

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2020-60444
(P2020-60444A)

(43) 公開日 令和2年4月16日(2020.4.16)

(51) Int.Cl.	F I	テーマコード (参考)
G 0 1 N 27/62 (2006.01)	G O 1 N 27/62	2 G O 4 1
H 0 1 J 49/26 (2006.01)	H O 1 J 49/26	4 B O 2 9
C 1 2 M 1/34 (2006.01)	C 1 2 M 1/34	4 B O 6 3
C 1 2 Q 1/04 (2006.01)	C 1 2 Q 1/04	5 C O 3 8
	G O 1 N 27/62	V
	審査請求 未請求 請求項の数 11 O L (全 15 頁)	

(21) 出願番号 特願2018-191764 (P2018-191764)
(22) 出願日 平成30年10月10日 (2018.10.10)

(71) 出願人 000001993
株式会社島津製作所
京都府京都市中京区西ノ京桑原町1番地

(71) 出願人 599002043
学校法人 名城大学
愛知県名古屋市天白区塙釜口1-501

(74) 代理人 100098305
弁理士 福島 样人

(74) 代理人 100108523
弁理士 中川 雅博

(74) 代理人 100125704
弁理士 坂根 剛

(74) 代理人 100187931
弁理士 鐘村 英幸

最終頁に続く

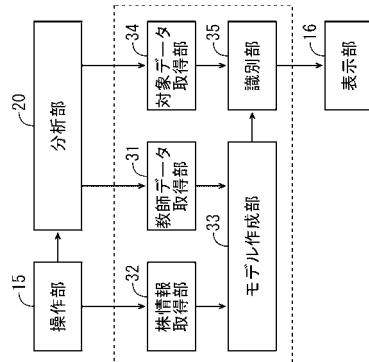
(54) 【発明の名称】 質量分析装置、質量分析方法および質量分析プログラム

(57) 【要約】

【課題】微生物の株の識別の精度を向上させることができ
可能な質量分析装置、質量分析方法および質量分析プログラ
ムを提供する。

【解決手段】株が既知の微生物を含む複数のマススペクトルデータの各々が教師データとして教師データ取得部31により取得される。各教師データに対応するサンプルは添加剤を含むとともにマトリックスが混合されている。機械学習を行うことにより、取得された複数の教師データに基づいて株を識別するための判別分析モデルがモデル作成部33により作成される。株が未知の微生物を含むマススペクトルデータが対象データとして対象データ取得部34により取得される。対象データに対応するサンプルは添加剤を含むとともにマトリックスが混合されている。作成された株ごとの判別分析モデルと取得された対象データとに基づいて当該対象データに対応する微生物の株が識別部35により識別される。

【選択図】図2



【特許請求の範囲】**【請求項 1】**

微生物の株を識別する質量分析装置であって、

株が既知の微生物を含みかつ添加剤を含むとともにマトリックスが混合された複数のサンプルについての複数のマススペクトルデータの各々を教師データとして取得する教師データ取得部と、

機械学習を行うことにより、前記教師データ取得部により取得された複数の教師データに基づいて株を識別するための判別分析モデルを作成するモデル作成部と、

株が未知の微生物を含みかつ前記添加剤を含むとともに前記マトリックスが混合されたサンプルについてのマススペクトルデータを対象データとして取得する対象データ取得部と、

前記モデル作成部により作成された株ごとの判別分析モデルと対象データ取得部により取得された対象データとに基づいて当該対象データに対応する微生物の株を識別する識別部とを備える、質量分析装置。

【請求項 2】

前記添加剤は、アルカリ金属付加イオン検出を抑制する化合物および界面活性剤の少なくとも一方を含む、請求項 1 記載の質量分析装置。

【請求項 3】

前記添加剤は、メチレンジホスホン酸またはデシル - - D - マルトピラノシドを含む、請求項 2 記載の質量分析装置。

【請求項 4】

前記モデル作成部は、サポートベクターマシンまたはニューラルネットワークにより判別分析モデルを作成する、請求項 1 ~ 3 のいずれか一項に記載の質量分析装置。

【請求項 5】

前記マトリックスはシナピン酸を含む、請求項 1 ~ 4 のいずれか一項に記載の質量分析装置。

【請求項 6】

微生物の株を識別する質量分析方法であって、

株が既知の微生物を含みかつ添加剤を含むとともにマトリックスが混合された複数のサンプルについての複数のマススペクトルデータの各々を教師データとして取得するステップと、

機械学習を行うことにより、取得された複数の教師データに基づいて株を識別するための判別分析モデルを作成するステップと、

株が未知の微生物を含みかつ前記添加剤を含むとともに前記マトリックスが混合されたサンプルについてのマススペクトルデータを対象データとして取得するステップと、

作成された株ごとの判別分析モデルと取得された対象データとに基づいて当該対象データに対応する微生物の株を識別するステップとを含む、質量分析方法。

【請求項 7】

前記添加剤は、アルカリ金属付加イオン検出を抑制する化合物および界面活性剤の少なくとも一方を含む、請求項 6 記載の質量分析方法。

【請求項 8】

前記添加剤は、チレンジホスホン酸またはデシル - - D - マルトピラノシドを含む、請求項 7 記載の質量分析方法。

【請求項 9】

前記判別分析モデルを作成するステップは、サポートベクターマシンまたはニューラルネットワークにより判別分析モデルを作成することを含む、請求項 6 ~ 8 のいずれか一項に記載の質量分析方法。

