



수치 모델링 및 머신러닝을 이용한 대기 오염 예측



수학과 2017010698 오서영

수학과 2018010705 신영민

목차

1
수행
개요

2
수행
과정

3
결과
분석

4
결론

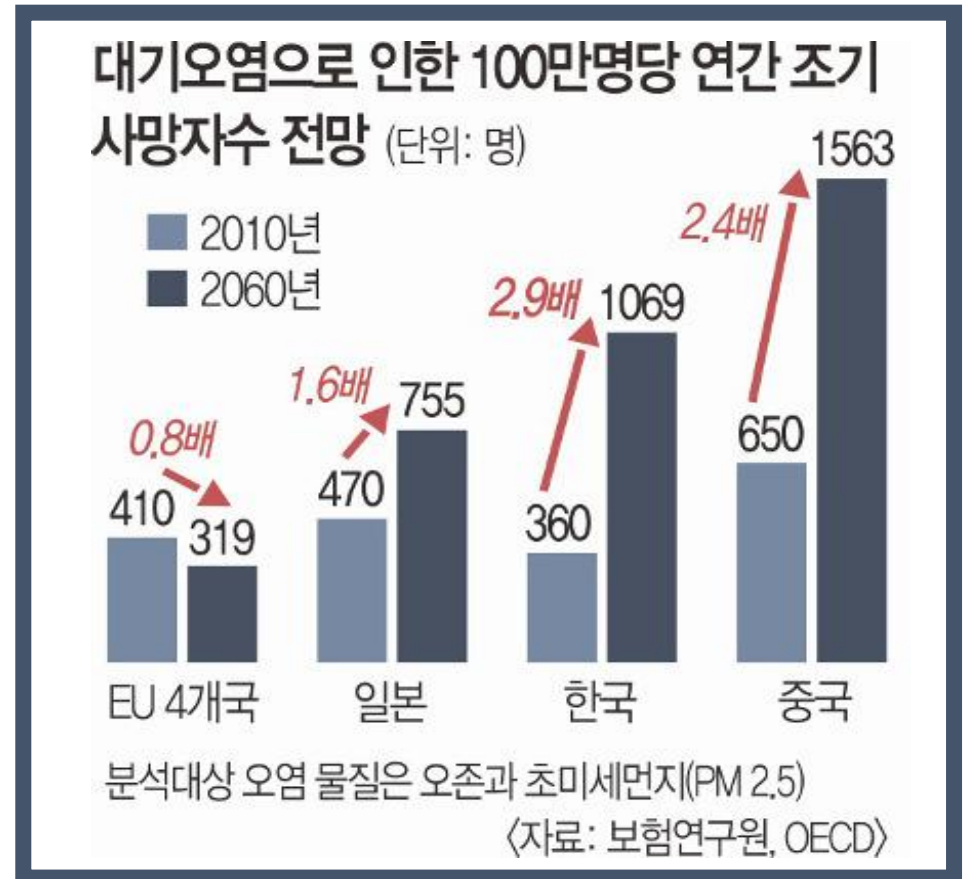
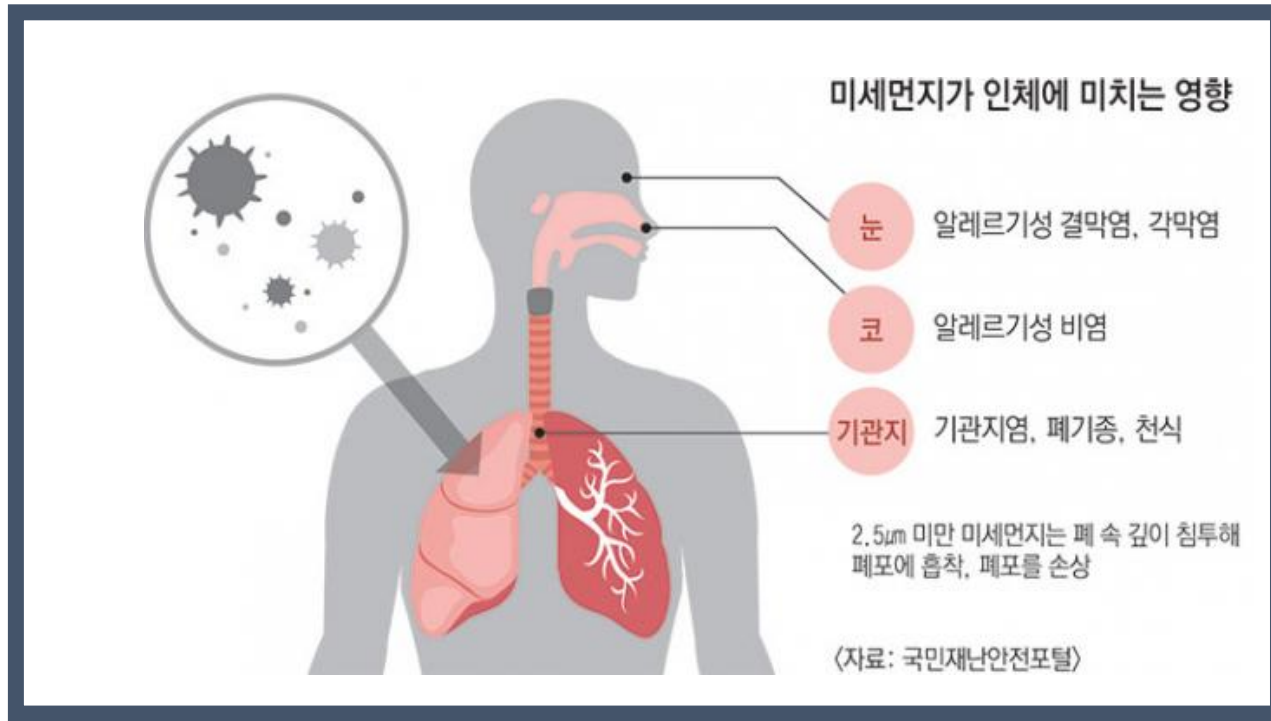
1. 수행개요

목적 및 필요성

미세먼지에 의한 사망률이 증가하는 추세



미세먼지 농도 **예측**의 필요성



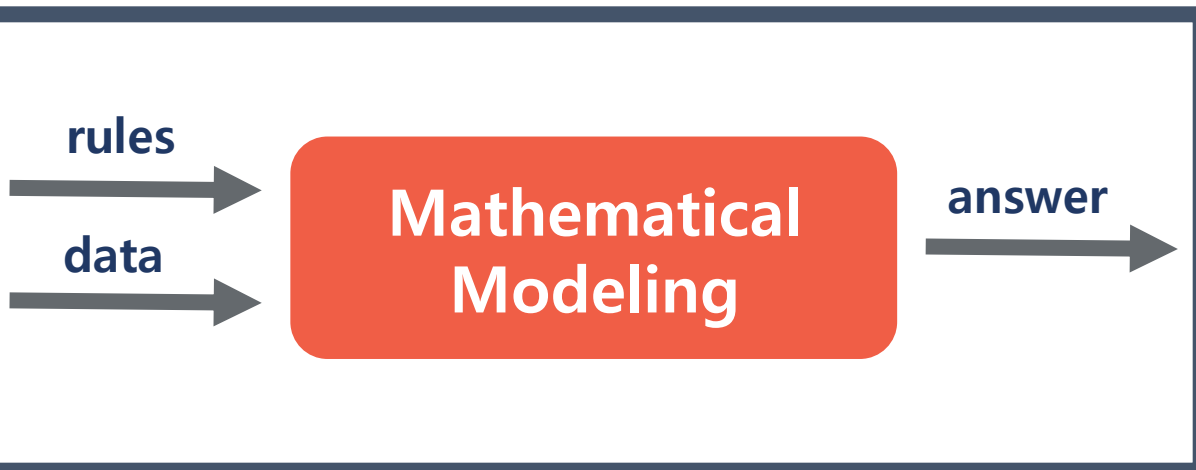
1. 수행개요

목표

두가지 방법론을 사용하여 미세먼지 농도 예측하기

수학적 모델링 (Mathematical modeling)

기계 학습 (Machine learning)



