

지지벡토르기를 리용한 서리발생예보방법

박 미 향

위대한 령도자 김정일동지께서는 다음과 같이 교시하시였다.

《기상수문부문에서 현대적인 기상관측설비들을 갖추어놓고 큰물예보와 기상예보를 정확히 하도록 하며 기상수문자료의 신속정확성과 과학성을 높이도록 하여야 합니다.》
(《김정일선집》 증보판 제23권 331페이지)

서리는 농작물의 생육을 억제하는것을 비롯하여 농업부문에 여러가지 영향을 미치므로 서리예보를 잘하는것은 중요한 문제로 나선다.

론문에서는 농작물의 생육에 피해를 주는 서리현상에 대한 예보를 실현하기 위한 지지벡토르기의 일반적원리와 그에 기초한 서리예보결과들에 대하여 서술하였다.

1. 지지벡토르기에 의한 서리예보방법

1) 예보인자

서리는 얼음점이하로 내려간 지물에 수증기가 승화되어 생긴 얼음결정이다. 따라서 일최저기온과 이슬점온도, 상대습도 등과 같은 기상요소들이 서리발생에 직접적으로 또는 간접적으로 영향을 준다. 특히 서리는 바람이 없는 고요한 날에 밤복사랭각이 심할 때 즉 일최저기온이 심하게 내려가 령하로 될 때 나타난다.

론문에서는 일최저기온 T_{\min} , 지면기온 T_s , 이슬점차 T_d , 서리발생전날의 최고기온 T_{\max} , 대기의 포화상태를 반영하는 기상요소인 상대습도 RH 를 서리발생예보를 위한 예보인자로 정하였다. 여기서 T_s , T_d 와 RH 는 03시 관측자료들, T_{\min} 은 24h전에 예보한 최저기온자료를 리용한다.

교지점에서 선택된 예보인자들이 서리발생에 주는 영향을 평가하기 위하여 일최저기온과 다른 예보인자들사이의 상관결수(표 1)를 계산하고 5%유의수준에서 유의성검증을 진행한 결과 우에서 언급한 5개의 예보인자들의 상관결수가 모두 유의하다는것이 검증되였다.

표 1. 일최저기온과 다른 예보인자들사이의 상관결수

일최저기온	예보인자			
	전날 최고기온	상대습도	이슬점차	지면기온
	0.91	0.31	0.28	0.96

표 1에서 보는바와 같이 일최저기온에 가장 큰 영향을 주는 인자가 바로 전날 최고기온과 지면기온이며 상대습도, 이슬점차도 무시할수 없는 예보인자라는것을 알수 있다.

2) 지지벡토르기를 리용한 서리예보방법

1960년대에 통계적학습리론에 기초한 지지벡토르기리론이 처음으로 소개되였다. 그 기본작상은 핵함수에 의하여 최대여백을 가진 최량초평면을 찾는것이다.[1]

선형분리가능한 문제에서는 최량초평면을 찾는것은 어렵지 않지만 선형분리불가능한 경우는 여백이 최대인 초평면을 구하는것이 매우 어렵다.

선형적으로 완전히 분리가가능한 학습자료가 다음과 같이 주어졌다고 하자.

$$(\mathbf{X}_1, y_1), \dots, (\mathbf{X}_k, y_k), \mathbf{X} \in R^n, y \in \{1, -1\} \quad (1)$$

여기서 \mathbf{X} 는 입력벡토르, y 는 출력값, n 은 입력공간의 차원수(예보인자의 개수), k 는 학습자료의 개수이다.

선형분리불가능한 문제 즉 비선형문제에 대하여서는 표본 X 를 기타 다른 고차원징표공간 H 로 넘긴 다음 공간 H 에서 선형분리가능한 문제로 고찰할수 있다. 즉 넘기기 $X \rightarrow \Phi(X) = (\phi(X_1), \phi(X_2), \dots, \phi(X_i), \dots)$ 을 실시하여 징표공간에서 선형분리가능한 문제로 취급한다. 여기서 $\phi(X_i)$ 는 실함수이다.

최량초평면을 구하자면 손실함수의 최소화에 관한 최량화문제가 제기된다. 이 최량화문제를 풀기 위하여 라그랑주승수를 리용하면 지지벡토르들은 정의 라그랑주승수에 대응하는 벡토르(학습자료)로 되는데 지지벡토르를 결정하자면 라그랑주승수를 계산하여야 한다. 라그랑주승수를 계산하기 위하여 다음의 2차계획문제를 푼다.[1, 2]

$$\max_{\alpha} W(\alpha) = \sum_{i=1}^k \alpha_i - \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^k \alpha_i \alpha_j y_i y_j (\Phi(\mathbf{X}_i) \cdot \Phi(\mathbf{X}_j)) \quad (2)$$

여기서 $\alpha_i (i=1, \dots, k)$ 는 라그랑주승수로서 부아닌 수이다.

