Efe can Dönmez

1220505047

Algoritma ödevi

dbscan tanıtım dokümanı

DBSCAN (Density-Based Spatial Clustering of Applications with Noise), yoğunluk tabanlı bir kümeleme algoritmasıdır. Veri madenciliği ve makine öğrenmesi alanında sıkça kullanılan bu algoritma, veri noktalarını yoğunluklarına ve komşuluk ilişkilerine dayalı olarak kümeleyen bir yaklaşımdır. DBSCAN algoritması, veri setinde farklı şekillerde ve yoğunluklarda kümeleme yapabilen esnek bir yapıya sahiptir. Bu araştırmada, DBSCAN algoritmasının çalışma prensipleri, avantajları, dezavantajları ve çeşitli uygulama alanları ayrıntılı bir şekilde incelenecektir.

DBSCAN Algoritmasının Çalışma Prensibi:

DBSCAN algoritması, iki temel parametre olan epsilon (ε) ve minimum nokta sayısı (MinPts) kullanır. Epsilon, bir noktanın komşuluk yarıçapını belirlerken, MinPts, bir noktanın çekirdek noktası olarak kabul edilebilmesi için gereken minimum komşu sayısını belirler.

Algoritmanın adımları şu şekildedir:

Başlangıçta, veri setindeki her nokta etiketsiz olarak işaretlenir.

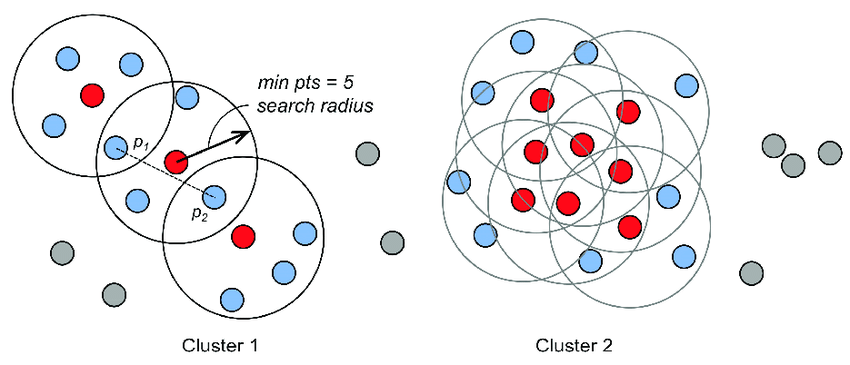
Bir nokta seçilir ve bu noktanın ε yarıçapı içindeki noktalar kontrol edilir. Eğer ε yarıçapı içindeki nokta sayısı MinPts'ten küçükse, bu nokta gürültü (noise) olarak işaretlenir.

Eğer ε yarıçapı içindeki nokta sayısı MinPts'ten büyük veya eşitse, bir yeni küme oluşturulur. Bu nokta çekirdek noktası olarak işaretlenir ve ε yarıçapı içindeki tüm noktaları bu kümeyle ilişkilendirir.

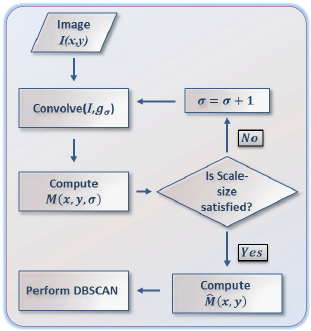
Her ilişkilendirilen noktanın ε yarıçapı içindeki noktaları kontrol edilir. Eğer bu noktalar çekirdek noktası değilse, gürültü olarak işaretlenir. Aksi halde, çekirdek noktasına ait kümeyle ilişkilendirilir.

İşlem tüm noktalara kadar tekrarlanır. Her nokta işaretlendiğinde, yeni bir küme oluşturulması veya gürültü olarak işaretlenmesi söz konusu olabilir.

DBSCAN algoritması, veri setindeki noktaların yoğunluklarına ve komşuluk ilişkilerine dayanarak kümeleme yapar. Çekirdek noktaları bir küme oluştururken, yoğunluğu düşük noktalar gürültü olarak işaretlenir. Çalışma prensibini daha iyi göstermek amaçlı aşağıda bulunan görsellere bakabilirsiniz;



Şekil 1: DBSCAN Mekanizması



Şekil 2: FLOWCHART ile DBSCAN

DBSCAN Algoritmasının Avantajları:

DBSCAN, küme sayısını önceden belirtme ihtiyacı olmadan otomatik olarak belirleyebilen bir algoritmadır. Veri setindeki yoğunluğa ve komşuluk ilişkilerine dayalı olarak farklı şekillerdeki kümelere uygulanabilir.

DBSCAN, aykırı değerleri tanımlama ve çıkarma yeteneği sayesinde veri temizleme süreçlerinde etkili bir araçtır.

DBSCAN, veri noktalarını doğrusal olmayan ilişkileri algılayarak farklı şekilli kümelere daha iyi adapte olabilir.

