

최적화된 다항식 방사형 기저함수 신경회로망을 이용한 수도권 여름철 초단기 강수예측 패턴 설계

Design of Summer Very Short-term Precipitation Forecasting Pattern in Metropolitan Area Using Optimized RBFNNs

김현기 · 최우용 · 오성권[†]

Hyun-Ki Kim , Woo-Yong Choi, and Sung-Kwun Oh[†]

수원대학교 공과대학 전기공학과

[†] Department of Electrical Engineering, The University of Suwon

요 약

최근 빈번히 일어나는 국지성 집중호우로 인해 피해가 급격히 증가하고 있다. 인구가 밀집한 수도권과 같은 경우 산사태와 토석류 및 홍수로 인해 인명 및 재산피해가 심각하다. 따라서 집중호우에 대한 예측의 중요성이 증가하고 있다. 우리나라 약천후 강수의 특징으로는 태풍과 집중호우로 구분된다. 이는 지속시간과 지역에 따라 차이를 보인다. 또한, 지역적인 강수는 계절에 따라 변동성이 크고 비선형적이기 때문에 강수를 예측하는데 어려움이 따른다. 본 논문에서는 기상청에서 현업으로 사용하는 초단기 기상 분석 및 예측시스템 (Korea Local Analysis and Prediction System; KLAPS)의 기상 관측 자료를 이용하여 초단기 호우 예측 패턴 모델을 구현한다. 그리고 약천후 시 피해가 큰 수도권을 중심으로 여름철 호우 특보를 예측한다. 유전자 알고리즘(Genetic Algorithm; GA) 기반 다항식 방사형 기저함수 신경회로망(Radial Basis Function Neural Networks; RBFNNs)을 이용하여 초단기 강수 예측 패턴 모델을 설계한다. 최적화된 분류기를 설계하기 위하여 유전자 알고리즘을 이용하여 주요 파라미터인 입력변수의 수, 다항식 차수, 퍼지화 계수, FCM(Fuzzy C-mean) 클러스터 수를 동조한다.

키워드 : 다항식 방사형 기저함수 신경회로망, 유전자 알고리즘, FCM 클러스터링 알고리즘

Abstract

The damage caused by Recent frequently occurring locality torrential rains is increasing rapidly. In case of densely populated metropolitan area, casualties and property damage is a serious due to landslides and debris flows and floods. Therefore, the importance of predictions about the torrential is increasing. Precipitation characteristic of the bad weather in Korea is divided into typhoons and torrential rains. This seems to vary depending on the duration and area. Rainfall is difficult to predict because regional precipitation is large volatility and nonlinear. In this paper, Very short-term precipitation forecasting pattern model is implemented using KLAPS data used by Korea Meteorological Administration. we designed very short term precipitation forecasting pattern model using GA-based RBFNNs. the structural and parametric values such as the number of Inputs, polynomial type, number of fcm cluster, and fuzzification coefficient are optimized by GA optimization algorithm.

Key Words : Radial Basis Function Neural Networks, Genetic Algorithm, Fuzzy C-means Algorithm

접수일자: 2013년 3월 31일

심사(수정)일자: 2013년 4월 7일

게재확정일자 : 2013년 9월 5일

[†] Corresponding author

이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(NRF-2012-003568) 그리고 경기도의 경기도지역협력연구센터사업의 일환으로 수행하였음[GRRC 수원 2013-B2, U-city 보안감시 기술협력센터].

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.

1. 서 론

지구온난화가 진행됨에 따라 세계적으로 대기에 비구름이 형성되기 쉬워지고 대기가 불안정해져 국지성 폭우 가능성이 높아져가고 있다. 국지성 폭우로 인한 인명 및 재산피해는 해가 지날수록 증가하고 있다. 기상 및 기후 변화로 인한 피해를 사전에 감지하기 위한 초단기 강수 예측의 필요성은 증가하고 있다.

