

기계학습법을 이용한 서리 발생 구분 추정 연구

김용석 · 심교문* · 정명표 · 최인태

국립농업과학원 농업환경부 기후변화생태과

(2017년 8월 21일 접수; 2017년 9월 23일 수정; 2017년 9월 23일 수락)

Study on the Estimation of Frost Occurrence Classification Using Machine Learning Methods

Yongseok Kim, Kyo-Moon Shim*, Myung-Pyo Jung and In-tae Choi

National Institute of Agricultural Sciences, RDA, Wanju, Korea

(Received August 21, 2017; Revised September 23, 2017; Accepted September 23, 2017)

ABSTRACT

In this study, a model to classify frost occurrence and frost free day was developed using the digital weather forecast data provided by Korea Meteorological Administration (KMA). The minimum temperature, average wind speed, relative humidity, and dew point temperature were identified as the meteorological variables useful for classification frost occurrence and frost-free days. It was found that frost-occurrence date tended to have relatively low values of the minimum temperature, dew point temperature, and average wind speed. On the other hand, relatively humidity on frost-free days was higher than on frost-occurrence dates. Models based on machine learning methods including Artificial Neural Network (ANN), Random Forest(RF), Support Vector Machine(SVM) with those meteorological factors had >70% of accuracy. This results suggested that these models would be useful to predict the occurrence of frost using a digital weather forecast data.

Key words: Frost, Artificial neural network, Random forest, Support vector machine

I. 서 론

서리는 가을부터 봄까지 발생하는데 가을철 서리는 발작물의 생육에 피해를 발생시키는 일이 많고 봄철에 내린 서리는 과수작물의 생육에 피해를 발생시키는 일이 많아 농민들이 작물을 재배하는데 큰 어려움을 겪게 한다. 일반적으로 서리는 날씨가 맑으면서 춥고 바람이 잘 불지 않을 때 많이 발생한다고 알려져 있다. 좀 더 구체적으로 Temeyer *et al.*(2003)은 서리가 발생 되는 상황을 다음 두 가지로 설명하였다. 첫째는 복사 냉각에 의해 역전층이 형성되었을 때 지표층이 15~20m의 상층부보다 차가워진 상태에서 수증기가 유입

되면서 발생한다고 하였고, 둘째는 이류현상에 의해 습윤한 기단에 차가운 공기가 유입되면서 서리가 발생한다고 하였다. 이러한 서리가 발생할 조건에서 기상 인자의 특성을 분석하여 서리 발생에 따른 피해를 줄이기 위한 연구가 수행되어 왔다. 이와 관련하여 Kwon *et al.*(2008)은 서리가 발생한 날의 기상 요소별 특성을 분석하는 연구를 수행하였으며, Han *et al.*(2009)은 서리발생 전날의 기상특성을 분석하여 서리발생 전날의 24시기온, 구름량, 5일간의 강수량 등을 이용한 판별분석(DA, discriminant analysis)을 통해 서리발생을 예측한 연구를 수행하였다. 그리고, Chung *et al.*(2004)은 최저기온의 공간단위의 상세화



* Corresponding Author : Kyo-Moon Shim
(kmschim@korea.kr)

를 통해 -3°C 기준으로 지역별 서리피해 경보의 가능성에 대해 연구하였다.

또한, Temeyer *et al.*(2003)은 서리가 발생하는데 중요한 기상인자인 기온, 이슬점온도, 풍속, 지표면온도를 기계학습법 중 하나인 인공신경망으로 예측하는 연구를 수행하기도 하였는데, 최근 기상자료의 규모가 커지고 다양한 형태의 자료가 생산되면서 효과적으로 기상현상을 예측하고 분석하기 위해서 기계학습법이 많이 사용되고 있다. 그 예로서, Shank *et al.*(2008)는 상대습도, 일사량, 기온, 풍속과 수증기압을 입력인자로 하여 인공신경망(ANN, artificial neural network)을 통한 이슬점온도 예측 연구를 수행하였으며, Smith *et al.*(2007)은 상해와 관련하여 인공신경망으로 기온을 예측하는 연구를 수행하였다. Abhishek *et al.*(2012)도 인공신경망의 여러 설정값을 변경하여 최고기온의 예측력을 비교 분석하였다. Naing and Htike(2015)은 월최고기온과 최저기온을 예측하기 위해 K-최근접 이웃(K-nearest neighbors) 알고리즘과 인공신경망, 랜덤 포레스트(RF, random forest) 등의 정확도를 비교하는 연구를 수행하였다.