식 (2)로부터 알수 있는바와 같이 목적함수는 징표공간에서의 스칼라적 $\Phi(\mathbf{X}_i) \cdot \Phi(\mathbf{X}_j)$ 에 의하여 결정된다. 따라서 최량분류실현에 적합한 스칼라적함수 $K(\mathbf{X}_i, \mathbf{X}_j) = \Phi(\mathbf{X}_i) \cdot \Phi(\mathbf{X}_j)$ 를 리용하면 일정한 비선형변환을 진행하여 선형분류를 실현할수 있다. 이때 식별함수는 다음과 같다.

$$f(\mathbf{X}) = \text{sgn} \left(\sum_{i=1}^{SV} y_i \alpha_i^0 K(\mathbf{X}_i, \mathbf{X}) - b_0 \right) \quad (3)$$

이상에서 보는바와 같이 지지벡토르기리론에서는 입력공간의 자료를 징표공간으로 넘겨서 분류문제를 취급하는데 이때 핵기술을 리용하면 징표공간을 구체적으로 다룰 필요는 없으며 다만 징표공간에서 스칼라적을 정의하는 핵만을 적당히 선택하면 된다.[1]

지지벡토르기방법에서는 스칼라적에 대한 기본함수에 따라 계산방법이 서로 다르게 된다. 여기에서 흔히 쓰이는 기본함수들은 다항식함수, 방사토대함수, 다층퍼셉트론함수, 동적함수 등이다.[1]

우리는 식 (4)로 정의되는 방사토대함수를 핵함수로 리용하였다.

$$K(X_i, X_j) = \exp(-\gamma \|X_i - X_j\|^2) \quad (4)$$

예보인자는 우에서 언급한 기상요소들을 리용하였으며 계산은 MATLAB 8.5.0에서 진행하였다.

지지벡토르기를 리용한 서리예보모형을 식으로 표시하면 다음과 같다.

$$y = f_{SVM}(T_{\min}, T_s, T_d, T_{\max}, RH) \quad (5)$$

여기서 y 는 서리가 나타나면 1, 서리가 나타나지 않으면 -1을 가진다.

2. 적용실례

1) 성능평가지표

예보성적은 관측자료와 예보자료에 기초하여 맞힘률, 공보률과 같은 여러가지 평가지표를 가지고 평가할수 있다. 여기서는 일기예보성적평가에서 널리 알려진 맞힘률(*HIR*)과 공보률(*FAR*), 성공지수(*CSI*), 일반성적(*PC*)과 오보확률(*POFD*)을 리용하여 개발된 모형의 예보성적을 평가한다.

성적평가지수들은 다음과 같은 식들에 의하여 계산된다.[3]

$$HIR = \frac{A}{A+C} \times 100 \quad (6)$$

$$FAR = \frac{B}{A+B} \times 100 \quad (7)$$

$$CSI = \frac{A}{A+B+C} \times 100 \quad (8)$$

$$PC = \frac{A+D}{A+B+C+D} \times 100 \quad (9)$$

$$POFD = \frac{B}{B+D} \times 100 \quad (10)$$

지수들을 계산하기 위한 변수 *A*, *B*, *C*, *D*의 의미는 표 2와 같다.

식 (6)–(10)으로부터 *HIR*, *CSI*, *PC*값이 각각 크면 클수록(최대로 100), *FAR*와 *POFD*값이 작으면 작을수록(최소 0) 개발된 예보모형은 더 좋은 예보결과를 준다는것을 알수 있다.

표 2. 변수 *A*, *B*, *C*, *D*의 의미

예보	관측	
	있다.	없다.
있다.	<i>A</i>	<i>B</i>
없다.	<i>C</i>	<i>D</i>

2) 표지방의 서리예보결과

표 3. 학습과 검증기간 서리발생수

기간	서리일수	비서리일수	학습/검증
2004년–2013년	283	1 527	학습
2014년–2015년	51	311	검증

이 기간 표지방에서 334회의 서리현상이 관측되었다. 모형작성을 위하여 2004년–2013년까지의 자료가 학습자료로 리용되었으며 2014년–2015년기간의 자료는 검증에 리용되었다. 학습자료와 독립자료에 대한 서리예보결과에 대한 통계자료를 표 4에 제시하였다.

표 4에서 보는바와 같이 학습자료에 대한 *HIR*와 *CSI*가 다 검증자료에 대한 성적보다 높다는것을 알수 있다.

한편 공보률(*FAR*)에서는 학습자

지지벡토르기에 기초한 서리예보모형을 작성하기 위하여 2004년–2015년까지 봄철(3월–5월)과 가을철(9월–11월)기간의 서리관측자료가 리용되었다. 표 3에서 보는바와 같이

표 4. 학습자료와 검증자료에 대한 통계자료

성적지수	학습자료	검증자료
<i>HIR</i>	64	58.8
<i>FAR</i>	25.8	28.5
<i>CSI</i>	52.3	47.6
<i>PC</i>	90.9	90.9
<i>POFD</i>	4.1	3.9

료성적이 검증자료보다 작은 값을 가지지만 오보확률(*POFD*)은 학습자료성적이 검증자료 성적보다 더 큰 값을 가진다. 그리고 일반예보성적(*PC*)은 학습자료와 검증자료에 대하여 정확도가 90.9%로서 비교적 높은 성적으로 지지벡토르기를 리용한 서리예보모형이 봄철과 가을철서리현상예보에 적용될수 있다는것을 암시해준다.

맺 는 말

앞으로 서리예보인자로서 우에서 리용한 5개의 인자외에도 바람속도와 밤길이 등 서리현상에 영향을 미치는 가능한 인자들을 다 리용한다면 예보정확성을 더 높일수 있을것이다.

참 고 문 헌

- [1] 김승남 등; 지식발견, 김책공업종합대학출판사, 186~220, 주체99(2010).
- [2] S. M. Mousavia et al.; Robotics and Computer-Integrated Manufacturing, 29, 157, 2013.
- [3] Luca Ghielmi et al.; Computers and Electronics in Agriculture, 54, 101, 2006.

주체107(2018)년 7월 5일 원고접수

A Method of Frost Occurrence Forecast Using Support Vector Machine

Pak Mi Hyang

This paper shows the general principle of Support Vector Machine of forecasting frost occurrence which led to damage of growth and development of the crops and results of frost forecasting based on it.

Key words: Support Vector Machine, frost, minimum air temperature