(19)대한민국특허청(KR)
(12) 등록특허공보(B1)

(51) 。Int. Cl.[7]
G01N 21/25

(45) 공고일자 2003년09월13일
(11) 등록번호 10-0397612
(24) 등록일자 2003년08월28일

(21) 출원번호 10-2001-0025223
(22) 출원일자 2001년05월09일

(65) 공개번호 특2002-0085624
(43) 공개일자 2002년11월16일

(73) 특허권자 삼성전자주식회사
경기도 수원시 팔달구 매탄3동 416번지

(72) 발명자 김연주
경기도수원시팔달구영통동신안아파트534동1801호

윤길원
경기도수원시팔달구영통동미주아파트651동607호

(74) 대리인 이영필
이해영

심사관 : 정호근

(54) 다변량 스펙트라 분석을 통한 성분 농도 결정 방법

요약

다변량 스펙트라 분석을 통한 성분 농도 결정 방법이 개시된다. 복잡한 매질내의 관심있는 성분의 농도를 예측하는 이 방법은, 다수개의 표본들에 대해 선정된 파장 영역에서 스펙트라 및 관심있는 성분의 기준 농도를 측정하는 (a) 단계와, 스펙트라와 기준 농도를 이용하여 다변량 보정 모델을 수립하는 (b) 단계와, 새로운 표본의 스펙트라로부터 다변량 보정 모델을 이용하여 관심있는 성분의 농도를 예측하고, 관심있는 성분의 예측된 농도와 기준 농도간의 오차인 제1 예측 오차를 구하는 (c) 단계 및 제1 예측 오차가 전 파장을 이용한 제2 예측 오차보다 더 작은 값들중 가장 작은 값인가를 판단하고, 제1 예측 오차가 제2 예측 오차보다 더 작은 값들중 가장 작은 값이 아닌 것으로 판단되면, (a) 단계로 진행하는 (d) 단계를 구비하는 것을 특징으로 한다. 그러므로, 첫번째 벡터를 이용한 파장 선정법은 다변량 모델에 바탕을 두고, 파장을 줄여 가도 다음 단계의 첫 번째 적재벡터에서 각 파장간의 관계가 계속 유지될 수 있고, 짧은 계산 시간을 갖고, 평탄한 형태를 가져 smoothing을 거치지 않고도 파장 영역을 선정할 수 있는 효과를 갖는다.

대표도

도 1

명세서

도면의 간단한 설명

도 1은 본 발명에 의한 다변량 스펙트라 분석을 통한 성분 농도 결정 방법을 설명하기 위한 플로우차트이다.
도 2는 도 1에 도시된 성분 농도 결정 방법에 대한 본 발명에 의한 바람직한 일 실시예를 설명하기 위한 플로우차트이다.
도 3은 전 파장을 이용한 PLSR 보정 모델을 수립하는 과정을 설명하기 위한 플로우차트이다.

도 4는 도 1에 도시된 성분 농도 결정 방법에 대한 본 발명에 의한 바람직한 다른 실시예를 설명하기 위한 플로우차트이다.
도 5는 97명 환자들의 혈청 스펙트라를 나타내는 그래프이다.
도 6은 총 단백질량을 예측하는 전 파장 PLSR 보정 모델에 의해 얻어진 적재 벡터들을 나타내는 그래프이다.
도 7은 총 단백질량을 예측하는 전 파장 PLSR 보정 모델에 의해 얻어진 회귀 벡터들을 나타내는 그래프이다.
도 8은 총 단백질량을 예측하는 선정된 파장을 이용한 PLSR 보정 모델에 의해 얻어진 적재 벡터들을 나타내는 그래프이다.
도 9는 총 단백질량을 예측하는 선정된 파장을 이용한 PLSR 보정 모델에 의해 얻어진 회귀 벡터들을 나타내는 그래프이다.

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

본 발명은 스펙트라를 측정하여 성분 농도를 결정하는 것에 관한 것으로서, 특히, 다변량 스펙트라 분석을 통해 복잡한 매질의 성분 농도를 결정하는 방법에 관한 것이다.
복잡한 매질이란, 생물학적 매질(biological material)이나 석유 등과 같이 복잡하면서도 전부 알려지기 어려운 구성 성분들을 갖는 매질을 의미한다. 이 때, 복잡한 매질의 성분이란, 복잡한 매질이 생물학적 매질이면 글루코즈, 총단백질, 콜레스테롤, 알콜 또는 약을 의미하며, 복잡한 매질이 석유이면 각종 탄화수소를 의미한다.
한편, 스펙트라 측정에서, 일반적으로 흡수도(absorbance) 및 반사도(reflectance) 측정은 한 개 이상의 파장을 다루고, 매질의 많은 표본(sample)들에 대해 행해지며, 표본내의 성분 농도는 시약을 사용하는 습식 화학(wet chemistry) 방식에 의해 결정된다. 이 때, 습식 화학 방식에 의해 결정된 성분 농도를 기준 농도로 사용하여, 측정된 스펙트라와 성분의 기준 농도 사이의 관계를 설명하는 보정 모델을 수립한다. 따라서, 보정 모델을 이용함으로써 스펙트라 측정만으로 미지 표본의 성분 농도를 예측할 수 있다. 이와 같이, 측정된 스펙 트라에 바탕을 둔 성분 농도의 정량 분석은 전통적인 습식 화학 방식에 비해 시간, 노력 및 비용적 측면에서 이롭우며, 비파괴적(nondestructive) 또는 비침습적인(noninvasive) 측정 방식으로서 이용된다.
일반적으로 흡수 분광학을 사용한 많은 보정 모델은 비어의 법칙(Beer's Law)에 바탕을 두고 있다. 여기서, Beer's Law은, Daniel C. Harris에 의해 "Quantitative Chemical Analysis"라는 제목으로 저술된 책의 20장 spectrophotometry에서 페이지 498쪽에 개시되어 있다. 비흡수 매질에서 성분 농도가 묽은 경우에 파장 λ에서 표본의 흡수도[A(λ)]는 다음 수학식 1과 같이 표현되며 다중 성분의 농도에 의존한다.

수학식 1

$$A(\lambda)=a_{\lambda}+y_1k_1(\lambda)+y_2k_2(\lambda)+\ldots+y_pk_p(\lambda)+e_{\lambda}$$

여기서, a_{λ} 는 스펙트라의 절편을 나타내고, y_i 는 측정 매질의 i번째 구성 성분을 나타내며, $k_i(\lambda)$는 빛이 지나간 거리와 i번째 구성 성분의 파장 λ에서의 흡수 계수를 승산한 결과를 나타내고, e_{λ} 는 파장 λ에서의 측정에 의한 흡수도 오차를 나타낸다.
이하, 스펙트라 측정을 통해 매질 내의 성분 농도를 추정하는 종래의 성분 농도 결정 방법들에 대해 다음과 같이 살펴본다.
종래의 제1 성분 농도 결정 방법은 초창기의 분광학자들에 의해 이용되는 방법으로서, 보정 모델에 단지 한 파장만을 사용한다. 즉, 단변량 보정 모델(Univariate Calibration)을 이용한다. 이러한 종래의 제1 성분 농도 결정 방법은 선택된 한 파장에서 흥미 있는 성분만이 흡수도가 있고 매질내의 다른 성분은 흡수도가 없거나, 분석하고자 하는 성분의 흡수도가 다른 성분의 흡수도에 비해 상당히 큰 경우에만 이용 가능한 이용상의 한계점을 갖는다.
종래의 제2 성분 농도 결정 방법은, 다변량 보정 모델을 이용한다. 이 방법은 성분 농도와 몇 개 파장들의 선형 조합(linear combination)사이의 관계를 규정지으고, 분석하고자 하는 성분의 스펙트라 특성이 매질 내 다른 성분의 스펙트라 특성과 겹치는 경우에 특히 유용하게 적용된다. 그러나, 이 방법은 관심 성분의 성질 뿐만 아니라 매질 내 간섭 물질의 스펙트라에 대한 특별한 지식을 요구하는 문제점을 갖는다.
한편, 단지 몇 개의 성분들만을 갖는 비교적 간단한 매질의 경우에서 분광학자들은 각 성분에 특징적이고, Beer's Law를 따르는 파장 영역을 선정하였다. 그러나, 생물학적 매질과 같은 복잡한 성질의 매질을 분석할 때 파장 선정은 훨씬 더 어렵다. 왜냐하면, 물질의 구성 성분에 대해 알 수 없고, 구성 성분이 알려져 있다 하더라도 다양한 실험 조건들에 의해서 그 성분의 특징적인 스펙트라 형태가 변경되고, 성분에 대한 스펙트라 형태가 알려져 있더라도 스펙트럼이 겹쳐져서 파장 선정을 매우 어렵게 하고, 성분간의 물리적 및 화학적 간섭들이 파장을 선정하는 능력을 방해하기 때문이다. 이러한 파장 선정의 어려움을 극복하기 위해 사용되는 종래의 제3 성분 농도 결정 방법은, 매우 큰 파장 수 모두를 동시에 이용할 수 있는 다변량 보정 모델인 부분 최소 제곱 회귀(PLSR:Partial Least Squares Regression) 법

