다중 목적 입자 군집 최적화 알고리즘 이용한 방사형 기저 함수 기반 다항식 신경회로망 구조 설계

논 문
61-1-23

Structural Design of Radial Basis Function-based Polynomial Neural Networks by Using Multiobjective Particle Swarm Optimization

김 옥 동* · 오 성 권†
(Wook-Dong Kim and Sung-Kwun Oh)

Abstract - In this paper, we proposed a new architecture called radial basis function-based polynomial neural networks classifier that consists of heterogeneous neural networks such as radial basis function neural networks and polynomial neural networks. The underlying architecture of the proposed model equals to polynomial neural networks(PNNs) while polynomial neurons in PNNs are composed of Fuzzy-c means-based radial basis function neural networks(FCM-based RBFNNs) instead of the conventional polynomial function. We consider PNNs to find the optimal local models and use RBFNNs to cover the high dimensionality problems. Also, in the hidden layer of RBFNNs, FCM algorithm is used to produce some clusters based on the similarity of given dataset. The proposed model depends on some parameters such as the number of input variables in PNNs, the number of clusters and fuzzification coefficient in FCM and polynomial type in RBFNNs. A multiobjective particle swarm optimization using crowding distance (MoPSO-CD) is exploited in order to carry out both structural and parametric optimization of the proposed networks. MoPSO is introduced for not only the performance of model but also complexity and interpretability. The usefulness of the proposed model as a classifier is evaluated with the aid of some benchmark datasets such as iris and liver.

Key Words : Polynomial neural networks, Radial basis function neural networks, Fuzzy C-means multiobjective particle swarm optimization, Pattern classifier

1. 서 론

최근 컴퓨터 지능(Computational Intelligence; CI) 분야에 서 패턴분류 및 인식에 대한 연구가 주목을 받고 있으며, 패턴분류를 응용하여 사물 및 얼굴인식을 바탕으로 보안시스템 및 상황인지 시스템을 개발하기 위해 다각도로 연구가 진행되고 있다. 컴퓨터 지능분야에서 패턴분류를 위한 알고리즘은 신경회로망이 대표적이다. 신경회로망의 장점은 다차원 입력 및 출력력을 갖는 패턴분류의 특성을 가장 잘 표현할 수 있으며, 유연한 모델 구조를 설계하여 데이터의 특징이나 적용분야에 적합한 최적의 모델을 구축할 수 있다 [1-2].

신경회로망을 바탕으로 다양한 형태의 신경회로망 구조가 제안되었으며, 그중 가장 대표적인 모델은 방사형 기저 함수 신경회로망이 있다. 방사형 기저 함수 신경회로망(radial basis function neural networks)[3]은 신경회로망과 동일하게 입력층, 은닉층, 출력층과 같이 세 개의 층으로 이루어져 있다. 방사형 기저 함수 신경회로망은 역전 오류와 이외에 최소자승법을 사용할 수 있다. 최소자승법을 사용하면 학습률이나 모멘텀 계수 같은 제어 파라미터가 줄어드는 장점이

있다.

방사형 기저 함수 신경회로망 이외에 신경회로망의 네트워크를 이용한 다항식 신경회로망이 제안되었다[4]. 다항식 신경회로망은 신경회로망의 다층구조를 유지하며, 노드에 활성함수 대신 1차이상의 다항식으로 구성된다. 각 노드가 다항식으로 표현되기 때문에 노드를 연결하는 연결가중치가 불필요하다. 다항식 신경회로망은 데이터의 특성에 따라 다양한 형태의 모델 구조가 생성되며 모델을 설계하는데 필요한 제어 파라미터가 비교적 적게 사용되어 기존 모델보다 학습방법이 간단하며 다른 모델과 결합이 용이하다.

본 연구에서는 다항식 신경회로망과 방사형 기저 함수를 결합한 구조를 바탕으로 패턴분류기를 설계한다. 제안된 모델은 다항식 신경회로망을 기본 구조로 하고 기존의 다항식 노드 대신 방사형 기저 함수 신경회로망으로 대체한다. 이는 방사형 기저 함수 신경회로망의 다차원 처리 능력을 부각시키고 다항식 신경회로망을 통해 최적의 유연한 모델 구조를 형성한다. 제안된 모델의 최적화를 위해 다중 목적 입자 군집 최적화(Multiobjective Particle Swarm Optimization; MoPSO) 알고리즘을 사용하였다[5]. 기존의 입자 군집 최적화 알고리즘과 달리 두 개 이상의 목적함수를 사용하여 모델의 성능뿐만 아니라 모델의 구조도 간소화하는데 중점을 두었다.

본 논문의 순서는 2장에서 방사형 기저 함수 신경회로망 및 다항식 신경회로망에 대해 서술하며, 3장에서는 다중 목적 입자 군집 최적화 알고리즘에 대해 설명한다. 4장에서는

† 교신저자, 시니어회원 : 수원대 공대 전기공학과 교수 · 공박

E-mail : ohsk@suwon.ac.kr

* 정 회 원 : 수원대 전기공학과 박사과정

접수일자 : 2011년 9월 22일

최종완료 : 2011년 11월 16일

제안된 분류기의 성능을 평가하기 위해 패턴분류에서 많이 사용되는 데이터를 이용한 실험결과를 분석하였다. 마지막으로 5장에 결론을 서술하여 본 연구에 대한 내용을 정리한다.

2. 방사형 기저 함수 기반 다항식 신경 회로망

본 장에서는 다항식 신경회로망 및 방사형 기저 함수 다항식 신경회로망의 구조 및 특징에 대해 설명하고 본 연구에서 제안한 방사형 기저 함수 다항식 신경회로망 분류기에 대해 서술한다.

