

Fuzzy Logic – FUZZY CLUSTERING

FUZZY CLUSTERING

Bölütleme (Clustering)

- Clustering-Segmentation-Partitioning-Classification
- Veri kümesini altküme-grup-sınıflara ayırır (bölüt)
- Aynı bölütteki elemanlar benzer olmalı; farklı bölüttekiler benzer olmamalıdır
- Değerlendirme kriteri: “benzerlik” ölçütü
 - Euclidian uzaklık
 - Açık uzaklığı

Bölütleme Çeşitleri

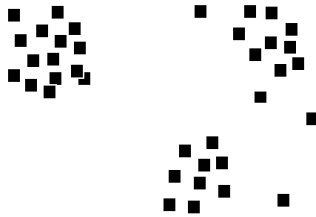
1. Hierarchical clustering
 - Bölütleme, ağaç dalları biçiminde olacak şekilde ve belirli bir hiyerarşiye göre gerçekleştirilir
2. Spectral clustering
 - Veri benzerlik matrisinin spektrumu kullanılarak bölütleme yapılır.
3. Partitional Clustering
 - Veriyi, bir benzerlik ölçütü kullanarak tanımlı olduğu uzayda gruplandırır. Örn: *k*-means (hard *c*-means), fuzzy *c*-means

Bölütleme Basamakları

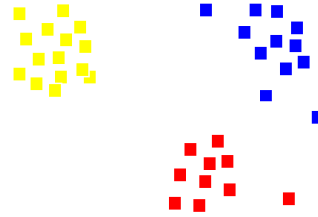
1. “Benzerlik” ölçüsü belirleme
2. Uygun bir algoritma biçimi kullanma
 - Yön göstericili (supervised) algoritmalar
 - Eğitici örnek veri kümesi, test veri kümesi (Classification, intelligent partitioning)
 - Yön göstericisiz (unsupervised) algoritmalar
 - Amaç fonksiyonu tanımlı algoritmalar (Clustering)

Bölütleme Örneği

Veri örnekleri



$c=3$ bölütlenmiş biçimi



Amaç Fonksiyonlu Bölütleme

D : Veri (data) kümesi

R : Sonuç (result) kümesi

$A(D,R)$: analiz uzayı

$J : A(D,R) \rightarrow$ "Reel sayı" haritalaması *amaç fonksiyonudur*

$$J(f) = \sum_{x \in X} \sum_{k \in K} f_{xk}^m d_{xk}^2$$

$f: X \rightarrow F(K)$

Amaç Fonksiyonlu Bölütleme Algoritmaları

- Klasik c-means (Hard c-means, k-means) HCM
- Bulanık c-means (Fuzzy c-means) FCM

HCM (Hard c-means)

Veri Uzayı : $X = \{x_1, x_2, \dots, x_n\}$ (n adet veri)

Herbir eleman : $x_k = [x_{k1} \ x_{k2} \ \dots \ x_{km}]$ (m boyutlu) ($k=1,2,\dots,n$)

Bölüt sayısı : c

Altgruplar A_i : (A_1, A_2, \dots, A_c) c tane altgrup (bölüt)

Koşul

$$(1) \bigcup_{i=1}^c A_i = X$$

$$(2) A_i \cap A_j = \emptyset \quad \forall i \neq j$$

$$(3) \emptyset \subset A_i \subset X \quad \forall i$$

$$2 \leq c < n$$

HCM (Hard c-means)

Karakteristik Fonksiyon: $X_{A_i}(x_k) = \begin{cases} 1, & x_k \in A_i \\ 0, & x_k \notin A_i \end{cases}$

Amaç: “n” tane “m” boyutlu elemanı “c” sınıfa bölmek

Bölüt Matrisi U: $c \times n \times m$ $\mathbf{U} = X_{A_i}(x_k)$

$$\text{Tanım} = \left\{ \mathbf{U} \mid X_{ij} \in \{0,1\}, \sum_{i=1}^c X_{ik} = 1, 0 < \sum_{k=1}^n X_{ik} < n \right\}$$

$$\mathbf{U} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 1 \\ 1 & 0 & \cdots & 0 \end{bmatrix}_{c \times n}$$

HCM Algoritma Basamakları

1. “c”yi sabitle, ilk U matrisini rastgele belirle
2. U matrisinin bölüt merkezi vektörlerini hesapla

$$v_{ij} = \frac{\sum_{k=1}^n X_{ik} \cdot x_{kj}}{\sum_{k=1}^n X_{ik}} = \frac{X_{i1} \cdot x_{1j} + X_{i2} \cdot x_{2j} + \cdots + X_{in} \cdot x_{nj}}{X_{i1} + X_{i2} + \cdots + X_{in}}$$

