

Phương pháp phân tích chùm mờ (FCM)

Hung Tran-Nam^{1, 2}

¹Laboratory for Applied and Industrial Mathematics, Institute for
Computational Science and Artificial Intelligence, Van Lang University,
Ho Chi Minh City, Vietnam

²Faculty of Fundamental Sciences, Van Lang University, Ho Chi Minh
City, Vietnam

1. Giới thiệu

Thuật toán Phương pháp phân tích chùm mờ (Fuzzy c-Means Clustering) (FCM) được đề xuất bởi Dunn vào năm 1973 và cải tiến sau đó bởi Bezdek. Mục tiêu của bài toán là tối thiểu hóa hàm mục tiêu sau

$$\text{Minimize : } J(\mathbf{U}, \mathbf{V}; \mathcal{X}, c, m) = \sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^m d^2(\mathbf{v}_i, \mathbf{x}_j). \quad (1)$$

1.1 Phương pháp Phân tích chùm mờ

Thuật toán phân tích chùm mờ đã được chứng minh là luôn dừng lại ở hữu hạn vòng lặp khi và chỉ khi trọng tâm chùm $\mathbf{V} = \mathbf{v}_i$ và ma trận phân vùng $\mathbf{U} = [\mu_{ij}]_{cn}$ được tính bởi công thức (U) và (V)

$$\mathbf{v}_i^{(t+1)} = \frac{\sum_{j=1}^n (\mu_{ij} \mathbf{x}_j)}{\sum_{j=1}^n \mu_{ij}} \quad (\text{U})$$

$$\mu_{ij}^{(t+1)} = \frac{1}{\sum_{s=1}^c \left(\frac{d(\mathbf{v}_i, \mathbf{x}_j)}{d(\mathbf{v}_s, \mathbf{x}_j)} \right)^{\frac{2}{m-1}}}, \quad (\text{V})$$

1.2 Các bước thực hiện

Cho bộ số liệu \mathcal{X} gồm các phần tử chưa được gán nhãn $\mathbf{x}_j = (a_j, b_j)^\top$ trong không gian \mathbb{R}^2 ($j = 1, 2, \dots, n$). Cho trước c và các thông số $m = 2$, khoảng cách giữa hai phần tử là khoảng cách Euclidean. Các bước thực hiện của thuật toán phân tích chùm mờ được thực hiện như sau

1. Khởi tạo c -trọng tâm chùm một cách ngẫu nhiên $\mathbf{v}_i, i = 1, 2, \dots, c$,
2. Tính ma trận khoảng cách \mathbf{D}_{cn} giữa c -trọng tâm với tất cả các điểm trong bộ số liệu, tức là $d(\mathbf{v}_i, \mathbf{x}_j)$,
3. Tính ma trận phân vùng $\mathbf{U}^{(t+1)}$ có các phần tử tương ứng là μ_{ij} theo công thức

$$\mu_{ij} = \frac{1}{\sum_{s=1}^c \left(\frac{d(\mathbf{v}_i, \mathbf{x}_j)}{d(\mathbf{v}_s, \mathbf{x}_j)} \right)^{\frac{2}{m-1}}},$$

4. Cập nhật lại ma trận trọng tâm chùm $\mathbf{V}^{(t+1)}$ từ ma trận phân vùng đã tính theo công thức

$$\mathbf{v}_i^{(t+1)} = \frac{\sum_{j=1}^n (\mu_{ij} \mathbf{x}_j)}{\sum_{j=1}^n \mu_{ij}},$$

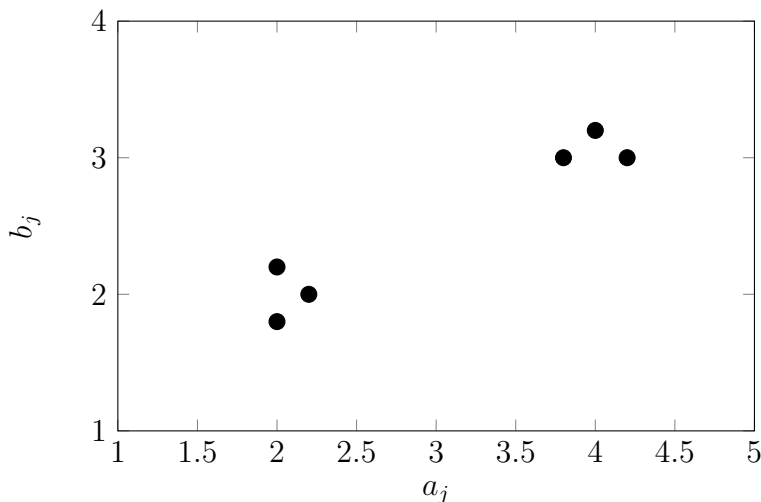
5. Tính điều kiện dừng và so sánh với ϵ . Nếu

