

기상 요소 데이터를 이용한 단기 강수량예측 향상을 위한 딥러닝 병렬 모델

김 창 복*

Deep Learning Parallel Model for Improving Short-term Precipitation Prediction using Meteorological Element Data

Chang-Bok Kim*

요 약

강수량예측은 누적 강수량보다는 짧은 시간에 얼마나 많은 비가 집중적으로 내렸는지 알려주는 시간당 강수량에 따라서 피해 정도가 달라진다. 본 연구는 딥러닝 기본모델인 DNN, LSTM, BiLSTM, 1D-CNN 모델과 성능 향상을 위해 기본모델을 병합한 병렬 구조를 이용하여 단기 강수량예측을 하였다. 병렬모델은 DNN 모델과 1D-CNN을 병합한 DCNN과 DNN 모델과 LSTM을 병합한 DLSTM 모델이다. 데이터셋은 강수일 만을 구축한 데이터셋, 6월부터 9월까지의 데이터셋, 5월부터 10월까지의 데이터셋 등 3개의 데이터셋을 이용하였으며, 각 데이터셋에 대해서 세부적으로 7개의 데이터셋으로 구분하여, 총 21개의 데이터셋에 대하여 강수량을 예측하고 비교 평가하였다. 실험 결과, 세 번째 데이터셋이 가장 예측 결과가 좋았다. 특히, 5 번째 세부 데이터셋인 DLSTM 병렬 모델의 RMSE가 0.25로서, 다른 모델보다 10배 정도 월등히 예측 결과가 좋았다.

Abstract

In precipitation forecasting, the degree of damage varies according to the amount of precipitation per hour, which tells how much rain has fallen intensively in a short period of time, rather than the cumulative amount of precipitation. In this study, short-term precipitation prediction was performed using deep learning basic models such as DNN, LSTM, BiLSTM, and 1D-CNN models and a parallel structure merging the basic models to improve performance. Parallel models include DCNN, which combines DNN and 1D-CNN, and DLSTM, which combines DNN and LSTM. As for the data set, three data sets were used: a data set built with only rainy days, a data set from June to September, and a data set from May to October. Each data set was divided into seven data sets in detail. The precipitation was predicted and compared for a total of 21 data sets. As a result of the experiment, the third data set had the best prediction results. In particular, the RMSE of the 5th detailed data set, the DLSTM parallel model, was 0.25, which was about 10 times better than other models.

Keywords

precipitation forecast, deep neural network, parallel model, 1D-CNN, ResNet

* 가천대학교 컴퓨터공학과 교수
- ORCID: <https://orcid.org/0000-0002-1755-5033>

· Received: Aug. 09, 2023, Revised: Sep. 01, 2023, Accepted: Sep. 04, 2023
· Corresponding Author: Chang-Bok Kim
Dept. of Computer Engineering, Gachon University, 1342, Seongnam-daero,
Sujeong-gu, Seongnam-si, Gyeonggi-do, Korea
Tel.: +82-31-750-4773, Email: cbkim@gachon.ac.kr

1. 서 론

지구 온난화로 인한 기후변화는 기상의 불확실성과 기상이변으로 여러 분야에 피해를 발생시키고 있다. 2022년 6월 기상청과 APEC 기후센터(Asia Pacific Economic Cooperation Climate Center)는 “온실가스를 감축하지 않는다면, 21세기 말 극한 강수량이 최대 70% 이상 증가할 것”이라 예측했다[1]. 이것은 기록적인 폭우와 홍수 재해 또는 극한의 가뭄 사태를 지금보다 훨씬 자주, 훨씬 험악한 양상으로 겪게 된다는 의미이다. 최근 10년간 호우, 가뭄, 폭염, 태풍, 대설, 한파 등 이상기후 현상을 더 자주 더 심하게 겪었다. 기상이변으로 인한 기상특보를 생성하고 시기적절한 조치를 취하기 위해서 기상예보의 중요성이 더욱 대두되고 있다[2].

기존의 수치 예보 모델은 물리적 특성을 고려하는 수학적 모델로, 복잡한 연산 때문에 처리시간이 오래 걸린다. 따라서 빠르고 정확한 기상예보를 목적으로 하는 단기 기상예보에는 적합하지 않다. 또한, 최근의 강수는 더욱 우발적이며, 집중적으로 변하고 있으므로 수학적 모델 강수량예측에는 한계가 있다[3].

예측 분야에서 딥러닝(Deep learning)은 인간의 능력을 능가하는 수준으로 진화하고 있으며, 실제계에 사용할 수 있도록 다양한 모델을 끊임없이 연구 개발하고 있다. 딥러닝은 DNN(Deep Neural Network), LSTM(Long Short Term Memory), CNN(Convolution Neural Network)을 기본으로, 입력 데이터의 특징에 따라 예측모델을 적용한다[3]-[6]. 기존에 기상예측을 위해 수치 데이터와 기상 영상, 레이더, 천리안 등 위성이 제공하는 영상을 가공, 처리하여 딥러닝 모델에 적용하는 연구가 활발히 이루어지고 있다 [7]-[10]. 그러나 대부분 각 데이터 특성에 따른 단독 모델을 적용하는 한계가 있었다. 이와 같은 서로의 예측 문제점을 보완하기 위해 기상요소 데이터와 영상 데이터를 이용하여, 모델을 병합하여 처리하는 연구도 일부 진행되었다[11].

본 연구는 기상요소 데이터만을 이용한 단기 강수량예측 향상을 위한 딥러닝 모델을 제안한다. 제안한 모델은 기본모델인 DNN, LSTM, 1D-CNN 모델

과 기본모델을 병합한 병렬모델이다[12]. DNN은 다소 깊은 중간층에 기울기 소실을 줄이기 위해 ResNet 모델을 적용하였다. CNN 모델은 수치 데이터를 처리하기 위하여 1D-CNN을 사용하였다. LSTM 모델은 학습 데이터가 시계열 데이터셋일 경우 사용하였다. 병렬모델은 DNN 모델과 1D-CNN을 병합한 DCNN 모델과 DNN 모델과 LSTM을 병합한 DLSTM 모델이다. 본 모델은 다양한 기상 영상 데이터를 이용한 2D CNN 모델과 병합한 멀티 모달(Multi modal) 형식의 모델에 대한 응용이 가능하다.

