(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 110000781 B
(45)授权公告日 2021.06.08

(21)申请号 201910255732.8
(22)申请日 2019.03.29
(65)同一申请的已公布的文献号
申请公布号 CN 110000781 A
(43)申请公布日 2019.07.12
(73)专利权人 郑州大学
地址 450001 河南省郑州市高新技术开发区科学大道100号
(72)发明人 王东署 杨凯 罗勇 辛健斌 王河山 马天磊
(74)专利代理机构 焦作市科彤知识产权代理事务所(普通合伙) 41133
代理人 秦贞明
(51)Int.Cl.
B25J 9/16(2006.01)
G06N 3/04(2006.01)
G06N 3/08(2006.01)

审查员 王京京

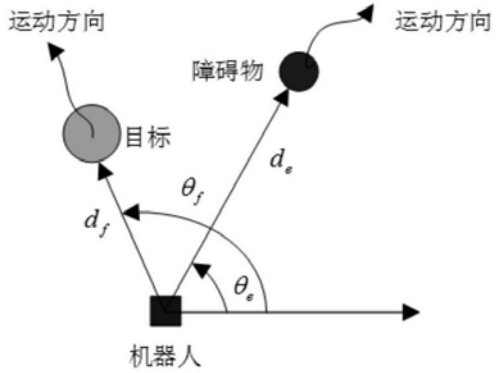
权利要求书3页 说明书9页 附图8页

(54)发明名称

基于发育网络的移动机器人运动方向预先决策方法

(57)摘要

本发明属于机器人智能控制技术领域，公开一种基于发育网络的移动机器人运动方向预先决策方法，包括以下步骤：1)发育网络的创建、训练和测试；2)在每次执行任务后的非工作状态下，由动作输出层激活次数最高的某个神经元触发发育网络中间层神经元的侧向激励机制，实现机器人运动方向的预先决策：计算Z层神经元的激活概率p_i：$p_i=\frac{2}{1+e^{-\gamma_i}}-1$，其中，$\gamma_i=\frac{n_{z区第i个经元激活次数}}{N_{z区所有神经元激活次数总和}}$；按照激活概率大小排序，激活前几个概率不为0的Z层神经元；激活每个Z层神经元时，依次执行如下过程：由Z层向Y层输入数据→激活Y层神经元→侧向激励→在新激活的神经元中保存新知识→建立保存有新知识的Y层神经元与Z层对应的神经元之间的权值连接。本发明可提高机器人的行为决策效率。



CN 110000781 B

1.基于发育网络的移动机器人运动方向预先决策方法，其特征在于，包括以下步骤：

1)发育网络的创建、训练和测试；

发育网络分为三层：X层、Y层、Z层；X层作为网络输入层，其神经元个数与输入矩阵元素个数相同；Y层为中间层，设置了10000个神经元，用于存储知识；Z层作为动作输出层，每个神经元分别代表8个方向之一；

2)在每次执行任务后的非工作状态下，由动作输出层激活次数最高的某个神经元触发发育网络中间层神经元的侧向激励机制，机器人将运动过程中遇到的新知识保存，最终实现机器人运动方向的预先决策：

计算Z层神经元的激活概率p_i：

$$p_i=\frac{2}{1+e^{-\gamma_i}}-1$$，其中，$\gamma_i=\frac{n_{z区第i个经元激活次数}}{N_{z区所有神经元激活次数总和}}$；

按照激活概率大小排序，激活前k个概率不为0的Z层神经元，根据Top-k竞争规则，激活前k个概率不为0的Z层神经元；

激活每个概率不为0的Z层神经元时，依次执行如下过程：由Z层向Y层输入数据→激活Y层神经元→侧向激励→在新激活的神经元中保存新知识→建立保存有新知识的Y层神经元与Z层对应的神经元之间的权值连接；

发育网络的训练包括：

设置多个训练数据，保证智能体不撞上障碍物，经过训练后的机器人将空间状态的机器人、障碍物和目标三者的相对位置情况转换为数据的形式：

网络输入数据：$[\cos(\theta_f),\sin(\theta_f),\cos(\theta_e),\sin(\theta_e),\frac{d_f}{d_f+d_e},\frac{d_e}{d_f+d_e}]$；

网络输出数据：n；

在任意时刻，以机器人为坐标原点建立坐标系，其中：

θ_f：目标与x轴的夹角；

θ_e：障碍物与x轴的夹角；

d_f：目标和机器人的距离；

d_e：障碍物和机器人的距离；

n：大小取值为1～8，代表机器人运动的八个方向，这八个方向将二维平面八等分。

2.根据权利要求1所述的基于发育网络的移动机器人运动方向预先决策方法，其特征在于：

发育网络中，X层到Y层以及Y层到Z层之间的权重更新公式为：

$$v_j\leftarrow\omega_1(n_j)v_j+\omega_2(n_j)y_j\dot{p}$$；

其中，v_j代表第j个神经元的权值向量，$\omega_1(n_j)+\omega_2(n_j)\equiv1$，$\omega_2(n_j)$是学习率，$\omega_1(n_j)$是保持率，对于发放的神经元，$y_j=1$，否则$y_j=0$，$\dot{p}$是归一化后的输入向量，$n_j$表示第j个神经元的发放次数。

3.根据权利要求1所述的基于发育网络的移动机器人运动方向预先决策方法，其特征在于：所述新知识是指新的环境位置信息，其确定依据为：发育网络训练好后，输入相应的

输入信息，计算输入信息与机器人已经学过知识的匹配度，匹配度低于设定阈值的便认为是新知识，匹配度计算公式如下：

$$r(v_b,b,v_t,t)=\frac{v_b}{\|v_b\|}\cdot\frac{b}{\|b\|}+\frac{v_t}{\|v_t\|}\cdot\frac{t}{\|t\|};$$

其中，v_b和v_t分别代表自底向上和自上而下的权值向量，b和t分别代表自底向上和自上而下的输入向量。

4.根据权利要求1所述的基于发育网络的移动机器人运动方向预先决策方法，其特征在于：步骤2)中，根据来自Z层的自顶向下的输入和其对应的权值，利用统一的区域函数，获得Y层神经元发放之前的能量值，根据Top-k竞争规则，激活前k个响应不为零的Y层神经元。

5.根据权利要求1或4所述的基于发育网络的移动机器人运动方向预先决策方法，其特征在于：步骤2)中，假设前四个Z层神经元概率不为0，概率从大到小排序为：[neuron1，neuron3,neuron2,neuron5]，则进入四次循环，依次执行所述过程；第一次循环中，Z层到Y层的输入为[1,0,0,0,0,0,0,0]，然后计算Y层神经元的响应值，激活响应值不为零的神经元，然后将这些神经元按如下公式进行能量值缩放：

$$r_i^{'}=\frac{k-i}{k}r_i;$$

其中，r_i'代表第i个神经元缩放后的能量值，k代表激活的神经元总数，r_i代表第i个神经元的能量值；

这些被激活的Y层神经元发生侧向激励，激活更多的神经元用于记忆新知识；

将机器人遇到的新知识进行分类整理，发育网络根据需要在与方向类别“1”对应的中间层神经元附近侧向激励出新的神经元，用于保存新的对应于方向类别“1”的新知识，侧向激励出的神经元能量值计算公式为：

$$r_{ij}^{'}=e^{-\frac{d^2}{2}}r_i;$$

其中，r'_{ij}表示第i个神经元激活的第j个神经元的能量值，d表示新激活的神经元j与激活它的神经元i之间的距离，r_i代表第i个神经元的能量值；

依次将对应于方向类别“1”中的新知识按照神经元能量值大小存入特定神经元中，这样机器人便学习到了对应于方向类别“1”中的新知识，依次执行以上过程将对应于其它方向类别中的新知识也进行记忆保存。