본 논문에서는 집중호우 발생 강수패턴의 변화로 인구밀도가 가장 높아 피해확률이 높은 수도권과 강수량이 제일 많은 계절인 여름철을 선택하였다. 지역적인 강수는 계절에 따라 변동성이 크고 비선형적이기 때문에 예측하기 어렵다. 이러한 강수 예측을 해결하기 위해 실제 기상 데이터가 이용하여 지능형 알고리즘인 다항식 방사형 기저함수 신경회

로망에 적용하는 초단기 강수 예측 패턴 모델을 제안한다. 이 다항식 방사형 기저함수 신경회로망은 고차원의 입-출력 형태를 갖는 클래스를 다루는데 유용하고 구조가 심플하여 데이터 전처리 및 클러스터링 방법과 최적화 알고리즘과 같이 서로 다른 분야에서 연구되고 있는 알고리즘과의 결합이 용이하다. 진화 최적화 알고리즘인 유전자 알고리즘을 이용하여 다항식 방사형 기저함수 신경회로망 성능의 주요 역할을 하는 파라미터를 동조한다.

2. 다항식 방사형 기저함수 신경회로망

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

2.1 다항식 방사형 기저 함수 신경회로망 구조

방사형 기저함수 신경회로망은 입력층과 은닉층, 출력층으로 구성되어 있고 조건부와 결론부, 추론부의 세 가지 기능적 모듈로 분리되어 동작한다. 조건부의 활성화 함수로는 가우시안 함수를 사용하고 결론부에는 상수항을 사용하며 최종출력은 조건부와 결론부의 상수항으로 표현된다. 그림 1은 기본적인 방사형 기저함수 신경회로망의 구조이다.

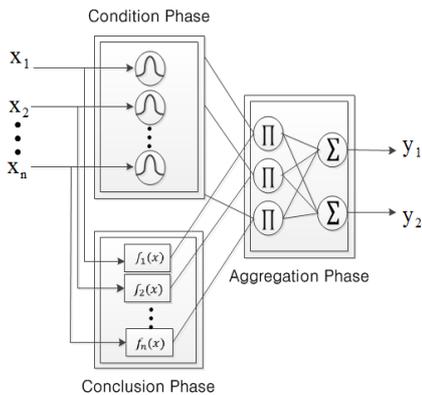


그림 1. 방사형 기저 함수 신경회로망
Fig. 1. Radial Basis Function Neural Networks

본 논문에서 제안된 다항식 방사형 기저함수 신경회로망[1,2,3,4]의 구조로는 조건부의 활성화 함수로 Fuzzy C-Mean 클러스터링 알고리즘을 사용하여 각 클러스터에 포함되는 데이터의 소속정도가 퍼지집합으로 출력되어 활성화 함수로 이용된다. 결론부는 표 1과 같이 1차식, 2차식, 변형된 2차식 3가지 중 하나의 형태를 갖는다. 결론부 파라미터 동정에는 표준 최소자승법(standard Least Squares Estimation; LSE)을 사용했다. 추론부는 네트워크의 최종출력으로서 조건부의 소속 함수(활성 함수)에 의해 활성화된 다항식 함수들을 연산을 통하여 출력한다. 그림 2는 다항식 방사형 기저함수 신경회로망의 구조이다.

FCM 클러스터링에 기반한 분할 함수를 활성화함수로

사용하며, 다항식 함수로 구성된 연결가중치를 사용함으로써 식 (1)의 퍼지 규칙 표현과 같이 해석될 수 있다.

$$\text{If } x \text{ is } A_i \text{ Then } f_{ji}(x) \quad (1)$$

x는 입력 벡터, A_i 는 FCM 클러스터링에 의한 $i(i=1, \dots, c)$ 번째 그룹의 소속 함수, $f_{ji}(x)$ 는 $j(j=1, \dots, c)$ 번째 출력에 대한 i 번째 퍼지 규칙의 다항식이다.

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

표 1. 다항식 유형
Table 1. Polynomial Type

Type	Polynomial Type
Linear	$y = a_0 + a_1x_1 + a_2x_2$
Quadratic	$y = a_0 + a_1x_1 + a_2x_2 + a_3x_1^2 + a_4x_2^2 + a_5x_1x_2$
Modified Quadratic	$y = a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2$