【請求項 10】

前記マトリックスはシナピン酸を含む、請求項 6 ~ 9 のいずれか一項に記載の質量分析方法。

10

20

30

40

50

【請求項 11】

処理装置により実行可能な微生物の株を識別する質量分析プログラムであって、

株が既知の微生物を含みかつ添加剤を含むとともにマトリックスが混合された複数のサンプルについての複数のマススペクトルデータの各々を教師データとして取得する処理と、

機械学習を行うことにより、取得された複数の教師データに基づいて株を識別するための判別分析モデルを作成する処理と、

株が未知の微生物を含みかつ前記添加剤を含むとともに前記マトリックスが混合されたサンプルについてのマススペクトルデータを対象データとして取得する処理と、

作成された株ごとの判別分析モデルと取得された対象データとに基づいて当該対象データに対応する微生物の株を識別する処理とを、

前記処理装置に実行させる、質量分析プログラム。

【発明の詳細な説明】**【技術分野】****【0001】**

本発明は、微生物を同定する質量分析装置、質量分析方法および質量分析プログラムに関する。

【背景技術】**【0002】**

種々の微生物のサンプルを同定するために質量分析装置が用いられる。複数のサンプルについて得られた複数のマススペクトルを比較することにより、各サンプルを識別するためのマーカピークを検出することができる。MALDI-MS(マトリックス支援レーザ脱離イオン化質量分析法)による微生物同定システム(以下、MALDI-MSシステムと呼ぶ。)は、迅速性および低コスト性に優れ、近年、臨床現場で急速に普及している。

【0003】

現時点では、臨床現場においては、MALDI-MSシステムによる微生物同定は種レベルの識別に留まっている。一方で、学術研究においては、微生物が菌株レベルで識別されたことが報告されている。例えば、非特許文献1では、遺伝子情報に基づいて特定の株にのみ発現するタンパク質(主にリボソームタンパク質)の理論質量が算出される。算出された理論質量に対応する質量電荷比におけるピーク(マーカピーク)の有無により株の識別が行われる。

【先行技術文献】**【非特許文献】****【0004】**

【非特許文献1】Yudai Hotta et al., "Classification of the Genus *Bacillus* Based on MALDI-TOF MS Analysis of Ribosomal Proteins Coded in S10 and spc Operons," Journal of Agricultural and Food Chemistry, 2011, Vol. 59, No. 10, pp. 5222-5230

【発明の概要】**【発明が解決しようとする課題】****【0005】**

臨床現場においても、微生物の株の識別が実用化されることにより、微生物の感染ルートの解明または毒性の有無の判別が可能になることが期待されている。しかしながら、微生物の株を高い精度で識別することは容易ではない。

【0006】

本発明の目的は、微生物の株の識別の精度を向上させることができるものである。

【課題を解決するための手段】**【0007】**

本発明者らは、複数のマススペクトルを用いて機械学習を行うことにより微生物の株を識別するための判別分析モデルを作成することを検討した。そして、種々の実験および考

10

20

30

40

50

察を行った結果、各マススペクトルにおけるピーク強度のばらつきを抑制することにより、株の識別に用いることが可能な判別分析モデルを作成することができるという知見を得た。これらの知見に基づいて、以下の本発明に想到した。

【0008】

(1) 第1の発明に係る質量分析装置は、微生物の株を識別する質量分析装置であって、株が既知の微生物を含みかつ添加剤を含むとともにマトリックスが混合された複数のサンプルについての複数のマススペクトルデータの各々を教師データとして取得する教師データ取得部と、機械学習を行うことにより、教師データ取得部により取得された複数の教師データに基づいて株を識別するための判別分析モデルを作成するモデル作成部と、株が未知の微生物を含みかつ添加剤を含むとともにマトリックスが混合されたサンプルについてのマススペクトルデータを対象データとして取得する対象データ取得部と、モデル作成部により作成された株ごとの判別分析モデルと対象データ取得部により取得された対象データとに基づいて当該対象データに対応する微生物の株を識別する識別部とを備える。10

【0009】

この質量分析装置においては、株が既知の微生物に対応する複数のマススペクトルデータの各々が教師データとして取得される。各教師データに対応するサンプルは添加剤を含むとともにマトリックスが混合されている。機械学習を行うことにより、取得された複数の教師データに基づいて株を識別するための判別分析モデルが作成される。また、株が未知の微生物に対応するマススペクトルデータが対象データとして取得される。対象データに対応するサンプルは添加剤を含むとともにマトリックスが混合されている。作成された株ごとの判別分析モデルと取得された対象データとに基づいて当該対象データに対応する微生物の株が識別される。20

【0010】

この構成によれば、各教師データにおけるピーク強度のばらつきが抑制される。そのため、取得される複数の教師データを機械学習することにより、株の識別に用いることが可能な判別分析モデルを作成することができる。また、各教師データと同様に、対象データにおけるピーク強度のばらつきが抑制される。これにより、作成された判別分析モデルと対象データとに基づいて対象データに対応する微生物の株を高い精度で識別することができる。その結果、微生物の株の識別の精度が向上する。30

【0011】

(2) 添加剤は、アルカリ金属付加イオン検出を抑制する化合物および界面活性剤の少なくとも一方を含んでもよい。この場合、複数の教師データおよび対象データの各々におけるピーク強度のばらつきを効率的に抑制することができる。

【0012】

(3) 添加剤は、メチレンジホスホン酸またはデシル- - D - マルトピラノシドを含んでもよい。この場合、複数の教師データおよび対象データの各々におけるピーク強度のばらつきをより効率的に抑制することができる。

【0013】

(4) モデル作成部は、サポートベクターマシンまたはニューラルネットワークにより判別分析モデルを作成してもよい。この場合、高い精度で株を識別するための判別分析モデルを容易に作成することができる。40