2. 수행과정

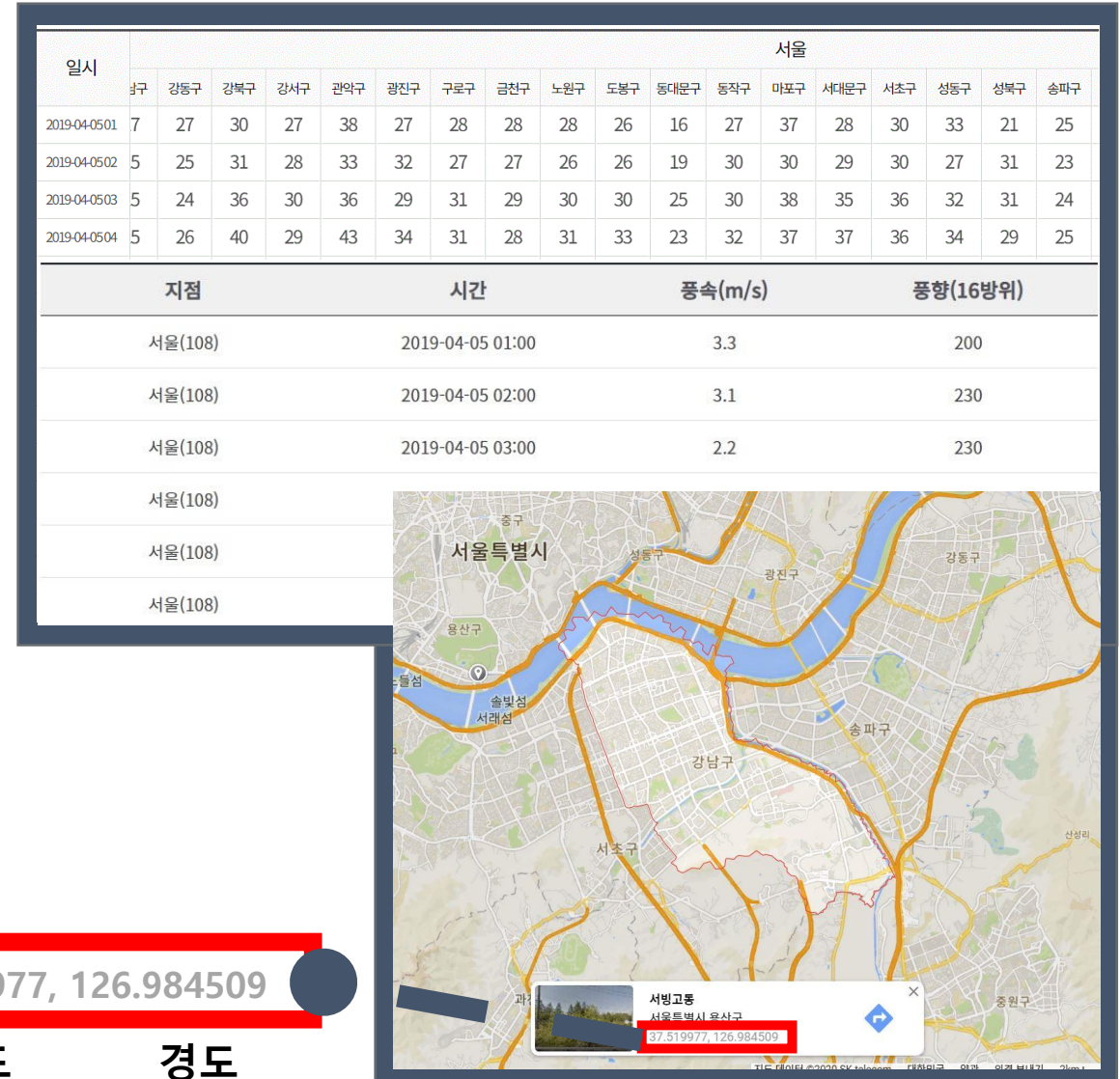
데이터 수집 및 정제

DATASET

지역별 시간당 미세먼지 농도 (Pm-10)

풍속, 풍향 데이터
(2019-04-05)

지역별 위도 경도 데이터



2. 수행과정

데이터 수집 및 정제

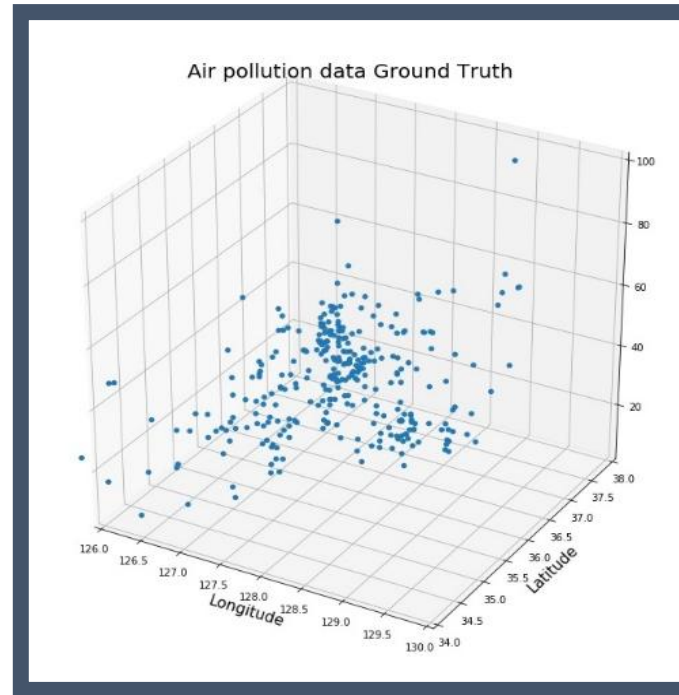
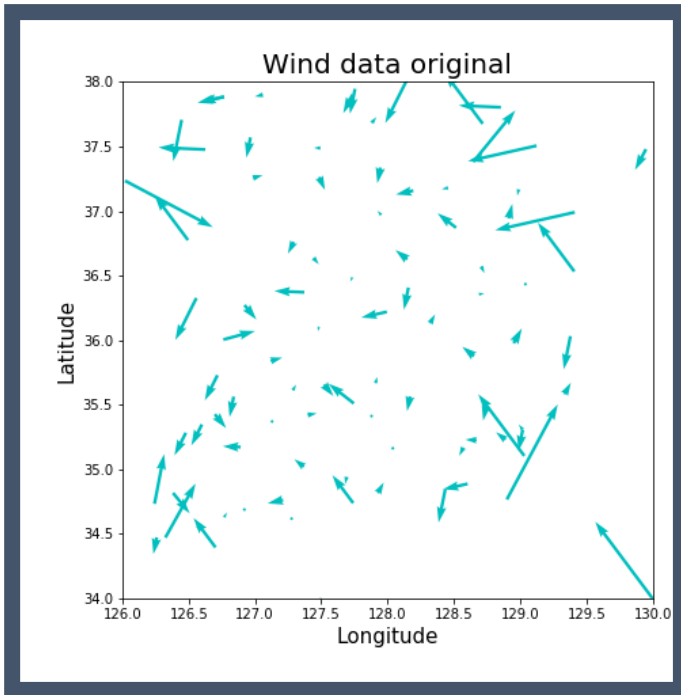
DATASET

데이터 정제 필요성

공간에 대한 통계자료를
모든 지점에서 획득하기는
현실적으로 불가능



보간법



1. 바람 데이터

3차 보간법 (Cubic Interpolation)

2. 미세먼지 데이터

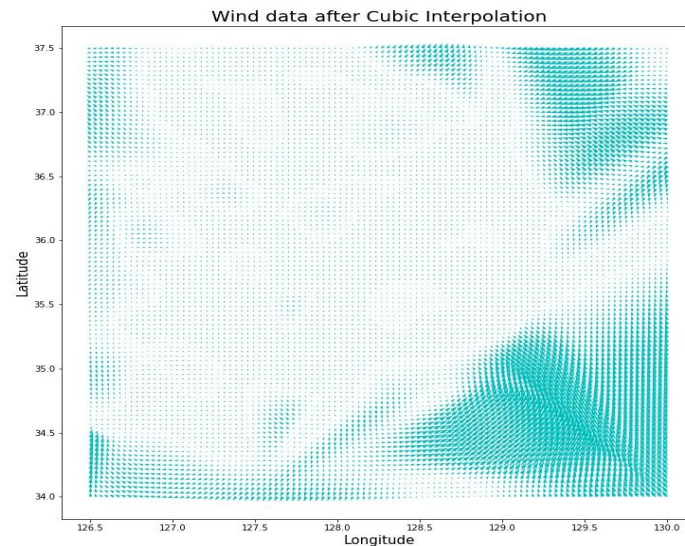
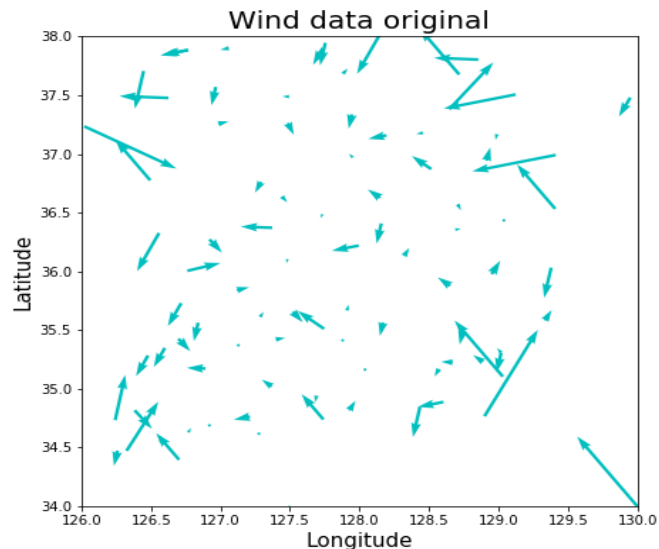
IDW(Inversed distance weighted)

$$\hat{u}(x) = \frac{\sum_{k=0}^N w_k(x) u_k}{\sum_{k=0}^N w_k}, w_k(x) = \frac{1}{d(x, x_k)}$$

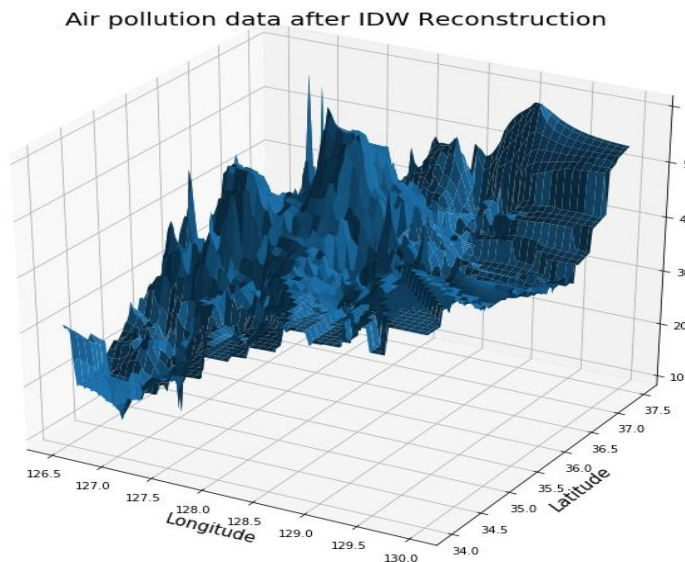
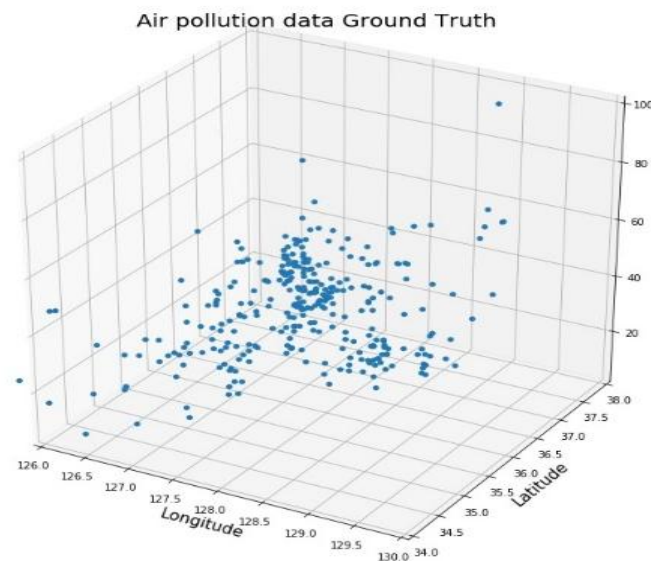
2. 수행과정

