

JEONGHYEON KIM

Data Carpenter

Master Course of Geography
Dept. Geography
Kyung Hee University

Phone: 010-6417-3765

E-Mail: zmsdkdle@khu.ac.kr, kjh3765@Hotmail.com

Github: <https://github.com/henney3765>

INTRODUCTION



김정현 (JEONGHYEON KIM)

1995.08.21

경상남도 통영시 멘데산업길 11, 1307호

경희대학교 일반대학원 지리학과 석사 재학
(2021.08 ~)

경희대학교 이과대학 지리학과 학사
(2015.03 ~ 2021.08)

관심 연구

공간 통계, 데이터 과학, 예측 모델링

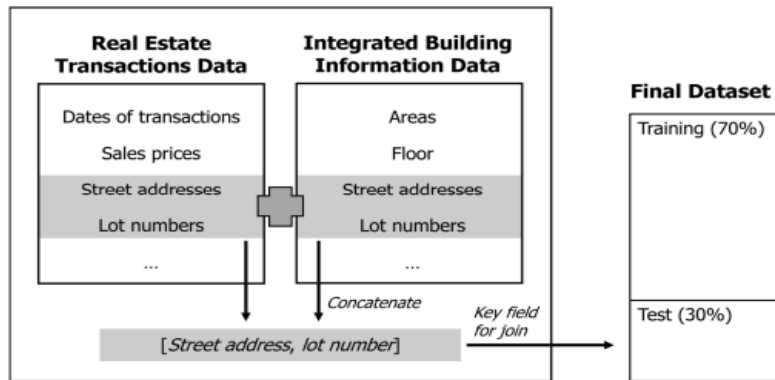
PROJECTS

1

A Comparative Study of Machine Learning and Spatial Interpolation Methods for Predicting House Prices

Data

GIS 통합건물정보 + 부동산 실거래가
건물면적, 계약연월일, 층, 위경도, 거래가격 등



Environment

R / gstat, rgdal, randomforest, keras, etc

Purpose

1. 전통적인 공간보간법을 머신러닝이 대체할 수 있는가?
2. 부동산 실거래가 예측에 영향을 미치는 변수 탐색

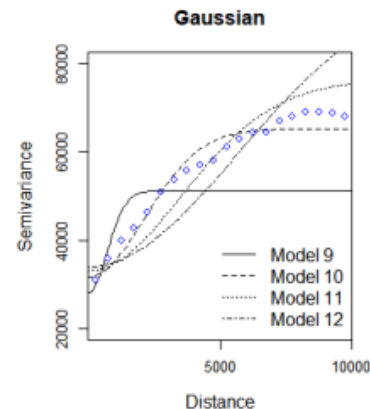
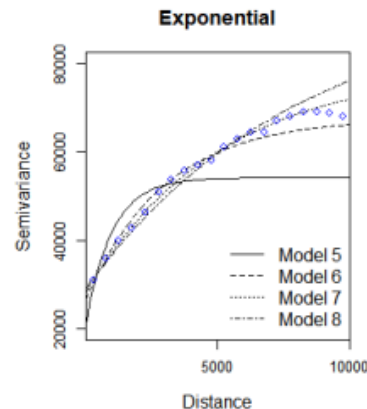
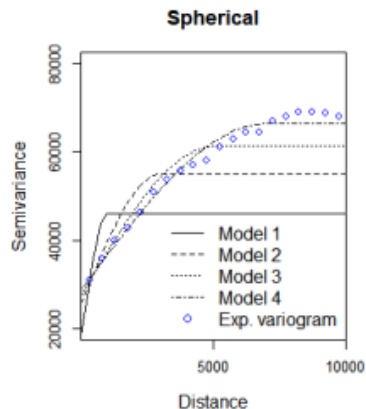
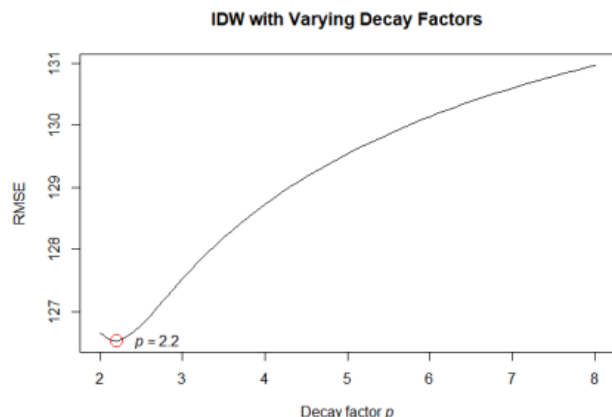
Methodology

