



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0040651
(43) 공개일자 2022년03월31일

(51) 국제특허분류(Int. Cl.)

G01N 21/65 (2006.01) G01J 3/44 (2006.01)
G01N 21/59 (2006.01) G06N 20/00 (2019.01)
G06N 3/08 (2006.01)

(52) CPC특허분류

G01N 21/65 (2013.01)
G01J 3/44 (2013.01)

(21) 출원번호 10-2020-0123589

(22) 출원일자 2020년09월24일

심사청구일자 2021년07월02일

(71) 출원인

서강대학교산학협력단

서울특별시 마포구 백범로 35 (신수동, 서강대학교)

(72) 발명자

김동철

서울특별시 마포구 백범로 35 서강대학교 AS관 701호

강태욱

서울특별시 마포구 백범로 35 서강대학교 R611호

곽대진

서울특별시 마포구 백범로 35 서강대학교 AS701호

(74) 대리인

특허법인 천지

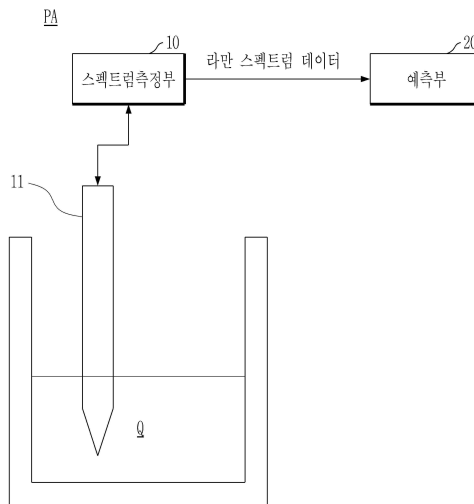
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 라만 분광 신호를 기초로 수용액의 용존 가스 농도를 예측하기 위한 장치 및 이를 위한 방법

(57) 요약

용존 가스 농도를 예측하기 위한 장치가 제공된다. 상기 장치는 수용액에 대한 라만 스펙트럼 데이터가 입력되면, 라만 스펙트럼 데이터로부터 가스의 종류에 상응하는 파장의 피크 데이터를 추출하는 피크추출부와, 상기 피크 데이터를 기초로 예측모델을 통해 상기 수용액에 포함된 가스 농도의 예측값을 산출하고, 산출된 예측값으로부터 가스 농도를 도출하여 출력하는 농도예측부를 포함한다.

대표도 - 도1



(52) CPC특허분류

G01N 21/5907 (2013.01)

G06N 20/00 (2021.08)

G06N 3/08 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711104218
과제번호	2018M3D3A1A01055759
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	기후변화대응기술개발(R&D)
연구과제명	표면증강라만분광법을 이용한 비평형 용존 C1 가스 및 대사산물 실시간 모니터링 기
술 개발	
기 여 율	7/10
과제수행기관명	서강대학교
연구기간	2020.01.01 ~ 2020.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711107800
과제번호	2020R1A2C2010986
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	개인기초연구(과기정통부)(R&D)
연구과제명	나노복합체의 복합 다중 자극에 의한 피로 및 부식 내구성 예측 및 설계 시스템 개
발	
기 여 율	3/10
과제수행기관명	서강대학교
연구기간	2020.03.01 ~ 2021.02.28

명세서

청구범위

청구항 1

라만 분광 신호를 기초로 수용액의 용존 가스 농도를 예측하기 위한 장치에 있어서,

수용액에 대한 라만 스펙트럼 데이터가 입력되면, 라만 스펙트럼 데이터로부터 가스의 종류에 상응하는 파장의 피크 데이터를 추출하는 피크추출부; 및

상기 피크 데이터를 기초로 예측모델을 통해 상기 수용액에 포함된 가스 농도의 예측값을 산출하고, 산출된 예측값으로부터 가스 농도를 도출하여 출력하는 농도예측부;

를 포함하는 것을 특징으로 하는

용존 가스 농도를 예측하기 위한 장치.

청구항 2

제1항에 있어서,

실험 데이터인 라만 스펙트럼 데이터로부터 가스의 종류에 상응하는 파장의 학습용 피크 데이터와 상기 학습용 피크 데이터에 대응하는 가스 농도를 추출하여 학습 데이터를 마련하고,

상기 학습 데이터 중 일부를 검사 데이터로 설정하고,

상기 학습 데이터를 이용하여 수용액에 포함된 가스 농도의 예측값을 산출하도록 예측모델을 학습시키고,

상기 검사 데이터를 이용하여 상기 학습 결과를 검사하는 모델생성부;

를 더 포함하는 것을 특징으로 하는

용존 가스 농도를 예측하기 위한 장치.

청구항 3

제2항에 있어서,

상기 모델생성부는

상기 학습 데이터 및 상기 검사 데이터 각각의 정확도 양자 모두가 기 설정된 조건을 만족할 때까지 상기 예측모델의 은닉계층 및 은닉노드의 수를 조절하면서, 상기 학습 데이터를 이용한 상기 예측모델에 대한 학습과, 상기 검사 데이터를 이용한 상기 예측모델에 대한 학습 성과의 검사를 반복하는 것을 특징으로 하는

용존 가스 농도를 예측하기 위한 장치.

청구항 4

제2항에 있어서,

상기 모델생성부는

상기 학습용 피크 데이터에 대응하는 가스 농도를 기 설정된 경계값을 기준으로 원핫인코딩 벡터로 변환하여 경계 레이블로 설정하고,

상기 학습용 피크 데이터를 예측모델에 입력하고,

상기 예측모델이 상기 학습용 피크 데이터에 대해 복수의 계층 간 가중치가 적용되는 복수의 연산을 수행하여 예측값을 산출하면,

경계손실함수

$$L_{borderselected} = \sum_i v_i (-\log(O_i))$$

에 따라 상기 예측모델의 출력인 예측값과 경계 레이블의 차이인 경계손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 경계 최적화를 수행하고,

상기 Lborderselected는 경계 손실함수이고,

상기 O_i 는 상기 예측모델의 예측값이고,

상기 v_i 는 상기 출력값에 대응하는 경계 레이블이고,

상기 i 는 상기 예측모델의 출력층의 출력노드에 대응하는 인덱스인 것을 특징으로 하는

용존 가스 농도를 예측하기 위한 장치.

청구항 5

제4항에 있어서,

상기 모델생성부는

기 설정된 경계값을 기준으로 학습용 피크 데이터에 대응하는 가스 농도를 원핫인코딩 벡터로 변환하여 상기 경계레이블로 설정하고,

상기 학습용 피크 데이터의 은닉벡터에 대응하는 기준벡터를 증강 레이블로 설정하고,

상기 학습용 피크 데이터를 예측모델에 입력하고,

상기 예측모델이 상기 학습용 피크 데이터에 대해 복수의 계층 간 가중치가 적용되는 복수의 연산을 수행하여 마지막 은닉계층의 복수의 은닉노드의 출력값과, 출력층의 출력인 예측값을 산출하면,

경계증강손실함수

$$L_{boaderenhanced} = \frac{1}{2} \sum_i v_i (-\log(O_i)) + \frac{1}{2} \frac{1}{N} \sum_i \sum_j |r_{ij} - h_{ij}|^2$$

에 따라

상기 예측모델의 출력인 예측값과 경계 레이블의 차이인 경계 손실 및 상기 예측모델의 마지막 은닉계층의 복수의 은닉노드의 출력값과 증강 레이블과의 차이를 나타내는 증강 손실을 포함하는 경계증강 손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 경계증강 최적화를 수행하고,

상기 Lboaderenhanced는 경계증강 손실함수이고,

상기 O_i 는 상기 예측모델의 출력층의 출력인 예측값이고,

상기 v_i 는 상기 예측값에 대응하는 경계 레이블이고,

상기 h_{ij} 는 상기 예측모델의 마지막 은닉계층의 복수의 은닉노드의 출력값이고,

상기 r_{ij} 는 상기 복수의 은닉노드의 출력값에 대응하는 증강 레이블이고,

상기 i 는 상기 예측모델의 출력층의 출력노드에 대응하는 인덱스이고,

상기 j 는 상기 예측모델의 마지막 은닉계층의 복수의 은닉노드에 대응하는 인덱스인 것

을 특징으로 하는

용존 가스 농도를 예측하기 위한 장치.

청구항 6

제5항에 있어서,

상기 모델생성부는

상기 학습용 피크 데이터에 대응하는 가스 농도를 수치 레이블로 설정하고,

상기 학습용 피크 데이터를 예측모델에 입력하고,

상기 예측모델이 상기 학습용 피크 데이터에 대해 복수의 계층 간 가중치가 적용되는 복수의 연산을 수행하여 예측값을 산출하면,

상기 예측모델의 상기 예측값과 수치 레이블의 차이인 수치 손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 수치 최적화를 수행하는 것을 특징으로 하는

용존 가스 농도를 예측하기 위한 장치.

청구항 7

제6항에 있어서,

상기 모델생성부는

수치손실함수

$$L_{value} = \frac{1}{N} \sum_{i=1}^N |O_i - C_i|^2 + \frac{1}{N} \sum_{i=1}^N (1 - SSIM(O_i, C_i))$$

에 따라 상기 예측모델의 출력값과 수치 레이블의 차이인 수치 손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 수치 최적화를 수행하며,

상기 Lvalue는 수치손실함수이고,

상기 O_i 는 상기 예측모델의 출력인 예측값이고,

상기 C_i 는 상기 예측값에 대응하는 수치 레이블이고,

상기 i 는 상기 예측모델의 출력층의 출력노드에 대응하는 인덱스인 것을 특징으로 하는

용존 가스 농도를 예측하기 위한 장치.

청구항 8

제1항에 있어서,

상기 가스가 CO이면, 상기 파장은 1948 cm^{-1} 및 2063 cm^{-1} 이고,

상기 가스가 아세트산염(Acetate)이면 상기 파장은 928 cm^{-1} 이고,

상기 가스가 낙산염(Butyrate)이면, 상기 파장은 877 cm^{-1} 인 것을 특징으로 하는

용존 가스 농도를 예측하기 위한 장치.

청구항 9

라만 분광 신호를 기초로 수용액의 용존 가스 농도를 예측하기 위한 장치에 있어서,

실험 데이터인 라만 스펙트럼 데이터로부터 가스의 종류에 상응하는 파장의 학습용 피크 데이터와 상기 학습용 피크 데이터에 대응하는 가스 농도를 추출하여 학습 데이터를 마련하고,

상기 학습 데이터를 이용하여 수용액에 포함된 가스 농도의 예측값을 산출하도록 인공지능망 알고리즘인 예측모델을 학습시키는 모델생성부;

를 포함하는 것을 특징으로 하는

용존 가스 농도를 예측하기 위한 장치.

청구항 10

제9항에 있어서,

상기 모델생성부는

상기 학습용 피크 데이터에 대응하는 가스 농도를 수치 레이블로 설정하고,

상기 학습용 피크 데이터를 예측모델에 입력하고,

상기 예측모델이 상기 학습용 피크 데이터에 대해 복수의 계층 간 가중치가 적용되는 복수의 연산을 수행하여 예측값을 산출하면,

상기 예측모델의 상기 예측값과 수치 레이블의 차이인 수치 손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 수치 최적화를 수행하되,

수치손실함수

$$L_{value} = \frac{1}{N} \sum_{i=1}^N |O_i - C_i|^2 + \frac{1}{N} \sum_{i=1}^N (1 - SSIM(O_i, C_i))$$

에 따라 상기 예측모델의 출력값과 수치 레이블의 차이인 수치 손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 수치 최적화를 수행하며,

상기 Lvaule는 수치 손실함수이고,

상기 O_i 는 상기 예측모델의 출력인 예측값이고,

상기 C_i 는 상기 예측값에 대응하는 수치 레이블이고,

상기 i 는 상기 예측모델의 출력층의 출력노드에 대응하는 인덱스인 것을 특징으로 하는

용존 가스 농도를 예측하기 위한 장치.

청구항 11

제9항에 있어서,

상기 모델생성부는

학습용 피크 데이터에 대응하는 가스 농도를 기 설정된 경계값을 기준으로 원핫인코딩 벡터로 변환하여 경계 레이블로 설정하고,

상기 학습용 피크 데이터를 예측모델에 입력하고,

상기 예측모델이 상기 학습용 피크 데이터에 대해 복수의 계층 간 가중치가 적용되는 복수의 연산을 수행하여 예측값을 산출하면,

경계손실함수

$$L_{borderselected} = \sum_i v_i (-\log(O_i))$$

에 따라 상기 예측모델의 출력인 예측값과 경계 레이블의 차이인 경계손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 경계 최적화를 수행한 후,

기 설정된 경계값을 기준으로 학습용 피크 데이터에 대응하는 가스 농도를 원핫인코딩 벡터로 변환하여 상기 경계레이블로 설정하고,

상기 학습용 피크 데이터의 은닉벡터에 대응하는 기준벡터를 증강 레이블로 설정하고,

상기 학습용 피크 데이터를 예측모델에 입력하고,

상기 예측모델이 상기 학습용 피크 데이터에 대해 복수의 계층 간 가중치가 적용되는 복수의 연산을 수행하여 마지막 은닉계층의 복수의 은닉노드의 출력값과, 출력층의 출력인 예측값을 산출하면,

경계증강손실함수

$$L_{boaderenhanced} = \frac{1}{2} \sum_i v_i (-\log(O_i)) + \frac{1}{2} \frac{1}{N} \sum_i \sum_j |r_{ij} - h_{ij}|^2$$

에 따라

상기 예측모델의 출력인 예측값과 경계 레이블의 차이인 경계 손실 및 상기 예측모델의 마지막 은닉계층의 복수의 은닉노드의 출력값과 증강 레이블과의 차이를 나타내는 증강 손실을 포함하는 경계증강 손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 경계증강 최적화를 수행하고,

상기 Lboaderselected는 경계 손실함수이고,

상기 Lboaderenhanced는 경계증강 손실함수이고,

상기 O_i 는 상기 예측모델의 출력층의 출력인 예측값이고,

상기 v_i 는 상기 예측값에 대응하는 경계 레이블이고,

상기 h_{ij} 는 상기 예측모델의 마지막 은닉계층의 복수의 은닉노드의 출력값이고,

상기 r_{ij} 는 상기 복수의 은닉노드의 출력값에 대응하는 증강 레이블이고,

상기 i 는 상기 예측모델의 출력층의 출력노드에 대응하는 인덱스이고,

상기 j 는 상기 예측모델의 마지막 은닉계층의 복수의 은닉노드에 대응하는 인덱스인 것

을 특징으로 하는

용존 가스 농도를 예측하기 위한 장치.