6.根据权利要求1所述的基于发育网络的移动机器人运动方向预先决策方法，其特征在于，发育网络的测试包括：

在机器人实际运行中的每一步都存在奖励值或惩罚值的调节，从而影响最终的运动方向决策，决定惩罚值和奖励值的公式如下：

$$\alpha=\begin{cases}0 & d_f<d_{2f}\\ \dfrac{1}{d_{1f}-d_{2f}}d_f-\dfrac{d_{2f}}{d_{1f}-d_{2f}} & d_{2f}<d_f<d_{1f}\\ 1 & d_f>d_{1f}\end{cases};$$

其中，α为奖励值，d_{1f}为机器人与目标的初始距离，d_{2f}为机器人追上目标时的距离，d_f为机器人与目标的实时距离；

$$\beta=\begin{cases}0 & d_e>d_s\\ -\dfrac{1}{d_s-d_{ms}}d_e+\dfrac{d_s}{d_s-d_{ms}} & d_{ms}<d_e<d_s\\ 1 & 0<d_e<d_{ms}\end{cases};$$

其中，β为惩罚值大小，d_s为机器人扫描范围，d_e为机器人与障碍物的实时距离，d_{ms}为机器人与障碍物的最小安全距离；

惩罚的方向时刻与机器人扫描到的最近的障碍物的方向相反，惩罚的方向和大小一直处于不断的变化之中，惩罚对机器人根据已掌握知识所做出的决策方向朝着远离障碍物的一侧进行微调，同时使得机器人的行动速度放慢；

奖励的方向时刻指向目标，且只有在其扫描范围内没有障碍物的时候才会存在奖励，奖励的方向和大小一直处于不断的变化之中，奖励机制的存在使得机器人快速地接近目标，同时对机器人做出的决策方向朝着目标方向进行了微调；

机器人在运动过程中会同时受到奖励和惩罚的影响，机器人的最终决策方向由以下公式决定：

$$z=z_i+\alpha*\vec{e}_\alpha+\beta*\vec{e}_\beta;$$

其中，z为最终决策方向，z_i为机器人根据已学到的知识做出的决策，$\vec{e}_\beta$为惩罚方向的单位向量，$\vec{e}_\alpha$为奖励方向的单位向量。

基于发育网络的移动机器人运动方向预先决策方法

技术领域

[0001] 本发明属于机器人智能控制技术领域，具体涉及一种基于发育网络的移动机器人运动方向预先决策方法。

背景技术

[0002] 神经生物学研究表明，感知学习不是与感知皮层的神经活动变化相关联，而是与决策相关的高级区域的神经活动变化相关联。研究发现，在运动方向辨识任务中，行为的改进与侧顶叶内皮层(决策区域)的神经元可塑性相关，但与颞中回(感觉区域)的神经元可塑性无关。

[0003] 既然感知学习与与决策相关的高级区域内的神经元活动变化相关联，很自然就产生了将人脑的感知学习机理引入到机器人的行为决策中来的想法，模仿人脑感知环境过程中的迁移学习机理，使机器人在进行感知学习的过程中，在执行任务的间隙，仍可以进行思考。该过程类似于人类在无外界输入信号时，仍在进行思考，对未来的事情进行预演或彩排，这种预演会对人类后续的行为产生影响。同理，机器人在非任务状态下的思考，可以对后续的运动行为进行预先决策，同时将决策的结果写到数据库中，不断更新数据库，使机器人在下次执行任务时得到更好的决策指导。

[0004] 若能将迁移学习思想应用于机器人领域，使机器人在工作间隙也能进行思考，必将大大提高机器人行为学习的效率。但这些关于感知学习中状态迁移的思想，目前都只应用于模式识别领域，在机器人行为学习中未见使用。

[0005] 近年来，随着人工智能的发展和硬件水平的不断进步，智能机器人也有了很大的进步和更加广阔的应用，对于移动机器人自主行为学习的研究也越来越引起人们的重视。本发明在移动机器人领域引入自主发育网络，使机器人在环境感知过程的间隙，通过非工作状态下的“思考”，将机器人已经学习过的环境位置信息与对应的运动方向预先建立联系，使机器人在后续的环境认知过程中，遇到类似的环境信息时，通过迁移学习，迅速确定运动方向，实现对运动行为的自主决策，不断提高机器人的智能。

[0006] 现有的机器人运动方向预测方法，大多是采用模型预测方法，通过对运动过程进行建模来预测其运动角度、位置或者姿态。但基于模型预测控制的机器人运动控制方法需要辨识模型，分析干扰，确定性能指标，整个问题集合了众多信息，在线计算量较大，难以实时控制，且开环控制+滚动优化的实施需要闭环特性的分析，甚至是标称稳定性的分析，一定程度上限制了该方法的使用。

发明内容

[0007] 针对上述情况，本发明的目的是提供一种基于发育网络的移动机器人运动方向预先决策方法，在机器人工作的间隙，即非工作状态下，通过发育网络中间层神经元的侧向激励机制，在发放的神经元周围激活(或招募)更多的神经元来存储类似的环境位置信息，并将这些位置信息与机器人最佳的运动方向预先建立联系，当机器人在后续环境认知过程

中，遇到类似的环境位置信息时，机器人就可以从已经学习的知识中，迅速确定运动方向，提高行为决策的效率。

[0008] 本发明提供了一种基于发育网络的移动机器人运动方向预先决策方法，包括以下步骤：

[0009] 1)发育网络的创建、训练和测试；

[0010] 发育网络分为三层：X层、Y层、Z层；X层作为网络输入层，其神经元个数与输入矩阵元素个数相同；Y层为中间层，设置了10000个神经元，用于存储知识；Z层作为动作输出层，每个神经元分别代表8个方向之一；

[0011] 2)在每次执行任务后的非工作状态下，由动作输出层激活次数最高的某个神经元触发发育网络中间层神经元的侧向激励机制，机器人将运动过程中遇到的新知识保存，最终实现机器人运动方向的预先决策：

[0012] 计算Z层神经元的激活概率p_i：

[0013] $$p_i=\frac{2}{1+e^{-\gamma_i}}-1$$，其中，$$\gamma_i=\frac{n_{z区第i个经元激活次数}}{N_{z区所有神经元激活次数总和}}$$；

[0014] 按照激活概率大小排序，激活前k个(一般取k＝1)概率不为0的Z层神经元；

[0015] 激活每个概率不为0的Z层神经元时，依次执行如下过程：由Z层向Y层输入数据→激活Y层神经元→侧向激励→在新激活的神经元中保存新知识→建立保存有新知识的Y层神经元与Z层对应的神经元之间的权值连接。

