

순서형 대설 예보를 위한 통계 모형 개발¹⁾

손건태²⁾, 이정형³⁾, 류찬수⁴⁾

요약

호남지역에 대한 대설특보 예보를 위한 통계모형 개발을 수행하였다. 일 신적설량에 따라 세 범주(0: 비발생, 1: 대설주의보, 2: 대설경보)로 구분되는 순서형 자료 형태를 지니고 있다. 두 가지 통계 모형(다등급 로지스틱 회귀모형, 신경회로망 모형)을 고려하였으며, 수치모델 출력 자료를 이용한 역학-통계모형 기법의 하나인 MOS(model output statistics)를 적용하여 축적된 수치모델 예보자료와 관측치의 관계를 통계모형식으로 추정하여 예측모형을 개발하였다. 군집분석을 사용하여 훈련자료와 검증자료를 구분하였으며, 예보치 생성을 위하여 문턱치를 고려하였다.

주요 용어: 순서형 대설특보 예보, MOS, 다등급 로지스틱 회귀모형, 신경회로망 모형, 문턱치

1. 서론

지구온난화를 포함한 다양한 원인들로 해수면 상승을 비롯한 해양 에너지 증가, 사막화, 열대 폭풍의 대형화 등으로 전세계적으로 악기상과 기상이변으로 인한 피해규모가 급증하고 있다. 우리나라 경우에 기상과 관련하여 1993년부터 2002년까지 10년간의 재산피해 중 91.7%가 태풍과 집중호우에 의한 피해이며, 7.6%가 대설에 의한 피해로 보고되고 있다. 2002년과 2003년 여름철 태풍 루사와 매미의 피해는 기록적인 것이었으며, 2004년 3월 충청·강원지역의 대설과 2005년 3월 영동·영남지역의 대설은 교통두절 및 막대한 재산피해를 가져오는 등 점차 기상피해의 정도가 커져가고 있다. 기상피해를 최소화시키기 위하여 국지적인 기상특성과 기상재해와의 관계를 연구하여 예보의 정확도를 향상시키는 것이 무엇보다 중요한 단계가 된다.

강수(강우 및 강설)예보는 (1) 강수량 예보(quantitative precipitation forecast), (2) 강수발생 확률예보(probability of precipitation forecast), (3) 등급별 강수예보(probability of classified precipitation forecast)가 있다.

본 연구는 대설예측을 위하여 세범주로 구분한 순서형 반응변수(ordinary response variable)에 대한 예측 기법을 적용하여 한후기(11월~3월)에 강설이 많은 호남지방에서의 대설 예측을 위한 통계적 모형을 개발하는데 목적을 두고 있다. 기상청에서 사용하는 수치모델 RDAPS (Regional Data Assimilation and Prediction System) 출력자료를 이용한 객관적 해석법으로 역학-통계모형화 기법의 하나인 MOS (model output statistics)를 적용하여, 축적된 수치모델 예보자료와 관측치의 관계를 통계모형식으로 추정하여 예측모형을 개발하고자 하였다. 예측모형

- 1) 본 연구는 기상청에서 시행하는 기상지진기술개발사업의 하나인 “국지기상예측기술개발” 2005년 과제 지원으로 수행되었다.
- 2) 부산대학교 통계학과 교수, 609-735 부산 금정구 장전동, ktsohn@pusan.ac.kr
- 3) 동아대학교 경영정보과학부 부교수, 604-714 부산 사하구 하단동
- 4) 조선대학교 과학교육학부 교수, 501-759 광주 동구 서석동

으로는 순서형 범주자료에 대한 다등급 로지스틱 회귀모형과 신경회로망을 적용하였다. 모형 개발은 SAS/E-Miner를 활용하였으며 관측과 예측의 결과로 이루어지는 3×3 결합빈도표를 비교하여 최적 문턱치를 구한 후 예보치를 생성하였다.

