

삼교호 유입량 예측을 위한 LSTM 모형의 적용성 평가

황병기
상명대학교 건설시스템공학과

Evaluation of LSTM Model for Inflow Prediction of Lake Sapgye

Byung-Gi Hwang
Department of Civil Engineering, Sangmyung University

요약 삼교호로 유입하는 곡교천 지역의 홍수시 유출량을 추정하기 위해서 Tensorflow를 활용하여 파이썬 기반의 LSTM 모형을 구축하였다. 층의 깊이가 성능에 미치는 영향을 분석하기 위해, 은닉층의 깊이를 2, 4, 6층으로 증가시키면서, 선행시간 1시간부터 5시간까지 예측을 수행하였으며, 은닉층의 개수가 4개일 때가 가장 우수한 성능을 나타내었다. 학습에 사용하는 입력자료의 길이 즉, 시퀀스길이가 모형의 성능에 미치는 영향을 파악하고자 시퀀스길이를 3시간, 5시간, 7시간으로 증가시키면서 모형을 실행한 결과, 시퀀스길이가 3시간 일 때, 전 시간대에 걸쳐 예측 성능이 우수한 것으로 분석되었다. 모형 검증에서 극한 강우 3건에 대하여 예측을 수행한 결과 선행시간 1시간에 대하여 평균 NSE 0.96 이상의 높은 정확도를 나타내었으며, 선행시간 2시간 이상에 대하여 정확도는 점차적으로 낮아지는 것으로 확인되었다. 결론적으로 시퀀스길이 3시간을 사용하여 선행시간 1시간에 대한 예측을 수행한다면 곡교천 강청 관측소의 홍수위를 높은 수준의 정확도로 예측할 수 있음을 확인하였다.

Abstract A Python-based LSTM model was constructed using a Tensorflow backend to estimate the amount of outflow during floods in the Gokgyo-cheon basin flowing into the Sapgyo Lake. To understand the effects of the length of input data used for learning, i.e., the sequence length, on the performance of the model, the model was implemented by increasing the sequence length to three, five, and seven hours. Consequently, when the sequence length was three hours, the prediction performance was excellent over the entire period. As a result of predicting three extreme rainfall events in the model verification, it was confirmed that an average NSE of 0.96 or higher was obtained for one hour in the leading time, and the accuracy decreased gradually for more than two hours in the leading time. In conclusion, the flood level at the Gangcheong station of Gokgyo-cheon can be predicted with high accuracy if the prediction is performed for one hour of leading time with a sequence length of three hours.

Keywords : Python, LSTM, Tensorflow, Sequence length, leading time

1. 서론

최근 수년간 게릴라성 폭우, 태풍을 포함한 극한 강우가 빈번히 발생하고 있어 이로 인한 하천 범람 우려가 증가하고 있다. 특히 하천 하류 지역은 지반고가 낮아 홍수

로 인한 인명 피해와 재산 손실에 취약하다. 홍수예보에 대한 정확도를 개선하고 신속하게 관심 지점에서의 수위를 예측할 수 있다면, 손실을 최소화할 수 있을 것이다. 홍수로 인한 피해를 저감하기 위해서는 강우-유출 과정에 대한 적절한 해석과 정확도 높은 강우의 예측과 홍

본 논문은 상명대학교 2020년도 교내연구비 지원으로 수행되었음.
*Corresponding Author : Byung-Gi Hwang(Sangmyung Univ.)

email: bh4a@smu.ac.kr

Received January 6, 2021

Accepted April 2, 2021

Revised February 17, 2021

Published April 30, 2021

수량 산정이 필요하다. 홍수량 산정 방안으로 수문모형을 이용하여 유출 거동을 모의하는 방법이 있으나 이러한 물리적 모형은 유역의 물리적인 현상을 매개변수를 이용하여 유역의 특성을 반영하여 모의하게 되므로, 매개변수 추정치의 정확도에 따라 모형의 예측값이 바뀌게 되는 경향이 있다. 또한, 대부분의 물리적 수문모형은 비선형인 강우-유출 과정을 선형으로 가정하여 해석하므로 그로 인하여 예측 결과의 정확성에 문제가 발생할 수 있다[1]. 최근 수문학 분야에는 인공신경망을 이용하여 강우예측, 자료보정, 기후자료 상세화 등 다양한 문제점을 해결해 나가고 있고[2], 적용성이 확대되어 가고 있는 추세이다.

