

TASNIFLASH VA KLASTERLASH

Onarqulov Maqsadjon Karimberdiyevich

Farg'ona davlat universiteti amaliy matematika va informatika kafedrası dotsenti

maxmaqsad@gmail.com**Abduhakimova Xusnora Abdusamad qizi**

Farg'ona davlat universiteti talabasi

abdujabborparpiyev6@gmail.com

Annotatsiya: Ushbu tezısdı tasnıflash va klasterlash usulları ma'lumotlarnı tahlil qilishda qo'llaniladigan muhim texnikalar sifatida ko'rib chiqiladi. Tasnıflash nazorat ostidagi o'rgatish usuli bo'lib, ma'lumotlarnı oldından belgilangan kategoriyalarga ajratishni o'z ichiga oladi. Bu jarayonda o'rgatish namunalari asosida model yaratiladi va yangi ma'lumotlar tegishli sinflarga taqsimlanadi. Klasterlash esa nazorat ostida bo'lmagan o'rgatish usuli bo'lib, ma'lumotlarnı yashirin o'xshashliklar yoki tuzilmalarga asoslangan holda guruhlash imkonini beradi. Ushbu usul K-means, hierarchik klasterlash va DBSCAN kabi algoritmlar yordamida amalga oshiriladi. Maqolada tasnıflash va klasterlashning marketing, tibbiyot, texnologik nosozliklarnı aniqlash kabi sohalardagi qo'llanilishi ko'rib chiqilib, ushbu texnologiyalarning samaradorlik va ilmiy-texnik rivojlanishga qo'shayotgan hissasi yoritilgan. Ma'lumotlarnı tahlil qilishdagi bu usullar bilimlarnı aniqlash va qaror qabul qilish jarayonini yanada takomillashtiradi.

Kalit so'zlar: Tasnıflash, klasterlash, ma'lumotlarnı tahlil qilish, nazorat ostida o'rgatish, nazorat ostida bo'lmagan o'rgatish, kategoriyalar, K-means, Hierarchik klasterlash, DBSCAN.

Аннотация: В данной статье классификация и кластеризация рассматриваются как важные методы анализа данных. Классификация представляет собой метод обучения с учителем, который включает в себя распределение данных по заранее определённым категориям. В этом процессе на основе обучающих выборок создаётся модель, которая распределяет новые данные по соответствующим классам. Кластеризация, напротив, является методом обучения без учителя, который позволяет группировать данные на основе скрытых сходств или структур. Этот метод реализуется с использованием алгоритмов, таких как K-means, иерархическая кластеризация и DBSCAN. В статье рассмотрено применение классификации и кластеризации в таких областях, как маркетинг, медицина и выявление технических неисправностей, а также освещён их вклад в повышение эффективности и научно-технический прогресс. Эти методы анализа данных способствуют выявлению знаний и улучшению процесса принятия решений.

Ключевые слова: классификация, кластеризация, анализ данных, обучение с учителем, обучение без учителя, категории, K-means, иерархическая кластеризация, DBSCAN.

Annotation:

This article discusses classification and clustering as essential techniques for data analysis. Classification is a supervised learning method that involves categorizing data into predefined classes. During this process, a model is created based on training samples to assign new data to appropriate categories. Clustering, on the other hand, is an unsupervised learning method that groups data based on hidden similarities or structures. This method is implemented using algorithms such as K-means, hierarchical clustering, and DBSCAN. The article explores the applications of classification and clustering in fields such as marketing, medicine, and fault detection, highlighting their contribution to improving efficiency and scientific-technical

progress. These data analysis techniques enhance knowledge discovery and decision-making processes.

Keywords: classification, clustering, data analysis, supervised learning, unsupervised learning, categories, K-means, hierarchical clustering, DBSCAN.