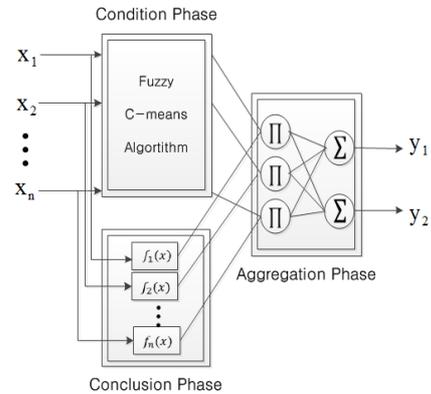


그림 2. 제안된 방사형 기저 함수 신경회로망
Fig. 2. Proposed Radial Basis Function Neural Networks

2.2 FCM 클러스터링 알고리즘

클러스터링은 데이터의 분류를 위해 사용되는 것으로 데이터에 형태나 속성 등의 기준으로 데이터를 분류하여 개체의 소속집단을 정하는 것이다. FCM 클러스터링[5,6,7,8]은 각 데이터의 거리를 기준으로 소속정도를 측정하여 데이터를 분류하는 방법이다. 다음은 FCM 클러스터링의 과정이다.

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

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

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

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

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{\|x_k - v_i^{(r)}\|}{\|x_k - v_j^{(r)}\|} \right)^{2/(m-1)}} \quad (4)$$

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

$$d_{ik} = d(x_k - v_i^{(r)}) = \left[\sum_{j=1}^1 (x_{kj} - v_{ij}^{(r)})^2 \right]^{1/2} \quad (5)$$

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

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

FCM 클러스터링은 데이터의 특성을 보다 더 잘 반영할 수 있다. FCM 클러스터링의 수는 은닉층 노드의 수를 대신하고 소속행렬의 값은 가우시안 활성화함수에 의한 적합도 값과 같으며 FCM의 퍼지화 계수는 가우시안 함수의 분포상수를 나타낸다.

2.3 최소자승법

다항식 연결가중치의 파라미터 계수를 동정하기 위해 최소자승법을 사용하여 계수를 추정하였다. 최소자승법의 성능 평가함수를 행렬식으로 표현하면 다음 식 (7)과 같다.

$$Q = \sum_{i=1}^c (Y - X_i a_i)^T (Y - X_i a_i) \quad (7)$$

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

$$X_i = \begin{bmatrix} 1 & x_{11} & \dots & x_{1m} \\ 1 & x_{21} & \dots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nm} \end{bmatrix} \quad (8)$$

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

$$a_i = (X_i^T X_i)^{-1} X_i^T Y \quad (9)$$

3. 진화 최적화 알고리즘

3.1 유전자 알고리즘

유전자 알고리즘[9,10,11]은 1975년 John Holland에 의해 개발 되었다. 생물학적 진화, 자연선택, 유전자 재조합의 생물학 원리에 바탕을 둔 확률적 탐색 방법으로, 잠재적으로 좋은 결과를 가져다줄 과정을 추적한다. 그리고 지역 극소에 빠질 가능성을 제거하기 위해 임의의 지점을 검사함으로써, 공간 전체를 탐색할 수 있는 능력을 가진다. 유전자 알고리즘의 특징은 주어진 값을 2진 코딩하여 코딩된 문자열을 하나의 개체로 그리고 각각의 비트를 유전자처럼 취급하며, 하나의 지점부터 최적의 지점으로가 아닌 한번에 여러 지점에서 최적의 지점을 찾아 나간다. 그리고 미분과 같은 수학적 연산이 아닌 결과의 적합도인 목적함수를 기준으로 수행되며, 결정적인 방법이 아닌 확률적인 방법이다. 유전자 알고리즘은 3개의 기본 연산인 생산, 교배, 돌연변이를 가지고 있다. 생산은 다음 세대를 위해 교배 유전자가 선택되어지는 과정으로, 적합도에 따라 다음세대를 생산할 수 있는 기회를 차등적으로 갖는다. 각각의 문자열들은 그들의 적합도를 가지고 다음세대에 전달된다. 교배는 두 단계의 과정을 가지는데 첫 번째 단계는 교배 유전자로부터 염색체들은 짝짓는다. 두 번째 단계로 짝지어진 문자열 쌍은 문자열의 k위치가 [1, l-1] 사이에서 확률적으로 균일하게 선택되어진다. 여기서 l은 문자열의 길이이다. 선택된 k와 l사이의 모든 특성들이 교차함으로써 두 개의 새로운 문자열이 생성되는 과정으로 교배가 이루어진다. 돌연변이는 문자열 위치의 값을 임의로 변경한다. 2진 코딩에서 돌연변이는 0을 1로 또는 1을 0으로 바꾸는 것을 의미한다. 적절하게 정의된 적합도 함수와 함께 조합된 연산자는 유전자 알고리즘의 주요 부분이 되며 만들어진 문자열의 군집을 가지고 평가, 선택, 재생산을 통해 반복 수행된다.

