

## 레귤러라이제이션 기반 개선된 밀도 무관 퍼지 클러스터링

한수환<sup>1</sup> · 허경용<sup>2\*</sup>

### Improved Density-Independent Fuzzy Clustering Using Regularization

Soowhan Han<sup>1</sup> · Gyeongyong Heo<sup>2\*</sup>

<sup>1</sup>Professor, Department of Applied Software Engineering, Dong-eui University, Busan, 47340 Korea

<sup>2\*</sup>Associate Professor, Department of Electronic Engineering, Dong-eui University, Busan, 47340 Korea

#### 요 약

FCM(Fuzzy C-Means)으로 대표되는 퍼지 클러스터링은 간단하면서도 효율적인 클러스터링 방법이지만, FCM에서 사용하는 목적 함수에서는 밀도가 높은 클러스터가 클러스터링 결과에 많은 영향을 미치도록 함으로써 클러스터 사이의 밀도 차에 의해 클러스터링 결과에 왜곡이 발생할 수 있다. 이러한 밀도 문제를 완화하는 방법의 하나로 FCM의 목적 함수에 밀도 차이를 보정할 수 있는 항을 추가한 EDI-FCM(Extended Density-Independent FCM)이 있다. 이 논문에서는 레귤러라이제이션을 이용하여 EDI-FCM을 보완한 Regularized EDI-FCM을 제안한다. 레귤러라이제이션은 해공간을 평탄화하고 잡음 민감성을 줄이기 위해 흔히 사용되는 방법으로, 클러스터링에서는 특정 클러스터가 클러스터링 결과에 미치는 영향을 줄이는 역할을 한다. 제안하는 방법은 FCM이나 EDI-FCM과 비교했을 때 실제 클러스터 중심에 빠르고 정확하게 수렴한다는 것을 실험 결과를 통해 확인할 수 있다.

#### ABSTRACT

Fuzzy clustering, represented by FCM(Fuzzy C-Means), is a simple and efficient clustering method. However, the object function in FCM makes clusters affect clustering results proportional to the density of clusters, which can distort clustering results due to density difference between clusters. One method to alleviate this density problem is EDI-FCM(Extended Density-Independent FCM), which adds additional terms to the objective function of FCM to compensate for the density difference. In this paper, proposed is an enhanced EDI-FCM using regularization, Regularized EDI-FCM. Regularization is commonly used to make a solution space smooth and an algorithm noise insensitive. In clustering, regularization can reduce the effect of a high-density cluster on clustering results. The proposed method converges quickly and accurately to real centers when compared with FCM and EDI-FCM, which can be verified with experimental results.

**키워드** : 퍼지 클러스터링, 유클리드 거리, 클러스터 밀도, 밀도 무관 클러스터링, 레귤러라이제이션

**Keywords** : Fuzzy clustering, Euclidean distance, Cluster density, Density-independent clustering, Regularization

Received 23 September 2019, Revised 29 September 2019, Accepted 19 October 2019

\* Corresponding Author Gyeongyong Heo(E-mail:hgycap@deu.ac.kr, Tel:+82-51-890-1675)

Associate Professor, Department of Electronic Engineering, Dong-eui University, Busan, 47340 Korea

Open Access <http://doi.org/10.6109/jkiice.2020.24.1.1>

print ISSN: 2234-4772 online ISSN: 2288-4165

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.  
Copyright © The Korea Institute of Information and Communication Engineering.

## I. 서론

클러스터링은 대표적인 비교사 학습 방법의 하나로 유사성을 기준으로 주어진 데이터로부터 클러스터 또는 그룹을 묶는 방법이다. 부분적인 소속을 나타내는 퍼지 집합이 클러스터링에 도입된 퍼지 클러스터링은 Bezdek[1]이 일반화한 Fuzzy C-Means(FCM)가 대표적이다. FCM은 간단하면서도 효과적인 방법으로 다양한 응용에 널리 사용되고 있지만, 해결해야 할 문제점 중 하나는 밀도가 높은 클러스터가 클러스터링 결과에 많은 영향을 주어 클러스터의 중심이 왜곡되는 현상이 발생할 수 있다는 것이다. 실제 데이터의 경우 모든 클러스터가 비슷한 밀도를 가지는 경우는 찾아보기 어렵다.

