



A Study on Rainfall Estimation of High Resolution Using Machine Learning

Min-Gyu Kim, Moo-Hun Lee*

WISE Institute, Hankuk University of Foreign Studies

ABSTRACT

Occurred rainfall as local torrential rainfalls among the climate change has a decisive effect on life. Rainfall is closely related to the living of people, its exact prediction is important to us. In this paper, we have improved the estimation accuracy of rainfall by comparing the applicable interpolation algorithms of meteorological field. The comparison algorithms are Kriging, MLP(Multi-Layer Perceptron), SVR(Support Vector Regression), Random Forest and Bagging. For the experiment, we used the AWS(Automatic Weather System) data which exists the event of rain on August 16th, 2015, 16:00~20:00. The AWS data is collected by a 1-minute cycle, the attributes related to rainfall include precipitation type, accumulated rainfall such as 10 minutes, 15 minutes, 30 minutes, and 1 day, etc. The dataset consists of training data and test data in 255 observation stations, Seoul. 230 stations are used as training data, the remaining stations are used as test data. Validation methods such as RMSE, R-square, Correlation Coefficient, CSI(Critical success index) and BIAS(frequency BIAS) were used to compare the algorithms. CSI and BIAS were measured for precision of rainfall and calculated using the rain contingency table. Our experimental results showed that bagging algorithm performed better than the others in estimating the rainfall information.

© 2017 KKITS All rights reserved

KEYWORDS: Machine learning, Spatial interpolation, Rainfall, Estimation, AWS

ARTICLE INFO: Received 20 February 2017, Revised 13 March 2017, Accepted 7 April 2017.

*Corresponding author is with the WISE Institute,
Hankuk University of Foreign Studies, 81, Oedae-ro
Mohyeon-myeon Cheoin-gu Yongin-si Gyeonggi-do,

17035, KOREA.
E-mail address: macbethe@gmail.com

1. 서론

우리나라는 봄, 여름, 가을, 겨울의 사계절이 뚜렷하면서 그에 대한 기상변화를 가지고 있다. 최근 들어 지구 온난화로 인해 발생하는 기후변화로 국지성 호우와 같은 위험기상이 있다. 국지성 호우로 인한 자연재해는 인명은 물론 막대한 재산상의 피해를 유발하고 있다. 이러한 국지성 호우로 인한 자연재해를 예방하기 위해서는 정확한 강우 예측이 매우 중요하다.

강우 예측을 위해 기상청은 다양한 기상변수를 활용하여 전지구예보모델, 지역예보모델, 국지모델과 같은 수치모델로부터 정보를 산출한다. 현재 전국을 5km 단위로 나눈 동네예보 서비스는 3일 이내의 날씨와 강수량, 강수확률 등 12개 요소를 산출하고 있다. 하지만, 스톱 규모(200m ~ 1km)에서 주로 발생하는 국지성 호우를 예측하는 것은 한계가 있다. 따라서, 이를 개선하기 위해 기상분야에서는 수치모델 결과의 후처리로 공간분석을 통한 다운스케일링(downscaling)을 수행하여 값을 추정하고 있다. 공간분석을 위한 알고리즘으로는 공간보간법이 주로 활용되고 있다. 다운스케일링을 위한 공간보간법은 위치좌표만을 가지고 고해상도의 강우량을 산출하기 때문에 강우량 추정에 필요한 지형적 특성과 같은 관련 요소를 고려하기 어렵다. 경험적 학습 방법인 기계학습은 축적해 놓은 자료를 기반으로 강우량을 추정함으로써 시공간적 특성을 반영할 수 있으며, 강우량과 상관성이 높은 다양한 관측 자료를 활용하여 학습할 수 있어 강우량 추정 정확도 향상을 기대할 수 있다.

따라서 본 논문에서는 고해상도의 강우량 추정을 위한 다양한 알고리즘을 적용하여 기상분야에서 활용되는 평가 방법을 통해 실험 및 검증을 하고, 적합한 알고리즘을 적용해 강우량 산출 정확도를 향상하고자 한다. 데이터는 SK Techx에서 구축

한 자동기상관측장비(AWS:Automatic Weather System)에서 수집되는 관측자료를 활용하고, 실험에는 기상분야에서 주로 활용되는 공간보간법인 크리깅(Kriging)과 기계학습에서는 MLP(Multi-Layer Perceptron), SVR(Support Vector Regression), 랜덤 포레스트(Random Forest), 배깅(Bagging)을 적용하여 비교분석 하였다.

