

Original Article

기계학습을 활용한 하절기 기온 및 폭염발생여부 예측

김영인¹ · 김동현² · 이승오^{3*}

¹홍익대학교 토목공학과 석사과정

²홍익대학교 토목공학과 박사과정

³홍익대학교 토목공학과 교수

Prediction of Temperature and Heat Wave Occurrence for Summer Season Using Machine Learning

Young In Kim¹, DongHyun Kim², and Seung Oh Lee^{3*}

¹Master Course Student, Dept. of Civil Engineering, Hongik Univ.

²Ph.D Student, Dept. of Civil Engineering, Hongik Univ.

³Professor, Dept. of Civil Engineering, Hongik Univ.

요약

최근 국내에서는 다양한 이상기후들이 발생하고 있으며 이로 인해 인명피해, 재산피해와 같은 큰 피해들이 발생하고 있다. 그 중에서도 폭염으로 인한 피해는 점점 증가하는 경향을 보인다. 이에 대처하기 위해서는 빠르고 정확한 기온 및 폭염발생여부 예측이 필수적이다. 현재 기상청에서는 폭염에 대한 정보를 단기예보를 통해 제공하는데, 단기예보를 위한 기온예측은 수치예보모델을 통해 수행된다. 과거 15년간(1998~2012년) 인구대비 폭염 사망률이 가장 높았던 ○○군에 대하여 2019년도 기온 예보자료와 관측 자료를 비교한 결과, 평균제곱근오차가 1.57°C 발생하였고, 관측 값이 33°C 이상에 해당하는 데이터만 비교한 결과, 평균제곱근오차가 1.96°C 발생하였다. 예보시간은 4시간이고 예보과정에는 약 3~4시간이 소요된다. 이에 본 연구에서는 소요시간과 예측 정확도를 고려하여, 기계학습방법의 일종인 LSTM을 이용한 기온 및 폭염발생 예측 방법론을 제시한다. 기계학습모델을 이용한 4시간 기온예측결과 1.71°C의 평균제곱근오차가 발생하였고, 관측 값이 33°C 이상에 해당하는 데이터만 비교한 결과 1.39°C의 평균제곱근오차가 발생하였다. 전 범위의 오차는 수치예보모델이 더 작은 값을 가지지만, 33°C 이상의 경우에는 기계학습모델 예측의 정확도가 더 높았다. 또한 수치예보를 이용한 경우 예상 소요시간이 4시간가량인 반면 기계학습을 이용한 기온예측에는 평균 9분26초의 시간이 소요되어 경제적이다 판단하였다. 향후 공간적인 범위를 확대하거나 대상 지역을 변경하는 일반적인 방안에 대해서 연구를 수행하고자 한다.

핵심용어: 기온예측, LSTM, 기계학습, 폭염, 수치예보

ABSTRACT

Climate variations have become worse and diversified recently, which caused catastrophic disasters for our communities and ecosystem including economic property damages in Korea. Heat wave of summer season is one of causes for such damages of which outbreak tends to increase recently. Related short-term forecasting information has been provided by the Korea Meteorological Administration based on results from numerical forecasting model. As the study area, the ○○ province was selected because of the highest mortality rate in Korea for the past 15 years (1998~2012). When comparing the forecasted temperatures with field measurements, it showed RMSE of 1.57°C and RMSE of 1.96°C was calculated when only comparing the data corresponding to the observed value of 33°C or higher. The forecasting process would take at least about 3~4 hours to provide the 4 hours advanced forecasting information. Therefore, this study proposes a methodology for temperature prediction using LSTM considering the short prediction time and the adequate accuracy. As a

*Corresponding author: Seung Oh Lee, seungoh.lee@hongik.ac.kr

Received: 15 April 2020, Revised: 24 June 2020, Accepted: 29 June 2020



result of 4 hour temperature prediction using this approach, RMSE of 1.71°C was occurred. When comparing only the observed value of 33°C or higher, RMSE of 1.39°C was obtained. Even the numerical prediction model of the whole range of errors is relatively smaller, but the accuracy of prediction of the machine learning model is higher for above 33°C. In addition, it took an average of 9 minutes and 26 seconds to provide temperature information using this approach. It would be necessary to study for wider spatial range or different province with proper data set in near future.

Keywords: Temperature prediction, LSTM, Machine learning, Heat wave, Numerical forecast