$$S = \|\mathbf{V}^{(t+1)} - \mathbf{V}^{(t)}\| < \epsilon$$

thì tiếp tục Bước 2 trở đi với $\mathbf{v}_i^{(t+1)}$ và $\mathbf{U}^{(t+1)}$ mới

2. Ví dụ

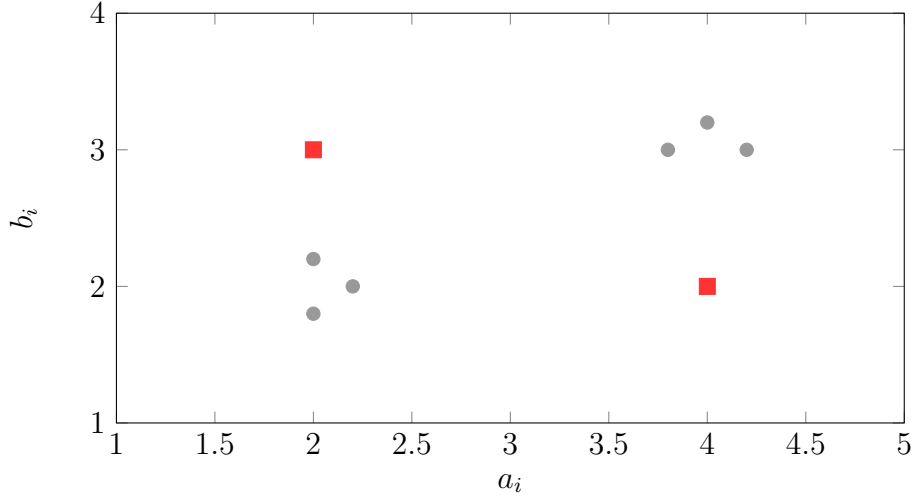
Cho dữ liệu $\mathbf{x}_j = (a_j, b_j) \in \mathbb{R}^2$ có tọa độ được cho ở Bảng dưới đây. Biết $m = 2$, khoảng cách giữa hai phần tử là khoảng cách Euclidean, và $\epsilon = 0.01$ hãy chia dữ liệu trên thành hai chùm bằng phương pháp phân tích chùm mờ.



Dữ liệu ví dụ		
ID	a_j	b_j
1	2.0	2.2
2	2.0	1.8
3	2.2	2.0
4	3.8	3.0
5	4.0	3.2
6	4.2	3.0

Bước 1. Vì bài toán cần phân tích dữ liệu thành 2 chùm nên ta khởi tạo ngẫu nhiên hai trọng tâm \mathbf{v}_i trong không gian \mathbb{R}^2 . Giả sử ta có

$$v_1 = \begin{pmatrix} 2.0 \\ 3.0 \end{pmatrix} \text{ và } v_2 = \begin{pmatrix} 4.0 \\ 2.0 \end{pmatrix}$$



Hình 1: Khởi tạo trọng tâm chùm. Ở đây, các điểm trong bộ số liệu được ký hiệu màu đen; các trọng tâm được ký hiệu màu đỏ.

Bước 2. Ta lần lượt tính khoảng cách từ trọng tâm đã khởi tạo đến mỗi điểm còn lại trong dữ liệu. Giả sử ma trận khoảng cách Euclidean là $\mathbf{D}_{cn} = d(\mathbf{v}_i, \mathbf{x}_j)$, ta có

