

클러스터링 기반 RBFNNs를 이용한 기상레이더 패턴분류기 설계 : 비교 연구 및 해석

Design of Meteorological Radar Pattern Classifier Using Clustering-based RBFNNs : Comparative Studies and Analysis

최우용 · 오성권[†]

Woo-Yong Choi, Sung-Kwun Oh[†]

수원대학교 전기공학과

Department of Electrical Engineering, The University of Suwon

요 약

기상레이더를 통해 취득된 데이터에는 지형에코, 파랑에코, 이상에코, 그리고 청천에코등이 존재한다. 각 에코는 여러 종류의 비강수에코이고, 이 비강수에코를 제거하기 위해 각 에코들의 특성을 분석하였다. 기상레이더 데이터는 매우 방대한 양이기 때문에 전처리 절차를 통해 분석된다. 본 논문에서는 클러스터링 기반 방사형 기저함수 신경회로망(RBFNNs : Radial Basis Function Neural Networks)과 에코 판단 모듈을 이용하여 기상레이더 데이터에서 강수에코와 비강수에코들을 구별하기 위한 에코 패턴분류기를 설계하였다. HCM(Hard C-Mean) 클러스터링 기반 RBFNNs 와 FCM(Fuzzy C-Mean) 클러스터링 기반 RBFNNs를 이용하여 출력성능은 비교 및 분석된다.

키워드 : 방사형 기저함수 신경회로망, 판단 모듈, FCM, HCM

Abstract

Data through meteorological radar includes ground echo, sea-clutter echo, anomalous propagation echo, clear echo and so on. Each echo is a kind of non-precipitation echoes and the characteristic of individual echoes is analyzed in order to identify with non-precipitation. Meteorological radar data is analyzed through pre-processing procedure because the data is given as big data. In this study, echo pattern classifier is designed to distinguish non-precipitation echoes from precipitation echo in meteorological radar data using RBFNNs and echo judgement module. Output performance is compared and analyzed by using both HCM clustering-based RBFNNs and FCM clustering-based RBFNNs.

Key Words : RBFNNs, Echo Judgement module, Fuzzy C-mean Clustering, Hard C-mean Clustering.

1. 서 론

기상레이더는 여러 기상에 대한 이동상황 등을 정확하게 관측하고 예보하기 위해 사용된다. 하지만 강수상황 및 날

이 맑은 상황에서 생성되는 비강수에코로 인해 정확한 예보를 어렵게 한다.

본 논문에서는 기상청에서 사용되고 있는 단일편파레이더인 오성산 레이더 데이터 자료를 입력데이터로 사용하였고, 2012년에 발생했던 비강수사례를 선정하여 데이터를 구성하였다. 강수에코와 비강수에코를 판별하여 제거하는 것은 쉽지 않기 때문에, 제안된 지능형 뉴로-퍼지인 HCM 및 FCM 클러스터링 기반 방사형 기저함수 신경회로망[1]과 판단 모듈을 이용하여 각각의 기상레이더 패턴분류기를 설계했다. 그리고 레이더의 필터링을 거치지 않은 DZ와 레이더에서 필터링을 거쳐 나오는 CZ와 기상청에서 현업에 사용되고 있는 QC 데이터와의 비교를 통하여 성능을 판단했다.[1]

접수일자: 2014년 3월 9일

심사(수정)일자: 2014년 4월 4일

게재확정일자 : 2014년 9월 23일

[†] Corresponding author

본 연구는 경기도의 경기도지역협력연구센터사업의 일환으로 수행하였음[GRRRC 수원2014-B2, U-city 보안감시 기술협력센터] 그리고 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (NRF-2012R1A1B3003568)

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.

2. 기상레이더 데이터의 특성분석 및 입력데이터 생성

기상레이더 자료인 UF(Universal Format)데이터의 구조

를 분석하였으며, 입력변수를 생성하여 각 입력변수에 의한 강수예코 및 비강수예코(지형, 청천, 파랑, 이상예코)등의 특성을 파악하였다. [2]

UF 기상레이더 자료는 레이더 설치 지역마다의 레이더에서 관측되는 자료들을 바이너리(binary)형태 자료로 저장한 데이터를 말한다. 저장되는 반사도는 dBZ(decibel Z)라는 단위를 사용한다.[3]

$$dBZ = 10 \cdot \log_{10} \left(\frac{Z}{1mm^6/m^3} \right) \quad (1)$$

이는 단위부피 $1m^3$ 안에 직경 $1mm$ 인 물방울이 한 개 있으면 0[dBZ], 10개 있으면 10[dBZ], 100개 있으면 20[dBZ], 1000개 있으면 30[dBZ]가 되는 것이다.

