

**OB'EKTNING O'XSHASHLIGI O'LCHOVLARI VA MA'LUMOTLARNI
KLASTERLASH MASALASI**

Rasuljonova Zarnigor Egamnazar qizi
Farg'ona davlat universiteti 2-kurs talabasi
zarnigorrasuljonova@gmsun`iy intellektl.com

Anotatsiya: Ushbu maqolada sun'iy intellektning muhim sohalaridan biri - obyektning o'xshashligi va ularni klasterlarga ajratish masalalari batafsil yoritilgan. Dastlab, obyektning o'xshashlik darajasini baholash uchun ishlatiluvchi o'lchamlar, jumladan, Evklidov masofasi, Manxetten masofasi va kosinus o'xshashlik o'lchami kabi eng mashhur usullar tavsiflab o'tilgan.

Kalit so'zlar: Ob'ektlarning o'xshashligi, ma'lumotlarni klasterlash, sun'iy intellekt, o'xshashlik o'lchovlari, klasterlash algoritmlari, yuqori o'lchamli ma'lumotlar, aralash ma'lumotlar turlari, shovqin va chegaralar, dinamik va rivojlanayotgan ma'lumotlar, bashoratli modellashtirish, anomaliyalarni aniqlash, yashirin naqshlar.

Abstract: This article describes in detail one of the important areas of artificial intelligence - the similarity of objects and their classification into clusters. First, the measures used to evaluate the degree of similarity between objects are described, including the most popular methods, such as the Euclidean distance, Manhattan distance, and the cosine similarity measure.

Keywords: Object similarity, data clustering, artificial intelligence, similarity measures, clustering algorithms, high-dimensional data, mixed data types, noise and thresholds, dynamic and evolving data, predictive modeling, anomaly detection, hidden patterns.

Аннотация: В данной статье подробно описано одно из важных направлений искусственного интеллекта — сходство объектов и их классификация на кластеры. Сначала описываются меры, используемые для оценки степени сходства между объектами, включая наиболее популярные методы, такие как евклидово расстояние, манхэттенское расстояние и косинусная мера сходства.

Ключевые слова: Сходство объектов, кластеризация данных, искусственный интеллект, меры сходства, алгоритмы кластеризации, многомерные данные, смешанные типы данных, шум и пороги, динамические и развивающиеся данные, прогнозное моделирование, обнаружение аномалий, скрытые закономерности.

Sun'iy intellekt sohasida asosiy muammolardan biri bu ob'ektlarni ular orasidagi o'xshashliklarni aks ettiradigan tarzda qanday qilib guruhlash yoki klasterlashni tushunishdir. Bu muammo tasvir aniqlashdan tortib ma'lumot olish va bashoratli modellashtirishgacha bo'lgan ko'plab SUN'YI INTELLEKT ilovalarining markazida yotadi. Samarali klasterlashning kaliti ob'ektning o'xshashligini o'lchash va solishtirish mumkin bo'lgan tegishli o'lchamlarni aniqlashdan iborat.

Asosan, klasterlash ob'ektlar to'plamini kichik to'plamlarga yoki "klasterlarga" bo'lishni o'z ichiga oladi, shunda har bir klaster ichidagi ob'ektlar boshqa klasterlardagi ob'ektlarga qaraganda bir-biriga o'xshashroq bo'ladi. Biroq, "o'xshashlik" ta'rifi oddiy emas, chunki u muayyan muammo sohasiga va tahlil qilinadigan ma'lumotlarning xususiyatlariga bog'liq.

Obyektlarning o'xshashlik o'lchovlari

Sun'iy intellektda da o'xshashlik o'lchovlari ob'ektlar orasidagi o'xshashlik darajasini aniqlash uchun ishlatilishi mumkin bo'lgan o'lchanadigan xususiyatlar yoki atributlarni anglatadi. Ushbu o'lchamlarni ikki turga bo'lish mumkin: raqamli va kategorik.

