

멤버십 함수와 DNN을 이용한 PM₁₀ 예보 성능의 향상

유숙현[†], 전영태^{**}, 권희용^{***}

Improvement of PM₁₀ Forecasting Performance using Membership Function and DNN

Suk Hyun Yu[†], Young Tae Jeon^{**}, Hee Yong Kwon^{***}

ABSTRACT

In this study, we developed a PM₁₀ forecasting model using DNN and Membership Function, and improved the forecasting performance. The model predicts the PM₁₀ concentrations of the next 3 days in the Seoul area by using the weather and air quality observation data and forecast data. The best model(RM14)'s accuracy (82%, 76%, 69%) and false alarm rate(FAR:14%,33%,44%) are good. Probability of detection (POD: 79%, 50%, 53%), however, are not good performance. These are due to the lack of training data for high concentration PM₁₀ compared to low concentration. In addition, the model dose not reflect seasonal factors closely related to the generation of high concentration PM₁₀. To improve this, we propose Julian date membership function as inputs of the PM₁₀ forecasting model. The function express a given date in 12 factors to reflect seasonal characteristics closely related to high concentration PM₁₀. As a result, the accuracy (79%, 70%, 66%) and FAR (24%, 48%, 46%) are slightly reduced in performance, but the POD (79%, 75%, 71%) are up to 25% improved compared with those of the RM14 model. Hence, this shows that the proposed Julian forecast model is effective for high concentration PM10 forecasts.

Key words: PM₁₀ Forecasting, Air Quality Index, Deep Neural Network, Membership Function, AI

1. 서 론

중국주요 공업도시지역에서 발생한 고농도 미세먼지가 편서풍을 타고 국내로 유입되는 상황에서 국내 대기 정체로 인해 배출된 다양한 오염원이 더해지면서 고농도 미세먼지 현상이 발생하고 있다[1]. 이러한 미세먼지는 국민 건강을 직·간접적으로 위협하고 있어서 이를 예방하기 위해서는 정확한 미세먼지 예보가 필요하다.

현재 환경부에서 운영 중인 미세먼지 예보의 지수 적중률은 70%이나, 나쁨 이상의 고농도 감지확률은 이보다 낮아서 실제 시민이 느끼는 체감 오염도와

많은 차이를 보이고 있다. 미세먼지 예보가 국민 건강 보호를 위한 예방적 기능을 수행하기 위해서는 예보의 정확도 개선이 시급하다. 시행중인 미세먼지 예보는 수치모델의 결과인 수치 예보 값을 중심으로 하고 있으나, 수치 모델의 입력 자료인 배출량 및 기상자료 등이 내포한 불확실성과 복잡한 대기현상을 완전하게 반영하지 못하는 수치모델의 특성으로 인하여 예보 정확도에 한계가 있다[2].

본 연구에서는 이러한 한계를 극복하기 위해서 과거의 측정 자료와 수치 예보 자료를 심층신경망 (DNN: Deep Neural Network)에 학습시켜 불완전한 수치모델 결과를 보정하고, 예보 정확도를 향상시킬 수 있

* Corresponding Author: Hee Yong Kwon, Address: (430-714) 22 Samdeokro 37beon-gil, Manan-gu, Anyang-si, Gyeonggi-do, Korea, TEL: +82-31-467-0878, FAX: +82-31-463-1249, E-mail: hykwon@anyang.ac.kr
Receipt date: July 31, 2019, Revision date: Sep. 3, 2019
Approval date: Sep. 5, 2019

[†] Dept. of Information & Communication Eng., Anyang University

(E-mail: shy@anyang.ac.kr)

^{**} Dept. of Computer Eng., Anyang University
(E-mail: mohty@daum.net)

^{***} Dept. of Computer Eng., Anyang University

는 예보 모델을 제안, 개발하였다. 심층신경망은 기존의 인공신경망(ANN: Artificial Neural Network)에서 심층학습이 가능하도록 입력층과 출력층 사이에 다수의 은닉층을 포함하도록 확장한 모델이다[3]. 심층신경망은 인공신경망과 마찬가지로 복잡한 비선형 관계의 모델링이 가능하며, 계층이 깊기 때문에 기존의 인공신경망보다 더 적은 노드들만으로도 복잡한 데이터를 모델링 할 수 있게 해준다[4]. 이러한 심층신경망의 특징은 미세먼지 농도 예측과 같이 복잡하고, 비선형적이며 방대한 자료를 대상으로 하는 문제에서 효율적이다.

이미 대기질 예보에 인공신경망의 효과는 입증되어 다양한 관련 연구가 이루어지고 있다. 관련연구로는 인공신경망에서 가장 범용적으로 사용되는 MLP (Multi Layer Perceptron)를 이용하여 PM₁₀과 PM_{2.5} 등으로 대표되는 대기오염 물질의 농도를 예측한 연구[5-10]와 RBF(Radial Based Function)를 이용하여 PM₁₀의 농도를 예측한 연구[11], 시계열 대기질

자료의 결측치를 처리하기 위해서 DRNN(Deep Recurrent Neural Network)를 기반으로 시공간 예측 프레임워크를 제안한 연구[12], 대기질 예보의 성능 향상을 위해 커널 삼중대각 희소행렬을 이용해 고속으로 자료동화 수행한 연구[13] 등이 있다.