본 논문은 2장에서 강수량예측과 딥러닝 모델에 관해서 서술하였다. 3장에서 제안모델과 실험방법을 서술하였으며, 4장에서 실험결과 및 성능평가를 하였다. 최종적으로 5장에서 결론 및 향후 과제에 대해서 나타냈다.

II. 강수량예측 및 딥러닝 모델

2.1 강수량예측

누적 강수량은 하루나 이틀 사이에 내린 비의 양을 의미하며, 이러한 누적 강수량보다는 짧은 시간에 얼마나 많은 비가 집중적으로 내렸는지 알려주는 시간당 강수량에 따라서 피해 정도가 달라진다. 시간당 강수량은 약한 비는 1시간에 3mm 미만, 보통 비는 3~15mm 미만, 강한 비는 1시간에 15mm 이상, 매우 강한 비는 1시간에 30mm 이상이다.

본 연구는 기상 자료개방 포털 사이트의 기상 데이터를 사용하였다[14]. 사용한 데이터는 기온(°C), 풍속(m/s), 습도(%), 증기압(hPa), 이슬점온도(°C), 현지기압(hPa), 전운량(10분위), 중하층운량(10분위), 지면온도(°C), 강수량(mm) 등 10개의 기상요소이다. 표 1에 2008~2021년 사이의 6월 1일~9월 30일까지의 기상요소에서 한 시간 후의 강수량에 대한 각 요소 간 상관관계를 나타냈다. 표 1에서 1~9는 기온, 풍속, 습도, 증기압, 이슬점온도, 현지기압, 전운량, 중하층 구름양, 지면온도 등 강수량을 제외한 기상요소며 10~14는 세 시간 이전의 강수량, 두 시간 이전의 강수량, 한 시간 이전의 강수량, 현재 시각의 강수량 그리고 다음 시간의 강수량이다.

표 1. 한 시간 후의 강수량과 기상요소 간 상관관계
Table 1. Correlation between 1-hour later precipitation and meteorological factors

1	2	3	4	5	6	7
-0.039	0.070	0.380	0.232	0.232	-0.194	0.403
8	9	10	11	12	13	14
0.352	-0.058	0.460	0.517	0.578	0.692	1.000

표 1에서 습도(3), 전운량(7), 중하층 구름양(8)을 제외한 다른 기상요소의 상관관계가 전반적으로 낮음을 알 수 있다. 이것은 기상요소로는 다소 강수량 예측에 한계점이 있다는 것을 의미한다. 따라서 본 연구는 다음 시간의 강수량예측(14)을 위해서 세 시간 이전의 강수량(10), 두 시간 이전의 강수량(11), 한 시간 이전의 강수량(12), 현재 시각의 강수량(13)이 다음 시간 강수량과 상관관계가 높음을 고려하여 입력 변수에 추가하였다. 그림 1에 2013년~2022년까지 10년간 여름철 강수량에 대해서 나타냈다.

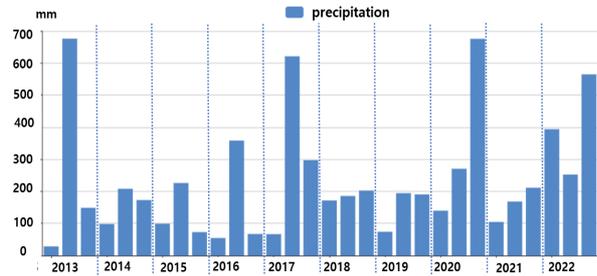


그림 1. 10년간 여름철 월 강수량
Fig. 1. Summer month precipitation over 10 years

1년 강수량 중 장마가 집중되는 7~8월에 강수량이 집중된다. 2020년 장마철은 1973년 이래로 가장 긴 기간의 장마이자 세 번째로 많은 강수량을 기록했으며, 2021년 6월은 강수량은 많지 않지만 한 달간 서울에 총 13차례의 강한 소나기가 내리며 기후 위기를 실감케 했다. 이처럼 기상이변으로 인한 갑작스러운 호우, 가뭄, 대설 등에 대한 기상 특보를 생성하고 시기적절한 조치하기 위해서 초단기 기상 예보의 중요성이 대두되고 있다. 그림 2에 여름철 시간당 최고 강수량과 평균 기온에 대해 나타냈다. 그림 2에서 2020년도의 시간당 강수량 보다 2021년도의 시간당 강수량이 월등히 높음을 알 수 있으며, 급격한 강수량으로 인해 많은 수해를 입었다.

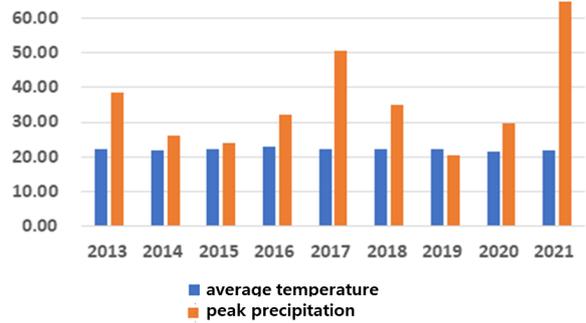


그림 2. 여름철 시간당 최고 강수량과 평균 기온
Fig. 2. Maximum precipitation and average temperature per hour in summer

2.2 딥러닝 모델

본 연구는 강수량예측 기본모델로서 DNN, 1D-CNN, Multi kernel 1D-CNN, LSTM, BiLSTM 모델을 사용하였다.

DNN은 ANN의 깊은 중간층 문제를 해결하여, 중간층을 늘려서 학습결과를 향상하는 모델이다. 일반적으로 3층 이상의 ANN을 딥러닝이라 한다.

