(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2022年12月29日(29.12.2022)

WIPO | PCT

(10) 国際公開番号
WO 2022/270289 A1

(51) 国際特許分類:
G01N 27/62 (2021.01)

(21) 国際出願番号： PCT/JP2022/022813

(22) 国際出願日： 2022年6月6日(06.06.2022)

(25) 国際出願の言語： 日本語

(26) 国際公開の言語： 日本語

(30) 優先権データ：
特願 2021-104510 2021年6月24日(24.06.2021) JP

(71) 出願人: 国立研究開発法人物質・材料研究機構 (**NATIONAL INSTITUTE FOR MATERIALS SCIENCE**) [JP/JP]; 〒3050047 茨城県つくば市千現一丁目２番地１ Ibaraki (JP).

(72) 発明者: 内 藤 昌 信 (**NAITO Masanobu**); 〒3050047 茨城県つくば市千現一丁目２番地１ 国立研究開発法人物質・材料研究機構内 Ibaraki (JP). 日比 裕理(**HIBI Yusuke**); 〒3050047 茨城県つくば市千現一丁目２番地１ 国立研究開発法人物質・材料研究機構内 Ibaraki (JP).

(74) 代理人: 金山 義信，外(**KANEYAMA Yoshinobu** et al.); 〒3001207 茨城県牛久市ひたち野東４丁目７－３ブリーズノース１０１ Ibaraki (JP).

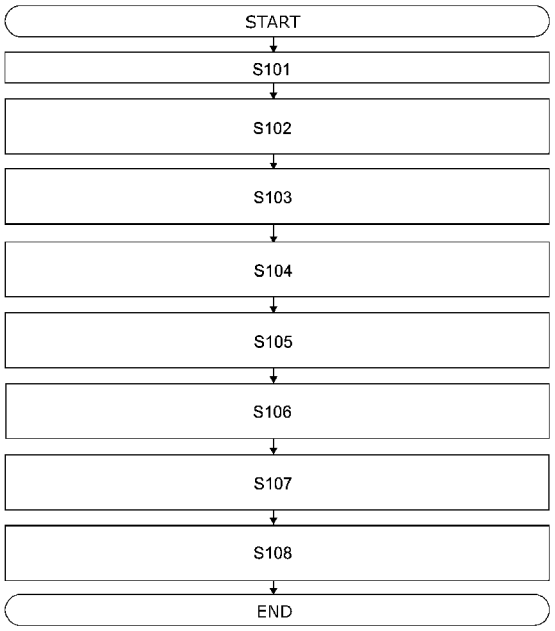
(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

WO 2022/270289 A1

(54) **Title:** METHOD FOR ESTIMATING CONTENT RATIO OF COMPONENTS CONTAINED IN SAMPLE, COMPOSITION ESTIMATING DEVICE, AND PROGRAM

(54) 発明の名称： 検体に含まれる成分の含有量比の推定方法、組成推定装置、及び、プログラム

[図1]



(57) **Abstract:** The method of the present invention is a method for estimating the content ratio of components in a sample containing components selected from among K types, and is capable of estimating the composition of the components of an unknown mixture, even from mass spectra obtained under ambient conditions, the method including: heating a sample set, sequentially ionizing the resulting gas components, and continuously observing a mass spectrum; obtaining a two-dimensional mass spectrum for each sample from the mass spectrum, and combining a plurality thereof to obtain a data matrix; performing non-negative matrix factorization of the data matrix; correcting an intensity distribution matrix by means of a canonical correlation analysis between a basal spectral matrix and the data matrix; obtaining a feature vector from the corrected intensity distribution matrix and representing the same in a vector space; setting a K-1 dimensional simple substance and determining an end member; and estimating the content ratio of the components in the sample from the end member and the feature vector.

[続葉有]

(84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

一 国際調査報告(条約第21条(3))

(57) 要約:本発明の方法は、Ｋ種類から選択される成分を含む検体における、成分の含有量比を推定する方法であって、検体セットを加熱し、生ずるガス成分を順次イオン化し、マススペクトルを連続的に観測することと、マススペクトルから、検体ごとの二次元マススペクトルを得て、これらを複数まとめてデータ行列を得ることと、データ行列を非負値行列因子分解することと、基底スペクトル行列とデータ行列との正準相関分析によって強度分布行列を補正することと、補正後強度分布行列から特徴ベクトルを得て、ベクトル空間内で表現することと、Ｋ－１次元単体を設定し、エンドメンバーを決定することと、エンドメンバーと特徴ベクトルとから検体中の成分の含有量比を推定することと、を含み、アンビエントな条件で得られるマススペクトルからであっても、未知の混合物の成分の組成を推定することができる。

明　細　書

発明の名称 ：

検体に含まれる成分の含有量比の推定方法、組成推定装置、及び、プログラム

技術分野

[0001] 本発明は、検体に含まれる成分の含有量比の推定方法、組成推定装置、及び、プログラムに関する。

背景技術

[0002] 従来、多成分の混合系である検体における成分組成を推定するための方法としては、例えば、検体中の各成分を何らかの方法で分離して、それぞれの分画成分を同定、定量する方法が知られている。この同定方法の一つとして、質量分析装置によって得られるマススペクトルが使われることがあった。

[0003] ところが、マススペクトルから得られる情報は、多数の相関関係を有する膨大な多次元データであるため、成分の同定につながる特徴的な信号を直感的、又は、経験的に抽出するのは困難であった。近年、非特許文献１に記載されているとおり、これらの多次元データを主成分分析によって次元削減し、特徴を探索する方法が用いられてきている。しかし、これら既知のデータ解析手法を用いても、ピーク強度がイオン化効率によって左右されるマススペクトルを用いて定量を行うことは本質的に困難であった。定量分析を行うためには、各化合物ごと有する未知変数であるイオン化効率をまず求める必要があり、そのためには検体量が既知、かつイオン化効率が同じとなる同位体検体を標準物質として用いる必要があった。

先行技術文献

非特許文献

[0004] 非特許文献1：Ａｎａｌｙｔｉｃａｌ　ｃｈｅｍｉｓｔｒｙ，２０２０，ｖｏｌ．９２，ｉｓｓｕｅ２，Ｐ．１９２５－１９３３

発明の概要

発明が解決しようとする課題

[0005] 非特許文献１に記載された方法は、次元削減の過程で、もとの多次元データが有する多くの情報が失われ、成分の定性的な分析はできても、定量分析を行うことはできなかった。

[0006] また、そもそも、成分の種類や試験環境等によって、成分ごとのイオン化効率が異なるという特性がある以上、得られるマススペクトルの大きさ（面積）の比は、混合物中における各成分の含有量の比を反映せず、これがマススペクトルによる定量分析の妨げの一つの要因となっていた。

[0007] 近年、質量分析方法の一つとして、アンビエントな条件（大気圧の開放空間にある）にある試料を前処理なしで質量分析するアンビエント脱離イオン化法（ＡＤＩ）が広く用いられるようになってきている。この方法は、未処理の試料や開放系の試料を固体のまま分析できるという利点がある一方で、各成分のイオン化率の不一致、及び、測定環境からの影響等によって、定量分析がより困難だと考えられてきた。

[0008] 上記事情を考慮し、本発明はアンビエントな条件で得られるマススペクトルからであっても、未知の混合物の成分の組成を推定することができる、成分の含有量比の推定方法を提供することを課題とする。

また、本発明は、組成推定装置、及び、プログラムを提供することも課題とする。

課題を解決するための手段

[0009] 本発明者らは、上記課題を達成すべく鋭意検討した結果、以下の構成により上記課題を達成することができることを見出した。

[0010] ［１］ Ｋを１以上の整数としたとき、Ｋ種類の成分から選択される少なくとも１種の上記成分を含む推定対象検体における、上記成分の含有量比を推定する方法であって、上記Ｋ種類の成分から選択される少なくとも１種の成分を含み、互いに組成の異なるＫ個以上の学習用検体、及び、上記成分を含まないバックグラウンド検体を準備することと、上記推定対象検体、上記学習用検体、及び、バックグラウンド検体を含む検体セットに含まれるそれ

ぞれの検体を加熱しながら、熱脱着、及び／又は、熱分解によって生ずるガス成分を順次イオン化し、マススペクトルを連続的に観測することと、加熱温度ごとに得られた上記マススペクトルを各行に格納し、上記検体ごとの二次元マススペクトルを得て、上記二次元マススペクトルの少なくとも２つ以上をまとめてデータ行列に変換することと、上記データ行列を非負値行列因子分解し、規格化された基底スペクトル行列とその強度分布行列の積に分解するＮＭＦ処理を行うことと、上記基底スペクトル行列と上記データ行列との正準相関分析によって強度分布行列に含まれるノイズ成分を抽出し、上記ノイズ成分の影響を減ずるよう上記強度分布行列を補正して、補正後強度分布行列を得ることと、上記補正後強度分布行列を上記検体のそれぞれに対応する小行列に分割し、上記小行列を特徴ベクトルとして、ベクトル空間内で上記検体のそれぞれを表現することと、上記特徴ベクトルの全てを内包するＫ－１次元単体を設定し、上記Ｋ－１次元単体におけるＫ個のエンドメンバーを決定することと、上記Ｋ個のエンドメンバーと、上記推定対象検体の特徴ベクトルとの、ユークリッド距離をそれぞれ計算し、上記ユークリッド距離の比により、上記推定対象検体中の上記成分の含有量比を推定することと、を含み、上記Ｋが３以上の場合、上記Ｋ－１次元単体に内接する超球体の外側の領域のそれぞれに、上記学習用検体の特徴ベクトルの少なくとも１つが位置する、又は、上記学習用検体が上記エンドメンバーの少なくとも１つを含む、方法。

［２］　上記Ｋが２以上の整数であり、上記推定対象検体が、上記成分の混合物である、［１］に記載の方法。

［３］　上記学習用検体が、上記エンドメンバーを含む、［１］又は［２］に記載の方法。

［４］　上記エンドメンバーの決定が、上記学習用検体に付された判別ラベルに基づき行われる、［３］に記載の方法。

［５］　上記エンドメンバーの決定が、上記学習用検体の上記特徴ベクトルの頂点成分分析によって行われる、［３］に記載の方法。

［６］　上記エンドメンバーの決定が、上記Ｋ－１次元単体が最小体積となるように頂点を定めるアルゴリズムによって実施される、［１］に記載の方法。

［７］　上記エンドメンバーの決定が、上記補正後強度分布行列を非負値行列因子分解し、上記検体中のＫ種類の上記成分の重量分率を表す行列と、Ｋ種類の個別の上記成分ごとのフラグメント存在強度を表す行列との積に分解する第２のＮＭＦ処理により行われる、［６］に記載の方法。

［８］　上記学習用検体が、上記エンドメンバーを含まない、［６］又は［７］に記載の方法。

［９］　上記補正後強度分布行列を得ることは、更に、上記強度分布行列の強度補正を行うことを含む、［１］～［８］のいずれかに記載の方法。

［１０］　上記検体セットが、更に、キャリブレーション検体を含み、上記キャリブレーション検体は、上記Ｋ種類の成分を全て含み、かつ、その組成が既知であり、上記強度補正は、上記強度分布行列を規格化することを含む、［９］に記載の方法。

［１１］　上記強度補正は、上記強度分布行列の少なくとも一部を、上記検体セットのうち対応する上記検体の質量と環境変数の積で割り付けることを含み、

上記環境変数は、上記観測の際における上記成分のイオン化効率への影響を表す変数である、［９］に記載の方法。

［１２］　上記環境変数が、上記観測の際に、雰囲気中に所定量含まれる分子量５０～１５００の化合物、又は、上記検体セットのそれぞれに所定量含まれる分子量５０～５００の有機低分子化合物のマススペクトルのピークの合計値である、［１１］に記載の方法。

［１３］　Ｋを１以上の整数としたとき、Ｋ種類の成分から選択される少なくとも１種の上記成分を含む推定対象検体における、上記成分の含有量比を推定する組成推定装置であって、上記Ｋ種類の成分から選択される少なくとも１種の成分を含み、互いに組成の異なるＫ個以上の学習用検体、上記成

分を含まないバックグラウンド検体、及び、上記推定対象検体を含む検体セットに含まれるそれぞれの検体を加熱しながら、熱脱着、及び／又は、熱分解によって生ずるガス成分を順次イオン化し、マススペクトルを連続的に観測する質量分析装置と、上記観測されたマススペクトルを処理する情報処理装置と、を備え、上記情報処理装置は、加熱温度ごとに得られた上記マススペクトルを各行に格納し、上記検体ごとの二次元マススペクトルを得て、上記二次元マススペクトルの少なくとも２つ以上をまとめてデータ行列に変換する、データ行列作成部と、上記データ行列を非負値行列因子分解し、規格化された基底スペクトル行列とその強度分布行列の積に分解するＮＭＦ処理を行う、ＮＭＦ処理部と、上記基底スペクトル行列と上記データ行列との正準相関分析によって強度分布行列に含まれるノイズ成分を抽出し、上記ノイズ成分の影響を減ずるよう上記強度分布行列を補正して、補正後強度分布行列を作成する、補正処理部と、上記補正後強度分布行列を上記検体セットのそれぞれに対応する小行列に分割し、上記小行列を特徴ベクトルとして、ベクトル空間内で上記検体のそれぞれを表現する、ベクトル処理部と、上記特徴ベクトルの全てを内包するＫ－１次元単体を設定し、上記Ｋ－１次元単体におけるＫ個のエンドメンバーを決定する、エンドメンバー決定部と、上記Ｋ個のエンドメンバーと、上記推定対象検体の特徴ベクトルとの、ユークリッド距離をそれぞれ計算し、上記ユークリッド距離の比により、上記推定対象検体中の上記成分の含有量比を推定する、含有量比計算部と、を含み、上記Ｋが３以上の場合、上記Ｋ－１次元単体に内接する超球体の外側の領域のそれぞれに、上記学習用検体の特徴ベクトルの少なくとも１つが位置する、又は、上記学習用検体が上記エンドメンバーの少なくとも１つを含む、組成推定装置。

［１４］　上記エンドメンバー決定部は、上記学習用検体に付された判別ラベルに基づいて上記エンドメンバーを決定する、［１３］に記載の組成推定装置。

［１５］　上記エンドメンバー決定部は、上記Ｋ－１次元単体が最小体積

となるように頂点を定めるアルゴリズムによって上記エンドメンバーを決定する、［１３］に記載の組成推定装置。

［１６］　上記エンドメンバー決定部は、上記補正後強度分布行列を非負値行列因子分解し、上記検体中のＫ種類の上記成分の重量分率を表す行列と、Ｋ種類の個別の上記成分ごとのフラグメント存在強度を表す行列との積に分解する第２のＮＭＦ処理により上記エンドメンバーを決定する、［１５］に記載の組成推定装置。

［１７］　上記補正処理部は更に、上記強度分布行列の強度補正を行う、［１５］に記載の組成推定装置。

［１８］　上記検体セットが、更に、キャリブレーション検体を含み、上記キャリブレーション検体が、上記Ｋ種類の成分を全て含み、かつ、その組成が既知であり、上記強度補正が、上記強度分布行列を規格化することを含む、［１７］に記載の組成推定装置。

［１９］　上記強度補正は、上記強度分布行列の少なくとも一部を、上記検体セットのうち対応する上記検体の質量と環境変数の積で割り付けることを含み、

上記環境変数は、上記観測の際における上記成分のイオン化効率への影響を表す変数である、［１７］に記載の組成推定装置。

［２０］　上記質量分析装置は減圧ポンプを更に備え、上記環境変数が、上記減圧ポンプの稼働によって生ずる物質のマススペクトルのピークの合計値である、［１９］に記載の組成推定装置。

［２１］　Ｋを１以上の整数としたとき、Ｋ種類の成分から選択される少なくとも１種の上記成分を含む推定対象検体における、上記成分の含有量比を推定する組成推定装置において用いられるプログラムであって、上記Ｋ種類の成分から選択される少なくとも１種の成分を含み、互いに組成の異なるＫ個以上の学習用検体、上記成分を含まないバックグラウンド検体、及び、上記推定対象検体を含む検体セットに含まれるそれぞれの検体を加熱しながら、熱脱着、及び／又は、熱分解によって生ずるガス成分を順次イオン化し

、マススペクトルを連続的に観測する質量分析装置と、上記観測されたマススペクトルを処理する情報処理装置と、を備え、上記質量分析装置によって、加熱温度ごとに取得された上記マススペクトルを各行に格納し、上記検体ごとの二次元マススペクトルを得て、上記二次元マススペクトルの少なくとも２つ以上をまとめてデータ行列に変換する、データ行列作成機能と、上記データ行列を非負値行列因子分解し、規格化された基底スペクトル行列とその強度分布行列の積に分解するＮＭＦ処理を行う、ＮＭＦ処理機能と、上記基底スペクトル行列と上記データ行列との正準相関分析によって強度分布行列に含まれるノイズ成分を抽出し、上記ノイズ成分の影響を減ずるよう上記強度分布行列を補正して、補正後強度分布行列を作成する、補正処理機能と、上記補正後強度分布行列を上記検体のそれぞれに対応する小行列に分割し、上記小行列を特徴ベクトルとして、ベクトル空間内で上記検体のそれぞれを表現する、ベクトル処理機能と、上記特徴ベクトルの全てを内包するＫ－１次元単体を設定し、上記Ｋ－１次元単体におけるＫ個のエンドメンバーを決定する、エンドメンバー決定機能と、上記Ｋ個のエンドメンバーと、上記推定対象検体の特徴ベクトルとの、ユークリッド距離をそれぞれ計算し、上記ユークリッド距離の比により、上記推定対象検体中の上記成分の含有量比を推定する、含有量比計算機能と、を含み、上記Ｋが３以上の場合、上記Ｋ－１次元単体に内接する超球体の外側の領域のそれぞれに、上記学習用検体の特徴ベクトルの少なくとも１つが位置する、又は、上記学習用検体が上記エンドメンバーの少なくとも１つを含む、プログラム。

発明の効果

[0011] 本発明の方法（以下「本方法」ともいう。）は、検体セットに含まれるそれぞれの検体を加熱しながら、熱脱着、及び／又は、熱分解によって生ずるガス成分を順次イオン化し、マススペクトルを連続的に観測し、得られた加熱温度ごとのマススペクトルを各行に格納した検体ごとの二次元マススペクトルを得て、これらをまとめたデータ行列を非負値行列因子分解（ＮＭＦ）によって、規格化された基底スペクトル行列とその強度分布行列の積に分解

し、正準相関分析によって強度分布行列に含まれるノイズ成分を抽出して補正し、各検体の特徴ベクトルを含む補正後強度分布行列を得て、その後の定量計算を行う点に特徴点の一つがある。

[0012] この正準相関分析によって強度分布行列を補正する処理を行うことによって、（詳細は後述するが）より容易に、かつ、高精度で定量分析が可能になった。

[0013] また、Ｋが２以上の整数であり、推定対象となる検体（推定対象検体）がＫ種類の成分の混合物である場合、本方法によって、推定対象検体中におけるＫ種類の成分の混合比（組成）を推定することができる。

[0014] また、学習用検体がエンドメンバーを含む場合、より正確、かつ、より容易に検体の組成を推定することができる。

[0015] また、学習用検体にエンドメンバーの判別するための判別ラベルが付されている場合、そのラベルの情報からエンドメンバーを決定できるため、より容易に検体の組成を推定することができる。

[0016] また、判別ラベルが付されていない場合であって、学習用検体がエンドメンバーを含んでいる場合には、学習用検体の特徴ベクトルの頂点成分分析によってエンドメンバーを決定でき、検体の組成が推定できる。

[0017] また、学習用検体がエンドメンバーを含んでいない場合であっても、Ｋ－１次元単体が最小体積となるように頂点を定めるアルゴリズムを用いることによってエンドメンバーを決定することで、検体の組成が推定できる。学習用検体にエンドメンバーを必須としない点で優れている。

[0018] また、補正後強度分布行列が、強度分布行列の強度補正後のものであると、推定の正確性がより高まる。

[0019] また、検体セットが更に、キャリブレーション検体を含む場合であって、強度補正が強度分布行列を規格化することを含む場合、より簡便な手順で、より高精度に検体の組成が推定できる。

[0020] また、強度分布行列における各検体に対応する部分を検体の質量と環境変数の積で割り付けることによれば、観測の際における各成分のイオン化効率

による影響が補正され、検体セットにキャリブレーション検体が含まれない場合であっても、検体の組成が高精度に推定できる。

[0021] また、上記割り付けに使用される環境変数が観測の際の雰囲気中に所定量含まれる分子量５０～１５００の化合物、又は、検体セットのそれぞれの検体に所定量含まれる分子量５０～５００の有機低分子化合物のマススペクトルのピークの合計値である場合、検体セットにキャリブレーション検体が含まれない場合であっても、より簡便な操作で、より正確な組成推定結果が得られる。

図面の簡単な説明

[0022] [図1]本発明の実施形態に係る組成推定方法のフローチャートである。

[図2]Ｋが３のときのＫ－１次元単体（２次元単体、三角形）を表すイメージ図である。

[図3]本発明の実施形態に係る組成推定装置のハードウェア構成図である。

[図4]本発明の実施形態に係る組成推定装置の機能ブロック図である。

[図5]「ＭＳＥ＿ｃａｌｉｂ」についてのオリジナルなスペクトル（Ａ）と、処理を行った後、再構築したスペクトル（Ｂ）を表す図である。

[図6]ＮＭＦ分解後のＣ行列の一部$C_{\delta(MSE_Calib)}$と、ＣＣＡ－Ｆｉｌｔｅｒによってバックグラウンド及び不純物のスペクトルを除いた後のS^{T}である。

[図7]Ｍ（ポリメチルメタクリレート）、Ｓ（ポリスチレン）、Ｅ（ポリエチルメタクリレート）三成分系の組成解析結果である。

[図8]ＮＭＦの条件から、直交制約を解除した場合の組成推定結果である。

[図9]ＭＮＦ最適化の条件から、ｍｅｒｇｉｎｇ　Ｓのみ解除した場合の組成推定結果である。

[図10]ＣＣＡ－ｆｉｌｔｅｒを解除した場合の結果（比較例）の組成推定結果である。

[図11]「ＭＳＥ＿ｃａｌｉｂ」についてのオリジナルなスペクトル（上、ｒａｗ　ｏｆ　ＭＳＥ＿ｃａｌｉｂ）と、ＣＣＡ－ｆｉｌｔｅｒを解除した上

記で処理を行った後、再構築したスペクトルを表す図である。

[図12]環境変数による補正を行わなかった場合の組成推定結果である。

[図13]Ｋが３のときのＫ－１次元単体（２次元単体、三角形）を表すイメージ図である。

[図14]本発明の第２実施形態におけるエンドメンバーの決定方法の仕組みの説明図である。

[図15]実施例２の推定結果を表す図である。

[図16]実施例３の推定結果を表す図である。

発明を実施するための形態

[0023] 以下、本発明について詳細に説明する。

以下に記載する構成要件の説明は、本発明の代表的な実施形態に基づいてなされることがあるが、本発明はそのような実施形態に制限されるものではない。

なお、本明細書において、「～」を用いて表される数値範囲は、「～」の前後に記載される数値を下限値及び上限値として含む範囲を意味する。

[0024] ［用語の定義］

本明細書において使用される用語について説明する。なお、以下に説明のない用語については、当業者の間で普通に理解される意味で使用される。

[0025] 本明細書において、「成分」とは、各検体を構成する物質を意味し、推定対象検体、学習用検体、及び、キャリブレーション検体は、成分の組合せによって構成される混合物を意味する。

[0026] 「推定対象検体」とは、組成を推定すべき検体で、含まれている成分の数、及び、各成分の含有量はいずれも未知であってよい。

[0027] 「学習用検体」とは、後述するＫ－１次元単体のＫ個のエンドメンバー決定に必要な検体を意味し、これらの検体には、Ｋ種類の成分から選択される少なくとも１種の成分が含まれ、それぞれ組成が異なっている。含まれる成分数は特に制限されず１～Ｋ種類であってよい。なお、学習用検体には、推定対象検体と同一の組成のものが含まれていてもよい。すなわち、推定対象

検体と学習用検体とは同一であってもよいが、学習用検体はそれぞれ異なる。

[0028] 「エンドメンバー」とは、Ｋ－１次元単体の頂点に対応するベクトルを意味し、このエンドメンバーは、Ｋ種類のうちの１成分のみからなる検体（学習用検体、又は、仮想の検体）の特徴ベクトルに対応する。

[0029] 「バックグラウンド検体」とは、推定対象検体、及び、学習用検体とは異なる検体であって、上記成分のいずれをも含まない検体を意味する。

[0030] 「キャリブレーション検体」とは、推定対象検体、学習用検体、及び、バックグラウンド検体とは異なる検体であって、Ｋ種類の成分を全て含み、その組成が明らかな検体を意味する。

[0031] ［本方法の第１実施形態］

図１は、本発明の実施形態に係る組成推定方法（以下、「本方法」ともいう。）のフローチャートである。

[0032] まず、推定対象検体、互いに組成の異なるＫ個以上の学習用検体、及び、バックグラウンド検体（以上をまとめて「検体セット」という。）が準備される（ステップＳ１０１）。なお、検体セットは、キャリブレーション検体を更に含んでもよい。

[0033] 推定対象検体は、Ｋ種類の成分のうち少なくとも１種の成分を含む。推定対象検体が１種類の成分を含む場合、本方法によってその成分の純度を推定することができる。この場合、推定対象検体には、上記「成分」と、それ以外の不純物とが含まれる。

[0034] Ｋが２以上の場合、推定対象検体は、成分の混合物である。本方法におけるＫは１以上であれば、上限は特に制限されない。典型的には、Ｋは、１０００以下が好ましい。

なお、学習用検体の個数は成分数Ｋに関係して、Ｋ個以上である必要がある。Ｋ個以上であれば、学習用検体の数としては特に制限されないが、より多い場合、後述するＫ－１次元単体の頂点の決定の精度がより高くなりやすく、結果として、より正確な組成推定結果が得られやすい。

すなわち、成分数Ｋに対する学習用検体の数の比（学習用検体の数／成分数）は、１以上であり、２以上が好ましく、３以上がより好ましく、５以上が更に好ましく、１０以上が特に好ましい。一方、上限値は特に制限されないが、一般に、１０００以下が好ましい。

[0035] 学習用検体は、互いに組成が異なる。組成は、含まれる成分の種類と数、及び、各成分の含有量によって決まる。組成が同一とは、含まれる成分の種類、数、及び、それぞれの含有量が同一であることを意味し、いずれか１つが異なれば、それは「互いに組成が異なる」状態となる。

[0036] なお、本方法において、学習用検体は、互いに組成が異なっていれば、それ以外の情報、すなわち、含まれる成分の種類、成分数、及び、各含有量は、推定対象検体の組成推定には必要ない。すなわち、学習用検体の組成は、未知であってよい。

推定対象検体と学習用検体との相違点は、学習用検体はそれぞれ組成が異なることが必要である点である。なお、学習用検体は、推定対象検体と組成が同一のものを含んでいてもよい。

[0037] 学習用検体は、エンドメンバーを含んでいてもよい。すなわち、学習用検体は、Ｋ種類の成分から選択される１種類の成分からなるもの（単味のもの、ピュア）を含んでもよい。学習用検体がエンドメンバーを含む場合、より容易に、高精度な組成推定が可能である。

[0038] このとき、エンドメンバーである学習用検体は、判別ラベルを含んでもよい。学習用検体が判別ラベルを含む場合、エンドメンバーの決定がより容易になる点で好ましい。

