

LSTM과 GRU를 활용한 도시공간 특성 기반의 평균기온 예측 모델: 강원도 원주시를 대상으로

Prediction Model of Average Temperature based on Characteristic of Urban-space Using LSTM and GRU: The Case of Wonju City

이우섭 Lee Wooseop*, 김형규 Kim Hyungkyoo**

Abstract

As the annual average temperature continues to rise due to climate change caused by global warming, the incidence of heat diseases and the number of deaths are also increasing, which is expected to require various alternatives and research. In this study, the average temperature rise-related variables are extracted through statistical analysis for Wonju City, where the average temperature increase rate and change are high, and the average temperature is predicted by utilizing deep learning-based LSTM and GRU based on the extracted variables. Three models were extracted through correlation and regression analysis for 26 variables collected based on prior research consideration, based on which LSTM and GRU analysis were conducted. The analysis showed the lowest MSE of LSTM - 0.4399(2.94°C), GRU - 0.4444(2.97°C) in the third model with 12 variables, with little MAE difference between validation and test data. This study is significant in that it extracted variables through statistical analysis and predicted average temperature rise using deep learning as a data acquisition method for adapting the annual average temperature rise problem. In addition, it is expected that urban space factors that affect the average temperature rise in Wonju City will be extracted along with predicting the trend of average temperature change, and appropriate measures will be prepared to take into account regional impact factors, not uniform climate change adaptation.

Keywords: Prediction of Average temperature, Deep-Learning, LSTM, GRU, Wonju

I. 서론

1. 연구의 배경 및 목적

우리나라는 지구온난화에 따른 기후변화로 인해 여름은 길어지고 겨울이 짧아지는 현상이 나타나고 있으며,

연평균기온은 +0.18°C/10년으로 상승하였다(기상청 2018). 또한, 전지구모델(HadGEM2-AO)을 통한 2050년 이전의 단기 기후변화 예측 결과, 한반도 평균기온은 모든 대표농도경로(Representative Concentration Pathways: RCP) 시나리오에서 +1.05°C~+1.95°C 상승하고, 폭염 등 고온 관련 극한지수가 증가할 전망이

* 홍익대학교 스마트도시전공 석사과정(제1저자) | Master's Student in Smart Cities Program, Hongik Univ. | Primary Author | wooseabi@naver.com

** 홍익대학교 도시공학과 조교수(교신저자) | Assistant Prof., Dept. of Urban Design & Planning, Hongik Univ. | Corresponding Author | gusailsang@gmail.com

다(기상청 2020). 기상청의 종관기상관측 자료를 살펴 보면 2001년부터 2018년까지 연평균기온 상승률이 높은 지역은 강원 홍천, 경북 의성, 강원 원주 등으로 분석되며, 연평균기온의 변화량은 원주에서 약 1.6°C로 높게 나타났다(이재호, 백희정, 현유경, 조천호 2011). 기온상승은 인간의 건강에 악영향을 미치며, 고온에 장시간 노출되면 온열질환 발생률과 그에 따른 사망자 수도 증가하기 때문에 기후변화에 대한 적응방안이 마련되어야 한다(장재연 2009). 따라서 높은 기온에 따른 폭염 등의 극한기후에 대한 사전 대책 마련을 위해 연평균기온 상승률과 변화량이 높은 지역에 대한 정확한 평균기온 예측 연구가 수행될 필요가 있다.

한편, 산업의 고도화에 따라 다양한 분야에서 인공지능과 관련한 많은 연구가 진행되고 있다. 특히 딥러닝 네트워크 중 하나인 순환신경망(Recurrent Neural Network: RNN)은 자연어, 날씨, 주가 등 순서가 있고, 시간의 흐름에 따라 변화하는 데이터를 처리하는 데 좋은 성과를 보인다(Karpathy, Johnson and Fei-Fei 2015). 그러나 RNN은 입력 데이터가 길어질수록 정확한 예측이 어려운 장기 의존성(Long-Term Dependency) 문제가 존재하였다(Bengio, Simard and Frasconi 1994). 이러한 장기 의존성 문제를 해결하기 위해 장단기메모리(Long Short Term Memory: LSTM) 모델과 LSTM의 변형인 게이트 순환 유닛(Gated Recurrent Unit: GRU) 모델이 제안되었다(Hochreiter and Schmidhuber 1997; Cho, Van Merriënboer, Gulcehre and Bahdanau et al. 2014). LSTM은 입출력 사이에 셀 상태(cell state)를 추가하여 앞선 레이어의 정보를 보존하거나 버림으로써 장기 의존성 문제를 해결하였고, GRU는 LSTM의 구조를 간소화하여 학습속도가 빠르고 LSTM과 비슷한 성능을 보인다(Chung, Gulcehre, Cho and Bengio 2014). 이처럼 시간의 흐름에 따라 변화하는 데이터의 분석

에 활용되는 딥러닝 기반의 LSTM과 GRU 모델은 향후 기후변화 예측에 적합할 것으로 사료된다. 또한 딥러닝 학습을 통한 평균기온 예측을 위해 상관분석 및 회귀분석을 통해 평균기온 상승에 영향을 주는 도시 공간 특성 변수를 추출하여 도시계획 차원에서의 극한기후에 대한 사전 대책 마련할 필요가 있다.

이에 본 논문에서는 연평균기온 상승에 따른 기후 문제의 사전 대책 마련을 위한 데이터 확보 방안으로 딥러닝 기반의 LSTM과 GRU를 통해 연평균기온 상승률과 변화량이 높은 강원 원주시를 대상으로 평균기온을 예측하는 모델을 제안하고자 한다. 이를 통해 원주시에 적합한 기후변화 적응대책으로서의 지역 및 도시계획 시사점을 제공할 수 있을 것이다.

본 연구는 다음과 같은 순서로 진행된다. 2장에서는 평균기온 및 극한일수, 도시환경에 따른 기온변화 연구와 인공지능 기반 기후분석 연구를 고찰하여 본 연구의 차별성을 도출한다. 다음으로 3장에서는 연구에 사용된 방법론을 개략적으로 기술하고, 4장에서 통계분석을 통해 평균기온 상승에 영향을 주는 변수를 추출한 후 LSTM과 GRU를 활용한 평균기온을 예측한다. 마지막으로 5장에서 학습 결과를 정리하고 시사점 및 향후 연구를 제시한다.

II. 선행연구 고찰

본 장에서는 평균기온 및 극한일수 분석 연구와 도시환경에 따른 기온변화 연구, 인공지능 기반 기후분석 연구를 고찰하여 본 연구의 차별성을 도출하고자 한다.

1. 평균기온 변화 및 극한일수 분석 연구

평균기온 변화 및 극한일수 분석을 위해 측정지점에 따라 과거부터 현재까지의 평균기온과 극한지수의 변

화 및 경향 연구가 진행되었다(윤진일 2002; 이승호, 허인혜 2011; 김진아, 김규량, 조창범, 김백조 2015). 또한, 분석된 데이터를 통해 미래 기후 전망을 제시하고, RCP 시나리오와의 비교분석도 연구되었다(김진아, 김규량, 조창범, 김백조 2015; 유동수, 권오창, 김흥기 2020).

윤진일(2002)은 30년 단위로 시기를 구분해 14개 지점에 대한 기온 변화량과 인구 증가 추세를 비교하여 도시화에 따른 기온상승 경향을 정량적으로 분석했다. 이를 토대로 월별 기온 변화량을 지구온난화에 의한 부분과 도시화에 기인한 부분으로 나누었으며, 순수 지구온난화 효과는 12월부터 4월 사이에서만 나타났고, 도시화에 따른 승온효과는 주로 일 최저기온의 상승을 주도한다는 것을 밝혀냈다.

이승호, 허인혜(2011)는 기온변화에 대한 도시화 효과를 분석하고자 우리나라의 60개 지점을 대도시, 중소도시, 비도시로 구분하여 총 16개의 극한지수 변화 경향을 분석하였고, 도시화 효과가 극한기온의 변화에 미치는 영향을 규명하였다. 분석 결과, 여름 관련 극한지수는 증가하였고, 겨울 관련 극한지수는 감소하였으며, 겨울철 변화폭이 더 크게 나타났다. 또한 도시 규모가 클수록 변화 경향이 더 크게 나타나 도시화가 기온변화에 큰 영향을 주고 있음을 밝혀냈다.