[0016] 本发明主要以发育网络为基础，结合机器人非工作状态下的“思考”和发育网络中间层神经元的侧向激励机制，通过迁移学习，实现机器人环境认知中运动方向的预先决策。其中的发育网络创建、训练等可参照现有技术中的常规方式进行，该发育网络是采用美国密歇根州立大学翁巨杨教授模拟人类大脑的发育规律而提出的一种智能网络，属于本领域的公知常识。另外，本发明中的Z 区、Y区分别是指Z层区域、Y层区域。

[0017] 根据本发明，发育网络中，X层到Y层以及Y层到Z层之间的权重更新公式为：

[0018] $$v_j \leftarrow \omega_1(n_j)v_j+\omega_2(n_j)y_j\,\dot{p}$$；

[0019] 其中，v_j代表第j个神经元的权值向量，$\omega_1(n_j)+\omega_2(n_j)\equiv 1$，$\omega_2(n_j)$是学习率，$\omega_1(n_j)$是保持率，p为输入矩阵，对于发放的神经元，$y_j$＝1，否则$y_j$＝0。

[0020] 本发明中，Z层作为动作输出层，每个神经元分别代表8个方向之一，也可以扩展为更多的运动方向。

[0021] 本发明中，所述新知识是指新的环境位置信息。新知识的确定依据为：发育网络训练好后，输入相应的输入信息，计算输入信息与机器人已经学过知识的匹配度，匹配度低于设定阈值的便认为是新知识，匹配度计算公式如下：

[0022] $$r(v_b,b,v_t,t)=\frac{v_b}{\|v_b\|}\cdot\frac{b}{\|b\|}+\frac{v_t}{\|v_t\|}\cdot\frac{t}{\|t\|}$$；

[0023] 其中，v_b和v_t分别代表自底向上和自上而下的权值向量，b和t分别代表自底向上和自上而下的输入向量。

[0024] 根据本发明，步骤2)中，将这些被激活的Z层神经元与所有Y层神经元建立权值连接，根据来自Z层的自顶向下的输入和其对应的权值，利用统一的区域函数，获得Y层神经元

发放之前的能量值，根据Top-k竞争规则，激活前k 个响应不为零的Y层神经元。

[0025] 进一步地，步骤2)中，假设前四个Z层神经元概率不为0，概率从大到小排序为：[neuron1,neuron3,neuron2,neuron5]，则进入四次循环，依次执行所述过程；第一次循环中，Z层到Y层的输入为[1,0,0,0,0,0,0,0]，然后计算Y层神经元的响应值，激活响应值不为零的神经元(这些被激活的Y层神经元均是属于第一类，即方向“1”所对应的神经元，其只与Z层第一个神经元有连接)，然后将这些神经元按如下公式进行能量值缩放：

[0026] $$r_i^{'}=\frac{k-i}{k}r_i;$$

[0027] 其中，r_i'代表第i个神经元缩放后的能量值，k代表激活的神经元总数，r_i代表第i个神经元的能量值；

[0028] 这些被激活的Y层神经元发生侧向激励，激活更多的神经元用于记忆新知识；

[0029] 将机器人遇到的新知识进行分类整理，发育网络根据需要在与方向类别“1”对应的中间层神经元附近侧向激励出新的神经元，用于保存对应于方向类别“1”的新知识，侧向激励出的神经元能量值计算公式为：

[0030] $$r_{ij}^{'}=e^{-\frac{d^2}{2}}r_i;$$

[0031] 其中，r'_{ij}表示第i个神经元激活的第j个神经元的能量值，d表示新激活的神经元j与激活它的神经元i之间的距离，r_i代表第i个神经元的能量值；

[0032] 依次将对应于方向类别“1”的新知识(新的环境位置信息)按照神经元能量值大小存入特定神经元中，这样机器人便学习到了对应于方向类别“1”中的新知识，依次执行以上过程将对应于其它方向的新知识也进行记忆保存。机器人在每次运行完之后，就可以将学习到的新知识与对应的运动方向建立连接，在后续的运动过程中，若遇到类似的环境位置信息，机器人就可以迅速确定运动方向，提高行为决策的效率。

[0033] 存储新知识时，根据机器人获得的新环境位置信息与这些被激活的神经元中已经存储的知识的匹配程度，确定这些新知识应该储存到哪一个被激活神经元周围的神经元中去，确定目标神经元之后，机器人将前一次环境探索过程中遇到的未训练过的新位置数据保存到这个目标神经元周围新激活的神经元中，并确定最佳运动方向(即与Z层中某个神经元建立联系)，然后在这些新激活的且保存了新知识的Y层神经元与Z层对应的神经元(代表不同的运动方向)之间建立权值连接(为后续的迁移学习创造条件)。

[0034] 当非工作状态结束，机器人已经将非工作状态过程中学习到的新知识进行了存储记忆。在机器人进行后续环境探索的过程中，当遇到类似位置情况的时候，机器人可以迅速根据数据库中已存储的位置信息与运动方向之间的连接，快速做出判断，选择最佳的运动方向(实现迁移学习)，提高行为决策的效率。

[0035] 优选地，发育网络的训练包括：

[0036] 设置多个训练数据，保证智能体不撞上障碍物，经过训练后的机器人将空间状态的机器人、障碍物和目标三者的相对位置情况转换为数据的形式：

[0037] 网络输入数据：$$[\cos(\theta_f),\sin(\theta_f),\cos(\theta_e),\sin(\theta_e),\frac{d_f}{d_f+d_e},\frac{d_e}{d_f+d_e}];$$

[0038] 网络输出数据:n;

[0039] 在任意时刻,以机器人为坐标原点建立坐标系,其中:

[0040] θ_f:目标与x轴的夹角;

[0041] θ_e:障碍物与x轴的夹角;

[0042] d_f:目标和机器人的距离;

[0043] d_e:障碍物和机器人的距离;

[0044] n:大小取值为1～8,代表机器人运动的八个方向,这八个方向将二维平面八等分。

[0045] 优选地,发育网络的测试包括:

[0046] 在机器人实际运行中的每一步都存在奖励值或惩罚值的调节,从而影响最终的运动方向决策,决定惩罚值和奖励值的公式如下:

[0047] $$\alpha=\begin{cases}0 & d_f<d_{2f}\\ \dfrac{1}{d_{1f}-d_{2f}}d_f-\dfrac{d_{2f}}{d_{1f}-d_{2f}} & d_{2f}<d_f<d_{1f}\\ 1 & d_f>d_{1f}\end{cases};$$

[0048] 其中,α为奖励值,d_{1f}为机器人与目标的初始距离,d_{2f}为机器人追上目标时的距离,d_f为机器人与目标的实时距离;

[0049] $$\beta=\begin{cases}0 & d_e>d_s\\ -\dfrac{1}{d_s-d_{ms}}d_e+\dfrac{d_s}{d_s-d_{ms}} & d_{ms}<d_e<d_s\\ 1 & 0<d_e<d_{ms}\end{cases};$$

[0050] 其中,β为惩罚值大小,d_s为机器人扫描范围,d_e为机器人与障碍物的实时距离,d_{ms}为机器人与障碍物的最小安全距离;