2.1 다항식 신경 회로망

다항식 신경회로망(Polynomial Neural Networks; PNNs)[4]은 구조의 간략성과 상호 연계성을 고려하여 모델을 구축한다. 다항식 신경회로망은 학습을 통해 유연한 구조를 형성한다. 이는 기존의 신경회로망과는 달리 층과 노드 수가 고정되어 있지 않고 생성 및 선택 그리고 소멸을 반복하여 최적의 네트워크 구조를 가진다.

다항식 신경회로망은 2입력 2차식으로 고정되어있는 GMDH 알고리즘[6]을 확장하여 2입력 이상의 입력변수와 1차식 이상의 다항식 차수로 이루어져 있으며, 그로인해 GMDH보다 다양하고 유연한 네트워크가 형성되며 기존의 퍼지시스템이나 신경회로망과의 결합이 용이하다.

다항식 파라미터 계수를 추정하기 위해서 최소자승법(Least Square Estimation; LSE)을 사용하였으며 식 (1)과 같다.

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

만약 2입력 2차식을 고려하면 최소자승법에 필요한 행렬은 식 (2)와 같다.

$$X_i = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{11}^2 & x_{12}^2 & x_{11}x_{12} \\ 1 & x_{21} & x_{22} & x_{21}^2 & x_{22}^2 & x_{21}x_{22} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{p1} & x_{p2} & x_{p1}^2 & x_{p2}^2 & x_{p1}x_{p2} \end{bmatrix} \quad Y = [y_1 \ y_2 \ \dots \ y_p]^T \quad (2)$$

$$C_i = [c_0 \ c_1 \ c_2 \ c_3 \ c_4 \ c_5]^T$$

여기서, $i(i=1, \dots, m)$ 는 노드 번호를 나타내며, p 는 데이터의 수를 나타낸다. 입력변수의 수가 늘어나거나 차수가 높아지면 행렬의 크기는 늘어나지만 행렬을 구성하는 방법은 동일하다.

다항식 신경회로망은 기존의 신경회로망과는 달리 적합도의 개념이 존재하지 않는다. 따라서 모델의 구조 및 성능에 영향을 미치는 파라미터로는 입력변수의 수와 다항식 차수만이 존재한다. 입력변수의 수와 다항식 차수를 설정하는 방법에 따라 모델의 구조가 결정되며 총 네 개의 경우의 수로 구분할 수 있다.

- **Case 1:** 각 노드에 동일한 입력변수 사용 및 동일한 다항식 차수 선택

- **Case 2:** 각 노드에 동일한 입력변수 사용 및 서로 다른 다항식 차수 선택

- **Case 3:** 각 노드에 서로 다른 입력변수 사용 및 동일한 다항식 차수 선택

- **Case 4:** 각 노드에 서로 다른 입력변수 사용 및 서로 다른 다항식 차수 선택

각 층에서 생성되는 노드의 수는 입력변수의 수와 관계가 있다. 만약 주어진 데이터의 입력변수의 수가 n 개라 가정하고 1층 각 노드에 선택되는 입력변수의 수를 3개로 정의하면 1층에 생성되는 노드의 수는 식 (3)과 같다.

$$No. of nodes = \frac{n!}{(n-2)!2!} \quad (3)$$

만약 1층에 생성되는 노드의 수가 15개면 각 노드의 출력이 2층의 입력이 된다. 식 (1)에 의해 2층에 생성되는 노드의 수는 105개가 된다. 2층의 105개의 출력이 3층의 입력으로 넘어가면 3층에서 생성되는 노드는 2층 노드의 수백 배가 넘는 개수가 생성된다. 이와 같이 다음 층으로 넘어갈수록 노드의 수가 기하급수적으로 증가한다. 이는 연산시간 및 구조적 문제점을 가진다. 이런 문제점을 극복하기 위해서 각 층에 들어오는 입력변수의 수를 제한하며 식 (4)와 같이 표현할 수 있다.

$$No. of nodes = \begin{cases} \frac{n!}{(n-2)!2!} \left(\frac{n!}{(n-2)!2!} \leq W \right) \\ W \left(\frac{n!}{(n-2)!2!} > W \right) \end{cases} \quad (4)$$

여기서, W 는 최대 입력변수의 수가 된다. 만약 W 가 15개면 2층의 105개 노드 중 성능이 우수한 노드 순으로 정렬시켜 15개를 선택하여 3층의 입력으로 넘겨준다. 이렇게 최대 입력변수를 정해줌으로써 연산시간 및 구조의 복잡성을 최소화 할 수 있으며 W 는 사용자에게 의해 자유롭게 설정할 수 있다.

모델의 층수 또한 너무 많거나 너무 적으면 최적의 네트워크 구조를 찾는 데 어려움을 겪는다. 모델을 설계하기 전 사용자의 의해 최대 층수를 정의 하거나 식 (5)으로 층수를 정의하여 최적의 네트워크를 구축 할 수 있다.

$$E(L)_{\min} \leq E(L+1)_{\min} \quad (5)$$

여기서 E 는 각 층의 오차를 나타내며, $L(L \geq 2)$ 은 층수를 나타낸다. 식 (3)의 조건을 만족하면 프로그램을 종료하고 그렇지 않으면 다음 층으로 넘어간다.

2.2 방사형 기저 함수 신경회로망

방사형 기저 함수 신경회로망(Radial Basis Function Neural Networks; RBFNNs)은 오류역전파 학습방법을 사용하는 신경회로망에 비해 수렴속도가 빠르고 하나의 은닉층(hidden layer)만을 가지는 간략한 구조로 이루어져있다. 방사형 기저 함수 신경회로망의 구성은 입력데이터를 네트워크와 연결하는 입력층(input layer), 입력데이터를 방사형 기

저 함수에 의해 비선형 변환을 시키는 은닉층(hidden layer), 은닉층의 출력과 연결가중치를 이용하여 최종 출력값을 계산하는 출력층(output layer) 등의 세 개의 서로 다른 층으로 이루어져 있다.

