

**O'ZBEKISTON RESPUBLIKASI ALOQA, AXBOROTLASHTIRISH VA
TELEKOMMUNIKASIYA TEXNOLOGIYALARI DAVLAT
QO'MITASI**

**TOSHKENT AXBOROT TEXNOLOGIYALAR UNIVERSITETI
SAMARQAND FILIALI**

Qo'l yozma huquqida

UDK 004.89, 378(372.8)

UMAROV HASAN ABDULLAYEVICH

Ma'lumotlarni intellektual tahlil etishning usullari va dasturiy ta'minoti

5A111001 – «Kasb ta'limi» (Informatika va mul'timedia texnologiyalari)

Magistr
akademik darajasini olish uchun yozilgan
dissertasiya

Ilmiy rahbar:

t.f.n., dots. Rahimov N.O.

SAMARQAND – 2014

MUNDARIJA

KIRISH.....	3
I BOB. Ma'lumotlarni intellektual tahlil etish usullarining nazariy asoslari.....	10
1. Ma'lumotlarni intellektual tahlil etish tizimning maqsad va vazifalari tahlili.....	10
2. Ma'lumotlar to'plami va ularning atributlari. Ma'lumotlar bazalari. Data Mining uslublari va bosqichlari.....	15
3. Tadqiqot maqsadi va vazifaning qo'yilishi.....	25
I bob bo'yicha xulosalar.....	26
II BOB. Ma'lumotlarni intellektual tahlilida klasterizatsiya masalasi algoritmlari	27
1. Klasterizatsiya masalasini tizimli yechishning formal qo'yilishi	27
2. Klasterizatsiya algoritmlarida qo'llaniladigan masofaga asoslangan yaqinlik darajasining o'lchovi	47
3. Klasterizatsiya masalalari usul va algoritmlarining qiyosiy tahlili	48
II bob bo'yicha xulosala.....	52
III BOB. Ma'lumotlarni intellektual tahlilida klasterizatsiya algoritmlari dasturiy ta'minotining amaliy tadbiri	53
1. Klasterizatsiya usullari uchun amaliy dasturiy paketlarda ishlash.....	53
2. Delphi dasturlash tilida yaratilgan klasterizatsiya dasturiy ta'minoti	61
3. Kompyuter xonalarini yoritishga qo'yilgan asosiy talablar.....	65
III bob bo'yicha xulosalar.....	70
XULOSA.....	71
FOYDALANILGAN ADABIYOTLAR RO'YXATI.....	73
ILOVA.....	76

KIRISH

Dissertasiya mavzusining asoslanishi va uning dolzarbligi.

Respublikamizda olib borilayotgan tub iqtisodiy islohotlar ta'lim sohasida ham izchil va uzluksiz o'zgarishlarni amalga oshirishni taqozo etmoqda. Iqtisodiyotning bozor munosabatlariga o'tishi kadrlarni tayyorlash sohasida jahon andozalarga mos ravishda tub o'zgarishlar qilinishini talab etadi. Ma'lumki, firmalar, korxonalar va tashkilotlarning samarali ishlashini undagi mavjud bo'lgan malakali va bilimli xodimlar ta'minlaydi. Shu jihatdan hozirgi vaqtda malakali va ishbilarmon xodimlarni shakllantirishga erishish muhim ijtimoiy-iqtisodiy ahamiyat kasb etadi[1,3].

Axborot tizimining keng tarmoqda yoyilishi aholining ijtimoiy-iqtisodiy sohalarni rivojlantirish borasida ishlab chiqilgan Prezident va Vazirlar Mahkamasining qarorlari, farmoyishlaridan o'z vaqtida xabar topishiga va uning mag'zini anglashiga hamda amalda tatbiq etishiga keng yo'l ochib bermoqda[2].

Shunindek, O'zbekiston Respublikasi Prezidentining 2012 yil 21 martdagi «Zamonaviy axborot-kommunikasiya texnologiyalarini yanada joriy etish va rivojlantirish chora-tadbirlari to'g'risida» PQ-1730 sonli qarori, 2012 yil 1 fevraldagi «Joylarda kompyuterlashtirish va axborot-kommunikasiya texnologiyalarini yanada rivojlantirish uchun shart-sharoitlar yaratish chora-tadbirlari to'g'risida»gi qarorlari Respublikamizda axborot-kommunikasiya texnologiyalari taraqqiyotida, bu sohadagi amaliy-ilmiy va fundamental tadqiqotlarni rivojlantirishida yana bir muhim bosqich bo'lib qoladi. Prezidentimiz I.Karimov XXI asrni ma'naviy-ma'rifat asri, ilm-fan madaniyat va axborotlar asri deb ta'riflar ekan, bu bilan hozirgi davrning muhim belgisi va xususiyatini ham belgilab berganini payqab olish qiyin emas[1].

O'zbekistonda ta'lim-tarbiya sohasini isloh qilishning asosiy omillaridan biri bu ta'lim jarayoniga zamonaviy axborot va kommunikasiya texnologiyalarini joriy etish bilan bog'liq bo'lib, respublikamiz Prezidenti I.A.Karimov bu masalada "... bugun hayotimizga chuqur kirib borayotgan Internet tizimini keng joriy etish, yoshlarimizni O'zbekistonning qadimiy va boy

tarixi, ezgu qadriyatlarimiz, yuksak axloqiy fazilatlar ruhida tarbiyalashga xizmat qiladigan milliy axborot resurslarini shakllantirish va rivojlantirish, bu borada o'zbek tilining imkoniyatlaridan samarali foydalanish masalasi doimo e'tiborimiz markazida turishi lozim" ligini ta'kidlaydi[3].

Bu kabi vazifalarni bajarish kadrlar tayyorlash milliy modelini amalga oshirishning bosqichlarida belgilab berilgan bo'lib, uning kelajakdagi istiqboli Prezidentimiz tomonidan ilmiy asoslab berildi. Modelni amaliyotga tatbiq etish o'quv jarayonini texnologiyalashtirish bilan uzviy bog'liqdir. Ilmiy texnikaviy taraqqiyot ishlab chiqarishning ko'p sonli tarmoqlari bilan bir qatorda ta'lim sohasiga ham zamonaviy axborot texnologiyalarini joriy etishni taqozo etmoqda. Shu boisdan, Kadrlar tayyorlash milliy dasturida "o'quv-tarbiyaviy jarayonni ilg'or pedagogik va axborot texnologiyalari bilan ta'minlash" zarurati e'tirof etildi, uning ikkinchi va uchinchi bosqichlarida bajariladigan jiddiy vazifalardan biri sifatida belgilandi"[5]. Bu yo'nalishlar Prezidentimizning 2002 yil 30 mayda qabul qilingan "Kompyuterlashtirishni yanada rivojlantirish va axborot-kommunikasiya texnologiyalarini joriy yetish" to'g'risidagi farmonida ko'rsatilgan vazifalarni bevosita to'ldirib, ularning dolzarbligini yanada oshirmoqda.

Mamlakatimizda ta'lim sohasida ro'y berayotgan tub yangilanishlar har bir ta'lim muassasasida o'quv jarayoni metodik ta'minotini rivojlantirishni talab etadi. Ta'lim oluvchilar uchun mustaqil bilim olish imkoniyatlarini oshirish, ta'limning elektron axborot resurslarini shakllantirish va rivojlantirish uchun zarur sharoitlarni yaratish ta'lim mazmunini takomillashtirishning zaruriy shartlaridan biri sanaladi.

Hozirgi kunda axborot-kommunikasiya texnologiyalari insonlarning intellektual faoliyatiga kirib kelib, ilmiy texnik taraqqiyotning o'sishiga asosiy sababchilardan biri bo'lib kelmoqda [4].

Axborot xajmining jadal suratlar bilan ortib borishi, axborotlarni yig'ish, saqlash hamda ma'lumotlar bazasi va omborini tashkil qilish, shu jumladan, internet texnologiyalarini rivojlanishi natijasida axborotlarni tahlil qilish va

tadqiqot obyektlarini modellashtirishning aniq usullari real hayot talablaridan qisman ortda qolmoqda. Bulardan kelib chiqqan holda shuni ta'kidlash mumkinki, turli soha axborotlarini qayta ishlash uchun yaroqli bo'lgan, shu jumladan, yaqin kelajakda vujudga keladigan muammolarni hal qilishga mo'ljallangan universal va ishonchli yondoshuvlarni ishlab chiqish talab qilinadi[5].

O'zbekiston Respublikasining «Ta'lim to'g'risida»gi Qonuni, Kadrlar tayyorlash milliy dasturi, «Axborotlashtirish to'g'risida»gi Qonun, O'zbekiston Respublikasi Prezidentining «O'zbekiston Respublikasi jamoat ta'lim axborot tarmog'ini tashkil etish to'g'risida» 2005 yil 28 sentyabrdagi PK-191-son qarori hamda O'zbekiston Respublikasi Vazirlar Mahkamasi Kompyuterlashtirish va axborot - kommunikasiya texnologiyalarini rivojlantirish bo'yicha Muvofiqlashtiruvchi kengashning 2008 yil 20-iyundagi 22-son majlis bayonlarida yangi axborot-kommunikasion texnologiyalarni ta'lim berish sohasida qo'llash muhim vazifa sifatida keltirilgan[1,5].

O'zbekistonda axborotlashtirish masalalari qatorida kompyuter vositali ta'lim tizimini rivojlantirishga katta e'tibor berilmoqda, bunga misol sifatida xalq ta'limi tizimida maktablar uchun elektron axborot resurslarini tashkil etish borasida 2000 yildan to hozirgacha qator ko'zga ko'rinadigan ishlar amalga oshirilmoqda. Oliy va o'rta maxsus ta'lim tizimida ham shunga o'xshash kompyuter parkini kengaytirish va yangilash, dasturiy vositalarni ko'paytirish shular jumlasidandir.

Ko'pchilik tadqiqotchilar o'z faoliyatlari davomida juda katta hajmdagi ma'lumotlarni to'playdilar. Lekin ularning yagona istagi – bu to'plangan ma'lumotlar bazasidan o'zlari uchun kerakli bo'lgan ma'lumotlarni tez va aniq qidirib topish hisoblanadi. Qisqacha aytganda ma'lumotlar bazasi xotira yoki murakkab yozuvlar daftari vazifasini bajaradi. Foydalanuvchilarning ma'lumotlar bazasiga kirishi aniq qo'yilgan savollarga javob olish uchun saqlanayotgan ma'lumotlarning unchalik katta bo'lmagan qismini olishini ta'minlaydi. Ulkan axborot oqimiga, to'plangan axborotning ulkan zaxiralariga

ega bo'lganimizda qandaydir jarayonlarni boshqarishni optimallashtirish, tashkilot faoliyatini yaxshilash murakkab obyektlarga, masalan, ta'lim tizimida pedagog-xodimlar, OTMlarning ko'rsatkichlari, talabalar haqida katta xajmdagi ma'lumotlarda mavjud bilimni chiqarish uchun bu ma'lumotdan maksimal maqsadga muvofiq foydalanish masalasi paydo bo'ladi.

Berilgan ma'lumotlar ba'zasidagi ma'lumotlarni yaqinlik xususiyatiga ko'ra guruhlariga ajratish va ulkan axborot oqimidan kerakli ma'lumotni tez va aniq ajratib olish masalasi hozirgi zamonning dolzarb masalasi hisoblanadi.

Mazkur dissertatsiya ishida ma'lumotlarni intellektual tahlil etishning usullari va dasturiy ta'minoti masalalaridan kelib chiqqan holda intellektual tahlil etishning klasterizatsiya masalasi algoritmlarini qiyosiy tahlil etish hamda dasturiy ta'minotlar orqali ifodasini keltirib, o'qitish tizimiga joriy etishni tashkil etish masalalariga bag'ishlangan.

Tadqiqot obyekti va predmetining belgilanishi. Ma'lumotlarni intellektual tahlil etishning sinflashtirish usullari, algoritmlari tadqiqot obyekti bo'lib, intellektual tahlil etish, statistik tahlil, ma'lumotlar strukturasi va algoritmlash tadqiqotning predmeti hisoblanadi.

Tadqiqot maqsadi va vazifalari. Ma'lumotlarni intellektual tahlil etishning klasterizatsiya usullari va algoritmlarini qiyosiy tahlil qilish, ta'lim tizimida shakllanadigan ma'lumotlarni klasterizatsiya usullari orqali intellektual tahlil etishdan iborat.

Dissertatsiya ishida tadqiqot maqsadini amalga oshirilishi uchun quyidagi vazifalar qo'yildi:

- intellektual tahlil bo'yicha adabiyotlarni tahlil qilish;
- ma'lumotlar intellektual tahlilida klasterizatsiya usullarini tavsiflash va qiyosiy tahlil qilish;
- ma'lumotlarni intellektual tahlilning klasterizatsiya usullari uchun dasturiy ta'minotlarini tadqiq qilish;
- klasterizatsiya usulida ma'lumotlarni intellektual tahlilini amaliy masalalarga taqbiq etish.

Tadqiqotning asosiy masalalari va farazlari. Ma'lumotlarni intellektual tahlil etishning klasterizatsiya usullari, ularning qiyosiy tahlillari va o'qitish jarayoni uslubiyatini ilmiy-amaliy asoslash imkoniyatini beradi.

Muammoni o'rganganlik darajasi. O'qitish jarayonida ma'lumotlarning intellektual tahlilini qo'llash, avtomatlashtirilgan shu sohada o'qitishni tashkil etish, qolaversa intellektual elementlarni o'zida mujassamlashtirgan intellektual o'qitish tizimlarni tashkil etishining nazariy asoslari hamda ularning amaliy tadbiq'i borasida bir qator ishlar amalga oshirilgan bo'lib, ular usullar, algoritmik vositalar ko'rinishda o'z aksini topgan. Ushbu ishlar Grigoriy Piatetskiy, L.Zade, R.A.Fisher, A.A.Barsegyan, V.I.Vasilev, Ayvazyan S.A, N.G.Zagoruyko, I.A.Chulkova, N.Paklin, Jan Jantzen va vatandoshlarimiz – T.F.Bekmuratov, O.J.Bobomurodov, Komarov V.A, M.M.Kamilov, N.A.Ignatev, A.R.Axatov, O.N.Raximov izlanish natijalari hamda amaliy tadbiq jarayonlarida o'z aksini topgan.

O'rganib chiqilgan ishlarda, intellektual tahlillar asosida o'qitishni tashkil etish elementlarining modellari tadqiq qilingan, biroq intellektual tahlillar asosida o'qitish tizimini bilimlar bazasida ifodalash masalalari yetarlicha o'rganilmagan. Shu sababli, tadqiqot ishida intellektual o'qitishni tashkil etishda bilimlar bazasini shakllantirish masalalari, shu orqali tizimning samarali ishini yo'lga qo'yish vazifalarini chuqurroq tahlil etish va tadqiq qilishga asos bo'ldi.

Tadqiqotda qo'llanilgan uslublarning qisqacha tavsifi. Ma'lumotlarni shakllantirish, intellektual tahlil etishda klasterizatsiya masalasi algoritmlarining qiyosiy tahlili, o'qitish jarayonida bilimlarni aks ettirish va saralash tizimini tashkil etish masalasini yechishda mantiqiy xulosalash, bilimlar qoidasini qurish, produksion hamda aralash modellar, modulli usullaridan foydalanildi.

Tadqiqotning ilmiy yangiligi. Tadqiqot ishining yangiligi ma'lumotlarni intellektual tahlil etishda klasterizatsiya usullarini sinfi, yutug'i, kamchiligi, ishlov berish ma'lumoti turi hamda hisoblash tezliklari bo'yicha nazariy va amaliy tahlili, o'qitish jarayonida ma'lumotlarni intellektual tahlil etishning klasterizatsiya usullarini o'rgatishda dasturiy ta'minotlaridan foydalanish

uslublari ishlab chiqilganligi, k-means usulida 3 ta klasterga klasterizatsiyalash uchun uslubiy dasturiy vosita ishlab chiqilganligi bilan asoslanadi.

Tadqiqot natijalarining nazariy va amaliy ahamiyati. Tadqiqot natijalarining *nazariy ahamiyati* - ma'lumotlarni intellektual tahlil etishning klasterizatsiya usuli algoritmlarini qiyosiy tahlillari, ma'lumotlar turi va xajmiga mos algoritmi tanlash ahamiyati kasb etadi. *Amaliy ahamiyati* – ta'lim sohasidagi katta xajmli ma'lumotlarni intellektual tahlil etishda klasterizatsiya usullarining dasturiy instrumentlaridan foydalanish taribi, usullarni tahlil qilish uslubiyati hamda ishlab chiqilgan dasturiy ta'minotning ta'lim berish jarayonida uslubiy dasturiy vosita sifatida taqdim etilishi bilan ahamiyatlidir.

Ishning tuzilishi va qisqacha mazmuni: Dissertasiya ishi kirish, 3 ta bob, xulosa, foydalanilgan adabiyotlar ro'yxati hamda ilovadan iborat. Ishning asosiy matni 72 - betdan iborat. 24 ta rasm, 4 ta jadval, 37 nomdan iborat adabiyotlar ro'yxati va ilovadan tashkil topgan.

Birinchi bobda Ma'lumotlarni intellektual tahlil etish usullarining nazariy va amaliy asoslariga bag'ishlangan bo'lib, unda asosiy tushunchalar va ta'riflar, ma'lumotlarning turlari, ularning saqlash formatlari va ma'lumotlar to'plami atributlari, ma'lumotlar bazasi, Data Mining uslublari va bosqichlari nazariy-amaliy usullari, vositalari keltirilgan hamda tahlil etilgan.

Ikkinchi bobda Ma'lumotlarni intellektual tahlil etishda klasterizatsiya usullari algoritmlari va klasterizatsiya masalasini tizimli yechishning formal qo'yilishi va klasterizatsiya masalalari usul va algoritmlarining qiyosiy tahlili to'liq tasnifi keltirilgan.

Uchinchi bobda Ikkinchi bobda keltirilgan klasterizatsiya algoritmlari asosida ma'lumotlarni intellektual tahlilida klasterizatsiya algoritmlari dasturiy ta'minotlarining amaliy tadbirlari asoslab keltirilgan. Klasterizatsiya usullari uchun amaliy dasturiy paketlarda ishlash paragrafidagi turli dasturiy modulardan, MATLAB, SPSS, WEKA dasturiy paketlarini misol qilib keltirilgan va bu algoritmlarning qiyosiy tahlili asoslangan.

Xulosada tadqiqot ishining asosiy natijalari bayon etilgan.

Ilovada dissertasiya ishining tajribaviy tadqiqot natijalarini aks ettiruvchi ekran formalari natijalari hamda dastur kodi keltirilgan.

Maqolalar. Dissertasiyaning asosiy natijalari 3 ta ilmiy maqolada nashr etilgan.

№	Maqola nomi	Hajmi (byet)	Chop etilgan to'plam
1	Intellectual o'qitish tizimlarida bilim oluvchi modeli asosida o'qitishni tashkil etish	2	Axborot kommunikatsiya texnologiyalarining hozirgi zamon rivojlanish bosqichda mutaxassisning kasbiy kompetentligini mukammallashtirish" ilmiy-amaliy konferensiyasi (II qism 2013 yil 6-8 may)
2	Подход к решению задачи приобретения знаний	3	Axborot kommunikatsiya texnologiyalarining hozirgi zamon rivojlanish bosqichda mutaxassisning kasbiy kompetentligini mukammallashtirish" ilmiy-amaliy konferensiyasi (II qism 2013 yil 6-8 may)
3	DATA MINIGA- Ma'lumotlarning intellektual tahlili	3	"Intellectual salohiyatli iqtidorli yoshlar" mavzusidagi yosh olimlar, magistrant va talabalarning an'anaviy XI respublika ilmiy-amaliy konferensiyasi (pedagogika ,psixologiya, ta'lim va tarbiya III qism 2014 yil 23-24 may)

I BOB. MA'LUMOTLARNI ITINTELLEKTUAL TAHLIL ETISH USULLARINING NAZARIY ASOSLARI

1. Ma'lumotlarni intellektual tahlil etish tizimning maqsad va vazifalari tahlili

Hozirgi kunda ma'lumotlar ummonida turli toifa va turlarga tegishli bo'lgan ma'lumotlar hajmi jadal sur'atlar bilan oshib bormoqda. Ma'lumotlar hajmi juda ham katta bo'lganligi sababli, ular ichidan foydalunuvchi o'ziga kerakli bo'lgan axborotlarni ajratib olishi masalasi murakkablashib bormoqda.

Insoniyat o'ziga kerakli axborotlarni izlash va ularni ajratib olishi uchun ma'lumotlarni qayta ishlashi, ularni tahlil qilishi, aniqrog'i ma'lumotlarni intellektual tahlil qilishi lozim bo'lmoqda. Ma'lumotlarni tahlil qilishning an'anaviy usullari, asosan, ma'lumotlar haqidagi oldindan mavjud bo'lgan gipotezalarni tekshirishga qaratilgan bo'lib, intellektual tahlil esa ma'lumotlar tuzilmasini, ma'lumotlar orasidagi ilgari ma'lum bo'lmagan bog'liqliklar va qonuniyatlarni aniqlashga qaratilgan[7,8].

Ma'lumotlar to'plamining turlari va ularni saqlash formatlari. Eng ko'p uchratiladigan ma'lumotlar bu yozuvlardan tashkil topgan ma'lumotlardir. Bunday ma'lumotlar to'plamiga jadvalli ma'lumotlar, matrisali ma'lumotlar, hujjatli ma'lumotlar, tranzaksiyali yoki operatsiyali ma'lumotlarni kiritamiz.

Jadvalli ma'lumotlar – fiksirlangan atributlar to'plamidan tuzilgan yozuvlardan iborat bo'lgan ma'lumotlardir.

Tranzaksiyali ma'lumotlar – har bir yozuv qiymatlari to'plami bilan tranzaksiya bo'lib keladigan ma'lumotlarning alohida turini anglatadi. Tranzaksiyali ma'lumotlar bazasiga-magazinda xaridorlarning qilgan savdolaridan tuzilgan ro'yxatni misol qilib olishimiz mumkin.

Grafikli-ma'lumotlarga misol sifatida *www-ma'lumotlari*, molekulalar strukturasi, grafalar, kartalar va shu kabilarni ayta olamiz.