[0039] 次に、推定対象検体をそれぞれ加熱して、熱脱着、及び／又は、熱分解によって生ずるガスを順次イオン化し、マススペクトルを取得する（ステップＳ１０２）。

[0040] マススペクトルを取得する方法としては特に制限されないが、アンビエントな条件にある試料を前処理なしで質量分析する方法が好ましい。このようなイオン化、及び、質量分析方法としては、ＤＡＲＴ（Ｄｉｒｅｃｔ　Ａｎ

ａｌｙｓｉｓ　ｉｎ　Ｒｅａｌ　Ｔｉｍｅ）イオン源と呼ばれるイオン源と質量分析計とを組合わせたＤＡＲＴ－ＭＳと呼ばれる質量分析装置が知られている。

質量分析計としては特に制限されず、精密質量分析が可能なものが好ましく、四重極型、及び、飛行時間（ＴＯＦ）型等のいずれでもよい。

[0041] マススペクトルの取得について、具体的な条件は特に制限されないが、非限定的な一例を挙げると、検体セットに含まれるそれぞれの検体を順次５０℃／ｍｉｎの昇温速度で加熱し、５０～５５０℃の温度範囲で生じる熱分解ガスに対して５０ｓｈｏｔ／ｍｉｎの間隔でヘリウムイオンを噴射しガスをイオン化させ、横軸にｍ／ｚ、縦軸に温度を有する二次元ＭＳスペクトルを得るという手順が挙げられる。

[0042] 次に、加熱温度ごとに得られたマススペクトルを各行に格納し、検体ごとの二次元マススペクトルを得て、この二次元マススペクトルの少なくとも２つ以上をまとめてデータ行列に変換する（ステップＳ１０３）。

[0043] データ行列の取得方法を具体的に説明する。

[0044] ステップＳ１０２では、所定の昇温間隔ごとに連続的にマススペクトルが取得される。これらのマススペクトルは、そのままデータ行列の作成に使用されてもよいが、所定の昇温温度範囲ごとに平均して使用されてもよい。所定の昇温温度範囲ごとに平均してマススペクトルを１本にまとめることで、データ量を圧縮することができる。このような昇温温度範囲としては、例えば、１０～３０℃程度が挙げられる。

[0045] また、各スペクトルにおいて、ピーク強度は規格化されていてもよい。規格化の方法としては、例えば、ピーク強度の二乗和が１になるように規格化する方法が挙げられる。

[0046] このようにすることで、ある検体について、一回の測定することで、加熱温度ごと（又は、所定の範囲ごとにまとめられた加熱温度ごと）に、所定の本数（まとめ方によって異なる、例えば、２０本等）のマススペクトルが得られる。

このマススペクトルが各行に格納され、加熱温度が各列に格納されると、検体ごとの二次元マススペクトルが得られる。

[0047] こうして、検体ごとの二次元マススペクトルが得られたら、これらの少なくとも２つ以上をまとめてデータ行列Ｘに変換する。

データ行列Ｘの作成に用いられる二次元マススペクトルの個数は２つ以上であれば特に制限されないが、検体セットに含まれる全検体の二次元マススペクトルが用いられることが好ましい。

なお、１つの検体について２回以上、測定が行われる場合には、２回以上の測定で得られた二次元マススペクトルの一部、又は、全てがデータ行列Ｘの作成に用いられてもよい。

[0048] 次に、得られたデータ行列を非負値行列因子分解（ＮＭＦ）し、規格化された基底スペクトル行列と、その強度分布行列との積に分解する（ステップＳ１０４）。

[0049] 非負値行列因子分解について説明する。

[0050] スペクトルの強度とサンプルの成分組成の間に線形加成性が成り立つという仮定の下で、スペクトルのデータ行列は一般に、

[0051] [数1]

$$X = CS^T$$

[0052] と行列分解できる。ここでデータ行列Ｘはチャンネル数Ｍのスペクトルを Ｎ本積み重ねてできるＮ×Ｍのデータ行列、Ｓは基底スペクトル行列、Ｃはその強度分布行列である。データセットに含まれるすべての成分数をK_1とすればＣは$N \times K_1$行列、Ｓは$M \times K_1$行列になる。実際の計測にはノイズが含まれるので、

[0053]

[数2]

$$X = CS^T + N$$

[0054] と誤差行列Ｎを考慮にいれ、Ｎを最小化するようにＣ，Ｓ行列を定める。行列右肩のＴは転置行列を意味する。ここでガウスノイズを仮定すれば

[0055] [数3]

$$\min_{C,\ S} \left\| X - CS^T \right\|_F^2$$

[0056] なるＣ、Ｓについての最適化問題として定式される。なお、

[0057] [数4]

$$\|matix\|_F$$

[0058] は、行列のフロベニウスノルム、

[0059] [数5]

$$\|vec\|_1$$

[0060] は、ベクトルのＬ１ノルム、

[0061] [数6]

$$\|vec\|$$

[0062] は、ベクトルのＬ２ノルムを表すものとする。

[0063] なお、上式におけるＣ、Ｓは唯一解が定まらないので、物理的要請からスペクトルが満たすであろう拘束条件を課す。ＮＭＦは、組成比やスペクトル強度が非負の値をとるという自明な要請から、

[0064] [数7]

$$\min_{C\geq 0,\ S\geq 0} \left\| X - CS^{T} \right\|_{F}^{2}$$

[0065] として立式される。しかし、非負条件だけでは唯一解が定まらないので、測定装置に応じたさらなる拘束条件を適用することが求められる。以下では、本ステップにおけるＮＭＦの拘束条件について説明する。

[0066] （ＮＭＦ更新式について）

本ステップにおけるＮＭＦ処理は、以下に述べる制約を組み込みやすいＨｉｅｒａｒｃｈｉｃａｌ ａｌｔｅｒｎａｔｉｎｇ ｌｅａｓｔ ｓｑｕａｒｅ （ＨＡＬＳ）により、Ｃ、Ｓの列ベクトルワイズに交互に更新していくことで、上記式（数７）をＣ、Ｓの各列についての凸最適化問題として解くことが好ましい。

[0067] 本方法では、各検体を加熱しながら、熱脱着、及び／又は、熱分解させて生じるガスについて質量分析を行うため、熱分解後の成分数K_1を事前情報として得ることは難しく、データから推定する必要がある。推定方法としては、Ａｕｔｏｍａｔｉｃ ｒｅｌｅｖａｎｃｅ ｄｅｔｅｒｍｉｎａｔｉｏｎ （ＡＲＤ）が好ましい。これは強度分布行列Ｃにスパース性を仮定するもので、ある成分の熱分解はその成分を含む検体の特定温度でのみ起こることを考えれば、自然な仮定といえる。これによりなるべく少ない成分数にする力と上記式（数７）を最小化する力が釣り合い、最適な成分数を求めることができる。

[0068] （直交制約）

質量分析計は、一般に０．１～０．００１ｍ／ｚ程度の分解能を有する。

したがって、異なるガス成分のｍ／ｚが小数点以下まで一致し、同一のチャネルを占めるという状況は稀にしか生じない。したがって、基底スペクトル行列Ｓの列ベクトル間にはある程度の直交性を仮定するのが妥当である。

[0069] これには良い局所解に導く効果のみならず、強度の強いピークにのみフィッティングが集中することを防ぐ効果も期待できる。一方で、完全な直交かつ非負は強すぎる制約となってしまうので、ラグランジュ未定乗数を有するペナルティー項を加える形でソフトに直交制約を課すことが好ましい（ソフト直交制約）。

[0070]（成分のマージング）

ＣとＳの更新途中で、ほとんど同じ基底スペクトルが複数得られた場合は一つのスペクトルにまとめることで正しい成分数ｎを得ることができる。ＥＡＬＳのスペクトライメージをＮＭＦに適用すると、Ｓのｋ番目の列ベクトルを$S_{:k}$と書いて、$S^{T}_{:k}S_{:m}$＞ｔｈｒｅｓｈｏｌｄなるｋ＜ｍが存在するときに、

[0071] [数8]

$$C_{:k} \leftarrow C_{:k} + C_{:m}$$

$$C_{:m} \leftarrow \mathbf{0}$$

[0072] によって一本化できる（Ｍｅｒｇｉｎｇ Ｃ）。ここでＴｈｒｅｓｈｏｌｄは０．９～０．９９で与えられるハイパーパラメータである。これと同様に類似Ｓのマージングを行い、更に、Ｃの複数の列ベクトルが平行であったときも、対応するＳの列ベクトルをＣの重みに応じて重ね合わせることが好ましい。すなわち、

[0073] [数9]

$$C_{:k}^{T}C_{:m} > threshold\|C_{:k}\|\|C_{:m}\|$$

[0074] なるｋ≠ｍが存在するときに、

[0075] [数10]

$$\boldsymbol{S}_{:k} \leftarrow \boldsymbol{S}_{:k} + \frac{\|\boldsymbol{C}_{:m}\|_1}{\|\boldsymbol{C}_{:k}\|_1}\boldsymbol{S}_{:m}$$

$$s = \|\boldsymbol{S}_{:k}\|$$

$$\boldsymbol{S}_{:k} \leftarrow \frac{\boldsymbol{S}_k}{\|\boldsymbol{S}_{:k}\|}$$

$$\boldsymbol{C}_{:k} \leftarrow s\boldsymbol{C}_{:k}$$

$$\boldsymbol{C}_{:m} \leftarrow \boldsymbol{0}$$

[0076] と一本化する（Ｍｅｒｇｉｎｇ　Ｓ）。これには同位体ピークやオリゴマーピークを一つの基底スペクトルにまとめる効果があると考えられる。

[0077] （更新式の立式）

上記によって得られる更新式について説明する。

モデルの全同時分布は

[0078] [数11]

$$p(\boldsymbol{X}, \boldsymbol{C}, \boldsymbol{S}) = p(\boldsymbol{X}|\boldsymbol{C}, \boldsymbol{S}, \sigma^2)p(\boldsymbol{C}|\boldsymbol{\lambda})p(\boldsymbol{S})p(\boldsymbol{\lambda}|a, b)$$

[0079] ここではσ^2は分散パラメータ、

[0080]

[数12]

$$\boldsymbol{\lambda} \in \mathbb{R}^{\boldsymbol{K}}$$

[0081] の要素λ_kは成分ｋがデータ中に占める重要度を表す逆ガンマ分布に従うパラメータである。ノイズにガウス分布を仮定したので

[0082] [数13]

$$p(X|C,S) = \prod_{i=1}^{N}\prod_{j=1}^{M}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{(X_{ij}-[CS^T]_{ij})^2}{2\sigma^2}\}$$

[0083] Ｃが従う指数分布は、Ｋを初期値として入力する十分大きな任意の成分数（５０程度）として、

[0084] [数14]

$$p(\boldsymbol{C}|\boldsymbol{\lambda}) = \prod_{i=1}^{N}\prod_{k=1}^{K} p(C_{ik}|\lambda_k)$$

$$p(C_{ik}|\lambda_k) = \frac{1}{\lambda_k}\exp\left(-\frac{C_{ik}}{\lambda_k}\right)$$

[0085] λ_kが従う逆ガンマ分布は、

[0086] [数15]

$$p(\boldsymbol{\lambda}|a,b) = \prod_{k=1}^{K} p(\lambda_k|a,b) = \frac{b^a}{\Gamma(a)}\lambda_k^{-(a+1)}\exp\left(-\frac{b}{\lambda_k}\right), k = 1,\ldots,K$$

[0087] ここでａはａ－１が極めて小さな正の数となるようなハイパーパラメータ、ｂはデータ行列の平均値μ_Xにより

[0088] [数16]

$$b = \frac{\mu_X(a-1)\sqrt{N}}{K}$$

[0089] として定まるパラメータである。また、Ｃの要素の期待値は

[0090] [数17]

$$E[C_{ik}] = \frac{\mu_X\sqrt{N}}{K}$$

[0091] ｐ（Ｘ，Ｃ，Ｓ）の負の対数尤度は

[0092] [数18]

$$\begin{aligned} L(\boldsymbol{X},\boldsymbol{C},\boldsymbol{S},\boldsymbol{\lambda}) &= -\log[p(\boldsymbol{X}|\boldsymbol{C},\boldsymbol{S})p(\boldsymbol{C}|\boldsymbol{\lambda})p(\boldsymbol{\lambda}|a,b)] \\ &= \frac{1}{2\sigma^2}\|\boldsymbol{X}-\boldsymbol{C}\boldsymbol{S}^T\|_F^2 + \frac{MN}{2}log2\pi\sigma^2 + (N+a+1)\sum_{k=1}^{K} log\boldsymbol{\lambda}_k \\ &+ \sum_{k=1}^{K}\frac{1}{\lambda_k}(b+\sum_{i=1}^{N}C_{ik}) + K(log\Gamma(a) - alogb) \end{aligned}$$

[0093] ここで、ガウス分布を仮定したときの分散パラメータ σ^2は、

[0094] [数19]

$$\sigma^2 = \frac{1}{MN}\|\boldsymbol{X}-\boldsymbol{C}\boldsymbol{S}^T\|_F^2$$

[0095] で与えられ、また、λについて凸関数なので

[0096] [数20]

$$\lambda_k = \frac{b + \sum_{i=1}^{N} C_{ik}}{N + a + 1}$$

[0097] を得ることができる。

次に、Sの更新式について説明する。まず、負対数尤度関数から$S_{:k}$に関連する項J（$S_{:k}$）のみを取り出すことを考える。Xに対するk成分の寄与、

[0098] [数21]

$$\boldsymbol{X}^{(k)} = \boldsymbol{X} - \boldsymbol{C}\boldsymbol{S}^T + \boldsymbol{C}_{:k}\boldsymbol{S}_{:k}^T$$

[0099] を用いて、

[0100] [数22]

$$\|\boldsymbol{X} - \boldsymbol{C}\boldsymbol{S}^T\|_F^2 = \left\|\boldsymbol{X}^{(k)} - \boldsymbol{C}_{:k}\boldsymbol{S}_{:k}^T\right\|_F^2 = Tr[(\boldsymbol{X}^{(k)} - \boldsymbol{C}_{:k}\boldsymbol{S}_{:k}^T)^T(\boldsymbol{X}^{(k)} - \boldsymbol{C}_{:k}\boldsymbol{S}_{:k}^T)]$$

[0101] と書けるので、定数項を除いて、

[0102] [数23]

$$\boldsymbol{J}(\boldsymbol{S}_{:k}) = \frac{1}{2\sigma^2}[-2\boldsymbol{S}_{:k}^T\boldsymbol{X}^{(k)}\boldsymbol{C}_{:k} + \|\boldsymbol{C}_{:k}\|^2\|\boldsymbol{S}_{:k}\|^2]$$

[0103] 上記は、Ｓｈｉｇａ，Ｍ．ｅｔ ａｌ．Ｓｐａｒｓｅ ｍｏｄｅｌｉｎｇ ｏｆ ＥＥＬＳ ａｎｄ ＥＤＸ ｓｐｅｃｔｒａｌ ｉｍａｇｉｎｇ ｄａｔａ ｂｙ ｎｏｎｎｅｇａｔｉｖｅ ｍａｔｒｉｘ ｆａｃｔｏｒｉｚａｔｉｏｎ．Ｕｌｔｒａｍｉｃｒｏｓｃｏｐｙ １７０，４３－５９（２０１６）．をベースに導出された。ここで、上述したSに加える直交制約

によるペナルティー項を加える。

[0104] [数24]

$$J(S_{:k}) = \frac{1}{2\sigma^2}\left[-2S_{:k}^T X^{(k)\ T} C_{:k} + \|C_{:k}\|^2 \|S_{:k}\|^2\right] + \xi_k s^{(k)\ T} S_{:k}$$

[0105] ここで、ξ_kはラグランジュ未定乗数、

[数25]

$$s^{(k)} = \sum_{j \neq k}^{K} S_{:j}$$

[0106] である。$s_{:k}$で微分して

[0107] [数26]

$$\frac{\partial J(S_{:k})}{\partial S_{:k}} = \frac{1}{\sigma^2}\left[X^{(k)\ T} C_{:k} + \|C_{:k}\|^2 S_{:k}\right] + \xi s^{(k)}$$

[0108] [数27]

$$\frac{\partial J(S_{:k})}{\partial S_{:k}} = 0$$

[0109] とおいて、$s_{:k}$についての更新式

[0110] [数28]

$$S_{:k} \leftarrow \frac{-\sigma^2 \xi_k s^{(k)} + X^{(k)\ T} C_{:k}}{\|C_{:k}\|^2}$$

[0111] を得る。完全直交条件を与えるξ_kは、

[0112]

[数29]

$$\boldsymbol{s}^{(k)\ T}\boldsymbol{S}_{:k} = \boldsymbol{0}$$

[0113] より、

[0114] [数30]

$$\xi_k = \frac{\boldsymbol{s}^{(k)\ T}\boldsymbol{X}^{(k)\ T}\boldsymbol{C}_{:k}}{\sigma^2\boldsymbol{s}^{(k)\ T}\boldsymbol{s}^{(k)}}$$

[0115] より見積もることができる。ソフト直交条件はξ_kに［０，１］のハイパーパラメータｗをかけて、完全直交と直交制約なしとの間で自由に直交度合いを調整することができる。最終的な更新式は、

[0116] [数31]

$$\boldsymbol{S}_{:k} \leftarrow \frac{-\sigma^2 w\xi_k\boldsymbol{s}^{(k)} + \boldsymbol{X}^{(k)\ T}\boldsymbol{C}_{:k}}{\|\boldsymbol{C}_{:k}\|^2}$$

$$\boldsymbol{S}_{:k} \leftarrow \frac{\boldsymbol{S}_{:k}}{\|\boldsymbol{S}_{:k}\|}$$

[0117] 次に、Ｃの更新式について説明する。負対数尤度関数から$C_{:k}$に関連する項Ｊ（$C_{:k}$）のみを取り出すことを考える。

[0118]

[数32]

$$J(\boldsymbol{C}_{:k}) = \frac{1}{2\sigma^2}\left[-2\boldsymbol{S}_{:k}^{T}\boldsymbol{X}^{(k)}\boldsymbol{C}_{:k} + \|\boldsymbol{C}_{:k}\|^2\|\boldsymbol{S}_{:k}\|^2\right] + \frac{1}{\lambda_k}\sum_{i=1}^{N} C_{ik}$$

$$\frac{\partial J(\boldsymbol{C}_{:k})}{\partial \boldsymbol{C}_{:k}} = \frac{1}{\sigma^2}\left[\boldsymbol{X}^{(k)}\boldsymbol{S}_{:k} + \|\boldsymbol{S}_{:k}\|^2\boldsymbol{C}_{:k}\right] + \frac{1}{\lambda_k}$$

[0119] [数33]

$$\|\boldsymbol{S}_{:k}\|^2 = 1$$

[0120] と規格化しておけば、更新式は

[0121] [数34]

$$\frac{\partial J(\boldsymbol{C}_{:k})}{\partial \boldsymbol{C}_{:k}} = \boldsymbol{0}$$

[0122] より、

[0123] [数35]

$$\boldsymbol{C}_{:k} \leftarrow \boldsymbol{X}^{(k)}\boldsymbol{S}_{:k} - \frac{\sigma^2}{\lambda_k}$$

[0124] 具体的なアルゴリズムは以下のとおりである。

[0125]

[数36]

Algorithm 1: Pseudo-code of ARD-SO-NMF for TDMS

Input: Data $\boldsymbol{X} \in \mathbb{R}^{\boldsymbol{N}\times\boldsymbol{M}}$, orthogonal constrain weight w, initial component number K, small positive number a, iteration number L, $threshold \in [0.9, 0.99]$

Output: optimized component number K^*, Cardinality matrix $\boldsymbol{C} \in \mathbb{R}^{\boldsymbol{N}\times\boldsymbol{K}}$, base spectra $\boldsymbol{S} \in \mathbb{R}^{\boldsymbol{M}\times\boldsymbol{K}}$, components importance $\boldsymbol{\lambda} \in \mathbb{R}^{\boldsymbol{K}}$

//Initialization

set $\boldsymbol{C}, b$ by Eq. (1-2)

set $\boldsymbol{S}$ by randomly selecting raw vector from $\boldsymbol{X}$

//Main loop

for l from 1 to L:

  for k from 1 to K:

    update $\boldsymbol{S}_{:\boldsymbol{k}}$ by Eq. (6)

    $\boldsymbol{S}_{:\boldsymbol{k}} \leftarrow (\boldsymbol{S}_{:\boldsymbol{k}} + |\boldsymbol{S}_{:\boldsymbol{k}}|)/2$

  end for

  for k from 1 to K:

    update $\boldsymbol{C}_{:\boldsymbol{k}}$ by Eq. (7)

    $\boldsymbol{C}_{:\boldsymbol{k}} \leftarrow (\boldsymbol{C}_{:\boldsymbol{k}} + |\boldsymbol{C}_{:\boldsymbol{k}}|)/2$

  end for

  for k from 1 to $K-1$:

    for m from $k+1$ to K:

      //Merging C

      if $\boldsymbol{S}_{:k}^{T}\boldsymbol{S}_{:m} > t$:

        $\boldsymbol{C}_{:k} \leftarrow \boldsymbol{C}_{:k} + \boldsymbol{C}_{:m}$

        $\boldsymbol{C}_{:m} \leftarrow \boldsymbol{0}$

      end if

      //Merging S

      if $\boldsymbol{C}_{:k}^{T}\boldsymbol{C}_{:m} > threshold\|\boldsymbol{C}_{:k}\|\|\boldsymbol{C}_{:m}\|$:

        $\boldsymbol{S}_{:k} \leftarrow \boldsymbol{S}_{:k} + \frac{\|\boldsymbol{C}_{:m}\|_1}{\|\boldsymbol{C}_{:k}\|_1}\boldsymbol{S}_{:m}, \quad s = \|\boldsymbol{S}_{:k}\|, \quad \boldsymbol{S}_{:k} \leftarrow \frac{\boldsymbol{S}_k}{\|\boldsymbol{S}_{:k}\|}$

        $\boldsymbol{C}_{:k} \leftarrow s\boldsymbol{C}_{:k}, \boldsymbol{C}_{:m} \leftarrow \boldsymbol{0}$

      end if

    end for

  end for

  update λ by Eq. (4)

  update σ^2 by Eq. (3)

end

[0126] 図１に戻り、次に、基底スペクトル行列とデータ行列とを正準相関分析し

て補正後強度分布行列を得る（ステップＳ１０５）。

具体的には、基底スペクトル行列とデータ行列との正準相関分析によって強度分布行列に含まれるノイズ成分を抽出し、ノイズ成分の影響を減ずるよう強度分布行列を補正して、補正後強度分布行列を得る。

[0127] ＮＭＦはデータ行列の低ランク近似であるため、実際にはｉ番目のスペクトルには存在しない成分ｋであっても、存在するとしたほうが最小二乗の意味で近似が良くなるのであれば$C_{ik}>0$とする。多くの場合このようなC_{ik}は非常に小さく、ＮＭＦ解析において問題にならない場合が多い。

[0128] しかし、微量成分検出においては、ｊ番目のスペクトルに微量実在する$C_{jk}>0$と、ＮＭＦアーティファクトである$C_{ik}>0$を区別し、ゴーストピークであるC_{ik}には０を代入することが好ましい。ＮＭＦアルゴリズムのアーティファクトに由来する偽ピークをノイズの１つとして除くことによりより精度の高い推定結果が得られるからである。

[0129] 本ステップでは、上記を解決するために、正準相関分析（Ｃａｎｏｎｉｃａｌ Ｃｏｒｒｅｌａｔｉｏｎ Ａｎａｌｙｓｉｓ ）を用いる。本ステップは本方法の特徴点の一つであり、本発明者らは本ステップを、Ｃａｎｏｎｉｃａｌ Ｃｏｒｒｅｌａｔｉｏｎ Ａｎａｌｙｓｉｓ （ＣＣＡ） Ｆｉｌｔｅｒと命名した。

[0130] ＣＣＡ ｆｉｌｔｅｒは、概念的には、ＮＭＦから出力される基底スペクトルS^Tの各成分が実際に元データに含まれていたかをサンプルワイズにスキャンし、元データに類似のピークパターンが見られなければそのサンプルのスペクトルから消去することによる。Ｌ番目のサンプルから生成されるスペクトルがデータ行列Ｘのｌ行からｌ′行までに格納されていた場合、その部分を抜き出した小行列を$X_{\delta(L)}$と書く。すなわち$\delta(L)=\{l, l+1, \cdots, l'-2, l'-1\}$ある。

[0131] ＣＣＡは従来、二系統のデータ間の相関を調べる手法として知られており、以下で定式される。

N個のｐ次元ベクトルからなる行列を

[0132] [数37]

$$\boldsymbol{Y} \in \mathbb{R}^{p \times N}$$

[0133] Ｎ個のｑ次元ベクトルデータからなる行列を

[0134] [数38]

$$\boldsymbol{Z} \in \mathbb{R}^{q \times n}$$

[0135] として、

[0136] [数39]

$$\rho = \frac{\boldsymbol{a}^T \boldsymbol{V}_{yz} \boldsymbol{b}}{\sqrt{\boldsymbol{a}^T \boldsymbol{V}_{yy} \boldsymbol{a}} \sqrt{\boldsymbol{b}^T \boldsymbol{V}_{zz} \boldsymbol{b}}}$$

$$\max_{\boldsymbol{a},\boldsymbol{b}} \rho$$

[0137] ここで、

[0138] [数40]

$$\boldsymbol{V}_{YY} = \boldsymbol{Y}\boldsymbol{Y}^T/N, \boldsymbol{V}_{ZZ} = \boldsymbol{Z}\boldsymbol{Z}^T/N, \boldsymbol{V}_{YZ} = \boldsymbol{Y}\boldsymbol{Z}^T/N$$

[0139] である。

[0140] すなわち、ある軸にＹとＺを投影して、その相関係数ρが最大となるような解（ａ^T，ｂ^T）^{T}を見つけることがＣＣＡの問題設定である。

[0141] （ａ^T，ｂ^T）^{T}と相関係数ρは次式の一般固有値問題の固有ベクトルと固有値として与えられ、すべての固有ベクトル（ａ^T，ｂ^T）^{T}をまとめて（Ａ^T，Ｂ^T）^{T}と書くと、

[数41]

$$\begin{pmatrix} \boldsymbol{O} & \boldsymbol{V}_{yz} \\ \boldsymbol{V}_{yz}^T & \boldsymbol{O} \end{pmatrix} \begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{pmatrix} = \begin{pmatrix} \boldsymbol{V}_{yy} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{V}_{zz} \end{pmatrix} \begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{pmatrix} \begin{pmatrix} \rho_1 & & \\ & \ddots & \\ & & \rho_{p+q} \end{pmatrix}, \rho_1 \geq \cdots \geq \rho_{p+q}$$

[0142] いま、ある成分ｋの基底スペクトルＳ$_{:k}$が、サンプルＬに含まれているか

、いないかを判定したいとする。まず、データ行列Ｘから該当の部分データ行列$X_{\delta(L)}$を抜き出す。一方S^Tは、$S_{:k}{}^T$と似たすべてのスペクトル群を$S_{sim}{}^T$、それとは大きく異なるスペクトル群を$S_{out}{}^T$と、二つの行列に分け、

[0143] [数42]

$$Y = \begin{pmatrix} S_{out}{}^T \\ X_{\delta(L)} \end{pmatrix}$$

$$Z = \begin{pmatrix} S_{:k}{}^T \\ S_{sim}{}^T \end{pmatrix}$$

[0144] としてＣＣＡを行う。ここである値$t \in (0, 1]$に対して、$\rho > t_1$なるすべての固有値を取り出し、それに対応する固有ベクトルｂの第一成分がその最大成分に対して$b_1 > t_2 b_{max}$を満たせば、成分ｋの基底スペクトル$S_{:k}$がサンプルＬに含まれていると判定する。そうでなければ、対応する$C_{\delta(L)k} = 0$とする。

