(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 118284802 A
(43)申请公布日 2024.07.02

(21)申请号 202280077216.0
(22)申请日 2022.10.14
(30)优先权数据
2021-189571 2021.11.22 JP
(85)PCT国际申请进入国家阶段日
2024.05.21
(86)PCT国际申请的申请数据
PCT/JP2022/038480 2022.10.14
(87)PCT国际申请的公布数据
WO2023/090015 JA 2023.05.25
(71)申请人 富士胶片株式会社
地址 日本
(72)发明人 杉田惟
(74)专利代理机构 永新专利商标代理有限公司 72002
专利代理师 徐殿军
(51)Int.Cl.
G01N 21/65(2006.01)
C12M 1/34(2006.01)
C12M 3/00(2006.01)
G01N 33/483(2006.01)
G06N 3/02(2006.01)
G06N 20/00(2006.01)

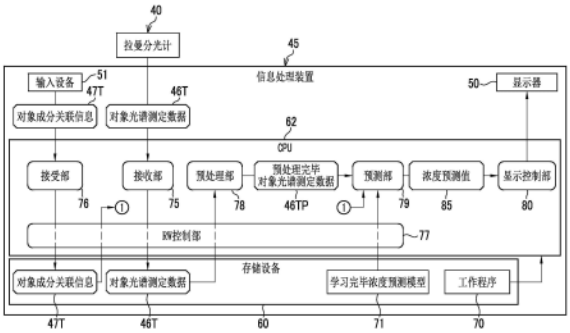
权利要求书4页 说明书16页 附图15页

(54)发明名称

信息处理装置、信息处理装置的工作方法、信息处理装置的工作程序、校准完毕状态预测模型的生成方法以及校准完毕状态预测模型

(57)摘要

一种信息处理装置,其根据测定从生物分子作为成分分散在液体中而成的悬浊液发出的电磁波的光谱而得到的光谱测定数据,预测悬浊液中的成分的状态,其中,所述信息处理装置具备处理器,处理器进行如下处理:使用校准完毕状态预测模型,所述校准完毕状态预测模型是使用第1校准数据和第2校准数据中至少2种校准数据进行了校准的校准完毕状态预测模型,第1校准数据将作为从包含第1成分的第1悬浊液获得的光谱测定数据的第1光谱测定数据及与第1成分有关的第1成分关联信息设为解释变量,并且,将与第1成分的状态有关的第1状态关联信息设为目标变量,第2校准数据将作为从包含第2成分的第2悬浊液获得的光谱测定数据的第2光谱测定数据及与第2成分有关的第2成分关联信息设为解释变量,并且,将与第2成分的状态有关的第2状态关联信息设为目标变量;获取与作为状态的预测对象的对象成分有关的对象成分关联信息、及作为从包含对象成分的对象悬浊液获得的光谱测定数据的对象光谱测定数据;将对象成分关联信息及对象光谱测定数据适用于校准完毕状态预测模型,从校准完毕状态预测模型输出预测出对象悬浊液中的对象成分的状态的对象状态预测结果。



CN 118284802 A

1.一种信息处理装置,其根据测定从生物分子作为成分分散在液体中而成的悬浊液发出的电磁波的光谱而得到的光谱测定数据,预测所述悬浊液中的成分的状态,其中,

所述信息处理装置具备处理器,

所述处理器进行如下处理:

使用校准完毕状态预测模型,所述校准完毕状态预测模型是使用第1校准数据和第2校准数据中至少2种校准数据进行了校准的校准完毕状态预测模型,

所述第1校准数据将作为从包含第1成分的第1悬浊液获得的所述光谱测定数据的第1光谱测定数据及与所述第1成分有关的第1成分关联信息设为解释变量,并且,将与所述第1成分的状态有关的第1状态关联信息设为目标变量,

所述第2校准数据将作为从包含第2成分的第2悬浊液获得的所述光谱测定数据的第2光谱测定数据及与所述第2成分有关的第2成分关联信息设为解释变量,并且,将与所述第2成分的状态有关的第2状态关联信息设为目标变量;

获取与作为所述状态的预测对象的对象成分有关的对象成分关联信息、及作为从包含所述对象成分的对象悬浊液获得的所述光谱测定数据的对象光谱测定数据;

将所述对象成分关联信息及所述对象光谱测定数据适用于所述校准完毕状态预测模型,从所述校准完毕状态预测模型输出预测出所述对象悬浊液中的所述对象成分的状态的对象状态预测结果。

2.根据权利要求1所述的信息处理装置,其中,

所述校准完毕状态预测模型具备:

第1模型,根据所述对象光谱测定数据,输出所述对象成分的状态的临时预测结果;及

第2模型,根据所述对象成分关联信息及所述临时预测结果,输出所述对象状态预测结果。

3.根据权利要求1或2所述的信息处理装置,其中,

所述对象成分与为了获得所述校准数据而使用的成分不同。

4.根据权利要求1至3中任一项所述的信息处理装置,其中,

所述处理器进行如下处理:

在对所述对象光谱测定数据实施以噪声去除、峰值分离及峰值强调中的至少任一个为目的的预处理之后,适用于所述校准完毕状态预测模型。

5.根据权利要求1至4中任一项所述的信息处理装置,其中,

所述第1成分、所述第2成分及所述对象成分为蛋白质。

6.根据权利要求5所述的信息处理装置,其中,

所述第1成分关联信息、所述第2成分关联信息及所述对象成分关联信息包含所述蛋白质的氨基酸的组成比的信息。

7.根据权利要求5或6所述的信息处理装置,其中,

所述蛋白质为抗体。

8.根据权利要求7所述的信息处理装置,其中,

所述第1成分关联信息、所述第2成分关联信息及所述对象成分关联信息包含所述抗体的亚类的信息。

9.根据权利要求1至8中任一项所述的信息处理装置,其中,

所述光谱为拉曼光谱。

10.根据权利要求1至9中任一项所述的信息处理装置,其中,

所述校准完毕状态预测模型是使用所述校准数据作为教师数据而进行了学习的机器学习模型。

11.根据权利要求1至10中任一项所述的信息处理装置,其中,

所述状态为浓度,

所述第1状态关联信息为所述第1成分的浓度的测定值,

所述第2状态关联信息为所述第2成分的浓度的测定值,

所述对象状态预测结果为所述对象成分的浓度的预测值。

12.一种信息处理装置的工作方法,所述信息处理装置根据测定从生物分子作为成分分散在液体中而成的悬浊液发出的电磁波的光谱而得到的光谱测定数据,预测所述悬浊液中的成分的状态,其中,

所述信息处理装置的工作方法包括如下:

使用校准完毕状态预测模型,所述校准完毕状态预测模型是使用第1校准数据和第2校准数据中至少2种校准数据进行了校准的校准完毕状态预测模型,

所述第1校准数据将作为从包含第1成分的第1悬浊液获得的所述光谱测定数据的第1光谱测定数据及与所述第1成分有关的第1成分关联信息设为解释变量,并且,将与所述第1成分的状态有关的第1状态关联信息设为目标变量,

所述第2校准数据将作为从包含第2成分的第2悬浊液获得的所述光谱测定数据的第2光谱测定数据及与所述第2成分有关的第2成分关联信息设为解释变量,并且,将与所述第2成分的状态有关的第2状态关联信息设为目标变量;

获取与作为所述状态的预测对象的对象成分有关的对象成分关联信息、及作为从包含所述对象成分的对象悬浊液获得的所述光谱测定数据的对象光谱测定数据;以及

将所述对象成分关联信息及所述对象光谱测定数据适用于所述校准完毕状态预测模型,从所述校准完毕状态预测模型输出预测出所述对象悬浊液中的所述对象成分的状态的对象状态预测结果。

13.一种信息处理装置的工作程序,所述信息处理装置根据测定从生物分子作为成分分散在液体中而成的悬浊液发出的电磁波的光谱而得到的光谱测定数据,预测所述悬浊液中的成分的状态,其中,

所述信息处理装置的工作程序用于使计算机执行包括如下的处理:

使用校准完毕状态预测模型,所述校准完毕状态预测模型是使用第1校准数据和第2校准数据中至少2种校准数据进行了校准的校准完毕状态预测模型,

所述第1校准数据将作为从包含第1成分的第1悬浊液获得的所述光谱测定数据的第1光谱测定数据及与所述第1成分有关的第1成分关联信息设为解释变量,并且,将与所述第1成分的状态有关的第1状态关联信息设为目标变量,

所述第2校准数据将作为从包含第2成分的第2悬浊液获得的所述光谱测定数据的第2光谱测定数据及与所述第2成分有关的第2成分关联信息设为解释变量,并且,将与所述第2成分的状态有关的第2状态关联信息设为目标变量;

获取与作为所述状态的预测对象的对象成分有关的对象成分关联信息、及作为从包含

所述对象成分的对象悬浊液获得的所述光谱测定数据的对象光谱测定数据;以及

将所述对象成分关联信息及所述对象光谱测定数据适用于所述校准完毕状态预测模型,从所述校准完毕状态预测模型输出预测出所述对象悬浊液中的所述对象成分的状态的对象状态预测结果。

14.一种校准完毕状态预测模型的生成方法,所述校准完毕状态预测模型根据测定从生物分子作为成分分散在液体中而成的悬浊液发出的电磁波的光谱而得到的光谱测定数据,预测所述悬浊液中的成分的状态,其中,

所述校准完毕状态预测模型的生成方法包括如下:

获取校准数据,所述校准数据是第1校准数据和第2校准数据中至少2种校准数据,

所述第1校准数据将作为从包含第1成分的第1悬浊液获得的所述光谱测定数据的第1光谱测定数据及与所述第1成分有关的第1成分关联信息设为解释变量,并且,将与所述第1成分的状态有关的第1状态关联信息设为目标变量,

所述第2校准数据将作为从包含第2成分的第2悬浊液获得的所述光谱测定数据的第2光谱测定数据及与所述第2成分有关的第2成分关联信息设为解释变量,并且,将与所述第2成分的状态有关的第2状态关联信息设为目标变量;以及

使用所述校准数据生成所述校准完毕状态预测模型。

15.根据权利要求14所述的校准完毕状态预测模型的生成方法,其中,

所述校准完毕状态预测模型的生成方法包括如下:

将所述校准数据的所述解释变量作为学习用输入数据输入到机器学习模型中,并从所述机器学习模型输出预测出所述状态的学习用状态预测结果;以及

根据所述学习用状态预测结果与所述校准数据的所述目标变量的比较结果,更新所述机器学习模型,

通过一边变更所述校准数据一边重复所述解释变量向所述机器学习模型的输入、来自所述机器学习模型的所述学习用状态预测结果的输出、及所述机器学习模型的更新,从而将所述机器学习模型设为所述校准完毕状态预测模型。

16.一种校准完毕状态预测模型,该校准完毕状态预测模型根据测定从生物分子作为成分分散在液体中而成的悬浊液发出的电磁波的光谱而得到的光谱测定数据,预测所述悬浊液中的成分的状态,其中,

所述校准完毕状态预测模型使用校准数据而生成,所述校准数据是第1校准数据和第2校准数据中至少2种校准数据,

所述第1校准数据将作为从包含第1成分的第1悬浊液获得的所述光谱测定数据的第1光谱测定数据及与所述第1成分有关的第1成分关联信息设为解释变量,并且,将与所述第1成分的状态有关的第1状态关联信息设为目标变量,

所述第2校准数据将作为从包含第2成分的第2悬浊液获得的所述光谱测定数据的第2光谱测定数据及与所述第2成分有关的第2成分关联信息设为解释变量,并且,将与所述第2成分的状态有关的第2状态关联信息设为目标变量;

所述校准完毕状态预测模型用于使计算机执行如下功能:

在被适用了与作为所述状态的预测对象的对象成分有关的对象成分关联信息、及作为从包含所述对象成分的对象悬浊液获得的所述光谱测定数据的对象光谱测定数据的情况

下,输出预测出所述对象悬浊液中的所述对象成分的状态的对象状态预测结果。

17.一种信息处理装置,其存储了权利要求16所述的校准完毕状态预测模型。

信息处理装置、信息处理装置的工作方法、信息处理装置的工作程序、校准完毕状态预测模型的生成方法以及校准完毕状态预测模型

技术领域

[0001] 本发明的技术涉及一种信息处理装置、信息处理装置的工作方法、信息处理装置的工作程序、校准完毕状态预测模型的生成方法以及校准完毕状态预测模型。

背景技术

[0002] 例如,已知有以单克隆抗体等蛋白质之类的生物分子为有效成分的生物医药品的制造工艺。在这种制造工艺中,常常产生以有效成分为代表的各种成分分散在液体中而成的悬浊液。为了成功地引导进行中的制造工艺,在生产线内监视该悬浊液中的对象成分(例如蛋白质或者源自蛋白质的掺杂物)的状态是重要的。

[0003] 在日本特开2016-128822号公报中,记载有在生产线内预测作为对象成分的状态的浓度的技术。具体而言,在日本特开2016-128822号公报中,在生产线内测定悬浊液的拉曼光谱,并使用线性模型根据拉曼光谱预测对象成分的浓度。

发明内容

[0004] 发明要解决的技术课题

[0005] 日本特开2016-128822号公报中记载的线性模型例如是专用于某一个对象成分A的专用模型。因此,在预测与对象成分A不同的对象成分B的浓度时,需要新生成预测对象成分B的浓度的专用模型。在以不同的抗体为有效成分的各种生物医药品的制造中,需要按每个对象成分生成专用模型,效率非常差。

[0006] 本发明的技术所涉及的1个实施方式提供一种能够有效地预测生物分子作为成分分散在液体中而成的悬浊液中的对象成分的状态的信息处理装置、信息处理装置的工作方法、信息处理装置的工作程序、校准完毕状态预测模型的生成方法、以及校准完毕状态预测模型。

[0007] 用于解决技术课题的手段

[0008] 本发明的信息处理装置是一种根据测定从生物分子作为成分分散在液体中而成的悬浊液发出的电磁波的光谱而得到的光谱测定数据来预测悬浊液中的成分的状态的信息处理装置,其中,所述信息处理装置具备处理器,处理器进行如下处理:使用校准完毕状态预测模型,所述校准完毕状态预测模型是使用第1校准数据和第2校准数据中至少2种校准数据进行了校准的校准完毕状态预测模型,第1校准数据将作为从包含第1成分的第1悬浊液获得的光谱测定数据的第1光谱测定数据及与第1成分有关的第1成分关联信息设为解释变量,并且,将与第1成分的状态有关的第1状态关联信息设为目标变量,第2校准数据将作为从包含第2成分的第2悬浊液获得的光谱测定数据的第2光谱测定数据及与第2成分有关的第2成分关联信息设为解释变量,并且,将与第2成分的状态有关的第2状态关联信息设为目标变量;获取与作为状态的预测对象的对象成分有关的对象成分关联信息、及作为从

包含对象成分的对象悬浊液获得的光谱测定数据的对象光谱测定数据;将对象成分关联信息及对象光谱测定数据适用于校准完毕状态预测模型,从校准完毕状态预测模型输出预测出对象悬浊液中的对象成分的状态的对象状态预测结果。

