

# 의사 결정 구조에 의한 오존 농도예측

## Forecasting Ozone Concentration with Decision Support System

°김재용, 김성신, 이종범\*\*, 김신도\*\*\*, 김용국\*\*

Jaeyong Kim, Sungshin Kim, Chong-Bum Lee\*\*, Shin-Do Kim\*\*\*, Yong-Goog Kim\*\*

부산대학교 전기공학과, \*\*강원대학교 환경학과, \*\*\*서울시립대학교 환경공학과

### ABSTRACT

In this paper, we present forecasting ozone concentration with decision support system. Forecasting ozone concentration with decision support system is acquired to information from human knowledge and experiment data. Fuzzy clustering method uses the acquisition and dynamic polynomial neural network gives us a good performance for ozone prediction with ability of superior data approximation and self-organization.

### I. 서론

산업의 발전과 기상 변화에 따른 대기중의 오존 농도 메커니즘은 질소산화물 및 탄화수소류 등의 오염 물질로 인한 광화학적 작용과 일사량, 풍속, 기온 등의 기상학적인 변수들의 상호작용으로 생성되어 최근 국내외를 막론하고 하계중 6월부터 8월 사이에 집중적인 고농도 현상을 보이는 것에 관심을 가지고 있다.

오존은 강력한 산화력으로 인해 적당량의 경우 인간에게 이로우나 기준농도 이상의 경우에는 인체의 건강이나 농작물의 수확 및 생태계에 피해를 주고 있다. 우리나라에서는 산업화의 영향으로 각종 오염물질의 배출이 증가함에 따라 지표오존의 고농도현상의 발생이 해마다 빈번해지고 있으며, 이런 고농도 발생으로 인한 오존의 피해를 줄이기 위해 대기오염 예보제 실시 및 정확한 고농도 오존예측에 관한 연구가 진행되고 있다.

기존의 고농도 오존예측을 위한 방법으로 통계적 방법에 의한 선형 회귀 모델[9] 그리고 다변량 통계분석[7]에 의한 방법과 신경회로망을 이용한 예측모델[6]에 의한 연구가 이루어져 왔으나, 기상학적인 변수와 오존 오염물질간의 강한 비선형성과 대류권내에서의 오존생성

에 관한 매우 복잡한 반응기의 동작으로 인하여 고농도 오존 예측에 있어서는 정확한 모형화에 많은 문제들을 가지고 있으며, 대부분의 예측 결과가 실제 오존농도보다 좋은 성능을 보이지 못하고 있다.

따라서, 본 논문에서는 퍼지 클러스터링[2]와 동적 다항식 신경회로망 (Dynamic Polynomial Neural Network, DPNN) 구조를 이용하여 의사결정 구조에 의한 오존 농도 예측 시뮬레이션을 하였다. 또한, *PC* (performance criterion)를 사용하여 모델링 과정 중에 발생할 수 있는 학습 데이터에 대한 의존 영향을 감소시키기 위해서 방법으로 사용되었다.

### II. 퍼지 클러스터링

퍼지 클러스터링은 일반적으로 데이터내부에 존재하는 고유의 특징을 알기 위한 방법으로 Bezdek[2]의 FCM(Fuzzy C-Means)을 많이 사용하고 있다.

본 논문에서는 출력 오존농도의 비슷한 특성에 의한 퍼지 분할공간을 통하여 입력 변수들에 미치는 분류 정도를 알기 위해 FCM을 사용하였다. 그림 1은 출력공간에 대한 퍼지 분할을 통하여 입력공간의 퍼지 클러스터링에 의해 관

련된 특성을 찾아낸다.