[0051] 惩罚的方向时刻与机器人扫描到的最近的障碍物的方向相反,惩罚的方向和大小一直处于不断的变化之中,惩罚对机器人根据已掌握知识所做出的决策方向朝着远离障碍物的一侧进行微调,同时使得机器人的行动速度放慢;

[0052] 奖励的方向时刻指向目标,且只有在其扫描范围内没有障碍物的时候才会存在奖励,奖励的方向和大小一直处于不断的变化之中,奖励机制的存在使得机器人快速地接近目标,同时对机器人做出的决策方向朝着目标方向进行了微调;

[0053] 机器人在运动过程中会同时受到奖励和惩罚的影响,机器人的最终决策方向由以下公式决定:

[0054] $$z=z_i+\alpha*\vec{e}_\alpha+\beta*\vec{e}_\beta;$$

[0055] 其中,z为最终决策方向,z_i为机器人根据已学到的知识做出的决策,$\vec{e}_\beta$为惩罚方向的单位向量,$\vec{e}_\alpha$为奖励方向的单位向量。

[0056] 由于每一步的决定,机器人都是根据已经记忆的知识做出的决策,实际的位置情况和识别出来的位置情况是有差别的,假设实际输入为$x=\{x_1,x_2,x_3,x_4,x_5,x_6\}$,Y层激活神经元的权重信息为$w=\{w_1,w_2,w_3,w_4,w_5,w_6\}$,据此可以定义某一步的识别精度值

$e=\sum_{i=1}^{6}|x_i-w_i|$，e越小，代表识别率越高，e越大，代表识别率越低。

[0057] 另外，本发明中未加以限定的具体操作步骤均可参照现有技术进行设定，如发育网络的创建、训练，数据在神经元中的保存、权值连接的建立等。

[0058] 与现有技术相比，本发明具有如下有益效果：

[0059] 本发明通过发育网络算法，对机器人进行训练，对探索的环境进行认知学习，通过环境感知过程中的迁移学习，使其在后续的环境认知过程中，在遇到类似的环境信息时，可以迅速确定运动方向，提高运动方向决策的效率；具体地，本发明通过机器人发育网络中间层神经元的侧向激励机制，在认知的环境位置信息和运动方向之间预先建立连接，使机器人在后续的运动过程中，遇到类似的环境位置情况时，机器人可以迅速确定最佳的运动方向，提高行为决策的效率。与传统的基于模型预测的方法相比，本发明提出了一种更具前瞻性的运动方向预测方法，可以更高效地预测机器人的运动方向，在提高机器人智能的同时也有效地提高了机器人的工作效率。

附图说明

[0060] 图1：机器人、目标和障碍物之间位置关系示意图；

[0061] 图2：惩罚机制对机器人下一步决策的影响示意图；

[0062] 图3：奖励机制对机器人下一步决策的影响示意图；

[0063] 图4：神经元侧向激励范围示意图；

[0064] 图5：静态环境下Y区域中存储知识的神经元分布示意图；

[0065] 图6：机器人五次运行的路径图；

[0066] 图7：图6中路径的局部放大图；

[0067] 图8：保存知识的神经元个数变化情况；

[0068] 图9：非工作状态结束后发育网络中间层神经元中知识存储情况；

[0069] 图10：运行识别误差折线图；

[0070] 图11：动态环境下Y区域中存储知识的神经元分布示意图；

[0071] 图12：动态环境下机器人中存储知识的神经元个数与运行次数之间的关系；

[0072] 图13：运行完成后机器人中存储知识的神经元分布情况；

[0073] 图14：动态环境下机器人五次运动路径示意图；

[0074] 图15：动态环境下机器人五次运行的误差折线图。

具体实施方式

[0075] 为了使本技术领域的研究与应用人员更好地理解本发明方案，下面将应用一个本方案的仿真结果进行分析，同时也是对本方案具体应用场景的一种验证。同时，下面的实例只是该方案在某一个场景的应用，并非方案的全部应用场景。基于本发明中的实例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实例，都属于本发明保护的范围。

[0076] 实施例

[0077] 一种基于发育网络的移动机器人运动方向预先决策方法，该方法包括以下步骤：

[0078] 1:网络创建

[0079] X层6个神经元分别代表输入数据x向量的6个元素。

[0080] Y层10000个神经元,X层到Y层权重向量、Y层到Z层权重向量初始化为随机数,响应值向量初始化为0年龄初始化为1。

[0081] Z层设置8个神经元,Z层到Y层权重向量初始化为0,年龄为1。

[0082] 2:网络训练

[0083] 训练数据设置了152个,这152个数据可以保证智能体不撞上障碍物,但接近目标的路径不是最优的。在二维平面上机器人、障碍物、目标的相对位置情况有无数多种,因此,经过训练后的机器人只是学到了部分知识,将空间状态的三者相对位置情况转换为数据的形式为:

[0084] 网络输入数据:$[\cos(\theta_f),\sin(\theta_f),\cos(\theta_e),\sin(\theta_e),\frac{d_f}{d_f+d_e},\frac{d_e}{d_f+d_e}]$;

[0085] 网络输出数据:n;

[0086] 在任意时刻,以机器人为坐标原点建立坐标系,如图1所示,其中:

[0087] θ_f:目标与x轴的夹角;

[0088] θ_e:障碍物与x轴的夹角;

[0089] d_f:目标和智能体的距离;

[0090] d_e:敌人和智能体的距离;

[0091] n:大小取值为1～8,代表机器人运动的八个方向,这八个方向将二维平面八等分。

[0092] 3:网络测试

[0093] 创建机器人,障碍物和目标,机器人用蓝色的正方形表示,障碍物为半径为6的黑色圆形,目标为半径为10的绿色圆形,在机器人实际运行中的每一步都存在奖励或惩罚值的调节,从而影响最终的运动方向决策。当机器人接近目标时均存在奖励值,并且距离目标越远,奖励值越大,随着机器人与目标的距离越来越小,奖励值也随之减小,最终当追上目标时候,奖励值减为0,奖励值的存在,使得机器人可以更快的接近目标。机器人在接近目标时,若遇到障碍物(机器人周围距离80像素内扫描到障碍物),此时,仅存在惩罚值的调节,机器人距离最近的障碍物为80像素时候,惩罚值为0,距离障碍物越近,惩罚值越大,从而使机器人更加有效地躲避障碍物。决定惩罚和奖励值的公式如下:

[0094] $$\alpha=\begin{cases}0 & d_f<d_{2f}\\ \frac{1}{d_{1f}-d_{2f}}d_f-\frac{d_{2f}}{d_{1f}-d_{2f}} & d_{2f}<d_f<d_{1f}\\ 1 & d_f>d_{1f}\end{cases};$$

[0095] 其中,α为奖励值,d_{1f}为机器人与目标的初始距离,d_{2f}为机器人追上目标时的距离,d_f为机器人与目标的实时距离。

[0096] $$\beta=\begin{cases}0 & d_e>d_s\\ -\dfrac{1}{d_s-d_{ms}}d_e+\dfrac{d_s}{d_s-d_{ms}} & d_{ms}<d_e<d_s\\ 1 & 0<d_e<d_{ms}\end{cases};$$

