



딥러닝 기법을 이용한 내일강수 예측

Forecasting the Precipitation of the Next Day Using Deep Learning

하지훈* · 이용희** · 김용혁***†

Ji-Hun Ha, Yong Hee Lee, and Yong-Hyuk Kim†

*광운대학교 임베디드SW공학과, **국립기상과학원, ***광운대학교 컴퓨터공학과

*Department of Embedded Software Engineering, Kwangwoon University

**National Institute of Meteorological Science

***Department of Computer Science and Engineering, Kwangwoon University

요약

정확한 강수예측을 위해서는 예측인자 선정과 예측방법에 대한 선택이 매우 중요하다. 최근에는 강수예측 방법으로 기계학습 기법이 많이 사용되고 있으며, 그 중에서도 특히 인공신경망을 사용한 강수예측 방법은 좋은 성능을 보였다. 본 논문에서는 딥러닝 기법 중 하나인 DBN(deep belief network)를 이용한 새로운 강수예측 방법을 제안한다. DBN는 비지도 사전 학습을 통해 초기 가중치를 설정하여 기존 인공신경망의 문제점을 보완한다. 예측인자로는 기온, 전일-전주 강수일, 태양과 달 궤도 관련 자료를 선정하였다. 기온과 전일-전주 강수일은 서울에서의 1974년부터 2013년까지 총 40년간의 AWS(automatic weather system) 관측 자료를 사용하였고, 태양과 달의 궤도 관련 자료는 서울을 중심으로 계산한 결과를 사용하였다. 전체 기간에서 일부는 학습 자료로 사용하여 예측모델을 생성하였고, 나머지를 생성한 모델의 검증 자료로 사용하였다. 모델 검증 결과로 나온 예측값들은 확률값을 가지며 임계치를 이용하여 강수유무를 판별하였다. 강수 정확도의 척도로 양분예보기법 중 CSI(critical successive index)와 Bias(frequency bias)를 계산하였다. 이를 통해 DBN과 MLP(multilayer perceptron)의 성능을 비교한 결과 DBN의 강수 예측 정확도가 높았고, 수행속도 또한 2배 이상 빨랐다.

키워드: 딥러닝, 딥 벨리프 네트워크, 강수, 예보

Abstract

For accurate precipitation forecasts the choice of weather factors and prediction method is very important. Recently, machine learning has been widely used for forecasting precipitation, and artificial neural network, one of machine learning techniques, showed good performance. In this paper, we suggest a new method for forecasting precipitation using DBN, one of deep learning techniques. DBN has an advantage that initial weights are set by unsupervised learning, so this compensates for the defects of artificial neural networks. We used past precipitation, temperature, and the parameters of the sun and moon's motion as features for forecasting precipitation. The dataset consists of observation data which had been measured for 40 years from AWS in Seoul. Experiments were based on 8-fold cross validation. As a result of estimation, we got probabilities of test dataset, so threshold was used for the decision of precipitation. CSI and Bias were used for indicating the precision of precipitation. Our experimental results showed that DBN performed better than MLP.

Key Words: Deep learning, Deep belief network, Precipitation, Forecast

Received: Nov. 16, 2015

Revised: Dec. 24, 2015

Accepted: Dec. 28, 2015

† Corresponding authors
yhdfly@kw.ac.kr

1. 서론

국민들의 생활과 사회·경제활동에 커다란 영향을 끼칠 수 있기 때문에 강수를 정확히 예측하여 정보를 전달하는 것은 매우 중요하다. 강수예측을 위해서는 강수에 영향을 미칠 수 있는 요소들을 파악해야 하지만 수많은 관측 자료들로부터 그것들을 파악해내는 것은 매우 어려운 일이다 [1].

최근에는 방대한 관측자료 내에서 강수패턴을 찾아내기 위한 방법으로 기계학습 기법이 사용되고 있으며, 그 중에서도 특히 인공신경망을 이용한 강수예측 연구가 많이 있었다. 인공신경망기법을 이용한 강수예측 연구는 1990년대부터 수행되어 왔다.

