

예보인자의 효과적 추출을 위한 다항식 방사형 기저 함수 신경회로망 기반 초단기 강수예측 분류기의 설계

Design of Very Short-term Precipitation Forecasting Classifier Based on Polynomial Radial Basis Function Neural Networks for the Effective Extraction of Predictive Factors

김 현 명* · 오 성 권† · 김 현 기*
(Hyun-Myung Kim · Sung-Kwun Oh · Hyun-Ki Kim)

Abstract - In this study, we develop the very short-term precipitation forecasting model as well as classifier based on polynomial radial basis function neural networks by using AWS(Automatic Weather Station) and KLAPS(Korea Local Analysis and Prediction System) meteorological data. The polynomial-based radial basis function neural networks is designed to realize precipitation forecasting model as well as classifier. The structure of the proposed RBFNNs consists of three modules such as condition, conclusion, and inference phase. The input space of the condition phase is divided by using Fuzzy C-means(FCM) and the local area of the conclusion phase is represented as four types of polynomial functions. The coefficients of connection weights are estimated by weighted least square estimation(WLSE) for modeling as well as least square estimation(LSE) method for classifier. The final output of the inference phase is obtained through fuzzy inference method. The essential parameters of the proposed model and classifier such as input variable, polynomial order type, the number of rules, and fuzzification coefficient are optimized by means of Particle Swarm Optimization(PSO) and Differential Evolution(DE). The performance of the proposed precipitation forecasting system is evaluated by using KLAPS meteorological data.

Key Words : RBFNNs(Radial Basis Function Neural Networks), FCM(Fuzzy C-means Method), PSO(Particle Swarm Optimization), DE(Differential Evolution), precipitation forecasting model and classifier

1. 서 론

강수 예측 시 초단기 강수는 그 특성이 돌발적이고 비선형적이다. 이 특성에 의한 집중 호우는 최근 들어 더욱 빈번하게 발생하며 인명 및 재산피해가 심각해지고 있다. 기상 예보는 예보관이 각 기상 자료 및 예측 시스템의 출력을 참고하여 이뤄진다. 만약 호우 특보 발령에 문제가 생긴다면 국가 및 개인의 피해는 심각할 것이다. 따라서 호우 특보의 발령은 매우 정확하고 신속해야 한다. 본 논문에서는 예보관의 판단을 보조하는 예측 시스템의 개발과 개발된 예측시스템의 성능을 높이는 데 목적을 둔다. 본 논문에서 제안한 예측 시스템은 지능형 뉴로-퍼지 알고리즘을 기반으로 하며 최적화 알고리즘과 결합하여 최적화된 성능의 예측 시스템을 개발한다. 본 연구에서는 뉴로-퍼지 알고리즘인 방사형 기저 함수 신경회로망 RBFNN(Radial Basis Function Neural Networks)의 두 가지 접근법 즉 패턴분류에 목적을 두는

RBFNN 과 정량적인 출력을 가지는 모델링 기법을 각각 호우 예측 시스템에 맞게 개발하여 비교한다. 또한 진화 최적화 알고리즘인 PSO와 DE 알고리즘을 적용한 비교도 함께 진행한다. 본 논문에서는 KLAPS(Korea Local Analysis and Prediction System) 재분석 기상자료를 AWS(Automatic Weather Station)지점에 매칭시켜 사용한다. 또한 방대한 기상 자료를 적절하게 샘플링 하여 데이터의 수를 축소 시켜야 한다. 또한 기상데이터는 호우특보 예측 시 불필요한 인자가 있을 수 있다. 이는 최적화 알고리즘을 통해 선별이 가능하다. 기상 자료는 시간에 따른 변화와 연속적인 특성이 있기 때문에 시계열 데이터로 인자를 추가하여 사용하며 기존에 주어진 1시간 누적강수량을 이용하여 2~6시간 누적강수량을 생성한다. 예측 시스템 개발 과정은 크게 기상 데이터 전처리 부분과 RBFNN기반 모델 및 분류기 구축, 테스트를 통한 성능 검증으로 나뉜다. 모델 및 분류기의 최종 출력은 호우 주의보의 발령 기준을 따른다.

2. 다항식 방사형 기저함수 신경회로망 기반 초단기 강수예측 모델 및 분류기 설계

본 절에서는 지능형 뉴로-퍼지 알고리즘인 다항식 방사형 기저함수 신경회로망을 기반으로 한 강수 예측 모델 및 분류기에 대해서 설명한다. 강수 예측 시스템은 크게 강수 예측 패턴분류

† Corresponding Author : Dept. of Electrical Engineering, The University of Suwon, Korea

E-mail : ohsk@suwon.ac.kr

* Dept. of Electrical Engineering, The University of Suwon, Korea

Received : November 26, 2013; Accepted : December 22, 2014

기와 강수예측 모델 두 가지 개념으로 개발된다. 강수 예측 패턴 분류기는 출력이 0,1로 이뤄지며 모델의 출력은 정량적인 강수량으로 출력된다. 본 연구에서는 두 시스템의 장단점을 파악하고 어떤 시스템이 호우 예측 시스템에 더 적합한지 판단한다. 전체적인 호우 예측 모델의 시스템 구성도는 그림 1과 같다.

