

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2019 年 6 月 27 日 (27.06.2019)



(10) 国际公布号
WO 2019/120174 A1

(51) 国际专利分类号:
G05D 1/00 (2006.01)

中国广东省深圳市龙岗区坂田华为总部
办公楼, Guangdong 518129 (CN)。

(21) 国际申请号: PCT/CN2018/121519

(22) 国际申请日: 2018 年 12 月 17 日 (17.12.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201711408965.4 2017 年 12 月 22 日 (22.12.2017) CN

(71) 申请人: 华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.) [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(72) 发明人: 钱俊 (QIAN, Jun); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。王新宇 (WANG, Xinyu); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。陈晨 (CHEN, Chen);

(74) 代理人: 北京三高永信知识产权代理有限公司 (BEIJING SAN GAO YONG XIN INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国北京市海淀区学院路蓟门里和景园 A 座 1 单元 102 室, Beijing 100088 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(54) Title: ACTION CONTROL METHOD AND APPARATUS

(54) 发明名称: 动作控制方法及装置

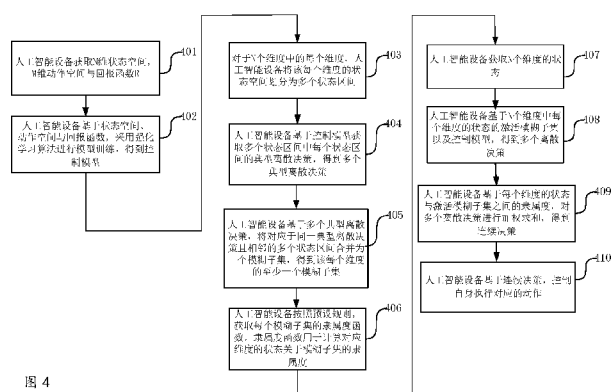


图 4

401 An artificial intelligence device obtains an N-dimensional state space, an M-dimensional action space, and a reward function R
402 On the basis of the state space, the action space, and the reward function, the artificial intelligence device performs model training by using a reinforcement learning algorithm to obtain a control model
403 The artificial intelligence device divides the state space of each of N dimensions into a plurality of state intervals
404 The artificial intelligence device obtains, on the basis of the control model, the typical discrete decision of each of the plurality of the state intervals to obtain a plurality of typical discrete decisions
405 The artificial intelligence device combines, on the basis of the plurality of typical discrete decisions, the plurality of the state intervals that are adjacent and correspond to the same typical discrete decision into a fuzzy subset to obtain at least one fuzzy subset of each dimension
406 The artificial intelligence device obtains, according to a preset rule, the membership degree function of each fuzzy subset, the membership degree function being used for calculating the membership degree of the state of a particular dimension for each fuzzy subset
407 The artificial intelligence device obtains the states of the N dimensions
408 The artificial intelligence device obtains a plurality of discrete decisions on the basis of the activated fuzzy subset of the state of each of the N dimensions and the control model
409 The artificial intelligence device performs, on the basis of the membership degree between the state of each dimension and the activated fuzzy subset, weighing and summing on the plurality of discrete decisions to obtain continuous decisions
410 The artificial intelligence device controls, on the basis of the continuous decisions, itself to execute corresponding actions

(57) Abstract: An action control method and apparatus, relating to the field of artificial intelligence. The method comprises: obtaining the states of N dimensions of an artificial intelligence device (407); obtaining a plurality of discrete decisions on the basis of the activated fuzzy subset of the state of each of the N dimensions and a control model (408), the activated fuzzy subset of one state referring to a fuzzy subset of which the state membership degree is not 0, each fuzzy subset comprising a state interval corresponding to the same discrete decision in one dimension, the membership degree being used for indicating the degree to which the state belongs to a fuzzy subset, and the control model being used for outputting a corresponding discrete decision according to the state of an input; on the basis of the membership degree between each dimension and the activated fuzzy subset, weighing and summing the plurality of discrete decisions to obtain continuous decisions (409); and controlling, on the basis of the continuous decisions, the artificial intelligence device to perform corresponding actions (410). The continuous output decisions can ensure the smooth control of the artificial intelligence device and ensure the smoothness of the actions.

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

(57) 摘要: 一种动作控制方法及装置, 属于人工智能领域。方法包括: 获取人工智能设备的N个维度的状态 (407); 基于N个维度中每个维度的状态的激活模糊子集以及控制模型, 得到多个离散决策 (408), 一个状态的激活模糊子集是指状态的隶属度不为0的模糊子集, 每个模糊子集包括一个维度内对应于同一个离散决策的状态区间, 隶属度用于表示状态隶属于模糊子集的程度高低, 控制模型用于根据输入的状态输出对应的离散决策; 基于每个维度的状态与激活模糊子集之间的隶属度, 对多个离散决策进行加权求和, 得到连续决策 (409); 基于连续决策, 控制人工智能设备执行对应的动作 (410)。输出的决策为连续量, 能够保证对人工智能设备的平顺控制, 保证动作的平滑性。

动作控制方法及装置

本申请要求于 2017 年 12 月 22 日提交的申请号为 201711408965.4、发明名称为“动作控制方法及装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本公开涉及人工智能领域，特别涉及一种动作控制方法及装置。

背景技术

随着人工智能领域的发展，无人车、机器人等各种人工智能设备纷纷诞生，为人们的生活带来了极大便利。目前通常由人工智能方法来为人工智能设备提供智能决策，充当人工智能设备的机器大脑，从而控制人工智能设备执行对应的动作，例如可以控制无人车在马路上行驶，可以控制机器人在仓库中移动等。

目前，通常采用强化学习算法控制人工智能设备的动作：以无人车为例，首先根据实际应用场景定义强化学习算法中的必要元素： N 维状态空间 S ， m 维动作空间 A 与回报函数 R ，再基于 S 、 A 、 R 在模拟器环境中进行模拟驾驶或在真实道路上进行驾驶，训练出一个控制模型，该控制模型用于根据输入的 S 内的 N 维状态，输出 A 内的 m 维离散决策。之后，无人车在实际驾驶过程中，会采集当前的 N 维状态，将 N 维状态输入到控制模型中，得到控制模型输出的 m 维离散决策，基于该 m 维离散决策，控制执行对应的动作。

在实现本公开的过程中，发明人发现相关技术至少存在以下问题：

基于离散决策控制人工智能设备执行动作时，由于其所输出的决策为离散量，难以保证对人工智能设备的平顺控制，导致动作的平滑性较差。

发明内容

本公开实施例提供了一种动作控制方法及装置，能够解决相关技术中难以对人工智能设备进行平顺控制的技术问题。所述技术方案如下：

第一方面，提供了一种动作控制方法，所述方法包括：

获取人工智能设备的 N 个维度的状态，所述 N 为大于或等于 1 的正整数；

基于所述 N 个维度中每个维度的状态的激活模糊子集以及控制模型，得到多个离散决策，一个状态的激活模糊子集是指所述状态的隶属度不为 0 的模糊子集，每个模糊子集包括一个维度内对应于同一个离散决策的状态区间，所述隶属度用于表示状态隶属于模糊子集的程度高低，所述控制模型用于根据输入的状态输出对应的离散决策；

基于所述每个维度的状态与激活模糊子集之间的隶属度，对所述多个离散决策进行加权求和，得到连续决策；

基于所述连续决策，控制所述人工智能设备执行对应的动作。

本实施例提供的方法，基于每个维度的状态与激活模糊子集之间的隶属度，对多个离散决策进行加权求和，得到连续决策，由于输出的决策为连续量，能够保证对人工智能设备的

平顺控制，保证动作的平滑性。同时，通过隶属度来获取连续决策，提供了一种合理地连续化离散决策的方式，保证连续决策的变化趋势与状态的变化趋势匹配，从而保证连续决策具有高准确性。进一步地，当人工智能设备为无人车时，可以保证控制无人车的平顺性，提升乘客的舒适度。

在一种可能的设计中，所述基于所述每个维度的状态与激活模糊子集之间的隶属度，对所述多个离散决策进行加权求和，得到连续决策，包括：

对于所述多个离散决策中的每个离散决策，获取所述每个离散决策对应的 N 个激活模糊子集的隶属度，得到 N 个隶属度；

基于所述 N 个隶属度，计算所述每个离散决策的权重；

基于所述每个离散决策的权重，对所述多个离散决策进行加权求和，得到所述连续决策。

在一种可能的设计中，所述基于所述 N 个维度中每个维度的状态的激活模糊子集以及控制模型，得到多个离散决策，包括：

获取所述 N 个维度中每个维度的激活模糊子集的中心值，得到多个中心值；

对不同维度的中心值进行组合，得到多个中间状态，每个中间状态包括 N 个维度的中心值；

分别将所述多个中间状态输入到所述控制模型中，得到所述控制模型输出的多个离散决策。

在一种可能的设计中，所述基于所述 N 个维度中每个维度的状态的激活模糊子集以及控制模型，得到多个离散决策之前，所述方法还包括：

对于所述 N 个维度中的每个维度，当所述每个维度的状态与所述每个维度的任一模糊子集之间的隶属度不为 0 时，将所述每个模糊子集作为所述每个维度的激活模糊子集；或，

对于所述 N 个维度中的每个维度，从所述每个维度的多个模糊子集中选取中心值位于所述每个维度的状态左右的两个模糊子集，作为所述每个维度的激活模糊子集。

在一种可能的设计中，所述获取人工智能设备的 N 个维度的状态之前，所述方法还包括：

对于所述 N 个维度中的每个维度，将所述每个维度的状态空间划分为多个状态区间；

基于所述控制模型，获取所述多个状态区间中每个状态区间的典型离散决策，得到多个典型离散决策；

基于所述多个典型离散决策，将对应于同一典型离散决策且相邻的多个状态区间合并为一个模糊子集，得到所述每个维度的至少一个模糊子集。

本设计提供了一种基于强化学习训练得到的控制模型自动划分模糊子集的方式，无需依赖人工确定决策，效率极高。进一步地，可以采用过分割的方式，将每个状态空间分割为大量状态区间，基于大量状态区间的典型离散决策进行合并得到模糊子集时，由于模糊子集的边界非常精确，能够保证模糊子集的准确性较高。进一步地，适用于为高维状态空间划分模糊子集的场景，可以方便快速的扩展到高维空间，以便应用于实际应用中复杂的运行情况，实用性强。

在一种可能的设计中，所述基于所述控制模型，获取所述多个状态区间中每个状态区间的典型离散决策，包括：

对于所述多个状态区间中的每个状态区间，获取所述每个状态区间的多个代表状态，每个代表状态包括所述每个维度上的所述每个状态区间的中心值以及其他每个维度上的任一个

状态；

分别将所述多个代表状态输入到所述控制模型中，得到所述控制模型输出的多个离散决策；

从所述多个离散决策中选取出现次数最多的离散决策，作为所述每个状态区间的典型离散决策。

在一种可能的设计中，所述获取人工智能设备的 N 个维度的状态之后，所述方法还包括：

对于所述 N 个维度中每个维度的每个模糊子集，采用所述每个模糊子集对应的隶属度函数，对所述每个维度的状态进行计算，得到所述每个模糊子集的隶属度。

在一种可能的设计中，所述获取所述人工智能设备的 N 个维度的状态之前，所述方法还包括：

按照预设规则，获取每个模糊子集的隶属度函数，所述隶属度函数用于计算对应模糊子集的隶属度，所述预设规则为：隶属度函数在所述每个模糊子集的中心值取 1，在所述每个模糊子集的边界值的隶属度取 0.5，在所述每个模糊子集的相邻的两个模糊子集的中心值取 0。

基于本设计，隶属度函数具有高解释性，较为有效。同时，步骤简单，能够提高构建隶属度函数的效率。

第二方面，提供了一种动作控制装置，所述装置包括多个功能模块，以实现上述第一方面以及第一方面的任一种可能设计中的动作控制方法。

第三方面，提供了一种人工智能设备，所述人工智能设备包括处理器和存储器，所述存储器中存储有至少一条指令，所述指令由所述处理器加载并执行以实现上述第一方面以及第一方面的任一种可能设计中的动作控制方法。