김진아, 김규량, 조창범, 김백조(2015)는 기후변화 시나리오에 따른 과거 제주도 지역 지점별 폭염 및 열대야 일수 차이에 대한 원인을 최저수온과 운량을 이용하여 분석하고, 미래 전망을 제시하였다. 분석 결과, 제주는 최저수온이 높아 폭염일수가 가장 많았고, 서귀포는 새벽 시간 운량이 상대적으로 많아 열대야 일수가 가장 많이 나타났다.

김장경, 정창삼, 권현한(2019)은 시간적 특성 및 공간적 특성(읍면동별 토지이용 및 평균 표고, 도시계획

시설 등)과 센서스 통계(인구, 가구, 사업체 등)를 연계한 폭염 취약성 관련 빅데이터 DB를 구축하고, 온열 질환 신고현황과 연계하여 합리적 폭염 취약성 지수를 개발하였다. 이를 통해 인공지능 기반의 의사결정 지원시스템과 연계하여 빅데이터를 활용한 잠재적 위험과 불확실성 요소를 분석하여 폭염에 대한 선제적 대응이 가능하다고 판단하였다.

유동수, 권오창, 김흥기(2020)는 미래 기후변화 시나리오에 대한 불확실성 자료를 보완하기 위해 특정 지점의 평균기온에 대해 2차 방정식을 이용한 회귀모형으로 자료를 보완하고, 비선형 회귀모형을 통해 적산온도를 산출해 변화를 추정하는 RCP 시나리오 기반 생장도일 보간프로그램(RCP scenario-based Growing degree days Interpolation program: RGI)를 개발하였다. 프로그램 성능 분석 결과, 잔차표준오차에서 우수한 성능을 보여 RCP 시나리오의 불확실성에 대한 오류를 보완할 수 있을 것으로 판단하였다.

대부분의 평균기온 변화 및 폭염, 열대야 등의 극한 일수 분석 연구는 기간 및 지점별 변화에 대한 분석과 함께 미래 기후 전망을 제시하고, RCP 시나리오와의 비교분석으로 이루어지고 있다. 그러나 많은 연구가 기후변화 현황 및 대책과 관련하여 진행되었고, 미래 기후변화 전망 예측을 위한 RCP 시나리오의 불확실성 자료 보완 연구는 다소 미비한 것으로 나타났다. 따라서 미래 기후변화 시나리오를 보완하기 위해 현재까지의 기후변화 추세와 다른 기상 변수에 따른 미래 기후변화 예측을 위한 연구 등 다양한 방법이 시도 될 필요가 있다.

2. 도시환경에 따른 기후변화 연구

도시공간 및 도시환경에 따른 기후변화를 분석하기 위해 대기 온도와 폭염 등에 영향을 미치는 도시환경

요소 추출 연구가 진행되었다(김학열 2003; 차영화, 김학열, 허태영 2009). 또한, 회귀분석 등을 통해 도시 환경요소와 기후변화 간의 상관관계 분석도 함께 이루어졌다(오규식, 홍재주 2005; 박경훈, 송봉근, 박재은 2016; 고동원, 박승훈 2019).

김학열(2003)은 서울시 자동기상측정망(Automatic Weather Station: AWS)으로부터 반경 500m의 원형 지역을 경계로 도시환경인자를 추출하고, 대기 온도를 활용한 단순회귀 및 다중회귀 모델을 통해 영향력을 분석하는 모델을 개발하였다. 분석 결과, 교통시설 지면적, 녹지, 불투수포장면적 등이 기온 저감과 관련하여 주요 영향 인자임을 도출하였다.

오규식, 홍재주(2005)는 위성영상을 이용하여 서울시의 기온분포도를 작성하고, 이를 토대로 토지이용, 용적률, 녹지 등과의 상관관계 및 회귀분석을 통해 도시 구성요소(토지이용, 녹지 등)와 기온의 관계를 파악하였다. 또한 열섬현상 저감을 위한 공원 및 녹지의 규모와 식생 특성에 따른 기온 저감효과를 분석하였고, 이를 토대로 효과적인 정책 수립과 도시 지속성 관련 대안을 마련할 수 있을 것으로 판단하였다.

차영화, 김학열, 허태영(2009)은 대도시권 도시 온도의 공간적 분포를 토지이용과 피복특성 등을 설명 변수로 2005년 월평균기온에 대한 회귀모형을 통해 대기 온도 영향인자를 추출하고, 영향력을 분석하였다. 분석 결과, 녹지 및 오픈스페이스 면적은 기온 저감에 영향을 주었고, 교통시설지와 주택지 면적은 기온상승에 영향을 주는 것으로 나타났다.

박경훈, 송봉근, 박재은(2016)은 밀양시를 대상으로 RCP 보고서 기반 폭염일수 자료를 산정해 지수화하고, 토지 피복과 지형 등의 공간자료를 구축하여 상관분석을 진행하여 공간 특성이 폭염일수에 미치는 영향을 분석하였다. 분석 결과, 도심 외곽지역의 시설

재배 중심 농경지의 폭염일수가 가장 많았고, 2000년 이후 산림지역이 감소하고 시가지지역과 농업지역이 증가하며 발생한 인공열 등의 영향으로 다른 지역에 비해 폭염이 많이 발생한 것으로 분석되었다.

고동원, 박승훈(2019)은 공간정보시스템을 활용하여 Landsat 8 위성영상 자료를 기반으로 도시열섬을 측정하고, Moran's I 테스트로 서울 지역 행정동 간 근린환경특성과 도시열섬현상의 공간적 자기상관성을 분석하였다. 분석 결과, 버스정류장 밀도는 도시열섬에 부정적인 영향을 미쳤고, 정규화식생지수(Normalized Difference Vegetation Index: NDVI)와 주거용 건물의 연면적은 도시열섬현상 완화에 긍정적 영향을 미쳤으며, 사회적 취약계층의 비율이 다소 양의 방향성을 나타내고 있음을 밝혀냈다.

송봉근, 박경훈, 김경아, 김정현 외(2020)는 경남 창원을 대상으로 노인인구 분포지역의 공간적 특성을 토지이용특성을 고려한 군집화 분석을 통해 5개 유형으로 나누고, 공간 유형에 따른 폭염 특성을 분석하였다. 노인 인구수는 도시지역에서 가장 많았으나, 전체 인구 대비 노인인구 비율은 산림지역과 농촌지역에서 더 높았다. 폭염일수 또한 농촌지역에서 가장 많아 농촌지역의 노인인구가 폭염에 더 많이 노출된 것으로 나타났다.

대기오염 및 폭염 등 기후변화에 영향을 미치는 도시환경 요소와 관련한 연구는 꾸준히 진행되고 있으며, 주로 회귀모형을 통한 분석이 진행되었다. 또한 분석을 통해 토지 피복, 토지이용, 녹지면적 등의 요소가 기온변화에 영향을 많이 미치는 것으로 나타났고, 녹지 식생을 통한 저감효과 분석도 같이 이루어지고 있다.

3. 인공지능 기반 기후분석 연구

인공지능을 활용한 기후분석 연구는 최근 미세먼지 농도 및 도시열섬현상 예측 등 다양하게 진행되고 있다(조경우, 정용진, 이종성, 오창현 2019; 정용진, 조경우, 이종성, 오창현 2019; 오진우 2020). 또한, 단순 기후 예측과 함께 기존 통계모형을 통한 기후분석과의 비교와 기온변화 요소 간의 관계성 분석 등도 함께 진행되고 있다(이슬기, 정성관, 이우성, 박경훈 2011; 조수현, 정미리, 이진향, 오일석 외 2020).

이슬기, 정성관, 이우성, 박경훈(2011)은 창원지역의 지표온도와 관측온도를 상관분석 및 회귀분석을 통해 평균기온과 공업지역, 녹지지역 등 8개의 토지 이용 유형의 관계성 분석하고, 신경망 및 회귀모형을 활용해 토지이용에 따른 도시 기온 예측모형을 구축하였다. 분석 결과, 공업지역의 평균기온이 가장 높게 나타났고, 녹지지역의 평균기온이 가장 낮게 나타났다. 또한 신경망 모형을 통한 예측에서 84.7%의 설명력을 보이며, 81.6%의 설명력을 보인 회귀모형보다 높은 정확도를 보였다.