[0097] 其中，β为惩罚值大小，d_s为机器人扫描范围，d_e为机器人与障碍物的实时距离，d_{ms}为机器人与障碍物的最小安全距离。

[0098] 惩罚的方向时刻与机器人扫描到的最近的障碍物的方向相反，它的方向和大小一直处于不断的变化之中，惩罚对机器人根据已掌握知识所做出的决策方向朝着远离障碍物的一侧进行微调，同时，使得机器人的行动速度放慢，惩罚对机器人最终方向的影响如图2所示。

[0099] 奖励的方向时刻指向目标，且只有在其扫描范围内没有障碍物的时候才会存在奖励，它的方向和大小一直处于不断的变化之中，奖励机制的存在使得机器人快速地接近目标，同时，对机器人做出的决策方向朝着目标方向进行了微调，奖励对机器人最终运动方向的影响如图3所示。

[0100] 图2和图3是在惩罚和奖励不同时存在的情况下的运动情况分析，但机器人在运动过程中，一般会同时受到奖励和惩罚的影响，因此机器人的最终决策方向由以下公式决定：

[0101] $$z=z_i+\alpha*\vec{e}_\alpha+\beta*\vec{e}_\beta;$$

[0102] 其中，z为最终决策方向，z_i为机器人根据已学到的知识做出的决策，$\vec{e}_\beta$为惩罚方向的单位向量，$\vec{e}_\alpha$为奖励方向的单位向量。

[0103] 由于每一步的决定，机器人都是根据已经记忆的知识做出的决策，实际的位置情况和识别出来的位置情况是有差别的，假设实际输入为x＝{x_1,x_2,x_3, x_4,x_5,x_6}，Y层激活神经元的权重信息为w＝{w_1,w_2,w_3,w_4,w_5,w_6}，据此可以定义某一步的识别精度值 $e=\sum_{i=1}^{6}|x_i-w_i|$，e越小，代表识别率越高，e越大，代表识别率越低。

[0104] 4:非工作状态

[0105] 首先计算Z层神经元激活概率：

[0106] $$p_i=\frac{2}{1+e^{-\gamma_i}}-1 \qquad \gamma_i=\frac{n_{z区第i个经元激活次数}}{N_{z区所有神经元激活次数总和}};$$

[0107] 按照激活概率大小排序，激活前k个概率不为0的Z层神经元，假设前四个神经元概率不为0，概率从大到小排序为:[neuron1,neuron3,neuron2,neuron5]，则进入四次循环，依次执行如下过程:由Z层向Y层输入数据→激活Y层神经元→侧向激励→保存数据→建立新的位置关系与机器人运动方向之间的权值连接。如第一次循环，Z层到Y层的输入为[1,0,0,0,0,0,0,0]，然后计算Y层神经元响应，激活响应不为零的神经元(这些被激活的Y层神经元均是属于第一类，既方向“1”对应的神经元，既只与Z层第一个神经元有连接)，然后将这些神经元按如下公式进行能量值缩放：

[0108] $$r_i'=\frac{k-i}{k}r_i;$$

[0109] 其中，r_i'代表第i个神经元缩放后的能量值，k代表激活的神经元总数，r_i代表第i个神经元的能量值。这些被激活的神经元发生侧向激励，激活更多的神经元用于记忆新的知识，侧向激励的激活范围如图4所示。

[0110] 图4中的数字表示与激活神经元的距离，颜色越深代表激活的神经元能量值越大，反之，则越小。侧向激励出的神经元能量值计算公式如下：

[0111] $$r_{ij}^{'}=e^{-\frac{d^2}{2}}r_i;$$

[0112] 其中，r'_{ij}表示第i个神经元激活的第j个神经元的能量值，d表示新激活的神经元j与激活它的神经元i之间的距离，r_i代表第i个神经元的能量值。其中，侧向激励范围可根据实际数据量大小灵活改变。

[0113] 将这些被激活的Z层神经元与所有Y层神经元建立权值连接，根据来自Z 层的自顶向下的输入和其对应的权值，利用统一的区域函数，获得Y层神经元发放之前的能量值。根据Top-k竞争规则，激活前k个响应不为零的Y层神经元，这些被激活的Y层神经元发生侧向激励，在自身周围激活更多的神经元用于记忆新的知识。

[0114] 存储新知识时，根据机器人获得的新环境位置信息与这些被激活的神经元中已经存储的知识的匹配程度，确定这些新知识应该储存到哪一个被激活神经元周围的神经元中去，确定目标神经元之后，机器人将前一次环境探索过程中遇到的未训练过的新位置数据保存到这个目标神经元周围新激活的神经元中，并确定最佳运动方向(即与Z层中某个神经元建立联系)，然后在这些新激活的且保存了新知识的Y层神经元与Z层对应的神经元(代表不同的运动方向)之间建立权值连接(为后续的迁移学习创造条件)。

[0115] 当非工作状态(线下过程)结束，机器人已经将非工作状态过程中学习到的新知识进行了存储记忆。在机器人进行后续环境认知的过程中，当遇到类似位置情况的时候，机器人可以迅速根据数据库中已存储的位置信息与运动方向之间的连接，快速做出判断，选择最佳的运动方向(实现迁移学习)，提高行为决策的效率。

[0116] 5:结果分析

[0117] 5.1静态环境试验

[0118] 设置了13个静态障碍物和一个目标，障碍物为半径为6的黑色圆形，目标为半径为10的绿色圆形，机器人为蓝色正方形。训练之后，Y区域存储知识的神经元分布情况如图5所示，每个方格代表一个神经元，白色代表空白神经元，即没有存储知识，蓝色代表存储有知识的神经元，训练完成之后有152 个神经元存储了相应的知识。

[0119] 在训练数据的基础上第一次运行路径如图6中带“+”标记的路线，机器人行走187步后追上目标，机器人是在原有152个训练数据的基础上做出的决策。机器人第二次运行的路径用红色带“*”号标记的路线所示，行走步数为176步，且机器人选取了不同的路径。原因在于第一次测试运行之后，在非工作状态即线下过程中，机器人将第一次运行过程中学习到的新知识进行了整理记忆，即把遇到的类似的情况提取出来保存到存储了相似特征的神经元周围的神经元之中，这样机器人在第二次行走的时候已经具备了新的知识，存储的知识量更多，机器人在行走过程中遇到了新的情况便做出了与第一次运行时不同的决策。同理，在第二次运行之后，由于机器人走了不同的路径，所以又学习到了新的知识，在非工作状态下，机器人将第二次运行过程中学习到的新知识进行了整理记忆。在机器人第三次运

行时，由于学习了新的知识，所以选择了图6中带黄色“口”标记的路径，共181步。第四次运行时候，机器人已经进行了三次迁移学习，选择了新的路径，但与第三次的路径差别不大，仅在中间部分存在差异。第五次运行时机器人选择了与第四次基本一样的路径，均为171步，可以看出，机器人在接近目标的过程中遇到的新知识越来越少，意味着机器人在接近该静态目标时候已经学习到了足够的数据，因此第五次和第四次的运行轨迹基本一致。注意：对新的知识定义为匹配度低于0.99的数据。用另一种方法解释第四次和第五次运行轨迹重合的原因，在于第四次运行时，遇到的新的位置情况与此前已经遇到过的位置情况(既学过的知识)的匹配度大部分都高于0.99，因此第四次学习到的新知识很少，第五次运行路径和第四次基本重合，但不是完全重合，如在机器人运行的后段轨迹与第四次略有差别，局部放大图如图7所示。

