



(12) 发明专利

(10) 授权公告号 CN 101923663 B

(45) 授权公告日 2015. 08. 12

(21) 申请号 201010199038. 8

审查员 张雯

(22) 申请日 2010. 06. 04

(30) 优先权数据

2009-140064 2009. 06. 11 JP

(73) 专利权人 索尼公司

地址 日本东京都

(72) 发明人 野田邦昭 佐部浩太郎

吉池由纪子 河本献太 日台健一

(74) 专利代理机构 北京集佳知识产权代理有限公司 11227

代理人 康建峰 李春晖

(51) Int. Cl.

G06N 7/00(2006. 01)

(56) 对比文件

CN 1716280 A, 2006. 01. 04,

US 2007265811 A1, 2007. 11. 15,

CN 101410855 A, 2009. 04. 15,

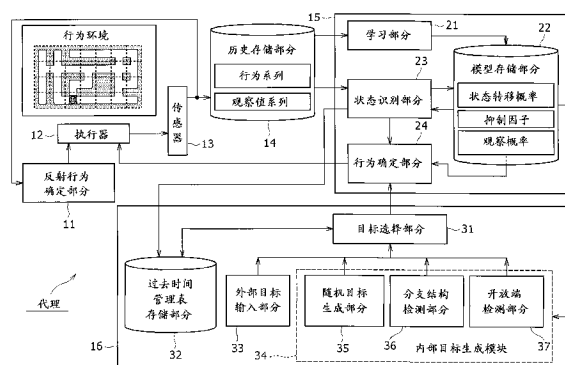
权利要求书5页 说明书70页 附图50页

(54) 发明名称

信息处理设备、信息处理方法和程序

(57) 摘要

本发明公开了一种信息处理设备、信息处理方法和程序。所述信息处理设备包括学习部分，被配置为使用由能够执行行为的代理执行的行为和当代理已经执行了行为时在代理中观察到的观察值来学习由以下定义的状态转移概率模型：状态由于代理执行的行为而进行状态转移的每个行为的状态转移概率；以及从所述状态观察到预定观察值的观察概率。



1. 一种信息处理设备,包括:

学习装置,用于使用由能够执行行为的代理执行的行为和当所述代理已经执行了行为时在所述代理中观察到的观察值学习由以下定义的状态转移概率模型:

状态由于由所述代理执行的行为而进行状态转移的每个所述行为的状态转移概率,其中特定行为的状态转移概率表示在任一状态下由于所述代理执行所述特定行为而转移至另一不同状态的状态转移概率;以及

从所述状态观察到预定观察值的观察概率;

其中所述状态转移概率模型是单个模型,并且

所述观察值在所述行为的环境内的下述观察单元中的至少一个中进行观察:

在上、下和左具有墙壁并且在右具有通道;

在上、左和右具有墙壁并且在下具有通道;

在上和左具有墙壁并且在下和右具有通道;

在上、下和右具有墙壁并且在左具有通道;

在上和下具有墙壁并且在左和右具有通道;

在上和右具有墙壁并且在下和左具有通道;

在上具有墙壁并且在下、左和右具有通道;

在下、左和右具有墙壁并且在上具有通道;

在下和左具有墙壁并且在上和右具有通道;

在左和右具有墙壁并且在上和下具有通道;

在左具有墙壁并且在上、下和右具有通道;

在下和右具有墙壁并且在上和左具有通道;

在下具有墙壁并且在上、左和右具有通道;

在右具有墙壁并且在上、下和左具有通道;以及

在所有的上、下、左和右具有通道。

2. 根据权利要求 1 所述的信息处理设备,

其中,所述学习装置在一个状态一个观察值约束下学习所述状态转移概率模型,在所述一个观察值的一个状态一个观察值约束下,在所述状态转移概率模型的一个状态中观察到一个观察值。

3. 根据权利要求 2 所述的信息处理设备,

其中,所述学习装置通过重复下面所述进行满足所述一个状态一个观察值约束的学习,直到检测不到划分对象状态为止:

在学习后的所述状态转移概率模型中检测其中观察到多个观察值的状态作为要划分的划分对象状态;

将所述划分对象状态划分为多个状态,在所述多个状态的每一个中,观察到所述多个观察值之一;并且

在所述划分对象状态被划分为所述多个状态后,重新学习所述状态转移概率模型。

4. 根据权利要求 3 所述的信息处理设备,

其中,所述学习装置通过下面所述将所述划分对象状态划分为多个划分后状态:

将所述多个观察值之一分配给通过划分所述划分对象状态而获得的划分后状态;

将在所述划分后状态中观察到被分配给所述划分后状态的观察值的观察概率设置为 1, 并且将在所述划分后状态中观察到其他观察值的观察概率设置为 0; 并且

将以所述划分后状态作为转移源的状态转移的状态转移概率设置为以所述划分对象状态作为转移源的状态转移的状态转移概率, 并且将以所述划分后状态作为转移目的地的状态转移的状态转移概率设置为通过以被分配给所述划分后状态的观察值在所述划分对象状态中的观察概率修正以所述划分对象状态为转移目的地的状态转移的状态转移概率所获得的值。

5. 根据权利要求 2 所述的信息处理设备,

其中, 当在执行预定行为时存在多个状态作为状态转移的转移源状态或者转移目的地状态、并且在所述多个状态的每一个中观察到相同的观察值时, 所述学习装置将所述多个状态合并为一个状态。

6. 根据权利要求 5 所述的信息处理设备,

其中, 所述学习装置通过下面所述将作为合并对象状态的多个状态合并为代表性状态:

当在执行预定行为时在状态转移的所述状态转移概率模型中存在多个状态作为转移源状态或者转移目的地状态、并且在多个相应状态中观察到的具有最大所述观察概率的观察值彼此一致时, 检测所述多个状态作为要合并的所述合并对象状态;

将当作为所述合并对象状态的多个状态被合并为所述一个状态时在作为所述一个状态的所述代表性状态中观察到每个观察值的观察概率设置为在作为所述合并对象状态的多个相应状态中观察到每个观察值的观察概率的平均值, 并且将在除了所述代表性状态之外的所述合并对象状态中观察到每个观察值的观察概率设置为 0;

将以所述代表性状态作为转移源的状态转移的状态转移概率设置为以作为所述合并对象状态的多个相应状态作为转移源的状态转移的状态转移概率的平均值, 并且将以所述代表性状态作为转移目的地的状态转移的状态转移概率设置为以作为所述合并对象状态的多个相应状态作为转移目的地的状态转移的状态转移概率的和; 并且

将以除了所述代表性状态之外的所述合并对象状态作为转移源的状态转移的状态转移概率和以除了所述代表性状态之外的所述合并对象状态作为转移目的地的状态转移的状态转移概率设置为 0。

7. 根据权利要求 6 所述的信息处理设备,

其中, 所述学习装置通过重复下面所述执行满足所述一个状态一个观察值约束的学习, 直到检测不到合并对象状态为止:

从学习后的所述状态转移概率模型中检测多个状态作为所述合并对象状态;

将作为所述合并对象状态的多个状态合并为所述代表性状态; 并且

在合并后, 重新学习所述状态转移概率模型。

8. 根据权利要求 2 所述的信息处理设备,

其中, 所述状态转移概率模型是扩展隐马尔可夫模型, 通过将隐马尔可夫模型的状态转移概率扩展到由所述代理执行的每个行为的状态转移概率而获得所述扩展隐马尔可夫模型, 并且

所述学习装置执行所述扩展隐马尔可夫模型的学习, 以根据 Baum-Welch 重评估法估

计关于每个行为的所述状态转移概率和所述观察概率。

9. 一种信息处理设备的信息处理方法,所述信息处理方法包括步骤:

使用由能够执行行为的代理执行的行为和当所述代理已经执行了行为时在所述代理中观察到的观察值来学习由以下定义的状态转移概率模型:

状态由于由所述代理执行的行为而进行状态转移的每个所述行为的状态转移概率,其中特定行为的状态转移概率表示在任一状态下由于所述代理执行所述特定行为而转移至另一不同状态的状态转移概率;以及

从所述状态观察到预定观察值的观察概率;

其中所述状态转移概率模型是单个模型,并且

所述观察值在所述行为的环境内的下述观察单元中的至少一个中进行观察:

在上、下和左具有墙壁并且在右具有通道;

在上、左和右具有墙壁并且在下具有通道;

在上和左具有墙壁并且在下和右具有通道;

在上、下和右具有墙壁并且在左具有通道;

在上和下具有墙壁并且在左和右具有通道;

在上和右具有墙壁并且在下和左具有通道;

在上具有墙壁并且在下、左和右具有通道;

在下、左和右具有墙壁并且在上具有通道;

在下和左具有墙壁并且在上和右具有通道;

在左和右具有墙壁并且在上和下具有通道;

在左具有墙壁并且在上、下和右具有通道;

在下和右具有墙壁并且在上和左具有通道;

在下具有墙壁并且在上、左和右具有通道;

在右具有墙壁并且在上、下和左具有通道;以及

在所有的上、下、左和右具有通道。

10. 一种信息处理设备,包括:

状态识别装置,用于根据状态转移概率模型,使用由能够执行行为的代理执行的行为和当所述代理已经执行了行为时在所述代理中观察到的观察值来识别所述代理的当前状况,并且获得当前状态作为所述状态转移概率模型的状态,所述状态转移概率模型的状态对应于所述当前状况,所述状态转移概率模型是通过使用由所述代理执行的行为和当所述代理已经执行了行为时在所述代理中观察到的观察值来学习由以下定义的状态转移概率模型而获得的:

状态由于由所述代理执行的行为而进行状态转移的每个所述行为的状态转移概率,其中特定行为的状态转移概率表示在任一状态下由于所述代理执行所述特定行为而转移至另一不同状态的状态转移概率;以及

从所述状态观察到预定观察值的观察概率;

目标确定装置,用于将所述状态转移概率模型的状态之一确定为被设置为目标的目标状态;以及

行为确定装置,用于根据所述状态转移概率模型来计算行为计划为最大化从所述当前

状态向所述目标状态的状态转移的可能性的行为系列,并且根据所述行为计划确定接下来要由所述代理执行的行为,

其中所述观察值在所述行为的环境内的下述观察单元中的至少一个中进行观察:

在上、下和左具有墙壁并且在右具有通道;

在上、左和右具有墙壁并且在下具有通道;

在上和左具有墙壁并且在下和右具有通道;

在上、下和右具有墙壁并且在左具有通道;

在上和下具有墙壁并且在左和右具有通道;

在上和右具有墙壁并且在下和左具有通道;

在上具有墙壁并且在下、左和右具有通道;

在下、左和右具有墙壁并且在上具有通道;

在下和左具有墙壁并且在上和右具有通道;

在左和右具有墙壁并且在上和下具有通道;

在左具有墙壁并且在上、下和右具有通道;

在下和右具有墙壁并且在上和左具有通道;

在下具有墙壁并且在上、左和右具有通道;

在右具有墙壁并且在上、下和左具有通道;以及

在所有的上、下、左和右具有通道。

11. 根据权利要求 10 所述的信息处理设备,

其中,所述状态识别装置还更新用于抑制状态转移的抑制因子,以便关于在从紧邻在所述当前状态之前的紧邻在前状态向所述当前状态的状态转移时由所述代理执行的行为,抑制在所述紧邻在前状态和除了所述当前状态之外的状态之间的状态转移,并且

所述行为确定装置使用所述抑制因子来修正所述状态转移概率模型的所述状态转移概率,并且根据修正后的所述状态转移概率来计算所述行为计划。

12. 根据权利要求 11 所述的信息处理设备,

其中,所述状态识别装置还更新所述抑制因子,以便随着时间的经过而放开所述状态转移的抑制。

13. 根据权利要求 10 所述的信息处理设备,还包括:

开放端检测装置,用于检测开放端为在可以与作为转移源的其中观察到预定观察值的状态进行的状态转移中具有还没有执行的状态转移的其他状态,在所述其他状态中观察到与所述预定观察值相同的观察值,

其中,所述目标确定装置将所述开放端确定为所述目标状态。

14. 根据权利要求 13 所述的信息处理设备,

其中,所述开放端检测装置

使用所述状态转移概率和所述观察概率来获得作为当观察到每个观察值时所述代理执行每个行为的概率的行为概率,

通过将所述行为概率乘以所述观察概率,计算基于观察概率的行为概率作为所述代理在其中观察到每个观察值的每个状态中执行每个行为的概率,

通过关于每个状态将以所述状态作为在每个行为中的转移源的状态转移的所述状态

转移概率相加在一起,计算基于状态转移概率的行为概率作为所述代理在每个状态中执行每个行为的概率,并且

将其中基于所述观察概率的行为概率和基于所述状态转移概率的行为概率之间的差等于或者大于预定阈值的状态检测为所述开放端。

15. 根据权利要求 10 所述的信息处理设备,还包括:

分支结构检测装置,用于根据所述状态转移概率检测分支结构的状态为下述状态:当执行一个行为时,从这个状态可以进行向不同状态的状态转移,

其中,所述目标确定装置将所述分支结构的状态确定为所述目标状态。

16. 一种信息处理方法的信息处理方法,所述信息处理方法包括以下步骤:

根据状态转移概率模型,使用由能够执行行为的代理执行的行为和当所述代理已经执行了行为时在所述代理中观察到的观察值来识别所述代理的当前状况,并且获得当前状态作为所述状态转移概率模型的状态,所述状态转移概率模型的状态对应于所述当前状况,所述状态转移概率模型是通过使用由所述代理执行的行为和当所述代理已经执行了行为时在所述代理中观察到的观察值来学习由以下定义的所述状态转移概率模型而获得的:

状态由于由所述代理执行的行为而进行状态转移的每个所述行为的状态转移概率,其中特定行为的状态转移概率表示在任一状态下由于所述代理执行所述特定行为而转移至另一不同状态的状态转移概率;以及

从所述状态观察到预定观察值的观察概率;

将所述状态转移概率模型的状态之一确定为被设置为目标的目标状态;并且

根据所述状态转移概率模型计算行为计划为最大化从所述当前状态向所述目标状态的状态转移的可能性的行为系列,并且根据所述行为计划确定接下来要由所述代理执行的行为,

其中所述观察值在所述行为的环境内的下述观察单元中的至少一个中进行观察:

在上、下和左具有墙壁并且在右具有通道;

在上、左和右具有墙壁并且在下具有通道;

在上和左具有墙壁并且在下和右具有通道;

在上、下和右具有墙壁并且在左具有通道;

在上和下具有墙壁并且在左和右具有通道;

在上和右具有墙壁并且在下和左具有通道;

在上具有墙壁并且在下、左和右具有通道;

在下、左和右具有墙壁并且在上具有通道;

在下和左具有墙壁并且在上和右具有通道;

在左和右具有墙壁并且在上和下具有通道;

在左具有墙壁并且在上、下和右具有通道;

在下和右具有墙壁并且在上和左具有通道;

在下具有墙壁并且在上、左和右具有通道;

在右具有墙壁并且在上、下和左具有通道;以及

在所有的上、下、左和右具有通道。

信息处理设备、信息处理方法和程序

技术领域

[0001] 本发明涉及一种信息处理设备、信息处理方法和程序，具体地涉及一种使得例如可以确定能够自主执行各种行为的代理（自主代理）的适当行为的信息处理设备、信息处理方法和程序。

背景技术

[0002] 作为状态预测和行为确定方法，例如存在一种方法，用于应用部分观察的马尔可夫决策过程并且根据学习数据自动地构造静态的部分观察的马尔可夫决策过程（参见例如日本待审专利公布第 2008-186326 号，以下称为专利文献 1）。

[0003] 另外，作为用于自主移动机器人或者钟摆（pendulum）的操作计划方法，存在一种方法，用于建立在马尔可夫状态模型中离散化的行为计划，并且进一步向控制器输入计划的目标以及得出要提供给控制对象的输出从而执行期望的控制（参见例如日本待审专利公布第 2007-317165 和 2006-268812 号，分别被称为专利文献 2 和 3）。

发明内容

[0004] 虽然已经提出了各种方法作为用于确定能够自主执行各种行为的代理的适当行为的方法，但是存在提出另外的新方法的要求。

[0005] 考虑到这种情况提出了本发明。期望能够确定代理的适当行为，即确定适当行为作为要由代理执行的行为。

[0006] 根据本发明的第一实施例的信息处理设备和程序是一种信息处理设备和用于使计算机用作信息处理设备的程序，所述信息处理设备包括学习部分，被配置为使用由代理执行的行为和当代理已经执行了该行为时代理中观察到的观察值来学习由每个行为的状态转移概率和观察概率定义的状态转移概率模型，所述每个行为是状态由于能够执行行为的代理执行的行为而进行状态转移的行为，所述观察概率是从所述状态观察到预定观察值的观察概率。

[0007] 根据本发明的第一实施例的信息处理方法是一种信息处理设备的信息处理方法，所述方法包括步骤：使用由代理执行的行为和当代理已经执行了行为时代理中观察到的观察值来学习由每个行为的状态转移概率和观察概率定义的状态转移概率模型，所述每个行为是状态由于能够执行行为的代理执行的行为而进行状态转移的行为，所述观察概率是从所述状态观察到预定观察值的观察概率。

[0008] 在如上所述的第一实施例中，使用由代理执行的行为和当代理已经执行了行为时代理中观察到的观察值来学习由每个行为的状态转移概率和观察概率定义的状态转移概率模型，所述每个行为是状态由于能够执行行为的代理执行的行为而进行状态转移的行为，所述观察概率是从所述状态观察到预定观察值的观察概率。

[0009] 根据本发明的第二实施例的信息处理设备或程序是一种信息处理设备或用于使计算机用作信息处理设备的程序。所述信息处理设备包括：状态识别部分，被配置为根据状

态转移概率模型,使用由能够执行行为的代理执行的行为和当代理已经执行了行为时代理中观察到的观察值来识别代理的当前状况,并且获得作为状态转移概率模型的状态的当前状态,所述状态转移概率模型的状态对应于当前状况,所述状态转移概率模型是通过使用由代理执行的行为和当代理已经执行了行为时代理中观察到的观察值来学习由每个行为的状态转移概率和观察概率定义的状态转移概率模型而获得的,所述每个行为是状态由于能够执行行为的代理执行的行为而进行状态转移的行为,所述观察概率是从所述状态观察到预定观察值的观察概率;目标确定部分,被配置为将所述状态转移概率模型的状态之一确定为被设置为目标的目标状态;以及行为确定部分,被配置为根据所述状态转移概率模型计算行为计划为最大化从所述当前状态向所述目标状态的状态转移的可能性的行为系列,并且所述行为确定部分被配置为根据所述行为计划确定接下来要由代理执行的行为。

[0010] 根据本发明的第二实施例的信息处理方法是一种信息处理设备的信息处理方法。所述信息处理方法包括步骤:根据状态转移概率模型,使用由能够执行行为的代理执行的行为和当代理已经执行了行为时代理中观察到的观察值来识别代理的当前状况,并且获得作为状态转移概率模型的状态的当前状态,所述状态转移概率模型的状态对应于当前状况,所述状态转移概率模型是通过使用由代理执行的行为和当代理已经执行了行为时代理中观察到的观察值来学习由每个行为的状态转移概率和观察概率定义的状态转移概率模型而获得的,所述每个行为是状态由于能够执行行为的代理执行的行为而进行状态转移的行为,所述观察概率是从所述状态观察预定观察值的观察概率;将状态转移概率模型的状态之一确定为被设置为目标的目标状态;并且根据状态转移概率模型计算行为计划为最大化从所述当前状态向所述目标状态的状态转移的可能性的行为系列,并且根据所述行为计划确定接下来要由代理执行的行为。

[0011] 在如上所述的第二实施例中,根据状态转移概率模型,使用由能够执行行为的代理执行的行为和当代理已经执行了行为时代理中观察到的观察值识别了所述代理的当前状况,并且获得作为状态转移概率模型的状态的当前状态,所述状态转移概率模型的状态对应于当前状况,所述状态转移概率模型是通过使用由代理执行的行为和当所述代理已经执行了行为时代理中观察到的观察值来学习由每个行为的状态转移概率和观察概率定义的状态转移概率模型而获得的,所述每个行为是状态由于能够执行行为的代理执行的行为而进行状态转移的行为,所述观察概率是从所述状态观察预定观察值的观察概率。此外,状态转移概率模型的状态之一被确定为被设置为目标的目标状态。然后,基于状态转移概率模型计算行为计划为最大化从所述当前状态向所述目标状态的状态转移的可能性的行为系列,并且接下来要由代理执行的行为根据行为计划确定。

[0012] 顺便提及,所述信息处理设备可以是独立的设备,或者可以是构成一个设备的内部模块。

[0013] 程序可以通过经由传输介质发送或者处于被记录在记录介质中的状态被提供。

[0014] 根据本发明的第一和第二实施例,可以将适当的行为确定为要由代理执行的行为。

附图说明

[0015] 图 1 是示出了行为环境的图;

- [0016] 图 2 是示出了其中行为环境的结构改变的状态的图；
- [0017] 图 3A 和 3B 是示出了代理执行的行为和代理观察到的观察值的图；
- [0018] 图 4 是示出了应用根据本发明的信息处理设备的代理的一个实施例的配置的示例的框图；
- [0019] 图 5 是帮助说明反射行为模式中的处理的流程图；
- [0020] 图 6A 和 6B 是帮助说明扩展 HMM 的状态转移概率的图；
- [0021] 图 7 是帮助说明学习扩展 HMM 的处理的流程图；
- [0022] 图 8 是帮助说明识别行为模式中的处理的流程图；
- [0023] 图 9 是帮助说明确定目标状态的处理的流程图，该处理由目标确定部分执行；
- [0024] 图 10A、10B 和 10C 是帮助说明行为确定部分 24 计算行为计划的图；
- [0025] 图 11 是帮助说明使用抑制因子来修正扩展 HMM 的状态转移概率的图，该修正由行为确定部分进行；
- [0026] 图 12 是帮助说明更新抑制因子的处理的流程图，该处理由状态识别部分执行；
- [0027] 图 13 是帮助说明由开放端检测部分检测为开放端的扩展 HMM 的状态的图；
- [0028] 图 14A 和 14B 是帮助说明开放端检测部分列出状态 S_i 的处理的图，其中以等于或者高于阈值的概率观察到观察值 O_k ；
- [0029] 图 15 是帮助说明使用针对观察值 O_k 列出的状态 S_i 来生成行为模板的方法的图；
- [0030] 图 16 是帮助说明计算基于观察概率的行为概率的方法的图；
- [0031] 图 17 是帮助说明计算基于状态转移概率的行为概率的方法的图；
- [0032] 图 18 是示意性地示出了差分行为概率的图；
- [0033] 图 19 是帮助说明检测开放端的处理的流程图；
- [0034] 图 20 是帮助说明一种用于由分支结构检测部分检测分支结构的状态的方法的图；
- [0035] 图 21A 和 21B 是示出了模拟中使用的行为环境的图；
- [0036] 图 22 是示意性地示出了模拟中学习后的扩展 HMM 的图；
- [0037] 图 23 是示出了模拟的结果的图；
- [0038] 图 24 是示出了模拟的结果的图；
- [0039] 图 25 是示出了模拟的结果的图；
- [0040] 图 26 是示出了模拟的结果的图；
- [0041] 图 27 是示出了模拟的结果的图；
- [0042] 图 28 是示出了模拟的结果的图；
- [0043] 图 29 是示出了模拟的结果的图；
- [0044] 图 30 是示出了应用代理的清洁机器人的一般外观的图；
- [0045] 图 31A 和 31B 是帮助说明用于实现一个状态一个观察值约束的状态划分的概要的图；
- [0046] 图 32 是帮助说明检测划分对象状态的方法的图；
- [0047] 图 33A 和 33B 是帮助说明一种将划分对象状态划分为划分后状态的方法的图；
- [0048] 图 34A 和 34B 是帮助说明用于实现一个状态一个观察值约束的状态合并的概要的图；

- [0049] 图 35A 和 35B 是帮助说明检测合并对象状态的方法的图；
- [0050] 图 36A 和 36B 是帮助说明一种用于将多个分支目的地状态合并为一个代表性状态的方法的图；
- [0051] 图 37 是帮助说明一种用于学习扩展 HMM 的处理的流程图，该处理在所述一个状态一个观察值约束下执行；
- [0052] 图 38 是帮助说明检测划分对象状态的处理的流程图；
- [0053] 图 39 是帮助说明划分状态的处理的流程图；
- [0054] 图 40 是帮助说明检测合并对象状态的处理的流程图；
- [0055] 图 41 是帮助说明检测合并对象状态的处理的流程图；
- [0056] 图 42 是帮助说明合并状态的处理的流程图；
- [0057] 图 43A、43B 和 43C 是帮助说明模拟在一个状态一个观察值约束下扩展 HMM 的学习的图；
- [0058] 图 44 是帮助说明识别行为模式中的处理的流程图；
- [0059] 图 45 是帮助说明计算当前状况状态系列的候选的处理的流程图；
- [0060] 图 46 是帮助说明计算当前状况状态系列的候选的处理的流程图；
- [0061] 图 47 是帮助说明根据第一策略确定行为的处理的流程图；
- [0062] 图 48 是帮助说明根据第二策略确定行为的概要的图；
- [0063] 图 49 是帮助说明根据第二策略确定行为的处理的流程图；
- [0064] 图 50 是帮助说明根据第三策略确定行为的概要的图；
- [0065] 图 51 是帮助说明根据第三策略确定行为的处理的流程图；
- [0066] 图 52 是帮助说明当根据多个策略确定行为时选择要遵循的策略的处理的流程图；
- [0067] 图 53 是帮助说明当根据多个策略确定行为时选择要遵循的策略的处理的流程图；以及
- [0068] 图 54 是示出了应用本发明的计算机的一个实施例的配置的示例的框图。

具体实施方式

[0069] [代理执行行为的环境]

[0070] 图 1 是示出了其中应用根据本发明的信息处理设备的代理执行行为的行为环境的示例的图。

[0071] 代理是诸如机器人（其可以是在真实世界中行动的机器人或者可以是在虚拟世界中行动的虚拟机器人）的设备，能够自主地执行诸如移动等的行为（能够执行行为）。

[0072] 代理能够通过执行行为、观察外部可观察的信息并且使用作为观察结果的观察值识别状况来改变代理本身的状况。

[0073] 另外，代理构造其中代理执行行为的行为环境的模型（环境模型），以便识别状况并且确定（选择）在每种状况下要执行的行为。

[0074] 代理不仅执行具有固定结构的行为环境的、而且执行具有不固定的而是概率性地改变的结构的行为环境的有效建模（构造环境模型）。

[0075] 图 1 中的行为环境是二维平面内的迷宫，并且行为环境的结构概率性地改变。在

图 1 的行为环境中,代理可以在图中作为通道的空白部分中移动。

[0076] 图 2 是示出了其中行为环境的结构改变的状态的图。

[0077] 在图 2 的行为环境中,在时刻 $t = t_1$ 处,位置 p_1 是墙壁,位置 p_2 是通道。因此,在时刻 $t = t_1$ 处,行为环境具有其中代理不能通过位置 p_1 但是可以通过位置 p_2 的结构。

[0078] 其后,在时刻 $t = t_2$ 处 ($> t_1$),位置 p_1 从墙壁改变为通道,从而行为环境具有其中代理可以通过位置 p_1 和 p_2 两者的结构。

[0079] 另外,在随后的时刻 $t = t_3$ 处,位置 p_2 从通道改变为墙壁,从而行为环境具有其中代理可以通过位置 p_1 但是不能通过位置 p_2 的结构。

[0080] [代理执行的行为和代理观察到的观察值]

[0081] 图 3A 和 3B 是示出了在行为环境中代理执行的行为和代理观察到的观察值的示例的图。

[0082] 代理将图 1 中所示的行为环境中由图中的虚线以正方形的形式划分的区域设置为在其中观察观察值的单元(观察单元)。代理在观察单元中执行移动的行为。

[0083] 图 3A 示出了由代理执行的行为的种类。

[0084] 在图 3A 中,代理总共可以执行 5 种行为 U_1-U_5 ,行为 U_1 是沿观察单元在图中的向上方向上移动,行为 U_2 是沿观察单元在图中的向右方向上移动,行为 U_3 是沿观察单元在图中的向下方向上移动,行为 U_4 是沿观察单元在图中的向左方向上移动,行为 U_5 是不移动(不做任何事情)。

[0085] 图 3B 示意性地示出了代理在观察单元中观察到的观察值的种类。

[0086] 本实施例中的代理在观察单元中观察到 15 种观察值(符号) 0_1-0_{15} 之一。

[0087] 在上、下和左具有墙壁并且在右具有通道的观察单元中观察到观察值 0_1 。在上、左和右具有墙壁并且在下具有通道的观察单元中观察到观察值 0_2 。

[0088] 在上和左具有墙壁并且在下和右具有通道的观察单元中观察到观察值 0_3 。在上、下和右具有墙壁并且在左具有通道的观察单元中观察到观察值 0_4 。

[0089] 在上和下具有墙壁并且在左和右具有通道的观察单元中观察到观察值 0_5 。在上和右具有墙壁并且在下和左具有通道的观察单元中观察到观察值 0_6 。

[0090] 在上具有墙壁并且在下、左和右具有通道的观察单元中观察到观察值 0_7 。在下、左和右具有墙壁并且在上具有通道的观察单元中观察到观察值 0_8 。

[0091] 在下和左具有墙壁并且在上和右具有通道的观察单元中观察到观察值 0_9 。在左和右具有墙壁并且在上和下具有通道的观察单元中观察到观察值 0_{10} 。

[0092] 在左具有墙壁并且在上、下和右具有通道的观察单元中观察到观察值 0_{11} 。在下和右具有墙壁并且在上和左具有通道的观察单元中观察到观察值 0_{12} 。

[0093] 在下具有墙壁并且在上、左和右具有通道的观察单元中观察到观察值 0_{13} 。在右具有墙壁并且在上、下和左具有通道的观察单元中观察到观察值 0_{14} 。

[0094] 在所有的上、下、左和右具有通道的观察单元中观察到观察值 0_{15} 。

[0095] 顺便提及,行为 U_m ($m = 1, 2, \dots, M$ (M 是行为(行为种类)的总数)) 和观察值 0_k ($k = 1, 2, \dots, K$ (K 是观察值的总数)) 每一个均是离散值。

[0096] [代理的配置的示例]

[0097] 图 4 是示出了应用根据本发明的信息处理设备的代理的一个实施例的配置的示

例的框图。

[0098] 代理获得通过学习对行为环境建模得到的环境模型。

[0099] 另外,代理使用一系列观察值(观察值系列)识别代理自身的当前状况。

[0100] 此外,代理制订从当前状况移动到某一目标要执行的行为的计划(行为计划),并且根据该行为计划确定接下来要执行的行为。

[0101] 顺便提及,代理所执行的学习、状况的识别和行为计划的制订(行为的确定)不仅可以应用于代理在观察单元中向上、下、左或右移动的问题(任务),而且可以应用于一般被作为强化学习的问题并且能够在马尔可夫决策过程(MDP)的框架中公式化的问题。

[0102] 在图 4 中,代理通过在行为环境中执行图 3A 所示的行为 U_m 而在观察单元中移动,并且获得在移动后的观察单元中观察到的观察值 O_k 。

[0103] 然后,代理学习行为环境(行为环境的结构(通过对行为环境建模得到的环境模型)),并且使用行为系列和观察值系列确定接下来要执行的行为,所述行为系列是目前为止执行的一系列行为 U_m (用于代表行为 U_m 的符号),并且观察值系列是目前为止观察到的一系列观察值 O_k (用于代表观察值 O_k 的符号)。

[0104] 存在其中代理执行行为的两种模式,即反射行为模式(反射行动模式)和识别行为模式(识别行动模式)。

[0105] 在反射行为模式中,将用于根据过去获得的观察值系列和行为系列确定接下来要执行的行为的规则设计为固有规则。

[0106] 在这种情况下,作为固有规则,可以例如使用用于确定行为以便不撞墙(允许在通道中前后移动)的规则或者用于确定行为以便不撞墙和以便不折回采取过的路径直到来到死胡同的规则。

[0107] 根据该固有规则,代理重复针对代理中观察到的观察值确定接下来要执行的行为和在执行该行为后在观察单元中观察观察值。

[0108] 由此当代理已经在行为环境中移动时代理获得行为系列和观察值系列。在反射行为模式中这样获得的行为系列和观察值系列用于学习行为环境。即,反射行为模式主要用于获取用作学习行为环境中所用的学习数据的行为系列和观察值系列。

[0109] 在识别行为模式中,代理确定目标、识别当前状况并且确定用于从当前状况达到目标的行为计划。然后代理根据该行为计划确定接下来要执行的行为。

[0110] 顺便提及,可以例如根据用户的操作来执行反射行为模式和识别行为模式之间的切换。

[0111] 在图 4 中,代理包括反射行为确定部分 11、执行器 12、传感器 13、历史存储部分 14、行为控制部分 15 和目标确定部分 16。

[0112] 反射行为确定部分 11 被提供有在行为环境中观察到并且由传感器 13 输出的观察值。

[0113] 在反射行为模式中反射行为确定部分 11 根据固有规则确定针对从传感器 13 提供的观察值接下来要执行的行为,并且控制执行器 12。

[0114] 执行器 12 例如当代理是在真实世界中行走的机器人时是用于使代理行走的电动机。执行器 12 在后面要描述的反射行为确定部分 11 或者行为确定部分 24 的控制下执行驱动。在执行器执行驱动的情况下,代理在行为环境中执行由反射行为确定部分 11 或者行

为确定部分 24 确定的行为。

[0115] 传感器 13 感测外部可观察的信息,并且输出观察值作为感测结果。

[0116] 即,传感器 13 观察在行为环境中代理所位于的观察单元,并且输出代表观察单元的符号作为观察值。

[0117] 顺便提及,图 4 中的传感器 13 还观察执行器 12,由此还输出由代理执行的行为(代表行为的符号)。

[0118] 由传感器 13 输出的观察值被提供给反射行为确定部分 11 和历史存储部分 14。另外,由传感器 13 输出的行为被提供给历史存储部分 14。

[0119] 历史存储部分 14 顺序存储由传感器 13 输出的观察值和行为。由此历史存储部分 14 存储一系列观察值(观察值系列)和一系列行为(行为系列)。

[0120] 顺便提及,虽然在这种情况下代表代理所位于的观察单元的符号被用作外部可观察的观察值,但是代表代理所位于的观察单元的符号和代表由代理执行的行为的符号的符号组可以被用作观察值。

[0121] 行为控制部分 15 使用历史存储部分 14 中存储的观察值系列和行为系列来学习状态转移概率模型作为环境模型以存储(获取)行为环境的结构。

[0122] 另外,行为控制部分 15 根据学习后的状态转移概率模型来计算行为计划。此外,行为控制部分 15 根据该行为计划确定代理接下来要执行的行为,并且根据该行为来控制执行器 12,由此使得代理执行该行为。

[0123] 具体地,行为控制部分 15 包括学习部分 21、模型存储部分 22、状态识别部分 23 和行为确定部分 24。

[0124] 学习部分 21 使用历史存储部分 14 中存储的行为系列和观察值系列来学习模型存储部分 22 中存储的状态转移概率模型。

[0125] 在这种情况下,由学习部分 21 学习的状态转移概率模型是由每个行为的状态转移概率和观察概率定义的状态转移概率模型,所述每个行为是状态由于代理执行的行为而进行状态转移的行为,所述观察概率是从状态观察到预定观察值的观察概率。

[0126] 例如有 HMM(隐马尔可夫模型)作为状态转移概率模型。但是,不是对于每个行为都存在普通 HMM 的状态转移概率。因此,在本实施例中,HMM(隐马尔可夫模型)的状态转移概率被扩展为由代理执行的每个行为的状态转移概率。具有如此扩展的状态转移概率的 HMM(所述 HMM 也被称为扩展 HMM)被用作学习部分 21 学习的对象。

[0127] 模型存储部分 22 存储扩展 HMM(状态转移概率和观察概率等作为用于定义扩展 HMM 的模型参数)。另外,模型存储部分 22 存储后面将要描述的抑制因子。

[0128] 在识别行为模式中状态识别部分 23 使用历史存储部分 14 中存储的行为系列和观察值系列、根据模型存储部分 22 中存储的扩展 HMM 来识别代理的当前状况,并且将当前状态获得(识别)作为扩展 HMM 的状态,所述状态对应于当前状况。

[0129] 然后状态识别部分 23 将当前状态提供给行为确定部分 24。

[0130] 另外,状态识别部分 23 更新模型存储部分 22 中存储的抑制因子,并且根据当前状态等更新在后面将要描述的时间管理表存储部分 32 中存储的时间管理表。

[0131] 行为确定部分 24 用作计划者,用于计划在识别行为模式中要由代理执行的行为。

[0132] 具体地,行为确定部分 24 被提供有来自状态识别部分 23 的当前状态,并且还提

供有作为来自目标确定部分 16 的作为目标的目标状态的、存储在模型存储部分 22 中的扩展 HMM 的状态之一。

[0133] 行为确定部分 24 根据模型存储部分 22 中存储的扩展 HMM 来计算（确定）为一系列行为的行为计划，所述一系列行为最大化从来自状态识别部分 23 的当前状态到来自目标确定部分 16 的目标状态的状态转移的可能性。

[0134] 此外，行为确定部分 24 根据行为计划确定接下来要由代理执行的行为，并且根据所确定的行为来控制执行器 12。

[0135] 在识别行为模式中目标确定部分 16 确定目标状态，然后将目标状态提供给行为确定部分 24。

[0136] 具体地，目标确定部分 16 包括目标选择部分 31、过去时间管理表存储部分 32、外部目标输入部分 33 和内部目标生成模块 34。

[0137] 目标选择部分 31 被提供有来自外部目标输入部分 33 的作为目标状态的外部目标和来自内部目标生成模块 34 的作为目标状态的内部目标。

[0138] 目标选择部分 31 选择来自外部目标输入部分 33 的作为外部目标的状态或者来自内部目标生成模块 34 的作为内部目标的状态，将所选择的状态确定为目标状态，然后将该目标状态提供给行为确定部分 24。

[0139] 过去时间管理表存储部分 32 存储过去时间管理表。对于存储在模型存储部分 22 中的扩展 HMM 的每个状态，在过去时间管理表中登记自从状态变为当前状态起过去的过去时间等。

[0140] 外部目标输入部分 33 将从（代理的）外部提供的状态设置为作为目标状态的外部目标，然后将该外部目标提供给目标选择部分 31。

[0141] 即，外部目标输入部分 33 例如当用户在外部分指定状态作为目标状态时由用户操作。外部目标输入部分 33 将由用户的操作指定的状态设置为作为目标状态的外部目标，然后将该外部目标提供到目标选择部分 31。

[0142] 内部目标生成模块 34 在（代理的）内部生成内部目标作为目标状态，然后将该内部目标提供给目标选择部分 31。

[0143] 具体地，内部目标生成模块 34 包括随机目标生成部分 35、分支结构检测部分 36 和开放端检测部分 37。

[0144] 随机目标生成部分 35 从存储在模型存储部分 22 中的扩展 HMM 的状态中随机地选择一个状态作为随机目标，将该随机目标设置为作为目标状态的内部目标，然后将该内部目标提供给目标选择部分 31。

[0145] 分支结构检测部分 36 根据存储在模型存储部分 22 中的扩展 HMM 的状态转移概率来检测分支结构的状态，所述状态可以在执行同一行为时进行向不同状态的状态转移，分支结构检测部分 36 将分支结构的状态设置为作为目标状态的内部目标，然后将该内部目标提供给目标选择部分 31。

[0146] 顺便提及，当分支结构检测部分 36 从扩展 HMM 检测到多个状态作为分支结构的状态时，目标选择部分 31 查找过去时间管理表存储部分 32 中的过去时间管理表，并且在分支结构的多个状态中选择其过去时间是最大值的分支结构的状态作为目标状态。

[0147] 开放端检测部分 37 检测开放端，开放端在模型存储部分 22 中存储的扩展 HMM 中

具有可以以其中观察到预定观察值的状态作为转移源进行的状态转移中要进行的状态转移,并且开放端是其中观察到与预定观察值相同的观察值的其他状态。然后,开放端检测部分 37 将开放端设置为作为目标状态的内部目标,并且将该内部目标提供给目标选择部分 31。

[0148] [反射行为模式中的处理]

[0149] 图 5 是帮助说明反射行为模式中的处理的流程图,该处理由图 4 的代理执行。

[0150] 在步骤 S11 中,反射行为确定部分 11 将用于计时的变量 t 设置为例如 1 作为初始值。然后处理进行到步骤 S12。

[0151] 在步骤 S12 中,传感器 13 从行为环境中获得当前观察值(时刻 t 处的观察值) o_t ,并且输出观察值 o_t 。然后处理进行到步骤 S13。

[0152] 在本实施例中时刻 t 处的观察值 o_t 是图 3B 中所示的 15 个观察值 0_1 到 0_{15} 之一。

[0153] 在步骤 S13 中,代理将由传感器 13 输出的观察值 o_t 提供给反射行为确定部分 11。然后处理进行到步骤 S14。

[0154] 在步骤 S14 中,反射行为确定部分 11 根据固有规则确定针对来自传感器 13 的观察值 o_t 在时间 t 处要执行的行为 u_t ,并且根据行为 u_t 控制执行器 12。然后处理进行到步骤 S15。

[0155] 在本实施例中在时间 t 处的行为 u_t 是图 3A 中所示的 5 个行为 U_1 至 U_5 之一。

[0156] 以下也将步骤 S14 中所确定的行为 u_t 称为被确定行为 u_t 。

[0157] 在步骤 S15 中,执行器 12 在反射行为确定部分 11 的控制下执行驱动。代理由此执行被确定行为 u_t 。

[0158] 此时,传感器 13 观察执行器 12,并且输出由代理执行的行为 u_t (代表行为 u_t 的符号)。

[0159] 然后处理从步骤 S15 进行到步骤 S16,在步骤 S16 中,历史存储部分 14 以被添加到已经被存储为观察值和行为的历史的一系列观察值和行为中的形式存储由传感器 13 输出的观察值 o_t 和行为 u_t 。然后处理进行到步骤 S17。

[0160] 在步骤 S17 中,反射行为确定部分 11 确定代理是否已经执行了行为多次,次数预先被指定(设置)为反射行为模式要执行的行为的数量。

[0161] 当在步骤 S17 中确定代理还没有执行行为预先指定的次数时,处理进行到步骤 S18,在步骤 S18 中,反射行为确定部分 11 将时刻 t 递增 1。然后处理从步骤 S18 返回到步骤 S12,以其后重复类似的处理。

[0162] 当在步骤 S17 中确定代理已经执行了行为预先指定的次数时,即当时刻 t 等于预先指定的次数时,在反射行为模式中的处理结束。

[0163] 根据在反射行为模式中的处理,在历史存储部分 14 中存储一系列观察值 o_t (观察值系列)和当观察到观察值 o_t 时由代理执行的一系列行为 u_t (行为系列)(一系列行为 u_t 和当已经执行了行为 u_t 时在代理中观察到的一系列值 o_{t+1})。

[0164] 然后,代理中的学习部分 21 使用历史存储部分 14 中存储的观察值系列和行为系列作为学习数据来学习扩展 HMM。

[0165] 在扩展 HMM 中,普通(现有)HMM 的状态转移概率被扩展为由代理执行的每个行为的状态转移概率。

[0166] 图 6A 和 6B 是帮助说明扩展 HMM 的状态转移概率的图。

[0167] 图 6A 示出了普通 HMM 的状态转移概率。

[0168] 现在,假定采用其中从某一状态到任意状态的状态转移均是可能的遍历 HMM 作为包括扩展 HMM 的 HMM。还假定 HMM 的状态数量为 N 。

[0169] 在这种情况下,普通 HMM 具有从 N 个状态 S_i 的每一个向 N 个状态 S_j 的每一个的 $N \times N$ 个状态转移的状态转移概率 a_{ij} 作为模型参数。

[0170] 普通 HMM 的所有状态转移概率可以用二维表表示,在该二维表中,从状态 S_i 向状态 S_j 的状态转移的状态转移概率 a_{ij} 布置在从顶部数的第 i 行和从左数的第 j 列中。

[0171] 以下也将 HMM 的状态转移概率的表描述为状态转移概率 A 。

[0172] 图 6B 示出了扩展 HMM 的状态转移概率 A 。

[0173] 扩展 HMM 具有针对由代理执行的每个行为 U_m 的状态转移概率。

[0174] 以下也将关于某一行行为 U_m 的从状态 S_i 向状态 S_j 的状态转移的状态转移概率描述为状态转移概率 $a_{ij}(U_m)$ 。

[0175] 状态转移概率 $a_{ij}(U_m)$ 代表当代理执行行为 U_m 时从状态 S_i 向状态 S_j 发生状态转移的概率。

[0176] 扩展 HMM 的所有状态转移概率可以用三维表表示,在该三维表中,关于行为 U_m 的从状态 S_i 向状态 S_j 的状态转移的状态转移概率 $a_{ij}(U_m)$ 布置在从顶部数的第 i 行、从左数的第 j 列和在从前侧起的深度方向上的第 m 平面中。

[0177] 以下,在状态转移概率 A 的三维表中,将垂直方向上的轴称为 i 轴,将水平方向上的轴称为 j 轴,并且将深度方向上的轴称为 m 轴或者行为轴。

[0178] 另外,由状态转移概率 $a_{ij}(U_m)$ 构成的平面也将被称为关于行为 U_m 的状态转移概率平面,所述平面是通过使用行为轴上在位置 m 处的垂直于行为轴的平面来切割状态转移概率 A 的三维表而获得的。

[0179] 此外,由状态转移概率 $a_{ij}(U_m)$ 构成的平面也将被称为关于状态 S_i 的行为平面,所述平面是通过使用 i 轴上在位置 I 处的垂直于 i 轴的平面来切割状态转移概率 A 的三维表而获得的。

[0180] 形成关于状态 S_i 的行为平面的状态转移概率 $a_{ij}(U_m)$ 代表当以状态 S_i 作为转移源发生状态转移时执行每个行为 U_m 的概率。

[0181] 顺便提及,如同普通 HMM 的情况,扩展 HMM 不仅具有针对每个行为的状态转移概率 $a_{ij}(U_m)$,而且具有在初始时刻 $t = 1$ 处位于状态 S_i 中的初始状态概率 π_i 和在状态 S_i 中观察到观察值 O_k 的观察概率 $b_i(O_k)$ 作为模型参数。

[0182] [学习扩展 HMM]

[0183] 图 7 是帮助说明学习扩展 HMM 的处理的流程图,所述处理由图 4 中的学习部分 21 使用作为学习数据存储在历史存储部分 14 中的观察值系列和行为系列来执行。

[0184] 在步骤 S21 中,学习部分 21 初始化扩展 HMM。

[0185] 具体地,学习部分 21 初始化作为存储在模型存储部分 22 中的扩展 HMM 的模型参数的初始状态概率 π_i 、(每个行为的)状态转移概率 $a_{ij}(U_m)$ 和观察概率 $b_i(O_k)$ 。

[0186] 顺便提及,假定扩展 HMM 的状态的数量(总数)是 N ,则初始状态概率 π_i 被初始化为例如 $1/N$ 。假定在这种情况下,作为二维平面中的迷宫的行为环境由 $a \times b$ 个观察单元、

即 a 个水平观察单元和 b 个垂直观察单元构成,则可以采用 $(a+\Delta) \times (b+\Delta)$ 个状态的数量作为扩展 HMM 的状态的数量 N ,其中 Δ 是作为裕量的整数。

[0187] 另外,状态转移概率 $a_{ij}(U_m)$ 和观察概率 $b_i(O_k)$ 被初始化为例如可以取值为概率值的随机值。

[0188] 在这种情况下,初始化状态转移概率 $a_{ij}(U_m)$,以使得在关于每个行为 U_m 状态转移概率平面中每行的状态转移概率 $a_{ij}(U_m)$ 的总和 $(a_{i,1}(U_m)+a_{i,2}(U_m)+\cdots+a_{i,N}(U_m))$ 是 1.0。

[0189] 类似地,初始化观察概率 $b_i(O_k)$,以使得从每个状态 S_i 观察到观察值 $O_1, O_2, \cdots, O_K$ 的观察概率的总和 $(b_i(O_1)+b_i(O_2)+\cdots+b_i(O_K))$ 是 1.0。

[0190] 顺便提及,当执行所谓的追加学习时,存储在模型存储部分 22 中的扩展 HMM 的初始状态概率 π_i 、状态转移概率 $a_{ij}(U_m)$ 和观察概率 $b_i(O_k)$ 被原样用作初始值。即,不执行步骤 S21 中的初始化。

[0191] 在步骤 S21 后,处理进行到步骤 S22。从步骤 S22 向下,执行扩展 HMM 的学习,其使用作为学习数据存储在历史存储部分 14 中的行为系列和观察值系列、根据 Baum-Welch 重评估法(通过关于行为扩展 Baum-Welch 重评估法而获得的方法)估计初始状态概率 π_i 、每个行为的状态转移概率 $a_{ij}(U_m)$ 和观察概率 $b_i(O_k)$ 。

[0192] 具体地,在步骤 S22 中,学习部分 21 计算前向概率 $\alpha_{t+1}(j)$ 和后向概率 $\beta_t(i)$ 。

[0193] 在扩展 HMM 中,当在时刻 t 处执行行为 u_t 时,发生从当前状态 S_i 向状态 S_j 的状态转移,并且在下一个时刻 $t+1$ 处在状态转移后的状态 S_j 中观察到观察值 o_{t+1} 。

[0194] 在这样的扩展 HMM 中,前向概率 $\alpha_{t+1}(j)$ 是在模型 Λ 中观察到学习数据的行为系列 $u_1, u_2, \cdots, u_t$, 观察到观察值系列 $o_1, o_2, \cdots, o_{t+1}$ 并且代理在时刻 $t+1$ 处位于状态 S_j 中的概率 $P(o_1, o_2, \cdots, o_{t+1}, u_1, u_2, \cdots, u_t, s_{t+1}=j | \Lambda)$, 模型 Λ 是当前的扩展 HMM(扩展 HMM 由实际存储在模型存储部分 22 中的初始状态概率 π_i 、状态转移概率 $a_{ij}(U_m)$ 和观察概率 $b_i(O_k)$ 定义)。前向概率 $\alpha_{t+1}(j)$ 通过式 (1) 来表达。

[0195] $\alpha_{t+1}(j) = P(o_1, o_2, \dots, o_{t+1}, u_1, u_2, \dots, u_t, s_{t+1} = j | \Lambda)$

$$[0196] = \sum_{i=1}^N \alpha_t(i) a_{ij}(u_t) b_j(o_{t+1}) \quad \dots (1)$$

[0197] 顺便提及,状态 s_t 代表在时刻 t 处的状态,并且当扩展 HMM 的状态的数量是 N 时是状态 S_1 到 S_N 之一。另外,式 $s_{t+1}=j$ 表示在时刻 $t+1$ 处的状态 s_{t+1} 是状态 S_j 。

[0198] 式 (1) 的前向概率 $\alpha_{t+1}(j)$ 代表当代理观察到学习数据的行为系列 $u_1, u_2, \cdots, u_{t-1}$ 和观察值系列 $o_1, o_2, \cdots, o_t$ 并且在时刻 t 处位于状态 s_t 时,在通过执行(观察)行为 u_t 来进行状态转移后,代理在时刻 $t+1$ 处位于状态 S_j 中并且观察到观察值 o_{t+1} 的概率。

[0199] 顺便提及,前向概率 $\alpha_{t+1}(j)$ 的初始值 $\alpha_1(j)$ 通过式 (2) 来表达。

$$[0200] \alpha_1(j) = \pi_j b_j(o_1) \dots (2)$$

[0201] 式 (2) 的初始值 $\alpha_1(j)$ 代表起初(时刻 $t=0$) 处于状态 S_j 中并且观察到观察值 o_1 的概率。

[0202] 另外,在扩展 HMM 中,后向概率 $\beta_t(i)$ 是在模型 Λ 中代理在时刻 t 处位于状态 S_i 中,其后观察到学习数据的行为系列 $u_{t+1}, u_{t+2}, \cdots, u_{T-1}$ 并且观察到观察值系列 $o_{t+1}, o_{t+2}, \cdots, o_T$ 的概率 $P(o_{t+1}, o_{t+2}, \cdots, o_T, u_{t+1}, u_{t+2}, \cdots, u_{T-1}, s_t = i | \Lambda)$, 模型 Λ 是当前的扩展 HMM。后向概率 $\beta_t(i)$ 通过式 (3) 来表达。

$$[0203] \quad \beta_t(i) = P(o_{t+1}, o_{t+2}, \dots, o_T, u_{t+1}, u_{t+2}, \dots, u_{T-1}, s_t = j | \Lambda)$$

$$[0204] \quad = \sum_{i=1}^N a_{ij}(u_{t+1}) b_j(o_{t+1}) \beta_{t+1}(j) \quad \dots (3)$$

[0205] 顺便提及, T 表示学习数据的观察值系列的观察值的数量。

[0206] 式 (3) 的后向概率 $\beta_t(i)$ 代表当在时刻 t+1 处的状态 s_{t+1} 是状态 S_j 并且在通过代理在时刻 t+1 处位于状态 S_j 的情况下在时刻 t 处在状态 S_i 中执行 (观察) 行为 u_t 而进行状态转移后观察到观察值 o_{t+1} , 其后观察到学习数据的行为系列 $u_{t+1}, u_{t+2}, \dots, u_{T-1}$ 并且观察到观察值系列 $o_{t+2}, o_{t+3}, \dots, o_T$ 时, 在时刻 t 处的状态 s_t 是状态 S_i 的概率。

[0207] 顺便提及, 后向概率 $\beta_t(i)$ 的初始值 $\beta_T(i)$ 通过式 (4) 来表达。

$$[0208] \quad \beta_T(i) = 1 \dots (4)$$

[0209] 式 (4) 的初始值 $\beta_T(i)$ 表示在结尾处 (时刻 t = T) 处于状态 S_i 的概率是 1.0, 即代理在结尾处总是处于状态 S_i 。

[0210] 扩展 HMM 与普通 HMM 的不同在于: 扩展 HMM 使用每个行为的状态转移概率 $a_{ij}(u_t)$ 作为从某一状态 S_i 向某一状态 S_j 的状态转移的状态转移概率, 如式 (1) 和式 (3) 中所示。

[0211] 在步骤 S22 中计算了前向概率 $\alpha_{t+1}(j)$ 和后向概率 $\beta_t(i)$ 后, 处理进行到步骤 S23, 在步骤 S23 中, 学习部分 21 使用前向概率 $\alpha_{t+1}(j)$ 和后向概率 $\beta_t(i)$ 重评估初始状态概率 π_i 、每个行为 U_m 的状态转移概率 $a_{ij}(U_m)$ 和观察概率 $b_i(o_k)$ 作为扩展 HMM 的模型参数 Λ 。

[0212] 在这种情况下, 当状态转移概率被扩展到每个行为 U_m 的状态转移概率 $a_{ij}(U_m)$ 时, 如下通过扩展 Baum-Welch 重评估法执行模型参数的重评估。

[0213] 使用前向概率 $\alpha_t(i)$ 和后向概率 $\beta_{t+1}(j)$, 通过式 (5) 来表达状态转移的概率 $\xi_{t+1}(i, j, U_m)$, 其中通过在作为当前扩展 HMM 的模型 Λ 中观察到行为系列 $U = u_1, u_2, \dots, u_{T-1}$ 和观察值系列 $O = o_1, o_2, \dots, o_T$ 的情况下在时刻 t 处在状态 S_i 中执行行为 U_m 来进行向时刻 t+1 处状态 S_j 的所述状态转移。

$$[0214] \quad \xi_{t+1}(i, j, U_m) = P(s_t = i, s_{t+1} = j, u_t = U_m | O, U, \Lambda)$$

$$[0215] \quad = \frac{\alpha_t(i) a_{ij}(U_m) b_j(o_{t+1}) \beta_{t+1}(j)}{P(O, U | \Lambda)} \quad (1 \leq t \leq T-1) \quad \dots (5)$$

[0216] 另外, 在时间 t 处在状态 S_i 中执行行为 $u_t = U_m$ 的概率 $\gamma_t(i, U_m)$ 可以被计算为通过针对时间 t+1 处的状态 S_j 边缘化概率 $\xi_{t+1}(i, j, U_m)$ 而获得的概率。概率 $\gamma_t(i, U_m)$ 通过式 (6) 来表达。

$$[0217] \quad \gamma_t(i, U_m) = P(s_t = i, u_t = U_m | O, U, \Lambda)$$

$$[0218] \quad = \sum_{j=1}^N \xi_{t+1}(i, j, U_m) \quad (1 \leq t \leq T-1) \quad \dots (6)$$

[0219] 学习部分 21 使用式 (5) 的概率 $\xi_{t+1}(i, j, U_m)$ 和式 (6) 的概率 $\gamma_t(i, U_m)$ 来重新估计扩展 HMM 的模型参数 Λ 。

