



(12)发明专利

(10)授权公告号 CN 109002012 B

(45)授权公告日 2020.05.29

(21)申请号 201810570347.8

(22)申请日 2018.06.05

(65)同一申请的已公布的文献号  
申请公布号 CN 109002012 A

(43)申请公布日 2018.12.14

(30)优先权数据  
2017-112191 2017.06.07 JP

(73)专利权人 发那科株式会社  
地址 日本山梨县

(72)发明人 上田哲士

(74)专利代理机构 北京银龙知识产权代理有限公司 11243  
代理人 范胜杰 王立杰

(51)Int.Cl.

G05B 19/414(2006.01)

(56)对比文件

JP H09258814 A,1997.10.03,  
CN 106393102 A,2017.02.15,  
JP 2017030137 A,2017.02.09,  
CN 101036984 A,2007.09.19,

审查员 贾奇峰

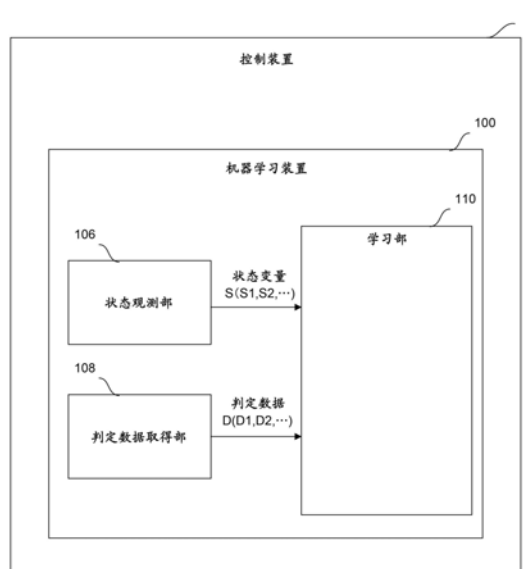
权利要求书2页 说明书11页 附图9页

(54)发明名称

控制装置以及机器学习装置

(57)摘要

本发明提供一种控制装置以及机器学习装置。控制装置所具备的机器学习装置将表示按照示教位置进行机器人的控制时的该示教位置的校正量的示教位置校正量数据和表示进行机器人的控制时的该机器人所具备的各电动机的干扰值的数据作为表示环境的当前状态的状态变量而进行观测,此外,取得表示进行机器人的控制时的该机器人所具备的各电动机的干扰值的适当与否判定结果的判定数据。并且,使用该观测到的状态变量和所取得的判定数据,将机器人的示教位置的校正量与电动机干扰值数据关联起来进行学习。



1. 一种控制装置,其决定按照包含在示教数据中的示教位置进行机器人的控制时的该示教位置的校正量,其特征在于,

该控制装置具备:机器学习装置,其学习按照所述示教位置进行所述机器人的控制时的该示教位置的校正量,

所述机器学习装置具备:

状态观测部,其将示教位置校正量数据、电动机干扰值数据和示教位置数据作为表示环境的当前状态的状态变量而进行观测,其中,所述示教位置校正量数据表示按照所述示教位置进行所述机器人的控制时的该示教位置的校正量,所述电动机干扰值数据表示进行所述机器人的控制时的该机器人所具备的各电动机的干扰值,所述示教位置数据包括基于示教数据的示教位置;

判定数据取得部,其取得表示进行所述机器人的控制时的该机器人所具备的各电动机的干扰值的适当与否判定结果的判定数据;以及

学习部,其使用所述状态变量和所述判定数据,将所述机器人的示教位置的校正量与所述电动机干扰值数据和示教位置关联起来进行学习。

2. 根据权利要求1所述的控制装置,其特征在于,

所述判定数据中,除了进行所述机器人的控制时的该机器人所具备的各电动机的干扰值的适当与否判定结果外,还包括最终所述机器人移动的示教位置的适当与否判定结果、由传感器检测出的值的适当与否判定结果以及按照校正后的示教位置进行所述机器人的控制时的周期时间的适当与否判定结果中的至少任一个。

3. 根据权利要求1或2所述的控制装置,其特征在于,

所述学习部具备:回报计算部,其求出与所述适当与否判定结果相关联的回报;以及价值函数更新部,其使用所述回报,更新表示针对进行所述机器人的控制时的该机器人所具备的各电动机的干扰值的、所述示教位置的校正量的价值的函数。

4. 根据权利要求1或2所述的控制装置,其特征在于,

所述学习部以多层构造对所述状态变量和所述判定数据进行运算。

5. 根据权利要求1或2所述的控制装置,其特征在于,

该控制装置还具备:决策部,其根据所述学习部的学习结果,输出基于按照所述示教位置进行所述机器人的控制时的该示教位置的校正量的指令值。

6. 根据权利要求1或2所述的控制装置,其特征在于,

所述学习部使用针对多个机器人的每一个得到的所述状态变量以及所述判定数据,来学习按照该多个机器人的每一个的所述示教位置进行所述机器人的控制时的该示教位置的校正量。

7. 根据权利要求1或2所述的控制装置,其特征在于,

所述机器学习装置存在于云服务器中。

8. 一种机器学习装置,其学习按照包含在示教数据中的示教位置进行机器人的控制时的该示教位置的校正量,其特征在于,该机器学习装置具备:

状态观测部,其将示教位置校正量数据、电动机干扰值数据和示教位置数据作为表示环境的当前状态的状态变量而进行观测,其中,所述示教位置校正量数据表示按照所述示教位置进行所述机器人的控制时的该示教位置的校正量,所述电动机干扰值数据表示进行

所述机器人的控制时的该机器人所具备的各电动机的干扰值,所述示教位置数据包括基于示教数据的示教位置;

判定数据取得部,其取得表示进行所述机器人的控制时的该机器人所具备的各电动机的干扰值的适当与否判定结果的判定数据;以及

学习部,其使用所述状态变量和所述判定数据,将所述机器人的示教位置的校正量与所述电动机干扰值数据和示教位置关联起来进行学习。

## 控制装置以及机器学习装置

### 技术领域

[0001] 本发明涉及一种控制装置以及机器学习装置,尤其涉及将示教位置最佳化的控制装置以及机器学习装置。

### 背景技术

[0002] 一般的工业用机器人按照预先制作的动作程序被驱动,或者,以通过由示教盒(Teaching pendant)等预先示教的示教点的方式驱动机器人。即,沿着预先决定的轨道驱动机器人。作为与机器人的示教动作有关的现有技术,例如在日本特开平02-284886号公报中公开了如下发明:针对来自控制装置的定位目标值与实际的位置数据的偏差进行统计处理,决定用于消除该偏差的校正量。

[0003] 以工件的设置、取出已加工完的工件、向加工机安装工具等为目的,利用工业用机器人。通过工业用机器人将工件设置于加工机的情况下,例如如上述那样,通过对机器人示教示教点,来示教从机器人握持工件起至将该工件设置于加工区域的夹具为止的动作。例如,如图9所示,在凸型的夹具上设置筒状工件的情况下,在使工件的孔平行于夹具方向的状态下,从接近位置(Approach Position)向目标位置与夹具方向水平地下降,从而可以将工件设置于夹具上。

[0004] 然而,对机器人示教动作的作业者是示教初学者的情况下,无法准确定位到接近位置,或接近位置中的工件的方向倾斜,或从接近位置向目标位置下降的方向发生偏差,从而工件被夹具卡住,或工件和夹具接触而产生摩擦,从而有时在握持工件的机器人的腕轴等产生干扰(负荷)。进行这样的示教时,每当根据被示教的动作反复进行工件的设置时,向机器人的各关节施加负荷,而成为在持续运转中发生故障、或机器人发生故障的原因。

### 发明内容

[0005] 因此,本发明的目的在于,提供一种能够使示教位置最佳化的控制装置以及机器学习装置以便不会向机器人的机械手施加多余的干扰。

[0006] 在本发明的控制装置中,机器学习针对向驱动机器人的各关节的电动机施加的干扰的、该机器人的示教位置的校正量,根据机器学习到的结果,对示教位置进行校正并控制以便在机器人向示教位置移动时抑制干扰,从而解决上述问题。

[0007] 本发明的一方式的控制装置决定按照包含在示教数据中的示教位置进行机器人的控制时的该示教位置的校正量,并具备学习按照所述示教位置进行所述机器人的控制时的该示教位置的校正量的机器学习装置。并且,所述机器学习装置具备:状态观测部,其将示教位置校正量数据和电动机干扰值数据作为表示环境的当前状态的状态变量而进行观测,其中,示教位置校正量数据表示按照所述示教位置进行所述机器人的控制时的该示教位置的校正量,电动机干扰值数据表示进行所述机器人的控制时的该机器人所具备的各电动机的干扰值;判定数据取得部,其取得表示进行所述机器人的控制时的该机器人所具备的各电动机的干扰值的适当与否判定结果的判定数据;以及学习部,其使用所述状态变量

和所述判定数据,将所述机器人的示教位置的校正量与所述电动机干扰值数据关联起来进行学习。

[0008] 所述状态观测部还将包含基于示教数据的示教位置的示教位置数据作为所述状态变量而进行观测,所述学习部将所述状态观测部观测到的数据与所述电动机干扰值数据关联起来进行学习。

[0009] 所述判定数据中,除了进行所述机器人的控制时的该机器人所具备的各电动机的干扰值的适当与否判定结果外,还包括最终所述机器人移动的示教位置的适当与否判定结果、传感器检测出的值的适当与否判定结果、以及按照校正后的示教位置进行所述机器人的控制时的周期时间的适当与否判定结果中的至少任一个。