Hozirgi vaqtda ma'lumotlarning asosiy xususiyatlaridan biri ular juda ko'p o'zgarishi natijalarida qaytadan tuzilishidir. Ma'lumotlar bilan ishlashning to'rtta jihatida mavjud: ma'lumotni aniqlash, hisoblash, manipulyasiya qilish va

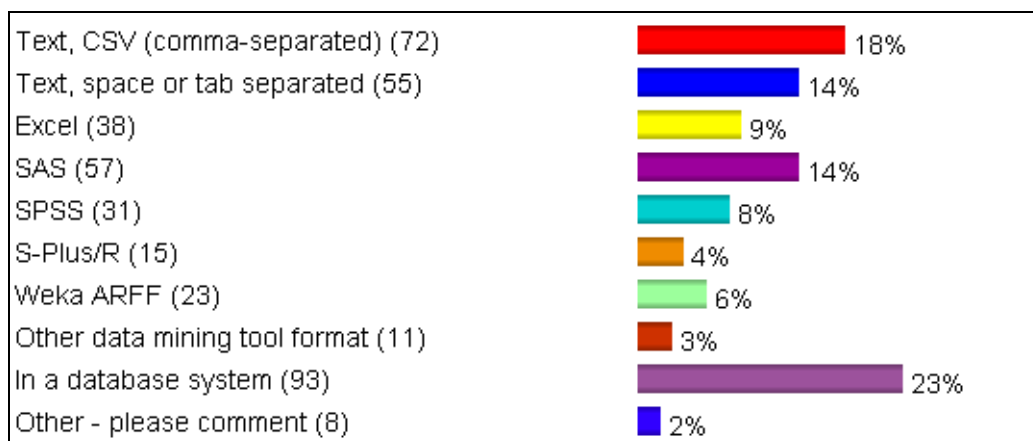
qayta ishlash (yig'ish, uzatish va h.k.)

Manipulyasiya qilingan ma'lumotlar orqali «fayl» tipidagi ma'lumotlar strukturasiidan foydalaniladi. Fayllar har xil formatga ega bo'lishi mumkin.

Data Mining (DM) aksariyat instrumentlari turli xil manbalardan ma'lumotlarni import qilishga ruxsat beradi hamda natija sifatida olingan ma'lumotlarni turli xil formatga eksport qiladi[8].

Tajribalar uchun ma'lumotlarni qandaydir yagona formatda saqlash qulay bo'ladi. DM ning ba'zi instrumentlaridagi proseduralar ma'lumotlarning importi/eksporti deb ataladi. Boshqalari esa turli ma'lumotlar manbalarini to'g'ridan-to'g'ri ochish imkonini beradi va DM natijalarini ko'rsatilgan formatlardan biriga saqlaydi.

Ma'lumotlarni saqlash formatlari eng keng tarqalgani 1.1-rasmda tasvirlangan:



1.1-rasm. Ma'lumotlarni saqlash bo'yicha eng ko'p tarqalgan formatlar.

So'roqlar sonning eng ko'pi 23% bo'lib, ular ma'lumotlar bazasi formatida saqlanadigan ma'lumotlarni tashkil etgan. Text, CSV formatlari – 18%, 14% li so'roqlarni Text, space or tab separated I SAS formatlarida saqlanadigan ma'lumotlar; 9% Excel formatida, SPSS da – 8%, S-Plus/R da – 4%, Weka ARFF da- 6% va Data Mining (DM) instrumentlarining boshqa formatlarida - 2%ni tashkil etgan. So'roqlar natijasida ko'rinib turganidek, DM uchun eng ko'p saqlangan ma'lumotlar formati bu ma'lumotlar bazasini ko'rsatadi [8].

Data Mining ma'lumotlarni intellektual tahlil qiluvchi texnologiya hisoblanib, amaliyotda keng qo'llanilib kelinmoqda.

DM odatda ikki xil ma'noni bildiradi, ya'ni katta hajmdagi ma'lumotlar bazasi (MB)dan kerakli ma'lumotlarni qidirib topish hamda katta hajmdagi ishlov berilmagan materialni mazmunan tadqiq qilish demakdir.

DM ma'lumotlarni intellektual tahlili, qonuniyatlarini topish muhiti, bilimlarni kengaytirish, shablonlarni tahlil qilish, MBdan bilimlarni axborot tarkibini aniqlash va h.k. kabi ma'noni anglatadi.

DM tushunchasi 1978 yillarda vujudga kela boshladi va 1990 yillarning birinchi yarim yillarida zamonaviy talqinda yuqori ommaboplikka erishildi. Bu vaqtlarda ma'lumotlarni qayta ishlash va tahlil qilish amaliy statistikaning tarkibida bo'lgan, shuning uchun qayta ishlash masalalari asosan katta bo'lmagan ma'lumotlar bazalarida olib borilgan edi[9].

DM– ma'lumotlardan yashirin qonuniyatlarni (axborotlar shablonlarini) aniqlab qaror qabul qilishga asoslangan jarayonidir.

Bu texnologiyaning mohiyati va maqsadi katta hajmdagi ma'lumotlardan ma'lum bo'lmagan obyektiv va amaliy foydali qonuniyatlarini aniqlash uchun mo'ljallangan.

Ma'lum bo'lmaslik – topilgan qonuniyat axborotni qayta ishlashning standart metodlari yoki ekspert yo'llar bilan aniqlanmaydi.

Obyektiv – aniqlangan qonuniyatlar haqiqatga to'liq mos keladi va ekspertli mulohazalardan farqi uning hamma vaqt subyektiv bo'lmasligidadir.

Amaliy foydali – amaliy qo'llanilish qonuniyatlari topilganda tahlil qilinayotgan ma'lumotlar aniq qiymatlarga ega bo'ladi.

Bilimlar – tavsiflanadigan manbalar, fan sohalari, muammolar va h.k. haqida ba'zi bir xulosalarni beradigan ma'lumotlar yig'indisidir.

Bilimlardan foydalanish aniq afzalliklarga erishish uchun topilgan bilimlarning amaliy qo'llanilishi tushuniladi.

Data Mining “topilma” yoki “topilmani qazib olish” kabi tarjima qilinadi. Ko'pincha Data Mining bilan birgalikda “ma'lumotlar bazasida bilimlarni izlab

topish” (knowledge discovery in databases) va “ma’lumotlarning intellektual tahlili” soʻzlari ishlatiladi. Ular Data Mining ning sinonimlari hisoblanadi. Yuqorida keltirilgan barcha terminlar ma’lumotlarni qayta ishlash vositalari va usullari bilan bogʻlangan.

Shu bilan birgalikda, amaliyotchilar amaliy masalalarni echish uchun nazariy fikrlarni qoʻllash koʻpincha natijasiz ekanligini bilishardi. Ammo ular bular bilan ishi yoʻq-oʻzlarining katta boʻlmagan lokal ma’lumotlar bazasini qayta ishlash kabi xususiy masalalari ustida bosh qotirishar edilar[9,10].

Ma’lumotlarni yozish va saqlash texnologiyalarining mukammallashuvi natijasida insonlarga turli sohalaridan koʻplab axborot oqimlari yopirildi. Ixtiyoriy tashkilot (tijorat, ishlab chiqaruvchi, meditsina, ilmiy va sh.oʻ.) ning ishi uning ish faoliyatini toʻliq roʻyxatdan oʻtkazish va yozish bilan bogʻliq boʻlib qoldi. Bu axborotlar bilan endi nima qilish kerak? Ma’lum boʻldiki, ma’lumotlar oqimi qayta ishlanmasa, ular keraksiz matohga aylanib qoladi.

Bunday qayta ishlashlar uchun zamonaviy talablar spetsifikasi quyidagicha:

- Ma’lumotlar cheklanmagan hajmga ega;
- Ma’lumotlar turli xil boʻladi (sonli, sifatli, matnli);
- Natijalar aniq va tushunarli boʻlishi kerak;
- Ma’lumotlarni qayta ishlash vositalari ishlatish uchun sodda boʻlishi kerak.

Ma’lumotlar analizining asosiy vositasi roliga koʻp vaqt davomida daʼvogarlik qilgan anʼanaviy matematik statistika paydo boʻlgan muammo oldida ochiqchasiga taslim boʻldi. Asosiy sabab – soxta kattaliklar ustida amallar bajarishga olib keladigan *tanlash boʻyicha yaqinlashtirish konsepsiyasi* (kasalxonadagi bemorning oʻrtacha temperaturasi, koʻchadagi qasr va kulbalardan iborat uylarning oʻrtacha balandligi va kabi amallar). Matematik statistika usullari oldindan ifodalangan gipotezlarni tekshirish uchun (verification-driven data mining) va ma’lumotlarni operativ analitik qayta ishlashga (online analytical processing, OLAP) asoslangan “qoʻpol” razvedka analizi uchun foydali boʻlib chiqdi[10].

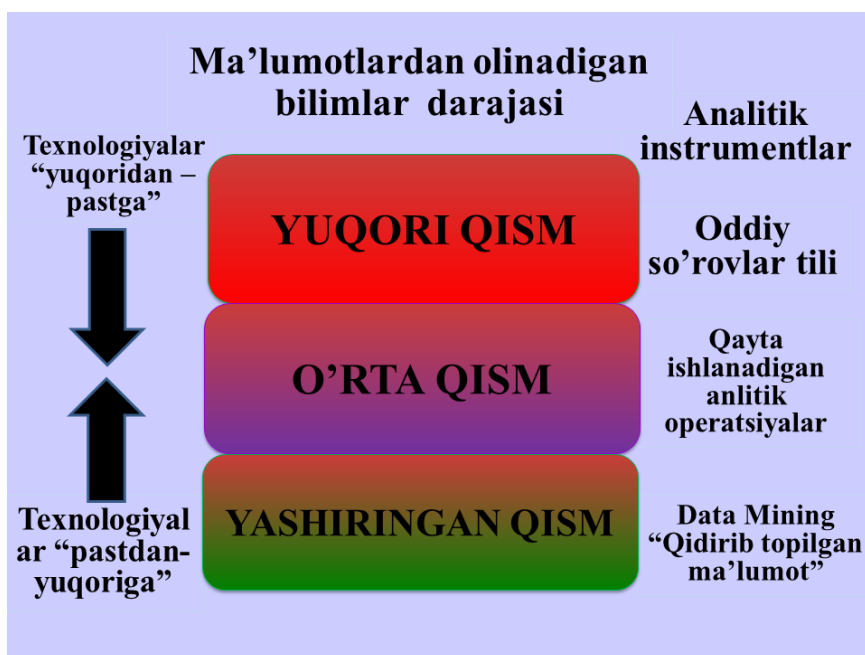
Zamonaviy Data Mining (discovery-driven data mining) asosiga ma'lumotlarning o'zaro munosabatida ko'p aspektli fragmentlarini tasvirlaydigan shablonlar (patternlar) konsepsiyasi qo'yilgan. Bu shablonlar o'zida inson uchun tushunarli bo'lgan ma'lumotlarni tanlab olishga asoslangan qonuniyatlarni tasvirlaydi. Shablonlarni izlash tanlash strukturasi bo'yicha tajribada ko'rib chiqilgan va analiz qilinadigan ko'rsatkichlar qiymatlarini taqsimlash ko'rinishidagi usullar bilan amalga oshiriladi. Bunday izlashga misol quyidagi 1-jadvalda keltirilgan.

1-jadval. OLAP va Data Mining metodlarini qo'llab masalalarni yechishga misol

OLAP	Data Mining
Chekuchilar va chekmaydiganlar orasida jarohatlanishlarning o'rtacha ko'rsatgichi qancha?	Yuqori jarohatlanishga duchor bo'lgan odamlarni tasvirlashda aniq shablonlar mavjudmi?
Mavjud mijozlarning telefon hisoblarining sobiq mijozlar hisoblariga taqqoslaganda o'rtacha miqdori qancha? (telefon kompaniyasi xizmatlaridan voz kechgan)	Telefon kompaniyasi xizmatlaridan voz kechadigan mijozlarning umumiyliги mavjudmi?
O'g'irlangan va o'g'irlanmagan kredit kartochkalari bo'yicha kunlik xaridlarning o'rtacha qiymati qancha?	Kredit kartochka bo'yicha qalloblik holatlari uchun bir xillik sxemasi mavjudmi?

Data Mining uchun muhim holat – izlab topilgan shablonlarning ma'noliligi. Bu topilgan shablonlar maxfiy bilimlar (hidden knowledge) deb ataluvchi ma'lumotlarda tasavvur qilinmagan, kutilmagan (unexpected) muntazamlilikni tasvirlashi kerak. Jamiyat shunday fikrga keldiki, zarur (сырые) ma'lumotlar (raw data) chuqur bilim manbai bo'lib, uni savodli

izlashdan haqiqiy bilimlar topiladi (1.2-rasm).



1.2-rasm. Ma'lumotlardan olinadigan bilimlar darajasi

Data Mining texnologiyasiga, bu texnologiya asoschilaridan biri Grigoriy Piatetskiy-Shapiro etarlicha to'liq ta'riflaydi:

Data Mining - bu zarur ma'lumotlarda inson faoliyatining turli sohalarida echimlarni qabul qilish uchun zarur bo'lgan noma'lum, ma'noli, amaliy foydali va kirish mumkin bo'lgan bilimlar interpretatsiyasini topib olish jarayonidir[9].

2. Ma'lumotlar to'plami va ularning atributlari. Ma'lumotlar bazalari.

Data Mining usullari va bosqichlari

Keng ma'noda ma'lumot tushunchasi faktlar, matnlar, chizmalar, rasmlar, ovozlar, analogli yoki raqamli video tasvirlarni o'zida aks ettiradi.

Ma'lumot o'lchovlar, tajribalar, arifmetik va mantiqiy amallarni bajarish jarayonida olinishi mumkin.

Ma'lumotlarni saqlash, uzatish va qayta ishlash qulay bo'lishi uchun ularni ma'lum bir obyektlar majmuasi shaklida tasvirlash kerak.

Obyektni atributlar to'plami deb ta'riflash mumkin. Obyekt yozuv, hodisa, misol, jadvalning qatori va h.k. bo'lishi mumkin va ular ma'lum bir o'zgaruvchilar orqali namoyon etiladi.

Atribut obyektning xususiyatini ifodalovchi ma'lumot yoki ma'lumotlar majmuasidir.

O'zgaruvchi – bir obyekt xususiyatlariga ko'ra boshqa bir obyektни hosil qiluvchi tushunchadir.

Ma'lumotlar tahlili davomida juda katta hajmdagi ma'lumotlarni o'rganish qimmatli jarayonlarni, ko'p vaqt sarflanishini talab etishini hamda inson faktorlari hisobga olinsa ushbu jarayonda muqarrar xatoliklar yuz berishi mumkin.

Ma'lumotlarning barcha to'plamlarining ba'zi bir qismlari to'liq qarab chiqilib ya'ni tanlanmalar asosida ulardan qiziqarli axborotlar olinib xulosa hosil qilinadi.

O'lchovlar–aniqlangan qoidaga muvofiq o'rganilayotgan obyektlar xususiyatlarining miqdoriy ko'rsatkichidir.

Ma'lumotlarni tayyorlanish jarayoni obyekt bilan emas uning xususiyatlari bilan o'lchanadi [12].

Shkala–DMning ko'p uskunolari boshqa manbalardan ma'lumotlarni import qilishi davomida har bir o'zgaruvchi uchun shkalalar turini tanlashi va kiruvchi hamda chiquvchi (belgili, sonli, diskretli va uzluksiz) o'zgaruvchilar uchun ma'lumotlar turini tanlash ko'rsatkichidir. O'lchovlarning besh turdagi shkalalari mavjud: *nominal, tartibli, oraliqli, nisbiy va dixotomik*.

Nominal skala (nominal scale) – faqat kategoriyalardan tuziladigan shkaladir, uning ma'lumotlari tartiblanmasligi mumkin va bu shkalalar ustida hech qanday arifmetik harakatlarni o'tkazib bo'lmasligi mumkin.

Bu *shkala* nomlanishlar, kategoriyalar, obyektlarning klassifikatsiyasi va tartiblanishi yoki ba'zi belgilar bo'yicha kuzatishlar uchun quyilgan nomlardan tuziladi.

Unga misol: kasblar, yashash joyi, oilaviy ahvoli va shu kabilar.

Bu *shkalalar* uchun faqat tenglik ($=$) va teng emas ($\neq$) operatsiyalari qo'llaniladi.

Tartiblangan shkala (ordinal scale) – obyektning nisbiy holatini belgilash uchun xizmat qiladigan obyektlardan o'zlashtiriladigan sonlar shkalasidir, lekin ular orasidagi farq kattaliklari mavjud emas.

Shkala o'lchovlarga o'zgaruvchilar qiymatlarini safga tizish (ranjirovat) imkoniyatini beradi. Tartibli shkaladagi o'lchovlar faqat keladigan kattaliklarning tartiblanganligi haqidagi informatsiyalardan tuziladi, lekin «bir kattalik boshqasidan qanchalik katta» yoki «u qanchalik boshqasidan kichik» deb mulohaza yuritishga yo'l qo'ymaydi.

Tartiblangan shkalaga misol: guruhning musobaqada olgan o'rni (1-, 2-, 3-), talabning qobiliyatlik reytingidagi tartib raqami (1-, 15- va h.k.) keltirilgan bo'lsin. Bu bilan bir talabning boshqasidan qanchalik qobiliyatligini aniqlab bo'lmaydi, uning faqat reytingdagi raqamigina ma'lum xolos.

Bu shkalalar uchun faqat tenglik ($=$), teng emas ($\neq$), katta ($>$), kichik ($<$) operatsiyalari ishlatish mumkin.

Oraliqli shkala (interval scale) – bu shkala ikki kattlik orasidagi farqni topishga ruxsat beradi, nominal va tartibli shkalalarning xususiyatlariga ega bo'ladi, hamda belgilarning miqdoriy o'zgarishlarini aniqlashga ruxsat beradi.

Nominal va tartibli shkalalar diskret, oraliqli shkala esa uzluksiz bo'ladi. U belgining aniq o'lchovlarini amalga oshirishga ruxsat beradi va qo'shish, ayirish, ko'paytirish, bo'lish arifmetik operatsiyalari bajariladi.[36]

Bu shkala uchun faqat tenglik ($=$), teng emas ($\neq$), katta ($>$), kichik ($<$), qo'shish (+), ayirish (-) operatsiyalaridan foydalaniladi.

Nisbiy shkala (ratio scale) – aniqlangan sanoq boshida va shkalalar qiymatlari orasidagi munosabatlarda mavjud bo'lgan shkaladir[11,12].

Ma'lumotlar bazalari. Ma'lumotlar bazasi (Database) – bu ma'lumotlarning tashkil etilishi va saqlanishining o'ziga xos elektron ko'rinishidir.

Tashkil etilishining o'ziga xosligi – bu ma'lumotlarni qandaydir aniq usullar bilan tashkil etilishidir, bu usullar ularni qidirish, bir yoki bir nechta ilovalarning murojaat etishini yengillashtiradi, hamda ma'lumotlarni bunday tashkil etish bilan ularning ortiqchaligini kamaytirish ko'zda tutiladi.

MB informasion texnologiyalarning turli ko'rinishlaridan biri hamda ma'lumotlarni saqlash formasi hisoblanadi.

MBni yaratishdan maqsad EHMda ma'lumotlarni dasturiy ta'minotga, texnik qo'llanilish muhitiga va fizik joylashuviga bog'liq bo'lmaydigan tizimlarini qurishdan iboratdir. Ma'lumotlarning bu tizimlarini qurish ma'lumotlarning bir-biriga zid emasligi va yaxlitligini ta'minlashi kerak. MBni loyihalashda uni ko'pmaqsadli foydalanish mo'ljallanadi

Ma'lumotlar sxemasi – ma'lumotlarni maxsuslashgan tilda tasvirlanadigan va MBBT da qayta ishlanadigan mantiqiy strukturasini bildiradi.

Foydalanuvchi sxemasi – jadval maydonining bir variantidagi tartibi aniq foydalanuvchilar uchun fiksirlanadi.

Ma'lumotlar bazasini boshqarish tizimi(MBBT)- bu MBni tashkil etilishini, saqlanishini, yaxlitligini, o'zgarishlarni kiritish, o'qish va informasiyalarni xavfsizligini nazorat qiluvchi dasturiy ta'minotdir.

MBBT jadval strukturalarini qurish va ularni ma'lumotlarga to'ldirish yordamida u yoki bu MB olinadigan qolipni taqdim etadi.

Relyasion ma'lumotlar bazasini boshqarish tizimi – bu relyasion modellar asosidagi MBBTdir[11].

Ma'lumotlarning relyasion modelida har qanday ko'rinishdagi ma'lumotlar relyasion jadvallarda (o'ziga xos turdagi ikki o'lchovli jadvallarda) yig'iladi. Relyasion ma'lumotlar bazasini boshqarish tizimidan ma'lumotlar omborini qurish uchun foydalaniladi.

Quyida ma'lumotlarning bir nechta sinflanishlarini keltiramiz:

Relyasion ma'lumotlar – bu relyasion bazadagi ma'lumotlardir.

Ko'p o'lchovli ma'lumotlar– bu OLAP ning kublarida keltiriladigan ma'lumotlar.

O'lchov yoki o'q– ko'p o'lchovli ma'lumotlarda bir yoki boshqa tip bo'yicha ko'p o'lchovli MBni tizimlashtirishga xizmat qiladigan ma'lumotlarni yig'ishdir.

Doimiylik mezon bo'yicha masalalarni yechilishida ma'lumotlarning qiymatlari quyidagicha bo'lishi mumkin:

- o'zgaruvchan;
- doimiy;
- shartli-doimiy

O'zgaruvchan ma'lumotlar – masalalarni yechish jarayonida qiymatlari o'zgarib boradigan ma'lumotlardir.

Doimiy ma'lumotlar - masalalarni yechish jarayonida qiymatlarini saqlab qoladigan va tashqi faktorlarga bog'liq bo'lmaydigan ma'lumotlardir.

Shartli-doimiy ma'lumotlar – o'z qiymatlarini gohida o'zgartirib turadi, lekin bu o'zgarishlar masalaning yechilishidagi faktorlarga bog'liq bo'lmay, tashqi faktorlar ta'sirida yuzaga keladi[11].

Ma'lumotlar shunday funksiyalarga bog'liq bo'ladiki, ularning vazifalari yordamida ma'lumotlar – ma'lumotnomali, tezkor, arxivli bo'lishi mumkin.

Ma'lumotlarni – davriy ma'lumotlar va ayni vaqtdagi ma'lumotlarga bo'linishiga olib keladi. Bu bo'linish axborotlarni yig'ish sistemalarini loyihalashtirish uchun muhim hisoblanadi.

Davriy ma'lumotlar – qandaydir vaqt oralig'ini tavsiflaydi. Bu ma'lumotlarga misol sifatida korxonaning bir oylik foydasi, oyning o'rtachi harorati kabilarni qarash mumkin.

Ayni vaqtdagi ma'lumotlar – bir nechta o'zgaruvchilarning aniq paytdagi qiymatlarini taqdim etadi.

Ma'lumotlar birlamchi va ikkilamchi bo'ladi. Ikkilamchi ma'lumotlar aniqlangan hisoblashlar natijasida birlamchi ma'lumotlarga o'zgaradigan ma'lumotlardir. Ikkilamchi ma'lumotlar qoidadagidek, saqlanayotgan axborotlar hajmini kengayishi hisobiga foydalanuvchining so'roviga tezkor javob olishga olib keladi[12].