1. 서론

최근 국내에서 다양한 이상기후들로 인하여 계절의 변화가 발생하였다. 이상기후보고서(KMA et al., 2019)에 따르면, 1912년부터 2017년까지 106년 동안 여름은 19일 길어지고 겨울은 18일이 짧아졌으며 연평균 기온은 10년 마다 0.18°C 상승하는 등 변화가 발생하였으며 이는 폭염이 주요한 원인이다. 특히 2018년도에는 폭염일수 31.4일(평년 9.8일), 열대야일수 17.7일(평년 5.1일) 등 기록적인 폭염이 발생하였고 이로 인해 농업, 해양수산, 환경, 건강 등 다양한 분야에서 인명피해 및 재산피해가 발생하였다. 국내에서 폭염이 일 사망률에 미치는 영향에 대해 연구한 결과, 일 최고기온이 임계 값을 넘을 때 기온의 증가에 따라 일 사망자수는 급격하게 증가하는 등 폭염으로 인한 인명피해의 정도가 뚜렷하게 나타났다(Kim et al., 2006). 또한, 폭염에 의한 지역별 인명피해에 대한 연구가 수행된 바 있으며, 폭염발생기간과 지역별 인명피해 자료를 분석한 결과 대부분의 지역에서 상관관계가 있음을 확인하였다(Park et al., 2016). 이와 같은 피해를 예방하기 위해 국가기관들은 예·경보 서비스 제공확대, 폭염재해영향모델 개발 등 다양한 대응방안을 마련하였으며, 이러한 시스템의 구축 및 활용에는 기상예측자료가 기초자료로 사용된다. 기상청에서는 현재 수치예보모델을 이용하여 기상예측 및 예보를 실시하며 단기예보를 통하여 기온, 최고기온, 최저기온 등의 폭염에 관한 기상정보를 제공한다. 단기예보의 기간은 예보발표시각으로부터 4시간~67시간이며 기온의 경우 3시간 간격의 데이터를 제공한다. 기상청의 기상예보는 관측 장비로부터 온도, 습도 등 대기 상태의 관측 값을 수신하여 수치예보모델을 통해 분석 자료를 만들어내고, 이를 예보관이 분석하여 발표하는 과정으로 이루어진다. 이 과정에서 데이터를 수신하여 자료동화 하는데 약 2~3시간이 소요되며 수치예보모델을 통해 분석 자료를 만드는 데 약 1시간 정도 소요된다(KMA, 2018). 이에 더하여 예보관의 분석과정에도 시간이 소요되며 이는 예보관에 따라 차이가 있다. 따라서 기상예보까지 최소 3~4시간이 소요되며 예보관의 데이터 가공 과정이 필수적이다. 기온예측에 소요되는 계산 시간을 단축시킨다면 기상예보 및 피해예방에 도움이 될 것이다. 이에 본 연구에서는 기계학습을 이용하여 수치예보모델보다 계산 시간이 짧고, 데이터 가공 과정이 불필요한 기온예측 방법론을 제시하고자 한다. 국내외로 기계학습을 이용한 기온예측모델에 관한 연구는 수차례 수행된 바 있다. Sharma et al.(2011)은 기계학습을 이용하여 태양 강도예측 모델을 생성하는 방법을 제안하였으며(Sharma et al., 2011), Zhang et al.(2017)은 기계학습을 이용하여 해수면 온도를 예측하는 모델을 제안하였다(Zhang et al., 2017). 국내의 경우 Cho et al.(2019)은 기계 학습을 통한 미세먼지 예측을 수행하였고(Cho et al., 2019), Yoon et al.(2017)은 LSTM의 성능을 검증하고 이를 이용하여 국내 광주지역의 24시간 기온예측을 수행한 바 있다(Yoon et al., 2017). 그러나 지역별로 폭염으로 인한 영향에는 차이가 있으며 이에 본 연구에서는 인구대비 폭염 사망률이 가장 높은 지역으로 기온예측을 수행하였고 4시간 예측을 통해 오차를 최소화하고자 하였다. 예보업무규정 부칙 제 17조에 따르면 현재 국내 폭염특보발표기준은 폭염주의보와 폭염경보가 있으며 폭염주의보는 일 최고기온 33°C 이상인 상태가 2일 이상 지속될 것으로 예상될 경우, 폭염경보는 일 최고기온 35°C 이상인 상태가 2일 이상 지속될 것으로 예상될 경우 발령한다(KMA, 2019). 2°C의 근소한 차이로 특보의 종류가 달라질 수 있다. 이처럼 폭염 대비를 위한 기온예측에서 가장 중요한 것은 정확도이다. 따라서 본 연구에서는 기온예측모델의 정확도를 우선적으로 고려하였고 결과의 정확도를 분석하여 수치

예보모델을 이용한 기온예측의 정확도와 비교하였다. 본 연구에서 제안하는 기온예측 모델은 과거의 데이터를 이용하여 미래의 데이터를 예측한다. 과거의 입력 값과 결과 값의 패턴을 분석하고 이를 학습하여 현재 시점의 입력 값을 제공하였을 때 분석한 패턴에 맞게 결과 값을 출력한다. 본 모델의 입력 값으로는 기상변수들을 사용하였으며 결과 값은 기온이다.

2. 본론

2.1 연구 내용

2.1.1 적용 대상

과거 15년(1998~2012년)간 인구대비 폭염 사망률이 가장 높은 지역은 ○○도 ○○군이다(Park et al., 2016). 따라서 대상 지역을 ○○군으로 선정하였다. 또한 우리나라 여름철(6~8월) 평균기온은 1979년 이후 현재까지 강한 변동성과 함께 증가하는 추세를 보이며, 여름철 기온의 중요성이 강조되고 있다(Won et al., 2017). 따라서 기온예측을 위한 과거 기상자료는 2010년부터 2018년까지 9년간의 여름철의 자료를 이용하였다. 또한 예측의 정확도 검토를 위한 비교자료는 2019년 여름철 기상자료를 이용하였으며, 4시간 후의 기온예측을 수행하였다. 기상자료는 기온, 강수량, 풍속, 습도, 증기압, 일사량 데이터를 이용하였다.

2.1.2 연구 방법론

현재 기상청에서는 수치예보모델을 이용하여 기상예보를 수행하며, 단기예보를 통해 폭염에 대한 정보를 전달한다. 관측 장비로부터 대기상태의 관측 값을 수신하고 수치예보모델이 관측 값을 이용하여 예측자료를 생성한다. 동네예보는 관측 자료 및 수치예보모델의 예측결과를 기반으로 전국을 5 km 공간해상도 격자로 나눈 뒤, 예보관이 읍·면·동에 대한 날씨 예측 정보를 편집하여 동네예보로써 제공한다(KMA, 2019).

본 연구에서는 기계학습을 이용한 기온예측모델을 제안한다. 기계학습의 학습방법으로는 지도학습, 비지도 학습, 강화학습이 있다. 본 연구에서는 지도학습방법을 적용하였는데 이는 입력된 데이터에 대한 답이 주어진 상태에서 학습시키는 방법이다. 지도학습은 분류와 회귀로 나뉘는데, 분류는 데이터가 주어졌을 때 데이터의 카테고리를 판단하는 방식이며 회귀는 데이터가 주어졌을 때 이의 특징을 기반으로 하여 연속된 값을 예측하는 방식이다. 따라서 과거의 데이터를 학습하여 미래의 데이터를 예측할 수 있다. 모델을 실행시키기 위해 입력해야하는 과거 데이터로는 학습 데이터 셋(train set)과 검증 데이터 셋(test set)이 있다. 본 연구에서의 학습 데이터 셋은 과거 수년간의 기상자료들과 그로부터 일정 시간 후의 기온을 포함한다. 검증 데이터 셋은 최근의 기상자료들과 일정시간 후의 기온을 포함한다. 이를 이용하여 모델을 실행하면 모델은 학습 데이터 셋의 기상자료들이 일정시간 후의 기온을 형성하기까지 어떤 패턴을 이루는지를 분석하고 이를 학습한다. 그 후 검증 데이터 셋의 기상자료에 학습한 패턴을 적용하여 일정시간 후의 기온을 예측하고, 이를 이미 입력되어있던 기온 데이터와 비교하여 모델 자체적으로 정확도를 검토하며 반복계산을 통해 오차를 줄여나간다(Fig. 1).