데이터 수집 및 정제

1. 바람 데이터 3차 보간법



2. 미세먼지 데이터 IDW



2. 수행과정

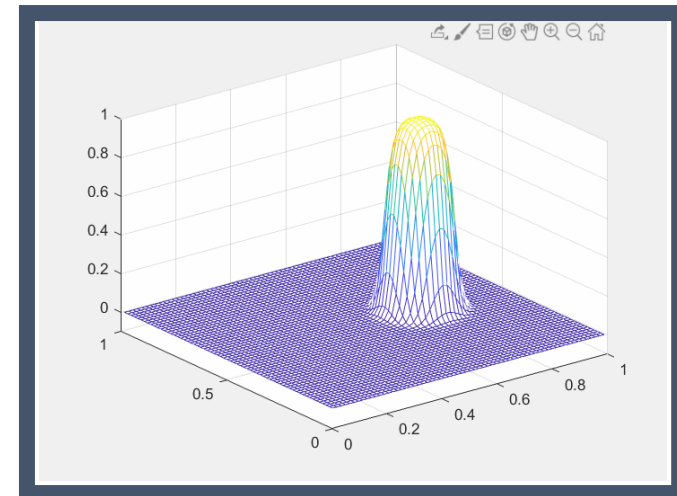
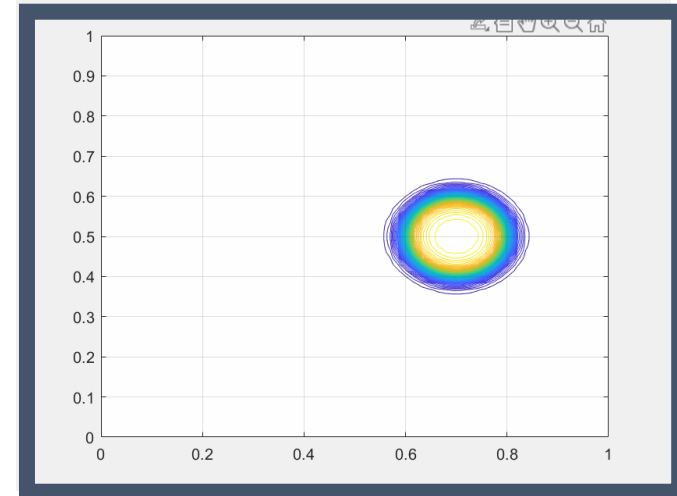
수학적 모델링

- 가정 : 물질들 사이의 화학적인 변화가 없다
- 공기 질 모델링 ~ 대류 + 확산

대류-확산방정식 (Convection-Diffusion equation)

$$\text{대류 방정식 : } \frac{\partial c}{\partial t}(x, t) + \frac{\partial(uc)}{\partial x}(x, t) = 0$$

$$\text{확산 방정식 : } \frac{\partial c}{\partial t}(x, t) = D \frac{\partial^2 c}{\partial x^2}(x, t)$$

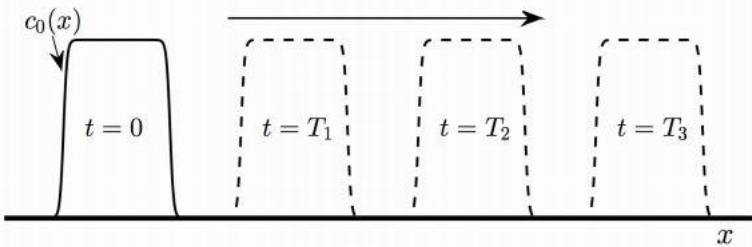


2. 수행과정

수학적 모델링

대류 방정식

$C_0(x)$: 초기 물질 농도분포



상수 u 의 속도로 물질이 흘러가고 있다면 물질의 분포는

$$C(x, t) = C_0(x - ut)$$

로 쓸 수 있다.

양변을 x 와 t 로 각각 편미분을 하면 연쇄법칙에 의하여 아래와 같은 결과가 나온다.

$$\frac{\partial C}{\partial x} = C'_0(x - ut)$$

$$\frac{\partial C}{\partial t} = C'_0(x - ut)(-u)$$

이를 정리하면,

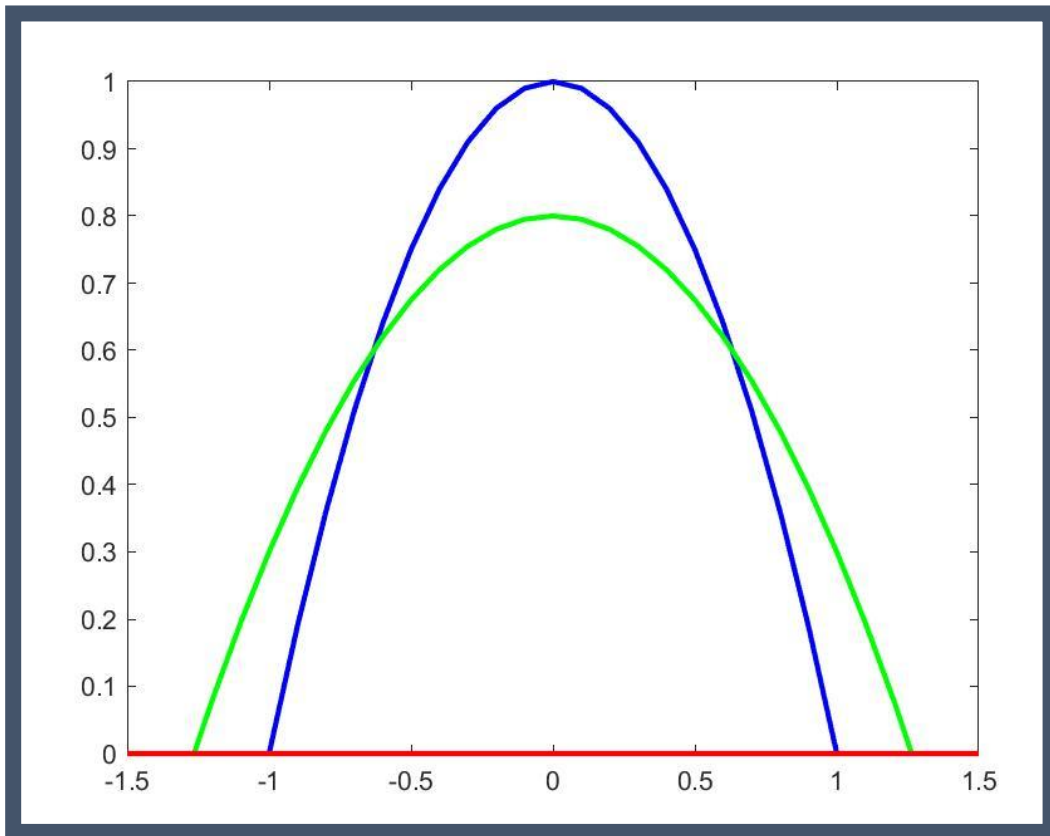
$$\frac{\partial C}{\partial t}(x, t) + \frac{\partial(uC)}{\partial x}(x, t) = 0$$

와 같은 대류 방정식을 얻을 수 있다.

2. 수행과정

수학적 모델링

확산 방정식



$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial x^2} \quad (D : \text{확산계수})$$

$$\frac{\partial C}{\partial t} \rightarrow 0 \quad \frac{\partial^2 C}{\partial x^2} \rightarrow 0$$

2. 수행과정

수학적 모델링

대류-확산방정식 (Convection-Diffusion equation)

$c(x, y, t)$ 와 $(u(x, y), v(x, y))$ 를 2차원 공간 (x, y) 와 시간 t 에서의
어떤 물질의 농도와 속도장이라고 하면

$$\begin{aligned} \frac{\partial c(x, y, t)}{\partial t} + \frac{\partial}{\partial x} [u(x, y)c(x, y, t)] + \frac{\partial}{\partial y} [v(x, y)c(x, y, t)] \\ = D \left[\frac{\partial^2 c(x, y, t)}{\partial x^2} + \frac{\partial^2 c(x, y, t)}{\partial y^2} \right] \end{aligned}$$

2. 수행과정

수학적 모델링

대류-확산방정식 (Convection-Diffusion equation)

이산화를 위한 차분 공식

$$\text{테일러 전개 : } f(x+h) = f(x) + hf'(x) + h^2 \frac{f''(x)}{2!} + h^3 \frac{f'''(x)}{3!} + \dots$$

