

3A4) 변수변환을 통한 미세먼지(PM10) 예보모형 구축

Development of a Statistical Forecasting Model for PM10 by the Variable Transformation

이영섭 · 김현구¹⁾ · 김희경

동국대학교 통계학과, ¹⁾한국에너지기술연구원

1. 서 론

포항지역은 강우량이 전국 최저 수준인 반면 일조량이 많은 등의 매우 건조한 기후특성과 철강산업공단과 시가지가 형상강을 경계로 인접하고 있는 지형학적 배치구조 때문에 원활한 대기유동에 의한 환기구조를 가졌음에도 불구하고 미세먼지 농도가 전국적으로 높은 수준을 보이고 있다. 포항의 연평균 미세먼지 농도를 살펴보면 1999년부터 2005년까지 7년 동안의 미세먼지 농도가 꾸준한 상승세를 보이고 있으며, 최근에는 장흥동과 대도동의 미세먼지 농도가 전국평균을 상회하고 있다. 한편 포항시의 2004년도 연평균 미세먼지 농도의 전국 측정소별 순위를 살펴보면 전체 177개 측정소 중 포항시 장흥동이 26위, 죽도동이 86위, 대도동이 56위로 포항시의 미세먼지 농도가 중상위권 수준에 해당되어 포항의 미세먼지에 의한 오염은 매우 심각한 상황인 것을 알 수 있다. 본 연구에서는 포항시의 이들 세 곳의 환경측정소에서의 관측자료를 이용하여 포항시 미세먼지 농도 예보모형을 개발하고, 다음날의 최대농도와 평균농도를 예측하여 실생활의 지침으로 제시하고자 한다. 특히 예보모형은 요즘 활발히 연구가 진행되고 있는 SVR(Support Vector Regression)기법을 적용하고, 기존의 연구에서 많이 다루었던 회귀분석, 신경망분석 방법들과 정량적 평가를 통한 상호비교를 함으로써 최종적으로 예측도가 가장 우수한 예측모형을 선별하는 것이 목적이다.

2. 연구 방법

본 연구에서는 당일의 대기오염측정자료와 지면기상자료를 이용하여 다음날의 미세먼지 최대농도와 평균농도를 예측하는 2개의 모형을 구축하였다. 모형에 사용된 독립변수는 김운수(2004)의 연구결과를 참조하여 선택하였으며, 분석자료 및 분석방법은 이영섭 등(2005)에서와 동일하게 적용하였다. 또한 본 연구에서는 미세먼지 최대농도 뿐만 아니라 평균농도를 예측하기 위해 당일 00:00부터 11:00까지의 PM10 평균값(PM10-mean1)과 당일 12:00부터 23:00까지의 PM10 평균값(PM10-mean2)으로 나누어 추가적인 입력자료로 사용하였다. 한편 각 측정소별 PM10 최대농도와 평균농도의 농도값에 따른 분포를 살펴본 결과 각 환경측정소의 경우에서 와이블 분포에 가까운 것으로 나타났다. 따라서 PM10 최대농도와 평균농도가 정규분포에 근접하도록 로그(log) 변환을 시켰다. 로그변환을 통한 자료의 정규화는 분포의 치우침을 완화시킴으로서 회귀모형의 경우 정규성 가정을 만족시켜 예측력을 높일 수 있으며, 신경망 분석이나 SVR과 같이 분포를 가정하지 않는 모형의 경우도 치우친 특정 자료에 대해 모형의 적합이 집중되는 것을 완화시켜 주므로 모형의 예측력을 높이는 효과를 기대할 수 있다.

통계 예측모형은 전통적인 방법인 선형 회귀분석 방법(Regression)과 인공지능기법을 사용한 비선형 방법인면서 정확도가 뛰어난 신경망 분석(Neural Networks, NN)방법, 그리고 신경망 분석이 과도적합과 신경망 구조의 설계에 많은 시간과 노력이 필요한 단점을 해결하는 방법으로 많이 쓰이는 SVR 모형을 적용하였다(Ripley, 1996). SVR은 구조적 위험을 최소화함으로써 과대적합문제에서 벗어날 수 있으며, 불록함수를 최소화하는 학습을 진행하기 때문에 국부적 최적해를 구할 수 있다는 점에서 신경망의 단점을 보완할 수 있는 학습기법으로 알려져 있다(Smola and Schölkopf, 1998).

