

딥러닝 오픈 라이브러리를 이용한 하천수위 예측

Prediction of River Water Level Using Deep-Learning Open Library

정성호* · 이대업** · 이경상***

Jung, Sungho*, Lee, Daeep**, and Lee, Kyoungsang***

Abstract

This research aims to predict the water level using deep-learning algorithm. To achieve this goal, we applied the TensorFlow, a deep-learning open source library to predict the water level of Okcheon station in the Guem-river. For model training and prediction, two hourly water level data sets of the 3 water level stations (Sutong, Hotan, Songcheon) are prepared: from 2002 to 2013 (training); from 2014 to 2016 (prediction). Even if many of physical data are necessary to understand water cycle system, in particular, model rainfall-runoff process, we used only upstream observed water level information to predict downstream water level using multi-linear regression and Long Short-Term Memory(LSTM) models based on the TensorFlow. The results showed that the weights(or regression coefficients) of multi-linear regression model were very fluctuated due to training trials, then, the predicted high water level were too much underestimated than the observations. On the other hand, the LSTM model predicted the downstream water level very stably regardless of water level height for the study period because sequence length of the LSTM memorize antecedent water level information for model training and updating.

Key words : Deep-learning, TensorFlow, Multi Linear Regression, LSTM, Water Level

요 지

본 연구에서는 딥러닝 알고리즘을 이용한 관측자료 기반의 하천수위 예측을 목적으로 하고 있으며, 금강유역 내 대청댐 상류부에 위치한 옥천 관측소 지점의 수위 예측을 위하여 구글의 딥러닝 오픈소스 라이브러리인 텐서플로를 활용하여 다중선형회귀모형과 LSTM 인공신경망 모형을 구축하였다. 모형의 학습과 예측을 위하여 입력자료(금강 본류의 수통과 호탄 관측소, 금강 지류 초강의 송천 관측소 수위)와 출력자료(옥천 수위관측소 수위)의 2002년부터 2016년까지의 15년간의 시수위자료를 획득한 후, 2002년부터 2013년까지의 12년간 수위자료를 이용하여 개별 모형의 학습을 진행하고 2014년부터 2016년까지 3개년에 대한 수위예측을 수행하였다. 본 연구는 하천 수위에 영향을 미치는 수많은 물리적 매개변수를 배제하고 오직 상류의 관측 수위만을 사용하였으며 학습된 모형의 수위예측 결과를 비교·분석한 결과, 다중회귀분석모형의 경우 최적가중치 추정을 위한 반복시행횟수 증가에 따라 가중치 변동성이 매우 크게 발생하여 수위예측의 오차 변동성이 매우 크고, 고수위에서 과소 추정된 반면, 선행시간의 수위 정보를 학습에 반영하여 수위 예측을 수행하는 LSTM 모형은 매우 안정적이고, 우수하게 수위 예측결과를 제공해주는 것으로 분석되었다.

핵심용어 : 딥러닝, 텐서플로, 다중선형회귀모형, LSTM, 하천수위

1. 서 론

하천의 수위 및 유량 정보는 가뭄과 홍수 등 수재해 예방과

수자원 개발 및 이용, 생태계 보존 등 다양한 목적을 위한 기초자료이며, 이에 대한 정확한 계측 및 예측이 요구된다. 특히 우리나라의 홍수예보의 경우, 「하천법시행규칙 제23조

*정회원, 경북대학교 건설방재공학과 석사과정(E-mail: wjtdjdgh1547@gmail.com)

Member, Master's Course, Dept. of Construction and Disaster Prevention Engineering, Kyungpook National University

**교신저자, 정회원, 경북대학교 건설방재공학과 박사과정(Tel: +82-54-535-1360, Fax: +82-54-535-1360, E-mail: hydroeop@gmail.com)

Corresponding Author, Member, Ph.D. Candidate, Dept. of Construction and Disaster Prevention Engineering, Kyungpook National University

***정회원, 경북대학교 건설방재공학과 석사과정(E-mail: lgs_1119@naver.com)

Member, Master's Course, Dept. of Construction and Disaster Prevention Engineering, Kyungpook National University

제2항]에 의거하여 홍수예보 발령지점과 당해 지점의 주의 보수위 및 경보수위를 미리 정하여 고시하고 있다. 홍수예보를 발령하는 지점의 수위가 계속 상승하여 주의보 경계홍수위(계획홍수량의 50%가 흐를 때의 수위)를 초과할 것이 예상되는 경우 홍수주의보를 발령하고, 홍수예보를 발령하는 지점의 수위가 계속 상승하여 경보위험 홍수위(계획홍수량의 70%가 흐를 때의 수위)를 초과할 것이 예상되는 경우 홍수경보를 발령하고 있다.

이러한 홍수예보의 일반적인 절차는 강수량관측소 및 수위관측소에서 매 10분단위로 수문자료 수집(한국수자원공사 관측자료 포함)하고 강우에 따른 유출량 계산 및 댐 저수량을 고려하여 주요 지점의 수위와 홍수 규모를 판단한 후 기상과 하류수위를 감안하여 댐 예비방류 등 홍수량 조절하도록 하고 있다. 즉, 일반적으로 홍수예보는 강우-유출 모형을 이용하여 예측 강우를 입력 자료로 하여 수리모형 등을 이용한 하도추적을 통해 예측수위를 추정한 후 수행되게 된다.

하천의 수위변화 및 강우-유출 현상은 매우 비선형적(non-linear)이어서 단위도(unit hydrograph) 등과 같은 선형모형(linear model)을 이용하여 모의하거나 예측하는 것은 한계가 있다. 이에 다양한 개념적(conceptual), 물리적(physics-based) 수문모형이 개발되었으며, 이들 모형은 유출생성에 필요한 수문학적 특성들을 매개변수화(parameterization)하여 강우-유출 해석을 수행하게 된다. 그러나 아무리 정교한 모형일지라도 강우-유출에 필요한 모든 변수(기상학적, 지형학적, 지질학적)를 반영할 수 없으며, 정교한 모형일수록 매개변수의 양이 증가하고, 제대로 검 · 보정되지 않은 매개변수로 인해 모의결과에 불확실성을 초래할 수 있으며, 모의시간이 길어짐에 따라 홍수예보에 필요한 선행시간 확보에 제약이 있다. 실제로 아직까지도 우리나라 하천의 주요지점의 홍수예보에 물리적 분포형 수문모형대신 개념적 수문모형과 하도추적 모형이 결합된 시스템이 실무에서 사용되는 이유는 여기서 기인한다고 할 수 있다.

하천수위 변화에 영향을 미치는 모든 인자가 고려된 물리적 모형을 구축하여 모든 영향인자들과 수위(또는 유출)와의 인과관계 또는 상관관계를 분석하는 것은 실무적으로 불가능한 영역으로 여겨졌다. 그러나 방대한 디지털 자료의 수집이 가능해지고, 이와 같은 빅데이터를 처리하는 컴퓨터 기술이 발달함에 따라 인공지능(Artificial Intelligence) 또는 인공신경망(Artificial Neural Network) 등과 같이 입 · 출력 자료 기반의 black-box 형태의 모형을 이용한 수문해석 분야가 재조명받고 있다.