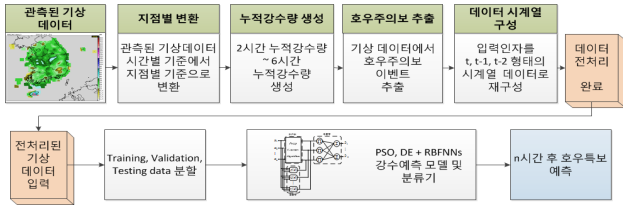


그림 1 호우 예측 시스템의 전체적인 개요

Fig. 1 Overall schematic diagram for heavy rain prediction system

개발된 시스템의 전체 흐름은 우선 관측된 KLAPS 재분석 기상 자료를 입력으로 사용한다. 학습하는데 필요한 정보, 즉 입력 데이터나 주요 파라미터들은 개발자가 적절하게 처리하여 만들어 주는 작업이 필요하다. 우선 기상데이터의 특성 분석을 통한 주요 입력인자의 선택이 이루어진다. 이 후 선택된 입력인자의 전처리를 통해 모델 및 분류기의 입력인자로 사용한다. 전처리 과정은 총 4단계를 거친다. 먼저 기상데이터를 시간 기준에서 지점기준으로 변환하고, 강수의 변화량을 입력으로 사용하기 위해 누적강수량을 추가 생성한다. 이 후 호우 주의보 추출을 통한 데이터의 출력에 대한 집중도를 높였으며 입력데이터 시계열 데이터 구성을 통해 시간의 변화에 따른 강수의 변화를 데이터에 적용하였다.

2.1 KLAPS 재분석 기상 데이터

제안된 모델 및 분류기의 입력데이터로 사용되는 KLAPS 기상 데이터는 기상청에서 현업으로 사용하는 초단기 예측 시스템의 입력데이터를 동일하게 사용한다. KLAPS는 한반도 주위의 영역에 대하여 수평, 수직 5Km 해상도의 기상데이터를 재분석 및 예측하는 시스템이다. KLAPS의 기상데이터는 기압, 습도, 기온, 수평 풍속, 수직 풍속, 1시간 누적강수량이며 이 6가지 인자를 1시간 간격으로 생성한다. 표 1은 KLAPS 재분석 기상 데이터의 인자이다. 본 연구에서 사용한 KLAPS 재분석 기상 데이터는 전체 격자 253*283 중 한반도 육지에 해당하는4,527개의 격자점을 선

표 1 KLAPS 데이터의 시간별 날씨 인자

Table 1 Hourly weather factors being used in KLAPS data

번호	설명
1	기압(Pa)
2	습도(%)
3	기온(K)
4	수평 풍속(UWND)
5	수직 풍속(VWND)
6	1시간 누적 강수량(mm)

택하여 사용하였다.

2.2 KLAPS 재분석 기상 데이터 전처리

기상데이터의 전처리 과정은 총 4단계로 이뤄진다. KLAPS 재분석 기상데이터의 초기 포맷은 시간별로 저장되어있다. 이는 본 연구에서 개발된 초단기 호우 예측 시스템에서 테스트 할 시 적합하지 않다. 테스트는 지점별로 이뤄지기 때문이다. 이를 지점별로 데이터를 재 저장한다. 다음으로 초기 변수 중 누적강수량의 추가 생성이다. 초기 1시간 누적강수량으로 2~6시간 누적강수량을 추가 생성한다. 추가 누적강수량을 생성하여 사용하는 이유는 시스템 개발 중 1시간 누적강수량의 입력변수만으로는 한계가 있어 추가 생성하여 사용하여 성능의 개선을 보았기 때문이다. 추가 누적강수량 생성에서 6시간 누적강수량의 생성은 개발된 모델 및 분류기의 출력을 만드는데 중요하다. 아래 표 2는 출력인 호우주의보의 기준을 보여준다.

표 2 각 형태별 야간 감시 장비의 비교

Table 2 Comparison of each type of night vision equipment

6시간 누적강수량	호우주의보 발령 기준	모델 출력	분류기 출력
70mm 미만	미발령	70mm 미만	0
70mm 이상	발령	70mm 이상	1

KLAPS 레이더 관측 자료는 격자구조 235*283으로 남북한에 해상까지 포함되어 있는 좌표이다. 전체 육지 좌표점은 9,291개 격자점이며 여기서 한반도 전체 격자점(4,627개)만 추출하여 해당 데이터만 사용하였다. 초기 KLPAS 레이더 관측 자료는 시간별로 저장되어있기 때문에 최종적인 출력인 지점(격자)별 호우 예측에 맞지 않는 형태로 되어있다. 이를 아래 그림 3와 같이 각 지점(좌표)별 데이터로 변환한다. 데이터의 기간은 2006년~2011년이며 년도 별로 구분할 수 있도록 데이터를 구성하였다. 이는 이후 훈련, 검증, 테스트 데이터 분할이 용이하도록 하기 위함이다. 지점별 변환은 전체 남한 격자점에 대해서 변환을 실행하였고 초기 KLAPS 레이더 관측 자료에서 기압, 습도, 온도, 1시간 누적강수량만 추출하였다. 이 외에도 풍향, 풍속, 등 여러 사전인자 존재하지만 강수에 직접적인 영향이 없는 인자는 제외하였다.