【0014】

(5) マトリックスはシナピン酸を含んでもよい。この場合、複数の教師データおよび対象データの各々を容易に取得することができる。また、複数の教師データおよび対象データの各々におけるピーク強度のばらつきを効率的に抑制することができる。

【0015】

(6) 第2の発明に係る質量分析方法は、微生物の株を識別する質量分析方法であって、株が既知の微生物を含みかつ添加剤を含むとともにマトリックスが混合された複数のサンプルについての複数のマススペクトルデータの各々を教師データとして取得するステップと、機械学習を行うことにより、取得された複数の教師データに基づいて株を識別する50

ための判別分析モデルを作成するステップと、株が未知の微生物を含みかつ添加剤を含むとともにマトリックスが混合されたサンプルについてのマススペクトルデータを対象データとして取得するステップと、作成された株ごとの判別分析モデルと取得された対象データとに基づいて当該対象データに対応する微生物の株を識別するステップとを含む。

【0016】

この質量分析方法によれば、作成された判別分析モデルと対象データとに基づいて対象データに対応する微生物の株を高い精度で識別することができる。その結果、微生物の株の識別の精度が向上する。

【0017】

(7) 添加剤は、アルカリ金属付加イオン検出を抑制する化合物および界面活性剤の少なくとも一方を含んでもよい。この場合、複数の教師データおよび対象データの各々におけるピーク強度のばらつきを効率的に抑制することができる。10

【0018】

(8) 添加剤は、チレンジホスホン酸またはデシル- - D - マルトピラノシドを含んでもよい。この場合、複数の教師データおよび対象データの各々におけるピーク強度のばらつきをより効率的に抑制することができる。

【0019】

(9) 判別分析モデルを作成するステップは、サポートベクターマシンまたはニューラルネットワークにより判別分析モデルを作成することを含んでもよい。この場合、高い精度で株を識別するための判別分析モデルを容易に作成することができる。20

【0020】

(10) マトリックスはシナピン酸を含んでもよい。この場合、複数の教師データおよび対象データの各々を容易に取得することができる。また、複数の教師データおよび対象データの各々におけるピーク強度のばらつきを効率的に抑制することができる。

【0021】

(11) 第3の発明に係る質量分析プログラムは、処理装置により実行可能な微生物の株を識別する質量分析プログラムであって、株が既知の微生物を含みかつ添加剤を含むとともにマトリックスが混合された複数のサンプルについての複数のマススペクトルデータの各々を教師データとして取得する処理と、機械学習を行うことにより、取得された複数の教師データに基づいて株を識別するための判別分析モデルを作成する処理と、株が未知の微生物を含みかつ添加剤を含むとともにマトリックスが混合されたサンプルについてのマススペクトルデータを対象データとして取得する処理と、作成された株ごとの判別分析モデルと取得された対象データとに基づいて当該対象データに対応する微生物の株を識別する処理とを、処理装置に実行させる。30

【0022】

この質量分析プログラムによれば、作成された判別分析モデルと対象データとに基づいて対象データに対応する微生物の株を高い精度で識別することができる。その結果、微生物の株の識別の精度が向上する。

【発明の効果】

【0023】

本発明によれば、微生物の株の識別の精度を向上させることができる。40

【図面の簡単な説明】

【0024】

【図1】本発明の一実施の形態に係る質量分析装置の構成を示す図である。

【図2】株識別装置の構成を示す図である。

【図3】図2の株識別装置により作成される判別分析モデルを説明するための図である。

【図4】株識別プログラムにより行われる株識別処理のアルゴリズムを示すフローチャートである。

【図5】サルモネラ菌のマススペクトルを示す図である。

【図6】複数のサンプルについての主成分分析の結果を示す図である。50

【図7】複数のサンプルについての主成分分析の結果を示す図である。

【図8】ホールドアウト検証における教師データおよび対象データの組み合わせを説明するための図である。

【図9】ホールドアウト検証による実施例および比較例における誤識別率を示す図である。

【図10】交差検証における教師データおよび対象データの組み合わせを説明するための図である。

【図11】交差検証による実施例および比較例における誤識別率の平均を示す図である。

【発明を実施するための形態】

【0025】

(1) 質量分析装置の構成

以下、本発明の実施の形態に係る質量分析装置、質量分析方法および株識別プログラムについて図面を参照しながら詳細に説明する。図1は、本発明の一実施の形態に係る質量分析装置の構成を示す図である。図1においては、主として質量分析装置100のハードウェアの構成が示される。図1に示すように、質量分析装置100は、処理装置10および分析部20を含む。

【0026】

処理装置10は、CPU(中央演算処理装置)11、RAM(ランダムアクセスメモリ)12、ROM(リードオンリーメモリ)13、記憶装置14、操作部15、表示部16および入出力I/F(インターフェイス)17により構成される。CPU11、RAM12、ROM13、記憶装置14、操作部15、表示部16および入出力I/F17はバス18に接続される。CPU11、RAM12およびROM13が株識別装置30を構成する。

【0027】

RAM12は、CPU11の作業領域として用いられる。ROM13にはシステムプログラムが記憶される。記憶装置14は、ハードディスクまたは半導体メモリ等の記憶媒体を含み、株識別プログラムを記憶する。CPU11が記憶装置14に記憶された株識別プログラムをRAM12上で実行することにより、後述する株識別処理が行われる。