1계 중앙차분

$$*f(x+h) = f(x) + hf'(x) + h^2 \frac{f''(x)}{2!} + h^3 \frac{f'''(\xi)}{3!}$$

$$*f(x-h) = f(x) - hf'(x) + h^2 \frac{f''(x)}{2!} - h^3 \frac{f'''(\xi)}{3!}$$

위 두식을 빼면

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h} + O(h^2)$$

2. 수행과정

수학적 모델링

대류-확산방정식 (Convection-Diffusion equation)

이산화를 위한 차분 공식

테일러 전개 : $f(x+h) = f(x) + hf'(x) + h^2 \frac{f''(x)}{2!} + h^3 \frac{f'''(x)}{3!} + \dots$

2계 중앙차분

$$*f(x+h) = f(x) + hf'(x) + h^2 \frac{f''(x)}{2!} + h^3 \frac{f'''(\xi)}{3!}$$

$$*f(x-h) = f(x) - hf'(x) + h^2 \frac{f''(x)}{2!} - h^3 \frac{f'''(\xi)}{3!}$$

위 두식을 더하면

$$f''(x) = \frac{f(x+h) - f(x) + f(x-h)}{h^2} + O(h^2)$$

2. 수행과정

수학적 모델링

대류-확산방정식 (Convection-Diffusion equation)

이산화한 결과

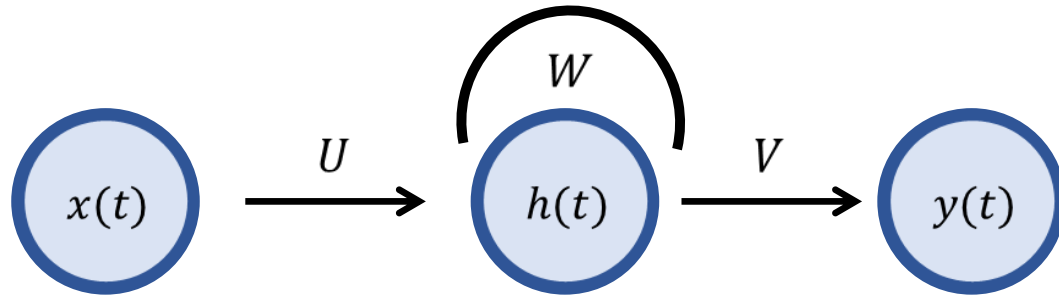
$$\begin{aligned} c_{ij}^{n+1} &= c_{ij}^n - \Delta t \left[\frac{cu_{i+1,j}^n - cu_{i-1,j}^n}{2h} + \frac{cv_{i,j+1}^n - cv_{i,j-1}^n}{2h} \right] \\ &+ D \frac{\Delta t}{h^2} [c_{i+1,j}^n + c_{i,j+1}^n - 4c_{ij}^n + c_{i-1,j}^n + c_{i,j-1}^n] \end{aligned}$$

2. 수행과정

머신러닝

RNN (Recurrent Neural Network)

시계열 데이터 예측에 적합한 딥러닝 모델



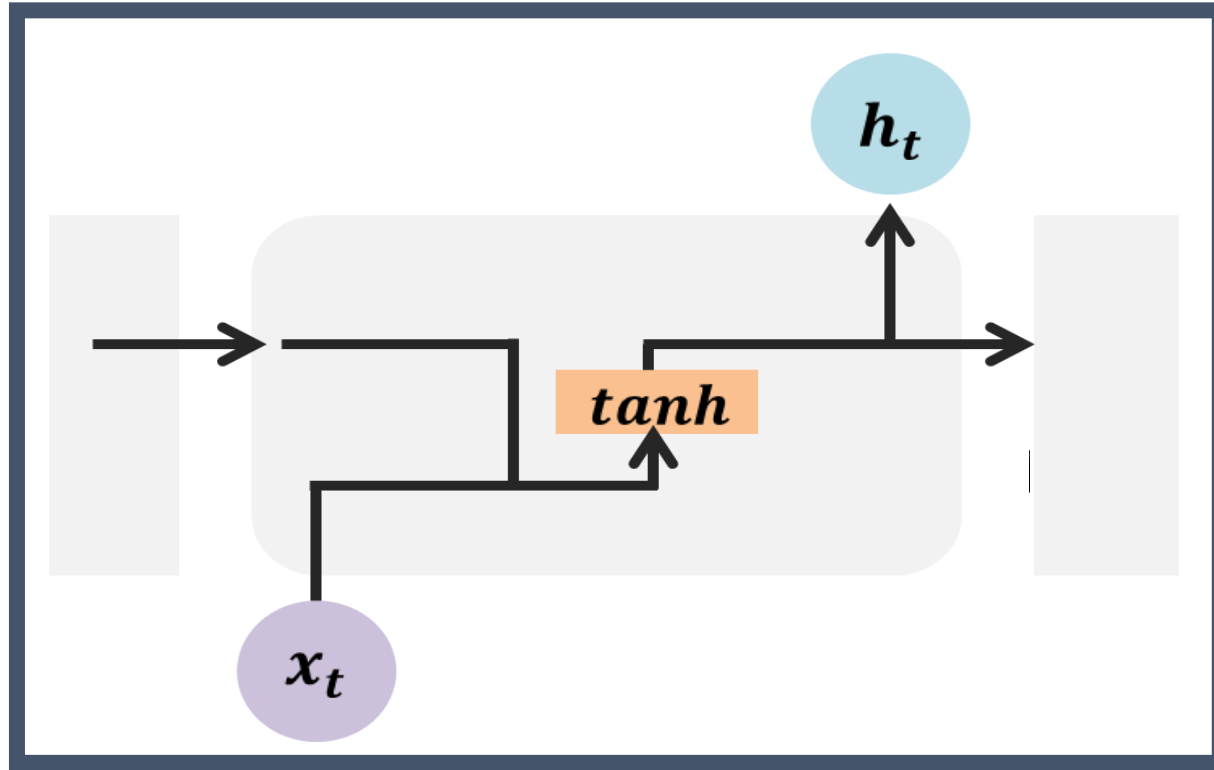
$$\begin{aligned}h_t &= f_w(h_{t-1}, x_t) \\ &= \tanh(w_{hh}h_{t-1} + w_{xh}x_t) \\ y_t &= w_{hy}h_t\end{aligned}$$

x_t : 현재 입력, h_{t-1} : 과거 기억, h_t : 현재기억

2. 수행과정

머신러닝

RNN



RNN (Recurrent Neural Network)



LSTM (Long Short Term Memory)

장기 의존성 문제

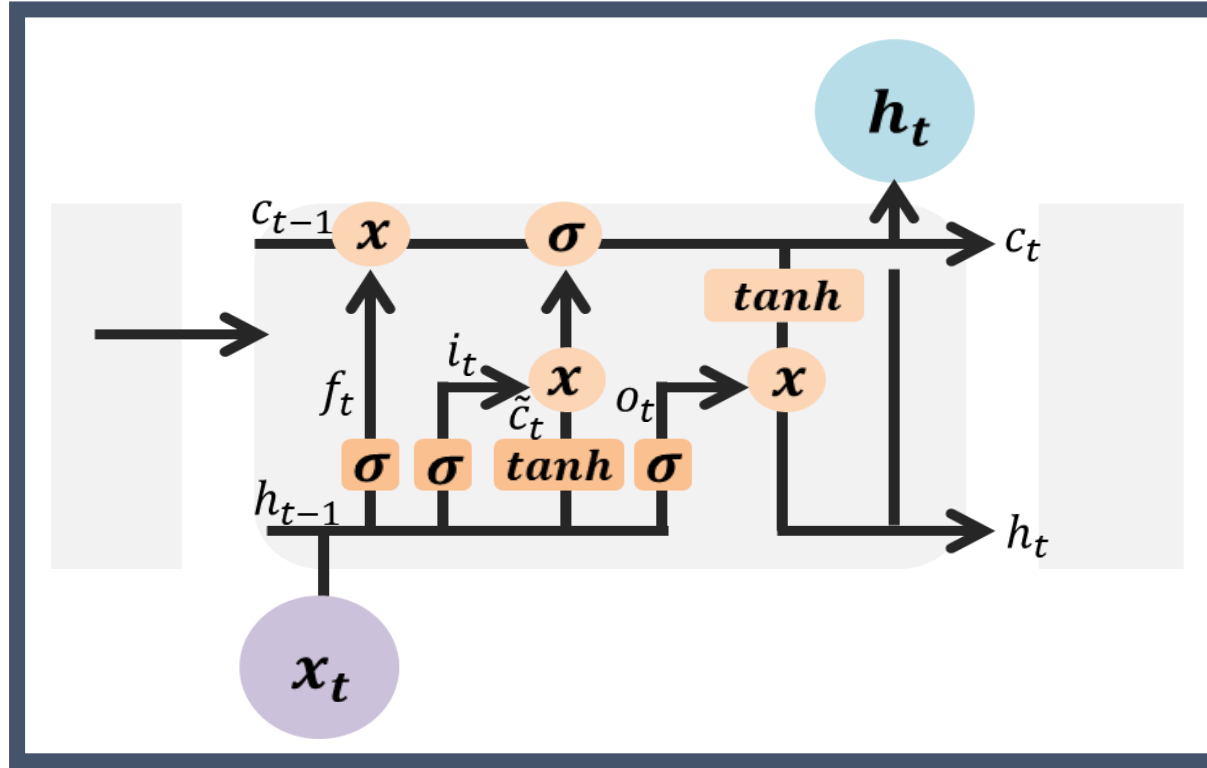
: 은닉층의 과거 정보가 마지막까지 전달되지 못하는 현상

장기 의존성 학습을 할 수 있는 RNN의 한 종류

2. 수행과정

머신러닝

LSTM



$$\begin{aligned} f_t &= \sigma(W_{xh_f}x_t + W_{hh_f}h_{t-1} + b_{h_f}) \\ i_t &= \sigma(W_{xh_i}x_t + W_{hh_i}h_{t-1} + b_{h_i}) \\ o_t &= \sigma(W_{xh_o}x_t + W_{hh_o}h_{t-1} + b_{h_o}) \\ \hat{c}_t &= \tanh(W_{xh_{\hat{c}}}x_t + W_{hh_{\hat{c}}}h_{t-1} + b_{h_{\hat{c}}}) \\ c_t &= f_t \odot c_{t-1} + i_t \odot \hat{c}_t \\ h_t &= o_t \odot \tanh(c_t) \end{aligned}$$