클러스터의 밀도 차이에 의한 영향을 줄이는 방법으로는 밀도 무관 FCM(Density-Independent FCM, DI-FCM)이 있다[2]. DI-FCM은 FCM의 목적 함수에 클러스터 중심이 멀리 떨어지도록 하는 항을 추가하여 FCM에 비해 나은 성능을 얻었다. 이를 더 보완한 것이 EDI-FCM(Extended DI-FCM)으로 클러스터 중심이 밀도가 높은 영역에 위치하도록 하는 항이 추가되었다[3].

이 논문에서는 클러스터의 밀도 차이에 효과적으로 대응할 수 있는 방법의 하나로 EDI-FCM에 레귤러라이제이션을 적용한 REDI-FCM(Regularized EDI-FCM)을 제안한다. 레귤러라이제이션은 해공간을 평탄화하므로 밀도에 따른 클러스터의 영향을 줄이는 역할을 할 수 있으며[4], 실험 결과 이전 방법들보다 정확한 클러스터링 결과를 얻을 수 있었다.

이 논문의 구성은 다음과 같다. 먼저 2장에서는 밀도 무관 클러스터링에 대해 살펴보고, 3장에서는 클러스터링에서의 레귤러라이제이션에 대해 알아본다. 4장에서는 이 논문에서 제안하는 REDI-FCM을 유도하고, 5장 실험 결과를 통해 제안하는 방법의 유효성을 보인다. 결론 및 향후 연구 방향은 6장에서 언급한다.

## II. 밀도 무관 퍼지 클러스터링

### 2.1. FCM : Fuzzy C-Means

퍼지 클러스터링은 역사가 오랜 만큼 다양한 방법이 제안되었지만, 가장 기본적인면서도 대표적인 방법이 FCM이다[5].  $N$ 개의  $d$ 차원 데이터  $X = \{x_i | 1 \leq i \leq N,$

$x_i \in R^d\}$ 가 주어졌을 때, 이를  $C$ 개 클러스터로 묶기 위해 FCM은 식 (1)의 목적 함수를 최소화 하는 클러스터 중심과 소속도를 구한다.

$$J_{FCM} = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m d_{ij}^2 \quad (1)$$

$$= \sum_{i=1}^N \sum_{j=1}^C u_{ij}^2 \|v_j - x_i\|^2$$

식 (1)에서  $u_{ij}$ 는  $i$ 번째 데이터 포인트  $x_i$ 가  $j$ 번째 클러스터에 소속되는 정도를 나타내는 소속도 값을,  $v_j$ 는  $j$ 번째 클러스터의 중심을,  $m$ 은 퍼지화 정도를 나타내는 상수로 일반적으로 2를 사용한다( $1 < m < \infty$ ).  $d_{ij}$ 는  $i$ 번째 데이터 포인트와  $j$ 번째 클러스터 중심 사이의 거리를 나타내며 여기서는 유클리드 거리를 사용한다.

퍼지 클러스터링에서 하나의 데이터 포인트  $x_i$ 는  $C$ 개의 클러스터에 소속될 수 있지만,  $C$ 개 클러스터에 소속되는 소속도 합은 1이 되어야 한다(sum-to-one)는 제약조건을 만족시켜야 한다.

$$\sum_{j=1}^C u_{ij} = 1 \quad (2)$$

식 (2)의 제약조건을 만족시키면서 식 (1)의 목적 함수를 최소화하는 소속도와 클러스터 중심은 반복 최적화를 통해 구할 수 있으며, 최적화 과정에서의 갱신식은 라그랑주 승수법(Lagrange multiplier method)을 이용하여 얻을 수 있다. 식 (3)은 클러스터 중심  $v_j$ 의 갱신식을, 식 (4)는 소속도  $u_{ij}$ 의 갱신식을 나타낸다.

$$v_j = \frac{\sum_{i=1}^N u_{ij}^2 x_i}{\sum_{i=1}^N u_{ij}^2} \quad (3)$$

$$u_{ij} = \frac{1}{\|v_j - x_i\|^2} \frac{1}{\sum_{k=1}^C \frac{1}{\|v_k - x_i\|^2}} \quad (4)$$

### 2.2. EDI-FCM : Extended Density Independent FCM

주어진 데이터가 밀도가 다른 몇 개의 클러스터로 구성되는 경우 클러스터의 중심은 밀도가 높은 클러스터

쪽으로 쏠리는 현상이 발생한다. 이는 FCM과 같이 목적 함수를 사용하는 방법에서는 밀도가 높은 클러스터, 즉, 많은 데이터를 가진 클러스터가 클러스터링 결과에 많은 영향을 미치기 때문이다. 퍼지 클러스터링에서 흔히 사용되는 유클리드 거리 역시 밀도가 높은 클러스터의 영향이 커지도록 하는 역할을 한다.