본 연구에서는 기존의 대기질 예보에 사용한 인공신경망보다 일반적으로 더 좋은 성능을 가진다고 평가되는 심층신경망을 기반으로 서울권역의 PM₁₀ 예측 농도를 생성하는 예보모델을 개발했다. 입력인자는 대기질 및 기상 측정자료와 예보자료의 6시간 평균값을 사용하였고, D+0(당일), D+1(내일), D+2(모레)까지의 PM₁₀ 농도를 6시간 단위로 생성했다. 제안한 예보모델의 성능 평가는 Table 1에 제시된 통합대기환경지수(CAI: Comprehensive air-quality index)를 기준으로 한 지표(Fig. 1)에 따랐다. Fig. 1에서 기술한 지수적중률(Accuracy)은 예보모델의 모든 지수 구간의 적중률을 평가하고, 감지확률(POD: Probability of Detection)은 고농도의 감지정도를 평

Table. 1 PM₁₀ index section

Index	Range	Meaning
1	0~30	Levels that will not be affected by air pollution-related patients
2	31~80	Levels that are less affected when chronic exposure to the patient group
3	81~150	Levels that will have a serious impact to patient and sensitive groups (children, the old and the infirm etc.) and may have a experience the health discomfort to general
4	151~	Levels that will have a significant impact when acute exposed to patient and sensitive groups and may have a weak impact on the general

Level		Prediction value			
		1	2	3	4
Observation value	1	a1	b1	c1	d1
	2	a2	b2	c2	d2
	3	a3	b3	c3	d3
	4	a4	b4	c4	d4

I: , II: , III: , IV: 

Evaluation item	Equation
Accuracy	$= \frac{(a1 + b2 + c3 + d4)}{N} \times 100(\%)$
POD: Probability of Detection	$= \frac{IV}{(III + IV)} \times 100(\%)$
FAR: False Alarm Rate	$= \frac{(II)}{(II + IV)} \times 100(\%)$

Fig. 1. Items and methods for evaluating the performance of the forecast model.

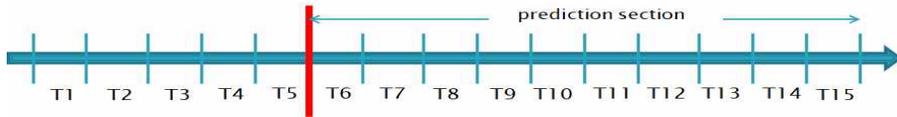


Fig. 2. Time unit concept of forecast model.

가하고, 오경보율(FAR: False Alarm Rate)은 고농도 예보의 정확성을 나타내는 지표로 측정 값이 저농도 구간에 속하는데 고농도 예보를 한 경우에 해당된다.

이어지는 2장에는 제안한 예보모델의 네트워크 구조와 입력인자에 대해 설명하고, 이를 기반으로 개발된 ORIM, RM14 모델들을 소개한다. 3장에서는 고농도 예보 성능을 향상시키기 위해 날짜인자를 학습에 추가한 Cos 예보모델과 이를 개선한 S_Cos 예보모델, 제안한 멤버십 함수를 이용하여 날짜 인자를 추가한 Julian 예보 모델에 대해 설명한다. 4장에서는 개발된 예보모델들의 성능을 평가한 실험결과를 제시하고 5장에서는 결론 및 향후계획을 기술한다.

2. 심층신경망을 이용한 PM10 예보모델

2.1 입력인자와 DNN 예보모델

제안한 DNN 예보 모델은 주어진 자료들의 6시간 평균값을 입력으로 하여 서울 권역의 당일(D+0), 내

일(D+1), 모레(D+2) 구간의 PM₁₀ 농도를 6시간 단위로 예측한다. 예보 모델에 사용한 자료들은 대기질 및 기상 측정자료와 예보자료이며 상세한 내용은 Table 2에 기술했다.

Table 2에 제시된 인자들을 보면 각각 서로 다른 범위의 값을 가짐을 알 수 있다. 이러한 값을 그대로 입력 자료로 사용하면 값의 극심한 차이로 인해 모델의 성능이 저하되므로 이를 방지하기 위해서 데이터를 정규화 하는 전처리 작업이 필요하다. 식(1)~식(4)에 정규화 방법을 기술했다. 기술한 정규화 수식에서 x_i 는 i 번째 인자이고, Max_i 는 i 번째 인자의 최대값, Min_i 는 i 번째 인자의 최소값이다. 정규화는 우선 입력인자가 최대값을 초과하거나 최소값 미만이면 각각 최대값 혹은 최소값으로 대체하고, 그 외의 경우는 주어진 수식에 의해 모두 일정한 범위의 값이 되도록 한다.

$$x_i > Max_i, normal(x_i) = Max_i \tag{1}$$

$$x_i < Min_i, normal(x_i) = Min_i \tag{2}$$

Table 2. Input Data

Variable	Unit	Min	Max	Variable	Unit	Min	Max
PM ₁₀	μg/m ³	0	180	Solar radiation	0.01MJ/hr-m ²	0	2
PM _{2.5}	μg/m ³	0	110	CASE4_PM ₁₀	μg/m ³	0	180
SO ₂	ppm	0	0.01	Mixed highs	m	100	2000
NO ₂	ppm	0	0.01	Geopotential height_850hpa	m	1250	1500
O ₃	ppm	0	0.08	Pressure_850hpa	Pa	97000	102500
CO	ppm	0	1.4	U_850hpa	m/s	-15	20
U	m/s	-6	6	V_850hpa	m/s	-15	15
V	m/s	-5	5	Relative humidity_850hpa	%	0	50
Pressure	Pa	970	1025	Temperature_850hpa	K	270	310
Temperature	K	260	310	Geopotential height_925hpa	m	710	860
Dew point temperature	K	260	310	U_925hpa	m/s	-15	20
Relative humidity	%	0	100	V_925hpa	m/s	-15	15
Accumulated Precipitation	mm	0	10	Temperature difference_850-925	K	0	8