Gate : 정보가 전달될 수 있는 추가적인 방법

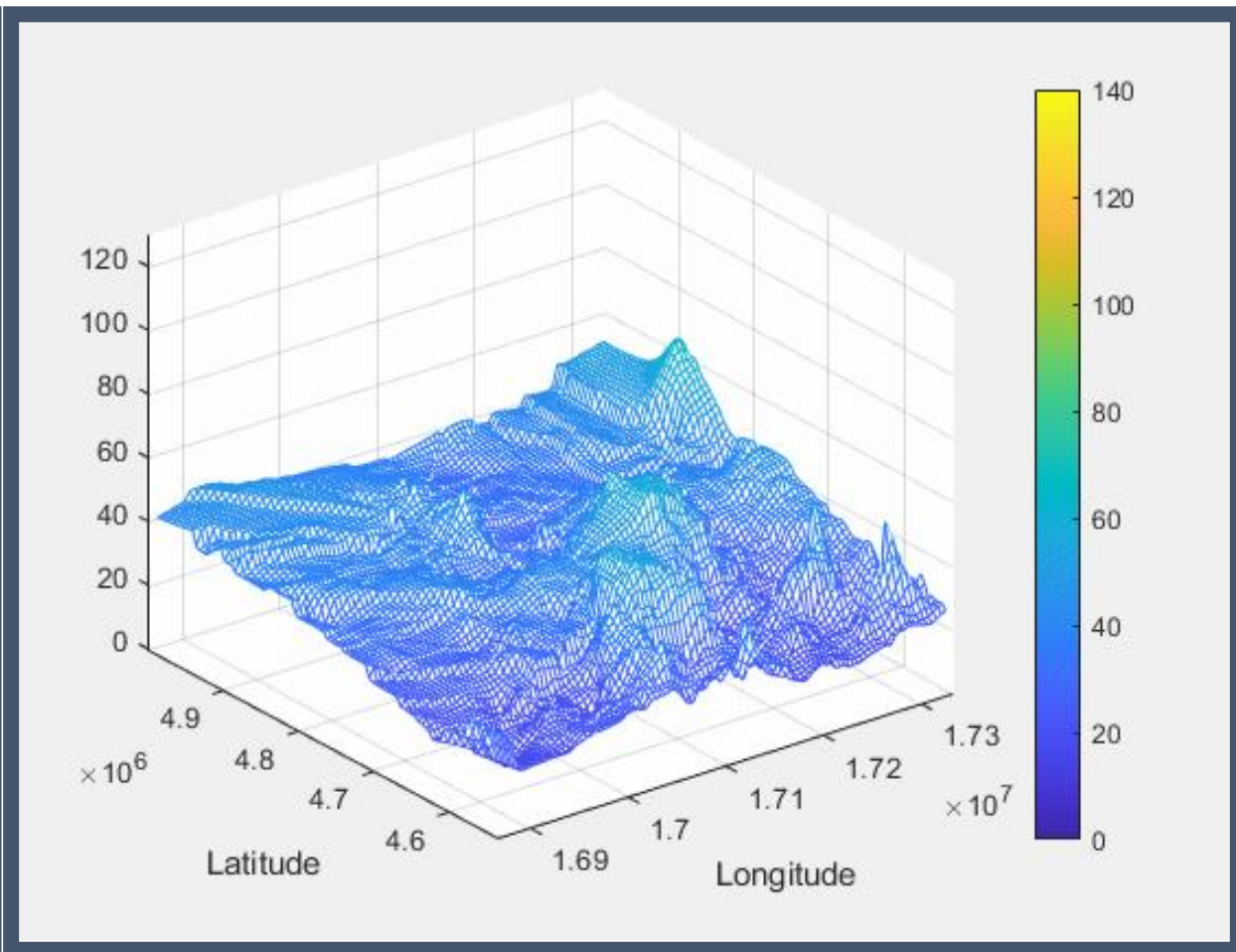
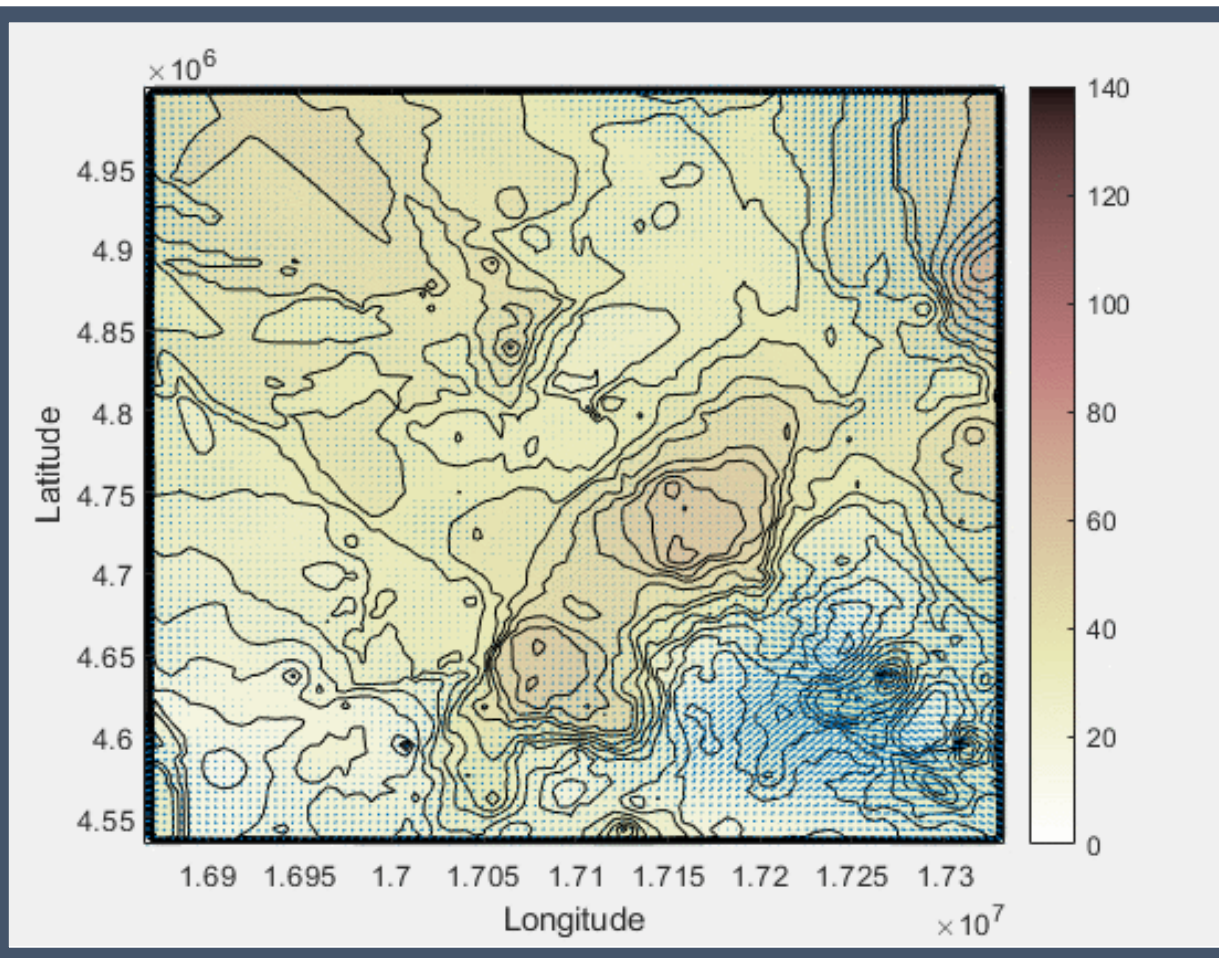


1. **Forget Gate** : 과거 정보를 잊기 위한 게이트
2. **Input Gate** : 현재 정보를 기억하기 위한 게이트
3. **Output Gate**