인공지능은 인간과 같이 생각하고 학습하고 판단하는 인간의 사고회로를 모방하여 만든 컴퓨터 모형이며, 인공신경망은 생물학적 뉴런(neuron)의 동작원리와 뉴런간의 연결 관계를 모방하여 만든 알고리즘이다. 인공신경망은 McCulloch and Pitts(1943)가 수학적 기법으로 처음 신경망의 기초논리

를 연구하였고, 이후 Rosenblatt(1957)가 많은 기계학습(Machine Learning, 이하 머신러닝)과 심층학습(Deep Learning, 이하 딥러닝)의 기초가 되는 신경망 알고리즘인 퍼셉트론(Perceptron)을 제안하였다. 그러나 그 당시 연구됐었던 신경망으로는 복잡한 논리회로 연산이 힘들다는 등의 이유로 Minsky and papert(1969)의 머신러닝에 관한 연구 이후로 관련 연구가 침체되어 암흑기를 맞았으나 1980년대 Rumelhart et al.(1986)에 의해 출력값을 역으로 전파하여 오차의 경사를 줄여나가는 역전파(back propagation) 알고리즘이 개발되면서 다시 각광받기 시작하였다. 이후 빅데이터 수집 및 처리에 관한 다양한 기술들이 개발됨과 동시에 사람의 학습능력을 모방하는 딥러닝 기법이 소개되고, 순환신경망(Recurrent Neural Network, 이하 RNN), 합성곱신경망(Convolutional Neural Network, 이하 CNN) 등의 새로운 알고리즘의 개발로 영상인식, 음성인식, 자율주행자동차 개발 등 다양한 분야에서 그 적용성을 인정받고 있다.

국외 수문 · 수자원 분야에서의 인공신경망 기법이 적용된 연구로 Thirumalaiah and Deo(1998)은 신경망 모형의 3가지 알고리즘을 이용하여 홍수기 시 홍수위를 예측하고 분석을 통해 종속상관알고리즘이 우수한 결과를 도출함을 확인하였으며, Dawson and Wilby(1998)은 강우-유출 모델링을 위해 역전파 알고리즘 신경망을 이용하여 데이터 최적화에 따라 홍수예측의 정확도가 향상되는 것을 입증하였다. 또한 Coulibaly and Anctil(1999)은 효율적 저수지운영을 위한 단기 유입량 예측을 위해 RNN을 적용하였으며 RNN이 효과적인 실시간 수문학적 예측결과를 도출하는 것을 확인하였다. Adnan et al.(2012)은 복합 지하수 수위 예측을 위해 기계학습 알고리즘 중 SVM(support vector machine)과 ANN을 비교하였고 예측기간을 달리 하였을 때 두 알고리즘 모두 지하수예측에 정확한 결과를 도출함을 확인하였다.

국내에서는 Shin et al.(1998)이 기본적인 다층전방향신경망을 역전파법을 사용하여 장기 강우예측에 적용한 바 있으며, Ahn et al.(2000)은 면적강수량 산정에 있어 역전파 알고리즘을 사용한 다층신경망을 적용하여 RDS방법, Thiessen방법, 산술평균법 등 보다 실측치에 근사한 결과를 내었으며 Park et al.(2006)은 인공신경망 이론을 이용하여 강우-유출과정의 비선형 모형을 개발하였다. 인공신경망을 이용한 하천의 수위 및 유량 예측과 관련된 연구로 Son and Lee(2003)는 신경망 알고리즘 중 역전파 기법, Fletcher-Reeves, BFGS 알고리즘을 이용하여 강원도 원통 지역의 인북천에 대한 홍수위의 예측을 수행하고 실제 관측유량과의 비교를 통해 신경망 알고리즘에 의한 홍수위 예측이 우수한 결과를 도출하는 것을 확인하였다. Yoon et al.(2004)은 낙동강의 남강댐 유역을 대상으로 인공신경망의 역전파 기법을 사용하여 홍수기 댐 유입량을 예측하고 이를 통해 실시간 댐 유입량 예측의 가능성을 검토하였다. 또한 Kim(2005)은 낙동강의 위천 유역을 대상으로 EDRNNM을 개발하여 홍수위예측을

수행하고 FFNNM과의 비교를 통해 개발된 모델을 중소하천 유역에서 홍수위예측을 위한 우수한 모델로 판단하였으며, Yeo et al.(2010)은 감천유역의 선산 수위표지점을 대상으로 인공신경망과 유전자 알고리즘을 결합한 모델을 이용하여 단시간 수위예측을 수행하였다.

본 연구에서는 구글(Google)의 딥러닝 오픈소스 라이브러리인 텐서플로우(Tensorflow)의 다중선형회귀모델과 시계열 데이터처리에 강점을 가진 RNN-LSTM(Long Short-Term Memory)모델을 사용하여 대청댐 상류의 금강분류 옥천관측소의 하천수위 예측을 위해 인공신경망 모델을 구축하고 그 적용성을 평가하였다. 홍수기 시 고수위뿐만 아니라 갈수기 시 저수위 변동성을 예측하기 위해 모델의 학습(training) 및 예측(prediction)을 위한 입력 자료는 옥천관측소 상류에 위치한 수통 및 호탄관측소, 지류인 초강에 위치한 송천관측소의 2002년부터 2016년까지의 15년 전 기간에 대한 시수위 자료를 이용하였다. 2002년부터 2013년까지의 12년간 수위 자료를 이용하여 모델을 학습시킨 후 2014년부터 2016년까지의 옥천관측소에 대한 수위예측을 수행하였으며, 이를 실제 수위관측 자료와 비교·분석하여 각 인공신경망 모델의 재현성을 평가하였다.

2. 이론적 배경

2.1 딥러닝

컴퓨터 공학에서 딥러닝은 신경망 알고리즘을 주로 사용하는 머신러닝의 한 분야이며, 기술적으로 인공신경망이라는 표현이 더 적절하나 포괄적인 의미로 딥러닝으로 불리고 있다. 인공신경망은 인간의 뇌구조의 신경세포인 뉴런의 동작을 모방하여 고안되었으며, Fig. 1과 같이 가장 단순한 인공신경망 구조는 입력층(input layer), 은닉층(hidden layer), 출력층(output layer)으로 구성되어 있다. 반면에 일반적으로

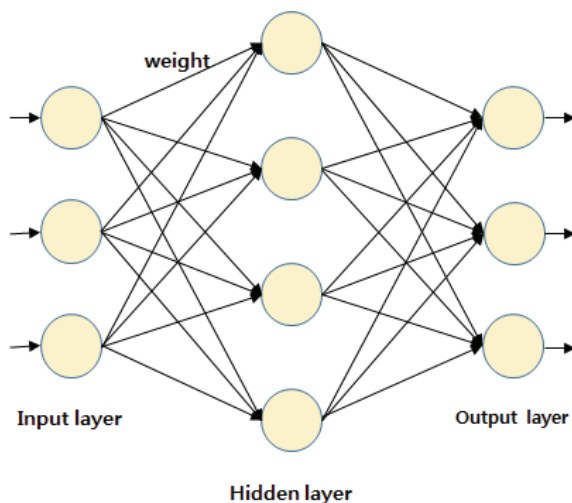


Fig. 1. Structure of Artificial Neural Network (Kwon et al., 2017)

딥러닝 모델의 구조는 다수의 은닉층이 존재하는 다층신경망(multi-layer neural network)으로 구성되어 있다. 각각의 데이터들은 자료 및 학습효과에 따라 가중치가 변화하며, 활성화 함수(activation function)에 의해 출력값이 도출된다. 이러한 활성화 함수는 Sigmoid(또는 Logistic) 함수, Softmax 함수, ReLU(Rectified Linear Unit) 함수 등 다양한 함수가 존재하며, 인공신경망 모델의 목적에 따라 함수의 이용도 달라지게 된다.