$$Mn_i > 0, normal(x_i) = \frac{x_i - Mn_i}{Max_i - Mn_i} \quad (3)$$

$$Mn_i < 0, normal(x_i) = \frac{x_i}{Max_i - |Mn_i|} \quad (4)$$

이렇게 정규화를 수행한 자료들을 입력으로 하여 심층신경망을 학습시키고, 6시간 단위의 PM₁₀ 예보 값을 생성한다. 예보 시간 프레임을 아래의 Fig. 2와 Table 3에 나타냈다. 여기서 각 T는 6시간 단위로 구성되어 있고, 15시에 예보를 시행하는 모델이므로 T1~T5는 과거에 해당하는 구간이고, T6~T15는 미래에 해당하는 구간으로 PM₁₀ 예보 농도를 생성해야 하는 구간이다. 이 중 T6~T7은 D+0에 해당하고, T8~T11은 D+1, T12~T15는 D+2에 해당한다.

DNN 예보 모델은 위에서 기술한 6시간 단위의 예보 값을 생성하기 위해 예보 구간(T6~T15)별로 총 10개의 개별 네트워크 구성했다. Fig. 3에 나타난 네트워크는 1개의 입력층과 다수의 은닉층, 1개의 출력층을 가진다. 네트워크 구성은 예보 모델별로 상이하나, 입력층은 기본적으로 대기질 및 기상 측정자료와 예보자료를 입력한다. 은닉층의 노드수는 입력층에 따라 결정되고, 출력층의 노드는 1개이며 해당 시간 구간의 PM₁₀ 측정값을 목표로 학습한다. 모델별 입력 인자 및 네트워크 구성의 차이는 다음절에서 자세히 기술한다.

Table 3. Time frame of 15 hour forecast model

Time	Time corresponding to section
T1	D-1 07~12 o'clock, Measure at forecast time
T2	D-1 13~18 o'clock, Measure at forecast time
T3	D-1 19~24 o'clock, Measure at forecast time
T4	D+0 01~06 o'clock, Measure at forecast time
T5	D+0 07~12 o'clock, Measure at forecast time
T6	D+0 13~18 o'clock
T7	D+0 19~24 o'clock
T8	D+1 01~06 o'clock
T9	D+1 07~12 o'clock
T10	D+1 13~18 o'clock
T11	D+1 19~24 o'clock
T12	D+2 01~06 o'clock
T13	D+2 07~12 o'clock
T14	D+2 13~18 o'clock
T15	D+2 19~24 o'clock

제안한 DNN 모델의 학습은 비용함수를 정의하고, 경사 하강법을 사용하여 가중치 및 파라메타들을 조정하여 비용을 최적화하는 과정을 거친다. 비용함수란 예측 값과 목표 값과의 차이를 계산하는 역할을 하며, 이를 통해 신경망의 파라메타들을 얼마나 조정할 것인지 결정하게 된다. 아래의 수식에 학습에 사용한 MSE(Mean Square Error) 비용함수를 기술했다.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (5)$$

본 연구에서는 지금까지 기술한 DNN 모델을 기반으로 먼저 2개의 예보 모델을 개발했다. 첫째는 과거 30시간의 측정 자료와 예보 자료를 입력으로 하는 ORIM, 두 번째는 과거 6시간의 측정 자료와 예보 자료를 입력으로 하는 RM14이다. 이어지는 절에서 각 모델별 특징을 설명한다.

2.2 ORIM 예보 모델

ORIM 예보 모델은 Fig. 3에서 제안한 DNN 모델을 기본으로 하되, 예보 시점에서 과거구간인 T1~T5에 해당하는 대기질 및 기상 측정 자료와 예보 구간에 해당하는 T6~T15의 예보 자료를 입력으로 하여 학습하고 PM₁₀ 예보 농도를 생성하는 모델이다. 현재 혹은 예보 시점의 PM₁₀ 농도는 과거의 기상 및 대기질의 영향을 받아서 생성된 결과라는 점에 의거하여 과거 30시간의 측정 자료를 사용한 것이 이 모델의 특징이다. ORIM의 인자는 T1~T5 구간에 해당하는 과거의 측정 자료가 각각 14개씩 총 70개이고, 예보구간인 T6~T15의 예보자료가 16개로 총 86개의 인자가 입력층에 입력되고, 3개의 은닉층과 1개의 출력층을 가진 5층의 네트워크로 구성된다. 각각 Table 4와 Table 5에 ORIM의 입력인자와 네트워크 파라메타를 상세히 기술했다.