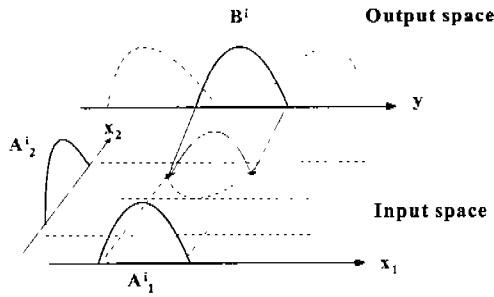


그림 1. 출력공간에 대한 입력공간의 클러스터링.

### III. 다항식 신경회로망

#### 3.1 다항식 신경회로망의 기본구조

다량의 관측자료와 변수들로부터 시스템의 모델을 구성하기 위해 GMDH(Group Method of Data Handling)[1][3][4]를 이용한 다항식 신경회로망은 비선형적이고 복잡한 동적인 시스템의 모델링과 예측 및 지능제어에 응용되어지고 있다. 각 노드에 대해서 두 개의 입력변수로부터 하나의 출력을 생성하는 일반적인 형태의 다항식 신경회로망의 구조를 그림 2에 나타내었으며 다항식 신경회로망의 각 노드에서 입력 변수와 출력과의 관계는 다음과 같은 함수식으로 표현할 수 있다.

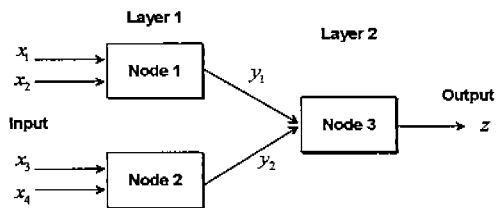


그림 2. 다항식 신경회로망의 기본구조.

각 노드에서의 출력  $y_1, y_2$ 는

$$y_1 = w_{01} + w_{11}x_1 + w_{21}x_2 + w_{31}x_1x_2 + w_{41}x_1^2 + w_{51}x_2^2 \quad (3)$$

$$y_2 = w_{02} + w_{12}x_3 + w_{22}x_4 + w_{32}x_3x_4 + w_{42}x_3^2 + w_{52}x_4^2 \quad (4)$$

와 같은 형태의 다항식으로 나타내어지며, 최종 출력  $z$ 는 각 노드의 출력 값  $y_1, y_2$ 의 다항식으로 나타내어진다.

$$z = w_{03} + w_{13}y_1 + w_{23}y_2 + w_{33}y_1y_2 + w_{43}y_1^2 + w_{53}y_2^2 \quad (5)$$

단,  $w_{ij}(i=0, 1, 2, \dots, n; j=0, 1, 2, \dots, k)$

각 노드의 파라미터를 결정하기 위해 Least Square Method를 사용하며 목적함수인 실제 측정된 값과 학습된 출력된 값과의 차이를 최소화하는 파라미터를 구하게 된다.

$$J = \sum_{k=1}^n (z(k) - \hat{z}(k))^2 = \|z - \Phi w\|^2 \quad (6)$$

$$w = (\Phi^T \Phi)^{-1} \Phi^T z \quad (7)$$

이렇게 해서 구해진 파라미터가 현재노드의 출력함수를 구성하고, 다음 층 노드의 입력 값으로 들어가게 되며 이러한 과정을 반복하여 최종적으로 우수한 성능을 보이는 출력 값을 나타내는 함수를 얻을 수 있다.

#### 3.2 Self Organization

다항식 신경회로망의 또 다른 특징으로서 자기 조직화하는 능력을 들 수 있는데, 다항식 신경회로망은 GMDH방법을 사용하여 입력 자료를 분산에 따라 학습 데이터와 테스트 데이터로 나누어서 시스템을 모델링한다.

두 개로 나누어진 데이터집합을 이용하여 모델의 구조와 각 노드에서의 파라미터를 구하게 되는데, 일반적으로 학습데이터를 이용하여 파라미터를 구하고, 테스트 데이터를 이용하여 모델의 성능을 평가하며 학습 오차와 테스트 오차의 관계를 이용하여 모델의 최종적인 구조를 결정하게 된다. 따라서 다항식 신경회로망은 각 층에서 선택기준에 의해 다음 층의 입력을 선택하고, 그림 3에 나타낸 것처럼 학습 오차와 테스트 오차를 이용한 성능기준에 따라서 과도 학습을 방지하고, 모델의 입력변수와 구조를 스스로 결정한다.