[0120] 图8为机器人每次运行后的知识存储量，即存储了知识的神经元个数，从图8可以看出，每次运行后机器人都学习了新的知识，且学习的新知识越来越少，其原因是由于设置的目标和障碍物均为静态的，随着运行次数的增加，机器人对环境越来越熟悉，新学习的知识就相应减少。

[0121] 图9所示为运行完之后，机器人的知识存储分布情况。从图9可以看出，其中多了一些聚合在一起的数据，这是由于发育网络中间层神经元侧向激励的作用，使某些神经元激活周围的神经元存储了新的环境位置数据。

[0122] 图10是5次运行的误差折线图，可以计算出每次运行的平均误差分别为：0.8602，0.3663，0.2179，0.2444和0.2319，折线上的某个点代表了机器人在某一步的识别误差值，由图10可以看出，在前三次运行中，机器人每运行一次后，其误差折线图都会降低一点，即其平均误差越来越低，到第三次运行时，平均误差达到极限，之后的每次运行其平均误差都位于0.23左右。可以这样解释，把机器人接近目标时能够遇到的所有新位置情况当作有限集合A，机器人测试运行一次就会从中学习一定量的新的知识，集合A的容量就会减少一部分，直到第三次运行时，集合A基本上接近空集，此时，机器人在接近目标的过程中遇到的位置情况都在自己的大脑(发育网络)中储存着，识别出的位置情况与实际遇到的位置情况基本一致，因此识别精度很高，误差很小。

[0123] 5.2动态环境试验

[0124] 与静态环境下的实验过程类似，将上述非工作状态下机器人的迁移学习应用于动态环境下的场景，可以获得类似的结果，如图11至图15所示。

[0125] 动态环境下，机器人获取的新知识量会随着运行时间的增加而增加，这是由于机器人运行的环境一直在变化，每次运行，机器人都会遇到新的环境位置情况，学习到新的知识，相应的，其存储知识的神经元个数会随之增加，如图 12所示。

[0126] 以上对本发明所提供的机器人运动方向预先决策的原理和实施方式进行了详细介绍。本文中应用了具体个例对本发明的原理及实施方式进行了阐述，上述实例的目的旨在对帮助理解本发明的具体原理和实施方式。应当声明的是，对于本专业领域的普通技术人员，在使用本发明解决问题和科学研究时，在不改变本发明原理和核心思想的前提下，允许做出若干技术改进，这些改进的技术也落入本发明专利权利要求的保护范围内。

