



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2023년06월30일  
(11) 등록번호 10-2549313  
(24) 등록일자 2023년06월26일

(51) 국제특허분류(Int. Cl.)  
G06F 30/27 (2020.01) G06F 113/08 (2020.01)  
G06F 18/10 (2023.01) G06F 18/214 (2023.01)  
G06Q 50/10 (2012.01) G06Q 50/26 (2012.01)  
(52) CPC특허분류  
G06F 30/27 (2020.01)  
G06F 18/10 (2023.01)  
(21) 출원번호 10-2023-0021417  
(22) 출원일자 2023년02월17일  
심사청구일자 2023년02월17일  
(56) 선행기술조사문헌  
CN115269675 A

(73) 특허권자  
대한민국(환경부 국립환경과학원장)  
인천 서구 환경로 42, 종합환경연구단지 (경서동)  
포항공과대학교 산학협력단  
경상북도 포항시 남구 청암로 77 (지곡동)  
(72) 발명자  
강필구  
경기도 김포시 김포한강8로 173-277, 804동 802호  
박재홍  
서울시 관악구 문성로 30길 9  
(뒷면에 계속)  
(74) 대리인  
이은철, 이수찬

전체 청구항 수 : 총 10 항

심사관 : 김호진

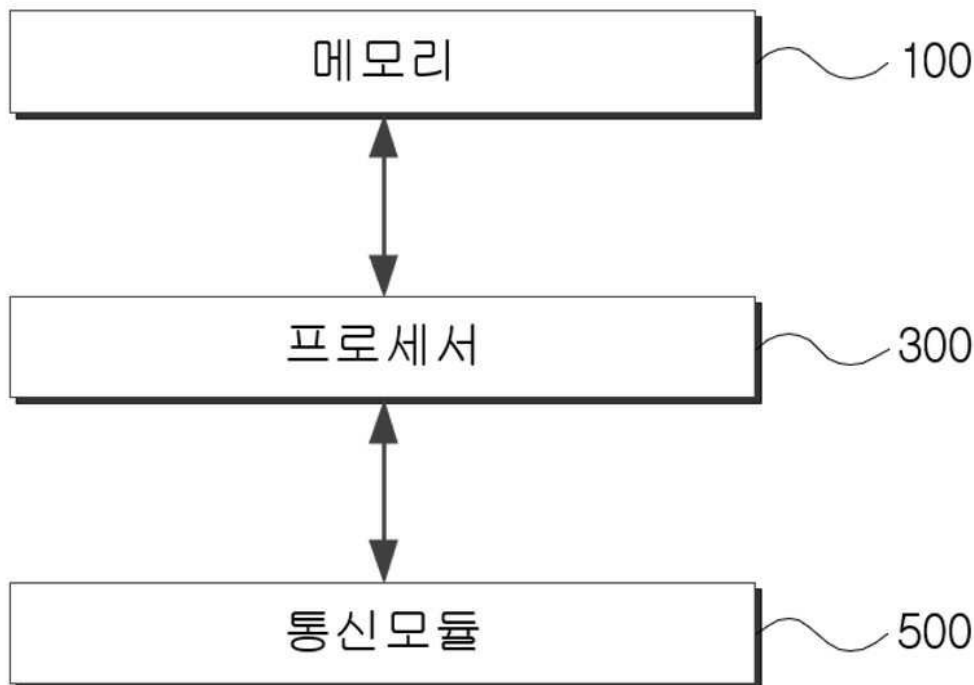
(54) 발명의 명칭 오염물질 배출수준 산정 시스템 및 방법

(57) 요약

본 발명은 오염물질 배출수준 산정 시스템 및 방법에 관한 것으로, 컴퓨터에서 판독가능한 명령어가 저장되는 메모리; 및 메모리에 저장된 명령어를 실행하도록 구현되는 적어도 하나의 프로세서를 포함하고, 적어도 하나의 프로세서가 오염물질을 배출하는 적어도 하나의 설비로부터 서로 다른 유형의 운영인자 데이터를 시점별로 수집하

(뒷면에 계속)

대표도 - 도1



고, 수집된 서로 다른 유형의 운영인자 데이터로부터 기 설정된 항목에 의거하여 적어도 하나의 항목에 대한 로우 데이터(Raw data)를 추출하고 추출된 항목별 로우 데이터의 이상치(Outlier)를 제거하고, 이상치가 제거된 항목별 로우 데이터를 랜덤 포레스트(Random Forest) 알고리즘을 이용하여 운영인자 데이터에 대한 중요도를 산출하고 산출된 중요도의 누적합이 기 설정된 기준값 이상인 운영인자 데이터를 주요 운영인자 데이터로 설정하며, 이상치가 제거된 항목별 로우 데이터 및 설정된 주요 운영인자 데이터를 딥러닝 알고리즘을 기반으로 학습하여 오염물질의 배출농도를 도출하는 구성으로 사람의 의사결정 개입을 최소화하여 오염물질 배출농도 수준에 대한 데이터의 신뢰성이 향상되며, 오염물질 배출 농도를 저감할 수 있다.

(52) CPC특허분류

*G06F 18/214* (2023.01)

*G06Q 50/10* (2015.01)

*G06Q 50/26* (2013.01)

*G06F 2113/08* (2020.01)

(72) 발명자

**김태우**

인천시 미추홀구 낙섬중로 18, 6-304

**서진원**

인천광역시 서구 승학로506번길 24-1, 203호

**전태완**

경기도 고양시 일산 서구 대산로 183, 문촌마을

603동 703호

**변상운**

대전광역시 서구 둔산북로 160, 한마루아파트 109

동 801호

**김중훈**

경상북도 포항시 남구 청암로 77 포항공과대학교

**민승기**

경상북도 포항시 남구 청암로 77 포항공과대학교

**이광훈**

경상북도 포항시 남구 청암로 77 포항공과대학교

**최승희**

경상북도 포항시 남구 청암로 77 포항공과대학교

## 명세서

### 청구범위

#### 청구항 1

오염물질 배출수준 산정 시스템에 있어서,

컴퓨터에서 판독가능한 명령어가 저장되는 메모리; 및

상기 메모리에 저장된 명령어를 실행하도록 구현되는 적어도 하나의 프로세서를 포함하고,

상기 적어도 하나의 프로세서는,

오염물질을 배출하는 적어도 하나의 설비로부터 서로 다른 유형의 운영인자 데이터를 시점별로 수집하는 제1 프로세스;