[0010] 所述学习部具备:回报计算部,其求出与所述适当与否判定结果相关联的回报;以及价值函数更新部,其使用所述回报,更新表示针对进行所述机器人的控制时的该机器人所具备的各电动机的干扰值的、所述示教位置的校正量的价值的函数。

[0011] 所述学习部以多层构造对所述状态变量和所述判定数据进行运算。

[0012] 所述控制装置还具备:决策部,其根据所述学习部的学习结果,输出基于按照所述示教位置进行所述机器人的控制时的该示教位置的校正量的指令值。

[0013] 所述学习部使用针对多个机器人的每一个得到的所述状态变量以及所述判定数据,来学习按照该多个机器人的每一个的、所述示教位置进行所述机器人的控制时的该示教位置的校正量。

[0014] 所述机器学习装置能够存在于云服务器中。

[0015] 本发明的其他方式的机器学习装置学习按照包含在示教数据中的示教位置进行机器人的控制时的该示教位置的校正量。该机器学习装置具备:状态观测部,其将示教位置校正量数据和电动机干扰值数据作为表示环境的当前状态的状态变量而进行观测,其中,示教位置校正量数据表示按照所述示教位置的进行所述机器人的控制时的该示教位置的校正量,电动机干扰值数据表示进行所述机器人的控制时的该机器人所具备的各电动机的干扰值;判定数据取得部,其取得表示进行所述机器人的控制时的该机器人所具备的各电动机的干扰值的适当与否判定结果的判定数据;以及学习部,其使用所述状态变量和所述判定数据,将所述机器人的示教位置的校正量与所述电动机干扰值数据关联起来进行学习。

[0016] 根据本发明,可以根据学习结果对机器人的示教位置进行校正,来抑制向驱动该机器人的各关节的电动机施加的干扰值,防止机器人的故障发生等的伤害。

## 附图说明

[0017] 图1是第一实施方式的控制装置的概要性的硬件结构图。

[0018] 图2是第一实施方式的控制装置的概要性的功能框图。

[0019] 图3是表示控制装置的一方式的概要性的功能框图。

[0020] 图4是表示机器学习方法的一方式的概要性的流程图。

[0021] 图5A是说明神经元的图。

[0022] 图5B是说明神经网络的图。

[0023] 图6是第二实施方式的控制装置的概要性的功能框图。

- [0024] 图7是表示具备多个机器人的系统的一方式的概要性的功能框图。
- [0025] 图8是表示具备多个机器人的系统的其他方式的概要性的功能框图。
- [0026] 图9是说明现有技术中的示教的问题点的图。

### 具体实施方式

[0027] 图1是表示第一实施方式的控制装置和由该控制装置控制的机床的主要部分的概要性的硬件结构图。

[0028] 控制装置1例如能够安装为,对进行工件的设置、已加工完的工件的取出、向加工机安装工件等的工业用机器人(未图示)进行控制的控制装置。本实施方式的控制装置1所具备的CPU 11是对控制装置1进行整体控制的处理器。CPU11经由总线20读出存储在ROM12中的系统程序,并按照该系统程序对控制装置1整体进行控制。RAM 13中临时存储临时的计算数据、显示数据以及操作者经由后述的示教操作盘60输入的各种数据等。

[0029] 非易失性存储器14构成为,例如通过未图示的电池进行备份等,而即使控制装置1的电源被断开也保持存储状态的存储器。非易失性存储器14中存储有经由接口19从示教操作盘60输入的示教数据、经由未图示的接口输入的机器人控制用程序等。存储在非易失性存储器14中的程序、各种数据也可以在执行时/利用时在RAM13中展开。此外,预先在ROM12中写入用于执行与机器人控制、示教位置的示教相关的处理等的各种系统程序(包括用于控制与后述的机器学习装置100的交换的系统程序)。

[0030] 示教操作盘60是具备显示器、手柄、硬键盘等的手动数据输入装置,经由接口19接受来自控制装置1的信息并显示,并且将从手柄、硬键盘等输入的脉冲、指令、各种数据传递给CPU 11。

[0031] 用于控制机器人所具备的关节等的轴的轴控制电路30接受来自CPU 11的轴的移动指令量,并将轴的指令输出到伺服放大器40。伺服放大器40接受该指令,驱动使机器人所具备的轴移动的伺服电动机50。轴的伺服电动机50内置有位置/速度检测器,并将来自该位置/速度检测器的位置/速度反馈信号反馈给轴控制电路30,进行位置/速度的反馈控制。此外,在图1的硬件结构图中,轴控制电路30、伺服放大器40、伺服电动机50仅示出了一个,但实际上按照成为控制对象的机器人所具备的轴的数量进行准备。例如,在具备六个轴的机器人的情况下,与六个轴对应地分别准备轴控制电路30、伺服放大器40和伺服电动机50。

[0032] 接口21是用于连接控制装置1与机器学习装置100的接口。机器学习装置100具备对机器学习装置100整体进行控制的处理器101、存储有系统程序等的ROM102、用于进行与机器学习有关的各处理中的临时存储的RAM103、以及用于存储学习模型等的非易失性存储器104。机器学习装置100能够经由接口21观测可通过控制装置1获取的各信息(伺服电动机50的位置信息和电流值、存储于RAM13等的执行中的程序、与示教位置有关的设定信息等)。此外,控制装置1接受从机器学习装置100输出的用于控制伺服电动机50、机器人的周边装置的指令,并进行基于程序、示教数据的机器人控制命令的修正等。

[0033] 图2是第一实施方式的控制装置1和机器学习装置100的概要性的功能框图。

[0034] 机器学习装置100包括用于通过所谓的机器学习,自学针对向驱动机器人的各关节的电动机施加的干扰值的、该机器人的示教位置的校正量的软件(学习算法等)以及硬件(处理器101等)。控制装置1所具备的机器学习装置100学习的内容相当于表示向驱动机器

人的各关节的电动机施加的干扰值与该机器人的示教位置的校正量的相关性的模型构造。

[0035] 如图2的功能模块所示那样,控制装置1所具备的机器学习装置100具备:状态观测部106,其对状态变量S进行观测,该状态变量S表示包含示教位置校正量数据S1和电动机干扰值数据S2的环境的当前状态,其中,示教位置校正量数据S1表示按照包含在示教数据中的示教位置对机器人进行控制时该机器人所具备的各电动机的示教位置的校正量,电动机干扰值数据S2表示按照包含在示教数据中的示教位置对机器人进行控制时该机器人所具备的各电动机的干扰值;判定数据取得部108,其取得表示按照校正后的示教位置对机器人进行控制时各电动机的干扰值的适当与否判定结果的判定数据D;以及学习部110,其使用状态变量S和判定数据D,将示教位置校正量数据S1与各电动机的干扰值关联起来进行学习。

[0036] 状态观测部106例如既可以构成为处理器101的一功能,也可以构成为用于使处理器101发挥功能的存储于ROM102的软件。状态观测部106观测的状态变量S中的示教位置校正量数据S1,可以作为针对包含在示教数据中的示教位置(机器人所具备的各电动机的坐标值的组合)的校正量而取得。示教位置校正量数据S1既可以是最终机器人移动的示教位置(例如,在图9的例子中所说的用夹具夹住工件时的最终的位置)的校正量,除此之外也可以包括最终机器人移动前的示教位置即接近位置的校正量。此外,示教位置校正量数据S1既可以使用向示教位置移动过程中对各个电动机施加的干扰值的最大值,也可以使用每隔预定周期(例如,10ms)取得向示教位置移动过程中对各个电动机施加的干扰值而得的系列值。

[0037] 示教位置校正量数据S1例如可以使用由熟练的作业者告知,而对控制装置1施加的机器人所具备的各电动机的示教位置的校正量。另外,在此所说的机器人所具备的各电动机的示教位置的校正量包括对电动机进行校正的方向(正负值)。此外,示教位置校正量数据S1在学习推进至某种程度的阶段,可使用机器学习装置100根据学习部110的学习结果在上一个学习周期决定的机器人所具备的各电动机的示教位置的校正量,在这样的情况下,机器学习装置100在每个学习周期将所决定的机器人所具备的各电动机的示教位置的校正量临时存储于RAM103,状态观测部106从RAM103取得在上一个学习周期由机器学习装置100决定的机器人所具备的各电动机的示教位置的校正量。

[0038] 此外,状态变量S中的电动机干扰值数据S2,例如可以使用根据按照包含在示教数据中的示教位置对机器人进行控制时的伺服电动机50中流动的电流值等间接求出的负荷值等。电动机干扰值数据S2也可以使用组入到电动机的扭矩测量仪等来直接求出。

[0039] 判定数据取得部108例如既可以构成为处理器101的一个功能,也可以构成为用于使处理器101发挥功能的存储于ROM102的软件。判定数据取得部108可以将针对按照校正后的示教位置对机器人进行控制时的各电动机的干扰值的、适当与否判定值D1作为判定数据D而使用。判定数据取得部108可以使用与状态观测部106观测电动机干扰值数据S2时同样的手段,来取得针对按照校正后的示教位置对机器人进行控制时的各电动机的干扰值。判定数据D是表示按照状态变量S下的示教位置进行机器人的控制时的结果的指标。

[0040] 以学习部110的学习周期进行考虑的情况下,同时输入到学习部110的状态变量S成为基于取得了判定数据D的一个学习周期前的数据的变量。这样,在控制装置1所具备的机器学习装置100推进学习的期间,在环境下,反复实施电动机干扰值数据S2的获取,按照

