

Deep learning 이론을 이용한 증발접시 증발량 모형화

Pan evaporation modeling using deep learning theory

서영민* · 김성원**

Youngmin Seo · Sungwon Kim

요 지

본 연구에서는 일 증발접시 증발량 산정을 위한 딥러닝 (deep learning) 모형의 적용성을 평가하였다. 본 연구에서 적용된 딥러닝 모형은 deep belief network (DBN) 기반 deep neural network (DNN) (DBN-DNN) 모형이다. 모형 적용성 평가를 위하여 부산 관측소에서 측정된 기상자료를 활용하였으며, 증발량과의 상관성이 높은 기상변수들 (일사량, 일조시간, 평균기상온도, 최대기온)의 조합을 고려하여 입력변수집합 (Set 1, Set 2, Set 3)별 모형을 구축하였다.

DBN-DNN 모형의 성능은 통계학적 모형성능 평가지표 (coefficient of efficiency, CE; coefficient of determination, r^2 ; root mean square error, RMSE; mean absolute error, MAE)를 이용하여 평가되었으며, 기존의 두 가지 형태의 ANN (artificial neural network), 즉 모형학습 시 SGD (stochastic gradient descent) 및 GD (gradient descent)를 각각 적용한 ANN-SGD 및 ANN-GD 모형과 비교하였다. 효과적인 모형학습을 위하여 각 모형의 초매개변수들은 GA (genetic algorithm)를 이용하여 최적화하였다. 그 결과, Set 1에 대하여 ANN-GD1 모형, Set 2에 대하여 DBN-DNN2 모형, Set 3에 대하여 DBN-DNN3 모형이 가장 우수한 모형 성능을 나타내는 것으로 분석되었다. 비록 비교 모형들 사이의 모형성능이 큰 차이를 보이지는 않았으나, 모든 입력집합에 대하여 DBN-DNN3, DBN-DNN2, ANN-SGD3 순으로 모형 효율성이 우수한 것으로 나타났다.

핵심용어 : Deep neural network, Deep belief network, Stochastic gradient descent, Artificial neural network, Pan evaporation

1. 서 론

신뢰성 있는 증발접시 증발량의 산정은 수문학적 설계 및 수자원 계획, 특히 지역 수자원 계획, 저수지 관리, 용수공급 및 배분, 가뭄관리 등의 측면에서 중요한 역할을 한다. 지금까지 증발접시 증발량의 산정은 경험식, 통계학적 모형, 자료기반 모형 등에 의해서 수행되어 왔으나 최근 인공지능경망, 퍼지기반모형 등을 포함하는 다양한 기계학습 모형의 수문학적 활용 등으로 인하여 보다 정확성이 높은 모형 개발이 이루어지고 있다. 한편, 2012년 이후 여러 분야에서 활발한 적용이 이루어지고 있고 급격한 성장을 보이고 있는 딥러닝 모형들은 그것의 수문학적 분야에서의 적용 가능성이 최근 주목되고 있다 (Bai et al. 2016; Li et al. 2016; Tao et al. 2016). 본 연구에서는 일 증발접시 증발량 산정을 위한 deep learning 모형의 적용성을 평가하였으며, 딥러닝 모형으로서 deep belief network (DBN) 기반 deep neural network (DNN) (DBN-DNN) 모형을 적용하였다.

2. Deep neural network

2.1 Restricted Boltzmann machines (RBMs)

RBMs은 가시층 (visible layer) (v)과 은닉층 (hidden layer) (h)의 두 개의 층으로 구성된 네트워크이다. 가

* 정회원 · 공학박사 · 경북대학교 과학기술대학 건설환경공학과 · E-mail: ymseoo@knu.ac.kr

** 정회원 · 발표자 · 공학박사 · 수자원개발기술사 · 동양대학교 철도건설안전공학과 부교수 · E-mail: swkim1968@dyu.ac.kr

시층 및 은닉층 노드의 결합확률은 $\mathbf{v} = \{v_1, v_2, \dots, v_N\}$, $\mathbf{h} = \{h_1, h_2, \dots, h_N\}$ 에 대하여 다음과 같이 추정된다.

$$p(\mathbf{v}, \mathbf{h} | \boldsymbol{\theta}) = \frac{\exp(-E(\mathbf{v}, \mathbf{h} | \boldsymbol{\theta}))}{Z} \quad (1)$$

여기서, $E(\mathbf{v}, \mathbf{h} | \boldsymbol{\theta})$ 는 에너지 함수 (energy function), $Z = \sum_{\mathbf{v}, \mathbf{h}} \exp(-E(\mathbf{v}, \mathbf{h} | \boldsymbol{\theta}))$ 이다.