클러스터 중심의 쏠림 현상을 방지하는 방법은 여러 가지가 있을 수 있지만, 이 논문에서는 FCM의 목적 함수에 항을 추가하는 방법을 사용한다. FCM의 목적 함수에 항을 추가하여 수정한 방법들은 FCM과 같은 연산 복잡도를 가지며, FCM과 같은 입력 및 출력을 사용하므로 기존 FCM이 적용된 시스템에 제안하는 방법을 쉽게 적용할 수 있는 장점이 있지만, 밀도 무관 클러스터링 기법들 이외에는 관련 연구를 찾아보기가 어렵다.

식 (5)는 EDI-FCM(Extended Density Independent FCM)의 목적 함수로 FCM의 목적 함수에 클러스터의 중심이 멀리 떨어져 있도록 하는 첫 번째 항과 클러스터의 중심이 밀도가 높은 곳에 있도록 하는 두 번째 항이 추가되었다[3]. 이때  $\alpha$ 와  $\beta$ 는 각 항이 목적 함수에 반영되는 정도를 결정하는 상수다.

$$\begin{aligned}
 J_{EDI} &= \sum_{i=1}^N \sum_{j=1}^C u_{ij}^2 \|v_j - x_i\|^2 \\
 &\quad - \alpha \sum_{a=1}^C \sum_{b=1}^C \|v_a - v_b\|^2 \\
 &\quad + \beta \sum_{i=1}^N \sum_{j=1}^C \|v_j - x_i\|^2 \\
 &= \sum_{i=1}^N \sum_{j=1}^C (u_{ij}^2 + \beta) \|v_j - x_i\|^2 \\
 &\quad - \alpha \sum_{a=1}^C \sum_{b=1}^C \|v_a - v_b\|^2
 \end{aligned} \tag{5}$$

EDI-FCM을 위한 소속도 갱신식은 FCM의 경우와 같지만, 클러스터 중심의 갱신식은 식 (6)과 같이 추가된 항이 반영되어 나타난다.

$$v_j = \frac{\sum_{i=1}^N (u_{ij}^2 + \beta) x_i - 2\alpha \sum_{b=1}^C v_b}{\sum_{i=1}^N (u_{ij}^2 + \beta) - 2\alpha C} \tag{6}$$

### III. 레귤러라이제이션

많은 패턴인식 문제들은 식 (7)과 같은 선형 연립 방정식을 푸는 문제로 변환될 수 있다.

$$Az = y \tag{7}$$

이때  $A$ 는 연산자,  $z$ 는 해를 구하고자 하는 변수, 그리고  $y$ 는 입력 변수를 나타낸다. 식 (7)에서 해  $z$ 는  $A$ 의 역행렬을 통해 구할 수 있지만, 모든 문제에서 역행렬을 구할 수 있는 것은 아니다. 이러한 문제점을 해결하기 위해 Tikhonov는 레귤러라이제이션을 사용하여 근사해를 구하는 방법을 제안하였다[6]. 레귤러라이제이션을 이용한 근사해는 식 (8)의 최적화 문제로 나타낼 수 있다.

$$\min_z \|Az - y\| + \gamma \Phi(z) \tag{8}$$

이때  $\gamma$ 는 레귤러라이제이션 상수를,  $\Phi$ 는 제약조건을 나타낸다. 제약조건은 가능한 해의 범위를 줄이는 역할을 하지만, 제약조건을 선택하는 일반적인 방법은 알려지지 않았다. 레귤러라이제이션은 클러스터링에도 적용되어 몇 가지 제약조건이 제안되었으며[7-9], 이 논문에서는 가장 많이 사용되는 소속도 값의 제곱 항을 제약조건으로 사용하였다. 식 (9)는 FCM의 목적 함수에 레귤러라이제이션 항을 추가한 목적 함수를 나타낸다.

$$J_R = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^2 \|v_j - x_i\|^2 + \gamma \sum_{i=1}^N \sum_{j=1}^C u_{ij}^2 \tag{9}$$