수집된 서로 다른 유형의 상기 운영인자 데이터로부터 기 설정된 항목에 의거하여 적어도 하나의 항목에 대한 로우 데이터(Raw data)를 추출하고 추출된 상기 항목별 로우 데이터의 이상치(Outlier)를 제거하는 제2 프로세스;

이상치가 제거된 상기 항목별 로우 데이터를 랜덤 포레스트(Random Forest) 알고리즘에 적용하여 상기 운영인자 데이터에 대한 중요도를 산출하고 산출된 상기 중요도의 누적합이 기 설정된 기준값 이상인 운영인자 데이터를 주요 운영인자 데이터로 설정하는 제3 프로세스; 및

상기 이상치가 제거된 항목별 로우 데이터 및 설정된 주요 운영인자 데이터를 딥러닝 알고리즘을 기반으로 학습하여 오염물질의 배출농도를 도출하는 제4 프로세스를 포함하는 오염물질 배출수준 산정 시스템.

#### 청구항 2

제 1 항에 있어서,

상기 제2 프로세스는,

상기 로우 데이터로부터 제1 시점의 오염물질 농도값 및 상기 제1 시점 바로 이전인 제2 시점의 오염물질 농도값의 차이가 기 설정된 임계값보다 크면 이전 시점의 오염물질의 농도값을 제거하는 것을 특징으로 하는 오염물질 배출수준 산정 시스템.

#### 청구항 3

제 1 항에 있어서,

상기 제4 프로세스는,

설정된 시간범위에 따른 적어도 하나의 오염물질의 평균절대오차(Mean Absolute Error, MAE)를 시점별로 도출하고 도출된 평균절대오차가 기 설정된 수치 미만인 시점을 학습에 사용할 시퀀스 길이(Sequence length)로 설정하는 것을 특징으로 하는 오염물질 배출수준 산정 시스템.

#### 청구항 4

제 1 항에 있어서,

상기 적어도 하나의 프로세서는,

상기 도출된 오염물질의 배출농도에 따른 자가진단평가를 수행하는 제5 프로세스를 더 포함하고,

상기 제5 프로세스는,

설정된 기간동안 배출되는 오염물질의 이상 배출 횟수가 기 설정된 제1 기준치를 초과하는 제1 목표기간을 도출하고, 전체 기간동안 배출되는 오염물질의 이상 배출 횟수가 기 설정된 제2 기준치를 초과하는 제2 목표기간을 도출하여 도출된 상기 제1 목표기간 및 도출된 상기 제2 목표기간의 이상치 비율(False Alarm Rate)이 1에 가까운 제3 목표기간을 도출하는 것을 특징으로 하는 오염물질 배출수준 산정 시스템.

**청구항 5**

제 4 항에 있어서,

상기 제5 프로세스는,

도출된 상기 제3 목표기간의 오염물질 배출 변화량으로 자기진단계수(Autocorrelation)을 산출하여 산출된 자기진단계수에 따라 상기 제3 목표기간의 정상, 주의, 및 경고 중 어느 하나의 상태를 도출하는 것을 특징으로 하는 오염물질 배출수준 산정 시스템.

**청구항 6**

오염물질 배출수준 산정 시스템에서 수행되는 오염물질 배출수준 산정 방법에 있어서,

적어도 하나의 프로세서가 오염물질을 배출하는 적어도 하나의 설비로부터 서로 다른 유형의 운영인자 데이터를 시점별로 수집하는 제1 단계;

상기 적어도 하나의 프로세서가 수집된 서로 다른 유형의 상기 운영인자 데이터로부터 기 설정된 항목에 의거하여 적어도 하나의 항목에 대한 로우 데이터(Raw data)를 추출하고 추출된 상기 항목별 로우 데이터의 이상치(Outlier)를 제거하는 제2 단계;

상기 적어도 하나의 프로세서가 이상치가 제거된 상기 항목별 로우 데이터를 랜덤 포레스트(Random Forest) 알고리즘에 적용하여 상기 운영인자 데이터에 대한 중요도를 산출하고 산출된 상기 중요도의 누적합이 기 설정된 기준값 이상인 운영인자 데이터를 주요 운영인자 데이터로 설정하는 제3 단계; 및

상기 적어도 하나의 프로세서가 상기 이상치가 제거된 항목별 로우 데이터 및 설정된 주요 운영인자 데이터를 딥러닝 알고리즘을 기반으로 학습하여 오염물질의 배출농도를 도출하는 제4 단계를 포함하는 오염물질 배출수준 산정 방법.

**청구항 7**

제 6 항에 있어서,

상기 적어도 하나의 프로세서가 상기 도출된 오염물질의 배출농도에 따른 자기진단평가를 수행하는 제5 단계를 더 포함하고,

상기 제5 단계는,

설정된 기간동안 배출되는 오염물질의 농도값이 기 설정된 제1 기준치를 초과하는 제1 목표기간을 도출하고, 전체 기간동안 배출되는 오염물질의 농도값이 기 설정된 제2 기준치를 초과하는 제2 목표기간을 도출하여 도출된 상기 제1 목표기간 및 도출된 상기 제2 목표기간의 이상치 비율(False Alarm Rate)이 1에 가까운 제3 목표기간을 도출하는 것을 특징으로 하는 오염물질 배출수준 산정 방법.

**청구항 8**

제 7 항에 있어서,

상기 제5 단계는,

도출된 상기 제3 목표기간의 오염물질 배출 변화량으로 자가진단계수(Autocorrelation)을 산출하여 산출된 자가진단계수에 따라 상기 제3 목표기간에 대한 정상, 주의, 및 경고 중 어느 하나의 상태를 도출하는 것을 특징으로 하는 오염물질 배출수준 산정 방법.

**청구항 9**

제 6 항 내지 제 7 항 중 어느 한 항의 오염물질 배출수준 산정 방법을 컴퓨터에 실행시키기 위한 프로그램이 기록된 것을 특징으로 하는 컴퓨터에서 판독가능한 기록매체.