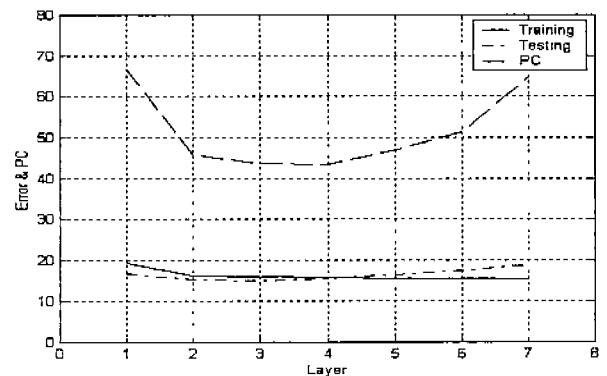


그림 3. 층수의 증가에 따른 모델성능의 변화.

식(8)에 학습 오차와 테스트 오차를 이용하여 PC를 결정하기 위한 식을 나타내었으며,  $\eta$ 는 0에서 1사이의 값을 사용하였다. 식(8)에 의한 성능평가방법은 구성된 모델이 학습데이터 뿐만 아니라 테스트 데이터에도 잘 적응하는가를 평가하며, 준비되지 않은 새로운 데이터에도 적용되는 모델평가방법이다.

$$\begin{aligned} e_1^2 &= \sum_{i=1}^{n_A} (y_i^A - f_A(x_i^A))^2 / n_A, \\ e_2^2 &= \sum_{i=1}^{n_B} (y_i^B - f_B(x_i^B))^2 / n_B, \\ PC &= e_1^2 + e_2^2 + \eta(e_1^2 - e_2^2)^2 \end{aligned} \quad (8)$$

여기서  $e_1$ 은 학습오차,  $e_2$ 는 테스트 오차,  $n_A$ 는 학습 데이터의 개수,  $y_i$ 는 실제 측정된 값,  $f_A(x_i^A)$ 와  $f_B(x_i^B)$ 는 학습에 의한 결과 값과 테스트에 의한 결과 값이다. 그리고  $n_A$ 는 학습 데이터의 개수,  $n_B$ 는 테스트 데이터의 개수이며, 전체 데이터의 개수는  $n = n_A + n_B$ 이다. 따라서 최적의 모델은 학습 오차와 테스트 오차에 대해 PC가 최소화되는 곳에서 얻어진다.

#### IV. 시뮬레이션

대기오염물질의 시간에 따른 변화량 예측은 관측된 변수들과 그 변수들로 구성된 예보시스템 모델의 정확성에 의하여 결정된다. 예측모델의 구성을 위하여 물리화학적인 현상의 수학적 표현을 이용할 수 있으나 대기의 현상은 매우 복잡한 동적인 시스템이므로 본 연구에서는 대기현상과 오염물질에 대한 축적된 데이터로부터 의사결정 시스템 구조를 가진 동적인 다항식 신경회로망을 이용하여 모델을 구성하여 성능을 향상시키고자 한다.

일반적으로 오존의 농도예측을 위해 사용되는 대기오염자료로서는 오존, 일산화탄소, 이산화질소, 아황산가스, 부유먼지등이 사용되며, 기상자료로서 풍속, 풍향, 기온, 일사량, 강수량, 습도 및 운량 등이 사용된다. 이러한 자료 중에서 100PPB이상의 고농도 오존에서의 강수량은 대부분 0mm이기 때문에 고농도 오존예측에 거의 영향을 미치지 못하며, 또한 풍향과 운량은 수치화의 어려움, 그리고 아황산가스와 부유먼지는 대기오염규제로 인해 그 배출량이 줄어들

고 있기 때문에 입력자료에서 삭제하였다. 따라서 학습될 변수로 오존, 일산화탄소, 이산화질소, 기온, 일사량, 습도, 풍속을 사용하였다.