【0028】

操作部15は、キーボード、マウスまたはタッチパネル等の入力デバイスである。使用者は、操作部15を操作することにより、分析部20または株識別装置30に所定の指示を与えることができる。表示部16は、液晶表示装置等の表示デバイスであり、株識別装置30による株識別の結果を表示する。入出力I/F17は、分析部20に接続される。

【0029】

分析部20は、MALDI(マトリックス支援レーザ脱離イオン化法)を用いて、微生物の種々のサンプルのマススペクトルを示すマススペクトルデータを生成する。サンプルは、株が既知であるサンプル(以下、教師サンプルと呼ぶ。)と、株が未知である識別対象のサンプル(以下、対象サンプルと呼ぶ。)とを含む。教師サンプルおよび対象サンプルの各々には、マトリックスが混合される。また、教師サンプルおよび対象サンプルの各々は、所定の添加剤を含む。

【0030】

マトリックスは、例えばシナピン酸を含む。添加剤は、アルカリ金属付加イオンの検出を抑制する化合物および界面活性剤の少なくとも一方を含む。具体的には、アルカリ金属付加イオンの検出を抑制する化合物は、メチレンジホスホン酸(MDPNA:methylene phosphonic acid)を含む。界面活性剤は、デシル- -D-マルトピラノシド(DMP:decyl-B-D-maltopyranoside)を含む。これらにより、生成されたマススペクトルデータにおけるピーク強度のばらつきを抑制することができる。

【0031】

株識別装置30は、複数の教師サンプルにそれぞれ対応する複数のマススペクトルデータに基づいて判別分析モデルを作成する。また、株識別装置30は、作成された判別分析

10

20

30

40

50

モデルに基づいて対象サンプルの株を識別する。以下、株識別装置30の動作について説明する。

【0032】

(2) 株識別装置

図2は、株識別装置30の構成を示す図である。図3は、図2の株識別装置30により作成される判別分析モデルを説明するための図である。図2に示すように、株識別装置30は、機能部として、教師データ取得部31、株情報取得部32、モデル作成部33、対象データ取得部34および識別部35を含む。図1のCPU11が記憶装置14に記憶された株識別プログラムを実行することにより、株識別装置30の機能部が実現される。株識別装置30の機能部の一部または全てが電子回路等のハードウェアにより実現されてもよい。

10

【0033】

教師データ取得部31は、分析部20により生成された複数の教師サンプルにそれぞれ対応する複数のマススペクトルデータ（以下、教師データと呼ぶ。）を取得する。使用者は、操作部15を操作することにより、分析部20に所望の複数の教師データを教師データ取得部31に与えるように指示することができる。図2の例では、教師データ取得部31は複数の教師データを分析部20から直接取得するが、本発明はこれに限定されない。分析部20により生成された複数の教師データが図1の記憶装置14に記憶されている場合には、教師データ取得部31は、複数の教師データを記憶装置14から取得してもよい。

20

【0034】

株情報取得部32は、教師データ取得部31により取得された複数の教師データにそれぞれ対応する複数の教師サンプルの各々の株を示す株情報を操作部15から取得する。使用者は、操作部15を操作することにより、複数の教師データにそれぞれ対応する複数の教師サンプルの各々の株情報を株情報取得部32に与えることができる。

30

【0035】

なお、使用者は、分析部20により教師データが生成される際に、当該教師データに対応する株情報を分析部20に登録してもよい。この場合、各教師データと、これに対応する株情報を紐付けた状態で一体的に取り扱うことが可能となる。したがって、教師データ取得部31により教師データが取得される際に、当該教師データに対応する株情報が株情報取得部32により自動的に分析部20または記憶装置14から取得される。

30

【0036】

モデル作成部33は、株情報取得部32により取得された株情報を基づいて、教師データ取得部31により取得された複数の教師データを株ごとに分類する。また、モデル作成部33は、同一の株に分類された複数の教師データを用い、機械学習（教師あり学習）を行うことにより当該株を識別するためのマススペクトルのパターンを判別分析モデルとして作成する。判別分析モデルは、サポートベクターマシン（SVM：support vector machine）またはニューラルネットワーク（NN：neural network）により作成されることが好ましい。

40

【0037】

図3(a)の左欄には、第1の株に分類された複数の教師データに基づく複数のマススペクトルが示される。図3(a)の右欄には、図3(a)左欄の複数の教師データを機械学習することにより作成された第1の株を識別するための判別分析モデルが示される。図3(b)の左欄には、第2の株に分類された複数の教師データに基づく複数のマススペクトルが示される。図3(b)の右欄には、図3(b)左欄の複数の教師データを機械学習することにより作成された第2の株を識別するための判別分析モデルが示される。

【0038】

図3の例では、判別分析モデルは連続波形を対象とするが、本発明はこれに限定されない。判別分析モデルは離散的なピークリスト（ピークの質量電荷比とピーク強度との組）を対象としてもよい。なお、理解を容易にするために、図3(a)における各マススペク

50

トルと図3(b)における各マススペクトルとは、明確に識別できるように脚色して異なるパターンで図示されている。しかしながら、実際には、一の株に対応するマススペクトルと他の株に対応するマススペクトルとは、類似したパターンを有し、識別することが困難であることが多い。

【0039】

対象データ取得部34は、分析部20により生成された対象サンプルに対応するマススペクトルデータ(以下、対象データと呼ぶ。)を取得する。使用者は、操作部15を操作することにより、分析部20に所望の対象データを対象データ取得部34に与えるように指示することができる。図2の例では、対象データ取得部34は対象データを分析部20から直接取得するが、本発明はこれに限定されない。分析部20により生成された対象データが記憶装置14に記憶されている場合には、対象データ取得部34は、対象データを記憶装置14から取得してもよい。