조경우, 정용진, 이종성, 오창현(2019)은 천안지역의 10년간 기상 데이터 및 대기오염 물질 데이터를 활용해 LSTM 모델을 구축하고, PM10 미세먼지 농도를 예측한 후 4단계의 미세먼지 대기환경지수(Air Quality Index: AQI)와 정확도를 비교하였다. 분석 결과, AQI 예측 정확도는 87%의 성능을 보였고, 저농도 미세먼지와 비교하여 고농도 미세먼지에 대한 예측 정확도가 상대적으로 낮았으나, 실제 값과 근사한 수치로 예측하는 것을 확인하였다.

정용진, 조경우, 이종성, 오창현(2019)은 10년간의 천안 지역 기상 데이터 및 대기오염 물질 데이터를 활용해 GRU 모델을 구축하고, PM10 미세먼지 농도를 예측한 후 4단계의 미세먼지 AQI와 정확도를 비교

하였다. 분석 결과, AQI 예측 정확도가 87.56%로 나타났다. 그러나, 고농도의 미세먼지에서 예측이 저조하였고, 이는 표본의 수가 많지 않은 것으로 판단하였다.

조수현, 정미리, 이진향, 오일석 외(2020)는 임의로 선정한 17개 지역의 최대 풍향 및 평균 풍속 데이터와 미세먼지 데이터 간의 피어슨 상관관계 분석으로 입력변수를 선정하여, LSTM 모델을 통해 PM10 미세먼지 농도를 예측하였다. 분석을 통해 기상정보와 미세먼지의 양적 선형관계를 파악하였고, 풍향/풍속 데이터를 활용한 LSTM 모델에서 미세먼지 예측에 높은 성능을 확인하였다.

오진우(2020)는 서울 지역 54개의 자동기상측정망(AWS)을 기준으로 수집된 기후, 토지피복, 건축물 공간 데이터 등을 활용하여 도시열섬현상의 현황을 분석하고, 열섬현상에 영향을 주는 요인을 추출하여 시계열 기반 딥러닝을 통해 도시열섬현상을 예측하였다. 평균 도시열섬 강도가 최대, 최소, 중간값에 해당하는 3개 지역에 대한 시계열 예측 결과, 자기회귀누적이동평균(Autoregressive Integrated Moving Average: ARIMA)모형보다 딥러닝을 통한 예측의 정확도가 더 높게 분석되었다.

인공지능 기반의 연구는 단순 예측과 함께 각 변수 간의 관계성 분석 및 기존 모형과의 예측 정확도 비교 분석이 함께 진행되고 있었다. 딥러닝 기반의 RNN 모델은 시간의 흐름에 따라 변화하는 시계열 데이터의 분석에 적합하며 특히, LSTM과 GRU 모델의 경우 앞선 레이어의 정보를 계속해서 기억하기 때문에 정확한 예측이 가능하다. 그러나 현재 인공지능 기반 기후 변화 예측 연구는 미세먼지 농도 예측에 집중되어 있어 평균기온 등 다양한 기후변화 예측을 위한 연구가 수행될 필요가 있다.

4. 소결 및 연구의 차별성

선행연구 고찰결과, 평균기온 및 극한일수 관련 연구는 RCP 시나리오 기반의 상관관계 분석이 주를 이루고 있으나, 미래 시나리오 자료의 불확실성이 존재하여 이를 보완하기 위한 다양한 방법의 연구가 수행될 필요가 있다. 또한 회귀모형을 통해 기후변화에 영향을 미치는 도시환경요소를 추출하고 관계성을 분석하고 있으며, 주로 토지피복과 녹지면적 등이 기후변화에 영향을 주는 것으로 나타나, 이를 기반으로 다양한 도시공간 특성 기반의 독립변수가 고려되어야 한다. 인공지능 기반 기후분석 연구를 살펴보면 단순 예측과 변수 간의 관계성 분석이 활발히 진행 중이며, RNN 모델을 통한 시계열 분석이 높은 정확도를 보이지만 미세먼지 농도 예측에 집중되어 있어 다른 기후요소에 대한 예측이 필요할 것으로 사료된다.

이에 본 논문은 상관분석 및 회귀분석을 통해 도시공간 및 환경 변수들과 대기오염, 기상 데이터에서 평균기온에 영향을 미치는 변수를 추출하고, 연평균기온 상승률이 높은 원주시를 대상으로 딥러닝 기반의 LSTM과 GRU를 통해 평균기온을 예측하고자 한다. 이를 통해 해당지역의 평균기온 상승에 영향을 미치는 요소를 확인하여 향후 적절한 기후변화 적응방안을 마련할 수 있을 것으로 기대된다.

III. 연구의 범위 및 방법

1. 연구의 범위

본 논문에서는 연평균기온 상승률과 변화량이 높은 강원 원주시를 공간적 범위로 설정하였으며, 시계열 예측을 위해 2001년부터 2018년까지 1시간 단위의 데이터를 확보하였고, 연속성을 위해 2월 29일의 데

이터(2004년, 2008년, 2012년, 2016년)는 제외하였다.

분석에 앞서 선행연구 고찰을 통해 평균기온에 영향을 주는 변수와 원주시의 토지이용 및 기본 통계, 대기오염 및 기상 데이터로부터 26개의 독립변수와 1개의 종속변수를 수집하였다(<Table 1> 참조). 원주시는 3개(문막읍, 중앙동, 명륜동)의 도시 대기 측정소가 설치되어 있지만, 문막읍 측정소는 최근에 설치되어 2018년 이전 데이터가 존재하지 않아 제외하였다. 또한 중앙동 측정소는 SO₂, PM₁₀ 등 도시 대기는 측정하지만, 기온과 풍속 등 종관기상관측(ASOS)은

Table 1_ Total_Variable

	Variables	Unit
Independent Variable	Population	person
	Population_density	person/km ²
	Ruminants	head
	forest_land	ha
	Residential_zone	km ²
	Commercial_zone	km ²
	Industrial_zone	km ²
	Green_zone	km ²
	Natural_parks	1,000km ²
	Urban_parks	1,000km ²
	Rivers_length	km
	Rivers_Improvement_rate	%
	Roads_length	km
	Roads_paved_rate	%
	Road_facilities	m ²
	Vehicles	car
	SO ₂	ppm
	CO	ppm
	O ₃	ppm
	NO ₂	ppm
	PM ₁₀	µg/m ³
	Precipitation	mm
	Wind_speed	m/s
Wind_direction	16th direction	
Humidity	%	
Pressure	hPa	
Dependent Variable	Temperature	°C

Figure 1 _Locations of National Air Pollution Monitoring Stations in Wonju (Provided by Airkorea)

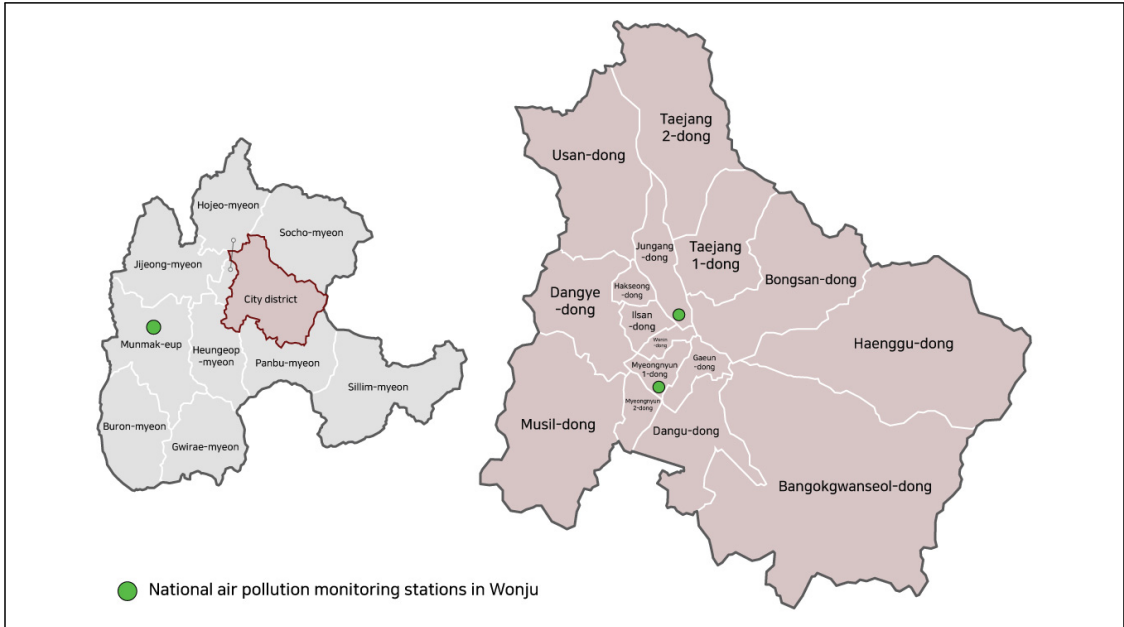


Table 2 _Dataset Example

Time (YYYY-MM-DD-hh)	...	Green_zone (km ²)	Natural_parks (1,000km ²)	Urban_parks (1,000km ²)	...	SO ₂ (ppm)	CO (ppm)	O ₃ (ppm)	...
2001-01-01-01	...	45.25	107720	5214.4	...	0.0115	1.6500	0.0085	...
2009-06-15-12	...	54.26	181572	6724.0	...	0.0030	0.2500	0.0455	...
2018-12-31-24	...	53.13	105506	8228.0	...	0.0067	1.3000	0.0063	...

