



(12)发明专利申请

(10)申请公布号 CN 111476257 A

(43)申请公布日 2020.07.31

(21)申请号 201910066435.9

(22)申请日 2019.01.24

(71)申请人 富士通株式会社

地址 日本神奈川县

(72)发明人 钟朝亮 孙俊

(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 王萍 王鹏

(51)Int.Cl.

G06K 9/62(2006.01)

G06N 3/04(2006.01)

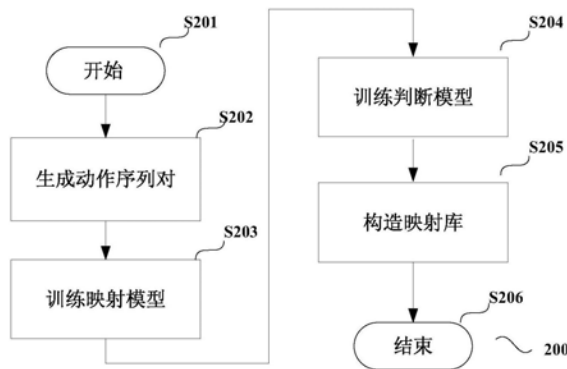
权利要求书2页 说明书15页 附图5页

(54)发明名称

信息处理方法和信息处理装置

(57)摘要

本公开涉及信息处理方法和信息处理装置，用于将能够根据观测信息执行相应的动作序列的第一智能体的处理知识转移到第二智能体。该信息处理方法包括：生成第一智能体的第一动作序列和第二智能体的第二动作序列的动作序列对，第一动作序列和第二动作序列完成相同的任务；使用所生成的动作序列对训练映射模型以使其能够根据第一智能体的动作序列生成第二智能体的动作序列；使用第一智能体的第一动作序列训练判断模型以使其能够判断第一智能体的动作序列的当前动作是否是该动作序列的最后一个动作；以及使用经训练的映射模型和经训练的判断模型构造映射库，其中映射库包括从第二智能体的观测信息到第二智能体的动作序列的映射。



1. 一种信息处理方法,用于将第一智能体的处理知识转移到第二智能体,其中所述第一智能体能够根据所述第一智能体的观测信息执行相应的动作序列,所述信息处理方法包括:

生成所述第一智能体的第一动作序列和所述第二智能体的第二动作序列的动作序列对,其中所述第一动作序列和所述第二动作序列完成相同的任务;

使用所生成的动作序列对训练映射模型,其中所述映射模型能够根据所述第一智能体的动作序列生成所述第二智能体的动作序列;

使用所述第一智能体的第一动作序列训练判断模型,其中所述判断模型能够判断所述第一智能体的动作序列的当前动作是否是该动作序列的最后一个动作;以及

使用经训练的映射模型和经训练的判断模型构造映射库,其中所述映射库包括从所述第二智能体的观测信息到所述第二智能体的动作序列的映射。

2. 根据权利要求1所述的信息处理方法,其中所述第一智能体的动作的自由度不同于所述第二智能体的动作的自由度。

3. 根据权利要求1所述的信息处理方法,其中通过使用不同的任务构造不同的所述动作序列对。

4. 根据权利要求1所述的信息处理方法,其中使用所述动作序列对训练映射模型的步骤进一步包括:

设置所述第一智能体的动作的第一索引,使用表示所述第一索引的第一索引向量表示所述第一智能体的第一动作序列;

设置所述第二智能体的动作的第二索引,使用表示所述第二索引的第二索引向量表示所述第二智能体的第二动作序列;以及

使用所述第一索引向量和所述第二索引向量训练所述映射模型。

5. 根据权利要求1所述的信息处理方法,其中使用所述第一动作序列训练判断模型的步骤进一步包括:

设置所述第一智能体的动作的第一索引,使用表示所述第一索引的第一索引向量表示所述第一智能体的第一动作序列;以及

使用所述第一索引向量训练所述判断模型。

6. 根据权利要求1所述的信息处理方法,其中

所述映射模型包括编码单元和解码单元,

所述编码单元被配置成将所述第一智能体的动作序列编码为固定长度向量,以及

所述解码单元被配置成将所述固定长度向量解码为所述第二智能体的动作序列。

7. 根据权利要求1所述的信息处理方法,其中

所述映射模型包括编码单元和解码单元,

所述编码单元被配置成将所述第一智能体的动作序列的逆序列编码为固定长度向量,以及

所述解码单元被配置成将所述固定长度向量解码为所述第二智能体的动作序列的逆序列。

8. 根据权利要求1所述的信息处理方法,其中使用经训练的映射模型和经训练的判断模型构造映射库的步骤进一步包括:

所述第一智能体根据与所述第一智能体的观测信息相关的环境信息执行由所述第一智能体的动作序列构成的动作流；

使用经训练的判断模型从所述动作流中提取所述第一智能体的动作序列；

使用经训练的映射模型根据所提取的所述第一智能体的动作序列生成所述第二智能体的动作序列；以及

构造从所述第二智能体的观测信息到所生成的所述第二智能体的动作序列的映射。

9. 根据权利要求1所述的信息处理方法, 进一步包括:

使用所述映射库对第二智能体进行训练。

10. 一种信息处理装置, 用于将第一智能体的处理知识转移到第二智能体, 其中所述第一智能体能够根据所述第一智能体的观测信息执行相应的动作序列, 所述信息处理装置包括:

生成单元, 被配置成生成所述第一智能体的第一动作序列和所述第二智能体的第二动作序列的动作序列对, 其中所述第一动作序列和所述第二动作序列完成相同的任务;

第一训练单元, 被配置成使用所生成的动作序列对训练映射模型, 其中所述映射模型能够根据所述第一智能体的动作序列生成所述第二智能体的动作序列;

第二训练单元, 被配置成使用所述第一智能体的第一动作序列训练判断模型, 其中所述判断模型能够判断所述第一智能体的动作序列的当前动作是否是该动作序列的最后一个动作; 以及

构造单元, 被配置成使用经训练的映射模型和经训练的判断模型构造映射库, 其中所述映射库包括从所述第二智能体的观测信息到所述第二智能体的动作序列的映射。

## 信息处理方法和信息处理装置

### 技术领域

[0001] 本发明总体上涉及智能体的迁移学习的技术领域,更具体地,涉及用于将第一智能体对任务的处理知识转移到具有不同动作空间的第二智能体的信息处理方法和信息处理装置。

### 背景技术

[0002] 目前,作为智能体的一种示例的智能机器在例如工业制造、外科医疗等领域具有极为广泛的应用。智能机器通常具有多关节机械手或多自由度的动作装置,能够依靠自身动力和控制能力根据观测信息智能化地执行一系列动作以完成预定的任务。

[0003] 对智能机器进行训练以使其能够根据观测信息自主完成预定的任务通常需要大量的训练样本并且耗用大量的时间,因此如果能够将经训练的智能机器所具备的处理知识转移到未经训练的智能机器上以使得该未经训练的智能机器具备相同的处理知识,那么将是极为有利的。

[0004] 然而,即使拥有相同或相近的处理能力的智能机器,其动作空间也可能不一样。例如,对于机械臂,即使它们的动作可以抵达相同的范围,但是由于它们的动作自由度(Degree of Freedom; DoF)不同,导致它们的动作空间也不同。此外,即使对于具有相同DoF的机械臂,由于连杆的尺寸不同、关节的种类不同等原因,动作空间仍可能不同。这里,将机械臂的连杆和关节等参与机械臂的动作的部件统称为执行机构。

[0005] 具体地,例如,对于4DoF机械臂,其动作空间可以是由4个关节的状态所组成的向量所构成的空间:(状态1,状态2,状态3,状态4),而对于6DoF机械臂,其动作空间可以是6个关节的状态所组成的向量所构成的空间:(状态1,状态2,状态3,状态4,状态5,状态6),其中每个关节的状态可以用例如角度来表示。

[0006] 对于上述示例,经训练的4DoF机械臂能够根据观测信息自主完成预定的任务,然而目前该4DoF机械臂所具备的处理知识难以转移到6DoF机械臂上。如果为了完成相同的任务对6DoF机械臂进行重新训练,则需要耗用大量的时间。

[0007] 因此,需要一种能够将经训练的智能体对任务的处理知识转移到具有不同动作空间的未经训练的智能体的技术。

### 发明内容

[0008] 本文提出了信息处理方法和信息处理装置,能够将经训练的智能体对任务的处理知识转移到具有不同动作空间的未经训练的智能体,从而简化具有不同的动作空间的未经训练的智能体的训练过程,从而降低训练成本,提高训练效率。

[0009] 在下文中将给出关于本公开的简要概述,以便提供关于本公开的某些方面的基本理解。应当理解,这个概述并不是关于本公开的穷举性概述。它并不是意图确定本公开的关键或重要部分,也不是意图限定本公开的范围。其目的仅仅是以简化的形式给出某些概念,以此作为稍后论述的更详细描述的前序。

[0010] 本公开的目的之一在于提供能够将经训练的智能体对任务的处理知识转移到具有不同动作空间的未经训练的智能体的信息处理方法和信息处理装置。通过根据本公开的信息处理方法和信息处理装置,能够简化具有不同的动作空间的未经训练的智能体的训练过程,从而降低训练成本,提高训练效率。

