



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0160975
(43) 공개일자 2022년12월06일

(51) 국제특허분류(Int. Cl.)
G06F 11/30 (2006.01) G01N 15/00 (2017.01)
G06F 11/32 (2006.01)
(52) CPC특허분류
G06F 11/3072 (2013.01)
G01N 33/0004 (2013.01)
(21) 출원번호 10-2021-0069289
(22) 출원일자 2021년05월28일
심사청구일자 2021년05월28일

(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)
이선우
서울특별시 강서구 우장산로 82, 101동 102호(화곡동, 우장산길성그랑프리텔)
(뒷면에 계속)
(74) 대리인
양성보

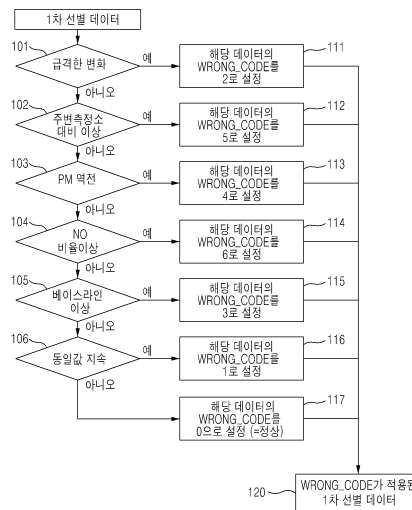
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 규칙 기반 알고리즘을 이용한 대기질 데이터의 이상 판정 방법 및 시스템

(57) 요약

규칙 기반 알고리즘을 이용한 대기질 데이터의 이상 판정 방법 및 시스템이 개시된다. 일 실시예에 따른 대기질 데이터의 이상 판정 방법은, 복수 개의 측정소로부터 측정된 대기질 데이터를 입력받는 단계; 및 각 측정소 별로 미리 설정된 판정 절차에 따라 상기 입력받은 대기질 데이터에 대한 이상을 판정하는 단계를 포함할 수 있다.

대표도 - 도1



(52) CPC특허분류

G06F 11/3089 (2013.01)
G06F 11/322 (2013.01)
G06F 11/324 (2013.01)
G06N 20/20 (2021.08)
G06N 5/025 (2019.01)
G01N 2015/0096 (2013.01)

(72) 발명자

최정무

전라남도 광양시 금호로 73, 11동 401호(금호동, 사랑아파트)

양호준

인천광역시 미추홀구 인주대로366번길 23, 11동 102호(주안동, 남광로얄아파트)

이문형

인천광역시 부평구 경원대로 1269, 104동 104호(산곡동, 현대아파트)

임용재

인천광역시 부평구 안남로 272, 103동 503호(청천동, 부평1차 금호타운)

박지훈

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

정동희

인천광역시 서구 서곶로 818, 112동 906호(당하동, 당하탐스빌아파트)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711125779
과제번호	2017-0-01642-005
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	대학 IT연구센터 육성·지원
연구과제명	[Ezbaro][정부] 인공지능을 활용한 콘텐츠 창작 기술(5차년도)
기 여 율	1/2
과제수행기관명	인하대학교 산학협력단
연구기간	2021.01.01 ~ 2021.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1000637901
과제번호	63790-1
부처명	환경부
과제관리(전문)기관명	국립환경과학원
연구사업명	학술(기술)연구용역사업
연구과제명	인공지능을 활용한 대기오염측정망 자료확정 알고리즘 개발
기 여 율	1/2
과제수행기관명	국립환경과학원
연구기간	2020.09.23 ~ 2021.03.22

명세서

청구범위

청구항 1

대기질 데이터의 이상 판정 방법에 있어서,
복수 개의 측정소로부터 측정된 대기질 데이터를 입력받는 단계; 및
각 측정소 별로 미리 설정된 판정 절차에 따라 상기 입력받은 대기질 데이터에 대한 이상을 판정하는 단계를 포함하는 대기질 데이터의 이상 판정 방법.

청구항 2

제1항에 있어서,
상기 판정하는 단계는,
상기 각 측정소 별로 미리 설정된 판정 절차에 따라 급격한 변화, 주변 측정소 대비 이상, PM역전, NO비율이상, 베이스라인 이상, 동일값 지속 순으로 상기 대기질 데이터에 대한 이상 판정을 수행하는 단계를 포함하는 대기질 데이터의 이상 판정 방법.

청구항 3

제2항에 있어서,
상기 판정하는 단계는,
상기 수행된 이상 판정을 통해 상기 대기질 데이터에 상기 판단된 이상 판정에 대응되는 이상 코드를 설정하고, 상기 설정된 이상 코드가 적용된 선별 데이터를 획득하는 단계를 포함하는 대기질 데이터의 이상 판정 방법.

청구항 4

제1항에 있어서,
상기 판정하는 단계는,
현 시점의 데이터의 모든 대기질 성분값에 대해 1)상기 현 시점의 데이터를 기준으로 이전의 직전 데이터와의 차이값의 절대값이 임계값을 넘을 경우, 2)상기 현 시점의 데이터를 기준으로 이후의 직후 데이터와의 차이값의 절대값이 임계값을 넘을 경우, 3)상기 현 시점의 데이터와 상기 직전 데이터와의 차이값과 상기 직후 데이터와 상기 현 시점의 데이터와의 차이값의 곱이 음수인 경우를 포함하는 복수 개의 조건을 모두 충족하였을 때, 대기질 데이터에 급격한 변화가 발생하였다고 판단하는 단계를 포함하는 대기질 데이터의 이상 판정 방법.

청구항 5

제1항에 있어서,
상기 판정하는 단계는,
현 시점의 데이터에 대해 1)PM 10이 35이하일 때, PM 2.5가 PM 10보다 10이상 큰 경우, 2)PM 10이 35초과일 때, PM 2.5가 PM 10보다 5이상 큰 경우를 포함하는 복수 개의 조건을 모두 충족하였을 때, 대기질 데이터에 PM 역전이 발생하였다고 판단하는 단계를 포함하는 대기질 데이터의 이상 판정 방법.

청구항 6

제1항에 있어서,

상기 판정하는 단계는,

현 시점의 데이터에 포함된 대기질 성분값에 대해 복수 개의 주변 측정소 데이터의 평균값 간의 차이의 절대값이 사용자로부터 지정된 임계값 이상일 때 대기질 데이터를 주변 측정소 대비 이상값으로 분류하는 단계

를 포함하는 대기질 데이터의 이상 판정 방법.

청구항 7

제1항에 있어서,

상기 판정하는 단계는,

현 시점의 데이터에 대해 NO₂, NO, NOX의 값이 (NO₂+NO>NOX+0.005)를 만족하는 경우, 대기질 데이터를 NO 비율 이상으로 분류하는 단계

를 포함하는 대기질 데이터의 이상 판정 방법.

청구항 8

제1항에 있어서,

상기 판정하는 단계는,

지도 학습 모델과 비지도 학습 모델이 앙상블된 학습 모델을 이용하여 상기 대기질 데이터에 대한 베이스라인 이상 여부를 판단하는 단계

를 포함하는 대기질 데이터의 이상 판정 방법.

청구항 9

제1항에 있어서,

상기 판정하는 단계는,

상기 대기질 데이터에 포함된 동일한 값이 사용자로부터 지정된 임계값 이상으로 반복될 경우, 대기질 데이터를 동일값 지속 데이터로 판단하는 단계

를 포함하는 대기질 데이터의 이상 판정 방법.

청구항 10

대기질 데이터의 이상 판정 시스템에 있어서,

복수 개의 측정소로부터 측정된 대기질 데이터를 입력받는 데이터 입력부; 및

각 측정소 별로 미리 설정된 판정 절차에 따라 상기 입력받은 대기질 데이터에 대한 이상을 판정하는 이상 판정부

를 포함하는 이상 판정 시스템.

발명의 설명

기술 분야

아래의 설명은 규칙 기반 알고리즘을 이용하여 대기질 데이터의 이상을 판정하는 기술에 관한 것이다.