및 주요 성분 회귀(PCR:Principal Component Regression)법을 이용한다. 전 파장 보정 모델(full spectrum calibration model)을 이용하는 이 방법은 비교적 몇 개의 파장만을 이용하는 방법에 비해 신호의 averaging 및 outlier 분별 등의 많은 장점을 갖고, 많은 파장을 사용할 수 있기 때문에 파장 선정의 필요나 간섭 성분에 대한 스펙트라 지식을 필요로 하지 않고, 아주 많은 예를 들면, 종종 수천에 이르는 파장 수들을 사용할 수 있다. Hall-JW 등에 의해 1992년도에 "Near-infrared spectrophotometry: a New dimension in clinical chemistry"라는 제목으로 발표되고, clinical chemistry라는 잡지의 Vol 38, No 9, 페이지 1623-1631쪽에 실린 논문을 참조하면, 혈청내의 글루코즈, 총단백, 알부민 등의 농도를 근적외선 스펙트라만으로 알아내는 이 방법에 대해 알 수 있다. 이 논문에 개시된 방법은 성분 농도를 추정하기 위해 단변량 보정 모델에서 단지 몇 개의 파장만을 이용하는 다변량 모델과 마지막으로 PLSR을 이용하였다. PLSR은 혈청내의 다른 성분에 비해 양이 작고 흡수 파장의 선택성이 가장 작은 글루코즈에 대한 예측 모델로 이용되고 있다. 대부분의 분광학자들은 전 파장 보정 모델을 사용할 때 사용 가능한 모든 파장을 다 이용한다. 그러나, 많은 경우에 다수의 스펙트라 측정은 어렵기도 하고 또한 어떤 파장들은 비 선형성을 갖기 때문에 모델에 포함시키기가 어려워 진다. PLS 따위의 전파장 보정 모델은 어느 정도 비 선형성을 수용하지만 관계없는 파장 영역을 포함할 수도 있기 때문에, 보정 모델의 예측력을 심각하게 떨어뜨릴 수 있는 문제점을 갖는다.

종래의 제4 성분 농도 결정 방법은, 측정이 어렵거나 관계없는 스펙트라 영 역들을 때때로 전처리 단계에서 제거한다. 즉, 전처리 단계에서 스펙트라 정보가 부족한 영역, 상당히 비선형성을 갖는 영역 및 신호 대 잡음비(SNR)가 작은 영역 등이 제거된다. 이러한 간단한 전 처리는 잘 알려져 있으며, 예전부터 사용되어 왔다. 생물학적 매질의 근적외선 스펙트라의 경우에 용매인 물에 의한 흡수도가 상대적으로 몹시 크고, 흡수도가 다양한 실험 조건에 따라 민감하게 반응한다. 따라서, 전처리 단계에서 물의 흡수 파장 영역을 제외시킨다. Marbach-R 등에 의해 1993년도에 "Noninvasive Blood glucose Assay by Near-Infrared Diffuse Refelcetance Spectroscopy of the Human Inner Lip"라는 제목으로 발표되고, Applied Spectroscopy란 잡지의 Vol 47, No 7, 페이지 875-881쪽에 실린 논문을 참조하면, 이 방법은 글루코즈를 잘 예측하기 위해서 물과 지질의 흡수 파장 영역을 제외하고 다변량 모델을 적용하였다. 이러한 전처리 이후에 전 파장 보정 모델이 사용되면, 파장 선정은 추천되지 않거나, 필요한 것으로 여겨지지 않는다. 그러나, 이러한 종래의 방법은 종종 관계 없는 파장 영역을 포함하여 모델의 예측력을 심각하게 떨어뜨릴 수 있으므로, 사용되는 PLSR 다변량 보정 모델에 따라 적절한 파장 선정 방법을 이용하는 것이 좋다. 주로 성분 농도를 예측하는 경우에 사용되는 PLSR 다변량 보정 모델은 Martens-H 및 Næs-T의 Chichester에 의해 1989년도에 "Multivariate Calibration"라는 제목으로 John wiley Sons 출판사에 의해 출간된 책에 개시되어 있다.

종래의 제5 성분 농도 결정 방법은 미국 특허 번호 US5,857,462에 "Systematic Wavelength Selection for improved multivariate spectral analysis" 라는 제목으로 개시되어 있다. 이 특허에 개시된 종래의 방법에 의하면, PLSR 다변량 모델을 만들 때 쓸모없는 파장을 포함하는 경우 모델의 예측력은 성분 농도의 평균값에서 멀어지는 표본에서 급격히 떨어짐을 알 수 있다. 특히 글루코즈 농도 추정의 경우 80 mg/deciliter 이하의 낮은 농도의 예측이 특히 중요한데, 낮은 농도는 평균 값에서 멀어지므로 관계없는 파장이 포함된 경우 예측 오차를 크게 한다. PLSR을 이용한 글루코즈 진단기의 개발을 위해서 적절한 파장의 선정이 특히 중요하다. 이외에, 이 방법은 체계적인 파장 선정법으로서 유전자 모델을 개시하고 있다. 그러나, 이 유전자 모델에 의한 파장 선정법은 파장을 선정하는 데 많은 계산 시간을 요구하며, 카운트 스펙트럼(count spectrum)을 얻기 위한 유전자 모델 과정 중에 결정해 주어야 하는 요소들을 가지며, 파장을 선정하는데 기준이 되는 카운트 스펙트럼이 잡음에 민감해서 smoothing 과정을 수행해야 하는 문제점들을 갖는다.

종래의 제6 성분 농도 결정 방법에서 파장을 선정하는 손쉬운 방식은 상관 관계 그래프를 이용하는 방법이다. 상관 관계 그래프는 각 파장에서의 흡수도와 관심 성분의 농도 사이의 상관계수(correlation coefficient:r)를 표시한다. 이 방법의 PLSR은 다변량 보정 모델인 데 반해 단변량 상관계수만을 고려하여 파장을 선정하므로 PLSR에서 이용하는 다변량 사이의 관계를 적절히 고려할 수 없는 문제점을 갖는다.

종래의 제7 성분 농도 결정 방법에서 파장을 선정하는 방법은 PLSR 보정 모델 결과 얻어지는 회귀 벡터(regression vector)를 이용하는 것이다. 그러나, 회귀 벡터의 경우 사용한 인자 수에 따라서 현저히 다른 값을 갖는다. 즉, 최적의 인자수가 달라짐에 따라 선정되는 파장이 상이하게 될 가능성이 많다. 또, 최적의 인자수에 의해 얻어진 회귀 벡터로 파장을 선정한 후, 선정된 파장을 이용하여 PLSR 농도 추정 모델을 세웠을 때 얻어지는 회귀벡터는 전 파장(full spectrum)을 이용한 회귀 벡터와 다른 양상을 보인다. 따라서 파장을 단계적으로 더 줄일 필요가 있는 경우, 어느 단계의 회귀 벡터를 사용해야 할 지 판단하기가 어려운 문제점을 갖는다.

발명이 이루고자 하는 기술적 과제

본 발명이 이루고자 하는 기술적 과제는, 향상된 예측력을 갖는 보정 모델을 이용하여 복잡한 매질의 성분 농도를 빠른 시간내에 결정할 수 있는 다변량 스펙트라 분석을 통한 성분 농도 결정 방법을 제공하는 데 있다.

