



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0027068
(43) 공개일자 2025년02월25일

- (51) 국제특허분류(Int. Cl.)
G06Q 50/26 (2024.01) G06N 3/042 (2023.01)
G06N 3/045 (2023.01) G06N 3/0895 (2023.01)
- (52) CPC특허분류
G06Q 50/26 (2024.01)
G06N 3/042 (2023.01)
- (21) 출원번호 10-2023-0108363
(22) 출원일자 2023년08월18일
심사청구일자 2023년08월18일
- (71) 출원인
인하대학교 산학협력단
인천광역시 미추홀구 인하로 100(용현동, 인하대학교)
대한민국 (관리부서 : 환경부 국립환경과학원장)
인천 서구 환경로 42, 종합환경연구단지 (경서동)
- (72) 발명자
권장우
부산광역시 남구 오륙도로 85, 109동 2503호(용호동, 오륙도 에스케이뷰 아파트)
최정무
전라남도 광양시 금호로 73, 11동 401호(금호동, 사랑아파트)
(뒷면에 계속)
- (74) 대리인
양성보

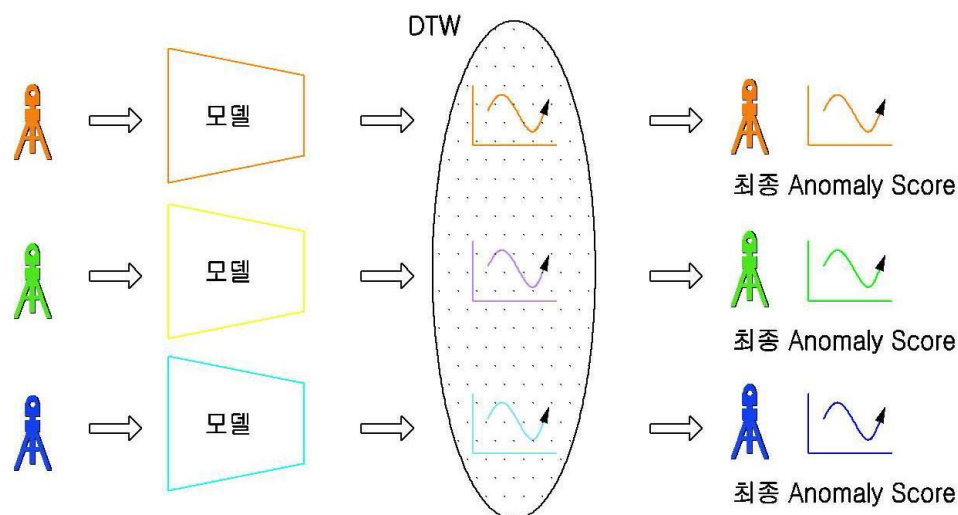
전체 청구항 수 : 총 12 항

(54) 발명의 명칭 대기관리권역별 미세먼지 고농도 우심지역에 대한 대기질 평가 시스템 및 방법

(57) 요약

대기관리권역별 미세먼지 고농도 우심지역에 대한 대기질 평가 시스템 및 방법이 개시된다. 일 실시예에 따른 대기질 평가 시스템에 의해 대기질 평가 방법은, 준지도 기반의 이상 데이터 검출 모델에 각 측정소별 대기질 데이터를 입력받는 단계; 상기 준지도 기반의 이상 데이터 검출 모델을 이용하여 각 측정소별 대기질 데이터로부터 각 측정소별 이상 데이터를 검출하는 단계; 상기 검출된 각 측정소별 이상 데이터에 대해 동적 시간 할당법(DTW, Dynamic Time Warping)을 적용함에 따라 각 측정소별 DTW 이상 데이터를 획득하는 단계; 및 상기 획득된 각 측정소별 DTW 이상 데이터에 대한 각 시점에서의 유사성 비교를 통해 이상 판정 결과를 제공하는 단계를 포함할 수 있다.

대표도



(52) CPC특허분류

G06N 3/045 (2023.01)

G06N 3/0895 (2023.01)

(72) 발명자

이선우

서울특별시 강서구 우장산로 82, 101동 102호(화곡동, 우장산길성그랑프리텔)

박지훈

인천광역시 서구 원당대로685번길 23, 204동 901호(마전동, 당하3차풍림아이원아파트)

정해진

인천광역시 서구 서곶로 12, 31동 206호 (가정동, 한신그랜드힐빌리지)

신혜정

서울특별시 영등포구 선유동2로 56, 202동 107호(당산동5가, 유원제일2차아파트)

박정민

인천광역시 서구 검암로 53, 105동 704호 (검암동, 풍림아이원2차)

명세서

청구범위

청구항 1

대기질 평가 시스템에 의해 대기질 평가 방법에 있어서,
 준지도 기반의 이상 데이터 검출 모델에 각 측정소별 대기질 데이터를 입력받는 단계;
 상기 준지도 기반의 이상 데이터 검출 모델을 이용하여 각 측정소별 대기질 데이터로부터 각 측정소별 이상 데이터를 검출하는 단계;
 상기 검출된 각 측정소별 이상 데이터에 대해 동적 시간 할당법(DTW, Dynamic Time Warping)을 적용함에 따라 각 측정소별 DTW 이상 데이터를 획득하는 단계; 및
 상기 획득된 각 측정소별 DTW 이상 데이터에 대한 각 시점에서의 유사성 비교를 통해 이상 판정 결과를 제공하는 단계
 를 포함하는 대기질 평가 방법.

청구항 2

제1항에 있어서,
 상기 제공하는 단계는,
 상기 획득된 각 측정소별 DTW 이상 데이터에 CPD(Change Point Detection) 알고리즘을 적용하여 추세가 변하는 구간을 산정하고, 상기 산정된 구간 내에서 이상으로 판정된 데이터가 존재할 경우, 구간 전체를 이상으로 판단하는 단계
 를 포함하는 대기질 평가 방법.