3. 결과 및 고찰

본 연구에서는 미세먼지의 최대농도와 평균농도에 대한 로그변환을 적용하고, 기존의 대기오염 예측 연구에서 많이 사용하지 않았던 새로운 통계모형인 SVR을 적용하여 보았다. 또한 기존의 다른 연구에서 많이 사용해왔던 회귀모형 및 신경망모형과 그 성능을 비교해 본 결과 표 1에서 나타난 바와 같이

SVR모형이 미세먼지 최대농도의 예측에 있어 예측력 및 적합성이 뛰어난 것을 알 수 있었다. 미세먼지 평균농도의 예측에 있어서는 표 2에 나타난 바와 같이 신경망모형이 가장 예측력이 뛰어났으나 IOA를 기준으로 한 경우는 SVR모형이 가장 적합한 것으로 나타났다. 특히 로그변환을 적용한 경우, 최대농도 예측에서는 IOA의 상승률이 회귀모형은 12.7%, NN이 22.5%, SVR이 42.7%로 특히 SVR의 정확도가 대폭 상승한 것을 알 수 있었으며, 평균농도의 경우에는 IOA의 상승률이 회귀모형은 5.1%, NN이 6.5%, SVR이 6.3%인 것으로 나타났다. 본 연구에서 사용한 통계적 예측모형 뿐만 아니라 시계열 분석기법 또는 분류회귀나무(CART) 기법 등을 사용하여 PM10 예측모형을 구축하고 그 정확도를 본 연구에서 사용한 통계적 기법들과 비교하여 보는 것을 향후과제로 남겨두고자 한다.

Table 1. Quantitative assessment of the maximum of PM10 for each models after log-transformation

Model	Measure	Monitoring Site		
		KME112	KME113	KME114
Regression	RMSE	0.364220	0.508314	0.329076
	CORR	0.576301	0.480560	0.501586
	IOA	0.714179	0.666054	0.653574
NN	RMSE	0.360829	0.458125	0.314520
	CORR	0.581710	0.568367	0.553694
	IOA	0.706885	0.711421	0.696096
SVR	RMSE	0.361543	0.457236	0.305469
	CORR	0.580891	0.574518	0.583573
	IOA	0.696129	0.721888	0.711607

Table 2. Quantitative assessment of the mean of PM10 for each models after log-transformation

Model	Measure	Monitoring Site		
		KME112	KME113	KME114
Regression	RMSE	0.26876	0.41716	0.28688
	CORR	0.68890	0.66266	0.69348
	IOA	0.80990	0.77317	0.80560
NN	RMSE	0.26846	0.40816	0.27477
	CORR	0.68971	0.68092	0.72859
	IOA	0.81151	0.78959	0.80948
SVR	RMSE	0.26883	0.41726	0.275244
	CORR	0.68697	0.66652	0.72235
	IOA	0.80239	0.79574	0.82623

참 고 문 헌

구운서, 윤원정, 권희용, 양재분, 최종혁, 윤희영 (2005) 전일 미세먼지(PM10) 예보시스템 개발, 한국대기 환경학회 춘계학술대회논문집.

김운수 (2004) 서울시 미세먼지 배출량 조사·분석 및 관리방안 연구, 서울시정개발연구원.

김현구 (2005) 기상조건별 비산먼지 관리체계 최적화 연구, 한국대기환경학회지, 21(5) 573-583.

이영섭, 박종석, 김현구 (2005) 미세먼지(PM10)의 통계적 예보모형에 관한 연구-포항지역을 중심으로, 한국대기환경학회 추계학술대회논문집.

Smola, A.J. and Schölkopf, B. (1998) A Tutorial on Support Vector Regression, Royal Holloway College, U.K, Neuro COLT Tech. Rep. TR-1998-030.