



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0038070
(43) 공개일자 2025년03월19일

- | | |
|---|---|
| <p>(51) 국제특허분류(Int. Cl.)
 <i>G01W 1/10</i> (2006.01) <i>G06N 3/042</i> (2023.01)
 <i>G06N 3/0464</i> (2023.01) <i>G06N 3/08</i> (2023.01)
 <i>G06Q 50/26</i> (2024.01)</p> <p>(52) CPC특허분류
 <i>G01W 1/10</i> (2013.01)
 <i>G06N 3/042</i> (2023.01)</p> <p>(21) 출원번호 10-2023-0120701
 (22) 출원일자 2023년09월11일
 심사청구일자 2023년09월11일</p> | <p>(71) 출원인
 경희대학교 산학협력단
 경기도 용인시 기흥구 덕영대로 1732 (서천동, 경희대학교 국제캠퍼스내)</p> <p>(72) 발명자
 유창규
 경기도 용인시 기흥구 덕영대로 1732 경희대학교 국제캠퍼스 공과대학 354호</p> <p>사접 타리크
 경기도 용인시 기흥구 덕영대로 1732, 경희대학교 국제캠퍼스 공과대학 275호
 (뒷면에 계속)</p> <p>(74) 대리인
 특허법인지담</p> |
|---|---|

전체 청구항 수 : 총 9 항

(54) 발명의 명칭 공기질 예측을 위한 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치 및 방법

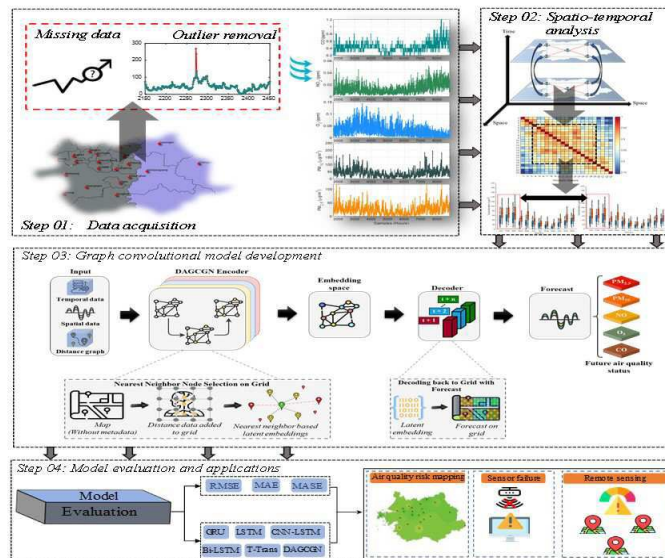
(57) 요약

본 발명은 공기질 모니터링 및 예측 기술에 관한 것으로서, 더욱 상세하게는 장기간 센서 고장 지역과 공기질 측정소 부재지역의 공기질 예측을 위한 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치 및 방법에 관한 것이다.

본 발명의 일 실시예에 따르면 기존 측정소에서 거리가 먼 지역을 원격 감지하여 공기질을 예측할 수 있는 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 시스템을 제공한다.

대표도 - 도16

600



(52) CPC특허분류

G06N 3/0464 (2023.01)

G06N 3/08 (2023.01)

G06Q 50/26 (2024.01)

G01W 2201/00 (2013.01)

(72) 발명자

김상윤

경기도 용인시 기흥구 덕영대로 1732, 경희대학교
국제캠퍼스 공과대학 275호

우대용

경기도 용인시 기흥구 덕영대로 1732, 경희대학교
국제캠퍼스 공과대학 275호

이 발명을 지원한 국가연구개발사업

과제고유번호	1711127448
과제번호	2021R1A2C2007838
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	개인기초연구(과기정통부)(R&D)
연구과제명	DeepAI 기반 하폐수처리장 스마트워터 자율운전 & 자율설계 원천기술 개발
기 여 율	1/2
과제수행기관명	경희대학
연구기간	2021.03.01 ~ 2026.02.28

이 발명을 지원한 국가연구개발사업

과제고유번호	1615013239
과제번호	RS2019KA152306
부처명	국토교통부
과제관리(전문)기관명	국토교통과학기술진흥원
연구사업명	지하철미세먼지저감기술개발사업
연구과제명	인공지능 기반 지하철 이용객의 (초)미세먼지 노출 저감기술 개발
기 여 율	1/2
과제수행기관명	한국철도기술연구원
연구기간	2023.01.01 ~ 2023.12.31

명세서

청구범위

청구항 1

인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치에 있어서,
측정소의 공기질 데이터를 입력하는 공기질 데이터 입력부;
상기 입력된 공기질 데이터를 전처리하는 공기질 데이터 전처리부;
상기 전처리한 공기질 데이터를 입력하여 측정소가 없는 지점의 예측 공기질 데이터를 생성하는 거리 적응형 그래프 컨볼루션 게이트 네트워크부; 및
상기 예측 공기질 데이터를 출력하는 예측 공기질 데이터 출력부를 포함하는
인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치

청구항 2

제1항에 있어서,
상기 거리 적응형 그래프 컨볼루션 게이트 네트워크부의 모델을 생성하는 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 생성부를 더 포함하는
인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치

청구항 3

제2항에 있어서,
상기 측정소의 공기질 데이터를 학습하여 거리 적응형 그래프 컨볼루션 게이트 네트워크를 개발하는 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 개발부; 및
상기 측정소의 공기질 데이터를 이용하여 상기 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델의 성능을 검증하는 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 검증부를 포함하는
인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치

청구항 4

제 3항에 있어서,
상기 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 개발부는,
상기 측정소의 공기질 데이터를 슬라이딩 윈도우 기법을 이용하여 과거 일정 시간 동안의 공기질 데이터를 데이터 세트로 구성하여 입력변수로 이용하는 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치

청구항 5

제1항에 있어서,
복수의 측정소의 공기질 데이터를 이용하여 상기 복수의 측정소의 공기질 데이터의 시공간 상관 관계를 분석하는 시공간 분석부를 더 포함하는 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치

청구항 6

인공지능-그래프 신경망 기반 스마트 공기질 모니터링 방법에 있어서,
 측정소의 공기질 데이터를 입력하는 단계;
 상기 입력된 공기질 데이터를 전처리하는 단계;
 상기 전처리한 공기질 데이터를 입력하여 측정소가 없는 지점의 예측 공기질 데이터를 생성하는 단계; 및
 상기 예측 공기질 데이터를 출력하는 단계를 포함하는
 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 방법

청구항 7

제6항에 있어서,
 상기 거리 적응형 그래프 컨볼루션 게이트 네트워크부의 모델을 생성하는 단계를 더 포함하는
 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 방법

청구항 8

제6항에 있어서,
 복수의 측정소의 공기질 데이터를 이용하여 상기 복수의 측정소의 공기질 데이터의 시공간 상관 관계를 분석하는 단계를 더 포함하는 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 방법

청구항 9

인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치는
 상기 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델을 구현할 컴퓨터 프로그램이 기록된 컴퓨터가 판독 가능한 기록매체를 포함하는 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치

발명의 설명

기술 분야

[0001] 본 발명은 공기질 모니터링 및 예측 기술에 관한 것으로서, 더욱 상세하게는 공기질 센서 고장 지역과 측정소 부재 지역의 공기질 예측을 위한 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치 및 방법에 관한 것이다.

배경 기술

[0003] 최근 20년 동안 급속한 산업화, 사회 변화 및 에너지 소비로 인해 상당한 대기 오염 문제가 발생했다. 대기질은 주로 미세먼지(PM 10), 초미세먼지(PM 2.5), 일산화탄소(CO), 이산화질소(NO2), 오존(O3)을 포함하는 1차 대기 오염 물질의 확산을 고려하여 측정한다. 1차 대기오염물질에 지속적으로 노출되면 호흡기 및 심혈관 건강에 장 단기적으로 해로운 영향을 미칠 수 있는데 예를 들어, NOx에 단기간 노출되면 호흡기 질환 및 폐 기능 장애가 발생할 수 있는 것으로 나타났다.

[0004] 따라서 도시의 지속가능하고 건강한 발전을 위해서는 대기질 모니터링, 예측 및 추적을 위한 신뢰할 수 있는

시스템 개발이 필요하다

[0005] 현재 대기질 예측 모델에서 방대한 공기질 데이터, 원격 감지 및 정교한 알고리즘의 도움으로 심층 신경망(DNN)은 특히 시계열 예측과 관련하여 기존 기계 학습 아키텍처의 한계를 극복해왔다. 이에 대기 오염 물질의 조기 예측을 제공하는 하이브리드 시공간 예측 모델의 연구가 제안되었으나 공기질 측정소 부재 지역이나 센서의 장기적 장애가 있는 경우 실시간 모니터링 및 예측이 매우 어려운 문제점이 있었다. 이러한 측면에서 센서 융합에 의한 대기질 패턴의 시공간적 구현을 기반으로 오염 물질 역학을 정확하게 식별하면 효과적인 대기 오염 관리 전략을 채택하는 데 도움이 될 수 있다.

[0006] 본 발명의 배경 기술은 대한민국 등록 특허 제2019-0178098호에 게시되어 있다.

발명의 내용

해결하려는 과제

[0008] 본 발명은 공기질 모니터링 및 예측 기술에 관한 것으로서, 장기간 센서 장애가 있거나 공기질 측정소가 없는 지역의 공기질을 예측할 수 있는 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 시스템을 제공한다.

[0009] 본 발명은 1차 대기 오염 물질을 여러 시간 간격에서 동시에 예측할 수 있는 건강 위험을 위한 조기 경고 시스템을 제공한다.

[0010] 본 발명은 기존 측정소에서 거리가 먼 지역을 원격 감지하여 공기질을 예측할 수 있는 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 시스템을 제공한다.

과제의 해결 수단

[0012] 본 발명의 일 측면에 따른, 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치를 제공한다.

[0013] 본 발명의 일 실시예에 따른 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치는 측정소의 공기질 데이터를 입력하는 공기질 데이터 입력부, 입력된 공기질 데이터를 전처리하는 공기질 데이터 전처리부, 전처리한 공기질 데이터를 입력하여 측정소가 없는 지점의 예측 공기질 데이터를 생성하는 예측 공기질 데이터 출력부를 포함할 수 있다.

[0014] 본 발명의 일 실시예에 따른 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치는 거리 적응형 그래프 컨볼루션 게이트 네트워크의 모델을 생성하는 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 생성부, 복수의 측정소의 공기질 데이터를 이용하여 복수의 측정소의 공기질 데이터의 시공간 상관 관계를 분석하는 시공간 분석부를 더 포함할 수 있다.

[0015] 본 발명의 다른 일 측면에 따르면, 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 방법을 제공한다.

[0016] 본 발명의 일 실시예에 따른 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 방법은 측정소의 공기질 데이터를 입력하는 단계, 입력된 공기질 데이터를 전처리하는 단계, 전처리한 공기질 데이터를 입력하여 측정소가 없는 지점의 예측 공기질 데이터를 생성하는 단계, 예측 공기질 데이터를 출력하는 단계를 포함할 수 있다.

[0017] 본 발명의 일 실시예에 따른 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 방법은 거리 적응형 그래프 컨볼루션 게이트 네트워크의 모델을 생성하는 단계, 복수의 측정소의 공기질 데이터를 이용하여 복수의 측정소의 공기질 데이터의 시공간 상관 관계를 분석하는 단계를 더 포함할 수 있다.

[0018] 본 발명의 일 실시예에 따른 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치는 거리 적응형 그래프 컨볼루션 게이트 네트워크의 모델을 구현할 컴퓨터 프로그램이 기록된 컴퓨터가 판독 가능한 기록매체를 포함할 수 있다.

발명의 효과

[0020] 본 발명의 일 실시예에 따르면, 공기질 예측을 위한 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치

및 방법은 장기간 센서 장애가 발생하거나 공기질 측정소가 없는 지역의 공기질을 예측할 수 있다.

[0021]

본 발명의 일 실시예에 따르면, 공기질 예측을 위한 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치 및 방법은 기존 측정소에서 거리가 먼 지역을 원격 감지하여 조기 경보 시스템을 발령함으로 건강 위험성 경고 및 공기질 규제 계획을 세우는 데 도움을 준다.

도면의 간단한 설명

[0023]

도 1은 본 발명의 일 실시예에 따른 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치를 도시한 도면.

도 2는 본 발명의 일 실시예에 따른 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 생성부의 구성을 도시한 도면.

도 3은 본 발명의 일 실시예에 따른 거리 적응형 그래프 컨볼루션(Distance adaptive graph convolutiion, 이하 DAGC)의 이웃 노드 임베딩을 그래픽으로 표현한 도면.

도 4는 본 발명의 일 실시예에 따른 측정소의 지리적 위치를 나타낸 도면.

도 5는 본 발명의 일 실시예에 따른 측정소 S1에서의 대기 오염 물질 CO, NO₂, O₃, PM₁₀, PM_{2.5}의 시계열 변화에 대한 도면.

도 6은 본 발명의 일 실시예에 따른 측정소 S1의 기본 통계자료에 대한 도면.

도 7은 본 발명의 일 실시예에 따른 PM₁₀과 PM_{2.5}의 측정소간의 공간 의존성을 표현한 도면.

도 8은 본 발명의 일 실시예에 따른 측정소 S1~ S9에서 월별 PM₁₀과 PM_{2.5}를 농도를 박스 플롯으로 비교한 도면.

도 9는 본 발명의 일 실시예에 따른 종합대기질 지수에 대한 도면.

도 10은 본 발명의 일 실시예에 따른 전체 연구 지역에서의 2019년 4월 한 달간 CO, NO₂, O₃, PM₁₀, PM_{2.5}의 AQIp 및 CAI의 공간적 분포를 표현한 도면.

도 11은 본 발명의 일 실시예에 따른 거리 적응형 그래프 컨볼루션 게이트 네트워크부에 관한 도면.

도 12는 본 발명의 일 실시예에 따른 ray tune 라이브러리를 사용하여 무작위 검색방법으로 하이퍼파라미터를 선택에 대한 도면.

도 13은 데이터 전처리를 위한 의사 코드 (pseudo code)와 거리 적응형 그래프 컨볼루션 게이트 네트워크부의 학습 절차를 나타낸 도면.

도 14는 본 발명의 일 실시예에 따른 측정값과 구간의 넓이가 다른 결측 데이터에 대한 (a) CO, NO₂, O₃, PM₁₀, PM_{2.5}의 시계열 분포 (b) 결측 구간이 있는 측정소(S7, S11, S17)와 센서가 없는 측정소(S3, S10, S15, S19, S20)에 대한 도면.

도 15는 본 발명의 일 실시예에 따른 측정소 S7, S11, S17에 결측치 데이터에 대해 나타낸 도면.