청구항 3

제1항에 있어서,
 상기 획득하는 단계는,
 지리적 특성에 기초하여 상기 획득된 각 측정소별 DTW 이상 데이터의 패턴 정보가 유사한 측정소들을 매칭하는 단계
 를 포함하는 대기질 평가 방법.

청구항 4

제1항에 있어서,
 상기 동적 시간 할당법은, 특정 시점을 기준으로 각 측정소별 대기질 데이터에 대한 특정 시점 이전의 구간 데이터와 특정 시점 이후의 구간 데이터의 유사도 점수가 도출되는 것을 특징으로 하는 대기질 평가 방법.

청구항 5

제1항에 있어서,
 상기 준지도 기반의 이상 데이터 검출 모델은,
 GNN(Graph Neural Network) 네트워크를 통해 상기 대기질 데이터에 포함된 복수 개의 성분 간의 관계성을 파악하도록 학습된 것을 특징으로 하는 대기질 평가 방법.

청구항 6

제1항에 있어서,

상기 준지도 기반의 이상 데이터 검출 모델은,

상기 대기질 데이터에 포함된 복수 개의 성분에 대한 관계성을 학습하는 제1 GAT 네트워크와 과거 데이터와의 관계성을 학습하는 제2 GAT 네트워크를 통해 추출된 각 특징 정보를 이용하여 상기 대기질 데이터의 이후 값을 예측하는 예측 모델과 상기 대기질 데이터의 원본 값을 복원하는 복원 모델을 동시에 학습시키는 것을 특징으로 하는 대기질 평가 방법.

청구항 7

대기질 평가 시스템에 있어서,

준지도 기반의 이상 데이터 검출 모델에 각 측정소별 대기질 데이터를 입력받는 데이터 입력부;

상기 준지도 기반의 이상 데이터 검출 모델을 이용하여 각 측정소별 대기질 데이터로부터 각 측정소별 이상 데이터를 검출하는 이상 데이터 검출부;

상기 검출된 각 측정소별 이상 데이터에 대해 동적 시간 할당법(DTW, Dynamic Time Warping)을 적용함에 따라 각 측정소별 DTW 이상 데이터를 획득하는 동적 시간 할당법 적용부; 및

상기 획득된 각 측정소별 DTW 이상 데이터에 대한 각 시점에서의 유사성 비교를 통해 이상 판정 결과를 제공하는 이상 판정 결과 제공부

를 포함하는 대기질 평가 시스템.

청구항 8

제7항에 있어서,

상기 이상 판정 결과 제공부는,

상기 획득된 각 측정소별 DTW 이상 데이터에 CPD(Change Point Detection) 알고리즘을 적용하여 추세가 변하는 구간을 산정하고, 상기 산정된 구간 내에서 이상으로 판정된 데이터가 존재할 경우, 구간 전체를 이상으로 판단하는

것을 특징으로 하는 대기질 평가 시스템.

청구항 9

제7항에 있어서,

상기 동적 시간 할당법 적용부는,

지리적 특성에 기초하여 상기 획득된 각 측정소별 DTW 이상 데이터의 패턴 정보가 유사한 측정소들을 매칭하는

것을 특징으로 하는 대기질 평가 시스템.

청구항 10

제7항에 있어서,

상기 동적 시간 할당법은, 특정 시점을 기준으로 각 측정소별 대기질 데이터에 대한 특정 시점 이전의 구간 데이터와 특정 시점 이후의 구간 데이터의 유사도 점수가 도출되는 것을 특징으로 하는 대기질 평가 시스템.

청구항 11

제7항에 있어서,

상기 준지도 기반의 이상 데이터 검출 모델은,

GNN(Graph Neural Network) 네트워크를 통해 상기 대기질 데이터에 포함된 복수 개의 성분 간의 관계성을 파악하도록 학습된 것을 특징으로 하는 대기질 평가 시스템.

청구항 12

제7항에 있어서,

상기 준지도 기반의 이상 데이터 검출 모델은,

상기 대기질 데이터에 포함된 복수 개의 성분에 대한 관계성을 학습하는 제1 GAT 네트워크와 과거 데이터와의 관계성을 학습하는 제2 GAT 네트워크를 통해 추출된 각 특징 정보를 이용하여 상기 대기질 데이터의 이후 값을 예측하는 예측 모델과 상기 대기질 데이터의 원본 값을 복원하는 복원 모델을 동시에 학습시키는 것을 특징으로 하는 대기질 평가 시스템.

발명의 설명

기술 분야

[0001] 아래의 설명은 이상 데이터 검출 기술에 관한 것이다.

배경 기술

[0003] 대기환경측정망 데이터는 경보체계 구축, 방지대책 수립의 기초자료, 대기예측모형의 보정자료, 환경 기준의 달성 여부, 대기배출원 감시 등 다양한 목적으로 활용될 수 있으며, 활용을 위해서는 데이터의 신뢰성 확보가 매우 중요하다. 대기환경 측정망 자료의 신뢰도 확보를 위하여 기존 대기오염측정망 자료확정 시스템에 적용할 수 있는 기술이 개발되었으며, 이러한 기술을 통해 대기환경 측정망 데이터의 기존 이상 레이블 중 상당수를 검출하였지만, 비교 분석을 위한 주변 측정소 데이터가 필수적으로 필요한 점, 비교 대상인 주변 측정소와 데이터 패턴이 상이할 경우 오탐지가 많이 발생할 수 있는 점 등 개선이 필요한 부분이 존재하였다.

[0004] 이에, 비교할 주변 측정소를 탐색하는 알고리즘을 개선하고, 베이스라인 이상 레이블의 기준을 정립하여 모델 학습을 개선함으로써 이상 데이터 검출 알고리즘의 적용성 및 신뢰도를 향상시키기 위한 기술이 요구된다.

발명의 내용

해결하려는 과제

[0006] 준지도 학습을 기반으로 대기관리권역별 미세먼지 고농도 우심지역에 대한 대기질을 평가하는 방법 및 시스템을 제공할 수 있다.