수문학 분야에 인공신경망 적용 사례를 살펴보면 다음과 같다. Jung(2018a)는 금강 유역 내 대청댐 상류부에 위치한 옥천 관측소 지점의 수위예측을 위하여 다중선형 회귀모형과 LSTM 인공신경망 모형을 구축하여[3], LSTM 모형이 안정적이고 우수하게 수위를 예측하는 것을 확인하였다. Jung(2018b)은 LSTM 모형을 이용하여 감조하천에서의 수위예측을 실행하였으며[4], 선행시간이 6시간 이하에서는 첨두수위에 대해 어느 정도의 정확도가 확보되었으나, 9시간 이후에서는 수위곡선이 과소 모의된 것을 확인하였다. Park(2018)는 댐유입량 예측 모형의 적용성 검토를 목적으로 소양강댐 및 충주댐 유역을 대상으로 인공신경망 모형과 엘만 순환신경망 모형을 구축하였으며[5], 엘만 순환신경망 모형이 댐 유입량 예측에 있어서 인공신경망 모형에 비해 예측결과가 우수한 것으로 나타났으며, 또한, 다층 은닉층 구조가 단일 은닉층 구조보다 예측 성능이 개선됨을 입증하였다.

Zhang(2018a)는 합류식 하수관거 월류수(CSOs: Combined Sewer Overflow)의 수위를 예측하기 위하여, 다층퍼셉트론(MLP: Multi-Layer Perceptron), 웨이브렛 신경망(WNN: Wavelet Neural Network), LSTM, 게이트드 순환 유닛(GRU: Gated Recurrent Unit)를 포함하는 다수의 딥러닝 기법의 적용성을 비교하여, LSTM, GRU와 같은 순환신경망 모형이 다른 모형에 비하여 우수함을 입증하였다[6]. 그리고 Zhang(2018b)은 하수처리장의 유입량 예측을 위해서 LSTM 모형을 적용하였으며, 노르웨이 남동부에 위치하고 있는 Muusoya 처리장에 적용하여 예측의 우수성을 검증하였다[7].

Qi(2019)은 안강 저수지(중국) 유입량을 예측하기 위해서 LSTM 모형을 개선한 DEL-LSTM 모형을 사용하였으며[8], 이 모형은 유입량 자료를 경향(trend), 주기(period), 임의(random) 향으로 분해하여 산정한 후 각 항의 결과를 합치는 방법을 사용하여 정규화 평균제곱오

차를 4%까지 줄일 수 있었다. He(2019)은 유출 모형의 예측 정확도를 높이기 위해서 하이브리드 모형을 개발하였으며[9], 이는 분해모드(VMD)와 심층신경망(DNN)으로 구성되었으며 Zhangjiashan 관측소(중국) 자료를 이용하여 평가를 수행하여 모형의 성능이 Nash-Sutcliffe 효율 0.95 이상을 입증하였다. Zhang(2018c)은 극한 강우시 하수처리장에서 월류수를 줄이기 위해서 수리모형과 순환신경망 모형을 결합하여 사용하였다[10]. 여기서 순환신경망 모형은 Elman, NARX, LSTM을 적용하였으며, Druammen 시(노르웨이) 하수관 시스템에 테스트한 결과 LSTM이 가장 성능이 우수함을 보였다. Kao(2020)은 홍수 시 다단계 선행시간 예측을 위해 인코더(Encoder)와 디코더(Decoder)를 LSTM과 결합한 LSTM-ED 모형을 제안하였으며, 이 모형을 태풍 4건에 대하여 Shihmen Reservoir (타이완)에 적용한 결과 NSE 0.96~0.99의 높은 성능을 나타냄을 입증하였다[11].

본 연구는 곡교천 하류 말단 강청 관측소에서 삽교호로 유입하는 홍수시 유입량 예측을 위해 시계열 자료에 우수성이 입증된 LSTM 모형을 구축하여 수위 예측을 수행하고 선행시간에 따른 예측 정확도를 분석하고자 한다. 모형 학습과 검증을 위한 입력자료는 강청 관측소 상류에 위치한 충무교 및 한내다리 관측소, 지류인 온천교 관측소의 2011년부터 2019년까지 9년간 전 기간 시간별 수위 자료와 유역 내 강수량 측정소 5곳의 시간별 강수량 자료를 활용하였다.

2. 본론

2.1 이론적 배경

인공신경망(Artificial Neural Network)은 여러 층으로 구성되어 있으며 일반적으로 입력층(input layer), 출력층(output layer), 은닉층(hidden layer)을 포함하며, 은닉층은 목적에 따라 다수의 은닉층이 존재할 수 있다. 각각의 층에는 뉴런(Neuron) 들이 존재하고 이 들이 층과 층 사이를 연결시키는 구조로 되어 있다.