3. X_{ik} değerlerini belirle, bu değerlerden yeni bölüt matrisi \mathbf{U}^* ’ı oluştur

$$X_{ik} = \begin{cases} 1, & d_{ik} = \min(d_{jk}) \quad \forall j \in c \text{ ise} \\ 0, & \text{aksidurumda} \end{cases}$$

4. $\mathbf{U} = \mathbf{U}^*$ ise DUR, değilse $\mathbf{U} = \mathbf{U}^*$ yap, basamak 2’ye dön

FCM (Fuzzy c-means)

Veri Uzayı : $X = \{x_1, x_2, \dots, x_n\}$ (n adet veri)

Herbir eleman : $x_k = [x_{k1} \ x_{k2} \ \dots \ x_{km}]$ (m boyutlu) ($k=1,2,\dots,n$)

Bölüt sayısı : c

Bölüt merkez vektörleri : $v_i = [v_{i1} \ v_{i2} \ \dots \ v_{im}]$, $i=1,2, \dots, c$

Bölüt matrisi : $U = [u_{ik}]_{c \times n}$ $i=1,2, \dots, c$ $k=1,2, \dots, n$

Elemanlar bulanık üyelik değerleri mertebesinde bölütlere dahil edilir

$$\text{Koşul : } \sum_{i=1}^c u_{ij} = 1$$

$$2 \leq c < n$$

FCM (Fuzzy c-means)

Amaç Fonksiyonu:

$$J(U, v) = \sum_{k=1}^n \sum_{i=1}^c u_{ik}^{m'} d_{ik}^2$$

u_{ik} : k 'nci verinin i 'nci bölüte ait olma üyelik değeri

d_{ik} : k 'nci veri ile i 'nci bölüt merkezi arasındaki uzaklık

m' : Bulanıklaştırıcı sabiti ($m' > 1$)

Algoritma Giriş Parametreleri: c : Bölüt sayısı

m' : Bulanıklaştırıcı sabiti ($m' > 1$)

ϵ : Algoritma durdurma kriteri (hata ölçüsü)

Giriş verisi

Algoritma Çıkış Parametreleri: $center$: Bölüt merkez vektörleri

U : Bölüt matrisi

Obj_fcn : Amaç fonksiyonu

FCM Algoritma Basamakları

1. "c"yi sabitle, ilk U matrisini rastgele belirle
2. U matrisinin bölüt merkezi vektörlerini hesapla $J(U,v)$ amaç fonksiyonu değerini belirle

$$J(U,v) = \sum_{k=1}^n \sum_{i=1}^c u_{ik}^{m'} d_{ik}^2 \quad v_{ij} = \frac{\sum_{k=1}^n u_{ik}^{m'} x_{kj}}{\sum_{k=1}^n u_{ik}^{m'}} \quad u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{\|x_k - v_{jk}\|}{\|x_k - v_{ik}\|} \right)^{\frac{2}{m'-1}}} = \frac{1}{\sum_{j=1}^c \left(\frac{d_{jk}}{d_{ik}} \right)^{\frac{2}{m'-1}}}$$

3. Yeni üyelik değerlerini belirle (u_{ij}^*),
Bu değerlerden yeni bölüt matrisi U^* 'ı oluştur,
Yeni bölüt merkezi v^* belirle,
Yeni amaç fonksiyonu $J^*(U^*,v^*)$ hesapla
4. $|J(U,v) - J^*(U^*,v^*)| < \epsilon$ ise DUR, değilse $U = U^*$, $v = v^*$,
 $J(U,v) = J^*(U^*,v^*)$ olarak güncelle, basamak 3'e dön.

Dağ Bölütleme (Mountain clustering)

1. Veri Uzayını eşit ızgaralara böl
2. Veri yoğunluk ölçüsünü belirleyen dağ fonksiyonunu ($m(v)$) herbir ızgara için hesapla

$$m(v) = \sum_{i=1}^n e^{-\frac{\|v-x_i\|^2}{2\sigma^2}}$$

3. Dağın zirve noktasını belirle. Bu nokta bölüt merkezidir.

DÖNGÜ Bulunan merkezin etkisini aşağıdaki formülle bastır

$$m_{yeni} = m(v) - m(c_1) e^{-\frac{\|v-c_1\|^2}{2\beta^2}}$$

Makul sayıda merkez bulunmuşsa DUR, değilse DÖNGÜ'ye dön.

Dağ Bölütleme

