

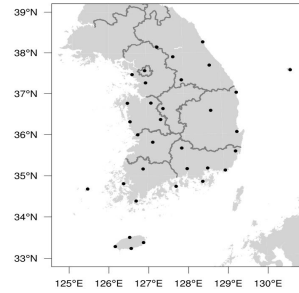
유전 알고리즘 기반의 특징선택을 이용한 SVR 모델의 시정 예측 모델 개발

임승준^{***}, 안광득^{*}, 하종철^{*}, 임은하^{*}, 이용희^{*}, 오성권^{**}
국립기상과학원^{*}, 수원대학교^{**}

Development of SVR model for Visibility Forecasting by using Feature Selection based on Genetic Algorithm

Sung-Joon Lim^{***}, Kwang-Deuk Ahn^{*}, Jong-Chul Ha^{*}, Eun-Ha Lim^{*}, Yong Hee Lee^{*}, Sung-Kwon Oh^{**}
National Institute of Meteorological Research^{*}, The University of Suwon^{**}

Abstract - 본 연구에서는 관측자료 기반의 안개 예보를 수행하기 위해 특징선택을 이용한 SVR 회귀분석 기반 시정 예측 가이드선스를 개발하였다. 예측에 필요한 인자를 사전에 선택하는 유전알고리즘 기반의 최적화 방법을 적용하여, 관측된 여러 기상인자의 입력인자 중 실제 시정을 예측하기 위한 입력인자를 선택하여 준다. 지점별 안개발생에 필요한 입력인자 및 예측 모델을 구성하여 통합적인 예측 모델이 아닌 각 지점에 최적화된 정보를 제공할 수 있도록 예측을 수행한다. 자료의 수집 특성상 3시간 간격으로 3시간 예보를 위한 시정을 예측하고, 예측 모델의 검증을 위해 현업의 수치모델 기반의 시정예측 정보와의 비교를 통해 실제 안개 시점에 대해 비교 분석하였고 그 결과를 통해 긍정적인 효과를 보였다. 예측모델을 적용하여 지도에 예측시정 정보를 제공하는 표출 시스템을 통해 실시간 가이드선스를 제공하고자 연구를 수행하였다.



〈그림 1〉 시정 예측을 위한 지상관측 기상대 34개 지점

1. 서 론

한반도에서는 안개 발생으로 인한 시정 악화로 지상 및 항공과 선박 교통에도 심각한 사고를 유발하여 인명, 재산의 피해를 야기하고 있고 특히 최근에는 이에 대한 대책을 강조한다. 이러한 시정악화 현상은 국지적인 현상에 야기되는 재해의 모습으로, 이를 사전에 예측하기 위한 초단기 예보를 통해 사전에 방지하는 효율적인 대응에 대해 분석을 수행하는 가이드선스가 제안되고 있고 그 필요성은 점차 증가하고 있다. 단순한 분석뿐만 아니라 상황에 따른 초단기적 대응을 필요로 하는 예보 기술의 개발 부분에서 다양한 연구들이 수행되고 있다.

기상분야에서는 현재 기상청에서 영국의 통합모델(Unified Model; UM)의 수치모델을 활용하여 시정예측이 수행되고 있다. 기상청(2010)은 UM기반 역학모델을 이용하여 안개예측 기술을 개발하였지만, 수치예보 가이드선스에 의존한 안개특보 운영의 한계가 있다고 판단되었고, 이에 대한 다양한 분야에서의 예측체계의 개발을 필요로 한다. 수치모델을 통한 예측은 초단기적으로 발생하는 안개에 대한 시정악화의 대응이 쉽지 않고, 이런 위험기상에 효율적인 대응을 위해 초단기예보와 특보의 연계와 실시간 활용 가능한 시정 예측 가이드선스의 개발이 필요하다.

이에 본 연구에서는 목적시정의 지점의 관측 자료를 기반으로 하여 시정 예측을 위해 유전 알고리즘(Genetic Algorithm, GA)의 최적화 알고리즘을 사용하고 분류기로는 SVM(Support Vector Machine)의 회귀분석을 적용한 SVR(Support Vector Regression) 방법을 사용하였다. 본 연구에서 제안하는 유전 알고리즘은 정보이론기반 기상인자를 선택하는 특징선택 방법을 적용하여 지점별 예측에 필요한 정보를 분류기의 입력인자로 사용하여 예측을 수행하는 시정 예측 가이드선스를 제안한다.