인공신경망은 구조에 따라 전방향 신경망(Feedforward Neural Network, FFNN)과 회귀 또는 순환 신경망(Recurrent Neural Network, RNN)으로 나눈다. 전방향 신경망은 패턴인식 등에 강점을 나타내나, 주식이나 홍수위 예측과 같은 시계열 패턴에 적용할 경우에는 만족할 만한 결과를 얻지 못하는 경우가 많다. 하천수위 예측과 같은 시계열 패턴은 이전 정보를 현재의 문제해결

에 활용하기 위해서 신경망의 입력층과 출력층 사이에 되돌림(순환)이 존재하는 RNN을 이용하게 된다. 그러나 RNN은 현재 노드와 먼 과거 상태를 처리할 경우 기울기 소실 문제가 생겨서 학습이 효과적으로 이루어지지 못하는, 즉 장기의존성 문제라는 단점이 있다.

2.2 LSTM 모형

RNN에서 발생하는 장기의존성 문제를 해결하기 위해서 장단기 메모리(Long-Short Term Memory, LSTM) 기법이 Hochreiter and Schmidhuber(1997)에 의해 개발되었다[12]. 이 기법은 일반적인 RNN의 단점인 오차 경사의 기울기 소실(Gradient vanishing) 문제를 해결할 수 있어 시계열 자료 예측에 유리한 것으로 알려져 있다[13].

LSTM 모형의 주요 메커니즘은 시간에 따라 상태를 유지할 수 있는 메모리 셀과 셀 안 또는 셀 바깥으로 데이터의 흐름을 조절하는 3개의 비선형 게이트이다. LSTM 모형 구조에 대한 모식도는 다수의 참고문헌에 주어져 있어서[4-7] 여기서는 생략한다. LSTM의 첫 번째 단계는 셀 상태에서 전단계 상태 h_{t-1} 와 새로운 입력 x_t 을 받아서 어떤 정보를 버릴 것인지 결정하는 것으로서 망각게이트(forget gate layer) f_t 에 의해 이루어진다. 두 번째, 입력 셀 상태에서 어떤 새로운 정보를 저장할 것인지 결정하는 것으로서, 입력게이트(input gate) i_t 에 의해 수행되며, 2단계로 구분한다. 먼저, 시그모이드 함수가 업데이트할 값을 결정하고, 다음으로, 하이퍼탄젠트(hyper tangent, tanh) 함수를 이용해 새로운 셀 상태(후보 셀 상태) \tilde{C}_t 를 생성한다. 마지막 단계인 출력게이트(output gate) o_t 는 시그모이드 함수를 이용하여 무엇을 출력할지 결정하게 된다. 그리고 하이퍼탄젠트 함수를 이용하여 현재 시점의 출력값 h_t 을 업데이트하게 된다.

2.3 모형 평가지수

모의결과의 정확도를 정량화하기 위해 평균 제곱근 오차 (Root Mean Square Error, RMSE), N-S 효율계수 (Nash-Sutcliffe Efficiency, NSE), 첨두유량오차 (Peak flow Error, PE) 같은 평가지수를 사용하였으며 관련 수식은 다수의 선행연구에[3-6] 나타나 있다. RMSE 값은 0에서 1까지 분포하며, '0'은 실측치와 모의치 간의 완전한 적합을 뜻하며, '1'은 그와 반대이다.

NSE는 음의 무한대에서 1까지 범위의 값을 가지며, 관측값과 산정값 간의 정확성을 검증하는데 주로 사용되는 계수로 1에 근접할수록 우수한 결과, 1이면 가장 이상적인 결과이다. PE는 모의치와 실측치의 차에 대한 실측치의 백분율로 나타내며, 0에 가까울수록 우수한 결과로 평가한다.

3. 모형 구축

3.1 대상유역

곡교천은 삼교호로 유입하는 제1지류로서 유역 면적은 545.1 km², 유로 연장은 50.6 km, 하천 경사는 1/1,675 ~ 1/4,148로 완만하며, 유역의 평균 고도는 103.8(EL.m)으로 전체적으로 200(EL.m) 이하의 지역이 85.4% 이상을 차지하는 비교적 평탄한 지형을 이루고 있다[14-15]. Fig. 1은 곡교천 유역의 주요 관측소의 위치를 보여주고 있으며, 본류에는 상류로부터 한내다리, 충무교, 강청 관측소가 위치하고 있으며, 지류인 매곡천에 매곡, 온천천에 온천 관측소가 위치하고 있다.

3.2 시나리오 설정 및 입력자료 구성

2011년부터 2019년까지 시간 단위로 측정된 강우량과 수위의 시계열 자료를 입력자료로 이용하였으며, 국가수자원정보시스템[16]에서 수집하였다. 자료의 특성에 따른 예측의 변화를 파악하기 위해서 2개의 시나리오를 설정하였으며, 시나리오 1은 강우량 5곳과 유량 4곳의 총 9열의 시계열 자료를, 시나리오 2는 유량 4곳의 시계열 자료를 모형의 입력자료로 사용하였다.

