

SVM을 이용한 계절별 호우 상황 예측 기법

이재동*, 이성우*, 김재광*, 이지형*

*성균관대학교 전자전기컴퓨터공학과

e-mail: {ultrajaepo, lsmoney}@skku.edu, {linux, jhlee}@ece.skku.ac.kr,

Seasonal Heavy Rain Forecast Using SVMs

Jaedong Lee*, Sungwoo Lee*, Jaekwang Kim*, Jee-Hyong Lee*

*Dept of Electronic Electrical and Computer Engineering, Sungkyunkwan University

요 약

본 연구에서는 날씨를 나타내는 속성들의 값을 이용하여 현재로부터 6시간 후의 호우/비호우를 예측하기 위한 기법을 연구한다. 본 연구를 통해 호우/비호우 예측을 할 때 각 속성 값들이 호우, 비호우를 나타내는 일기도의 특정 패턴에 영향을 받는지 혹은 계절별로 영향을 받는지 살펴보았다. 실험을 위하여 20년 누적 일기도를 SVM으로 학습하고 호우와 비호우 각각의 정답 집합을 이용하여 테스트 하였다. 실험 결과 SVM의 호우 예측도는 최대 70% 정도의 정확률을 보였으며 예측에 영향을 주는 것은 특정 패턴보다는 계절에 따른 변화임을 알아내었다.

1. 서론

집중호우, 대설 등의 위험기상 일기를 예측하는 것은 다양한 분야의 많은 사람들에게 지대한 영향을 미치기 때문에 매우 중요한 연구 분야이다[1]. 오늘날 호우를 예측하는 방법 중 대표적인 방법은 현재의 일기도와 유사한 과거 일기도를 기반으로 하여 그 때의 상황을 이용하여 일기를 예측하는 것이다. 그러나 이 방법은 방대한 양의 과거 일기도에서 현재 일기도와 유사한 패턴을 찾아내는 데에 많은 비용이 소모되는 문제가 있다. 이에 효과적으로 과거 일기도를 분석하는 데이터마이닝적 연구가 필요하다 [3].

본 연구에서는 과거의 기상자료를 기반으로 하여 6시간 후에 호우가 발생하는 경우와 발생하지 않는 경우를 학습하여, 현재의 기상 자료가 입력되었을 때 6시간 후에 호우가 발생 할지 안할지를 예측 하였다. 또한 이를 위한 분석적 연구로서 날씨를 나타내는 각 속성들이 호우/비호우에 따른 일기도의 특징 패턴에 영향을 많이 받는지 혹은 계절별 영향을 많이 받는지 비교분석 하였다. 실험을 통하여 6시간 후의 호우/비호우를 예측하는 것은 최대 70%의 정확도를 보였고 각 속성들은 계절에 영향을 더 많이 받는 것을 확인하였다. 그림 1은 호우 예측 시스템의 개념도이다.

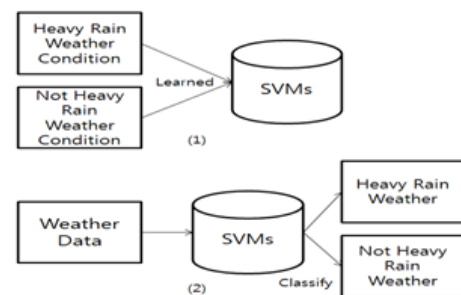


그림 1. 시스템 개념도

2. 과거 일기도

사용한 과거 일기도는 6시간 단위(00, 06, 12, 18 UTC)로 되어 있는 ECMWF 1.5도 해상도의 일기도를 사용하였으며 GHGT(고도), SHUM(철대습도), RHUM(상대습도), TEMP(온도), UWND(동서풍), VWND(북남풍) 6개의 속성과 각 속성별 850hPa, 700hPa, 500hPa, 200hPa 4개의 등압면 층을 사용 하였다. 호우를 판단하는 기준은 6시간 동안 누적 강수량이 70밀리미터 이상의 비가 온 경우이고, 비호우를 판단하는 기준은 6시간 동안의 누적 강수량이 5밀리미터 이하로 정의하였다.

또한 1년 중 호우가 집중되는 5, 6, 7, 8, 9월 중에서 순차적으로 정의된 기준에 따라 호우/비호우 날씨를 추출하였다. 사용한 자료의 총 기간은 1989년부터 2009년 까지 20년의 기간이다. ECMWF 1.5도 자료는 동아시아 지역을 나타내는 60x31 격자 형태로 구성되어 있다. 한반도의 날

씨를 예측하는데 동아시아 전체의 지도를 다 관찰하는 것은 불필요하기 때문에 그림 2가 보이는 바와 같이 한반도 지역과 그 주변을 포함하는 10x10 격자를 추출하여 사용하였다.