Klasterlash usuli - bu bir guruhdagi ob'ektlar to'plamini boshqa tarmoqlardagi ob'ektlarga qaraganda bir-biriga o'xshash tarzda guruhlash vazifasi. Bu ma'lumotlarni qidirishning asosiy vazifasi va ko'plab sohalarda qo'llaniladigan umumiy statistik tahlil texnikasi, jumladan, mashinani o'rganish, naqshni aniqlash, tasvirni aniqlash, axborotni qidirish, ma'lumotlarni siqish va kompyuter grafikasi. Klasterlash usulining o'zi bitta aniq algoritmlar emas, balki hal qilinishi kerak bo'lgan umumiy vazifadir. Bunga guruh nimadan iboratligini va uni qanday samarali topishni tushunishda sezilarli darajada farq qiluvchi turli xil algoritmlar yordamida erishish mumkin. Metamavzularni shakllantirish uchun klasterlash usulidan foydalanish guruhdan foydalanishni o'z ichiga oladigan orasidagi kichik masofalar, bo'shliqning zich hududlari, intervallar yoki ma'lum statistik taqsimotlar. Shuning uchun klasterlash ko'p maqsadli optimallashtirish muammosi sifatida shakllantirilishi mumkin. Tegishli usul va parametr sozlamalari (jumladan, foydalaniladigan masofa funksiyasi, zichlik chegarasi yoki kutilayotgan klasterlar soni kabi elementlar) individual ma'lumotlar to'plamiga va natijalardan maqsadli foydalanishga bog'liq. Tahlil avtomatik vazifa emas, balki bilimlarni kashf qilish yoki interaktiv ko'p maqsadli optimallashtirishning iterativ jarayonidir. Ushbu klasterlash usuli sinov va xato urinishlarini o'z ichiga oladi. Natija kerakli xususiyatlarga erishilgunga qadar tez-tez ma'lumotlarni qayta ishlash va model parametrlarini o'zgartirish kerak bo'ladi. "Klasterlash" atamasi bilan bir qatorda avtomatik tasnif, sonli taksonomiya, botriologiya va tipologik tahlil kabi ma'nolari o'xshash bir qancha so'zlar ham mavjud. Nozik farqlar ko'pincha metasub'ekt munosabatlarini shakllantirish uchun klasterlash usulidan foydalanishda yotadi. Ma'lumot olishda natijada paydo bo'lgan guruhlar qiziqish uyg'otsa-da, avtomatik tasniflashda bu funktsiyalarni bajaradigan diskriminatsion kuch allaqachon mavjud.

Klaster tahlili 1932-yilda Kroberning ko'plab ishlariga asoslangan edi. U 1938 yilda Zubin va 1939 yilda Robert Tryon tomonidan psixologiyaga kiritilgan. Va bu ishlardan Cattell 1943 yildan beri nazariy jihatdan klasterlash usullarining tasnifini ko'rsatish uchun ishlatib kelinmoqda. "Klaster" tushunchasiga aniq ta'rif berib bo'lmaydi. Bu juda ko'p klasterlash usullari mavjudligining sabablaridan biridir. Umumiy maxraj mavjud: ma'lumotlar ob'ektlari guruhi. Biroq, turli tadqiqotchilar turli modellardan foydalanadilar. Klasterlash usullaridan foydalanishning har biri turli ma'lumotlarni o'z ichiga oladi. Turli xil algoritmlar tomonidan topilgan tushuncha o'z xususiyatlarida sezilarli darajada farqlanadi.

Klasterlash usuli-ushbu ittifoq o'z nomi bilan ham tanilgan, ierarxik model. U ob'ektlar uzoqroqda joylashgan qismlarga qaraganda qo'shni qismlarga ko'proq bog'langan degan odatiy fikrga asoslanadi. Bu algoritmlar ob'ektlarni bir-biriga bog'lab, ularning masofasiga qarab turli klasterlarni hosil qiladi. Klasterlashning asosiy xususiyatlari:

1. Nazoratsiz o'rganish: Ma'lumotlarda oldindan belgilar (label) mavjud emas.
2. Guruhlash: Ma'lumot bir-biriga o'xshashlik yoki masofaga qarab guruhlanadi.
3. O'xshashlik o'lchovi: Klasterlashda ob'ektlar orasidagi o'xshashlik masofa (masalan, Evklid masofasi) orqali aniqlanadi.

Klasterlashning qo'llanilishi:

Mijoz segmentatsiyasi: Xaridorlarni ularning xarid qilish xatti-harakatlariga qarab guruhlash.

Bozor tahlili: Bir xil xususiyatga ega mahsulotlarni bir guruhga ajratish.

Tibbiyot: Kasallikni simptomlariga qarab klasterlash.

Rasm va video tahlili: Bir xil ob'ektlar mavjud rasmlarni guruhlash.

Geografiya: Shaharlarni iqlim, aholi yoki hudud xususiyatlariga qarab bo'lish.

Mashhur klasterlash algoritmlari:

1. K-means: Klasterlar markazini aniqlab, ma'lumotlarni unga yaqinlik bo'yicha ajratadi.
2. Ierarxik klasterlash (Hierarchical Clustering): Daraxt shaklidagi klasterlar tuzadi.
3. DBSCAN (Density-Based Spatial Clustering): Yig'ilgan (kuchli zichlikdagi) klasterlarni aniqlaydi.
4. Gaussian Mixture Models: Ma'lumotlarni statistik ehtimollik asosida bo'ladi.