2. 자료

2.1 예측량과 예측인자들

호남지역 17개 지점(군산, 전주, 광주, 목포, 여수, 흑산도, 완도, 진도, 부안, 임실, 정읍, 남원, 장수, 순천, 장흥, 해남, 고흥)에 대한 관측치(일 신적설량)와 RDAPS 예측인자들을 수집하여 모형개발에 사용하였다. 잠재적 예측인자로는 조주영·최준태(2002)와 Sohn *et al.*(2005a, 2005b)에서 사용된 45종류의 RDAPS 출력값(Table 1)을 고려하였다. 연구를 위하여 2002년 1월 1일부터 2005년 3월 31일까지 8,194 경우의 한후기 자료를 사용하였으며, 예측량인 강설량은 일 신적설량이며, 대설특보 기준(대설주의보: 일 신적설량이 50mm미만, 대설경보: 200mm 이상)을 참조하여 일 신적설량이 50mm미만, 50mm~150mm, 150mm이상의 세 범주로 구분하여 예측모형개발을 수행하였다. 200mm이상 경우가 한 건밖에 없어 150mm 이상으로 범주를 정하였다.

Table 1. Potential Predictors

Symbol	Predictors
E850, E700, E500 S850, SE700, S500 NW850, NW700, NW500 NE850, NE700, NE500 VV850, VV700, VV500	East wind speed on 850 hPa, 700 hPa and 500 hPa North wind speed on 850 hPa, 700 hPa and 500 hPa North-east wind speed on 850 hPa, 700 hPa and 500 hPa North-west wind speed on 850 hPa, 700 hPa and 500 hPa Wind speed on 850 hPa, 700 hPa and 500 hPa
VOR850, VOR700, VOR500	Relative vorticity on 850 hPa, 700 hPa and 500 hPa
QAD850, QAD700 Q84 Q74 TAD850, TAD700 RH850, RH700, RH500 CCL DWL PCWT CTOP CBAS BBX1 BBX2	Advection of specific humidity at 850 hPa and 700 hPa Difference of specific humidity at 850 hPa and 700 hPa Difference of specific humidity at 700 hPa and 700 hPa Thermal advection on 850 hPa and 700 hPa RH on 850 hPa, 700 hPa and 500 hPa Convective condensation level Depth of wet level Potential precipitation Level of cloud top Level of cloud base Black box index 1 Black box index 2
SSI KYID KIDX LR87 LR85	Showalt stability index KY index K index lapse rate between 850 hPa and 700 hPa lapse rate between 850 hPa and 500 hPa
T850, T700, T500 ET850, ET700 ET87	Temperature on 850 hPa, 700hPa and 500 hPa Equivalent potential temperature at 850 hPa and 700 hPa Difference of equivalent potential temperature at 850 hPa and 700 hPa

2.2 군집분석

강설은 한후기에서 발생하므로 한후기 기간의 일신적설량을 사용하여 17개 지점에 대하여 군집분석을 수행하였다. Figure 1에 나타난 군집의 결과를 기초로 모형훈련자료와 모형검증자료를 구분하였다. 군집분석의 결과훈련자료로 사용되는 지점은 {광주, 흑산도, 완도, 진도, 부안, 순천, 해남, 고흥, 남원, 임실}의 10개 지점이며, 검증자료로 사용되는 지점은 {군산, 전주, 장수, 정읍, 목포, 여수, 장흥}의 7개 지점으로 선정하였다.