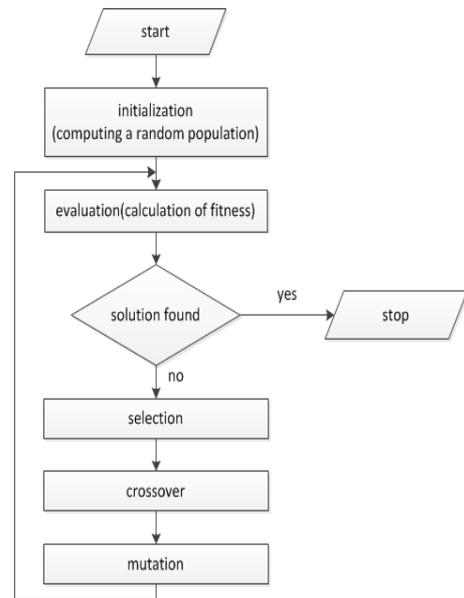


그림 3. 유전자 알고리즘의 순서도
Fig. 3. Flow Chart of Genetic Algorithm

본 논문에서는 선택연산자는 roulette-wheel 방식, 돌연변이연산자는 invert 방식, 교차연산자는 one-crossover 방식을 사용하였다. 제안된 다항식 방사형 기저함수 신경회로망에서 입력변수의 수와 퍼지화 계수, 다항식 차수, FCM 클러스터의 수를 최적화 알고리즘인 유전자 알고리즘을 통해 탐색하여 구조의 최적화를 효율적으로 구축하게 한다. 유전자 알고리즘의 순서도는 그림 3과 같다.

3.2 유전자 알고리즘을 이용한 최적화 과정

제안된 분류기의 성능은 입력변수의 수, 다항식 차수, FCM 클러스터링 알고리즘의 퍼지화 계수와 같은 파라미터에 의해 영향을 받는다. 따라서 이러한 파라미터들의 최적의 값을 찾고자 유전자 알고리즘을 사용하였다. 최적화 파라미터를 찾기 위한 제안된 분류기의 최적화 과정은 다음과 같다.

- 단계 1: 출력변수 y에 관계되는 n개의 시스템 입력변수를 결정.
- 단계 2: 전체 입력력 데이터를 training, validation, testing 데이터로 분할.
- 단계 3: RBFNNs에서 사용할 초기 파라미터를 랜덤하게 선택(퍼지화 계수, 다항식 차수, 입력 개수, 클러스터 수) 및 적합도 평가.
- 단계 4: 유전자 알고리즘을 이용한 RBFNNs 파라미터의 입력변수의 수, 다항식 차수, 퍼지화 계수, FCM 클러스터의 수를 선택.
- 단계 5: training 데이터를 통해 패턴분류기를 최적 모델을 구축하고 validation 데이터로 검증한 후 testing 데이터로 n 시간 후 강수예측을 통한 강수 패턴 분류를 통하여 성능지수를 평가.

4. 시뮬레이션 및 결과

본 논문에서는 유전자 알고리즘 기반 다항식 방사형 기저함수 신경회로망 패턴모델을 이용하여 초단기 강수 예측 시스템을 설계한다. 선택된 지역은 약천후 시 인구 밀도가 높아 집중호우에 의한 재산 및 인명피해가 높은 수도권 지역을 선정하여 실험을 진행한다. 사용된 기상 데이터는 기상청에서 현업으로 사용하고 있는 KLAPS 시스템의 데이터를 이용하였고, 연도는 2006년 ~ 2011년이며 우리나라의 여름철로 1년 동안 강수량이 가장 많은 6월 ~ 8월의 데이터를 선택한 후 데이터를 추출하여 사용하였다. 데이터의 구성은 다음 표 2와 같다.

표 2. KLAPS 데이터 구성
Table 2. Composition of KLAPS Data

	Year	Month
Training Data	2006 ~ 2008	6 ~ 8
Validation Data	2008 ~ 2009	6 ~ 8
Testing Data	2010 ~ 2011	6 ~ 8

데이터의 입력인자 구성으로는 기압, 습도, 기온, 1시간 ~ 6시간 누적강수량을 시계열로 확장하여 구성하였다.