2. 본 론

2.1 지상 관측 자료

우리나라는 시정을 목적수행을 통한 관측이 수행되고 꾸준히 관측되는 지점은 많지 않다. 대부분의 목적시정은 기상대 지점에서 자료가 수집되는데, 기계학습방법의 예측을 수행하기 위해 어느정도 기간의 자료가 필요하다. 이에 본 연구에서는 지속적으로 목적을 통해 시정 관측이 꾸준히 수집되는 기상대의 34개 지점 자료를 활용하였다(그림 1). 목적 관측의 특성상 3시간 간격으로 수행을 하기 때문에, 모든 자료는 3시간 간격으로 수집이 되었다. 자료기간은 1989년에서부터 2014년까지 26년의 기간의 자료를 사용하였다. 입력변수는 월, 시간, 풍향, 풍속, 풍속(남북), 풍속(동서), 현지기압, 해면기압, 지상온도(Ta), 이슬점온도(Td), 상대습도, 수증기압(hPa), 1시간강수, 일강수, 적설, 전운량, 중하층운량, 최저운고, 일조, 일사, 현재시정, 일사, 노점편차(Ta-Td), 3시간 간격 과거 3*24시간의 시정, 3시간 간격 과거 3*24시간 지상기온의 39개 인자 정보를 사용하였다.

2.2 샘플링

훈린 데이터 생성을 위해 예측 목표값인 시정을 대상으로 회귀 모델을 생성하는데, 여기서 예측하고자 하는 대상은 시정이 좋지 않은 시간을 목표로 한다. 하지만 한반도의 안개가 발생하는 시간은 극히 짧아 전체 시간에 대해 훈련을 수행하면 시정이 좋은 상황에 대해서 적합화될 위험이 있다. 이를 방지하여 시정이 좋은 않은, 즉 시정 1km 미만인 상황과 그렇지 않은 상황의 자료 비율을 1:1로 조절하여 훈린 데이터를 샘플링하는 전처리 방법을 적용하였다.

2.3 GA 기반의 특징선택 최적화 방법

특징선택의 최적화 방법을 통해 각 지점에서 사용하는 입력인자를 사전에 선택하기 위한 GA 기반의 NMIFS-GA의 특징선택방법을 사용하였다. NMIFS-GA방법은 GA의 진화과정에 기초하는 모델을 이용해 정보이론 기반의 사전인자 추출 필터를 최적화하는 방법으로, 선택한 인자 집합과 예측하고자 하는 인자와의 관계를 통해 엔트로피(Entropy)라는 정보에 기준하여 상호정보량(Mutual Information)의 정도를 최적화하는 값을 선택하는데, 이때 그 정보증가량과 엔트로피의 계산을 통해 최적의 적합도(Fitness) 값을 가지는 최적의 입력인자를 찾아주는 방법이다. 이때, 입력인자를 선택하는 방법에 대해 GA의 진화 방법에 의한 사전인자 추출을 수행하게 되고 그 적합도는 식 (3)과 같이 표현된다. 이때 적합도의 값에서 상호정보량 값으로 인한 과적합의 문제를 해결하기 위해 정규화를 통해 적합도 값의 과적합을 완화하도록 사용하였다.

$$H(X) = - \sum_{i=1}^n p(x_i) \log p(x_i) \tag{1}$$

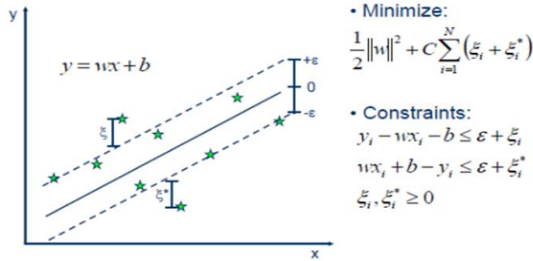
$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right) \tag{2}$$

$$J_{nmifs} = I(X_n; Y) - \frac{1}{|S|} \sum_{k=1}^{n-1} \frac{I(X_n; X_k)}{H(X_k)} \tag{3}$$

여기서 n은 선택된 입력인자의 개수, I는 상호 정보량, H는 정보의 엔트로피값, S는 선택된 인자 집합이다.