$$\begin{aligned} d(\mathbf{v}_1, \mathbf{x}_1) &= \left\| \begin{pmatrix} 2.0 \\ 3.0 \end{pmatrix}, \begin{pmatrix} 2.0 \\ 2.2 \end{pmatrix} \right\| = 0.80 & d(\mathbf{v}_2, \mathbf{x}_1) &= \left\| \begin{pmatrix} 4.0 \\ 2.0 \end{pmatrix}, \begin{pmatrix} 2.0 \\ 2.2 \end{pmatrix} \right\| = 2.01 \\ d(\mathbf{v}_1, \mathbf{x}_2) &= \left\| \begin{pmatrix} 2.0 \\ 3.0 \end{pmatrix}, \begin{pmatrix} 2.0 \\ 1.8 \end{pmatrix} \right\| = 1.20 & d(\mathbf{v}_2, \mathbf{x}_2) &= \left\| \begin{pmatrix} 4.0 \\ 2.0 \end{pmatrix}, \begin{pmatrix} 2.0 \\ 1.8 \end{pmatrix} \right\| = 2.01 \\ d(\mathbf{v}_1, \mathbf{x}_3) &= \left\| \begin{pmatrix} 2.0 \\ 3.0 \end{pmatrix}, \begin{pmatrix} 2.2 \\ 2.0 \end{pmatrix} \right\| = 1.02 & d(\mathbf{v}_2, \mathbf{x}_3) &= \left\| \begin{pmatrix} 4.0 \\ 2.0 \end{pmatrix}, \begin{pmatrix} 2.2 \\ 2.0 \end{pmatrix} \right\| = 1.80 \\ d(\mathbf{v}_1, \mathbf{x}_4) &= \left\| \begin{pmatrix} 2.0 \\ 3.0 \end{pmatrix}, \begin{pmatrix} 3.8 \\ 3.0 \end{pmatrix} \right\| = 1.80 & d(\mathbf{v}_2, \mathbf{x}_4) &= \left\| \begin{pmatrix} 4.0 \\ 2.0 \end{pmatrix}, \begin{pmatrix} 3.8 \\ 3.0 \end{pmatrix} \right\| = 1.02 \\ d(\mathbf{v}_1, \mathbf{x}_5) &= \left\| \begin{pmatrix} 2.0 \\ 3.0 \end{pmatrix}, \begin{pmatrix} 4.8 \\ 3.0 \end{pmatrix} \right\| = 2.01 & d(\mathbf{v}_2, \mathbf{x}_5) &= \left\| \begin{pmatrix} 4.0 \\ 2.0 \end{pmatrix}, \begin{pmatrix} 4.8 \\ 3.0 \end{pmatrix} \right\| = 1.20 \\ d(\mathbf{v}_1, \mathbf{x}_6) &= \left\| \begin{pmatrix} 2.0 \\ 3.0 \end{pmatrix}, \begin{pmatrix} 4.0 \\ 3.2 \end{pmatrix} \right\| = 2.20 & d(\mathbf{v}_2, \mathbf{x}_6) &= \left\| \begin{pmatrix} 4.0 \\ 2.0 \end{pmatrix}, \begin{pmatrix} 4.0 \\ 3.2 \end{pmatrix} \right\| = 1.02 \end{aligned}$$

Từ đây, ta có ma trận khoảng cách \mathbf{D} là

$$\mathbf{D} = \begin{pmatrix} 0.80 & 1.20 & 1.02 & 1.80 & 2.01 & 2.20 \\ 2.01 & 2.01 & 1.80 & 1.02 & 1.20 & 1.02 \end{pmatrix}$$

Bước 3. Ta tính toán các phần tử μ_{ij} trong ma trận phân vùng $\mathbf{U}^{(0)}$

36

$$\begin{aligned}\mu_{11} &= \left(\left(\frac{d_{11}}{d_{11}} \right)^2 + \left(\frac{d_{11}}{d_{21}} \right)^2 \right)^{-1} = 0.86 & \mu_{21} &= \left(\left(\frac{d_{21}}{d_{11}} \right)^2 + \left(\frac{d_{21}}{d_{21}} \right)^2 \right)^{-1} = 0.14 \\ \mu_{12} &= \left(\left(\frac{d_{12}}{d_{12}} \right)^2 + \left(\frac{d_{12}}{d_{22}} \right)^2 \right)^{-1} = 0.74 & \mu_{22} &= \left(\left(\frac{d_{22}}{d_{12}} \right)^2 + \left(\frac{d_{22}}{d_{22}} \right)^2 \right)^{-1} = 0.26 \\ \mu_{13} &= \left(\left(\frac{d_{13}}{d_{13}} \right)^2 + \left(\frac{d_{13}}{d_{23}} \right)^2 \right)^{-1} = 0.76 & \mu_{23} &= \left(\left(\frac{d_{23}}{d_{13}} \right)^2 + \left(\frac{d_{23}}{d_{23}} \right)^2 \right)^{-1} = 0.24 \\ \mu_{14} &= \left(\left(\frac{d_{14}}{d_{14}} \right)^2 + \left(\frac{d_{14}}{d_{24}} \right)^2 \right)^{-1} = 0.24 & \mu_{24} &= \left(\left(\frac{d_{24}}{d_{14}} \right)^2 + \left(\frac{d_{24}}{d_{24}} \right)^2 \right)^{-1} = 0.76 \\ \mu_{15} &= \left(\left(\frac{d_{15}}{d_{15}} \right)^2 + \left(\frac{d_{14}}{d_{24}} \right)^2 \right)^{-1} = 0.26 & \mu_{25} &= \left(\left(\frac{d_{25}}{d_{15}} \right)^2 + \left(\frac{d_{25}}{d_{25}} \right)^2 \right)^{-1} = 0.74 \\ \mu_{16} &= \left(\left(\frac{d_{15}}{d_{15}} \right)^2 + \left(\frac{d_{16}}{d_{26}} \right)^2 \right)^{-1} = 0.18 & \mu_{26} &= \left(\left(\frac{d_{26}}{d_{16}} \right)^2 + \left(\frac{d_{26}}{d_{26}} \right)^2 \right)^{-1} = 0.82\end{aligned}$$