Source: Wonju Statistical Information; Air Korea; Korea Meteorological Administration.

Table 3 _Missing Value Interpolation

Date	CO	CO_Interpolation
2001010710	0.8	0.8
2001010711	-999	0.9
2001010712	-999	0.9
2001010713	0.9	0.9
Average_Day	0.9	0.9

Table 4 _Missing Value Statement

Total Value	Normal Value	Missing_Value_Interporation	
		mean_weekly	mean_daily
	145,785	5,367	6,528

하지 않아 최종적으로 명륜동 측정소에서 수집된 데이터를 활용하였다(<Figure 1>, <Table 2> 참조). 기기 고장 등의 이유로 결측이 발생한 시간은 일간 및 주간 평균값으로 보간하였다(<Table 3, 4> 참조).

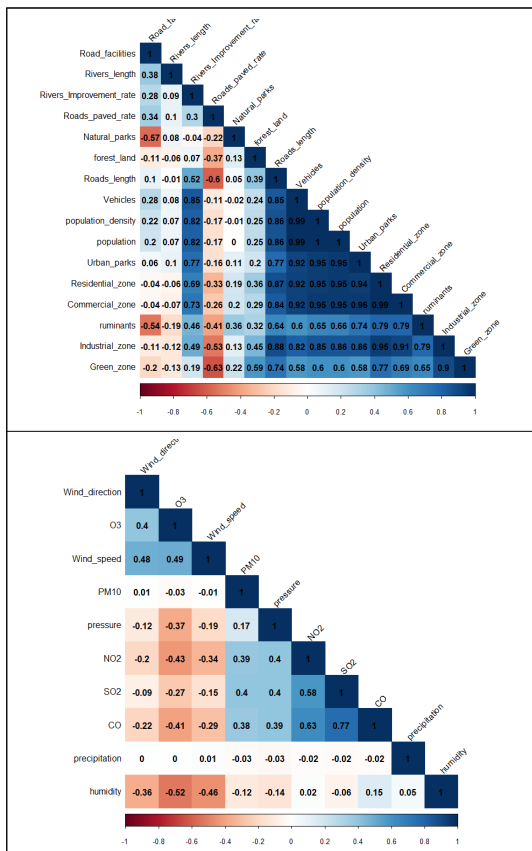
2. 연구의 방법

1) 상관분석 및 회귀분석

본 논문에서는 평균기온 상승을 예측하기 위해 평균 기온 상승에 영향을 미치는 변수를 추출하였다. 먼저

Stata와 R 통계분석 프로그램을 이용하여 앞서 수집한 독립변수 간 상관분석에 따라 상관계수가 0.4 이상인 강한 상관관계를 보이는 변수들을 확인하였다(Rea and Parker 2014). 다음으로 회귀분석을 통해 분산팽창계수(Variance Inflation Factor: VIF)가 10 이상인 변수는 다중공선성 문제를 고려하여 제외하고 첫 번째 모형을 만들었다(Hair, Anderson, Tatham and Black 1995). 다음으로 회귀분석 결과를 토대로 유의하지 않은 변수를 첫 번째 모형에서 제거하여 두 번째 모형을 만들었고, 추가로 보다 엄격한 VIF 기준을 적용하여 VIF가 4 이상인 변수를 제거하여 세 번째 모형을 만들었다(Ringle, Christian, Wende, Sven, and Becker 2015)(<Figure 2> 참조).

Figure 2_Correlation Analysis



2) 딥러닝 기반의 LSTM과 GRU 분석

본 논문에서는 상관분석 및 회귀분석을 통해 추출된 변수에 따라 세 가지 모형을 구성하여 딥러닝 기반의 RNN 모델의 변형인 LSTM과 GRU를 통해 평균기온을 예측하였다. RNN은 자연어 처리나 음성 인식, 날씨 추가처럼 시간의 흐름에 따라 변화하는 순서가 있는 데이터를 처리하는 데 적합한 딥러닝 기법으로 (Karpathy, Johnson and Fei-Fei 2015), 데이터 일부가 아닌 전체를 이해하는 방식으로 처리된다. 딥러닝의 기본 구조인 인공신경망(Artificial Neural Network: ANN), 심층신경망(Deep Neural Network: DNN)과 달리 RNN은 되먹임 구조를 가져 현재의 상태가 다음 상태에 영향을 미친다(<Figure 3> 참조).

RNN은 되먹임 구조를 통해 이전 레이어의 출력(hidden state, $ht-1$)을 현재 레이어에 입력(xt)으로 다시 받아 활성화 함수와 입력력 가중치(U, W)를 반영하여 값을 갱신해 출력하고(ht), 최종 결과 값(yt)을 출력한다(<Figure 4> 참조).

Figure 3_DNN & RNN Architecture

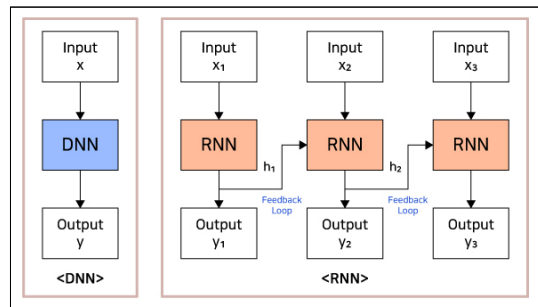
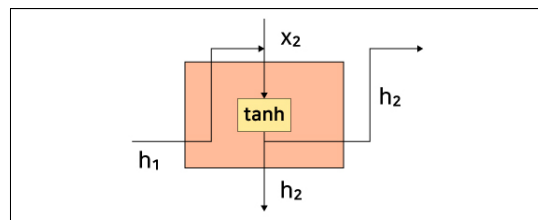


Figure 4_RNN Network



이때, 활성화 함수는 기울기 소멸(Vanishing Gradient)로 인한 학습 오류를 방지하기 위해 주로 비선형 함수인 쌍곡탄젠트(tanh)와 정류 선형 유닛(Rectified Linear Unit: ReLu)이 사용된다(Hochreiter 1998)(<식 1> 참조).

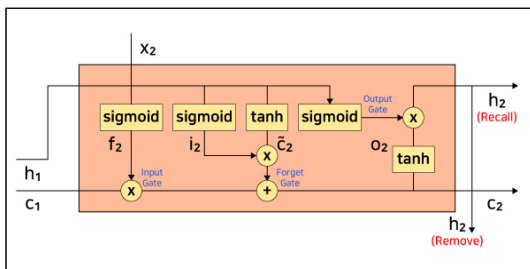
$$h_t = \tanh(Ux_t + Wh_{t-1}) \quad \text{<식 1>}$$

$$\begin{pmatrix} h_t = \text{이전 레이어의 출력} \\ U = \text{입력 가중치} \\ W = \text{출력 가중치} \end{pmatrix}$$

단순한 형태의 RNN은 짧은 시퀀스 데이터에선 효과를 보이나, 시퀀스가 길어질수록 이전 데이터와의 연관성이 떨어져 정확한 예측이 어려워지는 장기 의존성 문제를 가진다(Bengio, Simard and Frasconi 1994). 이러한 장기 의존성 문제를 해결하기 위해 단기 기억인 이전 레이어의 출력(h_t)과 함께 학습보다 정보 기억에 중점을 두며 입력층 사이에 공유되는 장기 기억 셀 상태(c_t)를 추가해 정보를 기억하는 형태인 LSTM 모델이 등장하였다(Hochreiter and Schmidhuber 1997) (<Figure 5> 참조).