2.4 SVR 모델

SVR은 기존 SVM의 회귀분석을 위한 방법이다. SVM은 주어진 여러 입력인자에 대한 주어진 공간에서보다 높은 차원의 공간으로 변환한 후 각 클래스를 2개의 클래스로 구분해내는 초평면을 형성하는 분류기법이다. 이때의 초평면을 서포트 벡터라고 부르는데, 높은 차원으로 특정공간에 표현하는 특성으로 인해 일반화 능력이 뛰어나다고 알려진 분류방법이다. 특정공간을 통한 훈련방법으로 인해 복잡하지만 비선형의 결정경계(decision boundary)를 가지기 때문에 정확성이 높은 편이다. 또한 일반화 능력으로 인해 과적합(overfitting)이 잘 발생하지 않는 장점을 가진다(Yu and Kim, 2009).



〈그림 2〉 회귀분석을 위한 SVR 모델

비선형 분류로의 적용하기 위해 커널 트릭방법이 제안되었고, 여기서는 RBF커널을 사용하여 예측모델을 구현하였다.

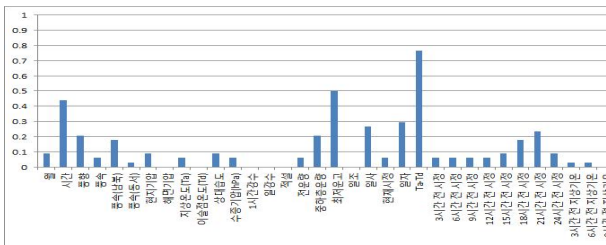
3. 실험 설정 및 결과

시정이 좋지 않은 상황을 예측을 목표로 실험을 수행하는데, 훈련데이터의 예측모델을 생성하는데 샘플링의 전처리 방법을 적용하였다. 또한 정규화를 수행하여 각 입력인자가 가지는 범위를 대상으로 0-1사이의 실수값을 적용하였다.

SVR 기법의 예측 모델을 구축하기 위해 1989~2010년의 자료에 대해 학습을 수행하였고, 2011~2014년 8월의 기간에 대해 검증을 수행하였다. 예측 대상은 안개의 상황으로 예상되는 시정값이기 때문에, 시정이 1km 미만으로 떨어지는 상황을 대상으로 예측성을 확인하고자 하였다. 훈련데이터는 NMIFS_GA의 특징선택을 수행하였는데 GA의 파라미터 설정은 다음 표와 같고, 특징선택의 수행에 따른 선택된 입력인자의 선택물은 다음 그림과 같다.

〈표 1〉 유전 알고리즘 파라미터

인코딩	바이너리 (39차원)
해집단 크기	16 개
세대수	256 세대
선택	토너먼트
교차	다중포인트(3 포인트)
변이	12.5%



〈그림 3〉 전지점에 대한 NMIFS_GA의 특징선택 방법에 의한 입력인자 선택물

생성된 예측모델의 검증을 위해 UM 수치모델 기반의 지역모델로 안개예측을 수행한 RFOG 결과와 비교하였다(그림 4). 예측검증은 안개가 자주 발생하는 06(KST)에 대한 SVR의 03(KST)의 +3hr과 RFOG 12(UTC)예측의 +9hr 예측시간에 대한 결과를 다음의 예보평가식을 통해 비교 검증을 수행하였다.

〈표 2〉 예보평가 기준표

		예보	
		Y	N
관측	Y	H (Hits)	M (Misses)
	N	F (False alarms)	C (Corrects)

SR : $\frac{H}{(H+F)}$ (4)

POD : $\frac{H}{(H+M)}$ (5)

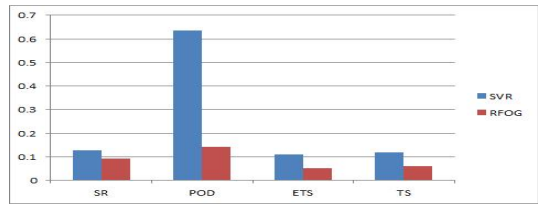
TS : $\frac{H}{(H+F+M)}$ (6)