본 연구에서는 기존 Han *et al.*(2009)에서 제시한 서리 발생 전날의 기상특성을 반영하는 기상인자로 다음날의 서리발생을 예측하는 모형과 다르게 2일(48시간) 후의 기상예측값을 제공하는 기상청의 동네예보자료를 이용한다는 전제로 서리발생 당일의 기상특성을 분석하여 최근 분류분석에 많이 사용되는 기계학습법인 인공신경망, 랜덤포레스트, 서포트벡터머신(SVM, support vector machine)을 이용한 서리발생을 예측하는 모형에 대해서 활용가능성을 검토해 보았다.

II. 재료 및 방법

2.1. 기상자료 수집 및 기상인자 선택

기상청 종관기상관측지점 30곳(Table 1)의 2004년부터 2013년까지의 초상일의 자료를 이용하여 해당되는 날의 서리 발생과 관련된 기상인자들의 자료를 정리하였고, 서리가 발생하지 않은 날은 초상일의 전날로 설정하고 기상인자들의 자료를 정리하였다. 기상인자는 기존 문헌들에서 나온 최저기온, 평균풍속, 구름량, 5일간 강수량, 3일간 강수량, 상대습도, 이슬점온도로 정하였으며(Han *et al.*, 2009; Robinson and Mort, 1996; Temeyer *et al.*, 2003), 기상청 동네예보자료로 대체가 가능한 인자로 선택하였다. 일반적으로 서리와 관련성이 많을 것으로 예상되지만 기상청 동네예보가 제공되지 않는 초상온도나 지면온도는 사용하지 않았다. 수집한 데이터는 서리가 발생한 날의 데이터 248세트와 서리가 발생하지 않은 날의 248세트로 총 496세트였으며, 이중 서리가 발생한 날의 데이터와 서리가 발생하지 않은 날의 데이터 비율을 같게 하여 초모형 구축에 326세트, 정확도 검증에 170세트를 사용하였다.

기상인자의 선택은 통계프로그램 R(3.3.1)을 사용하여 단계적 방법(stepwise method)으로 판별분석과 로지스틱 회귀분석을 실시하였다. 판별분석에서는 전체변동에 대한 집단 내 변동의 비율인 wilk's lambda 값으로 변수를 선택하였으며, 로지스틱 회귀분석을 이용하여 아카이케의 정보기준(AIC, akaike informaiton criterion)를 이용하여 변수를 선정하였다(Kim *et al.*, 2014; Koo *et al.*, 2013).

Table 1. Synoptic weather stations collected the frost data

Region	Station
Gangwondo	Daegwallyeong, Chuncheon, Wonju, Sokcho, Gangneung, Cheorwon, Bukgangneung, Donghae, Yeongwol
Gyeonggido	Suwon, Dongducheon
Chungcheongdo	Chungju, Chupungryong, Cheongju, Seosan
Gyeongsangdo	Ulsan, Pohang, Andong, Sangju, Changwon
Jeollado	Jeonju, Gunsan,
Metropolitan city	Daejeon, Daegu, Incheon, Seoul, Gwangju
Island	Baengnyeongdo, Ulleungdo

2.2. 기계학습법

서리가 발생한 날과 발생하지 않는 날의 구분하는 모형을 구동하기 위하여 통계프로그램 R(3.3.1)을 사용하였다. 인공신경망은 “nnet” 패키지의 nnet 함수를 사용하였고, 은닉층의 노드 수는 3개, 활성화함수는 로지스틱함수, 훈련 시 최적화를 위한 반복횟수는 200회로 설정하여 수행하였다. 랜덤포레스트는 “randomForest” 패키지의 randomforest 함수를 사용하였고, 트리 갯수를 500으로 설정하여 수행하였다. 서포트벡터머신의 경우는 “kernlab” 패키지의 ksvm 함수를 사용하였으며, 커널함수는 선형함수(vanilladot)를 사용하였다.

