

데이터 정보를 이용한 흑색 플라스틱 분류기 설계

Design of Black Plastics Classifier Using Data Information

박 상 범* · 오 성 권*
(Sang-Beom Park · Sung-Kwun Oh)

Abstract - In this paper, with the aid of information which is included within data, preprocessing algorithm-based black plastic classifier is designed. The slope and area of spectrum obtained by using laser induced breakdown spectroscopy(LIBS) are analyzed for each material and its ensuing information is applied as the input data of the proposed classifier. The slope is represented by the rate of change of wavelength and intensity. Also, the area is calculated by the wavelength of the spectrum peak where the material property of chemical elements such as carbon and hydrogen appears. Using informations such as slope and area, input data of the proposed classifier is constructed. In the preprocessing part of the classifier, Principal Component Analysis(PCA) and fuzzy transform are used for dimensional reduction from high dimensional input variables to low dimensional input variables. Characteristic analysis of the materials as well as the processing speed of the classifier is improved. In the condition part, FCM clustering is applied and linear function is used as connection weight in the conclusion part. By means of Particle Swarm Optimization(PSO), parameters such as the number of clusters, fuzzification coefficient and the number of input variables are optimized. To demonstrate the superiority of classification performance, classification rate is compared by using WEKA 3.8 data mining software which contains various classifiers such as Naivebayes, SVM and Multilayer perceptron.

Key Words : Slope, Area, Spectrum peak, Chemical elements, Particle Swarm Optimization(PSO), Classifier

1. 서 론

분류기를 설계하는데 있어서 중요한 것은 분류기에 사용하는 데이터를 어떻게 구성하느냐 일 것이다. 획득한 raw data의 전체를 사용하거나 혹은 전체가 아닌 어느 특정 데이터만을 사용하는가에 따라 분류율은 다르게 나타날 수 있다. 부분방전의 경우, 부분방전의 해석을 Phase Resolved Partial Discharge Analysis (PRPDA)를 많이 사용하는데 왜도(skewness), 첨쇄도(kurtosis)와 같은 파라미터를 사용하거나 혹은 PRPD의 데이터 그대로 사용하는 방식에 따라 분류율은 달라진다[1]. 분류기의 분류율을 향상시키기 위해서는 알고리즘도 좋아야 하지만 알고리즘에 사용하는 데이터가 가지고 있는 특성을 잘 사용하여 이를 입력데이터로 사용하는 것도 하나의 방법이 될 수 있다. 그래서 본 연구에서는 흑색 플라스틱을 분류하기 위해 데이터가 가지고 있는 정보를 분석하여 이를 입력데이터로 사용한 전처리 알고리즘 기반 방사형 기저함수 신경회로망(Radial Basis Function Neural Networks: RBFNNs)분류기를 설계한다. 기존에는 전체 데이터를 가지고 주 성분 분석법을 사용하여 분류율을 확인하였다[2]. 하지만 본 논문에서는 Laser Induced Breakdown Spectroscopy(LIBS) 장비를

를 이용해 획득한 스펙트럼의 기울기, 면적과 같은 정보를 RBFNNs 분류기의 입력데이터로 사용한다. 기울기, 면적 방법을 이용해 raw data를 가공하고 분류기의 연산량을 줄이기 위해 전처리 알고리즘을 사용하여 입력변수를 축소시킨다. 입력변수가 축소된 입력데이터를 RBFNNs 분류기를 사용해 분류율을 확인한다. 또한 최적화 알고리즘을 사용하여 최적화한 분류율을 확인한다. 또한 WEKA 3.8 데이터 마이닝 소프트웨어를 사용해 본 연구에서 얻은 분류율과 비교한다.

2. 전처리 알고리즘을 사용한 입력변수의 축소

2장에서는 전처리 알고리즘으로 사용된 주성분 분석법과 fuzzy transform에 대해 설명하고,

2.1 주성분 분석법을 이용한 특징추출

주성분 분석법은 과원의 데이터를 저차원으로 축소시키는 차원 축소 알고리즘이다[3]. 축소시키는 방법은 먼저 축소시키고자 하는 차원의 개수를 정한다. 그리고 기존 데이터의 평균 벡터를 구하고 데이터와 평균벡터의 차를 구하여 공분산행렬을 계산한다. 공분산행렬의 고유벡터(Eigenvector)와 고유값(Eigenvalue)을 구하고 고유벡터에서 가장 큰 고유값을 축소시키고자하는 입력변수의 개수만큼 선택한다. 그리고 기존의 고차원 데이터와 축소하

† Corresponding Author : School of Electrical and Electronic Engineering, The University of Suwon, Korea.

E-mail : ohsk@suwon.ac.kr

* School of Electrical and Electronic Engineering, The University of Suwon, Korea.

Received : August 25, 2017; Accepted : March 27, 2018

고차하는 차원의 개수만큼 선택한 고유값과의 선형변환을 통해 축소된 데이터를 추출한다. 이와 같은 작업을 통해 축소된 입력 데이터는 분류기의 입력으로 사용된다. 고차원의 데이터를 사용하면 연산해야할 계산량이 많아지기 때문에 차원 축소 알고리즘을 전처리로 사용하여 프로그램의 연산량을 줄여준다. 연산량이 줄어든다는 것은 분류기의 처리속도가 향상된다는 것을 의미한다. 주성분 분석법의 알고리즘 순서는 다음과 같다[4].

