



2025학년도 1학기 SW 캡스톤디자인 경진대회

팀 명

GAIA (Gaussian AI Air-Impact Analysis)

지도교수

이경수

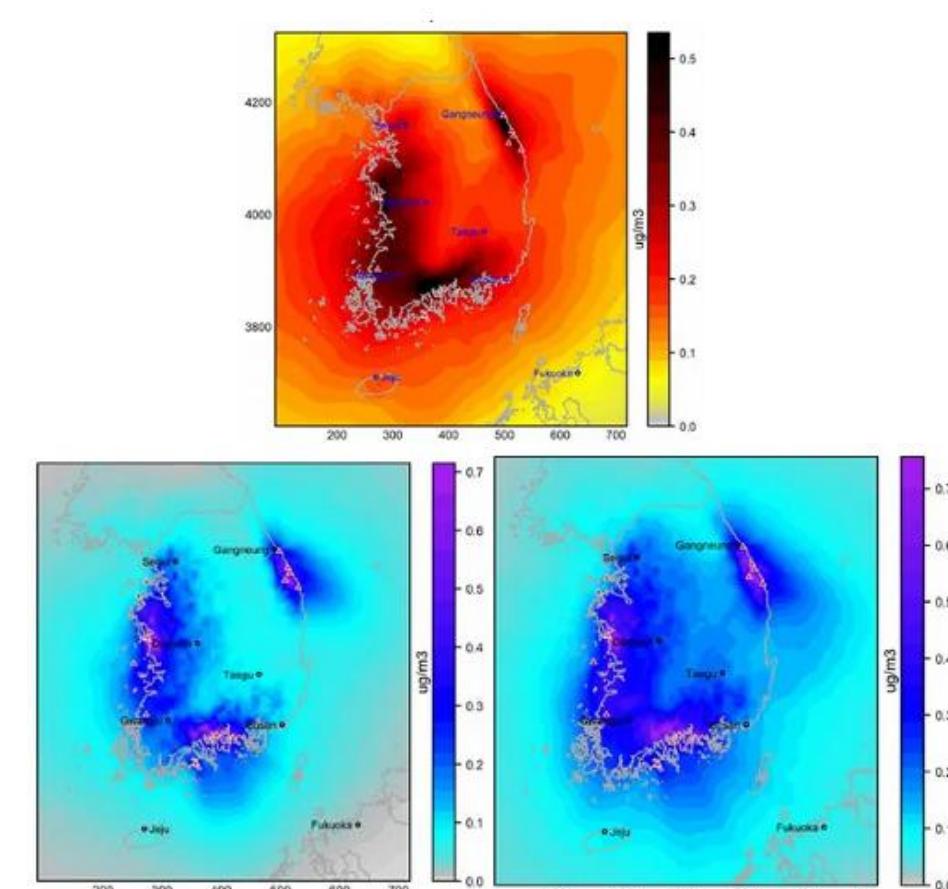
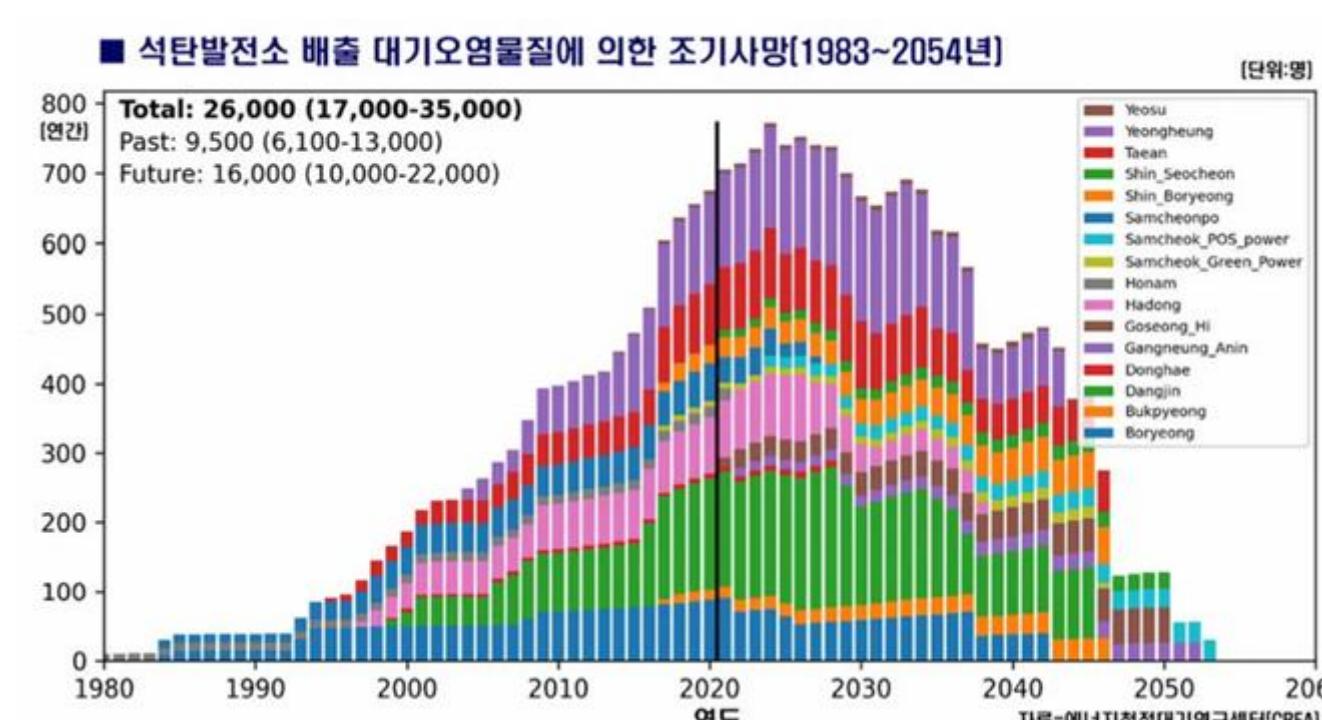
팀 원

