



KLASTERLASH. K -O'RTACHA ALGORITIMI. MAKSIMUM MASOFA ALGORITMI. ISODATA ALGORITMI. MAKSIMALLASHTIRISHNI QUTISH ALGARITMI.

Tojimamatov Israiljon Nurmatovich

Farg'ona davlat universiteti amaliy matematika
va informatika kafedrasi katta o'qituvchisi,
israiltojimatov@gmail.com

Tohirov Shohjahon Ilhomjon og'li

Farg'ona davlat universiteti 2-bosqich talabasi
shohjahontohirov31@gmail.com
<https://doi.org/10.5281/zenodo.11198157>

Annotatsiya

Ushbu maqola klasterlash metodlari va ularning turli sohalarda qo'llanilishini tahlil etadi. Klasterlash - bu ma'lumotlar to'plamini o'xshashlik yoki masofa asosida bir-biriga yaqin guruhlariga ajratish jarayoni. Bizning tadqiqotimiz klasterlashning o'rtacha algoritmi, maksimum masofa algoritmi, ISODATA algoritmi va maksimallashtirishni kutish algoritmi kabi asosiy usullariga qaratilgan. Har bir algoritmning matematik asoslari, afzalliklari va cheklovlari ko'rib chiqiladi, shuningdek ularning amaliy dasturlari misollar orqali tushuntiriladi.

Kalit So'zlar:

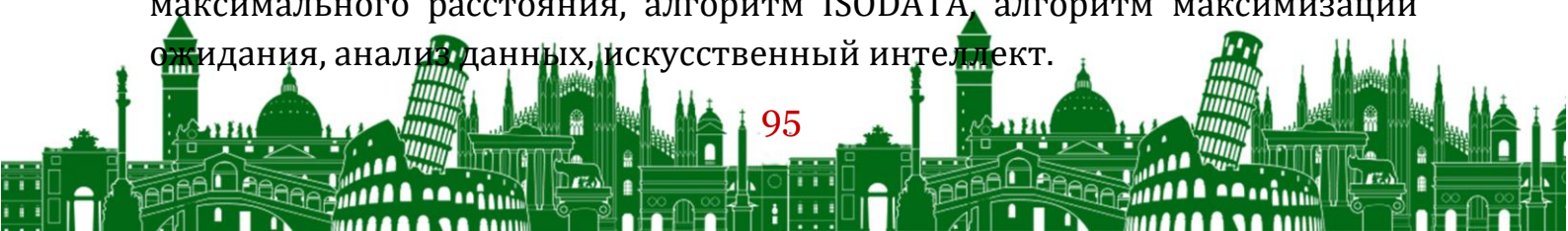
Klasterlash, O'rtacha algoritmi, Maksimum masofa algoritmi, ISODATA algoritmi, Maksimallashtirishni kutish algoritmi, Ma'lumotlar tahlili, Sun'iy intellekt.

Аннотация

В данной статье анализируются методы кластеризации и их применение в различных областях. Кластеризация — это процесс разделения набора данных на тесно связанные группы на основе сходства или расстояния. Наши исследования сосредоточены на основных методах кластеризации, таких как алгоритм усреднения, алгоритм максимального расстояния, алгоритм ISODATA и алгоритм ожидания максимизации. Рассмотрены математическая основа, преимущества и ограничения каждого алгоритма, а их практическое применение объяснено на примерах.

Ключевые слова:

Кластеризация, алгоритм среднего значения, алгоритм максимального расстояния, алгоритм ISODATA, алгоритм максимизации ожидания, анализ данных, искусственный интеллект.



Annotation

This article analyzes clustering methods and their applications in various fields. Clustering is the process of dividing a set of data into closely related groups based on similarity or distance. Our research focuses on the main methods of clustering, such as the averaging algorithm, the maximum distance algorithm, the ISODATA algorithm, and the maximization expectation algorithm. The mathematical basis, advantages, and limitations of each algorithm are reviewed, and their practical applications are explained through examples.