[0001]

배경 기술

- [0003] 최근 지구 온난화 및 이상기후에 대한 관심이 높아지면서, 전국의 대기오염측정망 측정소에 대한 숫자가 늘어나고 있는 추세이다. 대기오염측정망 측정소로부터 관측된 데이터 셋에는 기록 장치의 결함이나 자연재해 등의 이유로 결측치나 이상치가 포함될 수가 있다.
- [0004] 이러한 이상치나 결측값이 많아질 경우, 정보량이 정보의 질을 담보하지 못한다는 점에서, 정보의 질의 문제 또한 정보량의 문제만큼 데이터 분석에 있어서 중요한 역할을 한다. 그 이유는 첫째, 정보량이 많더라도 개개 정보의 내용을 신뢰할 수 없으면 그 정보를 이용한 통계적 추론의 신뢰성 또한 보장할 수 없기 때문이다. 둘째, 대부분 통계자료가 가지는 근원적 특성인 표본으로서의 조건을 충족하지 못하기 때문이다. 어떤 현상을 설명하기 위한 통계적 방법은 모집단을 대표할 수 있는 표본(sample)을 구성하고 그로부터 통계적 추론을 수행하는 것을 요체로 한다. 그러나 표본으로부터 조사한 자료가 불완전하여 자료의 질을 담보하지 못할 경우 자료의 대표성을 담보하기 어렵다.
- [0005] 시간과 공간의 자료를 포함하는 시공간 자료의 결측치는 시공간 분석의 주요한 장애물이다. 측정소 데이터의 고신뢰성을 위하여 측정망 데이터에서 이상치 탐지와 결측치 처리는 주요 관심사 중 하나이다. 결측치가 자료를 공간적, 시간적으로 단절시켜 공간적 현상의 변화를 분석하는데 어려움을 주기 때문이다. 이상 및 결측치 발생에 의한 불완전한 자료는 자료 분석 시 모델링에서의 편향된 모수 추정 등의 요소에서 문제가 발생 문제가 발생하여 잘못된 결과를 초래할 수 있기 때문에, 적절한 결측치 처리는 분석을 할 때 중요한 요소라고 할 수 있다. 또한 연구 대상 공간에 속한 일부 지역의 속성값이 결측되어 있다면, 그 자료를 이용한 대상공간 전체에 대한 통계적 추론은 완전한 자료를 이용할 때에 비해서 부정확할 가능성이 높다. 또한 결측이 발생한 어떤 시점의 공간표본이 다른 시점의 공간표본들로부터 얻은 값들과 다른 특이성을 가진다면, 공간분석의 결과가 왜곡될 수 있기 때문에 높은 시공간 자료는 신뢰성을 요구한다.

발명의 내용

해결하려는 과제

- [0007] 규칙 기반 알고리즘을 이용하여 대기질 데이터의 이상 판정을 수행하는 방법 및 시스템을 제공할 수 있다.
- [0008] 인공 지능을 이용하여 대기질 데이터에 대한 베이스라인 이상을 판정하고, 규칙 기반 알고리즘을 이용하여 대기질 데이터에 대한 급격한 변화, 주변 측정소 대비 이상, PM 역전, NO 비율이상, 동일값 지속을 판단하는 방법 및 시스템을 제공할 수 있다.
- [0009] 각 측정소 별로 미리 설정된 판정 절차에 따라 대기질 데이터의 이상 판정을 수행하는 방법 및 시스템을 제공할 수 있다.

과제의 해결 수단

- [0011] 대기질 데이터의 이상 판정 방법은, 복수 개의 측정소로부터 측정된 대기질 데이터를 입력받는 단계; 및 각 측정소 별로 미리 설정된 판정 절차에 따라 상기 입력받은 대기질 데이터에 대한 이상을 판정하는 단계를 포함할 수 있다.
- [0012] 상기 판정하는 단계는, 상기 각 측정소 별로 미리 설정된 판정 절차에 따라 급격한 변화, 주변 측정소 대비 이상, PM역전, NO비율이상, 베이스라인 이상, 동일값 지속 순으로 상기 대기질 데이터에 대한 이상 판정을 수행하는 단계를 포함할 수 있다.
- [0013] 상기 판정하는 단계는, 상기 수행된 이상 판정을 통해 상기 대기질 데이터에 상기 판단된 이상 판정에 대응되는 이상 코드를 설정하고, 상기 설정된 이상 코드가 적용된 선별 데이터를 획득하는 단계를 포함할 수 있다.
- [0014] 상기 판정하는 단계는, 현 시점의 데이터의 모든 대기질 성분값에 대해 1)상기 현 시점의 데이터를 기준으로 이전의 직전 데이터와의 차이값의 절대값이 임계값을 넘을 경우, 2)상기 현 시점의 데이터를 기준으로 이후의 직후 데이터와의 차이값의 절대값이 임계값을 넘을 경우, 3)상기 현 시점의 데이터와 상기 직전 데이터와의 차이값과 상기 직후 데이터와 상기 현 시점의 데이터와의 차이값의 곱이 음수인 경우를 포함하는 복수 개의 조건을

모두 충족하였을 때, 대기질 데이터에 급격한 변화가 발생하였다고 판단하는 단계를 포함할 수 있다.

- [0015] 상기 판정하는 단계는, 현 시점의 데이터에 대해 1)PM 10이 35이하일 때, PM 2.5가 PM 10보다 10이상 큰 경우, 2)PM 10이 35초과일 때, PM 2.5가 PM 10보다 5이상 큰 경우를 포함하는 복수 개의 조건을 모두 충족하였을 때, 대기질 데이터에 PM 역전이 발생하였다고 판단하는 단계를 포함할 수 있다.
- [0016] 상기 판정하는 단계는, 현 시점의 데이터에 포함된 대기질 성분값에 대해 복수 개의 주변 측정소 데이터의 평균 값 간의 차이의 절대값이 사용자로부터 지정된 임계값 이상일 때 대기질 데이터를 주변 측정소 대비 이상값으로 분류하는 단계를 포함할 수 있다.
- [0017] 상기 판정하는 단계는, 현 시점의 데이터에 대해 NO₂, NO, NO_x의 값이 (NO₂+NO>NO_x+0.005)를 만족하는 경우, 대기질 데이터를 NO 비율 이상으로 분류하는 단계를 포함할 수 있다.
- [0018] 상기 판정하는 단계는, 지도 학습 모델과 비지도 학습 모델이 앙상블된 학습 모델을 이용하여 상기 대기질 데이터에 대한 베이스라인 이상 여부를 판단하는 단계를 포함할 수 있다.
- [0019] 상기 판정하는 단계는, 상기 대기질 데이터에 포함된 동일한 값이 사용자로부터 지정된 임계값 이상으로 반복될 경우, 대기질 데이터를 동일값 지속 데이터로 판단하는 단계를 포함할 수 있다.
- [0020] 대기질 데이터의 이상 판정 시스템은, 복수 개의 측정소로부터 측정된 대기질 데이터를 입력받는 데이터 입력부; 및 각 측정소 별로 미리 설정된 판정 절차에 따라 상기 입력받은 대기질 데이터에 대한 이상을 판정하는 이상 판정부를 포함할 수 있다.

발명의 효과

- [0022] 기존 대기오염망 자료확정에 사용되는 규칙 기반 알고리즘을 보완 및 규칙 정립화를 통해 기존 레이블보다 더욱 명확한 레이블링이 가능해진다.
- [0023] 분류 작업이 제대로 되어있지 않은 기존 데이터에 규칙 기반 알고리즘을 적용함으로써 잘못 분류된 데이터를 적절하게 재분류할 수 있다.

도면의 간단한 설명

- [0025] 도 1은 일 실시예에 따른 이상 판정 시스템에서 측정소 별 규칙기반 알고리즘을 적용한 순서도를 설명하기 위한 도면이다.
- 도 2는 일 실시예에 따른 이상 판정 시스템에서 동일값 지속 감지 알고리즘의 동작을 설명하기 위한 흐름도이다.
- 도 3은 일 실시예에 있어서, 동일값 지속 감지 알고리즘을 적용한 결과를 나타낸 그래프이다.
- 도 4는 일 실시예에 따른 이상 판정 시스템에서 급격한 변화 감지 알고리즘의 동작을 설명하기 위한 흐름도이다.
- 도 5는 일 실시예에 있어서, 급격한 변화 감지 알고리즘을 적용한 결과를 나타낸 그래프이다.
- 도 6은 일 실시예에 따른 이상 판정 시스템에서 PM 역전 감지 알고리즘의 동작을 설명하기 위한 흐름도이다.
- 도 7은 일 실시예에 있어서, PM 역전 감지 알고리즘을 적용한 결과를 나타낸 그래프이다.
- 도 8은 일 실시예에 따른 이상 판정 시스템에서 주변 측정소 대비 이상 감지 알고리즘의 동작을 설명하기 위한 흐름도이다.
- 도 9는 일 실시예에 있어서, 주변 측정소 대비 이상 감지 알고리즘을 적용한 결과를 나타낸 그래프이다.
- 도 10은 일 실시예에 따른 이상 판정 시스템에서 NO 비율 이상 감지 알고리즘의 동작을 설명하기 위한 흐름도이다.
- 도 11은 일 실시예에 있어서, NO 비율 이상 감지 알고리즘을 적용한 결과를 나타낸 그래프이다.

도 12는 일 실시예에 따른 이상 판정 시스템에서 베이스라인 이상을 판정하는 동작을 설명하기 위한 도면이다.

도 13은 일 실시예에 따른 이상 판정 시스템의 구성을 설명하기 위한 블록도이다.

도 14는 일 실시예에 따른 이상 판정 시스템에서 대기질 데이터의 이상 판정 방법을 설명하기 위한 흐름도이다.