ここで、t_1、t_2はハイパーパラメータで、ＮＭＦのフィッティング精度、すなわちＮＭＦハイパーパラメータであるｔｈｒｅｓｈｏｌｄの影響を強く受けるので、慎重に吟味する必要がある。経験的に$t_1 = 0.9$ｔｈｒｅｓｈｏｌｄ，$t_2 \in [0.1, 0.3]$程度がよい結果を与える。

[0145] 上記は、ＣＣＡ ｆｉｌｔｅｒをＮＭＦアーティファクト（ノイズ）に適用する例について説明したが、上記のノイズ除去は、ＮＭＦアーティファクトだけでなく、バックグラウンドやコンタミネーションの除去にも適用できる。

[0146] すなわち、ＣＣＡ ｆｉｌｔｅｒで除去できるノイズとしては、例えば、（１）ＮＭＦのアーティファクト（いわゆる偽ピーク）、（２）装置の周辺環境から紛れ込む微量化学物質由来のピーク（バックグラウンド）、（３）

検体中に含まれるが、分析に含めるとその本質をゆがめる恐れのある物質由来のピーク（コンタミネーション）が挙げられる。

[0147] 本方法で扱う検体セットには、バックグランド検体が含まれているため、

[0148] [数43]

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{S}_{out}{}^{T} \\ \boldsymbol{X}_{\delta(BG)} \end{pmatrix}$$

[0149] とすれば、基底スペクトル$S_{:k}$がバックグランドファイルに含まれていたどうかが判定できる。もし含まれていたのであれば、それは環境由来のシステムノイズであるから、$C_{:k}=0$とすればサンプルから当該スペクトルを取り除くことができる。

[0150] なお、更に、バックグラウンド検体がコンタミネーション物質を含む場合、上記と同様の手順によって、コンタミネーション由来のノイズも除去することができる。コンタミネーション物質としては特に制限されないが、例えば、検体の作製やロードに用いた溶媒等が挙げられる。

[0151] 本ステップは更に、強度分布行列の強度補正を行うステップを含んでいてもよい。

[0152] 例えば、ＤＡＲＴ－ＭＳは、空気中の水やアンモニアをヘリウムイオンで一次的にイオン化し、これによりサンプルを二次的にイオン化する。そのため、湿度等、大気条件によってピーク強度が変化する場合がある。またロードする検体質量も検体ごとに異なっているのが普通であり、これもピーク強度に影響を及ぼす。また、検体に含まれる各成分のイオン化効率も一般にはそれぞれ異なり、これもピーク強度に影響を及ぼす。

強度分布行列の強度補正によれば、上記の影響を減じて、より正確な定量結果を得ることができる。

[0153] 全検体が、Ｋ種類の成分を全て含み、かつ、その組成が既知であるキャリブレーション検体を含む場合、上記強度補正は、上記強度分布行列を規格化

することによって実施できる。

[0154] 強度分布行列を規格化するとは、Ｃ行列を検体ごとに規格化することであり、より具体的には、検体ごとにすべての要素を足して１になるようにする操作を意味する。

Ｃ行列をサンプルワイズに分割し、サンプルＬに由来するＣ行列を$C_{\delta(L)}$とかくと、

[0155] [数44]

$$\widehat{C_{\delta(L)}} = \frac{C_{\delta(L)}}{\sum_{i,j}(C_{\delta(L)})_{ij}}, L \ for\ all\ samples$$

[0156] 組成分析においては分解温度の情報は不要であるので、

[0157] [数45]

$$\widehat{C_{\delta(L)}}$$

[0158] を縦方向に加算して得られるベクトル

[0159] [数46]

$$\widehat{c_L} \in \mathbb{R}^K$$

[0160] をサンプルＬの特徴量ベクトルとする。次にエンドメンバーとして測定したｐ個の単成分サンプルの特徴ベクトルを並べて、エンドメンバー行列

[0161] [数47]

$$P \in \mathbb{R}^{K \times p}$$

[0162] を作る。混合サンプルの特徴ベクトル

[0163] [数48]

$$\widehat{c_L}$$

[0164] はエンドメンバーの特徴ベクトルの線形和で書けるので、その係数ベクトル

[0165] [数49]

$$\boldsymbol{r}_L \in \mathbb{R}^p$$

[0166] を使って、

[0167] [数50]

$$\widehat{\boldsymbol{c}_L} = \boldsymbol{P}\boldsymbol{r}_L, subject\ \ to\ \ \boldsymbol{r}_L \geq \boldsymbol{0}, \|\boldsymbol{r}_L\|_1 = 1$$

[0168] と書ける。

一つのキャリブレーション検体は必ず一つ以上の熱分解サンプルを与えるので、熱分解後の成分数はＰは列フルランクな縦長の行列である。したがって一般的には解は不能で、最小二乗の意味での近似解は二次計画法で解くことができる。

[0169] 本方法において得られるマススペクトルでは、イオン化効率が化合物ごとに異なるうえ、熱分解効率も異なり、化合物量とピーク強度は比例しない。

しかし、ＮＭＦによって成分ｋのイオン化効率を含めたピーク強度比は$S_{:k}^{T}$の中にすでに反映されており、スペクトル間の強度比は$C_{:k}$に反映されている。

このＮＭＦによって形成された階層的データ構造のおかげで、エンドメンバーにおけるピーク強度比さえ考慮にいれれば、未知のイオン化効率を求めなくても、より正確な定量的な組成分析が可能になる。

[0170] 一方、検体セットがキャリブレーション検体を含まない場合、強度分布行列の少なくとも一部（好ましくは全部）を、対応する検体の質量と環境変数の積で割り付けることによって、より正確な定量分析が可能となる。

ここで、環境変数とは、観測の際における各成分のイオン化効率への影響を表す変数である。

[0171] イオン化効率が高い成分であれば、ピーク強度は大きくなり、イオン化効率が低い成分であれば、ピーク強度が小さくなる。

また、検体の質量についても、これが大きくなれば、一般的にピーク強度は大きく成り、小さく成れば、ピーク強度は小さくなる。

[0172] 検体セットがキャリブレーション検体を含まない場合、強度分布行列を規格化せず、強度分布行列が保持している検体間の強度情報を用いることができるよう、Ｃ行列を次のように補正する。

[0173] [数51]

$$\widetilde{C_{\delta(L)}} = \frac{C_{\delta(L)}}{m_L I_L}, L \ for \ all \ samples$$

[0174] ここでm_LはサンプルＬの質量、I_Lは６１０ｍ／ｚに全温度域で観測される、ＤＡＲＴ－ＭＳ付属のオイルポンプから生じる（ポリ）ジメチルシロキサンのピーク強度である。I_Lはその測定時の待機状態を反映しており、実際雨の日と晴れの日ではピーク強度に３倍程度の差がある。これにより補正した

[0175] [数52]

$$\widetilde{C_{\delta(L)}}$$

[0176] 縦方向に加算して得られるベクトル

[0177] [数53]

$$\widetilde{C_{\delta(L)}} \in \mathbb{R}^K$$

[0178] をサンプルＬの特徴ベクトルとする。あとは同様にエンドメンバーの特徴ベクトルからエンドメンバー行列

[0179] [数54]

$$P \in \mathbb{R}^{K \times p}$$

[0180] をつくり、

[0181]

[数55]

$$\widetilde{c_{\delta(L)}} = Pr_L, subject\ to\ r_L \geq 0, \|r_L\|_1 = 1$$

[0182] よりr_Lを求める。

[0183] なお、上記ではl_Lは、（ポリ）ジメチルシロキサン由来の６１０ｍ／ｚとしたが、上記に制限されず、観測の際に雰囲気中に所定量に含まれる分子量５０～１５００の化合物に由来するもの、及び／又は、全検体のそれぞれに所定量含まれる分子量５０～５００の有機低分子化合物に由来するものであることが好ましい。

[0184] 図１に戻り、次に、補正後強度分布行列を検体セットに含まれる検体のそれぞれの検体に対応する小行列に分割し、これを特徴ベクトルとしてベクトル空間内で各検体を表現する（ステップＳ１０６）。

さらに、これらの特徴ベクトル（それぞれ各検体に対応する）の全てを内包するＫ－１次元単体を設定し、Ｋ－１次元単体におけるＫ個のエンドメンバーを決定する（ステップＳ１０７）。

[0185] 学習用検体がエンドメンバーを含む場合であって、学習用検体に判別ラベル（当該学習用検体が「エンドメンバー」であると判別するためのラベル）が付されている場合、そのラベルを元にエンドメンバーが決定される。

[0186] 学習用検体がエンドメンバーを含む場合であって、学習用検体に判別ラベルが付されていない場合、すなわち、学習用検体にエンドメンバーが含まれてはいるが、どの学習用検体がエンドメンバーかが不明である場合、特徴ベクトルの頂点成分分析（Ｖｅｒｔｅｘ　Ｃｏｍｐｏｎｅｎｔ　Ａｎａｌｙｓｉｓ）によってエンドメンバーが決定される。

[0187] 学習用検体がエンドメンバーを含まない場合には、Ｋ－１次元単体が最小体積となるように頂点を定めるアルゴリズムによってエンドメンバーが決定される。なお、Ｋ－１次元単体が最小体積になる、とは、２次元単体（３角形）の場合は、最小面積を意味する。

[0188] 本方法の一形態においては、学習用検体がエンドメンバーを含まない場合でも、上記Ｋ－１次元単体に内接する超球体の外側の領域のそれぞれに、学習用検体の特徴ベクトルの少なくとも１つが位置しているため（Ｋが３以上の場合）、精度よくエンドメンバーを決定することができる。

[0189] 図２は、Ｋが３のときのＫ－１次元単体（２次元単体、三角形）を表すイメージ図である。Ｋ－１次元単体１０は、学習用検体１６、１７、及び、１８によって決定されたエンドメンバー１３、１４、及び、１５を頂点とする三角形である。

[0190] 図２のように、学習用検体１６、１７、及び、１８が、Ｋ－１次元単体の内接超球１２（この場合「円」）の外側の領域１９（ハッチング）のそれぞれに位置していると、精度の高い定量分析結果が得られる。

なお、上記はＲｏｂｕｓｔ Ｖｏｌｕｍｅ Ｍｉｎｉｍｉｚａｔｉｏｎ－Ｂａｓｅｄ Ｍａｔｒｉｘ Ｆａｃｔｏｒｉｚａｔｉｏｎ ｆｏｒ Ｒｅｍｏｔｅ Ｓｅｎｓｉｎｇ ａｎｄ Ｄｏｃｕｍｅｎｔ Ｃｌｕｓｔｅｒｉｎｇ，ＩＥＥＥ ＴＲＡＮＳＡＣＴＩＯＮＳ ＯＮ ＳＩＧＮＡＬ ＰＲＯＣＥＳＳＩＮＧ，ＶＯＬ．６４，ＮＯ．２３，ＤＥＣＥＭＢＥＲ １，２０１６をベースに考案された。

[0191] Ｋ－１次元単体が最小体積となるように頂点を定めるアルゴリズムとしては特に制限されず、公知のアルゴリズムが適用できる。このようなアルゴリズムとしては、例えば、Ｍｉｎｉｍｕｍ ｖｏｌｕｍｅ ｃｏｎｓｔｒａｉｎｔ－ＮＭＦ法、Ｍｉｎｉｍｕｍ Ｄｉｓｔａｎｃｅ Ｃｏｎｓｔｒａｉｎｔ－ＮＭＦ法、Ｓｉｍｐｌｅｘ ｉｄｅｎｔｉｆｉｃａｔｉｏｎ ｖｉａ ｓｐｌｉｔ ａｕｇｍｅｎｔｅｄ Ｌａｇｒａｎｇｉａｎ法、及び、Ｓｉｍｐｌｅｘ ｖｏｌｕｍｅ ｍｉｎｉｍｉｚａｔｉｏｎ法等が挙げられる。

[0192] 次に、Ｋ個のエンドメンバーと、推定対象検体の特徴ベクトルとのユークリッド距離をそれぞれ計算し、検体中の各成分の含有量比を推定する（ステップＳ１０８）。

[0193] なお、本方法の他の一形態として、学習用検体が、エンドメンバーを少な

くとも１つ含む（学習用検体の少なくとも１つがエンドメンバーである）場合、他の学習用検体の特徴ベクトルは、Ｋ－１次元単体に内接する超球体の内側の領域に位置していてもよい。図１３は、図２と同様に、Ｋが３のときのＫ－１次元単体（２次元単体、三角形）を表すイメージ図である。Ｋ－１次元単体６０は、エンドメンバーである学習用検体６１、それ以外の学習用検体６２、及び、学習用検体６３によって決定されたエンドメンバー６４、及び、６５を頂点とする三角形である。図２と異なるのは、学習用検体６２、及び、６３が、Ｋ－１次元単体に内接する超球体の内側の領域に位置していることである。

[0194] なお、学習用検体の少なくとも１つがエンドメンバーである場合であっても、他の学習用検体は、Ｋ－１次元単体に内接する超球体上に位置していてもよいし、外側の領域に位置していてもよい。すなわち、この場合、他の学習用検体の位置は任意でよい。

なかでも、より優れた本発明の効果が得られる点では、他の学習用検体は、エンドメンバーである学習用検体とは異なるエンドメンバーの成分（他のエンドメンバーの成分）を２０質量％以上含むことが好ましく、４０質量％以上含むことがより好ましい。

[0195] 本方法によれば、溶媒に溶解しにくく、従来の組成分析が行いにくい固体の検体であっても、特段の前処理なしに分析に供することができ、かつ、正確な定量分析が可能になる。本方法は、特に、複数の成分の混合物である検体の品質管理、及び、不良原因の調査等に適用した場合、結果を得るために必要な時間を大幅に短縮することができる。

[0196] ［本方法の第２実施形態］

本方法の第２実施形態は、すでに説明した第１実施形態におけるステップＳ１０７において、学習用検体がエンドメンバーを含まない場合、すなわち、Ｋ－１次元単体頂点に位置する学習用検体が１つもない場合のエンドメンバー（頂点）の決定方法を含む。

[0197] 本実施形態に係る方法の特徴点の一つは、エンドメンバーの決定が、検体

ごとのフラグメント存在強度を表す行列を非負値行列因子分解し、検体中のＫ種類の成分の重量分率を表す行列と、Ｋ種類の成分の各単体ごとのフラグメント存在強度を表す行列との積に分解する第２のＮＭＦ処理により行われる点にある。

[0198] 図１４は、その仕組みを説明するための模式図である。

各検体は、Ｋ種類の成分の混合物（以下では、「ポリマー」の混合物として説明する）であるが、実際に観測できるのは、各ポリマーが熱分解して生じたM種類のフラグメントガスのスペクトルSの適当な線形和である。

[0199] 本実施形態に係る上記特徴点は、その係数行列を

[数56]

$$\tilde{\boldsymbol{A}} \in \mathbb{R}_{+}^{N\times M}$$

とすれば、

[数57]

$$\tilde{\boldsymbol{X}} \approx \tilde{\boldsymbol{A}}\boldsymbol{S}$$

であり、仮にＫ個のポリマーの単体について測定できたとして、それが

[数58]

$$\boldsymbol{BS} \in \mathbb{R}_{+}^{K\times D}$$

と観測されたとすると、

[数59]

$$\tilde{\boldsymbol{A}} \approx \boldsymbol{CB}$$

と、二度目のＮＭＦ処理を行えば、組成Ｃが得られる、という着想に基づいている。

[0200] なお、ここで

[数60]

$$\boldsymbol{B} \in \mathbb{R}_{+}^{K\times M}$$

はＫ個の単体（ポリマー）から生じるＭ個のフラグメントの存在量を表す。

以下では、図１のフローにおける、ステップＳ１０４のＮＭＦ処理（以下、「１回目のＮＭＦ処理」ともいう。）と、ステップＳ１０７におけるエンドメンバーの決定において実施されるＮＭＦ処理（以下「２回目のＮＭＦ処理」ともいう。）の違い等について詳述する。なお、以下に説明のないステップ等については、原則として、第１実施形態における各ステップと同様である。

[0201] 以下の説明において、第１実施形態における説明で使用した各記号等を以下の表のとおり置き換えて説明する。

[0202] [表1]

項目	置き換え前	置き換え後
データ行列	$\boldsymbol{X}$:N スペクトル×M チャネル	（N サンプル数×T 温度帯域数）×D チャネル これを$\boldsymbol{X} \in \mathbb{R}_{+}^{NT\times D}$とする （非負の$NT$×$D$行列を意味する）
基底フラグメント数	K	M
強度分布行列	$\boldsymbol{C} \in \mathbb{R}_{+}^{N\times K}$	$\boldsymbol{A} \in \mathbb{R}_{+}^{NT\times M}$
基底スペクトル行列	$\boldsymbol{S} \in \mathbb{R}_{+}^{M\times K}$	$\boldsymbol{S} \in \mathbb{R}_{+}^{M\times D}$
固有ベクトル	$(\boldsymbol{a}^T, \boldsymbol{b}^T)^T$	$(\boldsymbol{u}^*, \boldsymbol{v}^*)$
固有ベクトルの行列表示	$\begin{pmatrix}\boldsymbol{A}\\ \boldsymbol{B}\end{pmatrix}$	$\begin{pmatrix}\boldsymbol{U}\\ \boldsymbol{V}\end{pmatrix}$

[0203] また、以下の説明において、多くは信号処理で一般的に使われている表記法に従う。

[数61]

$$\mathbb{R}_{+}^{N\times M},\ \mathbb{R}^{N\times M}$$

[0204] は、は非負、又は、実数のＮ×Ｍ行列を表す。

ある行列

[数62]

$$\boldsymbol{X} \in {\mathbb{R}_+}^{N \times M}$$

に対して、

[数63]

$$\boldsymbol{X}_{n:} \in {\mathbb{R}_+}^{1 \times M}, \boldsymbol{X}_{:m} \in {\mathbb{R}_+}^{N \times 1}, X_{nm} \in \mathbb{R}_+$$

はそれぞれn番目の行ベクトル、m番目の列ベクトル、(n, m)-要素を表す。X^TはＸの転置行列、

[数64]

$$\|\boldsymbol{X}\|_F$$

はフロベニウスノルムを表す。

[数65]

$$\|\boldsymbol{X}_{n:}\|_1 \ , \ \|\boldsymbol{X}_{n:}\|_2$$

はＸのｎ行目の

[数66]

$$\ell_1\text{-norm}, \quad \ell_2\text{-norm}$$

を表す。Ｘが正方行列のとき、Ｔｒ（Ｘ）はトレース、ｄｉａｇ（Ｘ）は対角成分からなる対角行列を表す。1_Nと11_Nはすべての要素が１のＮ次元ベクトル又は（Ｎ，Ｎ）行列を表す。I_NはＮ次元の単位行列を表す。

[0205] （１回目のＮＭＦ）

[数67]

$$\boldsymbol{X} \approx \boldsymbol{AS}$$

[0206] 入力:データ行列

[数68]

$$\boldsymbol{X} \in {\mathbb{R}_+}^{NT \times D}$$

[0207] 出力：分布行列

[数69]

$$\boldsymbol{A} \in \mathbb{R}_{+}^{NT \times M}$$

基底フラグメントスペクトル

[数70]

$$\boldsymbol{S} \in \mathbb{R}_{+}^{M \times D}$$

ここで、Ｎ：サンプル数、Ｔ：温度帯域分割数、Ｄ：チャネル数、Ｍ：基底スペクトル数である。

[0208] 一回目のＮＭＦは、志賀らによって提案されたＡＲＤ－ＳＯ－ＮＭＦに対して以下３点を主な変更点として加え、ＭＳ（Ｍａｓｓ　ｓｐｅｃｔｒｏｍｅｔｒｙ）のデータ解釈によく適合するように開発した。

[0209] １）チャネルごとのガウスノイズの分散共分散行列

[数71]

$$\boldsymbol{R} \in \mathbb{R}^{D \times D}$$

を自然同位体ピークに基づき見積もること

[0210] ２）ソフト直交制約を基底フラグメントスペクトル間に適用すること

[0211] ３）似た強度分布を有するフラグメントスペクトル同士を統合すること（マージング条件の拡張）

[0212] １）については、以下しばらくの間、分散σ^2で独立等分散（ｉ．ｉ．ｄ）のガウスノイズを仮定し、途中で分散共分散行列Ｒを導入する。ノイズにｉ．ｉ．ｄを仮定すると、データ行列Ｘの確率生成モデルは、

[数72]

$$p(\boldsymbol{X}|\boldsymbol{A},\boldsymbol{S},\sigma^2) = \prod_{i=1}^{NT}\prod_{m=1}^{M}\frac{1}{\sqrt{2\pi\sigma^2}}exp\left\{-\frac{(X_{im}-[\boldsymbol{AS}]_{im})^2}{2\sigma^2}\right\}$$

と書ける。基底スペクトル数Ｍは未知なので、自動的に推定できるように、

ａｕｔｏｍａｔｉｃ　ｒｅｌｅｖａｎｃｅ　ｄｅｔｅｒｍｉｎａｔｉｏｎ（ＡＲＤ）を分布行列Ａのスパース性に基づいて導入する。まず、基底成分ごとにλ$_m$でパラメトライズされた指数分布をＡの事前分布として仮定する。

[0213] すなわち、

[数73]

$$\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_M)^T \in \mathbb{R}^M$$

に対して、

[数74]

$$p(A_{im}|\lambda_m) = \frac{1}{\lambda_m}\exp\left(-\frac{A_{im}}{\lambda_m}\right) s.t. \lambda_m > 0, for\ i = 1, \ldots, NT, m = 1, \ldots, M,$$

$$p(\boldsymbol{A}|\boldsymbol{\lambda}) = \prod_{i=1}^{NT}\prod_{m=1}^{M} p(A_{im}|\lambda_m),$$

全体の確率モデルとしては、

[数75]

$$p(\boldsymbol{X}, \boldsymbol{A}, \boldsymbol{S}) = p(\boldsymbol{X}|\boldsymbol{A}, \boldsymbol{S}, \sigma^2)p(\boldsymbol{A}|\boldsymbol{\lambda})p(\boldsymbol{S})p(\boldsymbol{\lambda}|a, b),$$

と書ける。ここで、ｐ（Ｓ）は、

[数76]

$$\|\boldsymbol{S}_{n:}\|_2 = 1$$

の超球面上の一様分布を、ｐ（λ｜ａ，ｂ）は（ａ，ｂ）でパラメトライズされる逆ガンマ分布、すなわち

[数77]

$$p(\boldsymbol{\lambda}|a, b) = \prod_{m=1}^{M} p(\lambda_m|a, b) = \frac{b^a}{\Gamma(a)}{\lambda_m}^{-(a+1)}\exp\left(-\frac{b}{\lambda_m}\right) for\ m = 1, \ldots, M,$$

とする。ａはスパース性を調整するハイパーパラメータで、ａ＝１＋１０$^{-16}$

、ｂは経験的にＡ$_{im}$の期待値Ｅ（Ａ$_{im}$）から見積もれて、

[数78]

$$E(A_{im}) = \frac{b}{a-1} \qquad (1)$$

と関係づけられる。これは更にＥ（Ｘ$_{id}$）と

[数79]

$$E(X_{id}) = \sum_{m=1}^{M} E(A_{im})E(S_{md}) = M \cdot E(A_{im})E(S_{md}),$$

なる関係があり、Ｅ（Ｘ$_{id}$）をＸの平均μ_Xで近似すると、

[数80]

$$\mu_X = M\frac{b}{(a-1)\sqrt{D}}$$

よりｂは、

[数81]

$$b = \frac{\mu_X(a-1)\sqrt{D}}{M}, \qquad (2)$$

と定まる。したがって負対数尤度関数は、

[数82]

$$L(\boldsymbol{X},\boldsymbol{A},\boldsymbol{S},\boldsymbol{\lambda}) = -\log[p(\boldsymbol{X}|\boldsymbol{A},\boldsymbol{S})p(\boldsymbol{A}|\boldsymbol{\lambda})p(\boldsymbol{\lambda}|a,b)]$$

$$= \frac{1}{2\sigma^2}\|\boldsymbol{X}-\boldsymbol{A}\boldsymbol{S}\|_F^2 + \frac{DNT}{2}log2\pi\sigma^2 + (NT+a+1)\sum_{m=1}^{M} log\lambda_m$$

$$+ \sum_{m=1}^{M}\frac{1}{\lambda_m}\left(b + \sum_{i=1}^{NT} A_{im}\right) + M(log\Gamma(a) - alogb).$$

と書ける。これはλに対して下に凸な関数なので、

[数83]

$$\frac{\partial L}{\partial \boldsymbol{\lambda}} \equiv \mathbf{0}$$

すなわち、

[数84]

$$\lambda_m = \frac{b + \sum_{i=1}^{NT} A_{im}}{NT + a + 1}, for\ m = 1, \ldots, M \qquad (S3)$$

としてλに関する更新式が得られる。ここまでは志賀らの報告したARD-SO-NMFの導出に完全に従った。

[0214] ノイズに対するｉ．ｉ．ｄガウス分布の仮定はＭＳデータについては当てはまらず、大きいシグナルほど大きなノイズを有する傾向があることが知られている。そこで、ノイズ分布

[数85]

$$\mathbf{E} \in \mathbb{R}^{NT \times D}$$

を自然同位体ピークによる線形回帰の残渣成分として求めた。すなわち、

[数86]

$$\boldsymbol{E}_{:d} = \boldsymbol{X}_{:d} - \boldsymbol{M}^{(d)}\left(\boldsymbol{M}^{(d)T}\boldsymbol{M}^{(d)}\right)^{-1}\boldsymbol{M}^{(d)T}\boldsymbol{X}_{:d}, for\ d = 1, \ldots, D, \qquad (S4)$$

ここで

[数87]

$$\boldsymbol{M}^{(d)} = \boldsymbol{X}_{:[d-30, d-20, d-10, d+10, d+20, d+30]}$$

はチャネルｄを中心にした［ｄ－３０，ｄ－２０，ｄ－１０，ｄ＋１０，ｄ＋２０，ｄ＋３０］なるチャネルで、チャネルｄの±３ｍ／ｚの同位体ピークである（チャネル間隔が０．１ｍ／ｚであることに留意）。チャネルごとの分散共分散行列

[数88]

$$\boldsymbol{R} \in \mathbb{R}^{D\times D}$$

は

[数89]

$$\boldsymbol{R} = \frac{1}{NT}\boldsymbol{E}^{\boldsymbol{T}}\boldsymbol{E}$$

として求まる。これを用いて尤度関数を書き直すと、

[数90]

$$p(\boldsymbol{X}|\boldsymbol{A},\ \boldsymbol{S},\boldsymbol{R}) = \frac{1}{\sqrt{2\pi}^{DNT}\sqrt{|\boldsymbol{R}|}^{NT}}\exp\{-Tr[(\boldsymbol{X}-\boldsymbol{AS})\boldsymbol{R}^{-1}(\boldsymbol{X}-\boldsymbol{AS})^{T}]\}.$$

となる。Rは定数行列であることに注意する。全体の負対数尤度関数は、

[数91]

$$\begin{aligned} L(\boldsymbol{X},\boldsymbol{A},\boldsymbol{S},\boldsymbol{\lambda}) &= -\log[p(\boldsymbol{X}|\boldsymbol{A},\boldsymbol{S},\boldsymbol{R})p(\boldsymbol{A}|\boldsymbol{\lambda})p(\boldsymbol{\lambda}|a,b)] \\ &= Tr[(\boldsymbol{X}-\boldsymbol{AS})\boldsymbol{R}^{-1}(\boldsymbol{X}-\boldsymbol{AS})^{T}] + \frac{DNT}{2}log2\pi + \frac{NT}{2}log|\boldsymbol{R}| \\ &+ (NT+a+1)\sum_{m=1}^{M} log\lambda_m + \sum_{m=1}^{M}\frac{1}{\lambda_m}\left(b+\sum_{i=1}^{NT}A_{im}\right) \\ &+ M(log\Gamma(a) - alogb). \end{aligned}$$