第四方面，提供了一种计算机可读存储介质，所述存储介质中存储有至少一条指令，所述指令由处理器加载并执行以实现上述第一方面以及第一方面的任一种可能设计中的动作控制方法。

附图说明

图 1 是本公开实施例提供的一种实施环境的示意图；

图 2 是本公开实施例提供的一种应用场景的示意图；

图 3 是本公开实施例提供的一种人工智能设备的结构示意图；

图 4 是本公开实施例提供的一种动作控制方法的流程图；

图 5 是本公开实施例提供的一种隶属度函数的示意图；

图 6 是本公开实施例提供的一种隶属度函数的示意图；

图 7 是本公开实施例提供的一种隶属度函数的示意图；

图 8 是本公开实施例提供的一种动作控制装置的结构示意图。

具体实施方式

为使本公开的目的、技术方案和优点更加清楚，下面将结合附图对本公开实施方式作进一步地详细描述。

为了方便理解，下面先对本公开实施例中涉及的名词进行解释：

状态空间：是指人工智能设备所有可能的状态的集合，状态空间可以包括 N 个维度（ n 为正整数），状态空间的维度可以为速度维度、角度维度、距离维度等，人工智能设备在任

一时刻的状态可以采用状态空间中的一个 N 维向量表示。

动作空间：是指人工智能设备所有可以执行的动作的集合、动作空间可以包括 m 个维度 (m 为正整数)，例如包括油门维度、转向角维度、刹车维度等，人工智能设备在任一时刻的动作可以采用动作空间中的一个 m 维向量表示。

回报函数：以状态为输入、以奖励值为输出的函数，奖励值越大，表示对应的状态越理想，奖励值为负时表示对应的状态不理想。

强化学习(reinforcement learning)：又称再励学习、评价学习，是指以奖励值最大化为目标，从环境映射到行为的学习。强化学习中对产生动作的好坏均采用奖励值进行评价，在训练过程中控制模型必须靠自身的经历进行学习，在行动-评价的环境中获得知识，不断改进行动以适应环境。

模糊子集(fuzzy set)：也称模糊集合、模糊集，是指表达模糊性概念的集合，任一状态与模糊子集之间的隶属关系不是绝对的肯定或否定，而通过隶属度的大小表征隶属关系的高低程度。本公开实施例中，模糊子集是指一个维度内对应于同一个离散决策的状态区间，即，模糊子集内的所有状态对应的离散决策相同。

隶属度以及隶属度函数：若对论域 U 中的任一元素 x ，都有一个数 $A(x) \in [0, 1]$ 与之对应，则称 A 为 U 上的模糊集， $A(x)$ 称为 x 对 A 的隶属度。当 x 在 U 中变动时， $A(x)$ 就是一个函数，称为 A 的隶属函数。隶属度 $A(x)$ 越接近于 1，表示 x 属于 A 的程度越高， $A(x)$ 越接近于 0 表示 x 属于 A 的程度越低，通过取值于区间 $(0, 1)$ 的隶属函数 $A(x)$ 可以表征 x 属于 A 的程度高低。本公开实施例中，隶属度用于表示状态隶属于模糊子集的程度高低，隶属度越大，表示状态隶属于模糊子集的程度越高，可以认为状态强属于模糊子集，隶属度越小，表示状态隶属于模糊子集的程度越低，可以认为状态弱属于模糊子集。

激活模糊子集：当状态与某个模糊子集之间的隶属度不为 0 时，可以理解为该模糊子集被激活，则记该模糊子集为激活模糊子集。

图 1 是本公开实施例提供的一种实施环境的示意图，该实施环境包括人工智能设备 101 和多个服务器 102，人工智能设备 101 通过无线或者有线网络与多个服务器 102 连接，该人工智能设备 101 可以为无人车或机器人，每个服务器 102 可以为是一台服务器，或者由若干台服务器组成的服务器集群，或者是一个云计算服务中心。

在通过强化学习训练控制模型的过程中，人工智能设备训练所用的状态可以来自服务器 102，人工智能设备可以感知各个维度的状态，从而进行学习。以无人车为例，服务器可以将无人车当前的位置信息、无人车当前行驶的道路信息提供给无人车，以便无人车获知所处的位置、所行驶的道路信息。在通过控制模型实际使用的过程中，人工智能设备获取的状态也可以来自服务器，人工智能设备可以基于获取到的状态以及控制模型确定对应的决策，以便执行对应的动作。可选地，服务器 102 还可以具有至少一种数据库，例如道路交通路网数据库、地理信息数据库等，用于向人工智能设备 101 提供各个维度的状态。

本实施例提供的动作控制方法可以应用在多种实际应用场景，以下结合两种示例性应用场景进行阐述：

(1) 可以应用在无人车进行无人驾驶的场景：参见图 2，无人车在驾驶过程中可以通过当前的速度、与道路中轴线之间的距离、当前道路交通等各个维度的状态以及控制模型，输

出连续决策，该连续决策可以为转向角大小、加速度、油门大小、刹车大小等，通过连续决策能够保证无人车的动作的平滑性，避免出现速度忽高忽低等情况，提升乘客的舒适度。进一步地，输出的连续决策可以为底层数据，即连续决策可以细化到左拐的角度、加速度的数值等数据，从而进行高精度控制，实现无人车可以执行沿着道路轴线直行、换道、超车、跟车、泊车等高难度动作的效果，实现无人车自动驾驶的任务。

(2) 可以应用在机器人执行任务的场景：本发明实施例可以应用在仓储机器人在仓库、工地等各种地点搬运货物、除草机器人在草坪进行除草、扫地机器人在办公室进行扫地、收割机器人在田地收割庄稼、围棋机器人下围棋等场景。以仓储机器人在仓库搬运某货物为例，仓储机器人在运行过程中可以获取当前的速度、当前的位置、与相邻货架的距离、与放置该货物的货架之间的距离等状态以及控制模型，输出连续决策，该连续决策可以为捡起商品、抬手、将货物放在储物筐中、转向角大小、加速等，通过连续决策能够保证机器人的动作的平滑性，避免出现机器人动作之间的波动性大导致机器人摔倒的情况。进一步地，输出的连续决策可以为底层数据，即连续决策可以细化到捡起商品的速度、左拐的角度、加速度的数值等数据，从而进行高精度控制，实现机器人可以执行高难度动作的效果，提高机器人的性能。

图 3 是本公开实施例提供的一种人工智能设备的结构示意图，参见图 3，该人工智能设备包括：接收器 301、发射器 302、存储器 303 和处理器 304，该接收器 301、该发射器 302 和该存储器 303 分别与该处理器 304 连接，该存储器 303 存储有程序代码，该处理器 304 用于调用该程序代码，执行下述实施例中人工智能设备所执行的操作。

在示例性实施例中，还提供了一种计算机可读存储介质，例如包括指令的存储器，上述指令可由电子设备中的处理器执行以完成下述实施例中的动作控制方法。例如，所述计算机可读存储介质可以是只读存储器 (Read-Only Memory, ROM)、随机存取存储器 (Random Access Memory, RAM)、CD-ROM、磁带、软盘和光数据存储设备等。

本公开实施例提供了一种动作控制方法，该方法主要包括三部分，第一部分为强化学习的过程，详见以下步骤 401-步骤 402，通过强化学习的过程可以得到一个输入为状态，输出为离散决策的控制模型；第二部分为空间模糊化的过程，详见以下步骤 403-步骤 406，通过空间模糊化的过程可以将状态空间的每一维度划分为多个模糊子集，并得出各模糊子集的隶属度函数。第三部分是反模糊化的过程，详见以下步骤 407-步骤 410，通过在线控制过程中基于实际状态所属的模糊子集和隶属度函数，计算连续决策进而控制自身执行对应的动作。通过本公开实施例提供的方法，由于基于连续决策控制人工智能设备执行动作，可以保证对人工智能设备的平顺控制，保证动作的平滑性。

图 4 是本公开实施例提供的一种动作控制方法的示意图，该方法的执行主体为人工智能设备，如图 4 所示，该方法包括：

401、人工智能设备获取 N 维状态空间 S ， M 维动作空间 A 与回报函数 R ， N 为大于或等于 1 的正整数， M 也为大于或等于 1 的正整数。

由于后续过程会以状态空间、动作空间、回报函数为输入，各种离散决策为输出，基于强化学习算法构建训练模型，本步骤 401 中，人工智能设备会预先获取状态空间、动作空间、

回报函数，以便定义强化学习算法的必要元素。以状态空间用 S 表示、动作空间用 A 表示，回报函数用 R 表示为例，本步骤 401 具体可以包括以下步骤一至步骤三。

步骤一、人工智能设备生成 N 维状态空间 S 。

每个维度的状态空间的概念以及设计：每个维度的状态空间是指人工智能设备对应维度的状态的集合，每个维度的状态空间可以根据人工智能设备在运行过程中可以获取到的状态设计，即可以根据人工智能设备可以实时采集的状态或者通过处理得到的状态设计。每个维度的状态空间包括两个边界，左边界可以表征对应维度的状态的最小值，右边界可以表征对应维度的状态的最大值。其中，可以直接将状态的最小值设计为状态空间的左边界，直接将状态的最大值设计为状态空间的右边界。或者，可以对状态的最小值和最大值进行归一化，将归一化的最小值和归一化的最大值分别设计为状态空间的左边界和右边界。

示例性地，以人工智能设备为无人车为例，状态空间 S 可以包括 θ 维度的状态空间、 V 维度的状态空间、 P 维度的状态空间、 V 前维度的状态空间、 V 后维度的状态空间、 P 前维度的状态空间、 P 后维度的状态空间的任意组合，其中 θ 是指无人车的中轴线与道路的中轴线之间的夹角、 V 是指无人车当前的速度、 P 是指无人车与道路的中轴线之间的距离、 V 前是指无人车的前一辆车的速度， V 后是指无人车的后一辆车的速度、 P 前是指无人车与前一辆车之间的距离、 P 后是指无人车与后一辆车之间的距离。以人工智能设备为机器人为例，状态空间可以包括速度维度的状态空间、左脚维度的状态空间、右脚维度的状态空间等，本实施例对不做限定。

在一种可能的设计中，参见表 1，无人车可以对应 $N=3$ 维状态空间，包括 θ 维度的状态空间、 V 维度的状态空间和 P 维度的状态空间，其中 θ 维度的状态空间的边界可以为 -1 和 1 ， -1 表示 -180° ， 1 表示 180° 。 V 维度的状态空间的边界为 0 至 300 ， 0 表示最小速度， 300 表示最大速度， P 维度的状态空间的边界为 -1 和 1 ，该边界采用道路宽度进行归一化的结果。当 $P=0$ 时表示无人车正好在道路中轴线上，当 P 大于 1 或小于 -1 表示无人车已经驶出道路。

表 1

变量名	解释
θ	自车的中轴与道路中轴的夹角，范围为 -1 至 1 。
V	自车的速度，范围为 0 至 400
P	自车和道路中轴线之间的距离，范围为 -1 至 1

针对生成 N 维状态空间 S 的具体过程，人工智能设备可以获取状态空间生成指令，基于该状态空间生成指令生成 N 维状态空间。该状态空间生成指令指示状态空间的维度数量，还可以指示每个维度的状态空间的名称、每个维度的状态空间的边界值，人工智能设备可以基于状态空间生成指令中的维度数量、每个维度的状态空间的名称、每个维度的状态空间的边界值生成 N 维状态空间。其中，人工智能设备可以在运行定义状态空间的代码时，获取状态空间生成指令。该定义状态空间的代码由开发人员预先根据实际需求进行编写，并由开发人员预先在人工智能设备中进行存储。

步骤二、人工智能设备生成 M 维动作空间 A 。

每个维度的动作空间的概念以及设计：每个维度的动作空间是指人工智能设备对应维度的动作的集合，每个维度的动作空间可以根据人工智能设备在实际运行过程中能够执行的动

作确定。每个维度的动作空间包括两个边界，左边界可以表征对应维度的动作的最小值，右边界可以表征对应维度的动作的最大值。其中，可以直接将动作的最小值设计为动作空间的左边界，直接将动作的最大值设计为动作空间的右边界。或者，可以对动作的最小值和最大值进行归一化，将归一化的最小值和归一化的最大值分别设计为动作空间的左边界和右边界。