Kuligowski 와 Barros [2]는 인공신경망을 이용한 수치예보모델로부터 국지 강수예측을 시도하였으며, 인공신경망기법이 선형회귀보다 좋은 성능을 보였음을 나타냈다. Lee 등 [3]은 인공신경망을 이용한 분할 정복 방식의 강수 예측 방법을 제안하였다. Hall 등 [4]은 인공신경망을 이용하여 텍사스 지역의 강수확률 및 강수량 예측을 수행하였다. Luk 등 [5]

본 논문은 예보기술지원 및 활용 연구(NIMR-2012-B-1)의 일환으로 수행되었습니다.

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.

은 세 가지 종류의 인공신경망을 이용하여 단기 강우 예측을 연구하였다. Ramirez 등 [6] 은 인공신경망을 이용한 브라질 상파울루 지역의 강우예측을 연구했으며, 연구에 따르면 인공신경망이 다중선형회귀보다 좋은 성능을 보였다. Kim 등 [7] 은 방사형 신경회로망 지저 함수 신경회로망 분류기를 이용하여 여름철 초단기 강수예측 패턴 모델을 제안하였다. 또한 Kim 등 [8]은 호우특보 판별 모델 설계를 위해 RBFNNs를 이용하였다.

본 논문에서는 딥러닝 기법 중 하나인 DBN(deep belief network)을 이용한 새로운 강수 예측 방법을 제안한다. DBN는 비지도 사전 학습을 통해 초기 가중치를 설정하여, 기존 인공신경망의 단점을 보완한다. 예측인자로써는 기온, 전일-전주 강수일, 태양과 달의 궤도 관련 자료를 사용하였다.

본 논문의 구성은 다음과 같다. 2절에서는 DBN의 이론적 배경에 대해 설명한다. 3절에서는 예측인자 선정, 파라미터 설정, 검증방법, 실험결과에 대해 다루고, 4절에서 결론 및 향후연구방향을 제시한다.

2. Deep Belief Network

DBN는 그림 1과 같이 오류 역전파 알고리즘(error back-propagation algorithm)을 사용하는 인공신경망과 RBM(restricted Boltzmann machine)이 결합되어 있는 구조이다.

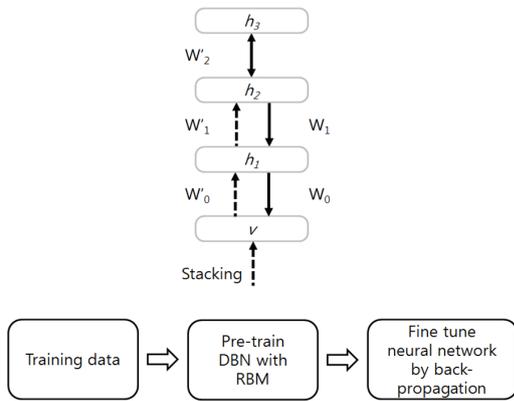


그림 1. Deep belief network 개념도
Fig. 1. Concept of deep belief network

학습 자료가 입력되면 먼저 RBM을 통해 비지도 학습(unsupervised learning)으로 가중치 값을 사전학습하여 인공신경망의 초기 가중치 값으로 사용한다. 초기 가중치 값이 결정되면 인공신경망에서는 지도학습(supervised learning)으로 오류 역전파 알고리즘을 이용한 미세조정(fine tune)을 하게 된다.

RBM은 기본적인 볼츠만 머신과 달리 같은 층의 뉴런간의 연결을 제거한 구조이다. 기본적으로 가시뉴런과 은닉뉴런으로 구성되어

은닉뉴런을 쌓아올리는 방식을 사용한다 [9]. RBM은 입력한 자료에 대하여 학습 결과를 얻은 뒤, 그 결과를 다시 입력 자료로 사용하여 학습한 결과를 기존에 있던 결과 위에 쌓는다. 이 작업을 모든 은닉층(h)에 반복하고, 그 결과로 얻은 가중치(w)들은 신경망에서의 초기 가중치로 사용한다.

