FCE 클러스터링 알고리듬을 이용한 3차원 데이터의 정점 검출

Vertex Detection of 3-D Data Using FCV Clustering Algorithm

최병걸, 이원희, 강훈
중앙대학교 제어계측공학과
서울시 동작구 흑석동 221번지
bkchoi@sirius.cie.cau.ac.kr, mystery@sirius.cie.cau.ac.kr, hkang@cau.ac.kr

ABSTRACT

최근 컴퓨터의 속도 및 용량의 확장과 더불어 3차원 정보에 대한 연구의 필요성이 요구되고 있다. 본 논문에서는이 여기에 관한 연구의 하나로 FCV(Fuzzy c-Varieties) 클러스터링의 방법을 써서 3차원 데이터의 변과 정점을 찾아 3차원 물체를 구성하여 중복된 자료의 크기를 압축하는 방법을 제시한다. 여기에 따른 문제점으로 클러스터의 개수를 결정하는 문제가 있는데 이는 fuzzy classification entropy로 해결하였다.

I. 서론

최근 컴퓨터의 속도 및 용량의 확장과 더불어 3차원 정보에 대한 연구의 필요성이 요구되고 있다. 디지털 영상과 같은 2차원 정보는 이미 디지털 카메라나 스캐너와 같은 도구를 사용하여 여러 가지 이미지형식으로 저장 및 교환이 가능하나, 3차원 정보는 자료를 얻는 방법과 도구가 그리 많지 않은 실정이다.

여기에서는 3차원 스캐너 연구 과정의 하나로 클러스터링 방법을 이용하여 물체의 정점을 찾아냄으로써 데이터의 압축효과와 처리속도를 향상시킬 수 있도록 하였다.

여기서 사용하게될 FCV(Fuzzy c-Varieties) 알고리듬에 대해 간단히 살펴보면, 많이 알려져 있는 FCM(Fuzzy c-means) 알고리듬[1]을 확장한 FCL(Fuzzy c-lines) 알고리듬[2], 즉 클러스터의 프로토타입(prototype)을 점에서 선으로 확장시켰고, 또 이 FCL 알고리듬을 n 차원으로 확장시킨 FCV 알고리듬[3]이 있다.

이 알고리듬의 결과로 나온 membership 값

※ 본 연구는 1996년도 한국학술진흥재단 대학부설연구소과제 연구비에 의하여 연구되었음.

을 분석하여 정점 및 변을 찾을 수 있다.

클러스터링의 일반적인 문제점으로 초기값 설정과 클러스터의 개수 문제가 있는데, 클러스터 개수에 관한 문제는 classification entropy[4]로 해결하고자 하였고, 초기값 문제는 여기에서도 여전히 문제점으로 남아 있는데, weighting exponent를 증가함으로써 어느 정도 local minimum에 빠지는 것을 막을 수 있다.

하지만 보다 근본적인 해결은 데이터에 대해 초기값군들에 대한 수렴 영역을 찾아내던가, 대략의 특징 추출을 해서 여기에 따른 초기값을 찾아내야 할 것으로 생각된다. 현재 이 부분에 대해 많은 연구중이다.

II. 본론

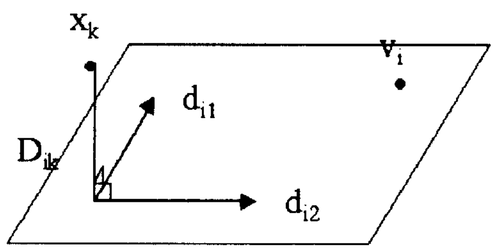
II-1 FCV(Fuzzy c-Varieties) 알고리듬

$$X=\{ x_1, x_2, \cdots, x_N \}$$

소속값 u_{ij} 은 다음 조건을 만족한다.

$$0 \le u_{ij} \le 1$$

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



$P_i(v_i, d_{i1}, d_{i2})$

그림 1 데이터와 클러스터간의 거리

▷ 클러스터의 프로토타입

$$V_2(\mathbf{v};\mathbf{d}_1,\mathbf{d}_2) = \left\{ \mathbf{y}\in R^s \mid \mathbf{y} = \mathbf{v} + \sum_{j=1}^{2} t_j \mathbf{d}_j;\ t_j \in R \right\}$$