**청구항 10**

컴퓨터와 결합되어 오염물질 배출수준 산정 방법을 컴퓨터에 실행시키기 위해 컴퓨터에서 판독가능한 기록매체에 저장된 컴퓨터 프로그램에 있어서,

상기 오염물질 배출수준 산정 방법은,

오염물질을 배출하는 적어도 하나의 설비로부터 서로 다른 유형의 운영인자 데이터를 시점별로 수집하는 제1 단계;

수집된 서로 다른 유형의 상기 운영인자 데이터로부터 기 설정된 항목에 의거하여 적어도 하나의 항목에 대한 로우 데이터(Raw data)를 추출하고 추출된 상기 항목별 로우 데이터의 이상치(Outlier)를 제거하는 제2 단계;

이상치가 제거된 상기 항목별 로우 데이터를 랜덤 포레스트(Random Forest) 알고리즘에 적용하여 상기 운영인자 데이터에 대한 중요도를 산출하고 산출된 상기 중요도의 누적합이 기 설정된 기준값 이상인 운영인자 데이터를 주요 운영인자 데이터로 설정하는 제3 단계; 및

상기 이상치가 제거된 항목별 로우 데이터 및 설정된 주요 운영인자 데이터를 딥러닝 알고리즘을 기반으로 학습하여 오염물질의 배출농도를 도출하는 제4 단계를 포함하는 오염물질 배출수준 산정 프로그램.

**발명의 설명**

**기술 분야**

[0001] 본 발명은 오염물질 배출수준 산정 시스템 및 방법에 관한 것으로, 공장별 또는 설비별로 배출되는 오염물질의 배출수준을 산정할 수 있도록 한 기술에 관한 것이다.

**배경 기술**

[0003] 한국환경관리공단은 굴뚝원격감시체계(CleanSYS)를 시행하고 이를 온라인에 공개하고 있다. 이때, 각 공장별로 기 설정된 시간마다 굴뚝에서 오염물질의 배출농도를 측정하는 굴뚝자동측정기기(TMS)로부터 측정된 데이터는 매년 수백만개 이상으로 생성된다. 이러한 방대한 데이터는 사람이 직접 수작업을 거쳐 범용 프로그램을 이용하여 정리하고 있으며, 데이터를 기반으로 배출허용기준의 초과를 판단하여 행정처분을 진행한다.

[0004] 종래의 오염물질 배출농도 평가방식은 해당 업종과 오염물질에 따라 다른 정책에 의해 각기 다른 유형정보를 사용해야 하나 인간의 의사결정이 개입되어 이상치 판단 결과의 상이함에 따라 최종 도출되는 데이터의 객관성이 저하되는 문제가 있다.

[0005] 또한, 종래의 평가방식은 공통된 정보를 가진 데이터를 취합하는 과정에서 데이터의 필터링 작업은 필수적으로 이루어지나 사람의 손으로 직접 작업함에 따라 비효율적인 작업 방식 및 방대한 자료 검수로 인한 실수가 발생하여 데이터 손실의 우려가 존재하며, 공장의 오염물질 배출 저감에 초점을 두지 아니하여 배출기준 또는 배출량의 변화에 따른 사전 대처가 불가능한 문제가 있다.

[0006] 이러한 문제를 해결하기 위해, 각 공장 또는 설비별로 오염물질의 배출수준을 산정하여 실시간으로 평가하고 공장 개별에 따른 진단기준을 최적화하여 의사판단을 지원할 수 있는 기술 개발이 필요하다.

[0007] 본 발명은 2022년도 환경부의 재원으로 국립환경과학원의 지원을 받아 수행된 스마트 통합환경관리체계 구축 시범사업의 결과이다(과제번호 : NIER-2021-03-02-031)

**선행기술문헌**

**특허문헌**

[0009] (특허문헌 0001) 대한민국 등록특허 제10-1756791호(2017.07.12. 공고)

**발명의 내용**

**해결하려는 과제**

[0010] 본 발명은, 사람의 의사결정 개입을 최소화하고, 데이터를 기반으로 한 평가방식으로 객관성을 확립하며, 실시간으로 필요한 오염물질의 양을 판단하여 공장의 가동 및 운영 방식에 대한 조언 또는 자료를 제공하며, 배출 기준 또는 배출량의 변화 발생 시 사전에 대처할 수 있도록 각 사업장에 따른 최적화된 모델과 진단 기준을 제시할 수 있는 오염물질 배출수준 산정 시스템 및 방법을 제공할 수 있다.

**과제의 해결 수단**

[0012] 본 발명의 일 측면에 따른 오염물질 배출수준 산정 시스템은 컴퓨터에서 관독가능한 명령어가 저장되는 메모리; 및 상기 메모리에 저장된 명령어를 실행하도록 구현되는 적어도 하나의 프로세서를 포함하고, 상기 적어도 하나의 프로세서는, 오염물질을 배출하는 적어도 하나의 설비로부터 서로 다른 유형의 운영인자 데이터를 시점별로 수집하는 제1 프로세스; 수집된 서로 다른 유형의 상기 운영인자 데이터로부터 기 설정된 항목에 의거하여 적어도 하나의 항목에 대한 로우 데이터(Raw data)를 추출하고 추출된 상기 항목별 로우 데이터의 이상치(Outlier)를 제거하는 제2 프로세스; 이상치가 제거된 상기 항목별 로우 데이터를 랜덤 포레스트(Random Forest) 알고리즘에 적용하여 상기 운영인자 데이터에 대한 중요도를 산출하고 산출된 상기 중요도의 누적합이 기 설정된 기준값 이상인 운영인자 데이터를 주요 운영인자 데이터로 설정하는 제3 프로세스; 및 상기 이상치가 제거된 항목별 로우 데이터 및 설정된 주요 운영인자 데이터를 딥러닝 알고리즘을 기반으로 학습하여 오염물질의 배출농도를 도출하는 제4 프로세스를 포함한다.

[0013] 바람직하게는, 상기 제2 프로세스는, 상기 로우 데이터로부터 제1 시점의 오염물질 농도값 및 상기 제1 시점 바로 이전인 제2 시점의 오염물질 농도값의 차이가 기 설정된 임계값보다 크면 이전 시점의 오염물질의 농도값을 제거할 수 있다.

[0014] 바람직하게는, 상기 제4 프로세스는, 설정된 시간범위에 따른 적어도 하나의 오염물질의 평균절대오차(Mean Absolute Error, MAE)를 시점별로 도출하고 도출된 평균절대오차가 기 설정된 수치 미만인 시점을 학습에 사용할 시퀀스 길이(Sequence length)로 설정할 수 있다.