LSTM은 DNN을 병렬로 연결한 형태이고, 하나의 DNN 모델을 타임스탬프(Time Stamp)라 하며, 타임스탬프에서 학습한 결과를 다음 타임스탬프로 전달하는 구조로 순차적인 시계열 데이터학습에 특화된 모델이다. 그림 3에 LSTM 모델에 대해서 나타냈다.

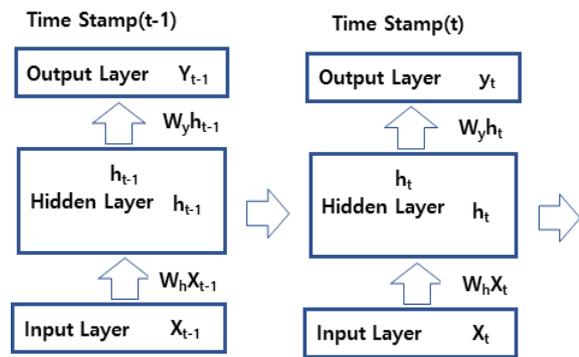


그림 3. LSTM 모델
Fig. 3. LSTM model

DNN과 LSTM은 데이터에서 지식을 추출하는 학습이라면, CNN은 데이터에서 특징을 추출하여, 패턴을 파악하는 구조이다. CNN은 수치 데이터에서 특징을 추출하는 1D-CNN과 영상 데이터에서 특징을 추출하는 2D-CNN이 있다.

본 연구에서는 수치 데이터셋이므로 1D-CNN을 사용하였다. 그림 4에 입력이 12개인 1D-CNN에 대해서 나타냈다.

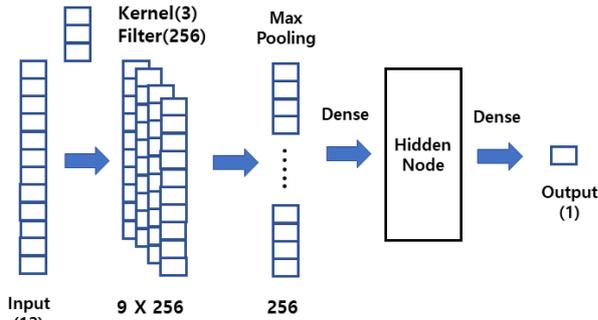


그림 4. 1D-CNN 모델
Fig. 4. 1D-CNN model

입력이 12개이고 커널이 3이고 필터가 256개 일 경우, 합성곱(Convolution) 결과는 9개 값을 가진 256개의 특성맵(Feature map)이 산출된다. 특성맵에 대해서 맥스 풀링(Max pooling)을 하게 되면 특성맵에서 가장 큰 값으로 256개의 스칼라값을 갖게되며 이를 연결하여 중간층을 거쳐 출력층에 연결되는 구조를 가진다.

III. 강수량 예측모델

3.1 강수량 예측모델

본 연구에서 강수량예측 모델은 기본모델로서 DNN, 1D-CNN, Multi kernel 1D-CNN, LSTM, BiLSTM으로 구축하였다. 또한, 병렬모델로서 DNN과 Multi kernel 1D-CNN을 병렬 연결한 DMCNN 모델 그리고 DNN과 LSTM을 병렬 연결한 DLSTM 모델을 구축하였다.

DNN 모델은 입력층은 기상요소를 입력하기 위한 노드로 9-13 노드, 중간층 6층, 중간층 노드 256 노드로 구성되어 있으며, 4층과 6층에 ResNet으로 구성하여 short cut으로 두 개층을 지름길로 하여 기울기 감소 효과를 줄이도록 구성하였다. 또한, 출력층은 예측할 강수량으로 1개의 노드로 구성하였다. LSTM과 BiLSTM은 입력층의 노드로 9-13, 타임 스탬프(Time stamp)는 12, 은닉층은 128개, 은닉층의

결과를 256 노드의 밀집(Dense) 중간층과 32개의 밀집 중간층으로 구성하였다. 또한, 출력층은 예측할 강수량으로 1개의 노드로 구성하였다. 표 2에 DNN과 LSTM 기본모델에 대해서 나타냈다.

표 2. DNN, LSTM, BiLSTM 기본모델
Table 2. DNN, LSTM, BiLSTM basic model

Model	Input layer		Hidden		Dense		Output layer
	Time stamp	Input	Layer	Node	Dense 1	Dense 2	
DNN		9-13	6	256			1
LSTM	12	9-13	1	128	256		1
					32		
BiLSTM	12	9-13	1	128	256		1
					32		

그림 5와 그림 6에 DNN과 LSTM 기본모델에 대해서 나타냈다.

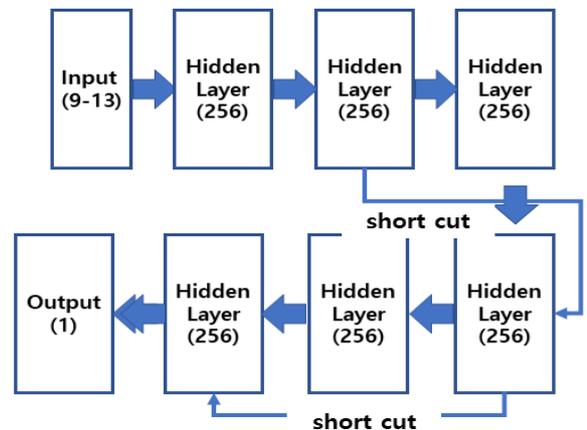


그림 5. DNN 기본모델
Fig. 5. DNN basic models

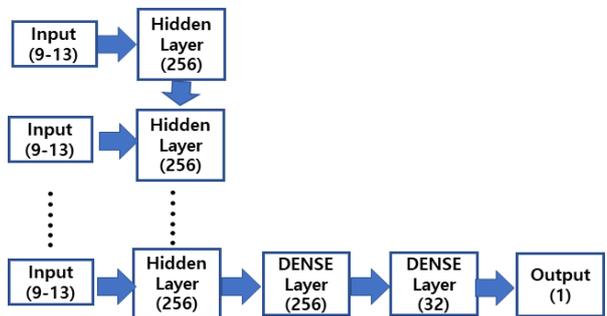


그림 6. LSTM 기본모델
Fig. 6. LSTM basic models

1D-CNN 모델은 2-D CNN과 달리 거널의 움직임인 스트라이드가 입력의 아래로만 이동한다. 또한, 입력의 차원은 2차원으로, 본 모델에서는 (9~13, 1)으로 입력된다. 입력층의 노드로 9~13, 합성곱 층으로 필터 256개, 커널 3, 스트라이드 1로 구성하였으며, 풀링층으로 맥스 풀링을 이용하였다. 또한, 풀링 결과는 커널 수와 동일한 256개의 스칼라 값이며, 이를 256 노드의 밀집 중간층과 32개의 밀집 층간층으로 구성하였다. 그림 3에 1-D CNN에 대해서 나타냈다.