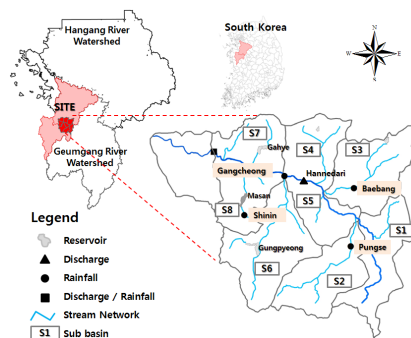


Fig. 1. Study site and its sub-basins.

가용한 자료를 세 그룹으로 구분하여, 각각 학습(Train), 보정(Validation), 검증 (Verification) 자료로 사용하였다. 2011년부터 2016년까지, 2017년과 2019년 자료는 각각 학습, 보정에 사용하였으며, 검증 자료는 전기간 자료 중에서 가장 큰 규모의 강우 3건을 추출하였으며, 학습 및 보정 자료에서 제외하였다. 본 연구에서 구축한 모형은 전처리, 메인 엔진, 후처리로 구성되었으며, 구글에서 제공하는 오픈 소스 소프트웨어 라이브러리 텐서플로와 파이썬 언어를 사용하여 구축하였다. 파이썬에 기초한 Pandas, Numpy, Scikit-learn과 같은 라이브러리를 활용하여 입력자료를 업로드하여 모형에 입력하기에 적합한 형태로 가공하는 전처리 작업부터 정규화, 모형 결과에 대한 평가지수 산정, 결과에 대한 그래프를 포함한 후처리 작업을 수행한다.

3.3 모형 보정 결과

학습된 LSTM 모형으로 보정 자료를 이용하여 시나리오별로 5개의 선행시간(T+1 ~ T+5)에 대한 예측을 수행하고 그 성능을 평가하였다. Table 1은 선행시간에 따른 RMSE와 NSE 분석 결과를 나타내고 있다.

시나리오 1은(SC1), 선행시간 T+1, T+3, T+5에서 RMSE가 각각 6.7 m³/s, 10.0 m³/s, 14.7 m³/s로 선행시간이 길어짐에 따라 예측 오차가 증가하였으며, NSE는 각각 0.9971, 0.9885, 0.9667로 예측 정확도가 감소하였다. 시나리오 2의 경우(SC2)도 값의 차이는 다소 있으나, 선행시간이 길어질수록 오차가 증가하고, 정확도는 감소하는 시나리오 1과 유사한 경향을 나타내었다. 유량과 강우량 자료를 동시에 사용하는 시나리오 1의 경우, 유량 자료 만을 사용하는 시나리오 2에 비하여 모든 선행시간에서 예측 성능이 우수하였으며, 단시간(T+1 ~ T+3) 예측에서 시나리오별 성능의 차이는 크지 않았으나, 장시간(T+4, T+5) 예측에는 성능의 차이가 상당히 있는 것으로 조사되었다.

Fig. 2은 선행시간 T+1, T+3, T+5에 대한 예측 결과를 산포도를 이용하여 비교하였다. Fig. 2(a), (c)와 (e)는 시나리오1에 대한, Fig. 2(b), (d) 와 (f)는 시나리오2에 대한 예측 결과이다. 선행시간이 T+1일 때 저유량에서 고유량까지 중심대각선에서 이격한 점이 거의 없이 안정적이며, 시나리오별로 차이가 없는 것으로 나타났다. 선행시간이 T+3, T+5로 길어지면서 중심대각선에서 이격되는 점이 점점 많아지는 것으로 나타났으며, 특히 저유량부터 중규모 유량까지 구간에서 중심선 아래에 더 많은 점이 분포하는 것으로 보아 모의값이 실측값을 과소

평가하는 것으로 분석되었다. 선행시간이 T+5와 같이 긴 경우 중심선에서 이격한 점이 시나리오 1이 시나리오 2보다 상대적으로 적은 것으로 보아 시나리오 1의 예측 성능이 나은 것으로 분석되었다. 시나리오별 차이는 크지 않았으나, 강우량과 유량 자료를 동시에 사용하는 시나리오 1의 결과가 유량 자료만을 사용하는 시나리오 2보다 모든 선행시간대에서 정확도가 우수한 것으로 분석되었다.

Table 1. Summary of model results at leading time steps T+1 ~ T+5 for validation phase.

