

MA'LUMOTLARNI KLASTERLASH ALGORITMLARINI KO'RIB CHIQISH

Toxtaovunova Xojalibi Abdumuxtor qizi

Andijon davlat universiteti

Kompyuter ilmlari va dasturlash texnologiyalari

yo'nalishi 2-bosqich magistranti

REJA:

1. KLASTERLASH TUSHUNCHASI
2. MASOFA O'LCHOVLARI
3. ALGORITMLARNING TASNIFI
4. KLASTERLARNI BIRLASHTIRISH
5. ALGORITMLARNI KO'RIB CHIQISH
6. FOYDALANILGAN ADABIYOTLAR

Annotatsiya: Mazkur maqola, ma'lumotlarni klasterlash algoritmlarini ko'rib chiqish haqida bo'lib ushbu mavzu bo'yicha ma'lumotlar qisqacha mazmunda umumiy tushunchalar keltirilgan.

Kalit so'zlar: Rasmiy ta'riflar, miqdoriy xususiyatlar, klasterlash algoritmlari, ob'ektlarning butun namunasi.

1 Klasterlash

1.1 Klasterlash nima

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

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

1) Ma'lumotlarni tahlil qilish

- Axborot bilan ishlashni soddalashtirish
- Ma'lumotlarni vizuallashtirish

2) Axborotni olish va qidirish

- Qulay klassifikatorlarni qurish

3) Ob'ektlarni guruhlash va tanib olish

- Shaklni aniqlash
- Ob'ektlarni guruhlash

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

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

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

Rasmiy ta'riflar

Qo'shimcha mulohaza yuritish uchun biz ishlaydigan tushunchalarni kiritamiz.

Biz ob'ektni klasterlash algoritmi ishlaydigan elementar ma'lumotlar to'plami deb ataymiz. Har bir ob'ekt uchun uni tavsiflovchi parametrlar aniqlanadi,

xususiyat vektoriga birlashtirilgan

$x_i = x_1, x_2, \dots, x_m$, bu yerda m xarakteristikalar makonining o'lchami va x_i ob'ekt (sifat yoki miqdoriy).

Ikki ob'ektning o'xshashligi o'lchovi

$d(u,v)$ berilgan ko'rsatkich bo'yicha hisoblanadi. d - jismlar orasidagi masofa deb ataladi, bu erda u, v to'plam elementlari.

Xususiyat vektorini ajratib olish

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

$[0;1]$ yoki $[1;1]$, vazifaga qarab.

Masofaviy o'lchovlar

Xarakteristikalar vektorini aniqlagandan so'ng, masofa o'lchovi deb ataladigan ikkita ob'ektning o'xshashlik darajasini aniqlash uchun funktsiyani tanlash kerak.

Tanlangan funksiya metrikaning barcha shartlarini qondirishi kerak [2, 5-bet]. Ob'ektlarning yaqinligini hisoblash uchun turli ko'rsatkichlar mavjud.

Belgilovchi

u, v - orasidagi masofa hisoblangan ob'ektlar

$d(u, v)$ va u_i, v_i - ularning koordinatalari, biz ba'zi mavjud ko'rsatkichlarni tavsiflaymiz:

1) Evklid masofasi. Ko'p o'lchovli fazoda geometrik masofa bo'lgan klassik Evklid metrikasi

$$d(u, v) = \sqrt{\sum_{i=1}^m (u_i - v_i)^2}.$$

2) Evklid masofasining kvadrati, Evklid masofasining kvadratiga teng. Uzoqroq ob'ektlarning og'irligini oshirish uchun ishlatiladi.

3) Shahar bloki masofasi yoki Manxetten masofasi. U koordinatalar bo'yicha o'rtacha farq sifatida hisoblanadi va ko'pincha odatdagi Evklid masofasiga o'xshash natijalarga olib keladi.

$$d(u, v) = \sum_{i=1}^m |u_i - v_i|.$$

4) Tegishli ob'ektlar juda farq qiladigan o'lchamga ishora qilib, og'irlikni yuqoriga yoki pastga o'zgartirish zarur bo'lganda qo'llaniladigan kuch-qonun masofasi. Quvvat masofasi Evklidning masofa formulasiga o'xshash formula yordamida hisoblanadi:

$$d(u, v) = r \sqrt{\sum_{i=1}^m (u_i - v_i)^p},$$

Bu erda r va p foydalanuvchi tomonidan belgilangan parametrlar. p parametri individual koordinatalardagi farqlarni bosqichma-bosqich tortish uchun javob beradi va r parametri ob'ektlar orasidagi katta masofalarni progressiv tortish uchun javob beradi. Yuqorida aytib o'tilganidek, ikkalasi ham ikkitaga teng bo'lsa, bu ko'rsatkich Evklid masofasi bilan bir xil bo'ladi.