본 연구에서는 4시간 예측을 수행하였다. ○○군의 2010년도부터 2018년도까지 9년간의 여름(6월~8월)의 시간 단위 기상자료와 각 시간별 4시간 후의 기온을 입력하여 이를 학습 데이터 셋으로 사용하였다. 검증 데이터 셋으로는 2019년도의 6월부터 8월까지의 기상자료와 4시간 후의 기온을 입력한 자료를 사용하였다. 기상자료로는 기온, 강수량, 풍속, 습도, 증기압, 일사량자료를 수집하였고, t-test를 수행하여 각 기상변수의 기온예측과정에서의 유의성을 판단하였다. 유의하다고 판단된 변수는 예측하고자하는 데이터인 기온과의 관계를 분석하기위해 상관분석을 수행하였다. 상관계수가 양수일 경우 두 변수는 양의 상관관계를, 상관계수가 음수일 경우에는 음의상관관계를 가지며, 절댓값이 클수록 상관성도 크다고 볼 수 있다. 본 연구에서는 기온, 강수량, 풍속, 습도, 증기압, 일사량이 모두 유의하다고 판단하였고, 각 변수들과 기온과의 상관성 분

석을 수행하였다. 이후, 기온예측모델에서 사용하는 매개변수들의 최적조건을 찾기 위해 매개변수의 값을 다르게 하여 몇 가지 경우를 설정하였다. 설정한 경우대로 매개변수의 값을 입력하고 모델을 실행시켜 각 경우마다 오차를 분석하고 오차가 가장 작은 경우를 매개변수의 최적조건으로 선정하였다. 이를 통해 찾은 매개변수의 최적 값을 입력하고 모델을 실행시켜 도출한 결과가 최종 예측데이터이다. 최종 데이터의 정확도를 검토를 위해 오차지표 RMSE(Root Mean Square Error), MAE (Mean Absolute Error), R^2 를 이용하였다. 이와 같은 과정을 거쳐 도출해낸 최적 조건에서의 기온예측 결과와 수치예보모델을 이용한 기온예측 결과의 정확도를 비교하였다.

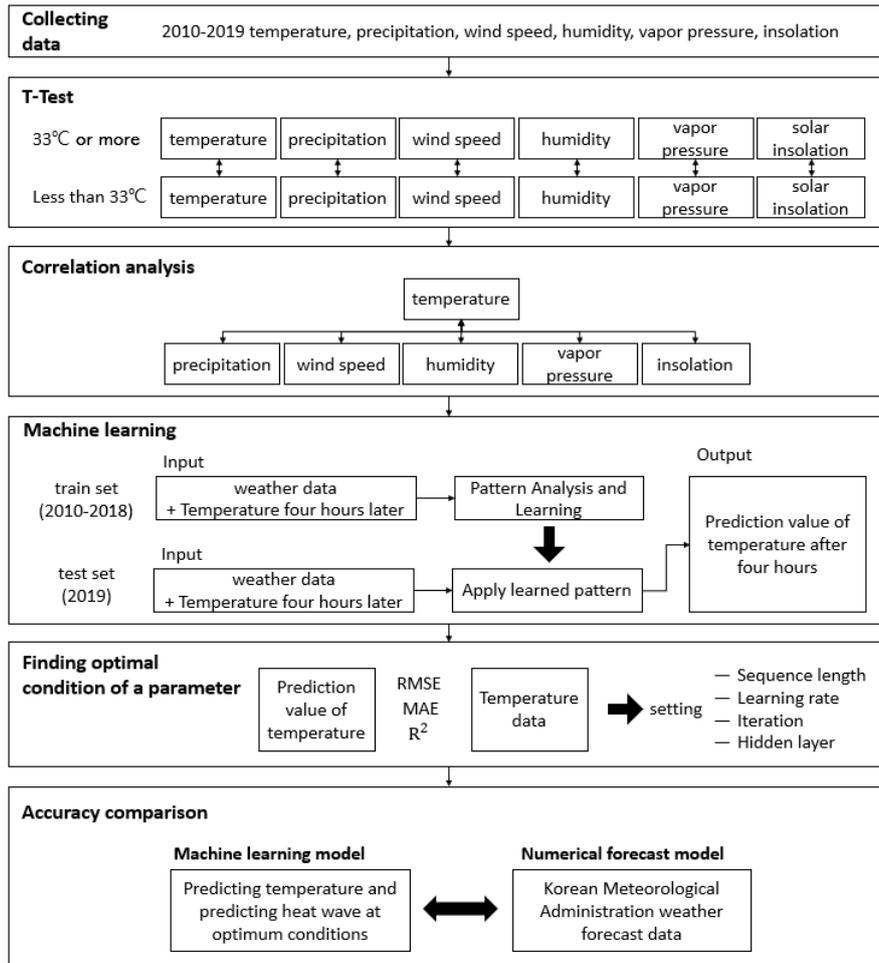


Fig. 1. Schematic process of this study

2.1.2.1 T-test

T-test는 두 집단의 평균을 비교하는 검정 방법이다. 집단을 대표할 수 있는 값인 평균을 비교하여 두 집단 간의 차이를 보고 평균차이가 유의미한 차이인지를 판단한다. 본 연구에서는 폭염발생여부를 기준으로 기상변수를 각각 두 집단으로 나누었다. 기온 관측 자료가 폭염발생기준인 33°C 이상의 값을 가진 경우와 33°C미만의 값을 가진 경우의 두 가지 집단으로 나누었으며, 각 기상변수별로 두 집단 간의 T-test를 통하여 자료의 유의성을 판단하였다. 이때, p-value(유의확률)값을 기준으로 판단하며, 이는 귀무가설이 관찰된 데이터와 양립하는 정도를 나타낸다.