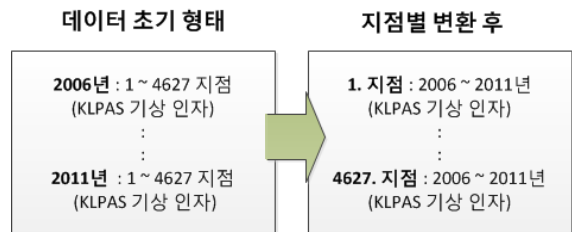


그림 2 KLAPS 데이터의 변환 방법

Fig. 2 Data format of individual branches using KLAPS data

2.2.1 KLAPS 기상데이터의 누적강수량 생성

위 표 2의 기상특보 발령기준을 보면 6시간 강우량과 12시간 강우량을 기준으로 하여 특보를 발령한다. 이에 따라 호우 예측 모델의 입, 출력에서는 각 누적강수량이 필요하며 시간에 흐름에 따라 강수의 누적 변화량을 입력으로 사용할 수 있도록 초기 1시간 누적강수량을 이용하여 2시간 누적강수량~6시간 누적강수량을 새롭게 생성하였다. 본 논문에서는 호우 주의보를 예측하므로 6시간 이상의 누적강수량을 사용하지 않았다. 새롭게 누적강수량이 적용된 KLAPS 기상 데이터의 형태는 다음 그림 3과 같다.

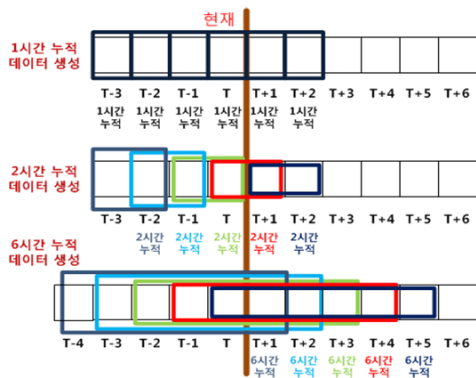


그림 3 KLAPS데이터의 누적 강수량(2시간 ~ 6시간)의 생성
Fig. 3 Method to build up accumulated rainfall amount (1 hour to 6 hours) using KLAPS Data

2.2.2 KLAPS 기상데이터 호우 주의보 추출

본 논문에서 사용한 KLAPS 기상데이터의 기간은 2006년~2011년 총 6년간의 데이터를 사용한다. 기상데이터 1년간의 데이터는 시간단위로 환산하면 8,760개 이다. 이를 6년으로 하면 52,560줄이다. 여기에 다양한 입력변수를 고려하면 그 수는 급격하게 늘어나게 된다. 이와 같이 방대한 기상데이터가 바로 입력으로 사용될 시 연산시간의 문제 및 성능저하의 원인이 될 수 있다. 이는 전처리의 세 번째 단계인 호우주의보 추출을 통해 데이터의 축소를 통해 해결할 수 있다. 호우 주의보 이벤트는 호우 주의보의 시작점과 끝점을 하나의 이벤트로 치며 시작점 전의 4시간과 끝점 후의 4시간의 비호우 데이터를 포함하여 2006년~2011년, 4,527 격자점에 대해 추출하여 입력 데이터를 구성하였다.

2.2.3 KLAPS 기상데이터의 시계열 데이터로 재구성

위 2.2.1절의 누적강수량 생성에 의하여 시간에 따른 강우의 변화를 잘 나타낼 수 있도록 재구성한 후 성능개선을 위하여 시간에 따른 강우의 변화표현을 더욱 강화하기 위하여 그림 4와 같이 각 인자별 기압 : t, t-1, 각 누적 강수량 : t, t-1, t-2 으로 시계열 데이터로 재구성 하였다. 이로 인하여 기존의 KLAPS 기상데이터에 비하여 시간의 변화에 따른 강수량의 변화 특징을 데이터 상에서 더욱 확실하게 표현할 수 있게 하였다.

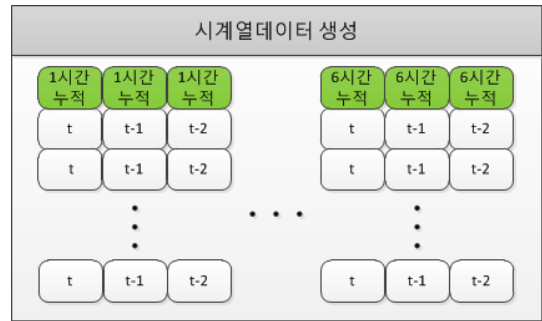


그림 4 KLAPS데이터를 이용한 시계열 데이터(t~t2)의 구성
Fig. 4 Construction of time series data(t to t-2) using KLAPS data

3. RBFNNs 이용한 초단기 호우주의보 예측 모델 설계

본 절에서는, 제공받은 기상관측자료에서 실험을 통하여 선정된 기온, 습도, 기압, 누적강수량(1시간~6시간)을 입력으로 하고, 출력은 1시간 후 6시간 누적강수량을 기준으로 하는 지능형알고리즘을 이용한 모델링에 대해 설명한다. 모델을 위해 제안한 FCM 기반 퍼지 모델의 최적화를 위해 멤버십 함수로서 효율적인 데이터 처리를 하기 위한 FCM 클러스터링 방법과, 제안한 모델의 성능향상을 위한 주요 입력인자 탐색 및 선택, 주요 파라미터 및 규칙, 후반부 다항식 형태의 최적화를 위한 입자군집화 알고리즘(Particle Swarm Optimization: PSO)을 적용한 지능형알고리즘에 대해 알아본다.