Time step	Scenario	RMSE (m ³ /s)	NSE
T+1	SC1	6.70	0.9971
	SC2	6.76	0.9970
T+2	SC1	8.48	0.9932
	SC2	8.75	0.9925
T+3	SC1	10.03	0.9885
	SC2	10.72	0.9861
T+4	SC1	11.90	0.9812
	SC2	13.44	0.9738
T+5	SC1	14.71	0.9667
	SC2	17.37	0.9493

3.4 모형 민감도 분석 결과

3.4.1 층의 깊이가 성능에 미치는 영향 분석

은닉층은 입력층과 출력층 사이에 있는 층으로, 이 층은 모든 입력노드부터 입력값을 받아 가중합을 계산하고, 이 값을 전이함수에 적용하여 출력층에 전달한다. 은닉층의 깊이를 2, 4, 6층으로 증가 시키면서(L2, L4, L6), 선행시간 1시간부터 5시간까지 (T+1 ~ T+5) 예측을 수행하였으며, 그 결과를 Fig. 3에 도시하였다. 평가 지표로 RMSE를 사용하였으며, 도출된 결론은 다음과 같다.

첫째, 층의 깊이가 깊은 경우(L6), 모든 선행시간대에 걸쳐 예측에 안정성을 보이나, 비교적 단시간(T+1 ~ T+3) 예측에는 다른 은닉층의 경우(L2, L4)보다 정확도가 낮았으며, 둘째, 층의 깊이가 가장 얇은 경우(L2), T+1, T+2 와 같은 짧은 시간 예측에 우수한 결과를 주었으나 T+5와 같은 긴 시간 예측에 정확도는 가장 낮았으며, 셋째, 층이 4개인 경우(L4)가, 모든 선행시간대에 대하여 전반적으로 우수한 결과를 주었다. 또한, 예측 선행시간이 길어질수록 은닉층의 깊이와 무관하게 정확도가 떨어지는 것으로 분석되었다.

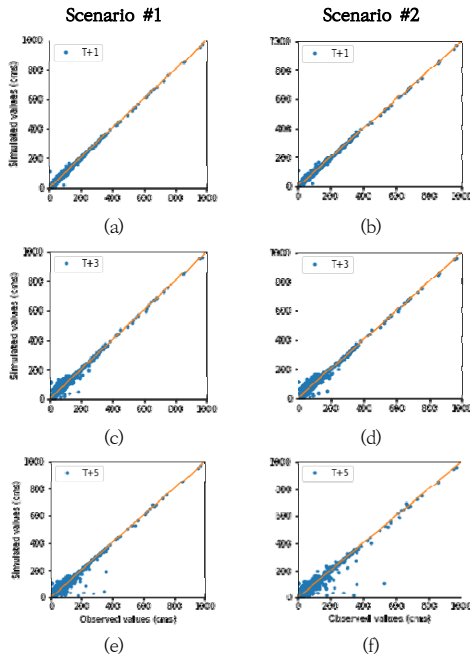


Fig. 2. Scatter plot at leading time steps T+1, T+3, and T+5 for scenarios 1 (a, c, e) and 2 (b, d, f) in validation phase

3.4.2 시퀀스길이에 따른 예측 성능 분석

시퀀스길이(Sequence Length, SL)는 모형이 어떤 시점의 자료를 학습하는 데 과거 몇 시간(또는 몇 단위) 길이의 자료를 사용할 지를 결정한다. 학습에 사용하는 입력자료의 길이 즉, 시퀀스 길이가 모형의 성능에 미치는 영향을 파악하고자 시퀀스길이를 3시간(SL=3), 5시간(SL=5), 7시간(SL=7)로 증가시키면서 모형을 실행하였다. 기본 자료가 1시간 단위로 구성되어 있기 때문에 시퀀스길이가 3시간일 경우 과거 2시간 전(t-2) 시간부터 현재 시간(t)까지 3시간 자료가 학습에 활용되었음을 의미한다. 은닉층의 개수는 앞서 수행한 층의 깊이에 따른 성능 분석 결과에 따라 가장 우수한 성능을 나타내는 층 4개로 고정하였다. Fig. 4에서 시퀀스길이가 3시간인 경우 선행시간 1시간부터 5시간(T+1 ~ T+5)까지 예측을 수행한 결과, RMSE는 각각 8.9, 12.3, 17.8, 24.1, 30.5로, 선행시간이 길어지면 오차가 증가하여 예측 정확도가 떨어지는 것으로 분석되었다. 이런 경향은 시퀀스길이가 5시간, 7시간 경우에 대하여도 유사한 것으로 나타났다. 한편, 선행시간 1시간(T+1) 경우, 시퀀스길이 3시간, 5시간, 7시간에 대하여 각각 RMSE는 8.9, 13.6, 15.4 m³/s로 나타나, 시퀀스길이가 늘어 남에 따라 오차는 증

가하여 예측 정확도가 감소하는 것으로 확인되었다. 다른 선행시간 (T+2 ~ T+5)에 대해서도 유사한 결과를 주는 것으로 나타났다. 이는 시퀀스길이가 길어지면, 먼 곳에 있는 자료는 학습할 기회가 적어져서 예측 정확도가 떨어지는 것으로 분석되었다. 결론적으로, 시퀀스 길이가 3시간일 때 모든 선행시간의 예측에 가장 우수한 성능을 나타내었으며, 반면에 시퀀스 길이가 7시간인 경우가 가장 낮은 성능을 나타내는 것으로 나타나, 시퀀스길이가 짧을 수록 예측 정확도는 높은 것으로 분석되었다.