[단계 1] 데이터의 벡터 집합(X)을 구성한다.

$$X = \{x_1, x_2, \dots, x_n\} \tag{1}$$

[단계 2] 벡터 집합의 평균벡터(M)를 계산한다.

$$M = \sum_{i=1}^n x_i \tag{2}$$

[단계 3] 벡터 집합과 평균벡터의 차 벡터(Φ)를 구한다.

$$\Phi_i = x_i - M \tag{3}$$

[단계 4] 차 벡터(Φ)의 공분산행렬(Cov)을 계산한다.

$$Cov = \frac{1}{n} \sum_{i=1}^n \Phi_i \Phi_i^T \tag{4}$$

[단계 5] 공분산행렬에서 고유값(Eigenvalue)과 고유벡터(Eigenvector)를 계산하고, 고유벡터에서 가장 큰 고유값을 가지는 고유벡터를 선정한다.

[단계 6] 선정된 고유벡터를 사용하여 특징벡터를 추출한다.

2.1 Fuzzy transform을 사용한 특징추출

Fuzzy transform도 주성분 분석법과 마찬가지로 차원 축소 알고리즘 중 하나이다. Fuzzy transform은 축소하고자하는 입력변

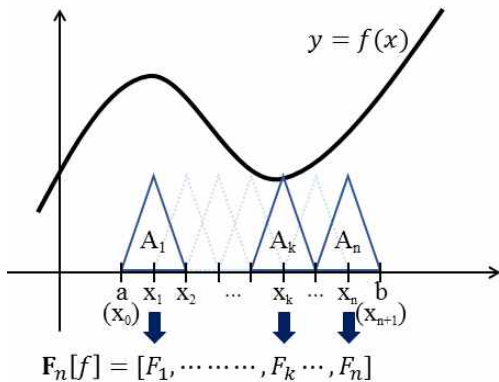


그림 1 Fuzzy transform을 사용한 입력변수 축소방법
 Fig. 1 Method of reduction of input variables using fuzzy transform

수의 개수만큼 퍼지공간의 개수를 결정하고 각각의 퍼지공간마다 대꽃값을 계산하여 축소한다[5]. Fuzzy transform에서 중요한 것은 멤버쉽함수의 형태이다. 멤버쉽함수의 형태는 삼각형, sin함수 형태 등 다양한 멤버쉽함수의 형태를 가지고 있고, 본 논문에서는 삼각형 모양의 멤버쉽 함수를 사용했다. 아래 그림 1에 Fuzzy transform을 사용해 입력변수를 축소하는 방법이다.

Fuzzy transform을 사용해 입력변수를 축소하는 방법과 같다[6].

[단계 1] 삼각형형태의 멤버쉽함수값을 계산한다.

$$A_1(x) = \begin{cases} 1 - \frac{(x-x_1)}{h_1}, & x \in [x_1, x_2] \\ 0 & otherwise \end{cases}$$

$$A_k(x) = \begin{cases} \frac{(x-x_{k-1})}{h_{k-1}}, & x \in [x_{k-1}, x_k] \\ 1 - \frac{(x-x_k)}{h_k}, & x \in [x_k, x_{k+1}] \\ 0 & otherwise \end{cases} \tag{5}$$

$$A_n(x) = \begin{cases} \frac{(x-x_{n-1})}{h_{n-1}}, & x \in [x_{n-1}, x_n] \\ 0 & otherwise \end{cases}$$

[단계 2] 식 (6)을 사용하여 F값을 구한다.

$$F_k = \frac{\sum_{j=1}^l f(p_j) A_k(p_j)}{\sum_{j=1}^l A_k(p_j)}, \quad k = 1, \dots, n \tag{6}$$

$$F_k = \frac{\int_a^b f(x) A_k(x) dx}{\int_a^b A_k(x) dx}, \quad k = 1, \dots, n \tag{7}$$

식 (6)은 함수 $y = f(x)$ 가 discrete인 경우, 식 (7)은 continuous한 경우에 사용한다. Fuzzy transform뿐만 아니라 Inverse fuzzy transform을 사용하기도 하는데, 이는 Fuzzy transform을 사용해 축소된 데이터를 복원시키기 위해 사용한다. 또한, Inverse fuzzy transform의 경우 데이터가 가지고 있는 노이즈를 제거할 때에도 사용한다[6]. 본 논문에서는 Fuzzy transform만 사용하여 입력변수 축소 알고리즘으로 사용하였다.

3. 데이터 정보기반 입력 데이터 구성

3장에서는 스펙트럼의 기울기, 기울기의 특정 peak, 화학원소의 면적계산을 이용하여 입력데이터를 어떻게 구성하는지에 대해 설명한다.

3.1 스펙트럼의 기울기를 이용한 입력데이터 구성

스펙트럼의 기울기는 아래와 같은 식을 사용하였다.

$$S = \frac{\Delta I}{\Delta W} \quad (8)$$

위의 식에서 S 는 기울기(slope), ΔI 는 강도(intensity)의 변화량, ΔW 는 파장(wavelength)의 변화량을 의미한다. 식 (8)을 적용하여 얻어진 스펙트럼의 기울기는 아래의 그림으로 제시하였다.

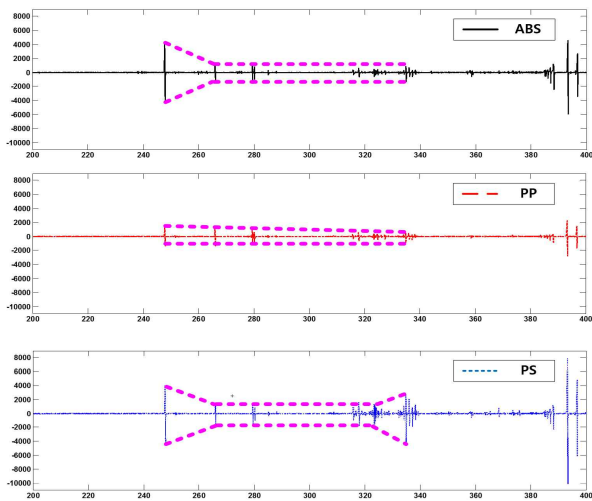


그림 2 각 재질별 스펙트럼의 기울기
Fig. 2 The slope of spectrum each material

위에 제시한 기울기 그래프에서 x축이 250~335nm 사이인 경우 ABS, PP, PS 재질별로 기울기의 개형이 다르게 나타나는 것을 확인할 수 있다. 이처럼 기울기의 개형이 다르게 나타나는 부분을 분류기의 입력 데이터로 사용하였다.