10

【0040】

識別部35は、モデル作成部33により作成された判別分析モデルと対象データ取得部34により取得された対象データとに基づいて、対象サンプルの株を識別する。具体的には、対象データに基づくマススペクトルと、複数の株にそれぞれ対応する複数の判別分析モデルの各々とのパターン認証が行われる。マススペクトルへの一致度が最も高い判別分析モデルに対応する株が、対象サンプルの株として識別される。識別部35は、識別された株を表示部16に表示させる。

20

【0041】

(3) 株識別処理

図4は、株識別プログラムにより行われる株識別処理のアルゴリズムを示すフローチャートである。以下、図2の株識別装置30および図4のフローチャートを用いて株識別処理を説明する。なお、以下の説明では、教師データおよび対象データは分析部20から取得されるが、記憶装置14から取得されてもよい。

【0042】

まず、教師データ取得部31は、分析部20から教師データを取得する(ステップS1)。本実施の形態においては、各教師データと、これに対応する株情報とが紐付けられた状態で分析部20に登録されている。そのため、ステップS1において、株情報取得部32は株情報を分析部20から取得する。

30

【0043】

次に、教師データ取得部31は、教師データの取得の終了が指示されたか否かを判定する(ステップS2)。使用者は、操作部15を操作することにより教師データの取得の終了を教師データ取得部31に指示することができる。教師データの取得の終了が指示されない場合、教師データ取得部31は、ステップS1に戻る。教師データの取得の終了が指示されるまでステップS1, S2が繰り返される。これにより、複数の教師データが取得される。

【0044】

教師データの取得の終了が指示された場合、モデル作成部33は、ステップS1で取得された教師データと株情報を基づいて、判別分析モデルを作成する(ステップS3)。なお、ステップS1で複数の株の各々について複数組の教師データと株情報を取得された場合には、モデル作成部33は判別分析モデルを株ごとに作成する。対象データ取得部34は、分析部20から対象データを取得する(ステップS4)。ステップS4は、ステップS3と同時に実行されてもよいし、ステップS4よりも前の時点に実行されてもよい。

40

【0045】

識別部35は、ステップS3で作成された判別分析モデルと、ステップS4で取得された対象データに基づくマススペクトルとのパターン認証を行う(ステップS5)。その後、識別部35は、ステップS3で作成された全部の判別分析モデルについてパターン認証が行われたか否かを判定する(ステップS6)。全部の判別分析モデルについてパターン

50

認証が行われていない場合、識別部 35 はステップ S5 に戻る。全部の判別分析モデルについてパターン認証が行われるまでステップ S5, S6 が繰り返される。

【0046】

全部の判別分析モデルについてパターン認証が行われた場合、識別部 35 は、ステップ S5 の認証結果に基づいて対象サンプルの株を識別する（ステップ S7）。最後に、識別部 35 は、ステップ S7 で識別された株を表示部 16 に表示させ（ステップ S8）、株識別処理を終了する。

【0047】

(4) 効果

本実施の形態に係る質量分析装置 100においては、株が既知の微生物に対応する複数のマススペクトルデータの各々が教師データとして教師データ取得部 31 により取得される。各教師データに対応するサンプルは添加剤を含むとともにマトリックスが混合されている。機械学習を行うことにより、教師データ取得部 31 により取得された複数の教師データに基づいて株を識別するための判別分析モデルがモデル作成部 33 により作成される。

10

【0048】

また、株が未知の微生物に対応するマススペクトルデータが対象データとして対象データ取得部 34 により取得される。対象データに対応するサンプルは添加剤を含むとともにマトリックスが混合されている。モデル作成部 33 により作成された株ごとの判別分析モデルと対象データ取得部 34 により取得された対象データとに基づいて当該対象データに対応する微生物の株が識別部 35 により識別される。

20

【0049】

この構成によれば、各教師データにおけるピーク強度のばらつきが抑制される。そのため、取得される複数の教師データを機械学習することにより、株の識別に用いることが可能な判別分析モデルを作成することができる。また、各教師データと同様に、対象データにおけるピーク強度のばらつきが抑制される。これにより、作成された判別分析モデルと対象データとに基づいて対象データに対応する微生物の株を高い精度で識別することができる。その結果、微生物の株の識別の精度が向上する。

【0050】

(5) 参考例

30

微生物の株の識別の手法として、非特許文献 1 に記載されたようにマーカピークを判定することが考えられる。図 5 は、サルモネラ菌のマススペクトルを示す図である。図 5 の横軸は質量電荷比を示し、縦軸はピーク強度を示す。遺伝子情報に基づいて図 5 のサルモネラ菌の株にのみ発現するタンパク質の理論質量を算出すると、質量電荷比 23000 の付近にマーカピークが存在することが予想される。

40

【0051】

しかしながら、図 5 に示すように、質量電荷比 23000 の付近には複数のピークが存在する。また、各ピーク強度は比較的低い。このように、ピーク強度が低い場合、またはマーカピークが他のピークと近接している場合には、マーカピークの有無を高い精度で安定的に判定することは困難である。

【0052】

微生物の株の識別の他の手法として、主成分分析が考えられる。具体的には、第 1 ~ 第 6 の株のいずれかに分類される微生物のサンプルが複数準備され、各サンプルについてマススペクトルが測定された。また、サンプルごとにピーク強度の列からなるベクトルが作成され、作成された複数のベクトルを入力とする主成分分析が行われた。なお、主成分分析の演算方法についてはよく知られているので、ここでは説明を省略する。