5) Chebishev masofasi. O'lchov qo'llaniladi, chunki ikkita ob'ektni har xil, bir xil koordinatada aniqlash kerak. Chebyshev masofasi formula bo'yicha hisoblanadi

$$d(u, v) = \max(|u_i - v_i|).$$

Klasterlash natijalari ko'p jihatdan metrikaning tanlashga bog'liq va turli ko'rsatkichlar uchun ular sezilarli darajada farq qilishi mumkin.

Klasterlash (yoki Klaster tahlili) — bu ob'ektlar to'plamini klasterlar deb ataladigan guruhlar bo'lish vazifasi. Har bir guruh ichida "o'xshash" ob'ektlar bo'lishi kerak va

har xil ob'ektlar guruhlar iloji boricha boshqacha bo'lishi kerak. Klasterlash va tasniflash o'rtasidagi asosiy farq shundaki, guruhlar ro'yxati aniq belgilanmagan va algoritm jarayonida aniqlanadi.

Klasterlash algoritmlari

Ierarxik klasterlash algoritmlari

Ierarxik klasterlash algoritmlari odatda ikki turga bo'linadi: yuqoridan pastga va pastdan yuqoriga algoritmlar. Pastga tushish, printsip bo'yicha harakat qilish

"kattadan eng kichigigacha": jarayonning boshida barcha ob'ektlar bitta klasterga, daraxtning tepasiga joylashtiriladi, shundan so'ng har bir qadamda mavjud klasterlardan biri ikkita kichikroqqa bo'linadi. ob'ekt o'z klasteriga tegishli.

Algoritmlarning ikkinchi turi, ortib boruvchi, ko'proq tarqalgan va birinchisiga nisbatan qarama-qarshi yo'nalishda ishlaydi. Birinchidan, ob'ektlarning har biri o'z klasteriga joylashtiriladi va algoritmnining har bir bosqichida ob'ektlarning butun namunasini o'z ichiga olgan faqat bitta klaster qolguncha eng yaqin ikkita klaster bittaga birlashtiriladi [5, 40-bet].

Ushbu algoritm bo'yicha klasterlash natijasi dendrogramma deb ataladigan bo'linish daraxtidir. Ierarxik klasterlardan foydalanishning eng mashhur namunasi hayvonlar va o'simliklarning tasnifidir.

Kvadrat xatolik algoritmlari

Klasterlash muammosini boshqacha talqin qilish mumkin: ob'ektlarning guruhlariga optimal bo'linishini qurish kerak. Bunday holda, optimallik sharti bo'limning o'rtacha kvadrat xatosini minimallashtirish talabi bilan belgilanishi mumkin:

$$e^2(X) = \sum_{j=1}^k \sum_{i=1}^{n_j} \|x_i^{(j)} - c_j\|^2,$$

Bu erda X - ob'ektlar to'plami, x_i^j, x_i^j - ularning koordinatalari, c_j - "massa markazi" j klaster (har bir nuqtaning massasi bittaga teng bo'lsa).

Bu toifadagi algoritmlar yassi algoritmlar sinfiga kiradi. K-means usuli ushbu turkumda eng mashhur hisoblanadi, chunki algoritm berilgan ob'ektlar to'plamini bir-biridan iloji boricha uzoqroqda joylashgan ma'lum miqdordagi klasterlarga ajratadi. Ushbu usulning ishi bir necha bosqichlarga bo'lingan:

- 1) Tasodifiy ravishda k boshlang'ich "massa markazlari" klasterlarini tanlang.
- 2) Har bir ob'ektni eng yaqin "massa markazi" bo'lgan klasterga tayinlang.
- 3) Klasterlarning "massa markazlari" ni hozirgi tarkibiga ko'ra qayta hisoblang.
- 4) To'xtatish mezonini tekshiring va agar u bajarilmasa, 2-bosqichga qayting.