3. 결과 및 문제점

수학적 모델링 : Convection-Diffusion model

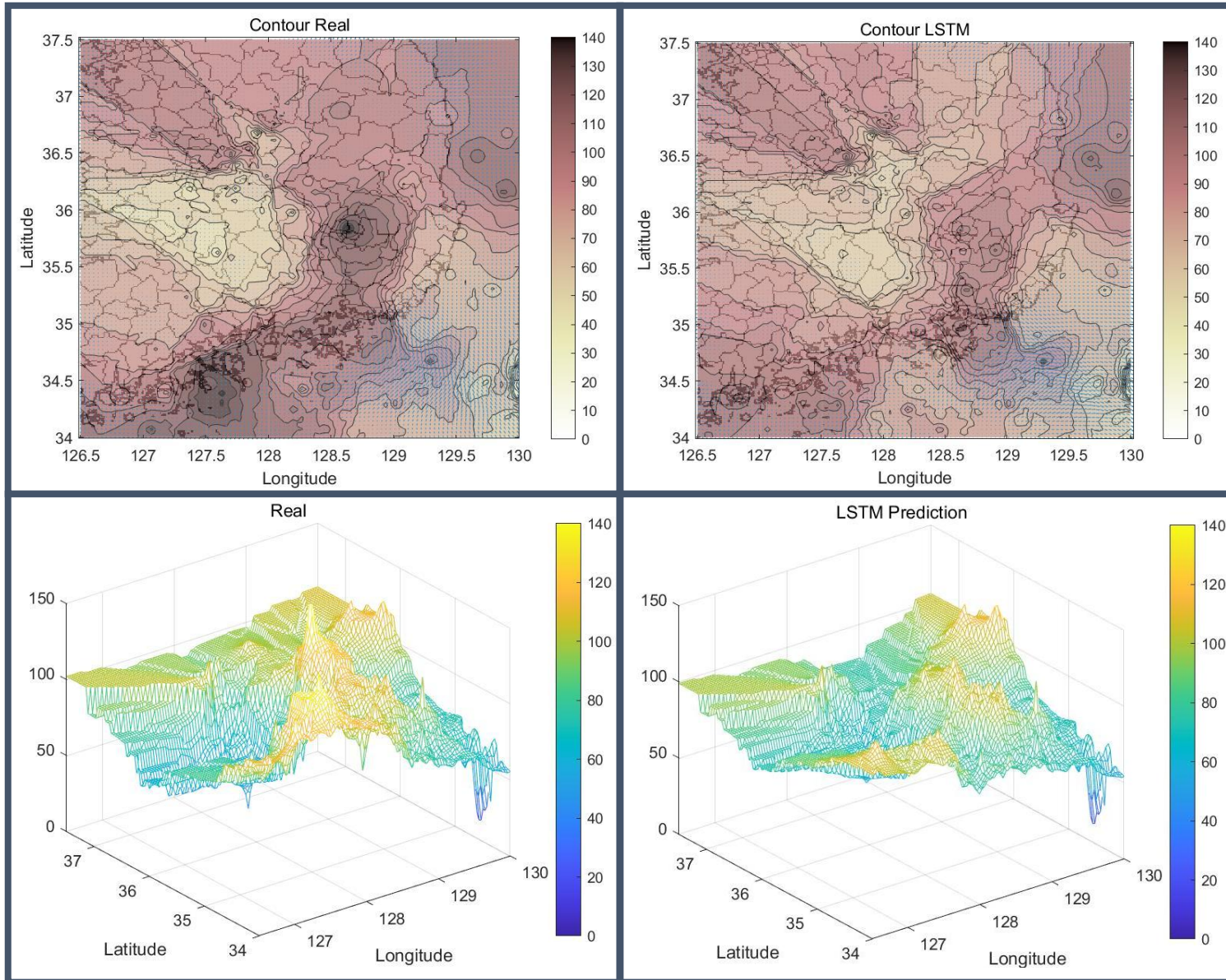
24시간 대기오염 농도



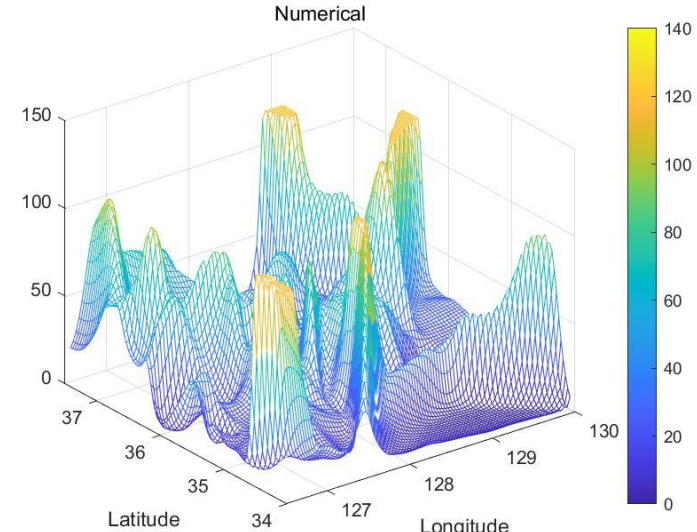
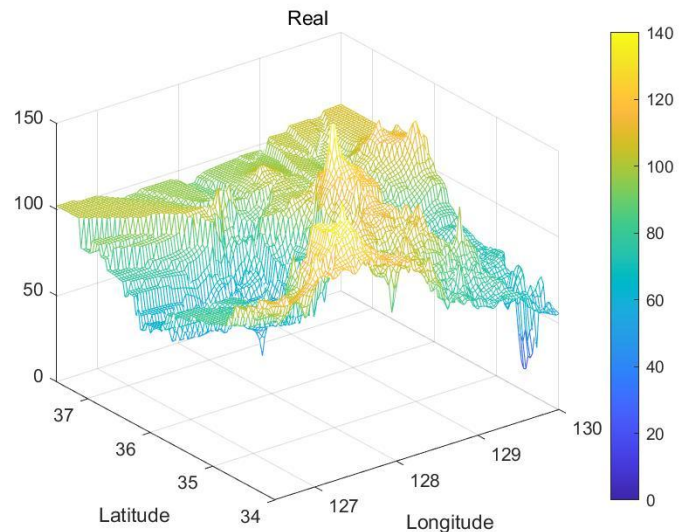
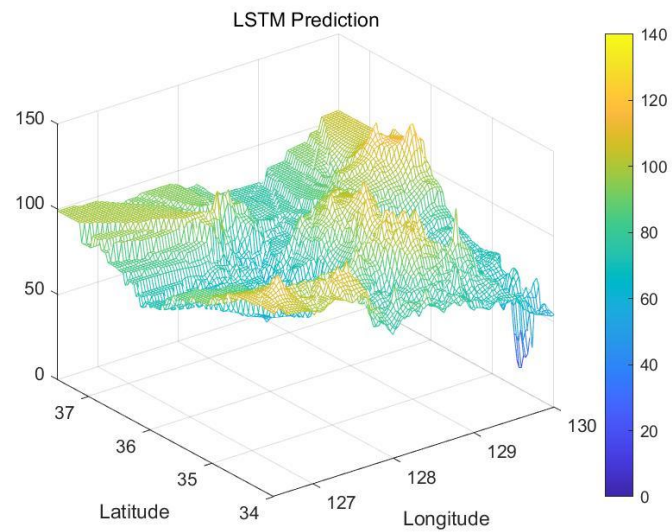
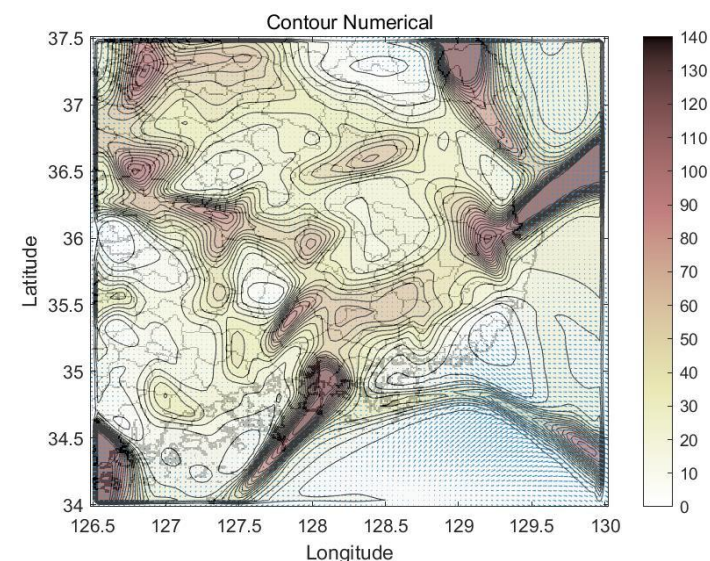
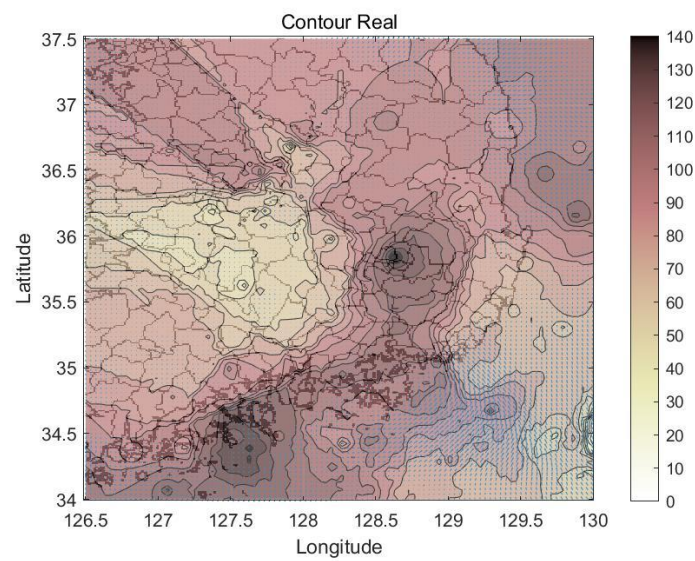
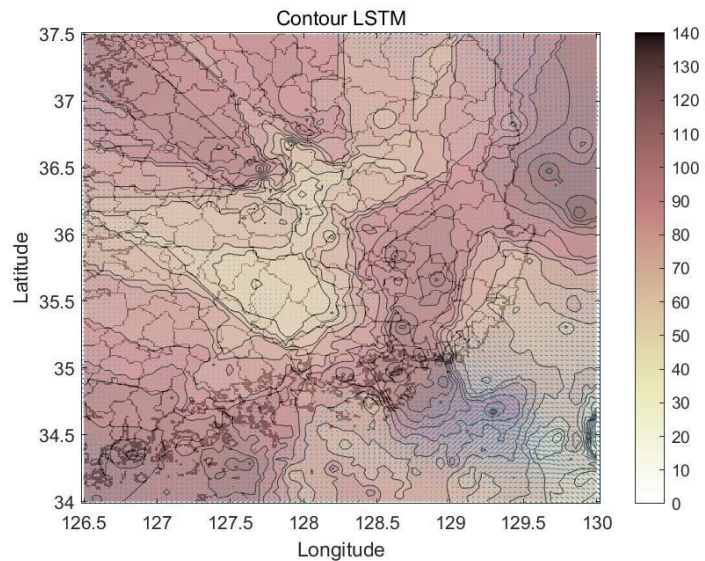
3. 결과 및 문제점