방사형 기저함수는 가우시안(Gaussian) 함수, Thin plate spline 함수, 역 다항 2차(Inverse Multiquadratic) 함수, 다항 2차(Multiquadratic) 함수, 3차(Cubic) 함수등이 사용되고 있으며 그 중에서도 가우시안 함수를 일반적으로 사용한다. 가우시안(Gaussian) 함수를 활성화함수로 사용한 은닉층의 출력은 식 (6)과 같이 표현할 수 있다.

$$R_j(\mathbf{x}) = \exp\left\{-\frac{(x_1 - v_{1j})^2 + (x_2 - v_{2j})^2 + \dots + (x_n - v_{nj})^2}{2\sigma_j^2}\right\} \quad (6)$$

$$= \exp\left\{-\frac{1}{2}\left(\sum_{i=1}^n \frac{(x_i - v_{ij})^2}{\sigma_j^2}\right)\right\}$$

여기서, $v_j(j=1, \dots, c)$ 번째 노드에서 활성화함수의 중심점을 나타내며, σ 는 활성화함수의 분포상수를 나타낸다.

방사형 기저 함수 신경회로망 또한 앞서 식 (1)과 같은 최소자승법을 사용하지만, 활성화함수의 출력을 고려하기 때문에 행렬을 구성하는데 있어 다항식 신경회로망과 차이가 있다. 식 (2)와 같은 조건을 주기 위해 은닉층 노드의 수가 2개이며 2입력 2차식(Quadratic) 일 때의 행렬은 아래와 같이 구성된다.

$$\mathbf{X} = \begin{bmatrix} \hat{w}_{11} & \hat{w}_{21} & x_{11}\hat{w}_{11} & x_{21}\hat{w}_{21} & x_{11}^2\hat{w}_{11} & x_{21}^2\hat{w}_{21} & \dots & x_{11}x_{21}\hat{w}_{12} \\ \hat{w}_{21} & \hat{w}_{22} & x_{21}\hat{w}_{21} & x_{22}\hat{w}_{22} & x_{21}^2\hat{w}_{21} & x_{22}^2\hat{w}_{22} & \dots & x_{21}x_{22}\hat{w}_{22} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{w}_{p1} & \hat{w}_{p2} & x_{p1}\hat{w}_{p1} & x_{p2}\hat{w}_{p2} & x_{p1}^2\hat{w}_{p1} & x_{p2}^2\hat{w}_{p2} & \dots & x_{p1}x_{p2}\hat{w}_{p2} \end{bmatrix} \quad (7)$$

$$\mathbf{Y} = [y_1 y_2 \dots y_p]^T \quad \mathbf{C}_i = [c_0 \ c_1 \ c_2 \ c_3 \ c_4 \ c_5 \dots c_{10} \ c_{11}]^T$$

입력변수 행렬에 적합도가 고려되기 때문에 식 (2)의 행렬보다 좀 더 복잡한 형태로 이루어져 있다. 위와 같이 방사형 기저 함수의 최소자승법은 각 노드와 연결되는 연결가중치 파라미터를 한꺼번에 구하지만 다항식 신경회로망은 노드 하나의 다항식에 대해서만 최소자승법을 수행하기 때문에 행렬을 구성하는데 차이점이 있다.

모델의 최종출력은 식(8)과 같이 은닉층의 출력과 다항식 형태의 연결가중치의 선형 조합으로 표현된다.

$$\hat{y} = \frac{\sum_{j=1}^c R_j(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x})_j}{\sum_{j=1}^c R_j(\mathbf{x})} = \sum_{j=1}^c \hat{w}_j \cdot \mathbf{f}_j(\mathbf{x}) \quad \hat{w}_j = \frac{R_j(\mathbf{x})}{\sum_{j=1}^c R_j(\mathbf{x})} \quad (8)$$

여기서 $j(j=1, \dots, c)$ 는 은닉층 노드의 수를 나타내며 $\hat{w}_j$ 는 정규화된 적합도를 나타낸다.

기존의 방사형 기저 함수 신경회로망은 연결가중치가 단일 상수항이라 모델 출력을 표현하는데 어려움이 있다. 하지만 연결가중치를 다항식 형태로 표현하여 입력-출력의 관계식으로 모델 추론식을 도출하였다. 이는 모델링분야뿐만 아니라 패턴분류에서도 우수한 성능을 나타낸다.

2.3 방사형 기저 함수 기반 다항식 신경회로망

본 논문에서는 2.1과 2.2절에서 서술한 서로 다른 두 개의 신경회로망을 결합한 방사형 기저 함수 기반 다항식 신경회로망(Radial Basis Function-based Polynomial Neural Networks; RBFNNs) 패턴분류기를 제안하였으며 모델의 구조는 그림 1과 같다.

그림 1에서 I 는 입력변수의 수, C 는 클러스터의 수, m 은 퍼지화 계수, P 는 다항식 타입을 나타낸다. 제안된 모델은 각 층의 노드가 Fuzzy C-Means(FCM)기반 방사형 기저 함수 신경회로망으로 구성된다. 그림에서 보듯이 전체적인 구조는 다항식 신경회로망을 유지하지만 기존의 다항식 노드(Polynomial Node)대신 FCM기반 방사형 기저 함수 신경회로망이 삽입되었다. 따라서 각 노드에 필요한 파라미터가 입력변수의 수, 클러스터 수, 다항식 차수와 같이 구조를 결정하는 요소와 더불어 FCM 알고리즘에서 학습에 필요한 퍼지화 계수까지 고려하여 총 네 개의 파라미터가 필요하다. FCM 알고리즘을 이용하여 노드에 들어오는 입력변수 사이의 유사성을 고려하였으며, FCM 알고리즘의 소속행렬을 활성화함수의 출력(적합도)으로 사용가능하다.