Algoritmnining ishlashini to'xtatish mezonini sifatida, qoida tariqasida, o'rtacha kvadrat xatosining minimal o'zgarishi qo'llaniladi. Bundan tashqari, agar 2-bosqichda klasterini o'zgartirgan ob'ektlar bo'lmasa, algoritm tugaydi.

Bu algoritmnini tekislikda ishlatish natijasi Voronoi diagrammasiga yaqin. Ushbu

algoritmning kamchiliklari bo'linish uchun klasterlar sonini ko'rsatish zaruriyatini o'z ichiga oladi.

Noaniq algoritmlar

Yuqorida aytib o'tilganidek, loyqa klasterlash algoritmlari aniq usullardan farqli o'laroq, har bir ob'ektni har bir klasterga ma'lum bir ehtimollik bilan tayinlaydi. Ushbu turkumda juda ko'p algoritmlar mavjud emas, shuning uchun keling, eng mashhurini ko'rib chiqaylik - c-means usuli. Algoritmning bosqichlari k-o'rtacha usulining bosqichlariga o'xshaydi:

1) $n \times k$ o'lchamdagi U a'zolik matritsasini tanlab, n ta obyektning dastlabki noaniq bo'limini k klasterga o'rnatib.

2) U matritsasi yordamida loyqa xatolik mezonining qiymatini toping

$$E^2(X, U) = \sum_{i=1}^N \sum_{j=1}^K U_{ij} \|x_i^j - c_j\|^2, \quad c_k = \sum_{i=1}^N U_{ki} x_i,$$

Bu erda X - ob'ektlar to'plami x_i^j, x_i^j - ularning koordinatalari, c_j - "massa markazi" j klaster (har bir nuqtaning massasi birga teng bo'lsa), aksessuarlar.

U_{ij} - a'zolik matritsasi

3) Loyqa xatolik mezonining ushbu qiymatini kamaytirish uchun ob'ektlarni qayta guruhlang.

4) U matritsadagi o'zgarishlarga aylanmaguncha 2-bosqichga qayting ahamiyatsiz.

Bu algoritmdan faqat klasterlar soni oldindan ma'lum bo'lsa va har bir ob'ektning barcha klasterlarga nisbatini hisoblash zarur bo'lganda foydalanish kerak.

Grafiklar nazariyasiga asoslangan algoritmlar

Grafik algoritmlarining o'ziga xos xususiyati shundaki, ob'ektlarning butun tanlovi grafik sifatida ifodalanadi

$G = (V, E)$, bu yerda V cho'qqilar to'plami, E esa uchlari tanlash ob'ektlari bo'lgan qirralarning to'plami va qirralarning og'irligi ular bog'laydigan ob'ektlar orasidagi masofaga teng. Ushbu toifadagi algoritmlarning afzalliklari ularning aniqligi va geometrik mulohazalar asosida yangilanishlarni amalga oshirishning nisbatan qulayligidir. Ushbu turkumda eng ommaboplar bog'langan komponentlarni ajratib olish algoritmi, minimal oraliq (qo'ltiq) daraxtini qurish algoritmi va qatlamli klasterlash algoritmidir.

Bog'langan komponentlarni ajratib olish algoritmi

Ushbu algoritm ishlashi uchun R parametri talab qilinadi, bu chekka og'irliklari uchun chegara qiymatini belgilaydi. Ushbu algoritmning ishlashi davomida og'irligi chegara qiymatidan oshib ketadigan barcha qirralar asta-sekin olib tashlanadi. Ish natijasida grafik olinadi, unda faqat eng yaqin ob'ektlarni bog'laydigan qirralar qoladi.

Klasterlarni olish uchun faqat R qiymatini tanlash qoladi, shunda grafik bir nechta

bog'langan komponentlarga bo'linadi, ular klasterlar bo'ladi.

Ko'pincha, R parametrini tanlash uchun ular juftlik masofalarining taqsimlanishi gistogrammasini qurishdan foydalanadilar. Agar klaster tuzilishi nisbatan yaxshi ifodalangan bo'lsa, u holda gistogramma ikkita cho'qqiga ega bo'ladi, ulardan biri klaster ichidagi masofalarga, ikkinchisi esa klasterlararo masofalarga mos keladi. R parametri ushbu cho'qqilar orasidagi minimal zonadan tanlanadi.