[0011] 为了实现本公开的目的,根据本公开的一个方面,提供了一种信息处理方法,用于将第一智能体的处理知识转移到第二智能体,其中第一智能体能够根据第一智能体的观测信息执行相应的动作序列,该信息处理方法包括:生成第一智能体的第一动作序列和第二智能体的第二动作序列的动作序列对,其中第一动作序列和第二动作序列完成相同的任务;使用所生成的动作序列对训练映射模型,其中映射模型能够根据第一智能体的动作序列生成第二智能体的动作序列;使用第一智能体的第一动作序列训练判断模型,其中判断模型能够判断第一智能体的动作序列的当前动作是否是该动作序列的最后一个动作;以及使用经训练的映射模型和经训练的判断模型构造映射库,其中映射库包括从第二智能体的观测信息到第二智能体的动作序列的映射。

[0012] 根据本公开的另一方面,提供了一种信息处理装置,用于将第一智能体的处理知识转移到第二智能体,其中第一智能体能够根据第一智能体的观测信息执行相应的动作序列,该信息处理装置包括:生成单元,被配置成生成第一智能体的第一动作序列和第二智能体的第二动作序列的动作序列对,其中第一动作序列和第二动作序列完成相同的任务;第一训练单元,被配置成使用所生成的动作序列对训练映射模型,其中映射模型能够根据第一智能体的动作序列生成第二智能体的动作序列;第二训练单元,被配置成使用第一智能体的第一动作序列训练判断模型,其中判断模型能够判断第一智能体的动作序列的当前动作是否是该动作序列的最后一个动作;以及构造单元,被配置成使用经训练的映射模型和经训练的判断模型构造映射库,其中映射库包括从第二智能体的观测信息到第二智能体的动作序列的映射。

[0013] 根据本公开的另一方面,提供了一种能够实现上述的信息处理方法的计算机程序。此外,还提供了具有至少计算机可读介质形式的计算机程序产品,其上记录有用于实现上述的信息处理方法的计算机程序代码。

[0014] 根据本公开的技术,能够将经训练的智能体对任务的处理知识转移到具有不同动作空间的未经训练的智能体,从而简化具有不同的动作空间的未经训练的智能体的训练过程,从而降低训练成本,提高训练效率。

## 附图说明

[0015] 参照下面结合附图对本公开实施方式的说明,会更加容易地理解本公开的以上和其他目的、特点和优点,在附图中:

[0016] 图1A和图1B分别是示出作为智能体的示例的4自由度(DoF)机械臂和6DoF机械臂及其任务空间的示意图;

[0017] 图2示出了根据本公开的实施方式的用于将第一智能体的处理知识转移到第二智能体的信息处理方法的流程图;

[0018] 图3示出了根据本公开的实施方式的使用动作序列对训练映射模型的示例处理的流程图;

[0019] 图4示出了根据本公开的实施方式的使用动作序列对训练映射模型的示例处理的示意图；

[0020] 图5示出了使用第一动作序列训练判断模型的示例处理的示意图；

[0021] 图6示出了根据本公开的实施方式的使用经训练的映射模型和判断模型构造映射库的示例处理的流程图；

[0022] 图7示出了根据本公开的实施方式的使用经训练的映射模型和判断模型构造映射库的示例处理的示意图；

[0023] 图8示出了根据本公开的实施方式的信息处理装置的结构框图；以及

[0024] 图9是示出可用来实现根据本公开的实施方式的信息处理方法和信息处理装置的通用机器的结构简图。

### 具体实施方式

[0025] 在下文中,将参照所附的说明性示图详细描述本公开的一些实施方式。在用附图标记指示附图的元件时,尽管相同的元件在不同的附图中示出,但相同的元件将由相同的附图标记表示。此外,在本公开的以下描述中,在有可能使本公开的主题不清楚的情况下,将省略对并入于本文中的已知功能和配置的详细描述。

[0026] 本文中使用的术语仅用于描述特定实施方式的目的,而非旨在限制本公开。如本文所使用的,除非上下文另外指出,否则单数形式旨在也包括复数形式。还将理解的是,说明书中使用的术语“包括”、“包含”和“具有”旨在具体说明所陈述的特征、实体、操作和/或部件的存在,但是并不排除一个或更多个其他的特征、实体、操作和/或部件的存在或添加。

[0027] 除非另有定义,否则本文中使用的包括技术术语和科学术语的所有术语具有与本发明构思所属领域技术人员通常理解的含义相同的含义。将进一步理解的是,诸如在常用词典中定义的那些术语应该被解释为具有与其在相关领域的上下文中的含义一致的含义,除非在此明确定义否则不应以理想化或过于正式的意义来解释。

[0028] 在下面的描述中,阐述了许多具体细节以提供对本公开的全面理解。本公开可以在没有这些具体细节中的一些或所有具体细节的情况下实施。在其他实例中,为了避免因不必要的细节而模糊了本公开,在附图中仅仅示出了与根据本公开的方案密切相关的部件,而省略了与本公开关系不大的其他细节。

[0029] 在下文中,将参照附图详细描述根据本公开的用于将经训练的智能体对任务的处理知识转移到具有不同动作空间的未经训练的智能体的信息处理技术。

[0030] 本公开的信息处理技术的核心思想在于建立具有不同动作空间的智能体的动作空间之间的映射关系。具体地,假设第一智能体是能够根据其观测信息执行相应的动作序列的经训练的智能体,而第二智能体是具有不同于第一智能体的动作空间的未经训练的智能体。根据本公开的技术,需要训练映射模型,用于将第一智能体的第一动作序列转换为第二智能体的第二动作序列,其中第一动作序列和和第二动作序列能够完成相同的任务。为了训练该映射模型,需要构造映射模型的训练样本集合,该训练样本集合由第一智能体的第一动作序列和第二智能体的第二动作序列的动作序列对构成。此外,由于动作序列中不存在表示动作序列结束的标志,因此还需要训练判断模型,用于判断动作序列的结束。在这一点上,可以使用第一智能体的第一动作序列作为判断模型的训练样本集合对判断模型进

行训练。最后,使用经训练的映射模型和判断模型构造映射库,第二智能体可以基于该映射库根据其观测信息自发地执行相应的动作序列,从而完成与第一智能体相同的任务。

[0031] 接下来,将参照图1至图6描述根据本公开的实施方式的用于将第一智能体的处理知识转移到第二智能体的信息处理方法。

[0032] 智能体的示例可以包括机械臂、机器人等。不同的智能体可以具有因不同的动作自由度,不同的连杆尺寸、不同的关节种类而导致的不同的动作空间。

[0033] 作为智能体的具体示例,图1A和图1B分别示出了4DoF机械臂和6DoF机械臂及其任务空间的示意图。在本公开的实施方式中,任务(Task)可以被限定为包含起始位置和结束位置的对。具体地,如图1中所示,这里所说的位置可以由三维空间中机械臂的执行机构的末端能够达到的范围内的坐标来表示。例如,以机械臂的底座为原点,可以定义以下任务:

[0034]  $\text{Task}\langle P1, P2\rangle = \langle (0.2, 0.4, 0.3), (0.1, 0.2, 0.4)\rangle$

[0035] 该任务的含义是,将机械臂的执行机构的末端从坐标P1(0.2,0.4,0.3)(起始位置)移动到坐标P2(0.1,0.2,0.4)(结束位置)。这里,可以取任意长度量纲作为单位。这里,将表示所有任务的起始位置坐标和结束位置坐标组成的对的集合限定为任务空间。任务空间是由起始位置和结束位置构成的二维空间。

[0036] 这里,4DoF机械臂是经训练的第一智能体的具体示例,其在下文中还被称为源机械臂,而6DoF机械臂是未经训练的第二智能体的具体示例,其在下文中还被称为目标机械臂。第一智能体和第二智能体可以具有相同的任务空间。

[0037] 图2示出了根据本公开的实施方式的用于将第一智能体的处理知识转移到第二智能体的信息处理方法200的流程图。这里,第一智能体能够根据其观测信息执行相应的动作序列。根据本公开的信息处理方法200开始于步骤S201。在步骤S202中,生成第一智能体的第一动作序列和第二智能体的第二动作序列的动作序列对,其中第一动作序列和第二动作序列完成相同的任务。接下来,在步骤S203中,使用所生成的动作序列对训练映射模型,其中映射模型能够根据第一智能体的动作序列生成第二智能体的动作序列。随后,在步骤S204中,使用第一智能体的第一动作序列训练判断模型,其中判断模型能够判断第一智能体的动作序列的当前动作是否是该动作序列的最后一个动作。随后,在步骤S205中,使用经训练的映射模型和经训练的判断模型构造映射库,其中映射库包括从第二智能体的观测信息到第二智能体的动作序列的映射。最后,信息处理方法200结束于步骤S206。

[0038] 下面使用图1中所示的分别作为第一智能体和第二智能体的具体示例的4DoF机械臂和6DoF机械臂详细描述根据本公开的实施方式的信息处理方法200的各个步骤S202至S205的示例实施方式。

[0039] 在步骤S202中,生成第一智能体的第一动作序列和第二智能体的第二动作序列的动作序列对,其中第一动作序列和第二动作序列完成相同的任务。如上文所述,为了训练映射模型,需要构建动作序列对集合作为映射模型的训练样本集合。动作序列对由第一智能体的第一动作序列和第二智能体的第二动作序列的对,其中第一动作序列和第二动作序列可以完成相同的任务。此外,为了便于处理,成对的第一动作序列和第二动作序列通过相同的形式语法来表示。特别地,成对的第一动作序列和第二动作序列可以具有不同的长度,因而这两个动作序列中的动作可以不具有一一对应的关系。

