilish quyidagi maqsadda amalga oshiriladi:

- qatorning tabiatini aniqlash;
- qatorning kelgusi qiymatlarini bashoratlash.

Davriy qator tuzilishini va qonuniyatlarini aniqlash jarayonida quyidagilar aniqlanadi: shovqinlar va chiqindilar, trend, mavsumiy komponentlar, siklik komponentlar. Davriy qator tabiatini aniqlash ma'lumotlarning o'ziga xos "razvedkasi" sifatida foydalanilishi mumkin. Mavsumiy komponent mavjudligi to'g'risida tahlil qiluvchining bilimi bashoratning tuzishda ishtirok etadigan tanlanma yozuvlarini miqdorini aniqlash uchun zarur.

Shovqinlar va chiqindilar davriy qatorni qiyinlashtirib yuboradi. Chiqindilarni aniqlash va filtrlashning turlicha usullari mavjud, bu sifatliroq tahlil qilish maqsadida ularni istisno qilish imkonini beradi.

Davriy qatorni asosiy tarkibi trend va mavsumiy komponentdan iborat. Trend - vaqtda o'zgarishi mumkin bo'lgan davriy qatorning tizimli komponenti bo'lib hisoblanadi.

Trend davriy qatorga ta'sir qiluvchi umumiy va uzoq muddatli tendensiyalarni ta'sirida shakllanuvchi tasodifiy bo'lmagan funksiya deb nomlanadi. Davriy qatorlarda trendlarni topishning avtomatik usuli mavjud emas.

Bir seriyali qatorlarni bashoratlash guruhiga boshqa o'zgaruvchilarni va omillarni hisobga olmagan holda faqat ushbu o'zgaruvchiga retrospektiv ma'lumot bo'yicha bir o'zgaruvchini bashoratini qurish masalalari kiradi.

Qatorlarni ko'p seriyali yoki o'zaro ta'sir ko'rsatadigan bashoratlash guruhi bir yoki bir nechta o'zgaruvchilarga o'zaro ta'sir qiluvchi omillarni hisobga olish zarur bo'lgan paytdagi tahlillash masalasini o'z ichiga oladi.

Bir seriyali va ko'p seriyali sinflarga bo'lishdan tashqari qatorlar mavsumiy va mavsumiy bo'lmagan bo'lishi mumkin.

Oxirgi bo'linish davriy qatorda mavsumiylik kabi tarkibiy qismning mavjudligi yoki yo'qligini bildiradi, ya'ni mavsumiy komponentni kiritishni. Davriy qatorning mavsumiy tarkibiy qismi davriy qatorning takrorlanadigan tarkibiy qismidir.

Mavsumiylik xususiyati taxminan teng vaqt oralig'ida bog'liq o'zgaruvchi harakatini tavsiflovchi egri shakli o'zining o'ziga xos shaklini takrorlashni anglatadi.

Agar uning tashqi ko'rinishini ko'rib chiqishda egri shaklni takrorlanishi to'g'risidagi teng vaqt oralig'ida taxminlar qilish mumkin bo'lmasa, qatomi mavsumiy emas deb hisoblash mumkin. Ba'zan qatomi egri tashqi ko'rinishi bo'yicha uning mavsumiy yoki mavsumiy emasligini aniqlash mumkin emas.

Mavsumiy ko'pqator tushunchasi mavjud. Unda har bir qator bog'liq (maqsad) o'zgaruvchiga ta'sir qiluvchi omillar harakatini tavsiflaydi.

Bashoratlash bo'yicha masalani yechish uchun ma'lumotlarni yig'ish va omillarni tanlashda shuni yodda tutish kerakki, bu yerda tovarlarni sotish hajmining ta'siri mavsumiylik omili ta'siridan ancha past.

Ko'pincha trend va mavsumiylik davriy qatorda bir vaqtning o'zida mavjud.

Siklik qator va mavsumiylikni ajratish odatiy holdir. ssiklning davomiyligi, haqiqatda, bir mavsumiy davrdan uzunroq, ssikllar mavsumiy davrdan farqli ma'lum bir davomiylikka ega emas.

Har qanday o'zgarishlarni amalga oshirishda davriy qator tabiatini tushunish sezilarlicha osonlashadi, bunday o'zgarishlar bilan trendni olib tashlash yoki qatomi tekislash.

Bashoratlashni boshlashdan avval quyidagi savollarga javob berish zarur:

1. Bashoratlash uchun nima zarur?
2. Qanday davriy elementlarda (parametrlarda)?
3. Qanday bashorat aniqligi bilan?

Birinchi savolga javob berishda bashoratlanuvchi o'zgaruvchilarni aniqlaymiz. Bu, masalan, navbatdagi kvartaldagi mahsulotning aniq bir turini

ishlab chiqarish darajasi, ushbu mahsulotning sotish narxining bashorati va boshqalar bo'lishi mumkin. O'zgaruvchilarni tanlashda retrospektiv ma'lumotlarning mavjudligi, qaror qabul qiluvchi shaxslarning afzalligi, Data Miningning yakuniy qiymati hisobga olinadi. Ko'pincha bashoratlash masalasini yechishda o'zgaruvchini o'zini emas, balki uning qiymatlarini o'zgarishini oldindan aytish zaruriyati tug'iladi.

Bashoratlashning ikkinchi masalasini hal qilishda quyidagi parametrlarni aniqlanadi:

- bashoratlash davri;
- bashoratlash gorizonti;
- bashoratlash intervali.

Bashoratlash davri– bashorat qiluvchi asosiy vaqt birligi.

Bashoratlash gorizonti – kelajakda bashoratni qoplovchi davrlar soni.

Bashoratlash intervali– yangi bashorat qilinadigan chastota.

Parametrlarni tanlashda bashoratlash gorizonti ushbu bashorat asosida qabul qilingan qarorni amalga oshirish uchun zarur bo'lgan vaqtga nisbatan kichik bo'lmasligi kerakligini hisobga olish zarur. Faqat shu holatdagina bashoratlash mantiqqa ega bo'ladi.

Bashoratlash aniqligi bashoratlash gorizonti bilan qoidaga ko'ra pasayadi, gorizontni kamayishi bilan esa – ko'tariladi.

Bashoratlash uchun yechimni amalga oshirishga zarur bo'lgan vaqtini kamaytirgan holda, shu bilan birga bashoratlash gorizontini va xatosini kamaytirgan holda ham bashoratlash sifatini oshirish mumkin.

Bashoratlash oralig'ini tanlashda ikkita xavf-xatar tanlanadi: tahlil qilinuvchi jarayondagi o'zgarishlarni vaqtida aniqlamaslik va bashoratning yuqori qiymatini. Uzun bashoratlash oralig'ida jarayonda sodir bo'lgan o'zgarishlarni aniqlamaslik xavfi tug'iladi, qisqasida esa – bashoratlash xarajatlari oshadi. Oraliqni tanlashda tahlil qilinuvchi jarayon barqarorligini va bashoratlash xarajatlarini ham hisobga olish zarur.

Bashorat aniqligi.

Muayyan masalani yechish uchun zarur bo'lgan bashorat aniqligi bashoratlash tizimiga katta ta'sir ko'rsatadi. Bashorat xatoligi foydalaniluvchi bashorat tizimiga bog'liq. Bunday tizim qanchalik ko'p resurslarga ega bo'lsa, shunchalik aniq bashoratni olish imkoniyatlari kattaroq bo'ladi. Biroq bashorat

qilish qaror qabul qilishda xavf-xatarlarni butunlay yo'q qila olmaydi. Shuning uchun ham har doim mumkin bo'lgan bashorat xatosi hisobga olinadi.

Bashoratning aniqligi bashorat xatoligi bilan tavsiflanadi. Xatolarning eng keng tarqalgan turlari:

1. O'rtacha xato ($O^{\prime}X$). Bu har bir qadamda xatolarni o'rtachasini hisoblaydi. Xatoning ushbu turining kamchiligi – musbat va manfiy xatolar bir-birini bekor qiladi.

2. O'rtacha absolyut xato ($O^{\prime}AX$). Bu o'rtacha absolyut xatolar sifatida hisoblanadi. Agar u nolga teng bo'lsa, u holda mukammal bashoratga ega bo'lamiz. O'rtacha kvadratik xatolik bilan solishtirishda ushbu o'lcham chiqindilarga "o'ta katta qiymatni bermaydi".

3. Xatolar kvadratlari yig'indisi (XKY), o'rta kvadratik xato. Bu xatolar kvadratlari yig'indisi (yoki o'rtacha) sifatida hisoblanadi. Bu eng ko'p foydalaniladigan bashorat xatoligi bahosi.

4. Nisbiy xato (NX). Avvalgi o'lchovlarda xatolarning haqiqiy qiymatlari foydalanilgan. Nisbiy xato nisbiy xatolar terminlarida natijaga keltirish sifatini ifodalaydi.

Bashoratlar turlari.

Bashorat qisqa muddatli, o'rta muddatli va uzoq muddatli bo'lishi mumkin.

Qisqa muddatli bashorat o'zi bilan bir nechta qadam oldingi bashoratni ifodalaydi, ya'ni bashoratni qurish kuzatish hajmining 3% dan ko'p bo'lmagan yoki 1-3 qadam oldin amalga oshiriladi.

O'rta muddatli bashorat – bu kuzatuvlar hajmining 3-5% i uchun, biroq 7-12 qadamdan oshmasligi kerak; shuningdek bashoratning ushbu turi bir yoki yarim mavsumiyssiklning bashorati sifatida tushuniladi. Qisqa muddatli va o'rta muddatli bashoratlarni qurish uchun statistik usullar thla mos keladi.

Uzoq muddatli bashorat – bunda kuzatish hajmini 5% dan ko'prog'ini bashorat qilishni tushuniladi. Bashoratlarning berilgan turini qurishda amaliy jihatdan statistik usullar foydalanilmaydi, bashoratni "tuzish" mumkin bo'lgan juda "yaxshi" qatorlar holatlari bundan mustasno.

Vizualizatsiya masalasi

Vizualizatsiya – hisoblarning yakuniy natijasini ko'rish, hisoblash jarayonini boshqaruvini tashkil etish, kelgusi harakatning ratsional yo'nalishini

aniqlash maqsadida hattoki dastlabki ma'lumotlarga qaytishga imkon beruvchi vositalar to'plamidir.

Vizualizatsiyadan foydalanish natijasida ma'lumotlarning grafik tasviri yaratiladi. Vizualizatsiyani qo'llash ma'lumotlarni tahlil qilish jarayonida anomalialar, tuzilmalar, tendensiyalarni ko'rishga yordam beradi. Tasniflash va klasterlash masalalarida ikki o'lchamli fazoda obyektlarni taqsimlanishini tasvirlash uchun vizualizatsiya ham qo'llaniladi.

Vizualizatsiyani qo'llash qonuniyatlarni keltirib chiqarishni yanada iqtisodiy va sodda usulidir (tadqiqotning birinchi bosqichida). Tarqalish diagrammasida trend chizig'i yoki nuqtalar to'plami tahlil qiluvchiga qonuniyatni tezroq aniqlashga va kerakli qarorga kelishga imkon beradi. Shunday qilib, bu yerda so'z Data Miningda belgilardan emas, balki timsollardan foydalanish to'g'risida yuritiladi.

Ma'lumotlarning vizualizatsiyasi quyidagi ko'rinishda berilishi mumkin bo'ladi: grafiklar, sxemalar, gistogrammalar, diagrammalar va boshqalar.

Qisqacha vizualizatsiya rolini quyidagi imkoniyatlar bilan tavsiflash mumkin:

- interfaol va izchil tadqiqotni qo'llab-quvvatlash;
- natijalarni taqdim etishga ko'maklashish;
- vizual tasvirlarni yaratish va ularni anglash uchun aniqlik.

Bunda shuni yodda tutish zarurki, ilmsizlik bilan amalga oshirilgan vizualizatsiya kelgusidagi tahlillarda odatda aniqlanuvchi noto'g'ri xulosalarga olib kelishi mumkin.

6. BANK ISHI. SUG'URTA. TELEKOMMUNIKATSIYA. MARKETING. FOND BOZORI. BIOINFORMATIKA. MEDITSINA. FARMATSEVIKA

Sun'iy intellektni an'anaviy algoritmdan nimasi bilan ajratib turishini bilasizmi? Berilgan dastlabki ma'lumotlar to'plami uchun algoritmdan har qanday urinishlar uchun bir xil natijani beradi, ammo SI. Chunki o'quv jarayonida u o'tgan natija hozirgi natijadagidek yaxshi emasligini "hal qilishi" mumkin. Ya'ni, algoritmdan natijasini tekshirishimiz va asoslashimiz mumkin, ammo sun'iy intellekt natijasi qila olmaydi. Bu sun'iy intellektning asosiy cheklovi va asosiy imkoniyatidir.

Nima uchun cheklash kerak? Chunki sun'iy intellektning natijasini oldindan aytib bo'lmaydi. Muayyan vaziyatda u o'zini qanday tutishini oldindan ayta olmaymiz. Taxminan, xuddi odam kabi: kim qaysi daqiqada va nima asosida u yoki bu qarorni qabul qilishini biladi. Algoritmdan eng mashxur xatti-harakati sifatida juda murakkab avtomatlashtirish yordamida amalga oshirilgan bo'lib, avtomatizatsiya ob-havo sharoitiga qarab qo'nish trayektoriyasini optimallashtirish to'g'risida qaror qabul qilganida, Buran kosmik kemasining avtomatik ravishda qo'nishini esga olish mumkin.

Nima uchun imkoniyat? Chunki sun'iy intellektning asosi qat'iy algoritmdan uzozqlashish, "insonparvarlik" ga qadam qo'yish, ya'ni ma'lum darajada gumanoidga aylanishdir. Bu shuni anglatadiki, qarorlar nafaqat algoritmgacha, balki tajribaga asoslangan holda ham qabul qilinishi kerak. Aslida klassik ma'noda "agar - keyin" ustunlik qilmaydi. Shunga ko'ra, u barcha keyingi oqibatlarga olib keladigan algoritmdan yuqori bo'lishi mumkin (ijobiy ham, unchalik ham emas).

Neyron tarmoqlari va sun'iy intellekt: xuddi shu narsa?

Juda keng tarqalgan bayonot shundaki, asab tarmog'i aslida sun'iy aqldir. Bu butunlay to'g'ri emas: neyron tarmoq SI ni amalga oshirishning asosiy algoritmlaridan biri bo'lishi mumkin. Shuningdek, media-makonda ma'lum bo'lgan mavzular orasida SI sifatida mashinani o'rganish qo'llaniladi (ma'lumotlar tahlili asosida model tuzilganda). Umuman olganda, bu sun'iy intellekt deb ataladigan neyron tarmoqlari va mashinasozlik kombinatsiyasi asosida yaratilgan murakkab dasturiy ta'minot tizimlari.

Bu yerda cheklovlar haqida yana bir narsani ta'kidlash kerak: neyron tarmoqlarda qurilgan sun'iy intellekt va mashinani o'rganish, o'qitish uchun

ishlatiladigan ma'lumotlar sifatiga juda sezgir. Bir misol bilan tushuntirib beray: agar biz kameraga va uning ustiga o'rnatilgan LED belgisiga biron bir odam yaqinlashganda, unda "Xush kelibsiz, Ivan" yozuvi yonishini xohlaymiz. Buning uchun biz sun'iy intellekt Ivanni "ajratib" olishi va belgini yoqish buyrug'ini berishi uchun biz Ivanning 500 ta fotosuratini SI ga yuklaymiz. Va bu yerda bir nozik narsa bor: agar Ivanning fotosuratlari bir burchakdan olingan bo'lsa, unda SI katta ehtimollik bilan Ivanni faqat va faqat shu burchakdan ajrata oladi. Va agar u boshqa burchakdan chiqsa, belgi yonmaydi. Agar biz Ivanning har xil tomondan olingan fotosuratlarini tanlasak, u holda SI, ehtimol Ivanni har qanday nuqtadan taniydi. Amaliy masalalar:

Sun'iy aql qayerda ishlatiladi? Deyarli hamma joyda "odamdan yaxshiroq qilish" va "charchamaslik" kerak. Birinchi, sodda va tushunarli misol - "aqlli yordamchilar", chat botlari, avtomatlashtirilgan qo'llab-quvvatlash xizmatlari. Ularning yadrosida SI juda muvaffaqiyatli ishlaydi, unga mos keladigan interfeyslar "murvatlangan". Gapiyasizmi? Arzimaydi. Suhbat? Siz ham mumkin. Iltimos qilasizmi? Muammosiz.

Shu bilan birga, agar ilgari bunday dasturlar algoritmik bo'lsa (ya'ni, ular marker so'zlariga munosabat bildirishgan bo'lsa), endi to'laqonli sun'iy intellektga qadam qo'yilmoqda, ya'ni odamlar hatto boshqa tomoni - bu sun'iy intellekt.

Bundan tashqari, sun'iy intellekt yangiliklar ishlab chiqarishda dasturni topdi (hozircha faqat tajribaviy). Yangiliklar bilan 500 ta saytni ko'rib chiqing, eng qiziqarini tanlang, reytingini o'rnatib va foydalanuvchiga taqdim eting - nega SI uchun vazifa emas? Va agar siz nafaqat "o'zingizning" saytingizda, balki boshqalarda ham yoqtirishlarni tahlil qilsangiz, siz juda kuchli vositaga ega bo'lasiz, uni samaradorligi bilan tahrirlovchilarni chiqaradigan katta xodimlarining ishi bilan taqqoslash mumkin. Va agar biz sun'iy intellektga mos rasmlarni topishni "ishonib topshirsak", biz inson aralashuviz yangiliklar yig'uvchisi g'oyasiga yaqinlashamiz.

Bundan tashqari, sun'iy intellekt matnlar va she'rlar yozishi mumkin. Biroq, bu yerda muvaffaqiyatlar juda kam - sun'iy intellekt inson uchun mantiqiy bo'lgan matn yoki she'r yozishi mumkin, hatto mazmunli bo'lishi mumkin, ammo unda ijodning asosiy elementi - hissiyotlar bo'lmaydi. Muayyan mantiqqa ega quruq matn uchun bu odatiy holdir (masalan, tvit formatidagi qisqa satrdagi yangiliklar), ammo katta matnlar uchun bu

kamchilik.

Bundan tashqari, sun'iy intellekt grafikalar sohasida "o'zini ko'rsata oladi" tasvirlar bilan ishlashga imkon beradigan ko'plab davlat va xususiy xizmatlar mavjud. Aniqrog'i, shunchaki ishlamang, balki to'liq siklni bajaring - tasvirlarni yaratishdan tortib, ularni aqlli qayta ishlashgacha. Xususan, matnlar tarkibini tahlil qilish asosida (odatdagidek, beta-rejimda) matnlar uchun illyustratsiyalar yaratishga imkon beruvchi xizmat mavjud. Bu g'oya formatida ham qiziqarli ko'rinadi. Sun'iy intellekt fonni o'zi oldingi plandagi obyektlarni tanlab olib tashlashi mumkin.

Yoki, masalan, aynan shu fon xira bo'lishi mumkin (bunday funktsiya allaqachon smartfonlarda mavjud, bir so'z bilan aytganda, sun'iy intellekt biz o'ylagandan ham yaqinroq).

SI o'yinlar va ko'ngil ochish sohasida ham muvaffaqiyatli qo'llanilishi mumkin. Aniqrog'i, u yerda uzoq vaqt davomida muvaffaqiyatli ro'yxatdan o'tdi, chunki bu soha aqlli algoritmlarni va umuman SI ni rivojlantirish uchun kuchli haydovchi bo'lgan va bo'lib qolmoqda. "Odam mashinaga qarshi" o'yini 30 yil davomida tadqiqot laboratoriyalari ostonasidan chiqib ketdi. Shuning uchun, hatto telefonda oddiy shaxmatni ishga tushirganda ham, bugungi kunda sizga qarshi aniq nima o'ynayotganini aniq aytish mumkin emas: "shunchaki algoritm" yoki sun'iy intellekt.

Muhim dastur muammolari

6.1. Sun'iy intellekt moliya sohasida muvaffaqiyatli qo'llanilishi.

Bank sohasida turli xil firibgarlik holatlarini aniqlash, operatsiyalarni boshqarish va h.k. Men firibgarlikning eng oddiy holatlarini ushlab turadigan filtrlar haqida gapirmayapman. Men murakkab ishlarni o'rganadigan va aniqlaydigan va bunday tahdidlarga qanday javob berishni biladigan algoritmlar haqida gapirayapman, ham odam bilan, ham mustaqil ravishda.

Bundan tashqari, moliya sohasida sun'iy intellekt tendentsiyalarni tahlil qilish va statistika asosida statistikani bashorat qilish uchun ishlatilishi mumkin. Xususan, bunday tahlillar har xil bashoratlarni ishlab chiqarish uchun foydali bo'ladi (masalan, valyuta kurslari). Shunga qaramay, ta'kidlash joizki, "inson" prognozi kabi sun'iy intellekt amalga oshishi yoki bo'lmasligi mumkin.

Moliya bilan bog'liq bo'lgan keyingi yo'nalish - avtomatlashtirilgan

savdo, masalan, fond birjasidagi aktsiyalar. Bu yerda, ular aytganidek, sharh yo'q - bashorat qilish va sof shaklda taxmin qilish. To'g'ri, bitta lekin bor: xato bo'lsa, so'raydigan odam yo'q.

Ishga qabul qilishda sun'iy intellektdan ham foydalaniladi. Bundan tashqari, bu nafaqat ochiq kirish rezyumellarini skanerlash, balki belgilangan parametrlarga ega nomzodlarni qidirishdir. Ideal holda, ijtimoiy tarmoqlarda nomzod haqidagi ochiq ma'lumotlarni tahlil qilish va aniq ko'rsatilgan mutaxassisni tanlash bilan.

Ammo sun'iy intellekt nafaqat moliya va ishga yollashda ishlatilmoqda. Bu tibbiyotda ham keng tarqalgan. SI diagnostika, davolash rejalarini tuzish va boshqalar uchun ishlatilishi mumkin, moliya yoki kadrlar sohasidan farqli o'laroq, zararlar rubl bilan o'lchanadi, tibbiyotda hamma narsa jiddiyroq yo'qotishlar, agar bo'lsa, "odamlarda" o'lchanadi. Shuning uchun ham, va yetarli qonunchilik bazasi bo'lmasligi sababli, Rossiya Federatsiyasida tibbiyotda sun'iy intellekt juda kamdan-kam qo'llaniladi (shuni ta'kidlaymanki, sun'iy intellekt dastlabki tavsiyalar bergan va shifokor ularni tasdiqlagan yoki rad etgan holatlar bo'lishi mumkin).

Sun'iy intellektdan intensiv foydalaniladigan yana bir yo'nalish - bu mashinalar. Ushbu segmentning eng taniqli vakili sifatida o'zini o'zi boshqaradigan mashinalar. Va turli xil yordamchilar, yordamchilar va "mukammallikchilar". Masalan, avtoturargoh yordamchilari, yuzaga kelishi mumkin bo'lgan xavfli vaziyatlarni oldindan ogohlantirish tizimlari va boshqalari qat'iy algoritmik bo'lishi mumkin emas va bu yerda sun'iy intellekt paydo bo'ladi. Biroq, ishlab chiqaruvchilar o'zlarining tizimlari SI "elementlari bilan" ekanligini da'vo qilishlari ehtimoli ko'proq. Ehtimol, bu haqiqatdir va bu tizimlarda qat'iy algoritmlar ham, sun'iy intellekt ham mavjud (bu haqda aniq narsa ma'lum emas, chunki bunday o'zgarishlar haqida aniqroq ma'lumot yo'q, aniqrog'i ularni to'ldirish haqida). Bu yerda xatoning qiymati ham katta - inson hayoti.

Avtoulovlardan tashqari, sun'iy intellekt ham yo'l harakatini tartibga solish uchun foydali bo'lishi mumkin. Ikkala kameralardan va navigatorlardan olingan ma'lumotlar kirish ma'lumotlari sifatida ishlatiladi (bunday monitoringning qonuniyligi to'g'risida alohida savol). Ma'lumotlarga asoslanib, sun'iy intellekt svetoforlarning optimal ishlash rejimini va orqaga qaytish yo'llarining ochilishini tanlashi mumkin. Va eng ilg'or holatlarda -

vaqtinchalik belgilarni yoqish va o'chirish orqali harakatni boshqarish (yana, tegishli qonunga savol).

Uy sharoitida foydalanish masalalari

Borgan sari biz sun'iy aql bilan kundalik hayotda uchrashamiz. Masalan, "aqlli" uy tizimida. U yuzni tahlil qilish asosida eshiklarni qulfdan chiqarishi yoki muqobil ravishda ma'lum raqamga ega bo'lgan ma'lum markalardagi mashinalar oldida eshiklarni ochishi yoki xonadagi haroratni sozlash, egalarining xohishiga qarab sozlash, bolalar va uy hayvonlarini boshqarish, musiqa tanlashi mumkin. u kayfiyat va uni xarakterli xususiyatlar bilan aniqlaydi.

SI shuningdek tozalik va tartibni kuzatishi mumkin, masalan, egalari bo'lmagan va tozalik buzilgan bir vaqtda robot changyutgichni ishga tushirish. Yoki odamlar hayoti va sog'lig'iga jiddiy tahdid aniqlansa, favqulodda xizmatlarga qo'ng'iroq qiling.

SI banklarda nima qilishi mumkin. Bugungi kunda mashina katta ma'lumotlarga ishlov berish bilan bog'liq har qanday ishni odamga qaraganda tezroq va yaxshiroq bajarishi mumkin. Shunday qilib, banklar uzoq vaqt davomida qarz oluvchilarning asosiy skorlarini avtomatik ravishda amalga oshirib kelmoqdalar. "Shubhasiz, yaqinda biometrik ma'lumotlarni qayta ishlash bilan bog'liq yechimlar paydo bo'ladi va bu borada SI qobiliyatlari odamlarga qaraganda ancha yuqori va ishonchli bo'ladi", deb taklif qiladi Kulikov.

Sberbank sun'iy intellekt yordamida chakana savdoda kredit berish bo'yicha 100% qarorlarni qabul qiladi va ularning 95% avtomatik ravishda, inson aralashuvisiz ishlab chiqariladi. Davlat banki matbuot xizmati Frank Media bank vaqtini, xarajatlarni qisqartirishi va qarorlarni qabul qilish sifatini oshirishi mumkin bo'lgan kompyuterlarni ko'rish, tillarni qayta ishlash, nutqni tahlil qilish texnologiyalarini ishlab chiqayotganini (jarayonlarni tezlashtirish o'n baravar) ta'kidladi. Sberbank Hujjatlarni tanib olish va xodimlarning avtomatik jadvalini tuzish uchun SIDan foydalanadi. "Biz qarorlarni qabul qilish jarayonlarini avtomatlashtirish uchun aqlli boshqaruv tizimidan foydalanamiz", - deya ta'kidladi matbuot xizmati. Modellar risklarni boshqarish, biznesni rejalashtirish, muvofiqlik sohasida ishlaydi. Shuningdek, biz sun'iy intellektni mijozlar bilan chat botlari orqali muloqot qilish va shaxsiy takliflarni ishlab chiqarish uchun joriy etdik. Sun'iy intellektidagi

xatolar banklar uchun qimmatga tushishi mumkin. Sberbank rahbari German Grefning aytishicha, bank sun'iy intellekt xatolarida milliardlab rubl yo'qotgan. "Sun'iy intellekt katta tizimlarda qaror qabul qilishga moyildir. Algoritmga kirib qolgan kichik xato juda katta oqibatlariga olib kelishi mumkin", dedi Gref ushbu texnologiyani joriy etish xavfi to'g'risida savolga javob berarkan.

SI yo'nalishi Tinkoff tomonidan ham ishlab chiqilmoqda: bank SI banki savdo belgisini ro'yxatdan o'tkazishni boshladi texnologik kompaniya kontseptsiyasini kuchaytirish maqsadida, dedi Tinkoff rahbari Oliver Xyuz "Vedomosti" ga bergan intervyusida. Tinkoffda ovozli yordamchilar, chat-botlar, call-botlar va tavsiyalar beruvchi dvigatellar mavjud.

2018-yilda Gazprombank robotlashtirilgan biznes-jarayon platformasini taqdim etdi, dedi GPB bank jarayonlari boshqarmasi boshlig'ining o'rinbosari Aleksey Golubintsev Frank Mediaga. Bankda mijozlarga xizmat ko'rsatishda va ichki jarayonlar samaradorligini oshirishda yordam beradigan 20 dan ortiq robot mavjud. "Masalan, chakana kreditlashda robotlar andarrayterlar uchun ariza yuboradi", deydi Golubintsev. Bank kartalarini chiqarishda ular xodimlarni odatdagi biznes operatsiyalaridan ozod qiladilar. Va agar xodimga 2-NDFL sertifikatini kerak bo'lsa, u robot tomonidan maxsus dasturlarga murojaat qilish orqali yaratiladi.

2020-yil iyul oyida Rosbank mijozlar ma'lumotnomasini shakllantirish uchun zarur bo'lgan hujjatlarni qayta ishlash uchun aqlli dvigatellardan SI tizimini joriy qildi. Integratsiyalashgan tanib olish texnologiyalari hisobga olishni tasdiqlash talab qilinadigan hisob varaqlarni ochishda va bank operatsiyalarini amalga oshirishda mijozlar ma'lumotlarini avtomatik ravishda qayta ishlash va kiritish imkoniyatini beradi. "Sun'iy aql 2 soniyada har bir mijoz uchun skanerlash va fotosuratlardagi 70 dan ortiq ma'lumotlarni aniqlaydi va 15 ga yaqin avtomatik ma'lumotlarni tekshirishni amalga oshiradi", dedi Rosbank vakili Frank Mediaga. Avvalroq Rosbank, shuningdek, filial tarmoqlarini boshqarish uchun Atlas geografik axborot tizimining yangi joylashuv razvedka texnologiyasidan foydalanishni boshladi. Ushbu texnologiya barcha bank filiallari to'g'risidagi ma'lumotlarni to'playdi, potentsial va ish hajmini baholaydi, mijozlar, raqobatdosh banklar, aholi, shahar ko'chalaridagi transport harakati va boshqa statistik ma'lumotlarga asoslangan holda ofislarning samaradorligini hisoblab chiqadi. Natijada,

bankda har bir mavjud bo'lgan shahar uchun "issiqlik" xaritasi mavjud bo'lib, u filialni piyoda (100 m) o'tish darajasida joylashtirish imkoniyatlarini baholaydi.

Home Credit Bank mijozlarning afzalliklari asosida mahsulot takliflarini shaxsiylashtirish uchun SI dan foydalanadi. Suhbat botlarida va ovozli botlarda sun'iy intellektdan foydalanish natijalari shuni ko'rsatdiki, mijozlarning 91 foizi bot bilan maslahatlashganidan keyin 24 soat ichida bankning call-markaziga murojaat qilmaydi. "Bu shuni anglatadiki, botdan olingan javob ularga mos keldi va bot bu masalani hal qila oldi" - deydi Yuriy Popov, Home Credit Bank raqamli texnologiyalarni rivojlantirish departamenti direktori. SI shuningdek, call-markazlar va telemarketing operatorlari sifatini tahlil qilishga yordam beradi, dedi u. Masalan, Natural Language Processing algoritmlari va "Ovozdan matnga" tizimlari dialogdagi asosiy iboralarni aniqlashga, operatorlar dialoglari uchun skriptlarni optimallashtirishga va operator ssenariyni qanchalik to'g'ri bajarishini tushunishga imkon beradi. Bu bank savdosi samaradorligini sezilarli darajada oshirishga imkon beradi, deya qo'shimcha qiladi Popov.

PSB SI texnologiyasidan uchta asosiy yo'nalishda foydalanadi: risklarni boshqarish, kichik va o'rta biznes va chakana savdo, deya xabar beradi bank matbuot xizmati. Xavfni boshqarishda sun'iy intellekt firibgarlikning oldini olishga yordam beradi, kichik va o'rta biznesda, shuningdek chakana savdoda, sun'iy intellektdan foydalangan holda, bank mijozlarning xohish-istaklarini modellashtirish va prognozlash tizimlaridan foydalangan holda sotuvlar konversiyasini oshirish muammosini hal qiladi. Banklarda sun'iy intellektni tartibga solish hozirgi sharoitda nima uchun kerak va u umuman zarurmi degan savolga hali ham aniq javob yo'q

Bu yerda siz sun'iy intellekt va mashinada o'qitishni ko'rishingiz mumkin bo'lgan kundalik bank xizmatlari:

1. Telefon banki: bankka qo'ng'iroq qilganda, shaxsingizni tasdiqlashingiz kerak. SI u yerda PIN-kodlarni taxmin qilish yoki autentifikatsiya qilish uchun ovozni aniqlash yordamida ishlaydi.

2. Veb-saytlar va mobil bank xizmatlari: Ko'pgina banklar sizning hisob qaydnomangizni boshqarish, kundalik onlayn-arizalarni to'ldirish va mobil ilovalarga takliflarni yo'lakay kirish, hisob-kitoblarni to'lash, pul yuborish va boshqalarni taklif qilishadi. Sun'iy intellekt xizmatlar o'rtasidagi barcha o'zaro

aloqalarni bog'laydi va boshqaradi va sizning operatsiyalaringiz xavfsizligini ta'minlaydi. Buyuk Britaniyaning ba'zi banklari o'zlarining shaxsini tasdiqlash uchun selfi yuborish uchun hisob ochmoqchi bo'lgan mijozlarga ruxsat berishadi. Bu shuni anglatadiki, siz filialni ziyorat qilmasdan yoki qo'lda tekshirish kunlarini kutmasdan, hisobni bir necha daqiqada ochishingiz mumkin.

3. Chatbotlar: Siz bankingizga onlayn tashrif buyurganingizda yoki mobil ilovani ochganingizda, sizni ma'lumotlar yoki xizmatlarni topishga yordam beradigan sun'iy intellekt bilan ishlaydigan chatbot kutib olishi mumkin. Suhbatlashuvchi bank sizga real vaqt rejimida savollar berishga va ularga javob olishga imkon beradi. O'zaro aloqalaringiz saqlanib qoladi, shuning uchun botlar har safar ulanganingizda sizga o'zingizning xatti-harakatlaringiz va ishlatgan xizmatlaringiz asosida mahsulotlarni tavsiya qilganingizda sizga ko'proq moslashtirilgan xizmatni taqdim etishi mumkin.

4. Byudjet: Agar siz YNAB yoki EveryDollar kabi onlayn byudjet dasturlaridan foydalanayotgan bo'lsangiz, ushbu tizimlarning sizning bank operatsiyalari bilan o'zaro aloqasi, pul mablag'larini avtomatik ravishda ma'lum toifalarga o'tkazish va moliyaviy ahvolingiz haqida bashorat qilishda SIni ko'rasiz.

5. Raqamli yordamchilar bilan integratsiya: Ba'zi mamlakatlarda siz Google yoki Alexa-dan oddiy operatsiyalarni bajarishni so'rashingiz mumkin, masalan, hisoblaringizni tekshirish yoki hisobingizni to'lash, oddiy tilda.

6. Chakana to'lovlar: Xaridlarni to'lash uchun kredit yoki debet kartalaridan foydalanish o'miga, ko'p odamlar o'zlarining mobil qurilmalaridan kontaktsiz xaridlar qilishadi. Apple Pay va Google Pay kabi xizmatlar foydalanuvchilarga biometrik yuz yoki barmoq izlarini skanerlash texnologiyalari yordamida mobil qurilmalardan xavfsiz to'lovlarni amalga oshirishga imkon beradi.

7. Kredit berish: Sizning sevimli kredit tashkilotlaringizdagi sun'iy intellekt jarayonlari sizning kredit kartalaringiz, kreditlaringiz va hatto ipotekangizga murojaat qilishingiz mumkinligini anglatadi. SI ma'lum bosqichlarda yoki butun jarayonda ishtirok etadimi, bankka bog'liq. Ammo bu tezroq, aniqroq qaror qabul qilishni va siz va qarz beruvchi uchun ko'proq xavfsizlikni anglatishi mumkin.

8. Investitsiya: Katta miqdordagi ma'lumotlarni qayta ishlash qobiliyatiga

ega bo'lgan AI investitsiya banklarini ham o'zgartiradi. Tizimlar sanoatdagi o'zgarishlarga hamroh bo'lishi va vaqt o'tishi bilan yaxshiroq bashorat qilishi mumkin. Investorlar hatto ertalab ovoqli bitimlar tuzishlari mumkin.

6.2. Sun'iy intellektni tibbiyotda qo'llash g'oyasi.

1972-yilda Stenford Universitetida MYCIN ishlab chiqarilishi boshlanganda paydo bo'lgan. Bu qon bilan zaharlanish masalasini o'rganish uchun ishlatiladigan prototip SI dasturi edi. Erta sun'iy intellekt tadqiqotlari asosan Amerika muassasalarida davom etdi (MIT-Tufts birgalikda ishladilar, Stenford va Rutgers universitetlarida texnologiyani faol rivojlantirdilar. 1980-yillarda Stenford universiteti sun'iy intellekt sohasida "Tibbiy eksperimental kompyuter sun'iy intellekti" loyihasi doirasida davom etdi. tibbiyot (SUMEX-AIM).

Hisoblash quvvatining o'sishi va yangi sun'iy intellekt texnologiyalarining paydo bo'lishi tufayli bu yo'nalishdagi ishlar ancha faollashdi. Neyron tarmoqlari va mashinalarni o'rganish yordamida amalga oshirilgan navbatdagi ilmiy kashfiyot haqidagi yangiliklar muntazam ravishda paydo bo'ladi. Bugungi kunda tibbiyotda sun'iy sun'iy intellektning imkoniyatlari va istiqbollari to'g'risida qanday qiziqarli narsalarni aytib bera olasiz?

Radiologiyada SI

Ko'p sonli tibbiy tasvirlash ma'lumotlari kichik, mahalliy tizimlarda juda ko'p miqdorda saqlanadi. Agar siz ma'lumotni bulutga yuklash va ularni SIga oziqlantirish orqali chuqur o'rganishni qo'llasangiz nima bo'ladi? Mashinalar va algoritmlar tasvirlash ma'lumotlarini samarali ravishda izohlashi, naqsh va anomaliyalarni aniqlashi mumkin.

Eng aniq foydalanish holati: trening shubhali shikastlanishlari, o'smalari, ichki qonashlari, miyaning shikastlanishi va boshqalarni aniqlaydigan va lokalizatsiya qiladigan rentgenolog rentgenolog yordamchisi. Kompyuter tezroq va aniqroq ishlaydi va shuning uchun ma'lumotni qayta ishlagandan bir necha soniya o'tgach kasallik haqida aniq ma'lumotlarni berishga qodir. Inson buni qila olmaydi.

Yana bir nuqta bor. Yuqori malakali mutaxassislar qimmat va juda katta talabga ega. Ular jiddiy bosimga duchor bo'lmoqdalar, tom ma'noda ularga har tomondan quyilayotgan ma'lumotlar oqimiga botib ketishdi. Ushbu

maqolaga ko'ra, bunday mutaxassis har 3-4 soniyada tashxis qo'yishi kerak. Mashina intellekti oddiy mutaxassisning malakasini oshirishi, qiyin vaziyatlarni tartibga solishda yordam berishi mumkin. Shunday qilib, yolg'on tashxislar sonini kamaytirish va hayotni saqlab qolish.

Noyob yoki tashxisi qo'yilishi qiyin bo'lgan kasalliklarni aniqlash ko'pincha shifokorning tajribasiga, shuningdek kasallikning "beparvoligi" darajasiga bog'liq. Oddiy qilib aytganda, yara chiqmaguncha, u tan olinmasligi mumkin. Xom tasvirlarni va ayrim kasalliklar bilan bog'liq patologiyalarning ko'plab shakllarini o'z ichiga olgan katta ma'lumotlar to'plamiga kompyuterni o'rgatish orqali diagnostika sifatini va aniqlangan kasalliklar sonini yaxshilash mumkin. Ushbu g'oya AIDOC startapi tomonidan ishlab chiqilmoqda.

SI vrachlar ishining vaqt va mas'uliyatli qismini avtomatlashtirish orqali tibbiyot muassasalari ishlarining sifatini oshirishga qodir. Kompyuter algoritmlari yordamida davolash samaradorligi va bajarilgan operatsiya sifatini boshqarish, organizmning tiklanish tezligini taxmin qilish mumkin.

Microsoftning InnerEye loyihasi bu kabi texnologiyalarning yorqin namunasidir. U 3D rentgen nurlari yordamida o'smalarni segmentirovka qilish va aniqlash uchun ML texnikasidan foydalanishni taklif qiladi. Bu aniq jarrohlik rejalashtirish, navigatsiya va radiatsiya terapiyasini rejalashtirish uchun samarali o'sma konturini yaratishda yordam beradi.

Shuni ham ta'kidlash kerakki, saratonni erta bosqichda aniqlash uchun ishlatiladigan MRI va boshqa zamonaviy tasvirlash tizimlari ML bilan ishlaydi. Algoritmlar rivojlangan tasvir tahliliga yordam beradi. Masalan, prostata bezi segmentatsiyasini bajaring yoki aniqroq rasm olish uchun bir nechta turli xil rasmlarni (masalan, ultratovush, KT va MRG) birlashtiring. Mashina razvedkasi, shuningdek, rejalashtirilgan tibbiy muolajalar paytida va hatto jarrohlik aralashuvida onkologiyani taniy oladi (ko'pincha operatsiya davomida yana bir malign shakllanish sezilmasdan qoladi).

Patologiyada SI

Patologik tashxis mikroskop ostida to'qima qismini tekshirishni o'z ichiga oladi. Tasvirni aniqlash algoritmini inson tajribasi bilan birlashtirgan holda tayyorlash uchun Deep Learningdan foydalanish aniqroq diagnostikani ta'minlaydi. Raqamli tasvirlarni piksel darajasida tahlil qilish inson ko'zi osongina o'tkazib yuboradigan patologik o'zgarishlarni aniqlashga yordam

beradi. Va bu yanada samarali tashxis qo'yish imkonini beradi.

Bunday texnologiya, masalan, Garvard tibbiyot maktabi tomonidan ishlab chiqilmoqda. Algoritm nutq va tasvimi aniqlash texnologiyasidan foydalanib, anomalliklari bo'lgan rasmlarni taniy oladi va kompyuterlarni saraton va saraton bo'lmagan turlarini ajratishga o'rgatadi. Ushbu algoritmnin inson ishi bilan birlashtirish 99,5% aniqlikka olib keldi.

Mashinasozlik va tibbiyot fanlari

Petabayt ma'lumotlari barcha turdagi tibbiyot muassasalarida ishlab chiqariladi. Ushbu ma'lumotlar, afsuski, odatda tasodifiy tarqoq va tuzilmasiz. Bu hech qanday tarzda shifokorlarga nisbatan malomat emas. Ularda davolanish uchun shunchalik ko'p narsa yo'qki, ular davolanish to'g'risida xabar berishadi. Biroq, betartiblik ma'lum bir mamlakat yoki umuman dunyo sog'lig'ini rejalashtirish va global monitoringini o'tkazishga katta xalaqit beradi.

Qo'shimcha murakkablik shundaki, standart biznes ma'lumotlaridan farqli o'laroq, bemor ma'lumotlari oddiy statistik modellashirish va tahlilga mos kelmaydi. Tibbiy ma'lumotlar bazalariga kirish imkoniyatiga ega kuchli sun'iy intellekt bilan ishlaydigan bulutli platforma aralashgan ma'lumotlarni (masalan, qon patologiyasi, genetik xususiyatlar, rentgen nurlari, kasallik tarixi) samarali tahlil qilishga qodir. Shuningdek, u (nazariy jihatdan) kirish ma'lumotlarini tahlil qilishga va haddan tashqari katta miqdordagi tibbiy ma'lumot tufayli ko'rinmaydigan yashirin naqshlarni ochib berishga qodir.

Ushbu vazifalar uchun talqin qilinadigan sun'iy intellekt modellari va tarqatilgan mashinalarni o'rganish tizimlari juda yaxshi. Ular nafaqat tibbiyot fanini samarali rivojlantirishga, odamlarning yangi shakllarini va irqiy, jinsi, yosh xususiyatlarini topishga, balki aniq mintaqalarda aholining sog'lig'i holati to'g'risida aniqroq ma'lumotlarni shakllantirishga imkon beradi.

Jarrohlik robot yordamchilari

Zotan, ko'plab operatsiyalar kompyuterni ko'rish va jarroh tomonidan boshqariladigan manipulyatorlar yordamida amalga oshiriladi. Bu tibbiy texnologiyalarni rivojlantirishning muhim qismidir, inson charchash omilini tenglashtiradi va protseduralar samaradorligini oshiradi. SI-robotlar an'anaviy jarrohlarga yordam berishda juda yaxshi. Masalan;

- E'tibor berilmagan taqdirda sug'urta vazifasini bajaruvchi shifokor

ishini nazorat qilish;

• Jarroh uchun ko'rinishni yaxshilang, unga protsedura davomida harakatlar ketma-ketligini eslatib qo'ying;

• To'qimalarning aniq, minimal invaziv kesmalarini yarating;

• Kesish va tikishning optimal geometriyasini tanlash tufayli bemor uchun og'riq darajasini pasaytirish.

Ammo bunday loyihani muvaffaqiyatli amalga oshirish uchun tajriba to'plash kerak. Robot va jarrohning o'zaro aloqasi uchun dasturiy ta'minotni ishlab chiqish. Haqiqatan ham amalga oshirilgan operatsiyalar asosida (faqat odamlar ishtirokida va odam + robot ligamentida) bir qator ma'lumot to'plang.

Jarrohning harakatlarini real vaqtda boshqarish uchun kompyuter tomonidan virtual haqiqat makonini yaratish yaxshi variant bo'ladi. Nisbatan sodda operatsiyalar uchun telemeditsina va masofaviy jarrohlikdan ham foydalanishingiz mumkin.

6.3. Intellektual tizimlarni qishloq xo'jaligida qo'llanilishi

Xorijiy tajribalarning tahlili SI sohasida erishilayotgan yutuqlar va rejalashtirilgan barcha chora-tadbirlar mamlakatlarning SI texnologiyalarini rivojlantirish orqali ijtimoiy, iqtisodiy va siyosiy sohalarida milliy xavfsizligini ta'minlash va himoyalashga qaratilganligini ko'rsatadi.

Sun'iy intellekt texnologiyalari (bundan keyin SI deb ataladi; muqobilga yaqin sharoitlarda kechadigan avtomatlashtirilgan jarayonlar va hodisalar bo'lib, statistik ma'lumotlarning kritik massasi to'planishi natijasida yaxshilanish imkoniyatiga ega) ular katta hajmda va katta tezlikda xalq xo'jaligining turli sohalarida (iqtisodiy faoliyat turlari) ga kirib boradilar.

Bu kompyuter texnologiyalarining hisoblash quvvatining eksponensial rivojlanishi va odamlar tomonidan mavjud muammolarni hal qilishga qodir bo'lgan eng yaxshi yechimlarni izlab topish bilan izohlanadi. Sun'iy intellekt o'zini to'g'ri va ishonchli baholash va diagnostika masalalari (moliyaviy xavflarni, tabiiy va iqlimiy hodisalarni baholash, qishloq xo'jalik mahsulotlarini hosildorligini bashoratlash) da eng aniq va samarali usul sifatida namoyon qildi. Ushbu texnologiya ma'lumotlarni qayta ishlash va tahlil qilishning klassik usullari bilan taqqoslaganda bashoratli natijalarni yaxshiroq namoyish etadi. Hozirgi vaqtda sun'iy intellekt tizimlarining joriy etilishi boshlang'ich bo'sag'ada bo'lsada, ammo bunday dasturlarning muvaffaqiyati

texnologiyadagi kamchiliklarga qaramay (modellashirilgan sun'iy neyron tarmoqlarining optimal arxitekturasini aniqlash, mehnat talab qiladigan hisoblash algoritmlarini engib o'tishi mumkin bo'lgan kompyuter uskunalarga ehtiyoj, zarur axborot kutubxonalarini hamda ma'lumotlar bazasining yo'qligi va boshqalar), avval ishlatilgan usullarga qaraganda yaxshiroq natijalarni ko'rsatmoqda. SI qo'llab-quvvatlaydigan hisoblash tizimlari aniqroq, ishonchli va samarali natijalarni beradi.

Fermer xo'jaligi va agrosanoat majmui har qanday ijtimoiy-iqtisodiy tizimning muhim qismidir. Buning sababi shundaki, ushbu tarmoqlar va boshqaruv sohalari majmui odamlarni oziq-ovqat bilan ta'minlaydi, bu esa o'z navbatida inson hayotining muhim sharti bo'lib, uning asosiy ehtiyojlarini ta'minlaydi. Bundan tashqari, shuni ta'kidlash kerakki, yuqori sifatli fermer xo'jaligi mahsulotlari odamlarning hozirgi sog'lig'i va uning davomiyligi bilan bevosita bog'liqdir, chunki agrosanoat majmuasi faoliyati oziq-ovqat bo'lib, ular o'z navbatida inson tanasi uchun qurilish materialini vazifasini bajaradi. Dunyo aholisining ko'payishi va qonuniy adolatga intilish bilan oziq-ovqat tanqisligi muammosi yanada keskinlashadi va shu bilan birga ularning sifatini oshirish zarurati kuchayadi. Sun'iy intellekt texnologiyalari ushbu vazifalarni hal qilishga ko'p jihatdan yordam beradi, bu esa iqtisodiy va boshqaruv jarayonini, shu jumladan ekish ishlarini optimallashtirish, o'rim-yig'im, ekin o'sadigan tuproqlarning holatini kuzatish, ozuqa aralashmalarining to'g'ri tarkibi va zararkunandalarni yo'q qilish, hayvonlarni oziqlantirishni avtomatlashtirish va boshqalarni avtomatlashtiradi. Yuqorida bayon qilinganlar mazkur tadqiqot ishining maqsadini asoslab beradi. Bu fermer xo'jaligida istiqbolli sun'iy intellekt texnologiyalari to'g'risidagi bilimlarni umumlashtirish va tizimlashtirishdan iborat bo'lib, ulardan agrosanoat majmuasining asosiy vazifalari uchun (aholini sifatli oziq-ovqat bilan ta'minlash) va fermer xo'jaligining qisqa muddatli istiqbolda bozordagi raqiblariga nisbatan yangi raqobatdosh ustunliklarini olish uchun muvaffaqiyatli foydalanish mumkin. Shunga muvofiq, ushbu ish doirasida bir qator vazifalar bajarildi:

Fermer xo'jaligi va agrosanoat sohasida muvaffaqiyatli qo'llanilishi mumkin bo'lgan istiqbolli sun'iy intellekt texnologiyalari aniqlandi, ulardan foydalanishning aniq yo'nalishlari namoyish etildi; tahlil qilingan texnologiyalarning o'ziga xos xususiyatlari aniqlandi va fermer xo'jaligida

ulardan foydalanishning haqiqiy misollari namoyish etildi.

SI texnologiyalaridan foydalanishning va hozirgi kunda ularni fermer xo'jaligida ishlatishga urinishlar mavjud istiqbolli yo'nalishlaridan biri - bu kompyuterli tanib olish asosida ma'lumotlarni tahlil qilish va qayta ishlatishdir. Bu fanda noravshan to'plamlar va sun'iy neyron tarmoqlari deb ataladigan ma'lum bir matematik modellarni shakllantirishga asoslangan matematik kompyuter algoritmlarini ishlab chiqishdan iborat (elementlarning o'zaro ta'siri tirik organizmlarning biologik neyron tarmoqlariga o'xshash). Ushbu texnologiya o'simliklar xosildorligini bashoratlashni o'z ichiga olgan bir qator vazifalarni muvaffaqiyatli hal qiladi, bu esa to'g'ri va o'z vaqtida vujudga keladigan muammolarni hal etish va ularning paydo bo'lishini oldini olishga imkon beradi. Bunday hodisalarni tahlil qilish uchun material mikro va makro darajada (hayvonlar populyatsiyasi, ekin maydonlari) olingan ma'lumotlar bo'lishi mumkin. Ikkala holatda ham progressiv patologiyalar aniqlanishi mumkin, buning asosida ularni yo'q qilish uchun zarur bo'lgan tezkor choralar ko'riladi.

Shuningdek, fermer xo'jaligidagi obyektlarda o'rnatilgan elektron kompyuter sensorlar asosida statistik ma'lumotlarni to'plashni o'z ichiga olishi kerak, bu esa dinamikada talab qilinadigan ko'rsatkichlarning o'zgarishini kuzatib boradi va vaziyatga to'g'ri yo'lda aralashish va uzatish uchun menejerni belgilangan standartlardan chetga chiqish to'g'risida xabardor qiladi. Bugungi kunda bozorda aniq qishloq xo'jaligini amalga oshiruvchi tashkilot sifatida namoyish etilayotgan IntelinAir hozirda shunday tadbirlarni amalga oshirmoqda. Doimiy qayd etiladigan ko'rsatkichlar asosida biznes yuritish va oshirmoqda. Deganda aynan shu tushuniladi. IntelinAir hozirda uni optimallashtirish deganda aynan shu tushuniladi. IntelinAir hozirda Illinoys, Ayova, Ogayo va Indiana shtatidagi makkajo'xori va soya ishlab chiqaruvchilarini qo'llab-quvvatlashga qaratilgan. Kompyuterli tanib olishga asoslangan ma'lumotlarni tahlil qilish va ishlov berish sohasidagi yana bir tashkilot - Saskachevan universitetida joylashgan Kanadalik o'simliklarni fototiplash va tasvirlash tadqiqot markazi. Raqamli texnologiyalar rivojlanishi bilan turli xil agrosanoat loyihalarida va ilmiy-tadqiqot loyihalarida muvaffaqiyatli ishlatilishi va qo'llanilishi mumkin bo'lgan turli xil qishloq xo'jalik ekinlari bo'yicha ma'lumotlar parametrlarining statistik kutubxonalarini shakllantirilmoqda.

Insonning kundalik faoliyatida elektron kompyuterlarning imkoniyatlari

rivojlanishi bilan robototexnika tobora ko'proq jalb qilinmoqda, ular asta-sekin qishloq xo'jaligida paydo bo'la boshladi va tarqala boshladi. Ushbu texnologiya tadqiqot uchun qimmatli vaqtni bo'shashtirib, noan'anaviy, ammo hal qilinmagan muammolarni hal qilish uchun zarur bo'lgan ishlanmalar, biznes-jarayonlarni avtomatlashtirish imkonini beradi.

Shuni ta'kidlash kerakki, ushbu texnologiya oldingi yo'nalish bilan chambarchas bog'liq va robot tizimlarning ichki jarayonlarini optimallashtirish uchun ishlatilishi mumkin, shuningdek, qishloq xo'jaligi va agrosanoat obyektlaridan kerakli ma'lumotlarni to'plash uchun platforma vazifasini bajaradi. Prospero roboti qishloq xo'jaligidagi inson funksiyalarini taqlid qiladigan robot texnologiyasiga misoldir. U quyidagi imkoniyatlarga ega: ekin uchun teshik qazish va uni keyinchalik joylashtirish, kuzatish va yig'ib olishgacha ko'chatlarni individual ravishda ekish. Kompyuter tizimida landshaftni aniqlash algoritmlari mavjud bo'lib, ularning asosida kerakli ekinlarni etishtirish uchun ideal sharoitlar tanlangan. Daniyaning Orxus universiteti olimlari va tadqiqotchilari "Hortibot" robotini ishlab chiqdilar, u qishloq xo'jalik ekinlarini tanib olish va shunga muvofiq, begona o'tlarni mexanik ta'sir yoki kimyoviy reagentlar yordamida yo'q qilish qobiliyatiga ega. Ushbu rivojlanish ilmiy dunyoda muhim yutuq deb hisoblanadi, chunki ilgari foydali qishloq xo'jalik ekinlarini begona o'tlardan sifatli ajratish imkoni bo'lmagan. Qishloq xo'jaligidagi robototexnika va sun'iy intellekt texnologiyalarini qo'llab-quvvatlovchi tizimlarni joriy etish kabi noyob unikal loyihalarga qo'shimcha ravishda, hozirda quyidagilar mavjud.

Fermer xo'jaligidagi amaliyotga o'simliklar va hayvonlar uchun zarur bo'lgan resurslarni mahalliy yetkazib berish, inson hayotiga tahdid solmaydigan xavfli kimyoviy moddalarni tashish, shuningdek, kompyuter tizimlarida muvaffaqiyatli ishlatilishi mumkin bo'lgan fotosuratlarini olish mumkin bo'lgan uzoq masofaga uchadigan avtotransport vositalarini (dronlar) joriy etish;

traktorlar, kombaynlar, ekskavatorlar ko'rinishidagi ancha yuqori texnologiyali va murakkab uskunalar bo'lgan avtomatlashtirilgan qishloq xo'jaligi komplekslar: turli xil funksiyalarni bajarishga qodir: ekish va o'rim-yig'im, o'simlik va hayvonot ekinlarining rivojlanishini nazorat qilish, zararkunandalarni va tabiiy-iqlimiy hodisalarning oqibatlarini bartaraf etish.

Mavjud amaliyotga ko'ra, hozirgi paytda qishloq xo'jaligi hayvonlarida

olib borilayotgan va ularning sog'lig'ini baholash uchun ishlatiladigan protseduralar hayvonlarning holatini og'irlashtirmoqda, ularni stressga uchrashiga sabab bo'lmoqda, bu esa agrosanoatning tayyor mahsulotlar sifatiga salbiy ta'sir ko'rsatmoqda (mahsulot yaratishda ishlatiladigan hayvonlarning foydali vaznini yo'qotishi). Ishlab chiquvchilar tomonidan rejalashtirilgan algoritmlar go'sht mahsulotlari vaznini aniqlashga imkon beradi va mahsulot ishlab chiqarishda nafaqat katta hajmdagi savdo-sotiqni olish bilan birga bir vaqtning o'zida ularning sifatini yaxshilash uchun ozuqa ratsionini tuzish bo'yicha tavsiyalar berishga qodir.

SI texnologiyalaridan foydalanish bilan bevosita bog'liq bo'lgan ushbu ishda ta'kidlangan uchta texnologiya endi tobora ko'proq fermer xo'jaligining oldingi qatoriga o'tadi. Buning sababi, fermer xo'jaligini avtomatlashtirish agrosanoat majmuasidagi mavjud muammolarni maqbul ravishda hal qiladi, bu esa tayyor mahsulot sifatini pasaytirish xavfisiz kerakli hajmdagi ekinlar va oziq-ovqat mahsulotlarini yetishtirish jarayonini tezlashtirish imkonini beradi. Tahlil jarayonida quyidagi ijobiy tomonlarga e'tibor qaratish lozim:

Rejalashtirilgan ko'rsatkichlarga erishmaslik xavfini pasaytirish yoki ularni o'z vaqtida belgilash va yangi standart qiymatlarni asoslash.

Qishloq va agrosanoat obyektlarini himoya qilish uchun zarur choralarni o'z vaqtida ko'rishga imkon beradigan tabiiy-iqlim sharoitidagi o'zgarishlarga tezkor javob qaytarish.

Sayyoramizning o'sib borayotgan aholisini oziq-ovqat bilan ta'minlashga imkon beradigan qishloq xo'jaligi ekinlari hosildorligining o'sishi va hayvonlarning mahsuldorligi hajmining ko'payishi.

Aniq ishlab chiqarish va uni samarali va avtomatlashtirilgan boshqarish uchun operatsion ma'lumotlarni to'plash tamoyillarini amalga oshirish asosida ishlab chiqarish xarajatlarini kamaytirish.

Ishlab chiqaruvchidan yakuniy iste'molchigacha bo'lgan vositachilar sonini kamaytiradigan logistika masalalarini yechish, bu iste'molchi mahsulot tannaxsini pasaytirishda aks ettirilishi kerak.

Vaqt o'tishi bilan malakali ishchi kuchi tanqisligini kamaytirish va sun'iy intellekt texnologiyasini qo'llab-quvvatlaydigan kompyuter tizimlarini rivojlantirish uchun zarur bo'lgan yuqori texnologiyali ish joylarini yaratish.

Fermer xo'jaligi mahsulotlari ishlab chiqaruvchilarini ham, ularning mijozlarini zarur ma'lumotlar bilan tezda xabardor qilish.

Rejada agrosanoat majmuasida SI texnologiyalarining ta'kidlangan xususiyatlariga qo'shimcha ravishda, ushbu ishda sun'iy intellekt usullaridan foydalanish mumkin bo'lgan fermer xo'jaligi profilining o'ziga xos yo'nalishlari keltirilgan.

SI texnologiyalari juda katta miqdordagi sonli ma'lumotlarni avtomatik tahlil qilishga asoslangan iqtisodiyotni global raqamlashtirish bilan bog'liq. Xulosa qilib shuni ta'kidlaymizki, sun'iy intellekt texnologiyalari miqdoriy ma'lumotlarni yaxshiroq qayta ishlash sharoitida o'zini muvaffaqiyatli namoyon etadi. Shu bilan birga, tahlil qilingan ma'lumot tarqoq va noravshan ko'rinishda bo'lishi mumkin. Shu munosabat bilan hozirgi paytda SI texnologiyasi hamma joyda joriy etilmoqda. Fermer xo'jaligining tarmoqlari bundan mustasno emas, chunki bu yo'nalishda kombinator portlashi muammosiga asoslangan muammolarni samarali hal qilish mumkin (tez tahlil qilib bo'lmaydigan ma'lumotlarning katta miqdori, natijada muammoni shu tarzda hal qilish muhim emas). Bularga fotosuratlar natijasida olingan ma'lumotlarga bog'langan vazifalar, qishloq xo'jaligini samarali bajarishga qodir bo'lgan inson singari avtonom tizimlarni modellashtirish va olingan ma'lumotlarni tezkor tahlil qilish kiradi. Ushbu yo'nalishda SI texnologiyalaridan foydalanish, yuqorida aytib o'tilganidek, nafaqat fermer xo'jaligi mahsulotlarining hajmi va sifatini oshirishga, balki agrosanoat majmui korxonalariga o'z faoliyat sohasida yetakchi bo'lishga imkon beradi. Amalga oshirilgan ishlarning natijalari, mavjud bo'lgan sun'iy intellekt texnologiyalari to'g'risidagi bilimlarni tizimlashtirmoqda, ular turli darajada fermer xo'jaligida va agrosanoatda muammolarni hal qilishning muvaffaqiyatli usullari ekanliklarini isbotladilar va yaqin 5-7 yil ichida ushbu yo'nalishda faol tarqaladilar. Ularning rivojlanishi endi xo'jalik yurituvchi sub'ektga tadqiqot ishida ko'rib chiqilgan texnologiyalarning innovatsiyasi tufayli ustuvor raqobatdosh ustunliklarni va tegishli iqtisodiy samarani olish imkonini beradi. Tadqiqotning keyingi bosqichlari inson hayotining turli sohasida (qishloq xo'jaligida va agrosanoatda) eng yaxshi foydalanishni topish, ularning bir-biri bilan samarali kombinatsiyasini aniqlash va eng maqbul ichki tuzilmani (matematik va dasturiy ta'minotni) asoslash maqsadida sun'iy intellektning turli xil turlari texnologiyalari va usullarini batafsil o'rganish bo'ladi (matematiklashtirish va dasturiy kod).

6.4. Intellektual tizimlarni transport sohasida qo'llanilishi

2017-2021-yillarda O'zbekiston Respublikasini rivojlantirishning beshta ustuvor yo'nalishi bo'yicha Harakatlar strategiyasini "Ilm, ma'rifat va raqamli iqtisodiyotni rivojlantirish yili"da transport infratuzilmasini rivojlantirish sohasida ishlar sifatini oshirish, transport oqimini boshqarishda zamonaviy axborot-kommunikatsiya texnologiyalarini joriy etishda eskirgan normalarni qayta ko'rib chiqish, transport xizmatlari standartlari va sifatini yaxshilash hamda ularni jahon talablari darajasiga ko'tarish, transport xizmatlari xavfsizligi, ommabopligi, ishonchliligi va barqarorligini ta'minlash, infratuzilmani yaxshilash, moslashuvchan tariflar qo'llash va yangi istiqbolli yo'nalishlarni shakllantirish, transport tizimini raqamlashtirish, aholining shahar yo'lovchi tashish transporti xizmatlariga bo'lgan talabini to'liq qondirish bo'yicha bir qator maqsadli ishlarni amalga oshirish zaruriyati yuzaga kelmoqda.

Shahar yo'lovchi transportida tashish xavfsizligi va sifatini oshirish, tashish samaradorligini oshirish, avtobus yo'nalishlari harakatini yo'lovchi transportining boshqa turlari, jumladan, Toshkent metropolitenining amaldagi va yangidan barpo etilayotgan liniyalari bilan o'zaro uzviy bog'liqlikda yanada optimallashtirish, harakatlanadigan tarkibning ratsional miqdorini aniqlash maqsadida yo'nalishlar bo'yicha yo'lovchilar oqimini o'rganib chiqish, yo'lovchi tashishlarni tashkil etishning prinsipial yangi tizimini joriy etish, yo'lovchilar tashishni boshqarishga zamonaviy axborot-kommunikatsiya texnologiyalarini keng joriy etish yuzasidan bir qator maqsadli ishlar amalga oshirilmoqda.