발명을 실시하기 위한 구체적인 내용

- [0026] 이하, 실시예를 첨부한 도면을 참조하여 상세히 설명한다.
- [0028] 도 1은 일 실시예에 따른 이상 판정 시스템에서 측정소 별 규칙기반 알고리즘을 적용한 순서도를 설명하기 위한 도면이다.
- [0029] 이상 판정 시스템은 각 측정소마다 미리 정해진 임계치를 기반으로 대기질 데이터의 이상 여부를 판정할 수 있다. 규칙 기반으로 판별하는 이상 사례들은 동일값 지속, 급격한 변화, PM 역전, 주변 측정소 대비 이상, NO 비율 이상이 있으며, 각각 이상의 종류를 나타는 이상 코드(wrong code)를 가질 수 있다. 예를 들면, 이상 코드 1은 동일값 지속, 이상 코드 2는 급격한 변화, 이상 코드 4는 PM10, PM2.5 역전(PM 역전), 이상 코드 5는 주변 측정소 대비 이상, 이상 코드 6은 NOX, NO, NO2 비율 이상으로 정의될 수 있다.
- [0030] 이상 판정 시스템은 복수 개의 측정소로부터 측정된 대기질 데이터를 입력받을 수 있다. 이때, 대기질 데이터는 각 측정소에서 측정된 데이터를 의미할 수 있고, 자동 또는 수동으로 대기질 데이터로부터 이상치 또는 결측값이 선별된 데이터를 의미할 수 있다.
- [0031] 이상 판정 시스템은 사용자로부터 측정소 별로 미리 설정된 판정 절차에 따라 급격한 변화, 주변 측정소 대비 이상, PM 역전, NO 비율 이상, 베이스라인 이상, 동일값 지속 순으로 대기질 데이터에 대한 이상 판정을 수행할 수 있다. 이때, 이상 판정 시스템은 인공지능을 이용하여 베이스라인 이상 판정을 수행하고, 나머지 이상에 대해서는 규칙 기반 알고리즘을 이용하여 이상 판정을 수행할 수 있다.
- [0032] 상세하게는, 이상 판정 시스템은 1차 선별 데이터에 대한 급격한 변화 여부를 판단할 수 있다(101). 이상 판정 시스템은 1차 선별 데이터에 대한 급격한 변화가 감지될 경우, 해당 데이터의 이상 코드를 2로 설정할 수 있다(111). 이상 판정 시스템은 1차 선별 데이터에 대한 급격한 변화가 감지되지 않는 경우, 주변 측정소 대비 이상 여부를 판단할 수 있다(102).
- [0033] 이상 판정 시스템은 1차 선별 데이터에 대한 주변 측정소 이상이 감지될 경우, 해당 데이터의 이상 코드를 5로 설정할 수 있다(112). 이상 판정 시스템은 1차 선별 데이터에 대한 주변 측정소 대비 이상이 감지되지 않는 경우, PM 역전 여부를 판단할 수 있다(103).
- [0034] 이상 판정 시스템은 1차 선별 데이터에 대한 PM 역전이 감지될 경우, 해당 데이터의 이상 코드를 4로 설정할 수 있다(113). 이상 판정 시스템은 1차 선별 데이터에 대한 PM 역전이 감지되지 않을 경우, NO 비율 이상 여부를 판단할 수 있다(104).
- [0035] 이상 판정 시스템은 1차 데이터에 대한 NO 비율 이상이 감지될 경우, 해당 데이터의 이상 코드를 6으로 설정할 수 있다(114). 이상 판정 시스템은 1차 선별 데이터에 대한 NO 비율 이상이 감지되지 않을 경우, 베이스라인 이상 여부를 판단할 수 있다(105).
- [0036] 이상 판정 시스템은 1차 선별 데이터에 대한 베이스라인 이상이 감지될 경우, 해당 데이터의 이상 코드를 3으로 설정할 수 있다(115). 이상 판정 시스템은 1차 선별 데이터에 대한 베이스라인 이상이 감지되지 않을 경우, 동일값 지속 여부를 판단할 수 있다(106).
- [0037] 이상 판정 시스템은 1차 선별 데이터에 대한 동일값 지속이 감지될 경우, 해당 데이터의 이상 코드를 1로 설정할 수 있다(116). 이상 판정 시스템은 1차 선별 데이터에 대한 동일값 지속이 감지되지 않을 경우, 해당 데이터의 이상 코드를 0으로 설정(=정상)할 수 있다(117).
- [0038] 이상 판정 시스템은 이상 코드가 적용된 1차 선별 데이터를 획득할 수 있다(120).
- [0039] 도 2는 일 실시예에 따른 이상 판정 시스템에서 동일값 지속 감지 알고리즘의 동작을 설명하기 위한 흐름도이다.

- [0040] 이상 판정 시스템은 대기질 데이터에 포함된 동일한 값이 사용자로부터 지정된 임계값 이상으로 반복될 경우, 동일값 지속으로 판단할 수 있다. 상세하게는, 이상 판정 시스템은 전체 데이터에 대하여 동일값 반복 횟수를 0으로 초기화할 수 있다(201). 이상 판정 시스템은 현 시점의 데이터가 마지막인지 여부를 판단할 수 있다(202). 이상 판정 시스템은 현 시점의 데이터가 마지막으로 판단됨에 따라 동일값 반복 횟수가 임계값 이상인지 여부를 판단할 수 있다(206).
- [0041] 또한, 이상 판정 시스템은 현 시점의 데이터가 마지막이 아닌 것으로 판단됨에 따라 현 시점의 데이터가 현 시점 데이터를 기준으로 직전 시점의 데이터와 동일한 지 여부를 판단할 수 있다(203). 이상 판정 시스템은 현 시점의 데이터가 직전 시점의 데이터와 동일하지 않을 것으로 판단됨에 따라 동일값 반복 횟수를 증가(204)시키고 다음 데이터로 시점을 이동(205)시킬 수 있다. 이상 판정 시스템은 현 시점의 데이터가 직전 시점의 데이터와 동일한 것으로 판단됨에 따라 동일값 반복 횟수가 임계값 이상인지 여부를 판단할 수 있다(206).
- [0042] 이상 판정 시스템은 동일값 반복 횟수가 임계값 이상으로 판단됨에 따라 동일값 반복 횟수 전부터의 시점부터 현 시점까지 데이터들을 동일값 지속 데이터로 판단할 수 있다(207). 이상 판정 시스템은 동일값 반복 횟수가 임계값 이상이 아닌 것으로 판단됨에 따라 동일값 반복 횟수 전부터의 시점부터 현 시점까지 데이터들을 정상 데이터로 판단할 수 있다(208).
- [0043] 이상 판정 시스템은 현 시점의 데이터가 마지막 데이터인지 판단할 수 있다(209). 이상 판정 시스템은 현 시점의 데이터가 마지막 데이터로 판단됨에 따라 다음 데이터로 시점을 이동(210)시킬 수 있고, 현 시점의 데이터가 마지막 데이터가 아닌 것으로 판단됨에 따라 동일값 지속 알고리즘이 적용된 전체 데이터를 획득할 수 있다.
- [0044] 도 3을 참고하면, 동일값 지속 감지 알고리즘을 적용한 결과를 나타낸 그래프이다. 원본 데이터와 동일값 지속 감지 알고리즘을 적용한 결과 데이터를 나타낸 것이다. 기존 데이터에서 동일값 지속 이상 데이터임에도 불구하고 정상으로 분류된 데이터가 동일값 지속 데이터로 분류됨을 확인할 수 있다.
- [0045] 도 4는 일 실시예에 따른 이상 판정 시스템에서 급격한 변화 감지 알고리즘의 동작을 설명하기 위한 흐름도이다.
- [0046] 이상 판정 시스템은 하나의 시점의 데이터의 모든 대기질 성분값(SO₂, CO, O₃, PM₁₀, PM_{2.5}, NO₂, NO, NO_x)에 대해 복수 개의 조건을 모두 충족하였을 때 대기질 데이터에 급격한 변화가 발생하였다고 판단할 수 있다. 이상 판정 시스템은 현 시점의 데이터의 모든 대기질 성분값에 시점의 데이터를 기준으로 이전의 직전 데이터와의 차이값의 절대값이 임계값을 넘을 경우(제1 조건), 현 시점의 데이터를 기준으로 이후의 직후 데이터와의 차이값의 절대값이 임계값을 넘을 경우(제2 조건), 현 시점의 데이터와 직전 데이터와의 차이값과 직후 데이터와 현 시점의 데이터와의 차이값의 곱이 음수인 경우(제3 조건)를 포함하는 복수 개의 조건을 모두 충족하였을 때, 대기질 데이터에 급격한 변화가 발생하였다고 판단할 수 있다. 이때, 제3 조건은 직전과 현재, 현재와 직후의 데이터 변화 추이가 계속 증가하거나, 계속 감소하는 것이 아닌 증가 이후 감소, 혹은 감소 이후 증가인 경우를 의미한다.
- [0047] 사용자로부터 급격한 변화 감지를 위한 임계값이 설정될 수 있다. 한 측정소의 모든 데이터의 각 대기질 성분값에 대해 현재 데이터와 직전 데이터 간의 차이의 절대값의 평균(Mean; M), 분산(Standard Deviation; SD)을 구한 후 분산으로부터 표준편차가 도출될 수 있다. 예를 들면, 데이터의 분포가 정규분포를 따른다고 가정했을 시 데이터의 평균에 표준편차의 3배를 곱한 값을 감산한 값과 평균에 표준편차의 3배를 곱한 값을 가산한 값 사이의 데이터가 전체 데이터의 99%를 따른다는 것에 착안하여 임계값이 $(SD * 3 + M)$ 으로 설정될 수 있다. 이때, 분산에 곱해지는 수는 사용자에게 의해 지정될 수 있다.
- [0048] 상세하게는, 이상 판정 시스템은 전체 데이터에 대해 직전 데이터 간 차이의 절대값의 평균과 표준편차를 계산할 수 있다(401). 이상 판정 시스템은 계산된 평균과 표준편차를 이용하여 급격한 변화 감지를 위한 임계값을 설정할 수 있다(402). 이상 판정 시스템은 현 시점의 데이터가 조건 3개를 만족하는지 판단할 수 있다(조건 1: (현 시점의 데이터 - 직전 시점의 데이터)의 절대값이 임계값 이상, 조건 2: (현 시점의 데이터 - 직후 시점의 데이터)의 절대값이 임계값 이상, 조건 3: (현 시점의 데이터 - 직전 시점의 데이터)와 (직후 시점의 데이터 - 현 시점의 데이터)의 곱이 음수인 경우)(403).
- [0049] 이상 판정 시스템은 현 시점의 데이터가 조건 3개를 모두 만족하는 것으로 판단됨에 따라 현 시점의 데이터를 급격한 변화 데이터로 판단(404)할 수 있고, 현 시점의 데이터가 조건 3개 중 어느 하나 이상을 만족하지 못하는 것으로 판단됨에 따라 현 시점의 데이터를 정상 데이터로 판단(405)할 수 있다. 이상 판정 시스템은 현 시점의 데이터가 마지막 데이터인지 판단할 수 있다(406). 이상 판정 시스템은 현 시점의 데이터가 마지막 데이