그림 1 방사형 기저 함수 기반 다항식 신경회로망의 구조

Fig. 1 Architecture of radial basis function-based polynomial neural networks

방사형 기저 함수 신경회로망에서 활성화함수를 구성할 때 필요한 중심점 및 분포상수의 값을 고려하지 않아 좀 더 간소화된 모델을 구축할 수 있는 장점이 있다. FCM 클러스터링 알고리즘의 학습 과정은 다음과 같다[7].

[단계 1] 클러스터의 개수 $c(2 \leq c \leq h)$ 을 정하고 퍼지화계수 $m(1 < m < \infty)$ 을 선택한다. 초기 소속행렬(partition matrix) $\mathbf{U}^{(r)}$ 을 초기화 한다. 알고리즘 반복 횟수를 $r(r=0, 1, \dots)$ 로 표시한다.

$$\mathbf{U}^{(r)} = \left\{ u_{jk} \in [0, 1], \sum_{j=1}^c u_{jk} = 1 \forall k, 0 < \sum_{k=1}^n u_{jk} < n \forall j \right\} \quad (9)$$

[단계 2] 식 (13)을 이용하여 FCM 클러스터 중심 $\mathbf{v}$ 을 계산한다.

$$J(u_{jk}, \mathbf{v}_j) = \sum_{j=1}^c \sum_{k=1}^p u_{jk}^m (d_{jk})^2 \quad (10)$$

여기서, u_{jk} 는 0과 1사이의 소속을 나타내는 정도의 값으로 $j(j=1, \dots, c)$ 번째 클러스터에 속해져 있는 $\mathbf{x}_q$ 의 $q(q=1, \dots, n)$ 번째 데이터의 소속정도를 나타낸다. $\mathbf{v}$ 는 $j(j=1, \dots, c)$ 번째 클러스터 중심 벡터이다. m 은 소속행렬에서 애매함의 정도에 대한 영향을 나타내는 퍼지화 계수이다. 이 값은 $m \in [1, \infty)$ 와 같은 범위를 가지고 있다.

식 (12)의 목적함수를 최소화하기 위해서 먼저 $\mathbf{v}_i$ 와 u_{jk} 와 관련된 식의 목적함수를 다음과 같이 분화시켜야 한다.

$$\mathbf{v}^{(r)} = \{v_{q1}^{(r)}, \dots, v_{qc}^{(r)}\}, \quad v_{qj}^{(r)} = \frac{\sum_{k=1}^N (u_{jk})^m \cdot x_{kj}}{\sum_{k=1}^N (u_{jk})^m} \quad (11)$$

$$u_{jk} = \frac{(1/\|\mathbf{x} - \mathbf{v}_j\|^2)^{1/m-1}}{\sum_{j=1}^c (1/\|\mathbf{x} - \mathbf{v}_j\|^2)^{1/m-1}} = \frac{1}{\sum_{j=1}^c (d_{jk})^{2/m-1}} \quad (12)$$

식 (12)에서 주어진 목적함수를 만족하기 위하여 식 (13)과 식 (14)에 의해서 구해진 값들을 이용하여 각각의 클러스터 중심과 데이터와의 거리계산을 통해 새로운 소속행렬을 생성한다.

$$d_{sk} = d(x_{kj} - v_{qs}) = [\sum_{s=1}^c (x_{kj} - v_{qs})^2]^{1/2} \quad (13)$$

[단계 3] 다음과 같이 새로운 소속행렬 $\mathbf{U}^{(r+1)}$ 을 계산한다.

$$u_{jk}^{(r+1)} = \frac{1}{\sum_{j=1}^c \left(\frac{d_{sk}^{(r)}}{d_{jk}^{(r)}} \right)^{2/(m-1)}} \quad (14)$$

[단계 4] 다음 식 (10)을 계산해서 만일 $\Delta > \varepsilon$ 이면 $r=r+1$ 로 정하고 [단계 2]로 가서 다시 알고리즘을 반복 수행하고 그렇지 않고 $\Delta \leq \varepsilon$ 이면 알고리즘을 종료한다. 여기서, ε 는 임계값이다.

$$\Delta = \|\mathbf{U}^{(r+1)} - \mathbf{U}^{(r)}\| = \max_{j,k} |u_{jk}^{(r+1)} - u_{jk}^{(r)}| \quad (15)$$

FCM 기반 방사형 기저 함수 신경회로망은 은닉층에 FCM 알고리즘을 사용한 부분을 제외하고는 2.2절에 설명한 학습 및 추론 방법이 동일하다.

일반적으로 패턴분류 실험에 사용되는 데이터는 출력의 개수가 곧 클래스의 수가 된다. 따라서 출력의 개수가 최소 2개 이상 되며 모델에서는 출력층의 노드가 클래스의 수만큼 생성되며 그로 인해 고려해야 될 파라미터가 많아진다. 하지만 제안된 모델은 출력층에서 하나의 노드로 클래스를 분류하였으며 식 (16)과 같다.

$$Class_{cs} = \min((y_{cs} - \hat{y})^2) \quad (16)$$

여기서, $y_{cs}(cs=1, \dots, \text{No. of class})$ 는 실제 데이터의 클래스를 나타내며, $\hat{y}$ 는 모델의 출력을 나타낸다. 각 클래스와 모델 출력의 오차가 가장 작은 y_{cs} 을 모델의 최종 출력으로 선택한다. 만약 클래스를 세 개로 가정하면 첫 번째 클래스는 -1, 두 번째 클래스는 0, 그리고 세 번째 클래스는 1로 치환하여 모델을 학습시키고 모델의 출력이 -1, 0, 1중 가장 오차가 작은 값에 해당하는 클래스를 선택하게 된다.

3. 다중목적 입자 군집 최적화 알고리즘을 이용한 최적화 방법

본 장에서는 제안된 모델의 구조 및 파라미터를 최적화하기 위해 적용된 다중목적 입자 군집 최적화 알고리즘의 대해 설명한다.