USHBU ALGORITMNING SALBIY TOMONI SHUNDAKI, NATIJADA KLASTERLAR SONINI BOSHQARISH ANCHA MURAKKAB. IERARXIK KLASTERLASH ALGORITMLARI

Ierarxik klasterlash algoritmlari orasida ikkita asosiy tur ajralib turadi: yuqoriga va pastga algoritmlar. Pastga tushadigan algoritmlar yuqoridan pastga qarab ishlaydi: boshida barcha ob'ektlar bitta klasterga joylashtiriladi, so'ngra ular tobora kichikroq klasterlarga bo'linadi. Yuqoriga qarab algoritmlar keng tarqalgan bo'lib, ular ish boshida har bir ob'ektni alohida klasterga joylashtiradi va keyin barcha namuna olish ob'ektlari bitta klasterda bo'lguncha klasterlarni tobora kattaroq qilib birlashtiradi. Shunday qilib, ichki bo'linmalar tizimi quriladi. Bunday algoritmlarning natijalari odatda daraxt-Dendrogramma shaklida bo'ladi. Bunday daraxtning klassik namunasi hayvonlar va o'simliklarning tasnifidir.

Klasterlar orasidagi masofani hisoblash uchun hamma ko'pincha ikkita masofadan foydalanadi: bitta bog'lanish yoki to'liq bog'lanish (klasterlar orasidagi masofa o'lchovlarini ko'rib chiqishga qarang).

Ierarxik algoritmlarning kamchiliklari to'liq bo'limlar tizimini o'z ichiga oladi, bu hal qilinayotgan muammo kontekstida ortiqcha bo'lishi mumkin.

IERARXIK KLASTERLASH ALGORITMLARI

Ierarxik klasterlash algoritmlari orasida ikkita asosiy tur ajralib turadi: yuqoriga va pastga algoritmlar. Pastga tushadigan algoritmlar yuqoridan pastga qarab ishlaydi: boshida barcha ob'ektlar bitta klasterga joylashtiriladi, so'ngra ular tobora kichikroq klasterlarga bo'linadi. Yuqoriga qarab algoritmlar keng tarqalgan bo'lib, ular ish boshida har bir ob'ektni alohida klasterga joylashtiradi va keyin barcha namuna olish ob'ektlari bitta klasterda bo'lguncha klasterlarni tobora kattaroq qilib birlashtiradi. Shunday qilib, ichki bo'linmalar tizimi quriladi. Bunday algoritmlarning natijalari odatda daraxt-Dendrogramma shaklida bo'ladi. Bunday daraxtning klassik namunasi hayvonlar va o'simliklarning tasnifidir.

Klasterlar orasidagi masofani hisoblash uchun hamma ko'pincha ikkita masofadan foydalanadi: bitta bog'lanish yoki to'liq bog'lanish (klasterlar orasidagi masofa o'lchovlarini ko'rib chiqishga qarang).

Ierarxik algoritmlarning kamchiliklari to'liq bo'limlar tizimini o'z ichiga oladi, bu hal qilinayotgan muammo kontekstida ortiqcha bo'lishi mumkin.

KVADRAT XATO ALGORITMLARI

Klasterlash muammosini ob'ektlarning guruhlarga optimal bo'linishini qurish deb hisoblash mumkin. Bunday holda, optimallikni standart bo'linish xatosini minimallashtirish talabi sifatida aniqlash mumkin:

$$e^2(X, L) = \sum_{j=1}^K \sum_{i=1}^{n_j} \|x_i^{(j)} - c_j\|^2$$

bu erda c_j -j klasterining "massa markazi" (berilgan Klaster uchun o'rtacha xarakterli nuqta).

Kvadratik xato algoritmlari tekis algoritmlar turiga kiradi. Ushbu toifadagi eng keng tarqalgan algoritmlar k-o'rta usuli hisoblanadi. Ushbu algoritmlar bir-biridan iloji boricha uzoqroq joylashgan ma'lum miqdordagi klasterlarni yaratadi. Algoritmning ishlashi bir necha bosqichlarga bo'linadi:

- Klasterlarning boshlang'ich "massa markazlari" bo'lgan k nuqtalarini tasodifan tanlang.

- Har bir ob'ektni eng yaqin "massa markazi" bo'lgan klasterga olib boring.
- Klasterlarning "massa markazlari" ni hozirgi tarkibiga qarab qayta hisoblang.
- Agar algoritmni to'xtatish mezonlari bajarilmasa, 2-bandga qayting.