[0007] CPD 알고리즘을 적용하여 이상 판정 결과를 제공하는 방법 및 시스템을 제공할 수 있다.

과제의 해결 수단

[0009] 대기질 평가 시스템에 의해 대기질 평가 방법은, 준지도 기반의 이상 데이터 검출 모델에 각 측정소별 대기질 데이터를 입력받는 단계; 상기 준지도 기반의 이상 데이터 검출 모델을 이용하여 각 측정소별 대기질 데이터로부터 각 측정소별 이상 데이터를 검출하는 단계; 상기 검출된 각 측정소별 이상 데이터에 대해 동적 시간 할당법(DTW, Dynamic Time Warping)을 적용함에 따라 각 측정소별 DTW 이상 데이터를 획득하는 단계; 및 상기 획득된 각 측정소별 DTW 이상 데이터에 대한 각 시점에서의 유사성 비교를 통해 이상 판정 결과를 제공하는 단계를 포함할 수 있다.

[0010] 상기 제공하는 단계는, 상기 획득된 각 측정소별 DTW 이상 데이터에 CPD(Change Point Detection) 알고리즘을 적용하여 추세가 변하는 구간을 산정하고, 상기 산정된 구간 내에서 이상으로 판정된 데이터가 존재할 경우, 구간 전체를 이상으로 판단하는 단계를 포함할 수 있다.

[0011] 상기 획득하는 단계는, 지리적 특성에 기초하여 상기 획득된 각 측정소별 DTW 이상 데이터의 패턴 정보가 유사

한 측정소들을 매칭하는 단계를 포함할 수 있다.

- [0012] 상기 동적 시간 할당법은, 특정 시점을 기준으로 각 측정소별 대기질 데이터에 대한 특정 시점 이전의 구간 데이터와 특정 시점 이후의 구간 데이터의 유사도 점수가 도출되는 것일 수 있다.
- [0013] 상기 준지도 기반의 이상 데이터 검출 모델은, GNN(Graph Neural Network) 네트워크를 통해 상기 대기질 데이터에 포함된 복수 개의 성분 간의 관계성을 파악하도록 학습된 것일 수 있다.
- [0014] 상기 준지도 기반의 이상 데이터 검출 모델은, 상기 대기질 데이터에 포함된 복수 개의 성분에 대한 관계성을 학습하는 제1 GAT 네트워크와 과거 데이터와의 관계성을 학습하는 제2 GAT 네트워크를 통해 추출된 각 특징 정보를 이용하여 상기 대기질 데이터의 이후 값을 예측하는 예측 모델과 상기 대기질 데이터의 원본 값을 복원하는 복원 모델을 동시에 학습시키는 것일 수 있다.
- [0015] 대기질 평가 시스템은, 준지도 기반의 이상 데이터 검출 모델에 각 측정소별 대기질 데이터를 입력받는 데이터 입력부; 상기 준지도 기반의 이상 데이터 검출 모델을 이용하여 각 측정소별 대기질 데이터로부터 각 측정소별 이상 데이터를 검출하는 이상 데이터 검출부; 상기 검출된 각 측정소별 이상 데이터에 대해 동적 시간 할당법(DTW, Dynamic Time Warping)을 적용함에 따라 각 측정소별 DTW 이상 데이터를 획득하는 동적 시간 할당법 적용부; 및 상기 획득된 각 측정소별 DTW 이상 데이터에 대한 각 시점에서의 유사성 비교를 통해 이상 판정 결과를 제공하는 이상 판정 결과 제공부를 포함할 수 있다.
- [0016] 상기 이상 판정 결과 제공부는, 상기 획득된 각 측정소별 DTW 이상 데이터에 CPD(Change Point Detection) 알고리즘을 적용하여 추세가 변하는 구간을 산정하고, 상기 산정된 구간 내에서 이상으로 판정된 데이터가 존재할 경우, 구간 전체를 이상으로 판단할 수 있다.
- [0017] 상기 동적 시간 할당법 적용부는, 지리적 특성에 기초하여 상기 획득된 각 측정소별 DTW 이상 데이터의 패턴 정보가 유사한 측정소들을 매칭할 수 있다.
- [0018] 상기 동적 시간 할당법은, 특정 시점을 기준으로 각 측정소별 대기질 데이터에 대한 특정 시점 이전의 구간 데이터와 특정 시점 이후의 구간 데이터의 유사도 점수가 도출되는 것일 수 있다.
- [0019] 상기 준지도 기반의 이상 데이터 검출 모델은, GNN(Graph Neural Network) 네트워크를 통해 상기 대기질 데이터에 포함된 복수 개의 성분 간의 관계성을 파악하도록 학습된 것일 수 있다.
- [0020] 상기 준지도 기반의 이상 데이터 검출 모델은, 상기 대기질 데이터에 포함된 복수 개의 성분에 대한 관계성을 학습하는 제1 GAT 네트워크와 과거 데이터와의 관계성을 학습하는 제2 GAT 네트워크를 통해 추출된 각 특징 정보를 이용하여 상기 대기질 데이터의 이후 값을 예측하는 예측 모델과 상기 대기질 데이터의 원본 값을 복원하는 복원 모델을 동시에 학습시키는 것일 수 있다.

발명의 효과

- [0022] 준지도 학습과 CPD 알고리즘을 적용하여 이상 판정 결과의 오탐지율을 감소시킬 뿐 아니라, 이상 판정 결과의 정확도를 향상시킬 수 있다.
- [0023] 준지도 학습 이후에 동적 시간 할당법을 적용하여 인근에 다른 측정소가 존재하지 않는 측정소에 대해서도 대기질 데이터의 이상 판정을 적용할 수 있다.