UF 데이터는 필터링 후 반사도(CZ), 필터링 후 시선속도(VR), 필터링 후 스펙트럼 폭(SW), 필터링 전 반사도(DZ)가 저장된다. 본 연구에서는 반사도의 값이 나타나는 CZ와 DZ를 이용하여 입력데이터를 구성하였고, 확장된 데이터는 다음과 같다.[4]

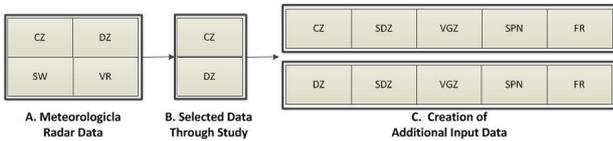


그림 1. 기상레이더 데이터를 이용한 입력데이터 구성
Fig. 1. Composition of input data using meteorological radar data

SDZ (Standard Deviation of Reflectivity)	VGZ (Vertical Gradient of Reflectivity)
$SDZ = \left[\frac{1}{n} \sum_{i=1}^n (dBZ_i - \overline{dBZ})^2 \right]^{1/2}$	$VGZ = \frac{dBZ_{\theta_{(i,j,k+1)}} - dBZ_{\theta_{(i,j,k)}}}{\theta_{(i,j,k+1)} - \theta_{(i,j,k)}}$
SPN (SPIN Change)	FR (Frequency of Reflectivity)
$SPN = \left(\frac{SPN \text{ counts}}{\text{All counts}} \right) \times 100$ ※ $ dBZ_i - dBZ_{i+1} \leq dBZ_{\text{thresh}}$	$FR = dBZ \text{ counts}$

그림 2. 추가 입력인자 연산
Fig. 2. Calculation of additional input data

3. 클러스터링 기반 방사형 기저함수 신경회로망

본 장에서는 다항식 방사형 기저 함수 신경회로망의 구조 및 은닉층을 구성하는 클러스터링 알고리즘과 다항식 연결가중치 파라미터를 동정하는 최소자승법에 대해 서술한다.

3.1 클러스터링 기반 방사형 기저 함수 신경회로망 구조

제안된 클러스터링 기반 방사형 기저함수 신경회로망(RBFNN)은 식(1)과 같이 If-then의 퍼지 규칙을 사용하였다.

$$R: \text{If } x_1, x_2 \dots \text{ and } x_n \text{ are } \overline{u_i}, \text{ then } y = f_i(x) \quad (1)$$

여기서 x 는 입력 데이터, $\overline{u_i}$ 는 클러스터링에 의한 $i=(1, \dots, c)$ 번째 그룹의 소속 함수(membership function), $f_{ji}(x)$ 는 $j=(1, \dots, s)$ 번째 출력에 대한 i 번째 퍼지 규칙의 다항식이다.[5]

RBFNN은 입력층, 은닉층, 출력층으로 구성된 기존의 신경회로망의 기본 틀을 두고 있지만 크게 기능적 구조로 그림 1과 같으며 입력 데이터를 클러스터링을 통해 데이터의 중심점과의 거리로 소속행렬을 구하는 조건부, 입력 데이터를 표 1과 같이 다항식으로 표현하는 결론부, 마지막으로 조건부의 소속행렬과 결론부에서 구한 다항식을 결합하여 최소자승법(LSE)으로 다항식의 계수를 구하여 최종 출력을 하는 추론부 이 세 가지로 나타낸다.[6][7][8]

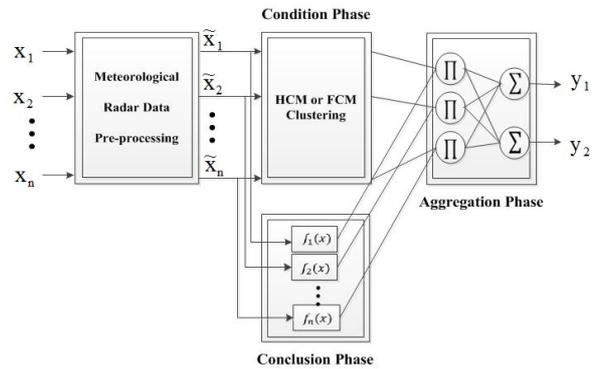


그림 3. 클러스터링 기반 RBFNNs의 구조
Fig. 3. Structure of clustering-based RBFNNs.