본 논문의 구성은 다음과 같다. 2장에서는 기상분야에서 다운스케일링에 활용되는 보간 알고리즘을 소개하고, 3장에서는 실험에 대한 데이터 구성과 실험 방법을 설명한다. 4장에서는 실험에 대한 결과를 분석하고, 마지막 5장에서는 결론 및 향후 연구방향을 제시한다.

2. 기상분야 보간 알고리즘

2.1 공간보간법

공간보간법이라고 하면, 공간상에 관심 있는 지점에 대한 특정 값을 알기 위해서 기존에 알고 있는 주위 값을 가지고 추정하는 방법을 말한다. 여기에는 결정론적 방법과 통계적 방법으로 나누어지는데, 결정론적 방법의 대표적인 것으로는 IDW(Inverse Distance Weighting)가 있고 통계적 방법으로는 크리깅이 있다. IDW는 알고자하는 위치에서 역거리로 계산하여 가까이 있는 관측값에 더 큰 가중치를 주어 보간하는 방법이다. 크리깅은 통계 모형을 이용해 특정 반경 내에 값들을 선형 조합하여 주변 값들 간에 거리와 상관강도를 반영해 값을 추정하는 방법이다.

Jung[1]은 한강유역 내의 108개 자동기상관측소(AWS) 자료를 활용해 호우 및 태풍사상을 대상으로 IDW 방법을 이용해 유역강우량 산정시 고도 영향 분석에 활용하였다. Lee[2]은 강우산도를 추정하기 위하여 IDW, 크리깅을 비교하여 가장 적합한

방법을 비교분석 하였으며, Glenn[3]는 크리깅을 이용하여 위성영상 데이터로부터 토양의 수분을 추정하고자 하였다. Simbahan[4]는 크리깅을 활용하여 토양의 유기탄소를 추정하는데 정확도 개선을 위한 연구를 진행하였고, Chof[5]는 기상청 강우 관측소 413개소를 대상으로 공간보간법을 적용하여 그 정확도를 검증하였다. 또한 Park[6]은 기상관측소의 기온 및 강수 관측치를 이용한 공간적 분포도 작성을 위해 수치표고모델과 다변량 크리깅의 적용 가능성을 검토하였고, Park[7]은 지역적 강수량 차이 분석의 중요성을 강조하면서 크리깅을 활용해 지형사면방향이라는 변수 사용의 활용성 여부를 파악하고자 하였다.

IDW는 추정하고자 하는 위치를 기준으로 가까이 위치하는 값에 큰 가중치를 주고 멀어질수록 가중치를 적게 주는 보간 방법이다. 그러므로 거리가 가까운 값에 영향을 많이 받고 멀어질수록 영향을 적게 받게 되어 관측값이 적을 경우 지역적 특성이 잘 나타낸다. 크리깅은 통계적 방법을 기반으로 하여 알고 있는 값들의 선형 조합을 통해 원하는 곳의 값을 추정하는 기법이다. IDW와 달리 값을 추정할 때 거리만을 반영하는 것이 아니라 주변 값들의 상관관계를 더해서 추정하게 된다. 크리깅은 스무딩 이펙트로 인해 전체적인 경향을 파악하기 적합하다.

공간보간법에 대한 연구동향을 살펴보았을 때, 일반적으로 실측값이 많을 때 더 정확도가 높은 크리깅을 기반으로 한 공간적 분포 추정에 대한 연구가 주로 진행되고 있다.