터로 판단됨에 따라 급격한 변화 알고리즘이 적용된 전체 데이터를 획득할 수 있고, 현 시점의 데이터가 마지막 데이터로 판단되지 않음에 따라 다음 데이터로 시점을 이동(407)시킬 수 있다.

- [0050] 도 5를 참고하면, 급격한 변화 감지 알고리즘을 적용한 결과를 나타낸 그래프이다. 원본 데이터와 급격한 변화 감지 알고리즘을 적용한 결과 데이터를 나타낸 것이다. 급격한 변화가 발생했음에도 정상 데이터로 분류된 원본 데이터와는 달리, 급격한 변화 알고리즘을 적용한 데이터의 경우 정상적으로 급격한 변화 데이터가 분류된 것을 확인할 수 있다.
- [0051] 도 6은 일 실시예에 따른 이상 판정 시스템에서 PM 역전 감지 알고리즘의 동작을 설명하기 위한 흐름도이다.
- [0052] 이상 판정 시스템은 하나의 시점의 데이터에 대해 복수 개의 조건을 만족하였을 때 PM 역전이 발생한 것으로 판단할 수 있다. 이상 판정 시스템은 현 시점의 데이터에 대하여 PM 10이 35이하일 때, PM 2.5가 PM 10보다 10이상 큰 경우(조건 1), PM 10이 35초과일 때, PM 2.5가 PM 10보다 5이상 큰 경우(조건 2)를 포함하는 복수 개의 조건을 모두 충족하였을 때, 대기질 데이터에 PM 역전이 발생하였다고 판단할 수 있다.
- [0053] 상세하게는, 이상 판정 시스템은 전체 데이터의 PM 10의 값이 35이하인지 판단할 수 있다(601). 이상 판정 시스템은 전체 데이터의 PM 10의 값이 35이하일 경우, PM 2.5의 값이 PM 10보다 10이상 큰 지 판단할 수 있다(602). 이상 판정 시스템은 전체 데이터의 PM 2.5의 값이 PM 10보다 10이상 클 경우, 현 시점의 데이터를 정상 데이터로 판단할 수 있다(607). 이상 판정 시스템은 전체 데이터의 PM 2.5의 값이 PM 10보다 10보다 작을 경우, 현 시점의 데이터를 PM 역전 데이터로 판단할 수 있다(604).
- [0054] 이상 판정 시스템은 전체 데이터의 PM 10의 값이 35이하가 아니라고 판단됨에 따라 PM2.5의 값이 PM10보다 5 이상 큰 지 판단할 수 있다(603). 이상 판정 시스템은 PM2.5의 값이 PM10보다 5보다 작은 것으로 판단됨에 따라 현 시점의 데이터를 정상 데이터로 판단할 수 있다(607). 이상 판정 시스템은 PM2.5의 값이 PM10보다 5이상 큰 것으로 판단됨에 따라 현 시점의 데이터를 PM 역전 데이터로 판단할 수 있다(604).
- [0055] 이상 판정 시스템은 현 시점의 데이터를 PM 역전 데이터로 판단함에 따라 현 시점의 데이터가 마지막 데이터인지 판단할 수 있다(605). 이상 판정 시스템은 현 시점의 데이터가 마지막 데이터로 판단됨에 따라 PM 역전 감지 알고리즘이 적용된 전체 데이터를 획득할 수 있다. 이상 판정 시스템은 현 시점의 데이터가 마지막 데이터가 아닌 것으로 판단됨에 따라 다음 데이터로 이점을 이동시킬 수 있다(606).
- [0056] 도 7을 참고하면, PM 역전 감지 알고리즘을 적용한 결과를 나타낸 그래프이다. 원본 데이터와 PM 역전 감지 알고리즘을 적용한 결과 데이터를 나타낸 것이다. 기존 데이터 중 PM 역전이 발생하지 않았음에도 불구하고 PM 역전으로 분류된 데이터들을 PM 역전 감지 알고리즘을 사용하여 정상 데이터로 재분류할 수 있음을 확인할 수 있다.
- [0057] 도 8은 일 실시예에 따른 이상 판정 시스템에서 주변 측정소 대비 이상 감지 알고리즘의 동작을 설명하기 위한 흐름도이다.
- [0058] 이상 판정 시스템은 하나의 시점의 데이터의 모든 대기질 성분값(SO2, CO, O3, PM10, PM2.5, NO2, NO, NOX)에 대해 특정 조건을 만족할 때 주변 측정소 대비 이상값으로 분류할 수 있다. 이상 판정 시스템은 현 시점의 데이터에 포함된 대기질 성분값이 복수 개의 주변 측정소 데이터의 평균값 간의 차이의 절대값이 사용자로부터 지정된 임계값 이상일 때 대기질 데이터를 주변 측정소 대비 이상값으로 분류할 수 있다. 이상 판정 시스템은 데이터의 값과 주변 3개 측정소 데이터의 평균값 간의 차이의 절대값이 사용자로부터 지정된 임계값 이상일 때, 주변 측정소 대비 이상값으로 분류할 수 있다.
- [0059] 이상 판정 시스템은 현 시점에 대한 주변 3개의 측정소의 데이터를 수집할 수 있다(801). 이상 판정 시스템은 수집한 주변 3개의 측정소의 데이터의 평균을 계산할 수 있다(802). 이상 판정 시스템은 현 시점의 데이터가 현 시점의 데이터와 주변 3개 측정소 데이터의 평균값과의 차이의 절대값이 사용자로부터 지정된 임계값 이상인 조건을 만족하는지 판단할 수 있다(803). 이상 판정 시스템은 현 시점의 데이터가 상기 조건을 만족할 경우, 현 시점의 데이터를 주변 측정소 대비 이상 데이터로 판단할 수 있다(804). 이상 판정 시스템은 현 시점의 데이터가 상기 조건을 만족하지 못할 경우, 현 시점의 데이터를 정상 데이터로 판단할 수 있다(805).
- [0060] 이상 판정 시스템은 현 시점의 데이터를 주변 측정소 대비 이상 데이터로 판단함에 따라 현 시점의 데이터가 마지막 데이터인지 판단할 수 있다(806). 이상 판정 시스템은 현 시점의 데이터가 마지막 데이터로 판단됨에 따라 주변 측정소 대비 이상 감지 알고리즘이 적용된 전체 데이터를 획득할 수 있다. 이상 판정 시스템은 현 시점의 데이터가 마지막 데이터가 아닌 것으로 판단됨에 따라 다음 데이터로 시점을 이동시킬 수 있다(807).