2.2.1. 인공신경망

인공신경망은 인간의 두뇌가 학습하는 과정을 모방해서 만들어진 모형이다. 일반적인 구조는 입력층, 은닉층, 출력층의 3가지 층으로 구성되어 있다. 입력층은 단순히 변수값을 입력받으며, 은닉층은 입력층으로 전달된 값들을 결합해서 다른 은닉층이나 출력층에 전달한다. 그리고 출력층은 최종적으로 우리가 알고자 하는 결과값을 출력해 준다.

인공신경망의 입력층과 은닉층에 들어온 값(x_i)들은 가중치(w_i)와 결합되어 합성함수(식 1)를 통해 계산되고, 그 결과값은 활성화함수(식 2)를 통해 변환된다 (Smith *et al.*, 2007; Yong *et al.*, 2007; Koo *et al.*, 2013; Oh, 2013).

$$u_i = \sum_{j=1}^n w_j x_j + b \quad (1)$$

$$y_i = f(u_i) \quad (2)$$

활성함수는 로지스틱(logistic) 함수와 하이퍼볼릭 탄젠트(hyperbolic tangent) 함수, 가우스(gauss) 함수, 선형(linear) 함수, 문턱값(threshod) 등이 사용된다.

가중치(w_i)는 목적함수(식 3)의 실측값과 신경망의 출력값 사이의 차이가 최소가 되도록 갱신되며, 최소화를 수행하기 위해서 경사 하강법의 일종인 역전파(back propagation) 알고리즘이 주로 사용되고 있다.

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

2.2.2. 랜덤포레스트

랜덤포레스트는 다수의 의사결정 트리 모형을 결합

하여 최적의 모형을 형성하는 앙상블기법이다. 의사결정 트리모형은 각 노드에 속하는 자료가 분할하면서 자라나게 되는데, 상위노드에서 하위노드로 가지분할을 수행할 때 오류율을 나타내는 불순도함수를 이용하여 분류변수와 분류기준값을 선택한다. 불순도함수는 식 (4)의 지니 지수(gini index)와 식 (5)의 엔트로피 지수(entropy index) 등이 주로 사용되며, 선택된 기준에 의해 분할이 일어날 때 순수도가 높아지는 쪽으로 재귀적(recursively)으로 분할이 이루어진다. 일반적인 의사결정 트리 모형은 각 노드마다 모든 예측변수 안에서 최적의 분할을 선택하는 방법을 사용하지만, 랜덤포레스트는 예측변수들을 임의로 추출하고 추출된 변수 내에서 최적의 분할을 수행하는 방법으로 다수의 의사결정 트리모형을 생성해 나간다(Yoo, 2015; Yong *et al.*, 2007).

$$G(x) = \sum_{i=1}^n \hat{p}(x_i)(1 - \hat{p}(x_i)) \quad (4)$$

$$E(x) = - \sum_{i=1}^n \hat{p}(x_i) \log \hat{p}(x_i) \quad (5)$$

2.2.3. 서포트벡터머신

다른 모형들은 주로 분류 범주들을 분리할 때 오류율을 최소화하도록 계산되지만, 서포트벡터머신은 분리간격이 최대화하도록 다차원 공간을 구분하는 초평면(hyperplane)이 형성되기 때문에 다른 모형들에 비해 일반화 능력이 우수하다고 알려져 있다.

서포트벡터머신의 초평면은 식 (6)과 같이 나타낸다(Chen *et al.*, 2010; Yong *et al.*, 2007; Oh, 2013; Lee, 2016).

$$w^t x_i + b = 0 \quad (6)$$

여기서, b 는 바이어스이며, w 는 개체 x_i 에 대한 가중치다.