도면의 간단한 설명

- [0025] 도 1은 일 실시예에 있어서, 대기질 평가 시스템의 개괄적인 동작을 설명하기 위한 도면이다.
- 도 2는 일 실시예에 있어서, 대기질 평가 시스템의 구성을 설명하기 위한 블록도이다.
- 도 3은 일 실시예에 있어서, 대기질 평가 방법을 설명하기 위한 흐름도이다.
- 도 4는 일 실시예에 있어서, 포항 대도동 측정소 리매칭 결과의 예이다.
- 도 5는 일 실시예에 있어서, CPD 알고리즘을 이용한 포항 대도동 측정소 리매칭 결과를 나타낸 예이다.

도 6은 일 실시예에 있어서, CPD 기반 패턴 변화 구간을 적용한 예이다.

도 7은 일 실시예에 있어서, 준지도 기반의 이상 데이터 검출 모델의 예이다.

도 8은 일 실시예에 있어서, 준지도 기반의 이상 데이터 검출 모델의 예이다.

도 9는 일 실시예에 있어서, 준지도 기반의 이상 데이터 검출 모델의 입력 데이터를 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0026] 이하, 실시예를 첨부한 도면을 참조하여 상세히 설명한다.

[0028] 도 1은 일 실시예에 있어서, 대기질 평가 동작을 설명하기 위한 도면이다.

[0029] 대기질 평가 시스템은 대기질 데이터로부터 1)동일값 지속, 2)급격한 변화, 3)베이스라인 및 추세 이상, 4)PM 10, PM 2.5 역전, 5)주변측정소 대비 이상, 6)NOx, NO, NO2 비율 이상 등의 이상 증상을 검출할 수 있다. 이때, 대기질 데이터는 대기오염측정망 데이터 또는 대기환경 데이터를 포함한다. 대기오염측정망 데이터는 전국 가지에 500개소 이상이 설치되어 운영되는 대기오염측정망으로부터 측정된 데이터를 의미하고, 대기환경 데이터는 대기환경연구소(예를 들면, 백령도, 수도권, 호남권, 중부권, 영남권, 충청권, 경기권, 제주권)로부터 측정된 데이터를 의미한다. 예를 들면, 대기질 평가 시스템은 권역별 대기질 특성을 파악하고 고농도 대기오염 현상의 발생 원인을 규명하기 위한 목적으로 전국 각 권역에 구축 및 운영되고 있는 대기환경연구소에서 측정된 대기환경 데이터를 대상으로 이상 증상을 검출할 수 있다. 대기환경연구소의 주요 측정 대상은 대기환경연구소의 주요 측정 대상은 미세먼지(PM 1.0, PM 10, PM 2.5), 탄소성분(OC, EC), 이온성분(황산염, 질산염, 암모늄염 등), 중금속 성분(납, 칼슘, 망간, 니켈, 아연 등)이다.

[0030] 일례로, 환경과학원과의 협의를 통해 서울 44개 측정소에 대해 베이스라인 이상 증상에 대해 이상 기준을 정립하고 2021년도 데이터에 대해 레이블이 제작될 수 있다. 이를 기반으로 대기질 평가 시스템은 준지도 기반의 이상 데이터 검출 모델에 각 측정소별 대기오염측정망 데이터를 입력할 수 있다. 준지도 기반의 이상 데이터 검출 모델은 각 측정소별 대기오염측정망 데이터의 주변 환경에 따른 변화 패턴의 비교를 통해 이상 데이터가 검출되도록 학습된 것일 수 있다. 여기서, 준지도 학습이란 일부만 레이블이 적용된 데이터를 사용하여 학습을 진행하는 기법으로, 처음에는 레이블된 데이터 위주로 학습을 진행하다가 이를 확장하여 레이블되지 않은 데이터에 적용하는 방식으로 진행된다. 예를 들면, 대기질 평가 시스템은 이상 레이블이 정립된 2021년도 대기질 데이터와 이전 연도(2016~2020) 대기질 데이터를 합쳐 준지도 기반의 데이터 검출 모델에 입력할 수 있다. 이에, 이상 레이블을 정립하고 준지도 기법의 적용이 가능해짐에 따라 노이즈가 많은 원본 데이터도 입력으로 사용이 가능해졌으며, 도 1과 같이 주변 측정소와 비교하는 단계를 모델 판정 이후로 변경하여 모델이 검출한 이상 패턴이 기기 이상으로 인한 것인지 인근에서 발생한 특정 이벤트(예를 들면, 공사, 행사 등) 때문인지를 판단할 수 있다.

[0031] 이때, 대기질 평가 시스템은 대기질 데이터의 결측 데이터를 처리를 수행한 후, 준지도 기반의 데이터 검출 모델에 입력할 수 있다. 대기오염측정망 데이터와 동일하게 대기환경 데이터 또한 계절 변화에 따른 데이터 변화 패턴이 뚜렷하게 나타나기 때문에 결측 데이터의 대체에 이전 또는 이후 연도의 같은 시기 데이터가 활용될 수 있다. 대기질 평가 시스템은 이전, 이후 연도의 데이터를 사용하되 이전, 이후 3년 이내의 데이터를 모두 사용하고, 더 이전 및 이후의 데이터일수록 낮은 가중치를 적용하는 Exponential Smoothing 기법을 적용할 수 있다.

[0032] 대기질 평가 시스템은 준지도 기반의 이상 데이터 검출 모델을 통해 이상 데이터를 검출할 수 있다. 대기질 평가 시스템은 준지도 기반의 이상 데이터 검출 모델을 통해 검출된 이상 데이터에 대해 동적 시간 할당법을 적용할 수 있다. 예를 들면, A 측정소, B 측정소 및 C 측정소가 존재하고, A 측정소의 이상 구간을 산출하는 것을 가정하기로 한다. 대기질 평가 시스템은 준지도 기반의 이상 데이터 검출 모델을 통해 검출된 A측정소에 대한 이상 데이터와 B 측정소에 대한 이상 데이터에 동적 시간 할당법을 적용하고, A 측정소에 대한 이상 데이터와 C 측정소에 대한 이상 데이터에 동적 시간 할당법을 적용할 수 있다. 마찬가지로, 대기질 평가 시스템은 B 측정소에 대한 이상 데이터와 C 측정소에 대한 이상 데이터에 동적 시간 할당법을 적용할 수 있다. 그러나, A 측정소의 이상 구간을 산출하는 것을 예를 들어 설명하기로 하며, 이하에서는 B 측정소와 C측정소에 대한 설명은 생략하기로 한다.