인공신경망은 지속적인 학습을 통해 가중치를 변화시키면서 적합한 모델을 구성할 수 있는 적응성(adaptability)과 다수의 퍼셉트론이 상호작용을 통해 불완전한 자료일지라도 학습과 의사결정을 할 수 있는 결점포용력(fault tolerance) 특징을 가지고 있다(Kwon et al., 2017). 그러나 심층구조의 신경망 모델의 경우, 매개변수들이 전역최적해(global optimum)를 찾지 못하고 국부최적해(local optimum)에 수렴하거나 학습 자료에 대하여 과하게 학습하여 실제자료에 대한 오차가 증가하는 과적합(overfitting) 문제가 발생할 수 있다. 그러나 현재 개발된 많은 딥러닝 알고리즘은 이러한 문제점을 해결하기 위해 다양한 최적화(optimization)기법 및 규제화(regularization), 가중치 감쇠(weight decay), 드롭아웃(drop-out) 등의 기법을 활용하고 있다.

2.2 텐서플로우

텐서플로우는 구글이 개발하고 배포한 오픈소스 라이브러리이며, 자료의 흐름그래프를 이용하여 수치계산을 수행할 수 있다. 2015년 구글은 머신러닝을 이용하고 싶은 개발자나 연구자들에게 누구나 텐서플로우를 사용할 수 있게 텐서플로우를 포함한 튜토리얼과 예제파일을 공개하였다

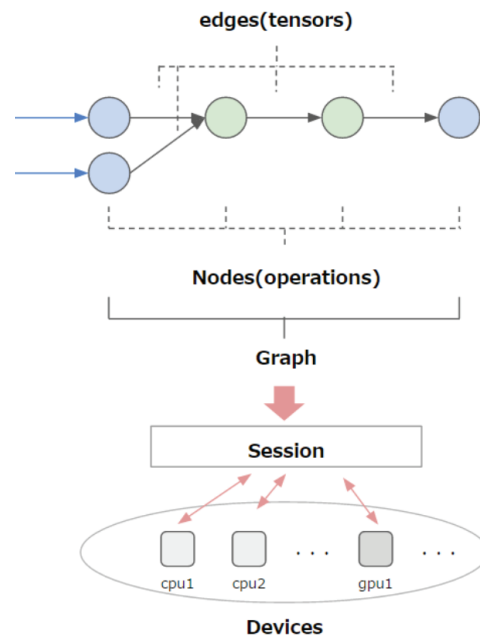


Fig. 2. Structure of TensorFlow (Ankit, S. 2017)

(<https://www.tensorflow.org/>). Fig. 2와 같이 기본적으로 텐서플로우에서 노드(node)는 수학적 계산, 데이터 입·출력, 그리고 데이터의 읽기·저장 등의 작업을 수행하며, 엣지(edge)는 노드들 간 데이터의 입·출력 관계를 나타내며, 동적 사이즈의 다차원 데이터 배열(=텐서)의 이동경로가 된다. 또한 텐서플로우는 CPU뿐만 아니라 GPU 활용도 가능하여 대용량 빅데이터 연산이 가능하며, Python과 C++뿐만 아니라 다양한 프로그램 언어의 지원도 가능하다 (Abadi et al., 2016).

2.3 다중선형회귀 모형

인공신경망 지도학습(supervised training)의 하나인 선형회귀(linear regression)는 종속변수와 한 개 이상의 독립변수의 선형상관관계를 분석하는 기법이다. 하나의 독립변수를 고려하려 종속변수와 상관관계를 분석하는 것을 단순선형회귀(simple linear regression), 두 개 이상의 독립변수를 고려하는 것을 다중선형회귀(multi linear regression)라고 한다. 종속 변수 y 와 i 개의 독립 변수 x_i 와의 다중선형회귀식은 Eq. (1)로 나타낼 수 있으며, $\beta_0, \beta_1, \dots, \beta_i$ 는 각 독립변수 x_i 의 회귀계수, ϵ_i 는 오차 변수로 관찰되지 않은 종속변수와 독립 변수 사이에 오차를 의미한다. 회귀계수를 통하여 각 주어진 독립변수가 종속변수에 미치는 영향력(가중치)를 유추할 수 있다.

$$Y_i = \beta_0 + \beta_1 \times x_1 + \beta_2 \times x_2 + \dots + \beta_i \times x_i + \epsilon_i \quad (1)$$

회귀계수 추정기법으로는 최소자승법(least squares method), 최대우도추정법(maximum likelihood estimation) 등이 있으며 본 연구에서는 최소자승법을 사용한다. 최소자승법 Eq. (2)와 같이 변수 x_i 와 회귀계수 β_i 인 방정식이 $f(x, \beta)$ 일 때 모형의 예측값과 실제 데이터와의 잔차(residual)제곱의 합 또는 평균이 최소화되도록 매개변수를 결정하는 방법이다.

$$\sum_i r^2 = \sum_i (y_i - f(x_i, \beta))^2 \quad (2)$$

인공신경망 모형에서 대상으로 하는 하루 하천수위에 영향을 줄 수 있는 요소들과 수위 사이의 관계식에 따른 예측자료를 생성할 수 있으며(Son and Lee, 2003), 관측자료 y 와 회귀식으로 이루어진 모형과의 오차를 구하고 경사하강법(gradient descent optimizer)을 활용하여 학습과정에서 오차를 줄여 나가며 최적의 가중치 β 값을 찾을 수 있다. 이때의 오차함수는 Eq. (3)과 같다.

$$Cost(\beta) = \frac{1}{m} \sum_{i=1}^m (y_i - f(x_i, \beta))^2 \quad (3)$$

2.4 LSTM 모형

RNN은 시계열 데이터처리에 특화된 알고리즘으로 알려져 있으며, 입력데이터와 전 단계에서 출력된 데이터를 동시에 이용하여 학습이 가능하며, Hochreiter and Schmidhuber (1997)는 일반적인 RNN의 장기 의존성 문제를 해결하기 위해 LSTM 기법을 제안하였다. LSTM은 여러 개의 게이트(gate)가 붙어있는 셀(cell)로 구성되며, 게이트는 셀 상태의 보호 및 제어를 위한 망각, 입력, 출력의 세 종류로 구분된다. LSTM은 일반적인 RNN의 고질적인 가중치 소실(vanishing gradient) 문제를 방지할 수 있어 시계열 자료 예측에 유리한 것으로 알려져 있다(Tran and Song, 2017).