도 16은 본 발명의 일 실시예에 따른 거리 적응형 그래프 컨볼루션 순환 네트워크(Distance addaptive graph convolutional gated network, DAGCGN) 모델 생성부를 그래픽으로 표현한 도면.

도 17은 본 발명의 일 실시예에 따른 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 생성부의 의사 코드 (pseudo code)를 나타낸 도면.

도 18 내지 도 19는 본 발명의 일 실시예에 따른 학습 세트 및 테스트 세트에서의 서로 다른 시간 범위에서 예측 모델의 RMSE를 나타낸 도면.

도 20은 본 발명의 일 실시예에 따른 DAGCGN 모델의 테스트 세트에서의 서로 다른 시간 범위에서 예측 모델의 MASE를 나타낸 도면.

도 21은 본 발명의 일 실시예에 따른 DAGCGN 모델의 (a) S4 (b) S7 (c) S12 (d) S17 (e) S12 측정소의 예측된 PM_{2.5}의 농도에 관한 비교 도면.

도 22는 본 발명의 일 실시예에 따른 DAGCGN 모델의 T_{+1} , T_{+3} , T_{+6} , and T_{+12} 의 대상 지역 각 측정소에서 평균 1차 대기 오염물질 농도 분포 및 예측 MAE 값에 관한 도면.

도 23은 본 발명의 일 실시예에 따른 DAGCGN 모델의 대상 지역 전체에 걸친 다중 시간 범위에서의 CO, NO₂, O₃, PM₁₀, PM_{2.5}의 예측 MAE를 비교한 도면.

도 24는 본 발명의 일 실시예에 따른 DAGCGN 모델과 기준 모델의 복잡성과 학습시간을 비교한 도면.

도 25는 본 발명의 일 실시예에 따른 측정소 S7, S11, S17에서 부분 및 전체 센서 고장의 MAE 결측치 성능을 나타낸 도면.

도 26 내지 도 28은 본 발명의 일 실시예에 따른 S-7, S-11 및 S-17 측정소에서 부분 및 전체 센서 고장에 대한 DAGCGN 모델의 결측치 처리 성능에 관한 도면.

도 29는 본 발명의 일 실시예에 따른 대상 지역에서의 DAGCGN 모델의 공기질 비교에 관한 관측된 CAI와 예측된 CAI로, (a) 알려진 측정소 (b) 알려진 측정소 및 미지의 측정소에 관한 도면(기간:2019.12.24~2019.12.28).

도 30은 본 발명의 일 실시예에 따른 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 방법에 관하여 도시한 도면.

발명을 실시하기 위한 구체적인 내용

- [0024] 본 발명은 다양한 변형을 가할 수 있고 여러 가지 실시 예를 가질 수 있는 바, 특정 실시 예를 도면에 예시하고 이를 상세한 설명을 통해 상세히 설명하고자 한다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 본 발명을 설명함에 있어서, 관련된 공지 기술에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다.
- [0025] 이하에서는 첨부한 도면을 참조하여 본 발명을 설명하기로 한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며, 따라서 여기에서 설명하는 실시예로 한정되는 것은 아니다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.
- [0026] 본 발명은 공기질 측정소 부재지역과 장기적인 센서 고장이 있는 경우에도 공기질을 예측할 수 있는 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치 및 방법에 관한 것이다.
- [0027] 본 발명은 센서 융합을 통합한 거리 적응형 그래프 컨볼루션 순환 네트워크(DAGCGN) 시스템을 사용하여 공기질 조기 경보 시스템의 한계를 극복하는 것을 목표로 한다.
- [0028] 도 1은 본 발명의 일 실시예에 따른 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치를 도시한 도면이다.
- [0029] 도 1을 참조하면, 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치(10)는 공기질 데이터 입력부, 공기질 데이터 전처리부, 시공간 분석부, 거리 적응형 그래프 컨볼루션 게이트 네트워크부, 예측 공기질 데이터 출력부, 거리 적응형 그래프 컨볼루션 게이트 모델 생성부를 포함하여 구성될 수 있다.
- [0030] 공기질 데이터 입력부(100)는 대상 지역 측정소의 공기질 데이터를 입력한다.
- [0031] 공기질 데이터 전처리부(200)는 입력된 공기질 데이터를 전처리 할 수 있다. 공기질 데이터 전처리부(200)는 입력된 공기질 데이터에서 누락된 값을 채우고 명백한 이상치를 제거할 수 있다.
- [0032] 공기질 데이터 전처리부(200)는 누락된 오염 물질의 결측치를 shape-preserving 조각별 3차 스플라인 보간법을 사용하여 채운다.
- [0033] 공기질 데이터 전처리부(200)는 3차 에르미트 다항식 변형 보간법으로 사용하여 명백한 이상치를 제거한다.
- [0034] 시공간 분석부(300)는 측정된 공기질 데이터의 시계열 패턴을 분석하고, 공기질 측정소간의 시공간 관계에 대한 의존성을 분석한다.
- [0035] 시공간 분석부(300)는 공기질 데이터 사이의 시계열 패턴을 분석하여 공기질 데이터의 시간별, 계절별, 대기 오염 물질간의 패턴을 분석한다.

- [0036] 시공간 분석부(300)는 박스플롯을 이용하여 공기질 데이터의 시공간적 상관 관계를 분석한다.
- [0037] 시공간 분석부(300)는 공기질 데이터를 집합적으로 학습하여 정확한 공기질 예측 모델의 선정을 위한 지상 측정 소간의 공간적 관계를 분석한다.
- [0038] 거리 적응형 그래프 컨볼루션 게이트 네트워크부(400)는 전처리한 공기질 데이터를 입력하여 측정소가 없는 지점의 예측 공기질 데이터를 생성한다.
- [0039] 거리 적응형 그래프 컨볼루션 게이트 네트워크부(400)는 공기질 데이터와 대상 위치까지의 거리의 입력으로 원격 위치의 대기질 분포를 학습시켜 오염 물질을 예측한다.
- [0040] 거리 적응형 그래프 컨볼루션 게이트 네트워크부(400)는 그래프 컨볼루션 연산 및 노드 샘플링의 시너지 효과를 활용하여 오염 물질을 예측한다.
- [0041] 예측 공기질 데이터 출력부(500)는 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델의 예측 공기질 데이터를 출력한다.
- [0042] 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 생성부(600)는 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델을 개발하고 모델의 성능을 검증한다.
- [0043] 도 2는 본 발명의 일 실시예에 따른 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 생성부의 구성을 도시한 도면이다.
- [0044] 도 2를 참조하면, 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 생성부(600)는 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 개발부(620)와 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 검증부(640)를 포함하여 구성될 수 있다.
- [0045] 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 개발부(620)는 측정소의 공기질 데이터를 학습하여 거리 적응형 그래프 컨볼루션 게이트 네트워크를 개발한다.
- [0046] 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 개발부(620)는 경기도 전역에 위치한 23개 공기질 측정소에서 중요한 대기질 지표에 대한 데이터를 수집하여 결측치를 확인하고 이상치를 제거한다.
- [0047] 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 개발부(620)는 결측치를 확인하기 위해 조각별 3차 스플라인 보간법을 사용하여 데이터 세트를 검사하고 결과를 편향시킬 수 있는 다양한 이상치를 제거하기 위해 3차 에르미트 다항 변환을 사용한다.
- [0048] 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 개발부(620)는 대기 오염 물질이 대상지역에서 공간적 관계가 있는 지를 알아보기 위해 상관 분석을 하고, 건강 위험 분포를 가시화하기 위해 CAI 지도를 작성한다. 이후 데이터를 정규화하고 80%: 20% 비율로 학습 및 테스트 세트로 분할한다.
- [0049] 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 개발부(620)는 5개의 대기질 측정소(S3, S10, S15, S19, S20)를 원래 데이터 세트에서 분리하여 미지의 위치에서의 원격 감지에 대한 모델 성능을 평가한다. 이 때, 학습을 위해 '알려진 위치'로 레이블이 지정된 18개의 측정소와 DAGCGN 모델 시연을 위해 '미지의 위치'로 레이블이 지정된 5개의 측정소가 생성한다.
- [0050] 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 개발부(620)는 다단계 공기질 예측을 제공하기 위해 슬라이딩 윈도우를 활용하여 데이터를 입력 및 출력 윈도우로 변환하여 DAGCGN 모델을 설계한다.
- [0051] 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 개발부(620)는 설계한 DAGCGN 모델을 슬라이딩 윈도우 아키텍처에 통합하고 알려진 위치에서 먼저 학습하여 5개의 대기 오염 물질을 예측한다.
- [0052] 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 검증부(640)는 측정소의 공기질 데이터를 이용하여 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델의 성능을 평가하고 검증한다.
- [0053] 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 검증부(640)는 T_{+1} , T_{+3} , T_{+6} , and T_{+12} 와 같은 시간 범위에 걸쳐 오염 물질 수준의 조기 감지를 하여 센서 데이터가 있는 알려진 측정소의 기준 모델과 성능을 비교한다.
- [0054] 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 검증부(640)는 다양한 시간 길이 동안 센서 고장과 관련된 공기질 모니터링의 성능을 검증한다.

- [0055] 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 검증부(640)는 공기질 모니터링을 위해 센서가 없는 미지의 위치에 대한 원격 감지를 하여 모델의 성능을 검증한다.
- [0056] 도 3은 본 발명의 일 실시예에 따른 거리 적응형 그래프 컨볼루션(Distance adaptive graph convolution, DAGC)의 이웃 노드 임베딩을 그래픽으로 표현한 도면이다.
- [0057] 도 3을 참조하면, 종래의 그래프 컨볼루션 네트워크를 고차원으로 확장하고 그래프 컨볼루션 네트워크의 학습 능력을 향상시키기 위해 자동으로 데이터에서 숨겨진 상호 의존성을 제시하기 위한 데이터 적응형 그래프 생성(DAGG)을 제안하였다.
- [0058] 대기 오염 농도를 광범위하게 예측하려면 지상 측정소간의 공간적 관계를 분석하고 집합적으로 학습해야 한다. 측정소 및 관련 데이터의 분포는 시퀀스 또는 그리드에서와 같이 본질적으로 규칙적이지 않고 비유클리드 데이터와 유사한 불규칙한 그래프와 같은 구조 토폴로지 구조를 갖는다. 수많은 연구에서 성공적으로 그래프 컨볼루션 네트워크를 사용하여 기상 측정소 정보, 교통 흐름 정보 및 PM 2.5 농도 관계에 관한 비유클리드 토폴로지 정보를 활용하여 매력적인 예측 결과를 얻었다. 본 발명에서도 유사한 불규칙 구조를 가진 데이터를 검토하므로, 그래프 컨볼루션 네트워크 프레임워크를 사용한다. 공간 위치 파악 기능과 수학적 단순성으로 인해 스펙트럼 필터로 표현되는 그래프 컨볼루션 네트워크는 상당한 관심을 끌었는데, 그래프 푸리에 변환에 기반한 그래프 G 의 정규화된 라플라시안 변환은 수학적 1로 나타낸다. 필터 $f_{\theta} = \text{diag}(\theta)$ 즉, $\theta \in \mathbb{R}^n$ 를 사용하여 시그널 'x'의 스펙트럼 그래프 변환은 수학적 2로 나타낸다.
- [0059]
$$L = I_n - D^{(-1/2)} A D^{(-1/2)} = U \Lambda U^T \quad (1)$$
- [0060]
$$x \cdot f_{\theta} = U f_{\theta} U^T x \quad (2)$$
- [0061] 여기서, I_n 과 A및 D는 각각 항등 행렬, 인접 행렬 및 차수 행렬을 나타내고, $U \in \mathbb{R}^n$ 와 $\Lambda \in \mathbb{R}^n$ 는 각각 L 고유값의 고유 벡터와 대각 행렬을 포함하는 행렬로 구성되고, 또한 $x \cdot f_{\theta}$ 는 컨볼루션 연산을 나타내며 $U^T x$ 는 x의 푸리에 변환을 나타낸다. 수학적 2는 스펙트럼 그래프 컨볼루션을 식별하는 데 사용할 수 있으나, 고유값 분해의 높은 계산으로 인해 적용 가능성이 낮아지고 서로 다른 근사값이 도출된다. 그래프 컨볼루션 프로세스는 1차 체비쇼프 다항식 확장을 사용하여 근사화될 수 있으며 다음과 같이 고차원 그래프 컨볼루션 네트워크로 확장될 수 있다.
- [0062]
$$Z = (I_N + D^{-1/2} A D^{-1/2}) X \Theta + b \quad (3)$$
- [0063] $X \in \mathbb{R}^n$ 과 $Z \in \mathbb{R}^n$ 은 각각 GCN레이어의 입력과 출력을 $\theta \in \mathbb{R}^n$ 과 $b \in \mathbb{R}^n$ 은 모델에 대해 학습 가중치와 편향 매개변수를 나타낸다. 그래프 컨볼루션 네트워크의 학습 능력을 향상시키기 위해 자동으로 데이터에서 숨겨진 상호 의존성을 제시하기 위한 데이터 적응형 그래프 생성(DAGG)이 제안되었다.
- [0064] 본 발명에서는 학습 노드 사전 $C_A \in \mathbb{R}^n$ 는 모든 노드에 대해 초기화되고 C_A 는 노드 임베딩을 나타낸다. 각 노드 쌍 사이의 공간 의존성은 C_A^T 와 C_A 의 곱에 의해 추론되고, 소프트맥스 함수는 적응 행렬을 정규화하는 데 사용된다. 네트워크는 서로 다른 측정소간의 공간 관계를 자세히 설명하기 위해 C_A 를 자동으로 업데이트하고, 데이터에 따라 적응형 그래프 토폴로지를 구축한다. 데이터 적응형 그래프 생성(DAGG) 모듈은 수학적 3을 기반으로 수학적 4로 나타낸다. 그래프 컨볼루션 네트워크를 개량한 데이터 적응형 그래프 생성(DAGG) 모듈은
- [0065] 다음과 같이 공식화할 수 있다.
- [0066]
$$D^{-1/2} A D^{-1/2} = \text{softmax}(\text{ReLU}(C_A C_A^T)) \quad (4)$$
- [0067]
$$Z = (I_N + \text{softmax}(\text{ReLU}(C_A C_A^T))) X \Theta + b \quad (5)$$
- [0068] 그러나, 적응형 그래프 컨볼루션 네트워크는 공간적으로 분산된 노드 네트워크에서 자동화된 그래프 구성 경로에서의 숨겨진 관계를 효율적으로 발견하지만 몇 가지 단점이 있다. 노드 수가 증가함에 따라 그래프 형성을 위한 컨볼루션 비용이 엄청나게 비싸고, 큰 그래프에서는 각 대상 노드에 대한 모든 이웃 정보를 활용하는 것이

이상적이지 않다. 결과적으로 네트워크 학습을 효과적으로 최적화하기 위해서는 그래프 분할 및 노드 샘플링 기술을 탐색해야 한다. 따라서, 여러 하위 그래프를 샘플링하고 각 하위 그래프 내에서 이웃 노드를 제한하는 방법이 제안되었는데, 그럼에도 불구하고 측정소의 불균일한 분포로 인해 이전 연구는 선택한 그리드 크기에 따라 불안정한 노드 수로 인해 공간 정보를 이해할 수 없거나 이웃 폭발이 발생하는 문제점이 있었다.