본 논문에서는 하루 중 14~16시경에 최고농도를 나타내기 때문에 이 시간대의 오존 농도를 예측하는 것을 목표로 하였으며, 오염자료중 이산화질소, 오존, 일산화탄소는 오전의 자료를 사용하고 나머지 자료는 오존 농도 예측 시간대의 자료를 사용하였다. 특히, 서울지역 중 하계에 고농도 오존이 자주 발생하는 지역인 쌍문동 8시간 전을 예측(기준시간: 1997년 8월~9월, 14시 예측)하기 위해서 기본 학습자료(초기 자료)를 1996년 5월~9월, 1997년 5월~7월(980개)을 사용하였고, 결측 데이터는 spline 보간하여 사용하였다.

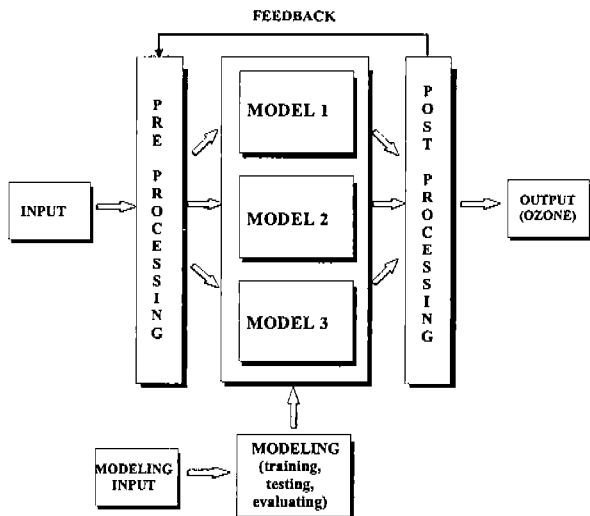


그림 4. 오존 예측시스템의 전체 구조.

의사결정을 가진 모델을 만들기 위하여 기본 학습자료를 퍼지 클러스터링에 의해 고농도, 저농도 모델로 구성하였다. 각 모델에 대한 입력 변수들에 대하여 평균과 편차를 구한 후 소속함수를 구하고 다음으로 각 입력 소속함수에 따른 상대적인 거리를 이용하여 우리가 원하는 소속함수를 찾는다.

예측자료는 전처리기를 통한 후 DPNN (Dynamic Polynomial Neural Network)의 고농도, 저농도모델에 의해서 예측이 수행되어지고 다음으로 후처리에 의해 두 모델에서 나온 결과를 퍼지 클러스터링에서 구한 소속함수를 이용하여 최종결과를 출력한다.

본 시뮬레이션에서는 클러스터의 개수를 2에

서부터 시작하여 4일 때까지를 기준으로 하였다. 기본적으로는 고농도를 최고로 높은 것을 사용하고, 저농도를 나머지의 집합으로 구성하였다. 아래 그림 5는 학습데이터를 클러스터링 후 만들어진 모델에서 나온 RMSE중 최저인 모델을 이용하여 1997년 8월 1일부터 8월 10일까지의 오존예측을 나타내고 있다. 그림 5에서 학습 RMSE가 최저인 경우는 클러스터의 수가 4일 때 10.890이고 예측시의 RMSE는 14.649이다. 성능의 다른 평가방법으로써 예측결과에 대해서 기울기와 절편을 구한다. 기울기는 1, 절편은 0이 되는 경우가 예측이 잘되었다고 할 수 있다. 본 실험에서 기울기가 0.772이고, 절편은 6.860이다.