3.1 다중목적 입자 군집 최적화 알고리즘

입자 군집 최적화 알고리즘(Particle Swarm Optimization; PSO)은 새, 벌, 그리고 물고기같이 집단생활을 하는 무리들의 행동양식을 기반으로 알고리즘 개발되었다. PSO는 기존의 진화 알고리즘에 비해 내용이 간결하고 적용이 용이한 장점을 있으며 성능 또한 우수하여 최적화 분야에서 많은 주목을 받고 있다. 그러나 실제 응용분야에서는 다양한 문제를 동시에 고려하는 특성상 기존의 최적화 알고리즘으로는 해결할 수 없는 문제들이 존재한다. 이런 문제를 해결하고자 다중목적 최적화 알고리즘이 제안되었으며, 그 중에서 다중 목적 입자 군집 최적화 알고리즘은 2002년 coello-coello et. al.[5] 의해 제안되었다. 그 후에 여러 이론들이 결합되면서 다양한 형태의 다중 목적 최적화 알고리즘이 개발되고 있다[8]. 본 연구에서는 다양한 MoPSO 중에서 crowding distance를 이용한 MoPSO-CD 알고리즘을 사용하였으며 최적화 과정은 아래와 같다[8].

[Step 1] 최적화를 이용해 구하고자하는 파라미터의 차원(Dimension; D) 수로 이루어진 개체(Particle; P)를 집단(Swarm; S)의 크기만큼 랜덤하게 생성한다.

[Step 2] 목적함수에 대입하여 각 개체의 적합도(Fitness; F)를 구한다. 초기 Swarm은 바로 Pbest가 된다.

[Step 3] 각 개체들의 적합도를 비교하여 모든 목적함수에 대해 적합도가 우수한 개체를 선별하여 외부 저장소(Archive; A)에 저장한다.

[Step 4] A에 저장된 개체들을 가지고 있는 적합도를 이용하여 Crowding distance를 계산하고 내림차순으로 개체를 재 정렬한다.

$$CD_{F_i}(j) = CD_{F_i}(j) + (F_{(i+1)} - F_{(i-1)}) \quad (17)$$

여기서, i 는 개체의 번호를 나타내며, j 는 목적함수의 수

를 나타낸다.

[Step 5] Crowding distance로 재 정렬된 저장소(A)의 적합도를 가지고 상위 10% 이상 적합도를 gbest(i)로 설정한다. 이동속도를 구할 때 하나의 gbest만 사용하지 않고 gbest(i) 중에서 랜덤하게 선택하여 이동속도를 계산 한다.

$$v_i(t+1) = w \cdot v_i(t) + c_1 \cdot r_1(pbest_i(t) - swarm_i(t)) + c_2 \cdot r_2(A[gbest_i(t)] - swarm_i(t)) \quad (18)$$

여기서, w 는 하중계수로 0.9~0.4로 선형적으로 감소시켰으며, c 는 가속상수로 일반적으로 $c_1, c_2 = 2.0$ 으로 설정하며, r 은 랜덤상수로 $r_1, r_2 = [0, 1]$ 사이의 값을 나타낸다.

[Step 6] 새로운 개체 $P(t+1)$ 를 생성한다.

$$P_i(t+1) = P_i(t) + v_i(t+1) \quad (19)$$

새로운 개체 또한 초기 탐색 범위 ($P_{\min} \leq P_i(t+1) \leq P_{\max}$)을 만족해야 된다.

[Step 7] $IF(t < Maxgen \times Mutation)$ 을 만족하면 식 (20)의 돌연변이를 수행한다.

$$P'_{ik}(t+1) = rand \times (P_k^{\max} - P_k^{\min}) + P_k^{\min} \quad (20)$$

여기서, t 는 현재 세대를 나타내며, $Maxgen$ 은 전체 세대를 나타낸다. $Mutation$ 은 돌연변이 확률로 조건식을 만족할 때만 돌연변이를 수행한다. 여기서 k 는 개체에 선택된 차원수로 랜덤하게 선택된다.

[Step 8] 새로운 개체 $P(t+1)$ 의 적합도를 계산한다.

[Step 9] 새로운 적합도 $F(t+1)$ 와 저장소(A)에 저장된 적합도를 비교하여 A를 갱신한다. A의 크기는 고정되어 있기 때문에 새로운 적합도의 crowding distance를 구해 내림차순으로 정렬하고 A의 하위 10%에 있는 적합도 중 랜덤하게 선택하여 제거하고 그 자리에 새로운 적합도가 삽입되는 방식으로 A를 갱신한다.

[Step 10] pbest(t)와 swarm(t+1)을 비교하여 pbest를 갱신하여 pbest(t+1)에 저장한다.

[Step 11] 종료조건을 만족하면 프로그램을 종료하고, 그렇지 않으면 [Step 5]로 이동해서 반복한다.

위와 같은 과정을 거쳐 여러 개의 목적함수를 모두 만족하는 최적의 파라미터를 찾을 수 있다. 다중 목적 최적화 알고리즘은 주어진 탐색 공간 중 하나의 최적해를 찾는게 아니라 탐색공간에서 여러개의 목적함수를 만족하는 탐색영역을 표현한다. 제안된 모델에서 고려해야할 파라미터는 총 네 개이며 목적함수는 총 3개를 사용하였다. 모델의 최적화 방법은 3.2절에서 자세히 서술한다.