Bunda Toshkent shahar jamoat transporti yo'lovchi tashish yo'nalishlari tizimini kompleks rejalashtirish maqsadida, modellashtirishni joriy etish bo'yicha ko'riladigan tadbirlarda transport tizimining tegishli ko'rsatkichlari bo'yicha yangilanib boriladigan raqamli bazani takomillashtirish, transport tizimining matematik modellarini ishlab chiqish, shahar yo'lovchi tashish transport tizimini qisqa va o'rta muddatli rejalashtirish tizimini joriy etishda sun'iy intellekt mexanizmlaridan foydalanishni amaliyotga tadbiriq etishni nazarda tutish alohida ahamiyatga ega.

Sun'iy intellekt texnologiyalaridan foydalanmasdan shahar yo'lovchilar transportini rejalashtirishni raqamlashtirish uchun yuqorida keltirilgan platformalarning samarali ishlashini ta'minlash mumkin emas. Bu ehtiyoj

jamiyat taraqqiyotining zamonaviy bosqichi ichki va tashqi muhitdagi o'zgarishlarning dinamikasi, o'rganilayotgan jarayonlar to'g'risidagi ma'lumotlar miqdori va turlarining ko'payishi bilan tavsiflanganligi bilan bog'liq. Bundan tashqari, bunday ma'lumotlar lingvistik shaklda sifatli baholash tufayli noaniqliklarni (to'liqsizligi, statistik mohiyati, noaniqligi, nomuvofiqligi va noravshanligi) o'z ichiga oladi. Ushbu turdagi ma'lumotlar nomukammal (to'liqsiz) ma'lumotlar toifasiga kiradi. Bu intellektual muammolar sinfiga kiruvchi, aniq algoritmlar va mavjud an'anaviy usullar bilan hal qilishning iloji bo'lmagan, tuzilmaviy bo'lmagan muammolar deb nomlanishiga olib keldi.

Ushbu muammolarni hal qilish uchun sun'iy intellekt nazariyasiga asoslangan yangi aqlli, bilimga yo'naltirilgan (kognitiv) algoritmlar va axborot texnologiyalaridan foydalanish zarur.

Shunday qilib, Toshkent shahar jamoat transporti yo'lovchi tashish yo'nalishlari tizimini kompleks rejalashtirish va transport tarkibini ratsional miqdorini aniqlash maqsadida transport tizimining matematik modeli, algoritmlari va dasturiy majmuasini ishlab chiqish, sun'iy intellekt mexanizmi tarkibiga kiruvchi blokcheyn texnologiyasidan foydalanishni amaliyotga tadbiq etish bilan bog'liq holda taqdim etilayotgan loyiha doirasida hal qilinadigan muammo dolzarbdir.

Kelajakdagi shahar harakatlanish tizimini muvaffaqiyatli qurish, rejalashtirish va ekspluatatsiya qilish uchun jamoat transporti sohasi eng yangi texnologik yutuqlardan xabardor bo'lib turishi shart. So'nggi yillardagi eng istiqbolli ishlanmalardan biri bu sun'iy intellektni (SI) amalga oshirishdir.

Sun'iy intellekt barcha sohalarda ishlash uslubimizni o'zgartirdi va hatto kundalik hayotga ta'sir qila boshladi. Ushbu texnologiyalar jamoat transporti sohasiga olib keladigan imkoniyatlar turlichadir: mijozlarga xizmat ko'rsatishdan tortib, yanada sifatli xizmat ko'rsatishga qadar. Sun'iy intellekt allaqachon transport xizmatining kelajagini o'zgartira boshladi.

Hozirgi bosqichda raqamli texnologiyalar odatda ma'lumotlarni taqdim etish, saqlash va qayta ishlash uchun ishlatiladi. Shuning uchun shakllangan jamiyat raqamli deb nomlandi.

Raqamli jamiyatda asosiy manba shunchaki axborot emas, balki katta hajmdagi ma'lumotlarni, shu jumladan nomukammal ma'lumotlarni qayta ishlash natijasida olingan ma'lumotlar va bilimlarni o'z ichiga olgan axborot

resursidir. Shuning uchun bunday bilimlarning asosiy qismi lingvistik ko'rinishda shakllantirilgan sust shakllangan, notuzilmaviy xususiyatga ega. Bu intellektual bilimga asoslangan axborot texnologiyalarini (AT) takomillashtirish, rivojlantirish va jamiyatning barcha sohalarida keng qo'llash bo'yicha tadqiqotlarning dolzarbligiga olib keldi.

Ma'lumotlarning murakkabligi, katta hajmi va nomukammalligi (baholarning noaniqligi, noravshanligi, nomuvofiqligi, noaniqligi), shuningdek atrof-muhitda noaniqliklar mavjudligi quyidagi xususiyatlar bilan tavsiflangan muammolarning paydo bo'lishiga olib keldi: qiyin shakllantirish va ularni yechish uchun deterministik algoritmlarning yo'qligi; dastlabki ma'lumotlarda va vaziyatlarda har xil noaniqliklar (noaniqlik, noravshanlik, qisman haqiqat va boshqalar) mavjudligi; statistik ma'lumotni olish imkoniyati yo'qligi; o'rganilayotgan obyektlar va jarayonlarning parametrlari o'rtasidagi noma'lum tabiat va munosabatlar turi; parametrik va tuzilmaviy-funksional xususiyatlarning o'ta katta o'lchovi va noaniqligi.

Bunday vazifalar intellektual (sust shakllangan) sinfga tegishli bo'lib, ularni qat'iy algoritmlar va mavjud an'anaviy IT-vositalar yordamida hal qilish mumkin emas. Bu yangi intellektual axborot texnologiyalarini (Int AT), shuningdek ularni sun'iy intellekt (SI) tushunchalari asosida amalga oshiradigan intellektual axborot tizimlarining (Int AT) funksional quyi tizimlarini algoritmik, dasturiy ta'minot va texnik qo'llab-quvvatlashning rivojlanishini faollashtirishga olib keldi.

O'tgan yillar davomida transport ta'minot zanjirini boshqarishda sun'iy intellektdan foydalanish butun dunyoda ma'lumotlarning shaffofligi va kuzatilishi mumkinligi ortganligi, shuningdek mijozlarga xizmat ko'rsatishni yaxshilash zarurati tufayli sezilarli darajada o'sdi. Ta'minot zanjirida sun'iy intellektni qabul qilish bo'yicha yetakchi tarmoqlar (2020- yil boshidagi ma'lumotlarga ko'ra) telekommunikatsiya (26%), yuqori texnologiyalar (23%), sog'liqni saqlash (21%), professional xizmatlar (19%), shuningdek sayohat, transport va logistika (18%).

SI ni integrasiyalashtirishning afzalliklariga qaramay, ba'zi tashkilotlar quyidagi muammolar tufayli uni amalga oshira olmaydilar:

Yuqori sifatli, izchil va yangilangan (real vaqtda) ma'lumotlarning cheklanganligi.

Turli bo'limlarda ta'minot zanjiri ma'lumotlarining mavjudligi (masalan,

marketing bo'limi, inventarizatsiya bo'limi, sotib olish bo'yicha menejer va boshqalar o'zlarining ma'lumotlar bazalariga ega).

Ma'lumotlarga kirish, tozalash va tahlil qilish uchun tizimlar va ma'lumotlar bazalari o'rtasida cheklangan integrasiya.

Kengaytirilgan ta'minot zanjiri bilan bog'liq cheklangan ma'lumotlarni boshqarish siyosati.

Xarid qilish bo'yicha mutaxassislar, COVID -19 pandemiyasi sababli yetkazib berish zanjiridagi so'nggi uzilishlar operatsiyalarni har qachongidan ham ko'proq optimallashtirish uchun SI ni ta'minot zanjiriga qo'shish zarurligini ta'kidlamogda. Ta'minot zanjirining muhim buzilishining oldini olish uchun tashkilot butun ekotizim haqida to'liq tasavvurga ega bo'lishi kerak; talab va taklifni aniq bashorat qilish; va boshqa narsalar qatorida logistika va yetkazib berishni maqbul rejalashtirish. SI, mashinali o'qitish bilan bir qatorda, tashkilotlarga yetkazib berishdan foydalanish muammolarni aniq taxmin qilish va shunga muvofiq oldindan (ogohlantirish tuzatish) choralarni ko'rish imkoniyatini beradi.

SI yechimlari yordamida siz qaror qabul qiluvchilar uchun juda foydali bo'lgan ish haqida kengroq va chuqurroq tushunchalar berish uchun bir nechta ulangan qurilmalar va tizimlardan (shu jumladan SCM, ERP va CRM tizimlaridan) real vaqt va tarixiy ma'lumotlarni to'plashingiz va tahlil qilishingiz mumkin. Ushbu yechimlardan foydalangan holda, ta'minot guruhi ta'minot zanjiri haqida ma'lumotga ega bo'lishi mumkin, muammolarni (tashkilot ichida bo'ladimi, masalan, uzilishlar sababli yoki tashqi, masalan, yetkazib berishning kechikishi kabi) va yetkazib berish zanjiriga ta'sirini minimallashtirish uchun muqobil choralarni ko'rish mumkin. Haqiqiy javobning kechikishi ta'minot zanjirlariga salbiy ta'sir qiladi va o'z navbatida sof foydaga.

SI yechimlari tashkilotlarga real vaqt rejimida bir nechta turli xil pudratchilar, mijozlar va ichki funktsiyalar (shu jumladan yetkazib beruvchilar, mijozlar, inventarizatsiya va mahsulotlar) dan ma'lumotlarni to'plash va aniq bashorat qilish uchun foydalanish imkoniyatini beradi. An'anaga ko'ra, bashorat qilish real vaqtda tafsilotlarni o'z ichiga olmaydi va faqat tarixiy ma'lumotlarga asoslanadi. Biroq, sun'iy intellektdan foydalangan holda prognozning aniqligi sezilarli darajada yaxshilandi, bu rahbarlarga nafaqat yaxshi rejalashtirish, balki samaradorlikni oshirishga ham imkon beradi.

Bundan tashqari, yuqori darajadagi qarorlarni qabul qilishni avtomatlashtirish uchun sun'iy intellektdan foydalanish menejrlarning quyi darajadagi strategiya va qarorlar qabul qilishga e'tibor qaratishlari uchun o'tkazuvchanlik qobiliyatini bo'shatishi mumkin.

SI vositalari va yechimlari ulkan ma'lumotlar to'plamlarini real vaqt rejimida tahlil qilishga, talab va taklifdagi bo'shliqlarni muvozanatlashga, ishlab chiqarishni samarali rejalashtirishga, ishlab chiqarish faoliyatini samarali rejalashtirishga va ta'minot zanjiri rejaları va strategiyalarini ishlab chiqishga yordam beradi. Sun'iy intellekt bozorni ehtiyojlarini to'g'ri baholashda va ishlab chiqarishni boshqarishda ortiqcha ishlab chiqarish yoki mahsulot yetishmovchiligini oldini olish uchun yordam berishi mumkin, bu esa yo'qotishlarga olib kelishi mumkin.

SI yechimlari turli xil ma'lumotlar to'plamini tahlil qilish (yetkazib berish ko'rsatkichlari, auditorlik tekshiruvlari, reytinglar va kredit ballari kabi) va yetkazib beruvchilar bilan munosabatlarni boshqarish bo'yicha moslashtirilgan maslahatlar berish uchun ishlatilishi mumkin. O'zaro manfaatli aloqalarni o'rnatish uchun potentsial yoki mavjud yetkazib beruvchilar to'g'risida dolzarb va muntazam ma'lumotlardan foydalanish mumkin.

SI yechimlari qaror qabul qiluvchilarga mavjud yo'nalishlarni tahlil qilish, to'siqlarni aniqlash va eng yaxshi yo'nalishga e'tibor berish imkoniyatini beradi; bu vaqtni ham, umumiy jamlanma va transport xarajatlarini ham kamaytiradi. Ma'lumotlarni qayta ishlashni sun'iy intellekt va mashinasozlik vositalari real vaqtda harakatlanuvchi tovarlarning tafsilotlarini olish va yetkazib berish vaqtlarini to'g'ri baholashga yordam beradi.

Sun'iy intellekt barcha sohalarda ishlash uslubimizni o'zgartirdi va hatto kundalik hayotga ta'sir qila boshladi. Ushbu texnologiyalar jamoat transporti sohasiga olib keladigan imkoniyatlar mijozlarga xizmat ko'rsatishdan tortib, yanada sifatli xizmat ko'rsatishga qadar. Sun'iy intellekt allaqachon transport xizmatining kelajagini o'zgartira boshladi.

Osiyo-Tinch okeani seksiyasining Texnik jihatdan mukammalligi MSOT, Singapurining yer usti transporti agentligi (LTA) bilan hamkorlikda o'tgan yil davomida "Ommaviy jamoat transportida sun'iy intellekt" bo'yicha tadqiqotlar olib bordi. Loyihaning maqsadi - sun'iy intellektdan foydalanishning dastlabki natijalarini har tomonlama tahlil qilish, tajriba to'plash va jamoat transporti sektorining istiqbollari to'g'risida xabardorligini

oshirish edi.

Loyiha davomida mutaxassis maslahatchilar tomonidan yozilgan bir qator muammoli blog yozuvlaridan tashqari, ushbu tadqiqot natijalari jamoat transportida sun'iy intellektning dolzarb misollari va jamoat transporti tizimlarida sun'iy intellektdan foydalanish istiqbollari taqdim etilgan maqolada umumlashtirildi.

Ushbu tadqiqot loyihasi doirasida o'tkazilgan onlayn so'rov natijalariga ko'ra, so'rovda qatnashgan jamoat transporti tashkilotlarining 62 foizi sun'iy intellekt texnologiyalari sohasidagi loyihalar va yechimlar bilan shug'ullanmoqdalar.

2018-yil iyul oyida bo'lib o'tgan Singapur Xalqaro transport kongressi va ko'rgazmasida (SITCE) MSOT a'zosi Sharqiy Yaponiya temir yo'l kompaniyasi (JR East) o'zining sayohatchilarga stantsiyalarda savollariga javob berishda yordam beradigan mijozlarga xizmat ko'rsatuvchi roboti Emiewni namoyish qildi.

Xizmat ko'rsatish sifatini oshirish uchun NEC Gonkong kabi soha o'yinchilari yo'lovchilar soni va yo'nalishdagi kunlik sayohatlarning umumiy soni kabi ma'lumotlarga asosanib real vaqt rejimida avtobus yo'nalishlari va jadvalarini optimallashtiradigan avtobuslar jadvalini tuzish tizimini o'rnatdilar.

SI tadqiqotlari va innovasiyalari jamoat transportida keskin burilishdir. Evropa Ittifoqi tomonidan moliyalashtiriladigan My-Trac (My TRAVel Companion) loyihasi yangi transport dasturini ishlab chiqishga qaratilgan bo'lib, u sun'iy intellekt algoritmlari va xatti-harakatlarning transport analitikasi kabi texnologiyalar yordamida yo'lovchilarga xizmat ko'rsatishni yaxshilaydi. MSOT ushbu loyihada yana sakkizta sherik bilan hamkorlik qilmoqda.

Ma'lum bo'lgan ishlarda, modellashtirish va yarim tuzilgan qarorlarni qabul qilish muammolarini hal qilish, aksariyat hollarda noravshan to'plamlar nazariyasi asosida amalga oshiriladi. Ko'p agentli taqsimlangan intellektual tizimlarning alternativ konsepsiyasi agentlar o'rtasidagi o'zaro ta'sir va raqobat bilan klassik yondashuvdan quyidagicha farq qiladi: har bir intellektual agent to'liq avtonom harakat qiladi; har bir intellektual agent umumiy muammoga yechim taklif qiladi (nafaqat o'zi), har bir agent mavjud bo'lgan barcha ma'lumotlarga to'liq kirish huquqiga ega; muammoning umumiy yechimi

raqobat asosida parallel ishlaydigan agentlardan birining taklifi sifatida tavsiflanadi (agentlarning shaxsiy yechimlarini muvofiqlashtirish va integrasiyalash yo'li bilan emas, ko'pincha takroriy rejimda amalga oshiriladi); agentlarning hamkorligi butun tizimning zaruriy xatti-harakatlarini shakllantiradi; tizimdagi hamkorlik va raqobat bilan bog'liq harakatlar bir vaqtning o'zida (ketma-ket emas) amalga oshiriladi. Shu bilan birga, sun'iy intellekt tizimlarini yaratish uchun evolyusion algoritmlardan foydalangan holda, yumshoq modellarni, shu jumladan noravshan-neyronlarni sintez qilishning istiqbolli masalalari yetarlicha o'rganilmagan.

Qo'yilgan maqsadga erishish uchun quyidagi tadqiqot vazifalarini hal qilish zarur:

1. Sun'iy intellekt, murakkab tuzilmaviy axborotni qayta ishlash tizimlarini yaratish bo'yicha xulosalarni tahlil qilish, umumlashtirish va tizimlashtirish hamda shahar yo'lovchi transportini rejalashtirish intellektual tizimlarini ishlab chiqish usul va yondoshuvlarini qurish tamoyillarini shakllantirish.

2. Harakatlanadigan tarkibning ratsional miqdorini aniqlash maqsadida yo'nalishlar bo'yicha yo'lovchilar oqimini o'rganib chiqish.

3. Shahar yo'lovchi transport tarkibni ratsional miqdorini aniqlash maqsadida yo'nalishlar bo'yicha yo'lovchilar oqimini optimal baholash dasturini ishlab chiqish.

4. Sun'iy intellekt mexanizmlaridan foydalangan holda transport oqimini boshqarish dasturini ishlab chikish.

5. Shahar yo'lovchi transportida tashish xavfsizligi va sifatini oshirish dasturini ishlab chiqish.

6. Transport logistikasining o'ziga xos xususiyatlari, infratuzilmasi tamoyillari va rivojlantirish omillarini aniqlash, jumladan, jamiyat faoliyatidagi ahamiyati, o'rni va istiqbolini belgilash.

7. Moddiy oqimlar harakatini amalga oshirish jarayonlarini makon va vaqtda modellashtirish asosida, ularni tizimli tahlillash va bashoratlash uslubiyatini ishlab chiqish.

8. Logistik tizimlarning, ekspluatatsion faoliyatini tashkil etish va integral xususiyatlarini inobatga olish orqali texnik-texnologik elementlarini rivojlantirish konsepsiyasini shakllantirish. Atrof-muhitning noaniqliligi va o'zgaruvchanligi sharoitlarida shahar yo'lovchi transportini qisqa va muddatli

rejalashtirish, qaror qabul qilish, bashoratlash, tahlil qilish va modellashtirish uchun intellektual axborot texnologiyalarining algoritm, usul, model va dasturlarini ishlab chiqish.

9. Avtomobil transporti tizimi faoliyatining va an'anaviy tashkiliy tuzilmalarni qayta tashkil etishning modernizatsiyalash sharoitlarga yo'naltirilgan konsepsiyasini ishlab chiqish. Ma'lumotlarni intellektual tahlil qilish (Data Mining), axborot manbalardan bilimlarni qidirish va olish (Knowledge Discovery in Databases) gibrid algoritmlari va dasturiy vositalari ishlab chiqish.

10. Avtomobil transporti tizimini rivojlantirish uchun innovatsion loyihalarni shakllantirish, rejalashtirish va boshqarish algoritmlari va usullarini ishlab chiqish. Axborot, shuningdek noformal, tuzilmagan bilimlarni qayta ishlashga asoslangan, raqamli iqtisodiyotning bilimlariga (Digital Knowledge Economy) yo'naltirilgan instrumental vositalar (algoritmlar va dasturlar) ishlab chiqish.

11. Avtobus yo'nalishlari harakatini yo'lovchi transportining boshqa turlari, jumladan, Toshkent metropolitenining amaldagi va yangidan barpo etilayotgan liniyalari bilan o'zaro uzviy bog'liqlikda yanada optimallashtirishda sun'iy intellekt mexanizmlaridan foydalanish. Yumshoq hisoblashlar (Soft Computing), chuqur neyron to'rlar, chuqur o'qitish (Deep Learning), optimallashtirishning genetik va evolyusion algoritmlari (Swarm intelligence – gala intellekti: chumoli, asalari) mos texnologiyalarini amalga oshiruvchi dasturiy ilovalar ishlab chiqish.

12. Avtobus yo'nalishlari harakatini yo'lovchi transportining boshqa turlari, jumladan, Toshkent metropolitenining amaldagi va yangidan barpo etilayotgan liniyalari bilan o'zaro uzviy bog'liqlikda yanada optimallashtirish dasturini ishlab chiqish; blokcheyn texnologiyasi asosida shahar yo'lovchi transportida innovatsion talabni boshqarish tamoyillari va innovatsiyalar samaradorligini baholash uslubini ishlab chiqish.

7. KLASTERLI TAHLIL. OLAP VA BOSHQA MA'LUMOT SAQLAGICHLAR

Klasterlash usullari klasterlar soni avvaldan berilganligi yoki yo'qligi bo'yicha ham tasniflanadi. Klasterlar soni avvaldan ma'lum bo'lmagan hollarda klasterlar soni algoritmnining bajarilish davomida berilgan ma'lumotlarni taqsimlash asosida aniqlanadi.

Klasterlash – bu qandaydir to'plam elementlarini ularning o'xshashliklari asosida guruhlarga ajratishdir. Klasterlash masalasi X dan olingan obyektlarni boshqa klasterdan olingan obyektlar bilan solishtirilgandagiga qaraganda o'zaro o'xshashroq bo'lgan bir nechta qism-to'plamlarga (klasterlarga) ajratishdan iborat. Metrik fazoda "o'xshashlik" odatda masofa bilan o'lchanadi.

Klasterlash algoritmlari obyektlar ustida amal bajaradi. X ning har bir obyektiga $X = (x_1, \dots, x_d)$ belgilar vektori mos qo'yiladi.

$x_i, i = 1, \dots, d$ komponentlar obyektning alohida belgilari hisoblanadi. Belgilar soni d belgilar fazosining o'lchamini belgilab beradi.

Belgilarning barcha vektorlaridan tashkil topgan to'plam $M = (X_1, \dots, X_n)$ bilan belgilanadi, bu yerda $X_i = (x_{i1}, \dots, x_{id})$.

Klaster M dan olingan o'zaro "bir-birlariga yaqin bo'lgan" obyektlarning qism-to'plamidir. X_i va X_j obyektlar o'rtasidagi $D(X_i, X_j)$ masofa tanlangan metrika asosida belgilar fazosida aniqlanadi.

Ravshan (kesishmaydigan) klasterlash – M dan olingan har bir X_i faqat bitta klasterga kiritiladigan klasterlashdir.

Klasterlash natijalarini tahlil qilishda foydalanilgan algoritmlarning o'ziga xos xususiyatlarini hisobga olish lozim.

Klasterlashning ayrim algoritmlarida obyektни navbatdagi klasterga kiritishda klasterlarning umumlashgan belgilaridan foydalaniladi. Bunday usullarga k -o'rtachalar iterativ algoritmi misol bo'ladi. k -o'rtachalar algoritmidagi klaster markazi tushunchasi kiritiladi. Obyekt va klaster o'rtasidagi masofa deganda obyekt va klaster markazi o'rtasidagi masofa tushuniladi. Sinflashtirilayotgan obyekt ungacha bo'lgan masofa eng kichik bo'lgan klasterga kiritiladi. Odatda masofa deganda Evklid masofasi tushuniladi, ya'ni obyektlar Evklid fazosidagi nuqtalar sifatida qaraladi.

Oddiy k -o'rtachalar algoritmiga nisbatan asosiy o'zgarish hisoblash

formulasiga masshtablanuvchi A matritsa vektorlari o'rtasidagi masofaning kiritilishidan iborat. Masshtablanuvchi sifatida odatda simmetrik bo'lgan musbat aniqlangan matritsa, ya'ni xos sonlari haqiqiy va musbat bo'lgan matritsa qo'llaniladi.

Gustafson-Kessel (GK) algoritmi har bir klaster uchun moslashuvchan normadan foydalanadi, ya'ni har bir j-klaster uchun o'zining norma-yaratuvchi matritsasi mavjud. Ushbu algoritmda klasterlashda nafaqat klaster markazlarining koordinatalari va noravshan bo'laklarga bo'lish matritsasi, balki barcha klasterlar uchun norma-yaratuvchi matritsalar ham optimallashtiriladi. Bu turli geometrik shakllarga ega bo'lgan klasterlarni ajratishga imkon beradi.

GK – ellips shaklga ega bo'lgan klasterlarni topishga imkon beruvchi oddiy noravshan klasterlash algoritmidir. k – o'rtachalar algoritmi bilan birgalikda undan ko'pincha klasterlashning boshqa noravshan algoritmlarini dastlabki ishga tushirishda foydalaniladi.

Noravshan k-o'rtachalar asosiy algoritmi va Gustafson-Kessel algoritmining qiyosiy tahlili quyidagi 7.1-jadvalda keltirilgan.

7.1-jadval.

Noravshan k-o'rtachalar asosiy algoritmi	Gustafson-Kessel algoritmi
<p>1-qadam. Algoritm parametrlarini kiritish: s – klasterlar soni; m – eksponensial vazn; ε - algoritmni to'xtatish parametri.</p> <p>2-qadam. Noravshan bo'laklarga bo'lish matritsasini tasodifiy asosda tarzda hosil qilish.</p>	
<p>3-qadam. Klasterlar markazini hisoblash:</p> $c_j = \frac{\sum_{i=1}^N (u_{ij})^m \cdot x_i}{\sum_{i=1}^N (u_{ij})^m}.$	

<p>4-qadam. X dan olingan obyektlar va klasterlar markazlari o'rtasidagi masofani hisoblash:</p> $D_{ij} = \sqrt{\ x_i - c_j\ ^2}.$	<p>4-qadam. j –matritsa uchun kovariatsiya matritsasini hisoblash:</p> $A_j = \frac{\sum_{i=1}^N (u_{ij})^m \cdot (x_i - c_j)^T \cdot (x_i - c_j)}{\sum_{i=1}^N (u_{ij})^m}.$
<p>5-qadam. Noravshan bo'laklarga bo'lish matritsasi elementlarini qaytadan sanab chiqish:</p> <p>Agar $D_{ij} > 0$ bo'lsa</p> $u_{ij} = \frac{1}{\left(D_{ij}^2 \sum_{k=1}^K \frac{1}{D_{ik}^2}\right)^{\frac{1}{m-1}}}$ <p>agar: $D_{ij} = 0$ bo'lsa</p> $u_{ik} = \begin{cases} 1, & k = j \\ 0, & k \neq j \end{cases} \quad k = \overline{1, K}.$	<p>5-qadam. X dan olingan obyektlar va klasterlar markazlari o'rtasidagi masofani hisoblash:</p> $D_{A_j} = (x_i - c_j) \cdot \left[(\det(A_j))^{-\frac{1}{N}} \cdot A_j^{-1} \right] \cdot (x_i -$
<p>6-qadam. $\ U - U^*\ ^2 < \varepsilon$, shartni tekshirish, bu yerda U^* - algoritmnining avvalgi iteratsiyasidagi noravshan bo'laklarga bo'lish matritsasi. Agar "ha" bo'lsa, 7-qadamga, aks holda 3-qadamga o'tilsin.</p>	<p>6-qadam. Noravshan bo'laklarga bo'lish matritsasi elementlarini qaytadan sanab chiqish:</p> <p>agar $D_{ij} > 0$ bo'lsa</p> $u_{ij} = \frac{1}{\left(D_{ij}^2 \sum_{k=1}^K \frac{1}{D_{ik}^2}\right)^{\frac{1}{m-1}}}$ <p>agar: $D_{ij} = 0$ bo'lsa</p> $u_{ik} = \begin{cases} 1, & k = j \\ 0, & k \neq j \end{cases} \quad k = \overline{1, K}.$
<p>7-qadam. Tamom.</p>	<p>7-qadam. $\ U - U^*\ ^2 < \varepsilon$, shartni tekshirish, bu yerda U^* - algoritmnining avvalgi iteratsiyasidagi noravshan bo'laklarga bo'lish matritsasi. Agar "ha" bo'lsa, 8-qadamga, aks holda 3-qadamga o'tilsin.</p>
	<p>8-qadam. Tamom.</p>

Tog'li klasterlash usuli 1993- yili R.Yaner va D.Filevlar tomonidan taklif qilingan. Tog'li usul bo'yicha klasterlash noravshan hisoblanmaydi, biroq undan ko'pincha ma'lumotlardan noravshan qoidalarni sintez qilishda foydalaniladi. Ushbu usulning o'ziga xos xususiyati algoritmi ish boshlagunicha klasterlar sonini berishga ehtiyojning yo'qligi hisoblanadi.

Tog'li klasterlashning birinchi qadamida klasterlarning markazlari bo'lishi mumkin bo'lgan nuqtalar aniqlanadi. Ikkinchi qadamda har bir nuqta uchun uning atrofida klasterni shakllantirish mumkinligini ko'rsatuvchi ehtimollik hisoblanadi. Klasterning ehtimoliy markazi atrofida obyektlar qanchalik zich joylashgan bo'lsa, uning ehtimolligi shunchalik yuqori bo'ladi. Shundan keyin eng yuqori darajadagi ehtimollikka ega bo'lgan nuqtalar orasidan klasterlarning markazlari iteratsiya asosida tanlab olinadi.

Tayanch vektorlar usuli (SVM - support vector machines) - bu sinflashtirish va regression tahlil masalalari uchun foydalaniladigan "o'qituvchi bilan o'qitish" ko'rinishidagi o'xshash algoritmlar to'plami. Bu usul chiziqli sinflashtiruvchilar oilasiga tegishli. U yana A.N.Tixonov bo'yicha sinflarizatsiyaning maxsus holati sifatida ham qaralishi mumkin. Tayanch vektorlar usulining o'ziga xos xususiyati sinflashtirish empirik xatoligining uzluksiz kamayishi va tanish oralig'ining ortishi hisoblanadi. Shuning uchun ushbu usul yana eng katta oraliqqa ega bo'lgan sinflashtirish usuli sifatida ham mashhur.

Tayanch vektorlar usulining asosiy g'oyasi — berilgan vektorlarni yuqoriroq o'lchamga ega bo'lgan fazoga o'tkazish va ushbu fazoda eng katta tirqishga ega bo'lgan ajratuvchi gipertekislikni qidirishdan iborat. Ikkita parallel gipertekisliklar sinflarni ajratuvchi gipertekislikning har ikki tarafiga quriladi. Ajratuvchi gipertekislik bo'lib ikki parallel gipertekislikkacha bo'lgan masofani maksimalashtiruvchi gipertekislik hisoblanadi. Algoritmi ushbu parallel gipertekisliklar o'rtasidagi farq yoki masofa qanchalik katta bo'lsa, sinflashtiruvchining o'rtacha xatosi shunchalik kichik bo'ladi, degan farazga asoslangan holda ishlaydi.

1963-yili Vladimir Vapnik tomonidan taklif qilingan optimal ajratuvchi gipertekislikni qurish algoritmi — chiziqli sinflashtirish algoritmi hisoblanadi. Biroq 1992-yili Bernxard Boser, Izabel Giyon va Vapniklar nochiqli sinflashtiruvchini qurish usulini taklif qildilar, uning asosida skalyar ko'paytirishlardan nochiqli ajratuvchilarni qurishga imkon beruvchi kernel

trick (ilk marta M.A.Ayzerman, E.M.Broverman va L.V.Rozonoerlar tomonidan potensial funksiyalar usuli uchun taklif qilingan) deb ataluvchi ixtiyoriy yadrolarga o'tish yotadi. Natijaviy algoritmi chiziqli sinflashtirish algoritmiga juda o'xshab ketadi, ular o'rtasidagi farq faqatgina yuqorida keltirilgan formulalarda skalyar ko'paytma yadroning nochiqli funksiyasi (katta o'lchamli fazodagi skalyar ko'paytma) bilan almashtirilishidan iborat. Ushbu fazoda allaqachon optimal ajratuvchi gipertekislik mavjud bo'lishi mumkin. Olinayotgan fazoning o'lchami berilgan fazo o'lchamidan katta bo'lishi mumkin, shuning uchun skalyar ko'paytmalarni solishtiruvchi almashtirish nochiqli bo'ladi, demak berilgan fazoda optimal ajratuvchi gipertekislikka mos bo'lgan funksiya ham nochiqli bo'ladi.

K-eng yaqin qo'shnilar (KNN - K-nearest neighbour) algoritmi ma'lumotlarni chuqur tahlil qilish, timsollarni statistik tanib olish kabi sohalardagi ko'pgina tadbirlarda foydalaniladigan fanning bir qismi hisoblanadi.

Eng yaqin qo'shnini qidirish masalasi ko'p o'lchamli metrik fazoda joylashgan elementlar to'plami orasidan ma'lum yaqinlik funksiyasiga ko'ra berilgan elementga yaqin bo'lgan elementlarni qidirib topishdan iborat.

Obyekt o'z qo'shnilarining ko'pchiligi ovozi bilan sinflashtiriladi. "K" doimo musbat butun son bo'ladi. Qo'shnilar to'g'ri sinflashtirilishi ma'lum bo'lgan obyektlar to'plamidan tanlab olinadi.

Algoritmi afzalliklari va kamchiliklari sifatida quyidagilarni keltirish mumkin.

Afzalliklari:

1. Ushbu algoritmi muhim afzalligi bo'lib o'qitishning o'quv tanlanmasini eslab qolishdan iboratligi hisoblanadi.
2. Amalga oshirishning osonligi va sozlash uchun qo'shimcha parametrlarni kiritish imkoniyati.
3. Algoritmi ishlashining pretsedentli mantiqi predmet sohasidagi (tibbiyot, biometriya, huquqshunoslik) ekspertlarga yaxshi darajada tushunarli.

Kamchiliklari:

1. O'quv tanlanmasini butunligicha saqlashga to'g'ri keladi. Xotiraning samarali bo'lmagan usulda ishlatilishi.
2. Timsollarni sinflashtirishdagi operatsiyalar sonining ko'pligi. katta tanlanmalar bilan ishlashning oqibati sifatida an'anaviy NT lariga nisbatan

anchagina ko'proq vaqt talab qilinadi.

Gravitatsiya asosida sinflashtirish (GBC – gravitation based classification) algoritmlarining bir qancha turlari mavjud bo'lib, ularning ba'zilari sinflar markazlarini evristik tarzda tanlab olib shu markazlar o'rtasidagi tortishish kuchidan foydalansa, ba'zilari esa sinflarga kiruvchi har bir obyekt hamda sinflarni ajratib turuvchi "chiziq" nuqtalarining tortishish kuchidan foydalangan holda optimal sinflash masalasini yechishga qaratilgan. Garchi ularning barchasi ham sinflash masalalarini echishga mo'ljallangan bo'lsada, har ikkala turdagi usul ham fizikaviy tortishish kuchi tamoyillaridan foydalanilgan holda yaratilgan.

So'nggi vaqtlarda noravshan algoritmlar keng ommalashib bormoqda, ular orasida "noravshan o'rtachalar" – FCMA (Fuzzy C-Means Algorithm) nomini olgan algoritmlar ayniqsa mashhurdir. Shuni ta'kidlab o'tish kerakki, k - va s - o'rtachalar algoritmlarining eng asosiy kamchiligi klasterlarning talab qilingan soni hamda klasterlash natijalari qiymatlariga jiddiy ravishda bog'liq bo'ladigan boshqa sonli parametrlarni oldindan berilishining zarurligi hisoblanadi. Bundan klasterlash algoritmlaridan foydalanganda u yoki bu parametrlarni qo'llash natijasini sonli baholashga imkon beruvchi obyektlarni klasterlarga bo'linishi sifatining qo'shimcha mezonlariga ega bo'lish zarurligi kelib chiqadi.

Noravshan s -o'rtachalar algoritmining eng muhim xususiyati bo'lib uni qo'llash natijasida noravshan klasterlarning shakllantirilishi hisoblanadi. Bunda obyektning klasterga tegishlilik darajasi tegishlilik funksiyasi yordamida ifodalanadi.

Afsuski, noravshan o'rtachalar algoritmlari klasterlash va sinflashtirishning ko'pgina mashhur dasturiy majmualari tarkibida mavjud emas.

Hozirgi vaqtda SNT dan ko'proq turli tipdagi obyektlarni sinflashtirish va klasterlashda foydalanilayapti. SNT algoritmlarining o'ziga xos xususiyati SNT ni o'qitish jarayonida kompyuterlarning tezligiga va tezkor xotirasiga yuqori talablarning qo'yilishi hisoblanadi.

Turli tipdaligi (miqdoriy va sifatiy) belgilar bilan tavsiflanuvchi obyektlarni sinflashtiruvchi tizimlarni qurishda quyidagi usullar eng ko'p darajada qo'llaniladi:

- binar (ravshan) mantiq usullaridan foydalanuvchi mantiqiy;

- noravshan yoki noaniq mantiq usullaridan foydalanuvchi mantiqiy;
- neyroto'rli.

Ushbu barcha usullarning ma'lum afzalliklari bilan bir qatorda kamchiliklari ham mavjud. Masalan, sinflashtirishning mantiqiy binar algoritmlari har doim ham yaqinlashavermaydigan juda uzun o'qitish jarayoni bilan ajralib turadilar, neyrotarmoqli algoritmlar mavjud apriori ma'lumotlarni hisobga olmaydi va odatda juda katta o'quv tanlanmasini talab qiladi. Noravshan mantiq apparatidan foydalanuvchi algoritmlar parametrlar, ekspert odam tomonidan shakllantiriladigan qo'yilayotgan noravshan qoidalar dastlabki to'plamining to'raligi va o'zaro ziddiyatli emasligi kabilarga jiddiy ravishda bog'liq bo'ladi.

Klasterlashning asosiy algoritmlari, (masalan, Microsoft Analysis Services 2005 da amalga oshirilgan K-Means, Expectation Maximization algoritmlarining modifikatsiyalari) jumladan, har bir klasterni alohida qavariq to'plam bilan qamrab olish imkoniyatini talab qilgan holda olinadigan klasterlarning geometrik shakllariga cheklanishlar qo'yadi. Bunday cheklanish ushbu algoritmlardagi klasterlarning markazlarining mavjudligi (K-Means) yoki har bir klaster uchun matematik kutilma va dispersiyaning tegishli qiymatlariga ega bo'lgan ehtimoliy taqsimot funksiyasining mavjudligi (Expectation Maximization) to'g'risidagi farazlar bilan qo'yiladi. Natijada ushbu algoritmlar klasterlarni mos holda nafaqat qavariq bo'lmagan to'plamlarga bo'lish, xattoki o'zaro ichma-ich joylashgan tuzilmalarga bo'lish imkoniyatiga ham ega bo'lmaydi.

Ushbu muammoni quyida bayon qilinadigan klasterlarga oralarida bir-birlariga "yaqinroq" bo'lgan elementlar ketma-ketliklari mavjud bo'lgan elementlarni guruhlashga imkon beruvchi noravshan munosabatlar asosidagi klasterlash algoritmi hal qiladi, bu esa guruhlash to'g'risidagi intuitiv tasavvurlarga ham mos keladi.

Ushbu maqsadda to'plamlarni noravshan munosabat bo'yicha ekvivalentlik sinflariga bo'lish asosida ixtiyoriy metrik fazoning chekli sondagi elementlari to'plamini klasterlashga bo'lgan yondoshuv bayon qilinadi.

Metrika asosida aniq reflektivlik va oddiy α -simmetriklik xususiyatlariga ega bo'lgan noravshan munosabat aniqlanadi. α ning 0 dan 1 gacha bo'lgan oraliqdagi har bir qiymati uchun berilgan to'plamdagi ekvivalentlik munosabatini aniqlashga imkon beruvchi munosabatlarning

tranzitiv halqasi quriladi. Munosabatlarning qurilishi bo'yicha ikki elementi faqat va faqat ular o'rtasida o'zaro bir-birlariga "yaqin" bo'lgan elementlar ketma-ketligi mavjud bo'lgan hollardagina bitta ekvivalentlik sinfiga kiradi.

Klasterlanayotgan elementlar fazosi induksiyalashgan metrikali topologiyaga ega bo'lgan vektor fazosi bo'lgan hollarda bayon qilingan algoritm yordamida olingan klasterlar ixtiyoriy geometrik shaklga ega bo'lishi, jumladan qabariq bo'lmagan to'plamlardan tashkil topishi mumkin. Bu yondashuvning klasterlashning boshqa ko'pgina mashhur algoritmlariga (masalan, K-Means, Expectation Maximization algoritmlarining modifikatsiyalari va b.) nisbatan jiddiy afzalligi hisoblanadi.

Algoritmning afzalliklari sifatida quyidagilarni ko'rsatish mumkin:

- Ma'lumotlarning tarkibi to'g'risidagi apriori farazlarga (klasterlar bo'yicha ehtimoliy taqsimotning ko'rinishi va parametrlari, zichlik markazlari, klasterlar soni) bo'lgan zaruriyatning yo'qligi.

- Klasterlar bo'yicha bo'lish natijalarining tushunarli talqini: o'zaro bir-birlariga yaqin bo'lgan elementlar ketma-ketligi mavjud bo'lgan hollarda elementlar bitta klasterga kiritiladi.

- Klasterlarning geometrik shakllariga bo'lgan cheklanishlarning mavjud emasligi.

- Algoritmning bajarilish vaqti kiruvchi vektorlardagi komponentlar soniga kam darajada bog'liqligi.

Algoritm o'quv tanlanmasini aniqlashdan boshlanadi:

$$\begin{array}{c|c} \begin{array}{ccc} x_1^1 & x_2^1 & \dots & x_n^1 \\ x_1^2 & x_2^2 & \dots & x_n^2 \\ \cdot & \cdot & \cdot & \cdot \\ x_1^m & x_2^m & \dots & x_n^m \end{array} & \begin{array}{c} y_1 \\ y_2 \\ \dots \\ y_m \end{array} \end{array}$$

Keyingi qadamda o'quv tanlanmasini normallashtirish amalga oshiriladi:

$$u_i^k = l \frac{x_i^k - x^{\min}}{x^{\max} - x^{\min}}; \quad i = \overline{1, n}, \quad k = \overline{1, m}.$$

Faziffikatsiyalash operatori bajariladi:

$$\mu^j(u_i^k) = \frac{1}{1 + \frac{u_i^k - c_j}{\sigma_j}}; \quad \sigma_j, c_j - \text{parametrlar,}$$

$j = \overline{1, l}$. l - termlar soni.

Sinflarni normallashtirish amalga oshiriladi:

$$v^k = l \frac{y^k - y^{\min}}{y^{\max} - y^{\min}}$$

Faziffikatsiyalash operatori bajariladi:

$$\mu^j(v^k) = \frac{1}{1 + \frac{v^k - c_j}{\sigma_j}};$$

$$\mu^*(u_i^k) = \max_j \mu^j(u_i^k);$$

$$\mu^*(v^k) = \max_j \mu^j(v^k);$$

$$SP_k = \prod_{j=1}^n \mu^*(u_j^k) \cdot \mu^*(v^k)$$

Noravshan qoidalarni tanlash amalga oshiriladi:

SP^k ni normallashtirish:

$$\eta^k = l \frac{SP_k - SP^{\min}}{SP^{\max} - SP^{\min}}.$$

Faziffikatsiya:

$$\mu^j(\eta^k) = \frac{1}{1 + \frac{\eta^k - c_j}{\sigma_j}};$$

$$\mu^*(\eta^k) = \max_j \mu^j(\eta^k).$$

Aytaylik, X - metrik fazo va $d: X \rightarrow R$ unda aniqlangan metrika, $(SP_1, \dots, SP_n) \subset X - X$ dan olingan elementlar ketma-ketligi bo'lsin.

Keyin biz quyidagicha faraz qilamiz

$$\forall i \in \{1, \dots, n\} \exists j \in \{1, \dots, n\} : SP_i \neq SP_j.$$

(7.1)

(7.1) shartdan $\forall i \in \{1, \dots, n\}$ uchun quyidagining o'rinaligi kelib chiqadi:

$$\max\{d(SP_i, SP_k) \mid k \in \{1, \dots, n\}\} > 0.$$

Shunday qilib har bir i indeks uchun ketma-ketlikning j -elementi bilan i -elementi o'rtasidagi o'xshashlik o'lchovini ifodalab beruvchi funksiyani aniqlashimiz mumkin: 1-ta'rif:

$$\xi_i : \{1, \dots, n\} \rightarrow [0, 1],$$

$$\xi_i(j) = 1 - \frac{d(SP_i, SP_j)}{\max\{d(SP_i, SP_k) \mid k \in \{1, \dots, n\}\}}.$$

Har bir i indeks uchun k - va l -elementlarni i -elementga nisbatan o'xshashlik o'lchovini ifodalovchi funksiyani aniqlaymiz: 2-ta'rif:

$$\zeta_i : \{1, \dots, n\}^2 \rightarrow [0, 1],$$

$$\zeta_i(k, l) := 1 - |\xi_i(SP_k) - \xi_i(SP_l)|.$$

Endi ixtiyoriy elementlarining ketma-ketligining boshqa barcha elementlariga nisbatan o'xshashlik o'lchovini ifodalovchi funksiyani aniqlaymiz:

3-ta'rif:

$$\mu : \{1, \dots, n\}^2 \rightarrow [0, 1],$$

$$\mu(i, j) := \min\{\zeta_k(i, j) \mid k \in \{1, \dots, n\}\}$$

$$\mu(i, i) = 1, \forall i \text{ ekanligini ko'rsatamiz.} \quad (7.2)$$

Barcha k va i lar uchun

$$\zeta_k(i, i) = 1 - |\xi_k(SP_i) - \xi_k(SP_i)| = 1,$$

shuning uchun:

$$\mu(i, i) := \min\{\zeta_k(i, i) \mid k \in \{1, \dots, n\}\} = 1.$$

$$\mu(i, j) = \mu(j, i), \forall i, j \text{ ekanligini ko'rsatamiz.} \quad (7.3)$$

$$\zeta_k(i, j) = 1 - |\xi_k(SP_i) - \xi_k(SP_j)| = 1 - |\xi_k(SP_j) - \xi_k(SP_i)| = \zeta_k(j, i),$$

shuning uchun:

$$\mu(i, j) := \min\{\zeta_k(i, j) \mid k \in \{1, \dots, n\}\} = \min\{\zeta_k(j, i) \mid k \in \{1, \dots, n\}\} = \mu(j, i).$$

4-ta'rif:

$k=1, 2, \dots, n$ uchun quyidagi funksiyani rekursiv ravishda aniqlaymiz:

$$\mu^{(k)} : \{1, \dots, n\}^2 \rightarrow [0, 1],$$

$$\begin{cases} \mu^{(1)}(i, j) := \mu(i, j), \\ \mu^{(k)}(i, j) := \max\{\min\{\mu^{(k-1)}(i, s), \mu^{(k-1)}(s, j)\} \mid s \in \{1, \dots, n\}\}. \end{cases} \quad (7.4)$$

$$\mu^{(k)}(i, i) = 1, \forall i, k \text{ ekanligini ko'rsatamiz.} \quad (7.5)$$

k bo'yicha induksiya asosidagi isboti.

$k=1$ bo'lganda tasdiq $\mu^{(1)}$ ta'rif va (7.2) dan kelib chiqadi.

Aytaylik, tasdiq $k-1$ uchun ham o'rinli bo'lsin. U holda (7.4) dan quyidagi kelib chiqadi:

$$1 \geq \mu^{(k)}(i, i) = \max\{\min\{\mu^{(k-1)}(i, s), \mu^{(k-1)}(s, i)\} \mid s \in \{1, \dots, n\}\} \geq \min\{\mu^{(k-1)}(i, i), \mu^{(k-1)}(i, i)\} = \mu^{(k-1)}(i, i) = 1.$$

Tasdiq isbotlandi.

$$\mu^{(k)}(i, j) = \mu^{(k)}(j, i), \forall i, j, k \quad (7.6)$$

ekanligini ko'rsatamiz.

k bo'yicha induksiya asosidagi isboti. $k=1$ bo'lganda tasdiq $\mu^{(1)}$ ta'rifdan va (7.3) dan kelib chiqadi.

Aytaylik, tasdiq $k-1$ uchun ham o'rinli bo'lsin. U holda (7.4) dan quyidagi kelib chiqadi:

$$\begin{aligned} \mu^{(k)}(i, j) &= \max\{\min\{\mu^{(k-1)}(i, s), \mu^{(k-1)}(s, j)\} \mid s \in \{1, \dots, n\}\} = \\ &= \max\{\min\{\mu^{(k-1)}(s, j), \mu^{(k-1)}(i, s)\} \mid s \in \{1, \dots, n\}\} = \\ &= \max\{\min\{\mu^{(k-1)}(j, s), \mu^{(k-1)}(s, i)\} \mid s \in \{1, \dots, n\}\} = \mu^{(k)}(j, i). \end{aligned}$$

Tasdiq isbotlandi.

$$\mu^{(k)}(i, j) \geq \mu^{(k-1)}, \forall k > 2, \quad (7.7)$$

ekanligini ko'rsatamiz.

$$\begin{aligned} \mu^{(k)}(i, j) &= \max\{\min\{\mu^{(k-1)}(i, s), \mu^{(k-1)}(s, j)\} \mid s \in \{1, \dots, n\}\} \geq \\ &\geq \min\left\{\underbrace{\mu^{(k-1)}(i, i)}_{=1}, \underbrace{\mu^{(k-1)}(i, j)}_{\geq \mu^{(k-1)}(i, j)}\right\} \geq \min\left\{1, \underbrace{\mu^{(k-1)}(i, j)}_{\geq \mu^{(k-1)}(i, j)}\right\} = \mu^{(k-1)}(i, j). \end{aligned}$$

Tasdiq isbotlandi.

$\alpha \in [0, 1]$ va $i, j_1, \dots, j_k, j \in \{1, \dots, n\}$ uchun quyidagini o'rinli ekanligini ko'rsatamiz:

$$\mu(i, j_1), \dots, \mu(j_k, j) \geq \alpha \Rightarrow \mu^{(k)}(i, j) \geq \alpha. \quad (7.8)$$

k bo'yicha induksiya asosidagi isboti. $k=1$ bo'lganda tasdiqlash shart emas. Aytaylik, tasdiq $k-1$ uchun ham o'rinli bo'lsin. U holda induksiya to'g'risidagi farazdan kelib chiqadi.

$$\mu^{(k-1)}(i, j_k) \geq \alpha$$

(7.9)

(7.7) dan quyidagi kelib chiqadi:

$$\mu^{(k-1)}(j_k, j) \geq \mu^{(1)}(j_k, j) = \mu(j_k, j) \geq \alpha. \quad (7.10)$$

U holda $\mu^{(k)}(i, j)$ ta'rif, (7.9) va (7.10) dan quyidagi kelib chiqadi:

$$\begin{aligned} \mu^{(k)}(i, j) &= \max\{\min\{\mu^{(k-1)}(i, s), \mu^{(k-1)}(s, j)\} \mid s \in \{1, \dots, n\}\} \geq \\ &\geq \min\left\{\underbrace{\mu^{(k-1)}(i, j_k)}_{\geq \alpha}, \underbrace{\mu^{(k-1)}(j_k, j)}_{\geq \alpha}\right\} \geq \alpha. \end{aligned}$$

Tasdiq isbotlandi.

Olaylik

$$\alpha \in [0, 1], i, j, k \in \{1, \dots, n\} \text{ va } \mu^{(k)}(i, j) \geq \alpha$$

bo'lsin. U holda shunday m soni va $j_1, \dots, j_m \in \{1, \dots, n\}$ indekslari mavjud bo'ladi,

$$\mu(i, j_1), \dots, \mu(j_m, j) \geq \alpha \quad (7.11)$$

o'rinli bo'ladi.

 k bo'yicha induksiya asosidagi isboti. $k = 1$ bo'lganda tasdiq j indeks uchun o'rinli bo'ladi. Aytaylik, tasdiq $k-1$ uchun ham o'rinli bo'lsin. $\mu^{(k)}(i, j)$ ta'rifdan shunday p indeksning mavjudligi kelib chiqadiki,

$$\begin{aligned} \mu^{(k)}(i, j) &= \max\{\min\{\mu^{(k-1)}(i, s), \mu^{(k-1)}(s, j)\} \mid s \in \{1, \dots, n\}\} = \\ &= \min\{\mu^{(k-1)}(i, p), \mu^{(k-1)}(p, j)\}. \end{aligned} \quad (7.12)$$

o'rinli bo'ladi.

(7.12) va tasdiq shartlaridan shu narsa kelib chiqadiki,

$$\alpha \leq \mu^{(k)}(i, j) = \min\{\mu^{(k-1)}(i, p), \mu^{(k-1)}(p, j)\} \Rightarrow \mu^{(k-1)}(i, p) \geq \alpha, \mu^{(k-1)}(p, j) \geq \alpha$$

o'rinli bo'ladi.

 $\mu^{(k-1)}(i, p) \geq \alpha$ sharti va induksiya shartlaridan shunday s_1, \dots, s_{m_1} indekslarining mavjudligi kelib chiqadiki, $\mu(i, s_1), \dots, \mu(s_{m_1}, p) \geq \alpha$ o'rinli bo'ladi.Xuddi shunga o'xshash tarzda $\mu^{(k-1)}(p, j) \geq \alpha$ sharti va induksiya shartlaridan shunday p_1, \dots, p_{m_2} indekslarining mavjudligi kelib chiqadiki, $\mu(p_1, p_1), \dots, \mu(p_{m_2}, j) \geq \alpha$ o'rinli bo'ladi.Ushbu xulosalarni birlashtirgan holda tasdiqning $m := m_1 + m_2$ va j_1, \dots, j_m uchun isbotini olamiz, bu yerda

$$j_u := \begin{cases} s_u, & u \leq m_1, \\ p_{u-m_1}, & m_1 < u \leq m_2. \end{cases}$$

Aytaylik,

$$\alpha \in [0, 1], i, j, m \in \{1, \dots, n\} \text{ va } \mu^{(n)}(i, j) \geq \alpha, \mu^{(n)}(j, m) \geq \alpha \text{ bo'lsin.}$$

U holda

$$\mu^{(n)}(i, m) \geq \alpha. \quad (7.13)$$

ekanligini ko'rsatamiz.

 $\mu^{(n)}(i, j) \geq \alpha, \mu^{(n)}(j, m) \geq \alpha$ shart va (7.11) dan shunday $j_1, \dots, j_p \in \{1, \dots, n\}$ va $j_{p+1}, \dots, j_{p+q} \in \{1, \dots, n\}$ indekslarining mavjudligi kelib chiqadiki,

$$\mu(i, j_1), \dots, \mu(j_p, j) \geq \alpha \text{ i } \mu(i, j_{p+1}), \dots, \mu(j_{p+q}, m) \geq \alpha$$

o'rinli bo'ladi.

 $\{j_1, \dots, j_p, j, j_{p+1}, \dots, j_{p+q}\} \subset \{1, \dots, n\}$ bo'lgani uchun ushbu indekslar to'plamidan quyidagi o'rinli bo'ladigan n tadan ko'p bo'lmagan indekslar $\{u_1, \dots, u_r\} \cup \{j_1, \dots, j_p, j, j_{p+1}, \dots, j_{p+q}\}, r \leq n$ dan tashkil topgan to'plamni (takrorlanuvchi elementlarni olib tashlash yo'li bilan) tanlash mumkinligi kelib chiqadi:

$$\mu(i, u_1), \dots, \mu(u_r, m) \geq \alpha$$

(7.8) dan $\mu^{(r)}(i, m) \geq \alpha$ ekanligi kelib chiqadi. $r \leq n$ bo'lganligi uchun (7.7) dan $\mu^{(n)}(i, m) \geq \mu^{(r)}(i, m) \geq \alpha$ ekanligi kelib chiqadi, shuni isbotlash talab qilingan edi.5-ta'rif: $\alpha \in [0, 1]$ uchun $\{SP_1, \dots, SP_n\}$ to'plamda $R_\alpha \subset \{SP_1, \dots, SP_n\}^2$ binar munosabatni quyidagicha aniqlaymiz:

$$(SP_i, SP_j) \in R_\alpha \Leftrightarrow \mu^{(n)}(i, j) \geq \alpha.$$

 R_α munosabatning ekvivalentlik munosabati ekanligini ko'rsatamiz.1. Refleksivlik. (7.5) dan $\mu^{(n)}(i, i) = 1 \geq \alpha \Rightarrow (SP_i, SP_i) \in R_\alpha$ ekanligi kelib chiqadi.2. Simmetriklik. Aytaylik, $(SP_i, SP_j) \in R_\alpha \Rightarrow \mu^{(n)}(i, j) \geq \alpha$ bo'lsin. (7.6)

dan $\mu^{(n)}(j, i) = \mu^{(n)}(i, j) \geq \alpha \Rightarrow (SP_j, SP_i) \in R_\alpha$ kelib chiqadi.

3. Tranzitivlik. Aytaylik, $(SP_j, SP_i) \in R_\alpha, (SP_i, SP_m) \in R_\alpha \Rightarrow \mu^{(n)}(i, j), \mu^{(n)}(j, m) \geq \alpha$ bo'lsin. (7.13) dan $\mu^{(n)}(i, m) \geq \alpha \Rightarrow (SP_i, SP_m) \in R_\alpha$ ekanligi kelib chiqadi.

Shunday qilib R_α ekvivalentlik munosabati $\{SP_1, \dots, SP_n\}$ to'plamni o'zaro kesishmaydigan ekvivalentlik sinflariga ajratadi. Ikki SP_i, SP_j element faqat va faqat ushbu elementlarning $\mu^{(n)}$ funksiyalari qiymatlari (7.8), (7.11) lar asosida $\mu(i, j)$ funksiya qiymatlari juda katta bo'ladigan $(SP_i, SP_j), (SP_j, SP_i), \dots, (SP_i, SP_i)$ elementlar juftliklari ketma-ketligining mavjudligiga ekvivalent bo'ladigan tarzda katta bo'ladigan hollardagina bitta ekvivalentlik sinfiga kiritiladi. Ta'rif bo'yicha $\mu(i, j)$ har bir juftlik elementlarining bir-birlariga yaqinliklarini bildiradi. Ya'ni, ikki element faqat va faqat ular o'rtasida juft holda bir-birlariga yaqin bo'lgan elementlar ketma-ketligi mavjud bo'lgan hollardagina bitta ekvivalentlik sinfiga kiritiladi.

8. BERILGANLARNI INTELLEKTUAL TAHLILINING ASOSIY USULLARI

Data Mining usullarini sinflashtirish

Data Miningning barcha usullari dastlabki o'qitiluvchi ma'lumotlar bilan ishlash tamoili bo'yicha ikkita katta guruhlariga ajratiladi. Ushbu sinflashtirishda yuqori daraja Data Miningdan keyin ma'lumotlar saqlanganmi yoki keyinchalik foydalanish uchun distillanganmi yo'qmi aniqlanadi.

1. Ma'lumotlardan bevosita foydalanish yoki ma'lumotlarni saqlash.

Ushbu holatda dastlabki ma'lumotlar batafsil tekshirilgan oshkora holatda saqlanadi va istisnolarni tahlil qilishning va/yoki bashoratli modellashtirish bosqichlarida bevosita foydalaniladi. Usullarni ushbu guruh muammosi — ulardan foydalanishda o'ta katta hajmli ma'lumotlar bazasini tahlil qilishning murakkabligi yuzaga kelishi mumkin. Ushbu guruh usullari: klaster tahlil, yaqin qo'shnilar usuli, k-yaqin qo'shnilar usuli, o'xshashliklar bo'yicha mulohaza.

2. Shakllangan qonuniyatlarni yoki shablonlarni distilyasiasini aniqlash va foydalanish.

Shablonlarni distillash texnologiyasi yordamida ma'lumotlarning bitta namunasi dastlabki ma'lumotlardan olinadi va ko'rinishi foydalaniladigan Data Mining usuliga bog'liq ba'zi formal konstruksiyalarga aylantiriladi.

Ushbu jarayon erkin qidiruv bosqichida amalga oshiriladi, usullarning birinchi guruhida bu bosqich umuman yo'q.

Istisnolarni bashoratli modellashtirish va tahlil qilish bosqichlarida erkin qidiruv bosqichlari natijalaridan foydalaniladi, ular ma'lumotlar bazalariga qaraganda sezilarlicha ixchamdir.

Eslatib o'tamiz, ushbu modellarni konstruksiyalari tahlilchi tomonidan izohlanishi mumkin yoki izohlanmasligi mumkin ("qora qutilar" tomonidan).

Ushbu guruh usullari: mantiqiy usullar; vizuallashtirish usullari; kross-tabulyasiya usullari; tenglamalarga asoslangan usullar.

Mantiqiy usullar, yoki mantiqiy induksiya usullari o'z ichiga oladi: noravshan so'rovlar; belgili qoidalar; daraxt yechimlar; genetik algoritmlar.

Ushbu guruh usullari, ehtimol, eng ko'p talqin qilingan bo'lishi mumkin ular topilgan qonuniyatlarni aksariyat hollarda foydalanuvchi nuqtai nazaridan

yetarlicha shaffof tarzda tasvirlaydi. Olingan qoidalar uzluksiz va diskret o'zgaruvchilarni o'z ichiga oladi. Daraxt yechimlar daraxt ildizidan terminal uchigacha bo'lgan yo'l bo'ylab bir qoidani yaratish orqali belgili qoidalar to'plamiga osongina aylantirilishi mumkin. Daraxt yechimlar va qoidalar haqiqatda bir masalani yechishni turli xil usullari bo'lib hisoblanadi va ular imkoniyatlari bilangina farqlanadi. Bundan tashqari qoidalarni amalga oshirish daraxt yechimlarni induksiyasiga ko'ra sekinroq ishlaydigan algoritmlar bilan amalga oshiriladi.

Kross-tabulyasiya usullari: agentlar, bayes (ishonch) to'rlari, kross-jadvalli vizualizatsiya. Oxirgi usul Data Mining xossalaridan biriga – tahliliy tizim qonuniyatlarini mustaqil qidirishga unchalik javob bermaydi. Biroq ma'lumotni kross jadval ko'rinishida taqdim etish Data Miningning asosiy masalasi bo'lgan shablonlarni qidirishni ishlab chiqishni ta'minlaydi, shuning uchun ushbu usulni ham Data Miningning usullaridan biri deb hisoblash mumkin.

Tenglamalar asosidagi usullar. Ushbu guruh usullari matematik ifodalar tenglamalar ko'rinishidagi aniq qonuniyatlarni ifodalaydi. Ular sonli o'zgaruvchilar bilangina ishlashi mumkin, va boshqa tiplar o'zgaruvchilari mos holda kodlashtirilishi kerak bo'ladi. Bu guruh usullarini qo'llashni biroz cheklaydi, shunga qaramay ular turlicha masalalarni, ayniqsa bashoratlash masalalarini hal qilishda keng foydalaniladi. Berilgan guruhning asosiy usullari: statistik usullar va neyron to'rlar. Statistik usullar bashoratlash masalalarini yechish uchun tez-tez qo'llaniladi. Ular orasida ma'lumotlarni statistik tahlil qilish usullari to'plami mavjud, masalan korrelyasion - regression tahlil, bir qator dinamikalar korrelyasiyasi, dinamik qatorlardagi tendensiyalarni, garmonik tahlillarni aniqlash.

Boshqa sinflashtirish Data Miningning barcha ko'p qirrali usullarini ikkita guruhga ajratadi: statistik va kibernetik usullar. Ushbu bo'lish sxemasi matematik modellarni o'qitish uchun turlicha yondashuvlarga asoslangan.

Data miningning statistik usullari.

Ushbu usullar o'zi bilan quyidagi to'rtta o'zaro bog'liq bo'limlarni ifodalaydi:

- statistik ma'lumotlar tabiatini dastlabki tahlili (statsionarlik, mo'tadillik, erkinlik, bir jinslilik farazlarini tekshirish, taqsimot funksiyasi turlarini va

parametrlarini baholash va hk);

- bog'lanishlar va qonuniyatlarni aniqlash (chiziqli va nochiziqli regressiyaga oid tahlil, korrellatsiyaga oid tahlil, va boshq);

- ko'po'lchamli statistik tahlil (chiziqli va nochiziqli diskriminantga oid tahlil, klasterga oid tahlil, komponent tahlil, omilga oid tahlil va boshqa);

- davriy qatorlar asosida dinamik modellar va bashorat.

Data Miningning arsenal statistik usullari usullarni to'rtta guruhga sinflashtirilgan:

1) Dastlabki ma'lumotlarni deskriptiv tahlili va tavsifi.

2) Bog'lanishlar tahlili (korrelyasiyaga oid va regressiyaga oid tahlil, omilga oid tahlil, dispersiyaga oid tahlil).

3) Ko'po'lchamli statistik tahlil (komponent tahlil, discriminant tahlil, ko'p o'lchamli regressiyaga oid tahlil, kanonik korrelyasiya va boshqa).

4) Davriy qatorlarni tahlillash (dinamik modellar va bashorat qilish).

Data Miningning kibernetik usullari.

Data Miningning ikkinchi yo'nalishi – bu kompyuter matematikasi va sun'iy tafakkur nazariyasidan foydalanishni birlashtirilgan g'oyalarni yondashuvlari to'plamidir.

Ushbu guruhga quyidagi usullar kiradi:

a. sun'iy neyron to'rlari (tanib olish, klasterizatsiya, bashorat);

b. evolyutsion dasturlashtirish;

c. genetik algoritmlar (optimallashtirish);

d. assotsiativ xotara (analoglar qidiruvi, prototiplarini);

e. noravshan mantiq;

f. daraxt yechimlar;

g. ekspert bilimlarini qayta ishlash.

Data Mining masalalari bo'yicha sinflashtirish mumkin. Bunday sinflashtirishlarga muvofiq ikkita guruhni ajratib ko'rsatamiz.