발명의 구성 및 작용

상기 과제를 이루기 위해, 복잡한 매질내에 포함된 관심있는 성분의 농도를 예측하는 본 발명에 의한 다변량 스펙트라 분석을 통한 성분 농도 결정 방법은, 상기 매질내의 다수개의 표본들에 대해 선정된 파장 영역에서 스펙트라 및 상기 관심있는 성분의 기준 농도를 측정하는 (a) 단계와, 상기 스펙트라와 상기 기준 농도를 이용하여 다변량 보정 모델을 수립하는 (b) 단계와, 새로운 표본의 스펙트라로부터 상기 다변량 보정 모델을 이용하여 관심있는 성분의 농도를 예측하고, 상기 관심있는 성분의 예측된 농도와 상기 기준 농도간의 오차인 제1 예측 오차를 구하는 (c) 단계 및 상기

제1 예측 오차가 전 파장을 이용한 제2 예측 오차보다 더 작은 값들 중 가장 작은 값인가를 판단하고, 상기 제1 예측 오차가 상기 제2 예측 오차보다 더 작은 값들중 가장 작은 값이 아닌 것으로 판단되면, 상기 (a) 단계로 진행하는 (d) 단계로 이루어지는 것이 바람직하다.

이하, 본 발명에 의한 다변량 스펙트라 분석을 통한 성분 농도 결정 방법을 첨부한 도면들을 참조하여 다음과 같이 설명한다.

도 1은 본 발명에 의한 다변량 스펙트라 분석을 통한 성분 농도 결정 방법을 설명하기 위한 플로우차트로서, 측정한 스펙트라와 관심있는 성분의 기준 농도를 이용하여 다변량 보정 모델을 수립하는 단계(제10 및 제12 단계들), 관심있는 성분의 농도를 예측하고, 예측된 농도와 기준 농도간의 오차인 제1 예측 오차를 구하는 단계(제14 단계) 및 제1 예측 오차와 제2 예측 오차간의 크기를 비교하는 단계(제16 단계)로 이루어진다.

본 발명에 의한 성분 농도 결정 방법은, 복잡한 매질내에 포함된 관심있는 성분의 농도를 예측하기 위해, 먼저, 다수개의 표본들에 대해 선정된 파장 영역에서 스펙트라 및 관심있는 성분의 기준 농도를 측정한다(제10 단계). 이 때, 본 발명에 따르면, 복잡한 매질이란, 전술한 바와 같이 복잡하면서 모두가 알려지기 어려운 구성 성분들을 갖는 생물학적 매질이나 석유 같은 것을 의미한다. 또한, 관심있는 성분이란, 글루코즈, 알부민, 콜레스테롤, 알코올, 약 또는 탄화수소류가 될 수 있다.

제10 단계후에, 스펙트라와 기준 농도를 이용하여 다변량 보정 모델을 수립한다(제12 단계). 여기서, 다변량 보정 모델은 부분 최소 제곱 회귀(PLSR)를 의미 한다.

제12 단계후에, 새로운 표본의 스펙트라로부터 제12 단계에서 수립한 다변량 보정 모델을 이용하여 관심있는 성분의 농도를 예측하고, 관심있는 성분의 예측된 농도와 기준 농도간의 오차인 제1 예측 오차를 구한다(제14 단계).

제14 단계후에, 제1 예측 오차가 전 파장을 이용한 제2 예측 오차보다 더 작은 값들중 가장 작은 값인가를 판단한다(제16 단계). 만일, 제1 예측 오차가 제2 예측 오차보다 더 작은 값들중 가장 작은 값이 아닌 것으로 판단되면, 제10 단계로 진행한다.

이하, 도 1에 도시된 다변량 스펙트라 분석을 통한 성분 농도 결정 방법에 대한 본 발명에 의한 바람직한 실시예들을 첨부한 도면들을 참조하여 다음과 같이 설명한다.

도 2는 도 1에 도시된 성분 농도 결정 방법에 대한 본 발명에 의한 바람직한 일 실시예를 설명하기 위한 플로우차트로서, 제10 단계를 수행하는 단계(제28 제34 단계들), 제12 단계를 수행하는 단계(제36 단계), 제14 단계를 수행하는 단계(제38 제42 단계들) 및 제16 단계를 수행하는 단계(제44 단계)로 이루어진다.

도 2에 도시된 성분 농도 결정 방법은 도 1에 도시된 제10 단계를 다음과 같이 수행한다(제28 제34 단계들). 먼저, s(≥2)개의 표본들에 대해 n(≥2)개의 파장들에서 스펙트라(X)를 측정하고, 표본들에 대해 관심있는 성분의 기준 농도(y)를 습식 화학 방식으로 측정한다(제28 단계).

제28 단계후에, 측정된 스펙트라(X)로부터 스펙트라(X)의 평균값[mean(X)]을 다음 수학식 2와 같이 감산하여 X의 평균 보정값(X^c)을 구하고, 기준 농도(y)로부터 기준 농도(y)의 평균값[mean(y)]을 다음 수학식 2와 같이 감산하여 y의 평균 보정값(y^c)을 구한다(제30 단계).

수학식 2

$$X^c = X - mean(X)$$

$$y^c = y - mean(y)$$

제30 단계후에, 전 파장을 이용한 부분 최소 제곱 회귀(PLSR) 보정 모델을 수립하여 각 인자(factor)에 대한 적재 벡터(loading vector)를 구한다(제32 단계).

이하, 전 파장을 이용한 PLSR 보정 모델을 수립하는 과정을 첨부한 도면을 참조하여 다음과 같이 설명한다.

도 3은 전 파장을 이용한 PLSR 보정 모델을 수립하는 과정을 설명하기 위한 플로우차트로서, 스펙트라와 기준 농도의 평균 보정값들을 구하는 단계(제60 및 제62 단계들), 스펙트라의 평균 보정값을 새로운 인자로 변수 변환하고 농도 추정 모델을 세우는 단계(제64 및 제66 단계) 및 새로운 표본의 스펙트라로부터 관심 있는 성분의 농도를 예측하는 단계(제68 단계)로 이루어진다.

도 3을 참조하면, 먼저, s개의 표본들에 대해 n개의 파장들에서 다음 수학식 3과 같이 표현되는 스펙트라(X)를 측정하고, s개의 표본들에 대한 관심있는 성분의 다음 수학식 4와 같이 표현되는 기준 농도(y)를 습식 화학 방식으로 측정한다(제60 단계).

수학식 3

$$X=\{x_1,x_2,...,x_j,...,x_n\}=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{in} \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ x_{s1} & x_{s2} & \cdots & x_{sj} & \cdots & x_{sn} \end{bmatrix}$$

수학식 4

$$y=\{y_1,y_2,\ldots,y_s\}$$

x_j가 갖는 원소의 수 및 y가 갖는 원소의 수는 각각 s개이고, X가 갖는 원소의 수는 n ×s개임을 알 수 있다. 제60 단계후에, 스펙트라(X)와 기준 농도(y)의 평균 보정값들(X^c 및 y^c)을 전술한 수학식 2와 같이 구한다(제62 단계). 제62 단계에서 구한 평균 보정값들(X^c 및 y^c)을 고려하여 x_j^c의 선형 조합으로부터 새로운 인자(factor)들(f_1, f_2, ..., f_k)로 이루어진 F를 다음 수학식 5와 같이 생성한다(제64 단계).

수학식 5

$$F=\{f_1,f_2,\ldots,f_k\}$$

여기서, f_1 및 f_k는 다음 수학식 6과 같이 각각 표현된다.

수학식 6

$$f_1=a_{11}x_1^c+a_{12}x_2^c+\ldots+a_{1n}x_n^c$$
$$f_k=a_{k1}x_1^c+a_{k2}x_2^c+\ldots+a_{kn}x_n^c$$

이 때, 인자들(f_1, f_2, ..., f_k)의 설명력은 다음 수학식 7과 같이 표현되고, 첫 번째 인자(f_1)에 대한 적재 벡터(l_1)는 다음 수학식 8과 같이 표현된다.

수학식 7

$$f_1 \ggg f_2 \gg \ldots > f_k$$

수학식 8

$$l_1=\{a_{11},a_{12},\ldots,a_{1n}\}$$

제64 단계후에, 농도 추정 모델을 세운다(제66 단계). 먼저, 농도 추정 모델은 수학식 3에 표현된 X를 이용하여 다음 수학식 9와 같이 표현될 수도 있고, 수학식 5에 표현된 F를 이용하여 다음 수학식 10과 같이 표현될 수도 있고, 여러 잡음 요소에서 영향을 덜 받도록 F에 포함되는 인자들중 최적의 v(여기서, v≤k)개수의 인자들만을 포함하여 다음 수학식 11과 같이 표현될 수도 있다.