3.2. 제안된 모델의 최적화 방법

이번 절에서는 제안된 모델의 최적화 방법에 대해 서술한다. 모델에 필요한 파라미터는 입력변수의 수, 클러스터의 수, 퍼지화 계수, 다항식 차수 등 네 개의 파라미터를 고려해야 되며 이를 다중 목적 최적화 알고리즘을 이용하여 동조한다. 앞서 말했듯이 목적함수는 모델의 패턴분류율 및 정확성 그리고 다항식을 이루는 파라미터 계수의 수 등을 고려하였으며 식 (21) ~ (23)과 같이 각각의 목적함수를 표현하였다.

$$Objective\ function\ 1 = \frac{False}{N} \quad (21)$$

첫 번째 목적함수에서 F 는 분류실패 횟수를 나타내며, N 은 전체 데이터의 수를 나타낸다.

$$Objective\ Function\ 2 = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{class} (y_{ij} - \hat{y}_{ij})^2 \quad (22)$$

두 번째 목적함수는 정확성을 측정하기 위한 식으로 y_i 는 실제 데이터의 출력이 되며 $\hat{y}_i$ 는 모델 출력이 된다.

$$Objective\ Function\ 3 = C + I + NP \quad (23)$$

세 번째 목적함수는 구조의 복잡성을 판단하기 위한 식으로 C 는 클러스터의 수, I 는 입력변수의 수, NP 는 다항식의 파라미터 계수를 나타낸다.

다중목적 입자 군집 최적화 알고리즘을 이용하여 모델의 최적화 학습공정은 그림 2와 같다.

그림 2 제안된 모델의 최적화 과정

Fig. 2 Optimization process of proposed model

제안된 모델에서 최적화는 층수만큼 반복된다. 만약 총 3층으로 모델을 구성하고자 하면 입력변수와 1층, 1층과 2층, 2층과 3층에서 최적화가 진행된다. 이렇게 3층으로 모델을 구성하고자 하면 총 3번의 최적화 알고리즘이 사용된다. 위의 그림 5는 제안된 모델에서 한 층의 최적화 과정을 나타낸 것으로 위의 과정을 층 수 만큼 반복해야 된다. 개체의 파라미터 수는 입력변수의 수에 따라 달라지며 각 파라미터

의 탐색범위도 데이터의 특징에 따라 다르게 설정해줘야 좀 더 효율적인 모델을 구성할 수 있다. 따라서 파라미터의 탐색범위에 대한 설명은 4장 실험결과 및 고찰에서 다루었다.

4. 실험결과 및 고찰

본 장에서는 제안된 분류기의 성능 평가를 위해 Iris, pima, liver 데이터를 사용하였다. 또한 기존 분류기와 비교를 통해 제안된 모델의 우수성을 입증하였다. 표 3는 다중 목적 입자 군집 최적화 알고리즘의 초기 설정 파라미터와 탐색범위를 나타낸다. 초기 설정된 수치적 값들은 학습과 성능의 수렴에 따른 페이스에 따라서 실험적 값과 모니터링의 운행에 의해서 시행착오적으로 선택하였다.

표 1 다중 목적 입자 군집 최적화 알고리즘의 설정 파라미터

Table 1 Set-up parameters of MoPSO

<i>Parameters of MoPSO</i>	
Maximum generations per each layer	60
Swarm size per each layer	30
Archive size per each layer	20
acceleration constant (c_1, c_2)	2.0
random constant (r_1, r_2)	[0 1]
Inertia weight (w)	[0.4 0.9]
Mutation rate	0.3
<i>Range of search space to find optimal parameters</i>	
input variables	[2 ~ 3]
Number of clusters	[2 ~ 5]
Fuzzification coefficient	[1.1 ~ 3.0]
Polynomial type	[1 ~ 4]

4.1 Iris 데이터

Iris 데이터는 4입력 3클래스 150개의 데이터로 구성되어 있다. 이 데이터는 패턴분류 분야에서 모델의 성능을 평가할 때 많이 쓰이는 데이터이며, 제안된 모델에서는 통계학적인 성능을 구하기 위해 5 fold cross validation(fcv)을 사용하여 모델의 패턴분류를 및 정확성을 평균과 표준편차로 나타냈다.

표 2 Iris 데이터의 패턴분류율 및 정확성

Table 2 Pattern classification rate and accuracy for Iris dataset

L	PCR of training data	PCR of testing data	Accuracy of training data	Accuracy of testing data
1	96.83 ± 1.49	96.67 ± 3.33	0.0291 ± 0.0136	0.0333 ± 0.0000
2	99.33 ± 0.69	97.33 ± 2.78	0.0079 ± 0.0071	0.0257 ± 0.0233
3	100.00 ± 0.00	94.67 ± 1.82	0.0030 ± 0.0004	0.0506 ± 0.0182

표 2는 각 층에서의 패턴분류율 및 정확도를 나타낸 것이다. Iris 데이터는 3층까지만 최적화를 통해 모델을 구축하였으며 학습 데이터의 경우 층이 증가할수록 성능이 좋아지는 반면 테스트 데이터의 경우 2층에서 가장 좋은 성능지수를 나타낸다. 따라서 이 경우는 3층으로 모델을 구성하기보다 2층으로 모델을 설계함으로써 좀 더 안정적인 패턴분류 및 정확성을 보인다.

그림 3은 다중 목적 입자 군집 최적화에서 저장

A(Archive)에 저장된 적합도의 분포를 나타낸다. 1층에 저장되는 개체는 총 20개이며 각 적합도의 분포는 그림 7(a)와 같다. 이 중에서 적합도가 우수한 개체 10개를 선택하여 개체들의 출력이 다음 2층의 입력으로 넘어간다. 그림 7(b)는 2층의 저장소 A에 저장된 적합도의 분포로 1층과 마찬가지로 우수한 개체 10개를 선택하여 출력이 3층의 입력으로 넘어간다. 3층에서는 마지막 층으로 이 중에서 패턴분류와 정확성 그리고 파라미터 계수가 가장 우수한 개체를 선택하여 모델의 최종출력으로 선택한다.