Ulardan birinchisi – bu Data Mining usullarini **hal qiluvchi segmentlash masalasiga** (ya'ni sinflashtirish va klasterlash masalalariga) va bashoratlash masalasiga bo'linishi.

Data Mining usullari masalalari bo'yicha ikkinchi sinflashtirish bilan muvofiq holda tavsifli va bashoratli natijalarni olishga imkon bo'lishi mumkin.

Tavsifiy usullar talqinlari tahlil qiluvchi nuqtai nazaridan beriluvchi

yetarlicha shaffof tarzda tasvirlaydi. Olingan qoidalar uzluksiz va diskret o'zgaruvchilarni o'z ichiga oladi. Daraxt yechimlar daraxt ildizidan terminal uchigacha bo'lgan yo'l bo'ylab bir qoidani yaratish orqali belgili qoidalar to'plamiga osongina aylantirilishi mumkin. Daraxt yechimlar va qoidalar haqiqatda bir masalani yechishni turli xil usullari bo'lib hisoblanadi va ular imkoniyatlari bilangina farqlanadi. Bundan tashqari qoidalarni amalga oshirish daraxt yechimlarni induksiyasiga ko'ra sekinroq ishlaydigan algoritmlar bilan amalga oshiriladi.

Kross-tabulyasiya usullari: agentlar, bayes (ishonch) to'rlari, kross-jadvalli vizualizatsiya. Oxirgi usul Data Mining xossalaridan biriga – tahliliy tizim qonuniyatlarini mustaqil qidirishga unchalik javob bermaydi. Biroq ma'lumotni kross jadval ko'rinishida taqdim etish Data Miningning asosiy masalasi bo'lgan shablonlarni qidirishni ishlab chiqishni ta'minlaydi, shuning uchun ushbu usulni ham Data Miningning usullaridan biri deb hisoblash mumkin.

Tenglamalar asosidagi usullar. Ushbu guruh usullari matematik ifodalar tenglamalar ko'rinishidagi aniq qonuniyatlarni ifodalaydi. Ular sonli o'zgaruvchilar bilangina ishlashi mumkin, va boshqa tiplar o'zgaruvchilari mos holda kodlashtirilishi kerak bo'ladi. Bu guruh usullarini qo'llashni biroz cheklaydi, shunga qaramay ular turlicha masalalarni, ayniqsa bashoratlash masalalarini hal qilishda keng foydalaniladi. Berilgan guruhning asosiy usullari: statistik usullar va neyron to'rlar. Statistik usullar bashoratlash masalalarini yechish uchun tez-tez qo'llaniladi. Ular orasida ma'lumotlarni statistik tahlil qilish usullari to'plami mavjud, masalan korrelyasion - regression tahlil, bir qator dinamikalar korrelyasiyasi, dinamik qatorlardagi tendensiyalarni, garmonik tahlillarni aniqlash.

Boshqa sinflashtirish Data Miningning barcha ko'p qirrali usullarini ikkita guruhga ajratadi: statistik va kibernetik usullar. Ushbu bo'lish sxemasi matematik modellarni o'qitish uchun turlicha yondashuvlarga asoslangan.

Data miningning statistik usullari.

Ushbu usullar o'zi bilan quyidagi to'rtta o'zaro bog'liq bo'limlarni ifodalaydi:

- statistik ma'lumotlar tabiatini dastlabki tahlili (statsionarlik, mo'tadillik, erkinlik, bir jinslilik farazlarini tekshirish, taqsimot funksiyasi turlarini va

parametrlarini baholash va hk);

- bog'lanishlar va qonuniyatlarni aniqlash (chiziqli va nochiziqli regressiyaga oid tahlil, korrellatsiyaga oid tahlil, va boshq);

- ko'po'lchamli statistik tahlil (chiziqli va nochiziqli diskriminantga oid tahlil, klasterga oid tahlil, komponent tahlil, omilga oid tahlil va boshqa);

- davriy qatorlar asosida dinamik modellar va bashorat.

Data Miningning arsenal statistik usullari usullarni to'rtta guruhiga sinflashtirilgan:

- 1) Dastlabki ma'lumotlarni deskriptiv tahlili va tavsifi.

- 2) Bog'lanishlar tahlili (korrelyasiyaga oid va regressiyaga oid tahlil, omilga oid tahlil, dispersiyaga oid tahlil).

- 3) Ko'po'lchamli statistik tahlil (komponent tahlil, discriminant tahlil, ko'p o'lchamli regressiyaga oid tahlil, kanonik korrelyasiya va boshqa).

- 4) Davriy qatorlarni tahlillash (dinamik modellar va bashorat qilish).

Data Miningning kibernetik usullari.

Data Miningning ikkinchi yo'nalishi – bu kompyuter matematikasi va sun'iy tafakkur nazariyasidan foydalanishni birlashtirilgan g'oyalarini yondashuvlari to'plamidir.

Ushbu guruhga quyidagi usullar kiradi:

- a. sun'iy neyron to'rlari (tanib olish, klasterizatsiya, bashorat);

- b. evolyutsion dasturlashtirish;

- c. genetik algoritmlar (optimalashtirish);

- d. assosiativ xotara (analoglar qidiruvi, prototiplarini);

- e. noravshan mantiq;

- f. daraxt yechimlar;

- g. ekspert bilimlarini qayta ishlash.

Data Mining masalalari bo'yicha sinflashtirish mumkin. Bunday sinflashtirishlarga muvofiq ikkita guruhni ajratib ko'rsatamiz.

Ulardan birinchisi – bu Data Mining usullarini **hal qiluvchi segmentlash masalasiga** (ya'ni sinflashtirish va klasterlash masalalariga) va bashoratlash masalasiga bo'linishi.

Data Mining usullari masalalari bo'yicha ikkinchi sinflashtirish bilan muvofiq holda tavsifli va bashoratli natijalarni olishga imkon bo'lishi mumkin.

Tavsifiy usullar talqinlari tahlil qiluvchi nuqtai nazaridan beriluvchi

tavsiflanuvchi ma'lumotlarni shablonlari yoki namunalarini topish uchun xizmat qiladi.

Tavsifiy natijalarni olishga yo'naltirilgan usullarga klasterga oid tahlilning iterativ usullari kiradi, jumladan: k- o'rtacha algoritmi, k-medianalar, klaster tahlilning ierarxik usullari, o'zini boshqara oladigan Koxonen xaritalari, kross-jadvalli vizuallashtirish usullari, turlicha vizuallashtirish usullari va boshqalar.

Bashorat qiluvchi usullar boshqa o'zgaruvchilarning (maqsad) noma'lum (tushirib qoldirilgan) yoki kelgusi qiymatlarini avvaldan aytish/bashorat qilish uchun bir o'zgaruvchi qiymatlaridan foydalanadi.

Bashorat qilinuvchi natijalarni olishga qaratilgan usullarga quyidagi usullar kiradi: neyron to'rlari, daraxt yechimlar, chiziqli regressiya, yaqin qo'shnilar usuli, tayanch vektorlar usuli va boshqa.

Intellektual texnologiyalar quyidagi istiqbollarga ega.

- Fan sohasini evristikalariga mos ravishda tipini aniqlash va shakllantirish ushbu sohalarga tegishli bo'lgan ma'lumotlarni intellektual tahlil qilish masalasini yechishni soddalashtiradi.

- Rasmiy tillar va mantiqiy vositalarni yaratish. Bular yordamida mulohazalar va avtomatlashtirish shakllantirilib, ular aniq bir predmet soha ma'lumotlarni intellektual tahlil qilish masalasini yechishda vosita bo'lib xizmat qiladi.

- Ma'lumotlarni intellektual tahlil qilishning nafaqat ma'lumotlardagi qonuniyatlarni aniqlovchi balki, empirik ma'lumotlarga asoslangan ayrim nazariyalarni shakllantiruvchi usullarini ishlab chiqish.

- Ma'lumotlarni intellektual tahlil qilish vositalari.

9. TASNIFLASH MASALASI

Tasniflash (Sinflashtirish) eng oddiy va shu bilan birga eng tez-tez hal qilinadigan Data Mining masalasi bo'lib hisoblanadi. Tasniflash masalasi keng tarqalganligi tufayli ushbu tushunchaning mohiyatini aniq tushunish kerak. Bu yerda ayrim ta'riflar mavjud.

Tasniflash – o'rganilayotgan predmetlar, hodisalar, jarayonlarni jinsi, turi, tipi bo'yicha, ularni o'rganish qulayligi uchun har qanday muhim belgilar bo'yicha tizimli taqsimlash; ushbu o'xshashlik darajasini aks ettiruvchi dastlabki tushunchalarni guruhlash va ularni belgilangan tartibda joylashishi.

Tasniflash – o'xshash sinflashtirish belgilariga (bir yoki bir nechta belgilar) ega ba'zi bir tamoilga muvofiq obyektlar to'plami bo'lib, ushbu obyektlar o'rtasidagi o'xshashlik yoki farqni aniqlash uchun tanlangan.

Tasniflash quyidagi qoidalarni o'z ichiga oladi:

- bo'linmaning har bir aktida faqat bir asosni qo'llash zarur;
- bo'linma mutanosib bo'lishi kerak, ya'ni turga oid tushunchalarning hajmi bo'linuvchi turga oid tushunchaga teng bo'lishi kerak;
- bo'linma hadlari o'zaro bir-birini chiqarib tashlashi kerak, ularning hajmi bir-biriga zid bo'lmasligi kerak;
- bo'linma ketma-ket bo'lishi kerak. Shu bilan farqlanadiki: tashqi belgi bo'yicha amalga oshiriluvchi va predmetlar (jarayonlar, hodisalar) to'plamini kerakli tartibda berish uchun xizmat qiluvchi yordamchi (sun'iy) tasniflash;
- predmetlar va hodisalarni ichki umumiyligini tavsiflovchi muhim belgilar bo'yicha amalga oshiriluvchi tabiiy tasniflash.

Oxirgisi ilmiy tadqiqotni muhim vositasi va natijasi bo'lib hisoblanadi, shuningdek sinflashtiriluvchi obyektlar qonuniyatlarini o'rganish natijalarini taxmin qiladi va mustahkamlaydi.

Tanlangan belgilarga, ularning birlashishi va bo'lish tushunchalariga bog'liq holda tasniflash quyidagicha bo'lishi mumkin:

- oddiy – turga oid tushunchani faqatgina belgi bo'yicha va barcha turlarni ochishdan avval faqat bir marta bo'lish. Bunday tasniflashga misol bo'lib dixotomiya bo'lib, unda faqat ikkita tushuncha bo'linma hadlari bo'lib, har biri boshqasiga ziddir (ya'ni tamoilga rioya qilinadi: "A va A emas");
- murakkab – bir tushunchani turlicha asoslar bo'yicha bo'lish va bunday bo'lishlarni bir butun holda sintez qilish qo'llaniladi. Bunday tasniflashga misol bo'lib, kimyoviy elementlarni davriy sistemasi bo'lib hisoblanadi.

Tasniflash deganda obyektlarni (kuzatishlarni, hodisalarni) ma'lum bir sinflardan biriga tayinlash tushuniladi.

Tasniflash – bu ma'lum bir guruhning xususiyatlarini aniqlash to'g'risidagi nisbiy xulosa chiqarishga imkon beruvchi qonuniyat. Shunday qilib, tasniflashni o'tkazish uchun u yoki boshqa hodisa yoki obyektga tegishli bo'lgan guruhni tavsiflovchi belgilar mavjud bo'lishi zarur (odatda, bunda allaqachon sinflashtirilgan hodisalarning tahlili natijasida ba'zi qoidalar tuziladi).

Tasniflash nazorat qilinuvchi yoki boshqariluvchi o'qitish deb ham nomlanuvchi o'qituvchili o'qitish (supervised learning) strategiyasini anglatadi.

Tasniflash masalalari odatda uzluksiz va/yoki kategoriyali o'zgaruvchilar asosida kategoriyali bog'liq o'zgaruvchini avvaldan aytish (ya'ni, kategoriya bo'lib hisoblangan bog'liq o'zgaruvchi) deb nomlanadi.

Masalan, oldindan aytish mumkinki, kim firmaning mijozlaridan bo'lsa ma'lum bir mahsulotning potensial xaridori bo'lib hisoblanadi, kimdir esa – yo'q, kimdir firmaning xizmatidan foydalanadi, kimdir esa – yo'q. Masalalarni ushbu turi binarli tasniflashga tegishli, ularda bog'liq o'zgaruvchi faqat ikkita qiymatni qabul qilishi mumkin (masalan, ha yoki yo'q, 0 yoki 1).

Tasniflashning boshqa turi, agar bog'liq o'zgaruvchi oldindan belgilangan sinflarni ba'zi to'plamlaridan qiymatni qabul qilishi mumkin bo'lsa, yuzaga kelishi mumkin. Masalan, mijoz avtomobilni qanday markalisini xohlashini oldindan aytib berish mumkin bo'lgan paytda. Bunday hollarda bog'liq o'zgaruvchi uchun sinflar to'plami ko'rib chiqiladi.

Tasniflash bir o'lchamli (bir belgisi bo'yicha) va ko'po'lchamli (ikki yoki undan ko'proq belgilar bo'yicha).

Inson tomonidan noaniq va noravshan muhitda qaror qabul qilishni amalga oshiruvchi mashinalarni yaratish ko'pgina hollarda, sun'iy intellekt sohasida noaniqlik bilan xarakterlanuvchi masalalarni yechishning aniq modellari va algoritmlarini ishlab chiqish bilan uzviy bog'liq. Sun'iy intellekt usullari ikkita xarakterli xususiyatlarga asoslangan:

1. Belgili ko'rinishdagi axborotlar: harf, so'z, ibora, belgi, rasmlardan foydalanish.

2. Belgili mantiqdan foydalangan holda qidirish.

Belgili sonli bo'lmagan axborotni qayta ishlashda kompyuterlar so'z va

iboralarni ikkilik shaklda kodlashtirilgan belgilarga almashtiradi. Sun'iy intellekt klassiklarining qayd etishlaricha, kompyuterlar tomonidan belgilarni boshqarish ham shu qadar osondir. So'ngra, keyingi amallarni bajarish yoki bajarmasligini solishtirish natijalari mashinada inson ongining eng yuqori namoyondasi- deduktiv mantiqiy mulohazalar funksiyalarini amalga oshirish imkonini beradi. Axborotni belgili qayta ishlashga asoslangan sun'iy intellektni yaratish yo'lidagi kompyuterning potensial imkoniyatlari chegarasiz ko'rinishi mumkin. Klassik ma'nodagi sun'iy intellektning ulkan muvaffaqiyatlariga qaramay, masalalarni yechish, teoremlarni avtomatik isbotlash, o'yin tizimlari, ekspert tizimlari, tabiiy tilni tushunish tizimlarining keng spektrini yaratishda unga yuklangan umidlar o'zini oqlamadi. An'anaviy sun'iy intellekt masalalarni inson ularni ongli ravishda yechishi kabi yecha olmaydi, insonning anglash va mulohaza yuritish qobiliyatiga o'xshash protseduralar bilan muvozanatda bo'la olmaydi. An'anaviy sun'iy intellekt robotshunoslik, kompyuterning ko'rish qobiliyati, nutqni va qo'lyozmalarni aniqlashtirish, mashina tarjimai, ta'lim berish va dunyoviy boshqa muhim masalalarni yechishda o'zini ko'rsata olmadi. Qayd etilgan muammolar va haqiqiy dunyoning boshqa murakkab masalalari nazardan chetda qoldirib bo'lmaydigan noaniqlik va noravshanlik bilan bog'liqdir. Professor L.Zadening qayd etishicha, an'anaviy sun'iy intellekt maqsadlariga erishish yo'lida o'zini faqatgina belgili axborotni qayta ishlash va birinchi tartibli mantiq (first order logic) dan foydalanish bilan cheklab qo'ymaganida katta muvaffaqiyatlarga erishishi mumkin edi.

Sun'iy intellektning barcha an'anaviy tizimlari Hard Computing texnologiyasi (ya'ni dasturni zaruriy funksionallashtirishga erishishga qaratilgan sun'iy intellektual protsedura ta'siri ostida yangilanmaydigan operatorlarning qat'iy ketma-ketligi sifatida qabul qiluvchi kompyuter texnologiyasidan foydalanish) asosida ishlab chiqilgan, bu esa shu tizimlarning imkoniyatlarini anchagina cheklab qo'ygan. Yana bir narsa. An'anaviy sun'iy intellekt yuqorida keltirilgan xususiyatlari hisobiga noaniqlik va noravshanlikni hisobga olishda juda muhim hisoblangan sonli usullardan foydalanmadi. Qayd etilgan holatlarga ko'ra an'anaviy sun'iy intellekt tizimlarining mashina intellekt darajasi – MIQ (Machine Intelligence Quotient) uncha yuqori emas ekan. Shuning uchun intellektual tizimlarning mashina intellekt darajasini oshirish masalasi qo'yilgan edi. Jumladan bunda kalit uslubiyat vazifasini

sun'iy intellektning usullari va umuman olganda boshqa usullar bilan ham yechib bo'lmaydigan haqiqiy dunyoning ko'pgina masalalarini yechishga imkon beruvchi noravshan mantiq, neyron tarmoqlar, evolyutsion hisoblashlar kabi yangi sonli yondashuvlarni avtonom ravishda emas, birgalikda qo'llashni nazarda tutuvchi Soft Computing o'taydi [1-4].

Har tomonlama noaniqlik, hamda atrofimizni o'rab olgan haqiqiy dunyoning to'liq bo'lmagan aniqlik bilan ish ko'ruvchi Genetik Algoritm-lar-GA (genetic algorithms), Xaos Nazariyasi-CT (chaos theory) va Ta'lim Nazariyasi-LT (learning theory) ning bir qismini o'z ichiga olgan Noravshan Mantiq-FL (Fuzzy Logic), Neyron Tarmoqlar-NN (neural networks), Ehtimolli Mulohazalar-PR (probabilistic reasoning) kabi intellektual paradigmlar (kompyuterli qayta ishlash vositalari sifatida) majmui Soft Computing – SC (so'zma-so'z tarjima qilganda, "yumshoq", egiluvchan, hisoblash qayta ishlash) ni tashkil etadi. An'anaviy hisoblash qayta ishlashdan (Hard Computing (HC)-ya'ni "qattiq", qat'iy hisoblash) farqli o'laroq SC ishlab chiqaruvchanlik va samaradorlikni yo'qotmagan holda noaniqlik, noravshanlik va qisman rostlik bilan ish ko'rishga qodirdir. Bu vaqt masalasidir, 10 yil o'tmay sun'iy intellekt an'anaviy HC ga emas, Soft Computing (SC) ga asoslanadi. L.Zadening qayd etishicha, an'anaviy HC dan farqli o'laroq, SC ning maqsadi haqiqiy dunyoning har joydagi noaniqlik bilan hamohanglikda yashashdir. SC ning asosiy tamoyili noaniqlik, noravshanlik va qisman rostlikni hisobga olgan holda bayon etilishga (sifatli belgili yoki lingvistik ko'rinishda ifodalash), robustlikka (kichik xatolikalarga bardoshlilikka) va yechimning past narxiga erishishdir [5].

Aniqlik o'zining narxiga egadir, afsuski bu tamoyil ko'pincha hisobga olinmaydi, ya'ni me'yoriy narxdagi masalani yechish uchun yetarli darajada aniq, lekin keragidan ortiq bo'lmagan yechimga e'tiborni qaratish kerak. Aynan shu noaniqlik va qisman aniqlikka (partial truth) ham taaluqlidir. Aytib o'tilganlarning yaqqol misollari vertolyotning qo'nish masalasi yoki mashinaning avtoto'xtovda joylashish masalasidir. Ikkinchi holni ko'rib o'tamiz. Mashinani uning yakuniy vaziyati berilmaganligi tufayli masofalarni va burchakni o'lchamay turib ham joylashtirish mumkin.

Agar bu o'rinli bo'lsa, u holda millimetrlar va bir nechta burchak sekundalar chegarada o'lchashlar olib borishga to'g'ri kelsa, u holda chamalash uchun ko'pvaqt va qurilmaviy o'lchashlar talab etilgan. Undan

tashqari, yechimning narxi aniqlikni oshirish bilanoq geometrik tarzda ortib ketardi.

Soft Computing texnologiyasi ma'lumotlarni siqishda ayniqsa HDTV (katta ruhsat beruvchi qobiliyatli televizorlar), sifatli ovoz yozish, tasvir va nutqni aniqlashtirish hamda shular bilan bog'liq sohalarda ulkan ahamiyat kasb etadi. Aslida, Soft Computingga asoslangan qonun va vositalar yuqori darajadagi MIQ li mahsulotlar va tizimlarni loyihalashtirish va ishlab chiqarishda katta rol o'ynaydi. L.Zade qayd etganidek, SC ning eng yaxshi modeli inson miyasidir.

9.1. Soft Computing (SC) ning tuzilmasi va tarkibi

Yuqorida qayd etilganidek, sun'iy intellektning barcha an'anaviy tizimlari, shu jumladan inson faoliyatining turli xil jabhalarida keng qo'llaniluvchi ekspert tizimlari HC asosida, ko'pincha kompyuterlar asosida ishlab chiqarilgan. Hisoblashlarning bunday asosi esa, tabiiyki samaradorlikni va umuman olganda har xil maqsadlarga yo'naltirilgan sun'iy intellekt tizimlarini yaratish imkoniyatini chegaralab qo'ygan. Bugungi kunda belgili qayta ishlash va an'anaviy HC ga emas, neyron tarmoqlar, noravshan qayta ishlash (fuzzy computing), evolyutsion hisoblashlar, belief-tarmoqlarga asoslangan sun'iy intellektning amaliy tizimlari soni sezilarli darajad ortib bormoqda. Shuningdek, ilmiy konferensiyalarning ishlari, sun'iy intellektga oid jumallarda noravshan mantiq (fuzzy logic), genetik algoritmlar, sun'iy hayot (artificial life), biologik hisoblashlar (biological computing), neyron hisoblashlar (neural computing) va boshqalarga oid maqolalar soni sezilarli darajada ortib bormoqda.

Qayd etilganlarning guvoh berishicha, sun'iy intellekt tizimlarini ishlab chiqish va o'rganishning og'irlik markazi Soft Computing tomonga siljib bormoqda.

Hozirgi vaqtda Soft Computingning asosiy qismlari quyidagilardir: Noravshan Mantiq-FL (fuzzy logic), Neyron Tarmoqlar Nazariyasi-NN (neural networks), Ehtimolli Mulohazalar - PR (probabilistic reasoning), ular o'z ichiga Genetik Algoritm-lar (genetic algorithms), Xaos Nazariyasi-CT (chaos theory) ni oladi. Soft Computing tarkibidagi FL asosan noaniqlik va taxminiy mulohazalar, NN-ta'limot bilan va PR-noaniqlik bilan ish ko'radi. Umuman olganda FL, NN va PR lar ustuvor yondashuvlardan ko'ra, bir-birini

to'ldiruvchi yondashuvlardir

9.2. SC tarkibiy qismlarining taqqoslama tavsiflari

Soft Computing qismlari mustaqil ravishda, masalan noravshan hisoblashlar (fuzzy computing), neyron hisoblashlar (neural computing), evolyutsion hisoblashlar (evolutionary computing) da, hamda ko'pincha kombinatsiyali ko'rinishda qo'llanilishi mumkin. Mustaqil qo'llanilishga asoslanib, Soft Computingni tashkil etuvchilar, noravshan texnologiya, neyron texnologiya, xaos texnologiya va boshqalar ayni vaqtda sanoat sohasida ham, undan tashqarida ham keng qo'llanilib kelinmoqda.

Soft Computingning old tarkibiy qismi noravshan mantiq (fuzzy logic) dir. Soft Computingda noravshan mantiq (FL) sonli vazifani o'taydi. FL so'z bilan ifodalanishni va hisoblashlarning interpretatsiyalanishini ta'minlaydi [5]. Noravshan mantiq ko'pgina sanoat sohaslarida muvaffaqiyatli qo'llanilgan. Robot texnikasi, mayatnik tizimi (inverted pendulum system) ni muvozanatlash, qaror qabul qilish va tashxis qo'yishning murakkab tizimlarida, ma'lumotlarni siqishda, TV va boshqa sohalarda, lingvistik shaklda yoki noravshan sonli ma'lumotlar ko'rinishida berilgan bilimlarni boshqarishga qaratilgan tizimli protsessorni loyihalashtirishda biz tizimning noravshan modeliga muhtojlik sezamiz. Noravshan to'plamlarni mukammal approksimatsiya tariqasida ishlatish mumkin, bu esa noma'lum obyektlarni modellashtirishda, hamda operator berilgan vaziyatda qanday turdagi amaldan foydalanishi to'g'risida aniq tasavvurga ega bo'lmaganida o'ta muhimdir, uning boshqaruv faoliyatini sonli ma'lumotlardan foydalangan holda modellashtirish foydalidir.

Lekin sof ko'rinishdagi noravshan mantiq intellektual tizimlarni yaratishda har doim ham o'rinli bo'lavermaydi. Xususan, loyihalashtiruvchi tizim to'g'risidagi yetarli aprior axborot (bilim) ga ega bo'lmasa, noravshan qoidalarining o'rinli bazasini qurishning iloji bo'lmaydi. Tizim murakkablashib borishi bilan tizimning xatti-harakatini to'g'ri ta'riflash uchun qoidalar va tegishlilik funksiyalarning to'g'ri to'plamini aniqlash bilan bog'liq qiyinchilik paydo bo'ladi. Noravshan tizimlar ham tajriba natijalari bo'yicha qo'shimcha bilimlarni ajratib olish va tizimning funksionallashtirish sifatini aniqlash uchun noravshan qoidalarini to'g'irlashga taaluqli kamchiliklardan aziyat chekadilar.

Soft Computing ning boshqa muhim qismi neyron tarmoqlardir. Sun'iy

neyron tarmoqlar parallel hisoblash modellari bo'lib, nochiqli statistik va dinamik tizimlarning mayda sodda parallel ishlab chiqarilishni ifodalaydilar. Ushbu tarmoqlarning eng muhim omili ularning "misolga qarab o'rganish" ning o'rmini an'anaviy "dasturlash" egallovchi moslanadigan tabiatidir. Boshqa kalit omil biriktirilgan parallelizmdir, u parallel sonli kompyuterlar asosida tez yechadigan sxemalarni ishlab chiqarishga imkon beradi. Sun'iy neyron tarmoqlar aniqlashtirish, egri chiziq va funksiyalarni approksimatsiyalash, ma'lumotlarni siqish, bog'liqlik xotira, nochiqli no'malum tizimlarni modellashtirish va boshqarish kabi masalarni yechishning zaruriy hisoblash modellardir. [6,7].

Neyron tarmoqlar hisoblash samaradorligi va ularning samarali apparatli amalga oshirilish jihatidan ajralib turadi. Ular umumlashtirish xossasi-yangi obrazlarni to'g'ri sinflashtirish qobiliyatiga egadirlar. Neyron tarmoqlarning kamchiligi yomon interpretatsiyalanishidir. Neyron tarmoqlarning kamchiligi qayd etayotganda ularni "qora quti" bilan solishtiradilar [5].

Evolutsion hisoblashlar-EC (evolutionary computing)-muqobillashtirishga revolyutsion yondashuv. EC ning tarkibiy qismi-genetik algoritmlar natural seleksiya va natural gentika mexanizmlariga asoslangan global muqobillashtirish algoritmlardir [8]. Genetik algoritmlarning ulkan ustuvorliklaridan biri ularni parallel ko'p o'lchovli qidiruv asosida samarali ishlab chiqarish imkoniyatidir.

Genetik algoritmnini bajarish mexanikasi juda sodda. Amallarning oddiyligi va quvvatli hisoblash samarasi genetik algoritmlarning ikkita asosiy ustuvorligidir. Genetik algoritmlarning kamchiligi, birinchidan yaqinlashish muammosi va umuman nazariy tuzilmaning yo'qligidir. Haqiqiy (natural) o'zgaruvchilarni bitli qatorlarga kodlash zarurati ham genetik algoritmlarning kamchiligidir. 9.1-jadvalda Soft Computing qismlarining taqqoslama tavsiflari keltirilgan.

Soft Computing qismlarini qo'llash sohasiga taaluqli o'ziga xos muammolar mavjud. Masalan, avtomobilni joylashtirish masalasi neyron tarmoqlar, GA va boshqalarni qo'llamasdan, faqatgina noravshan mantiqdan foydalanish asosida yechilishi mumkin. Perexvat masalasini FL, GA, NN da emas, CN da yechish maqsadga muvofiq.

9.1-jadval.

Soft Computing qismlarining taqqoslama tavsiflari			
	Noravshan to'plamlar	Sun'iy neyron tarmoqlar	Evolutsion hisoblashlar
Kamchilik jihatlari	Bilimlarni ajratib olish, O'rganish	Interpretatsiyalaruv-chilik "Qora quti" ning tabiati	Kodlash Hisoblash tezligi
Ustuvor jihatlari	Interpretatsiyalaruvchilik Shaffoflik, Chinlik, Navbatma-navbatlilik Modellashtrish Mantiqiy chiqarish Noaniqlikka tolerantlik	O'rganuvchanlik Moslashuvchanlik Xatolarga tolerantlik Egri chiziqlarning approksimatsiyasi Umumlashtirish qobiliyati Approksimatsiyalash qobiliyati	Hisoblash samaradorligi Global muqobillashtirish

9.1. SC tarkibiy qismlarining intellektual kombinatsiyasi Gibrid tizimlar

Yuqorida qayd etilganidek, Soft Computingning qismlari- noravshan mantiq, neyronli qayta ishlash va ehtimolli mulohazalar o'zaro bellashmasdan, bir-birini to'ldirib turadi. FL, NN, PR ni GA bilan avtonom tarzda emas, kombinatsiyada qo'llash maqsadga muvofiq ekanligi aniq bo'lmoqda. L.Zadening qayd etishicha "bugungi kunda Gibrid Intellektual Tizimlar atamasi FL, NN va PR lar kombinatsiyada qo'llanilgan tizimlarga nisbatan odatiy bo'lib bormoqda. Bizning nuqtai nazarimizga ko'ra Gibrid Intellektual Tizimlar kelajakning tezkor suratlar bilan rivojlanib boruvchi tizimlaridir" [5].

Soft Computing qismlarining quyidagi kombinatsiyalari ma'lum:

neuro computing + fuzzy logic (neuro fuzzy - NF);

fuzzy logic + genetic algorithms (FG);

fuzzy logic + chaos theory (FCh);

neural networks + genetic algorithms (NG);

neural networks + chaos theory (NCh);

neural networks + fuzzy logic + genetic algorithms (NFG);

fuzzy logic + neural networks + genetic algorithms (FNG).

Soft Computing tarkibiy qismlarining boshqa kombinatsiyalari ham kuzatilishi mumkin.

Neyron tarmoqlarning yomon interpretatsiyalanish qobiliyati bir

tomondan, va boshqa tomondan noravshan tizimlarda bilim olishga oid qiyinchiliklar FL&NC ning gibridlashuviga sabab bo'ladi. Neyro-noravshan tizimlar shu muammoni noravshan tizimlarning interpretatsiyalashuvi xossasiga ega bo'lgan konneksionistik qobiliyatlarini birlashtirish yordamida yechishga harakat qiluvchi gibrid tizimlardir.

Yuqorida qayd etilganidek, (2.1), dinamik ishchi muhit holida noravshan tizimlarda bilimlar bazasini avtomatik ravishda to'g'irlash zarur bo'ladi.

Boshqa tomondan, sun'iy neyron tarmoqlar misollarga qarab o'rganish asosida zaruriy darajadagi aniqlikda boshqarish va muqobillashtirish jarayonlari uchun bilimlar ajratib olish bilan bog'liq muammolarda muvaffaqiyatli qo'llanilib kelinmoqda. Noravshan qoidalarda tegishlilik funksiyalari egri chiziqlarining shaklini muqobillashtirish uchun neyron ta'limot tamoyili hamda zaruriy aniqlikka erishish uchun yetarli bo'lgan qoidalar sonining minimallashtirish- neyro-noravshan yondashuvning mazmunidir.

Keng qamrovli funksiyalarning muqobilini global qidirish imkonini beruvchi qarorlarni va genetik algoritmlarning inson tomonidan qabul qilinishida ishlatiluvchi empirik, intiutiv qoidalarga asoslangan(rule-based) noravshan tizimlar kombinatsiyasi samarali, ishchan, moslashuvchan boshqaruv tizimini yaratishga imkon beradi.

Noravshan tizimlar bilimlar bazasining noravshan qoidalarda tegishlilik funksiyalari LR-turdagi noravshan sonlar, masalan trapetsiya, uchburchak ko'rinishida tasvirlanadi. Noravshan bilimlar bazasini to'g'irlash, ya'ni tegishlilik funksiyalari markazlarining qiymatlari va ko'rinishlarini aniqlash uchun gradientli usullardan foydalanish o'zini oqlamaydi. Bunda samarali vosita GA dir. FL va GA ning kombinatsiyasi noravshan nazoratgichning noravshan bilimlar bazasini MB dagi qoidalarning muqobil sonini va tegishlilik funksiyalarining ko'rinishini aniqlash orqali muqobillashtirish imkonini beradi. Bunda GA loyihaluvchi noravshan tizimning munosabatlar matrisasi va tegishlilik funksiyalarini qurish uchun ishlatiladi.

O'z navbatida, FL va GA birgalikdagi ko'rinishida noravshan to'plamlar nazariyasi genetik operatorlar va umuman olganda genetik algoritmlarning xatti-harakatini yaxshilash uchun qo'llanilishi mumkin, ya'ni GA ning samaradorligini oshirish maqsadida noravshan vositalar (tools) ni yaratish mumkin.

Genetik algoritmnining neyron tarmoq bilan turmushi ham samarali natija

beradi. Ma'lumki, sun'iy neyron tizimlarni ishlab chiqishning asosiy masalalaridan biri neyro-tarmoqning parametrlarini sozlash uchun mos usulni tanlashdir. Ushbu usullardan eng mashhuri "backpropagation" algoritmi. Afsuski backpropagation bilan bog'liq ayrim muammolar mavjud. Birinchidan, ta'limot samaradorligi neyron tarmoq og'irliklarining boshlang'ich majmuasiga bog'liq bo'lib, u tasodifiy tarzda aniqlanadi. Ikkinchidan, backpropagation boshqa gradient usul singari lokal minimumlarni bartaraf etishga imkon bermaydi. Uchinchidan, agar o'quv tezligi juda ham kichkina bo'lsa, yechimni topish uchun ko'p vaqt talab etiladi. To'rtinchidan, backpropagation aktivatsiya funksiyalari differensiallanuvchan bo'linishini talab etadi. Bu shart neyron tarmoqlarning ko'pgina turlariga nisbatan bajarilmaydi. Ko'pgina masalalarni muqobillashtirish uchun "kuchli" usullar yaxshi yechim topa olmaganida qo'llaniladigan genetik algoritmlar yuqorida qayd etilgan kamchiliklardan holi ravishda neyron tarmoqlarni o'rganish uchun ishlatiladi. Biologik hisoblashlar (biological computation) ga yaqinroq bo'lgan noravshan xaotik neyron tarmoqlardan foydalanish orqali to'g'riroq natijalarni olish mumkin.

Neyron tarmoqlarni o'rganish uchun, ya'ni noravshan signalli va/yoki, noravshan og'irlikli neyron tarmoqlar ko'zdan kechirilganda, ularni o'rganishda har xil usullardan foydalaniladi. Xususan, standart δ -qoidani to'g'ridan-to'g'ri fazzifikatsiyalash, noravshan og'irliklarni qidirish uchun bacjpropagation algoritmdan foydalaniladi. Shuningdek uchburchak, trapetsiyalar ko'rinishida berilgan neyron tarmoqlarning noravshan og'irliklarini qidirish uchun α -cut protseduradan foydalaniladi. Bunda qidirish masalasi backpropagation va oraliq arifmetikani qo'llashga asoslanadi.

Birinchi holda algoritmnin ishi og'irliklarning to'g'ri qiymatiga yaqinlashish ma'nosida muvaffaqiyatsiz bo'lib chiqishi mumkin. Har bir holatda xato gradientini (error measure) hisoblash zarur, lekin hosilalarni olish juda qiyin (ayniqsa kiruvchi, chiquvchi signallar va og'irliklar uchun umumiy noravshan tarmoqlar holida). FL, NC va GA turmushga asoslangan noravshan neyron tarmoqlarni o'rganish algoritmlari yuqorida qayd etilgan kamchiliklardan holi bo'lib, juda samarali natijalar beradi.

Shubha yo'qki, Soft Computingning yuqorida qayd etilgan intellektual kombinatsiyalari rivojlanib boradi va yangilari ham paydo bo'lib boraveradi. Ular hisoblash intellekti (Computational Intelligence), biologik hisoblashlar (Biological Computing) va sun'iy hayotning (Imitating Life) fundamenti bo'lib

qoladi.

10. TASNIFLASH JARAYONI.

Tasniflash jarayonining maqsadi shundan iboratki, kirish parametrlari sifatida bashorat qilinuvchi atributlardan foydalaniladigan va bog'liq atribut qiymati olinadigan modelni qurish. Tasniflash jarayoni obyektlar to'plamini ma'lum mezon bo'yicha sinflarga bo'lishdan iborat.

Tasniflagich belgilarning vektori bo'yicha obyektini qaysi oldindan belgilangan sinflarga tegishligini aniqlaydi.

Matematik usullar bilan tasniflashni o'tkazish uchun tasniflashning matematik apparatidan foydalangan holda bajarishi mumkin bo'lgan obyektning formal tavsifiga ega bo'lishi zarur. Bizni holatda bunday tavsifi bilan ma'lumotlar bazasi chiqadi.

Har bir obyekt (ma'lumotlar bazasini yozib olish) obyektini ayrim xossalari to'g'risidagi axborotni tashiydi.

Daslabki ma'lumotlar to'plami (yoki ma'lumotlar tanlanmasi) ikkita to'plamga bo'linishi mumkin: o'qitiluvchi va test qilinuvchi.

O'qitiluvchi to'plam (training set) – modelni o'qitish (qurish) uchun foydalaniladigan ma'lumotlarni o'z ichiga olgan to'plam. Bunday to'plam misollarni kirish va chiqish qiymatlaridan iborat.

Chiqish qiymatlari modelni o'qitish uchun avvaldan belgilangan.

Testli (test set) to'plam misollarni kirish va chiqish qiymatlaridan iborat. Bu yerda chiqish qiymatlaridan modelni ishlash layoqatini tekshirish uchun foydalaniladi.

Tasniflash jarayoni ikki bosqichdan iborat: modelni qurish va undan foydalanish.

Modelni qurilishi: oldindan ma'lum sinflar to'plamini tavsiflash. Ma'lumotlar to'plamini har bir misoli oldindan belgilangan sinfga tegishli bo'ladi.

Ushbu bosqichda o'qitiluvchi to'plamdan foydalaniladi, unda modelni qurish amalga oshiriladi.

Olingan model daraxt yechim yoki matematik formulaning tasniflashga oid qoidalari bilan berilgan.

Modeldan foydalanish: yangi yoki noma'lum qiymatlarni tasniflash.

Modelning to'g'riligini (aniqligini) baholash.

Olingan modeldan foydalanish natijalari bilan ma'lum testga oid misol qiymatlari taqqoslanadi.

Aniqlik darajasi – testli to'plamda to'g'ri tasniflangan misollarni foizi.

Testli to'plam, ya'ni qurilgan model test qilinuvchi to'plam, o'qitiluvchi to'plamga bog'liq bo'lmasligi zarur.

Agar model aniqligi yo'l qo'yilgan bo'lsa, sinfi noma'lum bo'lgan yangi misollarni tasniflash uchun modeldan foydalanish mumkin.

Sinflashtirish masalasi obrazni bir nechta sinflardan biriga kiritishdan iborat.

Aytaylik, X-obyektlar bayonlarining to'plami, Y - sinflar raqamlari (yoki nomlari) to'plami bo'lsin. Ma'lum: $X^m = \{(x_1, y_1), \dots, (x_m, y_m)\}$ - o'quv tanlanmasi. Talab qilinadi: quyidagi akslantirishni qurish: $f: X \rightarrow Y, f(x_i) = y_i$.

Tanib olinayotgan no $X_r = (x_{r,1}, x_{r,2}, \dots, x_{r,n})$ chiziqli bog'liqlik "kirishlar-chiqish" ma'lumotlari tanlanmasi bilan ifodalanadi:

$$(X_r, Y_r), r = \overline{1, M},$$

bu yerda- kirishlar vektoriva Y_r - r - juftlikdagi chiqish vektori; M - tanlanma hajmi.

11. TASNIFLASH MASALALARINI YECHISH UCHUN QO'LLANILUVCHI USULLAR

Tasniflash uchun turlicha usullardan foydalaniladi. Ulardan asosiylari:

- Daraxt yechimlar yordamida tasniflash;
- bayes (sodda)tasniflash;
- sun'iy neyron to'rlari yordamida tasniflash;
- tayanch vektorlar usuli bilan tasniflash;
- statistik usullar, jumladan, chiziqli regressiya;
- yaqin qo'shnilar usuli tasniflash;
- CBR-usul bilan tasniflash;
- Genetik algoritmlar yordamida tasniflash;
- Tasniflash aniqligi: xatolar darajasini baholash;

Tasniflash aniqligini baholash kross-tekshirish yordamida o'tkazilishi mumkin. Kross-tekshirish (Cross-validation) – bu kross-tekshirish to'plami deb ham nomlanuvchi testli to'plam ma'lumotlarida tasniflash aniqligini baholash jarayonidir. Testli to'plamni tasniflash aniqligi o'qitiluvchi to'plam tasniflash aniqligi bilan solishtiriladi. Agar testli to'plamni tasniflash o'qitiluvchi to'plamni tasniflash kabi tahminan aniqlik bo'yicha shunday natijalarni bersa, berilgan model kross-tekshiruvdan o'tgan bo'lib hisoblanadi.

O'qitiluvchi va testli to'plamga bo'lish ma'lum nisbatda tanlanmani bo'lish yo'li bilan amalga oshiriladi, masalan, o'qitiluvchi to'plam – uchdan ikki qism, testli to'plam - uchdan bir qism ma'lumotlardan iborat. Ushbu usuldan ko'p sonli misollarni tanlash uchun foydalaniladi. Agar tanlanma kichik hajmlga ega bo'lsa, qisman kesishishi mumkin bo'lgan o'qitiluvchi va testli tanlanmadan foydalanishda maxsus usullarni qo'llash tavsiya qilinadi.

MIT tizimining asosiy masalalaridan biri klassifikatsiya masalasi hisoblanadi. Klassifikatsiya masalasi turkumiga o'qitish orqali klassifikatsiyalash, avtomatik klassifikatsiya (taksonomiya – klasterizatsiya), bashoratlash kabi jarayonlarni boshqarishning asosiy elementlari kiradi. Noravshan yondashuv orqali bemalol timsollarni tanib olish, ovozni tanib olish, yozuvlarni tanib olish kabi bir qator masalalarni yechish mumkinligini noravshan to'plamlar nazariyasi asoschisi Lutfi Zade ta'kidlab o'tgan.

Noravshan yondashuvli klassifikatsiyalash borasida dunyoda bir qator ishlar amalga oshirilmoqda, hususan, ularning aksariyati noravshan k - o'rtacha, noravshan neyron to'rlari, genetik algoritmlarning noravshan

modellarda qo'llanilishi, noravshan daraxt kabilarda o'z aksini topmoqda. Ushbu izlanishlar noravshan yondashuvlar asosida hamda kombinatsiyalangan ko'rinishlardan foydalanish an'anaviy yondashulardan ko'ra 3-5% samaraliroq ekanligini ko'rsatmoqda. Bunday yondashuvlarga asoslangan ishlar bilan quyida tanishib chiqamiz.

Noravshan k - o'rtacha borasida qilingan ishlar birinchilaridan deb olishimiz mumkin. Unda polimer moddalar namunalarini infra qizil spektrlar yordamida aniqlab olish usuli keltirilgan bo'lib, toza polimer va aralashgan polimerlarni ajratish masalasi muvofaqiyatli yechilgan.

Noravshan modellarni yana bir keng tarqalgan vakili bu neyro-noravshan yondashuv hisoblanadi. Unda noravshan to'plam va noravshan qoidalarni aniqlab olish uchun lokal o'qitish strategiyasidan foydalanilgan. Neyro-noravshan yondashuvning ishlatilishi noravshan kontrolerlar va noravshan klassifikatorlarni ishi samaradorligini oshirish maqsadida qo'llanilgan. Ushbu ishda klassifikatsiya masalasini yechish uchun noravshan qoidalarni o'qitish usuli ishlab chiqilib, uning yordamida tegishlilik funksiyasi parametrlarini o'zgartirish orqali noravshan qoidalarni o'quv ma'lumotlaridan tezda chiqishi va shu bilan birga aniqlikni kamaytirmaslikka erishish ko'rsatib berilgan. Ushbu algoritmlar ham evristik algoritmlar jumlasiga kiradi. Ushbu yondashuv asosida NEFCLASS dasturiy ishlanmasi taklif etilgan bo'lib, bir qator model masalalarda 0.8-6.7% hatoliklar bo'lgan.

Neyro-noravshan yondashuvning adaptiv usuli (ANFIS) elektro ensefalogramma (EEG) signallarini klassifikatsiya qilishda qo'llanilishi yoritilgan. Bu yerda qaror qabul qilish ikki bosqichda amalga oshirilgan: birinchi bosqichda veyvlet-almashtirish (WT) ni qo'llash orqali belgilarni ajratish va ANFISni gradient usulni eng kam kvadratlar usuli bilan birgalikdagi to'g'ridan-to'g'ri o'qitish orqali amalga oshirilgan. EEGning beshta turdagi signallari ANFIS klassifikatorining beshta kirish ma'lumotlari sifatida olingan. Taklif etilgan usul kombinatsiyalangan adaptiv neyron tarmoq va noravshan mantiq yondashuvlari asosida shakllantirilgan. Olingan natijalar EEG signallarini klassifikatsiya qilishda 3-5% aniqlikni oshirish imkonini yaratgan. Unda gibridlashgan ANFIS usulidan foydalanilgan bo'lib, belgilarni kiruvchi EKG signallaridan ajratib olish va uni sinflarga ajratish orqali amalga oshirilgan. O'tkazilgan tajribaviy tadqiqotlar ushbu usulning qo'llanilishi 97% aniqlikka erishilganligini tadsdiqlagan.

Klassifikatsiya masalasini yechishda qo'llaniladigan usullardan yana biri noravshan o'rta usuli bo'lib, uning yordamida shahar muammolarini besh tipga ajratgan holda uni noravshan yondashuv orqali yechish taklif etilgan. Ushbu yondashuv boshqa usullarni qo'llanilishida muvofaqiyatsizlikka erishilganda, ya'ni ajratilgan beshta tipdagilarning aksariyatida qo'llanilgan usullar yaxshi natija bermaganda bu usul yuqori samaradorlikka erishilishini ko'rsatgan.

Hozirgi kunda noravshan yondashuvni qo'llanilishi sohasida kombinatsiyalangan ko'rinishdagi usullarni ishlab chiqilishi va qo'llanilishi muhim ahamiyat kasb etmoqda. Bunday yondashuvga genetik algoritmlar (GA)ning noravshan klasterizatsiya bilan, eng yaqin k -noravshan qo'shni va GAning kombinatsiyalangan ko'rinishini [2] klassifikatsiya masalasida qo'llanilishini keltirishimiz mumkin. Unda mos noravshan tizimni qurish uchun noravshan taqsimot GA orqali amalga oshirilishi keltirilgan bo'lib, unda kiruvchi ma'lumotlarni ajratish qo'pol taqsimot orqali amalga oshiriladi.

12. NORAVSHAN TO'PLAMLAR USULI

1-ta'rif. Noravshan to'plamlar.

X elementlari x orqali belgilangan mukammal to'plam bo'lsin. A to'plamda X dagi elementlarning tegishliligi X dan olingan $\{0,1\}$ dagi μ_A tavsifli funksiya sifatida qaraladi, ya'ni

$$\mu_A(x) = \begin{cases} 1, & \text{agar } x \in A, \\ 0, & \text{agar } x \notin A. \end{cases}$$

$\{0,1\}$ to'plam baholar to'plami deyiladi.

Agar $\{0,1\}$ to'plam emas, $[0,1]$ haqiqiy oraliq baholar to'plami deb olinsa, u holda A noravshan to'plam bo'ladi. Bunday holatda $\mu_A(x)$ tegishlilik funksiyasi deb ataladi. $\mu_A(x)$ 1 ga qanchalik yaqin bo'lsa, x shunchalik $A[1]$ ga tegishli bo'ladi.

A noravshan to'plam quyidagi juftliklar to'plami bilan xarakterlanishi mumkin:

$$A = \{(x, \mu_A(x)), x \in X\}$$

Umumiy holda, $A \subseteq X$ noravshan to'plam deb $(x, \mu_A(x))$ ko'rinishidagi juftliklar majmuiga aytiladi, bu yerda $x \in X, \mu_A(x): X \rightarrow [0,1]$ (ayrim hollarda $\mu_A(x): X \rightarrow L$ - to'r turidagi tuzilma).

Noravshan to'plam chekli va cheksiz bo'lishi mumkin. X-chekli to'plam, ya'ni $X = \{x_1, \dots, x_n\}$ bo'lsa, u holda X dagi noravshan to'plam quyidagi tarzda aniqlanadi:

$$A = \mu_A(x_1)/x_1 + \dots + \mu_A(x_n)/x_n = \sum_{i=1}^n \mu_A(x_i)/x_i.$$

X - cheksiz bo'lsa, u holda

$$A = \int \mu_A(x) / x$$

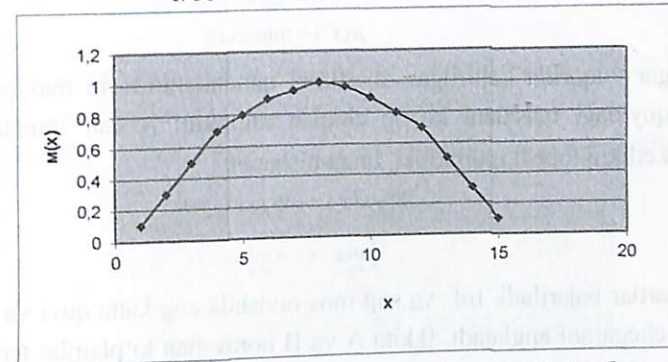
Misol. A noravshan to'plam quyidagi ko'rinishda berilgan bo'lsin:

$$A = 0.1/1 + 0.3/2 + 0.5/3 + 0.7/4 + 0.8/5 + 0.9/6 + 0.95/7 + 1.0/8 + 0.97/9 + 0.9/10 + 0.8/11 + 0.7/12 + 0.5/13 + 0.3/14 + 0.1/15.$$

U holda shu to'plamning klassik versiyasi quyidagi ko'rinishga ega bo'ladi:

$$A = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15\} \text{ yoki} \\ A = 1/1 + 1/2 + 1/3 + 1/4 + 1/5 + 1/6 + 1/7 + 1/8 + 1/9 +$$

$$+ 1/10 + 1/11 + 1/12 + 1/13 + 1/14 + 1/15.$$



Noravshan to'plamning grafik tasviri 12.1-rasmda keltirilgan.

Noravshan to'plam analitik shaklda tasvirlanishi mumkin. Misol, $A = \langle 8 \text{ ga yaqin haqiqiy son} \rangle$:

$$A = \int_{\mathbb{R}} \frac{1}{1+(x+8)^2} / x.$$

Berilgan turdagi noravshan to'plam, ya'ni uning tegishlilik funksiyasi oddiy (crisp) funksiya bo'lgan yoki tegishlilik darajasi oddiy sonlar bo'lgani 1-turdagi noravshan to'plam deyiladi. Noravshan to'plamning tegishlilik funksiyasi o'zi ham noravshan to'plam bo'lishi mumkin. U holda 2 turdagi noravshan to'plam tariqasida aniqlanadi

$$\mu: X \times [0,1] \rightarrow [0,1].$$

M turdagi noravshan to'plam deb, tegishlilik funksiyasining qiymatlari m-1 turdagi noravshan to'plam bo'lgan X dagi noravshan to'plamga aytiladi.

Noravshan to'plamlarning tegishlilik funksiyalarining qiymatlari tasodifiy o'zgaruvchi bo'lgan tur ko'rib chiqilsa, u holda X dagi A ehtimolli to'plam quyidagi xarakterli funksiya yordamida aniqlanadi

$$\mu_A: (x, \omega) \in X \times \Omega \rightarrow \mu(x, \omega) \in Q_c,$$

bu yerda $\mu_A(x^*)$ - har bir $x \in X$ uchun (B, B_c) -o'lchanadigan funksiyadir.

Noravshan to'plamlarning boshqa kengaytmalari ham mavjud. Agar $\forall x \in X, \exists! x^*$ ga nisbatan

$$\mu(x^*) \geq \mu(x), \\ \mu(x) \geq \mu(x^*)$$

shartlar bajarilsa, u holda x^* ni $\mu(x)$ tegishlilik funksiyasining maksimal va minimal qiymati deyiladi

$$\mu(x^*) = \max_{x \in X} \mu(x)$$

$$\mu(x^*) = \min_{x \in X} \mu(x)$$

Agar yuqorida keltirilgan shartlarni qanoatlantiruvchi mavjud bo'lsa, u holda quyidagi masalani ko'rib chiqish mumkin: X dan shunday x_1, x_2, \dots ketma-ketlikni topish mumkinki, unga nisbatan

$$\lim_{i \rightarrow \infty} \mu(x_i) = \sup_{x \in X} \mu(x), \text{ yoki}$$

$$\lim_{i \rightarrow \infty} \mu(x_i) = \inf_{x \in X} \mu(x)$$

shartlar bajariladi. Inf va sup mos ravishda eng katta quyi va eng kichik yuqori chegarani anglatadi. Ikkita A va B noravshan to'plamlar teng deyiladi, agar ularning tegishlilik funksiyasi teng, ya'ni

$$\forall x \in X, \mu_A(x) = \mu_B(x), A = B, \text{ bo'lsa.}$$

2-ta'rif. To'plam tashuvchisi, o'tish nuqtasi va singleton.

Noravshan to'plamning tashuvchisi $\mu_A(x) > 0$ ga ega bo'lgan x elementlardan iboratdir:

$$\text{supp } A = \{x \in X, \mu_A(x) > 0\}.$$

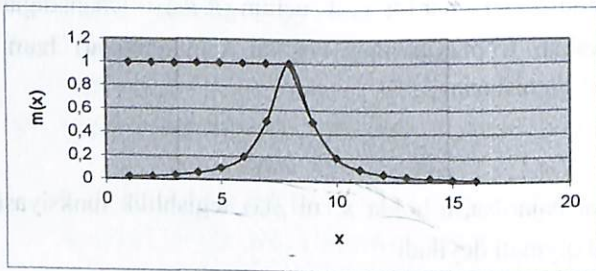
$\mu_A(x) = \frac{1}{2}$ bo'lgan $x \in X$ element A noravshan to'plamning o'tish nuqtasi deyiladi. Tashuvchisi X dan olingan $\mu_A = 1.0$ li bitta nuqta bo'lgan noravshan to'plam singleton deyiladi.

3-ta'rif. Noravshan to'plamning balandligi, normal noravshan to'plam.

A noravshan to'plamning balandligi deb tegishlilik funksiyasining eng yuqori chegarasiga aytiladi:

$$\text{hgt}(A) = \sup_{x \in X} \mu_A(x)$$

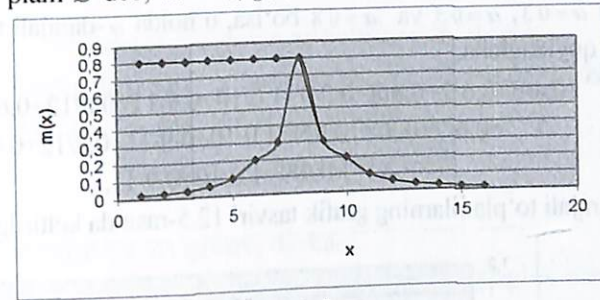
F noravshan to'plam, agar $\exists x \in X, \mu_A(x) = 1$ bo'lsa normal hisoblanadi. A normal noravshan to'plamning balandligi 1 ga teng (12.2-rasm), ya'ni $\text{hgt}(A) = 1$.



12.2-rasm. Normal noravshan to'plam

Agar $\max(\mu_A(x), \min(\mu_A(x), \mu_B(x))) = \mu_A(x) \Rightarrow A \cup (A \cap B) = A$. bo'lsa, u holda A noravshan to'plam subnormal deyiladi (12.3-rasm).

Bo'sh to'plam \emptyset deb, $\forall x \in X, \mu_{\emptyset}(x) = 0$ bo'lgan to'plamga aytiladi.



12.3-rasm. Subnormal noravshan to'plam

Misol. Noravshan to'plam berilgan (12.4-rasm).

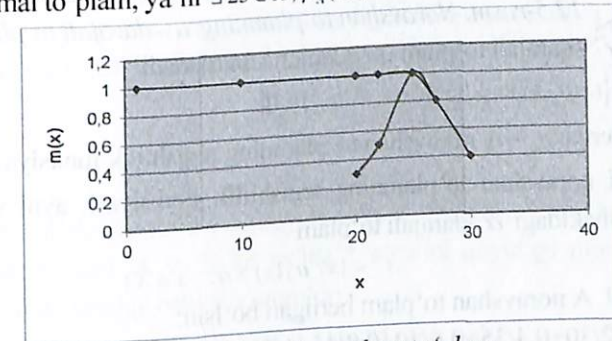
$$A = 0.3/20 + 0.5/22 + 1.0/25 + 0.8/27 + 0.4/30$$

Bu yerda

$$X = \{15, 20, 22, 25, 27, 30, 33, 35\} - \text{mukammal,}$$

$$\text{Supp } A = \{20, 22, 25, 27, 30\},$$

A - normal to'plam, ya'ni $\exists 25 \in X, \mu_A(25) = 1$.



12.4-rasm. Noravshan to'plam

4-ta'rif. α - darajali noravshan to'plam.

Tegishlilik qiymatlari ma'lum $\alpha \in [0, 1]$ ostonadan yuqori bo'lgan elementlarning oddiy to'plami A to'plamning α -kesimi A^α deyiladi:

$$A^\alpha = \{x \in X, \mu_A(x) \geq \alpha\}.$$

Qat'iy α -kesim

$$A^\alpha = \{x \in X, \mu_A(x) > \alpha\}$$

tariqasida aniqlanadi.

Misol. Noravshan to'plam berilgan bo'lsin.

$$A = 0.2/5 + 0.4/6 + 0.6/7 + 0.8/8 + 0.9/9 + 1.0/10 + 0.9/11 + 0.8/12 + 0.6/13 + 0.4/14 + 0.2/15.$$

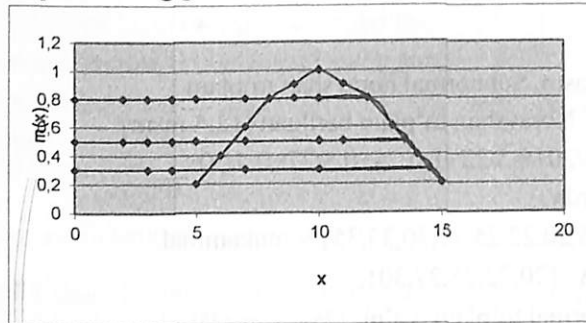
Agar $\alpha = 0.3$, $\alpha = 0.5$ va $\alpha = 0.8$ bo'lsa, u holda α -darajali to'plamlarning ko'rinishi quyidagicha:

$$A^{0.3} = 0.4/6 + 0.6/7 + 0.8/8 + 0.9/9 + 1.0/10 + 0.9/11 + 0.8/12 + 0.6/13 + 0.4/14,$$

$$A^{0.5} = 0.6/7 + 0.8/8 + 0.9/9 + 1.0/10 + 0.9/11 + 0.8/12 + 0.6/13,$$

$$A^{0.8} = 0.9/9 + 1.0/10 + 0.9/11.$$

α -darajali to'plamlarning grafik tasviri 12.5-rasmda keltirilgan.



12.5-rasm. Noravshan to'plamning α -darajali to'plamlari

A^α α -darajali to'plam quyidagicha aniqlanadi

$$A^\alpha = \{(x, \mu_{A^\alpha}(x) = \mu_A(x)) / x \in A^\alpha, \alpha \in [0, 1]\}$$

Bu yerda μ_A - A noravshan to'plamning tegishlilik funksiyasi va μ_{A^α} - A^α α -darajali noravshan to'plamning tegishlilik funksiyasi, ayni vaqtda A^α - quyidagi shakldagi α -darajali to'plam

$$A^\alpha = \{x / \mu_A(x) \geq \alpha, x \in X\}$$

Misol. A noravshan to'plam berilgan bo'lsin:

$$A = 0.2/30 + 0.4/35 + 0.6/40 + 0.8/45 + 1/50 + 0.7/55.$$

$A^{\alpha=0.5}$ 0.5-darajali to'plam deb, quyidagi to'plamga aytiladi

$$A^{\alpha=0.5} = 0.6/40 + 0.8/45 + 1/50 + 0.7/55.$$

12.1. Noravshan sonlar va noravshan sonlar ustida amallar

Avvalo noravshan sonlarning asosiy ta'riflarini va ular ustidagi amallarni ko'rib chiqamiz [1-4].

5-ta'rif. Noravshan son. Haqiqiy to'g'ri chiziqdagi noravshan A soni-

$\mu_A: R \rightarrow [0, 1]$ tegishlilik funksiyasi bilan xarakterlanuvchi noravshan to'plamdir. A noravshan son quyidagicha aniqlanishi mumkin

$$A = \int_R \mu_A(x) / x, \quad (12.1)$$

Bu yerda \int - barcha $x \in R$ bo'yicha birlashtirish ishorasi.

6-ta'rif. Noravshan sonning qavariqligi. Haqiqiy to'g'ri chiziqdagi A noravshan son haqiqiy $x, y, z \in R$ sonlarga nisbatan $x \leq y \leq z$ da

$$\mu_A(y) \geq \min(\mu_A(x), \mu_A(z)) \quad (12.2)$$

Munosabat bajarilganda qavariq deyiladi.

7-ta'rif. Noravshan sonning normalligi. haqiqiy o'qdagagi A noravshan son normal deyiladi, agar $\max_x \mu_A(x) = 1$ bo'lsa.