▷ 거리 함수

$$D_A(\mathbf{x}_k, V_{2i}) = \left(\| \mathbf{x}_k - \mathbf{v}_i \|_A^2 - \sum_{j=1}^{2} (\langle \mathbf{x}_k - \mathbf{v}_i, \mathbf{d}_{ij} \rangle_A)^2 \right)^{1/2} \triangleq D_{ik}.$$

▷ cost function

$$J_{V_2m}(U, V) = \sum_{k=1}^{N} \sum_{i=1}^{c} (u_{ik})^m (D_{ik})^2, \quad 1 \le m < \infty,$$

A : positive definite matrix

c : 클러스터의 개수 ($2 \le c < N$)

N : 데이터 개수

m : weight exponent ($1 < m < \infty$)

▷ 새로운 소속값

$$\hat{u}_{ik} = \frac{1}{\sum_{j=1}^{c} (\hat{D}_{ik} / \hat{D}_{jk})^{2/(m-1)}} \quad \forall i, k.$$

▷ 새로운 클러스터 중심

$$\hat{\mathbf{v}}_i = \frac{\sum_{k=1}^{n} (\hat{u}_{ik})^m \mathbf{x}_k}{\sum_{k=1}^{n} (\hat{u}_{ik})^m} \quad \forall i,$$

$$\hat{\mathbf{d}}_{ij} = A^{-1/2} \hat{\mathbf{y}}_{ij},$$

$\hat{\mathbf{y}}_{ij}$는 scatter matrix S_i의 j번째 unit eigenvector

$$\hat{S}_i = A^{1/2} \left[\sum_{k=1}^{n} (\hat{u}_{ik})^m (\mathbf{x}_k - \hat{\mathbf{v}}_i) \times (\mathbf{x}_k - \hat{\mathbf{v}}_i)^T \right] A^{1/2}.$$

II-2 Classification Entropy

$$E=-\frac{1}{N}\sum_{i}u_{ik}\log_{2}u_{ik}$$

위 식은 클러스터의 평가함수로 개수를 추정하는 척도로 쓰인다. 계산된 값이 작을수록 분류가 잘되었다고 판단할 수 있다.

II-3 Vertex detection

소속값 u_{ik} 를 분석함으로써 정점을 찾을 수 있다. 예를 들어 두 면과 만나는 점들은 변이라고 할 수 있는데 한 점에서 두 면과의 각각의 거리가 조건이하가 되면 소속값을 0.5 로 나누어 갖게 된다. 마찬가지로 3면이 만나는 점은 소속값을 0.33 으로 나누어 갖게 되는데 이 점들을 정점이라 할 수 있다.[5]

그리고 각 정점이 어떤 면들과 이루어진 알 수 있으므로 최종적으로 3차원 물체를 복원할 수 있다.

II-4 예제 및 실험결과

예제-1. 정육면체

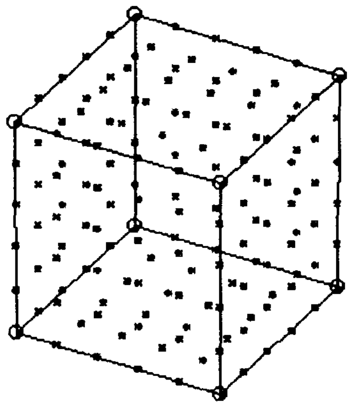


그림 2 정육면체 데이터 및 수행결과

0 :	0.333	0.000	0.333	0.000	0.333	0.000
25 :	0.333	0.333	0.333	0.000	0.000	0.000
70 :	0.000	0.000	0.333	0.333	0.333	0.000
75 :	0.000	0.333	0.333	0.333	0.000	0.000
100 :	0.000	0.333	0.000	0.333	0.000	0.333
125 :	0.000	0.000	0.000	0.333	0.333	0.333
126 :	0.333	0.000	0.000	0.000	0.333	0.333
147 :	0.333	0.333	0.000	0.000	0.000	0.333
148 :	0.000	0.500	0.000	0.000	0.000	0.500
149 :	0.000	0.500	0.000	0.000	0.000	0.500
150 :	0.000	0.500	0.000	0.000	0.000	0.500
151 :	0.000	0.500	0.000	0.000	0.000	0.500