머신러닝 : LSTM

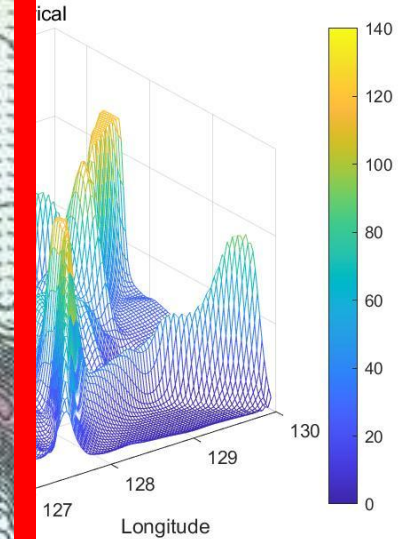
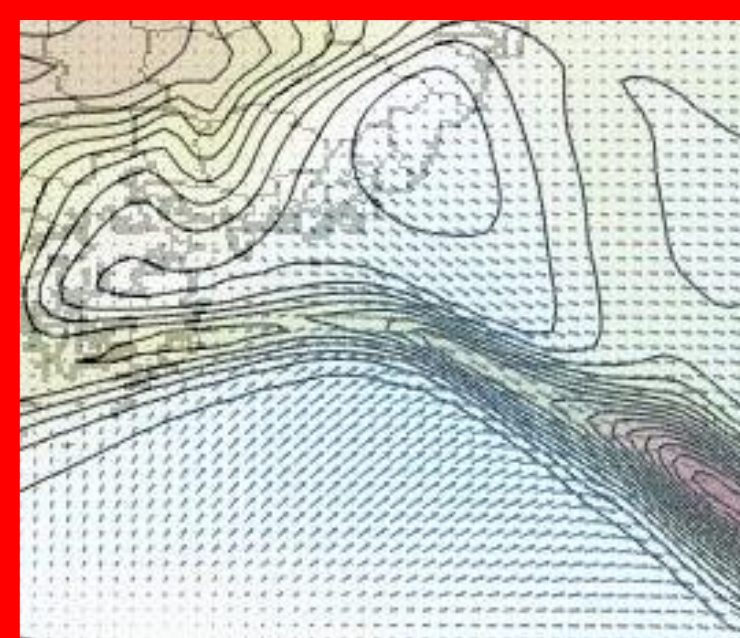
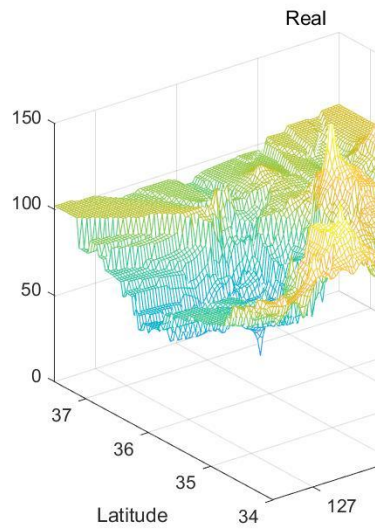
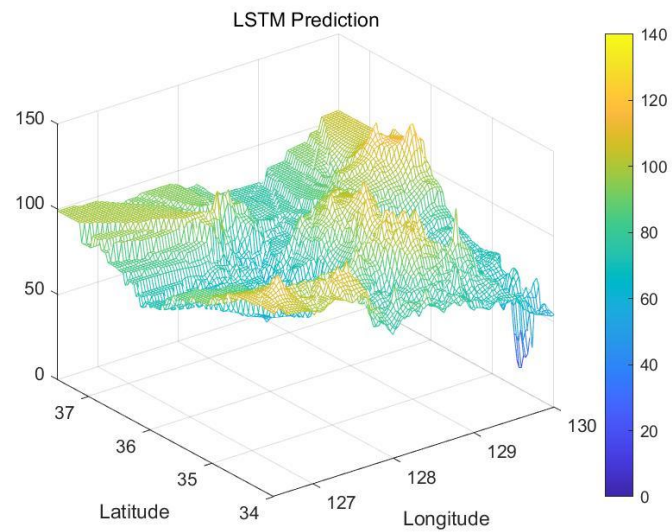
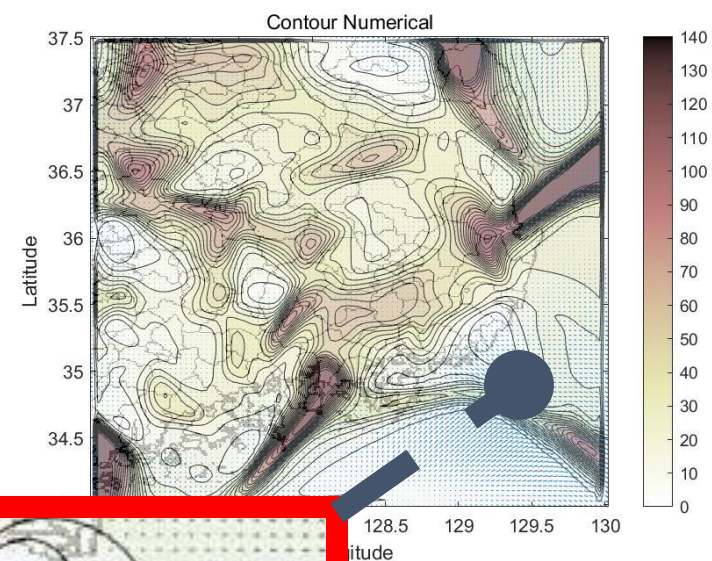
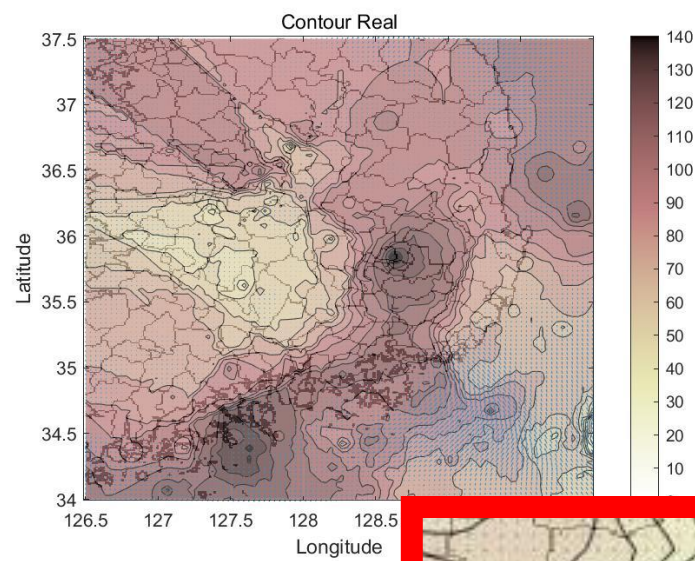
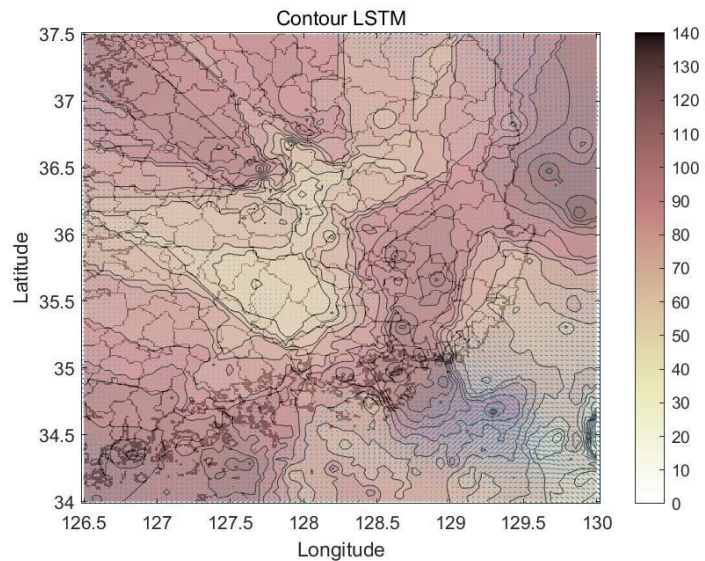
마지막 한시간에 대한
대기오염 농도



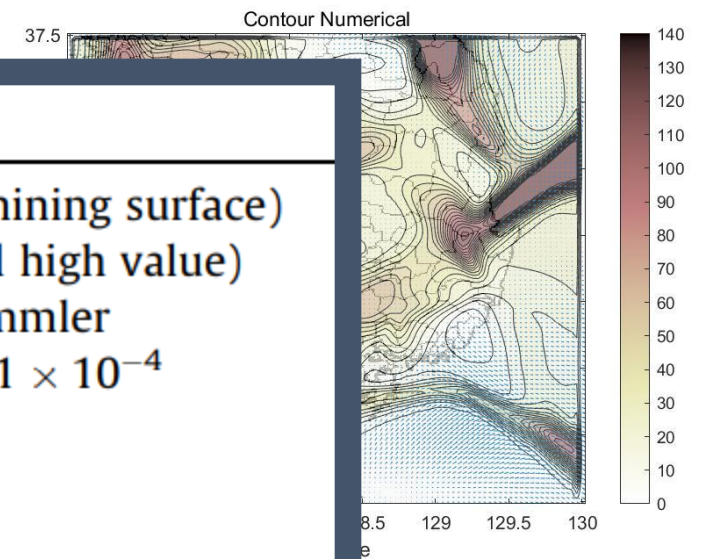
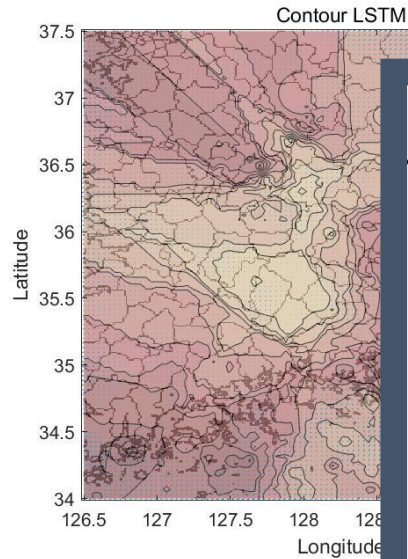
3. 결과 및 문제점