3.2 기울기의 peak만을 사용한 입력데이터 구성

3.1에서는 해당부분 전체를 사용했다면 이번에는 기울기 그래프에서 두드러지게 나타나는 peak들을 입력데이터로 선정하는 방법을 적용하였다. 본 연구에서 선택한 peak는 총 14개이고, 사용한 peak는 아래와 같다.

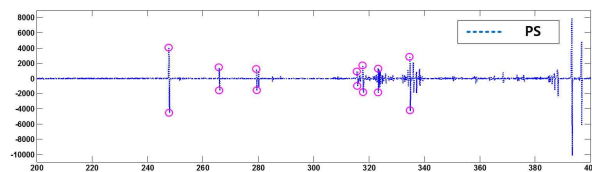
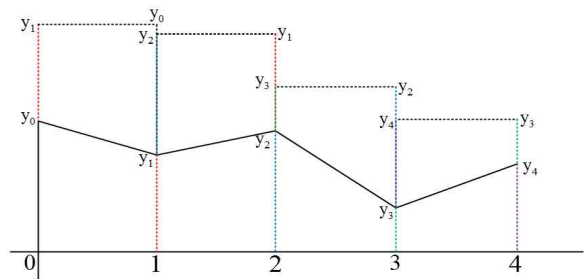


그림 3 분류기의 입력데이터로 사용한 peak 선정
Fig. 3 Selection of peak which is used as input data of classifier

전체 영역을 사용하는 것이 아니고 두드러지게 나타나는 peak만을 선정하여 분류기의 입력데이터로 사용한다. 이러한 방법을 사용하는 이유는 입력변수의 개수가 3.1에 비해 많이 줄어들게 되고 ABS, PP, PS 재질별로 그림 2와 같이 peak가 같은 x축상에서 나타나기 때문에 이와 같은 방법을 사용할 수 있다.

3.3 화학원소의 면적을 이용한 입력데이터 구성

세 번째 입력데이터 구성방법은 스펙트럼에서 나타나는 화학원소 peak를 기준으로 면적을 계산하여 면적값을 입력데이터로 구성하는 방법이다. 본 연구에서는 탄소, 수소, 질소, 산소원소가 나타나는 peak에 해당하는 부분의 면적을 계산하였다. 면적 계산하는 방법은 다음과 같다.



$$\begin{aligned} \text{Area} &= \frac{1}{2}(y_0 + y_1) + \frac{1}{2}(y_1 + y_2) + \frac{1}{2}(y_2 + y_3) + \frac{1}{2}(y_3 + y_4) \\ &= y_0 + y_1 + y_2 + y_3 + y_4 - \left(\frac{y_0}{2} + \frac{y_4}{2}\right) \end{aligned}$$

그림 4 면적 계산 방법 및 식
Fig. 4 Calculation method of the area and equation

면적 계산은 사다리꼴의 면적을 구하는 방법을 사용하였다. 사각형을 만들어 사각형의 면적을 구하고 구한 면적을 2로 나누어 주면 원하는 면적을 계산할 수 있다. 그림 4에 제시한 방법으로 각각의 원소에 해당하는 부분의 면적을 계산하였다.

3.3.1 탄소 peak

탄소 peak가 나타나는 부분은 3군데로 가장 먼저 245~250nm의 영역에서 나타난다.

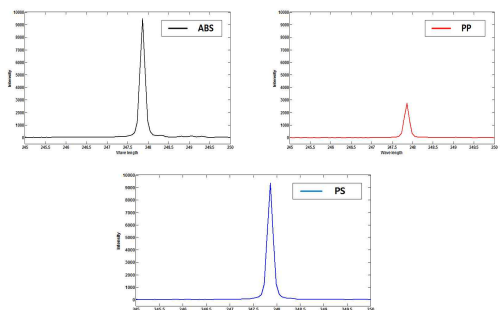


그림 5 245~250nm 파장의 탄소 peak
Fig. 5 Carbon peak of 245~250nm wavelength

약 247.7nm에서 탄소 peak가 나타나는 것을 확인할 수 있다. 두 번째로 탄소 peak가 나타나는 파장의 영역은 375~390nm이다.

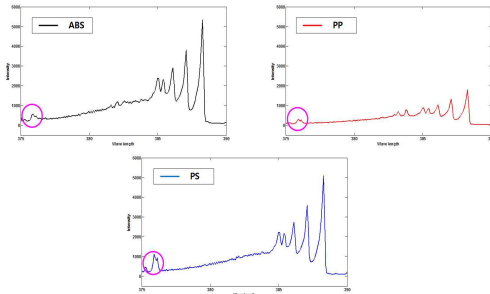


그림 6 375~390nm 파장의 탄소 peak
Fig. 6 Carbon peak of 375~390nm wavelength

탄소 peak는 385~390nm에서 나타나지만 약 376nm에서 재질 별로 차이를 보여주는 부분이 존재하여 동그랗게 표시한 부분의 넓이를 계산하였다. 마지막으로 탄소 peak가 나타나는 부분은 495~505nm이다.