Metama'lumotlar – bu ma'lumotlar haqidagi ma'lumotlardir. Uning tarkibiga kataloglar, ma'lumotnomalar, reyestrlarni kiritishimiz mumkin.

Metama'lumotlar – ma'lumotlar tarkibidagi xabarlarni tashkil etadi; tuzilishi, kelib chiqishi, joylashishi, sifati, tasvirlanishi, formatlari va shakllari, murojaat etish mumkinligi shartlari, olinishi va foydalanilishi va boshqalar.

Ma'lumotlarni saqlanish joyida qo'llaniladigan metama'lumotlar uning o'rnatilishi va foydalanilishi uchun zaruriy axborotlardan tuziladi. Ular biznes-metama'lumotlar va tezkor metama'lumotlarga bo'linadi.

Biznes-metama'lumotlar – biznes-terminlari va ta'riflari, ma'lumotlarning belgilari va saqlanish xizmatida to'lov qoidalaridan tashkil etiladi.

Tezkor metama'lumotlar – ma'lumotlarni saqlanishi vaqtida yig'ilgan axborotlardir:

- Ma'lumotlarning ko'chirilishi va o'zgartirilishini;
- Ma'lumotlardan foydalanish haq-huquqi (faollashtirish, arxivlangan va o'chirilgan);
- Monitoring ma'lumotlari - statistika foydalanadigan xatoliklar haqidagi xabar va boshqalar[12].

Ma'lumotlar bazasi texnologiyasining rivojlanishi 1968 yilda IBM firmasining birinchi MBBT IMS tizimi ishlab chiqarishga kiritilgan. 1975 yilda ma'lumotlarning tarmoqli modeli uchun hozirgacha asos bo'lib xizmat qiladigan ma'lumotlar bazasi tizimining nazariyasida bir qator fundamental tushunchalarni aniqlaydigan ma'lumotlarni qayta ishlash tizimining tillari bo'yicha assotsiatsiyasining birinchi standarti – Conference of Data System Languages (CODASYL) hisoblanadi. Ma'lumotlar bazasining keyingi rivojlanishiga ma'lumotlarning relyasion modelining yaratuvchisi hisoblanadigan amerikalik matematik E.F. Koddning hissasi katta.

1980 yillar bu vaqt mobaynida tadqiqotchilar MB sini qurishning yo'nalishida yangi qurilishni va unga ruxsatli kirishni ta'minlashni tadqiq etadilar. Bu izlanishlarning maqsadi ma'lumotlarni yanada oddiy modellashtirish uchun relyasion prototipni olish hisoblanadi. Natijada 1985 yilda SQL deb

ataladigan til yaratildi. Bugungi kunda barcha MBBT lar asosan shu interfeysni ta'minlaydi.

1990 yillar ma'lumotlarning spetsifik turlari paydo bo'ladi – "grafli tasvir", "ovo", "hujjat", "xarita". SQL tiliga vaqt uchun ma'lumotlar turlari, vaqtlar intervali, belgilarni tasvirlashning 2 baytli belgisi qatori qo'shilgan. Data Mining texnologiyasi, ma'lumotlar saqlanadigan joy, multimediyali ma'lumotlar bazasi va ma'lumotlarning web-bazalari paydo bo'ldi.

Data Mining ning paydo bo'lishi va rivojlanishi turli faktrlarga asoslangan, ularning asosiylari quyidagilar hisoblanadi:

- apparatli va dasturiy ta'minotning mukammalligi;
- ma'lumotlarni saqlash va yozish texnologiyasining mukammalligi;
- retrospektiv ma'lumotlarning katta sonini yig'ish;
- axborotlarni qayta ishlash algoritmlarining mukammalligi[12].

Data Mining uslublari va bosqichlari. DMning asosiy xossasi bu – keng matematik instrumentlarni (avvalgi statistik tahlildan hozirgi yangi kibernetik uslublarga bo'lgan) birgalikda olib borish va IT sohasidagi yutuqlarga erishish. DM texnologiyasida qat'iy formallashtirilgan uslublarning formallashtirilmagan tahlillari birlashadi.

DM ning uslublari va algoritmlariga quyidagilar aloqador bo'ladi: sun'iy neyron tarmoqlari, qaror yoki yechimlar daraxti, simvolli qoidalar, qo'shniga yaqinlashish va k ta yaqin qo'shni uslublari, tayanch vektorlar uslubi, bayes tarmoqlari, chiziqli regressiya, korrelyatsion-regression tahlil; klaster tahlilida iyerarxik uslublarning klaster tahlilida iyerarxik bo'lmagan uslublari, shu bilan birga k-o'rtacha va k-mediana algoritmlari; assosiyativ qoidalarni qidirish uslublari, shu bilan birga Aprior algoritmi; chegaralangan perebor uslubi, evolyusion dasturlash va genetik algoritmlar, ma'lumotlarni vizuallashtirishning har xil turdagi uslublari va boshqa uslublarning to'plamlari[12].

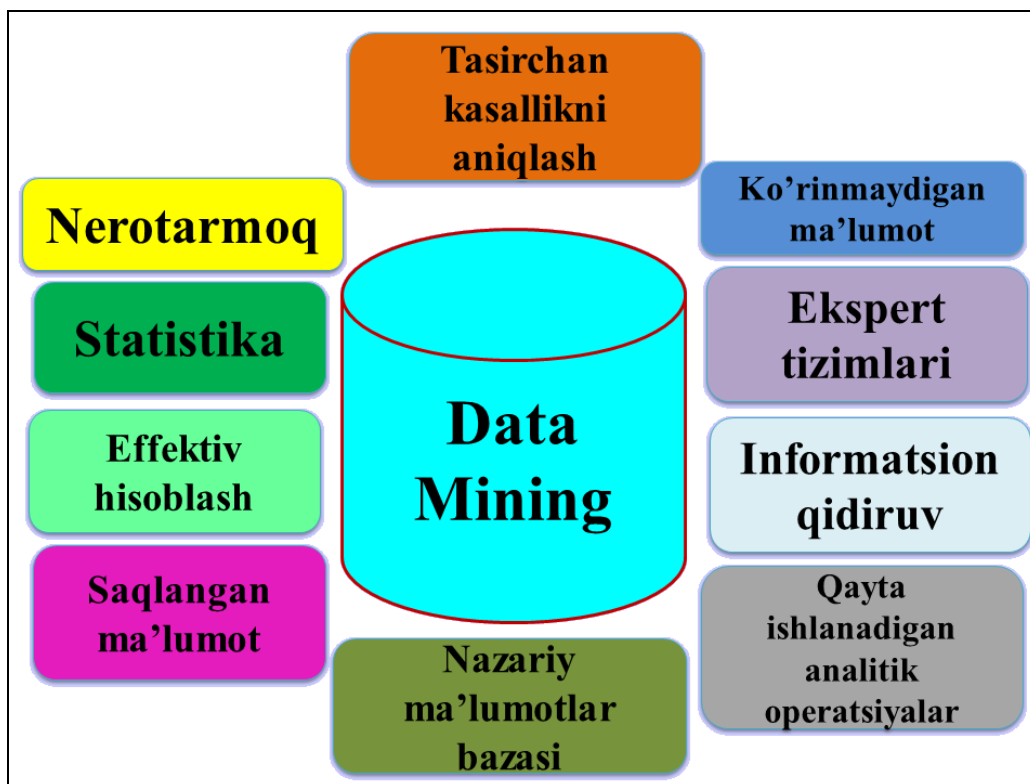
DM texnologiyalarida ishlatiladigan aksariyat analitik uslublarning - bu aniq bo'lgan matematik algoritmlar va uslublardir. Ularning qo'llanilishidagi yangiliklar ularga texnikoviy va dasturiy muhitlarning shartli ravishda paydo

bo'ladigan imkoniyatlari u yoki bu aniq muammolarni yechishda foydalanadigan imkoniyatlarni vujudga keltiradi. Ta'kidlab o'tish kerakki, DMning aksariyat uslublari sun'iy intellekt nazariyasi doirasida ishlab chiqilgan.

Data Mining ni ishlatish sohasi chegaralanmagan – u hamma erda mavjud. Ammo birinchi navbatda Data Mining metodlariga hozirgi vaqtda axborotli ma'lumotlar ombori (Data Warehousing) ga asoslangan loyihalarni ishlab chiqarayotgan tijorat kompaniyalarining qiziqishi ortdi. Bunday tashkilotlarning tajribasi shuni ko'rsatdiki, Data Mining ni ishlatishdan unumdorlik 1000% ga ko'tariladi. Masalan, birinchi marta 350 dan 750 ming dollargacha sarflashlar keyinchalik 10-70 marta ko'p foyda bergan. 20 mln. dollarlik loyiha 4 oyda harajatlarini qoplagan. Boshqa misol - Data Mining ni qo'llash natijasida Buyuk Britaniyadagi universamlar tarmog'i yillik 700 ming dollar foyda ko'ra boshlagan[12].

Data Mining boshliqlar va analitiklarning kunlik faoliyati uchun katta ahamiyatga ega. Ishbilarmon odamlar Data Mining metodlari yordamida katta muvaffaqiyatlarga erishishlari mumkinligini tushunib etdilar. Data Mining ning ba'zi biznes-ilovalariga ta'rif beramiz.

Data Mining amaliy statistika, tasvirlarni tanish, sun'iy intellekt usullari, ma'lumotlar bazasi nazariyasi va bularning rivojlanishi natijasida paydo bo'lgan va rivojlanayotgan ko'p tartibli soha hisoblanadi. Bundan esa Data Mining sistemalarida qo'llaniladigan usullar va algoritmlar ko'pchilikni tashkil etadi. Ko'pgina bunday tizimlar o'z ichiga bir nechta yondashuvlarni birlashtiradi. Shunga qaramasdan, qoidaga ko'ra, har bir tizimda qandaydir kalitli komponenta bo'lib, unga asosiy talab qo'yiladi. Quyida ish asosida ko'rsatilgan kalitli komponentalar sinflanishi keltirilgan. Ajratilgan sinflarga qisqacha ta'rif beriladi[12].



1.3-rasm. Data Mining — ko‘p tartibli soha

Xozirgi kunda mavjud Data Mining uchun mahsulotlar sifatida quyidagilarni qarashimiz mumkin.



1.4-rasm. Data Mining uchun taniqli mahsulotlar

Statistika – bu ma’lumotlarni yig‘ish usullari, ularni qayta ishlash va o‘rganilayotgan holatlarda qonuniyatni ko‘rsatish uchun tahlil etish haqidagi fandir[12].

Statistika – tadqiqotlarni rivojlantirish, ma'lumotlarni yig'ish, ularni tasvirlash va kengaytirish hamda bu ma'lumotlar asosida natijalarni olish va tahlil qilish usullari to'plami hisoblanadi.

Statistika izlanishlar yoki tadqiqotlar natijasida olingan ma'lumotlarga asoslanadi. Keyingi boblardan biri ma'lumotlar tushunchasiga bag'ishlanadi.

Mashinali o'qitish tushunchasi - yagona mashinali o'qitish ta'rifi hozirgi kunda yo'q. Mashinali o'qitishni Yangi bilimlarni olish jarayoni kabi tavsiflash mumkin. Mitchell 1996 yilda quyidagi ta'rifni berdi: "Mashinali o'qitish – bu ish vaqtida avtomatik Yaxshilanadigan kompyuter algoritmlarini o'rgatadigan fandır".

Sun'iy intellekt – ilmiy yo'nalish bo'lib, uning sohasida intellektual hisoblanadigan inson ish faoliyati ko'rinishidagi apparatli yoki dasturli modellashtirish masalalari qo'yiladi va echiladi.

Intellekt tushunchasi (intelligence) lotincha intellectus so'zidan kelib chiqqan bo'lib, aql, faoliyat, ong, insonning o'ylash imkoniyati degan ma'noni bildiradi.

Shunga mos holda, sun'iy intellekt (AI, Artificial Intelligence) avtomatik tizim xususiyatlari sifatida inson intellektining alohida funksiyalarini o'zida olishni bildiradi. Sun'iy intellekt deb inson prerogativi (ayrim huquqlari) hisoblanadigan amaliy funksiyalarni bajaruvchi intellektual tizim xususiyatlariga aytiladi.

Data Mining ni shakllantiradigan har bir yo'nalish o'zining xususiyatlariga ega. Ularning ba'zilarini taqqoslaymiz.

Statistika, mashinali o'qitish va Data Mining ni taqqoslash va boshqalar.

Statistika. Data Mining ga ko'ra nazariyaga asoslanadi. Gipotezni tekshirishda muhimroq.

Mashinali o'qitish. Yanada evristik. O'qitish agentlari ishini yaxshilashga mo'ljallangan.

Data Mining. Nazariya va evristikaning integratsiyasi. O'z ichiga ma'lumotlarni tozalash, o'qitish, integratsiya va natijalarni vizuallashtirishni oladi.

Data Mining tushunchasi MB texnologiyasi va keyingi ma'ruzalarda ko'rib chiqiladigan ma'lumotlar tushunchasi bilan bog'langan.

Data Mining – bu yopiq qonuniyatdagi (axborotlar shablonlari) ma'lumotlarni izlashga asoslangan echimlarni qabul qilish jarayonidir.

Data Mining texnologiyasini shu yo'nalishning asoschilaridan biri – Grigoriy Piatetskiy-Shapiro (Grigoriy Piatetskiy-Shapiro) aniq ta'riflab bergan.

Data Mining – bu xom ashyo ma'lumotlarida hozircha noma'lum, noaniq, amaliy jihatdan inson ish faoliyatining turli sohalarida echimlarni qabul qilish uchun zarur bo'lgan amaliy jihatdan foydali va bilimlar interpretatsiyasi uchun mumkin bo'lgan jarayondir[12].

3. Tadqiqot maqsadi va vazifaning qo'yilishi

Yuqoridagi ma'lumotlardan ma'lumotlarni intellektual tahlil etish choralaridan kelib chiqqan holda dissertatsiyaning asosiy maqsadi ma'lumotlarni intellektual tahlil etishda klasterizatsiya masalasi algoritmlarini qiyosiy tahlil qilish hamda dasturiy ta'minotlar asosida tahlilini berishdan iborat.

Bu maqsadni amalga oshirish uchun dissertatsiya ishida quyidagi masalalar echilgan:

- intellektual tahlil (Data Mining) bo'yicha adabiyotlarni tahlil qilish;
- ma'lumotlar intellektual tahlilida klasterizatsiya usullarini tavsiflash va qiyosiy tahlil qilish;
- ma'lumotlarni intellektual tahlilning klasterizatsiya usullari uchun dasturiy ta'minotlarini tadqiq qilish;
- klasterizatsiya usulida ma'lumotlarni intellektual tahlilini amaliy masalalarga taqbiq etish.

Amalga oshirilgan tadqiqot masalalarida ma'lumotlarni intellektual tahlil etishni tashkil etishning klasterizatsiya masalasi asosiy algoritmlarini qiyosiy tahlil etish hamda o'qitishda qo'llanilishi mumkin bo'lgan dasturiy vositasi asosida qilingan tahlillar samarasini oshirishi ko'rsatib berilgan[11,12].

I bob bo'yicha xulosalar

1. Ma'lumotlarni intellektual tahlil etish tizimlarini turli xildagi ko'rinishlari Data Mining ning qo'llanilish sohalari hamda tahlil etish jarayoni usullari asosi keltirilgan.
2. Ma'lumotlarni intellektual tahlil etishning usullari va dasturiy maxsulotlari keltirilgan. Bu maxsulotlar jaxonda qo'llanilish xamda vazifasi jixatdan muammoli ekanligi keltirib o'tilgan, qurilish kontseptsiyasi tadqiq qilindi.
3. Ma'lumotlar to'plamining turlari, ma'lumotlar bazasi berilish usullari xamda Data Mining usullari va bosqichlari keltirilgan.
4. Ushbu ishlarning tahlili natijalari asosida dissertatsiya ishining asosiy maqsad va vazifalari ishlab chiqildi.

II BOB. MA'LUMOTLARNI INTELLEKTUAL TAHLIL ETISHDA KLAUSTERIZATSIYA MASALASI ALGORITMLARI

Ushbu bobda ma'lumotlarni intellektual tahlil etish jarayonida klasterizatsiya masalasi algoritmlari aks ettirilgan bo'lib, paragrafning birinchi qismida klasterizatsiya masalasini tizimli yechishning formal qo'yilishi va algoritmlar tahlillari to'g'risida ma'lumotlar berilgan va ular qarab o'tilgan. Ikkinchi qismida esa klasterizatsiya algoritmlarida qo'llaniladigan masofaga asoslangan yaqinlik darajasining o'lchovi keltirilgan. Uchunchi qismida esa bu keltirilgan algoritmlarning qiyosiy tahlili aks ettirilgan.

1. Klasterizatsiya masalasini tizimli yechishning formal quyilishi

Klasterlar analizi haqidagi birinchi nashr o'tgan yuz yillikning 30-yillarida paydo bo'ldi. Ammo bu usulning faol rivojlanishi va uning keng qo'llanilishi 60-yillar oxiri, 70-yillar boshiga to'g'ri keladi. Keyinchlik bu ko'p o'lchamli analizni yo'nalish intensiv ravishda tarqaldi, yangi usullari paydo bo'ldi, ma'lum algoritmlar modifikatsiyalandi, klasterlar analizining tadbqiq etilish sohasi sezilarli darajada kengaydi. Agar dastlab ko'p o'lchamli sinflash psixalogiya, bialogiya, arxeologiya sohalarida qo'llanilgan bo'lsa, hozir esa ular sotsialogiya, iqtisod, statistika va tarixiy izlanishlarda ham faol qo'llanilmoqda. Hisoblash mashinalari paydo bo'lgandan keyin ularning qo'llanilishi alohida bir ko'rinishda kengayib bordi. Bu katta hajmdagi axborotlarni ishlash bilan bog'liq[13].

Klasterizatsiya klassifikatsiyadan farqi shundaki, o'rganiladigan analizda ajratilgan maqsad o'zgaruvchilari talab etilmaydi. Shu nuqtai nazardan u unsupervised learning sinfiga qarashli bo'ladi. Bu masalani o'rganishning birinchi bosqichida ma'lumot haqida juda kam ma'lum bo'lganda echiladi. Uning yechimi ma'lumotni yaxshiroq tushunishga yordam beradi. Bu nuqtai nazardan klasterizatsiya masalasi tavsifiy masala bo'ladi (discriptive). Klasterizatsiya bosqichlari uchun yozuvlar va o'zgaruvchilar orasidagi farq

yo'qligidir, aksincha yaqinroq guruhlar va o'xshash yozuvlar izlanadi. Avtomatik klasterlarga bo'linish usuli to'g'ridan – to'g'ri kam ishlatiladi, faqat o'xshash ob'ektlar guruhini hosil qilish uchun ishlatiladi. Klasterlarga ajratish bilan analiz boshlanadi. Klasterlarni aniqlagandan so'ng boshqa Data Mining usullaridan foydalanib, klasterlarga bo'linish nimani bildirishi, u nima bilan bog'liqligini aniqlashga harakat qiladi[13,14].

Klasterlar analizining katta ahamiyatga egaligi shundaki, u ob'ektlar bo'linishini bitta parametr bo'yicha olmaydi, balki butun belgilar majmuasini qamrab oladi. Bundan tashqari klasterlar analizi boshqa ko'pgina matematik – statistik usullardan farqli ravishda, qaralayotgan ob'ektlarga hech qanday chegaralash quyilmaydi va ma'lumotlarning boshlang'ich to'plami sifatida tabiatdagi ixtiyoriy to'plamni qarashga yo'l beradi.

Klasterlar analizi katta hajmdagi axborotlarni ko'rish va keskin qisqartirish, katta massivli axborotlarni siqish, ularni kompakt va yaqqol qilish imkoniyatini beradi.

Klasterizatsiya masalasi o'rganilayotgan ob'ektlar to'plamini klasterlar deb ataluvchi "o'xshash" ob'ektlar guruhlariga ajratishdan iborat. Klaster so'zi ingliz tilidan kelib chiqqan bo'lib (cluster), zichlik, dasta, guruh kabi tarjima qilish mumkin. Adabiyotda qo'llaniladigan o'xshash ma'nolari sinf, takson, zichlanish degan ma'nolarni beradi. Ba'zan, to'plam elementlarini klasterlarga ajratish masalasi klasterlar analizi deb ataladi. Klassefikatsiya masalasining yechimida har bir ma'lumotlar ob'ekti oldindan aniqlangan bir (yoki bir necha) sinfga oid bo'ladi va ma'lumotlar ob'ekti to'plamini sinflarga ajratish aniq hisoblarga asoslanadi. Klasterlash masalasida esa har bir ma'lumotlar ob'ektlari oldindan aniqlangan bir (yoki bir necha) sinflarga oidligi aniqlanadi. Ma'lumotlar ob'ektlarini klasterlarga ajratish ham ularni shakllantirish bilan bir vaqtda amalga oshiriladi. Klasterlarni aniqlash va ma'lumotlar ob'ektlari bo'yicha bo'linish ma'lumotlarning yakuniy modelini beradi. Bu model o'z vaqtida klasterizatsiya masalasining yechimi bo'ladi[14].

Qaralayotgan klasterizatsiya masalasining qator xususiyatlarini qaraymiz.

- Birinchidan, ob'ektlar ma'lumotlari yechmi tabiatiga (va ular atributiga) kuchli bog'liq. Demak, boshqa tamondan bu ob'ektlarning qat'iy miqdoriy qiyofasini aniqlaydi, boshqa tamondan esa ehtimollikka ega yoki noqat'iy tavsifli ob'ektlarni bildiradi.

- Ikkinchidan, yechim sinfining ifodalanishi va faraz qilingan ma'lumotlar ob'ekti munosabatiga va sinflarga ham katta bog'liq. Ob'ektlarning bir necha sinfga qarashli bo'lish imkoniyati borligi yoki imkoni yo'qligini bilish zarur. Sinfga qarashlilik xossasining o'zini ham aniqlash zarur: bir qiymatli (qarashli, qarashli emas), ehtimollik (qarashlilik ehtimoli), noqat'iy (qarashlilik darajasi).

Klasterizatsiya masalasi ma'lumotlarning intellektual analizida muhim o'rin egallab, uning yechimi uchun ko'pgina usullar ishlab chiqilgan. Ulardan biri – ma'lumotlar ob'ektining berilgan sinfga qarashli yoki qarashli emasligini ko'rsatuvchi sinflarning xarakteristik funksiyalari majmuasini qurish.

Sinflarning xarakteristik funksiyasi ikki ko'rinishda bo'lishi mumkin:

1. Ma'lumotlar ob'ektini berilgan sinfga qarashli yoki qarashli emasligi ma'nolariga teng kuchli aniq ikki qiymatdan birini qabul qiladigan diskret funksiya.

2. $0...1$ intervaldagi haqiqiy qiymatlar qabul qiladigan funksiya. Funksiya qiymati 1 ga qancha yaqin bo'lsa, ma'lumotlar ob'ekti berilgan sinfga shuncha ko'p qarashli bo'ladi.