제안된 분류기의 성능은 입력인자 개수, 다항식 차수, 퍼지화 계수, FCM 클러스터 수에 의해 영향을 받는다. 따라서 파라미터들의 최적의 값을 찾기 위해 유전자알고리즘을 이용하여 최적화한다. 다음 표 3과 표 4는 설계된 모델에서 유전자 알고리즘의 초기 설정 파라미터 정보와 다항식방사형 기저함수 신경회로망에서 조절하고자 하는 초기 파라미터 탐색 범위를 보여준다.

표 3. 유전자 알고리즘 설정 파라미터
Table 3. Set-up Parameters of Genetic Algorithm

GA	No. of Generations	60
	No. of Populations	30
	Mutation Rate	0.01
	Crossover Rate	0.65
	Feature Selection	0.5 이상
	String Length	130

표 4. RBFNNs의 파라미터 탐색 범위
Table 4. Parameters Search Space of Radial Basis Function Neural Networks

RBFNNs	No. of Inputs	[10 22]
	No. of Clusters	[2 5]
	Fuzzification Coefficient	[1.1 3]
	Polynomial Type	Type1. Constant Type2. Linear Type3. Quadratic Type4. Modified Quadratic

최적의 파라미터일 때 모델의 6시간 누적 예측 강수량을 출력하여 6시간 누적 실제 강수량과 예측 강수량을 패턴적으로 분석하고 비교하여 강수예측능력인 ETS(Equitable Threat Score)지수를 성능평가로 사용하여 제안된 패턴분류기의 성능지수를 확인했다. 호우 판별기준으로는 강수시간에 관계없이 누적 강수량이 70mm인 것을 기준으로 하였다. 6시간 누적 실제 강수량이 호우 주의보인 70mm를 넘었을 때 6시간 누적 예측 강수량이 70mm를 넘는다면 Hit(a), 아니라면 Miss(c)로 판별한다. 6시간 누적 실제 강수량이 70mm를 넘지 않고 6시간 누적 예측 강수량이 70mm를 넘었다면 False Alarm(b)으로, 아니라면 Correct Non-Event(d)로 판별한다. 강수예측 분할 표는 다음 표 5와 같다.

pp. 54-77, 2011.

[5] 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.

[6] A. Aiyer, K. Pyun, Y. Z. Huang, D. B. O'Brien, and R. M. Gray, "Lloyd clustering of Gauss mixture models for image compression and classification," *Signal Processing: Image Communication*, vol. 20, pp. 459-485, 2005.

[7] H.-S. Park, W. Pedrycz and S.-K. Oh, "Granular Neural Networks and Their Development Through Context-Based Clustering and Adjustable Dimensionality of Receptive Fields", *IEEE Trans. on Neural Networks*, vol. 20, no. 10, pp. 1604-1616, 2009.

[8] A. Staiano, J. Tagliaferri, W. Pedrycz, "Improving RBF networks performance in regression tasks by means of a supervised fuzzy clustering" Automatic structure and parameter," *Neurocomputing*, Vol. 69, pp. 1570-1581, 2006

[9] Z. Michalewicz, *Genetic Algorithms+Data Structure =Evolution Programs*, Springer-Verlag, 1992.

[10] D. E. Goldberg, *Genetic Algorithm in search, Optimization & Machine Learning*, Addison wesley, 1989.

[11] Holland, J. H., *Adaptation In Natural and Artificial Systems*. The University of Michigan Press, Ann Arbour. 1975.



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(선임 연구원)
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

저 자 소 개



김현기(Hyun-Ki Kim)

1977년 : 연세대학교 전기공학과 공학사
1985년 : 연세대학교 전기공학과 공학석사
1991년 : 연세대학교 전기공학과 공학박사

1997년~2010년 : 경기지역 산학연 기술지원단 단장
1998년~2000년 : 사업기술정책연구소 심의위원
2000년~2004년 : 사단법인 산학연 컨소시엄 적국협의회 회장
1989년~현재 : 수원대학교 전기공학과 교수

관심분야 : 자동화 시스템 및 감시제어, 지능형 모델링 및 제어 등.
Phone : +82-31-220-2666
E-mail : hkkim@suwon.ac.kr