8-ta'rif. Noravshan sonning chegaralari. Agar a songa nisbatan quyidagi munosabat bajarilsa

$$\forall \delta \mu_A = 0; \quad \mu(a - \delta) \neq 0, \mu(a + \delta) \neq 0, \quad (12.3)$$

U holda u tegishlilik funksiyaning chegarasi deyiladi. Agar bunday chegaralar ikkita: yuqori (b) va quyi(a) ekanligini hisobga olsak, A noravshan sonni quyidagi ko'rinishda yozib olish mumkin

$$A = \int_a^{\bar{a}} (x - a) / x + \int_{\bar{a}}^b (b - x) / x. \quad (12.4)$$

Umumlashtirish tamoyili. Haqiqiy R to'g'ri chiziqda A va B noravshan sonlar berilgan bo'lsin. A va B lar ustida * amalini quyidagi munosabatdan foydalangan holda amalga oshirish mumkin

$$A * B = \int_R \min(\mu_A(x), \mu_B(y)) / (x * y). \quad (12.5)$$

* gipotetik amalining o'rtinga arifmetik +, -, *, : lardan foydalangan holda A va B ustidagi to'rtta arifmetik amalni hosil qilib olish mumkin:

$$A + B = \int_R \min(\mu_A(x), \mu_B(y)) / (x + y). \quad (12.6)$$

$$A - B = \int_R \min(\mu_A(x), \mu_B(y)) / (x - y). \quad (12.7)$$

$$A \times B = \int_R \min(\mu_A(x), \mu_B(y)) / (x \times y). \quad (12.8)$$

$$A : B = \int_R \min(\mu_A(x), \mu_B(y)) / (x : y). \quad (12.9)$$

(7.4) dan foydalangan holda quyidagini hosil qilib olish mumkin

$$\begin{aligned} A * B &= \left(\int_a^{\bar{a}} \mu_A(x) / x + \int_{\bar{a}}^b \mu_A(x) / x \right) * \left(\int_a^{\bar{b}} \mu_B(x) / x + \int_{\bar{b}}^b \mu_B(x) / x \right) = \\ &= \int_a^{\bar{a} * \bar{b}} \mu_{A * B}(x) / x + \int_{\bar{a} * \bar{b}}^b \mu_{A * B}(x) / x, \end{aligned} \quad (12.10)$$

Bu yerda a'', b'' a, b dan hosil qilib olinadi a', b' esa ma'lum amalga qarab hosil qilinadi, $\mu_{A * B}(x)$ amalga qarab va μ ning normallashtiruviga qarab aniqlanadi. $A + V$ ni hisoblaymiz

$$\begin{aligned} A + B &= \left(\int_a^{\bar{a}} \mu_A(x) / x + \int_{\bar{a}}^b \mu_A(x) / x \right) + \left(\int_a^{\bar{b}} \mu_B(x) / x + \int_{\bar{b}}^b \mu_B(x) / x \right) = \\ &= \int_a^{\bar{c}} \mu_C(x) / x + \int_{\bar{c}}^b \mu_C(x) / x = C, \end{aligned} \quad (12.11)$$

Bu yerda

$$\bar{c} = \bar{a} + \bar{b}, \quad a'' = a + a', \quad b'' = b + b' \quad (12.12)$$

μ_C $\mu_C = k_1 x + k_2$ ko'rinishda aniqlanadi. Normallashtirishdan kelib chiqqan holda $a'' \leq x \leq \bar{c}$ ga nisbatan [4] ni quyidagicha yozib olish mumkin

$$A + B = \int_a^{\bar{c}} \frac{x - a''}{\bar{c} - a''} / x + \int_{\bar{c}}^b \frac{b'' - x}{b'' - \bar{c}} / x = C \quad (12.13)$$

Qolgan arifmetik amallar uchun shunga o'xshash usulda quyidagilarni hosil qilish mumkin [4]

$$A - B = \int_a^{\bar{c}} \frac{x - a''}{\bar{c} - a''} / x + \int_{\bar{c}}^b \frac{b'' - x}{b'' - \bar{c}} / x = C \quad (12.14)$$

Bu yerda

$$a'' = a - b', \quad b'' = b - a', \quad \bar{c} = \bar{a} - \bar{b} \quad (12.15)$$

Tegishlilik funksiyasini $\mu_c = k_1 \sqrt{x} + k_2$ ko'rinishda qabul qilib, quyidagiga ega bo'lamiz

$$A * B = \int_a^{\bar{c}} \frac{\sqrt{x} - \sqrt{a''}}{\sqrt{\bar{c}} - \sqrt{a''}} / x + \int_{\bar{c}}^b \frac{\sqrt{b''} - \sqrt{x}}{\sqrt{b''} - \sqrt{\bar{c}}} / x = C \quad (12.16)$$

Bu yerda

$$a'' = a * a', \quad b'' = b * b', \quad \bar{c} = \bar{a} * \bar{b} \quad (12.17)$$

μ_C tegishlilik funksiyasini $\mu_c = \frac{k_1}{x} + k_2$ ko'rinishda qabul qilib, quyidagiga ega bo'lamiz

$$A : B = \int_a^{\bar{c}} \frac{(x - a'')\bar{c}}{(\bar{c} - a'')x} / x + \int_{\bar{c}}^b \frac{(b'' - x)\bar{c}}{(b'' - \bar{c})x} / x = C \quad (12.18)$$

Bu yerda

$$a'' = a' : a, \quad b'' = b' : b, \quad \bar{c} = \bar{b} : \bar{a} \quad (12.19)$$

Misollar. $A =$ taxminan $6 = \bar{6}$; $V =$ taxminan $8 = \bar{8}$.

$$\bar{6} = \int_5^6 (x-5) / x + \int_6^7 (7-x) / x,$$

$$x=5 \text{ da: } \bar{6}|_{x=5} = \int_5^6 (x-5)_{x=5} = 5-5=0;$$

$$x=5,5 \text{ da: } \bar{6}|_{x=5,5} = \int_5^6 (x-5)_{x=5,5} = 5,5-5=0,5;$$

$$x=6 \text{ da: } \bar{6}|_{x=6} = \int_5^6 (x-5)_{x=6} = 6-5=1;$$

$$x=6,5 \text{ da: } \tilde{6}\big|_{x=6,5} = \int_6^7 (7-x)_{x=6,5} = 7-6,5=0,5;$$

$$x=7 \text{ da: } \tilde{6}\big|_{x=7} = \int_6^7 (7-x)_{x=7} = 7-7=0.$$

Demak $\tilde{6} = \{0/5; 0,5/5,5; 1/6; 0,5/6,5; 0/7\}$.

$\tilde{8}$ uchun $x=6, 7, 8, 9$ va 10 da shunga o'xshash usulda quyidagiga ega bo'lish mumkin

$$\tilde{8} = \{0/6; 0,5/7; 1/8; 0,5/9; 0/10\}$$

Noravshan sonlarning grafiklari 12.6-rasmda keltirilgan. Quyi va yuqori chegaralar hamda ushbu sonlarning balandliklari quyidagicha: $\tilde{6}$ uchun: $a=5, b=7, \bar{a}=6$; $\tilde{8}$ uchun: $a'=6, b'=10, \bar{b}=8$. Quyida $\tilde{6}$ va $\tilde{8}$ ustidagi to'rtta amal keltirilgan.

Qo'shish. Avvalo (12.12) ga ko'ra ($\tilde{6}+\tilde{8}$) yig'indining chegaralari va balandligini hisoblaymiz:

$$a'' = a + a' = 5 + 6 = 11; \quad b'' = b + b' = 7 + 10 = 17; \quad \bar{c} = \bar{a} + \bar{b} = 6 + 8 = 14$$

(12.13) ni hisobga olgan holda quyidagicha yozib olish mumkin

$$\tilde{6} + \tilde{8} = \int_{11}^{14} \frac{x-11}{3} / x + \int_{14}^{17} \frac{17-x}{3} / x = \tilde{14}$$

$\tilde{14}$ ni har xil x larda hisoblab: $x=12,5; x=15,5$ $\tilde{14} = \{0/11; 0,5/12,5; 1/14; 0,5/15; 0/17\}$ larni hosil qilib olish mumkin, ular 7.1-rasmda grafik tasvirlangan.

Ayirish. $\tilde{8}-\tilde{6}$ ayirmaning chegaralari va balandligi (12.15) ga binoan quyidagicha aniqlanadi:

$$a'' = a' - a = 6 - 5 = 1; \quad b'' = b' - b = 10 - 7 = 3; \quad \bar{c} = \bar{b} - \bar{a} = 8 - 6 = 2.$$

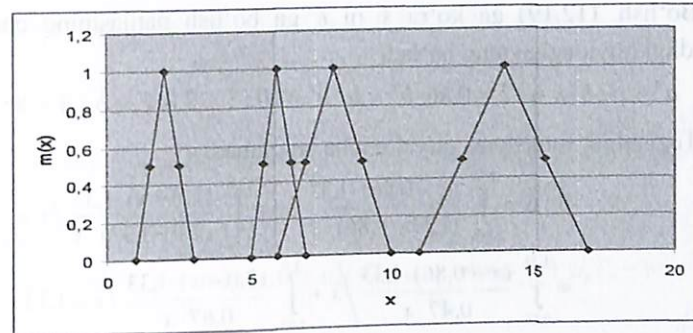
(12.15) ni hisobga olib, quyidagilarni hosil qilish mumkin

$$\tilde{8} - \tilde{6} = \int_1^2 \frac{x-1}{3-2} / x + \int_2^3 \frac{3-x}{3-2} / x = \tilde{2}$$

Tegishlilik funksiyasining $x=1,5$ va $x=2,5$ dagi qiymatlari mos ravishda $0,5$ va $0,5$ ga teng bo'ladi. U holda

$$\tilde{2} = \{0/1; 0,5/1,5; 1/2; 0,5/2,5; 0/3\}$$

U 12.6-rasmda grafik tasvirlangan.



12.6. Noravshan sonlar.

Taxminan 2. Taxminan 6.

Taxminan 8. Taxminan 14.

Ko'paytirish. (12.17) munosabat orqali

$\tilde{6} \times \tilde{8} : a'' = a \times a' = 5 \times 6 = 30; \quad b'' = b \times b' = 7 \times 10 = 70; \quad \bar{c} = \bar{a} \times \bar{b} = 6 \times 8 = 48$
cheagara va balandlikni aniqlaymiz. (12.16) ni hisobga olib $\mu_{\tilde{6}, \tilde{8}}$ tegishlilik

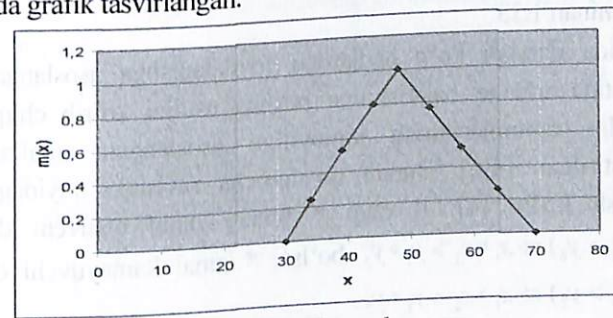
funksiyasini quyidagicha tasvirlash mumkin

$$\tilde{6} \times \tilde{8} = \int_{30}^{48} \frac{\sqrt{x}-\sqrt{30}}{\sqrt{48}-\sqrt{30}} / x + \int_{48}^{70} \frac{\sqrt{70}-\sqrt{x}}{\sqrt{70}-\sqrt{48}} / x =$$

$$= \int_{30}^{48} \frac{\sqrt{x}-5,48}{1,45} / x + \int_{48}^{70} \frac{8,37-\sqrt{x}}{1,44} / x = 4\tilde{8}$$

Tegishlilik funksiyasining $x=34; x=39; x=53; x=58$ nuqtalardagi qiymati mos ravishda $0,24; 0,53; 0,76; 0,52$ ga teng bo'ladi. Shunday qilib

$\tilde{6} \times \tilde{8} = 4\tilde{8} = \{0/30; 0,24/34; 0,53/39; 0,8/44; 1/48; 0,76/53; 0,52/58; 0,26/64; 0/70\}$
U 12.7-rasmda grafik tasvirlangan.



12.7-rasm. Noravshan son

Taxminan 48.

Bo'lish. (12.19) ga ko'ra $\tilde{8}$ ni $\tilde{6}$ ga bo'lish natijasining chegarasi va balandligi quyidagiga teng bo'ladi

$$a'' = a : b' = 6 : 7 = 0,86; b'' = b : a' = 10 : 5 = 2,0; \bar{c} = \bar{b} : \bar{a} = 8 : 6 = 1,33.$$

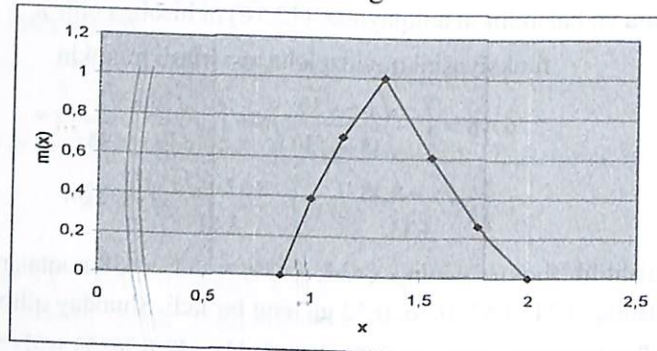
Tegishlilik funksiyasi quyidagicha aniqlanadi

$$\begin{aligned} \tilde{8} : \tilde{6} &= \int_{0,86}^{1,33} \frac{(x-0,86) \cdot 1,33}{(1,33-0,86) \cdot x} / x + \int_{1,33}^{2,0} \frac{(2,0-x) \cdot 1,33}{(2,0-1,33) \cdot x} / x = \\ &= \int_{0,86}^{1,33} \frac{(x-0,86) \cdot 1,33}{0,47 \cdot x} / x + \int_{1,33}^{2,0} \frac{(2,0-x) \cdot 1,33}{0,67 \cdot x} / x = 1,33. \end{aligned}$$

Tegishlilik funksiyasining $x=1,0$; $x=1,14$; $x=1,55$; $x=1,77$ nuqtalardagi qiymatlarini hisoblab va mos ravishda $0,39$; $0,70$; $0,60$; $0,26$ natijalarni olib, quyidagicha yozib olish mumkin

$$\tilde{8} : \tilde{6} = \{0/0,86; 0,39/1,0; 0,70/1,14; 1/1,33; 0,60/1,55; 0,26/1,77; 0/2,0\}.$$

U grafik ravishda 12.8-rasmda tasvirlangan.



12.8-rasm. Noravshan son

Taxminan 1,33

Quyida darajali ko'p hadlardan foydalanishga asoslangan noravshan sonlar ustida amallar bajarishning boshqa usulini ko'rib chiqamiz, undagi hisoblashlar umumlashtirish tamoyiliga asoslangan amallarga nisbatan soddalashtirilgan [4-6]. Bunda qo'shimcha ravishda quyidagi ta'riflardan foydalanish lozim [4]: R dagi * binary amal o'suvchi deyiladi, agar $(x_1 > y_1, x_2 > y_2) \Rightarrow x_1 * x_2 > y_1 * y_2$ bo'lsa. * amal kamayuvchi deyiladi, agar $(x_1 > y_1, x_2 > y_2) \Rightarrow x_1 * x_2 < y_1 * y_2$.

Agar μ_A va μ_B tegishlilik funksiyali noravshan A va B sonlar berilgan bo'lsa, u holda ular ustidagi umumlashgan * amalning natijasi quyidagi

tegishlilik funksiyasi orqali berilgan $C = A * B$ noravshan sondir

$$\mu_C(z) = \sup_{Z=X*Y} \min(\mu_A(x), \mu_B(y)) \quad (12.20)$$

Aniqroq qilib aytganda, to'rtta arifmetik amalni quyidagicha tasvirlash mumkin:

Qo'shish.

$$\mu_{A+B}(z) = \sup_{Z=X+Y} \min(\mu_A(x), \mu_B(y)) = \sup_X \min(\mu_A(x), \mu_B(z-x)) \quad (12.21)$$

Ayirish.

$$\mu_{A-B}(z) = \sup_{Z=X-Y} \min(\mu_A(x), \mu_B(y)) = \sup_X \min(\mu_A(x), \mu_B(x-z)) \quad (12.22)$$

Ko'paytirish.

$$\mu_{A \cdot B}(z) = \sup_{Z=X \cdot Y} \min(\mu_A(x), \mu_B(y)) = \sup_X \min(\mu_A(x), \mu_B(z/x)), x \neq 0 \quad (12.23)$$

Bo'lish.

$$\begin{aligned} \mu_{A/B}(z) &= \sup_{Z=X/Y} \min(\mu_A(x), \mu_B(y)) = \sup_X \min(\mu_A(x), \mu_B(x/z)) = \\ &= \sup_Y \min(\mu_A(yz), \mu_B(y)). \end{aligned} \quad (12.24)$$

Agar A va B noravshan sonlar quyidagicha tasvirlansa

$$A = \{\omega_1 / x_{11}; \omega_2 / x_{21}; \omega_1 / x_{12}\}; \quad B = \{\omega_1 / y_{11}; \omega_2 / y_{21}; \omega_1 / y_{12}\},$$

U holda ular ustidagi * umumlashgan amalning natijasi quyidagi noravshan son bo'ladi

$$C = A * B = \{\omega_1 / (x_{11} * y_{11}); \omega_2 / (x_{21} * y_{21}); \omega_1 / (x_{12} * y_{12})\} \quad (12.25)$$

Bu * amal o'suvchi yoki kamayuvchi bo'lganida o'rinlidir. Ayirish va bo'lish amallari bunday emas, lekin ularni quyidagicha tasvirlash mumkin [4]:

$$A - B = A + (-B); \quad A : B = A \times (B^{-1}) \quad (12.26)$$

Misollar. Ikkita noravshan son berilgan

$$\tilde{2} = \{0/1; 0,5/1,5; 1/2; 0,5/2,5; 0/3\},$$

$$\tilde{3} = \{0/2; 0,5/2,5; 1/3; 0,5/3,5; 0/4\}$$

Quyida ular ustida to'rtta umumlashgan amal bajariladi (+, -, x, :).
Qo'shish.

$$\tilde{3} + \tilde{2} = \{0/(2+1); 0,5/(2,5+1,5); 1/(3+2); 0,5/(3,5+2,5); 0/(4+3)\} = \{0/3; 0,5/4; 1/5; 0,5/6; 0/7\}$$

Ko'paytirish.

$$\tilde{3} \times \tilde{2} = \{0/2; 0,5/3,75; 1/6; 0,5/8,75; 0/12\}$$

Ayirish.

$$-\tilde{2} = \{0/(-3); 0,5/(-2,5); 1/(-2); 0,5/(-1,5); 0/(-1)\};$$

$$\tilde{3} - \tilde{2} = \tilde{3} + (-\tilde{2}) = \{0/(2-3); 0,5/(2,5-2,5); 1/(3-2); 0,5/(3,5-1,5); 0/(4-1)\} = \{0/(-1); 0,5/0; 1/1; 0,5/2; 0/3\}$$

Bo'lish.

$$\tilde{2}^{-1} = \{0/(1:1); 0,5/(1:1,5); 1/(1:2); 0,5/(1:2,5); 0/(1:3)\} = \{0/1; 0,5/0,66; 1/0,5; 0,5/0,4; 0/0,33\} = \{0/0,33; 0,5/0,4; 1/0,5; 0,5/0,66; 0/1\};$$

$$\tilde{3} : \tilde{2} = 3 \times (\tilde{2}^{-1}) = \{0/0,66; 0,5/1; 1/1,5; 0,5/2,33; 0/4\}$$

Qo'shimcha ayirish va bo'lish amallari. Noravshan tenglamalarni yechishda qarama-qarshi va teskari sonlarni hisoblash kerak bo'ladi [4]. Yuqorida ko'rib chiqilgan arifmetik amallar, umumlashtirish tamoyiliga asoslangan bo'lib, qarama-qarshi A' ($A+A'=0$ bo'ladigan) va teskari sonni

A ($A \times A'=1$) topishga imkon bermaydi. Shuningdek, quyidagi tengsizliklar o'rinlidir.

$$(A-B)+B \neq A; (A:B) \times B \neq A$$

Quyidagi tenglikni aniq yechish uchun

$$AX+B=D$$

(12.27)

Bu yerda A, B, D – noravshan sonlar, X -noma'lum, qo'shimcha ayirish (-) va qo'shimcha bo'lish (/) amalidan foydalaniladi.

Xususiyl holda (7.27) ning yechimi quyidagi ko'rinishda bo'ladi

$$X=D--B$$

(12.28)

B va D to'plamning tashuvchilari mos ravishda $S_B = [b_1, b_2]$ va $S_D = [d_1, d_2]$ oraliqlardir. Qo'shimcha ayirish yordamida aniqlanuvchi X to'plamning tashuvchisi quyidagi ko'rinishda bo'ladi.

$$S_X = [d_1 - b_1, d_2 - b_2]$$

(12.29)

Tegishlilik funksiyasi yordamida ifodalangan ko'rinishi esa [4]

$$\mu_X(x) = \inf_z \begin{cases} 1, & \text{agar } \mu_D(z-x) < \mu_B(z); \\ \mu_D, & \text{agar } \mu_B(z-x) \geq \mu_D(z). \end{cases} \quad (12.30)$$

Ko'rib chiqilayotgan ayirish amali kamayuvchining tashuvchisi uzunligi ayiriluvchilikidan kichik bo'lgandagina aniqlangan.

Qo'shimcha bo'lish. $AX=B$ tenglamaning yechimi $X=D//A$ to'plam bo'ladi. Agar A va D to'plamning tashuvchilari $S_A = [a_1, a_2]$ va $S_D = [d_1, d_2]$ bo'lsa, u holda X to'plamning tashuvchisi quyidagicha aniqlanadi.

$$S_X = [d_1, d_2] // [a_1, a_2] = \begin{cases} d_1 : a_1, d_2 : a_2, & \text{agar } S_A > 0; S_D > 0 \\ d_1 : a_2, d_2 : a_1, & \text{agar } S_A > 0; S_D < 0 \\ d_2 : a_1, d_1 : a_2, & \text{agar } S_A < 0; S_D > 0 \\ d_2 : a_2, d_1 : a_1, & \text{agar } S_A < 0; S_D < 0 \end{cases} [4]$$

Yoki uning tegishlilik funksiyasi orqali ifodalangan ko'rinishi:

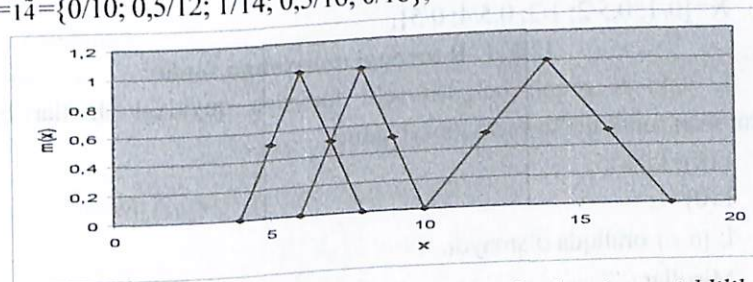
$$\mu_X(x) = \inf_t \begin{cases} 1, & \text{agar } \mu_A(t/x) < \mu_D(t); \\ \mu_D(t), & \text{agar } \mu_A(t/x) \geq \mu_D(t). \end{cases}$$

Bu amal ixtiyoriy A va D sonlarning ixtiyoriysi uchun aniqlanmagan, u oraliq tashuvchilari ma'lum shartlarni qanoatlantiruvchi sonlarga nisbatan aniqlangandir [4].

Misollar. Quyidagi tenglamani yeching
 $X+B=D$

(12.33)

bunda $B=\tilde{8}=\{0/6; 0,5/7; 1/8; 0,5/9; 0/10\}$,
 $D=1\tilde{4}=\{0/10; 0,5/12; 1/14; 0,5/16; 0/18\}$,



12.9-rasm. Qo'shimcha ayirish amali uchun to'plamlarning tegishlilik

$$\tilde{3} = \{0/2; 0,5/2,5; 1/3; 0,5/3,5; 0/4\}$$

Quyida ular ustida to'rtta umumlashgan amal bajariladi (+, -, x, :).

Qo'shish.

$$\tilde{3} + \tilde{2} = \{0/(2+1); 0,5/(2,5+1,5); 1/(3+2); 0,5/(3,5+2,5); 0/(4+3)\} = \{0/3; 0,5/4; 1/5; 0,5/6; 0/7\}.$$

Ko'paytirish.

$$\tilde{3} \times \tilde{2} = \{0/2; 0,5/3,75; 1/6; 0,5/8,75; 0/12\}.$$

Ayirish.

$$-\tilde{2} = \{0/(-3); 0,5/(-2,5); 1/(-2); 0,5/(-1,5); 0/(-1)\};$$

$$\tilde{3} - \tilde{2} = \tilde{3} + (-\tilde{2}) = \{0/(2-3); 0,5/(2,5-2,5); 1/(3-2); 0,5/(3,5-1,5); 0/(4-1)\} = \{0/(-1); 0,5/0; 1/1; 0,5/2; 0/3\}.$$

Bo'lish.

$$\tilde{2}^{-1} = \{0/(1:1); 0,5/(1:1,5); 1/(1:2); 0,5/(1:2,5); 0/(1:3)\} =$$

$$= \{0/1; 0,5/0,66; 1/0,5; 0,5/0,4; 0/0,33\} = \{0/0,33; 0,5/0,4; 1/0,5; 0,5/0,66; 0/1\};$$

$$\tilde{3} : \tilde{2} = 3 \times (\tilde{2}^{-1}) = \{0/0,66; 0,5/1; 1/1,5; 0,5/2,33; 0/4\}.$$

Qo'shimcha ayirish va bo'lish amallari. Noravshan tenglamalarni yechishda qarama-qarshi va teskari sonlarni hisoblash kerak bo'ladi [4]. Yuqorida ko'rib chiqilgan arifmetik amallar, umumlashtirish tamoyiliga asoslangan bo'lib, qarama-qarshi A' ($A + A' = 0$ bo'ladigan) va teskari sonni

A ($A \times A' = 1$) topishga imkon bermaydi. Shuningdek, quyidagi tengsizliklar o'rinalidir.

$$(A - B) + B \neq A; (A : B) \times B \neq A.$$

Quyidagi tenglikni aniq yechish uchun

$$AX + B = D \quad (12.27)$$

Bu yerda A, B, D – noravshan sonlar, X – noma'lum, qo'shimcha ayirish (-) va qo'shimcha bo'lish (/) amalidan foydalaniladi.

Xususiyl holda (7.27) ning yechimi quyidagi ko'rinishda bo'ladi

$$X = D \cdot -B \quad (12.28)$$

B va D to'planning tashuvchilari mos ravishda $S_B = [b_1, b_2]$ va $S_D = [d_1, d_2]$ oraliqlardir. Qo'shimcha ayirish yordamida aniqlanuvchi X to'planning tashuvchisi quyidagi ko'rinishda bo'ladi.

$$S_X = [d_1 - b_1, d_2 - b_2]$$

(12.29)

Tegishlilik funksiyasi yordamida ifodalangan ko'rinishi esa [4]

$$\mu_X(x) = \inf_z \begin{cases} 1, & \text{agar } \mu_D(z-x) < \mu_B(z); \\ \mu_D, & \text{agar } \mu_B(z-x) \geq \mu_D(z). \end{cases} \quad (12.30)$$

Ko'rib chiqilayotgan ayirish amali kamayuvchining tashuvchisi uzunligi ayiriluvchilikidan kichik bo'lgandagina aniqlangan.

Qo'shimcha bo'lish. $AX=B$ tenglamaning yechimi $X=D//A$ to'plam bo'ladi. Agar A va D to'planning tashuvchilari $S_A = [a_1, a_2]$ va $S_D = [d_1, d_2]$ bo'lsa, u holda X to'planning tashuvchisi quyidagicha aniqlanadi.

$$S_X = [d_1, d_2] // [a_1, a_2] = \begin{cases} d_1 : a_1, d_2 : a_2, & \text{agar } S_A > 0; S_D > 0 \\ d_1 : a_2, d_2 : a_1, & \text{agar } S_A > 0; S_D < 0 \\ d_2 : a_1, d_1 : a_2, & \text{agar } S_A < 0; S_D > 0 \\ d_2 : a_2, d_1 : a_1, & \text{agar } S_A < 0; S_D < 0 \end{cases} [4]$$

Yoki uning tegishlilik funksiyasi orqali ifodalangan ko'rinishi:

$$\mu_X(x) = \inf_t \begin{cases} 1, & \text{agar } \mu_A(t/x) < \mu_D(t); \\ \mu_D(t), & \text{agar } \mu_A(t/x) \geq \mu_D(t). \end{cases}$$

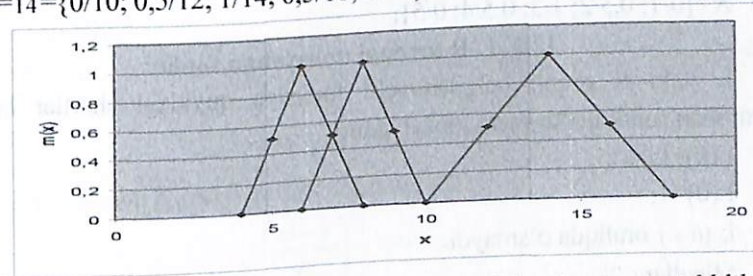
Bu amal ixtiyoriy A va D sonlarning ixtiyoriysi uchun aniqlanmagan, u oraliq tashuvchilari ma'lum shartlarni qanoatlantiruvchi sonlarga nisbatan aniqlangandir [4].

Misollar. Quyidagi tenglamani yeching

$$X + B = D \quad (12.33)$$

bunda $B = \tilde{8} = \{0/6; 0,5/7; 1/8; 0,5/9; 0/10\}$,

$D = 1\tilde{4} = \{0/10; 0,5/12; 1/14; 0,5/16; 0/18\}$,



12.9-rasm. Qo'shimcha ayirish amali uchun to'plamlarning tegishlilik

funksiyasi.

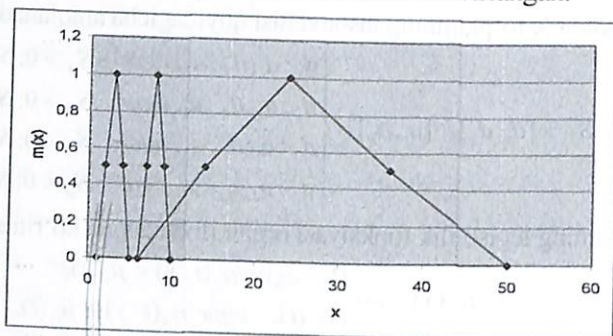
B va D uchun oraliq-tashuvchilar $S_B = [6, 10]$; $S_D = [10, 18]$. (7.29) ga ko'ra $S_X = [4, 8]$. (7.30) formulaga ko'ra 12.9-rasmda grafik usulda tasvirlangan $\mu_X(x)$ tegishlilik funksiyasini aniqlash mumkin.

Mos ravishda, $X = (0/4; 0,5/5; 1,0/6; 0,5/7; 0/8)$.

$A = \bar{8} = \{0/6; 0,5/7; 1/8; 0,5/9; 0/10\}$ va $D = 2\bar{4} = \{0/6; 0,5/14; 1,0/24; 0,5/36; 0/50\}$ da quyidagi tenglamani yeching

$$AX = D \quad (12.34)$$

μ_A va μ_D tegishlilik funksiyalari 12.10-rasmda tasvirlangan.



12.10-rasm. Qo'shimcha bo'lish amali uchun to'plamlarning tegishlilik funksiyalari

A va D to'plamlarning oraliq-tashuvchilari $S_A = [6, 10]$; $S_D = [6, 50]$. (12.31) ga ko'ra $S_X = [6; 6, 50; 10] = [1, 5]$.

(12.32) ga ko'ra, 12.10-rasmda keltirilgan $\mu_X(x)$ tegishlilik funksiyasining qiymantini aniqlash mumkin.

Tenglamani yechimi

$$X = \{0/1; 0,5/2; 1/3; 0,5/4; 0/5\}.$$

12.2. L-R turdagi noravshan sonlar

L yoki R orqali belgilanuvchi funksiya quyidagi shartlar bajarilsa, noravshan sonni ifodalovchi funksiyadir:

$$L(x) = L(-x);$$

$$L(0) = 1;$$

L $[0, \infty)$ oraliqda o'smaydi.

Misollar.

$$a) L(x) = \begin{cases} 1, & x \in [-1, 1]; \\ 0, & \text{aks holda.} \end{cases}$$

$$b) L(x) = \max(0; 1 - |x|^p), \quad p \geq 0.$$

$$v) L(x) = e^{-|x|^p}, \quad p > 0.$$

$$g) L(x) = 1 / (1 + |x|^p), \quad p \geq 0.$$

M noravshan son L-R turda deyiladi, agar:

$$\mu_M(x) = \begin{cases} L\left(\frac{m-x}{\alpha}\right) & x \leq m \text{ da, } \alpha > 0; \\ R\left(\frac{x-m}{\beta}\right) & x \geq m \text{ da, } \beta > 0. \end{cases}$$

L chap ifoda deyiladi; R-o'ng ifoda; m-M ning o'rta qiymati; α va β mos ravishda chap va o'ng kengaytma deyiladi. Kengaytmalar nolga teng bo'lganida, M sodda son bo'ladi. Kengaytmani oshirib borgan sari, M shunchalik noravshanlashib boradi. L-R turdagi sonning belgisi yozuvi

$$M = (m, \alpha, \beta)_{LR}.$$

L-R turdagi sonlar ustida algebraik amallarni ko'rib chiqamiz.

1-teorema. [6] M va N - uzluksiz tegishlilik funksiyali ikkita noravshan son; * - uzluksiz o'suvchi binary amal; $[\lambda_M, \rho_M]$ - M noravshan sonning tegishlilik funksiyasi kamaymaydigan ma'lum bir oraliq ($\lambda_m = \rho_m$); $[\lambda_N, \rho_N]$ - N ga nisbatan o'xshash oraliq, jumladan

$$\mu_M(x) = \mu_N(y) = \omega \text{ uchun } \forall x \in [\lambda_M, \rho_M], \forall y \in [\lambda_N, \rho_N].$$

Tegishlilik funksiyasi umumlashtirish tamoyili bo'yicha aniqlangan $M * N$ noravshan to'plam bo'lsa, undagi uzluksiz tegishlilik funksiyali noravshan son ixtiyoriy $t \in [\lambda_{M*N}, \rho_{M*N}]$ nuqtada $\mu_{M*N}(t) = \omega$ formula bo'yicha aniqlanadi.

Noravshan sonlarni qo'shish. Ikkita noravshan sonning o'suvchi qismlarini ko'rib chiqamiz

$$M = (m, \alpha, \beta)_{LR}; \quad N = (n, \gamma, \delta)_{LR}.$$

x va y -yagona haqiqiy sonlar bo'lib, bunda

$$L\left(\frac{m-x}{\alpha}\right) = \omega = L\left(\frac{n-y}{\gamma}\right),$$

Bu yerda $\omega \in [0,1]$ kesmadagi qo'zg'almas nuqta. Bu quyidagiga ekvivalent

$$x = m - \alpha L^{-1}(\omega), \quad y = n - \gamma L^{-1}(\omega) \quad (12.35)$$

(12.35) dan quyidagilar kelib chiqadi

$$z = x + y = m + n - (\alpha + \gamma)L^{-1}(\omega) \quad \text{va} \quad L\left(\frac{m+n-z}{\alpha+\gamma}\right) = \omega.$$

Huddi shunday, M va N ning kamyuvchi qismlariga nisbatan quyidagilarga ega bo'lamiz

$$R\left(\frac{z - (m+n)}{\beta + \delta}\right) = \omega$$

Bu yerdan toremaga ko'ra quyidagilar kelib chiqadi

$$(m, \alpha, \beta)_{LR} \oplus (n, \gamma, \delta)_{LR} = (m+n, \alpha + \gamma, \beta + \delta)_{LR} \quad (12.36)$$

Umumiy ko'rinishda

$$(m, \alpha, \beta)_{LR} \oplus (n, \gamma, \delta)_{LR} = (m+n, 1, 1)_{LR},$$

Bu yerda $L^n = (\alpha L^{-1} + \gamma L^{-1})^{-1}$, $R^n = (\beta L^{-1} + \delta L^{-1})^{-1}$.

Noravshan sonni inkor etish formulasining ko'rinishi quyidagicha

$$-(m, \alpha, \beta)_{LR} = (-m, \beta, \alpha)_{RL} \quad (12.37)$$

Noravshan sonni ifodalovchi L va R funksiyalarning o'rni almashdi. (12.36) va (12.37) dan ayirish formulasini keltirib chiqarish mumkin

$$(m, \alpha, \beta)_{LR} - (n, \gamma, \delta)_{RL} = (m-n, \alpha + \delta, \beta + \gamma)_{LR}.$$

Ko'paytirish. Oldindagi kabi, quyidagiga ega bo'lish mumkin

$$z = xy = mn + (m\gamma + n\alpha)L^{-1}(\omega) + \alpha\gamma(L^{-1}(\omega))^2 \quad (12.38)$$

(12.38) tenglama $L^{-1}(\omega)$ ga nisbatan ikkinchi tartibli tenglama deyiladi va $z \leq mn$ holda bitta musbat ildizga ega. Teoremdan foydalangan holda, $\mu_{M \otimes N}$ ko'paytmaning tegishlilik funksiyasini aniqlash mumkin. Odatda bu ko'paytma $M \otimes N$ L-R turdagi noravshan son bo'lmaydi. Shunga qaramay, agar α va β m va n va n ga nisbatan kichik bo'lsa hamda (yoki) $\omega \in I$ atrofida joylashgan bo'lsa, $\alpha\gamma(L^{-1}(\omega))^2$ hadni hisobga olmasa bo'ladi. Bunda tenglama soddalashadi va yaqinlashgan formulalarni hosil qilish mumkin

$$\begin{aligned} (m, \alpha, \beta)_{LR} \otimes (n, \gamma, \delta)_{LR} &= (mn, m\gamma + n\alpha, m\delta + n\beta)_{LR}, \quad M > 0, N > 0 \text{ da}, \\ (m, \alpha, \beta)_{LR} \otimes (n, \gamma, \delta)_{LR} &= (mn, n\alpha - m\delta, n\beta - m\gamma)_{LR}, \quad M < 0, N > 0 \text{ da}, \\ (m, \alpha, \beta)_{LR} \otimes (n, \gamma, \delta)_{LR} &= (mn, -n\beta - m\delta, n\alpha - m\gamma)_{LR}, \quad M < 0, N < 0 \text{ da}. \end{aligned}$$

Agar kengaytmalar o'rta qiymatlarga nisbatan kichik bo'lmasa, boshqa taxminiy formulalardan foydalanish mumkin. Masalan, agar $M > 0$ va $N > 0$ bo'lsa, u holda

$$(m, \alpha, \beta)_{LR} \otimes (n, \gamma, \delta)_{LR} \approx (mn, m\gamma + n\alpha - \alpha\gamma, m\delta + n\beta + \beta\delta)_{LR}.$$

Ushbu tenglik orqali aniqlangan tegishlilik funksiyasi μ bilan kamida uchta nuqtada ustma-ust tushsa, u holda:

$$(mn, 1); [(m-\alpha)(n-\gamma), L(1)]; [(m+\beta)(n+\delta), R(1)].$$

Skalyarga ko'paytirish. Quyidagi formulalar o'rinni:

$$\forall \lambda > 0 \lambda \in R: \lambda \otimes (m, \alpha, \beta)_{LR} = (\lambda m, \lambda \alpha, \lambda \beta)_{LR},$$

$$\forall \lambda < 0 \lambda \in R: \lambda \otimes (m, \alpha, \beta)_{LR} = (\lambda m, -\lambda \alpha, -\lambda \beta)_{LR}.$$

Bo'lish. Musbat, noravshan L-R va R-L turdagi sonlar uchun quyidagi taxminiy natijalar olingan:

$$(m, \alpha, \beta)_{LR} \div (n, \gamma, \delta)_{LR} \approx \left(\frac{m}{n}, \frac{\delta m + \alpha n}{n^2}, \frac{\gamma m + \beta n}{n^2} \right)_{LR}.$$

12.3. Noravshan sonlarni solishtirish.

Noravshan sonlarni solishtiranda ikki turdagi savol paydo bo'ladi:

Berilgan noravshan sonlar oilasidagi eng katta (eng kichik) sonning noravshan qiymati qanday?

“ \bar{m} \bar{n} dan katta (kichik)” mulohazaning chinlik qiymati nimaga teng”. Boshqa so'z bilan aytganda, \bar{m} ning \bar{n} dan katta (kichik) bo'lishi ehtimoli qanday?

Birinchi savolga javob berish uchun quyidagi funksiyaga nisbatan qo'llanilgan umumlashtirish tamoyilidan foydalanamiz

$$z(m, n) = \max\{m; n\};$$

$$t(m, n) = \min\{m; n\}.$$

Agar $\mu_1(m), \mu_2(n)$ - \bar{m} va \bar{n} noravshan sonlarning tegishlilik funksiyalari bo'lsa, u holda umumlashtirish tamoyiliga binoan quyidagiga ega bo'lamiz

$$\mu_3(z) = \sup_{z = \max(m, n)} \min\{\mu_1(m), \mu_2(n)\};$$

$$\mu_4(t) = \sup_{z=\min(m,n)} \min\{\mu_1(m), \mu_2(n)\}$$

μ_1 va μ_2 tegishlilik funksiyali \tilde{z} va \tilde{t} noravshan to'plamlar \tilde{m} va \tilde{n} dan farqli qavariq normal sonlar bo'ladi. Ikkinchi savolga javob berishga harakat qilamiz " \tilde{m} \tilde{n} dan katta" mulohazaning chinlik darajasi $v(\tilde{n} > \tilde{m})$ ko'rinishda yozib olinadi va quyidagicha aniqlanadi:

$$v(\tilde{n} > \tilde{m}) = \max_{n \geq m} \min\{\mu_1\}$$

« \tilde{n} \tilde{m} dan katta» hol uchun

$$v(\tilde{n} > \tilde{m}) = \min\{\mu_1(A), \mu_2(B)\} = \min\{1, 1\} = 1;$$

$$v(\tilde{m} > \tilde{n}) = \min\{\mu_1(C), \mu_2(D)\} = d.$$

D soni \tilde{m} va \tilde{n} sonlarning farq qilish darajasini xarakterlaydi. C nuqtaning ordinatasi 1 ga qanchalik yaqin bo'lsa, \tilde{m} va \tilde{n} sonlardan qaysi biri kattaroq ekanligiga javob berish shunchalik qiyin bo'ladi.

\tilde{m} va \tilde{n} L-R va R-L turdagi noravshan sonlar bo'ladi; $\tilde{m} = (m, \alpha, \beta)$, $\tilde{n} = (n, \gamma, \delta)$, bunda C quyidagi tenglikdan topiladi

$$(C - m) / \beta = (n - C) / \gamma,$$

Bu yerdan

$$C = (\beta n + \gamma m) / (\beta + \gamma) \text{ i } d = R[(n - m) / (\beta + \gamma)] \quad (12.39)$$

Shunday qilib, ikkita noravshan sonning kesishmasi bo'lgan noravshan to'planning balandligi (12.39) formula bo'yicha aniqlanadi (biz $m < n$ holni ko'rib chiqmoqdamiz). Demak, \tilde{m} va \tilde{n} ni solishtirish uchun ham $v(\tilde{m} > \tilde{n})$, ham $v(\tilde{m} < \tilde{n})$ ni bilish kerak. Agar, masalan $v(\tilde{m} > \tilde{n}) = 1$ bo'lsa, bu degani yoki \tilde{m} \tilde{n} dan katta, yoki ikkala noravshan son ularni ajratish mumkin bo'lishi uchun, juda uzoqda joylashgan. Bunday holda ma'lum bir τ ostonaviy qiymatni tanlab, \tilde{m} \tilde{n} dan katta deb olish mumkin, agar $v(\tilde{n} > \tilde{m}) \leq \tau$ bo'lsa. Bu quyidagicha belgilanadi $\tilde{m} \geq \tilde{n}$.

L-R turdagi noravshan sonlar uchun qoida quyidagicha aniqlanishi mumkin:

$$\tilde{n} > \tilde{m} \Leftrightarrow n - m \geq \beta + \gamma; (\tau = R(1));$$

$$\tilde{m} > \tilde{n} \Leftrightarrow m - n \geq \alpha + \delta; (\tau = L(1));$$

Bu yerda qisqalik uchun \geq ning o'rniga $>$ belgilash kiritilgan. " n m dan katta" uchun boshqa ta'rifni kiritish ham mumkin:

$\tilde{n} \geq \tilde{m} \Leftrightarrow \max(\tilde{m}, \tilde{n}) = \tilde{n}$ va $\min(\tilde{m}, \tilde{n}) = \tilde{m}$. Misollar.

1. Noravshan \tilde{n} soni noravshan \tilde{m} sonidan katta, agar

$$n > m$$

$$n - \underline{n} \geq m - \underline{m}$$

$$n + \overline{n} \geq m + \overline{m}$$

(12.40)

2. \tilde{m} noravshan son \tilde{n} noravshan sonidan kichik, agar

$$m \leq n$$

$$m - \underline{m} \leq n - \underline{n}$$

$$m + \overline{m} \leq n + \overline{n}$$

(12.41)

3. \tilde{n} noravshan son \tilde{m} noravshan songa teng, agar

$$n = m$$

$$n - \underline{n} = m - \underline{m}$$

$$n + \overline{n} = m + \overline{m}$$

(12.42)

Misol. Ikkita L-R turdagi son berilgan:

a) $A = (1, 5, 2)_{LR}$

$B = (3, 8, 4)_{LR}$

b) $A = (8, 10, 6)_{LR}$

$B = (6, 9, 5)_{LR}$

c) $A = (20, 35, 30)_{LR}$

$B = (18, 20, 25)_{LR}$

d) $A = (4, 7, 6)_{LR}$

$B = (8, 10, 5)_{LR}$

e) $A = (2, 8, 6)_{LR}$

$B = (4, 8, 6)_{LR}$

Ushbu sonlarni qo'shish, ayirish, ko'patirish va bo'lishni hisoblang.

Javob:

Qo'shish	Ayirish	Ko'paytirish	Bo'lish
a) $(4, 13, 6)_{LR}$	$(-2, 9, 10)_{LR}$	$(3, 23, 10)_{LR}$	$(0.3, 2.11, 1.44)_{LR}$
b) $(14, 19, 11)_{LR}$	$(-2, 15, 15)_{LR}$	$(48, 132, 76)_{LR}$	$(1.33, 2.78, 3)_{LR}$
c) $(38, 55, 55)_{LR}$	$(2, 60, 50)_{LR}$	$(360, 1030, 1040)_{LR}$	$(1.11, 3.49, 2.9)_{LR}$
d) $(12, 17, 11)_{LR}$	$(-4, 12, 16)_{LR}$	$(32, 96, 68)_{LR}$	$(0.5, 1.19, 1.38)_{LR}$
e) $(6, 16, 12)_{LR}$	$(-2, 14, 14)_{LR}$	$(8, 48, 36)_{LR}$	$(0.5, 2.75, 2.5)_{LR}$

12.4. Noravshan lingvistik modellashtirish

Noravshan to'plamlar murakkab nohiziq, noaniq obyektlarni modellashtirish va funksiyalarni approksimatsiyalashda keng qo'llaniladi [1-8]. Noravshan modellashtirish g'oyasi shundan iboratki, obyekt parametrlari o'rtasidagi matematik bog'lanish odatda lingvistik qoidalarning atamaları bilan ifodalanuvchi sifatli munosabatlar bilan almashtiriladi.

9-ta'rif. Noravshan modellashtirish. Noravshan modellashtirish-tizim tavsiflarini noravshan qoidalardan foydalangan holda ta'riflash usulidir.

Noravshan modellar asosan uchta sinfga ajratiladi:

- lingvistik noravshan modellar;
- noravshan relyasion modellar;
- Sugeno-Takagi-Kang (TSK) modellari.

Noravshan modellarni qurish 1) tuzilmani; 2) parametrlarni obyektning kiruvchi va chiquvchi ma'lumotlaridan foydalangan holda aniqlashtirishga olib kelinadi.

12.4.1. Lingvistik noravshan modellar

Ko'p kirishli (n) va ko'p chiqishli obyektlar yoki tizimlarga nisbatan lingvistik noravshan modellarning (MIMO-model) tuzilmasi:

AGAR $X_1, A_{11}, VA, X_2, A_{12}, VA, \dots, X_n, A_{1n}$ bo'lsa u holda

$Y_1, B_{11}, VA, Y_2, B_{12}, VA, \dots, Y_m, B_{1m}$ dir

AGAR $X_1, A_{21}, VA, X_2, A_{22}, VA, \dots, X_n, A_{2n}$ u holda

$Y_1, B_{21}, VA, Y_2, B_{22}, VA, \dots, Y_m, B_{2m}$ dir

(12.43)

Shuningdek

Agar $X_1, A_{11}, VA, X_2, A_{12}, VA, \dots, X_n, A_{1n}$ bo'lsa u holda

$Y_1, B_{11}, VA, Y_2, B_{12}, VA, \dots, Y_m, B_{1m}$ dir

Bitta kirishli (x) va bitta chiqishli (y) obyektlar uchun (SISO-model) model quyidagi shaklda redusiyalanadi: Agar X, A_i bo'lsa u holda

$$B_i, \quad i = \overline{1, n}$$

(12.44)

(12.43) va (12.44) noravshan modellarni noravshan graf, lingvistik qoidalar jadvali (look-up-table with interpolation) va kirish va chiqishlarning ixtiyoriy akslantirilishi (structure-free) ko'rinishida tuzilmaviy ta'riflash mumkin. (12.43) va (12.44) modellarni qurish quyidagi bosqichlarni o'z ichiga

oladi:

1. Qoidalar sonini aniqlash. Kirishlar fazosi uchun ham, chiqishlar fazosi uchun ham noravshan klasteringni qo'llash kirish va chiqishlar fazolari bo'linuvchi noravshan tarmoqlar hosil bo'ladi. Bo'linishdan so'ng hosil bo'lgan noravshan tarmoqlar soni (12.43) yoki (12.44) modellardagi qoidalar sonini aniqlaydi.

2. Tegishlilik funksiyasini tanlash. Noravshan modellarni konstruksiyalashda qidiruvning asosiy parametrlari tegishlilik funksiyasining markazi va shakli (shape) dir. Bu parametrlar ko'pincha noravshan modellarni adaptatsiyalash uchun ishlatiladi. Modellashtirilayotgan obyektning kirishlari va chiqishlari to'g'risidagi tajribaviy ma'lumotlarini qo'llab, (12.43) yoki (12.44) modellar lingvistik qoidalarning ham antesedent, ham konsekvent qismlarida qo'llab, nazariy jihatdan approksimatsiyaning ixtiyoriy aniqligiga erishish mumkin. Shuni qayd etish joizki, bu masalani yechishning to'g'ridan-to'g'ri usuli, ya'ni muqobil tegishlilik funksiyalar tanlovi yo'q. Oxirgi paytda bu muammo ko'pincha sonli muqobillashtirish, ta'lim muammosi sifatida qo'yiladi. Neyron texnika, genetik algoritmlardan foydalangan holda tegishlilik funksiyasining shunday parametrlari tanlanadiki, ular asosida modellarning aniqligi va to'g'riligi ta'minlanadi. Noravshan modellarning ko'rib chiqilgan turi yaxshi semantic impretatsiyalashuvga, modelning aniqligi va ishonchliligida muhim omil hisoblangan boshqariluvchi interpolyatsiyaga, evristikani umumlashuvga qo'shish imkoniyatiga ega bo'lishi kerak.

12.4.2. Munosabatning noravshan modeli.

Noravshan relyatsion modelni qurish noma'lum noravshan R munosabatli relyatsion tenglamalar sistemasini yechish muammosi sifatida qaraladi:

$$B = A \circ R$$

(12.45)

Bu yerda $A \in F(x)$ va $B \in F(y)$ - universumlar fazosida aniqlanuvchi noravshan to'plamlar va $R \in F(X * Y)$ - modellashtiriluvchi obyekt yoki sistemaning kirishlari va chiqishlari orasidagi munosabatni ifodalovchi karteziyan ko'paytma orqali aniqlanuvchi noravshan munosabat, kompozitsiyaning sup-min operatori.

(12.45) bilan berilgan noravshan tizimni identifikatsiyalash masalasi A va $B_i, i = \overline{1, n}$ noravshan tarmoqlar ko'rinishida tasvirlanuvchi kiruvchi va

chiquvchi ma'lumotlar asosida R baholashga olib kelinadi.

$$R_i = A_i \circ R, \quad i = \overline{1, n} \quad (12.46)$$

Agar noravshan A va B to'plamlar (12.45) tenglamani qanoatlantirsa, eng katta noravshan R munosabat

$$B = A \circ R \quad \text{da} \quad R = A \alpha B \quad (12.47)$$

ga teng bo'ladi, bu yerda α A va B ning α -kompozitsiyasini anglatadi.

R tegishlilik funksiyasi quyidagi munosabatlardan aniqlanadi:

$$\begin{aligned} R(x, y) &= 1, \text{ agar } A(x) \leq B(y) \\ R(x, y) &= Y(y), \text{ agar } A(x) > B(y) \end{aligned} \quad (12.48)$$

$n \geq 2$ da istalgan R^* noravshan munosabat quyidagi formulaga asosan kelib chiqadi:

$$R^* = \bigcap_{i=1}^n R_i, \quad R_i = A_i \alpha B_i, \quad i = \overline{1, n} \quad (12.49)$$

Agar (12.46) sistema yechimga ega bo'lsa, u holda (12.49) munosabat (12.46) ning eng katta yechimi bo'ladi.

(12.46) sistema yechimga ega bo'lishining zaruriy va yetarli sharti

$$\forall i \forall y \exists x ((A_i(x) \geq B_i(y)) \wedge \forall j \neq i (B_j(y) < B_i(y) \rightarrow A_j(x) \leq B_j(y))) \quad \text{dir} \quad (12.50)$$

12.4.3. TSK-model

N ta kirishli va bitta chiqishli noravshan modellashtiriluvchi sistemani ko'rib chiqamiz. U holda bunday turdagi sistemalarga nisbatan TSK-model

R' : AGAR x_1, A'_1 VA $x_2, A'_2 \dots$ VA x_n, A'_n bo'lsa u holda

$$Y' = f'(X_1, X_2, \dots, X_n), \quad (12.51)$$

ko'rinishda bo'ladi, bu yerda R' , $i = \overline{1, m}$ - i-implikasiya, m-noravshan modelning qoidalar soni, x_1, x_2, \dots, x_n - modelning kiruvchi o'zgarishlari, A'_j , $i = \overline{1, n}$ - kiruvchi o'zgaruvchilarning noravshan qism to'plamlari, y' kirishlarning nochiziqli (yoki chiziqli) funksiyasi sifatida aniqlanuvchi i-chiqishi.

$(x_1^0, x_2^0, \dots, x_n^0)$ kiruvchi o'zgaruvchilarning qiymatlari ma'lum bo'lsa, noravshan modelning yakuniy chiqishi

$$Y' = \sum_{i=1}^n W_i y' / \sum_{i=1}^n W_i \quad (12.52)$$

Sifatida aniqlanishi mumkin, bu yerda W_i - i-implikasiya shartli qismining chinlik qiymati bo'lib, u

$$W_i = \bigwedge_{j=1}^n \mu_{A'_j}(x_j^0) \quad (12.53)$$

Bu yerda $\mu_{A'_j}(x_j)$ - A'_j noravshan qism to'plamlarning tegishlilik funksiyasi.

TSK-model, umuman olganda bir nechta qoidalardan foydalangan holda murakkab nochiziqli funksional bog'lanishni ta'riflashi mumkin. Bunday modelni qurish qoidalari shartli qismining muqobil tuzilmasini (ya'ni kiruvchi o'zgaruvchilarning tegishlilik funksiyasini), tuzilmaning konsenkventini (ya'ni qanday termlar nochiziqli yoki chiziqli tenglamaga qo'shilishini aniqlash va ularning parametrlari bahosi) aniqlashni nazarda tutadi.

Shuni qayd etish joizki, amaliyotda TSK-modelni qurish oson masala emas. Bunday modellarni qurishning qat'iy tizimli prosedurasini mavjud emas. Bundan tashqari industrial jarayonlarni modellashtirishda tuzilmaviy va parametric off-line indentifikatsiya uchun kiruvchi va chiquvchi ma'lumotlarni to'plash qiyindir.

TSK-modellarni qurish prosedurasini yaxshilash uchun bir qator g'oyalar va usullar taklif etilgan. R.Langary va L.Wang da TSK-modellarni aniqlashga yangi yondashuv taklif etilgan, unda noravshan to'plamlarni yaratish uchun noravshan diskretlashtirish vositalaridan foydalangan holda avvalo modellarning shartli qismi aniqlanadi. Noravshan to'plamlar soni modelning konsenkvent qismidagi qoidalar sonini va chiziqli tenglamalar sonini aniqlaydi. Ushbu chiziqli tenglamalarning parametrlari orthogonal estimatordan foydalangan holda baholanadi. Quyida bunday modellarni qurishning turli xil uslublarini solishtirish jadvali keltiriladi.

12.1-jadval.

Modelning nomi	Kirishlar soni	Qoidalar soni	O'rta kvadratik xatolik
Tong modeli (Tong, 1979)	2	19	0,469
Pedruks modeli (Pedruks(1984)	2	81	0,320
Xu modeli (Xu va Lu, 1987)	2	25	0,328
Boks modeli (Boks va Djekins, 1976)	6	-	0,202
Sugeno modeli (Sugeno, Yasukava, 1993)	3	6	0,190
Vang modeli (Vang va Langari, 1994)	2	5	0,158
Sugno modeli (Sugeno va Tanaka, 1991)	6	2	0,068
Langari va Vang modeli (Langari va Vang, 1994)	6	2	0,066

TSK-modellarning aniqlash va tasvirlashning samarali bo'lishi uchun zarur bo'lgan qoidalarining kam soni va ma'lumotlarning kichik hajmi kabi ustuvorliklariga qaramay, ular lingvistik va relyatsion modellarga qaraganda, evristikalarni qo'shishning semantik qiyinchiligidan aziyat chekadilar TSK va lingvistik modellar xossalari nisbiy baholanishi quyida keltirilgan (12.2-jadval).

12.2-jadval

Xossa	TSK-model	Lingvistik model
1. Tasvirlash	Samarali	Samarali emas
2. Umumlashtirish	Past	Yuqori
3. Qoidalar soni	Past	Yuqori
4. Semantik qiymat	Chegaralangan	Yaxshi
5. Evristikani qo'shish	Qiyin	Oson
6. Interpolyatsiya	Boshqarilmaydi	Boshqariluvchi
7. Aniqlashtirish uchun ma'lumotlar	Kam	Ko'p

U yoki bu modellarning ustuvorliklari va kamchiliklarini hisobga olgan holda noravshan chiziqli modelni bo'lishga asoslangan to'ldiruvchi noravshan bo'linish qonuniyatini hisobga oluvchi yangi yondashuv taklif etiladi. U quyidagi misolda tasvirlanib o'tiladi.

Misol. Fermenterda bosim dinamikasining noravshan modeli aniqlashtiriladi. Noravshan TSK-model quyidagi ko'rinishda qabul qilinadi:

R_1 : AGAR $x(k)$ kichik va $y(k)$ o'rta bo'lsa u holda

$$y(k+1) = a_1 y(k) + b_1 x(k) + c_1 \quad (12.54)$$

R_2 : AGAR $x(k)$ taxminan nol va $y(k)$ katta bo'lsa u holda

$$y(k+1) = a_2 y(k) + b_2 x(k) + c_2$$

Bu yerda $x(k)$, $y(k)$ -* shartli qismning o'zgaruvchilari, a, b, c - konsekvant parametrlar, k - diskret vaqt. Yuqorida ko'rsatilib o'tilganidek, (12.54) noravshan modelni qurish modeli 1) (12.54) modeldagi qoidalarining shartli qismida tegishlilik funksiyasini aniqlash va 2) o'ng qism tuzilmasini va a, b, c parametrlarning qiymatini baholashga asoslanadi. Birinchi bosqichning mos vositasi giperplanar noravshan sinflashtirish, ikkinchi bosqichda esa aniqlashtirish nazariyasining mashhur usullari qo'llanilishi mumkin.

Giperplamar noravshan sinflashtirishdan foydalangan holda quyidagi uchta qoida aniqlangan:

Agar $y(k)$ past va $x(k)$ ochiq bo'lsa u holda

$$y(k+1) = 0,67y(k) + 0,0007x(k) + 0,35$$

Agar $y(k)$ o'rta va $x(k)$ yarim yopiq bo'lsa u holda

$$y(k+1) = 0,80y(k) + 0,0028x(k) + 0,07$$

Agar $y(k)$ yuqori va $x(k)$ yopiq bo'lsa u holda

$$y(k+1) = 0,90y(k) + 0,0071x(k) - 0,39$$

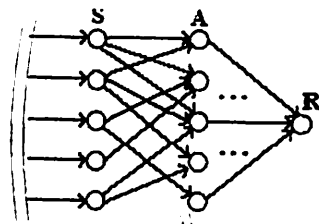
Ko'rinib turganidek, fermenterdagi bosimning dinamik modeli birinchi tartibli nochiziqli regression model orqali tasvirlanadi.

Asl noravshan chiziqli bo'linishdan klapan va bosim holati uchun to'ldiruvchi bo'linishning tegishlilik funksiyalari keltirib chiqarildi.

13. NEYRON TO'RLAR USULI. PERSEPTRON

«Perseptron» tushunchasini birinchi bo'lib 1957 yilda neyrofiziolog Frenk Rozenblatt (Rosenblatt) kiritdi. Perseptron miya va uning alohida tizimini ayrim sinfi bo'lib hisoblanadi (masalan, ko'rishga oid).

Uning sxematik qurilmasi 13.1-rasmda ko'rsatilgan. Bu yerda S – sezuvchi sensor to'plami (setchatka), A – biriktiruvchi elementlar to'plami (neyronlarni), R – sezgir element (ya'ni mushaklarga va bezlarga boshqaruv signalini uzatuvchi neyron). Neyronning ko'plab kirishlari va bitta chiqishi bor. Neyronga kirishlar tormozlashuvchi va qo'zg'aluvchi turga bo'linadi. Sensor elementlari, agar qo'zg'alish ta'siri natijasida (masalan, yorug'lik) kirish signalining kattaligi ayrim bo'sag'aviy qiymatdan katta bo'lishi ma'lum bo'lsa, qo'zg'aladi. S –elementlar tasodifiy tarzda A-neyronlar bilan bog'liq. Bunda agar qo'zg'aluvchi signallarni soni tormozlashuvchi signallar sonidan katta bo'lsa u holda neyron qo'zg'aladi va reaksiya qiluvchi elementga signal jo'natadi.



13.1-rasm.

Neyronlardan farqli reaksiya qiluvchi elementga kelayotgan signallar, bir qancha vaznlar bilan yig'indisi olinadi. Reaksiya qiluvchi element ba'zi harakatlarni tanlaydi, agar $R(x) = \sum_{i=0}^n w_i x_i = (w, x) > 0$, bo'lsa, bu yerda $w = (w_0, w_1, \dots, w_n)$, $x = (1, x_1, \dots, x_n)$, x_i – i – neyrondan reaksiya qiluvchi elementga kelayotgan signal ($x_0 \equiv 1$ – siljish signali).

Perseptron yordamida obyektlarni ikkita sinf bo'yicha sinflashtirishni amalga oshirish mumkin. Taxmin qilamizki, $x \in \varpi_1$ deb, agar $R(x) > 0$, va aks holda - $x \in \varpi_2$ $R(x) = (w, x)$ funksiya chiziqli hal qiluvchi funksiya bo'ladi (CHHF), gipertekislik esa $(w, x) = 0$ –chiziqli ajratuvchi yuza.

O'ylaymizki, ayrim o'quv tanlanma $\{x_1, \dots, x_n\}$ mavjud bo'lsin, shuningdek $\{x_1, \dots, x_n\} \subset \varpi_1$ va $\{x_{m+1}, \dots, x_n\} \subset \varpi_2$. O'quv tanlanma elementlarini to'g'ri

ajratuvchi chiziqli hal qiluvchi funksiyani $R(x) = (w, x)$ qurish talab etiladi

$$(w, x_i) > 0 \quad \forall i = 1, \dots, m;$$

$$(w, x_i) < 0 \quad \forall i = m+1, \dots, n;$$

(13.1)

Bunday CHHFni qurishni iteratsiyali algoritmi yordamida amalga oshiramiz. Buning uchun o'quv tanlanmani cheksizssiklik ketma-ketlik $\{x_1, \dots, x_n, x_1, \dots, x_n, \dots\}$ ko'rinishida yozib olamiz va ayrim vaznlarni $w^{(k)} = (w_1^{(k)}, \dots, w_n^{(k)})$ boshlang'ich qiymati tanlab olamiz. Bundan tashqari algoritmning har bir k -qadamida xk o'quv tanlanmaning k -vektori «taqdim etiladi» va $w^{(k)}$ vazn vektorining qiymati quyidagi qoidaga muvofiq tuzatiladi yoki tuzatilmaydi

$$w^{k+1} = \begin{cases} w^k, & \text{agap } (w^{(k)}, x_k) > 0 \text{ esa } x_k \in \varpi_1; \\ w^k, & \text{agap } (w^{(k)}, x_k) < 0 \text{ esa } x_k \in \varpi_2; \\ w^k + x_k, & \text{agap } (w^{(k)}, x_k) \leq 0 \text{ esa } x_k \in \varpi_1; \\ w^k - x_k, & \text{agap } (w^{(k)}, x_k) \geq 0 \text{ esa } x_k \in \varpi_2. \end{cases}$$

(13.2)

Ya'ni vazn vektori o'zgarmaydi, agar «taqdim etilgan» vektor to'g'ri sinflashtirilsa va noto'g'ri sinflashtirishda xk ga oshirilsa yoki kamaytirilsa. Algoritm o'z ishini tugatadi, agar o'quv tanlanma timsollarini to'g'ri sinflashtirishni n – karra amalga oshirilsa.

Perseptronni o'qitish algoritmini soddalashtirish uchun $\{x_1, \dots, x_n\}$ o'quv tanlanmani o'miga $\{y_1, \dots, y_n\}$ tanlanmani ko'rib chiqamiz, bu yerda $i = m+1, \dots, n$ uchun $i = 1, \dots, m$ va $y_i = -x_i$ uchun $y_i = x_i$. U holda (8) sistema o'miga w^* vektor sistemani qondirishi zarur

$$(w, y_i) > 0 \quad \forall i = 1, \dots, n;$$

(13.3)

va (13.2) formulani o'miga perseptron algoritmidagi vazn vektorini to'g'rilash quyidagi soddalashtirilgan formula bo'yicha amalga oshiriladi:

$$w^{k+1} = \begin{cases} w^k, & \text{agap } (w^{(k)}, y_k) > 0, \\ w^k + y_k, & \text{agap } (w^{(k)}, x_k) \leq 0. \end{cases}$$

Misol. Timsollarni ikkita sinf $x \in \varpi_1$ va $x \in \varpi_2$ bo'yicha ajratish uchun perseptronni o'qiting, agar $\{x_1, x_2\} \subset \varpi_1$ va $\{x_3, x_4\} \subset \varpi_2$ ma'lum bo'lsa, bu yerda $x_1 = (1, 0, 1, 0)$, $x_2 = (1, 1, 1, 0)$, $x_3 = (0, 0, 1, 1)$, $x_4 = (1, 1, 0, 0)$ (bunday binar vektorlar

yordamida kodlashtirish mumkin, masalan, binar tasvirlar). O'qitish jarayonini.