(a) 1 Layer

(b) 2 Layer

(c) 3 Layer

그림 3 제안된 모델의 최적화 과정

Fig. 3 Optimization process of proposed model

하지만 표 2에서 설명했듯이 2층에서 학습 데이터와 테스트 데이터의 성능이 더 안정적이기 때문에 3층으로 넘어가지 않고 2층의 가장 우수한 개체를 선택하여 2층으로 이루어진 모델을 설계하였으며,

그림 4 제안된 모델의 최적화 과정

Fig. 4 Optimization process of proposed model

그림 4과 같다. 제안된 모델은 다항식 신경 회로망을 기본 구조로 사용하기 때문에 최종출력은 각 층과 연결과 노드로 표현된다. 따라서 그림 4는 2층과 연결된 1층 노드 2개와 1층 노드와 연결된 입력변수를 모두 사용하여 모델을

나타내었다. 그림에서 보면 1층에서 선택된 노드는 총 10개의 노드중 4번째와 8번째의 노드가 선택되었다. 4번째 노드의 파라미터는 입력변수의 수는 3개(x_1, x_2, x_3), 클러스터의 수는 5개, 퍼지화 계수는 2.5, 그리고 다항식 차수는 quadratic이 선택되었으며, 8번째 노드의 파라미터는 입력변수의 수는 3개(x_2, x_3, x_4), 클러스터의 수는 5개 퍼지화 계수는 2.1, 그리고 다항식 차수는 modified quadratic이 선택되었다.

표 3은 기존 모델의 분류기와 제안된 모델의 평균 패턴분류율을 비교하였다. NEFCLASS는 neuro-fuzzy를 이용한 패턴분류기로 모델의 학습은 오류역전파(back propagation)를 사용하였다. C4.5는 decision tree를 생성하여 패턴분류를 하는 모델이며 FID3.1은 decision tree에 fuzzy 개념을 적용한 분류기로 C4.5나 FID3.1 모두 ID3(Iterative Dichotomiser 3) 알고리즘을 기반으로 확장 및 응용한 알고리즘이다. K-nearest neighbor(KNN) 알고리즘은 K의 영역에서 가장 많이 포함된 클래스로 데이터의 패턴을 인식하는 알고리즘이며 Tabu search-k nearest neighbor(TS-KNN)와 Locally adaptive metric nearest neighbor(ADMENN)은 KNN을 개선시킨 모델이다.

표에서 보듯이 제안된 모델의 성능이 기존의 모델보다 우수함을 알 수 있으며 특히 2층에서 97.3% 가장 높은 패턴분류율을 보인다.

표 3 기존모델과의 성능지수 비교

Table 3 Comparison of pattern classification rate for several classifiers

Classifier	Pattern classification rate [%]
RBFNNs [9]	90.1
NEFCLASS [10]	96.0
C 4.5 [11]	94.0
FID 3.1 [12]	96.0
KNN [13]	94.7
TS-KNN [13]	96.7
LAMNN [13]	97.0
Proposed model 2 Layer	97.3
Proposed model 3 Layer	94.7

4.3 Liver 데이터

Liver 데이터는 6입력 2클래스의 데이터로 총 345개로 구성되어있다. 이 데이터도 pima 데이터와 마찬가지로 5 fold cross validation을 사용하여 모델의 성능을 평균과 표준편차로 나타내었다. 이번 실험에서는 학습 데이터를 한 번 더 나누어 총 데이터에서 학습데이터 173개(50%) 검증데이터 103개(30%) 그리고 테스트데이터 69개(20%)가 되도록 나누었다. 데이터를 세 개로 나누어도 실제 평가되는 테스트데이터는 기존의 5fcv와 동일하다. 다만 검증데이터를 추가하여 학습데이터와 검증데이터로 모델을 구축함으로써 테스트데이터의 성능이 저하 및 발산하는 현상을 줄이고 보다 안정적인 성능을 나타낸다. 총 4층까지 학습하였으며 모델의 패턴분류율 및 정확성은 표 4와 같다.

표 4를 통해 각 층에서의 평균 패턴분류율 및 정확성을 알 수 있으며, 그림 12를 통해 각 fcv에서의 패턴분류율 나타내었다. 학습데이터의 패턴분류율은 층수가 올라갈수록

좋아지지만 테스트데이터의 패턴분류율은 3층에 77.31%로 가장 우수하다. 4층의 경우 학습데이터는 3층에 비해 크게 개선되었지만 테스트데이터의 성능이 떨어지고 표준편차가 큰 것을 알 수 있다.

표 4 Liver 데이터의 패턴분류율 및 정확성

Table 4 Pattern classification rate and accuracy for Liver dataset

L	PCR of training dataset	PCR of validation dataset	PCR of testing dataset	Accuracy of training dataset	Accuracy of validation dataset	Accuracy of testing dataset
1	63.9±7.4	64.9±3.6	73.6±7.8	0.222±0.021	0.229±0.017	0.232±0.047
2	80.9±2.4	63.7±6.4	77.3±2.4	0.140±0.026	0.641±0.843	0.284±0.082
3	88.1±1.3	63.5±6.1	77.3±1.9	0.088±0.014	0.229±0.078	0.366±0.199
4	93.6±1.8	60.6±3.9	74.7±5.6	0.052±0.019	0.649±0.797	10e+4±20e+4

표 5는 기존 분류기와 제안된 분류기의 평균 패턴분류율을 비교하였다. 제안된 분류기의 경우 2층과 3층에서 테스트데이터의 평균 분류율이 77.31%로 가장 좋았으며 다른 분류기보다 우수함을 알 수 있다.

