

# 동적 GBFCM(Gradient Based FCM) 알고리즘

김명호\*<sup>o</sup> 박동철\*\*

\* 명지대학교 전기공학과 \*\* 명지대학교 제어계측공학과

## Dynamic GBFCM(Gradient Based FCM) Algorithm

Myoung-Ho Kim\*<sup>o</sup> Dong C. Park\*\*

\* Myongji Univ. Electrical Eng. \*\* Myongji Univ. Control & Instru. Eng.

### Abstract

A clustering algorithms with dynamic adjustment of learning rate for GBFCM(Gradient Based FCM) is proposed in this paper. This algorithm combines two idea of dynamic K-means algorithms and GBFCM : learning rate variation with entropy concept and continuous membership grade. To evaluate dynamic GBFCM, we made comparisons with Kohonen's Self-Organizing Map over several tutorial examples and image compression. The results show that DGBFCM(Dynamic GBFCM) gives superior performance over Kohonen's algorithm in terms of signal-to-noise.

### 1. 서론

클러스터링(Clustering) 알고리즘의 목적은 유사한 것들과 그렇지 않은 것들을 서로 분리하여 그룹(group)들을 만드는 것이다. 이러한 것들중에 잘 알려진 것으로는 Kohonen 네트워크(network)이 있다. Kohonen은 학습(learning)이 Winner - Take - All 규칙(rule)에 기반을 둔 네트워크를 소개했다. Winner-Take-All 규칙은 오직 하나의 뉴런(neuron)만이 입력 데이터에 대해서 반응한다는 것이다. 즉 워너(winner)가 선택되면 나머지들은 모두 루저(looser)가 되는 것이다[1].

Kohonen 네트워크[7]은 몇가지 단점이 있는데 우선 가장 심각한 문제가 기준 벡터(reference vecot)들의 초기치 문제인데, Kohonen 네트워크는 초기치에 많이 의존한다. 어떤 초기 기준 벡터들은 적은 입력 패턴(pattern)을 갖는 입력 영역(domain)에서 곤경에 빠질지도 모른다. 그리고 또 다른 문제점은 시변(time varying)문제에 대해서 대처하지 못한다는 것이다. 왜냐하면 Kohonen 네트워크는 학습률(learning rate)이 일정하게 감소하므로 증가해야하는 상황에 대해서는 대처할 수 없기 때문이다. 그래서 학습률이 동적으로 움직이는 것이 요구되는데 이러한 것을 동적 K-means[5]에서 구현하였다. 그러나 이 알고리즘은 variation의 초기치가 매우 작다는 조건에 많이 좌우되므로 보다 덜 민감한 variation 초기치를 갖도록 하기위해 동적 GBFCM 알고리즘[2-4]을 개발했다. 이 알고리즘은 variation의 초기치에 덜 민감하고 멤버십 등급이 연속적인 값을 갖도록 구현했다.

본 논문의 알고리즘인 동적 GBFCM 알고리즘은 GBFCM 알고리즘의 원리와 동적 K-means의 개념을 조합하였다.

### 2. 최소화 과정

Kohonen 네트워크의 목적은 입력 패턴 벡터들을 K개의 영역으로 나누고 그 K개의 영역을 대표하는 최상의 기준벡터들을 구하고자 하는 것이다.

최소화 과정(minimization procedure)은 기준벡터  $\vec{c}_k$ 에 관 계되는 입력 패턴 벡터  $\vec{x}$ 의 변형(variation)을 최소화 하기 위한 것이다. 가장 최적의 클러스터링은 각각의 그룹들이 모두 같은 변형을 갖으면 된다. 목적함수를 아래와 같이 정의하면:

$$MSE = \sum_{k=1}^K v_k = \sum_{k=1}^K \sum_{i=1}^N \cdot \|\vec{x}_i - \vec{c}_k\|^2 \quad (1)$$

$$NMSE = \frac{MSE}{\sum_{i=1}^N \|\vec{x}_i - \vec{m}\|^2} \quad (2)$$

$\vec{x}$  : 입력 패턴 벡터  
 $n$  : 영역내의 데이터 개수  
 $N$  : 전체 데이터 개수

NMSE(Normalization Mean Square Error)은 MSE의 표준화(normalization)로서 분모에서  $\vec{m}$ 은 패턴 벡터 집합의 평균을 의미하고  $v_k$ 는 k영역내의 변형을 의미한다. 그리고  $v_k$ 의 값들은  $\vec{c}$ (기준 벡터)의 위치에 의존한다.