4가지 알고리즘 사용 / RMSE로 모델 평가

역거리 가중법(Inverse Distance Weighting),
크리깅(Ordinary kriging), 랜덤 포레스트(Random
Forest), 심층신경망(Deep Neural Network)

1

A Comparative Study of Machine Learning and Spatial Interpolation Methods for Predicting House Prices



역거리 가중법, 크리깅의 최적 파라미터 탐색

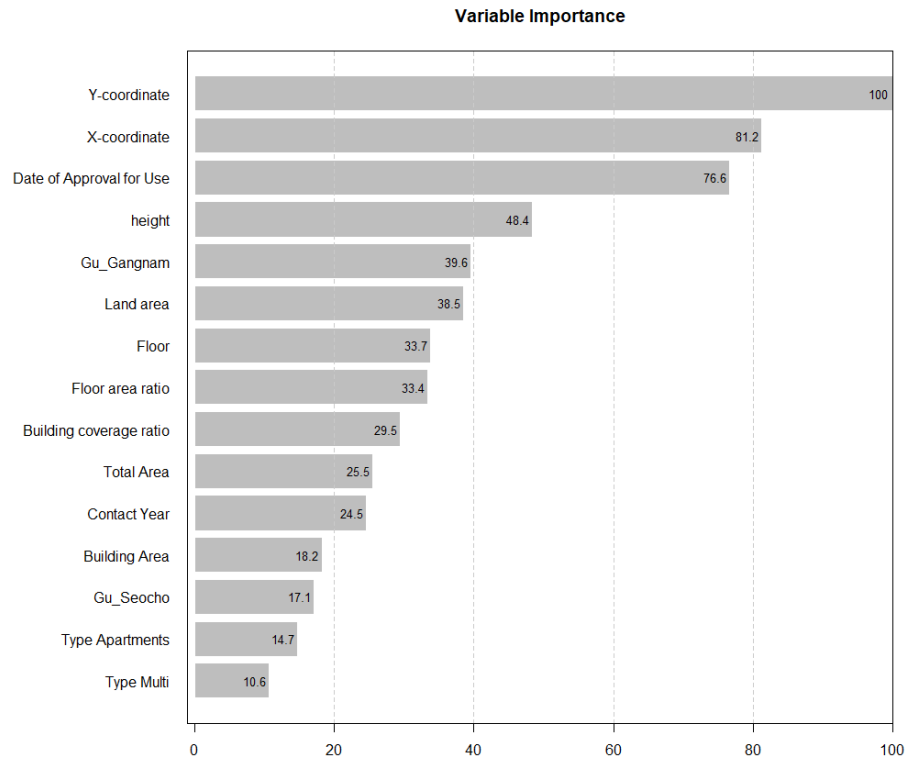
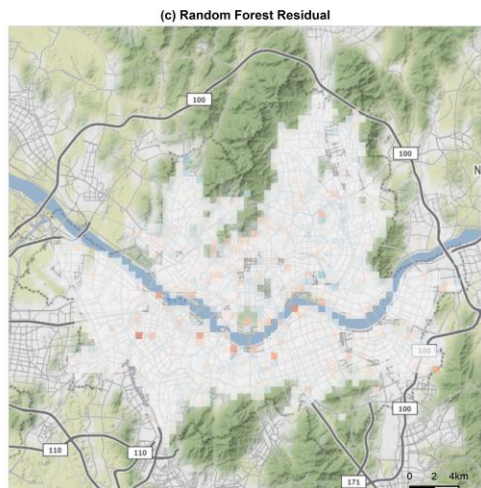
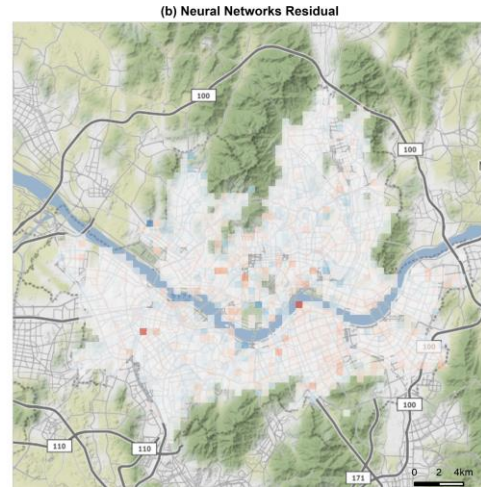
Method	MAE	RMSE	MAPE	MASE
Neural network	64.4615	102.0348	0.1260	0.2759
Random forest	48.1897	80.8789	0.0952	0.2063
IDW	78.4847	126.5288	0.1544	0.3360
Ordinary kriging	83.5145	131.0042	0.1655	0.3575