3. 결과 및 문제점



3. 결과 및 문제점



Property

Injection type

Density (kg/m^3)

Diameter distribution

Diameter range (m)

Mean diameter (m)

Spread parameter

Velocity (m/s)

Total flow rate (kg/s)

Duct dimension (m)

Tunnel high (m)

Tunnel length (m)

Tunnel width (m)

Fan diameter (m)

Fan position (exhaust) (m)

Fan position (blowing) (m)

Value

Surface (mining surface)

1400 (coal high value)

Rosin-Rammler

$2 \times 10^{-6} - 1 \times 10^{-4}$

1×10^{-4}

2.78

1

0.0062

0.6

2.9

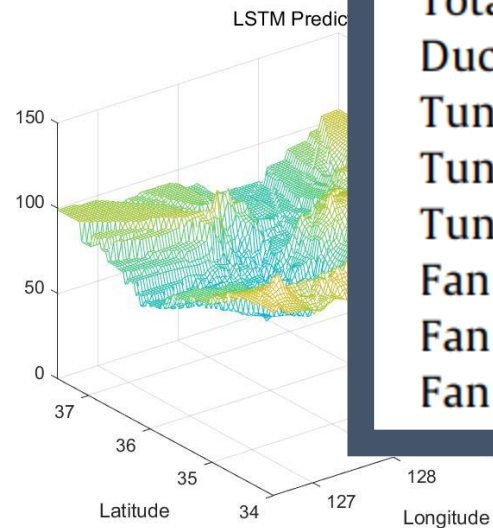
20

6

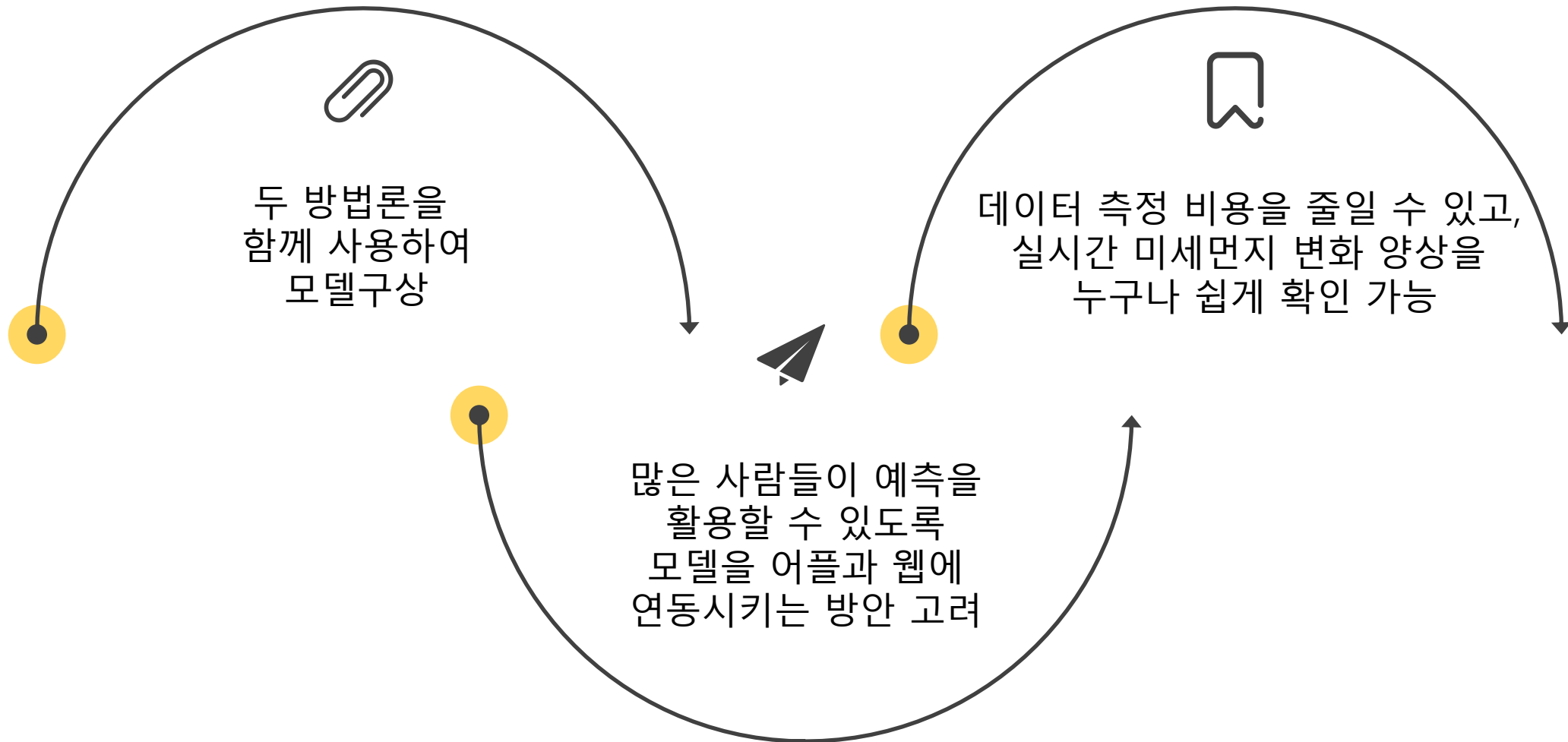
0.6

11.4, 1, 0.65

8.25, 1, 2



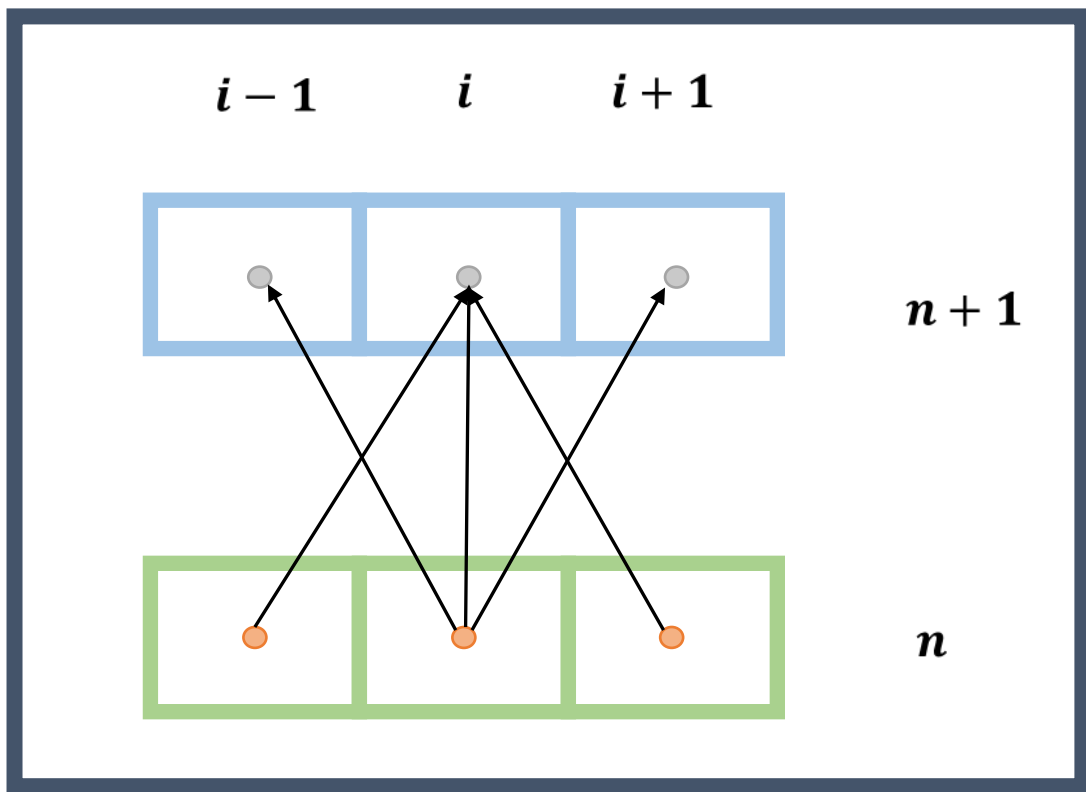
4. 결론





질의응답

확산 방정식



$$\begin{aligned}C_i^{n+1} &= C_i^n + kC_{i-1}^n - kC_i^n + kC_{i+1}^n - kC_i^n \\&= C_i^n + k(C_{i-1}^n - 2C_i^n + C_{i+1}^n)\end{aligned}$$

$$C_i^{n+1} - C_i^n = k(C_{i-1}^n - 2C_i^n + C_{i+1}^n)$$

$$\frac{C_i^{n+1} - C_i^n}{\Delta t} = \frac{k}{\Delta t} (C_{i-1}^n - 2C_i^n + C_{i+1}^n)$$

$$\frac{C_i^{n+1} - C_i^n}{\Delta t} = \frac{kh^2}{\Delta t} \frac{(C_{i-1}^n - 2C_i^n + C_{i+1}^n)}{h^2}$$

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial x^2} \quad \left(D = \frac{kh^2}{\Delta t}\right)$$

대류-확산방정식 (Convection-Diffusion equation)

경계 조건은 노이만 경계 조건 (Neumann boundary condition) 을 사용