3. 실험

3.1 실험자료

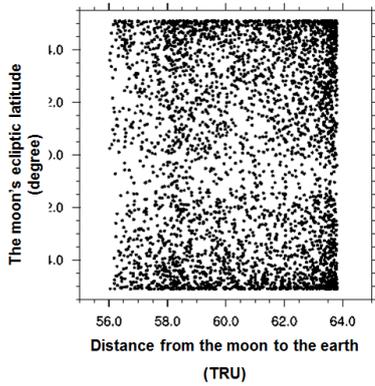
서울에서의 1974년부터 2013년까지 (40년) 자료를 사용하였다. 기상청에서 제공하는 서울(108) 지점의 강수와 기온 관측자료를 사용하였고, R프로그램의 lunar 패키지를 참고하여 서울을 기준으로 달의 위상, 달의 황도 위도, 달과 지구 사이의 거리를 계산하였다.

강수유무 예측을 위한 예측인자는 표 1과 같이 선정하였다.

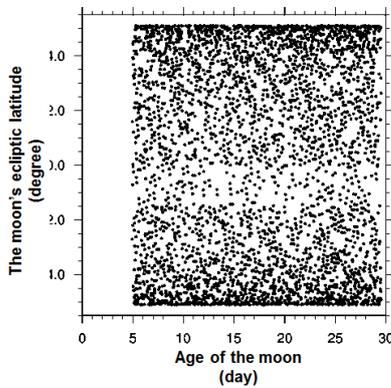
표 1. 예측인자
Table 1. Features for forecasting precipitation

No	Feature	No	Feature
1	before noon precipitation in d+0	13	day
2	before noon precipitation in d+1	14	month
3	before noon precipitation in d+2	15	afternoon precipitation in d+0
4	before noon precipitation in d+3	16	afternoon precipitation in d+1
5	before noon precipitation in d+4	17	afternoon precipitation in d+2
6	before noon precipitation in d+5	18	afternoon precipitation in d+3
7	before noon precipitation in d+6	19	afternoon precipitation in d+4
8	before noon precipitation in d+7	20	afternoon precipitation in d+5
9	age of the moon	21	afternoon precipitation in d+6
10	distance from the moon to the earth	22	afternoon precipitation in d+7
11	the moon's ecliptic latitude	23	daily temperature range
12	day of the week	24	daily mean temperature

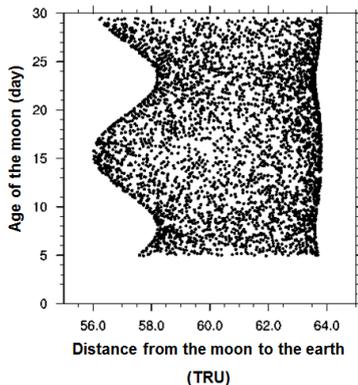
강수의 1주일 단위 연관성을 알아보기 위해 당일부터 그 전주인 7일전까지의 오전, 오후의 강수유무를 예측인자로 선정하였다. 기온은 강수발생의 주요원인 중 하나로 기온에 따라 상승류가 발생하여 비구름을 생성할 수 있다. 구름, 강수 그리고 대기 중의 수증기가 일교차에 영향을 준다는 연구 [10]가 있었으며 이에 일평균기온과 일교차를 예측인자로 선정하였다.



(a) The relation between the moon's ecliptic latitude and distance from the moon to the earth in precipitation



(b) The relation between the moon's ecliptic latitude and age of the moon in precipitation



(c) The relation between distance from the moon to the earth and age of the moon in precipitation

그림 2. 달의 변화에 따른 강수유무
Fig. 2. Precipitation of the moon's moving

양력날짜에 따른 강수의 주기성을 알아보기 위하여 요일, 일, 월을 추가하였다. 일 자료의 경우 1부터 365까지 1년 단위의 값으로 나타냈으며, 요일의 경우 월요일부터 일요일까지를 각각 0부터 6으로 나타냈다.

달의 변화와 강수유무의 연관성을 알아보기 위해 1974년부터 2013년까지의 달의 위상, 달의 황도 위도, 달과 지구 사이의 거리에 따른 강수유무를 그림 2에 나타냈다.