청구항 12

제10항에 있어서,

수용액에 내에 프로브를 통해 빛을 발광하고, 발광된 빛이 산란되어 나오는 것을 측정하여 라만 스펙트럼 데이터를 취득하고, 취득한 라만 스펙트럼 데이터를 제공하는 스펙트럼측정부;

상기 라만 스펙트럼 데이터로부터 가스의 종류에 상응하는 파장의 피크 데이터를 추출하는 피크추출부; 및

상기 피크 데이터를 기초로 상기 예측모델을 통해 상기 수용액에 포함된 가스 농도의 예측값을 산출하고, 산출된 예측값으로부터 가스 농도를 도출하여 출력하는 농도예측부;

를 더 포함하는 것을 특징으로 하는

용존 가스 농도를 예측하기 위한 장치.

청구항 13

라만 분광 신호를 기초로 수용액의 용존 가스 농도를 예측하기 위한 방법에 있어서,

스펙트럼측정부가 수용액에 내에 프로브를 통해 빛을 발광하고, 발광된 빛이 산란되어 나오는 것을 측정하여 라만 스펙트럼 데이터를 취득하는 단계;

피크추출부가 라만 스펙트럼 데이터로부터 가스의 종류에 상응하는 파장의 피크 데이터를 추출하는 단계;

농도예측부가 상기 피크 데이터를 기초로 예측모델을 통해 상기 수용액에 포함된 가스 농도의 예측값을 산출하는 단계; 및

상기 농도예측부가 산출된 예측값으로부터 가스 농도를 도출하여 출력하는 단계;

를 포함하는 것을 특징으로 하는

용존 가스 농도를 예측하기 위한 방법.

청구항 14

제13항에 있어서,

상기 라만 스펙트럼 데이터를 취득하는 단계 전,

상기 모델생성부가 실험 데이터로부터 학습 데이터 및 검사 데이터를 추출하는 단계;

상기 모델생성부가 상기 예측모델의 은닉계층 및 은닉노드의 수를 설정하는 단계;

상기 모델생성부가 상기 학습 데이터를 이용하여 설정된 수의 은닉계층 및 은닉노드를 가지는 예측모델에 대한 학습을 수행하고, 상기 검사 데이터를 이용하여 예측모델(PM)의 학습 성과를 검사하고, 상기 모델생성부가 상기 학습 데이터의 정확도 및 상기 검사 데이터의 정확도를 산출하는 단계;

상기 모델생성부가 상기 학습 데이터 정확도가 상기 검사 데이터의 정확도 보다 높으면서 상기 학습 데이터 정확도 및 상기 검사 데이터의 정확도 양자 모두가 임계치 이상인 조건을 만족하는지 여부를 판별하는 단계; 및

상기 판별 결과, 상기 조건을 만족하면, 상기 모델생성부가 현재 설정에 따라 은닉계층 및 은닉노드의 수를 결정하는 단계;

를 포함하는 것을 특징으로 하는

용존 가스 농도를 예측하기 위한 방법.

청구항 15

제14항에 있어서,

상기 조건을 만족하는지 여부를 판별하는 단계후,

상기 판별 결과, 상기 조건을 만족하지 않으면,

상기 모델생성부가 은닉계층 및 은닉노드의 수를 재설정한 후, 상기 정확도를 산출하는 단계 및 상기 조건을 만족하는지 여부를 판별하는 단계를 반복하는 단계;

더 포함하는 것을 특징으로 하는

용존 가스 농도를 예측하기 위한 방법.

청구항 16

제13항에 있어서,

상기 라만 스펙트럼 데이터를 취득하는 단계 전,

상기 모델생성부가 실험 데이터인 라만 스펙트럼 데이터로부터 가스의 종류에 상응하는 파장의 학습용 피크 데이터와 상기 학습용 피크 데이터에 대응하는 가스 농도를 추출하여 학습 데이터를 마련하는 단계;

상기 모델생성부가 상기 학습용 피크 데이터에 대응하는 가스 농도를 수치 레이블로 설정하는 단계;

상기 모델생성부가 상기 학습용 피크 데이터를 예측모델에 입력하는 단계;

상기 예측모델이 상기 학습용 피크 데이터에 대해 복수의 계층 간 가중치가 적용되는 복수의 연산을 수행하여 예측값을 산출하는 단계;

상기 모델생성부가 상기 예측모델의 상기 예측값과 수치 레이블의 차이인 수치 손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 수치 최적화를 수행하는 단계;

를 더 포함하는 것을 특징으로 하는

용존 가스 농도를 예측하기 위한 방법.

청구항 17

제16항에 있어서,

상기 수치 최적화를 수행하는 단계는

상기 모델생성부가

수치손실함수

$$L_{value} = \frac{1}{N} \sum_{i=1}^N |O_i - C_i|^2 + \frac{1}{N} \sum_{i=1}^N (1 - SSIM(O_i, C_i))$$

에 따라 상기 예측모델의 출력값과 수치 레이블의 차이인 수치 손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 수치 최적화를 수행하며,

상기 Lvalue는 수치손실함수이고,

상기 O_i 는 상기 예측모델의 출력인 예측값이고,

상기 C_i 는 상기 예측값에 대응하는 수치 레이블이고,

상기 i 는 상기 예측모델의 출력층의 출력노드에 대응하는 인덱스인 것을 특징으로 하는

용존 가스 농도를 예측하기 위한 방법.

청구항 18

제13항에 있어서,

상기 라만 스펙트럼 데이터를 취득하는 단계 전,

상기 모델생성부가 상기 학습용 피크 데이터에 대응하는 가스 농도를 기 설정된 경계값을 기준으로 원핫인코딩 벡터로 변환하여 경계 레이블로 설정하는 단계;

상기 모델생성부가 상기 학습용 피크 데이터를 예측모델에 입력하는 단계;

상기 예측모델이 상기 학습용 피크 데이터에 대해 복수의 계층 간 가중치가 적용되는 복수의 연산을 수행하여 예측값을 산출하는 단계;

상기 모델생성부가

경계손실함수

$$L_{borderselected} = \sum_i v_i (-\log(O_i))$$

에 따라 상기 예측모델의 출력인 예측값과 경계 레이블의 차이인 경계손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 경계 최적화를 수행하는 단계;

를 더 포함하며,

상기 Lborderselected는 경계 손실함수이고,

상기 O_i 는 상기 예측모델의 출력인 예측값이고,

상기 v_i 는 상기 예측값에 대응하는 경계 레이블이고,

상기 i 는 상기 예측모델의 출력층의 출력노드에 대응하는 인덱스인 것을 특징으로 하는

용존 가스 농도를 예측하기 위한 방법.

청구항 19

제18항에 있어서,

상기 경계 최적화를 수행하는 단계 후,

상기 모델생성부가 기 설정된 경계값을 기준으로 학습용 피크 데이터에 대응하는 가스 농도를 원핫인코딩 벡터로 변환하여 상기 경계레이블로 설정하고, 상기 학습용 피크 데이터의 은닉벡터에 대응하는 기준벡터를 증강 레이블로 설정하는 단계;

상기 모델생성부가 상기 학습용 피크 데이터를 예측모델에 입력하는 단계;

상기 예측모델이 상기 학습용 피크 데이터에 대해 복수의 계층 간 가중치가 적용되는 복수의 연산을 수행하여 마지막 은닉계층의 복수의 은닉노드의 출력값과, 출력층의 출력인 예측값을 산출하는 단계;

상기 모델생성부가 경계증강손실함수

$$L_{\text{boaderenhanced}} = \frac{1}{2} \sum_i v_i (-\log(O_i)) + \frac{1}{2} \frac{1}{N} \sum_i \sum_j |r_{ij} - h_{ij}|^2$$

에 따라 상기 예측모델의 출력인 예측값과 경계 레이블의 차이인 경계 손실 및 상기 예측모델의 마지막 은닉계층의 복수의 은닉노드의 출력값과 증강 레이블과의 차이를 나타내는 증강 손실을 포함하는 경계증강 손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 경계증강 최적화를 수행하는 단계;

를 더 포함하며,

상기 Lboaderenhanced는 경계증강 손실함수이고,

상기 O_i 는 상기 예측모델의 출력인 예측값이고,

상기 v_i 는 상기 예측값에 대응하는 경계 레이블이고,

상기 i 는 상기 예측모델의 출력층의 출력노드에 대응하는 인덱스이고,

상기 h_{ij} 는 상기 예측모델의 마지막 은닉계층의 복수의 은닉노드의 출력값이고,

상기 r_{ij} 는 상기 복수의 은닉노드의 출력값에 대응하는 증강 레이블이고,

상기 j 는 상기 예측모델의 마지막 은닉계층의 복수의 은닉노드에 대응하는 인덱스인 것을 특징으로 하는

용존 가스 농도를 예측하기 위한 방법.

청구항 20

제13항에 있어서,

상기 가스가 CO이면, 상기 파장은 1948 cm^{-1} 및 2063 cm^{-1} 이고,

상기 가스가 아세트산염(Acetate)이면 상기 파장은 928 cm^{-1} 이고,

상기 가스가 낙산염(Butyrate)이면, 상기 파장은 877 cm^{-1} 인 것을 특징으로 하는

용존 가스 농도를 예측하기 위한 방법.

발명의 설명

기술 분야

[0001] 본 발명은 용존 가스 농도를 예측 기술에 관한 것으로, 보다 상세하게는, 라만 분광 신호를 기초로 수용액의 용존 가스 농도를 예측하기 위한 장치 및 이를 위한 방법에 관한 것이다.

배경 기술

[0002] 라만 산란법은 고감도 검출 기술이라는 장점으로 인해 기존에 측정할 수 없었던 수용액 상의 소량의 가스(gas) 검출을 가능하게 한다. 하지만 가스의 농도와 라만 산란법의 라만 스펙트럼의 관계는 선형적이지 않고, 또한 그 관계를 쉽게 도출할 수 없다. 그러므로 기존에는 실험을 통해 도출한 라만 스펙트럼 데이터와 용존 가스의 농도 데이터를 1:1로 비교하여 그 관계를 도출하였다. 하지만 이러한 방법의 경우 수많은 실험 데이터베이스가 필요

하다.

선행기술문헌

특허문헌

- [0003] (특허문헌 0001) 한국공개특허 제2001-0022918호 2001년 03월 26일 공개 (명칭: 실시간 가스 분석 방법 및 장치)

발명의 내용

해결하려는 과제

- [0004] 본 발명의 목적은 라만 분광 신호를 기초로 수용액의 용존 가스 농도를 예측하기 위한 장치 및 이를 위한 방법을 제공함에 있다.

과제의 해결 수단

- [0005] 상술한 바와 같은 목적을 달성하기 위한 본 발명의 바람직한 실시예에 따른 라만 분광 신호를 기초로 수용액의 용존 가스 농도를 예측하기 위한 장치는 수용액에 대한 라만 스펙트럼 데이터가 입력되면, 라만 스펙트럼 데이터로부터 가스의 종류에 상응하는 파장의 피크 데이터를 추출하는 피크추출부와, 상기 피크 데이터를 기초로 예측모델을 통해 상기 수용액에 포함된 가스 농도의 예측값을 산출하고, 산출된 예측값으로부터 가스 농도를 도출하여 출력하는 농도예측부를 포함한다.
- [0006] 상기 장치는 실험 데이터인 라만 스펙트럼 데이터로부터 가스의 종류에 상응하는 파장의 학습용 피크 데이터와 상기 학습용 피크 데이터에 대응하는 가스 농도를 추출하여 학습 데이터를 마련하고, 상기 학습 데이터 중 일부를 검사 데이터로 설정하고, 상기 학습 데이터를 이용하여 수용액에 포함된 가스 농도의 예측값을 산출하도록 예측모델을 학습시키고, 상기 검사 데이터를 이용하여 상기 학습 결과를 검사하는 모델생성부를 더 포함한다.
- [0007] 상기 모델생성부는 상기 학습 데이터 및 상기 검사 데이터 각각의 정확도 양자 모두가 기 설정된 조건을 만족할 때까지 상기 예측모델의 은닉계층 및 은닉노드의 수를 조절하면서, 상기 학습 데이터를 이용한 상기 예측모델에 대한 학습과, 상기 검사 데이터를 이용한 상기 예측모델에 대한 학습 성과의 검사를 반복하는 것을 특징으로 한다.
- [0008] 상기 모델생성부는 상기 학습용 피크 데이터에 대응하는 가스 농도를 기 설정된 경계값을 기준으로 원핫인코딩 벡터로 변환하여 경계 레이블로 설정하고, 상기 학습용 피크 데이터를 예측모델에 입력하고, 상기 예측모델이 상기 학습용 피크 데이터에 대해 복수의 계층 간 가중치가 적용되는 복수의 연산을 수행하여 예측값을

$$L_{borderselected} = \sum_i v_i (-\log(O_i))$$

산출하면, 경계손실함수에 따라 상기 예측모델의 출력인 예측값과 경계 레이블의 차이인 경계손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 경계 최적화를 수행하고, 상기 $L_{borderselected}$ 는 경계 손실함수이고, 상기 O_i 는 상기 예측모델의 예측값이고, 상기 v_i 는 상기 출력값에 대응하는 경계 레이블이고, 상기 i 는 상기 예측모델의 출력층의 출력 노드에 대응하는 인덱스인 것을 특징으로 한다.