3.4 모형 검증 결과

모형의 민감도 분석을 통해서 최적의 조건을 도출하였으며, 입력 시계열 자료는 강우량과 유량을 모두 사용하는 시나리오 2를 적용하여 모형을 구성하였다. 기간 내 첨두유량이 가장 커서, 학습할 기회가 적은 강우 사상 3개를 추출하였으며, 이 자료는 학습 과정에서 제외하였다.

선행시간 1시간부터 5시간까지 모의 결과를 RMSE, NSE 및 PE 지표를 이용하여 나타내었으며 Table 2와 같다. Fig. 5은 각 강우 사상에 대한 강정 지점의 유량 예측 결과를 선행시간별로 나타내었다.

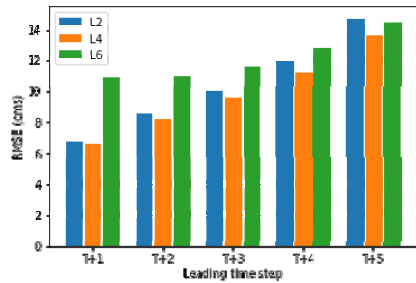


Fig. 3. Performance at leading time steps T+1 through T+5 based on different numbers of hidden layers, where L2 denotes two hidden layers.

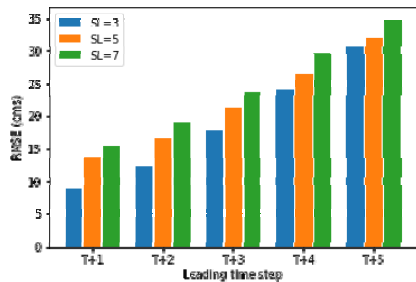


Fig. 4. Performance at leading time steps T+1 through T+5 based on different numbers of sequence lengths, where SL=3 denotes three sequence length.

Fig. 5(a), (b), (c)는 Table 2의 강우사상 FE1, FE2, FE3에 대한 수문곡선이며 이때 각 강우 사상의 첨두유량은 각각 1,335 m³/s, 963 m³/s, 1,005 m³/s이다. 강우사상1(FE1)에서 선행시간이 T+1에서 T+5까지 길어질수록 RMSE는 76 m³/s에서 202 m³/s로 오차가 증가하였으며, NSE는 0.97에서 0.68로 모형의 예측 정확도가 감소하였다. 다른 2개 강우 사상(FE2, FE3)에 대해서도 오차 및 정확도의 차이는 다소 있으나 이러한 경향은 유사한 것으로 조사되었다. 또한, 선행시간 T+1, T+2일 때 전 강우 사상에 걸쳐 NSE가 높았으나, T+3 이상일 때 NSE가 상당히 낮아지는 것으로 분석되었다.

첨두유량의 예측 정확도를 평가하기 위해서 평가 지표 PE(첨두유량 오차)를 사용하였다. 선행시간 T+1일 때 -2.9 ~ -4.3 % 범위로서 평균 -3.4 % 의 높은 정확도를 보였으며, 모두 마이너스로 나타나 모의 첨두유량이 관측 유량을 과소 추정하는 것으로 분석되었다. 이는 타 연구 결과와 유사한 것으로 조사되었다[17]. 또한, 모든 연구 결과와 유사한 것으로 조사되었다[17]. 또한, 모든 강우 사상에서 선행시간이 길어짐에 따라 PE가 점차적으로 음수에서 양수로 바뀌는 경향이 있으며 이 과정에서 선행시간이 긴, 가령 T+3 이상일 때 예기치 않게 PE가 절대값 기준으로 T+1, T+2보다 낮아지는 현상이 나타나게 된다. 그러나 이것은 오차가 낮아져서 정확도가 높아졌다고 볼 수 없으며, 이유는 RMSE, NSE와 같은 평가 지표 모두에서 오차가 크다는 것을 확인할 수 있기 때문이다.