Multi kernel 1D-CNN은 1D-CNN과 동일한 구조이나 커널을 3, 4, 5로 구성하였다. 따라서 3개의 커널을 거친 특성값은 9개 노드를 가진 256개의 벡터가 3개로 구성되며, 맥스 풀링의 값은 768개의 스칼라 값이다. 이를 256 노드의 밀집 중간층과 32개의 밀집 층간층으로 구성하였다. 또한, 출력층은 다음시간 강수량으로 1개의 노드로 구성되었다. 표 3에 1-CNN과 Multi kernel 1D-CNN 기본모델에 대해서 나타냈다. 또한, 그림 7에 Multi kernel 1D-CNN 구조에 대해서 나타냈다.

표 3. CNN 기본모델
Table 3. CNN basic model

model	input layer	Convolution		Pooling result	Dense		output layer
		kernel	filter		Dense1	Dense2	
1D-CNN	9-13	3	256	256	256 32	1	
1D-MCNN	9-13	3,4,5	256	768	256 32	1	

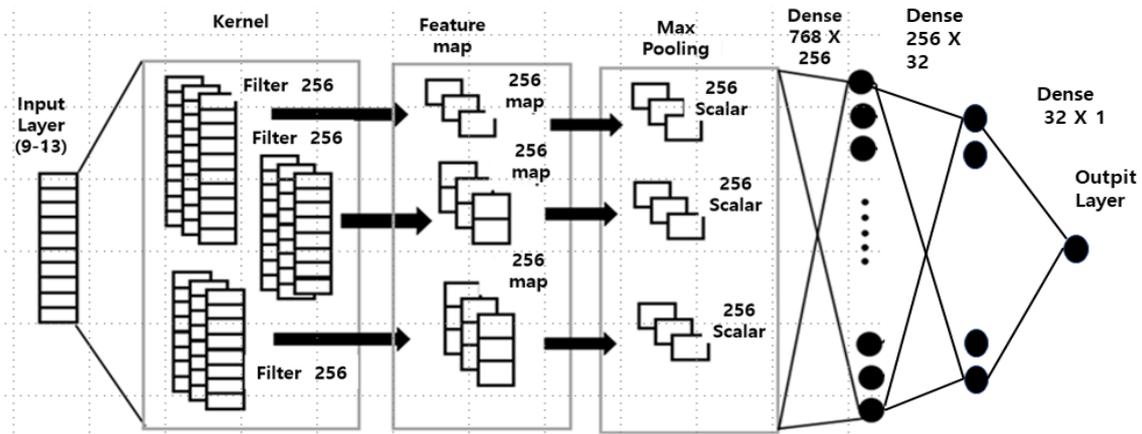


그림 7. Multi kernel 1D-CNN 구조
Fig. 7. Multi kernel 1D-CNN structure

병렬모델은 동일한 모델 또는 서로 다른 모델을 Keras Functional API로 병렬로 구성하는 방법이다. 기본모델은 순차적으로 중간층을 구성하지만, 병렬 모델은 각 기본모델로 중간층을 구성하고, 모델의 결과를 연결해야 한다. 그림 8에 기본모델과 병렬 모델에 대해서 나타냈다.

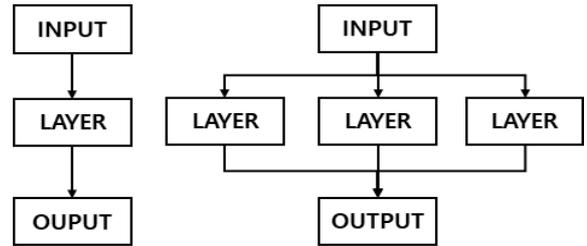


그림 8. 기본모델과 병렬모델
Fig. 8. Basic and parallel models

병렬모델은 DMCNN 모델과 DLSTM 등 2개의 모델로 구성되어 있다. DMCNN 모델은 DNN 모델과 Multi kernel 1D-CNN 모델을 병렬한 모델로서 DNN 모델의 출력을 8노드로 하고, Multi kernel 1D-CNN 모델의 출력을 8노드로 하여, 출력 결과를 연결한 16노드에 대해서 256노드의 밀집층과 32개의 밀집층으로 구성하였다. DLSTM 모델은 DNN 모델과 LSTM 모델을 병렬한 모델로서 DNN 모델의 출력을 8노드로 하고, LSTM 모델의 출력을 8노드로 하여, 출력 결과를 연결한 16노드에 대해서 256노드의 밀집층과 32개의 밀집층으로 구성하였다. 표 4에 병렬모델 구조에 대해서 나타냈다. 그림 9에 병렬모델에 대해서 나타냈다.