となる。λについての更新式Ｅｑ．Ｓ３を代入し、定数項を落として簡略化すると、

[数92]

$$L(\boldsymbol{X},\boldsymbol{A},\boldsymbol{S},\boldsymbol{\lambda}) = Tr[(\boldsymbol{X}-\boldsymbol{AS})\boldsymbol{R}^{-1}(\boldsymbol{X}-\boldsymbol{AS})^{T}] + (NT+a+1)\sum_{m=1}^{M} log\lambda_m,$$

となる。このＡ、Ｓに関する最小化はｈｉｅｒａｒｃｈｉｃａｌ　ａ

Ｉｔｅｒｎａｔｉｎｇ ｌｅａｓｔ ｓｑｕａｒｅ （ＨＡＬＳ）により行う。簡便のため、以下のベクトル表記を用いる。

[数93]

$$\boldsymbol{a}_m \equiv \boldsymbol{A}_{:m}, \boldsymbol{s}_m \equiv \boldsymbol{S}_{m:}^{T}, for\ m = 1, .., M.$$

ＨＡＬＳでは残差Ｘ－ＡＳを

[数94]

$$\boldsymbol{X}^{(m)} - \boldsymbol{a}_m \boldsymbol{s}_m{}^{T}\ (m = 1, \ldots, M)$$

として表す。ここで

[数95]

$$\boldsymbol{X}^{(m)} = \boldsymbol{X} - \boldsymbol{AS} + \boldsymbol{a}_m \boldsymbol{s}_m{}^{T}$$

である。これによりＬ（Ｘ，Ａ，Ｓ，λ）は成分ｍごとに分離して書けて、

[数96]

$$L\left(\boldsymbol{X}, \boldsymbol{a}_m, \boldsymbol{s}_m, \lambda_m(\boldsymbol{a}_m)\right)$$

$$= Tr\left[\left(\boldsymbol{X}^{(m)} - \boldsymbol{a}_m \boldsymbol{s}_m{}^{T}\right)\boldsymbol{R}^{-1}\left(\boldsymbol{X}^{(m)} - \boldsymbol{a}_m \boldsymbol{s}_m{}^{T}\right)^{T}\right] + (NT + a + 1) log \lambda_m,$$

$$for\ m = 1, \ldots, M.$$

Ｓについてのソフト直交制約はペナルティー項

[数97]

$$w_O \xi_m \boldsymbol{s}_m^{T} \boldsymbol{s}^{(m)}$$

として目的関数Ｌに組み込むことができる。ここで

[数98]

$$\boldsymbol{s}^{(m)} \equiv \sum_{j \neq m}^{M} \boldsymbol{s}_j$$

である。この項はｍ－成分とその他成分の非直交性を表し、ξ$_m$は厳密に直交性が満たされる際のラグランジェ未定乗数を表す。この項は更にｗ$_O$∈［０，１］によって緩和される。したがって最小化すべき目的関数は、

[数99]

$$L(\boldsymbol{X}, \boldsymbol{a}_m, \boldsymbol{s}_m, \lambda_m(\boldsymbol{a}_m)) = Tr\left[\left(\boldsymbol{X}^{(m)} - \boldsymbol{a}_m \boldsymbol{s}_m{}^T\right)\boldsymbol{R}^{-1}\left(\boldsymbol{X}^{(m)} - \boldsymbol{a}_m \boldsymbol{s}_m{}^T\right)^T\right] + (NT + a + 1)log\lambda_m + w_O\, \xi_m \boldsymbol{s}_m^T \boldsymbol{s}^{(m)},$$

となり、ａ$_m$とｓ$_m$に関する勾配は、

[数100]

$$\frac{\partial L}{\partial \boldsymbol{a}_m} = \left(\boldsymbol{a}_m \boldsymbol{s}_m{}^T - \boldsymbol{X}^{(m)}\right)\boldsymbol{R}^{-1}\boldsymbol{s}_m + \frac{1}{\lambda_m}\mathbf{1}_{NT},$$

$$\frac{\partial L}{\partial \boldsymbol{s}_m} = \boldsymbol{R}^{-1}\left(\boldsymbol{s}_m \boldsymbol{a}_m^T - \boldsymbol{X}^{(m)^T}\right)\boldsymbol{a}_m + w_O\, \xi_m \boldsymbol{s}^{(m)},$$

となり、これらをゼロと置けば、更新式は

[数101]

$$\boldsymbol{a}_m = \frac{\boldsymbol{X}^{(m)}\boldsymbol{R}^{-1}\boldsymbol{s}_m - \frac{1}{\lambda_m}\mathbf{1}_{NT}}{\boldsymbol{s}_m^T \boldsymbol{R}^{-1}\boldsymbol{s}_m}, \qquad (S5)$$

$$\boldsymbol{s}_m = \boldsymbol{X}^{(m)^T}\boldsymbol{a}_m - w_O\, \xi_m \boldsymbol{R}\boldsymbol{s}^{(m)}, \qquad (S6)$$

となる。ｓ$_m$は更新毎に規格化されるので、定数係数は省略した。非負制約は、更新毎に非負象限に射影することにより満たした。具体的には例えば、

[数102]

$$\boldsymbol{a}_m \leftarrow \frac{\boldsymbol{a}_m + |\boldsymbol{a}_m|}{2},$$

によって射影できる。ここで

[数103]

$$|\boldsymbol{a}_m|$$

は要素ごとに絶対値を取ったベクトルを意味する。未定乗数ξ$_m$は

[数104]

$$\boldsymbol{s}^{(m)^T}\boldsymbol{R}$$

をＥｑ．Ｓ６の左側から乗じ、厳密な直交条件、

[数105]

$$\boldsymbol{s}^{(m)^T}\boldsymbol{s}_m = 0$$

とｗ$_0$＝１を適用することで、

[数106]

$$-\boldsymbol{s}^{(m)^T}\boldsymbol{X}^{(m)^T}\boldsymbol{a}_k + \xi_m \boldsymbol{s}^{(m)^T}\boldsymbol{R}\boldsymbol{s}^{(m)} = 0,$$

$$\xi_m = \frac{\boldsymbol{s}^{(m)^T}\boldsymbol{X}^{(m)^T}\boldsymbol{a}_k}{\boldsymbol{s}^{(m)^T}\boldsymbol{R}\boldsymbol{s}^{(m)}} \qquad (S7)$$

と得られる。以上Ｅｑ．Ｓ１－Ｓ７を用いて以下のアルゴリズム１を提案する。

[0215]

[数107]

Algorithm 1: Pseudo-code of the first NMF

Input: ℓ_1-normalized data; $\boldsymbol{X} \in \mathbb{R}_+^{NT\times D}$, orthogonal constraint weight; w_O, initial component number; M, iteration number; itr, margining threshold; $t \in [0.9, 0.99]$.

Output: spectra-wise fragment abundance matrix $\boldsymbol{A} \in \mathbb{R}_+^{NT\times M}$, M-fragment spectra $\boldsymbol{S} \in \mathbb{R}_+^{M\times D}$ with optimized M.

Initialization

initialize $\boldsymbol{A}$ by Eq. S1

initialize $\boldsymbol{S}$ by random selection of M-row vectors from $\boldsymbol{X}$ and ℓ_2-normalization.

calculate b by Eq. S2

initialize $\boldsymbol{\lambda}$ by Eq. S3

calculate $\boldsymbol{E}$ by Eq. S4 and $\boldsymbol{R} = \frac{1}{NT}\boldsymbol{E}^T\boldsymbol{E}$

Repeat until convergence criteria are satisfied:

for $m = 1, \ldots, M$:

calculate $\boldsymbol{s}^{(m)} \leftarrow \sum_{j\neq m}^{M} \boldsymbol{s}_j$ and calculate ξ_m by Eq. S7

update $\boldsymbol{s}_m$ by Eq. S6.

project $\boldsymbol{s}_m$ to the non-negative orthant

ℓ_2-normalize $\boldsymbol{s}_m$

for $m = 1, \ldots, M$:

update $\boldsymbol{a}_m$ by Eq. S5

project $\boldsymbol{a}_m$ to the non-negative orthant

Merging very similar components

for $k = 1, \ldots, M-1$:

for $m = k, \ldots, M$:

if $\boldsymbol{s}_k^T\boldsymbol{s}_m > t$:

$\boldsymbol{a}_k \leftarrow \boldsymbol{a}_k + \boldsymbol{a}_m,\ \boldsymbol{a}_m \leftarrow \boldsymbol{0}$

if $\boldsymbol{a}_k^T\boldsymbol{a}_m > t\|\boldsymbol{a}_k\|\|\boldsymbol{a}_m\|$:

$\boldsymbol{s}_k \leftarrow \boldsymbol{s}_k + \frac{\|\boldsymbol{a}_m\|_1}{\|\boldsymbol{a}_k\|_1}\boldsymbol{s}_m,\ s = \|\boldsymbol{s}_k\|_2,\ \boldsymbol{s}_k \leftarrow \frac{\boldsymbol{s}_k}{s}$

$\boldsymbol{a}_k \leftarrow s\boldsymbol{a}_k, \boldsymbol{a}_m \leftarrow \boldsymbol{0}$

update $\boldsymbol{\lambda}$ by Eq. S3

[0216] ここで、ＡとＳの更新毎に、似た成分同士をマージングするようにする。同様のスペクトルを有する成分同士をマージすることが志賀らに提案されているが、ここではさらに、同様の強度分布を有する成分同士もマージングす

ることにした。これにより、同位体ピーク、異なるイオン付加によりイオン化されたフラグメントシリーズ、モノマー数違いのオリゴマーピークシリーズなどを一つの成分とすることで、より解釈性の高い結果を与えることができる。

[0217] （Ｃａｎｏｎｉｃａｌ ｃｏｒｒｅｌａｔｉｏｎ ａｎａｌｙｓｉｓ ｆｉｌｔｅｒ （ＣＣＡ－ｆｉｌｔｅｒ）の導出）

[0218] 入力：１回目ＮＭＦの出力

[数108]

$$\boldsymbol{S} \in \mathbb{R}_{+}^{M \times D}$$

、バックグラウンドスペクトル

[数109]

$$\boldsymbol{X}_{BG} \in \mathbb{R}_{+}^{T \times D}$$

[0219] 出力：バックグランド由来と判定された成分番号のリスト

[0220] １回目のＮＭＦで得られたＭ－成分の中には、バックグランドや混入物に由来する成分が含まれ、これらは組成分析結果をゆがめる可能性があるため、ＡやＳから取り除くことが好ましい。Ｍ′－成分がバックグラウンド由来の成分だとＣＣＡ－ｆｉｌｔｅｒにより判定されたとすると、

[数110]

$$\boldsymbol{A} \in \mathbb{R}_{+}^{NT \times M}$$

及び

[数111]

$$\boldsymbol{S} \in \mathbb{R}_{+}^{M \times D}$$

はそれぞれ、

[数112]

$$\boldsymbol{A} \in \mathbb{R}_{+}^{NT \times (M - M')}$$

と

[数113]

$$\boldsymbol{S} \in \mathbb{R}_{+}^{(M-M')\times D}$$

になる。単純のため、ＣＣＡ－ｆｉｌｔｅｒ適用後の成分数はＭ－Ｍ′になるが、一貫してＭを用いる。

ＣＣＡ－ｆｉｌｔｅｒを用いるためには、バックグラウンドスペクトル

[数114]

$$\boldsymbol{X}_{BG} \in \mathbb{R}_{+}^{T\times D}$$

をデータセットに含めて、サンプルスペクトルと併せて１回目のＮＭＦを行う必要がある。もし何らかの混入物が予想できる場合は、その混入物について測定したスペクトルをX_{BG}として用いることができる。ＣＣＡ－ｆｉｌｔｅｒは成分ｍ＝１，…，Ｍについてひとつずつ、そのスペクトル$S_{m:}$がX_{BG}に含まれているか確認していく。

最初のステップは、Ｍ－スペクトルからなる

[数115]

$$\ell_2\text{-normalized}$$

されたSを、$S_{m:}$に似たスペクトルセット

[数116]

$$\boldsymbol{Y} \in \mathbb{R}_{+}^{M_{sim}\times D}$$

と、似ていないスペクトルセット

[数117]

$$\boldsymbol{Z} \in \mathbb{R}_{+}^{M_{dis}\times D}$$

に分割することである。この分割は、

[数118]

$$\boldsymbol{S}_{m:}\boldsymbol{Y}_{m':}^{T} \geq t_1, for\ m' = 1, \ldots, M_{sim},$$
$$\boldsymbol{S}_{m:}\boldsymbol{Z}_{m':}^{T} < t_1, for\ m' = 1, \ldots, M_{dis},$$

が満たされるように行う。ｔ$_1$∈［０，１］はある閾値で本発明では一貫してｔ$_1$＝０．２を用いた。なお、Ｓ$_{m:}$は常にＹに含まれるので、Ｙの一列目にＳ$_{m:}$を格納した。ＺはＸ$_{BG}$と結合し、

[数119]

$$\boldsymbol{Z} \leftarrow \begin{pmatrix} \boldsymbol{Z} \\ \boldsymbol{X}_{BG} \end{pmatrix},$$

とした。ＹとＺは、平均ゼロとなるように、

[数120]

$$\bar{\boldsymbol{Y}} = \boldsymbol{Y}\left(\boldsymbol{I}_D - \frac{1}{D}\boldsymbol{1}\boldsymbol{1}_D\right) \in \mathbb{R}^{M_{sim} \times D}, \qquad (S8)$$

$$\bar{\boldsymbol{Z}} = \boldsymbol{Z}\left(\boldsymbol{I}_D - \frac{1}{D}\boldsymbol{1}\boldsymbol{1}_D\right) \in \mathbb{R}^{(M_{dis}+T) \times D}, \qquad (S9)$$

とした。これら二つのスペクトルセットに対して、ＣＣＡを行う。ＣＣＡは、二つのスペクトルセット内の線形結合により、なるべく似たペアのスペクトルを作り出す。

[数121]

$$\bar{\boldsymbol{Y}}$$

[0221] 及び

[数122]

$$\bar{\boldsymbol{Z}}$$

の線形結合の係数がベクトル

[数123]

$$\boldsymbol{u} \in \mathbb{R}^{M_{sim}}$$

及び

[数124]

$$\boldsymbol{v} \in \mathbb{R}^{M_{dis}+T}$$

に格納されているとする。従ってスペクトルペアはそれぞれ、

[数125]

$$\boldsymbol{y} \equiv \boldsymbol{u}^T\overline{\boldsymbol{Y}} \in \mathbb{R}^{1\times D}$$

及び

[数126]

$$\boldsymbol{z} \equiv \boldsymbol{v}^T\overline{\boldsymbol{Z}} \in \mathbb{R}^{1\times D}$$

と書ける。その類似度は相関係数ρによって評価され、

[数127]

$$\rho = \frac{\boldsymbol{u}^T\boldsymbol{V}_{yz}\boldsymbol{v}}{\sqrt{\boldsymbol{u}^T\boldsymbol{V}_{yy}\boldsymbol{u}}\sqrt{\boldsymbol{v}^T\boldsymbol{V}_{zz}\boldsymbol{v}}},$$

ここで、

[数128]

$$\boldsymbol{V}_{YY} = \overline{\boldsymbol{Y}}\overline{\boldsymbol{Y}}^T/D, \boldsymbol{V}_{ZZ} = \overline{\boldsymbol{Z}}\overline{\boldsymbol{Z}}^T/D, \boldsymbol{V}_{YZ} = \overline{\boldsymbol{Y}}\overline{\boldsymbol{Z}}^T/D.$$

ＣＣＡの問題設定は、

[数129]

$$(\boldsymbol{u}^*, \boldsymbol{v}^*) = \underset{\boldsymbol{u},\boldsymbol{v}}{arg\,max}\,\rho.$$

と書ける。その解

[数130]

$$\begin{pmatrix}\boldsymbol{u}\\ \boldsymbol{v}\end{pmatrix} \in \mathbb{R}^{M_{sim}+M_{dis}+T}$$

は一般固有値問題の解として与えられて、

[数131]

$$\begin{pmatrix} \boldsymbol{O} & \boldsymbol{V}_{yz} \\ \boldsymbol{V}_{yz}^{T} & \boldsymbol{O} \end{pmatrix}\begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{V} \end{pmatrix} = \begin{pmatrix} \boldsymbol{V}_{yy} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{V}_{zz} \end{pmatrix}\begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{V} \end{pmatrix}\begin{pmatrix} \rho_1 & & \\ & \ddots & \\ & & \rho_{M_{sim}+M_{dis}+T} \end{pmatrix}, \rho_1 \geq \cdots \geq \rho_{M_{sim}+M_{dis}+T}, (S10)$$

ここで

[数132]

$$\begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{V} \end{pmatrix}$$

は固有ベクトル

[数133]

$$\begin{pmatrix} \boldsymbol{u}* \\ \boldsymbol{v}* \end{pmatrix}$$

をその固有値が大きい順に列ベクトルとして並べて得られる行列である。

各固有値

[数134]

$$(\rho_1, \cdots, \rho_{M_{sim}+M_{dis}+T})$$

は、対応する（ｕ*，ｖ*）を係数として線形結合で作られたｙとｚの相関係数になっている。ここではρ＞ｔ$_2$を満たすすべての固有値に対応するｕ*を取り出し、もしＳ$_{m:}$の係数に対応する１番目の要素がｕ*に大きく寄与する大きな成分であれば、すなわち

[数135]

$$\frac{|u_1|}{\|\boldsymbol{u}^*\|_1} \geq t_3$$

であれば、成分ｍはバックグラウンドに由来する成分と判定して、システムから除く。ここでｔ$_2$∈［０．９，０．９９］、ｔ$_3$∈［０，１］である。これを以下にアルゴリズム２としてまとめた。

[数136]

Algorithm 2: CCA-filter

Input: ℓ_1-normalized basis-spectra of the 1stNMF output; $\boldsymbol{S} \in \mathbb{R}_+^{M \times D}$, a background spectrum; $\boldsymbol{X}_{BG} \in \mathbb{R}_+^{T \times D}$, thresholds: $t_1 \in [0,1], t_2 \in [0,1], t_3 \in [0,1]$, in this study $(t_1, t_2, t_3) = (0.2, 0.9, 0.5)$ is consistently used.

Output: a list of the components judged as background components.

for $m = 1, \ldots, M$:

  for $m' = 1, \ldots, M$:

    classify $\boldsymbol{S}_{m':}$ into $\boldsymbol{Y}$ if $\boldsymbol{S}_{m:}\boldsymbol{S}_{m':}^T \geq t_1$ else into $\boldsymbol{Z}$.

  $\boldsymbol{Z} \leftarrow \begin{pmatrix} \boldsymbol{Z} \\ \boldsymbol{X}_{BG} \end{pmatrix}$

  calculate $\overline{\boldsymbol{Y}}$ and $\overline{\boldsymbol{Z}}$ by Eq. S8 and Eq. S9

  calculate $\boldsymbol{V}_{YY} = \overline{\boldsymbol{Y}}\overline{\boldsymbol{Y}}^T/D, \boldsymbol{V}_{ZZ} = \overline{\boldsymbol{Z}}\overline{\boldsymbol{Z}}^T/D, \boldsymbol{V}_{YZ} = \overline{\boldsymbol{Y}}\overline{\boldsymbol{Z}}^T/D$

  obtain $\boldsymbol{U}^* = (\boldsymbol{u}^*{}_1, \ldots, \boldsymbol{u}^*{}_{M_{sim}+M_{dis}+T})$ and $(\rho_1, .., \rho_{M_{sim}+M_{dis}+T})$ by solving Eq. S10

  find Q such that $\rho_Q \geq t_2$ *and* $\rho_{Q+1} < t_2$

  for $q = 1, \ldots, Q$:

    if $\frac{|\boldsymbol{U}^*{}_{1q}|}{\|\boldsymbol{U}^*{}_{:q}\|_1} \geq t_3$, add m in the background-component list

return the background-component list

[0222] バックグラウンド成分が特定されたあとは、対応するＡの列ベクトル、及びＳの行ベクトルを削除することで、システムから除いた。

ＣＣＡ－ｆｉｌｔｅｒから出力された、バックグラウンドに由来するＭ′－成分除去済みの

[数137]

$$\boldsymbol{A} \in \mathbb{R}_+^{NT \times (M - M')}$$

（以下、簡単のためＭ－Ｍ′ をＭと置き換えて、

[数138]

$$\boldsymbol{A} \in \mathbb{R}_+^{NT \times M}$$

と表す）は、サンプル量及び内部標準ピークで割り付け強度を補正する。その後、サンプルｎに関するＡの小行列部分Ａ(n)を一次元ベクトル化した

[数139]

$$\boldsymbol{a}_n = vec(\boldsymbol{A}^{(n)})$$

をサンプルｎの特徴ベクトルとして二回目のＮＭＦの入力としてもよいが、組成分析においてはＭ－フラグメントの温度分布は不要な情報のため、サンプル毎にすべての温度帯域を足し合わせてサンプル毎のＭ－フラグメント存在強度（ｆｒａｇｍｅｎｔ　ａｂｕｎｄａｎｃｅ；　ＦＡ）を表す

[数140]

$$\widetilde{\boldsymbol{A}} \in \mathbb{R}_{+}^{N\times M}$$

を入力としてもよい。ここでは簡単のため

[数141]

$$\widetilde{\boldsymbol{A}}$$

について二回目のＮＭＦを行うことにする。

[0223] 以下のセクションでは、Ｎｏｎ－ｎｅｇａｔｉｖｅ　ｌｅａｓｔ　ｓｑｕａｒｅ　（ＮＮＬＳ）によるフィッティングを頻繁に使う。この問題は、最適な非負係数

[数142]

$$\boldsymbol{x} \in \mathbb{R}_{+}^{n}$$

を用いて、定数行列

[数143]

$$\boldsymbol{\Phi} \in \mathbb{R}^{m\times n}$$

の列ベクトル線形結合により

[数144]

$$\boldsymbol{y} \in \mathbb{R}^{m}$$

を近似するもので、

[数145]

$$\boldsymbol{x}^{*} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \|\boldsymbol{y} - \boldsymbol{\Phi x}\|^{2}, s.t. \boldsymbol{x} \geq \boldsymbol{0},$$

により求められる。これを解くには、ａｌｔｅｒｎａｔｉｎｇ　ｄｉｒｅｃｔｉｏｎ　ｍｕｌｔｉｐｌｉｅｒ　ｍｅｔｈｏｄｓ　（ＡＤＭＭ）（１）を含む数多くの最適化手法が使えるが、ここではＦｕ．　ｅｔ．　ａｌ．　（２）により開発されたＡＤＭＭ－ＮＮＬＳを用いてこの問題をとき、解を

[数146]

$$\boldsymbol{x}^* = NNLS(\boldsymbol{y}, \boldsymbol{\Phi})$$

と表記する。近似するベクトルセットＹ＝［ｙ$_1$，．．．，ｙ$_L$］に対応する非負係数ベクトルｘ$_l$*（ｌ＝１，．．．，Ｌ）は個別に計算できて、

[数147]

$$\boldsymbol{x}_l{}^* = NNLS(\boldsymbol{Y}_{:l}, \boldsymbol{\Phi})\ for\ l = 1,..,L,$$

または行列形式で、

[数148]

$$\boldsymbol{X}^* = NNLS(\boldsymbol{Y}, \boldsymbol{\Phi}),$$

と書くことにする。ここで

[数149]

$$\boldsymbol{X}^* = [\boldsymbol{x}_1{}^*,..,\boldsymbol{x}_L{}^*]$$

である。同様の問題で、係数ベクトルの総和が１になるような拘束条件付の問題はｆｕｌｌｙ　ｃｏｎｓｔｒａｉｎｅｄ　ｌｅａｓｔ　ｓｑｕａｒｅ（ＦＣＬＳ）とよばれ、ＡＤＭＭを用いて同じく解くことができて、

[数150]

$$\boldsymbol{x}_l{}^* = FCLS(\boldsymbol{Y}_{:l}, \boldsymbol{\Phi})\ for\ l = 1,..,L,$$

または行列表記で、

[数151]

$$\boldsymbol{X}^* = FCLS(\boldsymbol{Y}, \boldsymbol{\Phi}),$$

と書くことにする。

[0224] （２回目のＮＭＦ）

[0225] 入力：サンプル毎のＦＡを表す

[数152]

$$\widetilde{\boldsymbol{A}} \in \mathbb{R}_{+}^{N\times M}$$

、基底Ｍ－フラグメントごとのスペクトル

[数153]

$$\boldsymbol{S} \in \mathbb{R}_{+}^{M\times D}$$

、ポリマー成分数K

[0226] 出力：サンプル毎のポリマー重量分率

[数154]

$$\boldsymbol{C} \in \mathbb{R}_{+}^{N\times K}$$

、単位重量Ｋ－基底ポリマーごとのＦＡ

[数155]

$$\boldsymbol{B} \in \mathbb{R}_{+}^{K\times M}$$

[0227] ２回目のＮＭＦ：

[数156]

$$\widetilde{\boldsymbol{A}} \approx \boldsymbol{CB}$$

は、リーマン計量を用いてその近似残差

[数157]

$$D_{\boldsymbol{G}}\left(\widetilde{\boldsymbol{A}}\middle|\boldsymbol{CB}\right)$$

を評価する、すなわち、

[数158]

$$D_G\left(\widetilde{\boldsymbol{A}}\middle|\boldsymbol{CB}\right) = Tr\left[\left(\widetilde{\boldsymbol{A}} - \boldsymbol{CB}\right)\boldsymbol{G}\left(\widetilde{\boldsymbol{A}} - \boldsymbol{CB}\right)^T\right],$$

ここで、

[数159]

$$\boldsymbol{G} = \boldsymbol{SS}^T \in \mathbb{R}_+^{M\times M}$$

である。下三角行列

[数160]

$$\boldsymbol{L} \in \mathbb{R}^{M\times M}$$

をコレスキー分解Ｇ＝ＬＬ^{T}により得れば、

[数161]

$$D_G(\widetilde{\boldsymbol{A}}|\boldsymbol{CB}) = Tr\left[(\widetilde{\boldsymbol{A}} - \boldsymbol{CB})\boldsymbol{LL}^T(\widetilde{\boldsymbol{A}} - \boldsymbol{CB})^T\right] = Tr\left[(\widehat{\boldsymbol{A}} - \boldsymbol{C}\widehat{\boldsymbol{B}})(\widehat{\boldsymbol{A}} - \boldsymbol{C}\widehat{\boldsymbol{B}})^T\right] = \|\widehat{\boldsymbol{A}} - \boldsymbol{C}\widehat{\boldsymbol{B}}\|_F^2,$$

と書ける。ここで

[数162]

$$\widehat{\boldsymbol{A}} = \widetilde{\boldsymbol{A}}\boldsymbol{L} \in \mathbb{R}^{N\times M}, \widehat{\boldsymbol{B}} = \boldsymbol{BL} \in \mathbb{R}^{K\times M}$$