모델 별 성능 비교

랜덤 포레스트, 심층 신경망, 역거리가중법, 크리깅 순으로 성능이 좋을 수 있다.

1

A Comparative Study of Machine Learning and Spatial Interpolation Methods for Predicting House Prices



랜덤 포레스트 변수 중요도 시각화

랜덤 포레스트 변수 중요도를 활용하여
실거래가 예측에 영향을 미치는 변수 탐색하였다.

랜덤 포레스트, 인공 신경망 잔차의 지도 시각화

2

판교제로시티 데이터와 U-net 구조를 활용한 가로수 영역 추출

Data

판교제로시티 CCTV 2D 세그멘테이션

JSON 파일을 OpenCV를 통해 전처리하여
데이터 세트 구축



카카오 로드뷰 데이터

카카오 API를 활용하여 로드뷰 데이터 수집



Purpose

1. 기존 녹지피복률의 한계를 보완하는 녹시율을
딥러닝을 활용하여 측정
2. CCTV 데이터를 통해 학습한 모델이 로드뷰에도
적용될 수 있는가?

Methodology

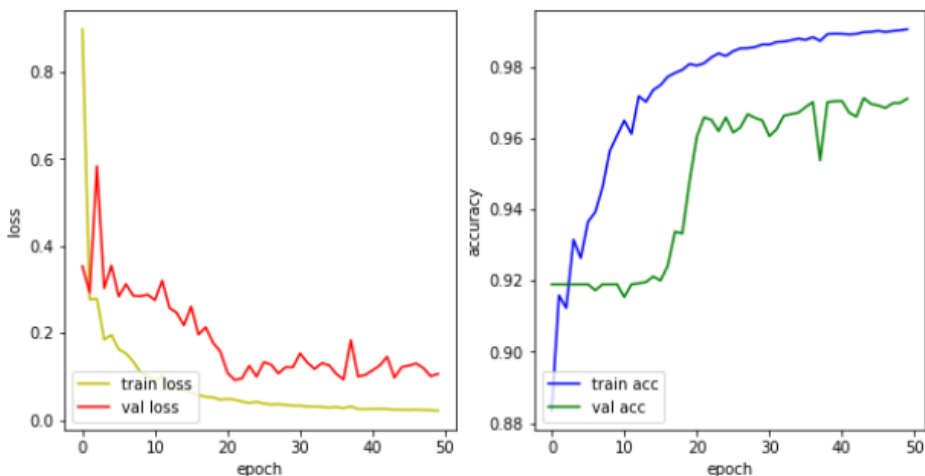
U-net 구조의 CNN 모델 활용

Environment

Python / OpenCV, keras, etc

2

판교제로시티 데이터와 U-net 구조를 활용한 가로수 영역 추출