[0033] 대기질 평가 시스템은 검출된 각 측정소별 이상 데이터에 대해 동적 시간 할당법(DTW, Dynamic Time Warping)을

적용함에 따라 각 측정소별 DTW 이상 데이터를 획득할 수 있다. 이때, 동적 시간 할당법이란, 시계열 분석에서 비슷한 두 개의 서로 다른 시간축의 유사성을 측정하는 알고리즘으로, 특정 시점을 기준으로 각 측정소별 대기질 데이터에 대한 특정 시점 이전의 구간 데이터와 특정 시점 이후의 구간 데이터의 유사도 점수를 도출하며, 두 개의 다른 시점(인접 시점)의 데이터 쌍을 비교하기 때문에 시점이 틀어지거나 길어가 다른 데이터의 비교가 가능하다. 예를 들면, 대기질 평가 시스템은 A 측정소에 대한 이상 데이터와 B 측정소에 대한 이상 데이터에 동적 시간 할당법을 적용한 제1 DTW 데이터와 A 측정소에 대한 이상 데이터와 C 측정소에 대한 이상 데이터에 동적 시간 할당법을 적용한 제2 DTW 데이터를 획득할 수 있다.

[0034] 대기질 평가 시스템은 획득된 각 측정소별 DTW 이상 데이터에 대한 각 시점에서의 유사성 비교를 통해 이상 판정 결과를 제공할 수 있다. 대기질 평가 시스템은 시점 별 측정값과 예측값의 차이를 이상 점수로 나타낼 수 있다. 예를 들면, 대기질 평가 시스템은 제1 DTW 데이터의 예측값과 측정값과의 차이를 이상 점수로 나타낼 수 있고, 제2 DTW 데이터의 예측값과 측정값과의 차이를 이상 점수로 나타낼 수 있다. 대기질 평가 시스템은 이상 점수가 일정 수치 일정 수치(threshold)를 초과한 부분을 이상 구간으로 선정할 수 있다. 이상 데이터 검출 시스템은 선정된 복수 개의 이상 구간에 AND 연산을 적용하여 특정 측정소(예를 들면, A 측정소)의 이상 구간을 확정할 수 있다.

[0035] 도 2는 일 실시예에 있어서, 대기질 평가 시스템의 구성을 설명하기 위한 블록도이고, 도 3은 일 실시예에 있어서, 대기질 평가 방법을 설명하기 위한 흐름도이다.

[0036] 대기질 평가 시스템(100)의 프로세서는 데이터 입력부(210), 이상 데이터 검출부(220), 동적 시간 할당법 적용부(230) 및 이상 판정 결과 제공부(240)를 포함할 수 있다. 이러한 프로세서의 구성요소들은 대기질 평가 시스템에 저장된 프로그램 코드가 제공하는 제어 명령에 따라 프로세서에 의해 수행되는 서로 다른 기능들(different functions)의 표현들일 수 있다. 프로세서 및 프로세서의 구성요소들은 도 3의 대기질 평가 방법이 포함하는 단계들(310 내지 340)을 수행하도록 대기질 평가 시스템을 제어할 수 있다. 이때, 프로세서 및 프로세서의 구성요소들은 메모리가 포함하는 운영체제의 코드와 적어도 하나의 프로그램의 코드에 따른 명령(instruction)을 실행하도록 구현될 수 있다.

[0037] 프로세서는 대기질 평가 방법을 위한 프로그램의 파일에 저장된 프로그램 코드를 메모리에 로딩할 수 있다. 예를 들면, 대기질 평가 시스템에서 프로그램이 실행되면, 프로세서는 운영체제의 제어에 따라 프로그램의 파일로부터 프로그램 코드를 메모리에 로딩하도록 대기질 평가 시스템을 제어할 수 있다. 이때, 데이터 입력부(210), 이상 데이터 검출부(220), 동적 시간 할당법 적용부(230) 및 이상 판정 결과 제공부(240) 각각은 메모리에 로딩된 프로그램 코드 중 대응하는 부분의 명령을 실행하여 이후 단계들(310 내지 340)을 실행하기 위한 프로세서의 서로 다른 기능적 표현들일 수 있다.

[0038] 단계(310)에서 데이터 입력부(210)는 준지도 기반의 이상 데이터 검출 모델에 각 측정소별 대기질 데이터를 입력받을 수 있다. 이때, 준지도 기반의 이상 데이터 검출 모델은 GNN(Graph Neural Network) 네트워크를 통해 대기질 데이터에 포함된 복수 개의 성분 간의 관계성을 파악하도록 학습된 것일 수 있다. 또는, 준지도 기반의 이상 데이터 검출 모델은 대기질 데이터에 포함된 복수 개의 성분에 대한 관계성을 학습하는 제1 GAT 네트워크와 과거 데이터와의 관계성을 학습하는 제2 GAT 네트워크를 통해 추출된 각 특징 정보를 이용하여 대기질 데이터의 이후 값을 예측하는 예측 모델과 대기질 데이터의 원본 값을 복원하는 복원 모델을 동시에 학습시킬 수 있다.

[0039] 단계(320)에서 이상 데이터 검출부(220)는 준지도 기반의 이상 데이터 검출 모델을 이용하여 각 측정소별 대기질 데이터로부터 각 측정소별 이상 데이터를 검출할 수 있다.