Klasterizatsiya masalasi yechimiga umumiy yo'nalish L. Zadening noqat'iy to'plamlar nazariyasi rivojidan kelib chiqqan. Berilgan yondashish chegarasi sifatida tushunchalarni formallashtirish mumkin. Real jarayonlardagi va ma'lumotlardagi noaniqlikni yo'qotish mumkin. Bu yondashuvning afzallik jihatlari shundaki, ma'lumotlarning anashu jarayonida inson qatnashadi, baholar va xulosalar sub'ekti bo'ladi. Noqat'iy to'plamlar nazariyasi asoschisi L. Zade aytganidek: “noqat'iylikni insonning mavjudligi kabi universal reallik deb qabul qiluvchi yangi tushunchalar va usullar kompleksi bo'yicha yangi nuqtai nazar zarur edi”[14].

Klasterizatsiya masalasini yechishda noqat'iy to'plamlar nazariyasini

qo'llash bilan masalani ijobiy hal qiluvchi turli usullarni hosil qilish mumkin. Noqat'iylikni xuddi ma'lumotlarning ifodalanishi va ularning o'zaro aloqasini yozish kabi o'rganish mumkin. Bundan tashqari ma'lumotlar miqdoriy tabiatga ega bo'lishi ham, bo'lmasligi ham mumkin. Bundan tashqari ko'pgina tajribaviy masalalarni o'rganishni talab qiladi va inson tamonidan to'plangan tajribalarga asoslagan bo'lib, ko'p hollarda miqdoriy ifodaga ega bo'ladi. O'rganilayotgan ma'lumotlarning noqat'iyligini umumiy holda hisoblash juda muammo. Shuning uchun maxsus algoritmlar va yondashuvlar mavjud bo'lib, boshlang'ich ma'lumotlarning noqat'iy bo'linishiga yo'l qo'ymaydi. Ma'lumotlar qat'iy va miqdoriy deb hisoblanadi.

Noqat'iy ma'lumotlarning o'zaro aloqasini turli usullarda tasniflash mumkin. Xuddi shunday usullardan biri, noqat'iy ma'lumotlarni klasterlash algoritmlarini qo'llashda keng tarqalganlaridan biri algoritmlarning o'zaro aloqasini klasterizatsiyalash markazi va munosabatlar orqali tavsiflanishidir[15].

Masalaning formal quyilishi. Ma'lumotlar to'plamining quyidagi xossalari berilgan.

- ma'lumotlarning har bir nusxasi qat'iy sonli qiymat bilan ifodalanadi.
- har bir aniq ma'lumotning nusxasi uchun sinf aniq emas.

Quyidagilarni topish kerak.

- ma'lumatlarni o'zaro taqqoslash usuli (o'xshashlik darajasi).
- klasterizatsiya usuli.
- ma'lumotlarning klaster bo'yicha ajratish.

Formal klasterizatsiya masalasi quyidagi ko'rinishda ifodalanadi.

Har bir atributlar majmuasi bilan ifodalangan ma'lumotlar ob'ektlar to'plami I berilgan. Klasterlar to'plami C va I to'plamni C da akslantiruvchi F akslantirish $F: I \rightarrow C$ ni qurish talab qilinadi. F akslantirish masala yechimi bo'lgan ma'lumotlar modelini beradi. Masala yachimi sifatida to'g'ri sinflangan ma'lumotlar ob'ektlar soni aniqlanadi.

I to'plamni quyidagicha aniqlaymiz:

$$I = \{i_1, i_2, \dots, i_j, \dots, i_n\},$$

bu yerda i – o'rganilayotgan ob'ektlar.

Bunday to'plamga misol iris haqidagi ma'lumotlar majmuasini keltirish mumkin. O'tgan asrning 30-yillari o'rtalarida mashxur statist R.A. Fisher iris haqidagi ma'lumotlar ustida ishlagan (bu ma'lumotlar ba'zan Fisher irisleri deb ataladi). U uch sinf irislarini o'rgangan Iris setosa, Iris versicolor va Iris virginika. Ularning har biri uchun to'rt parameter: kosachabargning uzunligi va kengligi, gulbargning uzunligi va kengligi bo'yicha har xil qiymat 50 nusxada taqdim etildi. Har bir sinf uchun 5 nusxa bo'yicha ma'lumotlar taqdim etildi.

Ob'ektlarning har bir parametrlar majmuasi bilan xarakterlanadi.

$i_j = x_1, x_2, \dots, x_\eta, \dots, x_m$ Irislar qatnashgan misollarda bunday parametrlar sifatida Iris kosachabargning uzunligi va kengligi Iris gulbargning uzunligi va kengligi olingan. x_η o'zgaruvchi bir necha to'plamlardan qiymat qabul qilishi mumkin.

$$x_\eta = v_\eta^1, v_\eta^2, \dots$$

Qaralgan misol uchun qiymatlar haqiqiy bo'ladi.

Klasterizatsiya masalasi quyidagi to'plamni qurishdan iborat:

$$C = \{c_1, c_2, \dots, c_k, \dots, c_g\}$$

Bu yerda c_k – I to'plamning bir – biriga o'xshash ob'ektlarni saqlovchi klaster

$$C_k = \{i_j, i_p \mid i_j \in I, i_p \in I \text{ va } d(i_j, i_p) < \sigma\}.$$

Bu yerda σ yaqinlik darajasini aniqlovchi kattalik bo'lib, ob'ektlarning bir klasterga qarashligini bildiradi. $d(i_j, i_p)$ – ob'ektlar orasidagi yaqinlik darajasini bildiruvchi masofa deb ataladi.

$d(i_j, i_p)$ – manfiy bo'lmagan qiymat quyidagi shartlarni qanoatlantirsa, elementlar orasidagi masofa deyiladi:

$$a) \quad d(i_j, i_p) \geq 0, \text{ barcha } i_j \text{ va } i_p \text{ lar uchun;}$$

- b) $d(i_j, i_p) = 0$, faqat va faqat $i_j = i_p$ bo'lganda;
- c) $d(i_j, i_p) = d(i_p, i_j)$;
- d) $d(i_j, i_p) = d(i_j, i_r) + d(i_r, i_p)$;

Agar $d(i_j, i_p)$ – masofa σ qiymatdan kichik bo'lsa, u holda elementlar yaqin va bir klasterda joylashgan deyiladi. Aksincha bo'lganda esa elementlar bir biridan farqli va ular turli klasterlarda joylashadi.

Klasterizatsiya masalasini yechish uchun keng tarqalgan algoritmlarning ko'pchiligida kirish ma'lumotlari formati sifatida D farq matritsasi ishlatiladi. Matritsaning satr va ustun elementlari I to'plam elementlariga mos bo'ladi. Matritsaning elementi sifatida j qatordagi va p ustundagi $d(i_j, i_p)$ qiymatlarni olamiz. Bundan esa bosh diogonal elementlari nolga tengligi kelib chiqadi:

$$D = \begin{pmatrix} 0 & d(e_1, e_2) & \dots & d(e_1, e_n) \\ d(e_2, e_1) & 0 & \dots & d(e_2, e_n) \\ d(e_n, e_1) & d(e_n, e_2) & \dots & 0 \end{pmatrix}$$

Ko'pgina algoritmlar simmetrik matritsalar bilan ishlaydi. Agarda matritsa simmetrik bo'lmasa,

$$(D + D^m) / 2$$

almashtirish yordamida simmetrik ko'rinishga keltiriladi[14].

Aglomerativ algoritmlar

Birinchi qadamda butun I to'plam klasterlar to'plamidek ifodalanadi.

$$c_1 = \{i_1\}, \quad c_2 = \{i_2\}, \quad \dots, \quad c_m = \{i_m\}$$

Keyingi qadamda bir-biriga ancha yaqin bo'lgan ikkitasi olinadi va bitta umumiy klasterga birlashtiriladi {masalan C_p va C_q }. Yangi to'plam $m-1$ ta klasterdan iborat bo'ladi,

$$c_1 = \{i_1\}, \quad c_2 = \{i_2\}, \quad \dots, \quad c_p = \{i_p, i_q\}, \quad \dots, \quad c_m = \{i_m\}.$$

Jarayonni qaytarib klasterlarni $(m-2)$, $(m-3)$, $(m-4)$ va x , y ta o'lchovli to'plamlari ketma-ketligini hosil qilamiz.

Jarayon oxirida m ta obektdan tashkil topgan va boshlang'ich I to'plam

bilan mos keluvchi klasterlarni hosil qilamiz.

Klasterlar orasidagi masofani topish uchun turli usullardan foydalanish mumkin.

Ularga bog'liq holda turli xossaligi algoritmlarni olamiz. Birlashtirilgan klasterlar uchun masofani eski qiymatlaridan foydalanib hisoblaydigan bir necha usullar mavjud bo'lib, quyidagi formuladagi koeffitsientlardan farqlanadi

$$d_{rs} = \alpha_p d_{ps} + \alpha_q d_{qs} + \beta d_{pq} + \gamma |d_{ps} - d_{qs}|, \quad (1).$$

Agar p va q klasterlar r klasterga birlashtarilsa va yangi klasterdan S klastergacha bo'lgan masofa hisoblanishi talab qilinsa, qaysi usulni qo'llashimiz klasterlar orasidagi masofani topish usuliga bog'liq bo'lib, bu usullar α_p , α_q , β va γ larni qiymatlari bilan farqlanadi[16].

Divizim algoritmlar

Divizim klasterlash algoritmidagi aglomerativ algoritmdan farqli birinchi qadamda butun I to'plamning barcha elementlari yagona klaster kabi ifodalanadi. Algoritmlarning har bir qadamida mavjud bo'lgan klasterlardan biri rekursiv ravishda ikkiga bo'linadi. Xuddi shu usulda klasterlar balanddan pastga iteratsion ifodalanadi. Bu usul klasterlar analizi adabiyotlarida aglomerativ algoritmlari kabi to'la tavsiflamaydi. Bu usulni qo'llash zarur bo'lganda obektlar to'plami I ga nisbatan kam sonli klasterlarga ajratiladi.

Birinchi divizim algoritmlardan biri 1965- yili Smit Makneomo tomonidan taklif qilingan.

Birinchi qadamda barcha elementlar birgina klasterga joylanadi $C_1 = I$.

Keyin bu klasterdagi elementlardan boshqa elementlarga nisbatan masofani o'rtacha qiymatining kattasi tanlab olinadi. O'rtacha qiymatlarni misol uchun quyidagi formula bilan hisoblash mumkin:

$$D_{C_1} = 1/N_{C_1} \times \sum \sum d(i_p, i_q), \quad \forall \quad i_p, i_q \in C, \quad (2).$$

C_1 klasterdan tanlangan element o'chiriladi va C_2 klasterning birinchi elementi sifatida olinadi.

Keyingi har bir qadamda C_1 klaster elementlaridan C_2 klasterdagi

elementlarigacha bo'lgan o'rtacha masofa va C_1 klasterdagi qolgan elementlargacha bo'lgan o'rtacha masofalar orasidagi farq katta bo'ladigan element C_2 ga o'tkaziladi.

Elementlarni C_1 dan C_2 ga o'tkazish o'rtacha masofalar orasidagi farq manfiy bo'lguncha davom etadi, ya'ni C_1 da C_1 dagi elementlarga qaraganda C_2 dagi elementlarga yaqinroq bo'ladigan elementlar qolmaguncha davom etadi. Natijada bitta klaster ikkitaga bo'linadi va ulardan biri ierarxiyaning keying darajasida birlashtiriladi. Keyingi har bir darajada klasterlardan birining bo'linishida tadbqiqilinadi.

Birlashgan klasterlarni tanlash turlicha bajariladi. 1990-yil Kaufman va Rouzev birlashtirilgan klasterlardan diametri kattasini olishni taklif qildi. Diametr quyidagi formula biln topiladi:

$$D_C = \max(l_p, l_q), \quad \forall \quad l_p, l_q \in C \quad (3).$$

Klasterlarning rekursiv bo'linishi yoki barcha klasterlar bir ob'ektdan tashkil topgan bo'lguncha yoki klasterning barcha a'zolari bir biridan farqi nol bo'lguncha davom etadi.[68]

Noierarxik algoritmlar

Klasterizatsiya masalasini yechishda masalaning ma'nosidan kelib chiqqan holda bo'linish hosil qiluvchi algoritmlardan foydalaniladi. Berilgan algoritmlar ma'lumotlarni shunday guruhlashga harakat qiladiki, bo'lish algoritmining maqsad funksiyasi ekstremumga (minimumga) erishsin.

Klasterizatsiyaning bo'lishlar usuliga asoslangan uchta asosiy algoritmini qaraymiz. Berilgan algoritmlarni tahlil qilishda quyidagi bazaviy tushunchalardan foydalanamiz[17].

- M o'rganilayotgan bo'linishlar to'plami;
- Quyidagi formula masofa matritsasi formulasi deyiladi:

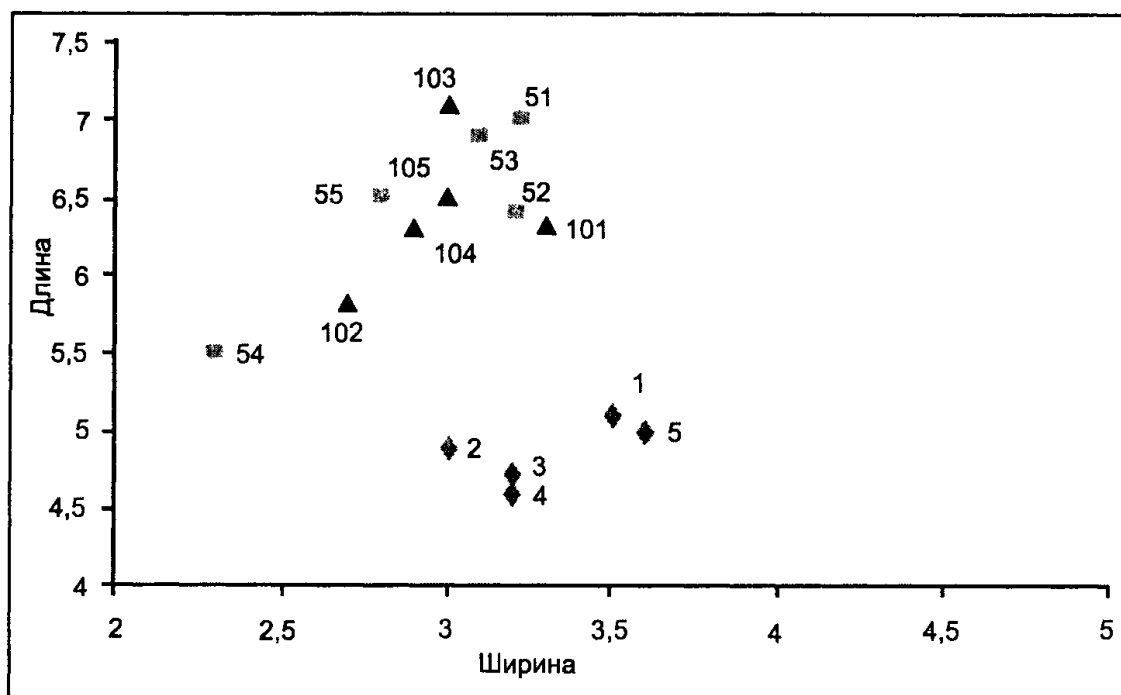
$$d_A^2(m_j, c^{(i)}) = \|m_j, c^{(i)}\|_A^2 = (m_j, c^{(i)})^i A (m_j, c^{(i)})^{\sim} \quad (4)$$

bu yerda A matritsa – masofani hisoblash usulini aniqlaydi. Masalan, birlik matritsa bo'yicha Evklad masofasidan foydalanish.

- C – klasterlar markazlari vektori;
- U – klasterizatsiya bo'yicha bo'linish matritsasi;
- $J = J(M, d, C, U)$ - maqsad funksiyasi;
- Chegaralanishlar majmuasi[18].

k – means algoritmi (K-means)

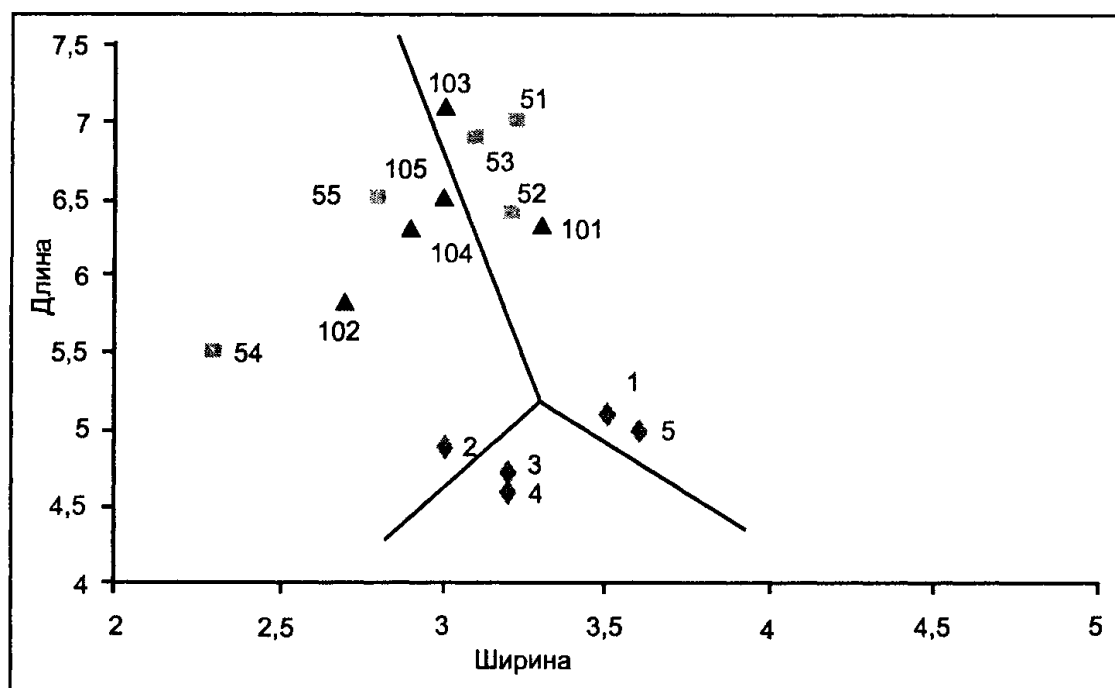
Yaxshiroq ko'rish uchun ikkita parameter – kosacha bargning uzunligi va kengligi bilan chegaralanamiz. Bu ma'lumotni ikki o'lchamli fazoda ifodalashga imkon beradi. Nuqtalar ob'ektlar nomeri bilan belgilangan. Dastlab, barcha ob'ektlar fazosidagi nuqtalar markazlardan ixtiyoriy k boshlang'ich nuqtalar olinadi. Bu markazlarni bo'lish usuli muhim emas, boshlang'ich nuqtalarni tanlash hisoblash vaqtida namoyon bo'ladi. Masalan, I to'plamdagi dastlabki k ta ob'ekt bo'linishi mumkin. Berilgan misolda bu nuqta 1,2 va 3. Ikkinchi qadamda jarayon iteratsion bajariladi.



2.1- rasm. Ikki o'lchamli fazoda irislarni tasvirlash.

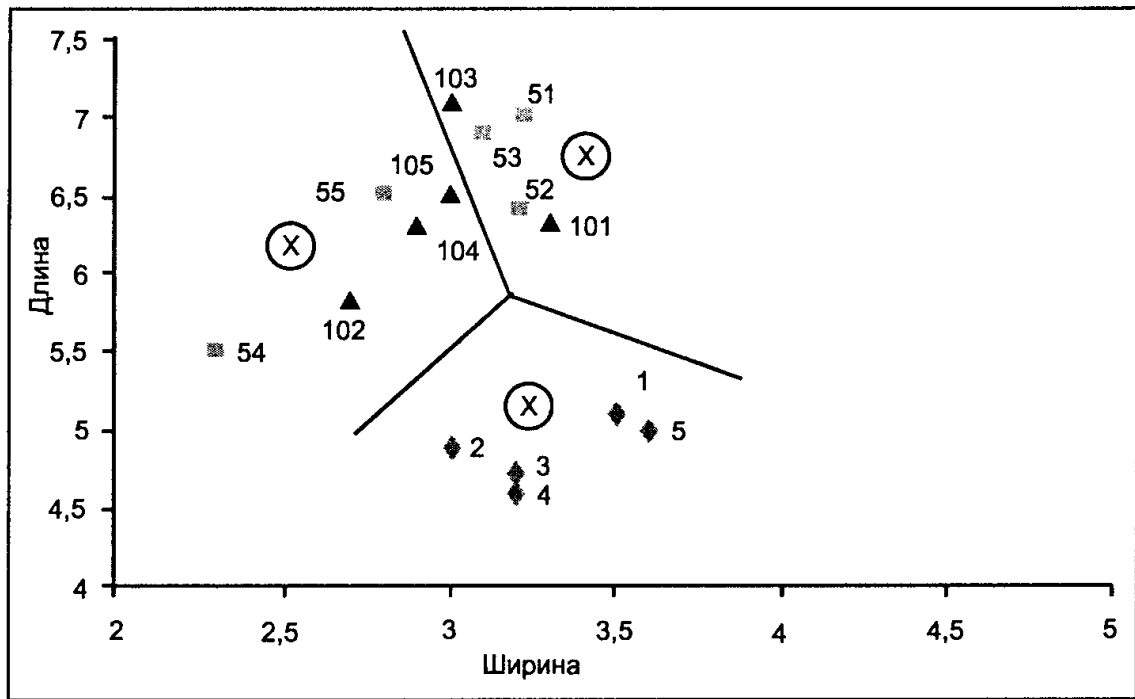
Birinchi qadamda hamma ob'ektlarning biror markazga yaqinroq joylashganini hisobga olib, k ta guruhga ajratiladi. Yaqinlik, berilgan usullarni

bir biri biln masofa orqali hisoblanadi (masalan, Evklid bo'yicha masofa olinadi). 4-rasmda irislarining uch guruhga ajratilishi ko'rsatilgan.



2.2-rasm. Irislarining uch klasterga boshlang'ich bo'linishi.

Ikkinchi qadamda klasterlarning yangi markazlari hisoblanadi. Markazlarni guruh shakllantirgan ob'ektlarning o'zgaruvchilarning o'rtacha qiymatini hisoblash kabi hisoblash mumkin. Tabiiyki, yangi markazlar oldingilaridan farq qiladi. 5-rasmda yangi markazlar va ularga mos yangi bo'linishlar ko'rsatilgan. Tabiiyki, avvalgi klasterga tegishli bo'lgan ba'zi bir nuqtalar yangi bo'linishda boshqasiga tushadi (berilgan holatda 4,2,5 va boshqalar shunaqa nuqtalar bo'ladi). Rasmda yangi markazlar aylanaga olingan "x" belgi bilan berilgan.