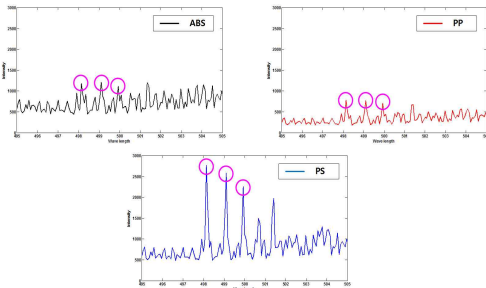


그림 7 495~505nm 파장의 탄소 peak
Fig. 7 Carbon peak of 495~505nm wavelength

해당 영역의 경우 498nm, 499nm, 500nm에서 나타난 peak에 해당하는 면적을 계산하였다.

3.3.2 수소 peak

수소 peak는 아래에 제시한 그림 8과 같이 650~665nm에서 나타난다.

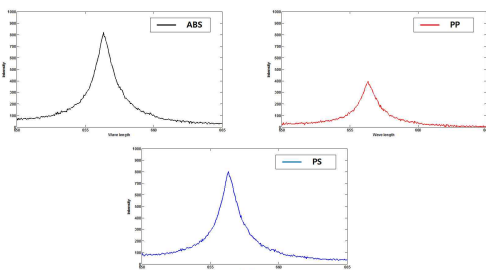


그림 8 650~665nm 파장의 수소 peak
Fig. 8 Hydrogen peak of 650~665nm wavelength

3.3.3 질소 peak

질소 peak는 740~750nm 영역에서 나타난다. 질소 peak는 3개의 peak가 나타나기 때문에 3개의 peak를 모두 사용하여 해당 부분의 면적을 계산하였다.

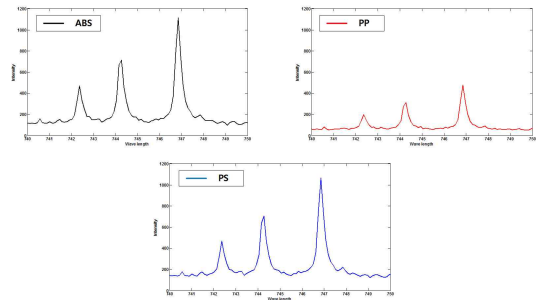


그림 9 740~750nm 파장의 탄소 peak
Fig. 9 Nitrogen peak of 740~750nm wavelength

3.3.4 산소 peak

산소 peak는 775~780nm에서 나타난다.

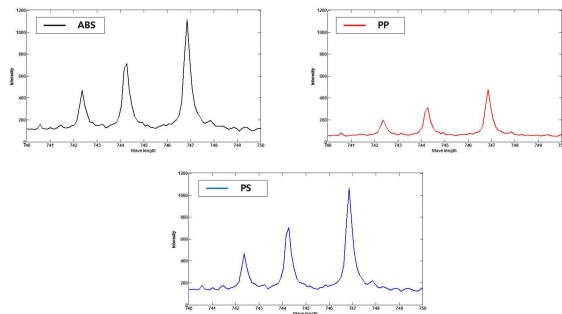


그림 10 775~780nm 파장의 탄소 peak
Fig. 10 Oxygen peak of 775~780nm wavelength

4. 전처리 알고리즘 기반 방사형 기저함수 신경회로망과 입자군집 최적화

4장에서 전처리 알고리즘을 사용한 방사형 기저함수 신경회로망 분류기의 구조와 데이터 그룹화 방법 중 하나인 FCM 클러스터링에 대해 기술한다. 그리고 파라미터 최적화 알고리즘 중 하나인 입자군집 최적화에 대해 설명한다.

4.1 전처리 알고리즘 기반 방사형 기저함수 신경회로망의 구조

위에 제시한 그림 11은 본 논문에서 사용한 분류기의 구조를 나타낸다. 제안된 분류기는 총 4개의 부분으로 구성되어있다. 전처리 부분에는 입력변수 축소 알고리즘을 사용하였고 각각 구성

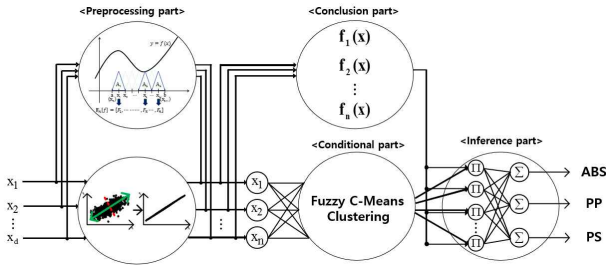


그림 11 전처리 알고리즘 기반 방사형 기저함수 신경회로망의 구조

Fig. 11 Structure of RBFNNs based on preprocessing algorithm

본 분석법, Fuzzy transform을 사용하였다. 조건부 부분에는 Fuzzy C-Means(FCM) 클러스터링을 사용하여 적합도를 구하고 결론부에는 연결가중치를 1차 형식을 사용하였다. 그리고 추론부에는 퍼지추론법을 사용하여 분류기의 출력을 계산한다.

4.1.1 Fuzzy C-Means(FCM) 클러스터링

FCM 클러스터링은 데이터의 소속정도를 0과 1사이의 값으로 나타내는 방법이다. 방사형 기저함수 신경회로망의 조건부에 가우시안 함수대신 FCM 클러스터링을 사용한 이유는 FCM 클러스터링을 사용함으로써 데이터가 가지고 있는 특성을 좀 더 반영하기 위함이다[7].