LSTM에는 장기 의존성 문제 해결을 위해 3개의 게이트(Input Gate, Forget Gate, Output Gate)가 존재하여 이전 레이어의 정보를 보존하거나 버리는 선택을 하고, 다음 레이어에 넘겨줄 정보와 출력할 값을 결정

Figure 5 _ LSTM Network



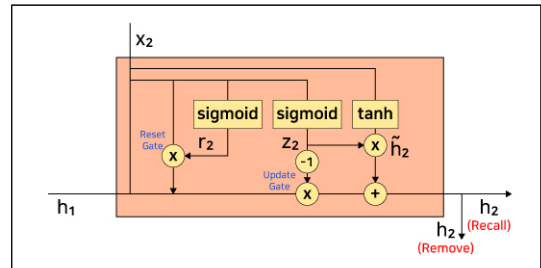
함으로써 과거 데이터를 통해 미래 데이터를 예측한다(<식 2> 참조).

$$\begin{aligned} i_t &= \text{sigmoid}(x_t U^i + h_{t-1} W^i) \\ f_t &= \text{sigmoid}(x_t U^f + h_{t-1} W^f) \\ o_t &= \text{sigmoid}(x_t U^o + h_{t-1} W^o) \\ \tilde{c}_t &= \tanh(x_t U^c + h_{t-1} W^c) \\ c_t &= f_t \times c_{t-1} + i_t \times \tilde{c}_t \\ h_t &= \tanh(c_t) \times o_t \end{aligned} \quad \text{<식 2>}$$

$$\begin{pmatrix} i_t = \text{Input Gate의 출력} \\ f_t = \text{Forget Gate의 출력} \\ o_t = \text{Output Gate의 출력} \\ \tilde{c}_t = c_t \text{ 이전의 cell state} \end{pmatrix}$$

이후 장기 의존성 문제를 해결하면서 LSTM의 구조를 간소화하여 성능은 비슷하지만 처리속도가 빠른 GRU 모델이 제안되었다(Cho, Van Merriënboer, Gulcehre and Bahdanau et al. 2014)(<Figure 6> 참조).

Figure 6 _ GRU Network



GRU는 LSTM과 달리 단기 기억(h_t)과 장기 기억(c_t)을 나누지 않고 한 번에 처리하며(h_t), 2개의 게이트(Reset Gate, Update Gate)가 존재하여 이전 레이어의 정보를 보존하거나 버리는 선택을 하고, 다음 레이어에 넘겨줄 정보와 출력할 값을 결정한다(<식 3> 참조).

$$\begin{aligned} z_t &= \text{sigmoid}(x_t U^z + h_{t-1} W^z) \\ r_t &= \text{sigmoid}(x_t U^r + h_{t-1} W^r) \\ \tilde{h}_t &= \tanh(x_t U^h + (h_{t-1} \times r_t) W^h) \\ h_t &= (1 - z_t) \times h_{t-1} + z_t \times \tilde{h}_t \end{aligned} \quad \text{<식 3>}$$

$$\begin{pmatrix} z_t = \text{Update Gate의 출력} \\ r_t = \text{Reset Gate의 출력} \\ \tilde{h}_t = h_t \text{이전의 cell state} \end{pmatrix}$$

3) 분석 환경 및 모델 구축

본 연구는 하드웨어 RAM 16GB, Nvidia GeForce GTX 1650, 소프트웨어 Python 3.7 및 Tensorflow 2.30 환경에서 이루어졌다. 분석에 앞서 2001년부터 2018년까지 1시간 단위 데이터를 구축하였고, 모델 경량화를 위해 평균기온 상승 관련 극한지수인 여름일수의 기준에 따라 최고기온이 25°C 이상인 4월부터 10월까지의 데이터만 추출하여 모델을 구축하였다.

전체 데이터는 모델의 학습 및 평가를 위해 시간 순서에 따라 6:2:2=학습:검증:테스트 데이터로 분할하였으며, 데이터의 60%로 모델을 학습시키고, 학습시킨 모델에 대해 다시 20%의 데이터로 검증을 한 후 최종 성능 평가를 위해 나머지 20%의 데이터로 테스트하였다. 또한, 학습 시간을 줄이기 위하여 전체 데이터 세트를 여러 그룹으로 나눈 후 그룹의 크기 (Batch Size)를 64로 설정하였다. 다음으로, 전체 데이터 반복 학습 횟수(Epoch)는 100으로 설정하였고, 학습의 과적합을 방지하기 위해 연속 5회 이상 손실 값(loss)이 감소하지 않는 경우 학습의 조기 종료를 명령하였다. 마지막으로 모델의 평가를 위한 손실 값(loss)을 ‘평균제곱오차(Mean Squared Error: MSE)’로 설정하고, ‘아담(Adam)’ 함수를 사용하여 최적화를 진행한 후 평가지표로 ‘평균절대오차(Mean Absolute Error: MAE)’를 사용하였다.

IV. 분석 결과

1. 상관분석 및 회귀분석을 통한 변수 추출

앞서 추출한 26개의 독립변수 간 상관분석을 통해 상관관계가 높은 변수들이 확인되었고, 회귀분석을 통해 VIF가 10 이상인 8개의 변수(등록 인구수, 인구밀도, 반추동물 사육 수, 주거지역 면적, 상업지역 면적, 공업지역 면적, 도로 연장, 자동차 등록대수)를 제거하여 18개의 독립변수로 구성된 첫 번째 모형을 만들었다. 다음으로 회귀분석 결과를 토대로 유의확률(p-value) 2.2e-16을 기준으로 통계적으로 유의하지 않은 2개의 변수(자연공원 면적, 도로시설물 면적)를 제거하여 두 번째 모형을 만들었고, 결정계수(R-square) 0.7754로 높은 설명력을 보였으며, p-value 값도 유의한 값을 가졌다. 마지막으로 예측 정확도를 높이기 위해 VIF가 4 이상인 4개의 변수(녹지지역 면적, 도시공원 면적, 하천개수율, 일산화탄소 농도)를 제거하여 세 번째 모형을 만들었다(<Table 5> 참조).

Table 5_ Analysis Model

Variables	Unit	Model 1	Model 2	Model 3
Population	person	×	×	×
Population_density	person/km ²	×	×	×
Ruminants	head	×	×	×
Forest_land	ha	○	○	○
Residential_zone	km ²	×	×	×
Commercial_zone	km ²	×	×	×
Industrial_zone	km ²	×	×	×
Green_zone	km ²	○	○	×
Natural_parks	1,000km ²	○	×	×
Urban_parks	1,000km ²	○	○	×
Rivers_length	km	○	○	○
Rivers_Improvement_rate	%	○	○	×
Roads_length	km	×	×	×
Roads_paved_rate	%	○	○	○
Road_facilities	m ²	○	×	×
Vehicles	car	×	×	×
O ₃	ppm	○	○	○

(continued)

Table 5_Analysis Model (continued)

Variables	Unit	Model 1	Model 2	Model 3
NO ₂	ppm	○	○	○
PM10	μg/m ³	○	○	○
Precipitation	mm	○	○	○
Wind_speed	m/s	○	○	○
Wind_direction	16th direction	○	○	○
Humidity	%	○	○	○
Pressure	hPa	○	○	○
Total		18	16	12

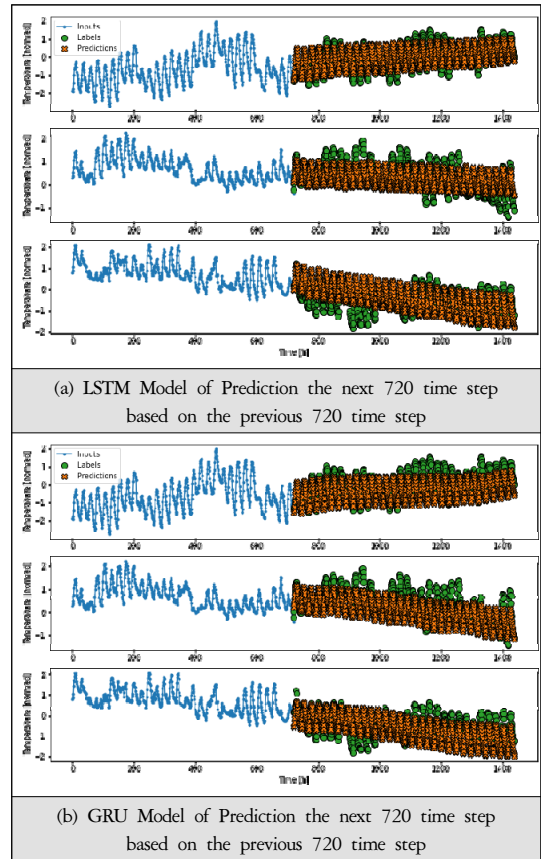
2. 딥러닝 기반 평균기온 예측

학습에 앞서 데이터의 단위를 맞추기 위해 평균 및 표준편차를 통해 데이터 정규화를 진행하였고, 기온은 +2.76°C~−3.69°C로 정규화되었다. 다음으로 평균기온 예측을 위해 앞선 30일 데이터를 기준으로 다음 30일을 예측하는 모델을 구성하여 딥러닝 기반의 LSTM과 GRU를 이용한 학습을 진행하였다.