### 3. 동적 K-means 알고리즘

#### 3.1 최적 분류를 위한 편향(Biasing)

편향은 입력 패턴 벡터  $\vec{x}$ 와 기준 벡터  $\vec{c}_k$ 사이의 거리에 자신의 그룹 변형까지 곱해준다.

$$d(\vec{x} - \vec{c}_k) = v_k \cdot \|\vec{x} - \vec{c}_k\|^2 \quad (3)$$

그리고 편향 멤버십 척도(membership indicator)는

$$\mu_{k, bias}(\vec{x}) = \begin{cases} 1 & \text{if } (v_k \parallel \vec{x} - \vec{c}_k \parallel^2) \leq (v_k \parallel \vec{x} - \vec{c}_k \parallel^2) \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

각각의 영역에 대해서  $\vec{v}_k$ 를 갱신하기 위한 식은:

$$v_k(new) = a \cdot v_k(old) + (1-a) \cdot \mu_{k, bias}(\vec{x}(old)) \cdot \parallel \vec{x}(old) - \vec{c}_k(old) \parallel^2 \quad (5)$$

식 (5)에서  $a$ 는 1보다 작은 값으로 보통 0.99에서 0.9999사이의 값들을 사용한다[5].

### 3.2 학습률의 동적 조절

영역 분류가 원하는 것과 크게 다르면 학습률은 커져야 하며, 반대로 클러스터링이 잘 되었으면 학습률은 작아야 한다. 그리고 아래의 수식들은 엔트로피(entropy)[9]수식으로 부터 나온 것이다.

$$v_{k, norm} = \frac{v_k}{\sum_{i=1}^K v_i}, \quad \sum_{k=1}^K v_{k, norm} = 1 \quad (6)$$

$$H(v_1, v_2, \dots, v_K) = - \sum_{k=1}^K v_{k, norm} \ln(v_{k, norm}) \quad (7)$$

$$\eta = \frac{\ln(K) - H(v_1, v_2, \dots, v_K)}{\ln(K)} \quad (8)$$

식 (8)에서 학습률  $\eta$ 는 각 영역들의 변형들에 의존한다. 그리고 최적 분류에서 멀어질수록 학습률은 1에 가까워지고 최적 분류에 가까워 질수록 0에 가까운 값을 갖는다.

## 4. Gradient-Based FCM(Fuzzy C\_means)

GBFCM이 FCM과 크게 다른 점은 FCM에서는 모든 데이터를 사용하여 목적 함수를 최소화 하려고 하나 GBFCM은 한개의 데이터에 대해서 목적 함수를 최소화 하려고 한다[2].

### 4.1 알고리즘

한 개의 데이터  $x_i$ 와 임의의 센터  $c_j (j=1,2,\dots,K)$ 가 주어졌을 때 목적 함수는 다음과 같이 정의된다[10]:

$$J_i = \mu_{1i}^2(c_1 - x_i) + \mu_{2i}^2(c_2 - x_i) + \dots + \mu_{Ki}^2(c_K - x_i) \quad (9)$$

$$\mu_{1i} + \mu_{2i} + \dots + \mu_{Ki} = 1 \quad (10)$$

기준 벡터 갱신 식은 :

$$c_{k+1} = c_k - \eta \cdot \mu_{ki}^2 \cdot (c_k - x_i) \quad (11)$$

$\eta$  : 학습률, 한 주기당 1/16배로 감소된다.

멤버십 등급에 대해서 :

$$\mu_{ik} = \frac{1}{\sum_{j=1}^K \left( \frac{d_j(x_k)}{d_j(x_k)} \right)} \quad (12)$$

식 (12)에서  $\sum_{j=1}^K d_j(x_k)$ 은 하나의 데이터와 모든 센터들 사이의 거리의 합이고,  $d_j(x_k)$ 은 하나의 데이터와  $j$ 번째 센터와의

거리이다. 식(12)에서 입력 데이터와 센터사이의 거리가 가까우면  $\mu$ 값이 커지고 멀면 작아진다. 즉, 모든 센터들이 자신의 가중치를 갖고 움직이게 되는 것이다.

## 5. 동적 GBFCM

동적 GBFCM은 동적 K-means와 Kohonen의 단점을 보완하기 위해서 고안되었다. 이 알고리즘은 GBFCM에서 정의된 멤버십 등급  $\mu$ 를 사용한다.

### 5.1 동적 GBFCM 알고리즘

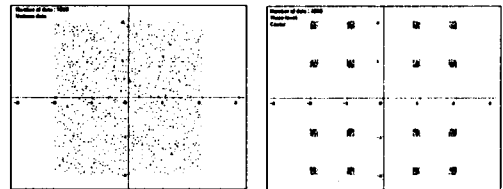
구현하기위한 알고리즘은 다음과 같다 :

```
while (NMSE > limit error)
for (Input file is not empty)
  Read one datum  $x_i$ 
  Variation Update
   $c_{k+1} = c_k - \eta \cdot \mu_{ki} \cdot (c_k - x_i)$ 
   $\mu_{ik} = \frac{1}{\sum_{j=1}^K \left( \frac{d_j(x_k)}{d_j(x_k)} \right)}$ 
  e(k) = NMSE(k)
End for
Output  $\mu_{ik}$  and  $\eta$ 
```

## 6. 실험 및 결과

우리는 2개의 입력 데이터 형태에 대해서 실험을 해보았다. 그리고 2개의 데이터 형태를 가지고 서로 비교해 보았다.