[0009] 校准完毕状态预测模型优选具备:第1模型,根据对象光谱测定数据,输出对象成分的状态的临时预测结果;及第2模型,根据对象成分关联信息及临时预测结果,输出对象状态预测结果。

[0010] 对象成分优选与为了获得校准数据而使用的成分不同。

[0011] 处理器优选在对对象光谱测定数据实施以噪声去除、峰值分离及峰值强调中的至少任一个为目的的预处理之后,适用于校准完毕状态预测模型。

[0012] 第1成分、第2成分及对象成分优选为蛋白质。

[0013] 第1成分关联信息、第2成分关联信息及对象成分关联信息优选包含蛋白质的氨基酸的组成比的信息。

[0014] 蛋白质优选为抗体。

[0015] 第1成分关联信息、第2成分关联信息及对象成分关联信息优选包含抗体的亚类的信息。

[0016] 光谱优选为拉曼光谱。

[0017] 校准完毕状态预测模型优选为使用校准数据作为教师数据(teaching data)而进行了学习的机器学习模型。

[0018] 优选状态为浓度,第1状态关联信息为第1成分的浓度的测定值,第2状态关联信息为第2成分的浓度的测定值,对象状态预测结果为对象成分的浓度的预测值。

[0019] 本发明的信息处理装置的工作方法是一种根据测定从生物分子作为成分分散在液体中而成的悬浊液发出的电磁波的光谱而得到的光谱测定数据来预测悬浊液中的成分的状态的信息处理装置的工作方法,其中,所述信息处理装置的工作方法包括如下:使用校准完毕状态预测模型,所述校准完毕状态预测模型是使用第1校准数据和第2校准数据中至少2种校准数据进行了校准的校准完毕状态预测模型,第1校准数据将作为从包含第1成分的第1悬浊液获得的光谱测定数据的第1光谱测定数据及与第1成分有关的第1成分关联信息设为解释变量,并且,将与第1成分的状态有关的第1状态关联信息设为目标变量,第2校准数据将作为从包含第2成分的第2悬浊液获得的光谱测定数据的第2光谱测定数据及与第2成分有关的第2成分关联信息设为解释变量,并且,将与第2成分的状态有关的第2状态关联信息设为目标变量;获取与作为状态的预测对象的对象成分有关的对象成分关联信息、及作为从包含对象成分的对象悬浊液获得的光谱测定数据的对象光谱测定数据;以及将对象成分关联信息及对象光谱测定数据适用于校准完毕状态预测模型,从校准完毕状态预测模型输出预测出对象悬浊液中的对象成分的状态的对象状态预测结果。

[0020] 本发明的信息处理装置的工作程序是一种根据测定从生物分子作为成分分散在液体中而成的悬浊液发出的电磁波的光谱而得到的光谱测定数据来预测悬浊液中的成分的状态的信息处理装置的工作程序,其中,所述信息处理装置的工作程序使计算机执行包括如下的处理:使用校准完毕状态预测模型,所述校准完毕状态预测模型是使用第1校准数据和第2校准数据中至少2种校准数据进行了校准的校准完毕状态预测模型,第1校准数据将作为从包含第1成分的第1悬浊液获得的光谱测定数据的第1光谱测定数据及与第1成分

有关的第1成分关联信息设为解释变量,并且,将与第1成分的状态有关的第1状态关联信息设为目标变量,第2校准数据将作为从包含第2成分的第2悬浊液获得的光谱测定数据的第2光谱测定数据及与第2成分有关的第2成分关联信息设为解释变量,并且,将与第2成分的状态有关的第2状态关联信息设为目标变量;获取与作为状态的预测对象的对象成分有关的对象成分关联信息、及作为从包含对象成分的对象悬浊液获得的光谱测定数据的对象光谱测定数据;以及将对象成分关联信息及对象光谱测定数据适用于校准完毕状态预测模型,从校准完毕状态预测模型输出预测出对象悬浊液中的对象成分的状态的对象状态预测结果。

[0021] 本发明的校准完毕状态预测模型的生成方法是一种根据测定从生物分子作为成分分散在液体中而成的悬浊液发出的电磁波的光谱而得到的光谱测定数据来预测悬浊液中的成分的状态的校准完毕状态预测模型的生成方法,其中,所述校准完毕状态预测模型的生成方法包括如下:获取校准数据,所述校准数据是第1校准数据和第2校准数据中至少2种校准数据,第1校准数据将作为从包含第1成分的第1悬浊液获得的光谱测定数据的第1光谱测定数据及与第1成分有关的第1成分关联信息设为解释变量,并且,将与第1成分的状态有关的第1状态关联信息设为目标变量,第2校准数据将作为从包含第2成分的第2悬浊液获得的光谱测定数据的第2光谱测定数据及与第2成分有关的第2成分关联信息设为解释变量,并且,将与第2成分的状态有关的第2状态关联信息设为目标变量;以及使用校准数据生成校准完毕状态预测模型。

[0022] 优选包括如下:将校准数据的解释变量作为学习用输入数据输入到机器学习模型中,并从机器学习模型输出预测出状态的学习用状态预测结果;以及根据学习用状态预测结果与校准数据的目标变量的比较结果,更新机器学习模型,通过一边变更校准数据一边重复向机器学习模型的解释变量的输入、来自机器学习模型的学习用状态预测结果的输出、及机器学习模型的更新,从而将机器学习模型设为校准完毕状态预测模型。

[0023] 本发明的校准完毕状态预测模型是一种根据测定从生物分子作为成分分散在液体中而成的悬浊液发出的电磁波的光谱而得到的光谱测定数据来预测悬浊液中的成分的状态的校准完毕状态预测模型,其中,所述校准完毕状态预测模型使计算机执行如下功能:使用校准数据生成,所述校准数据是第1校准数据和第2校准数据中至少2种校准数据,第1校准数据将作为从包含第1成分的第1悬浊液获得的光谱测定数据的第1光谱测定数据及与第1成分有关的第1成分关联信息设为解释变量,并且,将与第1成分的状态有关的第1状态关联信息设为目标变量,第2校准数据将作为从包含第2成分的第2悬浊液获得的光谱测定数据的第2光谱测定数据及与第2成分有关的第2成分关联信息设为解释变量,并且,将与第2成分的状态有关的第2状态关联信息设为目标变量;在适用了与作为状态的预测对象的对象成分有关的对象成分关联信息、及作为从包含对象成分的对象悬浊液获得的光谱测定数据的对象光谱测定数据的情况下,输出预测出对象悬浊液中的对象成分的状态的对象状态预测结果。

[0024] 本发明的信息处理装置是存储了上面记载的校准完毕状态预测模型的信息处理装置。

[0025] 发明效果

[0026] 根据本发明的技术,能够提供一种能够有效地预测生物分子作为成分分散在液体

中而成的悬浊液中的对象成分的状态的信息处理装置、信息处理装置的工作方法、信息处理装置的工作程序、校准完毕状态预测模型的生成方法、以及校准完毕状态预测模型。

附图说明

[0027] 图1是表示生物医药品的制造工艺的概要的图。

[0028] 图2是表示通过信息处理装置获取用拉曼分光计测定对象第1纯化液的拉曼光谱而获得的对象光谱测定数据及对象第1纯化液内的对象抗体的对象成分关联信息的情况的图。

[0029] 图3是表示对象光谱测定数据的图。

[0030] 图4是表示对象成分关联信息的图。

[0031] 图5是构成信息处理装置的计算机的框图。

[0032] 图6是构成信息处理装置的计算机的CPU的框图。

[0033] 图7是表示预处理部进行的处理的概要的图。

[0034] 图8是表示作为峰值强调处理的稀疏化处理的图。

[0035] 图9是表示预测部进行的处理的概要的图。

[0036] 图10是表示构成第1模型及第2模型的神经网络的图。

[0037] 图11是表示教师数据组的结构的图。

[0038] 图12是表示教师数据组的图。

[0039] 图13是表示对教师数据的光谱测定数据实施预处理而设为预处理完毕光谱测定数据的情况的图。

[0040] 图14是表示第1模型的学习阶段中的处理的概要的图。

[0041] 图15是表示第2模型的学习阶段中的处理的概要的图。

[0042] 图16是表示获取教师数据并存储于存储设备的顺序的流程图。

[0043] 图17是表示第1模型的学习阶段中的处理顺序的流程图。

[0044] 图18是表示第2模型的学习阶段中的处理顺序的流程图。

[0045] 图19是表示信息处理装置的处理顺序的流程图。

具体实施方式

[0046] 作为一例,如图1所示,适用了本发明的技术所涉及的信息处理装置45(参考图2)的生物医药品的制造工艺2大致区分为第1工艺10、第2工艺11及第3工艺12。第1工艺10是将抗体基因14编入中国仓鼠卵巢细胞(CHO(Chinese Hamster Ovary cells)细胞)等宿主细胞13并建立抗体生产细胞15的工艺。第2工艺是在培养槽16中对抗体生产细胞15进行细胞培养的工艺。

[0047] 第3工艺12是从培养上清液17中纯化生物医药品的原料药18的工艺。培养上清液17是从结束了第2工艺11的培养槽16内的培养液中去除细胞而获得的溶液。在培养上清液17中分散有抗体生产细胞15所生产的免疫球蛋白、即抗体19。抗体19例如为单克隆抗体,成为生物医药品的有效成分。并且,在培养上清液17中,除了抗体19以外,还分散有源自细胞的蛋白质·源自细胞的DNA(Deoxyribonucleic Acid:脱氧核糖核酸)20、及抗体19的凝聚物21等掺杂物、或者病毒22等。抗体19是本发明的技术所涉及的“生物分子”及“蛋白质”的

一例。另外,“生物分子”是指从细胞、细胞器官、细胞分子、基因重组体、源自天然的非合成化学物质的生物等获得的物质。

[0048] 第3工艺12中,使用免疫亲和层析装置25、阳离子层析装置26及阴离子层析装置27。在免疫亲和层析装置25中导入有培养上清液17。免疫亲和层析装置25通过使用将与抗体19具有亲和性的蛋白A等配体固定在载体的柱从培养上清液17中提取抗体19而生成第1纯化液28。虽然省略了图示,但对第1纯化液28实施使病毒22灭活的处理(以下,标记为病毒灭活处理)。

[0049] 在阳离子层析装置26中导入有病毒灭活处理后的第1纯化液28。阳离子层析装置26通过使用以阳离子交换体为固定相的柱从第1纯化液28中提取抗体19而生成第2纯化液29。在阴离子层析装置27中导入有第2纯化液29。阴离子层析装置27通过使用以阴离子交换体为固定相的柱从第2纯化液29中提取抗体19而生成第3纯化液30。虽然省略了图示,但对第3纯化液30实施去除病毒的处理。之后,对第3纯化液30实施基于超滤(UF:Ultrafiltration)及渗滤(DF:Diafiltration)的浓缩/过滤处理,由此纯化原料药18。通过依次进行这种基于多种层析装置25~27的成分分离处理,阶段性地去除掺杂物及病毒22,并阶段性地提高抗体19的纯度。

[0050] 以下,为了简化说明,以预测第1纯化液28内的抗体19的浓度的情况为例进行说明。以下,将预测浓度的抗体19标记为对象抗体19T,将包含对象抗体19T的第1纯化液28标记为对象第1纯化液28T。对象抗体19T是本发明的技术所涉及的“对象成分”的一例。对象第1纯化液28T是本发明的技术所涉及的“对象悬浊液”的一例。并且,浓度是本发明的技术所涉及的“状态”的一例。另外,“状态”是指表示对象成分的物理化学特征的指标。

[0051] 作为一例,如图2所示,本实施方式中,在第3工艺12内使用拉曼分光计40测定对象第1纯化液28T的拉曼光谱。拉曼分光计40是利用拉曼散射光的特性进行物质的评价的设备。若将激励光照射到物质,则激励光与物质相互作用,从而产生具有与激励光不同的波长的拉曼散射光。激励光与拉曼散射光的波长差相当于物质具有的分子振动的能量部分。因此,在分子结构不同的物质之间,能够获得具有不同波数的拉曼散射光。拉曼散射光在斯托克斯线及反斯托克斯线中,优选使用斯托克斯线。拉曼光谱是本发明的技术所涉及的“电磁波的光谱”的一例。

[0052] 拉曼分光计40由探针41和分析仪42构成。探针41其前端浸渍在对象第1纯化液28T内。探针41从前端的射出口射出激励光,并在配设于前端的光接收部中接收通过激励光与对象第1纯化液28T的相互作用产生的拉曼散射光。探针41将接收到的拉曼散射光输出到分析仪42中。另外,本实施方式中,作为激励光使用激光束,将激光束的输出设为200mW,将中心波长设为785nm,将照射时间设为1秒钟。另外,拉曼分光计40并不限于将具有光接收部的探针41浸渍于液内而使用的类型,也可以是将具有光接收部的流动池设置在流路中而使用的类型。

[0053] 分析仪42按每个波数分解拉曼散射光,并导出每个波数的拉曼散射光的强度,从而生成光谱测定数据。分析仪42通过LAN(Local Area Network:局域网)等计算机网络,与信息处理装置45可相互通信地连接。分析仪42将所生成的光谱测定数据作为对象光谱测定数据46T发送到信息处理装置45。信息处理装置45接收来自分析仪42的对象光谱测定数据46T。并且,信息处理装置45接受作为与对象抗体19T有关的信息的对象成分关联信息47T。

在此,“成分关联信息”是指该成分固有的信息,与成分的状态无关。并且,“成分关联信息”是对拉曼光谱等光谱产生影响的信息。

[0054] 信息处理装置45例如为桌面型个人计算机,具备显示配设有GUI(GraPhical User Interface:图形用户界面)的各种画面的显示器50及用于进行通过GUI的操作的键盘及鼠标等输入设备51。对象成分关联信息47T例如经由输入设备51由信息处理装置45的操作者输入。另外,信息处理装置45可以是笔记本型个人计算机,也可以是平板终端。

[0055] 作为一例,如图3所示,对象光谱测定数据46T是记录有相对于各波数的拉曼散射光的强度的数据。图3中,对象光谱测定数据46T是以$1cm^{-1}$刻度导出波数$500cm^{-1}$～$3000cm^{-1}$为止的范围的散射光的强度的数据。另外,图3的下部所示的曲线图G是按每个波数标绘该对象光谱测定数据46T的强度并用线连接而成的曲线图。

[0056] 作为一例,如图4所示,对象成分关联信息47T包含对象抗体19T的氨基酸的组成比的信息(以下,标记为对象氨基酸组成比信息)55T及对象抗体19T的亚类的信息(以下,标记为对象亚类信息)56T。对象氨基酸组成比信息55T是记录有组氨酸、亮氨酸等各种氨基酸的组成比(%)的信息。对象亚类信息56T是记录有Ig(Immunoglobulin:免疫球蛋白)G1、IgG2、IgG3、IgG4等对象抗体19T的亚类的信息。图4中,例示出作为亚类记录有IgG2的情况。作为对象氨基酸组成比信息55T,可以是对象抗体19T的重链的氨基酸及轻链的氨基酸中任一氨基酸的组成比的信息,但更优选为重链及轻链两者的氨基酸的组成比的信息。