[0220] 假定通过重新估计模型参数 Λ 而获得的估计值被使用符号 (') 表示为模型参数 Λ' , 根据式 (7) 来获得初始状态概率的估计值 π'_i , 所述估计值是模型参数 Λ' 。

$$[0221] \quad \pi'_i = \frac{\alpha_1(i)\beta_1(i)}{P(O, U | \Lambda)} \quad (1 \leq i \leq N) \quad \dots(7)$$

[0222] 另外,根据式 (8) 来获得每个行为的状态转移概率的估计值 $a'_{ij}(U_m)$, 所述估计值是模型参数 Λ' 。

$$[0223] \quad a'_{ij}(U_m) = \frac{\sum_{t=1}^{T-1} \xi_{t+1}(i, j, U_m)}{\sum_{t=1}^{T-1} \gamma_t(i, U_m)} = \frac{\sum_{t=1}^{T-1} \alpha_t(i) a_{ij}(U_m) b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{t=1}^{T-1} \sum_{j=1}^N \alpha_t(i) a_{ij}(U_m) b_j(o_{t+1}) \beta_{t+1}(j)} \quad \dots(8)$$

[0224] 式 (8) 的状态转移概率的估计值 $a'_{ij}(U_m)$ 的分子代表通过在状态 S_i 中执行行为 $u_t = U_m$ 而向状态 S_j 进行状态转移的次数的期望值, 并且式 (8) 的状态转移概率的估计值 $a'_{ij}(U_m)$ 的分母代表通过在状态 S_i 中执行行为 $u_t = U_m$ 而进行状态转移的次数的期望值。

[0225] 根据式 (9) 来获得观察概率的估计值 $b'_j(o_k)$, 所述估计值是模型参数 Λ' 。

$$[0226] \quad b'_j(o_k) = \frac{\sum_{t=1}^{T-1} \sum_{i=1}^N \sum_{m=1}^M \xi_{t+1}(i, j, U_m, o_k)}{\sum_{t=1}^{T-1} \sum_{i=1}^N \sum_{m=1}^M \xi_{t+1}(i, j, U_m)} = \frac{\sum_{t=1}^{T-1} \alpha_{t+1}(j) b_j(o_k) \beta_{t+1}(j)}{\sum_{t=1}^{T-1} \alpha_{t+1}(j) \beta_{t+1}(j)} \quad \dots(9)$$

[0227] 式 (9) 的观察概率的估计值 $b'_j(o_k)$ 的分子代表向状态 S_j 进行状态转移并且在状态 S_j 中观察到观察值 o_k 的次数的期望值, 并且式 (9) 的观察概率的估计值 $b'_j(o_k)$ 的分母代表向状态 S_j 进行状态转移的次数的期望值。

[0228] 在步骤 S23 中重新估计初始状态概率、状态转移概率和观察概率的估计值 π'_i 、 $a'_{ij}(U_m)$ 和 $b'_j(o_k)$ 作为模型参数 Λ' 后, 学习部分 21 将以覆盖的方式分别将估计值 π'_i 作为新初始状态概率 π_i 、估计值 $a'_{ij}(U_m)$ 作为新状态转移概率 $a_{ij}(U_m)$ 和估计值 $b'_j(o_k)$ 作为新观察概率 $b_j(o_k)$ 存储在模型存储部分 22 中。然后处理进行到步骤 S24。

[0229] 在步骤 S24 中, 确定扩展 HMM 的模型参数即存储在模型存储部分 22 中的 (新) 初始状态概率 π_i 、(新) 状态转移概率 $a_{ij}(U_m)$ 和 (新) 观察概率 $b_i(o_k)$ 是否已经收敛。

[0230] 当在步骤 S24 中确定扩展 HMM 的模型参数还没有收敛时, 处理返回到步骤 S22, 以使用存储在模型存储部分 22 中的新初始状态概率 π_i 、新状态转移概率 $a_{ij}(U_m)$ 和新观察概率 $b_i(o_k)$ 重复类似的处理。

[0231] 当在步骤 S24 中确定扩展 HMM 的模型参数已经收敛时, 即例如当步骤 S23 中重新估计后的扩展 HMM 的模型参数几乎未从步骤 S23 中的重新估计之前的扩展 HMM 的模型参数改变时, 学习扩展 HMM 的处理结束。

[0232] 如上所述, 通过使用由代理执行的行为的行为系列和当代理已经执行了行为时在代理中观察到的观察值的观察值系列, 学习由每个行为的状态转移概率 $a_{ij}(U_m)$ 定义的扩展 HMM, 通过在扩展 HMM 中的观察值系列获得行为环境的结构, 并且获得每个观察值和当观察到该观察值时执行的行为之间的关系 (由代理执行的行为和当已经执行了行为时观察到的观察值 (在所述行为之后观察到的观察值) 之间的关系)。

[0233] 结果, 如下文中将描述的, 通过使用这种学习后的扩展 HMM 可以在识别行为模式中将适当的行为确定为要由行为环境中的代理执行的行为。

[0234] [识别行为模式中的处理]

[0235] 图 8 是帮助说明识别行为模式中的处理的流程图, 所述处理由图 4 中的代理执行。

[0236] 在识别行为模式中,如上所述,代理确定目标、识别当前状况并且计算行为计划以从当前状况达到目标。此外,代理根据该行为计划确定接下来要执行的行为,并且执行该行为。然后代理重复上述处理。

[0237] 在步骤 S31 中,状态识别部分 23 将用于计时的变量 t 设置为例如 1 作为初始值。然后处理进行到步骤 S32。

[0238] 在步骤 S32 中,传感器 13 从行为环境中获得当前观察值(在时刻 t 处的观察值) o_t ,并且输出观察值 o_t 。然后处理进行到步骤 S33。

[0239] 在步骤 S33 中,历史存储部分 14 将由传感器 13 获得的在时刻 t 处的观察值 o_t 和当观察到观察值 o_t 时(紧邻在传感器 13 获得观察值 o_t 之前)由传感器 13 输出的行为 u_{t-1} (由代理在紧邻在前的时刻 $t-1$ 处执行的行为 u_{t-1})以被添加到已经存储的观察值和行为系列中的形式存储为观察值和行为的历史。然后处理进行到步骤 S34。

[0240] 在步骤 S34 中,状态识别部分 23 使用由代理执行的行为和当已经执行了行为时在代理中观察到的观察值、根据扩展 HMM 识别代理的当前状况,并且获得作为对应于当前状况的扩展 HMM 的状态的当前状态。

[0241] 具体地,状态识别部分 23 从历史存储部分 14 中读出 0 个或者更多个最新行为的行为系列以及一个或更多个最新观察值的观察值系列作为用于识别的行为系列和观察值系列,所述系列用于识别代理的当前状况。

[0242] 此外,状态识别部分 23 观察用于识别的行为系列和观察值系列,并且在模型存储部分 22 中存储的所学习的扩展 HMM 中,根据维特比算法(通过将维特比算法扩展到行为而获得的算法)获得最佳状态概率 $\delta_t(j)$ 和最佳路径 $\psi_t(j)$,最佳状态概率 $\delta_t(j)$ 是在时刻(当前时刻) t 处位于状态 S_j 中的状态概率的最大值,最佳路径 $\psi_t(j)$ 是提供最佳状态概率 $\delta_t(j)$ 的状态系列。

[0243] 根据维特比算法,可以从当在普通 HMM 中观察到某一观察值系列时跟踪的状态的系列(状态系列)中估计最大化观察到该观察值系列的可能性的状态系列(最大可能状态系列)。

[0244] 但是,因为在扩展 HMM 中状态转移概率关于行为进行了扩展,因此需要关于行为扩展维特比算法,以便使其应用于扩展 HMM。

[0245] 因此,状态识别部分 23 分别根据式 (10) 和式 (11) 来获得最佳状态概率 $\delta_t(j)$ 和最佳路径 $\psi_t(j)$ 。

[0246]

$$\delta_t(i) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}(u_{t-1}) b_{ij}(o_t)] \quad (1 \leq t \leq T, 1 \leq j \leq N) \quad \dots(10)$$

[0247]

$$\psi_t(j) = \arg \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}(u_{t-1}) b_{ij}(o_t)] \quad (1 \leq t \leq T, 1 \leq j \leq N) \quad \dots(11)$$

[0248] 在这种情况下,式 (10) 中的 $\max[X]$ 表示当指示状态 S_i 的下标 i 改变为范围 1 到 N 中的整数时所获得的 X 的最大值, N 是状态的数量。另外,式 (11) 中的 $\arg \max[X]$ 表示当下标 i 改变为范围 1 到 N 中的整数时最大化所获得的 X 的下标 i 。

[0249] 状态识别部分 23 观察用于识别的行为系列和观察值系列,并且获得最大可能状态系列,其是根据式 (11) 的最佳路径 $\psi_t(j)$ 在时间 t 处到达状态 S_j 并且最大化式 (10) 的

最佳状态概率 $\delta_t(j)$ 的状态系列。

[0250] 此外,状态识别部分 23 将该最大可能状态系列设置为当前状况的识别结果并且获得(估计)最大可能状态系列的最后状态作为当前状态 s_t 。

[0251] 在获得当前状态 s_t 后,状态识别部分 23 根据当前状态 s_t 更新存储在过去时间管理表存储部分 32 中的过去时间管理表。然后处理从步骤 S34 进行到步骤 S35。

[0252] 具体地,与扩展 HMM 的每个状态相关联,在过去时间管理表存储部分 32 的过去时间管理表中登记自从状态变为当前状态起过去的过去时间。状态识别部分 23 将已经变为当前状态 s_t 的状态的过去时间重置为例如 0,并且在过去时间管理表中将其他状态的过去时间递增例如 1。

[0253] 如上所述,当目标选择部分 31 选择目标状态时,在需要时查找过去时间管理表。

[0254] 在步骤 S35 中,状态识别部分 23 根据当前状态 s_t 更新存储在模型存储部分 22 中的抑制因子。下面将描述抑制因子的更新。

[0255] 此外,在步骤 S35 中,状态识别部分 23 将当前状态 s_t 提供给行为确定部分 24。然后处理进行到步骤 S36。

[0256] 在步骤 S36 中,目标确定部分 16 从扩展 HMM 的状态中确定目标状态,并且将目标状态提供给行为确定部分 24。然后处理进行到步骤 S37。

[0257] 在步骤 S37 中,行为确定部分 24 使用模型存储部分 22 中存储的抑制因子(在紧邻在前的步骤 S35 中更新的抑制因子)来修正存储在同一模型存储部分 22 中的扩展 HMM 的状态转移概率。由此行为确定部分 24 将修正的转移概率计算为修正后的状态转移概率。

[0258] 在后面将要描述的由行为确定部分 24 对行为计划的计算中,被修正的转移概率被用作扩展 HMM 的状态转移概率。

[0259] 在步骤 S37 之后,处理进行到步骤 S38,在步骤 S38 中,行为确定部分 24 基于存储在模型存储部分 22 中的扩展 HMM、根据例如维特比算法(通过将维特比算法扩展到行为而获得的算法)计算行为计划为最大化来自状态识别部分 23 的当前状态向来自目标确定部分 16 的目标状态的状态转移的可能性的一系列行为。

[0260] 根据维特比算法,在普通 HMM 中,可以例如在从两个状态之一向另一个状态的状态系列、即从当前状态向目标状态的状态系列中,估计最大化观察到某一观察值系列的可能性的最大可能状态系列。

[0261] 但是,如上所述,因为在扩展 HMM 中状态转移概率关于行为进行了扩展,因此需要关于行为扩展维特比算法,以便应用于扩展 HMM。

[0262] 因此,行为确定部分 24 根据式 (12) 来获得状态概率 $\delta'_t(j)$ 。

[0263]

$$\delta'_t(j) = \max_{1 \leq i \leq N, 1 \leq m \leq M} [\delta'_{t-1}(i) a_{ij}(U_m)] \quad \dots(12)$$

[0264] 在这种情况下,式 (12) 中的 $\max[X]$ 表示当指示状态 S_i 的下标 i 改变为范围 1 到 N 中的整数并且指示行为 U_m 的下标 m 改变为范围 1 到 M 中的整数时获得的 X 的最大值,其中 N 是状态的数量, M 是行为的数量。

[0265] 式 (12) 通过从用于获得最佳状态概率 $\delta_t(j)$ 的式 (10) 中删除观察概率 $b_j(o_t)$ 获得。另外,在式 (12) 中,状态概率 $\delta'_t(j)$ 考虑到行为 U_m 而获得。这对应于维特比算法关于行为的扩展。

[0266] 行为确定部分 24 在前向上执行式 (12) 的计算, 并且暂时存储取最大状态概率 $\delta'_t(j)$ 的下标 i 和指示当每个时刻发生向由下标 i 表示的状态 S_i 的状态转移时执行的行为 U_m 的下标 m 。

[0267] 顺便提及, 通过由抑制因子修正学习的扩展 HMM 的状态转移概率 $a_{ij}(U_m)$ 而获得的修正的转移概率在式 (12) 的计算中用作状态转移概率 $a_{ij}(U_m)$ 。

[0268] 行为确定部分 24 以当前状态 s_t 作为第一状态来计算式 (12) 的状态概率 $\delta'_t(j)$, 并且当目标状态 S_{goal} 的状态概率 $\delta'_t(S_{goal})$ 变为预定阈值 δ'_{th} 或者更大时结束式 (12) 的状态概率 $\delta'_t(j)$ 的计算, 如式 (13) 中所示。

$$[0269] \quad \delta'_t(S_{goal}) \geq \delta'_{th} \quad \cdots (13)$$

[0270] 顺便提及, 例如, 根据式 (14) 来设置式 (13) 中的阈值 δ'_{th} 。

$$[0271] \quad \delta'_{th} = 0.9^{T'} \quad \cdots (14)$$

[0272] 在这种情况下, 式 (14) 中的 T' 表示式 (12) 的计算的次数 (从式 (12) 获得的最大可能状态系列的系列长度)。

[0273] 根据式 (14), 当发生一个可能的状态转移时通过采用 0.9 作为状态概率来设置阈值 δ'_{th} 。

[0274] 因此, 根据式 (13), 当连续发生 T' 个可能的状态转移时, 结束式 (12) 的状态概率 $\delta'_t(j)$ 的计算。

[0275] 结束式 (12) 的状态概率 $\delta'_t(j)$ 的计算, 行为确定部分 24 通过下述方式来获得从当前状态 s_t 向目标状态 S_{goal} 的最大可能状态系列 (在许多情况下为最短路径) 和当提供最大可能状态系列的状态转移发生时执行的一系列行为 U_m : 通过从结束式 (12) 的状态概率 $\delta'_t(j)$ 的计算时开始, 即在相反的方向上从目标状态 S_{goal} 向当前状态 s_t 跟踪针对状态 S_i 和行为 U_m 存储的下标 i 和 m 。

[0276] 具体地, 如上所述, 行为确定部分 24 存储取最大状态概率 $\delta'_t(j)$ 的下标 i 和表示当在前向上计算式 (12) 的状态概率 $\delta'_t(j)$ 时每个时刻发生向由下标 i 表示的状态 S_i 的状态转移时执行的行为 U_m 的下标 m 。

[0277] 每个时刻的下标 i 表示当在暂时倒退的方向上从状态 S_j 返回状态 S_i 时哪个状态 S_i 具有最大状态概率。每个时刻的下标 m 表示进行提供最大状态概率的状态转移的行为 U_m 。

[0278] 因此, 当每个时刻的下标 i 和 m 从结束式 (12) 的状态概率 $\delta'_t(j)$ 的计算的时刻逐个时刻地回溯、并且到达开始式 (12) 的状态概率 $\delta'_t(j)$ 的计算的时刻时, 获得通过布置从当前状态 s_t 向目标状态 S_{goal} 的状态系列的状态的一系列下标和当以暂时倒退的顺序发生所述状态系列的状态转移时执行的行为系列的行为的一系列下标的每个而形成的系列。

[0279] 行为确定部分 24 获得从当前状态 s_t 向目标状态 S_{goal} 的状态系列 (最大可能状态系列) 和当通过重新布置按时间顺序以暂时倒退的顺序布置的系列发生状态系列的状态转移时执行的行为系列。

[0280] 在发生从当前状态 s_t 向目标状态 S_{goal} 的最大可能状态系列的状态转移时执行的行为系列是行为计划, 这些行为系列如上所述由行为确定部分 24 获得。

[0281] 在这种情况下, 在行为确定部分 24 中获得的最大可能状态系列与行为计划一起是在代理根据行为计划执行行为时发生 (应该发生) 的状态转移的状态系列。因此, 如果

当代理根据行为计划执行行为时发生不按照最大可能状态系列的状态布置的状态转移,则即使代理根据行为计划执行行为代理也可能无法到达目标状态。

[0282] 在行为确定部分 24 如上所述在步骤 S38 中获得行为计划后,处理进行到步骤 S39,在步骤 S39 中,行为确定部分 24 根据行为计划确定接下来要执行的行为 u_t 。然后处理进行到步骤 S40。

[0283] 即,行为确定部分 24 将作为行为计划的行为系列的第一行为设置为接下来要执行的被确定行为 u_t 。

[0284] 在步骤 S40 中,行为确定部分 24 根据紧邻在前的步骤 S39 中确定的行为(被确定行为) u_t 控制执行器 12,并且代理由此执行行为 u_t 。

[0285] 其后处理从步骤 S40 进行到步骤 S41,其中状态识别部分 23 将时刻 t 递增 1。处理返回到步骤 S32,以重复从步骤 S32 向下类似的处理。

[0286] 顺便提及,例如,当操作代理以便结束识别行为模式时、当关断对代理的供电时或者当代理的模式从识别行为模式改变为另一个模式(反射行为模式等)时,结束图 8 的识别行为模式中的处理。

[0287] 如上所述,状态识别部分 23 使用由代理执行的行为和当已经执行了行为时在代理中观察到的观察值来识别代理的当前状况,并且根据扩展 HMM 获得对应于当前状况的当前状态;目标确定部分 16 确定目标状态;并且行为确定部分 24 计算行为计划,所述行为计划是基于扩展 HMM 最大化从当前状态向目标状态的状态转移的可能性(状态概率)的一系列行为,并且行为确定部分 24 根据该行为计划确定接下来要由代理执行的行为。因此,可以将适当的行为确定为要由代理执行的行为以便代理到达目标状态。

[0288] 现有的行为确定方法彼此单独地准备了学习观察值系列的状态转移概率模型和作为用于实现状态转移概率模型的状态转移的行为的模型的行为模型,并且进行学习。

[0289] 因此,由于学习了状态转移概率模型和行为模型这两个模型,因此学习需要大量的计算成本和存储资源。

[0290] 另一方面,图 4 的代理学习在扩展 HMM 中彼此相关联的观察值系列和行为系列,所述扩展 HMM 是单个模型。因此,可以使用少量的计算成本和存储资源来执行学习。

[0291] 另外,现有的行为确定方法需要使用状态转移概率模型来计算直到目标状态的状态系列,并且使用行为模型来计算获得该状态系列的行为。即,现有的行为确定方法需要使用单独的模型计算直到目标状态的状态系列,并且计算获得该状态系列的行为。

[0292] 因此,现有的行为确定方法需要直到行为的计算的大量的计算成本。

[0293] 另一方面,图 4 的代理可以同时确定从当前状态到目标状态的最大可能状态系列和用于获得最大可能状态系列的行为系列。因此,可以以少量的计算成本来确定接下来要由代理执行的行为。

[0294] [目标状态的确定]

[0295] 图 9 是帮助说明确定目标状态的处理的流程图,所述处理由图 4 中的目标确定部分 16 在图 8 中的步骤 S36 中执行。

[0296] 在目标确定部分 16 中,在步骤 S51 中,目标选择部分 31 确定是否设置了外部目标。

[0297] 当在步骤 S51 中确定设置了外部目标时,即当例如用户操作外部目标输入部分 33

以将模型存储部分 22 中存储的扩展 HMM 的一个状态指定为作为目标状态的外部目标,并且目标状态(指示目标状态的下标)从外部目标输入部分 33 提供给目标选择部分 31 时,处理进行到步骤 S52,在步骤 S52 中,目标选择部分 31 从外部目标输入部分 33 选择外部目标,然后将外部目标提供给行为确定部分 24。然后处理返回。

[0298] 顺便提及,用户不仅可以通过操作外部目标输入部分 33 而且可以通过操作诸如 PC(个人计算机)等的终端来指定作为目标状态的状态(状态的下标)。在这种情况下,外部目标输入部分 33 通过与由用户操作的终端通信来识别由用户指定的状态,然后将该状态提供给目标选择部分 31。

[0299] 另一方面,当在步骤 S51 中确定未设置外部目标时,处理进行到步骤 S53,在步骤 S53 中,开放端检测部分 37 根据模型存储部分 22 中存储的扩展 HMM 从扩展 HMM 的状态检测开放端。然后处理进行到步骤 S54。

[0300] 在步骤 S54 中,目标选择部分 31 确定是否检测到开放端。

[0301] 当开放端检测部分 37 从扩展 HMM 的状态检测到开放端时,开放端检测部分 37 将作为开放端的状态(指示状态的下标)提供给目标选择部分 31。目标选择部分 31 根据是否从开放端检测部分 37 提供了开放端来确定是否检测到开放端。

[0302] 当在步骤 S54 中确定检测到开放端时,即当一个或多个开放端从开放端检测部分 37 提供给目标选择部分 31 时,处理进行到步骤 S55,在步骤 S55 中,目标选择部分 31 从来自开放端检测部分 37 的一个或多个开放端中选择例如具有最小的指示状态的下标的开放端作为目标状态,然后将该目标状态提供给行为确定部分 24。然后处理返回。

[0303] 当在步骤 S54 中确定未检测到开放端时,即当没有开放端从开放端检测部分 37 提供给目标选择部分 31 时,处理进行到步骤 S56,在步骤 S56 中,分支结构检测部分 36 根据模型存储部分 22 中存储的扩展 HMM 从扩展 HMM 的状态中检测分支结构的状态。然后处理进行到步骤 S57。

[0304] 在步骤 S57 中,目标选择部分 31 确定是否检测到分支结构的状态。

[0305] 在这种情况下,当分支结构检测部分 36 从扩展 HMM 的状态中检测到分支结构的状态时,分支结构检测部分 36 将该分支结构的状态(指示分支结构的状态的下标)提供给目标选择部分 31。目标选择部分 31 根据是否从分支结构检测部分 36 提供了分支结构的状态来确定是否检测到分支结构的状态。

[0306] 当在步骤 S57 中确定检测到分支结构的状态时,即当一个或多个分支结构的状态从分支结构检测部分 36 提供给目标选择部分 31 时,处理进行到步骤 S58,在步骤 S58 中,目标选择部分 31 从来自分支结构检测部分 36 的一个或多个分支结构的状态中选择一个状态作为目标状态,然后将目标状态提供给行为确定部分 24。然后处理返回。

[0307] 具体地,查找过去时间管理表存储部分 32 中的过去时间管理表,目标选择部分 31 识别来自分支结构检测部分 36 的一个或多个分支结构的状态的过去时间。

[0308] 此外,目标选择部分 31 从来自分支结构检测部分 36 的一个或多个分支结构的状态中检测具有最长过去时间的状态,并且选择该状态作为目标状态。

[0309] 另一方面,当在步骤 S57 中确定未检测到分支结构的状态时,即当没有分支结构的状态从分支结构检测部分 36 提供给目标选择部分 31 时,处理进行到步骤 S59,在步骤 S59 中,随机目标生成部分 35 随机选择模型存储部分 22 中存储的扩展 HMM 的一个状态,然

后将该状态提供给目标选择部分 31。

[0310] 此外,在步骤 S59 中,目标选择部分 31 选择来自随机目标生成部分 35 的状态作为目标状态,然后将目标状态提供给行为确定部分 24。然后处理返回。

[0311] 顺便提及,后面将详细描述由开放端检测部分 37 的开放端检测和由分支结构检测部分 36 的分支结构的检测。

[0312] [行为计划的计算]

[0313] 图 10A、10B 和 10C 是帮助说明由图 4 中的行为确定部分 24 计算行为计划的图。

[0314] 图 10A 示意性地示出了用于计算行为计划的扩展 HMM。

[0315] 在图 10A 中,圆圈 (o) 标记代表扩展 HMM 的状态,并且在圆圈标记中输入的数字是由圆圈标记代表的状态的下标。指示由圆圈标记代表的状态的箭头表示可能的状态转移 (其状态转移概率不是 0.0 (可以视为 0.0 的值) 的状态转移)。

[0316] 在图 10A 的扩展 HMM 中,在对应于状态 S_i 的观察单元的位置处布置状态 S_i 。

[0317] 其间可能进行状态转移的两个状态代表代理可以在对应于两个相应的状态之间的两个观察单元之间移动。因此,指示扩展 HMM 的状态转移的箭头表示通道,代理可以在行为环境中移动通过这个通道。

[0318] 在图 10A 中,存在布置两个 (多个) 状态 S_i 和 $S_{i'}$, 以便在一个观察单元的位置处彼此重叠的情况。这表示这两个 (多个) 状态 S_i 和 $S_{i'}$ 对应于所述一个观察单元。

[0319] 例如,在图 10A 中,状态 S_3 和 S_{30} 对应于一个观察单元,并且状态 S_{34} 和 S_{35} 也对应于一个观察单元。类似地,状态 S_{21} 和 S_{23} 对应于一个观察单元,状态 S_2 和 S_{17} 对应于一个观察单元,状态 S_{37} 和 S_{48} 对应于一个观察单元,状态 S_{31} 和 S_{32} 对应于一个观察单元。

[0320] 当使用在其结构改变的行为环境中获得的观察值系列和行为系列作为学习数据来学习扩展 HMM 时,获得如图 10A 所示的其中多个状态对应于一个观察单元的扩展 HMM。

[0321] 即,在图 10A 中,使用例如在其中在对应于状态 S_{21} 和 S_{23} 的观察单元与对应于状态 S_2 和 S_{17} 的观察单元之间存在墙壁和通道之一的结构的行为环境中获得的观察值系列和行为系列作为学习数据来学习扩展 HMM。

[0322] 另外,在图 10A 中,还使用在其中在对应于状态 S_{21} 和 S_{23} 的观察单元与对应于状态 S_2 和 S_{17} 的观察单元之间存在墙壁和通道中的另一个的结构的行为环境中获得的观察值系列和行为系列作为学习数据来学习扩展 HMM。

[0323] 结果,在图 10A 的扩展 HMM 中,通过状态 S_{21} 和状态 S_{17} 获得在对应于状态 S_{21} 和 S_{23} 的观察单元与对应于状态 S_2 和 S_{17} 的观察单元之间具有墙壁的结构的行为环境。

[0324] 即,在扩展 HMM 中,获得下述行为环境的结构:其中,由于在对应于状态 S_{21} 和 S_{23} 的观察单元的状态 S_{21} 与对应于状态 S_2 和 S_{17} 的观察单元的状态 S_{17} 之间存在墙壁,因此不进行状态转移并且不允许通道。

[0325] 另外,在扩展 HMM 中,通过状态 S_{23} 和状态 S_2 来获得在对应于状态 S_{21} 和 S_{23} 的观察单元与对应于状态 S_2 和 S_{17} 的观察单元之间具有通道的结构的行为环境。

[0326] 即,在扩展 HMM 中,获得下述行为环境的结构:其中,进行状态转移,并且在对应于状态 S_{21} 和 S_{23} 的观察单元的状态 S_{23} 与对应于状态 S_2 和 S_{17} 的观察单元的状态 S_2 之间允许通道。

[0327] 如上所述,在扩展 HMM 中,即使当行为环境的结构改变时,也可以获得行为环境的

结构,所述结构发生改变。

[0328] 图 10B 和图 10C 示出了由行为确定部分 24 计算的行为计划的示例。

[0329] 在图 10B 和图 10C 中,图 10A 中的状态 S_{30} (或者状态 S_3) 是目标状态,对应于代理所位于的观察单元的状态 S_{28} 被设置为当前状态,并且计算用于从当前状态到达目标状态的行为计划。

[0330] 图 10B 示出了由行为确定部分 24 在时刻 $t = 1$ 处计算的行为计划 PL1。

[0331] 在图 10B 中,图 10A 中的一系列状态 S_{28} 、 S_{23} 、 S_2 、 S_{16} 、 S_{22} 、 S_{29} 和 S_{30} 被设置为从当前状态到达目标状态的最大可能状态系列,并且计算当提供最大可能状态系列的状态转移发生时执行的的行为系列作为行为计划 PL1。

[0332] 行为确定部分 24 将行为计划 PL1 的从第一状态 S_{28} 向下一个状态 S_{23} 移动的行为设置为被确定行为。代理执行被确定行为。

[0333] 结果,代理从对应于作为当前状态的状态 S_{28} 的观察单元在向右方向上向对应于状态 S_{21} 和 S_{23} 的观察单元移动 (执行图 3A 的行为 U_2),并且从时刻 $t = 1$ 经过一个时刻而导致时刻 t 变为时刻 $t = 2$ 。

[0334] 在图 10B 中 (在图 10C 中同样),在对应于状态 S_{21} 和 S_{23} 的观察单元与对应于状态 S_2 和 S_{17} 的观察单元之间有墙壁。

[0335] 如上所述,在其中获得在对应于状态 S_{21} 和 S_{23} 的观察单元与对应于状态 S_2 和 S_{17} 的观察单元之间有墙壁的结构的状态中,对应于状态 S_{21} 和 S_{23} 的观察单元是状态 S_{21} 。在时刻 $t = 2$ 处,状态识别部分 23 识别出当前状态是状态 S_{21} 。

[0336] 状态识别部分 23 更新用于抑制状态转移的抑制因子,以便关于由代理在从当前状态紧邻在前的状态向当前状态的状态转移时执行的行为,抑制在紧邻在前的状态和当前状态以外的状态之间的状态转移,并且不抑制紧邻在前的状态和当前状态之间的状态转移 (不抑制以下也将称为使能)。

[0337] 即,在这种情况下,由于当前状态是状态 S_{21} 并且紧邻在前的状态是状态 S_{28} ,因此抑制因子被更新以便抑制在紧邻在前的状态 S_{28} 和当前状态 S_{21} 以外的状态之间的状态转移,即例如在时刻 $t = 1$ 处获得的行为计划 PL1 中的第一状态 S_{28} 和下一个状态 S_{23} 之间的状态转移。

[0338] 此外,抑制因子被更新以便使能紧邻在前的状态 S_{28} 和当前状态 S_{21} 之间的状态转移。

[0339] 然后,在时刻 $t = 2$ 处,行为确定部分 24 将状态 S_{21} 设置为当前状态,将状态 S_{30} 设置为目标状态,获得用于从当前状态到达目标状态的最大可能状态系列 S_{21} 、 S_{28} 、 S_{27} 、 S_{26} 、 S_{25} 、 S_{20} 、 S_{15} 、 S_{10} 、 S_1 、 S_{17} 、 S_{16} 、 S_{22} 、 S_{29} 和 S_{30} ,并且计算当提供最大可能状态系列的状态转移发生时执行的的行为系列作为行为计划。

[0340] 此外,行为确定部分 24 将从行为计划的第一状态 S_{21} 向下一个状态 S_{28} 的移动的行为设置为被确定行为。代理执行被确定行为。

[0341] 结果,代理从对应于作为当前状态的状态 S_{21} 的观察单元在向左方向上向对应于状态 S_{28} 的观察单元移动 (执行图 3A 中的行为 U_4),并且从时刻 $t = 2$ 经过一个时刻而导致时刻 t 变为 $t = 3$ 。

[0342] 在时刻 $t = 3$ 处,状态识别部分 23 识别出当前状态是状态 S_{28} 。

[0343] 然后,在时刻 $t = 3$ 处,行为确定部分 24 将状态 S_{28} 设置为当前状态,将状态 S_{30} 设置为目标状态,获得从当前状态到达目标状态的最大可能状态系列,并且计算当提供最大可能状态系列的状态转移发生时执行的行为的行为系列作为行为计划。

[0344] 图 10C 示出了由行为确定部分 24 在时刻 $t = 3$ 处计算的行为计划 PL3。

[0345] 在图 10C 中,获得一系列状态 S_{28} 、 S_{27} 、 S_{26} 、 S_{25} 、 S_{20} 、 S_{15} 、 S_{10} 、 S_1 、 S_{17} 、 S_{16} 、 S_{22} 、 S_{29} 和 S_{30} 为最大可能状态系列,并且计算当提供最大可能状态系列的状态转移发生时执行的行为的行为系列为行为计划 PL3。

[0346] 即,在时刻 $t = 3$ 处,即使当前状态是与时刻 $t = 1$ 处相同的状态 S_{28} ,并且目标状态是与时刻 $t = 1$ 处相同的状态 S_{30} ,也计算得到与时刻 $t = 1$ 处的行为计划 PL1 不同的行为计划 PL3。

[0347] 这是因为如上所述在时刻 $t = 2$ 处抑制因子被更新以便抑制状态 S_{28} 和状态 S_{23} 之间的状态转移,并且在时刻 $t = 3$ 处获得最大可能状态系列中,抑制选择状态 S_{23} 作为从作为当前状态的状态 S_{28} 的状态转移的转移目的地,由此选择作为可以从状态 S_{28} 向其进行状态转移的状态的状态 S_{27} ,而不是状态 S_{23} 。

[0348] 在计算行为计划 PL3 后,行为确定部分 24 将从行为计划 PL3 的第一状态 S_{28} 向下一个状态 S_{27} 移动的行为设置为被确定行为。代理执行被确定行为。

[0349] 结果,代理从对应于作为当前状态的状态 S_{28} 的观察单元在向下方向上向对应于状态 S_{27} 的观察单元移动(执行图 3A 中的行为 U_3)。其后,在每个时刻类似地计算行为计划。

[0350] [使用抑制因子修正状态转移概率]

[0351] 图 11 是帮助说明使用抑制因子来修正扩展 HMM 的状态转移概率的图,由行为确定部分 24 在图 8 中的步骤 S37 中进行所述修正。

[0352] 如图 11 中所示,行为确定部分 24 通过将扩展 HMM 的状态转移概率 A_{1tm} 乘以抑制因子 $A_{inhibit}$ 来修正扩展 HMM 的状态转移概率 A_{1tm} ,由此获得修正转移概率 A_{stm} ,其是修正后的状态转移概率 A_{1tm} 。

[0353] 然后行为确定部分 24 使用修正转移概率 A_{stm} 作为扩展 HMM 的状态转移概率而计算行为计划。

[0354] 在这种情况下,在计算行为计划时,因为下面的原因,用于计算的状态转移概率由抑制因子修正。

[0355] 学习后的扩展 HMM 的状态可以包括分支结构的状态,所述状态允许当执行一个行为时向不同状态的状态转移。

[0356] 例如,在如上所述的图 10A 中的状态 S_{29} 的情况下,当执行用于在向左方向上移动的行为 U_4 (图 3A) 时,可以进行向左侧的状态 S_3 的状态转移,或者可以进行向在同一左侧的状态 S_{30} 的状态转移。

[0357] 因此,当执行一个特定行为时可以发生从状态 S_{29} 的不同状态转移,并且状态 S_{29} 是分支结构的状态。

[0358] 当关于一个特定行为可以发生不同的状态转移时,即,例如当在执行一个特定行为时可以发生向某一状态的状态转移或者也可以发生向其他状态的状态转移时,抑制因子抑制除了可以发生的不同状态转移的一个状态转移之外的状态转移的发生,以使得仅仅发

生这一个状态转移。

[0359] 即,假定关于一个特定行为可以发生的不同状态转移被称为分支结构,当使用从结构上改变的行为环境获得的观察值系列和行为系列作为学习数据而学习扩展 HMM 时,扩展 HMM 获得行为环境的结构上的改变作为分支结构。结果,分支结构的状态发生。

[0360] 因此,由于分支结构的态的发生,因此即使当行为环境的结构改变为各种结构时,扩展 HMM 也获得行为环境的所有各种结构。

[0361] 通过扩展 HMM 获得的、在结构上改变的行为环境的各种结构是要长时间存储而不被忘记的信息。因此,已经获得这样的信息(具体地为扩展 HMM 的状态转移概率)的扩展 HMM 也将被称为长期记忆。

[0362] 当当前状态是分支结构的状态时,在结构上改变的行为环境的当前结构确定作为分支结构的不同状态转移的哪一个有可能作为从当前状态的状态转移。

[0363] 即,即使当从作为长期记忆的扩展 HMM 的状态转移概率看可能进行状态转移时,取决于在结构上改变的行为环境的当前结构状态转移也可能无法进行。

[0364] 因此,代理根据通过独立于长期记忆而识别代理的当前状况获得的当前状态来更新抑制因子。然后,代理通过使用抑制因子修正作为长期记忆的扩展 HMM 的状态转移概率来获得修正转移概率作为修正后的状态转移概率,修正转移概率抑制在行为环境的当前结构中不能进行的状态转移并且使能在行为环境的当前结构中可以进行的状态转移。代理使用修正转移概率来计算行为计划。

[0365] 修正转移概率是通过使用根据每个时刻的当前状态更新的抑制因子修正作为长期记忆的状态转移概率而在每个时刻获得的信息,并且是满足短时间存储的信息。因此修正转移概率也将被称为短期记忆。

[0366] 行为确定部分 24(图 4)执行通过使用抑制因子修正扩展 HMM 的状态转移概率而获得修正转移概率的处理如下。

[0367] 当通过如图 6B 中所示的三维表表达扩展 HMM 的所有状态转移概率 $A_{l_{tm}}$ 时,抑制因子 $A_{inhibit}$ 也通过具有与扩展 HMM 的状态转移概率 $A_{l_{tm}}$ 的三维表相同尺寸的三维表来表达。

[0368] 用于表达扩展 HMM 的状态转移概率 $A_{l_{tm}}$ 的三维表也将被称为状态转移概率表。用于表达抑制因子 $A_{inhibit}$ 的三维表也将被称为抑制因子表。

[0369] 当扩展 HMM 的状态的数量是 N ,并且可以由代理执行的行为的数量是 M 时,状态转移概率表是 $N \times N \times M$ 个元素、即 N 个元素宽、 N 个元素高和 M 个元素深的三维表。因此,在这种情况下,抑制因子表也是 $N \times N \times M$ 个元素的三维表。

[0370] 顺便提及,除了抑制因子 $A_{inhibit}$ 之外,也可以通过 $N \times N \times M$ 个元素的三维表来表达修正转移概率 A_{stm} 。用于表达修正转移概率 A_{stm} 的三维表也将被称为修正转移概率表。

[0371] 例如,假定状态转移概率表中在从顶部起的第 i 行、从左起的第 j 列和在从前侧起的深度方向上的第 m 个平面的位置表达为 (i, j, m) ,则行为确定部分 24 通过下述方式获得修正转移概率 A_{stm} 作为修正转移概率表的位置 (i, j, m) 中的元素:通过根据式 (15) 将作为状态转移概率表的位置 (i, j, m) 中的元素的状态转移概率 $A_{l_{tm}} (= a_{ij}(U_m))$ 乘以作为抑制因子表的位置 (i, j, m) 中的元素的抑制因子 $A_{inhibit}$ 。

[0372] $A_{stm} = A_{l_{tm}} \times A_{inhibit} \cdots (15)$

[0373] 顺便提及,抑制因子在每个时刻在代理的状态识别部分 23(图 4)中更新如下。

[0374] 状态识别部分 23 更新抑制因子,以便关于由代理在从紧邻在前的状态 S_i 向当前状态 S_j 的状态转移时执行的行为 U_m ,抑制在当前状态 S_j 紧邻在前的状态 S_i 和当前状态 S_j 以外的状态之间的状态转移,并且不抑制(使能)在紧邻在前的状态 S_i 和当前状态 S_j 之间的状态转移。

[0375] 具体地,假定通过在行为轴的位置 m 处使用垂直于行为轴的平面切割抑制因子表而获得的平面被称为关于行为 U_m 的抑制因子平面,状态识别部分 23 用 1.0 来覆盖 $N \times N$ 个抑制因子即关于行为 U_m 的抑制因子平面的 N 个水平抑制因子和 N 个垂直抑制因子中的、作为从顶部起的第 i 行和从左起的第 j 列中的位置 (i, j) 中的元素的抑制因子,并且使用 0.0 来覆盖从顶部起的第 i 行中的 N 个抑制因子中的、作为除了位置 (i, j) 之外的位置中的元素的抑制因子。

[0376] 结果,根据通过使用抑制因子来修正状态转移概率而获得的修正转移概率,在来自分支结构的的状态的状态转移(分支结构)中,仅仅可以执行最近的经历、即已经进行的最近的状态转移,而不能进行其他状态转移。

[0377] 扩展 HMM 表达代理到当前为止所经历(通过学习获得)的行为环境的结构。此外,当行为环境的结构改变为各种结构时,扩展 HMM 将行为环境的各种结构表达为分支结构。

[0378] 另一方面,抑制因子表达作为由作为长期记忆的扩展 HMM 拥有的分支结构的多个状态转移的哪一个建模行为环境的当前结构。

[0379] 因此,通过将作为长期记忆的扩展 HMM 的状态转移概率乘以抑制因子而修正状态转移概率,并且使用修正转移概率作为修正后的转移概率(短期记忆)来计算行为计划,即使当行为环境的结构改变时,也可以考虑到改变后的结构(当前结构)而获得行为计划,而扩展 HMM 不用重新学习改变后的结构。

[0380] 即,当在结构上改变的行为环境的结构是已经通过扩展 HMM 获得的结构时,通过根据当前状态来更新抑制因子,并且使用更新后的抑制因子来修正扩展 HMM 的状态转移概率,可以考虑到行为环境的改变后的结构来获得行为计划,而不用重新学习扩展 HMM。

[0381] 即,可以迅速而高效地获得适应于行为环境的结构上的改变的行为计划,降低了计算成本。

[0382] 顺便提及,当行为环境被改变为未由扩展 HMM 获得的结构时,为了确定在改变后的结构的行为环境中的适当行为需要使用改变后的行为环境中观察到的观察值系列和行为系列来重新学习扩展 HMM。

[0383] 在行为确定部分 24 原样地使用扩展 HMM 的状态转移概率来计算行为计划的情况下,假定即使当行为环境的当前结构是其中只能进行作为分支结构的多个状态转移的一个而不能进行其他状态转移的结构时也可以根据维特比算法来进行作为分支结构的所有多个状态转移,当发生从当前状态 s_t 向目标状态 S_{goal} 的最大可能状态系列的状态转移时要执行的行为系列被计算作为行为计划。

[0384] 另一方面,在行为确定部分 24 通过抑制因子修正扩展 HMM 的状态转移概率,并且使用修正转移概率作为修正后的状态转移概率来计算行为计划的情况下,假定不能进行由抑制因子抑制的状态转移,并且当发生从当前状态 s_t 向目标状态 S_{goal} 的最大可能状态系列的状态转移而没有抑制因子抑制的状态转移时要执行的行为系列可以被计算作为行为计划。

[0385] 具体地,例如,在如上所述的图 10A 中,状态 S_{28} 是分支结构的状态,当执行用于在向右方向上移动的行为 U_2 时,从该状态可以向两个状态 S_{21} 和状态 S_{23} 进行状态转移。

[0386] 在图 10A 中,如上所述,在时刻 $t = 2$ 处,状态识别部分 23 更新抑制因子,以便关于由代理在从当前状态 S_{21} 紧邻在前的状态 S_{28} 向当前状态 S_{21} 的状态转移时执行的用于在向右方向上移动的行为 U_2 ,抑制从紧邻在前的状态 S_{28} 向当前状态 S_{21} 以外的状态 S_{23} 的状态转移,并且使能从紧邻在前的状态 S_{28} 向当前状态 S_{21} 的状态转移。

[0387] 结果,在时刻 $t = 3$ 处,在图 10C 中,即使当前状态是状态 S_{28} 并且目标状态是状态 S_{30} ,并且当前状态和目标状态都与图 10B 中在时刻 $t = 1$ 时相同,由于抑制因子抑制了当执行在向右方向上的移动的行为 U_2 时从状态 S_{28} 向状态 S_{21} 以外的状态 S_{23} 的状态转移,因此与时刻 $t = 1$ 处的不同的状态系列、即其中不进行从状态 S_{28} 向状态 S_{23} 的状态转移的状态系列 S_{28} 、 S_{27} 、 S_{26} 、 S_{25} 、 $\dots$ 、 S_{30} 被获得作为用于从当前状态到达目标状态的最大可能状态系列,并且当发生提供该状态系列的状态转移时要执行的行为系列被计算作为行为计划 PL3。

[0388] 抑制因子被更新,以便使能从作为分支结构的多个状态转移中代理经历过的状态转移,并且抑制除了代理经历过的状态转移之外的状态转移。

[0389] 即,抑制因子被更新,以便关于由代理在从紧邻在前的状态向当前状态的状态转移时执行的行为,抑制在当前状态紧邻在前的状态和当前状态以外的状态之间的状态转移(从紧邻在前的状态向当前状态以外的状态的状态转移),并且使能在紧邻在前的状态和当前状态之间的状态转移(从紧邻在前的状态向当前状态的状态转移)。

[0390] 当作为抑制因子的更新执行了仅使能在作为分支结构的多个状态转移中的代理经历过的状态转移和抑制除了代理经历过的状态转移之外的状态转移时,通过更新抑制因子而抑制的状态转移保持被抑制,除非代理其后经历这些状态转移。

[0391] 当根据使用修正转移概率计算的行为计划确定接下来要由代理执行的行为时,不计算包括引起由抑制因子抑制的状态转移的行为的任何行为计划,其中所述修正转移概率是通过如上所述在行为确定部分 24 中由抑制因子修正扩展 HMM 的状态转移概率而获得的。因此,由抑制因子抑制的状态转移保持被抑制,除非通过根据行为计划确定接下来要执行的行为的方法以外的方法来确定接下来要执行的行为,或者除非代理偶然经历了由抑制因子抑制的状态转移。

[0392] 因此,即使当行为环境的结构从其中不能进行被抑制因子抑制的状态转移的结构改变为其中可以进行所述状态转移的结构时,也不能计算包括引起所述状态转移的行为的行为计划,直到代理幸运地经历了由抑制因子抑制的状态转移为止。

[0393] 因此,状态识别部分 23 不仅随着抑制因子的更新使能在作为分支结构的多个状态转移中代理经历过的状态转移并抑制代理经历过的状态转移之外的状态转移,而且根据时间的经过而放开状态转移的抑制。

[0394] 即,状态识别部分 23 更新抑制因子,以便使能在作为分支结构的多个状态转移中代理经历过的状态转移,以及抑制代理经历过的状态转移以外的状态转移,并且还更新抑制因子以便根据时间的经过来放开状态转移的抑制。

[0395] 具体地,状态识别部分 23 例如根据式 (16) 将时刻 t 处的抑制因子 $A_{\text{inhibit}}(t)$ 更新为时刻 $t+1$ 处的抑制因子 $A_{\text{inhibit}}(t+1)$,以使得抑制因子根据时间的经过而收敛到 1.0。

[0396] $A_{\text{inhibit}}(t+1) = A_{\text{inhibit}}(t) + c(1 - A_{\text{inhibit}}(t))$ ($0 \leq c \leq 1$) $\dots$ (16)

[0397] 在式 (16) 中, 系数 c 高于 0.0 而低于 1.0。系数 c 越高, 则抑制因子越快收敛到 1.0。

[0398] 根据式 (16), 随着时间的经过放开曾经抑制的状态转移 (其抑制因子被设置为 0.0 的状态转移) 的抑制, 以使得即使当代理未经历该状态转移时也计算包括引起该状态转移的行为的行为计划。

[0399] 以下也将为了根据时间的经过而放开状态转移的抑制执行的抑制因子更新称为对应于由于自然衰减引起的遗忘的更新。

[0400] [抑制因子更新]

[0401] 图 12 是帮助说明更新抑制因子的处理的流程图, 所述处理由图 4 中的状态识别部分 23 在图 8 的步骤 S35 中执行。

[0402] 顺便提及, 当在图 8 中在识别行为模式中的处理的步骤 S31 中时刻 t 被初始化为 1 时, 将抑制因子初始化为 1.0 作为初始值。

[0403] 在更新抑制因子的处理的步骤 S71 中, 状态识别部分 23 更新模型存储部分 22 中存储的所有抑制因子 A_{inhibit} 作为对应于由于自然衰减引起的遗忘的更新, 即根据式 (16) 的更新。然后处理进行到步骤 S72。

[0404] 在步骤 S72 中, 状态识别部分 23 根据模型存储部分 22 中存储的扩展 HMM (扩展 HMM 的状态转移概率) 确定当前状态 S_j 紧邻在前的状态 S_i 是否是分支结构的状态、并且当前状态 S_j 是否是可以通过执行同一行为从作为紧邻在前的状态 S_i 的分支结构的状态向其进行状态转移的不同状态之一。

[0405] 在这种情况下, 可以像在分支结构检测部分 36 (图 4) 检测分支结构的状态的情况中那样确定紧邻在前的状态 S_i 是否是分支结构的状态。

[0406] 当在步骤 S72 中确定紧邻在前的状态 S_i 不是分支结构的状态或者确定紧邻在前的状态 S_i 是分支结构的状态、但是当前状态 S_j 不是可以通过执行同一行为从作为紧邻在前的状态 S_i 的分支结构的状态向其进行状态转移的不同状态之一时, 处理跳过步骤 S73 和 S74, 并且返回。

[0407] 当在步骤 S72 中确定紧邻在前的状态 S_i 是分支结构的状态并且当前状态 S_j 是可以通过执行同一行为从作为紧邻在前的状态 S_i 的分支结构的状态向其进行状态转移的不同状态之一时, 处理进行到步骤 S73, 在步骤 S73 中, 状态识别部分 23 将模型存储部分 22 中存储的抑制因子 A_{inhibit} 中关于紧邻在前的行为 U_m 的、从紧邻在前的状态 S_i 向当前状态 S_j 的状态转移的抑制因子 $h_{ij}(U_m)$ (在抑制因子表的位置 (i, j, m) 中的抑制因子) 更新为 1.0。然后处理进行到步骤 S74。

[0408] 在步骤 S74 中, 状态识别部分 23 将模型存储部分 22 中存储的抑制因子 A_{inhibit} 中关于紧邻在前的行为 U_m 的、从紧邻在前的状态 S_i 向当前状态 S_j 以外的状态 $S_{j'}$ 的状态转移的抑制因子 $h_{ij'}(U_m)$ (在抑制因子表的位置 (i, j', m) 中的抑制因子) 更新为 0.0。然后处理返回。

[0409] 在现有的行为确定方法中, 采用静态结构的建模来学习诸如 HMM 等的状态转移概率模型。因此, 当在学习了状态转移概率模型后学习对象的结构发生改变时, 需要将改变后的结构作为对象重新学习状态转移概率模型。因此, 需要高的计算成本来处理学习对象的结构上的改变。

[0410] 另一方面,在图 4 的代理中,扩展 HMM 获取行为环境结构上的改变作为分支结构。当紧邻在前的状态是分支结构的状态时,更新抑制因子,以便关于在从紧邻在前的状态向当前状态的状态转移时执行的行为抑制在紧邻在前的状态和当前状态以外的状态之间的状态转移。通过使用更新后的抑制因子来修正扩展 HMM 的状态转移概率。根据作为修正后的状态转移概率的修正转移概率来计算行为计划。

[0411] 因此,当行为环境的结构改变时,可以以低计算成本来计算适应于(按照)改变的结构的行为计划(而不用重新学习扩展 HMM)。

[0412] 另外,更新抑制因子以便根据时间的经过而放开状态转移的抑制。因此,即使当代理未偶然经历在过去抑制的状态转移时,随着时间的经过也可以计算包括引起在过去抑制的状态转移的行为的行为计划。结果,当行为环境的结构被改变为与当在过去抑制状态转移时的结构不同的结构时,可以迅速地计算适合于改变后的结构的行为计划。

[0413] [开放端的检测]

[0414] 图 13 是帮助说明由图 4 的开放端检测部分 37 检测为开放端的扩展 HMM 的状态的图。

[0415] 开放端广义而言是作为转移源的状态,在扩展 HMM 中预先知道可能会发生代理未经历过的从其的状态转移。

[0416] 具体地,未被分配状态转移概率(其状态转移概率是 0.0(视为 0.0 的值))的、并且因为下述原因而不能从其进行状态转移的状态对应于开放端,所述原因:虽然在所述状态的状态转移概率和被分配了与在观察的状态中相同的观察值的观察概率(所述观察概率不是值 0.0(不视为 0.0 的值))的其他状态的状态转移概率之间的比较显示当执行某一行为时可以进行向下一个状态的状态转移,但是在该状态中还没有执行该行为。

[0417] 因此,当在扩展 HMM 中检测到在可以以其中观察到预定观察值的状态作为转移源的状态转移中具有还没有执行的状态转移的其他状态,并且在所述其他状态中观察到与预定观察值相同的观察值时,所述其他状态是开放端。

[0418] 如图 13 中所示,开放端在概念上例如是对应于通过将代理放置在房间中并且以房间中的某一区域作为对象进行学习由扩展 HMM 获得的结构的端部(房间中所学习区域的端部)的状态,或者,开放端在概念上例如是对应于到新房间的入口的状态,所述入口作为以代理所位于的房间的全部区域为对象来执行学习、并且其后将代理可以移向的新房间被加到代理所位于的房间的邻近的结果而出现。

[0419] 所检测的开放端表示代理未知的区域扩展到由扩展 HMM 获得的结构的部分之外。因此,通过以开放端作为目标状态来计算行为计划,代理积极地执行步入未知区域中的行为。结果,代理可以更广泛地学习行为环境的结构(获得用作用于学习行为环境的学习数据的观察值系列和行为系列),并且高效地获得补充其结构未在扩展 HMM 中获得(行为环境中接近对应于作为开放端的状态的观察单元的结构)的模糊部分所需要的经历。

[0420] 开放端检测部分 37 首先生成行为模板,以便检测开放端。

[0421] 在生成行为模板时,开放端检测部分 37 将扩展 HMM 的观察概率 $B = \{b_i(O_k)\}$ 进行阈值处理,并且建立状态 S_i 的每个观察值 O_k 的列表,其中,以等于或者高于阈值的概率观察到观察值 O_k 。

[0422] 图 14A 和 14B 是帮助说明列出状态 S_i 的处理的图,其中,以等于或者高于阈值的

概率观察到观察值 O_k 。

[0423] 图 14A 示出了扩展 HMM 的观察概率 B 的示例。

[0424] 具体地,图 14A 示出了扩展 HMM 的观察概率 B 的示例,其中状态 S_i 的数量 N 是 5,并且观察值 O_k 的数量 M 是 3。

[0425] 开放端检测部分 37 将阈值设置在例如 0.5,并且执行检测等于或者高于阈值的观察概率 B 的阈值处理。

[0426] 在这种情况下,在图 14A 中,阈值处理检测关于状态 S_1 观察到观察值 O_3 的观察概率 $b_1(O_3) = 0.7$ 、关于状态 S_2 观察到观察值 O_2 的观察概率 $b_2(O_2) = 0.8$ 、关于状态 S_3 观察到观察值 O_3 的观察概率 $b_3(O_3) = 0.8$ 、关于状态 S_4 观察到观察值 O_2 的观察概率 $b_4(O_2) = 0.7$ 和关于状态 S_5 观察到观察值 O_1 的观察概率 $b_5(O_1) = 0.9$ 。

[0427] 其后,对于每个观察值 O_1 、 O_2 和 O_3 ,开放端检测部分 37 列出并且检测状态 S_i ,其中以等于或者高于阈值的概率观察到观察值 O_k 。

[0428] 图 14B 示出了针对每个观察值 O_1 、 O_2 和 O_3 列出的状态 S_i 。

[0429] 对于观察值 O_1 ,将状态 S_5 列出为其中以等于或者高于阈值的概率观察到观察值 O_1 的状态。对于观察值 O_2 ,将状态 S_2 和 S_4 列出为其中以等于或者高于阈值的概率观察到观察值 O_2 的状态。对于观察值 O_3 ,将状态 S_1 和 S_3 列出为其中以等于或者高于阈值的概率观察到观察值 O_3 的状态。

[0430] 其后,开放端检测部分 37 使用扩展 HMM 的状态转移概率 $A = \{a_{ij}(U_m)\}$ 来计算关于每个观察值 O_k 的每个行为 U_m 的转移概率对应值,所述转移概率对应值是对应于在针对观察值 O_k 列出的从状态 S_i 的状态转移中其状态转移概率 $a_{ij}(U_m)$ 是最大值的状态转移的状态转移概率 $a_{ij}(U_m)$ 的值;并且,开放端检测部分 37 将针对关于每个观察值 O_k 的每个行为 U_m 计算的转移概率对应值设置为当观察到观察值 O_k 时执行的行为 U_m 的行为概率,并且生成行为模板 C,行为模板 C 是以行为概率为元素的矩阵。