Key words:

Clustering, Average Algorithm, Maximum Distance Algorithm, ISODATA Algorithm, Maximization Expectation Algorithm, Data Analysis, Artificial Intelligence.

Kirish

Klasterlash - bu katta hajmdagi ma'lumotlar to'plamida mantiqiy guruhlar yoki "klasterlar"ni aniqlash uchun ishlatiladigan usullardan biridir. Bu jarayon ma'lumotlarni tahlil qilish, tasniflash va ular orasidagi o'xshashliklarni aniqlashda muhim ahamiyatga ega. Klasterlashning turli algoritmlari mavjud bo'lib, har biri ma'lum bir vazifani yechish uchun mo'ljallangan.

O'rtacha algoritmi

O'rtacha algoritmi (K-means algoritmi deb ham tanilgan) eng ko'p qo'llaniladigan klasterlash usullaridan biridir. Ushbu algoritmi har bir klaster uchun markaz (centroid) hisoblaydi va har bir nuqta uchun eng yaqin markazga asoslanib klasterlarni shakllantiradi. Algoritmning asosiy qadamlari quyidagicha:

1. Tasodifiy markazlarni tanlash.
2. Har bir nuqtani eng yaqin markazga ajratish.
3. Har bir klaster uchun yangi markazlarni hisoblash.
4. 2-3 qadamlarni klasterlar o'zgarmaguncha takrorlash.

Maksimum masofa algoritmi

Maksimum masofa algoritmi (Maximum distance algorithm), klasterlarni shakllantirishda nuqtalar orasidagi maksimal masofadan foydalanadi. Bu usul, o'rtacha algoritmdan farqli o'laroq, klasterlarning chegaralarini aniqroq belgilash imkonini beradi, ayniqsa ma'lumotlar to'plamida aniq ajratilgan guruhlar mavjud bo'lsa.

ISODATA algoritmi

ISODATA (Iterative Self-Organizing Data Analysis Technique Algorithm) algoritmi o'rtacha algoritmnining kengaytirilgan versiyasi hisoblanadi. U



ma'lumotlar to'plamining o'zgaruvchan tabiatini hisobga oladi va klasterlarni birlashtirish yoki bo'lish orqali dinamik ravishda moslashadi. Bu, murakkab ma'lumotlar to'plamlari bilan ishlashda qo'shimcha moslashuvchanlik beradi.

Maksimallashtirishni kutish algoritmi

Maksimallashtirishni kutish (Expectation Maximization, EM) algoritmi, ma'lum bir taqsimotga ega bo'lgan ma'lumotlar to'plamida klasterlarni topish uchun ishlatiladi. Bu yondashuv, har bir nuqtaning klasterga tegishlilikini ehtimoliylik asosida baholaydi va bu ehtimollarni maksimallashtirish orqali klasterlar markazlarini yangilaydi. EM algoritmi, ayniqsa, statistik modellash sohasida keng qo'llaniladi.

Xulosa

Klasterlash algoritmlari ma'lumotlar tahlilida muhim rol o'ynaydi va turli xil ilmiy va amaliy sohalarida keng qo'llaniladi. Har bir algoritmnining o'ziga xos xususiyatlari mavjud bo'lib, ularning qo'llanilishi ma'lumotlar to'plamining xususiyatlariga va tadqiqot maqsadlariga bog'liq. Ushbu algoritmlar sun'iy intellekt va ma'lumotlar tahlilining turli sohalarida muhim qo'llanma hisoblanadi.

Foydalanilgan adabiyotlar:

1. Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys (CSUR)*, 31(3), 264-323.
2. MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, 1(14), 281-297.
3. Ball, G. H., & Hall, D. J. (1965). *ISODATA, a novel method of data analysis and pattern classification*. Stanford Research Institute.
4. Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), 1-38.