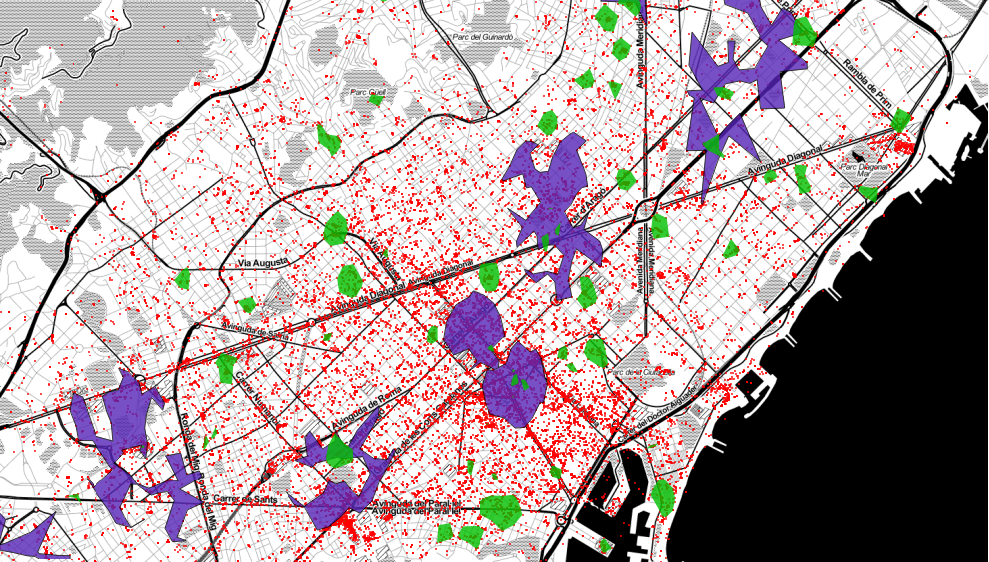
DBSCAN Algoritmasının Dezavantajları:

DBSCAN, epsilon (ε) ve minimum nokta sayısı (MinPts) parametrelerini doğru bir şekilde ayarlamak gerektirir. Bu parametrelerin yanlış ayarlanması, yanlış kümeleme sonuçlarına yol açabilir.

Veri setindeki yoğunluğun farklı olduğu alanlarda, epsilon (ε) parametresinin ayarlanması zor olabilir. Homojen olmayan yoğunluklara sahip veri setleri üzerinde DBSCAN performansı düşebilir.

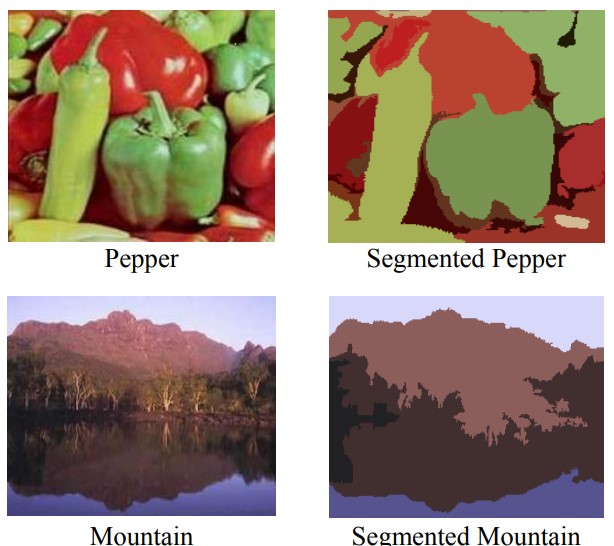
DBSCAN Algoritmasının Uygulama Alanları:

Coğrafi Veri Analizi: DBSCAN, coğrafi verilerdeki noktaların kümeleme analizinde sıkça kullanılır. Örneğin, hava istasyonları, jeolojik bölgeler veya nüfus yoğunluğu gibi coğrafi verilerdeki desenleri belirlemek için DBSCAN kullanılabilir.



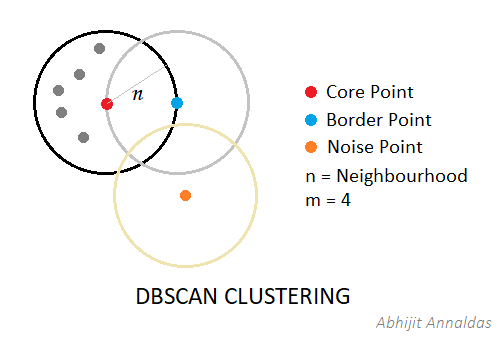
Şekil 3: DBSCAN kullanılarak elde edilmiş coğrafi veri

Görüntü İşleme: Görüntü işleme alanında, DBSCAN, benzer özelliklere sahip piksellerin kümeleme analizi için kullanılabilir. Nesne tespiti, görüntü segmentasyonu ve yüz tanıma gibi uygulamalarda DBSCAN yöntemleri başarılı sonuçlar vermiştir.



Şekil 4: DBSCAN ile Görüntü İşleme

Veri Madenciliği: DBSCAN, veri madenciliği uygulamalarında yaygın olarak kullanılan bir algoritmadır. Örneğin, müşteri segmentasyonu, pazar analizi, anomaliden kaçma ve kullanıcı davranışı analizi gibi alanlarda DBSCAN kullanılarak değerli bilgiler elde edilebilir.



Şekil 5: DBSCAN ve Veri Madenciliği

DBSCAN, yoğunluk tabanlı bir kümeleme algoritmasıdır ve veri noktalarını yoğunluklarına ve komşuluk ilişkilerine dayalı olarak kümeleyen bir yaklaşımdır. Algoritma, esnekliği ve aykırı değerleri tanımlama yeteneği sayesinde birçok uygulama alanında etkili bir araçtır. Ancak, epsilon (ε) ve minimum nokta sayısı (MinPts) gibi parametrelerin doğru ayarlanması önemlidir. DBSCAN, coğrafi veri analizi, görüntü işleme ve veri madenciliği gibi birçok alanda başarıyla uygulanmıştır.

KAYNAKÇA:

https://www.researchgate.net/figure/Flowchart-of-MS-DBSCAN-algorithm\_fig1\_262326121

https://dergipark.org.tr/tr/download/article-file/384439

https://doublebyteblog.wordpress.com/2014/05/16/clustering-geospatial-data/

https://forum.knime.com/t/color-image-segmentation-with-dbscan/14313

https://mineracaodedados.wordpress.com/2018/02/09/a-gentle-introduction-to-dbscan/