Algoritmning ishlashini to'xtatish mezonlari sifatida odatda standart xatoning minimal o'zgarishi tanlanadi. Agar 2-bosqichda klasterdan klasterga o'tgan ob'ektlar bo'lmasa, algoritmni to'xtatish ham mumkin.

Ushbu algoritmning kamchiliklari bo'linish uchun klasterlar sonini belgilash zarurligini o'z ichiga oladi.

LOYQA ALGORITMLAR

Eng mashhur loyqa klasterlash algoritmi C-o'rta (C-means) algoritmidir. Bu k-o'rta usulining modifikatsiyasi. Algoritmning ishlash bosqichlari:

- Boshlang'ich loyqa bo'linishni tanlang n ob'ektlar k ga tegishli matritsani tanlash orqali klasterlar u hajmi n x k.

- U matritsadan foydalanib, noaniq xato mezonining qiymatini toping:

$$E^2(X, U) = \sum_{i=1}^N \sum_{k=1}^K U_{ik} \|x_i^{(k)} - c_k\|^2$$

bu erda C_k loyqa k klasterining "massa markazi" dir:

$$c_k = \sum_{i=1}^N U_{ik} x_i$$

- Loyqa xato mezonining ushbu qiymatini kamaytirish uchun ob'ektlarni qayta guruhlash.

- Qaytish p. 2 u matritsasi o'zgarishlar ahamiyatsiz bo'lguncha.

Agar klasterlar soni oldindan ma'lum bo'lmasa yoki har bir ob'ektni bitta klasterga noyob tarzda kiritish zarur bo'lsa, ushbu algoritmlar mos kelmasligi mumkin.

GRAFIK NAZARIYASIGA ASOSLANGAN ALGORITMLAR

Bunday algoritmlarning mohiyati shundaki, ob'ektlarni tanlash g \ u003d(V, E) grafik shaklida taqdim etiladi, uning tepalari ob'ektlar bilan mos keladi va qirralarning og'irligi ob'ektlar orasidagi "masofa" ga teng. Grafik klasterlash algoritmlarining afzalligi-bu aniqlik, amalga oshirishning nisbatan soddaligi va geometrik mulohazalarga asoslangan turli xil yaxshilanishlarni kiritish imkoniyati. Asosiy algoritmlar-bu bog'langan tarkibiy qismlarni ajratish algoritmi, minimal qoplama (skelet) daraxtini qurish algoritmi va qatlamli klasterlash algoritmi.

BOG'LANGAN TARKIBIY QISMLARNI AJRATISH ALGORITMI

Bog'langan tarkibiy qismlarni tanlash algoritmidagi kirish parametri r o'rnatiladi va grafada "masofalar" R dan katta bo'lgan barcha qirralar o'chiriladi, faqat eng yaqin ob'ektlar juftlari ulanadi. Algoritmning ma'nosi shundaki, unda joylashgan R qiymatini tanlash barcha "masofalar" oraligi, unda grafik bir nechta ulangan komponentlarga "parchalanadi". Olingan komponentlar klasterlardir.

R parametrini tanlash uchun odatda juft masofalar taqsimotining gistogrammasi tuziladi. Yaxshi aniqlangan klasterli ma'lumotlar tuzilishiga ega bo'lgan muammolarda gistogrammada ikkita tepalik bo'ladi – biri Klaster ichidagi masofalarga, ikkinchisi klasterlararo masofalarga to'g'ri keladi. R parametri ushbu cho'qqilar orasidagi minimal zonadan tanlanadi. Shu bilan birga, masofa chegarasi yordamida klasterlar sonini boshqarish juda qiyin.

Kvadrat xato algoritmlari

Klasterlash muammosini ob'ektlarning guruhlariga optimal bo'linishini qurish deb hisoblash mumkin. Bunday holda, optimallikni standart bo'linish xatosini minimallashtirish talabi sifatida aniqlash mumkin:

$$e^2(X, L) = \sum_{j=1}^K \sum_{i=1}^{n_j} \|x_i^{(j)} - c_j\|^2$$

bu erda c_j -j klasterining "massa markazi" (berilgan Klaster uchun o'rtacha xarakterli nuqta).