- [0061] 도 9를 참고하면, 주변 측정소 대비 이상 감지 알고리즘을 적용한 결과를 나타낸 그래프이다. 원본 데이터와 주변 측정소 대비 이상 감지 알고리즘을 적용한 결과 데이터를 나타낸 것이다. 주변 측정소 대비 이상을 세부적으로 감지하지 못한 원본 데이터와는 달리, 주변 측정소의 평균값을 이용하여(오렌지색) 주변 측정소 대비 이상을 세부적으로 분류할 수 있음을 확인할 수 있다.
- [0062] 도 10은 일 실시예에 따른 이상 판정 시스템에서 NO 비율 이상 감지 알고리즘의 동작을 설명하기 위한 흐름도이다.
- [0063] 이상 판정 시스템은 하나의 시점의 데이터에 대해 특정 조건을 만족했을 경우, NO₂, NO_x, NO비율 이상으로 분류할 수 있다. 이상 판정 시스템은 현 시점의 데이터에 대해 NO₂, NO, NO_x의 값이 (NO₂+NO>NO_x+0.005)를 만족하는 경우, 대기질 데이터를 NO₂, NO_x, NO 비율 이상으로 분류할 수 있다.
- [0064] 이상 판정 시스템은 현 시점의 데이터의 NO₂+NO가 NO_x+0.005를 초과하는지 판단할 수 있다(1001). 이상 판정 시스템은 현 시점의 데이터의 NO₂+NO가 NO_x+0.005를 초과할 경우, 현 시점의 데이터를 NO 비율 이상 데이터로 판단할 수 있다(1002). 이상 판정 시스템은 현 시점의 데이터의 NO₂+NO가 NO_x+0.005를 초과하지 않을 경우, 현 시점의 데이터를 정상 데이터로 판단할 수 있다(1003).
- [0065] 이상 판정 시스템은 현 시점의 데이터가 마지막 데이터인지 판단할 수 있다(1004). 이상 판정 시스템은 현 시점의 데이터가 마지막 데이터로 판단됨에 따라 다음 데이터로 시점을 이동시킬 수 있다(1005). 이상 판정 시스템은 현 시점의 데이터가 마지막 데이터가 아닌 것으로 판단됨에 따라 NO 비율 이상 알고리즘이 적용된 전체 데이터를 획득할 수 있다.
- [0066] 도 11을 참고하면, NO 비율 이상 감지 알고리즘을 적용한 결과를 나타낸 그래프이다. 원본 데이터와 NO 비율 이상 감지 알고리즘을 적용한 결과 데이터를 나타낸 것이다. 기존 데이터에서 NO 비율 이상이 발생하였음에도 불구하고 NO 비율 이상을 비정상적으로 적게 분류한 기존 데이터와는 달리, NO 비율 이상 감지 알고리즘을 적용한 결과 데이터에서는 정상적으로 NO 비율 이상 데이터를 감지한 것을 확인할 수 있다.
- [0067] 도 12는 일 실시예에 따른 이상 판정 시스템에서 베이스라인 이상을 판정하는 동작을 설명하기 위한 도면이다.
- [0068] 이상 판정 시스템은 베이스라인 이상의 경우, 기존 데이터의 레이블링으로 학습된 지도 모델의 결과와 비정상 데이터를 제외한 정상 데이터만으로 학습이 진행된 비지도 모델의 결과를 통합하여 대기질 데이터의 이상 여부를 판단할 수 있다.
- [0069] 이상 판정 시스템은 지도 학습 모델과 비지도 학습 모델이 앙상블된 학습 모델에 대기질 데이터를 입력 데이터로 입력받을 수 있다. 이상 판정 시스템은 지도 학습 모델과 학습 모델이 앙상블된 학습 모델을 이용하여 대기질 데이터에 대한 이상을 판정할 수 있다.
- [0070] 도 13은 일 실시예에 따른 이상 판정 시스템의 구성을 설명하기 위한 블록도이고, 도 14는 일 실시예에 따른 이상 판정 시스템에서 대기질 데이터의 이상 판정 방법을 설명하기 위한 흐름도이다.
- [0071] 이상 판정 시스템(100)의 프로세서는 데이터 입력부(1310) 및 이상 판정부(1320)를 포함할 수 있다. 이러한 프로세서의 구성요소들은 이상 판정 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 14의 대기질 데이터의 이상 판정 방법이 포함하는 단계들(1410 내지 1420)을 수행하도록 이상 판정 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.
- [0072] 프로세서는 대기질 데이터의 이상 판정 방법을 위한 프로그램의 파일에 저장된 프로그램 코드를 메모리에 로딩할 수 있다. 예를 들면, 이상 판정 시스템에서 프로그램이 실행되면, 프로세서는 운영체제의 제어에 따라 프로그램의 파일로부터 프로그램 코드를 메모리에 로딩하도록 이상 판정 시스템을 제어할 수 있다. 이때, 프로세서 및 프로세서가 포함하는 데이터 입력부(1310) 및 이상 판정부(1320) 각각은 메모리에 로딩된 프로그램 코드 중 대응하는 부분의 명령을 실행하여 이후 단계들(1410 내지 1420)을 실행하기 위한 프로세서의 서로 다른 기능적 표현들일 수 있다.
- [0073] 단계(1410)에서 데이터 입력부(1310)는 복수 개의 측정소로부터 측정된 대기질 데이터를 입력받을 수 있다.
- [0074] 단계(1420)에서 이상 판정부(1320)는 각 측정소 별로 미리 설정된 판정 절차에 따라 입력받은 대기질 데이터에 대한 이상을 판정할 수 있다. 이상 판정부(1320)는 각 측정소 별로 미리 설정된 판정 절차에 따라 급격한

변화, 주변 측정소 대비 이상, PM역전, NO비율이상, 베이스라인 이상, 동일값 지속 순으로 상기 대기질 데이터에 대한 이상 판정을 수행할 수 있다. 이상 판정부(1320)는 수행된 이상 판정을 통해 대기질 데이터에 판단된 이상 판정에 대응되는 이상 코드를 설정하고, 설정된 이상 코드가 적용된 선별 데이터를 획득할 수 있다. 이상 판정부(1320)는 현 시점의 데이터의 모든 대기질 성분값에 대해 1)현 시점의 데이터를 기준으로 이전의 직전 데이터와의 차이값의 절대값이 임계값을 넘을 경우, 2)현 시점의 데이터를 기준으로 이후의 직후 데이터와의 차이값의 절대값이 임계값을 넘을 경우, 3)현 시점의 데이터와 직전 데이터와의 차이값과 직후 데이터와 현 시점의 데이터와의 차이값의 곱이 음수인 경우를 포함하는 복수 개의 조건을 모두 충족하였을 때, 대기질 데이터에 급격한 변화가 발생하였다고 판단할 수 있다. 이때, 현 시점의 데이터는 측정소에서 현재 측정된 대기질 데이터를 의미할 수 있다. 이상 판정부(1320)는 현 시점의 데이터에 대해 1)PM 10이 35이하일 때, PM 2.5가 PM 10보다 10이상 큰 경우, 2)PM 10이 35초과일 때, PM 2.5가 PM 10보다 5이상 큰 경우를 포함하는 복수 개의 조건을 모두 충족하였을 때, 대기질 데이터에 PM 역전이 발생하였다고 판단할 수 있다. 이상 판정부(1320)는 현 시점의 데이터에 포함된 대기질 성분값에 대해 복수 개의 주변 측정소 데이터의 평균값 간의 차이의 절대값이 사용자로부터 지정된 임계값 이상일 때 대기질 데이터를 주변 측정소 대비 이상값으로 분류할 수 있다. 이상 판정부(1320)는 현 시점의 데이터에 대해 NO₂, NO, NO_x의 값이 (NO₂+NO>NO_x+0.005)를 만족하는 경우, 대기질 데이터를 NO 비율 이상으로 분류할 수 있다. 이상 판정부(1320)는 지도 학습 모델과 비지도 학습 모델이 앙상블된 학습 모델을 이용하여 대기질 데이터에 대한 베이스라인 이상 여부를 판단할 수 있다. 이상 판정부(1320)는 대기질 데이터에 포함된 동일한 값이 사용자로부터 지정된 임계값 이상으로 반복될 경우, 대기질 데이터를 동일값 지속 데이터로 판단할 수 있다.

[0075] 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, 프로세서, 콘트롤러, 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 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0076] 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.

[0077] 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.