[0431] 图 15 是帮助说明使用针对观察值 O_k 列出的状态 S_i 来生成行为模板的方法的图。

[0432] 开放端检测部分 37 在三维状态转移概率表中从针对观察值 O_k 列出的状态 S_i 的状态转移的状态转移概率中检测最大状态转移概率,所述状态转移概率布置在列(横向)方向(j 轴方向)上。

[0433] 具体地,例如,现在将关注观察值 O_2 ,并且假定针对观察值 O_2 列出状态 S_2 和 S_4 。

[0434] 在这种情况下,开放端检测部分 37 将注意力导向关于状态 S_2 的行为平面,关于状态 S_2 的行为平面通过在 i 轴上的位置 $i = 2$ 处用垂直于 i 轴的平面来切割三维状态转移概率表而获得;并且,开放端检测部分 37 在关于状态 S_2 的行为平面内检测在执行行为 U_1 时发生的、从状态 S_2 的状态转移的状态转移概率 $a_{2,j}(U_1)$ 的最大值。

[0435] 即,开放端检测部分 37 在关于状态 S_2 的行为平面中的行为轴上的位置 $m = 1$ 处检测在 j 轴方向上布置的状态转移概率 $a_{2,1}(U_1)$ 、 $a_{2,2}(U_1)$ 、 $\cdots$ 、 $a_{2,N}(U_1)$ 的最大值。

[0436] 开放端检测部分 37 类似地从关于状态 S_2 的行为平面中检测在执行其他行为 U_m 时发生的、从状态 S_2 的状态转移的状态转移概率的最大值。

[0437] 此外,关于作为针对观察值 O_2 列出的另一个状态的状态 S_4 ,开放端检测部分 37 类似地从关于状态 S_4 的行为平面中检测在执行每个行为 U_m 时发生的、从状态 S_4 的状态转移的状态转移概率的最大值。

[0438] 如上所述,开放端检测部分 37 检测当关于针对观察值 O_2 列出的状态 S_2 和 S_4 的每一个执行每个行为 U_m 时发生的状态转移的状态转移概率的最大值。

[0439] 其后,开放端检测部分 37 平均如上所述关于针对每个行为 U_m 的观察值 O_2 列出的状态 S_2 和 S_4 检测的状态转移概率的最大值,并且将通过平均获得的平均值设置为对应于关于观察值 O_2 的状态转移概率的最大值的转移概率对应值。

[0440] 针对每个行为 U_m 获得关于观察值 O_2 的转移概率对应值。关于观察值 O_2 获得的每个行为 U_m 的转移概率对应值表示当观察到观察值 O_2 时执行行为 U_m 的概率(行为概率)。

[0441] 开放端检测部分 37 类似地获得转移概率对应值作为关于其他观察值 O_k 的每个行为 U_m 的行为概率。

[0442] 然后开放端检测部分 37 生成作为行为模板 C 的矩阵,其在从上数的第 k 行和从左数的第 m 列中以当观察到观察值 O_k 时执行行为 U_m 的行为概率作为元素。

[0443] 因此行为模板 C 是 K 行 M 列的矩阵,在该矩阵中,行数等于观察值 O_k 的数量,列数等于行为 U_m 的数量。

[0444] 在生成行为模板 C 后,开放端检测部分 37 使用行为模板 C 计算基于观察概率的行为概率 D。

[0445] 图 16 是帮助说明计算基于观察概率的行为概率 D 的方法的图。

[0446] 假定在第 i 行第 k 列中以在状态 S_i 中观察到观察值 O_k 的观察概率 $b_i(O_k)$ 作为元素的矩阵被称为观察概率矩阵 B,则观察概率矩阵 B 是 N 行 K 列的矩阵,在该矩阵中,行数等于状态 S_i 的数量 N,列数等于观察值 O_k 的数量 K。

[0447] 开放端检测部分 37 通过根据式 (17) 将 N 行 K 列的观察概率矩阵 B 乘以是 K 行 M 列的矩阵的行为模板 C,来计算基于观察概率的行为概率 D,所述行为概率 D 是在第 i 行第 m 列中以在其中观察到观察值 O_k 的状态 S_i 中执行行为 U_m 的概率为元素的矩阵。

[0448] $D = BC \quad \dots (17)$

[0449] 除了如上所述计算基于观察概率的行为概率 D 之外,开放端检测部分 37 还计算基于状态转移概率的行为概率 E。

[0450] 图 17 是帮助说明计算基于状态转移概率的行为概率 E 的方法的图。

[0451] 开放端检测部分 37 通过将在由 i 轴、j 轴和行为轴构成的三维状态转移概率表 A 中关于在 i 轴方向上的每个状态 S_i 的每个行为 U_m 的状态转移概率 $a_{ij}(U_m)$ 相加在一起,来计算基于状态转移概率的行为概率 E,所述行为概率 E 是在第 i 行第 m 列中以在状态 S_i 中执行行为 U_m 的概率作为元素的矩阵。

[0452] 即,开放端检测部分 37 通过下述方式计算基于状态转移概率的行为概率 E,所述行为概率 E 是 N 行 M 列的矩阵:当注意力转向 i 轴上的位置 i 和行为轴上的位置 m 时获得在由 i 轴、j 轴和行为轴构成的三维状态转移概率表 A 中在水平方向(列方向)上布置的状态转移概率 $a_{ij}(U_m)$ 的总和,即在与 j 轴平行的、通过点 (i, m) 的直线中布置的状态转移概率 $a_{ij}(U_m)$ 的总和,并且将所述总和设置为矩阵第 i 行第 m 列中的元素。

[0453] 在如上所述计算基于观察概率的行为概率 D 和基于状态转移概率的行为概率 E 后,开放端检测部分 37 根据式 (18) 计算差分行为概率 F,其是基于观察概率的行为概率 D 和基于状态转移概率的行为概率 E 之间的差。

[0454] $F = D - E \quad \dots (18)$

[0455] 像基于观察概率的行为概率 D 和基于状态转移概率的行为概率 E 一样,差分行为概率 F 是 N 行 M 列的矩阵。

[0456] 图 18 是示意性地示出了差分行为概率 F 的图。

[0457] 图 18 中的小正方形代表矩阵的元素。没有图案的正方形表示其值是 0.0 (视为 0.0 的值) 的元素。以黑色填充的正方形表示其值不是 0.0 (不是视为 0.0 的值) 的元素。

[0458] 根据差分行为概率 F, 当存在作为其中观察到观察值 O_k 的状态的多个状态, 并且已知可以从多个状态的一部分 (其中代理已经执行了行为 U_m 的状态) 执行行为 U_m 时, 可以检测到其中当执行行为 U_m 时发生的状态转移未反映在状态转移概率 $a_{ij}(U_m)$ 中的其他状态 (代理没有执行过行为 U_m 的状态), 即开放端。

[0459] 具体地, 当执行行为 U_m 时发生的状态转移反映在了状态 S_i 的状态转移概率 $a_{ij}(U_m)$ 中时, 在基于观察概率的行为概率 D 的第 i 行第 m 列中的元素与基于状态转移概率的行为概率 E 的第 i 行第 m 列中的元素取值相互近似。

[0460] 另一方面, 当执行行为 U_m 时发生的状态转移未反映在状态 S_i 的状态转移概率 $a_{ij}(U_m)$ 中时, 在基于观察概率的行为概率 D 的第 i 行第 m 列中的元素是由于其中观察到与状态 S_i 中相同的观察值并且已经执行了行为 U_m 的状态的状态转移概率的影响而不能视为 0.0 的某一值, 而在基于状态转移概率的行为概率 E 的第 i 行第 m 列中的元素是 0.0 (包括可以视为 0.0 的小值)。

[0461] 因此, 当在状态 S_i 的状态转移概率 $a_{ij}(U_m)$ 中未反映当执行行为 U_m 时发生的状态转移时, 在差分行为概率 F 的第 i 行第 m 列中的元素具有不能视为 0.0 的值 (绝对值)。因此, 可以通过检测差分行为概率 F 中具有不能视为 0.0 的值的元素来检测开放端和还没有在开放端中执行的行为。

[0462] 即, 当差分行为概率 F 的第 i 行第 m 列中的元素具有不能视为 0.0 的值时, 开放端检测部分 37 检测状态 S_i 为开放端, 并且检测行为 U_m 为还没有在作为开放端的状态 S_i 中执行的行为。

[0463] 图 19 是帮助说明检测开放端的处理的流程图, 所述处理在图 9 的步骤 S53 中由图 4 中的开放端检测部分 37 执行。

[0464] 在步骤 S81 中, 开放端检测部分 37 将模型存储部分 22 (图 4) 中存储的扩展 HMM 的观察概率 $B = \{b_i(O_k)\}$ 进行阈值处理, 并且由此建立状态 S_i 的每个观察值 O_k 的列表, 其中观察到观察值 O_k 的概率等于或者高于阈值, 如参照图 14A 和 14B 所述。

[0465] 在步骤 S81 后, 处理进行到步骤 S82, 在步骤 S82 中, 如参照图 15 所述, 开放端检测部分 37 使用模型存储部分 22 中存储的扩展 HMM 的状态转移概率 $A = \{a_{ij}(U_m)\}$ 计算关于每个观察值 O_k 的每个行为 U_m 的转移概率对应值, 所述转移概率对应值是对应于在从针对观察值 O_k 列出的从状态 S_i 的状态转移中其状态转移概率 $a_{ij}(U_m)$ 是最大值的状态转移的状态转移概率 $a_{ij}(U_m)$ 的值, 并且开放端检测部分 37 将关于每个观察值 O_k 针对每个行为 U_m 计算的转移概率对应值设置为当观察到观察值 O_k 时执行行为 U_m 的行为概率, 并且生成行为模板 C, 所述行为模板 C 是以行为概率作为元素的矩阵。

[0466] 其后, 处理从步骤 S82 进行到步骤 S83, 在步骤 S83 中, 开放端检测部分 37 根据式 (17) 通过将观察概率矩阵 B 乘以行为模板 C 而计算基于观察概率的行为概率 D。然后处理进行到步骤 S84。

[0467] 在步骤 S84 中,如参照图 17 所述,开放端检测部分 37 通过将状态转移概率表 A 中关于 i 轴方向上的每个状态 S_i 的每个行为 U_m 的状态转移概率 $a_{ij}(U_m)$ 相加在一起,来计算基于状态转移概率的行为概率 E ,所述行为概率 E 是在第 i 行第 m 列中以在状态 S_i 中执行行为 U_m 的概率作为元素的矩阵。

[0468] 然后处理从步骤 S84 进行到步骤 S85,在步骤 S85 中,开放端检测部分 37 根据式 (18) 计算差分行为概率 F ,其是基于观察概率的行为概率 D 和基于状态转移概率的行为概率 E 之间的差。然后处理进行到步骤 S86。

[0469] 在步骤 S86 中,开放端检测部分 37 将差分行为概率 F 进行阈值处理,由此检测差分行为概率 F 中其值等于或者高于阈值的元素为作为检测的对象的检测对象元素。

[0470] 此外,开放端检测部分 37 检测检测对象元素的第 i 行第 m 列,检测作为开放端的状态 S_i ,并且检测作为在开放端 S_i 中还没有执行的未经历的行为的行为 U_m 。然后处理返回。

[0471] 通过执行开放端中未经历的行为,代理可以探查在开放端之外延续的未知区域。

[0472] 现有的行为确定方法在确定代理的目标时,等同地(无区分地)对待已知区域(学习过的区域)和未知区域(还未学习的区域)而不考虑代理的经验。因此,需要执行多个行为来积累未知区域的经验。结果,需要许多尝试和许多时间来广泛地学习行为环境的结构。

[0473] 另一方面,图 4 的代理检测开放端,并且以开放端作为目标状态来确定行为。因此图 4 的代理可以高效地学习行为环境的结构。

[0474] 即,开放端是代理未经历的未知区域延伸到其外的状态。因此,通过检测开放端并且以开放端为目标状态来确定行为,代理可以积极地进入未知区域中。由此,代理可以高效地积累经验以更广泛地学习行为环境的结构。

[0475] [分支结构的状态的检测]

[0476] 图 20 是帮助说明一种由图 4 的分支结构检测部分 36 检测分支结构的状态的方法的图。

[0477] 扩展 HMM 获得行为环境中结构上改变的部分作为分支结构的状态。可以通过参考作为长期记忆的扩展 HMM 的状态转移概率来检测对应于代理已经经历的结构上的改变的分支结构的状态。当检测到分支结构的状态时,代理可以识别出行为环境中结构上发生改变的部分的存在。

[0478] 当在行为环境中存在结构上发生改变的部分时,期望定期或者不定期地主动查看这样的部分的当前结构,并且在抑制因子中反映当前结构,或者继而在作为短期记忆的修正转移概率中反映当前结构。

[0479] 因此,图 4 的代理可以在分支结构检测部分 36 中检测分支结构的状态,并且在目标选择部分 31 中选择分支结构的状态作为目标状态。

[0480] 如图 20 中所示,分支结构检测部分 36 检测分支结构的状态。

[0481] 状态转移概率表 A 中关于每个行为 U_m 的状态转移概率平面被归一化,以使得在每行的水平方向(列方向)上的总和是 1.0。

[0482] 因此,当注意力转向关于行为 U_m 的状态转移概率平面中的某一行 i ,并且状态 S_i 不是分支结构的状态时,第 i 行中的状态转移概率 $a_{ij}(U_m)$ 的最大值是 1.0 或者很接近 1.0 的值。

[0483] 另一方面,当状态 S_i 是分支结构的状态时,第 i 行中的状态转移概率 $a_{ij}(U_m)$ 的最大值远远小于 1.0,诸如图 20 中所示的 0.6 或者 0.5 等,并且大于通过将其总和是 1.0 的状态转移概率均等地除以状态的数量 N 而获得的值(平均值) $1/N$ 。

[0484] 因此,当关于每个行为 U_m 的状态转移概率平面中的每行 i 中的状态转移概率 $a_{ij}(U_m)$ 的最大值比小于 1.0 的阈值 $a_{\max_th}$ 小并且比平均值 $1/N$ 大时,分支结构检测部分 36 根据式 (19) 检测状态 S_i 为分支结构的状态。

$$[0485] \quad 1/N < \max_{j,i=S,m=U} (A_{ijm}) < a_{\max_th} \quad \dots(19)$$

[0486] 在式 (19) 中, A_{ijm} 表示三维状态转移概率表 A 中在 i 轴方向上从上起的第 i 位置、在 j 轴方向上从左起的第 j 位置和在行为轴方向上从前起的第 m 位置中的状态转移概率 $a_{ij}(U_m)$ 。

[0487] 另外,在式 (19) 中, $\max(A_{ijm})$ 表示状态转移概率表 A 中在 j 轴方向上从左起的第 S 位置(作为从状态 S_i 的状态转移的转移目的地的状态是状态 S)和在行为轴方向上从前起的第 U 位置(当从状态 S_i 的状态转移发生时执行的行为是行为 U)中的 N 个状态转移概率 $A_{1,S,U}$ 到 $A_{N,S,U}$ ($a_{1,S}(U)$ 到 $a_{N,S}(U)$) 的最大值。

[0488] 顺便提及,可以根据检测分支结构的状态的灵敏度在范围 $1/N < a_{\max_th} < 1.0$ 中调整式 (19) 中的阈值 $a_{\max_th}$ 。阈值 $a_{\max_th}$ 越接近 1.0,则可以越灵敏地检测分支结构的状态。

[0489] 当检测到一个或更多个分支结构的状态时,分支结构检测部分 36 将一个或更多个分支结构的状态提供给目标选择部分 31,如参照图 9 所述。

[0490] 查找过去时间管理表存储部分 32 中的过去时间管理表,目标选择部分 31 识别来自分支结构检测部分 36 的一个或更多个分支结构的状态的过去时间。

[0491] 然后,目标选择部分 31 来自分支结构检测部分 36 的一个或更多个分支结构的状态中检测其过去时间最长的状态,并且选择该状态作为目标状态。

[0492] 通过如上所述从一个或更多个分支结构的状态中检测其过去时间最长的状态并且选择该状态作为目标状态,可以暂时均等地将一个或更多个分支结构的状态中的每一个设置为目标状态,并且执行行为以检查对应于分支结构的状态的结构如何。

[0493] 在现有的行为确定方法中,确定目标时不将注意转向分支结构的状态,因此经常将不是分支结构的状态的状态设置为目标。因此,当要掌握行为环境的最新结构时,经常执行不必要的行为。

[0494] 另一方面,图 4 的代理以分支结构的状态设置为目标状态来确定行为。因此,可以较早地掌握对应于分支结构的状态的部分的最新结构,并且将其反映在抑制因子中。

[0495] 顺便提及,当将分支结构的状态设置为目标状态时,代理在到达了作为目标状态的分支结构的状态(对应于分支结构的状态的观察单元)后可以根据扩展 HMM 来识别通过其可以从分支结构的状态向不同的状态进行状态转移的行为,并且通过执行该行为来移动。代理由此可以识别(掌握)对应于分支结构的状态的部分的结构,该状态即为可以从当前分支结构的状态向其进行状态转移的状态。

[0496] [模拟]

[0497] 图 21A 和 21B 是示出了图 4 的代理的模拟中使用的行为环境的图,该模拟由本发明人执行。

[0498] 具体地,图 21A 示出了第一结构的行为环境,图 21B 示出了第二结构的行为环境。

[0499] 在第一结构的行为环境中,位置 pos1、pos2 和 pos3 是代理可以通过的通道,而在第二结构的行为环境中,位置 pos1 到 pos3 是代理不能通过的墙壁。

[0500] 顺便提及,可以使位置 pos1 到 pos3 的每一个单独地为通道或者墙壁。

[0501] 在模拟中,使代理在第一结构和第二结构的行为环境的每一个中在反射行为模式(图 5)中执行行为,获得用作 4000 步(次)的学习数据的观察值系列和行为系列,并且学习扩展 HMM。

[0502] 图 22 是示意性地示出了学习后的扩展 HMM 的图。

[0503] 在图 22 中,圆圈(o)标记表示扩展 HMM 的状态,并且圆圈标记中输入的数字是由圆圈标记代表的状态的下标。用于指示由圆圈标记代表的状态的箭头代表可能的状态转移(其状态转移概率不是 0.0(可以视为 0.0 的值)的状态转移)。

[0504] 在图 22 的扩展 HMM 中,状态 S_1 被布置在对应于状态 S_1 的观察单元的位置处。

[0505] 其间可能进行状态转移的两个状态表示代理可以在对应于该两个相应的状态的两个观察单元之间移动。因此,用于表示扩展 HMM 的状态转移的箭头表示代理可以在行为环境中移动通过的通道。

[0506] 在图 22 中,存在两个(多个)状态 S_1 和 $S_{1'}$ 被布置为在一个观察单元的位置处彼此重叠的情况。这表示该两个(多个)状态 S_1 和 $S_{1'}$ 对应于所述一个观察单元。

[0507] 在图 22 中,如同在图 10A 的情况中,状态 S_3 和 S_{30} 对应于一个观察单元,并且状态 S_{34} 和 S_{35} 也对应于一个观察单元。类似地,状态 S_{21} 和 S_{23} 对应于一个观察单元,状态 S_2 和 S_{17} 对应于一个观察单元,状态 S_{37} 和 S_{48} 对应于一个观察单元,以及状态 S_{31} 和 S_{32} 对应于一个观察单元。

[0508] 在图 22 中,状态 S_{29} 、状态 S_{39} 、状态 S_{28} 、状态 S_1 、状态 S_{16} 、状态 S_{12} 、状态 S_{42} 、状态 S_{36} 和状态 S_{25} 是分支结构的状态,其中,当执行在向左方向上的移动的行为 U_4 (图 3B) 时可以进行从状态 S_{29} 向不同状态 S_3 和 S_{30} 的状态转移,当执行在向右方向上的移动的行为 U_2 时可以进行从状态 S_{39} 向不同状态 S_{34} 和 S_{35} 的状态转移,当执行在向左方向上的移动的行为 U_4 时可以进行从状态 S_{28} 向不同状态 S_{34} 和 S_{35} 的状态转移(状态 S_{28} 也是下述状态:当执行在向右方向上的移动的行为 U_2 时可以进行从状态 S_{28} 向不同状态 S_{21} 和 S_{23} 的状态转移),当执行在向上方向上的移动的行为 U_1 时可以进行从状态 S_1 向不同状态 S_2 和 S_{17} 的状态转移,当执行在向下方向上的移动的行为 U_3 时可以进行从状态 S_{16} 向不同状态 S_2 和 S_{17} 的状态转移,当执行在向左方向上的移动的行为 U_4 时可以进行从状态 S_{12} 向不同状态 S_2 和 S_{17} 的状态转移,当执行在向下方向上的移动的行为 U_3 时可以进行从状态 S_{42} 向不同状态 S_{37} 和 S_{48} 的状态转移,当执行在向下方向上的移动的行为 U_3 时可以进行从状态 S_{36} 向不同状态 S_{31} 和 S_{32} 的状态转移,以及当执行在向左方向上的移动的行为 U_4 时可以进行从状态 S_{25} 向不同状态 S_{31} 和 S_{32} 的状态转移。

[0509] 顺便提及,图 22 中的虚线箭头代表仅在第二结构的行为环境中可能的状态转移。因此,当行为环境的结构是第一结构(图 21A) 时,不能进行由图 22 中虚线箭头代表的状态转移。

[0510] 在模拟中,进行初始设置以将对应于由图 22 中的虚线箭头代表的状态转移的抑制因子设置为 0.0,并且将对应于其他状态转移的抑制因子设置为 1.0。由此防止代理模拟开始后立即计算包括实施仅在第二结构的行为环境中可能的状态转移的行为的行为计划。

[0511] 图 23 至 29 是示出了代理的图,所述代理根据学习后的扩展 HMM 来计算到达目标状态的行为计划,并且执行根据行为计划确定的行为。

[0512] 顺便提及,图 23 至 29 在上侧示出了行为环境中的代理和目标状态(对应于目标状态的观察单元),并且在下侧示出了扩展 HMM。

[0513] 图 23 示出了时刻 $t = t_0$ 处的代理。

[0514] 在时刻 $t = t_0$ 处,行为环境的结构是第一结构(图 21A),其中位置 pos1 到 pos3 是通道。

[0515] 此外,在时刻 $t = t_0$ 处,目标状态(对应于目标状态的观察单元)是左下状态 S_{37} ,并且代理位于状态 S_{20} (对应于状态 S_{20} 的观察单元)中。

[0516] 然后,代理计算用于去往作为目标状态的状态 S_{37} 的行为计划,并且作为根据行为计划确定的行为而从作为当前状态的状态 S_{20} 在向左方向上移动。

[0517] 图 24 示出了在时刻 $t = t_1 (> t_0)$ 处的代理。

[0518] 在时刻 $t = t_1$ 处,行为环境的结构从第一结构改变到其中位置 pos1 是代理可以通过的通道而位置 pos2 和 pos3 是代理不能通过的墙壁的结构。

[0519] 此外,在时刻 $t = t_1$ 处,与时刻 $t = t_0$ 处相同,目标状态是左下状态 S_{37} 并且代理位于状态 S_{31} 中。

[0520] 图 25 示出了在时刻 $t = t_2 (> t_1)$ 处的代理。

[0521] 在时刻 $t = t_2$ 处,行为环境的结构是其中位置 pos1 是代理可以通过的通道而位置 pos2 和 pos3 是代理不能通过的墙壁的结构(下文中也该结构称为改变后的结构)。

[0522] 此外,在时刻 $t = t_2$ 处,目标状态是上侧的状态 S_3 ,并且代理位于状态 S_{31} 中。

[0523] 然后,代理计算用于去往作为目标状态的状态 S_3 的行为计划,并且将作为根据行为计划确定的行为而从作为当前状态的状态 S_{31} 在向上方向上移动。

[0524] 在这种情况下,在时刻 $t = t_2$ 处,计算用于进行状态系列 S_{31} 、 S_{36} 、 S_{39} 、 S_{35} 和 S_3 的状态转移的行为计划。

[0525] 顺便提及,当行为环境是第一结构时,对应于状态 S_{37} 和 S_{48} 的观察单元与对应于状态 S_{31} 和 S_{32} 的观察单元之间的位置 pos1(图 21A)、对应于状态 S_3 和 S_{30} 的观察单元与对应于状态 S_{34} 和 S_{35} 的观察单元之间的位置 pos2 以及对应于状态 S_{21} 和 S_{23} 的观察单元与对应于状态 S_2 和 S_{17} 的观察单元之间的位置 pos3 的每一个均是通道。因此代理可以通过位置 pos1 到 pos3。

[0526] 但是,当行为环境改变为改变后的结构时,位置 pos2 和 pos3 是墙壁,从而代理不能通过位置 pos2 和 pos3。

[0527] 如上所述,在模拟的初始设置中,只有对应于仅在第二结构的行为环境中可能的状态转移的抑制因子被设置为 0.0。在时刻 $t = t_2$ 处,不抑制在第一结构的行为环境中可能的状态转移。

[0528] 因此,在时刻 $t = t_2$ 处,虽然对应于状态 S_3 和 S_{30} 的观察单元与对应于状态 S_{34} 和 S_{35} 的观察单元之间的位置 pos2 是代理不能通过的墙壁,但是代理计算包括进行从状态 S_{35} 向状态 S_3 的状态转移以通过对应于状态 S_3 和 S_{30} 的观察单元与对应于状态 S_{34} 和 S_{35} 的观察单元之间的位置 pos2 的行为的行为计划。

[0529] 图 26 示出了在时刻 $t = t_3 (> t_2)$ 处的代理。

[0530] 在时刻 $t = t_3$ 处, 行为环境的结构保持为改变后的结构。

[0531] 此外, 在时刻 $t = t_3$ 处, 目标状态是上侧的状态 S_3 , 并且代理位于状态 S_{28} 中。

[0532] 然后, 代理计算用于去往作为目标状态的状态 S_3 的行为计划, 并且将作为根据行为计划确定的行为而从作为当前状态的状态 S_{28} 在向右方向上移动。

[0533] 在这种情况下, 在时刻 $t = t_3$ 处, 计算用于进行状态系列 S_{28} 、 S_{23} 、 S_2 、 S_{16} 、 S_{22} 、 S_{29} 和 S_3 的状态转移的行为计划。

[0534] 在时刻 $t = t_2$ 之后代理还计算出与时刻 $t = t_2$ 处计算的用于进行状态系列 S_{31} 、 S_{36} 、 S_{39} 、 S_{35} 和 S_3 的状态转移的行为计划 (图 25) 类似的行为计划, 并且通过执行根据行为计划确定的行为移动到对应于状态 S_{35} 的观察单元。此时, 代理认识到代理不能通过对应于状态 S_3 (和 S_{30}) 的观察单元与对应于状态 (S_{34} 和) S_{35} 的观察单元之间的位置 pos2, 即认识到通过执行根据行为计划确定的行为而从对应于行为计划的状态系列 S_{31} 、 S_{36} 、 S_{39} 、 S_{35} 和 S_3 的状态 S_{39} 能够到达的状态不是临近状态 S_{39} 的状态 S_{35} , 而是状态 S_{34} 。代理更新对应于不能进行的、从状态 S_{39} 向状态 S_{35} 的状态转移的抑制因子为 0.0。

[0535] 结果, 在时刻 $t = t_3$ 处, 代理计算用于进行状态系列 S_{28} 、 S_{23} 、 S_2 、 S_{16} 、 S_{22} 、 S_{29} 和 S_3 的状态转移的行为计划, 该行为计划不引起其中代理可以通过位置 pos2 的从状态 S_{39} 向状态 S_{35} 的状态转移。

[0536] 顺便提及, 当行为环境是改变后的结构时, 对应于状态 S_{21} 和 S_{23} 的观察单元与对应于状态 S_2 和 S_{17} 的观察单元之间的位置 pos3 (图 21) 是代理不能通过的墙壁。

[0537] 如上所述, 在模拟的初始设置中, 只有对应于仅在第二结构的行为环境中可能的状态转移的抑制因子被设置为 0.0, 其中位置 pos1 到 pos3 是代理不能通过的墙壁。在时刻 $t = t_3$ 处, 不抑制从状态 S_{23} 向状态 S_2 的状态转移, 所述状态转移在第一结构的行为环境中是可能的, 并且对应于通过位置 pos3。

[0538] 因此, 在时刻 $t = t_3$ 处, 代理计算行为计划, 所述行为计划进行从状态 S_{23} 向状态 S_2 的状态转移以通过对应于状态 S_{21} 和 S_{23} 的观察单元与对应于状态 S_2 和 S_{17} 的观察单元之间的位置 pos3。

[0539] 图 27 示出了在时刻 $t = t_4 (= t_3 + 1)$ 处的代理。

[0540] 在时刻 $t = t_4$ 处, 行为环境的结构是改变后的结构。

[0541] 此外, 在时刻 $t = t_4$ 处, 目标状态是上侧的状态 S_3 , 并且代理位于状态 S_{21} 中。

[0542] 代理通过执行根据在时刻 $t = t_3$ 处计算的用于进行状态系列 S_{28} 、 S_{23} 、 S_2 、 S_{16} 、 S_{22} 、 S_{29} 和 S_3 的状态转移的行为计划 (图 26) 确定的行为, 从对应于状态 S_{28} 的观察单元移动到对应于状态 S_{21} 和 S_{23} 的观察单元。此时, 代理认识到, 通过执行根据行为计划确定的行为而从对应于行为计划的状态系列 S_{28} 、 S_{23} 、 S_2 、 S_{16} 、 S_{22} 、 S_{29} 和 S_3 的状态 S_{28} 可以到达的状态不是临近状态 S_{28} 的状态 S_{23} , 而是状态 S_{21} 。代理将对应于从状态 S_{28} 向状态 S_{23} 的状态转移的抑制因子更新为 0.0。

[0543] 结果, 在时刻 $t = t_4$ 处, 代理计算不包括从状态 S_{28} 向状态 S_{23} 的状态转移 (并且继而不包括代理通过对应于状态 S_{21} 和 S_{23} 的观察单元与对应于状态 S_2 和 S_{17} 的观察单元之间的位置 pos3) 的行为计划。

[0544] 在这种情况下, 在时刻 $t = t_4$ 处, 计算用于进行状态系列 S_{28} 、 S_{27} 、 S_{26} 、 S_{25} 、 S_{20} 、 S_{15} 、 S_{10} 、 S_1 、 S_2 、 S_{16} 、 S_{22} 、 S_{29} 和 S_3 的状态转移的行为计划。

[0545] 图 28 示出了在时刻 $t = t_5 (= t_4 + 1)$ 处的代理。

[0546] 在时刻 $t = t_5$ 处, 行为环境的结构是改变后的结构。

[0547] 此外, 在时刻 $t = t_5$ 处, 目标状态是上侧的状态 S_3 , 并且代理位于状态 S_{28} 中。

[0548] 代理通过执行根据在时刻 $t = t_4$ 处计算的用于进行状态系列 S_{28} 、 S_{27} 、 S_{26} 、 S_{25} 、 S_{20} 、 S_{15} 、 S_{10} 、 S_1 、 S_2 、 S_{16} 、 S_{22} 、 S_{29} 和 S_3 的状态转移的行为计划 (图 27) 确定的行为, 从对应于状态 S_{21} 的观察单元移动到对应于状态 S_{28} 的观察单元。

[0549] 图 29 示出了在时刻 $t = t_6 (> t_5)$ 处的代理。

[0550] 在时刻 $t = t_6$ 处, 行为环境的结构是改变后的结构。

[0551] 此外, 在时刻 $t = t_6$ 处, 目标状态是上侧的状态 S_3 , 并且代理位于状态 S_{15} 中。

[0552] 然后, 代理计算用于去往作为目标状态的状态 S_3 的行为计划, 并且将作为根据行为计划确定的行为而从作为当前状态的状态 S_{15} 在向右方向上移动。

[0553] 在这种情况下, 在时刻 $t = t_6$ 处, 计算用于进行状态系列 S_{10} 、 S_1 、 S_2 、 S_{16} 、 S_{22} 、 S_{29} 和 S_3 的状态转移的行为计划。

[0554] 如上所述, 即使当行为环境的结构改变时, 代理也观察改变后的结构 (确定 (识别) 当前状态), 并且更新抑制因子。然后, 代理使用更新后的抑制因子重新计算行为计划, 并且可以最终到达目标状态。

[0555] [代理的应用的示例]

[0556] 图 30 是示出了应用图 4 的代理的清洁机器人的一般外观的图。

[0557] 图 30 中的清洁机器人 51 包括用作清洁器的模块、对应于图 4 的代理中的执行器 12 和传感器 13 的模块和用于执行无线电通信的模块。

[0558] 图 30 中的清洁机器人将起居室设置为行为环境, 执行移动作为行为, 并且在起居室中执行清洁。

[0559] 计算机主机 52 用作图 4 中的反射行为确定部分 11、历史存储部分 14、行为控制部分 15 和目标确定部分 16 (具有对应于反射行为确定部分 11、历史存储部分 14、行为控制部分 15 和目标确定部分 16 的模块)。

[0560] 计算机主机 52 安装在起居室或者其他房间中, 并且连接到用于控制通过无线 LAN (局域网) 等的无线电通信的接入点 53。

[0561] 计算机主机 52 通过经由接入点 53 与清洁机器人 51 执行无线电通信来与清洁机器人 51 交换必要数据。由此清洁机器人 51 移动来作为与图 4 的代理的行为类似的行为。

[0562] 顺便提及, 在图 30 中, 为了小型化清洁机器人 51 并且考虑到在清洁机器人 51 中结合足够的电源和足够的计算能力的困难, 清洁机器人 51 仅仅设置有对应于执行器 12 和传感器 13 的模块, 该模块是在形成图 4 的代理的模块中的最小必需模块, 与清洁机器人 51 分离的计算机主机 52 设置有其他模块。

[0563] 但是, 向清洁机器人 51 和计算机主机 52 的每一个提供的形成图 4 的代理的模块不限于上述模块。

[0564] 具体地, 例如, 清洁机器人 51 可以设置有不仅对应于执行器 12 和传感器 13 的模块, 而且可以设置有对应于不要求具有高级计算功能的反射行为确定部分 11 的模块, 并且计算机主机 52 可以设置有对应于需要高级计算功能和高存储容量的历史存储部分 14、行为控制部分 15 和目标确定部分 16 的模块。

[0565] 根据扩展 HMM,在其中在不同位置处的观察单元中观察到同一观察值的行为环境中,可以使用观察值系列和行为系列来识别代理的当前状况,并且可以唯一地识别当前状态或者继而唯一地识别代理所位于的观察单元(位置)。

[0566] 图 4 的代理根据当前状态更新抑制因子,并且继续计算行为计划同时通过更新后的抑制因子修正扩展 HMM 的状态转移概率。由此代理即使在其结构概率性地改变的行为环境中也可以到达目标状态。

[0567] 这样的代理例如适用于诸如在例如生活环境中工作的清洁机器人等的实际机器人,所述生活环境由于人的生活活动而在结构上动态地改变,并且人生活在该环境中。

[0568] 例如,可以通过开关房间的门或者改变房间中家具的布置等在结构上改变诸如房间等的生活环境。

[0569] 但是,因为房间本身的形状不改变,因此结构上改变的部分和结构上不改变的部分共存于生活环境中。

[0570] 根据扩展 HMM,结构上改变的部分可以被存储为分支结构的状态。因此,可以高效地(以低存储容量)表达包括结构上改变的部分的生活环境。

[0571] 另一方面,对于在生活环境中被用作由人操作的清洁器的替代装置以实现清洁整个房间的目的的清洁机器人,该清洁机器人必须识别清洁机器人自身的位置,并且在其结构概率性地改变的房间(其结构可以改变的房间)中自适应地改变路径的同时移动。

[0572] 图 4 的代理特别有益于在如此识别清洁机器人自身的位置并且在其结构概率性地改变的生活环境中自适应地改变路径的同时实现目的(清洁整个房间)。

[0573] 顺便提及,从降低清洁机器人的制造成本的视点看,期望避免在清洁机器人中结合作为高级传感器的摄像机和用于执行诸如识别由摄像机输出的图像等图像处理的图像处理装置作为用于观察观察值的装置。

[0574] 即,为了降低清洁机器人的制造成本,期望使用便宜的装置作为清洁机器人用来观察观察值的装置,诸如通过在多个方向上输出超声波或者激光等来测量距离的距离测量装置。

[0575] 但是,当采用诸如距离测量装置等便宜装置作为用于观察观察值的装置时,经常在生活环境的不同位置处观察到同一观察值,难以仅用单一时刻的观察值唯一地识别清洁机器人的位置。

[0576] 即使在因此难以仅用单一时刻的观察值唯一地识别清洁机器人的位置的生活环境中,也可以根据扩展 HMM 使用观察值系列和行为系列来唯一地识别位置。

[0577] [一个状态一个观察值约束]

[0578] 图 4 中的学习部分 21 使用学习数据来学习扩展 HMM,以便最大化根据 Baum-Welch 重评估法观察到学习数据的可能性。

[0579] Baum-Welch 重评估法本质上是一种通过梯度方法来收敛模型参数的方法。从而模型参数可以落入到局部最小值。

[0580] 通过初始值依赖、即模型参数对于初始值的依赖来确定模型参数是否落入到局部最小值。

[0581] 本实施例使用遍历 HMM 作为扩展 HMM。遍历 HMM 具有特别显著的初始值依赖。

[0582] 学习部分 21(图 4)可以在一个状态一个观察值约束下学习扩展 HMM,以便减小初

始值依赖。

[0583] 在这种情况下,一个状态一个观察值约束使得在扩展 HMM(包括扩展 HMM 的 HMM)的一个状态中观察到(仅仅)一个观察值。

[0584] 顺便提及,当在结构上改变的行为环境中没有任何约束地学习扩展 HMM 时,可以在学习后的扩展 HMM 中混合通过具有在观察概率上的分布而表示行为环境的结构上的改变的情况和通过具有状态转移的分支结构而表示行为环境的结构上的改变的情况。

[0585] 通过具有在观察概率上的分布而表示行为环境的结构上的改变的情况是在某一状态中观察到多个观察值的情况。通过具有状态转移的分支结构而表示行为环境的结构上的改变的情况是由同一行为引起向不同状态的状态转移的情况(当执行某一行为时有可能进行从当前状态向某一状态的状态转移,也有可能进行向与某一状态不同的状态的状态转移)。

[0586] 根据一个状态一个观察值约束,行为环境的结构上的改变仅仅通过在扩展 HMM 中具有状态转移的分支结构表示。

[0587] 顺便提及,当行为环境的结构不改变时,可以不施加一个状态一个观察值约束而学习扩展 HMM。

[0588] 通过向扩展 HMM 的学习中引入状态划分,并且更期望地向扩展 HMM 的学习中引入状态合并(综合),可以施加一个状态一个观察值约束。

[0589] [状态划分]

[0590] 图 31A 和 31B 是帮助说明用于实现一个状态一个观察值约束的状态划分的概要的图。

[0591] 在状态划分中,当在其模型参数(初始状态概率 π_i 、状态转移概率 $a_{ij}(U_m)$ 和观察概率 $b_i(O_k)$) 通过 Baum-Welch 重评估法收敛的扩展 HMM 中的一个状态中观察到多个观察值时,该状态被划分为多个状态,状态的数量等于多个观察值的数量,以便在一个状态中观察到多个观察值的每一个。

[0592] 图 31A 示出了紧接着模型参数通过 Baum-Welch 重评估法收敛后的扩展 HMM(扩展 HMM 的一部分)。

[0593] 在图 31A 中,扩展 HMM 具有三个状态 S_1 、 S_2 和 S_3 。可以在状态 S_1 和 S_2 之间进行状态转移,并且可以在状态 S_2 和 S_3 之间进行状态转移。

[0594] 此外,在图 31A 中,在状态 S_1 中观察到一个观察值 O_{15} ,在状态 S_2 中观察到两个观察值 O_7 和 O_{13} ,在状态 S_3 中观察到一个观察值 O_5 。

[0595] 因为图 31A 中在状态 S_2 中观察到作为多个观察值的两个观察值 O_7 和 O_{13} ,所以状态 S_2 被划分为在数量上等于两个观察值 O_7 和 O_{13} 的数量的两个状态。

[0596] 图 31B 示出了状态划分后的扩展 HMM(扩展 HMM 的一部分)。

[0597] 在图 31B 中,图 31A 中划分之前的状态 S_2 被划分为两个状态,即划分后的状态 S_2 和状态 S_4 ,状态 S_4 是紧接着模型参数已经收敛后的扩展 HMM 中的无效状态之一(例如其状态转移概率和观察概率全部是 0.0(可以视为 0.0 的值)的状态)。

[0598] 此外,在图 31B 中,在划分后的状态 S_2 中仅观察到观察值 O_{13} ,其是在划分之前的状态 S_2 中观察的两个观察值 O_7 和 O_{13} 之一,并且在划分后的状态 S_4 中仅观察到观察值 O_7 ,其是在划分之前的状态 S_2 中观察的两个观察值 O_7 和 O_{13} 的另一个。

[0599] 另外,在图 31B 中,在划分后的状态 S_2 与状态 S_1 和 S_3 的每一个之间,与划分前的状态 S_2 相同,状态转移是可能的。在划分后的状态 S_4 与状态 S_1 和 S_3 的每一个之间,与划分前的状态 S_2 相同,状态转移也是可能的。

[0600] 在状态划分中,学习部分 21 (图 4) 首先检测其中在学习后 (紧邻模型参数已经收敛后) 的扩展 HMM 中观察到多个观察值的状态作为要划分的划分对象状态。

[0601] 图 32 是帮助说明用于检测划分对象状态的方法的图。

[0602] 图 32 示出了扩展 HMM 的观察概率矩阵 B。

[0603] 如参照图 16 所述,观察概率矩阵 B 是将在状态 S_i 中观察到观察值 O_k 的观察概率 $b_i(O_k)$ 作为第 i 行第 k 列的元素的矩阵。

[0604] 在学习扩展 HMM (包括扩展 HMM 的 HMM) 中,观察概率矩阵 B 中在某一状态 S_i 中观察到观察值 O_1 到 O_K 的观察概率 $b_i(O_1)$ 到 $b_i(O_K)$ 的每一个均被归一化,以使得观察概率 $b_i(O_1)$ 到 $b_i(O_K)$ 的总和是 1.0。

[0605] 因此,当在一个状态 S_i 中观察到 (仅仅) 一个观察值时,状态 S_i 的观察概率 $b_i(O_1)$ 到 $b_i(O_K)$ 的最大值是 1.0 (可以视为 1.0 的值),并且除了最大值以外的观察概率是 0.0 (可以视为 0.0 的值)。

[0606] 另一方面,当在一个状态 S_i 中观察到多个观察值时,状态 S_i 的观察概率 $b_i(O_1)$ 到 $b_i(O_K)$ 的最大值远远小于 1.0,诸如图 32 中所示的 0.6 或者 0.5 等,并且大于通过将其总和是 1.0 的观察概率除以观察值 O_1 到 O_K 的数量 K 而获得的值 (平均值) $1/K$ 。

[0607] 因此,可以根据式 (20) 通过关于每个状态 S_i 搜索小于比 1.0 小的阈值 $b_{\max_th}$ 并且大于平均值 $1/K$ 的观察概率 $B_{ik} = b_i(O_k)$ 来检测划分对象状态。

[0608] $\arg \underset{k, i=S}{\text{find}} (1/K < B_{ik} < b_{\max_th}) \quad \dots (20)$

[0609] 在式 (20) 中, B_{ik} 表示在观察概率矩阵 B 的第 i 行第 k 列中的元素,并且等于在状态 S_i 中观察到观察值 O_k 的观察概率 $b_i(O_k)$ 。

[0610] 在式 (20) 中, $\arg \text{find}(1/K < B_{ik} < b_{\max_th})$ 表示当在状态 S_i 的下标 i 是 S 的情况下可以检索到 (找到) 满足括号中的条件表达式 $1/K < B_{ik} < b_{\max_th}$ 的观察概率 B_{ik} 时所有满足括号中的条件表达式 $1/K < B_{ik} < b_{\max_th}$ 的观察概率 B_{ik} 的下标 k。

[0611] 顺便提及,式 (20) 中的阈值 $b_{\max_th}$ 可以根据检测划分对象状态的灵敏度在范围 $1/K < b_{\max_th} < 1.0$ 中调整。使得阈值 $b_{\max_th}$ 越接近 1.0,则可以越灵敏地检测到划分对象状态。

[0612] 当可以检索到 (找到) 满足式 (20) 中括号中的条件表达式 $1/K < B_{ik} < b_{\max_th}$ 的观察概率 B_{ik} 时,学习部分 21 (图 4) 检测其下标 i 是 S 的状态作为划分对象状态。

[0613] 此外,学习部分 21 检测式 (20) 中表达的所有下标 k 的观察值 O_k 作为划分对象状态 (其下标 i 是 S 的状态) 中观察到的多个观察值。

[0614] 然后学习部分 21 将划分对象状态划分为在数量上等于划分对象状态中观察到的多个观察值的数量的多个状态。

[0615] 假定在划分划分对象状态的划分后的状态被称为划分后状态,则可以将划分对象状态用作划分后状态之一,并且可以采用在划分时在扩展 HMM 中的无效状态作为划分后的其他状态。

[0616] 具体地,例如,当划分对象状态被划分为三个划分后状态时,可以将划分对象状态

用作三个划分后状态之一,并且可以采用在划分时在扩展 HMM 中的无效状态作为其他两个划分后状态。

[0617] 另外,可以采用在划分时在扩展 HMM 中的无效状态作为所有的多个划分后状态。但是,在这种情况下,在状态划分后必须使得划分对象状态成为无效状态。

[0618] 图 33A 和 33B 是帮助说明一种将划分对象状态划分为划分后状态的方法的图。

[0619] 在图 33A 和 33B 中,扩展 HMM 具有七个状态 S_1 到 S_7 ,其中两个状态 S_6 和 S_7 是无效状态。

[0620] 此外,在图 33A 和 33B 中,状态 S_3 被设置为其中观察到两个观察值 O_1 和 O_2 的划分对象状态,并且划分对象状态 S_3 被划分为:划分后状态 S_3 ,其中观察到观察值 O_1 ;以及划分后状态 S_6 ,其中观察到观察值 O_2 。

[0621] 学习部分 21 (图 4) 如下将划分对象状态 S_3 划分为两个划分后状态。

[0622] 学习部分 21 向划分对象状态 S_3 的划分后的状态 S_3 分配多个观察值 O_1 和 O_2 之一,例如观察值 O_1 ,并且将在划分后状态 S_3 中观察到被分配给划分后状态 S_3 的观察值 O_1 的观察概率设置为 1.0,将观察到其他观察值的观察概率设置为 0.0。

[0623] 此外,学习部分 21 将以划分后状态 S_3 为转移源的状态转移的状态转移概率 $a_{3,j}(U_m)$ 设置为以划分对象状态 S_3 为转移源的状态转移的状态转移概率 $a_{3,j}(U_m)$,并且将以划分后状态 S_3 为转移目的地的状态转移的状态转移概率设置为通过下述方式获得的值:通过在划分对象状态 S_3 中观察到被分配给划分后状态 S_3 的观察值的观察概率来修正以划分对象状态 S_3 为转移目的地的状态转移的状态转移概率。

[0624] 学习部分 21 关于另一个划分后状态 S_6 类似地设置观察概率和状态转移概率。

[0625] 图 33A 是帮助说明设置划分后状态 S_3 和 S_6 的观察概率的图。

[0626] 在图 33A 和 33B 中,作为通过划分对象状态 S_3 而获得的两个划分后状态 S_3 和 S_6 之一的划分后状态 S_3 被分配有观察值 O_1 ,观察值 O_1 是在划分对象状态 S_3 中观察到的观察值 O_1 和 O_2 之一,并且另一个划分后状态 S_6 被分配有另一个观察值 O_2 。

[0627] 在这种情况下,如图 33A 中所示,学习部分 21 将在被分配了观察值 O_1 的划分后状态 S_3 中观察到观察值 O_1 的观察概率设置为 1.0,并且将观察到其他观察值的观察概率设置为 0.0。

[0628] 此外,如图 33A 中所示,学习部分 21 将在被分配了观察值 O_2 的划分后状态 S_6 中观察到观察值 O_2 的观察概率设置为 1.0,并且将观察到其他观察值的观察概率设置为 0.0。

[0629] 通过式 (21) 来表达如上所述的观察概率的设置。

$$[0630] \quad B(S_3, :) = 0.0$$

$$[0631] \quad B(S_3, O_1) = 1.0$$

$$[0632] \quad B(S_6, :) = 0.0$$

$$[0633] \quad B(S_6, O_2) = 1.0 \quad \dots (21)$$

[0634] 在式 (21) 中, $B(,)$ 是二维阵列,阵列的元素 $B(S, O)$ 表示在状态 S 中观察到观察值 O 的观察概率。

[0635] 作为阵列的下标的冒号 (:) 表示作为冒号的维度的所有元素。因此,在式 (21) 中,式 $B(S_3, :) = 0.0$ 表示在状态 S_3 中观察到相应观察值 O_1 到 O_k 的观察概率全部被设置为 0.0。

[0636] 根据式 (21), 在状态 S_3 中观察到相应观察值 O_1 到 O_k 的观察概率全部被设置为 $0.0 (B(S_3, :) = 0.0)$, 然后只有观察到观察值 O_1 的观察概率被设置为 $1.0 (B(S_3, O_1) = 1.0)$ 。

[0637] 此外, 根据式 (21), 在状态 S_6 中观察到相应观察值 O_1 到 O_k 的观察概率全部被设置为 $0.0 (B(S_6, :) = 0.0)$, 然后只有观察到观察值 O_2 的观察概率被设置为 $1.0 (B(S_6, O_2) = 1.0)$ 。

[0638] 图 33B 是帮助说明设置划分后状态 S_3 和 S_6 的状态转移概率的图。

[0639] 应当将与以划分对象状态 S_3 作为转移源的状态转移类似的状态转移作为以划分后状态 S_3 和 S_6 的每一个作为转移源的状态转移。

[0640] 因此, 如图 33B 中所示, 学习部分 21 将以划分后状态 S_3 为转移源的状态转移的状态转移概率设置为以划分对象状态 S_3 作为转移源的状态转移的状态转移概率。此外, 如图 33B 中所示, 学习部分 21 还将以划分后状态 S_6 为转移源的状态转移的状态转移概率设置为以划分对象状态 S_3 作为转移源的状态转移的状态转移概率。

[0641] 另一方面, 应当将使得以划分对象状态 S_3 作为转移目的地的状态转移按照划分对象状态 S_3 中观察到相应观察值 O_1 和 O_2 的观察概率的比率 (比) 划分的状态转移作为以被分配了观察值 O_1 的划分后状态 S_3 和被分配了观察值 O_2 的划分后状态 S_6 作为转移目的地的状态转移。

[0642] 因此, 如图 33B 中所示, 学习部分 21 通过将以划分对象状态 S_3 为转移目的地的状态转移的状态转移概率乘以在划分对象状态 S_3 中观察到被分配给划分后状态 S_3 的观察值 O_1 的观察概率来修正以划分对象状态 S_3 为转移目的地的状态转移的状态转移概率, 由此获得通过观察值 O_1 的观察概率来修正状态转移概率而产生的修正值。

[0643] 然后学习部分 21 将以被分配了观察值 O_1 的划分后状态 S_3 为转移目的地的状态转移的状态转移概率设置为通过观察值 O_1 的观察概率来修正状态转移概率而产生的修正值。

[0644] 此外, 如图 33B 中所示, 学习部分 21 通过将以划分对象状态 S_3 为转移目的地的状态转移的状态转移概率乘以在划分对象状态 S_3 中观察到被分配给划分后状态 S_6 的观察值 O_2 的观察概率来修正以划分对象状态 S_3 为转移目的地的状态转移的状态转移概率, 由此获得通过观察值 O_2 的观察概率来修正状态转移概率而产生的修正值。

[0645] 然后学习部分 21 将以被分配了观察值 O_2 的划分后状态 S_6 为转移目的地的状态转移的状态转移概率设置为通过观察值 O_2 的观察概率来修正状态转移概率而产生的修正值。

[0646] 通过式 (22) 来表达如上所述的状态转移概率的设置。

$$[0647] \quad A(S_3, :, :) = A(S_3, :, :)$$

$$[0648] \quad A(S_6, :, :) = A(S_3, :, :)$$

$$[0649] \quad A(:, S_3, :) = B(S_3, O_1) A(:, S_3, :)$$

$$[0650] \quad A(:, S_6, :) = B(S_3, O_2) A(:, S_3, :) \quad \cdots (22)$$

[0651] 在式 (22) 中, $A(, ,)$ 是三维阵列, 并且阵列的元素 $A(S, S', U)$ 表示当执行行为 U 时从作为状态转移源的状态 S 向状态 S' 进行状态转移的状态转移概率。

[0652] 像在式 (21) 中那样, 作为阵列的下标的冒号 (:) 代表作为冒号的维度的所有元素。

[0653] 因此, 在式 (22) 中, 例如 $A(S_3, :, :)$ 表示当执行每个行为时从作为转移源的状态 S_3 向每个状态 S 的状态转移的所有状态转移概率。在式 (22) 中, 例如 $A(:, S_3, :)$ 表示当

执行每个行为时从每个状态向作为转移目的地的状态 S_3 的状态转移的所有状态转移概率。

[0654] 根据式 (22), 对于所有的行为, 以划分后状态 S_3 作为转移源的状态转移的状态转移概率被设置为以划分对象状态 S_3 作为转移源的状态转移的状态转移概率 ($A(S_3, :, :) = A(S_3, :, :)$)。

[0655] 另外, 对于所有的行为, 以划分后状态 S_6 作为转移源的状态转移的状态转移概率被设置为以划分对象状态 S_3 作为转移源的状态转移的状态转移概率 ($A(S_6, :, :) = A(S_3, :, :)$)。

[0656] 此外, 根据式 (22), 对于所有的行为, 将以划分对象状态 S_3 作为转移目的地的状态转移的状态转移概率 $A(:, S_3, :)$ 乘以在划分对象状态 S_3 中观察到被分配给划分后状态 S_3 的观察值 O_1 的观察概率 $B(S_3, O_1)$, 由此获得修正以划分对象状态 S_3 作为转移目的地的状态转移的状态转移概率 $A(:, S_3, :)$ 而产生的修正值 $B(S_3, O_1)A(:, S_3, :)$ 。

[0657] 然后, 对于所有的行为, 将以被分配了观察值 O_1 的划分后状态 S_3 作为转移目的地的状态转移的状态转移概率 $A(:, S_3, :)$ 设置为修正值 $B(S_3, O_1)A(:, S_3, :)$ ($A(:, S_3, :) = B(S_3, O_1)A(:, S_3, :)$)。

[0658] 另外, 根据式 (22), 对于所有的行为, 将以划分对象状态 S_3 作为转移目的地的状态转移的状态转移概率 $A(:, S_3, :)$ 乘以在划分对象状态 S_3 中观察到被分配给划分后状态 S_6 的观察值 O_2 的观察概率 $B(S_3, O_2)$, 由此获得修正以划分对象状态 S_3 作为转移目的地的状态转移的状态转移概率 $A(:, S_3, :)$ 而产生的修正值 $B(S_3, O_2)A(:, S_3, :)$ 。

[0659] 然后, 对于所有的行为, 将以被分配了观察值 O_2 的划分后状态 S_6 作为转移目的地的状态转移的状态转移概率 $A(:, S_6, :)$ 设置为修正值 $B(S_3, O_2)A(:, S_3, :)$ ($A(:, S_6, :) = B(S_3, O_2)A(:, S_3, :)$)。

[0660] [状态合并]

[0661] 图 34A 和 34B 是帮助说明用于实现一个状态一个观察值约束的状态合并的概要的图。

[0662] 在状态合并中, 在通过 Baum-Welch 重评估法收敛其模型参数的扩展 HMM 中, 当在执行某一行为时存在多个状态 (不同状态) 作为以一个状态作为转移源的状态转移的转移目的地、并且在所述多个状态的每个中观察到相同的观察值时, 其中观察到相同观察值的所述多个状态被合并为一个状态。

[0663] 此外, 在状态合并中, 在其模型参数收敛了的扩展 HMM 中, 当在执行某一行为时存在多个状态作为以一个状态作为转移目的地的状态转移的转移源、并且在所述多个状态的每一个中观察到相同的观察值时, 其中观察到相同观察值的所述多个状态被合并为一个状态。

[0664] 即, 在状态合并中, 在其模型参数收敛了的扩展 HMM 中, 当针对每个行为发生以相同状态作为转移源或者转移目的地的状态转移、并且存在其中观察到相同观察值的多个状态时, 这样的多个状态是冗余的, 从而被合并为一个状态。

[0665] 在这种情况下, 状态合并包括: 前向合并, 当在执行某一行为时存在多个状态作为从一个状态的状态转移的转移目的地的状态时, 其合并作为转移目的地的多个状态; 以及后向合并, 当在执行某一行为时存在多个状态作为向一个状态的状态转移的转移源的状态时, 其合并作为转移源的多个状态。