클러스터링에 사용되는 레귤러라이제이션 항은 약간씩 차이가 있지만, 모든 레귤러라이제이션 항은  $u_{ik}$ 가 모두 동일한 값,  $1/C$ 을 가질 때 최소가 된다. 즉, 모든 레귤러라이제이션 항은  $u_{ik}$ 가 0이나 1의 극단적인 값을 가지지 않도록 함으로써 해공간(solution space)을 평탄화하고 잡음 민감성을 줄이는 역할을 한다. 밀도가 다른 클러스터로 이루어진 데이터의 경우에는 특정 클러스터, 특히 밀도가 높은 클러스터가 클러스터링 결과에 미치는 영향을 감소시키는 효과를 얻을 수 있다. 또한, 해공간의 평탄화를 통해 국부 최적해에 빠지는 확률을 줄여 안정적인 동작을 얻을 수 있고, 수렴 속도 역시 빨라지는 효과를 얻을 수 있다.

#### IV. REDI-FCM

식 (10)은 식 (5)의 EDI-FCM에 레귤러라이제이션 항을 추가한 REDI-FCM의 목적 함수다.

$$J_{REDI} = \sum_{i=1}^N \sum_{j=1}^C (u_{ij}^2 + \beta) \|v_j - x_i\|^2 - \alpha \sum_{a=1}^C \sum_{b=1}^C \|v_a - v_b\|^2 + \gamma \sum_{i=1}^N \sum_{j=1}^C u_{ij}^2 \quad (10)$$

이때 소속도는 식 (2)의 제약조건을 만족시켜야 하므로 라그랑지 방정식은 식 (11)과 같이 나타낼 수 있다. 이때  $\lambda = [\lambda_1, \dots, \lambda_N]$ 는 라그랑지 상수를 나타낸다.

$$L = \sum_{i=1}^N \sum_{j=1}^C (u_{ij}^2 + \beta) \|v_j - x_i\|^2 - \alpha \sum_{a=1}^C \sum_{b=1}^C \|v_a - v_b\|^2 + \gamma \sum_{i=1}^N \sum_{j=1}^C u_{ij}^2 - \sum_{i=1}^N \lambda_i \left( \sum_{j=1}^C u_{ij} - 1 \right) \quad (11)$$

먼저 라그랑지 방정식을  $u_{ij}$ 로 편미분하고 영이 되도록 하는  $u_{ij}$ 를 구하면 식 (12)와 같다.

$$u_{ij} = \frac{\lambda_i / 2}{\|v_j - x_i\|^2 + \gamma} \quad (12)$$

식 (12)와 식 (2)의 제약조건을 통해 라그랑지 상수는 식 (13)과 같이 구할 수 있다.

$$\frac{\lambda_i}{2} = \frac{1}{\sum_{j=1}^C \frac{1}{\|v_j - x_i\|^2 + \gamma}} \quad (13)$$

식 (13)을 사용하여 소속도 갱신식 식 (12)는 식 (14)와 같이 나타낼 수 있다.

$$u_{ij} = \frac{1}{\|v_j - x_i\|^2 + \gamma} \frac{1}{\sum_{k=1}^C \frac{1}{\|v_k - x_i\|^2 + \gamma}} \quad (14)$$

식 (11)의 라그랑지 방정식을  $v_j$ 에 대해 편미분하고 영의 값을 가지는  $(v_j - v_j)$  항을 추가하여 정리하면 식 (15)를 얻을 수 있다.

$$\begin{aligned} \frac{\partial L}{\partial v_j} &= 2 \sum_{i=1}^N (u_{ij}^2 + \beta)(v_j - x_i) - 2\alpha \sum_{\substack{b=1 \\ b \neq j}}^C (v_j - v_b) \\ &= 2 \sum_{i=1}^N (u_{ij}^2 + \beta)(v_j - x_i) - 2\alpha \sum_{\substack{b=1 \\ b \neq j}}^C (v_j - v_b) - 2\alpha(v_j - v_j) \\ &= 2 \sum_{i=1}^N (u_{ij}^2 + \beta)(v_j - x_i) - 2\alpha \sum_{\substack{b=1 \\ b \neq j}}^C (v_j - v_b) \end{aligned} \quad (15)$$

식 (15)가 영이 되도록 하는  $v_j$ 를 구하면 클러스터 중심에 대한 갱신식 식 (16)을 얻을 수 있다.