ETS : $\frac{(H-H_c)}{(H+F+M-H_c)}$ (7)

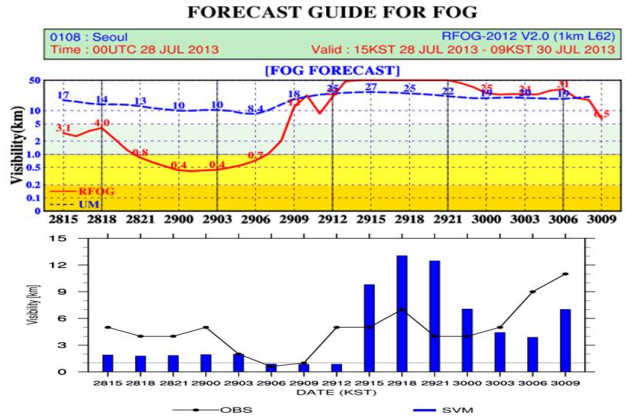
where, $H_c = (H+F)*(H+C)/(H+F+M+C)$

예측 결과에서 보면 SVR기반 예측모델이 RFOG에 비해 ETS가 0.5 정도의 향상을 보였다. 이를 실제 안개 상황에 대해 시정 예측한 결과를 비교하기 위해 현업으로 표출하는 시계열 데이터와 비교 분석해 보았다.

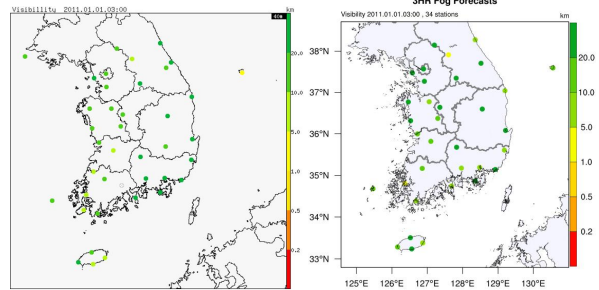
또한 이를 실제 지도상에 표출하는 시스템을 적용하여 예측 시정에 대한 정보를 제공하고자 한다.



〈그림 4〉 RFOG와 SVR 시정 예측 모델의 예측성 비교(2013.5.1.~2014.8.31., 06KST)



〈그림 5〉 서울 지점의 안개 예측 사례 비교(위: RFOG, 아래: SVR 모델)2013.7.28.15KST~2013.7.30.09KST



〈그림 6〉 지점별 시정 예측에 따른 한반도 표출 예시

4. 결 론

현재 기상대에서 목측을 수행하는 지점에 대해서 과거의 정보가 충분한 양으로 수집된 지점들에 대한 예측모델을 생성하고 검증을 수행하였다. 목측시정의 사용으로 인해 예측시간이 3시간 간격으로 생성되는 한계가 있지만, 안개가 주로 발생하는 시간에 대해 RFOG와의 비교검증을 통해 전체 기상대 지점에서의 SVR기반의 안개 예측 모델의 개선된 결과를 확인한다. 하지만 안개의 국지적 특성상 그 예측성 자체는 높지 않았다. 하지만 실제 적용을 통해 안개 상황에 대한 사례를 통해 긍정적인 효과를 감당 할 수 있을 것으로 판단된다.

감사의 글

본 연구는 경기도의 경기도지역협력연구센터사업의 일환으로 수행하였음(GRRR 수원2015-B2, U-city 보안감시 기술협력센터)

[참 고 문 헌]

[1] 손희정, 김맹기, 김좌겸, 황명균, 우상천, “우리나라 안개 발생의 장기 변동 특성 분석”, 한국기상학회 학술대회 논문집, 229-230, 2010
 [2] 기상청, “역학모델을 이용한 안개예측기술 개발”, 88pp, 2010
 [3] Yu H. and Kim S., “SVM Tutorial: Classification, Regression, and Ranking. Handbook of Natural computing”, 2009
 [4] Seo J. H., Lee Y. H. and Kim Y. H., “Feature selection for Very Short-Term Heavy Rainfall Prediction Using Evolutionary Computation”, Advances in Meteorology, Vol. 2014, Article ID 203545, 15pp, 2014