2.1.2.2 상관성분석

상관성 분석은 두 변수간의 상관관계를 파악하기 위해 수행한다. 이를 통해 도출되는 상관계수는 -1과 1사이의 값을 가진다. 상관계수가 양수일 경우 두 변수는 양의 상관관계를, 상관계수가 음수일 경우에는 음의상관관계를 가지며, 절댓값이 클수록 뚜렷한 상관관계를 가진다고 볼 수 있다. 상관계수를 판단하는 기준에는 피어슨 상관계수, 스피어만 상관계수 등 다양한 기준이 존재하며, 본 연구에서의 해석에는 보편적으로 사용되는 피어슨 상관계수를 적용하였다.

2.1.2.3 LSTM

인공신경망은 기계학습 방법의 일종이며 입력 층과 출력 층으로 구성된다. 입력 층과 출력 층에 은닉 층이 존재하기도 하는데, 2개 이상의 은닉 층이 존재하는 경우를 심층신경망이라고 하며, 이를 학습시키는 것을 딥러닝이라고 한다. RNN (Recurrent Neural Networks)은 은닉 층 결과 값을 출력 층으로만 보내는 기존의 인공신경망과 달리 은닉 층의 결과 값을 출력 층으로 보내는 동시에 은닉 층의 다음 계산 입력 값으로도 사용한다. LSTM(Long Short Term Memory)은 RNN의 일종이며, 기존 RNN의 장기 의존성문제를 보완하는 모델이다. 기존의 RNN은 이전의 정보가 현재의 계산결과에 영향을 미치지만 이는 단기에만 해당되며 장기적인 영향을 미칠 수 없다. 반면 LSTM은 장기적인 정보 저장이 가능하다. LSTM cell은 현재 시점의 입력 값 x_t 와 과거시점 은닉상태 h_{t-1} 를 이용하여 현재시점의 출력 값 \hat{y}_t , 현재시점의 은닉상태 h_t 를 출력한다. LSTM cell의 구조는 Fig. 2과 같다. 입력게이트와 망각게이트는 과거정보와 현재정보에서 필요한 정보를 저장하고 불필요한 정보는 삭제하며 이를 셀 상태 C_t 로 표현한다. 출력게이트는 셀 상태 정보를 이용하여 출력정보를 결정한다. 본 연구 대상인 폭염여부를 결정하는 기온과 같이 과거의 정보를 기반으로 하고 현재의 정보를 이용하여 미래의 결과를 도출하고자 하는 연구에서는 이러한 장기적 기억이 필수적이다. 따라서 본 연구에서는 LSTM을 이용하여 기계학습을 수행하였다

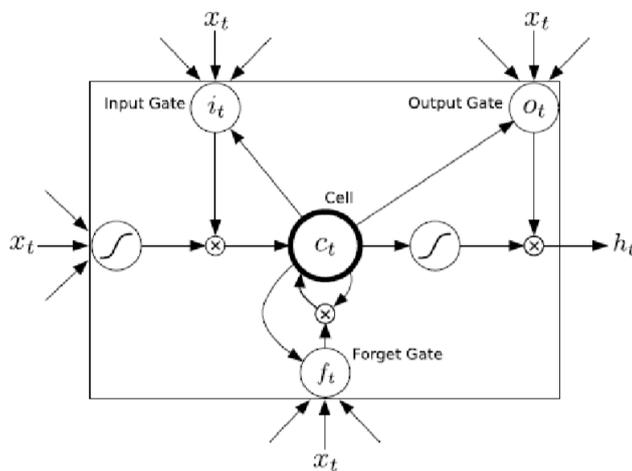


Fig. 2. Structure of LSTM cell (Qin Zhang, 2017)

2.1.2.4 오차지표

본 연구에서 사용된 오차지표는 RMSE, MAE 및 결정계수 R^2 이다. 일반적으로 RMSE와 MAE는 추정 값과 관찰 값을 비교하는 지표이다. RMSE는 오차제곱평균에 루트를 취한 것으로, 큰 오차에 민감하고, MAE는 오차 절댓값의 평균으로 RMSE에 비해 작은 오차에 민감하다. 기상예보 평가에 있어서, 기상변수별로 예보 검증지표가 다르며 일반적으로 기온에 대해서는 RMSE, MAE 등을 이용한다(Ahn, 2017).

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}} \quad (1)$$

$$MAE = \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{n} \quad (2)$$

여기서 \hat{y}_i : 추정 값, y_i : 관찰 값이다.

결정계수 R^2 은 선형회귀분석에서 추정된 모델이 주어진 자료에 적합한 정도를 나타내는 지표이다. 이는 분산을 기반으로 하여, 예측 데이터와 관측 데이터의 분산정도를 비교할 수 있다. 0에서 1 사이의 값을 가지며 1에 가까울수록 적합한 모델임을 의미한다. RMSE와 MAE의 경우, 데이터의 스케일에 따라 값에 큰 차이가 발생할 수 있으므로 상대적 값인 R^2 을 이용하였다.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (3)$$

여기서 \hat{y}_i : 추정 값, y_i : 관찰 값, \bar{y}_i : 관찰 값 평균이다.

2.2. 연구결과

2.2.1 T-test 결과

95%신뢰구간에서 p-value가 0.05이하이면 그 변수는 유의하다고 판정할 수 있다. 기계학습에 입력 값으로 사용할 기상변수인 기온, 강수량, 풍속, 습도, 증기압, 일사량을 대상으로 t-test를 수행한 결과, 모든 기상변수의 p-value 값이 약 2.2e-16으로 판정되었다. 따라서 모든 변수들이 유의하다고 판단하였으며 모든 변수를 기온예측에 사용하였다.