RBM는 Bernoulli RBM 및 Gaussian RBM과 같은 두 가지 타입이 있으며, RBM의 학습은 대조발산 (contrastive divergence, CD) 알고리즘을 이용하여 학습된다.

2.2 Deep belief networks (DBNs)

DBNs는 확률발생모형 (probabilistic generative model)으로서 딥러닝 모형의 일종이다. 딥러닝 모형 중에서 가장 일반적이고 효과적인 모형 중 하나는 다수의 은닉층을 가지는 전방향 신경망 (feed forward neural network)이다. DBNs은 여러 개의 RBMs를 쌓아올린 구조를 가지고 있다. 일반적인 신경망에서 층의 수가 증가하면 경사소실 (gradient vanishing) 문제로 인하여 대수우도함수가 국부 최소점에 빠지는 문제점이 발생한다. 그러나 DNN에서는 DBN을 이용한 사전학습 (pretraining)을 통해 이러한 문제점을 개선한다. 즉, DBN으로부터 학습된 모형의 가중치들은 DNN을 위한 초기 가중치로 사용되며, DBN을 사용하여 사전학습된 DNN은 개선된 모형성능을 나타내게 된다.

첫 번째 RBM은 가시층에서의 입력자료와 CD 알고리즘을 이용하여 학습된다. 학습된 첫 번째 RBM의 위에 두 번째 RBM이 위치하게 되며, 가시층에 대한 값으로서 이전 RBM의 은닉노드로부터의 출력값을 사용한다. 이러한 과정을 반복하며, 최상위 층에 대한 은닉층의 출력값은 감독모형 (supervised model)의 입력으로 사용될 수 있다. 이를 위하여 DBN의 최상위 층에 신경망을 위치시키고 DBN의 최상위 층으로부터의 출력값을 신경망의 입력으로 사용한다.

2.3 Genetic algorithm (GA)

GA는 자연 세계의 진화과정에 기초한 전역 최적화 기법 (global optimization method)으로서 선택 (selection), 교차 (crossover), 변이 (mutation), 대체 (replacement) 등과 같은 주요 연산을 통해 최적 매개변수를 탐색한다.

3. 적용 및 결과

DBN 기반 DNN 모형 (DBN-DNN)을 이용한 일 증발접시 증발량을 모형화하기 위하여 부산 관측소에서 측정된 기상자료 (일사량, 일조시간, 평균지상온도, 최대기온) (training data, 2010-2014; testing data, 2015)를 적용하였다. 최적 모형을 결정하기 위해 증발량과의 상관성이 높은 변수들을 조합하여 Table 1과 같이 세 가지 입력집합 (Set 1, Set 2, Set 3)을 구성하였다. DBN-DNN 모형의 성능을 평가하기 위하여 통계학적 모형성능평가 지표 (CE, r^2 , RMSE, MAE)를 적용하였으며, 기존의 두 가지 유형의 신경망 모형인 확률적 경사하강법 (stochastic gradient descent, SGD)을 적용한 ANN-SGD 모형과 경사하강법 (gradient descent, GD)을 적용한 ANN-GD 모형과 비교하였다. 모형별 초매개변수를 최적화하기 위하여 본 연구에서는 GA (genetic algorithm)를 적용하였다. DBN-DNN의 경우 은닉층 개수, 은닉노드 개수, 반복수, 배치크기 (batch size)에 대하여 GA를 적용하였으며, ANN-SGD의 경우 은닉노드 개수, 반복수, 배치크기에 대하여, ANN-GD의 경우 은닉노드 개수에 대하여 GA를 적용하였다.