[0057] 作为一例,如图5所示,构成信息处理装置45的计算机除了前述显示器50及输入设备51以外,还具备存储设备(storage)60、存储器(memory)61、CPU(Central Processing Unit:中央处理器)62及通信部63。它们经由总线64彼此连接。

[0058] 存储设备60是内置于构成信息处理装置45的计算机或通过电缆、网络与构成信息处理装置45的计算机连接的硬盘驱动器。或者,存储设备60是联装多台硬盘驱动器而成的磁盘阵列。存储设备60中存储有操作系统等的控制程序、各种应用程序及这些程序所附带的各种数据等。另外,也可以使用固态驱动器来代替硬盘驱动器。

[0059] 存储器61是用于CPU62执行处理的工作存储器。CPU62将存储在存储设备60中的程序加载到存储器61中,并按照程序执行处理。由此,CPU62集中控制计算机的各部。CPU62是本发明的技术所涉及的“处理器”的一例。通信部63进行与拉曼分光计40等外部装置的各种信息的传输控制。另外,存储器61可以内置于CPU62中。

[0060] 作为一例,如图6所示,在信息处理装置45的存储设备60中存储有工作程序70。工作程序70为用于使计算机作为信息处理装置45发挥作用的应用程序。即,工作程序70是本发明的技术所涉及的“信息处理装置的工作程序”的一例。在存储设备60中,还存储有学习完毕浓度预测模型71。学习完毕浓度预测模型71是本发明的技术所涉及的“校准完毕状态预测模型”的一例。除此以外,在存储设备60中,存储有显示在显示器50中的各种画面的数据等。

[0061] 若工作程序70被启动,则构成信息处理装置45的计算机的CPU62与存储器61等协同而作为接收部75、接受部76、读写(以下,简称为RW(Read Write))控制部77、预处理部78、预测部79及显示控制部80发挥作用。

[0062] 接收部75接收来自拉曼分光计40的对象光谱测定数据46T。如此通过由接收部75接收对象光谱测定数据46T,从而CPU62获取对象光谱测定数据46T。接收部75将对象光谱测

定数据46T输出到RW控制部77中。

[0063] 接受部76接受经由输入设备51由操作者输入的对象成分关联信息47T。如此通过由接受部76接受对象成分关联信息47T,从而CPU62获取对象成分关联信息47T。接受部76将对象成分关联信息47T输出到RW控制部77中。

[0064] RW控制部77控制存储在存储设备60中的各种数据的读出及各种数据向存储设备60的存储。例如RW控制部77将来自接收部75的对象光谱测定数据46T及来自接受部76的对象成分关联信息47T存储在存储设备60中。RW控制部77从存储设备60读出对象光谱测定数据46T,并将所读出的对象光谱测定数据46T输出到预处理部78中。并且,RW控制部77从存储设备60读出对象成分关联信息47T,并将所读出的对象成分关联信息47T输出到预测部79中。而且,RW控制部77从存储设备60读出学习完毕浓度预测模型71,并将所读出的学习完毕浓度预测模型71输出到预测部79中。

[0065] 预处理部78对对象光谱测定数据46T实施预处理,将对象光谱测定数据46T设为预处理完毕对象光谱测定数据46TP。预处理部78将预处理完毕对象光谱测定数据46TP输出到预测部79中。

[0066] 预测部79将预处理完毕对象光谱测定数据46TP及对象成分关联信息47T适用于学习完毕浓度预测模型71,并从学习完毕浓度预测模型71输出预测出对象抗体19T的浓度的浓度预测值85。预测部79将浓度预测值85输出到显示控制部80中。浓度预测值85是本发明的技术所涉及的“对象状态预测结果”的一例。

[0067] 显示控制部80控制各种画面向显示器50的显示。例如显示控制部80使对象成分关联信息47T的输入画面显示在显示器50中。并且,显示控制部80使向操作者通知来自预测部79的浓度预测值85的通知画面显示在显示器50中。

[0068] 作为一例,如图7所示,预处理部78将噪声去除处理90、峰值分离处理91及峰值强调处理92作为预处理对对象光谱测定数据46T进行实施。噪声去除处理90例如是基于SG(Savitzky-Golay)法的平滑处理及基线校正处理等。峰值分离处理91是微分处理(也称为导数计算处理)等。峰值强调处理92是归一化处理(也称为标准化处理)、平均化处理、例如基于主成分分析的维度削减处理及稀疏化处理等。

[0069] 作为一例,如图8所示,作为峰值强调处理92的稀疏化处理是对象光谱测定数据46T的各波数的强度中排除对浓度预测值85的相关性相对低的波数的强度的处理。通过该稀疏化处理,预处理完毕对象光谱测定数据46TP的强度的数量、即数据数量比对象光谱测定数据46T的数据数量(本例种为2501)大幅减少。预处理完毕对象光谱测定数据46TP的数据数量例如优选为5以上且小于1000,更优选为5以上且小于800,进一步优选为5以上且小于500。

[0070] 作为一例,如图9所示,学习完毕浓度预测模型71具备第1模型95及第2模型96。预测部79将预处理完毕对象光谱测定数据46TP输入到第1模型95中,并从第1模型95输出作为对象抗体19T的浓度的临时预测值的临时浓度预测值85T。即,第1模型95是根据预处理完毕对象光谱测定数据46TP输出临时浓度预测值85T的模型。临时浓度预测值85T是本发明的技术所涉及的“临时预测结果”的一例。

[0071] 接着,预测部79将第1模型95所输出的临时浓度预测值85T和对象成分关联信息47T输入到第2模型96,并从第2模型96输出浓度预测值85。即,第2模型96是根据对象成分关

联信息47T及临时浓度预测值85T输出浓度预测值85的模型。如此,预测部79使用第1模型95及第2模型96分2个阶段预测浓度预测值85。

[0072] 作为一例,如图10所示,第1模型95及第2模型96由神经网络100构建。众所周知,神经网络100具有输入层101、隐匿层(也称为中间层)102及输出层103。这些输入层101、隐匿层102及输出层103分别具有多个节点ND。在输入层101的节点ND与隐匿层102的节点ND之间、在隐匿层102内的节点ND之间、以及在隐匿层102的节点ND与输出层103的节点ND之间,设定有表示各节点ND的结合的强度的系数。在输出层103的节点ND设定有线性函数、ReLu(Rectified Linear Unit:线性整流函数)函数等适当的激活函数。

[0073] 在第1模型95的输入层101的各节点ND,输入有预处理完毕对象光谱测定数据46TP的各波数的强度。并且,从第1模型95的输出层103的节点ND,输出临时浓度预测值85T。另一方面,在第2模型96的输入层101的各节点ND,输入有临时浓度预测值85T和对象成分关联信息47T的对象氨基酸组成比信息55T的各组成比及对象亚类信息56T。并且,从第2模型96的输出层103的节点ND,输出浓度预测值85。

[0074] 作为一例,如图11所示,用于生成学习完毕浓度预测模型71的教师数据组110是多个教师数据111A、111B、……(以下,有时统一标记为教师数据111)的集合。教师数据组110例如通过使用教师数据111学习机器学习模型,从而存储于将机器学习模型设为学习完毕浓度预测模型71的学习装置的存储设备112中。另外,学习装置可以是与信息处理装置45不同的装置,也可以是信息处理装置45。在信息处理装置45负责学习装置的功能的情况下,存储设备112是信息处理装置45的存储设备60。教师数据111是本发明的技术所涉及的“校准数据”的一例。

[0075] 教师数据111能够从市售的生物医药品、或者用小规模设备实际制造出的生物医药品等中收集。除此以外,例如,也能够在过去的生物医药品的制造工艺2中获取。教师数据111由光谱测定数据46、成分关联信息47及浓度测定值113的组构成。光谱测定数据46与图2中所示的对象光谱测定数据46T的情况相同地为通过拉曼分光计40对过去的生物医药品的制造工艺2中纯化的第1纯化液28的拉曼光谱进行测定而获得的数据。成分关联信息47是与过去的生物医药品的制造工艺2中的抗体19有关的信息。成分关联信息47包含过去的生物医药品的制造工艺2中的抗体19的氨基酸组成比信息55及亚类信息56。浓度测定值113是例如使用HPLC(High Performance Liquid ChromatograPhy:高效液相层析法)等方法而实际测定过去的生物医药品的制造工艺2中纯化的第1纯化液28中的抗体19的浓度而获得的值。

[0076] 在图11中,作为教师数据111,例示出2个教师数据111A及111B。教师数据111A是从抗体19A及包含抗体19A的第1纯化液28A中获取的数据,教师数据111B是从抗体19B及包含抗体19B的第1纯化液28B中获取的数据。具体而言,教师数据111A由通过拉曼分光计40测定第1纯化液28A的拉曼光谱而获得的光谱测定数据46A、与抗体19A有关的成分关联信息47A、及第1纯化液28A中的抗体19A的浓度测定值113A的组构成。教师数据111B由通过拉曼分光计40测定第1纯化液28B的拉曼光谱而获得的光谱测定数据46B、与抗体19B有关的成分关联信息47B、及第1纯化液28B中的抗体19B的浓度测定值113B的组构成。

[0077] 根据亚类信息56A,抗体19A的亚类为IgG1。相对于此,根据亚类信息56B,抗体19B的亚类为IgG4。即,教师数据111A与教师数据111B中,抗体19的亚类不同。因此,教师数据111A与教师数据111B的种类不同。并且,根据图4中所示的对象亚类信息56T,对象抗体19T

的亚类为IgG2。即,对象抗体19T与抗体19A及19B不同。

[0078] 抗体19A是本发明的技术所涉及的“第1成分”的一例,第1纯化液28A是本发明的技术所涉及的“第1悬浊液”的一例。教师数据111A是本发明的技术所涉及的“第1校准数据”的一例。并且,光谱测定数据46A是本发明的技术所涉及的“第1光谱测定数据”的一例,成分关联信息47A是本发明的技术所涉及的“第1成分关联信息”的一例。而且,浓度测定值113A是本发明的技术所涉及的“第1状态关联信息”的一例。并且,抗体19B是本发明的技术所涉及的“第2成分”的一例,第2纯化液28B是本发明的技术所涉及的“第2悬浊液”的一例。教师数据111B是本发明的技术所涉及的“第2校准数据”的一例。并且,光谱测定数据46B是本发明的技术所涉及的“第2光谱测定数据”的一例,成分关联信息47B是本发明的技术所涉及的“第2成分关联信息”的一例。而且,浓度测定值113B是本发明的技术所涉及的“第2状态关联信息”的一例。

[0079] 作为一例,如图12所示,教师数据组110以数据表的形式存储在存储设备112中。教师数据111中,光谱测定数据46及成分关联信息47用作学习用输入数据(解释变量),浓度测定值113用作正解数据(目标变量)。另外,在亚类信息56中记录有作为对象抗体19T的亚类的IgG2的教师数据111不存在。

[0080] 作为一例,如图13所示,学习装置对教师数据111的光谱测定数据46实施与基于预处理部78的预处理相同的预处理,并设为预处理完毕光谱测定数据46P。

[0081] 预处理中,图8中所示的稀疏化处理根据对教师数据111的光谱测定数据46的稀疏建模来进行。本实施方式中的稀疏建模是指,对于将教师数据111的光谱测定数据46中所包含的每个波数的强度设为解释变量并将浓度预测值85设为目标变量的回归模型,分选解释变量,即,排除解释变量的一部分。作为稀疏建模的方法,例如能够使用基于Lasso(Least absolute shrinkage and selection oPerator:最小绝对收缩和选择算子)回归的方法。Lasso回归是以通过将惩罚项(也称为处罚项)相加于均方根误差(RMSE:Root Mean Squared Error)而计算出的损失函数成为最小的方式分选解释变量的方法。惩罚项例如通过以K-分割交叉验证为代表的交叉验证(Cross-Validation)等来确定。

[0082] 稀疏建模按以下顺序进行。首先,对教师数据111的光谱测定数据46,进行间隔剔除随机确定的波数的强度的处理,构建表示间隔剔除处理后的光谱测定数据46与对应的浓度预测值85的关联性的回归模型。并且,对于所构建的回归模型,导出将惩罚项相加于RMSE而得到的损失函数。通过重复进行规定次数的这些间隔剔除、回归模型的构建及损失函数的导出的各处理,对于被间隔剔除的波数相互不同的多个光谱测定数据46中的每一个,生成回归模型,对于各回归模型,导出多个损失函数。并且,将能够使损失函数最小化的最少数的强度分选为对浓度预测值85的相关性相对高的强度,并将除此以外的强度作为对浓度预测值85的相关性相对低的强度而排除。

[0083] 作为一例,如图14所示,在第1模型95的学习阶段中,学习装置将教师数据111中的预处理完毕光谱测定数据46P输入到第1模型95中,并从第1模型95输出学习用临时浓度预测值85TL。学习用临时浓度预测值85TL是本发明的技术所涉及的“学习用状态预测结果”的一例。

[0084] 学习装置根据学习用临时浓度预测值85TL与浓度测定值113的比较结果,进行使用了损失函数的第1模型95的损失运算。学习装置根据损失运算的结果进行第1模型95的节

点ND之间的系数的更新设定,并根据更新设定更新第1模型95。

[0085] 在第1模型95的学习阶段中,学习装置一边变更教师数据111一边重复进行预处理完毕光谱测定数据46P向第1模型95的输入、来自第1模型95的学习用临时浓度预测值85TL的输出、损失运算、更新设定及第1模型95的更新的上述一系列处理。学习装置在相对于浓度测定值113的学习用临时浓度预测值85TL的预测精度达到预先规定的设定水平时,结束上述一系列处理的重复。如此预测精度达到设定水平的第1模型95作为学习完毕浓度预测模型71的一部分存储于存储设备60,并在预测部79中进行使用。另外,也可以与相对于浓度测定值113的学习用临时浓度预测值85TL的预测精度无关地,在重复进行了设定次数的上述一系列处理时,结束学习。

[0086] 作为一例,如图15所示,在第2模型96的学习阶段中,学习装置将教师数据111中的成分关联信息47和第1模型95所输出的学习用临时浓度预测值85TL输入到第2模型96中,并从第2模型96输出学习用浓度预测值85L。学习用浓度预测值85L与学习用临时浓度预测值85TL相同地为本发明的技术所涉及的“学习用状态预测结果”的一例。

[0087] 学习装置根据学习用浓度预测值85L与浓度测定值113的比较结果,进行使用了损失函数的第2模型96的损失运算。学习装置根据损失运算的结果进行第2模型96的节点ND之间的系数的更新设定,并根据更新设定更新第2模型96。

[0088] 在第2模型96的学习阶段中,学习装置一边变更教师数据111一边重复进行成分关联信息47及学习用临时浓度预测值85TL向第2模型96的输入、来自第2模型96的学习用浓度预测值85L的输出、损失运算、更新设定及第2模型96的更新的上述一系列处理。学习装置在相对于浓度测定值113的学习用浓度预测值85L的预测精度达到预先规定的设定水平时,结束上述一系列处理的重复。如此预测精度达到设定水平的第2模型96作为学习完毕浓度预测模型71的一部分存储于存储设备60,并在预测部79中进行使用。另外,也可以与相对于浓度测定值113的学习用浓度预测值85L的预测精度无关地,在重复进行了设定次数的上述一系列处理时,结束学习。