2.3-rasm. Irislar to'plamining klasterga ikkinchi bo'linishi.

Qaralayotgan operatsiyalar klasterlar markazlari (mos holda ular orasidagi chegara) o'zgarishdan to'xtamaguncha davom etadi. Bizning misolda ikkinchi bo'linish oxirigisi bo'ladi. Agar klasterizatsiya natijasiga qarasak, quyi klasterlar to'lig'icha Iris Setos sinfiga to'g'ri keladi. Qolgan boshqa klasterlarda ikki sinfnig ham vakillari bor. Yuqorida berilgan holda klasterizatsiya to'rtta emas, balki ikkita parameter bo'yicha amalga oshirilganligi sababli ro'u berdi. Ob'ektlar ko'p o'lchovli fazodagi nuqtalar kabi aks ettirilsa, klasterlarga avtomatik bo'lish amali ancha osonlashadi[20].

Muammo shundaki, kirish ob'ektlarini har doim ham nuqtalar ko'rinishida ifodalab bo'lmaydi. Geometriyada hamma o'zgaruvchilar har xil qiymatli bo'ladi, real holatda berilgan o'zgarishlarda o'zgaruvchilardan biri boshqalariga nisbatan sezilarli darajada katta o'zgarishi mumkin. Haqiqiy o'zgaruvchilarni taxminan har xil qiymatli masshtabda o'zgaradi, ularning tabiiy xarakterli mashtabni bo'lish orqali yoki, agar u aniq bo'lmasa, bu o'zgaruvchilarning o'rtacha qiymati bilan ularning o'zgarish diapazonida (o'zgaruvchilarning minimum va maksimum qiymatlari orasidagi farq) yoki

standart cheklanishi bolsa, u holda nuqtalar orasidagi geometrik masofa taxminan yozuvlar haqidagi intuitiv ifodalanishga mos keladi.

Berilgan algoritm noqat'iy klasterizatsiyaning tajribadagi barcha algoritmlarning asl qiyofasi bo'lib, uni o'rganish qiyinroq algoritmlarning prinsiplarini yaxshiroq tushunishga yordam beradi.

Berilgan algoritmni umumiyashtiruvchi ta'rif va tushunchalar quyidagilar:

- $M = \{m_j\}_{j=1}^d$ - o'rganilayotgan to'plam, d- ma'lumotlar vektori soni.
- (7.6) formula bilan hisoblanuvchi masofa metrikasi
- $C = \{c^{(i)}\}_{i=1}^c$ - klasterlar markazi vektori, bu yerda

$$c^{(i)} = \frac{\sum_{j=1}^d u_{ij} m_j}{\sum_{j=1}^d u_{ij}}, \quad 1 \leq i \leq c \quad (5)$$

$U = \{u_{ij}\}$ –bo'linish matritsasi, bu yerda

$$u_{ij}^{(i)} = \begin{cases} 1, & d(m_i, c_i^{(i)}) = \max_{1 \leq l \leq c} d(m_i, c_l^{(i)}) \\ 0, & \text{qolgan hollarda} \end{cases} \quad (6)$$

- Maqsad funksiyasi:

$$J(M, U, C) = \sum_{i=1}^c \sum_{j=1}^d u_{ij} d_A^2(m_i, c_i) \quad (7)$$

- Chegaralanishlar majmuasi:

$$u_{ij} \in \{0,1\}; \quad \sum_{i=1}^c u_{ij} = 1, \quad 0 < \sum_{j=1}^d u_{ij} < d. \quad (8)$$

Bu chegaralanishlar majmuasi har bir ma'lumotlar vektori faqat bitta sinfga qarashli ekanligini, qolganlari esa qarashli emasligini bildiradi.

Har bir klaster bittadan kam bo'lmagan, lekin barcha nuqtalar sonidan kam nuqtani o'zida saqlaydi[20].

Konstruktiv algoritm o'zi bilan quyidagi iteratsion protsedurani ifodalaydi:

1-Qadam. Boshlang'ich bo'linishni initsializatsiya qilish (masalan,

tasodifiy holatda),

δ aniqlikni tanlash (algoritmning tugash shartidan foydalangan holda), iteratsiya nomeri $l = 0$ ni initsializatsiya qilish.

2-Qadam. Quyidagi formula yordamida klasterlar markazini aniqlash:

$$c_l^{(i)} = \frac{\sum_{j=1}^d u_{ij}^{(l-1)} m_j}{\sum_{j=1}^d u_{ij}^{(l-1)}}, \quad 1 \leq i \leq c \quad (9)$$

3-Qadam. Quyidagi formula bo'yicha bo'linishlar matritsasini xatolar kvadrati minimumlashadigan qilib almashtirish:

$$U_{ij}^{(i)} = \begin{cases} 1, & d(m_i, c_i^{(i)}) = \max_{1 \leq l \leq c} d(m_j, c_l^{(i)}) \\ 0, & \text{qolgan hollarda} \end{cases} \quad (10)$$

4-Qadam. $\|U^{(l)} - U^{(l-1)}\| < \delta$ shartni tekshirish. Agar shart bajarilsa, jarayonni to'xtatish, aks holda iteratsiya nomerini $l = l + 1$ deb, 2- qadamga qaytadi.

Qaralgan algoritmning asosiy kamchiligi bo'linishlar matritsasi elementlarining diskret xarakter kuchida bo'lib, bo'linishlar fazasining ata o'lchamdaligidir. Bu kamchilikni to'g'rilash usullaridan biri bo'linishlar matritsasi elementlarini birlik intervaldagi sonlar bilan ifodalashdir. Ya'ni ma'lumotlar elementining berilgan klasterga qarashliligi qarashlilik funksiyasi bilan aniqlanishi kerak. Ma'lumotlar elementlari bir necha klasterga turlicha darajalar bilan qarashli bo'lishi mumkin. Bu yondashuv noqat'iy klasterizatsiya algoritmi *Fuzzy C-Means* da o'z o'rnini topgan[19,20].

Fuzzy C-Means algoritmi

Bu algoritm avval o'tilgan algoritmning umumlashmasidir. Bu algoritmning farqi shundaki, endi klasterlar noqat'iy to'plamda bo'ladi va har bir nuqta har xil klasterlarga har xil qarashlilik darajalari bilan qarashli bo'ladi. Nuqta biror bir klasterga maksimum qarashlilik kriteriyasi bo'yicha qarashli bo'ladi.

Berilgan hol uchun bazaviy tushunchalar quyidagilar bo'ladi:

- $M = \{m_j\}_{j=1}^d$ – o'rganilayotgan to'plam, d – ma'lumotlardagi nuqtalar (vektorlar) soni;

- Masofa metrikasi ((3) formula);

- $C = \{c^{(i)}\}_{i=1}^c$ – klasterlar markazi vektori,

bu yerda,

$$c^{(i)} = \frac{\sum_{j=1}^d (u_{ij})^w m_j}{\sum_{j=1}^d (u_{ij})^w}, \quad 1 \leq i \leq c \quad (11)$$

- $U = \{u_{ij}\}$ – bo'linoish matritsasi, bu yerda

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_A^2(m_j, c^{(i)})}{d_A^2(m_j, c^{(k)})} \right)^{\frac{1}{w-1}}}, \quad (12)$$

- Maqsad funksiyasi:

$$J(M, U, C) = \sum_{i=1}^c \sum_{j=1}^d u_{ij}^w d_A^2(m_j, c^{(i)}), \quad (13)$$

bu yerda, $w \in (1, \infty)$ noqat'iylik koeffitsenti (tortish koeffitsenti) bo'lib, noqat'iy bo'linishlarni tartibga soladi. Odatda $w=2$ deb olinadi.

- Chegaralanishlar majmuasi:

$$u_{ij} \in [0,1]; \quad \sum_{i=1}^c u_{ij} = 1, \quad 0 < \sum_{j=1}^d u_{ij} < d. \quad (14)$$

Har bir ma'lumotlar vektori turli xil klasterlarga turlicha qarashlilik darajasi bilan qarashli bo'lishini ifodalaydi[16]. Ma'lumotlar elementlarining bo'linishlari fazasida barcha klasterlarga qarashliligining yig'indisi birga teng.

Bu algoritmnining bajarilish ketma-ketlagi quyidagicha amalga oshiriladi:

1-Qadam. Klasterlar sonini tanlash: $2 \leq c \leq d$.

2-Qadam. Ma'lumotlar vektorini haqiqiy sonlar o'qida tasvirlash uchun skalyar metrikan tanlash.

3-Qadam. To'xtash qadami δ ni tanlash.

4-Qadam. Noqat'iylik koeffitsenti $w \in (1, \infty)$ ni tanlash, masalan $w=2$.

5-Qadam. Bo'linishlar matritsasini initsializatsiya qilish (masalan, tasodifiy qiymatlarni).

6-Qadam. Klasterlar markazlarini quyidagi formula yordamida hisoblash:

$$c_l^{(i)} = \frac{\sum_{j=1}^d \mu_{ij}^{(l-1)} m_j}{\sum_{j=1}^d \mu_{ij}^{(l-1)}}, \quad 1 \leq i \leq c \quad (15)$$

7-Qadam. Barcha ma'lumotlar elementlari uchun barcha klasterlargacha (markazlarigacha) bo'lgan masofa kvadratini quyidagi formula bo'yicha hisoblash:

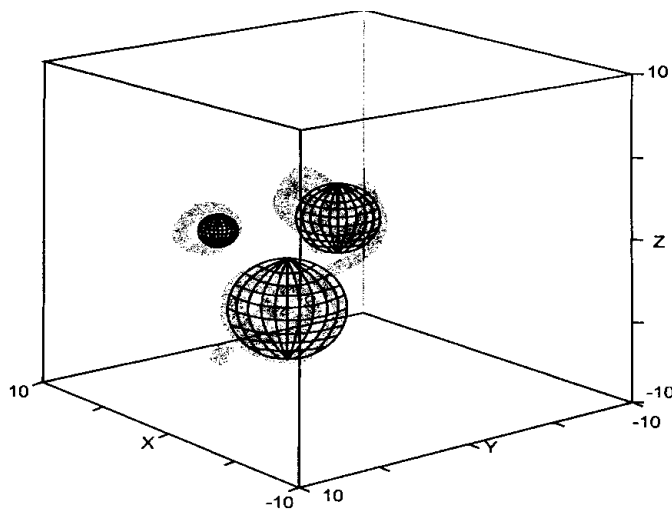
$$d_A^2(m_j, c_l^{(i)}) = (c_l^{(i)} - m_j)^T A (c_l^{(i)} - m_j), \quad (16)$$

8-Qadam. Bo'linishlar matritsasini quyidagi formula bo'yicha aniqlash:

$$u_{ij}^{(l)} = \frac{1}{\sum_{k=1}^c \left(\frac{d_A^2(m_j, c^{(i)})}{d_A^2(m_j, c^{(k)})} \right)^{\frac{1}{w-1}}}, \quad (17)$$

(16) formuladagi cheklanishlarni hisobga olib, barchasi uchun $1 \leq i \leq c$, $1 \leq j \leq d$.

9-Qadam. $\|U^{(l)} - U^{(l-1)}\| < \delta$ shartni tekshirish. Agar shart bajarilsa, jarayon to'xtatiladi, aks holda iteratsiya nomerini $l = l + 1$ deb, 7-qadamga o'tish.

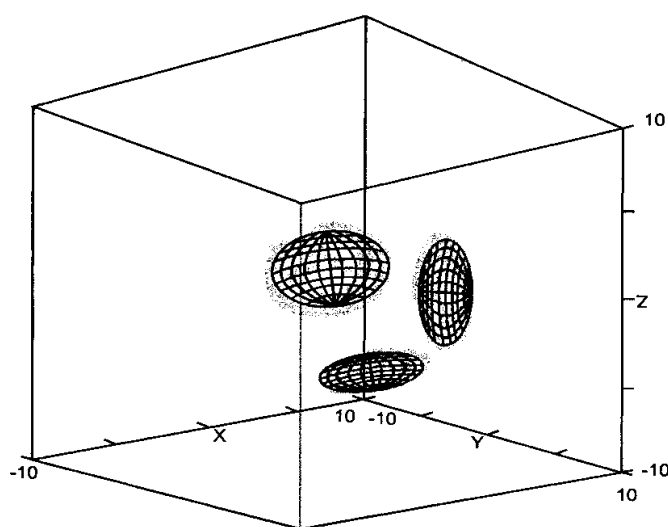


2.4-rasm. Fuzzy *k-means* algoritmidagi klasterlar shakli.

Berilgan algoritm k – means algoritmgiga qaraganda afzalliklarga ega, ammo shunday kamchilikka egaki, sferik formadagi klasterlarni izlaydi. Bu esa masalalarga mos kelavermayda va shuning uchun yechim tuzatib bo'lmaydigan darajaga keltiradi. Bu kamchilik quyidagi algoritmda kuzatilmaydi.

Gustafson – Kassel bo'yicha klasterizatsiyalash

Quyida beriladigan noqat'iy klasterizatsiyalash algoritmi ellipsoida (2.5-rasm) ko'rinishidagi algoritmlarni izlaydi. Bu esa turli xil masalalarni yechishda ancha egiluvchan qiladi. Algoritmnı tavsiflashdan oldin unda qo'llaniladigan qo'shimcha tushunchalarni keltiramiz[21].



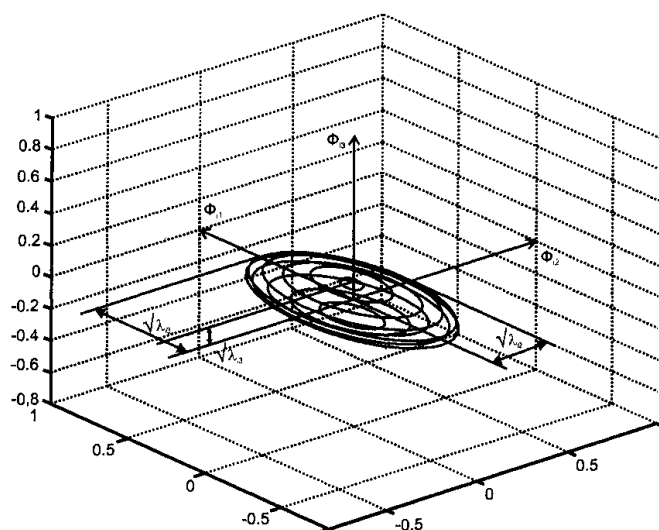
2.5-rasm. Gustafson – Kassel bo'yicha klasterizatsiyalash algoritmidagi klaster formasi.

Klaster faqat o'zining markazi bilangina emas, balki kovariatsion matritsa bilan ham xarakterlanadi.

$$F^{(i)} = \frac{\sum_{j=1}^d (u_{ij})^w (m_j - c^{(i)})(m_j - c^{(i)})^i}{\sum_{j=1}^d (u_{ij})^w}, \quad (18)$$

bu yerda, λ_{ik} berilgan $F^{(i)}$ matritsaning k – xos qiymatlarini belgilaydi, Φ_{ik} esa $F^{(i)}$ matritsaning k – birlik xos vektori. λ_{ik} ning xos qiymatlari kamayish tartibida tartiblangan. U holda $\Phi_{i1}, \dots, \Phi_{i(n-1)}$ xos vektorlar i -chi chiziqli qism fazoni o'raydi, Φ_{in} esa bu chiziqli qism fazoning normali

bo'ladi. Yuqorida aytilganlarning barchasi 2.6-rasmدا tasvirlangan.



2.6-rasm. Gustafson – Kassel bo'yicha klasterizatsiyalash algoritmining geometrik tasviri.

Bu algoritm masofani hisoblash uchun o'zining normallashtiruvchi matritsasidan foydalanadi. Bu masofani hisoblash formulasi quyidagicha bo'ladi:

$$d_{A^{(i)}}^2 = (c^{(i)} - m_j)^i (c^{(i)} - m_j), \quad (19)$$

bu yerda,

$$A^{(i)} = (| F^{(i)} |)^{\frac{1}{r+1}} \left(F^{(i)} \right)^{\frac{1}{r+1}}, \quad (20)$$

$F^{(i)}$ esa (18) formula bilan aniqlanadi.

Asosiy tushunchalarni umumlashtiramiz:

- $M = \{m_j\}_{j=1}^d$ – o'rganilayotgan to'plam, d – ma'lumotlardagi nuqtalar (vektorlar) soni;
- (19) va (20) formulalar bilan hisoblanadigan masofa metrikasi.
- $C = \{c^{(i)}\}_{i=1}^c$ – klasterlar markazi vektori,

bu yerda,

$$c^{(i)} = \frac{\sum_{j=1}^d (u_{ij})^w m_j}{\sum_{j=1}^d (u_{ij})^w}, \quad 1 \leq i \leq c \quad (21)$$

- $U = \{u_{ij}\}$ – bo'linish matritsasi, bu yerda

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{A^{(i)}}^2(m_j, c^{(i)})}{d_{A^{(i)}}^2(m_j, c^{(k)})} \right)^{\frac{1}{w-1}}}, \quad (22)$$

- Maqsad funksiyasi:

$$J(M, U, C) = \sum_{i=1}^c \sum_{j=1}^d u_{ij}^w d_{A^{(i)}}^2(m_j, c^{(i)}), \quad (23)$$

bu yerda, $w \in (1, \infty)$ noqat'iylik koeffitsenti (tortish koeffitsenti).

- Chegaralanishlar majmuasi:

$$u_{ij} \in [0,1]; \quad \sum_{i=1}^c u_{ij} = 1, \quad 0 < \sum_{j=1}^d u_{ij} < d. \quad (24)$$

Har bir ma'lumotlar vektori turli xil klasterlarga turlicha qarashlilik darajasi bilan qarashli bo'lishini ifodalaydi. Ma'lumotlar elementlarining bo'linishlari fazasida barcha klasterlarga qarashliligining yig'indisi birga teng[21,22].

Bu algoritmnining bajarilish ketma-ketlagi quyidagicha amalga oshiriladi:

1-Qadam. Klasterlar sonini aniqlash: $2 \leq c \leq d$.

2-Qadam. To'xtash qadami $\delta > 0$ ni tanlash.

3-Qadam. Noqat'iylik parametri $w \in (1, \infty)$ ni tanlash, masalan $w=2$ deb olish mumkin.

4-Qadam. Bo'linishlar matritsasini initsializatsiya qilish (masalan, tasodifiy qiymatlarni).

5-Qadam. Klasterlar markazlarini quyidagi formula yordamida hisoblash:

$$c_l^{(i)} = \frac{\sum_{j=1}^d u_{ij}^{(l-1)} m_j}{\sum_{j=1}^d u_{ij}^{(l-1)}}, \quad 1 \leq i \leq c \quad (25)$$

6-Qadam. Klasterlar kovariatsion matritsasini quyidagi formula bo'yicha hisoblash:

$$F^{(i)} = \frac{\sum_{j=1}^d (u_{ij})^w (m_j - c^{(i)})(m_j - c^{(i)})^i}{\sum_{j=1}^d (u_{ij})^w}, \quad (26)$$

7-Qadam. Quyidagi formula bo'yicha masofani aniqlash:

$$d_{F^{(i)}}^2 = (c_l^{(i)} - m_j) = (c_l^{(i)} - m_j)^i \left| F^{(i)} \right|^{\frac{1}{r+1}} (F^{(i)})^{-1} \left| c_l^{(i)} - m_j \right| \quad (27)$$

8-Qadam. (14) formuladagi cheklanishlarni hisobga olib, bo'linishlar matritsasini quyidagi formula bo'yicha aniqlash:

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{F^{(i)}}^2(m_j, c^{(i)})}{d_{F^{(i)}}^2(m_j, c^{(k)})} \right)^{\frac{1}{w-1}}}, \quad (28)$$

9-Qadam. $\|U^{(l)} - U^{(l-1)}\| < \delta$ shartni tekshirish. Agar shart bajarilsa, jarayon to'xtatiladi, aks holda iteratsiya nomerini $l = l + 1$ deb, 5-qadamga o'tish.

Xulosa sifatida aytish kerakki, berilgan algoritmlar klasterizatsiyaga yondashish jihatdan bir biridan farq qilmaydi. Bu maqsad funksiyalarini taqqoslashda aniqroq ravshanlashadi, ularni minimallashtirish berilgan algoritm asosida yotadi. Farq esa kirish ma'lumotlari fazasidagi nuqtalar orasidagi masofa turlicha usullar bilan hisoblanishidagina iborat bo'ladi. Berilgan algoritmlar qiyinlashib borish tartibida joylashgan. Har bir algoritm o'zidan oldingilaridan ko'ra ko'proq ma'lumotlarning o'zaro aloqasini, aspektlarini o'rganishga harakat qiladi[23].

Yuqorida aytilganidek, bir qancha algoritmlar mavjudki, ular bir biridan

maqsad funksiyasidagi qo'shimcha qo'shiluvchilar bilan farq qiladi. Shuni ta'kidlash kerakki, bu maqsad funksiyalarida Bezdekov ikkilangan yig'indisi Φ bilan aniqlanuvchi qo'shiluvchi o'zgartirilmaydigan qo'shiluvchi bo'lib, maqsad funksiyasining qurilishiga asos bo'ladigan asosiy faraz o'gartirilmasligidan dalolat beradi.

Bu farazlardan asosiylari quyidagi ko'rinishda bo'ladi:

- Umumiy holda klasterlar ellipsoida shaklida bo'ladi;
- Klasterda doimo markaz mavjud bo'ladi ;
- Nuqtaning klasterga qarashliligi, nuqtadan klasterlar markazlarigacha bo'lgan bir necha masofalarga asoslanadi;

Bu uchta qismni o'zi berilgan algoritmlar kamchiliklarini aniqlash uchun yetarli bo'ladi.

- Faraz qilinadiki, barcha klasterlar har doim bir necha aniqlovchi algoritm shakllariga ega bo'ladi, shunisi aniqki, bu har doim ham bajarilavermaydi. Ma'lumotlarda bir necha figuralarda berilgan kirish ma'lumotlar fazasining aproksimatsiyasi interpretatsiyalanmagan yechimlarga olib keladi.

- Klasterda doimo bir nechta tugun nuqtalar (klaster markazlari) mavjud bo'lib, uning klasterga qarashlilik darajasi birga teng, bir vaqtda boshqa nuqtalar kabi klasterga bunday yuqori qarashlilik darajasi bilan qarashli bo'lmaydi va yana nuqtalarining o'zaro qiyin joylanishi o'lchamsiz bo'ladi.

- Berilgan algoritmlar nuqtalarning o'zaro jaylashishiga asoslanmaydi, balki nuqtalarning klasterlar markazlariga bo'lgan munosabatiga asoslanadi.

Bunaqa klasterlash algoritmlarining kuchsiz tamoni shundaki, kirish ma'lumotlari ikkita doiraga joylashgan shaklda bo'ladi. Fuzzy C-Means algoritmi sferik klasterlar quradi, lekin hech qanday shartda bu sferalarni saqlovchi ma'lumotlar fazasini ikkita klasterga ajratmaydi[22,23].