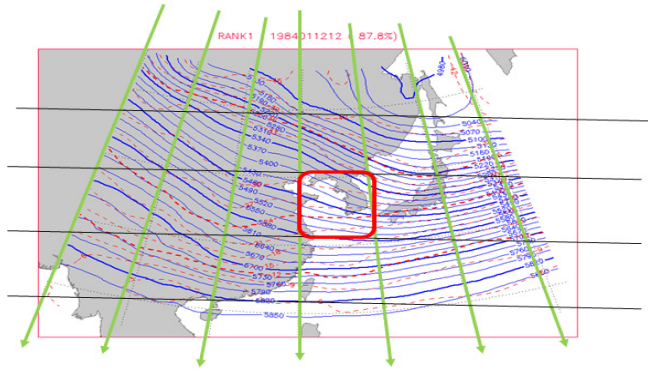


그림 2. 한반도 영역 추출

3. 호우 예측 실험

6시간 후의 호우/비호우 예측 실험을 위하여 위에서 정의된 기준으로 호우/비호우 각 100개씩 총 200개의 날짜를 이용하여 트레이닝 셋을 구성하였다. 실험은 구성된 트레이닝 셋으로부터 160개의 트레이닝 자료, 40개의 테스트 자료로 구성되는 5-fold cross validation으로 진행되었다.

<표 1>은 각 속성, 등압면 별 실험 결과를 보인다. 각 등압면별 가장 좋은 결과를 보인 속성은 굵은 글씨로 표시하였다. 대부분의 등압면에서 예측 정확도가 가장 높은 경우 65~70%를 나타냈다. 등압면별 차이의 특징을 알아보면 모든 등압면에서 GHGT와 SHUM은 좋은 결과를 보였으며, 낮은 고도에서 TEMP 역시 좋은 결과를 보였다.

표 1. 호우/비호우 예측 정확도

	GHGT	RHUM	SHUM	TEMP	UWND	UWND
850 hPa	62.00	53.50	69.50	66.47	50.50	53.00
700 hPa	66.50	49.00	61.00	66.50	53.50	56.00
500 hPa	68.50	53.50	60.00	68.00	59.00	52.00
200 hPa	65.50	65.00	64.30	47.50	59.00	46.00

4. 계절에 따른 호우 상황 예측 실험

계절에 따른 호우 상황 예측 실험을 위하여 위에서 정의된 기준으로 5월과 9월에서 기준에 맞는 호우/비호우 날짜를 각 50개씩 100개 선정하였고, 마찬가지로 7월과 8월 기준에 맞는 호우/비호우 날짜를 각 50개씩 100개를 선정하였다. 5, 9월과 7, 8월을 각기 그룹 하여 두 개의 계절, 즉 한 여름과 주변 여름으로 구분하였다. 총 4개의 세트 구성된 월별 날짜 100개를 이용하여 80개의 트레이닝 자료, 20개의 테스트 자료로 구성하여 5-fold cross validation을 수행하였다. 실험은 850hPa 등압면에서 진행되었으며 결과를 평가하는 지수로는 Accuracy, Precision, Recall, F1 measure가 사용되었다. 호우와 비호우의 구분 없이 같은 계절로 분류된 자료를 이용하여 예측을 한 결과값보다, 서로 다른 계절로 분류된 날짜를 이용한 실험의 결과가 훨씬 높게 나온 것을 알 수 있다.



그림 3. 계절별 호우/비호우 예측 정확도

5. 결과

6시간 후의 호우 상황 예측 실험에서는 대부분의 층에서 RHUM, SHUM, TEMP에 의한 예측 정확도가 높았으며, 저층으로 갈수록 GHGT 속성에 의한 예측 정확도가 높았다. 이전 연구였던 현재의 날씨를 판별하는 것보다는 확실히 예측 정확도가 떨어졌으며, 예측 정확도를 높이기 위한 다양한 방법이 필요하다고 분석하였다[1].

계절에 따른 호우 상황 예측 실험에서는 1개의 속성만으로는 호우와 비호우의 구분이 어렵다는 사실을 알 수 있었고 호우 구분은 대부분의 속성에서 F1 measure 50% 정도를 나타낸다는 것을 알 수 있었다. 구분된 계절에서의 대부분의 속성은 높은 값을 나타냈고 특히 SHUM 속성에서는 F1 measure 가 90% 정도를 나타내었다. 따라서

각 속성 값은 호우/비호우 보다 계절에 더 큰 영향을 받은 것으로 분석하였다. 더 좋은 호우/비호우 예측을 위하여 더 정확률을 높일 필요가 있으며, 분류가 잘 된 속성과 고도를 조합하여 새롭게 추출된 속성을 이용하여 본다면 더 좋은 결과가 나타날 수 있을 것으로 기대된다.

감사의 글

본 연구는 2012년 선진예보기술 시스템 사업의 일환으로 수행하였습니다. 감사드립니다.

참고문헌

- [1] J. Lee, J. Kim, J.-H. Lee, "Parameter Evaluation to Classify Heavy Rain using SVMs," ISIS 2011.
- [2] Korea Meteorological Administration
<http://www.kma.go.kr>
- [3] 이정덕, 유재홍 "주성분 분석을 이용한 유사일기도 검색 알고리즘 개발", 한국기상학회 가을 학술대회 논문집 2010, pp. 214-215.