[0040] 단계(330)에서 동적 시간 할당법 적용부(230)는 검출된 각 측정소별 이상 데이터에 대해 동적 시간 할당법(DTW, Dynamic Time Warping)을 적용함에 따라 각 측정소별 DTW 이상 데이터를 획득할 수 있다. 동적 시간 할당법 적용부(230)는 지리적 특성에 기초하여 획득된 각 측정소별 DTW 이상 데이터의 패턴 정보가 유사한 측정소들을 매칭할 수 있다. 여기서, 동적 시간 할당법이란, 특정 시점을 기준으로 각 측정소별 대기질 데이터에 대한 특정 시점 이전의 구간 데이터와 특정 시점 이후의 구간 데이터의 유사도 점수가 도출되는 것을 의미한다.

[0041] 단계(340)에서 이상 판정 결과 제공부(240)는 획득된 각 측정소별 DTW 이상 데이터에 대한 각 시점에서의 유사성 비교를 통해 이상 판정 결과를 제공할 수 있다. 이상 판정 결과 제공부(240)는 획득된 각 측정소별 DTW 이상 데이터에 CPD(Change Point Detection) 알고리즘을 적용하여 추세가 변하는 구간을 산정하고, 산정된 구간 내에서 이상으로 판정된 데이터가 존재할 경우, 구간 전체를 이상으로 판단할 수 있다.

- [0042] 도 4는 일 실시예에 있어서, 포항 대도동 측정소 리매칭 결과를 나타낸 예이다.
- [0043] 대기질 평가 시스템은 지리적 특성에 기초하여 획득된 각 측정소별 DTW 이상 데이터의 패턴 정보가 유사한 측정소들을 매칭할 수 있다. 지리적 특성이 고려되지 않은 매칭으로 인하여 오탐지가 발생하였던 포항 대도동 측정소의 매칭 결과를 지리적 특성을 반영한 매칭 결과 기존 매칭 대비 유사한 데이터 패턴을 가지는 측정소들과 매칭된 것을 확인할 수 있다.
- [0044] 도 5는 일 실시예에 있어서, CPD 알고리즘을 이용한 포항 대도동 측정소 리매칭 결과를 나타낸 예이다.
- [0045] 대기질 평가 시스템은 획득된 각 측정소별 DTW 이상 데이터에 CPD(Change Point Detection) 알고리즘을 적용하여 이상 구간을 판단할 수 있다. CPD 알고리즘은 시계열 데이터에서 급격하게 추세(trend)가 변하는 지점을 찾는 알고리즘으로, 주로 최대 우도 추정(Maximum Likelihood estimation, MLE)을 기반으로 진행되며, 실시예에서는 이진 분류(Binary Segmentation) 기반의 알고리즘이 적용될 수 있다.
- [0046] 이진 분류 알고리즘은 이진 탐색 알고리즘과 유사하게 진행되며, 입력된 시계열 데이터에서 변화 지점을 찾은 뒤 변화 지점을 중심으로 시계열 데이터를 분할하여 두 하위 신호에 대해 해당 과정을 반복한다. 도 5를 참고하면, 이진 분류 알고리즘의 진행 과정을 나타낸 것이다.
- [0047] 도 6을 참고하면, CPD 기반 패턴 변화 구간을 적용한 예이다. 대기질 평가 시스템은 명확한 이상 구간 제시를 위해, 획득된 각 측정소별 DTW 이상 데이터에 CPD(Change Point Detection) 알고리즘을 적용하여 추세가 변하는 구간을 산정하고, 산정된 구간 내에서 이상으로 판정된 데이터가 존재할 경우, 구간 전체를 이상으로 판단할 수 있다. 이와 같이, 기존 알고리즘 대비 명확한 이상 판정 결과를 제시할 수 있음을 확인할 수 있다.
- [0048] 또한, 대기질 평가 시스템은 전이 학습(Transfer Learning)을 통한 전체 측정소를 적용할 수 있다. 레이블 기준 정립 및 재 레이블링 작업은 주로 서울 지역 측정소에 대하여 진행 되었으며, 이러한 과정을 전체 측정소에 적용하는 것은 많은 인력을 요구하기 때문에 비효율적이다. 따라서 전체 측정소에 새로 정립한 레이블 기준을 적용하기 위해, 대기질 평가 시스템은 서울 측정소 데이터 및 레이블을 통해 학습한 모델의 가중치 정보를 다른 모델에 전이(Transfer)시킨 후 재학습을 진행하는 전이 학습 기법을 사용할 수 있다.
- [0049] 도 7은 일 실시예에 있어서, 준지도 기반의 이상 데이터 검출 모델의 예이다.
- [0050] 대기질 평가 시스템은 대기환경 데이터가 대기오염측정망 데이터의 8개 항목 대비 4배 이상 많은 30개의 측정 항목을 가지고 있으며, 각 항목 간의 상관성을 파악하기 위해 준지도 기반의 이상 데이터 검출 모델에 GNN(Graph Neural Network)을 구성할 수 있다. 대기질 평가 시스템은 GNN을 통해 각 노드 간의 관계성을 파악할 수 있다. GNN은 일반적인 1차원 또는 2차원 데이터와는 다른 그래프 형태의 데이터를 통해 학습하며, 주로 그래프의 각 노드 간의 연결 정보를 나타낸 인접 행렬(adjacency matrix) 값을 입력으로 사용한다.
- [0051] 도 8은 일 실시예에 있어서, 준지도 기반의 이상 데이터 검출 모델의 예이다.
- [0052] 대기질 평가 시스템은 준지도 기반의 이상 데이터 검출 모델에 GAT(Graph Attention network)을 구성할 수 있다. GAT는 GNN의 기존 노드 임베딩 과정에 셀프-어텐션(self-attention) 알고리즘을 적용한 네트워크로써, 첫 단계에선 각 노드들이 모두 연결된 완전연결 그래프 상태에서 시작하며, 각 계층(layer)마다 인접한 각 노드들 간의 어텐션 스코어(attention score)를 산출한다. 이 과정을 통해 관계성이 높은 노드(항목) 간에 더 많은 가중치가 걸리고, 결과적으로 학습 과정에서 관계성이 높은 항목들의 데이터를 더 많이 참고하게 된다.
- [0053] 대기질 평가 시스템은 대기질 데이터에 포함된 복수 개의 성분에 대한 관계성을 학습하는 제1 GAT 네트워크와 과거 데이터와의 관계성을 학습하는 제2 GAT 네트워크 구분하여 별도로 학습시키고, 제1 GAT와 제2 GAT를 통해 추출된 각 특징 정보를 이용하여 대기질 데이터의 이후 값을 예측하는 예측 모델과 대기질 데이터의 원본 값을 복원하는 복원 모델을 동시에 학습시킬 수 있다. 이를 통해 모델의 견고성을 확보할 수 있다.
- [0054] 도 9는 일 실시예에 있어서, 준지도 기반의 이상 데이터 검출 모델의 입력 데이터를 설명하기 위한 도면이다.
- [0055] 도 9의 왼쪽 도면은 기존에 장기적 주기의 특성을 학습하기 위해 패치 단위의 입력 데이터의 구성을 사용한 것을 나타낸 것이다. 이에 반해, 실시예에서는 실시간성을 반영하기 위해 동작 시간을 최소화하고, 단기적 특성 학습에 더 집중하기 위해 입력 데이터에 대해 윈도우(window)를 1단위(1시간씩) 움직이는 보다 일반적인 형태를 적용할 수 있다.
- [0056] 이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로