示例性地，以人工智能设备为无人车为例，动作空间 A 可以包括转向角维度的动作空间、油门维度的状态空间、刹车维度的状态空间的任意组合等。以人工智能设备为机器人为例，动作空间可以包括扫地维度的动作空间、除草维度的动作空间、转向角维度的动作空间、搬运物品维度的动作空间的任意组合。

其中，每个维度对应一种类型的动作，每个维度的动作空间为人工智能设备对应类型的动作的集合，每个维度的动作空间的边界由对应类型的动作的最小值和最大值确定。例如速度维度的动作空间为人工智能设备行驶的速度的集合，速度维度的动作空间的边界由人工智能设备的最小速度和最大速度确定，例如为 0 至 400。

针对生成 M 维动作空间 A 的具体过程，人工智能设备可以获取动作空间生成指令，基于该动作空间生成指令生成 M 维动作空间 A 。该动作空间生成指令指示动作空间的维度数量，还可以指示每个维度的动作空间的名称、每个维度的动作空间的边界值，人工智能设备可以基于动作空间生成指令中的维度数量、每个维度的动作空间的名称、每个维度的动作空间的边界值生成 M 维动作空间。其中，人工智能设备可以在运行定义动作空间的代码时，获取动作空间生成指令，该定义动作空间的代码由开发人员预先根据实际需求进行编写，并由开发人员预先在人工智能设备中进行存储。

在一种可能的设计中，人工智能设备可以对 M 维动作空间 A 进行离散化，即，对于每个维度的动作空间，可以按照一定的粒度从该维度的动作空间提取 q 个动作 $\{a_i, i=1, \dots, p\}$ ，作为该维度的离散化的动作空间。其中，可以按照统一的粒度对每个维度的动作空间进行离散化，也可以获取粒度和动作空间的维度之间的对应关系，按照维度对应的粒度对每个维度的动作空间进行离散化。另外，粒度的具体数值可以根据实际需求确定，本实施例对此不做限定。

示例性地，参见表 2，无人车可以对应 $M=1$ 维动作空间（转向角维度的动作空间），该动作空间的范围为 $[-1, 1]$ ，其中 -1 表示最大右转角， 1 表示最大左转角， 0 表示直行。可以按照 0.1 的粒度对该动作空间进行离散化，得到 $[-1, -0.9, -0.8 \dots 0.8, 0.9, 1]$ ，共计 21 个动作，组成转向角维度的离散化的动作空间。

表2

变量名	解释
转向角	范围是 $[-1, 1]$ ，按照 0.1 的粒度进行离散化后为 $[-1, -0.9, -0.8 \dots 0.8, 0.9, 1]$ ，一共 21 维输出。

本实施例中，通过对动作空间进行离散化，可以达到以下技术效果：

提高控制模型的收敛速度，保证快速训练出控制模型：动作空间的设计会直接影响后续模型训练的过程中控制模型的收敛速度，动作空间的动作越多，控制模型确定某个状态下的决策时可供选择的动作越多，而面临的选择越多则计算量越大，则控制模型的收敛速度越慢，在线使用时确定决策的速度也越慢。相关技术中，对于深度确定性策略梯度（Deep DeterMinistic Policy Gradient, DDPG）以及其他类型的连续决策的强化学习方案来说，其设计的动作空间均是连续的，导致模型训练的计算量过大，难以保证控制模型快速收敛，实用性差。而本实施例中，将动作空间离散化，缩小了强化学习过程的探索空间，减少了训练

控制模型的计算量，提高了控制模型的收敛速度，保证快速训练出控制模型。进一步地，减少了在线使用控制模型确定决策的计算量，提高了确定决策的速度，保证快速确定决策。

步骤三、人工智能设备获取回报函数。

回报函数可以根据对人工智能设备实际运行过程中状态的期望设计，可以设计回报函数为理想状态输出的奖励值为正数，且与理想程度正相关，即，若某个状态越理想，则向回报函数输入该状态时回报函数输出的奖励值越大。同时可以设计回报函数为不理想的状态输出的奖励值为负数，即相当于输出了惩罚值。

通过上述设计回报函数的方式，可以达到以下技术效果：

帮助控制模型学习每个状态对应的决策，进而提升控制模型在使用过程中确定决策的准确性：后续模型训练过程中，控制模型对每个状态的认知均通过回报函数所确定的状态的奖励值确定，控制模型会以奖励值最大化为趋向确定决策。通过设计不理想的状态的奖励值为负，当控制模型时触发这种状态时会得到负奖励值，可以理解为受到了惩罚，则控制模型会认知到后续确定决策时应当避免出现该状态，那么在实际使用控制模型确定决策时，可以达到避免出现不理想的状态的效果。而通过设计理想状态的奖励值为正，且与理想程度正相关，当控制模型时触发这种状态时会得到正奖励值，可以理解为受到了鼓励，则控制模型会认知到后续确定决策时应当趋向该状态，那么在实际使用控制模型确定决策时，可以达到趋向理想状态的效果。

以无人车为例，由于无人车驾驶过程的理想状态包括无人车不发生碰撞、无人车快速行驶、无人车沿着道路行驶等，可以设计回报函数输出的奖励值与发生碰撞负相关、与速度正相关、与 θ 负相关。示例性地，可以设计无人车在未发生碰撞时，回报函数为 $R = V \cos \theta - V \sin \theta$ ，当发生碰撞时， $R = -200$ 。

基于这种回报函数，由于训练过程中发生碰撞时控制模型会被惩罚，从而督促控制模型输出决策避免碰撞。由于速度越大则奖励越大，能够鼓励控制模型输出决策来提高速度，从而保证无人车尽量快速行驶。由于 θ 越小则奖励越大，能够鼓励控制模型输出决策来减小 θ ，从而保证无人车尽量沿着道路行驶，而不会驶离车道、产生偏移的行为。

需要说明的是，本实施例对上述步骤一至步骤三的执行顺序不做限定，可以按照任意顺序分别执行每个步骤。

402、人工智能设备基于状态空间、动作空间与回报函数，采用强化学习算法进行模型训练，得到控制模型。

本实施例中，控制模型用于根据输入的 S 内的 N 维状态，输出 A 内的 M 维离散决策。针对模型训练的具体过程，当人工智能设备进行第 t 次学习时，可以获取 N 个维度的状态 S_t ， S_t 为 N 维状态空间中的一个 N 维状态向量，采用回报函数 R 对 S_t 进行计算，得到奖励值 R_{t-1} ，向当前的控制模型输入 S_t ，得到输出的 A_t ，执行 A_t 后，再获取 S_{t+1} 及 R_t ，将 S_t 、 R_{t-1} 加入经验池，通过经验池中的数据进行模型训练，从而对控制模型进行更新，后续再基于更新的控制模型进行下一次学习。

其中，针对获取 N 个维度的状态的过程，人工智能设备可以配置摄像头、各种传感器、网络通信模块等各种部件，通过各种部件可以感知外部环境的状态以及自身的状态，例如人工智能设备可以通过前置摄像头，采集前方的图像，通过该图像感知前方环境的状态，又如人工智能设备可以通过超声波雷达，采集与周围物体的距离，从而感知自身与障碍物的远近，

又如人工智能设备可以通过加速度传感器,感知自身的加速度以及速度,又如人工智能设备可以通过陀螺仪,感知自身的旋转角度,又如人工智能设备可以通过全球定位系统(Global Positioning System, GPS)传感器,获取道路交通路网数据以及车辆的位置信息等,从而感知自身与道路中轴线之间的距离、自身朝向与道路朝向之间的夹角等。

以无人车为例,在驾驶过程中GPS传感器可以采集位置信息,无人车根据位置信息可以确定当前的朝向与当前行驶的道路的中轴线的方向,根据两个方向可以计算出无人车的中轴线与道路的中轴线之间的夹角 θ 。另外无人车根据位置信息可以计算自身与道路的中轴线之间的距离P。另外驾驶过程中加速度传感器可以采集加速度,无人车可以通过加速度计算出无人车的速度V。

在一种可能的设计中,强化学习的过程中可以采用探索策略进行探索:以进行第t次学习为例,可以以一定的概率按照奖励值 R_{t-1} 确定的 A_t 执行动作,以一定的概率随机执行某个动作,以便增强控制模型的泛化性,锻炼控制模型探索未知环境的能力,保证控制模型能够在实际使用过程中应对复杂多变的实际情况。其中,探索策略可以为 ϵ -greedy策略、乐观初始估计策略、衰减 ϵ -greedy策略、不确定行为优先探索策略、概率匹配策略、信息价值策略等,本实施例对此不做限定。

需要说明的第一点是,针对上述步骤401将动作空间A离散为q个动作的设计,本步骤402中训练模型基于离散的动作空间进行训练时,每次学习时只需从q个动作中选择一个动作,即可作为输出的离散决策,探索空间较小,能够快速收敛。

需要说明的第二点是,进行模型训练时可以在虚拟的模拟器环境中训练模型,也可以在真实环境中训练模型。以无人车为例,可以由某个电子设备运行仿真软件,构建模拟器环境,控制模型可以理解为模拟器环境中运行的进程或线程,通过在模拟器环境中不断进行模拟驾驶,控制模型会不断成熟,直至训练完毕。或者可以由无人车在真实环境上驾驶,则要求真实环境为带有车道线的多车道真实道路,且道路上有其他行驶的移动车辆给无人车制造换道机会,同时其他移动车辆的布置满足一定的随机性,保证模型可以在多样的数据集上训练,以增强模型的泛化能力。

需要说明的第三点是,本实施例中,模型训练的episode的个数、停止规则、具体计算方式均可以根据实际需求确定,episode是指完成一次预设动作的过程,停止规则是指触发模型训练时停止的规则,例如为跑完一圈时停止、到达目的地时停止等。模型训练的具体算法可以包括深度增强学习(Deep Reinforcement Learning, DQN)算法、强化学习(Q-learning)算法等,在实施中,当状态空间的维数较大时可以采用DQN算法,当状态空间的维数较小时可以采用Q-learning算法。本实施例对此均不做限定。

需要说明的第四点是,针对于相关技术中采用卷积神经网络训练出高层决策模型的方案来说,本实施例中采用强化学习算法训练控制模型的过程具有以下技术效果:

第一,控制模型具有底层控制的能力:

相关技术中会收集大量人类开车时的第一视角图像和对应的转向角标注数据,基于卷积神经网络学习出第一视角图像从第一视角图片到转向角的映射函数,以便无人车实际驾驶时根据实时采集的第一视角图片以及映射函数确定转向角。然而,该方案中确定的转向角目前只能达到转向角大小这种高层决策的程度,而无法达到确定左拐的速度、右拐的速度、左拐的加速度、右拐的加速度这种底层控制的程度。而本实施例中设计了每个维度的动作空间,

还可进一步将每个维度的动作空间细化为多个离散的动作，从而将控制模型决策的动作底层化、清晰化、具体化，则可以实现控制模型具有底层控制的能力，即控制模型输出的离散决策能够指示每个维度的具体动作，例如精确到加速度的数值、抬脚的数值等，提高了决策的精确性，实用性强。

第二，控制模型的稳健性强。

相关技术中由于第一视角图像与车辆、道路强相关，导致映射函数的泛化能力差：映射函数只适用于控制拍摄了学习过程的第一视角图片的车辆进行驾驶的情况，实际应用中换了别的车辆进行驾驶时，映射函数确定的转向角的可靠性就会很差。同时，映射函数只适用于学习过程拍摄第一视角图片的车辆所处的道路，实际应用中在别的道路进行驾驶时映射函数确定的转向角的可靠性也会很差，因而该方法难以面对复杂多变的道路情况，也无法适用于其他车辆，稳健性很差。

而本实施例提供的方法，由于控制模型以底层的状态为输入进行训练，底层状态的普适性强，与采集状态的自主体与所处的环境的相关度弱，能够保证针对某种确定的状态所确定的决策可以适用于各种自主体和环境。以无人车为例，不必要求实际使用过程中应用的车辆与学习过程中应用的车辆相同，更不必要求实际应用过程中应用的道路与学习过程中应用的道路相同，在实际使用过程中确定的决策可以适用于各种车辆与道路，从而可以保证可以面对复杂多变的道路情况，稳健性强。