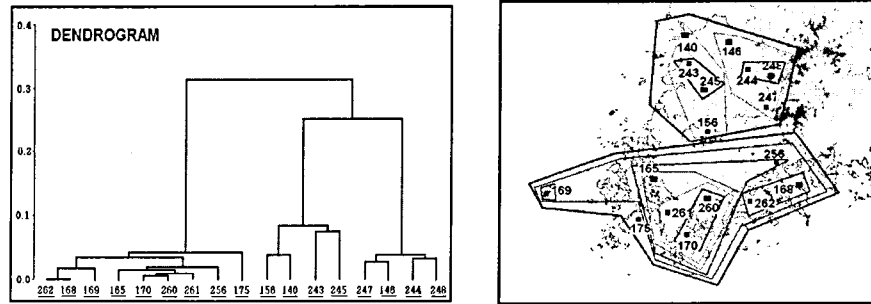


Figure 1. Dendrogram(left) and Clusters(right)

Table 2. Frequency Table of Predictand

Category	value	FREQ	%	Training	Test
(50 mm 미만)	0	8067	98.45	4745	3322
(50 mm ~ 150 mm)	1	114	1.39	70	44
(150 mm 이상)	2	13	0.16	5	8
		8194		4820	3374

2.3 예측모델 개발 전략

세범주로 이루어진 순서형 자료로 이루어진 대설 예측을 위하여 다등급 로지스틱회귀모형 (multi-grade logistic regression model)과 신경회로망(neural networks)을 적용하였다. 세 범주로 이루어진 순서형 반응자료에 대한 다등급 로지스틱회귀모형식은 다음과 같으며, 45개 설명 변수 중 유의변수를 선택하기 위하여 단계별 회귀방법을 적용하였다.

$$\log\left(\frac{P(Y \leq j)}{1 - P(Y \leq j)}\right) = b_j + \beta'x, j = 1, 2$$

고려된 신경회로망 구조는 Multilayer perceptron으로 Table 1의 45개 잠재적 예측인자들을 입력으로, 하나의 은닉층, 세개의 출력값으로 이루어진다. 은닉층의 노드 수는 식별통계량 AIC(Akaike Information Criterion)를 최소로 하는 노드수로 결정하였으며, 최적모형 선택기준은 평균오차(Average Error)로 하였다. 훈련자료로 모수 추정이 수행되었으며, 표준화된 입력자료 ⇨ 정보전달함수(LBF) ⇨ 정보활성화 함수(logistic function) ⇨ (출력층: 다등급 로지스틱 회귀모형) ⇨ 출력의 과정으로 이루어진다. 출력치는 각 범주가 발생할 확률로 (p_0, p_1, p_2) 인 다항분포값으로 주어진다. 그러나 강수확률예보처럼 추정된 강설발생확률을 직접 사용하지 않고, 세 종류의 문턱치(T1, T2, T3)를 구하여 다음과 같이 예측치를 생성하였다.

```

if (1-P_0) < T1 then snow_index=0;
  else if (T1 ≤ (1-P_0) < T2) or (P_1 ≥ T3)
    then snow_index=1;
  else snow_index=2;
  
```

순서형 대설 예보를 위한 통계 모형 개발

문턱치를 변화시켜가며 관측값과 예측치로 이루어진 3×3 빈도표를 비교하여 문턱치를 결정하고, 훈련자료로 추정된 모형과 결정된 문턱값을 사용하여 모형검증을 수행하였다.

4. 개발 결과

4.1 다등급 로지스틱 회귀모형 개발 결과

45개 잠재적 예측인자를 설명변수로 고려하고 단계별변수선택으로 추정된 설명변수는 Table 3과 같다.

Table 3. Estimated parameters

Parameter	Estimate	Standard	
		Error	P_value
Intercept 2	96.0712	13.3182	<.0001
Intercept 1	99.3411	13.3597	<.0001
CTQP	0.00698	0.00179	<.0001
DWL	0.00438	0.00177	0.0136
ET700	-0.4744	0.0775	<.0001
NE700	-0.0485	0.0224	0.0302
NW500	0.0238	0.00658	0.0055
NW650	-0.5065	0.1416	0.0003
S500	-0.1876	0.0577	0.0012
S850	-0.4125	0.1381	0.0028
T650	-0.4583	0.0516	<.0001
VV650	0.5889	0.1437	<.0001

추정된 다등급 로지스틱 회귀모형에 의하여 산출된 발생확률에 대한 분포에 기초한 3×3 빈도 표 비교를 통하여 결정된 문턱치는 T1=0.02, T2=0.6, T3=0.63이다. 결정된 문턱치에 의한 모형 훈련결과는 Table 4에 요약하였다. 모형훈련에서 얻어진 추정식과 문턱치를 사용하여 모형검증 자료에 대한 예측결과를 Table 5에 요약하였다.