[0015] 바람직하게는, 상기 적어도 하나의 프로세서는, 상기 도출된 오염물질의 배출농도에 따른 자가진단평가를 수행하는 제5 프로세스를 더 포함하고, 상기 제5 프로세스는, 설정된 기간동안 배출되는 오염물질의 이상 배출 횟수가 기 설정된 제1 기준치를 초과하는 제1 목표기간을 도출하고, 전체 기간동안 배출되는 오염물질의 이상 배출 횟수가 기 설정된 제2 기준치를 초과하는 제2 목표기간을 도출하여 도출된 상기 제1 목표기간 및 도출된 상기 제2 목표기간의 이상치 비율(False Alarm Rate)이 1에 가까운 제3 목표기간을 도출할 수 있다.

[0016] 바람직하게는, 상기 제5 프로세스는, 도출된 상기 제3 목표기간의 오염물질 배출 변화량으로 자기진단계수(Autocorrelation)을 산출하여 산출된 자기진단계수에 따라 상기 제3 목표기간의 정상, 주의, 및 경고 중 어느 하나의 상태를 도출할 수 있다.

[0017] 본 발명의 다른 측면에 따른 오염물질 배출수준 산정 시스템에서 수행되는 오염물질 배출수준 산정 방법은 적어도 하나의 프로세서가 오염물질을 배출하는 적어도 하나의 설비로부터 서로 다른 유형의 운영인자 데이터를 시점별로 수집하는 제1 단계; 상기 적어도 하나의 프로세서가 수집된 서로 다른 유형의 상기 운영인자 데이터로부터 기 설정된 항목에 의거하여 적어도 하나의 항목에 대한 로우 데이터(Raw data)를 추출하고 추출된 상기 항목별 로우 데이터의 이상치(Outlier)를 제거하는 제2 단계; 상기 적어도 하나의 프로세서가 이상치가 제거된 상기 항목별 로우 데이터를 랜덤 포레스트(Random Forest) 알고리즘에 적용하여 상기 운영인자 데이터에 대한 중요도

를 산출하고 산출된 상기 중요도의 누적합이 기 설정된 기준값 이상인 운영인자 데이터를 주요 운영인자 데이터로 설정하는 제3 단계; 및 상기 적어도 하나의 프로세서가 상기 이상치가 제거된 항목별 로우 데이터 및 설정된 주요 운영인자 데이터를 딥러닝 알고리즘을 기반으로 학습하여 오염물질의 배출농도를 도출하는 제4 단계를 포함한다.

[0018] 바람직하게는, 상기 적어도 하나의 프로세서가 상기 도출된 오염물질의 배출농도에 따른 자가진단평가를 수행하는 제5 단계를 더 포함하고, 상기 제5 단계는, 설정된 기간동안 배출되는 오염물질의 농도값이 기 설정된 제1 기준치를 초과하는 제1 목표기간을 도출하고, 전체 기간동안 배출되는 오염물질의 농도값이 기 설정된 제2 기준치를 초과하는 제2 목표기간을 도출하여 도출된 상기 제1 목표기간 및 도출된 상기 제2 목표기간의 이상치 비율(False Alarm Rate)이 1에 가까운 제3 목표기간을 도출할 수 있다.

[0019] 바람직하게는, 상기 제5 단계는, 도출된 상기 제3 목표기간의 오염물질 배출 변화량으로 자가진단계수(Autocorrelation)을 산출하여 산출된 자가진단계수에 따라 상기 제3 목표기간에 대한 정상, 주의, 및 경고 중 어느 하나의 상태를 도출할 수 있다.

[0020] 오염물질 배출수준 산정 방법을 컴퓨터에 실행시키기 위한 프로그램이 기록된 것을 특징으로 하는 컴퓨터에서 판독가능한 기록매체를 제공한다.

[0021] 컴퓨터와 결합되어 오염물질 배출수준 산정 방법을 컴퓨터에 실행시키기 위해 컴퓨터에서 판독가능한 기록매체에 저장된 컴퓨터 프로그램을 제공한다.

**발명의 효과**

[0023] 본 발명에 따르면, 사람의 의사결정 개입을 최소화하여 오염물질 배출농도 수준에 대한 데이터의 신뢰성이 향상되며, 오염물질 배출 농도를 저감할 수 있다.

**도면의 간단한 설명**

[0025] 도 1은 일 실시예에 따른 오염물질 배출수준 산정 시스템의 구성도이다.

도 2는 일 실시예에 따른 프로세서의 세부구성도이다.

도 3은 일 실시예에 따른 오염물질 배출수준 산정 시스템의 흐름도이다.

도 4는 일 실시예에 따른 배출 오염물질 배출수준을 산정하는 세부 흐름도이다.

도 5는 일 실시예에 따른 자가진단평가를 수행하는 세부 흐름도이다.

도 6은 일 실시예에 따른 이상치가 제거된 그래프이다.

도 7은 일 실시예에 따른 운영인자 데이터의 중요도를 도시한 그래프이다.

도 8은 일 실시예에 따른 BiLSTM-AE를 계략적으로 도시한 도이다.

도 9는 일 실시예에 따른 평균절대오차를 나타낸 그래프이다.

도 10은 일 실시예에 따른 오염물질 배출수준 산정 방법을 나타낸 순서도이다.

**발명을 실시하기 위한 구체적인 내용**

[0026] 이하에서는 본 발명에 따른 오염물질 배출수준 산정 시스템 및 방법을 첨부된 도면들을 참조하여 상세하게 설명한다. 이러한 과정에서 도면에 도시된 선들의 두께나 구성요소의 크기 등은 설명의 명료성과 편의상 과장되게 도시되어 있을 수 있다. 또한, 후술되는 용어들은 본 발명에서의 기능을 고려하여 정의된 용어들로서 이는 운용자의 의도 또는 관례에 따라 달라질 수 있다. 그러므로, 이러한 용어들에 대한 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다.

[0027] 본 발명의 목적 및 효과는 하기의 설명에 의해서 자연스럽게 이해되거나 보다 분명해질 수 있으며, 하기의 기재만으로 본 발명의 목적 및 효과가 제한되는 것은 아니다. 또한, 본 발명을 설명함에 있어서 본 발명과 관련된 공지 기술에 대한 구체적인 설명이, 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략하기로 한다.