만약 두 부류로 나뉘는 데이터들을 식 (7)과 같이 분류가 된다면

$$\begin{cases} w^t x_i + b \geq 1, & \text{if } y_i = 1 \\ w^t x_i + b \leq -1, & \text{if } y_i = -1 \end{cases} \quad (7)$$

위 조건을 만족하는 초평면은 식 (8)로 나타낼 수

있다.

$$\begin{cases} w^t x_i + b = -1 \\ w^t x_i + b = 1 \end{cases} \quad (8)$$

이때 식 (8)의 중심이 되는 초평면이 식 (6)이 되며, 식 (8)의 두 초평면으로부터 식 (6)의 초평면까지의 유클리드 거리는 각각 $\frac{1}{\|w\|}$ 이며, 이 거리가 분리간격의 넓이가 된다. 분리간격이 최대화되고 분류오차 최소화되는 것을 표현한 조건식이 식 (9)이며, 식 (9)가 최소화되도록 초평면 식 (6)이 구해 지게 된다.

$$\min J(w, b, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad (9)$$

$$\text{Subject to } \begin{cases} w^t x_i + b \geq 1 - \xi_i, \text{ if } y_i = 1 \\ w^t x_i + b \leq -1 + \xi_i, \text{ if } y_i = -1 \\ \xi_i \geq 0 \end{cases}$$

식 (9)의 마진오차 ξ_i 는 오분류한 점들에 대해서는 양의 수가 되고 올바르게 분류된 자료들에 대해서는 0을 갖게 된다.

III. 결 과

서리가 발생한 날과 서리가 발생하지 않은 날의 기상특징을 비교한 결과 Table 2와 같았다.

기상요소별로 평균을 분석한 결과 최저기온은 서리가 발생한 날이 1.0°C로 서리가 발생하지 않은 날의 3.2°C보다 2.2°C가 낮았으며, 서리가 발생한 날에 0°C

이하인 날은 30%정도로 나타났다. 평균풍속은 서리가 발생한 날은 1.7m/s로 서리가 발생하지 않은 날의 2.6m/s보다 0.9m/s 약했으며, 서리가 발생한 날에는 2.0m/s이하의 풍속이 71%로 나타났다. 구름량은 서리가 발생한 날이 20.1%로 서리가 발생하지 않은 날의 23.1%보다 3.0% 낮았으며, 서리가 발생한 날에는 20% 이하인 날이 58%를 나타냈다. 서리가 발생한 날을 기준으로 5일간 강수량은 7.5mm로 서리가 발생하지 않은 날을 기준일 때의 8.5mm보다 1mm 적었다. 3일간 강수량은 서리가 발생한 날 기준으로 3.3mm로 서리가 발생하지 않은 날의 5.0mm보다 1.7mm 적었다. 상대습도는 서리가 발생한 날은 59.3%로 서리가 발생하지 않은 날의 56.1%보다 3.3% 높았다. 이슬점온도는 서리가 발생한 날이 -1.1°C로 서리가 발생하지 않은 날의 -1.3°C보다 0.2°C 높았다.

서리가 발생한 날의 기상인자와 서리가 발생하지 않은 날의 기상인자 간에 T-test를 실시한 결과 유의수준 0.05에서 최저기온과 평균풍속, 상대습도는 유의한 결과가 나타났지만, 구름량, 5일간 강수량과 3일간 강수량, 이슬점온도는 유의하지 않은 것으로 나타났다. 이것은 최저기온과 평균풍속, 상대습도는 서리가 발생한 날과 서리가 발생하지 않은 날의 기상값이 통계적으로 차이가 있다는 것을 나타낸다.