Olah(2015)에 따르면 첫 번째 망각게이트(forget gate, f_t)는 전 단계 출력값인 h_{t-1} 과 새로운 값 x_t 를 통하여 0과 1을 사이의 값을 출력하며 이 정보를 유지할지 제거할지를 선택하게 된다.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (4)$$

여기서, σ 는 활성화함수를 나타내며 W_f 는 게이트의 가중치, b_f 는 편향값을 나타낸다.

다음 단계인 입력게이트(input gate, i_t)에서는 어떤 값들을 입력할지 결정하고 새로운 셀 상태를 갱신한다.

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (5)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (6)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (7)$$

여기서 C_t 는 셀 상태를 나타내며, \tilde{C}_t 는 활성화함수를 통해 만들어진 새로운 셀 상태, W_c 는 셀 상태의 가중치, b_c 는 편향값을 나타낸다.

마지막 단계인 출력게이트(output gate, o_t)는 무엇을 출력할지 결정하게 된다.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (8)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (9)$$

여기서 h_t 는 새로운 출력값을 나타낸다.

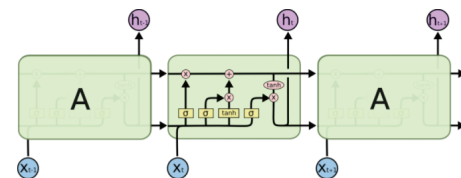


Fig. 3. LSTM Network (Olah, C. 2015)

3. 모형 적용 및 분석

3.1 대상유역 및 분석절차

본 연구는 금강 대청댐 상류를 대상으로 상류 3개 수위관측소(수통, 호탄, 송천)의 수위정보를 이용하여 하류 관측소(옥천)의 수위를 예측하기 위한 인공신경망 모형을 구축하고 그 재현성을 평가하는 것을 목적으로 하고 있다. Table 1은 대상 관측소에 대한 정보를 포함하고 있으며, 각 관측소에 대한 위치는 Fig. 4와 같다. 금강 본류의 상류로부터 수통, 호탄, 옥천 관측소가 위치해 있으며, 지류인 초강에 송천 관측소가 위치해 있다. 옥천관측소 상류에 위치한 3개의 관측소에 대해 2002년부터 2016년까지의 15년간의 시(hr)단위 수위자료를 사용하여 다중선형회귀분석과 LSTM을 활용하여 옥천 관측소의 수위의 학습 및 예측을 수행하였다. 모형의 학습을 위한 훈련자료로는 2002년부터 2013년인 12년간의 자료를 사용하였으며, 2014년부터 2016년까지의 3년간의 모형의 예측자료와 실제 수위관측 자료와 비교하여 각 인공신경망 모형의 적용성을 비교·분석하였다.

다중선형회귀분석 모형과 LSTM 모형의 학습속도(learning_rate)는 0.01, 반복시행횟수는 500번, 3000번, 10000번으로 3번씩 실행하였다. LSTM의 경우, Sequence length의 길이 3, 5, 7시간에 대한 3가지 경우와 은닉층의 개수 8, 10, 12 3개로 구성하여 조합에 따른 재현성을 비교·분석하였다. 최종적으로 Eqs. (10) and (11)의 RMSE(Root Mean Square

Error)와 NSE(Nash-Sutcliffe Efficiency) 평가지수를 사용하여 각 모형으로부터 예측된 3개년(2014년 ~ 2016년)의 실측수위자료와의 비교·분석을 통해 모형의 재현성을 평가하였다.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (O_i - P_i)^2} \quad (10)$$

$$NSE = 1 - \frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (O_i - \bar{O})^2} \quad (11)$$

여기서 O_i 는 관측소별 관측수위를 나타내며, \bar{O} 는 관측수위의 평균, P_i 는 모형별 예측수위를 나타낸다.

3.2 인공신경망 모형의 결과 비교·분석

3.2.1 다중선형회귀모형 분석결과

옥천 관측소를 기준으로 최상류에 위치한 수통 관측소의 수위자료를 x_1 , 호탄 관측소 수위자료를 x_2 , 송천 수위관측소 수위자료를 x_3 로 하여 다중선형회귀분석을 수행하였으며, 그 결과는 Table 2와 같다. 반복 수행횟수에 따른 RMSE 결과는 0.159~0.231의 범위로 나타났으며, 반복횟수의 증가

Table 1. Information of Water Level Stations

| Water level station | Location | | Elevation (EL.m) | Commencement of observation |
|---------------------|-----------|----------|------------------|-----------------------------|
| | Longitude | Latitude | | |
| Sutong | 127-35-37 | 36-02-56 | 146.130 | 1992-07-09 |
| Hotan | 127-39-00 | 36-07-55 | 113.960 | 1987-06-30 |
| Songcheon | 127-48-18 | 36-13-47 | 120.800 | 1992-07-09 |
| Okcheon | 127-40-04 | 36-14-22 | 88.288 | 1979-09-15 |

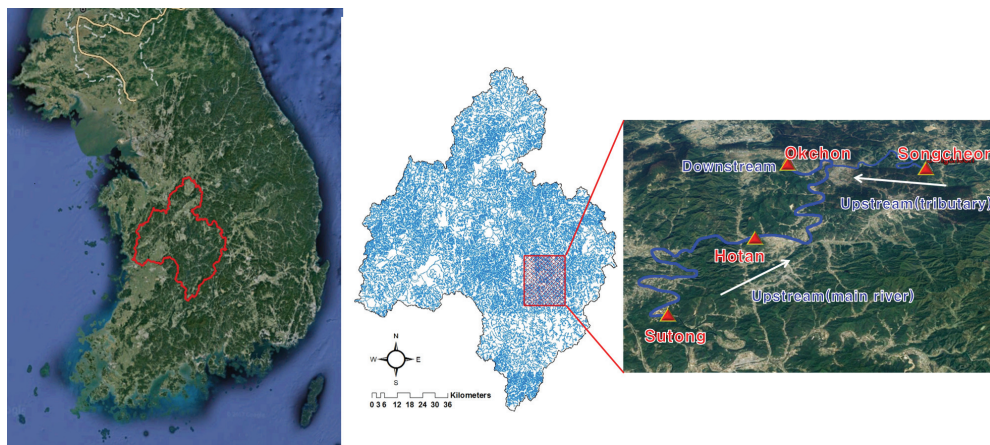
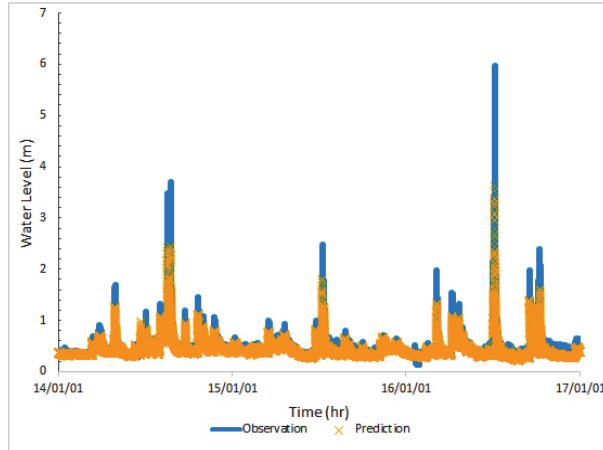