2.2 기계학습

공간데이터를 처리하는데 있어서는 앞에서 언급한 공간보간법들 외에 기계학습 방법도 적용되고 있다. 예를 들어 기상분야에서는 기온, 태양에너지, 바람, 강우, 하천유량 등을 예측하는데 인공신경망

(Artificial Neural Network)과 SVR과 같은 기법들이 활용되고 있다. Lee[8]는 도시기온 예측모형을 구축하기 위해 인공신경망 및 회귀분석을 비교분석하였고 인공신경망의 우수성을 나타내었다. Kuligowski[9]는 대서양 중부 지역 네 곳을 대상으로 6시간 강수예측 정보를 생산하기 위해 인공신경망을 이용한 연구를 수행하였고, Hall[10]은 인공신경망을 통해 정량적 강수 정보와 강수 확률을 예측하였다. Luk[11]은 세 가지 형태의 인공신경망을 가지고 15분 후에 대한 공간적 강우분포를 예측하기 위한 연구를 진행하였다.

이처럼 기계학습 알고리즘은 분류나 클러스터링 알고리즘뿐만 아니라 회귀분석이 가능한 알고리즘을 통해 기온, 강우, 바람 등의 다양한 요소를 예측 및 추정하고자 하는 연구가 이루어지고 있다. 기상분야에 적용되고 있는 알고리즘으로는 앞서 언급한 인공신경망, SVR을 주로 사용하고 있다. 이 알고리즘들은 정교화 되어 성능이 좋은 편이지만 학습 파라미터의 수가 많아지면 학습 시간이 많이 소요되기도 하며 최적의 해를 찾기 또한 어려워진다. 기상관측 데이터에서 측정되는 다양한 관측요소들은 그 것들 간의 유의한 관계를 가지고 있음에 따라 특정 값 추정에 다양한 요소들을 활용하기에는 적절하지 못할 수 있다. 또한 추정하고자 하는 값과 그 파라미터에 따라 적합한 알고리즘이 다르기 때문에 강우량 추정에 대한 정확도 향상을 위해 다양한 알고리즘 적용이 필요하다.

3. 강우 추정 알고리즘의 검증

3.1 데이터 구성

본 연구에서 사용되는 데이터는 SK Techx에서 구축한 서울시 전역의 1~2km 해상도를 가지는 255개소 자동기상관측장비에서 수집된 것으로, 수집되

표 1. AWS 수집 데이터 유형
Table 1. Data type of observed AWS

No	Variable	No	Variable
1	observation point number	15	last 1 hour accumulated rainfall
2	observation point name	16	last 6 hour accumulated rainfall
3	observation point type	17	last 12 hour accumulated rainfall
4	observation point latitude	18	last 24 hour accumulated rainfall
5	observation point longitude	19	present temperature
6	Sky status code	20	maximum temperature
7	Sky status name	21	minimum temperature
8	precipitation type	22	wind direction
9	1 hour accumulated precipitation	23	wind speed
10	1 hour accumulated rainfall	24	relative humidity
11	1 day accumulated rainfall	25	spot atmospheric pressure
12	last 10 minute accumulated rainfall	26	sea-level pressure
13	last 15 minute accumulated rainfall	27	lighting
14	last 30 minute accumulated rainfall	28	observation time

는 데이터 요소는 다음 <표 1>과 같다. 해당 데이터는 1분 주기로 수집되며 강우와 관련된 속성으로 강수 형태와 함께 10분·15분·30분·1일 등의 누적 강우량을 포함하고 있다. 본 연구에서는 강우가 나타난 2015년도 8월 16일을 사례로 16시 ~ 20시 까지 10분 주기의 데이터 6,344개의 인스턴스로 실험을 수행하였다. 여기에 강우와 관련된 다양한 속성을 활용할 수 있으나 공간보간법과 기계학습의 알고리즘별 비교분석을 위해 위치정보인 위도, 경도 두 가지를 속성으로 학습데이터를 구성하였다. 추정하고자 하는 대상은 강우 요소 중 10분 누적 강우량이다. 이렇게 구성한 데이터로부터 학습 모델을 구축하고 검증하여 강우량 추정을 위한 알고리즘 비교분석을 수행하였다.