수학식 9

$$y^c=b_1x_1^c+b_2x_2^c+\ldots+b_nx_n^c$$

수학식 10

$$y^c=B_1f_1+B_2f_2+\ldots+B_kf_k$$

수학식 11

$$\mathbf{y}^{c} \fallingdotseq \mathbf{B}_1\mathbf{f}_1+\mathbf{B}_2\mathbf{f}_2+\ldots+\mathbf{B}_v\mathbf{f}_v$$

이 때, v개의 인자들만을 포함하는 경우의 회귀 벡터(regression vector)(b_v)는 다음 수학식 12와 같이 표현된다.

수학식 12

$$\mathbf{b}_v=\{b_1,b_2,\ldots,b_n\}$$

제66 단계후에, 농도 추정 모델을 이용하여 새로운 표본의 스펙트라(x_{new})로부터 관심있는 성분의 농도(y_{new})를 다음 수학식 13과 같이 예측한다(제68 단계).

수학식 13

$$\mathbf{y}^{c}_{new}=\mathbf{b}\mathbf{x}^{c}_{new}$$

여기서, x_{new}는 다음 수학식 14와 같으므로, 관심있는 성분의 농도(y_{new})는 다음 수학식 15와 같이 표현될 수 있다.

수학식 14

$$\mathbf{x}_{new}=\{\ x_{new1}{}^{c},\ x_{new2}{}^{c},\ \ldots,\ x_{newn}{}^{c},\}$$

수학식 15

$$y_{new}{}^{c}=b_1 x_{new1}{}^{c}+b_2 x_{new2}{}^{c}+\ldots+b_n x_{newn}{}^{c}$$

전술한 수학식들 2 15에서, 이탤릭체의 소문자는 스칼라(scalar)량을 표시하고, 이탤릭체가 아닌 소문자는 벡터(vector)량을 표시하고, 대문자는 매트릭스(matrix)를 각각 표시한다.
한편, 제32 단계후에, 첫 번째 인자에 대한 적재 벡터(l_1)를 이용하여 적재 벡터의 절대값이 큰 파장을 선정한다(제34 단계).
이와 같이 도 2에 도시된 방법은 도 1에 도시된 제10 단계를 수행한 다음, 제12 단계를 수행하기 위해, 제34 단계에서 선정된 파장을 이용하여 부분 최소 제곱 회귀(PLSR) 보정 모델을 전술한 도 3에 도시된 과정을 통해 다시 수립하여 다변량 보정 모델을 구한다(제36 단계).
이 때, 도 2에 도시된 방법은 다음과 같이 도 1에 도시된 제14 단계를 수행한다(제38 제42 단계들).
제36 단계후에, s개의 표본들에서 s-1개의 표본만을 이용하여 다변량 보정 모델이 구해졌을 때 나머지 한 개의 표본을 예측 자료로서 이용하여 관심있는 성분의 농도(y)를 예측한다(제38 단계). 제38 단계후에, 모든 표본들에 대해 관심있는 성분의 농도가 예측되었는가를 판단한다(제40 단계). 만일, 제38 단계가 모든 표본들에 대해 수행되지 않은 것으로 판단되면 제38 단계로 진행한다. 그러나, 제38 단계가 모든 표본들에 대해 수행된 것으로 판단되면, 즉, 모든 표본들에 대해 관심있 는 성분의 농도가 예측되었으면, 각 표본에 대해 예측된 농도와 기준 농도간의 오차를 구하고, 구해진 오차들을 평균하여 제1 예측 오차를 구한다(제42 단계).
전술한 바와 같이 도 2에 도시된 방법은 도 1에 도시된 제14 단계를 수행한 후, 제16 단계를 수행하기 위해, 제1 예측 오차가 제2 예측 오차보다 더 작은 값들중 가장 작은 값인가를 판단한다(제44 단계). 만일, 제1 예측 오차가 제2 예측 오차보다 작은 값들중 가장 작은 값이 아닌 것으로 판단되면, 제34 단계로 진행한다.
결국, 제1 예측 오차가 제2 예측 오차보다 작고 최소가 될 때까지 제34 제42 단계들이 반복적으로 수행되어 선정된 파장의 수는 점점 더 줄어들게 된다.
도 4는 도 1에 도시된 성분 농도 측정 방법에 대한 본 발명에 의한 바람직한 다른 실시예를 설명하기 위한 플로우차트로서, 제10 단계를 수행하는 단계(제78 제84 단계들), 제12 단계를 수행하는 단계(제86 단계), 제14 단계를 수행하는 단계(제88 및 제90 단계들) 및 제16 단계를 수행하는 단계(제92 단계)로 이루어진다.
도 4에 도시된 다변량 스펙트라 분석을 통한 성분 농도 결정 방법은 도 1에 도시된 제10 단계를 다음과 같이 수행한다(제78 제84 단계들). 먼저, s개의 표본들에 대해 n개의 파장들에서 스펙트라(X)를 측정하고, 표본들에 대해 관심있는 성분의 기준 농도를 습식 화학 방식으로 측정한다(제78 단계). 이 때, s개의 표본들은 c개의 보정 표본들과 p개의 예측 표본들로 나뉜다. 따라서, 보정 자료는 c개의 보정 표본들에 대한 n개의 파장들에서의 스펙트라(X_c)와 관심있는 성분의 기준 농도(y_c)로 이루어지고, 에측 자료는 p개의 예측 표본들에 대한 n개의 파장들에서의 스펙트라(X_p)와 관심있는 성분의 기준 농도(y_p)로 이루어진다.
제78 단계후에, s개의 표본들중에서 c개의 보정 표본들에 대해 n개의 파장들에서 측정된 스펙트라(X_c)로부터 스펙트라(X_c)의 평균값[mean(X_c)]을 다음 수학식 16과 같이 감산하여 스펙트라(X_c)의 평균 보정값($X_c{}^{c}$)을 구하

고, 기준 농도(y_c)로부터 기준 농도(y_c)의 평균값[mean(y_c)]을 다음 수학식 16과 같이 감산하여 기준 농도(y_c)의 평균 보정값(y_c^c)을 구한다(제80 단계).

수학식 16

$$X_c^c = X_c - \text{mean}(X_c)$$

$$y_c^c = y_c - \text{mean}(y_c)$$

제80 단계후에, 전 파장을 이용한 부분 최소 제곱 회귀(PLSR) 보정 모델을 도 3에 도시된 바와 같이 수립하여 각 인자에 대한 적재 벡터(l_1 l_k)를 구한다(제82 단계). 제82 단계후에, 첫 번째 인자(f_1)에 대한 적재 벡터(l_1)를 이용하여 적재 벡터의 절대값이 큰 파장을 선정한다(제84 단계).

이와 같이 도 4에 도시된 방법은 도 1에 도시된 제10 단계를 수행한 다음, 제12 단계를 수행하기 위해, 제84 단계에서 선정된 파장을 이용하여 부분 최소 제 곱 회귀(PLSR) 보정 모델을 도 3에 도시된 바와 같은 과정을 거쳐 다시 수립하여 다변량 보정 모델을 구한다(제86 단계).

이 때, 도 4에 도시된 방법은 다음과 같이 도 1에 도시된 제14 단계를 수행한다(제88 및 제90 단계들).

제86 단계후에, p개의 표본들에 대해 n개의 파장들에서 측정된 스펙트라(X_p)로부터 전 파장을 이용한 PLSR 보정 모델 및 선정된 파장을 이용한 PLSR 보정 모델을 이용하여 각 표본에 대한 관심있는 성분의 농도를 예측한다(제88 단계). 제88 단계후에, 각 표본에 대해 예측된 농도와 기준 농도간의 오차를 구하고, 구해진 오차들을 평균하여 제1 예측 오차를 구한다(제90 단계).

전술한 바와 같이 도 4에 도시된 방법은 도 1에 도시된 제14 단계를 수행한 후, 제16 단계를 수행하기 위해, 제1 예측 오차가 제2 예측 오차보다 작은 값들중 가장 작은 값인가를 판단한다(제92 단계). 만일, 제1 예측 오차가 제2 예측 오차보다 작은 값들중 가장 작은 값이 아닌 것으로 판단되면, 제84 단계로 진행한다.