[0028] 도 1은 일 실시예에 따른 오염물질 배출수준 산정 시스템의 구성도이다.



- [0029] 도 1에서 나타낸 바와 같이, 일 실시예에 따른 오염물질 배출수준 산정 시스템은 컴퓨터에서 판독가능한 명령어가 저장되는 메모리(100) 및 상기 메모리(100)에 저장된 명령어를 실행하도록 구현되는 적어도 하나의 프로세서(300)를 포함하고, 적어도 하나의 프로세서(300)는, 제1 프로세스, 제2 프로세스, 제3 프로세스, 및 제4 프로세스를 포함하며, 제5 프로세스를 더 포함할 수 있다.
- [0030] 제1 프로세스는 오염물질을 배출하는 적어도 하나의 설비로부터 서로 다른 유형의 운영인자 데이터를 시점별로 수집한다. 운영인자 데이터는 대기배출원관리시스템(Stack Emission Management System, SEMS)에서 측정된 SEMS 데이터, 굴뚝원격감시체계(tele-Monitoring System, TMS)에서 측정된 TMS 데이터, WTMS 데이터, 및 WEMS 데이터일 수 있다. 여기서, SEMS 데이터로부터 사업장명, 배출구번호, 분류체계, 오염물질, 대기법 배출시설 상세구분, 연도, 농도, 단위의 항목의 로우 데이터를 추출하고, 상기 TMS 데이터로부터 연도, 사업장명, 배출구번호, TSP 농도, SOX 농도, NOX 농도, HCl 농도, NH3 농도, 및 CO 농도의 항목의 로우 데이터를 추출할 수 있다.
- [0031] 제2 프로세스는 수집된 서로 다른 유형의 상기 운영인자 데이터로부터 기 설정된 항목에 의거하여 적어도 하나의 항목에 대한 로우 데이터(Raw data)를 추출하고 추출된 상기 항목별 로우 데이터의 이상치(Outlier)를 제거한다. 여기서, 제2 프로세스는 상기 로우 데이터로부터 제1 시점의 오염물질 농도값 및 상기 제1 시점 바로 이전인 제2 시점의 오염물질 농도값의 차이가 기 설정된 임계값보다 크면 이전 시점의 오염물질의 농도값을 제거할 수 있다.
- [0032] 제3 프로세스는 이상치가 제거된 상기 항목별 로우 데이터를 랜덤 포레스트(Random Forest) 알고리즘에 적용하여 상기 운영인자 데이터에 대한 중요도를 산출하고 산출된 상기 중요도의 누적합이 기 설정된 기준값 이상인 운영인자 데이터를 주요 운영인자 데이터로 설정한다.
- [0033] 제4 프로세스는 상기 이상치가 제거된 항목별 로우 데이터 및 설정된 주요 운영인자 데이터를 딥러닝 알고리즘을 기반으로 학습하여 오염물질의 배출농도를 도출한다. 여기서, 제4 프로세스는 설정된 시간 범위에 따른 적어도 하나의 오염물질의 평균절대오차(Mean Absolute Error, MAE)를 시점별로 도출하고 도출된 평균절대오차가 기 설정된 수치 미만인 시점을 학습에 사용할 시퀀스 길이(Sequence length)로 설정할 수 있다.
- [0034] 이때, 적어도 하나의 프로세서(300)는, 상기 도출된 오염물질의 배출농도에 따른 자가진단평가를 수행하는 제5 프로세스를 더 포함할 수 있다.
- [0035] 제5 프로세스는 설정된 기간동안 배출되는 오염물질의 이상 배출 횟수가 기 설정된 제1 기준치를 초과하는 제1 목표기간을 도출하고, 전체 기간동안 배출되는 오염물질의 이상 배출 횟수가 기 설정된 제2 기준치를 초과하는 제2 목표기간을 도출하여 도출된 상기 제1 목표기간 및 도출된 상기 제2 목표기간의 이상치 비율(False Alarm Rate)이 1에 가까운 제3 목표기간을 도출할 수 있다. 또한, 제5 프로세스는 도출된 상기 제3 목표기간의 오염물질 배출 변화량으로 자가진단계수(Autocorrelation)을 산출하여 산출된 자가진단계수에 따라 상기 제3 목표기간에 대한 정상, 주의, 및 경고 중 어느 하나의 상태를 도출할 수 있다.
- [0036] 일 실시예에 따른 상기 적어도 하나의 프로세서(300)는 추출된 상기 로우 데이터를 최적가용기법 연계배출수준(Best Available Techniques Associate Emission Level, BAT-AEL) 산정 계산식으로 연산하여 상기 설비의 정상 운영 중 배출되는 오염물질의 배출농도 수준을 도출할 수 있다. 상기 적어도 하나의 프로세서(300)는 도출된 상기 배출농도 수준으로 기 설정된 분류기준에 따른 피벗 테이블(Pivot table) 및 상기 항목별로 로우 데이터를 비교한 그래프 중 어느 하나를 생성할 수 있다. 상기 적어도 하나의 프로세서(300)는 서로 다른 유형의 운영인자 데이터의 항목들을 기 설정된 텍스트의 항목으로 변경하고, 상기 항목에 따라 로우 데이터를 취합할 수 있다. 또한, 상기 적어도 하나의 프로세서(300)는 로우 데이터가 추출된 적어도 하나의 항목을 그룹으로 설정하고 설정된 그룹별로 그룹식별정보(Identification)를 생성할 수 있다.
- [0037] 오염물질은 먼지, 황산화물, 질소산화물, 염화수소, 불소화합물, 암모니아, 및 일산화탄소일 수 있고, 오염물질 배출농도는 먼지, 황산화물, 질소산화물, 염화수소, 불소화합물, 암모니아, 및 일산화탄소의 배출농도일 수 있다.
- [0038] 도 2는 일 실시예에 따른 프로세서의 세부구성도이다.
- [0039] 도 2에서 나타낸 바와 같이, 일 실시예에 따른 프로세서(300)의 세부구성도는 자료 수집부(310), 전처리부(320), 주요인자 설정부(330), 학습부(340), 자가진단부(350)를 포함할 수 있다.
- [0040] 자료 수집부(310)는 오염물질을 배출하는 적어도 하나의 설비로부터 서로 다른 유형의 운영인자 데이터를 시점별로 수집한다.