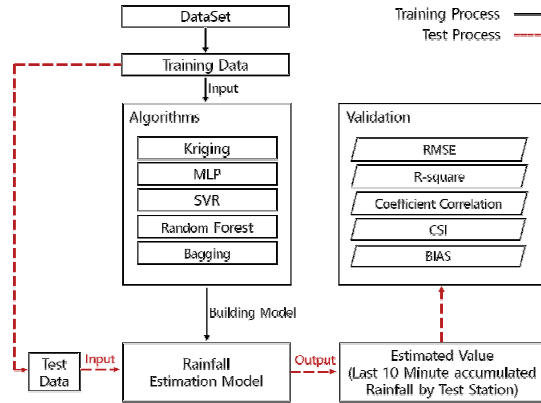


그림 1. 알고리즘 평가 구성도
Figure 1. Flow chart for algorithm evaluation

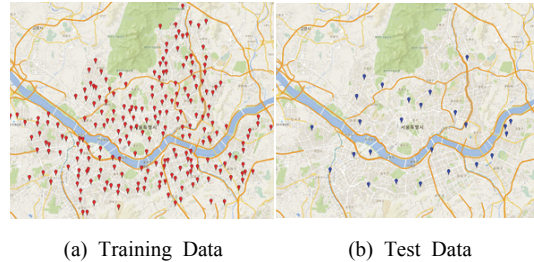


그림 2. 데이터 셋 설명
Figure 2. Dataset description

3.2 실험 방법

본 연구에서는 강우량 추정을 위한 알고리즘 비교분석을 수행한다. 이에 대한 전체적인 평가 구성도는 <그림 1>과 같다.

알고리즘 평가에는 훈련(training) 과정과 테스트(test) 과정으로 나누어지며, 여기에는 각각의 훈련 데이터와 테스트 데이터를 구분해서 실험을 수행한다. 본 실험에서는 <그림 2>와 같이 서울시 전역의 255개소 중 230개소 5,739개의 인스턴스를 훈련 데이터로 하며, 나머지 25개소 605개의 인스턴스를 테스트 데이터로 구분하였다.

본 실험에서 비교분석할 알고리즘은 다섯 가지

로 크리깅, MLP, SVR, 랜덤포레스트, 배깅이다. 여기서 배깅의 분류기는 k-최근접 이웃 분류기(k-Nearest Neighbor classifier)를 사용한다. 동일한 훈련 데이터를 이용해 각각의 알고리즘별 학습 모델을 생성하고, 생성된 모델에 테스트 데이터를 입력 데이터로 하여 강우량을 추정하였다.

이렇게 산출된 알고리즘별 강우량 추정은 RMSE(Root Mean Square Error), R-square, 상관계수(Correlation Coefficient), CSI(Critical Success Index), BIAS(frequency BIAS)의 총 다섯 가지 검증 방법을 통해 비교한다. 여기서 CSI와 BIAS는 양분 예보기법이며 강수 정확도의 척도로 활용되는 지표이다. 이 두 가지 방법은 강우량 추정에 대한 검증을 위해 임계값(thresholds)을 0.5mm 간격으로 분할하고 0.5~12.5mm까지의 구간으로 정하여 실험을 수행한다. 양분예보기법은 임계값을 기준으로 그 값이 클 경우 Yes, 작을 경우 No로 나누고, <표 2>의 강우 분할표를 이용해 강우추정 성능을 평가하여 모델의 추정 능력과 오차의 종류를 파악한다. 여기서 추정값은 알고리즘별로 생성된 모델을 통해 산출된 값이고, 관측값은 테스트 데이터의 실측값으로 하여 임계값과의 비교를 통해 Yes와 No로 판별하고 각각의 강수 정확도 평가를 위한 CSI와 BIAS를 계산한다. 각각 식(1)과 식(2)의 수식을 통해 값을 산출하게 된다.

표 2. 강우 분할표
Table 2. Rain contingency table

Observed	Forecast		
	Yes	No	Total
Yes	H(hits)	M(misses)	observed yes
No	F(false alarms)	C(correct negatives)	observed no
Total	forecast yes	forecast no	N= H+M+F+C

$$CSI = \frac{hits}{hits + misses + false} \quad (1)$$

$$BIAS = \frac{hits + false}{hits + misses} \quad (2)$$

CSI는 0에서 1사이의 값을 가지며, 1에 가까울수록 강우 현상에 대한 추정을 정확하게 한 것을 의미한다. BIAS 또한 0에서 1 사이의 값을 가지게 되며 관측 빈도수와 추정 빈도수의 비를 통해 과소추정과 과대추정에 대한 추정경향을 나타내게 된다. BIAS도 1에 가까울수록 보다 정확한 추정경향을 나타내는 것으로 볼 수 있다.