Raqamli o'lchamlar:

Raqamli o'lchamlar matematik operatsiyalar yordamida to'g'ridan-to'g'ri taqqoslanadigan miqdoriy o'lchovlarni o'z ichiga oladi. Bunga misollar kiradi:

1. Fazoviy o'lchamlar (masalan, uzunlik, kenglik, balandlik, koordinatalar)
2. Vaqtinchalik o'lchamlar (masalan, davomiylik, vaqt belgilari)
3. Skalyar xossalari (masalan, vazn, harorat, zichlik)
4. Olingan sonli belgilar (masalan, statistik momentlar, chastotalar)

Kategorik o'lchamlar:

Kategorik o'lchovlar sifatli yoki tavsiflovchi atributlarni o'z ichiga oladi, ularni raqamli jihatdan bevosita taqqoslab bo'lmaydi. Buning o'rniga, bu o'lchovlar korrelyatsiya, assotsiatsiya yoki masofa o'lchovlariga tayanadi. Bunga misollar kiradi:

1. Rang tavsiflovchilari (masalan, qizil, ko'k, yashil)
2. Matn yorliqlari yoki toifalari (masalan, mahsulot turlari, janrlari)
3. Strukturaviy yoki relyatsion atributlar (masalan, grafik aloqasi, ierarxik munosabatlar)
4. Olingan toifali xususiyatlar (masalan, bir-hot kodlash, o'rnatish)

Tegishli o'xshashlik o'lchovlarini aniqlash samarali klasterlash uchun juda muhimdir, chunki u ob'ektlarni qanday taqqoslash va guruhlashni belgilaydi. Ko'pgina real ilovalarda ma'lumotlarning to'liq murakkabligini olish uchun raqamli va kategorik o'lchamlarning kombinatsiyasi talab qilinishi mumkin.

Ma'lumotlarni klasterlash masalasi

Tegishli o'xshashlik o'lchovlari aniqlangandan so'ng, klasterlash jarayonining keyingi bosqichi ob'ektlarni ushbu o'lchamlar bo'yicha o'xshashligidan kelib chiqqan holda guruhlashdir. Biroq, bu vazifa ko'pincha bir necha omillar bilan murakkablashadi:

1. Yuqori o'lchamli ma'lumotlar: O'lchovlar soni ortib borishi bilan klasterlash masalasining murakkabligi eksponent ravishda o'sib boradi. "O'lchovlilik la'nati" deb nomlanuvchi bu hodisa klasterlar o'rtasida kamtarlik va kam diskriminatsiyaga olib kelishi mumkin, bu esa mazmunli naqshlarni aniqlashni qiyinlashtiradi.
2. Chiziqsiz aloqalar: Haqiqiy dunyoning ko'plab ma'lumotlar to'plamlari xususiyatlar o'rtasidagi chiziqli bo'lmagan munosabatlarni namoyish etadi, bu esa chiziqli bo'linish yoki oddiy geometrik shakllarni (masalan, sferik yoki elliptik klasterlar) qabul qiladigan an'anaviy klasterlash algoritmlarini chalkashtirib yuborishi mumkin.

3. Aralash ma'lumotlar turlari: Ma'lumotlar to'plamlari ko'pincha sonli va toifali o'lchamlarning aralashmasini o'z ichiga oladi, bu heterojen ma'lumotlar turlarini samarali boshqarish uchun maxsus texnikani talab qiladi.

4. Shovqin va chegaralar: Haqiqiy dunyo ma'lumotlari ko'pincha shovqin, etishmayotgan qiymatlar yoki o'xshashlik ko'rsatkichlarini buzishi va noto'g'ri klaster tayinlanishiga olib kelishi mumkin bo'lgan chegaralar bilan ifloslangan.

5. Klasterlarning o'zaro kelishi va ierarxialari: Ko'pgina domenlarda klasterlar bir-birining ustiga chiqishi yoki ierarxik tuzilmalarni namoyish qilishi mumkin, bu esa ma'lumotlarni aniq, yaxshi ajratilgan guruhlariga bo'lish vazifasini murakkablashtiradi.

6. Dinamik va rivojlanayotgan ma'lumotlar: Doimiy ravishda ma'lumotlar oqimi yoki mavjud ma'lumotlar vaqt o'tishi bilan rivojlanib borar ekan, klasterlash algoritmlari kontseptsianing siljishi va asosiy ma'lumotlar taqsimotidagi o'zgarishlarni boshqarish uchun moslashishi kerak.