【0053】

図 6 および図 7 は、複数のサンプルについての主成分分析の結果を示す図である。図 6 の横軸は第 1 主成分を示し、縦軸は第 2 主成分を示す。図 7 の横軸は第 1 主成分を示し、縦軸は第 3 主成分を示す。第 1 ~ 第 3 の主成分は、複数のピーク強度の線形結合量である

50

。また、図 6 および図 7 においては、同一の株に分類される微生物のサンプルについての主成分分析の結果が同一の指標により示されるように、複数の指標「○」、「」、「」、「×」、「+」および「・」がプロットされる。

【0054】

図 6 および図 7 に示すように、同一の株に対応する指標はクラスタを形成する傾向がある。しかしながら、同一の指標により形成されるクラスタは、複数の領域に分離して存在する。また、一の指標により形成されるクラスタと他の指標により形成されるクラスタとは互いに重複する。そのため、主成分分析のようにピーク強度の線形結合量等を評価関数とする単純な識別方法によっては、微生物の株を高い精度で識別するのは困難であることが示唆される。

10

【0055】

(6) 実施例

以下の実施例では、上記実施の形態に基づいて SVM により作成された判別分析モデルを用いてサンプルの株が識別された。一方で、比較例では、一般的な線形判別法により作成された線形モデルを用いてサンプルの株が識別された。実施例および比較例の各々における誤識別率がホールドアウト検証および交差検証 (cross validation) の各々により評価された。以下に詳細を説明する。

10

【0056】

(a) ホールドアウト検証

2 日間にわたって、株が既知である 205 個のサンプルの各々についてのマススペクトルデータ（以下、単にデータと呼ぶ。）が生成された。具体的には、第 1 日目に 107 個のデータが生成され、第 2 日目に 98 個のデータが生成された。生成された 205 個のデータの一部または全部を用いて、教師データと対象データとの組み合わせが複数定義された。

20

【0057】

図 8 は、ホールドアウト検証における教師データおよび対象データの組み合わせを説明するための図である。図 8 に示すように、第 1 の組み合わせにおいては、同日に生成された 49 個のデータが教師データとされ、教師データと同日に生成された他の 49 個のデータが対象データとされた。第 2 の組み合わせにおいては、第 1 日目に生成された 107 個のデータが教師データとされ、第 2 日目に生成された 98 個のデータが対象データとされた。第 3 の組み合わせにおいては、2 日間にわたって生成された 102 個のデータが教師データとされ、他の 103 個のデータが対象データとされた。

30

【0058】

実施例では、第 1 の組み合わせの教師データを用いて SVM により作成された判別分析モデルに基づいて第 1 の組み合わせの各対象データの株が識別された。同様に、第 2 の組み合わせの教師データを用いて SVM により作成された判別分析モデルに基づいて第 2 の組み合わせの各対象データの株が識別された。第 3 の組み合わせの教師データを用いて SVM により作成された判別分析モデルに基づいて第 3 の組み合わせの各対象データの株が識別された。

30

【0059】

なお、上記の 205 個のデータの生成の際には、全部のサンプルにマトリックスが混合され、添加剤が配合された。サンプルにマトリックスが混合されない場合、または添加剤が配合されない場合には、データにおけるノイズ成分が増加するとともに、ピーク強度のばらつきが大きくなり、判別分析モデルを作成することができなかった。

40

【0060】

比較例では、第 1 の組み合わせの教師データを用いて線形判別法により作成された線形モデルに基づいて第 1 の組み合わせの各対象データの株が識別された。同様に、第 2 の組み合わせの教師データを用いて線形判別法により作成された線形モデルに基づいて第 2 の組み合わせの各対象データの株が識別された。第 3 の組み合わせの教師データを用いて線形判別法により作成された線形モデルに基づいて第 3 の組み合わせの各対象データの株が

50

識別された。

【0061】

また、実施例および比較例における誤識別率がホールドアウト検証により評価された。図9は、ホールドアウト検証による実施例および比較例における誤識別率を示す図である。図9(a)に示すように、実施例における第1～第3の組み合わせに対応する誤識別率は、それぞれ12%、5%および3%となった。図9(b)に示すように、比較例における第1～第3の組み合わせに対応する誤識別率は、それぞれ12%、44%および27%となった。ホールドアウト検証による実施例と比較例との比較の結果、SVMにより作成された判別分析モデルを用いることにより、高い精度で株を識別することが可能であることが確認された。

10

【0062】

(b) 交差検証

図10は、交差検証における教師データおよび対象データの組み合わせを説明するための図である。図10に示すように、第4の組み合わせにおいては、第1の組み合わせの教師データのうち、無作為に選択された1/10のデータが対象データとされ、他のデータが教師データとされた。第5の組み合わせにおいては、第2の組み合わせの教師データのうち、無作為に選択された1/10のデータが対象データとされ、他のデータが教師データとされた。第6の組み合わせにおいては、第3の組み合わせの教師データのうち、無作為に選択された1/10のデータが対象データとされ、他のデータが教師データとされた。

20

【0063】

交差検証においては、上記の教師データの無作為の選択は複数回繰り返される。したがって、選択が行われるごとに対象データが変化するとともに、教師データが変化する。

【0064】

実施例では、第4の組み合わせの教師データが選択されるごとに、当該教師データを用いてSVMにより作成された判別分析モデルに基づいて第4の組み合わせの各対象データの株が識別された。同様に、第5の組み合わせの教師データが選択されるごとに、当該教師データを用いてSVMにより作成された判別分析モデルに基づいて第5の組み合わせの各対象データの株が識別された。第6の組み合わせの教師データが選択されるごとに、当該教師データを用いてSVMにより作成された判別分析モデルに基づいて第6の組み合わせの各対象データの株が識別された。