- [0009] 상기 모델생성부는 기 설정된 경계값을 기준으로 학습용 피크 데이터에 대응하는 가스 농도를 원핫인코딩 벡터로 변환하여 상기 경계레이블로 설정하고, 상기 학습용 피크 데이터의 은닉벡터에 대응하는 기준벡터를 증강 레이블로 설정하고, 상기 학습용 피크 데이터를 예측모델에 입력하고, 상기 예측모델이 상기 학습용 피크 데이터에 대해 복수의 계층 간 가중치가 적용되는 복수의 연산을 수행하여 마지막 은닉계층의 복수의 은닉노드의 출력값과, 출력층의 출력인 예측값을 산출하면, 경계증강손실함수

$$L_{boaderenhanced} = \frac{1}{2} \sum_i v_i (-\log(O_i)) + \frac{1}{2} \frac{1}{N} \sum_i \sum_j |r_{ij} - h_{ij}|^2$$

에 따라 상기 예측모델

의 출력값인 예측값과 경계 레이블의 차이인 경계 손실 및 상기 예측모델의 마지막 은닉계층의 복수의 은닉노드의 출력값과 증강 레이블과의 차이를 나타내는 증강 손실을 포함하는 경계증강 손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 경계증강 최적화를 수행하고, 상기 Lboaderehanced는 경계증강 손실함수이고, 상기 O_i 는 상기 예측모델의 출력층의 출력인 예측값이고, 상기 v_i 는 상기 예측값에 대응하는 경계 레이블이고, 상기 h_{ij} 는 상기 예측모델의 마지막 은닉계층의 복수의 은닉노드의 출력값이고, 상기 r_{ij} 는 상기 복수의 은닉노드의 출력값에 대응하는 증강 레이블이고, 상기 i 는 상기 예측모델의 출력층의 출력노드에 대응하는 인덱스이고, 상기 j 는 상기 예측모델의 마지막 은닉계층의 복수의 은닉노드에 대응하는 인덱스인 것을 특징으로 한다.

[0010] 상기 모델생성부는 상기 학습용 피크 데이터에 대응하는 가스 농도를 수치 레이블로 설정하고, 상기 학습용 피크 데이터를 예측모델에 입력하고, 상기 예측모델이 상기 학습용 피크 데이터에 대해 복수의 계층 간 가중치가 적용되는 복수의 연산을 수행하여 예측값을 산출하면, 상기 예측모델의 상기 예측값과 수치 레이블의 차이인 수치 손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 수치 최적화를 수행하는 것을 특징으로 한다.

[0011] 상기 모델생성부는 수치손실함수

$$L_{value} = \frac{1}{N} \sum_{i=1}^N |O_i - C_i|^2 + \frac{1}{N} \sum_{i=1}^N (1 - SSIM(O_i, C_i))$$

에 따라 상기 예측모델의 출력값과 수치 레이블의 차이인 수치 손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 수치 최적화를 수행하며, 상기 L_{value} 는 수치손실함수이고, 상기 O_i 는 상기 예측모델의 출력인 예측값이고, 상기 C_i 는 상기 예측값에 대응하는 수치 레이블이고, 상기 i 는 상기 예측모델의 출력층의 출력노드에 대응하는 인덱스인 것을 특징으로 한다.

[0012] 상기 가스가 CO이면, 상기 파장은 1948 cm^{-1} 및 2063 cm^{-1} 이고, 상기 가스가 아세트산염(Acetate)이면 상기 파장은 928 cm^{-1} 이고, 상기 가스가 낙산염(Butyrate)이면, 상기 파장은 877 cm^{-1} 인 것을 특징으로 한다.

[0013] 상기 장치는 수용액에 내에 프로브를 통해 빛을 발광하고, 발광된 빛이 산란되어 나오는 것을 측정하여 라만 스펙트럼 데이터를 취득하고, 취득한 라만 스펙트럼 데이터를 제공하는 스펙트럼측정부를 더 포함한다.

[0014] 상술한 바와 같은 목적을 달성하기 위한 본 발명의 바람직한 실시예에 따른 라만 분광 신호를 기초로 수용액의 용존 가스 농도를 예측하기 위한 장치는 실험 데이터인 라만 스펙트럼 데이터로부터 가스의 종류에 상응하는 파장의 학습용 피크 데이터와 상기 학습용 피크 데이터에 대응하는 가스 농도를 추출하여 학습 데이터를 마련하고, 상기 학습 데이터를 이용하여 수용액에 포함된 가스 농도의 예측값을 산출하도록 인공지능망 알고리즘인 예측모델을 학습시키는 모델생성부를 포함한다.

[0015] 상기 모델생성부는 상기 학습용 피크 데이터에 대응하는 가스 농도를 수치 레이블로 설정하고, 상기 학습용 피크 데이터를 예측모델에 입력하고, 상기 예측모델이 상기 학습용 피크 데이터에 대해 복수의 계층 간 가중치가 적용되는 복수의 연산을 수행하여 예측값을 산출하면, 상기 예측모델의 상기 예측값과 수치 레이블의 차이인 수치 손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 수치 최적화를 수행하는 것을 특징으로 한다.

[0016] 상기 모델생성부는 수치손실함수

$$L_{value} = \frac{1}{N} \sum_{i=1}^N |O_i - C_i|^2 + \frac{1}{N} \sum_{i=1}^N (1 - SSIM(O_i, C_i))$$

에 따라 상기 예측모델의 출력값과 수치 레이블의 차이인 수치 손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 수치 최적화를 수행하며, 상기 L_{value} 는 수치 손실함수이고, 상기 O_i 는 상기 예측모델의 출력인 예측값이고, 상기 C_i 는 상기 예측값에 대응하는 수치 레이블이고, 상기 i 는 상기 예측모델의 출력층의 출력노드에 대응하는 인덱스인 것을 특징으로 한다.

[0017] 상기 모델생성부는 학습용 피크 데이터에 대응하는 가스 농도를 기 설정된 경계값을 기준으로 원핫인코딩 벡터로 변환하여 경계 레이블로 설정하고, 상기 학습용 피크 데이터를 예측모델에 입력하고, 상기 예측모델이 상기 학습용 피크 데이터에 대해 복수의 계층 간 가중치가 적용되는 복수의 연산을 수행하여 예측값을 산출하면, 경

$$L_{borderselected} = \sum_i v_i (-\log(O_i))$$

계손실함수에 따라 상기 예측모델의 출력인 예측값과 경계 레이블의 차이인 경계손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 경계 최적화를 수행하고, 상기 Lborderselected는 경계 손실함수이고, 상기 O_i 는 상기 예측모델의 예측값이고, 상기 v_i 는 상기 출력값에 대응하는 경계 레이블이고, 상기 i 는 상기 예측모델의 출력층의 출력노드에 대응하는 인덱스인 것을 특징으로 한다.

[0018] 상기 모델생성부는 기 설정된 경계값을 기준으로 학습용 피크 데이터에 대응하는 가스 농도를 원핫인코딩 벡터로 변환하여 상기 경계레이블로 설정하고, 상기 학습용 피크 데이터의 은닉벡터에 대응하는 기준벡터를 증강 레이블로 설정하고, 상기 학습용 피크 데이터를 예측모델에 입력하고, 상기 예측모델이 상기 학습용 피크 데이터에 대해 복수의 계층 간 가중치가 적용되는 복수의 연산을 수행하여 마지막 은닉계층의 복수의 은닉노드의 출력값과, 출력층의 출력인 예측값을 산출하면, 경계증강손실함수

$$L_{boaderenhanced} = \frac{1}{2} \sum_i v_i (-\log(O_i)) + \frac{1}{2} \frac{1}{N} \sum_i \sum_j |r_{ij} - h_{ij}|^2$$

에 따라 상기 예측모델의 출력인 예측값과 경계 레이블의 차이인 경계 손실 및 상기 예측모델의 마지막 은닉계층의 복수의 은닉노드의 출력값과 증강 레이블과의 차이를 나타내는 증강 손실을 포함하는 경계증강 손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 경계증강 최적화를 수행하고, 상기 Lboaderenhanced는 경계증강 손실함수이고, 상기 O_i 는 상기 예측모델의 출력층의 출력인 예측값이고, 상기 v_i 는 상기 예측값에 대응하는 경계 레이블이고, 상기 h_{ij} 는 상기 예측모델의 마지막 은닉계층의 복수의 은닉노드의 출력값이고, 상기 r_{ij} 는 상기 복수의 은닉노드의 출력값에 대응하는 증강 레이블이고, 상기 i 는 상기 예측모델의 출력층의 출력노드에 대응하는 인덱스이고, 상기 j 는 상기 예측모델의 마지막 은닉계층의 복수의 은닉노드에 대응하는 인덱스인 것을 특징으로 한다.

[0019] 상기 모델생성부는 상기 학습 데이터 및 상기 검사 데이터 각각의 정확도 양자 모두가 기 설정된 조건을 만족할 때까지 상기 예측모델의 은닉계층 및 은닉노드의 수를 조절하면서, 상기 학습 데이터를 이용한 상기 예측모델에 대한 학습과, 상기 검사 데이터를 이용한 상기 예측모델에 대한 학습 성과의 검사를 반복하는 것을 특징으로 한다.

[0020] 상기 장치는 수용액에 내에 프로브를 통해 빛을 발광하고, 발광된 빛이 산란되어 나오는 것을 측정하여 라만 스펙트럼 데이터를 취득하고, 취득한 라만 스펙트럼 데이터를 제공하는 스펙트럼측정부와, 상기 라만 스펙트럼 데이터로부터 가스의 종류에 상응하는 파장의 피크 데이터를 추출하는 피크추출부와, 상기 피크 데이터를 기초로 상기 예측모델을 통해 상기 수용액에 포함된 가스 농도의 예측값을 산출하고, 산출된 예측값으로부터 가스 농도를 도출하여 출력하는 농도예측부를 더 포함한다.

[0021] 상술한 바와 같은 목적을 달성하기 위한 본 발명의 바람직한 실시예에 따른 라만 분광 신호를 기초로 수용액의 용존 가스 농도를 예측하기 위한 방법은 스펙트럼측정부가 수용액에 내에 프로브를 통해 빛을 발광하고, 발광된 빛이 산란되어 나오는 것을 측정하여 라만 스펙트럼 데이터를 취득하는 단계와, 피크추출부가 라만 스펙트럼 데이터로부터 가스의 종류에 상응하는 파장의 피크 데이터를 추출하는 단계와, 농도예측부가 상기 피크 데이터를 기초로 예측모델을 통해 상기 수용액에 포함된 가스 농도의 예측값을 산출하는 단계와, 상기 농도예측부가 산출된 예측값으로부터 가스 농도를 도출하여 출력하는 단계를 포함한다.

[0022] 상기 방법은 상기 라만 스펙트럼 데이터를 취득하는 단계 전, 상기 모델생성부가 실험 데이터로부터 학습 데이터 및 검사 데이터를 추출하는 단계와, 상기 모델생성부가 상기 예측모델의 은닉계층 및 은닉노드의 수를 설정하는 단계와, 상기 모델생성부가 상기 학습 데이터를 이용하여 설정된 수의 은닉계층 및 은닉노드를 가지는 예측모델에 대한 학습을 수행하고, 상기 검사 데이터를 이용하여 예측모델(PM)의 학습 성과를 검사하고, 상기 모델생성부가 상기 학습 데이터의 정확도 및 상기 검사 데이터의 정확도를 산출하는 단계와, 상기 모델생성부가 상기 학습 데이터 정확도가 상기 검사 데이터의 정확도 보다 높으면서 상기 학습 데이터 정확도 및 상기 검사 데이터의 정확도 양자 모두가 임계치 이상인 조건을 만족하는지 여부를 판별하는 단계와, 상기 판별 결과, 상기 조건을 만족하면, 상기 모델생성부가 현재 설정에 따라 은닉계층 및 은닉노드의 수를 결정하는 단계를 포함한다.

[0023] 또한, 상기 방법은 상기 조건을 만족하는지 여부를 판별하는 단계후, 상기 판별 결과, 상기 조건을 만족하지 않

으면, 상기 모델생성부가 은닉계층 및 은닉노드의 수를 재설정한 후, 상기 정확도를 산출하는 단계 및 상기 조건을 만족하는지 여부를 판별하는 단계를 반복하는 단계를 더 포함한다.

[0024] 상기 방법은 상기 라만 스펙트럼 데이터를 취득하는 단계 전, 상기 모델생성부가 실험 데이터인 라만 스펙트럼 데이터로부터 가스의 종류에 상응하는 파장의 학습용 피크 데이터와 상기 학습용 피크 데이터에 대응하는 가스 농도를 추출하여 학습 데이터를 마련하는 단계와, 상기 모델생성부가 상기 학습용 피크 데이터에 대응하는 가스 농도를 수치 레이블로 설정하는 단계와, 상기 모델생성부가 상기 학습용 피크 데이터를 예측모델에 입력하는 단계와, 상기 예측모델이 상기 학습용 피크 데이터에 대해 복수의 계층 간 가중치가 적용되는 복수의 연산을 수행하여 예측값을 산출하는 단계와, 상기 모델생성부가 상기 예측모델의 상기 예측값과 수치 레이블의 차이인 수치 손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 수치 최적화를 수행하는 단계를 더 포함한다.

[0025] 상기 수치 최적화를 수행하는 단계는 상기 모델생성부가 수치손실함수

$$L_{value} = \frac{1}{N} \sum_{i=1}^N |O_i - C_i|^2 + \frac{1}{N} \sum_{i=1}^N (1 - SSIM(O_i, C_i))$$

에 따라 상기 예측모델의 출력값과 수치 레이블의 차이인 수치 손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 수치 최적화를 수행하며, 상기 Lvalue는 수치손실함수이고, 상기 O_i 는 상기 예측모델의 출력인 예측값이고, 상기 C_i 는 상기 예측값에 대응하는 수치 레이블이고, 상기 i 는 상기 예측모델의 출력층의 출력노드에 대응하는 인덱스인 것을 특징으로 한다.