표 4. DMCNN과 DLSTM 병렬모델

Table 4. DMCNN and DLSTM parallel model

Model	Network	Output1	Hidden1	Hidden2	Output
DMCNN	DNN	8	256	32	1
	MCNN	8			
DLSTM	DNN	8	2564	32	1
	LSTM	8			

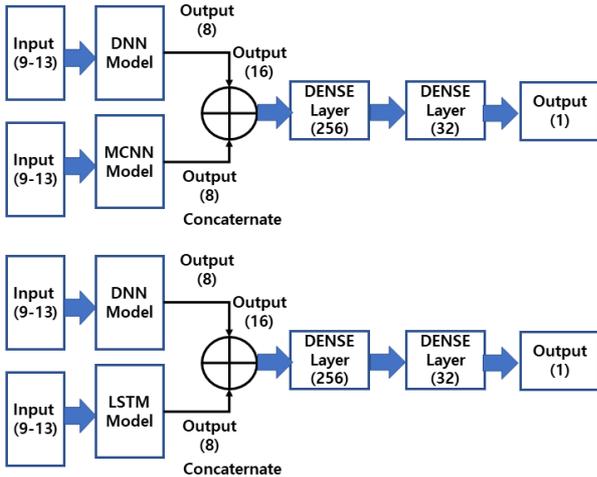


그림 9. DMCNN과 DLSTM 병렬모델

Fig. 9. DMCNN and DLSTM parallel model

병렬모델은 모델마다 입력 데이터의 형태가 다르기 때문에, 입력 데이터의 매칭이 필요하다. LSTM 데이터 구축 시 12번째 샘플까지 타임스탬프에 입력되는 데이터이며, 12번째 샘플이 라벨이다. 따라서 DNN과 CNN의 12개 샘플을 제거해야만 LSTM의 실제 데이터와 일치하기 때문이다.

3.2 강수량예측 데이터

데이터셋은 2008~2021년 사이의 6월 1일~9월 30일 까지 10개의 기상요소를 기본으로 하였다, 표 5에 본 연구에서 사용한 데이터의 기술 통계에 대해서 나타냈다.

표 5. 데이터셋의 기술 통계

Table 5. Descriptive statistics of the data set

Average	0.257898	Min value	0
Standard deviation	1.77207	Max value	71
variance	3.140232	Sample	65880

전체 샘플수는 65880개이며, 평균은 0.258, 표준편차는 1.77, 분산은 3.14이다, 또한 최대 시간당 강수량은 71 mm이다, 샘플의 내역은 다음과 같다

$$Sample = 24(\text{시간}) \times (31(\text{일}) + 30(\text{일}) + 31(\text{일}) + 31(\text{일}) + 30(\text{일}) + 31(\text{일})) \times 15 \quad (1)$$

단, 여기서 10월 31일은 0시까지만 데이터를 추출하였다. 실험 전체 데이터는 크게 3 개의 데이터셋으로 구성하였다.

첫 번째 데이터셋은 강수일이 매우 적음에 인하여 예측 문제를 분석하기 위해 강수일이 없는 날을 제거한 데이터셋이다. 전체 65880 샘플에서 강수가 측정된 강수일수는 5680 샘플로 강수일수는 8.6% 정도여서 강수가 없는 날을 제거하여 다음 시간 강수량을 예측하였다. 본 데이터셋은 시계열 특성이 제거되므로 LSTM 모델은 사용하지 않았다.

두 번째 데이터셋은 여름철 기상예측 데이터셋으로 2008년부터 2022년까지 6월 1일부터 9월 31까지 10개의 기상 요소를 기본으로 하였다.

세 번째 데이터셋은 2008~2022년 사이의 5월 1일~10월 31까지 10개의 기상 요소를 기본으로 하였다. 표 6에 데이터셋에 대해서 나타냈다.

표 6. 데이터셋

Table 6. Dataset

DataSet	contents
DataSet 1	Dataset from which days with no precipitation are removed
DataSet 2	Data set between June 1st and September 30th
DataSet 3	Data set between May 1st and October 31st

각 데이터셋은 7개의 세부 데이터셋을 구성하였다. 표 7에 세부 데이터셋에 대해서 나타냈다.

1. 세부 데이터셋 1의 입력은 기온, 풍속, 습도, 증기압, 이슬점온도, 현지기압, 전운량, 중하층운량, 지면온도 등 9개 기본 기상요소이며, 라벨은 현재 시각 강수량이다.

2. 세부 데이터셋 2의 입력은 9개 기본 기상요소, 현재 시각 강수량이며, 라벨은 다음 시간 강수량이다.

표 7. 세부 데이터셋

Table 7. Detailed data set

Dataset	Input	Label
1	Weather elements(9)	Current time precipitation
2	Weather elements(9), present time precipitation	1-hr later precipitation
3	Weather elements(9), 1-hr before precipitation, present time precipitation	1-hr later precipitation
4	Weather elements(9), 2-hr, 1-hr before precipitation, present time precipitation	1-hr later precipitation
5	Weather elements(9), 3-hr, 2-hr, 1-hr before precipitation, present time precipitation	1-hr later precipitation
6	Weather elements(9), 3-hr, 2-hr, 1-hr before precipitation, present time precipitation	2-hr later precipitation
7	Weather elements(9), 3-hr, 2-hr, 1-hr before precipitation, present time precipitation	3-hr later precipitation

3. 세부 데이터셋 3의 입력은 9개 기본 기상요소, 1시간 전, 현재 시각 강수량이며, 라벨은 다음 시간 강수량이다.

4. 세부 데이터셋 4의 입력은 9개 기본 기상요소, 2시간 전, 1시간 전, 현재 시각 강수량이며, 라벨은 다음 시간 강수량이다.

5. 세부 데이터셋 5의 입력은 9개 기본 기상요소, 3시간 전, 2시간 전, 1시간 전, 현재 시각 강수량이며, 라벨은 다음 시간 강수량이다.

6. 세부 데이터 세트 6의 입력은 데이터셋 5의 입력과 동일하며, 라벨은 두시간 후 강수량이다.

7. 세부 데이터셋 7의 입력은 데이터셋 5의 입력과 동일하며, 라벨은 세시간 후 강수량이다.

IV. 실험 및 성능평가

본 실험은 윈도우 10 환경에서 Anaconda, Python 3.6, Keras 2.6.0을 사용하였다, 학습 알고리즘은 Adam을 사용하였으며, 학습률은 0.001, 배치 크기는 10으로 하여, 100번 반복 학습하였다. 모델의 중간층 활성화 함수는 ReLU를 사용하였으며, 출력층 활성화 함수는 linear를 사용하였다. 표 8에 실험에 사용한 하이퍼 파라미터에 대해서 나타냈다.