Từ đây, ta có ma trận phân vùng $\mathbf{U}^{(1)}$

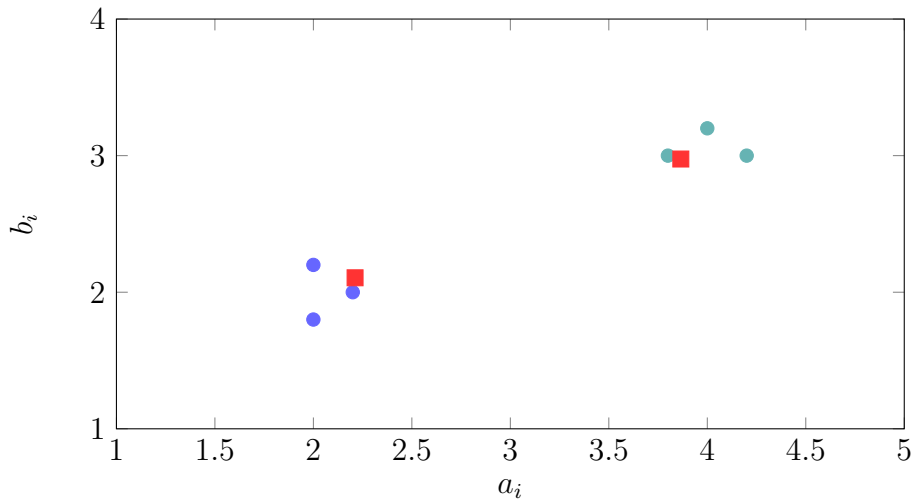
37

$$\mathbf{U}^{(1)} = \begin{pmatrix} 0.86 & 0.74 & 0.76 & 0.24 & 0.26 & 0.18 \\ 0.14 & 0.26 & 0.24 & 0.76 & 0.74 & 0.82 \end{pmatrix}$$

Bước 4. Ta cập nhật lại ma trận trọng tâm $\mathbf{V}^{(1)}$ với các thành phần \mathbf{v}_i

38

$$\mathbf{v}_1 = \frac{\sum_{i=1}^c \mu_{ij} \mathbf{x}_j}{\sum_{i=1}^c \mu_{ij}} = \begin{pmatrix} 2.21 \\ 2.11 \end{pmatrix} \quad \mathbf{v}_2 = \frac{\sum_{i=1}^c \mu_{ij} \mathbf{x}_j}{\sum_{i=1}^c \mu_{ij}} = \begin{pmatrix} 3.87 \\ 2.98 \end{pmatrix}$$