[0089] 学习装置在结束第1模型95的学习之后进行第2模型96的学习。当然,也可以并行进行第1模型95的学习和第2模型96的学习。

[0090] 接着,参考图16～图19所示的流程图,对基于上述结构的作用进行说明。

[0091] 作为一例,如图16所示,首先,在学习装置中获取用于生成学习完毕浓度预测模型71的多个教师数据111(步骤ST100)。更详细而言,如图11中所示,获取从亚类为IgG1的抗体19A和包含抗体19A的第1纯化液28A中所获取的教师数据111A、从亚类为IgG4的抗体19B和包含抗体19B的第1纯化液28B中所获取的教师数据111B等至少2种教师数据111。教师数据111存储于学习装置的存储设备112(步骤ST110)。

[0092] 如图13中所示,对教师数据111的光谱测定数据46实施预处理,并设为预处理完毕光谱测定数据46P。

[0093] 在学习装置中,进行图14中所示的第1模型95的学习。作为一例,如图17所示,预处理完毕光谱测定数据46P被输入到第1模型95中,由此从第1模型95输出学习用临时浓度预测值85TL(步骤ST200)。接着,根据学习用临时浓度预测值85TL与浓度测定值113的比较结果,更新第1模型95(步骤ST210)。在相对于浓度测定值113的学习用临时浓度预测值85TL的预测精度未达到设定水平时(步骤ST220中“否”),一边变更教师数据111(步骤ST230)一边

重复进行这些步骤ST200及步骤ST210的处理。在相对于浓度测定值113的学习用临时浓度预测值85TL的预测精度达到设定水平时(步骤ST220中“是”),结束处理。

[0094] 并且,在学习装置中,进行图15中所示的第2模型96的学习。作为一例,如图18所示,学习用临时浓度预测值85TL及成分关联信息47被输入到第2模型96中,由此从第2模型96输出学习用浓度预测值85L(步骤ST300)。接着,根据学习用浓度预测值85L与浓度测定值113的比较结果,更新第2模型96(步骤ST310)。在相对于浓度测定值113的学习用浓度预测值85L的预测精度未达到设定水平时(步骤ST320中“否”),一边变更教师数据111(步骤ST330)一边重复进行这些步骤ST300及步骤ST310的处理。在相对于浓度测定值113的学习用浓度预测值85L的预测精度达到设定水平时(步骤ST320中“是”),结束处理。如此学习的第1模型95及第2模型96一同作为学习完毕浓度预测模型71存储于信息处理装置45的存储设备60。

[0095] 若在信息处理装置45中工作程序70被启动,则如图6所示,信息处理装置45的CPU62作为接收部75、接受部76、RW控制部77、预处理部78、预测部79及显示控制部80发挥作用。

[0096] 作为一例,如图19所示,通过接收部75接收来自拉曼分光计40的对象光谱测定数据46T。并且,通过接受部76,接受经由输入设备51由操作者输入的对象成分关联信息47T。由此,由CPU62获取对象光谱测定数据及对象成分关联信息47T(步骤ST400)。对象光谱测定数据46T从接收部75输出到RW控制部77中,通过RW控制部77存储于存储设备60。对象成分关联信息47T从接受部76输出到RW控制部77中,通过RW控制部77存储于存储设备60。

[0097] 对象光谱测定数据46T通过RW控制部77从存储设备60读出,并输出到预处理部78中。如图7及图8中所示,在预处理部78中,对对象光谱测定数据46T实施预处理,并设为预处理完毕对象光谱测定数据46TP(步骤ST410)。预处理完毕对象光谱测定数据46TP从预处理部78输出到预测部79中。

[0098] 对象成分关联信息47T通过RW控制部77从存储设备60读出,并输出到预测部79中。如图9中所示,在预测部79中,首先,预处理完毕对象光谱测定数据46TP被输入到第1模型95中,由此从第1模型95输出临时浓度预测值85T(步骤ST420)。接着,临时浓度预测值85T及对象成分关联信息47T被输入到第2模型96中,由此从第2模型96输出浓度预测值85(步骤ST430)。浓度预测值85从预测部79输出到显示控制部80中。

[0099] 在显示控制部80的控制下,浓度预测值85的通知画面显示在显示器50中(步骤ST440)。由此向操作者通知浓度预测值85。操作者根据浓度预测值85,验证是否适当地进行了基于免疫亲和层析装置25的纯化,或者研究免疫亲和层析装置25的纯化条件的变更。纯化条件例如是将培养上清液17注入柱时的流速、或者从柱洗提抗体19时使用的缓冲液的量及组成等。另外,可以从预测部79向RW控制部77输出浓度预测值85,并通过RW控制部77将浓度预测值85存储于存储设备60。

[0100] 如以上所说明,信息处理装置45的CPU62使用利用教师数据111A和教师数据111B中至少2种教师数据111而进行了学习的学习完毕浓度预测模型71。教师数据111A将从包含抗体19A的第1纯化液28A获得的光谱测定数据46A及与抗体19A有关的成分关联信息47A设为学习用输入数据(解释变量),并且,将实际测定出第1纯化液28A中的抗体19A的浓度的浓度测定值113A设为正解数据(目标变量)。教师数据111B将从包含抗体19B的第1纯化液28B

获得的光谱测定数据46B及与抗体19B有关的成分关联信息47B设为学习用输入数据(解释变量),并且,将实际测定出第1纯化液28B中的抗体19B的浓度的浓度测定值113B设为正解数据(目标变量)。

[0101] CPU62获取与对象抗体19T有关的对象成分关联信息47T、及从包含对象抗体19T的对象第1纯化液28T获得的对象光谱测定数据46T。CPU62将对象成分关联信息47T及对象光谱测定数据46T适用于学习完毕浓度预测模型71,从学习完毕浓度预测模型71输出对象第1纯化液28T中的对象抗体19T的浓度预测值85。

[0102] 不仅参考光谱测定数据46还参考成分关联信息47,因此也能够高精度地预测教师数据111中不存在的亚类的对象抗体19T的浓度。因此,无需生成专用于各抗体19的专用的学习完毕浓度预测模型71,能够用1个学习完毕浓度预测模型71进行多种抗体19的浓度的预测。因此,能够有效地预测对象抗体19T的浓度。

[0103] 学习完毕浓度预测模型71具备第1模型95及第2模型96。第1模型95根据对象光谱测定数据46T输出临时浓度预测值85T。第2模型96根据对象成分关联信息47T及临时浓度预测值85T输出浓度预测值85。如此,以通过对象成分关联信息47T来校正根据对象光谱测定数据46T预测出的临时浓度预测值85T的形式,分2个阶段进行对象抗体19T的浓度的预测,由此能够提高浓度预测值85的预测精度。

[0104] 对象抗体19T与为了获得教师数据111而使用的抗体19(抗体19A及19B等)不同。因此,能够进一步发挥能够有效地预测对象抗体19T的浓度之类的效果。另外,也能够预测亚类与教师数据111的抗体19相同的对象抗体19T(本例中,与抗体19A相同地亚类为IgG1的对象抗体19T、或者与抗体19B相同地亚类为IgG4的对象抗体19T)的浓度。

[0105] 预处理部78对对象光谱测定数据46T实施包含噪声去除处理90、峰值分离处理91及峰值强调处理92的预处理。预测部79将预处理完毕对象光谱测定数据46TP适用于学习完毕浓度预测模型71。因此,能够进一步提高浓度预测值85的预测精度。

[0106] 另外,预处理部78只要将噪声去除处理90、峰值分离处理91及峰值强调处理92中的至少任一个作为预处理进行实施即可。更具体而言,预处理部78只要将上述中例示出的平滑处理、基线校正处理、微分处理、归一化处理、平均化处理、维度削减处理及稀疏化处理中的至少任一个处理(也可以是2个以上的组合)作为预处理实施即可。

[0107] 包含抗体19作为蛋白质的生物医药品称为抗体医药品,以癌症、糖尿病、类风湿性关节炎之类的慢性疾病的治疗为代表,在血友病、克罗恩氏病之类的罕见疾病的治疗中被广泛地使用。因此,根据将第1成分、第2成分及对象成分作为蛋白质而且将蛋白质作为抗体19的本例,能够促进在各种疾病的治疗中被广泛地使用的抗体医药品的开发。

[0108] 另外,生物分子及对象成分并不限于蛋白质。可以是肽、核酸(DNA、RNA(Ribonucleic Acid:核糖核酸))、脂质、病毒、病毒亚单位及病毒样颗粒等。并且,也可以是通过化学合成法合成的分子。

[0109] 对象成分并不限于抗体19。也可以将源自细胞的蛋白质·源自细胞的DNA20及凝聚物21等掺杂物设为对象成分。若在生物医药品中混入目标蛋白质以外的掺杂物,则即使其量为微量也可能会影响生物医药品的药效,因此预测掺杂物的状态也是重要的。

[0110] 第1悬浊液、第2悬浊液及对象悬浊液并不限于例示的第1纯化液28。也可以是培养上清液17、第2纯化液29或者第3纯化液30。并且,也可以是培养槽16内的培养液。

[0111] 成分关联信息47A、成分关联信息47B及对象成分关联信息47T包含抗体19A的氨基酸组成比信息55A、抗体19B的氨基酸组成比信息55B及对象抗体19T的对象氨基酸组成比信息55T。氨基酸组成比信息55A、氨基酸组成比信息55B及对象氨基酸组成比信息55T是良好地表现抗体19A、抗体19B及对象抗体19T的特性的信息。并且,成分关联信息47A、成分关联信息47B及对象成分关联信息47T包含抗体19A的亚类信息56A、抗体19B的亚类信息56B及对象抗体19T的对象亚类信息56T。亚类信息56A、亚类信息56B及对象亚类信息56T也是良好地表示抗体19A、抗体19B及对象抗体19T的特性的信息。因此,能够进一步提高浓度预测值85的预测精度。

[0112] 另外,对象成分关联信息47T可以代替对象氨基酸组成比信息55T及对象亚类信息56T,或者除此以外,包含IgA、IgD、IgE、IgG、IgM等对象抗体19T的同种型信息。并且,对象成分关联信息47T也可以包含从氨基末端朝向羧基末端记载了构成对象抗体19T的氨基酸的肽键的顺序的氨基酸排列信息。而且,对象成分关联信息47T也可以包含对象抗体19T的分子量、等电点及摩尔吸光系数等。在对象成分为DNA或者RNA的情况下,能够将DNA或者RNA的核酸排列的信息设为对象成分关联信息47T。

[0113] 拉曼光谱容易反映源自蛋白质的氨基酸的官能团的信息。因此,如本例通过将光谱设为拉曼光谱,能够进一步提高作为蛋白质的对象抗体19T的浓度预测值85的预测精度。

[0114] 另外,光谱并不限于拉曼光谱。也可以是红外吸收光谱、核磁共振光谱、紫外可见分光(UV-Vis:Ultraviolet Visible AbsorPtion SPectroscoPy)光谱、或者荧光光谱。

[0115] 学习完毕浓度预测模型71是使用教师数据111进行了学习的机器学习模型。机器学习模型一般用于未知参数的预测,能够通过学习将预测精度提高至一定程度的水平。因此,能够简单生成预测精度比较高的学习完毕浓度预测模型71。

[0116] 另外,校准完毕状态预测模型并不限于学习完毕浓度预测模型71。也可以是通过多变量分析、统计分析生成的校准完毕状态预测模型。作为多变量分析、统计分析的例子,为多元回归、偏最小二乘回归(PLS:Partial Least Squares Regression)、主成分回归、逻辑回归、Lasso回归、岭回归、支持向量回归及高斯过程回归等。通过这种多变量分析、统计分析生成的校准完毕状态预测模型中,根据至少2种校准数据确定回归式系数的情况相当于本发明的技术所涉及的“校准”。

[0117] 作为学习完毕浓度预测模型71(第1模型95及第2模型96),例示出神经网络100,但并不限于此。也可以是决策树、随机森林、朴素贝叶斯及梯度提升决策树等。

[0118] 浓度是在了解对象抗体19T的物理化学特征时最常见的指标。因此,如本例作为对象抗体19T的状态预测浓度,则能够使操作者容易理解对象抗体19T的物理化学特征。

[0119] 另外,对象抗体19T的状态并不限于浓度。也可以是对象抗体19T的纯度或者密度。关于纯度,将对象抗体19T的量与掺杂物的量的相加值作为分母,并将对象抗体19T的量作为分子而进行计算。可以预测例如浓度和密度等2个以上的指标。并且,并不限于浓度、纯度等定量指标,也可以是对象抗体19T的品质的水平(可以是好坏的2个阶段、或者1～5的5个阶段等)等定性指标。

[0120] 在对象成分例如为肽的情况下,作为对象成分的状态,可以采用肽中的氨基酸的组成比、生产的混合物的含有组成比(例如相对于谷氨酸的葡萄糖的含有组成比)等。并且,在对象成分为例如从细胞产生的醇等化学物质的情况下,作为对象成分的状态,可以采用

聚合率。这些情况也与对象抗体19T的情况相同,可以预测2个以上的指标,也可以预测定性指标。

[0121] 在第1模型95及第2模型96的学习阶段,将教师数据111的光谱测定数据46及成分关联信息47(校准数据的解释变量)作为学习用输入数据输入到第1模型95及第2模型96中,并从第1模型95及第2模型96输出学习用临时浓度预测值85TL及学习用浓度预测值85L。接着,根据学习用临时浓度预测值85TL及学习用浓度预测值85L与作为教师数据111的正解数据的浓度测定值113(校准数据的目标变量)的比较结果,更新第1模型95及第2模型96。并且,通过一边变更教师数据111一边重复光谱测定数据46及成分关联信息47向第1模型95及第2模型96的输入、来自第1模型95及第2模型96的学习用临时浓度预测值85TL及学习用浓度预测值85L的输出、以及第1模型95及第2模型96的更新,从而生成学习完毕浓度预测模型71。因此,能够简单生成预测精度比较高的学习完毕浓度预测模型71。

[0122] 以下,记载本发明的技术的实施例。

[0123] 本实施例中,为了收集包含各种浓度的抗体的悬浊液的拉曼光谱,实施了模拟第3工艺12的蛋白A亲和层析。作为导入亲和层析的悬浊液,第一,使用了生产亚类IgG1的抗体mAbA的CHO细胞的培养上清液。将蛋白A柱(Cytiva Corporation制制品名MabSelect SuRe)连接于亲和层析装置(Cytiva Corporation制制品名AKTA Pure 25)而使用。

[0124] 在亲和层析装置中的层析洗提步骤中,不是将酸性洗提液切换为阶梯状,而是一边施加浓度梯度一边进行以梯度状通液的梯度洗提。并且,梯度洗提的条件设为如下3个模式。即,在5CV、10CV及15CV的容量范围内设定了蛋白A柱的柱体积(CV:Column Volume)。并且,通过拉曼分光计,对通过亲和层析纯化的第1纯化液的拉曼光谱进行了测定。另外,拉曼光谱的测定在亲和层析的进行中继续进行。并且,通过第1纯化液的离线分析,对第1纯化液中的抗体的浓度进行了实际测定。将亲和层析的进行中所测定的拉曼光谱的光谱测定数据与测定拉曼光谱时的浓度测定值建立关联而进行了存储。