방사형 기저함수 신경회로망은 Robust 네트워크 특성을 갖으며 Multi Dimension의 문제를 해결할 수 있고 Predictive Ability가 우수한 특징을 갖는다.

표 1. 다항식 타입.
Table 1. Polynomial type.

Type	Polynomial Type
Constant	$y = a_0$
Linear	$y = a_0 + a_1x_1 + a_2x_2$

3.2 Hard C-mean 클러스터링 알고리즘

클러스터링은 데이터의 분류를 위해 사용되는 것으로 데이터에 형태나 속성 등의 기준으로 데이터를 분류하여 개체의 소속집단을 정하는 것이다. HCM 클러스터링은 각 데이터의 거리를 기준으로 소속정도를 측정하여 데이터를 분류하는 방법이다. 다음은 HCM 클러스터링의 과정이다.[9]

단계 1 : 클러스터의 개수를 결정하고, 소속행렬 $U(r)$ 을 초기화 한다.

$$U^{(r)} = u_{ij} \in [0, 1], \sum_{i=1}^c u_{ik} = 1, 0 < \sum_{k=1}^n u_{ik} < m \quad (2)$$

단계 2 : $U(r)$ 값을 기반으로 각각의 클러스터에 대한 중심 값과 소속함수의 값을 구한다.

$$v_{ij}^{(r)} = \frac{\sum_{k=1}^n u_{ik} \cdot x}{\sum_{k=1}^n u_{ik}}, \quad v_i^{(r)} = v_{i1}^{(r)}, \dots, v_{ic}^{(r)} \quad (3)$$

단계 3 : 각각의 클러스터 중심과 데이터와의 거리를 계산하여 새로운 소속행렬을 생성한다.

$$d_{ik} = d(x_k - v_i^{(r)}) = [\sum_{j=1}^l (x_{kj} - v_{ij}^{(r)})^2]^{1/2} \quad (4)$$

$$u_{ik}^{(r+1)} = \begin{cases} 1 & d_{ik}^{(r)} = \min_j d_{jk}^{(r)} \text{ for } all \ j \in c \\ 0 & otherwise \end{cases} \quad (5)$$

단계 4 : 오차가 허용범위 안에 도달하면 종료하고, 그렇지 않으면 단계 2로 돌아간다.

$$\|U^{(r+1)} - U^{(r)}\| = \epsilon \text{ (tolerance level)} \quad (6)$$

3.3 Fuzzy C-mean 클러스터링 알고리즘

FCM과 HCM 클러스터링 방법의 중요한 차이점은 FCM 클러스터링 방법은 0과1사이의 소속정도에 의해서 나타난 소속감의 정도를 가지고 주어진 데이터 점이 몇 개의 그룹에 속할 수 있다는 퍼지 분할을 사용한다는 점이다. 즉 퍼지 분할을 적용하기 위해서, 소속함수 U는 0과 1사이의 값을 가지는 요소들을 가진다.[10]

다음은 FCM 클러스터링의 과정이다.

단계 1 : 클러스터의 개수를 결정하고, 소속행렬 U(r)을 초기화 한다.

$$U^{(r)} = u_{ij} \in [0, 1], \sum_{i=1}^c u_{ik} = 1 \forall k, 0 < \sum_{k=1}^n u_{ik} < n \forall i \quad (7)$$

단계 2 : U(r)값을 기반으로 각각의 클러스터에 대한 중심 값과 소속함수의 값을 구한다.

$$v_{ij}^{(r)} = \frac{\sum_{k=1}^n u_{ik}^m \cdot x}{\sum_{k=1}^n u_{ik}^m}, \quad v_i^{(r)} = v_{i1}^{(r)}, \dots, v_{ic}^{(r)} \quad (8)$$

단계 3 : 각각의 클러스터 중심과 데이터와의 거리를 계산하여 새로운 소속행렬을 생성한다.

