

신경망 모델을 이용한 미세 먼지 예보시스템

권희용, 구윤서
안양대학교 컴퓨터공학과, 환경공학과

PM-10 Forecasting using Neural Network Model

Heeyong Kwon, Yoonseo Koo
Dept. of Computer Eng., Environmental Eng., Anyang University

요약

미세 먼지(PM-10)는 인체에 깊은 영향을 미치는 대기 오염원 중의 하나로써, 발생 원인에 대한 규명이 어렵고 기초 자료의 부족으로 통계적 분석 역시 어려워 예보가 지난한 실정이다. 본 연구에서는 기상 및 환경 인자들과 PM-10과의 상관 관계를 분석하고, 다층 인식자 신경망 모델을 이용하여 미세 먼지 예경보 시스템을 구축하였다. 실험 결과 RMSE가 0.04~0.09로 매우 정확한 성능을 보였다.

1. 서론

대기 오염 물질은 인체, 자연 생태계 및 재산상 등 다방면에 직·간접으로 영향을 미치는 것으로 알려져 있으며, 그중에서도 호흡성 먼지, 또는 미세 먼지라고 하는 PM-10 (Particulates Matter < 10 μ m)은 건강에 직접적으로 나쁜 영향을 끼칠 수 있으며, 비교적 대기 중 체류 기간이 길며, 호흡기나 심장 질환이 있는 어린이나 노인 등에게는 지대한 영향을 미칠 뿐만 아니라, 빛의 시정 장애 유발과 산성비 등의 2차적인 영향을 가져올 수 있다. 따라서 미세 먼지의 예경보 시스템이 시급한 실정이다.

미세 먼지는 다른 오염 물질과는 달리 정확한 오염원에 대한 규명이 이루어지지 못하고 있고, 단지 직접적인 배출원으로 공사장, 차량 및 도로의 비산 먼지 등이 있고, 2차 오염 물질로 직접 배출원과 가스상 오염 물질이 반응하여 생성되는 경우가 있으며, 특수한 경우로 중국에서 월경하는 물질 등이 있는 것 정도가 알려져 있다. 이같은 배출원에 대한 기초 자료의 부족과 2차 생성 기구에 대한 이해 부족으로, 물리적 수치 모델 사용에 한계를 갖게 하였으며, 일반적으로 대기 오염 모델에 사용하는 대기 확산 모델의 적용을 어렵게 한다. 다른 접근 방법으로 관측한 측정망 자료와 기상과의 상관 관계를 분석하여 이를 적용한 통계

모델을 사용하는 경우도 있지만 이 역시 비슷한 이유로 한계를 보이고 있다.

미세 먼지에 대한 해외의 연구 및 운용 사례는 미국 Salt Lake City의 미세 먼지 예경보 시스템,[1] 캐나다 몬트리올 지역의 미세 먼지 예경보 시스템,[2] 미국의 Short Course on Air Quality Forecasting,[3] 영국의 UK Air Forecasting System[4] 등이 있으며 정확도가 20% 정도로서 그리 높지 않은 형편이다. Ian의 연구는 신경망에 의한 일 최대 오존 농도 예보의 성공 사례를 기초로, 신경망을 PM 예보에 적용하였으며, 그 결과 PM10 과 PM2.5 예보는 가능성을 보였으나, 다양한 PM의 발생원과 자연 환경, PM과 기상간의 복잡한 관계로 인해 오존의 경우보다는 복잡한 접근 방식이 필요한 것으로 결론을 맺고 있다.[5]

국내의 경우 오존을 비롯한 대기 오염 농도에 관한 연구가 대부분이며 시정과 미세 입자상 물질 화학 조성과의 관계 해석에 관한 연구가 있었으며,[6] 최근 PM-2.5에 관한 연구가 행해지고 있다.

본 연구에서는 이러한 미세 먼지에 영향을 끼치는 기상 및 환경 인자를 연구하고, 그 결과를 이용하여 우리 실정에 맞는 신경망 미세 먼지 예경보 모델을 개발하였다.