로지스틱회귀분석과 판별분석을 통해 변수를 선정한 결과는 두 방법 모두 최저기온, 평균풍속, 상대습도, 이슬점온도를 선정되었다(Table 3). 비록 T-test에서 이슬점온도가 서리가 발생한 날과 서리가 발생하지 않은 날에 차이가 나타나지 않았지만, 다른 기상인자 함께 고려한다면 의미가 있을 것으로 판단되어 서리 발생을 구분하는 모형 구동할 때 최저기온, 평균풍속, 상대습도, 이슬점온도를 사용하였다.

4가지 기상인자를 이용하여 인공신경망, 랜덤포레

Table 2. Characteristic of climate factors of frost occurrence days and frost free days

		Tmin (°C)	WS (m/s)	Cloud (%)	Pre5 (mm)	Pre3 (mm)	RH (%)	DP (°C)
FOD	Average	1.0	1.7	20.1	7.5	3.3	59.3	-1.1
	(SD)	(2.3)	(0.9)	(20.9)	(13.2)	(9.2)	(11.1)	(4.2)
FFD	Average	3.2	2.6	23.1	8.5	5.0	56.1	-1.3
	(SD)	(2.8)	(1.2)	(22.7)	(13.6)	(9.9)	(12.9)	(5.1)

Tmin: Minimum temperature, WS: Average wind speed, Cloud: Amount of cloud, Pre5: Amount of precipitation within 5 days, Pre3: Amount of precipitation within 3 days, RH: Average relative humidity, DP: Average Dew point temperature, SD: Standard deviation, FOD: Frost occurrence day, FFD: Frost free day

Table 3. Selected Independent variables by statistical methods for modeling

Method	Independent variable
T-test	Minimum temperature, Average wind speed, Relative humidity,
LR	Minimum temperature, Average wind speed, Relative humidity, Dew point temperature
DA	Minimum temperature, Average wind speed, Relative humidity, Dew point temperature

LR: Logistic Regression, DA: Discriminant Analysis

스트, 서포트벡터머신에 대한 서리 예측 모형을 구축하였고, Han *et al.*(2009)에서 제시한 전날의 24시간온, 구름량, 5일간 강수량을 이용한 판별분석 모형(Table 4)과 함께 정확도(accuracy)와 민감도(sensitivity)를 비교하였다. 정확도는 전반적으로 모든 경우 중에서 서리가 발생하는 것과 서리가 발생하지 않는 것을 올바르게 추정했는지를 비교할 수 있으며, 민감도에서는 서리의 발생 모형의 서리의 발생을 알리는 것이 더 중요하기 때문에 실제 서리가 발생한 경우에서 올바르게 서리가 발생하는 것을 추정했는지를 비교하는 것이다. 정확도에서는 인공신경망은 검증데이터 170일 중 127일을 올바르게 분류하여 74.7%로 가장 높았으며, 랜덤포레스트는 74.1%, 서포트벡터머신은 72.9%, 판별분석 모형은 59.4%로 나타났다. 반면, 민감도에서는 판별분석 모형이 85일 중 70일을 올바르게 분류하여 82.4%로 가장 높았으며, 랜덤포레스트가 74.1%, 인공신경망이 71.8%, 서포트벡터머신이 68.2%를 나타냈다(Table 5).

Table 4. Coefficients of linear discriminant function for frost prediction (Han *et al.*, 2009)

Parameters	Coefficients of functions	
	FOD	FFD
Amount of cloud(0.1%)	0.08217	0.25916
Temperature at 24:00(°C)	0.80399	1.88597
Amount of rainfall within 5 days(mm)	-0.05519	-0.06492
Constant	-3.01817	-10.77937

FOD: Frost Occurrence Day, FFD: Frost Free Day

이러한 결과에서 기존 판별분석 모형과 본 연구에서 사용된 모형을 정확도면에서 비교했을 때는 전반적으로 서리 발생 유무를 구분하는 능력은 기계학습법에 의한 모형이 높게 나타나 그 활용 가능성이 있는 것으로 판단되지만, 민감도면에서 판별분석 모형이 더 높게 나타나기 때문에 기계학습법을 이용한 모형과 판별분석 모형 결과를 함께 고려할 수 있는 방법의 모색이 필요할 것으로 판단된다.