4. 실험 결과 및 분석

본 실험에서는 추정성능을 평가하기 위해 앞서 언급한 다섯 가지의 검증 방법으로 비교분석을 수행한다. 먼저 비교를 위한 성능 지표로서 RMSE, R-square, 상관계수 검증방법별 결과를 <표 3>에 나타내었다. RMSE를 봤을 때 추정 오차가 가장 낮은 성능을 보이는 알고리즘은 배깅과 크리깅 순으로 나타났다. 그에 비해 일반적으로 기상분야에서 많이 쓰이는 MLP나 SVR은 상대적으로 오차가 높은 것으로 나타났다. R-square는 회귀분석에 있어

표 3. 알고리즘별 RMSE, R-square, 상관계수 비교 결과
Table 3. Comparison result for RMSE, R-square, Correlation Coefficient by algorithms

	RMSE	R-square	Correlation Coefficient
Kriging	1.027	0.768	0.884
MLP	1.827	0.339	0.613
SVR	1.677	0.372	0.671
Random Forest	1.208	0.669	0.832
Bagging	0.935	0.811	0.906

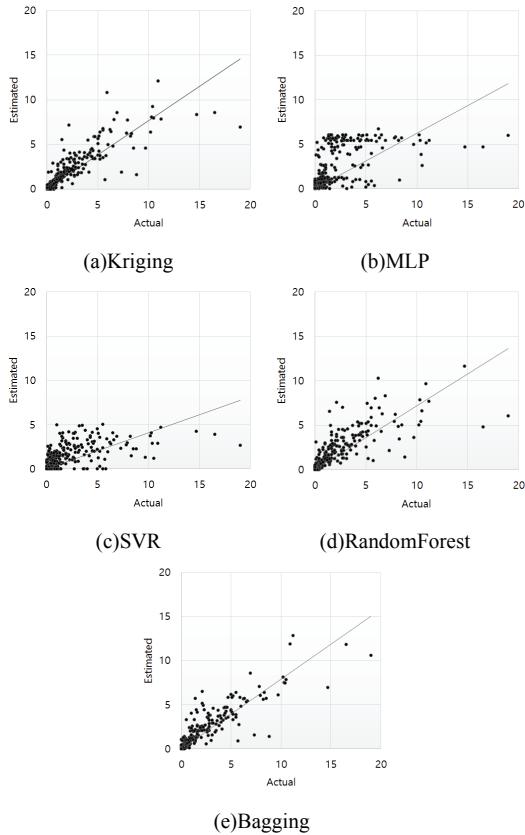


그림 3. 관측값과 추정값 간의 산점도
Figure 3. Scatter plot between actual data and estimated data

서 정확도에 대한 설명을 나타내는 것으로, 배깅의 경우 실제 자료를 81%의 정확도로 설명해주고 있음을 알 수 있다. 그리고 실측값과 추정값에 대한 상관계수를 보면 배깅과 크리깅, 랜덤포레스트의 경우 0.8 이상으로 상관관계가 높은 것으로 확인된다. 이와 같이 세 가지 검증방법에 대해서 분석해 본 결과, 배깅이 가장 좋은 결과를 나타내는 것으로 확인됐다. <그림 3>은 테스트 데이터를 대상으로 관측값과 추정값의 산점도를 나타낸 것인데, 배깅이 다른 알고리즘에 비해 퍼짐의 정도가 가장 낮은 것으로 확인되었다. 그만큼 강우를 추정하는데 있어서 보다 정밀하고 정확한 성능을 가지는 알고리즘이라고 볼 수 있다.