표 8. 하이퍼 파라미터

Table 8. Hyper parameter

Hyper parameter	Value
Learning algorithm	Adam
Learning rate	0.001
Batch size	10
Epoch	100
Hidden activation function	Relu
Output activation function	Linear

기상요소는 데이터의 크기가 모두 다르기 때문에, 데이터의 범위를 일치시켜야 한다. 또한, 딥러닝은 입력 데이터의 크기가 작아야 학습이 효과적이기 때문에, 기상요소 입력 값을 0과 1 사이로 정규화하였다. 그러나 출력 값은 그대로 실제 값(0~71)으로 사용하였다. 모델의 성능평가는 RMSE (Root Mean Square Error)를 사용하였다. 또한 실험은 3번 반복하였으며, 각 실험에 대한 표준 편차를 구하였다.

$$Normalize = \frac{d_i - MIN(d)}{MAX(d) - MIN(d)} \quad (2)$$

$$RMSE(x', x) = \frac{1}{N} \sqrt{\sum_{n=1}^N (x'_n - x_n)^2} \quad (3)$$

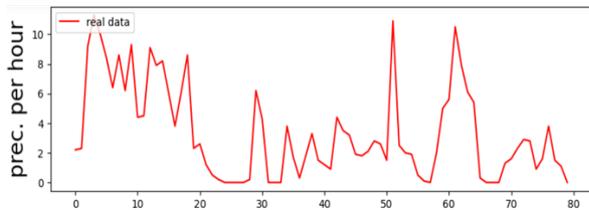
$$STDEV = \sqrt{\frac{\sum |x - \mu|^2}{N}} \quad (4)$$

첫 번째 데이터셋은 강수가 측정된 강수일수는 5680 샘플이고, 학습 데이터는 5112 샘플로 하였으며, 테스트 데이터는 샘플의 10%인 568 샘플로 하였다. 또한, 강수일을 제거한 시계열이 아닌 데이터셋이므로 LSTM 계열의 모델을 사용하지 않았다. 표 9에 첫 번째 데이터셋 결과에 대한 표준 편차 (Stdev)와 RMSE를 나타냈다. 각 데이터 셋의 결과는 3회 반복 실험 후 평균 값이다. 그림 10에 두 번째 세부 데이터셋에 대한 DNN 결과 그래프를 나타냈다. 그림 10에서 30번째 부근의 샘플에 다소 높은 오차를 보임을 수 있으며, 전체적으로 모든 세부 데이터셋의 오차가 높았음을 알 수 있다.

표 9. 첫 번째 데이터셋 결과

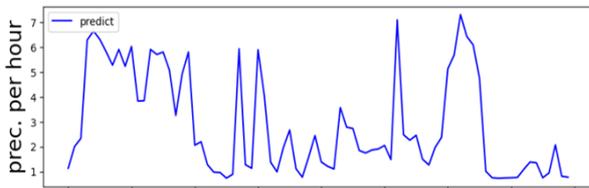
Table 9. First data set result

model		1	2	3	4	5
DNN	STDEV	0.066	0.093	0.202	0.148	0.222
	RMSE	5.310	3.750	3.770	3.930	5.600
CNN	STDEV	0.090	0.093	0.018	0.063	0.152
	RMSE	5.440	3.780	3.930	3.860	5.290
MCNN	STDEV	1.117	0.277	0.543	0.039	0.397
	RMSE	4.770	4.020	4.130	3.980	5.530
DMCNN	STDEV	0.269	0.037	0.047	0.209	0.187
	RMSE	5.610	3.810	3.920	4.060	5.640



(a) 실제 데이터 그래프

(a) Real data graph



(b) DNN 모델 그래프

(b) DNN model graph

그림 10. 첫 번째 세부 데이터셋 DNN 모델 결과

Fig. 10. First detailed dataset DNN model results

두 번째 데이터셋은 33814샘플이며, 학습 데이터는 2008년도~2021년까지 30912샘플이며, 테스트 데이터는 2022년도 2902샘플이다. 표 10에 두 번째 데이터셋 기본 모델 결과에 대해 나타냈다.

표 10. 두 번째 데이터셋 기본 모델 결과

Table 10. Second data set basic model results

model	1	2	3	4	5	6	7
DNN	0.042	0.181	0.208	0.075	0.007	0.042	0.016
	2.620	2.710	2.880	2.830	2.620	2.590	2.530
CNN	0.284	0.086	0.041	0.110	0.057	0.055	0.026
	2.850	2.800	2.810	2.770	2.640	2.640	2.550
DMCNN	0.112	0.130	0.053	0.220	0.245	0.106	0.013
	2.630	2.900	2.760	2.980	3.020	2.700	2.530
LSTM	0.234	0.316	0.360	0.456	0.520	0.273	0.131
	2.329	2.044	1.997	2.211	2.211	2.438	2.277
BiLSTM	0.131	0.124	0.014	0.122	0.130	0.121	0.003
	2.320	2.110	2.110	2.110	2.100	2.300	2.380

표에서 흰색 셀이 표준 편차이고 회색 부분이 RMSE를 나타낸다. 두 번째 데이터 셋은 전반적으로 첫 번째 데이터셋 보다는 오차가 낮았음을 알 수 있었다. 본 데이터셋에서는 시계열 데이터 셋이므로 당연히 LSTM과 BiLSTM은 오차가 낮았으나 높은 강수량을 보인 부분에 대해 다소 예측이 안 되었다.

세 번째 데이터셋은 전체 데이터셋에서 학습 데이터는 2008~2021년까지 61488 샘플이며, 테스트 데이터는 2022년도 4392 샘플이다. 표 11에 세 번째 데이터셋 결과에 대해 나타냈다.