[0125] 第二,使用生产亚类IgG4的抗体mAbB的CHO细胞的培养上清液,与上述抗体mAbA相同地实施试验,获取了光谱测定数据及浓度测定值。但是,层析洗提步骤仅在10CV的1个条件下进行了实施。

[0126] 从抗体mAbA及mAbB各自的氨基酸排列信息,获取抗体mAbA及mAbB各自的氨基酸组成比信息。将所获取的氨基酸组成比信息设为成分关联信息,与先前获取的光谱测定数据及浓度测定值一同设为教师数据。

[0127] 将抗体mAbA的光谱测定数据设为学习用输入数据,并将抗体mAbA的浓度测定值设为正解数据,从而构建了第1模型。并且,将抗体mAbA及mAbB各自的成分关联信息以及从第1模型输出的学习用临时浓度预测值设为学习用输入数据,并将抗体mAbA及mAbB各自的浓度测定值设为正解数据,从而构建了第2模型。如此,在适用了光谱测定数据及成分关联信息的情况下,生成了输出浓度预测值的学习完毕浓度预测模型。

[0128] 为了验证所生成的学习完毕浓度预测模型的预测精度,使用生产与亚类为IgG2的抗体mAbA及mAbB不同的抗体mAbC的CHO细胞的培养上清液,与上述抗体mAbA等相同地实施试验,获取了光谱测定数据及浓度测定值。但是,层析洗提步骤仅在10CV的1个条件下进行了实施。而且,作为抗体mAbC的成分关联信息,从抗体mAbC的氨基酸排列信息获取了抗体mAbC的氨基酸组成比信息。

[0129] 将抗体mAbC的光谱测定数据及成分关联信息适用于如上所述生成的学习完毕浓度预测模型,并从学习完毕浓度预测模型输出浓度预测值。并且,对浓度预测值与浓度测定值进行比较,计算出学习完毕浓度预测模型的预测误差,结果为RMSE＝0.54。因此,确认到对于与用于教师数据的抗体不同种类的抗体,也能够以较高的精度预测浓度。另外,抗体mAbA及抗体mAbB是本发明的技术所涉及的“第1成分”及“第2成分”的一例。并且,抗体mAbC是本发明的技术所涉及的“对象成分”的一例。

[0130] 也可以按抗体19的每个同种型或者每个亚类,生成学习完毕浓度预测模型。在按每个亚类生成学习完毕浓度预测模型的情况下,除了亚类信息56以外,还将例如λ、κ之类的抗体19的轻链的信息设为成分关联信息。

[0131] 上述实施方式中,通过亚类不同,设为“种类不同”,但可以将同种型不同的情况设为“种类不同”,也可以将轻链不同的情况设为“种类不同”。

[0132] 校准数据也可以是与1种抗体19有关的数据。

[0133] 在上述实施方式中,例如,作为接收部75、接受部76、RW控制部77、预处理部78、预测部79及显示控制部80之类的执行各种处理的处理部(Processing Unit:处理单元)的硬件结构,能够使用以下所示的各种处理器(Processor)。如上所述,各种处理器中,除了执行软件(工作程序70)来作为各种处理部发挥作用的通用的处理器即CPU62以外,还包括FPGA(Field Programmable Gate Array:现场可编程门阵列)等在制造之后能够变更电路结构的处理器即可编程逻辑器件(Programmable Logic Device:PLD)、ASIC(Application Specific Integrated Circuit:专用集成电路)等具有为了执行特定处理而专门设计的电路结构的处理器即专用电路等。

[0134] 一个处理部可以由这些各种处理器中的一个构成,也可以由相同种类或不同种类的两个以上处理器的组合(例如,多个FPGA的组合和/或CPU与FPGA的组合)构成。并且,也可以由1个处理器构成多个处理部。

[0135] 作为由一个处理器构成多个处理部的例子,第一,有如下方式:如以用户端及服务器等计算机为代表那样,由一个以上的CPU与软件的组合构成一个处理器,且该处理器作为多个处理部发挥作用。第二,有如下方式:如以片上系统(System On Chip:SoC)等为代表那样,使用由一个IC(Integrated Circuit(集成电路))芯片实现包括多个处理部的系统整体的功能的处理器。如此,各种处理部作为硬件结构,使用上述各种处理器的1个以上而构成。

[0136] 此外,作为这些各种处理器的硬件结构,更具体而言,能够使用将半导体元件等电路元件组合而成的电路(circuitry)。

[0137] 本发明的技术也能够适当地组合上述的各种实施方式和/或各种变形例。并且,并不限于上述实施方式,只要不脱离宗旨,则当然也能够采用各种结构。此外,本发明的技术除了程序以外,还涉及非临时性地存储程序的存储介质。

[0138] 以上所示的记载内容及图示内容为关于本发明的技术所涉及的部分的详细说明,只不过是本发明的技术的一例。例如,与上述的结构、功能、作用及效果有关的说明为与本发明的技术所涉及的部分的结构、功能、作用及效果的一例有关的说明。因此,在不脱离本发明的技术宗旨的范围内,当然可以对以上所示的记载内容及图示内容删除不必要的部分,或者追加或替换新的要件。并且,为了避免错综复杂的情况,并且便于理解本发明的技术所涉及的部分,在以上所示的记载内容及图示内容中省略了与在使得能够实施本发明的

技术的方面不特别需要说明的技术常识等有关的说明。

[0139] 在本说明书中,“A和/或B”的含义与“A及B中的至少一个”相同。即,“A和/或B”是可以仅为A,也可以仅为B,也可以为A与B的组合的含义。并且,在本说明书中,将3个以上的事项用“和/或”连结而表述的情况也适用与“A和/或B”相同的思路。

[0140] 本说明书中所记载的所有文献、专利申请及技术标准,与具体地且分别地记载通过参考而被并入的各个文献、专利申请及技术标准的情况相同程度地,通过参考而被并入本说明书中。

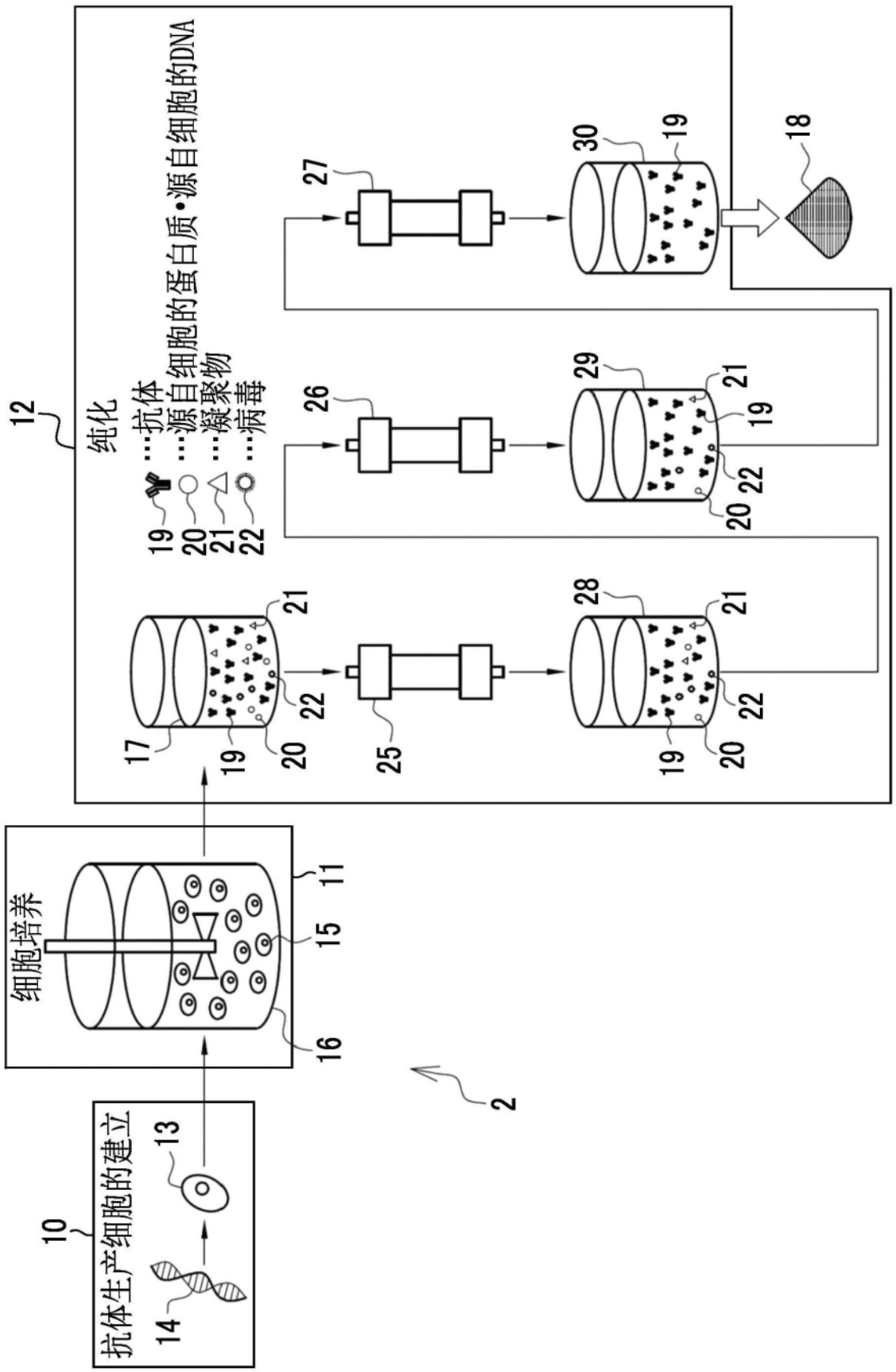

抗体生产细胞的建立
10
13
14
细胞培养
11
15
16
2
12
纯化
…抗体
…源自细胞的蛋白质
…凝聚物
…病毒
…源自细胞的DNA
19
20
21
22
17
25
26
27
28
29
30
18

图1

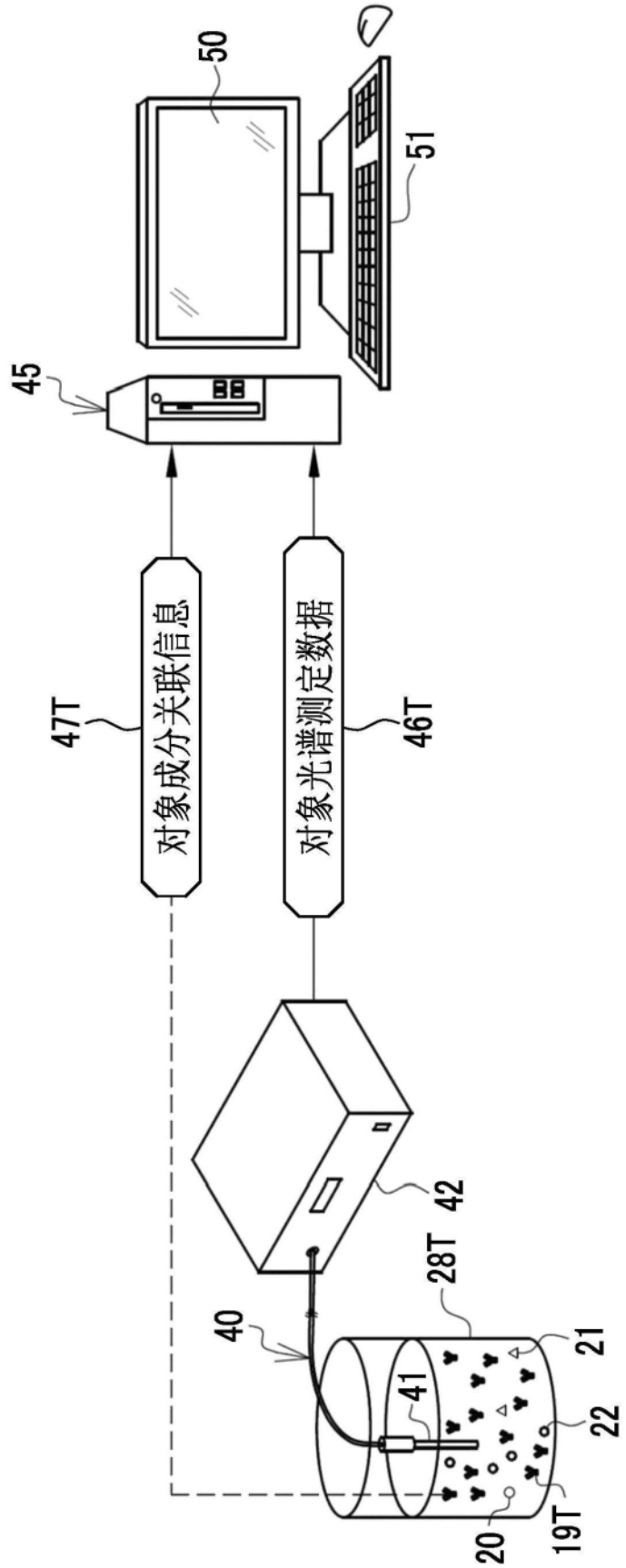


图2

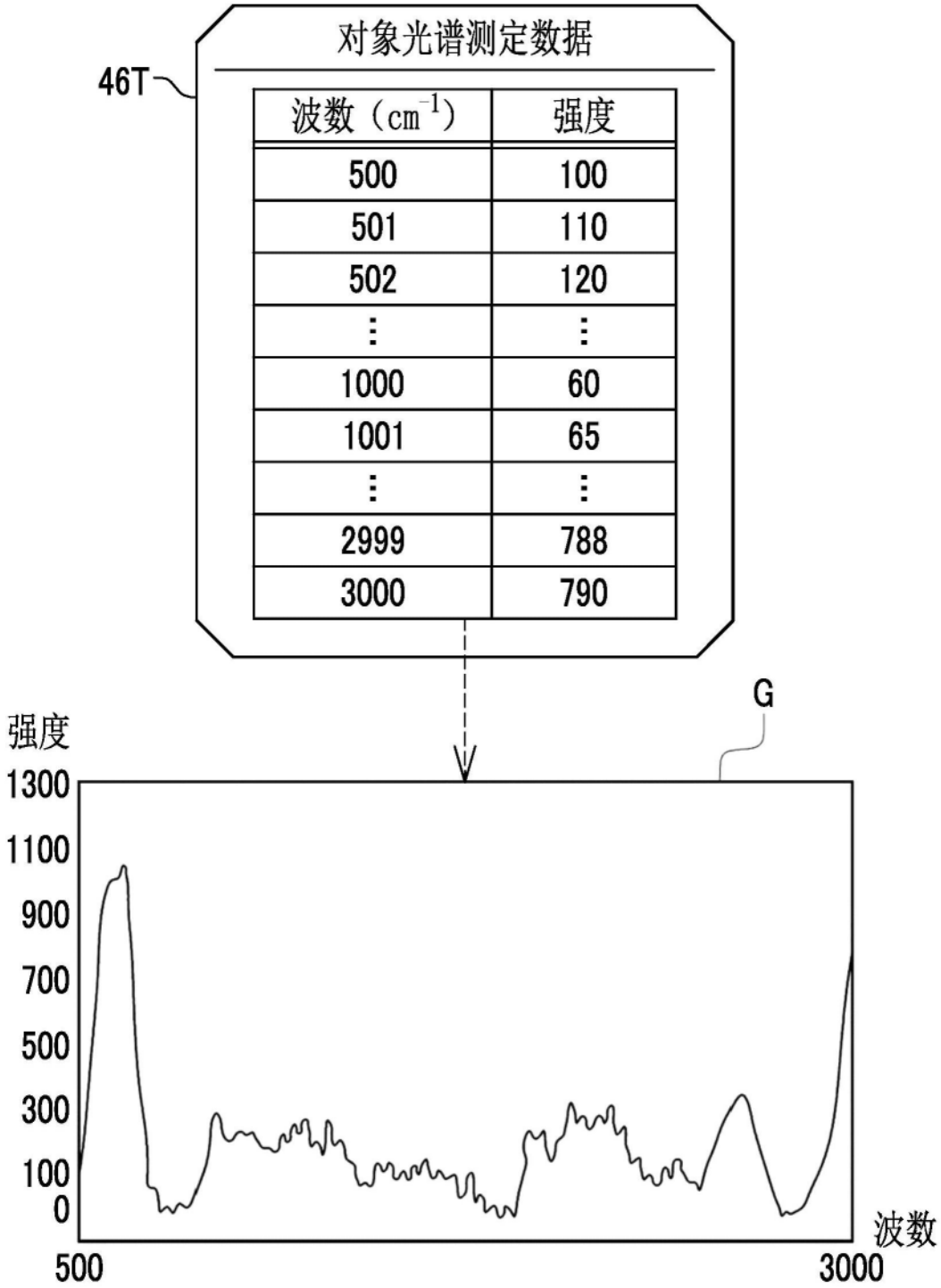

对象光谱测定数据
46T
波数（cm-1）
强度
500
100
501
110
502
120
1000
60
1001
65
2999
788
3000
790
G
强度
1300
1100
900
700
500
300
100
0
500
3000
波数