[0040] 为了构建动作序列对集合作为映射模型的训练样本集合,需要随机地从任务空间

中对任务进行采样。根据本公开的实施方式,可以通过使用不同的任务构造不同的动作序列对。

[0041] 具体地,针对从任务空间中采样的每个任务,获得该任务的起始位置和结束位置。随后,将起始位置和结束位置输入到动作规划工具中,该动作规划工具能够根据表示任务的起始位置和结束位置自动规划出相对应的动作轨迹,而该动作轨迹中的每个动作所组成的序列即为动作序列。这里,动作规划工具可以使用本领域已知的动作规划工具,例如 MoveIt,因而不作进一步的详细描述。

[0042] 对于图1所示的示例,作为第一智能体的示例的4DoF源机械臂的动作序列是第一动作序列,又被称为源动作序列,而作为第二智能体的示例的6DoF目标机械臂的动作序列是第二动作序列,又被称为目标动作序列。

[0043] 对于所采用的每个任务,分别由第一智能体和第二智能体分别执行该任务,以分别获得第一动作序列和第二动作序列,从而构成动作序列对。根据本公开的实施方式,在所获得第一动作序列和第二动作序列的末尾添加动作序列结束标志EOS。

[0044] 例如,对于从任务空间中采样的任务 $\langle (0.2, 0.4, 0.3), (0.1, 0.2, 0.4) \rangle$ ,分别利用作为第一智能体的示例的4DoF源机械臂和作为第二智能体的示例的6DoF目标机械臂执行该任务。

[0045] 这里,假设机械臂的各个关节的状态用角度来表示,精度为 $1^\circ$ 。每个动作各个关节的角度的最大活动行程为 $2^\circ$ 。

[0046] 4DoF源机械臂执行该任务后,可以产生源动作序列,即第一动作序列 $S = [a_{11}, a_{12}, a_{13}]$ 。此外,6DoF目标机械臂执行该任务后,可以产生目标动作序列,即第二动作序列 $T = [a_{21}, a_{22}, a_{23}, a_{24}]$ 。

[0047] 其中源动作序列 $S$ 中各个动作的取值如下:

[0048]  $a_{11} = (55^\circ, 62^\circ, 71^\circ, 43^\circ)$ ;

[0049]  $a_{12} = (53^\circ, 64^\circ, 69^\circ, 42^\circ)$ ;

[0050]  $a_{13} = (51^\circ, 66^\circ, 67^\circ, 41^\circ)$ 。

[0051] 目标动作序列 $T$ 中各个动作的取值如下:

[0052]  $a_{21} = (42^\circ, 11^\circ, 27^\circ, 78^\circ, 52^\circ, 30^\circ)$ ;

[0053]  $a_{22} = (40^\circ, 13^\circ, 28^\circ, 79^\circ, 54^\circ, 32^\circ)$ ;

[0054]  $a_{23} = (38^\circ, 15^\circ, 30^\circ, 80^\circ, 56^\circ, 34^\circ)$ ;

[0055]  $a_{24} = (36^\circ, 17^\circ, 32^\circ, 80^\circ, 58^\circ, 35^\circ)$ 。

[0056] 对于源动作序列 $S$ ,动作 $a_{11}$ 为源机械臂在该任务的起始位置执行的动作,然后依次执行动作 $a_{12}, a_{13}$ 。在4DoF源机械臂执行完动作 $a_{13}$ 之后,其执行机构的末端到达结束位置,从而完成该任务。具体地,以动作 $a_{11}$ 为例, $(55^\circ, 62^\circ, 71^\circ, 43^\circ)$ 依次为4DoF源机械臂的4个关节的关节状态。当4DoF源机械臂执行动作 $a_{12}$ 时,第一个关节角度减小 $2^\circ$ ,第二个关节角度增大 $2^\circ$ ,第三个关节角度减小 $2^\circ$ ,并且第四个关节角度减小 $1^\circ$ 。

[0057] 6DoF目标机械臂的目标动作序列中的各个动作与此类似,但是其关节的数目为6。

[0058] 随后,将 $S$ 和 $T$ 组合成动作序列对 $\langle S, T \rangle$ ,然后将其加入动作序列对集合 $C$ 。 $C = \{\langle S, T \rangle\}$ ,其中 $S$ 为源机械臂执行采样任务后产生的第一动作序列, $T$ 为目标机械臂执行同一采样任务后产生的第二动作序列。



[0059] 通过从任务空间中采样不同的任务并且分别使第一智能体和第二智能体执行该任务,可以获得动作序列对以构成动作序列对集合用作映射模型的训练样本集合。组成映射模型的训练样本集合的动作序列对的数目可以是任意的。相对较多的动作序列对的数目可以获得对映射模型的更好的训练效果,但是训练成本也相应较高。因此,可以根据具体应用确定需要获得的动作序列对的数目。

[0060] 随后,在步骤S203中,使用所生成的动作序列对训练映射模型,训练的目的在于使得映射模型能够根据第一智能体的动作序列生成第二智能体的动作序列。

[0061] 图3示出了根据本公开的实施方式的使用动作序列对训练映射模型的示例处理300的流程图。该处理300开始于步骤S301。

[0062] 随后,在步骤S302中,设置第一智能体的动作的第一索引,使用表示第一索引的第一索引向量表示第一智能体的第一动作序列。此外,在步骤S303中,设置第二智能体的动作的第二索引,使用表示第二索引的第二索引向量表示第二智能体的第二动作序列。第一索引向量和第二索引向量是分别表示第一智能体的动作和第二智能体的动作的长度相同的固定长度向量。应注意,步骤S302和步骤S303的执行顺序可以是任意的,即可以首先执行步骤S302,随后执行步骤S303,也可以首先执行步骤S303,随后执行步骤S302,或者还可以并行执行步骤S302和S303。

[0063] 根据本公开的实施方式,为了对映射模型进行训练,基于所构建的动作序列对集合,针对每个序列对中的源动作序列(即第一动作序列)中的每个动作,为其在词典中设置第一索引,从而构建源动作词典。类似地,针对每个序列对中的目标动作序列(即第二动作序列)中的每个动作,建立索引,构建目标动作词典。

[0064] 针对第一智能体,可以针对所获得的所有第一动作序列中的每个动作设置相应的第一索引。例如,对于上文所述的作为第一智能体的示例的4DoF源机械臂的第一动作序列 $S = [a_{11}, a_{12}, a_{13}]$ ,可以设置如下的第一索引

[0065]  $(55^\circ, 62^\circ, 71^\circ, 43^\circ) \rightarrow 1$

[0066]  $(53^\circ, 64^\circ, 69^\circ, 42^\circ) \rightarrow 2$

[0067]  $(51^\circ, 66^\circ, 67^\circ, 41^\circ) \rightarrow 3$

[0068] .....

[0069] 此外,针对第二智能体,可以针对所获得的所有第二动作序列中的每个动作设置相应的第二索引。例如,对于上文所述的作为第二智能体的示例的6DoF目标机械臂的第二动作序列 $T = [a_{21}, a_{22}, a_{23}, a_{24}]$ ,可以设置如下的第二索引。

[0070]  $(42^\circ, 11^\circ, 27^\circ, 78^\circ, 52^\circ, 30^\circ) \rightarrow 1$

[0071]  $(40^\circ, 13^\circ, 28^\circ, 79^\circ, 54^\circ, 32^\circ) \rightarrow 2$

[0072]  $(38^\circ, 15^\circ, 30^\circ, 80^\circ, 56^\circ, 34^\circ) \rightarrow 3$

[0073]  $(36^\circ, 17^\circ, 32^\circ, 80^\circ, 58^\circ, 35^\circ) \rightarrow 4$

[0074] .....