であり、

[数163]

$$\widehat{\boldsymbol{A}} \approx \boldsymbol{C}\widehat{\boldsymbol{B}}$$

と行列分解することと等値である。

[数164]

$$\widehat{\boldsymbol{B}}$$

の行ベクトルで張られるシンプレックスの体積項を

[数165]

$$vol\left(\widehat{\boldsymbol{B}}\right), \widehat{\boldsymbol{B}}$$

の行ベクトル間の非直交性項を

[数166]

$$nonorth(\widehat{\boldsymbol{B}})$$

として、α＞０とβ∈［０，１］を重みとすれば、

[数167]

$$(\boldsymbol{C}^*, \widehat{\boldsymbol{B}}^*) = \underset{\boldsymbol{C},\widehat{\boldsymbol{B}}}{arg\,min} \frac{1}{2}\left\|\widehat{\boldsymbol{A}} - \boldsymbol{C}\widehat{\boldsymbol{B}}\right\|_F^2 + \frac{\alpha}{2} vol(\widehat{\boldsymbol{B}}) + \frac{\beta}{2} nonorth(\widehat{\boldsymbol{B}}), \qquad (S11)$$
$$s.t. \widehat{\boldsymbol{B}} = \boldsymbol{B}\boldsymbol{L}, \boldsymbol{B} \geq 0, \boldsymbol{C} \geq 0, \boldsymbol{C}\boldsymbol{1}_K = \boldsymbol{1}_N.$$

により

[数168]

$$(\boldsymbol{C}^*, \widehat{\boldsymbol{B}}^*)$$

が求まる。ここで、Ｆｕらは（３）外れ値へのロバスト性を付与するために、

[数169]

$$w_n = \frac{p}{2}\left(\left\|\widehat{\boldsymbol{A}}_{n:} - \boldsymbol{C}_{n:}\widehat{\boldsymbol{B}}\right\|^2 + \varepsilon\right)^{\frac{p-2}{2}}$$
$$\boldsymbol{W} = diag(w_1, \ldots w_N)$$

なる重み行列の導入を提案している。ここで、ｐ∈（０，２］、かつεは小さな正則化パラメータである。本発明では、一貫してｐ＝１．５かつε＝１０$^{-8}$とした。最適化問題は

[数170]

$$\min_{\boldsymbol{C},\widehat{\boldsymbol{B}}} \frac{1}{2} Tr\left[\boldsymbol{W}\left(\widehat{\boldsymbol{A}} - \boldsymbol{C}\widehat{\boldsymbol{B}}\right)\left(\widehat{\boldsymbol{A}} - \boldsymbol{C}\widehat{\boldsymbol{B}}\right)^T\right] + \frac{\alpha}{2} vol(\widehat{\boldsymbol{B}}) + \frac{\beta}{2} nonorth(\widehat{\boldsymbol{B}}),$$
$$s.t. \widehat{\boldsymbol{B}} = \boldsymbol{B}\boldsymbol{L}, \boldsymbol{B} \geq 0, \boldsymbol{C} \geq 0, \boldsymbol{C}\boldsymbol{1}_K = \boldsymbol{1}_N.$$

となる。Ｃと

[数171]

$\widehat{\boldsymbol{B}}$

はｂｌｏｃｋ　ｃｏｏｒｄｉｎａｔｅ　ｄｅｓｃｅｎｔ　（ＢＣＤ）　理論により交互に最適化する。今、ｔ回の最適化を行い、

[数172]

$(\boldsymbol{C}^{(t)}, \widehat{\boldsymbol{B}}^{(t)})$

を得たとする。なお、初期化には、ｖｅｒｔｅｘ　ｃｏｍｐｏｎｅｎｔ　ａｎａｌｙｓｉｓ　（ＶＣＡ）（４）を用いて、

[数173]

$\widehat{\boldsymbol{A}}$

のＮ－行ベクトルの中からシンプレックスのＫ－頂点に近いＫ個の行ベクトルを選択し、これを

[数174]

$\widehat{\boldsymbol{B}}^{(0)}$

とした。

[数175]

$\widehat{\boldsymbol{B}}^{(t)}$

に基づくＣのアップデートは、

[数176]

$$\boldsymbol{C}^{(t+1)^T} = FCLS\left(\widehat{\boldsymbol{A}}^T, \widehat{\boldsymbol{B}}^{(t)^T}\right). \qquad (S13)$$

により単純に行える。そこでＣ（ｔ）に基づく

[数177]

$\widehat{\boldsymbol{B}}^{(t)}$

から

[数178]

$$\widehat{\boldsymbol{B}}^{(t+1)}$$

へのアップデートについて考える。なお、

[数179]

$$nonorth(\widehat{\boldsymbol{B}}) = Tr\left(\boldsymbol{\Lambda}\left(\widehat{\boldsymbol{B}}\widehat{\boldsymbol{B}}^{T} - diag(\widehat{\boldsymbol{B}}\widehat{\boldsymbol{B}}^{T})\right)\right),$$

で、

[数180]

$$\boldsymbol{\Lambda} \in \mathbb{R}^{K\times K}$$

は未定乗数である。また

[数181]

$$vol(\widehat{\boldsymbol{B}}) = \log\left|\det\left(\widehat{\boldsymbol{B}}\widehat{\boldsymbol{B}}^{T} + \tau\boldsymbol{I}_{K}\right)\right| = \log|\det(\boldsymbol{H})|,$$

ここで、

[数182]

$$\boldsymbol{H}(\widehat{\boldsymbol{B}}) = \widehat{\boldsymbol{B}}\widehat{\boldsymbol{B}}^{T} + \tau\boldsymbol{I}_{K}, \tau$$

は小さな正則化パラメータで、一貫して$\tau = 10^{-8}$を用いた。ここで、

[数183]

$$vol(\widehat{\boldsymbol{B}})$$

のＭａｊｏｒｉｚｅｒ関数を導入する。接線不等式より、一期前の

[数184]

$$\boldsymbol{H}^{(t)} = \boldsymbol{H}(\widehat{\boldsymbol{B}}^{(t)})$$

に基づいて、

[数185]

$$\log|\det(\boldsymbol{H})| \leq \log\left|\det\left(\boldsymbol{H}^{(t)}\right)\right| + Tr\left[\left(\nabla_{\boldsymbol{H}^{(t)}}\log|\det(\boldsymbol{H})|\right)^{T}\left(\boldsymbol{H} - \boldsymbol{H}^{(t)}\right)\right],$$

が成立する。ここで

[数186]

$$\nabla_{H^{(t)}} \log|\det(H)|$$

は

[数187]

$$\log|\det(H)|$$

のＨ（ｔ）における勾配で、

[数188]

$$\nabla_{H^{(t)}} \log|\det(H)| = {H^{(t)}}^{-T}$$

より、

[数189]

$$vol(\hat{B}) = \log|\det(H)| \leq Tr\left[{H^{(t)}}^{-1} H\right] + const = Tr\left[F^{(t)} \hat{B} \hat{B}^{T}\right] + const,$$

ここで、

[数190]

$$F^{(t)} = {H^{(t)}}^{-1}$$

かつ、ｃｏｎｓｔは定数項である。したがって、

[数191]

$$vol(\hat{B})$$

を

[数192]

$$Tr\left[F^{(t)} \hat{B} \hat{B}^{T}\right]$$

で置き換えて、すべてのペナルティー項を合わせて書くと、

[数193]

$$\frac{\alpha}{2}vol(\widehat{\boldsymbol{B}})+\frac{\beta}{2}nonorth(\widehat{\boldsymbol{B}})\leq\frac{\alpha}{2}Tr\left(\boldsymbol{F}^{(t)}\widehat{\boldsymbol{B}}\widehat{\boldsymbol{B}}^{T}\right)+\frac{\beta}{2}Tr\left(\boldsymbol{\Lambda}^{(t)}\widehat{\boldsymbol{B}}\widehat{\boldsymbol{B}}^{T}\right)+const$$

$$=\frac{1}{2}Tr\left(\boldsymbol{V}\widehat{\boldsymbol{B}}\widehat{\boldsymbol{B}}^{T}\right)+const,$$

ここで

[数194]

$$\boldsymbol{V}=\alpha\boldsymbol{F}^{(t)}+\beta\boldsymbol{\Lambda}^{(t)}.$$

従って

[数195]

$$\widehat{\boldsymbol{B}}^{(t+1)}$$

のアップデートは以下を解くことで得られる。

[数196]

$$\widehat{\boldsymbol{B}}^{(t+1)}=\arg\min_{\widehat{\boldsymbol{B}}}\frac{1}{2}Tr\left[\boldsymbol{W}\left(\widehat{\boldsymbol{A}}-\boldsymbol{C}^{(t)}\widehat{\boldsymbol{B}}\right)\left(\widehat{\boldsymbol{A}}-\boldsymbol{C}^{(t)}\widehat{\boldsymbol{B}}\right)^{T}\right]+\frac{1}{2}Tr\left(\boldsymbol{V}\widehat{\boldsymbol{B}}\widehat{\boldsymbol{B}}^{T}\right)\qquad(S14)$$

$$s.t.\,\widehat{\boldsymbol{B}}=\boldsymbol{B}\boldsymbol{L},\boldsymbol{B}\geq 0.$$

複雑な制約条件を整理するために、

[数197]

$$\widehat{\boldsymbol{B}}=\boldsymbol{B}\boldsymbol{L}$$

なる制約をラグランジェ未定乗数

[数198]

$$\boldsymbol{Z}\in\mathbb{R}^{K\times M}$$

を用いて、目的関数に組み込み、ＡＤＭＭの枠組みでこの問題を解く。

[数199]

$$\left(\boldsymbol{B}^{(t+1)},\widehat{\boldsymbol{B}}^{(t+1)}\right)=\arg\min_{\boldsymbol{B},\widehat{\boldsymbol{B}}}\max_{\boldsymbol{Z}}\left\{f(\widehat{\boldsymbol{B}})+Tr\left[\boldsymbol{Z}^{T}\left(\widehat{\boldsymbol{B}}-\boldsymbol{B}\boldsymbol{L}\right)\right]-\frac{1}{2\mu}\|\boldsymbol{Z}-\boldsymbol{Z}'\|_{F}^{2}\right\},$$

$$s.t.\,\boldsymbol{B}\geq 0,$$

ここで、

[数200]

$$f(\widehat{\boldsymbol{B}}) \equiv \frac{1}{2}Tr\left[\boldsymbol{W}\left(\widehat{\boldsymbol{A}}-\boldsymbol{C}^{(t)}\widehat{\boldsymbol{B}}\right)\left(\widehat{\boldsymbol{A}}-\boldsymbol{C}^{(t)}\widehat{\boldsymbol{B}}\right)^{T}\right]+\frac{1}{2}Tr\left(\boldsymbol{V}\widehat{\boldsymbol{B}}\widehat{\boldsymbol{B}}^{T}\right)$$

かつμはＡＤＭＭのハイパーパラメータである（ここでは一貫してμ＝１とした）。Ｚ′は一期前のＺを表し、

[数201]

$$\boldsymbol{Z}=\boldsymbol{Z}'+\mu\left(\widehat{\boldsymbol{B}}-\boldsymbol{B}\boldsymbol{L}\right), \qquad (S15)$$

のときに目的関数はＺについて最大化された、

[数202]

$$g\left(\boldsymbol{B},\widehat{\boldsymbol{B}};\boldsymbol{Z}\right)\equiv f\left(\widehat{\boldsymbol{B}}\right)+\frac{\mu}{2}\left\|\widehat{\boldsymbol{B}}-\boldsymbol{B}\boldsymbol{L}+\frac{1}{\mu}\boldsymbol{Z}\right\|_{F}^{2},\boldsymbol{B}\geq 0,$$

を与える。これを

[数203]

$$\left(\boldsymbol{B},\widehat{\boldsymbol{B}}\right)$$

について最適化すればよい。

[数204]

$$\left(\boldsymbol{B},\widehat{\boldsymbol{B}},\boldsymbol{Z}\right)$$

をサイクリックに最適化し、アルゴリズムをAlgorithm3にまとめた。

[数205]

Algorithm3: ADMM for solving the optimization of Eq. S14

input: $\boldsymbol{B}^{(t)}, \widehat{\boldsymbol{B}}^{(t)}$, hyperparameter μ, function $g\left(\boldsymbol{B}, \widehat{\boldsymbol{B}}\right)$

output: $\boldsymbol{B}^{(t+1)}, \widehat{\boldsymbol{B}}^{(t+1)}$

initialize: $q = 0, \boldsymbol{B}_q \leftarrow \boldsymbol{B}^{(t)}, \widehat{\boldsymbol{B}}_q \leftarrow \widehat{\boldsymbol{B}}^{(t)}, \boldsymbol{Z}_q \leftarrow \boldsymbol{0}$

repeat until convergence:

$\widehat{\boldsymbol{B}}_{q+1} = \underset{\widehat{\boldsymbol{B}}}{\operatorname{argmin}}\, g\left(\widehat{\boldsymbol{B}}; \boldsymbol{B}_q, \boldsymbol{Z}_q\right)$, solved by Eq. S16

$\boldsymbol{B}_{q+1} = \underset{\boldsymbol{B} \geq \boldsymbol{0}}{\operatorname{argmin}}\, g\left(\boldsymbol{B}; \widehat{\boldsymbol{B}}_{q+1}, \boldsymbol{Z}_q\right)$, solved by Eq. S17

$\boldsymbol{Z}_{q+1} = \boldsymbol{Z}_q + \mu\left(\widehat{\boldsymbol{B}}_{q+1} - \boldsymbol{B}_{q+1}\boldsymbol{L}\right)$

$q \leftarrow q + 1$

if $\left\|\widehat{\boldsymbol{B}}_q - \boldsymbol{B}_q\boldsymbol{L}\right\|_F^2 < 10^{-6}$:

end

$\boldsymbol{B}^{(t+1)} \leftarrow \boldsymbol{B}_q, \widehat{\boldsymbol{B}}^{(t+1)} \leftarrow \widehat{\boldsymbol{B}}_q$.

[0228] 目的関数

[数206]

$$g\left(\widehat{\boldsymbol{B}}; \boldsymbol{B}_q, \boldsymbol{Z}_q\right)$$

は

[数207]

$$\widehat{\boldsymbol{B}}$$

については拘束条件なしの二次関数の最適化なので、

[数208]

$$\frac{\partial g}{\partial \widehat{\boldsymbol{B}}} = \left(\boldsymbol{C}^{(t)^T}\boldsymbol{W}\boldsymbol{C}^{(t+1)} + \boldsymbol{V} + \mu\boldsymbol{I}_K\right)\widehat{\boldsymbol{B}} - \boldsymbol{C}^{(t)^T}\boldsymbol{W}\widehat{\boldsymbol{A}} - \boldsymbol{B}_q\boldsymbol{L} + \frac{1}{\mu}\boldsymbol{Z} \equiv \boldsymbol{0}$$

$$\widehat{\boldsymbol{B}}_{q+1} = \left(\boldsymbol{C}^{(t)^T}\boldsymbol{W}\boldsymbol{C}^{(t)} + \boldsymbol{V} + \mu\boldsymbol{I}_K\right)^{-1}\left(\boldsymbol{C}^{(t)^T}\boldsymbol{W}\widehat{\boldsymbol{A}} + \boldsymbol{B}_q\boldsymbol{L} - \frac{1}{\mu}\boldsymbol{Z}\right), \qquad (S16)$$

より最小化される。また

[数209]

$$g\left(\boldsymbol{B}; \widehat{\boldsymbol{B}}_{q+1}, \boldsymbol{Z}_q\right)$$

はＮＮＬＳにより解けて、

[数210]

$$\boldsymbol{B}_{q+1}{}^T = NNLS\left(\left(\widehat{\boldsymbol{B}}_{q+1} + \frac{1}{\mu}\boldsymbol{Z}\right)^T, \boldsymbol{L}^T\right), \qquad (S17)$$

と更新される。以上より元の問題Ｅｑ．Ｓ１１を解くための

[数211]

$$\left(\widehat{\boldsymbol{B}}, \boldsymbol{B}, \boldsymbol{C}\right)$$

の更新式が得られた。最後にＶ＝αＦ（ｔ）＋βΛ（ｔ）の更新について考える。Ｆ（ｔ）については

[数212]

$$\boldsymbol{F}^{(t)} = \left(\widehat{\boldsymbol{B}}^{(t)}\widehat{\boldsymbol{B}}^{(t)^T} + \tau\boldsymbol{I}_K\right)^{-1}, \qquad (S18)$$

と簡単に求まる。Λ（ｔ）は厳密な直交条件、すなわち

[数213]

$$\widehat{\boldsymbol{B}}\widehat{\boldsymbol{B}}^T = diag\left(\widehat{\boldsymbol{B}}\widehat{\boldsymbol{B}}^T\right) \equiv \boldsymbol{D},\ \beta = 1$$

と

[数214]

$$\frac{\partial f\left(\widehat{\boldsymbol{B}};\boldsymbol{C}^{(t)}\right)}{\partial \widehat{\boldsymbol{B}}} = \boldsymbol{C}^{(t)^T}\boldsymbol{W}\left(\boldsymbol{C}^{(t)}\widehat{\boldsymbol{B}} - \widehat{\boldsymbol{A}}\right) + \left(\alpha\boldsymbol{F}^{(t)} + \boldsymbol{\Lambda}^{(t)}\right)\widehat{\boldsymbol{B}} \equiv 0. \qquad (S19)$$

を組み合わせて、

[数215]

$$\widehat{\boldsymbol{B}}^T$$

をＥｑ．Ｓ１９の右側からかけると、

[数216]

$$\boldsymbol{C}^{(t)^T}\boldsymbol{W}\left(\boldsymbol{C}^{(t)}\boldsymbol{D} - \widehat{\boldsymbol{A}}\widehat{\boldsymbol{B}}^T\right) + \left(\alpha\boldsymbol{F}^{(t)} + \boldsymbol{\Lambda}^{(t)}\right)\boldsymbol{D} \equiv 0,$$
$$\boldsymbol{\Lambda}^{(t)} = \boldsymbol{C}^{(t)^T}\boldsymbol{W}\left(\widehat{\boldsymbol{A}}\widehat{\boldsymbol{B}}^T\boldsymbol{D}^{-1} - \boldsymbol{C}^{(t)}\right) - \alpha\boldsymbol{F}^{(t)}. \qquad (S20)$$

と得られる。Ｅｑ．Ｓ１８とＥｑ．Ｓ２０より、

[数217]

$$\boldsymbol{V} = \alpha(1-\beta)\left(\widehat{\boldsymbol{B}}^{(t)}\widehat{\boldsymbol{B}}^{(t)^T} + \tau\boldsymbol{I}_K\right)^{-1} + \beta\boldsymbol{C}^{(t)^T}\boldsymbol{W}\left(\widehat{\boldsymbol{A}}\widehat{\boldsymbol{B}}^T\boldsymbol{D}^{-1} - \boldsymbol{C}^{(t)}\right). \qquad (S21)$$

が与えられる。以上を用いて、問題Ｅｑ．Ｓ１１を解くためのアルゴリズムは以下のとおりである。

[数218]

Algorithm 4: Pseudo-code for the second NMF (solving optimization Eq. S11)

Input: output of the first NMF (sample-wise FA; $\widetilde{\boldsymbol{A}} \in \mathbb{R}_+^{N\times M}$, basis-fragment spectra; $\boldsymbol{S} \in \mathbb{R}_+^{M\times D}$), basis-polymer number; K, weights for penalty terms; (α, β), weight for outliers; p

Output: polymer weight-fraction; $\boldsymbol{C} \in \mathbb{R}_+^{N\times K}$, FAs of unit-weight basis-polymers $\boldsymbol{B} \in \mathbb{R}_+^{K\times M}$

Initialization

calculate $\boldsymbol{L}$ via Cholesky decomposition of $\boldsymbol{SS}^T$

set $\widehat{\boldsymbol{A}} = \widetilde{\boldsymbol{A}}\boldsymbol{L}$

initialize $\widehat{\boldsymbol{B}}^{(0)}$ by selecting K-rows of $\widehat{\boldsymbol{A}}$ via VCA algorithm

set $\boldsymbol{B}^{(0)} = \widehat{\boldsymbol{B}}^{(0)}\boldsymbol{L}^{-1}$

initialize $\boldsymbol{C}^{(0)}$ by $\boldsymbol{C}^{(0)^T} = FCLS(\widehat{\boldsymbol{A}}^T, \widehat{\boldsymbol{B}}^{(0)^T})$

initialize $\boldsymbol{W}$ by Eq. S12

initialize $\boldsymbol{V}$ based on $\boldsymbol{B}^{(0)}$ and $\boldsymbol{C}^{(0)}$ by Eq. S21

Repeat until criteria for convergence satisfied:

update $\widehat{\boldsymbol{B}}, \boldsymbol{B}$ by algorithm 3

update $\boldsymbol{C}$ by Eq. S13

update $\boldsymbol{W}, \boldsymbol{V}$ by Eq. S12 and S21, respectively

[0229] （ＳとＢの張る超平面へのテストデータの射影）

データセットからＳとＢを推定したのちに、ＳとＢの推定には用いなかったデータ（ここではテストデータと呼ぶ）

[数219]

$$\boldsymbol{X}_{test} \in \mathbb{R}_+^{T\times D}$$

を、ＳとＢの張る超平面へ射影する方法について述べる。

まずＳ－超平面への射影により、

[数220]

$$\boldsymbol{A}_{test} \in \mathbb{R}_+^{T\times M}$$

が得られる。すなわち、

[数221]

$$\boldsymbol{A}_{test}{}^{T} = NNLS\left(\boldsymbol{R}^{-\frac{1}{2}}\boldsymbol{X}_{test}{}^{T}, \boldsymbol{R}^{-\frac{1}{2}}\boldsymbol{S}^{T}\right),$$

温度帯域について総和をとり、

[数222]

$$\widetilde{\boldsymbol{A}}_{test} \in \mathbb{R}_{+}{}^{1\times M}$$

に変換したのちに、Ｂ－超平面への射影と規格化により、

[数223]

$$\boldsymbol{C}_{test} \in \mathbb{R}_{+}{}^{1\times K}$$

が得られる。すなわち、

[数224]

$$\boldsymbol{C}_{test}{}^{T} = NNLS\left(\boldsymbol{L}^{T}\widetilde{\boldsymbol{A}}_{test}{}^{T}, \boldsymbol{L}^{T}\boldsymbol{B}^{T}\right),$$

$$\boldsymbol{C}_{test} \leftarrow \frac{\boldsymbol{C}_{test}}{\|\boldsymbol{C}_{test}\|_{1}}.$$

[0230] なお、データセットとハイパーパラメータは以下のとおりである。

[0231] [表2]

Summary of dataset and hyperparameters.

Sample number N	Temp. range (°C) Mass range(m/z)	First NMF					Second NMF		
		w_o	Merging thres.	Initial M	Prior	iteration	K	$\alpha = \beta$	p
24	[50, 450] [50, 1000]	0.2	0.99	30	Exp	3000	3	0.1	1.5

[0232] また、引用文献は、以下のとおりである。

[表3]

1. S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3**, 1–122 (2010).
2. X. Fu, W. Ma, K. Huang, N. D. Sidiropoulos, ROBUST VOLUME MINIMIZATION-BASED MATRIX FACTORIZATION VIA ALTERNATING OPTIMIZATION Department of ECE , University of Minnesota , Minneapolis , MN , USA Department of EE , The Chinese University of Hong Kong , Shatin , N . T ., Hong Kong. *Icassp 2016.* **MI**, 2534–2538 (2016).
3. X. Fu, K. Huang, B. Yang, W. K. Ma, N. D. Sidiropoulos, Robust Volume Minimization-Based Matrix Factorization for Remote Sensing and Document Clustering. *IEEE Trans. Signal Process.* **64**, 6254–6268 (2016).
4. J. M. B. Dias, J. M. P. Nascimento, Vertex component analysis: A geometric-based approach to unmix hyperspectral data. *Signal Image Process. Remote Sens.* **43**, 415–439 (2006).

[0233] （組成推定装置）

次に、本発明の実施形態に係る組成推定装置について、図面を参照しながら説明する。図３は、本発明の実施形態に係る組成推定装置のハードウェア構成図である。

[0234] 組成推定装置３０は、質量分析装置３１と、情報処理装置３３とを備え、質量分析装置３１は、減圧ポンプ３２を備えている。情報処理装置３３は、プロセッサ３４と、記憶デバイス３５と、表示デバイス３６と、入力デバイス３７とを有し、質量分析装置３１と情報処理装置３３とは相互にデータを送受信できるよう構成されている。

[0235] プロセッサ３４は、例えば、マイクロプロセッサ、プロセッサコア、マルチプロセッサ、ＡＳＩＣ（ａｐｐｌｉｃａｔｉｏｎ－ｓｐｅｃｉｆｉｃ ｉｎｔｅｇｒａｔｅｄ ｃｉｒｃｕｉｔ）、ＦＰＧＡ（ｆｉｅｌｄ ｐｒｏｇｒａｍｍａｂｌｅ ｇａｔｅ ａｒｒａｙ）、及び、ＧＰＧＰＵ（Ｇｅｎｅｒａｌ－ｐｕｒｐｏｓｅ ｃｏｍｐｕｔｉｎｇ ｏｎ ｇｒａｐｈｉｃｓ ｐｒｏｃｅｓｓｉｎｇ ｕｎｉｔｓ）等である。

[0236] 記憶デバイス３５は、各種プログラム、及び、データを一時的に、及び／又は、非一時的に記憶する機能を有し、プロセッサ３４の作業エリアを提供する。

記憶デバイス３５は、例えば、ＲＯＭ（Ｒｅａｄ　Ｏｎｌｙ　Ｍｅｍｏｒｙ）、ＲＡＭ（Ｒａｎｄｏｍ　Ａｃｃｅｓｓ　Ｍｅｍｏｒｙ）、ＨＤＤ（Ｈａｒｄ　Ｄｉｓｋ　Ｄｒｉｖｅ）、フラッシュメモリ、及び、ＳＳＤ（Ｓｏｌｉｄ　Ｓｔａｔｅ　Ｄｒｉｖｅ）等である。

[0237] 入力デバイス３７は、各種情報入力を受け付け、また、組成推定装置３０への指示の入力を受け付けることができる。入力デバイス３７は、キーボード、マウス、スキャナ、及び、タッチパネル等でよい。

[0238] 表示デバイス３６は、組成推定装置３０のステータス、分析の進捗、及び、組成推定結果等を表示できる。表示デバイス３６は、液晶ディスプレイ、及び、有機ＥＬ（Ｅｌｅｃｔｒｏ　Ｌｕｍｉｎｅｓｃｅｎｃｅ）ディスプレイ等でよい。

また、表示デバイス３６は、入力デバイス３７と一体として構成されていてもよい。この場合、表示デバイス３６がタッチパネルディスプレイであって、ＧＵＩ（Ｇｒａｐｈｉｃａｌ　Ｕｓｅｒ　Ｉｎｔｅｒｆａｃｅ）を提供する形態が挙げられる。

[0239] データバスにより相互にデータを通信可能なプロセッサ３４、記憶デバイス３５、入力デバイス３７、及び、表示デバイス３６を備える、情報処理装置３３は、典型的にはコンピュータである。

[0240] 質量分析装置３１は、減圧ポンプ３２を備え、典型的には、ＤＡＲＴイオン源、検体加熱装置、及び、飛行時間型質量分析計を備える質量分析装置である。質量分析装置３１が備えるイオン源、及び、質量分析計は、いずれも非限定的な例であり、本組成推定装置が備える質量分析装置の構成は上記に限定されない。

[0241] 図４は、本発明の実施形態に係る組成推定装置３０の機能ブロック図である。組成推定装置３０は、全検体の質量分析を行う質量分析部４１と、質量分析部４１によって得られたマススペクトルを処理する情報処理部４２とを有する。

[0242] 質量分析部４１は、質量分析装置３１を含んで構成され、質量分析装置３

１が後述する制御部４３に制御されて実現する機能である。

質量分析部４１は、ロードされた検体セット（検体、学習用検体、及び、バックグラウンド検体、図４中、符号ａ）について、加熱しながら熱脱着、及び／又は、熱分解されて生ずるガス成分をイオン化し、順次質量分析して、マススペクトルを出力する（図４中、符号ｃ）。

この際、質量分析装置３１の設置されている環境に起因した環境要因（図４中、符号ｂ、例えば、減圧ポンプ３２から生ずる（ポリ）ジメチルシロキサン）も同時に質量分析装置３１にロードされ、マススペクトルに影響を与える。