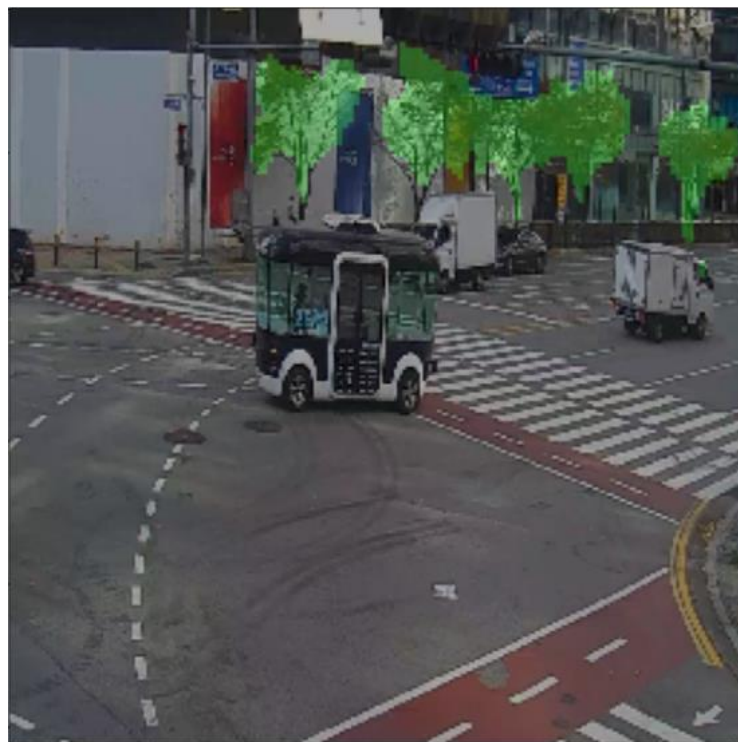
모델의 학습과정

배치사이즈는 8,

Optimizer는 rmsprop,

손실함수는 binary_crossentropy로

설정하여 총 50회 학습하였다.



결과 예시

검증 데이터셋에 대해서 0.96%의 정확도를 보임

2

판교제로시티 데이터와 U-net 구조를 활용한 가로수 영역 추출



로드뷰 데이터에 모델을 적용한 결과 예시

CCTV 데이터로 학습한 모델이 로드뷰에 적용되기에는 다소 정확도가 낮다고 판단된다.

보완

Cityscape 데이터셋과 같은 로드뷰 시점의 정밀한 학습 데이터를 활용한다면
성능의 향상이 가능할 것으로 생각한다.

3

Xgboost 를 활용한 지면/지상온도 예측

Data

국가 기상 위성센터에서 제공하는 천리안위성
2A호 기상위성 데이터 및 지상관측데이터
- 날씨 빅데이터 콘테스트 제공

Environment

Python, ArcGIS pro / sklearn, xgboost, etc

Preprocessing

1. 예측 변수가 되는 지면/지상 온도와 각 독립변수와의 **상관분석**을 통해 학습에 사용할 변수를 선택
2. 기상 데이터의 특성상 전체 데이터에 대해 상관계수가 낮더라도 **일별, 월별 상관계수는 높을 수 있기 때문에** 이를 고려