[0666] 图 34A 示出了前向合并的示例。

[0667] 在图 34A 中, 扩展 HMM 具有状态 S_1 到 S_5 , 并且从状态 S_1 向状态 S_2 和 S_3 的状态转移、从状态 S_2 向状态 S_4 的状态转移和从状态 S_3 向状态 S_5 的状态转移是可能的。

[0668] 当在状态 S_1 中执行相同的行为时, 进行从状态 S_1 向作为转移目的地的多个状态 S_2 和 S_3 的状态转移的每一个, 即从状态 S_1 向作为转移目的地的状态 S_2 的状态转移和从状态 S_1 向作为转移目的地的状态 S_3 的状态转移。

[0669] 此外, 在状态 S_2 和状态 S_3 中观察到相同的观察值 O_5 。

[0670] 在这种情况下, 学习部分 21 (图 4) 将作为通过相同的行为进行的从一个状态 S_1 的状态转移的转移目的地的、并且在其中观察到相同观察值 O_5 的多个状态 S_2 和 S_3 设置为要合并的合并对象状态, 并且将合并对象状态 S_2 和 S_3 合并为一个状态。

[0671] 以下也将通过合并多个合并对象状态而获得的一个状态称为代表性状态。在图 34A 中, 两个合并对象状态 S_2 和 S_3 被合并为一个代表性状态 S_2 。

[0672] 当执行某一行为时可以从某一状态向其中观察到相同观察值的状态发生的多个状态转移看起来是从一个转移源状态向多个转移目的地状态的分支。因此, 这样的状态转移将被称为前向分支。在图 34A 中, 从状态 S_1 向状态 S_2 的状态转移和从状态 S_1 向状态 S_3 的状态转移是前向分支。

[0673] 顺便提及, 在前向分支的情况下, 作为分支源的状态是作为转移源的状态 S_1 , 并且作为分支目的地的状态是作为其中观察到相同观察值的转移目的地的状态 S_2 和 S_3 。也作为转移目的地状态的分支目的地状态 S_2 和 S_3 是合并对象状态。

[0674] 图 34B 是后向合并的示例。

[0675] 在图 34B 中, 扩展 HMM 具有状态 S_1 到 S_5 , 并且从状态 S_1 向状态 S_3 的状态转移、从状态 S_2 向状态 S_4 的状态转移、从状态 S_3 向状态 S_5 的状态转移和从状态 S_4 向状态 S_5 的状态转移是可能的。

[0676] 当在状态 S_3 和 S_4 中执行相同的行为时, 进行从作为转移源的多个状态 S_3 和 S_4 向状态 S_5 的状态转移的每一个, 即进行从作为转移源的状态 S_3 向状态 S_5 的状态转移和从作为转移源的状态 S_4 向状态 S_5 的状态转移。

[0677] 此外, 在状态 S_3 和状态 S_4 中观察到相同的观察值 O_7 。

[0678] 在这种情况下, 学习部分 21 (图 4) 将作为通过相同的行为进行的从一个状态 S_5 的状态转移的转移源的、并且在其中观察到相同观察值 O_7 的多个状态 S_3 和 S_4 设置为合并对象状态, 并且将合并对象状态 S_3 和 S_4 合并为作为一个状态的代表性状态。

[0679] 在图 34B 中, 作为两个合并对象状态 S_3 和 S_4 之一的状态 S_3 是代表性状态。

[0680] 当执行某一行为时从其中观察到相同观察值的多个状态向作为转移目的地的某一状态的状态转移看起来是从一个转移目的地状态向多个转移源状态的分支。因此, 这样的状态转移将被称为后向分支。在图 34B 中, 从状态 S_3 向状态 S_5 的状态转移和从状态 S_4 向状态 S_5 的状态转移是后向分支。

[0681] 顺便提及, 在后向分支的情况下, 作为分支源的状态是作为转移目的地的状态 S_5 , 并且作为分支目的地的状态是其中观察到相同观察值的、作为转移源的状态 S_3 和 S_4 。也作为转移源状态的分支目的地状态 S_3 和 S_4 是合并对象状态。

[0682] 在状态合并中, 学习部分 21 (图 4) 首先将作为学习后 (紧邻在模型参数已经收敛

之后)的扩展 HMM 中的分支目的地状态的多个状态检测为合并对象状态。

[0683] 图 35A 和 35B 是帮助说明用于检测合并对象状态的方法的图。

[0684] 当在执行预定行为时存在作为扩展 HMM 的状态的、为状态转移的转移源或者转移目的地的多个状态并且在所述多个相应状态中观察到的、具有最大观察概率的观察值彼此一致时,学习部分 21 将所述多个状态检测为合并对象状态。

[0685] 图 35A 表示一种用于将作为前向分支的分支目的地的多个状态检测为合并对象状态的方法。

[0686] 具体地,图 35A 示出了关于某一行行为 U_m 的状态转移概率平面 A 和观察概率矩阵 B。

[0687] 在关于每个行为 U_m 的状态转移概率平面 A 中,状态转移概率被归一化以使得关于每个状态 S_i ,以状态 S_i 为转移源的状态转移概率 $a_{ij}(U_m)$ 的总和(当下标 i 和 m 固定、下标 j 从 1 向 N 改变时所取的 $a_{ij}(U_m)$ 的总和)是 1.0。

[0688] 因此,当在以某一状态 S_i 为分支源的前向上没有分支时关于某一行行为 U_m 以状态 S_i 为转移源的状态转移概率(在关于行为 U_m 的状态转移概率平面 A 中的水平方向上的某一行 i 中布置的状态转移概率)的最大值是 1.0(可以视为 1.0 的值),并且除了所述最大值之外的状态转移概率是 0.0(可以视为 0.0 的值)。

[0689] 另一方面,当在以某一状态 S_i 为分支源的前向上有分支时关于某一行行为 U_m 以状态 S_i 为转移源的状态转移概率的最大值远远小于 1.0,诸如图 35A 中所示的 0.5 等,并且大于通过将其总和是 1.0 的状态转移概率等地除以状态 S_1 至 S_N 的数量 N 而获得的值(平均值) $1/N$ 。

[0690] 因此,可以通过搜索状态 S_i 来检测作为前向分支的分支源的状态,以使得根据式 (19),在关于行为 U_m 的状态转移概率平面中行 i 中的状态转移概率 $a_{ij}(U_m)$ ($= A_{ijm}$) 的最大值小于比 1.0 小的阈值 $a_{\max_th}$ 并且大于平均值 $1/N$,就像在如上所述检测分支结构的的状态的情况下那样。

[0691] 顺便提及,在这种情况下,可以根据检测作为前向分支的分支源的状态的灵敏度在范围 $1/N < a_{\max_th} < 1.0$ 中调整式 (19) 中的阈值 $a_{\max_th}$ 。使阈值 $a_{\max_th}$ 越接近 1.0,则可以越灵敏地检测作为分支源的状态。

[0692] 当如上所述检测作为前向分支的分支源的状态(以下也将所述状态称为分支源状态)时,学习部分 21(图 4)检测多个状态作为从分支源状态的前向分支的分支目的地。

[0693] 即,根据式 (23),当行为 U_m 的下标 m 是 U 并且前向分支的分支源状态 S_i 的下标 i 是 S 时,学习部分 21 检测多个状态作为从分支源状态的前向分支的分支目的地。

[0694] $\arg \underset{j, i=S, m=U}{find}(a_{\min_th1} < A_{ijm}) \quad \dots(23)$

[0695] 在式 (23) 中, A_{ijm} 表示在三维状态转移概率表中在 i 轴方向上从上起的第 i 位置、在 j 轴方向上从左起的第 j 位置和在行为轴方向上从前起的第 m 位置上的状态转移概率 $a_{ij}(U_m)$ 。

[0696] 另外,在式 (23) 中, $\arg \underset{j, i=S, m=U}{find}(a_{\min_th1} < A_{ijm})$ 表示:在行为 U_m 的下标 m 是 U 并且分支源状态 S_i 的下标 i 是 S 的情况下,当可以检索到(找到)满足括号中的条件表达式 $a_{\min_th1} < A_{ijm}$ 的状态转移概率 $A_{s, j, U}$ 时所有满足括号中的条件表达式 $a_{\min_th1} < A_{ijm}$ 的状态转移概率 $A_{s, j, U}$ 的下标 j 。

[0697] 顺便提及,可以根据检测作为前向分支的分支目的地的多个状态的灵敏度在范围

$0.0 < a_{\min_th1} < 1.0$ 中调整式 (23) 中的阈值 $a_{\min_th1}$ 。使阈值 $a_{\min_th1}$ 越接近 1.0, 则可以越灵敏地检测作为前向分支的分支目的地的多个状态。

[0698] 当可以检索到 (找到) 满足式 (23) 中的括号中的条件表达式 $a_{\min_th1} < A_{ijm}$ 的状态转移概率 A_{ijm} 时, 学习部分 21 (图 4) 将其下标是 j 的状态 S_j 检测为作为前向分支的分支目的地的状态的候选 (以下也将所述状态称为分支目的地状态)。

[0699] 其后, 当检测到多个状态作为前向分支的分支目的地状态的候选时, 学习部分 21 确定在分支目的地状态的多个相应候选中观察到的具有最大观察概率的观察值是否彼此一致。

[0700] 然后学习部分 21 在作为在前向分支的分支目的地状态的多个分支目的地状态候选中检测其具有最大观察概率的观察值彼此一致的候选。

[0701] 具体地, 学习部分 21 根据式 (24) 获得关于多个分支目的地状态的候选的每一个的具有最大观察概率的观察值 $O_{\max}$ 。

$$[0702] \quad O_{\max} = \arg \max_{k, j=S} (B_{ik}) \quad \dots(24)$$

[0703] 在式 (24) 中, B_{ik} 表示在状态 S_i 中观察到观察值 O_k 的观察概率 $b_i(O_k)$ 。

[0704] 在式 (24) 中, $\arg \max (B_{ik})$ 表示观察概率矩阵 B 中其下标是 S 的状态的最大观察概率 $B_{S,k}$ 的下标 k 。

[0705] 当关于作为多个分支目的地状态候选的多个状态 S_i 的每一个的下标 i 在式 (24) 中获得的具有最大观察概率 $B_{S,k}$ 的下标 k 彼此一致时, 学习部分 21 在作为前向分支的分支目的地状态的多个分支目的地状态候选中检测其在式 (24) 中获得的具有最大观察概率的下标 k 彼此一致的候选。

[0706] 在图 35A 中, 状态 S_3 被检测为前向分支的分支源状态, 并且其从分支源状态 S_3 的状态转移的状态转移概率都是 0.5 的状态 S_1 和 S_4 被检测为前向分支的分支目的地状态的候选。

[0707] 关于作为前向分支的分支目的地状态的候选的状态 S_1 和 S_4 , 在状态 S_1 中观察到的具有最大观察概率 1.0 的观察值 O_2 和在状态 S_4 中观察到的具有最大观察概率 0.9 的观察值 O_2 彼此一致。因此, 状态 S_1 和 S_4 被检测为前向分支的分支目的地状态。

[0708] 图 35B 表示一种用于将作为后向分支的分支目的地的多个状态检测作为合并对象状态的方法。

[0709] 具体地, 图 35B 示出了关于某一行行为 U_m 的状态转移概率平面 A 和观察概率矩阵 B 。

[0710] 在关于每个行为 U_m 的状态转移概率平面 A 中, 状态转移概率被归一化以使得关于每个状态 S_i , 以状态 S_i 为转移源的状态转移概率 $a_{ij}(U_m)$ 的总和是 1.0, 如上参照图 35A 所述。但是, 不执行使得以状态 S_j 为转移目的地的状态转移概率 $a_{ij}(U_m)$ 的总和 (在下标 j 和 m 固定、下标 i 从 1 向 N 改变时所取的 $a_{ij}(U_m)$ 的总和) 是 1.0 的归一化。

[0711] 但是, 当有可能进行从状态 S_i 向状态 S_j 的状态转移时, 以状态 S_j 为转移目的地的状态转移概率 $a_{ij}(U_m)$ 是非 0.0 (可以视为 0.0 的值) 的正值。

[0712] 因此, 可以根据式 (25) 来检测后向分支的分支源状态 (可以是分支源状态的状态) 和分支目的地状态的候选。

$$[0713] \quad \arg \text{find}(a_{\min_th2} < A_{ijm}) \quad \dots(25)$$

$i, j=S, m=U$

[0714] 在式 (25) 中, A_{ijm} 表示三维状态转移概率表中在 i 轴方向上从顶部起的第 i 位置、

在 j 轴方向上从左起的第 j 位置和在行为轴方向上从前起的第 m 位置上的状态转移概率 $a_{ij}(U_m)$ 。

[0715] 另外,在式 (25) 中, $\text{argfind}(a_{\min_th2} < A_{ijm})$ 表示在行为 U_m 的下标 m 是 U 并且分支源状态 S_j 的下标 j 是 S 的情况下,当可以检索到 (找到) 满足括号中的条件表达式 $a_{\min_th2} < A_{ijm}$ 的状态转移概率 $A_{i,S,U}$ 时所有满足括号中的条件表达式 $a_{\min_th2} < A_{ijm}$ 的状态转移概率 $A_{i,S,U}$ 的下标 i 。

[0716] 顺便提及,可以根据用于检测后向分支的分支源状态和分支目的地状态 (分支目的地状态的候选) 的灵敏度在范围 $0.0 < a_{\min_th2} < 1.0$ 中调整式 (25) 中的阈值 $a_{\min_th2}$ 。使得阈值 $a_{\min_th2}$ 与 1.0 越接近,则可以越灵敏地检测后向分支的分支源状态和分支目的地状态。

[0717] 当可以检索到 (找到) 满足式 (25) 中的括号中的条件表达式 $a_{\min_th2} < A_{ijm}$ 的多个状态转移概率 A_{ijm} 时,学习部分 21 (图 4) 将其下标 j 是 S 的状态检测为后向分支的分支源状态 (作为可能为分支源状态的状态)。

[0718] 学习部分 21 还将下述状态检测为分支目的地状态的候选:当可以检索到满足式 (25) 中的括号中的条件表达式 $a_{\min_th2} < A_{ijm}$ 的多个状态转移概率 A_{ijm} 时与满足式 (25) 中的括号中的条件表达式 $a_{\min_th2} < A_{ijm}$ 的多个状态转移概率 A_{ijm} 对应的状态转移的多个分支源状态,即当可以检索到满足条件表达式 $a_{\min_th2} < A_{ijm}$ 的多个状态转移概率 $A_{i,S,U}$ 时以满足括号中的条件表达式 $a_{\min_th2} < A_{ijm}$ 的多个状态转移概率 $A_{i,S,U}$ 的每一个的 i (由式 (25) 表示的多个 i) 为下标的多个状态 S_i 。

[0719] 其后,学习部分 21 确定在后向分支的分支目的地状态的多个相应候选中观察到的具有最大观察概率的观察值是否彼此一致。

[0720] 然后,像在检测前向分支的分支目的地状态的情况中那样,学习部分 21 在作为后向分支的分支目的地状态的多个分支目的地状态候选中检测其具有最大观察概率的观察值彼此一致的候选。

[0721] 在图 35B 中,状态 S_2 被检测为后向分支的分支源状态,并且其向分支源状态 S_2 的状态转移的状态转移概率都是 0.5 的状态 S_2 和 S_5 被检测为后向分支的分支目的地状态候选。

[0722] 关于作为后向分支的分支目的地状态候选的状态 S_2 和 S_5 ,在状态 S_2 中观察到的具有最大观察概率 1.0 的观察值 O_3 和在状态 S_5 中观察到的具有最大观察概率 0.8 的观察值 O_3 彼此一致。因此,状态 S_2 和 S_5 被检测为后向分支的分支目的地状态。

[0723] 当学习部分 21 如上所述检测到前向和后向分支的分支源状态以及从分支源状态分支出来的多个分支目的地状态时,学习部分 21 将所述多个分支目的地状态合并为一个代表性状态。

[0724] 在这种情况下,学习部分 21 例如将在多个分支目的地状态中具有最小下标的分支目的地状态设置为代表性状态,并且将多个分支目的的状态合并为该代表性状态。

[0725] 即,例如,当三个状态被检测为从某一支源状态分支出来的多个分支目的地状态时,学习部分 21 将在多个分支目的地状态中具有最小下标的目的地状态设置为代表性状态,并且将多个分支目的地状态合并为该代表性状态。

[0726] 另外,学习部分 21 使得三个分支目的地状态中未被设置为代表性状态的其他两

个状态为无效状态。

[0727] 顺便提及,可以从无效状态而不是分支目的地状态中选择状态合并中的代表性状态。在这种情况下,在多个分支目的地状态被合并为代表性状态后,使得多个分支目的地状态全部为无效状态。

[0728] 图 36A 和 36B 是帮助说明一种用于将从某一分支源状态分支出来的多个分支目的地状态合并为一个代表性状态的方法的图。

[0729] 在图 36A 和 36B 中,扩展 HMM 具有 7 个状态 S_1 到 S_7 。

[0730] 此外,在图 36A 和 36B 中,两个状态 S_1 和 S_4 被设置为合并对象状态,两个合并对象状态 S_1 和 S_4 中具有最小下标的状态 S_1 被设置为代表性状态,并且两个合并对象状态 S_1 和 S_4 被合并为一个代表性状态 S_1 。

[0731] 学习部分 21 (图 4) 如下将两个合并对象状态 S_1 和 S_4 合并为一个代表性状态 S_1 。

[0732] 学习部分 21 将在代表性状态 S_1 中观察到相应观察值 O_k 的观察概率 $b_1(O_k)$ 设置为在作为合并对象状态的多个状态 S_1 和 S_4 的每一个中观察到相应观察值 O_k 的观察概率 $b_1(O_k)$ 和 $b_4(O_k)$ 的平均值,并且将作为合并对象状态的多个状态 S_1 和 S_4 的、除了代表性状态 S_1 之外的状态 S_4 中观察到相应观察值 O_k 的观察概率 $b_4(O_k)$ 设置为 0。

[0733] 另外,学习部分 21 将以代表性状态 S_1 作为转移源的状态转移的状态转移概率 $a_{1,j}(U_m)$ 设置为以作为合并对象状态的多个状态 S_1 和 S_4 的每一个作为转移源的状态转移的状态转移概率 $a_{1,j}(U_m)$ 和 $a_{4,j}(U_m)$ 的平均值,并且将以代表性状态 S_1 作为转移目的地的状态转移的状态转移概率 $a_{1,1}(U_m)$ 设置为以作为合并对象状态的多个状态 S_1 和 S_4 的每一个作为转移目的地的状态转移的状态转移概率 $a_{1,1}(U_m)$ 和 $a_{1,4}(U_m)$ 的和。

[0734] 此外,学习部分 21 将以作为合并对象状态的多个状态 S_1 和 S_4 的、除代表性状态 S_1 外的状态 S_4 为转移源的状态转移的状态转移概率 $a_{4,j}(U_m)$ 和以状态 S_4 为转移目的地的状态转移的状态转移概率 $a_{1,4}(U_m)$ 设置为 0。

[0735] 图 36A 是帮助说明设置观察概率的图,所述设置在状态合并中进行。

[0736] 学习部分 21 将在代表性状态 S_1 中观察到观察值 O_1 的观察概率 $b_1(O_1)$ 设置为在合并对象状态 S_1 和 S_4 中观察到观察值 O_1 的观察概率 $b_1(O_1)$ 和 $b_4(O_1)$ 的平均值 $(b_1(O_1) + b_4(O_1)) / 2$ 。

[0737] 类似地设置在代表性状态 S_1 中观察到另一个观察值 O_k 的观察概率 $b_1(O_k)$ 。

[0738] 此外,学习部分 21 将在合并对象状态 S_1 和 S_4 的除了代表性状态 S_1 之外的状态 S_4 中观察到相应观察值 O_k 的观察概率 $b_4(O_k)$ 设置为 0。

[0739] 通过式 (26) 来表达如上所述的观察概率的设置。

$$[0740] \quad B(S_1, :) = (B(S_1, :) + B(S_4, :)) / 2$$

$$[0741] \quad B(S_4, :) = 0.0 \quad \cdots (26)$$

[0742] 在式 (26) 中, $B(,)$ 是二维阵列,并且阵列的元素 $B(S, O)$ 表示在状态 S 中观察到观察值 O 的观察概率。

[0743] 作为阵列的下标的冒号 (:) 表示作为冒号的维度的所有元素。因此,在式 (26) 中,式 $B(S_4, :) = 0.0$ 表示在状态 S_4 中观察到相应观察值的观察概率全部被设置为 0.0。

[0744] 根据式 (26),在代表性状态 S_1 中观察到相应观察值 O_k 的观察概率被设置为在合并对象状态 S_1 和 S_4 的每一个中观察到相应观察值 O_k 的观察概率 $b_1(O_k)$ 和 $b_4(O_k)$ 的平均值

$(B(S_1, :) = (B(S_1, :) + B(S_4, :)) / 2)$ 。

[0745] 此外,根据式 (26),在合并对象状态 S_1 和 S_4 的、除了代表性状态 S_1 之外的状态 S_4 中观察到相应观察值 0_k 的观察概率 $b_4(0_k)$ 被设置为 0 ($B(S_4, :) = 0.0$)。

[0746] 图 36B 是帮助说明设置状态转移概率的图,所述设置在状态合并中进行。

[0747] 以作为合并对象状态的多个状态的每一个作为转移源的状态转移不一定彼此一致。以作为合并对象状态的多个状态的每一个作为转移源的状态转移应当有可能作为以通过合并所述合并对象状态而获得的代表性状态作为转移源的状态转移。

[0748] 因此,如图 36B 中所示,学习部分 21 将以代表性状态 S_1 作为转移源的状态转移的状态转移概率 $a_{1,j}(U_m)$ 设置为以多个合并对象状态 S_1 和 S_4 的每一个作为转移源的状态转移的状态转移概率 $a_{1,j}(U_m)$ 和 $a_{4,j}(U_m)$ 的平均值。

[0749] 以作为合并对象状态的多个状态的每一个作为转移目的地的状态转移也不一定彼此一致。以作为合并对象状态的多个状态的每一个作为转移目的地的状态转移应当有可能作为以通过合并所述合并对象状态而获得的代表性状态作为转移目的地的状态转移。

[0750] 因此,如图 36B 中所示,学习部分 21 将以代表性状态 S_1 为转移目的地的状态转移的状态转移概率 $a_{i,1}(U_m)$ 设置为以合并对象状态 S_1 和 S_4 的每一个作为转移目的地的状态转移的状态转移概率 $a_{i,1}(U_m)$ 和 $a_{i,4}(U_m)$ 的和。

[0751] 顺便提及,虽然以合并对象状态 S_1 和 S_4 作为转移源的状态转移的状态转移概率 $a_{1,j}(U_m)$ 和 $a_{4,j}(U_m)$ 的平均值被用作以代表性状态 S_1 作为转移源的状态转移的状态转移概率 $a_{1,j}(U_m)$,但是以合并对象状态 S_1 和 S_4 作为转移目的地的状态转移的状态转移概率 $a_{i,1}(U_m)$ 和 $a_{i,4}(U_m)$ 的和被用作以代表性状态 S_1 作为转移目的地的状态转移的状态转移概率 $a_{i,1}(U_m)$,因为在关于每个行为 U_m 的状态转移概率平面 A 中,虽然归一化状态转移概率 $a_{ij}(U_m)$ 以使得以状态 S_i 为转移源的状态转移概率 $a_{ij}(U_m)$ 的总和是 1.0,但是不执行使得以状态 S_j 为转移目的地的状态转移概率 $a_{ij}(U_m)$ 的总和是 1.0 的归一化。

[0752] 除了以代表性状态 S_1 为转移源的状态转移概率和以代表性状态 S_1 为转移目的地的状态转移概率的设置之外,学习部分 21 将以合并对象状态 S_4 (除了代表性状态 S_1 之外的合并对象状态) 为转移源的状态转移概率和以合并对象状态 S_4 为转移目的地的状态转移概率设置为 0,其中,通过将合并对象状态 S_1 和 S_4 合并为代表性状态 S_1 而使得合并对象状态 S_4 对于行为环境的结构的表达不必要。

[0753] 通过式 (27) 来表达如上所述的状态转移概率的设置。

[0754] $A(S_1, :, :) = (A(S_1, :, :) + A(S_4, :, :)) / 2$

[0755] $A(:, S_1, :) = A(:, S_1, :) + A(:, S_4, :)$

[0756] $A(S_4, :, :) = 0.0$

[0757] $A(:, S_4, :) = 0.0 \quad \dots (27)$

[0758] 在式 (27) 中, $A(, ,)$ 是三维阵列,并且阵列的元素 $A(S, S', U)$ 表示当执行行为 U 时从作为转移源的状态 S 向状态 S' 进行状态转移的状态转移概率。

[0759] 像在式 (26) 中那样,作为阵列的下标的冒号 (:) 表示作为冒号的维度的所有元素。

[0760] 因此,在式 (27) 中, $A(S_1, :, :)$ 例如表示当执行每个行为时从作为转移源的状态 S_1 向每个状态的状态转移的所有状态转移概率。在式 (27) 中, $A(:, S_1, :)$ 例如表示当执

行每个行为时从每个状态向作为转移目的地的状态 S_1 的状态转移的所有状态转移概率。

[0761] 根据式 (27), 关于所有行为, 以代表性状态 S_1 为转移源的状态转移的状态转移概率被设置为以合并对象状态 S_1 和 S_4 为转移源的状态转移的状态转移概率 $a_{1,j}(U_m)$ 和 $a_{4,j}(U_m)$ 的平均值 ($A(S_1, :, :) = (A(S_1, :, :) + A(S_4, :, :))/2$)。

[0762] 另外, 关于所有的行为, 以代表性状态 S_1 为转移目的地的状态转移的状态转移概率被设置为以合并对象状态 S_1 和 S_4 为转移目的地的状态转移的状态转移概率 $a_{i,1}(U_m)$ 和 $a_{i,4}(U_m)$ 的和 ($A(:, S_1, :) = A(:, S_1, :) + A(:, S_4, :)$)。

[0763] 此外, 根据式 (27), 关于所有的行为, 以合并对象状态 S_4 为转移源的状态转移概率和以合并对象状态 S_4 为转移目的地的状态转移概率被设置为 0 ($A(S_4, :, :) = 0.0$, $A(:, S_4, :) = 0.0$), 其中, 通过将合并对象状态 S_1 和 S_4 合并为代表性状态 S_1 而使得合并对象状态 S_4 对于行为环境的结构的表达不必要。

[0764] 如上所述, 以合并对象状态 S_4 为转移源的状态转移概率和以合并对象状态 S_4 为转移目的地的状态转移概率被设置为 0.0, 其中, 通过将合并对象状态 S_1 和 S_4 合并为代表性状态 S_1 而使得合并对象状态 S_4 对于行为环境的结构的表达不必要, 并且在被使得不必要的合并对象状态 S_4 中将观察到每个观察值的观察概率设置为 0.0。由此使得不必要的合并对象状态 S_4 变为无效状态。

[0765] [在一个状态一个观察值约束下的扩展 HMM 的学习]

[0766] 图 37 是帮助说明一种用于学习扩展 HMM 的处理的流程图, 所述处理由图 4 中的学习部分 21 在一个状态一个观察值约束下执行。

[0767] 在步骤 S91 中, 学习部分 21 使用历史存储部分 14 中存储的作为学习数据的观察值系列和行为系列, 根据 Baum-Welch 重评估法执行扩展 HMM 的初始学习, 即与在图 7 中的步骤 S21 至 S24 类似的处理。

[0768] 在模型参数已经在步骤 S91 中的初始学习中收敛后, 学习部分 21 将扩展 HMM 的模型参数存储在模型存储部分 22 (图 4) 中。然后处理进行到步骤 S92。

[0769] 在步骤 S92 中, 学习部分 21 从模型存储部分 22 中存储的扩展 HMM 中检测划分对象状态。然后处理进行到步骤 S93。

[0770] 当在步骤 S92 中学习部分 21 不能检测到划分对象状态时, 即当模型存储部分 22 中存储的扩展 HMM 中没有划分对象状态时, 处理跳过步骤 S93 和 S94, 并且进行到步骤 S95。

[0771] 在步骤 S93 中, 学习部分 21 执行状态划分, 其将在步骤 S92 中检测到的划分对象状态划分为多个划分后状态。然后处理进行到步骤 S94。

[0772] 在步骤 S94 中, 学习部分 21 使用历史存储部分 14 中存储的作为学习数据的观察值系列和行为系列, 根据 Baum-Welch 重评估法执行在模型存储部分 22 中存储的、并且其中已经在紧邻在前的步骤 S93 中执行了状态划分的扩展 HMM 的学习, 即与图 7 的步骤 S22 至 S24 类似的处理。

[0773] 顺便提及, 在步骤 S94 中的学习中 (也在后述的步骤 S97 中), 存储在模型存储部分 22 中的扩展 HMM 的模型参数被原样用作模型参数的初始值。

[0774] 在扩展 HMM 的模型参数已经在步骤 S94 的学习中收敛后, 学习部分 21 将扩展 HMM 的模型参数存储 (覆盖) 在模型存储部分 22 (图 4) 中。然后处理进行到步骤 S95。

[0775] 在步骤 S95 中, 学习部分 21 从存储在模型存储部分 22 中的扩展 HMM 中检测合并

对象状态。然后处理进行到步骤 S96。

[0776] 当在步骤 S95 中学习部分 21 不能检测到合并对象状态时,即当存储在模型存储部分 22 中的扩展 HMM 中没有合并对象状态时,处理跳过步骤 S96 和 S97,并且进行到步骤 S98。

[0777] 在步骤 S96 中,学习部分 21 执行状态合并,用于将在步骤 S95 中检测到的合并对象状态合并为代表性状态。然后处理进行到步骤 S97。

[0778] 在步骤 S97 中,学习部分 21 使用历史存储部分 14 中存储的作为学习数据的观察值系列和行为系列,根据 Baum-Welch 重评估法执行模型存储部分 22 中存储的、并且其中在紧邻在前的步骤 S96 中执行了状态合并的扩展 HMM 的学习,即与图 7 中的步骤 S22 至 S24 类似的处理。

[0779] 在扩展 HMM 的模型参数已经在步骤 S97 中的学习中收敛后,学习部分 21 将扩展 HMM 的模型参数存储(覆盖)在模型存储部分 22(图 4)中。然后处理进行到步骤 S98。

[0780] 在步骤 S98 中,学习部分 21 确定是否在步骤 S92 中检测划分对象状态的处理中未检测到划分对象状态并且在步骤 S95 中检测合并对象状态的处理中未检测到合并对象状态。

[0781] 当在步骤 S98 中确定检测到划分对象状态和合并对象状态的至少一个时,处理返回到步骤 S92,以重复从步骤 S92 向下的类似处理。

[0782] 当在步骤 S98 中确定检测不到划分对象状态和合并对象状态两者的任何一个时,结束学习扩展 HMM 的处理。

[0783] 通过重复如上所述的状态划分、在状态划分后的扩展 HMM 的学习、状态合并和在状态合并后的扩展 HMM 的学习直到检测不到划分对象状态和合并对象状态的任何一个时,执行满足一个状态一个观察值约束的学习,并且可以获得其中在一个状态中观察到(仅仅)一个观察值的扩展 HMM。

[0784] 图 38 是帮助说明检测划分对象状态的处理的流程图,所述处理由图 4 中的学习部分 21 在图 37 的步骤 S92 中执行。

[0785] 在步骤 S111 中,学习部分 21 将指示状态 S_i 的下标的变量 i 初始化为例如 1。然后处理进行到步骤 S112。

[0786] 在步骤 S112 中,学习部分 21 将指示观察值 O_k 的下标的变量 k 初始化为例如 1。然后处理进行到步骤 S113。

[0787] 在步骤 S113 中,学习部分 21 确定在状态 S_i 中观察到观察值 O_k 的观察概率 $B_{ik} = b_i(O_k)$ 是否满足式 (20) 中括号中的条件表达式 $1/K < B_{ik} < b_{\max_th}$ 。

[0788] 当在步骤 S113 中确定观察概率 $B_{ik} = b_i(O_k)$ 不满足条件表达式 $1/K < B_{ik} < b_{\max_th}$ 时,处理跳过步骤 S114,并且进行到步骤 S115。

[0789] 当在步骤 S113 中确定观察概率 $B_{ik} = b_i(O_k)$ 满足条件表达式 $1/K < B_{ik} < b_{\max_th}$ 时,处理进行到步骤 S114,在步骤 S114 中,学习部分 21 将观察值 O_k 作为要划分的观察值(要在一对一基础上分配给划分后状态的观察值)与状态 S_i 相关联地暂时存储在图中未示出的存储器中。

[0790] 其后的处理从步骤 S114 进行到步骤 S115,在步骤 S115 中,确定下标 k 是否等于观察值的数量 K (以下也称为符号的数量)。

[0791] 当在步骤 S115 中确定下标 k 不等于符号的数量 K 时,处理进行到步骤 S116,在步

骤 S116 中,学习部分 21 将下标 k 递增 1。然后处理从步骤 S116 返回到步骤 S113,以重复从步骤 S113 向下的类似处理。

[0792] 当在步骤 S115 中确定下标 k 等于符号的数量 K 时,处理进行到步骤 S117,在步骤 S117 中,学习部分 21 确定下标 i 是否等于状态的数量 N (扩展 HMM 的状态的数量)。

[0793] 当在步骤 S117 中确定下标 i 不等于状态的数量 N 时,处理进行到步骤 S118,在步骤 S118 中,学习部分 21 将下标 i 递增 1。然后处理从步骤 S118 返回到步骤 S112,以重复从步骤 S112 向下的类似处理。

[0794] 当在步骤 S117 中确定下标 i 等于状态的数量 N 时,处理进行到步骤 S119,在步骤 S119 中,学习部分 21 将在步骤 S114 中与要划分的观察值相关联地存储的每个状态 S_i 检测为划分对象状态。然后处理返回。

[0795] 图 39 是帮助说明划分状态(划分划分对象状态)的处理的流程图,所述处理由图 4 中的学习部分 21 在图 37 中的步骤 S93 中执行。

[0796] 在步骤 S131 中,学习部分 21 将其状态还没有被设置为感兴趣的状态的划分对象状态之一选择为感兴趣的状态。然后处理进行到步骤 S132。

[0797] 在步骤 S132 中,学习部分 21 将与感兴趣的状态相关联的、要划分的观察值的数量设置为感兴趣的状态被划分成的划分后状态的数量 C_s (以下也所述数量称为划分数量),并且总共选择 C_s 个状态作为划分后状态,所述 C_s 个状态即在扩展 HMM 的状态中的感兴趣的状态和 C_s-1 个无效状态的状态。

[0798] 其后处理从步骤 S132 进行到步骤 S133,在步骤 S133 中,学习部分 21 将与感兴趣的状态相关联的、 C_s 个要划分的观察值中的一个分配给 C_s 个划分后状态的每一个。然后处理进行到步骤 S134。

[0799] 在步骤 S134 中,学习部分 21 将用于计数 C_s 个划分后状态的变量 c 初始化为例如 1。然后处理进行到步骤 S135。

[0800] 在步骤 S135 中,学习部分 21 在 C_s 个划分后状态中选择第 c 个划分后状态作为划分后感兴趣的状态,注意力转向该划分后感兴趣的状态。然后处理进行到步骤 S136。

[0801] 在步骤 S136 中,学习部分 21 将在划分后感兴趣的状态中观察到被分配给划分后感兴趣的状态的要划分的观察值的观察概率设置为 1.0,并且将观察到其他观察值的观察概率设置为 0.0。然后处理进行到步骤 S137。

[0802] 在步骤 S137 中,学习部分 21 将以划分后感兴趣的状态为转移源的状态转移的状态转移概率设置为以感兴趣的状态为转移源的状态转移的状态转移概率。然后处理进行到步骤 S138。

[0803] 在步骤 S138 中,学习部分 21 通过在感兴趣的状态中观察到被分配给划分后感兴趣的状态的划分对象状态的观察值的观察概率来修正以感兴趣的状态为转移目的地的状态转移的状态转移概率,如参照图 33A 和 33B 所述,并且学习部分 21 由此获得状态转移概率的修正值。然后处理进行到步骤 S139。

[0804] 在步骤 S139 中,学习部分 21 将以划分后感兴趣的状态为转移目的地的状态转移的状态转移概率设置为在紧邻在前的步骤 S138 中获得的修正值。然后处理进行到步骤 S140。

[0805] 在步骤 S140 中,学习部分 21 确定变量 c 是否等于划分数量 C_s 。

[0806] 当在步骤 140 中确定变量 c 不等于划分数量 C_s 时,处理进行到步骤 S141,在步骤 S141 中,学习部分 21 将变量 c 递增 1。然后处理返回到步骤 S135。

[0807] 当在步骤 S140 中确定变量 c 等于划分数量 C_s 时,处理进行到步骤 S142,在步骤 S142 中,学习部分 21 确定是否所有的划分对象状态已经被选择为感兴趣的状态。

[0808] 当在步骤 S142 中确定未将所有的划分对象状态选择为感兴趣的状态时,处理返回到步骤 S131,以重复从步骤 S131 向下的类似处理。

[0809] 当在步骤 S142 中确定已经将所有的划分对象状态选择为感兴趣的状态时,即当已经完成了所有划分对象状态的划分时,处理返回。

[0810] 图 40 是帮助说明检测合并对象状态的处理的流程图,所述处理由图 4 中的学习部分 21 在图 37 的步骤 S95 中执行。

[0811] 在步骤 S161 中,学习部分 21 将指示行为 U_m 的下标的变量 m 初始化为例如 1。然后处理进行到步骤 S162。

[0812] 在步骤 S162 中,学习部分 21 将指示状态 S_i 的下标的变量 i 初始化为例如 1。然后处理进行到步骤 S163。

[0813] 在步骤 S163 中,学习部分 21 在模型存储部分 22 中存储的扩展 HMM 中检测关于行为 U_m 以状态 S_i 为转移源向每个状态 S_j 的状态转移的状态转移概率 $A_{ijm} (= a_{ij}(U_m))$ 的最大值 $\max(A_{ijm})$ 。然后处理进行到步骤 S164。

[0814] 在步骤 S164 中,学习部分 21 确定最大值 $\max(A_{ijm})$ 是否满足式 (19),即式 $1/N < \max(A_{ijm}) < a_{\max_th}$ 。

[0815] 当在步骤 S164 中确定最大值 $\max(A_{ijm})$ 不满足式 (19) 时,处理跳过步骤 S165,并且进行到步骤 S166。

[0816] 当在步骤 S164 中确定最大值 $\max(A_{ijm})$ 满足式 (19) 时,处理进行到步骤 S165,在步骤 S165 中,学习部分 21 检测状态 S_i 作为前向分支的分支源状态。

[0817] 此外,学习部分 21 检测状态 S_j 作为前向分支的分支目的地状态,状态 S_j 是在关于行为 U_m 以前向分支的分支源状态 S_i 为转移源的状态转移中其状态转移概率 $A_{ijm} = a_{ij}(U_m)$ 满足式 (23) 中括号中的条件表达式 $a_{\min_th1} < A_{ijm}$ 的状态转移的转移目的地。然后处理从步骤 S165 进行到步骤 S166。

[0818] 在步骤 S166 中,学习部分 21 确定下标 i 是否等于状态的数量 N 。

[0819] 当在步骤 S166 中确定下标 i 不等于状态的数量 N 时,处理进行到步骤 S167,在步骤 S167 中,学习部分 21 将下标 i 递增 1。然后处理返回到步骤 S163。

[0820] 当在步骤 S166 中确定下标 i 等于状态的数量 N 时,处理进行到步骤 S168,在步骤 S168 中,学习部分 21 将指示状态 S_j 的下标的变量 j 初始化为例如 1。然后处理进行到步骤 S169。

[0821] 在步骤 S169 中,学习部分 21 确定是否存在多个状态 $S_{i'}$ 作为在关于行为 U_m 从每个状态 $S_{i'}$ 向作为转移目的地的状态 S_j 的状态转移中其状态转移概率 $A_{i'jm} = a_{i'j}(U_m)$ 满足式 (25) 中括号中的条件表达式 $a_{\min_th2} < A_{i'jm}$ 的状态转移的转移源。

[0822] 当在步骤 S169 中确定没有多个状态 $S_{i'}$ 作为满足式 (25) 中括号中的条件表达式 $a_{\min_th2} < A_{i'jm}$ 的状态转移的转移源时,处理跳过步骤 S170,并且进行到步骤 S171。

[0823] 当在步骤 S169 中确定存在多个状态 $S_{i'}$ 作为满足式 (25) 中括号中的条件表达式

$a_{\min_th2} < A_{i', j_m}$ 的状态转移的转移源时, 处理进行到步骤 S170, 在步骤 S170 中, 学习部分 21 检测状态 S_j 作为后向分支的分支源状态。

[0824] 此外, 学习部分 21 检测多个状态 $S_{i'}$ 作为后向分支的分支目的地状态, 所述多个状态 $S_{i'}$ 作为在关于行为 U_m 从每个状态 $S_{i'}$ 向以后向分支的分支源状态 S_j 作为转移目的地的状态转移中其状态转移概率 $A_{i', j_m} = a_{i', j}(U_m)$ 满足式 (25) 中括号中的条件表达式 $a_{\min_th2} < A_{i', j_m}$ 的状态转移的转移源。然后处理从步骤 S170 进行到步骤 S171。

[0825] 在步骤 S171 中, 学习部分 21 确定下标 j 是否等于状态的数量 N 。

[0826] 当在步骤 S171 中确定下标 j 不等于状态的数量 N 时, 处理进行到步骤 S172, 在步骤 S172 中, 学习部分 21 将下标 j 递增 1。然后处理返回到步骤 S169。

[0827] 当在步骤 S171 中确定下标 j 等于状态的数量 N 时, 处理进行到步骤 S173, 在步骤 S173 中, 学习部分 21 确定下标 m 是否等于行为 U_m 的数量 M (以下也将所述数量称为行为数量)。

[0828] 当在步骤 S173 中确定下标 m 不等于行为 U_m 的数量 M 时, 处理进行到步骤 S174, 在步骤 S174 中, 学习部分 21 将下标 m 递增 1。然后处理返回到步骤 S162。

[0829] 当在步骤 S173 中确定下标 m 等于行为 U_m 的数量 M 时, 处理进行到图 41 中的步骤 S191。

[0830] 图 41 是从图 40 继续的流程图。

[0831] 在图 41 中的步骤 S191 中, 学习部分 21 将在由图 40 中的步骤 S161 至 S174 的处理检测到的分支源状态中还没有被设置为感兴趣的状态的一个分支源状态选择为感兴趣的状态。然后处理进行到步骤 S192。

[0832] 在步骤 S192 中, 关于与感兴趣的状态相关地检测到的多个分支目的地状态的每一个 (分支目的地状态的候选), 即关于从作为分支源的感兴趣的状态分支出的多个分支目的地状态 (分支目的地状态的候选) 的每一个, 学习部分 21 根据式 (24) 检测在分支目的地状态中观察到的具有最大观察概率的观察值 $O_{\max}$ (以下也将该观察值称为最大概率观察值)。然后处理进行到步骤 S193。

[0833] 在步骤 S193 中, 学习部分 21 确定与感兴趣的状态相关地检测到的多个分支目的地状态是否包括其最大概率观察值 $O_{\max}$ 彼此一致的分支目的地状态。

[0834] 当在步骤 S193 中确定与感兴趣的状态相关地检测到的多个分支目的地状态不包括其最大概率观察值 $O_{\max}$ 彼此一致的分支目的地状态时, 处理跳过步骤 S194, 并且进行到步骤 S195。

[0835] 当在步骤 S193 中确定与感兴趣的状态相关地检测到的多个分支目的地状态包括其最大概率观察值 $O_{\max}$ 彼此一致的分支目的地状态时, 处理进行到步骤 S194, 在步骤 S194 中, 学习部分 21 将在与感兴趣的状态相关地检测到的多个分支目的地状态中其最大概率观察值 $O_{\max}$ 彼此一致的多个分支目的地状态检测为一组的合并对象状态。然后处理进行到步骤 S195。

[0836] 在步骤 S195 中, 学习部分 21 确定是否已经将所有分支源状态选择为感兴趣的状态。

[0837] 当在步骤 S195 中确定还没有将所有分支源状态选择为感兴趣的状态时, 处理返回到步骤 S191。

[0838] 当在步骤 S195 中确定已经将所有的分支源状态选择为感兴趣的状态时,处理返回。

[0839] 图 42 是帮助说明合并状态(合并合并对象状态)的处理的流程图,所述处理由图 4 中的学习部分 21 在图 37 的步骤 S96 中执行。

[0840] 在步骤 S211 中,学习部分 21 将合并对象状态的组中还没有被选择为感兴趣的组的一个组选择为感兴趣的组。然后处理进行到步骤 S212。

[0841] 在步骤 S212 中,学习部分 21 将在感兴趣的组的多个合并对象状态中例如具有最小下标的合并对象状态选择为感兴趣的组的代表性状态。然后处理进行到步骤 S213。

[0842] 在步骤 S213 中,学习部分 21 将在代表性状态中观察到每个观察值的观察概率设置为在感兴趣的组的多个相应的合并对象状态中观察到每个观察值的观察概率的平均值。

[0843] 此外,在步骤 S213 中,学习部分 21 将在感兴趣的组的除了代表性状态之外的合并对象状态中观察到每个观察值的观察概率设置为 0.0。然后处理进行到步骤 S214。

[0844] 在步骤 S214 中,学习部分 21 将以代表性状态为转移源的状态转移的状态转移概率设置为以感兴趣的组的合并对象状态的每一个作为转移源的状态转移的状态转移概率的平均值。然后处理进行到步骤 S215。

[0845] 在步骤 S215 中,学习部分 21 将以代表性状态作为转移目的地的状态转移的状态转移概率设置为以感兴趣的组的合并对象状态的每一个作为转移目的地的状态转移的状态转移概率的和。然后处理进行到步骤 S216。

[0846] 在步骤 S216 中,学习部分 21 将以感兴趣的组的除了代表性状态之外的合并对象状态作为转移源的状态转移和以感兴趣的组的除了代表性状态之外的合并对象状态作为转移目的地的状态转移的状态转移概率设置为 0.0。然后处理进行到步骤 S217。

[0847] 在步骤 S217 中,学习部分 21 确定是否已经将合并对象状态的所有组选择为感兴趣的组。

[0848] 当在步骤 S217 中确定还没有将所有的合并对象状态的组选择为感兴趣的组时,处理返回到步骤 S211。

[0849] 当在步骤 S217 中确定已经将所有的合并对象状态的组选择为感兴趣的组时,处理返回。

[0850] 图 43A、43B 和 43C 是帮助说明在一个状态一个观察值约束下的扩展 HMM 的学习的模拟的图,所述模拟由本发明人执行。

[0851] 图 43A 是示出了模拟中使用的行为环境的图。

[0852] 在模拟中,其结构被转换为第一结构和第二结构的环境用作行为环境。

[0853] 在第一结构的行为环境中,位置 pos 是墙壁,从而不能被通过,而在第二结构的行为环境中,位置 pos 是通道,从而可以被通过。

[0854] 在模拟中,在第一结构的行为环境和第二结构的行为环境的每一个中获得用作学习数据的观察值系列和行为系列,并且执行扩展 HMM 的学习。

[0855] 图 43B 示出了在没有一个状态一个观察值约束的情况下作为学习的结果获得的扩展 HMM。图 43C 示出了在一个状态一个观察值约束下作为学习的结果获得的扩展 HMM。

[0856] 在图 43B 和图 43C 中,圆圈(o)标记代表扩展 HMM 的状态,并且在圆圈标记中输入的数是由圆圈标记代表的状态的下标。指示由圆圈标记代表的状态的箭头代表可能的状态

转移（其状态转移概率不是 0.0（可以视为 0.0 的值）的状态转移）。

[0857] 另外，在图 43B 和图 43C 中，在左侧位置以垂直方向布置的状态（代表状态的圆圈标记）是扩展 HMM 中的无效状态。

[0858] 图 43B 中的扩展 HMM 确认在没有一个状态一个观察值约束的学习中，模型参数落入局部最小值中，在学习后的扩展 HMM 中通过具有在观察概率上的分布而代表在结构上改变的行为环境的第一结构和第二结构的情况和通过具有状态转移的分支结构而代表在结构上改变的行为环境的第一结构和第二结构的情况相互混合，从而不能通过扩展 HMM 的状态转移适当地代表在结构上改变的行为环境的结构。

[0859] 另一方面，图 43C 中的扩展 HMM 确认在一个状态一个观察值约束下的学习中，在学习后的扩展 HMM 中通过仅仅具有状态转移的分支结构而代表在结构上改变的行为环境的第一结构和第二结构，从而可以通过扩展 HMM 的状态转移来适当地代表在结构上改变的行为环境的结构。

[0860] 根据在一个状态一个观察值约束下的学习，当行为环境的结构改变时，在扩展 HMM 中公共地存储结构上不改变的部分，并且在扩展 HMM 中用状态转移的分支结构（作为当执行某一行行为时发生的状态转移的向不同状态的（多个）状态转移）来代表结构上改变的部分。

[0861] 因此，由于可以用单个扩展 HMM 来适当地代表结构上改变的行为环境而不用针对每种结构准备模型，因此可以用少量存储资源对结构上改变的行为环境建模。

[0862] [在用于根据预定策略来确定行为的识别行为模式中的处理]

[0863] 在图 8 的识别行为模式中的处理中，图 4 的代理假定代理位于行为环境的已知区域（当使用在该区域中观察的观察值系列和行为系列学习了扩展 HMM 时的区域（学习区域））中识别代理的当前状况，获得对应于当前状况的作为扩展 HMM 的状态作为当前状态，并且确定用于从当前状态到达目标状态的行为。但是，代理不一定位于已知区域中，而是可以位于未知区域（未学习的区域）中。

[0864] 在代理位于未知区域的情况下，即使在如参照图 8 所述确定了行为时，该行为也不一定是用于到达目标状态的适当行为，反而可能实际上是不必要或者冗余的行为，诸如在未知区域中的迷路等。

[0865] 因此，在识别行为模式中的代理可以确定代理的当前状况是未知状况（其中观察到目前为止还未观察过的观察值系列和行为系列的状况）（没有在扩展 HMM 中获得的状况）还是已知状况（其中观察到目前为止观察过的观察值系列和行为系列的状况）（在扩展 HMM 中获得的状况），并且根据确定的结果来确定适当的行为。

[0866] 图 44 是帮助说明在这样的识别行为模式中的处理的流程图。

[0867] 在图 44 中的识别行为模式中，代理执行与图 8 中步骤 S31 至 S33 的处理类似的处理。

[0868] 其后，处理进行到步骤 S301，在步骤 S301 中，代理的状态识别部分 23（图 4）从历史存储部分 14 中读出其系列长度（形成系列的值的数量） q 是预定长度 Q 的最新观察值系列和当观察到所述观察值系列的每个观察值时执行的行为的行为系列作为用于识别出代理的当前状况的用于识别的观察值系列和行为系列，并且由此获得用于识别的观察值系列和行为系列。

[0869] 然后,处理从步骤 S301 进行到步骤 S302,在步骤 S302 中,状态识别部分 23 在模型存储部分 22 中存储的学习过的扩展 HMM 中观察用于识别的观察值系列和行为系列,并且根据基于维特比算法的上述式 (10) 和 (11) 来获得最佳状态概率 $\delta_t(j)$ 和最佳路径 $\psi_t(j)$,最佳状态概率 $\delta_t(j)$ 是在时刻 t 处位于状态 S_j 中的状态概率的最大值,最佳路径 $\psi_t(j)$ 是提供最佳状态概率 $\delta_t(j)$ 的状态系列。

[0870] 此外,状态识别部分 23 观察用于识别的观察值系列和行为系列,并且获得最大可能状态系列,该最大可能状态系列是用于从式 (11) 中的最佳路径 $\psi_t(j)$ 到达状态 S_j 的状态系列,所述状态系列最大化时刻 t 处式 (10) 中的最佳状态概率 $\delta_t(j)$ 。

[0871] 其后处理从步骤 S302 进行到步骤 S303,在步骤 S303 中,状态识别部分 23 根据最大可能状态系列来确定代理的当前状况是已知状况还是未知状况。

[0872] 以下,用于识别的观察值系列(或者用于识别的观察值系列和行为系列)将被表示为 O ,并且其中观察到用于识别的观察值系列 O 和行为系列的最大可能状态系列将被表示为 X 。顺便提及,形成最大可能状态系列 X 的状态的数量等于用于识别的观察值系列 O 的系列长度 q 。

[0873] 观察到用于识别的观察值系列 O 的第一观察值的时刻 t 被设置为例如 1。最大可能状态系列 X 在时刻 t 处的状态(从开始起的第 t 个状态)将被表示为 X_t ,并且从状态 X_t 向时刻 $t+1$ 处的状态 X_{t+1} 的状态转移的状态转移概率将被表示为 $A(X_t, X_{t+1})$ 。

[0874] 此外,在最大可能状态系列 X 中观察到用于识别的观察值系列 O 的可能性将被表示为 $P(O|X)$ 。

[0875] 在步骤 S303 中,状态识别部分 23 确定是否满足式 (28) 和式 (29)。

[0876] $A(X_t, X_{t+1}) > \text{Thres}_{\text{trans}} (0 < t < q) \quad \dots (28)$

[0877] $P(O|X) > \text{Thres}_{\text{obs}} \quad \dots (29)$

[0878] 式 (28) 中的 $\text{Thres}_{\text{trans}}$ 是用于区分从状态 X_t 向状态 X_{t+1} 的状态转移是否可以发生的阈值。式 (29) 中的 $\text{Thres}_{\text{obs}}$ 是用于区分在最大可能状态系列 X 中是否可以观察到用于识别的观察值系列 O 的阈值。例如,通过模拟等将可以适当地进行上述区分的值设置为阈值 $\text{Thres}_{\text{trans}}$ 和 $\text{Thres}_{\text{obs}}$ 。

[0879] 当不满足式 (28) 和式 (29) 的至少一个时,状态识别部分 23 在步骤 S303 中确定代理的当前状况是未知状况。

[0880] 当满足式 (28) 和式 (29) 两者时,状态识别部分 23 在步骤 S303 中确定代理的当前状况是已知状况。

[0881] 当在步骤 S303 中确定当前状况是已知状况时,状态识别部分 23 获得最大可能状态系列 X 的最后状态作为当前状态 S_t (将最大可能状态系列 X 的最后状态估计为当前状态 S_t)。然后处理进行到步骤 S304。

[0882] 在步骤 S304 中,状态识别部分 23 根据当前状态 S_t 更新存储在过去时间管理表存储部分 32(图 4) 中的过去时间管理表,就像图 8 的步骤 S34 中那样。

[0883] 其后代理执行与图 8 中从步骤 S35 向下的处理类似的处理。

[0884] 另一方面,当在步骤 S303 中确定当前状况是未知状况时,处理进行到步骤 S305,在步骤 S305 中,状态识别部分 23 根据模型存储部分 22 中存储的扩展 HMM 计算用于使代理到达当前状况的状态系列的当前状况状态系列的一个或更多个候选。

[0885] 此外,状态识别部分 23 将当前状况状态系列的一个或多个候选提供给行为确定部分 24(图 4)。然后处理从步骤 S305 进行到步骤 S306。

[0886] 在步骤 S306 中,行为确定部分 24 使用来自状态识别部分 23 的当前状况状态系列的一个或多个候选根据预定策略来确定接下来要由代理执行的行为。

[0887] 其后代理执行与图 8 中从步骤 S40 向下的处理类似的处理。

[0888] 如上所述,在当前状况是未知状况时,代理计算当前状况状态系列的一个或多个候选,并且使用当前状况状态系列的一个或多个候选根据预定策略来确定代理的行为。

[0889] 即,在当前状况是未知状况时,代理从可以从过去的经历获得的状态系列、即在学习的扩展 HMM 中发生的状态转移的状态系列(以下也将该状态系列称为经历的状态系列)中获得导致当前状况的状态系列作为当前状况状态系列的候选,在所述状态系列中观察到某一系列长度 q 的最新观察值系列和最新行为系列。

[0890] 然后代理使用(重新使用)当前状况状态系列作为经历的状态系列根据预定策略来确定代理的行为。

[0891] [当前状况状态系列的候选的计算]

[0892] 图 45 是帮助说明计算当前状况状态系列的候选的处理的流程图,所述处理由图 4 的状态识别部分 23 在图 44 中的步骤 S305 中执行。

[0893] 在步骤 S311 中,状态识别部分 23 从历史存储部分 14(图 4)中读出其系列长度 q 是预定长度 Q' 的最新观察值系列和当观察到所述观察值系列的每个观察值时执行的行为系列(由代理执行的行为的最新行为系列,其系列长度 q 是预定长度 Q' ;以及,当执行所述行为系列的行为时在代理中观察到的观察值的观察值系列)作为用于识别的观察值系列和行为系列,并且由此获得用于识别的观察值系列和行为系列。