Table 4. 3x3 Table for Training case Table 5. 3x3 Table for Validation case

관측 \ 예측	0	1	2	Total
0	4354 91.76%	388 8.18%	3 0.06%	4745
1	6 8.57%	59 84.29%	5 7.14%	70
2	0 0.00%	0 0.00%	5 100%	5
Total	4360	447	5	4820

관측 \ 예측	0	1	2	Total
0	3000 90.31%	314 9.45%	8 0.24%	4745
1	3 6.82%	39 88.64%	2 4.55%	44
2	0 0.00%	6 75.00%	2 25.00%	8
Total	3003	359	12	3374

4.2 신경회로망 개발 결과

노드수를 변화시키며 AIC를 구한 결과(Table 6)에 따라 최적 노드수는 4개로 정하였다.

Table 6. AIC

nodes	2	3	4	5	6	7
AIC	627	580	573	673	760	796

추정된 다등급 로지스틱 회귀모형에 의하여 산출된 발생확률에 대한 분포에 기초한 3×3 빈도

표 비교를 통하여 결정된 문턱치는 $T1=0.001$, $T2=0.9$, $T3=0.9$ 이다. 결정된 문턱치에 의한 모형 훈련결과는 Table 8에 요약하였으며, 모형훈련에서 얻어진 추정식과 문턱치를 사용하여 모형 검증자료에 대한 예측결과를 Table 9에 요약하였다.

Table 8. 3x3 Table for Training case

예측 관측	0	1	2	Total
0	4447 93.72%	298 6.28%	0 0.00%	4745
1	0 0.00%	63 90.00%	7 10.00%	70
2	0 0.00%	1 20.00%	4 80%	5
Total	4447	362	11	4820

Table 9. 3x3 Table for Validation case

예측 관측	0	1	2	Total
0	3069 92.38%	246 7.41%	7 0.21%	3322
1	8 18.18%	28 63.64%	8 18.18%	44
2	1 12.50%	6 75.00%	1 12.50%	8
Total	3003	359	12	3374

5. 결론

Sohn *et al.* (2005)에서는 집중호우발생 또는 대설발생에 대한 2범주 예측모형을 고려하였으나, 본연구에서는 3범주로 대설발생예측을 확장하였으며, 다등급 로지스틱 회귀모형을 적용된 신경회로망 모형을 적용하였다. 동일한 모형화 전략은 타 지역에도 적용가능하다고 기대된다.

대설 경보(일 신적설량 200mm이상)에 해당하는 경우가 너무 적으므로 좀더 장기간의 자료를 사용한 모형 개발이 요구되며, 3범주에 대한 다른 통계모형을 적용하여 비교 검토하여 예측의 정확도를 향상시키는 것이 향후 연구목표가 된다.

참고문헌

- 조주영, 최준태 (2002), PPM을 이용한 객관적 강수확률예보법, *한국기상학회지*, 38권 2호, 119-127.
- Sohn, K. T., J. H. Lee and C. S. Ryu (2005), Statistical prediction of heavy snow in Honam area, Korea, *Proceedings of the 5th IAP-METRI Joint Research Workshop*, Yanji, China, 16-17.
- Sohn, K. T., J. H. Lee, S. H. Lee and C. S. Ryu (2005) Statistical prediction of heavy rain in South Korea, *Advanced in Atmospheric Sciences*, Vol. 22, No. 5, 703-710.