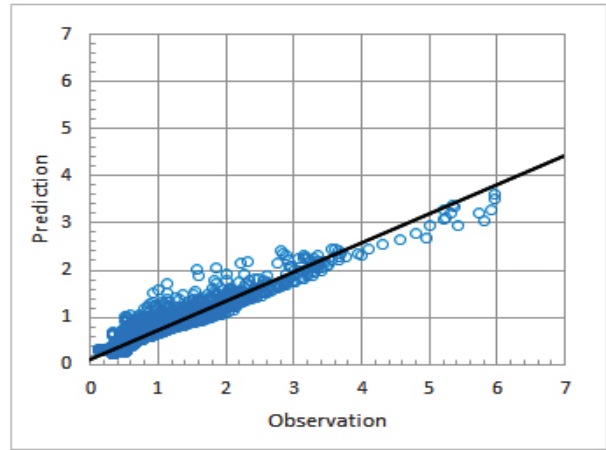
Fig. 4. Location of the Study Area and the Water Level Stations

Table 2. Model Performance Results of the Multi Linear Regression Model

| Case | iterations | RMSE | NSE |
|------|------------|-------|--------|
| M1 | 500 | 0.159 | 0.7112 |
| M2 | 1000 | 0.216 | 0.4660 |
| M3 | 3000 | 0.231 | 0.4065 |



(a) Water level time series



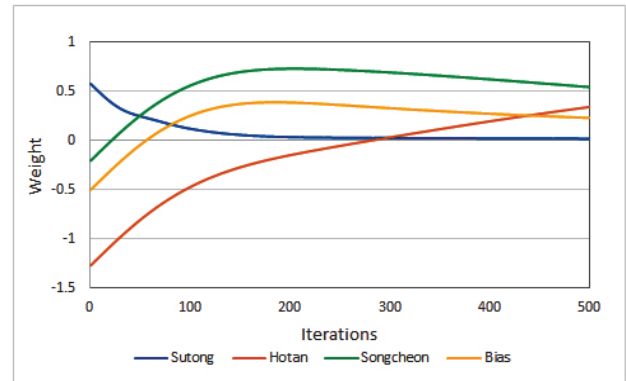
(b) Scatter plot of water level comparison

Fig. 5. Comparison of Water Level in Okcheon Station – Multi Linear Regression Model (Case M1)

에 따라 RMSE가 증가하는 것을 확인할 수 있다. 즉, 텐서플로우 다중선형회귀모형의 경우, 반복시행횟수에 따른 최적 가중치의 변동성이 매우 크며, 이에 따라 오차가 매우 민감하게 반응하는 것으로 분석되었다. NSE 결과의 경우, Case M1에서 비교적 높은 재현성을 보이는 것으로 나타났다.

실제로 Fig. 5와 같이 예측기간 (2014년~2017년) 기간 수위 시계열 자료의 비교결과를 살펴보면 Case M1의 다중회귀분석모형의 경우, 관측수위곡선의 시간적 변동성을 비교적 우수하게 모의하고 있으나 전 기간에 걸쳐 관측수위에 비해 예측수위가 과소 산정되는 것을 확인할 수 있으며, 특히 고수위의 경우 약 2 m 이상의 차이가 발생하는 것으로 분석되었다.

다중회귀분석모형 분석결과에서 또 하나 주목할 만한 사항은 다음과 같다. Case M1의 다중선형회귀모형의 경우, Fig. 6과 같이 반복시행횟수가 최초 130회까지는 오차(bias)가 증가하다가 이후 점차적으로 감소하고 각 관측소 수위의 가중치가 일정한 값으로 수렴하고 있음을 확인할 수 있다. 이에 대한 다중선형회귀모형은 Eq. (12)와 같으며, 옥천지점과 가까운 관측소순으로 가중치가 0.5440(송천), 0.346(호탄), 0.0197(수통)로 감소하는 것으로 분석되었다. 이는 예측 3년 기간 동안 대해 옥천 수위에 가장 영향을 많이 미치는 지점은 금강 지류 초강에 위치한 송천 관측소임을 추정할 수 있다.

**Fig. 6.** Weight Variation of the Multi Linear Regression Model (Case M1)

$$H(x) = 0.0197x_1 + 0.3460x_2 + 0.5440x_3 + 0.2312 \quad (12)$$

다만, 홍수기(5월~9월) 시 금강 본류의 유량이 지류보다 월등히 크므로 지류인 송천 관측소보다 본류에 위치한 호탄과 수통 관측소의 영향이 더 클 것으로 추정되는 바, 대상기간 (2002년~2016년) 홍수기 시수위자료를 이용하여 동일한 조건으로 다중회귀분석모형을 재학습시키고, 이에 대한 결과를 도출하였다.

Table 3은 홍수기에 대해 학습된 다중회귀분석모형의 분석결과를 나타내고 있으며, 전 기간에 대해 학습된 다중회귀분석에 비해 RMSE와 NSE 결과가 모두 향상된 것을 알

Table 3. Model Performance Results of the Multi Linear Regression Model (only flood season)

| Case | iterations | RMSE | NSE |
|------|------------|-------|--------|
| MF1 | 500 | 0.166 | 0.8296 |
| MF2 | 1000 | 0.184 | 0.7893 |
| MF3 | 3000 | 0.230 | 0.6698 |

수 있다. 재현성이 가장 우수하게 분석된 Case MF1의 다중회귀분석모형은 Eq. (13)과 같으며, 본류 수통 관측소의 수위의 영향이 지류 송천 관측소 수위보다 큰 것으로 분석되었으며 가중치는 1.300(수통), -0.106(호탄), 0.250(송천)으로 계산되었다. Case MF1의 반복시행횟수에 따른 가중치 및 오차 변화곡선은 Fig. 7과 같으며, 관측수위와 예측수위 비교결과 는 Fig. 8과 같다.

$$H(x) = 1.2995x_1 + (-0.1062)x_2 + 0.2495x_3 + 0.0644 \quad (13)$$

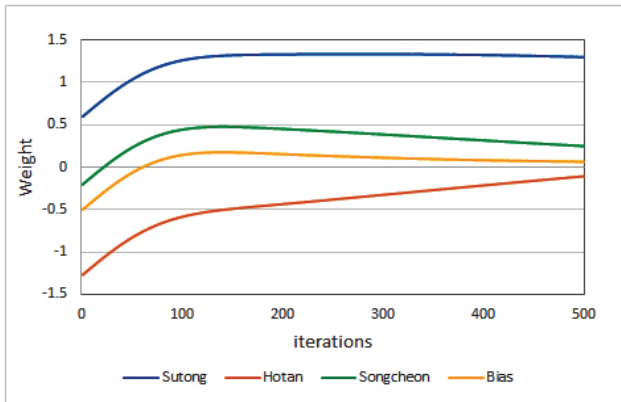


Fig. 7. Weight Variation of the Multi Linear Regression Model (Case MF1)