[0243] 次に、情報処理装置３３を含んで構成される情報処理部４２の各部の機能について説明する。

制御部４３は、プロセッサ３４を含んで構成される。制御部４３は、各部を制御して、組成推定装置３０の各機能を実現する。

[0244] 記憶部４４は、記憶デバイス３５を含んで構成される。記憶部４４には予め各部の制御のためのプログラムが記憶され、また、各部によって入出力されるデータの一時的、又は、非一時的な記憶領域を提供する。記憶部４４は、制御部４３によって制御され、データを読み出したり、書き込んだりすることができる。

[0245] 入力部４５は入力デバイス３７を含んで構成される。また、出力部４６は、表示デバイス３６を含んで構成される。制御部４３がこれらを制御することで、組成推定装置３０の使用者からの入力を受け付けて、これを記憶部４４に記憶させたり、記憶部４４に記憶された情報を各部に送信したり、出力部４６を介して組成推定装置３０の使用者に対して表示したりできる。

[0246] データ行列作成部４７は、記憶部４４に記憶されたプログラムが制御部４３により実行されて実現される機能である。

データ行列作成部４７は、質量分析部４１によって加熱温度ごとに取得されたマススペクトル（図４中、符号ｃ）を各行に格納した二次元マススペクトルからデータ行列Ｘを作成し、後述するＮＭＦ処理部４８に渡す（図４中

、符号ｄ）。なお、データ行列Ｘの詳細については、すでに説明したとおりである。

[0247] ＮＭＦ処理部４８は、記憶部４４に記憶されたプログラムが制御部４３により実行されて実現される機能である。

ＮＭＦ処理部４８は、データ行列作成部４７によって作成されたデータ行列Ｘ（図４中、符号ｄ）を非負値行列因子分解（ＮＭＦ）し、規格化された基底スペクトル行列（Ｓ）とその強度分布行列（Ｃ）との積（$X=C\times S^{T}$）に分解する。その結果は、補正処理部４９に渡される（図４中、符号ｅ）。なお、ＮＭＦ処理の詳細はすでに説明したとおりである。

[0248] 補正処理部４９は、記憶部４４に記憶されたプログラムが制御部４３により実行されて実現される機能である。

補正処理部４９は、ＮＭＦ処理部４８によって作成された基底スペクトル行列とデータ行列との正準相関分析によって強度分布行列を補正して、補正後強度分布行列を作成する機能である。作成された補正後強度分布行列はベクトル処理部５０に渡される（図４中、符号ｆ）。

[0249] なお、補正処理部４９は、上記の正準相関分析による補正に加えて、すでに説明した強度補正を更に行ってもよい。この際、強度補正に用いるキャリブレーション検体の組成（ｃａｌｉｂ検体組成）がある場合は、入力部４５を介して補正処理部４９に渡される（図４中、符号ｋ）。正準相関分析による補正（ＣＣＡ　ｆｉｌｔｅｒ）、及び、強度補正の詳細はすでに説明したとおりである。

[0250] ベクトル処理部５０は、記憶部４４に記憶されたプログラムが制御部４３により実行されて実現される機能である。

ベクトル処理部５０は、補正処理部４９により作成された補正後強度分布行列を全検体のそれぞれに対応する小行列に分割し、これを特徴ベクトルとしてベクトル空間内で全検体を表現する。この結果は、エンドメンバー決定部５１に渡される（図４中、符号ｇ）。

[0251] エンドメンバー決定部５１は、記憶部４４に記憶されたプログラムが制御

部４３により実行されて実現される機能である。

エンドメンバー決定部５１は、ベクトル処理部５０によって作成された特徴ベクトルの全てを内包するＫ－１次元単体を設定し、Ｋ－１次元単体におけるＫ個のエンドメンバーを決定する。この結果は、含有量比計算部５２に渡される（図４中、符号ｈ）。

[0252] この際、学習用検体にエンドメンバーが含まれる場合であって、判別ラベルを有する場合には、入力部４５から判別ラベルの情報がエンドメンバー決定部５１に渡される（図４中、符号ｊ）。

しかし、すでに説明したとおり、エンドメンバーの決定（エンドメンバー決定部５１の処理）には判別ラベルは必須ではない。

[0253] 含有量比計算部５２は、記憶部４４に記憶されたプログラムが制御部４３により実行されて実現される機能である。

含有量比計算部５２は、エンドメンバー決定部５１によって決定されたＫ個のエンドメンバーと、ベクトル処理部５０により作成された推定対象検体の特徴ベクトルとのユークリッド距離をそれぞれ計算し、Ｋ個のエンドメンバーと検体の特徴ベクトルとのユークリッド距離の比から、推定対象検体中のＫ個の成分の含有量比を推定する。この含有量比の推定結果（図４中、符号ｉ）は、出力部４６から出力される。

なお、この際、学習用検体の組成推定結果もあわせて出力してもよい。

[0254] 本組成推定装置よれば、溶媒に溶解しにくく、従来の組成分析が行いにくい固体の検体であっても、特段の前処理なしに分析に供することができ、かつ、正確な定量分析結果が得られる。本組成推定装置は、特に、複数の成分の混合物である検体の品質管理、及び、不良原因の調査等に適用した場合、結果を得るために必要な時間を大幅に短縮することができる。

実施例

[0255] 以下では、実施例をもとに、本組成分析方法について詳述する。なお、本発明の組成分析方法は下記の特定の実施例によって限定的に解釈されるものでない。

[0256] ［実施例１］

（検体の準備）

ポリメチルメタクリレート（ｐｏｌｙ（ｍｅｔｈｙｌ　ｍｅｔｈａｃｒｙｌａｔｅ）、記号「M」）、ポリスチレン（ｐｏｌｙｓｔｙｒｅｎｅ、記号「Ｓ」）、ポリエチルメタクリレート（ｐｏｌｙ（ｅｔｈｙｌ　ｍｅｔｈａｃｒｙｌａｔｅ）、記号「Ｅ」）の三種類の高分子のジオキサン溶液（約４質量%、高分子の固形分として０．４ｍｇ程度）を、サンプルポット上にキャストした後に一晩風乾して検体を得た。表４は、各検体の組成である。なお、質量分析は、ＳＨＩＭＡＤＺＵ社製ＬＣＭＳ－２０２０を検出器として、ＩｏｎＳｅｎｓｅ社製ＤＡＲＴイオン源として、バイオクロマト社製イオンロケットをサンプルの加熱部として用いて行った。

[0257]

[表4]

	Composition		
SampleName	M	S	E
M	1.000		
S		1.000	
E			1.000
EM	0.475		0.525
MS	0.510	0.490	
SE		0.502	0.498
MS01	0.884	0.116	
MS005	0.958	0.042	
MS001	0.990	0.010	
MS0001	0.991	0.009	
M1_S2_E2	0.201	0.397	0.402
M2_S1_E2	0.360	0.228	0.413
M2_S2_E1	0.408	0.394	0.197
MSE_calib	0.337	0.326	0.337

[0258] 得られたマススペクトルについて、以下の（ｉ）～（ｉｖ）を実装したアルゴリズムによって解析を行った。

[0259] （ｉ）Ｃ行列ＡＲＤとＳ行列へのソフト直交制約

（ｉｉ）ＮＭＦ最適化中のマージング条件

（ｉｉｉ）ＣＣＡ－ｆｉｌｔｅｒによるＮＭＦアーティファクト、及び、バックグラウンド、コンタミネーション除去

（ｉｖ）環境変数を補正項にいれることで、キャリブレーション不要の定量分析

[0260] 解析に使用したハイパーパラメータは以下のとおりである。

ＮＭＦ：

ｏｒｔｈｏｇｏｎａｌ ｃｏｎｓｔｒａｉｎ ｗｅｉｇｈｔ ｗ＝０．２，ｉｎｉｔｉａｌ ｃｏｍｐｏｎｅｎｔ ｎｕｍｂｅｒ Ｋ＝５０，ｓｍａｌｌ ｐｏｓｉｔｉｖｅ ｎｕｍｂｅｒ ａ＝１＋０．１$^{-15}$，ｉｔｅｒａｔｉｏｎ ｎｕｍｂｅｒ Ｌ＝２０００，ｔｈｒｅｓｈｏｌｄ＝０．９

ＣＣＡ－Ｆｉｌｔｅｒ：

t_1＝０．８，t_2＝０．２，

[0261] 図５は、「ＭＳＥ＿ｃａｌｉｂ」についてのオリジナルなスペクトル（Ａ）と、上記処理を行った後、再構築したスペクトル（Ｂ）を表す図である。両者の波形は類似していることがわかる。すなわち、上記処理によってオリジナルなスペクトルの特徴は保存され、組成分析に用いやすいデータとなったことが示された。

[0262] また、オリジナルデータにはエアコン等から排出されるフタル酸エステル等のバックグラウンドが３９０ｍ／ｚに観測される（図５（Ａ）中、「ＢＧ」と表示した部分）が、再構築されたピーク（Ｂ）では、これが除去されている。また、検体の作成に用いられた溶媒であるジオキサン（図５（Ａ）中、「ＤＯＸ」と表示した部分）も除去されていた。これらはいずれもノイズに該当するものであり、本方法によって、効果的にノイズが除去されたことが示された。

[0263] また、図６は、ＮＭＦ分解後のＣ行列の一部$C_{\delta\,(MSE_Calib)}$と、ＣＣＡ－Ｆｉｌｔｅｒによってバックグラウンド及び不純物のスペクトルを除いた後のS^Tである。

[0264] 図７は、Ｍ、Ｓ、Ｅ三成分系の組成解析結果である。図７中、中空円で表され、凡例に「Ｃａｌｉｂｒａｔｅｄ」とあるのは、ＭＳＥ＿ｃａｌｉｂをキャリブレーション検体とした場合、つまり、ＭＳＥ＿ｃａｌｉｂの組成情報を既知として与え、強度分布行列を強度補正した場合の検体の組成推定結果である。

[0265] また、図７中、中実円で表され、凡例に「Ｎｏｎ－ｃａｌｉｂｒａｔｅｄ」とあるのは、環境変数（すなわち、６１０ｍ／ｚ：ポリジメチルシロキサンのピーク強度）と検体質量の積を用いて、強度分布行列を強度補正した場合の検体の組成推定結果である。

[0266] また、図７中、星印で表され、凡例に「Ｋｎｏｗｎ」とあるのは、実際の組成（真の値）を表している。

[0267] 図７に示した結果から、「Ｃａｌｉｂｒａｔｅｄ」、「Ｎｏｎ－ｃａｌｉｂｒａｔｅｄ」のいずれの場合にも真の値に近い、優れた推定結果が得られた。

[0268] 次に、図８は、ＮＭＦの条件から、（１）の直交制約を解除した場合の組成推定結果である。各プロットの定義は、図７と同様である。

図８の結果から、ＮＭＦが直交制約を有する方が、より優れた推定結果が得られることがわかった。

[0269] 次に、図９は、ＭＮＦ最適化の条件から、ｍｅｒｇｉｎｇ　Ｓのみ解除した場合の組成推定結果である。各プロットの定義は、図７と同様である。

図９の結果から、ＭＮＦ最適化の条件がｍｅｒｇｉｎｇ　Ｓを有する方が、より優れた推定結果が得られることがわかった。

なお、成分数はＡＲＤによって３６までしか減らせないが、Ｍｅｒｇｉｎｇ　Ｓの条件を維持したときは成分数９でより精度よく組成が推定できることから、Ｍｅｒｇｉｎｇ　Ｓの重要性がわかった。

[0270] 図１０は、ＣＣＡ－ｆｉｌｔｅｒを解除した場合の結果（比較例）の組成推定結果である。各プロットの定義は図７と同様である。

図１０においては、検体「ＳＥ」は、Ｍの含有量は０％にも関わらず１％程度入っているという計算結果となった。０％を１％と推定してしまうと、数％の推定精度もないということになり実用的ではない。

[0271] 図１１は、「ＭＳＥ＿ｃａｌｉｂ」についてのオリジナルなスペクトル（上、ｒａｗ　ｏｆ　ＭＳＥ＿ｃａｌｉｂ）と、ＣＣＡ－ｆｉｌｔｅｒを解除した上記で処理を行った後、再構築したスペクトル（下、ｒｅｃｏｎｓｔｒ

ｕｃｔｅｄ　ｏｆ　ＭＳＥ＿ｃａｌｉｂ）を表す図である。

この結果から、ＣＣＡ－ｆｉｌｔｅｒを解除すると、ＤＯＸ、ＢＧ（図５の説明を参照）のいずれのノイズも除去できていないことがわかった。

[0272] 図１２は、環境変数による補正を行わなかった場合の組成推定結果である。図１２の結果においては、中実円で表された「Ｎｏｎ－ｃａｌｉｂｒａｔｅｄ」がＫｎｏｗｎからのずれが大きく成っており、環境変数による補正により、得られる推定結果がより正確となることがわかった。

[0273] ［実施例２］

（エンドメンバーを含まない検体による推定）

学習用検体を含む検体（群）が、エンドメンバーを含まない場合にも精度よく推定できることを確かめるために、表５に記載したサンプルを調製して実施例１と同様の試験を行った。

なお、表中、記号「Ｍ」「Ｓ」「Ｅ」が意味するものは、表４（実施例１）と同様である。また、各サンプルは、単独の成分（ポリマー）の重量分率が８０％を超えないように調整されている。

[0274]

[表5]

sample	M(mass%)	S(mass%)	E(mass%)	mass(mg)
1	0	50	50	0.42
2	0	62	38	0.60
3	0	31	69	0.44
4	0	79	21	0.35
5	0	32	68	0.96
6	0	20	80	0.87
7	14	72	14	0.44
8	15	15	70	0.68
9	19	0	81	0.39
10	20	40	40	0.44
11	25	75	0	0.35
12	29	0	71	0.42
13	29	71	0	0.72
14	34	33	34	0.39
16	36	23	41	0.47
17	41	39	20	0.44
18	48	0	52	0.43
19	49	51	0	0.32
20	64	36	0	0.43
21	67	0	33	0.36
22	70	30	0	0.98
23	79	9	12	0.46
24	80	0	20	0.27
25	83	17	0	0.34

[0275] 上記検体について、実施例１と同様の方法で二次元マススペクトルを得て、データ行列に変換した。得られたデータ行列について、すでに説明した１回目のＮＭＦ処理、すなわち、

[数225]

$$\boldsymbol{X} \approx \boldsymbol{AS}$$

を実施した。なお、ノイズの分散共分散行列の見積もりは、Ｘのみを用いて計算できるので、１回目のＮＭＦ処理におけるモジュール内部処理とした。

[0276] 次に、Ａに対するＣＣＡ－Ｆｉｌｔｅｒ、重さや標準ピークによる強度補正を行った。なお、組成分析には温度分布情報は不要なため、温度軸に沿ってサンプル毎に総和を取って、

[数226]

$$\boldsymbol{A} \in {\mathbb{R}_+}^{NT \times M} \longrightarrow \widetilde{\boldsymbol{A}} \in {\mathbb{R}_+}^{N \times M}$$

としてもよい。

[0277] 次に、

[数227]

$$\widetilde{\boldsymbol{A}} \approx \boldsymbol{CB}$$

の２回目のＮＭＦ処理を実施した。なお、リーマン計量による評価等は、２回目のＮＭＦ処理の内部処理として実施した。図１５はその結果として、ＣとＢＳとを図示したものである。

図中Ｃでは、本方法によって、推定された組成を丸印で、使用したサンプルの組成（正しい組成）を星印で示している。いずれもよい一致を示している。

ここでは、使用した検体にエンドメンバー、すなわち、Ｅ、Ｓ、Ｍのピュアポリマーが含まれていないことが、Ｃの各頂点にプロットがないことからもわかる。

[0278] また、図中ＢＳは、推測によって得られたＥ、Ｍ、Ｓのピュアポリマーのスペクトルと、実測のＥ、Ｍ、Ｓのスペクトルとを重ねて示しているが、両者に差はなく、絶対強度を含めて、よく推定されていることが確かめられた。

[0279] このように、本発明の方法によれば、図１５のＢＳに示されたように、Ｅ、Ｍ、Ｓのピュアポリマーのスペクトル、及び、混合物の組成が、正確に推測され得ることが確かめられた。

なお、テストデータがある場合には、X_{test}をＳとＢに連続射影してC_{test}を得る。

[0280] ［実施例３］

（エンドメンバーを１つ含む検体による推定）

次に、（学習用）検体のうち１つがエンドメンバーを含む場合に、他の検体における成分含有量が任意となる（Ｋ－１次元単体の内接円の外側の領域でなくてもよい）ことを検証した。

[0281] 表５に記載したサンプルのうち、Ｓを４０質量％以上含むサンプルを除き、かつ、サンプルとしてＭピュアポリマーを加えたこと以外は、実施例２と同様にして、組成推定を行った。図１６は結果である。

[0282] 図１６において、中実円で表された点は、Ｂ、Ｓの推定に用いた検体の推定結果を表し、星印がその実際の組成である。この結果から、Ｓについては、内接円の外側の領域に検体はなかったものの（中空円は、推定に使用しなかった検体である）、精度よく推定することができることが確認された。

上記原因のひとつは、すでに説明した２回目のＮＭＦ処理（Ｅｑ．１１）において基底の非直交性項

[数228]

$$nonorth(\hat{\boldsymbol{B}})$$

をペナルティー項として組み込んだことによるものと考えらえる。

産業上の利用可能性

[0283] 材料開発の多くの場合で、材料性能を各構成成分に還元して考察することが求められる。これには、成分の同定と定量が必要だが、系が複雑化するほどその難易度は上がる。異なる成分のピークが互いにオーバーラップする可能性が低い高解像度の質量分析は、複雑な混合物系における成分の化学構造

を同定するうえで最も強力な分析手法と言える。一方で、質量分析は定量性を有さないことが問題となる。また、本質的には分離を必要としないものの、その結果得られるスペクトルは複雑で現実的には読解不能に陥りやすい。

[0284] 本発明では、材料の破片を大気下・前処理なしで直接熱分解させ、そこから生じるガスをＤＡＲＴ－ＭＳで分析し、そのスペクトルを教師無し学習することで、各成分のピュアなエンドメンバースペクトルを推定し、その線形和で検体のスペクトルを表す手法を開発した。これは、複雑な組成からなる材料であっても分離を必要とせずに、その成分の同定と定量を可能とする。

[0285] 同定については、他の成分に由来するピークをすべて排除した可読性の高いエンドメンバースペクトルを得られれば、あとは比較的容易である。一方で、定量分析の妨げとなる、ピークごとにイオン化効率が異なりピーク強度とその存在比が比例しない、という事実は本発明においても同様である。しかし、本発明ではエンドメンバースペクトルの各ピーク強度を、規格化された相対値ではなく、絶対強度として推定できるため、エンドメンバースペクトルの線形和で検体スペクトルを表現する際、その線形係数をそのまま組成比として求めることができる。

[0286] 検体を成分ごとに分解し、その含有比まで求める本発明は、材料の性能を発揮する要因となったキー成分の同定など、材料開発において重宝されるツールになり得るのは明らかであるが、さらには劣化成分の含有量検査等の品質管理、危険物含有量等の法令順守検査、及び、権利侵害の検出のための検査等、きわめて広範な応用が期待される。さらには樹脂の劣化・寿命予測、樹脂・有機材料における極微不純物成分の検出、嗅覚や味覚の定量分析、食品等のトレーサビリティー・産地偽装防止タグなどの技術が加速すると期待できる。

[0287] また、Ｐｌａｓｔｉｃｓ　Ｅｕｒｏｐｅは２０２１年９月に欧州委員会に対して２０３０年までにプラスチック容器包装に再生プラスチック３０％含有の義務化を提言するなど、再生材の利用が促進されているが、含有量を測定する一般的手法は確立されていない。本発明はバージン剤と再生材の比率

を正確に測定し得る技術としても、品質管理や法令順守検査などへの応用が期待される。

符号の説明

[0288] １０：Ｋ－１次元単体、　１２：内接超球、　１３－１５：エンドメンバー、　１６－１８：学習用検体、　１９：領域、　３０：組成推定装置、　３１：質量分析装置　、３２：減圧ポンプ、　３３：情報処理装置、　３４：プロセッサ、　３５：記憶デバイス、　３６：表示デバイス、　３７：入力デバイス、　４１：質量分析部、　４２：情報処理部、　４３：制御部、　４４：記憶部、　４５：入力部、　４６：出力部、　４７：データ行列作成部、　４８：ＮＭＦ処理部、　４９：補正処理部、　５０：ベクトル処理部、　５１：エンドメンバー決定部、　５２：含有量比計算部

請求の範囲

[請求項1] Ｋを１以上の整数としたとき、Ｋ種類の成分から選択される少なくとも１種の前記成分を含む推定対象検体における、前記成分の含有量比を推定する方法であって、

前記Ｋ種類の成分から選択される少なくとも１種の成分を含み、互いに組成の異なるＫ個以上の学習用検体、及び、前記成分を含まないバックグラウンド検体を準備することと、

前記推定対象検体、前記学習用検体、及び、バックグラウンド検体を含む検体セットに含まれるそれぞれの検体を加熱しながら、熱脱着、及び／又は、熱分解によって生ずるガス成分を順次イオン化し、マススペクトルを連続的に観測することと、

加熱温度ごとに得られた前記マススペクトルを各行に格納し、前記検体ごとの二次元マススペクトルを得て、前記二次元マススペクトルの少なくとも２つ以上をまとめてデータ行列に変換することと、

前記データ行列を非負値行列因子分解し、規格化された基底スペクトル行列とその強度分布行列の積に分解するＮＭＦ処理を行うことと、

前記基底スペクトル行列と前記データ行列との正準相関分析によって強度分布行列に含まれるノイズ成分を抽出し、前記ノイズ成分の影響を減ずるよう前記強度分布行列を補正して、補正後強度分布行列を得ることと、

前記補正後強度分布行列を前記検体のそれぞれに対応する小行列に分割し、前記小行列を特徴ベクトルとして、ベクトル空間内で前記検体のそれぞれを表現することと、

前記特徴ベクトルの全てを内包するＫ－１次元単体を設定し、前記Ｋ－１次元単体におけるＫ個のエンドメンバーを決定することと、

前記Ｋ個のエンドメンバーと、前記推定対象検体の特徴ベクトルとの、ユークリッド距離をそれぞれ計算し、前記ユークリッド距離の比

により、前記推定対象検体中の前記成分の含有量比を推定することと、を含み、

前記Ｋが３以上の場合、前記Ｋ－１次元単体に内接する超球体の外側の領域のそれぞれに、前記学習用検体の特徴ベクトルの少なくとも１つが位置する、又は、前記学習用検体が前記エンドメンバーの少なくとも１つを含む、方法。

[請求項2] 前記Ｋが２以上の整数であり、前記推定対象検体が、前記成分の混合物である、請求項１に記載の方法。

[請求項3] 前記学習用検体が、前記エンドメンバーを含む、請求項１又は２に記載の方法。

[請求項4] 前記エンドメンバーの決定が、前記学習用検体に付された判別ラベルに基づき行われる、請求項３に記載の方法。

[請求項5] 前記エンドメンバーの決定が、前記学習用検体の前記特徴ベクトルの頂点成分分析によって行われる、請求項３に記載の方法。

[請求項6] 前記エンドメンバーの決定が、前記Ｋ－１次元単体が最小体積となるように頂点を定めるアルゴリズムによって実施される、請求項１に記載の方法。

[請求項7] 前記エンドメンバーの決定が、前記補正後強度分布行列を非負値行列因子分解し、前記検体中のＫ種類の前記成分の重量分率を表す行列と、Ｋ種類の個別の前記成分ごとのフラグメント存在強度を表す行列との積に分解する第２のＮＭＦ処理により行われる、請求項６に記載の方法。

[請求項8] 前記学習用検体が、前記エンドメンバーを含まない、請求項６又は７に記載の方法。

[請求項9] 前記補正後強度分布行列を得ることは、更に、前記強度分布行列の強度補正を行うことを含む、請求項１～８のいずれか１項に記載の方法。

[請求項10] 前記検体セットが、更に、キャリブレーション検体を含み、

前記キャリブレーション検体は、前記Ｋ種類の成分を全て含み、かつ、その組成が既知であり、

前記強度補正は、前記強度分布行列を規格化することを含む、請求項９に記載の方法。

[請求項11] 前記強度補正は、前記強度分布行列の少なくとも一部を、前記検体セットのうち対応する前記検体の質量と環境変数の積で割り付けることを含み、

前記環境変数は、前記観測の際における前記成分のイオン化効率への影響を表す変数である、請求項９に記載の方法。

[請求項12] 前記環境変数が、前記観測の際に、雰囲気中に所定量含まれる分子量５０～１５００の化合物、又は、前記検体セットのそれぞれに所定量含まれる分子量５０～５００の有機低分子化合物のマススペクトルのピークの合計値である、請求項１１に記載の方法。

[請求項13] Ｋを１以上の整数としたとき、Ｋ種類の成分から選択される少なくとも１種の前記成分を含む推定対象検体における、前記成分の含有量比を推定する組成推定装置であって、

前記Ｋ種類の成分から選択される少なくとも１種の成分を含み、互いに組成の異なるＫ個以上の学習用検体、前記成分を含まないバックグラウンド検体、及び、前記推定対象検体を含む検体セットに含まれるそれぞれの検体を加熱しながら、熱脱着、及び／又は、熱分解によって生ずるガス成分を順次イオン化し、マススペクトルを連続的に観測する質量分析装置と、

前記観測されたマススペクトルを処理する情報処理装置と、を備え、

前記情報処理装置は、

加熱温度ごとに得られた前記マススペクトルを各行に格納し、前記検体ごとの二次元マススペクトルを得て、前記二次元マススペクトルの少なくとも２つ以上をまとめてデータ行列に変換する、データ行列

作成部と、

前記データ行列を非負値行列因子分解し、規格化された基底スペクトル行列とその強度分布行列の積に分解するＮＭＦ処理を行う、ＮＭＦ処理部と、

前記基底スペクトル行列と前記データ行列との正準相関分析によって強度分布行列に含まれるノイズ成分を抽出し、前記ノイズ成分の影響を減ずるよう前記強度分布行列を補正して、補正後強度分布行列を作成する、補正処理部と、

前記補正後強度分布行列を前記検体セットのそれぞれに対応する小行列に分割し、前記小行列を特徴ベクトルとして、ベクトル空間内で前記検体のそれぞれを表現する、ベクトル処理部と、

前記特徴ベクトルの全てを内包するＫ－１次元単体を設定し、前記Ｋ－１次元単体におけるＫ個のエンドメンバーを決定する、エンドメンバー決定部と、

前記Ｋ個のエンドメンバーと、前記推定対象検体の特徴ベクトルとの、ユークリッド距離をそれぞれ計算し、前記ユークリッド距離の比により、前記推定対象検体中の前記成分の含有量比を推定する、含有量比計算部と、を含み、

前記Ｋが３以上の場合、前記Ｋ－１次元単体に内接する超球体の外側の領域のそれぞれに、前記学習用検体の特徴ベクトルの少なくとも１つが位置する、又は、前記学習用検体が前記エンドメンバーの少なくとも１つを含む、組成推定装置。

[請求項14] 前記エンドメンバー決定部は、前記学習用検体に付された判別ラベルに基づいて前記エンドメンバーを決定する、請求項１３に記載の組成推定装置。

[請求項15] 前記エンドメンバー決定部は、前記Ｋ－１次元単体が最小体積となるように頂点を定めるアルゴリズムによって前記エンドメンバーを決定する、請求項１３に記載の組成推定装置。