2. PM-10과 기상 및 환경 인자

미세 먼지 예경보 모델을 신경망 모델로 구축하기 위해서는 신경망의 입력 방식을 결정하는 것이 중요하다. 따라서 본 연구에서는 기상 및 대기 오염원이 모델링 대상 영역 즉 미세 먼지의 오염도에 영향을 미치는 것으로 가정하고, 미세 먼지 발생에 영향을 미치는 것으로 추정되는 인자들을 선정한 후, 선정된 인자들과 미세 먼지간의 상관관계 분석을 통해 영향력을 분석하고, 그 결과 영향력이 큰 것들을 신경망의 입력 인자로 선정하였다.

미세 먼지 농도에 영향을 미치는 인자로서는 아래와 같은 인자들을 분석하였다.

지표 기상 측정 자료 : 풍향, 풍속, 온도, 습도, 기압, 일사, 강우

고층기상 측정 자료 : 풍향, 풍속, 기압, 온도, 고도 모델링 입력 계산자료 : 대기안정도, 혼합고

국립환경연구원 보유자료 : SO₂, NO₂, CO, 먼지, 오존자료

분석 결과는 표 1.과 같다.

표 1. 기상 및 환경 인자와 PM-10과의 상관도

인자	SO ₂	O ₃	NO ₂	CO	풍속	습도	혼합고	강우	전일강수
상관도	0.61	0.04	0.54	0.54	0.18	0.18	0.16	0.26	0.24

모든 자동 측정망에서 PM-10의 계절적 패턴이 주로 봄철에서 최대가 되며 여름철에 급격히 낮아지는 등 일정하게 나타났으며, 피크 날짜가 일정하였다. SO₂, O₃, NO₂, CO의 일평균의 시계열은 계절적 패턴이 유사하였으나, 피크가 발생하는 날짜는 다르게 나타났다. 오염 물질 항목간에는 SO₂가 가장 PM-10과 관계가 있는 것으로 나타났으며, O₃와는 거의 상관 관계가 없는 것으로 나타났다. PM-10 농도와 지표기상자료의 일평균값과는 큰 상관 관계가 없었다.

3. 신경망 PM-10 예경보 모델

신경망은 과거의 경험을 학습시켜 미지의 입력에 대한 출력을 생성하는 비모수 모델로, 본 과제와 같이 입력 인자 간, 또는 입력과 출력간의 상호 인과 관계가 불분명할 경우, 그 둘 간의 관계를 효율적으로 찾는 특성이 있다. 그러나 신경망은 문제 별로 적합한 신경망 모델 및 구조, 학습 방법이 모두 다르다. 가령 가장 많이 연구되고, 가장 널리 사용되고 있는 오류역전파 학습(Error Back Propagation) 기능을 갖는 다층 인식자(Multi-Layer Perceptron) 신경망 모델의

경우, 단층은 입력과 출력이 선형 관계인 문제에 한정되며, 두층인 경우는 블록다각형 또는 닫힌 영역을 찾는 문제에 한정된다. 이 신경망을 세층 이상으로 구성할 때 임의의 도형 또는 입력과 출력 관계를 해결할 수 있다는 점이 증명되어 있다. 또한 각 층 간의 연결 방법을 완전 연결을 만들 것인가, 아니면 부분 연결을 만들 것인가에 따라 문제에 내재한 입력 간, 입출력 간의 위상 관계를 반영시켜 문제 해결에 큰 영향을 주는 것으로 알려져 있다. 따라서 본 과제에 적합한 신경망 모델과 신경망 구조 및 학습 방법을 연구 개발하는 것이 필요하다.[7]