30

【0065】

比較例では、第4の組み合わせの教師データが選択されるごとに、当該教師データを用いて線形判別法により作成された線形モデルに基づいて第4の組み合わせの各対象データの株が識別された。同様に、第5の組み合わせの教師データが選択されるごとに、当該教師データを用いて線形判別法により作成された線形モデルに基づいて第5の組み合わせの各対象データの株が識別された。第6の組み合わせの教師データが選択されるごとに、当該教師データを用いて線形判別法により作成された線形モデルに基づいて第6の組み合わせの各対象データの株が識別された。

40

【0066】

また、実施例および比較例における誤識別率の平均が交差検証により評価された。図11は、交差検証による実施例および比較例における誤識別率の平均を示す図である。図11(a)に示すように、実施例における第4～第6の組み合わせに対応する誤識別率の平均は、それぞれ0%、1%および1%となった。図11(b)に示すように、比較例における第4～第6の組み合わせに対応する誤識別率の平均は、それぞれ61%、35%および49%となった。交差検証による実施例と比較例との比較の結果、SVMにより作成された判別分析モデルを用いることにより、高い精度で株を識別することが可能であることが確認された。

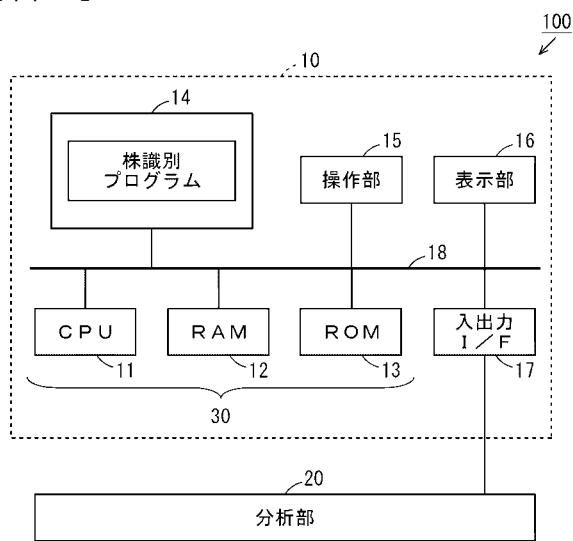
【符号の説明】

【0067】

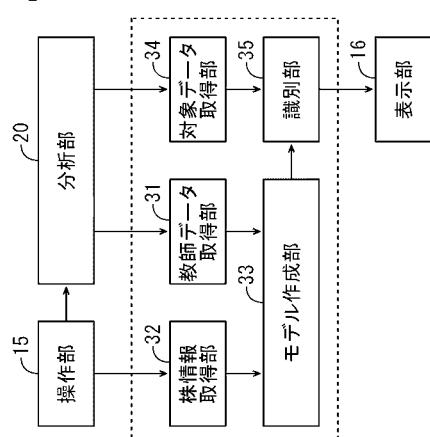
50

10 ... 処理装置 , 11 ... C P U , 12 ... R A M , 13 ... R O M , 14 ... 記憶装置 , 15 ... 操作部 , 16 ... 表示部 , 17 ... 入出力 I / F , 18 ... バス , 20 ... 分析部 , 30 ... 株識別装置 , 31 ... 教師データ取得部 , 32 ... 株情報取得部 , 33 ... モデル作成部 , 34 ... 対象データ取得部 , 35 ... 識別部 , 100 ... 質量分析装置