综上所述，通过上述步骤 401 至步骤 402，可以得到以当前状态为输入，以离散决策为输出的控制模型。在下述步骤 403 至步骤 405 中，会基于上述控制模型将状态空间 $\mathcal{S}$ 分成不相交的多个模糊子集 $\{S_i\}$ ，满足每个模糊子集 S_i 的所有状态在控制模型的作用下对应于同一个离散决策 a_i ，并为每个模糊子集生成对应的隶属度函数。

403、对于 N 个维度中的每个维度，人工智能设备将该每个维度的状态空间划分为多个状态区间。

对于 N 维状态空间 $\mathcal{S}$ 的第 i 维 ($i=1, \dots, n$)，可以按一定粒度将其分割为 m_i 个状态区间。可选地，可以采用过分割的方式对每个维度的状态空间进行划分，即要求划分的状态区间的数量尽量多，每个状态区间尽量窄，以保证后续得到的模糊子集的精确性。另外，由于后续会基于每个状态区间的中心值获取每个状态区间的典型离散决策，在划分状态区间时，可以记录每个状态区间的中心值 $s_{ij}, j=1, \dots, m_i$ ，得到每个维度的状态空间的中心值集合。

示例性地，以划分 θ 维度的状态空间为例，可以将 θ 维度的状态空间划分为 M 个状态区间 $\{\Theta_1^1, \Theta_1^2, \dots, \Theta_1^M\}$ ，另外可以记录第 i 个状态区间 Θ_i^j 的中心值 θ_i^j ，得到包含 M 个中心值 $\{\theta_i^1, \theta_i^2, \dots, \theta_i^M\}$ 的中心值集合。假设 θ 维度的状态空间为 $[-1, 1]$ ， $M=200$ ，则中心值集合为 $\{-0.995, -0.985, -0.975, \dots, 0.985, 0.995\}$ 。

404、人工智能设备基于控制模型获取多个状态区间中每个状态区间的典型离散决策，得到多个典型离散决策。

状态区间的典型离散决策是指以状态区间的中心值为控制模型的输入时，控制模型最可能输出的离散决策，即输出的概率最大的离散决策。人工智能设备会基于控制模型获取每个状态区间的典型离散决策，以便后续通过合并典型离散决策相同的状态区间来得到模糊子集。获取典型离散决策的过程具体可以包括以下步骤一至步骤三：

步骤一、对于多个状态区间中的每个状态区间，获取状态区间的多个代表状态，每个代

表状态包括维度上的状态区间的中心值以及其他维度上的状态。

状态区间的代表状态是指在该状态区间的维度的取值为状态区间的中心值，且维度等于状态空间的维度的状态，代表状态用于确定对应状态区间的典型离散决策，代表状态包括状态区间的中心值和其他维度上的状态，该其他维度上的状态可以为其他维度的状态空间的采样值。

以获取第 i 维的第 j 个状态区间的代表状态为例，获取代表状态的过程可以包括 (1) - (2)。

(1) 在 N 个维度中第 i 维以外的其他维度的状态空间进行采样，得到多个采样值。

采样方式可以为蒙特卡洛采样，则采样值为随机采样值，或者采样方式为等间隔采样、以一定的先验概率采样等。采样值的数量可以根据实际需求确定，为了保证精确性，可以要求采样值的数量尽量多，例如采样值的数量为 10000，本实施例对采样方式和采样值的数量不做限定。

(2) 对于该多个采样值中的每个采样值，将该采样值和状态区间的中心值组成一个 N 维向量，作为该状态区间的代表状态。其中，该 N 维向量在第 i 维的取值为 s_{ij} ，在第 i 个维度以外的其他维度取值为其他维度的状态空间的采样值。

可选地，可以为每个维度的所有状态区间进行一次采样，通过采样值分别确定该维度每个状态区间的代表状态：当获取第 i 维的每个状态区间的代表状态时，可以在 N 维状态空间中第 i 维以外的其他维度中每个维度的状态空间均进行采样，得到第 i 维以外的每个维度的状态空间的采样值，将第 i 维的 m 个状态区间的中心值集合中的 m 个中心值分别和第 i 维以外的每个维度的状态空间的采样值进行组合，得到 m 个中心值对应的 m 个代表状态，即得到了第 i 维的每个状态区间的代表状态，从而保证为在第 i 维进行一次采样即可，而无需为 m 个状态区间进行 m 次采样。

示例性地，以获取 θ 维度的状态空间的每个状态区间的代表状态为例，假设随机采样的点的个数为 w ，记状态空间 S 中 θ 以外的维度为 S^o ，可以在 S^o 的状态空间中随机采样 w 个点，则得到 $\{S^{o1}, S^{o2}, \dots, S^{ow}\}$ 这些采样值。对于 θ 维度的中心值集合 $\{\theta_i^1, \theta_i^2, \dots, \theta_i^M\}$ 中的每个中心值 θ^i ($i=1, \dots, M$)，令 θ^i 与 $\{S^{o1}, S^{o2}, \dots, S^{ow}\}$ 中的每个采样值组合，可以得到 w 个代表状态 $\{S^{i1}, S^{i2}, \dots, S^{iw}\}$ ，其中任意代表状态 $S^{ij} = (\theta^i, S^{oj})$ ，且 S^{ij} 的维数与状态空间 S 的维数一致。

步骤二、分别将多个代表状态输入到控制模型中，得到该控制模型输出的多个离散决策。

对于 w 个代表状态中的每个代表状态，可以将该代表状态输入到控制模型中，该控制模型会根据输入的代表状态输出对应的离散决策，通过 w 个代表状态可以得到 w 个离散决策。

步骤三、从多个离散决策中选取出现次数最多的离散决策，作为状态区间的典型离散决策。

在获取离散决策的过程中，可以统计每个离散决策的出现次数，从而选取出现次数最多的离散决策，作为状态区间的典型离散决策。

示例性地，以获取 θ 维度的状态区间 $\{\Theta_i^1, \Theta_i^2, \dots, \Theta_i^M\}$ 中 Θ_i^i 的典型离散决策为例，当通过 Θ_i^i 的 w 个代表状态得到 w 个离散决策时，可以统计这 w 个离散决策中出现次数最多的决策，作为 Θ_i^i 的典型离散决策，记为 A_i 。

405、人工智能设备基于多个典型离散决策，将对应于同一典型离散决策且相邻的多个状

态区间合并为一个模糊子集,得到该每个维度的至少一个模糊子集。

对于每个维度的状态空间来说,当得到该维度的多个状态区间以及对应的多个典型离散决策后,可以对典型离散决策的变化情况进行分析,获取典型离散决策在这些状态区间中发生变化的边沿位置,每确定一次边沿位置则在该边沿位置分割一次,从而将典型离散决策未发生变化的多个状态区间合并为一个模糊子集,进而将该维度的状态空间分割为至少一个模糊子集,这些模糊子集中每个模糊子集对应的离散决策相同,且相邻模糊子集对应的典型离散决策不同。

以人工智能设备为机器人为例,假设肢体动作维度的状态空间包括 100 个状态区间,状态区间 1-状态区间 10 对应的典型离散决策均为下蹲,状态区间 10-状态区间 40 对应的典型离散决策均为站起,状态区间 40-状态区间 80 对应的典型离散决策均为抬手,状态区间 80-状态区间 100 对应的典型离散决策均为跳跃,则根据典型离散决策的变化情况进行分析时,确定典型离散决策在状态区间 10、状态区间 40、状态区间 80 发生了变化,即边沿位置为状态区间 10、状态区间 40、状态区间 80,则状态区间 1-状态区间 10 会合并为一个模糊子集,状态区间 10-状态区间 40 会合并为一个模糊子集,状态区间 40-状态区间 80 会合并为一个模糊子集,状态区间 80-状态区间 100 会合并为一个模糊子集。

以人工智能设备为无人车为例,假设 θ 维度的状态空间得到了状态区间 $\{\Theta_t^1, \Theta_t^2, \dots, \Theta_t^M\}$ 以及对应的典型离散决策集合 $\{A_1, \dots, A_M\}$, 将对应于相同典型离散决策并且相邻的状态区间进行合并,会得到多个新的状态区间 $\{\bar{\Theta}_t^1, \bar{\Theta}_t^2, \dots, \bar{\Theta}_t^{k_\theta}\}$, 记为模糊子集,其中 k_θ 是 θ 维度的模糊子集的个数,从而完成了对 θ 维度的模糊子集的划分。

进一步地,以上阐述了将某一个维度的状态空间划分为多个模糊子集的过程,在实施中可以对每个维度的状态空间均采用此方法进行划分,得到每个维度的状态空间的模糊子集,以人工智能设备为无人车为例,可以对 V 维度的状态空间进行划分,得到 V 维度的模糊子集 $\{\bar{V}^1, \bar{V}^2, \dots, \bar{V}^{k_v}\}$, 对 P 维度的状态空间进行划分,得到 P 维度的模糊子集 $\{\bar{P}^1, \bar{P}^2, \dots, \bar{P}^{k_p}\}$, 其中 k_v 为 V 维度的模糊子集个数, k_p 为 P 维度的模糊子集个数。

综上所述,通过上述步骤 403-步骤 405,将每个维度的状态空间划分为多个模糊子集,这种划分模糊子集的方式可以达到以下技术效果:

相关技术通常依赖于专家经验,通过人工的方式完成模糊子集的划分:当要将某个维度的状态空间划分为多个模糊子集时,需要邀请多位专家,让每位专家依据个人经验人工填写各种状态下的决策,将相同决策对应的状态合并为模糊子集,这种方式极其繁琐、效率低下,准确性较差。进一步地,该方法难以应用在高维状态空间划分模糊子集的场景:由于高维状态空间的状态向量已经为不同维度的状态排列组合,代表的实际运行情况非常复杂,专家已经无法依据个人经验确定决策,故这种方法实用性较差,应用范围狭窄。

而本实施例提供了一种基于强化学习训练得到的控制模型自动划分模糊子集的方式,无需依赖人工确定决策,效率极高。进一步地,可以采用过分割的方式,将每个状态空间分割为大量状态区间,基于大量状态区间的典型离散决策进行合并得到模糊子集时,由于模糊子集的边界非常精确,能够保证模糊子集的准确性较高。进一步地,适用于为高维状态空间划分模糊子集的场景,可以方便快速的扩展到高维空间,以便应用于实际应用中复杂的运行情况,实用性强。

406、人工智能设备按照预设规则,获取每个模糊子集的隶属度函数,隶属度函数用于计算对应模糊子集的隶属度,预设规则为:隶属度函数在每个模糊子集的中心值取 1,在每个

模糊子集的边界值的隶属度取 0.5，在每个模糊子集的相邻的两个模糊子集的中心值取 0。

针对获取模糊子集的隶属度函数的方式，对于每个维度的每个模糊子集，会确定该模糊子集的 5 个点：模糊子集的中心值、模糊子集的左边界值、模糊子集的右边界值、模糊子集的前一个模糊子集的中心值、模糊子集的后一个模糊子集的中心值，将中心值的隶属度取 1，将左边界值、右边界值的隶属度取 0.5，将前一个模糊子集的中心值、后一个模糊子集的中心值的隶属度取 0，将这 5 个点中相邻的点以直线连接，得到一个分段线性函数，作为该模糊子集的隶属度函数。其中，模糊子集的隶属度函数可以如图 5 所示。

具体来说，对于任一维度的第一个模糊子集和最后一个模糊子集来说，由于这两个模糊子集只存在一个相邻的模糊子集，则只需确定 4 个点即可。参见图 6，对于某个维度的第一个模糊子集，可以确定该维度的状态空间的左边界值、模糊子集的中心值、模糊子集的右边界值、模糊子集的后一个模糊子集的中心值，将状态空间的左边界值、模糊子集的中心值的隶属度取 1，模糊子集的右边界值的隶属度取 0.5，后一个模糊子集的中心值取 0，将这 4 个点中相邻的点以直线连接，得到一个分段线性函数，作为第一个模糊子集的隶属度函数。参见图 7，对于某个维度的最后一个模糊子集来说，可以确定该维度的状态空间的右边界值、模糊子集的中心值、模糊子集的左边界值、模糊子集的前一个模糊子集的中心值，将状态空间的右边界值、模糊子集的中心值的隶属度取 1，模糊子集的左边界值的隶属度取 0.5，前一个模糊子集的中心值取 0，将这 4 个点中相邻的点以直线连接，得到一个分段线性函数，作为最后一个模糊子集的隶属度函数。