$$v_j = \frac{\sum_{i=1}^N (u_{ij}^2 + \beta)x_i - \alpha \sum_{b=1}^C v_b}{\sum_{i=1}^N (u_{ij}^2 + \beta) - \alpha \sum_{b=1}^C 1} = \frac{\sum_{i=1}^N (u_{ij}^2 + \beta)x_i - \alpha \sum_{b=1}^C v_b}{\sum_{i=1}^N (u_{ij}^2 + \beta) - \alpha C} \quad (16)$$

식 (12)와 식 (16)을 사용하여 REDI-FCM 알고리즘은 그림 1과 같이 나타낼 수 있다. 클러스터링은 동일한 소속도 값으로 초기화한 후 클러스터 중심을 먼저 갱신하는 순서로 이루어진다. 그림 1에서 수렴 조건은 클러스터 중심의 최대 이동 거리가  $\epsilon_1 (= 10^{-3})$  보다 작고, 소속도의 최대 변화량이  $\epsilon_2 (= 10^{-4})$  보다 작은 것으로 설정하였다.

<p>U : data point membership V : cluster center</p>
<p>1: initialize U with equal values 2: do 3: update V using Eq. (12) 4: update U using Eq. (10) 5: while U and V do not satisfy convergence qualification 6: return U and V</p>

Fig. 1 REDI-FCM algorithm

표 1은 FCM, EDI-FCM, REDI-FCM의 소속도 갱신식을 비교한 것이다. FCM과 EDI-FCM은 같은 소속도 갱신식을 사용한다. 하지만 REDI-FCM의 소속도 갱신식은 레귤러라이제이션 상수  $\gamma$ 가 반영된 것으로, 상수

를 더한 유클리드 거리를 새로운 거리 척도로 정의한다면 기본적으로 세 가지 방법은 같은 형태의 갱신식이 된다. REDI-FCM은 레귤러라이제이션 상수  $\gamma$ 를 통해 소속도 값이 일정 수준 이상의 값을 가지도록 보장함으로써 개별 데이터 포인트  $x_i$ 가 클러스터링 결과에 미치는 영향을 줄여준다.

**Table. 1** Update equation for membership values

Clustering method	Update equation ( $1 \leq i \leq N$ )
FCM	$u_{ij} = \frac{1}{\ v_j - x_i\ ^2 + \sum_{k=1}^C \frac{1}{\ v_k - x_i\ ^2}}$
EDI-FCM	
REDI-FCM	$u_{ij} = \frac{1}{\ v_j - x_i\ ^2 + \gamma + \sum_{k=1}^C \frac{1}{\ v_k - x_i\ ^2 + \gamma}}$

표 2는 FCM, EDI-FCM, REDI-FCM의 클러스터 중심 갱신식을 비교한 것으로 EDI-FCM과 REDI-FCM은 같은 갱신식을 사용한다. 갱신식에서는  $\alpha$ 를 통해 클러스터 중심 거리에 대한 제약이,  $\beta$ 를 통해 클러스터 중심 위치에 대한 제약이 반영되어 클러스터의 중심이 결정된다.

**Table. 2** Update equation for cluster centers

Clustering method	Update equation ( $1 \leq j \leq C$ )
FCM	$v_j = \frac{\sum_{i=1}^N u_{ij}^m x_i}{\sum_{i=1}^N u_{ij}^m}$
EDI-FCM	$v_j = \frac{\sum_{i=1}^N (u_{ij}^2 + \beta) x_i - 2\alpha \sum_{b=1}^C v_b}{\sum_{i=1}^N (u_{ij}^2 + \beta) - 2\alpha C}$
REDI-FCM	

표 1과 표 2에서도 알 수 있듯이 REDI-FCM은 상수 설정에 따라 EDI-FCM이나 FCM으로 동작할 수 있으므로 REDI-FCM은 이들 모두를 포함하는 클러스터링 기법으로 생각할 수 있다.