1) 분석 1 (독립변수 - 18개)

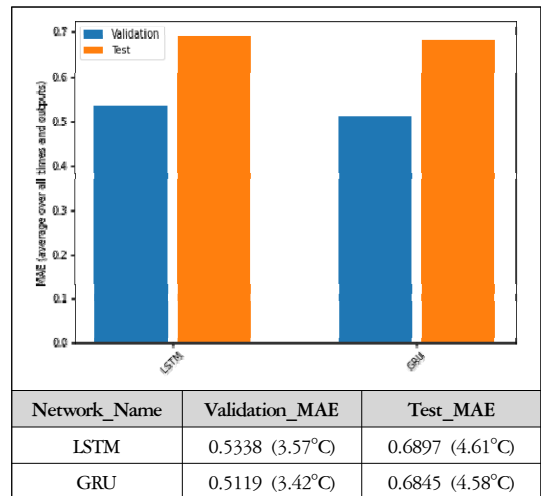
첫 번째 회귀모형에서 추출된 18개의 변수를 통해 평균기온 예측을 진행한 결과 검증 데이터에 대한 LSTM의 MAE는 0.5338 (3.57°C), GRU의 MAE는 0.6845(4.86°C)로 나타났다. 또한 학습의 성능평가를 위한 테스트 데이터 분석 결과 LSTM의 MAE는 0.6897(4.61°C), GRU의 MAE는 0.6845(4.58°C)로 나타나 전체적으로 MAE가 높게 나타났고, 검증 데이터와 테스트 데이터 간의 차이가 발생하였다(<Figure 7, 8> 참조). 따라서 MAE를 감소시키고, 검증 데이터와 테스트 데이터 간의 차이를 줄이기 위하여 변수 구성을 달리한 두 번째 모형을 만들었다.

Figure 7_Learning Result - 1st Model



Note: Inputs : Input data, Labels : Actual data, Predictions : Prediction data.

Figure 8_Analysis Result - 1st Model



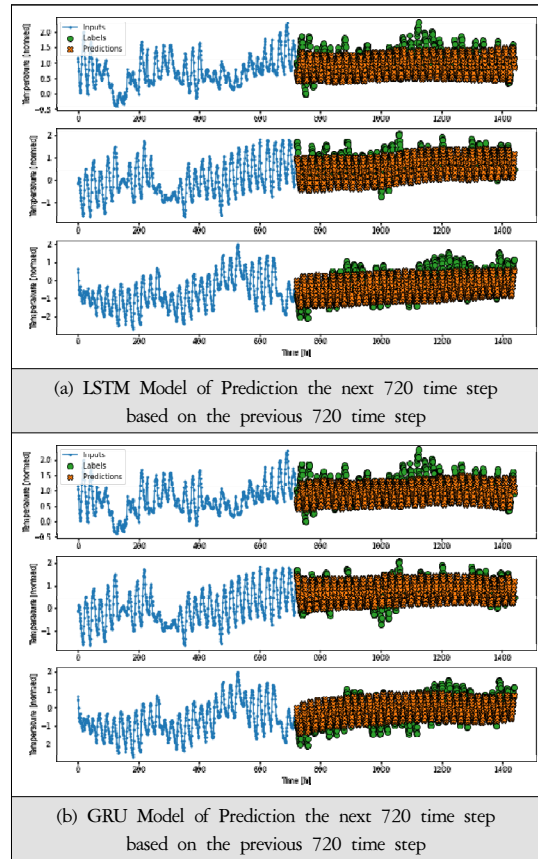
2) 분석 2 (독립변수 - 16개)

두 번째 회귀모형에서 추출된 18개의 변수를 통해 평균기온 예측을 진행한 결과 검증 데이터에 대한 LSTM의 MAE는 0.4207 (2.81°C), GRU의 MAE는 0.4256(2.85°C)으로 나타났다. 또한 학습의 성능평가를 위한 테스트 데이터 분석 결과 LSTM의 MAE는 0.5678(3.80°C), GRU의 MAE는 0.5933(3.97°C)으로 나타나 첫 번째 모형과 비교하여 MAE가 낮아졌으나 여전히 검증 데이터와 테스트 데이터 간의 차이가 발생하였다(<Figure 9, 10> 참조). 따라서 검증 데이터와 테스트 데이터 간의 차이를 줄이기 위하여 다시 변수 구성을 달리한 세 번째 모형을 만들었다.

3) 분석 3 (독립변수 - 12개)

세 번째 회귀모형에서 추출된 18개의 변수를 통해 평균기온 예측을 진행한 결과 검증 데이터에 대한 LSTM의 MAE는 0.4335 (2.90°C), GRU의 MAE는 0.4442(2.97°C)로 나타났다. 또한 학습의 성능평가를 위한 테스트 데이터 분석 결과 LSTM의 MAE는 0.4399(2.94°C), GRU의 MAE는 0.4444(2.97°C)로 나타나 두 번째 모형과 비교하여 검증 데이터의 MAE가 높아졌으나 차이가 크지 않았고, 검증 데이터와 테스트 데이터 간의 차이가 거의 없어 모델 구축이 잘 되었다. 특히 MAE에 따른 예측값과 실제값의 차이가 2.90~2.97°C로 많이 감소하면서 유의미한 결과를 도출하였다(<Figure 11, 12> 참조).

Figure 9 _ Learning Result - 2nd Model



Note: Inputs : Input data, Labels : Actual data,
Predictions : Prediction data.

Figure 10 _ Analysis Result - 2nd Model

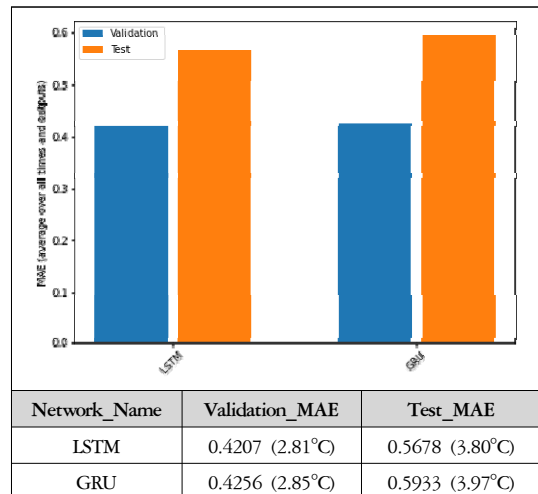
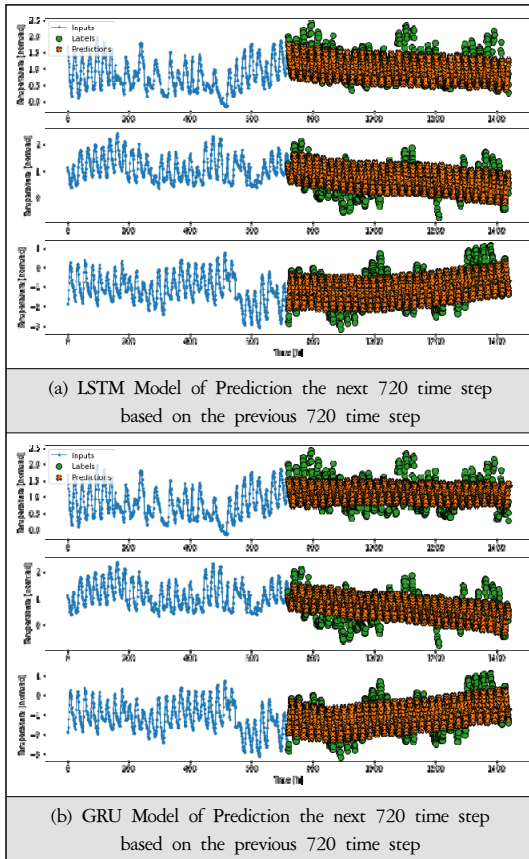
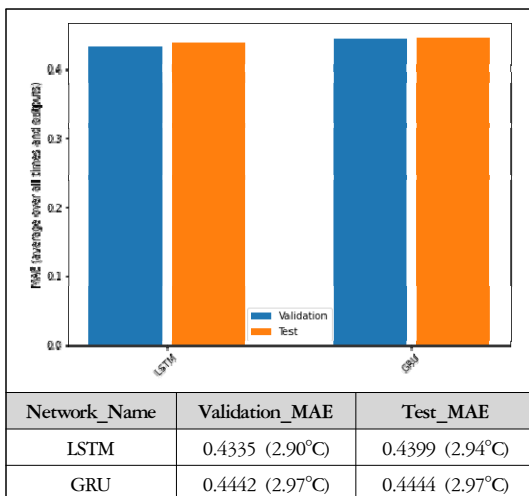


Figure 11_Learning Result – 3rd Model



Note: Inputs : Input data, Labels : Actual data,
Predictions : Prediction data.