달의 위상 주기는 29.5일이며, 달과 지구와의 거리는 지구 반지름 단위(terrestrial radius unit, TRU)이다. 그림 2(a)에서 달의 황도 위도가 약 5 또는 약 -5에 가깝고 달과 지구 사이의 거리가 멀어질수록 강수가 많이 온 것으로 나타났다. 그림 2(b)에서 달의 위상 주기가 0~5일 때 강수가 전혀 관측되지 않았으며 달의 위도가 약 5 또는 약 -5에 가까울수록 강수가 많은 것으로 나타났다. 그림 2(c) 또한 달의 위상 주기가 0~5일 때 강수가 전혀 관측되지 않았으며 달과 지구 사이의 거리가 멀어질수록 강수가 많이 왔다.

그림 2에 나타난 결과를 고려해 볼 때 달의 변화가 강수유무에 미치는 영향이 있다고 판단되며, 이를 근거로 달의 위상, 달의 황도 위도, 달과 지구 사이의 거리를 예측인자로 선정하였다.

모든 실험자료는 정규화(normalization) 과정을 거쳐 0에서 1 사이의 값으로 치환하였다. 이 과정을 통해 신경망 학습의 빠른 수렴과 잡음이나 왜곡에 의해 변형된 것을 바로 잡는 효과를 기대할 수 있다 [11].

3.2 실험방법

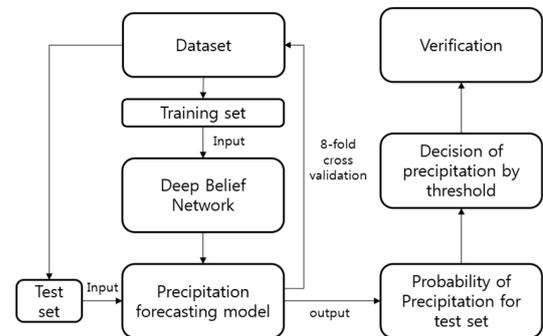


그림 3. 내일강수 예측 시스템

Fig. 3. System of forecasting the precipitation of the next day

내일강수 예측 시스템은 그림 3과 같다. DBN 학습을 통해 강수유무 예측모델을 생성하여 검증자료에 대해 강수유무 확률을 출력하는 단계와 출력된 강수유무 확률예측값을 임계값(threshold)에 따라 강수유무를 판별하고 검증하는 단계로 나누었다.

입력자료는 표 2와 같이 40년간의 실험자료를 5년 단위로 8개의 자료로 분할하여 사용하였다. 그 중 7개를 학습자료로 사용하고 나머지 1개를 검증자료로 사용하였다. 모든 자료에 대해 검증하기 위해 학습자료와 검증자료를 바꿔가며 8번 반복하여 수행하였다. (8-fold cross validation)

표 2. 데이터 구성
Table 2. Data set

Dataset	Period	Number of data
Dataset1	1974-1978	1826
Dataset2	1979-1983	1826
Dataset3	1984-1988	1827
Dataset4	1989-1993	1826
Dataset5	1994-1998	1826
Dataset6	1999-2003	1826
Dataset7	2004-2008	1827
Dataset8	2009-2013	1826

실험에 이용한 DBN는 R 프로그램의 deepnet 패키지를 참고하여 Fortran으로 작성하였다. 인공신경망과의 성능 비교를 위해 Weka 프로그램에 내장된 MLP(multilayer perceptron)을 이용하였다.

신경망 프로그램을 동작하기 위해서는 사전에 일부 변수들을 지정해주어야 한다. DBN과 MLP의 매개변수 설정값은 표 3과 같다. Batch size는 자료의 단위를 지정하여 처리하기 위한 변수로 속도향상을 위한 방법이다. Decay는 값이 수렴할수록 학습률이 점차 감소하도록 지정하는 변수이며, Dropout은 과적합을 막기 위한 방법으로 임의의 노드 비활성화 정도를 정하는 변수이다. 각 설정값은 공정한 비교를 위해 Weka의 MLP 기본설정을 따랐으며, 그 외의 변수는 성과와 속도를 고려하여 경험적으로 결정하였다.