[0894] 在这种情况下,作为由状态识别部分 23 在步骤 S311 中获得的用于识别的观察值系列的系列长度 q 的长度 Q' 例如是 1,其短于作为在图 44 的步骤 S301 中获得的观察值系列的系列长度 q 的长度 Q 。

[0895] 即,如上所述,代理从经历的状态系列中获得其中观察到用于识别的观察值系列和行为系列作为最新观察值系列和最新行为系列的状态系列作为当前状况状态系列的候选。当用于识别的观察值系列和行为系列的系列长度 q 太长时,可能在经历的状态系列中找不到其中观察到这样长系列长度 q 的用于识别的观察值系列和行为系列的状态系列(或者仅仅有找到这样的状态系列的大致等于 0 的可能性)。

[0896] 因此,状态识别部分 23 在步骤 S311 中获得短系列长度 q 的用于识别的观察值系列和行为系列,以便能够从经历的状态系列中获得其中观察到用于识别的观察值系列和行为系列的状态系列。

[0897] 在步骤 S311 后,处理进行到步骤 S312,在步骤 S312 中,状态识别部分 23 在模型存储部分 22 中存储的扩展 HMM 中观察在步骤 S311 中获得的用于识别的观察值系列和行为系列,并且根据基于维特比算法的上述式 (10) 和 (11) 获得最佳状态概率 $\delta_t(j)$ 和最佳路径 $\psi_t(j)$,最佳状态概率 $\delta_t(j)$ 是在时刻 t 处位于状态 S_j 中的状态概率的最大值,最佳路径 $\psi_t(j)$ 是提供最佳状态概率 $\delta_t(j)$ 的状态系列。

[0898] 即,状态识别部分 23 从经历的状态系列中获得最佳路径 $\psi_t(j)$,其是其系列长度

为 Q' 的状态系列,在所述状态系列中观察到用于识别的观察值系列和行为系列。

[0899] 以下也将作为根据维特比算法获得(估计)的最佳路径 $\psi_t(j)$ 的状态系列称为用于识别的状态系列。

[0900] 在步骤 S312 中,针对扩展 HMM 的 N 个状态 S_j 的每一个获得最佳状态概率 $\delta_t(j)$ 和用于识别的状态系列(最佳路径 $\psi_t(j)$)。

[0901] 在步骤 S312 中获得用于识别的状态系列后,处理进行到步骤 S313,在步骤 S313 中,状态识别部分 23 从在步骤 S312 中获得的用于识别的状态系列中选择一个或多个用于识别的状态系列作为当前状况状态系列的候选。然后处理返回。

[0902] 即,在步骤 S313 中,例如选择其可能性即最佳状态概率 $\delta_t(j)$ 等于或者高于阈值(例如最佳状态概率 $\delta_t(j)$ 的最大值(最大可能性)的 0.8 倍的值)的用于识别的状态系列作为当前状况状态系列的候选。

[0903] 或者,例如选择其最佳状态概率 $\delta_t(j)$ 在 R 个最高分级中的 R (R 是一或者更大的整数)个用于识别的状态系列作为当前状况状态系列的候选。

[0904] 图 46 是帮助说明计算当前状况状态系列的候选的处理的另一个示例的流程图,所述处理由图 4 中的状态识别部分 23 在图 44 中的步骤 S305 中执行。

[0905] 在图 45 中计算当前状况状态系列的候选的处理中,用于识别的观察值系列和行为系列的系列长度 q 被固定为短长度 Q' ,并且获得长度 Q' 的用于识别的状态系列或者继而获得长度 Q' 的当前状况状态系列的候选。

[0906] 另一方面,在图 46 中计算当前状况状态系列的候选的处理中,代理自适应(自主)地调整用于识别的观察值系列和行为系列的系列长度 q 。由此,获得具有最长系列长度 q 的状态系列作为当前状况状态系列的候选,在所述状态系列中,在经历的状态系列中观察到更像在通过扩展 HMM 获得的行为环境的结构中的代理的当前位置的结构,即用于识别的观察值系列和行为系列(最新观察值系列和最新行为系列)。

[0907] 在图 46 中计算当前状况状态系列的候选者的处理的步骤 S321 中,状态识别部分 23(图 4)将系列长度 q 初始化为例如最小的 1。处理进行到步骤 S322。

[0908] 在步骤 S322 中,状态识别部分 23 从历史存储部分 14(图 4)中读出其系列长度是长度 q 的最新观察值系列和当观察到所述观察值系列的每个观察值时执行的行为的行为系列作为用于识别的观察值系列和行为系列,由此获得用于识别的观察值系列和行为系列。然后处理进行到步骤 S323。

[0909] 在步骤 S323 中,状态识别部分 23 在模型存储部分 22 中存储的学习过的扩展 HMM 中观察具有系列长度 q 的用于识别的观察值系列和行为系列,并且根据基于维特比算法的上述式 (10) 和 (11) 获得最佳状态概率 $\delta_t(j)$ 和最佳路径 $\psi_t(j)$,最佳状态概率 $\delta_t(j)$ 是在时刻 t 处位于状态 S_j 中的状态概率的最大值,最佳路径 $\psi_t(j)$ 是提供最佳状态概率 $\delta_t(j)$ 的状态系列。

[0910] 此外,状态识别部分 23 观察用于识别的观察值系列和行为系列,并且从式 (11) 中的最佳路径 $\psi_t(j)$ 获得最大可能状态系列,其是用于到达状态 S_j 的状态系列,该状态系列最大化时刻 t 处式 (10) 中的最佳状态概率 $\delta_t(j)$ 。

[0911] 其后处理从步骤 S323 进行到步骤 S324,在步骤 S324 中,状态识别部分 23 根据最大可能状态系列来确定代理的当前状况是已知状况还是未知状况,就像在图 44 中的步骤

S303 中那样。

[0912] 当在步骤 S324 中确定当前状况是已知状况时,即当可以从经历的状态系列中获得其中观察到具有系列长度 q 的用于识别的观察值系列和行为系列(最新观察值系列和最新行为系列)的状态系列时,处理进行到步骤 S325,在步骤 S325 中,状态识别部分 23 将系列长度 q 递增 1。

[0913] 然后处理从步骤 S325 返回到步骤 S322,以重复从步骤 S322 向下的类似处理。

[0914] 当在步骤 324 中确定当前状况是未知状况时,即当不能从经历的状态系列中获得其中观察到具有系列长度 q 的用于识别的观察值系列和行为系列(最新观察值系列和最新行为系列)的状态系列时,处理进行到步骤 S326,在步骤 S326 中,状态识别部分 23 在随后的步骤 S326 到 S328 中获得具有最长系列长度的状态系列,在所述状态系列中,从经历的状态系列中观察到用于识别的观察值系列和行为系列(最新观察值系列和最新行为系列)作为当前状况状态系列的候选。

[0915] 即,在步骤 S322 到 S325 中,根据其中观察到用于识别的观察值系列和行为系列的最大可能状态系列确定代理的当前状况是已知状况还是未知状况,同时将用于识别的观察值系列和行为系列的系列长度 q 递增 1。

[0916] 因此,将最大可能状态系列提供为具有最长系列长度的、在经历的状态系列中观察到用于识别的观察值系列和行为系列的状态系列(状态系列之一),在该最大可能状态系列中,观察到具有系列长度 $q-1$ 的用于识别的观察值系列和行为系列,系列长度 $q-1$ 是紧邻在步骤 S324 中确定当前状况是未知状况后将系列长度 q 递减 1 而获得的。

[0917] 因此,在步骤 S326 中,状态识别部分 23 从历史存储部分 14(图 4)中读出其系列长度是长度 $q-1$ 的最新观察值系列和当观察到所述观察值系列的每个观察值时执行的行为的行为系列作为用于识别的观察值系列和行为系列,并且由此获得用于识别的观察值系列和行为系列。然后处理进行到步骤 S327。

[0918] 在步骤 S327 中,状态识别部分 23 在模型存储部分 22 中存储的学习过的扩展 HMM 中观察在步骤 S326 中获得的具有系列长度 $q-1$ 的用于识别的观察值系列和行为系列,并且根据基于维特比算法的上述式 (10) 和 (11) 获得最佳状态概率 $\delta_t(j)$ 和最佳路径 $\psi_t(j)$,最佳状态概率 $\delta_t(j)$ 是在时刻 t 处位于状态 S_j 中的状态概率的最大值,最佳路径 $\psi_t(j)$ 是提供最佳状态概率 $\delta_t(j)$ 的状态系列。

[0919] 即,状态识别部分 23 从在学习过的扩展 HMM 中发生的状态转移的状态系列中获得最佳路径 $\psi_t(j)$ (用于识别的状态系列),其是具有系列长度 $q-1$ 的其中观察到用于识别的观察值系列和行为系列的状态系列。

[0920] 在步骤 S327 中获得用于识别的状态系列后,处理进行到步骤 S328,在步骤 S328 中,状态识别部分 23 从在步骤 S327 中获得的用于识别的状态系列中选择一个或多个用于识别的状态系列作为当前状况状态系列的候选,就像在图 45 中的步骤 S313 中那样。然后处理返回。

[0921] 如上所述,通过递增系列长度 q 并且获得具有系列长度 $q-1$ 的用于识别的观察值系列和行为系列,可以从经历的状态系列中获得适当的当前状况状态系列的候选(与更早在通过扩展 HMM 获得的行为环境的结构中的代理的当前位置的结构的结构对应的状态系列),所述系列长度 $q-1$ 是通过在紧邻确定当前状况是未知状况后将系列长度 q 递减 1 而获

得的。

[0922] 即,当用于获得当前状况状态系列的候选的用于识别的观察值系列和行为系列的系列长度固定时,如果固定系列长度太短或者太长,则不能获得适当的当前状况状态系列的候选。

[0923] 具体地,当用于识别的观察值系列和行为系列的系列长度太短时,经历的状态系列包括具有观察到这样的观察长度的用于识别的观察值系列和行为系列的高可能性的许多状态系列,从而获得大量具有高可能性的用于识别的状态系列。

[0924] 结果,当从这样的大量具有高可能性的用于识别的状态系列中选择当前状况状态系列的候选时,可以提高在经历的状态系列中更适合地表示当前状况的状态系列未被选择为当前状况状态系列的候选的可能性。

[0925] 另一方面,当用于识别的观察值系列和行为系列的系列长度太长时,经历的状态系列可能不包括具有观察到这样太长的观察长度的用于识别的观察值系列和行为系列的高可能性的状态系列,从而可能提高不能获得当前状况状态系列的候选的可能性。

[0926] 另一方面,如参照图 46 所述,重复作为具有观察到用于识别的观察值系列和行为系列的最高可能性的状态转移的状态系列的估计,并且重复根据最大可能状态系列来确定代理的当前状况是在扩展 HMM 中获得的已知状况还是未在扩展 HMM 中获得的未知状况,直到确定代理的当前状况是未知状况为止,同时递增(增大)用于识别的观察值系列和行为系列的系列长度。估计一个或更多个用于识别的状态系列作为状态转移的状态系列,在所述状态转移中,观察到当确定代理的当前状况是未知状况时具有比系列长度 q 短一个样本的系列长度 $q-1$ 的用于识别的观察值系列和行为系列。从一个或更多个用于识别的状态系列中选择当前状况状态系列的一个或更多个候选。由此,可以将代表更像在由扩展 HMM 获得的行为环境的结构中的代理的当前位置的结构的结构的状态系列获得为当前状况状态系列的候选。

[0927] 结果,可以通过充分利用经历的状态系列来确定行为。

[0928] [根据策略的行为的确定]

[0929] 图 47 是帮助说明根据策略来确定行为的处理的流程图,所述处理由图 4 中的行为确定部分 24 在图 44 的步骤 S306 中执行。

[0930] 在图 47 中,行为确定部分 24 根据第一策略确定行为,以执行由代理在扩展 HMM 中获得的已知状况中类似代理的当前状况的已知状况中执行的行为。

[0931] 具体地,在步骤 S341 中,行为确定部分 24 将来自状态识别部分 23(图 4)的当前状况状态系列的一个或更多个候选的一个候选选择为感兴趣的状态系列,该候选还没有被设置为将注意力转向其的感兴趣的状态系列。

[0932] 在步骤 S342 中,行为确定部分 24 根据模型存储部分 22 中存储的扩展 HMM 获得针对每个行为 U_m 的以感兴趣的状态系列的最后状态作为转移源的状态转移的状态转移概率的和作为行为适合度,所述行为适合度用于指示与感兴趣的状态系列相关地执行行为 U_m (根据第一策略)的适合度。

[0933] 具体地,假定最后状态被表示为 S_I (I 是整数 1 到 N 之一),行为确定部分 24 获得在关于每个行为 U_m 的状态转移概率平面中的 j 轴方向(水平方向)上布置的状态转移概率 $a_{I,1}(U_m)$ 、 $a_{I,2}(U_m)$ 、 $\dots$ 、 $a_{I,N}(U_m)$ 的和作为行为适合度。

[0934] 然后处理从步骤 S342 进行到步骤 S343, 在步骤 S343 中, 行为确定部分 24 将关于在已经获得其行为适合度的 M (种) 行为 U_1 至 U_M 中的其行为适合度低于阈值的 U_m 获得的行为适合度设置为 0.0。

[0935] 即, 通过将关于其行为适合度低于阈值的 U_m 获得的行为适合度设置为 0.0, 行为确定部分 24 从关于感兴趣的状态系列根据第一策略要执行的下一个行为的候选中排除其行为适合度低于阈值的 U_m , 从而选择其行为适合度等于或者高于阈值的 U_m 作为根据第一策略要执行的下一个行为的候选。

[0936] 在步骤 S343 后, 处理进行到步骤 S344, 在步骤 S344 中, 行为确定部分 24 确定是否已经将当前状况状态系列的所有候选设置为感兴趣的状态系列。

[0937] 当在步骤 S344 中确定还没有将当前状况状态系列的所有候选设置为感兴趣的状态系列时, 处理返回到步骤 S341。然后, 在步骤 S341 中, 行为确定部分 24 从来自状态识别部分 23 的当前状况状态系列的一个或多个候选中选择一个还没有被设置为感兴趣的状态系列的候选作为新的感兴趣的状态系列。其后重复类似的处理。

[0938] 当在步骤 S344 中确定当前状况状态系列的所有候选者已经被设置为感兴趣的状态系列时, 处理进行到步骤 S345, 在步骤 S345 中, 行为确定部分 24 根据关于每个行为 U_m 的行为适合度从下一个行为的候选中确定下一个行为, 所述行为适合度是关于来自状态识别部分 23 的当前状况状态系列的一个或多个候选的每一个获得的。

[0939] 即, 行为确定部分 24 例如确定其行为适合度最大的候选为下一个行为。

[0940] 另外, 行为确定部分 24 获得关于每个行为 U_m 的行为适合度的期望值 (平均值), 并且根据期望值来确定下一个行为。

[0941] 具体地, 例如, 关于每个行为 U_m , 行为确定部分 24 获得关于每个行为 U_m 的行为适合度的期望值 (平均值), 所述行为适合度是关于当前状况状态系列的一个或多个候选的每一个获得的。

[0942] 然后行为确定部分 24 根据关于每个行为 U_m 的期望值确定例如其期望值最大的行为 U_m 为下一个行为。

[0943] 或者, 行为确定部分 24 根据关于每个行为 U_m 的期望值通过例如 SoftMax 方法来确定下一个行为。

[0944] 具体地, 行为确定部分 24 以对应于关于以整数 m 为其下标的行为 U_m 的期望值的概率随机生成 M 个行为 U_1 至 U_M 的下标 1 至 M 的范围中的整数 m , 并且确定以所生成的整数 m 为其下标的行为 U_m 是下一个行为。

[0945] 如上所述, 当根据第一策略确定行为时, 代理执行由代理在与代理的当前状况类似的已知状况中执行的行为。

[0946] 因此, 根据第一策略, 当代理处于未知状况中并且期望代理执行与代理在已知状况中采取的行为类似的行为时, 可以使代理执行适当的行为。

[0947] 根据这样的第一策略的行为的确定不仅可以在代理处于未知状况中的情况下进行, 而且可以在例如如上所述代理已经到达开放端之后确定要执行的行为的情况下进行。

[0948] 当代理处于未知状况中并且使代理执行与在已知状况中采取的行为类似的行为时, 代理可能在行为环境中迷路。

[0949] 当代理在行为环境中迷路时, 代理可以返回已知地点 (区域) (当前状况变为已知

状况),或者可以探索未知地点(当前状况继续保持在未知状况)。

[0950] 因此,当期望代理返回已知地点或者代理探索未知地点时,将难以断言使代理在行为环境中迷路的行为作为要由代理执行的行为是适当的。

[0951] 因此,行为确定部分 24 可以不仅根据第一策略而且根据下面的第二策略和第三策略来确定下一个行为。

[0952] 图 48 是帮助说明根据第二策略确定行为的概要的图。

[0953] 第二策略是用于增加使得能够识别代理的(当前)状况的信息的策略。通过根据第二策略确定行为,可以将适当的行为确定为代理要返回到已知地点的行为。结果,代理可以高效地返回到已知地点。

[0954] 具体地,在根据第二策略确定行为中,行为确定部分 24 将下述行为确定为下一个行为:所述行为进行从来自状态识别部分 23 的当前状况状态系列的一个或更多个候选的最后状态 s_t 向作为紧邻最后状态 s_t 在前的状态的紧邻在前的状态 s_{t-1} 的状态转移,例如如图 48 中所示。

[0955] 图 49 是帮助说明一种根据第二策略来确定行为的处理的流程图,所述处理由图 4 中的行为确定部分 24 在图 44 中的步骤 S306 中执行。

[0956] 在步骤 S351 中,行为确定部分 24 选择来自状态识别部分 23(图 4)的当前状况状态系列的一个或更多个候选的一个候选作为感兴趣的状态系列,该候选还没有被设置为注意力将要转向其的感兴趣的状态系列。然后处理进行到步骤 S352。

[0957] 在这种情况下,当来自状态识别部分 23 的当前状况状态系列的候选的系列长度是 1,从而没有紧邻最后状态在前的紧邻在前状态时,行为确定部分 24 在执行步骤 S351 的处理之前参照模型存储部分 22 中存储的扩展 HMM(扩展 HMM 的状态转移概率),并且获得一个状态作为转移目的地,可以从这个状态向来自状态识别部分 23 的当前状况状态系列的一个或更多个候选的每一个的最后状态进行状态转移。

[0958] 然后行为确定部分 24 将通过布置所述状态而获得的状态系列作为转移目的地,并且将最后状态作为当前状况状态系列的候选,可以从这个被布置的状态向来自状态识别部分 23 的当前状况状态系列的一个或更多个候选的每个的最后状态进行状态转移。这对于后面描述的图 51 同样成立。

[0959] 在步骤 S352 中,行为确定部分 24 根据模型存储部分 22 中存储的扩展 HMM 获得针对每个行为 U_m 从感兴趣的状态系列的最后状态向最后状态紧邻在前的紧邻在前状态的状态转移的状态转移概率作为行为适合度,所述行为适合度用于指示与感兴趣的状态系列相关地执行行为 U_m (根据第二策略)的适合度。

[0960] 具体地,行为确定部分 24 获得从最后状态 S_i 向紧邻在前状态 S_j 的状态转移的状态转移概率 $a_{ij}(U_m)$ 作为关于行为 U_m 的行为适合度,所述状态转移是当执行行为 U_m 时进行的。

[0961] 然后处理从步骤 S352 进行到步骤 S353,在步骤 S353 中,行为确定部分 24 将关于在 M (种)行为 U_1 至 U_M 中的、除了其行为适合度是最大值的行为之外的行为获得的行为适合度设置为 0.0。

[0962] 即,通过将关于除了其行为适合度是最大值的行为之外的行为获得的行为适合度设置为 0.0,从而行为确定部分 24 选择其行为适合度最大的行为作为要根据第二策略与感

兴趣的状态系列相关地执行的下一个行为的候选。

[0963] 在步骤 S353 后,处理进行到步骤 S354,在步骤 S354 中,行为确定部分 24 确定是否已经将当前状况状态系列的所有候选设置为感兴趣的状态系列。

[0964] 当在步骤 S354 中确定还没有将当前状况状态系列的所有候选设置为感兴趣的状态系列时,处理返回到步骤 S351。然后,在步骤 S351 中,行为确定部分 24 从来自状态识别部分 23 的当前状况状态系列的一个或更多个候选中将选择一个还没有被设置为感兴趣的状态系列的候选作为新的感兴趣的状态系列。其后重复类似的处理。

[0965] 当在步骤 S354 中确定已经将当前状况状态系列的所有候选者设置为感兴趣的状态系列时,处理进行到步骤 S355,在步骤 S355 中,行为确定部分 24 根据关于每个行为 U_m 的行为适合度从下一个行为的候选中确定下一个行为,就像图 47 中的步骤 S345 中那样,所述行为适合度是关于来自状态识别部分 23 的当前状况状态系列的一个或更多个候选的每一个获得的。然后处理返回。

[0966] 当如上所述根据第二策略确定行为时,代理执行从折回代理采取的路径开始的行为。结果,增加了使能识别代理的状况的信息(观察值)。

[0967] 因此,根据第二策略,当代理在未知状况下时可以使代理执行适当的行为,并且期望代理执行返回已知地点的行为。

[0968] 图 50 是帮助说明根据第三策略确定行为的概要的图。

[0969] 第三策略是用于增加在扩展 HMM 中未获得的未知状况的信息的策略。通过根据第三策略来确定行为,可以将适当的行为确定为代理探索未知地点的行为。结果,代理可以高效地探索未知地点。

[0970] 具体地,在根据第三策略确定行为中,行为确定部分 24 将下述行为确定为下一个行为:所述行为进行除了从来自状态识别部分 23 的当前状况状态系列的一个或更多个候选的最后状态 s_t 向作为紧邻最后状态 s_t 在前的状态的紧邻在前的状态 s_{t-1} 的状态转移之外的状态转移,例如如图 50 中所示。

[0971] 图 51 是帮助说明根据第三策略确定行为的处理的流程图,所述处理由图 4 中的行为确定部分 24 在图 44 中的步骤 S306 中执行。

[0972] 在步骤 S361 中,行为确定部分 24 选择来自状态识别部分 23(图 4)的当前状况状态系列的一个或更多个候选的一个候选作为感兴趣的状态系列,这个候选还没有被设置为注意力要转向其的感兴趣的状态系列。然后处理进行到步骤 S362。

[0973] 在步骤 S362 中,行为确定部分 24 根据模型存储部分 22 中存储的扩展 HMM 获得针对每个行为 U_m 从感兴趣的状态系列的最后状态向最后状态紧邻在前的紧邻在前状态的状态转移的状态转移概率来作为行为适合度,所述行为适合度用于指示与感兴趣的状态系列相关地执行行为 U_m (根据第三策略)的适合度。

[0974] 具体地,行为确定部分 24 获得从最后状态 S_i 向紧邻在前状态 S_j 的状态转移的状态转移概率 $a_{ij}(U_m)$ 作为关于行为 U_m 的行为适合度,所述状态转移是当执行行为 U_m 时进行的。

[0975] 然后处理从步骤 S362 进行到步骤 S363,在步骤 S363 中,行为确定部分 24 将在 M (种)行为 U_1 至 U_M 中其行为适合度是最大值的行为检测为关于感兴趣的状态系列实现将状态返回至紧邻在前的状态的状态转移的行为(以下也将所述行为称为返回行为)。

[0976] 在步骤 S363 后,处理进行到步骤 S364,在步骤 S364 中,行为确定部分 24 确定是否已经将当前状况状态系列的所有候选设置为感兴趣的状态系列。

[0977] 当在步骤 S364 中确定还没有将当前状况状态系列的所有候选设置为感兴趣的状态系列时,处理返回到步骤 S361。然后,在步骤 S361 中,行为确定部分 24 从来自状态识别部分 23 的当前状况状态系列的一个或更多个候选中选择一个还没有被设置为感兴趣的状态系列的候选作为新的感兴趣的状态系列。其后重复类似的处理。

[0978] 当在步骤 S364 中确定已经将当前状况状态系列的所有候选设置为感兴趣的状态系列时,行为确定部分 24 复位将当前状况状态系列的所有候选作为感兴趣的状态系列的选择。然后处理进行到步骤 S365。

[0979] 在步骤 S365 中,像在步骤 S361 中那样,行为确定部分 24 选择来自状态识别部分 23 的当前状况状态系列的一个或更多个候选的一个候选作为感兴趣的状态系列,所述候选还没有被设置为注意力要转向其的感兴趣的状态系列。然后处理进行到步骤 S366。

[0980] 在步骤 S366 中,像在图 47 中的步骤 S342 中那样,行为确定部分 24 根据模型存储部分 22 中存储的扩展 HMM 获得针对每个行为 U_m 以感兴趣的状态系列的最后状态作为转移源的状态转移的状态转移概率的和来作为行为适合度,所述行为适合度用于指示与感兴趣的状态系列相关地执行行为 U_m (根据第三策略) 的适合度。

[0981] 其后处理从步骤 S366 进行到步骤 S367,在步骤 S367 中,行为确定部分 24 将关于在已经获得其行为适合度的 M 个(种)行为 U_1 到 U_m 中的、其行为适合度低于阈值的行为 U_m 而获得的行为适合度和关于返回行为获得的行为适合度设置为 0.0。

[0982] 即,通过将关于其行为适合度低于阈值的行为 U_m 获得的行为适合度设置为 0.0,从而行为确定部分 24 选择其行为适合度等于或者高于阈值的行为 U_m 作为要关于感兴趣的状态系列根据第三策略执行的下一个行为的候选。

[0983] 此外,通过将关于在与感兴趣的状态系列相关地选择的、其行为适合度等于或者高于阈值的行为 U_m 中的返回行为获得的行为适合度设置为 0.0,从而行为确定部分 24 选择除了返回行为之外的行为作为要与感兴趣的状态系列相关地根据第三策略执行的下一个行为的候选。

[0984] 在步骤 S367 后,处理进行到步骤 S368,在步骤 S368 中,行为确定部分 24 确定是否已经将当前状况状态系列的所有候选设置为感兴趣的状态系列。

[0985] 当在步骤 S368 中确定还没有将当前状况状态系列的所有候选设置为感兴趣的状态系列时,处理返回到步骤 S365。然后,在步骤 S365 中,行为确定部分 24 从来自状态识别部分 23 的当前状况状态系列的一个或更多个候选中选择一个还没有被设置为感兴趣的状态系列的候选作为新的感兴趣的状态系列。其后重复类似的处理。

[0986] 当在步骤 S368 中确定已经将当前状况状态系列的所有候选设置为感兴趣的状态系列时,处理进行到步骤 S369,在步骤 S369 中,行为确定部分 24 根据关于每个行为 U_m 的行为适合度从下一个行为的候选中确定下一个行为,就像图 47 中的步骤 S345 中那样,所述行为适合度是关于来自状态识别部分 23 的当前状况状态系列的一个或更多个候选的每一个获得的。然后处理返回。

[0987] 当如上所述根据第三策略确定行为时,代理执行除了返回行为之外的行为,即探索未知地点的行为。结果,增加了关于在扩展 HMM 中未获得的未知状况的信息。

[0988] 因此,根据第三策略,当代理处于未知状况时,可以使代理执行适当的行为,并且期望代理探索未知地点。

[0989] 如上所述,代理根据扩展 HMM 将当前状况状态系列的候选计算为使代理到达当前状况的状态系列,并且使用当前状况状态系列的候选根据预定策略确定接下来要由代理执行的行为。由此即使当代理未被提供有要执行的行为的诸如用于计算代理的报酬的报酬函数等的度量时,代理也可以根据在扩展 HMM 中获得的经历来确定行为。

[0990] 顺便提及,作为用于消除状况的模糊的行为确定方法,例如,日本待审专利公开第 2008-186326 号描述了一种通过一个报酬函数来确定行为的方法。

[0991] 图 44 中的识别行为模式中的处理与日本待审专利公开第 2008-186326 号的行为确定方法的不同之处例如在于:根据扩展 HMM 来计算作为使代理到达当前状况的状态系列的当前状况状态系列的候选,并且使用当前状况状态系列的候选确定行为;可以在已经由代理经历过的状态系列中获得具有最长系列长度 q 的、其中观察到用于识别的观察值系列和行为系列的状态系列作为当前状况状态系列的候选(图 46);以及,可以改变(从多个策略中选择)在行为确定时遵循的策略,如下面将要描述的。

[0992] 如上所述,第二策略增加了使得能够识别代理的状况的信息,并且第三策略增加了关于在扩展 HMM 中未获得的未知状况的信息。因此第二和第三策略是增加某些信息的策略。

[0993] 除了参照图 48 至 51 所述的方法之外,还可以如下地进行根据第二和第三策略从而增加某些信息的行为的确定。

[0994] 通过式 (30) 来表达当代理在某一时刻 t 执行行为 U_m 时观察到观察值 O 的概率 $P_m(O)$ 。

$$[0995] \quad P_m(O) = \sum_{i=1}^N \sum_{j=1}^N \rho_i a_{ij}(U_m) b_j(O) \quad \dots(30)$$

[0996] 顺便提及, ρ_i 表示在时刻 t 处位于状态 S_i 中的状态概率。

[0997] 现在假定其发生概率由概率 $P_m(O)$ 表达的信息量被表示为 $I(P_m(O))$, 通过式 (31) 来表达当根据增加某些信息的策略来确定行为时行为 U_m 的下标 m' 。

$$[0998] \quad m' = \arg \max_m \{I(P_m(O))\} \quad \dots(31)$$

[0999] 式 (31) 中的 $\arg \max \{I(P_m(O))\}$ 表示在行为 U_m 的下标 m 中最大化括号中的信息量 $I(P_m(O))$ 的下标 m' 。

[1000] 现在假定使得能够识别代理的状况的信息(以下也将所述信息称为识别使能信息)被用作信息,则根据式 (31) 来确定行为 U_m , 是根据增加识别使能信息的第二策略来确定行为。

[1001] 另外,假定关于在扩展 HMM 中未获得的未知状况的信息(以下也将所述信息称为未知状况信息)被用作信息,则根据式 (31) 来确定行为 U_m , 是根据增加未知状况信息的第三策略来确定行为。

[1002] 假定其发生概率由概率 $P_m(O)$ 表达的信息的熵被表示为 $H^0(P_m)$, 则可以通过下式来等价地表达式 (31)。

[1003] 即,可以通过式 (32) 表达熵 $H^0(P_m)$ 。

$$[1004] \quad H^0(P_m) = \sum_{O=O_1}^{O_K} (-P_m(O) \times \log(P_m(O))) \quad \dots(32)$$

[1005] 当式 (32) 的熵 $H^0(P_m)$ 较大时, 观察到相应观察值 O 的概率 $P_m(O)$ 是均匀的。因此, 增大了使得不知道观察到哪个观察值或者继而不知道代理在哪里的模糊性, 并且可以说增大了获得代理未知的信息或者关于未知世界的信息的可能性。

[1006] 因此, 由于通过使熵 $H^0(P_m)$ 大而增加了未知状况信息, 所以当根据增加未知状况信息的第三策略来确定行为时的式 (31) 可以由最大化熵 $H^0(P_m)$ 的式 (33) 等价地表达。

$$[1007] \quad m' = \arg \max_m \{H^0(P_m)\} \quad \dots(33)$$

[1008] 式 (33) 中的 $\arg \max \{H^0(P_m)\}$ 表示在行为 U_m 的下标 m 中最大化括号中的熵 $H^0(P_m)$ 的下标 m' 。

[1009] 另一方面, 当式 (32) 的熵 $H^0(P_m)$ 较小时, 观察到相应观察值 O 的概率 $P_m(O)$ 仅在特定观察值处为高。因此, 消除了使得不知道观察到哪个观察值或者继而不知道代理在哪里的模糊性, 并且容易确定代理的位置。

[1010] 因此, 由于通过使熵 $H^0(P_m)$ 小而增加了识别使能信息, 所以当根据增大识别使能信息的第二策略来确定行为时的式 (31) 可以由最小化熵 $H^0(P_m)$ 的式 (34) 等价地表达。

$$[1011] \quad m' = \arg \min_m \{H^0(P_m)\} \quad \dots(34)$$

[1012] 式 (34) 中的 $\arg \min \{H^0(P_m)\}$ 表示在行为 U_m 的下标 m 中最小化括号中的熵 $H^0(P_m)$ 的下标 m' 。

[1013] 顺便提及, 另外, 根据概率 $P_m(O)$ 的最大值和阈值之间在幅度上的关系, 例如可以将最大化概率 $P_m(O)$ 的行为 U_m 确定为下一个行为。

[1014] 当概率 $P_m(O)$ 的最大值大于阈值 (等于或者大于阈值) 时将最大化概率 $P_m(O)$ 的行为 U_m 确定为下一个行为是为了消除模糊性而确定行为, 即根据第二策略来确定行为。

[1015] 另一方面, 当概率 $P_m(O)$ 的最大值等于或者小于阈值 (小于阈值) 时将最大化概率 $P_m(O)$ 的行为 U_m 确定为下一个行为是为了提高模糊性而确定行为, 即根据第三策略来确定行为。

[1016] 以上, 使用当代理在某一时刻 t 处执行行为 U_m 时观察到观察值的概率 $P_m(O)$ 来确定行为。另外, 可以使用当代理在某一时刻 t 处执行行为 U_m 时从状态向状态 S_j 进行状态转移的式 (35) 中的概率 P_{mj} 来确定行为。

$$[1017] \quad P_{mj} = \sum_{i=1}^N \rho_i a_{ij}(U_m) \quad \dots(35)$$

[1018] 具体地, 通过式 (36) 表达当根据增大通过概率 P_{mj} 来表示其发生概率的信息量 $I(P_{mj})$ 的策略确定行为时的行为 $U_{m'}$ 的下标 m' 。

$$[1019] \quad m' = \arg \max_m \{I(P_{mj})\} \quad \dots(36)$$

[1020] 式 (36) 中的 $\arg \max \{I(P_{mj})\}$ 表示行为 U_m 的下标 m 中最大化括号中的信息量 $I(P_{mj})$ 的下标 m' 。

[1021] 现在假定将识别使能信息用作信息, 则根据式 (36) 来确定行为 $U_{m'}$ 是根据增大识别使能信息的第二策略来确定行为。

[1022] 另外, 假定将未知状况信息用作信息, 则根据式 (36) 来确定行为 $U_{m'}$ 是根据增大

未知状况信息的第三策略来确定行为。

[1023] 假定通过概率 P_{mj} 来表达其发生概率的信息的熵被表示为 $H^j(P_m)$, 式 (36) 可以等价地由下式表达。

[1024] 即, 可以通过式 (37) 来表达熵 $H^j(P_m)$ 。

$$[1025] \quad H^j(P_m) = \sum_{j=1}^N (-P_{mj} \times \log(P_{mj})) \quad \dots(37)$$

[1026] 当式 (37) 的熵 $H^j(P_m)$ 较大时, 从状态 S_i 向状态 S_j 进行相应的状态转移的概率 P_{mj} 是均匀的。因此, 增大了使得不知道进行哪个状态转移或者继而不知道代理在哪里的模糊性, 并且增大了获得关于代理未知的未知世界的信息的可能性。

[1027] 因此, 由于通过使熵 $H^j(P_m)$ 大而增大了未知状况信息, 所以当根据增加未知状况信息的第三策略来确定行为时的式 (36) 可以由最大化熵 $H^j(P_m)$ 的式 (38) 等价地表达。

$$[1028] \quad m' = \arg \max_m \{H^j(P_m)\} \quad \dots(38)$$

[1029] 式 (38) 中的 $\arg \max \{H^j(P_m)\}$ 表示在行为 U_m 的下标 m 中最大化括号中的熵 $H^j(P_m)$ 的下标 m' 。

[1030] 另一方面, 当式 (37) 的熵 $H^j(P_m)$ 较小时, 从状态 S_i 向状态 S_j 进行相应状态转移的概率 P_{mj} 仅在特定状态转移时为高。因此, 消除了使得不知道观察到哪个观察值或者继而不知道代理在哪里的模糊性, 并且容易确定代理的位置。

[1031] 因此, 由于通过使熵 $H^j(P_m)$ 小而增大了识别使能信息, 所以当根据增大识别使能信息的第二策略来确定行为时的式 (36) 可以由最小化熵 $H^j(P_m)$ 的式 (39) 等价地表达。

$$[1032] \quad m' = \arg \min_m \{H^j(P_m)\} \quad \dots(39)$$

[1033] 式 (39) 中的 $\arg \min \{H^j(P_m)\}$ 表示在行为 U_m 的下标 m 中最小化括号中的熵 $H^j(P_m)$ 的下标 m' 。

[1034] 顺便提及, 另外, 根据概率 P_{mj} 的最大值和阈值之间在幅度上的关系, 例如可以将最大化概率 P_{mj} 的行为 U_m 确定为下一个行为。

[1035] 当概率 P_{mj} 的最大值大于阈值 (等于或者大于阈值) 时将最大化概率 P_{mj} 的行为 U_m 确定为下一个行为是为了消除模糊性而确定行为, 即根据第二策略来确定行为。

[1036] 另一方面, 当概率 P_{mj} 的最大值等于或者小于阈值 (小于阈值) 时将最大化概率 P_{mj} 的行为 U_m 确定为下一个行为是为了提高模糊性而确定行为, 即根据第三策略来确定行为。

[1037] 另外, 可以使用当观察到观察值 O 时处于状态 S_x 中的后验概率 $P(X|O)$ 来进行使得消除模糊性的行为确定, 即根据第二策略的行为确定。

[1038] 即, 通过式 (40) 来表达后验概率 $P(X|O)$ 。

$$[1039] \quad P(X|O) = \frac{P(X,O)}{P(O)} = \frac{\sum_{i=1}^N \rho_i a_{ix}(U_m) b_x(O)}{\sum_{i=1}^N \sum_{j=1}^N \rho_i a_{ij}(U_m) b_j(O)} \quad \dots(40)$$

[1040] 假定后验概率 $P(X|O)$ 的熵被表示为 $H(P(X|O))$, 可以通过确定使得减小熵 $H(P(X|O))$ 的行为来进行根据第二策略的行为确定。

[1041] 即, 可以通过根据式 (41) 确定行为 U_m 来进行根据第二策略的行为确定。

$$[1042] \quad m' = \arg \min_m \left\{ \sum_{O=O_1}^{O_k} P(O) H(P(X|O)) \right\} \quad \dots(41)$$

[1043] 式 (41) 中的 $\arg \min \{ \}$ 表示在行为 U_m 的下标 m 中最小化括号中的值的下标 m' 。

[1044] 式 (41) 中的 $\arg \min \{ \}$ 的括号中的 $\sum P(O) H(P(X|O))$ 是观察到观察值 O 的概率 $P(O)$ 和当观察到观察值 O 时处于状态 S_x 中的后验概率 $P(X|O)$ 的熵 $H(P(X|O))$ 的乘积的总和, 其中观察值 O 从观察值 O_1 改变到观察值 O_k , 并且 $\sum P(O) H(P(X|O))$ 表示当执行行为 U_m 时观察到观察值 O_1 到 O_k 的整个熵。

[1045] 根据式 (41), 最小化熵 $\sum P(O) H(P(X|O))$ 的行为, 即提高唯一地确定观察值 O 的可能性的行为被确定为下一个行为。

[1046] 因此, 根据式 (41) 确定行为是为了消除模糊性而确定行为, 即根据第二策略来确定行为。

[1047] 另外, 可以进行使得提高模糊性的行为确定, 即根据第三策略的行为确定, 以便最大化减少量, 假定所述减少量是未知状况信息的量, 所述减少量表示下述量: 后验概率 $P(X|O)$ 的熵 $H(P(X|O))$ 相对于处于状态 S_x 中的先验概率 $P(X)$ 的熵 $H(P(X))$ 降低这个量。

[1048] 即通过式 (42) 来表达先验概率 $P(X)$ 。

$$[1049] \quad P(X) = \sum_{i=1}^N \rho_i a_{ix}(U_m) \quad \dots(42)$$

[1050] 可以根据式 (43) 来确定最大化后验概率 $P(X|O)$ 的熵 $H(P(X|O))$ 相对于处于状态 S_x 中的先验概率 $P(X)$ 的熵 $H(P(X))$ 降低的量的行为 $U_{m'}$ 。

$$[1051] \quad m' = \arg \max_m \left\{ \sum_{O=O_1}^{O_k} P(O) (H(P(X)) - H(P(X|O))) \right\} \quad \dots(43)$$

[1052] 式 (43) 中的 $\arg \max \{ \}$ 表示行为 U_m 的下标 m 中最大化括号中的值的下标 m' 。

[1053] 根据式 (43), 将通过将观察到观察值 O 的概率 $P(O)$ 乘以先验概率 $P(X)$ 的熵 $H(P(X))$ 和当执行行为 U_m 时处于状态 S_x 中并且观察到观察值 O 的后验概率 $P(X|O)$ 的熵 $H(P(X|O))$ 之间的差 $H(P(X)) - H(P(X|O))$ 而获得的乘积值 $P(O) (H(P(X)) - H(P(X|O)))$ 的总和 $\sum P(O) (H(P(X)) - H(P(X|O)))$ 作为通过执行行为 U_m 而增加的未知状况信息的量, 其中观察值 O 从观察值 O_1 改变到观察值 O_k , 先验概率 $P(X)$ 是当观察值 O 未知时处于状态 S_x 中的状态概率。最大化未知状况信息的量的行为被确定为下一个行为。

[1054] [策略的选择]

[1055] 代理可以根据参考图 47 至 51 所述的第一到第三策略来确定行为。在确定行为时要遵循的策略不仅可以预先设置, 而且可以从作为多个策略的第一到第三策略中自适应地选择。

[1056] 图 52 是帮助说明当代理根据多个策略确定行为时选择要遵循的策略的处理的流程图。

[1057] 根据第二策略, 确定使得增加识别使能信息并且消除模糊性的行为, 即使代理返回已知地点 (区域) 的行为。

[1058] 另一方面, 根据第三策略, 确定使得增加未知状况信息并且提高模糊性的行为, 即使代理探索未知地点的行为。

[1059] 顺便提及, 根据第一策略, 不知道代理返回已知地点还是探索未知地点, 但是执行

由代理在与代理的当前状况类似的已知状况中执行的行为。

[1060] 为了广泛地获得行为环境的结构,即可以说是为了增加代理的知识(已知世界),必须确定使得代理探索未知地点的行为。

[1061] 另一方面,为了代理获得未知地点作为已知地点,代理需要从未知地点返回到已知地点,并且学习扩展 HMM(追加学习),以将未知地点与已知地点连接。因此,为了代理获得未知地点作为已知地点,必须确定使代理返回到已知地点的行为。

[1062] 通过以良好平衡(well-balanced)的方式确定使代理探索未知地点的行为并且确定使代理返回到已知地点的行为,可以高效地在扩展 HMM 中建模行为环境的整个结构。

[1063] 因此,代理可以根据从代理的状况变为未知状况起的过去时间从第二和第三策略中选择确定行为时要遵循的策略,如图 52 中所示。

[1064] 在步骤 S381 中,行为确定部分 24(图 4)根据状态识别部分 23 中的当前状况的识别结果来获得从未知状况起的过去时间(以下也将所述过去时间称为未知状况过去时间)。然后处理进行到步骤 S382。

[1065] 在这种情况下,未知状况过去时间是在状态识别部分 23 中连续获得当前状况是未知状况的识别结果的次数。当获得当前状况是已知状况的识别结果时,未知状况过去时间被复位为 0。因此,当当前状况不是未知状况时(当当前状况是已知状况时),未知状况过去时间是 0。

[1066] 在步骤 S382 中,行为确定部分 24 确定未知状况过去时间是否大于预定阈值。

[1067] 当在步骤 S382 中确定未知状况过去时间不大于预定阈值时,即当已经过去了代理的状况是未知状况的不太多的时间时,处理进行到步骤 S383,在步骤 S383 中,行为确定部分 24 从第二和第三策略中选择用于增加未知状况信息的第三策略作为在确定行为时要遵循的策略。然后处理返回到步骤 S381。

[1068] 当在步骤 S382 中确定未知状况过去时间大于预定阈值时,即当已经过去了代理的状况是未知状况的相当多的时间时,处理进行到步骤 S384,在步骤 S384 中,行为确定部分 24 从第二和第三策略中选择用于增加识别使能信息的第二策略作为在确定行为时要遵循的策略。然后处理返回到步骤 S381。

[1069] 在图 52 中,根据从代理的状况变为未知状况起的过去时间来选择在确定行为时要遵循的策略。另外,可以根据例如已知状况的时间或者未知状况的时间与预定邻近时间的比率来选择在确定行为时要遵循的策略。

[1070] 图 53 是帮助说明当根据已知状况的时间或者未知状况的时间与预定邻近时间的比率确定行为时选择要遵循的策略的处理的流程图。

[1071] 在步骤 S391 中,行为确定部分 24(图 4)从状态识别部分 23 获得针对预定邻近时间的状况的识别结果,并且根据识别结果计算状况是未知状况的比率(以下也将所述比率称为未知比率)。然后处理进行到步骤 S392。

[1072] 在步骤 S392 中,行为确定部分 24 确定未知比率是否大于预定阈值。

[1073] 当在步骤 S392 中确定未知比率不大于预定阈值时,即当代理的状况是未知状况的比率不很高时,处理进行到步骤 S393,在步骤 S393 中,行为确定部分 24 从第二和第三策略中选择用于增加未知状况信息的第三策略作为在确定行为时要遵循的策略。然后处理进行到步骤 S391。

[1074] 当在步骤 S392 中确定未知比率大于预定阈值时,即当代理的状况是未知状况的比率很高时,处理进行到步骤 S394,在步骤 S394 中,行为确定部分 24 从第二和第三策略中选择用于增加识别使能信息的第二策略作为在确定行为时要遵循的策略。然后处理返回到步骤 S391。

[1075] 顺便提及,虽然在图 53 中根据针对预定邻近时间的状况的识别结果中状况是未知状况的比率(未知比率)来选择策略,但是可以根据针对预定邻近时间的状况的识别结果中状况是已知状况的比率(以下也将这个比率称为已知比率)来选择策略。

[1076] 在根据已知比率来进行策略选择的情况下,当已知比率高于阈值时选择第三策略作为在确定行为时的策略,或者当已知比率不高于阈值时选择第二策略作为在确定行为时的策略。

[1077] 另外,在图 52 的步骤 S383 中和图 53 的步骤 S393 中,例如可以以一次对多次的比率选择第一策略取代第三策略作为在确定行为时的策略。

[1078] 通过如此选择策略,可以在扩展 HMM 中高效地建模行为环境的整个结构。

[1079] [应用本发明的计算机的描述]

[1080] 接着,上述系列处理不仅可以通过硬件,而且通过软件执行。当通过软件执行所述系列处理时,构成软件的程序安装到通用计算机等上。

[1081] 图 54 示出了其上安装了用于执行上述系列处理的程序的计算机的实施例的配置的示例。

[1082] 程序可以预先记录在作为计算机中包括的记录介质的硬盘 105 或者 ROM 103 上。

[1083] 或者,程序可以存储(记录)在可移除记录介质 111 上。这样的可移除记录介质 111 可以被提供作为所谓的套装软件。在这种情况下,可移除记录介质 111 包括例如软盘、CD-ROM(致密盘只读存储器)、MO(磁光盘)、DVD(数字通用盘)、磁盘和半导体存储器。

[1084] 除了如上所述从可移除记录介质 111 安装到计算机上之外,程序还可以经由通信网络或者广播网络下载到计算机,并且安装到内置的硬盘 105 上。具体地,程序可以从下载站点例如通过经由用于数字卫星广播的人造卫星的无线电传送到计算机,或者通过经由诸如局域网和互联网等网络的连线传送到计算机。

[1085] 计算机包括 CPU(中央处理单元)102。CPU 102 经由总线 101 与输入输出接口 110 连接。

[1086] 当 CPU 102 经由输入输出接口 110 例如通过用户对输入部分 107 的操作而被提供命令时,CPU 102 根据命令运行存储在 ROM(只读存储器)103 中的程序。或者,CPU 102 将硬盘 105 中存储的程序加载到 RAM(随机存取存储器)104 中,然后运行该程序。

[1087] 由此 CPU 102 执行根据上述流程图的处理或者通过上述方框图的配置执行的处理。根据需要,然后 CPU 102 例如从输出部分 106 输出处理结果、从通信部分 108 发送结果、或者经由输入输出接口 110 将结果记录在硬盘 105 上。

[1088] 顺便提及,输入部分 107 包括键盘、鼠标和麦克风等。输出部分 106 包括 LCD(液晶显示器)和扬声器等。

[1089] 在本说明书中,由计算机根据程序执行的处理不一定必须以被描述为流程图的顺序按照时序来执行。即,由计算机根据程序执行的处理也包括并行或者单独地执行的处理(例如并行处理或者基于对象的处理)

[1090] 另外,程序可以由一个计算机(处理器)处理,或者可以进行由多个计算机执行的分布式处理。此外,程序可以被传送到远程计算机并且由远程计算机运行。

[1091] 应当注意,本发明的实施例不限于上述实施例,并且在不背离本发明的精神的情况下,可以进行各种改变。

[1092] 本申请包含与在 2009 年 6 月 11 日提交于日本专利局的日本在先专利申请 JP 2009-140064 中公开的主题相关的主题,所述日本在先专利申请 JP 2009-140064 的全部内容通过引用合并于此。