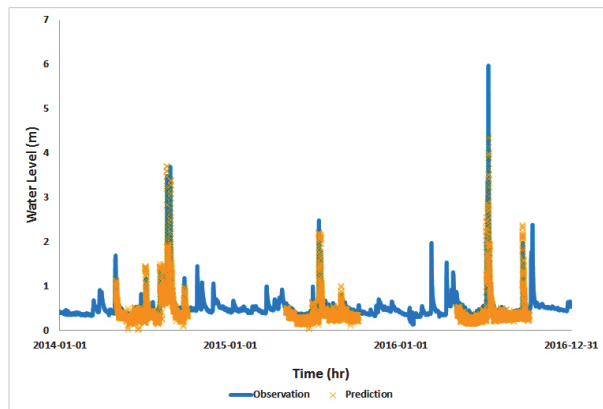
3.2.2 LSTM 모형 분석결과

전술한 바와 같이 LSTM 모형의 Sequence length를 3시간,

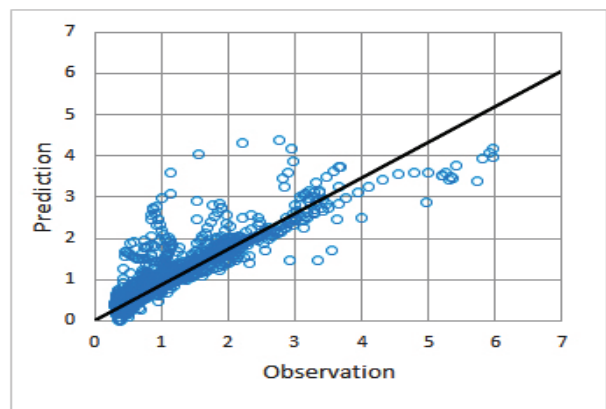
5시간, 7시간으로 구분하여 학습시켰으며, 예를 들어 Sequence length가 3시간일 경우 t 시간의 수위를 H_t 라 할 경우, 1시간 전의 수위 H_{t-1} , 2시간 전의 수위 H_{t-2} 를 이용하여 모형이 학습됐음을 의미한다. 또한 심층학습을 위한 은닉층 개수 역시 8, 10, 12개로 변화시켰으며, 활성화 함수는 LSTM에서 가장 흔히 쓰이는 Eq. (5)의 tanh(hyperbolic tangent) 함수를 사용하여 역전파 알고리즘에 의한 옥천 관측소 수위 예측을 수행하였다.

Table 4는 LSTM 모형의 각 Case별 분석결과를 나타내고 있으며, RMSE의 경우 0.012~0.017로 산정되었으며, LSTM 모형은 모든 Case에 대해 Table 2의 다중선형회귀분석모형에 비해 매우 높은 정확도를 보이는 것을 확인 할 수 있다. 그 중 Case L14, L23는 RMSE가 각각 0.0128, 0.0123로 가장 좋은 성능을 보여주었으며, Case L14, L23 모두 은닉층의 개수는 10개, 반복횟수 3000번으로 분석되었다.

Figs. 9 and 10에서 보여주는 바와 같이 LSTM 모형의 대부분은 대상기간 전체에 대해 저·중·고 수위에 상관없이 매우 우수하고, 안정적으로 예측하고 있는 것으로 분석되었으며, 전 기간에 대한 모형의 성능을 평가하는 NSE 값의 경우, Case L14와 L23 모두 0.99 이상의 매우 우수한 재현성을 나타내었다. 이는 현재 수위예측에 있어 선행 수위의 영향성을 반영할 수 있는 LSTM의 고유의 특성으로부터 기인한 것이며, 비선형 수문해석의 경우, 선형이론 기반의 다중회귀분석모형보다는 RNN-LSTM의 성능이 월등하게 우수하므로 다양한 수문시계열 분석 및 예측에 활용할 수 있을 것으로 판단된다.



(a) Water level time series

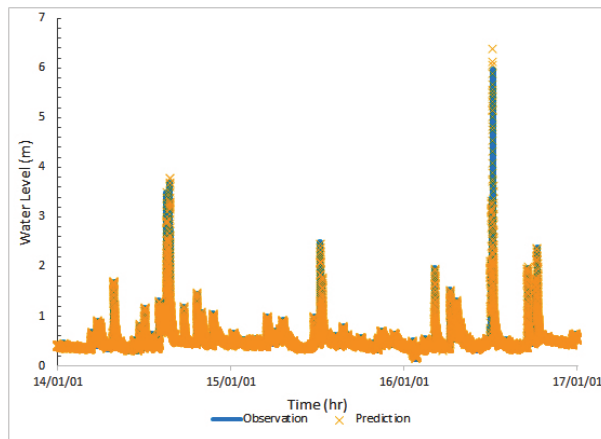


(b) Scatter plot of water level comparison

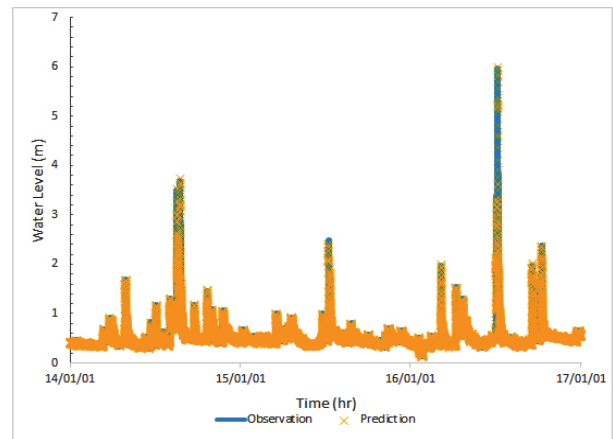
Fig. 8. Comparison of Water Level in Okcheon Station – Multi Linear Regression Model (Case MF1)