[請求項16] 前記エンドメンバー決定部は、前記補正後強度分布行列を非負値行列因子分解し、前記検体中のＫ種類の前記成分の重量分率を表す行列と、Ｋ種類の個別の前記成分ごとのフラグメント存在強度を表す行列との積に分解する第２のＮＭＦ処理により前記エンドメンバーを決定する、請求項１５に記載の組成推定装置。

[請求項17] 前記補正処理部は更に、前記強度分布行列の強度補正を行う、請求項１５に記載の組成推定装置。

[請求項18] 前記検体セットが、更に、キャリブレーション検体を含み、

前記キャリブレーション検体が、前記Ｋ種類の成分を全て含み、かつ、その組成が既知であり、

前記強度補正が、前記強度分布行列を規格化することを含む、請求項１７に記載の組成推定装置。

[請求項19] 前記強度補正は、前記強度分布行列の少なくとも一部を、前記検体セットのうち対応する前記検体の質量と環境変数の積で割り付けることを含み、

前記環境変数は、前記観測の際における前記成分のイオン化効率への影響を表す変数である、請求項１７に記載の組成推定装置。

[請求項20] 前記質量分析装置は減圧ポンプを更に備え、

前記環境変数が、前記減圧ポンプの稼働によって生ずる物質のマススペクトルのピークの合計値である、請求項１９に記載の組成推定装置。

[請求項21] Ｋを１以上の整数としたとき、Ｋ種類の成分から選択される少なくとも１種の前記成分を含む推定対象検体における、前記成分の含有量比を推定する組成推定装置において用いられるプログラムであって、

前記Ｋ種類の成分から選択される少なくとも１種の成分を含み、互いに組成の異なるＫ個以上の学習用検体、前記成分を含まないバックグラウンド検体、及び、前記推定対象検体を含む検体セットに含まれるそれぞれの検体を加熱しながら、熱脱着、及び／又は、熱分解によ

って生ずるガス成分を順次イオン化し、マススペクトルを連続的に観測する質量分析装置と、

前記観測されたマススペクトルを処理する情報処理装置と、を備え、

前記質量分析装置によって、加熱温度ごとに取得された前記マススペクトルを各行に格納し、前記検体ごとの二次元マススペクトルを得て、前記二次元マススペクトルの少なくとも２つ以上をまとめてデータ行列に変換する、データ行列作成機能と、

前記データ行列を非負値行列因子分解し、規格化された基底スペクトル行列とその強度分布行列の積に分解するＮＭＦ処理を行う、ＮＭＦ処理機能と、

前記基底スペクトル行列と前記データ行列との正準相関分析によって強度分布行列に含まれるノイズ成分を抽出し、前記ノイズ成分の影響を減ずるよう前記強度分布行列を補正して、補正後強度分布行列を作成する、補正処理機能と、

前記補正後強度分布行列を前記検体のそれぞれに対応する小行列に分割し、前記小行列を特徴ベクトルとして、ベクトル空間内で前記検体のそれぞれを表現する、ベクトル処理機能と、

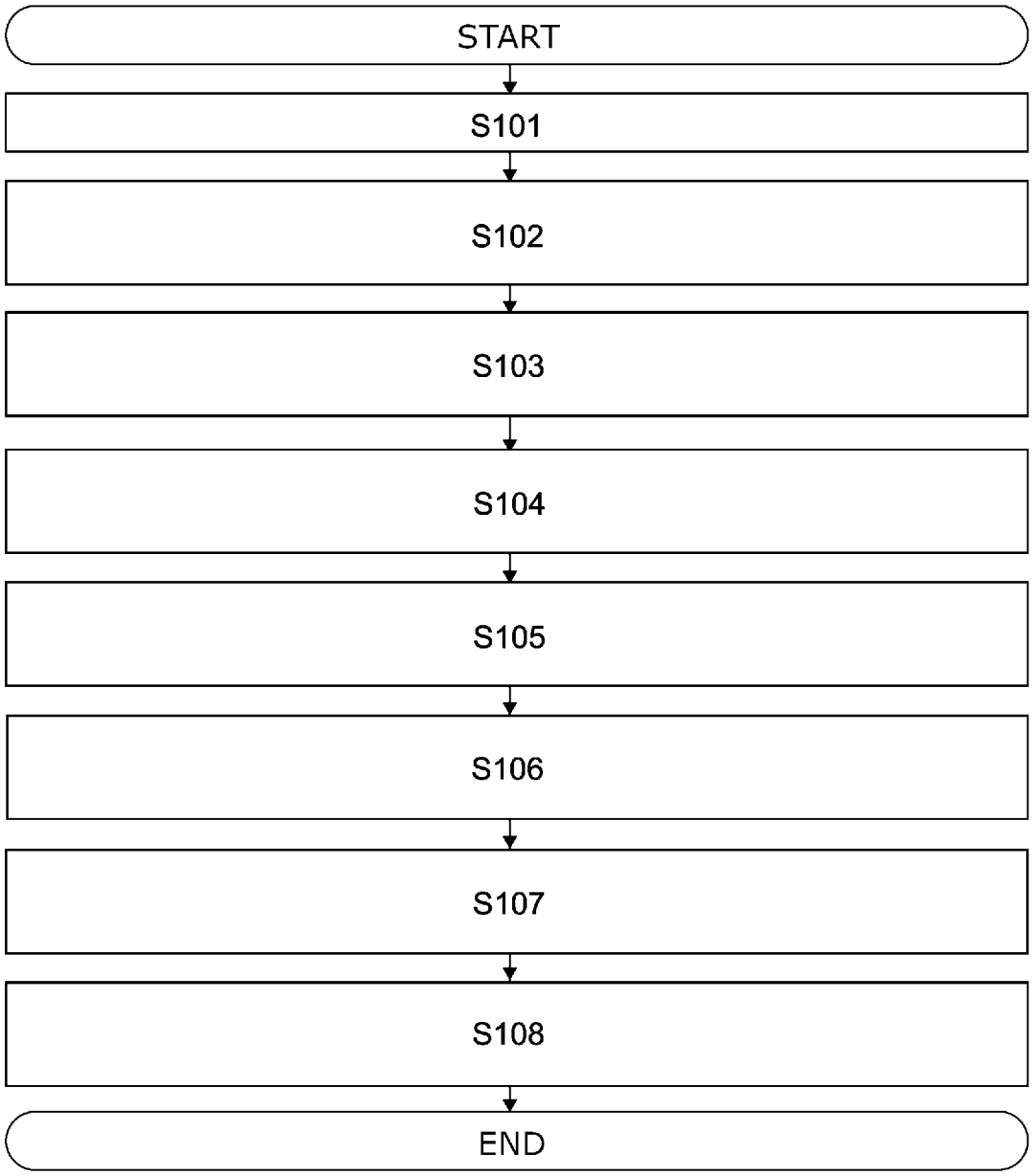
前記特徴ベクトルの全てを内包するＫ－１次元単体を設定し、前記Ｋ－１次元単体におけるＫ個のエンドメンバーを決定する、エンドメンバー決定機能と、

前記Ｋ個のエンドメンバーと、前記推定対象検体の特徴ベクトルとの、ユークリッド距離をそれぞれ計算し、前記ユークリッド距離の比により、前記推定対象検体中の前記成分の含有量比を推定する、含有量比計算機能と、を含み、

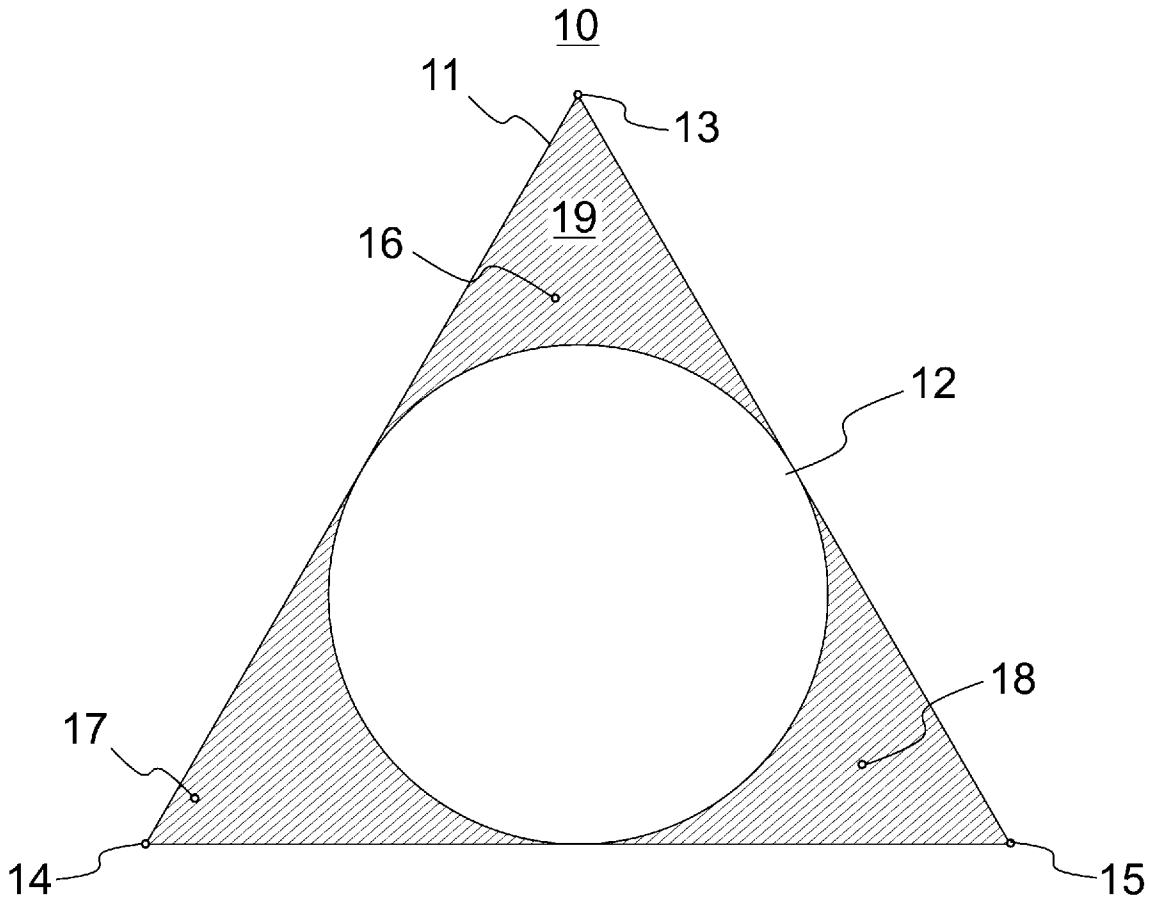
前記Ｋが３以上の場合、前記Ｋ－１次元単体に内接する超球体の外側の領域のそれぞれに、前記学習用検体の特徴ベクトルの少なくとも１つが位置する、又は、前記学習用検体が前記エンドメンバーの少な

くとも１つを含む、プログラム。

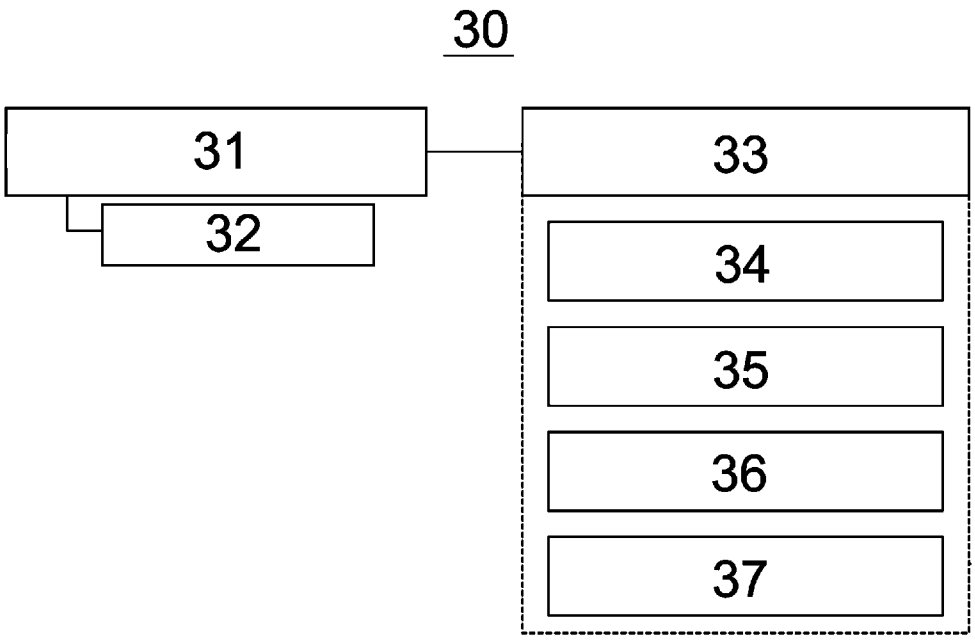
[図1]



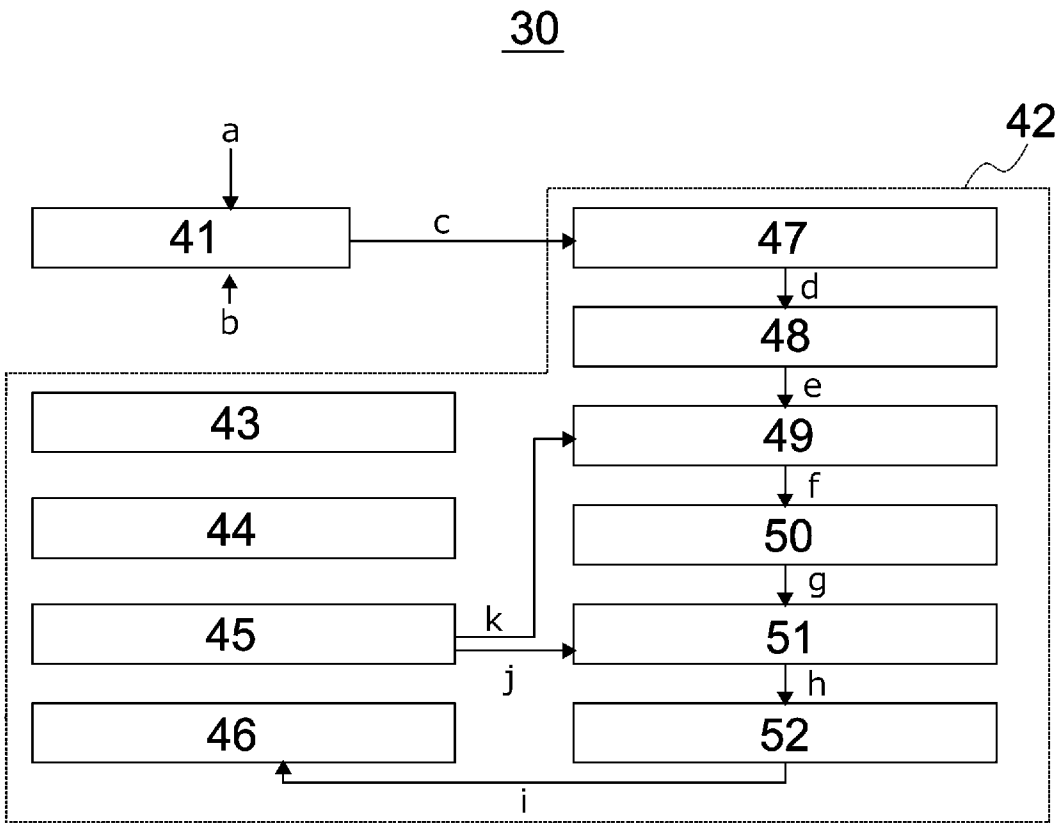
[図2]



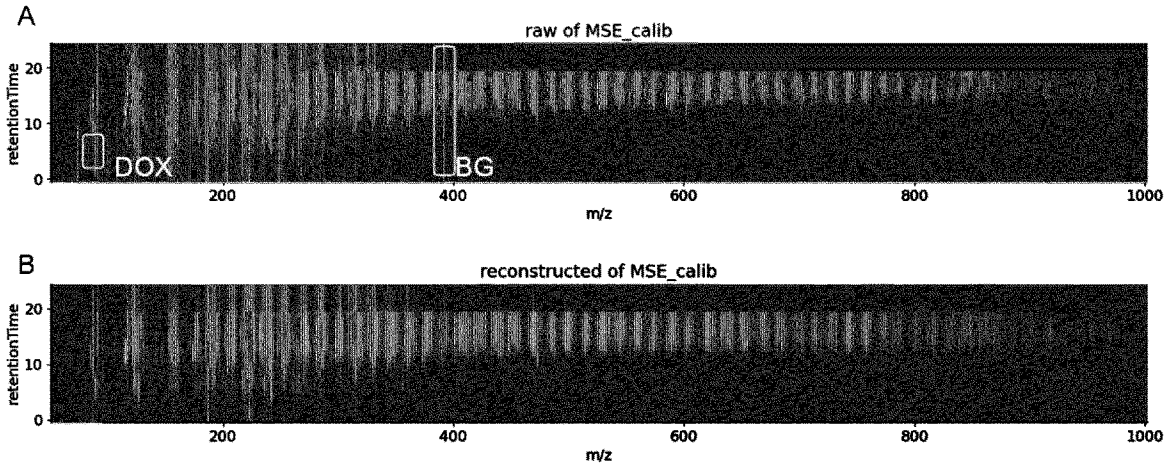
[図3]



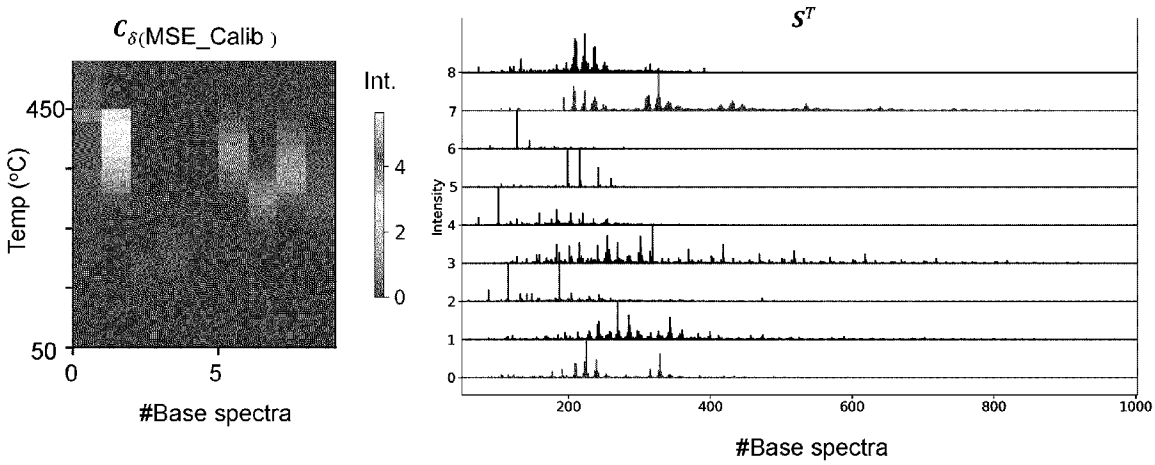
[図4]


30
42
a
41
b
c
47
d
48
e
49
f
50
g
51
h
52
43
44
45
46
k
j
i

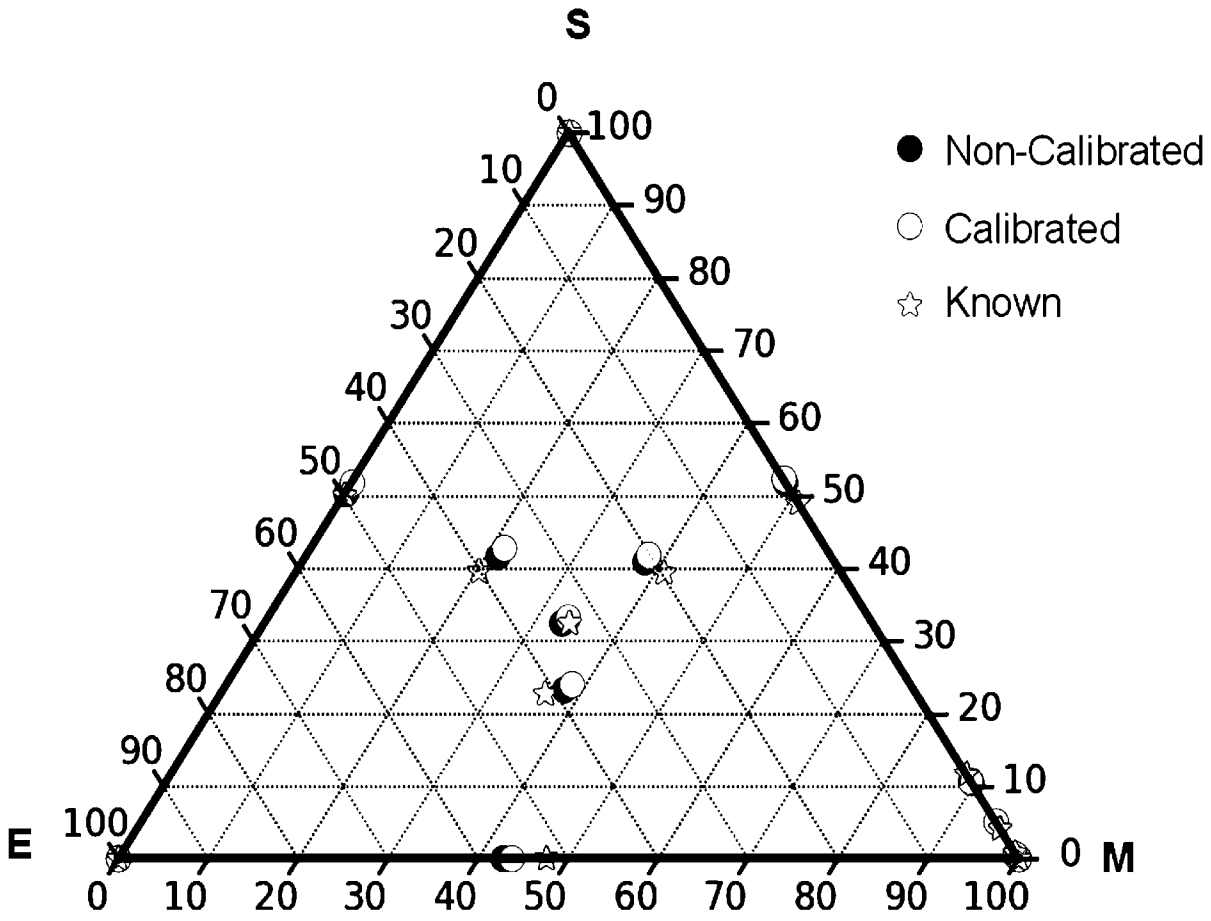
[図5]


A
raw of MSE_calib
retentionTime
DOX
BG
m/z
B
reconstructed of MSE_calib
retentionTime
m/z

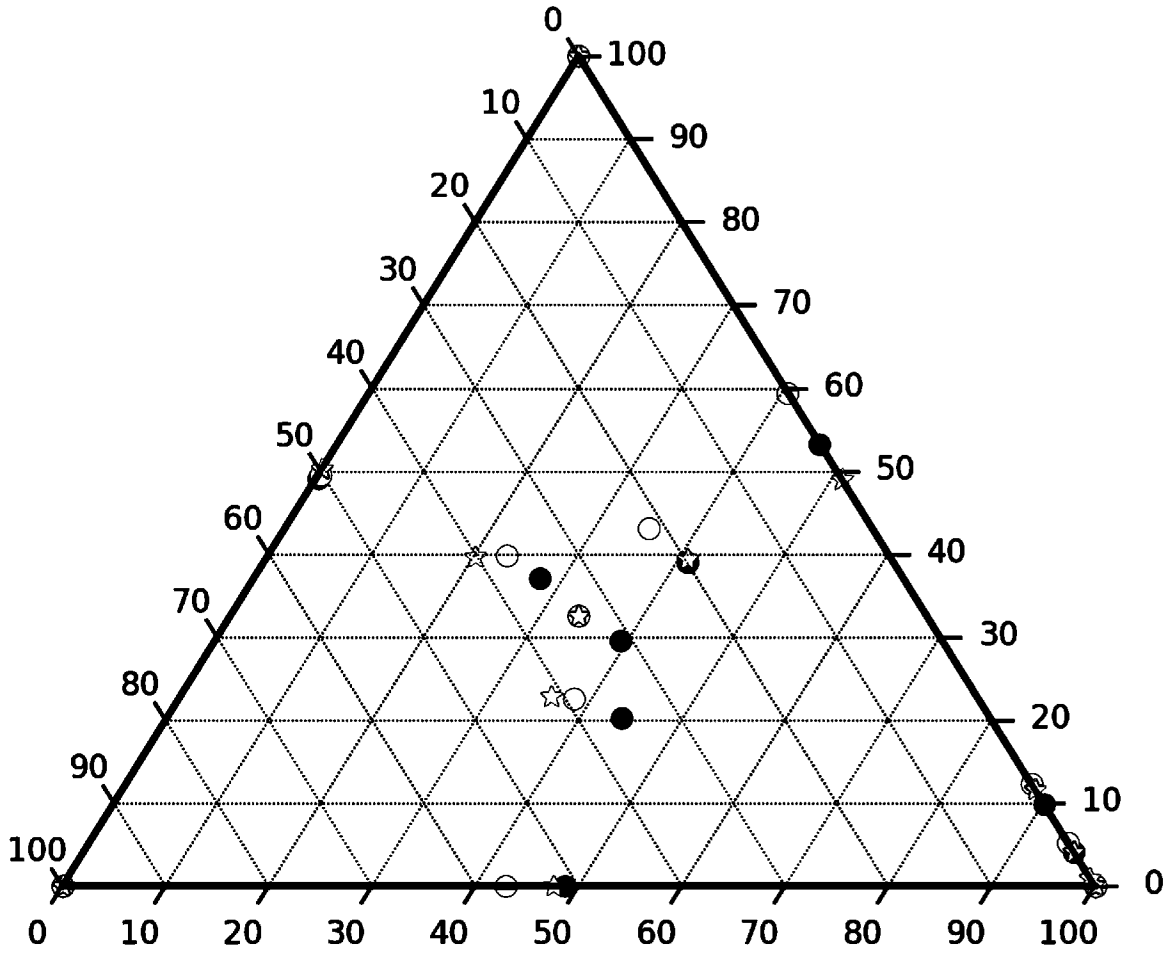
[図6]


$C_{\delta(\mathrm{MSE_Calib})}$
Temp (°C)
450
50
0
5
#Base spectra
Int.
4
2
0
S^T
Intensity
8
7
6
5
4
3
2
1
0
200
400
600
800
1000
#Base spectra