[0078] 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 다

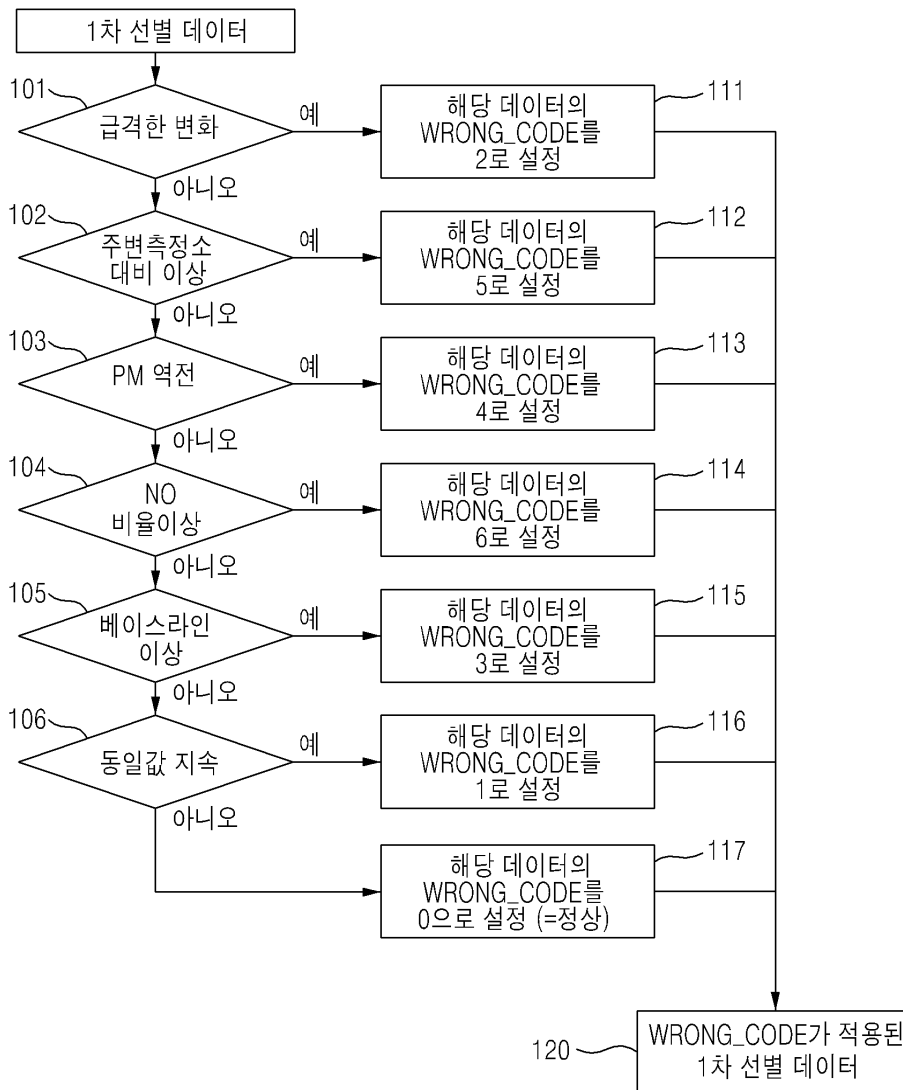
른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

[0079]

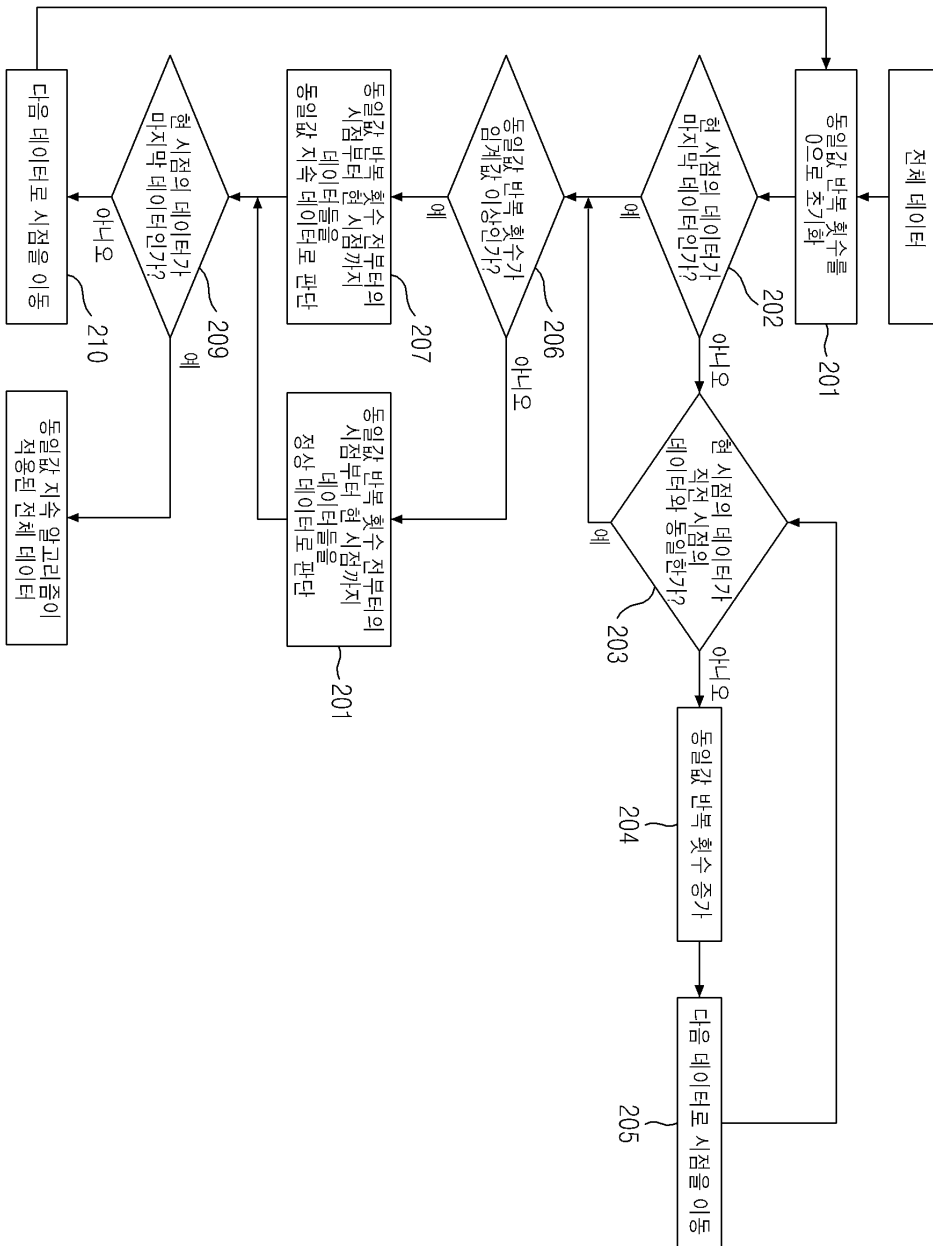
그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