[0075] 这里,所设置的第一索引和第二索引是一个整数,不便于映射模型的训练,因而可以将作为整数的第一索引和第二索引转换为向量。本领域中最简单的方法是独热编码技术(one-hot encoding),即索引向量维度等于全部索引数目,即与词典的大小相同,其中在索引向量中与相应的索引对应的元素取值为1,其他所有元素取值为0。

[0076] 然而,独热编码技术可能会占用大量的存储的存储空间。因此,优选地,可以采用词嵌入(word embedding)技术将第一索引和第二索引转换为固定长度的每个维度取值为实数的向量。这里,词嵌入技术可以使用本领域已知的词嵌入技术,例如Word2Vec,因而不作进一步的详细描述。

[0077] 例如,对于上文所述的作为第一智能体的示例的4DoF源机械臂的每个动作的第一索引,可以将其转换为如下的作为4维实数向量的第一索引向量。

[0078] 1→(0.6897,0.314,0.4597,0.6484)

[0079] 2→(0.6572,0.7666,0.8468,0.3075)

[0080] 3→(0.1761,0.0336,0.1119,0.7791)

[0081] ……

[0082] 此外,例如对于上文所述的作为第二智能体的示例的6DoF源机械臂的每个动作的第二索引,可以将其转换为如下的作为4维实数向量的第二索引向量。

[0083] 1→(0.494,0.6018,0.2934,0.0067)

[0084] 2→(0.0688,0.8565,0.9919,0.4498)

[0085] 3→(0.647,0.0328,0.7988,0.7429)

[0086] 4→(0.1579,0.2932,0.9996,0.0464)

[0087] ……

[0088] 通过上述处理,第一动作序列可以用第一索引向量来表示,而第二动作序列可以用第二索引向量来表示。

[0089] 接下来,在步骤S304中,使用第一索引向量和第二索引向量训练映射模型。

[0090] 根据本公开的实施方式,映射模型可以包括编码单元和解码单元,其中编码单元可以将第一智能体的动作序列编码为固定长度向量,并且解码单元可以将固定长度向量解码为第二智能体的动作序列。

[0091] 图4示出了根据本公开的实施方式的使用动作序列对训练映射模型的示例处理的示意图。

[0092] 如图4所示,映射模型包含编码单元和解码单元两部分。根据本公开的实施方式,编码单元和解码单元分别可以由递归神经网络(RNN)模型实现。递归神经网络是具有树状阶层结构且网络节点按其连接顺序对输入信息进行递归的人工神经网络,是深度学习算法之一。

[0093] 此外,根据本公开的实施方式,也可以使用作为改进的递归神经网络的长短期记忆(LSTM)模型或门控循环单元(GRU)模型来实现构成映射模型的编码单元和解码单元。

[0094] 鉴于RNN模型、LSTM模型和GRU模型对于本领域技术人员是已知的,因此为了简洁起见,本文仅对其在本公开的实施方式中的应用进行描述,而不对其原理进行更详细的描述。

[0095] 如图4中所示,例如,对于第一动作序列 $S=[a_{11}, a_{12}, a_{13}]$ ,在时间 $t_0$ 将与动作 $a_{11}$ 对应的第一索引向量,例如(0.6897,0.314,0.4597,0.6484)输入至编码单元,得到在时间 $t_0$ 的隐含状态 $v_0$ 。随后,在时间 $t_1$ 将与动作 $a_{12}$ 对应的第一索引向量,例如(0.6572,0.7666,0.8468,0.3075)和在时间 $t_0$ 的隐含状态 $v_0$ 输入解码单元,得到在时间 $t_1$ 的隐含状态 $v_1$ 。随后,在时间 $t_2$ 将与动作 $a_{13}$ 对应的第一索引向量,例如(0.1761,0.0336,0.1119,0.7791)和

在时间 $t_1$ 的隐含状态 $v_1$ 输入解码单元,得到在时间 $t_2$ 的隐含状态 $v_2$ 。随后,在时间 $t_2$ 将表示第一动作序列结束的结束标志<EOS>向量和在时间 $t_2$ 的隐含状态 $v_2$ 输入解码单元,此时编码单元结束运行,并且输出最后的隐含状态 $v$ 。

[0096] 接下来,对于第二动作序列 $T=[a_{21}, a_{22}, a_{23}, a_{24}]$ ,在时间 $t_0$ 将编码单元输出的隐含状态 $v$ 以及表示解码开始的开始标志<START>向量输入编码单元,得到在目标动作词典上的概率分布。根据该概率分布,以及动作 $a_{21}$ 的第二索引向量,可以得到预测出动作 $a_{21}$ 的概率 $P(a_{21}|v)$ 。依此类推,可以得到第二动作序列 $T$ 中剩余的每个动作 $a_{22}$ 、 $a_{23}$ 、 $a_{24}$ 的预测正确的概率 $P(a_{22}|v, a_{21})$ 、 $P(a_{23}|v, \dots, a_{22})$ 、 $P(a_{24}|v, \dots, a_{23})$ 。随后,将与每个动作对应的预测正确的概率相乘,便得到第二动作序列预测正确的概率。此外,与编码单元类似,在每个时间步骤,仅有隐含状态被传递到下一个时间步骤的解码处理。

[0097] 下面以LSTM模型为例对编码单元和解码单元的实现进行简单的说明,采用如GRU模型的其他RNN模型的实现方式与此相似,因而这里不作进一步的说明。

[0098] LSTM模型能够通过其记忆单元学习长时间范围的依赖关系,其通常包括四个单元,即输入门 $i_t$ ,输出门 $o_t$ ,遗忘门 $f_t$ 和存储状态 $c_t$ ,其中 $t$ 表示当前时间步骤。存储状态 $c_t$ 根据上一个时间步骤的状态影响其他单元的当前状态。遗忘门 $f_t$ 可用于确定应放弃哪些信息。上述过程可以用下式表示

$$[0099] \quad i_t = \sigma(W_{(i,x)}x_t + W_{(i,h)}h_{t-1} + b_i)$$

$$[0100] \quad f_t = \sigma(W_{(f,x)}x_t + W_{(f,h)}h_{t-1} + b_f)$$

$$[0101] \quad g_t = \tanh(W_{(g,x)}x_t + W_{(g,h)}h_{t-1} + b_g)$$

$$[0102] \quad c_t = i_t g_t + f_t \odot c_{t-1}$$

$$[0103] \quad o_t = \sigma(W_{(o,x)}x_t + W_{(o,h)}h_{t-1} + b_o)$$

$$[0104] \quad h_t = o_t \tanh(c_t)$$

[0105] 其中 $\sigma$ 是sigmoid函数,表示向量元素依次相乘, $x_t$ 表示当前时间步骤 $t$ 的输入, $h_t$ 表示当前时间步骤 $t$ 的中间状态, $o_t$ 表示当前时间步骤 $t$ 的输出。连接权重矩阵 $W_{(i,x)}$ 、 $W_{(f,x)}$ 、 $W_{(g,x)}$ 、 $W_{(o,x)}$ 和偏置向量 $b_i$ 、 $b_f$ 、 $b_g$ 、 $b_o$ 是待训练的参数。

[0106] 在使用上述LSTM模型实现编码单元时,第一动作序列的各个动作所对应的第一索引向量作为 $x_t$ 被输入输入门 $i_t$ ,而上一个时间步骤的隐藏状态作为 $h_{t-1}$ 也被输入输入门 $i_t$ 。在使用LSTM模型实现编码单元时,当前时间步骤的输出 $o_t$ 被弃用,仅有当前时间步骤 $t$ 的中间状态 $h_t$ 作为隐藏状态在下一时间步骤中被使用。

[0107] 此外,在使用上述LSTM模型实现解码单元时,第二动作序列的各个动作所对应的第二索引向量作为 $x_t$ 被输入输入门 $i_t$ ,而上一个时间步骤的隐藏状态作为 $h_{t-1}$ 也被输入输入门 $i_t$ 。然而,不同于编码单元,在使用LSTM模型实现解码单元时,当前时间步骤的输出 $o_t$ 作为相应的动作的预测正确的概率被输出。

[0108] 对于上述映射模型,训练目的在于使与第一动作序列 $S$ 对应的第二动作序列 $T$ (其中 $S$ 和 $T$ 构成动作序列对)被预测正确的概率最大,这可以由如下目标函数表示

$$[0109] \quad \frac{1}{|S|} \sum_{\langle T, S \rangle \in C} \log p(T|S)$$

[0110] 该目标函数表示对映射模型的训练样本集合 $C$ 中的每个动作序列对 $\langle S, T \rangle$ 得到的

预测正确的概率相加然后取平均值,优化目标是使得该平均预测正确的概率最大。经过多次迭代,可以得到映射模型的各个参数,其中迭代次数可以根据收敛情况确定,也可以人为设定。例如,在使用LSTM模型实现映射模型的编码单元和解码单元的情况下,经过训练(迭代)可以获得实现编码单元和解码单元的LSTM模型的连接权重矩阵 $W_{(i,x)}$ 、 $W_{(f,x)}$ 、 $W_{(g,x)}$ 、 $W_{(o,x)}$ 和偏置向量 $b_i$ 、 $b_f$ 、 $b_g$ 、 $b_o$ 的数值。

[0111] 基于上述示例扩展至普适的情况,假设给定的第一动作序列 $S = (x_1, \dots, x_T)$ ,与之对应的第二动作序列 $T = (y_1, \dots, y_{T'})$ ,其中 $T$ 是第一动作序列的长度, $T'$ 是第二动作序列的长度, $T$ 与 $T'$ 可以不同,在解码单元处,上式中的 $\log p(T|S)$ 可以表示为:

$$[0112] \quad \log p(T|S) = p(y_1, \dots, y_{T'} | x_1, \dots, x_T) = \prod_{t=1}^{T'} p(y_t | v, y_1, \dots, y_{t-1})$$

[0113] 其中 $p(y_t | v, y_1, \dots, y_{t-1})$ 表示第二动作序列中的动作 $y_t$ 基于其之前的动作 $y_1$ 至 $y_{t-1}$ 以及从编码单元输出的隐含状态 $v$ 而被预测正确的概率。

[0114] 应注意,在映射模型的训练过程中,每个动作序列在结束时需要附加结束标志<EOS>,这使得映射模型能够针对所有可能的动作序列长度被训练。换言之,例如,针对上述示例,针对编码单元的输入是 $[a_{11}, a_{12}, a_{13}, \langle \text{EOS} \rangle]$ ,而解码单元针对 $[a_{21}, a_{22}, a_{23}, a_{24}, \langle \text{EOS} \rangle]$ 计算预测正确的概率。