[0026] 상기 라만 스펙트럼 데이터를 취득하는 단계 전, 상기 모델생성부가 상기 학습용 피크 데이터에 대응하는 가스 농도를 기 설정된 경계값을 기준으로 원핫인코딩 벡터로 변환하여 경계 레이블로 설정하는 단계와, 상기 모델생성부가 상기 학습용 피크 데이터를 예측모델에 입력하는 단계와, 상기 예측모델이 상기 학습용 피크 데이터에 대해 복수의 계층 간 가중치가 적용되는 복수의 연산을 수행하여 예측값을 산출하는 단계와, 상기 모델생성부가

$$L_{borderselected} = \sum_i v_i (-\log(O_i))$$

경계손실함수 에 따라 상기 예측모델의 출력인 예측값과 경계 레이블의 차이인 경계손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 경계 최적화를 수행하는 단계를 더 포함한다. 여기서, 상기 Lborderselected는 경계 손실함수이고, 상기 O_i 는 상기 예측모델의 출력인 예측값이고, 상기 v_i 는 상기 예측값에 대응하는 경계 레이블이고, 상기 i 는 상기 예측모델의 출력층의 출력노드에 대응하는 인덱스이다.

[0027] 상기 방법은 상기 경계 최적화를 수행하는 단계 후, 상기 모델생성부가 기 설정된 경계값을 기준으로 학습용 피크 데이터에 대응하는 가스 농도를 원핫인코딩 벡터로 변환하여 상기 경계레이블로 설정하고, 상기 학습용 피크 데이터의 은닉벡터에 대응하는 기준벡터를 증강 레이블로 설정하는 단계와, 상기 모델생성부가 상기 학습용 피크 데이터를 예측모델에 입력하는 단계와, 상기 예측모델이 상기 학습용 피크 데이터에 대해 복수의 계층 간 가중치가 적용되는 복수의 연산을 수행하여 마지막 은닉계층의 복수의 은닉노드의 출력값과, 출력층의 출력인 예측값을 산출하는 단계와, 상기 모델생성부가 경계증강손실함수

$$L_{boaderenhanced} = \frac{1}{2} \sum_i v_i (-\log(O_i)) + \frac{1}{2} \frac{1}{N} \sum_i \sum_j |r_{ij} - h_{ij}|^2$$

에 따라 상기 예측모델의 출력인 예측값과 경계 레이블의 차이인 경계 손실 및 상기 예측모델의 마지막 은닉계층의 복수의 은닉노드의 출력값과 증강 레이블과의 차이를 나타내는 증강 손실을 포함하는 경계증강 손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 경계증강 최적화를 수행하는 단계를 더 포함한다. 여기서, 상기 Lboaderenhanced는 경계증강 손실함수이고, 상기 O_i 는 상기 예측모델의 출력인 예측값이고, 상기 v_i 는 상기 예측값에 대응하는 경계 레이블이고, 상기 i 는 상기 예측모델의 출력층의 출력노드에 대응하는 인덱스이고, 상기 h_{ij} 는 상기 예측모델의 마지막 은닉계층의 복수의 은닉노드의 출력값이고, 상기 r_{ij} 는 상기 복수의 은닉노드의 출력값에 대응하는 증강 레이블이고, 상기 j 는 상기 예측모델의 마지막 은닉계층의 복수의 은닉노드에 대응하는 인덱스인 것을 특징으로 한다.

[0028] 여기서, 상기 가스가 CO이면, 상기 파장은 1948 cm^{-1} 및 2063 cm^{-1} 이고, 상기 가스가 아세트산염

(Acetate)이면 상기 파장은 928 cm^{-1} 이고, 상기 가스가 낙산염(Butyrate)이면, 상기 파장은 877 cm^{-1} 인 것을 특징으로 한다.

[0029] 상술한 바와 같은 목적을 달성하기 위한 본 발명의 바람직한 실시예에 따른 라만 분광 신호를 기초로 수용액의 용존 가스 농도를 예측하기 위한 방법은 상기 모델생성부가 실험 데이터인 라만 스펙트럼 데이터로부터 가스의 종류에 상응하는 파장의 학습용 피크 데이터와 상기 학습용 피크 데이터에 대응하는 가스 농도를 추출하여 학습 데이터를 마련하는 단계와, 상기 모델생성부가 상기 학습용 피크 데이터에 대응하는 가스 농도를 수치 레이블로 설정하는 단계와, 상기 모델생성부가 상기 학습용 피크 데이터를 예측모델에 입력하는 단계와, 상기 예측모델이 상기 학습용 피크 데이터에 대해 복수의 계층 간 가중치가 적용되는 복수의 연산을 수행하여 예측값을 산출하는 단계와, 상기 모델생성부가 상기 예측모델의 상기 예측값과 수치 레이블의 차이인 수치 손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 수치 최적화를 수행하는 단계를 더 포함한다.

[0030] 상기 수치 최적화를 수행하는 단계는 상기 모델생성부가 수치손실함수

$$L_{value} = \frac{1}{N} \sum_{i=1}^N |O_i - C_i|^2 + \frac{1}{N} \sum_{i=1}^N (1 - SSIM(O_i, C_i))$$

에 따라 상기 예측모델의 출력값과 수치 레이블의 차이인 수치 손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 수치 최적화를 수행하며, 상기 Lvalue는 수치손실함수이고, 상기 O_i 는 상기 예측모델의 출력인 예측값이고, 상기 C_i 는 상기 예측값에 대응하는 수치 레이블이고, 상기 i 는 상기 예측모델의 출력층의 출력노드에 대응하는 인덱스인 것을 특징으로 한다.

[0031] 상기 방법은 학습 데이터를 마련하는 단계 후, 상기 가스 농도를 수치 레이블로 설정하는 단계 전, 상기 모델생성부가 상기 학습용 피크 데이터에 대응하는 가스 농도를 기 설정된 경계값을 기준으로 원핫인코딩 벡터로 변환하여 경계 레이블로 설정하는 단계와, 상기 모델생성부가 상기 학습용 피크 데이터를 예측모델에 입력하는 단계와, 상기 예측모델이 상기 학습용 피크 데이터에 대해 복수의 계층 간 가중치가 적용되는 복수의 연산을 수행하여 예측값을 산출하는 단계와, 상기 모델생성부가 경계손실함수

$$L_{borderselected} = \sum_i v_i (-\log(O_i))$$

에 따라 상기 예측모델의 출력인 예측값과 경계 레이블의 차이인 경계손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 경계 최적화를 수행하는 단계를 더 포함한다. 여기서, 상기 Lborderselected는 경계 손실함수이고, 상기 O_i 는 상기 예측모델의 출력인 예측값이고, 상기 v_i 는 상기 예측값에 대응하는 경계 레이블이고, 상기 i 는 상기 예측모델의 출력층의 출력노드에 대응하는 인덱스이다.

[0032] 상기 방법은 상기 경계 최적화를 수행하는 단계 후, 상기 모델생성부가 기 설정된 경계값을 기준으로 학습용 피크 데이터에 대응하는 가스 농도를 원핫인코딩 벡터로 변환하여 상기 경계레이블로 설정하고, 상기 학습용 피크 데이터의 은닉벡터에 대응하는 기준벡터를 증강 레이블로 설정하는 단계와, 상기 모델생성부가 상기 학습용 피크 데이터를 예측모델에 입력하는 단계와, 상기 예측모델이 상기 학습용 피크 데이터에 대해 복수의 계층 간 가중치가 적용되는 복수의 연산을 수행하여 마지막 은닉계층의 복수의 은닉노드의 출력값과, 출력층의 출력인 예측값을 산출하는 단계와, 상기 모델생성부가 경계증강손실함수

$$L_{boaderenhanced} = \frac{1}{2} \sum_i v_i (-\log(O_i)) + \frac{1}{2} \frac{1}{N} \sum_i \sum_j |r_{ij} - h_{ij}|^2$$

에 따라 상기 예측모델의 출력인 예측값과 경계 레이블의 차이인 경계 손실 및 상기 예측모델의 마지막 은닉계층의 복수의 은닉노드의 출력값과 증강 레이블과의 차이를 나타내는 증강 손실을 포함하는 경계증강 손실이 최소가 되도록 상기 예측모델의 가중치를 수정하는 경계증강 최적화를 수행하는 단계를 더 포함한다. 여기서, 상기 Lboaderenhanced는 경계증강 손실함수이고, 상기 O_i 는 상기 예측모델의 출력인 예측값이고, 상기 v_i 는 상기 예측값에 대응하는 경계 레이블이고, 상기 i 는 상기 예측모델의 출력층의 출력노드에 대응하는 인덱스이고, 상기 h_{ij} 는 상기 예측모델의 마지막 은닉계층의 복수의 은닉노드의 출력값이고, 상기 r_{ij} 는 상기 복수의 은닉노드의 출력값에 대응하는 증강 레이블이고, 상기 j 는 상기 예측모델의 마지막 은닉계층의 복수의 은닉노드에 대응하는 인덱스이다.

[0033] 상기 방법은 상기 수치 최적화를 수행하는 단계 후, 스펙트럼측정부가 수용액에 내에 프로브를 통해 빛을 발광

하고, 발광된 빛이 산란되어 나오는 것을 측정하여 라만 스펙트럼 데이터를 취득하는 단계와, 피크추출부가 라만 스펙트럼 데이터로부터 가스의 종류에 상응하는 파장의 피크 데이터를 추출하는 단계와, 농도예측부가 상기 피크 데이터를 기초로 예측모델을 통해 상기 수용액에 포함된 가스 농도의 예측값을 산출하는 단계와, 상기 농도예측부가 산출된 예측값으로부터 가스 농도를 도출하여 출력하는 단계를 더 포함한다.

발명의 효과

[0034] 본 발명에 따르면, 라만 분광 신호의 특정 파장의 증폭량을 기초로 머신러닝을 통해 수용액의 용존 가스 농도를 정확하게 예측할 수 있다. 이에 따라, 용존 가스 농도를 실시간으로 획득할 수 있다.

도면의 간단한 설명

[0035] 도 1은 본 발명의 실시예에 따른 라만 분광 신호를 기초로 수용액의 용존 가스 농도를 예측하기 위한 장치의 구성을 설명하기 위한 도면이다.

도 2는 본 발명의 실시예에 따른 라만 스펙트럼 데이터로부터 수용액의 농도를 예측하는 예측부의 구성을 설명하기 위한 블록도이다.

도 3은 본 발명의 실시예에 따른 가스 농도를 예측하는 예측모델의 구성을 설명하기 위한 도면이다.

도 4는 본 발명의 실시예에 따른 가스 농도를 예측하는 예측모델의 노드를 설명하기 위한 도면이다.

도 5는 본 발명의 실시예에 따른 예측모델의 원형을 생성하는 방법을 설명하기 위한 흐름도이다.

도 6은 본 발명의 실시예에 따른 예측모델의 원형을 생성하는 방법을 설명하기 위한 그래프이다.

도 7은 본 발명의 실시예에 따른 예측모델에 대한 학습 방법에 대해서 설명하기 위한 흐름도이다.

도 8은 본 발명의 실시예에 따른 실험 데이터에서 아세트산염 및 낙산염이 포함된 수용액에 대한 라만 스펙트럼 데이터를 나타내는 그래프이다.

도 9는 본 발명의 실시예에 따른 실험 데이터에서 아세트산염 및 낙산염의 혼합물 및 일산화탄소가 포함된 수용액에 대한 라만 스펙트럼 데이터를 나타내는 그래프이다.

도 10은 본 발명의 실시예에 따른 실험 데이터에서 수용액에 포함된 가스의 농도가 알려진 라만 스펙트럼 데이터를 나타내는 그래프이다.

도 11은 본 발명의 실시예에 따른 예측모델에 대한 경계학습을 설명하기 위한 흐름도이다.

도 12는 본 발명의 실시예에 따른 예측모델에 대한 경계학습을 설명하기 위한 도면이다.

도 13은 본 발명의 실시예에 따른 예측모델에 대한 수치학습을 설명하기 위한 흐름도이다.

도 14는 본 발명의 실시예에 따른 라만 분광 신호를 기초로 수용액의 용존 가스 농도를 예측하기 위한 방법을 설명하기 위한 흐름도이다.

도 15는 본 발명의 실시예에 따른 컴퓨팅 장치를 나타내는 도면이다.

발명을 실시하기 위한 구체적인 내용

[0036] 본 발명의 상세한 설명에 앞서, 이하에서 설명되는 본 명세서 및 청구범위에 사용된 용어나 단어는 통상적이거나 사전적인 의미로 한정해서 해석되어서는 아니 되며, 발명자는 그 자신의 발명을 가장 최선의 방법으로 설명하기 위해 용어의 개념으로 적절하게 정의할 수 있다는 원칙에 입각하여 본 발명의 기술적 사상에 부합하는 의미와 개념으로 해석되어야만 한다. 따라서 본 명세서에 기재된 실시예와 도면에 도시된 구성은 본 발명의 가장 바람직한 실시예에 불과할 뿐, 본 발명의 기술적 사상을 모두 대변하는 것은 아니므로, 본 출원시점에 있어서 이들을 대체할 수 있는 다양한 균등물과 변형 예들이 있을 수 있음을 이해하여야 한다.

[0037] 이하, 첨부된 도면을 참조하여 본 발명의 바람직한 실시예들을 상세히 설명한다. 이때, 첨부된 도면에서 동일한 구성 요소는 가능한 동일한 부호로 나타내고 있음을 유의해야 한다. 또한, 본 발명의 요지를 흐리게 할 수 있는 공지 기능 및 구성에 대한 상세한 설명은 생략할 것이다. 마찬가지로 이유로 첨부 도면에 있어서 일부 구성요소는 과장되거나 생략되거나 또는 개략적으로 도시되었으며, 각 구성요소의 크기는 실제 크기를 전적으로 반영하는 것이 아니다.