Purpose

1. 기상위성 자료를 활용하여 지면/지상 온도 산출기술 개발 및 개선

Methodology

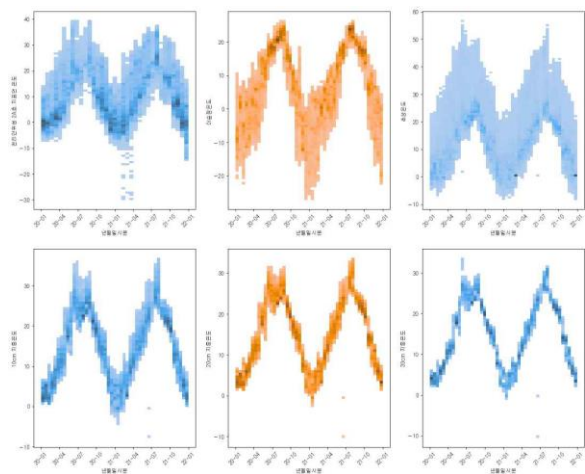
XGboost, Linear Interpolation, Inverse Distance Weighting / **RMSE로 모델 평가**



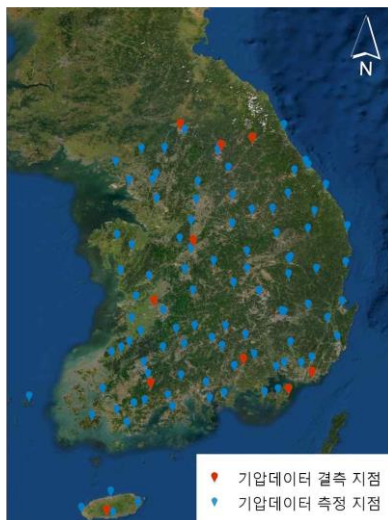
3 Xgboost 를 활용한 지면/지상온도 예측

Preprocessing (계속)

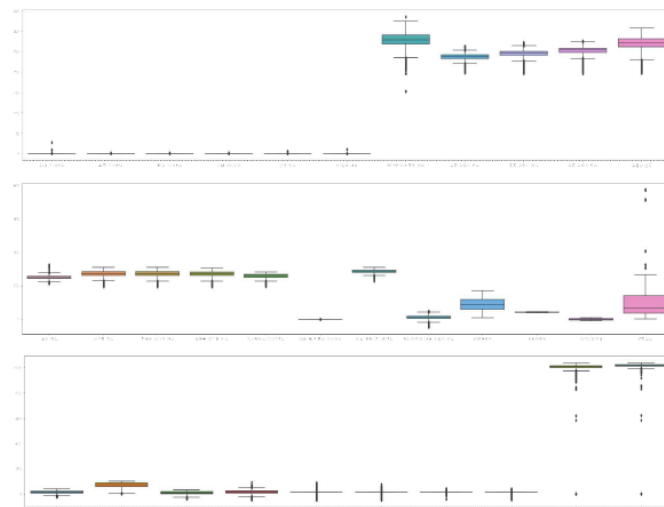
1. 예측 변수인 지면온도, 지상온도의 결측치를 제거
2. 시계열에 따라 일정 선형관계를 보이는 데이터는 **Linear Interpolation**을 활용하여 결측치 대체
3. 일부 자연현상의 경우 공간적 자기 상관성이 높다고 판단하여 **Inverse Distance Weighting**을 사용하여 결측치 대체
4. **Box-plot**을 활용한 이상치 제거 후, 모델링 작업을 수행



시계열 특성 확인



공간 보간



Box-plot 확인

3 Xgboost 를 활용한 지면/지상온도 예측

결과

각 모델에 대해서 RMSE를 통해 평가 이후, 날씨 빅데이터 콘테스트에 제출

train - RMSE	1.31185
test - RMSE	1.37100

train - RMSE	0.1590
test - RMSE	0.1828

보완

지상/지면 온도 RMSE

지면 온도와 큰 연관을 보이는 천리안 위성 2A호 지표면 온도, 지중 온도의 **결측치 처리 방식의 변경**을 통해 정확도 향상을 기대할 수 있다.

THANK YOU



Master Course of Geography
Dept. Geography
Kyung Hee University

Phone: 010-6417-3765

E-Mail: zmsdkdle@khu.ac.kr, kjh3765@Hotmail.com

Github: <https://github.com/henney3765>