[0115] 通过上述训练,经训练的映射模型能够将第一智能体的动作序列映射为第二智能体的动作序列。

[0116] 此外,根据本公开的实施方式,对于构成映射模型的编码单元和解码单元,可以使用不同的RNN模型来实现编码单元和解码单元,这可以同时针对多个第一智能体和第二智能体对编码单元和解码单元进行训练。具体而言,经过训练的编码单元和解码单元可以分开和组合使用。

[0117] 此外,根据本公开的实施方式,编码单元可以将第一智能体的动作序列的逆序列编码为固定长度向量,并且解码单元可以将固定长度向量解码为第二智能体的动作序列的逆序列。换言之,可以将第一动作序列中的顺序反转并且依次将相应的第一索引向量输入编码单元,此时解码单元的预测是针对第二动作序列的顺序反转的动作序列进行的。通过这样处理,可以引入第一动作序列与第二动作序列之间的短期依赖关系,从而有助于解决某些优化问题。

[0118] 此外,根据本公开的实施方式,为了进一步提高性能,也可以在映射模型中引入注意力(Attention)机制。

[0119] 使用动作序列对训练映射模型的处理300结束于步骤S305。

[0120] 接下来,回到图2,在步骤S204中,使用第一智能体的第一动作序列训练判断模型,其中判断模型能够判断第一智能体的动作序列的当前动作是否是该动作序列的最后一个动作。

[0121] 图5示出了使用第一动作序列训练判断模型的示例处理的示意图。

[0122] 由于在实际应用中,智能体有可能连续执行多个任务,在上一个任务的动作序列执行结束之后可能下一个任务的动作序列立即开始,在这两个动作序列之间不存在明确的表示上一个动作序列结束的标志。因此,需要判断模型来判断动作序列中的当前动作是否是该动作序列的最后一个动作。应注意,考虑到本公开的技术方案是将经训练的第一智能

体的处理知识转移到未经训练的第二智能体上,因此仅使用第一智能体的第一动作序列来判断模型进行训练。

[0123] 为了训练判断模型,将第一动作序列中的每个动作加上确定该动作是否是第一动作序列的最后一个动作的标签。例如,检查第一动作序列中的每个动作,如果该动作的后续动作是结束标志<EOS>,则该动作是结束动作,为该动作加上标签1,否则加上标签0,从而构造用于对判断模型进行训练的训练样本集合。

[0124] 根据本公开的实施方式,与映射模型的编码单元和解码单元类似,判断模型也可以由RNN模型实现。此外,根据本公开的实施方式,也可以使用作为改进的递归神经网络的长短期记忆(LSTM)模型或门控循环单元(GRU)模型来实现判断模型。

[0125] 鉴于RNN模型、LSTM模型和GRU模型对于本领域技术人员是已知的,因此为了简洁起见,本文仅对其在本公开的实施方式中的应用进行描述,而不对其原理进行更详细的描述。

[0126] 在判断模型的训练过程中,与映射模型的训练过程类似,使用由作为固定长度向量的第一索引向量来表示作为判断模型的训练样本集合的第一动作序列中的每个动作。

[0127] 如图5所示,在判断模型的训练过程中,在每个时间步骤,判断模型的输入是上一时间步骤的判断模型的隐藏状态以及第一动作序列中的当前动作的第一索引向量,判断模型的输出是表示该动作是结束动作的概率的值以及当前时间步骤的隐藏状态。

[0128] 用于判断模型的训练的损失函数被构造为
$$L = \frac{1}{N} (Y - Y')^2$$
。

[0129] 其中Y表示当前动作是否为结束动作的标签,如上文所述,如果当前动作是结束动作,则该标签为1,否则为0。Y'是判断模型预测的结果。N是所有第一动作序列中包含的动作用的数目的总和。通过在每一次迭代过程中,使得该损失函数最小来训练判断模型。

[0130] 经过多次迭代,可以得到判断模型的各个参数,其中迭代次数可以根据收敛情况确定,也可以人为设定。例如,在使用LSTM模型实现判断模型的情况下,经过训练(迭代)可以获得实现判断模型的LSTM模型的连接权重矩阵和偏置向量的数值。

[0131] 通过上述训练过程,经训练的判断模型能够确定第一智能体的动作序列中的结束动作。

[0132] 在映射模型和判断模型的训练完成之后,第二智能体例如6DoF目标机械臂仍然无法自主地完成任务。因此,为了使第二智能体能够自主地根据观测信息执行一系列动作以完成相同的任务,需要构造第二智能体的从观测信息到动作的映射库,即实现第一智能体的对任务的处理知识到第二智能体的转移。

[0133] 因此,在图2的步骤S205中,使用经训练的映射模型和经训练的判断模型构造第二智能体的映射库,其包括从第二智能体的观测信息到第二智能体的动作序列的映射。

[0134] 图6示出了根据本公开的实施方式的使用经训练的映射模型和判断模型构造映射库的示例处理600的流程图。此外,图7示出了根据本公开的实施方式的使用经训练的映射模型和判断模型构造映射库的示例处理的示意图。

[0135] 处理600开始于步骤S601。在步骤S602中,第一智能体根据与第一智能体的观测信息相关的环境信息执行由第一智能体的动作序列构成的动作流。如图7所示,第一智能体例如4DoF源机械臂是经过训练的智能体,因此能够根据观测信息自主地执行一系列动作以完

成预定的任务,所述一系列动作构成了动作流 $a_{11}, a_{12}, a_{13}, a_{14}, a_{15}, \dots$ 。

[0136] 这里所说的第一智能体的处理知识可以被理解为第一智能体的从观测信息到动作的映射库,因而经训练的第一智能体能够根据该映射库针对不同的观测信息执行相应的动作以完成预定任务。本公开的技术方案可被理解为基于经训练的第一智能体的映射库构造未经训练的第二智能体的映射库,从而实现第一智能体的处理知识到第二智能体的转移。然而,由于第一智能体与第二智能体的动作空间不同,因此需要使用上述映射模型和判断模型来实现第一智能体的动作与第二智能体的动作的转换。

[0137] 因此,随后在步骤S603中,使用经训练的判断模型从动作流中提取第一智能体的动作序列。如上文所述,由于第一智能体的动作流中不存在结束标志,因此需要使用经训练的判断模型找到动作流中的结束动作,从而能够将第一智能体的动作流划分成第一智能体的动作序列,以便进行后续处理。如图7所示,判断模型在动作流中判断 $a_{13}$ 为结束动作,因而从上一结束动作到 $a_{13}$ 的动作被提取作为第一智能体的动作序列 $[a_{11}, a_{12}, a_{13}]$ 。

[0138] 随后,在步骤S604中,使用经训练的映射模型根据所提取的第一智能体的动作序列生成第二智能体的动作序列。如图7所示,映射模型可以基于第一智能体的动作序列 $[a_{11}, a_{12}, a_{13}]$ 生成第二智能体的动作序列 $[a_{21}, a_{22}, a_{23}, a_{24}]$ 。

[0139] 随后,在步骤S605中,构造从第二智能体的观测信息到所生成的第二智能体的动作序列的映射。具体地,根据本公开的实施方式,如图7所示,可以在上述步骤S604的执行过程中,记录第二智能体执行动作序列 $[a_{21}, a_{22}, a_{23}, a_{24}]$ 中的每个动作之前的观测信息 $o_1, o_2, o_3, o_4$ ,然后将观测信息与获得的第二智能体的动作成对地记录在第二智能体的映射库中,例如 $o_1 \rightarrow a_{21}, o_2 \rightarrow a_{22}, o_3 \rightarrow a_{23}, o_4 \rightarrow a_{24}$ 。

[0140] 重复上述过程,从而能够基于经训练的第一智能体的映射库构造未经训练的第二智能体的映射库,从而实现第一智能体的处理知识到第二智能体的转移。

[0141] 使用经训练的映射模型和判断模型构造映射库的处理600结束于步骤S606。

[0142] 通过上述处理,第一智能体的处理知识可以被转移到第二智能体,从而第二智能体能够根据观测信息执行相应的动作以完成相同的任务。然而,由于第二智能体的映射库是基于第一智能体的映射库构造的,因此第二智能体仅具备与第一智能体相同的处理知识。换言之,对于第一智能体没有遇到过的观测信息,第二智能体不具备相应的处理知识。因此,为了进一步提高第二智能体的处理性能,根据本公开的实施方式,可以使用所构造的第二智能体的从观测信息到动作的映射库作为训练样本集合对第二智能体进行训练,从而使第二智能体能够应对第一智能体此前从未遇到过的观测信息。

[0143] 根据本公开的信息处理方法,能够将经训练的第一智能体对任务的处理知识转移到具有不同动作空间的未经训练的第二智能体,从而简化第二智能体的训练过程,从而降低训练成本,提高训练效率。

[0144] 此外,本公开还提出了一种信息处理装置,用于将第一智能体的处理知识转移到第二智能体,其中第一智能体能够根据第一智能体的观测信息执行相应的动作序列。

[0145] 图8示出了根据本公开的实施方式的信息处理装置800的结构框图。如图8所示,装置800包括生成单元801,其生成第一智能体的第一动作序列和第二智能体的第二动作序列的动作序列对,其中第一动作序列和第二动作序列完成相同的任务。例如,生成单元801能够执行上文所述的方法200的步骤S202的处理。