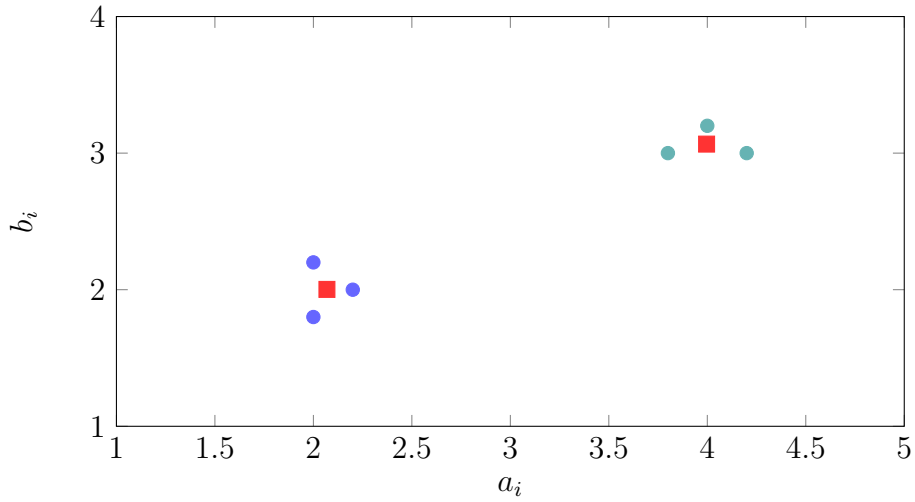


Hình 2: Cập nhật trọng tâm ở vòng lặp số 1

Bước 5. Ta tính điều kiện dừng S như sau

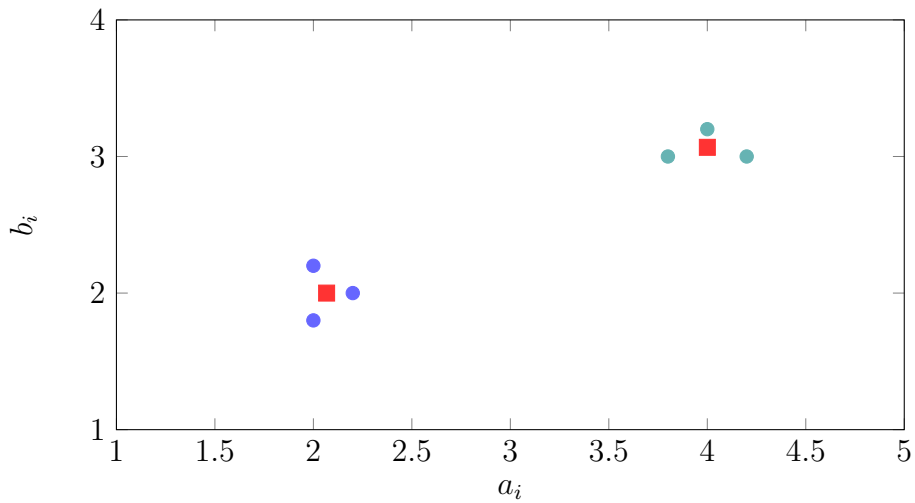
$$S = \|\mathbf{V}^{(0)} - \mathbf{V}^{(1)}\| = \left\| \begin{pmatrix} 2.21 & 3.87 \\ 2.11 & 2.98 \end{pmatrix} - \begin{pmatrix} 2 & 4 \\ 3 & 2 \end{pmatrix} \right\| = 1.11 > \epsilon$$

Ta tiếp tục vòng lặp thứ hai từ Bước 2 với ma trận trọng tâm chùm $\mathbf{V}^{(1)}$ và $\mathbf{U}^{(1)}$ đã biết. Hình 3 cho kết quả phân tích chùm ở vòng lặp thứ hai. Vì chưa thỏa mãn điều kiện dừng nên ta tiếp tục vòng lặp thứ ba.



Hình 3: Cập nhật trọng tâm ở vòng lặp số 2 (chưa dừng vì $S = 0.25$)

Vòng lặp thứ ba là vòng lặp cuối cùng với $S = 0.004 < \epsilon$. Thuật toán phân tích chùm mờ kết thúc, ta nhận được trọng tâm chùm và ma trận phân vùng tương ứng.