2.3 RM14 예보 모델

RM14 예보 모델은 ORIM과 마찬가지로 과거의 대기질 및 기상측정 자료와 예보자료를 사용하되, ORIM과 다른 점은 과거 30시간의 측정 자료가 아니라 과거 6시간에 해당하는 T5의 측정 자료만 학습에 사용한다는 점이다. RM14 모델의 인자는 T5의 측정 자료 14개와 예보구간의 예보 자료 16개로 총 30개의 인자가 입력층에 입력되고, 4개의 은닉층과 1개의 출

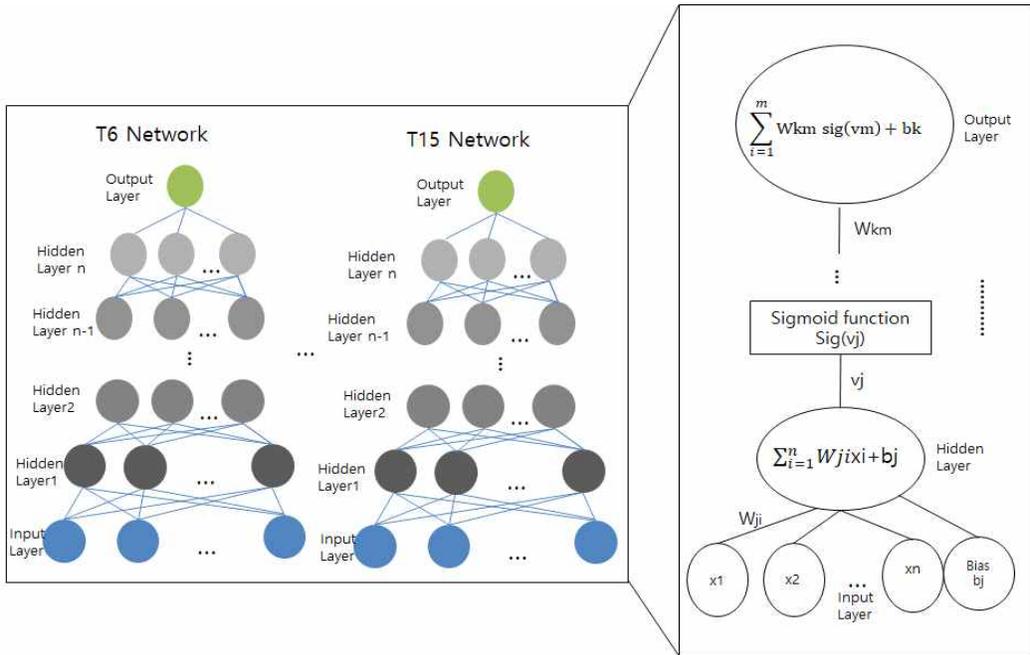


Fig. 3. Proposed DNN architecture.

Table 4. Input Data of ORIM

Time	Kind		Data
T1	Observation data(70)	Weather	U, V, Pressure, Temperature, Dew point temperature, Relative humidity, Precipitation, Solar radiation
		Air quality	PM _{2.5} , O ₃ , NO ₂ , CO, SO ₂ , PM ₁₀
T2	Observation data(70)	Weather	U, V, Pressure, Temperature, Dew point temperature, Relative humidity, Precipitation, Solar radiation
		Air quality	PM _{2.5} , O ₃ , NO ₂ , CO, SO ₂ , PM ₁₀
T3	Observation data(70)	Weather	U, V, Pressure, Temperature, Dew point temperature, Relative humidity, Precipitation, Solar radiation
		Air quality	PM _{2.5} , O ₃ , NO ₂ , CO, SO ₂ , PM ₁₀
T4	Observation data(70)	Weather	U, V, Pressure, Temperature, Dew point temperature, Relative humidity, Precipitation, Solar radiation
		Air quality	PM _{2.5} , O ₃ , NO ₂ , CO, SO ₂ , PM ₁₀
T5	Observation data(70)	Weather	U, V, Pressure, Temperature, Dew point temperature, Relative humidity, Precipitation, Solar radiation
		Air quality	PM _{2.5} , O ₃ , NO ₂ , CO, SO ₂ , PM ₁₀
T# #:6~15	Forecast data(16)	Air quality	CASE4_PM ₁₀
		Weather	Temperature, Pressure, Relative humidity, Mixed highs, U, V, Geopotential height_850hpa, Pressure_850hpa, U_850hpa, V_850hpa, Relative humidity_850hpa, Temperature_850hpa, Geopotential height_925hpa, U_925hpa, V_925hpa, Temperature difference_850-925

Table 5. Network parameter of ORIM

Input layer	Hidden layer1	Hidden layer2	Hidden layer3	Output layer	Learning rate	Epoch
86	15	8	4	1	0.009	160000

Table 6. Input Data of RM14

Time	Kind		Data
T5	Observation data(14)	Weather	U, V, Pressure, Temperature, Dew point temperature, Relative humidity, Precipitation, Solar radiation
		Air quality	PM _{2.5} , O ₃ , NO ₂ , CO, SO ₂ , PM ₁₀
T# #:6~15	Forecast data(16)	Air quality	CASE4_PM ₁₀
		Weather	Temperature, Pressure, Relative humidity, Mixed highs, U, V, Geopotential height_850hpa, Pressure_850hpa, U_850hpa, V_850hpa, Relative humidity_850hpa, Temperature_850hpa, Geopotential height_925hpa, U_925hpa, V_925hpa, Temperature difference_850-925

Table 7. Network parameter of RM14

Input layer	Hidden layer1	Hidden layer2	Hidden layer3	Hidden layer4	Output layer	Learning rate	Epoch
30	15	8	4	2	1	0.009	160000

력층을 갖는 6층의 네트워크로 구성된다. 각각 Table 6와 Table 7에 RM14의 입력인자와 네트워크 파라메타를 상세히 기술했다.