본 연구에서는 신경망 모델 중 가장 범용성이 높은 오류 역전파 학습(Error Back Propagation) 기능을 갖는 다층 인식자(Multi-Layer Perceptron) 신경망 모델을 기본 모델로 하여, 본 과제에 적합한 신경망 구조 및 학습 방법을 연구 개발코자 한다. 현재 사용 중인 신경망의 구조는 주어진 문제가 전형적인 비선형 관계인 점을 고려하여 입력층과 은닉층(1), 은닉층(2), 출력층의 네 개 층으로 구성하였으며 출력층을 제외한 각층의 노드 수는 실험 환경 즉 입력 인자들에 따라 다양하게 변화하도록 하였다. 출력은 먼지 농도의 수준을 단계로 표시하는 이산 구조로 구성할 수도 있지만, 현재 연구에서는 농도 값 자체와 예보 성능을 정확히 계산할 수 있도록 0.0 에서 1.0 사이의 연속량이 나오도록 하였다.

4. 실험

실험 자료는 추출 기간 2000년(366개), 2001년(365개), 2002.1.1.~ 2003.6.30(180개) 간의 환경 및 기상 자료를 사용하였으며, 대상 지역은 동 기간 중, 사이트별과 도시 권역별로 구분하여, 서울과 인천, 수원동수도권 3개 지역 평균값을 대상으로 하였다. 이때 도시 평균은 해당 도시 내 5개 사이트의 측정 자료를 평균하였다. 또한 권역별 자료의 경우 9시간 측정값과 전일 24시간 측정값, 당일 오전 9시를 기준으로 이전 24 시간 측정값 등 3가지 시간대 별로 실험을 하였다.

위 자료에 대한 PM-10의 연도별 분포 현황은 아래와 같다. 이때 각 경우에 대해 01시에서 24시까지만을 고려하였다. 자료는 PM-10의 농도가 연도별로 증가하는 경향이 있으며, 특히 2002년은 자료 추출 기간이 180일로 전년도에 비해 반으로 감소했음을 감안할 때, 2 등급 이상이 전년도에 비해 3 배 이상 증가하였다. 또한 1이하의 등급이 전체 자료의 80%이상으로 편중된 분포를 보이고 있다.

실제 입력은 위 각 경우에 대해 SO2(ppb), O3(ppb), NO2(ppb), CO(ppm), 온도(°C), 풍속(m/s), 평균강수량(mm), 습도(%), 누적강수량(mm), 전일 PM-10(ppb)의 10개 인자에 대한 3일치 자료를 입력으로 사용하여 되먹임 효과를 살리도록 하였고, 1일 자료만을 사용하는 경우는 되먹임 효과를 제거하도록 9개 인자만을 사용하였다.

출력은 금일 평균 미세 먼지 정도를 0에서 260으로 제한하고 이를 0.0 ~ 1.0의 실수로 정규화하고, 농도는 미국의 AQI지수를 2배로 상세화하여 등급을 정하였다. 따라서 실제 PM-10 농도와 해당 등급은 다음과 같다.

표 2. 최근 3년간 수도권 평균 PM-10 수준

PM-10	2000	2001	2002
0 - 54 (0)	109	145	36
55 - 104 (1)	215	168	106
105 - 154 (2)	30	32	29
155 - 204 (3)	9	13	3
205 - 254 (4)	2	4	1
255 - 304 (5)	1	2	1
305 - 354 (6)	0	1	0
355 - 404 (7)	0	0	0
405 - 454 (8)	0	0	0
455 - 504 (9)	0	0	1
505 - (10)	0	0	3

표 3. PM-10 농도와 정규화값과 등급

AQI	상세등급	PM-10	정규화값
0	0	0 ~ 54	0.0~0.21
1	1	55 ~ 104	0.21~0.41
	2	105 ~ 154	0.41~0.60
2	3	155 ~ 204	0.60~0.80
	4	205 ~ 260	0.80~0.99

실험은 2단계로 나누어 1단계에서는 실험에 영향을 미치는 여러 가지 경우에 대해 최적의 조합을 찾고, 2단계에서는 그 결과 최적의 조합에 대한 모델의 예보 성능을 실험하였다.