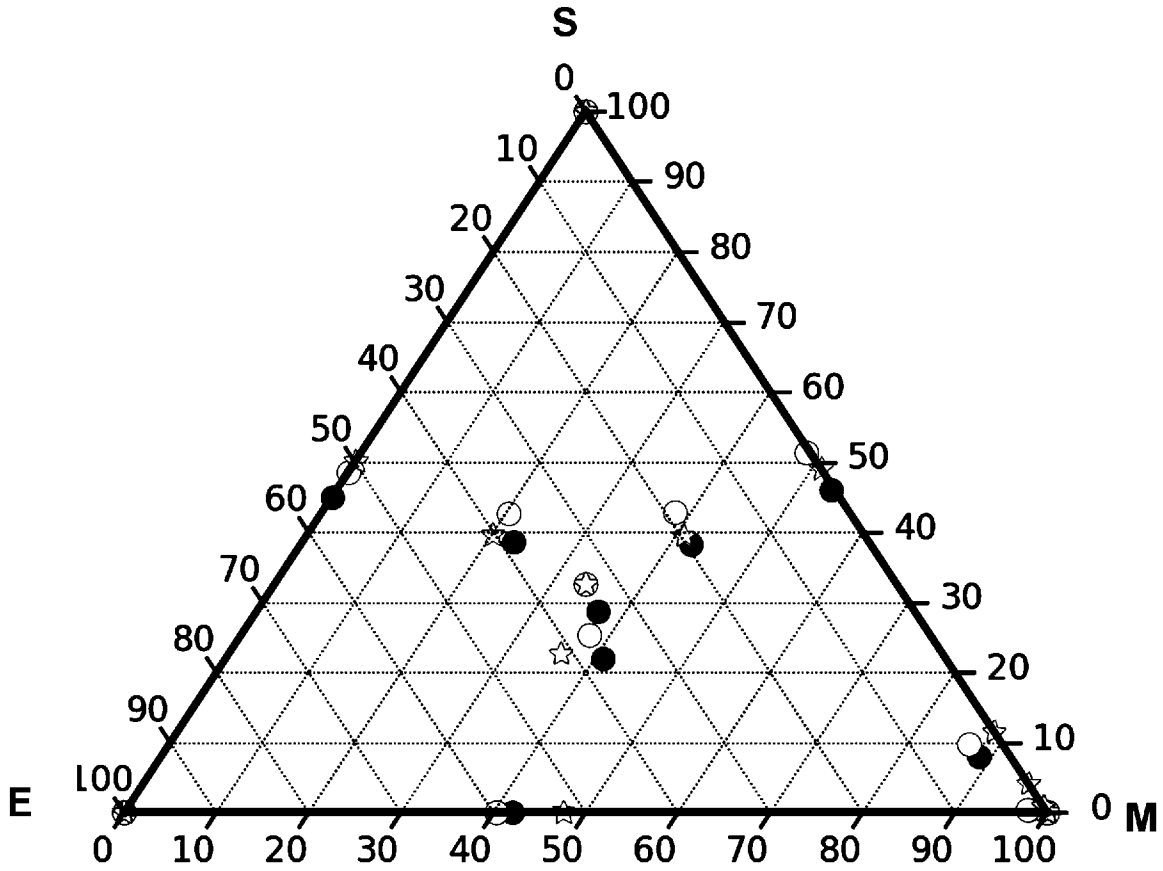
[図7]


S
E
M
Non-Calibrated
Calibrated
Known
0
10
20
30
40
50
60
70
80
90
100

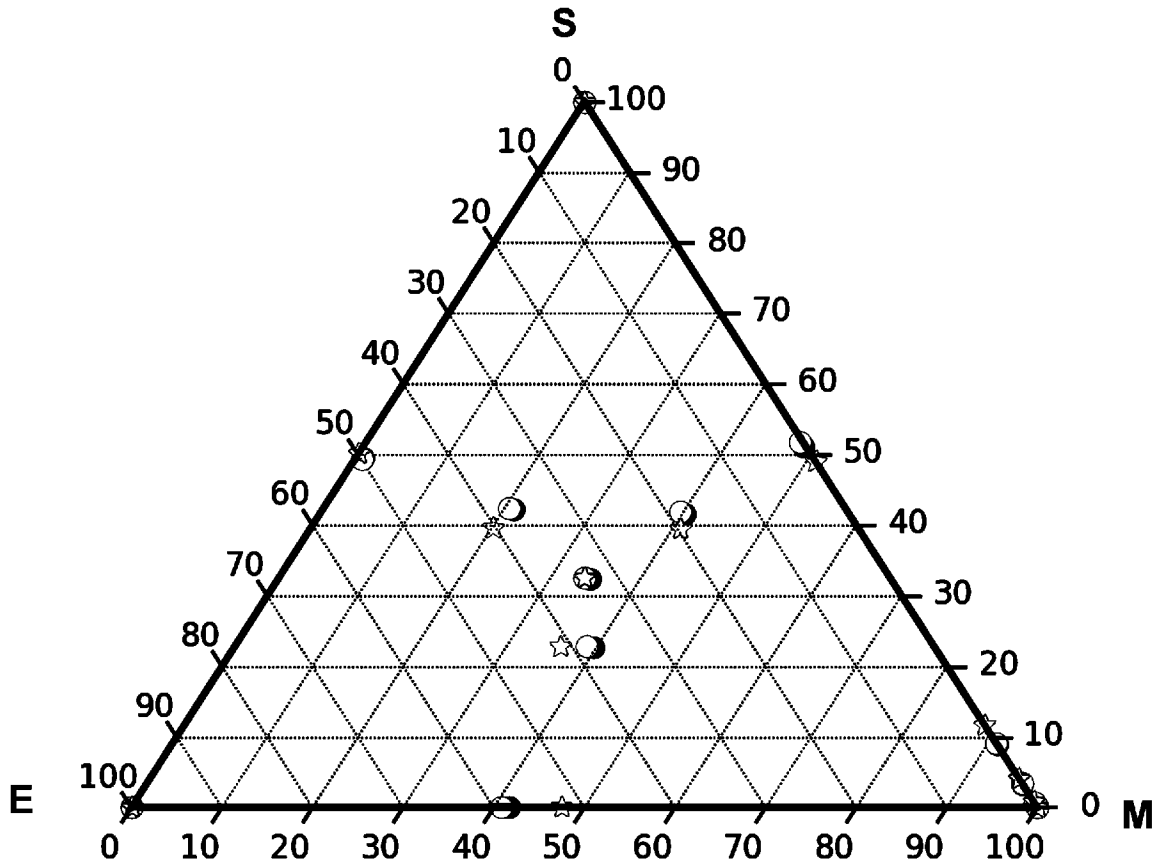
[図8]


0
100
10
90
20
80
30
70
40
60
50
50
60
40
70
30
80
20
90
10
100
0
0
10
20
30
40
50
60
70
80
90
100

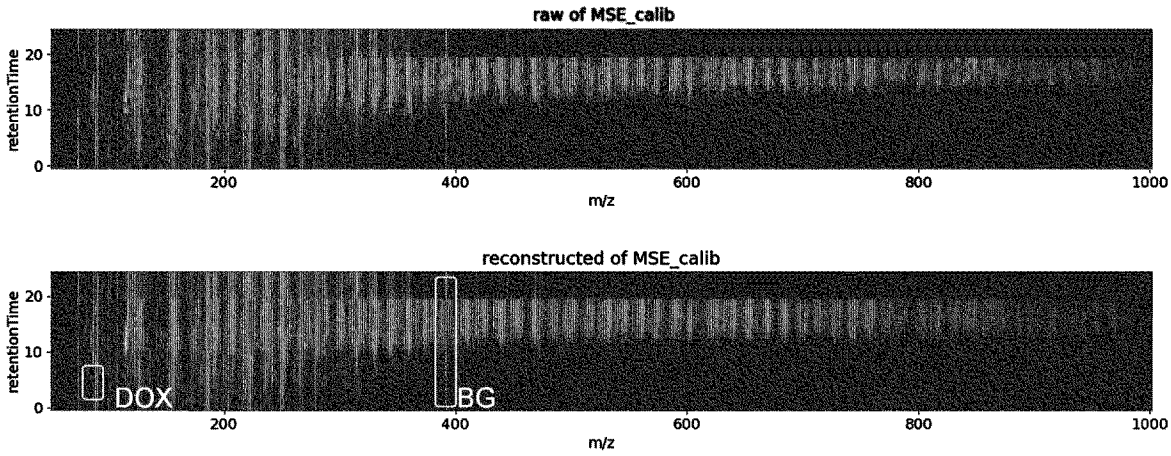
[図9]


S
E
M
0
10
20
30
40
50
60
70
80
90
100

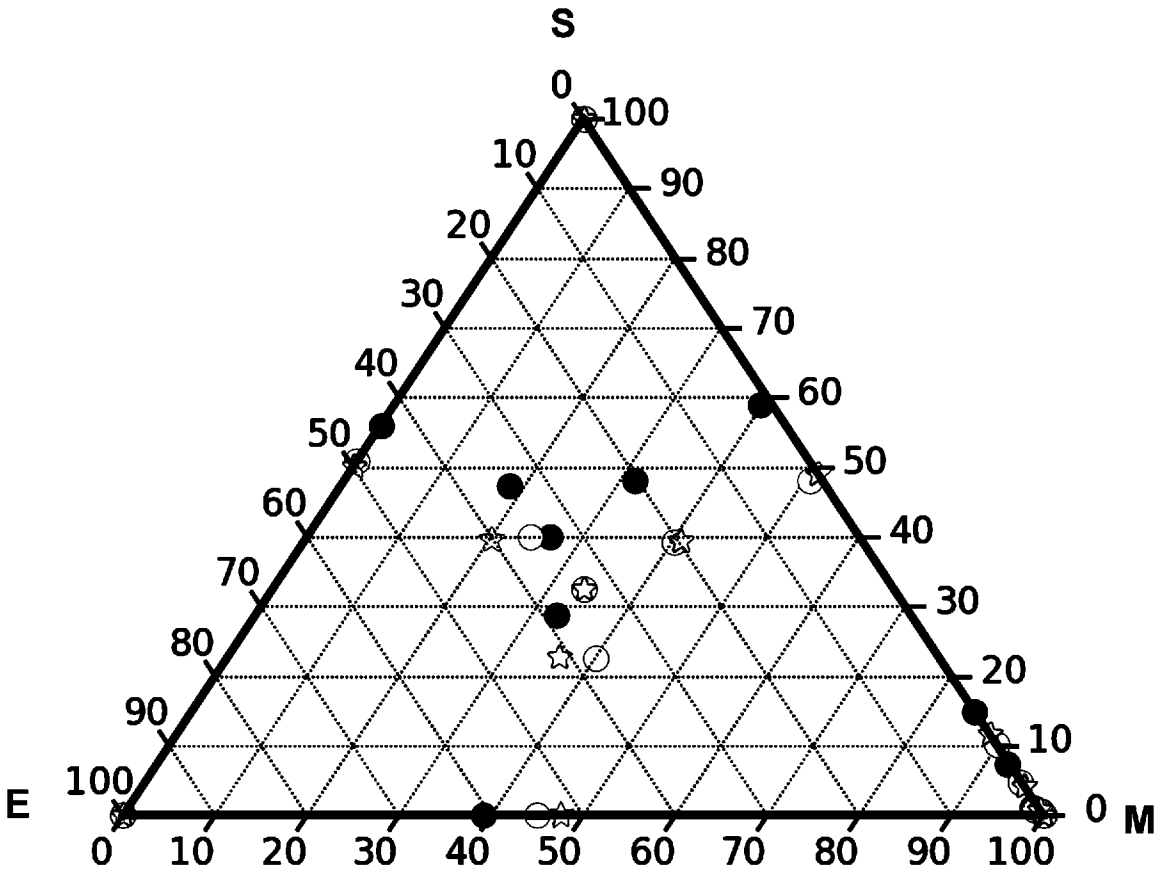
[図10]


S
E
M
0
10
20
30
40
50
60
70
80
90
100

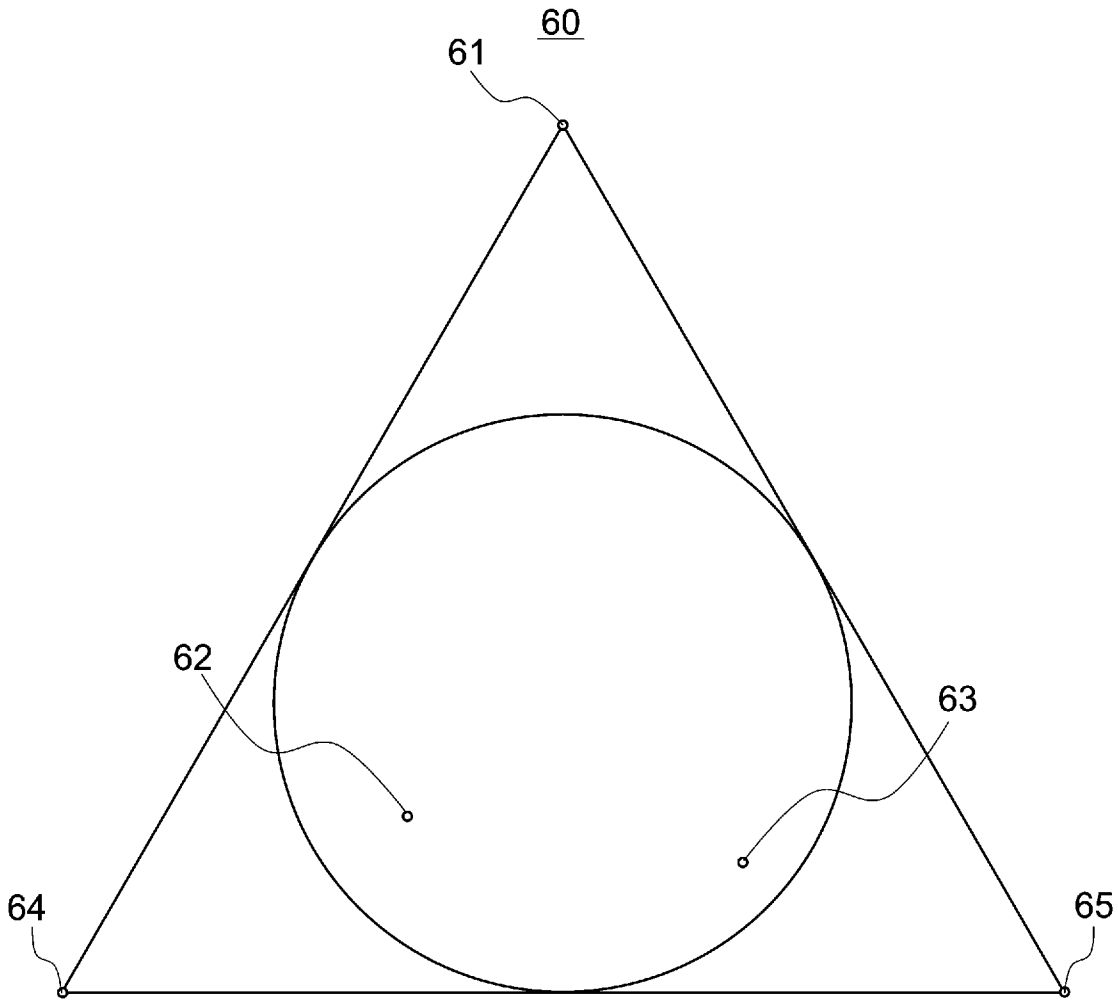
[図11]


raw of MSE_calib
retentionTime
0
10
20
200
400
600
800
1000
m/z
reconstructed of MSE_calib
retentionTime
DOX
BG
m/z

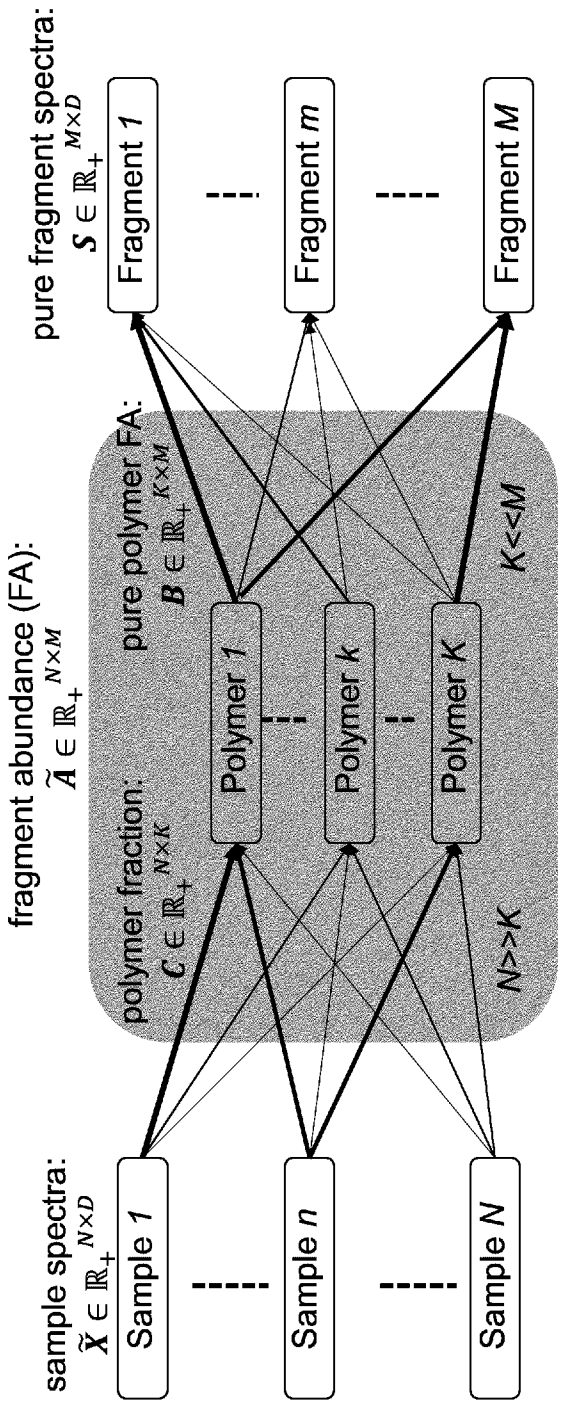
[図12]


S
E
M
0
10
20
30
40
50
60
70
80
90
100

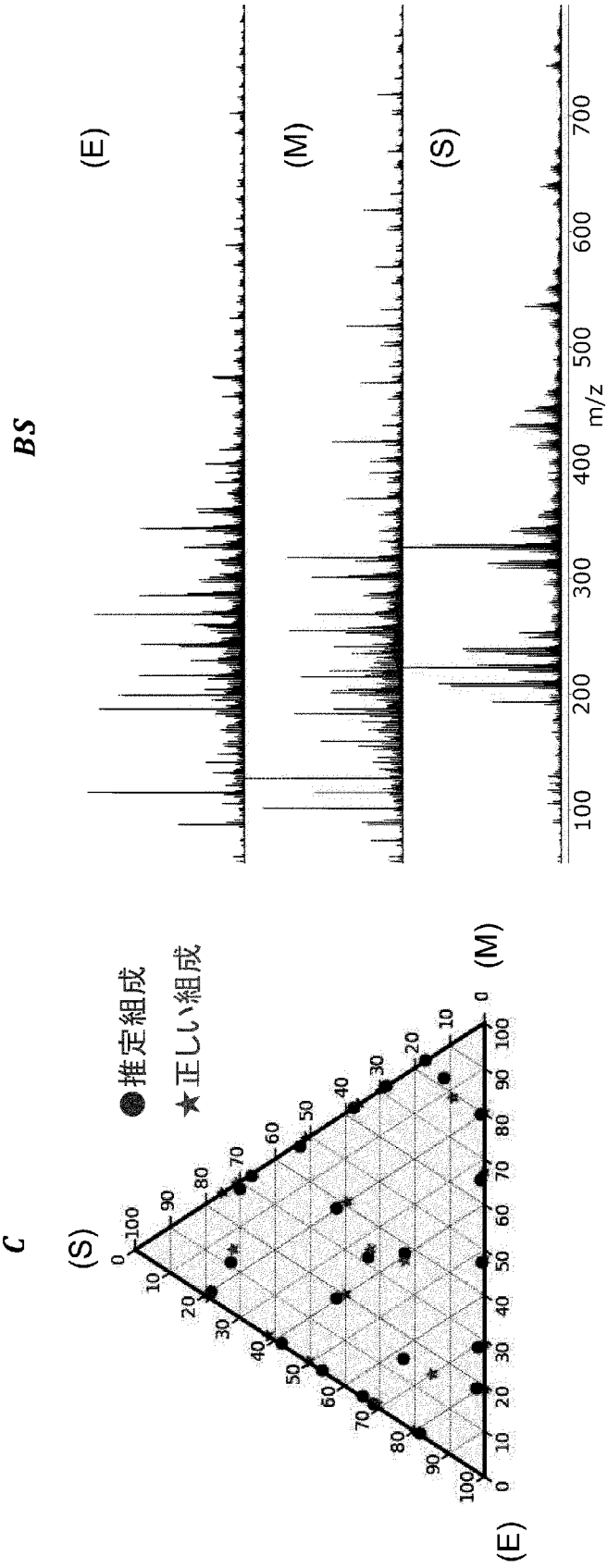
[図13]



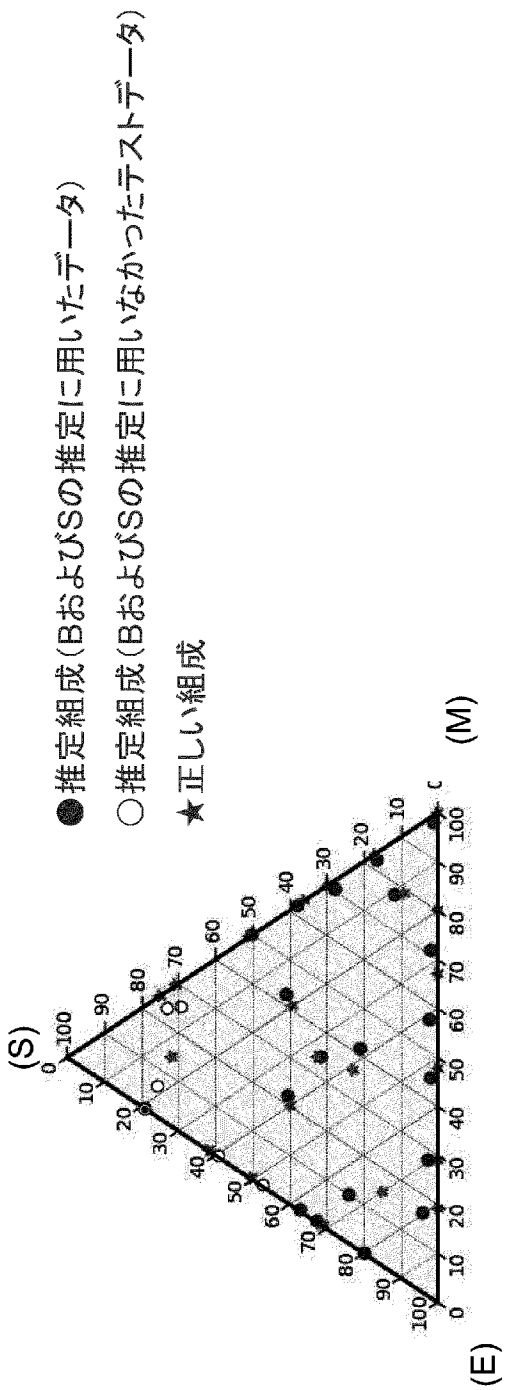
[図14]


pure fragment spectra:
$\boldsymbol{S} \in \mathbb{R}_+^{M \times D}$
Fragment 1
Fragment m
Fragment M
pure polymer FA:
$\boldsymbol{B} \in \mathbb{R}_+^{K \times M}$
fragment abundance (FA):
$\tilde{\boldsymbol{A}} \in \mathbb{R}_+^{N \times M}$
polymer fraction:
$\boldsymbol{C} \in \mathbb{R}_+^{N \times K}$
Polymer 1
Polymer k
Polymer K
K<<M
N>>K
sample spectra:
$\tilde{\boldsymbol{X}} \in \mathbb{R}_+^{N \times D}$
Sample 1
Sample n
Sample N

[図15]



[図16]


●推定組成（BおよびSの推定に用いたデータ）
○推定組成（BおよびSの推定に用いなかったテストデータ）
★正しい組成
(S)
(M)
(E)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2022/022813

A. CLASSIFICATION OF SUBJECT MATTER

G01N 27/62(2021.01)i
FI: G01N27/62 D

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G01N27/62

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan 1922-1996
Published unexamined utility model applications of Japan 1971-2022
Registered utility model specifications of Japan 1996-2022
Published registered utility model applications of Japan 1994-2022

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

JSTPlus/JMEDPlus/JST7580 (JDreamIII)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2016/120958 A1 (SHIMADZU CORP.) 04 August 2016 (2016-08-04)	1-21
A	WO 2018/158801 A1 (SHIMADZU CORP.) 07 September 2018 (2018-09-07)	1-21
A	JP 2012-516445 A (HOLOGIC, INC.) 19 July 2012 (2012-07-19)	1-21
A	JP 2013-528287 A (INTERNATIONAL BUSINESS MACHINES CORP.) 08 July 2013 (2013-07-08)	1-21
A	JP 2021-81365 A (SHIMADZU CORP.) 27 May 2021 (2021-05-27)	1-21

☐ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 June 2022	**19 July 2022**
Name and mailing address of the ISA/JP **Japan Patent Office (ISA/JP)** **3-4-3 Kasumigaseki, Chiyoda-ku, Tokyo 100-8915** **Japan**	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/JP2022/022813

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
WO 2016/120958 A1	04 August 2016	US 2017/0356889 A1	
WO 2018/158801 A1	07 September 2018	(Family: none)	
JP 2012-516445 A	19 July 2012	US 2010/0190652 A1	
		WO 2010/088187 A1	
		CA 2750818 A1	
		CN 102334033 A	
JP 2013-528287 A	08 July 2013	US 2013/0073219 A1	
		US 2014/0052385 A1	
		WO 2011/154219 A2	
		CN 102906851 A	
JP 2021-81365 A	27 May 2021	US 2021/0208117 A1	
		CN 112824894 A	

国際調査報告

国際出願番号
PCT/JP2022/022813

A. 発明の属する分野の分類（国際特許分類（ＩＰＣ））

G01N 27/62(2021.01)i
FI: G01N27/62 D

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（ＩＰＣ））

G01N27/62

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	１９２２－１９９６年
日本国公開実用新案公報	１９７１－２０２２年
日本国実用新案登録公報	１９９６－２０２２年
日本国登録実用新案公報	１９９４－２０２２年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

JSTPlus/JMEDPlus/JST7580 (JDreamIII)

C. 関連すると認められる文献

引用文献のカテゴリー*	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求項の番号
A	WO 2016/120958 A1 (株式会社島津製作所) 04.08.2016 (2016-08-04)	1-21
A	WO 2018/158801 A1 (株式会社島津製作所) 07.09.2018 (2018-09-07)	1-21
A	JP 2012-516445 A (ホロジック，インコーポレイテッド) 19.07.2012 (2012-07-19)	1-21
A	JP 2013-528287 A (インターナショナル・ビジネス・マシーンズ・コーポレーション) 08.07.2013 (2013-07-08)	1-21
A	JP 2021-81365 A (株式会社島津製作所) 27.05.2021 (2021-05-27)	1-21

☐ Ｃ欄の続きにも文献が列挙されている。

☑ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー
“A” 特に関連のある文献ではなく、一般的技術水準を示すもの
“E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
“L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
“O” 口頭による開示、使用、展示等に言及する文献
“P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献
“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの
“X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
“Y” 特に関連のある文献であって、当該文献と他の１以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
“&” 同一パテントファミリー文献

国際調査を完了した日	国際調査報告の発送日
30.06.2022	19.07.2022

名称及びあて先

日本国特許庁(ISA/JP)
〒100-8915
日本国
東京都千代田区霞が関三丁目４番３号

権限のある職員（特許庁審査官）

吉田　将志 2W 4636

電話番号 03-3581-1101 内線 3258

国際調査報告
パテントファミリーに関する情報

国際出願番号
PCT/JP2022/022813

引用文献	公表日	パテントファミリー文献	公表日
WO 2016/120958 A1	04.08.2016	US 2017/0356889 A1	
WO 2018/158801 A1	07.09.2018	（ファミリーなし）	
JP 2012-516445 A	19.07.2012	US 2010/0190652 A1	
		WO 2010/088187 A1	
		CA 2750818 A1	
		CN 102334033 A	
JP 2013-528287 A	08.07.2013	US 2013/0073219 A1	
		US 2014/0052385 A1	
		WO 2011/154219 A2	
		CN 102906851 A	
JP 2021-81365 A	27.05.2021	US 2021/0208117 A1	
		CN 112824894 A	