13.1-jadvalda tasvirlaymiz

Vektorlar																				
w																				
			1	1			1	1				1	1							1
(w,x)							1	1	1	1	2	2	2	2	3	3				
Tuzatmalar																				
r																				

Vektorlar	x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4
w	0	1	1	0	0	0	0	0
	-1	0	0	-1	-1	-1	-1	-1
	1	1	2	2	2	2	2	2
	-3	-3	-3	-3	-3	-3	-3	-3
(w,x)	1	0	-1	1	2	1	-1	-1
Tuzatmalar	-	+	-	+	-	-	-	-

Ushbu jadvalda birinchi qatorda o'quv tanlanma elementlari keltirilgan, navbatdagi 4 ta qatorlarda vazn vektori w koeffitsientlari keltiriladi, oltinchi qatorda – skalyar ko'paytma hisoblash natijalari, oxirgi qatorda "+" shuni bildiradiki, vektor noto'g'ri sinflashtirilgan va koeffitsientlarni to'g'irlashni amalga oshirish zarur, "-" esa shuni bildiradiki, vektor to'g'ri sinflashtirilgan va to'g'irlashni amalga oshirish zarur emas. To'rt marta ketma-ket to'g'ri sinflashtirilgandan so'ng algoritm tugaydi, natijada vazn vektor $w = (0, -1, 2, -3)$ hosil bo'ladi. Payqash mumkinki, berilgan misolda koordinata boshidan o'tuvchi to'rt o'lchamli fazoda ajratuvchi gipertekislikni olamiz. Umumiy holda ajratuvchi gipertekislikni topish uchun (agar u mavjud bo'lsa) n – o'lchovli fazoga siljishni kiritish zarur: (n+1) -o'lchov vektorlari $x = (1, x_1, \dots, x_n)$ ni ko'rib chiqish va qidirish (n+1) -vazn o'lchov vektorlarini $w = (w_0, w_1, \dots, w_n)$ izlash.

14. NEYRON TO'RLAR USULI. XOPFILD VA XEMMING NEYRON TO'RLARI

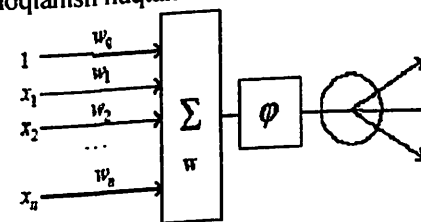
14.1. Xopfild neyron to'ri va o'qitish algoritmi

Neyroinformatika asosida (NT) neyron to'rlarini o'rganish bilan shug'ullanuvchi sun'iy intellekt bo'limi ikkita ko'rinish yotadi: miya tuzilishi to'g'risida va o'qitish jarayoni to'g'risida. Miyaning tuzilishini ko'rib chiqishda kalit element bo'lib, oddiy element tushunchasi, miya g'ishtlari – "neyron" bo'lib hisoblanadi. Ikkinchi ko'rinish tirik organizmlarga o'xshashlik bo'yicha ma'lum masalani hal qilish maqsadida neyronlar (neyron to'r) orasidagi aloqani o'qitish orqali oladigan shakllantirishga asoslanadi. NT perseptron bo'lib hisoblanadi.

NTga to'g'ridan-to'g'ri dasturlash uchun ham ruxsat beriladi, ya'ni aniq qoidalar bo'yicha bog'lanishlarni shakllantirish (aniq formula bo'yicha bog'lanishlarni shakllantirish uchun masalalarni katta sinfi mavjud), ma'lum masalalarni hal qilishga NTni noaniq sozlashga ham ruxsat beriladi. Ushbu jarayonni o'qitish ham deb ataladi.

O'qitish odatda «rag'batlantirish-jazo» tamoili bo'yicha quriladi: tizimda berilgan javoblar bilan misollar to'plami taqdim etiladi. Neyronlar kirish signallariga aylanadi, signallar to'plami ham shunday javob qaytaradi. To'g'ri javobdan og'ish to'rlarni ichki sozlashni o'zgartirish yo'li bilan jarimaga tortiladi.

Standart formal neyron chiqishdagi kirish summatori, chiziqsiz o'zgartiruvchi va tarmoqlanish nuqtalaridan tashkil topgan.



14.1-rasm.

Neyronning eng muhim elementi – bu adaptiv summatoridir, w sozlanuvchi parametrlar vektoriga x kiruvchi signal vektorini skalyar ko'paytmasini hisoblaydi (neyrofiziologiyaga o'xshashlik bilan neyronning kirishiga dendrit deb ataladi) (14.1-rasm)

Keyin signal aktivatsiya funksiyasi deb nomlanuvchi chiziqsiz o'zgaruvchiga o'tadi. U skalyar kirish signali x ni $\varphi(x)$ signalga o'zgartiradi (neyronning chiqishi akson deb nomlanadi). Tarmoqlanish nuqtasi bir qancha adreslar bo'yicha bir signalni yuborish uchun xizmat qiladi.

Chiziqli aloqa (sinaps) summatorlardan alohida uchramaydi. Sinaps «sinaps vazni» w ga x kiruvchi signalni ko'paytiradi.

NT to'plamlari orasida ikkita asosiy arxitekturalarni – qatlamli va to'liq bog'lanishli to'rlarni ajratib ko'rsatish mumkin.

Qatlamli to'rlarda neyronlar bir qancha qatlamlarda joylashgan. Birinchi qatlam neyronlari kirish signallarini qabul qiladi, ularni o'zgartiradi va tarmoqlanish nuqtasi orqali neyronning ikkinchi qatlamiga uzatadi. Keyin ikkinchi qatlam va h.k. chiqish signallarini beruvchi k -qatlamgacha ishlaydi. Agar aksi ko'rsatilmagan bo'lsa, u holda signal i -qatlamning har bir chiqish signali $(i+1)$ -qatlam neyronlarini kirishiga uzatiladi. Har bir qatlamdagi neyronlar soni ixtiyoriy bo'lishi mumkin va avvaldan boshqa qatlam neyronlari miqdoriga bog'lanmagan. Ko'pincha ikki va uch qatlamli NT uchraydi.

To'liq bog'lanishli to'rlarda har bir neyron o'zining chiqish signalini boshqa neyronlarga, shu jumladan o'ziga uzatadi. To'rlarning chiqish signallari to'rlarni funkcionallashtirishning o'lchovlaridan keyin bir qancha neyronlarni barcha yoki ayrim chiqish signallari bo'lishi mumkin. Barcha chiqish signallari barcha neyronlarga uzatiladi.

NT bir o'zgaruvchining chiziqli funksiyalarini, chiziqsiz funksiyalarini, shuningdek superpozitsiyalarning barcha turlarini – tarmoqlarning kaskadli birlashishi natijasida olingan funksiyalarni hisoblash mumkin.

Xopfild NT assotsiatiy xotira deb nomlanuvchi tarmoqlarga tegishli, ya'ni ushbu to'rlar timsollar to'g'risidagi ba'zi ma'lumotlarni saqlaydi va uning kirishiga kelgan o'sha timsolni «eslab qoladi». Kiruvchi vektor bipolyar bo'lishi zarur, ya'ni $x_i \in \{-1, 1\}$. Xopfild neyron to'ri har bir taqdim etilgan bipolyar x vektor uchun unga eng yaqin bo'lgan etalon-timsol topiladi va chiqishda uni chiqaradi. Etalon-timsol to'g'risidagi ma'lumotlar NT bilan sinaptik bog'lanishlarda «himoya» bo'ladi. Shuningdek, vektorlar bipolyar bo'lgani uchun x vektor va e_k etalon vektor orasidagi masofa $\|x - e_k\|^2 = \|x\|^2 + \|e_k\|^2 - 2(x, e_k)$ ga teng bo'ladi. U holda (x, e_k) skalyar ko'paytma

faqatgina maksimal bo'lganda $\|x - e_k\|$ masofa minimal bo'ladi. Xopfild to'ri funkcionallashtirilishi asosida ayrim funkcionallarni minimallashtirish g'oyasi yotadi (shunday deb ataluvchi energiya funkcionali)

$$H(x) = -0.5 \sum_k (x, e_k)^2 + \alpha \sum_k (x_i^2 - 1)^2 + \alpha \sum_k (x_i^2 - 1)^2, \quad \alpha > 0. \quad (14.1)$$

(14.1) ifodada birinchi qo'shiluvchi minimal bo'ladi shu holatdagi, agar x vektor etalon vektorlardan biriga (yoki bir qanchasiga) yaqin bo'lsa, ikkinchi qo'shiluvchi x vektor koordinatasi bipolyar vektorga yaqin bo'lgan paytda minimal bo'ladi. Koeffitsient $\alpha > 0$ ushbu ikki mezon ustivorligini boshqaradi. To'ring funkcionallashtirilishi jarayonida uni asta-sekin oshirish tavsiya etiladi. $H(x)$ funkcionallarni minimallashtirishni quyidagi formula bo'yicha gradient usuli bilan amalga oshirish mumkin:

$$x(t+1) = x(t) - h \text{grad} H(x(t)),$$

bu yerda t – iteratsiya nomeri; h – gradient qadam. Gradient vektor koordinatasini hisoblash uchun xususiy hosilasini topamiz:

$$\frac{\partial H}{\partial x_i} = -\sum_k (x, e_k) e_{ik} + 4\alpha (x_i^2 - 1) x_i = -\sum_k \left(\sum_j x_j e_{jk} \right) e_{ik} + 4\alpha (x_i^2 - 1) x_i =$$

$$-\sum_j \left(\sum_k e_{jk} e_{ik} \right) x_j + 4\alpha (x_i^2 - 1) x_i = -\sum_j w_{ij} x_j + 4\alpha (x_i^2 - 1) x_i$$

bu yerda $w_{ij} = \sum_k e_{jk} e_{ik}$ (k – etalon nomeri). Ikkinchi qo'shiluvchi tarmoq qatnashmagan holda i -neyronda bevosita hisoblanadi. i - va j - neyronlar orasidagi bog'lanishlar vazni $w_{ij} = \sum_k e_{jk} e_{ik}$ ga teng bo'ladi. Shunday qilib, energiya funkcionallarni minimallashtirish quyidagi formula bo'yicha amalga oshiriladi

$$x_i(t+1) = x_i(t) + h \sum_j w_{ij} x_j(t)$$

yoki matritsa ko'rinishida ushbu formula bo'yicha

$$X(t+1) = (I + hW)X(t) \quad (14.2)$$

bu yerda $X(t)$ – x vektor koordinatasini ustun-vektori; I – birlik matritsa. Bu yerda $W = EE^T$, bu yerda $E = [E_1, \dots, E_m]$ – etalon vektorlar koordinatasini E_k ustun vektorlaridan tuzilgan matritsa. W matritsa $n \times n$ o'lchamga ega va uning elementlari qiymatlari m sonlar moduli bo'yicha ustun bo'ladi. Shuning

uchun ham $\|W\| < nm$ iteratsiyali jarayon ham (12) mos keladi, agar $0 < h < 1/(nm)$. Ma'lumki, agar matritsaning asosiy diagonaliga W nol elementlarni qo'yilsa, iteratsiyali jarayon kafolatli mos keladi.

Xopfild to'ri bir qatlamli, to'liq bog'lanishli va 14.1-rasmga ko'rsatilgan tuzilishga ega bo'ladi. Qatlamdagi neyronlar soni vektorlar o'lchamiga teng bo'ladi. x boshlang'ich vektori har bir neyronni kirishiga uzatiladi. Boshlang'ich davrda sinapslarning vazn koeffitsientlari to'g'ri o'rnatiladi

$$q_{ij} = \begin{cases} h \sum_k e_{jk} e_{ik}, & i \neq j; \\ 0, & i = j. \end{cases}$$

Xopfild algoritmi quyidagi ketma-ket qadamlardan tashkil topadi:

1. Kirishga $x(x_i)$ signal uzatiladi va taxmin qilamizki, $y(0) = x$, $t = 0$.

2. Neyronlarning yangi holati quyidagi formula bilan hisoblanadi

$$s_i(t+1) = \sum_j q_{ij} y_j(t)$$

va aksonlarning yangi qiymati (aktivatsiya funksiyasi) quyidagi formula bo'yicha hisoblanadi

$$y_i(t+1) = f(s_i(t+1)), \text{ bu yerda } f(u) = \begin{cases} 1, & u \geq 0; \\ -1, & u < 0. \end{cases}$$

3. Aksonlarni barqarorlashtirish sharti tekshiriladi: agar aksonlar barqarorlashsa, (ya'ni $y_i(t+1) = y_i(t)$ barcha i lar uchun), u holda algoritm ishni tugatadi, aks holda 2-punktga o'tamiz.

Izoh. Ba'zan Xopfild to'ri to'g'ri tanib olishni bajarolmaydi va chiqishda mavjud bo'lmagan timsolni hosil qiladi. To'r timsollarni ishonchli tanib olishi uchun ular bir-biri bilan kuchsiz bog'lanishlari kerak va ularning soni $m < 0,15n$ dan oshmasligi kerak, bu yerda n – vektorlar o'lchami.

Misol. Ikkita olti o'lchamli etalon vektor mavjud bo'lsin, $e_1 = (1, 1, 1, -1, 1, -1)$ va $e_2 = (-1, -1, -1, 1, 1, 1)$. Ushbu vektorlar va $x = (1, 1, 1, -1, 1, 1)$ vektorni sinflashtirish uchun Xopfild to'rini qurish talab qilinadi.

Yechish. Payqaymizki, tanib olishning turg'unlik sharti $m < 0,15n$ ushbu misolda bajarilmaydi, shuningdek $m \square 2$, $n \square 6$. Sinaptik bog'lanishlarni quramiz $h \square 0,5$ uchun $Q = H + hW \square$. Quyidagilarga ega bo'lamiz:

$$E = [E_1, E_2] = \begin{bmatrix} 1 & 1 & 1 & -1 & 1 & -1 \\ -1 & -1 & -1 & 1 & 1 & 1 \end{bmatrix}^T; \quad Q + I + 0,5EE^T = \begin{bmatrix} 2 & 1 & 1 & -1 & 0 & -1 \\ 1 & 2 & 1 & -1 & 0 & -1 \\ 1 & 1 & 2 & -1 & 0 & -1 \\ -1 & -1 & -1 & 2 & 1 & 1 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ -1 & -1 & -1 & 1 & 0 & 2 \end{bmatrix}$$

U holda $x = (1, 1, 1, -1, 1, 1)$ vektorning taqdim etilishida birinchi iteratsiyadan so'ng to'rning chiqishida ustun-vektorni $S = QX = (4, 4, 4, -3, 2, -2)^T$ olamiz, aktivatsiya funksiyasini qo'llagandan keyin f – ustun-vektor $Y = (1, 1, 1, -1, 1, -1)^T$. Shunday qilib, x vektor birinchi iteratsiyadan so'ng birinchi sinfga tayinlanadi.

14.2. Neyron to'rlarini o'qitishning Xemming algoritmi

Ushbu to'rda bipolyar vektorlar uchun Xemming masofasi xossasidan foydalaniladi. Agar ikkita bipolyar vektor $x = (x_i)$ va $y = (y_i)$, $x_i, y_i \in \{-1, 1\}$ mavjud bo'lsa, u holda x va y orasidagi Xemming masofasi $\|x - y\| = \sum_i |x_i - y_i| = 2d(x, y)$ ga teng bo'ladi, bu yerda $d(x, y)$ – ushbu vektorlar turlicha komponentlari soni. Bundan tashqari, agar $a(x, y) = x$ va y vektorlarni ustma-ust tushuvchi komponentlari soni, a_n – ushbu vektorlar o'lchamligi, u holda

$$a(x, y) + d(x, y)n; \quad (x, y) = \sum_i x_i y_i = a(x, y) - d(x, y) = 2a(x, y) - n.$$

U holda ikkita bipolyar vektorlar orasidagi yaqinlik o'lchovi $a(x, y)$ $a(x, y) \approx \frac{n}{2} + \frac{1}{2} \sum_i x_i y_i$ ga teng bo'ladi. Agar m etalon vektorlar bo'lsin, u holda x kiruvchi vektor va e_k etalon vektor orasidagi yaqinlik o'lchami quyidagicha bo'ladi:

$$a(x, e_k) = \frac{n}{2} + \frac{1}{2} \sum_i e_{ki} x_i$$

Shunday qilib, agar k -neyronning sinapsi bo'lib $w_k = (w_{ik})_{i=1}^n$ vektor hisoblanadi, bu yerda $w_{0k} = n/2$ ($k=1, \dots, n$), u holda $\tilde{x} = (1, x_1, \dots, x_n)$ vektor neyronning kirishiga qabul qilishdan so'ng ushbu neyronning chiqishida esa k -vektor-etalonga berilgan vektorning yaqinlik o'lchami qiymatini olamiz.

Xemming NT ikkita qatlamga ega. m neyronlarning birinchi qatlamida (etalonlar soni bo'yicha), ularning har biri kirish vektori \tilde{x} ni barcha

komponentlarini qabul qiladi. Ushbu qatlamning har bir k-neyroni e_k etalon vektorga x vektorning a_k yaqinlik o'lchami qiymatini hisoblaydi. Ikkinchi qatlamda (Maxnet tarmog'i deb ataluvchi tarmoq) yaqinlik o'lchami vektori $a=(a_1, \dots, a_m)$ bitta nol bo'lmagan komponent bo'lgan vektor (y_1, \dots, y_m) ga aylanadi. Ushbu komponent raqami x kiruvchi vektorga eng yaqin bo'lgan (Xemming metrikasi ma'nosida) o'sha element raqamiga teng bo'lishi zarur. Maxnet tarmog'i to'liq bog'lanishli, bir qatlamli m neyronlardan tarkib topgan. Har bir neyronning chiqish signallari barcha neyron to'rlarini kirishiga o'tadi. Tarmoq chiquvchi neyronlar qiymati barqarorlashmaguniga qadar iteratsiya rejimida ishlaydi. Xemming algoritmi quyidagi ketma-ket qadamlardan iborat:

1. Birinchi qatlam kirishiga signal $\bar{x}=(1, x_1, \dots, x_n)$ uzatiladi va $(a_k = w_k \cdot \bar{x})$ yaqinlik o'lchamlari vektori $a=(a_1, \dots, a_m)$ hisoblanadi, bunda taxmin qilamizki, $y_i(0) = a_i$ ($i=1, \dots, m$) va $t=0$.

2. Ikkinchi qatlamning har bir k - neyronning kirishiga $y_i(t)$ vektor kiradi. Neyronlarni yangi holati $s_i(t+1) = \sum_j q_{ij} y_j(t)$ formula bilan hisoblanadi, bu yerda

$$q_{ij} = \begin{cases} 1, & i=j, \\ -\varepsilon, & i \neq j, \end{cases} \quad 0 < \varepsilon \leq 1/n. \text{ Yangi aksonlar qiymatlari (aktivatsiya qiymatlari)}$$

hisoblanadi $y_i(t+1) = f(s_i(t+1))$, bu yerda $f(u) = \begin{cases} u, & u \geq 0, \\ 0, & u < 0. \end{cases}$

3. Aksonlarni barqarorlashtirish sharti tekshiriladi: agar aksonlar barqarorlashsa (ya'ni $y_i(t+1) = y_i(t)$ barcha i lar uchun), u holda algoritm ishni yakunlaydi, aks holda 2-puktga o'tamiz.

Misol. Ikkita 6-o'lchamli etalon-vektor mavjud bo'lsin, $e_1 = (1, 1, 1, 1, -1, 1)$ va $e_2 = (-1, 1, -1, 1, 1, 1)$. Xemming to'ri yordamida $\square \square x \square (1, 1, 1, 1, -1, -1)$ vektorga eng yaqin etalon vektor raqamini topish talab qilinadi.

Yechish. x vektor va etalon vektorlar orasidagi yaqinlik o'lchami qiymatlarini hisoblaymiz:

$$a_1 = \frac{6}{2} + \frac{1}{2}(e_1, x) = 3 + \frac{1}{2} \cdot 4 = 5, \quad a_2 = \frac{6}{2} + \frac{1}{2}(e_2, x) = 3 + \frac{1}{2} \cdot (-2) = 2.$$

Keyin Maxnet to'ri yordamida yaqinlashish o'lchami vektoridan (5,2) a maksimal komponentni belgilaymiz:

$$\text{1-qadam} \begin{cases} s_1(1) = 5 - \frac{1}{6} \cdot 2 = \frac{14}{3}, \\ s_2(1) = 2 - \frac{1}{6} \cdot 5 = \frac{7}{6} \end{cases} \Rightarrow \begin{cases} y_1(1) = \frac{14}{3}, \\ y_2(1) = \frac{7}{6}; \end{cases}$$

$$\text{2-qadam} \begin{cases} s_1(2) = \frac{14}{3} - \frac{1}{6} \cdot \frac{7}{6} = \frac{161}{36}, \\ s_2(2) = \frac{7}{6} - \frac{1}{6} \cdot \frac{14}{3} = \frac{7}{18} \end{cases} \Rightarrow \begin{cases} y_1(2) = \frac{161}{36}, \\ y_2(2) = \frac{7}{18}; \end{cases}$$

$$\text{3-qadam} \begin{cases} s_1(3) = \frac{161}{36} - \frac{1}{6} \cdot \frac{7}{18} = \frac{119}{27}, \\ s_2(3) = \frac{7}{18} - \frac{1}{6} \cdot \frac{161}{36} = -\frac{77}{216} \end{cases} \Rightarrow \begin{cases} y_1(3) = \frac{119}{27}, \\ y_2(3) = 0; \end{cases}$$

$$\text{4-qadam} \begin{cases} s_1(4) = \frac{119}{27} - \frac{1}{6} \cdot 0 = \frac{119}{27}, \\ s_2(4) = 0 - \frac{1}{6} \cdot \frac{119}{27} = -\frac{119}{162} \end{cases} \Rightarrow \begin{cases} y_1(4) = \frac{119}{27}, \\ y_2(4) = 0; \end{cases}$$

Shunday qilib, Xemming algoritmi x vektorni birinchi etalon sinfiga tayinlaydi.

15. YECHIMLAR DARAXTINI QURISH JARAYONI

Daraxt yechimlar usuli (decision trees) sinflashtirish va bashoratlash masalalarni yechishning eng mashhur usullaridan biri bo'lib hisoblanadi.

Ba'zan Data Miningning ushbu usuli sinflashtirish va regressiya daraxtlari bilan hal qiluvchi qoida daraxtlari deb ham nomlanadi.

Agar bog'liq, ya'ni maqsad o'zgaruvchi diskret qiymatlarni qabul qilsa, daraxt yechimlar usuli yordamida tasniflash masalasi hal qilinadi.

Agar bog'liq o'zgaruvchi uzluksiz qiymatlarni qabul qilsa, u holda daraxt yechim mustaqil o'zgaruvchilardan olingan ushbu o'zgaruvchi bog'lanishini aniqlaydi, ya'ni sonli bashoratlash masalasini hal qiladi.

Daraxt yechimning eng sodda ko'rinishida – ushbu uslub ierarxik, ketma-ket tuzilishdagi qoidalarni berilishidir. Bunday tuzilishning negizi - qator savollarga “Ha” yoki “Yo'q” javoblar.

Tarmoqlanishni binarli daraxtlar tugunlarida faqatgina ikkita yo'nalishni yurgizish mumkin. Ya'ni qo'yilgan savolga (“ha” va “yo'q”) faqat ikkita javob berish imkoniyati mavjud. Binar daraxt daraxt yechimlarning eng sodda, xususiy holi bo'lib hisoblanadi. Qolgan hollarda javoblar ham daraxtning ichki tugunidan chiquvchi daraxt shoxlariga muvofiq holda ikkitadan ko'p bo'lishi mumkin.

Yanada murakkabroq misolni ko'rib chiqamiz. Bashoratlashni amalga oshirish kerak bo'lgan asosidagi ma'lumotlar bazasi ma'lumotlarni atributlari bo'lib hisoblangan bank mijozlari to'g'risidagi quyidagi retrospektiv ma'lumotlardan iborat: yosh, afzalligini mavjudligi, ma'lumot, o'rta oyli kirim, kreditni mijoz o'z vaqtida qaytarishligi. Masala yuqorida sanab o'tilgan ma'lumotlar asosida (oxirgi atributdan tashqari) yangi mijozga kredit berish arziyasligini aniqlashdan iborat.

Bunday masala ikki bosqichda hal qilinadi: tasniflash modelini qurish va undan foydalanish.

Modelni qurish bosqichida xususan, tasniflash daraxti ham quriladi va bir qancha qoidalar to'plami yaratiladi. Belgilangan mijoz uchun qoidalar to'plami bo'lib hisoblangan qurilgan daraxt modeldan foydalanish bosqichida, yoki uning ildizidan uchlardan biriga yo'ldan “Kredit berish kerakmi?” qo'yilgan savoliga javob berish uchun foydalaniladi.

“Agar u holda” ko'rinishida berilgan mantiqiy konstruktsiya qoida bo'lib hisoblanadi.

Daraxtni yechim usuli “sodda” yondashuv deb ataladi. Biroq qator afzalliklar sababli berilgan usul tasniflash masalasini yechish uchun eng tanilgan usullardan biri bo'lib hisoblanadi.

Daraxt yechimlarning afzalliklari.

Daraxt yechimlarni ichki his-tuyg'ularga oidligi. Daraxt yechim ko'rinishida tasvirlangan tasniflash modeli ichki his-tuyg'ularga oid bo'lib hisoblanadi va yechilayotgan masalani tushunishni osonlashtiradi.

Daraxt yechimlarni qurish algoritmlarini ishi natijasi o'zi bilan “qora qutilar” ni tasvirlovchi neyron to'rlaridan farqli foydalanuvchi uchun oson talqin etiladi. Daraxt yechimlarning ushbu xossasi yangi obyektning belgilangan sinfiga kiritishda nafaqat muhim, balki butunlay tasniflash modelini talqin etishda ham foydalidir. Daraxt yechim nima uchun aniq obyekt u yoki boshqa sinfga tegishli bo'lishini tushunish va tushuntirishga imkon beradi.

Daraxt yechim tabiiy tilda ma'lumotlar bazasidan qoida chiqarish imkonini beradi. Qoida misoli: Agar yosh > 35 va Kirim > 200, u holda kredit berish.

Daraxt yechim tasniflash modelini shunday sohalarda yaratishga imkon beradiki, bu yerda tahlil qiluvchi uchun bilimlarni shakllantirish etarlicha murakkab bo'lib hisoblanadi.

Daraxt yechimni qurish algoritmi foydalanuvchidan kiruvchi atributlarini (mustaqil o'zgaruvchilar) tanlashni talab qilmaydi. Algoritmga kirishda barcha atributlarni uzatish mumkin, algoritm o'zi ularning ichidan eng muhimini tanlab oladi, va ular faqatgina daraxtni qurish uchun foydalaniladi. Neyron to'rlari bilan solishtirishda, neyron to'rlarida kiruvchi atributlar miqdorini tanlash o'qitish vaqtiga sezilarli ta'sir ko'rsatishi sababli daraxt yechim foydalanuvchi ishini ancha yengillashtiradi.

Daraxt yechim yordamida yaratilgan modellar aniqligi boshqa tasniflash modellarini qurish usullari bilan solishtirilgan.

Tez o'qitish jarayoni. Daraxt yechimlarni qurish algoritmlari yordamida tasniflash modellarini qurishga neyron to'rlarini o'qitishga nisbatan sezilarlicha kam vaqt talab qilinadi.

Aksariyat daraxt yechimlarni qurish algoritmlari tushirib qoldirilgan qiymatlarni maxsus qayta ishlash imkoniyatiga ega bo'ladi.

Tasniflash masalalarini hal qiluvchi ko'plab klassik statistik usullar faqatgina sonli ma'lumotlar bilan ishlashi mumkin, va ayni paytda daraxt yechimlar ham sonli, ham ma'lumotlarning kategoriyali turi bilan ishlashi mumkin.

Ko'pgina statistik usullar parametrikdir va foydalanuvchi avvaldan ma'lum ma'lumotlarga ega bo'lishi kerak, masalan, model turini bilishi, o'zgaruvchilar orasidagi bog'lanishlar ko'rinishi to'g'risida farazga ega bo'lishi, ma'lumotlar qanday taqsimlash turiga ega ekanligini taxmin qilishi kerak. Daraxt yechimlar bunday usullardan farqli ravishda noparametrik modellarni quradi.

Shunday qilib, daraxt yechimlar Data Miningning bunday masalalarini hal qilish imkoniyatiga ega, bunda o'rganilayotgan ma'lumotlar orasidagi bog'lanishlar turi to'g'risidagi aprior ma'lumotlar mavjud emas.

Daraxt yechimlarni qurish jarayoni.

Ko'rib chiqilayotgan tasniflash masalasi ba'zan induktiv o'qitish deb nomlanuvchi o'qituvchili o'qitish strategiyasiga tegishli. Bunday hollarda o'quv ma'lumotlar bazasidagi barcha obyektlar avvaldan belgilangan sinflardan biriga tayinlanadi.

Daraxt yechimlarni qurish algoritmlari daraxtning "qurish" yoki "yaratish" (tree building) va daraxtni "qisqartirish" (tree pruning) bosqichlaridan iborat. Daraxtni yaratish jarayonida bo'linish mezonini tanlash va o'qitishni to'xtatish masalasi yechiladi (agar bu algoritmda nazarda tutilgan bo'lsa). Daraxtni qisqartirish bosqichida uning ba'zi shoxlarini kesish masalasi yechiladi.

Bo'linish mezonlari.

Daraxtni yaratish jarayoni yuqoridan pastgacha amalga oshirilgan, ya'ni pastga. Jarayon davomida algoritm shunday bo'linish mezonini topishi kerak, ba'zan uni bo'linish mezonini deb ham ataladi, bunda berilgan tekshirish tuguniga birlashuvchi to'plamni qism to'plamlarga bo'linadi. Har bir tekshiruv tuguni belgilangan atribut bilan belgilanishi kerak. Atributni tanlash qoidasi mavjud: u dastlabki ma'lumotlar to'plamini shunday tarzda bo'lishi kerakki, ushbu bo'linish natijasida olinuvchi qism to'plamlar obyektlari bir xil sinf vakillari yoki shunday bo'linishga imkon qadar yaqinroq bo'lsin.

Oxirgi jumla shuni bildiradiki, "aralashmalar" deb nomlanuvchi boshqa

sinflar obyektlari miqdori har bir sinfda minimumga intilishi kerak.

Bo'linishning turlicha mezonlari mavjud. Eng tanilganlari – entropiya o'lchami va Gini indeksi.

Bo'linish atributlarini tanlash uchun ayrim usullarda atributlarni qism fazolari informativligi o'lchamidan foydalaniladi, bu entropiya yondashuviga asoslanadi va "informatson yutish o'lchami" (information gain measure) yoki entropiya o'lchami nomi bilan ma'lum.

Breyman(Breiman) va boshqalar tomonidan taklif etilgan boshqa bo'linish mezonlari CART algoritmidan amalga oshirilgan va Gini indeksi deb nomlanadi. Ushbu indeks yordamida atributsinflarni taqsimlanishlari orasidagi masofalarga asoslangan holda tanlanadi. Agar T to'plam, shu jumladan n sinflardan olingan misollar, indeks Gini berilgan bo'lsa, ya'ni gini(T) quyidagi formula bo'yicha aniqlanadi:

$$gini(T) = 1 - \sum_{j=1}^n p_j^2$$

bu yerda T-joriy tugun, p_j - T tugunda j sinf ehtimolligi, n-sinflar miqdori.

Daraxt yechimlarda qanchalik ko'p xususiy hollar bayon qilingan bo'lsa, har bir xususiy holda shunchalik kam miqdordagi obyektlar uchraydi. Bunday daraxtlar "shoxli" yoki "butoqli" deb ataladi, ular asossiz ravishdagi ko'p sonli tugunlar va shoxlardan iborat, dastlabki to'plam juda kam sonli obyektlardan iborat ko'p sonli qism to'plamlarga bo'linadi. Bunday daraxtlarni "toshib ketishi" natijasida ularning umumlashtirishga bo'lgan qobiliyati pasayadi, va qurilgan modellar to'g'ri javoblarni berish mumkin bo'la olmaydi.

Daraxtni qurish jarayonida uning o'lchamini haddan ziyod kattalashib ketmasligi uchun "mos o'lchamlar" deb nomlanuvchi optimal daraxtlarni yaratishga imkon beruvchi maxsus protseduralardan foydalaniladi.

Qanday o'lchamli daraxtni optimal deb hisoblash mumkin? Daraxt tadqiq etilayotgan ma'lumotlar to'plamidan olingan ma'lumotlarni hisobga olish uchun yetarlicha murakkab bo'lishi kerak, biroq bir vaqtning o'zida yetarlicha oddiy bo'lishi zarur. Boshqacha aytganda daraxt model sifatini yaxshilovchi ma'lumotdan foydalanishi kerak va uni yaxshilamaydigan ma'lumotlarga e'tibor qilmasligi kerak. Bu yerda ikkita mumkin bo'lgan strategiya mavjud. Birinchisi foydalanuvchi tomonidan berilgan parametrlarga muvofiq ma'lum

o'Ichamgacha daraxtni o'stirishdan iborat.

Ushbu parametrlarni aniqlash tahlil qiluvchining tajribasi va ichki his-tuyg'ulariga, shuningdek daraxt yechimlarni quruvchi tizimning ayrim "tashxisga oid xabarlariga" asoslanadi.

Ikkinchi strategiya daraxtning "mos o'Ichamini" aniqlovchi bir qator protseduralardan foydalanishdan iborat. Haddan ziyod katta daraxtlarni paydo bo'lishiga yo'l qo'ymaslik uchun foydalaniluvchi protseduralar quyidagilardan iborat: shoxlarni kesish yo'li bilan daraxtni qisqartirish; o'qitishni to'xtatish qoidalaridan foydalanish.

Ta'kidlash kerakki, daraxtni qurishda barcha algoritmlar ham bitta sxema bo'yicha ishlayvermaydi. Ayrim algoritmlar ikkita alohida ketma-ket bosqichlarni o'z ichiga oladi: daraxtni qurish va uni qisqartirish; boshqalari ichki tugunlarni shakllanishiga yo'l qo'ymaslik uchun ish jarayonida ushbu bosqichlarni o'zaro almashadilar.

Daraxtni qurishni to'xtatish.

To'xtatish qoidasini ko'rib chiqamiz. Ko'rib chiqilayotgan tugun ichki tugun bo'lib hisoblanadimi, bunda u bo'lingan bo'ladimi, yoki u oxirgi tugun bo'lib hisoblanadimi, ya'ni yechim tugun bo'lishini aniqlash lozim.

To'xtatish-daraxt qurish jarayonidagi shunday o'zakki, bunda keyingi shoxlanish to'xtatiladi.

To'xtatish qoidalari variantlaridan biri – "erta to'xtash" (prepruning) bo'lib, bu tugunni taqsimlashni maqsadga muvofiqligini belgilaydi. Bunday variantdan foydalanish afzalligi – modelni o'qitishga sarflanadigan vaqtni kamaytirish. Biroq bu yerda tasniflash aniqligini kamaytirish xavf-xatari tug'iladi. Shuning uchun ham to'xtatish o'miga «kesish»dan foydalanish mazmunga ega bo'ladi.

O'qitishni to'xtatishni ikkinchi varianti – daraxtning chuqurligini cheklash. Bu holatda qurish berilgan chuqurlikka erishilganda tugallanadi.

To'xtatishni yana bir varianti – daraxtning oxirgi tugunlarida mavjud bo'lgan minimal miqdordagi misollar. Ushbu variantda shoxlanish daraxtning barcha oxirgi tugunlari toza bo'lmagunicha yoki belgilangan miqdordagi obyektlardan iborat bo'lmagunicha davom etadi.

Daraxtni qisqartirish yoki shoxlarni kesib tashlash.

O'ta shoxli daraxt masalasini yechimi bo'lib ayrim shoxlarni kesish (pruning) yo'li bilan uni qisqartirish hisoblanadi.

Daraxt yechimlar yordamida qurilgan tasniflash modelining sifati ikkita asosiy belgilar bilan tavsiflanadi: tanib olish aniqligi va xatoligi.

Tanib olish aniqligi o'qitishda ishtirok etuvchi ma'lumotlar to'plamini obyektlarini umumiy miqdoriga o'qitish jarayonidagi to'g'ri sinflashtirilgan obyektlar munosabati sifatida hisoblanadi.

Xatolik o'qitishda ishtirok etuvchi ma'lumotlar to'plamini obyektlarning umumiy miqdoriga o'qitish jarayonida noto'g'ri sinflashtirilgan obyektlar munosabati sifatida hisoblanadi.

Shoxlarni kesish yoki ayrim shoxlarni pastki qism bilan almashtirish agar ushbu protsedura xatoni o'sishiga olib kelmasa amalga oshirilishi kerak. Jarayon pastdan yuqoriga o'tadi, ya'ni ko'tariluvchi bo'lib hisoblanadi.

Ayrim shoxlarni kesishdan keyin hosil qilingan daraxtlar kesilgan deb nomlanadi. Agar bunday kesilgan daraxt hali ham intuitiv bo'lib hisoblanmasa va tushunish uchun qiyin bo'lsa, sinflarni tavsiflash uchun to'plamlarda birlashtirilgan qoidalarni olishdan foydalaniladi.

Daraxt ildizidan uning uchigacha yoki bargigacha bo'lgan har bir yo'l bitta qoidani beradi. Qoida shartlari bo'lib daraxtning ichki tugunlaridagi tekshirish hisoblanadi.

Algoritmlar.

Hozirgi kunda daraxt yechimlarni amalga oshiruvchi ko'p sonli algoritmlar mavjud: CART, C4.5, CHAID, CN2, NewId, ITrule va boshqalar.

Algoritm CART (Classification and Regression Tree), nomidan ko'rinib turganidek, tasniflash va regressiya masalalarini hal qiladi. Ma'lumotlar to'plami atributlari ham diskret, ham sonli qiymat bo'lishi mumkin. Algoritm CART algoritmi binar daraxt yechimlarni qurish uchun mo'ljallangan. CART algoritmining xususiyatlari:

- bo'lish sifatini baholash funksiyasi;
- daraxtni kesish mexanizmi;
- tushirib qoldirilgan qiymatlarni qayta ishlash algoritmi;
- regressiya daraxtlarini qurish.

Binar daraxtning har bir tuguni bo'linishida faqatgina ikkita avlodga ega

bo'lib, u qizlar shoxlari deb ataladi. Shoxlarning keyingi bo'linishi berilgan shoxlar dastlabki ma'lumotlar tomonidan tavsiflanganligiga bog'liq. Daraxtni qurishning har bir qadamida tugunda shakllanuvchi qoida misollarni berilgan to'plamini ikkita qismga bo'ladi. Uning o'ng qismi (shox right) bu to'plamning shunday qismiki, qoida bajariladi; chap (shox left) — esa, qoida bajarilmaydi.

Optimal qoidani tanlash uchun foydalaniluvchi bo'lish sifatini baholash Gini indeksiga asoslangan. Ushbu baholash funksiyasi tugunda noaniqlikni kamaytirish g'oyasiga asoslangan. Aytaylik, tugun bor va ikkita sinfga bo'lingan. Tugundagi maksimal noaniqlik uni 50% misollar bo'yicha ikkita qism to'plamga bo'lganda, maksimal aniqlik esa - 100% va 0% misollarga bo'lganda erishiladi.

Bo'lish qoidalari. CART algoritmi sonli va kategoriyali atributlar bilan ishlaydi. Agar sonli bo'lsa, u holda ichki tugunda atribut $x_i \leq c$ ko'rinishdagi qoida shakllanadi. «c» qiymat ko'p hollarda o'qitiluvchi ma'lumotlar to'plamini x_i o'zgaruvchisini ikkita qo'shni tartiblangan o'rta arifmetik qiymatlari sifatida tanlanadi. Agar atribut kategorial turga tegishli bo'lsa, u holda ichki tugunda $x_i \in V(x_i)$ qoida shakllanadi, bu yerda $V(x_i)$ — o'qitiluvchi ma'lumotlar to'plamida x_i o'zgaruvchining qiymatlari to'plamini ayrim bo'sh bo'lmagan qism to'plamlari.

Kesish mexanizmi. Minimal cost-complexity tree pruning nomga ega ushbu mexanizm bilan, CART algoritmi iprinsipial jihatdan daraxt yechimlarni qurishning boshqa algoritmlaridan farq qiladi. Ko'rib chiqilayotgan kesish algoritmidagi — bu “mos o'lchamli” daraxtni olish va tasniflash bahosini aniqroq olish o'rtasidagi kompromissdir. Usul kamayib boruvchi daraxtlar ketma-ketligini olishdan iborat, biroq daraxtlarni hammasini emas, balki faqatgina “eng yaxshi yakillar”ni ko'rib chiqiladi.

Chorahali tekshiruv (V-fold cross-validation) birmuncha murakkabroqdir va bir paytni o'zida CART algoritmini original qismi bo'lib hisoblanadi. U o'zi bilan yakuniy daraxtni tanlash yo'lini ifodalaydi, shunday shartdagi, ma'lumotlar to'plami katta bo'lmagan hajmga ega bo'ladi yoki ma'lumotlar to'plamining yozuvlari qanchalik o'ziga xos bo'lishi, bunda to'plamni o'qitish va tekshirish tanlanmasiga ajratish mumkin bo'lmaydi.

CART algoritmining asosiy xususiyatlari: binar bo'linish, bo'linish mezonlari - Gini indeksi, minimal cost-complexity tree pruning va V-fold cross-validation algoritmlari, “daraxtni o'stirish, keyin esa qisqartirish” tamoili,

qurishni yuqori tezligi, tushirib qoldirilgan qiymatlarni qayta ishlash.

C4.5-algoritmi tugundagi cheklanmagan sondagi shoxlar bilan daraxt yechimlarni quradi. Berilgan algoritmi faqatgina diskret bog'liq atributlar bilan ishlashi mumkin va shuning uchun ham faqat tasniflash masalalarini yechish mumkin bo'ladi. C4.5 tasniflash daraxtlarini qurishda eng tanilgan va keng foydalaniluvchi algoritmlardan biri bo'lib hisoblanadi.

C4.5-algoritmi ishlashi uchun quyidagi talablarni amal qilish zarurdir:

Ma'lumotlar to'plamini har bir yozuvi oldindan belgilangan sinflardan biri bilan bog'lanishi kerak, ya'ni ma'lumotlar to'plamining atributlaridan biri sinf metkasi bo'lib hisoblanadi.

Sinflar diskret bo'lishi lozim. Har bir misol sinflardan biriga bir qiymatli tegishli bo'lishi zarur.

Sinflar miqdori tadqiq etilayotgan ma'lumotlar bazasida yozuvlar miqdoridan sezilarli kam bo'lishi kerak.

C4.5-algoritmi juda katta hajmdagi va shovqinli ma'lumotlar bazasida sekin ishlaydi.

Ko'rib chiqilayotgan ikkala algoritmi ham mustahkam, ya'ni shovqin va ma'lumot chiqarib tashlovchilarga chidamli bo'lib hisoblanadi.

Daraxt yechimlarni qurish algoritmlari quyidagi xususiyatlari bilan farqlanadi:

- bo'linish turi - binar (binary), ko'plik (multi-way);
- bo'linish mezonlari - entropiya, Gini, boshqalar;
- tushirib qoldirilgan qiymatlarni qayta ishlash imkoniyati;
- shoxlarni qisqartirish va kesish protsedurasi;
- daraxtlardan qoidalarni olish imkoniyati.

Hech qanday daraxtni qurish algoritmi apriori eng yaxshi yoki mukammal bo'lib hisoblanishi mumkin emas, ma'lum bir algoritmdan foydalanishning maqsadga muvofiqligini tasdiqlanishi tajriba bilan tekshirilgan va tasdiqlangan bo'lishi kerak.

16. TAYANCH VEKTORLAR TUSHUNCHASI

Tayanch vektorlar usuli (mashinasi) (SVM – Support Vector Machine) V. Vapnikning qator ishlarida va o'tgan asrning 70-yillarida ishlab chiqilgan. Ushbu usul timsollarni ayrim ma'nosida eng yaxshi hal qiluvchi funksiya yordamida sinflarga ajratishning o'qituvchili o'qitish usulidir. Ikkita sinfni ajratish tezligi bo'yicha tayanch vektorlar usuli (TVU) eng yaxshi usul bo'lib hisoblanadi.

16.1. Chiziq bilan ajratiluvchi holat

Ayrim o'quv tanlanmasi $\Omega = \{x_1, \dots, x_m\}$ mavjud bo'lsin, bundan tashqari $\{x_1, \dots, x_k\} \subset \omega_1$ va $\{x_{k+1}, \dots, x_m\} \subset \omega_2$, bu yerda ω_1 va ω_2 – ikkita sinf. O'quv tanlanmasi bo'yicha chiziqli hal qiluvchi funksiyani qurish talab qilinadi, ya'ni shunday chiziqli $f(x)$ funksiyaki, quyidagi shartni qoniqtirsin:

$$f(x_i) > 0 \quad \forall x_i \in \omega_1, \quad (16.1)$$

$$f(x_i) \leq 0 \quad \forall x_i \in \omega_2.$$

O'quv tanlanma elementlarini birlashtirish uchun u yoki boshqa sinf tanlanuvchi elementlarini tegishlilik koeffitsientlarini hisobga olamiz.

$$y_i = \begin{cases} 1, & x_i \in \omega_1, \\ -1, & x_i \notin \omega_2. \end{cases}$$

Shunday qilib, quyida berilgan shartni qoniqtiruvchi, chiziqli hal qiluvchi $f(x)$ funksiyani topishni talab qiladi

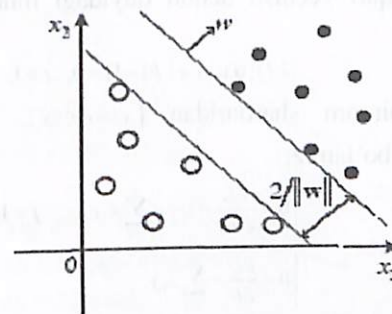
$$y_i f(x_i) > 0 \quad \forall x_i \in \Omega. \quad (16.2)$$

f funksiyani ayrim musbat songa ko'paytirib ko'rish mumkinki, tengsizliklar sistemasi teng kuchli sistemadir

$$y_i f(x_i) \geq 1 \quad \forall x_i \in \Omega. \quad (16.3)$$

Shuningdek $f(x)$ – chiziqli funksiya bo'lsa, u holda tenglamalar sistemasi (16.3) quyidagi ko'rinishni oladi

$$y_i((w, x_i) + b) \geq 1, \quad i = 1, \dots, m, \quad (16.4)$$



16.1-rasm.

bu yerda w – vazn koeffitsientlari vektori, b – ayrim son. U holda ikkita sinfga bo'linuvchi gipertekislik $(w, x_i) + b = 0$ bo'ladi. Ko'rish qiyinmaski, barcha gipertekislik ko'rinishi ham $(w, x_i) + b' = 0$ ko'rinishda bo'ladi, bu yerda $b' \in (b-1, b+1)$, ham taqsimlanuvchi bo'ladi (16.1-rasm). $(w, x) + b - 1 = 0$ va $(w, x) + b + 1 = 0$ chekli gipertekisliklar oraliq masofasi $2/\|w\|$ teng. Haqiqatan

ham $\left(\frac{w}{\|w\|}, x\right) + \frac{b+1}{\|w\|} = 0$ – ushbu gipertekislikning normal tenglamalaridir. U

holda $p_1 = \frac{b-1}{\|w\|}$ va $p_2 = \frac{b+1}{\|w\|}$ – ushbu gipertekislikdan koordinata boshigacha

bo'lgan masofalar, $p_2 - p_1 = 2/\|w\|$ esa – gipertekisliklar orasidagi masofa. Chegara tekisliklarining o'zida ayrim sonli (ikkidan kam bo'lmagan) o'qitiluvchi vektorlar joylashishi mumkin. Ushbu vektorlar tayanch vektorlar deb nomlanadi.

Sinflarni ishonchli ajratish uchun ajratuvchi gipertekisliklar orasidagi masofa imkon qadar katta bo'lishi, ya'ni $\|w\|$ kattalik imkon qadar kichik bo'lishi zarur. Shunday qilib, (16.4) tengsizliklar sistemasi bilan beriladigan qabariq ko'pburchakda (differensiallashga qulaylik bo'lishi uchun 0,5 koeffitsient kiritiladi) kvadrat funksionalning 0,5(w,w) minimumini topish masalasi qo'yilayapti. Qabariq to'plamda kvadratik funksional har doim yagona minimal qiymatga ega (agar ushbu to'plam bo'sh bo'lmasa). Ushbu optimallashtirish masalasi yechimi Lagranj ko'paytuvchilari $\lambda_i \geq 0 (i = 1, \dots, m)$ bo'yicha ortantlarda Lagranj kritik nuqtalarini topishga tengdir.

$$L(w, b, \lambda) = 0,5(w, w) - \sum_{i=1}^m \lambda_i (y_i((w, x_i) + b) - 1)$$

Bundan tashqari yechish uchun quyidagi munosabatlarni bajarishimiz zarur

$$\lambda_i(y_i((w, x_i) + b) - 1) = 0, \quad i = 1, \dots, m.$$

Zaruriy minimum shartlaridan $[x_i = (x_{i1}, x_{i2}, \dots, x_{im})]$ deb taxmin qilsak] quyidagilarga ega bo'lamiz:

$$\begin{cases} 0 = \frac{\partial L}{\partial w_j} = w_j - \sum_{i=1}^m \lambda_i y_i x_{ij}, & j = 1, \dots, n; \\ 0 = \frac{\partial L}{\partial b} = \sum_{i=1}^m \lambda_i y_i, \end{cases}$$

Bundan kelib chiqadiki, w vektorini quyidagi ko'rinishda izlanadi:

$$w = \sum_{i=1}^m \lambda_i y_i x_i, \quad (16.5)$$

$$\sum_{i=1}^m \lambda_i y_i = 0 \quad (16.6)$$

λ_i nolli koeffitsientlarga ega (16.5) yig'indiga $y_i((w, x_i) + b) - 1 = 0$ faqatgina tenglik uchun to'g'ri bo'lgan vektorlar, ya'ni tayanch vektorlar kiradi. Agar topilgan w qiymatni ko'rib chiqsak, b joy almashtirishni (смешение) x ni ixtiyoriy tayanch vektori s uchun $b = y_s^{-1} - (w, x_s)$ sifatida hisoblash mumkin.

Lagranj ko'paytuvchilari qiymatini Lagranj kritik nuqtalari sifatida topamiz. Buning uchun (16.5) va (16.6) ni Lagranjga qo'yamiz va quyidagilarni olamiz:

$$\begin{aligned} L(w, b, \lambda) &= 0,5(w, w) - \sum_{i=1}^m \lambda_i (y_i((w, x_i) + b) - 1) = 0,5(w, w) - \left((w, w) - \sum_{i=1}^m \lambda_i \right) = \\ &= \sum_{i=1}^m \lambda_i - 0,5(w, w) = \sum_{i=1}^m \lambda_i - 0,5 \sum_{i=1}^m \lambda_i \lambda_j y_i y_j (x_i, x_j) = \sum_{i=1}^m \lambda_i - 0,5 \left\| \sum_{i=1}^m \lambda_i y_i x_i \right\|^2. \end{aligned}$$

Shunday qilib, masala funksiyani kritik nuqtalarini topishga kelib qoladi.

$$\Phi(\lambda) = \sum_{i=1}^m \lambda_i - 0,5 \left\| \sum_{i=1}^m \lambda_i y_i x_i \right\|^2 \quad (16.7)$$

Shuningdek, (16.7) funksiya o'zida chiziqli va kvadratik funksiyalar farqini ifodalaydi, bundan tashqari kvadratik funksiya manfiy aniqlangan, u

holda $\lambda_i \geq 0$ ($i = 1, \dots, m$) sohada $\sum_{i=1}^m \lambda_i y_i = 0$ shartda $\Phi(\lambda)$ funksiyani eng katta qiymatini talab etiladi. Ushbu masalani yechish uchun ko'pgina algoritmlar (optimallashtirish nazariyasida) mavjud (masalan, gradient usullar, koordinatani tushirish usuli va boshqalar).

Izohlar.

1. (16.7) dagi yig'indi barcha vektorlar bo'yicha emas, balki o'qitiluvchilariga nisbatan anchagina kichik bo'lishi mumkin bo'lgan tayanch vektorlar bo'yicha amalga oshiriladi.

2. Natijada chiziqli hal qiluvchi funksiya quyidagi ko'rinishni oladi:

$$f(x) = \sum_{i=1}^m \lambda_i y_i (x_i, x) + y^{-1} - \sum_{i=1}^m \lambda_i y_i (x_i, x_i),$$

bu yerda λ_i faqat y_i ga va skalyar ko'paytirish (x_i, x_i) qiymatlariga bog'liq, bundan tashqari yig'indi faqat tayanch vektorlar bo'yicha amalga oshiriladi.

3. $f(x)$ funksiya hisoblangandan so'ng, x vektor ω_1 sinfga kiritiladi, agar $f(x) > 0$ bo'lsa, ω_2 sinfga kiritiladi, agar, $f(x) < 0$. Noto'g'ri sinflashtirish $p(x)$ ehtimolligini $t \rightarrow \infty$ da $\phi(0) = 0,5, \phi(t) \rightarrow 0$ shartni qoniqtiruvchi ayrim uzluksiz kamayuvchi funksiya $\phi(t)$ bilan baholash mumkin. U holda $p(x)$ ehtimolligi $\phi(p(x, L_i))$ ga teng, agar $x \in \omega_i$ ($i = 1, 2$) bo'lsa, bu yerda $L_i : (w, x) + b + \text{sgn}(\alpha - i) = 0, \quad 1 < \alpha < 2$. ya'ni

$$p(x) = \phi \left(\left(\frac{w}{\|w\|} \cdot x \right) + \frac{b + \text{sgn}(\alpha - i)}{\|w\|} \right), \quad \text{agar } x \in \omega_i, \quad (i = 1, 2).$$

Misol. Tayanch vektorlar usuli bilan $\omega_1 = \{x_1\}$ va $\omega_2 = \{x_2, x_3\}$ sinflarni taqsimlang, agar $x_1 = (1, 1), x_2 = (1, 2), x_3 = (2, 3)$ bo'lsa.

Yechish. $\Phi(\lambda)$ funksiya quyidagi ko'rinishni oladi:

$$\Phi(\lambda) = \sum_{i=1}^m \lambda_i - 0,5 \sum_{i,j=1}^m \lambda_i \lambda_j y_i y_j (x_i, x_j) = \lambda_1 + \lambda_2 + \lambda_3 - 0,5(2\lambda_1^2 + 5\lambda_2^2 + 13\lambda_3^2 - 6\lambda_1 \lambda_2 - 10\lambda_1 \lambda_3 + 16\lambda_2 \lambda_3),$$

Bundan tashqari $\lambda_1 - \lambda_2 - \lambda_3 = 0 \Rightarrow \lambda_3 = \lambda_1 - \lambda_2$. U holda

$$\Phi(\lambda_1, \lambda_2) = 2\lambda_1 - 2,5\lambda_1^2 - \lambda_2^2 + 3\lambda_1 \lambda_2.$$

$\Phi(\lambda_1, \lambda_2)$ funksiya uchun normal sistemani tuzamiz va yechamiz

$$\begin{cases} \partial \Phi / \partial \lambda_1 = 0, \\ \partial \Phi / \partial \lambda_2 = 0 \end{cases} \Leftrightarrow \begin{cases} 2 - 5\lambda_1 + 3\lambda_2 = 0, \\ -2\lambda_2 + 3\lambda_1 = 0 \end{cases} \Leftrightarrow \begin{cases} \lambda_1 = 4, \\ \lambda_2 = 6. \end{cases}$$

Natijada, $\lambda_1 = 4$, $\lambda_2 = 6$, $\lambda_3 = -2$. Shuningdek $\lambda_3 < 0$, u holda $\lambda_3 = \lambda_1 - \lambda_2$ shart bilan $\lambda_i \geq 0$ ($i=1,2,3$) chegarada $\Phi(\lambda)$ funksiyani tadqiq qilamiz.

Agar $\lambda_1 = 0$, u holda $\lambda_3 = -\lambda_2 \Rightarrow \lambda_i^{(1)} = 0$ ($i=1,2,3$) $\Rightarrow \Phi(\lambda^{(1)}) = 0$.

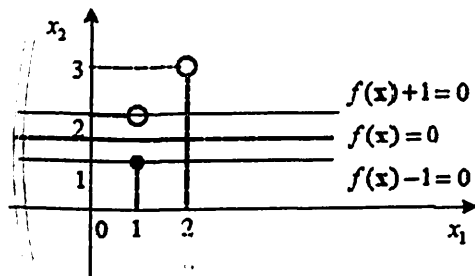
Faraz qilaylik, $\lambda_2 = 0$. U holda $\lambda^{(2)} = 2/5$ da $\lambda_1 = \lambda_3 = \lambda$ va $\Phi(\lambda) = 2\lambda - 2,5\lambda^2$, $\Phi'(\lambda) = 0$ Natijada, $\lambda_1^{(2)} = \lambda_3^{(2)} = 2/5$, $\lambda_2^{(2)} = 0$ va $\Phi(\lambda^{(2)}) = 2/5$.

Agar $\lambda_3 = 0$ bo'lsa, u holda $\lambda^{(3)} = 2$ da $\lambda_1 = \lambda_2 = \lambda$ va $\Phi(\lambda) = 2\lambda - 0,5\lambda^2$, $\Phi'(\lambda) = 0$ Shuning uchun, $\lambda_1^{(3)} = \lambda_2^{(3)} = 2$, $\lambda_3^{(3)} = 0$ va $\Phi(\lambda^{(3)}) = 2$.

Shunday qilib, $\lambda_3 = \lambda_1 - \lambda_2$ shart bilan $\lambda_i \geq 0$ ($i=1,2,3$) sohada $\Phi(\lambda)$ funksiya eng katta qiymatga $\lambda^{(3)} = (2, 2, 0)^T$ nuqtada erishadi. Ushbu holda

$$w = \sum_{i=1}^3 \lambda_i y_i x_i = 2x_1 - 2x_2 = 2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \end{bmatrix}.$$

$$b = \frac{1}{y_1} - (w, x_1) = 1 - (0 - 2) = 3$$



16.2-rasm.

Shunday qilib, $f(x) = (w, x) + b = -2x^{(2)} + 3$ va $f(x) = 0 \Leftrightarrow x^{(2)} = 1,5$. Taqsimlanuvchi yo'llar kengligi $h = 2/\|w\| = 2/2 = 1$, to'g'ri chiziqlar esa $f(x) + 1 = 0 \Leftrightarrow f(x) = 0 \Leftrightarrow x^{(2)} = 2$ va $f(x) - 1 = 0 \Leftrightarrow x^{(2)} = 1$ uning chegaralari bo'ladi.

16.2. Chiziqli ajralmas holat

Ushbu holatda $\phi: R^n \rightarrow H$ akslantirish yordamida katta o'lcham H fazosida R^n belgilar fazosini qo'yish zarur. U holda tayanch vektorlar algoritmini ko'rib chiqqan holda o'quv tanlanma $\phi(x_i)$ timsollari uchun masala yechimini chiziqli ajratiladigan holatiga keltiramiz, ya'ni ajratish funksiyasini quyidagi ko'rinishda izlaymiz:

$$f(x) = (w, \phi(x)) + b, \quad w = \sum_{i=1}^m \lambda_i y_i \phi(x_i).$$

bu yerda λ_i ko'effitsientlar y_i ga va $(\phi(x_i), \phi(x_j))$ qiymatlarga bog'liq. Shunday qilib, hal qiluvchi funksiyani topish uchun $(\phi(x_i), \phi(x_j))$ skalyar ko'paytma qiymatlari bo'lishi zarur. Yadro deb nomlanuvchi $K(x, y) = (\phi(x), \phi(y))$ funksiya uchun quyidagi xossa o'rinalidir:

1) $K(x, y) = K(y, x)$ (simmetriklik);

2) matritsa, $K = (K_{i,j})$, $K_{i,j} = K(x_i, x_j)$ ixtiyoriy vektor x_1, \dots, x_m uchun manfiy bo'lmagan aniqdir (ya'ni $K = (K_{i,j})$) - $(\phi(x_i))$ vektorlar sistemasini Grama matritsasi.

Misollar: 1) $K(x, y) = x, y$; 2) $K(x, y) = \phi(x), \phi(y)$; 3) $K(x, y) = C > 0$.

Ixtiyoriy $m+1$ vektorlarni ixtiyoriy ikkita sinfga m dan katta bo'lmagan monomial akslantirish darajasi yordamida ajratish mumkin. Shuning uchun ham agar $\phi: x \rightarrow \{x_1^i, x_2^i, \dots, x_n^i\}$, $i_1 + \dots + i_n \leq m$ shunday akslantirish bo'lsa, u holda ushbu akslantirishga mos keluvchi yadroni quyidagi ko'rinishda izlash mumkin:

$$K(x, y) = (\phi(x), \phi(y)) = ((x, y) + 1)^m.$$

16.3-rasm

Shunday qilib, ushbu yadro har qanday $m+1$ vektorlarni ixtiyoriy ikkita sinfga ajratishni kafolatlaydi. Ushbu holda ajratuvchi funksiyalarni topish quyidagi tarzda amalga oshiriladi:

1) $\lambda_i \geq 0$ ($i=1, \dots, m$) sohada $\sum_{i=1}^m \lambda_i y_i = 0$ shart bilan

$\Phi(\lambda) = \sum_{i=1}^m \lambda_i - 0,5 \sum_{i=1}^m \lambda_i \lambda_j y_i y_j K(x_i, x_j)$ eng katta qiymatini topamiz, natijada vektor

$$\lambda^{(0)} = (\lambda_1^{(0)}, \dots, \lambda_m^{(0)})$$
 ni hosil qilamiz.

2) ajratuvchi funksiyani quyidagi ko'rinishda topamiz:

$$f(x) = (w, \phi(x)) + b = \left(\sum_{i=1}^m \lambda_i^0 y_i \phi(x_i), \phi(x) \right) + b = \sum_{i=1}^m \lambda_i^0 y_i (\phi(x_i), \phi(x)) + y_r^- -$$

$$\sum_{i=1}^m \lambda_i^0 y_i (\phi(x_i), \phi(x)) = \sum_{i=1}^m \lambda_i^0 y_i K(x_i, x) + y_r^- - \sum_{i=1}^m \lambda_i^0 y_i K(x_i, x_i).$$

Misol. Tayanch vektorlar usuli bilan

$\omega_1 = \{x_1, x_2\}$ va $\omega_2 = \{x_3\}$ sinflarga ajrating, agar $x_1 = (0, 0)$, $x_2 = (2, 0)$, $x_3 = (1, 0)$

bo'lsa (16.3-rasm).

Yechish. Shuningdek o'quv tanlanma vektorlari miqdori uchta bo'lsa, u

Natijada, $\lambda_1 = 4$, $\lambda_2 = 6$, $\lambda_3 = -2$. Shuningdek $\lambda_3 < 0$, u holda $\lambda_3 = \lambda_1 - \lambda_2$ shart bilan $\lambda_i \geq 0$ ($i=1,2,3$) chegarada $\Phi(\lambda)$ funksiyani tadqiq qilamiz.

Agar $\lambda_1 = 0$, u holda $\lambda_3 = -\lambda_2 \Rightarrow \lambda_i^{(1)} = 0$ ($i=1,2,3$) $\Rightarrow \Phi(\lambda^{(1)}) = 0$.

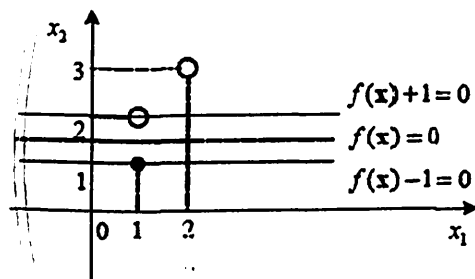
Faraz qilaylik, $\lambda_2 = 0$. U holda $\lambda^{(2)} = 2/5$ da $\lambda_1 = \lambda_3 = \lambda$ va $\Phi(\lambda) = 2\lambda - 2,5\lambda^2$, $\Phi'(\lambda) = 0$ Natijada, $\lambda_1^{(2)} = \lambda_3^{(2)} = 2/5$, $\lambda_2^{(2)} = 0$ va $\Phi(\lambda^{(2)}) = 2/5$.

Agar $\lambda_3 = 0$ bo'lsa, u holda $\lambda^{(3)} = 2$ da $\lambda_1 = \lambda_2 = \lambda$ va $\Phi(\lambda) = 2\lambda - 0,5\lambda^2$, $\Phi'(\lambda) = 0$ Shuning uchun, $\lambda_1^{(3)} = \lambda_2^{(3)} = 2$, $\lambda_3^{(3)} = 0$ va $\Phi(\lambda^{(3)}) = 2$.

Shunday qilib, $\lambda_3 = \lambda_1 - \lambda_2$ shart bilan $\lambda_i \geq 0$ ($i=1,2,3$) sohada $\Phi(\lambda)$ funksiya eng katta qiymatga $\lambda^{(3)} = (2, 2, 0)^T$ nuqtada erishadi. Ushbu holda

$$w = \sum_{i=1}^3 \lambda_i y_i x_i = 2x_1 - 2x_2 = 2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \end{bmatrix}$$

$$b = \frac{1}{y_1} - (w, x_1) = 1 - (0 - 2) = 3$$



16.2-rasm.

Shunday qilib, $f(x) = (w, x) + b = -2x^{(2)} + 3$ va $f(x) = 0 \Leftrightarrow x^{(2)} = 1,5$. Taqsimlanuvchi yo'llar kengligi $h = 2/\|w\| = 2/2 = 1$, to'g'ri chiziqlar esa $f(x) + 1 = 0 \Leftrightarrow f(x) = 0 \Leftrightarrow x^{(2)} = 2$ va $f(x) - 1 = 0 \Leftrightarrow x^{(2)} = 1$ uning chegaralari bo'ladi.