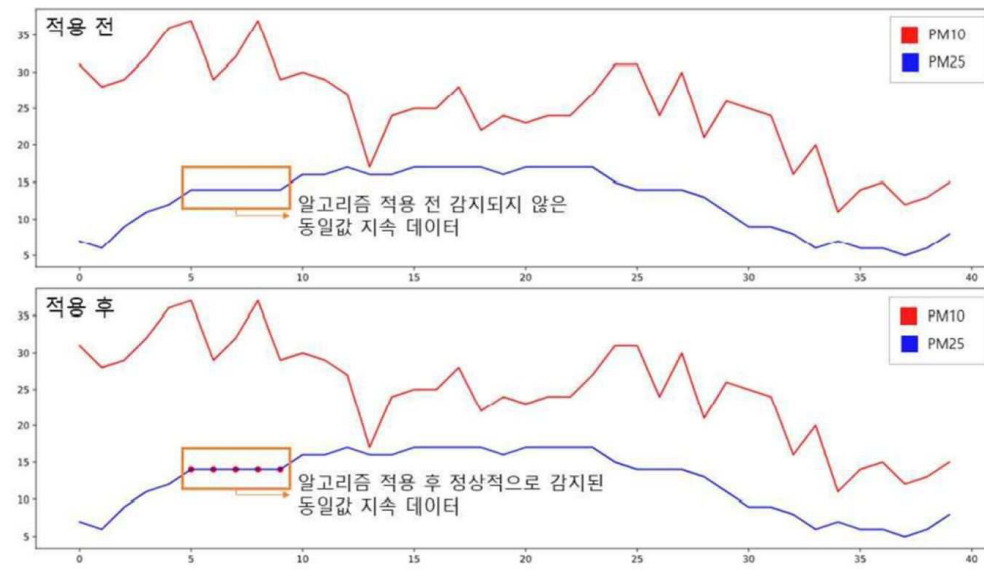
도면1



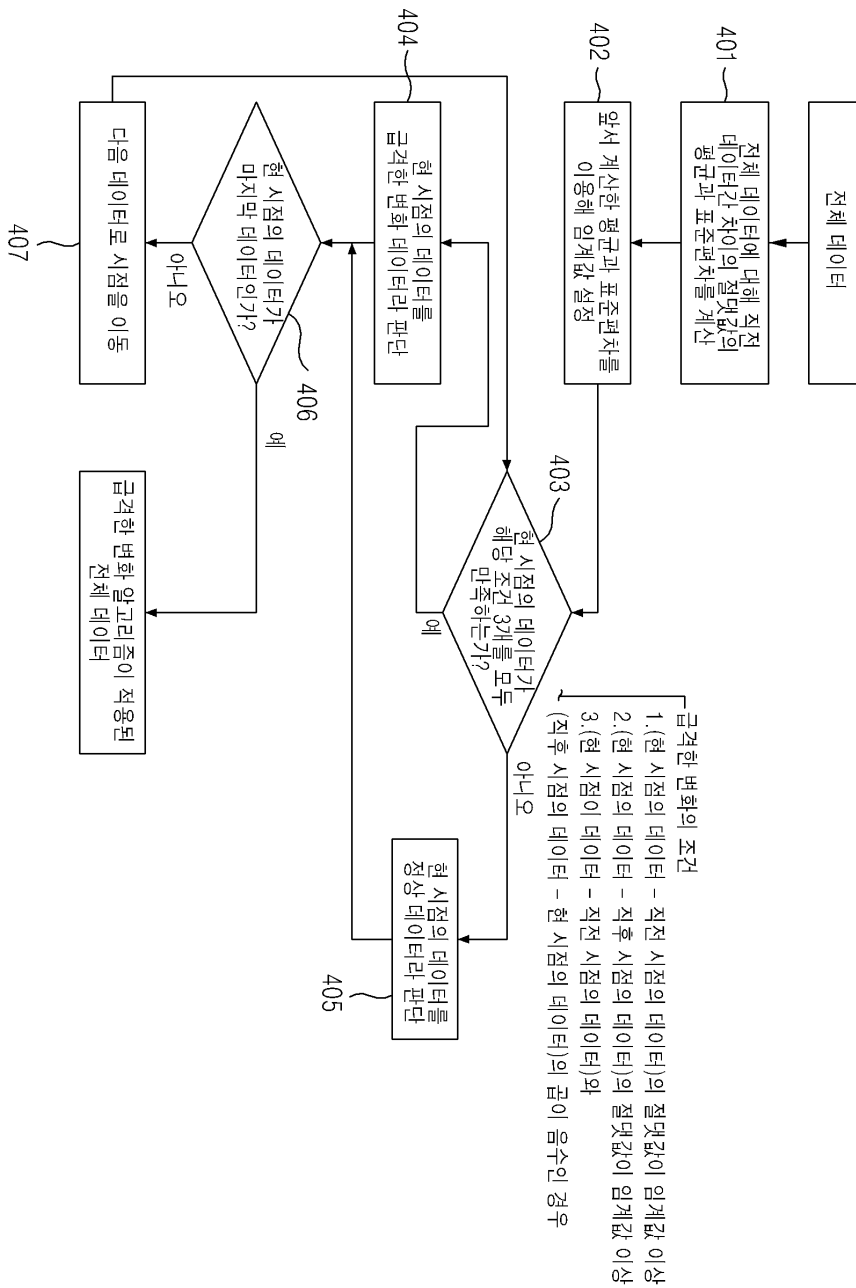
도면2



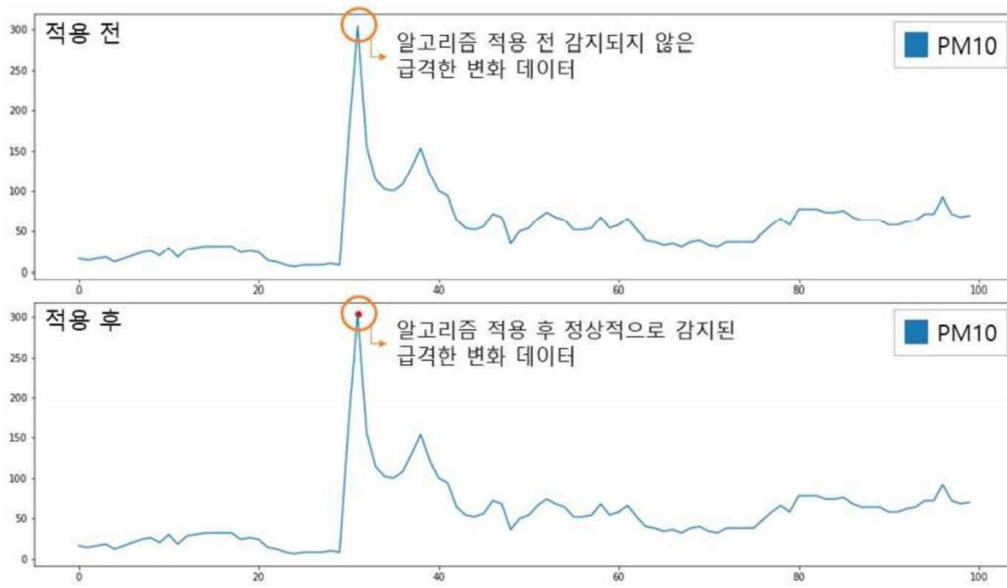
도면3



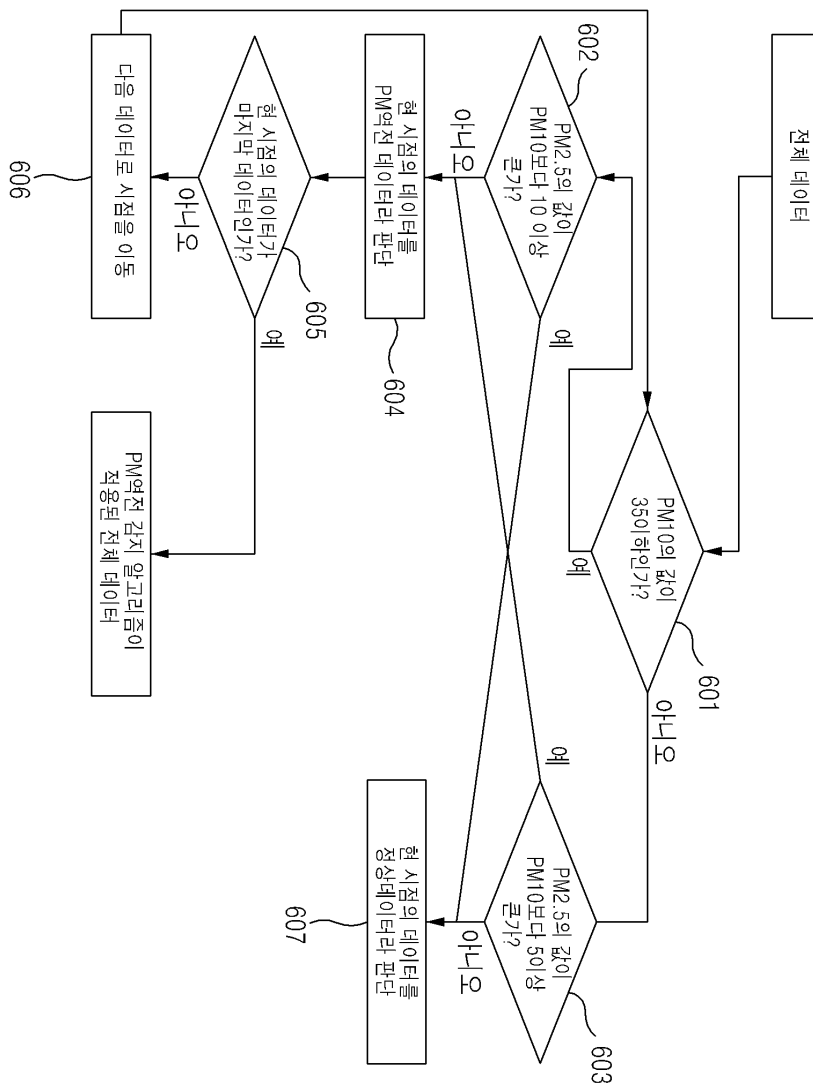
도면4



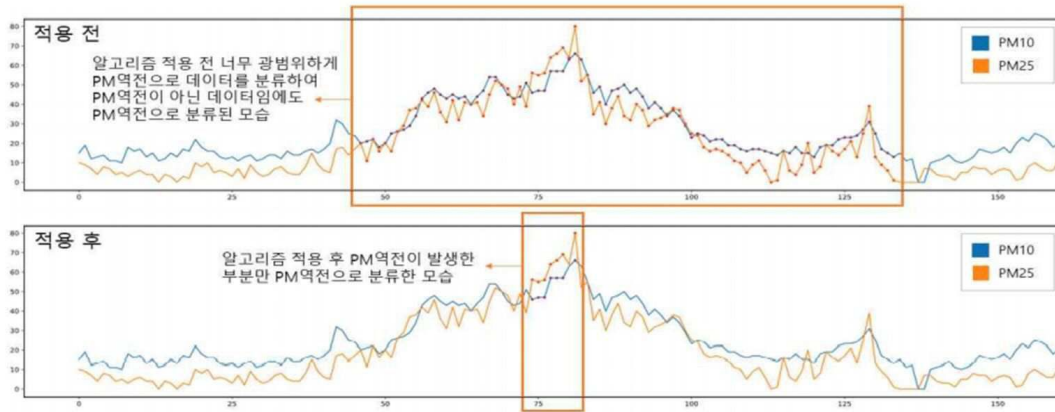
도면5



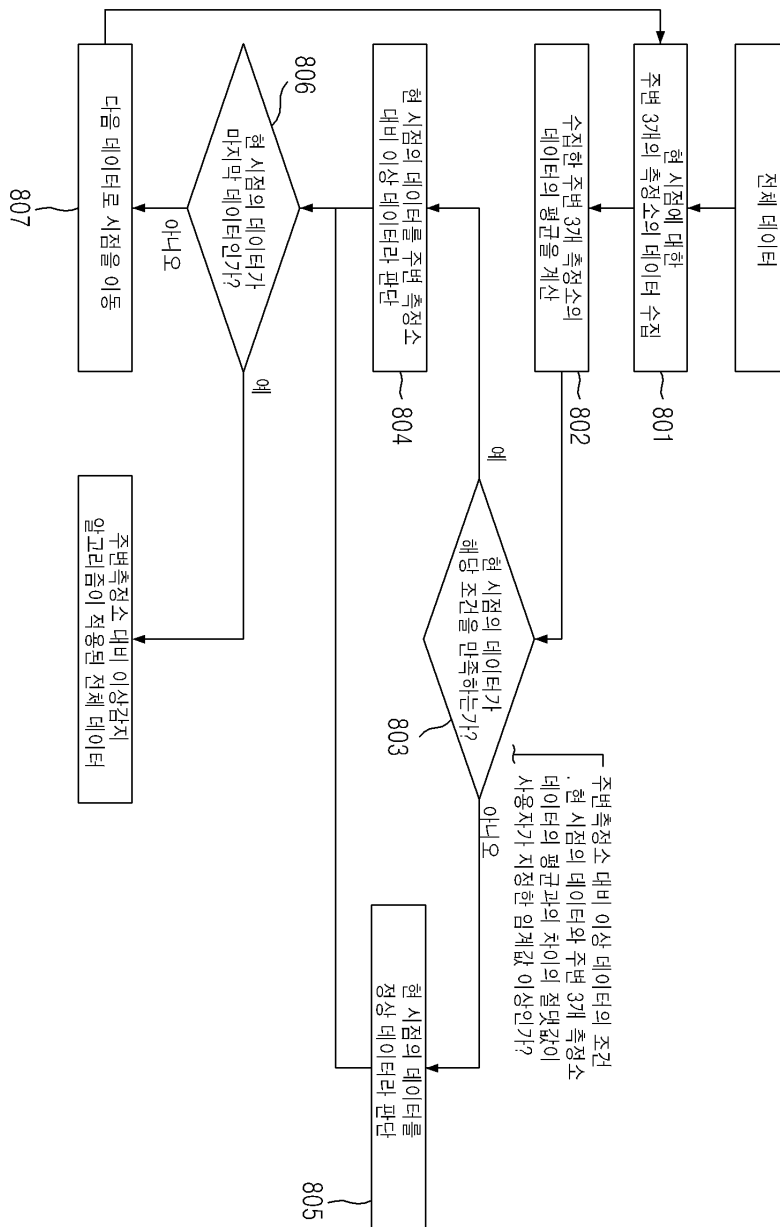