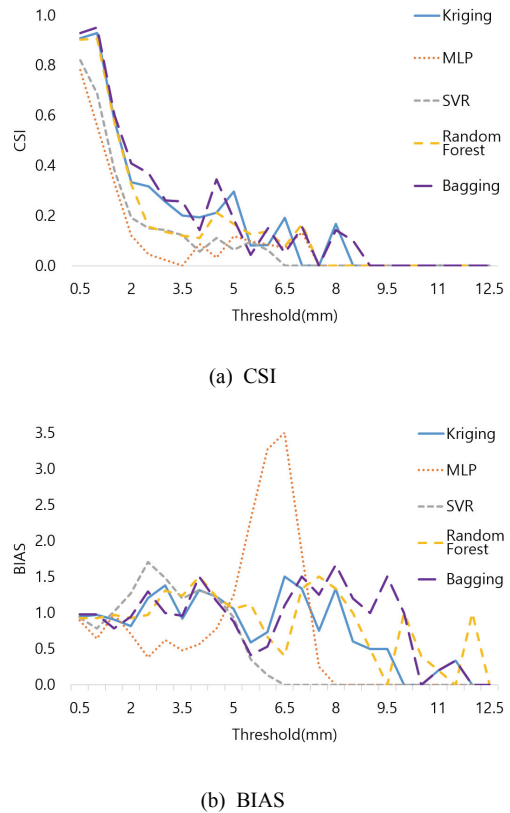


그림 4. 추정 강우량의 임계값별 CSI, BIAS 결과
Figure 4. Comparison result for CSI, BIAS by thresholds

양분예보기법인 CSI와 BIAS에 대한 비교 결과는 <그림 4.>와 같다. 먼저 <그림 4.>의 (a) CSI를 보면 임계값 1mm 이하에서는 1에 가까운 높은 정확도를 보이다 3mm에서 급격히 낮아지는 것을 볼 수 있다. 그러면서 5mm부터 8.5mm까지 0.0~0.2 사이를 유지하다가 9mm부터는 전부 0을 나타낸다. 해당 결과를 보면 대부분의 알고리즘은 강우량이 적을 때 추정 능력이 좋다는 것을 볼 수 있다. <그림 4.>의 (b) BIAS를 보면 임계값 0.5~1mm까지 1에 근접하고 2mm부터 과대추정의 경향을 보이다가 5~6.5mm까지 과소추정하는 경향을 보인다. 그 중 MLP의 경우 다른 알고리즘과 달리 과대추정하는 경향을 보이다 8mm부터 0을 나타낸다. 이는 해당

표 4. 알고리즘별 CSI, BIAS 결과
Table 4. CSI, Bias by algorithms

	Average	
	CSI	BIAS
Kriging	0.190	0.764
MLP	0.099	0.737
SVR	0.115	0.494
Random Forest	0.164	0.860
Bagging	0.204	0.887

임계값 부분의 강우 정보를 정확하게 추정하기 어렵다는 결과를 보인다. SVR은 지속적으로 과소추정을 하다가 6.5mm에서부터 0을 나타낸다. 나머지 알고리즘 중 크리깅과 랜덤포레스트는 임계값이 높아질수록 과소추정하는 경향이 발생되고, 배깅의 경우는 1에 가까운 값에서 과대추정을 반복하다 10mm 이후부터 과소추정하게 된다. 즉, 크리깅과 랜덤포레스트는 실제 강우의 발생수를 적게 추정하였고, 배깅의 경우 실제 강우 발생수 보다 강우를 추정한 횟수가 많았다는 특징을 나타낸다.

<표 4.>는 각 알고리즘별 CSI와 BIAS의 실험 결과에 따른 평균치를 나타낸다. 전체적으로 배깅 알고리즘의 CSI가 다른 알고리즘의 비해 가장 높은 결과를 나타냈다. BIAS의 경우는 전체적으로 과소추정하는 경향을 보이고 있으며, 이 경우도 배깅 알고리즘이 1에 가장 가까운 BIAS를 나타냄에 따라 비교한 알고리즘 중에서 추정 경향이 가장 정확하다는 것을 알 수 있다.

결과를 종합해보면, 다섯 가지 알고리즘 중 배깅이 전체적인 검증 결과에서 가장 좋은 성능을 나타낸다. 배깅은 비교 알고리즘 중 추정 오차가 적고 가장 정확한 추정 경향을 보이며, 추정한 값과 실제 관측값 간의 상관관계도 가장 높은 것으로 나타났다. 그 다음으로는 크리깅, 랜덤포레스트, SVR, MLP 순으로 나타난다.