5. 시뮬레이션 및 결과고찰

전처리 과정을 거치기 전에, LIBS 장비를 이용해 획득한 original 스펙트럼을 가공하여 새로운 데이터를 획득하였다. 가공한 방법은 3가지로 각각 스펙트럼의 기울기를 이용하여 입력데이터를 획득하는 방법, 스펙트럼의 기울기에서 특정 peak만을 입력데이터로 사용하는 방법, 그리고 화학원소가 나타나는 peakmf 기준으로 면적을 계산하여 입력데이터로 사용하는 방법이다. 본 연구에서 사용한 흑색 플라스틱 재질은 ABS, PP, PS 재질을 사용하였고 재질별로 400개의 시료를 획득하여 스펙트럼을 얻었다.

5.1 입력데이터 구성에 대한 실험 설정조건 및 실험결과

3.1, 3.2, 3.3절에 제시한 입력데이터 구성방법을 사용하여 획득한 데이터의 입력변수 개수를 아래의 표에 제시하였다.

표 1 입력데이터의 구성방법에 따른 입력변수의 개수

Table 1 Number of input variables according to method of construction of input data

Method	Slope	Peak	Area
No. of input variables	1,445	14	10

Method : Method of construction of input data
 Slope : Slope of the spectrum
 Peak : Selection of peak from slope of spectrum
 Area : Area of spectrum from chemical elements

original 스펙트럼의 입력변수의 개수는 10,240개이다.

스펙트럼의 기울기를 사용하여 획득한 입력데이터의 입력변수 개수는 1,445개이고, 기울기에서 두드러지게 나타나는 peak들을 입력데이터로 획득했을 경우의 입력변수의 개수는 14개이다. 그리고 화학원소 peak를 기준으로 스펙트럼의 면적을 계산하여 얻은 입력데이터의 입력변수 개수는 10개이다. 아래의 표 2는 실험을 진행하기 위해 설정한 실험조건들을 나열하였다.

표 2 실험 조건

Table 2 Experimental condition

Parameters	Values
Number of samples	1,200
Number of samples per class	400
Ratio of training data and test data	4 : 1
Number of fuzzy clusters	3, 4, 5
Fuzzification coefficient(FC)	2.0
Polynomial type	Linear

본 연구에서 사용하는 데이터의 개수는 총 1,200개로 ABS, PP, PS 각각 400개의 샘플을 수집하였다. 학습데이터와 테스트데이터는 4 : 1의 비율로 분할하여 5-fold cross validation을 사용하여 분류성능을 확인하였다. 그리고 FCM 클러스터링의 클러스터 개수는 3~5개로 설정하였다. 퍼지화계수는 2.0으로 고정하였고 연결가중치의 형태는 1차 선형식을 사용하였다. 스펙트럼의 기울기를 사용하여 획득한 데이터의 경우 입력변수의 개수가 많아 주성분 분석법과 선형판별 분석법을 사용하여 입력변수의 개수를 축소시켰다. 반면에 peak선택, 면적값을 계산하여 얻은 데이터는 입력변수의 개수가 많지 않아 전처리 알고리즘을 사용하지 않았다. 먼저 스펙트럼의 기울기를 이용하여 얻은 데이터의 분류성능은 다음과 같다.

표 3 입력데이터를 스펙트럼의 기울기로 사용한 경우의 RBFNNs 분류기의 분류율

Table 3 Classification rate of RBFNNs classifier in case of using input data as the slope of spectrum

(a) In case of PCA

Preprocessing	D	C	Training	Test
Principal Component Analysis	30	3	95.12±0.44	93.33±2.04
		4	96.15±0.69	94.08±1.75
		5	96.25±0.58	93.58±2.40
	40	3	95.85±0.36	93.50±2.07
		4	97.00±0.14	94.00±1.97
		5	97.19±0.33	93.58±2.12
	50	3	96.38±0.27	93.67±2.07
		4	97.25±0.30	93.50±1.88
		5	97.44±0.24	93.42±2.19

(b) In case of fuzzy transform

Preprocessing	D	C	Training	Test
Fuzzy transform	30	3	95.10±0.33	93.58±1.40
		4	96.16±0.32	94.25±1.82
		5	96.50±0.28	94.33±1.33
	40	3	95.46±0.50	93.66±1.51
		4	95.93±0.51	93.83±1.80
		5	96.79±0.30	94.50±2.17
	50	3	95.04±0.45	93.75±1.97
		4	95.93±0.54	93.08±1.40
		5	97.00±0.13	92.83±1.65

Preprocessing : Preprocessing algorithm which is used in RBFNNs classifier

D : Number of reduced dimensions using preprocessing algorithm

MFs : Number of membership functions

C : Number of fuzzy clusters

Training : Classification rate of training data and standard deviation

Test : Classification rate of test data and standard deviation

스펙트럼의 기울기를 입력변수로 사용한 경우, 전처리 알고리즘으로 주성분 분석법을 사용했을 때는 평균적으로 93%의 분류율을 확인하였다. Fuzzy transform을 사용했을 때는 멤버쉽함수의 개수가 30개, 35개일 때 주성분 분석법보다 분류율이 조금 높은 것을 확인하였다. 나머지는 주성분 분석법과 비슷한 분류율의 결과를 얻었다. 그리고 스펙트럼의 기울기에서 두드러지게 나타나는 peak를 선별하여 분류기의 입력데이터로 사용한 경우의 분류율은 아래의 표 4의 결과를 얻을 수 있었다.