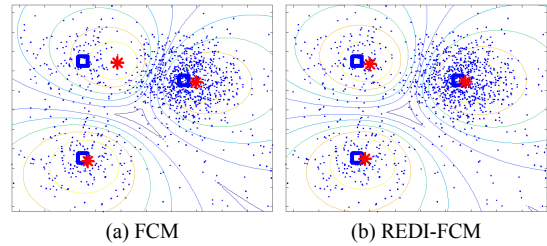
## V. 실험 결과

소개된 다양한 FCM의 변형 중 밀도 차이를 FCM의 목적 함수 변형을 통해 해결하고자 하는 시도는 EDI-FCM의 유일하므로 제안하는 REDI-FCM은 FCM 및 EDI-FCM과 그 결과를 비교하였다. 또한, 이전 방법들과의 비교를 위해 이전 논문[3]에서의 실험 방법 및 절차를 따랐다.

클러스터링은 비교사 학습 방법으로 데이터가 소속된 클러스터를 알 수 없으므로 정량적인 비교를 위해 이 논문에서는 식 (17)의 오류 함수를 사용하였다. 오류는 샘플 데이터 생성에 사용된 실제 클러스터 중심  $V_R = \{v_{R,j} | 1 \leq j \leq C\}$ 과 클러스터링 결과로 얻어진 클러스터 중심  $V_{FC} = \{v_{FC,j} | 1 \leq j \leq C\}$  사이의 거리 합을 나타낸다.

$$Error = \sum_{j=1}^C \|v_{R,j} - v_{FC,j}\|^2 \quad (17)$$

그림 2는 FCM과 밀도 무관 FCM을 적용한 전형적인 결과를 보여준다. 샘플 데이터는 3개의 클러스터로 구성되어 있으며 오른쪽 위의 밀도가 높은 클러스터는 500개의 데이터 포인트를, 나머지 2개의 클러스터는 100개의 데이터 포인트를 가우스 분포에서 생성하였다. 또한, 100개의 균일 분포를 갖는 잡음 데이터를 추가하였다.



**Fig. 2** Clustering results with 3-cluster data

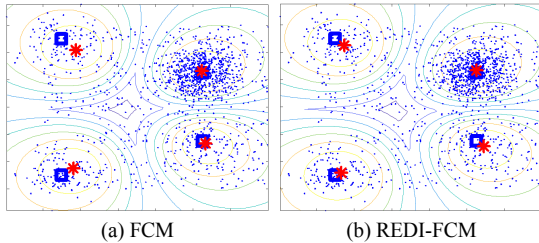
그림 2에서 ‘□’은 샘플 데이터 생성에 사용된 클러스터의 중심을, ‘\*’는 클러스터링 결과로 찾아진 클러스터의 중심을 나타낸다. 실험 결과에서 알 수 있듯이 FCM의 경우 클러스터의 중심이 밀도가 높은 클러스터 쪽으로 치우치지만, 밀도 무관 FCM의 경우 클러스터 중심이 멀어지면서 실제 클러스터 중심에 가까워지고 있다.

**Table. 3** Clustering results with 3-cluster data

Method	Error	Number of Iterations
FCM	1.652	60.17
EDI-FCM	1.572	51.53
REDI-FCM	1.552	47.85

표 3은 3가지 클러스터링 방법을 3개의 클러스터를 가지는 샘플 데이터에 적용한 결과를 요약한 것이다. 표에 나타난 값은 임의로 생성된 데이터에 대한 클러스터링을 500회 시행한 후 평균한 값이다. 표 3에서 볼 수 있듯이 REDI-FCM은 다른 두 가지 방법과 비교할 때 실제 클러스터 중심에 더 근접하고 있으며 더 빠른 속도로 수렴하고 있다.

그림 3은 4개의 클러스터로 구성되는 샘플 데이터에 FCM과 밀도 무관 클러스터링을 적용한 결과의 예다. 4개의 클러스터 중 오른쪽 위의 클러스터는 400개의 데이터 포인트를 가지고 나머지 3개 클러스터는 100개의 데이터 포인트를 가지며, 100개의 균일 분포를 갖는 잡은 데이터를 추가하였다.



**Fig. 3** Clustering results with 4-cluster data

표 4는 3가지 클러스터링 방법을 4개의 클러스터를 가지는 샘플 데이터에 적용한 결과를 요약한 것이다. 이전 실험 결과와 마찬가지로 레귤러라이제이션 항의 추가를 통해 빠르고 정확하게 실제 클러스터 중심을 찾아내고 있음을 확인할 수 있다.