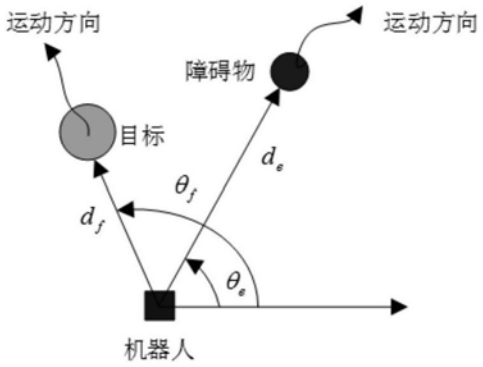

运动方向
运动方向
障碍物
目标
d_e
θ_f
d_f
θ_e
机器人

图1

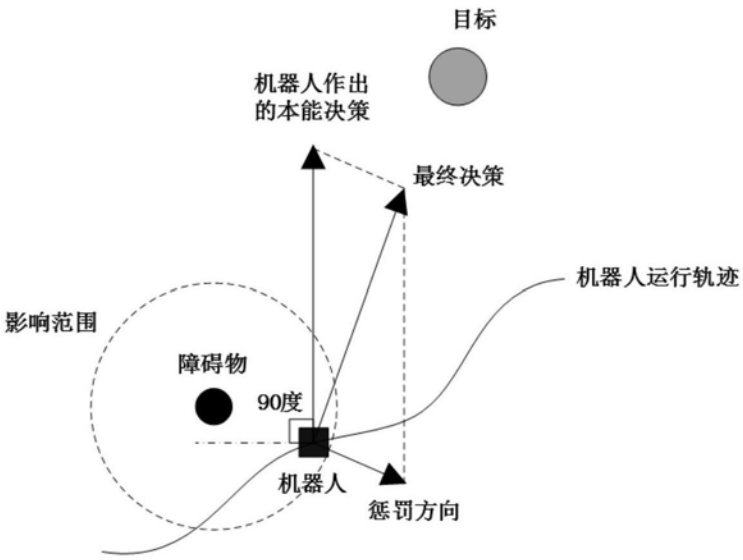

目标
机器人作出
的本能决策
最终决策
机器人运行轨迹
影响范围
障碍物
90度
机器人
惩罚方向

图2

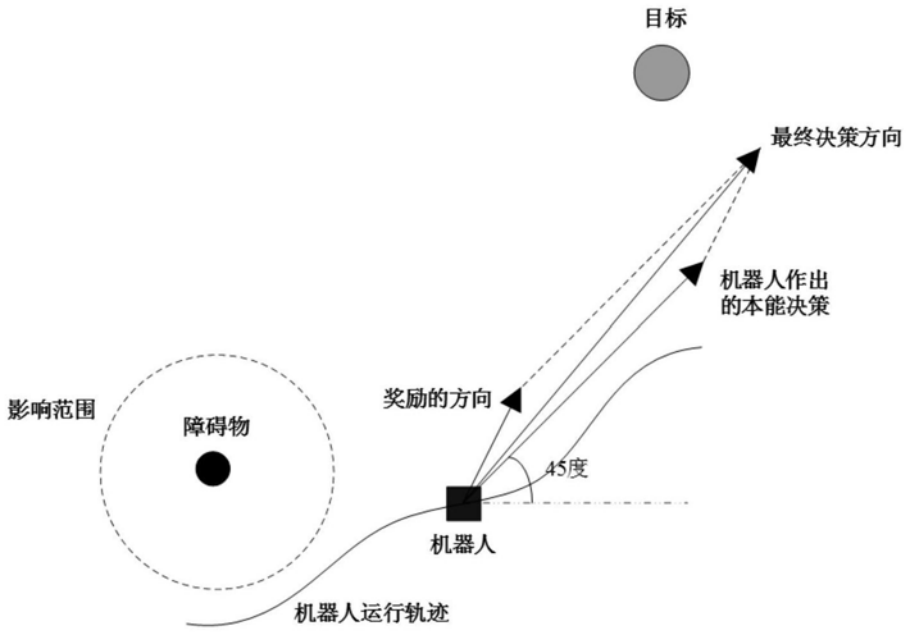

目标
最终决策方向
机器人作出
的本能决策
奖励的方向
影响范围
障碍物
45度
机器人
机器人运行轨迹

图3

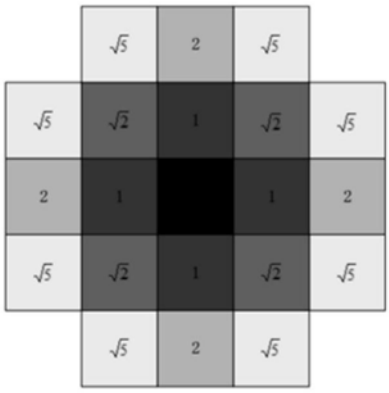

√5
2
√5
√5
√2
1
√2
√5
2
1
1
2
√5
√2
1
√2
√5
√5
2
√5

图4

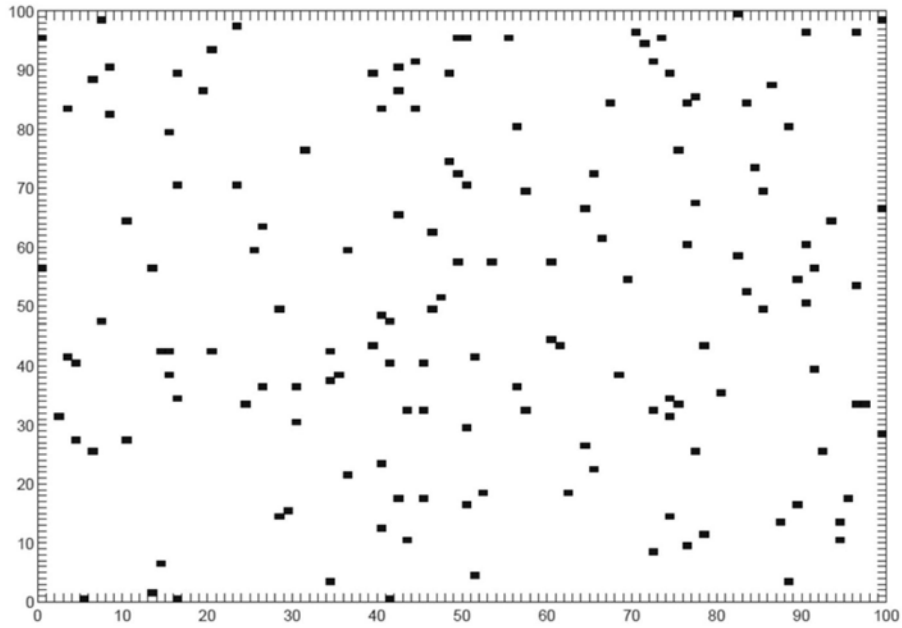

0
10
20
30
40
50
60
70
80
90
100

图5

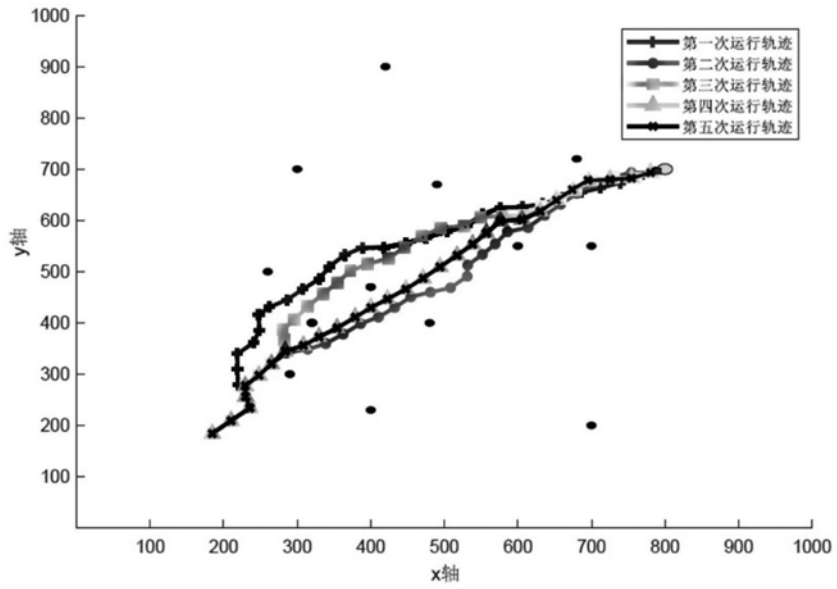

第一次运行轨迹
第二次运行轨迹
第三次运行轨迹
第四次运行轨迹
第五次运行轨迹
y轴
x轴
100
200
300
400
500
600
700
800
900
1000

图6

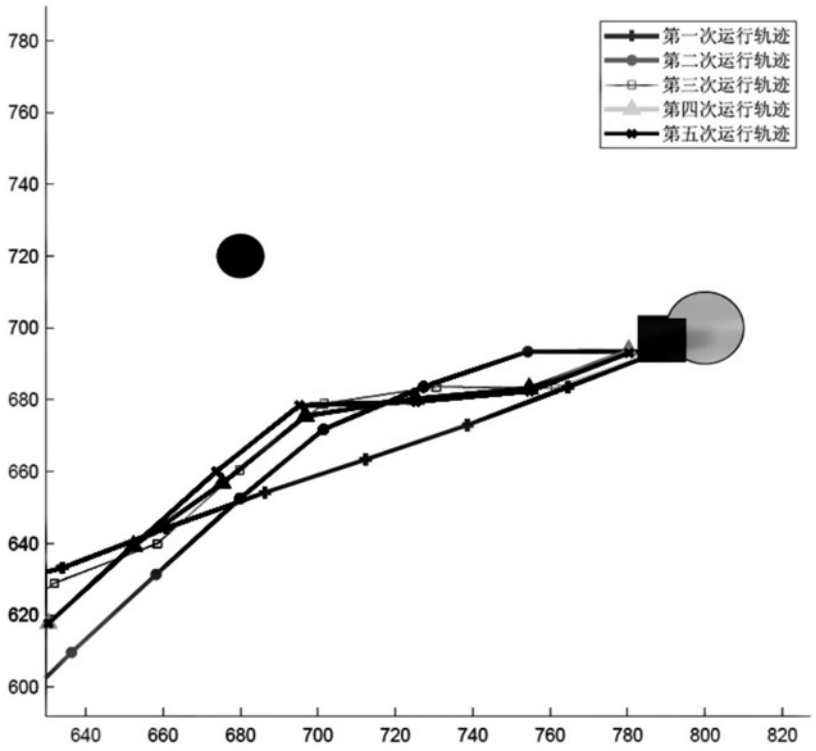

第一次运行轨迹
第二次运行轨迹
第三次运行轨迹
第四次运行轨迹
第五次运行轨迹
780
760
740
720
700
680
660
640
620
600
640
660
680
700
720
740
760
780
800
820