표 4 입력데이터를 peak로 사용한 경우의 RBFNNs 분류기의 분류율

Table 4 Classification rate of RBFNNs classifier in case of using input data as the peak

(a) In case of PCA

Preprocessing	D	C	Training	Test
Principal Component Analysis	5	3	91.21±0.64	90.67±2.18
		4	92.04±0.46	91.42±1.83
		5	92.90±0.49	91.92±1.24
	8	3	91.33±0.71	90.58±2.71
		4	91.98±0.69	91.08±2.73
		5	93.08±0.58	91.42±2.29
	10	3	91.48±0.60	90.42±2.52
		4	92.77±0.74	90.83±2.38
		5	93.42±0.64	91.42±2.24
	-	3	91.96±0.59	90.42±2.55
		4	92.92±0.41	91.08±2.01
		5	93.85±0.45	91.50±1.68

(b) In case of fuzzy transform

Preprocessing	D	C	Training	Test
Fuzzy transform	5	3	83.75±0.43	83.00±3.40
		4	83.93±0.64	82.16±4.49
		5	84.98±0.68	82.50±3.45
	8	3	89.83±0.65	87.75±1.96
		4	91.39±0.44	89.08±1.65
		5	92.06±0.73	90.16±2.21
	10	3	89.96±0.38	88.00±1.98
		4	90.68±0.49	88.08±2.07
		5	92.18±0.54	89.58±2.08
	-	3	91.96±0.59	90.42±2.55
		4	92.92±0.41	91.08±2.01
		5	93.85±0.45	91.50±1.68

입력데이터를 기울기의 peak로 사용한 경우, 전처리 알고리즘으로 주성분 분석법을 사용했을 때는 분류율이 약 91%로 Fuzzy transform을 사용했을 때보다 분류율이 높게 나오는 것을 확인하였다. 표 3에 제시한 분류율과 비교하면 2%낮게 나온 것을 알 수 있다. 입력데이터를 스펙트럼의 면적으로 사용한 경우의 분류율은 다음과 같다.

표 5 입력데이터를 스펙트럼의 면적으로 사용한 경우의 RBFNNs 분류기의 분류율

Table 5 Classification rate of RBFNNs classifier in case of using input data as the area of spectrum

(a) In case of PCA

Preprocessing	D	C	Training	Test
Principal Component Analysis	5	3	89.19±0.58	88.17±1.97
		4	89.29±0.77	88.50±2.01
		5	90.17±0.72	89.00±1.20
	8	3	89.85±0.90	88.92±3.41
		4	90.35±0.74	88.50±1.92
		5	90.37±0.84	89.33±1.73
	-	3	89.96±0.79	88.83±3.72
		4	90.31±0.56	88.83±1.90
		5	90.60±0.77	88.92±1.92

(b) In case of fuzzy transform

Preprocessing	D	C	Training	Test
Fuzzy transform	5	3	86.43±0.51	85.58±2.25
		4	87.31±0.61	86.33±1.23
		5	87.83±0.62	86.91±1.33
	8	3	89.31±1.06	88.41±3.41
		4	89.71±0.80	88.75±2.06
		5	90.02±0.83	88.58±2.72
	-	3	89.96±0.79	88.83±3.72
		4	90.31±0.56	88.83±1.90
		5	90.60±0.77	88.92±1.92

스펙트럼의 면적값을 입력데이터로 사용한 경우, 88%의 분류율을 나타냈고 표 3과 표 4에 제시한 분류율에 비해 조금 떨어지는 것을 확인하였다.

5.2 최적화 알고리즘을 사용한 실험 설정조건 및 실험결과

입자군집 최적화를 사용하여 최적화를 진행하였고, 최적화하기 위해 설정한 파라미터들은 아래 표 6에 제시하였다.

표 6 최적화를 위한 파라미터 설정 조건

Table 6 Setting condition of parameters for optimization

Parameters		Values		
Ratio of data (Training : Validation : Test)		5 : 3 : 2		
Number of generations		50		
Swarm size		100		
[W_{min} W_{max}]		[0.4 0.9]		
V_{max}		20%		
c_1, c_2		2.0		
r_1, r_2		[0 1]		
Search space	Number of reduced dimensions	Slope	Peak	Area
		[10 50]	[3 10]	[3 8]
	Number of fuzzy clusters	[2 7]		
	Fuzzification coefficient(FC)	[1.1 3.0]		
Polynomial type	Linear, Modified Quadratic, Quadratic			

입자군집 최적화를 사용해 최적화하고자 하는 파라미터의 개수는 총 4가지로 각각 축소된 입력변수의 개수, FCM 클러스터링의 클러스터 개수, 퍼지화계수, 연결가중치의 형태이다. 입자군집 최적화에 사용한 목적함수는 다음과 같다.

$$\text{Objective function} = (CR_{TR} + CR_{VA}) * 0.5 \quad (9)$$

$$\text{Objective function} = (MSE_{TR} + MSE_{VA}) * 0.5 \quad (10)$$

목적함수는 2가지를 사용하였다. 식 (17)에 제시한 분류율을 목적함수로 사용하는 방법과 식 (18)에 나타난 성능지수를 목적함수로 사용하는 방법이다. 분류율을 목적함수로 사용한 경우 학습데이터의 분류율과 검증데이터의 분류율을 더한 후 2로 나눈 값을 사용하였다. 그리고 각 particle마다 계산된 분류율값을 비교하여 더 큰 값에 해당하는 particle를 선정한다. 그리고 성능지수 (Performance Index, PI)를 목적함수로 사용한 경우 학습데이터의 성능지수와 검증데이터의 성능지수를 더하여 2로 나눈 값을 목적함수로 사용한다. 그리고 particle마다 성능지수를 계산하여 성능지수가 낮은 값에 해당하는 particle를 선정한다. 본 연구에서 사용한 성능지수 평가방법은 평균제곱오차(Mean Squared Error, MSE)를 사용하여 평가하였다. MSE 계산식은 다음과 같다.

$$MSE = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2 \quad (11)$$

식 (11)에서 \hat{y}_i 는 모델의 출력, y_i 는 original data의 출력을 의미한다. 즉, 모델의 출력과 original data의 출력을 빼서 제곱하

고 데이터의 개수만큼 모두 더한 후, 데이터의 개수로 나누어 얻은 결과가 성능지수가 된다. 기울기, peak, 면적을 각각 최적화 알고리즘의 입력데이터로 사용하여 얻은 결과는 아래의 표 7과 같다.