실험 1

실험에 영향을 미치는 여러 가지 경우에 대해 최적의 조합을 찾기 위해 1) 사이트별로 각기 학습하고 해당 사이트를 예보하는 경우와, 한 지역의 사이트들의 평균값을 학습하고 그 지역의 일 평균 PM-10의 농도를 예보, 2) 당일 01시에서 09시 까지 9시간 평균 자료를 가지고 학습하고 그날 일 평균치를 예보하는 경우와 전일 24시간 평균 자료를 사용하는 경우, 3)

PM-10 농도값이 254이상이면 강제로 254로 만들어 학습하는 경우와 아예 그럴 경우는 제거하고 학습과 예보를 수행, 4) 1일 자료를 가지고 학습하는 경우와 3일치 자료를 되먹임 효과(Feedback)를 고려하여 이용하는 경우, 5) 환경 자료가 측정 실패등의 이유로 0인 경우 전일 자료를 복사하여 사용하는 경우와 제거하고 학습과 예보를 실시하는 경우 등이다. 따라서 총 32가지 실험의 경우가 가능하지만 현재 이중 10가지 경우의 결과를 아래 표에서 보인다.

표 4. 평균별, 피드백 여부에 의한 실험 결과

	00-01		01-02		00-01		01-02	
	피드백 포함		피드백 포함		피드백 제외		피드백 제외	
평균방법\등급	0	1	0	1	0	1	0	1
24시간(121)	54	28	42	67	52	41	33	60
24시간(122)	74	44	74	40	72	38	91	25
9시간(121)	67	31	63	52	65	40	67	61
9시간(122)	67	64	52	58	55	69	46	49
24시간(Seoul)	74	45	57	46	94	35	68	58

실험 2

1단계 실험 결과 입력 자료의 추출 시간은 01-09시 즉 9시간 평균, 지역은 광역 즉 도시별, 피드백 등을 갖춘 경우가 바람직한 것으로 나타났다. 2단계 실험에서는 이 같은 경우에 대한 3개 도시 및 수도권 지역에 대해 2001년 자료를 가지고 학습시킨 후 2002년 자료를 가지고 실험을 실시하였으며 결과는 아래와 같다. 각 도시의 자료는 해당 도시에 속하는 5개의 측정소에서의 값을 평균하였다. 모델의 성능은 표 5에 정리하였다. 그림 1은 일일 예보 결과 추이이다. 표에서 Bias는 예측값과 목표값의 차이의 합을 평균한 것이고, Accuracy는 차이의 절대값을 평균한 것이다. RMSE는 차이의 자승값을 평균하여 제곱근을 구한 것이다.

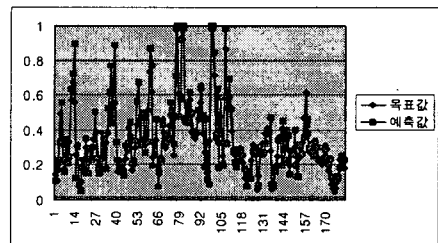


그림 1. 서울의 예보 결과 비교.

표 5. 신경망 모델을 이용한 예측 결과

	서울	인천	수원	수도권
Bias (mean error)	0.002	-0.007	0.002	-0.017
Accuracy (mean absolute error)	0.069	0.048	0.074	0.064
RMSE	0.093	0.073	0.048	0.095

5. 결론

실험 결과 예보 성능은 1) 2000년 자료를 학습하고 2001년을 예보하는 경우가 2001년에 대한 2002년의 예보 결과보다 우수하였으며, 2) 어느 경우나 PM-10의 농도가 높아질수록 성능이 저하하는 현상이 있었다. 3) 다양한 대비 실험 결과는 결과에 미치는 영향이 미미하였지만 특정 사이트에 대한 예보 대비 특정 도시의 평균의 예보의 경우는 현격한 차이가 있었다. 4) 또한 입력 자료의 추출 시간은 당일 01~09까지의 경우가 전일 24시간이나 9시 기준 이전 24시간의 경우 보다 우수하였다.