$$d_{ik} = d(x_k - v_i^{(r)}) = [\sum_{j=1}^l (x_{kj} - v_{ij}^{(r)})^2]^{1/2} \quad (9)$$

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{\|x_k - v_i^{(r)}\|}{\|x_k - v_j^{(r)}\|} \right)^{2/(m-1)}} \quad (10)$$

단계 4 : 오차가 허용범위 안에 도달하면 종료하고, 그렇지 않으면 단계 2로 돌아간다.

$$\Delta = \|U^{(r+1)} - U^{(r)}\| = \max_{i,k} |u_{ik}^{(r+1)} - u_{ik}^{(r)}| \quad (11)$$

FCM 클러스터링의 수는 은닉층 노드의 수를 대신하고 소속행렬의 값은 가우시안 활성함수에 의한 적합도 값과 같으며 FCM의 퍼지화 계수는 가우시안 함수의 분포상수를 나타낸다.

3.4 최소자승법

다항식 연결가중치의 파라미터 계수를 동정하기 위해 최소자승법을 사용하여 계수를 추정하였다. 최소자승법의 성능 평가함수를 행렬식으로 표현하면 다음 식 (12)과 같다.

$$Q = \sum_{i=1}^c (Y - X_i a_i)^T (Y - X_i a_i) \quad (12)$$

여기서 a_i 는 추정하고자 하는 j번째 다항식의 계수이고, Y는 출력데이터 이다. X_i 는 i번째 지역 모델의 계수를 추정하기 위한 입력 데이터 행렬을 의미하며 지역 모델의 선형식일 경우 다음 식 (8)처럼 정의된다.

$$X_i = \begin{bmatrix} 1 & x_{11} & \dots & x_{1m} \\ 1 & x_{21} & \dots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nm} \end{bmatrix} \quad (13)$$

식 (8)에서의 n은 전체 데이터의 수를 나타내며 I번째 규칙에 대한 지역 모델인 다항식 연결가중치의 계수는 식 (9)에 의해 구해진다.

$$a_i = (X_i^T X_i)^{-1} X_i^T Y \quad (14)$$

4. 에코 판단 모듈

기상레이더에 나타나는 비강수 에코 중에 가장 많이 나타나는 지형에코를 데이터로 사용하여 확실히 판단할 수 있는 부분을 사전에 판단함으로써 연산시간을 단축시키고, 성능을 향상시켰다. 판단 모듈의 판단 기준은 다음 표 2에 나타나있다.

표 2. 판단 기준.

Table 2. Criterion of judgement

		Representative ground echo data(DZ)	
		Null	DZ(dBZ)
Testing data(DZ)	Null	0 (Elimination)	0 (Elimination)
	DZ(dBZ)	Decision by classifier	Decision by classifier

5. 시뮬레이션 및 결과

본 논문에서는 제안된 클러스터링 기반 방사형 기저함수 신경회로망을 이용하여 기상레이더 패턴분류기를 설계하였다. 기상청에서 사용되고 있는 단일편파레이더인 오성산 레이더 데이터 자료를 입력데이터로 사용하였고, 테스트 데이터로 2012년에 발생했던 강수와 비강수사례를 선정하였다.

성능평가로는 패턴분류율을 이용하여 QC와의 유사도로 구하였다. 패턴분류율은 다음 표 3과 식 (2)와 같이 표현된다.

$$1 - \left(\frac{Error_count}{Total_data} \right) * 100 \quad (2)$$

표 3. 패턴분류율.
Table 3. Classification rate

Predictive Value	QC	
	Yes(1)	No(0)
Yes(1)	Correct	Error
No(0)	Error	Correct

설계된 패턴분류기의 파라미터 설정범위 및 선택된 파라미터 그리고 성능지수는 다음 표 4~ 6과 같다.

표 4. 설계된 패턴분류기의 파라미터.
Table 4. Parameters of designed pattern classifiers.