Figure 12_Analysis Result – 3rd Model



V. 결론 및 토의

1. 결론 및 시사점

본 논문에서는 연평균기온 상승에 따른 기후문제 사전 대응을 위한 데이터확보 방안으로 딥러닝을 통해 연평균기온 상승률과 변화량이 큰 강원 원주시를 대상으로 평균기온 상승을 예측하였다. 먼저, 상관분석 및 회귀분석을 통해 도시 공간 및 도시 변수들과 대기 오염, 기상 데이터에서 평균기온 상승에 영향을 미치는 변수를 추출했다. 다음으로 추출된 변수를 통해 딥러닝 기반의 LSTM과 GRU 모델을 통해 평균기온 상승을 예측하는 모델을 만들어 일반 선형추세 모형과 비교하였다. 분석을 위해 2001년부터 2018년까지의 통계데이터를 시간 단위로 수집하고, 모델 경량화를 위해 최고기온이 25°C 이상인 날이 존재하는 4월~10월까지로 데이터를 추출한 후 앞선 30일로 다음 30일을 예측하는 모델의 평가를 진행하였다.

분석 결과 상관분석 및 회귀분석을 통해 $R^2 = 0.7753$ 의 설명력을 보이며 18개, 16개, 12개의 변수를 추출해 학습을 진행하였고, 변수가 12개인 모델에서 MAE가 가장 낮게 분석되었으며, 검증 데이터와 테스트 데이터 간의 MAE 차이가 거의 발생하지 않았다.

본 논문은 연평균기온 상승에 따른 기후문제의 사전 대응 마련을 위한 데이터확보 방안으로 통계분석 방법을 통해 평균기온 상승에 영향을 미치는 변수를 추출하고, 딥러닝을 활용해 평균기온 상승을 예측하였다는 점에서 의의가 있다. 또한, 평균기온 변화 추세 예측과 함께 회귀 및 상관분석을 통해 원주시의 평균기온 상승에 영향을 미치는 도시공간 요소를 추출하여, 획일화된 기후변화 적응방안이 아닌 지역별 영향 요소를 고려한 적절한 방안을 마련할 수 있을 것으로 기대된다.

2. 연구의 한계 및 향후 과제

본 논문에서 사용한 딥러닝 기반의 LSTM과 GRU 모델은 앞선 데이터를 기억하거나 제거하면서 다음 학습을 진행하기 때문에 시계열 예측에 효과적인 만큼 많은 데이터를 기억할 수 있는 주기억장치(RAM)가 필요하다. 그러나, 본 논문에서 사용한 주기억장치의 용량은 16GB로 장기간 분석을 진행하는 데 한계가 있기 때문에 분석 기간을 30일 단위로 설정하고, 한 달 기준 평균기온을 예측하여 전체기간의 평균기온 상승 추세를 명확하게 제시하지 못하였다. 또한, 1440개(24시간×30일×2개월) 데이터를 Batch Size 64로 나누었고, 학습 횟수(Epoch)도 100으로 설정되어 더 깊은 학습을 진행하지 못하면서 학습의 정확도를 높이기 위한 다양한 모델 환경 구축을 진행하지 못하였다.

향후 예측 정확도를 높이기 위해 주기억장치의 용량 확대에 따른 메모리의 추가적인 확보가 필요하며, 이를 통해 모델 레이어 구성, Batch Size, 학습 횟수(Epoch) 수치 등을 바꿔가며 다양한 모델 환경을 구축할 수 있을 것으로 사료된다. 또한 메모리의 한계를 보완하여, 한 달 기준 분석이 아닌 1년 단위 분석 및 전체 기간 분석을 통해 거시적인 관점에서 평균기온 상승에 대한 데이터를 구축하고, 이를 토대로 기후변화에 대한 적응대책으로서의 지역 및 도시계획적 대응방안이 도출될 수 있을 것으로 기대된다.

참고문헌 ●●●●●

- 고동원, 박승훈. 2019. 근린환경특성과 도시열섬현상과의 상호관계에 관한 연구. 한국도시설계학회지 도시설계 20권, 3호: 55-67.
Ko Dong-Won and Park Seung-Hoon. 2019. How the neighborhood environment characteristics affect the urban heat island effect in Seoul, Korea. *Urban Design* 20, no.3:

- 55-67.
- 기상청. 2018. 한반도 100년의 기후변화. 서울: 기상청. Korea Meteorological Administration. 2018. *100 Years of Climate Change on the Korean Peninsula*. Seoul: Korea Meteorological Administration.
- 기상청. 2020. 한국기후변화평가보고서. 서울: 기상청 Korea Meteorological Administration. 2020. *Korean Climate Change Assessment Report 2020*. Seoul: Korea Meteorological Administration.
- 기상청 날씨누리. <https://www.weather.go.kr/> KMA Weather. <https://www.weather.go.kr/>
- 원주통계정보. <https://stat.wonju.go.kr> Wonju Statistical Information. <https://stat.wonju.go.kr>
- 김장경, 정창삼, 권현한. 2019. 공공 빅데이터를 이용한 폭염 취약성 지수 개발에 관한 연구. 대한토목학회 정기학술대회, 10월 17일. 평창: 알펜시아 컨벤션센터. Kim Jang-Gyeong, Jeong Chang-Sam and Kwon Hyun-Han. 2019. A Study on the development heat wave vulnerability index using public big data. Korean Society of Civil Engineers 2019 Convention, October 17. Pyeongchang: Alpensia.
- 김진아, 김규량, 조창범, 김백조. 2015. 제주의 과거와 미래 폭염 및 열대야일수 분석. 한국기상학회 추계학술대회 논문집: 876-877. Kim Jin-Ah, Kim Kyu-Rang, Cho Changbum and Kim Baek-Jo. 2015. Analysis of past and future heat waves and tropical nights in Jeju. *Proceedings of the Autumn Meeting of Korean Meteorological Society*: 876-877.
- 김학열. 2003. 도시환경인자에 따른 대기온도 변화 연구. 서울시를 대상으로. 국토계획 38권, 5호: 259-269. Kim Hag Yeol. 2003. The impacts of urban environmental factors on air temperatures in Seoul. *Journal of Korea Planning Association* 38, no.5: 259-269.
- 박경훈, 송봉근, 박재은. 2016. 토지피복유형과 지형특성이 폭염일수에 미치는 영향 분석. 한국지리정보학회지 19권, 4호: 76-91. Park Kyung-hun, Song Bong-geun and Park Jae-eun. 2016. Analysis on the effects of land cover types and topographic features on heat wave days. *Journal of the Korean Association of Geographic Information Studies* 19, no.4: 76-91.
- 송봉근, 박경훈, 김경아, 김성현, 박건웅, 문한솔. 2020. GIS