图3

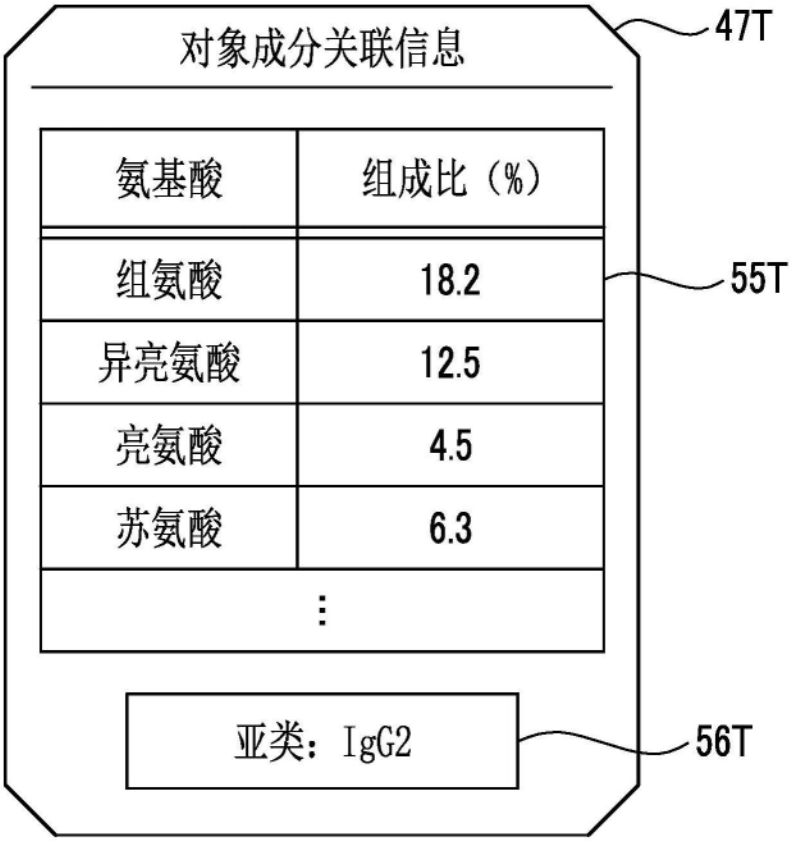

对象成分关联信息
47T
氨基酸
组成比（%）
组氨酸
18.2
55T
异亮氨酸
12.5
亮氨酸
4.5
苏氨酸
6.3
亚类：IgG2
56T

图4

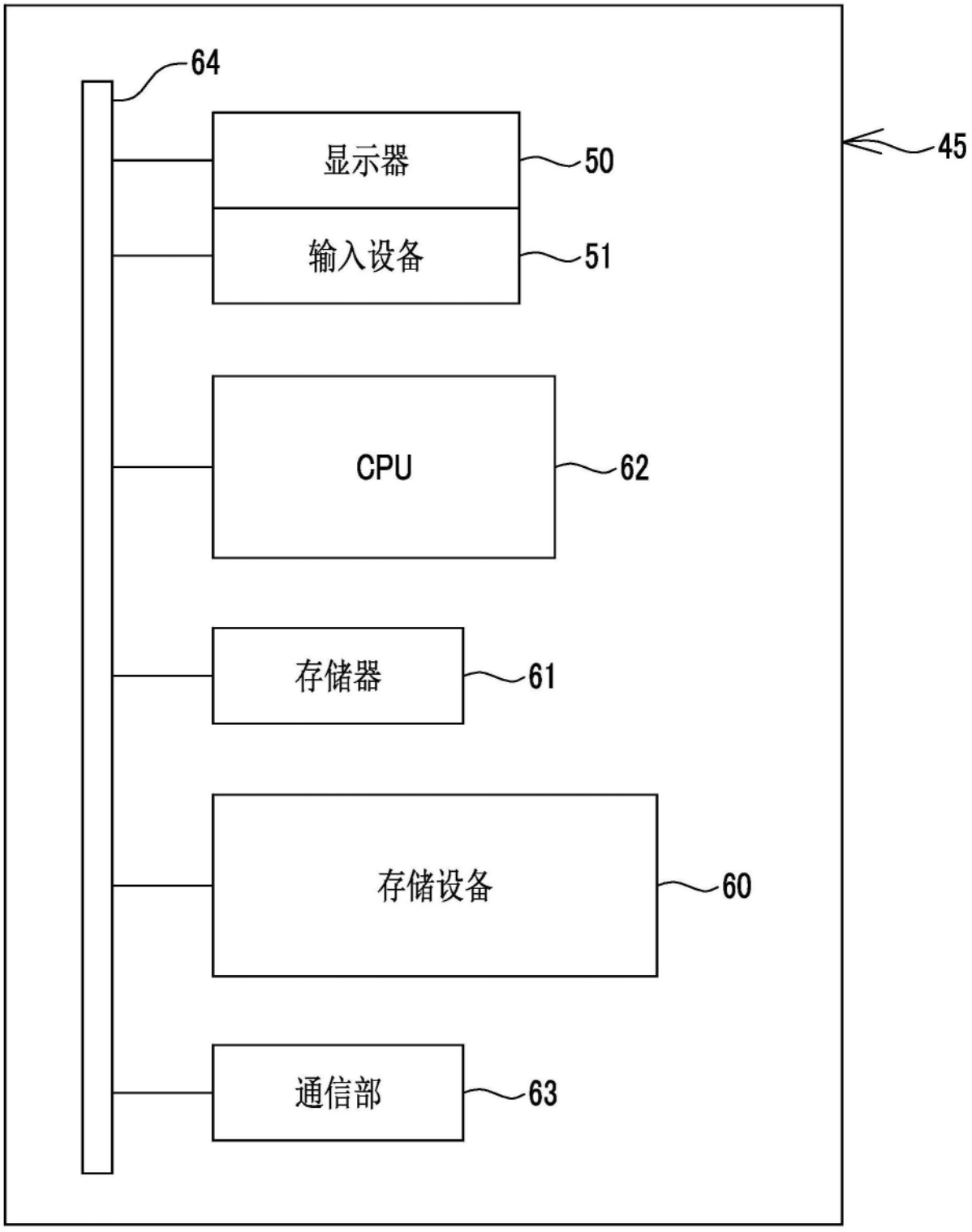


图5

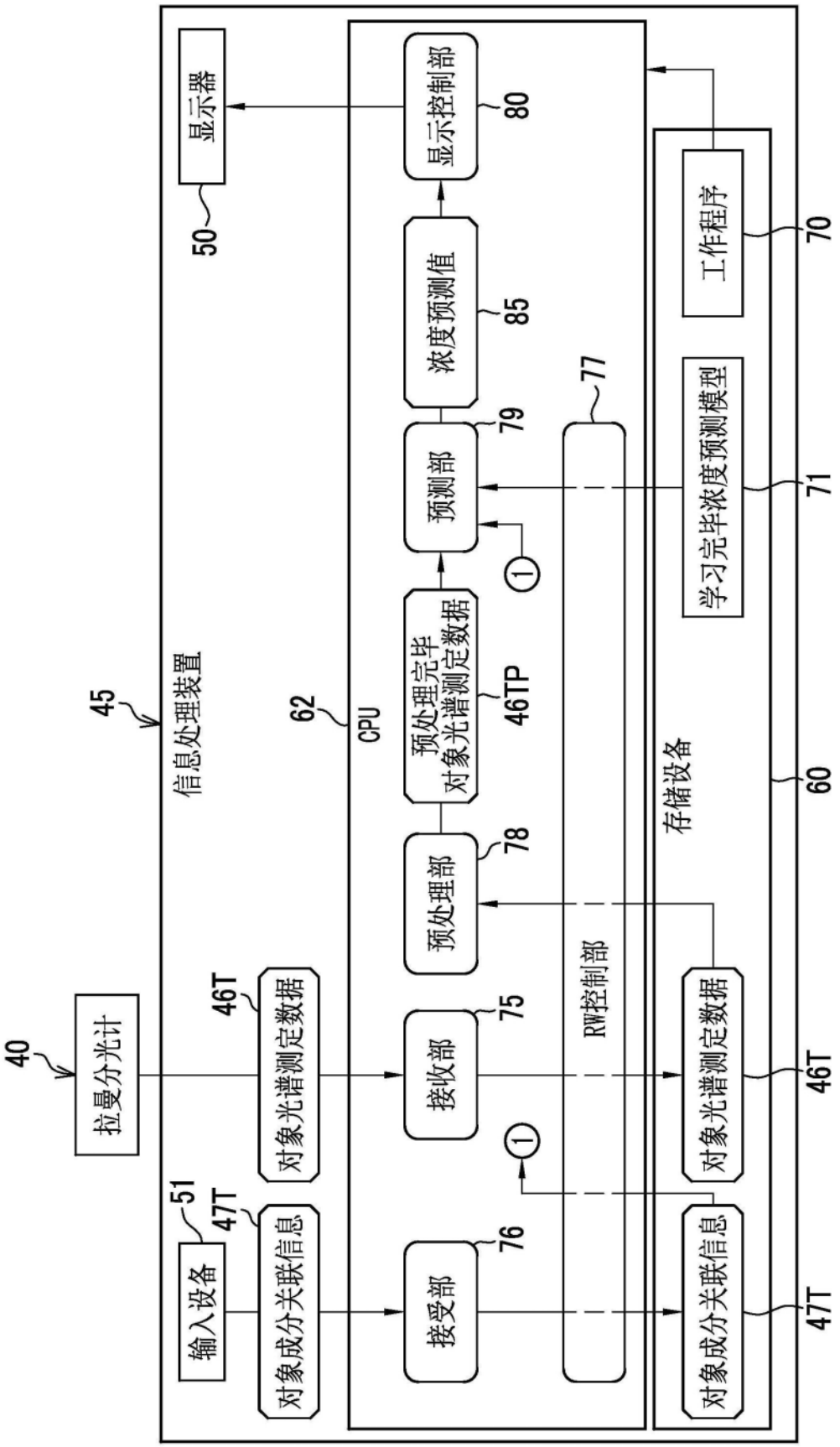

40
拉曼分光计
45
信息处理装置
62
CPU
50
显示器
80
显示控制部
85
浓度预测值
79
预测部
46TP
预处理完毕
对象光谱测定数据
78
预处理部
75
接收部
76
接受部
51
输入设备
47T
对象成分关联信息
46T
对象光谱测定数据
77
RW控制部
60
存储设备
71
学习完毕浓度预测模型
70
工作程序
①