결국, 제1 예측 오차가 제2 예측 오차보다 작고 최소가 될 때까지 제34 제42 단계들이 반복적으로 수행되어 선정된 파장의 수는 점점 더 줄어들게 된다.

이하, 도 2에 도시된 본 발명에 의한 다변량 스펙트라 분석을 통한 성분 농도 결정 방법을 시뮬레이션한 결과들을 첨부한 도면들을 참조하여 다음과 같이 설명한다.

도 5는 97명 환자들의 혈청 스펙트라를 나타내는 그래프로서, 횡축은 파장을 나타내고, 종축은 흡수도를 각각 나타낸다. 도 5에 도시된 그래프에서 하나의 선은 하나의 스펙트럼을 나타낸다.

97명 환자들로부터 얻은 혈청으로부터 화학 분석기(Chemistry Analyzer)를 이용하여 총 단백질량(total protein:g/deciliter)을 결정한 다음, 혈청의 일부를 가지고 1530~1850nm의 파장 영역에서 2nm의 해상도(resolution)로 각 파장에 대한 흡수도(absorbance:a.u.)를 측정하면 도 5에 도시된 바와 같은 결과를 얻을 수 있다. 총 단백질량은 5.1~7.7 g/deciliter의 분포를 나타내며, 0.5mm 두께의 큐벳에 혈청을 넣어 분광 광도계로부터 스펙트럼을 얻었다. 스펙트럼은 이중 빔 방식으로 기준 빔 쪽에 공기를 기준으로 각 혈청의 흡수 스펙트라를 얻었다. 결국, 도 5는 도 2에 도시된 제28 단계의 수행결과이다.

도 6은 총 단백질량을 예측하는 전 파장 PLSR 보정 모델에 의해 얻어진 적재 벡터들(l_1 l_4)을 나타내는 그래프로서, 횡축은 파장을 나타내고, 종축은 적재 벡터(l_i)를 각각 나타낸다.

전 파장(1530~1850 nm)을 사용한 PLSR 보정 모델을 도 3에 도시된 바와 같이 수립하면, 도 6에 도시된 바와 같이 제1 인자(f_1)에서 제4 인자(f_4)까지 적재 벡터들(l_1 l_4)이 나타낸다. 제1 인자(f_1)에 대한 적재 벡터(l_1)의 절대값이 큰 영역을 고려하면 1672~1754 nm의 파장 영역이 선정된다. 도 6은 도 2에 도시된 제30 및 제32 단계들의 수행결과이다. 최적의 인자 수(v)는 4개로 결정 되었다. 교차 타당성의 제1 예측 오차는 0.28 g/deciliter로 얻어졌다.

도 7은 총 단백질량을 예측하는 전 파장 PLSR 보정 모델에 의해 얻어진 회귀 벡터들(b_3 b_5)을 나타내는 그래프로서, 횡축은 파장을 나타내고, 종축은 회귀 벡터(b_i)를 각각 나타낸다.

전 파장 PLSR 보정 모델의 결과 인자 수가 3, 4, 5개일 때 회귀 벡터는 도 7에 도시된 바와 같이 나타난다. 도 7을 참조하면, 보정 모델에 포함되는 인자 수가 달라짐에 따라 회귀 벡터들이 현저히 다른 양상을 보인다. 최적의 인자수로 결정된 회귀벡터(b_4)의 경우, 상당히 잡음에 민감한 양상을 나타낸다.

도 8은 총 단백질량을 예측하는 선정된 파장을 이용한 PLSR 보정 모델에 의해 얻어진 적재 벡터들(l_1 l_4)을 나타내는 그래프로서, 횡축은 파장을 나타내고, 종축은 적재 벡터(l_i)를 각각 나타낸다.

도 8은 도 6에 도시된 첫 번째 적재 벡터(l_1)를 고려하여 도 2의 제34 단계에서 선정한 파장 영역으로 PLSR 보정 모델을 세운 결과로서, 선정된 파장 영역은 1672~1754 nm에 해당한다.

선정된 파장 영역 PLSR 결과 얻어진 제1 인자(f_1)에서 제4 인자(f_4)까지의 적재 벡터들(l_1 l_4)은 도 8에 도시된 바와 같이 나타난다. 도 6에 도시된 첫 번째 적재 벡터(l_1)의 파장에 따른 특성이 도 8의 첫 번째 적재 벡터(l_1)에 그대로 유지 되는 것을 알 수 있다. 그러나, 나머지 적재 벡터들(l_2, l_3 및 l_4)은 도 6과 도 8에 서 각기 다른 양상을 보이고 있다. 최적의 인자 수는 4개로 결정되었고, 교차 타당성의 제1 예측 오차는 0.27 g/deciliter로 얻어졌다.

도 9는 총 단백질량을 예측하는 선정된 파장을 이용한 PLSR 보정 모델에 의해 얻어진 회귀 벡터들(b_3 b_5)을 나타내는 그래프로서, 횡축은 파장을 나타내고, 종축은 휘귀 벡터(b_i)를 각각 나타낸다.

선정된 파장(1672~1754nm)을 이용한 PLSR 보정 모델을 얻고, 인자의 수를 각기 3, 4, 5개 포함했을 때 회귀 벡터들(b_3 b_5)은 도 9에 도시된 바와 같이 나타난다. 도 7과 마찬가지로 회귀 벡터들(b_3 b_5) 각각은 파장에 따라 다른 양상을 보이고 있다. 또한, 도 7에 도시된 회귀벡터들과 도 9에 도시된 회귀 벡터들은 다른 양상을 보인다. 전 파장(1530~1850nm: 161개 파장) PLSR 보정 모델의 교차 타당성에 대한 제1 예측 오차는 0.28 g/deciliter 이고, 첫 번째 적재 벡터(l_1)에 의해 선정된 파장 영역(1672~1754 nm: 82개 파장) PLSR 보정 모델의 교차 타당성에 대한 제1 예측 오차는 0.27g/deciliter로 얻어졌다. 두 PLSR 보정 모델의 제1 예측 오차는 거의 같다. 따라서 혈청 스펙트라로부터 1530~1850 nm 파장 영역을 전부 사용하기 보다는 1672~1754 nm 파장 영역을 이용한 PLSR 보정 모델로 총 단백질량을 예측할 수 있다. 첫 번째 적재벡터(l_1)를 기준으로 한 파장 선정법은 손쉽고, PLSR 보정 모델에 근거를 두고 있으며, 파장을 더 줄여야 하는 경우, 첫 번째 적재 벡터(l_1)가 전 파장을 이용한 경우와 선정된 파장을 이용한 경우 파장에 따라서 동일한 특성을 나타내므로 어떤 단계의 첫 번째 적재 벡터(l_1)를 기준으로 파장을 선정하여도 무방하다.

결국, 본 발명에 의한 다변량 스펙트라 분석을 통한 농도 결정 방법은, 스펙트라를 측정하고, 측정한 스펙트라를 이용하여 복잡한 매질내의 성분 농도를 추정하는 다변량 보정 모델을 세울 때, 전 파장 영역 대신에 특정한 파장 영역을 선정하므로서 다변량 보정 모델의 예측력을 향상시킬 수 있음을 알 수 있다.

발명의 효과

이상에서 설명한 바와 같이, 본 발명에 의한 다변량 스펙트라 분석을 통한 성분 농도 결정 방법은 PLSR 보정 모델의 결과 얻어지는 적재 벡터를 이용하여 파장을 선정하고, 특히 각각의 변환된 인자에 따라 존재하는 적재 벡터들중 가장 설명력 있는 변수인 첫 번째 적재벡터를 PLSR 농도 추정 모델에서 파장을 선정하는데 고려하므로, 상관 관계 그래프를 이용한 파장 선정법이 단변량에 근거를 두는 종래의 방법과 달리 다변량 모델에 바탕을 둔다. 또한, 회귀 벡터를 사용하는 종래의 파장 선정법에서 파장을 줄인 후에 회귀 벡터가 전 단계 회귀 벡터와 현저히 달라서 파장을 더 줄여야 하는 경우 파장 선정의 기준이 어느 단계의 회귀 벡터를 사용해야 할지를 결정하기 어려운 반면, 본 발명에 의한 방법은 파장을 줄여 가도 다음 단계의 첫 번째 적재벡터에서 각 파장간의 관계가 계속 유지될 수 있다. 게다가, 본 발명에 의한 방법은 종래의 유전자 모델에 비해 짧은 계산 시간을 갖고, 평탄한 형태를 가지는 첫 번째 적재 벡터를 이용하므로 smoothing을 거치지 않고도 파장 영역을 선정할 수 있는 효과를 갖는다.