- [0038] 먼저, 본 발명의 실시예에 따른 라만 분광 신호를 기초로 수용액의 용존 가스 농도를 예측하기 위한 장치에 대해서 설명하기로 한다. 도 1은 본 발명의 실시예에 따른 라만 분광 신호를 기초로 수용액의 용존 가스 농도를 예측하기 위한 장치의 구성을 설명하기 위한 도면이다. 도 1을 참조하면, 본 발명의 실시예에 따른 라만 분광 신호를 기초로 수용액의 용존 가스 농도를 예측하기 위한 장치(PA: Prediction Apparatus, 이하, '예측장치'로 축약함)는 프로브(11)를 포함하는 스펙트럼측정부(10)와, 예측부(20)를 포함한다.
- [0039] 스펙트럼측정부(10)는 도 1에 도시된 바와 같이, 수용액(Q)에 내에 프로브(11)를 통해 빛을 발광하고, 발광된 빛이 산란되어 나오는 것을 측정하여 라만 스펙트럼 데이터를 생성한다. 이러한 라만 스펙트럼 데이터는 예측부(20)에 제공된다.
- [0040] 예측부(20)는 인공지능망 알고리즘인 예측모델(PM: Prediction Model)을 통해 라만 스펙트럼 데이터로부터 수용액의 농도를 예측한다.
- [0041] 그러면, 예측부(20)의 구성에 대해서 보다 상세하게 설명하기로 한다. 도 2는 본 발명의 실시예에 따른 라만 스펙트럼 데이터로부터 수용액의 농도를 예측하는 예측부의 구성을 설명하기 위한 블록도이다. 도 2를 참조하면, 예측부(20)는 모델생성부(100), 피크추출부(200) 및 농도예측부(300)를 포함한다.
- [0042] 모델생성부(100)는 실험 데이터로부터 학습 데이터를 마련한다. 즉, 모델생성부(100)는 실험 데이터인 라만 스펙트럼 데이터로부터 가스의 종류에 상응하는 파장의 학습용 피크 데이터와 상기 학습용 피크 데이터에 대응하는 가스 농도를 추출하여 학습 데이터를 마련할 수 있다. 모델생성부(100)는 마련된 학습 데이터를 이용하여 학습(machine learning/deep learning)을 수행하여 수용액에 포함된 가스 농도의 예측값을 산출하도록 예측모델(PM)을 생성한다. 또한, 모델생성부(100)는 마련된 학습 데이터 중 일부를 검사 데이터로 설정하고, 검사 데이터를 이용하여 학습 결과를 검사할 수 있다. 모델생성부(100)의 학습 방법에 대해서는 아래에서 보다 상세하게 설명하기로 한다.
- [0043] 피크추출부(200)는 스펙트럼측정부(10)로부터 수용액에 대한 라만 스펙트럼 데이터가 입력되면, 라만 스펙트럼 데이터로부터 가스의 종류에 상응하는 파장(Raman Shift)의 피크 데이터를 추출한다. 여기서, 가스가 CO인 경우, 파장은 1948 cm^{-1} 및 2063 cm^{-1} 이다. 또한, 가스가 아세트산염(Acetate)이면 파장은 928 cm^{-1} 이고, 가스가 낙산염(Butyrate)이면, 파장은 877 cm^{-1} 이다.
- [0044] 농도예측부(300)는 피크추출부(200)가 추출한 피크 데이터를 기초로 예측모델을 통해 수용액에 포함된 가스 농도의 예측값을 산출하고, 산출된 예측값으로부터 가스 농도를 도출하여 출력한다. 이와 같이, 본 발명은 종래의 복잡한 실험을 통해 시료를 채취하여 도출되는 수용액의 용존 가스 농도를 스펙트럼측정부(10)의 한 번의 측정을 통해 도출된 라만 스펙트럼 데이터로부터 예측모델(PM)을 이용하여 예측할 수 있다. 이에 따라, 용존 가스 농도를 실시간으로 획득할 수 있다.
- [0045] 그러면, 본 발명의 실시예에 따른 가스 농도를 예측하는 예측모델(PM)의 구성에 대해서 보다 상세하게 설명하기로 한다. 도 3은 본 발명의 실시예에 따른 가스 농도를 예측하는 예측모델(PM)의 구성을 설명하기 위한 도면이다. 도 4는 본 발명의 실시예에 따른 가스 농도를 예측하는 예측모델(PM)의 노드를 설명하기 위한 도면이다.
- [0046] 본 발명의 실시예에서 예측모델(PM)은 복수의 계층으로 이루어지며, 이러한 복수의 계층은 입력계층(IL: Input Layer), 적어도 하나의 은닉계층(HL: Hidden Layer, HL1~HLk) 및 출력계층(OL: Output Layer)을 포함한다.
- [0047] 또한, 복수의 계층(IL, HL, OL) 각각은 복수의 노드를 포함한다. 예컨대, 도시된 바와 같이, 입력계층(IL)은 a개의 입력노드(i1 내지 ia)를 포함하며, 출력계층(OL)은 1개의 출력노드(o)를 포함한다. 또한, 은닉계층(HL) 중 제1 은닉계층(HL1)은 b개의 노드(g1 내지 gb)를 포함하고, 마지막 은닉계층인 제k 은닉계층(HLk)은 j개의 노드(h1 내지 hj)를 포함할 수 있다.
- [0048] 복수의 계층(IL, HL, OL)의 복수의 노드는 개별적으로 연산을 수행한다. 서로 다른 계층의 노드는 가중치(W: weight)를 가지는 채널(점선으로 표시)로 연결된다. 다른 말로, 어느 하나의 노드의 연산 결과는 가중치가 적용되어 다음 계층 노드의 입력이 된다. 도 4에 어느 하나의 노드(ND)의 구성이 상세하게 도시되었다. 이러한 노드(ND)의 연산에 대해 도 4를 참조하여 설명하기로 한다. 노드(ND)는 이전 계층의 복수(n개)의 노드의 출력값 각각이 $X=[X_1, X_2, \dots, X_n]$ 에 가중치 $W=[W_1, W_2, \dots, W_n]$ 를 적용한 값을 입력받고, 이를 모두 합산 후, 그 합산 값에 함수 F를 취한다. 여기서, 함수 F는 활성화 함수(activation function) 또는 전달함수(transfer function)라고 한다. 이러한 활성화함수는 계단(step)함수, 부호(sgn)함수, 시그모이드(Sigmoid), 하이퍼볼릭탄

젠트(tanh: Hyperbolic tangent), ELU(Exponential Linear Unit), ReLU(Rectified Linear Unit), Leaky ReLU, Maxout, Minout, Softmax 함수 등을 예시할 수 있다.

[0049] 노드(ND)의 출력은 다음의 수학적 식 1과 같다.

수학적 식 1

$$y = F\left(\sum_{i=0}^n x_n \cdot w_n - \theta\right)$$

[0050]

[0051] 설명되지 않은 변수 θ 는 임계치 혹은 바이어스이며, 이러한 임계치는 수학적 식 4에서 $(\sum x_n \cdot w_n)$ 의 값이 임계치 보다 작을 때 해당 노드가 활성화되지 않도록 하는 역할을 한다.

[0052] 수학적 식 1에 따라 노드(ND)가 수행하는 연산은 다음과 같이 수행된다. 일례로, 노드(ND)의 이전 계층의 노드가 3개라고 가정한다. 이에 따라, 이전 계층의 3개의 노드 각각의 출력이 해당 노드(ND)에 대해 3개의 입력(n=3) X1, X2, X3 각각에 3개의 가중치 W1, W2, W3가 적용되어 입력된다. 이에 따라, 노드(ND)는 3개의 입력 X1, X2, X3에 대응하는 가중치 W1, W2, W3을 곱한 값을 입력받고, 모두 합산한 후, 합산된 값을 전달 함수에 대입하여 출력을 산출한다. 구체적으로, 입력 [X1, X2, X3] = 0.5, -0.3, -0.1이라고 가정하고, 가중치 [W1, W2, W3] = 4, 5, 4라고 가정한다. 또한, 설명의 편의를 위하여 활성화 함수는 부호함수, 즉, 'sgn()'이라고 가정하면, 다음의 수학적 식 2와 같이 출력값이 산출된다.

수학적 식 2

$$X1 \times W1 = 0.5 \times 4 = 2$$

$$X2 \times W2 = -0.3 \times 5 = -1.5$$

$$X3 \times W3 = -0.1 \times 4 = -0.4$$

$$2.0 + (-1.5) + -0.4 = 0.1$$

$$\text{sgn}(0.1) = 1$$

[0053]

[0054] 도 4에서 설명된 바와 같이, 예측모델(PM)의 어느 한 계층의 어느 하나의 노드 또한 이전 계층의 복수의 노드의 출력에 가중치(W)를 적용한 값을 입력받는다. 그러면, 해당 노드는 입력을 합산하고, 합산한 값에 활성화함수에 의한 연산을 수행하여 해당 노드의 출력값을 산출한다. 출력값은 다음 계층의 입력이거나, 예측모델(PM)의 최종 출력값이 될 수 있다. 이에 따라, 예측모델(PM)은 피크 데이터가 입력되면, 입력된 피크 데이터에 대해 복수의 계층(IL, HL, OL) 간 가중치가 적용되는 복수의 연산을 수행하여 수용액의 가스 농도를 예측하는 예측값을 산출할 수 있다.

- [0055] 보다 구체적으로 설명하면, 농도예측부(300)는 피크추출부(200)로부터 피크 데이터를 수신하면, 피크 데이터를 바이너리 데이터로 변환하고, 예측모델(PM)의 입력계층(IL)의 입력노드(i1 내지 ia)의 수에 맞게 분할한 후, 입력계층(IL)의 복수의 입력노드(i1 내지 ia)에 분산하여 입력한다.
- [0056] 그러면, 제1 은닉계층(HL1)의 복수의 제1 은닉노드(g1 내지 gb) 각각은 복수의 입력노드(i1 내지 ia)에 분산되어 입력된 바이너리 데이터(즉, 피크 데이터) 각각에 가중치가 적용된 값을 입력받고(점선으로 표시), 입력된 값을 모두 합산한 후, 합산된 값에 대해 활성화함수에 따른 연산을 수행하여 복수의 제1 은닉노드의 출력값을 산출한다.
- [0057] 이어서, 도시되지는 않았지만, 제2 은닉계층의 복수의 제2 은닉노드 각각은 복수의 제1 은닉노드(g1 내지 gb)의 복수의 출력값 각각에 가중치가 적용된 값을 입력받고, 입력된 값을 모두 합산하고, 합산된 값에 대해 활성화함수에 따른 연산을 수행하여 복수의 제2 은닉노드의 출력값을 산출한다. 이와 같은 방식으로, 은닉계층(HL) 내에서 이전의 노드값이 가중치가 적용되어 전달되고, 연산을 통해 현재의 노드값이 산출된다. 이러한 과정을 반복하여, 마지막 은닉계층인 제k 은닉계층(HLk)의 복수의 제k 은닉노드(h1 ~ hj)의 복수의 제k 은닉노드의 출력값을 산출할 수 있다.
- [0058] 이에 따라, 도 3을 참조하면, 출력노드(O)는 제k 은닉계층(HLk)의 복수의 제k 은닉노드(h1 ~ hj)의 복수의 제k 은닉노드의 출력값에 가중치 $w=[w_1, w_2, \dots, w_j]$ 가 적용된 값을 입력받고(점선으로 표시), 입력된 값을 모두 합산한 후, 합산된 값에 대해 활성화함수에 따른 연산을 수행하여 출력값을 산출한다. 출력층(OL)의 출력노드(O)는 수용액의 가스 농도에 대응한다. 그리고 출력노드(O)의 출력값은 수용액의 가스 농도의 예측값이다. 예컨대, 출력노드(O)의 출력값이 0.089이면, 해당 가스의 농도는 9%이고, 출력노드(O)의 출력값이 0.911이면, 해당 가스의 농도는 91%를 의미한다. 이와 같이, 예측모델(PM)이 예측값(예컨대, 0.089, 0.911)을 산출하면, 농도예측부(300)는 예측값을 농도로 변환하여 출력한다.
- [0059] 그러면, 전술한 예측모델(PM)의 원형을 생성하는 방법에 대해서 설명하기로 한다. 도 5는 본 발명의 실시예에 따른 예측모델의 원형을 생성하는 방법을 설명하기 위한 흐름도이다. 도 6은 본 발명의 실시예에 따른 예측모델의 원형을 생성하는 방법을 설명하기 위한 그래프이다.
- [0060] 도 5를 참조하면, 모델생성부(100)는 S100 단계에서 실험 데이터로부터 학습 데이터 및 검사 데이터를 추출한다. 즉, 모델생성부(100)는 실험 데이터인 라만 스펙트럼 데이터로부터 가스의 종류에 상응하는 파장의 학습용 피크 데이터와 학습용 피크 데이터에 대응하는 가스 농도를 추출하여 학습 데이터를 마련하고, 실험 데이터인 라만 스펙트럼 데이터로부터 가스의 종류에 상응하는 파장의 검사용 피크 데이터와 검사용 피크 데이터에 대응하는 가스 농도를 추출하여 검사 데이터를 마련한다.
- [0061] 다음으로, 모델생성부(100)는 S110 단계에서 초기값에 따라 예측모델(PM)의 은닉계층 및 은닉노드의 수를 설정한다. 초기값은 예측모델(PM)의 은닉계층의 수가 1개인 것이 바람직하다. 이는 도 3의 예에서, 제1 은닉계층(HL1)만 존재하며, 나머지 은닉계층은 존재하지 않는 상태를 의미한다. 하지만, 본 발명을 이에 한정하는 것은 아니며, 신뢰되는 데이터를 근거로 초기값을 달리 설정할 수도 있다.
- [0062] 모델생성부(100)는 S120 단계에서 실험 데이터로부터 추출되는 학습 데이터를 이용하여 앞서 설정된 수의 은닉계층 및 은닉노드를 가지는 예측모델(PM)에 대한 학습을 수행한다. 그런 다음, 모델생성부(100)는 S130 단계에서 검사 데이터를 이용하여 예측모델(PM)의 학습 성과를 검사한다. 이어서, 모델생성부(100)는 S140 단계에서 학습 데이터 및 검사 데이터의 정확도를 산출한다.
- [0063] 모델생성부(100)는 S150 단계에서 학습 데이터 및 검사 데이터의 정확도가 조건을 만족하는지 여부를 판단한다. 여기서, 조건은 학습 데이터 정확도가 검사 데이터의 정확도 보다 높으면서 학습 데이터 정확도 및 검사 데이터의 정확도 양자 모두가 임계치 이상인 경우를 의미한다.
- [0064] S150 단계의 판단 결과, 학습 데이터 및 검사 데이터의 정확도가 설정 조건을 만족하지 않으면, 모델생성부(100)는 S160 단계로 진행하여 은닉계층 및 은닉노드의 수를 재설정한다. 이때, 모델생성부(100)는 은닉계층 및 은닉노드의 수를 증가 혹은 감소시킨다. 그런 다음, 모델생성부(100)는 전술한 S120 단계 내지 S150 단계를 반복한다.
- [0065] 반면, S150 단계의 판단 결과, 학습 데이터 및 검사 데이터의 정확도가 설정 조건을 만족하면, 모델생성부(100)는 현재 설정에 따라 은닉계층의 수 및 은닉노드의 수를 결정한다. 이로써, 예측모델(PM)의 원형이 결정된다.
- [0066] 예컨대, 도 6의 그래프에 도시된 바와 같이, 은닉계층이 1개 및 2개의 경우, 학습 데이터의 정확도 및 검사 데

이터 정확도 양자 모두 임계치 미만이다. 그리고 은닉계층의 수가 4개 및 5개의 경우, 학습 데이터의 정확도는 매우 높지만, 검사 데이터의 정확도는 임계치 미만이다. 이에 따라, 은닉계층의 수가 3개일 때에만 조건을 만족하기 때문에 예측모델(PM)의 원형은 은닉계층이 3개인 모델로 결정할 수 있다.