3.1 RBFNNs의 구조

지능형 알고리즘[2]으로는 RBFNNs[7]을 제안하였다. 일반 RBFNNs는 일반 역전파 학습규칙을 사용하는 다층퍼셉트론(MLP)에 비해 수렴속도가 빠르고 입력층, 1개의 은닉층, 출력층으로 구성된 네트워크로 단순하며 일반적인 NN(Neural Network) 구조에 기반을 둔다. RBFNNs 구조는 언어적 해석관점에서 식(1)과 같이 "If-then"의 퍼지 규칙[1]으로 표현되며 퍼지추론 메커니즘에 의해 구동된다.

$$\text{If } x \text{ is } A_i \text{ then } f_{ji}(x) \tag{1}$$

RBFNNs 구조는 RBF함수에 기반 한 네트워크 구조를 가지며, 조건부, 결론부, 추론부와 같이 세 가지 기능적 모듈로 분리되어 동작한다. RBFNNs의 구성은 입력데이터를 네트워크와 연결하는 입력층, 입력데이터를 RBFNNs에 의해 비선형 변환을 시키는 은닉층, 은닉층의 결과를 선형 조합하여 최종 출력값을 계산하는 출력층 등의 세 개의 다른 층으로 이루어져 있다. 은닉층의 노드수 즉, 커널함수인 RBF의 수는 보통 사용자가 결정하고 활성화함수는 정규분포형태의 방사형 기저함수를 사용한다.

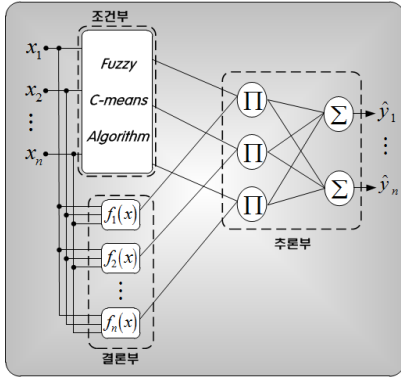


그림 5 방사형 기저함수 신경회로망의 전체 구조
Fig. 5 Overall architecture of Radial Basis Function Neural Networks

3.1.1 RBFNNs의 조건부

FCM 클러스터링을 사용하는 RBFNNs의 조건부 기능은 학습 데이터의 특성 반영을 위해 입력 공간을 c 개의 클러스터 수(퍼지 규칙 수)만큼 분리하고 각 영역의 소속정도를 퍼지 집합으로서 출력한다. FCM 클러스터 방법은 각 데이터와 특정 클러스터 중심과의 거리에 소속 정도를 부여하고 이 소속정도에 따라 데이터를 분류하는 알고리즘이다. FCM 클러스터링은 목적함수인 식(2)를 최소화함으로써 입력데이터의 각 클러스터에 대한 소속 값을 구한다.

$$Q = \sum_{i=1}^c \sum_{k=1}^N u_{ik}^m \| \mathbf{x}_k - \mathbf{v}_i \|^2 \quad (2)$$

여기서, c 는 클러스터의 개수(퍼지 규칙 수), N 은 입력 패턴 수, m 은 퍼지화 계수이다. x_k 는 k 번째 입력 벡터이고 v_i 는 i 번째 클러스터의 중심이다. u_{ik} 는 k 번째 데이터가 i 번째 클러스터에 속하는 소속정도를 나타내는 0과1사이의 실수로 식(3)과 식(4)의 조건을 만족한다.

$$\sum_{i=1}^c u_{ik} = 1 \quad 1 \leq k \leq N \quad (3)$$

$$0 < \sum_{k=1}^N u_{ik} < N \quad 1 \leq i \leq c \quad (4)$$

식(2)의 $\| \cdot \|$ 는 식(5)로 표현되는 가중 유클리디안 거리 (weighted Euclidean distance)를 사용한다.

$$\| \mathbf{x}_k - \mathbf{v}_i \|^2 = \sum_{j=1}^n \frac{(x_{kj} - v_{ij})^2}{\sigma_j^2} \quad (5)$$

여기서, σ_j 는 입력 패턴들의 j 번째 입력 차원의 표준편차이다. 가중 유클리디안 거리는 데이터 크기 분포에 크게 영향을 받지 않는 알맞은 거리 정보를 제공하므로 많이 사용되고 있다.

n 차원 유클리드 공간상의 N 개의 패턴으로 구성된 입력 벡터 집합 $X = \{x_1, x_2, \dots, x_N\}$, $x_k \in R_n$, $1 \leq k \leq N$ 과 클러스터 중심 $v = \{v_1, v_2, \dots, v_c\}$, $v_i \in R_n$, $1 \leq i \leq c$ 에 대한 소속정도를 소속행렬로 표현하면 $U = [u_{ik}]$ 로 표현되고 u_{ik} 와 V_i 는 식(6)과 식(7)에 의해 각각 계산된다.

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{\| \mathbf{x}_k - \mathbf{v}_i \|}{\| \mathbf{x}_k - \mathbf{v}_j \|} \right)^{\frac{2}{m-1}}} \quad 1 \leq k \leq N, 1 \leq i \leq c \quad (6)$$

$$\mathbf{v}_i = \frac{\sum_{k=1}^N u_{ik}^m \mathbf{x}_k}{\sum_{k=1}^N u_{ik}^m} \quad 1 \leq i \leq c \quad (7)$$

FCM 클러스터링은 식(6)과 식(7)을 반복적으로 수행하면서 소속행렬 U 와 각 클러스터의 중심 $V_i (i=1, \dots, c)$ 를 수정하며 식(2)의 목적함수 $Q(U, V_1, V_2, \dots, V_c)$ 를 특정 값으로 수렴시킨다.