2. Klasterizatsiya algoritmlarida qo'llaniladigan masofaga asoslangan yaqinlik darajasining o'lchovi

Ob'ektlar orasidagi masofa R^m o'chovli fazodagi nuqtalar ko'rinishida ifodalanishini faraz qiladi. Bu holda masofani hisoblash uchun turli usullardan foydalanish mumkin. Quyida qaraladigan daraja kirish o'zgaruvchilari fazosidagi ikki nuqta orasidagi masofani aniqlaydi.

Quyidagi belgilashlardan foydalanamiz:

- $X_Q \subseteq R^m$ – ma'lumotlar to'plami bo'lib, m – o'lchovli haqiqiy fazoning qism to'plami.

- $x_i = (x_{i_1}, \dots, x_{i_m}) \in X_Q$, $i = \overline{1, Q}$ – ma'lumotlar to'plamining elementlari.

- $S = \frac{1}{Q} \sum_{i=1}^Q (x_i - \bar{x})(x_i - \bar{x})^i$ $m \times m$ o'lchovli kovariatsion matritsa.

Endi bizga ancha ma'lum bo'lgan yaqinlik darajalari (o'chovlari, miqdorlari) ni o'rganamiz.

Evklid masofasi. Ba'zida bir biridan ancha uzoqlikdagi ob'ektlarga ko'proq kuch quyish uchun standart evklid masofani kvadratga ko'tarish ehtiyoji paydo bo'ladi. Bu masofa quyidagi ko'rinishda hisoblanadi:

$$d_2(x_i, x_j) = \sqrt{\sum_{i=1}^m (x_{ii} - x_{ji})}.$$

Xemmingu masofasi. Bu yerda oddiy koordinatalar bo'yicha o'rtacha farq masofa bo'ladi. Ko'pgina hollarda berilgan masofa o'lchovi (darajasi) beradigan natijalar Evklid masofasi o'lchovi natijalari bilan bir xil bo'ladi. Bu holda farqlar kamayadi. Xemmingu masofasi quyidagi formula bilan hisoblanadi:

$$d_L(x_i, x_j) = \frac{1}{m} \sum_{i=1}^m \frac{|x_{ii} - x_{ji}|}{x_{ii} + x_{ji}}.$$

Agar ma'lumotlarning klasterizatsiyalashga ta'sir etuvchi xarakterlari haqidagi axborotga ega bo'lsak, u holda qaralgan masofalar darajalaridan

ixtiyoriysini ishonch bilan olish mumkin. Xuddi shunday cho'qqili masofada tasodifiy o'zgaruvchilar orasidagi erkinlik so'ralishi kabi. Ammo tajribada qo'llashda bu o'zgaruvchilar mustaqil bo'la olmaydi[21,22].

3. Klasterizatsiya masalalari usul va algoritmlarining qiyosiy tahlili

Klasterizatsiyani bajarish natijasida nechta klaster qurilishi lozimligini bilish muhimdir. Klasterizatsiyada ob'ektlarning tabiiy lokal zichligini aniqlashtirish kerak deb faraz qilinadi. Shuning uchun klasterlar soni noaniq bo'ganda algoritmlarning ko'rinishini etarlicha qiyinlashtiruvchi, aniq bo'lganda esa yechim sifatiga kuchli ta'sir o'tkazuvchi parameter bo'ladi. Klasterlar sonini tanlash muammosi trivial emas. Ba'zan, qanoatlantiruvchi nazariy yechimni olish uchun oldindan berilgan bir necha taqsimlash xossalari haqida kuchli faraz qilishni talab qiladi. Ammo, ayniqsa izlanishning boshida ma'lumotlar haqida hech narsa aniq bo'lmasa, qanday faraz haqida gap borishi mumkin. Shuning uchun klasterizatsiya algoritmlari odatda klasterlar sonini tanlashning ba'zi usullaridek va uning optimal qiymatini tanlash jarayonida aniqlash kabi quriladi. To'plamni klasterlarga ajratish usularining soni katta. Ularning barchasini ierarxiklik va noierarxiklikka bo'lish mumkin. Noierarxik algoritmlarda, ularning ishlarida va to'xtalish shartlarida oldindan reglamentlash zarur. Ba'zan parametrlar soni etarlicha kattaligi boshlang'ich bosqichlarda materialni o'rganishni qiyinlashtiradi. Lekin bunday algoritmlarda klasterizatsiyani variatsiyalashda katta egiluvchanlikka erishiladi va odatda klasterlar soni aniqlanadi.

Boshqa tamondan, ob'ekt qachon ko'p sonli parametrlar bilan xarakterlansa, u holda alomatlarni guruhlash muhim ahamiyatga ega bo'ladi. Boshlang'ich axborotlarga bog'liq kvadrat matritsada, xususiyl holda korrelatsion matritsa saqlanadi. Guruhlash masalasining asosiy muvofaqqiyatli yechimi – yashirin faktorlarning katta bo'lmagan soni haqidagi formal bo'lmagan gipotezasi bo'lib, alomatlar orasidagi o'zaro aloqaning tuzilishini aniqlaydi.

Ierarxik algoritmlarda klasterlardan to'liq daraxt qurib, klasterlar sonini aniqlashni asosli ravishda inkor etadi. Farazdan relslar soni algoritm ishiga bog'liq bo'lmaslik prinsipida aniqlanadi. Misol uchun dinamika bo'yicha klasterlar ostonasini birlashishini o'zgarishi. Bunday algoritmlarning murakkabligi yaxshi o'rganilgan. Klasterlarning yaxlit darajasini tanlash dendrogrammada indeksleri inversiya muammosi ierarxik sinflashni egiluvchan emasligi, bu ko'p xollarda ko'ngilli emas.

Bundan tashqari klasterlashning dendrogramma ko'rinishida ifodalashni klasterlar tuzish haqida to'liqroq ta'surot olishga ijozat beradi.

Ierarxik algoritmlar dendrogrammalar ko'rinishi bilan bo'g'liq bo'ladi va quyidagilarga bo'linadi:

A) Algomerativ, boshlangich elementlarni klasterlar soni kamayib borishiga mos xolda ketma-ket birlashishning (klasterlarni pastdan yuqoriga qarab qurilishi)

B) Divizim (bo'linuvchi), klasterlar soni bittadan boshlab o'suvchi va natijada guruxlarni birlashtiruvchi ketma-ketlik hosil qiladi (balanddan pastga qarab klasterlar qurish)[24].

Ushbu 10 ta algoritmlarning qiyosiy tahlili keltirilgan. Bunda algoritmlarning CURE, BICH, CLARA, MST, k-means, PAM, CLOPE, Koxonena, Hard C – Means, Fuzzy C-means lar ko'rib qiyosiy tahlil qilindi. Qiyosiy tahlilda algoritmlarning sinfi, yutig'lik tarafi, kamchiligi, qanday turdagi ma'lumotlar bilan ishlash mumkinligi va hamda ishlash tezliklari aniqlandi(2-jadval). 3-jadvalda ushbu algoritmlarning ishlash vaqti (sekundlarda) ning ularga kirishdagi tanlov elementlari soniga bog'liqligi keltirilgan.

2-jadval. Klasterizatsiya usullarining qiyosiy tahlili

Usul	Sinfi	Yutug'i	Kamchiligi	Ma'lumot	Tezligi
CURE	Iyerarxik (aglomerativ)	Turli o'lchamli va murakkab shakldagi klasterlarni quradi	Porog qiymati va klasterlar sonini talab qiladi	sonli	$O(n^2)$
BIRCH	Iyerarxik (divizim)	Klasterlash tezligi baland, Ikki etapli	2-etapda boshqa algoritm ishlatiladi	sonli	$O(n^2)$
CLARA (Kafman-Rassel algoritmi)	Iterativ	Katta Mbda ishlaydi, yaxshi klasterlaydi	Obraz bo'yicha klasterlaydi va uning sifatiga bog'liq	sonli	$O(\max(n, m)), m < n(n-1)/2$
MST	Iyerarxik (divizim)	Qavariq va botiq klasterlar quradi.	Klaster sonini talab qilad	ixtiyoriy	$O(E \log V)$
k-means	Iterativ	Sodda va tez ishlaydi,	Tez xatoga yo'l qo'yadi	Kichik xajmli	$O(nkl)$, k – klaster soni, l – iteratsiya soni
PAM	Iterativ	Sodda va tez ishlaydi, xatolikka sezgirliги k-means dan kamroq	klasterlar sonini talab kiladi, katta Mbda sekin ishlaydi	sonli	$O(nkl)$, k – klaster soni, l – iteratsiya soni
CLOPE	Iterativ	Tez hisob-laydi va interpreta-siyalanadi, klasterlar sonini avtomatik tanlaydi	Tez xatoga yo'l qo'yadi	kategoriyali	$O(nkl)$, k – klaster soni, l – iteratsiya soni
Koxonena	Notiniq	O'zi tashkil qilinadagan neyrotarmoq, tadbiqu oson, tarmoqni o'rganish o'qituvchisiz, kafolatli	Faqat sonlar bilan ishlaydi, tarmoq o'lchami minimal, klasterlar sonini talab qiladi	sonli	$O(n^2 \log n)$
Hard C – Means	Iterativ	Oson qo'llaniladi, hisoblash sodda	klaster sonini talab qiladi va optimal yechim kafolatlanmagan	sonli	$O(nkl)$, k – klaster soni, l – iteratsiya soni
Fuzzy C-means	Notiniq	Klaster chegarasidagi obyektlarni aniqlaydi	Murakkab xisoblashlar, klaster sonini talab qiladi	sonli	$O(nkl)$, k – klaster soni, l – iteratsiya soni

3-jadval. Algoritmning ishlash vaqtini (sekundlarda) kiruvchi elementlar soniga bog'liqligi

Algoritm turi	500 Elementlar	1000 Elementlar	1500 Elementlar	2000 Elementlar	2500 Elementlar
CLARA (Kafman-Rassel algoritmi)	5,459	46,703	180,820	385,016	777,763
MST	0,344	2,832	12,212	23,095	55,579
k-means	0,020	0,084	0,221	0,375	0,596
PAM	0,357	2,837	10,782	22,940	49,787
CLOPE	0,339	2,813	10,054	22,505	43,669
Koxonena	0,347	2,828	10,789	23,195	55,752
Hard C – Means	0,017	0,081	0,217	0,363	0,580
Fuzzy C- means	0,036	0,162	0,483	0,717	1,352

II bob bo'yicha xulosalar

Ushbu bobda ma'lumotlarni intellektual tahlil etish jarayonida klasterizatsiya masalasi algoritmlari aks ettirilgan bo'lib, paragrafning birinchi qismida klasterizatsiya masalasini tizimli yechishning formal qo'yilishi va algoritmlar tahlillari to'g'risida ma'lumotlar berilgan va ular qarab o'tilgan. Bu paragrafda klasterizatsiya usuli algoritmlari qadamlari oydin yoritib berilgan. Ushbu algoritmlar Aglomerativ algoritmlar, Divizim algoritmlari, noierarxik algoritmlar, k-means, Fuzzy C-means, Gustafson-Kassel bo'yicha klasterlash kibi algoritmlar o'rganildi va ushbu algoritmlar asosida II bobning ikkinchi qismida esa klasterizatsiya algoritmlarida qo'llaniladigan masofaga asoslangan yaqinlik darajasining o'lchovi keltirildi. Uchunchi qismida esa o'rganilgan algoritmlarning qiyosiy tahlili qilindi. Tahlillar natijasida algoritmlarning bir-biridan farqli jihatlari algoritmning yutug'i, kamchigi, ma'lumotlar bilan ishlash turlari, ishlash tezligi, qulaylilik darajasi ushbu qismda jadvallar asosida keltirilgan.

III BOB. MA'LUMOTLARNI INTELLEKTUAL TAHLILIDA KLASTERIZATSIYA ALGORITMLARI DASTURIY TA'MINOTLARINING AMALIY TADBIQLARI

1. Klasterizatsiya usullari uchun amaliy dasturiy paketlarda ishlash

Ma'lumotlar intellektual tahlili usullarining ayrimlari uchun professional statistik va matematik paketlarda turli dasturiy modullar ishlab chiqilgan. Masalan, MATLAB – matritsaviy laboratoriya paketi va qiymatlar statistik tahlil qilishga mo'ljallangan SPSS paketi, ma'lumotlarni intellektual tahlil qilish WEKE (Waikato Environment for Knowledge Analysis) dasturiy paketlarini misol qilib keltirish mumkin. Bu paketlar yordamida turli sonli va analitik (simvulli) ma'lumotlarni klassifikatsiya va klasterizatsiya masalalari orqali tahlil qilish, dendogrammalarni hosil qilish, vizuallashtirish imkoniyatlari mavzud.

1.1. SPSS – ijtimoiy fanlar statistik paketida klasterizatsiya moduli

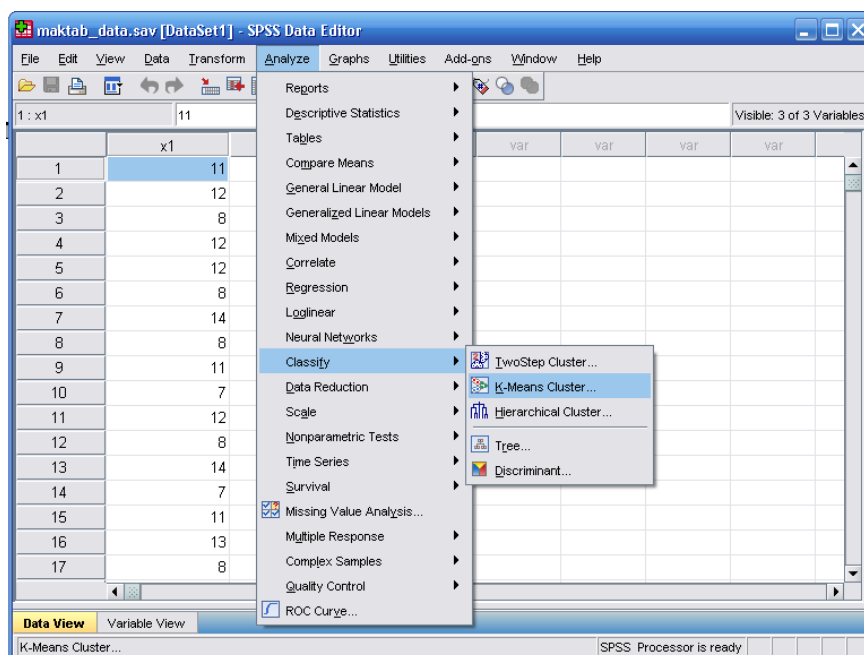
SPSS (inglizcha *«Statistical Package for the Social Sciences»* — «statistik fanlar uchun statistik paket») — ma'lumotlarni statistik qayta ishlash kompyuter dasturi bo'lib, ijtimoiy ilm-fanda amaliy tadqiqotlar o'tkazishga mo'ljallangan tijorat dasturlari bozorida peshqadamlaridan biri hisoblanadi¹. Bu dasturning birinchi versiyasi 1968 yilda Norman Nay, Xedli Xall va Deyl Bentlar (talabalar) tomonidan yaratilgan va Chikago universitetida rivojlantirilgan. 1975 yilda SPSS Inc alohida kompaniya bo'lib, 1992 yilda Microsoft Windows ostida paketning birinchi versiyasi chiqqan. Xozirgi kunda uning 10 dan ziyod paketlari Mac OS X va Linux OS lari uchun ham yaratilgan.

Ishda 2008 yil aprel oyida chiqqan SPSS16.0.1Evaluation versiyasida klasterizatsiyalash usullarini tadqiq qilindi[32].

Misol uchun, bizga tibbiyot muassasasi tomonidan yig'ilgan o'smirlar haqidagi ma'lumotlar 1-ilovadagi 1-jadval ko'rinishda berilgan bo'lsin, bu ma'lumot asosida ularni maktab sinflariga sinflashtirishni SPSS – ijtimoiy fanlar uchun statistik paketida bajarib ko'ramiz.

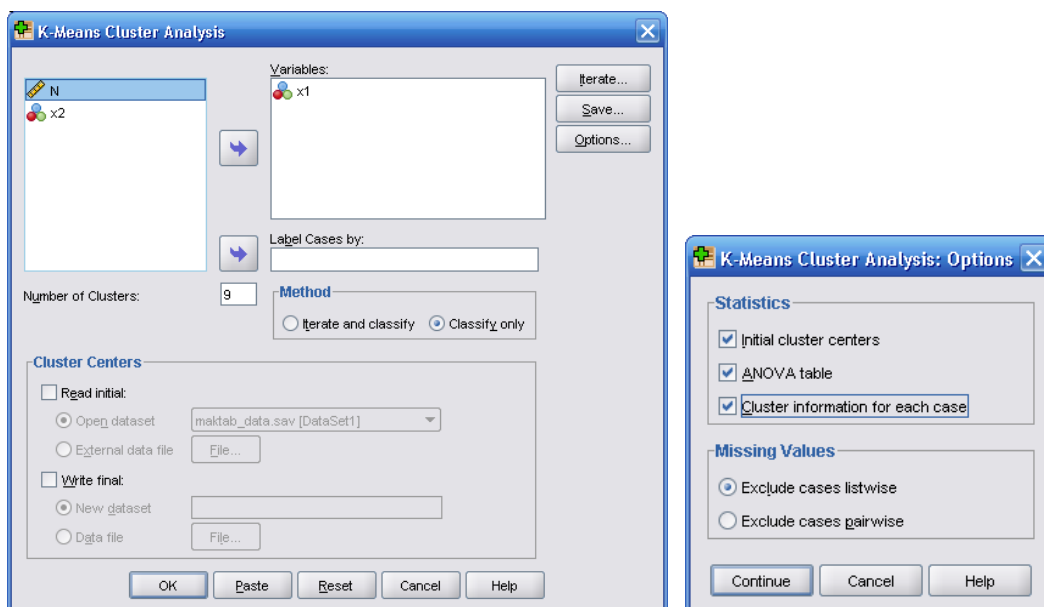
¹ <http://ru.wikipedia.org/wiki/SPSS>

Jadvaldagi x_1 – o'smirlar yoshini bildiradi, x_2 – o'smirlar jinsi 1 (o'gil bola) va 0 (qiz bola) bilan berilgan. Jami 861 ta satrni ob'jekt sifatida qaralganda, berilgan 861 ta ob'jekt 2 ta belgi bilan berilgan deyiladi.



3.1-rasm. SPSS statistik paketining ma'lumotlarni taxrirlash oynasi va k-means usulida klasterizatsilash modulini ishga tushirish

SPSS paketida bu ma'lumotlarni fayldan yuklash mumkin (3.1-rasm). Taxrirlash oynasi yoki asosiy oynaning menyusida Analyze/Classify/K-means bo'limi tanlash orqali yuklangan ma'lumotlarni k-means algoritmi asosida klasterizatsiyalash mumkin. k-means klassifikatorning parametrlar (3.2-rasm) oynasida natijalarni bayon qilish va asosiy faktorlarni tanlash mumkin.



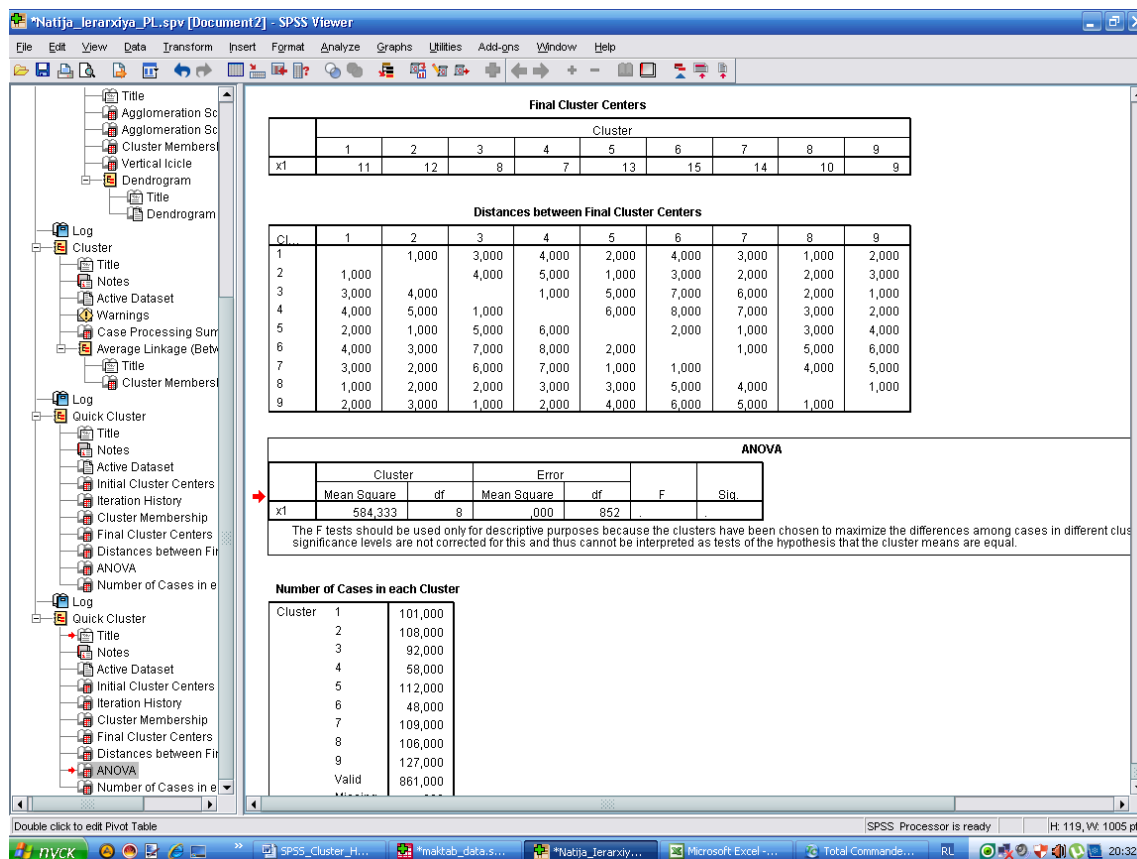
a) b)
3.2-rasm. k-means klasterizatsiya uchun parametrlarni tanlash oynasi: a) – parametr darchasi; b) – Options... - sozlash darchasi.

Number of Cluster: taxrir satriga bizga kerakli klasterlar soni kiritiladi, ushbu misolda o'smirlarni maktab sinflari bo'yicha klasterlashimiz uchun klasterizatsiya soni 9 deb olindi. Natijada SPSS paketi asosiy oynasida quyidagi natijaviy jadvallar hosil qilindi (3.3-rasm).

1.1-jadval. Natijaviy klasterlar markazlari.

	Cluster								
	1	2	3	4	5	6	7	8	9
x1	11	12	8	7	13	15	14	10	9

Bu jadvalda 9 ta klasterga x_1 belgi bo'yicha klasterizatsiya natijasi keltirilgan.