산업체

개발 동기 및 목적

[개발 동기]

- 현대 사회의 급격한 산업화는 대기환경에 큰 악영향을 미치고 있으며, 특히 발전소와 공장에서 배출되는 대기오염 물질(SOx, NOx, PM 등)은 도시의 대기질 악화에 주요한 원인이 되고 있음



- 그러나 현재까지 사용되어 온 전통적인 확산 모델(예: Gaussian Plume Model, AERMOD 등)은 실시간 기상 변화, 지형 조건, 배출량의 동적 변화 등을 충분히 반영하지 못하여 정밀한 오염 예측과 정책 대응에 한계를 드러내고 있음
- 이러한 한계를 극복하기 위해, 최근 인공지능(AI) 기술과의 융합이 주목받고 있음
- 딥러닝 및 시계열 예측 모델(예: LSTM, Transformer)을 활용하면, 기상·지형·배출 데이터를 통합적으로 분석하여 보다 정밀하고 실시간성 있는 예측이 가능함

[목적]

1. AI 기반 가우시안 확산 모델 구축

- 물리 기반 모델(AERMOD 등)에 AI 예측 알고리즘(Transformer 기반 등)을 결합하여, 발전소 및 공장에서 배출되는 오염물질의 확산 경로를 실시간으로 정밀 예측하는 모델을 개발

2. 정책 수립을 위한 의사결정 지원 시스템 구현

- 수집된 실측 데이터(TMS, 에어코리아, 기상청 등)를 기반으로 대기질 분석 시각화 및 시뮬레이션 기능을 제공하고, 이를 통해 도시 환경 정책 수립 및 규제 검토 시 데이터 기반 근거를 제시

3. 웹 기반 확산 시각화 플랫폼 제공

- 사용자와 정책 담당자가 직관적으로 활용할 수 있도록 지도 기반 확산 시뮬레이션, 시계열 농도 변화 그래프 등을 제공하는 웹 플랫폼을 개발하여 실무 적용 가능성 강화

가우시안 확산 모델을 활용한 인공지능 기반 공장 및 발전소 배출가스의 도시 대기 영향 분석

오픈소스 URL : <https://github.com/GAI AinCapstone>

주요 기술

핵심 AI/ML 기술

Deep Learning Framework

- PyTorch 2.0+: Transformer 기반 시계열 예측 모델 구현
 - torch.nn.Transformer: 시계열 대기질 데이터 패턴 학습
 - Custom Attention Mechanism: 기상 조건과 오염물질 확산 패턴 간 상관관계 모델링
 - CUDA 최적화를 통한 GPU 가속 학습

Advanced AI Models

Transformer Architecture:

- Multi-Head Attention으로 다변량 시계열(기상, TMS, 대기질) 데이터 처리
- Positional Encoding을 통한 시간적 패턴 인식
- 7년간(2018~2024) 시계열 데이터 기반 학습

Gaussian Diffusion Models:

- 대기 오염물질 확산 패턴의 불확실성 모델링
- Noise Scheduling을 통한 점진적 예측 정밀도 향상
- Monte Carlo 샘플링 기반 확률적 예측
- 결과 데이터 자동 파싱 및 DB 저장

데이터 엔지니어링

고성능 데이터 처리

Pandas 2.0:

- 대용량 시계열 데이터 최적화 처리
- 멀티인덱싱을 통한 다차원 데이터 관리
- 메모리 효율적인 데이터타입 최적화

Numpy:

- 가우시안 확산 수치 계산 벡터화
- 기상 격자 데이터 interpolation

Scikit-learn:

- Feature Engineering (PCA, 정규화, 스케일링)
- 모델 검증 (Cross-validation, GridSearch)
- 예측 성능 평가 메트릭

전문 기상/환경 데이터 라이브러리

MetPy: 기상학적 계산 및 단위 변환

Xarray: 다차원 기상 데이터 (NetCDF) 처리

Shapely/GeoPandas: 지리공간 데이터 분석

개발 내용

1. 데이터 수집 및 전처리 시스템

다중 소스 데이터 통합 파이프라인

굴뚝자동측정기기(TMS) 데이터 수집 - 수집 항목: TSP, SO₂, NO_x, CO, HCl, HF, NH₃
(7개 주요 오염물질) - 실시간 API 연동을 통한 자동 데이터 수집 - 데이터 품질 검증 및 이상치 처리 알고리즘 구현

시계열 데이터 전처리

- 2018-2024년 7년간 시계열 데이터 표준화 및 정규화
- 결측치 처리: Forward Fill, Interpolation, LSTM 기반 imputation
- Feature Engineering: 시간적 특성(시간, 요일, 계절), 기상 파생 변수 생성
- Multi-scale 시계열: 시간별/일별/월별 집계 데이터 생성

2. 하이브리드 AI 예측 모델 개발

AERMOD + Transformer 융합 모델
딥러닝 아키텍처 설계

- Multi-Head Attention: 기상 조건과 오염물질 농도 간 복합 관계 모델링
- Positional Encoding: 시간적 주기성(일주기, 계절 변화) 반영
- Residual Connections: 깊은 네트워크의 gradient vanishing 문제 해결
- Dropout & Batch Normalization: 과적합 방지 및 안정적 학습

3. 실시간 웹 대시보드 구현

Streamlit 기반 인터랙티브 플랫폼

1. 실시간 대기질 모니터링 대시보드
2. 사용자 정의 시나리오 시뮬레이션
3. 3D 오염 확산 시각화
4. 발전소별 배출량 비교 분석
5. 예측 vs 실측 데이터 검증 차트

지리정보 시각화

- Folium 지도: 실시간 오염 농도 히트맵
- 등차선 맵핑: 농도 구배 시각화
- 3D 플롯 모델링: 오염물질 확산 경로 3차원 시각화
- 시간대별 애니메이션: 확산 패턴의 시간적 변화

결과 및 분석

[결과]

1. 사용자 상호작용 기반, 기존 오염물질 배출 정도 그래프화

발전소 별 오염물질 배출 정도 확인

2. 과거 데이터를 기반으로 AERMOD 구조에 맞게 트랜스포머 모델 학습

학습 결과 기반으로 다음과 같은 파라미터 조정하여 예상 예측 결과 분석 가능

- 화력발전소

- 대기종 오염물질

- 대기 측정소

- 기상 데이터 연도

3. 학습 결과 기반으로, 추후 국내 화력발전소가 추가될 경우 오염물질 확산 정도를 예측할 수 있는 시스템 구축

추가될 발전소의 위치 정보, 풍향, 풍속을 고려 → 대기 확산 정도 예측 가능

[분석]

1. 활용방안

- 환경 정책 수립 지원 가능
- 산업체 맞춤형 전략 수립에 기여
- 재난 대응 및 사고 시뮬레이션
- 연구 및 교육 자료로 활용 가능

2. 확장 가능성

- 국토 전역 또는 주요 산업 지역 대상 플랫폼 확대
- 드론 및 IoT 센서와 연동한 실시간 오염 추적 시스템 구성
- 기후변화 시나리오를 반영한 장기 대기질 예측 모델로 확장

3. 기대효과

- 환경정책 및 규제 평가를 위한 과학적 분석 도구 제공
- 산업 지역과 도시 간 대기오염 확산 문제 해결 지원
- 탄소중립 및 기후대응 정책 기반 마련