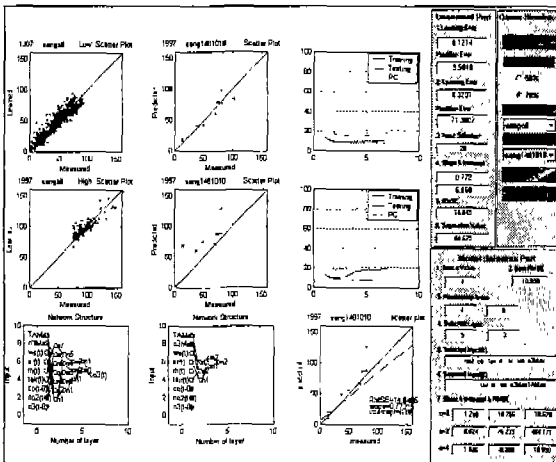


그림 5. 의사결정시스템에 의한 결과.

## V. 결론

본 논문에서 고농도 오존예측을 위해 의사결정구조를 가진 동적 다항식 신경회로망(Dynamic Polynomial Neural Network)을 이용한 모델링 결과가 우수함을 보였다. 오존 농도 예측에 있어서 우수한 예측을 위해서는 모델의 구조를 고정시키는 것보다는 매일의 새로운 자료를 모델의 새로운 입력변수로 사용하여 예측 모델의 구조를 조정해 나가는 것이 더 우수한 예측 결과를 낼 것으로 본다. 퍼지 클러스터링에서의 클러스터의 개수 선택에 의한 고농도 모델의 선정이 중요할 것으로 생각된다. 또한, 앞으로 두 가지 이상의 복합적인 모델을 통한 예측과 입력선택 기준의 다양화, 그리고 모델구조의 최적화를 행한다면 우수한 예측결과가 나올

것으로 본다.

## VI. 참고문헌

- [1] A. G. Ivahnenko. "Polynomial theory of complex system," *IEEE trans. Syst. Man and Cybern*, vol. SMC-12, pp. 364-378, 1971.
- [2] James C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum, 1981.
- [3] Duc Trung Pham and Liu Xing, *Neural Networks for Identification, Prediction and control*, Springer-Verlag Inc., 1995.
- [4] A. G. Ivakhnenko, "The Group Method of Data Handling in Prediction Problem," *Soviet Automatic Control*, vol. 9, no. 6, pp. 21-30, 1976.
- [5] S. Farlow, ed., *Self Organizing Method in Modeling: GMDH-Type Algorithms*, Marcel Dekker, Inc., New York, 1984.
- [6] 김용국, 이종범. "하계의 일 최고 오존농도 예측을 위한 신경망 모델의 개발," *한국 대기보전학회지*, vol. 10, no. 4, pp. 224-232, 1994.
- [7] 김영성. "1991~1993년 사이 우리나라 오존농도 변화," *한국 대기보전학회지*, vol. 12, no. 1, pp. 55-66, 1996.
- [8] 허정숙, 김동술. "다변량 통계분석을 이용한 서울시 고농도 오존의 예측에 관한 연구," *한국 대기보전학회지*, vol. 9, no. 2, 1993.
- [9] 이선기, 이종범, "서울지역의 광화학 오염농도추정을 위한 중회귀 모델의 개발," *기상연구논문집*, vol. 8, no. 1, pp. 71-85, 1991.
- [10] 김성신, "A Neuro-Fuzzy Approach to Integration and Control of Industrial Processes : Part I," *퍼지 및 지능시스템 학회 논문지*, vol. 8, no. 6, pp. 58-69, 1998.
- [11] 김용국, "수도권 지역의 광화학 오존농도 예측에 관한 연구," *강원대 박사학위 논문*, 1996.
- [12] 홍낙기, 이종범, 김용국, "다변량 해석을 이용한 서울시 하계스모그의 형태 분류," *한국 대기보전학회지*, vol. 9, no. 4, pp. 278-287, 1993.