[0067] 그러면, 본 발명의 실시예에 따른 예측모델(PM)에 대한 학습 방법에 대해서 보다 상세하게 설명하기로 한다. 도 7은 본 발명의 실시예에 따른 예측모델(PM)에 대한 학습 방법에 대해서 설명하기 위한 흐름도이다. 도 8은 본 발명의 실시예에 따른 실험 데이터에서 아세트산염 및 낙산염이 포함된 수용액에 대한 라만 스펙트럼 데이터를 나타내는 그래프이다. 도 9는 본 발명의 실시예에 따른 실험 데이터에서 아세트산염 및 낙산염의 혼합물 및 일산화탄소가 포함된 수용액에 대한 라만 스펙트럼 데이터를 나타내는 그래프이다. 도 10은 본 발명의 실시예에 따른 실험 데이터에서 수용액에 포함된 가스의 농도가 알려진 라만 스펙트럼 데이터를 나타내는 그래프이다.

[0068] 도 7을 참조하면, 모델생성부(100)는 S210 단계에서 학습 데이터를 마련한다. 도 8 내지 도 10에 도시된 바와 같이, 실험 데이터는 가스가 포함된 수용액에 대한 라만 스펙트럼 데이터 및 해당 수용액의 가스 농도를 포함한다. 따라서 모델생성부(100)는 실험 데이터인 라만 스펙트럼 데이터로부터 가스의 종류에 상응하는 파장의 학습용 피크 데이터와 학습용 피크 데이터에 대응하는 가스 농도를 추출하여 학습 데이터를 마련한다. 예컨대, 도 8의 (A)는 아세트산염(Acetate)이 포함된 물과 배양액의 혼합액에 대한 라만 스펙트럼 데이터이다. 이와 같이, 수용액에 포함된 가스의 종류가 아세트산염(Acetate)이면 928 cm^{-1} 의 피크 데이터를 추출한다. 또한, 도 8의 (B)는 낙산염(Butyrate)이 포함된 물과 배양액의 혼합액에 대한 라만 스펙트럼 데이터이다. 이와 같이, 수용액에 포함된 가스의 종류가 아세트산염(Acetate)이면 877 cm^{-1} 의 피크 데이터를 추출한다. 도 9의 (C)는 아세트산염(Acetate) 및 낙산염(Butyrate)의 혼합물 수용액에 대한 라만 스펙트럼 데이터이다. 수용액에 포함된 가스가 아세트산염(Acetate) 및 낙산염(Butyrate)이기 때문에 928 cm^{-1} 및 877 cm^{-1} 각각의 피크 데이터를 추출한다. 도 9의 (D)는 일산화탄소(CO)의 수용액에 대한 라만 스펙트럼 데이터이다. 이와 같이, 수용액에 포함된 가스의 종류가 일산화탄소(CO)이면 1948 cm^{-1} 및 2063 cm^{-1} 의 피크 데이터를 추출한다. 전술한 바와 같이, 실험 데이터는 해당 수용액의 가스 농도를 포함한다. 예를 들면, 도 10의 (E)는 2.5% 농도(concentration)의 일산화탄소를 포함하는 수용액의 라만 스펙트럼 데이터이고, (F)는 5% 농도의 일산화탄소를 포함하는 수용액의 라만 스펙트럼 데이터를 보인다. 이에 따라, 해당 실험 데이터로부터 수용액에 포함된 가스의 종류가 일산화탄소(CO)이기 때문에 1948 cm^{-1} 및 2063 cm^{-1} 의 피크 데이터를 추출하고, 그 피크 데이터에 대응하는 가스의 농도(2.5%, 5%)를 추출할 수 있다.

[0069] 학습 데이터를 마련한 후, 모델생성부(100)는 S220 단계에서 학습 데이터를 이용하여 경계 학습을 수행할 수 있다. 이러한 S220 단계는 선택적인 것으로, 필요에 따라 수행하거나, 생략할 수 있다. 경계 학습은 학습을 통해 명확하게 구분되지 않는 경계를 명확하게 위한 것이다. 도 10을 참조하면, 일산화탄소 농도 2.5%와 일산화탄소 농도 5%의 데이터는 학습을 통해 명확하게 구분되지 않는다고 가정한다. 이러한 경우, 2.5%와 5% 사이의 값을 경계값으로 설정하고(예컨대, 2.75%), 그 경계값을 기준으로 레이블을 설정하여 농도 2.5%와 5% 사이의 값들을 명확하게 구분되도록 경계학습을 수행한다. 경계학습의 구체적인 내용에 대해서는 아래에서 더 상세하게 설명될 것이다.

[0070] 다음으로, 모델생성부(100)는 S230 단계에서 학습 데이터를 이용하여 수치 학습을 수행한다. 수치학습은 목표값, 즉, 레이블을 실험 데이터에서 얻어진 농도(수치)로 설정하여 학습을 수행함으로써, 실제로 수용액에서 가스의 농도를 예측할 수 있도록 예측모델(PM)을 학습시키는 절차이다. 수치학습의 구체적인 내용에 대해서는 아래에서 더 상세하게 설명될 것이다.

[0071] 다음으로, 본 발명의 실시예에 따른 예측모델(PM)에 경계학습에 대해서 보다 상세하게 설명하기로 한다. 도 11은 본 발명의 실시예에 따른 예측모델(PM)에 대한 경계학습을 설명하기 위한 흐름도이다. 도 12는 본 발명의 실시예에 따른 예측모델(PM)에 대한 경계학습을 설명하기 위한 도면이다. 다시 강조하면, 도 11은 도 7의 S220 단계의 경계학습의 일 실시예를 보다 상세하게 설명하는 것이다.

[0072] 도 11을 참조하면, 모델생성부(100)는 S310 단계에서 가스의 종류에 따른 파장의 학습용 피크 데이터 및 대응하는 가스 농도를 포함하는 복수의 학습 데이터를 마련한다. 그런 다음, 모델생성부(100)는 S320 단계에서 학습용 피크 데이터에 대응하는 가스 농도를 기 설정된 경계값을 기준으로 원핫인코딩(One-hot-encoding) 벡터로 변환하여 경계 레이블을 설정한다. 일 실시예에 따르면, 일산화탄소 농도 2.5%와 5%의 구분을 명확하게 하기 위하여

경계값을 2.75%, 즉, 0.0275로 설정하였다고 가정한다. 그러면, 2.75% 이상의 가스 농도는 벡터 1로 변환되고, 2.75% 미만의 가스 농도는 벡터 0로 변환된다. 예컨대, 일산화탄소 농도 2.5%의 경우, 레이블은 벡터 0으로 설정되며, 일산화탄소 농도 5%의 경우, 레이블은 벡터 1로 설정된다.

[0073] 경계 레이블이 설정되면, 모델생성부(100)는 S330 단계에서 경계 손실을 최적화하는 학습을 수행한다. 이때, 모델생성부(100)가 학습용 피크 데이터를 예측모델(PM)에 입력하면, 예측모델(PM)은 학습용 피크 데이터에 대해 복수의 계층 간 가중치가 적용되는 복수의 연산을 통해 예측값을 산출한다. 그러면, 모델생성부(100)는 예측값과 경계 레이블의 차이인 경계손실이 최소가 되도록 예측모델(PM)의 가중치를 수정하는 경계손실 최적화를 수행한다. 이때, 경계손실은 다음의 수학적 식 3과 같은 경계손실함수를 통해 도출할 수 있다.

수학적 식 3

$$L_{borderselected} = \sum_i v_i (-\log(O_i))$$

[0074]

[0075] 수학적 식 3에서, $L_{borderselected}$ 는 경계손실함수를 나타낸다. 특히, O_i 는 예측모델(PM)의 출력층(OL)의 출력노드(0)의 출력인 예측값이고, v_i 는 예측값에 대응하는 경계 레이블이다. 여기서, i 는 예측모델(PM)의 출력층(OL)의 출력노드(0)에 대응하는 인덱스이다.

[0076] 정리하면, 모델생성부(100)는 수학적 식 3의 경계 손실함수를 통해 경계 손실을 도출한 후, 경계 손실이 최소가 되도록 예측모델(PM)의 가중치를 수정하는 경계 손실 최적화를 수행한다. 이러한 경계 손실 최적화는 복수의 서로 다른 학습용 데이터를 이용하여 반복하여 수행되며, 이러한 반복은 평가 지표를 통해 원하는 정확도에 도달할 때까지 이루어질 수 있다.

[0077] 전술한 바와 같은 경계 손실 최적화에 따른 학습이 완료되면, 모델생성부(100)는 S140 단계에서 예측모델(PM)을 통해 경계값을 기준으로 구분되는 그룹 각각의 복수의 은닉벡터 중 기준벡터를 선택한다. 복수의 학습 데이터 중 대응하는 가스 농도가 경계값 이상인 학습용 피크 데이터를 제1 그룹으로 분류하고, 대응하는 가스 농도가 경계값 미만인 학습용 피크 데이터를 제2 그룹으로 분류할 때, 기준벡터를 선택하는 방법은 다음과 같다. 앞서(S130) 사용된 학습 데이터의 학습용 피크 데이터를 경계 손실 최적화가 이루어진 예측모델(PM)에 다시 입력하여 예측모델(PM)의 마지막, 은닉계층, 즉, 제k 은닉계층(HLk)의 복수의 은닉노드(h_1 내지 h_j)의 출력값인 은닉벡터를 도출한다. 즉, 은닉벡터는 $H[h_1, h_2, h_3, \dots, h_j]$ 이다. 모델생성부(100)는 복수의 차원으로 이루어진 복수의 은닉벡터를 소정의 벡터공간에 임베딩한다. 도 12의 (가)에 소정의 벡터공간에 복수의 은닉벡터가 임베딩된 예를 도시하였다. 기본적으로, 경계 손실 최적화에 의해 벡터공간 상에서 경계값 이상의 농도를 가지는 제1 그룹에 대한 은닉벡터(원형 표시)와 경계값 미만의 농도를 가지는 제2 그룹에 대한 은닉벡터(사각형 표시)는 경계(BORDER)에 의해 구분이 이루어진다. 하지만, 도 12의 (가)에 도시된 바와 같이, 경계(BORDER)에 걸쳐 있는 은닉벡터도 존재한다. 모델생성부(100)는 이러한 벡터 공간에 임베딩된 복수의 은닉벡터 중 각 그룹 별로 기준벡터를 선택한다. 기준벡터는 9의 (A)에 도시된 바와 같이, 동일한 그룹 내의 은닉벡터 중 중간값을 선택할 수 있다.

[0078] 다음으로, 모델생성부(100)는 S350 단계에서 경계 레이블 및 증강 레이블을 설정한다. 경계 레이블은 앞서 S320 단계에서와 동일하게 학습용 피크 데이터에 대응하는 가스 농도를 기 설정된 경계값을 기준으로 원핫인코딩 벡터로 변환하여 설정된다. 증강 레이블은 앞서(S340) 그룹별로 선택된 기준벡터를 이용한다.

[0079] 경계 레이블 및 증강 레이블이 설정되면, 모델생성부(100)는 S360 단계에서 증강 손실을 최적화하는 학습을 수행한다. 이때, 모델생성부(100)가 학습용 피크 데이터를 예측모델(PM)에 입력하면, 예측모델(PM)은 학습용 피크 데이터에 대해 복수의 계층 간 가중치가 적용되는 복수의 연산을 통해 순차로 마지막 은닉계층(예컨대, 제k 은닉계층)의 출력값 및 출력층의 출력인 예측값을 산출한다. 그러면, 모델생성부(100)는 예측값과 경계 레이블의 차이를 나타내는 경계 손실과, 마지막 은닉계층(예컨대, 제k 은닉계층)의 출력값 $H[h_1, h_2, h_3, \dots, h_j]$ 과 증강 레이블과의 차이를 나타내는 증강 손실을 포함하는 경계증강 손실이 최소가 되도록 예측모델(PM)의 가중치를 수정하는 경계증강 손실 최적화를 수행한다. 이때, 경계증강 손실은 다음의 수학적 식 3과 같은 경계증강 손실함수를 통해 도출할 수 있다.

수학식 4

$$L_{boaderenhanced} = \frac{1}{2} \sum_i v_i (-\log(O_i)) + \frac{1}{2} \frac{1}{N} \sum_i \sum_j |r_{ij} - h_{ij}|^2$$

[0080]

[0081]

수학식 4에서, $L_{boaderenhanced}$ 는 경계증강 손실함수를 나타낸다. 특히, O_i 는 예측모델(PM)의 출력층(OL)의 출력노드(0)의 출력인 예측값이고, v_i 는 예측값에 대응하는 경계 레이블이다. 여기서, i 는 예측모델(PM)의 출력층(OL)의 출력노드(0)에 대응하는 인덱스이다. 또한, h_{ij} 는 예측모델(PM)의 마지막 은닉계층(예컨대, HLk)의 복수의 은닉노드의 출력값인 은닉벡터 $H[h_1, h_2, h_3, \dots, h_j]$ 이고, r_{ij} 는 마지막 은닉계층(예컨대, HLk)의 복수의 은닉노드의 출력값인 은닉벡터 $H[h_1, h_2, h_3, \dots, h_j]$ 에 대응하는 증강 레이블이다. 이 증강 레이블은 그룹 별로 선택된 기준벡터이다. j 는 예측모델(PM)의 마지막 은닉계층(예컨대, HLk)의 복수의 은닉노드에 대응하는 인덱스이다.