2.2.2 상관성분석 결과

본 연구에서는 예측하고자 하는 값이 기온이므로 기온과 각 변수들과의 관계의 선형성을 확인하기 위하여 상관성 분석을 수행하였다. Table 1은 상관성분석의 결과이며 각 값은 상관계수이다. 상관계수의 해석에는 보편적으로 쓰이는 피어슨 상관계수를 사용하였는데 상관계수가 0.3에서 0.7사이일 경우 뚜렷한 양적 선형관계임을, 상관계수가 -0.7과 -0.3 사이일 경우 뚜렷한 음의 상관관계를 가진다. 또한 상관계수가 -0.1과 0.1사이일 경우 거의 무시될 수 있는 선형관계임을 의미한다. 강수량과 기온과의 선형성은 비교적 낮으며 일사량과 기온과의 상관성이 높은 양의 상관관계를 가지고 있음을 확인할 수 있다. 강수량을 제외하고 나머지 변수들은 모두 기온과 뚜렷한 선형관계를 가진다.

Table 1. Result of correlation analysis

	Precipitation	Wind speed	Humidity	Vapor pressure	Solar insolation
Temperature	-0.07	0.44	-0.58	0.58	0.64

2.2.3 매개변수 분석결과

기계학습의 매개변수로는 학습률(Learning rate), 시퀀스길이(Sequence length), 은닉 층(Hidden layer), 반복횟수(Iteration)를 사용하였다. 시퀀스는 데이터에 번호를 지정하여 나열한 것이며 시퀀스길이는 한 번에 몇 개의 데이터를 이용하여 예측을 수행할 것인지를 지정하는 변수이다. 은닉 층은 기계학습에서 입력 층과 출력 층 사이에 존재하는 층이며 개수가 많을수록 많은 단계를 거쳐 출력이 이루어지는 것이므로 예측 측면에서는 일반적으로 정확도가 향상된다. Iteration은 반복횟수를 의미하며 Learning rate는 학습률을 의미한다. 학습률은 기계학습의 학습과정에서 어느 정도로 세세하게 학습을 하는지를 결정하는 변수이다. 매개변수들의 최적조건을 설정하기 위하여 Table 2와 같이 16개의 경우를 설정하였다. case L은 다른 매개변수들의 조건은 고정시키고 학습률의 조건만 바뀌며 최적의 학습률 조건을 찾도록 설정하였고 최적조건을 찾은 후 조건을 고정하고 다음 매개변수의 최적조건을 찾는다. case S의 경우, 학습률은 앞서 설정한 최적조건으로 고정하고, 4개의 시퀀스길이 조건 중에서 최적 조건을 찾는다. 이와 같이, case H, case I는 각각 은닉 층 개수, 반복횟수의 최적조건을 찾도록 설정하였다. 최적조건은 오차지표인 RMSE 와 MAE, R²를 이용하여 오차가 가장 적은 경우를 최적의 조건이라고 판단하였다. 본 연구의 매개변수 최적조건 분석결과는 Table 2에 나타내었다. 모델실행 결과, Learning rate=0.005, Sequence length=8, Hidden layer=20, Iteration=2000일 경우에 RMSE, MAE값이 가장 작고 R² 값이 가장 크게 산정되었다. 따라서 이와 같은 조건을 매개변수의 최적조건으로 판단하였다. 따라서 최적조건으로 기온예측을 수행하였을 때 기계학습모델을 이용한 기온예측은 RMSE 1.705°C, MAE 1.269°C가 발생하였다.

Table 2. Sensitivity analysis for each parameter

Case	Learning rate	Sequence length	Hidden layer	Iteration	RMSE (°C)	MAE (°C)	R ²
L-1	0.0005	4	5	1000	3.03	2.40	0.47
L-2	0.001	4	5	1000	2.57	1.98	0.64
L-3	0.005	4	5	1000	2.35	1.77	0.70
L-4	0.010	4	5	1000	2.47	1.88	0.67
S-1	0.005	1	5	1000	2.98	2.33	0.53
S-2	0.005	2	5	1000	2.62	2.03	0.63
S-3	0.005	4	5	1000	2.35	2.40	0.47
S-4	0.005	8	5	1000	2.17	1.66	0.75
H-1	0.005	8	5	1000	2.17	1.66	0.75
H-2	0.005	8	10	1000	1.88	1.41	0.81
H-3	0.005	8	15	1000	1.89	1.42	0.81
H-4	0.005	8	20	1000	1.82	1.36	0.82
I-1	0.005	8	20	500	2.00	1.51	0.78
I-2	0.005	8	20	1000	1.82	1.36	0.82
I-3	0.005	8	20	1500	1.72	1.30	0.84
I-4	0.005	8	20	2000	1.71	1.27	0.84

2.2.4 기계학습모델과 수치예보모델의 예측결과 비교

각 매개변수를 최적 조건으로 설정하고 2019년도 ○○군의 기온 예측 값과 관측 값을 비교한 결과 1.71°C의 RMSE 값을 가지며 결과는 Fig. 3과 Fig. 4와 같다. 또한 기온 관측 값이 폭염발생의 기준인 33°C 이상일 경우의 RMSE는 1.39°C로 더욱 낮은 오차가 발생하였다. 기계학습모델을 통한 예측은 데이터의 스케일이 클수록 극값 예측에 대한 정확도가 저하되는 경향

을 보였다. 따라서 1년의 기온데이터를 이용하는 대신 여름철의 기온데이터만을 사용하였다. 또한, 33°C 이하보다 33°C 이상에서 더 높은 정확도를 나타내었는데, 이는 33°C 이하보다 33°C 이상 데이터의 스케일이 작기 때문이라고 판단하였다. 모델 실행에 소요되는 시간은 평균적으로 9분 26초 정도로 측정되었다.