< 표 1 > 소속값 (u_{ik}) 결과

클러스터 개수	entropy
3	0.4384
4	0.4496
5	0.4904
→ 6	0.3992
7	0.4221
8	0.5078

< 표 2 > 클러스터 개수에 따른 entropy

예제-2. 오면체(피라미드)

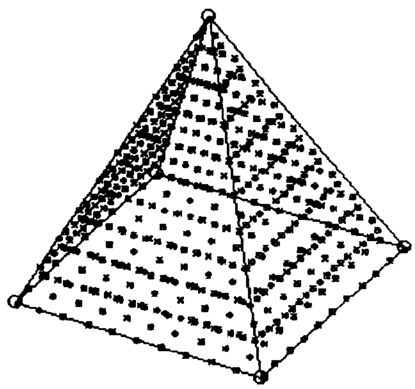


그림 3 오면체 데이터 및 수행결과

0	:	0.3333	0.0000	0.3333	0.0000	0.3333
1	:	0.5000	0.0000	0.0000	0.0000	0.5000
2	:	0.5000	0.0000	0.0000	0.0000	0.5000
9	:	0.5000	0.0000	0.0000	0.0000	0.5000
10	:	0.3333	0.3333	0.0000	0.0000	0.3333
11	:	0.5000	0.0000	0.5000	0.0000	0.0000
109	:	0.5000	0.5000	0.0000	0.0000	0.0000
110	:	0.3333	0.0000	0.3333	0.3333	0.0000
111	:	0.5000	0.0000	0.0000	0.5000	0.0000
119	:	0.5000	0.0000	0.0000	0.5000	0.0000
120	:	0.3333	0.3333	0.0000	0.3333	0.0000
121	:	0.0000	0.0000	0.5000	0.0000	0.5000
479	:	0.0000	0.5000	0.0000	0.5000	0.0000
480	:	0.0000	0.0000	0.0000	1.0000	0.0000
481	:	0.0000	0.2500	0.2500	0.2500	0.2500

< 표 3 > 소속값 (u_{ik}) 결과

클러스터 개수	entropy
3	0.3653
4	0.3647
→ 5	0.1688
6	0.2477
7	0.2926

< 표 4 > 클러스터 개수에 따른 entropy

III. 결론

FCV 알고리즘을 이용하여 3차원 데이터에 대해서 정점을 찾아 물체를 복원을 해보았다. 문제점으로는 초기값과 클러스터 개수를 들 수 있다. 여기서는 평가함수 entropy를 써서 개수 문제를 해결하였는데 클러스터의 개수를 증가시키면서 계속 계산하는데 개수가 많을 경우 시간이 많이 걸릴뿐 아니라 잘못된 개수를 추정할 가능성이 있다. 이 문제점은 대략의 특징 추출을 해서 개수와 초기값을 줌으로써 해결할 수 있을 것으로 생각된다. 현재 이 부분에 대하여 연구중이다.

IV. 참고문헌

[1] J. C. Bezdek, R. J. Hathaway, M. J. Sabin, W. Tucker, "Convergence theory for fuzzy c-means: counterexamples and repairs", IEEE Trans. SMC, vol. SMC-17, no. 5, pp.873-877, 1987

[2] J. C. Bezdek, C. Coray, R. Gunderson, and J. Watson, "Detection and characterization of cluster substructure Part I. Linear structure:Fuzzy c-lines," SIAM J. Appl. Math, vol.40, No.2, pp.339-357, 1981

[3] J. C Bezdek, C. Coray, R. Gunderson, and J. Watson, "Detection and characterization of cluster substructure Part II. Fuzzy c-Varieties and convex combinations thereof," SIAM J. Appl. Math, vol.40, No.2, pp.358-372, 1981

[4] M. P. Windham, "Cluster validity for the fuzzy c-means clustering algorithm," IEEE Trans. PAMI, vol. PAMI-4, no. 4, pp. 357-363, 1982

[5] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," J. Cybernetics, vol. 3, no. 3, pp. 32-57, 1973