

기상 IoT 센서 데이터를 활용한 서리 예측 연구

서재하¹, 안은지², 문애경³, 박주영³

¹경희대학교, ²안동대학교, ³ETRI

jaeha.5eo@khu.ac.kr, {akmoon, jypark}@etri.re.kr

Frost Prediction Research Using IoT Sensor Weather Data

Jaeha Seo¹, Eunji An², Aekyeong Moon³, Juyoung Park³

¹Kyunghee Univ., ²Andong Univ., ³ETRI

요약

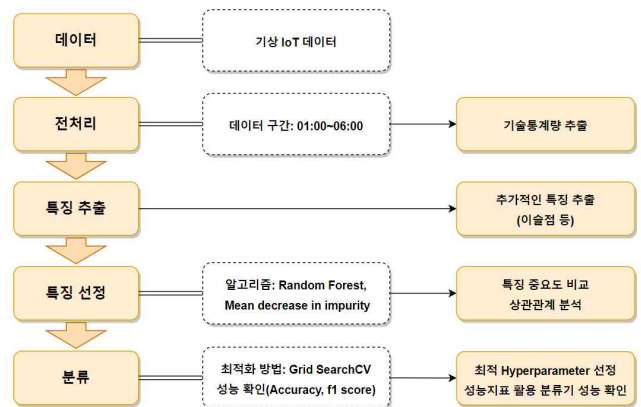
서리 피해에 미리 대응하지 못해서 발생하는 농가의 손실을 최소화하기 위해서는 정확한 서리 예보가 필수적이다. 본 논문에서는 서리 발생에 영향을 주는 기상인자에 머신러닝을 적용하여 서리 발생을 예측하는 방법을 제안한다. 농업 기상 IoT 데이터를 전처리한 후 랜덤 포레스트를 기반으로 한 특징 선정을 통해 서리 발생에 주요한 영향을 주는 변수들을 도출하고, 교차 검증과 적합한 성능지표를 통해 모델을 최적화하였다. 실제 서리 관측 데이터를 통해 검증을 수행한 결과 우수한 성능을 가지는 랜덤 포레스트 기반의 서리 예측 모델을 구현할 수 있었다.

I. 서론

최근 지구온난화 현상으로 인해 작물의 개화시기가 앞당겨지고 있고 이때 발생하는 서리는 작물 생육에 큰 영향을 미친다. 서리가 발생하는 시기를 정확히 예측할 수 있다면 농민들은 서리를 대비할 수 있고, 서리 발생으로부터 오는 손실을 막을 수 있다[1]. 따라서, 봄, 가을철 서리로 인한 농가 피해를 최소화하기 위해 정확한 서리 발생 예측이 필요하다. [2][3]에서는 서리 예측의 정확성을 높이기 위한 서리 발생일의 기온 위주 기상 특성을 제시하였다. 서리는 일반적으로 날씨가 맑으면서 춥고 바람이 잘 불지 않을 때 발생한다. [4]의 연구에서 서리 발생 예측을 위해 도출된 기상인자는 최저기온, 평균풍속, 상대습도, 이슬점 온도이다. 서리가 발생한 날이 서리가 발생하지 않은 날에 비해 최저기온, 이슬점 온도, 평균풍속은 낮게 나타나고 상대습도는 높게 나타난다.

본 논문에서는 농업 기상 IoT 데이터를 전제로 머신러닝 알고리즘을 사용하여 서리 발생을 예측한다. 그림 1은 본 논문에서 제시한 서리 예측을 위한 전체적인 데이터 분석 단계를 나타낸다. 특징 선정 단계에서는 랜덤 포레스트 기반의 특징 중요도를 활용하여 서리 발생에 주요한 영향을 주는 변수들을 결정한다. 특징 중요도를 계산하기 위하여 랜덤 포레스트 분류기 기반의 MDI(Mean Decrease in Impurity) 모델을 활용한다. 분류 단계에서는 서리 예측 모델의 구축을 위해 분류기 알고리즘을 구현하며 최적화 및 두 가지 성능지표를 활용하여 분류기 성능을 확인한다. 이

연구를 통해 서리 예측 정확도를 높임으로써 서리 피해를 사전 예방하고자 한다.



[그림 1] 데이터 분석 단계

II. 본론

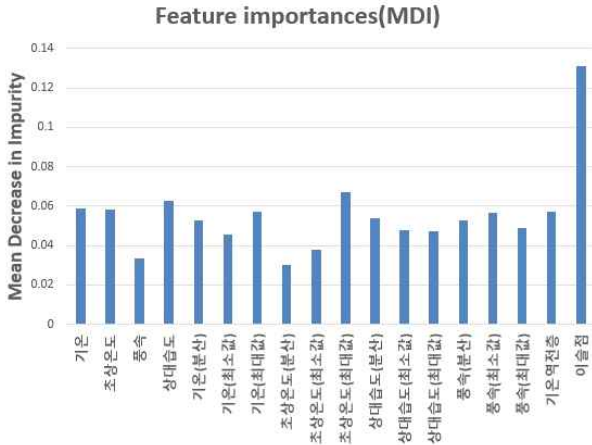
2.1 데이터 전처리 및 특징 선정

본 연구에서는 2020년 1월 1일부터 5월 31일까지 경상북도 영천시 화북면의 IoT 기상 관측 장비에서 수집한 기온, 상대습도, 강수량, 초상 온도 등의 센서 데이터를 활용한다. 관측 기간 152일 중 서리가 관측된 날은 29일이고, 서리 현상이 발생하는 새벽 시간대(01:00 ~ 06:00)의 데이터를 분류 모델 학습에 사용한다.