표 11. 세 번째 데이터셋 기본 모델 결과

Table 11. Third data set basin model result

model	1	2	3	4	5	6	7
DNN	0.102	0.055	0.049	0.024	0.097	0.046	0.024
	2.190	2.220	2.390	2.170	2.370	2.200	2.130
CNN	0.042	0.010	0.111	0.164	0.057	0.009	0.029
	2.240	2.250	2.350	2.370	2.300	2.130	2.090
MCNN	0.081	0.091	0.027	0.057	0.216	0.132	0.130
	2.250	2.370	2.230	2.350	2.460	2.230	2.180
LSTM	0.029	0.058	0.059	0.068	0.049	0.003	0.010
	1.800	1.810	1.730	1.750	1.810	1.850	1.900
BiLSTM	0.145	0.055	0.118	0.017	0.070	0.014	0.059
	1.940	1.820	1.840	1.920	1.890	1.850	1.960

세 번째 데이터셋은 전반적으로 가장 우수한 결과를 보였다. 또한, 표 12에 병렬 모델 데이터 셋에 대해 나타냈다.

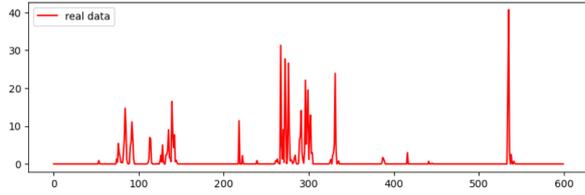
표 12. 병렬 모델 데이터셋 결과

Table 12. Parallel model dataset result

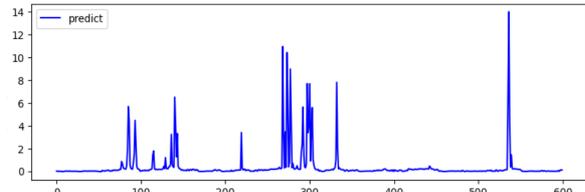
model	1	2	3	4	5	6	7
DMCNN	0.112	0.130	0.053	0.220	0.245	0.106	0.013
	2.630	2.900	2.760	2.980	3.020	2.700	2.530
DLSTM	0.149	0.279	1.289	0.103	0.259	0.093	
	0.870	0.950	1.560	0.830	2.280	2.360	
DMCNN	0.420	0.050	0.165	0.032	0.014	0.032	0.020
	2.370	2.480	2.210	2.210	2.230	2.170	2.130
DLSTM	0.001	0.082	0.103	0.081	0.058	0.172	0.075
	2.020	0.490	0.540	0.280	0.250	1.850	1.900

본 연구에서 DNN과 LSTM 모델을 병합한 DLSTM 모델에서 매우 우수한 결과를 나타냈다. 다른 모든 데이터 셋은 강수량에 대한 예측 결과가 실제 강수량 보다는 적게 예측하였으나, DLSTM은 강수량도 정확히 예측함을 알 수 있었다.

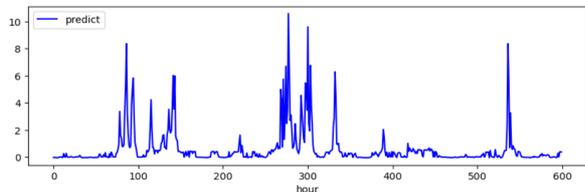
또한, 두시간 후, 세시간 후 예측 결과는 다소 오차가 높음을 알 수 있다. 그림 11에 세 번째 데이터셋에 대한 결과 그래프를 나타냈다.



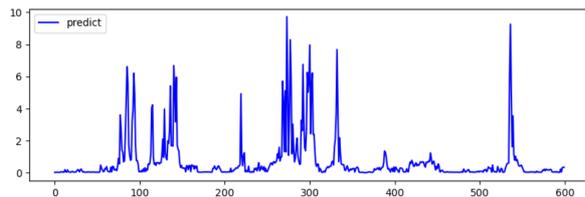
(a) 실제 데이터_1
(a) Real data_1



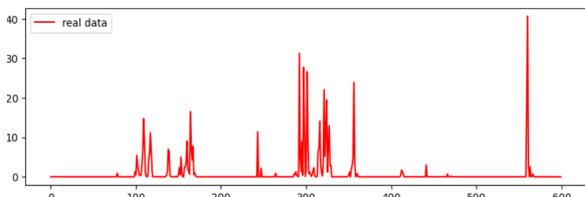
(b) DNN 예측
(b) DNN prediction



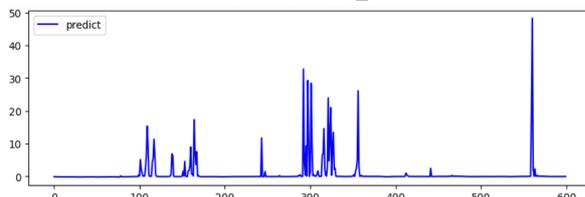
(c) CNN 예측
(c) CNN prediction



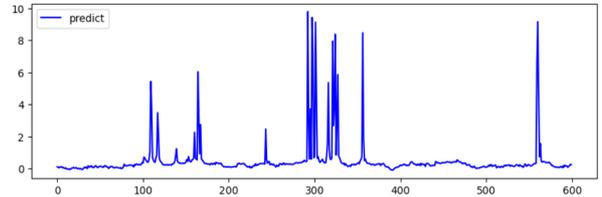
(d) MCNN 예측
(d) MCNN prediction



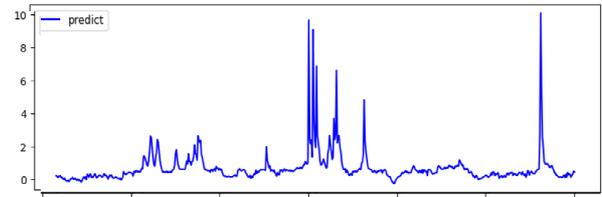
(e) 실제 데이터_2
(e) Real data_2



(f) DLSTM 예측
(f) DLSTM prediction



(g) 세부 데이터셋 6의 DLSTM 예측
(g) DLSTM prediction of detail dataset 6



(h) 세부 데이터셋 7의 DLSTM 예측
(h) DLSTM prediction of detail dataset 7

그림 11. 네 번째 세부 데이터셋에 대한 결과
Fig. 11. Results for the fourth detailed dataset

그림에서 (a)는 실제 데이터에 대한 그래프이며, DNN(b), CNN(c), MCNN(d) 데이터셋에 대한 결과 그래프이다. 전반적으로 강수가 있는 날에 대한 강수량이 실제 강수량보다는 적게 나타났다. 이것은 예측 결과값을 보정함으로써 보다 정확한 결과값을 얻을 수 있었다. (e)는 LSTM 계열 모델에 대한 실제 데이터값에 대한 그래프이다. (f)는 5번째 세부 데이터셋에 대한 그래프이며, 다른 모든 세부 데이터셋보다 오차가 매우 적었으며, 실제 강수량과 거의 정확하게 일치함을 알 수 있다. (g)와 (f)는 두 시간 후 강수량과 세 시간 후 강수량 예측으로 DNN(b), CNN(c), MCNN(d) 데이터셋과 비슷한 결과를 보였다.