3.3-rasm. k-means klassifikatori natijasi

Klasterizatsiya natijalari 1-ilova 2-jadvalda keltirilgan. 1-ustun ob'yekt raqami, 2-ustun klaster raqamini bildiradi.

1.2-jadvalda klasterlar markazlari orasidagi masofalar keltirilgan, 1.3-jadvalda esa tahlil keltirilgan. Bu tahlil faqat klasterlar markazlari orasidagi farqlar turlicha bo'lganda to'liqroq tahlilni beradi.

1.2-jadval. Klasterlar markazlari oarsidagi masofalar

Distances between Final Cluster Centers									
Cluster	1	2	3	4	5	6	7	8	9
1		1,000	3,000	4,000	2,000	4,000	3,000	1,000	2,000
2	1,000		4,000	5,000	1,000	3,000	2,000	2,000	3,000
3	3,000	4,000		1,000	5,000	7,000	6,000	2,000	1,000
4	4,000	5,000	1,000		6,000	8,000	7,000	3,000	2,000
5	2,000	1,000	5,000	6,000		2,000	1,000	3,000	4,000
6	4,000	3,000	7,000	8,000	2,000		1,000	5,000	6,000
7	3,000	2,000	6,000	7,000	1,000	1,000		4,000	5,000
8	1,000	2,000	2,000	3,000	3,000	5,000	4,000		1,000
9	2,000	3,000	1,000	2,000	4,000	6,000	5,000	1,000	

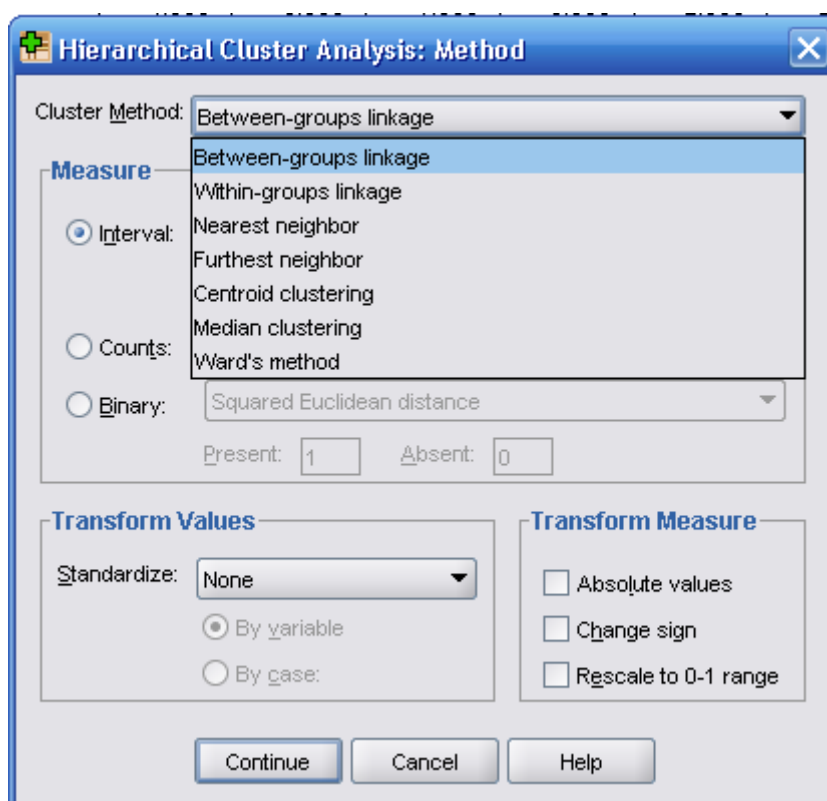
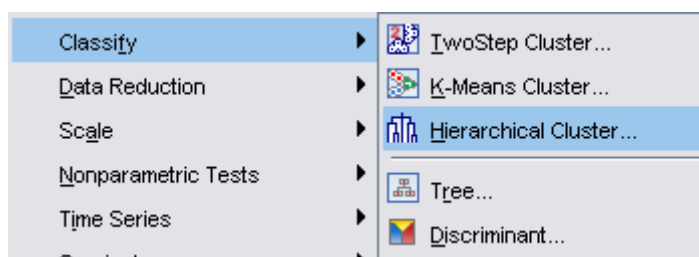
Har bir klasterdagi ob'yektlar soni quyidagi jadvalda berilgan. Demak, 1-klasterda 101 ta ob'yek sinflangan, 1.1-jadvalga ko'ra u 11 yoshlilar klasteri.

1.3-jadval. Tahliliy jadval.

Number of Cases in each Cluster		
Cluster	1	101,000
	2	108,000
	3	92,000
	4	58,000
	5	112,000
	6	48,000
	7	109,000
	8	106,000
	9	127,000
	Valid	861,000
	Missing	,000

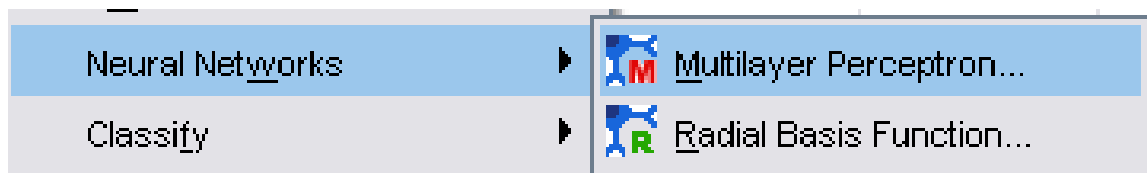
ANOVA						
	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
x1	584,333	8	,000	852	.	.

SPSS paketi orqali klassterizatsiya usullaridan TwoStep klassifikatori, Iyerarxik (aglomerativ) klasterizatsiyalar turiga mansub 7 ta usullarini tadqiq qilish mumkin.



3.4 a-rasm. Iyerarxik klasterizatsiya uslubi usullari

Bu usullarni dendogrammalari – yani iyerarxiyasini vizuallashtirish imkoni mavjud.

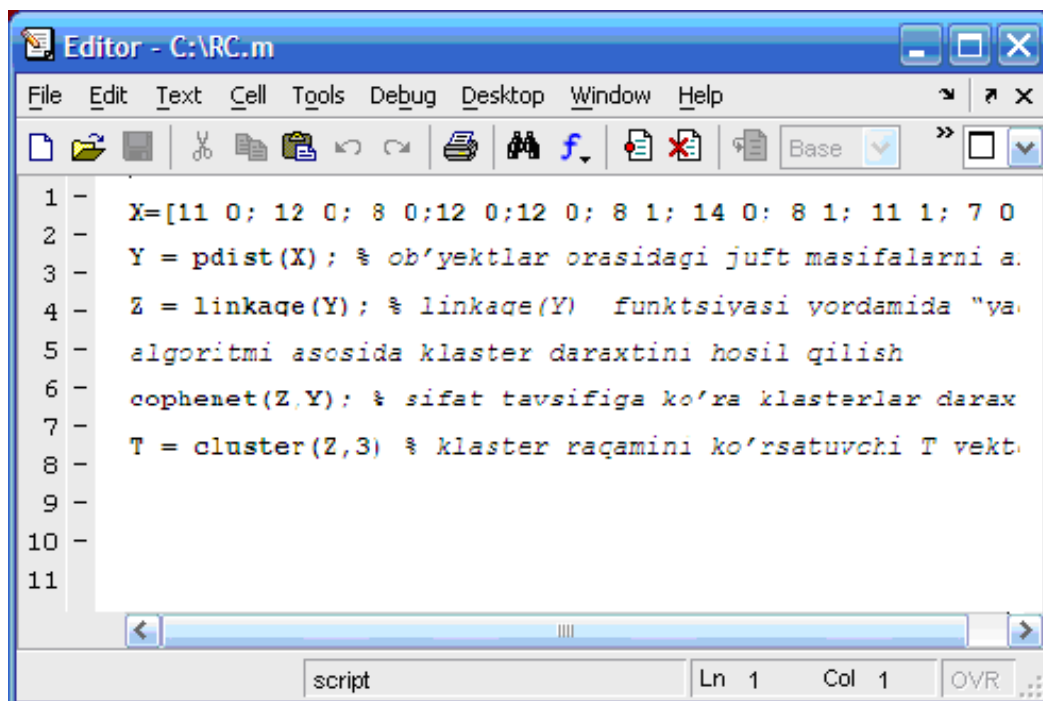


3.4 b-rasm. Neyrotarmoq usullari darchasi

Xatto neyrotarmoq usulida ma'lumotlarni intellektual tahlil qilish imkoniyati ham taqdim etilgan.

1.2. MATLAB tizimida klasterizatsiya dasturini yaratish

MATLAB – matritsaviy laboratoriya – hozirgi vaqtga kelib hisoblash matematikasi, axborotni qayta ishlash, elektr asboblarni loyihalashtirish, iqtisod va amaliy fanning boshqa bir qator bo'limlariga taalluqli o'ndan ortiq hususiy ilovalar bilan to'ldirilgan ilmiy-texnik hisoblashlar uchun eng rivojlangan dasturlash tizimidir. Shu sababli bu kodni MATLABda modellashtirish orqali samarali natijalarga erishish mumkin[6].



3.5-rasm. MATLAB tizimining dastur yozish (taxrirlash) oyansi

MATLAB – interpretator kabi ishlovchi hamda turli xil hisoblashlarni

bajarish, ma'lumotlar tarkibini berish va axborotni grafik ko'rinishda tasvirlash uchun xizmat qiladigan katta hajmli qo'llanmalar (buyruqlar) to'plamini o'z ichiga oluvchi yuqori darajali dasturlash tizimidir. Ushbu buyruqlar tizimning har xil direktoriyalarida joylashgan mavzuli guruxlarga bo'lingan. Endilikda tizimda 800 ga yaqin buyruq bo'lib ulardan yarmidan ko'pi boshlovchi foydalanuvchi uchun tushunarlidir. Bundan tashqari MATLAB – algoritmlarni tekshirish va taqribiy hisoblashlar o'tkazishdagi mehnat harajatlarini jadal suratda qisqartirish imkonini beradi. MATLAB da katta hisoblashlar o'tkazish imkoniyati foydalanuvchi yo'l qo'ya oladigan vaqt darajalariga qarab aniqlanadi. Tizim sonli usullarni o'zlashtirish va qo'llash uchun juda qulay[6].

MATLAB tizimida Statistics Toolbox paketi mavjud, unda klasterli tahlil funktsiyalari quyidagi jadvalda keltirilgan.

1.4-jadval. MATLAB tizimi Statistics Toolbox paketining klasterizatsiyalash funktsiyalari

Funktsiya nomi	Vazifasi
cluster	klasterlarni alohida klasterlarning iyerarxik daraxt ko'rinishad ajratish. (Linkage funktsiyasida chiqish ma'lumotlarni guruxlash)
clusterdata	Boshlang'ich ma'lumotlar matritsasini klasterlarga guruxlash
cophenet	Boshlang'ich ma'lumotlarni klasterlarga ajratish sifati ko'effitsiyentini hisoblash. (bu ko'effitsiyent korrelyatsiyani ko'effitsiyentiga o'xshatish mumkin, ya'ni qiymati qanchalik 1 ga yaqin bo'lsa, klasterlash shunchalik aniq ajratilgan bo'ladi)
dendrogram	Klasterlarning dendogrammasi
inconsistent	Klasterlarning iyerarxik daraxtida har bir bog'liqligining mostushmaslik ko'effitsiyentlarini hisoblash. Klasterlarni ajratish sifatini baholashda ham

	qo'llanilishi mumkin.
kmeans	Ichki guruxlar o'rtachalariga asoslangan klasterizatsiya
linkage	Binarli klasterlarning iyerarxik daraxtini shakllantirish
pdist	Boshlang'ich ma'lumotlar to'plamidagi ob'yektlar (vektorlar) o'rtasidagi juft masofalarni hisoblash
silhouette	Klasterlarning siluetlari (konturlari) grafigi
squareform	pdist funktsiyasi chiqishlaridagi ma'lumotlari vektorini simmetrik kvadrat matritsaga almashtirish

Bu funktsiyalar orqali MATLABda klasterizatsiya usullaridan uchtasining (iyerarxik usuli, k-means, Koxonen neyrotarmog'i) dasturini tuzamiz.

1. Iyerarxik usulda klasterizatsiyalash.

861 ta satr 2 ta ustunli X matritsaga ma'lumotlarni beramiz, ya'ni har biri 2 elementli 861 ta vektor.

```
X=[11 0; 12 0; 8 0;12 0;12 0; 8 1; 14 0; 8 1; 11 1; 7 0; 12 1; ....];
Y = pdist(X); % ob'yektlar orasidagi juft masifalarni aniqlash
Z = linkage(Y); % linkage(Y) funktsiyasi yordamida "yaqin qo'shni"
algoritmi asosida klaster daraxtini hosil qilish
cophenet(Z,Y); % sifat tavsifiga ko'ra klasterlar daraxtiga ajratish
T = cluster(Z,3) % klaster raqamini ko'rsatuvchi T vektorni hosil qilish
```

Ushbu dastur 1-ilova 2-jadvaldagidek Natija berdi.

2. k-means algoritmi amaliy tadqiqi

```
data=load('maktab_data.dat'); % ma'lumot fayldan yuklandi
[center,U,obj_fcn]=fcm(data,3); % klaster markazini aniqlash
maxU=max(U); % massiv elementlarini indekslash
index1=find(U(1,:)==maxU);
index2=find(U(2,:)==maxU);
index3=find(U(3,:)==maxU);
```

3. Koxonen neyrotarmoq usuli amaliy tadqiqi

```

net = newc([0 1; 0 1; 0 1; 0 1; 0 1; 0 1; 0 1; 0 1; 0 1; 0 1; 0 1],3,0.1); % 3
pog'ona
w0 = net.IW{1};
b0 = net.b{1};
c0 = exp(1)./b0;
tic, net.trainParam.epochs = 50; % 50 tsikl hosil qilinadi
[net,TR] = train(net,P); toc
w = net.IW{1};
bn = net.b{1};
cn = exp(1)./bn;
Y=sim(net,P);
disp(Y)

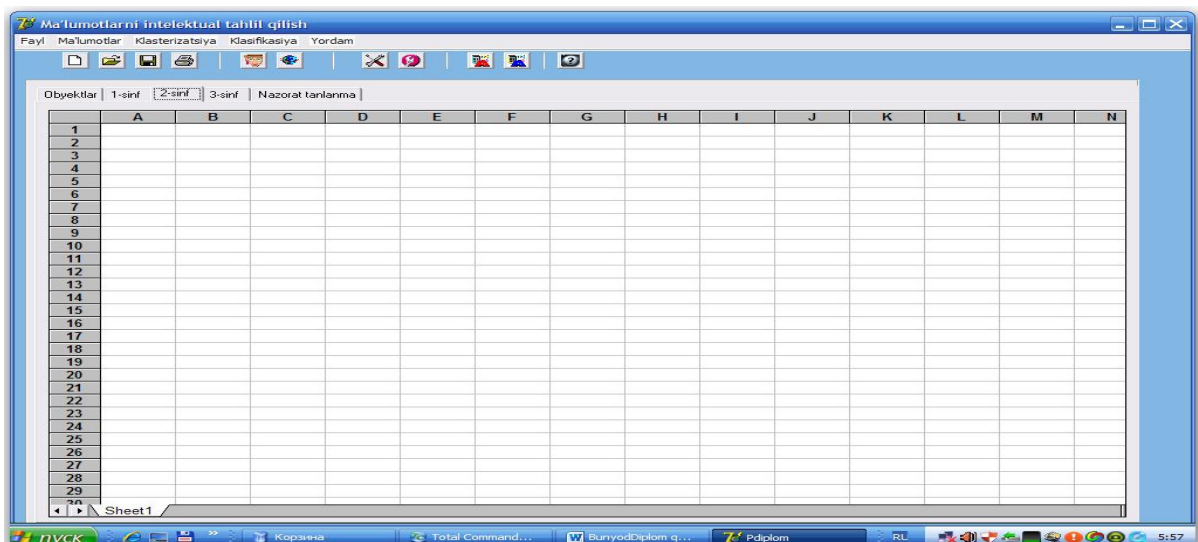
```

Bu usullar orqali olingan natijalar deyarli bir xil ko'rinishda hosil qilindi.

2. Delphi dasturlash tilida yaratilgan klasterizatsiya dasturiy ta'minoti

Klasterizatsiya masalasi asosida yaratilgan bu dasturimiz obektlarning belgilari asosida klasterizatsiyalash amalga oshiradi.

Dasturni bilan ishlash jarayonini ko'rib chiqamiz. Buning uchun Klaster1.exe ni yuklaymiz va quyidagi dastur oynasi hosil bo'ladi(3.1-rasm).



3.6-rasm. Dasturning bosh menyusi.

Bu dasturimiz ham menyu bandlaridan tashkil topgan:

- Fayl;
- Ma'lumotlar;
- Klasterizatsiya;
- Yordam.

Har bir menyu bir necha bandlardan tashkil topgan.

Dasturning ma'lumotlar menyusi ikkita banddan tashkil topgan.

-avtomatik;

-mutaxasis.

Bizlar avtomatik bandini tanlaymiz. Shundan so'ng bizlardan belgilar fazosi o'lchami va obektlar soni kiritishni suraydi(3.7-rasm).

3.7-rasm.Belgi va obektlar sonini kiritish qismi.

O'lchamlarni kiritganimizdan so'ng quyidagi jadval hosil bo'ladi.

(3.8-rasm).

Obyektlar	A	B	C	D	E	F	G	H	I	J	K	L	M
1	-0.119	-0.917	-0.872	0.068	-0.4	0.133	-0.654	-0.543	-0.629	-0.21			
2	0.989	-0.784	-0.374	-0.934	0.886	-0.185	-0.16	0.483	0.855	-0.1			
3	0.511	0.331	0.879	-0.463	0.237	0.446	-0.135	-0.608	0.494	0.474			
4	0.838	0.068	-0.791	0.089	-0.094	0.32	0.707	-0.866	0.853	-0.627			
5	0.148	0.159	-0.317	-0.963	-0.971	0.316	-0.019	-0.124	0.855	0.085			
6	0.42	0.114	0.569	0.798	0.259	0.517	-0.164	-0.565	-0.814	0.966			
7	0.773	-0.347	0.336	0.934	-0.781	0.516	-0.16	0.595	-0.853	0.907			
8	0.739	-0.226	0.295	0.078	-0.388	0.464	-0.377	-0.688	0.76	-0.875			
9	-0.312	0.163	-0.706	-0.511	-0.471	-0.749	0.962	-0.924	0.15	0.065			
10	-0.931	-0.837	-0.796	0.764	-0.676	0.399	-0.027	-0.58	-0.038	0.802			
11													
12													

3.8-rasm.Tasodifiy sonlar bilan to'ldirilgan jadval.

Keyin klasterizatsiya menyusidan masofa yoki k-o'rtacha yaqin bandlarini tanlaymiz.Shundan so'ng tahminiy sinflar sonini kiritish oynasi faollashadi. Sinf o'lchamini kiritganimizdan so'ng dastur obektlarning belgilari asosida sinflarga ajratib beradi(3.9, 3.10, 3.11-rasmlar,).

Ma'lumotlarni intellektual tahlil qilish

Fayl Ma'lumotlar Klasterizatsiya Klasifikatsiya Yordam

Obyektlar 1-sinf 2-sinf 3-sinf Nazorat tanlanma

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	-0.673	0.65	0.628	-0.147	-0.301	-0.718	0.249	-0.749	0.145	-0.256			
2	-0.235	-0.157	0.646	0.34	-0.178	-0.121	-0.16	0.741	0.763	-0.787			
3	0.304	-0.256	-0.596	0.585	0.242	0.504	0.314	-0.627	0.864	0.541			
4													
5													
6													

3.9-rasm. 1-sinfga tegishli ob'ektlar.

Ma'lumotlarni intellektual tahlil qilish

Fayl Ma'lumotlar Klasterizatsiya Klasifikatsiya Yordam

Obyektlar 1-sinf 2-sinf 3-sinf Nazorat tanlanma

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	-0.812	-0.2	0.626	-0.801	0.818	-0.119	0.523	-0.68	0.446	0.38				
2	0.615	0.831	0.832	0.345	0.447	-0.825	0.012	-0.764	0.301	-0.43				
3	-0.957	-0.755	0.711	0.935	0.128	0.959	0.378	-0.945	-0.788	0.552				
4														
5														
6														
7														

3.10-rasm. 2-sinfga tegishli ob'ektlar.

Ma'lumotlarni intellektual tahlil qilish

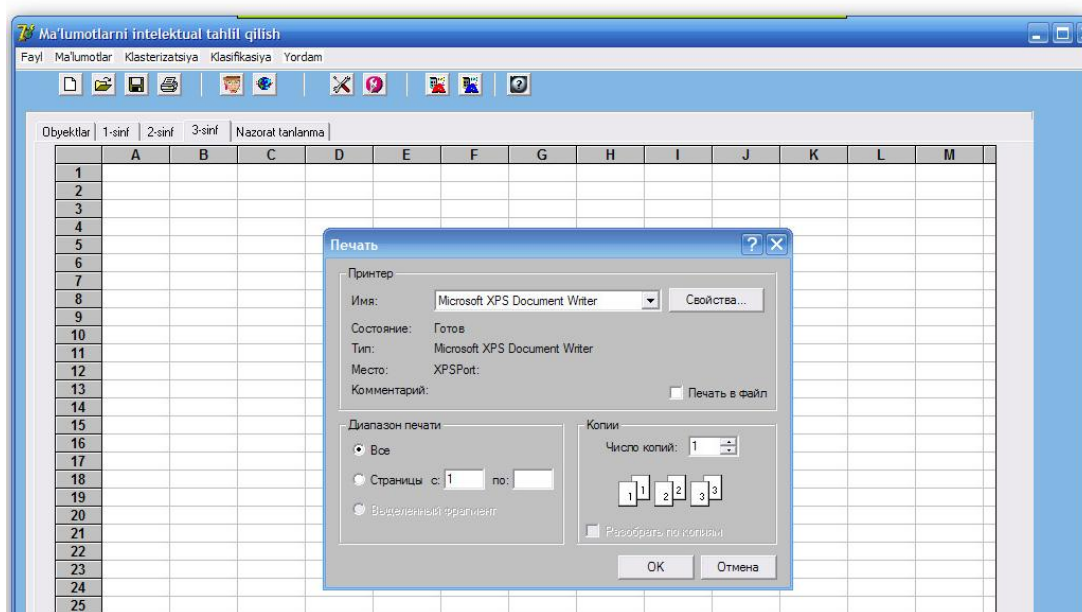
Fayl Ma'lumotlar Klasterizatsiya Klasifikatsiya Yordam

Obyektlar 1-sinf 2-sinf 3-sinf Nazorat tanlanma

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2	0.484	-0.166	0.993	0.134	-0.402	-0.708	0.017	-0.557	0.641	0.835			
3	0.771	-0.736	-0.48	-0.347	0.539	0.392	0.919	-0.962	-0.632	-0.147			
4	0.597	0.722	-0.843	-0.863	0.322	0.488	0.088	0.193	-0.864	-0.168			
5													
6													
7													

3.11-rasm. 3-sinfga tegishli ob'ektlar.