图6

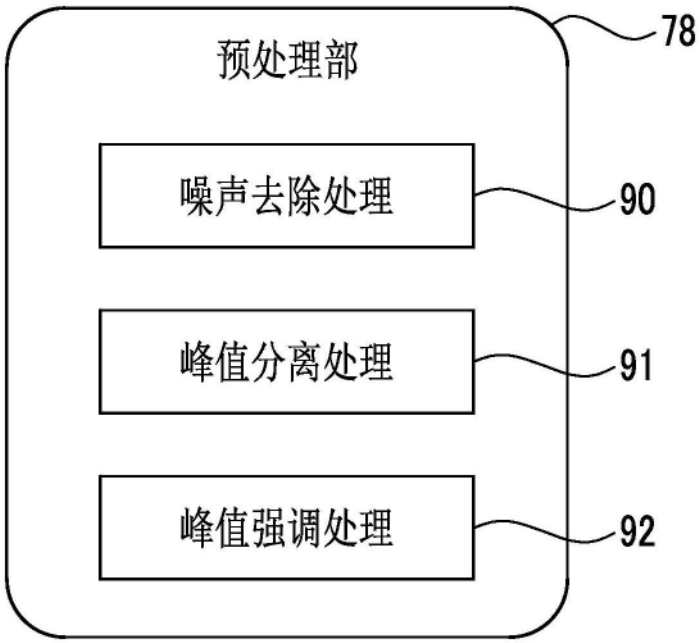


图7

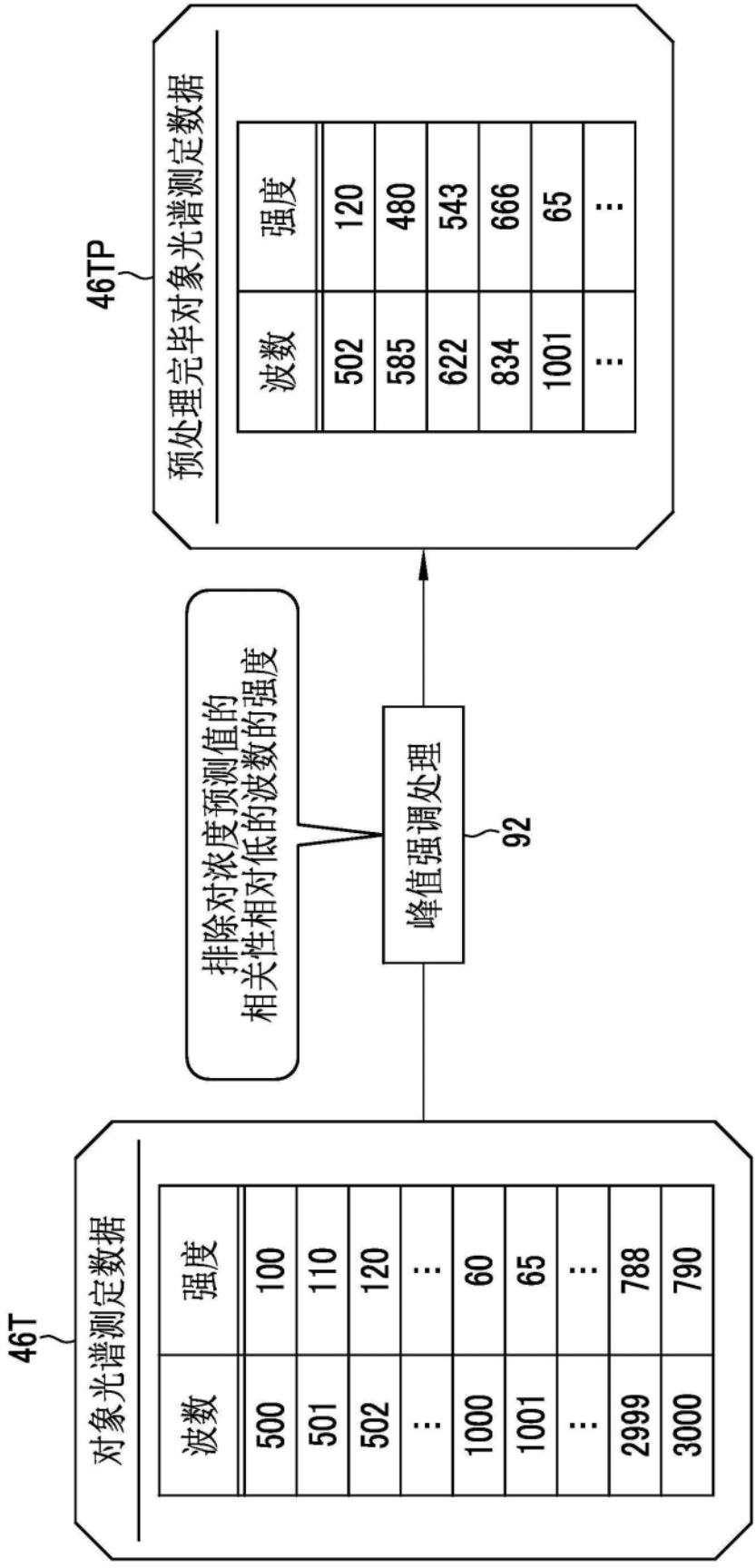

46T
对象光谱测定数据
波数
强度
500 100
501 110
502 120
... ...
1000 60
1001 65
... ...
2999 788
3000 790
峰值强调处理
92
排除对浓度预测值的
相关性相对低的波数的强度
46TP
预处理完毕对象光谱测定数据
波数
强度
502 120
585 480
622 543
834 666
1001 65
... ...

图8

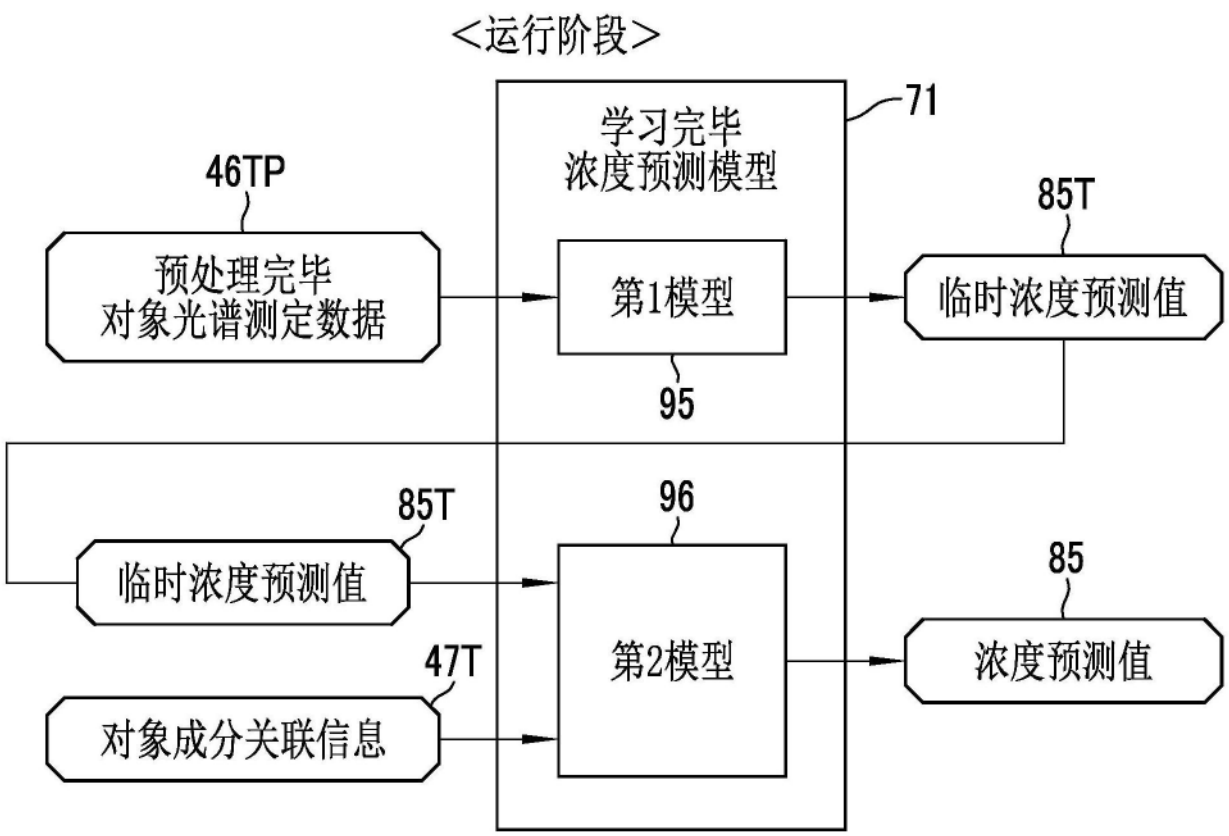

<运行阶段>
学习完毕
浓度预测模型
71
46TP
预处理完毕
对象光谱测定数据
第1模型
95
85T
临时浓度预测值
85T
临时浓度预测值
96
第2模型
47T
对象成分关联信息
85
浓度预测值

图9

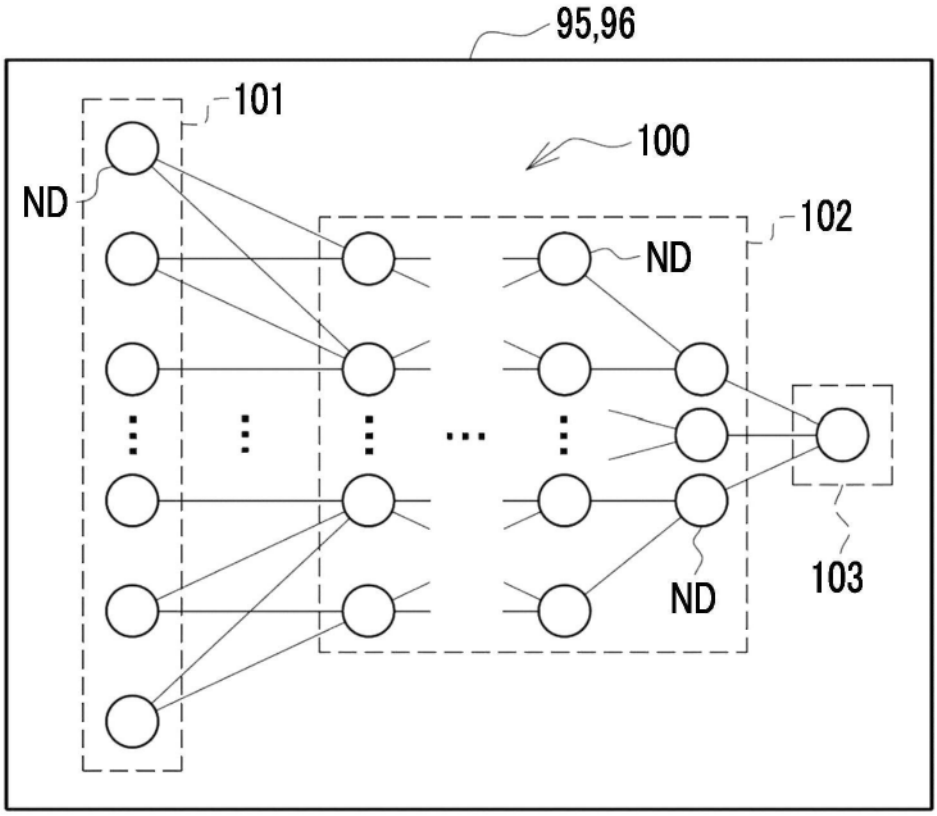


图10

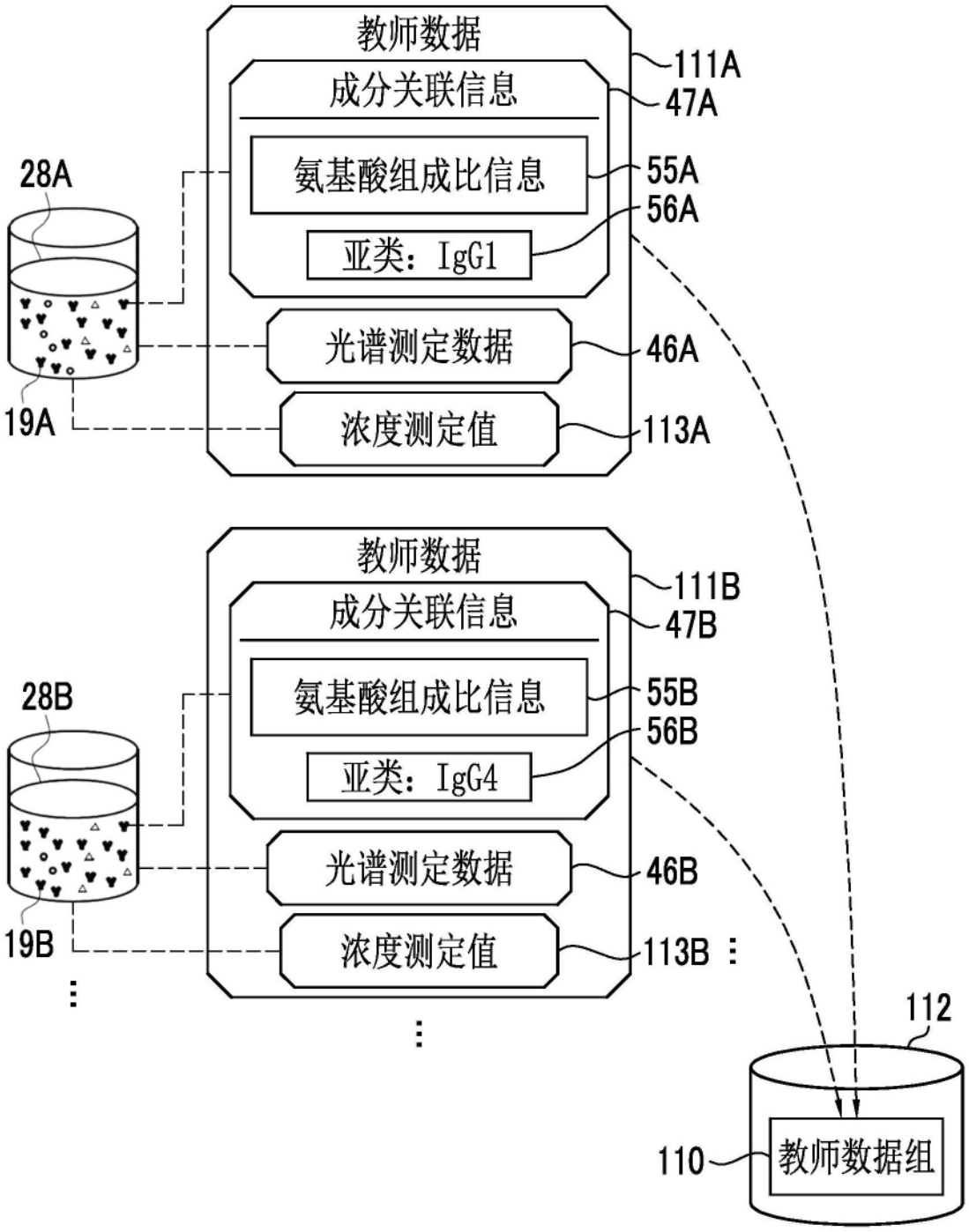

教师数据
111A
成分关联信息
47A
氨基酸组成比信息
55A
56A
亚类：IgG1
光谱测定数据
46A
浓度测定值
113A
28A
19A
教师数据
111B
成分关联信息
47B
氨基酸组成比信息
55B
56B
亚类：IgG4
光谱测定数据
46B
浓度测定值
113B
28B
19B
112
110
教师数据组

图11

学习用输入数据（解释变量） 正解数据（目标变量）

成分关联信息		光谱测定数据	浓度测定值（%）
氨基酸组成比信息	亚类信息		
组氨酸-16.5%、…	IgG1		79
组氨酸-7.2%、…	IgG4		64
组氨酸-18.3%、…	IgG3		82
⋮			