[0082]

정리하면, 모델생성부(100)는 수학식 4의 경계증강 손실함수를 통해 경계증강 손실을 도출한다. 경계증강 손실은 예측값과 경계 레이블의 차이를 나타내는 경계 손실과, 마지막 은닉계층(예컨대, 제k 은닉계층)의 출력값 $H[h_1, h_2, h_3, \dots, h_j]$ 과 증강 레이블과의 차이를 나타내는 증강 손실을 포함한다. 이와 같이, 경계증강 손실이 도출되면, 모델생성부(100)는 경계증강 손실이 최소가 되도록 예측모델(PM)의 가중치를 수정하는 경계증강 손실 최적화를 수행한다. 이러한 경계증강 손실 최적화는 복수의 서로 다른 학습용 데이터를 이용하여 반복하여 수행되며, 이러한 반복은 평가 지표를 통해 원하는 정확도에 도달할 때까지 이루어질 수 있다.

[0083]

이러한 경계증강 손실에 의한 학습이 완료되면, 벡터 공간 상의 복수의 은닉벡터는 도 12의 (B)와 같이 분류될 수 있다. 즉, 복수의 은닉벡터는 벡터 공간 상에서 기준벡터를 향하여 이동된다. 특히, 경계(BORDER)에 존재하는 은닉벡터가 기준벡터를 향하여 이동함에 따라 제1 그룹과 제2 그룹의 구분이 보다 명확해짐을 알 수 있다. 즉, 경계증강 손실 최적화를 통한 학습을 수행함에 따라 예측모델(PM)에서 수치의 차이의 구분 성능이 향상될 수 있다.

[0084]

다음으로, 본 발명의 실시예에 따른 예측모델(PM)에 수치학습에 대해서 보다 상세하게 설명하기로 한다. 도 13은 본 발명의 실시예에 따른 예측모델(PM)에 대한 수치학습을 설명하기 위한 흐름도이다. 다시 강조하면, 도 13은 도 7의 S230 단계의 수치학습을 보다 상세하게 설명하는 것이다.

[0085]

도 13을 참조하면, 모델생성부(100)는 S410 단계에서 가스의 종류에 따른 과장의 학습용 피크 데이터 및 대응하는 가스 농도를 포함하는 복수의 학습 데이터를 마련한다. 그런 다음, 모델생성부(100)는 S420 단계에서 학습용 피크 데이터에 대응하는 가스 농도의 수치를 수치 레이블로 설정한다. 예컨대, 일산화탄소 농도 2.5%의 경우, 수치 레이블은 0.025로 설정되며, 일산화탄소 농도 5%의 경우, 레이블은 0.050로 설정된다.

[0086]

수치 레이블이 설정되면, 모델생성부(100)는 S430 단계에서 수치 손실을 최적화하는 학습을 수행한다. 이때, 모델생성부(100)가 학습용 피크 데이터를 예측모델(PM)에 입력하면, 예측모델(PM)은 학습용 피크 데이터에 대해 복수의 계층 간 가중치가 적용되는 복수의 연산을 통해 예측값을 산출한다. 그러면, 모델생성부(100)는 예측값과 수치 레이블의 차이인 수치 손실이 최소가 되도록 예측모델(PM)의 가중치를 수정하는 수치 손실 최적화를 수행한다. 이때, 수치 손실은 다음의 수학식 5와 같은 수치손실함수를 통해 도출할 수 있다.

수학식 5

$$L_{value} = \frac{1}{N} \sum_{i=1}^N |O_i - C_i|^2 + \frac{1}{N} \sum_{i=1}^N (1 - SSIM(O_i, C_i))$$

[0087]

[0088]

수학식 3에서 L_{value} 는 수치손실함수를 나타낸다. O_i 는 출력층(OL)의 출력인 예측값이고, C_i 는 출력층(OL)의 출력인 예측값에 대응하는 수치 레이블이다. 그리고 i 는 예측모델(PM)의 출력층(OL)의 출력노드에 대응하는 인덱스이다. 수치손실함수 L_{value} 의 첫 번째 항(term)은 L1-norm 손실이며, 두 번째 항은 SSIM(Structural Similarity Index)을 나타낸다. 즉, L1-norm 및 SSIM을 합산한 값이 최소가 되도록 최적화를 수행할 수 있다.

- [0089] 정리하면, 모델생성부(100)는 수학적 5의 수치 손실함수를 통해 수치 손실을 도출한 후, 수치 손실이 최소가 되도록 예측모델(PM)의 가중치를 수정하는 수치 손실 최적화를 수행한다. 이러한 수치 손실 최적화는 복수의 서로 다른 학습용 데이터를 이용하여 반복하여 수행되며, 이러한 반복은 평가 지표를 통해 원하는 정확도에 도달할 때까지 이루어질 수 있다.
- [0090] 전술한 바와 같은 방법을 통해 예측모델(PM)이 생성되면, 본 발명의 실시예에 따른 예측장치(PA)는 실시간으로 수용액의 용존 가스 농도를 예측할 수 있다. 이러한 방법에 대해서 설명하기로 한다. 도 14는 본 발명의 실시예에 따른 라만 분광 신호를 기초로 수용액의 용존 가스 농도를 예측하기 위한 방법을 설명하기 위한 흐름도이다.
- [0091] 도 14를 참조하면, 스펙트럼측정부(10)는 S510 단계에서 도 1에 도시된 바와 같이, 수용액에 내에 프로브(11)를 통해 빛을 발광하고, 발광된 빛이 산란되어 나오는 것을 측정하여 라만 스펙트럼 데이터를 취득한다. 이러한 라만 스펙트럼 데이터는 예측부(20)에 제공된다.
- [0092] 예측부(20)의 피크추출부(200)는 S520 단계에서 스펙트럼측정부(10)로부터 수용액에 대한 라만 스펙트럼 데이터가 입력되면, 라만 스펙트럼 데이터로부터 수용액에 포함된 가스의 종류에 상응하는 파장(Raman Shift)의 피크 데이터를 추출한다. 도 8 내지 도 9를 참조하면, 가스가 CO인 경우, 파장은 1948 cm^{-1} 및 2063 cm^{-1} 이다. 또한, 가스가 아세트산염(Acetate)이면 파장은 928 cm^{-1} 이고, 가스가 낙산염(Butyrate)이면, 파장은 877 cm^{-1} 이다.
- [0093] 농도예측부(300)는 S530 단계에서 예측모델(PM)을 통해 피크 데이터로부터 가스 농도의 예측값을 산출한다. 이때, 농도예측부(300)는 피크추출부(200)가 추출한 피크 데이터를 예측모델(PM)에 입력한다. 그러면, 예측모델(PM)은 피크 데이터에 대해 복수의 계층(IL, HL, OL) 간 학습된 가중치가 적용되는 복수의 연산을 통해 수용액에 포함된 가스 농도의 예측값을 산출한다.
- [0094] 다음으로, 농도예측부(300)는 S540 단계에서 예측값을 농도로 변환하여 변환된 농도를 수용액의 용존 가스 농도를 출력한다. 예컨대, 예측모델(PM)이 일산화탄소(CO)에 대해 0.089, 0.911과 같은 예측값을 산출하면, 농도예측부(300)는 예측값을 9%, 91% 등의 농도로 변환하여 일산화탄소(CO)의 용존 가스 농도를 9%, 91% 등으로 출력한다.
- [0096] 도 15는 본 발명의 실시예에 따른 컴퓨팅 장치를 나타내는 도면이다. 도 15의 컴퓨팅 장치(TN100)는 본 명세서에서 기술된 장치(예, 예측장치(PA) 등) 일 수 있다.
- [0097] 도 15의 실시예에서, 컴퓨팅 장치(TN100)는 적어도 하나의 프로세서(TN110), 송수신 장치(TN120), 및 메모리(TN130)를 포함할 수 있다. 또한, 컴퓨팅 장치(TN100)는 저장 장치(TN140), 입력 인터페이스 장치(TN150), 출력 인터페이스 장치(TN160) 등을 더 포함할 수 있다. 컴퓨팅 장치(TN100)에 포함된 구성 요소들은 버스(bus)(TN170)에 의해 연결되어 서로 통신을 수행할 수 있다.
- [0098] 프로세서(TN110)는 메모리(TN130) 및 저장 장치(TN140) 중에서 적어도 하나에 저장된 프로그램 명령(program command)을 실행할 수 있다. 프로세서(TN110)는 중앙 처리 장치(CPU: central processing unit), 그래픽 처리 장치(GPU: graphics processing unit), 또는 본 발명의 실시예에 따른 방법들이 수행되는 전용의 프로세서를 의미할 수 있다. 프로세서(TN110)는 본 발명의 실시예와 관련하여 기술된 절차, 기능, 및 방법 등을 구현하도록 구성될 수 있다. 프로세서(TN110)는 컴퓨팅 장치(TN100)의 각 구성 요소를 제어할 수 있다.
- [0099] 메모리(TN130) 및 저장 장치(TN140) 각각은 프로세서(TN110)의 동작과 관련된 다양한 정보를 저장할 수 있다. 메모리(TN130) 및 저장 장치(TN140) 각각은 휘발성 저장 매체 및 비휘발성 저장 매체 중에서 적어도 하나로 구성될 수 있다. 예를 들어, 메모리(TN130)는 읽기 전용 메모리(ROM: read only memory) 및 랜덤 액세스 메모리(RAM: random access memory) 중에서 적어도 하나로 구성될 수 있다.
- [0100] 송수신 장치(TN120)는 유선 신호 또는 무선 신호를 송신 또는 수신할 수 있다. 송수신 장치(TN120)는 네트워크에 연결되어 통신을 수행할 수 있다.
- [0101] 한편, 앞서 설명된 본 발명의 실시예에 따른 방법은 다양한 컴퓨터수단을 통하여 판독 가능한 프로그램 형태로 구현되어 컴퓨터로 판독 가능한 기록매체에 기록될 수 있다. 여기서, 기록매체는 프로그램 명령, 데이터 파일, 데이터구조 등을 단독으로 또는 조합하여 포함할 수 있다. 기록매체에 기록되는 프로그램 명령은 본 발명을 위

하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 예컨대 기록매체는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광 기록 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 장치를 포함한다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어를 포함할 수 있다. 이러한 하드웨어 장치는 본 발명의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

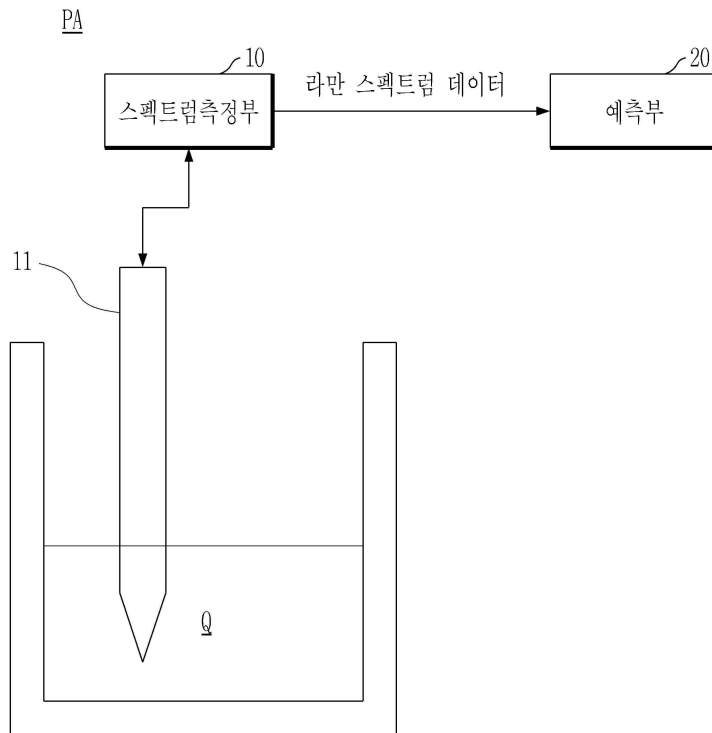
[0102] 이상 본 발명을 몇 가지 바람직한 실시예를 사용하여 설명하였으나, 이들 실시예는 예시적인 것이며 한정적인 것이 아니다. 이와 같이, 본 발명이 속하는 기술분야에서 통상의 지식을 지닌 자라면 본 발명의 사상과 첨부된 특허청구범위에 제시된 권리범위에서 벗어나지 않으면서 균등론에 따라 다양한 변화와 수정을 가할 수 있음을 이해할 것이다.