표 3. 신경망 매개변수 설정값
Table 3. Predetermined parameters of neural networks

Parameter	DBN	MLP
Hidden layer	1	1
Hidden node	100	100
#epochs	500	500
Batch size	100	-
Learning rate	0.3	0.3
Decay	0	0
Momentum	0.2	0.2
Dropout	0.1	-

3.3 검증방법

강수유무 판별을 위하여 각 데이터셋 검증과정에서 나온 예측값들의 중앙값을 평균하여 임계값(threshold)으로 정하였다. 예측값이 임계값보다 클 경우 강수가 있다고 분류하고 작을 경우 강수가 없다고 분류하였다.

강수유무 판별 정확도를 검증하기 위해 양분예보기법을 이용하였다. 양분예보기법은 표 4에서와 같이 예측과 관측을 yes와 no로 나누어 그 관계에 따라 hits, false alarms, misses, correct rejections로 나타낸다[12].

표 4. 양분예보기법
Table 4. Method for dichotomous forecasts

Event forecast	Event observed		
	Yes	No	Total
Yes	H (Hits)	F (False alarms)	$H+F$
No	M (Misses)	C (Correct rejection)	$M+C$
Total	$H+M$	$F+C$	$H+F+M+C$

양분예보기법을 이용한 여러 강수검증 지표들이 있으며, 본 연구에서는 그 중 하나인 CSI(critical successive index)를 사용하였다. CSI는 강수 상황에 대한 예측이나 관측이 있을 때만을 고려한 지표 [13]로 식 (1)과 같다. CSI가 무강수 예측을 고려하지 않는다는 단점을 보완하기 위하여 예측과 관측간의 치우침 정도를 나타내는 Bias(frequency bias)를 함께 사용하였으며, 식 (2)와 같다.

$$CSI = \frac{H}{H+M+F} \tag{1}$$

$$Bias = \frac{H+F}{H+M} \tag{2}$$

3.4 실험결과

MLP와 DBN를 이용하여 각 5회씩 동일한 실험방법을 수행하여 퍼포먼스 다이어그램(performance diagram)에 나타냈다. 퍼포먼스 다이어그램은 양분예보기법을 이용한 CSI, Bias 등 여러 강수검증 지표들을 하나의 그림에 표현할 수 있다 [14]. 퍼포먼스 다이어그램에서 왼쪽 아래에서 오른쪽 위로 뻗어있는 점선은 Bias

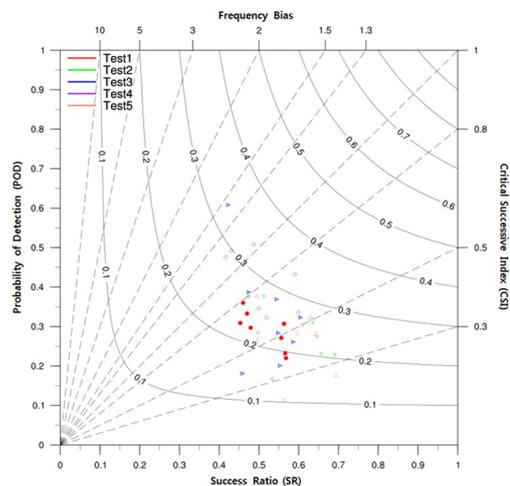


그림 4. MLP 실험결과를 나타낸 퍼포먼스 다이어그램
Fig. 4. Performance diagram showing experimental results of MLP

값을 나타내고, 오른쪽에서부터 나선형으로 생긴 선분은 CSI값을 나타낸다. 그림 4는 MLP의 실험수행 결과를, 그림 5는 DBN의 실험수행 결과를 나타낸다.

그림 4에서 CSI는 대체로 0.3이하이고 Bias는 1보다 작다. Bias가 1보다 작다는 것은 실제 강수 발생 관측보다 강수 발생 예측의 횟수가 적었다는 것을 의미한다.

그림 5에서 각 결과들은 CSI가 0.3에 가깝게 나타나며 Bias는 1부터 3까지 널리 퍼져있다. DBN에서는 MLP와는 반대로 Bias가 대체로 1보다 큰 것으로 나타나 실제 강수 발생 관측보다 강수 발생 예측 횟수가 많았다는 특징이 나타났다.