- [0041] 전처리부(320)는 수집된 서로 다른 유형의 상기 운영인자 데이터로부터 기 설정된 항목에 의거하여 적어도 하나의 항목에 대한 로우 데이터(Raw data)를 추출하고 추출된 상기 항목별 로우 데이터의 이상치(Outlier)를 제거한다.
- [0042] 주요인자 설정부(330)는 이상치가 제거된 상기 항목별 로우 데이터를 랜덤 포레스트(Random Forest) 알고리즘에 적용하여 상기 운영인자 데이터에 대한 중요도를 산출하고 산출된 상기 중요도의 누적합이 기 설정된 기준값 이상인 운영인자 데이터를 주요 운영인자 데이터로 설정한다.
- [0043] 학습부(340)는 상기 이상치가 제거된 항목별 로우 데이터 및 설정된 주요 운영인자 데이터를 딥러닝 알고리즘을 기반으로 학습하여 오염물질의 배출농도를 도출한다.
- [0044] 자가진단부(350)는 상기 도출된 오염물질의 배출농도에 따른 자가진단평가를 수행한다. 이때 자가진단부(350)는 설정된 기간동안 배출되는 오염물질의 이상 배출 횟수가 기 설정된 제1 기준치를 초과하는 제1 목표기간을 도출하고, 전체 기간동안 배출되는 오염물질의 이상 배출 횟수가 기 설정된 제2 기준치를 초과하는 제2 목표기간을 도출하여 도출된 상기 제1 목표기간 및 도출된 상기 제2 목표기간의 이상치 비율(False Alarm Rate)이 1에 가까운 제3 목표기간을 도출할 수 있다. 또한, 자가진단부(350)는 도출된 상기 제3 목표기간의 오염물질 배출 변화량으로 자기진단계수(Autocorrelation)을 산출하여 산출된 자기진단계수에 따라 상기 제3 목표기간의 정상, 주의, 및 경고 중 어느 하나의 상태를 도출할 수 있다.
- [0045] 도 3은 일 실시예에 따른 오염물질 배출수준 산정 시스템의 흐름도이고, 도 4는 일 실시예에 따른 배출 오염물질 배출수준을 산정하는 세부 흐름도이며, 도 5는 일 실시예에 따른 자가진단평가를 수행하는 세부 흐름도이다.
- [0046] 도 3에서 나타낸 바와 같이, 오염물질 배출수준 산정 시스템은 IoT를 통한 공장운영인자 및 배출오염물질 농도를 측정하고, 주요 공장운영인자를 분석하여 주요 공장운영인자를 이용한 각 배출 오염물질의 배출수준을 산정할 수 있다. 이후 자기진단계수를 통해 산정된 오염물질 배출수준을 평가를 수행할 수 있고, 평가결과를 실시간으로 제공할 수 있다.
- [0047] 인공지능을 이용하여 주요 운영인자 데이터를 설정할 수 있다. 각 오염물질의 배출수준을 산정할 때 및 하나의 시스템으로 보았을 때 랜덤 포레스트 모델의 운영 인자 변수 중요도를 순서대로 나열하여 중요도의 합이 전체의 80%를 차지하는 인자들을 해당 공장의 주요 인자로 설정할 수 있다.
- [0048] 도 4에서 나타낸 바와 같이, 설정된 주요 인자들을 입력으로 하여 인공지능 모델을 학습할 수 있다. 인공지능 모델의 학습 방법은 각 타임 스텝마다 또는 시점마다 각 시점으로부터 기 설정된 시간 이전의 시점 데이터부터 반영할 것인지를 의미하는 하이퍼파라미터(Hyperparameter)인 학습용 시퀀스 길이(Sequence length)를 1 분 내지 300 분 범위 내에서 설정할 수 있다.
- [0049] 시점 t=1에서 시작하여 평균이동한 시점(Moving average time step)의 데이터로 학습된 모델의 평균절대오차(Mean Absolute Error)값을 계산한다. 시점이 t=1 일 때의 평균절대오차가 기 설정된 수치인 10을 초과하는 경우, 시점 t를 5로 변경하여 평균이동한 데이터로 평균절대오차 값을 계산한다. 이후 반복 연산하여 평균절대오차 값이 10이하가 되면 해당 t를 학습에 사용할 시퀀스 길이로 정한다. 시점 t>300일 경우, 1 분 내지 300 분 범위 내의 MAE 값이 최저인 t를 학습에 사용할 시퀀스 길이로 정하여 오염물질 배출수준을 산정할 수 있다.
- [0050] 도 5에서 나타낸 바와 같이, 일 실시예에 따른 자가진단평가는 운영인자 데이터에 대한 간접운영인자 데이터를 생성하여 산정된 오염물질 배출수준 데이터(최소 기간 60일동안 측정된 데이터)에서 자가 진단 시 타겟 오염물질의 이상 배출 목표(95 or 99%이나 법적 기준치)를 설정할 수 있다. 이때, 사전 진단 지수의 경우, 과거와 현재의 오염물질 배출농도가 증가하는 경향을 진단할 수 있도록 자기상관계수(Autocorrelation)를 적용할 수 있고, 이는 다음 수학적식으로 도출될 수 있다.

[0051] [수학식 1]

$$R(t) = \frac{\text{autocorr}(t)}{\text{autocorr}(t - a)}$$

[0052]

[0054] 수학식 1에서, R(t)는 과거와 현재 시점의 오염물질 배출농도의 패턴 변화를 사전에 탐지할 수 있는 지수로 사용할 수 있다. 자기상관계수를 계산 시 일정 기간의 데이터를 사용할지 결정할 수 있다. 한달 기간의 자기상관계수 지수를 이용하여 다음 시점에서의 자기상관계수 지수와 비교한다. 즉, 이를 비율로 계산하여 한달 간격의 오염물질 배출농도 데이터와 현재 측정된 오염물질 배출농도 데이터의 차이를 계산하고 이를 비교하여 공장 또

는 설비의 가동 혹은 오염물질 배출농도에 관한 정보를 나타낼 수 있다.  $t$ 는 1분 내지 60분 간격일 수 있다.