16.2. Chiziqli ajralmas holat

Ushbu holatda $\phi: R^n \rightarrow H$ akslantirish yordamida katta o'lcham H fazosida R^n belgilar fazosini qo'yish zarur. U holda tayanch vektorlar algoritmini ko'rib chiqqan holda o'quv tanlanma $\phi(x_i)$ timsollari uchun masala yechimini chiziqli ajratiladigan holatiga keltiramiz, ya'ni ajratish funksiyasini quyidagi ko'rinishda izlaymiz:

$$f(x) = (w, \phi(x)) + b, \quad w = \sum_{i=1}^m \lambda_i y_i \phi(x_i),$$

bu yerda λ_i koefitsientlar y_i ga va $(\phi(x_i), \phi(x_i))$ qiymatlarga bog'liq. Shunday qilib, hal qiluvchi funksiyani topish uchun $(\phi(x_i), \phi(x_j))$ skalyar ko'paytma qiymatlari bo'lishi zarur. Yadro deb nomlanuvchi $K(x, y) = (\phi(x), \phi(y))$ funksiya uchun quyidagi xossa o'rinalidir:

1) $K(x, y) = K(y, x)$ (simmetriklik);

2) matritsa, $K = (K_{i,j})$, $K_{i,j} = K(x_i, x_j)$ ixtiyoriy vektor x_1, \dots, x_m uchun manfiy bo'lmagan aniqdir (ya'ni $K = (K_{i,j})$) - $(\phi(x_i))$ vektorlar sistemasini Grama matritsasi.

Misollar: 1) $K(x, y) = x, y$; 2) $K(x, y) = \phi(x), \phi(x)$; 3) $K(x, y) = C > 0$.

Ixtiyoriy $m+1$ vektorlarni ixtiyoriy ikkita sinfga m dan katta bo'lmagan monomial akslantirish darajasi yordamida ajratish mumkin. Shuning uchun ham agar $\phi: x \rightarrow \{x_1^i, \dots, x_n^i\}$, $i_1 + \dots + i_n \leq m$ shunday akslantirish bo'lsa, u holda ushbu akslantirishga mos keluvchi yadroni quyidagi ko'rinishda izlash mumkin:

$$K(x, y) = (\phi(x), \phi(y)) = ((x, y) + 1)^m.$$

16.3-rasm

Shunday qilib, ushbu yadro har qanday $m+1$ vektorlarni ixtiyoriy ikkita sinfga ajratishni kafolatlaydi. Ushbu holda ajratuvchi funksiyalarni topish quyidagi tarzda amalga oshiriladi:

1) $\lambda_i \geq 0$ ($i=1, \dots, m$) sohada $\sum_{i=1}^m \lambda_i y_i = 0$ shart bilan

$$\Phi(\lambda) = \sum_{i=1}^m \lambda_i - 0,5 \sum_{i=1}^m \lambda_i \lambda_j y_i y_j K(x_i, x_j)$$
 eng katta qiymatini topamiz, natijada vektor

$$\lambda^{(0)} = (\lambda_1^{(0)}, \dots, \lambda_m^{(0)})$$
 ni hosil qilamiz.

2) ajratuvchi funksiyani quyidagi ko'rinishda topamiz:

$$f(x) = (w, \phi(x)) + b = \left(\sum_{i=1}^m \lambda_i^0 y_i \phi(x_i), \phi(x) \right) + b = \sum_{i=1}^m \lambda_i^0 y_i (\phi(x_i), \phi(x)) + y_r^- -$$

$$\sum_{i=1}^m \lambda_i^0 y_i (\phi(x_i), \phi(x)) = \sum_{i=1}^m \lambda_i^0 y_i K(x_i, x) + y_r^- - \sum_{i=1}^m \lambda_i^0 y_i K(x_i, x_i).$$

Misol. Tayanch vektorlar usuli bilan

$\omega_1 = \{x_1, x_2\}$ va $\omega_2 = \{x_3\}$ sinflarga ajrating, agar $x_1 = (0, 0)$, $x_2 = (2, 0)$, $x_3 = (1, 0)$

bo'lsa (16.3-rasm).

Yechish. Shuningdek o'quv tanlanma vektorlari miqdori uchta bo'lsa, u

holda $m=2$, yadro sifatida $K(x, y) = ((x, y) + 1)^2$ funksiyani olamiz. U holda

$$\Phi(\lambda) = \lambda_1 + \lambda_2 + \lambda_3 - 0,5(\lambda_1^2 + 25\lambda_2^2 + 4\lambda_3^2 + 2\lambda_1\lambda_2 - 2\lambda_1\lambda_3 - 18\lambda_2\lambda_3) = |\lambda_1 + \lambda_2 = \lambda_3| = 2\lambda_1 + 2\lambda_2 + \lambda_3 - 0,5(3\lambda_1^2 + 11\lambda_2^2 + 4\lambda_3^2 + 10\lambda_1\lambda_2)$$

Normal sistemani tuzamiz:

$$\begin{cases} \partial\Phi / \partial\lambda_1 = 0, \\ \partial\Phi / \partial\lambda_{21} = 0 \end{cases} \Leftrightarrow \begin{cases} 2 - 3\lambda_1 + 5\lambda_2 = 0, \\ 2 - 11\lambda_2 + 5\lambda_1 = 0 \end{cases} \Leftrightarrow \begin{cases} \lambda_1^{(0)} = 4, \\ \lambda_2^{(0)} = 2, \\ \lambda_3^{(0)} = 6. \end{cases}$$

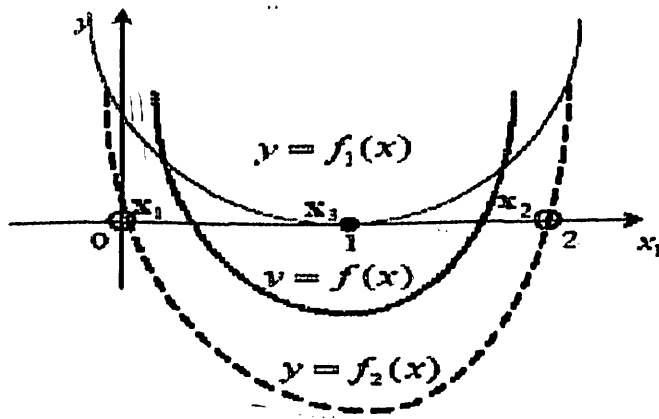
Aytish mumkinki, $\lambda^{(0)}$ nuqtada $\lambda_1 + \lambda_2 = \lambda_3$ shartda $\lambda_i \geq 0 (i=1,2,3)$ sohada $\Phi(\lambda)$ funksiya eng katta qiymatiga erishiladi. Natijada,

$$f(x) = 4K(x_1, x) + 2K(x_2, x) + 6K(x_3, x) + 1 - (4K(x_1, x_1) + 2K(x_2, x_1) - 6K(x_1, x_1)) = 4 \cdot 1 + 2(2x_1 + 1)^2 - 6(x_1 + 1)^2 + 1 - (4 + 2 - 6) = 2x_1^2 - 4x_1 + 1.$$

Shunday qilib, $f(x) = 2x_1^2 - 4x_1 + 1$, $f(x) = 0 \Leftrightarrow x_1 = 1 \pm \frac{\sqrt{2}}{2}$.

$f(x) = f(x) + 1 = (x_1 - 1)^2$ $f_2(x) = f(x) - 1 = 2x_1(x_1 - 2)$. U holda

$f_1(x) = 0 \Leftrightarrow x_1 = 1$, $f_2(x) = 0 \Leftrightarrow x_1 = 0$ yoki $x = 2$. Agar $(x_1, x_2) \rightarrow (x_1, x_2, y)$ qoida va x Oy tekislikdagi oxirgi proeksiya bo'yicha belgilarni ikki o'lchamli fazosini uch o'lchamli fazosiga qo'yishni ko'rib chiqilsa, u holda funksiya grafiklari $y = f(x_1)$, $y = f_1(x_1)$ ($i=1,2$) bilan o'quv tanlanmani ajratishni tasvirlash mumkin (16.4-rasm).



16.4-rasm.

17. "Eng yaqin qo'shni" usuli

a) x , o'quv tanlanmada berilgan x timsolga eng yaqin bo'lgan element aniqlanadi, ya'ni

$$\|x - x_i\| = \min \{\|x - x_i\| : i = 1, \dots, n\}.$$

b) Shart tekshiriladi: agar $x_i \in \omega_i$, u holda $x \in \omega_i$ bo'lib hisoblanadi. Ushbu holatda masofa funksiyasi quyidagi formula bo'yicha aniqlanadi:

$$d(x, \omega_i) = \min \|x - x_k\|, x_k \in \omega_i.$$

17.1. Timsoldan etalon timsolgacha bo'lgan masofani aniqlash

Ushbu uslub etalon timsollarga kirish va sinflarni tuzilmaviy va geometrik xossalari bilimlariga asoslangan. Bunda atrofi o'quv tanlanma boshqa timsollarini shakllantiruvchi o'quv tanlanma elementlari (shunday ataluvchi mos etalon timsollar) ayrim to'plamni ajratiladi. Bunday sinf ichidagi tuzilmalar klaster deb nomlanuvchi bir etalon timsoli doirasini guruhlarga ajratadi.

Agar timsollar ma'lum bo'lsa, u holda berilgan x timsol va ω_i sinf orasidagi masofani quyidagi formula bo'yicha aniqlash mumkin

$$d(x, \omega_i) = \min \|x - c_k\|, c_k \in \omega_i,$$

Bu yerda ω_i - etalon timsol. Klasterda o'quv tanlanmani taqsimlash masalasi klasterlash deb ataladi. Klasterlash algoritmi quyida keltiriladi.

Vektor kvantlash g'oyasi tanib olish masalalarini hal qilish uchun klasterning barcha nuqtalari ajralib turmasligi uchun belgilar maydonini taqsimlashdan (yoki o'quv tanlanmani) iborat. Klasterlar miqdori qoidaga ko'ra sinflar sonlari sezilarlicha katta.

Sinflarni klasterlarga taqsimlash masalalari deganda $\{x_1, \dots, x_n\} = \mu$, o'quv tanlanmadan shunday S_1, \dots, S_p qismto'plamni ajratish masalasi tushuniladiki, quyidagi shartlar bajariladi:

$$S_1 \cup \dots \cup S_p = \mu; S_i \cap S_j = \emptyset \text{ barcha } i \neq j \text{ uchun.}$$

Ushbu qism to'plamlar tez-tez foydalanib turiladigan ayrim optimal xossalarga ega, masalan klasterlar dispersiyasining minimalligi, ya'ni uning markazi atrofida klaster elementlarining tarqalish darajasini tavsiflovchi belgi (etalon timsol). Ixtiyoriy klasterdagi etalon timsol - klasterning markazidir:

$$c_i = \frac{1}{|S_i|} \sum_{x_k \in S_i} x_k.$$

Klaster markazi atrofi elementlarini taqsimlash kattaligi (klaster dispersiyasi) quyidagi formula bo'yicha hisoblanishi mumkin:

$$\Delta_i = \frac{1}{|S_i|} \sum_{x_k \in S_i} \|x_k - c_i\|^2.$$

bu yerda N_i - i-klasterdagi elementlar soni.

Elementlar to'plamini klasterlarga taqsimlashni muvaffaqiyatli deb hisoblash mumkin, agar $Q = \sum_{i=1}^p \Delta_i$ funksional minimal bo'lsa.

17.2. k-guruh ichidagi o'rtacha algoritmi (k-means)

Ushbu algoritim o'zi bilan klaster markazlarini bosqichma-bosqich (iterativ) topish va Q funksional kamayguncha o'quv tanlanmani klasterlarga bo'lishni ifodalaydi.

1. O'quv tanlanmadan bir qancha timsollarni belgilaymiz va ularni etalon timsollar - $c_1^{(0)}, \dots, c_p^{(0)}$ klasterlar markazi deb nomlaymiz.

2. Barcha yaqin qo'shnilar usuli bo'yicha tanlanmani p klasterlarga (Voronoy katakchasi) taqsimlaymiz. Bir qancha klasterlar $S_1^{(0)}, \dots, S_p^{(0)}$ hosil bo'ladi.

3. $Q_{(0)}$ funksional qiymatini hisoblaymiz.

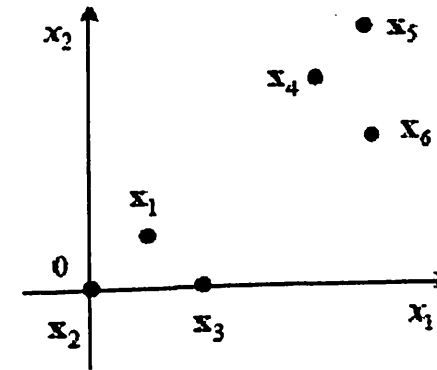
4. Klasterlarni yangi markazlarini 2 punktga o'tgan holda $c_i^{(1)} = \frac{1}{|S_i^{(0)}|} \sum_{x_k \in S_i^{(0)}} x_k$ formula bo'yicha hisoblaymiz. Ushbu algoritimning to'xtash sharti bo'lib $Q_{(k)} = Q_{(k-1)}$ tenglikni k-qadamda bajarish bo'lib hisoblanadi.

Izoh.

1. To'xtash sharti o'rnida $Q_{(k)} = Q_{(k-1)}$ boshqa shartlardan ham foydalanish mumkin, masalan, klaster markazlarini o'zgarish holati sharti.

2. k-means algoritmi Q funksionalni global emas, balki lokal minimallashtirishni amalga oshiradi. Shuning uchun ham "yaxshi" klasterlash kafolatini ushbu algoritim bermaydi.

1. n nuqtalarni k klasterlarga bo'lish soni $n^k/k!$ ga nisbatan katta. Shuning uchun ham k-means algoritmi juda sekin birlashishi mumkin.



17.1-rasm.

Misol. 2R tekislikda $x_1(1;1)$, $x_2(0;0)$, $x_3(2;0)$, $x_4(4;4)$, $x_5(5;5)$, $x_6(5;3)$ nuqtalar - timsollar, (17.1-rasm) ushbu timsollarni ikkita sinf bo'yicha klasterizatsiyasi berilgan bo'lsin. Buning uchun ko'rib chiqilgan algoritimning ketma-ket qadamlarini bajaramiz.

1. Klasterlarni boshlang'ich markazlari sifatida $c_1^{(0)} = x_1$ va $c_2^{(0)} = x_2$ timsollarni tanlab olamiz. U holda $\{x_1, \dots, x_6\}$ tanlanmani yaqin qo'shnilar usuli bilan ikkita qism to'plamga bo'lgan holda boshlang'ich klasterlarni olamiz

$$S_1^{(0)} = \{x_1, x_3, x_4, x_5, x_6\} \text{ va } S_2^{(0)} = \{x_2\}.$$

2. Klasterlarni yangi markazlarini hisoblaymiz:

$$c_1^{(1)} = \frac{1}{5} \begin{bmatrix} x_{11} + x_{31} + x_{41} + x_{51} + x_{61} \\ x_{12} + x_{32} + x_{42} + x_{52} + x_{62} \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 1+2+4+5+5 \\ 1+0+4+5+3 \end{bmatrix} = \begin{bmatrix} 3.4 \\ 2.6 \end{bmatrix}; \quad c_2^{(1)} = x_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

3. $c_1^{(0)} \neq c_1^{(1)}$ va $c_2^{(0)} = c_2^{(1)}$ ni qiyoslaymiz. Algoritimni bajarishda davom etamiz.

4. $\{x_1, \dots, x_6\}$ tanlanmani yaqin qo'shnilar usuli bo'yicha yangi markazlarga ega ikkita qism to'plamga taqsimlaymiz, quyidagi klasterlarga ega bo'lamiz:

$$S_1^{(1)} = \{x_4, x_5, x_6\} \text{ va } S_2^{(1)} = \{x_1, x_2, x_3\}.$$

5. Klasterlarni yangi markazlarini qaytadan hisoblaymiz:

$$c_1^{(2)} = \frac{1}{3} \begin{bmatrix} x_{41} + x_{51} + x_{61} \\ x_{42} + x_{52} + x_{62} \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 4+5+5 \\ 4+5+3 \end{bmatrix} = \begin{bmatrix} 14/3 \\ 4 \end{bmatrix}; \quad c_2^{(2)} = \frac{1}{3} \begin{bmatrix} x_{11} + x_{21} + x_{31} \\ x_{12} + x_{22} + x_{32} \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1+0+2 \\ 1+0+0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1/3 \end{bmatrix}$$

6. $c_1^{(1)} \neq c_1^{(2)}$ va $c_2^{(1)} \neq c_2^{(2)}$ ni qiyoslaymiz. Algoritimni bajarishni davom ettiramiz.

7. $\{x_1, \dots, x_6\}$ tanlanmani yaqin qo'shnilar usuli bo'yicha yangi markazlarga

ega ikkita qism to'plamga taqsimlaymiz, quyidagi klasterlarga ega bo'lamiz
 $S_1^{(1)} = \{x_1, x_2, x_3\}$ aa $S_2^{(1)} = \{x_1, x_2, x_3\}$.

8. Klasterlarni yangi markazlarini qaytadan hisoblaymiz
 $c_1^{(3)} \neq c_1^{(2)}$ aa $c_2^{(3)} \neq c_2^{(2)}$. Algoritmni to'xtatish.

Algoritmni amaliy jihatdan amalga oshirishda quyidagi muammolar yuzaga keladi:

- 1) klasterlar sonini topish;
- 2) algoritmning ish sifati klasterlar markazlarini boshlang'ich joylashuviga bog'liq.

18. BAYES TAQSIMOTI

18.1. Bayes klassifikatori

Qoidaga ko'ra o'quv tanlanma elementlarini tabiiy sharoitlarda olish mumkin. U yoki boshqa timsolni paydo bo'lishi tasodifiy hodisa bo'lib, uning paydo bo'lish ehtimolini ehtimollikni taqsimlash zichlik funksiyasi yordamida tavsiflash mumkin. Ushbu funksiyaning o'zi tanib olish tizimi ishlaydigan aniq muhitni belgilaydi. O'quv tanlanma elementlarini bilgan holda, haqiqatda ushbu muhitni ehtimolli xususiyatlarini tiklash mumkin. Bu xususiyatlar quyidagilar jumlasidandir:

- belgilar (timsollar) vektorlarini paydo bo'lishni ehtimollikni taqsimlash zichlik funksiyasi $f(x)$;
- sinflar ichida belgilarni ehtimolliklarini taqsimlash zichliklarini shartli funksiyalari $\varpi, f_i(x) = f(x|\varpi_i)$;
- sinflarni paydo bo'lish ehtimolligi $p_i = p(\varpi_i)$;
- ayrim timsolni paydo bo'lishida sinflarni paydo bo'lishning shartli ehtimolligi $p = p(\varpi_i|x^0)$.

Misol. Yog'ochni qayta ishlash fabrikasida yog'ochni tanib olish tizimini ko'rib chiqamiz. O'ylaymizki, fabrikaga daraxtning turi bo'yicha turlicha taxtalar kelgan bo'lsin. Tanib olish tizimi u yoki boshqa daraxt turi bo'yicha bir qancha xususiyatlari bilan tanib olishi va taxtalarni saralab olishi zarur. Bunday xususiyatlar sifatida tanlash mumkin bo'ladi: x_1 – yog'och rangi; x_2 – yog'och qattiqligi; x_3 – yog'ochning o'ziga xos zichligi. U holda $p = p(\varpi_i|x^0) \frac{1}{2}$ ehtimollikni ko'rib chiqish mumkinki, $x^0 = (x_1, x_2, x_3)$ xususiyatlarga ega bo'lgan berilgan yog'och i –daraxt turlarining yog'och partiyasida paydo bo'lish ehtimolligi ϖ_i ni yog'och turi bo'lib hisoblanadi.

Muhitning ko'rsatilgan ehtimolli xususiyatlari quyidagi formulalar bilan bog'liq:

– to'liq ehtimolli formula: $f(x) = \sum f_i(x)p_i$

– Bayes formulasi:

$$p(\varpi_i|x) = \frac{p_i f_i(x)}{f(x)}$$

(18.1)

Ushbu formulalarda ϖ_i sinflar hodisalarning to'liq guruhini tasvirlaydi

(faraz bo'lib hisoblanuvchi), ya'ni $\omega_i \cap \omega_j = \emptyset \quad i \neq j, \quad p(\omega_1) + p(\omega_2) + \dots + p(\omega_N) = 1$

Misol. ω_1 va ω_2 ikkita kesishmaydigan sinf mavjud bo'lsin, shuningdek berilgan ω_i sinfdagi x timsollarni ajratishni shartli zichligi $f_i(x) = \frac{1}{(2\pi\sigma_i^2)^{n/2}} e^{-\frac{(x-c_i)^2}{2\sigma_i^2}}$ ($i=1,2$) ga teng bo'lsin. $f_i(x)$ funksiyaning c_i va σ_i^2 parametrlarini $\{x_1, x_2, \dots, x_N\}$ tasodifiy tanlanma bo'yicha baholash maksimal o'xshashlik usuli bilan quyidagicha amalga oshiriladi:

– tarqalish markazini baholash:

$$c_i \approx \frac{1}{N_i} \sum_{k=1}^{N_i} x_k$$

– dispersiyani baholash

$$\sigma_i^2 \approx \frac{1}{N_i} \sum_{k=1}^{N_i} \|x_k - c_i\|^2$$

(18.2)

U yoki boshqa sinfni paydo bo'lish ehtimolligini $p(\omega_i) \approx N_i / N$ formula bo'yicha baholaymiz, bu yerda $N_i - \omega_i$ ga tegishli elementlar soni.

Keyin (18.1) va (18.2) formulalar bo'yicha topish va muhit xususiyatini baholash mumkin.

Masalani qo'yilishi. Noto'g'ri sinflashtirishni xato ehtimolligi minimal bo'lishi uchun ω_1 yoki ω_2 sinfni x kuzatish vektorining tegishlilikini aniqlash talab qilinadi. R_n belgilar fazosini shunday ikkita kesishmaydigan Ω_1 va Ω_2 sohalarga ajratamizki, $x \in \omega_i$, agar $x \in \Omega_i$ ($i=1,2$) bo'lsa. U holda noto'g'ri sinflashtirish xato ehtimolligini quyidagi formula bo'yicha hisoblash mumkin:

$$Q = \int_{\Omega_1} p(\omega_2|x)f(x)dx + \int_{\Omega_2} p(\omega_1|x)f(x)dx$$

(18.3)

(18.3) formuladagi birinchi qo'shiluvchi ω_2 sinf x vektorining tegishlilik ehtimoliga teng bo'ladi, agar aslida u esli ω_1 sinfga tegishli bo'lsa (ikkinchi turdagi xato Q_2 ehtimoli deb nomlanuvchi ehtimollik). Ikkinchi qo'shiluvchi – birinchi tur xato Q_1 ehtimolligi. (18.3) kattalik - noto'g'ri sinflashtirishni o'rtacha xato.

Bundan kelib chiqadiki, eng yaxshi klassifikatorni qurish masalasi noto'g'ri sinflashtirishni Q o'rtacha xatoligini minimallashtiruvchi

kesishmaydigan ikkita Ω_1 va Ω_2 sohaga R_n fazoni shunday taqsimlashga ekvivalent bo'ladi. Bayes sinflashtirish bilan muvofiqlikda ko'rsatildi

$$x \in \omega_1, \text{ agar } p_1 f_1(x) > p_2 f_2(x) \text{ va } x \in \omega_2, \text{ agar } p_1 f_1(x) < p_2 f_2(x).$$

Ushbu masala va uning yechimi u yoki boshqa sinflashtirish uchun sinflarni va turlicha jarimalarni ixtiyoriy holati uchun umumlashtiriladi.

Misol. ω_i ($i=1,2$) sinflar ichidagi belgilarni taqsimlashning shartli zichligi $x \geq 0, a_i > 0$ ($i=1,2$) da $f_i(x) = a_i e^{-a_i x}$ ko'rinishga ega bo'lsin (shunday deb nomlanuvchi taqsimlash ko'rsatkichi). U holda $x \in \omega_1$ bayes sinflashtirish qoidasi bilan muvofiqlikda, agar

$$a_1 e^{-a_1 x} p_1 > a_2 e^{-a_2 x} p_2 \Leftrightarrow x > \frac{1}{a_2 - a_1} \ln \left(\frac{a_2 p_2}{a_1 p_1} \right) \quad a_2 > a_1 \text{ bo'lganda yoki}$$

$$x < \frac{1}{a_2 - a_1} \ln \left(\frac{a_2 p_2}{a_1 p_1} \right) \quad a_2 < a_1 \text{ bo'lganda.}$$

18.2. Sinflashtirishning minimaks mezoni

Bayes sinflashtirish statistik nuqtai nazardan eng yaxshisi bo'lib hisoblanadi. Biroq u har doim ham qo'llanilmaydi. Masalan, p_i ehtimollik noma'lum bo'lganda u qo'llanilmaydi.

Noto'g'ri sinflashtirishni o'rtacha xatosini yozamiz

$$Q = p_1 \int_{\Omega_1} f_1(x) dx + p_2 \int_{\Omega_2} f_2(x) dx$$

Faraz qilaylik, $\int_{\Omega_1} f_1(x) dx \geq \int_{\Omega_2} f_2(x) dx$ bo'lsin. U holda

$$Q = (p_1 + p_2) \int_{\Omega_1} f_1(x) dx = \int_{\Omega_1} f_1(x) dx. \text{ Agar } \int_{\Omega_1} f_1(x) dx \leq \int_{\Omega_2} f_2(x) dx \text{ bo'lsa, u holda}$$

$$Q \leq (p_1 + p_2) \int_{\Omega_2} f_2(x) dx = \int_{\Omega_2} f_2(x) dx.$$

Shunday qilib,

$$Q \leq \max \left\{ \int_{\Omega_1} f_1(x) dx, \int_{\Omega_2} f_2(x) dx \right\}$$

Shuning uchun ham agar sinflarning aprior ehtimolligini paydo bo'lishi noma'lum bo'lsa, u holda sinflashtirilgan strategiya sifatida noto'g'ri sinflashtirishning maksimal mumkin bo'lgan o'rtacha xatosini minimallashtirishda Ω_1 va Ω_2 kesishmaydigan sohaga R_n belgilar fazosini shunday taqsimlash strategiyasini tanlash mumkin bo'ladi, ya'ni

$$\bar{Q}(\Omega_1, \Omega_2) = \max \left\{ \int_{\Omega_1} f_1(x) dx, \int_{\Omega_2} f_1(x) dx \right\} \rightarrow \min.$$

Misol. Faraz qilaylik, belgilar $x \geq 0$, $a_i > 0$ ($i=1,2$) bo'lganda $f_i(x) = a_i e^{-a_i x}$ isbotlangan qonunga ko'ra ω_i ($i=1,2$) ikkita sinfga ajratilgan bo'lsin.

$\Omega_1 = \{x < x_0\}$ ko'rinishdagi Ω_1 sohani izlaymiz. U holda $Q_2 = \int_0^{x_0} a_2 e^{-a_2 x} dx = 1 - e^{-a_2 x_0}$

, $Q_1 = \int_{x_0}^{\infty} a_1 e^{-a_1 x} dx = e^{-a_1 x_0}$. Funktsiya minimumi $\bar{Q}(x_0) = \max \{e^{-a_1 x_0}, 1 - e^{-a_2 x_0}\}$ nuqtasini

topamiz. Ko'rish qiyinmaski, ushbu nuqta sonli hisoblash mumkin bo'lgan $e^{-a_1 x_0} = 1 - e^{-a_2 x_0}$ tenglama ildizi bo'lib hisoblanadi. Xususiyl holda,

$a_1 = a_2 = a$, $x_0 = \frac{1}{a} \ln 2$ bo'lganda

18.3. Neyman – Pirson mezon

Ushbu mezon p_i sinflarni paydo bo'lish ehtimolligi noma'lum bo'lganda ham qo'llaniladi. Sinflashtirishni ikkita sinfga bo'yicha ko'rib chiqamiz. Ushbu holda muhimlik darajasi deb nomlanuvchi ayrim kichik musbat kattalik belgilanadi va $\alpha = \int_{\Omega_1} f_1(x) dx$ birinchi tur xato ehtimolligi deb nomlanuvchisiga

sonli teng bo'ladi (haqiqatda x ni ω_1 sinfga tegishli bo'lgan paytda ω_2 sinfga x ning munosabati ehtimolligi). α belgilangan qiymatda Ω_1 va Ω_2 kesishmaydigan sohada R_n belgili fazoni bo'linishi Neyman – Pirson mezoniga muvofiq amalga oshiriladi, shunday qilib, ikkinchi turdagi xato ehtimolligini minimallashtirish uchun

$$Q_2 = \int_{\Omega_2} f_2(x) dx \rightarrow \min.$$

Teorema (Neyman – Pirsonning). Ikkinchi turdagi xato ehtimolligi Q_2 minimal bo'ladi, agar

$$\Omega_1 = \left\{ x \in R^n : \frac{f_1(x)}{f_2(x)} > h \right\}; \quad \Omega_2 = \left\{ x \in R^n : \frac{f_1(x)}{f_2(x)} < h \right\},$$

bu yerda bo'sag'aviy kattalik h quyidagi munosabatlardan aniqlanadi:

$$\alpha = \int_{\Omega_1} f_1(x) dx, \quad \Omega_2(h) = \left\{ x \in R^n : r(x) < h \right\}, \quad r(x) = \frac{f_1(x)}{f_2(x)}.$$

Misol. ω_i ($i=1,2$) sinflar ichida belgilarni taqsimlashni shartli zichliklari $x \geq 0$, $a_i > 0$ ($i=1,2$) da ko'rsatkichli taqsimlash ko'rinishiga ega bo'lsin. Agar

$a_1 > a_2$ bo'lsa, u holda soha quyidagichadir:

$$\Omega_1 = \left\{ x : \frac{f_1(x)}{f_2(x)} > h \right\} = \left\{ x : \frac{a_1}{a_2} e^{(a_2 - a_1)x} > h \right\} = \left\{ x < \frac{1}{a_2 - a_1} \ln \left(\frac{a_2 h}{a_1} \right) \right\}.$$

h kattalikni $\alpha = \int_{\Omega_1(h)} a_1 e^{-a_1 x} dx$ tenglamadan quyidagi shartlar bo'yicha aniqlaymiz:

$$\Omega_2(h) = \left\{ x : \frac{f_1(x)}{f_2(x)} < h \right\} = \left\{ x > \frac{1}{a_2 - a_1} \ln \left(\frac{a_2 h}{a_1} \right) \right\}.$$

U holda

$$\alpha = \int_{\frac{1}{a_2 - a_1} \ln \left(\frac{a_2 h}{a_1} \right)}^{\infty} a_1 e^{-a_1 x} dx = \exp \left(\frac{a_1}{a_1 - a_2} \ln \left(\frac{a_2 h}{a_1} \right) \right),$$

bundan

$$h = \frac{a_1}{a_2} \alpha^{\frac{a_2 - a_1}{a_1}} \text{ va } \Omega_1 = \left\{ x < -\frac{\ln \alpha}{a_1} \right\}.$$

Umumiy holda

$$\Omega_1 = \left\{ x < -\frac{\ln \alpha}{\max(a_1, a_2)} \right\}.$$

Sodda-bayes yondashuvi quyidagi kamchiliklarga ega:

barcha kiruvchi o'zgaruvchilar haqiqatan statistik jihatdan mustaqil bo'lganda taqdiridagina shartli ehtimolliklarni to'g'ri ko'paytirish; statistik mustaqillik sharti bajarilmaganda berilgan usul yetarlicha yaxshi natijalar ko'rsatsa ham, bayes to'rlariga asoslangan murakkabroq usullarni nazariy jihatdan bunday holat qayta ishlashi zarur bo'ladi;

uzluksiz o'zgaruvchilarning to'g'ridan-to'g'ri qayta ishlashning iloji yo'q – atributlar diskret bo'lishi uchun interval shkalada ularni aylantirish talab qilinadi; biroq bunday aylantirishlar ba'zan muhim qonuniyatlarni yo'qolishiga olib kelishi mumkin;

Sodda-bayes yondashuvida tasniflash masalasi natijasiga faqatgina kiruvchi o'zgaruvchilarning individual qiymatlari ta'sir qiladi, bu yerda turlicha atributlarni qiymatlari juftlari uchliklarining qo'hishma ta'siri hisobga olinmaydi. Bu tasniflash modelini sifatini uning bashorat aniqligi nuqtai nazaridan yaxshilashi mumkin, biroq tekshiriluvchi variantlar miqdorini oshiradi.

19. BASHORATLASH MASALASI

Bashoratlash (forecasting), keng ma'noda ushbu so'z kelajakdagi oldindan akslanish sifatida belgilanadi. Bashoratlash maqsadi kelgusidagi voqealarni oldindan aytish bo'lib hisoblanadi.

Bashorat qilish (prognostics)- bashoratlash nazariyasi va amaliyotidir.

Bashoratlash retrospektiv ma'lumotlar asosida aniq obyekt yoki hodisaning dinamikasini tendensiyalarini aniqlashga yo'naltirilgan, ya'ni uning keyingi va hozirgi holatlarini tahlil qilish. Shunday qilib, bashoratlash masalasini yechish ma'lumotlarni ayrim o'qitiluvchi tanlanmasini talab qiladi.

Bashoratlash – bog'liq va bog'liq bo'lmagan o'zgaruvchilar o'rtasida funksional bog'lanish o'rnatish. Marketingda odatiy masala bo'lib bozorni bashorat qilish (market forecasting) hisoblanadi. Berilgan masalani yechish natijasida belgilangan bozorni kon'yunkturasini istiqboli, kelgusi davrga bozor sharoitlarini o'zgarishi baholanadi, bozor tendensiyalari aniqlanadi (tarkibiy o'zgarishlar, mijozlar talablari, narxlarning o'zgarishi).

Bashoratlash masalalarini hal qilishda eng umumiy tomonlari quyidagi qismmasalalarni yechishga keltiriladi:

- bashoratlash modelini tanlash;
- qurilgan bashoratning adekvatligi va aniqligini tahlil qilish.

Data Miningning ko'plab usullari tasniflash va bashoratlash masalalarini yechish uchun foydalaniladi: chiziqli regressiya, neyron to'rlari, daraxt yechimlar (ba'zan ularni bashoratlash va tasniflash daraxtlari deb ataladi).

Ikkala masalani yechishda o'qitiluvchi to'plam asosida modelni qurishning ikki bosqichli jarayonidan foydalaniladi va undan bog'liq o'zgaruvchini noma'lum qiymatlarini oldindan aytish uchun foydalaniladi.

Tasniflash va bashoratlash masalasining farqi shundan iboratki, birinchi masalada bog'liq o'zgaruvchini sinfi oldindan aytiladi, ikkinchisida esa – tushirib qoldirilgan yoki noma'lum (kelajakka tegishli) bog'liq o'zgaruvchining sonli qiymatlari.

Bashoratlash va davriy qatorlar

Bashoratlash uchun asos bo'lib davriy qatorlar ko'rinishidagi ma'lumotlar bazasida saqlanuvchi tarixiy axborot bo'lib xizmat qiladi.

Davriy qatorlar Data Mining tushunchasi mavjud (Time-Series Data Mining). Davriy qator – tasodifiy bo'lmagan vaqt onlarida tartiblangan

qandaydir belgining kuzatiluvchi qiymatlari ketma-ketligi.

- 1) Davriy qatorning o'ziga xos misoli – birja savdolari ma'lumotlari;
- 2) Davriy qatorni tahlil qilish quyidagi maqsadda amalga oshiriladi;
- 3) qatorning tabiatini aniqlash;
- 4) qatorning kelgusi qiymatlarini bashoratlash;

Davriy qator tuzilishini va qonuniyatlarini aniqlash jarayonida quyidagilar aniqlanadi: shovqinlar va chiqindilar, trend, mavsumiy komponentlar, siklik komponentlar. Davriy qator tabiatini aniqlash ma'lumotlarning o'ziga xos "razvedka"si sifatida foydalanilishi mumkin. Mavsumiy komponent mavjudligi to'g'risida tahlil qiluvchining bilimi bashoratning tuzishda ishtirok etadigan tanlanma yozuvlarini miqdorini aniqlash uchun zarur.

Shovqinlar va chiqindilar davriy qatorni qiyinlashtirib yuboradi. Chiqindilarni aniqlash va filtrlashning turlicha usullari mavjud, bu sifatliroq tahlil qilish maqsadida ularni istisno qilish imkonini beradi.

Davriy qatorni asosiy tarkibi trend va mavsumiy komponentdan iborat. Trend - vaqtda o'zgarishi mumkin bo'lgan davriy qatorning tizimli komponenti bo'lib hisoblanadi.

Trend davriy qatorga ta'sir qiluvchi umumiy va uzoq muddatli tendensiyalarni ta'sirida shakllanuvchi tasodifiy bo'lmagan funksiya deb nomlanadi. Davriy qatorlarda trendlarni topishning avtomatik usuli mavjud emas.

Bir seriyali qatorlarni bashoratlash guruhiga boshqa o'zgaruvchilarni va omillarni hisobga olmagan holda faqat ushbu o'zgaruvchiga retrospektiv ma'lumot bo'yicha bir o'zgaruvchini bashoratini qurish masalalari kiradi.

Qatorlarni ko'p seriyali yoki o'zaro ta'sir ko'rsatadigan bashoratlash guruhi bir yoki bir nechta o'zgaruvchilarga o'zaro ta'sir qiluvchi omillarni hisobga olish zarur bo'lgan paytdagi tahlillash masalasini o'z ichiga oladi.

Bir seriyali va ko'p seriyali sinflarga bo'lishdan tashqari qatorlar mavsumiy va mavsumiy bo'lmagan bo'lishi mumkin.

Oxirgi bo'linish davriy qatorlarda mavsumiylik kabi tarkibiy qismning mavjudligi yoki yo'qligini bildiradi, ya'ni mavsumiy komponentni kiritishni. Davriy qatorning mavsumiy tarkibiy qismi davriy qatorning takrorlanadigan tarkibiy qismidir. Mavsumiylik xususiyati taxminan teng vaqt oralig'ida bog'liq o'zgaruvchi harakatini tavsiflovchi egri shakli o'zining o'ziga xos shaklini takrorlashni anglatadi.

Agar uning tashqi ko'rinishini ko'rib chiqishda egri shaklni takrorlanishi to'g'risidagi teng vaqt oralig'ida taxminlar qilish mumkin bo'lmasa, qatorni mavsumiy emas deb hisoblash mumkin. Ba'zan qatorni egri tashqi ko'rinishi bo'yicha uning mavsumiy yoki mavsumiy emasligini aniqlash mumkin emas.

Mavsumiy ko'pqator tushunchasi mavjud. Unda har bir qator bog'liq (maqsad) o'zgaruvchiga ta'sir qiluvchi omillar harakatini tavsiflaydi.

Bashoratlash bo'yicha masalani yechish uchun ma'lumotlarni yig'ish va omillarni tanlashda shuni yodda tutish kerakki, bu yerda tovarlarni sotish hajmining ta'siri mavsumiylik omili ta'siridan ancha past.

Ko'pincha trend va mavsumiylik davriy qatorda bir vaqtning o'zida mavjud.

Siklik qator va mavsumiylikni ajratish odatiy holdir. siklning davomiyligi, haqiqatda, bir mavsumiy davrdan uzunroq, sikllar mavsumiy davrdan farqli ma'lum bir davomiylikka ega emas.

Har qanday o'zgarishlarni amalga oshirishda davriy qator tabiatini tushunish sezilarlicha osonlashadi, bunday o'zgarishlar bilan trendni olib tashlash yoki qatorni tekislash.

Bashoratlashni boshlashdan avval quyidagi savollarga javob berish zarur:

Bashoratlash uchun nima zarur?

Qanday davriy elementlarda (parametrlarda)?

Qanday bashorat aniqligi bilan?

Birinchi savolga javob berishda bashoratlanuvchi o'zgaruvchilarni aniqlaymiz. Bu, masalan, navbatdagi kvartaldagi mahsulotning aniq bir turini ishlab chiqarish darajasi, ushbu mahsulotning sotish narxining bashorati va boshqalar bo'lishi mumkin. O'zgaruvchilarni tanlashda retrospektiv ma'lumotlarning mavjudligi, qaror qabul qiluvchi shaxslarning afzalligi, Data Miningning yakuniy qiymati hisobga olinadi. Ko'pincha bashoratlash masalasini yechishda o'zgaruvchini o'zini emas, balki uning qiymatlarini o'zgarishini oldindan aytish zaruriyati tug'iladi.

Bashoratlashning ikkinchi masalasini hal qilishda quyidagi parametrlarni aniqlanadi:

1) bashoratlash davri;

2) bashoratlash gorizonti;

3) bashoratlash intervali.

4) bashoratlash davri– bashorat qiluvchi asosiy vaqt birligi.

5) bashoratlash gorizonti – kelajakda bashoratni qoplovchi davrlar soni.

6) bashoratlash intervali– yangi bashorat qilinadigan chastota.

Parametrlarni tanlashda bashoratlash gorizonti ushbu bashorat asosida qabul qilingan qarorni amalga oshirish uchun zarur bo'lgan vaqtga nisbatan kichik bo'lmasligi kerakligini hisobga olish zarur. Faqat shu holatdagina bashoratlash mantiqqa ega bo'ladi.

Bashoratlash aniqligi bashoratlash gorizonti bilan qoidaga ko'ra pasayadi, gorizontni kamayishi bilan esa – ko'tariladi.

Bashoratlash uchun yechimni amalga oshirishga zarur bo'lgan vaqtni kamaytirgan holda, shu bilan birga bashoratlash gorizontini va xatosini kamaytirgan holda ham bashoratlash sifatini oshirish mumkin.

Bashoratlash oralig'ini tanlashda ikkita xavf-xatar tanlanadi: tahlil qilinuvchi jarayondagi o'zgarishlarni vaqtida aniqlamaslik va bashoratning yuqori qiymatini. Uzun bashoratlash oralig'ida jarayonda sodir bo'lgan o'zgarishlarni aniqlamaslik xavfi tug'iladi, qisqasida esa – bashoratlash xarajatlari oshadi. Oraliqni tanlashda tahlil qilinuvchi jarayon barqarorligini va bashoratlash xarajatlarini ham hisobga olish zarur.

Bashorat aniqligi.

Muayyan masalani yechish uchun zarur bo'lgan bashorat aniqligi bashoratlash tizimiga katta ta'sir ko'rsatadi. Bashorat xatoligi foydalaniluvchi bashorat tizimiga bog'liq. Bunday tizim qanchalik ko'p resurslarga ega bo'lsa, shunchalik aniq bashoratni olish imkoniyatlari kattaroq bo'ladi. Biroq bashorat qilish qaror qabul qilishda xavf-xatarlarni butunlay yo'q qila olmaydi. Shuning uchun ham har doim mumkin bo'lgan bashorat xatosi hisobga olinadi. Bashoratning aniqligi bashorat xatoligi bilan tavsiflanadi.

Xatolarning eng keng tarqalgan turlari:

O'rtacha xato ($O'X$). Bu har bir qadamda xatolarni o'rtachasini hisoblaydi. Xatoning ushbu turining kamchiligi – musbat va manfiy xatolar bir-birini bekor qiladi.

O'rtacha absolyut xato ($O'AX$). Bu o'rtacha absolyut xatolar sifatida hisoblanadi. Agar u nolga teng bo'lsa, u holda mukammal bashoratga ega bo'lamiz. O'rtacha kvadratik xatolik bilan solishtirishda ushbu o'lcham chiqindilarga "o'ta katta qiymatni bermaydi".

Xatolar kvadratlari yig'indisi (XKY), o'rta kvadratik xato. Bu xatolar

kvadratlari yig'indisi (yoki o'rtacha) sifatida hisoblanadi. Bu eng ko'p foydalaniladigan bashorat xatoligi bahosi.

Nisbiy xato (NX). Avvalgi o'lchovlarda xatolarning haqiqiy qiymatlari foydalanilgan. Nisbiy xato nisbiy xatolar terminlarida natijaga keltirish sifatini ifodalaydi.

Bashoratlar turlari.

Bashorat qisqa muddatli, o'rta muddatli va uzoq muddatli bo'lishi mumkin.

Qisqa muddatli bashorat o'zi bilan bir nechta qadam oldingi bashoratni ifodalaydi, ya'ni bashoratni qurish kuzatish hajmining 3% dan ko'p bo'lmagan yoki 1-3 qadam oldin amalga oshiriladi.

O'rta muddatli bashorat – bu kuzatuvlar hajmining 3-5% i uchun, biroq 7-12 qadamdan oshmasligi kerak; shuningdek bashoratning ushbu turi bir yoki yarim mavsumiy siklning bashorati sifatida tushuniladi. Qisqa muddatli va o'rta muddatli bashoratlarni qurish uchun statistik usullar to'la mos keladi.

Uzoq muddatli bashorat – bunda kuzatish hajmini 5% dan ko'prog'ini bashorat qilishni tushuniladi. Bashoratlarning berilgan turini qurishda amaliy jihatdan statistik usullar foydalanilmaydi, bashoratni "tuzish" mumkin bo'lgan juda "yaxshi" qatorlar holatlari bundan mustasno.

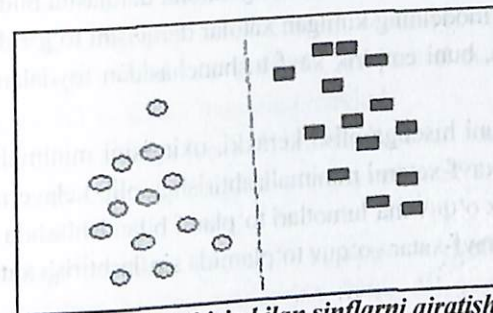
20.TASNIFLASH VA BASHORATLASH USULLARI

Tayanch vektorlar usuli

Tayanch vektorlar usuli (Support Vector Machine - SVM) chegarali usullarni guruhiga kiradi. U soha chegaralari yordamida sinflarni aniqlaydi. Berilgan usul yordamida binarli tasniflash masalasi yechiladi.

Usul asosida yechimlar tekisliklari tushunchasi yotadi.

Yechim tekisligi(plane) turlicha sinfga oid tegishlilikka ega obyektlarga ajratadi. 20.1-rasmda ikki turdagi obyektlar ishtirok etuvchi misol keltirilgan. Bo'linuvchi chiziq chegarani beradi, o'ng tomoni - brown (jigar rang-to'g'ri to'rtburchak) turining barcha obyektlari, chap tomoni esa - yellow (sariq- oval) turi. O'ng tomonga kiruvchi yangi obyekt brown yoki yellow obyekt sinfi sifatida sinflashtiriladi, agar u ajratuvchi to'g'ri chiziqning chap tomonida joylashgan bo'lsa. Bunday holda har bir obyekt ikkita o'lchov bilan tavsiflanadi.



20.1-rasm. To'g'ri chiziq bilan sinflarni ajratish

Tayanch vektorlar usulining maqsadi – obyektning ikkita to'plamini ajratuvchi tekislikni topish. Usul ikkita sinf orasidagi chegaralarda joylashgan namunalarni izlaydi, ya'ni tayanch vektorlarni.

Tayanch vektorlar bo'lib soha chegaralarida yotuvchi to'plam obyektlariga aytiladi. Agar chegaralar orasidagi soha bo'sh bo'lsa, tasniflash yaxshi deb hisoblanadi.

Chiziqli SVM.

Binarli tasniflash masalasini yechish tayanch vektorlar usuli yordamida ma'lumotlar to'plamini ikkita sinfga to'g'ri ajratuvchi ayrim chiziqli funksiyani izlashdan iborat.

Masalani bir sinf vektorlari uchun noldan kichik va boshqa sinf vektorlari

uchun noldan katta qiymatlarni qabul qiluvchi $f(x)$ funksiyani izlash sifatida shakllantirish mumkin. Qo'yilgan masalani yechish uchun dastlabki ma'lumotlar sifatida, ya'ni sinflashtiriluvchi $f(x)$ ni qidirish uchun fazo vektorlarining o'quv to'plami berilgan bo'lib, ularning sinflardan biriga tegishli ekani ma'lum.

Sinflashtiriluvchi funksiyalar oilasini $f(x)$ funksiyasi orqali tavsiflash mumkin. Gipertekislik «a» vektor «b» qiymat bilan aniqlangan, ya'ni $f(x)=ax+b$. Berilgan masalaning yechimi 20.2-rasmدا keltirilgan.

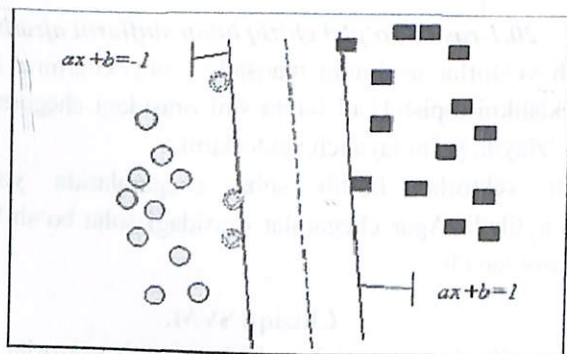
Masalani yechish ya'ni SVM-modelni qurish natijasida, bir sinf vektorlari uchun noldan kichik va boshqa sinf vektorlari uchun noldan katta qiymatlarni qabul qiluvchi funksiya topilgan. Har bir yangi obyekt uchun musbat yoki manfiy qiymat obyektini sinflardan biriga tegishligini aniqlaydi.

Sinflashtirishning eng yaxshi funksiyasi bo'lib kutilayotgan xavf-xatari minimal bo'lgan funksiya bo'lib hisoblanadi. Berilgan holatda kutilayotgan xavf-xatar kutilayotgan tasniflashning xatolik darajasini bildiradi.

Qurilgan modelning kutilgan xatolar darajasini to'g'ridan-to'g'ri baholash mumkin emas, buni empirik xavf tushunchasidan foydalanib amalga oshirish mumkin.

Biroq shuni hisobga olish kerakki, oxirgisini minimallashtirish har doim ham kutilgan xavf-xatarni minimallashtirishga olib kelavermaydi. Ushbu holat nisbatan kichik o'quv ma'lumotlari to'plami bilan ishlashda yodga olinadi.

Empirik xavf-xatar – o'quv to'plamida sinflashtirish xatolari darajasi.



20.2-rasm. Chiziqli SVM

Shunday qilib, chiziqli ajratiluvchi ma'lumotlar uchun tayanch vektorlar usuli masalasini yechish natijasida kutilayotgan xavf-xatarni yuqori bahosini

minimallashtiruvchi tasniflash funksiyasini olinadi.

Ko'rib chiqilayotgan usul bilan tasniflash masalasini yechish bilan bog'liq masalalardan biri bo'lib ikkita sinf orasidagi chiziqli chegarani har doim ham topish oson bo'lmagan holat hisoblanadi.

Bunday holatlarda variantlardan biri – o'lchamni oshirish, ya'ni ma'lumotlarni tekislikdan uch o'lchamli fazoga o'tkazish, bu yerda namunalar to'plamini ikkita sinfga ideal ajratuvchi shunday tekislikni qurish mumkin bo'ladi. Bu holatda tayanch vektorlar bo'lib ekstremal bo'lgan ikkala sinfdan olingan obyektlar hisoblanadi.

Shunday qilib, yadro operatori va qo'shimcha o'lchamlarni qo'shish yordamida gipertekislik ko'rinishidagi sinflar orasidagi tekisliklar topiladi.

SVM-modelni qurish murakkabligi shundan iboratki, fazoning o'lchami qanchalik katta bo'lsa, u bilan ishlash shunchalik qiyinlashadi. Berilgan katta hajmli ma'lumotlar bilan ishlash variantlaridan biri - bu eng muhim tarkibiy qismlarni aniqlash uchun ma'lumotlar o'lchamini kamaytirishning ba'zi usullaridan foydalanish, so'ngra tayanch vektorlar usulidan foydalanish.

Boshqa usullar singari SVM usul ham berilgan usulni tanlashda hisobga olinadigan o'zining kuchli va salbiy tomonlari mavjud.

Usulning kamchiligi shundan iboratki, tasniflash uchun barcha namunalar to'plamidan emas, balki ularning chegaralarda joylashuvchi kichik qismidan foydalanadi.

Usulning afzalligi shundan iboratki, tayanch vektorlar usuli bilan tavsiflash uchun aksariyat boshqa usullardan farqli o'laroq kichik hajmli ma'lumotlar to'plami etarlidir. Testli to'plamda qurilgan modelni to'g'ri ishlashida berilgan usulni haqiqiy ma'lumotlarga qo'llash batamom mumkin.

Tayanch vektorlar usuli quyidagilarga imkon beradi:

kutilayotgan xavf-xatarni minimal yuqori bahosi bilan tasniflash funksiyasini olish (tasniflash xatolar darajasi);

oddiylikni samaradorlik bilan birlashtirgan holda chiziqsiz bo'linuvchi ma'lumotlar bilan ishlash uchun chiziqli tasniflagichdan foydalanish.

“Yaqin qo'shnilar” usuli

“Yaqin qo'shnilar” usuli (“nearest neighbour”) yangi elementlar bilan solishtirish uchun xotirada ma'lumotlarni saqlashga asoslanuvchi ish usullari sinfga tegishli bo'ladi. Yangi yozuvni paydo bo'lishida bashoratlash uchun

ushbu yozuv va ma'lumotlarni o'xshash to'plamlari orasida og'ishlar topiladi, shuningdek eng o'xshashi aniqlanadi (yoki yaqin qo'shni).

Masalan, bankning yangi mijozini ko'rib chiqishda yangi atributlar berilgan bankning barcha mavjud mijozlari bilan solishtiriladi (kirim, yosh, va boshqalar). Bankning potensial mijozining "yaqin qo'shnilar" to'plami kirim, yosh, va boshqalarning yaqin qiymatlari asosida tanlanadi.

Bunday yondashuvda "k-yaqin qo'shnilar" ("k- nearest neighbour") terminidan foydalaniladi. Termin shuni bildiradiki, "yaqin qo'shnilar" to'plami sifatida ularni ko'rib chiqish uchun k "yuqori"(yaqinlashuvchi) qo'shnilar tanlanadi. Barcha ma'lumotlarni saqlash har doim ham qulay bo'lmasligi sababidan, ba'zan faqat "tipik" holatlar to'plami saqlanadi. Bunday holda foydalaniluvchi usul "CaseBased Reasoning" (CBR) deb nomlanadi, o'xshash holatlar asosida mulohaza, pretsedentlar tomonidan mulohaza yuritiladi.

Pretsedent –bu berilgan holatda oldindan qabul qilinuvchi amallarni to'liq birlashtirilish holatini tavsiflash.

Pretsedentlarga asoslangan yondashuvni shartli ravishda quyidagi bosqichlarga bo'lish mumkin:

- qo'yilgan masala to'g'risidagi to'liq ma'lumotlar to'plami;
- o'xshash holatlarni aniqlash uchun bazada saqlanuvchi pretsedentlar detallari bilan ushbu ma'lumotlarni solishtirish;
- pretsedentlar bazasidan joriy muammoga eng yaqin pretsedent tanlovi;
- joriy muammoda tanlangan yechimning adaptatsiyasi, agar bu zaruriyat bo'lsa;
- har bir qaytadan olingan yechimni tekshirish;
- pretsedentlar bazasiga yangi pretsedent to'g'risida batafsil ma'lumot kiritish.

Shunday qilib, pretsedentlarga asoslangan xulosa o'zi bilan pretsedentlar bazasida saqlanuvchi o'xshashliklarni qidiruv natijalari bo'yicha berilgan holatga nisbatan ma'lumotlarni tahlil qilishning shunday usulini tasvirlaydi.

Berilgan usul o'z ma'nosida "o'qituvchisiz o'qitish" kategoriyasiga tegishli bo'ladi, ya'ni "o'z-o'zini o'qitish" texnologiyasi, pretsedentlar har bir bazasini ish xususiyatlari vaqt o'tishi davomida va misollarni to'plash bilan yaxshilanadi. "Yaqin qo'shnilar" usulining afzalligi. Olingan natijalarni foydalanish qulayligi.

Aniq bir holat uchun yechim noyob emas, boshqa holatlar uchun undan foydalanish mumkin.

Qidirish maqsadi kafolatlangan to'g'ri yechim emas, balki eng maqbul yechimdir.

"Yaqin qo'shni" usulining kamchiliklari.

Berilgan usul qandaydir avvalgi tajribalarni umumlashtiruvchi model yoki qoidalarni yaratmaydi, – yechimni tanlashda ular barcha massivlarda yo'l qo'yiluvchi tarixiy ma'lumotlarga asoslanadi, shuning uchun ham javoblarni nimaga asoslanganligini aytish mumkin emas.

"Yaqinlik" (metrikalar)o'lchamlarini tanlashda qiyinchiliklar mavjud. Ushbu o'lchamdan asosan qoniqarli tasniflash va bashoratlash erishish uchun xotirada saqlanish kerak bo'lgan yozuvlar to'plami hajmiga bog'liq. Shuningdek tanlangan metrikaning tasniflash natijalarini yuqori bog'lanishi mavjud.

Usuldan foydalangan paytda ushbu hisoblash murakkabligi oqibatini tanib olishda o'qitiluvchi tanlanmani to'liq saralash zaruriyati yuzaga keladi.

Berilgan usulning tipik masalalari – bu sinflar va o'zgaruvchilar miqdori bo'yicha kichik hajmdagi masalalar.

Tasniflash va regressiya (bashoratlash) masalalari uchun k-yaqin qo'shnilar usulini ishlash tamoilini batafsil ko'rib chiqamiz.

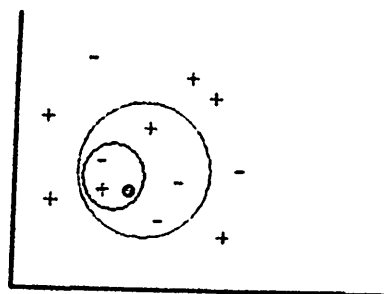
Yangi obyektlarni tasniflash masalalarini yechish.

Ushbu masala 20.3-rasmda sxematik tasvirlangan. Misollar (ma'lum nusxalar) mos sinfga ("+" yoki "-") tegishliligi aniqlovchi "+" yoki "-" belgi bilan belgilangan, tasniflash talab qilinayotgan yangi obyekt esa qizil yumaloq bilan belgilangan. Yangi obyektlar so'rov nuqtalari deb ham ataladi.

Maqsad yaqin qo'shnilarini maxsus tanlangan sonlardan foydalangan holda so'rov nuqtalarini javoblarini baholashdan (tasniflashdan) iborat. Boshqacha aytganda, quyidagi so'rovni qaysi sinfga kiritish kerakligini bilish kerak: "+" belgi sifatida yoki "-" belgi sifatida.

Boshlanishiga bir yaqin qo'shnidan foydalangan holda k-yaqin qo'shnilar usulining ishi natijasini ko'rib chiqamiz. Bu holatda so'rov nuqtalarining javobi ham qo'shuv, ham qo'shuv belgisiga ega yaqin qo'shni nuqtasi sifatida sinflashtiriladi.

Endi foydalaniluvchi yaqin qo'shnilar sonini ikkigacha oshiramiz. Bu safar k-yaqin qo'shnilar usuli so'rov nuqtalari javobini tasniflash mumkin bo'lmaydi, chunki ikkinchi yaqin nuqta ayiruv belgisiga ega va ikkala belgi teng qiymatli (ya'ni ovozlarni bir xil miqdori bilan g'alaba qozonish).



20.3-rasm. k parametrning turlicha qiymatlarida to'plam obyektlarini tasniflash

Yaqin qo'shnilar sonini 5 gacha oshirgan holda so'rov nuqtalarini atrofini olamiz (kul rang aylana). Shuningdek soha "+" belgisi bilan 2 ta nuqtadan va "-" belgisi bilan 3 ta nuqtadan iborat, algoritmi k -yaqin qo'shnilar algoritmi so'rov nuqtalari javobiga "-" belgisini tayinlaydi.

Bashoratlash masalasini yechish.

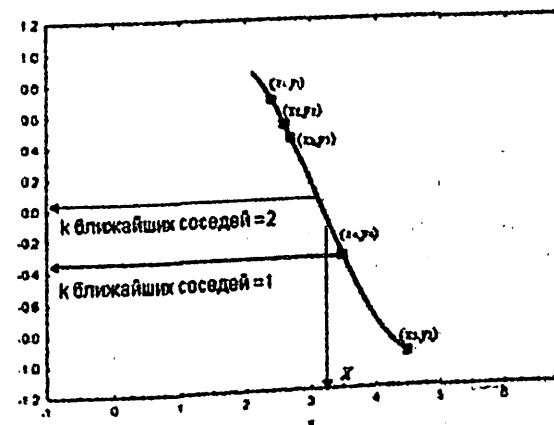
Ishning davomida regressiya masalasini yechish uchun k -yaqin qo'shnilar usulini ishlash tamoilini ko'rib chiqamiz. Regression masalalar ma'lumotlar to'plamini bog'liq o'zgaruvchilari qiymatlari bo'yicha bashorat qilish bilan bog'liq.

20.4-rasmida ko'rsatilgan grafikni ko'rib chiqamiz. Unda tasvirlangan nuqtalar to'plami (yashil to'g'ri to'rtburchak) x erkin o'zgaruvchi va y bog'liq o'zgaruvchi (egri qizil rang) orasidagi bog'lanish bo'yicha olingan. Yashil obyektlar to'plami berilgan (ya'ni misollar to'plami); biz k -yaqin qo'shnilar usulidan X so'rov nuqtalarini chiqishini bashorat qilish uchun berilgan misollar to'plami bo'yicha (yashil to'g'ri to'rtburchaklar) foydalanamiz.

Avvalo misol sifatida k -yaqin qo'shnilar usulini bir yaqin qo'shnidan ($k=1$) foydalangan holda ko'rib chiqamiz. Misollar to'plamini olamiz (yashil to'g'ri to'rtburchaklar) va so'rov nuqtasiga yaqin ularning sonidan belgilaymiz. Buning uchun yaqin misol holati - $(x_4; y_4)$ nuqtasidir. x_4 chiqish (ya'ni $y=4$), shunday qilib X (ya'ni Y) chiqishni bashorati natijasi sifatida qabul qilinadi. Binobarin, bir yaqin qo'shni uchun quyidagicha yozishimiz mumkin: Y chiqish teng y_4 ($Y=y_4$).

Keyin k ikkiga teng holatni, ya'ni ikki yaqin qo'shnilar uchun ko'rib chiqamiz. Bu holatda X nuqtalarga ikki yaqin qo'shnilar ajratiladi. Grafikda bu nuqtalar muvofiq holda y_3 va y_4 dir. Ularni chiqishlarini o'rtachasini

hisoblagan holda Y uchun ushbu ko'rinishdagi yechimni yozib olamiz $Y=(y_3 + y_4)/2$.



20.4-rasm k parametrning turlicha qiymatlarida bashoratlash masalasini yechish

Bashoratlash masalasini yechish yaqin qo'shnilarni ixtiyoriy sonlaridan foydalanishga yuqorida bayon qilingan amallarni o'tkazish yo'li bilan quyidagicha tarzda amalga oshiriladi, shunday qilib, X so'rov nuqtalarining Y chiqishi so'rov nuqtalarini k -yaqin qo'shnilar chiqishlari qiymatlarini o'rtacha arifmetiklari sifatida hisoblanadi.

Ma'lumotlar to'plamining erkin va bog'liq o'zgaruvchilari ham uzluksiz, ham kategorial bo'lishi mumkin. Uzluksiz bog'liq o'zgaruvchilar uchun masala bashoratlash masalasi sifatida, diskret o'zgaruvchilar uchun - tasniflash masalasi sifatida ko'rib chiqiladi.

Bashoratlash masalasida avvaldan aytish k -yaqin qo'shnilarni chiqishlarini o'rtachalashtirish bilan olinadi, tasniflash masalasining yechimi esa "ovozlarni ko'pligi bo'yicha" tamoiliga asoslangan.

k -yaqin qo'shnilar usulidan foydalanishda muhim nuqta bo'lib bashoratlash yoki tasniflash modelini sifatini aniqlashdan iborat.

Agar k parametrni o'ta kichik qiymati tanlansa, bashorat qiymatlarini ko'p tarqalish ehtimolligi yuzaga keladi. Agar tanlangan qiymat o'ta katta bo'lsa, bu modelni kuchli qo'zg'alishligiga (смещенности) olib kelishi mumkin. Shunday qilib, k parametrning optimal qiymati tanlangan bo'lishi zarur. U holda bu qiymat noto'g'ri tasniflash ehtimolligini keltirish uchun shunchalik

katta bo'lishi kerak, va bir paytni o'zida, so'rov nuqtasiga yetarlicha yaqin joylashgan k qo'shnilar uchun yetarlicha kichik bo'lishi kerak.

Bu yerda k modelni diapazoni (tarqalishi) va uning qo'zg'alishligi (смещенности) orasida kompromiss topilishi zarur bo'lgan silliq parametr sifatida ko'rib chiqiladi.

Kross-tekshiruv usuli bilan k parametrni baholash

k parametrni baholash variantlaridan biri - kross-tekshiruvni o'tkazish.

Usulning asosiy g'oyasi - ma'lumotlar tanlanmasini v-"burmalar"ga bo'lish. V-"burmalar" - bu tasodifiy tarzda belgilangan ajratilgan qism tanlanma.

Belgilangan k ning qiymati v-segmentda bashoratni olish uchun k-yaqin qo'shnilar modelini quradi (qolgan segmentlar bunda misol sifatida foydalaniladi) va tasniflash xatosini baholaydi. Regressiya masalalari uchun ko'pincha kvadratlar yig'indisi xatolarni baholash vazifasini bajaradi, tasniflash masalalari uchun esa aniqlikni ko'rib chiqish qulayroq. (to'g'ri tasniflangan kuzatishlarni foizi).