- [0069] 이에 비해 대상 노드에서 랜덤 워크를 시작하고 가장 높은 정규화된 방문 수를 기반으로 상위 N 개의 노드를 선택하는 것을 통합하는 중요도 기반 노드 샘플링 기술이 제안되었다. 공기질 측정소의 토폴로지(망구성방식), 거리 행렬 S 및 설정된 수의 인근 노드 K 가 주어지면 중요도 기반 노드 샘플링 기술은 추론과 적응형 그래프 생성을 위해 가장 높은 상관 관계가 있는 노드를 자동으로 식별할 수 있는 거리 적응형 그래프 컨볼루션(Distance adaptive graph convolution) 모듈을 생성할 수 있다.
- [0070] 또한, 이 모듈을 사용하면 신경망 모델이 유연한 메모리 풋프린트를 갖는 동시에 중요한 상위 K 노드의 공간 종속성으로부터 이점을 얻을 수 있다.
- [0071] 본 발명의 일 실시예에 따른 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치 및 방법은 장기적인 센서 고장이 있는 경우와 공기질 측정소 부재 지역의 공기질을 예측할 수 있는 거리 적응형 그래프 컨볼루션 게이트 네트워크 장치 및 방법에 관한 것으로, 우리나라 인구의 25%를 차지하는 경기도를 본 발명의 대상 지역으로 선정하였다. 최근 몇 년 동안 이 지역은 산업 생산량 및 에너지 소비의 증가와 함께 급속한 도시화를 경험하고 빈번하고 심각한 대기 오염 시나리오로 이어졌다.
- [0072] 도 4는 본 발명의 일 실시예에 따른 측정소의 지리적 위치를 나타낸 도면이다.
- [0073] 도 4를 참조하면, 경기도 남부의 지리적 지도로 다른 도시에 걸쳐 있는 23개의 선택된 측정소의 위치를 보여준다. 본 발명은 주로 국가의 수도에 인접한 남쪽 지방에 초점을 맞추고 있다.
- [0074] 환경부 산하 한국환경공단은 전국 대기질 모니터링을 위해 446개 측정소에서 기록된 오염물질에 대한 데이터베이스를 유지하고 있다. 이것들은 공개적으로 접근 가능하고 품질이 보장된 데이터 세트의 이전 연구에서도 활용되었다. 본 발명은 2019년 1월 1일부터 12월 31일까지 23개 측정소에서 PM2.5 PM10 NO2 CO 및 O3의 주요 대기 오염 물질에 대한 데이터를 수집했다.
- [0075] 도 5는 본 발명의 일 실시예에 따른 측정소 S1에서의 대기 오염 물질 CO, NO2, O3, PM10, PM2.5의 시계열 변화에 대한 도면이다.
- [0076] 도 6은 본 발명의 일 실시예에 따른 측정소 S1의 기본 통계자료에 대한 도면이다.
- [0077] 도 7은 본 발명의 일 실시예에 따른 PM10과 PM2.5의 측정소간의 공간 의존성을 표현한 도면이다.
- [0078] 도 8은 본 발명의 일 실시예에 따른 측정소 S1~ S9에서 월별 PM10과 PM2.5 농도를 박스 플롯으로 비교한 도면이다.
- [0079] 도 5 내지 도 8을 참조하면, 공기질 데이터 전처리부(200)를 통해 대상 지역에서 측정된 공기질 데이터를 전처리한다. 누락된 오염 물질의 결측치는 shape-preserving 조각별 3차 스플라인 보간법을 사용하여 검사하였다. 데이터에서 이상치는 거의 없었으나 3차 에르미트 다항식 변형 보간법으로 사용하여 명백한 이상치를 제거했으며, 모든 작업은 MathWorks에서 제공하는 MATLAB을 사용하여 수행되었다.
- [0080] 도 6을 참조하면, 대부분의 대기 오염 물질이 매우 왜곡되어 비선형 분포를 나타내었으며, NO2는 다소 작은 왜도를 나타내었다.
- [0081] 공기질 관련 작업에서의 그래프 컨볼루션 네트워크(GCN)는 숨겨진 시공간 종속성을 캡처하여 시간적 학습을 촉진하기 위해 주로 적용된다. 그러나 데이터 사이에 공간적 상관관계가 존재하는지 먼저 평가할 필요가 있다. 이를 해결하기 위해 피어슨 상관 계수(R)를 사용하여 측정소간의 공간적 상관 관계를 평가했다.
- [0082] 도 7을 참조하면, PM2.5에 대한 R 값은 0.7316에서 0.9610 사이의 관계로 대상 지역의 공기질 분포 사이에 강력한 공간적 상관 관계가 존재한다. 따라서 공간 종속성을 효율적으로 추출하는 그래프 신경망은 시간 학습을 완화하는데 도움이 될 수 있다. 추가 분석에 따르면 R 값은 측정소 사이의 거리와 반비례 관계가 있으며, 높은 상관관계를 가지는 더 작은 하위 영역들을 생성한다
- [0083] 이 분석에서는 멀리 있는 노드의 노이즈를 줄이면서 인접 노드의 예측 출력을 신중하게 분석해야 한다.
- [0084] 도 8을 참조하면, 박스플롯을 이용하여 PM10과 PM2.5의 시공간적 상관 관계도 검증할 수 있다. 박스플롯의 점들

은 최대값과 최소값 사이의 샘플 분포를 나타낸다. S10~S23 측정소(미도시)를 포함하여 전체 측정소에서는 약간 다른 오염물질 농도의 모양에도 불구하고 대체적으로 비슷한 추세가 일년 내내 관찰되었다.

[0085] 공간적 의존성을 파악하기 위해서는 대기질 모니터링과 오염물질 분석이 필수적이다. 하지만 모니터링만으로는 조기 경보 및 오염 통제를 위한 시스템을 구축하여 대중 또는 정책 입안자에게 건강 위험의 실제 상태를 전달하기 어렵다. 이러한 문제를 극복하기 위하여 많은 국가에서는 쉽게 계산되고 건강 위험 문제를 효과적으로 전달하는 대기질 기준을 구현하였다.

[0086] 도 9는 본 발명의 일 실시예에 따른 종합대기질 지수에 대한 도면이다.

[0087] 도 9를 참조하면, 한국환경공단은 여러 환경오염 물질의 농도를 고려하고 공기질 지수(AQI 지수)가 높은 오염 물질을 기준으로 건강오염도를 특성화한 종합대기질지수(CAI)를 도입했다.

$$[0088] \quad AQI_p = \frac{I_{HI} - I_{LO}}{BP_{HI} - BP_{LO}} (C_p - BP_{LO}) + I_{LO} \quad (6)$$

$$[0089] \quad CAI = \max [AQI_{(p_1)}, AQI_{(p_2)}, \dots, AQI_{(p_n)}] \quad (7)$$

[0090] 각 오염 물질에 대한 CAI는 수학적 식 6과 7을 사용하여 계산되었다.

[0091] AQIp는 각 오염 물질에 대한 대기질 지수를 나타내며, Cp는 현재 대기 오염 물질 농도이다. BP_{HI} 및 BP_{LO}는 각 농도의 브레이크포인트 $\geq C_p$ 및 $\leq C_p$ 이다.

[0092] I_{HI}, I_{LO}, BP_{HI}, BP_{LO}는 각각에 해당하는 인덱스 값이다.

[0093] 최근까지 대기질 평가는 개별 오염물질 특히 미세먼지(PM)의 경우 분석과 예측에 주로 초점을 맞췄다. 그러나 CAI에서 보는 것처럼 개별 오염 물질은 전체 대기질의 진정한 상태를 나타낼 수 없다.

[0094] 도 10은 본 발명의 일 실시예에 따른 전체 연구 지역의 2019년 4월 한 달간 CO, NO2, O3, PM10, PM2.5의 AQIp 및 CAI의 공간적 분포를 표현한 도면이다.

[0095] 도 10을 참조하면, AQIp와 CAI를 기준으로 대상 지역 전체의 건강위험도를 비교한 것으로, 일별 편향을 제거하기 위하여 AQI와 CAI를 한달 동안 평균한 것이다.

[0096] 도 9 및 도 10을 참조하면, AQI는 농도의 브레이크포인트가 특정 오염물질을 초과할때만 더 높은 위험 수준을 제공하므로 다른 오염 물질의 영향은 무시된다. CAI는 대기질 상태를 나타내기 위하여 각 단계에서 가장 심각한 오염 물질만을 독점적으로 고려하기 때문에 CAI의 공간분포를 통하여 정확한 조기 경보 시스템을 개발하는데 있어 모니터링 되는 모든 오염 물질에 대한 미래 예측이 중요함을 더욱 알 수 있다.

[0097] 그래프 컨볼루션 네트워크(GCN)은 CNN의 변형으로 그래프 구조 데이터에 최적화되어 노드 분류, 링크 예측 및 그래프 분류에 널리 사용된다. 최근 연구는 미리 정의된 그래프를 사용하여 공유 패턴을 캡처하기 위한 복잡한 그래프 신경망 아키텍처 개발에 중점을 두었다. 또한 학습 노드별 패턴은 시공간적 실외 공기질 예측에 중요하다.

[0098] 도 11은 본 발명의 일 실시예에 따른 거리 적응형 그래프 컨볼루션 게이트 네트워크부에 관한 도면이다.

[0099] 도 11을 참조하면, 오염물질 농도에 의한 건강 위험 수준의 조기 경보를 제공하기 위해 실외 대기질을 예측하는 다단계 및 다중 사이트 문제에 중점을 둔다.

[0100] 여기서, 데이터 세트는 $D = \{(S_i, Y_i) | i = 1, \dots, N\}$, 타임스텝 t에서 $S_i^t = \{s_{i,1}^t, \dots, s_{i,m}^t\} \in R^{m \times p}$, $Y_i^t = \{y_{i,1}^t, \dots, y_{i,m}^t\} \in R^{m \times p}$ 로서 m 개의 대기 오염 소스를 갖는 N개의 데이터 샘플이다. p는 대기질 행렬의 수이고, 각 데이터 샘플은 m X p-차원 다변량 시계열 입력 S_i^t 및 출력 Y_i^t 로 구성된다. 예를 들면 평균 제곱 오차 또는 절대값 평균 오차 같은 손실함수 $L(Y_i^t, \hat{Y}_i^t)$ 를 사용하고 과거 데이터 S_i^t 를 기반으로 미래를 예측할 수 있는 $\hat{Y}_i^t$ 를 찾는 $F_\theta(S_i^t) \in f: R^N \rightarrow \hat{Y}_i^t$ 를 찾는 것으로 공식화 되었다. 여기서, θ 는 모델에서 모든 학습가능한 매개 변수를 나타낸다

- [0101] 서로 다른 실외 측정소사이의 정확한 공간적 상관관계를 얻기 위해서 그래프는 $G = (V, E, A)$ 로 공식화된다.
- [0102] $|V| = m$ 으로 표현되는 공기질 센서의 수, E 는 엣지세트, $A \in \mathbb{R}^{m \times m}$ 는 공기질 센서의 근접성을 나타내는 그래프의 인접 행렬을 나타낸다.
- [0103] 다음과 같이 표현되는데, $\mathcal{F}_\theta(S_i^t; G) \in f: \mathbb{R}^N \rightarrow \hat{Y}_i^t$ 는 여기서 두 노드사이의 엣지는 가까운 이웃에 있는 경우에만 정의된다. 또한, 선택한 측정소의 대기질 정보와 함께 멀리있는 설치되지 않은 미지의 지점까지의 거리를 입력으로 추가하여, 모델이 원격 위치의 대기질 분포를 학습하고 기본 그래프를 발전시키도록 한다.
- [0104] 실외 공기질 데이터 세트의 시퀀스 원시 센서 데이터는 자연 상태에서 신경망 기반 방법을 학습하는데 적합하지 않다. 따라서, 과거 데이터(lag) L 의 크기와 다단계 미래 데이터의 크기(horizon) h 를 규정하는 전처리를 먼저 진행했다. 이 정보를 이용하여 데이터 세트(시계열 입력 S_i^t 및 출력 Y_i^t)가 슬라이딩 윈도우 기법을 사용하여 변환되었다. 각 단계에서는 일련의 기능 및 대상 변수를 반복하여 샘플을 만든다.
- [0105] t_0 에의 입력은 t_0 과 t_0+L 사이의 값을 사용하여 생성되며, y_0 은 대상 변수를 나타낸다. 여기서 L 는 과거 입력 샘플수와 대상이 되는 미래 시간 범위로 창 크기를 나타낸다. 예를 들어, 매시간 학습은 과거는 5로 미래 시간 범위는 2로 설정된 경우이다.
- [0106] 모든 시간 단계에서 데이터는 윈도우로 변환되는데, 입력으로 다른 측정소의 과거 5개 시간 단계의 5개 1차 대기 오염물질 농도를 포함하는 윈도우로 변환된다.
- [0107] 이러한 입력이 주어지면 모델은 5개의 1차 오염 물질에 대한 미래 2시간 후를 출력값으로 제공하려고 시도한다.
- [0108] 효과적인 실외 공기질 예측을 위해서는 노드별 시공간 상관 관계를 모두 캡처하는 것이 중요하다. 이전 연구에서 노드 적응 매개변수 학습, 데이터 적응 그래프 생성 및 순환 유닛을 결합하여 노드별 시공간 상관 관계를 캡처하는 일변량 시계열에 대한 적응형 그래프 컨볼루션 순환 네트워크에 대한 제안이 있었다. 본 발명은 이 개념을 다변량 시계열로 확장하고 여러 대기질 오염 물질의 동시 다단계 예측을 위해 거리 적응 그래프 컨볼루션 게이트 순환 네트워크 기반 인코더 디코더 프레임워크를 개발하였다.
- [0109]
$$\bar{A} = \text{softmax}(\text{ReLU}(CC^T)) \quad (8)$$
- [0110]
$$z_t = \sigma[\bar{A}[S_t^f, H_{t-1}]CW_z + Cb_z] \quad (9)$$
- [0111]
$$r_t = \sigma[\bar{A}[S_t^f, H_{t-1}]CW_r + Cb_r] \quad (10)$$
- [0112]
$$\hat{H}_t = \tanh(\bar{A}[S_t^f, r \odot H_{t-1}]CW_{\hat{H}} + Cb_{\hat{H}}) \quad (11)$$
- [0113]
$$H_t = z \odot H_{t-1} + (1-z) \odot \hat{H}_t \quad (12)$$
- [0114] S_t^f 와 H_t 는 타임스텝 t 에서의 입력과 출력값이고, $[\cdot]$ 는 연결작업, z 는 리셋게이트, r 은 업데이트 게이트를 나타낸다. C , W_z , W_r , $W_{\hat{H}}$, b_z , b_r , $b_{\hat{H}}$ 는 학습 매개변수이다. 표기를 단순화하기 위해 $GCM(S_t^f, A)$ 은 입력값 S_t^f 와 인접행렬 A 가 있는 GCN 모델의 출력값을 나타낸다.
- [0115]
$$\text{Encoder: } Z = GCM(S_t^f, A) \quad (13)$$
- [0116]
$$\text{Decoder: } Y_t^f = \text{BatchNorm}(\text{Conv2D}(Z)) \quad (14)$$
- [0117] 본 발명의, 일실시예에 따른 거리 적응형 그래프 컨볼루션 게이트 네트워크부(400)는 STGCN(spatio-temporal graph convolutional network), ASTGCN(Attention-based spatio-temporal graph convolutional network), STGCN(spatio-temporal synchronous graph convolutional network) 및 AGCRN(adaptive graph graph convolutional recurrent network)와 같은 여러 최신 GCN 프레임워크에서 영감을 받아 개발되었다. 이전 프레임워크는 GCN 기반 레이어를 쌓으면 노드별 시공간 역학을 캡처하는 모델의 기능을 향상시킬 수 있음을 시사한

다.