V. 결론 및 향후 과제

본 연구는 기상요소 데이터만을 이용한 단기 강수량 예측 향상을 위한 딥러닝 모델을 제안하였다. 기본 딥러닝 예측모델과 예측모델의 성능을 향상하기 위해 기본모델을 병렬로 병합하는 구조를 제안하였다. 병렬모델은 깊은 은닉층에 의한 기울기 감소로 모델의 성능을 저하하는 문제점을 제거하고, 더욱 많은 학습 파라미터로 학습하여, 모델의 성능을 향상하기 위함이다.

대부분 예측 모델은 강우시간에 대한 예측은 정

확했으나 강우량을 작게 예측함을 알 수 있었다.

이것은 예측값을 보정하여 최종 예측이 가능 할 것이다. 그러나 DNN과 LSTM을 병렬로 병합한 DLSTM 병렬 모델은 강우량도 정확하게 예측하였으며, RMSE가 0.25로서 성능이 다른 모델보다 10배 정도 성능이 좋았음을 보여주었다. 본 연구를 통해 강수량 데이터 만으로 우수한 예측 결과를 산출함을 확인하였고 모델의 하이퍼 파라미터를 다듬으면 좀 더 우수한 성능을 나타낼 것으로 생각된다.

추후 본 모델과 다양한 기상 영상 데이터를 이용한 2D CNN 모델과 병합한 멀티 모달(Multi modal) 형식의 모델에 대한 연구가 필요하다.

References

- [1] <https://www.greenpeace.org/korea/update/23683/blog-extreme-climate-hazards> [accessed: Aug. 8, 2023]
- [2] I. Y. Lee, J. Y. Park, M. S. Son, S. G. Cho, M. K. Jeong, and C. G. Kim, "Nowcasting in Korean Territory using Korean Total Precipitable Water Dataset and Deep Neural Network", Autumn Annual Conference of Journal of the Institute of Electronics and Information Engineers, pp. 847-851, Jun. 2022.
- [3] P. Zhang, Y. Jia, J. Gao, W. Song, and H. Leung, "Short-term rainfall forecasting using multi-layer perceptron", IEEE Transactions on Big Data, Vol. 6, No. 1, pp. 93-106, Mar. 2020. <https://doi.org/10.1109/TBDDATA.2018.2871151>.
- [4] L. Wei, A. Kiaghadi, and C. Dawson. "High temporal resolution rainfall-runoff modeling using long-short-term-memory (LSTM) networks", Neural Computing and Applications, Vol. 33, pp. 1261-1278, Jun. 2020. <https://doi.org/10.1007/s00521-020-05010-6>.
- [5] M. Patel, A. Patel, and D. R. Ghosh, "Precipitation nowcasting: Leveraging bidirectional lstm and 1d cnn", arXiv preprint arXiv:1810.10485, Oct. 2018. <https://doi.org/10.48550/arXiv.1810.10485>.
- [6] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a Convolutional Neural Network", International Conference on Engineering and Technology (ICET), Antalya, Turkey, pp. 1-6, Aug. 2017. <https://doi.org/10.1109/ICEngTechnol.2017.8308186>.
- [7] S. Agrawal, L. Barrington, C. Bromberg, J. Burge, C. Gazen, and J. Hickey, "Machine learning for precipitation nowcasting from radar images", arXiv preprint arXiv:1912.12132, Dec. 2019. <https://doi.org/10.48550/arXiv.1912.12132>.
- [8] T. Kevin, T. Stańczyk, and S. Mehrkanoon, "Smaat-unet: Precipitation nowcasting using a small attention-unet architecture", Pattern Recognition Letters, Vol. 145, pp. 178-186, May 2021. <https://doi.org/10.1016/j.patrec.2021.01.036>.
- [9] R. Barzegar, M. T. Aalami, and J. Adamowski, "Short-term water quality variable prediction using a hybrid CNN-LSTM deep learning model", Stochastic Environmental Research and Risk Assessment, Vol. 34, No. 2, pp. 415-433, Feb. 2020. <https://doi.org/10.1007/s00477-020-01776-2>.
- [10] J. J. Kim, Y. H. You, and C. B. Kim, "Radiation Prediction Based on Multi Deep Learning Model Using Weather Data and Weather Satellites Image", Journal of Advanced Navigation Technology, Vol. 25, No. 6, pp. 569-575, Dec. 2021. <https://doi.org/10.12673/jant.2021.25.6.569>.
- [11] C. B. Kim, "Deep Learning Model for Prediction Rate Improvement of Weather Data using Parallel Merge Structure", Journal of Korean Institute of Information Technology, Vol. 20, No. 4, pp. 131-140, Apr. 2022. <http://dx.doi.org/10.14801/jkiit.2022.20.4.131>.
- [12] <https://data.kma.go.kr/cmmn/main.do> [accessed: Aug. 7, 2023]

저자소개

김 창 복 (Chang-Bok Kim)



1986년 2월 : 단국대학교

전자공학과(공학사)

1989년 2월 : 단국대학교

전자공학과(공학석사)

2009년 2월 : 인천대학교

컴퓨터공학과(공학박사)

1994년 ~ 현재 : 가천대학교

IT대학 컴퓨터공학과 교수

관심분야 : 사물인터넷, 딥러닝, 클라우드 컴퓨팅