IV. 결 론

서리 발생 예측을 위해 도출된 기상인자는 최저기온, 평균풍속, 상대습도, 이슬점온도로 나타났다. 이러한 기상 인자들의 특징은 서리가 발생한 날이 서리가 발생하지 않은 날에 비해 최저기온, 이슬점온도, 평균풍속은 낮게 나타났고 상대습도는 높게 나타났다. 이러한 결과는 서리가 발생하기 위해서는 공기 중에 어느 정도 수증기가 포함되어 있는 상태에서 기온이 낮아지면서 서리가 발생하는 상황을 반영한 것으로 판단

Table 5. The accuracy/sensitivity comparison of frost occurrence classification estimation of models

Prediction		Observation		Total	
		FOD	FFD	Accuracy	Sensitivity
ANN	FOD	61	19	127/170 (74.7%)	61/85 (71.8%)
	FFD	24	66		
RF	FOD	63	22	126./170 (74.1%)	63/85 (74.1%)
	FFD	22	63		
SVM	FOD	58	19	124/170 (72.9%)	58/85 (68.2%)
	FFD	27	66		
DA	FOD	70	54	101/170 (59.4%)	70/85 (82.4%)
	FFD	15	31		

ANN: Artificial Neural Network, RF: Random Forest, SVM: Support Vector Machine, DA: Discriminant Analysis, FOD: Frost Occurrence Day, FFD: Frost Free Day

된다. 그렇지만, 일부 자료에서 서리가 발생한 날에 최저기온이 0°C 이상에서도 서리가 발생하는 것으로 나타났는데, 이것은 온도계가 설치되어 있는 위치의 온도보다 지면 부근의 온도가 더 낮았기 때문에 지면 부근에 서리가 발생할 것으로 예상되며, 추후 연구에서는 지면온도나 초상온도에 대한 분석이 함께 이루어진다면 서리 발생 예측 모형의 정확도가 향상될 것으로 예상된다.

서리 발생 예측 모형을 비교한 결과에서 기존 판별 분석 모형의 정확도가 낮게 나타난 이유는 나주 지역을 대상으로 연구가 진행되어 전국적인 서리 발생과 관련된 기상 환경을 반영하기에 제한적이기 때문일 것으로 예상된다. 그 예로서, 기존 판별분석 모형에서는 5일간 강수량이 서리 발생에 많은 영향을 미치는 것으로 판단되었지만 이번 연구에서는 강수량의 영향이 크지 않게 나타났기 때문에 전국의 강수량 데이터를 적용할 경우 기존 판별분석 모형의 정확도가 낮아 질 수 있을 것으로 예상된다. 또한, 기존 판별모형에서는 전날의 24시 이후에 서리 발생에 대한 예측이 이루어진다는 면에서 농민이 서리 발생에 미리 대처하기 힘든 부분이 있다고 판단된다. 이런 점에서 기상청의 동네예보를 통해 예측한 기상을 적용할 수 있다는 점에서 본 연구의 모형이 활용 가능성이 있다고 판단된다.

그렇지만, 본 연구에 사용된 모형은 기상청의 동네예보를 사용한다는 전제로 만들어졌기 때문에 기상청 동네예보의 오차가 발생한다면 모형의 정확도 또한 그 오차만큼 낮아질 것으로 예상되고, 민감도에서 기존 판별분석 모형이 더 높게 나타났기 때문에 기계학습법에 의한 모형과 기존 판별분석 모형을 함께 사용하면 더 활용성이 높아질 것으로 예상된다.