- [0118] 거리 적응형 그래프 컨볼루션 게이트 네트워크부(400)의 구조는 정규화를 돕고 학습 프로세스 속도를 높이는 네트워크의 입력 데이터를 표준화하기 위한 배치 정규화 계층도 포함되어 있다. 또한 오버피팅을 최소화하고 추가 정규화를 제공하기 위해 드롭 아웃 레이어가 도입되었다.
- [0119] 모델 출력의 경우, 하나의 컨볼루션 레이어와 배치 정규화 레이어를 사용하여 아키텍처에서 디코더를 매우 단순하게 유지했다. 따라서 간접적으로 인코더가 여러 오염 물질의 미래 시간 단계를 구성할 수 있도록 유익하고 충분히 단순한 잠재 공간 임베딩을 생성하도록 강제하였다. 이를 통해 본 발명의 일 실시예에 따른 거리 적응형 그래프 컨볼루션 게이트 네트워크부(400)는 입력 데이터의 잠재 공간을 보다 정확하게 표현할 수 있었고, 실외 대기 오염 물질을 더 정확하게 예측할 수 있었다. 거리 적응형 그래프 컨볼루션 게이트 네트워크부(400)는 모델에 가장 적합한 하이퍼파라미터를 선택하기 위해 무작위 검색 방법으로 pytorch의 ray tune 라이브러리를 사용하였다.
- [0120] 도 12는 본 발명의 일 실시예에 따른 ray tune 라이브러리를 사용하여 무작위 검색 방법으로 하이퍼파라미터의 선택에 대한 도면이다.
- [0121] 도 12를 참조하면, 최적화를 위한 하이퍼파라미터는 1) 히든레이어의 수 2) AGCGN 셀의 GRU 유닛 3) 드롭아웃율 4) 노드 임베딩 차원 5) 활성화 함수를 포함한다.
- [0122] 이 모델은 Adam을 옵티마이저로 사용하고 평균 제곱 오차(MSE)를 손실 함수로 사용하여 100epoch 동안 학습되었다.
- [0123] 그 결과, 손실이 가장 적은 선택된 하이퍼파라미터는 다음과 같다.
- [0124] 1) 히든 레이어의 수: 2
- [0125] 2) AGCGN 셀의 GRU 유닛: 64, 64
- [0126] 3) 드롭아웃율: 0.25
- [0127] 4) 노드 임베딩 차원: 10
- [0128] 5) 활성화 함수: 시그모이드, Tanh, ReLU, Softmax
- [0129] 6) 학습률: 0.003
- [0130] 도 13은 데이터 전처리를 위한 의사 코드 (pseudo code)와 거리 적응형 그래프 컨볼루션 게이트 네트워크부(400)의 학습 절차를 나타낸 도면이다.
- [0131] 센서는 데이터 및 신호 수집, 특히 환경 모니터링에 광범위하게 사용되는데, 공기질 모니터링에서는 실제 공기질 모니터링 장비의 데이터 누락 또는 센서 사용이 불가능한 경우, 기계 고장, 빈번한 유지 보수, 불안정한 전송 또는 고가의 모니터링 장비 등의 통제 불가에 의한 변수가 많다. 이러한 문제들은 지속적인 공기질 데이터를 통한 건강 위험을 모니터링하고 조기 경고 발령에 치명적인 결과를 초래할 수 있다.
- [0132] 본 발명의 일 실시예에 따른 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델부 생성부(600)는 DAGCGN 모델을 개발하고 성능을 검증한다.
- [0133] 도 14는 본 발명의 일 실시예에 따른 측정값과 구간의 넓이가 다른 결측 데이터에 대한 (a) CO, NO2, O3, PM10, PM2.5의 시계열 분포 (b) 결측 구간이 있는 측정소(S7, S11, S17)와 센서가 없는 측정소(S3, S10, S15, S19, S20)에 대한 도면이다.
- [0134] 도 15는 본 발명의 일 실시예에 따른 측정소 S7, S11, S17에 결측치 데이터에 대해 나타낸 도면이다.
- [0135] 도 16은 본 발명의 일 실시예에 따른 거리 적응형 그래프 컨볼루션 순환 네트워크(Distance addaptive graph convolutional gated network, DAGCGN) 모델 생성부를 그래픽으로 표현한 도면이다.
- [0136] 도 14 내지 도 16을 참조하면, 본 발명의 일 실시예에 따른 거리 적응형 그래프 컨볼루션 게이트 네트워크(DAGCGN)의 성능을 평가하기 위해 두 가지 실제 시나리오를 실시했다.
- [0137] 첫째, 측정소에서 장기간 센서 고장인 경우의 공기질 예측치를 출력한다.
- [0138] 둘째, 공기질 원격 감지를 사용하여 학습 단계에서 활용되지 않은 미지 지점에서의 공기질 예측을 생성한다.

[0139] 첫째 검증 시나리오는 S7, S11, S17에서 서로 다른 결측 기간(1주 또는 2주)을 갖는 CO, NO2, O3, PM10, PM2.5의 각 센서 측정에 대해 누락된 데이터 구간을 의도적으로 도입하였다.

[0140] 둘째 검증 시나리오는 원격 감지 및 미지 지점에서 전체 공기질 상태, 즉 5개의 대기 오염 물질 모두를 예측하는 것에 대한 것이다.

[0141] 예측을 위해 5개의 실제 측정소에 센서 데이터가 없다고 가정하고, 이를 미지의 위치로 표시하였다. 이 측정소는 원래 데이터 세트에서 분리되었으며, 모델 학습에서 활용되지 않았다. 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 생성부(600)는 개발된 동적 그래프를 기반으로 CO, NO2, O3, PM10, PM2.5를 동시에 예측했다.

[0142] 도 17은 본 발명의 일 실시예에 따른 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 개발부(620)의 의사 코드(pseudo code)를 나타낸 도면이다.

[0143] 이전에 그래프 컨볼루션 순환 네트워크는 시계열 예측 응용 프로그램에서 자주 사용되는 세 가지 고유한 기준을 사용하여 미래 대기질 상태를 캡처하는 성능에 대해 철저히 테스트되는 연구가 있었다. 여기에는 평균 제곱근 오차(RMSE), 평균 절대 오차(MAE), 평균 절대 척도 오차(MASE)가 포함되는데, 수학적식은 다음과 같다.

$$[0144] \text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - y_i^*)^2}{n}} \quad (15)$$

$$[0145] \text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - y_i^*| \quad (16)$$

$$[0146] \text{MASE} = \frac{1}{n} \sum_{i=1}^n \left(\frac{|y_i - y_i^*|}{\frac{1}{n-1} \sum_{i=2}^n |y_i - y_{i-1}^*|} \right) \quad (17)$$

[0147] 여기서, y 는 시간 단계 i 에서의 관찰값을, y^* 는 추정값을 나타낸다

[0148] 본 발명의 일 실시예에 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델의 개발은 다음의 과정을 통하여 진행된다.

[0149] 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 생성부(600)는 경기도 전역에 위치한 23개 대기질 측정소에서 중요한 대기질 지표에 대한 데이터를 수집하여 결측치를 확인하고 이상치를 제거하였다. 결측치를 확인하기 위해 조각별 3차 스플라인 보간법을 사용하여 데이터 세트를 검사하고 결과를 편향시킬 수 있는 다양한 이상치를 제거하기 위해 3차 에르미트 다항 변환을 사용했다.

[0150] 그리고, 대기 오염 물질이 대상지역에서 공간적 관계가 있는 지를 알아보기 위해 상관 분석을 하였고, 건강 위험 분포를 가시화하기 위해 CAI 지도를 작성하였다.

[0151] 이후 데이터를 정규화하고 80%: 20% 비율로 학습 및 테스트 세트로 분할하였다.

[0152] 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 개발부(620)는 5개의 대기질 측정소(S3, S10, S15, S19, S20)를 원래 데이터 세트에서 분리하여 미지의 위치에서의 원격 감지에 대한 모델 성능을 평가했다. 학습을 위해 '알려진 위치'로 레이블이 지정된 18개의 측정소와 DAGCGN 모델 시연을 위해 '미지의 위치'로 레이블이 지정된 5개의 측정소가 생성되었다.

[0153] 그리고, 단단계 공기질 예측을 제공하기 위해 슬라이딩 윈도우를 활용하여 데이터를 입력 및 출력 윈도우로 변환하였다. 그런 다음 제안된 DAGCGN 네트워크가 구성되어 슬라이딩 윈도우 아키텍처에 통합되었다. DAGCGN 모델은 시공간 역학을 자세히 설명하고 기본 그래프를 작성하기 위해 알려진 위치에서 먼저 학습되었다. 비교를 위해 스택형 공간 모델 및 인코더-디코더 프레임워크를 포함하는 시공간 예측 모델의 다양한 기준 모델을 선정하였다. 그래프 기반 프레임워크가 없기 때문에 스택형 시공간 기준 모델들을 위한 고유한 위치 기반 모델이 개발되었다.

[0154] 본 발명의 일 실시예에 따른 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 생성부는 각 측정소에서 동시에 5개의 대기 오염 물질을 모두 예측했다.

[0155] 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 검증부(640)는 개발된 모델의 성능을 평가하고 검증하는 과정을 다음과 같이 수행하였다.

- [0156] 첫번째 단계, 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 검증부(640)는 T_{+1} , T_{+3} , T_{+6} , and T_{+12} 와 같은 시간 범위에 걸쳐 오염 물질 수준의 조기 감지를 평가하기 위해 센서 데이터가 있는 알려진 측정소의 기준 모델과 성능을 비교했다.
- [0157] 두번째 단계, 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 검증부(640)는 다양한 시간 길이 동안 센서 고장과 관련된 공기질 모니터링의 중요한 문제를 해결하는데 활용되었다.
- [0158] 마지막으로, 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 검증부(640)는 공기질 모니터링을 위해 센서가 없는 미지의 위치에 대한 원격 감지에 대해 평가되었다. 또한 건강 위험 분포의 조기 식별을 위한 제한된 모니터링 조건에서의 모델의 효율성을 입증하기 위해 CAI 맵을 생성했다.
- [0159] 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 검증부(640)는 첫 번째 단계에서는, 대상 영역에서 대기 오염 물질의 시공간적 역학을 포착하기 위해 5가지 기계 학습 프레임워크를 기준 모델로 사용했다.
- [0160] 본 발명의 일 실시예에 따른 예측 공기질 데이터 출력부(500)는 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 생성부(600)에서 과거 12시간을 입력값으로 미래 12시간의 공기질을 예측한다.
- [0161] 모델 입력의 경우, 시계열이 과거 시퀀스와 미래 시간 단계로 변환되었는데,
- [0162] 과거 시간은 T_{+1} , T_{+3} , T_{+6} , and T_{+12} 를 포함하여 여러 시간 간격이 있는 12시간으로 설정되어 공기질 예측값을 제공한다.
- [0163] 알려진 측정소에서 평가된 각 신경망 모델에서 생성된 오차를 정량화했으며, 가장 정확한 모델은 가장 낮은 RMSE, MASE, MAE 값을 사용하여 식별되었다. 기준 모델은 두 가지 범주로 특징지어진 시공간 대기질 예측 아키텍처의 혼합으로 구성되었는데, 스택형 공간 모델과 인코더-디코더 공간 모델이다
- [0164] 스택형 공간 모델은 RNN 변형(LSTM, GRU) 및 attention 기반 시계열 변환기(T-Trans) 프레임 워크가 포함되며, LSTM 및 GRU 모델은 스택 아키텍처로 구성된다. 시계열 변환기에는 인코더에 GRU 계층도 포함되어 있다. 스택 구조와 어텐션 매커니즘은 대상 영역에 위치한 공간 노드의 공기질 데이터를 고려하고 로컬 노드에서만 예측을 생성함으로써 단기 및 장기 이벤트의 효과적인 학습을 가능하게 하여 출력에서 프레임워크의 복잡성이 줄어든다. 그러나 고유한 다중 시간 예측 모델은 각 모니터링 스테이션에 대해 개발된다. 스택형 공간 모델은 오버피팅 문제를 방지하기 위해 드롭아웃 레이어가 있다. 또한 다변량 누적 공간 모델에 대한 데이터 세트는 슬라이딩 윈도우를 사용하여 준비된다. 여기서 각 측정소에서 대상 지역에 위치한 모든 측정소의 지난 12시간 센서 데이터가 입력값으로 사용되며, 출력값은 서로 다른 시간, 즉 T_{+1} , T_{+3} , T_{+6} , and T_{+12} 에서의 5개 1차 오염 물질로 구성된다.
- [0165] 인코더- 디코더 공간 모델도 기준 모델로 제안된 GCN간의 비교를 위해 고려되었다. 본 발명에서의 인코더는 연구 대상 영역의 모든 노드에서 사용할 수 있는 오염 물질 데이터를 기반으로 각 센서에 대한 숨겨진 표현을 매핑했다. 디코더는 모든 공기질 오염 물질에 대한 여러 사이트 예측을 동시에 수행하도록 설계되었다. 따라서 서로 다른 시간 범위에서 대상 지역의 오염 물질 센서의 다중 사이트 예측을 위해 단일 인코더-디코더 프레임 워크가 개발되었다.
- [0166] 그러나 스택형 공간 모델에 비해 모델 복잡성이 증가하여 효과적인 시공간 표현을 보장하기 위해 두 가지 네트워크 아키텍처가 활용되었다. 첫 번째 모델은 CNN(Convolutional Neural Network) 레이어를 사용하여 주변 노드에서 공간 정보를 추출한 반면, 두 번째 모델은 양방향 LSTM 레이어, 즉 CNN-LSTM 및 Bi-LSTM을 사용했으며, 오버피팅 문제를 방지하기 위해 드롭아웃 레이어가 네트워크 구조에 통합되었다. 스택형 공간 모델에 대한 입력에는 대상 영역에 있는 모든 측정소의 과거 12시간 센서 데이터가 포함되고 디코더는 미래의 T_{+1} , T_{+3} , T_{+6} , and T_{+12} 시간에 따라 모든 센서 위치에서 동시 예측을 생성한다.
- [0167] 도 18 내지 도 19는 본 발명의 일 실시예에 따른 학습 세트 및 테스트 세트에서의 서로 다른 시간 범위에서 예측 모델의 RMSE를 나타낸 도면이다.
- [0168] 도 18 내지 도 19를 참조하면, 학습 및 테스트 세트에서 각각 1, 3, 6, 12시간 범위에서 각 오염 물질 센서에 대한 평균 RMSE 항목으로 정량화된 18개의 알려진 측정소에서의 대기질 예측 성능을 보여준다. 결과는 학습 및 테스트 세트에 대한 프레임 워크를 예측하여 얻은 RMSE 점수의 적당한 변동을 보여주었다. 1시간과 3시간의 더 짧은 시간 범위에 대해 기준 모델과 DAGCGN은 각각 최대 12%와 8%의 성능 차이를 보였다.