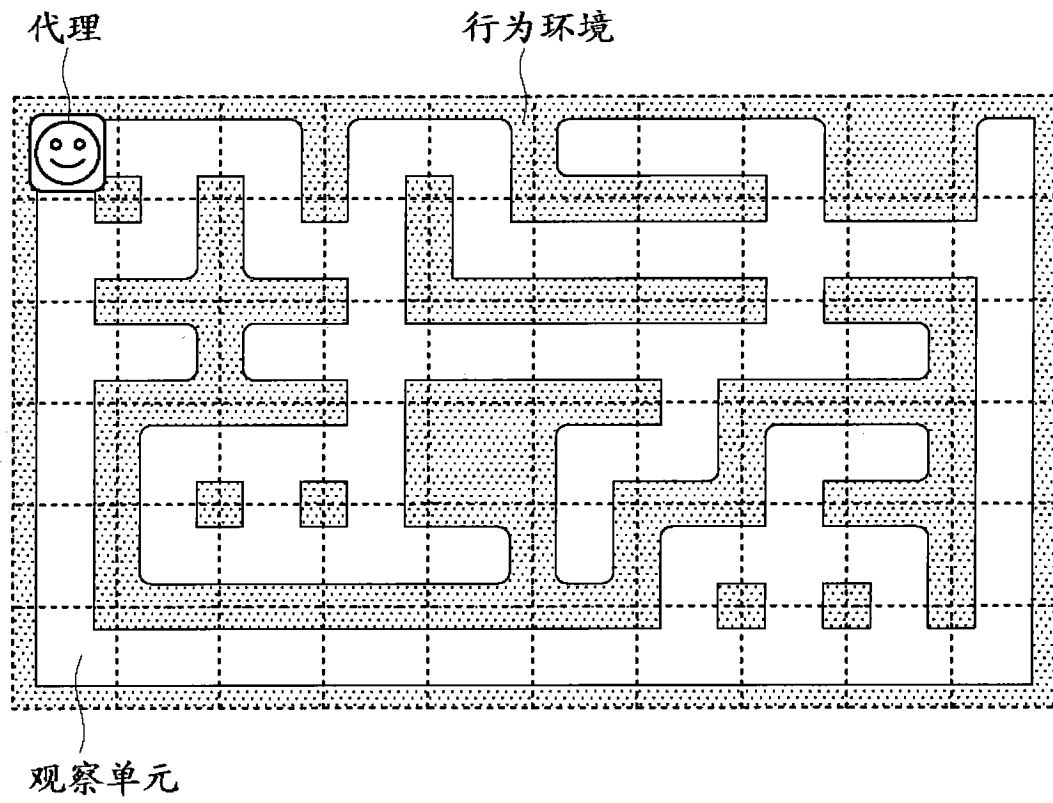


图 1

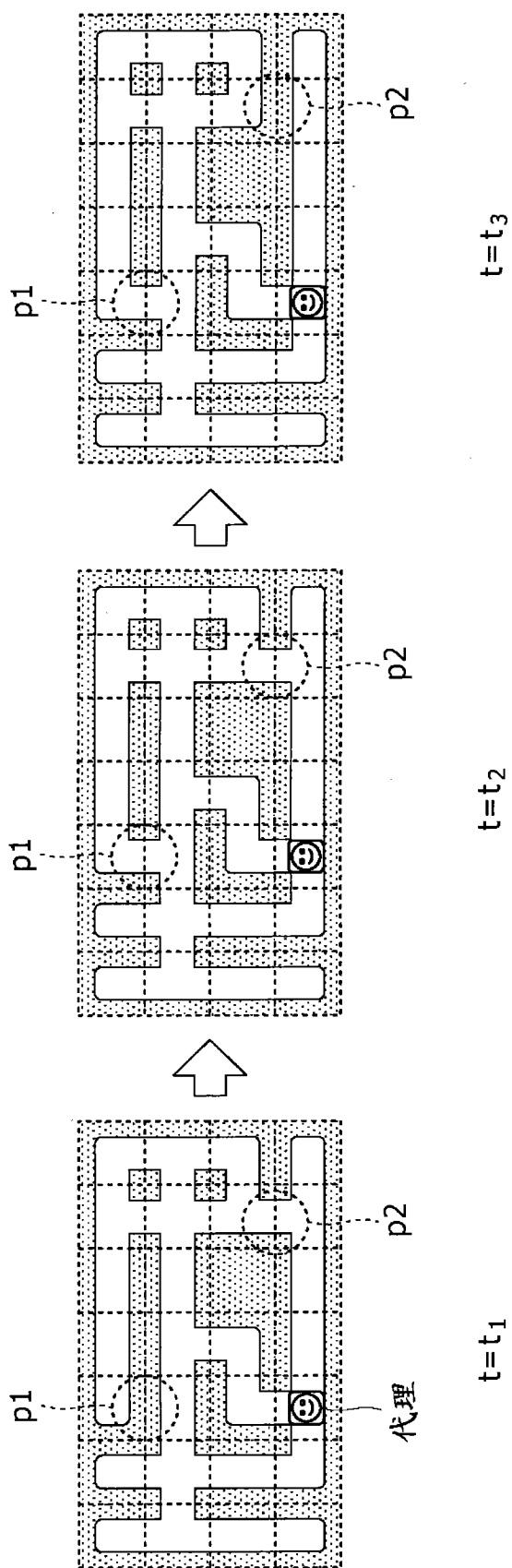


图 2

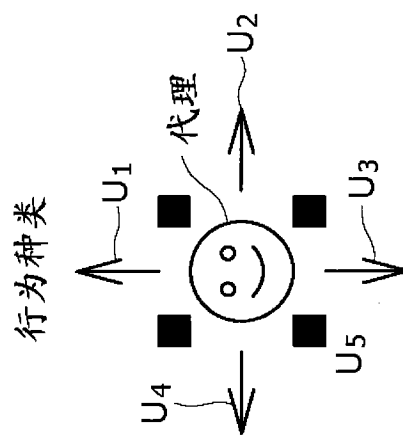


图 3A

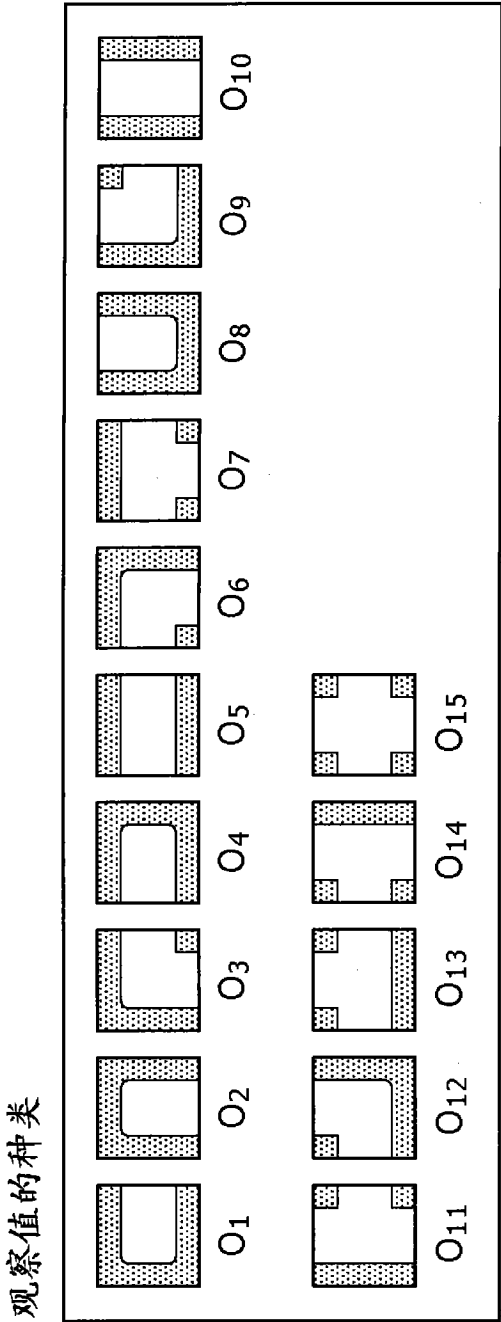


图 3B

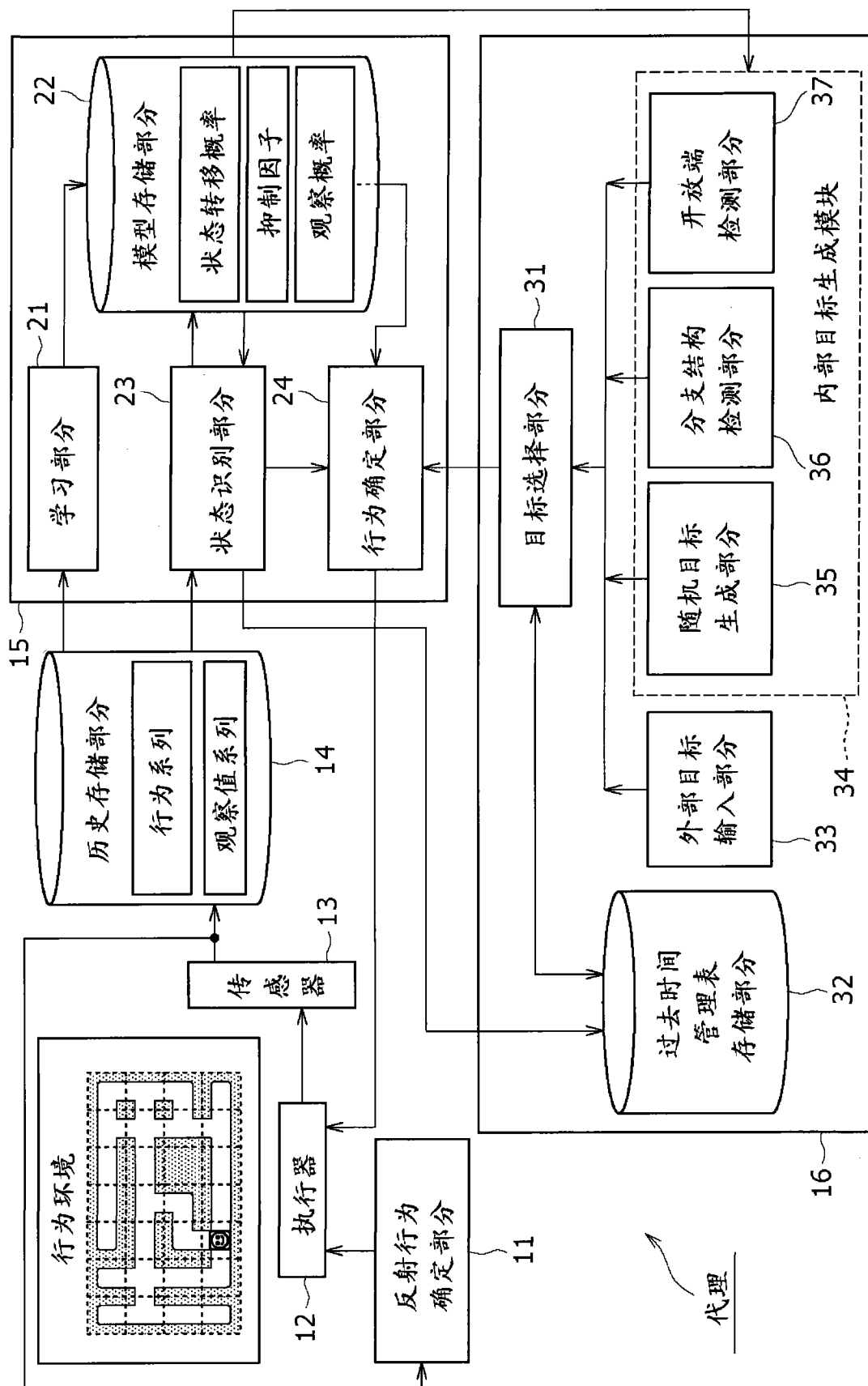


图 4

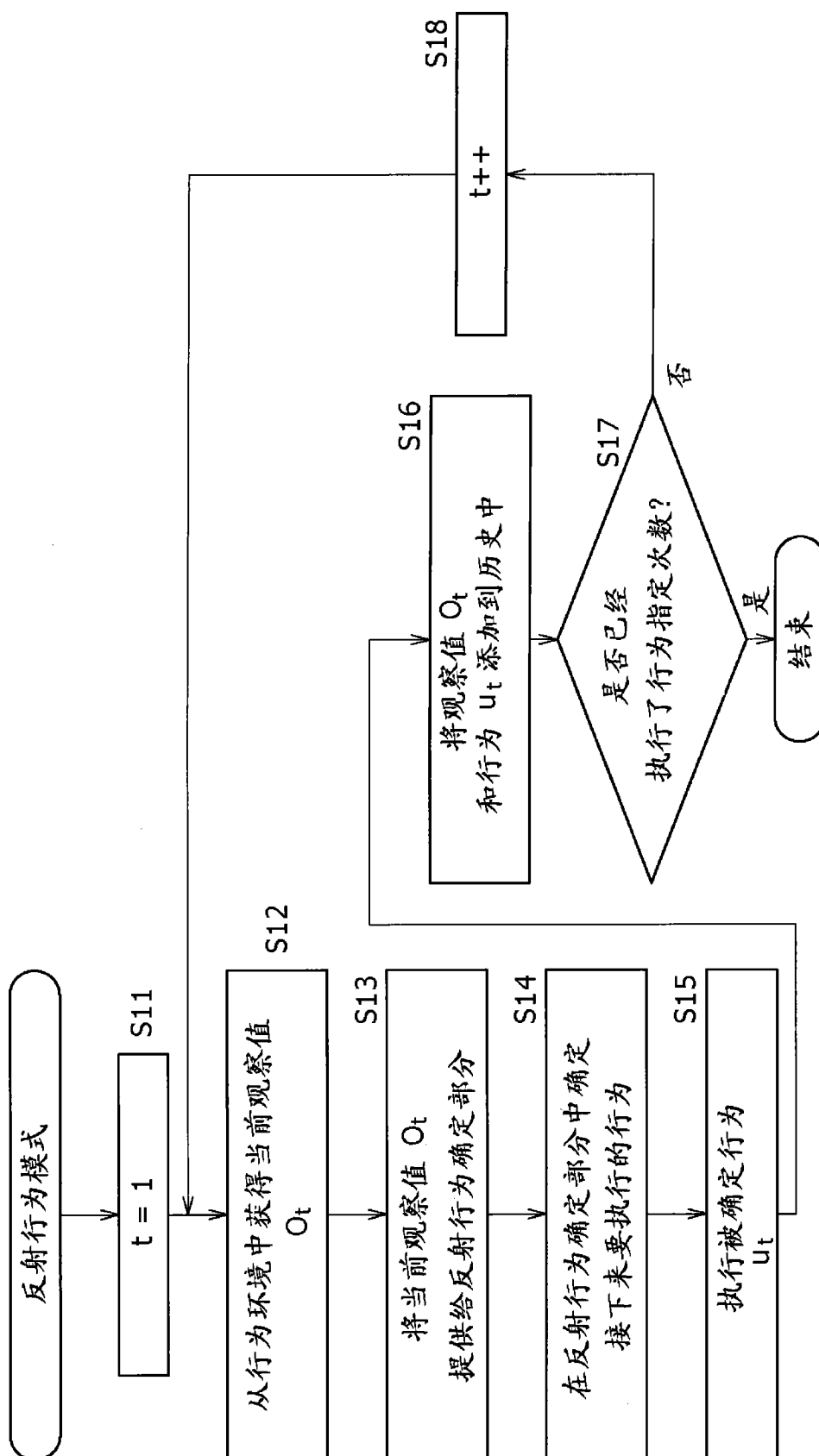
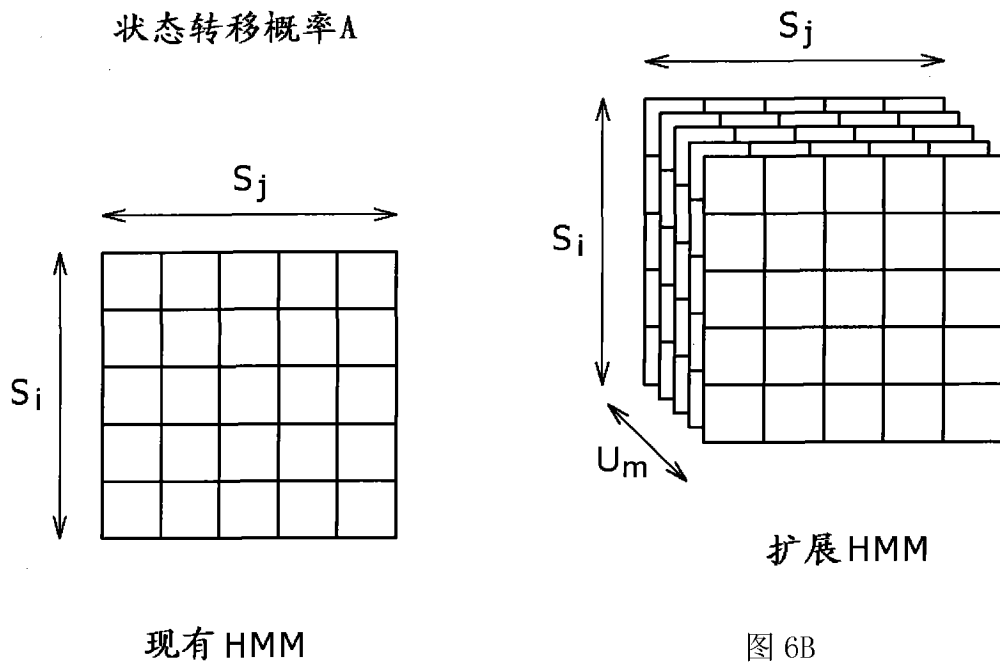


图 5



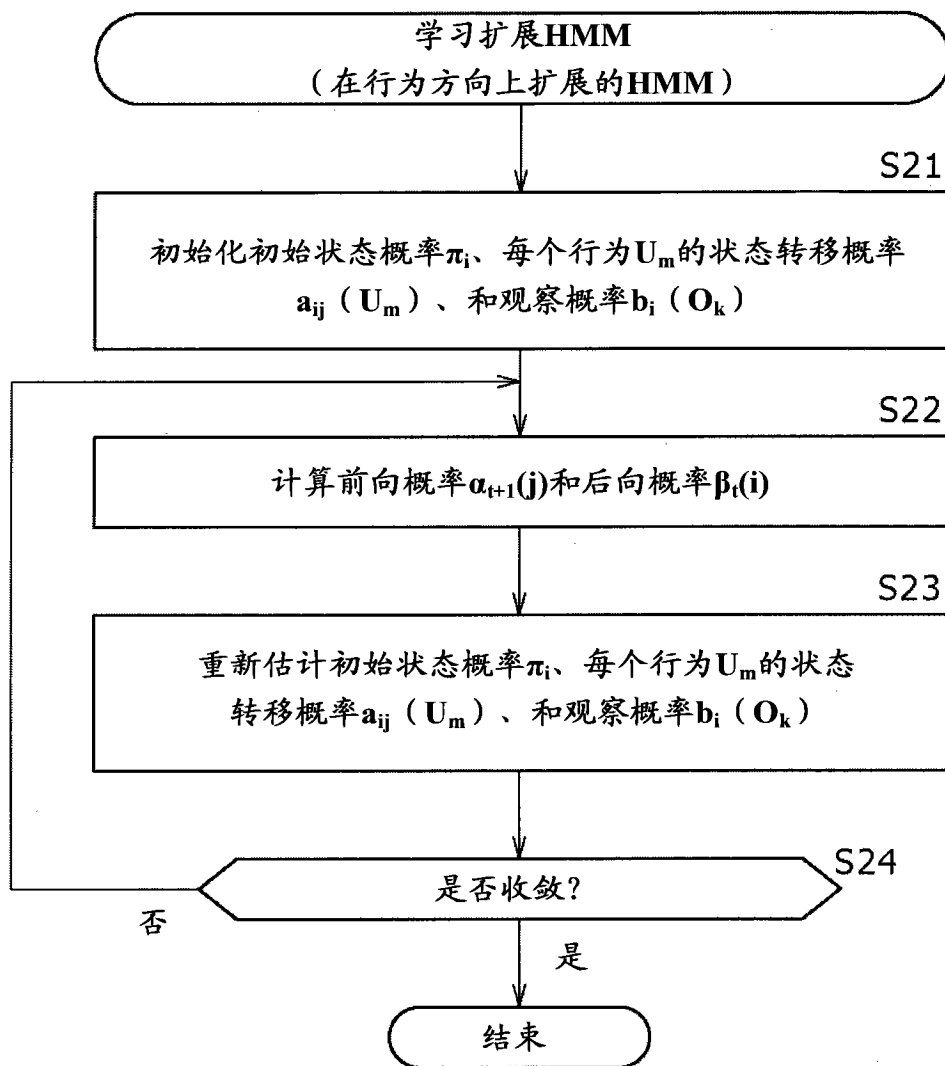


图 7

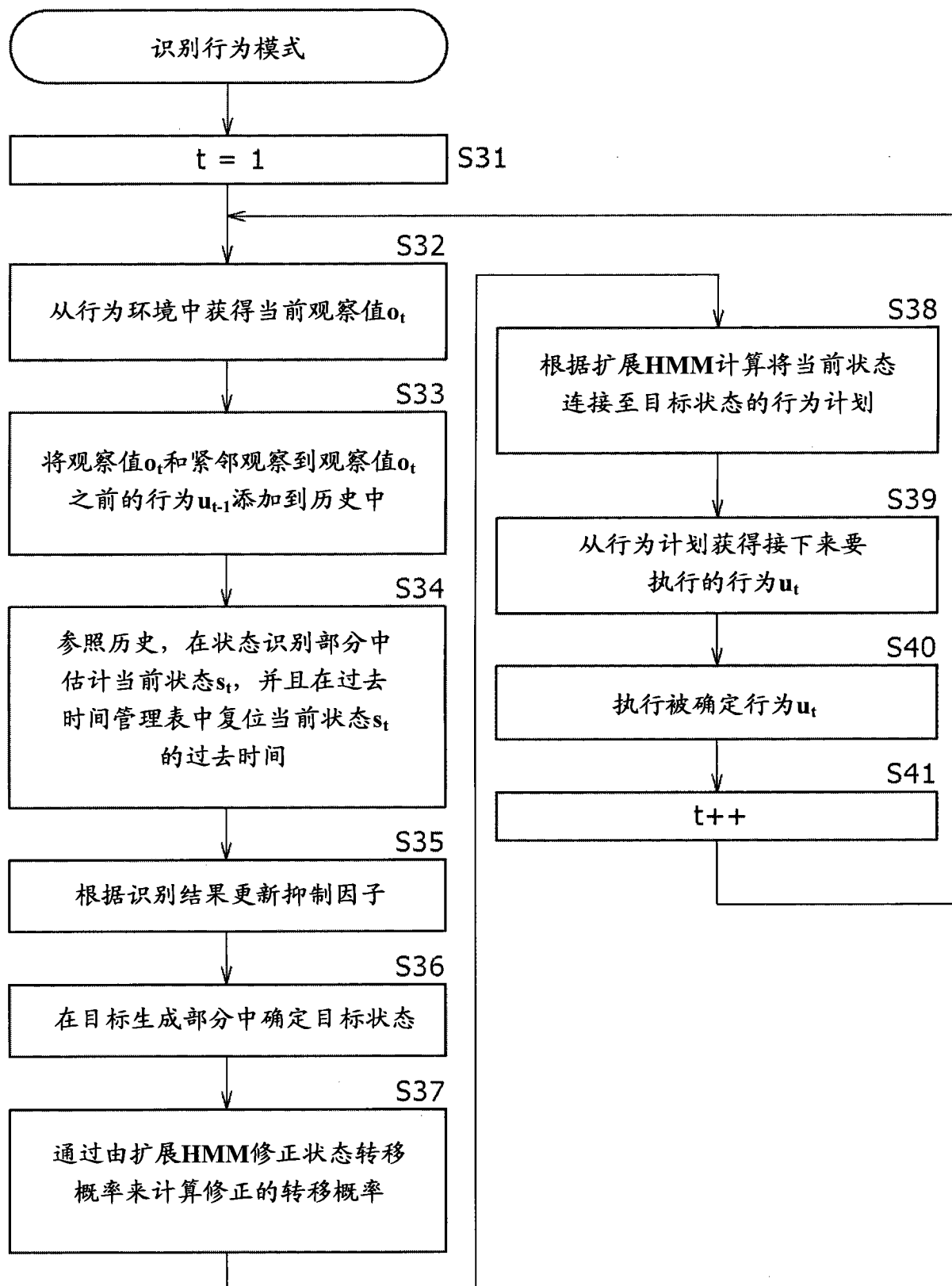


图 8

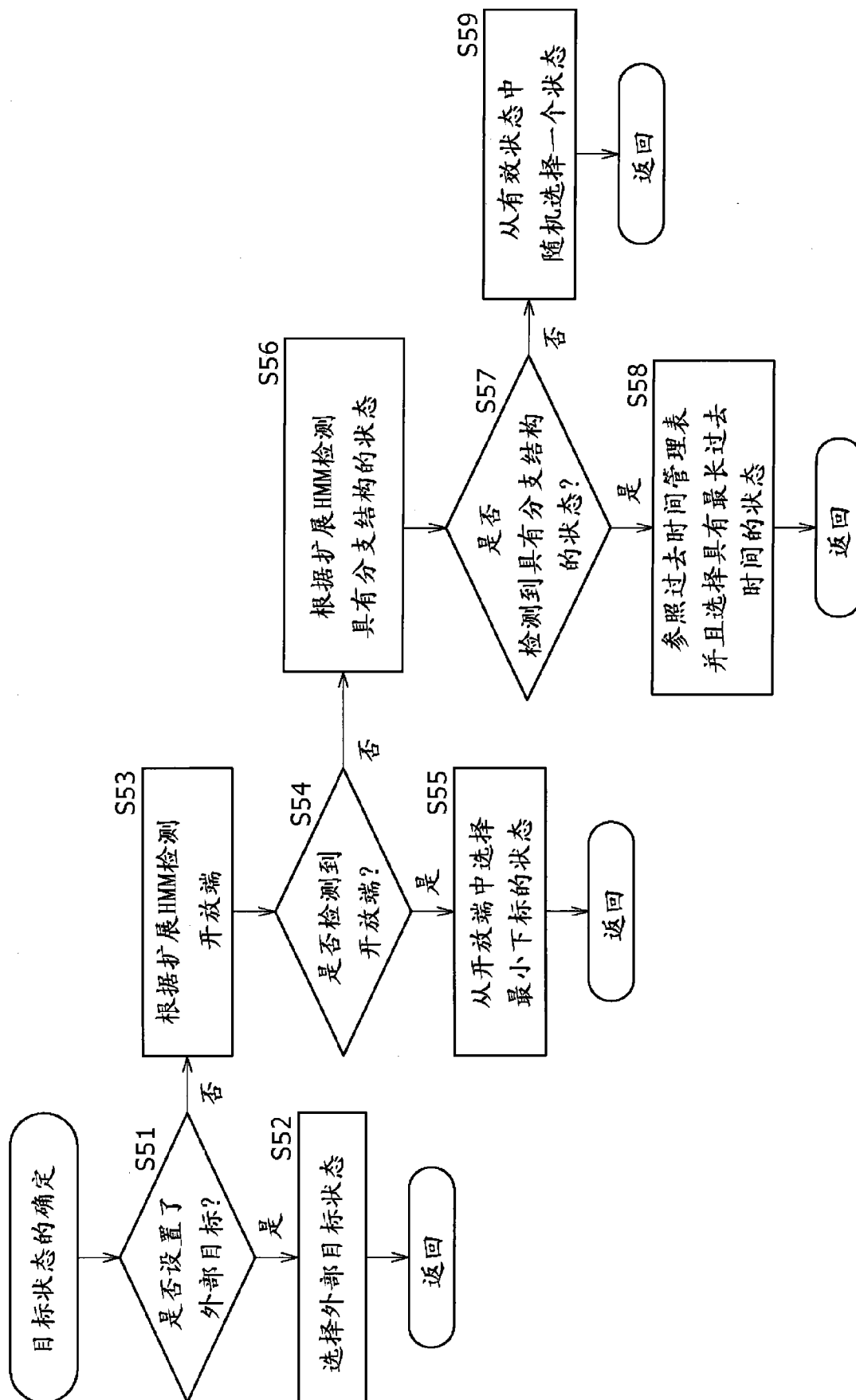
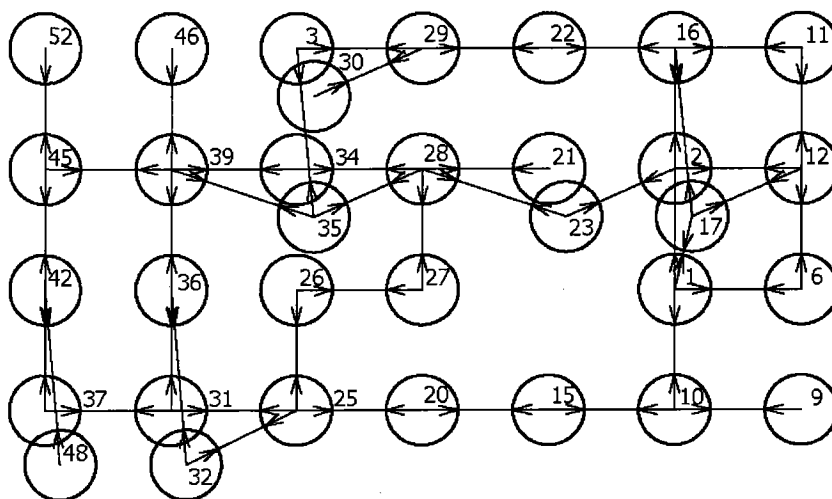