3.1.2 RBFNNs의 결론부

RBFNNs 구조의 결론부는 조건부에서 분리한 각 로컬 영역을 다항식 함수의 로컬 회기모델로서 표현하여 식(1)의 “then” 이후의 규칙을 형성한다. 식(1)의 $f_{ji}(x)$ 에서 출력 뉴런에 대한 첨자 $j(1, \dots, s)$ 를 생략한 $f_i(x)$ 는 표3과 같은 형태를 갖는 네 가지 타입의 함수 중 하나의 형태를 갖는다. 즉, 로컬 회기 모델은 상수항, 일차식 또는 이차식, 그리고 입력 공간의 차원이 매우 클 경우 2차 항에 대해 입력변수들의 조합수가 증가하여 계산량이 크게 증가되는 것을 방지하기 위해 감소된 2차항수를 포함한 총 네 가지 타입이다.

표 3 다항식 타입

Table 3 Polynomial type

Type	Polynomial type
Constant	$f_i(x) = a_{i0}$
Linear	$f_i(x) = a_{i0} + \sum_{j=1}^n a_{ij} x_j$
Quadratic	$f_i(x) = a_{i0} + \sum_{j=1}^n a_{ij} x_j + \sum_{j=1}^n \sum_{k=1}^n a_{ijk} x_j x_k$
Reduced quadratic	$f_i(x) = a_{i0} + \sum_{j=1}^n a_{ij} x_j + \sum_{k=1}^n a_{ijk} x_k^2$

본 논문에서 제안한 RBFNNs는 기존의 상수항의 형태를 갖는 연결가중치를 1차식과 2차식으로 확장함으로써 은닉층 뉴런의 출력간의 비선형 결합으로 인해 분류기로서의 성능을 향상시킨다.

3.1.3 RBFNNs의 결론부

추론부에서는 네트워크의 최종출력을 구하게 된다. 그림 5에서

“ Π ”로 표기된 뉴런은 각 다항식을 의미하며, “ Σ ”으로 표기된 뉴런에 의해 입력신호들은 합하여져 그 결과를 출력층 뉴런의 최종 출력으로 내보낸다. 이와 같은 일련 과정은 퍼지추론 과정과 같으며, 퍼지 뉴럴 네트워크와 같은 형태를 갖게 된다. 결론적으로 RBFNNs 구조를 보인 그림 5 $j(1, \dots, s)$ 번째 출력의 최종출력은 퍼지 추론에 의한 식 (8)과 같이 표현된다.

$$y_j = g_j(x) = \frac{\sum_{i=1}^c u_i f_{ji}(x)}{\sum_{k=1}^c u_k} = \sum_{i=1}^c u_i f_{ji}(x) \quad (8)$$

규칙 후반부 다항식의 계수를 동정하기 위한 후반부 파라미터 동정은 주로 LSE를 사용하여 후반부 파라미터 계수를 동정하였는데, LSE는 각 규칙에 파라미터를 동시에 구하는 전역학습 방법이다. 이로 인해 퍼지 규칙수가 많거나, 입력이 많은 경우에는 찾고자 하는 파라미터들이 늘어나서 연산시간이 오래 걸리는 단점이 있다. 본 연구의 실험에서는 WLSE를 사용하여 각 규칙에 대한 파라미터를 독립적으로 구하는 지역학습 방법을 적용하였다. 이로 인해 각 규칙에 대해 해석 할 수 있으며, 계산부하가 줄어들어 연산시간 또한 단축시킬 수 있다.

또한 모델의 성능평가를 위해 성능지수 식을 식(9)의 Mean Square Error(MSE)의 방법을 이용하여 실행하였다.

$$E_{MSE} = \frac{1}{N} \sum_{t=1}^N (y(x_t) - \hat{y}(x_t))^2 \quad (9)$$

여기서, $y(x)$ 는 주어진 실제 출력 데이터이고, $\hat{y}(x)$ 는 모델의 출력을 의미한다. 위 성능지수는 RBFNN의 모델링 기법에서 사용되는 성능지수이다. RBFNN패턴 분류기의 성능지수는 일반적으로 패턴분류율을 따르며 패턴분류를 위한 판별함수의 생성이 필요하다. 패턴분류의 문제는 2부류 분류문제(Two-class classification problem)와 다 부류 분류문제(Multi-class classification problem)로 나눌 수 있다. 분류기를 표현하는 여러 방법이 있지만 일반적으로 많이 사용하는 방법으로는 2부류 분류 문제에 대해 식(10)을 이용하고 다 부류 분류문제에 대해 식(11)과 같은 규칙을 이용한다.

$$w_1 \text{으로 판정 if } g(\mathbf{x}) > 0; w_2 \text{로 판정 otherwise} \quad (10)$$

$$w_i \text{으로 판정 if } g_i(\mathbf{x}) > g_j(\mathbf{x}) \text{ for all } i \neq j \quad (11)$$

여기서 $g_j(\mathbf{x})$, $j = 1, \dots, m$ 은 j 번째 클래스에 대한 판별함수이고 m 은 클래스 수이다. w_j 는 j 번째 클래스를 나타낸다. 2부류 분류문제는 출력층에 1개의 출력뉴런이 존재하고, 식(10)과 같이 1개의 판별 함수만을 사용하여 해당 클래스로 패턴을 분류한다. 다 부류 분류문제인 경우 m 개의 출력 뉴런이 존재하며, 식(11)와 같은 m 개의 판별함수를 생성하고 해당 클래스로 분류한다.