Keyin jarayon v tanlovning barcha mumkin bo'lgan variantlari uchun ketma-ket takrorlanadi. v-"burmalar" (sikllar) tugagandan so'ng hisoblangan xatolar modelni turg'unligi o'lchami sifatida foydalaniladi (ya'ni so'rov nuqtalarida bashoratning sifati o'lchami).

Yuqorida tavsiflangan amallar turlicha k uchun takrorlanadi va eng kichik xatoga mos (yoki eng katta tasniflash aniqligi) optimal kross-tekshirish usuli ma'nosida qabul qilinadi.

k parametrning qiymatlarini tanlashning boshqa varianti-mustaqil ravishda uning qiymatini beradi. Biroq ushbu usuldan agar parametrning mumkin bo'lgan nisbatan asoslangan taxminlariga, masalan, shunga o'xshash ma'lumotlar to'plamining avvalgi tadqiqotlari ega bo'lsa foydalaniladi.

k-yaqin qo'shnilar usuli turlicha masalalarda yetarlicha yaxshi natijalarni ko'rsatmoqda.

Bayes tasnifi

Alternativ nomlar: bayes modellashtirish, bayes statistikasi, bayes to'rlari usuli. Dastlab ekspert tizimlarida ekspert bilimlarni shakllantirish uchun bayes tasnifi foydalanilgan bo'lsa, hozir bayes tasnifi Data Mining usullaridan biri sifatida ham qo'llanilmoqda. Sodda deb nomlanuvchi tasniflash yoki sodda-

bayes yondashuvi (naive-bayes approach) bayes to'ridan foydalanuvchi usulning eng oddiy varianti bo'lib hisoblanadi. Bu yondashuvda tasniflash masalasi yechiladi, ishning natijasi bo'lib "shaffof" deb nomlanuvchi modellar hisoblanadi.

Tasniflash "Sodda" deb ataladi, chunki u belgilaming o'zaro mustaqilligi to'g'risidagi taxminlarga asoslangan.

Sodda tasniflashning xossalari:

Barcha o'zgaruvchilardan foydalanish va ular orasidagi bog'lanishlarni aniqlash.

O'zgaruvchilar to'g'risida ikkita taxminga ega bo'linadi:

- barcha o'zgaruvchilar bir xilda muhim;

- barcha o'zgaruvchilar statistik jihatdan mustaqil, ya'ni bir o'zgaruvchining qiymati boshqasining qiymati to'g'risida hech narsa gapirmaydi.

Tasniflashning ko'pgina boshqa usullari obyektning u yoki boshqa sinfga tegishli bo'lish ehtimolini tasniflashdan oldin bir xil deb taxmin qiladi, biroq bu har doim ham to'g'ri bo'lmaydi.

Data Mining usuli sifatidagi bayes to'rlarining quyidagicha afzalliklari belgilangan:

- modelda barcha o'zgaruvchilar orasidagi bog'lanishlar aniqlanadi, bu qiymatlari noma'lum bo'lgan ayrim o'zgaruvchilar holatini oson qayta ishlash imkoniyatini beradi;

- bayes to'rlari yetarlicha oddiy talqin etiladi va "nima, agar" senariysi bo'yicha tahlilni bashoratga oid modellashtirish bosqichida oson o'tkazishga imkon beradi;

- bayes usuli ma'lumotlardan kelib chiqadigan qonuniyatlarni va aniq shaklda olingan ekspert bilimlarni tabiiy ravishda birlashtirishga imkon beradi;

- bayes to'rlaridan foydalanish qayta tayyorlash (over fitting) muammolaridan, ya'ni modelni haddan ziyod murakkablashuvdan qochishga imkon beradi, bu ko'plab usullarning zaif tomoni bo'lib hisoblanadi (masalan, daraxt yechimlari va neyron to'rlari).

21. BASHORATLASH VA VAQT QATORLARI

Vaqt qatori- sonli ketma- ketlik bo'lib, uning elementlari vaqtga bog'liq jarayonning qiymatlaridan iborat. Ular vaqtning ketma- ket, ko'p hollarda teng oraliqdan iborat qiymatlarida o'lchangan. Odatda, vaqt qatorni tashkil etuvchi sonlar vaqt qatorning elementlari, vaqtning nomeriga mos holda belgilanadi. (misol uchun, x_1, x_2, x_3, \dots va hokozolar). Shunday qilib, vaqt qatorlarda elementlarning berilish tartibi juda muhim. Iqtisodiyotda vaqt qatorlarga misol tariqasida qimmatli qog'ozlarning har kungi bahosi, valyuta kursi, haftalik va oylik savdo miqdorlari, yillik ishlab chiqarish hajmi va boshqalarni ko'rish mumkin.

Doimo ta'sir etuvchi faktorlarni silliq o'zgarishini ifodalovchi komponenta trend deb ataladi. Huquqshunoslikda bunday faktorlarga misol qilib quyidagilarni olish mumkin:

- jinoyat ko'rsatkichlarining aholi farovon turmush tarziga qarab o'zgarishi;

- texnologik va iqtisodiy rivojlanish;

- malakali huquqshunoslarga bo'lgan talabning o'sishi.

Eng oddiy -chiziqli model $tr_i = b_0 + b_1 t$

polinomial: $tr_i = b_0 + b_1 t + b_2 t^2 + \dots + b_n t^n$ odatda polinom darajasi $n \leq 5$ bo'ladi.

ko'rsatkichli: $tr_i = \exp(b_0 + b_1 t)$. Bu model ko'p hollarda doimo o'sish xususiyatiga ega bo'lgan masalalarda ishlatiladi.

logistik: $tr_i = a / (1 + be^{-ct})$

Gompers: $\log(tr_i) = a + br^t$, $0 < r < 1$

Tajribalar shuni ko'rsatadiki, odatda huquqiy jarayonlarni o'zgarimas komponenta bilan ifodalash qulaydir.

T to'plamda aniqlangan, t ning funksiyasi bo'lib, har bir $t_0 \in T$ da $X(t_0)$ tasodifiy miqdor bo'lsa, $X(t)$ tasodifiy jarayon deb ataladi. Tasodifiy jarayonlar o'zlarining chekli o'lchovli taqsimotlari bilan to'laligicha aniqlanadi. Tasodifiy jarayonlarning sonli xarakteristikalari xuddi tasodifiy miqdorlarniki kabi aniqlanadi.

Tasodifiy jarayonlarning sonli xarakteristikalari:

Matematik kutilma:

$$m(t) = MX(t)$$

$X(t)$ tasodifiy jarayonining kovariatsiya funksiyasi:

$$B(s, t) \text{ cov}(X(t), X(s)) = M[(X(t) - m(t))(X(s) - m(s))].$$

Dispersiya: $DX(t) = \text{cov}(X(t), X(t))$.

Standart chetlanish: $\sigma(t) = \sqrt{\text{cov}(X(t), X(t))}$.

Korrelyasion funksiya $\text{cor}(X(t), X(s)) = \text{cov}(X(t), X(s)) / \sigma(t)\sigma(s)$.

22. BASHORATLASH MASALASINI YECHISH

22.1. Vaqt qatorining additiv modelini qurish

Additiv modelning umumiy ko'rinishi quyidagicha:

$$Y = T + S + E$$

Ushbu model vatga bog'liq qatorning har bir pog'onasi trendli(T), mavsumiy(S) hamda tasodifiy (E) komponentalarning yig'indisi ko'rinishida ifodalanishini nazarda tutadi.

22.1-misol. Tashkilotning 12 oylik daromadi to'g'risidagi inflatsiyani hisobga olgandagi ma'lumotlar asosida uning keyingi ikkita kvartali uchun daromadni bashorat qilish maqsadida additiv trend modelini quring. Model aniqligini tavsiflang va xulosalar chiqaring.

Yechish.

Vaqt qatori additiv modelining komponentalarini o'rganib chiqamiz.

1-qadam. Qatorning boshlang'ich pog'onalarini siljuvchi o'rta usuli yordamida tekislaymiz:

1.1. Siljuvchi o'rtalarni topamiz (3-jadval asosida). Shu yo'l bilan olingan qiymatlarda mavsumiy komponentalar qatnashmaydi.

1.2. Ushbu qiymatlarni vaqt momentlariga moslashtiramiz, buning uchun ikkita ketma-ket siljuvchi o'rtalardan o'rta qiymatlarni- markazlashgan siljuvchi o'rtalarni topamiz (4-jadval asosida)

t	y _t	Siljuvchi o'rta	Markazlashgan siljuvchi o'rta	Mavsumiy komponentaning bahosi
1	160	-	-	-
2	130	153.5	-	-
3	159	152.5	153	6
4	165	155.25	153.88	11.13
5	156	154.75	155	1
6	141	156.5	155.63	-14.63
7	157	156.75	156.63	0.38
8	172	157.75	157.25	14.75
9	157	159.25	158.5	-1.5

10	145	160.5	159.88	-14.88
11	163	-	-	-
12	177	-	-	-

2-qadam. Mavsumiy komponentaning baholashlarini rasmiy qator pog'onalari hamda siljuvchi o'rtalarning ayirmasi ko'rinishida topamiz (5-jadval). Ushbu baholashlardan S mavsumiy komponentaning qiymatlarini hisoblashda foydalanamiz. Buning uchun S_i mavsumiy komponentaning har bir kvartal uchun (barcha yillar bo'yicha) o'rtacha baholashlarni topamiz. Mavsumiy komponenta modellarida mavsumiy ta'sirlar o'zaro yutilib ketadi deb faraz qilinadi. Additiv modelda mavsumiy komponenta qiymatlarining barcha kvartallar bo'yicha yig'indisi nolga teng bo'ladi.

Ko'rsatkichlar	1	2	3	4
1	-	-	6	11.13
2	1	-14.63	0.38	14.75
3	-1.5	-14.88	-	-
Jami	-0.5	-29.5	6.38	25.88
Mavsumiy komponentaning o'rta bahosi	-0.25	-14.75	3.19	12.94
To'g'irlangan mavsumiy komponenta, S _i	-0.53	-15.03	2.91	12.66

Mazkur model uchun quyidagiga ega bo'lamiz: $-0.25 - 14.75 + 3.188 + 12.938 = 1.125$.

To'g'irlovchi koeffitsient: $k = 1.125/4 = 0.281$.

S_i mavsumiy komponentaning to'g'irlangan qiymatlarini hisoblaymiz va olingan ma'lumotlarni jadvalga kiritamiz.

3-qadam. Mavsumiy komponentaning ta'sirini bartaraf etamiz, buning uchun uning qiymatini boshlang'ich vaqt qatorning har bir pog'onasidan ayiramiz. $T + E = Y - S$ (4-jadval) kattaliklarni olamiz. Ushbu qiymatlar har bir vaqt momenti uchun hisoblanib, ularda tendensiya hamda tasodifiy komponenta qatnashadi xolos.

Tenglama parametrlarini kichik kvadratlar usuli yordamida topamiz.

KKU tenglamalar sistemasi:

$$\begin{cases} a_0 n + a_1 \sum t = \sum y \\ a_0 \sum t + a_1 \sum t^2 = \sum yt \end{cases}$$

Bizning ma'lumotlarimiz uchun tenglamalar sistemasi quyidagi ko'rinishga ega bo'ladi:

$$\begin{cases} 12a_0 + 78a_1 = 1882 \\ 78a_0 + 650a_1 = 12351.75 \end{cases}$$

Birinchi tenglamadan a_0 ni ifodalab olib, uni ikkinchi tenglamaga qo'yamiz.

Natijada $a_0 = 0,83$, $a_1 = 151,44$ ga ega bo'lamiz.

O'rta qiymatlar $y = 1882/12 = 156,83$.

t	y	t ²	y ²	t·y	y(t)	(y - y _{tpm}) ²	(y - y(t)) ²
1	160.53	1	25770.28	160.53	152.27	13.67	68.31
2	145.03	4	21034.06	290.06	153.1	139.29	65.05
3	156.09	9	24365.26	468.28	153.93	0.55	4.7
4	152.34	16	23208.62	609.38	154.76	20.16	5.83
5	156.53	25	24502.03	782.66	155.59	0.0912	0.89
6	156.03	36	24345.75	936.19	156.42	0.64	0.15
7	154.09	49	23744.88	1078.66	157.25	7.51	9.95
8	159.34	64	25390.43	1274.75	158.08	6.3	1.6
9	157.53	81	24816.09	1417.78	158.91	0.49	1.9
10	160.03	100	25610	1600.31	159.74	10.23	0.0849
11	160.09	121	25630.01	1761.03	160.57	10.63	0.23
12	164.34	144	27008.87	1972.13	161.4	56.41	8.66
13	1882	650	295426.29	12351.75	1882	265.96	167.35

4-qadam. Berilgan modelning T komponentasini aniqlaymiz. Buning uchun, chiziqli trend yordamida ($T + E$) qatorni analitik jihatdan tekislaymiz. Analitik tekislash natijalari quyidagicha: $T = 151,36 + 0,83t$. Ushbu tenglamaga $t = 1, \dots, 12$, ning qiymatlarini qo'yib, har bir vaqt momenti uchun T pog'onalarni topamiz (5-jadval).

t	y _t	S _t	y _t - S _t	T	T + S _t	E = y _t - (T + S _t)	E ²
1	160	-0.53	160.53	152.27	151.73	8.27	68.31
2	130	-15.03	145.03	153.1	138.07	-8.07	65.05
3	159	2.91	156.09	153.93	156.83	2.17	4.7
4	165	12.66	152.34	154.76	167.41	-2.41	5.83
5	156	-0.53	156.53	155.59	155.06	0.94	0.89
6	141	-15.03	156.03	156.42	141.39	-0.39	0.15
7	157	2.91	154.09	157.25	160.15	-3.15	9.95
8	172	12.66	159.34	158.08	170.74	1.26	1.6
9	157	-0.53	157.53	158.91	158.38	-1.38	1.9
10	145	-15.03	160.03	159.74	144.71	0.29	0.0849
11	163	2.91	160.09	160.57	163.48	-0.48	0.23
12	177	12.66	164.34	161.4	174.06	2.94	8.66

5-qadam. Additiv model asosida olingan qator pog'onalarining qiymatlarini topamiz. Buning uchun pog'onalariga mos kvartallarning T mavsumiy komponentalari qiymatlarini qo'shamiz (6-jadval). Qurilgan modelning sifatini baholash uchun olingan absolyut xatoliklar kvadrlarining yig'indisini tabdiq etamiz.

$$R^2 = 1 - \frac{\sum E^2}{\sum (y_t - y_{tpm})^2}$$

O'rta qiymat

$$y = \frac{1882}{12} = 156,83$$

ga teng.

t	y	(y _t - y _{tpm}) ²
1	160	10.03
2	130	720.03

3	159	4.69
4	165	66.69
5	156	0.69
6	141	250.69
7	157	0.0277
8	172	230.03
9	157	0.0277
10	145	140.03
11	163	38.03
12	177	406.69
13	1882	650

$$R^2 = \frac{1 - 167,347}{1867,666} = 0,91.$$

Demak, additiv model vaqt qatori pog'onalarining 91% umumiy varitsiyasini tushuntirib beradi deb aytish mumkin.

$$F = \frac{R^2}{(1 - R^2)(n - m - 1)} = \frac{0,91^2}{(1 - 0,91^2)(12 - 1 - 1)} = 101,6.$$

Bu erda m – trend tenglamasidagi faktorlar soni ($m=1$).

$$F_{kp} = 4,96.$$

$F > F_{kp}$ bo'lganligi bois, tenglama statistik jihatdan ahamiyatga ega bo'ladi.

6-qadam. Additiv model bo'yicha bashoratlash. Additiv modelda vaqt qatori F_t pog'onasining bashorat qiymati trend va mavsumiy komponentalarning yig'indisiga teng. Trend komponentasini aniqlash uchun trend tenglamasidan foydalanamiz:

$$T = 151,436 + 0,83t.$$

Natijada quyidagiga ega bo'lamiz

$$T_{13} = 151,436 + 0,83 \cdot 13 = 162,231.$$

Mavsumiy komponentaning berilgan davrdagi qiymati:

$$S_1 = -0,31$$

ga teng.

Shunday qilib,

$$F_{13} = T_{13} + S_1 = 162,231 - 0,531 = 161,7$$

$$T_{14} = 151,436 + 0,83 \cdot 14 = 163,061.$$

Mavsumiy komponentaning mos davr uchun qiymati quyidagiga teng:

$$S_2 = -15,031.$$

Natijada,

$$F_{14} = T_{14} + S_2 = 163,061 - 15,031 = 148,03$$

$$T_{15} = 151,436 + 0,83 \cdot 15 = 163,892.$$

Mavsumiy komponentaning mos davr uchun qiymati quyidagiga teng:

$$S_3 = 2,906.$$

Shunday qilib,

$$F_{15} = T_{15} + S_3 = 163,892 + 2,906 = 166,798.$$

$$T_{16} = 151,436 + 0,83 \cdot 16 = 164,722$$

Mavsumiy komponentaning mos davr uchun qiymati quyidagiga teng:

$$S_4 = 12,656.$$

Va nihoyat,

$$F_{16} = T_{16} + S_4 = 164,722 + 12,656 = 177,379.$$

22.2. Multiplikativ modelni qurish

Multiplikativ modelning umumiy ko'rinishi quyidagicha:

$$Y = T \times S \times E$$

Ushbu model vatga bog'liq qatorning har bir pog'onasi trendli(T), mavsumiy(S) hamda tasodifiy (E) komponentalarning ko'paytmasi ko'rinishida ifodalanishini nazarda tutadi.

22.2-misol. Tashkilotning 12 oylik daromadi to'g'risidagi inflatsiyani hisobga olgandagi ma'lumotlar asosida uning keyingi ikkita kvartali uchun daromadni bashorat qilish maqsadida multiplikativ trend hamda mavsumlik modelini quring. Model aniqligini tavsiflang va xulosalar chiqaring.

Yechish.

Vaqt qatori multiplikativ modelining komponentalarini o'rganib chiqamiz.

1-qadam. Qatorning boshlang'ich pog'onalarini siljувchi o'rta usuli yordamida tekislaymiz:

1.1. Siljувchi o'rtalarni topamiz(3-jadval asosida). Shu yo'l bilan olingan qiymatlarda mavsumiy komponentalar qatnashmaydi.

1.2. Ushbu qiymatlarni vaqt momentlariga moslashtiramiz, buning uchun ikkita ketma-ket siljuvchi o'rtalardan o'rta qiymatlarni- markazlashgan siljuvchi o'rtalarni topamiz (4-jadval asosida)

2-qadam. Mavsumiy komponentaning baholashlarini rasmiy qator pog'onalari hamda siljuvchi o'rtalarning nisbati ko'rinishida topamiz (5-jadval). Ushbu baholashlardan S mavsumiy komponentaning qiymatlarini hisoblashda foydalanamiz. Buning uchun S_t mavsumiy komponentaning har bir kvartal uchun (barcha yillar bo'yicha) o'rtacha baholashlarni topamiz. Mavsumiy komponenta modellarida mavsumiy ta'sirlar o'zaro yutilib ketadi deb faraz qilinadi.

t	y_t	Siljuvchi o'rta	Markazlashgan siljuvchi o'rta	Mavsumiy komponentaning bahosi
1	375	-	-	-
2	371	657.5	-	-
3	869	653	655.25	1.33
4	1015	678	665.5	1.53
5	357	708.75	693.38	0.51
6	471	710	709.38	0.66
7	992	718.25	714.13	1.39
8	1020	689.25	703.75	1.45
9	390	689.25	689.25	0.57
10	355	660.5	674.88	0.53
11	992	678.25	669.38	1.48
12	905	703	690.63	1.31
13	461	685	694	0.66
14	454	690.5	687.75	0.66
15	920	-	-	-
16	927	-	-	-

Multiplikativ modelda barcha kvartallar bo'yicha mavsumiy komponenta qiymatlarining yig'indisi sikldagi davrlar soniga teng bo'lishi kerak. Bizning misolimizda bittas siklning davrlar soni 4 ga teng.

Ko'rsatkichlar	1	2	3	4
1	-	-	1.33	1.53
2	0.51	0.66	1.39	1.45
3	0.57	0.53	1.48	1.31
4	0.66	0.66	-	-
Jami	1.74	1.85	4.2	4.28
Mavsumiy komponentaning o'rtacha bahosi	0.58	0.62	1.4	1.43
To'g'irlangan mavsumiy komponenta, S_t	0.58	0.61	1.39	1.42

Mazkur modelga nisbatan quyidagiga ega bo'lamiz: $0.582 + 0.617 + 1.399 + 1.428 = 4.026$.

To'g'irlanuvchi koeffitsient: $k = 4/4.026 = 0.994$.

S_t mavsumiy komponentaning to'g'irlangan qiymatlarini hisoblab chiqib, olingan ma'lumotlarni jadvalga kiritamiz.

3-qadam. Boshlang'ich qatorning har bir pog'onasini mavsumiy komponentaning mos qiymatlariga bo'lamiz. Natijada faqatgina tendensiya hamda tasodifiy komponentani o'z ichiga olgan $T \times E = \frac{Y}{S}$ kattaliklarni olamiz (4-jadval).

Tenglama parametrlarini kichik kvadratlar usuli yordamida topamiz.

Kichik kvadratlar usuli tenglamalar sistemasi

$$\begin{cases} a_0 n + a_1 \sum t = \sum y_t \\ a_0 \sum t + a_1 \sum t^2 = \sum yt \end{cases}$$

Bizning ma'lumotlarimiz uchun tenglamalar sistemasi quyidagi ko'rinishga ega bo'ladi:

$$\begin{cases} 16a_0 + 136a_1 = 10872.41 \\ 136a_0 + 1496a_1 = 93531.1 \end{cases}$$

Birinchi tenglamadan a_0 ni ifodalab olib, uni ikkinchi tenglamaga qo'yamiz.

Natijada $a_0 = 3.28$, $a_1 = 651.63$ ga ega bo'lamiz.

O'rta qiymatlar $y = 10872.41/16 = 679.53$

4-qadam. Berilgan modelning T komponentasini aniqlaymiz. Buning uchun, chiziqli trend yordamida ($T + E$) qatomi analitik jihatdan tekislaymiz.

Analitiki tekislash natijalari quyidagiga teng:

$$T = 651,634 + 3,281t$$

t	y	t ²	y ²	t · y	y(t)	(y - y _{ym}) ²	(y - y(t)) ²
1	648.87	1	421026.09	648.87	654.92	940.05	36.61
2	605.46	4	366584.89	1210.93	658.2	5485.32	2780.93
3	625.12	9	390770.21	1875.35	661.48	2960.37	1322.21
4	715.21	16	511519.56	2860.82	664.76	1273.1	2544.83
5	617.72	25	381577.63	3088.6	668.04	3819.95	2532.22
6	768.66	36	590838.18	4611.96	671.32	7944.97	9474.64
7	713.6	49	509219.75	4995.17	674.6	1160.83	1520.44
8	718.73	64	516571.58	5749.83	677.88	1536.93	1668.26
9	674.82	81	455381.82	6073.38	681.17	22.14	40.28
10	579.35	100	335647.52	5793.51	684.45	10034.93	11045.26
11	713.6	121	509219.75	7849.56	687.73	1160.83	66914
12	637.7	144	406656.13	7652.35	691.01	1749.71	2842.39
13	797.67	169	636280.07	10369.73	694.29	13958.53	10687.5
14	740.92	196	548957.15	10372.83	697.57	3768.85	1878.69
15	661.8	225	437983.3	9927.05	700.85	314.08	1524.97
16	653.2	256	426667.57	10451.17	704.14	693.14	2594.6
17	10872.41	1496	7444901.2	93531.1	10872.41	56823.71	53162.96

Ushbu tenglamaga $t = 1, \dots, 16$ ning qiymatlarini qo'yib, har bir vaqt momenti uchun T pog'onalarini topamiz (5-jadval).

t	y _t	S _t	y _t - S _t	T	T + S _t	E = y _t - (T + S _t)	E ²
1	375	0.58	648.87	654.92	378.5	0.99	12.23
2	371	0.61	605.46	658.2	403.31	0.92	1044.15
3	869	1.39	625.12	661.48	919.55	0.95	2555.16

4	1015	1.42	715.21	664.76	943.41	1.08	5125.42
5	357	0.58	617.72	668.04	386.08	0.92	845.78
6	471	0.61	768.66	671.32	411.36	1.14	3557.43
7	992	1.39	713.6	674.6	937.79	1.06	2938.24
8	1020	1.42	718.73	677.88	962.03	1.06	3359.96
9	390	0.58	674.82	681.17	393.67	0.99	13.45
10	355	0.61	579.35	684.45	419.4	0.85	4147.15
11	992	1.39	713.6	687.73	956.04	1.04	1293.1
12	905	1.42	637.7	691.01	980.66	0.92	5724.7
13	461	0.58	797.67	694.29	401.25	1.15	3569.68
14	454	0.61	740.92	697.57	427.44	1.06	705.39
15	920	1.39	661.8	700.85	974.29	0.94	2946.99
16	927	1.42	653.2	704.14	999.29	0.93	5225.65

5-qadam. Qator pog'onalarini T ning qiymatlarini mavsumiy komponentaning mos qiymatlariga ko'paytirish orqali topamiz (6-jadval).

Multiplikativ modeldagi xatolik $E = Y/(T \cdot S) = 16$ formula bo'yicha hisoblanadi.

Multiplikativ hamda boshqa modellarni taqqoslash uchun absolyut xatoliklar kvadratlarining yig'indisidan foydalanish mumkin:

t	y	(y - y _{ym}) ²
1	375	92796.39
2	371	95249.39
3	869	35862.89
4	1015	112476.39
5	357	104086.89
6	471	43524.39
7	992	97578.14
8	1020	115855.14
9	390	83882.64

10	355	105381.39
11	992	97578.14
12	905	50793.89
13	461	47796.89
14	454	50906.64
15	920	57780.14
16	927	61194.39
136	10874	1252743.75

$$R^2 = 1 - \frac{43064,467}{1252743,75} = 0,97.$$

Demak, multiplikativ model vaqt qatori pog'onalarining 97% umumiy varitsiyasini tushuntirib beradi deb aytish mumkin.

$$F = \frac{R^2}{(1 - R^2)(n - m - 1)} = \frac{0,97^2}{(1 - 0,97^2)(16 - 1 - 1)} = 393,26.$$

Bu yerda m – trend tenglamasidagi faktorlar soni ($m=1$).

$$F_{kp} = 4,6.$$

$F > F_{kp}$ bo'lganligi bois, tenglama statistik jihatdan ahamiyatga ega.

6-qadam. Multiplikativ model bo'yicha bashoratlash. Multiplikativ modelda vaqt qatori F_t pog'onasining bashorat qiymati trend va mavsumiy komponentalarning yig'indisiga teng. Trend komponentasini aniqlash uchun trend tenglamasidan foydalanamiz:

$$T = 651,634 + 3,281t.$$

$$T_{17} = 651,634 + 3,281 \cdot 17 = 707,416 \text{ ga ega bo'lamiz.}$$

Mavsumiy komponentaning mos davrdagi qiymati $S_1 = 0,578$ ga teng.

$$\text{Shunday qilib, } F_{17} = T_{17} + S_1 = 707,416 + 0,578 = 707,994.$$

$$T_{18} = 651,634 + 3,281 \cdot 18 = 710,698.$$

Mavsumiy komponentaning mos davrdagi qiymati $S_2 = 0,613$ ga teng.

$$\text{Shunday qilib, } F_{18} = T_{18} + S_2 = 710,698 + 0,613 = 711,311.$$

$$T_{19} = 651,634 + 3,281 \cdot 19 = 713,979.$$

Mavsumiy komponentaning mos davrdagi qiymati: $S_3 = 1,39$ ga teng.

$$\text{Natijada, } F_{19} = T_{19} + S_3 = 713,979 + 1,39 = 715,369.$$

$$T_{20} = 651,634 + 3,281 \cdot 20 = 717,26.$$

Mavsumiy komponentaning mos davrdagi qiymati: $S_3 = 1,419$ ga teng.

$$\text{Va nihoyat, } F_{20} = T_{20} + S_4 = 717,26 + 1,419 = 718,68.$$

23. KLASSTERLASH MASALASI

Klasterli tahlil – berilganlarni to'plash, tanlov obyektlari haqidagi ma'lumotlarni saqlovchi va ularni bir jinsli guruhlariga nisbatan tartiblashni bajaruvchi ko'p o'lchamli statistik protseduradir! **Ошибка! Источник ссылки не найден.**¹

Klasterizatsiya- bu tekshirilayotgan obyektlarni "o'xshash" obyektlar klasterlari deb nomlangan guruhlariga ajratilishi. "Klaster" atamasi (Inglizcha Klaster - paket, guruh) sinf, takson, guruh atamalariga sinusim atamadir. Klasterizatsiya usuli klassifikatsiya usulidan tubdan farq qiladi. Klassifikatsiya usuli yechimi predmetlarni har biriga predmet sinflardan biriga tayinlashdir. Klasterizatsiya vazifasida obyekt oldindan aniqlanmagan sinflardan biriga tayinlash. Klasterlardagi obyektlarning klasterlarga ajratish bir vaqtning o'zida klasterlar shakllanishi bilan amalga oshiriladi! **Ошибка! Источник ссылки не найден.**¹

Klasterizatsiya, ma'lumotlarni yig'ish vazifalarini bir qator hal qilishni osonlashtiradigan o'xshash ma'lumotlarni to'plash imkonini beradi.

• Ma'lumotlarni tahlil qilish, tahlil qilishni osonlashtirish. Olingan guruhlarini mazmunli tahlil qilish qonuniyatlarni aniqlash imkonini beradi. Masalan, siz uyali aloqa tarmog'i mijozlarini aniqlashingiz mumkin, ular uchun yangi tarif rejasini taklif qilishingiz mumkin. Boshqa misollar - savdo tarmog'ining xaridor guruhlarini aniqlash, bozorni segmentlash. Klaster tarkibining tahlili turli xil guruhlarining obyektlariga tahlilning turli usullarini qo'llash imkonini beradi;

• Prognozlash. Klasterlardan biriga yangi obyektning tayinlash orqali, uning xatti-harakati klaster obyektlarining xatti-harakati bilan o'xshash bo'lgani uchun, obyekt xatti-harakatlarini oldindan belgilashingiz mumkin;

• Anomaliyalarni aniqlash. Guruhlarni to'g'ri aniqlanishi anomaliyalarni aniqlashga yordam beradi. Odatda, bu klaster, yerda juda oz narsalar tushadi.

Har bir klasterni to'g'ri aniqlanishining ahamiyatini e'tiborga olish muhimdir. Har bir guruhga klaster xossalarini aks ettiruvchi nom berilishi kerak. Buning uchun obyektlarni klasterga birlashtiradigan xususiyatlarni aniqlash kerak. Bu klaster obyektining xususiyatlarini statistik tahlil qilishni talab qiladi.

Klaster tahlil usullari

"Klaster" tushunchasini matematik nuqtai nazardan tavsiflaymiz, shu bilan birga klaster tahlil usullarini ko'rib chiqamiz. Klaster tahlil o'z ichiga 100

dan ortiq turlicha algoritmlarni oladi. Tasniflash masalalaridan farqli o'laroq klaster tahlilda ma'lumotlar to'plami to'g'risidagi aprior taxminlarni talab qilmaydi, tadqiq qilinuvchi obyektlarni berilishida cheklanishlar qo'ymaydi, ma'lumotlarni turlicha tiplari (oraliqli ma'lumotlar, chastotalar, binar ma'lumotlar) ko'rsatkichlarini tahlil qilishga imkon beradi. Barcha o'zgaruvchilar solishtiriladigan shkalalarda o'lchanishi zarur.

Klaster tahlili davriy qatorlari to'plamiga qo'llanilishi mumkin, bu yerda ayrim ko'rsatkichlarni o'xshashlik davrlari ajratilishi va o'xshash dinamikaga ega davriy qatorlar guruhlarini aniqlanishi mumkin.

№ примера	признак X	признак Y
1	27	19
2	11	46
3	25	15
4	36	27
5	35	25
6	10	48
7	11	44
8	36	24
9	26	14
10	26	14
11	9	45
12	33	23
13	27	16
14	10	47

23.1-rasm. A ma'lumotlar to'plami

Klaster tahlil masalalarini quyidagi guruhlariga birlashtirish mumkin:

1. Tipologiya yoki tasniflashni ishlab chiqish.
2. Obyektlarni guruhlashni foydali konseptual sxemalarini tadqiq qilish.
3. Ma'lumotlarni tadqiq qilish asosida gipotezalarini taqdim etish.
4. Mavjud ma'lumotlarda u yoki boshqa usullar bilan belgilangan guruhlar haqiqatdan mavjudligini aniqlash uchun gipoteza yoki tadqiqotlarni tekshirish.

Qoida tariqasida, klaster tahlilidan amaliy foydalanishda ko'rsatilgan masalalarning bir qanchasi bir vaqtni o'zida hal etiladi.

Klaster tahlil protseduralariga misol ko'rib chiqamiz.

23. KLASTERLASH MASALASI

Klasterli tahlil – berilganlarni to‘plash, tanlov obyektlari haqidagi ma’lumotlarni saqlovchi va ularni bir jinsli guruhlariga nisbatan tartiblashni bajaruvchi ko‘p o‘lchamli statistik protseduradir. **Ошибка! Источник ссылки не найден.**

Klasterizatsiya- bu tekshirilayotgan obyektlarni “o‘xshash” obyektlar klasterlari deb nomlangan guruhlariga ajratilishi. “Klaster” atamasi (Inglizcha Klaster - paket, guruh) sinf, takson, guruh atamalariga sinosim atamadir. Klasterizatsiya usuli klassifikatsiya usulidan tubdan farq qiladi. Klassifikatsiya usuli yechimi predmetlarni har biriga predmet sinflardan biriga tayinlashdir. Klasterizatsiya vazifasida obyekt oldindan aniqlanmagan sinflardan biriga tayinlash. Klasterlardagi obyektlarning klasterlarga ajratish bir vaqtning o‘zida klasterlar shakllanishi bilan amalga oshiriladi. **Ошибка! Источник ссылки не найден.**

Klasterizatsiya, ma’lumotlarni yig‘ish vazifalarini bir qator hal qilishni osonlashtiradigan o‘xshash ma’lumotlarni to‘plash imkonini beradi.

• Ma’lumotlarni tahlil qilish, tahlil qilishni osonlashtirish. Olingan guruhlarini mazmunli tahlil qilish qonuniyatlarni aniqlash imkonini beradi. Masalan, siz uyali aloqa tarmog‘i mijozlarini aniqlashingiz mumkin, ular uchun yangi tarif rejasini taklif qilishingiz mumkin. Boshqa misollar - savdo tarmog‘ining xaridor guruhlarini aniqlash, bozorni segmentlash. Klaster tarkibining tahlili turli xil guruhlarining obyektlariga tahlilning turli usullarini qo‘llash imkonini beradi;

• Prognozlash. Klasterlardan biriga yangi obyektini tayinlash orqali, uning xatti-harakati klaster obyektlarining xatti-harakati bilan o‘xshash bo‘lgani uchun, obyekt xatti-harakatlarini oldindan belgilashingiz mumkin;

• Anomaliyalarni aniqlash. Guruhlarni to‘g‘ri aniqlanishi anomaliyalarni aniqlashga yordam beradi. Odatda, bu klaster, yerda juda oz narsalar tushadi.

Har bir klasterni to‘g‘ri aniqlanishining ahamiyatini e‘tiborga olish muhimdir. Har bir guruhga klaster xossalari aks ettiruvchi nom berilishi kerak. Buning uchun obyektlarni klasterga birlashtiradigan xususiyatlarni aniqlash kerak. Bu klaster obyektining xususiyatlarini statistik tahlil qilishni talab qiladi.

Klaster tahlil usullari

“Klaster” tushunchasini matematik nuqtai nazardan tavsiflaymiz, shu bilan birga klaster tahlil usullarini ko‘rib chiqamiz. Klaster tahlil o‘z ichiga 100

dan ortiq turlicha algoritmlarni oladi. Tasniflash masalalaridan farqli o‘laroq klaster tahlilda ma’lumotlar to‘plami to‘g‘risidagi aprior taxminlarni talab qilmaydi, tadqiq qilinuvchi obyektlarni berilishida cheklanishlar qo‘ymaydi, ma’lumotlarni turlicha tiplari (oraliqli ma’lumotlar, chastotalar, binar ma’lumotlar) ko‘rsatkichlarini tahlil qilishga imkon beradi. Barcha o‘zgaruvchilar solishtiriladigan shkalalarda o‘lchanishi zarur.

Klaster tahlili davriy qatorlari to‘plamiga qo‘llanilishi mumkin, bu yerda ayrim ko‘rsatkichlarni o‘xshashlik davrlari ajratilishi va o‘xshash dinamikaga ega davriy qatorlar guruhlarini aniqlanishi mumkin.

№ примера	признак X	признак Y
1	27	19
2	11	46
3	25	15
4	36	27
5	35	25
6	10	43
7	11	44
8	36	24
9	26	14
10	26	14
11	9	45
12	33	23
13	27	16
14	10	47

23.1-rasm. A ma’lumotlar to‘plami

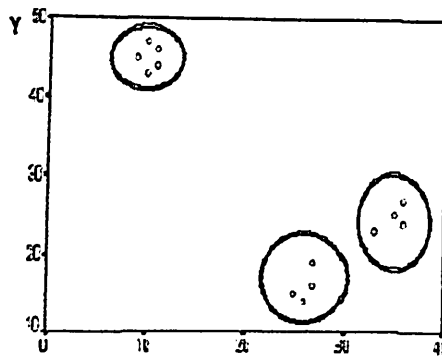
Klaster tahlil masalalarini quyidagi guruhlariga birlashtirish mumkin:

1. Tipologiya yoki tasniflashni ishlab chiqish.
2. Obyektlarni guruhlashni foydali konseptual sxemalarini tadqiq qilish.
3. Ma’lumotlarni tadqiq qilish asosida gipotezalarini taqdim etish.
4. Mavjud ma’lumotlarda u yoki boshqa usullar bilan belgilangan guruhlar haqiqatdan mavjudligini aniqlash uchun gipoteza yoki tadqiqotlarni tekshirish.

Qoida tariqasida, klaster tahlilidan amaliy foydalanishda ko‘rsatilgan masalalarning bir qanchasi bir vaqtini o‘zida hal etiladi.

Klaster tahlil protseduralariga misol ko‘rib chiqamiz.

Aytaylik, A ma'lumotlar to'plamiga egamiz, X va Y ikkita belgisi bo'yicha ega bo'luvchi 14 ta misollardan iborat. Ular bo'yicha ma'lumotlar 21-rasmda keltirilgan. Ma'lumotlar jadval shaklida informativ tabiatga ega bo'lmaydi. X va Y o'zgaruvchilarni 23.2-rasmda tasvirlangan tarqalish diagrammasi ko'rinishida taqdim etamiz.



23.2-rasm. X va Y o'zgaruvchilarni tarqalish diagrammasi

Rasmda ba'zi "o'xshash" misollar guruhini ko'ramiz. X va Y qiymatlar bo'yicha bir-biriga "o'xshash" bir guruhga (klasterga) tegishli misollar; turlicha klasterlar obyektlarini bir-biriga o'xshamagan misollari (obyektlar).

Klasterlarni o'xshashligini va farqini aniqlash uchun mezonlari – tarqalish diagrammasidagi nuqtalar orasidagi masofa. Ushbu o'xshashlikni "o'lchash" mumkin, u grafikdagi nuqtalar orasidagi masofaga teng. Yaqinlik o'lchami deb nomlanuvchi klasterlar orasidagi o'lchamni aniqlashning bir necha usullari mavjud. Eng ko'p tarqalgan usul – tekislikdagi I va j nuqtalar orasidagi Evklid masofasini hisoblash, qachonki ularning X va Y koordinatalari ma'lum bo'lgan paytda:

$$D_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}.$$

Klaster quyidagi matematik tavsiflarga ega bo'ladi: markaz, radius, o'rtacha kvadratik og'ish, klaster o'lchami.

Klaster markazi - bu o'zgaruvchilar fazosida nuqtalarni o'rtacha geometrik o'rni.

Klaster radiusi - klaster markazidan nuqtalarni maksimal masofasi. Klasterlar bir-birini qoplaydigan bo'lishi mumkin. Bunday holat klasterlarning bir-biriga ustma-ust tushishi aniqlanganda yuzaga keladi. Bu holatda obyektning ikkita klasterdan biriga bir qiymatli belgilash matematik protseduralardan

foydalanish mumkin emas.

Muhokamali obyekt – o'xshashlik o'lchami bo'yicha bir necha klasterga tegishli bo'lishi mumkin bo'lgan obyekt.

Klaster o'lchami klaster radiusi bo'yicha yoki ushbu klaster uchun obyektning o'rtacha kvadratik og'ishi bo'yicha aniqlanishi mumkin. Obyekt klasterga tegishli bo'ladi, agar obyektning klaster markazigacha bo'lgan masofa klaster radiusidan kichik bo'lsa. Agar ushbu shart ikkita yoki undan ko'p klasterlarda bajarilsa, obyekt bahsli bo'lib hisoblanadi.

Berilgan masalani bir qiymatli bo'lmashligini ekspert yoki tahlil qiluvchi tomonidan bartaraf etiladi.

Klaster tahlili ikkita taxminga asoslanadi. Birinchi taxmin –obyektning ko'rib chiqilayotgan belgilari obyektning yig'indisini klasterlarga kerakli taqsimlashini yo'lga qo'yadi. Ikkinchi taxmin – belgilarni o'lchov birliklari yoki masshtabni tanlash to'g'riligi.

Klaster tahlilda masshtabni tanlash katta ahamiyatga ega.

Misol ko'rib chiqamiz.

O'ylaymizki, A ma'lumotlar to'plamida X belgi ma'lumotlari U belgi ma'lumotlaridan ikki kattalikka katta (X o'zgaruvchi qiymatlari 100 dan 700 gacha diapazonda joylashgan, U o'zgaruvchi qiymatlari esa - 0 dan 1 gacha bo'lgan diapazonda).

U holda ularning xossalari fazosida obyektning holatini aks ettiruvchi nuqtalar orasidagi kattaliklarni hisoblashda katta qiymatlarga ega X o'zgaruvchi kichik qiymatlarga ega U o'zgaruvchi ustidan amaliy jihatdan to'liq holda ustun bo'lishini ko'rsatadi. Shunday qilib, belgilar o'lchamlari bir jinsli emasligi sababli nuqtalar orasidagi masofani to'g'ri hisoblash mumkin emas. Ushbu muammo o'zgaruvchilarni taxminiy standartlashtirish yordamida hal etiladi.

Standartlashtirish (standartization) yoki **normallashtirish** (normalization) barcha aylantirilgan o'zgaruvchilar qiymatlarini qiymatlarni yagona diapazoniga aniq belgining ma'lum xossalari aks ettiruvchi ma'lum bir kattalikka ushbu qiymatlarni munosabati orqali ifodalash yo'li bilan olib keladi.

Dastlabki ma'lumotlarni normallashtirishning turlicha usullari mavjud. Ikkita eng ko'p tarqalgan usullari quyidagicha:

- dastlabki ma'lumotlarni muvofiq o'zgaruvchilarni o'rtacha kvadratik

og'ishiga bo'lish;

- Z-hissani yoki standartlashtirilgan hissani hisoblash.

O'zgaruvchilarni standartlashtirish bilan bir qatorda ularning har bir ma'lum koeffitsientining muhimligini yoki mos o'zgaruvchining ahamiyatligini aks ettirishi mumkin bo'lgan vazni berish varianti mavjud. Vaznlar sifatida ekspertlarni, ya'ni predmet soha mutaxassislari so'rovida olingan ekspert baholar oldinga chiqadi.

Normallashtirilgan o'zgaruvchilarni muvofiq vaznlarda olingan ko'paytmasi o'zgaruvchilarning bir xil bo'lmagan vaznlarini hisobga olgan holda ko'p o'lchamli fazoda nuqtalar orasidagi masofani olishga imkon beradi.

Tajribalar davomida natijalarni solishtirish mumkin, natijalar ekspert baholarni hisobga olgan holda va ularni hisobga olmagan holda, shuningdek ular ichida eng yaxshisini tanlash bilan olinadi.

Klaster tahlil usullari

Klaster tahlil usullarini ikki guruhga bo'lish mumkin:

- ierarxik;
- ierarxik bo'lmagan.

Guruhlardan har biri yondashuvlar va algoritmlar to'plamini o'z ichiga oladi. Klaster tahlilning turlicha usullaridan foydalanilgan holda berilgan ma'lumotlar uchun turli yechimlarni olish mumkin. Bu normal hodisa bo'lib hisoblanadi.

Ierarxik va ierarxik bo'lmagan usullarni batafsil ko'rib chiqamiz.

24. KLASTER TAHLILINI TADBIQ ETISH

Klaster tahlilning ierarxik usullari.

Ierarxik klasterlashning mazmuni ketma-ket kichik klasterlarni kattalariga birlashtirish yoki kattalarini kichiklariga taqsimlashdan iborat.

Ierarxik aglomerativ usullar (Agglomerative Nesting, AGNES).

Usullarning ushbu guruhi dastlabki elementlarni ketma-ket birlashtirish va klasterlar sonini muvofiq kamaytirish bilan tavsiflanadi.

Algoritmnin boshlang'ich qismida barcha obyektlar alohida klasterlar bo'lib hisoblanadi. Birinchi qadamda eng ko'p o'xshash obyektlar klasterga birlashtiriladi. Birlashtirishning keyingi qadamlarida barcha obyektlar bitta klastermi tuzmaguncha davom ettiriladi.

Ierarxik bo'linuvchi usullar (Divisive ANALysis, DIANA).

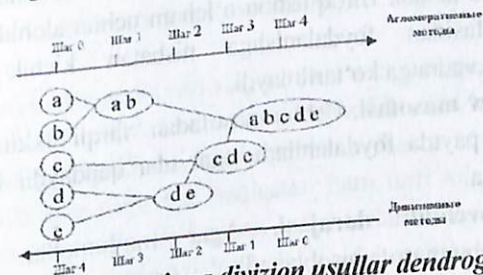
Ushbu usullar aglomerativ usullarni mantiqiy qarama-qarshiligi bo'lib hisoblanadi. Algoritmnin boshlang'ich qismida barcha obyektlar bitta klasterga tegishli bo'ladi, ushbu klaster keyingi qatorlarda kichik klasterlarga bo'linadi, natijada bo'linuvchi guruhlar ketma-ketligi shakllanadi.

Dendrogramma ko'rinishidagi usullarni yuqoridagi bayon qilingan guruhlarini ishlash tamoili 24.1-rasmda ko'rsatilgan.

Klaster tahlilning algoritmlarini dasturiy amalga oshirish yetarlicha katta hajmli masalalarni yechishga imkon beruvchi Data Miningning turlicha instrumentlarida keng ifodalangan.

Ierarxik klasterlash usullari klasterlarni qurish qoidalari bilan farq qiladi. Obyektlarni guruhlariga birlashtirishda (aglomerativ usullar) yoki guruhlariga bo'linishda (divizion usullar) obyektlarni "o'xshashliklari" to'g'risidagi masalani yechishda qoidalar sifatida foydalaniluvchi mezonlar oldinga chiqadi.

Klaster tahlilning ierarxik usullari kichik hajmli ma'lumotlar to'plamida foydalaniladi.



24.1-rasm. Aglomerativ va divizion usullar dendrogrammasi

Klasterlashning ierarxik usullarining afzalliklari bo'lib ularning ko'rinishidir.

Ierarxik algoritmlar ierarxik klaster tahlilning natijalari bo'lib hisoblangan dendogramm (grekcha dendron – “daraxt”) qurilishi bilan bog'liq.

Dendrogramma alohida nuqtalarni va klasterlarni bir-biriga yaqinligini tavsiflaydi, klasterlarni birlashtirish (ajratish) ketma-ketligini grafik ko'rinishini tasvirlaydi.

Dendrogramma (dendrogram) – klasterlarni ketma-ket jarayonining qadamlaridan biriga mos keluvchi n darajalardan iborat daraxt ko'rinishidagi diagramma.

Dendrogrammani daraxt ko'rinishli sxema, klasterlarni birlashtirish daraxti, ierarxik tuzilish daraxti deb ham atash mumkin. Dendrogramma o'zi bilan turlicha ierarxik darajalarda o'zgaruvchi obyektning ichki guruhlarini tasvirlaydi.

Dendogrammalar tuzilishining ko'pgina usullari mavjud. Dendogrammada obyektlar vertikal va gorizontal joylashtirilishi mumkin.

O'xshashlik o'lchamlari. Obyektlar orasidagi masofalarni hisoblash uchun metrikalar yoki masofalarni funksiyalari deb nomlanuvchi turlicha o'xshashlik o'lchamlari (o'xshashlik o'lchamlari) dan foydalaniladi. Evklid masofasi – o'xshashlikning eng taniqli o'lchamidir.

Evklidning kvadratik masofasi. Bir-biridan uzoqroq bo'lgan obyektlarga katta vaznlarni berish uchun standart evklid masofasini kvadratga ko'tarish yo'li bilan evklid masofasini kvadratidan foydalanish mumkin.

Manxetten masofasi (shahar kvartallari masofasi), shuningdek “xemming” yoki “siti-blok” masofasi deb nomlanadi. Ushbu masofa koordinatalar bo'yicha o'rtacha farqlari sifatida hisoblanadi. Masofaning ushbu o'lchovi aksariyat hollarda evklid masofasi o'xshash hisob-kitoblari natijalariga olib keladi. Biroq ushbu o'lcham uchun alohida chiqindilar ta'siri evklid masofasidan foydalanishga nisbatan kichik bo'ladi, chunki koordinatalar kvadratga ko'tarilmaydi.

Chebisev masofasi. Ushbu masofadan “farqli” ikkita obyektning aniqlash zarur bo'lgan paytda foydalaniladi, agar ular qandaydir bir o'lchamda farq qiladigan bo'lsa.

Kelishmovchilik darajasi. Agar ma'lumotlar kategorial bo'lib hisoblansa, ushbu masofa hisoblanadi.

Birlashtirish yoki bog'lanishlar usullari.

Har bir obyekt o'zi bilan alohida klasterni tasvirlagan paytda, ushbu obyektlar orasidagi masofalar tanlangan o'lcham bilan aniqlanadi. Quyidagi savol tug'iladi – klasterlar orasidagi masofa qanday aniqlanadi? Ikkita klaster uchun birlashtirishlar yoki bog'lanishlar usullari deb nomlanuvchi turli xil qoidalar mavjud.

Yaqin qo'shni yoki birlamchi bog'lanish. Bu yerda ikkita klaster orasidagi masofa turli xil klasterlarda ikki eng yaqin obyektlar (yaqin qo'shnilar) orasidagi masofa bilan aniqlanadi.

Ushbu usul har qanday murakkab shaklli klasterlarni ajratishga imkon beradi, agar bunday klasterlarning turli qismlari elementlarning bir-biriga yaqin zanjirlar bilan bog'langan bo'lsa. Ushbu klaster usuli natijasida uzun “zanjirli” yoki “tolali” klasterlar bilan ifodalanadi, “bir-biriga bog'langan” alohida elementlar bo'lib, ular tasodifiy tarzda boshqalariga qaraganda bir-biriga yaqinroq bo'lishadi.

Eng uzoq qo'shnilar yoki to'liq bog'lanish usuli. Bu yerda klasterlar orasida turli xil klasterlardagi ixtiyoriy ikki obyekt orasidagi eng katta masofa aniqlanadi (ya'ni “eng uzoq qo'shnilar orasi”). Turlicha “daraxtzor”dan obyektlar kelgan paytda usulni qo'llash yaxshi. Agar klasterlar qandaydir tarzda cho'zilgan shakli yoki ularning tabiiy turi “zanjirli” bo'lib hisoblanadi, u holda ushbu usuldan foydalanilmaydi.

Vard usuli (Ward's method). Klasterlar orasidagi masofa sifatida ularni birlashtirish natijasida olinuvchi klasterlar markazigacha obyektlar masofalari kvadratlari yig'indisini o'sishi olinadi. Klaster tahlilining boshqa usullaridan farqli klasterlar orasidagi masofani baholash uchun, bu yerda dispersion tahlil usullaridan foydalaniladi. Algoritmning har bir qadamida shunday maqsad funksiyasini minimal oshirishga, ya'ni ichki guruhlararo kvadratlari yig'indisiga olib keluvchi ikki klaster birlashtiriladi. Ushbu usul yaqin joylashgan klasterlarni birlashtirishga qaratilgan va kichik hajmli klasterlarni yaratishga “intiladi”.

Vaznsiz juft o'rtacha usul (vaznsiz juft o'rtacha arifmetik usul - unweighted pair-group method using arithmetic averages, UPGMA). Ikkita klaster orasidagi masofa sifatida ulardagi obyektning barcha juftlari orasidagi o'rtacha masofasi olinadi. Ushbu usul, agar obyektlar haqiqatan ham turli xil “daraxtzor”dan kelib chiqsa, klasterlar teng bo'lmagan o'lchamlari asosida klasterlarni “zanjirli” turi mavjud bo'lganda foydalaniladi.

Vaznli juft o'rtacha usul (weighted pair-group method using arithmetic

averages, WPGMA). Ushbu usul vaznsiz juft o'rtacha usulga o'xshash, farqi shundan iboratki, bu yerda vaznli koeffitsient sifatida klaster o'lchami foydalaniladi (klasterdagi obyektlar soni). Ushbu usuldan aynan turlicha o'lchamli klasterlar to'g'risidagi taxminlar mavjud bo'lgan paytda qo'llash tavsiya qilinadi.

Vaznsiz markaziy usul (vaznsiz juft markaziy o'rtachalashtirish usuli - unweighted pair-group method using the centroid average). Ikkita klaster orasidagi masofa sifatida ushbu usulda ularning og'irlik markazlari orasidagi masofa olinadi.

Vaznli markaziy usul (vaznli juft markaziy o'rtachalashtirish usuli - weighted pair-group method using the centroid average, WPGMC). Ushbu usul avvalgisiga o'xshash, farqi shundaki, klasterlar o'lchamlari (ulardagi obyektlar soni) orasidagi farqlarni hisobga olish uchun vazndan foydalaniladi. Klasterlar o'lchamlaridagi mavjud farqlarga nisbatan taxminlar mavjud bo'lgan hollarda ushbu usuldan foydalanish afzaldir.

Klasterlar miqdorini aniqlash.

Klasterlar sonini aniqlash masalasi mavjud. Ba'zan ushbu sonni aprior aniqlash mumkin. Biroq ko'p hollarda klasterlar soni obyektlar to'plamini aglomerativ bo'lish jarayonida aniqlanadi.

E deb nomlanuvchi mezon koeffitsientining bosqichma-bosqich o'sishi obyektlarni guruhlash jarayoniga muvofiq keladi. *E* mezon qiymatlarini keskin oshishini tadqiq etilayotgan ma'lumotlar to'plamida haqiqatan mavjud klasterlar sonining xarakteristikasi sifatida aniqlash mumkin. Shunday qilib, ushbu usul obyektlarni kuchli bog'langan holatidan zaif bog'langan holatiga o'tishni tavsiflovchi ma'lum bir koeffitsientning keskin oshishini aniqlashga olib keladi.

Klaster tahlilning iterativ usullari.

Ko'p sonli kuzatuvlar klaster tahlilning ierarxik usullari bilan mos emas. Bunday hollarda ierarxik bo'lmagan usullardan foydalaniladi, ular dastlabki jamlanmani maydalashning iterativ usullarini o'zi bilan tasvirlovchi bo'lishga asoslanadi. Bo'lish jarayonida yangi klasterlar to'xtatish qoidasi amalga oshirilmagunicha shakllantiriladi.

Bunday ierarxik bo'lmagan klasterlash ma'lumotlar to'plamini alohida klasterlarni ma'lum miqdoriga bo'lishdan iborat. Ikkita yondashuv mavjud. Birinchisi dastlabki ma'lumotlarning ko'p o'lchamli fazosida eng zich

maydonlar sifatidagi klasterlarni chegaralarini aniqlashdan iborat, ya'ni "nuqtalarni quyuqlashishi" ko'p bo'lgan klasterlarni aniqlash. Ikkinchi yondashuv obyektlarni farqlarini o'lchamlarini minimallashtirishdan iborat.

k-o'rtacha algoritmi (k-means).

Tez klasterlash tahlili k-o'rtacha algoritmining noierarxik usullari orasida eng ko'p tarqalganidir. Ushbu usuldan foydalanish imkoniyati uchun klasterlar soniga nisbatan dastlabki taxminlarni talab qilmaydigan ierarxik usullardan farqli klasterlarning eng ehtimolli miqdori to'g'risidagi farazga ega bo'lish zarur.

Taqrizni jo'natish

Tarix saqlangan

k-o'rtacha algoritmi bir-biridan mumkin bo'lgan katta masofada joylashgan k klasterlarni quradi. k-o'rtacha algoritmini hal qiluvchi masalalarning asosiy turi, - klasterlar soniga nisbatan taxminlarni (faraz) mavjudligi, bunda ular qanchalik turlicha bo'lishi shart bo'lsa, shunchalik bu mumkin bo'ladi. k sonlarni tanlash nazariy mulohazalar yoki ichki his-tuyg'ularni avvalgi tadqiqotlarining natijalariga asos bo'lishi mumkin.

Algoritmning umumiy g'oyasi: kuzatish klasterlarini ma'lum bir belgilangan k soni klasterlar bilan shunday solishtiriladiki, bunda klasterdagi o'rtacha (barcha o'zgaruvchilar uchun) bir-biridan maksimal darajada farq qilishi mumkin.

Algoritmning tavsifi:

1. Klasterlar bo'yicha obyektlarni birlamchi taqsimlash.

k soni tanlanadi, birinchi qadamda ushbu nuqtalar klasterlar "markazlari" bo'lib hisoblanadi. Har bir klasterga bitta markaz to'g'ri keladi. Boshlang'ich markazlarni tanlash quyidagi tarzda amalga oshiriladi:

- boshlang'ich masofani maksimallashtirish uchun k-kuzatuvlarni tanlash;
- k-kuzatishlarni tasodifiy tanlash;
- birinchi k-kuzatishlarni tanlash.

Natijada har bir obyekt ma'lum bir klasterga tayinlanadi.

2. Iterativ jarayon.

Kiyinroq klasterlarni koordinatalari bo'yicha o'rtachalari bo'lgan klasterlar markazlari hisoblanadi.

Markazlarni hisoblash jarayoni va obyektlarni qayta taqsimlash quyidagi shartlardan biri bajarilmagunicha davom etadi:

- klaster markazlari turg'unlashadi, ya'ni barcha kuzatuvlar;
- joriy iteratsiyadan oldin tegishli bo'lgan klasterga tegishli;
- iteratsiyalar soni iteratsiyalarning maksimal soniga teng bo'ladi.

Klasterlar sonini tanlash murakkab masala hisoblanadi. Ushbu son to'g'risida taxminlar yo'q bo'lsa, olingan natijalarni solishtirgan holda 2ta klasterni yaratishni tavsiya qiladi, so'ngra 3, 4, 5 ta va hokazo.

Klasterlash sifatini tekshirish.

k-o'rtacha usuli bilan klaster tahlil natijalarini olgandan so'ng klasterlashning to'g'riligini tekshirish kerak bo'ladi (ya'ni klasterlar bir-biridan qanchaga farq qiladi).

Buning uchun har bir klaster uchun o'rtacha qiymatlar hisoblanadi. Yaxshi klasterlashda barcha o'lchamlar yoki loaqal ularning aksariyat qismi uchun keskin farqlanuvchi o'rtachani olinishi zarur.

k-o'rtacha algoritmining afzalliklari:

- foydalanish qulayligi;
- foydalanish tezligi;
- algoritmnin tushunariligi va shaffofligi.

k-o'rtacha algoritmining kamchiliklari:

• algoritmnin o'rtacha qiymatni buzishi mumkin bo'lgan chiqindilarga ta'sirchanligi yuqori darajada. Ushbu muammoning mumkin bo'lgan yechimi algoritm modifikatsiyasi - k-mediana algoritmidan foydalanish hisoblanadi;

• katta hajmli ma'lumotlar bazasida sekin ishlashi mumkin. Berilgan muammoning mumkin bo'lgan yechimi ma'lumotlar tanlanmasidan foydalanish hisoblanadi.

PAM algoritmi (Partitioning Around Medoids).

PAM- k-o'rtacha algoritmi modifikatsiyasi, k- mediana algoritmi (k-medoids) bo'lib hisoblanadi. Algoritm k-o'rtacha algoritmiga nisbatan ma'lumotlar shovqiniga va chiqindilariga sezgirligi pastroq, chunki mediana chiqindilarga kamroq ta'sir qiladi. PAM kichik ma'lumotlar bazasida samarali, biroq undan katta hajmli ma'lumotlar to'plamida foydalanilmasligi kerak.

O'lchamni dastlabki qisqartirish.

Agar dastlabki o'zgaruvchilar to'plami o'rniga o'zgaruvchilar orasidagi bog'lanishlarni to'g'risidagi siqiq ko'rinishdagi axborotdan tashkil topgan umumlashgan o'zgaruvchilar yoki mezonlardan foydalanilsa, tushunarliroq va shaffofroq natijalarni olish mumkin bo'ladi. Ya'ni ma'lumotlar o'lchamini

kamaytirish masalasi tug'iladi. Bu masala turlicha usullar yordamida hal qilinishi mumkin; eng ko'p tarqalgan usullardan biri – omilga oid tahlil.

Omilga oid tahlil - bu o'zgaruvchilar qiymatlari orasidagi o'zaro bog'liqliklarni o'rganish uchun qo'llaniluvchi usul. Omilga oid tahlil ikkita maqsadni ilgari suradi:

- o'zgaruvchilar sonini qisqartirish;
- o'zgaruvchilarni tasniflash;
- o'zgaruvchilar orasidagi o'zaro bog'lanishlar tuzilishini aniqlash.

Shunga ko'ra, faktorli tahlildan ma'lumotlar o'lchamini qisqartirish yoki tasniflash masalalarini yechish uchun foydalanish mumkin.

Omili tahlil natijasida belgilangan mezonlar yoki asosiy omillar o'zgaruvchilar orasidagi mavjud bog'lanishlar to'g'risidagi siqiq axborot ko'rinishida bo'ladi. Ushbu axborot klasterlashning eng yaxshi natijasini olishga va klasterlar semantikasini eng yaxshi tushuntirishga imkon beradi. Omillarning o'ziga ma'lum ma'noni berish mumkin.

Omili tahlil yordamida ko'p sonli o'zgaruvchilar omillar deb nomlanuvchi oz sonli mustaqil ta'sir qiluvchi kattaliklarga keltiriladi.

"Siqiq" ko'rinishdagi omil bir qancha o'zgaruvchilar to'g'risidagi axborotni o'z ichiga oladi. Bir omilda o'zaro kuchli bog'liq bo'lgan o'zgaruvchilar birlashtiriladi. Omili tahlil natijasida ko'rib chiqilayotgan o'zgaruvchilar orasidagi bog'lanishlarni imkon boricha batafsil tushuntiruvchi shunday kompleks omillar qidiriladi.

Omili tahlilning birinchi qadamida o'zgaruvchilar qiymatlarini standartlashtirish amalga oshiriladi. Omili tahlil shunday taxminga asoslanadiki, tahlil qilinuvchi o'zgaruvchilar kam sonli ba'zi yashirin omillarga nisbatan bilvosita namoyon bo'ladi.

Omili tahlil – bu kuzatiluvchi o'zgaruvchilar orasidagi yashirin bog'lanishlarni chiqarish va tahlil qilishga mo'ljallangan usullar jamlanmasi. Yashirin bog'lanishlar latent deb ham ataladi.

Omili tahlil usullaridan biri – asosiy komponent usuli - omillarni bir-biridan mustaqilligi to'g'risidagi taxminlarga asoslangan.

Ierarxik va noierarxik usullardan birini tanlashda ularning quyidagi xususiyatlarini hisobga olish zarur.

Noierarxik usullar shovqinga va chiqindilarga bo'lgan munosabat bo'yicha eng yuqori turg'unlikni, metrikani nokorrekt tanlovini, klasterlashda ishtirok etuvchi to'plamdagi ahamiyatsiz o'zgaruvchilarni qo'shishni

aniqlaydi. Ushbu usulning afzalliklari uchun badal "apriorlar" so'zidir. Klasterlar sonini, iteratsiyalar miqdorini yoki to'xtatish qoidasini, shuningdek klasterlashning ayrim boshqa parametrlarini avvaldan aniqlash zarur.

Agar klasterlar sonlariga nisbatan taxminlar bo'lmasa, ierarxik algoritmlardan foydalanish tavsiya etiladi. Biroq tanlanma hajmi buni qilishga imkon bermasa, mumkin bo'lgan yo'l turli xil miqdorli klasterlar bilan qator tajribalarni o'tkazish, masalan ikkita guruh ma'lumotlar jamlanmasini bo'lishni boshlash va asta-sekin ularni miqdorini oshirgan holda natijalar solishtirish. Natijalarning ushbu "xilma-xilligi" tufayli klasterlashning yetarlicha katta egiluvchanligiga erishiladi.

Ierarxik usullar noierarxik usullardan farqli klasterlar sonini aniqlashdan bosh tortadi, lekin joylashtirilgan klasterlarning to'liq daraxtini quradi.