Guruhni asosan klasterning turli qismlarini ulash uchun zarur bo'lgan maksimal masofa bilan tavsiflash mumkin. Barcha mumkin bo'lgan masofalarda dendrogramma yordamida ifodalanishi mumkin bo'lgan boshqa guruhlar hosil bo'ladi. Bu "ierarxik klasterlash" umumiy nomi qayerdan kelganligini tushuntiradi. Ya'ni, bu algoritmlar ma'lumotlar to'plamining yagona bo'limini ta'minlamaydi, aksincha, vakolatning keng tartibini ta'minlaydi. Unga rahmat, ma'lum masofalarda bir-biri bilan drenaj mavjud. Dendrogrammada y o'qi klasterlarning birikish masofasini bildiradi. Guruhlar aralashmasligi uchun ob'ektlar X chizig'i bo'ylab joylashtirilgan.

Ulanishga asoslangan klasterlash - bu masofalarni hisoblashda farq qiluvchi usullarning butun oilasi. Masofa funktsiyalarining odatiy tanlovidan tashqari, foydalanuvchi ulanish mezonini ham hal qilishi kerak. Klaster bir nechta ob'ektlardan iborat bo'lganligi sababli, uni hisoblashning ko'plab variantlari mavjud. Ommabop tanlov bitta tutqichli guruhlash deb nomlanadi, bu usulUPGMA yoki WPGMA ni o'z ichiga olgan to'liq havola (o'rtacha arifmetik bo'lgan, shuningdek, o'rtacha havola klasteri sifatida ham tanilgan, vaznsiz yoki vaznli juftlik ansambli). Bundan tashqari, ierarxik tizim aglomerativ (alohida elementlardan boshlab va ularni guruhlariga birlashtirish) yoki bo'linuvchi (to'liq ma'lumotlar to'plamidan boshlab va uni bo'limlarga bo'lish) bo'lishi mumkin.

DBSCAN-ning yana bir qiziqarli xususiyati shundaki, uning murakkabligi ancha past - bu ma'lumotlar bazasiga nisbatan chiziqli qator so'rovlarni talab qiladi. Va g'ayrioddiy narsa shundaki, u har bir ishda deyarli bir xil natijalarni topadi (bu asosiy va shovqin nuqtalari uchun deterministik, lekin chegara elementlari uchun emas). Shuning uchun uni bir necha marta ishga tushirishning hojati yo'q.

DBSCAN va OPTICS ning asosiy kamchiligi shundaki, ular klaster chegaralarini aniqlash uchun zichlikning biroz pasayishini kutishadi. Masalan, Gauss taqsimotlari bir-biriga o'xshash bo'lgan ma'lumotlar to'plamlarida - sun'iy ob'ektlar uchun umumiy foydalanish holati - bu algoritmlar tomonidan yaratilgan klaster chegaralari ko'pincha o'zboshimchalik bilan ko'rinadi. Bu guruhlarining zichligi doimiy ravishda kamayib borayotganligi sababli sodir bo'ladi. Va Gauss aralashmasi ma'lumotlar to'plamida bu algoritmlar deyarli har doim bunday tizimlarni aniq modellashtirishga qodir bo'lgan EM klasterlash kabi usullardan ustun turadi.

O'rtacha siljish - bu klasterlash usuli bo'lib, unda har bir ob'yekt butun yadroni baholash asosida yaqin atrofdagi eng zich hududga o'tadi. Oxir-oqibat, ob'ektlar mahalliy o'tkazmaslik maksimumlariga yaqinlashadi. K-vositalari klasterlashiga o'xshab, bu "zichlik jalb qiluvchilar" ma'lumotlar to'plamining vakili bo'lib xizmat qilishi mumkin. Ammo o'rtacha siljishDBSCAN ga o'xshash o'zboshimchalik bilan shakllangan klasterlarni aniqlay oladi. Qimmatbaho iterativ protsedura va zichlikni baholash tufayli o'rtacha siljish odatda DBSCAN yoki k-Means ga qaraganda sekinroq bo'ladi. Bundan tashqari, odatdagi siljish algoritmini yuqori o'lchamli ma'lumotlarga qo'llash yadro zichligi taxminining bir xil bo'lmagan harakati tufayli qiyin, bu klaster dumlarining haddan tashqari parchalanishiga olib keladi.

Tasniflash (classification) – ma'lumotlarni oldindan belgilangan toifalarga ajratish jarayonidir. Bu usul nazoratli o'rganishga asoslangan bo'lib, ma'lumotning har bir bo'lagi o'ziga xos toifa (klass) bilan belgilanadi.