부호의 설명

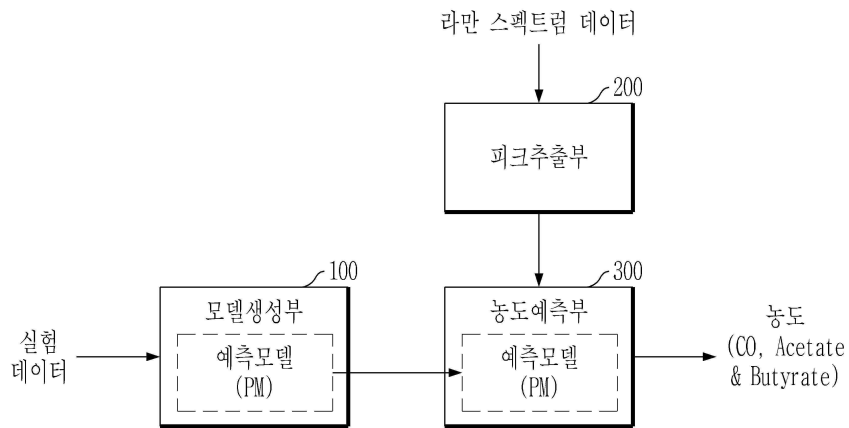
[0103] 10: 스펙트럼측정부
20: 예측부
100: 모델생성부
200: 피크추출부
300: 농도예측부

도면

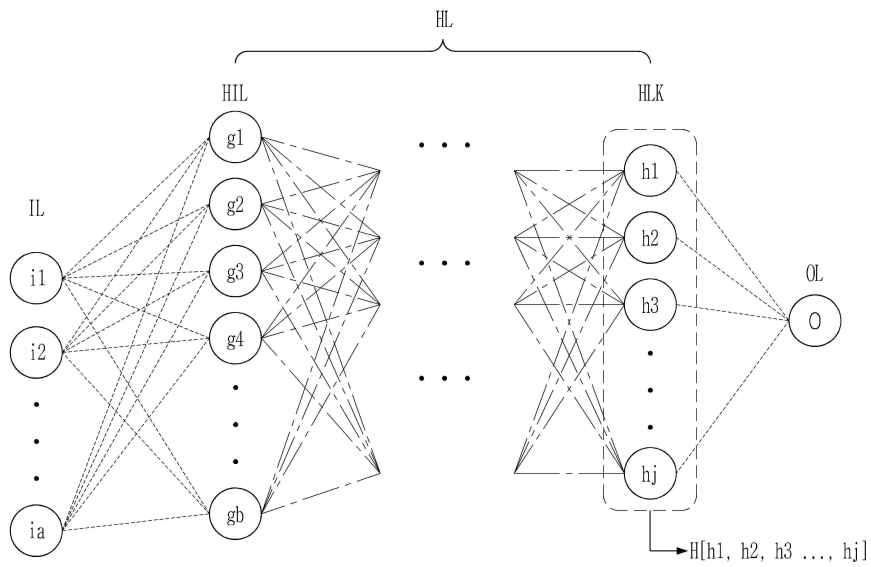
도면1



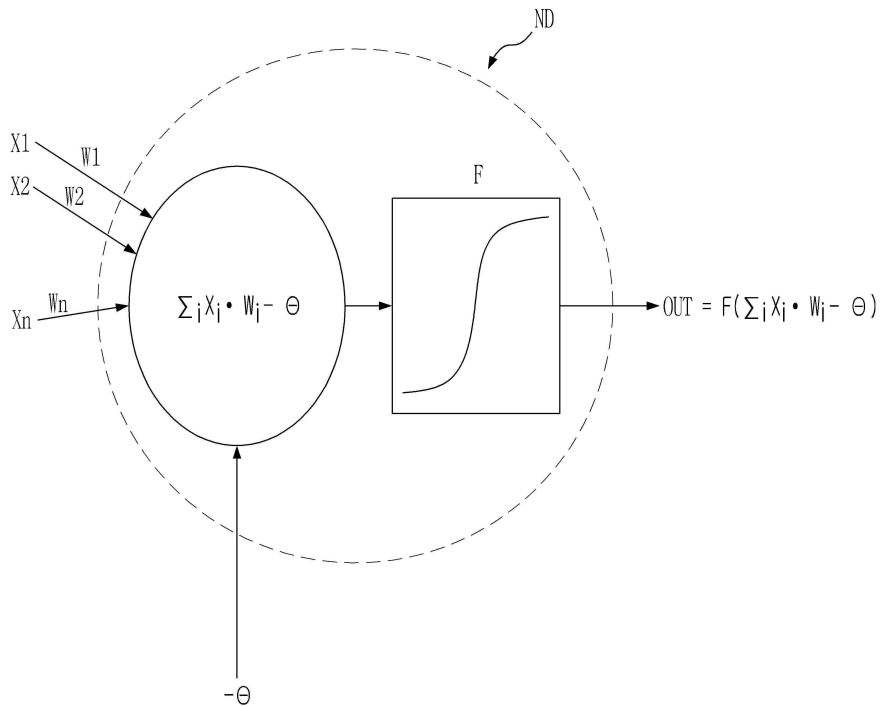
도면2



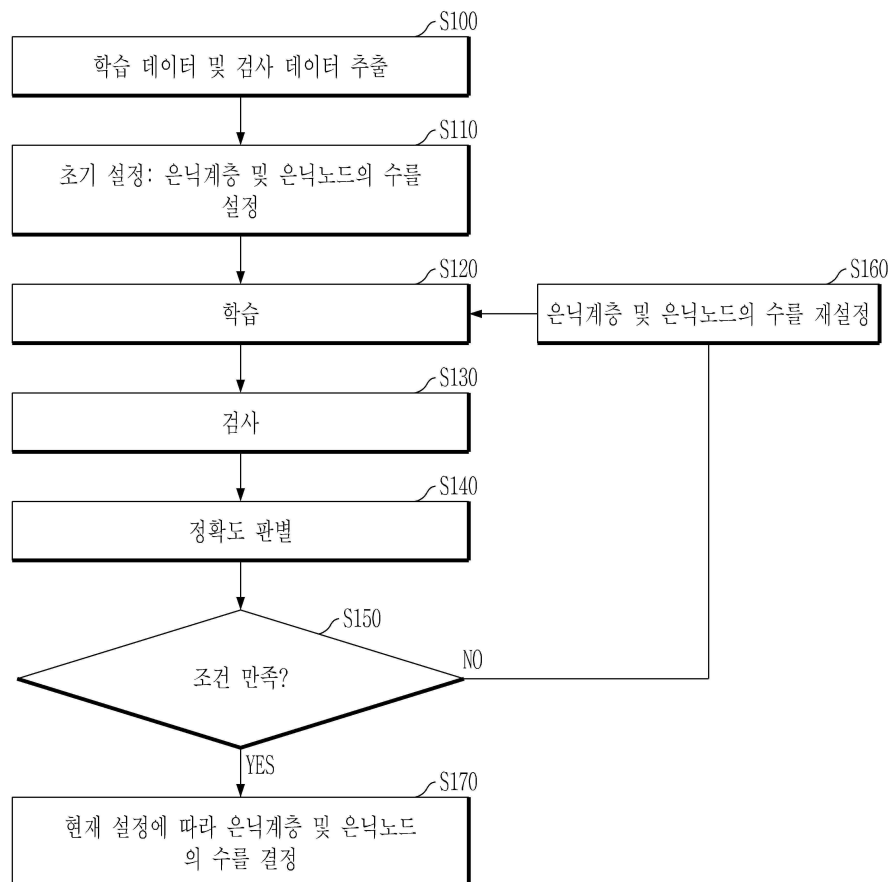
도면3



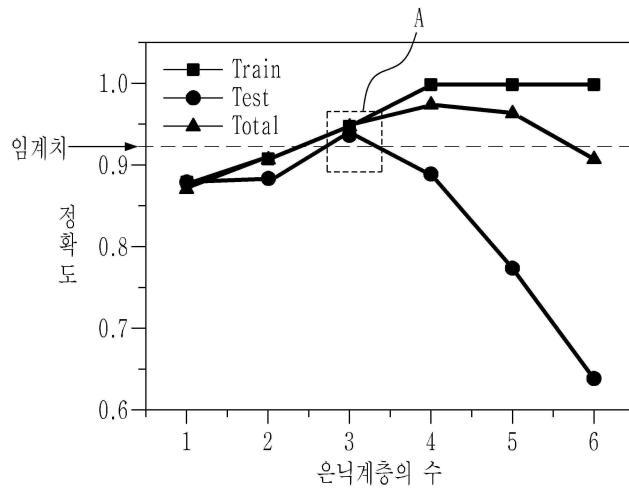
도면4



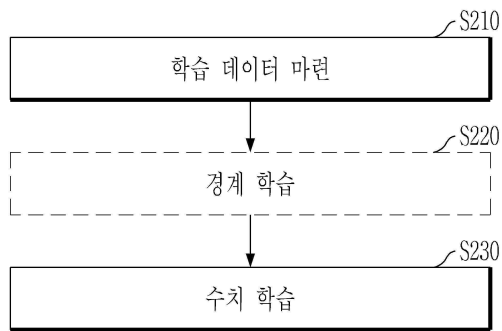
도면5



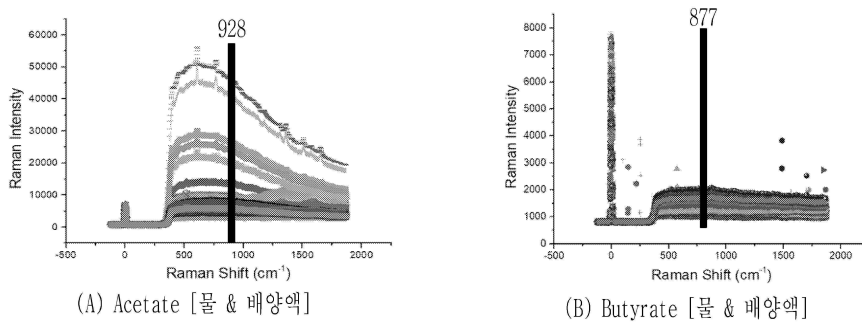
도면6



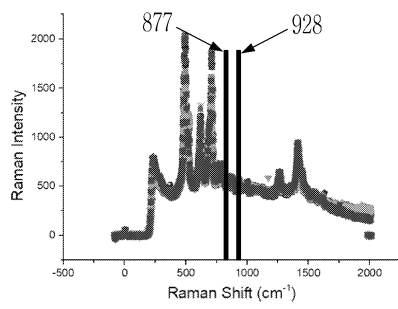
도면7



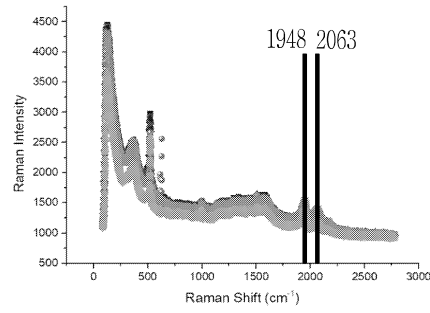
도면8



도면9

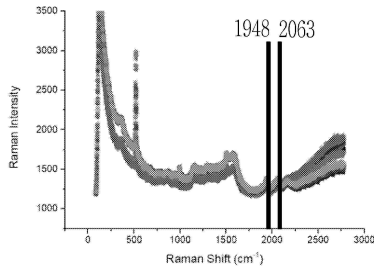


(C) Acetate & Butyrate [물]

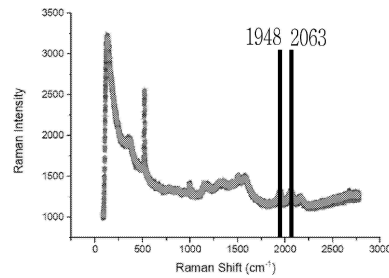


(D) CO [물]

도면10



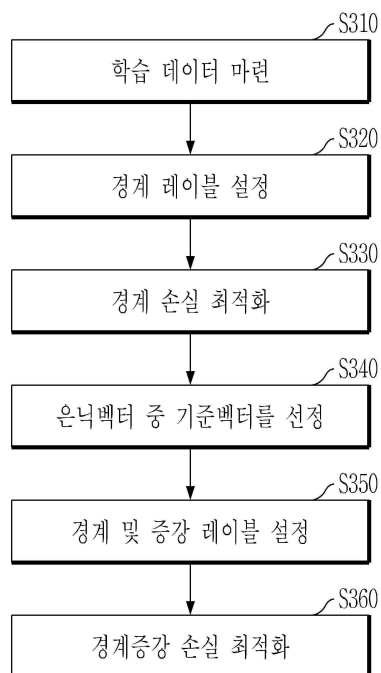
(E) CO 2.5%



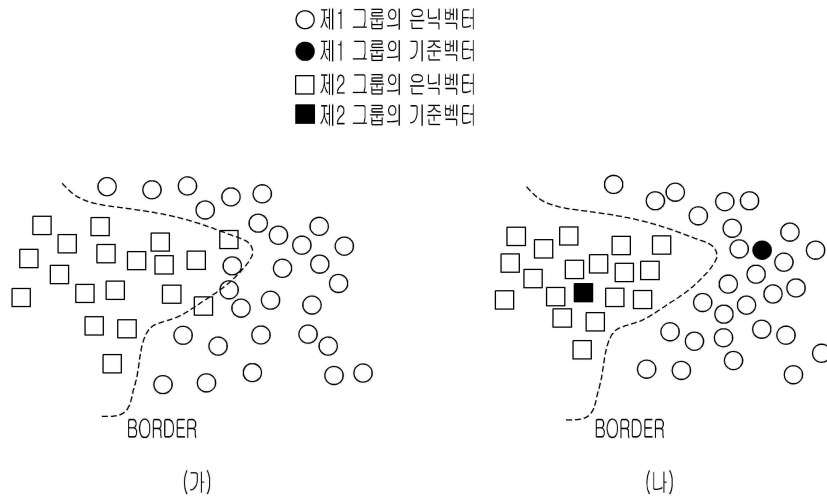
(F) CO 5%

도면11

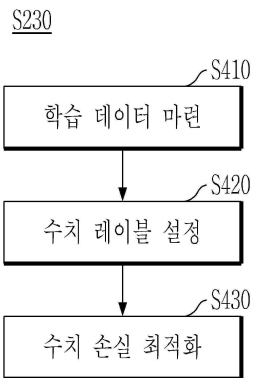
220



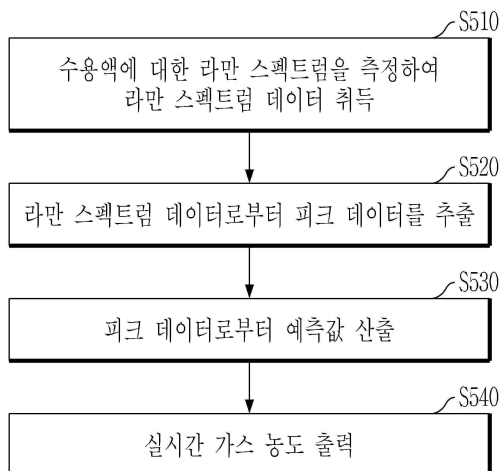
도면12



도면13



도면14



도면15