표 5 기존모델과의 성능지수 비교

Table 5 Comparison of pattern classification rate for several classifiers

Classifier	Pattern classification rate [%]
KNN [13]	62.90
TS-KNN [13]	73.80
MLP [14]	67.69
RBF1 [15]	68.32
RBF2 [15]	70.47
Proposed model 2 Layer	77.31
Proposed model 3 Layer	77.31
Proposed model 4 Layer	74.78

5. 결 론

본 논문에서는 다항식 신경 회로망과 방사형 기저 함수 신경회로망을 결합하여 새로운 형태의 분류기를 제안하였다. 방사형 기저 함수 기반 다항식 신경 회로망은 다항식 신경회로망을 기본 구조로 하고 각 노드가 FCM 기반 방사형 기저 함수 신경회로망으로 구성된다. 다항식 신경회로망의 유연한 구조와 고차원의 입력-출력변수에도 학습능력이 뛰어난 방사형 기저 함수 신경회로망을 사용함으로써 데이터 특성에 맞는 최적의 분류기를 구축하였다. 또한 다중 목적 입자 군집 최적화 알고리즘을 사용함으로써 분류기의 패턴분류뿐만 아니라 정확성 및 구조도 모두 고려한 최적의 모델 설계를 위해 파라미터를 동조하였다. 제안된 모델의 성능은 실험결과를 통해 기존의 분류기보다 우수한 성능을 보이며 데이터에 따라 서로 다른 구조로 이루어진 최적의 모델을 구축하였다.

감사의 글

본 연구는 중소기업청에서 지원하는 2011년도 산학연 공동 기술개발사업(No. 00047279)의 연구수행으로 인한 결과물임을 밝힙니다. 그리고 본 연구는 경기도의 경기도지역협력연구센터(GRRC) 사업의 일환으로 수행하였음 [GRRC수원2011-B2, U-city 보안감시 기술협력센터]

참 고 문 헌

- [1] Y. W. Wong, K. P. Seng and L. M. Ang "Radial Basis Function Neural Network with Incremental Learning for Face Recognition," *Systems Man, and Cybernetics, Part B: Cybernetics, IEEE Trans. on*, Vol. 41, No.4, pp. 940-949, 2011.
- [2] K. A. A. Aziz, R. A. Ramlee, S. S. Abdullah, and A.N. Jahari, "Face Detection Using Radial Basis Function Neural Networks with Variance Spread Value," *International Conference of Soft Computing and Pattern Recognition 2009(SOCPAR'09)*, 399-403.
- [3] M. Hou, X. Han, "Constructive Approximation to Multivariate Function by Decay RBF Neural Network," *IEEE Trans. on Neural Networks*, Vol. 21, No. 9, pp. 1517-1523, 2010.
- [4] S. B. Roh, S. K. Oh, and W. Pedrycz, "A Fuzzy Ensemble of Parallel Polynomial Neural Networks with Information Granules Formed by Fuzzy Clustering," *Knowledge-Based Systems*, Vol. 23, No. 3, pp. 202-219, 2010.
- [5] C. A. Coello Coello, and M. S. Lechuga, "MOPSO: A Proposal for Multiple Objective Particle Swarm Optimization," *Proceedings of the 2002 IEEE Congress on Evolutionary Computation*, pp.12-17.
- [6] I. Hayashi, and H. Tanaka, "The Fuzzy GMDH Algorithm by Possibility Models and Its Application," *Fuzzy Sets and Systems*, Vol. 36, pp. 245-258, 1990.
- [7] J. C. Bezdek, R. Ehrlich, and W. Full "FCM: The Fuzzy C-Means Clustering Algorithm" *Computers & Geosciences*, Vol. 10, No. 2-3, pp. 191-203, 1984.
- [8] C. R. Raquel and P. C. Naval Jr, "An Effective Use of Crowding Distance in Multiobjective Particle Swarm Optimization," *Proc. Genetic and Evol. Comput. (GECCO 2005)*, pp. 257-264, 2005.
- [9] M. H. Hassoun, "Fundamentals of Artificial Neural Networks," *The MIT Press, Cambridge, MA*, 1995.
- [10] F. Klawounn, D. Nauck, and R. Kruse, "Generating Rules from Data by Fuzzy and Neuro-fuzzy Methods," in *Poc. Fuzzy-Neuro-Systems (1999)*, pp. 223-230.
- [11] J. R. Quinlan, "Improved Use of Continuous Attributes in C4.5," *J. Artificial Intelligence Press 4 (1996)*, pp. 77-90.
- [12] C. Z. Janikow, and M. Faifer, "Fuzzy Partitioning with FID 3.1," *In proc. IEEE 18th Int. conf. North American Fuzzy Information Proceedings Society (1999)*, pp. 467-471.
- [13] M. A. Tahir, A. Bouridane, and F. Kurugollu, "Simulatneous Feature Selection and Feature Weighting Using Hybrid Tabu Search/K-Nearest Neighbor Classifier," *Pattern Recognition Letters*, Vol. 28, pp. 438-446, 2007.
- [14] R. O. Duda, P. E. Hart, "Pattern Classification and Science Analysis," *New York: Wiley*, 2002.
- [15] Z. R. Young, "A Novel Radial Basis Function Neural Network for Discriminant Analysis," *IEEE Trans. Neural Networks*, Vol. 17, No. 3, pp. 604-612, 2006.

저 자 소 개

김 옥 동 (金 旭 東)

2009년 수원대학교 전기공학과 졸업.
2011년 동 대학원 석사. 2011년~현재 동 대학원 박사과정. 관심분야는 퍼지추론 시스템, 뉴럴 네트워크, Granular Computing 진화론적 최적화 알고리즘.
Tel : 031-222-6544
E-mail : wdlim@suwon.ac.kr

오 성 권 (吳 聖 權)

1981년 연세대학교 전기공학과 졸업, 동 대학원석사(1983),박사(1993). 1983-1989년금성산전연구소(선임연구원). 1996-1997년 캐나다 Manitoba 대학 전기 및 컴퓨터공학과Post-Doc. 1993-2004년 원광대학교 전기전자 및 정보공학부 교수. 2005~현재 수원대학교 전기공학과 교수, 2002~현재 대한전기학회, 제어로봇시스템학회, 퍼지및지능시스템학회 편집위원. 관심분야는 퍼지 시스템, 퍼지-뉴럴 네트워크, 자동화 시스템, 고급 computational intelligence, 지능 제어등.
Tel : 031-229-8162
Fax : 031-220-2667
E-mail : ohsk@suwon.ac.kr