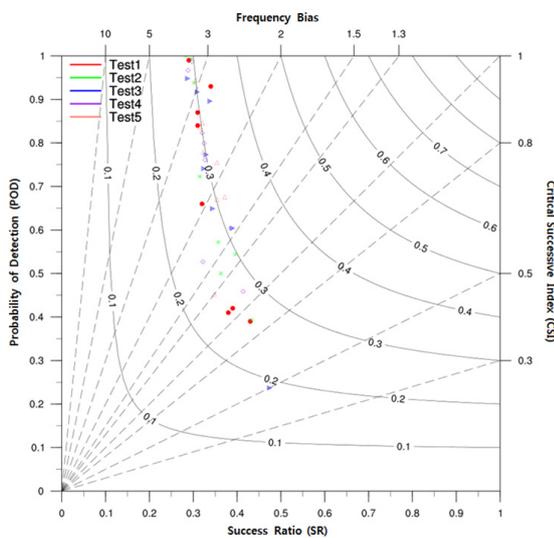


그림 5. DBN 실험결과를 나타낸 퍼포먼스 다이어그램
Fig. 5. Performance diagram showing experimental results of DBN

표 5와 표 6은 MLP와 DBN의 각 데이터셋 검증에서 5회 실험한 결과의 평균값과 전체 평균값을 나타낸다. 전체적으로 DBN의 CSI가 MLP의 CSI보다 높게 나타났으며, 수행속도는 DBNs이 MLP에 비해 2배 이상 빨랐다. Bias는 MLP의 경우 0.597, DBN은 2.203으로 서로 다른 예측경향을 보였다.

표 5. MLP 실험결과
Table 5. Experimental results of MLP

Test data	Average		
	CSI	Bias	Time(s)
Dataset1	0.206	0.522	453
Dataset2	0.214	0.557	454
Dataset3	0.227	0.686	457
Dataset4	0.236	0.471	457
Dataset5	0.242	0.491	454
Dataset6	0.251	0.498	456
Dataset7	0.248	0.563	454
Dataset8	0.284	0.985	461
Total ave	0.238	0.597	456

표 6. DBN 실험결과
Table 6. Experimental results of DBN

Test data	Average		
	CSI	Bias	Time(s)
Dataset1	0.290	2.856	196
Dataset2	0.267	2.123	194
Dataset3	0.297	2.725	195
Dataset4	0.281	1.912	201
Dataset5	0.267	2.217	203
Dataset6	0.294	2.442	202
Dataset7	0.297	1.441	197
Dataset8	0.302	1.910	197
Total ave	0.287	2.203	198

4. 결론 및 향후 연구

본 논문에서는 DBN를 이용한 새로운 강수예측 방법을 제안하고 기온, 전일-전주 강수일, 태양과 달 궤도 관련 자료를 예측인자로 선정하여 내일강수유무를 예측하였다. MLP와의 성능비교 결과 DBN의 강수 유무 예측 성능이 MLP보다 좋았으며, DBN의 수행속도는 MLP의 수행속도보다 2배 이상 빠르다는 것을 확인할 수 있었다.

향후 연구에서는 강수 예측 정확도를 높이기 위한 방법으로써 강수와 연관성이 높은 예측인자를 탐색 및 추가하고, 유전 알고리즘을 이용한 DBN 파라미터 최적화를 적용해보려고 한다. 또한 다양한 기계학습 기법과 DBN의 성능 비교를 통해 DBN를 이용한 강수예측 방법의 효용성을 검증해보려고 한다.

References

- [1] J.-H. Seo, Y. H. Lee, Y.-H. Kim, "Feature selection to predict very short-term heavy rainfall based on differential evolution", *Journal of Korean Institute of Intelligent Systems*, vol. 22, no. 6, pp. 706-714, 2012
- [2] R. J. Kuligowski, A. P. Barros, "Localized precipitation forecasts from a numerical weather prediction model using artificial neural networks", *Weather and Forecasting*, vol. 13, no. 4, pp. 1194-1204, 1998
- [3] S. Lee, S. Cho, P. M. Wong, "Rainfall prediction using artificial neural networks", *Journal of Geographic Information and Decision Analysis*, vol. 2, no. 2, pp. 233-242, 1998
- [4] T. Hall, H. E. Brooks, C. A. Doswell III, "Precipitation forecasting using a neural network", *Weather and Forecasting*, vol. 14, no.