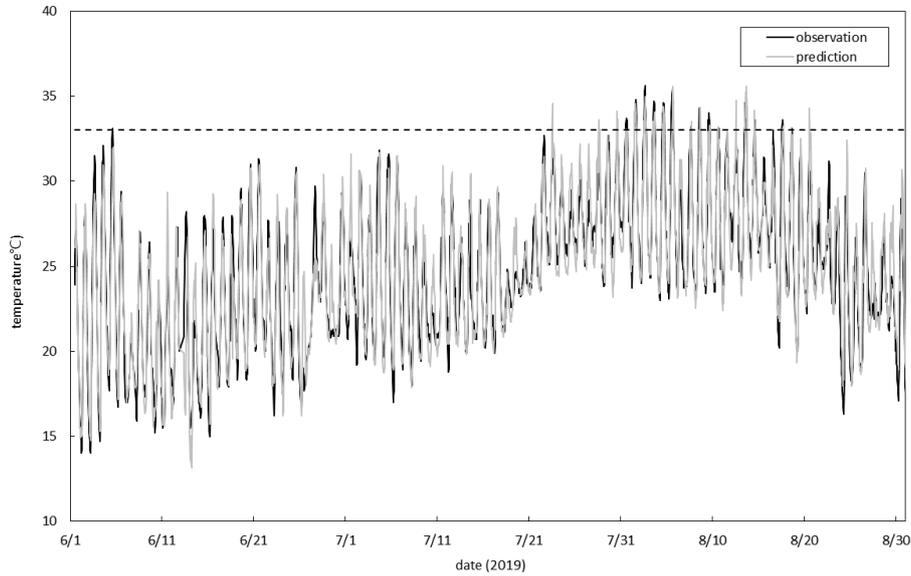


Fig. 3. Temperature prediction with LSTM (2019 ○○-gun)

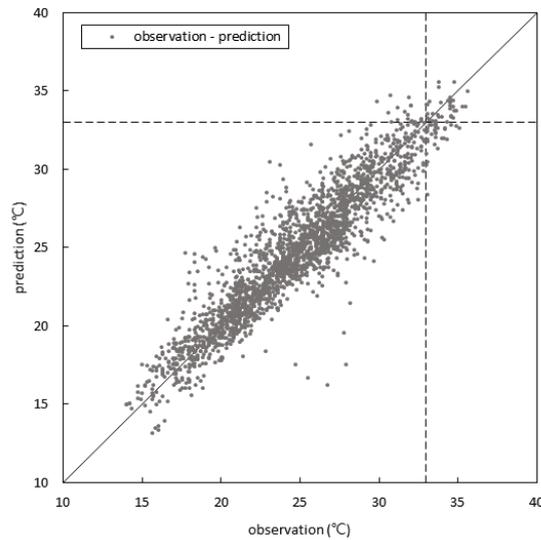


Fig. 4. Comparison of observations and machine learning model predictions (2019 ○○-gun)

○○도 ○○군의 기상관측지점은 중심점에서 약 20 km 떨어진 남쪽에 위치한 ○○읍에 있으므로 지리적 동질성을 확보하였다고 판단하였다. 기상청에서 제공하는 기상관측자료는 시군구 단위로 제공 되고 단기예보자료는 읍면리 단위까지 제공된다. 따라서 관측 자료는 ○○군의 자료를 수집하였고 예보자료는 ○○군에 속한 읍, 면별로 자료를 수집하였다. ○○군에는 ○○읍을 포함하여 총 13개의 읍과 면이 포함되어 있다. 관측 자료가 군단위로 제공되었기 때문에 전체적으로 예보의 정확도를 파악하기 위하여 읍, 면 각각의 예보자료의 오차를 분석하였으며 그 결과는 Table 3과 같다. ○○군 예보는 최소 3~4 시간 이상이 소요되며 RMSE 1.57°C, MAE 1.17°C의 값을 나타냈다.

Table 3. Result of numerical forecast model error analysis (2019 ○○-gun)

○○-gun	RMSE (°C)	MAE (°C)	R ²
○○-eup	1.53	1.15	0.87
○○-myeon	1.53	1.15	0.87
○○-myeon	1.82	1.44	0.82
○○-myeon	1.44	1.04	0.89
○○-myeon	1.51	1.11	0.88
○○-myeon	2.00	1.63	0.78
○○-myeon	1.54	1.14	0.87
○○-myeon	1.41	1.00	0.89
○○-myeon	1.54	1.14	0.87
○○-myeon	1.43	1.04	0.89
○○-myeon	1.46	1.06	0.89
○○-myeon	1.70	1.33	0.85
○○-myeon	1.44	1.04	0.89
Mean	1.57	1.17	0.87

Table 4는 수치예보모델과 기계학습모델의 기온예측 오차와 소요시간을 나타낸 표이다. 수치예보모델과 기계학습모델의 기온예측에 대한 오차는 RMSE 0.14°C, MAE 0.10°C로 근소한 차이를 보이며 R²값 또한 차이가 크지 않다. 그러나 예보 소요시간 측면에서 큰 차이를 보임을 확인할 수 있다. 수치예보모델을 이용한 예측의 소요시간을 3시간이라고 가정했을 때, 기계학습모델을 이용한 예측의 소요시간은 9분26초로 94.79%의 큰 감소율을 보였다. Table 5는 관측 값이 33°C 이상인 경우 두 모델의 기온예측 오차를 나타낸 표이다. 전범위의 오차는 수치예보모델이 더 작은 값을 가진 반면, 33°C 이상의 경우에는 기계학습모델을 통한 예측이 더 작은 오차를 가지며 높은 정확도를 보였다. 수치예보모델을 통한 기온예측과 기계학습을 통한 기온예측 결과데이터의 분포를 33°C 전후로 나누어 확인해본 결과를 Fig. 5에 나타내었다. 관측 값 33°C 미만의 범위에서는 기계학습모델과 수치예보모델의 예측 값이 비슷한 정도의 분포를 보인다. 관측 값 33°C 이상의 범위에서는 예측 값의 분포가 차이를 보인다. 기계학습모델에서 위로 뺀 막대의 길이가 아래로 뺀 막대보다 긴 반면, 수치예보모델에서는 위 아래 막대의 길이가 같으며 다른 형태의 분포를 보인다. 또한 기계학습모델의 제1,2,3 사분위수가 각각 수치예보모델의 제1,2,3 사분위수보다 큰 값을 확인할 수 있는데 이는 동일한 관측 값 범위에서 기계학습모델이 수치예보모델보다 큰 값으로 예측하였음을 의미하므로 기계학습을 이용한 예측이 더욱 보수적이라고 판단할 수 있다.