3. 제안한 멤버십 함수를 이용한 예보모델

3.1 관련연구

예보 모델의 성능을 평가하는데 4개의 농도 구간의 지수를 맞추는 지수적중률과 더불어 고농도를 감지하는 성능은 중요한 요소이다. 고농도의 발생 요인에는 여러 가지가 있지만, 계절적 특성은 가장 상관도가 높은 요인 중 하나이고, 계절을 반영하기에 가장 명확한 요소는 날짜이다. 따라서, 날짜를 입력자로 추가하면 고농도 예보 성능 향상에 기여할 수 있다. 하지만, 날짜 자료를 전처리 없이 그대로 사용하면 계절 특성이 왜곡되어 반영될 수 있다. 예를 들어, 12월과 1월을 거의 유사한 특성을 가짐에도 수의 정량적 차이로 인해 완전히 다른 자료로 입력된다. 이를 위해 계절적 특성이 정확히 반영될 수 있도록 적절히 날짜를 가공하여야 한다. 이와 관련된 연구는 날짜에 코사인 함수를 적용하여 날짜를 하나의 값으로

표현한 방법[14]이 있는데, 이 방법을 이용하면 겨울에는 최대 값을 갖고, 여름에는 최소 값을 갖게 된다. 이는 일반적으로 여름으로 갈수록 고농도의 출현이 희소하고, 겨울로 갈수록 고농도의 출현이 빈번한 계절적 특성이 표현된다. 수식 (6)과 Fig. 4에 Cos 함수를 나타냈다. 수식 (6)에서 dth는 서수(ordinal number)를 나타내고, T는 년 중 일수를 의미한다. 수식 (6)에 의하면 날짜는 -1에서 1 사이의 값으로 표현되고, 대략의 고농도의 계절적 특성을 반영하게 된다. 하지만, 보통 겨울부터 봄철까지 점점 고농도의 정도가 심해지고, 여름부터 가을까지 저농도를 유

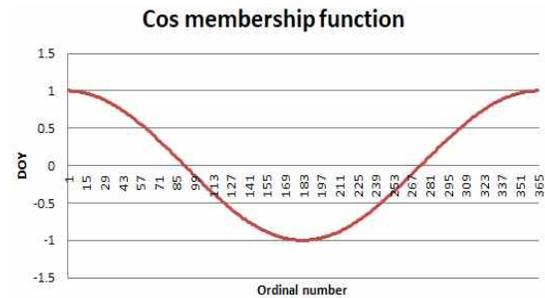


Fig. 4. Cos function.

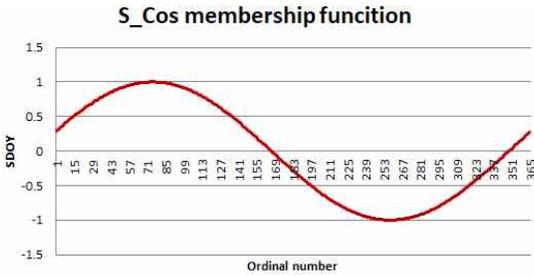


Fig. 5. S_Cos function.

지하는 현상을 Cos 함수만으로는 반영할 수 없다. 본 연구에서는 이런 계절적 특성을 반영하기 위해서 Cos 함수를 75일 시프트 시킨 S_Cos 함수를 제안했고, 이를 수식 (7)과 Fig. 5에 나타냈다.

$$DOY = \cos(2\pi d_{th}/T) \tag{6}$$

$$SDOY = \cos(2\pi (d_{th} + 75)/T) \tag{7}$$

3.2 Julian 멤버십 함수

본 연구에서는 3.1절에서 소개한 관련연구와 마찬가지로 감지확률을 높이기 위해 계절 특성이 반영되는 날짜 데이터를 학습 인자로 입력했다. Cos와 그를 개선한 S_Cos은 날짜 데이터를 하나의 수치 값으로 표현하나, 제안한 방법은 날짜 및 계절이 연속성을 갖는 특성에 의거하여 날짜 데이터를 1월에서 12월을 대표하는 12개의 연속적인 인자로 표현하는데, 이를 위해서 줄리안 멤버십 함수를 제안했다.

멤버십 함수는 퍼지이론에서 나오는 개념으로 퍼지집합에 속하는 정도를 정의한 함수이다. 제안한 줄리안 멤버십 함수(Julian membership function)는 날짜를 표현하는데 1월부터 12월에 해당하는 12개의 인자를 사용한다. 이 12개의 인자 중 2개만이 활성화되고, 나머지 10개의 인자는 0으로 설정한다. 활성화되는 2개의 인자는 각각 주어진 날짜의 월(Month)과 인접 월(Adjacent_Month)의 값을 표현한다. 예를 들어, 주어진 날짜가 4월 20일 이라면 제안한 줄리안 멤버십 함수에 의해서 인접 월(Adjacent_Month)은 5, 해당 월의 값(Month_value)은 0.8, 인접 월의 값(Adjacent_Month_value)은 0.2가 되어 날짜를 나타내는 12개 인자는 0, 0, 0, 0, 0.8, 0.2, 0, 0, 0, 0, 0, 0가 된다. 이 인자들의 의미는 주어진 날짜(4월 20일)가 4월에 해당되는 날짜이지만, 5월에 인접해있는 날짜이므로, 4월에 해당하는 4번째 인자는 0.8, 5월에

해당하는 5번째 인자는 0.2로 설정하여 날짜의 연속성을 표현한 것이다. 인접 월(Adjacent_Month)의 정의와 인접 월의 값(Adjacent_Month_Value), 해당 월의 값(Month_Value)을 계산하는 과정은 Fig. 6에 기술하고, 줄리안 멤버십 함수는 Fig. 7에 나타냈다.