3, pp. 338-345, 1999

[5] K. C. Luk, J. E. Ball, A. Shama, "An application of artificial neural networks for rainfall forecasting", *Mathematical and Computer Modelling*, vol. 33, no. 6-7, pp. 683-693, 2001

[6] M. C. V. Ramirez, H. F. C. Velho, N. J. Ferreira, "Artificial neural network technique for rainfall forecasting applied to the Sao Paulo region", *Journal of Hydrology*, vol. 301, nos. 1-4, pp. 146-162, 2005

[7] H.-K. Kim, W.-Y. Choi, S.-K. Oh, "Design of summer very short-term precipitation forecasting pattern in metropolitan area using optimized RBFNNs", *Journal of Korean Institute of Intelligent Systems*, vol. 23, no. 6, pp. 533-538, 2013

[8] H.-M. Kim, S.-K. Oh, Y.-H. Lee, "Design of heavy rain advisory decision model based on optimized RBFNNs using KLAPS reanalysis data", *Journal of Korean Institute of Intelligent Systems*, vol. 23, no. 5, pp. 473-478, 2013

[9] G. E. Hinton, S. Osindero, Y.-W. Teh, "A fast learning algorithm for deep belief nets", *Neural Computation*, vol. 18, no. 7, pp. 1527-1554, 2006

[10] A. Dai, K. E. Trenberth, T. R. Karl, "Effects of clouds, soil moisture, precipitation, and water vapor on diurnal temperature range", *Journal of Climate*, vol. 12, no. 8, pp. 2451-2473, 1999

[11] S. Samarasinghe, *Neural Networks for Applied Sciences and Engineering: From Fundamentals to Complex Pattern Recognition*, CRC Press, 2006

[12] I. T. Jolliffe, D. B. Stephenson, *Forecast Verification: a Practitioner's Guide in Atmospheric Science*, John Wiley & Sons, 2012

[13] R. J. Donaldson, R. M. Dyer, M. J. Kraus, "Objective evaluator of technique for predicting severe weather events", *Bulletin of the American Meteorological Society*, vol. 56, no. 7, 1975

[14] P. J. Roebber, "Visualizing multiple measures of forecast quality", *Weather and Forecasting*, vol. 24, no. 2, pp. 601-608, 2009

저자 소개



하지훈(Ji-Hun Ha)

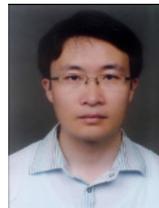
2014년 : 서울시립대학교 환경공학부 학사

2015년~현재 : 광운대학교 대학원 임베디드SW
공학과 석사과정

관심분야 : 기계학습, 딥러닝, 유전알고리즘

Phone : +82-2-940-5212

E-mail : onlyred@kw.ac.kr



이용희(Yong Hee Lee)

1997년 : 경북대학교 천문대기과학과 이학 석사

1998년~2007년 : 기상연구소 기상연구사

2007년~현재 : 국립기상과학원 기상연구관

관심분야 : 기상수치예보모델, 자료동화, 초단기 기상예보,
기상예보 · 특보

Phone : +82-70-7850-6838

E-mail : gonos2004@korea.kr



김용혁(Yong-Hyuk Kim)

1999년 : 서울대학교 전산과학전공 학사

2001년 : 서울대학교 전기컴퓨터공학부 석사

2005년 : 서울대학교 전기컴퓨터공학부 박사

2005년~2007년 : 서울대학교 반도체

공동연구소 연구원

2007년~현재 : 광운대학교 컴퓨터소프트웨어학과 부교수

관심분야 : 최적화, 진화 연산, 지식 공학

Phone : +82-2-940-5212

E-mail : yhdfly@kw.ac.kr