세서, 컨트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 컨트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0057] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치에 구체화(embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

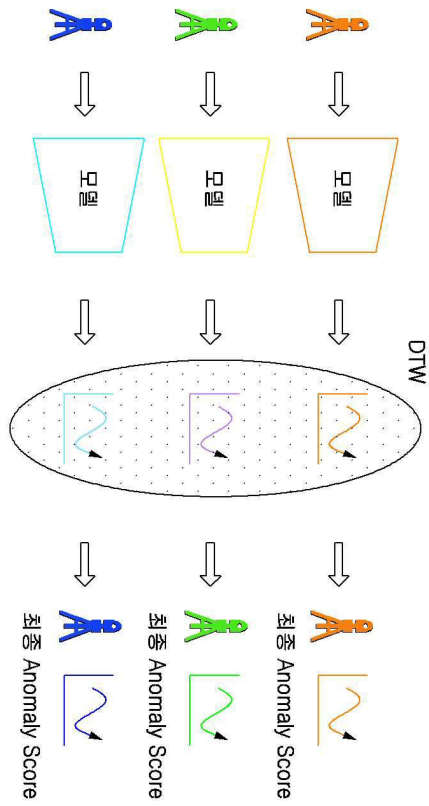
[0058] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

[0059] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

[0060] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

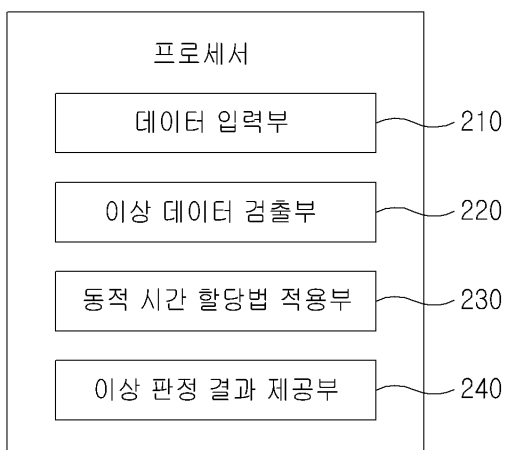
도면

도면1

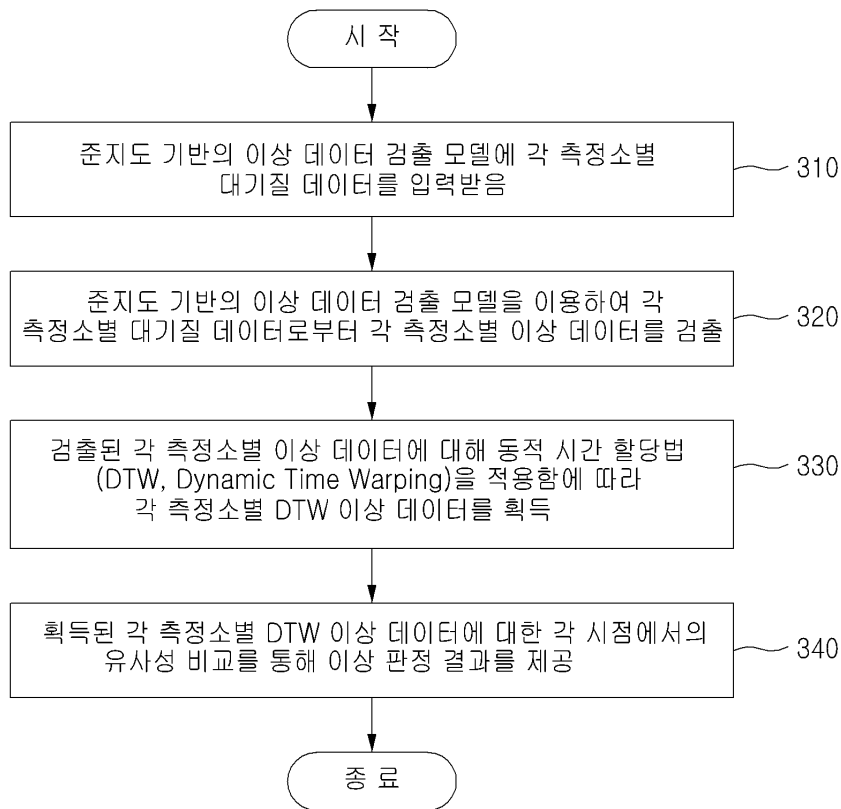


도면2

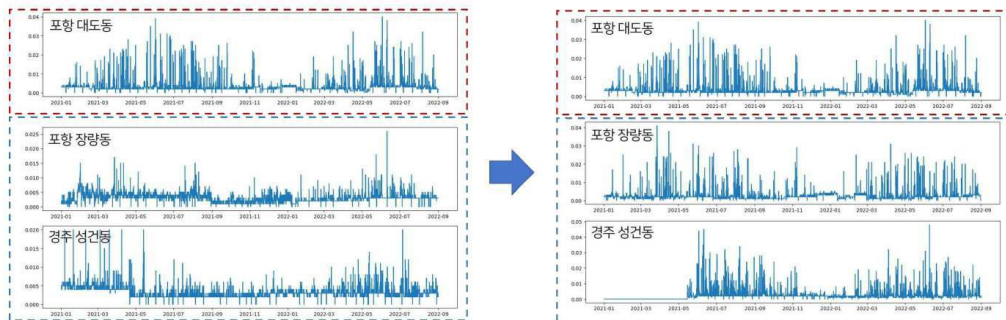
100



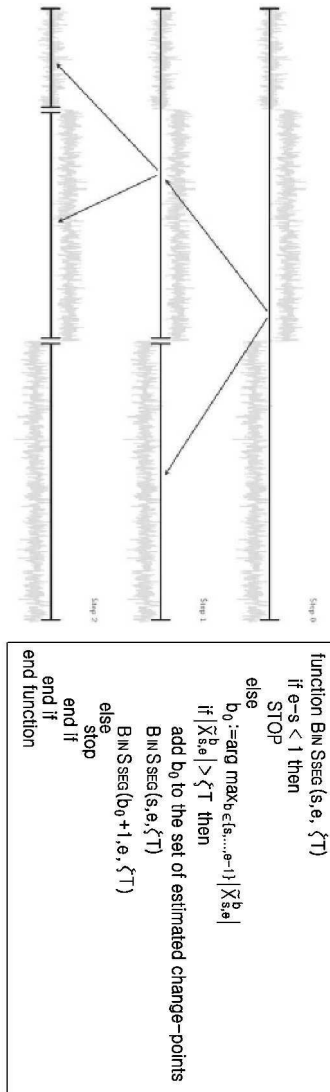
도면3



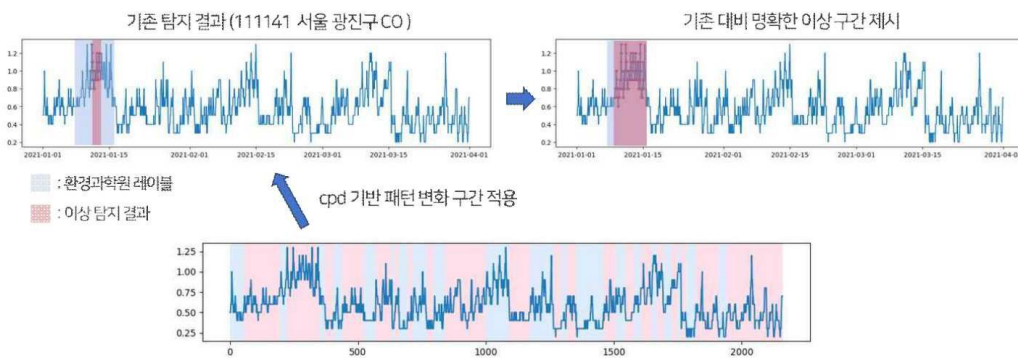
도면4



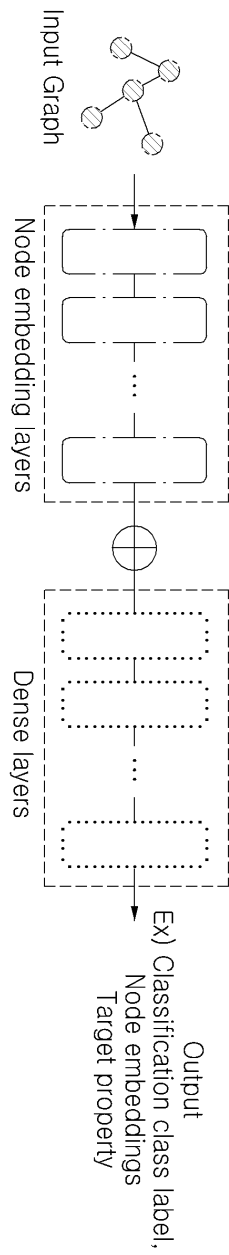
도면5



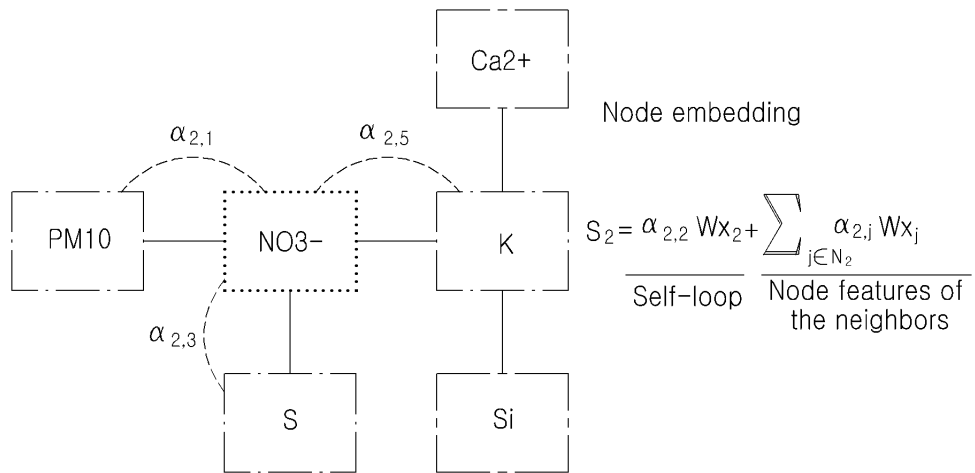
도면6



도면7



도면8



도면9