(57) 청구의 범위

청구항 1.

복잡한 매질내에 포함된 관심있는 성분의 농도를 예측하는 성분 농도 결정 방법에 있어서,

(a) 상기 매질의 다수개의 표본들에 대해 선정된 파장 영역에서 스펙트라 및 상기 관심있는 성분의 기준 농도를 측정하는 단계;

(b) 상기 스펙트라와 상기 기준 농도를 이용하여 다변량 보정 모델을 수립하는 단계;

(c) 새로운 표본의 스펙트라로부터 상기 다변량 보정 모델을 이용하여 관심있는 성분의 농도를 예측하고, 상기 관심있는 성분의 예측된 농도와 상기 기준 농도간의 오차인 제1 예측 오차를 구하는 단계; 및

(d) 상기 제1 예측 오차가 전 파장을 이용한 제2 예측 오차보다 더 작은 값들중 가장 작은 값인가를 판단하고, 상기 제1 예측 오차가 상기 제2 예측 오차보다 더 작은 값들중 가장 작은 값이 아닌 것으로 판단되면, 상기 (a) 단계로 진행하는 단계를 구비하는 것을 특징으로 하는 다변량 스펙트라 분석을 통한 성분 농도 결정 방법.

청구항 2.

제1 항에 있어서, 상기 (a) 단계는

(a11) 상기 다수개의 표본들에 대해 n(≥2)개의 파장들에서 상기 스펙트라를 측정하고, 상기 표본들에 대해 상기 기준 농도를 측정하는 단계;

(a12) 상기 스펙트라(X)로부터 스펙트라(X)의 평균값[mean(X)]을 감산하여 X의 평균 보정값(X_c)을 구하고, 상기 기준 농도(y)로부터 상기 기준 농도(y)의 평균값[mean(y)]을 감산하여 y의 평균 보정값(y_c)을 구하는 단계;

(a13) 전 파장을 이용한 부분 최소 제곱 회귀(PLSR) 보정 모델을 수립하여 각 인자에 대한 적재 벡터를 구하는 단계; 및

(a14) 첫 번째 상기 인자에 대한 적재 벡터를 이용하여 상기 적재 벡터의 절대값이 큰 파장을 선정하는 단계를 구비하여 상기 파장 영역을 선정하고,

상기 (d) 단계는 상기 제1 예측 오차가 상기 제2 예측 오차보다 더 작은 값들중 가장 작은 값이 아닌 것으로 판단되면, 상기 (a14) 단계로 진행하는 것을 특징으로 하는 다변량 스펙트라 분석을 통한 성분 농도 결정 방법.

청구항 3.

제2 항에 있어서, 상기 (b) 단계는

상기 (a14) 단계에서 선정된 파장을 이용하여 상기 부분 최소 제곱 회귀 보정 모델을 다시 수립하여 상기 다변량 보정 모델을 구하고, 상기 (c)단계로 진행하는 것을 특징으로 하는 다변량 스펙트라 분석을 통한 성분 농도 결정 방법.

청구항 4.

제3 항에 있어서, 상기 (c) 단계는

(c11) 상기 (b) 단계후에, s(≥2)개의 상기 표본들에서 s-1개의 표본만을 이 용하여 상기 다변량 보정 모델이 구해졌

을 때 나머지 한 개의 표본을 예측 자료로서 이용하여 상기 관심있는 성분의 농도를 예측하는 단계;
(c12) 상기 (c11) 단계가 모든 표본들에 대해 수행되었는가를 판단하여, 상기 (c11) 단계가 상기 모든 표본들에 대해 수행되지 않은 것으로 판단되면 상기 (c11) 단계로 진행하는 단계; 및
(c13) 상기 (c11) 단계가 상기 모든 표본들에 대해 수행된 것으로 판단되면, 상기 각 표본에 대해 예측된 농도와 상기 기준 농도간의 차를 구하고, 구해진 차들을 평균하여 상기 제1 예측 오차를 구하고, 상기 (d) 단계로 진행하는 단계를 구비하는 것을 특징으로 하는 다변량 스펙트라 분석을 통한 성분 농도 결정 방법.

청구항 5.

제1 항에 있어서, 상기 (a) 단계는
(a21) 상기 다수개의 표본들에 대해 n(≥2)개의 파장들에서 상기 스펙트라를 측정하고, 상기 표본들에 대해 상기 기준 농도를 측정하는 단계;
(a22) s(≥2)개의 상기 표본들중에서 c개의 보정에 관한 표본들에 대해 측정된 스펙트라(X_c)로부터 스펙트라(X_c)의 평균값[mean(X_c)]을 감산하여 X_c의 평균 보정값(X_c^c)을 구하고, 상기 기준 농도(y_c)로부터 상기 기준 농도(y_c)의 평균값[mean(y_c)]을 감산하여 y_c의 평균 보정값(y_c^c)을 구하는 단계;
(a23) 전 파장을 이용한 부분 최소 제곱 회귀(PLSR) 보정 모델을 수립하여 각 인자에 대한 적재 벡터를 구하는 단계; 및
(a24) 첫 번째 상기 인자에 대한 적재 벡터를 이용하여 상기 적재 벡터의 절대값이 큰 파장을 선정하는 단계를 구비하여, 상기 파장을 선정하고,
상기 (d) 단계는 상기 제1 예측 오차가 상기 제2 예측 오차보다 더 작은 값들중 가장 작은 값이 아닌 것으로 판단되면, 상기 (a24) 단계로 진행하는 것을 특징으로 하는 다변량 스펙트라 분석을 통한 성분 농도 결정 방법.

청구항 6.

제5 항에 있어서, 상기 (b) 단계는
상기 (a24) 단계에서 선정된 파장을 이용하여 상기 부분 최소 제곱 회귀 보정 모델을 다시 수립하여 상기 다변량 보정 모델을 구하고, 상기 (c)단계로 진행하는 것을 특징으로 하는 다변량 스펙트라 분석을 통한 성분 농도 결정 방법.

청구항 7.

제6 항에 있어서, 상기 (c) 단계는
(c21) 상기 (b) 단계후에, p(=s-c)개의 상기 표본들에 대해 n개의 파장들에서 측정된 스펙트라(X_p)로부터 전 파장을 이용한 PLSR 보정 모델 및 선정된 파장을 이용한 PLSR 보정 모델을 이용하여 각 표본에 대한 상기 관심있는 성분의 농도를 예측하는 단계; 및
(c23) 상기 각 표본에 대해 예측된 농도와 상기 기준 농도간의 차를 구하고, 구해진 차들을 평균하여 상기 제1 예측 오차를 구하며, 상기 (d) 단계로 진행하는 단계를 구비하는 것을 특징으로 하는 다변량 스펙트라 분석을 통한 성분 농도 결정 방법.

청구항 8.

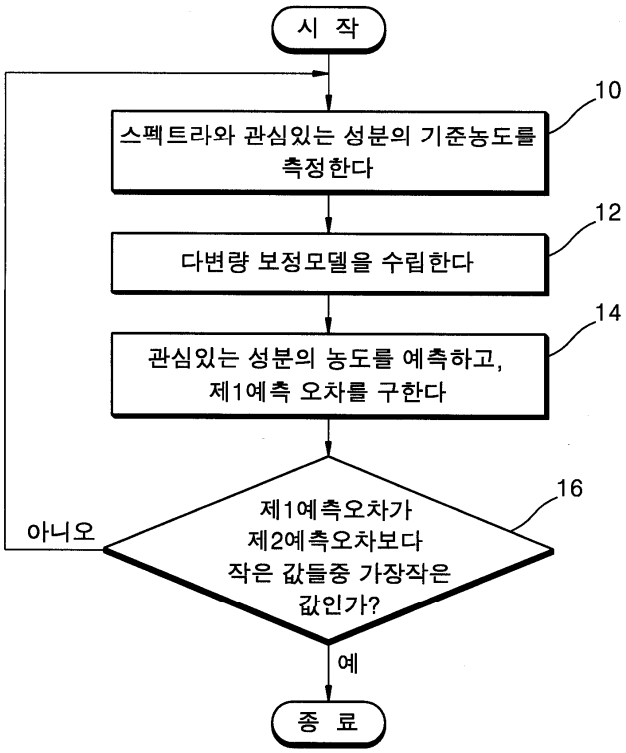
제1 항에 있어서, 상기 매질은 생물학적 매질이나 석유에 해당하는 것을 특징으로 하는 다변량 스펙트라 분석을 통한 성분 농도 결정 방법.

