

Phân cụm mờ khả thi

March 17, 2025

1 Giới thiệu về PFCM

- Mô hình **Possibilistic Fuzzy C-Means (PFCM)** là một cải tiến của **Fuzzy C-Means (FCM)** và **Possibilistic C-Means (PCM)**, kết hợp các ưu điểm của cả hai mô hình để cải thiện khả năng phân cụm trong môi trường nhiễu.

2 Các hàm trong PFCM

Hàm mục tiêu

$$\min_{(U,T,V)} \left\{ J_{m,\eta}(U, T, V; X) = \sum_{k=1}^n \sum_{i=1}^c (au_{ik}^m + bt_{ik}^\eta) \|\mathbf{x}_k - \mathbf{v}_i\|_A^2 + \sum_{i=1}^c \gamma_i \sum_{k=1}^n (1 - t_{ik})^\eta \right\} \quad (20)$$

Trong đó:

- N là số điểm dữ liệu.
- C là số cụm.
- u_{ij} là **độ thành viên fuzzy** của điểm dữ liệu x_i vào cụm j .
- t_{ij} là **độ khả năng possibilistic** của x_i vào cụm j .
- $d_{ij} = \|x_i - v_j\|$ là khoảng cách Euclidean từ x_i đến tâm cụm v_j .
- a, b là **hệ số điều chỉnh** cho u_{ij} và t_{ij} .
- m là **bậc fuzziness**, thường $m > 1$.
- η là **bậc possibilistic**, thường $\eta > 1$.
- γ_j (gamma) là **tham số điều chỉnh khả năng** của cụm j .

Hàm chọn tham số gamma

$$\gamma_i = K \frac{\sum_{k=1}^n u_{ik}^m D_{ik}^2 A}{\sum_{k=1}^n u_{ik}^m} \quad \text{thường chọn } K=1 \quad (11)$$

*giống với hàm khởi tạo eta của PCM

Hàm cập nhật điển hình

$$t_{ij} = \frac{1}{1 + (\frac{b}{\gamma_i} d_{ij}^2)^{\frac{1}{\eta-1}}} \quad (8)$$

+ Nếu d_{ij}^2 nhỏ thì $t_{ij} \rightarrow 1$

+ Nếu d_{ij}^2 lớn thì $t_{ij} \rightarrow 0$

*Điển hình trong PFCM giống với điển hình trong PCM

Hàm cập nhật tâm cụm

$$\mathbf{v}_i = \frac{\sum_{k=1}^n (au_{ik}^m + bt_{ik}^\eta) \mathbf{x}_k}{\sum_{k=1}^n (au_{ik}^m + bt_{ik}^\eta)} \quad (23)$$

$$1 \leq i \leq c.$$

Hàm cập nhật ma trận thành viên

$$u_{ik} = \left(\sum_{j=1}^c \left(\frac{D_{ikA}}{D_{jkA}} \right)^{\frac{2}{m-1}} \right)^{-1} \quad (21)$$

$$1 \leq i \leq c; \quad 1 \leq k \leq n$$

3 Đặc điểm

1. Kết hợp cả độ thành viên mờ (fuzzy) và độ khả năng (possibilistic)
- **F**CM chỉ sử dụng độ thành viên, buộc tổng các giá trị u_{ij} của một điểm dữ liệu phải bằng 1. Điều này có thể dẫn đến việc điểm nhiều lần bị ép vào một cụm.

-**PCM** chỉ sử dụng độ khả năng t_{ij} , cho phép một điểm có thể thuộc vào nhiều cụm hoặc không thuộc cụm nào, nhưng dễ bị lỗi "**cụm không có điểm nào**".

PFCM kết hợp cả hai u_{ij} (giúp mô hình duy trì tính chất phân cụm mờ) và t_{ij} giúp mô hình giảm ảnh hưởng của nhiễu và tránh lỗi cụm trùng lặp.

2. Giảm nhạy cảm với nhiễu:

- **FCM có thể bị nhiễu kéo lệch tâm cụm**, do mỗi điểm nhiễu vẫn có một độ thành viên u_{ij} đáng kể.

- **PCM** có thể gán giá trị t_{ij} lớn cho các điểm nhiễu, làm sai lệch kết quả phân cụm.

4 Điều chỉnh tham số a và b

Tham số a :

- Tham số a kiểm soát độ thành viên u_{ij} (**Fuzzy Membership**), giúp mỗi điểm dữ liệu phải có một mức độ thành viên trong tất cả các cụm.

- Nếu a lớn, thuật toán sẽ gần giống **FCM**, tức là **mọi điểm dữ liệu đều bị ràng buộc phải thuộc về một cụm nào đó**.

Tham số b

- b kiểm soát độ khả năng t_{ij} (**Possibility Membership**), giúp mô hình chống nhiễu tốt hơn.

Nếu b lớn, thuật toán sẽ gần giống PCM, tức là **một điểm có thể không thuộc về cụm nào cả (nếu nhiễu)**.

5 Triển khai thuật toán PFCM

Bước 1: Khởi tạo: -Chọn số cụm c , chỉ số mờ $m(m>1)$, n , chọn tham số điều chỉnh a và b , - Khởi tạo ma trận tâm cụm (centroid), ma trận thành viên (membership), ma trận khả năng (typically), tham số gamma

Bước 2: Cập nhật ma trận thành viên u và ma trận khả năng t

Cập nhật u (giống FCM)

$$u_{ik} = \left(\sum_{j=1}^c \left(\frac{D_{ikA}}{D_{jkA}} \right)^{\frac{2}{m-1}} \right)^{-1}$$

Cập nhật t (giống PCM): Trước tiên, tính tham số γ_i (giống công thức số (9) trong bài báo PCM):

$$\gamma_i = K \frac{\sum_{k=1}^n u_{ik}^m D_{ik}^2 A}{\sum_{k=1}^n u_{ik}^m} \quad \text{thường chọn } K=1 \quad (11)$$

sau đó cập nhật ma trận khả năng t:

$$t_{ij} = \frac{1}{1 + (\frac{b}{\gamma_i} d_{ij}^2)^{\frac{1}{\eta-1}}} \quad (8)$$

+ Nếu d_{ij}^2 nhỏ thì $t_{ij} > 1$

+ Nếu d_{ij}^2 lớn thì $t_{ij} > 0$

Bước 3: Cập nhật tâm cụm v theo cả u và t

$$\mathbf{v}_i = \frac{\sum_{k=1}^n (au_{ik}^m + bt_{ik}^\eta) \mathbf{x}_k}{\sum_{k=1}^n (au_{ik}^m + bt_{ik}^\eta)}$$

$$1 \leq i \leq c. \quad (23)$$

* nếu a = 1 , b=0 -> PFCM trở thành FCM

* nếu a =0, b=1 -> PFCM trở thành PCM

Bước 4: kiểm tra điều kiện dừng

lặp lại từ bước 2 đến khi sai số hội tụ