根据示教位置校正量数据S1校正后的示教位置的机器人的控制的执行、判定数据D的获取。

[0041] 学习部110例如既可以构成为处理器101的一个功能,也可以构成为用于使处理器101发挥功能的存储于ROM102的软件。学习部110按照被称为机器学习的任意的学习算法,学习针对按照包含在示教数据中的示教位置进行机器人的控制时的机器人所具备的各电动机的干扰值的、示教位置校正量数据S1。学习部110可以反复执行基于包含上述的状态变量S和判定数据D的数据集合的学习。

[0042] 按照包含在示教数据中的示教位置进行机器人的控制时的针对机器人所具备的各电动机的干扰值的、示教位置校正量数据S1的学习周期的反复中,将状态变量S中的电动机干扰值数据S2作为如上所述按照包含在一个学习周期前取得的示教数据中的示教位置进行机器人的控制时的该机器人所具备的各电动机的干扰值,将状态变量S中的示教位置校正量数据S1作为按照包含在基于上次为止的学习结果得到的示教数据中的示教位置进行机器人的控制时的该机器人所具备的各电动机的干扰值,此外,将判定数据D作为进行了基于示教位置校正量数据S1的示教位置的校正的状态下按照本次的学习周期中的(校正后的)示教位置进行机器人的控制时的针对该机器人所提供的各电动机的干扰值的适当与否判定结果。

[0043] 通过反复这样的学习周期,学习部110能够自动识别暗示如下内容的特征:按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该机器人所具备的各电动机的干扰值(电动机干扰值数据S2)和针对该示教位置的校正量的相关性。学习算法的开始时,实质上电动机干扰值数据S2和针对示教位置的校正量的相关性是未知的,但学习部110随着学习的推进逐渐识别特征来解析相关性。电动机干扰值数据S2和针对示教位置的校正量的相关性被解释到某种可信赖的水准时,在针对当前状态(也就是说,按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该机器人所具备的各电动机的干扰值),为了进行应将该示教位置的校正量设为何种程度的值的行为的选择(也就是是决策)而使用学习部110反复输出的学习结果。也就是说,学习部110随着学习算法的推进,能够使按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该机器人所具备的各电动机的干扰值与对于该状态应将机器人的示教位置的校正量设为何种程度的值的行为的相关性逐渐接近最佳解。

[0044] 如上所述,在控制装置1所具备的机器学习装置100中,使用由状态观测部106观测到的状态变量S和由判定数据取得部108取得的判定数据D,学习部110按照机械学习算法,对机器人的示教位置的校正量进行学习。状态变量S由示教位置校正量数据S1和电动机干扰值数据S2这样的难以受到干扰的影响的数据构成,此外,判定数据D通过取得由控制装置1实测的伺服电动机50的电流值等而被唯一地求出。因此,根据控制装置1所具备的机器学习装置100,通过使用学习部110的学习结果,不依赖于运算、估算而能够自动且准确地求出按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该机器人所具备的各电动机的干扰值所对应的机器人的该示教位置的校正量。

[0045] 并且,若能够不依赖于运算、估算而自动地求出按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该示教位置的校正量,则仅通过掌握按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该机器人所具备的各电动机的干扰值(电动机干扰值数据S2),就能够迅速决定机器人的该示教位置的校正量的恰当的值。因



此,能够有效地决定机器人的示教位置的校正量。

[0046] 在控制装置1所具备的机器学习装置100的第1变形例中,作为状态变量S,状态观测部106还可以观测包含基于示教数据的示教位置的示教位置数据S3。示教位置数据S3可以作为包含在示教数据中的示教位置(机器人所具备的各电动机的坐标值的组合)而取得。示教位置数据S3既可以是最终机器人移动的示教位置(例如,在图9的例子中所说的用夹具夹住工件时的最终位置),除此之外,也可以包括作为最终机器人移动前的示教位置的接近位置。

[0047] 根据上述变形例,机器学习装置100可以学习针对按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该机器人所具备的各电动机的干扰值以及该示教位置双方的、针对该示教位置的校正量。这样,在基于示教数据的机器人的控制中,可以一边考虑该机器人的姿势,一边将示教位置的校正量变更为适当的值,因此能够更适当地抑制向机器人所具备的电动机的干扰。

[0048] 作为控制装置1所具备的机器学习装置100的第2变形例,作为判定数据D,判定数据取得部108除了可以使用针对按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该机器人所具备的各电动机的干扰值的适当与否判定值D1外,还可以使用最终机器人移动的示教位置的适当与否判定值D2、由振动传感器和声音传感器等传感器检测出的值的适当与否判定值D3、按照校正后的示教位置进行机器人的控制时的周期时间的评价值D4等。

[0049] 根据上述变形例,机器学习装置100可以考虑学习针对按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该机器人所具备的各电动机的干扰值的、机器人的该示教位置的校正量时,可以考虑示教位置的偏移程度、振动或异音的发生、周期时间的上升等。

[0050] 作为控制装置1所具备的机器学习装置100的第3变形例,学习部110可以使用针对进行同一作业的多个机器人的每一个获得的状态变量S和判定数据D,来学习按照这些包含在机器人的示教数据中的示教位置进行控制时的该示教位置的校正量。根据该结构,可以增加包含在一定时间获得的状态变量S和判定数据D的数据集合的量,因此能够以更多样的数据集合为输入,来提高按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该示教位置的校正量的学习速度和可靠性。

[0051] 在具有上述结构的机器学习装置100中,并不特别限定学习部110执行的学习算法,作为机器学习可采用公知的学习算法。图3是图1所示的控制装置1的一个方式,作为学习算法的一例示出了具备执行强化学习的学习部110的结构。

[0052] 强化学习是如下的方法:观测学习对象所在的环境的当前状态(即输入),并且在当前状态下执行预定的行为(即输出),试错性地反复进行针对该行为给出何种回报的周期,将回报总计被最大化的方案(在本申请的机器学习装置中,按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该示教位置的校正量)作为最佳解而进行学习。

[0053] 在图3所示的控制装置1所具备的机器学习装置100中,学习部110具备回报计算部112和价值函数更新部114,其中,回报计算部112求出与按照根据状态变量S校正后的示教位置进行该机器人的控制时的该机器人所具备的各电动机的干扰值的适当与否判定结果(相当于在取得了状态变量S的下次的学习周期中所使用的判定数据D)相关联的回报R,价

值函数更新部114使用回报R来更新表示按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该示教位置的校正量的价值的函数Q。学习部110通过价值函数更新部114反复进行函数Q的更新,来学习针对按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该机器人所具备的各电动机的干扰值的、机器人的该示教位置的校正量。

[0054] 对学习部110所执行的强化学习的算法的一例进行说明。该例子的算法是作为Q学习(Q-learning)而公知的算法,是以行为主体的状态s和在该状态s下该行为主体可选择的行为a为独立变量,来学习表示在状态s下选择了行为a时的行为的价值的函数Q(s,a)的方法。通过在状态s下选择价值函数Q最高的行为a来得到最佳解。在状态s和行为a的相关性未知的状态下开始Q学习,反复进行在任意的状态s下选择各种行为a的试错,由此反复更新价值函数Q,而接近最佳解。在此,作为在状态s下选择了行为a的结果而环境(即状态s)变化时,得到与该变化对应的回报(即行为a的加权)r,对学习进行引导以便选择得到更高回报r的行为a,从而能够使价值函数Q在较短时间内接近最佳解。

[0055] 一般,可以如下式(1)所示地表示价值函数Q的更新式。在式(1)中, $s_t$ 以及 $a_t$ 为各个时刻t的状态以及行为,状态根据行为 $a_t$ 变化为 $s_{t+1}$ 。 $r_{t+1}$ 为通过将状态从 $s_t$ 变化为 $s_{t+1}$ 而获得的回报。 $\max Q$ 项表示进行了(在时刻t被认为)在时刻t+1成为最大的函数Q的行为a时的Q。 $\alpha$ 以及 $\gamma$ 分别是学习系数以及折扣率,在 $0 < \alpha \leq 1$ 、 $0 < \gamma \leq 1$ 的范围内被任意设定。

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha (r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t))$$

[0056]

..... (1)

[0057] 在学习部110执行Q学习的情况下,状态观测部106观测到的状态变量S以及判定数据取得部108取得的判定数据D相当于更新式的状态s,应如何决定针对当前状态(即,按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该机器人所具备的各电动机的干扰值)的机器人的该示教位置的校正量的行为相当于更新式的行为a,回报计算部112求出的回报R相当于更新式的回报r。因此,价值函数更新部114通过使用回报R的Q学习,来反复更新表示针对当前状态的机器人的示教位置的校正量的价值的函数Q。

[0058] 回报计算部112求出的回报R,例如在决定了按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该示教位置的校正量后,按照基于所决定的示教位置的校正量进行了校正后的示教位置来进行机器人的控制时,在该机器人的动作的适当与否判定结果被判定为“适当”时(例如,该机器人所具备的各电动机的干扰值在可允许的范围内)的情况下,最终该机器人移动的示教位置在可允许的范围内(例如,该机器人所具备的各电动机的干扰值在可允许的范围内)的情况下,设为正的回报R,在该机器人的动作的适当与否判定结果被判定为“否”时(例如,该机器人所具备的各电动机的干扰值在可允许的范围外的情况下,最终该机器人移动的示教位置在可允许的范围外的情况下),设为负的回报R。

[0059] 在上述的回报R的决定中,正负的回报R的绝对值既可以彼此相同也可以彼此不同。此外,作为判定的条件,也可以将包含在判定数据D中的多个值进行组合来进行判定。

[0060] 此外,不仅可以将机器人的动作的适当与否判定结果设定为“适当”和“否”,也可以设定为多个阶段。作为例子,在机器人所具备的各电动机的干扰值的允许范围的最大值