- [0169] 또한 기준 모델은 12시간후 예측에 있어서 13~15%의 약간 더 높은 RMSE 값의 차이를 가졌다. 반면 DAGCGN 모델은 12%의 적당한 성능 변화를 보였다. 학습 세트와 테스트 세트에 대한 공기질 조건 사이에 차이가 있기 때문에 12시간 후 예측 범위에서의 적당한 변동은 합리적인 수준이다. 학습 세트와 테스트 세트 간의 예측 성능 비교는 낮은 오버피팅 문제로 잘 학습된 모델임을 나타낸다.
- [0170] 테스트 세트의 결과에 따르면, 1시간, 3시간의 짧은 시간 범위에서의 DAGCGN 모델은 더 높은 복잡성이 있음에도 불구하고 스택형 공간 모델을 능가하는 성능을 가졌다. 이것은 대상 영역의 시공간 표현이 학습되는 성능과 관련이 있을 수 있다. 1, 3시간 후의 PM_{2.5}를 보면 CNN-LSTM은 GRU, LSTM 및 T-Trans에 비해 9.05~5.71%, 17.89~9.06%, 5.78~2.12%의 성능 향상을 나타냈다. CNN-LSTM 모델은 컨볼루션 피쳐맵을 활용하여 여러 인접 노드에서 공간 상호 작용을 추출하고 LSTM 레이어와 쌓을 이를 때 모델이 시간 종속성을 학습할 수 있다. 따라서 CNN-LSTM은 Bi-LSTM 프레임 워크에 비해 더 긴 예측시간에서 우수한 성능을 보인다.
- [0171] DAGCGN 모델의 예측 효율성은 예측 범위가 클 때 감소한다. CNN-LSTM 및 Bi-LSTM 프레임워크는 알려진 모든 측정소에서 동시 예측을 허용하지만, 모델은 3시간 이상의 예측 작업에 대한 시간적 종속을 잃어버린다. 기준 모델중에서 T-Trans는 6시간 및 12시간 예측 범위에서 다른 기준 모델을 능가했다. T-Trans는 보다 포괄적인 정보를 추출하기 위해 다중 하위 공간 표현에서 입력 정보를 공동으로 검사하는 다중 헤드 어텐션 매커니즘을 사용한다. 또한 GRU 계층과 결합하면 단기 이벤트 외에도 모델은 장기 종속성에 대해 우수한 성능을 나타낸다. 이러한 결과를 바탕으로 공간 모델은 더 긴 예측 범위에서 기준 모델을 능가했다. 이러한 모델은 오염 농도의 조기 추정을 제공하기 위해 측정소에서 교육, 업데이트 및 유지 관리를 수반한다.
- [0172] 노드 샘플링 기술로 구성된 본 발명의 일 실시예에 따른 DAGCGN 모델은 먼 노드로부터의 노이즈를 효율적으로 최소화하고 다양한 시간 범위에 걸쳐 다양한 대기질 오염 물질에 대한 시공간 패턴을 학습했다. 12시간 시간 범위에서 PM_{2.5}의 경우 프레임워크는 GRU, LSTM, T-Trans, CNN-LSTM 및 bi-LSTM에 비해 45.67%, 44.87%, 43.89%, 52.59% 및 49.71%의 RMSE 성능 향상을 보였다. 각각 CO, O₃, NO₂, PM₁₀에 대한 예측 성능 개선 범위는 각각 33~35%, 32~45%, 45~48%, 39~46%이다. MASE의 성능 평가는 표 7에 나와있다. 결과를 보면, 서로 다른 오염 물질 센서에 대해 기본 모델의 학습 능력에 변화가 관찰되었다. 1시간 예측의 경우 MASE값의 범위는 불안정한 학습 능력을 나타내는 다양한 기준 모델에서 1.13에서 최대 5.04를 보였다. 이에 반하여, DAGCGN 모델은 MASE값의 적은 변동으로 모든 대기 오염 물질의 시간적 역학을 감지하여 합리적으로 강력한 조기 경보 시스템을 생성했다. 또한 기준 모델과의 비교를 통해 RMSE와 유사한 평가를 수행할 수 있다.
- [0173] 도 20은 본 발명의 일 실시예에 따른 DAGCGN 모델의 테스트 세트에서의 서로 다른 시간 범위에서 예측 모델의 MASE를 나타낸 도면이다
- [0174] 도 20을 참조하면, 기준 모델과 본 발명의 DAGCGN 모델의 성능을 시각적으로 평가하기 위해 주요 오염 물질 농도의 동적 거동을 포착하는 능력을 먼저 비교했다.
- [0175] 도 21은 본 발명의 일 실시예에 따른 DAGCGN 모델의 (a) S4 (b) S7 (c) S12 (d) S17 (e) S12 측정소의 예측된 PM_{2.5}의 농도에 관한 비교 도면이다.
- [0176] 도 21의 (a)를 참조하면, S12 측정소에서 1시간 후 예측에 대하여 측정값과 모델들의 예측값을 비교했다. 결과적으로 모든 예측 프레임워크는 PM_{2.5}의 시간적 추세를 포착하려고 하지만 측정값과 예측값 사이의 특정한 편차는 기준 모델의 예측을 기반으로 특정 시간 간격에서 관찰될 수 있으며 이는 시공간 정보를 잘못 인식한 결과로 볼 수 있다. 그러나, DAGCGN 모델은 PM_{2.5}에서 역동적인 추세에서 상당한 안정성을 보였다.
- [0177] 편향되지 않은 평가를 위해 신경망 프레임워크의 MAE점수는 시간 범위에 걸쳐 알려진 측정소의 각 센서에서 비교된다.
- [0178] 도 22는 본 발명의 일 실시예에 따른 DAGCGN 모델의 T₊₁, T₊₃, T₊₆, and T₊₁₂의 대상 지역 각 측정소에서 평균 1차 대기 오염물질 농도 분포 및 예측 MAE에 관한 도면이다.
- [0179] 도 22를 참조하면, 왼쪽축은 MAE를 나타내고 오른쪽 축은 막대그래프를 사용하여 각 노드의 평균 오염 물질 농도를 나타낸다.
- [0180] 도 22(a)를 참조하면, MAE는 각 노드의 평균 오염 물질 농도에 비해 상당히 낮다. 예를 들어 PM₁₀ 및 Pm_{2.5}의 평균 농도 범위는 38.13~52.9 $\mu\text{g}/\text{m}^3$ 과 20.93~31.72 $\mu\text{g}/\text{m}^3$ 이다. PM₁₀과 Pm_{2.5}를 고려한 기준 모델의 MAE 값은 9~15 $\mu\text{g}/\text{m}^3$ 과 5~10 $\mu\text{g}/\text{m}^3$ 으로 예측 모델이 적절하게 학습되었음을 나타낸다. 또한 기준 모델의 성능은 각 측정

소의 농도 프로파일에 따라 변경되었다.

- [0181] 기준 모델과 비교하여 본 발명의 DAGCGN 모델은 대부분의 대기 오염 물질에 대해 각 측정소에서 최소한의 노드별 변동으로 안정적인 성능을 보였다. 그러나 S9 및 S18 측정소의 CO센서에 대해서는 특정 변동이 관찰되었다. DAGCGN 모델의 향상된 성능은 모델이 노드 샘플링을 채택하여 원거리의 측정소에서 오는 노이즈를 피하고 예측 분산을 줄이는 노드별 학습에 기인할 수 있다. 거리 적응형 그래프 생성에 의해 제공되는 이웃 정보의 상당한 성능 향상은 특히 노드 12, 13, 17 및 18에서 관찰될 수 있으며 여기서 기준 모델들은 5개의 오염 물질에서 가장 낮은 성능을 보였다.
- [0182] 마지막으로, DAGCGN 모델의 여러 시간대에 걸친 예측 방법의 직관적이고 동시적인 비교를 위해 예측 MAE의 확률 밀도 분포를 조사했다.
- [0183] 도 23은 본 발명의 일 실시예에 따른 DAGCGN 모델의 대상 지역 전체에 걸친 다중 시간 범위에서의 CO, NO2, O3, PM10, PM2.5의 예측 MAE를 비교한 도면이다.
- [0184] 도 23을 참조하면, 기준 모델과 본 발명의 DAGCGN 모델에 대한 예측 MAE의 확률 밀도 곡선을 나타낸 것이다. 세로축은 1시간에서 12시간 범위에서의 전체 예측 MAE의 확률을 나타낸다. 기준 모델의 CO 및 NO2에 대한 MAE 확률분포는 모두 유사한 곡선을 가지며 한 모델이 다른 모델보다 명확한 이점을 나타내진 않았다. 나머지 밀도 분포 결과에 따르면 GRU 및 T-Trans 예측 오류의 대부분은 LSTM과 비교하여 왼쪽에 위치하여 다양한 시간 범위에 걸쳐 더 작은 오류를 나타낸다. 또한 T-Trans는 작은 오류가 발생할 확률이 높은 좁은 곡선으로 표시되는 인코더-디코더 프레임워크보다 성능이 더 높은 것으로 보인다. Bi-LSTM 및 CNN-LSTM은 오류가 적을 확률이 낮고 분산이 상대적으로 높다. 인코더-디코더 프레임워크에 의해 생성된 더 큰 오류는 도 19에 나온 것처럼 예측 범위가 증가하는 더 먼 시간대의 예측에서 주로 발생한다. 여러 시간대에 걸쳐 DAGCGN 모델로 인한 낮은 오류는 상대적으로 높은 확률로 확률 밀도 분포를 왼쪽으로 이동한다. CO에 대한 MAE 분포는 기준 모델과 비슷한 확률 비율을 나타내었지만 오류값은 더 낮았다. 나머지 오염 물질에 대한 결과는 DAGCGN 모델이 좁은 곡선으로 여러 시간대에 걸쳐 더 작은 오류를 생성할 수 있음을 나타내며 안정적인 성능과 높은 예측 정확도를 보여준다.
- [0185] 마지막으로, DAGCGN 모델의 계산 비용을 평가하기 위해 모델 매개변수와 학습시간을 기준 모델과 비교했다.
- [0186] 도 24는 본 발명의 일 실시예에 따른 DAGCGN 모델과 기준 모델의 복잡성과 학습시간을 비교한 도면이다.
- [0187] 도 24를 참조하면, 스택형 공간 모델의 경우 각 측정소에 대해 고유한 다변량 모델이 구성된다. 결과적으로 학습 시간은 각 측정소에 대해 계산된 다음 합산되어 에포크당 총 학습시간을 계산했다. 결과에 따르면 인코더-디코더 CNN-LSTM 및 BI-LSTM 프레임워크는 각각 에포크당 12초에서 14초 사이로 가장 낮은 학습시간을 보였는데 이러한 프레임워크는 가장 높은 예측 오류도 생성했다. 또한 노드별 패턴을 학습하기 위한 절충안으로 DAGCGN 모델에는 기본 T-Trans 모델보다 두 배 많은 모델 매개변수가 포함되어 있다. 가장 성능이 좋은 기준 모델인 T-Trans와 비교하여 DAGCGN 모델의 교육 시간은 17초 정도 적당한 수준으로 증가한다. 그럼에도 불구하고 이 모델은 RMSE 값이 최대 52% 감소하여 예측 능력이 상당히 향상됨을 보여준다. 광범위한 성능 평가 및 계산 비용 분석을 기반으로 한 결과는 최신 공기질 조기 경보 모델과 비교하여 본 발명의 DAGCGN 모델의 프레임워크가 단기 및 장기 시간 간격에서 높은 예측 성능을 제공한다는 것은 나타낸다.
- [0188] 기존의 조기 경보 프레임워크는 과거 센서 데이터를 기반으로 일정 기간 동안 오염 물질 예측을 제공한다. 그러나 실제 상황에서 센서의 신뢰성은 기상 조건, 유지 보수 및 열악한 모니터링 환경과 같은 다양한 변수에 의해 크게 손상된다. 따라서 기존의 프레임워크는 모든 센서 데이터가 정상 상태인 경우에만 적용이 가능하기 때문에 실용성이 크게 훼손되고 기존의 프레임워크를 이용한 예보는 충분한 모니터링 장비를 사용할 수 있는 위치에서만 사용할 수 있다.
- [0189] 본 발명의 일 실시예에 따른 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치(10)는 다음은 장기 센서 고장 및 미지의 위치에서 원격 조기 건강 위험 경고라는 두 가지 시나리오에서 개발되었다.
- [0190] 장기 센서 고장 시나리오에 대해 대상 측정소 S7, S11 및 S17을 선정했으며, 측정소에서 센서 데이터는 세 개의 간격으로 의도적으로 삭제되었다. 측정소에 대한 전체 센서 데이터 오류를 반영하는데 2주간은 2-3개의 센서의 부분적인 고장, 한 주간은 완전한 고장의 상황이다. 센서가 NaN값을 나타내면 이를 센서 고장으로 간주하고 로컬 및 공간 위치에 대해 식별된 노드로부터 입력 수신을 중지한다. 또한 식별된 측정소에 대해 거리 적응 그리드를 생성하고 센서 장애가 해결될때까지 오염 물질 농도를 예측한다.
- [0191] 장기간 발생한 센서의 결측치 처리는 실제 값과 예측값의 불일치를 기반으로 먼저 평가되었다.