최대 주기 수(epoch)는 200이고,  $\eta$ 는 0.5이고  $a$ 는 0.99 그리고 변형의 초기치는 0.1로 놓았다.



(a) 데이터 1000개 (b) 데이터 1008개

그림 1. 입력 데이터 형태

예러비고

데이터 형태 (a)

Algorithm	NMSE(4)	NMSE(8)	NMSE(16)
Kohonen	0.970840	0.990393	0.927219
DGBFCM	0.984596	1.041096	0.955705

표 1

데이터 형태 (b)

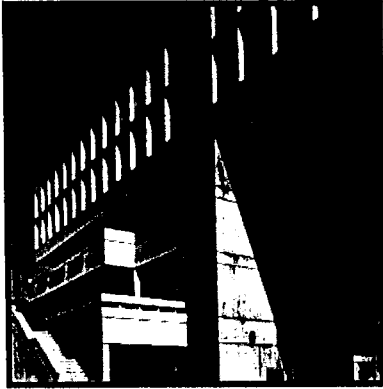
Algorithm	NMSE(4)	NMSE(8)	NMSE(16)
Kohonen	0.459136	0.461186	0.251669
DGBFCM	0.551691	0.411221	0.022024

표 2

위의 표에서 괄호 안의 숫자는 센터들의 개수를 의미한다. 표 1에서는 2가지 알고리즘의 에러가 큰 차이를 보이지 않고 최종 센터들의 위치도 거의 비슷하다. 그러나 표 2를 보면 센터가 16일 경우 에러의 차이가 크고 최종 센터들의 위치도 DGBFCM에비해 Kohonen's algorithm은 좋지 않은 결과를 나타냈다.

## 7. 이미지 압축

아래의 그림은 이미지 압축에 사용된 원래의 이미지이다.



Original

사용된 입력 이미지 크기는 256 by 256 이진 그레이 스케일 이미지이고, 입력 블록 크기는 2 by 2이다. 그래서 총 블록 크기는 16384 블록이고 센터들의 개수는 100개로 정했으므로대략(164:1)의 압축이 된다.

사용된 SNR 공식은

$$SNR = 10 \log \left( \frac{1000^2}{MSE} \right) \text{ dB} \quad (13)$$

이미지 압축 실험에서보면 DGBFCM이 Kohonen algorithm 보다 SNR(signal-to noise)이 크다. 즉 MSE가 적다는 의미가 된다.

## 8. 결론

본 논문에서 주로 강조된 부분은 이전의 Kohonen network의 단점을 극복하고자 하는 목적에서 GBFCM 알고리즘을 향상시켰다. 이 알고리즘의 장점은 센터들의 고립 현상을 방지할 수 있고 초기치 값들에 덜 민감하다는 것이다. 앞으로 계속하고자 하는 것은 Kohonen network의 적용 분야에서 DGBFCM을 적용해보고자 한다.

### Acknowledgement

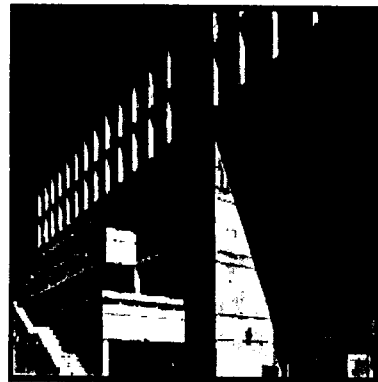
본 연구는 한국과학재단 95핵심전문연구(과제번호 951 - 0906 - 062 - 2)의 연구비 지원으로 수행되었습니다.

### 참고문헌

[1] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well separated clusters", J. Cybern, 1973. (3)32-75



Kohonen (SNR = 8.486958)



DGBFCM (SNR = 14.059511)

그림 2. 이미지 압축 결과

- [2] Dong C. Park and Issam Daugher, "GBFCM", IEEE Int. Conf. on Neural Networks, ICNN-94 vol.3, pp1626 - 1631
- [3] J.C.Bezdek, "A convergence theorem for the fuzzy ISODATA clustering algorithms", "IEEE Trans. Pattern Anal. Mach. Ini., 1980, (2):1-8. 1975. (24)835-838
- [4] J.C.Dezdek, Pattern recognition with fuzzy objective function algorithms", "New York : Plenum, 1981.
- [5] Chedsada Chinrungrueng and Carlo H. Sequin, "Optimal Adaptive K-Means Algorithm with Dynamic Adjustment of Learning Rate", "IEEE Trans. on Neural Networks", vol. 6, pp157 - 169, 1995.
- [6] J. Tou and R. C. Gonzalez, Pattern Recognition Principles. Reading, MA:Addison-Wesley, 1974
- [7] Laurene Fausett, Fundamentals of Neural Network, Prentice Hall, 1994
- [8] Robert Schalkloff, Pattern Recognition, Wiley, 1992
- [9] Ferrel G. Stremler, Introduction Communication System, Addison-Wesley 1990
- [10] B.S. Gottfried, J.Weisman, Introduction to optimization theory, Prentice Hall, 1973