为 $V_{\max}$ 的情况下,机器人所具备的各电动机的干扰值 $V$ 在 $0 \leq V < V_{\max}/5$ 时给予回报 $R=5$ , $V_{\max}/5 \leq V < V_{\max}/2$ 时给予回报 $R=2$ , $V_{\max}/2 \leq V \leq V_{\max}$ 时给予回报 $R=1$ 。并且,在学习的初期阶段也可以将 $V_{\max}$ 设定得较大,随着学习的推进将 $V_{\max}$ 缩小。

[0061] 价值函数更新部114可具有将状态变量 $S$ 、判定数据 $D$ 、回报 $R$ 与用函数 $Q$ 表示的行为价值(例如数值)关联起来整理而得的行为价值表。在该情况下,价值函数更新部114更新函数 $Q$ 的行为等同于价值函数更新部114更新行为价值表的行为。 $Q$ 学习开始时,环境的当前状态与按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该示教位置的校正量的相关性是未知的,因此在行为价值表中,以与随机确定的行为价值的值(函数 $Q$ )相关联的方式准备各种状态变量 $S$ 、判定数据 $D$ 和回报 $R$ 。另外,回报计算部112若知道判定数据 $D$ 则可立即算出与之对应的回报 $R$ ,并将算出的值 $R$ 写入到行为价值表中。

[0062] 使用按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该机器人所具备的各电动机的干扰值的适当与否判定结果所对应的回报 $R$ 来推进 $Q$ 学习时,向选择可得到更高回报 $R$ 的行为的方向引导学习,根据作为在当前状态下执行所选择的行为的结果而变化的环境的状态(即状态变量 $S$ 和判定数据 $D$ )重写针对当前状态下进行的行为的行为价值的值(函数 $Q$ )来更新行为价值表。通过反复进行该更新,显示于行为价值表的行为价值的值(函数 $Q$ )被重写为越是适当的行为越是大的值。这样,使未知环境的当前状态(按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该机器人所具备的各电动机的干扰值)和与之对应的行为(机器人的该示教位置的校正量的决定)的相关性逐渐明确。也就是说,通过行为价值表的更新,使按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该机器人所具备的各电动机的干扰值与机器人的该示教位置的校正量的决定之间的关系逐渐接近最佳解。

[0063] 参照图4,进一步说明学习部110所执行的上述的 $Q$ 学习的流程(即,机器学习方法的一方式)。

[0064] 首先,在步骤SA01中,价值函数更新部114一边参照该时间点的行为价值表,一边随机选择机器人的该示教位置的校正量作为在状态观测部106所观测到的状态变量 $S$ 所表示的当前状态下进行的行为。接着,价值函数更新部114在步骤SA02中取入由状态观测部106观测到的当前状态的状态变量 $S$ ,在步骤SA03中,取入由判定数据取得部108取得的当前状态的判定数据 $D$ 。接着,价值函数更新部114在步骤SA04中根据判定数据 $D$ 判断机器人的该示教位置的校正量是否适当,判断为适当时,在步骤SA05中,将回报计算部112求出的正的回报 $R$ 适用于函数 $Q$ 的更新式,接着,在步骤SA06中,使用当前状态下的状态变量 $S$ 以及判定数据 $D$ 、回报 $R$ 、行为价值的值(更新后的函数 $Q$ )来更新行为价值表。在步骤SA04中,在判断为机器人的该示教位置的校正量不适当时,在步骤SA07中,将回报计算部112求出的负的回报 $R$ 适用于函数 $Q$ 的更新式,接着在步骤SA06中,使用当前状态下的状态变量 $S$ 以及判定数据 $D$ 、回报 $R$ 、行为价值的值(更新后的函数 $Q$ )来更新行为价值表。学习部110通过反复进行步骤SA01~SA07来反复更新行为价值表,使机器人的该示教位置的校正量的学习推进。另外,针对包含在判定数据 $D$ 中的各个数据,执行从步骤SA04至步骤SA07为止的求出回报 $R$ 的处理以及价值函数的更新处理。

[0065] 推进上述的强化学习时,例如代替 $Q$ 学习,可使用神经网络。图5A示意性地表示神经元的模型。图5B示意性地表示将图5A所示的神经元组合而构成的三层神经网络的模型。

神经网络例如可以由模拟了神经元模型的运算装置、存储装置等构成。

[0066] 图5A所示的神经元输出针对多个输入x(在此,作为一个例子是输入 $x_1$ ~输入 $x_3$ )的结果y。各输入 $x_1$ ~ $x_3$ 被乘以与该输入x对应的权重w( $w_1$ ~ $w_3$ )。由此,神经元输出通过下式(2)表现的结果y。另外,在式(2)中,输入x、结果y和权重w均为向量。此外, $\theta$ 是偏置, $f_k$ 是激活函数。

$$[0067] \quad y = f_k \left( \sum_{i=1}^n x_i w_i - \theta \right) \quad \dots\dots (2)$$

[0068] 图5B所示的三层神经网络从左侧输入多个输入x(在此,作为一个例子为输入 $x_1$ ~输入 $x_3$ ),从右侧输出结果y(在此,作为一个例子是结果 $y_1$ ~结果 $y_3$ )。在图5B的例子中,输入 $x_1$ 、 $x_2$ 、 $x_3$ 被乘以与之对应的权重(统称为 $w_1$ )后,分别被输入到三个神经元N11、N12、N13。

[0069] 在图5B中,神经元N11~N13的各个输出统称为 $z_1$ 。可将 $z_1$ 视为提取输入向量的特征量而得的特征向量。在图5B的例子中,对各个特征向量 $z_1$ 乘以对应的权重(统称为 $w_2$ )后,各个特征向量 $z_1$ 均被输入到两个神经元N21、N22。特征向量 $z_1$ 表示权重 $w_1$ 与权重 $w_1$ 之间的特征。

[0070] 在图5B中,将神经元N21~N22的各个输出统称为 $z_2$ 。可将 $z_2$ 视为提取特征向量 $z_1$ 的特征量而得的特征向量。在图5B的例子中,特征向量 $z_2$ 被乘以与之对应的权重(统称为 $w_3$ )后,分别被输入到三个神经元N31、N32、N33。特征向量 $z_2$ 表示权重 $w_2$ 与权重 $w_3$ 之间的特征。最后,神经元N31~N33分别输出结果 $y_1$ ~结果 $y_3$ 。

[0071] 另外,也可以利用使用了构成三层以上的层的神经网络的、所谓的深度学习方法。

[0072] 在控制装置1所具备的机器学习装置100中,以状态变量S和判定数据D为输入x,学习部110进行遵照上述的神经网络的多层构造的运算,从而输出机器人的示教位置的校正量(结果y)。此外,在控制装置1所具备的机器学习装置100中,将神经网络用作强化学习中的价值函数,以状态变量S和行为a为输入x,学习部110进行遵照上述的神经网络的多层构造的运算,从而输出该状态下的该行为的价值(结果y)。另外,神经网络的动作模式有学习模式和价值预测模式,例如在学习模式中使用学习数据组来学习权重w,并使用所学习的权重w在价值预测模式下进行行为价值判断。另外,在价值预测模式中,也可以进行检测、分类、推论等。

[0073] 可以将上述的控制装置1的结构记述为处理器101执行的机器学习方法(或者软件)。该机器学习方法是学习按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该示教位置的校正量的机器学习方法,具备如下步骤:

[0074] 机器人的CPU将示教位置校正量数据S1和电动机干扰值数据S2作为表示进行机器人控制的环境的当前状态的状态变量S而进行观测,其中,示教位置校正量数据S1表示按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该示教位置的校正量,电动机干扰值数据S2表示按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该机器人所具备的各电动机的干扰值;

[0075] 取得表示按照校正后的示教位置进行该机器人的控制时的该机器人所具备的各电动机的干扰值的适当与否判定结果的判定数据D;

[0076] 使用状态变量S和判定数据D,将电动机干扰值数据S2和机器人的示教位置的校正量关联起来进行学习。

[0077] 图6表示第2实施方式的控制装置2。

[0078] 控制装置2具备机器学习装置120和状态数据取得部3,该状态数据取得部3将状态观测部106观测的状态变量S的示教位置校正量数据S1和电动机干扰值数据S2作为状态数据S0而取得。状态数据取得部3取得的状态数据S0可以包括包含基于示教数据的示教位置的示教位置数据S3等。状态数据取得部3可从控制装置2的各部、机床所具备的各种传感器、作业者进行的适当的数据输入等,取得状态数据S0。

[0079] 控制装置2所具有的机器学习装置120除了用于通过机器学习来自学按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该示教位置的校正量的软件(学习算法等)和硬件(处理器101等)外,还包括用于将按照学习到的包含在机器人的示教数据中的示教位置进行该机器人的控制时的该示教位置的校正量作为向控制装置2的指令而输出的软件(运算算法等)和硬件(处理器101等)。控制装置2所包含的机器学习装置120也可以具有一个共同的处理器执行学习算法、运算算法等全部软件的结构。

[0080] 决策部122例如也可以构成为处理器101的一功能,也可以构成为用于使处理器101发挥功能的存储于ROM102的软件。决策部122根据学习部110学习到的结果,生成并输出指令值C,该指令值C包含决定针对按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该机器人所具备的各电动机的干扰值的、机器人的该示教位置的校正值的指令。在决策部122将指令值C输出至控制装置2的情况下,环境状态随之变化。

[0081] 状态观测部106在下一个学习周期观测通过决策部122向环境输出指令值C后变化的状态变量S。学习部110使用该变化后的状态变量S,通过更新例如价值函数Q(即行为价值表),来学习按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该示教位置的校正量。另外,此时状态观测部106不是从状态数据取得部3取得的状态数据S0取得示教位置校正量数据S1,而是如在第1实施方式中说明的那样,从机器学习装置120的RAM103取得示教位置校正量数据S1。