3.3 날짜를 적용한 예보 모델들

Cos, S_Cos, Julian 예보 모델은 RM14 모델을 기반으로 하되, 고농도의 감지확률을 향상시키기 위해서 고농도 자료와 상관도가 높은 날짜 데이터를 각각의 방법에 의해 표현하여 학습 인자로 추가한 모델이다. Cos과 S_Cos 예보모델은 날짜 인자를 1개의 값으로 표현하고, Julian 예보모델은 12개의 연속된 인자로 표현한다.

날짜 인자를 추가한 3 개의 모델 모두 RM14를 기반으로 하므로 Cos과 S_COS 예보모델은 T5의 측정자료 14개와 예보구간의 예보 자료 16개를 사용하고, 거기에 1개의 날짜 인자가 추가되어 총 31개의 인자가 입력되고, Julian 예보모델은 12개의 날짜인자가 추가되어 42개의 입력인자를 갖는다. 각각 Table 8와 Table 9에 Cos, S_Cos, Julian 예보모델의 입력인자와 네트워크 파라메타를 상세히 기술했다.

▷ Definition of Adjacent_Month

- Ⓐ if(Day = 15) then There is no Adjacent_Month
- Ⓑ else-if(Month = 1 and Day < 15) then Adjacent_Month = 12
- Ⓒ else-if(Month = 12 and Day > 15) then Adjacent_Month = 1
- Ⓓ else-if(Day < 15) then Adjacent Month = Month - 1
- Ⓔ else-if(Day > 15) then Adjacent Month = Month + 1

▷ Month_value

- Ⓐ if(Day < 15) then Month_Value=

$$\frac{1}{28} \times Day + \frac{13}{28}$$
- Ⓑ else-if(Day > 15) then Month_Value=

$$-\frac{1}{30} \times Day + \frac{3}{2}$$
- Ⓒ Day = 15, Month_value=1

▷ Adjacent_Month_value

Adjacent_Month_value = 1 - Month_value

Fig. 6. Definition of Month and Adjacent Month.

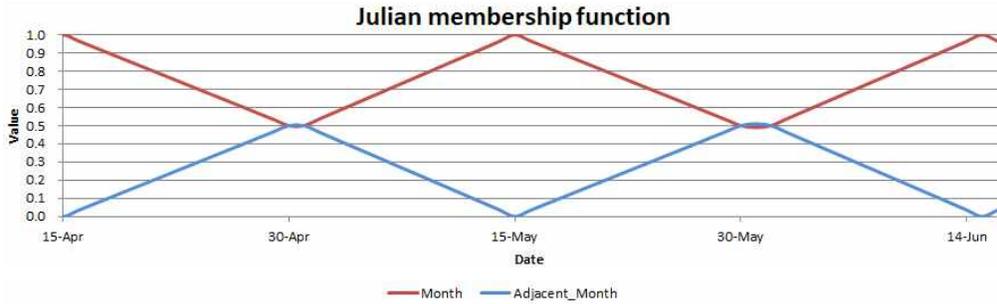


Fig. 7. Julian membership function.

4. 실험 결과 및 고찰

4.1 실험환경 및 실험 데이터

본 연구에서 개발한 5개의 예보모델과 평가 및 분석 프로그램은 windows 10 기반의 visual c++ 2010을 사용하여 구현했으며, 실험에 사용한 데이터는 2015년 01월 01일부터 2017년 12월 31일까지의 서울 권역 6시간 평균 데이터로 예보대상 물질은 PM₁₀이다. 이 중 학습에 2015년 01월 01일부터 2016년 12월 31일까지의 자료를 사용하고, 평가에는 2018년 01월 01일부터 2018년 12월 31일까지의 데이터를 이용했다.

각 DNN 예보모델에 사용한 인자는 상이하나, 과거의 대기질 및 기상 측정 자료와 예보구간의 대기질 및 기상 예보자료를 사용했다. 각 예보모델별로 사용한 인자는 2장과 3장에서 상세히 기술했다.

4.2 실험 결과

Table 10과 Fig. 8에 5개의 DNN 예보모델과 수치 예보모델(CASE4)의 예보 결과를 일별로 나타냈다. 예보 성능은 지수적중률(Accuracy), 감지확률(POD), 오경보율(FAR)로 평가했는데, 모델에 따라 일별, 평가지표별로 좋고 나쁨의 성능 차이가 있으나, 예보에

Table 8. Input Data of Cos, S_Cos and Julian

Time	Kind		Data
T5	Observation data(14)	Weather	U, V, Pressure, Temperature, Dew point temperature, Relative humidity, Precipitation, Solar radiation
		Air quality	PM _{2.5} , O ₃ , NO ₂ , CO, SO ₂ , PM ₁₀
T# (#:6~15)	Forecast data(16)	Air quality	CASE4_PM ₁₀
		Weather	Temperature, Pressure, Relative humidity, Mixed highs, U, V, Geopotential height_850hpa, Pressure_850hpa, U_850hpa, V_850hpa, Relative humidity_850hpa, Temperature_850hpa, Geopotential height_925hpa, U_925hpa, V_925hpa, Temperature difference_850-925
T# (#:6~15)	Date(1/1/12)		Cos: 1 factor expressed by the Cos function
			S_Cos: 1 factor expressed by the S_Cos function
			Julian: 12 factor expressed by the Julian membership function

Table 9. Network parameter of Cos, S_Cos and Julian

Model	Input layer	Hidden layer1	Hidden layer2	Hidden layer3	Hidden layer4	Output layer	Learning rate	Epoch
Cos	31	15	8	4	2	1	0.009	160000
S_Cos	31	15	8	4	2	1	0.009	160000
Julian	42	21	10	5	2	1	0.009	160000

서 가장 중요한 구간인 D+1과 감지확률의 우수성 등을 종합적으로 판단했을 때, Julian 예보 모델이 가장 우수한 결과를 보였다.