Berilgan natijalarni chop etmoqchi bo'lsak fayl menyusidan chop etish bandidini tanlaymiz (3.12-rasm).



3.12-rasm.Chop etish banddi.

Dasturning menyulari orqali mutaxassis tomonidan berilgan ob'ektlar belgilarini kiritish, saqlash, chiqish va hakazo bandlardan tashkil topgan.

Ushbu Delphi tilida yaratilgan dasturiy ta'minot klaterizatsiya masalasi k-means algoritmi asosida ishlaydi. Bunda asosan ta'lim sohasining ixtiyoriy yo'nalishlarida ma'lumotlarni kerakli jihatdari asosida sinflarga ajratib beradi. Shu bilan birga ma'lumotlarni tahlil etish jarayonlarida qo'llash mumkin.

3. Kompyuter xonalarini yoritishga qo'yilgan asosiy talablar

Ishlab chiqarish mikroiklimi normalari mehnat xafsizligi standartlari tizimi "Ish zonasi mikroiklimi" (GOST 12.1· 005-76) ga asosan belgilanadi. Ular gigiyenik, texnik va iqtisodiy negizlarga asoslangan [32].

Sanoat korxonalaridagi xonalar, yil fasllari va ish toifasiga qarab, ulardagi harorat, nisbiy namlik va havo harakatining ish joylari uchun ruxsat etilgan normalari belgilangan.

Ishlab chiqarish xonalari, kompyuter xonalari,
ish joylardagi havoning harorati, nisbiy namligi va
xarakat tezligining risoladagi normalari

Yil fasli	Ish toifalari	Havoning harorati, S	Nisbiy namlik %	Xarakat tezligi m/s
sovuq	Yengil – I	20-23	60-30	0,2
	O'rtacha og'irlikdagi – I	18-20	60-40	0,2
	Og'ir – III	16-18	60-40	0,3
iliq	Yengil – I	20-25	60-40	0,2
	O'rtacha og'irlikdagi – I	21-23	60-40	0,3
	Og'ir – II	18-21	60-40	0,5
issiq	Yengil – I	20-30	60-30	0,3
	o'rtacha og'irlikdagi – I	20-30	60-30	0,4-0,5
	Og'ir – III	20-30	60-30	0,5-1,0

Sanoat korxonalari va kompyuter xonalarida yoritish. Yorug'lik inson xayoti faoliyati davomida juda muxim rol o'ynaydi. Shuning uchun ham sanoat korxonalarini, kompyuter bor honalarni rasional yoritish sifatli maxsulot ishlab chiqarishni ta'minlash bilan birga ishlab chiqarish sharoitini yaxshilaydi, ishchilarni charchashdan saqlaydi va mehnat unumdorligini oshiradi. Oqilona yoritilgan zonalarda ishlayotgan ishchilarning kayfiyati yaxshi bo'ladi, buning natijasi baxtsiz xodisalar keskin kamayadi[32].

Bundan ko'rinib turibdidiki, kompyuter xonalarini yoritish faqatgina gigiyenik talab qo'yilmasdan, balki texnik-iqtisodiy talablar ham qo'yiladi.

Kompyuter xonalarida yoritishning mukammalligi sifat va son ko'rsatkichlari bilan tavsiflanadi. Son ko'rsatkichlari nur oqimi, yorug'lik kuchi, yorqinlik, nur qaytarish koeffitsiyentlari, yorug'lik kiradi.

1. Ish joylarini yoritish sanitariya – gigiyena normalari asosida ish toifalariga moslashgan bo'lishi kerak. Ish joylarini maksimal yoritish albatta ish sharoitini yaxshilashga olib keladi. Bundan ish olib borilayotgan obyektning kurinishi yaxshilanadi, natijada ish unumdorligi ortadi. Ba'zi bir aniq ishlarni bajarganda yoritilishni 50 lk dan 1000 lk gacha oshirish bilan ish unumi 25 foizga oshganligi ma'lum. Ko'z bilan ko'rib ishlash unchalik shart bo'lmagan ishlarni bajarganda ham yoritilishni 50 lk dan 300 lk ga oshirish ish unumini 5-7 foizga oshirgan [32].

2. Ish olib borilayotgan yuzaga va ko'zga ko'rinadigan atrof muxitga yorug'lik bir tekis tushadigan bo'lishi kerak. Chunki agar ish olib borilayotgan yuzada va atrof muxitda yaltiroq uchastkalar mavjud bo'lsa, unda ko'zning jimirlashi va ma'lum vaqt ko'nikishi kerak bo'ladi. Bu esa ko'zning tez charchashiga olib keladi.

3. Yoritish miqdori vaqt bo'yicha o'zgaras bo'lishi kerak. Yoritilishning ko'payib – kamayishi, agar u o'qtin-o'qtin ro'y beradigan bo'lsa, ko'zga zarar keltiradi, chunki ko'z yorug'lik o'zgarishlariga ko'nikishga to'g'ri keladi. Bu esa ko'zning tez charchashiga olib keladi. Yoritilishning o'zgarasligi muxim o'zgaras kuchlanishli manbalaridan foydalanish yo'li bilan erishish mumkin.

4. Yorug'lik qurilmalari qo'shimcha xavf va zararlar manbai bo'lmasligi kerak. Shuning uchun yoritish manbalari ajratadigan issiqlikni, tovush chiqarishni maksimal kamaytirish zarur.

5. Yoritish qurilmasi ishlatish uchun qulay, o'rnatish oson va iqtisodiy samarador bo'lishi kerak.

Yoritgichlar va ularni joylashtirish. Yorug'lik manbalari yoritish armaturasi joylashadi va ular birgalikda yoritgichlar yoki chiroqlar deb ataladi. Yoritgichlar konstruksiyalariga qo'ydagi talablar qo'yiladi:

- 1) nur oqimining yo'nalishini ishchi yuzalar tomonga qayta taqsimlanishini ta'minlash;
- 2) lampaning nur tarqatayotgan yuzalarining yaraqlab ko'zga ta'sir ko'rsatishidan muxofaza qilish;
- 3) lampani har xil sanoat iflosliklari va changdan himoya qilish;
- 4) lampani portlash, o't olish xaflaridan muxofaza qilish.

Shovqindan saqlanish. Hozirgi zamon texnika taraqqiyoti davrida sanoat korxonalarida shovqinga qarshi kurash masalalari muhim muammolar qatoriga kiradi. Bu – asosan mashinasozlik sanoati, transport vositalarini ishlatishda, energetika sanoatida va sanoat korxonalarida juda jiddiy muammo bo'lib turibdi. Shovqinning oqibatlari ma'lum. Katta shovqin ta'sirida insonning asab tizimlari charchab, eshitish faoliyati susayib ketadi. Shuning uchun ham sanoat korxonalarida shovqinni kamaytirish chora-tadbirlarini belgilash inson salomatligini saqlashdek juda muhim ijtimoiy ahamiyatga moliqdir.

Shovqinga qarshi kurash chora-tadbirlari qo'yidagi usullarda olib boriladi:

- shovqinni ajralib chiqayotgan manbaida kamaytirish;
- shovqinning tarqalish yo'nalishini o'zgartirish;
- sanoat korxonalari va kompyuter honalarini oqilona rejalashtirish;
- sanoat korxonalari va kompyuter honalariga akustik ishlov berish;
- shovqinni tarqalish yulida kamaytirish.

Elektr hafsizligini ta'minlash. Sanoat korxonalarida o'rnatilgan kompyuter, yordamchi mashina-mexanizmlar, shkaflar, boshqarish pultlari

hamda elektr asboblari va boshqa jihozlarning hammasi, agar ularni ishlatish uchun foydalanadigan elektr quvvati 42 V dan yuqori kuchlanishga ega bo'lsa, albatta yerga ulangan bo'lishi shart. Buning uchun har qanday kompyuter honalari yerga ulash vositalari qulay, oson va yengil bajariladigan bo'lishi kerak. Bu vositalar yerga yetarli darajada puxta ulangan yoki nolga ulangan bo'lishi shart. Bu vositalarda yerga ulash simlarini osongina mahkamlash mumkin bo'lgan vintsimon qurilma yoki qiskich bo'lishi kerak.

Yong'inga qarshi ishlatiladigan texnik qurilmalar. Sanoat korxonalarini binolarini yong'indan muhofaza qilish uchun ishlatiladigan asosiy texnik qurilmalar GOST 12,4009-75 asosida aniqlanadi. Har qanday yong'inni o'chirishda yong'inning kuchayishiga olib kelayotgan omillarni va sharoitni aniqlash muximdir. Bunda yonishning davom etishini to'xtatuvchi sharoit yaratish katta rol o'ynaydi. Yong'inni o'chirish paytida qattiq jismlar yonganda yong'inning tezligi 4m/min, suyuqliklar yuzasi bo'yicha esa 30 m/min bo'lishishini hisobga olish kerak[32].

Yong'in (o't) o'chirish vositalari va usullari. O't o'chirish usullari qo'ydagicha bo'lishi mumkin:

- yonayotgan zonani ko'p miqdorda issiqlik yutuvchi materiallar yordamida sovitish;
- yonayotgan materiallarni atmosfera havosidan ajratib qo'yish;
- yonayotgan zonaga kirayotgan kislorod miqdorini kamaytirish;
- maxsus kimyoviy vositalarni qo'llash.

O't o'chirish vositalari sifatida, suv bug'lari, kimyoviy va mexanik ko'piklar, inert va yonmaydigan gazlar, qattiq kukunsimon materiallar va aralashmalardan foydalaniladi.

Suv bilan o'chirish. Suv eng ko'p tarqalgan arzon va shuning bilan birga deyarli hamma yerda mavjud bo'lgan o't o'chirish vositasi bo'lib, suv bilan har qanday masshtabdagi yong'inlarni o'chirish mumkin.

Bug' bilan o'chirish. Bug' bilan o'chirishning asosiy mohiyati shuki, honalarga yuborilgan bug' kislorodga boy havoni siqib chiqarib, uning o'rnini

egallaydi. Bug'ning o't o'chirish samaradorligi uning ma'lum bir xonaga yuborilgan miqdoriga bog'lik bo'ladi.

Yong'inga qarshi suv ta'minoti. Odatda o't o'chirish uchun ishlatiladigan suv katta bosim ostida kuchli oqim sifatida alanganayotgan joyga yuboriladi. Buning uchun yetarli bo'lgan bosimni shahar sharoitiga umumiy vodoprovod tarmoqlari orqali xosil qilinadi. Yoki bo'lmasa ba'zi bir yerlarda maxsus tayorlangan xovuz va idishlardan foydalanilinish mumkin.

Isitish va shamollatish tizimlari. Sanoat korxonalarida odatda markaziy isitish tizimi orqali isitiladi. Shuning uchun bunday tizimlarning yong'in xafi bo'lgan uchastkalarida isitish radiatorlarining murakkab turlaridan foydalanish tavsiya etilmaydi. Chunki changlar truba va radiator ustki qismlarida yig'ilib qolishi, isitish natijasida qizib, yong'in chiqarish xafini kuchaytiradi[32].

Xonalari markaziy usulda isitilgan sanoat korxonalarida, havo asosan koloriferlarda isitilib, undan keyin korxona xonalariga yuboriladi. Bunday xollarda isitilgan havoning harorati 66°S dan oshmasligi kerak. Ammo bunday isitish tizimida havo kanallar orqali xonalarga tarqatilgani sababli yong'in bo'lgan taqdirda bu kanallar orqali alanga tarqatilishi mumkin. Chunki alanga va tutun butun bino bo'ylab tarqalib ketishi tufayli yong'in xavfi kuchayadi.

III bob bo'yicha xulosalar

Ushbu bobda ma'lumotlarni intellectual tahlilida klasterizatsiya algoritmlari dasturiy ta'minotining amaliy tadbiqi keltirilgan. Klasterizatsiya usuli amaliy dasturiy paketlarida ishlash uchun professional statistik va matematik paketlarni turli dasturiy modullar ishlab chiqilgan, bu paketlarda ma'lumot sifatida tibbiyot muassasasi tomonidan yig'ilgan o'smirlar haqidagi ma'lumotlar berilgan, shu ma'lumotlarni SPSS va MATLAB tizimi Statistics Toolbox paketlarida klasterizatsiya usullari algoritmlari orqali sinflarga ajratib intellektual tahlil etildi va natijalarga erishildi. Delphi dasturlash tilida yaratilgan klasterizatsiya dasturiy ta'minoti yaratildi va natijalar tahlil qilindi. Bu yaratilgan dasturiy ta'minot kata xajimli ma'lumotga ega bo'lgan ob'ektlarni sinflarga ajratadi. Ushbu dasturiy taminot ta'lim sohasida va boshqa sohalarda yig'ilgan ma'lumotlarni sinflaga ajratishni qadamlab, asoslab ajratadi.

XULOSA

Dissertasiya ishining asosiy natijalari ma'lumotlarni intellektual tahlil etishni nazariy asoslaridan foydalangan holda tashkil etildi, klasterizatsiya usuli algoritmlari qiyosiy tahlil qilindi. Shu qiyosiy tahlillarga tayangan holda Deiphi dasturlash tilida dasturiy ta'minoti yaratildi.

Ma'lumotlarni intellektual tahlil etish usullarining nazariy va amaliy asoslariga bag'ishlangan bo'lib, unda asosiy tushunchalar va ta'riflar, ma'lumotlarning turlari, ularning saqlash formatlari va ma'lumotlar to'plami atributlari, ma'lumotlarbazasi, Data Mining uslublari va bosqichlari nazariy-amaliy usullari, vositalari keltirilgan hamda tahlil etilgan.

Ma'lumotlarni intellektual tahlil etish jarayonida klasterizatsiya masalasi algoritmlari aks ettirilgan bo'lib, paragrafning birinchi qismida klasterizatsiya masalasini tizimli yechishning formal qo'yilishi va algoritmlar tahlillari to'g'risida ma'lumotlar berilgan va ular qarab o'tilgan. Bu paragrafda klasterizatsiya usuli algoritmlari qadamlari oydin yoritib berilgan. Ushbu algoritmlar Aglomerativ algoritmlar, Divizim algoritmlari, noierarxik algoritmlar, k-means, Fuzzy C-means, Gustafson-Kassel bo'yicha klasterlash kibi algoritmlar o'rganildi va ushbu algoritmlar asosida II bobning ikkinchi qismida esa klasterizatsiya algoritmlarida qo'llaniladigan masofaga asoslangan yaqinlik darajasining o'lchovi keltirildi. Uchunchi qismida esa o'rganilgan algoritmlarning qiyosiy tahlili qilindi. Tahlillar natijasida algoritmlarning bir-biridan farqli jihatlari algoritmnining yutug'i, kamchigi, ma'lumotlar bilan ishlash turlari, ishlash tezligi, qulaylilik darajasi ushbu qismda jadvallar asosida keltirilgan.

Ma'lumotlarni intellektual tahlilida klasterizatsiya algoritmlari dasturiy ta'minotining amaliy tadbiqi keltirilgan. Klasterizatsiya usuli amaliy dasturiy paketlarida ishlash uchun professional statistik va matematik paketlarni turli dasturiy modullar ishlab chiqilgan, bu paketlarda ma'lumot sifatida tibbiyot muassasasi tomonidan yig'ilgan o'smirlar haqidagi ma'lumotlar berilgan, shu

ma'lumotlarni SPSS va MATLAB tizimi Statistics Toolbox paketlarida klasterizatsiya usullari algoritmlari orqali sinflarga ajratib intellektual tahlil etildi va natijalarga erishildi. Delphi dasturlash tilida yaratilgan klasterizatsiya dasturiy ta'minoti yaratildi va natijalar tahlil qilindi. Bu yaratilgan dasturiy ta'minot kataxajimli ma'lumotga ega bo'lgan ob'ektlarni sinflarga ajratadi. Ushbu dasturiy taminot ta'lim sohasida va boshqa sohalarda yig'ilgan ma'lumotlarni sinflarga ajratishni qadamlab, asoslab ajratadi.

FOYDALANILGAN ADABIYOTLAR RO'YXATI

1. Zamonaviy avborot-kommunikatsiya texnologiyalarini yanada joriy etish va rivojlantirish chora tadbirlari to'g'risida. O'zbekiston Respublikasi Prezidentining 21.03.2012 y.dagi №PQ-1730 sonli qarori. O'zbekiston qonun hujjatlar to'plami, 2012y., 13-son, 139-modda. [www.lex.uz]
2. "Axborot erkinligi prinsiplari va kafolatlar to'g'risida" O'zbekiston Respublikasining 12.12.2002 y. № 439- II- son Qonuni.
3. "Axborotlashtirish to'g'risida" O'zbekiston Respublikasining 11.12.2003y. № 560-II- son Qonuni.
4. "Kompyuterlashtirishni yanada rivojlantirish va axborot-kommunikatsiya texnologiyalarini joriy etish to'g'risida" O'zbekiston Respublikasi Prezidentining 30.05.2002y. PF-3080-son Farmoni.
5. "Kompyuterlashtirishni yanada rivojlantirish va axborot-kommunikatsiya texnologiyalarini joriy etish to'g'risida" O'zbekiston Respublikasi Prezidentining 30.05.2002y. PF-3080-son farmoni.
6. Дадажонов Т. "МАТЛАВ асослари". Дарслик, ФарПИ, 2007. -570 б.
7. Алексеева И.Ю. Искусственный интеллект и рефлексия над знаниями.// "Философия науки и техники": журнал 1991 №9, с. 44-53.
8. Брусиловский П.Л. Интеллектуальные обучающие системы. // Информатика. Наука-технический сборник. Киев 1990.№2.
9. Бекмуратов Т.Ф., Бобомуродов О.Ж. Интеллектуальная обучающая система основам произвольной дисциплины.// Сборник тезисов международной конференции «Инновация -2002». Ташкент, 10-11 ноября 2002, с. 253-254.
10. Бекмуратов Т.Ф., Бобомуродов О.Ж. Интеллектуал электрон ўқув қўлланмалари ёрдамида ўқитиш жараёнини бошқаришни ташкиллаштириш.// Совместный выпуск Узбекский журнал "Проблемы информатики и энергетики", 2004, №1, с. 210-214.
11. "Intelektual salohiyatli iqtidorli yoshlar" mavzusidagi yosh olimlar, magistrant va talabalarning an'anaviy XI respublika ilmiy-amaliy konferensiyasi III qism.
12. Дюк В.А., Самойленко А.П. Data Mining: учебный курс.// СПб.:2001
13. Боровков А.А.Математическая статистика:Учебник.–М.:Наука,1984.- 472с.
14. Бекмуратов Т.Ф., Бобомуродов О. Ж. Интеллектуал ўқитиш тизимларида БО модели асосида ўқитишни ташкил этиш//Совместный

выпуск Узбекский журнал "Проблемы информатики и энергетики". – Ташкент.

15. Бобомуродов О.Ж., Бекмуратов Т.Ф. Структурно - функциональная модель интеллектуального электронного пособия.//Сборник тезисов республиканской профессора-педагогической конференции «Развития технических наук при рыночном отношении» 2003, 2-часть. С. 60-62.
16. Бобомуродов О.Ж., Ахатов А.Р.: Обработка информации в системах приобретения знаний. Издательство «ФАН» АН РУз, 2009г., 136с
17. Бобомуродов О.Ж., Рахимов Н.О. Объектный когнитив модели ва интеллектуал ўқитиш тизимини ташкил этиш масалалари.// Ёш математикларнинг янги теоремалари –2009.-Наманган. 2009. 136-138 б.
18. Монахов М.Ю. Информационная образовательная сеть.// Информационные технологии, №7, 2001.
19. Комаров В.А., Соловов А.В. АОС и инженерная интуиция.// "Вестник высшей школы", 1986, № 2.– с.30-33.
20. Комилов М.М., Бобомуродов О.Ж. Ўқитиш жараёнида ҳолатларни баҳолаш ва стратегияни белгилаш усул ва моделлари.// Совместный выпуск Узбекский журнал "Проблемы информатики и энергетики", 2006г, спец выпуск., стр. 12-19.
21. Камилов М.М., Бобомуродов О.Ж., Худайбердиев М.Х. О применении современных информационных технологий в образовательной системе. // Вопросы кибернетики, Ташкент, 2003, №165, с. 84-94.
22. Davis R. (1982). Expert systems: Where are we? And where do we go from here? *AI Magazine*, 3(2).
23. Kahn G. and McDermott J. (1984). The MUD system. In *Proc. 1st IEEE Conference on Artificial Intelligence Applications*.
24. Бобомуродов О.Ж., Рахимов Н.О. Объектный когнитив модели ва интеллектуал ўқитиш тизимини куриш масалалари// ТАТУ хабарлари. – Тошкент, 2008. №4. 24-28 б.
25. Kamilov M.M., Bobomurodov O.J., Khudayberdiev M.X. The recognition and adaptation in the intelligent systems of the distant learning.// WCIS-2004, Tashkent 2004, p.55-56.
26. Kamilov M.M., Bobomurodov O.J., Rahimov N.O. Method for formation of knowledge by subject// WCIS-2010. Sixth world conference on Intelligent Systems for Industrial Automation. Tashkent. 2010. pp.12-14.
27. Нишанов А.Х., Туляганов Ш.Е., Бобомуродов О.Ж., Худайбердиев М.Х. Распознавания и адаптация для дистанционного образования //

- Математические Методы Распознавания Образов/ Доклады XI Всероссийской конференции. Москва, 2003, с.387-389.
28. Bekmuratov T.F., Bobomurodov O.J., Dadaboeva R.A. Structural – functional model of intellektual textbooks.// WCIS-2002, Tashkent 4-5 june, p. 38-41.
 29. Bekmuratov T.F., Bobomurodov O.J., Dadaboeva R.A. The level of projection and modeling intelligent teaching programs. //WCIS-2004, Tashkent, 2004, pp.13-16.
 30. Bobomurodov O.J., Rahimov N.O. Shaping scholastic material in system of the intellektual education.// International Conference on IT Promotion in Asia 2009 .Tashkent. 2009, pp302-304.
 31. Рахмиов Н.О. Интеллектуал ўқитиш тизимларида билимларни ифодалаш моделлари // ТАТУ хабарлари. – Тошкент. №4. 2010. 64-68 б.
 32. Hayot faoliyati xavfsizligi. A. Qudratov, T. G'aniyev va boshqalar. Toshkent, Aloqachi, 2005.
 33. <http://ru.wikipedia.org/wiki/SPSS>
 34. <http://fastra.ua.ac.be/en/index.html>
 35. <http://www.external.ameslab.gov/hoomd/index.html>
 36. <http://www.springerlink.com/content/pk22n1632859082k/fulltext.html>
 37. <http://www.linzik.com/>