[0082] 决策部122在按照学习到的包含在机器人的示教数据中的示教位置进行该机器人的控制时的该示教位置的校正量下,根据状态变量S将指令值C输出至控制装置2。

[0083] 通过反复该学习周期,机器学习装置120推进按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该示教位置的校正量的学习,使按照自身决定的包含在机器人的示教数据中的示教位置进行该机器人的控制时的该示教位置的校正量的可靠性逐渐提高。

[0084] 具有上述结构的控制装置2所具备的机器学习装置120实现与上述的机器学习装置100同样的效果。尤其,机器学习装置120可根据决策部122的输出使环境状态变化。另一方面,在机器学习装置100中,可向外部装置请求相当于用于将学习部110的学习结果反映于环境的决策部的功能。

[0085] 图7表示具备多个机器人160的一实施方式的系统170。

[0086] 系统170至少具备进行同一作业的多个机器人160、160'和将这些机器人160、160'彼此连接的有线/无线网络172,多个机器人160中的至少一个构成为具备上述的控制装置2的机器人160。此外,系统170也可以包括不具备控制装置2的机器人160'。机器人160、160'具有相同目的作业所必要的机构。

[0087] 关于具有上述结构的系统170,在多个机器人160、160'中具备控制装置的机器人

160可以使用学习部110的学习结果,而不依赖运算、估算地能够自动且准确地求出针对按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的、该机器人所具备的各电动机的干扰值的、机器人的该示教位置的校正量。此外,至少一个机器人160的控制装置2根据针对其他多个机器人160、160'的每一个获得的状态变量S和判定数据D,来学习按照全部机器人160、160'共用的包含在机器人的示教数据中的示教位置进行该机器人的控制时的该示教位置的校正量,并由全部机器人160、160'共享该学习结果。因此,根据系统170,能够以更多样的数据集合(包含状态变量S和判定数据D)为输入,提高按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该示教位置的校正量的学习速度、可靠性。

[0088] 图8表示具备多个机器人160'的其他实施方式的系统170'。

[0089] 系统170'具备机器学习装置120(或100)、具有同一机械结构的多个机器人160'、将这些机器人160'和机器学习装置120(或100)相互连接的有线/无线网络172。

[0090] 关于具有上述结构的系统170',机器学习装置120(或100)根据针对多个机器人160'的每一个获得的状态变量S和判定数据D,来学习按照全部机器人160'共用的包含在机器人的示教数据中的示教位置进行该机器人的控制时的该示教位置的校正量,并使用该学习结果,而不依赖运算、估算地能够自动且准确地求出针对按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的、该机器人所具备的各电动机的干扰值的机器人的该示教位置的校正量。

[0091] 系统170'可以具有机器学习装置120(或100)存在于在网络172中准备的云服务器等中的结构。根据该结构,不取决于多个机器人160'各自所在的位置、时期,而可以在必要时将必要数量的机器人160'连接于机器学习装置120(或100)。

[0092] 从事系统170、170'的作业者在通过机器学习装置120(或100)开始学习后的恰当的时期,能够执行如下的判断:按照机器学习装置120(或100)的包含在机器人的示教数据中的示教位置进行该机器人的控制时的该示教位置的校正量的学习到达程度(即,按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该示教位置的校正量的可靠性)是否达到了要求水平的判断。

[0093] 以上,对本发明的实施方式进行了说明,但本发明并不限于上述的实施方式的例子,通过增加适当的变更能够以各种方式实施。

[0094] 例如,机器学习装置100、120所执行的学习算法、机器学习装置120所执行的运算算法、控制装置1、2所执行的控制算法等并不限于上述的说明,可采用各种算法。

[0095] 此外,在上述的实施方式中将控制装置1(或2)和机器学习装置100(或120)作为具有不同的CPU的装置而进行了说明,但机器学习装置100(或120)也可以通过控制装置1(或2)所具备的CPU11和存储在ROM12中的系统程序来实现。

[0096] 并且,在上述的实施方式中示出了根据机器学习装置120输出的指令值C,控制装置2控制按照包含在机器人的示教数据中的示教位置进行该机器人的控制时的该示教位置的校正量的结构,但也可以将机器学习装置120输出的指令值C显示于示教操作盘所具备的显示装置中,而成为作业者进行示教时的参考。