1)의 결과는 실험 자료의 특성상 2000년과 2001년은 유사하거나 사이트별로 농도값이 다소 증가하는 경향임에 비해 2002년은 현저하게 증가하여 학습 효과가 감소하는데 따른 결과로 파악된다. 2)의 결과는 농도 등급이 1이하인 경우가 80%이상임에 비해 2이상인 경우는 10%~20%에 불과하여 고농도 현상에 대한 학습 효율이 저하한데 기인하는 것으로 판단된다. 3)의 결과는 특정 측정소에서 측정된 환경 및 기상 자료가 PM-10 농도 예보에 미치는 영향이 국부적임을 보였다. 따라서 보다 광역 단위 예보인 도시 전체 평균 PM-10 농도의 예보가 안정적인 것으로 판단된다. 4)의 결과는 당일 평균 값을 예보하는데는 예보 시점에서 가장 가까운 자료를 적정량 만큼만 이용하는 것이 정확도를 향상시키는 것으로 생각된다.

실험 결과는 각 도시 평균에 대한 9시간 자료에 의한 예보의 경우, 신경망 모델은 평균 정확도가 0.064 (17), 회귀 모델의 경우 0.053 (14) 였으며, RMSE는 각각 0.077 (20), 0.084(22) 로 비교적 우수한 것으로 나타났다. 여기서 수치는 1.0으로 정규화한 경우이고, 팔호 속은 260을 최대로 한 경우이다. 즉 연구 결과 모델은 미세 먼지 농도를 최대 260으로 제한할 경우 5%~10%의 오차의 범위에서 예보가 가능하다.

이상의 실험 결과를 볼 때, 위에서 지적하였듯이

년도별로 고농도 PM-10의 분포가 매우 크게 증가한다는 점과 고농도와 저농도 자료의 빈도수가 크게 차이가 남에 따라 학습에 비효율적이거나 나쁜 영향을 미치는 경우를 극복하기 위한 연구가 필요하다. 따라서 연도 단위 학습 및 예보가 아닌 월 또는 분기 단위의 학습 및 예보 체제로 학습 및 예보 자료간의 시간적 간격을 좁히는 방법을 연구할 필요가 있고, 등급별로 학습 및 예보를 서로 다른 모듈로 구현하여 실시하는 방법에 대한 연구가 필요하다. 또한 예보 지역은 특정 사이트보다는 특정 도시에 대한 평균 농도를 예보하는 방법에 대한 연구가 필요하다.

참고문헌

- [1] T. S. Dye, D. S. Miller, C. B. Anderson, C. P. MacDonald, C. A. and Knoderer, B. S. Thompson, "PM2.5 Forecasting Method Development and Operations for Salt Lake City, Utah," *2003 National Air Quality Conference*, U.S. EPA, pp 1-18, 2003.
- [2] M. Benjamin and J. Rousseau, "Winter INFO-SMOG Program Forecast for the Greater Montral Area," *2003 National Air Quality Conference*, U.S. EPA, pp 19-23, 2003.
- [3] Use of Time-Series Analysis to Examine the Link Between Photochemistry and PM Concentrations in Chicago, <http://capita.wustl.edu/NEARDAT/WebLinks/pmupdate.htm>.
- [4] Air Pollution Forecasting in the UK, <http://www.airquality.co.uk/archive/reports/list.php>.
- [5] Ian G. McKendry, "Evaluation of Artificial Neural Networks for Fine Particulate Pollution (PM10 and PM2.5) Forecasting," *Journal of Air & Waste Management Association*, Sep., 2002.
- [6] 한진석, 김병곤, 김신도, "1994년 수도권 지역에서의 시정과 미세 입자상물질 화학조성과의 관계 해석," *한국대기보전학회지*, 제12권 4호, pp. 377 - 387, 1996.
- [7] J.L.McClelland, D.E.Rumelhart, *Parallel Distributed Processing, Vol 1: Foundations*, Cambridge, MA, MIT Press, 1986.