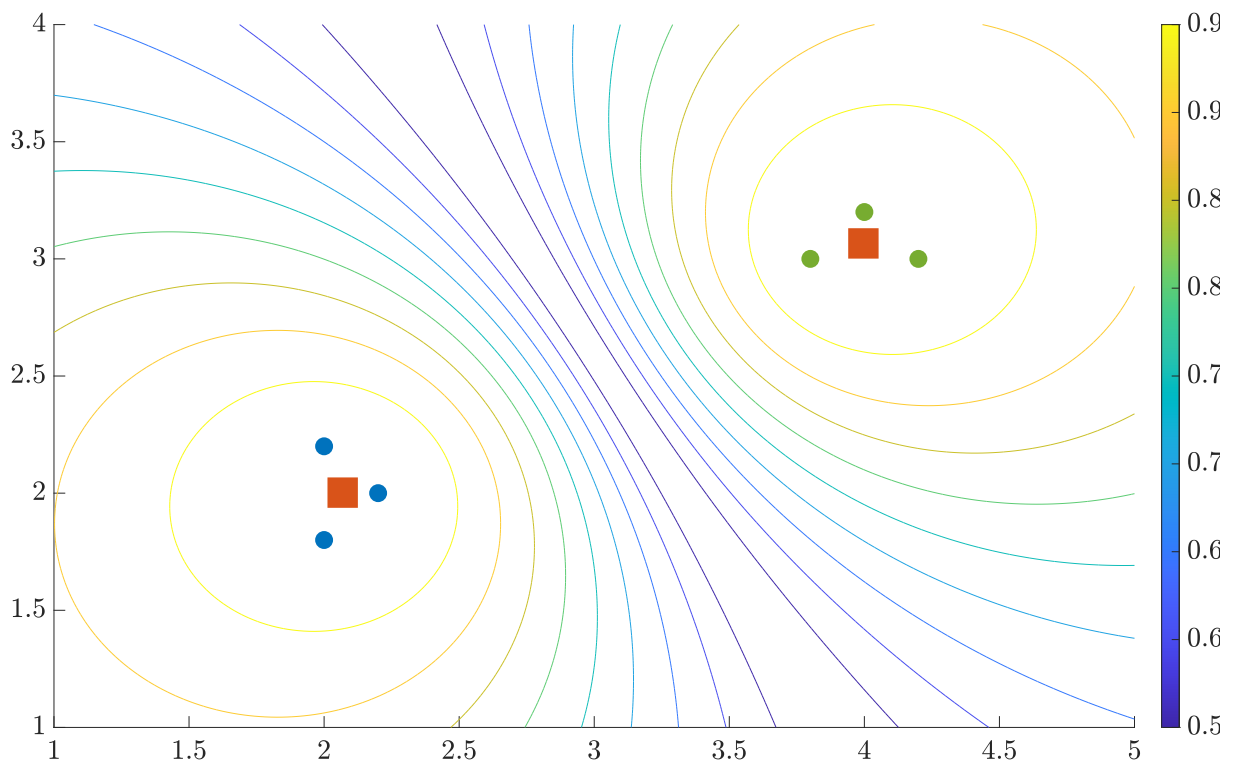


Hình 4: Cập nhật trọng tâm ở vòng lặp số 3 (vòng lặp cuối cùng với $S = 0.004$)

Thuật toán phân tích chùm mờ cho kết quả như sau

$$\mathbf{V} = \begin{pmatrix} 2.06 & 4.00 \\ 2.00 & 3.07 \end{pmatrix}$$

$$\mathbf{U} = \begin{pmatrix} 0.99 & 0.97 & 1.00 & 0.00 & 0.02 & 0.02 \\ 0.01 & 0.03 & 0.00 & 1.00 & 0.98 & 0.98 \end{pmatrix}$$



Hình 5: Kết quả phân tích chùm mờ

1	<i>%% MATLAB code for FCM</i>	46
2	a = [2.0 2.0 2.2 3.8 4.0 4.2];	47
3	b = [2.2 1.8 2.0 3.0 3.2 3.0];	48
4	X = [a; b];	49
5		50
6	V = [2 3; 4 2]';	51
7		52
8	m = 2;	53
9	c = 2;	54
10	n = size(X,2);	55
11	max_iter = 100;	56
12	epsilon = 0.01;	57
13	colors = parula(2)';	58
14		59
15	for iter = 1:max_iter	60
16	for i = 1:c	61
17	for j = 1:n	62
18	D(i,j) = sum((V(:,i) - X(:,j)).^2).^0.5;	63
19	end	64
20	end	65
21		66
22	for i = 1:c	67
23	for j = 1:n	68
24	S = 0;	69
25	for k = 1:c	70
26	S = S + (D(i,j)^(2/(m-1))) / (D(k,j)^(2/(m-1)));	71
27	end	72
28	U(i, j) = 1 / S;	73
29	end	74
30	end	75
31		76
32	V_new = (X * (U.^m)') ./ sum(U.^m,2)';	77
33		78
34	Stop = norm(V_new - V, 1);	79
35	if Stop < epsilon	80
36	break;	81
37	else	82
38	V = V_new;	83
39	end	84
40	end	85