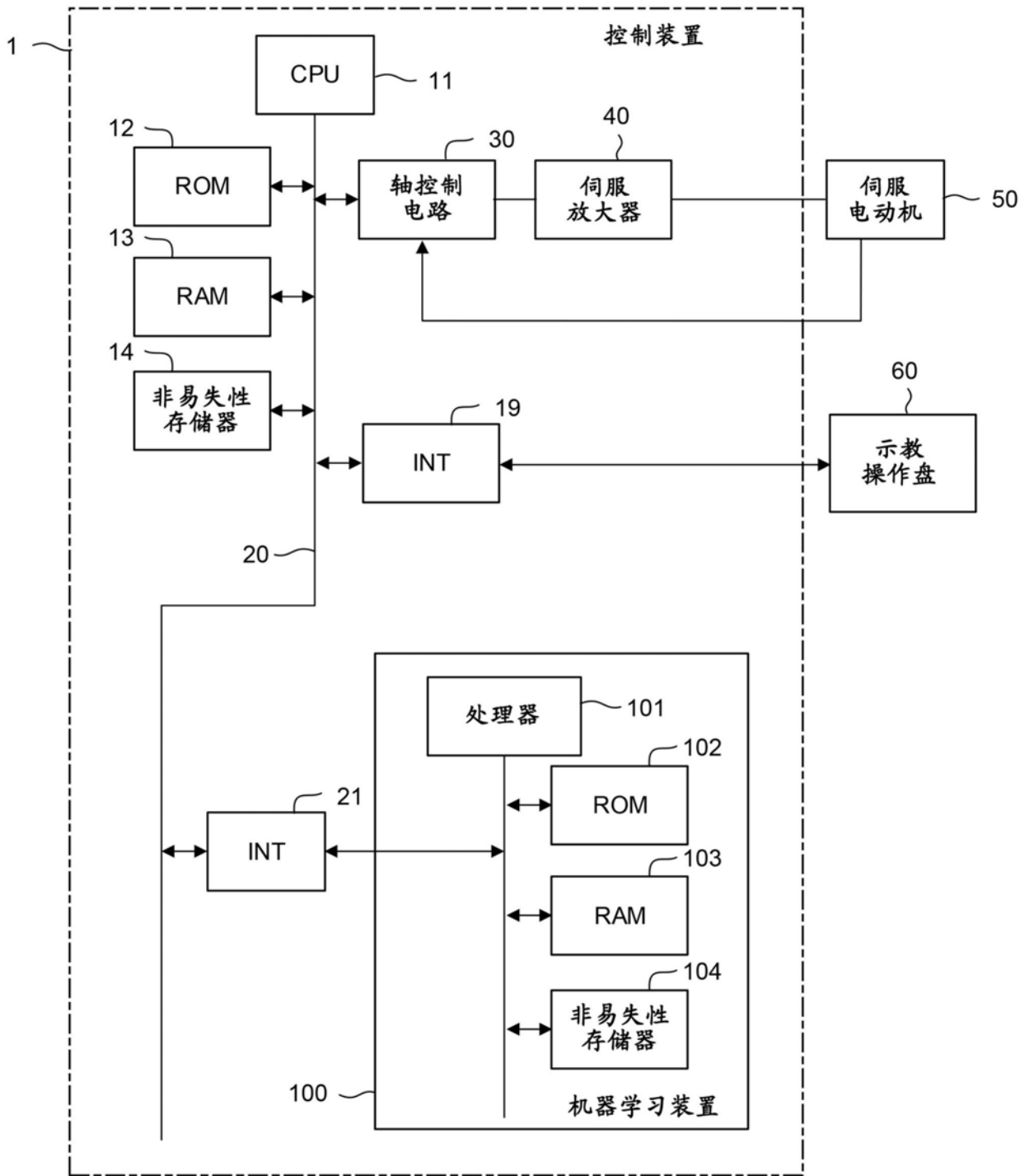


图1

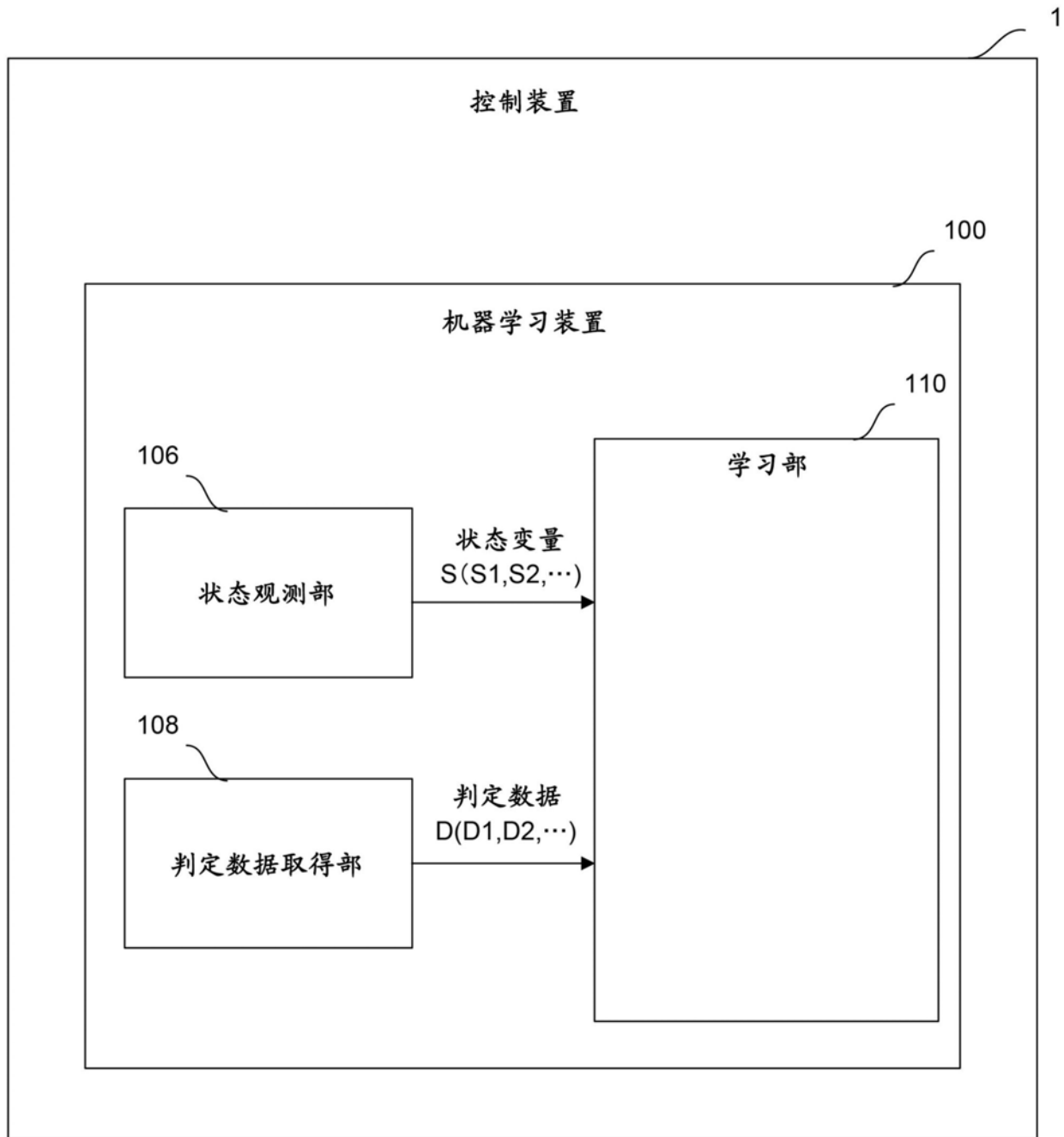


图2



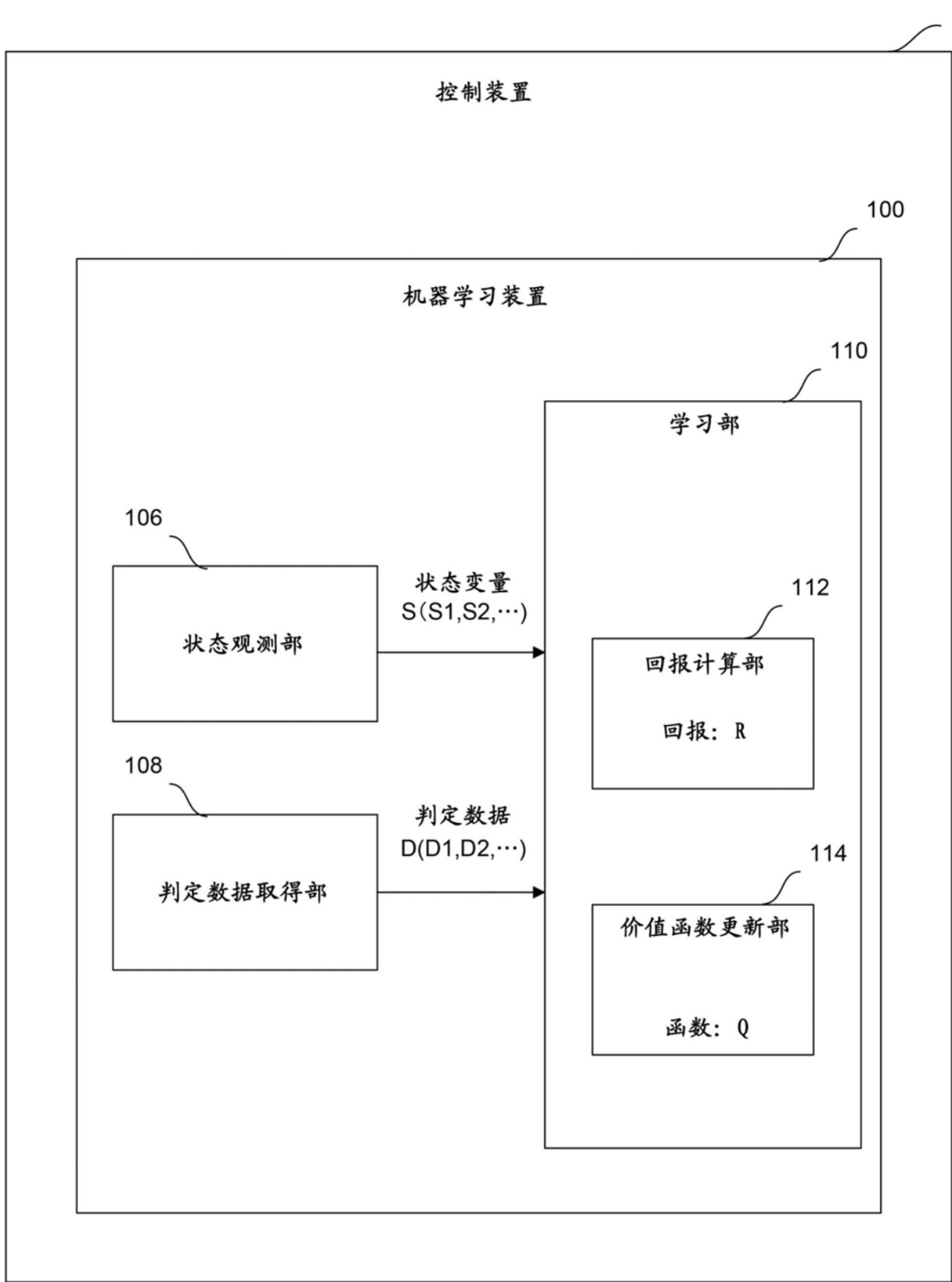


图3