Clustering-based RBFNNs	HCM	FCM
No. of Inputs	5	
No. of Clusters	[2 4]	
Fuzzification Coefficient	-	[1.1 3]
Polynomial Type	Type1. Constant Type2. Linear	

표 5. 파라미터 변화에 따른 분류기의 성능지수
Table 5. Performance Index of classifier by selected parameters

HCM	Pattern Classification Rate		
	PI for Training	PI for Testing (PRE+GRE)	PI for Testing (Non-PR)
	85.16	84.32	77.25
FCM	PI for Training	PI for Testing (PRE+GRE)	PI for Testing (Non-PR)
	90.65	90.27	87.98
	91.24	91.04	88.43
Fuzzification Coefficient	2	91.35	89.23
	2.5	91.08	88.75
	3	90.51	88.21

표 6. 패턴분류기의 선택된 파라미터.
Table 6. Selected parameters of pattern classifiers.

Clustering-based RBFNNs	HCM	FCM
No. of Inputs	5	

No. of Clusters	4	
Fuzzification Coefficient	-	2
Polynomial Type	Type2. Linear	

선택된 파라미터에 의한 봄철강수, 여름철강수, 과량에코 +이상예코, 청천예코 사례를 테스트하였다. 각 사례에 대하여 1시간(7번) 분량을 테스트하였고, 총 4사례로 4시간(28번)분량을 테스트 하였다.

표 7. 봄철강수/지형예코 사례의 성능지수
Table 7. Performance index of PRE/GRE in spring

Date	PI for testing	
	HCM	FCM
2012.04.03 00:30	92.13	93.44
2012.04.03 00:40	92.90	93.80
2012.04.03 00:50	92.25	93.83
2012.04.03 01:00	91.31	94.8
2012.04.03 01:10	90.61	94.59
2012.04.03 01:20	90.49	94.39
2012.04.03 01:30	91.71	94.24

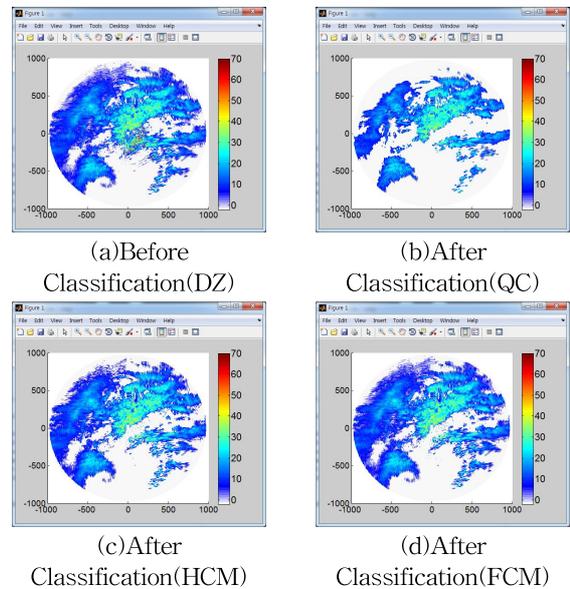


그림 4. 봄철강수/지형예코 사례 분류 결과
Fig. 4 Classification result of PRE/GRE in spring

표 8. 여름철강수/지형예코 사례의 성능지수
Table 8. Performance index of PRE/GRE in summer

Date	PI for testing	
	HCM	FCM
2012.08.10 07:00	90.11	92.41
2012.08.10 07:10	90.27	92.41
2012.08.10 07:20	90.34	92.10
2012.08.10 07:30	91.05	91.78
2012.08.10 07:40	90.73	91.68
2012.08.10 07:50	91.75	90.97
2012.08.10 08:00	92.10	90.91