표 7 목적함수에 따른 입자군집 최적화를 사용한 분류율 결과

Table 7 The result of classification rate using particle swarm optimization according to the objective function

(a) In case of classification rate

Data type	Slope	Peak	Area
Dimension	5	5	6
Clusters	4	5	5
FC	2.12	1.92	2.35
Polynomial type	Linear	Linear	Linear
Training	96.42±0.81	92.67±1.04	90.36±0.74
Validation	95.14±1.15	91.94±0.81	87.66±2.25
Test	94.82±2.11	92.25±1.83	89.83±2.23
MSE_{TR}	0.121±0.007	0.206±0.017	0.267±0.005
MSE_{VA}	0.714±0.521	0.828±0.931	0.316±0.046
MSE_{TE}	0.236±0.248	0.396±0.204	0.335±0.030

(b) In case of performance index

Data type	Slope	Peak	Area
Dimension	6	7	6
Clusters	5	6	4
FC	1.87	2.16	2.13
Polynomial type	Linear	Linear	Modified Quadratic
Training	97.70±0.68	93.26±0.34	93.16±0.65
Validation	93.11±1.47	93.94±0.91	89.94±1.78
Test	93.33±1.74	90.66±2.13	89.25±1.23
MSE_{TR}	0.097±0.009	0.198±0.009	0.194±0.008
MSE_{VA}	0.213±0.024	0.209±0.029	0.273±0.038
MSE_{TE}	0.269±0.073	0.286±0.053	0.399±0.082

MSE_{TR} : Performance index of training data

MSE_{VA} : Performance index of validation data

MSE_{TE} : Performance index of test data

목적함수를 식 (9)에 제시한 분류율로 사용한 경우 표 3, 표 4, 표 5에 제시한 분류율보다 개선되는 것을 확인하였다. 하지만 입력데이터를 스펙트럼의 기울기로 사용한 경우, 최적화된 분류율이 peak, 면적을 입력데이터로 사용했을 때보다 크게 개선되는 결과를 얻지 못하였다. 목적함수를 식 (10)에 제시한 성능지수로 사용한 경우, 목적함수를 분류율로 사용했을 때보다 최적화된 테스트 분류율이 전체적으로 낮게 나타나는 것을 확인하였다. 그리고 성능지도 분류율에 비해 더 낮게 나타나는 것을 확인하였다.

5.3 WEKA 3.8 data mining software를 사용한 분류율 비교

WEKA 3.8 data mining software는 뉴질랜드의 Waikato 대학에서 개발한 소프트웨어로 분류뿐만 아니라 회귀, 전처리 등과 같은 다양한 알고리즘을 포함하고 있다[10]. 스펙트럼의 기울기, 기울기에서 선별된 peak, 스펙트럼의 면적을 똑같이 WEKA 3.8

data mining software의 입력데이터로 사용하여 본 논문에서 제시한 분류율과 비교한다. 입력데이터 구성 방법에 대한 WEKA software와 RBFNN 분류기의 분류율을 비교한 결과를 아래의 표 8에 제시하였다.

표 8 WEKA software와 RBFNN 분류기의 분류율 비교
Table 8 Comparison of classification rate between WEKA software and the proposed classifier

(a) In case of the slope of spectrum as input data

Model of classifier		Training	Test
Naivebayes		N/A	72.33±3.13
Multilayer perceptron		N/A	93.08±2.66
SMO		N/A	88.58±1.20
IBK		N/A	87.25±4.83
PART		N/A	87.08±1.56
J4.8		N/A	90.17±1.68
Random forest		N/A	93.75±1.61
Random tree		N/A	82.92±2.78
Proposed classifier	PCA-based RBFNNs	96.15±0.69	94.08±1.75
	Fuzzy transform-based RBFNNs	96.50±0.28	94.33±1.33

(b) In case of the peak of slope as input data

Model of classifier		Training	Test
Naivebayes		N/A	78.08±2.07
Multilayer perceptron		N/A	90.50±1.92
SMO		N/A	89.75±2.10
IBK		N/A	90.58±1.80
PART		N/A	89.75±1.78
J4.8		N/A	89.08±3.19
Random forest		N/A	91.33±1.86
Random tree		N/A	90.00±0.72
Proposed classifier		93.85±0.45	91.50±1.68

(c) In case of the area of spectrum as input data

Model of classifier		Training	Test
Naivebayes		N/A	82.08±2.30
Multilayer perceptron		N/A	89.17±2.15
SMO		N/A	87.58±2.09
IBK		N/A	84.58±3.49
PART		N/A	86.67±0.72
J4.8		N/A	87.58±1.12
Random forest		N/A	88.25±0.78
Random tree		N/A	84.08±2.11
Proposed classifier		90.60±0.77	88.92±1.92

표 8의 (a)는 제시한 분류율의 경우 입력변수의 개수를 30개로 축소시킨 후 얻은 분류율을 비교하였다. WEKA 3.8 data mining software의 분류기를 사용했을 때보다 본 논문에서 사용한 RBFNNs 분류기의 분류율이 더 높은 것을 확인하였다. Multilayer perceptron의 경우 분류율이 93.08%로 본 논문에서 사용한 분류기보다 1%정도 낮은 것을 확인할 수 있는데 1%의 분류율은 큰 의미가 없을 것이라고 생각할 수 있다. 하지만 상업화를 목적으로 생각한다면 1%의 분류율도 크게 나타날 수 있기

때문에 1%의 분류율도 의미를 가지고 있는 결과라고 볼 수 있다. 표 8의 (b)의 경우에는 표 8의 (a)의 결과와 비슷하게 본 논문에서 사용한 분류기의 분류율이 더 높게 나타나는 것을 확인하였다. 다만 Random forest 분류기를 사용한 경우에는 91.33%로 RBFNNs 분류기의 분류율과 거의 비슷한 분류율을 보여주었다. 표 8의 (c)도 비슷한 결과를 나타내었다. Multilayer perceptron과 Random forest가 RBFNNs 분류기와 거의 비슷하게 나타나는 것을 확인하였으며 그 외 나머지 분류기의 분류율은 제안된 분류기의 분류율보다 낮게 나타나는 경향을 보여주었다.