Table 2는 모형 효율성 평가지표를 비교한 것이고, Figure 1은 관측치와 예측치에 대한 산점도를 나타낸다. 모델 효율성 측면에서, Set 1의 경우 ANN-GD1 모형, Set 2의 경우 DBN-DNN2 모형, Set 3의 경우 DBN-DNN3 모형이 가장 우수한 모형 성능을 나타내는 것으로 분석되었다. 비록 비교 모형들 사이의 모형성능이 큰 차이를 보이지는 않았으나, 모든 입력집합에 대하여 DBN-DNN3, DBN-DNN2, ANN-SGD3 순으로 모형 효율성이 우수한 것으로 나타났다. DBN-DNN3 모형의 경우, CE=0.795, $r^2=0.800$, RMSE=0.754, MAE=0.580으로서 모든 비교 모형 중 가장 우수한 성능을 가지는 것으로 분석되었다.

4. 결론

본 연구에서는 일 증발접시 증발량 산정에 있어서 딥러닝 모형의 적용성을 평가하기 위하여 DBN 기반 DNN (DBN-DNN) 모형을 적용하였다. 이를 위하여 부산 관측소에서 측정된 기상자료를 활용하였으며, 증발량과의 상관성이 높은 기상변수들로부터 세 개의 입력집합별 모형을 구축하였다.

모형 효율성 측면에서 DBN-DNN3, DBN-DNN2, ANN-SGD3의 순으로 우수한 성능을 나타내었으며, DBN-DNN3 모형은 $CE=0.795$, $r^2=0.800$, $RMSE=0.754$, $MAE=0.580$ 으로 모든 비교 모형 중 가장 우수한 성능을 가지는 것으로 분석되었다. 따라서 DBN-DNN 모형은 관측된 기상자료를 이용하여 일 증발접시 증발량을 산정하는데 효과적인 도구로 적용될 수 있을 것으로 판단된다.

참 고 문 헌

1. Bai Y, Chen Z, Xie J, Li C (2016). Daily reservoir inflow forecasting using multiscale deep feature learning with hybrid models. *Journal of Hydrology*, 532: 193-206.
2. Li C, Bai Y, Zeng B (2016). Deep feature learning architectures for daily reservoir inflow forecasting. *Water Resources Management*, 30: 5145-5161.
3. Tao Y, Gao X, Hsu K, Sorooshian S (2016). A deep neural network modeling framework to reduce bias in satellite precipitation products. *Journal of Hydrometeorology*, 17(3): 931-945.

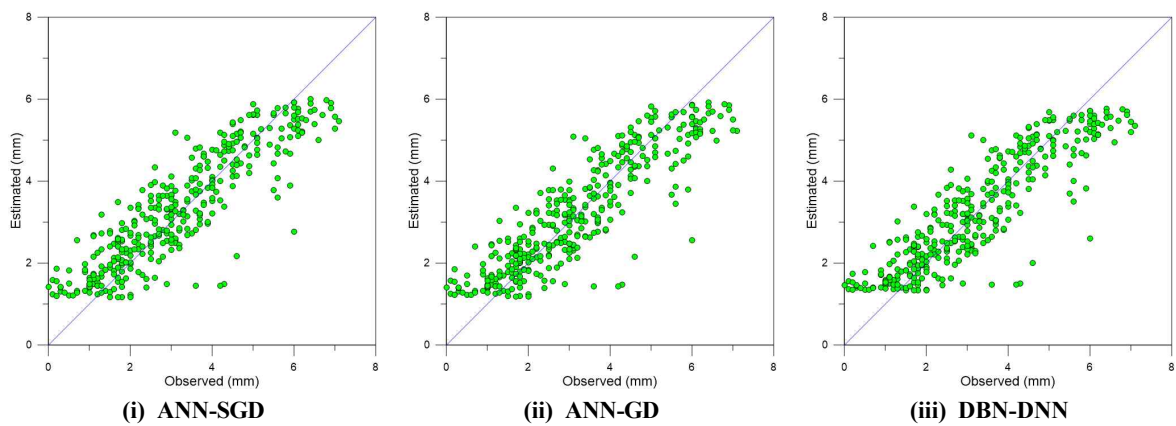
Table 1. Combination of input and output variables

Station	Input sets	Input variables	Output variables
Busan	Set 1	R_s, D	E_p
	Set 2	R_s, D, GT_{mean}	E_p
	Set 3	$R_s, D, GT_{mean}, T_{max}$	E_p