3.2 PSO 최적화 알고리즘

Particle Swarm Optimization(PSO)[4]는 Kennedy and

Eberhart에 의해 처음 제안되었으며, 새 떼와 물고기 떼와 같은 생체군집의 사회적 행동양식[5]을 바탕으로 한 최적화 알고리즘이다. 일반적으로 PSO는 이론의 간결성, 구현의 용이성, 연산의 효율성과 같은 특징이 있어 짧은 계산시간 안에 최적의 해를 생성할 수 있고, 다른 확률적 방법보다 안정적인 수렴특징을 갖는다. PSO 알고리즘에서 swarm은 m 개의 particle을 가지며 이들은 주어진 문제의 해가 된다. 각 particle은 n 차원 실수 벡터로 구성된다.

3.3 차분진화(DE) 알고리즘을 이용한 최적화

차분진화 알고리즘(Differential Evolution; DE)은 Price와 Storn에 의해 벡터 차분(Vector differential)을 사용하여 Chebychev 다항곡선의 내삽문제(Polynomial fitting problem)를 해결하는 과정에서 개발되었다. 차분진화 알고리즘은 현재 널리 사용되고 있는 통계적 임의 탐색법인 유전 알고리즘(Genetic Algorithm: GA)[11]과 유사한 알고리즘이다. DE와 GA는 초기 개체군을 이루는 개체를 샘플링하고 교배(Crossover), 변이(Mutation), 선택(Selection) 과정을 거쳐 새로운 개체들을 추출해 낸다는 공통점이 있다. 그러나 GA는 연산 과정에서 개체들의 표현형을 유전형으로 바꾸는 과정이 필요한 반면, DE는 코딩 과정이 필요 없으며 개체를 벡터로 표현하고 산술적 연산을 통해 새로운 개체를 생성한다는 이점이 있다. 또한 GA의 단점이었던 복잡한 구조와 연산과정에 비하여 간단하고 합리적인 구조와 연산 속도가 빠르다는 장점이 있다.

3.4 PSO 및 DE에 의한 RBFNNs의 최적화

본 절에서는 제안된 RBFNNs 모델의 성능 향상에 주요한 역할을 하는 중요 파라미터인 클러스터 수, FCM 클러스터링 방법에서의 퍼지화 계수, 그리고 규칙 후반부 다항식의 형태 및 입력인자를 최적화하기 위하여 점목된 입자 군집 최적화(Particle Swarm Optimization; PSO)알고리즘과 차분진화 최적화(Differential Evolution; DE)[12] 알고리즘의 설계 방법에 대해 설명한다.

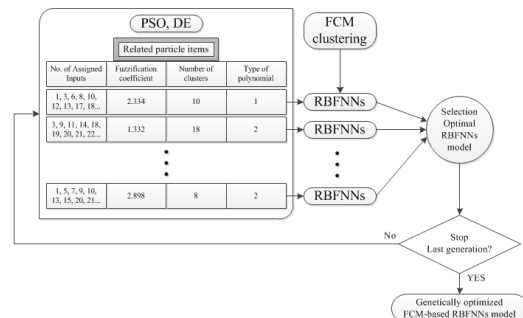


그림 6 PSO/DE알고리즘과 FCM클러스터링 방법에 의해 실현된 RBFNNs의 전체적인 흐름

Fig. 6 Overall flow for the design of RBFNNs realized with the aid of PSO/DE algorithm and FCM clustering method

그림 6는 PSO 및 DE 최적화 알고리즘에서 각각의 particle에 의해서 선택된 파라미터들과 각 진화론적 최적화 알고리즘을 통해 최적화된 RBFNNs[8][9][10]의 구축을 나타낸다.

4. 초단기 호우 예측 시뮬레이션 및 실험 결과고찰

초단기 호우 예측 시스템에서 입력으로 사용되는 기상 데이터는 기상연구소 예보연구과로부터 제공받아 사용하였다. 먼저 과거 기상 데이터를 이용하여 초단기 호우 예측 모델 및 분류기를 구축한다. 이 후 실시간 기상 데이터를 입력으로 받아 실시간으로 호우특보를 예측하는 실시간 호우 예측 모델 및 분류기를 구축한다. 그림 7은 전체적인 호우 예측 시스템의 개념도를 나타낸 것이다.

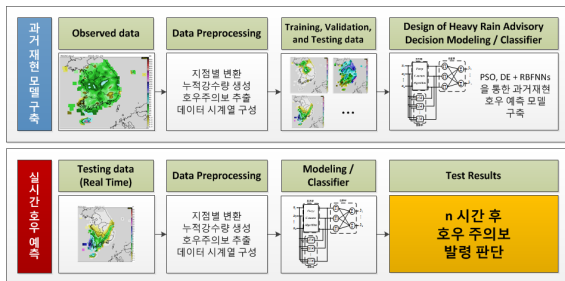


그림 7 RBFNNs 알고리즘에 의해 실현된 예측 모델의 구축
Fig. 7 Construction of predictive model realized by RBFNNs algorithm

4.1 호우 주의보 평가 척도

강수예측능력 ETS(Equitable Threat Score)은 기상 이벤트에 적합한 강수예측능력 측정 방법이다. 이 방법은 예측치의 적중 여부와 관측치의 적중 여부를 네 가지 유형으로 비교하여 실수 값을 도출한다. ETS 수치를 구하는 방법은 표 4의 강수예측 분할 표를 기준으로 a, b, c, d 각각의 기상 이벤트 수를 구한 후에 식(12)을 적용한다.