[0146] 此外,装置800还包括第一训练单元802,其使用所生成的动作序列对训练映射模型,其中映射模型能够根据第一智能体的动作序列生成第二智能体的动作序列。例如,第一训练单元802能够执行上文所述的方法200的步骤S203的处理。

[0147] 此外,装置800还包括第二训练单元803,其使用第一智能体的第一动作序列训练判断模型,其中判断模型能够判断第一智能体的动作序列的当前动作是否是该动作序列的最后一个动作。例如,第二训练单元803能够执行上文所述的方法200的步骤S204的处理。

[0148] 此外,装置800还包括构造单元804,其使用经训练的映射模型和经训练的判断模型构造映射库,其中映射库包括从第二智能体的观测信息到第二智能体的动作序列的映射。例如,构造单元804能够执行上文所述的方法200的步骤S205的处理。

[0149] 尽管上文以机械臂作为智能体的具体示例描述了本公开的实施方式,但是本公开不限于此。本领域技术人员应认识到,本公开可以应用于除机械臂之外的任何其他的具备执行机构的智能体,例如机器人、无人驾驶汽车、无人驾驶飞行器等。

[0150] 此外,尽管上文为了简化起见仅以机械臂的关节角度为示例描述了本公开的实施方式,但是本公开不限于此。本领域技术人员应认识到,这里所记载的智能体的动作除了涉及机械臂的关节角度之外,还可以涉及连杆的伸缩长度等。在其他智能体的示例中,例如在无人驾驶汽车中,智能体的动作还可能涉及制动踏板和/或油门踏板的下压量和下压行程、方向盘的转动角度等。所有上述内容均应涵盖于本公开的范围之内。

[0151] 此外,尽管上文基于作为4DoF机械臂的第一智能体和作为6DoF机械臂的第二智能体描述了本公开的具体实施方式,但是本领域技术人员在本公开的教导下,能够设想其他的第一智能体和第二智能体的示例,只要第一智能体和第二智能体具有不同的动作空间但是能够完成相同的任务即可。

[0152] 图9是示出可用来实现根据本公开的实施方式的信息处理方法和信息处理装置的通用机器900的结构简图。通用机器900可以是例如计算机系统。应注意,通用机器900只是一个示例,并非暗示对本公开的方法和装置的使用范围或者功能的局限。也不应将通用机器900解释为对上述装置或方法中示出的任一组件或其组合具有依赖或需求。

[0153] 在图9中,中央处理单元(CPU)901根据只读存储器(ROM)902中存储的程序或从存储部分908加载到随机存取存储器(RAM)903的程序执行各种处理。在RAM 903中,还根据需要存储当CPU 901执行各种处理等等时所需的数据。CPU 901、ROM 902和RAM 903经由总线904彼此连接。输入/输出接口905也连接到总线904。

[0154] 下述部件也连接到输入/输出接口905:输入部分906(包括键盘、鼠标等等)、输出部分907(包括显示器,例如阴极射线管(CRT)、液晶显示器(LCD)等,和扬声器等)、存储部分908(包括硬盘等)、通信部分909(包括网络接口卡例如LAN卡、调制解调器等)。通信部分909经由网络例如因特网执行通信处理。根据需要,驱动器910也可连接到输入/输出接口905。可拆卸介质911例如磁盘、光盘、磁光盘、半导体存储器等等可以根据需要被安装在驱动器910上,使得从中读出的计算机程序可以根据需要被安装到存储部分908中。

[0155] 在通过软件实现上述系列处理的情况下,可以从网络例如因特网或从存储介质例如可拆卸介质911安装构成软件的程序。

[0156] 本领域的技术人员应当理解,这种存储介质不局限于图9所示的其中存储有程序、与设备相分离地分发以向用户提供程序的可拆卸介质911。可拆卸介质911的例子包含磁盘

(包含软盘)、光盘(包含光盘只读存储器(CD-ROM)和数字通用盘(DVD))、磁光盘(包含迷你盘(MD)(注册商标))和半导体存储器。或者,存储介质可以是ROM 902、存储部分908中包含的硬盘等等,其中存有程序,并且与包含它们的设备一起被分发给用户。

[0157] 此外,本公开还提出了一种存储有机器可读取的指令代码的程序产品。所述指令代码由机器读取并执行时,可执行上述根据本公开的信息处理方法。相应地,用于承载这种程序产品的上面列举的各种存储介质也包括在本公开的范围之内。

[0158] 上面已通过框图、流程图和/或实施方式进行了详细描述,阐明了根据本公开的实施方式的装置和/或方法的具体实施方式。当这些框图、流程图和/或实施方式包含一个或多个功能和/或操作时,本领域的技术人员明白,这些框图、流程图和/或实施方式中的各功能和/或操作可以通过各种硬件、软件、固件或实质上它们的任意组合而单独地和/或共同地实施。在一种实施方式中,本说明书中描述的主题的几个部分可通过特定用途集成电路(ASIC)、现场可编程门阵列(FPGA)、数字信号处理器(DSP)或其他集成形式实现。然而,本领域的技术人员会认识到,本说明书中描述的实施方式的一些方面能够全部或部分地以在一个或多个计算机上运行的一个或多个计算机程序的形式(例如,以在一个或多个计算机系统上运行的一个或多个计算机程序的形式)、以在一个或多个处理器上运行的一个或多个程序的形式(例如,以在一个或多个微处理器上运行的一个或多个程序的形式)、以固件的形式、或以实质上它们的任意组合的形式等效地实施,并且,根据本说明书中公开的内容,设计用于本公开的电路和/或编写用于本公开的软件和/或固件的代码完全是在本领域技术人员的能力范围之内。

[0159] 应该强调,术语“包括/包含”在本文使用时指特征、要素、步骤或组件的存在,但并不排除一个或多个其他特征、要素、步骤或组件的存在或附加。涉及序数的术语“第一”,“第二”等并不表示这些术语所限定的特征、要素、步骤或组件的实施顺序或者重要性程度,而仅仅是为了描述清楚起见而用于在这些特征、要素、步骤或组件之间进行标识。

[0160] 综上,在根据本公开的实施方式中,本公开提供了如下方案,但不限于此:

[0161] 方案1.一种信息处理方法,用于将第一智能体的处理知识转移到第二智能体,其中所述第一智能体能够根据所述第一智能体的观测信息执行相应的动作序列,所述信息处理方法包括:

[0162] 生成所述第一智能体的第一动作序列和所述第二智能体的第二动作序列的动作序列对,其中所述第一动作序列和所述第二动作序列完成相同的任务;

[0163] 使用所生成的动作序列对训练映射模型,其中所述映射模型能够根据所述第一智能体的动作序列生成所述第二智能体的动作序列;

[0164] 使用所述第一智能体的第一动作序列训练判断模型,其中所述判断模型能够判断所述第一智能体的动作序列的当前动作是否是该动作序列的最后一个动作;以及

[0165] 使用经训练的映射模型和经训练的判断模型构造映射库,其中所述映射库包括从所述第二智能体的观测信息到所述第二智能体的动作序列的映射。

[0166] 方案2.根据方案1所述的信息处理方法,其中所述第一智能体和所述第二智能体是机械臂。

[0167] 方案3.根据方案1或2所述的信息处理方法,其中所述第一智能体的动作的自由度不同于所述第二智能体的动作的自由度。



[0168] 方案4.根据方案1至3中任一项所述的信息处理方法,其中通过使用不同的任务构造不同的所述动作序列对。

[0169] 方案5.根据方案1至4中任一项所述的信息处理方法,其中使用所述动作序列对训练映射模型的步骤进一步包括:

[0170] 设置所述第一智能体的动作的第一索引,使用表示所述第一索引的第一索引向量表示所述第一智能体的第一动作序列;

[0171] 设置所述第二智能体的动作的第二索引,使用表示所述第二索引的第二索引向量表示所述第二智能体的第二动作序列;以及

[0172] 使用所述第一索引向量和所述第二索引向量训练所述映射模型。

[0173] 方案6.根据方案1至4中任一项所述的信息处理方法,其中使用所述第一动作序列训练判断模型的步骤进一步包括:

[0174] 设置所述第一智能体的动作的第一索引,使用表示所述第一索引的第一索引向量表示所述第一智能体的第一动作序列;以及

[0175] 使用所述第一索引向量训练所述判断模型。

[0176] 方案7.根据方案1至4中任一项所述的信息处理方法,其中

[0177] 所述映射模型包括编码单元和解码单元,

[0178] 所述编码单元被配置成将所述第一智能体的动作序列编码为固定长度向量,以及

[0179] 所述解码单元被配置成将所述固定长度向量解码为所述第二智能体的动作序列。

[0180] 方案8.根据方案1至4中任一项所述的信息处理方法,其中

[0181] 所述映射模型包括编码单元和解码单元,

[0182] 所述编码单元被配置成将所述第一智能体的动作序列的逆序列编码为固定长度向量,以及

[0183] 所述解码单元被配置成将所述固定长度向量解码为所述第二智能体的动作序列的逆序列。

[0184] 方案9.根据方案7中所述的信息处理方法,其中所述编码单元和所述解码单元通过递归神经网络模型实现。

[0185] 方案10.根据方案1至4中任一项所述的信息处理方法,其中所述判断模型通过递归神经网络模型实现。

[0186] 方案11.根据方案10或11所述的信息处理方法,其中所述递归神经网络模型是长短期记忆模型或门控循环单元模型。

[0187] 方案12.根据方案1至4中任一项所述的信息处理方法,其中使用经训练的映射模型和经训练的判断模型构造映射库的步骤进一步包括:

[0188] 所述第一智能体根据与所述第一智能体的观测信息相关的环境信息执行由所述第一智能体的动作序列构成的动作流;

[0189] 使用经训练的判断模型从所述动作流中提取所述第一智能体的动作序列;

[0190] 使用经训练的映射模型根据所提取的所述第一智能体的动作序列生成所述第二智能体的动作序列;以及

[0191] 构造从所述第二智能体的观测信息到所生成的所述第二智能体的动作序列的映射。

[0192] 方案13.根据方案1至4中任一项所述的信息处理方法,进一步包括:

[0193] 使用所述映射库对第二智能体进行训练。

[0194] 方案14.一种信息处理装置,用于将第一智能体的处理知识转移到第二智能体,其中所述第一智能体能够根据所述第一智能体的观测信息执行相应的动作序列,所述信息处理装置包括:

[0195] 生成单元,被配置成生成所述第一智能体的第一动作序列和所述第二智能体的第二动作序列的动作序列对,其中所述第一动作序列和所述第二动作序列完成相同的任务;

[0196] 第一训练单元,被配置成使用所生成的动作序列对训练映射模型,其中所述映射模型能够根据所述第一智能体的动作序列生成所述第二智能体的动作序列;

[0197] 第二训练单元,被配置成使用所述第一智能体的第一动作序列训练判断模型,其中所述判断模型能够判断所述第一智能体的动作序列的当前动作是否是该动作序列的最后一个动作;以及

[0198] 构造单元,被配置成使用经训练的映射模型和经训练的判断模型构造映射库,其中所述映射库包括从所述第二智能体的观测信息到所述第二智能体的动作序列的映射。

[0199] 方案15.根据方案14所述的信息处理装置,其中所述第一智能体和所述第二智能体是机械臂。

[0200] 方案16.根据方案14或15所述的信息处理装置,其中所述第一智能体的动作的自由度不同于所述第二智能体的动作的自由度。

[0201] 方案17.根据方案14至16中任一项所述的信息处理装置,其中通过使用不同的任务构造不同的所述动作序列对。

[0202] 方案18.根据方案14至17中任一项所述的信息处理装置,其中所述第一训练单元被进一步配置成:

[0203] 设置所述第一智能体的动作的第一索引,使用表示所述第一索引的第一索引向量表示所述第一智能体的第一动作序列;

[0204] 设置所述第二智能体的动作的第二索引,使用表示所述第二索引的第二索引向量表示所述第二智能体的第二动作序列;以及

[0205] 使用所述第一索引向量和所述第二索引向量训练所述映射模型。

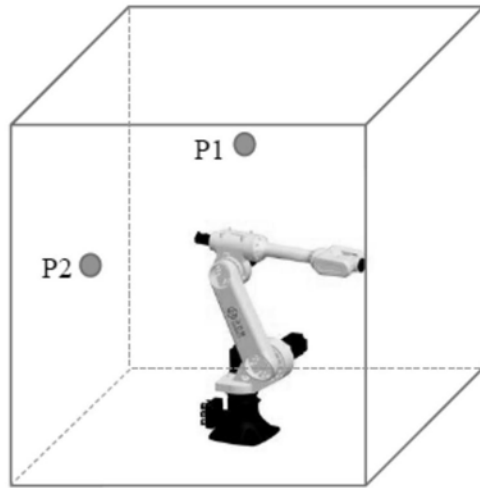
[0206] 方案19.根据方案14至17中任一项所述的信息处理装置,其中所述第二训练单元被进一步配置成:

[0207] 设置所述第一智能体的动作的第一索引,使用表示所述第一索引的第一索引向量表示所述第一智能体的第一动作序列;以及

[0208] 使用所述第一索引向量训练所述判断模型。

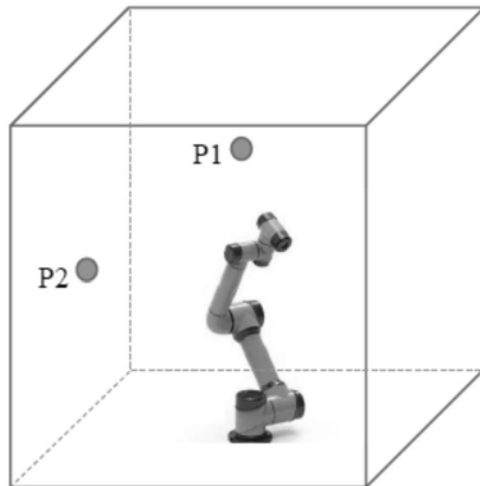
[0209] 方案20.一种计算可读存储介质,其上存储有计算机程序,所述计算机程序在被计算机执行时实现如方案1至13中任一项所述的信息处理方法。

[0210] 尽管上面已经通过对本公开的具体实施方式的描述对本公开进行了披露,但是,应该理解,本领域的技术人员可在所附权利要求的精神和范围内设计对本公开的各种修改、改进或者等同物。这些修改、改进或者等同物也应当被认为包括在本公开的保护范围内。