通过这种构建隶属度函数的方式，可以达到以下技术效果：

第一，隶属度函数具有高解释性，较为有效：当某个维度的状态为某个模糊子集的中心值时，人们通过主观感知通常会给出该状态强属于该模糊子集的评价，而采用该模糊子集的隶属度函数对该状态进行计算时，得到的状态与模糊子集之间的隶属度会为 1，达到最大的隶属度，从而准确地表征出这种强属于模糊子集的程度。而当某个维度的状态为某个模糊子集的边界值时，人们通过主观感知通常会给出该状态大概隶属于该模糊子集，也可能隶属于该模糊子集相邻的模糊子集的模糊评价，而采用该模糊子集的隶属度函数对该状态进行计算时，得到的状态与模糊子集之间的隶属度会为 0.5，状态与该模糊子集相邻的模糊子集之间的隶属度也会为 0.5，两者相等，从而准确地表征出这种大致隶属于模糊子集的程度。而当某个维度的状态为某个模糊子集相邻模糊子集的中心值时，人们通过主观感知通常会给出该状态弱隶属于该模糊子集的评价，而采用该模糊子集的隶属度函数对该状态进行计算时，得到的状态与模糊子集之间的隶属度会为 0，从而准确地表征出这种弱属于模糊子集的程度。也即是，保证通过隶属度函数对状态计算出的隶属度会与人为评价非常匹配，从而合理而准确地表征出状态隶属于模糊子集的程度高低。

第二，步骤简单，能够提高构建隶属度函数的效率：相关技术通常采用曲线拟合的方式构建隶属度函数：预先采集大量的样本，绘制隶属度函数的大致曲线，从抛物型分布、梯形分布等多种典型隶属度函数中确定与该曲线较为匹配的隶属度函数，调整隶属度函数的系数，使其逼近于样本，从而得到模糊子集的隶属度函数，这种方式较为繁琐，效率较低。而本实施例中，只需为每个模糊子集选取 5 个点，赋值、相连即可得到隶属度函数，构建隶属度函数的速度较快，提高了效率。

需要说明的是，上述仅是示例性地以绘制分段线性函数作为构建隶属度函数的方式进行

说明, 在实施中也可以采用其他方式构建隶属度函数, 例如可以采用曲线拟合的方式, 获取抛物型分布函数、梯形分布函数、三角形分布函数等常用隶属度函数, 根据模糊子集对隶属度函数进行曲线拟合, 得到最终的隶属度函数, 当然还可以采用根据模糊子集确定隶属度函数的任一种方式来构建隶属度函数, 本实施例对此不做限定。

综上所述, 上述步骤 403-步骤 406, 将每个维度的状态空间划分为多个模糊子集并得到了每个模糊子集的隶属度函数。下述步骤 407 中人工智能设备会结合每个维度的模糊子集、隶属度函数、控制模型以及当前的状态, 进行决策以控制自身的动作。

407、人工智能设备获取 N 个维度的状态。

本步骤与上述步骤 402 中训练模型时获取状态的过程类似, 在此不做赘述。

408、人工智能设备基于 N 个维度中每个维度的状态的激活模糊子集以及控制模型, 得到多个离散决策, 一个状态的激活模糊子集是指状态的隶属度不为 0 的模糊子集, 控制模型用于根据输入的状态输出对应的离散决策。

对于 N 个维度中的每个维度, 人工智能设备会根据该每个维度的状态, 获取该每个维度的多个模糊子集中的至少一个激活模糊子集, 进而得到 N 个维度的激活模糊子集。根据这些激活模糊子集以及控制模型可得到多个离散决策, 这些离散决策中每个离散决策均为针对确定的 N 个维度的状态时可供考虑的决策, 后续会对这些离散决策加权求和得到连续决策。

获取每个维度的激活模糊子集的具体过程可以包括以下两种可能的的设计。

设计一、对于 N 个维度中的每个维度, 当该每个维度的模糊子集的隶属度不为 0 时, 将该模糊子集作为该每个维度的激活模糊子集。

对于 N 个维度中每个维度的每个模糊子集, 可以采用该每个模糊子集对应的隶属度函数, 对维度的状态进行计算, 得到该每个模糊子集的隶属度, 当该隶属度不为 0 时, 将该模糊子集作为激活模糊子集。也即是, 可以获取各个维度内已划分的模糊子集中隶属度不为 0 的模糊子集, 作为激活模糊子集。

具体来说, 当人工智能设备在实际运行过程中得到 N 个维度的状态 $S = (s_1, \dots, s_n)$ 后, 对于每个维度的 S , 可以采用上述步骤 406 得到的该维度的每个模糊子集的隶属度函数, 分别计算 S 与每个模糊子集之间的隶属度, 记隶属度不为 0 的模糊子集为激活模糊子集。例如, 假设 θ 维度的状态空间具有 5 个模糊子集以及 5 个隶属度函数, 则得到 θ 后, 可以分别采用 5 个隶属度函数对 θ 进行计算, 得到 θ 与 5 个模糊子集中每个模糊子集之间的隶属度, 从这 5 个模糊子集中选取隶属度不为 0 的模糊子集, 作为 θ 维度的激活模糊子集。其中, 依照上述隶属度函数设计原则, 可看出任意一个维度上的激活模糊子集不会超过 2 个。

设计二、对于 N 个维度中的每个维度, 从该每个维度的多个模糊子集中选取中心值位于维度的状态左右的两个模糊子集, 作为该每个维度的激活模糊子集。

当人工智能设备在实际运行过程中得到 N 个维度的状态 $S = (s_1, \dots, s_n)$ 后, 对于每个维度的 S , 可以从该维度对应的模糊子集的中心值集合中找到位于 S 左侧的模糊子集的中心值, 即小于 S 的所有模糊子集的中心值中与 S 最接近的中心值, 将包含该中心值的模糊子集作为激活模糊子集。同理地, 从该维度对应的模糊子集的中心值集合中找到位于 S 右侧的模糊子集的中心值, 即大于 S 的所有模糊子集的中心值中与 S 最接近的中心值, 将包含该中心值的模糊子集作为激活模糊子集。以此类推, 每个维度获取两个激活模糊子集, N 个维度即可获取 n 对激活模糊子集。

示例性地, 以得到的 N 个维度的状态为 $S = (\theta, V, P)$ 为例, 可以从 θ 维度的中心值集合中

找到与 θ 最接近的 $(\bar{\theta}^i, \bar{\theta}^{i+1})$, 从 V 维度的中心值集合中找到与 V 最接近的 $(\bar{v}^j, \bar{v}^{j+1})$, 从 P 维度的中心值集合中找到与 P 最接近的 $(\bar{p}^k, \bar{p}^{k+1})$, 即每个维度找到三对激活模糊子集中心值, 分别为 $(\bar{\theta}^i, \bar{\theta}^{i+1}), (\bar{v}^j, \bar{v}^{j+1}), (\bar{p}^k, \bar{p}^{k+1})$, 则确定激活模糊子集包括 $\bar{\Theta}^i$ 、 $\bar{\Theta}^{i+1}$ 、 $\bar{V}^j$ 、 $\bar{V}^{j+1}$ 、 $\bar{P}^k$ 和 $\bar{P}^{k+1}$ 。

在得到 N 个维度的激活模糊子集后, 可以根据这些激活模糊子集获取多个离散决策。获取离散决策的具体过程可以包括以下步骤一至步骤三:

步骤一、获取每个维度的状态对应的激活模糊子集的中心值, 得到多个中心值。

步骤二、对不同维度的中心值进行组合, 得到多个中间状态, 每个中间状态包括 N 个维度的中心值。

对于 N 个维度中的第 i 维, 可以从第 i 维的至少一个激活模糊子集的中心值选取一个中心值, 遍历 N 个维度后会选取出 N 个中心值, 对这 N 个中心值进行组合, 可以得到一个中间状态, 该中间状态包括 N 个维度, 在任一维度的取值为该维度的某个激活模糊集合的中心值。

示例性地, 当每个维度对应 2 个激活模糊集合时, 得到的多个 n 维中间状态可以表示为 $\{(s_{1,k_1+x_1}, \dots, s_{n,k_n+x_n}), x_i = 0 \text{ 或 } 1\}$ 。其中, s_{i,k_i+x_i} 为 s_i 在第 i 维上激活模糊子集的中心值。示例性地, 假设 N 个维度的状态为 $S = (\theta, V, P)$, 则 3 个维度的模糊子集中心值组合在一起得到 8 个 3 维状态, 可以表示为 $\{S = (\bar{\theta}^{i+x}, \bar{v}^{j+y}, \bar{p}^{k+z}), x, y, z = 0, 1\}$ 。

步骤三、分别将多个中间状态输入到控制模型中, 得到控制模型输出的多个离散决策。

当得到多个 N 维中间状态后, 将任一 N 维中间状态输入到控制模型, 控制模型即可输出该 N 维中间状态的离散决策。示例性地, 以 $\{S_i = (\bar{\theta}^{i+x}, \bar{v}^{j+y}, \bar{p}^{k+z}), x, y, z = 0, 1\}$ 作为控制模型的输入, 可以得到输出的 8 个离散决策 $\{A_{xyz}, x, y, z = 0, 1\}$ 。

409、人工智能设备基于每个维度的状态与激活模糊子集之间的隶属度, 对多个离散决策进行加权求和, 得到连续决策。

由于隶属度函数为连续函数, 基于隶属度函数计算的隶属度对多个离散决策进行加权求和后, 得到的决策为连续决策, 通过连续决策控制人工智能设备执行动作时, 可以保证对人工智能设备的平顺控制, 提高动作的平滑性。

对多个离散决策加权求和得到连续决策的过程具体可以包括以下步骤一至步骤三:

步骤一、对于多个离散决策中的每个离散决策, 获取该每个离散决策对应的 N 个激活模糊子集的隶属度, 得到 N 个隶属度。

针对离散决策对应的 N 个激活模糊子集的概念以及获取 N 个激活模糊子集的过程, 对于每个离散决策, 可以获取之前向控制模型输入中间状态得到输出的离散决策的过程中, 输出该离散决策的中间状态, 再获取之前将多个激活模糊子集的中心值组成中间状态的过程中, 确定该中间状态的 N 个激活模糊子集, 将该 N 个激活模糊子集作为该离散决策对应的 N 个激活模糊子集。

当确定 N 个激活模糊子集后, 当人工智能设备在上述步骤 408 中采用了设计一时, 可以根据上述过程中计算的每个维度的每个模糊子集的隶属度, 获取该 N 个激活模糊子集的隶属度, 得到 N 个隶属度。而当人工智能设备在上述步骤 408 中采用了设计二时, 可以对于该 N 个维度中每个维度的每个模糊子集, 采用该模糊子集对应的隶属度函数, 对该维度的状态进行计算, 得到该模糊子集的隶属度, 从而得到 N 个隶属度。

以第 i 维第 j 个激活模糊子集的隶属度函数表示为 f_{ij} 为例, 每个维度的每个激活模糊子

集的隶属度函数为 $\{f_{ij}, i=1, \dots, n, j=1, \dots, m_i\}$, 可以计算 N 维状态 S 与每个激活模糊子集之间的隶属度, 其中 S 与第 j 个被激活子集之间的隶属度等于 $f_{1,k_1+x_1}^j(s_1) \times \dots \times f_{n,k_n+x_n}^j(s_n)$ 。