청구항 9.

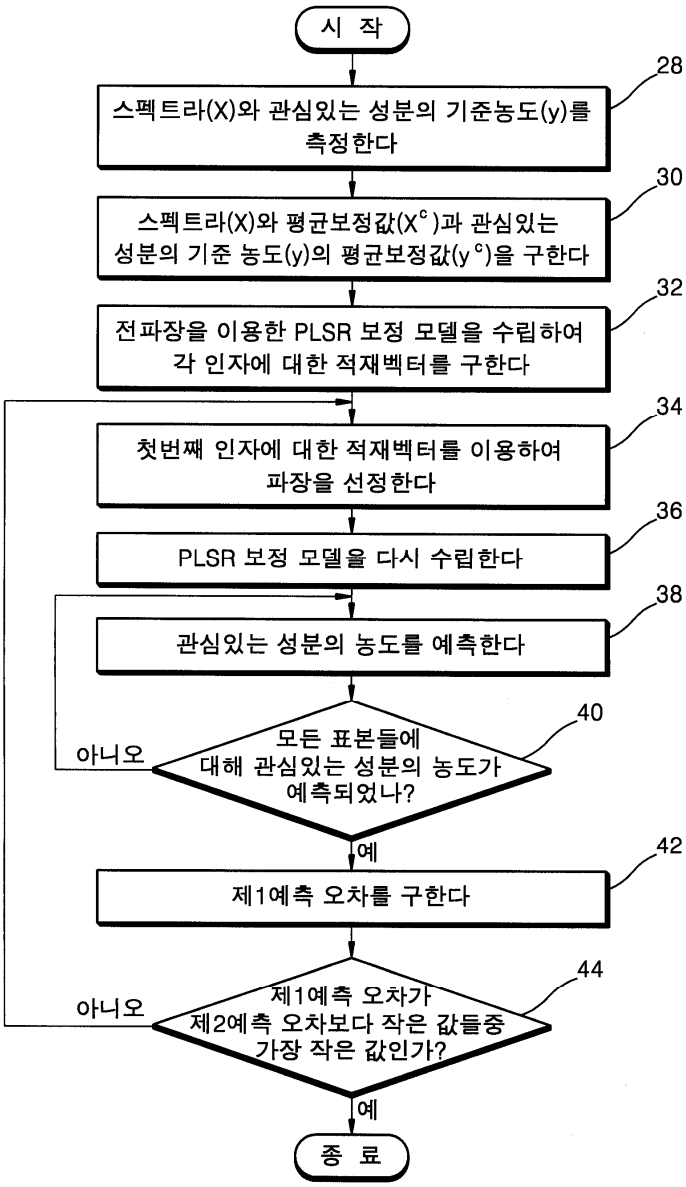
제8 항에 있어서, 상기 관심있는 성분은 글루코즈, 알부민, 콜레스테롤, 알코올, 약 또는 탄화수소류에 해당하는 것을 특징으로 하는 다변량 스펙트라 분석을 통한 성분 농도 결정 방법.

도면

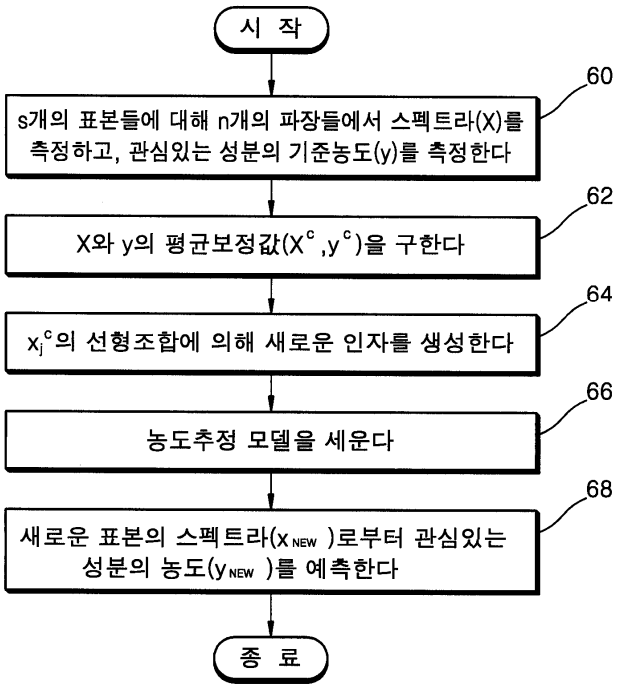
도면1


시 작
스펙트라와 관심있는 성분의 기준농도를 측정한다
10
다변량 보정모델을 수립한다
12
관심있는 성분의 농도를 예측하고, 제1예측 오차를 구한다
14
제1예측오차가 제2예측오차보다 작은 값들중 가장작은 값인가?
16
아니오
예
종 료

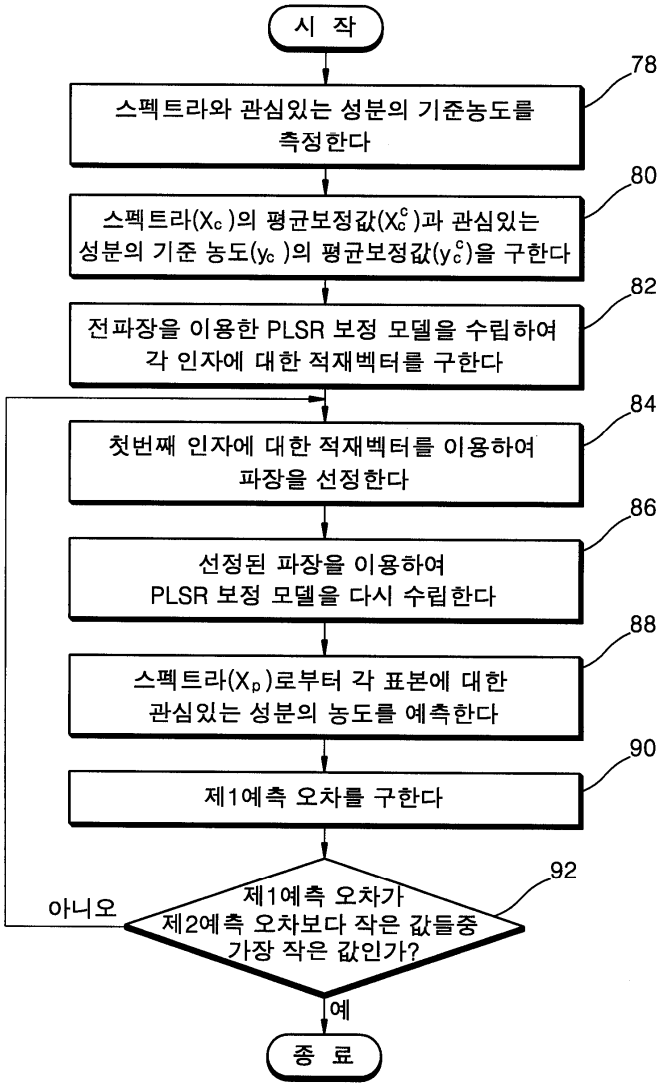
도면2


시 작
스펙트라(X)와 관심있는 성분의 기준농도(y)를 측정한다
28
스펙트라(X)와 평균보정값(X^c)과 관심있는 성분의 기준 농도(y)의 평균보정값(y^c)을 구한다
30
전파장을 이용한 PLSR 보정 모델을 수립하여 각 인자에 대한 적재벡터를 구한다
32
첫번째 인자에 대한 적재벡터를 이용하여 파장을 선정한다
34
PLSR 보정 모델을 다시 수립한다
36
관심있는 성분의 농도를 예측한다
38
모든 표본들에 대해 관심있는 성분의 농도가 예측되었나?
40
아니오
예
제1예측 오차를 구한다
42
제1예측 오차가 제2예측 오차보다 작은 값들중 가장 작은 값인가?
44
아니오
예
종 료