[0055] 과거 산정 오염 데이터에서 타겟 오염물질의 이상 배출 목표를 넘어선 시점들(A)을 도출하여 도출된 시점들로부터 하루부터 30일전까지 24시간 단위로 변경하면서  $R(t)$ 값을 도출할 수 있다. 도출된  $R(t)$ 값이 1.5를 초과하는 시점으로 자가상관계수를 도출할 수 있다. 또한, 전체 기간에서  $R(t)$ 값이 1.5이상인 시점들(B)을 도출하여 이상치 비율(False Alarm Rate)이  $B/A$ 일 때 1에 가까운 시점을  $R(t)$ 값을 설정하는 시점으로 변경하여 자가진단계수  $R(t)$ 값과 자가상관계수를 연산하여 0(정상), 1(주의), 2(경고) 중 어느 하나의 상태를 도출할 수 있다.

[0056] 도 6은 일 실시예에 따른 이상치가 제거된 그래프이다.

[0057] 도 6에서 나타낸 바와 같이, 일 실시예에 따른 이상치 제거는  $t$  시점의 오염물질 배출농도와  $t+1$  시점의 오염물질 배출농도의 차이가 전체 시점의 오염물질 배출농도의 0.1 백분위수와 99.9 백분위수의 차이의 1/2보다 클 경우 이상치로 판단할 수 있고, 이상치로 판단된 시점의 오염물질 배출농도를 삭제할 수 있다. 이는 다음 수학적 식으로 표현될 수 있다.

[0058] [수학적 식 2]

$$x_{t+1} - x_t > \frac{u_x - l_x}{2}$$

[0059]

[0061] 수학적 식 1에서,  $x_t$ 는  $t$  시점의 오염물질 배출농도 데이터  $x$ 이고,  $u_x$ 는 데이터  $x$ 의 99.9 백분위수이고,  $l_x$ 는 0.1 백분위수이다.

[0062] 도 7은 일 실시예에 따른 운영인자 데이터의 중요도를 도시한 그래프이다.

[0063] 도 7에서 나타낸 바와 같이, 일 실시예에 따른 운영인자 데이터의 중요도는 랜덤 포레스트 알고리즘에서 어떠한 운영인자가 오염물질 배출농도에 기여를 하였는지 나타낼 수 있다. 이때, 변수 중요도의 누적합은 1일 수 있으며, 정렬된 중요도의 누적합이 80% 이상에 해당하는 인자들을 해당 설비의 주요 운영인자 데이터로 설정할 수 있다.

[0064] 도 8은 일 실시예에 따른 BiLSTM-AE를 계략적으로 도시한 도이다.

[0065] 도 8에서 나타낸 바와 같이, 일 실시예에 따른 BiLSTM-AE(AutoEncoder)는 64개의 및 32개의 히든 유닛(Hidden unit)으로 이루어진 LSTM 층을 거쳐 입력을 압축된 내부 표현으로 바꾼 후, 32개 및 64개의 히든 유닛으로 이루어진 LSTM 층을 통하여 내부 표현을 복원하여 데이터를 효율적으로 학습하는 구조이다.

[0066] BiLSTM-AE는 LSTM-AE의 각 LSTM 층이 BiLSTM으로 구성되어 순방향과 역방향의 결과를 이용하여 학습하는 구조이며, 초기에는 인코더가 각각 64개, 32개의 히든 유닛의 BiLSTM으로 연결되어 있고, 두번째는 128, 64, 32의 BiLSTM으로 이루어져있다. 인코더와 디코더는 대칭으로 구축될 수 있다. 이를 통해 배치 사이즈(Batch Size) 및 히든 유닛을 최적화할 수 있다. 이때, 배치 사이즈는 학습으로 사용할 데이터를 단위 크기로 재조정하여 학습을 수행할 것인지를 나타내는 하이퍼 파라미터일 수 있다.

[0067] 도 9는 일 실시예에 따른 평균절대오차를 나타낸 그래프이다.

[0068] 도 9에서 나타낸 바와 같이, 평균절대오차가 기 설정된 수치보다 높을 경우 시점  $t$ 를 5씩 늘려가며  $t$ 분 이동평균에 대한 산정 성능을 도출할 수 있다. 평균절대오차가 10보다 낮은 경우의 시점  $t$ 를 시퀀스 길이로 설정하여 모델의 학습 및 산정을 수행할 수 있다. 또한,  $h$  시점의 오염물질 배출수준을 산정할 때  $h-t$  시점부터  $h$ 시점까지의 오염물질 배출농도를 반영할 수 있다.

[0069] 도 10은 일 실시예에 따른 오염물질 배출수준 산정 방법을 나타낸 순서도이다.

[0070] 도 10에서 나타낸 바와 같이, 일 실시예에 따른 오염물질 배출수준 산정 방법은 제1 단계(S100), 제2 단계(S300), 제3 단계(S500), 및 제4 단계(S700)를 포함하고, 제5 단계(S900)를 더 포함할 수 있다.

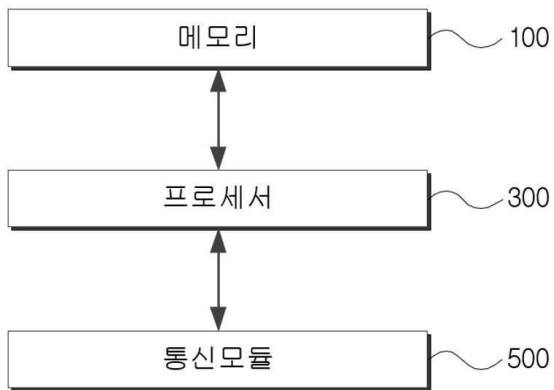
[0071] 제1 단계(S100)는 적어도 하나의 프로세서(300)가 오염을 배출하는 적어도 하나의 설비로부터 서로 다른 유형의 운영인자 데이터를 시점별로 수집한다.