以 $S = (\theta, V, P)$ 为例, 当上述步骤 404 得到了激活模糊子集 $\bar{\theta}^i$ 、 $\bar{\theta}^{i+1}$ 、 $\bar{V}^j$ 、 $\bar{V}^{j+1}$ 、 $\bar{P}^k$ 和 $\bar{P}^{k+1}$ 时, 可以采用 $\bar{\theta}^i$ 的隶属度函数对 θ 进行计算, 得到 θ 与 $\bar{\theta}^i$ 之间的隶属度 w_θ^1 , 采用 $\bar{\theta}^{i+1}$ 的隶属度函数对 θ 进行计算, 得到 θ 与 $\bar{\theta}^{i+1}$ 之间的隶属度 w_θ^2 , 采用 $\bar{V}^j$ 的隶属度函数对 V 进行计算, 得到 V 与 $\bar{V}^j$ 之间的隶属度 w_V^1 , 采用 $\bar{V}^{j+1}$ 的隶属度函数对 V 进行计算, 得到 V 与 $\bar{V}^{j+1}$ 之间的隶属度 w_V^2 , 采用 $\bar{P}^k$ 的隶属度函数对 P 进行计算, 得到 P 与 $\bar{P}^k$ 之间的隶属度 w_P^1 , 采用 $\bar{P}^{k+1}$ 的隶属度函数对 P 进行计算, 得到 P 与 $\bar{P}^{k+1}$ 之间的隶属度 w_P^2 。

步骤二、基于 N 个隶属度, 计算每个离散决策的权重。

可选地, 人工智能设备可以计算 N 个隶属度的乘积, 作为该离散决策的权重。例如假设离散决策为 A_{xyz} , 则该离散决策的权重为 $w_\theta^x w_V^y w_P^z$ 。

步骤三、基于每个离散决策的权重, 对多个离散决策进行加权求和, 得到连续决策。

以 $S = (\theta, V, P)$ 对应于 8 个离散决策为例为例, 当得到 8 个离散决策的权重后, 可以基于每个离散决策的权重, 对 8 个离散决策进行加权求和, 得到连续决策为 $\sum_{x,y,z=0,1} w_\theta^x w_V^y w_P^z A_{xyz}$ 。

410、人工智能设备基于连续决策, 控制自身执行对应的动作。

连续决策可以包括 M 个维度, 即上述步骤 401 中 M 维动作空间 A 的维度。人工智能设备可以获取连续决策在每个维度上的取值, 控制自身执行该每个维度的动作。

结合实际使用场景, 以人工智能设备为无人车为例, 对于转向角维度的连续决策, 无人车会按照该连续决策转动一定的角度, 例如转向角维度的连续决策为 -1, 映射为无人车能够右转的最大角度, 则无人车会按照最大角度右转, 又如连续决策为 0, 映射为 0° , 则无人车会直行。对于加速度维度的连续决策, 无人车会按照连续决策进行加速, 例如加速度维度的连续决策为 0.3, 映射的加速度为 $300M/s^2$, 则无人车会按照 $300M/s^2$ 进行加速。以人工智能设备为机器人为例, 对于左脚维度的连续决策, 机器人会按照该连续决策控制左脚动作, 例如左脚维度的连续决策为 5, 映射为抬起左脚 40cm, 则机器人会抬起左脚 40cm。

需要说明的第一点是, 本实施例仅是以执行主体为人工智能设备为例进行说明, 在实施中, 本实施例提供的动作控制方法也可以由服务器执行, 该服务器可以部署于云端, 服务器可以与人工智能设备之间建立网络连接, 通过该网络连接与人工智能设备进行实时通信, 人工智能设备可以将获取的每个维度的状态发送给服务器, 服务器可以基于每个维度的状态以及控制模型, 获取连续决策, 将连续决策发送给人工智能设备, 以便人工智能设备接收该连续决策后, 可以控制自身执行对应的动作, 从而达到服务器远程控制人工智能设备执行动作的效果。其中, 在服务器与人工智能设备进行交互的过程中, 为了保证安全性, 可以要求通信延迟尽量小、通信安全尽量高。

需要说明的第二点是, 通过状态与激活模糊子集之间的隶属度来确定连续决策, 提供了一种合理地连续化决策的方式, 保证连续决策具有高准确性: 由于隶属度能够反映状态在模糊区间内的变化趋势, 随着人工智能设备的运行, 当获取到状态的数值以一定的变化趋势变

化时，状态在模糊子集中的位置也会以该变化趋势变化，则模糊子集的隶属度也会以该变化趋势变化，则基于隶属度确定的连续决策也会以该变化趋势变化，也即是，连续决策的变化趋势会与状态的变化趋势匹配，准确性高。

本实施例提供的方法，基于每个维度的状态与激活模糊子集之间的隶属度，对多个离散决策进行加权求和，得到连续决策，由于输出的决策为连续量，能够保证对人工智能设备的平顺控制，保证动作的平滑性。同时，通过隶属度来获取连续决策，提供了一种合理地连续化离散决策的方式，保证连续决策的变化趋势与状态的变化趋势匹配，从而保证连续决策具有高准确性。进一步地，当人工智能设备为无人车时，可以保证控制无人车的平顺性，提升乘客的舒适度。

图8是本公开实施例提供的一种动作控制装置的结构示意图，如图8所示，该装置包括：获取模块801、计算模块802和控制模块803。

获取模块801，用于执行上述步骤407；

该获取模块801，还用于执行上述步骤408；

计算模块802，用于执行上述步骤409；

控制模块803，用于执行上述步骤410；

在一种可能的设计中，该获取模块801，包括：

获取子模块，用于执行上述步骤408的设计二中的步骤一；

组合子模块，用于上述步骤408的设计二中的步骤二；

输入子模块，用于上述步骤408的设计二中的步骤三。

在一种可能的设计中，该装置还包括：

划分模块，用于执行上述步骤401；

该获取模块801，还用于执行上述步骤404；

该获取模块801，还用于执行上述步骤405。

在一种可能的设计中，该获取模块801，包括：

获取子模块，用于执行上述步骤404中的步骤一；

输入子模块，用于执行上述步骤404中的步骤二；

选取子模块，用于执行上述步骤404中的步骤三。

在一种可能的设计中，该计算模块802，用于计算每个模糊子集的隶属度；

在一种可能的设计中，该获取模块801，还用于执行上述步骤406。

在上述实施例中，可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用软件实现时，可以全部或部分地以计算机程序产品的形式实现。该计算机程序产品包括一个或多个计算机程序指令。在计算机上加载和执行该计算机程序指令时，全部或部分地产生按照本申请实施例该的流程或功能。该计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。该计算机指令可以存储在计算机可读存储介质中，或者从一个计算机可读存储介质向另一个计算机可读存储介质传输，例如，该计算机程序指令可以从一个网站站点、计算机、服务器或数据中心通过有线或无线方式向另一个网站站点、计算机、服务器或数据中心进行传输。该计算机可读存储介质可以是计算机能够存取的任何可用介质或

者是包含一个或多个可用介质集成的服务器、数据中心等数据存储设备。该可用介质可以是磁性介质（例如软盘、硬盘、磁带）、光介质（例如，数字视频光盘（digital video disc, DVD）、或者半导体介质（例如固态硬盘）等。

以上该仅为本公开的可选实施例，并不用以限制本公开，任何熟悉本技术领域的技术人员在本申请揭露的技术范围内，可轻易想到变化或替换，都应涵盖在本申请的保护范围之内。

权 利 要 求 书

1、一种动作控制方法，其特征在于，所述方法包括：

获取人工智能设备的N个维度的状态，所述N为大于或等于1的正整数；

基于所述N个维度中每个维度的状态的激活模糊子集以及控制模型，得到多个离散决策，一个状态的激活模糊子集是指所述状态的隶属度不为0的模糊子集，每个模糊子集是指一个维度内对应于同一个离散决策的状态区间，所述隶属度用于表示状态隶属于模糊子集的程度高低，所述控制模型用于根据输入的状态输出对应的离散决策；

基于所述每个维度的状态与激活模糊子集之间的隶属度，对所述多个离散决策进行加权求和，得到连续决策；

基于所述连续决策，控制所述人工智能设备执行对应的动作。

2、根据权利要求1所述的方法，其特征在于，所述基于所述每个维度的状态与激活模糊子集之间的隶属度，对所述多个离散决策进行加权求和，得到连续决策，包括：

对于所述多个离散决策中的每个离散决策，获取所述每个离散决策对应的N个激活模糊子集的隶属度，得到N个隶属度；

基于所述N个隶属度，计算所述每个离散决策的权重；

基于所述每个离散决策的权重，对所述多个离散决策进行加权求和，得到所述连续决策。

3、根据权利要求1所述的方法，其特征在于，所述基于所述N个维度中每个维度的状态的激活模糊子集以及控制模型，得到多个离散决策，包括：

获取所述N个维度中每个维度的激活模糊子集的中心值，得到多个中心值；

对不同维度的中心值进行组合，得到多个中间状态，每个中间状态包括N个维度的中心值；

分别将所述多个中间状态输入到所述控制模型中，得到所述控制模型输出的多个离散决策。

4、根据权利要求1所述的方法，其特征在于，所述基于所述N个维度中每个维度的状态的激活模糊子集以及控制模型，得到多个离散决策之前，所述方法还包括：

对于所述N个维度中的每个维度，当所述每个维度的状态与所述每个维度的任一模糊子集之间的隶属度不为0时，将所述每个模糊子集作为所述每个维度的激活模糊子集；或，

对于所述N个维度中的每个维度，从所述每个维度的多个模糊子集中选取中心值位于所述每个维度的状态左右的两个模糊子集，作为所述每个维度的激活模糊子集。

5、根据权利要求1所述的方法，其特征在于，所述获取人工智能设备的N个维度的状态之前，所述方法还包括：

对于所述N个维度中的每个维度，将所述每个维度的状态空间划分为多个状态区间；

基于所述控制模型，获取所述多个状态区间中每个状态区间的典型离散决策，得到多个典型离散决策；

基于所述多个典型离散决策，将对应于同一典型离散决策且相邻的多个状态区间合并为一个模糊子集，得到所述每个维度的至少一个模糊子集。

6、根据权利要求5所述的方法，其特征在于，所述基于所述控制模型，获取所述多个状态区间中每个状态区间的典型离散决策，包括：

对于所述多个状态区间中的每个状态区间，获取所述每个状态区间的多个代表状态，每个代表状态包括所述每个维度上的所述每个状态区间的中心值以及其他每个维度上的任一个状态；

分别将所述多个代表状态输入到所述控制模型中，得到所述控制模型输出的多个离散决策；

从所述多个离散决策中选取出现次数最多的离散决策，作为所述每个状态区间的典型离散决策。

7、根据权利要求1所述的方法，其特征在于，所述获取人工智能设备的N个维度的状态之后，所述方法还包括：

对于所述N个维度中每个维度的每个模糊子集，采用所述每个模糊子集对应的隶属度函数，对所述每个维度的状态进行计算，得到所述每个模糊子集的隶属度。

8、根据权利要求7所述的方法，其特征在于，所述获取所述人工智能设备的N个维度的状态之前，所述方法还包括：

按照预设规则，获取每个模糊子集的隶属度函数，所述隶属度函数用于计算对应模糊子集的隶属度，所述预设规则为：隶属度函数在所述每个模糊子集的中心值取1，在所述每个模糊子集的边界值的隶属度取0.5，在所述每个模糊子集的相邻的两个模糊子集的中心值取0。

9、一种动作控制装置，其特征在于，所述装置包括：

获取模块，用于获取人工智能设备的N个维度的状态，所述N为大于或等于1的正整数；

所述获取模块，还用于基于所述N个维度中每个维度的状态的激活模糊子集以及控制模型，得到多个离散决策，一个状态的激活模糊子集是指所述状态的隶属度不为0的模糊子集，每个模糊子集包括一个维度内对应于同一个离散决策的状态区间，所述隶属度用于表示状态隶属于模糊子集的程度高低，所述控制模型用于根据输入的状态输出对应的离散决策；