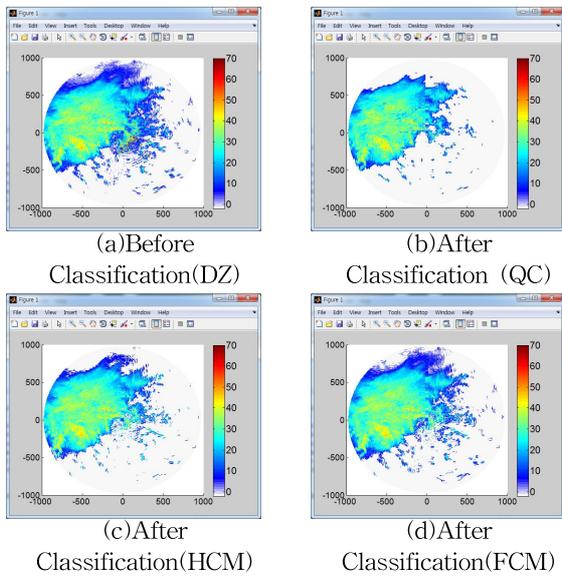


그림 5. 여름철강수/지형에코 사례 분류 결과
Fig. 5. Classification result of PRE/GRE in summer

표 9. 파랑에코+이상에코/지형에코 사례
Table 9. Performance index of sea-clutter+AP/GRE

Date	PI for testing	
	HCM	FCM
2012.05.21 00:00	99.99	99.99
2012.05.21 00:10	99.99	99.99
2012.05.21 00:20	99.99	99.99
2012.05.21 00:30	99.99	99.99
2012.05.21 00:40	99.97	99.99
2012.05.21 00:50	99.99	99.99
2012.05.21 01:00	99.98	99.99

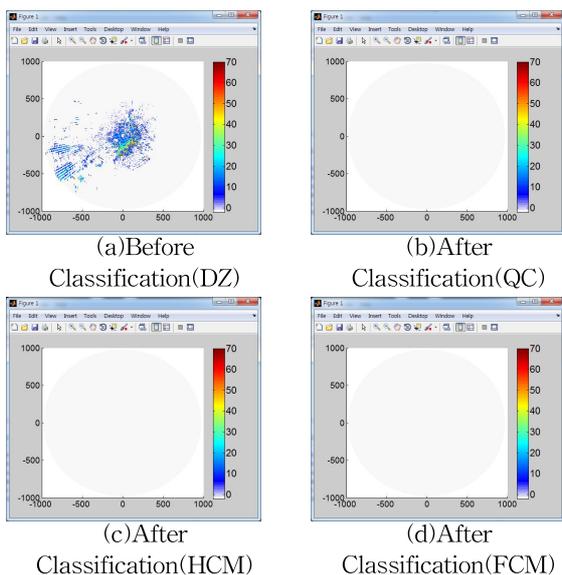


그림 6. 파랑에코+이상에코/지형에코 사례 분류 결과
Fig. 6. Classification result of sea-clutter+AP/GRE

표 10. 청천에코+지형에코 사례
Table 10. Performance index of clear echo+GRE

Date	PI for testing	
	HCM	FCM
2012.10.07 20:00	99.92	99.88
2012.10.07 20:10	99.89	99.85
2012.10.07 20:20	99.75	99.68
2012.10.07 20:30	99.85	99.80
2012.10.07 20:40	99.80	99.74
2012.10.07 20:50	99.60	99.47
2012.10.07 21:00	99.66	99.53

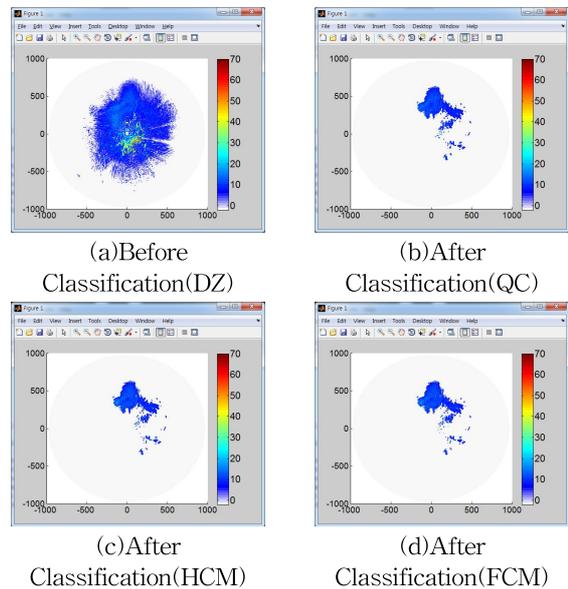


그림 7. 파랑에코+이상에코/지형에코 사례 분류 결과
Fig. 7. Classification result of sea-clutter+AP/GRE