Table 4. Model Performance Results of the LSTM Model

| Case | Sequence length | Hidden layer | iterations | RMSE | NSE |
|------|-----------------|--------------|------------|-------|--------|
| L1 | 3 | 8 | 500 | 0.017 | 0.9965 |
| L2 | | | 3000 | 0.014 | 0.9977 |
| L3 | | | 10000 | 0.013 | 0.9979 |
| L4 | | 10 | 500 | 0.017 | 0.9966 |
| L5 | | | 3000 | 0.013 | 0.9981 |
| L6 | | | 10000 | 0.013 | 0.9980 |
| L7 | | 12 | 500 | 0.016 | 0.9970 |
| L8 | | | 3000 | 0.013 | 0.9979 |
| L9 | | | 10000 | 0.014 | 0.9978 |
| L10 | 5 | 8 | 500 | 0.017 | 0.9967 |
| L11 | | | 3000 | 0.015 | 0.9960 |
| L12 | | | 10000 | 0.014 | 0.9972 |
| L13 | | 10 | 500 | 0.017 | 0.9970 |
| L14 | | | 3000 | 0.013 | 0.9964 |
| L15 | | | 10000 | 0.013 | 0.9958 |
| L16 | | 12 | 500 | 0.016 | 0.9970 |
| L17 | | | 3000 | 0.014 | 0.9961 |
| L18 | | | 10000 | 0.016 | 0.9970 |
| L19 | 7 | 8 | 500 | 0.016 | 0.9971 |
| L20 | | | 3000 | 0.015 | 0.9956 |
| L21 | | | 10000 | 0.014 | 0.9968 |
| L22 | | 10 | 500 | 0.016 | 0.9959 |
| L23 | | | 3000 | 0.012 | 0.9972 |
| L24 | | | 10000 | 0.014 | 0.9950 |
| L25 | | 12 | 500 | 0.016 | 0.9960 |
| L26 | | | 3000 | 0.013 | 0.9960 |
| L27 | | | 10000 | 0.013 | 0.9953 |

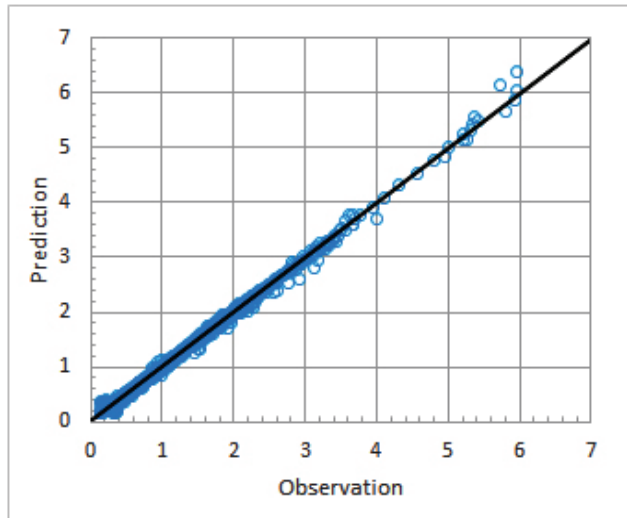


(a) Water level time series (Case 14)

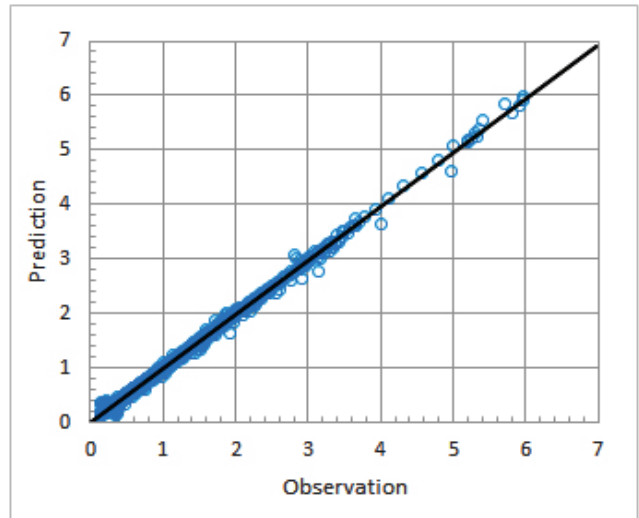


(b) Water level time series (Case 23)

Fig. 9. Comparison of Water Level in Okcheon Station – LSTM Model



(a) Scatter plot of water level (Case 14)



(b) Scatter plot of water level (Case 23)

Fig. 10. Comparison of Water Level in Okcheon Station - LSTM Model

4. 결론 및 고찰

본 연구에서는 딥러닝 알고리즘을 이용한 관측자료 기반의 하천수위 예측을 목적으로 하고 있으며, 이를 위해 금강유역 내 대청댐 상류부에 위치한 옥천 관측소 지점의 수위 예측을 위하여 구글의 딥러닝 오픈소스 라이브러리인 텐서플로를 활용하여 다중선형회귀모형과 LSTM 인공신경망 모형을 구축하였다. 모형의 학습과 예측을 위하여 입력자료(금강 본류의 수통과 호탄 관측소, 금강 지류 초강의 송천 관측소 수위)와 출력자료(옥천 수위관측소 수위)의 2002년부터 2016년까지의 15년간의 시수위자료를 획득하여 2002년부터 2013년까지의 12년간 수위자료를 이용하여 개별 모형의 학습을 진행 후 2014년부터 2016년까지 3개년에 대한 수위예측을 수행하였다. 개별 딥러닝 기반 인공신경망 모형의 재현성 평가를 위하여 예측수위자료와 관측수위자료와의 비교·분석을 수행하였으며, 주요 연구결과는 다음과 같다.

다중선형회귀모형의 최적 회귀계수(가중치) 추정을 위한 반복시행횟수를 500번, 1,000번, 3,000번으로 조정하였으며, 각 Case에 따른 예측수위와 관측수위의 RMSE 검토결과 오차는 0.159~0.231의 범위로 분포되었고, 반복시행횟수 500회인 Case M1의 RMSE가 0.1592로 오차가 가장 적은 것으로 나타났다. 그러나 Case M1의 전 기간 수위에 대한 모형의 또 다른 평가지표인 NSE를 검토한 결과 0.711로 산정되었으며, Case M1의 다중선형회귀모형의 가중치 검토결과 옥천 수위에 가장 영향을 많이 미치는 지점은 지류인 송천 수위로 분석되었다.

이에 다중선형회귀모형을 동일한 조건으로 대상기간의 홍수기(5월~9월)에 대해서만 학습을 시킨 후 모형의 재현성을 평가한 결과, RMSE 및 NSE 모두 전 기간 수위를 이용하여

학습한 Case에 비해 예측결과가 향상된 것을 볼 수 있었으며, 가중치 역시 지류 송천 수위보다는 본류 수통 수위에 많은 영향을 받는 것으로 분석되었다. 다만, 홍수기 시 본류의 유량이 지류에 비해 월등하므로 옥천 관측소를 기준으로 호탄, 수통, 송천 순으로 가중치가 낮아질 것으로 예측했지만, 본 연구에서는 수위-유량관계곡선식의 불확실성을 미포함하고자 유량자료가 아닌 수위자료로 모형을 학습시키다 보니 본류와 지류의 양적인 크기차이가 인공신경망에 충분히 반영되지 않았으며, 이에 대한 추가 연구가 필요할 것으로 판단된다.

반면에 LSTM 모형의 경우, Sequence length, 은닉층 개수, 반복시행횟수를 조정해가며 28개의 Case에 대한 수위예측을 수행한 결과, RMSE는 0.012~0.017의 범위로 분포되었고 NSE 결과 대부분의 경우, 매우 우수한 것으로 분석되었다. 최적의 모형성능을 나타낸 Case L14와 L23의 경우 NSE는 거의 1.0에 가깝게 예측능력이 우수함을 확인하였으며, 이 두 Case의 Sequence length는 각각 5시간과 7시간이며, 은닉층의 개수는 10, 반복횟수 3,000번인 것으로 분석되었다.