Ushbu muammolarni hal qilish SUN`IY INTELLEKTda har birining o'ziga xos kuchli, zSun`iy intellektf tomonlari va taxminlariga ega bo'lgan keng ko'lamli klasterlash algoritmlari va usullarini keltirib chiqardi.

Sun`iy intellektda mashhur klasterlash algoritmlari

Sun`iy intellektda eng ko'p ishlatiladigan klasterlash algoritmlaridan ba'zilari:

1. K-Klasterlash vositalari: Eng oddiy va eng ko'p qo'llaniladigan algoritmlardan biri bo'lgan K-ma'lumot nuqtalarini iterativ ravishda eng yaqin klaster markaziga tayinlaydi va belgilangan nuqtalar asosida markazlarni yangilaydi.

2. Ierarxik klasterlash: Ierarxik klaster algoritmlari kichikroq klasterlarni kattaroqlarga birlashtirish (aglomerativ) yoki kattaroq klasterlarni kichikroqlarga (bo'linuvchi) bo'lish orqali klasterlar ierarxiasini yaratadi.

3. Zichlikka asoslangan klasterlash: DBSCAN va OPTICS kabi algoritmlar klasterlarni zichroq hududlar bilan ajratilgan zich hududlar sifatida aniqlaydi, bu ularni ixtiyoriy shakldagi klasterlar va shovqinlarni boshqarish uchun juda mos keladi.

4. Tarqatishga asoslangan klasterlash: Gauss aralashmasi modellari (GMMs) va Mean-Shift kabi ushbu algoritmlar ma'lumotlar nuqtalari asosiy ehtimollik taqsimotlari aralashmasidan hosil bo'lishini taxmin qiladi.

5. Spektral klasterlash: o'ziga xos tuzilishidan foydalanadi, ko'pincha chiziqli bo'lmagan va murakkab ma'lumotlarni taqsimlashda yaxshi ishlaydi.

6. Neyron tarmoqqa asoslangan klasterlash: Chuqur o'rganish arxitekturalari, masalan, avtokoderlar va o'z-o'zini tartibga soluvchi xaritalar (SOM) klasterlashni osonlashtiradigan ma'lumotlarning past o'lchamli tasvirlarini o'rganishi mumkin.

7. Ansambl va gibrid usullar: Bir nechta klasterlash algoritmlarini birlashtirish yoki klasterlashni boshqa texnikalar bilan birlashtirish (masalan, o'lchamlarni kamaytirish, zichlikni baholash) unumdorlik va mustahkamlikni oshirishi mumkin.

Klasterlash algoritmini tanlash ma'lumotlarning o'ziga xos xususiyatlariga, kerakli klaster xususiyatlariga va mavjud hisoblash resurslariga bog'liq. Bundan tashqari, ko'pgina

algoritmlar klasterlar soni yoki zichlik chegaralari kabi giperparametrlarni ehtiyotkorlik bilan sozlashni talab qiladi, bu esa klasterlash natijalari sifatiga sezilarli ta'sir ko'rsatishi mumkin.

Ma'lumotlarni klasterlash masalasi sun'iy intellektning asosiy muammosi bo'lib, turli sohalarda keng ko'lamli ilovalar mavjud. Tegishli o'xshashlik o'lchamlarini aniqlash va mos klasterlash algoritmlarini tanlash murakkab ma'lumotlar to'plamidan mazmunli tushunchalarni olishda muhim qadamdir.

Sun'iy intellekt tizimlari rivojlanishda davom etar ekan va tobora murakkab bo'lgan real muammolarni hal qilishda samarali klasterlash usullariga bo'lgan ehtiyoj ortib boradi. Davom etilayotgan tadqiqot sa'y-harakatlari yanada mustahkam va kengaytiriladigan algoritmlarni ishlab chiqish, aralash ma'lumotlar turlari va chiziqi bo'lmagan munosabatlarni qayta ishlash va domen bilimlari va inson tajribasini klasterlash jarayoniga integratsiyalashga qaratilgan.