ORIM은 D+0 구간에서는 지수적중률 76%, 감지확률 83%, 오경보율 26%로 우수한 결과를 보이나, D+1과 D+2 구간에서는 지수적중률과 감지확률이 60%대로 크게 떨어지고, 오경보율이 60% 이상으로 크게 증가한다. 이것은 지나치게 긴 구간의 과거 측정 자료를 학습에 포함함으로써 오히려 예보 성능을 저하시킨 요인으로 분석된다.

RM14는 이를 개선하기 위해서 ORIM 모델에서

Table 10. PM10 prediction result for Seoul by DNN models (unit: %)

Model	Day	Accuracy	POD	FARM
ORIM	D+0	76	83	26
	D+1	63	69	61
	D+2	60	65	73
RM14	D+0	82	79	14
	D+1	76	50	33
	D+2	69	53	44
Cos	D+0	80	79	17
	D+1	76	63	33
	D+2	67	47	56
S_Cos	D+0	82	83	20
	D+1	72	63	38
	D+2	66	53	53
Julian	D+0	79	79	24
	D+1	70	75	48
	D+2	66	71	46
CASE4	D+0	72	54	32
	D+1	77	63	50
	D+2	75	47	47

과거 30시간의 측정 자료를 사용한 것을 축소하여 과거 6 시간의 측정 자료만 학습에 포함시켰다. 그 결과 RM14 모델은 지수적중률과 오경보율이 ORIM에 비해 개선되었다. 하지만, 감지확률이 79%, 50%, 53%로 좋은 성능을 보이지 못했다.

감지확률을 개선하기 위해 날짜인자를 추가한 예보모델은 Cos, S_Cos, Julian 모델로 Cos는 3장의 관련연구에서 소개한 방법을 적용한 예보모델이고, S_Cos은 Cos 함수를 고농도 발생 추이에 좀 더 적합하도록 개선한 방법이다. Cos는 RM14 대비 지수적중률은 -2%이내로 비슷하고, 오경보율은 D+2 구간에서 오히려 12% 증가했으며, 감지확률은 D+1 구간에서 63%로 13% 개선된 결과를 보인다. 이는 고농도 감지의 기대치가 70%이상이라는 기준으로 볼 때 목표에 도달하지 못한 결과이다. S_Cos는 Cos과 비교해 보면 지수적중률 ±4% 이내, 오경보율 ±5% 이내로 비슷하고, 감지확률이 D+2 구간에서 11% 향상되었다. Julian 예보 모델은 지수적중률이 79%, 70%, 66%로 S_Cos보다 3% 이내로 떨어지나 감지확률이 전 구간 70%대로 크게 향상되고, 오경보율도 전 구간 50% 이내로 기대치에 근접한 결과를 보였다. 또한, 수치모델인 CASE4와 비교했을 때도 지수적중률은 다소 떨어지나, 감지확률 및 오경보율이 우수하다. 이는 고농도와 상관도가 높은 날짜 데이터를 표현하는데 있어 제안한 줄리안 멤버십 함수가 효과적이고, 이를 적용한 Julian 예보모델이 고농도 감지에 적합한 모델임을 증명할 수 있는 결과이다.

5. 결론

본 논문에서는 서울 권역의 PM₁₀ 예보를 위해서 DNN을 기반으로 한 5가지 예보 모델을 개발했다. ORIM는 과거의 30시간 동안의 측정데이터와 예보 구간의 예보 데이터를 사용한 모델로 D+0의 예보 성능은 우수하나, 하루를 초과하는 과거 데이터를 사용함으로 인해 D+1 이상 구간의 성능이 떨어지고, 특히 오경보율이 증가하는 문제점을 보였다. 이 문제를 해결하기 위해서 RM14 모델에서는 과거 6시간의 측정 데이터만 사용했다. 그 결과 지수적중률과 오경보율이 개선되었으나 감지확률은 D+1이상의 구간에서 50%대로 기대치에 미치지 못했다. 이것은 학습데이터에 고농도를 감지하는 인자가 미비하여 발생하는 현상으로 이를 개선하기 위해 고농도의 발생과 밀접

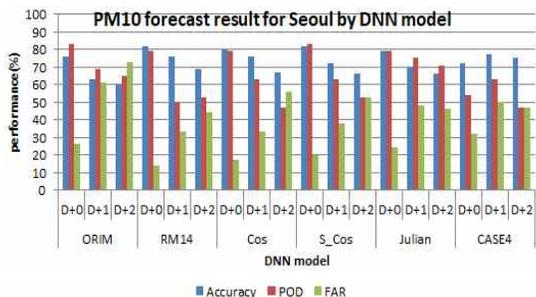


Fig. 8. PM10 prediction result for Seoul by DNN models.

한 영향을 미치는 날씨 인자를 추가하여 Cos, S_Cos, Julian 모델을 개발했다. Cos과 S_Cos는 기본적으로 날씨 데이터를 cos 함수 값으로 표현한 방법으로 RM14와 비교해 볼 때, 지수적증률은 비슷하고, S_Cos의 경우 감지확률이 좀 더 개선되었다. 하지만 여전히 예보에서 가장 중요한 구간인 D+1에서 60%대의 감지확률을 보였다. Julian 예보모델은 날씨를 표현하는데 제안한 줄리안 멤버십 함수를 사용해서 12개의 인자로 나타낸 모델로 S_Cos과 비교하면 지수적증률은 -3% 이내로 비슷하지만, 감지확률은 전 예보구간이 70%대로 최대 18% 성능이 향상되었으며, 오경보율도 전 구간 50% 이내로 기대치에 부합하는 결과를 보였다.