Klasterlarni birlashtirish

Ierarxik algoritmlardan foydalanganda klasterlarni qanday birlashtirish, ular orasidagi "masofalarni" qanday hisoblash kerakligi haqida savol tug'iladi. Bir nechta ko'rsatkichlar mavjud:

- Yagona bog'lanish (eng yaqin qo'shni masofalar) ushbu usulda ikkita Klaster orasidagi masofa turli klasterlardagi eng yaqin ikkita ob'ekt (eng yaqin qo'shnilar) orasidagi masofa bilan belgilanadi. Natijada paydo bo'lgan klasterlar zanjirlarga birlashishga moyildirlar.

- To'liq bog'lanish (eng uzoq qo'shnilarning masofasi) ushbu usulda klasterlar orasidagi masofa turli klasterlardagi har qanday ikkita ob'ekt

orasidagi eng katta masofa (ya'ni eng uzoq qo'shnilar) bilan belgilanadi. Ob'ektlar alohida guruhlardan kelganda, bu usul odatda juda yaxshi ishlaydi. Agar klasterlar cho'zilgan shaklga ega bo'lsa yoki ularning tabiiy turi "zanjirli" bo'lsa, unda bu usul yaroqsiz.

- Og'irlashtirilmagan juftlik o'rtacha

ushbu usulda ikki xil Klaster orasidagi masofa ulardagi barcha juft ob'ektlar orasidagi o'rtacha masofa sifatida hisoblanadi. Usul ob'ektlar turli guruhlarni tashkil qilganda samarali bo'ladi, ammo u kengaytirilgan ("zanjirli" turdagi) klasterlarda ham bir xil darajada yaxshi ishlaydi.

- Og'irlikdagi juftlik o'rtacha

usul vaznsiz juftlik o'rtacha usuli bilan bir xil, faqat hisoblashda tegishli klasterlarning kattaligi (ya'ni ulardagi ob'ektlar soni) og'irlik koeffitsienti sifatida ishlatiladi. Shuning uchun, bu usul klasterlarning teng bo'lmagan o'lchamlari taxmin qilinganda ishlatilishi kerak.

- Tortilmagan sentroid usuli

ushbu usulda ikkita Klaster orasidagi masofa ularning tortishish markazlari orasidagi masofa sifatida aniqlanadi.

- Vaznli sentroid usuli (median)

bu usul avvalgisiga o'xshaydi, faqat hisob-kitoblar Klaster o'lchamlari orasidagi farqni hisobga olish uchun og'irliklardan foydalanadi. Shuning uchun, agar klasterlar hajmida sezilarli farqlar mavjud bo'lsa yoki shubha qilinsa, bu usul avvalgisiga qaraganda afzalroq bo'ladi.

FOYDALANILGAN MANBALAR RO'YXATI:

1. Berikov V. S. Klaster tahlilining zamonaviy tendentsiyalari / V. S. Berikov, G. S. Lbov. - "Axborot va telekommunikatsiya tizimlari" ustuvor yo'nalishi bo'yicha ko'rib chiqish va tahliliy maqolalarning Butunrossiya tanlovi, 2008. - 26 p.

2. Ganenkova E. G. Funktsional tahlil: bo'shliqlarning asosiy sinflari / E. G. Ganenkova, K. F. Amozova. - Petrozavodsk: PetrGU, 2013. - 26 p.

3. Ershov, K. S. Klasterlash algoritmlarini tahlil qilish va tasniflash / K. S.

Ershov, T. N. Romanova. // Avtomatlashtirilgan tizimlarda yangi axborot texnologiyalari. - 2016 yil - 19-son. - S. 274-279.

4. Kotov, A. Ma'lumotlarni klasterlash [Elektron resurs]. – Kirish rejimi: <http://logic.pdmi.ras.ru/~yura/internet/02ia-seminar-note.pdf>.

5. Mandel, I. D. Klaster tahlili / I. D. Mandel. - M.: Moliya va statika, 1988. - 176 b.

6. Protasov, C.C. Qanday qilib katta ma'lumotlar IT sanoatining eng qiziqarli vazifalaridan biriga aylandi [Elektron resurs]. – Kirish rejimi: <http://andrew--r.github.io/bigdata/>.

7. Suslov, S. A. Klaster tahlili: mohiyati, afzalliklari va kamchiliklari

/ S. A. Suslov. // Vestnik NGIEI. - 2011. - 1-son. - S. 51-56.

8. Blanko-Silva, F.J. Raqamli va ilmiy hisoblash uchun fanni o'rganish /

F. J. Blanko-Silva. – Packt nashriyoti, 2015. – 150 b.