6. 결 론

본 논문에서는 레이저유도봉고분광법을 이용하여 획득한 스펙트럼을 스펙트럼의 기울기, 기울기의 peak선별, 탄소, 수소와 같은 화학원소의 면적값을 이용하여 스펙트럼 데이터를 가공 및 획득하였다. 그리고 획득한 데이터를 전처리 알고리즘 기반 방사형 기저함수 신경회로망 패턴분류기의 입력데이터로 사용하여 분류율을 확인하였다. 또한, 입자군집 최적화 알고리즘을 사용하여 FCM 클러스터링의 클러스터 개수, 퍼지화계수, 입력변수의 개수, 연결가중치의 형태를 최적화하여 분류율을 비교하였다. 그리고 WEKA 3.8 data mining software에 존재하는 다양한 분류기를 사용하여 본 연구에서 사용한 RBFNNs 분류기와 분류율을 비교하여 제안된 분류기의 우수성을 입증하였다. 향후에는 선형판별 분석법(Linear Discriminant Analysis, LDA), fusion PCA&LDA와 같은 전처리 알고리즘을 다양하게 사용하여 분류율을 확인하고 RBFNNs 패턴분류기뿐만 아니라 퍼지 집합, 퍼지추론기반 신경회로망(Fuzzy Neural Network, FNN) 분류기와 LDA 분류기를 사용하여 분류율을 비교하고자 한다.

감사의 글

This study was supported by the R&D Center for Valuable Recycling(Global-Top R&D Program) of Ministry of Environment(Project No.: 2016002250002)

References

- [1] G.-S. Kim, L.-S. Ahn and S.-K. Oh, "The Design of Optimized Type-2 Fuzzy Neural Networks and Its Application", The Transaction of KIEE, Vol. 58, No. 8, pp. 1615-1623, 2009.
- [2] S.-B. Park, J.-S. Bae, S.-K. Oh and H.-K. Kim, "Design of Pattern Classifier for Electrical and Electronic Waste Plastic Devices Using LIBS Spectrometer", Journal of Korean Institute of Intelligent Systems, Vol. 26, No. 6, pp. 477-484, 2016.

[3] Shengji Tee, Q. Liu and Zhongdong Wang, "Insulation condition ranking of transformers through principal component analysis and analytic hierarchy precess", IET Generation, Transmission and Distribution, Vol. 11, No. 1, pp. 110-117, 2017.

[4] S.-H. Wang, T.-M. Zhan, Y. Chen, Y. Zhang, M. Yang, H.-M. Lu, H.-N. Wang, B. Liu and P. Philips, "Multiple Sclerosis Detection Based on Biothogonal Wavelet Transform, RBF Kernel Principal Component Analysis, and Logistic Regression", IEEE Access special section on advanced signal processing methods in medical imaging, Vol. 4, pp. 7561-7576, 2016.

[5] V. Loia, S. Tomasiello and A. Vaccaro, "Fuzzy Transform Based Compression of Electrical Signal Waveforms for Smart Grids", IEEE Transactions on Systems, Man, and Cybernetics: Systems, Vol. 47, No. 1, pp. 121-132, 2017.

[6] I. Perfilieva, "Fuzzy transform: Theory and application", Fuzzy sets and systems, Vol. 157, No. 8, pp. 993-1023, 2006.

[7] J.-S. Bae, S.-K. Oh and H.-K. Kim, "Design of Fingerprints Identification Based on RBFNN Using Image Processing Techniques", The Transaction of the KIEE, Vol. 65, No. 6, pp. 1060-1069, 2016.

[8] M.-R. Bonyadi and Z. Michalewicz, "Impacts of Coefficients on Movement Patterns in the Particle Swarm Optimization Algorithm", IEEE Transactions on Evolutionary Computation, Vol. 21, No. 3, pp. 378-390, 2017.

[9] S.-H. Oh, S.-K. Oh and J.-Y. Kim, "Design of Face Recognition and Tracking System by Using RBFNNs Pattern Classifier with Object Tracking Algorithm", The Transactions of the KIEE, Vol. 64, No. 5, pp. 766-778, 2015.

[10] I. Zliobaite, A. Bifet, J. Read, B.Pfahringner and G. Holmes, "Evaluation method and decision theory for classification of streaming data with temporal dependence", Machine Learning, Vol. 98, No. 3, pp. 455-482, 2015.

저 자 소 개



박 상 범(Sang-Beom Park)

2016년 : 수원대학교 전기공학과 졸업
 2016년~현재 : 동 대학원 석사과정
 관심분야 : Fuzzy Inference System, Neural Networks, Evolutionary optimization, Granular computing etc.
 Phone : +82-31-222-6544
 E-mail : sangbeom91@suwon.ac.kr



오 성 권 (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 전기전자공학부 교수
 2002년~현재 : 대한전기학회 및 한국지능시스템학회 편집위원
 2013년~현재 : Information Sciences 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