**Table. 4** Clustering results with 4-cluster data

Method	Error	Number of Iterations
FCM	1.701	25.77
EDI-FCM	1.109	24.62
REDI-FCM	0.981	23.98

## VI. 결론

이 논문에서는 클러스터의 밀도 차이로 클러스터링 결과가 왜곡되는 현상을 줄이기 위해 FCM의 목적 함수를 수정한 새로운 퍼지 클러스터링 기법을 제안하였다. 제안하는 방법에서는 기존에 추가된 클러스터 중심의 위치를 제한하는 항에 더하여, 해공간을 평탄화하는 레귤러라이제이션 항이 추가되었다. 레귤러라이제이션이 클러스터링에 적용된 경우 소속도 값이 비슷한 값을 가지도록 함으로써 밀도가 높은 클러스터가 클러스터링 결과에 미치는 영향을 줄이는 결과를 가져온다. 또한, 평탄화 된 해공간으로 수렴 속도 역시 빨라지는 효과를 얻을 수 있다.

레귤러라이제이션 항은 잡음의 영향을 줄이기 위한 목적으로 흔히 사용되며 이 논문에서는 밀도에 따른 클러스터링 오류를 줄이는 목적으로 사용하였다. 밀도가 높은 클러스터의 데이터 포인트 역시 잡음으로 생각할 수 있지만, 레귤러라이제이션 항은 모든 데이터 포인트를 같은 정도의 잡음으로 간주한다는 한계는 있다. 잡음 정도를 결정하고 이를 레귤러라이제이션 항을 통해 반영할 수 있다면 밀도 차이에 의한 클러스터링 오류를 더 줄일 수 있을 것으로 생각되며, 이에 관한 연구가 현재 진행 중이다.

## ACKNOWLEDGEMENT

This Work was supported by Dong-eui University Foundation Grant(2019).

## REFERENCES

- [1] J. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, New York, Springer, 1981.
- [2] B. H. Yoo, W. W. Kim, and G. Heo, "An Improved Clustering Method with Cluster Density Independence," *Journal of the Korea Society of Computer and Information*, vol. 20, no. 12, pp. 15-20, Dec. 2015.
- [3] S. H. Kim, and G. Heo, "Improvement on Density Independent Clustering Method," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 21, no. 5, pp. 967-973, May. 2017.

- [ 4 ] X. Chang, Q. Wang, Y. Liu, and Y. Wang, "Sparse Regularization in Fuzzy c-Means for High-Dimensional Data Clustering," *IEEE Transactions on Cybernetics*, vol. 47, no. 9, pp. 2616-2627, Sep. 2017.
- [ 5 ] J. Nayak, "Fuzzy C-means(FCM) Clustering Algorithm: A Decade Review from 2000 to 2014," *Systems and Technologies*, vol. 32, no. 2, pp. 133-179, Dec. 2014.
- [ 6 ] M. Fuhry, and L. Reichel, "A new Tikhonov regularization method," *Numerical Algorithms*, vol. 59, no. 3, pp. 433-445, Mar. 2012.
- [ 7 ] S. Susan, P. Sharawat, S. Singh, R. Meena, A. Verma, and M. Kumar, "Fuzzy C-means with non-extensive entropy regularization," in *Proceedings of 2015 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems*, Kozhikode, India, pp. 1-5, 2015.
- [ 8 ] Y. Kanzawa, "Power-regularized fuzzy clustering for spherical data," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 22, no. 2, pp. 163-171, Mar. 2018.
- [ 9 ] E. Yasunori, T. Isao, H. Yukihiko, and M. Sadaaki, "Kernelized fuzzy c-means clustering for uncertain data using quadratic penalty-vector regularization with explicit mappings," in *Proceedings of 2011 IEEE International Conference on Fuzzy Systems*, Taipei, Taiwan, pp. 804-809, 2011.



**한수환(Soowhan Han)**

연세대학교 전자공학학사 (1986)  
 Florida Tech, 전기전자공학석사 (1990)  
 Florida Tech, 전기전자공학박사 (1993)  
 동의대학교 응용소프트웨어공학과 교수 (1997~현재)  
 ※ 관심분야 : 디지털신호처리, 영상처리, 패턴인식, 인공지능



**허경용(Gyeongyong Heo)**

연세대학교 전자공학과 공학석사 (1996)  
 University of Florida 컴퓨터공학과 공학박사 (2009)  
 동의대학교 전자공학과 교수 (2012~현재)  
 ※ 관심분야 : 인공지능, 패턴인식, IoT 시스템