Klasterlashning ierarxik usullarini murakkabligi: ma'lumotlar to'plami hajmining cheklanganligi; yaqinlashish o'lchamini tanlash; olingan tasniflarni egiluvchan emasligi.

Usullarning ushbu guruhlarini noierarxik usullari bilan solishtirishda afzalligi – ularning ravshanligi va ma'lumotlar tuzilishi to'g'risida batafsil tushuncha olish imkoniyatidir.

Ierarxik usullarni foydalanishda ma'lumotlar to'plamida chiqindilarni yetarlicha oson aniqlash imkoni mavjud va natijada ma'lumotlar sifatini yaxshilash mumkin bo'ladi. Ushbu protsedura klasterlashning ikki bosqichli algoritmi asosida yotadi. Ma'lumotlarning bunday to'plami keyinchalik noierarxik klasterlashni o'tkazish uchun foydalanilishi mumkin.

Ierarxik usullar katta hajmli ma'lumotlar to'plami bilan ishlashi mumkin emas, ayrim tanlanmalardan, ya'ni ma'lumotlar qismidan foydalanish esa ushbu usullarni qo'llashga imkon berishi mumkin.

Klasterlash natijalari yetarlicha statistik asoslanishlarga ega bo'lmasligi mumkin. Boshqa tomondan klasterlash masalasini yechishda olingan natijalarni nostatistik talqin qilish, shuningdek klaster tushunchasining ko'plab xilma-xil variantlari qabul qilinadi. Bunday nostatistik talqin qilish tahlilchi uchun uni qoniqtiruvchi klasterlash natijalarini olishga imkon beradiki, boshqa usullardan foydalanishda tez-tez qiyinchiliklar bo'ladi.

Juda katta ma'lumotlar bazasi paydo bo'lishi bilan bog'liq holda klasterlash algoritmini qoniqtirishi shart bo'lgan yangi talablar yuzaga keldi.

Ulardan asosiylari bu - algoritmni masshtablashtirish (bu shuni anglatadiki, ma'lumotlar hajmini o'sishi bilan o'qitishga va klasterlashga sarf

qilinadigan vaqt chiziqi o'sadi).

Klasterlash algoritmini qoniqtirishi shart bo'lgan boshqa xususiyatlarini ta'kidlash muhim: kiruvchi ma'lumotlar tartibidan natijalarni mustaqilligi; kiruvchi ma'lumotlardan algoritmning parametrlarini mustaqilligi.

Ierarxik klasterlash usullari boshqa usullar bilan birlashtirilgan algoritmlar ishlab chiqilgan. Bunday algoritmlarga quyidagilar tegishli bo'ladi: BIRCH, CURE, CHAMELEON, ROCK.

BIRCH algoritmi

(Balanced Iterative Reducing and Clustering using Hierarchies).

Klasterlarni umumlashgan ko'rinishlari tufayli klasterlash tezligi oshadi, bunda algoritm katta masshtablashtirishga ega bo'ladi. Ushbu algoritmda ikki bosqichli klasterlash jarayoni amalga oshirilgan.

Birinchi bosqichda taxminiy klasterlar to'plami shakllantiriladi. Ikkinchi bosqichda klasterlashning boshqa algoritmlari klasterlarni aniqlashga qo'llaniladi – operativ xotirada ishlash uchun qulay.

Ushbu algoritmni tavsiflovchi o'xshashlik. Agar ma'lumotlarning har bir elementini stol ustida yotgan sharik deb o'ylansa, u holda shariklar klasterlarini tennis shariklari bilan "almashish" mumkin va tennis shariklari klasterlarini batafsilroq o'rganishga o'tish mumkin. Shariklar soni yetarlicha ko'p, biroq tennis shariklarini diametrini shunday tanlash mumkinki, ikkinchi bosqichda an'anaviy klasterlash algoritmlarini qo'llagan holda klasterlarni haqiqiy murakkab shaklini aniqlash mumkin bo'ladi.

Wave Cluster algoritmi.

WaveCluster o'zi bilan to'liqlikni aylantirish asosida klasterlash algoritmini tasvirlaydi. Algoritmning boshida ma'lumotlarning fazosiga ko'p o'lchamli panjarani o'rnatish orqali ma'lumotlar umumlashtiriladi. Algoritmning keyingi qadamlarida alohida nuqtalarni emas, balki panjaraning bitta katakchasiga tushadigan nuqtalarning umumlashgan xususiyatlarini tahlil qilinadi. Bunday umumlashtirish natijasida zaruriy axborot operativ xotirada joylanadi. Keyingi qadamlarda klasterlarni aniqlash uchun algoritm umumlashgan ma'lumotlarga to'liqlikni aylantirishni qo'llaydi.

WaveClustering asosiy xususiyatlari:

- amalga oshirishning murakkabligi;
- algoritm ixtiyoriy shakldagi klasterlarni topishi mumkin;
- algoritm shovqinlarga sezgir emas;

- algoritmni kichik o'lchamli ma'lumotlarga qo'llaymiz.

CLARA algoritmi

(Clustering LARge Applications).

CLARA algoritmida ma'lumotlar bazasidan namunalari to'plamini oladi. Klasterlash namunalarning har biriga qo'llaniladi, algoritmning chiqishida eng yaxshi klasterlash taklif etiladi.

Katta ma'lumotlar bazasi uchun algoritm PAM algoritmiga nisbatan samaraliroq. Algoritm samaradorligi ma'lumotlar to'plamining namunasi sifatida tanlanganlarga bog'liq. Tanlangan to'plamda barcha ma'lumotlar to'plamiga yaxshi klasterlashni bermasligi mumkin.

Clarans, CURE, DBScan algoritmi

Clarans algoritmi (Clustering Large Applications based upon RANdomized Search) klasterlash masalasini grafdagi tasodifiy qidiruv sifatida shakllantiradi. Ushbu algoritm natijasida graf tugunlari jamlanmasi o'zi bilan ma'lumotlar to'plamini foydalanuvchi tomonidan belgilangan klasterlar soniga bo'linishini tasvirlaydi. Klasterlar olgan "Sifat" mezoniy funktsiya yordamida aniqlanadi. Clarans algoritmi muvofiq yechimni qidirishda ma'lumotlar to'plamini barcha mumkin bo'lgan bo'linishlarini saralaydi. Yechimni qidirish shunday tugunda to'xtatiladiki, bu tugun lokal minimumlarni avvaldan belgilangan soni orasida minimumga erishadi.

Yangi mashtablashtiriluvchi algoritmlar orasida CURE algoritmini belgilash mumkin- ierarxik klasterlash algoritmi va DBScan algoritmi, bu yerda klaster tushunchasi zichlik (density) tushunchasidan foydalanish bilan shakllantiriladi.

BIRCH, Clarans, CURE, DBScan algoritmlarini asosiy kamchiliklari shundan iboratki, ular nuqtalar zichligining ayrim bo'sag'alarini belgilashni talab qiladi, bu esa har doim ham qabul qilinmaydi. Ushbu cheklanishlar tavsiflangan algoritmlarning juda katta ma'lumotlar bazalariga yo'naltirilganligi va katta hisoblash resurslaridan foydalana olmasligi bilan ta'minlangan.

25. ASSOTSIATIV QOIDALARNI IZLASH USULI

Assotsiativ qoidalarini qidirishdan maqsad (association rule) ma'lumotlar bazasida bog'liq hodisalar orasida qonuniyatlarni aniqlash hisoblanadi.

Assotsiativ qoidalarini qidirish masalasi (association rule mining) birinchi marta supermarketlarda bajariluvchi odatiy xaridlar shablonlarini topish uchun taklif etilgan, shuning uchun ham uni bozor savatini tahlili deb ham atashadi

Bozor savati – xaridor tomonidan bitta bitim doirasida sotib olingan mahsulotlar to'plami.

Tranzaksiya - bu bir vaqtni o'zida sodir bo'lgan hodisalar to'plami. Savdo kompaniyalari o'zlarining butun faoliyati davomida barcha biznes-operatsiyalarni ro'yxatdan o'tkazish orqali ulkan bitimlar to'plamlarini to'plashadi. Har bir tranzaksiya o'zi bilan mijoz tomonidan bitta tashrif davomida sotib olingan tovarlar to'plamini ifodalaydi. Olinganlar shablonni tahlil qilish natijasida to'plam ma'lumotlaridan iborat tovarlar ro'yxatini va tranzaksiyalar sonini o'z ichiga oladi. Tranzaksion yoki operatsion ma'lumotlar bazasi (Transaction database) o'zida ikki o'lchamli jadvalni ifodalaydi, bu jadval tranzaksiya raqami (TID) va ushbu bitim vaqtida sotib olingan xaridlar ro'yxatidan iborat.

TID – kamyob identifikator bo'lib, har bir savdo yoki bitimni aniqlaydi. Xaridorga oid bitimdan iborat tranzaksion ma'lumotlar bazasi uchun 25-misolda keltirilgan. Jadvalda birinchi ustun bitim raqamini belgilaydi, ikkinchi ustunda belgilangan bitim vaqtida sotib olingan tovarlar keltirilgan.

TID	Приобретенные покупки
100	Хлеб, молоко, печенье
200	Молоко, сметана
300	Молоко, хлеб, сметана, печенье
400	Колбаса, сметана
500	Хлеб, молоко, печенье, сметана

25.1-rasm. Tranzaksion ma'lumotlar bazasiga doir misol
Mavjud ma'lumotlar bazasi asosida xaridlar orasidagi qonuniyatlarni topish zarur.

Tez-tez duch keladigan shablonlar va namunalar.

Aytaylik, tranzaksiya ma'lumotlari bazasi D mavjud bo'lsin. Tovarlar qiymatlariga o'zgaruvchilarni tayinlaymiz (25.1-rasm).

$Non=a$, $Sut=b$, $Pechenie=c$, $Smetana=d$, $Kolbasa=e$, $Konfetlar=f$. Tovarlar to'plamini ko'rib chiqamiz (Itemset), masalan, $\{Non, sut, pechenie\}$. Ushbu to'plamni o'zgaruvchilar yordamida ifodalaymiz: $abc=\{a,b,c\}$

Tovarlarni bu to'plamni bizni ma'lumotlar bazamizda uch marta uchraydi, ya'ni tovarlarni bu to'plamini qo'llab-quvvatlash 3 ga teng: $SUP(abc)=3$.

Qo'llab-quvvatlashning minimal darajasida tovarlar to'plami abc tez-tez duch kelinadigan shablon bo'lib hisoblanadi.

Ma'lum ma'lumotlar to'plamidan tashkil topgan tranzaksiyani miqdori yoki foizi qo'llab-quvvatlash deb nomlanadi. Berilgan tovarlar to'plami uchun foizli munosabatda ifodalangan qo'llab-quvvatlash 50% ga teng. Qo'llab-quvvatlashni ba'zan to'plamning ta'minoti deb ataladi.

Shunday qilib, to'plam qiziqish uyg'otadi, agar uning qo'llab-quvvatlashi foydalanuvchi tomonidan belgilangan minimal qiymatdan (min support) yuqori bo'lsa. Ushbu to'plamlar tez-tez duch keladigan (frequent) deb nomlanadi.

Assotsiativ qoidalar xususiyatlari.

Assotsiativ qoida quyidagi ko'rinishga ega: "A hodisadan B hodisa kelib chiqadi".

Bunday ko'rinishdagi tahlil natijasida quyidagi ko'rinishdagi qonuniyatni o'rnatish mumkin: "Agar tranzaksiyada A tovarlar to'plami (yoki elementlar to'plami) uchrasa, u holda shu xulosaga kelish mumkinki, ushbu tranzaksiyada B elementlar to'plami paydo bo'lishi kerak". Bunday qonuniyatlarni o'rnatilishi bizga assotsiativ deb nomlanuvchi juda sodda va tushunarli qoidalarni topish imkoniyatini beradi.

Assotsiativ qoidaning asosiy xususiyatlari bo'lib, qoidani qo'llab-quvvatlash va ishonchliligi hisoblanadi. 25- rasmda ma'lumotlar uchun "sutni xarididan pechenieni xaridi kelib chiqadi" qoidasini ko'rib chiqamiz.

Agar barcha to'plamdan tranzaksiya $s\%$ bir vaqtni o'zida A va B elementlar to'plamini tashkil qilsa, boshqacha aytganda, ikkala tovarni tashkil qilsa qoida s qo'llab-quvvatlashga ega bo'ladi.

Sut – bu A mahsulot, pechenie – bu B mahsulot. "Sutni xarididan pechenieni xaridi kelib chiqadi" qoidasini qo'llab-quvvatlash 3 ga teng, yoki

50%.

Qoidani ishonchliligi B hodisaning "A hodisadan kelib chiqish ehtimolligini ko'rsatadi. 'A dan B kelib chiqadi'" qoidasi γ ishonchlilik bilan amal qiladi, agar A elementlar to'plamidan iborat barcha to'plamlardan olingan γ tranzaksiyalar foizi B elementlar to'plamini o'z ichiga olsa.

Sutdan iborat tranzaksiya soni 4 ga, pecheniedan iborat tranzaksiya soni 3 ga teng, qoida ishonchliligi 75% ga teng.

Assotsiativ qoida qo'llab-quvvatlashi va ishonchliligini chegaralari.

Assotsiativ qoidalarni qidirish algoritmlaridan foydalanish yordamida "A dan B kelib chiqadi" ko'rinishidagi barcha mumkin bo'lgan qoidalarni qo'llab-quvvatlash va ishonchlilikning turlicha qiymatlari bilan olish mumkin. Biroq ko'p hollarda qoidalar miqdorini avvaldan qo'llab-quvvatlash va ishonchlilikning o'rnatilgan minimal va maksimal qiymatlarini cheklash zarur.

Agar qoidaning qo'llab-quvvatlashi o'ta katta bo'lsa, u holda algoritmi ishlash natijasida aniq va taniqli qoidalar topiladi. Qo'llab-quvvatlashning juda kichik qiymati juda ko'p sonli qoidalarni aniqlashga olib keladi, bu qoidalarning katta qismi asoslanmagan bo'lishi mumkin, ma'lum va aniq emas. Shunday qilib, bir tomondan aniq bo'lmagan qoidalarni topishni va ikkinchi tomondan ularning asoslanganligini ta'minlovchi oraliqni – "oltin o'rtachani" topish kerak.

Agar ishonchlilik darajasi o'ta kichik bo'lsa, u holda qoidaning bahosi sezilarli ikkilanishlarni chaqiradi. Masalan, ishonchliligi 3% li qoidani shartli ravishda qoida deb nomlash mumkin.

Assotsiativ qoidalarni qidirish usullari

AIS algoritmi. AIS algoritmdama'lumotlar bazasini skanerlash vaqtida to'plamlar jamlanmasi nomzodlari yaratiladi va tez hisoblab chiqiladi.

SETM algoritmi. Algoritm AIS kabi SETM ma'lumotlar bazasini aylantirishga asoslangan nomzodlarni ham shakllantiradi. Nomzodni shakllantirish uchun SQL tilni birlashtirishning standart amaldan foydalanish maqsadida SETM nomzodni shakllanishini ularni hisoblashdan ajratadi. AIS va SETM algoritmlar noqulayligi tez-tez duch kelmasligini ko'rsatishi natijasida juda ko'p nomzodlarning yaratilishi va hisoblashning ortiqchaligi. Ularning ishlashlarini yaxshilash uchun Apriori algoritmi taklif etiladi.

Apriori algoritmi berilgan algoritmning ishlashi bir nechta bosqichlardan tashkil topadi, bosqichlarning har biri quyidagi qadamlardan iborat:

- nomzodlarni shakllantirish;
- nomzodlarni hisoblash.

Nomzodlarni shakllantirish (candidate generation) - algoritm ma'lumotlar bazasini skaner qiladigan bosqich, i - elementli nomzodlar (i - nomer etapa) to'plamini yaratadi. Ushbu bosqichda nomzodlarni qo'llab-quvvatlash hisobga olinmaydi.

Nomzodlarni hisoblash (candidate counting) - har bir i - elementli nomzodlarni hisoblanuvchi bosqich. Bu yerda foydalanuvchi tomonidan o'rnatilgan qo'llab-quvvatlashi (min_sup) minimumdan kichik nomzodlar kesimi amalga oshiriladi. Qolgan i - elementli to'plamlar tez-tez duch kelinadigan deb nomlanadi.

D ma'lumotlar bazasi misolida Apriori algoritmini ishlashini ko'rib chiqamiz. Algoritm ishlashini 25.2- rasmda tasvirlangan. Qo'llab-quvvatlashning minimal darajasi 3 ga teng.

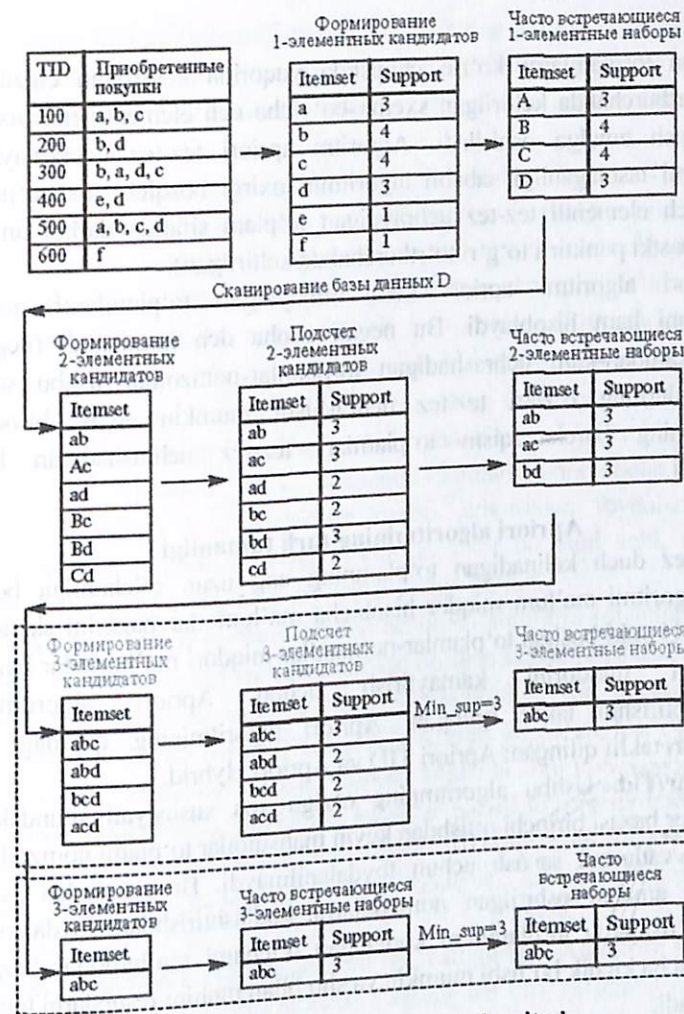
Birinchi bosqichda bir elementli nomzodlarni shakllantirish amalga oshiriladi. Keyin algoritm bir elementli to'plam qo'llab-quvvatlashini hisoblaydi.

Qo'llab-quvvatlash darajasi o'rnatilganidan kichik, ya'ni 3 ga teng to'plamlar bekor qilinadi. Bizni misolimizda 1 ga teng e va f to'plamlar bo'lib, mahsulotlarni qolgan to'plamlari mahsulotlarni tez-tez duch kelinadigan bir elementli to'plamlarini hisoblaydi: bu to'plamlar a, b, c, d .

Keyin ikki elementli nomzodlarni shakllantirish amalga oshiriladi, ularni qo'llab-quvvatlashi hisoblab chiqiladi va 3 dan kichik qo'llab-quvvatlashlar darajasi bilan to'plamlarni kesiladi. Tez-tez duch kelinadigan to'plamlar bo'lib hisoblanuvchi ab, ac, bd qolgan ikki elementli mahsulotlar to'plamlari algoritmnining keyingi ishlash jarayonida ishtirok etadi.

Agar algoritmnining ishlashini to'g'ri chiziqli qaraydigan bo'lsak, algoritmnining oxirgi bosqichida uch elementli mahsulotlar to'plamini shakllantiradi: abc, abd, bcd, acd , ularning qo'llab-quvvatlashini hisoblaydi va 3 dan kichik qo'llab-quvvatlash darajasiga ega to'plamlarni kesadi. abc mahsulotlar to'plamini tez-tez duch kelinadigan deb nomlash mumkin.

Biroq Apriori algoritmi avvalgi algoritmnining ishlash bosqichlarida nomzodlarni kesish to'g'risidagi axborotlar asosida tez-tez duch kelinmaydigan bo'lishini anglagan holda apriorlarni kesib, nomzodlar miqdorini kamaytiradi.



25.2-рasm. Apriori algoritmi

Nomzodlarni kesish shu to'g'risidagi taxminga asoslangan holda amalga oshiriladiki, mahsulotlarning tez-tez duch kelinadigan to'plaming barcha qism to'plamlari tez-tez duch keladigan bo'lishi shart.

Agar to'plamda avvalgi bosqichda tez-tez duch kelinmaydigan deb belgilangan qism to'plam topilgan bo'lsa, ushbu nomzod allaqachon shakllantirish va nomzodlarni sanashga kiritilmaydi.

Shunday qilib, ad, bc, cd mahsulotlar to'plamlari tez-tez uchramaydigan to'plam sifatida tashlab yuborildi, algoritm abd, bcd, acd mahsulotlarni ko'rib

chiqmadi.

Ushbu to'plamlarni ko'rib chiqishda yuqorida keltirilgan chiziqchali to'g'ri to'rtburchakda keltirilgan sxema bo'yicha uch elementli nomzodlarni shakllantirish amalga oshiriladi. Algoritm apriori tez-tez uchramaydigan to'plamlarni tashlaganligi sababli algoritmn oxirgi bosqichi abc to'plamni yagona uch elementli tez-tez uchraydigan to'plam sifatida darhol aniqladi (bosqich pastki punktirli to'g'ri to'rtburchakda keltirilgan).

Apriori algoritmi apriori kesa olmaydigan to'plamlarni qo'llab-quvvatlashni ham hisoblaydi. Bu negativ soha deb nomlanadi (negative border), kamdan-kam uchrashadigan to'plamlar-nomzodlar ushbu sohaga tegishli, ularning o'zlari tez-tez uchrashishi mumkin emas, biroq bu to'plamlarning barcha qism to'plamlari tez-tez uchrashadigan bo'lib hisoblanadi.

Apriori algoritmining turli tumanligi

Tez-tez duch kelinadigan to'plamning eng uzun o'lchamiga bog'liq Apriori algoritmi ma'lum miqdor hisobicha ma'lumotlar bazasini skanerdan o'tkazadi. U yoki boshqa to'plamlar-nomzodlar miqdori ma'lumotlar bazasini skanerlarini miqdorini kamaytirish uchun Apriori algoritmining optimallashtirishni taklif qilingan. Apriori algoritmining quyidagi turli ko'rinishlari taklif qilingan: Apriori TID va Apriori Hybrid.

AprioriTid. Ushbu algoritmnning o'ziga xos xususiyati shundaki, D ma'lumotlar bazasi birinchi o'tishdan keyin mahsulotlar to'plami nomzodlarini qo'llab-quvvatlashni sanash uchun foydalanilmaydi. Bu maqsadda avvalgi o'tishlarda amalga oshirilgan nomzodlarni kodlashtirishdan foydalaniladi. Keyingi o'tishlarda to'plamlarni kodlangan o'lchami ma'lumotlar bazasiga nisbatan ancha kichik bo'lishi mumkin va shu bilan muhim resurslarni tejashga imkon beradi.

AprioriHybrid. Apriori va AprioriTid algoritmlarni ishlash vaqtlarini tahlil qilish shuni ko'rsatadiki, AprioriTidga nisbatan Apriorieng avvalgi o'tishlarda yaxshiroq ishlaydi; biroq eng keyingi o'tishlarda AprioriTid Aprioridan ko'ra yaxshiroq ishlaydi. Bundan tashqari ular to'plamlar-nomzodlarni shakllantirishning xuddi shu protsedurasidan foydalanadilar. Ushbu kuzatishga asoslangan AprioriHybrid algoritmi Apriori va AprioriTid algoritmlarni eng yaxshi xususiyatlarini birlashtirish maqsadida taklif etilgan. Boshlang'ich o'tishlarda va AprioriTid algoritmgga o'tishda birlamchi

to'plamning kodlashtirilgan jamlanmasi o'tishning oxirida xotira imkoniyatiga muvofiq lashtirilishi kutilgan paytda AprioriHybrid algoritmi Aprioridan foydalanadi. Biroq, Aprioridan AprioriTidga o'tish qo'shimcha resurslarni talab qiladi.

DHP algoritmi xesh algoritmi deb ham ataladi. Uning ishlashi Apriori algoritmining har bir bosqichida nomzodlarni hisoblovchi sonlarni qisqartirish uchun amalga oshiriluvchi jamlanma-nomzodlarni ehtimolli hisoblashga asoslangan. Qisqartirish k - elementli jamlanmalar-nomzodlardan har biri kamaytirish bosqichidan tashqari xesh bosqichidan o'tishi bilan erishiladi. Algoritm $k-1$ bosqichda nomzodni tanlash vaqtida xesh-jadvali yaratiladi. Xesh-jadvalning har bir yozuvi xesh-jadvaldagi ushbu yozuvga muvofiq k - elementli jamlanmalarni barcha qo'llab-quvvatlashlarini hisoblagichi bo'lib hisoblanadi. Algoritm k - elementli jamlanma-nomzodlar to'plamini qisqartirish uchun k - bosqichda ushbu axborotdan foydalanadi. Qism to'plamni qisqartirgandan so'ng Apriorida sodir bo'lgani kabi, foydalanish uchun o'rnatilgan bo'sag'aviy qiymatlar kichik bo'ladi.

PARTITION algoritmi. Ushbu bo'linish algoritmi (qismlarga ajratish) har biri operativ xotiraga sig'ishi mumkin bo'lgan kesishmaydigan bo'limlarga ma'lumotlar bazasini ajratish orqali tranzaksion ma'lumotlar bazasini skanerlashdan iborat. Birinchi bosqichda Apriori algoritmi yordamida har bir bo'limlardan "mahalliy" tez-tez duch keluvchi ma'lumotlar jamlanmasi aniqlanadi. Ikkinchisida barcha ma'lumotlar bazasi to'g'risida bunday har bir jamlanmani qo'llab-quvvatlash hisoblanadi. Shunday qilib, ikkinchi bosqichda barcha potensial uchrashuvchi ma'lumotlar jamlanmalarini to'plami aniqlanadi.

DIC algoritmi (Dynamic Itemset Counting). Algoritm ma'lumotlar bazasini bir nechta bloklarga ajratadi, ularning har biri "boshlang'ich nuqtalar" (start point) deb nomlanadi va so'ngra ma'lumotlar bazasini davriy skaner qiladi.

26. EVOLYUSION ALGORITMLAR

Tabiiy evolyusion jarayonlarni modellashtirish qonuniyatlarini asosidagi genetik algoritmlar maqsad funksiya o'zgarishiga mos topilgan yechimni o'zgarishini talab qiladigan masalalarni yechish imkonini beradi.

Statsionar funksiyalar optimizatsiyasida qo'llaniladigan standart genetik algoritmdan berilgan sinf masalalarini yechishda to'g'ridan-to'g'ri foydalanib bo'lmaydi. Chunki genetik algoritmning boshlang'ich avlod yaratish jarayonida populyasiyaning xilma-xilligi tezda yo'qoladi. Bu xilma-xillik maqsad funksiyani o'zgarishi asosida algoritmning moslashuvchanligi uchun zarur. Mazkur muammoni hal etish uchun populyasiya xilma-xilligini saqlovchi maxsus usullar qo'llaniladi. Bu usullar uchta asosiy sinfga ajratiladi: populyasiyada xilma-xillikni saqlovchi algoritmlar, moslashuv tartibidan foydalanuvchi algoritmlar, qo'shimcha "genetik ma'lumotlar" ga ega bo'lgan algoritmlar.

Evolyusion qidiruv jarayoniga ega populyasiyada xilma-xillikni saqlovchi algoritmlarni quyidagilarga ajratish mumkin:

- "Tasodifiy migrant" (randomim migrants) usuliga asoslangan algoritmlar. Mazkur usulda har bir populyasiyaning qismi tasodifiy generatsiya orqali hosil qilingan yangi organizmlar bilan to'ldiriladi.

- Termodinamik xususiyatli algoritmlar. "Erkin energiya" deb ataluvchi boshqarish yo'li bilan populyasiyalar xilma-xilligini boshqarishni ta'minlash g'oyasi "termodinamik" genetik algoritmning asosini tashkil etadi.

- Organizmlarni "yosh" ni inobatga oluvchi algoritmlar. Organizm "yosh" ini inobatga olish orqali sifat funksiyasini takomillashtirish mazkur algoritmni asosiy g'oyasi hisoblanadi. Bunda "o'rta yoshli" li organizmlar muhim o'rin tutadi.

- Adaptatsiya protsedurasidan foydalanuvchi algoritmlar. Bu algoritmlar statistik yoki evristik qonun asosida o'z parametrlarini o'zgartiradi.

- "Adaptiv mutatsiya" usuli (adaptive hyper mutation). Bu usul algoritmning ishlash sifati yomonlashishi kuzatilganda, mutatsiya darajasini keskin oshirish imkonini beradi.

- "O'zgaruvchan lokal qidiruv" usuli (variable local search). Mazkur usul sifat funksiyasining o'rta qiyamati kamayishi kuzatilganda mutatsiya darajasini oshiradi.

- Chatishtirish ehtimolligi va mutatsiya darajasini moslashtiruvchi usullar. Moslashish algoritmning ishlash jarayonida uzluksiz amalga oshiriladi.

- "xotira" ning (qo'shimcha "genlar" yoki "xromosomal") qo'shimcha shaklidan foydalanuvchi usullar. Ushbu usullar oldindan aniqlangan eng yaxshi yechimlarni kelgusida foydalanish uchun saqlab qoladi.

- "tuzilmali genetik algoritm". Mazkur algoritm organizmlarning ko'p pog'onali tuzilmasidan foydalanishga asoslangan bo'lib, bunda birinchi pog'onali genlar quyi pog'ona genlarini faollashtiradi.

Fiksirlangan o'zgaruvchi muhit sharoitida yangi populyasiyani tashkil etish uchun eng yaxshi organizmlarni saqlash usuli taklif etilgan. O'zgarish sanoqli avlodlar almashishida yuzaga keladi. Bunda populyasiyaning eng yaxshi organizmlari berilgan vaqt oraliqlarida saqlanib boradi. Maqsad funksiya o'zgarganda yangi populyasiyani oldingi bosqichlarda saqlangan yechimlar bilan qisman (5-10%), qolgan organizmlar esa tasodifiy to'ldiriladi. Ushbu yondoshuvdan foydalanganda algoritmning sifati tasodifiy initsializatsiya algoritmidan yuqori ekanligi ko'rsatilgan. Biroq saqlangan yechimlarning katta qismini yangi populyasiyaga o'tkazish (50-100%) yoki funksiyaning keskin o'zgarishi algoritmning ishlash sifati pasaytirib yuboradi.

Populyasiyaning eng yaxshi organizmlarini saqlovchi "bilimlar ombori" deb ataluvchi maxsus yondashuv taklif etilgan bo'lib, bunda atrof-muhitning turli shartlari inobatga olingan va klassifikatsiyalashgan bo'lishi mumkin. Teng vaqt oralig'idagi eng yaxshi organizmlar omborda saqlanadi va muhit holatiga ko'ra tartiblanadi. Muhitda o'zgarish yuzaga kelsa, algoritm qaytadan ishga tushiriladi va bunda populyasiyaning bir qismi bilimlar omboridagi organizmlardan tashkil topadi. O'tkazilgan tadqiqotlar bilimlar omboridagi ma'lumotlardan samarali foydalanish tajribasini talab qilishini ko'rsatdi. Biroq bunday yondashuvdan turli tipli atrof-muhit holatlari klassifikatsiyalashganda foydalanish mumkin.

Uchinchi sinf usullari diskret va davriy o'zgaruvchan muhitlar uchun ancha samarali bo'lsada, biroq ularning ko'pchiligini uzluksiz o'zgaruvchan optimumli masalalarga qo'llab bo'lmaydi. Yuqorida keltirilgan barcha yondoshuvlar evristik bo'lib, ularning samaradorligi maqsad funksiyasi turli tipda bo'lgan masalalarda turlicha bo'lishi mumkin.

Ayni paytda adabiyotlarda statsionar bo'lmagan funksiyalar optimizatsiyasida genetik algoritmlarning turli variantlarini taqqoslashga kam

e'tibor berilmoqda. Shuning uchun mazkur dissertatsiyada bir qator tez-tez foydalaniladigan usullar taqqoslangan. Taqqoslash dinamik tizimni optimal boshqarishni qidirish masalasini yechish misolida o'tkazilgan bo'lib, maqsad funksiyaning turli tipli o'zgarishlariga yaxshi moslashish imkonini beradigan yuqorida ko'rib chiqilgan usullarning modifikatsiyalari taklif etilgan.

26.1. Evolyusion algoritmlarda genetik algoritmlarni tutgan o'rni.

Genetik algoritmlar 50- yillarning boshlarida bir necha biologlar biologik tizimlarni simulyatsiyasi uchun kompyuterdan foydalanishganda paydo bo'la boshlangan[1,4]. Biroq genetik algoritmlar hozirgi darajadagidek mashhurlikka 1960- yillar oxiri va 1970- yillar boshlarida Michigan Universitetidan Jon Xolland boshchiligidagi o'tkazilgan ilmiy ishlar natijasida erishdi.

Genetik algoritmlar potentsial yechimlar populyatsiyasini ko'rib chiqish asosida ko'p yo'nalishli qidiruvni amalga oshiradi va u ushbu yo'nalishlar o'rtasida axborot almashinuvi va shakllanishini ta'minlaydi. Populyatsiyalar simulyatsiyalashgan evolyusiyaga uchraydi: har bir generatsiyada nisbatan "yaxshi" yechimlar reproduksiya uchun saqlanib qoladi, "yomon" lari esa chiqarib yuboriladi. Turli yechimlarni farqlash uchun muhit vazifasini bajaruvchi maqsad (baholash) funksiyasidan foydalaniladi.

Genetik algoritmlarning ishlash mexanizmi juda sodda bo'lib, amallarning soddaligi va samaraning yuqoriligi uning asosiy yutuqlaridan biri hisoblanadi.

Genetik algoritmlarning asosiy xossalari ko'rib chiqaylik. Boshida individlar populyatsiyasi (xromosomalar deb ataluvchi bitlar satri) olinadi. Har bir individga qidiruv fazosidagi bitti nuqta mos qo'yiladi. Populyatsiya o'lchami masala o'lchamiga bog'liq bo'lib, u individlar soniga teng bo'ladi va uni tadqiqotchi beradi.

Har bir individ 0 va 1 lar yordamida kodlangan satrdir. Masalan, 0 dan 31 gacha bo'lgan butun sonlarni ikkilikda quyidagicha kodlashimiz mumkin.

(00000)

(00001)

(00010)

(11111)

Bunda har bir satr 5 ta belgidan iborat, ya'ni 5 bet. Ularni bitlar, satr esa xromosoma deb ataladi.

O'zgaruvchilarni kodlash turli usullarda amalga oshirilishi mumkin. Bundan tashqari o'zgaruvchilarni ikkilik bo'lgan masalalar ham uchrab turadi,

ularda kodlash shart emas. Shunday qilib, tasodifiy ravishda n ta individ olamiz va boshlang'ich populyatsiyani tuzamiz.

$$x_1 = (01100)$$

$$x_2 = (11001).$$

$$x_n = (00011)$$

Bu nuqtalarda funksiyaning qiymatlarini hisoblaymiz.

$$f(x_1), f(x_2), \dots, f(x_n).$$

$f(x)$ – tadqiq etiladigan funksiya. Bu optimallashtiriladigan funksiya ham bo'lishi mumkin.

Quyidagi hisoblaymiz:

$$F = \sum_{i=1}^n f(x_i).$$

$f(x_i)$ – (x_i) individ sifati (yaroqliligi) ning bahosi.

Barcha individlar va populyatsiyaning yaroqliligi aniqlangach, har bir individning yashab ketish ehtimolligi quyidagicha hisoblanadi.

$$P_s^1 = f(x_1) / F,$$

$$P_s^n = f(x_n) / F.$$

Har bir individning kumulyativ ehtimolliklari quyidagicha hisoblanadi.

$$P_{cum}^1 = P_s^1,$$

$$P_{cum}^2 = P_s^1 + P_s^2$$

$$P_{cum}^n = \sum_{i=1}^n P_s^i,$$

$$0 < P_{cum}^1 < P_{cum}^2 < \dots < P_{cum}^n = 1.$$

Bu seleksiyani amalga oshirish uchun bajariladi. Seleksiya sun'iy ravishda tabiiy genetik seleksiya kabi amalga oshiriladi. Seleksiya natijasi yaroqliligi yuqori bo'lgan individlar yashash va ko'payish imkoniyati ko'proq, yaroqliligi past bo'lgan individlar yashash va ko'payish imkoniyati kamroq bo'lishidan iborat.

Seleksiyani amalga oshirish uchun $[0,1]$ oraliqdan r_i tasodifiy son olinadi. Agar $P_{cum}^k < r_i < P_{cum}^{k+1}$, bo'lsa, bu x_{k+1} xromosomani navbatdagi nasl uchun tanlanishini bildiradi. Bu jarayon r tasodifiy son bilan n marta takrorlanadi va n ta xromosoma navbatdagi nasl uchun tanlanadi. Shuning uchun yaroqliligi

yuqori bo'lgan individlar yangi naslda bir nechta nusxalariga ega bo'lishi mumkin. Yaroqliligi kamlari esa yangi naslda umuman bo'lmasligi mumkin.

Genetik algoritmnning ikkinchi asosiy operatori chatishtirish hisoblanadi. Chatishtirish uchun populyasiyaning bir necha xromosomalari tanlanadi. Ularning soni P_c dastlabki chatishtirish ehtimolligiga bog'liq bo'ladi. Agar $P_c = 0.25$ bo'lsa, populyasiya o'lchami 16 ga teng va har bir nasl olishda 4 tadan xromosoma qatnashadi. Agar x_m va x_k chatishtirish uchun tasodifiy tanlangan va xromosoma uzunligi i ga teng bo'lsa, u holda qabul qiluvchi tasodifiy j son $1 \leq j < l$ tengsizlikni qanoatlantiradi.

Faraz qilaylik, $x_m = (1101101)$, $x_k = (1110010)$ va $j=3$ bo'lsin.

Bu x_m va x_k avlodlar chatishtirish natijasida quyidagi ko'rinishga ega bo'lishini bildiradi:

$$x'_m = (1100010),$$

$$x'_k = (1111101).$$

j ning qiymati chatishtirish nuqtasi deb ataladi. Ushbu nuqtadan xromosomalari kesiladi va avlodlar olish uchun ularning o'rinlari almashtiriladi. Mazkur amalni qolgan tanlangan juftliklar uchun ham qo'llab, chatishtirish jarayonini yakunlaymiz va olingan yangi avlodlarni ularning otasonasi bilan populyasiyaga kiritamiz.

Genetik algoritmlardagi uchinchi asosiy amal – bu mutatsiyadir. Tabiiy mutatsiyaga o'xshash bo'lgan sun'iy mutatsiyada xromosomadagi bitta bit o'zgaradi. Ya'ni, agar $x_k = (1110010)$ xromosoma mutatsiya uchun ajratib olingan bo'lsa, u holda yana j (j ning qiymati qatordagi bitlar sonidan oshib ketmasligi kerak) tasodifiy soni beriladi va 0 dan 1 ga o'zgaruvchi (yoki aksincha) genning tartib raqati aniqlanadi. Faraz qilaylik, $j=1$. Bu x_k xromosomaning birinchi geni 1 dan 0 ga o'zgarishini bildiradi, ya'ni $x'_k = (0110010)$, qolgan genlar esa o'zgarimasdan qoladi. Olingan yangi xromosoma populyasiyaga qo'shiladi. Keyin seleksiya o'tkaziladi, so'ngra boshlang'ich o'lchamli populyasiya tuziladi.

Bu amallarni ketma-ket o'tkazib va yangi generatsiyalarni yaratib, har bir generatsiyada populyasiyaning o'rtacha yaroqliligi yaxshilanganini kuzatish mumkin. Bir necha generatsiyalardan so'ng, populyasiyaning yaroqliligining yaxshilanishi juda kichikligi ko'rinadi, shunda oxirgi generatsiyaning eng

yaxshi organizmini masalaning yechimi deb tanlab, jarayonni to'xtatish mumkin.

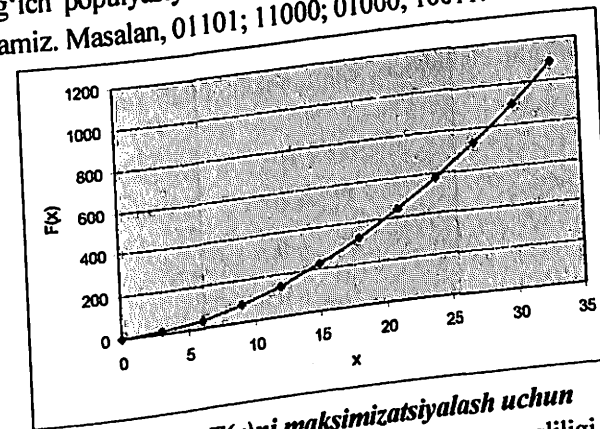
Odatda generatsiyalar soni avvaldan beriladi. Shuningdek P_m mutatsiyaning ehtimolligi ham oldindan beriladi. Bu qiymat mutatsiyaga uchraydigan xromosomalari sonini aniqlaydi. Ya'ni, agar populyasiyaning o'lchami 20 va $P_m = 0.05$ ga teng bo'lsa, bu mutatsiyada 1 ta xromosoma ($20 \cdot 0.05 = 1$) qatnashishini bildiradi.

26.2. Optimizatsiya masalasini genetik algoritm yordamida yechish

Mazkur bo'limda genetik algoritmnning asosiy xususiyati ko'rib chiqiladi. Sodaroq bo'lishi uchun dastlab genetik algoritm yordamida bir o'zgaruvchili funksiya optimizatsiyasini ko'rib chiqaylik. Masalan, $F(x) = x^2$ funksiyaning $x \in [0, 31]$ dagi maksimizatsiyani (2.2-rasm).

Genetik algoritmdan foydalanishdan oldin, birinchi navbatda o'zgaruvchini bitli satr ko'rinishida kodlab olishimiz zarur. 0 va 31 orasidagi ixtiyoriy butun sonni 5 xonali ikkilik son bilan ifodalash mumkin ($(00000)_2 = 0$ va $(11111)_2 = 31$). Bunda xromosoma uzunligi 5 ga teng bo'ladi.

Boshlang'ich populyasiya sifatida 5 ta genli ixtiyoriy 4 ta xromosoma yordamida olamiz. Masalan, 01101; 11000; 01000; 10011.



26.1-rasm. $F(x)$ ni maksimizatsiyalash uchun

Navbatdagi qadamda olingan xromosomalarning yaroqliligi funksiyaga o'zgaruvchining haqiqiy qiymatini qo'yish orqali tekshiriladi. Masalan, $x = 01000_2$ da $F(01000)_2 = F(8) = 64$. Shu tariqa barcha individlarni yaroqliligi tekshiriladi va quyidagi formula yordamida har bir individning yashab ketish ehtimolliklari hisoblanadi.

$$P_s^i = F_i / \sum_{j=1}^4 F_j, \quad i = \overline{1,4},$$

Quyidagi formula yordamida esa kumulyativ ehtimolliklari hisoblanadi.

$$P_{sum}^i = \sum_{j=1}^i F_s^j, \quad i = \overline{1,4}.$$

Olingan barcha ma'lumotlarni jadval ko'rinishida yozib olamiz.

26.1-jadval.

Boshlang'ich populyasiya	Butun qiymat	$F(x) = x^2$	P_s	P_{cum}	Seleksiyadan keyingi son
01101	13	169	0.14	0.14	1
11000	24	576	0.49	0.63	2
01000	8	64	0.06	0.69	0
10011	19	361	0.31	1	4
O'rtacha		293	0.25		1
Maksimal		576	0.49		2
Jami		1170	1		4

Seleksiyani amalga oshirish uchun [0,1] oraliqdagi ixtiyoriy 4 ta son olamiz. Faraz qilaylik biz 0.1; 0.25; 0.5 va 0.8 sonlarini tanlagan bo'laylik. U holda bu sonlarni kumulyativ ehtimolliklar bilan solishtirib quyidagiga ega bo'lamiz:

$$0.1 < P_{cum}^1,$$

$$P_{cum}^1 < 0.25 < P_{cum}^2,$$

$$P_{cum}^1 < 0.5 < P_{cum}^2,$$

$$P_{cum}^3 < 0.8 < P_{cum}^4.$$

Tengsizliklarning o'ng tomoniga qarab, 1-chi va 4-chi xromosomalar yangi generatsiyada birinchi o'rinni olib seleksiyadan o'tganligini ko'rish mumkin.

Navbatdagi qadamda chatishtirish amali bajariladi. Agar chatishtirish ehtimolligi ($p_c = 1$) 1 ga teng bo'lsa, chatishtirish jarayonida 4 ($4 \cdot 1 = 4$) ta xromosomaning ishtirok etishini bildiradi. Bu xromosomalarni ixtiyoriy tanlab olamiz. 1-chi va 2-chi satrlarda chatishtirish nuqtasi 4, 3-chi va 4-chi satrlarda esa chatishtirish nuqtasini 2 deb olaylik. Olingan natijalar 2.2-jadvalda keltirilgan.

26.2-jadval.

Chatishtirishdan keyingi populyasiya	Chatishtirish nuqtasi	Yangi populyasiya	X ning qiymati	$F(x) = x^2$
01101	4	01100	12	144
11000	4	11001	25	625
11000	2	11011	27	729
10011	2	10000	16	256
O'rtacha				439
Maksimal				729
Jami				1754

26.1 va 26.2- jadvallarni solishtirish orqali populyasiyaning yaroqliligi oshganligini ko'rishimiz mumkin, ya'ni biz yechimga yaqinlashdik.

Oxirgi amal mutatsiya deb ataladi va bu bitlarda amalga oshiriladi. Agar $p_m = 0.05$ bo'lsa, bu 20 bitli populyasidan faqat 1 ($20 \cdot 0.05 = 1$) tasi o'zgarishini bildiradi. Faraz qilaylik 4- satrning 3- biti mutatsiyani talab etsin, ya'ni $x_3 = 10100$. Yuqoridagi amallarni bir necha marta takrorlash orqali masalani optimal yechimni beruvchi xromosomaga (11111) ega bo'lamiz.

Shuni alohida ta'kidlash joizki, genetik algoritmdan ko'p ekstremal masalalarni global yechimni topishda foydalanish yaxshi samara beradi. Masalan, agar funksiya 26.1- rasmda keltirilgan ko'rinishga ega bo'lsa uni global maksimumkini odatiy usullar yordamida topish yetarli darajada murakkab bo'ladi.

27. ZAMONAVIY EVOLYUTSION ALGORITMLAR

27.1 Global optimallashtirish masalasini yechish uchun zarralar galasi algoritmi

Global optimallashtirish masalasini yechish uchun zarralar galasi protsedurasi to'rtta asosiy bosqichlardan tashkil topadi: initsializatsiya, jarlikni qurish, tezlik va holatni yangilash hamda qidiruvning lokal protsedurasi. Uchinchi bosqichda muvozanat koeffitsienti mexanizmi tavsiflangan.

1- bosqich. Initsializatsiya. Birinchi qadam dastlabki populyasiyani shakllantirish uchun tasodifiy zarralarni generatsiya qilish NP (populyasiya o'lchami) hisoblanadi. Barcha zarralar dastlab qidiruv fazosi atrofida tekis taqsimlangan bo'ladi. Birinchi qadam ikki qismdan tashkil topgan, ya'ni zarralar pozitsiyasini initsializatsiya qilish va ularning tezliklarini initsializatsiya qilish. Holatlar va tezliklar vektor ko'rinishida ifodalanadi.

Xususan, birinchi qadam quyidagi ifodaga mos ravishda bajariladi:

$$X_i^d = \text{random}_i^d(\text{Rang}L_i^d, \text{Rang}R_i^d),$$

$$V_{\max}^d = \text{Rang}R^d - \text{Rang}L^d,$$

$$V_i^d = \text{random}_i^d(((-0,5) * V_{\max}^d, 0,5 * V_{\max}^d),$$

bu yerda $i \in [1, NP]$ – zarra indeksi, $d \in [1, D]$ o'lchami. X_i^d – zarraning d -elementini anglatadi, V_i^d – zarraning d -elementi tezligini anglatadi. $\text{random}_i^d(a, b)$ $[a, b]$ oraliqda tasodifiy sonni generatsiya qiladi. Agar tezlik juda yuqori bo'lsa, zarra qidiruv sohasidan uchib chiqib ketishi mumkin. V_{\max}^d parametr i – zarraning d -elementi tezligini chegaralarini aniqlash uchun foydalaniladi. $\text{Rang}R_i^d$ va $\text{Rang}L_i^d$ qidiruv fazosi o'lchami d ning quyi va yuqori chegaralarini ifodalaydi.

2- bosqich. Jarlikni qurish.

Ko'p ekstremal masalani hal qilishda jarliklar bir nechta optimumlarni aniqlash uchun ishlatiladi. Har bir turdagi eng yaxshi zarra avlod deb ataladi. Turlar avlodlar atrofida yaratiladi va optimumlar atrofida yaqinlashish uchun javobgar hisoblanadi. Turlar har bir avlodda yangilanadi.

Har bir avlodda zarralar maqsad funksiyasining kamayish tartibida saralanadi. Keyin barcha zarralar naslga mos yoki yo'qligini aniqlash uchun shu tartibda tekshiriladi. Ushbu protsedurada ishlatiladigan muhim

parametrlardan biri r turning radiusi bo'lib, u jarliklarni belgilaydi. Eng yaxshi zarra to'g'ridan-to'g'ri nasl sifatida tanlanadi va S avlodlar majmuiga kiritiladi. S har bir avlodda bo'sh deb belgilanadi. Keyinchalik barcha avlodlar S ga qo'shiladi. Boshqa zarralar r turning radiusi yordamida tekshiriladi. Agar zarra va avlod orasidagi masofa r dan kichik (yoki unga teng) bo'lsa, u holda zarra avlod bilan bir xil turdagi a'zo hisoblanadi. Agar S da zarra va nasl orasidagi masofa r dan katta bo'lsa, u holda zarra yangi avlod deb hisoblanadi va S ga joylashtiriladi. Barcha zarralarni saralashdan so'ng, har bir zarra turga tegishli bo'ladi.

3- bosqich. Tezlik va holatni yangilash.

Har bir zarra har bir iteratsiya davomida $pbest$, $vbest$, zarralari tomonidan boshqariladi. i zarra tomonidan aniqlangan eng yaxshi vaziyatni shaxsiy eng yaxshi $pbest$ va uning turi avlodining eng yaxshi vaziyatini qo'shni eng yaxshi $vbest$ deb ataymiz. Ushbu ikki turdagi yetakchi zarralar galasi usulining tezlik tenglamasida qatnashadilar. Tenglama PSO tenglamasiga o'xshaydi va yagona farq $gbest$ $vbest$ bilan almashtiriladi.

Qidiruv jarayonida har xil jarliklardagi zarralar soni asta-sekin tekshiriladi. Farqni kamaytirish uchun muvozanat omili (EF) mexanizmi taklif qilinadi, unda katta jarlikning eng yomon zarralari turdan chiqarib yuboriladi va boshqa jarliklarga harakatlanadi.

Tezlik vektori (BC) bir nechta zarralarning tezlik tenglamasiga shunday qo'shiladiki, bu holda ular kamroq zarralari bo'lgan jarlikga qarab harakat qilishadi.

EF ni amalga oshirishda, tur hajmi bo'yicha kamayish tartibida saralanadi. Ularning eng katta va eng kichigi hisoblash uchun tanlanadi

$$DV^d = \text{seed}S^d - \text{seed}L^d,$$

bu yerda $\text{seed}S^d$ – eng katta tur avlodi, $\text{seed}L^d$ eng kichik turga tegishli avlod.

Shu bilan birga, eng katta turlarga tegishli bo'lgan barcha individumlar yaroqlilik qiymatlariga qarab saralanadi. Xatolik o'lchami sifatida aniqlangan parametri EF ni qabul qiladigan zarralar sonini aniqlash uchun foydalaniladi. DS quyidagi formula bo'yicha hisoblanadi:

$$DS = (\text{size}L + \text{size}S) / 2,$$

bu yerda $\text{size}L$ va $\text{size}S$ – mos ravishda eng katta va eng kichik turdagi

zarralar soni.

Muvozanat koeffitsienti mexanizmi har bir iteratsiya paytida DV ni aniqlaydi va bir vaqtning o'zida vektorning qaysi zarralari ta'sir qilishini aniqlaydi. Keyinchalik DV zarralar tezligini yangilash uchun ishlatiladi.

Tezlik vektori hisoblab chiqqandan so'ng, zarralarning tezligi va holatlari quyidagilarga muvofiq yangilanadi:

$$V_i^d(t+1) = w * V_i^d(t) + c_1 * random1_i^d * (pbest_i^d(t) - X_i^d(t)) + c_2 * random2_i^d * (nbest^d(t) - X_i^d(t)) + DV, \quad (27.1)$$

$$X_i^d(t+1) = X_i^d(t) + V_i^d(t+1), \quad (27.2)$$

bu yerda (27.1) ifoda muvozanat koeffitsienti (EF) yordamida aniqlanadigan zarralar uchun. Barcha zarralarning holati (27.2) ifoda bo'yicha yangilanadi.

4- bosqich. Lokal qidiruv.

To'rtinchi bosqichda barcha $pbest$ lokal qidirish protsedurasi bilan yangilanadi. Har bir $pbest$ o'z atrofida yangi zarra hosil qiladi va eng yaxshi zarra yangi $pbest$ bo'lish uchun boshqasini almashtiradi. Aniqroq aytganda, populyasiyadagi har bir $pbest$ uchun, operator $pbest$ va unga eng yaqin bo'lgan boshqa $pbest$ o'rtasidagi farqlarga muvofiq sozlanadi. Lokal qidiruv usuli, asl algoritmi sozlashni yaxshilaydi, bu ko'plab murakkab masalalarda samarali ekanligini isbotladi.

Amaliyotda turli amaliy masalalarni hal qilish uchun intellektual tizimlarni ishlab chiqishda gibriddan foydalanish tendensiyasi mavjud. Noravshan modellarni o'qitishda ba'zi mualliflar an'anaviy optimallashtirish usullari bilan bir qatorda immun algoritmidan ham foydalanishadi. O'tkazilgan tajribaviy tadqiqotlar shuni ko'rsatadiki, ushbu yondashuvdan foydalanish masalani yechishda tezlik va aniqlikni oshirishda yaxshi natijalar beradi. Shunday qilib, intellektual usullar va samarali evristik algoritmlar asosidagi noravshan modelni o'qitishning samarali algoritmini ishlab chiqish ekspert tizimlarining bilimlar bazasini shakllantirish muammosini hal qilishda dolzarb masala hisoblanadi.

27.2. Optimallashtirish masalasini yechish uchun noravshan madaniy algoritmi

Madaniy algoritmi (Cultural Algorithms, SA) – masalani yechish davomida to'plangan tajribani hisobga olgan holda optimallashtirishning populyasion algoritmlarining takomillashgan algoritmidir. Ushbu algoritmning asoschisi Reynolds (R.Reynolds) hisoblanadi.

Jamiyatni rivojlanishi jarayonida odamlar atrof-muhit haqidagi ma'lumotlarni to'playdilar. Ushbu ma'lumotlarni jamiyatning bilimlar ombori deb hisoblash mumkin, undan a'zolar o'z ko'rinishlarini optimallashtirish maqsadida foydalanishlari mumkin. Madaniy algoritmda ushbu bilimlar omborini ishonchlilik fazosi (belief space) ko'rinishida shakllantiriladi hamda populyasiyaga yana bir evolyutsion ta'sir elementi sifatida qo'llaniladi.

Optimallashtirishning ixtiyoriy masalasini umumiylikni yo'qotmagan holda quyidagi ko'rinishdagi chegaralangan global shartli minimallashtirish masalasi kabi ifodalanish mumkin

$$\min_{X \in D \cap P} f(X) = f(X^*) = f^*, \\ D = \{X \mid f(X) \geq 0, P = X \mid pl, \leq x, \leq pu, j \in [1:N]\}.$$

Faraz qilaylik, $D \cap P$ to'plam qabariq hamda ushbu to'plamda $f(X)$ fitness-funksiya nomanfiy qiymatlarni qabul qiladi, bunda maqsad funksiya optimallashtirilishi kerak. Optimallashtirish masalasining murakkabligi $f(X)$ ning matematik holatiga bog'liq bo'lib, optimallashtirishning an'anaviy usullarini qo'llash bunday masala uchun maqsadga muvofiq emas.

Ushbu mulohazalarni inobatga olgan holda bunday masalalarni yechish uchun evolyusion algoritmlar nomi bilan mashhur tabiatga asoslangan algoritmlar ishlab chiqilgan. Evolyusion algoritmning vazifasi tabiiy jarayonning jarayonlarini o'xshatish, ya'ni tiklash va tanlash, qidiruvning avvalgi bosqichlardagi nisbatan yaxshi natijalarni qo'llagan holda maqbul yechimlarni aniqlash. Buning uchun ko'p yillar davomida ko'plab evolyusion algoritmlar ishlab chiqilgan, bularga genetik algoritmi (GA), zarralar galasi optimizatsiyasi (PSO), arilar koloniyasi algoritmi (ABC), differensial evolyusiya (DE) va madaniy algoritmi (CA) misol bo'ladi.

Har bir evolyusion algoritmi uchun populyasion fazo mavjud bo'lib, qidiruv boshida populyasiya to'plami bilan initsializatsiya qilinadi, so'ngra individlar tiklash amalini qo'llagan holda evolyutsiya qilinadi. Har bir

evolyutsion algoritmlar turlari amallariga ega bo'lib, ularni mutatsion va crossover amallar va eng yaxshi populyatsiyalarning yangi avlodini yaratish maqsadi deb atash mumkin. Yuqorida keltirilgan algoritmlarni tahlil qilib, shunday xulosa qilish mumkinki, madaniy algoritmlar qidiruvni yaqqol boshqarish uchun ishonchlilik fazosi deb ataluvchi bilimlarning turli manbalaridan foydalanuvchi yagona evolyutsion algoritmlar hisoblanadi. Bu fazo evolyutsion algoritmlar asosida optimal qidiruv uchun populyasiyaga yangi maqbul yechimlarni yaratishga yordam berishi uchun avlodlar davomida yig'ilgan bilimlardan foydalanadi. Populyasiya fazosi hamda ishonchlilik fazosi o'rtasidagi o'zaro ta'sirni muvofiqlashtirish uchun bu fazolarni bog'lovchi algoritmlar zarur. Evolyutsion algoritmlar ko'pincha optimallashtirishning ko'p ekstremalli masalasini yechishda nokompetentlikni namoyish qiladi, ayniqsa o'lcham katta bo'lganda. Bu esa stagnatsiya hamda muddatdan avval yaqinlashishni keltirib chiqaradi. Stagnatsiya yaxshilanish bo'lmagan holda sodir bo'ladi, va muddatdan avval yaqinlashish paydo bo'lganda, populyasiya bitta yechimga yaqinlashsa, bu yechimning sifati kutilgandek bo'lmasa, yaxshi yechimni topish ketma-ket avlodlar uchun takrorlanadi.

Deyarli barcha evolyutsion algoritmlarda mavjud bo'lgan yuqorida keltirilgan muammolarni yechishda madaniy algoritmlar samaradorligini oshirish uchun uchta asosiy konsepsiyadan foydalaniladi. Birinchisi – gibridlash usullarini qo'llash, bunda madaniy evolyutsion algoritmlar boshqa usul yoki lokal qidiruv bilan gibridlanadi. Qidiruvning lokal algoritmlari populyasiyaga asoslangan stoxastik strategiyalardir. Ularning asosiy maqsadi – qidiruv sohasida mavjud bilimlarni qo'llashdir. Ikkinchi konsepsiya – yana bir evolyutsion algoritmlar madaniy algoritmlarning populyasiya fazosi sifatida qo'llashdan iborat. Ushbu maqsadda yumshoq hisoblashlarning ko'plab algoritmlaridan foydalaniladi.

Madaniy algoritmlar ishonchlilik fazosi komponentasida ifodalangan madaniy bilimlardan foydalanadi. Ushbu komponent madaniy bilimlarni maxsus ko'rinishlari bo'lgan bilimlarning turli manbalaridan tashkil topgan.

Algoritmlarning umumiy ko'rinishi

Madaniy algoritmlarning asosiy komponentalari populyasiya fazosi va ishonchlilik fazosi hisoblanadi.

Populyasiya fazosi va ishonchlilik fazosi bevosita qaror funksiyasi

(acceptance function) va ta'sir funksiyasi (influence function) bilan o'zaro ta'sirlashadi. Ko'rsatilgan birinchi funksiya yordamida ishonchlilik fazosini sozlash imkoniga ega bo'lgan eng yaxshi individlar to'plamini aniqlaydi. Ikkinchi funksiya ishonchlilikka populyasiya individlari evolyusiyasiga ta'sir etish imkoniyatini beruvchi qoidalarni beradi.

Madaniy algoritmlarning psevdokodi quyida keltirilgan[76].

Begin

$g = 0$;

Populyasiyani initsializatsiya qilish $S(g)$;

Ishonchlilik fazosini initsializatsiya qilish $B(g)$;

Populyasiyani fitnes-funksiya yordamida baholash $S(g)$;

repeat

Populyasiya fazosi va ishonchlilik fazosi o'rtasida bevosita qaror funksiyasi orqali o'zaro ta'sirlashishi ($S(g), B(g)$);

Ishonchlilik fazosini sozlash $B(g)$;

Populyasiya fazosi va ishonchlilik fazosi o'rtasida bevosita ta'sir funksiyasi orqali o'zaro ta'sirlashishi ($B(g), S(g)$);

$g = g + 1$;

populyasiyaning joriy iteratsiyada evolyusiyasi $S(g)$;

Until (termination condition)

End.

Ishonchlilik fazosini initsializatsiya qilish va sozlash.

ADABIYOTLAR RO'YXATI

1. https://www.texnoman.uz/post/big-dataning-asosiy-8-atamasi_.html
2. <https://www.colocationamerica.com/blog/big-data-and-education>
3. Ifeanyi.G.Ndukwe, Ben.K.Daniel and Russell.J.Butson. Data Science Approach for Simulating Educational Data: Towards the Development of Teaching Outcome Model (TOM). *Big Data Cogn. Comput.* 2018, 2(3), 24; <https://doi.org/10.3390/bdcc2030024>
4. <https://precisioncampus.com/blog/benefits-big-data-education/>
5. Ifeanyi.G.Ndukwe, Ben.K.Daniel and Russell.J.Butson. Data Science Approach for Simulating Educational Data: Towards the Development of Teaching Outcome Model (TOM). *Big Data Cogn. Comput.* 2018, 2(3), 24; <https://doi.org/10.3390/bdcc2030024>
6. Baig, M.I., Shuib, L. & Yadegaridehkordi, E. Big data in education: a state of the art, limitations, and future research directions. *Int J Educ Technol High Educ* 17, 44 (2020). <https://doi.org/10.1186/s41239-020-00223-0>
7. Nian-Shing Chen, Chengjiu Yin, Pedro Isaias & Joseph Psocka (2020) Educational big data: extracting meaning from data for smart education, *Interactive Learning Environments*, 28:2, 142-147, DOI: 10.1080/10494820.2019.1635395
8. https://www.jstor.org/stable/26388386?seq=1#metadata_info_tab_contents
9. Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, 37, 13–49.
10. Baker, R., & Siemens, G. (2014). Learning analytics and educational data mining. In: Keith-Sawyer R (ed) *Cambridge handbook of the leaning sciences* (2nd edn). Cambridge University Press: New York, NY, pp. 253–272.
11. Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), *Learning analytics* (pp. 61–75). Springer, New York, NY.
12. Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. US Department of Education, Office of Educational Technology, 1, 1–57.
13. Lee, LK., Cheung, S.K.S. & Kwok, LF. Learning analytics: current trends

- and innovative practices. *J. Comput. Educ.* 7, 1–6 (2020). <https://doi.org/10.1007/s40692-020-00155-8>
14. Абдыкаримова А. Т., «Технология больших данных», Наука и жизнь Казахстана, №2(78), 2019 г., с.223-226.

1.
2.
3.
A
T
ht
4.
5.
A
T
2
6.
o
E
7.
E
Ir
1
8
9
n
s
1
n
(
1
l
6
1
l
t
5
1

D.T.MUXAMEDIEVA, O.N.ABDURAXMONOV, M.H.PRIMOVA

BIG DATA VA MA'LUMOTLAR TAHLILI

O'quv qo'llamma

Muharrir: G.Beraliyeva

Bosishga ruxsat etildi 10.01.2023y. Bichimi 60X84 ¹/₁₆.
Bosma tabog'i 15,5. Shartli bosma tabog'i 15,5. Adadi 17 nusxa.
Buyurtma № 115. Bahosi kelishilgan narxda.
"Fidokor-Yosh Avlod" nashriyoti. Toshkent, Salorbo'yi kochasi, 35A.
O'zbekiston Milliy universiteti bosmaxonasida bosildi.
Toshkent, Talabalar shaharchasi, O'zMU

ISBN 978-9943-9119-6-3



9 789943 911963