고농도의 농도와 일수가 점점 증가하는 현 대기 상황으로 볼 때, 미세먼지 예보에 있어서 가장 중요한 성능은 고농도를 잘 감지하는 것으로 제안한 Julian 예보모델은 이러한 요구에 충족하여 예보 모델로써 활용도가 높을 것으로 사료된다.

향후 과제로는 고농도 감지에 효율적인 학습인자의 추가 선별과 예보 성능 향상을 위해서 주변 권역의 자료를 이용하는 연구를 진행할 예정이다.

REFERENCE

- [1] NIER, *A Study of Data Accuracy Improvement for National Air Quality Forecasting(III)*, NIER-RP2016-248, 11-1480523-002809-01, 2016.
- [2] NIER, *A Study of Construction of Air Quality Forecasting System using Artificial Intelligence(I)*, NIER-SP2017-148, 11-1480523-0003221-01, 2017.
- [3] Y. Bengio, A. Courville, and P. Vincent, "Representation Learning: A Review and New Perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, Issue 8, pp. 1798-1828, 2013.
- [4] J. Schmidhuber, "Deep Learning in Neural Networks: An Overview," *Neural Network*, Vol. 61, Issue C, pp. 85-117, 2015.
- [5] G. Grivas and A. Chaloulakou, "Artificial Neural Network Models for Prediction of PM₁₀ Hourly Concentrations, in The Greater Area of Athens, Greece," *Atmospheric Environment*, Vol. 40, No. 7, pp. 1216-1229, 2006.
- [6] A.B. Chelani, D.G. Gajghate, and M.Z. Hasan, "Prediction of Ambient PM₁₀ and Toxic Metals using Artificial Neural Networks," *Journal of the Air and Waste Management Association*, Vol. 52, No. 7, pp. 805-810, 2002.
- [7] I.G. McKendry, "Evaluation of Artificial Neural Networks for Fine Particulate Pollution (PM₁₀ and PM_{2.5}) Forecasting," *Journal of the Air and Waste Management Association*, Vol. 52, No. 9, pp. 1096-1101, 2002.
- [8] A. Chaloulakou, G. Grivas, and N. Spyrellis, "Neural Network and Multiple Regression Models for PM₁₀ Prediction in Athens: A Comparative Assessment," *Journal of the Air and Waste Management Association*, Vol. 53, No. 10, pp. 1183-1190, 2003.
- [9] G. Corani, "Air Quality Prediction in Milan: Feed-Forward Neural Networks, Pruned Neural Networks and Lazy Learning," *Ecological Modelling*, Vol. 185, Issue 2-4, pp. 513-529, 2005.
- [10] M. Cai, Y. Yin, and M. Xie, "Prediction of Hourly Air Pollutant Concentrations Near Urban Arterials using Artificial Neural Network Approach," *Transportation Research Part D: Transport and Environment*, Vol. 14, No. 1, pp. 32-41, 2009.
- [11] W. Lu, W. Wang, X. Wang, S. Yan, and J.C. Lam, "Potential Assessment of A Neural Network Model with PCA/RBF Approach for Forecasting Pollutant Trends in Mong Kok Urban Air, Hong Kong," *Environmental Research*, Vol. 96, No. 1, pp. 79-87, 2004.
- [12] J. Fan, Q. Li, J. Hou, X. Feng, H. Karimian, and S. Lin, "A Spatiotemporal Prediction Framework for Air Pollution Based on Deep RNN," *Proceeding of ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume IV-4/W2, 2017 2nd International Symposium on Spatiotemporal Computing*, pp. 15-22, 2017.

- [13] H. Bae, S. Yu, and H. Kwon, "Fast Data Assimilation using Kernel Tridiagonal Sparse Matrix for Performance Improvement of Air Quality Forecasting," *Journal of Korea Multimedia Society*, Vol. 20, No. 2, pp. 363-370, 2017.
- [14] X. Feng, Q. Li, J. Hou, L. Jin and J. Wang, "Artificial Neural Networks Forecasting of PM_{2.5} Pollution Using Air Mass Trajectory Based Geographic Model and Wavelet Transformation," *Atmospheric Environment*, Vol. 107, pp. 118-128, 2015.



유 숙 현

1999년 안양대학교 컴퓨터공학과
학사
2002년 안양대학교 컴퓨터공학과
석사
2011년 안양대학교 컴퓨터공학과
박사

2012년~현재 안양대학교 정보통신공학과 조교수
관심분야: 패턴인식, 신경망, 영상처리, 병렬처리응용,
딥러닝



전 영 태

2019년 안양대학교 컴퓨터공학과
학사
2019년~현재 안양대학교 컴퓨터
공학과 석사 과정중
관심분야: 패턴인식, 신경망, 영
상처리, 병렬처리응용



권 희 용

1983년 서울대학교 전자계산기
공학과 학사
1985년 서울대학교 전자계산기
공학과 석사
1993년 서울대학교 컴퓨터공학과
박사

1986년~1995년 한국통신 연구개발단 선임연구원
1995년~현재 안양대학교 컴퓨터공학과 교수
관심분야: 패턴인식, 신경망, 영상처리, 병렬처리응용