Table 4. Prediction accuracy and time required for summer season

	Prediction using numerical forecast model	Prediction using machine learning model	Increase/Decrease
RMSE (°C)	1.57	1.71	8.95% increase
MAE (°C)	1.17	1.27	8.18% increase
R ²	0.87	0.84	2.65% decrease
Time required	More than 3~4 hours	Average 9 minutes and 26 seconds	94.79% decrease

Table 5. Prediction accuracy and time required (33°C or higher)

	Prediction using numerical forecast model	Prediction using machine learning model	Increase/Decrease
RMSE (°C)	1.96	1.39	28.94% decrease
MAE (°C)	1.72	0.92	46.42% decrease

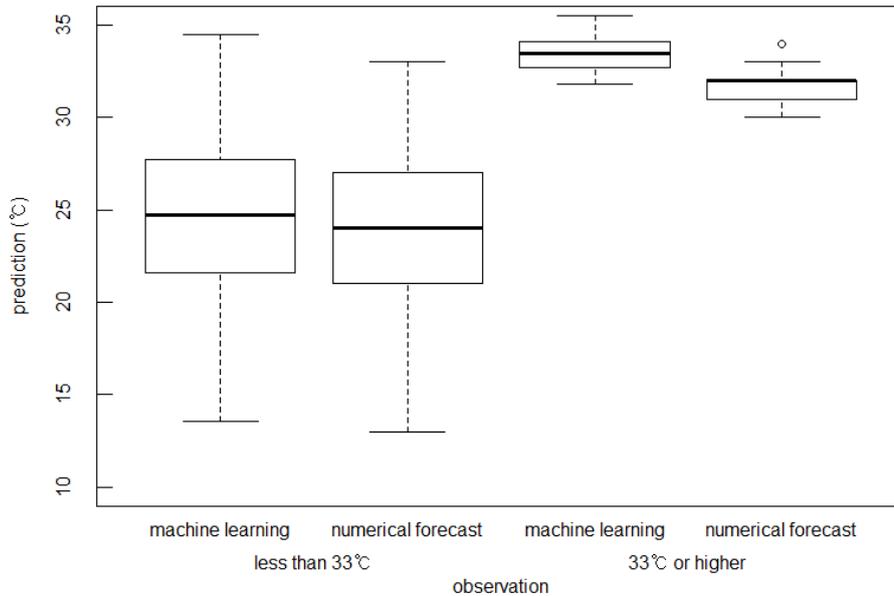


Fig. 5. Prediction distribution chart

3. 결론

본 연구에서는 기계학습 중에서도 LSTM을 이용하여 국내 ○○도 ○○군의 4시간 후의 기온예측을 수행하였다. 해당 지역의 2010-2018년도의 여름철 기상자료들을 이용하여 2019년도의 기온을 예측하였으며 다양한 오차지표를 이용하여 예측의 정확도를 검토하였다. 현재 기상청에서는 수치예보모델을 이용하여 기상예보를 실시한다. 이는 데이터를 수신하여 분석 자료를 출력하고 기상예보관의 분석을 거치는 단계까지 비교적 복잡한 과정을 거치며 3~4시간 이상의 시간이 소요되는 반면, 본 연구의 기온예측은 사람의 분석과정이 없는 간단한 과정과 약 10분 정도의 짧은 소요시간만을 필요로 한다. 본 모델의 4시간 기온예측 결과의 오차는 두 가지 경우로 나누어 분석하였다. 첫 번째 경우 전체 데이터의 오차를 분석한 결과는 기계학습모델이 약 1.71°C의 RMSE값을 가지는데, 이는 기상청예보모델의 오차인 1.57°C와 약 0.14°C, MAE값은 수치예보모델이 1.17°C, 기계학습모델이 1.27°C로 0.10°C의 근소한 차이를 보인다. 2019년 6월~8월 ○○군의 관측기온데이터가 최저기온 14°C, 최고기온 35.6°C로 21.6°C의 데이터 폭을 가지는데 1.71°C는 이의 약 7.89%, 1.27는 약 5.88%의 오차에 해당한다. 두 번째 경우 관측 값이 폭염주의보 기준인 33°C이상인 값들만 분석하였다. 그 결과 수치예보모델의 오차는 RMSE 1.96°C, MAE 1.72°C, 기계학습모델의 오차는 RMSE 1.39°C, MAE 0.92°C의 값을 가지며 수치예보모델보다 높은 정확도를 나타냄을 확인할 수 있었다. 본 연구에서 제안한 기계학습 모델을 통한 기온예측은 폭염대비를 목적으로 하므로, 폭염주의보 발령기준인 33°C이상의 데이터에서 수치예보모델을 이용한 기온예측보다 높은 정확도를 나타낸다는 것은 매우 유의미한 결과라고 판단할 수 있다.

본 연구의 기계학습모델은 ○○군을 대상으로 한 기온예측을 수행하였다. 따라서 ○○군이 아닌 다른 지역의 기온예측을 수행한다면 정확도가 저하될 가능성이 있다. 따라서 향후 지역별 기온 차이에 대한 검토를 수행하고 이를 반영할 수 있는 요

소를 포함하여 국내 모든 지역에 적용할 수 있는 보편적 모델을 개발한다면 기온예측모델의 효율성이 크게 증가될 것이라 기대한다. 또한 본 모델을 이용하여 4시간 예측을 수행하는데 있어서 8% 이내의 오차가 발생하였는데 향후, 이와 같은 수준의 정확도를 유지하며 12시간, 24시간 등으로 예측시간 증가가 가능하다면 폭염에 대하여 더욱 빠르고 철저한 대처가 가능할 것이라 기대한다.

Acknowledgments

This work was supported by Korea Environment Industry & Technology Institute (KEITI) through Water Management Research Program, funded by Korea Ministry of Environment (127572).