6. 결론

기상레이더 데이터에는 강수에코와 비강수에코가 섞여 있는 것과 맑은 날씨임에도 불구하고 비강수에코가 나타나는 것이 있다. 본 논문에서는 강수에코와 비강수에코를 분류하여 비강수에코를 제거하고 지도상에 필요 에코만을 나타내는 것을 목적으로 하고, HCM과 FCM 클러스터링을 이용하여 논문에 더 적합한 클러스터링을 찾는다. 강수사례에서 QC 데이터는 약한 반사도 값들이 제거되는 문제를 갖고 있지만 제안된 클러스터링 기반 패턴분류기에서는 비강수에코만을 제거하는 것을 확인할 수 있었고, 비강수사례에서는 QC 데이터정도의 성능을 보임을 확인할 수 있었다. 그리고 강수에코와 비강수에코의 경계점이 애매한 레이더 데이터와 같은 경우 FCM 클러스터링을 이용하는 것이 더 적합함을 실험을 통해 알 수 있었다. 클러스터링 기반 패턴분류기에서 나온 정량적인 값은 기상청에서 현업으로 사용하고 있는 QC(Quality Control)와 비교하여 구하였다.

References

- [1] Cho, Y. H., G. Lee, K. E. Kim, and I. Zawadzki, "Identification and removal of ground echoes and anomalous propagation using the characteristics of radar echoes." *J. Atmos. Oceanic Technol.*, 23, 1206 - 1222 (2006).
- [2] Tanvir Islam, Miguel A. Rico-Ramirez, Dawei Han, Prashant K. Srivastava, "Artificial intelligence techniques for clutter identification with polarimetric radar signatures." *Atmospheric Research*, Volumes 109 - 110, June, Pages 95 - 113 (2012).
- [3] Walther, A., Schröder, M., Fischer, J., & Bennartz, R. "Comparison of precipitation in the regional climate model BALTIMOS to radar observations." *Theoretical and Applied Climatology*, 1-14. (2009).
- [4] Berenguer, M., Sempere-Torres, D., Corral, C., & Sánchez-Diezma, R. "A fuzzy logic technique for identifying nonprecipitating echoes in radar scans." *Journal of Atmospheric and Oceanic Technology*, 23(9), 1157-1180. (2006).
- [5] G. Pajares, M. Guijarro, A. Ribeiro, "A Hopfield Neural Network for combining classifiers applied to textured images." *Neural Networks*, Vol.23, pp. 144-153, (2010).
- [6] S. Abbasbandy, M. Otadi, M. Mosleh, "Numerical solution of a system of fuzzy polynomials by fuzzy neural network." *Information Sciences* 178 (8) 1948-1960. (2008)
- [7] S.-K. Oh, W. Pedrycz, B.-J. Park, "Self-organizing neurofuzzy networks in modeling software data," *Fuzzy Sets and Systems*, vol. 145, pp. 165-181, (2004).
- [8] L. Sanchez, I. Couso, J.A. Corrales, "Combining GP operators with SA search to evolve fuzzy rule based classifiers." *Information Science*, 136 175-191. (2001).
- [9] S. K. Oh, W. D. Kim, and W. Pedrycz, "Polynomial based radial basis function neural networks(P-RBF NNs) realized with the aid of particle swarm optimization," *Fuzzy Sets and Systems*, Vol. 163, No. 1, pp. 54-77, 2011.

- [10] W. Pedrycz, "Conditional fuzzy clustering in the design of radial basis function neural networks," *IEEE Trans. Neural Networks*, vol. 9, pp. 601 - 612, July. (1998).

저 자 소 개



최우용(Woo-Yong Choi)

2013년 : 수원대학교 전기공학과 공학사
2013년~현재 : 수원대학교 전기공학과 석사과정

관심분야 : 뉴럴 네트워크, Granular Computing, 진화론적 최적화 알고리즘, 퍼지 추론 시스템 등.

Phone : +82-31-222-6544

E-mail : choiwy@suwon.ac.kr



오성권(Sung-Kwon Oh)

1981년 : 연세대학교 전기공학과 공학사
1983년~1989년 : 금성산전연구소(선임연구원)

1993년 : 연세대학교 전기공학과 공학박사

1996년~1997년 : 캐나다 Manitoba 대학 전기 및 컴퓨터 공학과

Post-Doc.

1993년~2004년 : 원광대학교 전기전자 및 정보공학부 교수

2005년~현재 : 수원대학교 전기공학과 교수

2002년~현재 : 대한전기학회, 제어로봇시스템학회, 퍼지 및 지능시스템학회 편집위원

관심분야 : 퍼지 시스템, 퍼지-뉴럴 네트워크, 자동화 시스템, 고급 Computational Intelligence, 지능제어 등.

Phone : +82-31-229-8162

E-mail : ohsk@suwon.ac.kr