Oxir oqibat, shunga o'xshash ob'ektlarni aniq aniqlash va guruhlash qobiliyati ko'plab sun'iy intellekt ilovalari uchun asosiy yordam beradi, bashoratli modellashtirish va anomaliyalarni aniqlashdan tavsiyalar tizimlari va shaxsiylashtirilgan tajribalargacha. Ma'lumotlarni klasterlash bilan bog'liq muammolarni hal qilish orqali tadqiqotchilar va amaliyotchilar yashirin naqshlarni ochish, qimmatli tushunchalarni olish va turli sohalarda innovatsiyalarni rivojlantirish uchun sun'iy intellektning to'liq imkoniyatlarini ochishlari mumkin.

Foydalanilgan adabiyotlar

1. MacQueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. 1. University of California Press. pp. 281–297.
2. Ester, M., Kriegel, H.P., Sander, J., & Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96).
3. Rokach, L., & M. Sun'iy intellektmon, O. (2005). Clustering Methods. In Data Mining and Knowledge Discovery Handbook (pp. 321-352). Springer..
4. Tan, P.N., Steinbach, M., & Kumar, V. (2005). Introduction to Data Mining. Pearson Education.
5. Aggarwal, C.C., & Reddy, C.K. (2013). Data Clustering: Algorithms and Applications. CRC Press.
6. Tojimatov, I. N., Olimov, A. F., Khaydarova, O. T., & Tojiboyev, M. M. (2023). CREATING A DATA SCIENCE ROADMAP AND ANALYSIS. PEDAGOGICAL SCIENCES AND TEACHING METHODS, 2(23), 242-250.
7. Тожимаматов, И. Н. (2023). ЗАДАЧИ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ. PEDAGOG, 6(4), 514-516.
8. Muqaddam, A., Shahzoda, A., Gulasal, T., & Isroil, T. (2023). NEYRON TARMOQLARDAN FOYDALANIB TASVIRLARNI ANIQLASH USULLARI. SUSTAINABILITY OF EDUCATION, SOCIO-ECONOMIC SCIENCE THEORY, 1(8), 63-74.
9. Raximov, Q. O., Tojimatov, I. N., & Xo, H. R. O. G. L. (2023). SUNIY NEYRON TARMOQLARNI UMUMIY TASNIFI. Scientific progress, 4(5), 99-107.
10. Ortiqovich, Q. R., & Nurmamatovich, T. I. (2023). NEYRON TARMOQNI O'QITISH USULLARI VA ALGORITMLARI. Scientific Impulse, 1(10), 37-46.

11. Tojimamatov, I. N., Mamalatipov, O., Rahmatjonov, M., & Farhodjonov, S. (2023). NEYRON TARMOQLAR. Наука и инновация, 1(1), 4-12.
12. Tojimamatov, I. N., Mamalatipov, O. M., & Karimova, N. A. (2022). SUN'IY NEYRON TARMOQLARINI O 'QITISH USULLARI. Oriental renaissance: Innovative, educational, natural and social sciences, 2(12), 191-203.
13. Muqaddam, A., Shahzoda, A., Gulasal, T., & Isroil, T. (2023). NEYRON TARMOQLARDAN FOYDALANIB TASVIRLARNI ANIQLASH USULLARI. SUSTAINABILITY OF EDUCATION, SOCIO-ECONOMIC SCIENCE THEORY, 1(8), 63-74.
14. Raximov, Q. O., Tojimamatov, I. N., & Xo, H. R. O. G. L. (2023). SUNIY NEYRON TARMOQLARNI UMUMIY TASNIFI. Scientific progress, 4(5), 99-107.
15. Raxmatjonova, M. N., & Tojimamatov, I. N. (2023). BIZNESDA SUNIY INTELEKT TECHNOLOGYALARI VA ULARNI AHAMIYATI. Лучшие интеллектуальные исследования, 11(3), 46-52.
16. Nurmamatovich, T. I. (2024, April). BIR QATLAMLI PERCEPTRONNI O 'QITISH. In " CANADA" INTERNATIONAL CONFERENCE ON DEVELOPMENTS IN EDUCATION, SCIENCES AND HUMANITIES (Vol. 17, No. 1).
17. Nurmamatovich, T. I. (2024, April). SUNIY NEYRONNING MATEMATIK MODELI HAMDA FAOLLASHTIRISH FUNKTSIYALARI. In " USA" INTERNATIONAL SCIENTIFIC AND PRACTICAL CONFERENCE TOPICAL ISSUES OF SCIENCE (Vol. 17, No. 1).