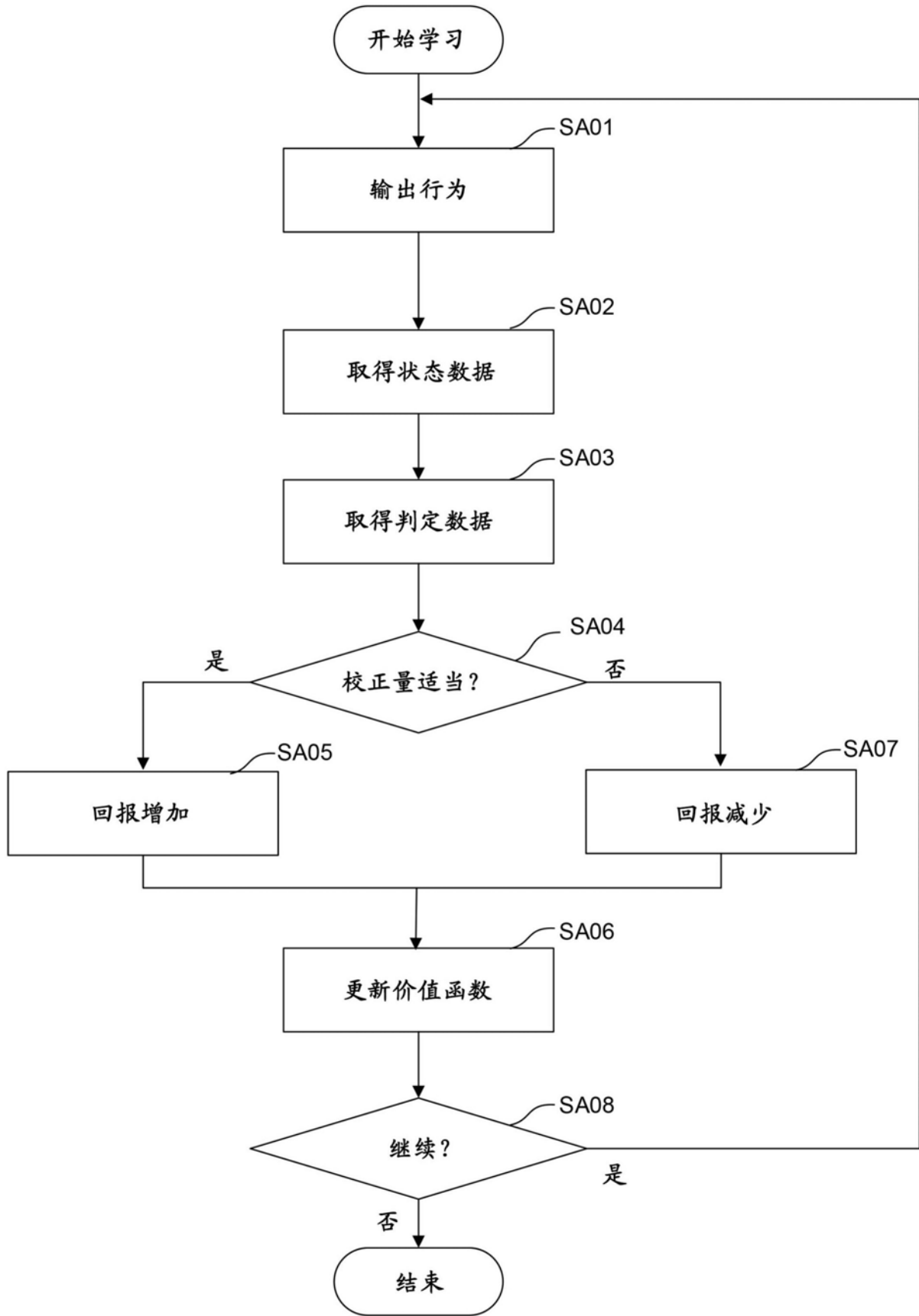


图4

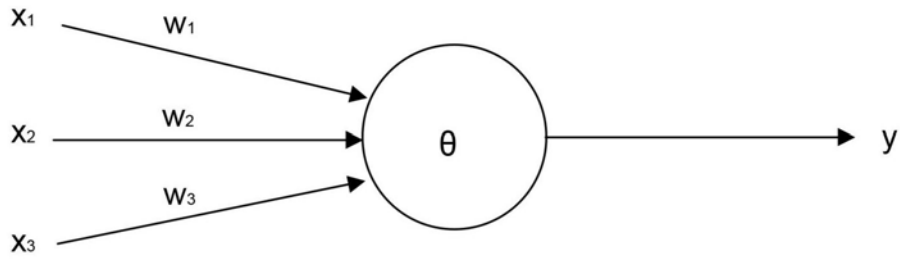


图5A

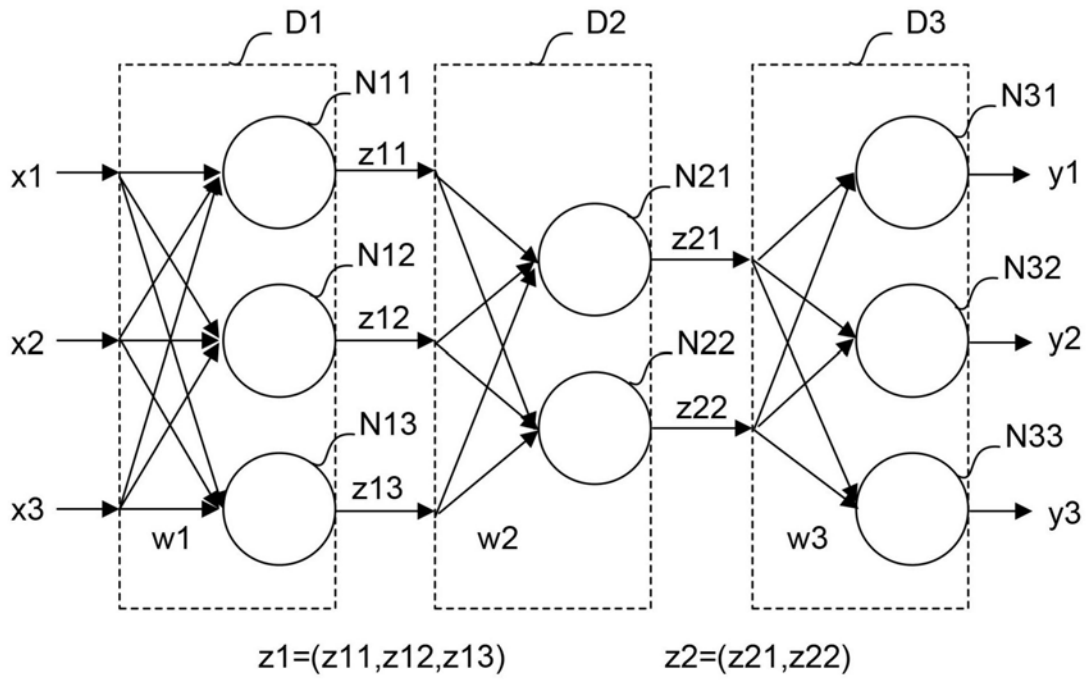


图5B

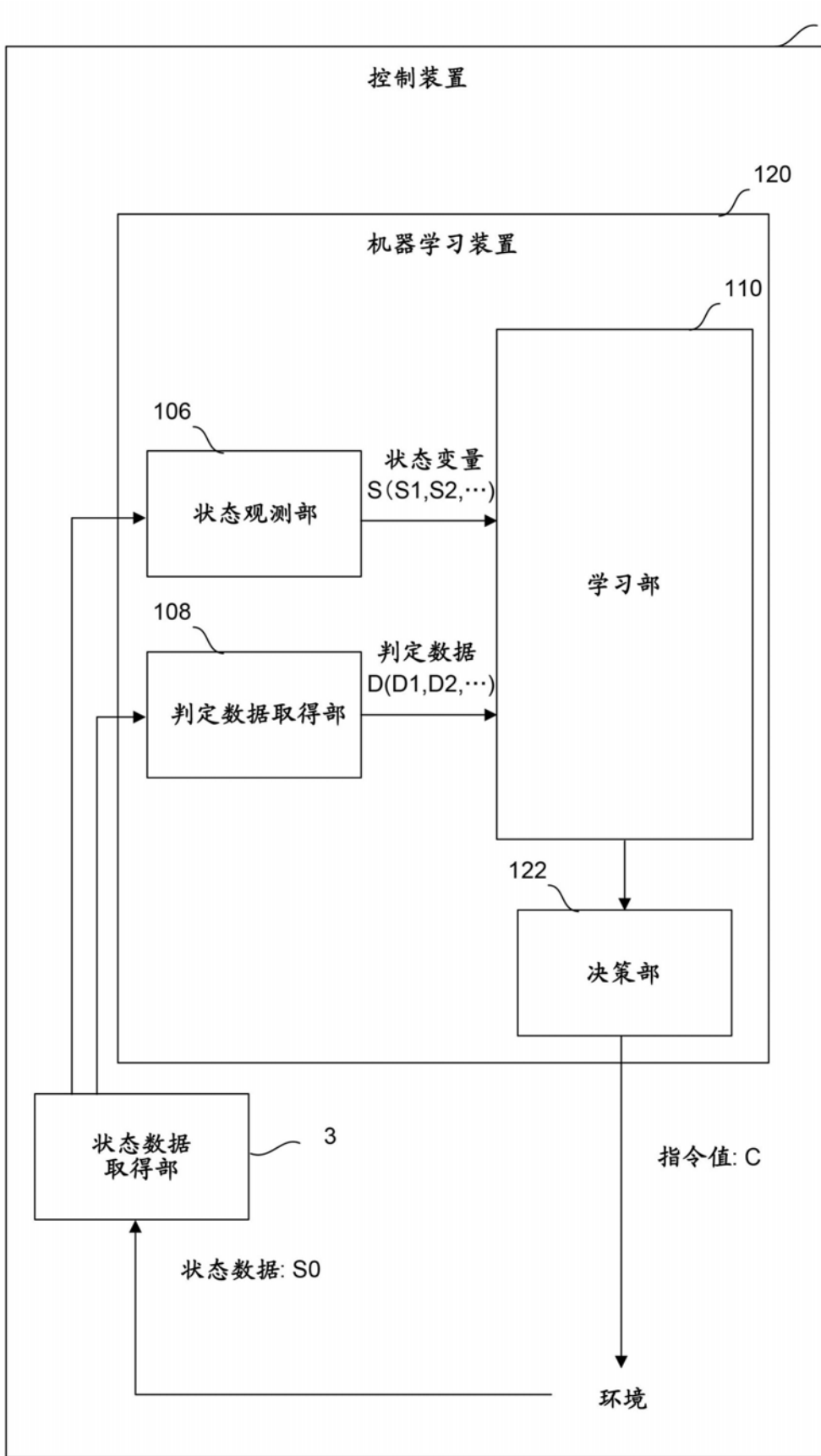