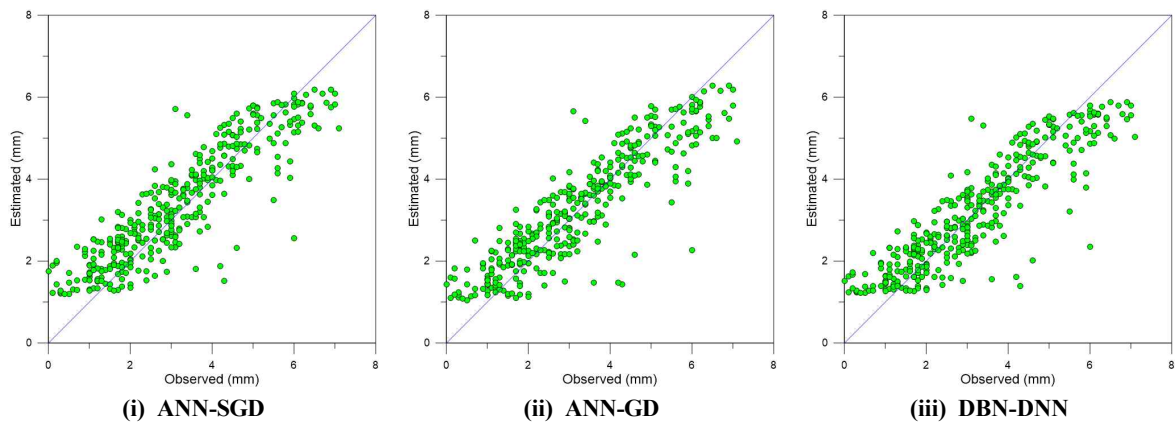
R_s : solar radiation, D : sunshine duration, GT_{mean} : mean ground temperature, T_{max} : maximum air temperature, E_p : evaporation

Table 2. Evaluation of model performance

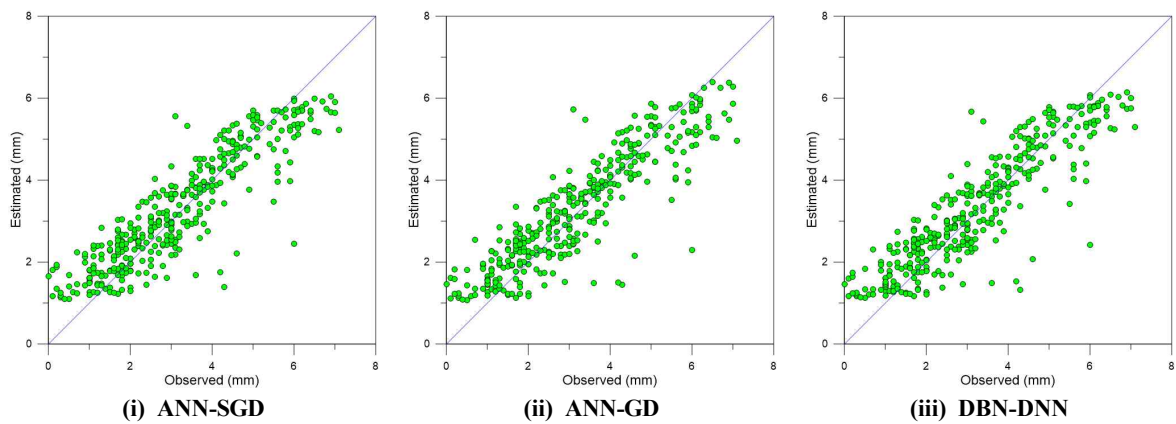
Input sets	Models	CE	r^2	RMSE (mm)	MAE (mm)
Set 1	ANN-SGD1	0.770	0.779	0.798	0.629
	ANN-GD1	0.774	0.778	0.791	0.619
	DBN-DNN1	0.773	0.778	0.793	0.622
Set 2	ANN-SGD2	0.777	0.797	0.785	0.620
	ANN-GD2	0.788	0.794	0.767	0.585
	DBN-DNN2	0.792	0.798	0.759	0.577
Set 3	ANN-SGD3	0.791	0.800	0.761	0.588
	ANN-GD3	0.786	0.793	0.770	0.590
	DBN-DNN3	0.795	0.800	0.754	0.580



(a) Set 1



(b) Set 2



(c) Set 3

Figure 1. Scatter plots