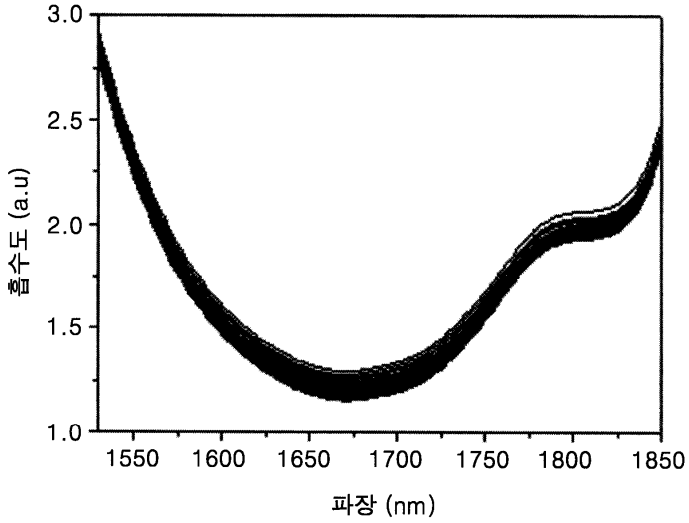
도면3


시 작
s개의 표본들에 대해 n개의 파장들에서 스펙트라(X)를 측정하고, 관심있는 성분의 기준농도(y)를 측정한다
60
X와 y의 평균보정값(X^c,y^c)을 구한다
62
x_j^c의 선형조합에 의해 새로운 인자를 생성한다
64
농도추정 모델을 세운다
66
새로운 표본의 스펙트라(x_{NEW})로부터 관심있는 성분의 농도(y_{NEW})를 예측한다
68
종 료

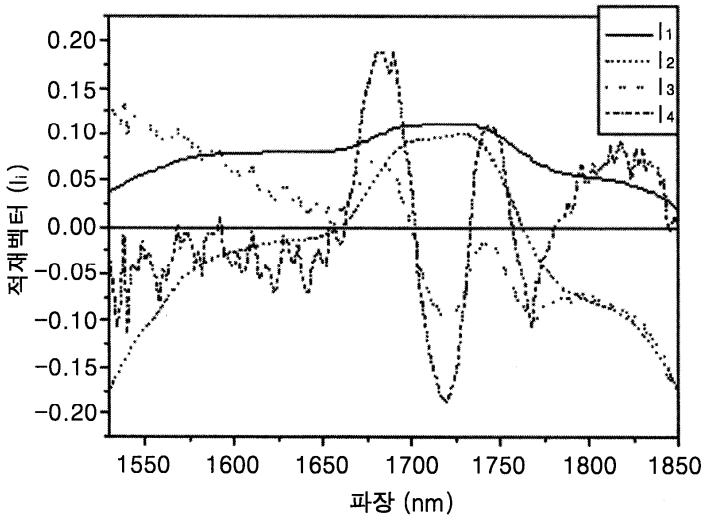
도면4


시 작
스펙트라와 관심있는 성분의 기준농도를 측정한다
78
스펙트라(X_c)의 평균보정값(X_c^c)과 관심있는 성분의 기준 농도(y_c)의 평균보정값(y_c^c)을 구한다
80
전파장을 이용한 PLSR 보정 모델을 수립하여 각 인자에 대한 적재벡터를 구한다
82
첫번째 인자에 대한 적재벡터를 이용하여 파장을 선정한다
84
선정된 파장을 이용하여 PLSR 보정 모델을 다시 수립한다
86
스펙트라(X_p)로부터 각 표본에 대한 관심있는 성분의 농도를 예측한다
88
제1예측 오차를 구한다
90
제1예측 오차가 제2예측 오차보다 작은 값들중 가장 작은 값인가?
92
아니오
예
종 료

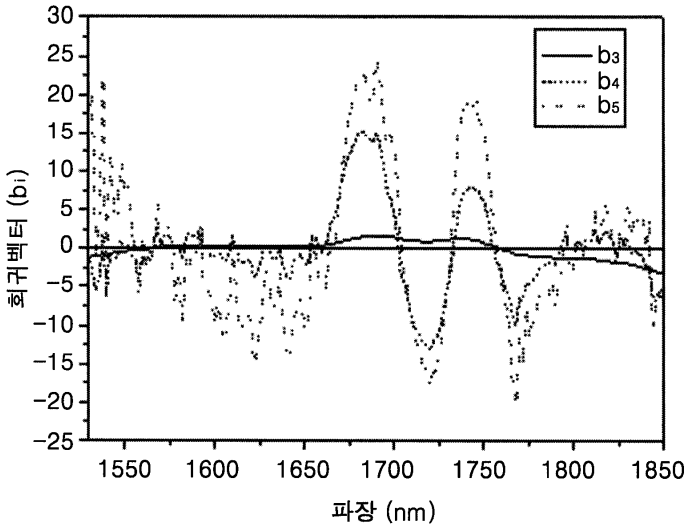
도면5


흡수도 (a.u)
3.0
2.5
2.0
1.5
1.0
1550
1600
1650
1700
1750
1800
1850
파장 (nm)

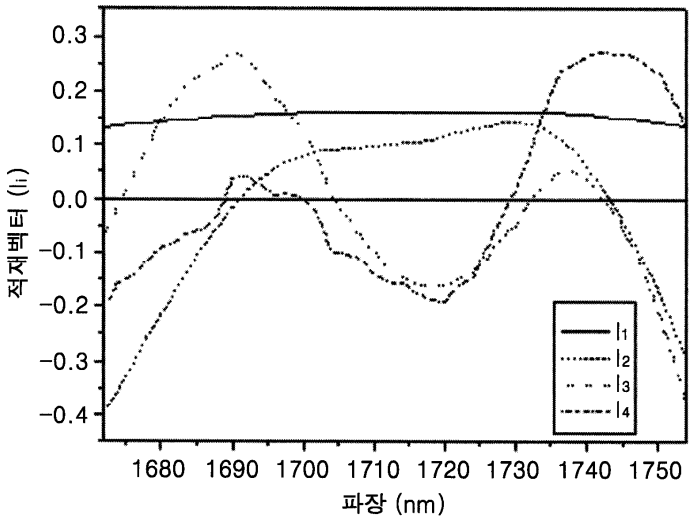
도면6


적재벡터 (li)
0.20
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
−0.20
l1
l2
l3
l4
1550
1600
1650
1700
1750
1800
1850
파장 (nm)

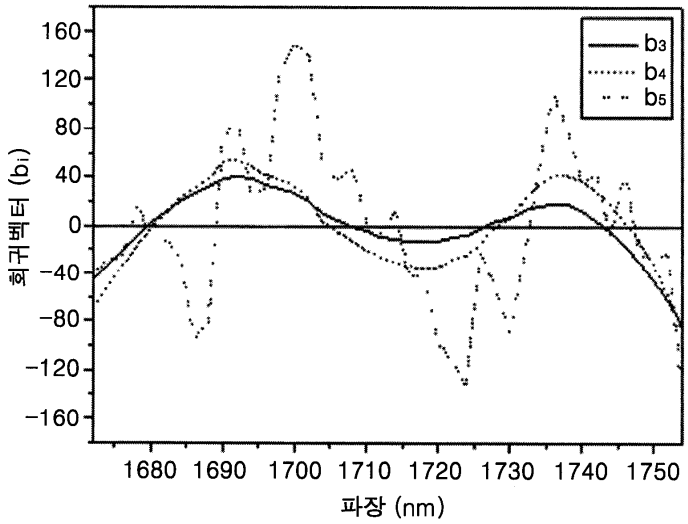
도면7


회귀벡터 (bi)
30
25
20
15
10
5
0
−5
−10
−15
−20
−25
b3
b4
b5
1550
1600
1650
1700
1750
1800
1850
파장 (nm)

도면8


0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
적재벡터 (li)
1680 1690 1700 1710 1720 1730 1740 1750
파장 (nm)
l1
l2
l3
l4

도면9


160
120
80
40
0
−40
−80
−120
−160
회귀벡터 (bi)
1680 1690 1700 1710 1720 1730 1740 1750
파장 (nm)
b3
b4
b5