Tasniflashning asosiy xususiyatlari:

1. Supervayz o'rganish: Modelni o'rgatish uchun belgilangan ma'lumot (label) kerak bo'ladi.
2. Toifalar: Oldindan aniqlangan kategoriyalar mavjud.
3. Masala turi: Model kiritilgan ma'lumotni qaysi kategoriya (klass)ga tegishli ekanini taxmin qiladi.

Qanday ishlaydi?

1. Ma'lumotlarni tayyorlash: Ma'lumotlar to'plami kirish xususiyatlari (features) va natija toifasiga (label) bo'linadi.
2. Modelni o'rgatish: Belgilangan ma'lumotlar yordamida algoritmlar o'qitiladi.
3. Test qilish: Model yangi ma'lumotni sinab ko'rish orqali to'g'ri toifalashni o'rganadi.
4. Foydalanish: Model yangi kelgan ma'lumotni avtomatik toifalash uchun ishlatiladi.

Tasniflashning mashhur algoritmlari:

1. Logistik regressiya
2. K yaqin qo'shnilar usuli (KNN)
3. Qaror daraxtlari (Decision Trees)
4. Random Forest
5. Neyron tarmoqlar (Neural Networks)

XULOSA:

Tasniflash va klasterlash texnikalari ma'lumotlar tahlilining nazoratli va nazoratsiz o'rganish usullariga asoslanib, bir-birini to'ldiradi. Tasniflash o'z aniq maqsadlariga erishish uchun oldindan belgilangan toifalardan foydalansa, klasterlash yangi ma'lumotlar guruhlarini aniqlash uchun ishlatiladi. Zamonaviy texnologiyalar bilan birga ushbu usullar tibbiyot, savdo, marketing, muhandislik, ilmiy tadqiqot va boshqa sohalarda keng qo'llanilmoqda. Tasniflash va klasterlash algoritmlarini samarali qo'llash orqali katta ma'lumotlardan (big data) yanada ko'proq foydali ma'lumotlar olish mumkin. Kelajakda ushbu usullarni rivojlantirish sun'iy intellekt, mashinaviy o'qitish va avtomatlashtirilgan tizimlar samaradorligini yanada oshiradi.

FOYDALANILGAN ADABIYOTLAR

1. Daliev, K. S., Ahmedov, M. M., & Onarkulov, M. K. (2021). INFLUENCE OF THE TEMPERATURE AND CYCLIC DEFORMATIONS OF $[(Bi_{x}Sb_{1-x})_2Te_3]$ FILMS ON THEIR RESISTANCE. Journal of Engineering Physics and Thermophysics, 94(5), 1369-1374.
2. Онаркулов, М. (2021). Influence of the temperature and cyclic deformations of $(Bi_xSb_{1-x})_2Te_3$ films on their resistance. Инженерно-Физический журнал. Беларусь, 94(5), 1403-1408.
3. Onarkulov MK, Otajonov SM. DEVICE FOR STUDYING TENZE SENSITIVITY IN PHOTSENSITIVE SEMICONDUCTOR FILMS. Euroasian Journal of Semiconductors Science and Engineering. 2021;3(1):5.
4. Далиев, Х. С., & Онаркулов, М. К. (2020). ДИФФУЗИОННЫЕ ПРОЦЕССЫ В ПОЛУПРОВОДНИКОВЫХ МИКРОСТРУКТУРАХ НА ОСНОВЕ A_4B_6 . О „ZBEKISTON RESPUBLIKASI OLIY VA O „RTA MAXSUS TA“ LIM VAZIRLIGI O „ZBEKISTON RESPUBLIKASI INNOVATSION RIVOJLANTIRISH VAZIRLIGI, 44.

5. Otazhonov, S. M., Onarkulov, M., Botirov, K., Yunusov, N., Mamadzhonov, U., & Kakhkhorova, B. (2020). Device for studying tenze sensitivity in photosensitive semiconductor films. *Universum technical science*, 2(71), 2.
6. Akhmedov, M. M., Gaynazarova, K. I., Kadyrov, K. S., & Onarkulov, M. K. (2020). The chemical composition of strainsensitive films based on the Bi–Sb–Te system, *Universum. Tech. Sci.: Electron. Sci. J*, (2).
7. Daliev, K. S., Azimov, T. M. R., & Onarkulov, M. K. (2020). Influence of doping impurities on mechanical and thermoelectric properties of cooling thermoels based on bismuth and antimony chalcogenides. *Euroasian Journal of Semiconductors Science and Engineering*, 2(1), 1.