대체로 적은 강우량일 경우 대부분의 알고리즘들의 추정 성능이 좋으나 2mm의 강우부터 성능이 현저히 떨어지는 것으로 나타났다. 이는 집중호우가 잦은 여름철 강수특성상 강한 강도의 강우가 발생하는 이벤트의 경우 이에 대한 추정 정확도가 낮다고 할 수 있다. 이를 개선하기 위해서는 본 실험에서 활용한 위치정보와 강우에 대한 요소뿐만 아니라, 강우와 관련된 습도나 시간별 강우누적량 등의 다른 요소들도 고려해야할 것으로 판단된다.

5. 결론 및 향후 연구

본 논문에서는 고해상도의 강우량 추정을 위해 공간보간법과 기계학습 알고리즘을 비교분석하였고, 그 중 가장 정확한 추정 능력을 나타내는 알고리즘을 분석하였다. 여기에 활용된 데이터는 2015년 8월 16일 16시부터 20시의 강우 사례를 대상으로, SK Techx의 서울시 AWS 255개소 중 230개소를 훈련 데이터, 25개소를 테스트 데이터로 선정하여 비교분석을 수행하였다. 이에 대한 검증방법은 RMSE, R-square, 상관계수, CSI, BIAS의 다섯 가지 방법을 이용하였다. 그 결과, 배깅 알고리즘이 강우 추정에 있어서 전체 검증에 가장 우수한 성능을 나타내었다. 실험에서의 강우량 추정 비교분석 결과를 토대로 기계학습 알고리즘을 적용하여 강우량 추정을 한다면 기존의 공간보간법 보다 정확한 산출 결과를 얻을 수 있을 것으로 기대된다. 또한 국지성 호우에 대한 고해상도 예측 자료 생산뿐만 아니라 도로에서 표준으로 사용하고 있는 노드링크 단위의 고해상도 도로기상 정보 산출에 활용할 수 있을 것이다.

향후 연구에서는 강한 강우 추정의 정확도 개선을 목표로, 기상변수, 시공간적 특성을 고려한 다변량 분석(Multivariate Analysis)을 수행하고, 기계학습 알고리즘별 파라미터 최적화하여 강우량 추정

정확도를 향상시키기 위한 연구를 수행하고자 한다.

References

- [1] H. Jung, H. J. Shin, J. Y. Park, I. K. Jung, and S. J. Kim, *A study of spatial interpolation impact on large watershed rainfall considering elevation*, Journal of The Korea Society of Agricultural Engineers. Vol. 53, No. 6, pp. 23-29, 2011.
- [2] K-S. Lee, *Spatial estimation of point observed environmental variables: A case study for producing rainfall acidity map*, Journal of the Korean Society of Remote Sensing, Vol. 11, No. 3, pp. 33-47, 1995.
- [3] N. F. Glenn, and J. R. Carr, *The use of geostatistics in relating soil moisture to RADARSAT-1 SAR data obtained over the great basin, Nevada, USA*, Computers & Geosciences, Vol. 29, No. 29, pp. 577-586, 2003.
- [4] G. C. Simbahan, A. Dobermann, P. Goovaerts, J. Ping, and M. L. Haddix, *Fine-resolution mapping of soil organic carbon based on multivariate secondary data*, Geoderma, Vol. 132, No. 3-4, pp. 471-489, 2006.
- [5] H-L. Cho, and J-C. Jeong, *Application of spatial interpolation to rainfall data*, The Journal of GIS Association of Korea, Vol. 14, No. 1, pp. 29-41, 2006.
- [6] N-W, Park, and D-H, Jang, *Mapping of temperature and rainfall using DEM and multivariate kriging*, Journal of the Korean Geographical Society, Vol. 43, No. 6, pp. 1002-1015, 2008.
- [7] J-C, Park, and M-K, Kim, *A study on the use of a terrain aspect variable in producing the precipitation distribution map applying cokriging: A case of Jeju Island*, Journal of the Korean Geomorphological Association, Vol. 16, No. 3, pp. 59-66, 2009.
- [8] S-G. Lee, S-G. Jung, W-S. Lee, and K-H. Park, *A predictive model for urban temperature using the artificial neural network*, Journal of Korea Planning Association, Vol. 46, No. 1, pp. 129-142, 2011.
- [9] R. J. Kuligowski, and A. P. Barros, *Localized precipitation forecasts from a numerical weather prediction model using artificial neural networks, weather and forecasting*. Vol. 13, pp. 1194-1204, 1998.
- [10] T. Hall, *Precipitation forecasting using a neural network, weather and forecasting*, Vol. 14, pp. 338-345, 1999.
- [11] K. C. Luk, J. E. Ball, and A. Sharma, *An application of artificial neural Networks for rainfall forecasting*, Mathematical and Computer Modelling, Vol. 33, No. 6-7, pp. 683-693, 2001.
- [12] Y. H. Lee, J-E. Nam, and S. W. Joo, *Optimization of Z-R relationship in the summer of 2014 using a micro genetic algorithm*, Journal of Korean Institute of Intelligent Systems, Vol. 26, No. 1, pp. 1-8. 2016.
- [13] Y-M. Kim, M-N. Hwang, T. H. Kim, C-H. Jeong, and D-H. Jeong, *Big data mining for natural disaster analysis*, Journal of the Korean Data & Information Science Society, Vol. 26, No. 5, pp. 1105-1115, 2015.
- [14] M. C. V. Ramirez, H. F. C. Velho, and N. J. Ferreira, *Artificial neural network technique*