图 9



存储的扩展HMM的结构

图 10A

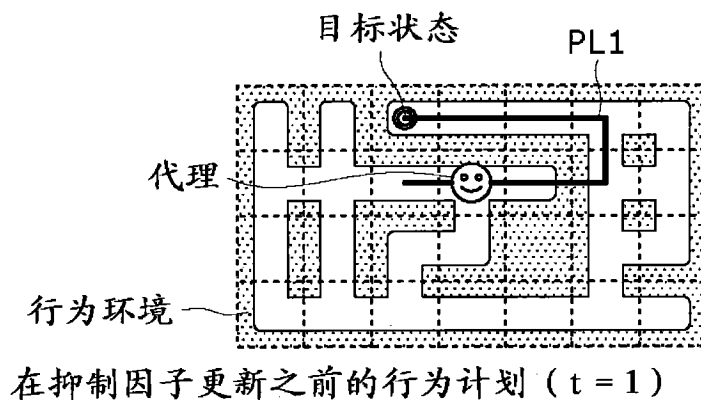


图 10B

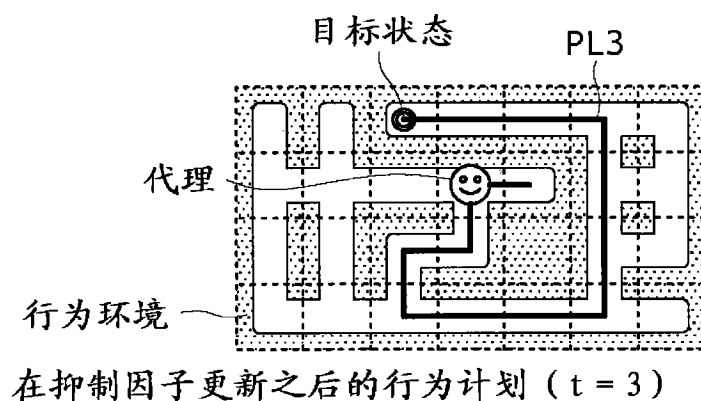


图 10C

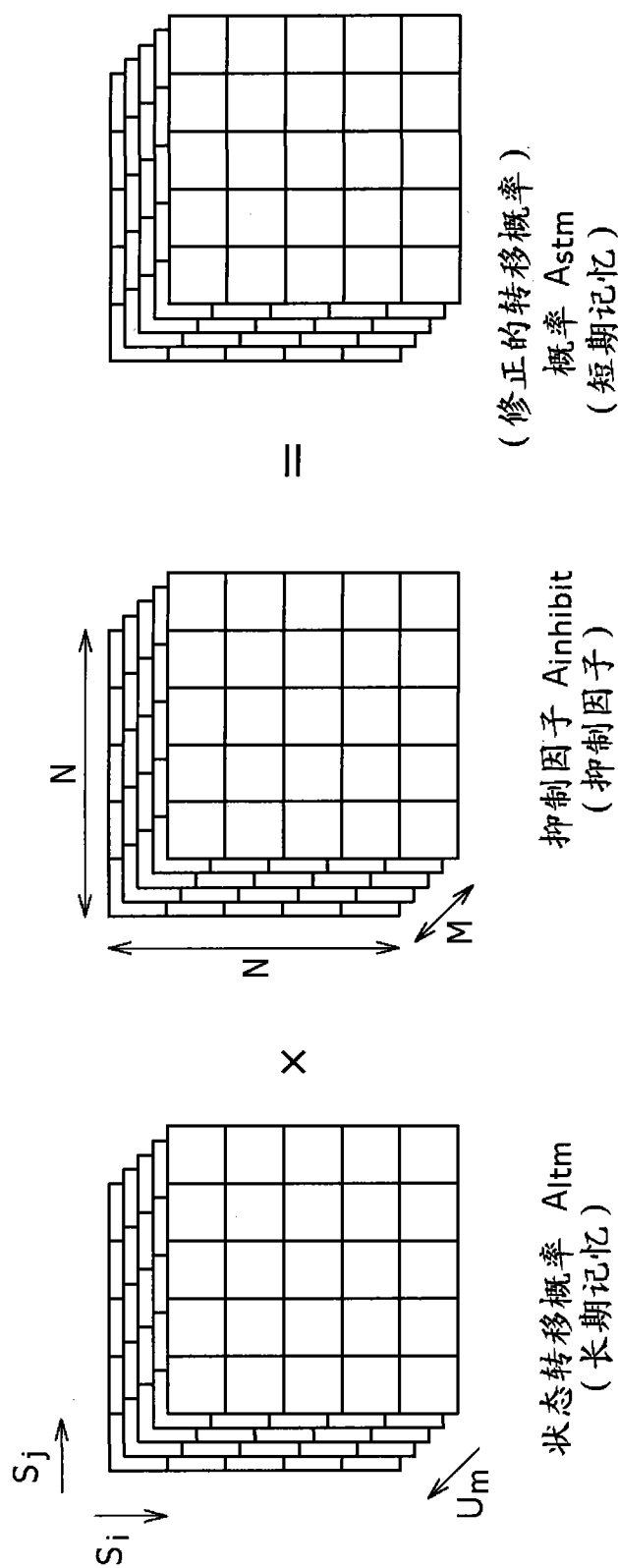


图 11

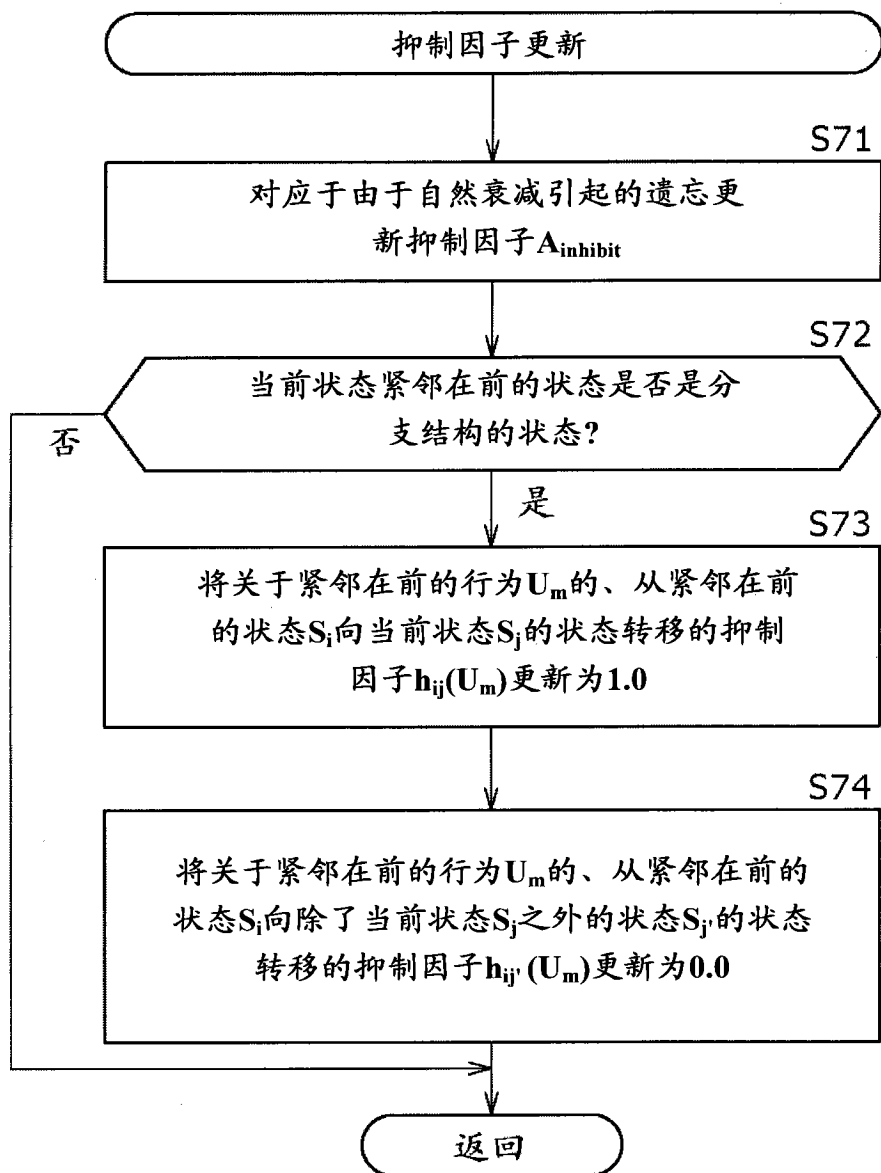


图 12

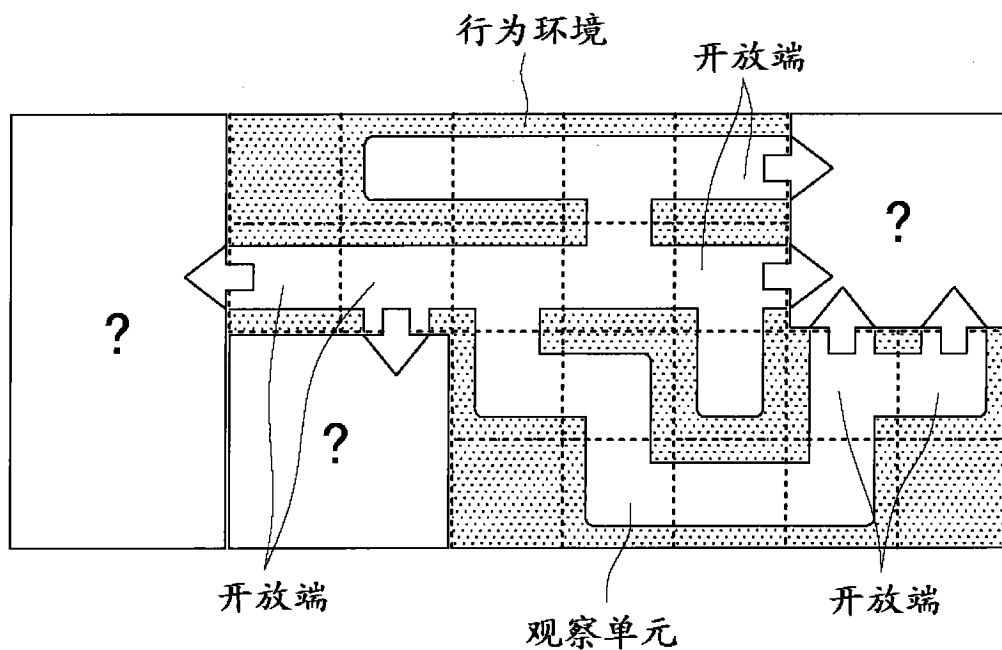


图 13

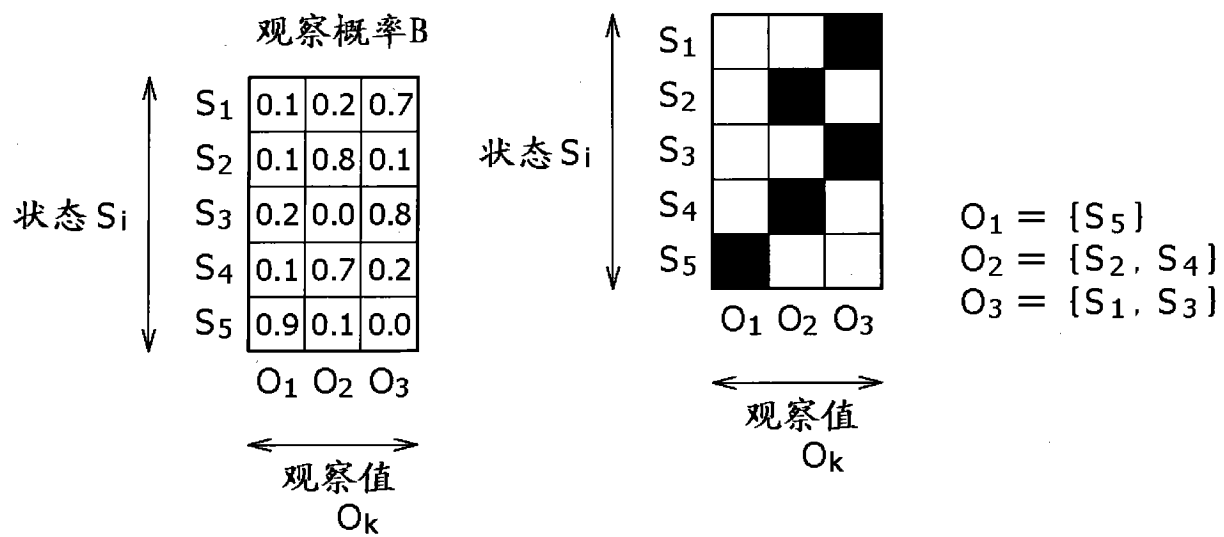


图 14B

图 14A

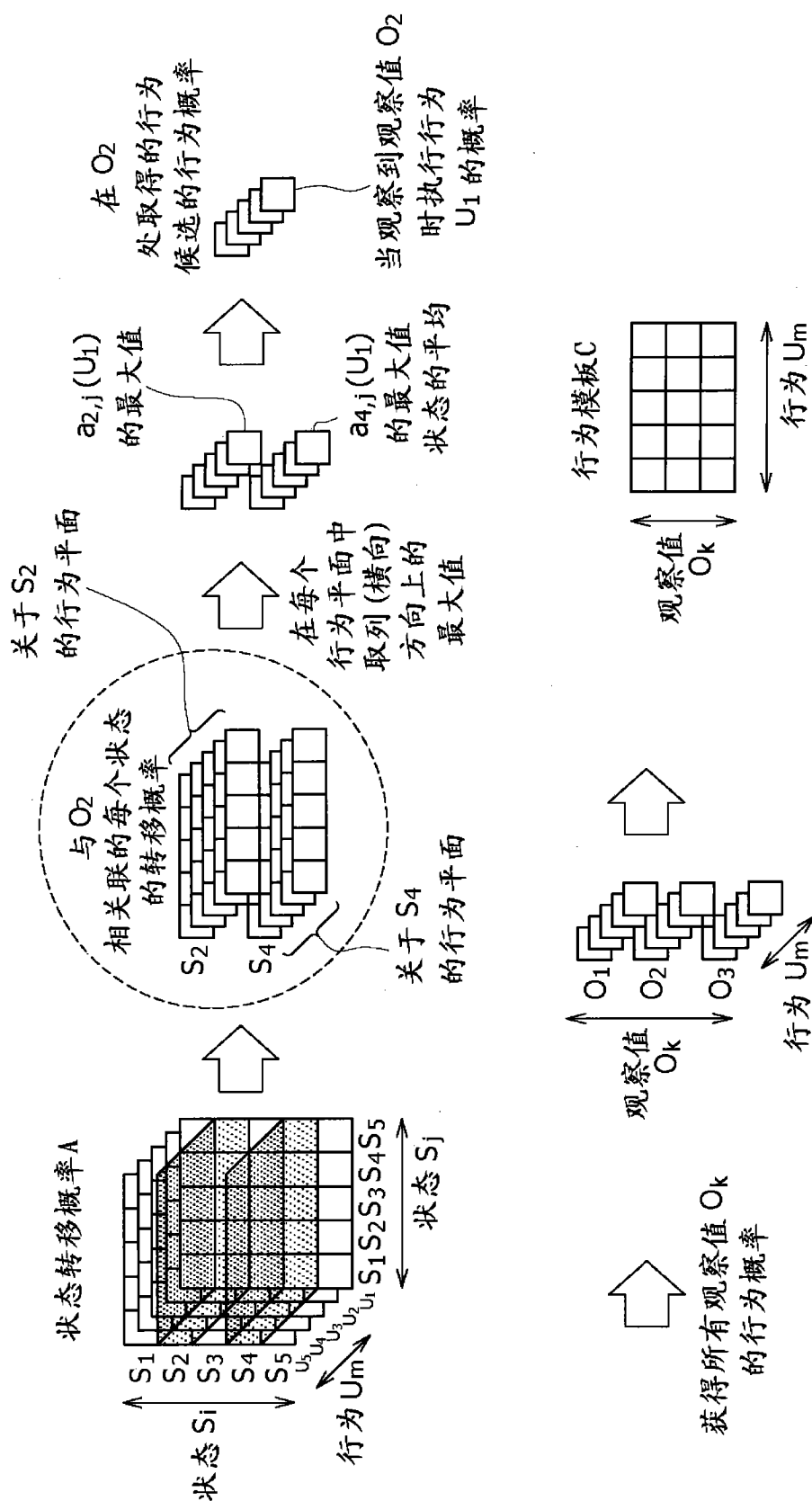


图 15

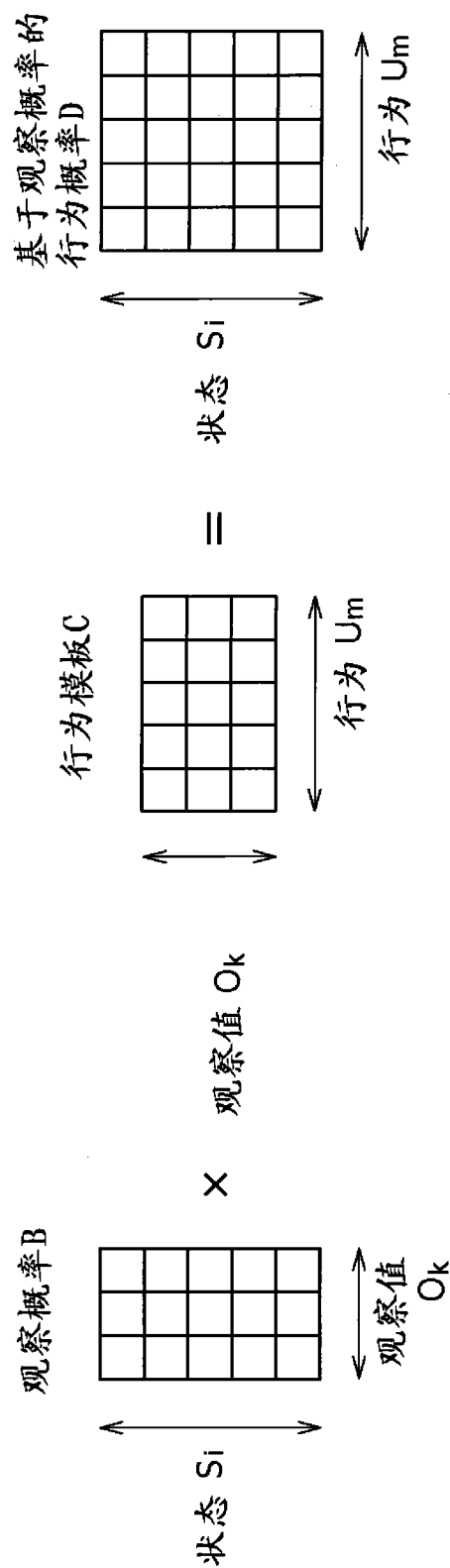


图 16

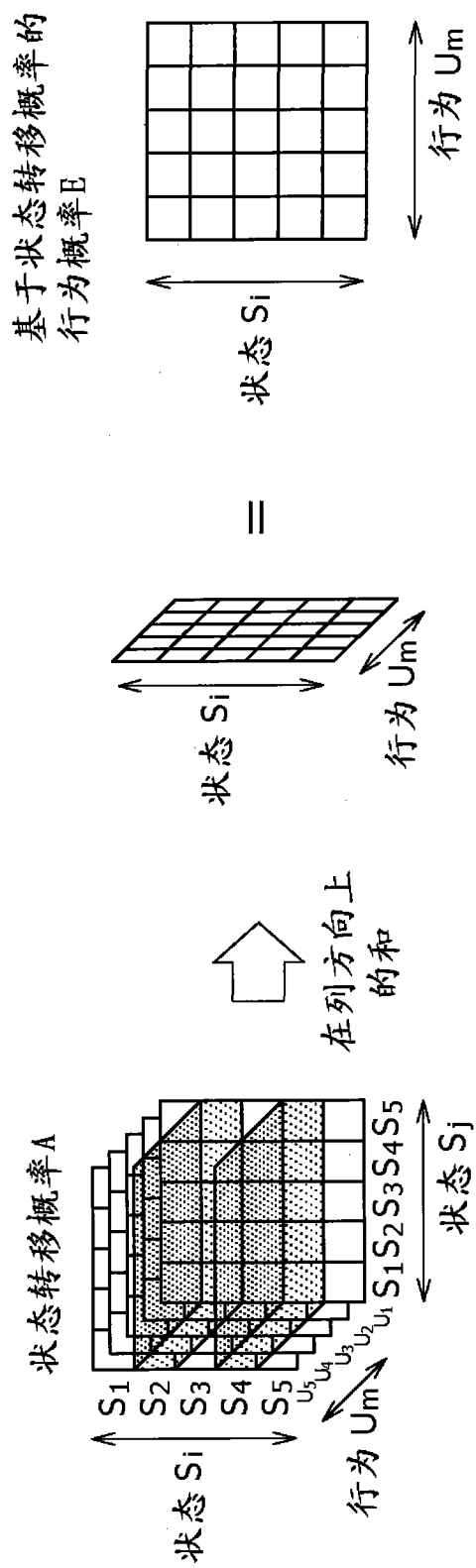


图 17

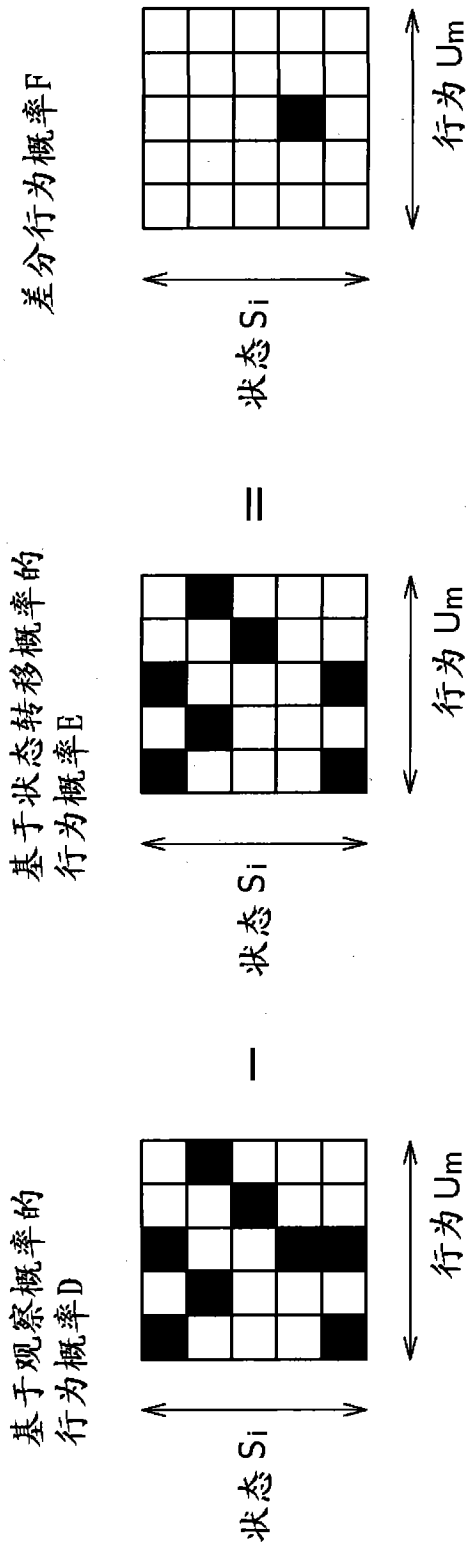


图 18

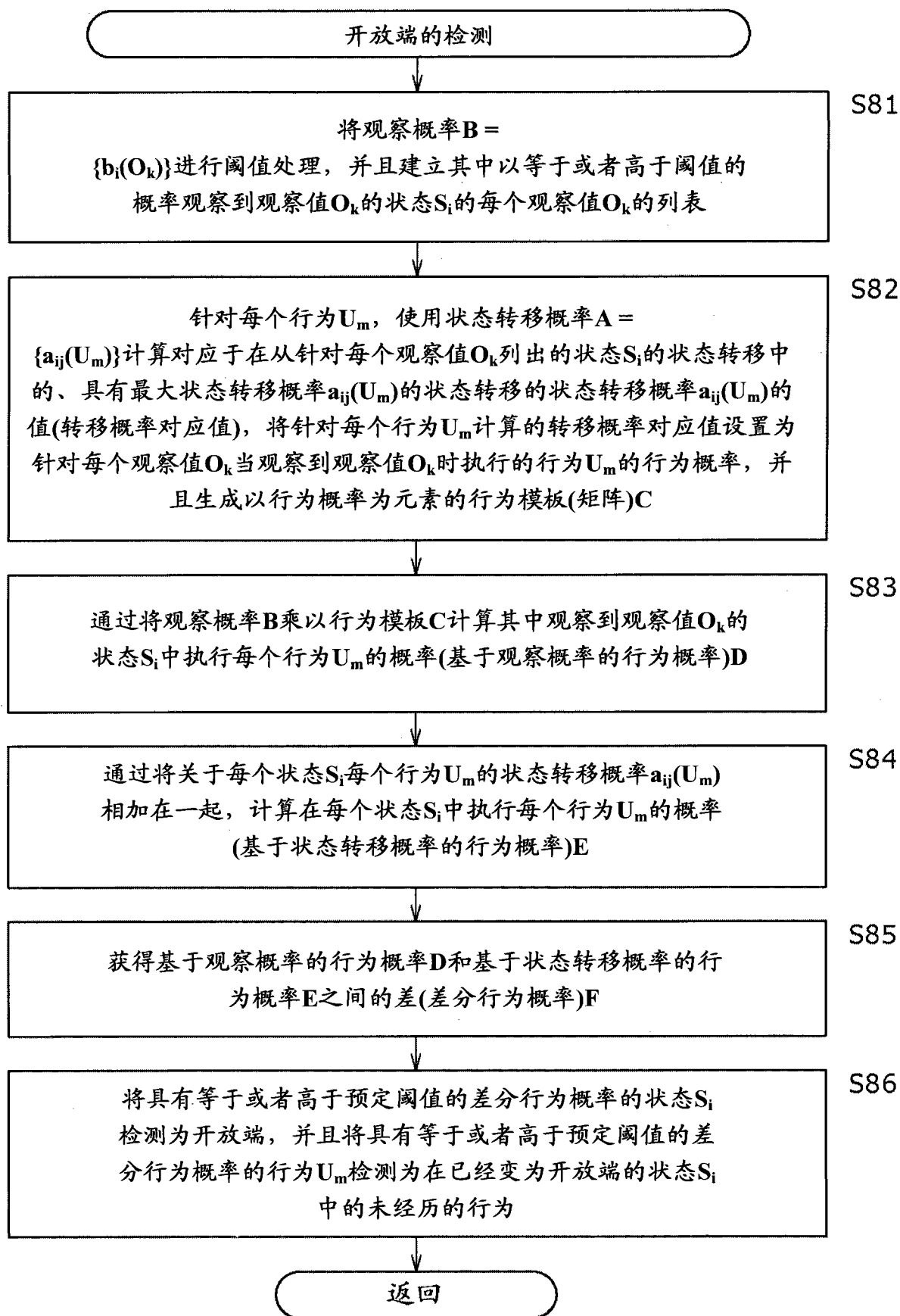


图 19

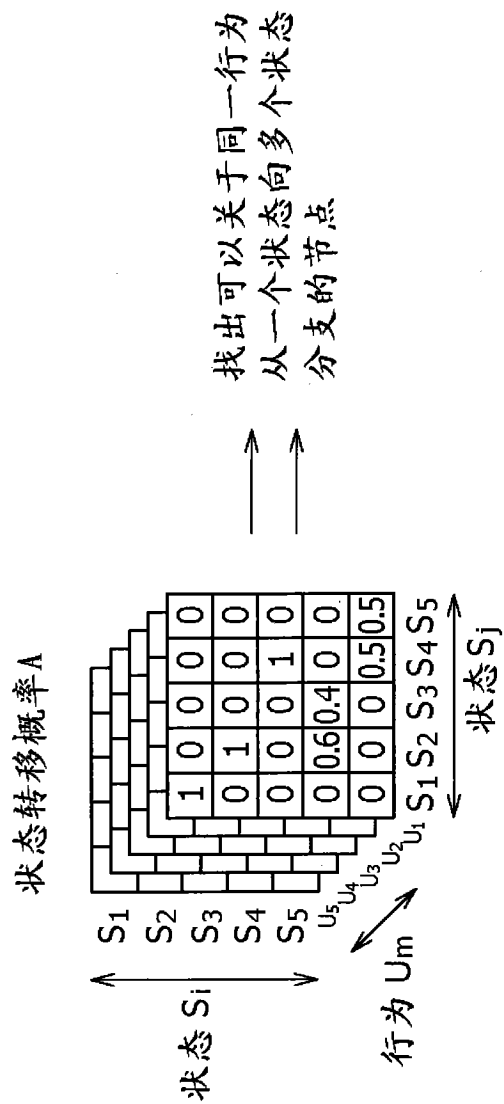


图 20

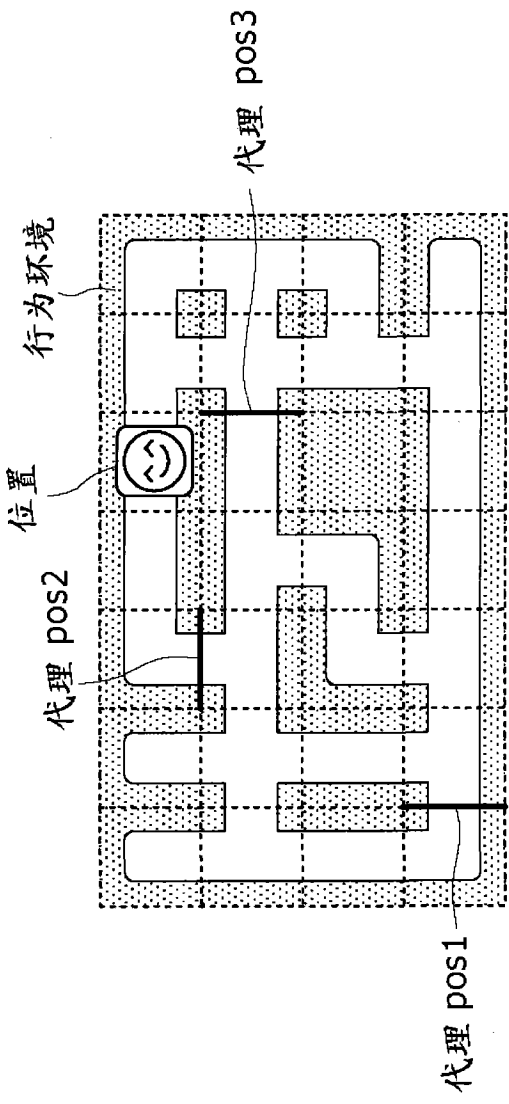


图 21A

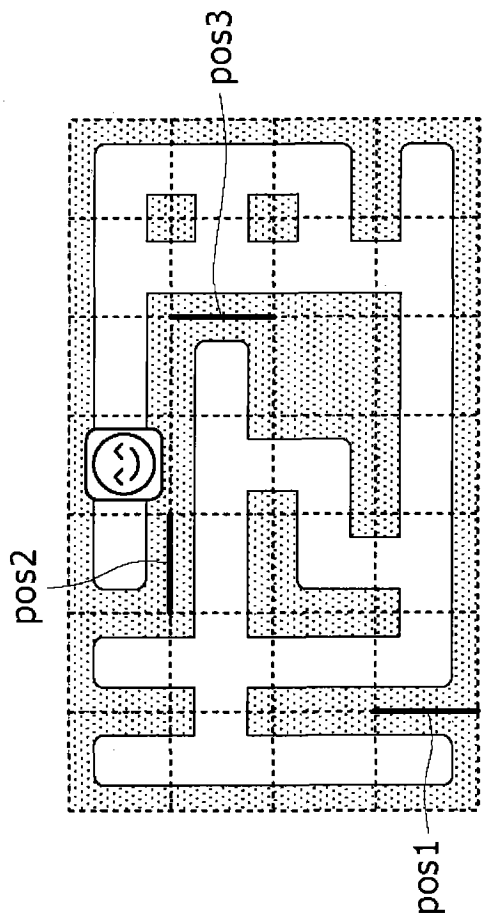


图 21B

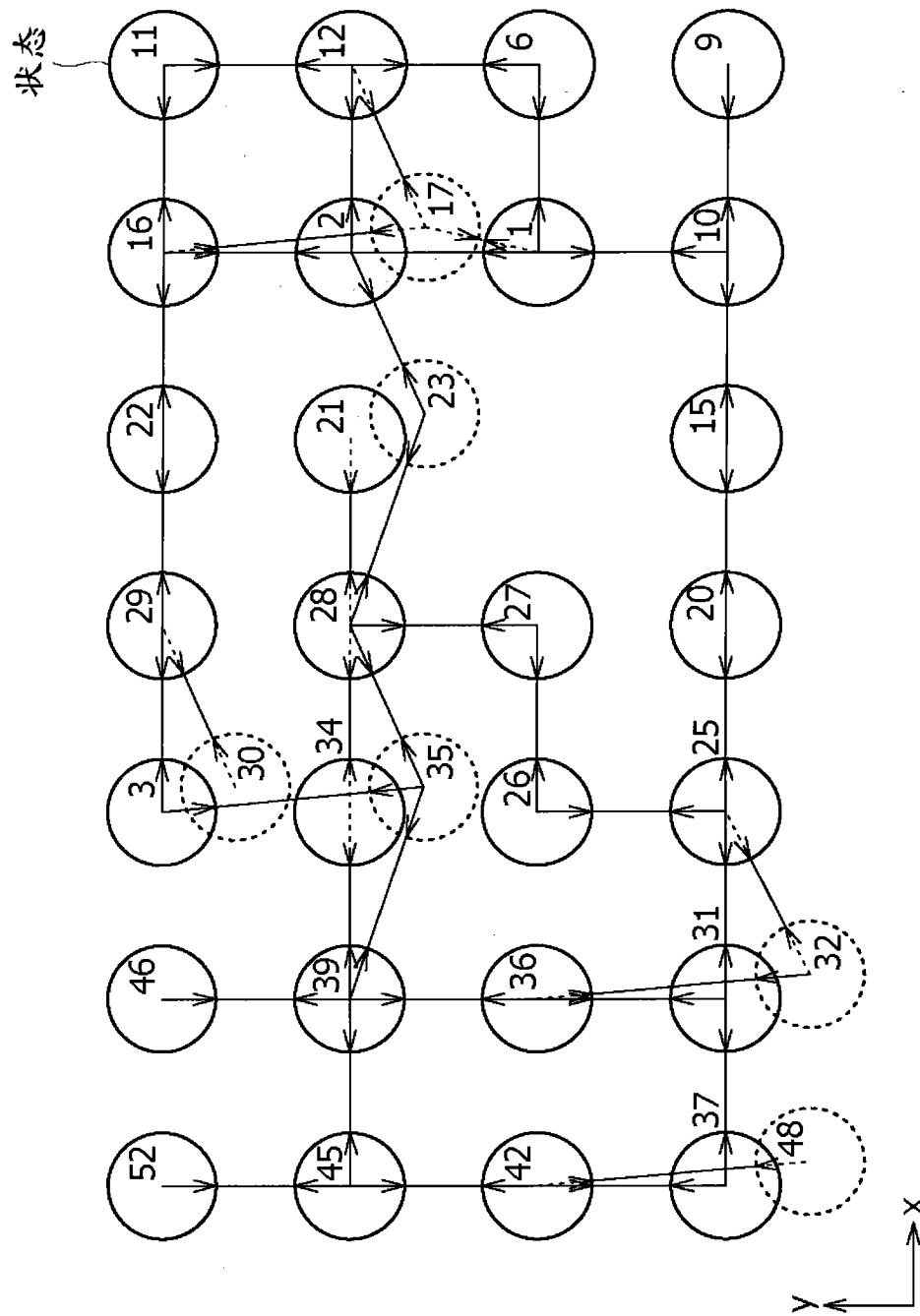


图 22

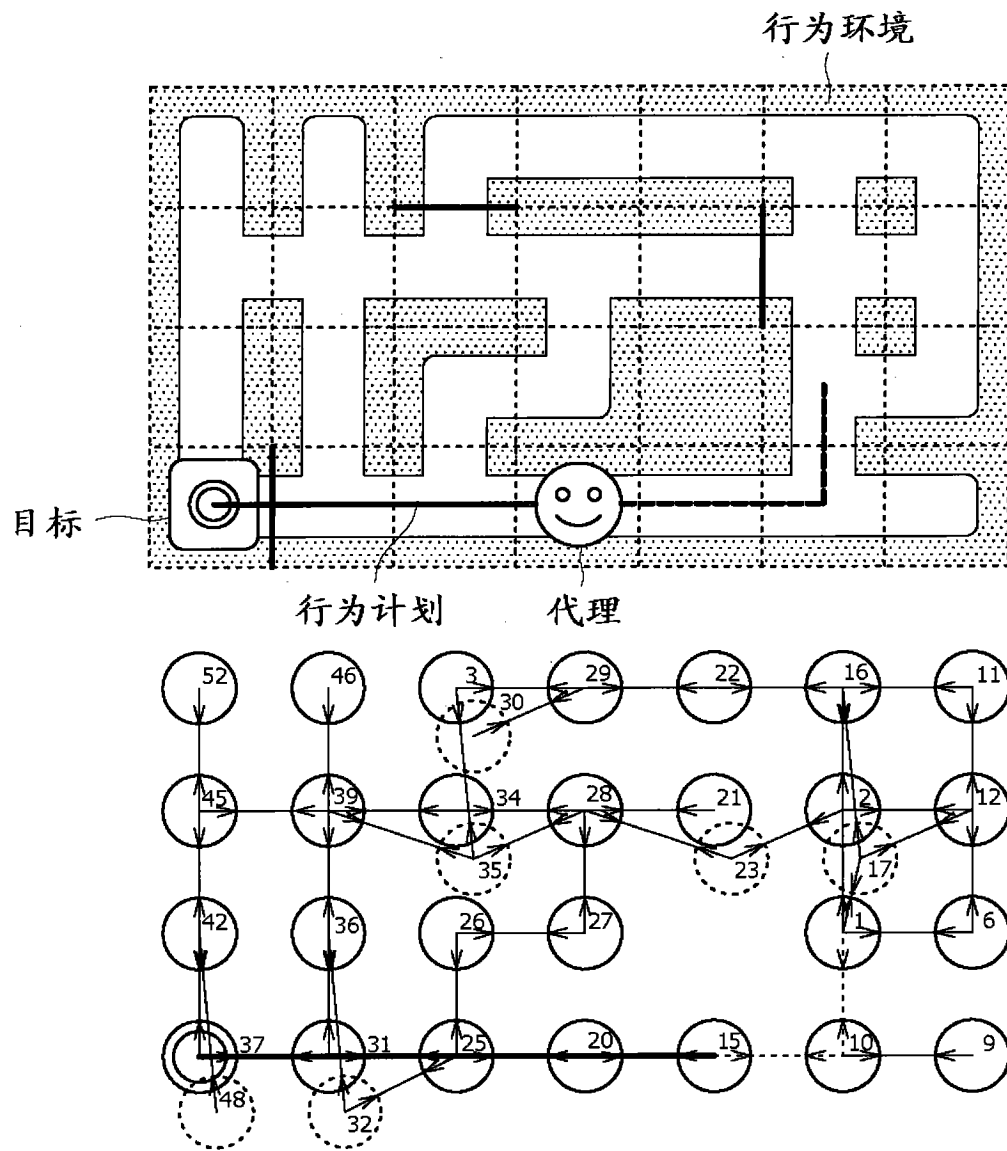


图 23

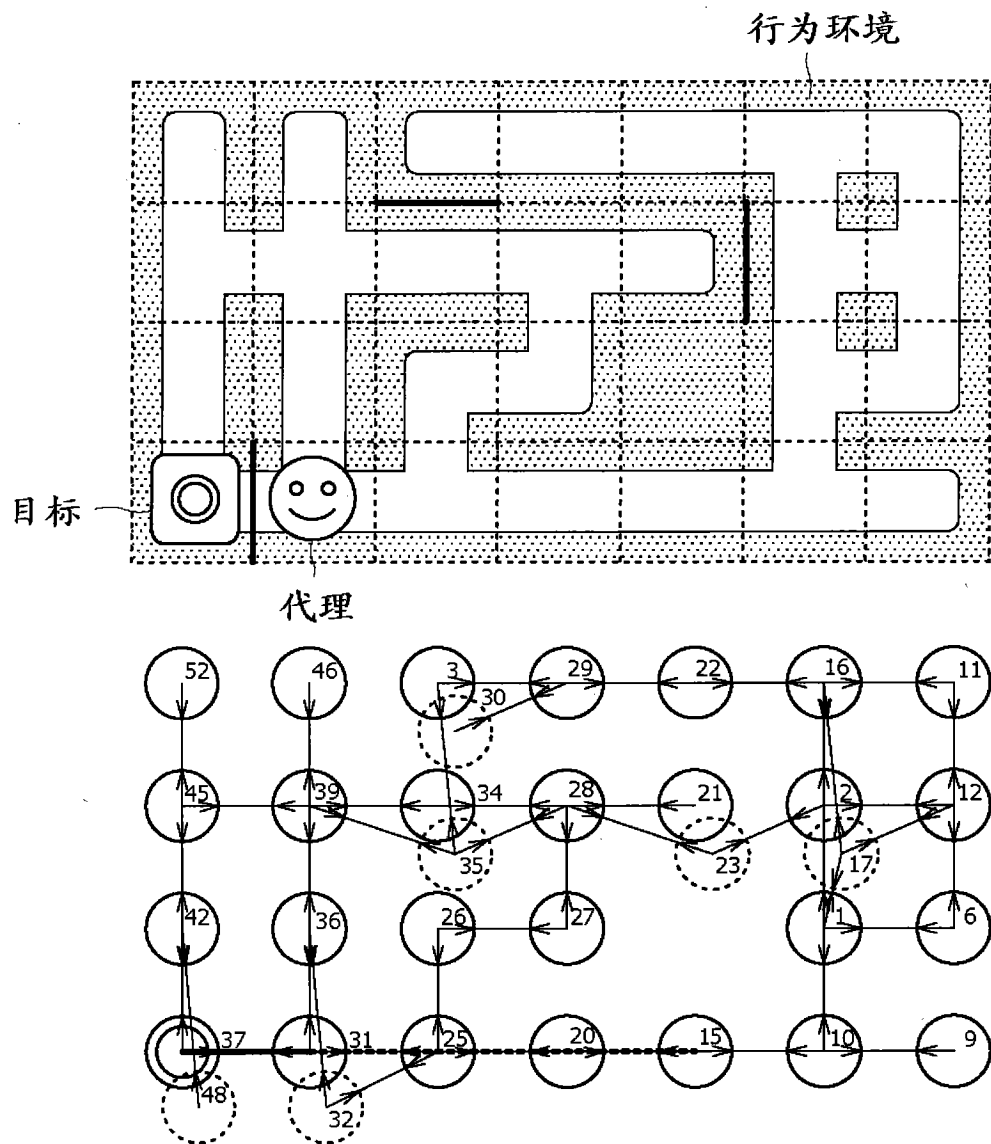


图 24

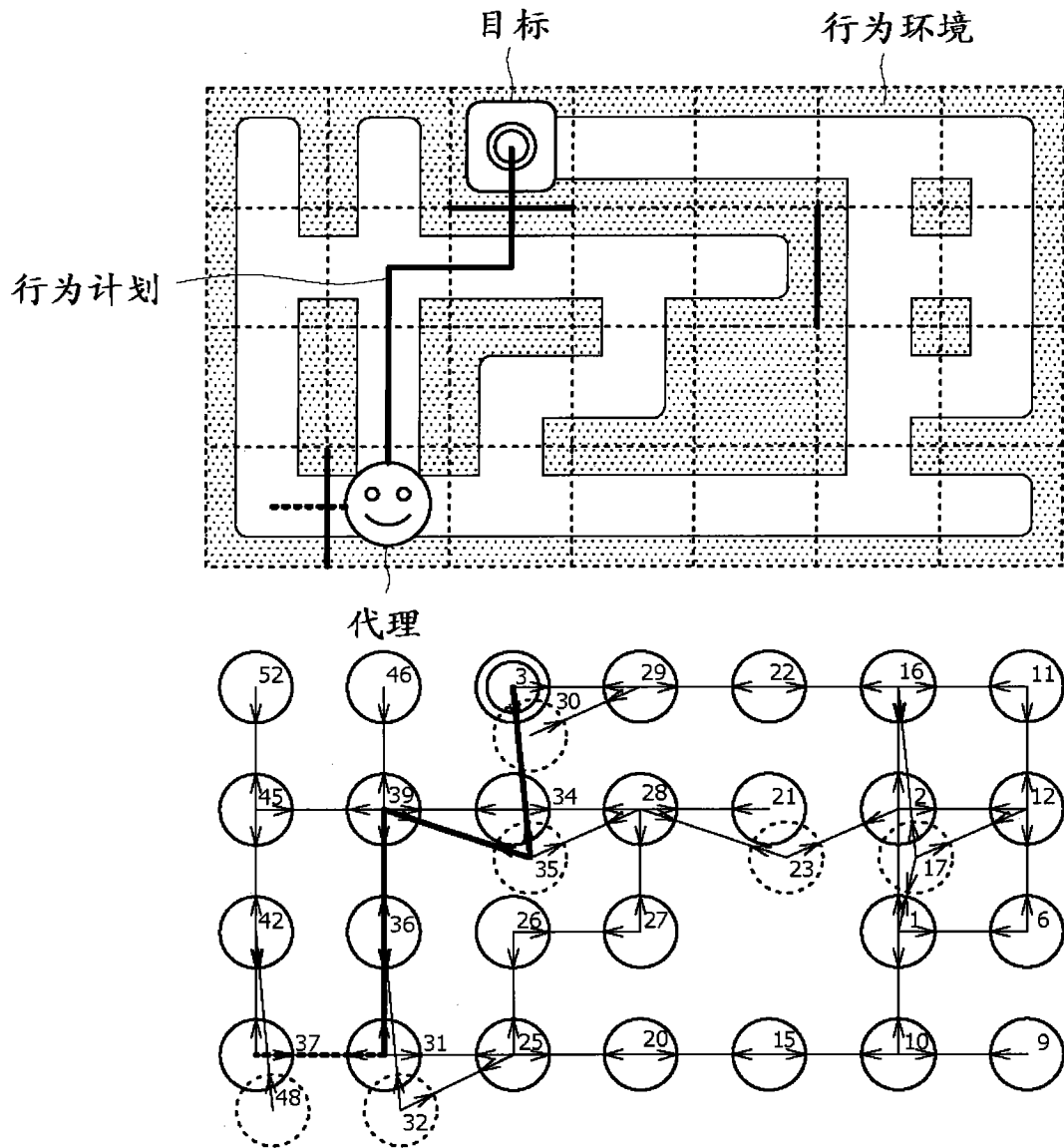


图 25

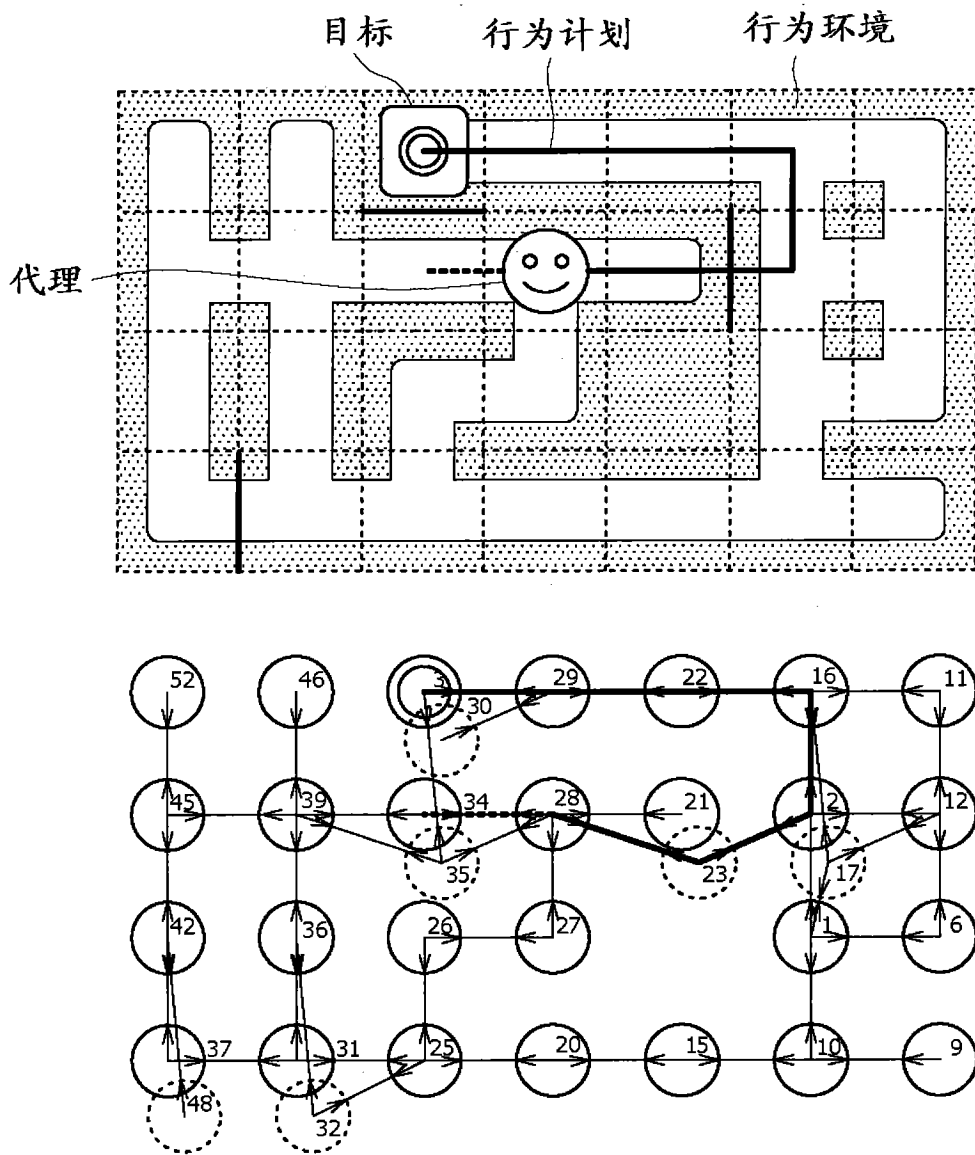


图 26

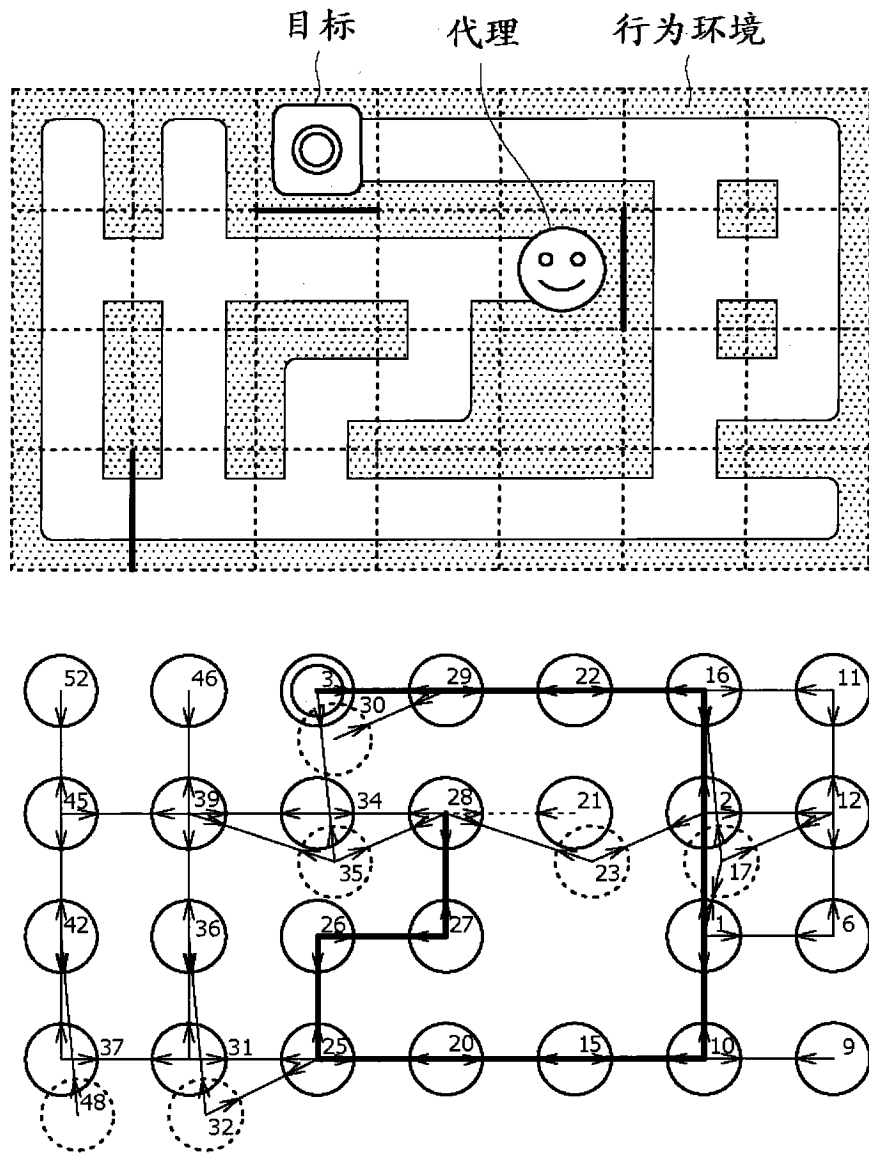


图 27

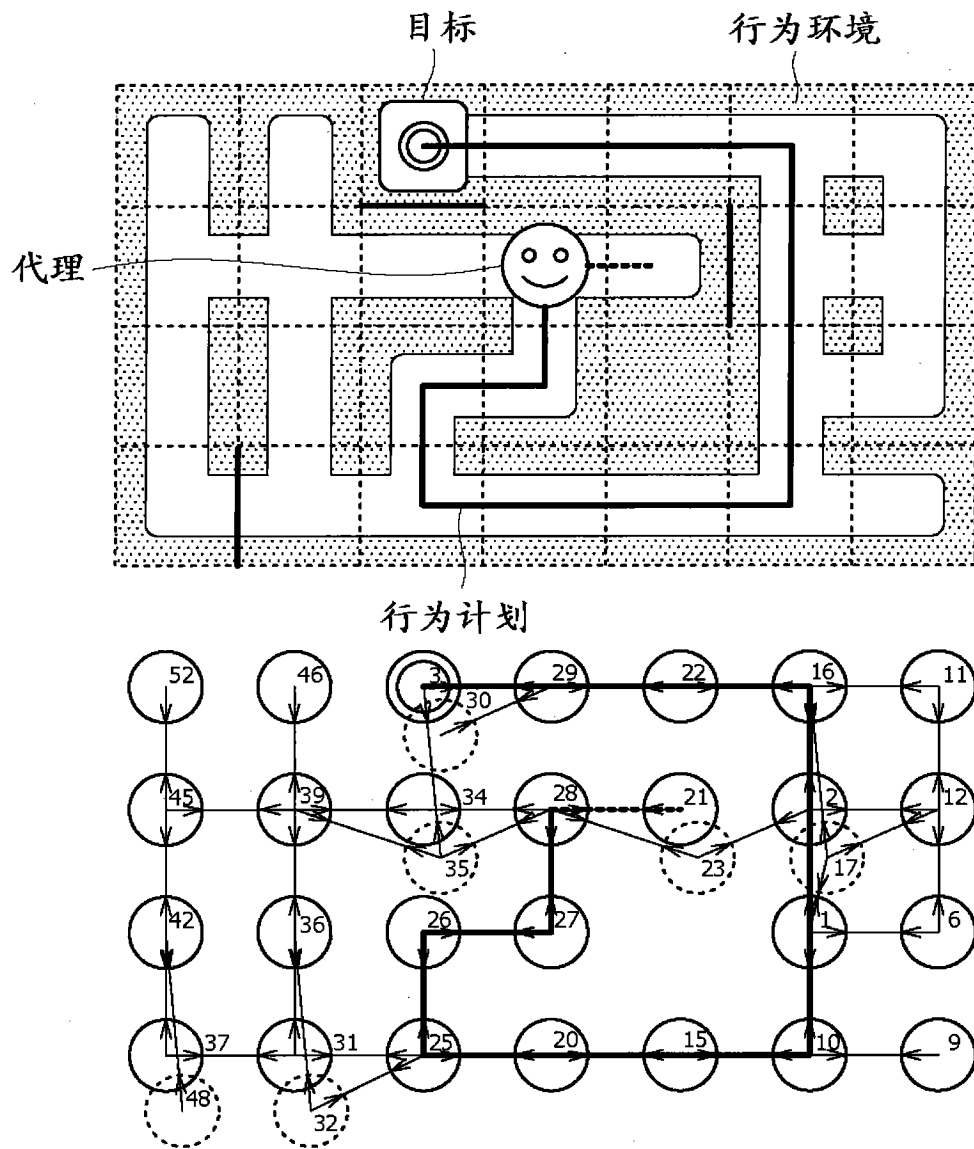


图 28

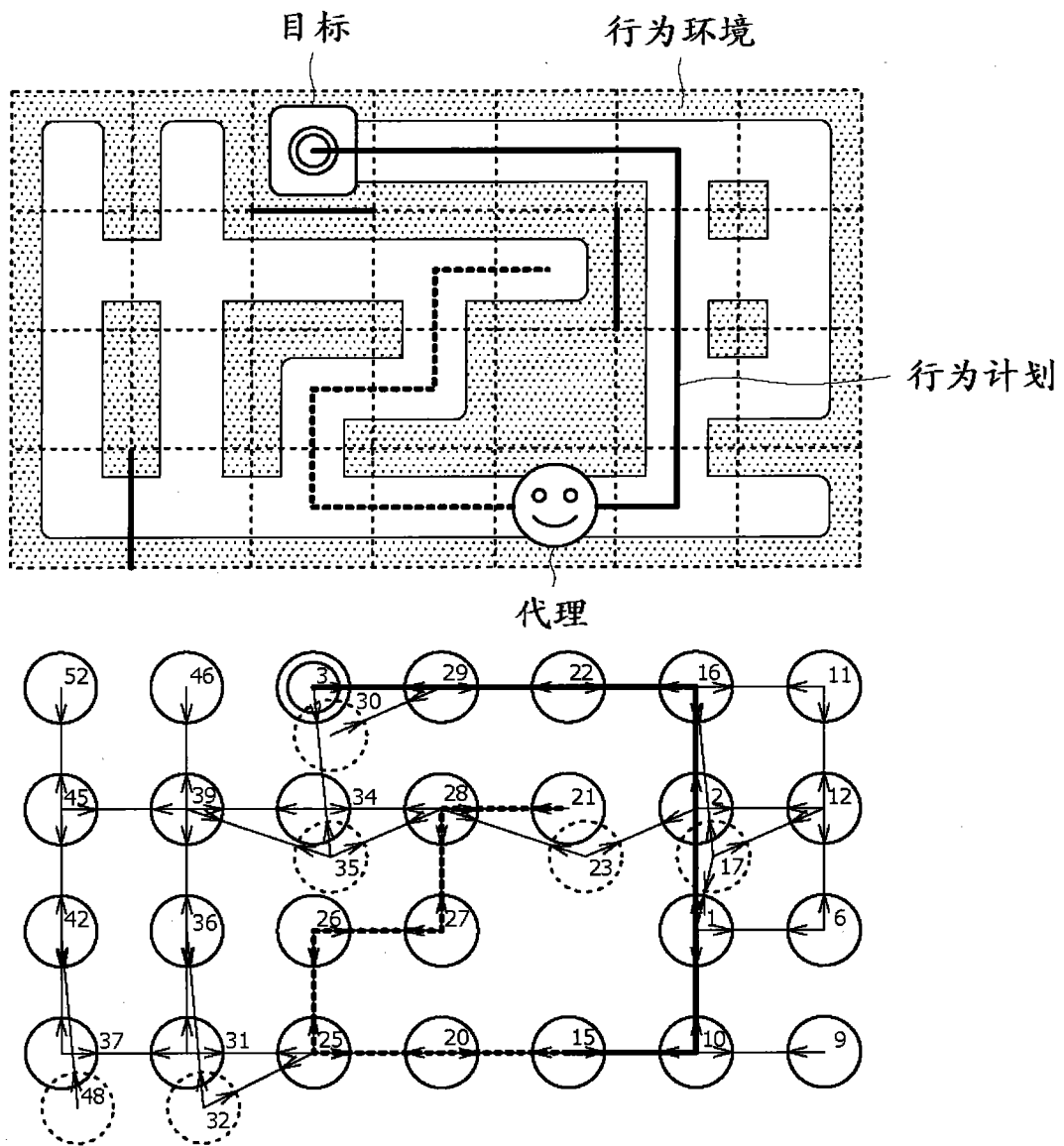


图 29

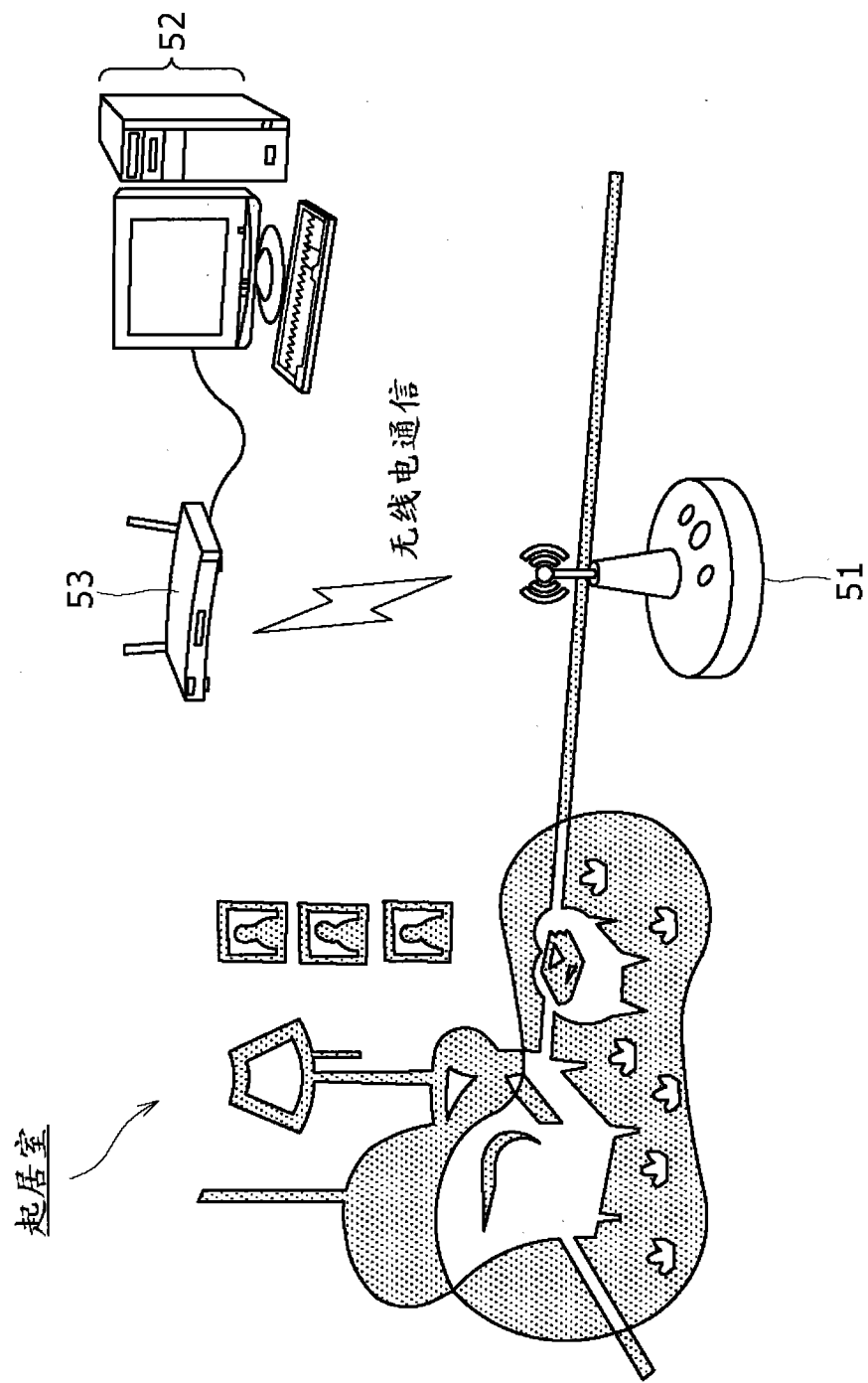


图 30

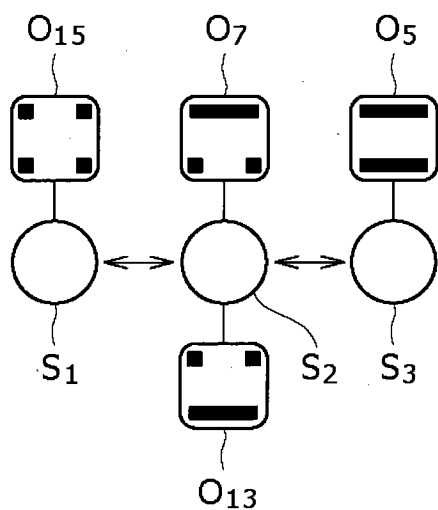


图 31A

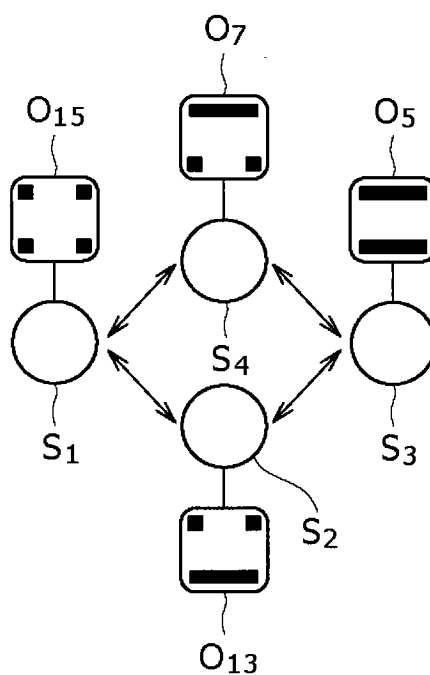


图 31B

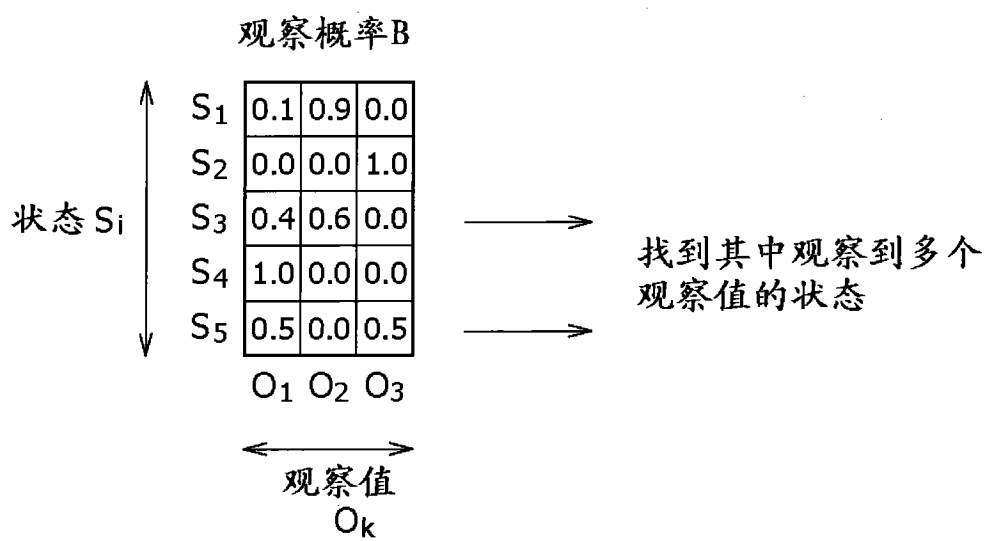


图 32

将其中观察到多个观察值 O_1 和 O_2 的要划分的状态 S_3 划分为状态 S_3 和 S_6 , 向每个状态分配一个观察值, 并且将在每个状态中观察值的观察分配给该状态的观察值的概率设置为 1.0

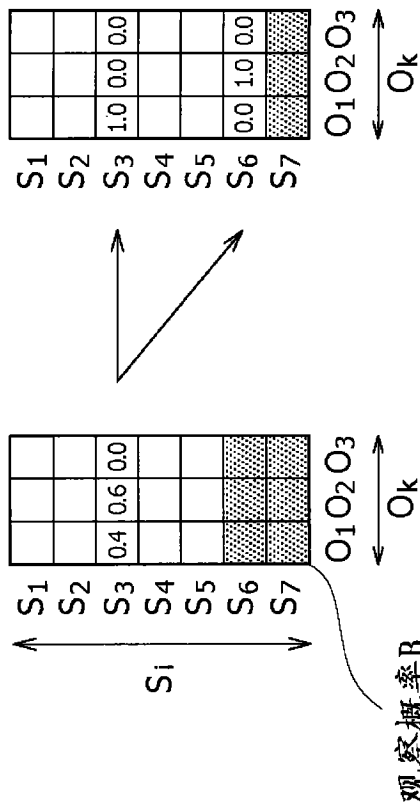


图 33A

与被分配给划分后状态的观察值的观察概率成比例地划分转移概率

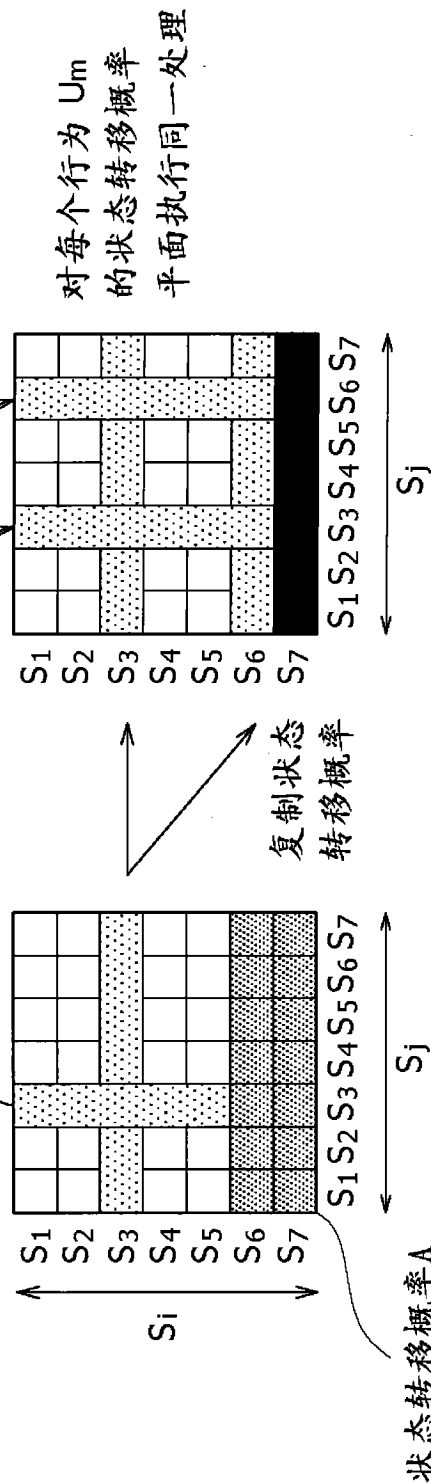


图 33B

前向合并

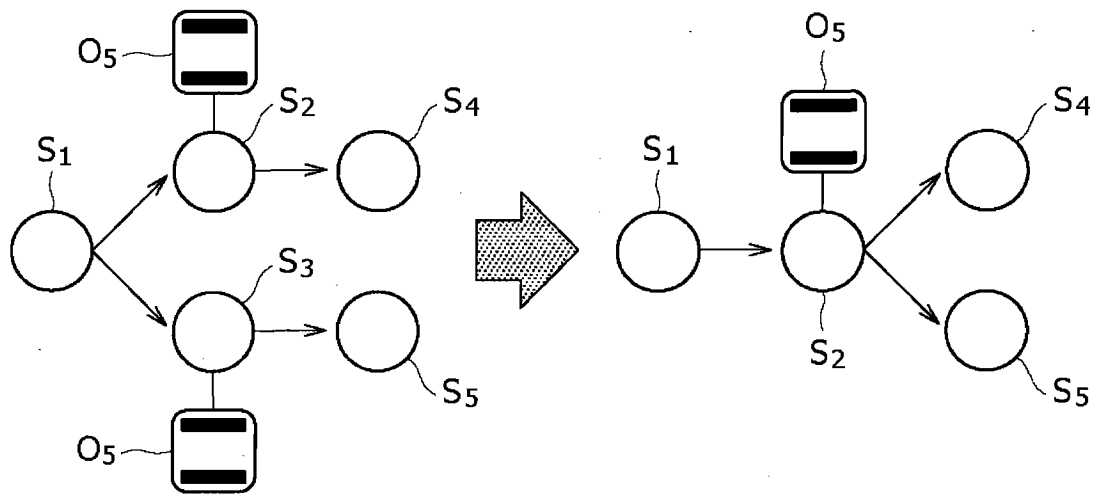


图 34A

后向合并

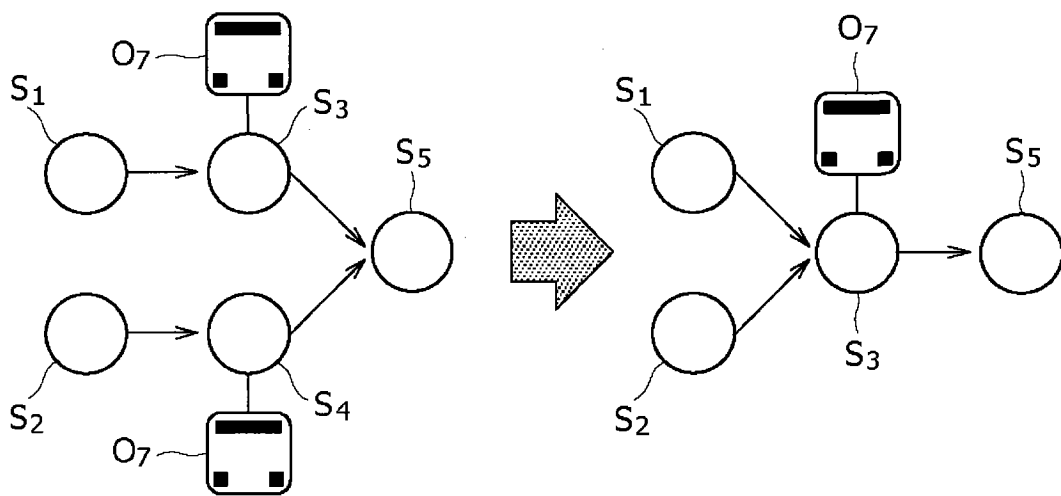
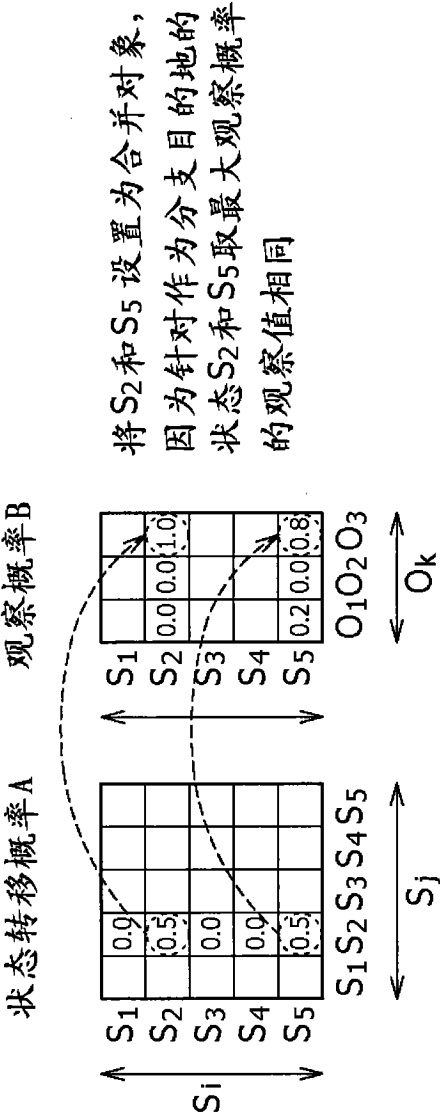
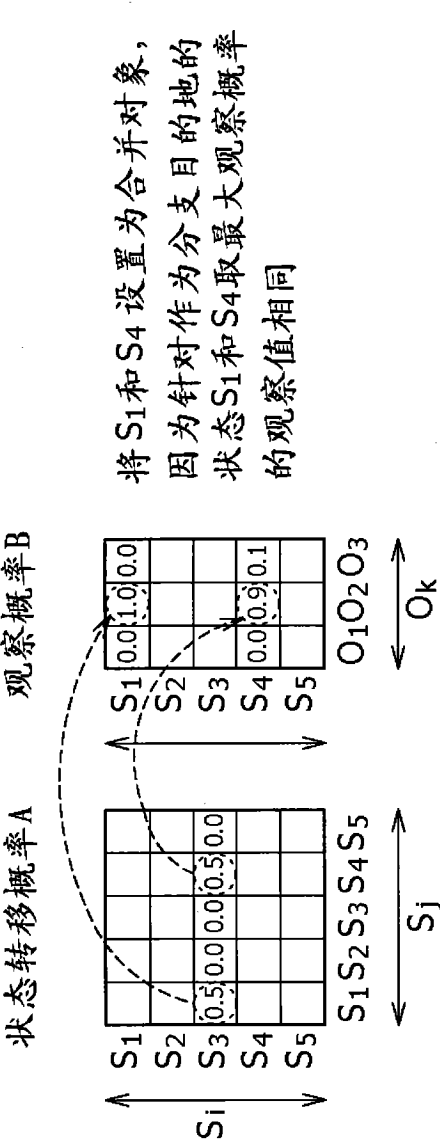