Table 2. Performance of LSTM model for flood forecasting at leading time steps from T+1 to T+5 for thr flood events. (test phase)

Flood event	Time step	RMSE (m ³ /s)	NSE	PE(%)
FE1	T+1	76	0.97	-4.3
	T+2	93	0.95	-3.1
	T+3	125	0.90	0.9
	T+4	161	0.81	10.5
	T+5	202	0.68	21.6
FE2	T+1	73	0.95	-3.0
	T+2	99	0.89	-4.4
	T+3	124	0.80	-2.9
	T+4	142	0.71	-3.2
FE3	T+1	162	0.61	5.6
	T+1	88	0.96	-2.9
	T+2	88	0.96	1.6
	T+3	105	0.93	12.6
	T+4	133	0.85	22.3
	T+5	168	0.71	32.6

FE1, FE2, and FE3 are flood events with maximum flow of 1335, 963, and 1005 m³/s, respectively

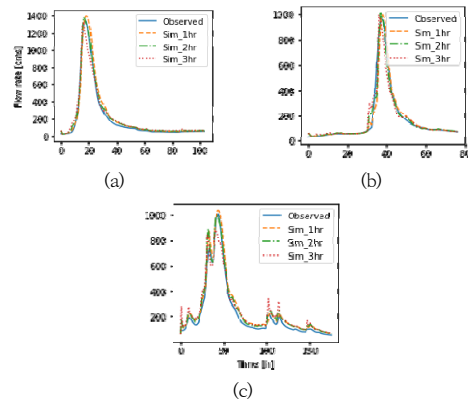


Fig. 6. Comparison of observed and simulated hydrographs at leading time steps T+1, T+2, and T+3 for three flood events. (test phase)

4. 결론

삼교호로 유입하는 곡교천 하류 강청 지점의 홍수량을 예측하기 위해서 LSTM모형을 구축하여 구글의 딥러닝 오픈소스 라이브러리 Tensorflow를 활용하여 LSTM 인공신경망 모형을 구축하였다. 모형 매개변수는 민감도 분석을 통해 은닉층의 개수는 4개, 시퀀스길이는 3시간, 학습율은 0.01, 학습횟수는 2000회로 하여 2011년부터 2019년까지 시간 단위의 강우량과 유량 자료를 이용하여 유출 모의를 수행하였으며 도출된 결론은 다음과 같다.

1. 입력자료의 특성에 따른 예측의 변화를 파악하기 위해서 2개의 시나리오를 설정하였으며, 단시간 예측에서 시나리오별 성능의 차이는 크지 않았으나, 장시간 예측에는 유량과 강우량을 모두 사용한 시나리오 1이 유량만 사용한 시나리오 2에 비하여 성능이 우수한 것으로 조사되었다.
2. 은닉층의 깊이가 성능에 미치는 영향을 파악하기 위해 층의 깊이를 증가시키면서, 선행시간 1시간 부터 5시간까지 예측을 수행하였다. 층이 4개인 경우가, 모든 선행시간대에 대하여 전반적으로 우수한 결과를 주었다. 또한, 층이 깊은 경우 전 선행 시간에 걸쳐 안정적인 결과를 주기는 하나, 단시간 예측에 있어서 정확도가 층이 다른 얇은 층에 비하여 낮아지는 것으로 분석되었다.
3. 시퀀스길이가 모형의 성능에 미치는 영향을 파악하고자 시퀀스길이를 3시간, 5시간, 7시간으로 증가

시키면서 모형을 실행하였으며, 시퀀스 길이가 늘어 남에 따라 오차는 증가하여 예측 정확도가 감소하였다. 이는 시퀀스길이가 길어지면 현재 시간에서 과거 먼 곳에 있는 자료는 학습할 기회가 적어져서 예측 정확도가 떨어지는 것으로 분석되었다. 시퀀스길이가 3시간일 때 모든 선행시간의 예측에 가장 우수한 성능을 나타내었다.

4. 모형 검증에서 선행시간이 T+1에서 T+5까지 길어질수록 모든 강우 사상에서 오차는 증가하였으며, 예측 정확도는 감소하였다. 또한, 선행시간 T+1, T+2일 때 모든 강우 사상에 걸쳐 예측 정확도가 높았으나, T+3 이상일 때 정확도가 상당히 낮아지는 것으로 조사되어, 선행시간 T+2까지 예측 결과를 사용하는 것이 바람직한 것으로 분석되었다.