110

图12

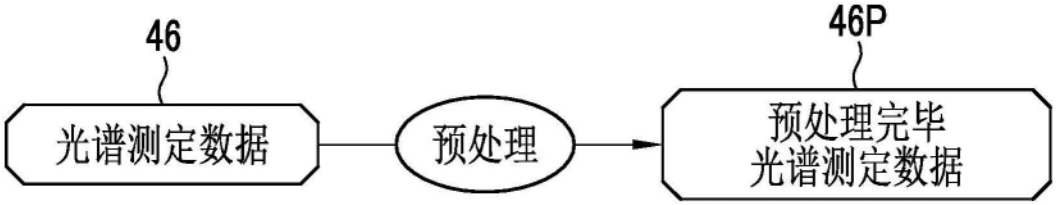


图13

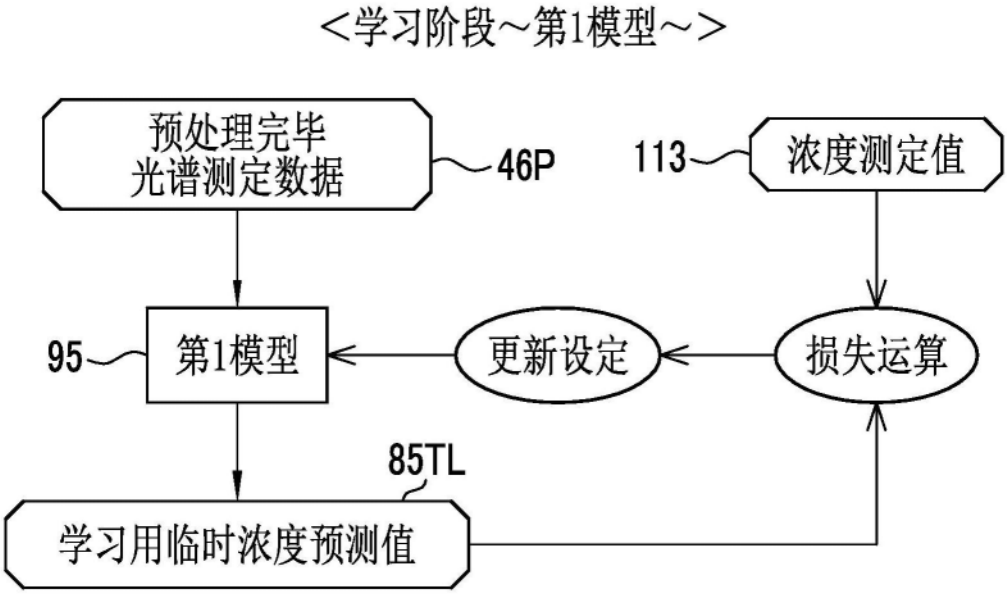


图14

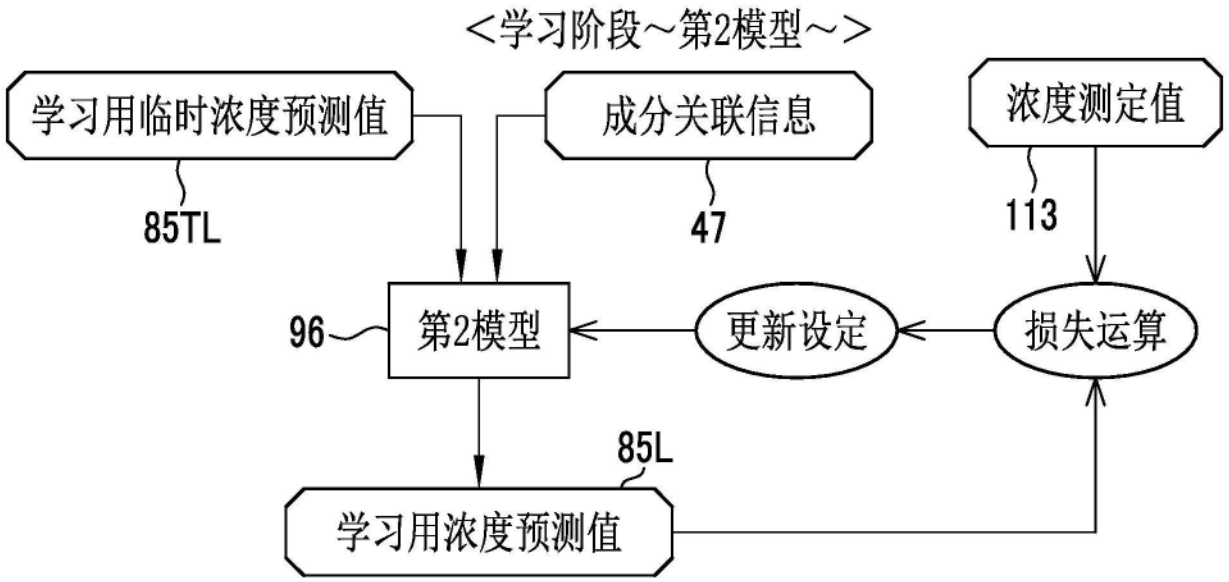

<学习阶段～第2模型～>
学习用临时浓度预测值
85TL
成分关联信息
47
浓度测定值
113
96
第2模型
更新设定
损失运算
85L
学习用浓度预测值

图15

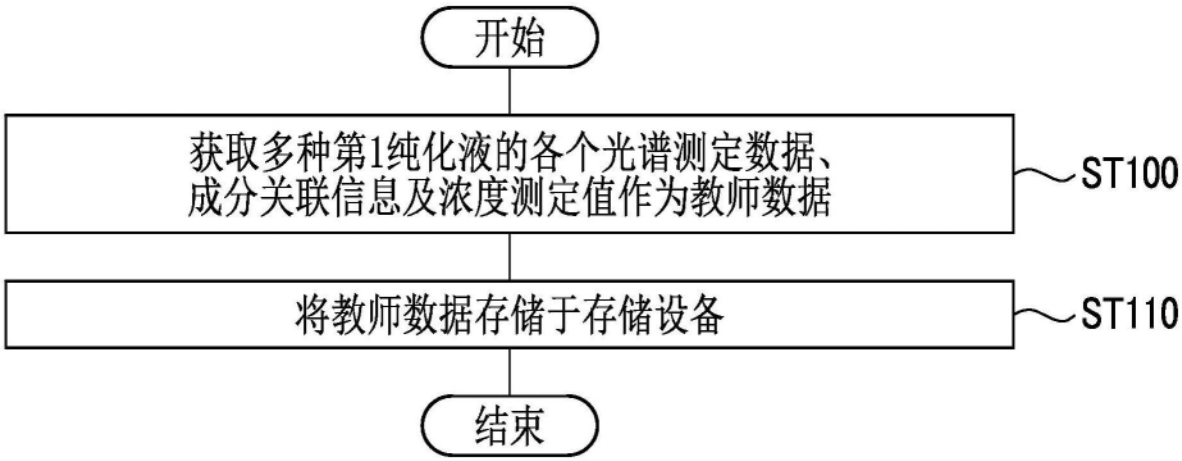

开始
获取多种第1纯化液的各个光谱测定数据、
成分关联信息及浓度测定值作为教师数据
ST100
将教师数据存储于存储设备
ST110
结束

图16

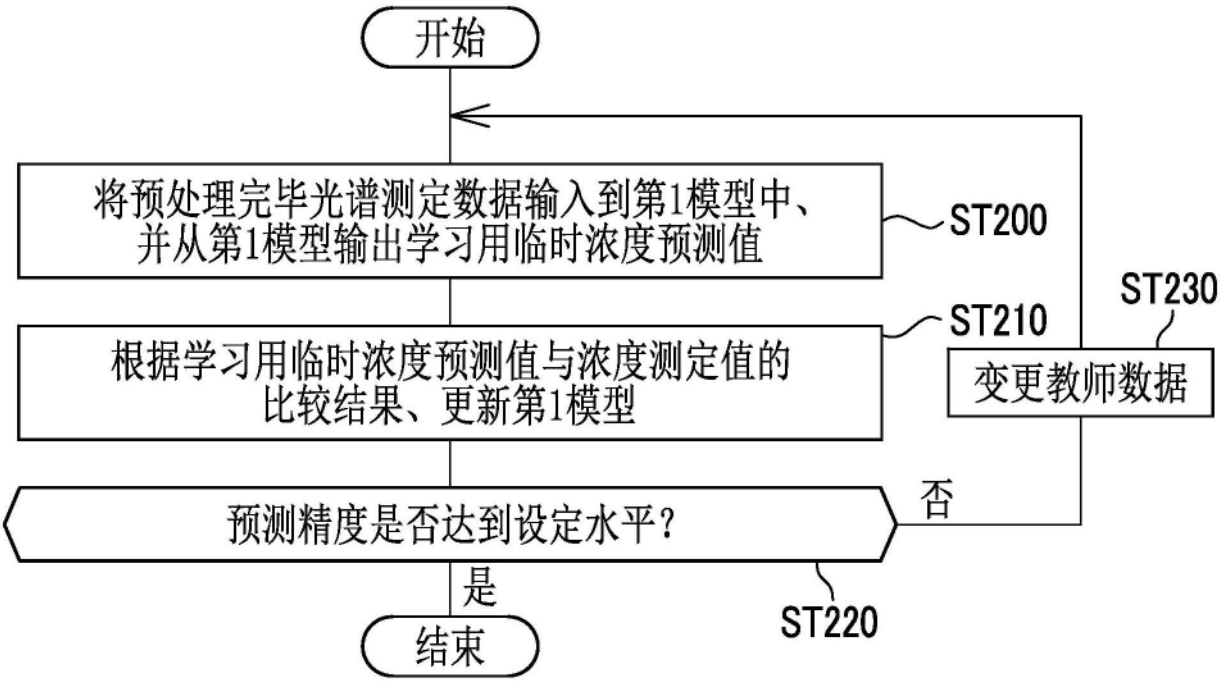

开始
将预处理完毕光谱测定数据输入到第1模型中、并从第1模型输出学习用临时浓度预测值
ST200
根据学习用临时浓度预测值与浓度测定值的比较结果、更新第1模型
ST210
变更教师数据
ST230
预测精度是否达到设定水平？
否
是
ST220
结束

图17

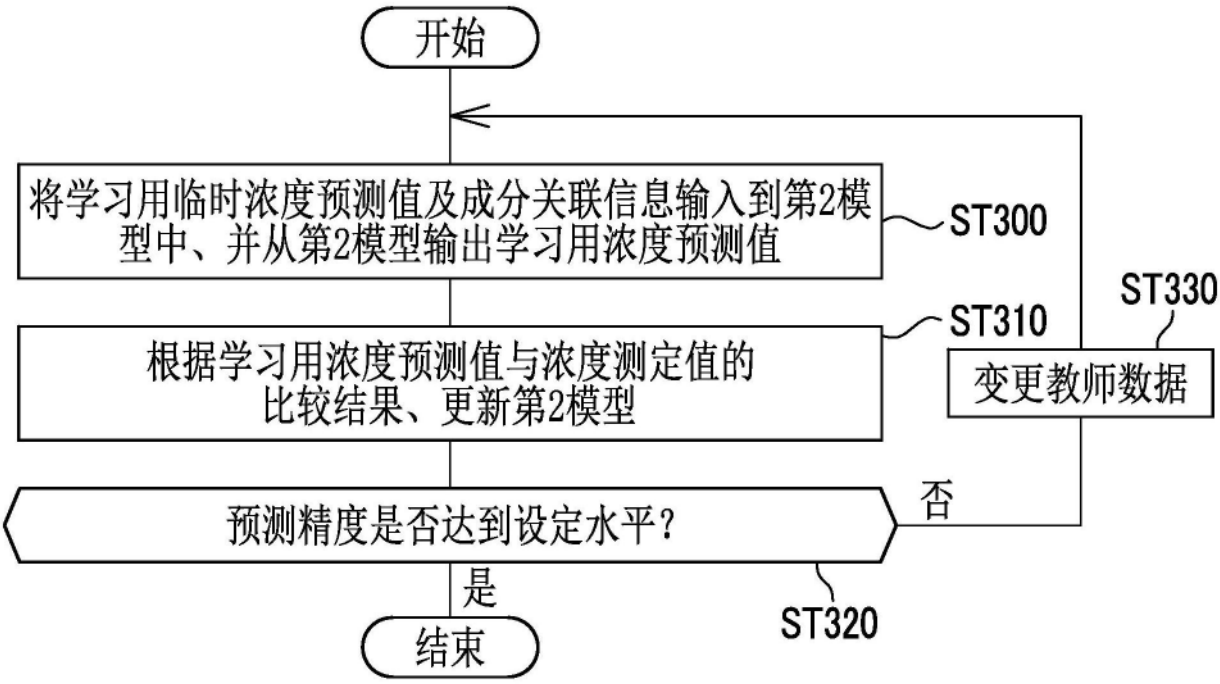

开始
将学习用临时浓度预测值及成分关联信息输入到第2模型中、并从第2模型输出学习用浓度预测值
ST300
根据学习用浓度预测值与浓度测定值的比较结果、更新第2模型
ST310
变更教师数据
ST330
预测精度是否达到设定水平？
否
是
ST320
结束

图18

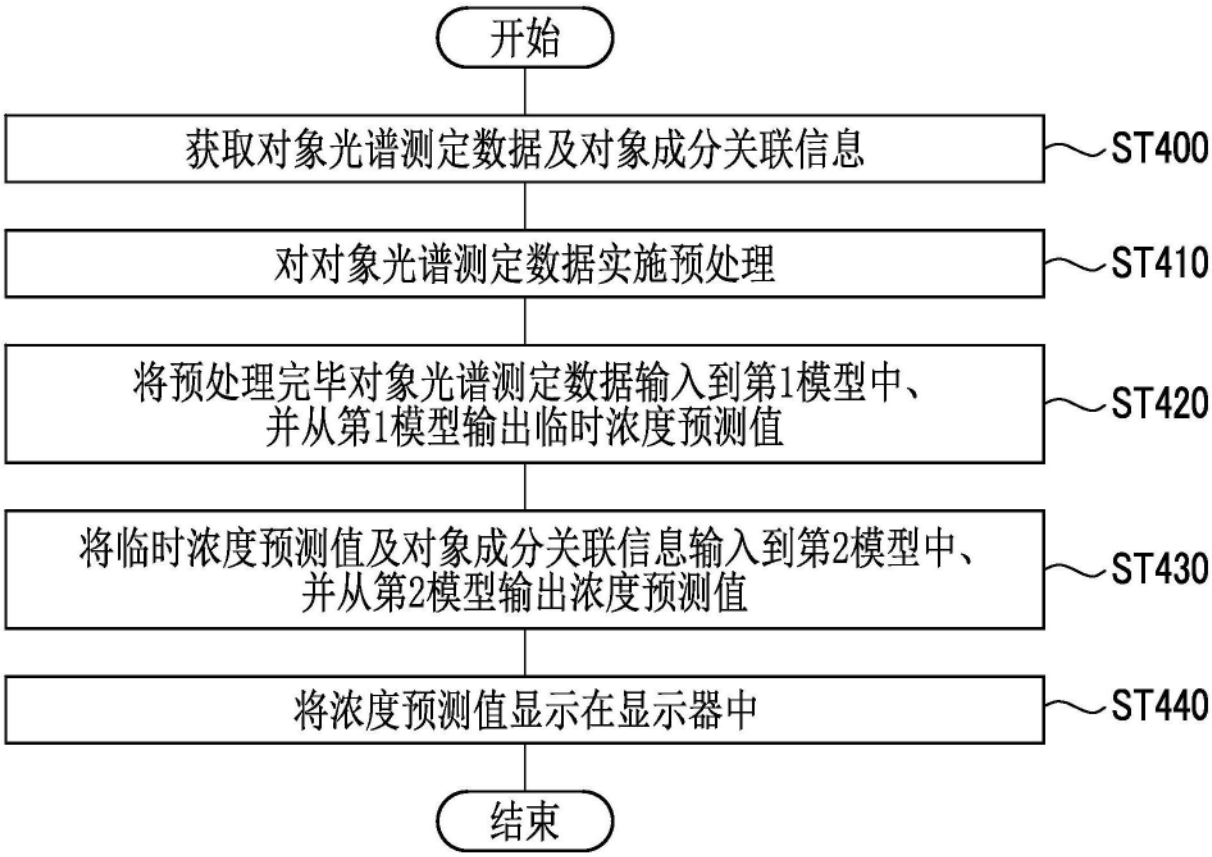

开始
获取对象光谱测定数据及对象成分关联信息
ST400
对对象光谱测定数据实施预处理
ST410
将预处理完毕对象光谱测定数据输入到第1模型中、
并从第1模型输出临时浓度预测值
ST420
将临时浓度预测值及对象成分关联信息输入到第2模型中、
并从第2模型输出浓度预测值
ST430
将浓度预测值显示在显示器中
ST440
结束

图19