计算模块，用于基于所述每个维度的状态与激活模糊子集之间的隶属度，对所述多个离散决策进行加权求和，得到连续决策；

控制模块，用于基于所述连续决策，控制所述人工智能设备执行对应的动作。

10、根据权利要求9所述的装置，其特征在于，所述计算模块，包括：

获取子模块，用于对于所述多个离散决策中的每个离散决策，获取所述每个离散决策对应的N个激活模糊子集的隶属度，得到N个隶属度；

计算子模块，用于基于所述N个隶属度，计算所述每个离散决策的权重；

所述计算子模块，还用于基于所述每个离散决策的权重，对所述多个离散决策进行加权求和，得到所述连续决策。

11、根据权利要求9所述的装置，其特征在于，所述获取模块，包括：

获取子模块，用于获取所述N个维度中每个维度的激活模糊子集的中心值，得到多个中心值；

组合子模块，用于对不同维度的中心值进行组合，得到多个中间状态，每个中间状态包括N个维度的中心值；

输入子模块，用于分别将所述多个中间状态输入到所述控制模型中，得到所述控制模型输出的多个离散决策。

12、根据权利要求9所述的装置，其特征在于，所述获取模块，还用于：对于所述N个维度中的每个维度，当所述每个维度的状态与所述每个维度的任一模糊子集之间的隶属度不为0时，将所述每个模糊子集作为所述每个维度的激活模糊子集；或，对于所述N个维度中的每个维度，从所述每个维度的多个模糊子集中选取中心值位于所述每个维度的状态左右的两个模糊子集，作为所述每个维度的激活模糊子集。

13、根据权利要求9所述的装置，其特征在于，所述装置还包括：

划分模块，用于对于所述N个维度中的每个维度，将所述每个维度的状态空间划分为多个状态区间；

所述获取模块，还用于基于所述控制模型，获取所述多个状态区间中每个状态区间的典型离散决策，得到多个典型离散决策；

所述获取模块，还用于基于所述多个典型离散决策，将对应于同一典型离散决策且相邻的多个状态区间合并为一个模糊子集，得到所述每个维度的至少一个模糊子集。

14、根据权利要求12所述的装置，其特征在于，所述获取模块，包括：

获取子模块，用于对于所述多个状态区间中的每个状态区间，获取所述每个状态区间的多个代表状态，每个代表状态包括所述每个维度上的所述每个状态区间的中心值以及其他每个维度上的任一个状态；

输入子模块，用于分别将所述多个代表状态输入到所述控制模型中，得到所述控制模型输出的多个离散决策；

选取子模块，用于从所述多个离散决策中选取出现次数最多的离散决策，作为所述每个状态区间的典型离散决策。

15、根据权利要求9所述的装置，其特征在于，所述计算模块，还用于对于所述N个维度中每个维度的每个模糊子集，采用所述每个模糊子集对应的隶属度函数，对所述每个维度的状态进行计算，得到所述每个模糊子集的隶属度。

16、根据权利要求9所述的装置，其特征在于，所述获取模块，还用于：按照预设规则，

获取每个模糊子集的隶属度函数，所述隶属度函数用于计算对应维度的模糊子集的隶属度，所述预设规则为：隶属度函数在所述每个模糊子集的中心值取 1，在所述每个模糊子集的边界值的隶属度取 0.5，在所述每个模糊子集的相邻的两个模糊子集的中心值取 0。

17、一种人工智能设备，其特征在于，所述人工智能设备包括处理器和存储器，所述存储器中存储有至少一条指令，所述指令由所述处理器加载并执行以实现如权利要求 1 至权利要求 8 中任一项所述的动作控制方法中所执行的操作。

18、一种计算机可读存储介质，其特征在于，所述存储介质中存储有至少一条指令，所述指令由处理器加载并执行以实现如权利要求 1 至权利要求 8 中任一项所述的动作控制方法中执行的操作。