표 4 각 형태별 야간 감시 장비의 비교
Table 4 Comparison of each type of night vision equipment

		Observation	
		Positive	Negative
Forecast	True	a(hits)	b(false alarm)
	False	c(misses)	d(non-events)

$$ETS = \frac{a - a_r}{a + b + c - a_r}, \quad a_r = \frac{(a+b)(a+c)}{a+b+c+d} \quad (12)$$

ETS 지표는 완벽한 예측일 때, 1의 값을 갖는다. 본 논문의 연구에서는 ETS를 주 지표로 사용하여 호우특보 예측 평가를 수행한다. 이 외에도 FBI(Frequency Bias Index)은 호우특보라고 예측한 것과 관측된 것의 비를 의미한다. FBI지수가 1보다 높은

경우 over-forecast 1보다 낮으면 under-forecast를 의미한다. POD(probability of detection)은 호우특보를 검출할 확률을 의미하며, PC(Proportion Correct), PAG(Post Agreement)와 같은 보조적인 평가 척도들을 사용하였다. 평가척도 ETS는 두 가지 측면으로 고려하여 ETS1(호우주의보를 연속시간의 이벤트로 적용)과 ETS2(호우주의보를 각 시간대별 이벤트로 적용)로 구하였다.

4.2 KLAPS 기상데이터를 이용한 호우주의보 예측 실험

제안된 모델 및 분류기의 성능평가를 위해 KLAPS 기상데이터를 사용하였으며 실험 지역은 2007년부터 2011년 까지 정부가 선포한 특별재난지역을 선정하여 실험하였다. 데이터는 Training, Validation, Testing으로 나누어 구축하였다. 이는 Supervised learning을 통해 만들어진 시스템이 Training data에 의해서는 출력이 잘 나오지만 새로운 Testing 데이터에서는 강인하게 대처할 수 없는 Overfitting이 나타나게 된다. 이를 대처하기 위해서 Validation개념을 추가하여 문제점을 개선하였다. 표 5는 실험에 사용된 KLAPS 데이터의 구축 정보이다.

표 5 실험에 사용된 지점의 구축 정보
Table 5 Point information was used in the experiment

Area	Partitioning of data	Year	No. of grid points	No. of data
특별재난지역	Training	06~08	8	19040
	Validation	08~09	8	9290
	Testing	10~11	56	19367

제안된 모델 및 분류기의 성능은 입력변수, 클러스터의 수, 퍼지화 계수, 다항식의 형태와 같은 파라미터에 의해 영향을 받는다. 따라서 이러한 파라미터들의 최적의 값을 찾고자 PSO, DE 최적화 알고리즘들을 사용하였다. 표 6는 최적화 알고리즘의 파라미터 설정 값을 보여주고, 표 7는 각 최적화 알고리즘의 파라미터 Vector 및 Particle과 탐색 범위를 보여준다.

표 6 최적화 알고리즘의 파라미터 설정 값
Table 6 Summary of the parameters in the optimization environment

Optimization Algorithm	Parameters	Values
PSO	No. of generations	50
	Swarm size	100
	Acceleration value	2.0
	Inertial weight	[0.4 0.9]
	v_{max}	탐색공간의 20%
DE	No. of generations	50
	No. of populations	100
	Crossover rate	1
	Scaling factor	0.5

표 7 모델 및 분류기의 최적화 파라미터 탐색 범위

Table 7 Search range for model and classifier optimization parameters

Parameters to be optimized	Range	
	Classifier	Modeling
No. of Assigned inputs	[15 22]	[15 22]
Polynomial Type	[Type1 Type2]	[Type1 Type2]
No. of clusters	[1 5]	[1 20]
Fuzzification coefficient	[1.1 3]	[1.1 3]

표 8 각 최적화 기반 분류기의 성능

Table 8 Output performance for each optimization-based classifier

Classifier	예측시간	PSO-based Classifier		DE-based Classifier	
		ETS1	ETS2	ETS1	ETS2
	1시간	0.32	0.77	0.32	0.74
2시간	0.15	0.68	0.15	0.65	

표 9 각 최적화 기반 모델의 성능

Table 9 Output performance for each optimization-based model

Modeling	예측시간	PSO-based Classifier		DE-based Classifier	
		ETS1	ETS2	ETS1	ETS2
	1시간	0.32	0.76	0.31	0.74
2시간	0.16	0.63	0.15	0.65	

표 8과 표 9를 비교해 보면 PSO 기반 분류기[13] 및 모델의 강수예측능력이 근소하게나마 우수함을 알 수 있다. ETS1와 ETS2의 성능 차이는 실험에 앞서 설명한바와 같이 모든 호우주의보의 적중률을 판단하는 ETS2가 ETS1에 비하여 높아 보일 수 있다. 하지만 호우주의보 예측 시 중요한 시작지점의 적중률을 판단하는 ETS1의 성능지수가 매우 중요하다. 다항식 방사형 기저함수 신경회로망의 학습은 가장 최소자승법을 사용하였으며, 제안된 모델 및 분류기는 각 진화 최적화 알고리즘 PSO, DE를 사용하여 각 모델 및 분류기의 최적의 파라미터를 동조하였다.

