Rozmycie grupy wyznaczonej algorytmem FCM

1 maja 2025

Przyjmijmy, że mamy zbiór danych $\mathbb{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_X\}$, gdzie X to liczba danych. Każda dana $\mathbf{x}_i$ jest zbudowana z deskryptorów (atrybutów, wymiarów) $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{ia}, \dots, x_{iA}]$, gdzie A to liczba atrybutów. Zapiszmy dane w macierzy $\mathbf{X}$ w ten sposób, że każdy wiersz to jedna dana $\mathbf{x}_i$, a każda kolumna do wartości atrybutu a-tego:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_X \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1A} \\ x_{21} & x_{22} & \dots & x_{2A} \\ \vdots & \vdots & \ddots & \vdots \\ x_{X1} & x_{X2} & \dots & x_{XA} \end{bmatrix}. \tag{1}$$

Algorytm FCM wypracowuje macierz $\mathbf{U}$ przynależności danych do grup (klastrów). Macierz ma tyle wierszy, ile jest grup C, i tyle kolumn, ile jest danych: $\mathbf{U}[C \times X]$:

$$\mathbf{U} = \begin{bmatrix} u_{11} & u_{12} & \dots & u_{1X} \\ u_{21} & u_{22} & \dots & u_{2X} \\ \vdots & \vdots & \ddots & \vdots \\ u_{C1} & u_{C2} & \dots & u_{CX} \end{bmatrix}. \tag{2}$$

Istotne jest, że suma każdej kolumny jest równa 1:

$$\forall_{c \in \mathbb{C}} \quad \sum_{x=1}^{X} u_{cx} = 1. \tag{3}$$

Gdy mamy wyznaczone przynależności każdej danej do grup, możemy wyznaczyć środki grup (klastrów). Znowu zastosujemy zapis wektorowy: środek grupy c-tej zapiszemy jako $\mathbf{v}_c = [v_{c1}, v_{c2}, \ldots, v_{ca}, \ldots, v_{cA}]$. Możemy teraz zapisać środki wszystkich grup jako macierz $\mathbf{V}$ o C wierszach (dla każdej grupy) i A kolumnach (dla każdego atrybutu):

$$\mathbf{V} = \begin{bmatrix} v_{11} & v_{12} & \ldots & v_{1A} \\ v_{21} & v_{22} & \ldots & v_{2A} \\ \vdots & \vdots & \ddots & \vdots \\ v_{C1} & v_{C2} & \ldots & v_{CA} \end{bmatrix}. \tag{4}$$

Jak wyznaczyć jednak środki grup? To jest w sumie proste. Środek $\mathbf{v}_c$ grupy c-tej wyznaczymy jako średnią ważoną wszystkich danych $\mathbf{x}_i$ z wartościami przynależności u_{ci} jako wagami. Zapiszmy to wektorowo:

$$\mathbf{v}_c = \frac{\sum_{i=1}^{X} u_{ci}^m \mathbf{x}_i}{\sum_{i=1}^{X} u_{ci}^m}, \tag{5}$$

gdzie m jest współczynnikiem, zwykle przyjmujemy, że $m = 2$ (to działa zaskakująco dobrze dla bardzo wielu danych, liczb grup itd).

No to zostało nam wyznaczenie rozmycia grup $\mathbf{s}_c$ grupy c-tej ze względu na każdy atrybut $\mathbf{s}_c = [s_{c1}, s_{c2}, \ldots, s_{ca}, \ldots, s_{cA}]$. Rozmycie s_{ca} grupy c-tej dla atrybutu a-tego wyznacza się jako wariancję:

$$s_{ca} = \sqrt{\frac{\sum_{i=1}^{X} u_{ci}^m \left(x_{ia} - v_{ca}\right)^2}{\sum_{i=1}^{X} u_{ci}^m}}. \tag{6}$$