References

- [1] H. Jun, J. Lee, "A Methodology for Flood Forecasting and Warning Based on the Characteristic of Observed Water Levels Between Upstream and Downstream", *Journal of Korea Society of Hazard Mitigation*, Vol.13, No.6, pp.367-374, 2013.
DOI: <http://dx.doi.org/10.9798/KOSHAM.2013.13.6.367>
- [2] M. Moradi, T. Lee, "Comparison of Optimization Algorithms in Deep Learning-Based Neural Networks for Hydrological Forecasting: Case Study of Nam River Daily Runoff", *Journal of Korea Society of Hazard Mitigation*, Vol.18, No.6, pp.377-384, 2018.
DOI: <http://dx.doi.org/10.9798/KOSHAM.2018.18.6.377>
- [3] S. Jung, D. Lee, K. Lee, "Prediction of River Water Level Using Deep-Learning Open Library", *Journal of Korea Society of Hazard Mitigation*, Vol.18, No. 1, pp.1-11, 2018(a).
DOI: <http://dx.doi.org/10.9798/KOSHAM.2018.18.1.1>
- [4] S. Jung, H. Cho, J. Kim, G. Lee, "Prediction of Water Level in a Tidal River Using a Deep-learning Based LSTM Model", *Journal of Korea Water Resources Association*, Vol.51, No.12, pp.1207-1216, 2018(b).
DOI: <http://doi.org/10.3741/JKWRA.2018.51.12.1207>
- [5] M. K. Park, Y. S. Yoon, H. H. Lee, J. H. Kim, "Application of Recurrent Neural Network for Inflow Prediction into Multi-purpose Dam Basin", *Journal of Korea Water Resources Association*, Vol.51, No.12, pp.1217-1227, 2018.
DOI: <http://doi.org/10.3741/JKWRA.2018.51.12.1217>
- [6] D. Zhang, G. Lindholm, H. Ratnaweera, "Use Long Short-Term Memory to Enhance Internet of Things for Combined Sewer Overflow Monitoring", *J. of Hydrology*, Vol.556, pp.409-418, 2018(a).
DOI: <http://doi.org/10.1016/j.jhydrol.2017.11.018>
- [7] D. Zhang, E.K. Holland, G. Lindholm, H. Ratnaweera, "Hydraulic Modelling and Deep Learning Based Flow Forecasting for Optimizing Inter Catchment Wastewater Transfer", *J. of Hydrology*, Vol.557, pp. 792-802, 2018(b).
DOI: <http://doi.org/10.1016/j.jhydrol.2017.11.029>
- [8] Y. Qi, Z. Zhou, L. Yang, Y. Quan, Q. Mao, "A decomposition ensemble learning model based on LSTM neural network for daily reservoir inflow forecasting", *Water Resources Management*, 2019.
DOI: <https://doi.org/10.1007/s11269-019-02345-1>
- [9] X. He, J. Luo, G. Zuo, J. Xie, "Daily runoff forecasting using a hybrid model based on variational mode decomposition and deep neural networks", *Water Resources Management*, 33: 1571-1590, 2019.
DOI: <https://doi.org/10.1007/s11269-2183-x>
- [10] D. Zhang, N. Martinez, G. Lindholm, H. Ratnaweera, "Manage sewer in-line storage control using hydraulic model and recurrent neural network", *Water Resources Management*, 2018(c).
DOI: <https://doi.org/10.1007/s11269-018-1919-3>
- [11] I-F. Kao, Y. Zhou, L-C. Chang, F-J. Chang, "Exploring a long short-term memory based on encoder-decoder framework for multi-step ahead flood forecasting", *J. of Hydrology*, 2020.
DOI: <https://doi.org/10.1016/j.jhydrol.2020.124631>
- [12] S. Hochreuter and J. Schmidhuber, "Long Short-Term Memory", *Neural Computation*, Vol.9, No.8, :pp.1735-1780, 1997.
DOI: <https://doi.org/10.1162/neco.1997.9.8.1735>
- [13] Q.-K. Tran, S.-K. Song, "Water Level Forecasting based on Deep Learning: A Use Case of Trinity River-Texas-The United States", *Journal of Korean Institute of Information Scientists and Engineers*, Vol. 44, No.6, pp.607-612, 2017.
DOI: <http://doi.org/10.3741/JKWRA.2018.51.12.1207>
- [14] Ministry of Construction and Transportation(MOCT), "River Master Plan Report on Sapgyechun Watershed", 2012.
- [15] Ministry of Environment(MOE), "Annual Report on Hydrological Survey", Geum River Flood Control Office, Gongju City, Chungnam Procince, 2019.
- [16] Ministry of Land and Transportation(MOLT), "Water Resources Management Information System(WAMIS)", <http://www.wamis.go.kr>
- [17] X.-H. Lee, H.V. Ho, G. Lee, S. Jung, "Application of Long Short-Term Memory (LSTM) Neral Network for flood forecasting", *Water*, 11, 2019.
DOI: <http://doi.org/10.3390/w11071387>

황 병 기(Byung-Gi Hwang)

[종신회원]



- 1988년 12월 : 미국 버지니아공대 토목공학과 (수자원공학 석사)
- 1995년 5월 : 미국 버지니아주립대학교 토목환경공학과 (수질모델링 박사)
- 1995년 6월 ~ 1995년 12월 : 미국 버지니아주립대 연구원
- 1996년 3월 ~ 현재 : 상명대학교 건설시스템공학과 교수

<관심분야>

수질관리, 비점오염원모델링, 수질환경영향평가