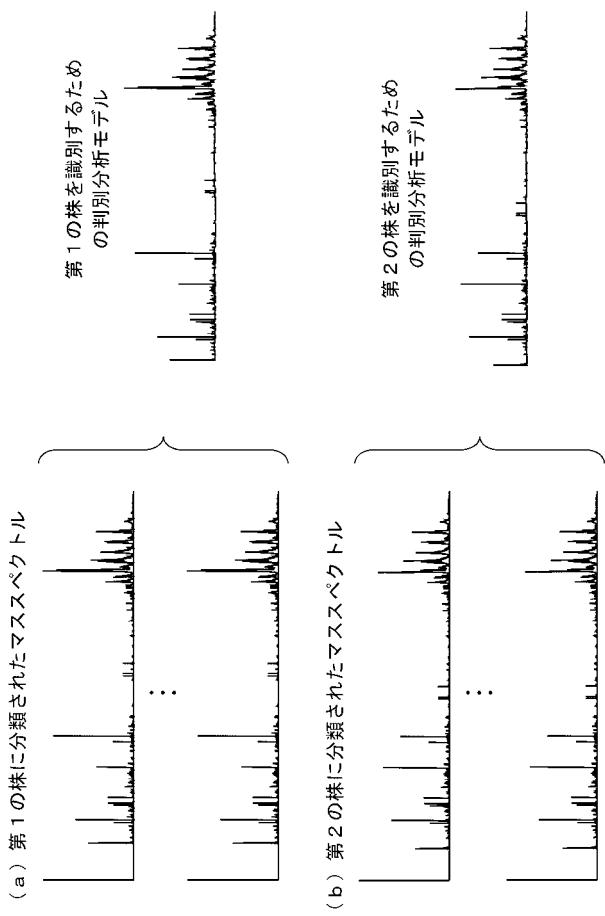
【図 1】



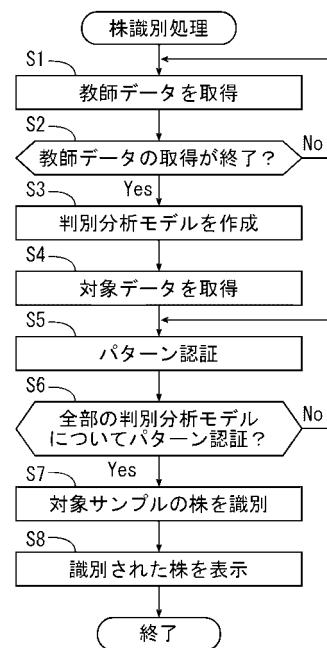
【図 2】



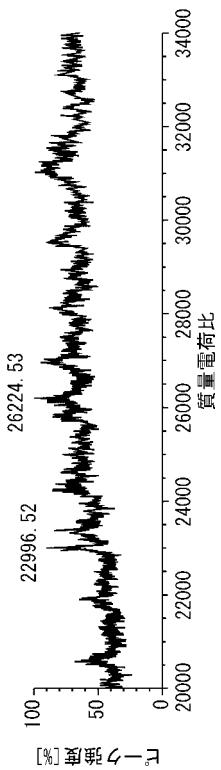
【図3】



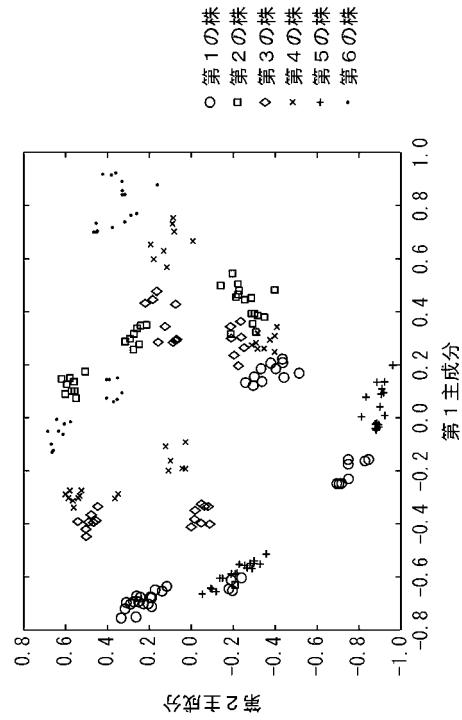
【図4】



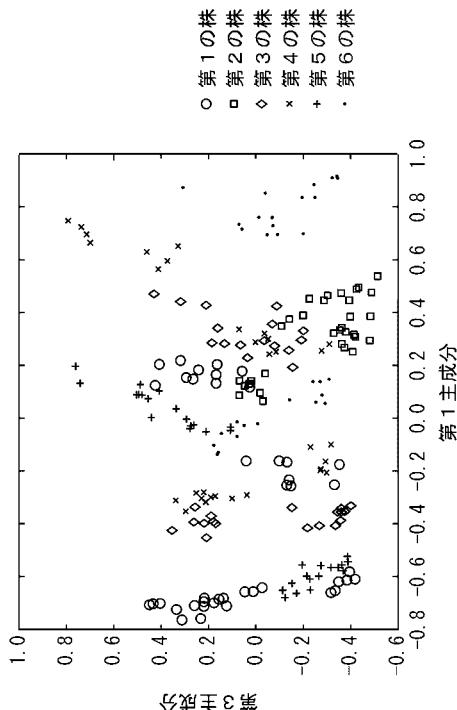
【図5】



【図6】



【図7】



【図8】

	教師データ	対象データ
第1の組み合わせ	同日に生成された49個のデータ	標準データと同日に生成された他の49個のデータ
第2の組み合わせ	第1日目に生成された107個のデータ	第2日目に生成された98個のデータ
第3の組み合わせ	2日間にわたって生成された102個のデータ	2日間にわたって生成された他の103個のデータ

【図9】

(a) 実施例

SVM	誤識別率
第1の組み合わせ	12%
第2の組み合わせ	5%
第3の組み合わせ	3%

(b) 比較例

線形判別	誤識別率
第1の組み合わせ	12%
第2の組み合わせ	44%
第3の組み合わせ	27%

【図10】

	教師データ	対象データ
第4の組み合わせ	第1の組み合わせの標準データのうち右記の対象データを除いたデータ	第1の組み合わせの標準データのうち無作為に選択された1/10のデータ
第5の組み合わせ	第2の組み合わせの標準データのうち右記の対象データを除いたデータ	第2の組み合わせの標準データのうち無作為に選択された1/10のデータ
第6の組み合わせ	第3の組み合わせの標準データのうち右記の対象データを除いたデータ	第3の組み合わせの標準データのうち無作為に選択された1/10のデータ

【図11】

(a) 実施例

SVM	誤識別率(平均)
第4の組み合わせ	0%
第5の組み合わせ	1%
第6の組み合わせ	1%

(b) 比較例

線形判別	誤識別率(平均)
第4の組み合わせ	61%
第5の組み合わせ	35%
第6の組み合わせ	49%

フロントページの続き

(72)発明者 山田 賢志
京都府京都市中京区西ノ京桑原町1番地 株式会社島津製作所内

(72)発明者 福山 裕子
京都府京都市中京区西ノ京桑原町1番地 株式会社島津製作所内

(72)発明者 田村 廣人
愛知県名古屋市天白区塩釜口1-501 学校法人名城大学内

F ターム(参考) 2G041 CA01 DA04 FA10 JA07 JA08 LA06 LA07
4B029 AA07 BB01 FA03
4B063 QA01 QA18 QQ05 QS10 QS12 QS39
5C038 HH28