图6

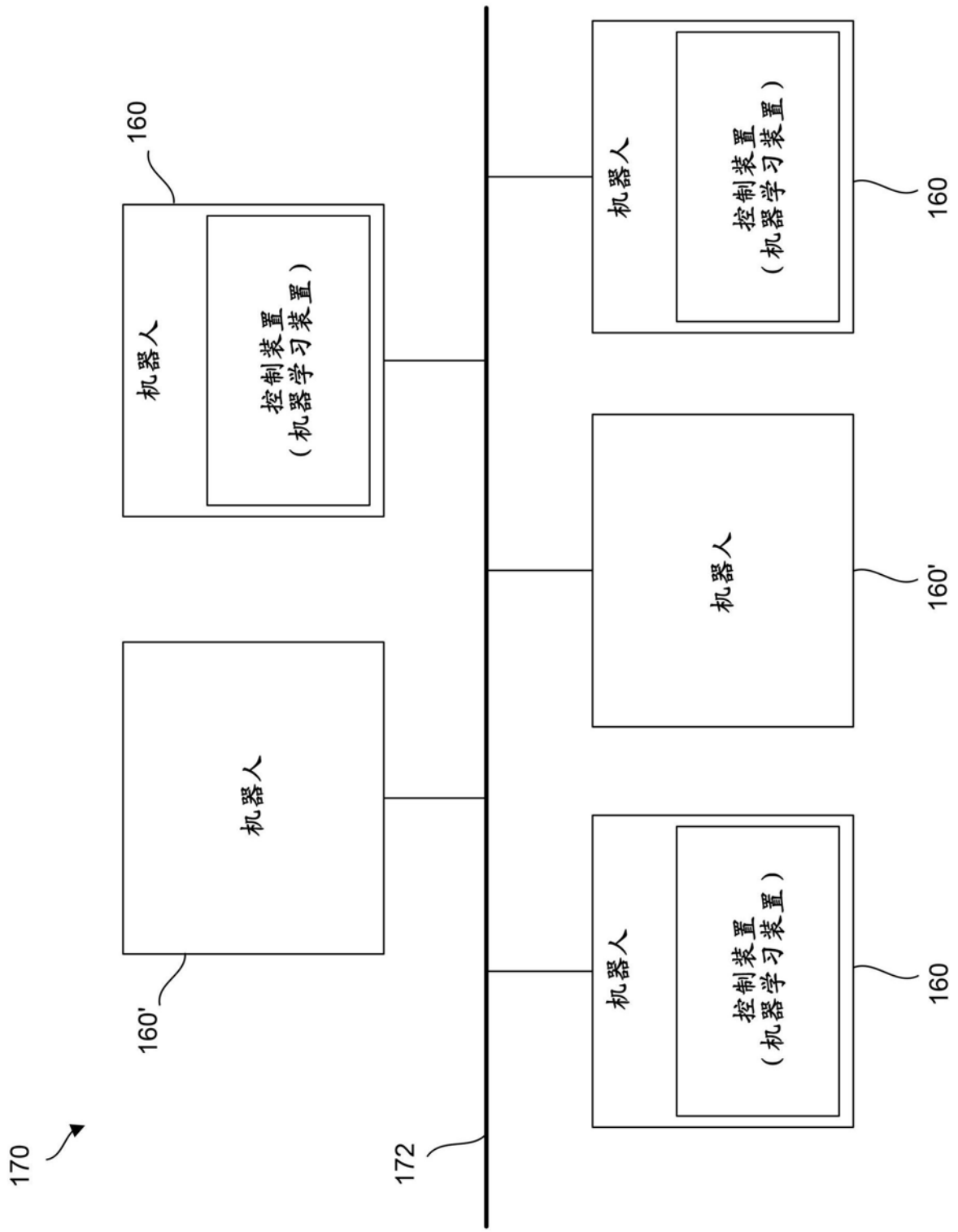


图7

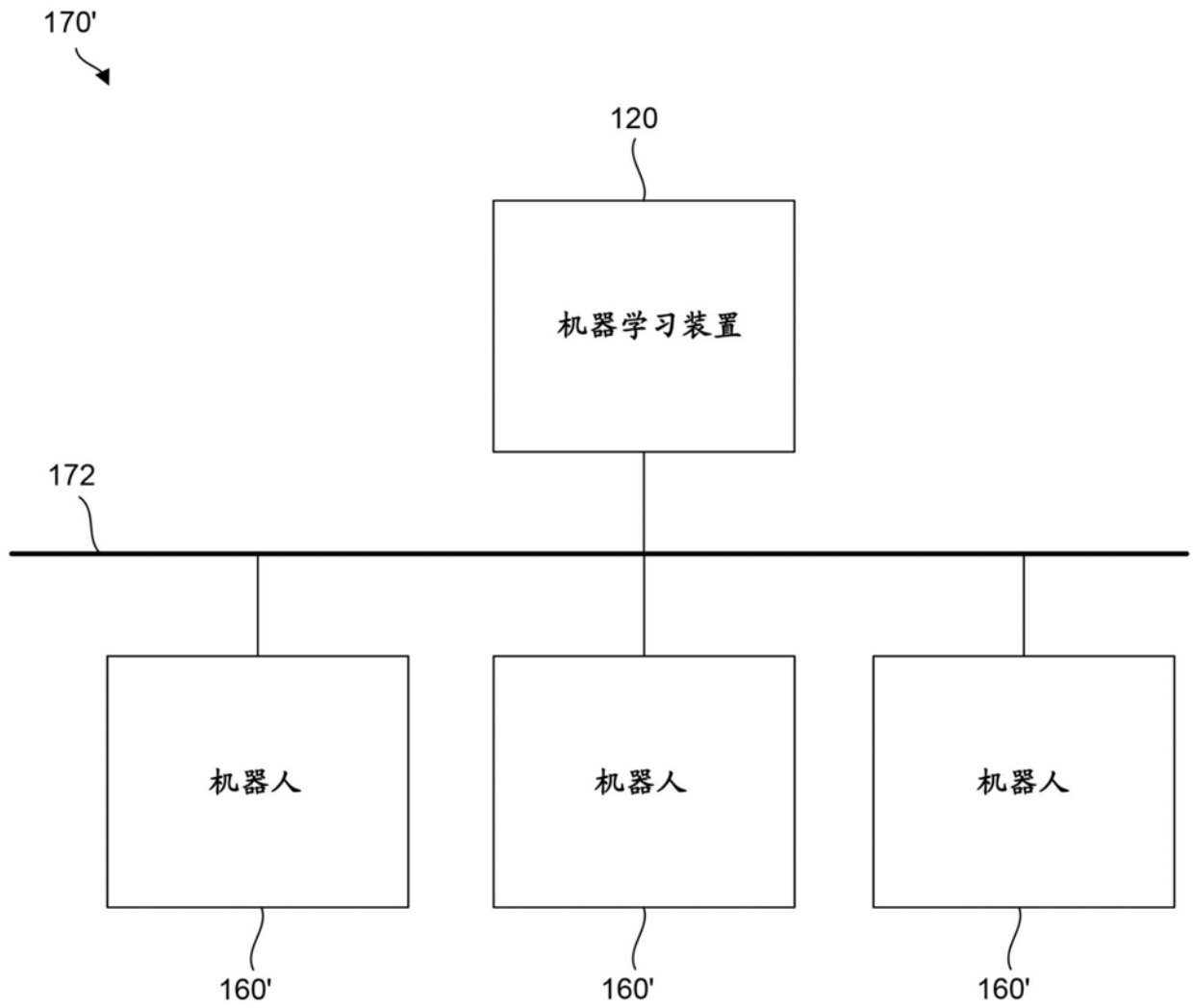


图8

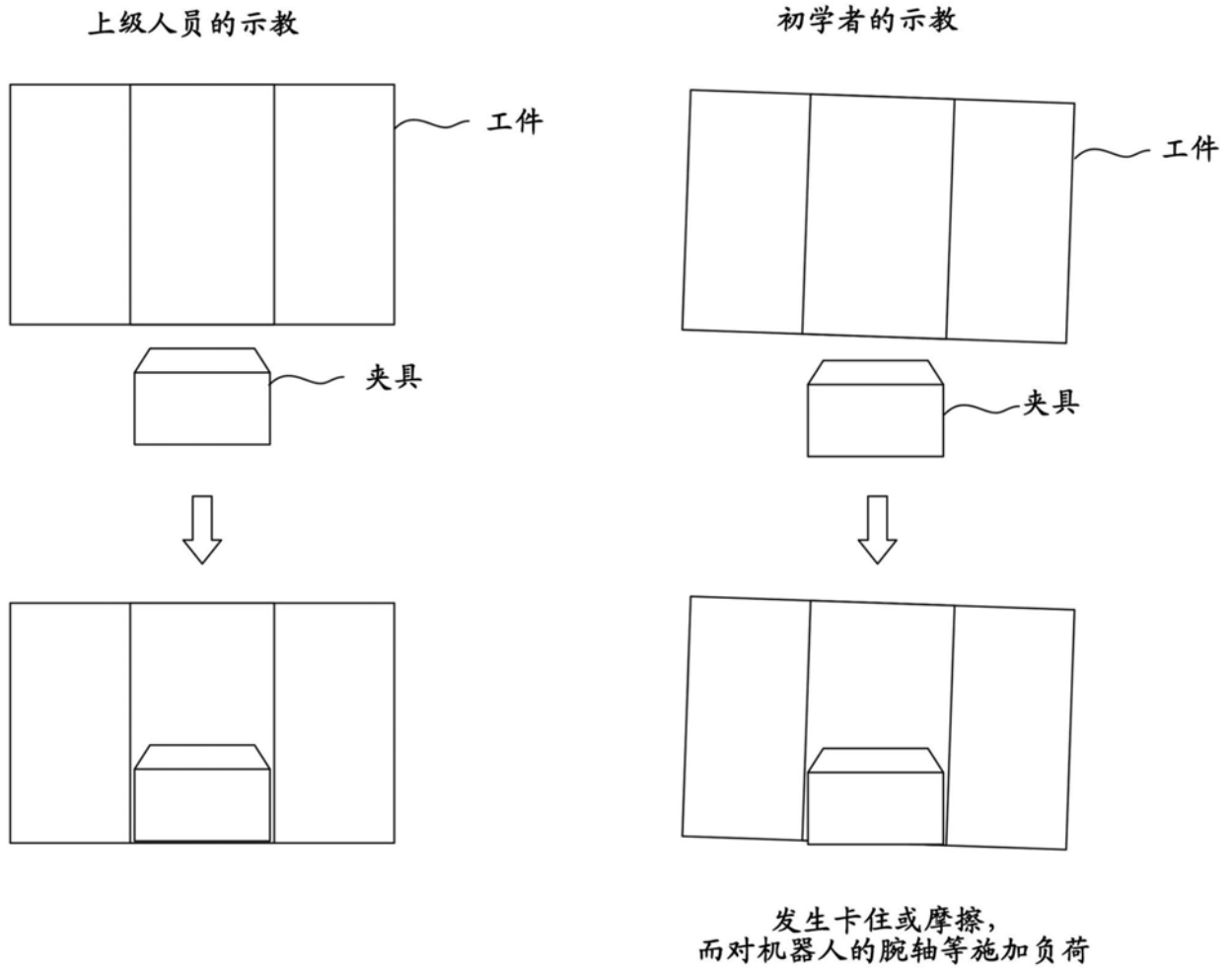


图9