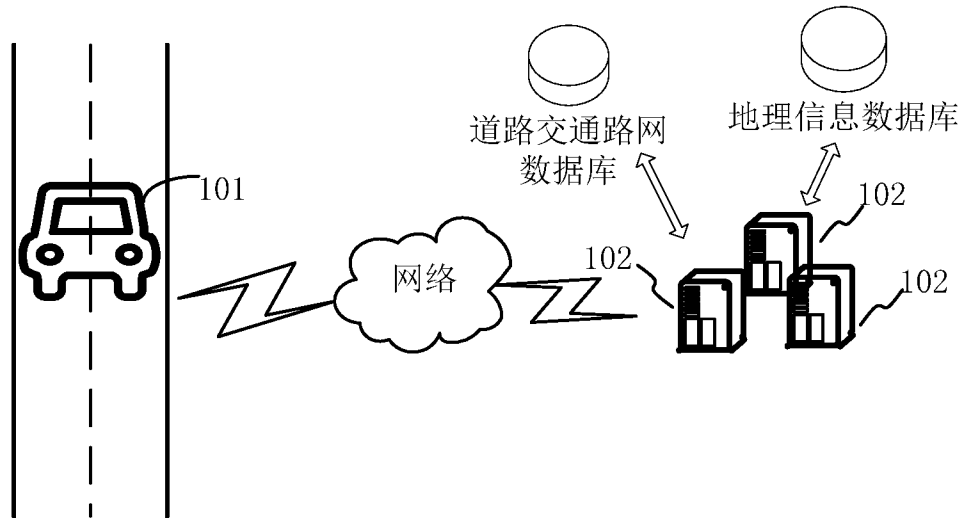


图 1

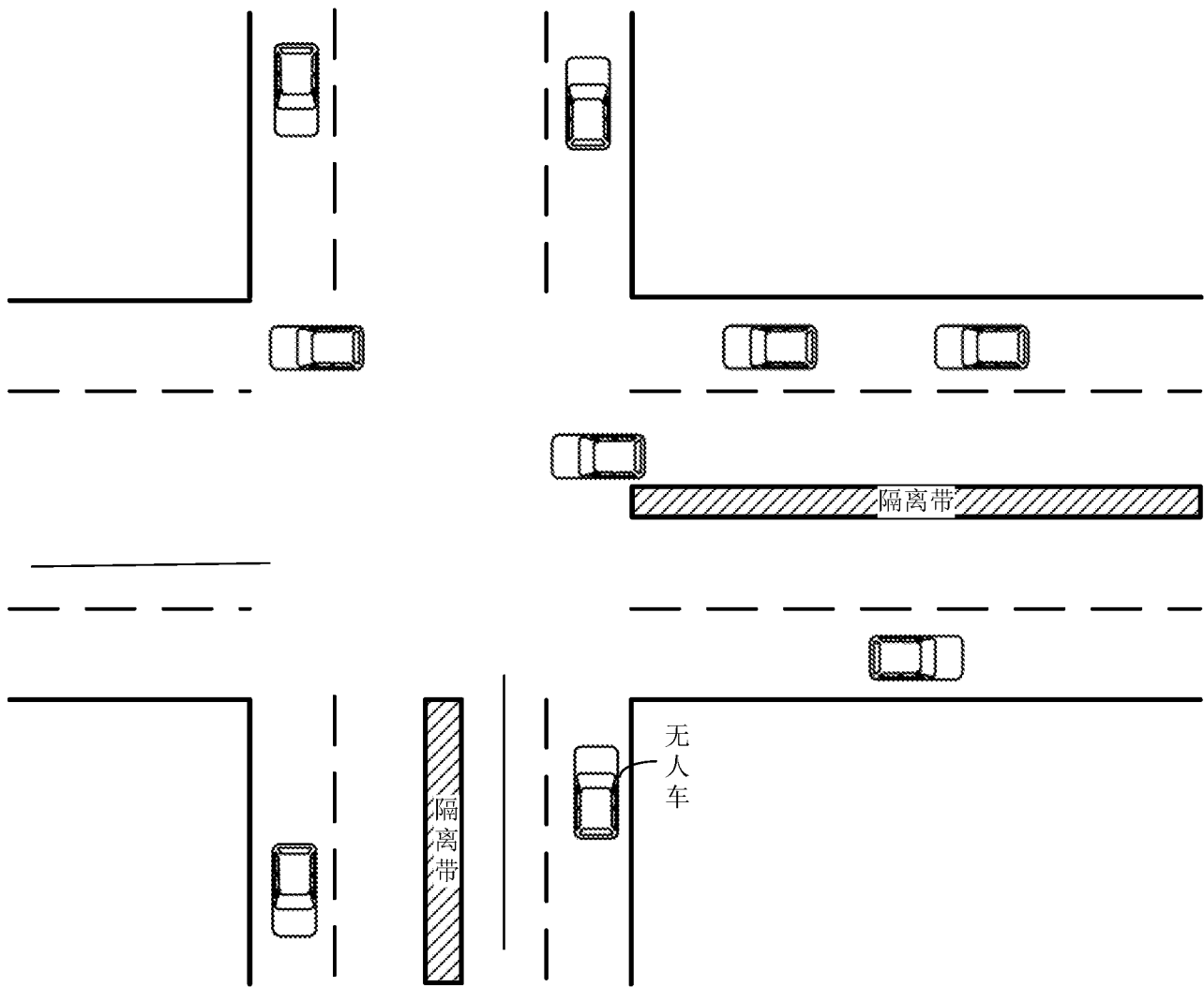


图 2

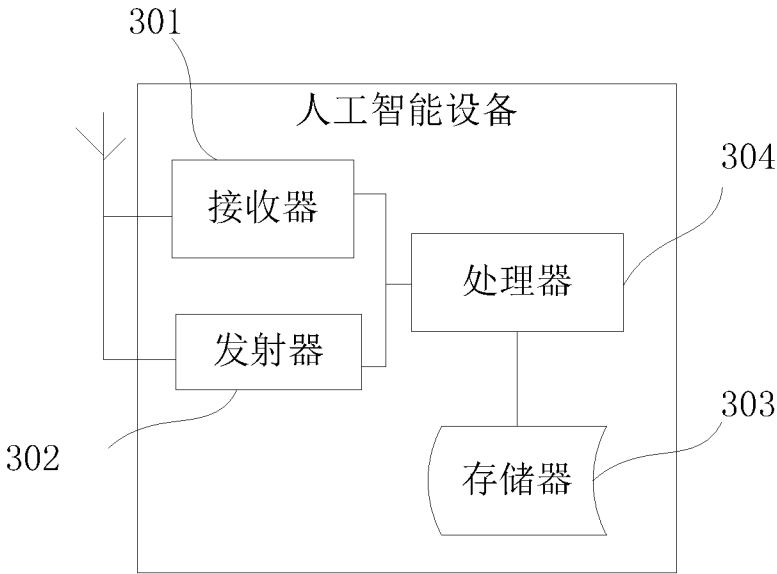


图 3

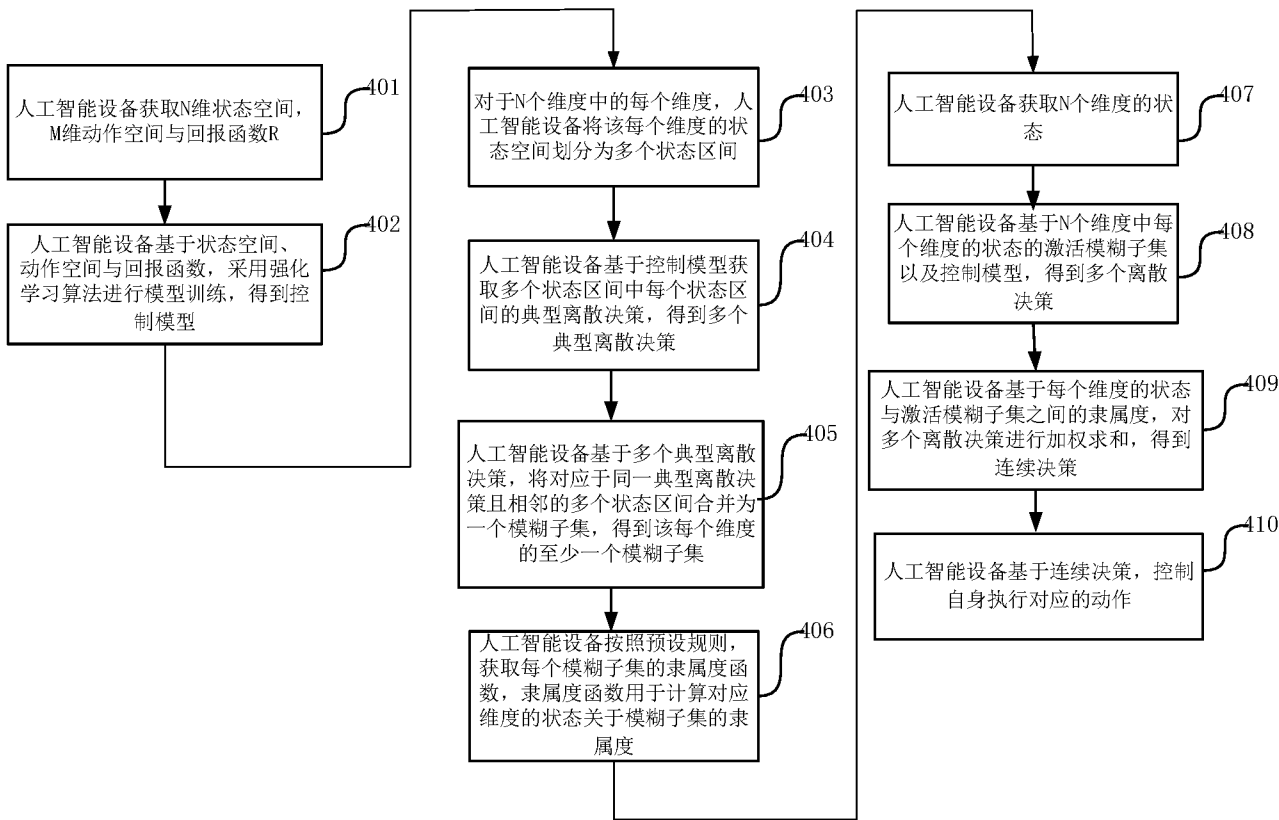


图 4

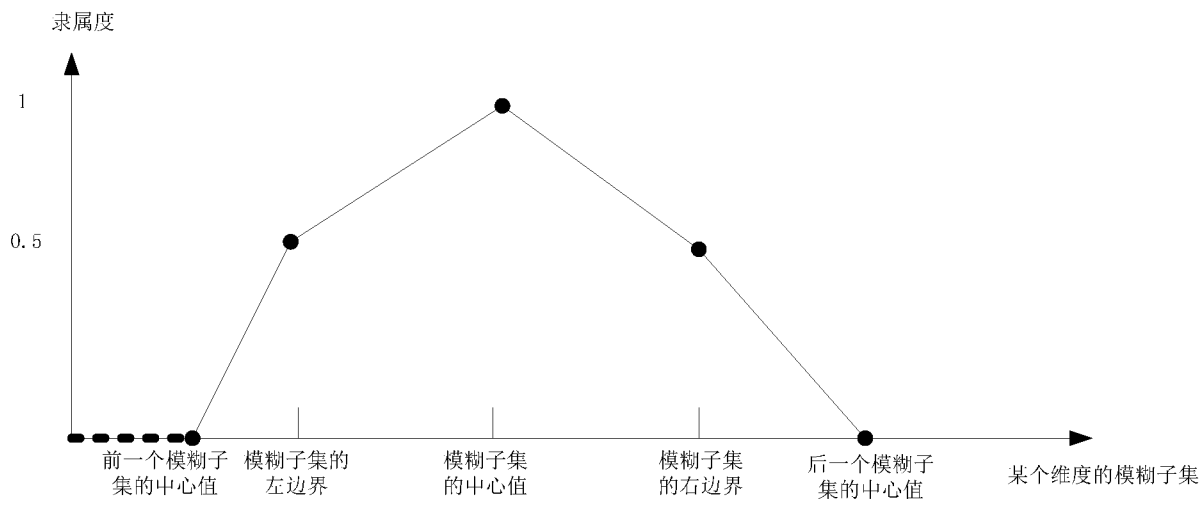


图 5

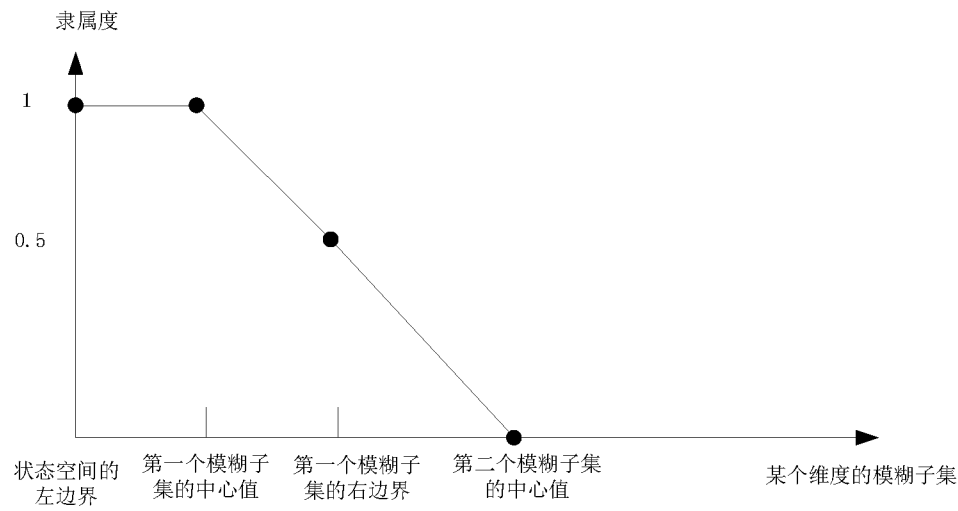


图 6

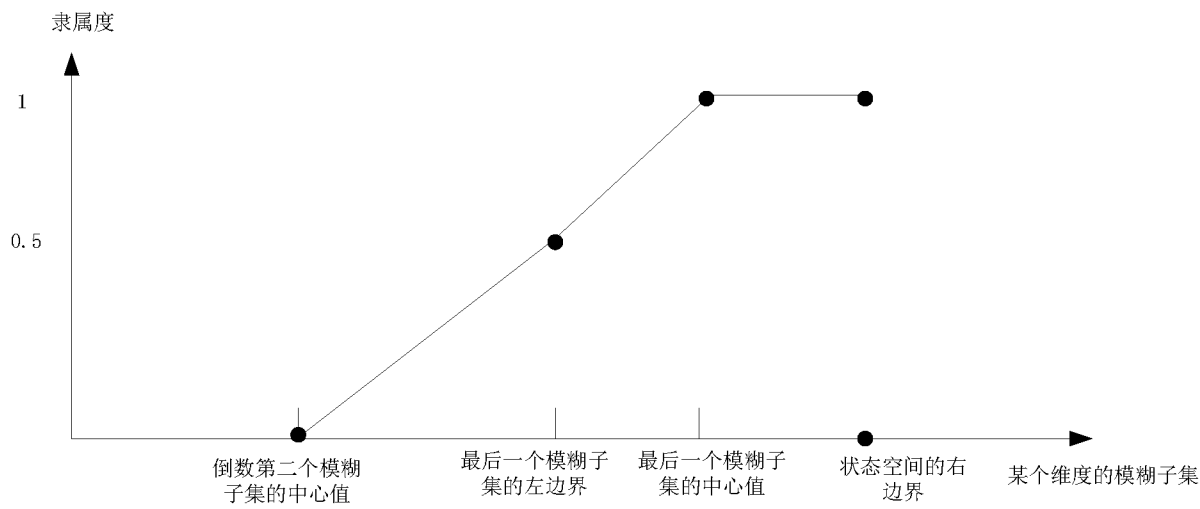


图 7

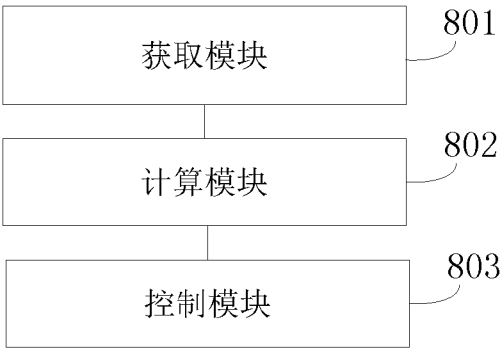


图 8

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/121519

A. CLASSIFICATION OF SUBJECT MATTER

G05D 1/00(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G05D

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS, CNTXT, CNKU, VEN: 连续决策, 离散决策, 离散, 加权求和, 模糊子集, 隶属度, 维度, 连续, 激活模糊子集, continu
+, discrete, weight+, sum, fuzzy.**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 101414159 A (CHONGQING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS) 22 April 2009 (2009-04-22) description page 4, paragraph 4 to page 9, paragraph 4, and figures 1-4	1-18
A	CN 103955930 A (PLA UNIVERSITY OF SCIENCE AND TECHNOLOGY) 30 July 2014 (2014-07-30) entire description	1-18
A	CN 106874668 A (FUDAN UNIVERSITY) 20 June 2017 (2017-06-20) entire description	1-18
A	EP 2933069 A1 (ALDEBARAN ROBOTICS ET AL.) 21 October 2015 (2015-10-21) entire description	1-18



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

07 March 2019

Date of mailing of the international search report

13 March 2019

Name and mailing address of the ISA/CN

State Intellectual Property Office of the P. R. China
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/121519

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	101414159	A	22 April 2009	CN	101414159	B	18 April 2012
CN	103955930	A	30 July 2014	CN	103955930	B	18 January 2017
CN	106874668	A	20 June 2017	None			
EP	2933069	A1	21 October 2015	RU	2016144008	A	17 May 2018
				CA	2946049	C	20 November 2018
				AU	2015248711	A1	03 November 2016
				JP	2017513726	A	01 June 2017
				HK	1216406	A1	11 November 2016
				SG	11201608204Q	A	28 October 2016
				NZ	725275	A	22 December 2017
				MX	2016013016	A	04 October 2017
				WO	2015158885	A2	22 October 2015
				KR	20170030078	A	16 March 2017
				WO	2015158885	A3	10 December 2015
				AU	2015248711	B2	01 February 2018
				CN	106573370	A	19 April 2017
				CA	2946049	A1	22 October 2015
				US	2017144299	A1	25 May 2017
				RU	2016144008	A3	17 May 2018

国际检索报告

国际申请号

PCT/CN2018/121519

A. 主题的分类

G05D 1/00 (2006.01) i

按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献 (标明分类系统和分类号)

G05D

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用))

CNABS, CNTXT, CNKU, VEN: 连续决策, 离散决策, 离散, 加权求和, 模糊子集, 隶属度, 维度, 连续, 激活模糊子集, continu+, discrete, weight+, sum, fuzzy.

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	CN 101414159 A (重庆邮电大学) 2009年 4月 22日 (2009 - 04 - 22) 说明书第4页第4段至第9页第4段、附图1-4	1-18
A	CN 103955930 A (中国人民解放军理工大学) 2014年 7月 30日 (2014 - 07 - 30) 说明书全文	1-18
A	CN 106874668 A (复旦大学) 2017年 6月 20日 (2017 - 06 - 20) 说明书全文	1-18
A	EP 2933069 A1 (ALDEBARAN ROBOTICS等) 2015年 10月 21日 (2015 - 10 - 21) 说明书全文	1-18

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2019年 3月 7日

国际检索报告邮寄日期

2019年 3月 13日

ISA/CN的名称和邮寄地址

中国国家知识产权局 (ISA/CN)
中国北京市海淀区蓟门桥西土城路6号 100088

受权官员

邓薇

传真号 (86-10) 62019451

电话号码 62085578

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/121519

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	101414159	A	2009年 4月 22日	CN	101414159	B	2012年 4月 18日
CN	103955930	A	2014年 7月 30日	CN	103955930	B	2017年 1月 18日
CN	106874668	A	2017年 6月 20日	无			
EP	2933069	A1	2015年 10月 21日	RU	2016144008	A	2018年 5月 17日
				CA	2946049	C	2018年 11月 20日
				AU	2015248711	A1	2016年 11月 3日
				JP	2017513726	A	2017年 6月 1日
				HK	1216406	A1	2016年 11月 11日
				SG	11201608204Q	A	2016年 10月 28日
				NZ	725275	A	2017年 12月 22日
				MX	2016013016	A	2017年 10月 4日
				WO	2015158885	A2	2015年 10月 22日
				KR	20170030078	A	2017年 3月 16日
				WO	2015158885	A3	2015年 12月 10日
				AU	2015248711	B2	2018年 2月 1日
				CN	106573370	A	2017年 4月 19日
				CA	2946049	A1	2015年 10月 22日
				US	2017144299	A1	2017年 5月 25日
				RU	2016144008	A3	2018年 5月 17日

表 PCT/ISA/210 (同族专利附件) (2015年1月)