- [0192] 도 25는 본 발명의 일 실시예에 따른 측정소 S7, S11, S17에서 부분 및 전체 센서 고장의 MAE 결측치 성능을 나타낸 도면이다.
- [0193] 도 25를 참조하면, 측정소 S7, S11 및 S17에서 부분 및 전체 센서 고장으로 결측치 성능의 MAE 값을 나타낸다. 시간 단계는 다양한 예측범위에서 이웃 노드를 활용하여 DAGCGN 모델에서 생성된 예측을 나타낸다. 결과에 따르면, 로컬 노드 정보 없이 생성된 장기 센서 고장에 대한 예측은 낮은 MAE 값으로 검증된 바와 같이 밀접한 유사성을 나타낸다.
- [0194] CO, NO2, O3, PM10, PM 2.5에 대한 MAE값은 각각 0.081~0.160, 0.004~0.007, 0.005~0.0155, 9.41~13.5, 5.75~8.12 범위였다. 오염 물질에 대한 MAE는 측정소에 따라 달랐으며, 이는 주로 이웃 지점과의 시공간적 관계 및 고장 기간 동안의 평균 오염 물질 농도와 관련이 있다. 예를 들어, S11은 PM2.5의 가장 낮은 농도
- [0195] 23.4 $\mu\text{g}/\text{m}^3$ 를 나타냈고, 5.91 $\mu\text{g}/\text{m}^3$ 의 가장 낮은 평균 MAE를 얻었다. 또한 평균 오염 물질 농도와 비교하여 낮은 MAE 값은 높은 결측치 처리 성능을 반영한다.
- [0196] 도 26 내지 도 28은 본 발명의 일 실시예에 따른 S-7, S-11 및 S-17 측정소에서 부분 및 전체 센서 고장에 대한 결측치 처리 성능에 관한 도면이다.
- [0197] 도 26 내지 도 28을 참조하면, DAGCGN 모델은 장기간의 센서 고장에 대한 주요 대기 오염 물질의 동적경향을 대부분 식별했음을 보여준다. PM10과 PM2.5에 대해 약간의 편차가 관찰되었지만, 이런 불일치는 복잡한 도시환경에서 높은 비선형성과 관련이 있으며 드문 경우이다. 이 조사는 실시간 모니터링을 위한 실제 센서 고장 문제를 해결하는데 본 발명의 방법론이 전반적인 성능을 보여준다. 이런 결측치 처리 성능은 다양한 실외 오염 물질의 농도가 높아짐에 따라 발생하는 정확한 조기 경보에 중요하다.
- [0198] 조기 경보 시스템은 사전에 사람들에게 알리고 오염을 방지하기 위한 규제 조치를 시작하는데 중요하다. 그러나 도시에는 불규칙적으로 배치되고 전체 지역을 포함하지 않는 제한된 수의 대기질 측정소가 있는 경우가 많으며 개발된 도시와 미개발된 도시의 모니터링 기능 차이가 있어 일반 대중이 정확한 건강에 대한 조기 경보를 받을 수 있는 능력이 제한된다. 새롭거나 미지의 위치에서 이러한 실제 문제를 해결하기 위해 5개의 미지의 위치에서 DAGCGN 모델의 원격 감지 및 예측 성능을 조사했는데 이 측정소의 데이터는 모델 학습의 어떤 부분에서도 활용되지 않았다. DAGCGN 모델은 동적 그래프에 의해 이미 학습된 은닉 표현을 활용하고, 알려진 위치에서 사용할 수 있는 공기질 정보로 미지의 위치까지의 거리를 할당하여 예측을 수행한다.
- [0199] 도 29는 본 발명의 일 실시예에 따른 대상 지역에서의 DAGCGN 모델의 공기질 비교에 관한 관측된 CAI와 예측된 CAI로, (a) 알려진 측정소 (b) 알려진 측정소 및 미지의 측정소에 관한 도면이다. (기간:2019.12.24~2019.12.28)
- [0200] 도 29를 참조하면, 센서가 없는 효과를 보여주기 위해 건강 위험 분포의 알려지지 않은 위치에 대한 비교는 전체 CAI 표면 분포 측면에서 이루어졌다. 대기질 분포는 대상 지역의 남쪽 부분이 보통에서 건강에 해로운 CAI를 갖는 경향이 있어 높은 PM10 및 Pm2.5농도를 초래한다는 것을 보여준다.
- [0201] 도 29를 참조하면, (a)는 미지의 측정소로 구분된 S3, S10, S15, S19, S20를 포함한 모든 측정소에 대한 CAI 분포를 보여준다. 모니터링 부족으로 인해 센서를 사용할 수 없는 경우 실제 건강 위험을 잘못 표현할 수 있음을 확인했다. CAI 분포의 차이는 특히 측정소 위치 S3, S10, S15, S19, S20에서 관찰할 수 있었는데 센서 비가용성은 S3, S10에서 높은 CAI 값을 보였으며, 실제 건강 위험은 중간 범위에 있었다. 또한 S15, S20에서는 반대 효과가 관찰될 수 있으며, 부적절한 추정은 건강에 해로운 대기질 조건을 과소평가하는 결과를 초래할 수 있다.
- [0202] 본 발명의 일 실시예에 따른 DAGCGN 모델은 미지의 위치에서 대기 오염 물질의 원격 감지 및 미래 예측을 용이하게 했다.
- [0203] 도 29를 참조하면, (b)는 조기 건강 위험 경고를 제공하는 3시간 후의 시간 범위에서 대상 영역의 CAI 분포를 보여준다. 관찰된 농도와 비교하여 CAI분포는 DAGCGN 모델은 대상 지역의 시공간 관계를 적절하게 설명하고 새로운 미지의 위치에서 예측할 수 있음을 보여준다. CAI분포는 주로 S3, S10, S15, S20 측정소에서 실제 CAI값과 예측된 CAI 값 사이의 밀접한 유사성을 관찰하여 검증되었다. 특히 그래프 기반 프레임워크의 성능은 측정소 사이의 거리가 멀면 관찰값과 예측값 사이의 불일치가 증가할 수 있는 인근 노드의 공간적 의존성에 민감하다. 광범위한 평가 및 시각적 비교를 통해 DAGCGN 모델은 오염 물질 예측 기능 측면에서 기존의 조기 경보 시스템을 모두 능가하고 장기간 센서 고장시 데이터 재구성을 허용하며 미지의 위치에서 원격 건강 위험의 추정을 용이하게 한다.

- [0204] 도 30은 본 발명의 일 실시예에 따른 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 방법에 관하여 도시한 도면이다.
- [0205] 도 30을 참조하면, 단계 S11에서 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치는 대상 지역 측정소의 공기질 측정 데이터를 입력한다
- [0206] 단계 S12에서 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치는 입력된 공기질 데이터를 전처리한다.
- [0207] 전처리 방법은 shape-preserving 조각별 3차 스플라인 보간법과 3차 에르미트 다항식 변형 보간법으로 사용하여 공기질 데이터를 전처리한다.
- [0208] 단계 S13에서 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치는 공기질 데이터의 시공간적 상관 관계를 분석한다.
- [0209] 단계 S13에서 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치는 시간별, 계절별, 대기 오염 물질간의 시계열 패턴을 분석하고 박스플롯을 이용하여 공기질 데이터의 시공간적 상관 관계를 분석한다.
- [0210] 단계 S14에서 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치는
- [0211] 전처리한 공기질 데이터를 입력하여 측정소가 없는 지점의 예측 공기질을 생성한다.
- [0212] 단계 S15에서 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치는 예측 공기질 데이터를 출력한다.
- [0213] 단계 S16에서 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치는
- [0214] 측정소의 공기질 데이터를 학습하여 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델을 생성한다.
- [0215] 단계 S16에서 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치는 측정소의 공기질 데이터를 학습하여 거리 적응형 그래프 컨볼루션 게이트 네트워크를 개발하고 모델의 성능을 검증한다.
- [0216] 본 발명의 일 실시예에 따른 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치는 잠재적으로 도시 대기질 측정소의 격차를 메우고 지속 가능하고 건강한 도시개발에 기여할 수 있다. 또한 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델의 효율성을 개선하기 위해 현재의 프레임워크는 건강 위험 추정과 관련된 정확한 대기질을 위해 전 지역으로 확장될 수 있으며, 이는 모니터링 역량이 부족한 지역에 특히 유용하다.
- [0217] 본 발명의 일 실시예에 따른 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치는 도시 환경의 주요 대기 오염 물질에 대한 조기 경보를 위해 센서 융합을 활용한 그래프 기반의 시공간 대기질 예측 프레임 워크를 개발했다. 이 발명은 여러 시간 범위에서 만족스러운 예측을 제공하기 위해 인접한 측정소 간의 복잡한 시공간 상호 작용을 식별할 수 있는 거리 적응형 그래프 컨볼루션 네트워크에 대한 설계이다. 또한 우리는 거리 적응형 그리드를 사용하여 새로운 사이트에서 장기 센서 오류 및 원격 감지와 같은 기존의 조기 경보 프레임워크와 관련된 문제를 해결한다. 거리 적응형 그래프 컨볼루션 네트워크에 여러 시공간 및 인코더-디코더 기반의 대기질 예측 프레임워크와 비교되었고, 결과는 다음과 같다.
- [0218] DAGCGN 모델은 기존 모델과 비교하여 강력한 대기질 예측 성능을 보여준다. 12시간 시간 범위에서 PM2.5의 경우 본 발명의 모델은 GRU, LSTM, T-Trans, CNN-LSTM, Bi-LSTM과 비교하여 45.67%, 44.87%, 43.89% 52.59%, 49.71%의 RMSE 감소로 성능을 증가한다.
- [0219] 예측 오류의 확률 밀도는 DAGCGN 모델이 여러 시간 범위에서 더 작은 예측 오류를 생성하여 꾸준한 성능과 높은 예측 정확도를 나타냄을 보여준다.
- [0220] DAGCGN 모델은 PM10과 PM2.5의 MAE값이 9.41-13.5 과 5.75-8.12로 평균 오염 물질 농도보다 5배 낮은 장기 센서 고장에 대한 결측치 처리 성능을 보여준다.
- [0221] DAGCGN 모델은 미지의 위치에서 원격 감지 및 공기질 예측이 대기질 지수가 나타내는 건강위험 분포와 밀접한 유사성을 보였다.
- [0222] 본 발명의 일 실시예에 따른 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치는 장기간 센서 장애가 발생하거나 공기질 측정소가 없는 지역의 공기질을 예측할 수 있다.
- [0223] 본 발명의 일 실시예에 따른 조기 경보 시스템은 거리 적응형 그래프 기반 방법론의 효율성을 입증하고 모니터

링이 없는 지역에서 대기질 관련 건강 위험을 평가할 수 있도록 한다

[0224] 본 발명에 의한 따른 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치는 위에서 실시예가 예시적으로 설명되었음에도 불구하고, 본 발명이 이의 취지 및 범주에서 벗어남 없이 여러 다른 형태로 구체화될 수 있다는 사실은 해당 기술에 통상의 지식을 가진 이들에게는 자명한 것이다.

[0225] 따라서, 상술된 실시예는 제한적인 것이 아닌 예시적인 것으로 여겨져야 하며, 첨부된 청구항 및 이의 동등 범위 내의 모든 실시예는 본 발명의 범주내에 포함된다.

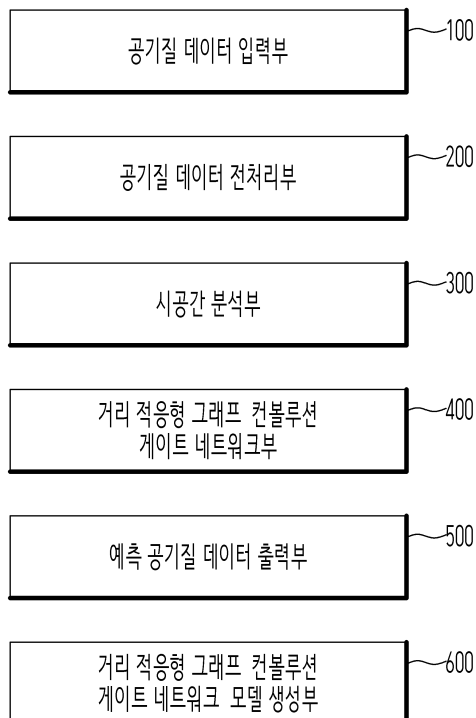
부호의 설명

[0227] 10 인공지능-그래프 신경망 기반 스마트 공기질 모니터링 장치
 100 공기질 데이터 입력부
 200 공기질 데이터 전처리부
 300 시공간 분석부
 400 거리 적응형 그래프 컨볼루션 게이트 네트워크부
 500 예측 공기질 데이터 출력부
 600 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 생성부
 620 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 개발부
 640 거리 적응형 그래프 컨볼루션 게이트 네트워크 모델 검증부