도면6



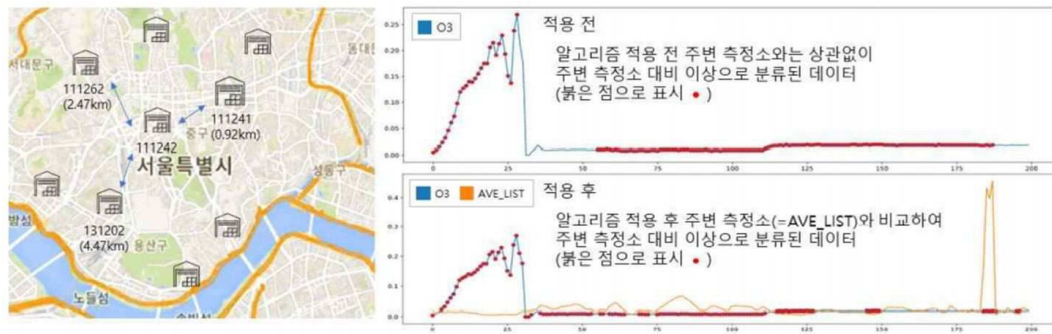
도면7



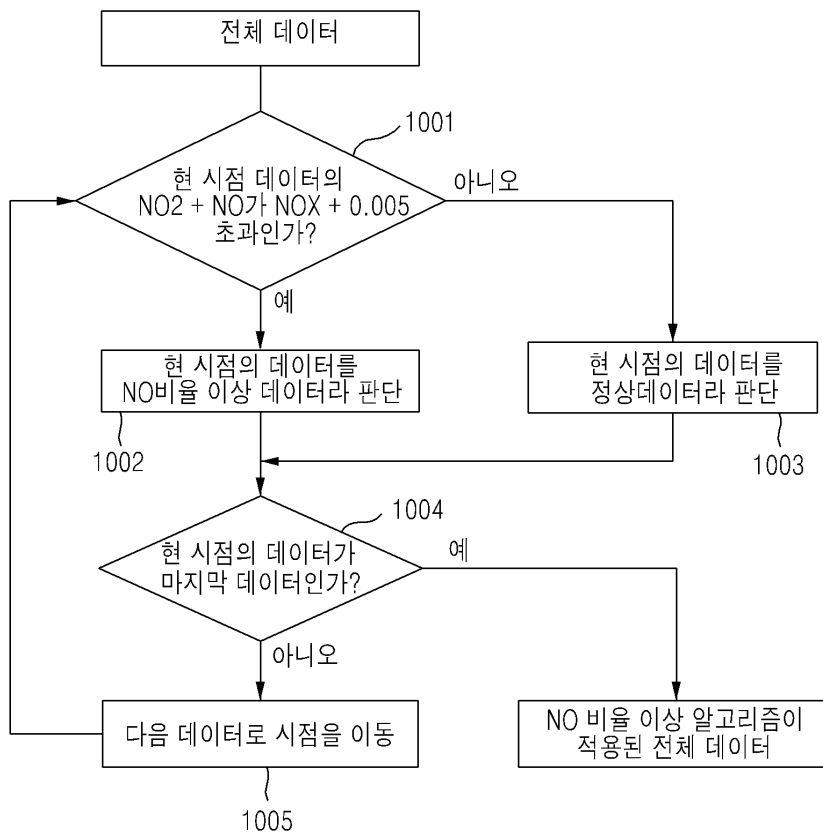
도면8



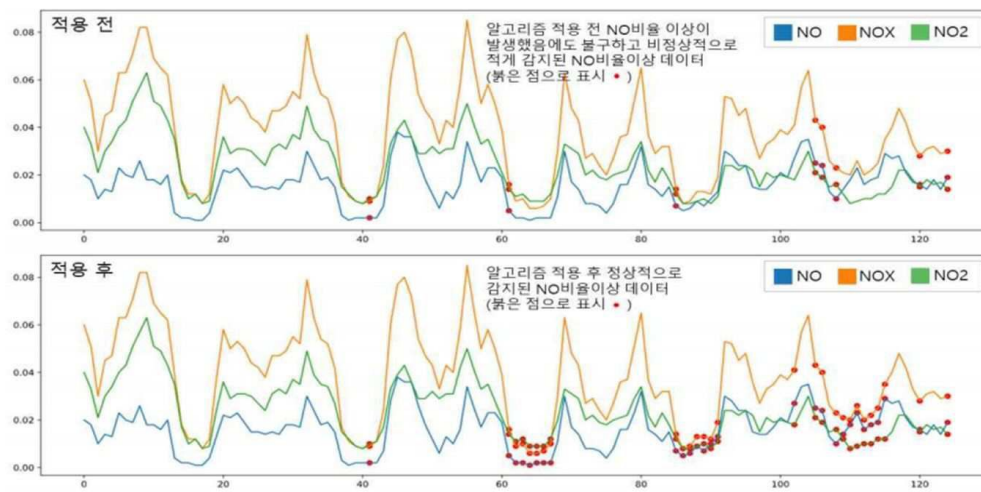
도면9



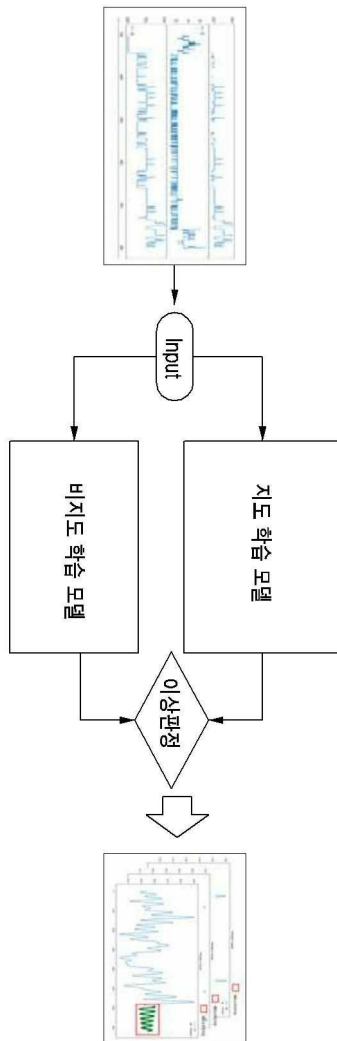
도면10



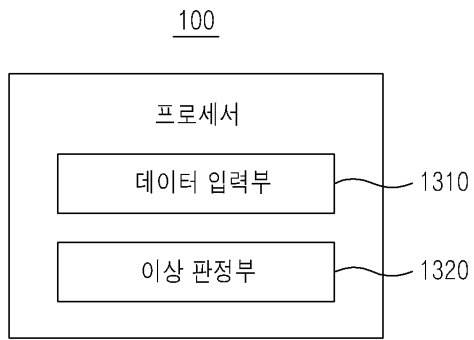
도면11



도면12



도면13



도면14