머신러닝 모델을 학습시키기 위해서 설정한 시간대의 데이터를 요약하여 대표할 수 있는 새로운 데이터 세트를 산

술 평균, 최대/최솟값, 분산 등의 기술 통계량을 활용하여 생성한다. 또한, 기존 연구를 바탕으로 서리 발생에 영향을 주는 변수로서 기온역전층[1]과 이슬점을 추가 정의한다.

본 연구에서는 MDI 알고리즘을 이용하여 각 특징의 중요도를 계산한 뒤, 높은 중요도의 특징들을 활용하여 모델을 학습한다. MDI 중요도는 트리형 알고리즘에서 각 데이터가 split될 때 불순도의 감소분에 따라 각 특징의 중요도를 계산한다.



[그림 2] 특징 중요도(MDI)

그림 2는 각 특징의 MDI 중요도를 나타낸다. 추가한 기온역전층과 이슬점이 가장 높은 중요도를 가졌고, 상대습도의 최솟값과 기온의 분산 값 또한 높은 중요도를 가진다. 이 결과에 따라 기온역전층과 이슬점, 상대습도의 최솟값, 그리고 기온의 분산 값을 활용하여 모델을 학습한다.

2.2 실험 및 성능 평가

본 장에서는 랜덤 포레스트의 기본 설정 모델과 초매개변수(Hyper Parameter) 최적화 모델의 성능 지표 비교를 통하여 머신러닝 모델의 발전 가능성과 이를 사용한 서리 예측 가능 여부를 확인한다.

랜덤 포레스트 분류기 모델을 생성할 때 설정하는 트리 개수, 트리의 최대 깊이 등의 여러 초매개변수에 따라 분류기의 성능 차이가 날 수 있다. 본 연구에서는 Scikit-learn에서 제공하는 GridSearchCV를 통해 초매개변수 변화에 따른 성능을 측정하여 초매개변수를 설정한다. 표 1은 기본 설정 모델과 최적화 모델의 초매개변수를 비교한 표이다.

	Parameter Name	Value
Default	Number of estimator	100
	Max depth of estimator	None
	Class weight	None
Optimized	Number of estimator	1000
	Max depth of estimator	3
	Class weight	'balanced'

[표 1] 초매개변수 비교

두 모델 성능 평가는 아래의 식으로 계산되는 두 가지 성능지표(Accuracy, F1-score)를 통해 진행한다.

$$(Accuracy) = \frac{TP + TN}{TP + FN + FP + TN} \quad (1)$$

$$(F1-score) = 2 \times \frac{Precision \times recall}{Precision + recall} \quad (2)$$

서리가 관측된 날의 데이터 세트보다 그렇지 않은 날보다 4배 정도 많은 편향된 데이터 세트를 모델의 학습에 활용했기 때문에, 이러한 편향된 데이터 세트의 성능을 잘 표현해 줄 수 있는 F1-score 값이 중요하다. 기본 설정 모델의 Accuracy와 F1-score는 각각 0.8, 0.25이고, 최적화 모델은 0.93, 0.83으로 두 지표 모두 최적화 이후 상승하였다.

III. 결론

본 연구에서는 서리 발생 예측에 필요한 중요 인자들을 도출해보고, 랜덤 포레스트 모델 최적화를 통해 서리의 발생을 예측해보았다. 전체 데이터 세트의 숫자가 크지 않았고 편향적인 데이터 세트를 활용하였지만 적절한 특징 데이터를 도출하고, 최적의 초매개변수를 설정함으로써 서리의 발생을 정확하게 예측할 수 있는 가능성을 실험을 통해 확인하였다. 보다 많고 다양한 기상 데이터를 확보한다면 본 연구에 기초하여 정확한 서리 예측으로 농작물 피해를 예방할 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

본 결과물은 농림축산식품부의 재원으로 농림식품기술기획평가원 첨단생산기술개발사업의 지원을 받아 연구되었음 (317015-6)

참 고 문 헌

- [1] Yosuke Tamura, Liya Ding, Kosuke Noborio, Kazuki Shibuya, "Frost Prediction for Vineyard Using Machine Learning", 2020 Joint 11th International Conference on Soft Computing and Intelligent Systems and 21st International Symposium on Advanced Intelligent Systems (SCIS-ISIS), Dec 2020
- [2] 권영아, 이효신, 권원태, 부경은, "서리 피해 방지를 위한 서리 발생일의 기상 특성에 대한 연구", 대한지리학회지 제43권 제6호, 2008.12
- [3] 노일석, 도해원, 김수옥, 이승재, "2018년 봄철 서리 발생일 기상 분석", 한국기상학회 학술대회 논문집, 2019.10
- [4] Y.S. Kim, K.M. Shim, M.P. Jung, I.T. Choi, "Study on the Estimation of Frost Occurrence Classification Using Machine Learning Methods", Korean Journal of Agricultural and Forest Meteorology, Vol. 19, No. 3, (2017), pp. 86~92, 2017