도면

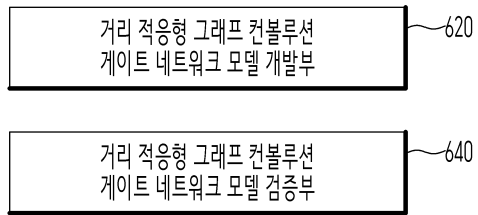
도면1

10

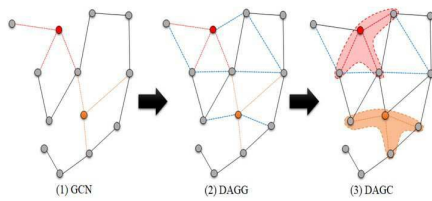


도면2

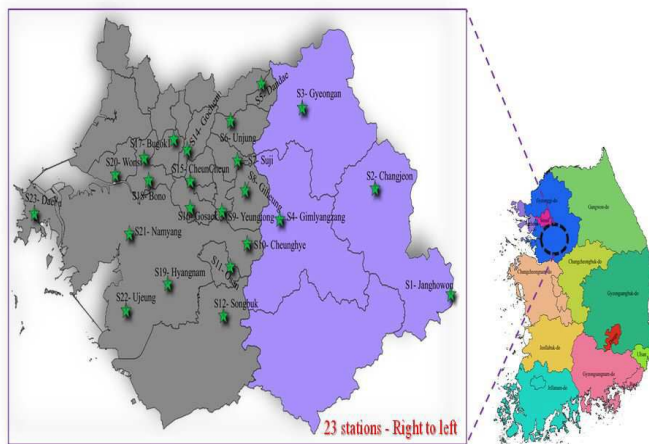
600



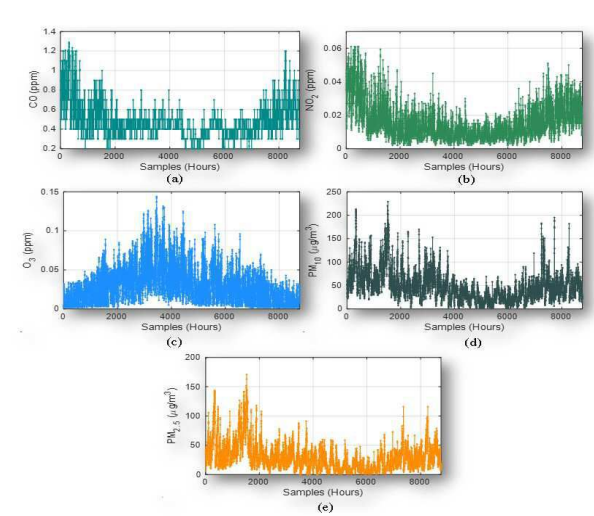
도면3



도면4



도면5

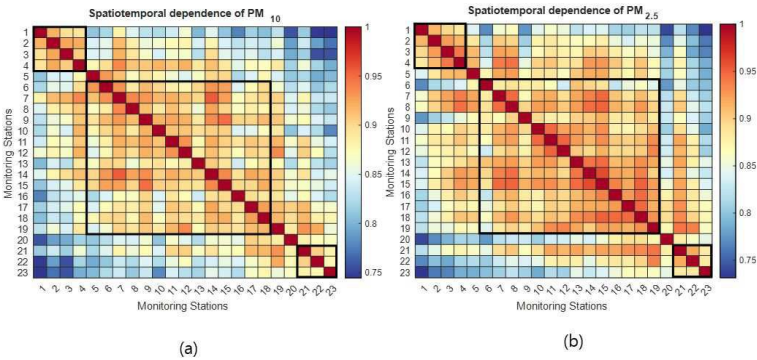


도면6

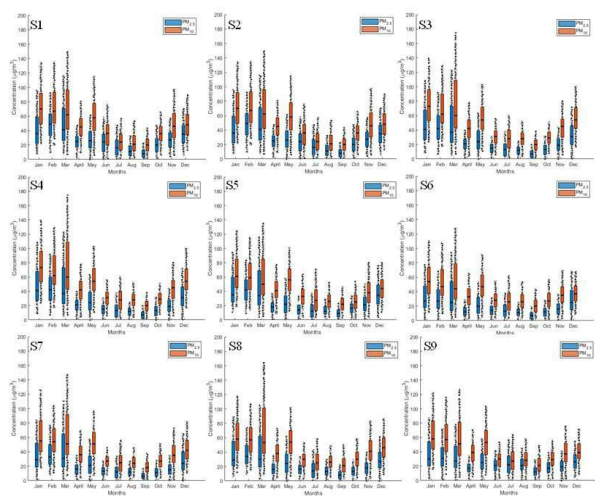
Basic statistics of collected variables. Unit, range, mean, standard deviation and Skewness from 1 January 2019 to 1 December 2019.

Variable	Unit	Range	Mean	St. dev.	Skewness
CO	ppm	[0.1,1.738]	0.51	0.19	1.34
NO ₂	ppm	[0.001,0.127]	0.02	0.01	0.92
O ₃	ppm	[0.001,0.144]	0.03	0.02	1.02
PM ₁₀	µg/m ³	[1,283]	45.30	31.19	1.88
PM _{2.5}	µg/m ³	[1,214]	25.41	22.26	2.24

도면7



도면8



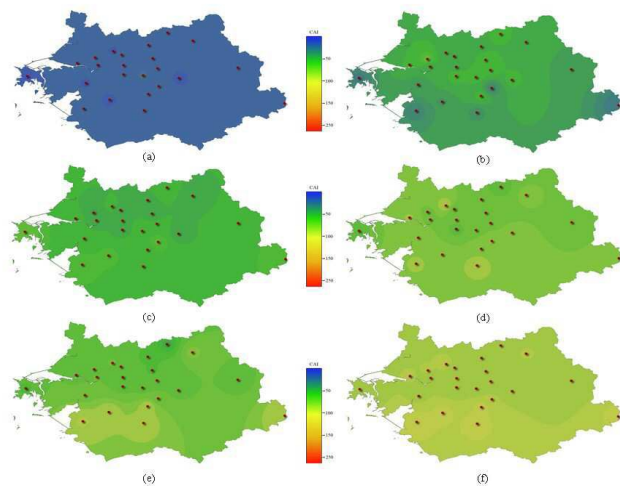
도면9

Comprehensive air quality index (CAI)

Description		Good	Moderate	Unhealthy	Very unhealthy
Values	I _{LO}	0	51	101	251
	I _{HI}	50	100	250	500
Concentration		BP _{LO}	BP _{HI}	BP _{LO}	BP _{HI}
SO ₂ (ppm)	1hr	0	0.02	0.021	0.05
CO (ppm)	1hr	0	2	2.1	9
O ₃ (ppm)	1hr	0	0.03	0.031	0.09
NO ₂ (ppm)	1hr	0	0.03	0.031	0.06
PM ₁₀ (μg/m ³)	24hr	0	30	31	80
PM _{2.5} (μg/m ³)	24hr	0	15	16	35

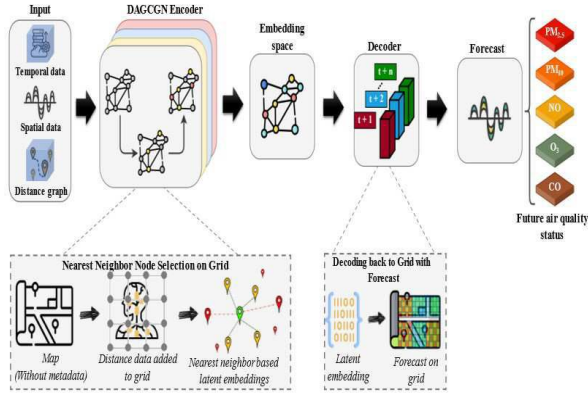
Note: I_{LO} is the index breakpoint corresponding to BP_{LO}, I_{HI} is the index breakpoint corresponding to BP_{HI}, and BP_{LO} and BP_{HI} are the concentration breakpoints of each health level.

도면10



도면11

400



도면12

Hyperparameter selection via random search in ray tune library.

No. of hidden layers	Embedding dimension	No. of GRU units in AGCGN cells	Dropout rate	Learning rate	Activation function	Loss (MSE)
3	10	128,64,32	0.1	0.0007	Tanh, Tanh, Tanh, Softmax	0.0081
2	2	32,32	0.2	0.005	Sigmoid, ReLU, Tanh, Softmax	0.0077
3	2	64,32,32	0.2	0.005	ReLU, ReLU, ReLU, Softmax	0.0065
3	6	64,32,32	0.3	0.001	Sigmoid, ReLU, ReLU, Softmax	0.0051
2	10	64,64	0.25	0.003	Sigmoid, Tanh, ReLU, Softmax	0.0047

도면13

Algorithm 1: Data preprocessing and training process of DAGCGN

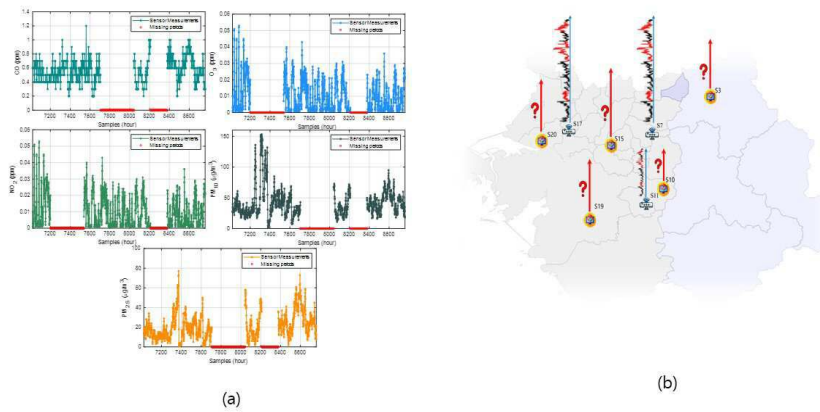
```

Input:  $D = \{(S_i, Y_i) | i = 1, \dots, N\}$ , /* dataset containing  $N$  data samples */
 $M = [m][m]$ , /* adjacency matrix containing distance between each sensor  $m$  */
 $K = \{k\}$ , /* list of  $k$  known sensors */
 $P \leftarrow$  number of nearest neighbors i.e., proximity
 $H \leftarrow$  number of historical timesteps used for training i.e., historical window
 $F \leftarrow$  number of future timesteps for forecast i.e., forecast horizon

1: for  $i$  in  $K$  do /* Data Preprocessing step */
2:   for  $j$  in  $K$  do
3:      $x, y = \text{transform}(D, i, j, M[i][j], P, H, F)$  /* transforms the dataset  $D$  into an input time-series  $S_i = \{s_{i,1}, \dots, s_{i,m}\} \in \mathbb{R}^{m \times p}$ , (i.e., Sensor values from  $m$  sources for  $H$  timesteps) and output time-series  $Y_i = \{y_{i,1}, \dots, y_{i,m}\} \in \mathbb{R}^{m \times p}$ , (i.e., Prediction for Sensors  $S_i$  for  $F$  timesteps) */
4:      $S_i.append(x)$ ,  $Y_i.append(y)$ 
5:   end
6: end
7: while  $i < N$  do /* Training step */
8:    $\mathcal{F}_\theta(S_i) \in \mathbb{R}^N \rightarrow \hat{Y}_i$ 
9:    $\mathcal{L}(Y_i, \hat{Y}_i)$ , /* where the loss is calculate based on the difference between forecast and ground truth i.e.,  $\mathcal{L}(Y_i, \hat{Y}_i)$ , using rmse, mae or mase */
10:  update weights using backpropagation based on the loss
11:   $i = i + 1$ 
12: end
Output:  $\mathcal{F}_\theta(\cdot)$  /* Trained Model */

```

도면14

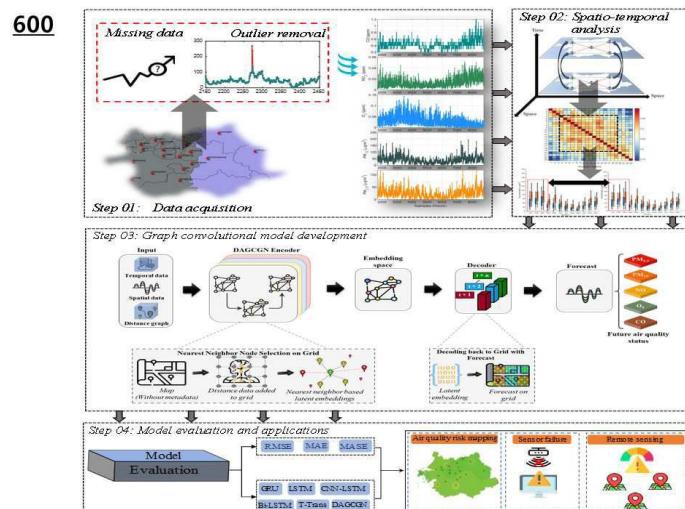


도면15

Partial and complete sensor failure intervals for application scenario 1.

Monitoring station	Air quality sensors					Missing interval	
	CO	O ₃	NO ₂	PM ₁₀	PM _{2.5}	Start	End
S7	✓	✗	✗	✓	✓	7201	7537
	✗	✓	✓	✗	✗	7707	8043
	✗	✗	✗	✗	✗	8209	8377
S11	✗	✗	✗	✗	✗	7201	7369
	✓	✗	✗	✓	✓	7707	8043
	✗	✓	✓	✗	✗	8209	8545
S17	✓	✗	✗	✓	✓	7201	7537
	✗	✗	✗	✗	✗	7707	7873
	✗	✓	✓	✗	✗	8209	8545

도면16



도면17

Algorithm 2: Forecast for sensor failure and remote sensing using trained DAGCGN

Input: $D = \{(S_i, Y_i) | i = 1, \dots, N\}$, /* dataset containing N data samples */
 $M = [m][m]$, /* adjacency matrix containing distance between each sensor m */
 $U = [u]$, /* list of u unknown stations */
 $F = [f]$, /* list of f faulty sensors */
 $P \leftarrow$ number of nearest neighbors i.e., proximity
 $H \leftarrow$ number of historical timestamps used for training i.e., historical window
 $F \leftarrow$ number of future timestamps for forecasts i.e., forecast horizon
 $\mathcal{F}_\theta(\cdot)$ /* Trained Model */

```

1: for  $i$  in  $K$  do /* Data Preprocessing step */
2:   for  $u$  in  $U$  do
3:      $x, y = \text{transform}(D, i, j, M[i][u], P, H, F)$ 
4:      $S_u^g.append(x), Y_u.append(y)$  /* here  $Y_u$  is the ground truth which can be
5:       used to validate the forecast  $\hat{Y}_u$  */
6:   end
7:   for  $f$  in  $F$  do
8:      $x, y = \text{transform}(D, i, j, M[i][f], P, H, F)$ 
9:      $S_f^f.append(x), Y_f.append(y)$  /* here  $Y_f$  is the ground truth which can be
10:      used to validate the forecast  $\hat{Y}_f$  */
11:   end
12: end
13:  $\hat{Y}_u = \mathcal{F}_\theta(S_u^g)$  /* Forecast for unknown locations */
14:  $\hat{Y}_f = \mathcal{F}_\theta(S_f^f)$  /* Forecast for faulty sensors */
Output:  $\hat{Y}_u, \hat{Y}_f$ 

```

도면18

RMSE of forecasting models at different temporal horizons on training set.

Time step	Model	CO (ppm)	O ₃ (ppm)	NO ₂ (ppm)	PM ₁₀ (μg/m ³)	PM _{2.5} (μg/m ³)
T=1	GRU	0.1207	0.0063	0.0082	12.9408	7.7823
	LSTM	0.1264	0.0066	0.0080	12.8433	8.5104
	T-Trans	0.1171	0.0059	0.0075	11.6309	7.5758
	Bi-LSTM	0.1038	0.0058	0.0067	10.4041	6.8155
	CNN-LSTM	0.1062	0.0058	0.0071	11.0244	7.0047
	DAGCGN	0.0895	0.0058	0.0052	9.7251	5.6128
T=3	GRU	0.1380	0.0095	0.0103	15.7429	9.0288
	LSTM	0.1382	0.0096	0.0100	15.7253	9.3886
	T-Trans	0.1316	0.0089	0.0100	14.5408	8.5553
	Bi-LSTM	0.1343	0.0107	0.0102	14.9342	9.5729
	CNN-LSTM	0.1327	0.0099	0.0101	14.0942	8.4906
	DAGCGN	0.1084	0.0065	0.0059	10.6319	6.1727
T=6	GRU	0.1623	0.0134	0.0123	19.2531	11.3557
	LSTM	0.1592	0.0142	0.0124	18.9854	11.5454
	T-Trans	0.1582	0.0129	0.0125	18.6336	10.8019
	Bi-LSTM	0.1598	0.0147	0.0123	20.6623	12.9867
	CNN-LSTM	0.1585	0.0156	0.0128	19.4076	11.9800
	DAGCGN	0.1214	0.0065	0.0057	12.4452	6.7033
T=12	GRU	0.1658	0.0126	0.0125	22.1422	13.3464
	LSTM	0.1665	0.0129	0.0122	21.9042	13.1019
	T-Trans	0.1660	0.0123	0.0125	21.6421	12.7739
	Bi-LSTM	0.1672	0.0144	0.0127	24.7832	15.2683
	CNN-LSTM	0.1699	0.0154	0.0130	22.9342	14.4250
	DAGCGN	0.1119	0.0084	0.0067	13.4816	7.3994

도면19

RMSE of forecasting models at different temporal horizons on test set.