4. 결 론

본 논문에서는 KLAPS 기상데이터에서 특별재난지역의 2006~2011년 기간의 데이터를 사용하여 다항식 방사형 기저 함수 신경회로망 모델 및 분류기를 설계하고 호우주의보를 예측하였다. 보다 정확한 호우주의보 예측을 위해 KLAPS 기상데이터 처리방법과 지능형 호우주의보 예측 모델 및 분류기의 성능 향상 방법에 대하여 연구하였다. 호우주의보 추출방법을 통한 데이터의 호우주의보 집중도를 높였으며 시계열 방식의 데이터 처리 방법으로 시간의 변화에 따른 데이터의 특성을 잘 나타내어 다항식 방사형 기저함수 신경회로망의 입력으로 사용하였으며 진화론적 최

적화 알고리즘인 PSO 및 DE를 적용하여 모델 및 분류기의 최적의 파라미터를 동조하였으며 ETS 성능을 비교하였다.

감사의 글

This work was supported by GRRC program of Gyeonggi province [GRRC Suwon 2014-B2, Center for U-city Security & Surveillance Technology] and supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology (NRF-2012R1A1B3003568)

References

- [1] L.A Zadeh, Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, *Fuzzy Sets Syst.* 90 (1997) 111 - 117.
- [2] Y. Qian, J. Linag, Combination entropy and combination granulation in rough set theory, *Int. J. Uncertain Fuzzy Knowledge-Based Syst.* 16 (2) (2008) 179 - 183.
- [3] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The Fuzzy C-Means Clustering Algorithm," *Computers & Geoscience*, Vol. 10, No. 2-3, pp. 191-203, 1984.
- [4] J. Kennedy, R. Eberhart, Particle Swarm Optimization, in: *Proceedings of the IEEE International Conference on Neural Networks*, 4 (1995) 1942 - 1948.
- [5] J. Kennedy, The Particle Swarm: Social adaptation of Knowledge, in: *Proceedings of the IEEE International Conference on Evolutionary Computation*, 1997, 303 - 308.
- [6] S. K. Oh, W. D. Kim, W. Pedrycz, and B. J. Park, "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.
- [7] S. K. Oh, W. Pedrycz and H. S. Park, "Hybrid Identification in Fuzzy-Neural Networks", *Fuzzy Sets and Systems*, Vol. 138, Issue 2, pp.399-426, 2003.
- [8] S. K. Oh, W. Pedrycz, and K. J. Park, "Structural Developments of Fuzzy Systems with The Aid of Information Granulation," *Simulation ModellingPractice and Theory*, Vol. 15, No. 10, pp. 1292-1309, 2007.
- [9] W. Shen, X. Guo, C. Wu, D. Wu, Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm, *Knowledge-Based Syst.* 24 (3) (2011) 378 - 385.
- [10] J. Li, X. Liu, Melt index prediction by RBF neural network optimized with an MPSO-SA hybrid algorithm,

Neurocomputing 74 (5) (2011) 735 - 740.

- [11] Michalewicz, Z, "Genetic Algorithms + Data Structures = Evolution Programs", Springer-Verlag, Berlin Heidelberg, 1996.
- [12] R. Storn, Differential Evolution, A Simple and Efficient Heuristic Strategy for Global Optimization over Continuous Spaces, Journal of Global Optimization, Vol. 11 pp.341-359, 1997.
- [13] R.O. Duda, P.E. Hart, D.G. Stork, "Pattern Classification," 2nd ed., Wiley-Interscience, 2000.



김 현 기(Kim, Hyun-Ki)

1977년 연세대학교 전기공학과 졸업, 동 대학원 석사(1985), 박사(1991). 1997~현재 경기지역 산학연 기술지원단 단장, 1998~현재 산업기술정책연구소 심의위원, 2000~현재 사단법인 산학연 컨소시엄 전국협의회 회장, 1989~현재 수원대학교 전기공학과 교수, 관심분야는 시스템 자동화 및 감시제어, 지능형 모델링 및 제어등.

Tel : 031-220-2666

Fax : 031-220-2667

E-mail : hkkim@suwon.ac.kr

저 자 소 개



김 현 명(Hyun-Myung Kim)

2012년 수원대학교 전기공학과 졸업, 동대학원 석사(2014), 관심분야는 퍼지추론 시스템, 뉴럴 네트워크, Granular Computing 등

Tel : 031-222-6544

E-mail : khm@suwon.ac.kr



오 성 권(Sung-Kwon Oh)

1981년 연세대학교 전기공학과 졸업, 동 대학원 석사(1983), 박사(1993). 1983-1989년 금성산전연구소(선임연구원). 1996-1997년 캐나다 Manitoba 대학 전기 및 컴퓨터공학과 Post-Doc. 1993-2004년 원광대학교 전기전자 및 정보공학부 교수. 2005~현재 수원대학교 전기공학과 교수, 2002~현재 대한전기학회, 제어로봇시스템학회, 퍼지및지능시스템학회 편집위원. 2012~현재 Information Sciences 편집위원. 관심분야는 퍼지 시스템, 퍼지-뉴럴 네트워크, 자동화 시스템, 고급 computational intelligence, 지능 제어 등.

Tel : 031-229-8162, Fax : 031-220-2667

E-mail : ohsk@suwon.ac.kr