4 DoF 机械臂

图1A



6 DoF 机械臂

图1B

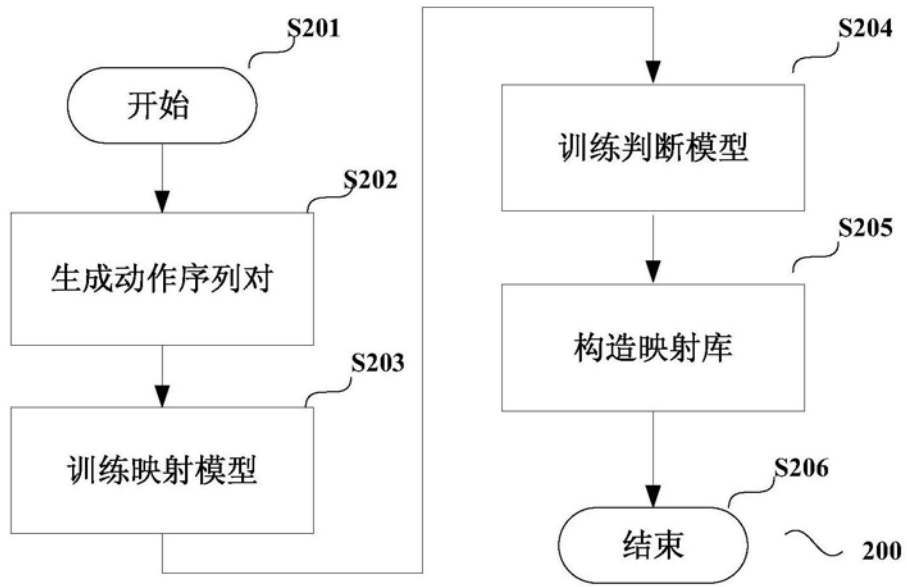


图2

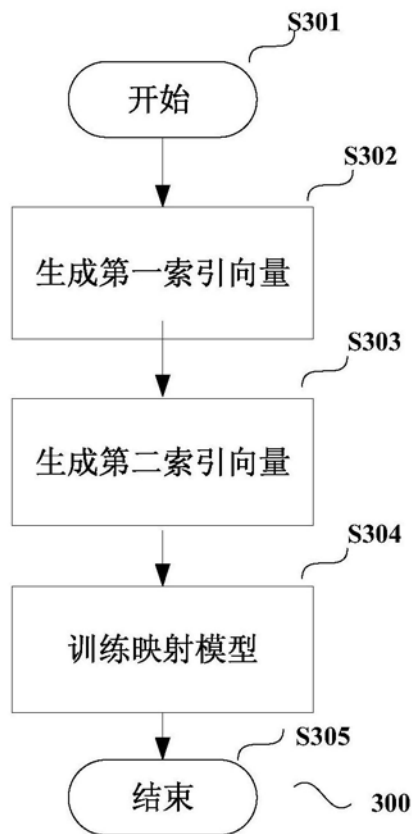


图3

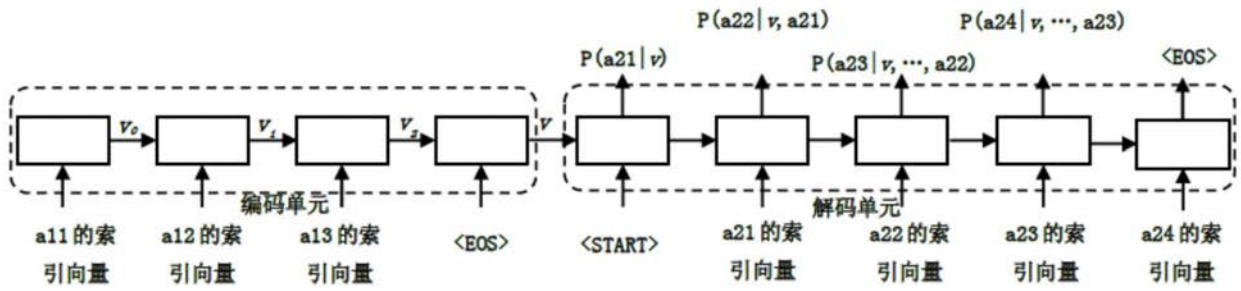


图4

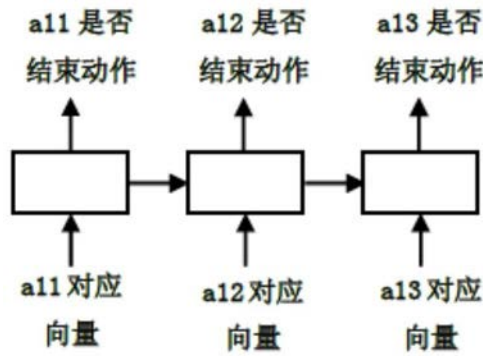


图5

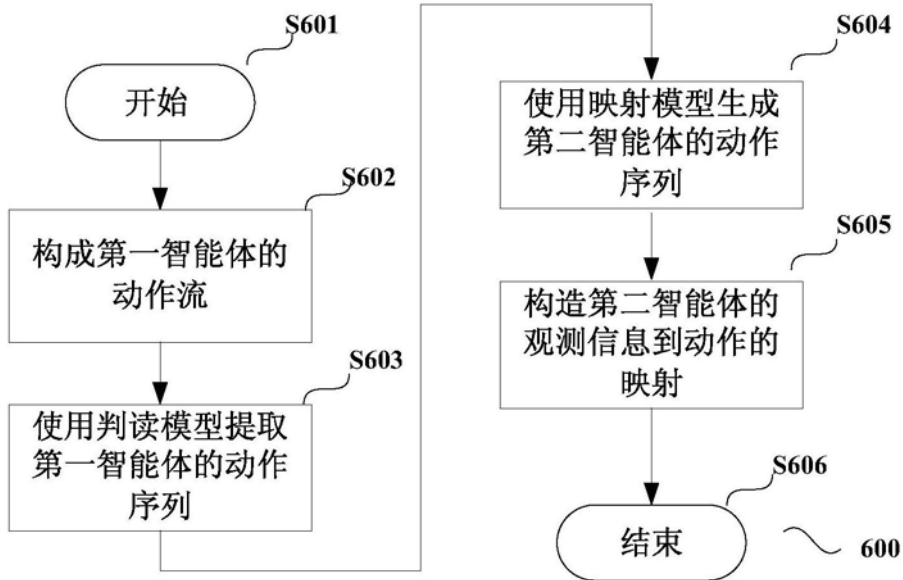


图6

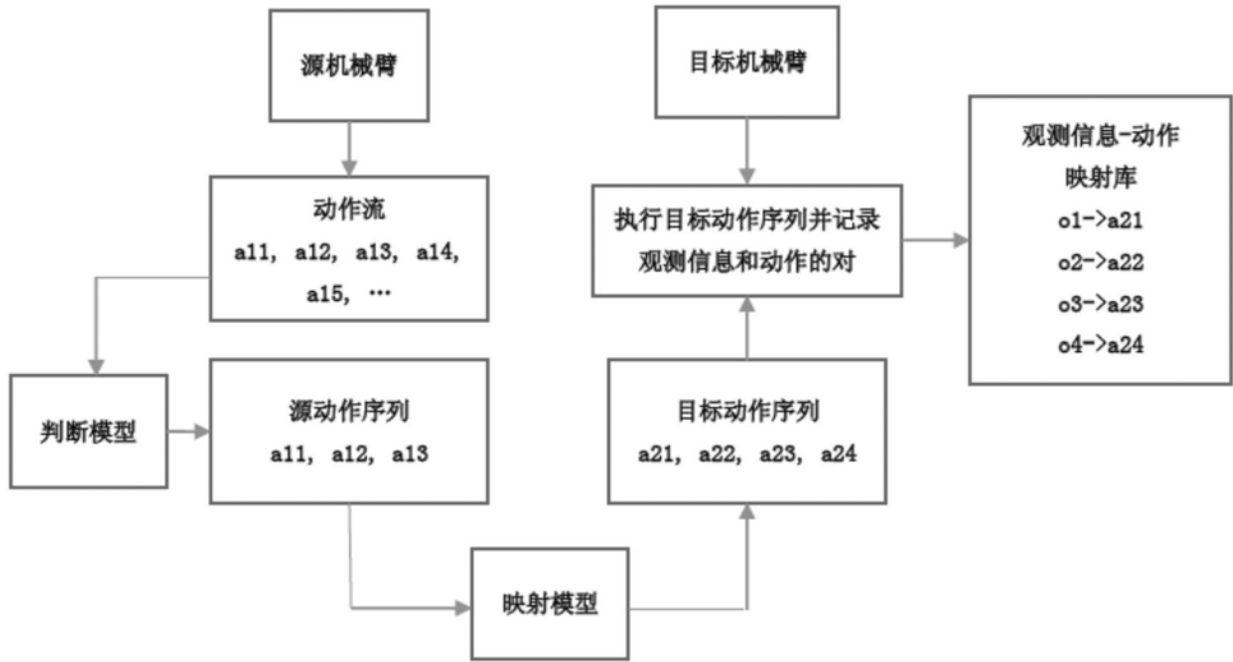


图7

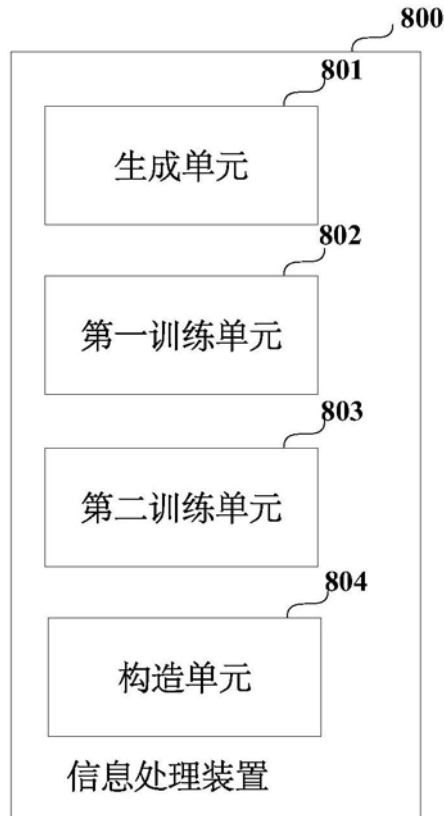


图8

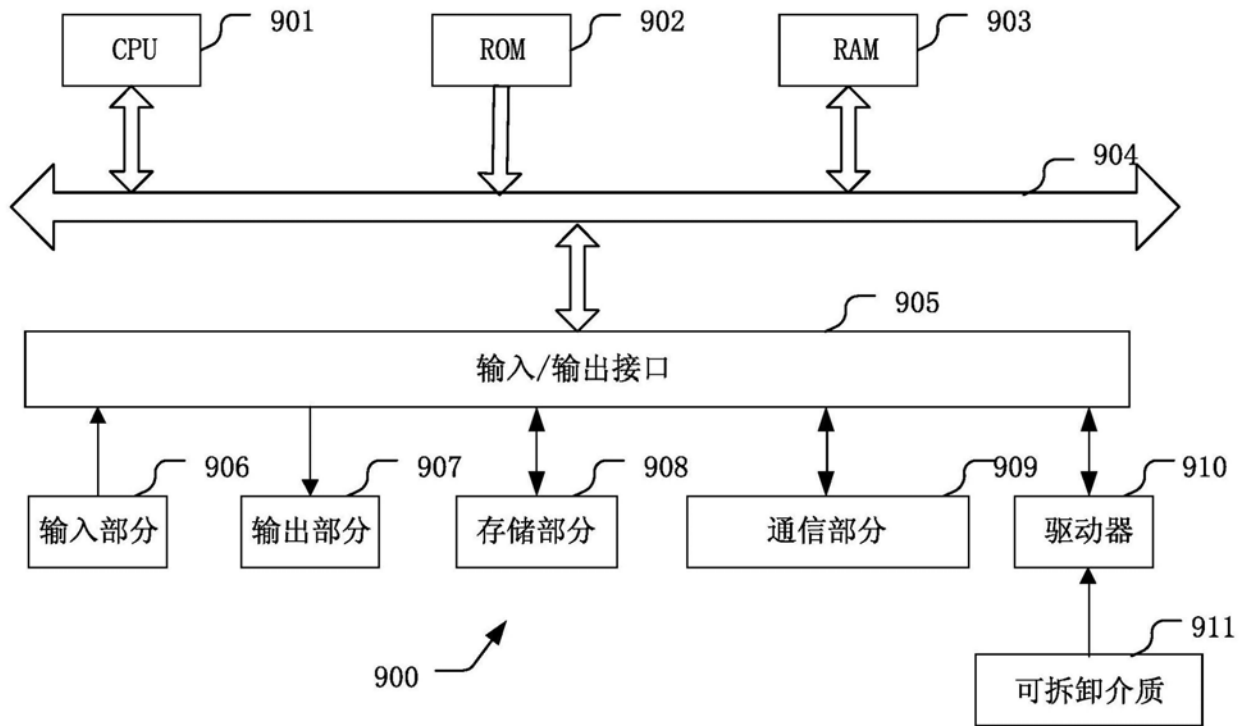


图9