다중선형회귀모형이 반복시행횟수에 따라 가중치의 변동성이 나타나고, 고수위에 대한 예측결과가 불안정한 반면, LSTM 모형의 경우, 수위의 고저에 상관없이 모든 수위에 대하여 매우 안정적인 예측결과를 제공해주는 것으로 분석되었다. 즉 Sequence length의 조정으로 인해 현재시간의 수위예측에 있어 선행 수위의 변동량을 기억하고, 이를 예측에 반영하는 LSTM 모형은 수문자료 시계열 예측에 탁월한 효과가 있음을 확인할 수 있다.

본 연구는 상류 측 수위자료만을 이용하여 인공신경망 모형을 학습하고, 하류 측 수위를 예측하는 비교적 단순한 구조의 수위예측 시스템을 구축하였으나, 비선형성이 매우

강한 하천수위 예측에 있어 수위변화에 영향이 있을 것으로 판단되는 모든 영향인자 자료를 입력 자료로 구축하여 LSTM 기반 인공신경망 모형을 구축할 경우 기존의 물리적 수문·수리모형을 충분히 대체할 수 있을 것으로 판단된다.

또한 텐서플로우는 일반 CPU에서가 아닌 GPU 환경에서 딥러닝을 구현할 수 있도록 라이브러리를 제공하므로 빅데이터 기반의 수문시계열 예측모형의 구축이 가능하며, 확장성이 매우 용이하다. 향후 비선형성이 보다 강한 강우-유사유출(rainfall-sedimentation runoff) 시스템 등 다양한 수리·수문분야의 딥러닝 알고리즘 적용을 위해 심층망 구성 및 활성화 함수 선택에 따른 인공신경망 모형의 민감도 분석연구가 요구되며, 모형의 정확도 향상과 계산시간의 감소를 위한 연구 및 학습 자료의 불확실성에 따른 예측결과 오차전파 등에 대한 추가적인 연구가 필요할 것으로 판단된다.

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C. et al. (2016) Tensorflow: Large-scale Machine Learning on Heterogeneous Distributed Systems. *12th USENIX Symposium on Operating Systems Design and Implementation*, USENIX Association.
- Adnan, R., Ruslan, F.A., Samad, A.M., and Zain, Z.M. (2012) Flood Water Level Modelling and Prediction Using Artificial Neural Network: Case Study of Sungai Batu Pahat in Johor. *Control and System Graduate Research Colloquium (ICSGRC) 2012*, IEEE, pp. 22-25.
- Ahn, S.J., Yeon, I.S., and Kim, K.I. (2000) Rainfall Forecasting Using Neural Network. *Journal of the Korean Society of Civil Engineers B*, Vol. 20, No. 5B, pp. 711-722.
- Ankit, S. (2017) *TensorFlow Tutorial: 10minutes Practical TensorFlow Lesson for Quick Learners*. Retrieved from <http://cv-tricks.com/artificial-intelligence/deep-learning/deep-learning-frameworks/tensorflow-tutorial/>
- Coulibaly, P., and Anctil, F. (1999) *Real-time Short-term Natural Water Inflows Forecasting using Recurrent Neural Networks*. *Neural Networks, 1999. IJCNN'99. International Joint Conference on*, IEEE, , pp. 3802-3805.
- Dawson, C.W., and Wilby, R. (1998) An Artificial Neural Network Approach to Rainfall-runoff Modelling. *Hydrological Sciences Journal*, Vol. 43, No. 1, pp. 47-66.
- Hochreiter, S., and Schmidhuber, J. (1997) Long Short-term Memory. *Neural Computation*, Vol. 9, No. 8, pp. 1735-1780.
- Kim, S.W. (2005) Reliability Analysis of Flood Stage Forecasting using Neural Networks Model - I. Model Development and Application. *Journal of the Korean Society of Civil Engineers B*, Vol. 25, No. 6B, pp. 473-482.
- Kwon, S.H., Lee, J.W., and Chung, G.H. (2017) Snow Damages Estimation using Artificial Neural Network and Multiple Regression Analysis. *J. Korean Soc. Hazard Mitig.*, Vol. 17, No. 2, pp. 315-325.
- McCulloch, W.S., and Pitts, W. (1943) A Logical Calculus of the Ideas Immanent in Nervous Activity. *The Bulletin of Mathematical Biophysics*, Vol. 5, No. 4, pp. 115-133.
- Minsky, M., and Papert, S. (1969) *Perceptrons: an Introduction to Computational Geometry*. Mit Press, Cambridge, Massachusetts, USA.
- Olah, C. (2015) Understanding LSTM Networks. GITHUB blog, Retrieved from <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>.
- Park, S.C., Oh, C.R., Kim, D.R., and Jin, Y.H. (2006) Development of Flood Runoff Forecasting System by using Artificial Neural Networks - Development & Application of GUI_FFS -. *Journal of the Korean Society of Civil Engineers B*, Vol. 26, No. 2B, pp. 145-152.
- Rosenblatt, F. (1958) The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, Vol. 65, No. 6, pp. 386-408.
- Rumelhart, D.E., Hintont, G.E., and Williams, R.J. (1986) Learning Representations by Back-propagating Errors. *NATURE*, Vol. 323, No. 9, pp. 533-536.
- Shin, H.S., Choi, J.N., and Park, M.J. (1998) Nonlinear Forecasting Models of Real Time Rainfall and Flood Runoff Based on Neural Networks. *Proceeding of the Korean Society of Civil Engineers Conference*, Vol. 3, pp. 185-188.
- Son, M.W., and Lee, K.S. (2003) Forecasting of Flood Stage Using Neural Networks and Regression Analysis. *Journal of the Korean Society of Civil Engineers B*, Vol. 23, No. 3B, pp. 147-155.
- Tensorflow <https://www.tensorflow.org/>
- Thirumalaiah, K., and Deo, M.C. (1998) Real-Time Flood Forecasting Using Neural Networks. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 13, No. 2, pp. 101-111.

Tran, Q.K., and Song, S.K. (2017) Water Level Forecasting based on Deep Learning: A Use Case of Trinity River-Texas-The United States. *Journal of KIISE*, Vol. 44, No. 6, pp. 607-612.

Yeo, W.K., Seo, Y.M., Lee, S.Y., and Jee, H.K. (2010) Study on Water Stage Prediction Using Hybrid Model of Artificial Neural Network and Genetic Algorithm. *Journal of Korea Water Resources Association*, Vol. 43, No. 8, pp. 721-731.

Yoon, K.H., Seo, B.C., and Shin, H.S. (2004) Dam Inflow Forecasting for Short Term Flood Based on Neural Networks in Nakdong River Basin. *Journal of Korea Water Resources Association*, Vol. 37, No. 1, pp. 67-75.

| | |
|-----------------|--------------------|
| Received | September 29, 2017 |
| Revised | October 18, 2017 |
| Accepted | November 20, 2017 |