적 요

본 연구에서는 기상청 예보자료를 이용할 것을 전제로 서리가 발생하는 날과 서리가 발생하지 않는 날을 구분하는 모형을 구축하였다. 서리가 발생한 날과 서리가 발생하지 않은 날을 구분할 수 있는 기상인자로서 최저기온, 평균풍속, 평균상대습도, 평균이슬점온도로 나타났다. 기상인자별로 두 날을 비교한 결과 서리가 발생한 날이 서리가 발생하지 않은 날에 비해 최저기온, 이슬점온도, 평균풍속은 낮게 나타났고 상대습도는 높게 나타났다. 이러한 기상인자로 인공신경망, 랜덤포레스트, 서포트벡터 머신의 기계학습법을

이용한 모형을 연구한 결과 70%이상의 정확도를 나타내어 활용 가능성이 있을 것으로 판단된다.

감사의 글

본 논문은 농촌진흥청 국립농업과학원 농업과학기술 연구개발사업(과제번호: PJ01000702)의 지원에 의해 이루어진 것임.

REFERENCES

- Abhishek, K., M. P. Singh, S. Ghosh, and A. Anand, 2012: Weather forecasting model using artificial neural network. *Procedia Technology* **4**, 311-318.
- Chen, J. L., H. B. Liu, W. Wu, and D. T. Xie, 2010: Estimation of monthly solar radiation from measured temperatures using support vector machines - A case study. *Renewable Energy* **36**(2), 413-420.
- Han, J. H., J. J. Choi., U. Chung., K. S. Cho, and J. P. Chun, 2009: Frostfall forecasting in the Naju pear production area based on discriminant analysis of climatic data. *Korean Journal of Agricultural and Forest Meteorology* **11**(4), 135-142.
- Chung, U., H. C. Seo, and J. I. Yun, 2004: Site-specific frost warning based on topoclimatic estimation of daily minimum temperature. *Korean Journal of Agricultural and Forest Meteorology* **6**(3), 164-169.
- Kim, S. S., H. J. Kim, S. S. Chung, and Y. K. Lee, 2014; *Multivariate Data Analysis with R*. Korea National Open University Press, 191-210.
- Koo, J. Y., H. J. Park, D. W. Choi, and S. S. Kim, 2013: *Data mining*. Korea National Open University Press, 211-248 and 261-279.
- Kwon, Y. A., H. S. Lee., W. T. Kwon., and K. O. Boo, 2008: The weather characteristics of frost occurrence days for protecting crops against frost damage. *Journal of the Korean Geographical Society* **43**(6), 824-842.
- Lee, J. G., 2016: *R Program Recipes for Multi-Variate Analysis & Data Mining*. Bullsbook, Seoul, 358pp.
- Naing, W. Y. N. and Z. Z. Htike, 2015: Forecasting of monthly temperature variations using random forests. *ARNP journal of Engineering and Applied Sciences* **10**(21).
- Oh, I. S., 2013: *Pattern Recognition*, Kyobo Book Centre, Seoul, 95-132, and 137-170.
- Robinson, C. and N. Mort, 1996: A neural network solution to the problem of frost Prediction. *UKACC International Conference on Control. Control '96*,

- 136-139.
- Shank, D. B., G. Hoogenboom, and R. W. McClendon, 2008: Dewpoint Temperature Prediction Using Artificial Neural Networks. *Journal of Applied Meteorology and Climatology* **47**(6), 1757-1769.
- Smith, B. A., R. W. McClendon, and G. Hoogenboom, 2007: Improving Air Temperature Prediction with Artificial Neural Networks. *International Journal of Computational Intelligence* **3**, 180-186.
- Temeyer, B. R., W. A. Gallus Jr, K. A. Jungbluth, D. Burkheimer, and D. McCauley, 2003: Using an artificial neural network to predict parameters for frost deposition on Iowa bridgeways. *Proceedings of the 2003 Mid-Continent Transportation Research Symposium*, Iowa State University, Ames, IA, 2003.
- Yoo, J. E., 2015: Random forests, an alternative data mining technique to decision tree. *Journal of Educational Evaluation* **28**(2), 427-448.
- Yong, H. S., Y. Nah, J. S. Park, H. W. Seung, M. Lee, S. Lee, and L. Choi, 2007: Data Mining, NFINITYBOOKS, 241-270 and 283-286.