[0072] 제2 단계(S300)는 상기 적어도 하나의 프로세서(300)가 수집된 서로 다른 유형의 상기 운영인자 데이터로부터 기 설정된 항목에 의거하여 적어도 하나의 항목에 대한 로우 데이터(Raw data)를 추출하고 추출된 상기 항목별

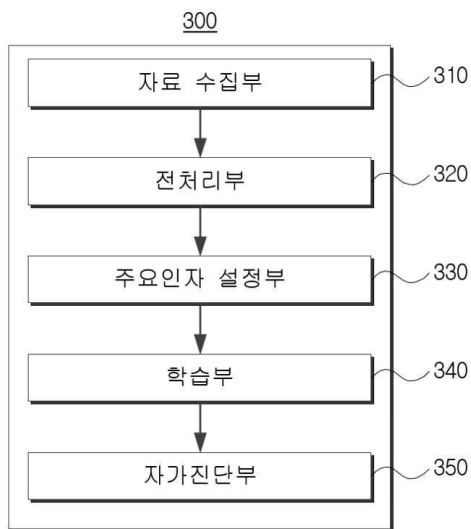


도면

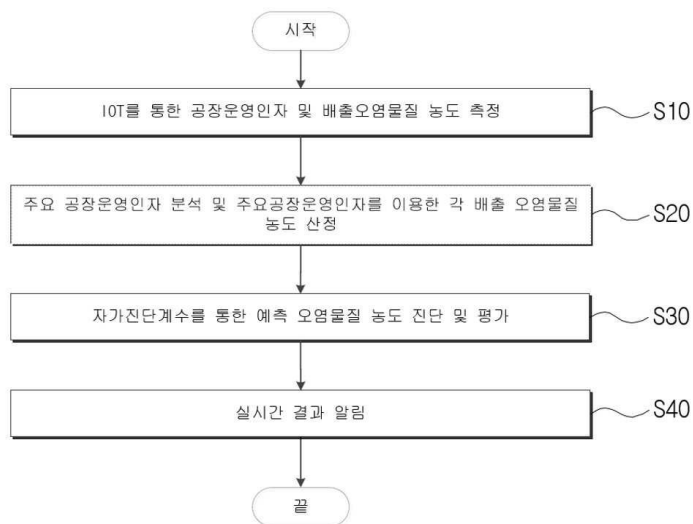
도면1



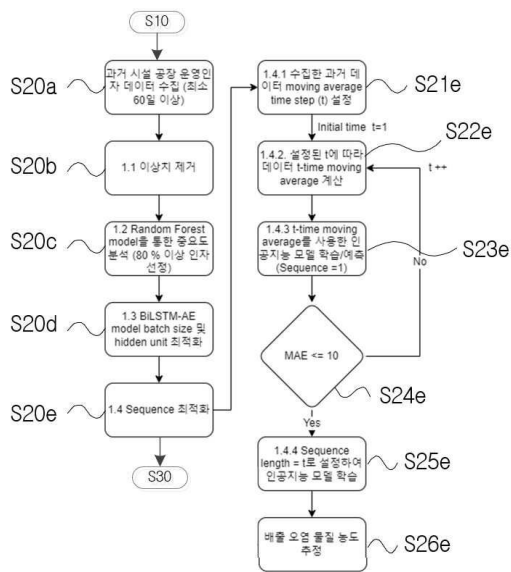
도면2



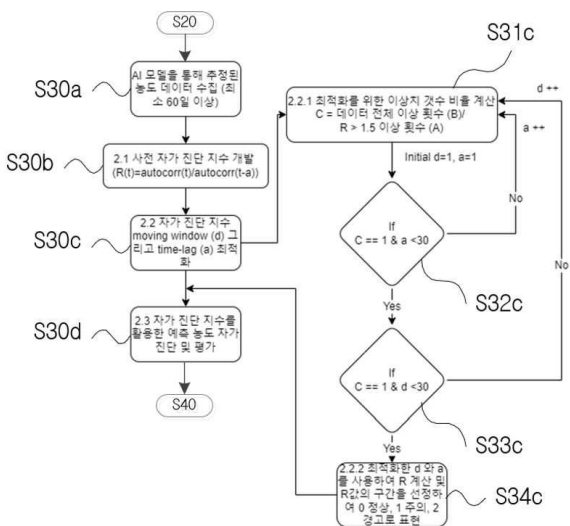
도면3



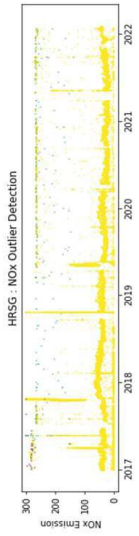
도면4



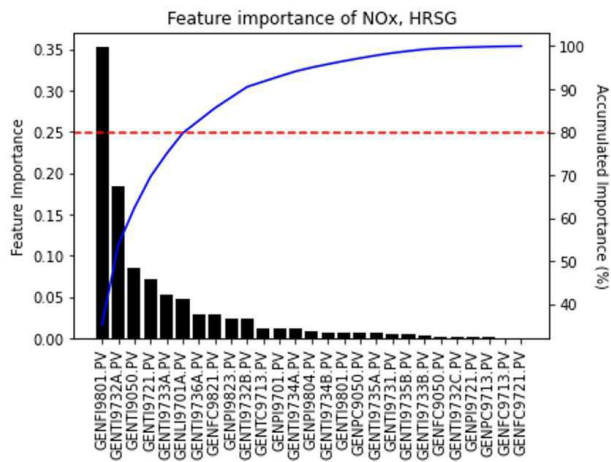
도면5



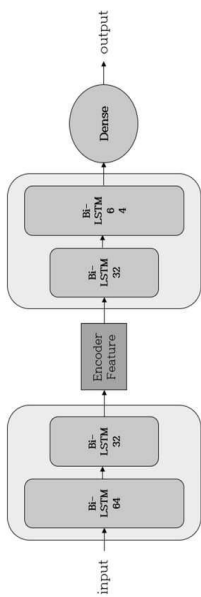
도면6



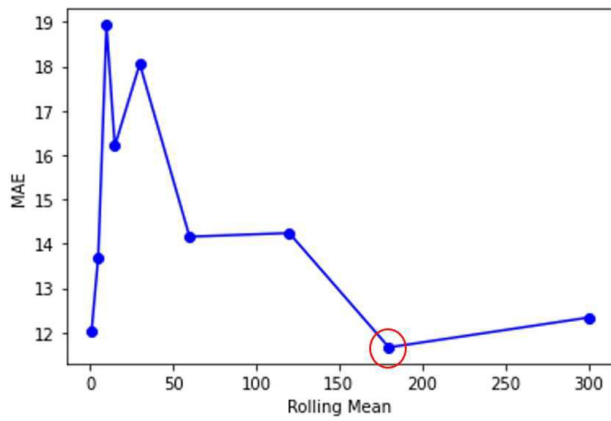
도면7



도면8



도면9



도면10