图 34B



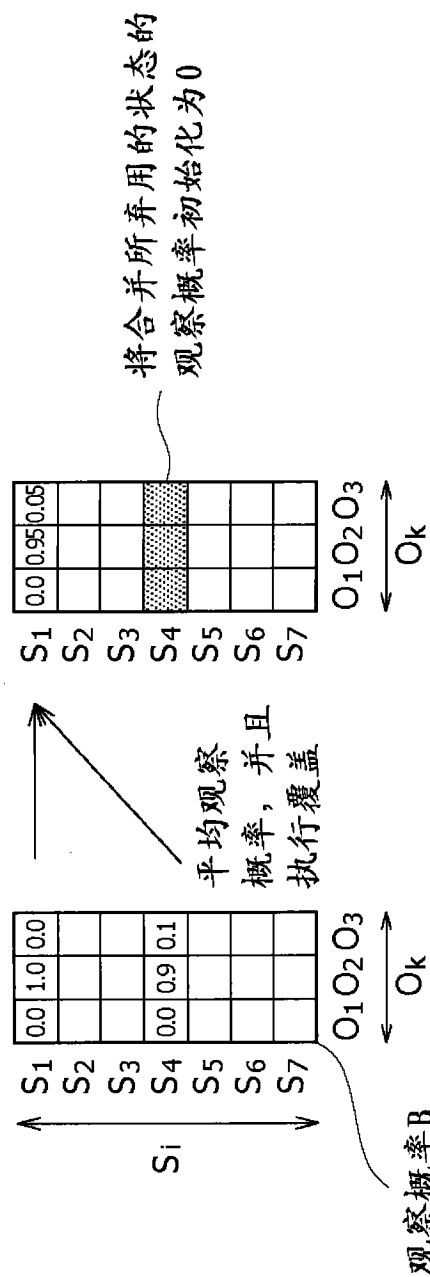


图 36A

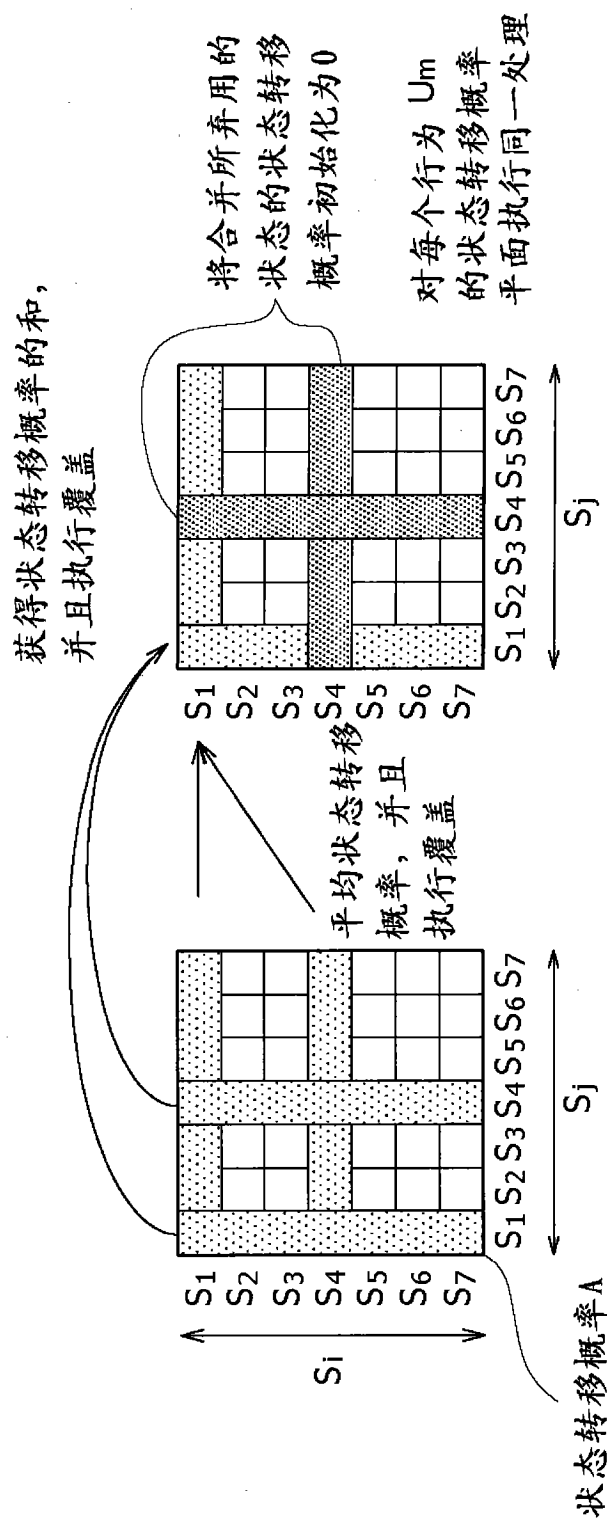


图 36B

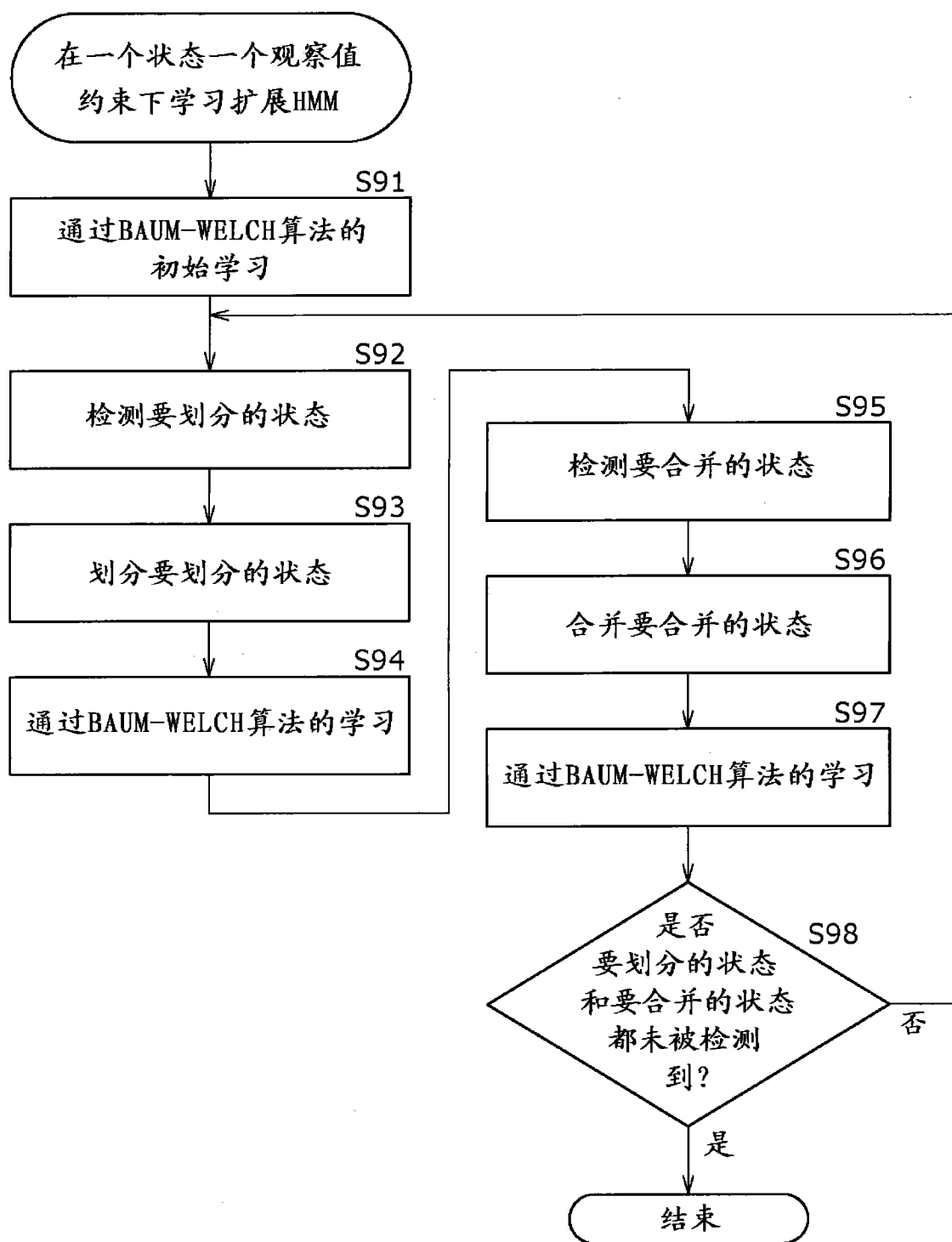


图 37

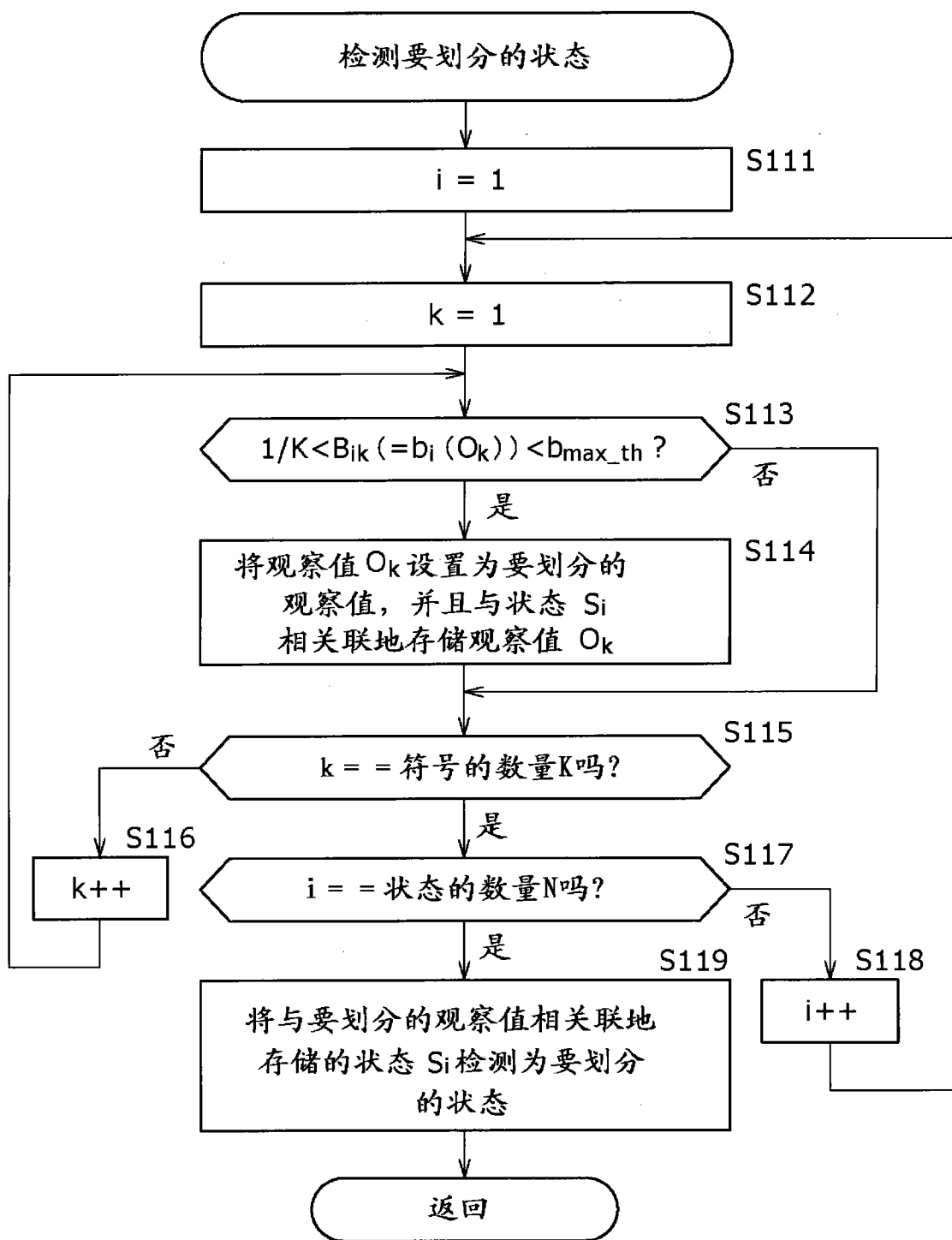


图 38

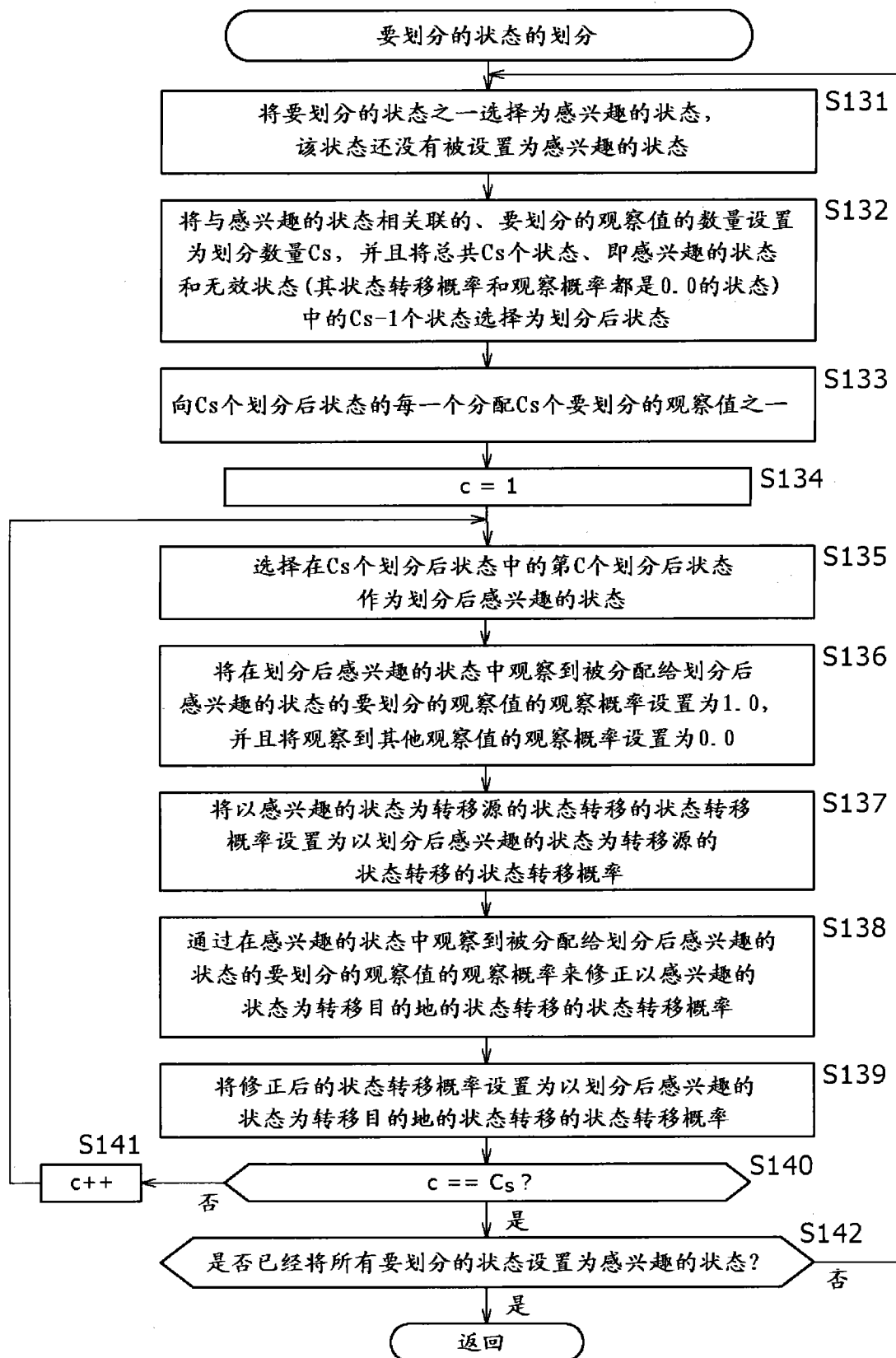


图 39

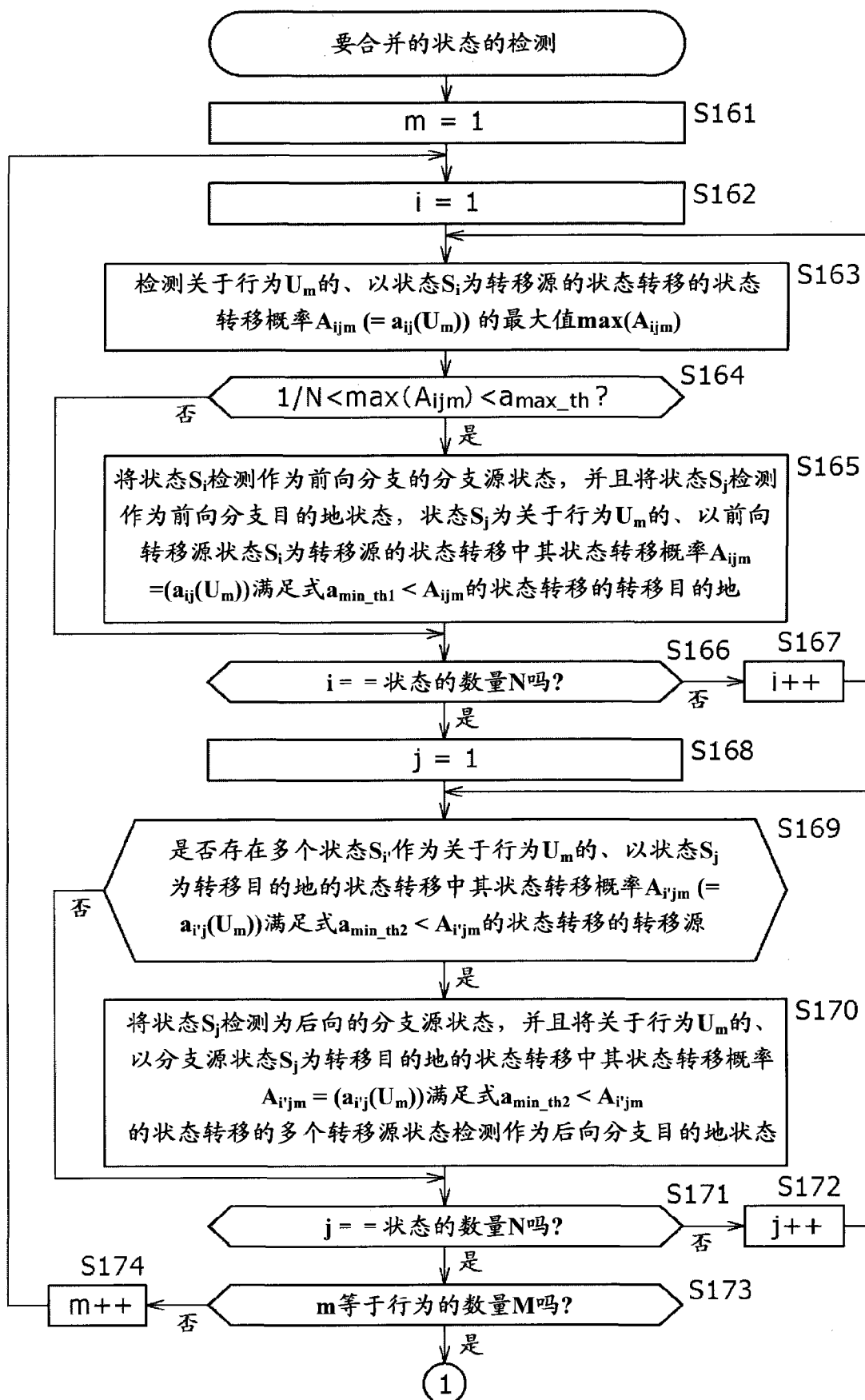


图 40

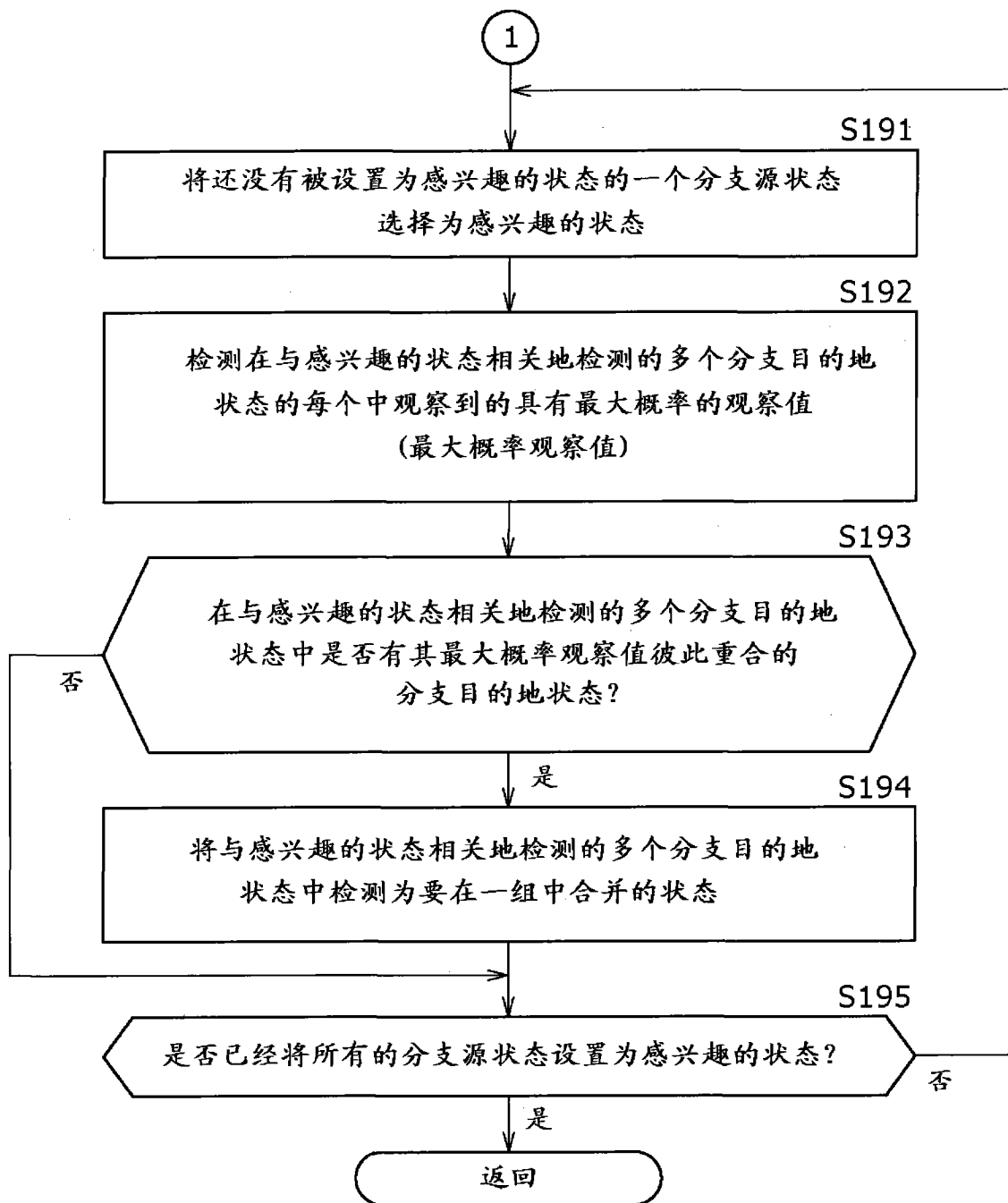


图 41

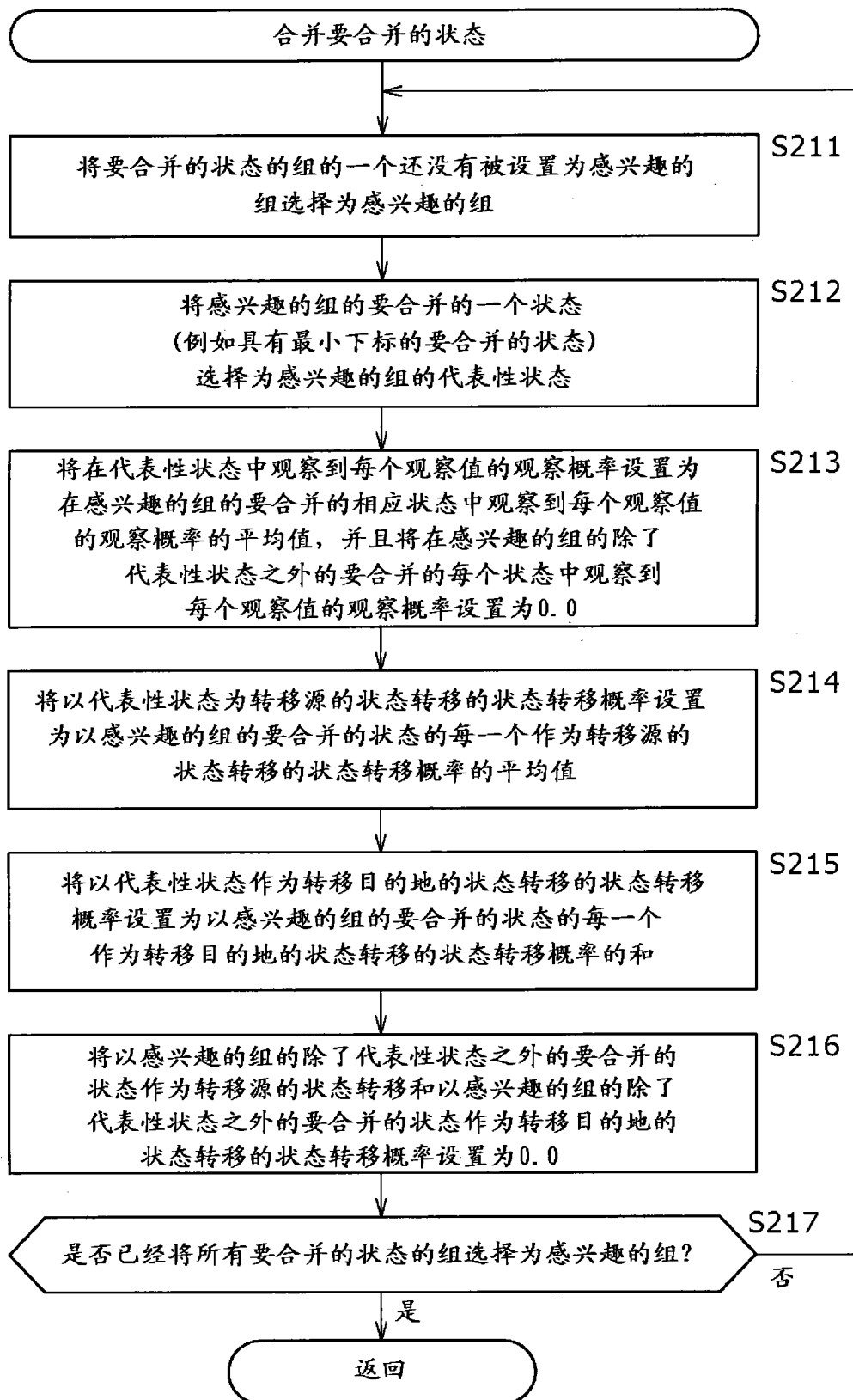


图 42

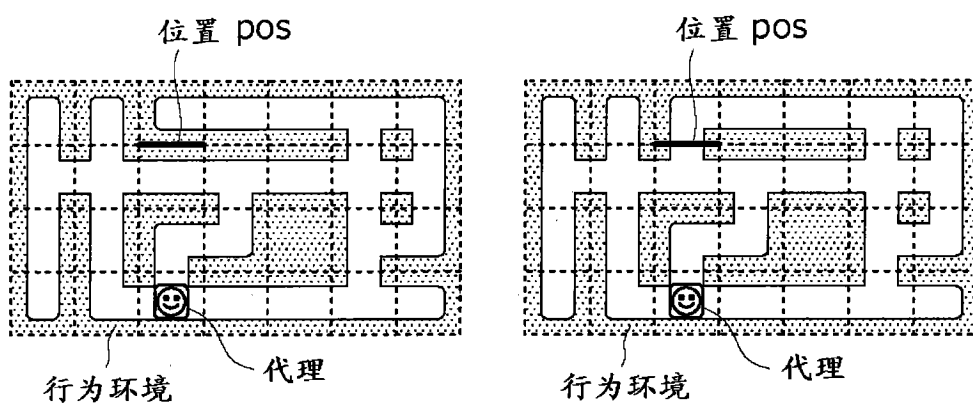


图 43A

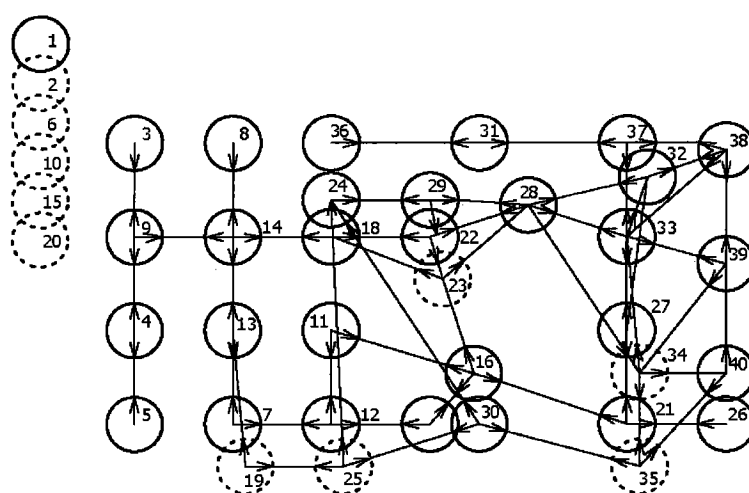


图 43B

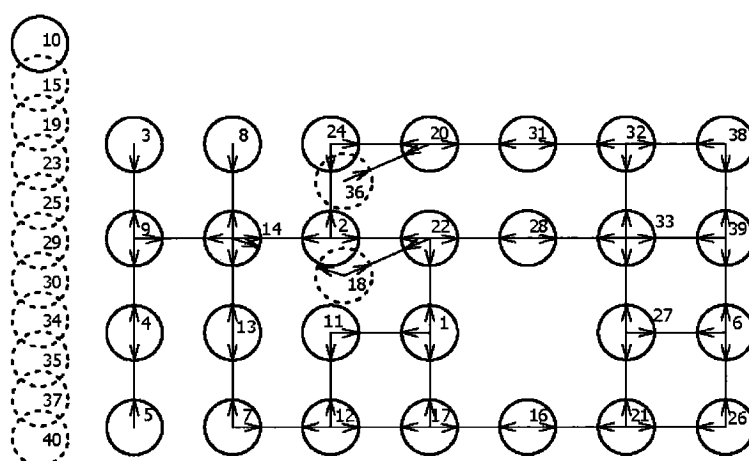


图 43C

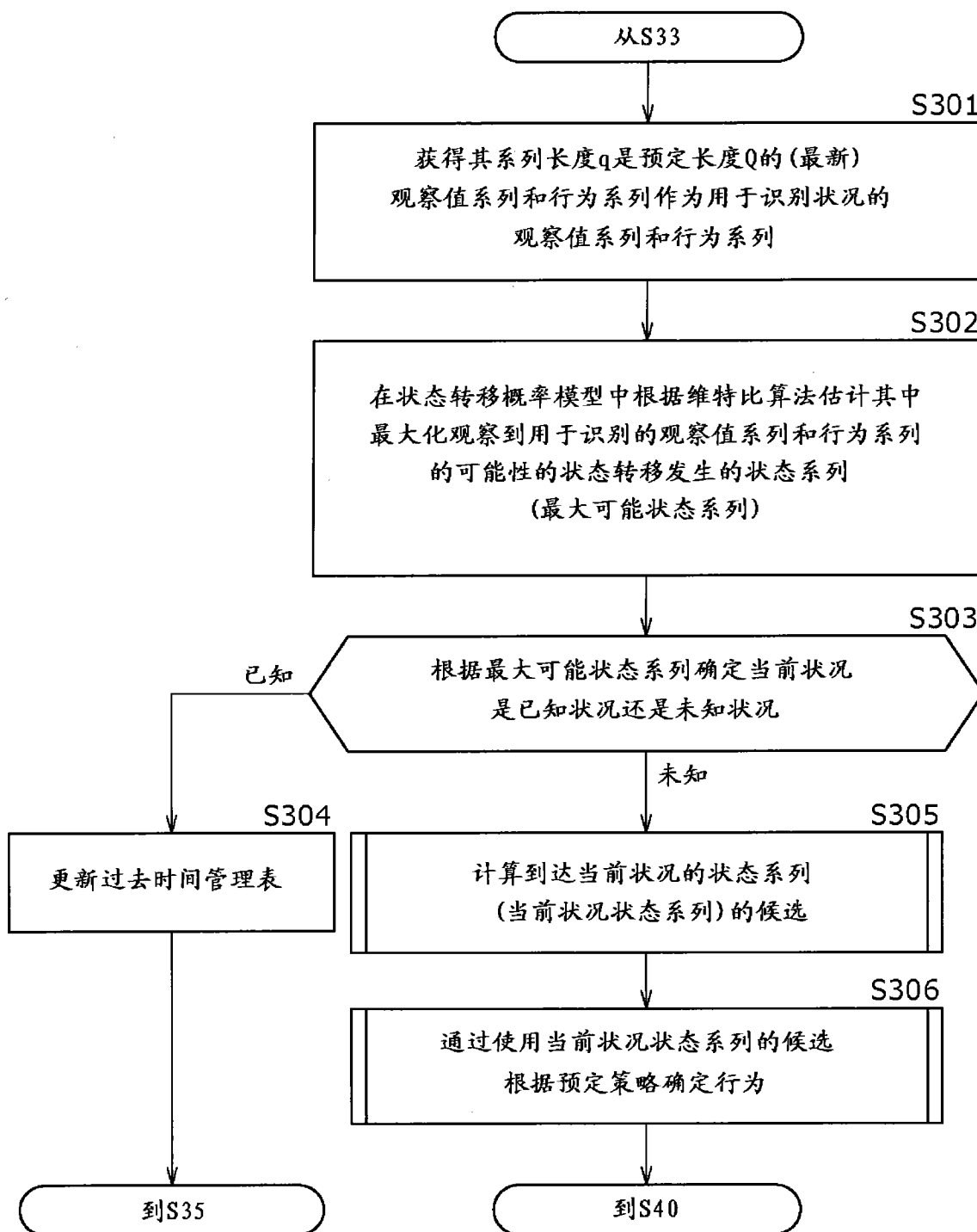


图 44

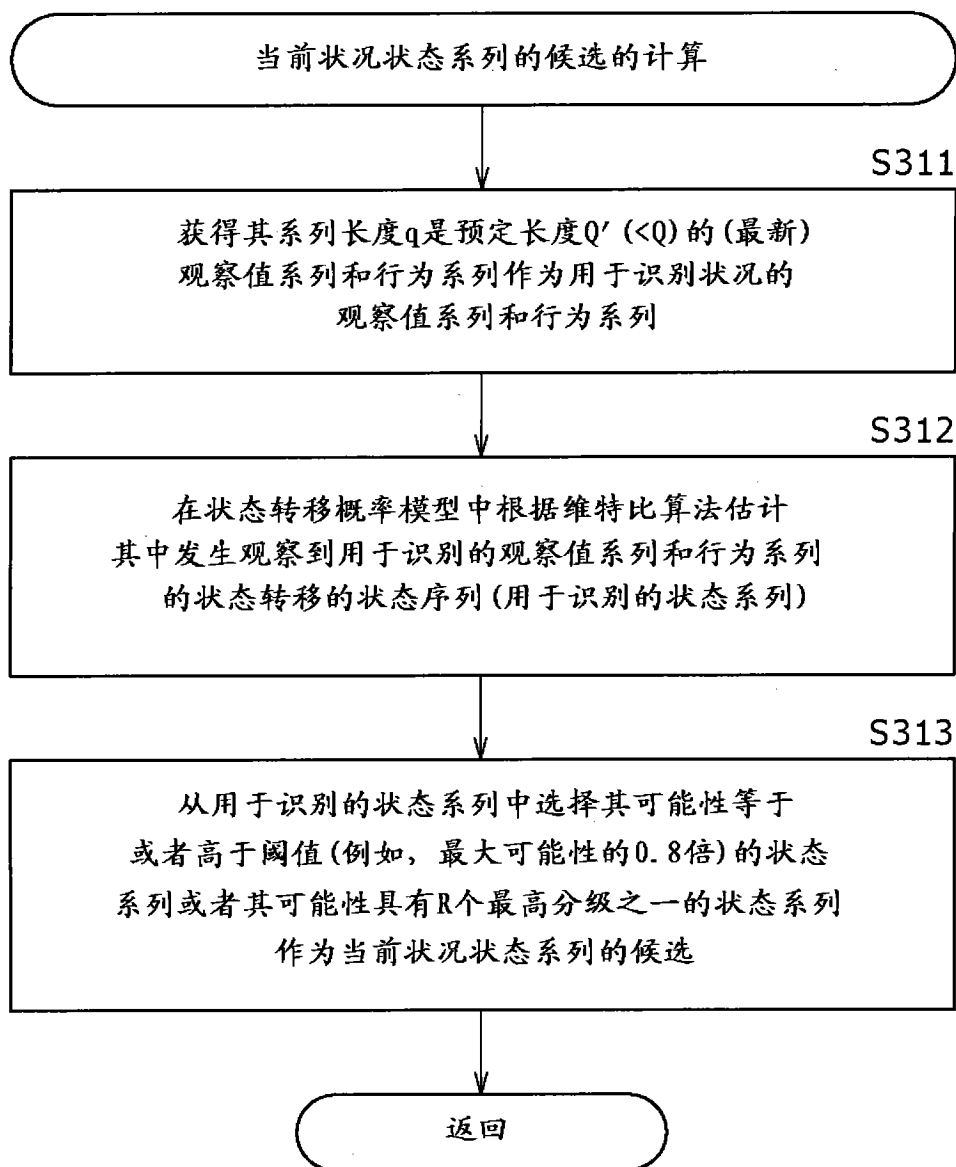


图 45

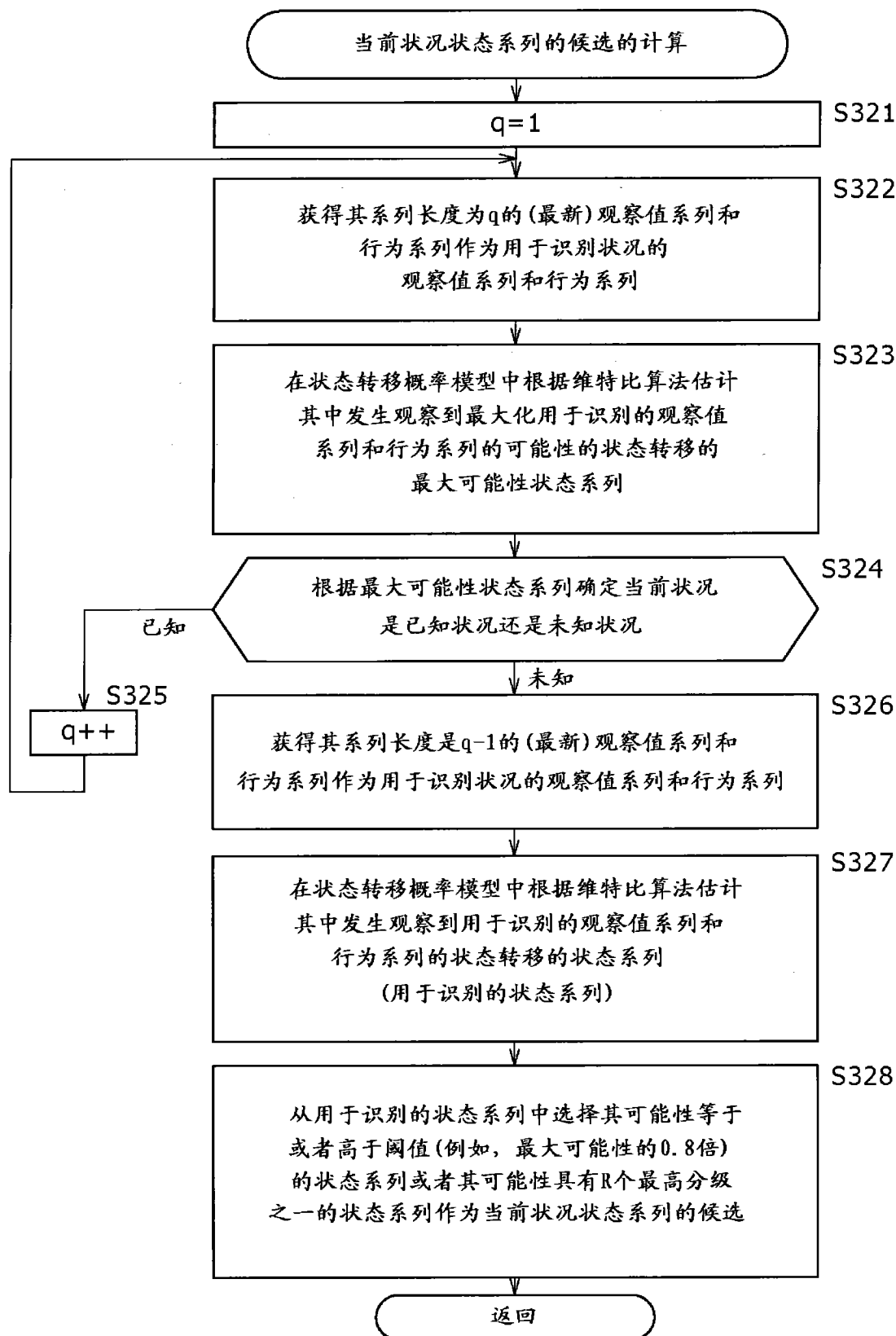


图 46

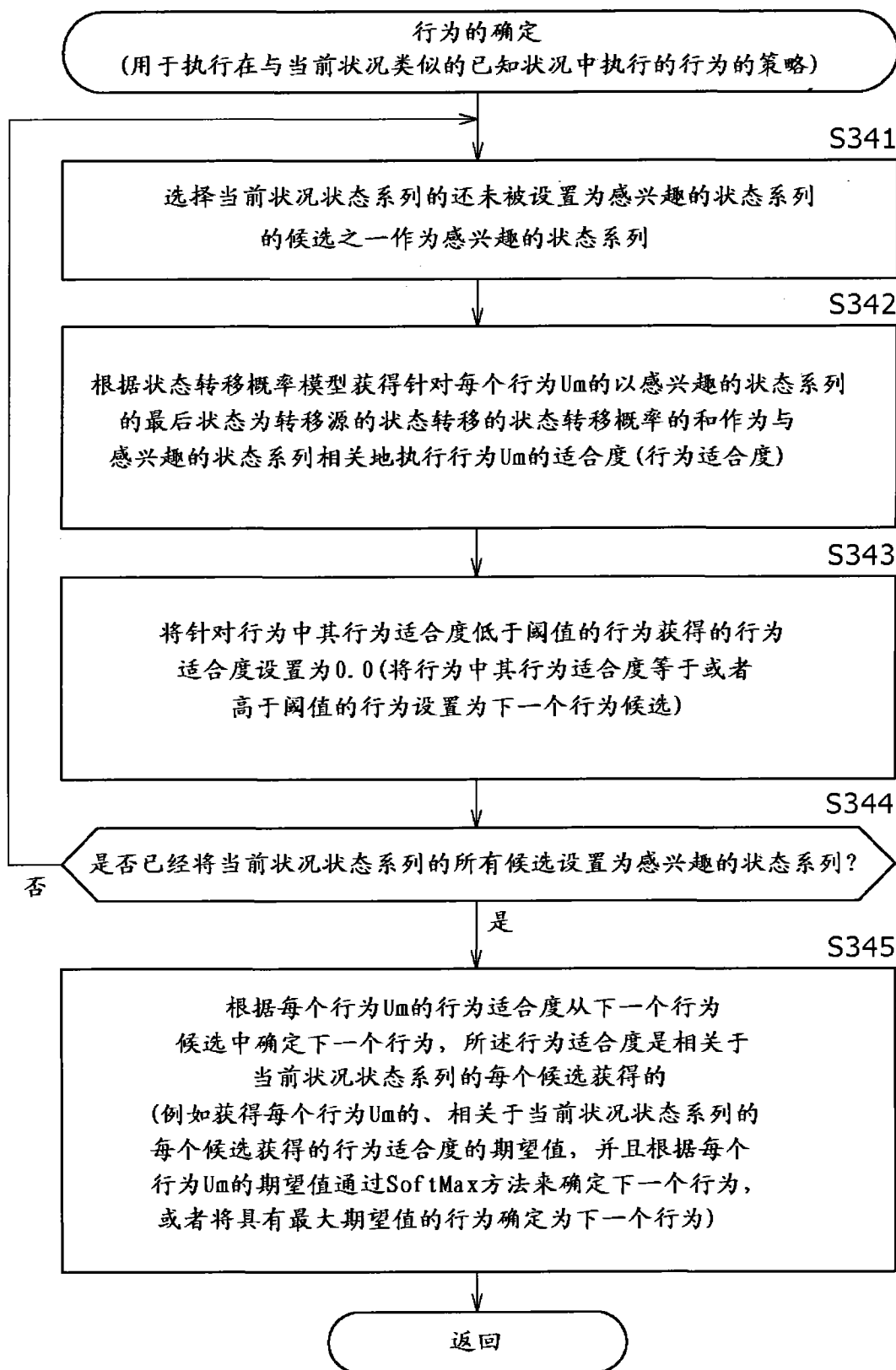


图 47

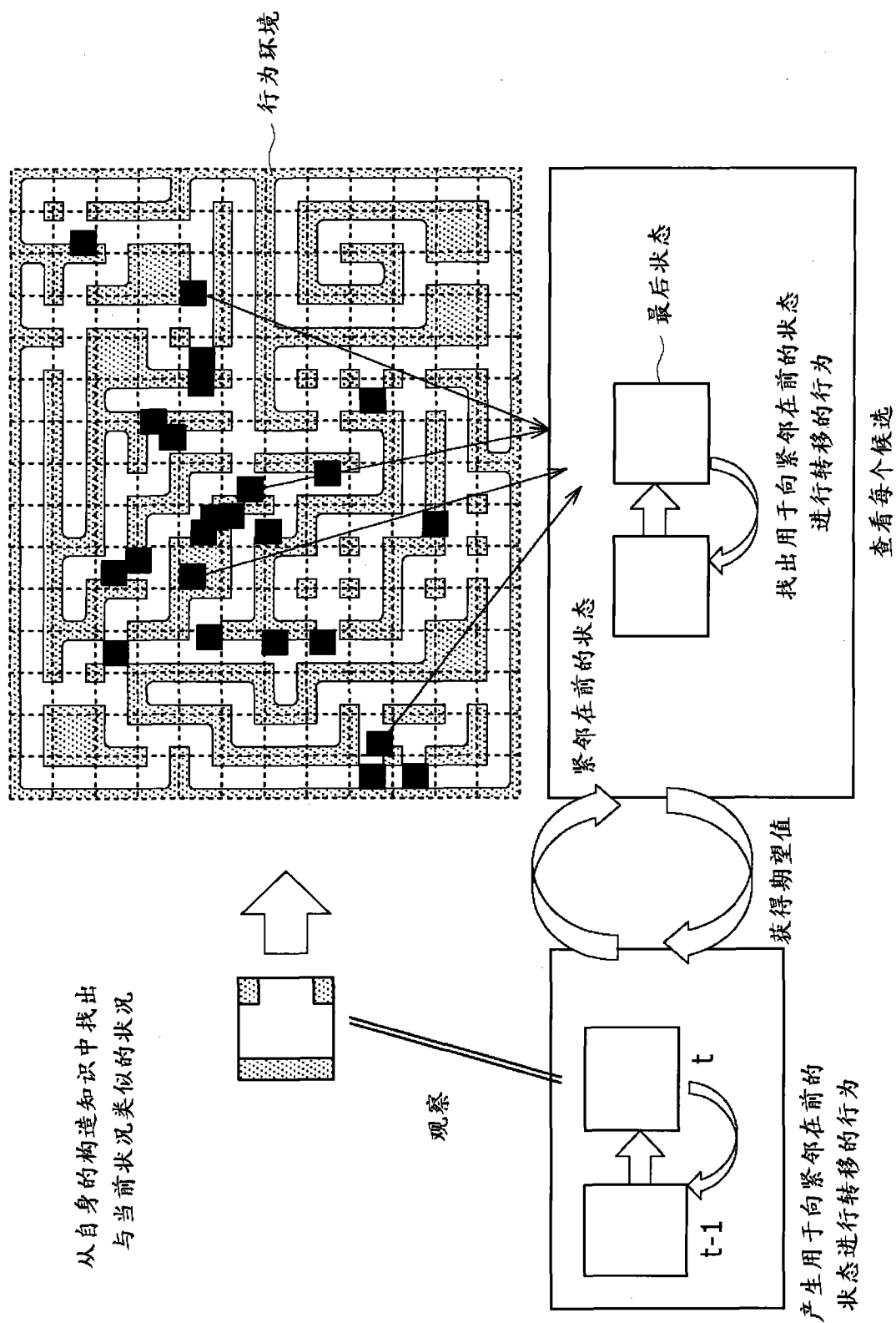


图 48

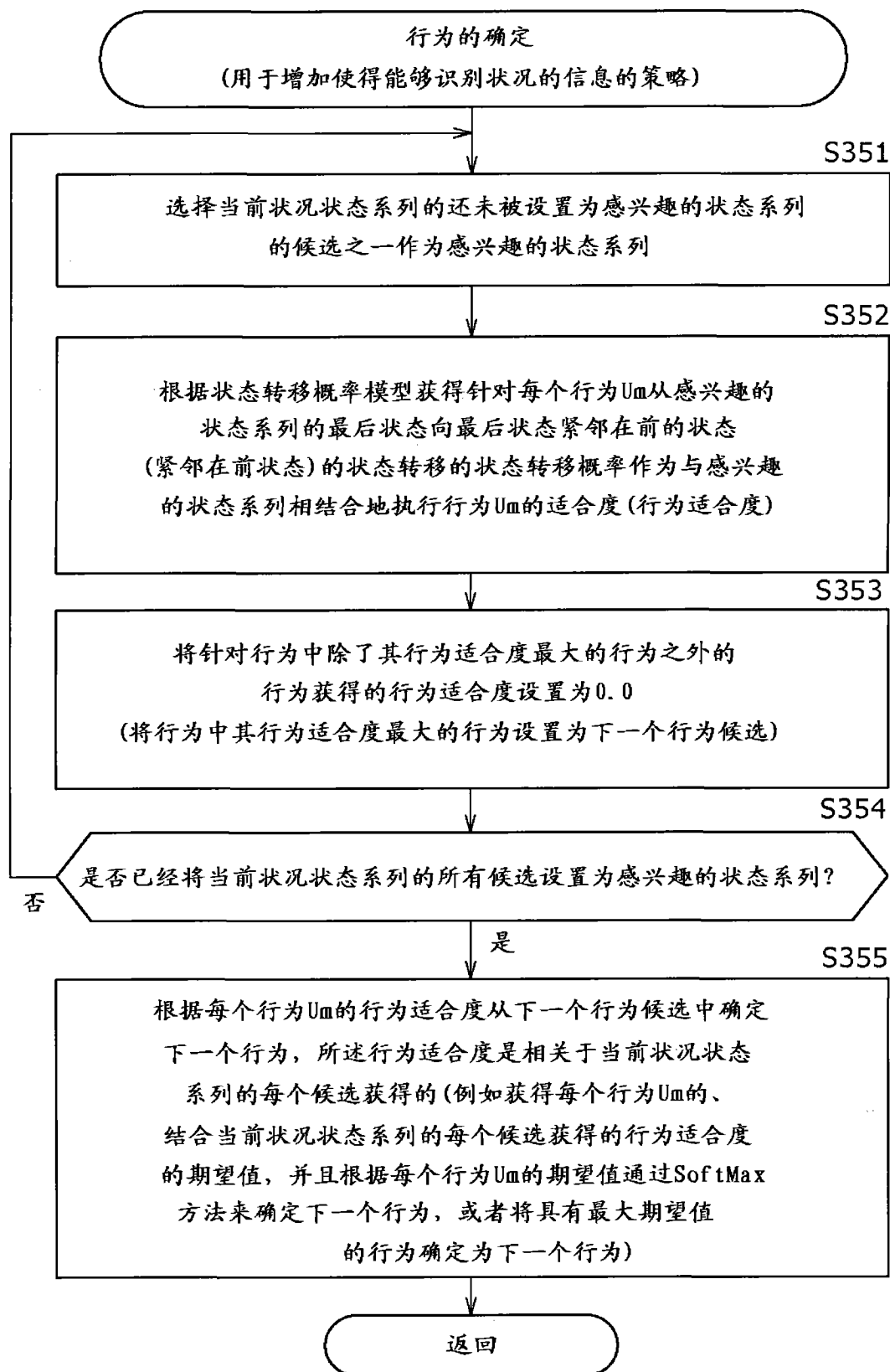


图 49

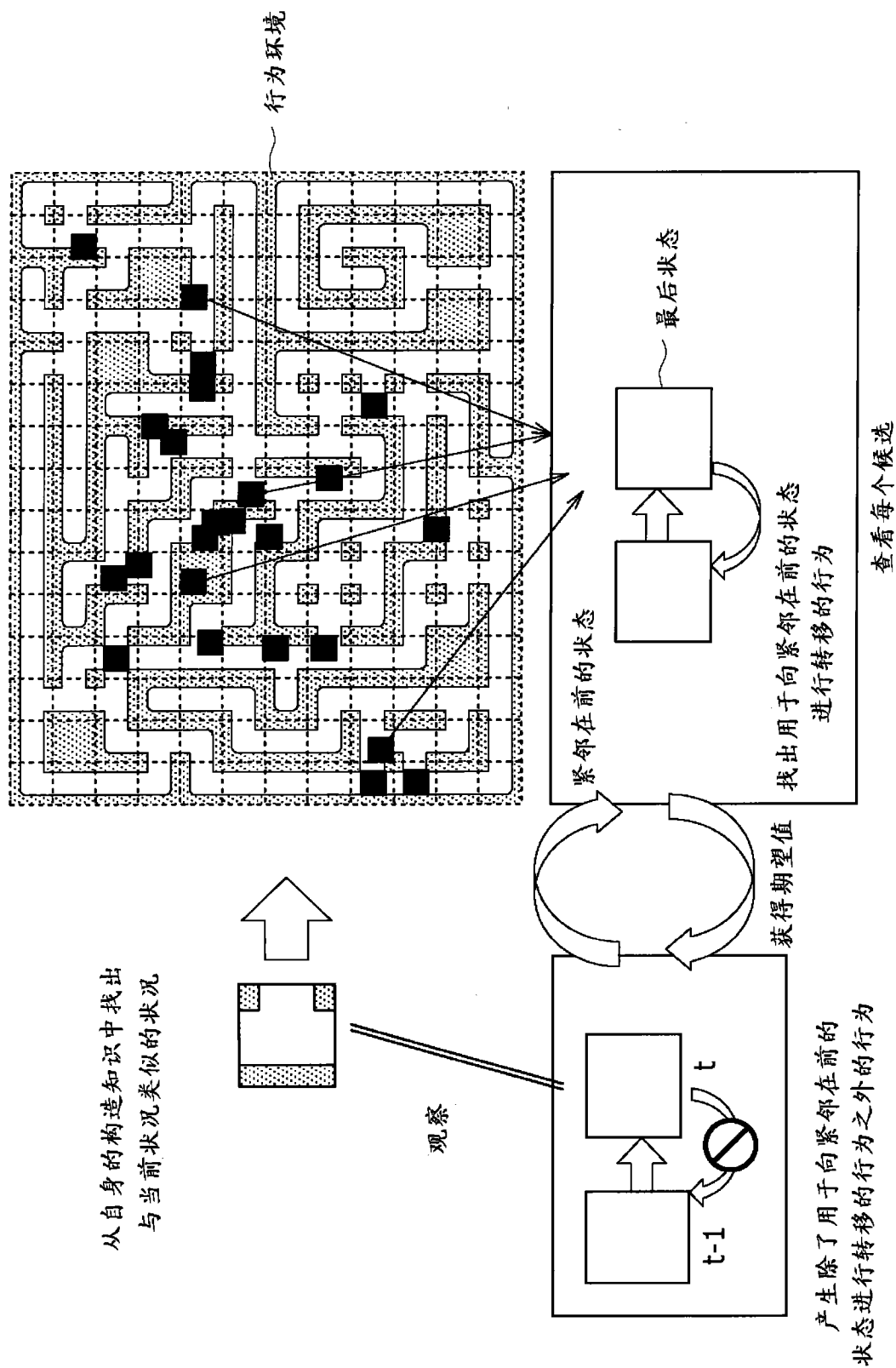


图 50

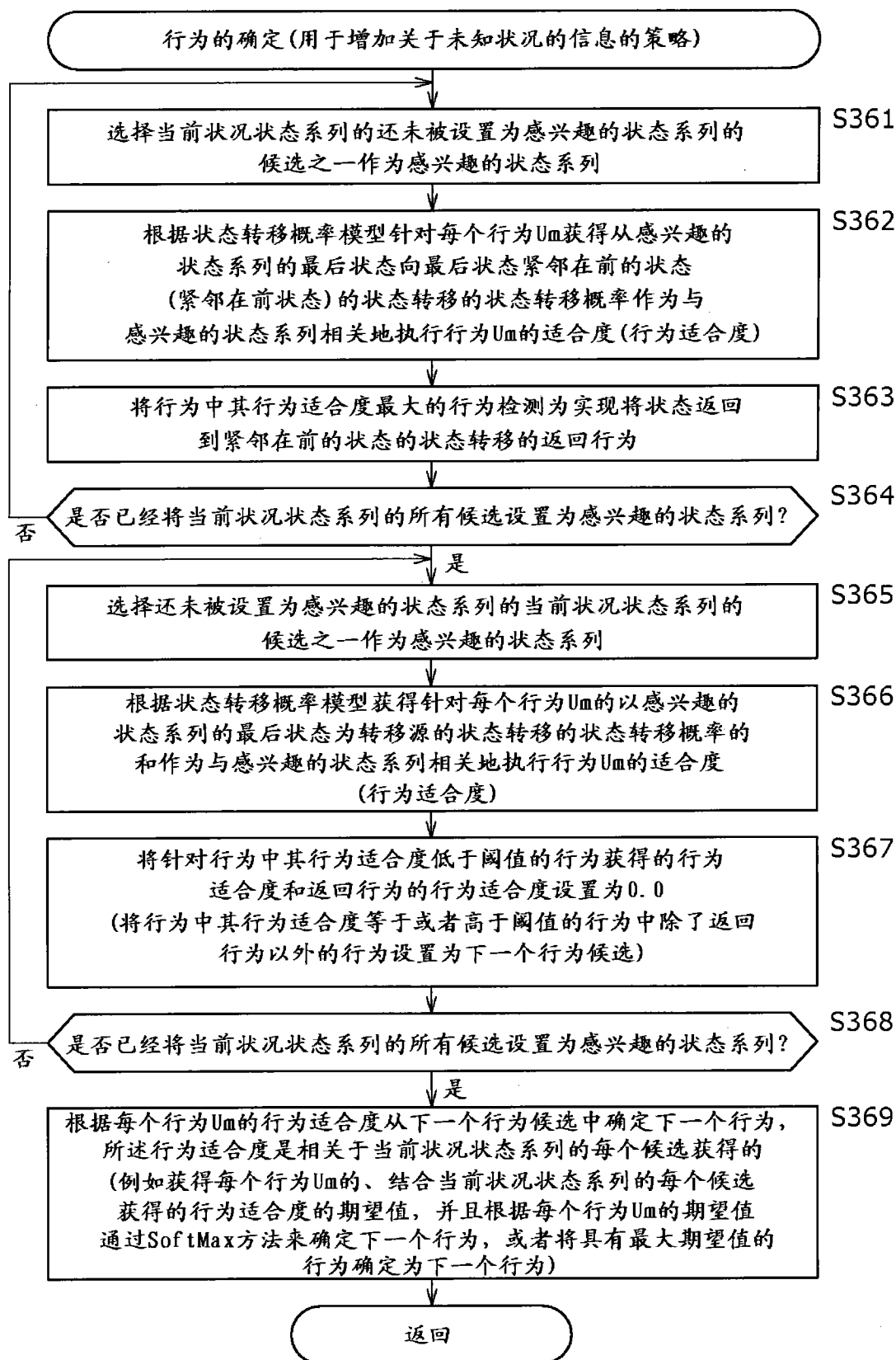


图 51

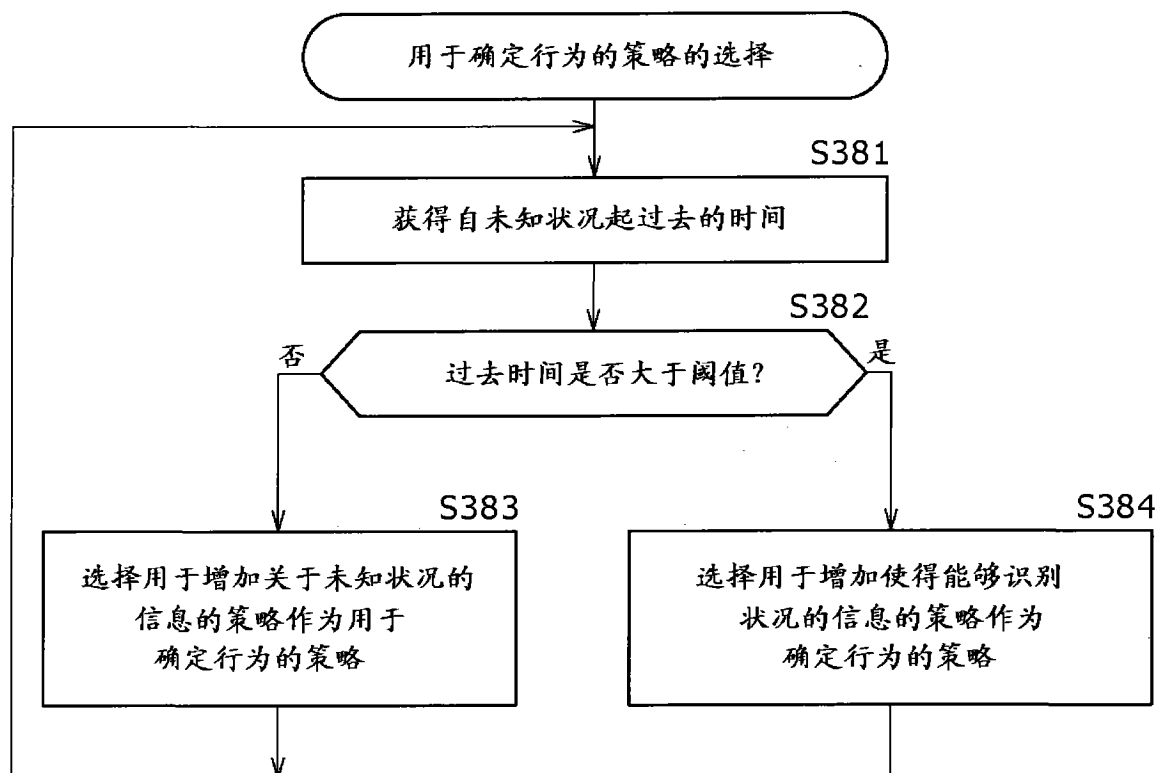


图 52

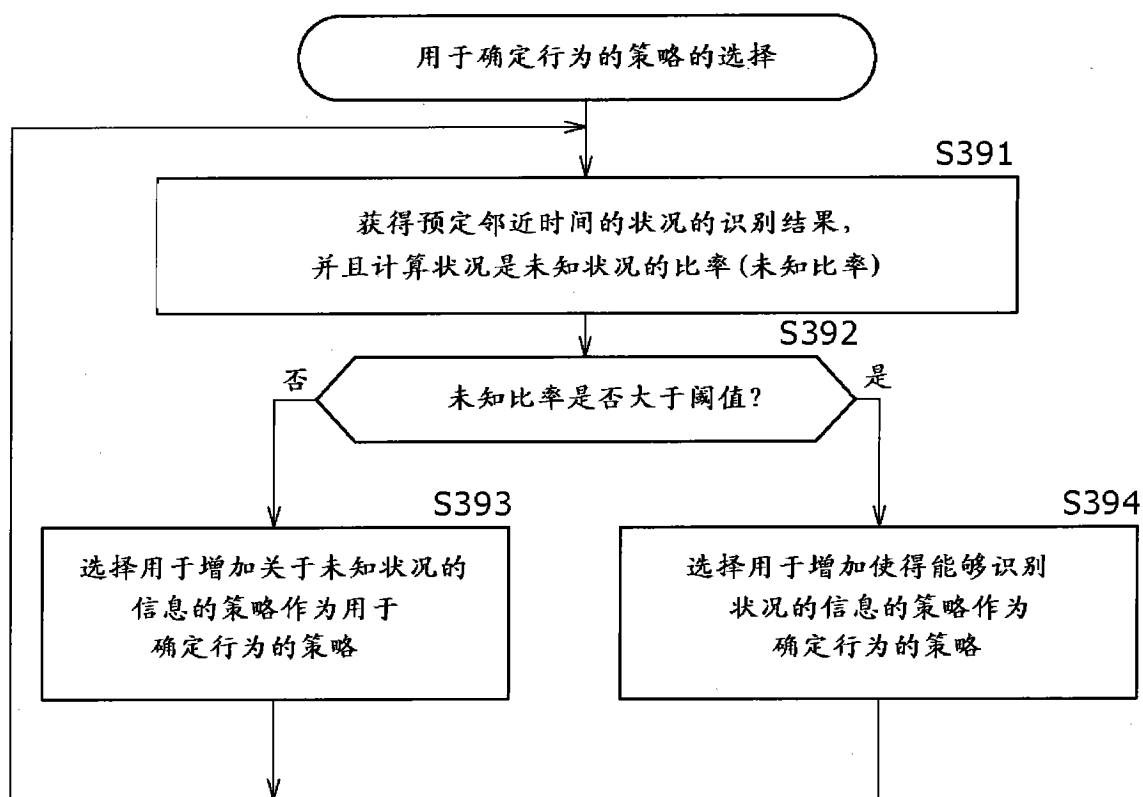


图 53

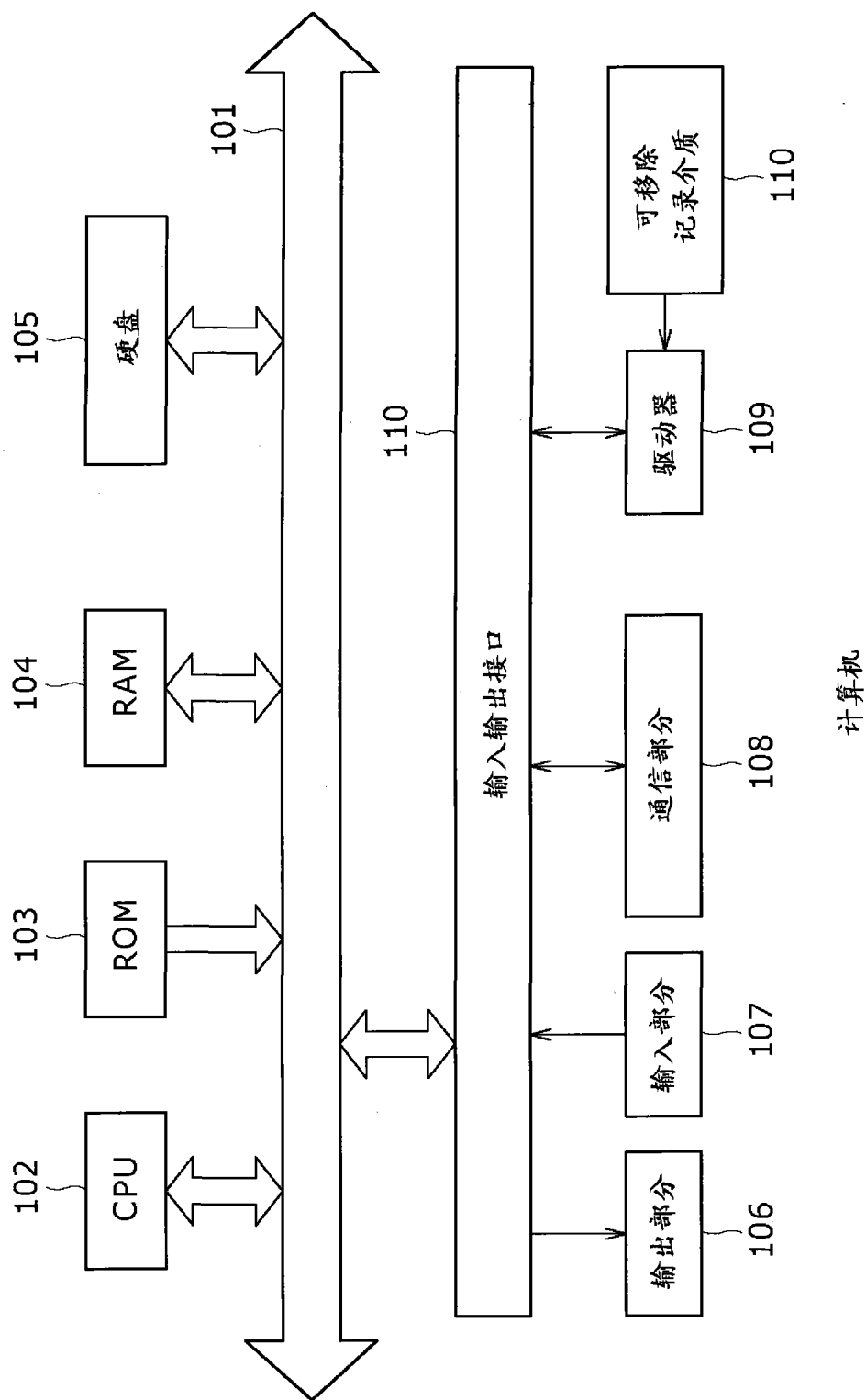


图 54