References

- Ahn, J. and Jeong, C. (2018). Numerical Simulation of the Flood Event Induced Temporally and Spatially Concentrated Rainfall - On August 17, 2017, the Flood Event of Cheonggyecheon. *Journal of Korean Society of Disaster and Security*. 11(2): 45-52.
- Ahn, S. (2016). Deep Learning Architectures and Applications. *Journal of Intelligence and Information Systems*. 22(2): 127-142.
- Choi, J. (2019). Proposal of Early-Warning Criteria for Highway Debris Flow Using Rainfall Frequency (1): Proposal of Rainfall Criteria. *Journal of Korean Society of Disaster and Security*. 12(2): 1-13.
- Choi, M. H. and Yun, J. I. (2009). On Recent Variations in Solar Radiation and Daily Maximum Temperature in Summer. *Korean Journal of Agricultural and Forest Meteorology*. 11(4): 185-191.
- Joins of Related Ministries (2018) Abnormal Weather Report. Seoul: Korea Meteorological Administration.
- Kim, J., Lee, D. G., Park, I. S., Choi, B. C., and Kim, J. S. (2006). Influences of Heat Waves on Daily Mortality in South Korea. *Atmosphere*. 16(4): 269-278.
- Korea Meteorological Administration (2018). Evaluation of Contribution of Meteorological Observation Data to Weather Forecasts. Seoul: Korea Meteorological Administration.
- Korea Meteorological Administration (2019). Forecast Services Regulations APPENDUM Article17. Seoul: Korea Meteorological Administration.
- Korea Meteorological Administration (2019). A Study on the Diagnosis and Development Direction of the Forecasting System. Seoul: Korea Meteorological Administration.
- Lee, S. G., Jung, S. G., Lee, W. S., and Park, G. H. (2011). A Predictive Model for Urban Temperature Using the Artificial Neural Network. *Korea Planners Association*. 46(1): 129-142.
- Park, J.E., Heo, B.Y., and Sunwoo, Y. (2016). A Study on Human Damage due to Heat Wave by Region. *Journal of the Korean Society of Hazard Mitigation*. 16(1): 103-109.
- Qing, X. and Niu, Y. (2018). Hourly Day-ahead Solar Irradiance Prediction Using Weather Forecasts by LSTM. *Energy*. 148: 461-468.
- Sharma, N., Sharma, P., Irwin, D., and Shenoy, P. (2011). Predicting Solar Generation from Weather Forecasts Using Machine Learning. In 2011 IEEE International Conference on Smart Grid Communications (SmartGridComm). 528-533.
- Won, Y. J., Yeh, S. W., Yim, B. Y., and Kim, H. K. (2017). Relationship between Korean Monthly Temperature during Summer and Eurasian Snow Cover during Spring. *Atmosphere*. 27(1): 55-65.
- Yoo, H., Lee, S. O., Choi, S., and Park, M. (2019). A Study on the Data Driven Neural Network Model for the Prediction of Time Series Data: Application of Water Surface Elevation Forecasting in Hangang River Bridge. *Journal of Korean Society of Disaster and Security*. 12(2): 73-82.

- Yoon, J.W. and Jeon, M.G. Temperature Forecasting Model by Using Deep Learning Technology based on LSTM. Proceeding of Institute of Electronics and Information Engineers Conference. 912-915.
- Zhang, Q., Wang, H., Dong, J., Zhong, G., and Sun, X. (2017). Prediction of Sea Surface Temperature Using Long Short-term Memory. IEEE Geoscience and Remote Sensing Letters. 14(10): 1745-1749.

Korean References Translated from the English

- 관계부처합동 (2019). 2018 이상기후보고서. 서울: 기상청.
- 기상청 (2018). 기상관측자료의 예보 기여도 평가. 서울: 기상청.
- 기상청 (2019). 동네예보체계 진단 및 발전방향에 관한 정책 연구. 서울: 기상청.
- 기상청 (2019). 예보업무규정 부칙 제17조. 서울: 기상청.
- 김지영, 이대근, 박일수, 최병철, 김정식 (2006). 한반도에서 여름철 폭염이 일 사망률에 미치는 영향. 대기. 16(4): 269-278.
- 박재은, 허보영, 선우영 (2016). 폭염에 의한 지역별 인명피해 발생에 관한 연구. 한국방재학회 논문집. 16(1): 103-109.
- 안성만 (2016). 딥러닝의 모형과 응용사례. 지능정보연구. 22(2): 127-142.
- 안정환, 정창삼 (2018). 시공간적으로 편중된 강우에 의한 홍수사상 수치모의 - 2017년 8월 17일 청계천 홍수사상을 대상으로. 한국방재안전학회 논문집. 11(2): 45-52.
- 안중배 (2017). 기상예보의 과정. 감사. 136: 40-46.
- 원유진, 예상욱, 임보영, 김현경 (2017). 우리나라 여름철 월별 기온 변동성과 유라시아 봄철 눈덮임 간의 상관성 분석. 대기. 27(1): 55-65.
- 유형주, 이승오, 최서혜, 박문형 (2019). 시계열 자료의 예측을 위한 자료 기반 신경망 모델에 관한 연구: 한강대교 수위예측 적용. 한국방재안전학회 논문집. 12(2): 73-82.
- 윤재용, 전문구 (2017). LSTM 기반 딥러닝 기술을 활용한 기온 예측 모델. 대한전자공학회 학술대회. 912-915.
- 이슬기, 정성관, 이우성, 박경훈 (2011). 인공신경망을 이용한 도시기온 예측모형 구축. 국토계획. 46(1): 129-142.
- 조경우, 정용진, 강철규, 오창현 (2019). 미세먼지 예측을 위한 기계 학습 알고리즘의 적합성 평가. 한국정보통신학회논문지. 23(1): 20-26.
- 최미희, 윤진일 (2009). 여름철 일최고기온과 일사량의 최근 변동에 관하여. 한국농림기상학회지. 11(4): 185-191.
- 최재순 (2019). 확률 강우량을 이용한 고속도로 토석류 조기경보기준 제안 (1) : 강우기준 제안. 한국방재안전학회 논문집. 12(2): 1-13.