- 기반 노인인구 분포지역의 공간적 특성과 폭염의 관계 분석: 창원시를 대상으로. *한국지리정보학회지* 23권, 3호: 68-84.
Song Bong-geun, Park Kyung-hun, Kim Gyeong-ah, Kim Seoung-hyeon, Park Geon-ung and Mun Han-sol. 2020. Analysis of relationship between the spatial characteristics of the elderly population distribution and heat wave based on GIS: Focused on Changwon City. *Journal of the Korean Association of Geographic Information Studies* 23, no.3: 68-84.
11. 에어코리아. <https://www.airkorea.or.kr/web>
Air Korea. <https://www.airkorea.or.kr/web>
 12. 오규식, 홍재주. 2005. 도시공간 구성요소와 도시열섬현상의 관련성 연구. *한국도시설계학회지 도시설계* 6권, 1호: 47-63.
Oh Kyu-Shik and Hong Jae-Joo. 2005. The relationship between urban spatial elements and the urban heat island effect. *Urban Design* 6, no.1: 47-63.
 13. 오진우. 2020. 도시환경특성에 따른 답러닝 기반 도시열섬현상 예측: 서울 지역을 중심으로. 석사학위논문, 경희대학교.
Oh Jinwoo. 2020. *Deep Learning Prediction of Urban Heat Island Effects in Urban Environments*. M.S. diss., Kyung Hee University.
 14. 유동수, 권오창, 김홍기. 2020. RCP 시나리오기반 평균기온, 적산온도 불확실성 보완 프로그램. *한국기후변화학회지* 11권, 2호: 113-122.
Yu Dong-Su, Kwon Oh-Chang and Kim Hong-Gie. 2020. A simple program improving uncertainly average temperature and growing degree days based on RCP scenario. *Journal of Climate Change Research* 11, no.2: 113-122.
 15. 윤진일. 2002. 우리나라의 기온상승-도시효과인가 지구온난화인가. *한국농림기상학회지* 4권, 1호: 58-63.
Yun Jinil. 2002. Urbanization effect on the observed warming in Korea during the recent half century. *Korean Journal of Agricultural and Forest Meteorology* 4, no.1: 58-63.
 16. 이슬기, 정성관, 이우성, 박경훈. 2011. 인공신경망을 이용한 도시기온 예측모형 구축. *국토계획* 46권, 1호: 129-142.
Lee Seul-Gi, Jung Sung-Gwan, Lee Woo-sung and Park Kyung-hun. 2011. A predictive model for urban temperature using the artificial neural network. *Journal of Korea Planning Association* 46, no.1: 129-142.
 17. 이승호, 허인혜. 2011. 한국의 도시화에 의한 극한기온의 변화. *대한지리학회지* 46권, 3호: 257-276.
Lee Seung-Ho and Heo In-Hye. 2011. The impacts of urbanization on changes of extreme events of air temperature in South Korea. *Journal of the Korean Geographical Society* 46, no.3: 257-276.
 18. 이재호, 백희정, 현유경, 조천호. 2011. 기상 관측 자료를 이용한 강릉의 기후변화 추세 분석. *한국기후변화학회지* 2권, 2호: 133-141.
Lee Jaeho, Baek Hee-Jeong, Hyun Yu-Kyung and Cho Chunho. 2011. Observation-based Analysis of Climate Change using Meteorological Data of Gangneung. *Climate Change Research* 2, no.2: 133-141.
 19. 장재연. 2009. 기후변화, 지구온난화가 인체 건강에 미치는 영향. *대우재단 지식의 지평* 6권: 159-175.
Jang Jae-yeon. 2009. Effects of global warming on human health the horizon of knowledge. *Daewoofound Orbis Sapientiae* 6: 159-175.
 20. 정용진, 조경우, 이종성, 오창현. 2019. GRU 를 이용한 미세먼지 (PM10) 농도 예측 모델. *한국정보통신학회 종합학술대회 논문집* 23권, 2호: 644-646.
Jung Yong-jin, Cho Kyoung-woo, Lee Jong-sung and Oh Chang-heon. 2019. Particulate Matter(PM10) concentration prediction model using GRU. *The Korea Institute of Information and Communication Engineering Proceedings* 23, no.2: 644-646.
 21. 조경우, 정용진, 이종성, 오창현. 2019. LSTM 을 이용한 PM10 미세먼지 농도 예측. *한국정보통신학회 종합학술대회 논문집* 23권, 2호: 632-634.
Cho Kyoung-woo, Jung Yong-jin, Lee Jong-sung and Oh Chang-heon. 2019. PM10 particulate matters concentration prediction using LSTM. *The Korea Institute of Information and Communication Engineering Proceedings* 23, no.2: 632-634.
 22. 조수현, 정미리, 이진향, 오일석, 한영태. 2020. 풍향풍속과 미세먼지의 상관관계 분석과 LSTM 을 이용한 미세먼지 예측. *한국정보과학회 학술발표논문집*: 1649-1651.
Jo Soohyun, Jeong Miri, Lee Jinhyang, Oh Ilseok and Han Yeongtae. 2020. Analysis of correlation of wind direction/speed and Particulate Matter(PM10) and prediction of Particulate Matter using LSTM. *The Korean Institute of Information Scientists and Engineers Proceedings*: 1649-1651.
 23. 차영화, 김학열, 허태영. 2009. 토지이용, 토지피복 특성이 서울시 도시기온에 미치는 영향. *서울도시연구* 10권, 4호: 107-120.
Cha Young Hwa, Kim Hag Yeol and Heo Tae Young. 2009. The effects of urban land use and land cover

- characteristics on air temperature in Seoul Metropolitan Area. *Seoul Studies* 10, no.4: 107-120.
24. Bengio, Y., Simard, P. and Frasconi, P. 1994. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks* 5, no.2: 157-166. <https://doi.org/10.1109/72.279181>
25. Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. and Bengio, Y. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078. <https://arxiv.org/abs/1406.1078v3>
26. Chung, J., Gulcehre, C., Cho, K. and Bengio, Y. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555. <https://arxiv.org/abs/1412.3555v1>
27. Hair, J. F., Anderson, R. E., Tatham, R. L. and Black, W. C. 1995. *Multivariate Data Analysis with Readings*. New Jersey: Prentice Hall.
28. Hochreiter, S. and Schmidhuber, J. 1997. Long short-term memory. *Neural Computation* 9, no.8: 1735-1780.
29. Hochreiter, S. 1998. Recurrent neural net learning and vanishing gradient. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 6, no.2: 107-116.
30. Karpathy, A., Johnson, J. and Fei-Fei, L. 2015. Visualizing and understanding recurrent networks. arXiv preprint arXiv:1506.02078. <https://arxiv.org/abs/1506.02078v2>
31. Rea, L. M. and Parker, R. A. 2014. *Designing and Conducting Survey Research: A comprehensive guide*. New Jersey: John Wiley & Sons.
32. Ringle, C. M., Wende, S. and Becker, J. M. 2015. SmartPLS 3. Bönningstedt: SmartPLS. Retrieved from <http://www.smartpls.com>

- 논문 접수일: 2021. 1. 14.
- 심사 시작일: 2021. 2. 3.
- 심사 완료일: 2021. 6. 18.

요약

주제어: 평균기온 예측, 딥러닝, LSTM, GRU, 원주시

지구온난화로 인한 기후변화로 연평균기온이 계속해서 상승하는 추세를 보이면서 온열질환 발생률과 사망자 수도 증가하고 있어 이를 위한 다양한 대안과 연구가 수행될 필요가 있다. 이에 본 논문에서는 연평균기온 상승률 및 변화량이 높은 원주시를 대상으로 통계분석을 통해 평균기온 상승 관련 변수를 추출하고, 추출된 변수를 토대로 딥러닝 기반의 LSTM과 GRU를 활용하여 평균기온을 예측하고자 한다. 선행연구 고찰을 토대로 수집한 26개의 변수에 대해 상관분석 및 회귀분석을 통해 3가지 모형을 추출하였고, 이를 바탕으로 LSTM과 GRU 분석을 진행하였다. 분석 결과, 변수가 12개인 세 번째 모형에서 테스트 데이

터 MSE가 LSTM - 0.4399(2.94°C), GRU - 0.4444(2.97°C)로 가장 낮게 나타났고, 검증 데이터와 테스트 데이터 간의 MAE 차이가 거의 발생하지 않았다. 본 논문은 연평균기온 상승 문제 적응을 위한 데이터 확보 방안으로 통계분석을 통해 변수를 추출하고, 딥러닝을 활용해 평균기온을 예측하였다는 점에서 의미가 있다. 또한, 평균기온 변화 추세 예측과 함께 원주시의 평균기온 상승에 영향을 미치는 도시공간 요소를 추출하여, 획일화된 기후변화 적응방안이 아닌 지역별 영향 요소를 고려한 적절한 방안을 마련할 수 있을 것으로 기대된다.