for rainfall forecasting applied to the São Paulo Region, Journal of Hydrology, Vol. 301, No. 1-4, pp. 146-162, 2005.

[15] S. G. Baek, B. J. Kim, and Y. T. Lim, Analysis of rain impacts on freeway trip characteristics, Journal of Korea Transportation Research Society, Vol. 26, No. 3, pp. 119-128, 2008.

[16] Numerical Weather Prediction Model, http://web.kma.go.kr/aboutkma/intro/supercom/model/model_category.jsp, Jan. 2017.

기계학습을 적용한 고해상도 강우량 추정에 관한 연구

김민규, 이무훈

한국의국어대학교 차세대도시농림융합기상사업단

요 약

다양한 기후변화 중 국지성 호우는 인명 피해 및 막대한 재산상의 피해를 유발하는 위험 기상이다. 기후변화에 따른 국지성호우는 정확한 예측이 우선되어야 한다. 현재 5km 해상도의 동네예보 서비스는 스톱 규모(200m~1km)에서 발생하는 국지성 호우를 예측하는데 한계가 있다. 다운스케일링을 위해 기상분야에서는 공간보간법을 활용하고 있으나 강우의 특성상 강우량 추정 정확도가 낮은 실정이다. 본 논문에서는 강우량 추정 정확도 향상을 위해 공간보간법과 기계학습 알고리즘을 대상으로 강우량을 추정하고 정확도를 비교분석하여 기상분야의 다운스케일링 기법으로 기계학습 알고리즘을 적용하고자 한다. 이를 위해 강우가 관측된 2015년 8월 16일 16시부터 20시까지의 자동기상관측장비 데이터를 활용하여 사례 분석을 수행하였다. 비교분석을 위한 검증방법은 RMSE, R-square, Correlation Coefficient, CSI, BIAS를 사용하였으며, 그 결과, 기계학습의 배경 알고리즘이 가장 높은 정확도로 강우량을 추정함을 확인할 수 있었다.

감사의 글

이 연구는 기상청 차세대도시농림융합스마트기상 서비스개발(WISE) 사업(KMIPA-2012-0001-1)의 지원으로 수행되었습니다.



Min Gyu Kim received the bachelor's degree in the Department of GIS from the Namseoul University in 2012. He received the M.S. degree in the Department of Geoinformatic Engineering from Inha University in 2014, respectively. He is currently a researcher at WISE Institute.

E-mail address: gisdev107@gmail.com



Moo Hun Lee received his B.S., M.S. and Ph.D. degree in computer engineering from Hannam University, Daejeon, Korea, in 2002, 2004, and 2013 respectively. From 2008 to 2015, he was a senior researcher at the Next Generation Smart TV Research Department of Electronics and Telecommunications Research Institute (ETRI), Korea. He is currently working as senior researcher in the Weather Information Service Engine(WISE) Institute of Hankuk University of Foreign Studies, Korea. His research interests include data mining, information retrieval, and semantic search.

E-mail address: macbethe@gmail.com