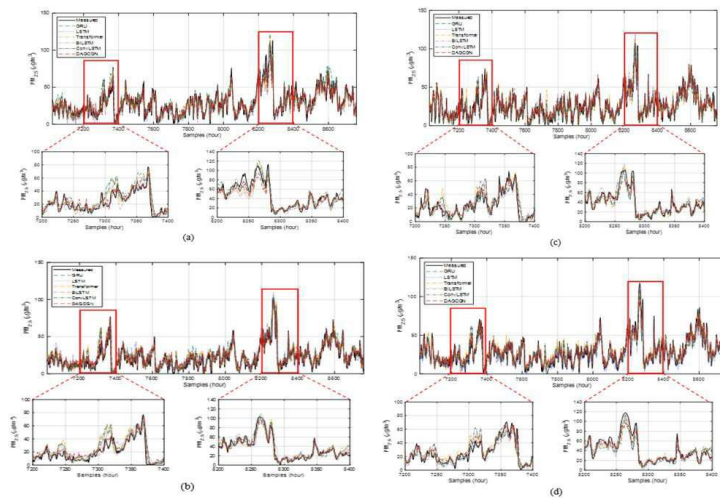
Time step	Model	CO (ppm)	O ₃ (ppm)	NO ₂ (ppm)	PM ₁₀ (μg/m ³)	PM _{2.5} (μg/m ³)
T=1	GRU	0.1325	0.0069	0.0090	13.8834	8.2870
	LSTM	0.1365	0.0072	0.0087	13.9166	9.1804
	T-Trans	0.1259	0.0065	0.0083	12.4826	8.0007
	Bi-LSTM	0.1131	0.0062	0.0073	11.1668	7.2429
	CNN-LSTM	0.1165	0.0064	0.0078	11.8501	7.5375
	DAGCGN	0.0939	0.0061	0.0055	10.1198	5.8345
T=3	GRU	0.1557	0.0107	0.0115	17.2531	9.8081
	LSTM	0.1542	0.0106	0.0111	17.1980	10.1694
	T-Trans	0.1480	0.0100	0.0111	16.0728	9.4485
	Bi-LSTM	0.1489	0.0119	0.0115	16.3197	10.3854
	CNN-LSTM	0.1476	0.0109	0.0112	15.4730	9.2479
	DAGCGN	0.1157	0.0069	0.0065	11.3710	6.5597
T=6	GRU	0.1832	0.0150	0.0139	21.7205	12.6986
	LSTM	0.1785	0.0158	0.0138	21.3737	12.8541
	T-Trans	0.1778	0.0145	0.0141	20.9687	11.9931
	Bi-LSTM	0.1814	0.0165	0.0139	22.9957	14.3716
	CNN-LSTM	0.1805	0.0178	0.0146	21.5322	13.2225
	DAGCGN	0.1327	0.0071	0.0063	13.5273	7.2783
T=12	GRU	0.1895	0.0144	0.0142	25.2917	15.1690
	LSTM	0.1910	0.0149	0.0140	24.9522	14.9485
	T-Trans	0.1902	0.0141	0.0143	24.7803	14.6876
	Bi-LSTM	0.1908	0.0164	0.0144	28.1421	17.3821
	CNN-LSTM	0.1938	0.0175	0.0148	25.9419	16.3861
	DAGCGN	0.1258	0.0095	0.0076	15.0464	8.2399

도면20

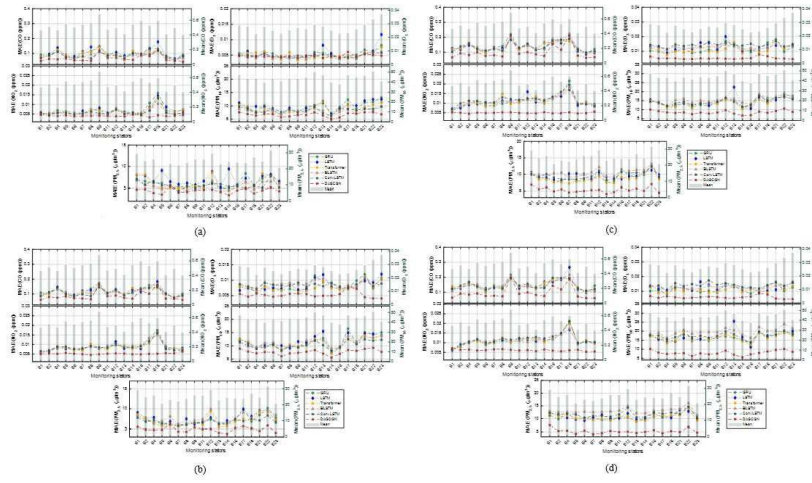
MASE of forecasting models at different temporal horizons on the test set.

Time step	Model	CO (ppm)	O ₃ (ppm)	NO ₂ (ppm)	PM ₁₀ (μg/m ³)	PM _{2.5} (μg/m ³)
T=1	GRU	5.0435	1.3695	3.5794	3.8341	3.4855
	LSTM	4.4820	1.3723	3.2065	3.1617	3.2490
	T-Trans	3.9986	1.1695	2.6531	2.7263	2.6632
	Bi-LSTM	3.7938	1.1357	2.4361	2.7117	2.7791
	CNN-LSTM	3.7626	1.1691	2.5654	2.6805	2.7532
	DAGCGN	1.8332	1.3774	0.8317	1.2500	1.2228
T=3	GRU	6.2253	2.1765	4.8476	4.5790	3.9713
	LSTM	5.3417	2.0319	4.3588	3.7492	3.4497
	T-Trans	4.7977	1.8260	3.9077	3.3305	2.9639
	Bi-LSTM	5.5327	2.2901	4.0742	4.0855	4.1536
	CNN-LSTM	5.2899	2.0130	4.1601	3.3323	3.2427
	DAGCGN	2.3522	1.4959	1.0302	1.4221	1.3469
T=6	GRU	8.2195	3.4637	7.1043	6.2023	5.3309
	LSTM	6.9995	3.4775	6.7133	5.2155	4.7077
	T-Trans	6.6562	3.0547	6.4254	4.6901	4.0059
	Bi-LSTM	6.9620	3.4087	5.2787	5.9812	5.6788
	CNN-LSTM	7.1100	3.6676	6.4663	4.9816	4.8383
	DAGCGN	2.6811	1.5596	1.0104	1.6043	1.4826
T=12	GRU	8.5225	3.5832	7.3970	7.5829	6.5461
	LSTM	7.8678	3.6801	6.8246	6.9339	5.8296
	T-Trans	7.4327	3.2954	6.7071	6.1767	5.2733
	Bi-LSTM	7.2707	3.4354	5.2928	7.1403	6.7850
	CNN-LSTM	7.6742	4.0352	6.5207	6.6361	6.2113
	DAGCGN	2.5418	2.0800	1.2425	1.7320	1.6681

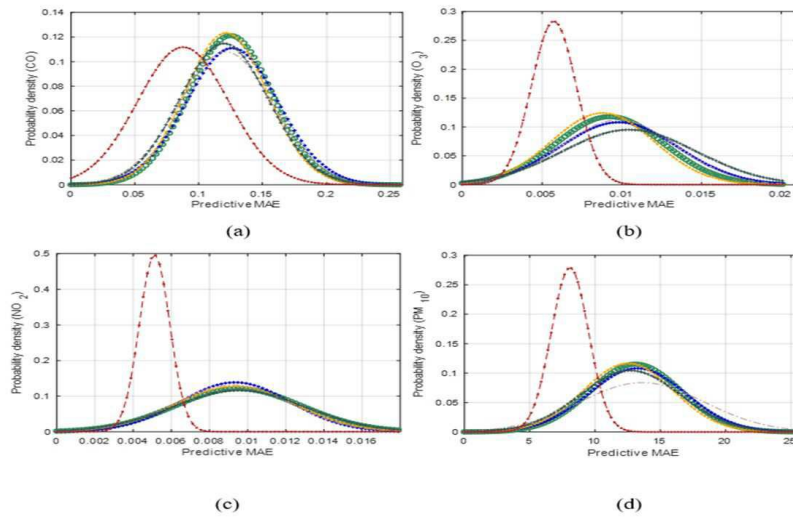
도면21



도면22



도면23



도면24

Comparison of computational complexity and training time of DAGCGN and baselines.

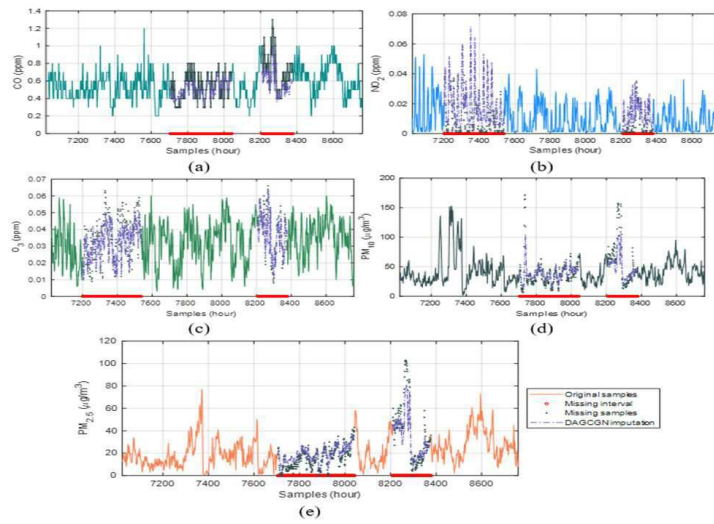
Model	#Parameters	Training Time (Epoch)
LSTM	79,164	54s (3s per epoch per station)
GRU	60,348	54s (3s per epoch per station)
T-Trans	335,814	72s (4s per epoch per station)
Bi-LSTM	297,656	14s
CNN-LSTM	197,656	12s
DAGCGN	745,099	95s

도면25

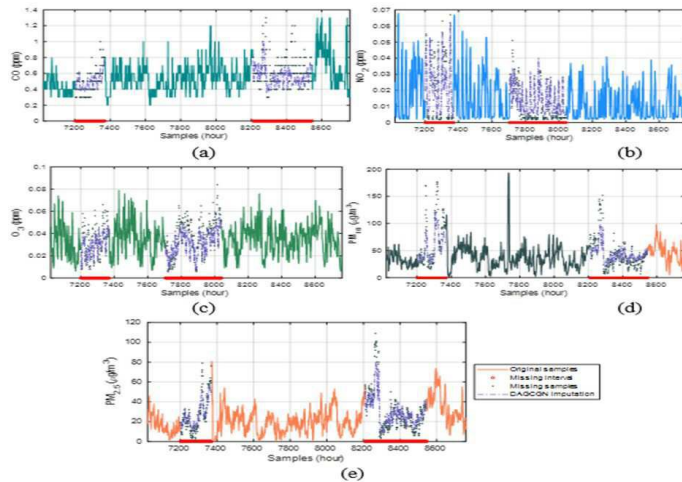
MAE based imputation performance for partial and complete sensor failure.

Time step	Station	CO	O ₃	NO ₂	PM ₁₀	PM _{2.5}
T+1	S7	0.0811	0.0064	0.0063	9.53	5.75
	S11	0.1146	0.0041	0.0054	11.16	5.84
	S17	0.1476	0.006	0.0114	11.70	6.68
T+3	S7	0.0867	0.0045	0.0066	9.41	5.84
	S11	0.1198	0.0056	0.0064	12.29	5.93
	S17	0.1373	0.0055	0.0135	11.12	7.20
T+6	S7	0.095	0.0065	0.0062	10.50	5.93
	S11	0.1107	0.0052	0.0067	12.73	6.06
	S17	0.1609	0.0059	0.0129	13.18	7.66
T+12	S7	0.0892	0.0075	0.0078	10.42	6.11
	S11	0.1248	0.0059	0.0091	13.47	6.33
	S17	0.1515	0.0067	0.0156	13.19	8.12

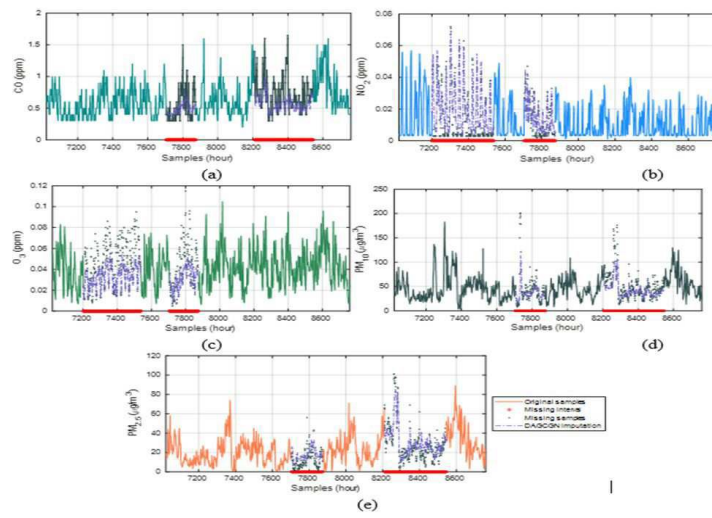
도면26



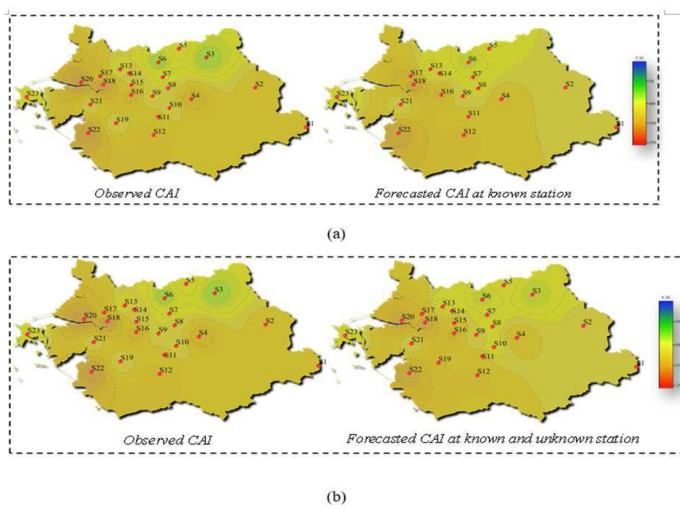
도면27



도면28



도면29



도면30