图7

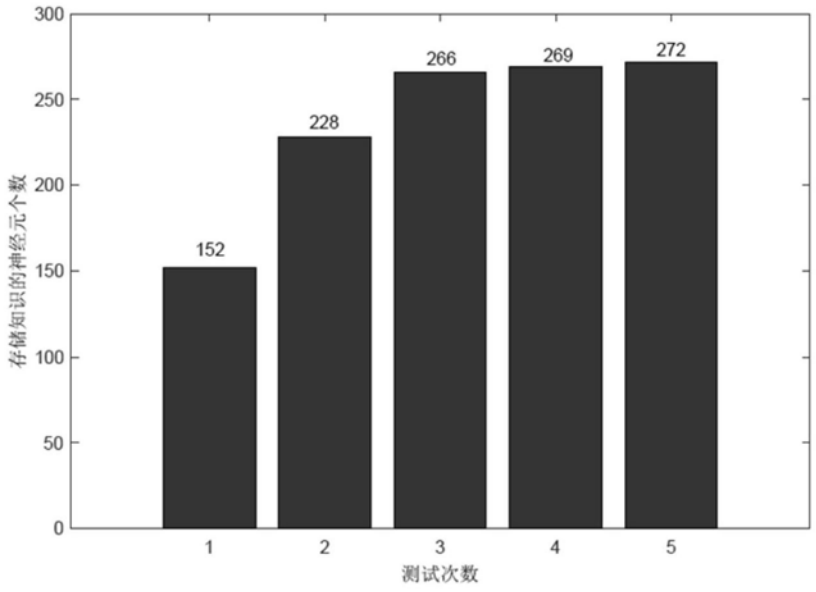

300
250
200
150
100
50
0
存储知识的神经元个数
152
228
266
269
272
1
2
3
4
5
测试次数

图8

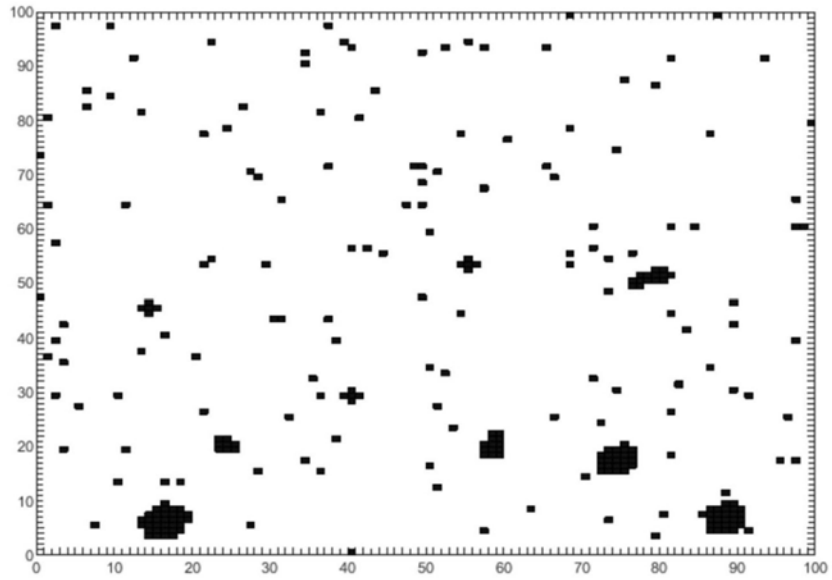

0
10
20
30
40
50
60
70
80
90
100

图9

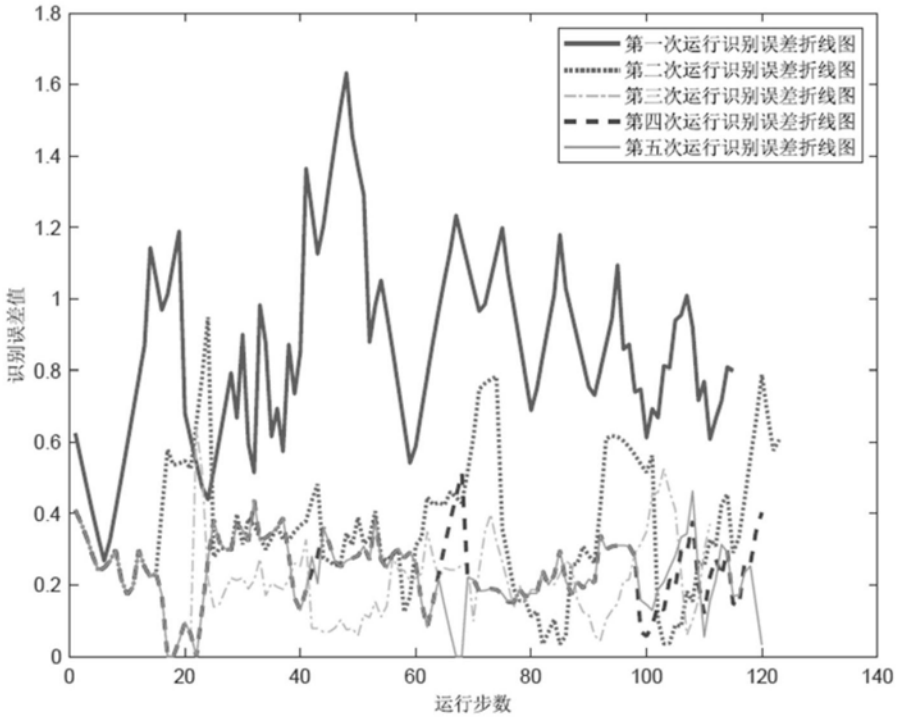

第一次运行识别误差折线图
第二次运行识别误差折线图
第三次运行识别误差折线图
第四次运行识别误差折线图
第五次运行识别误差折线图
识别误差值
运行步数
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
0
20
40
60
80
100
120
140

图10

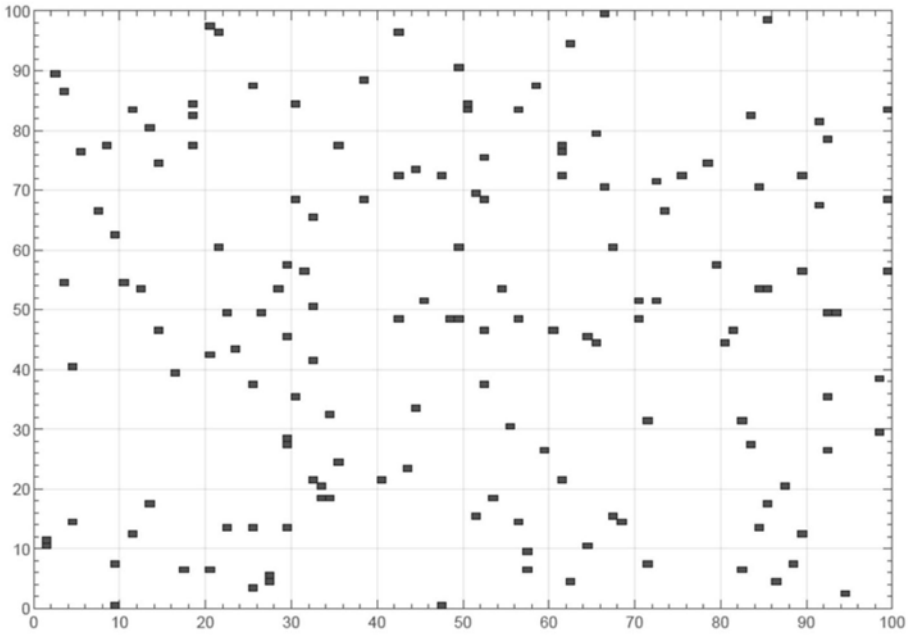

0
10
20
30
40
50
60
70
80
90
100

图11

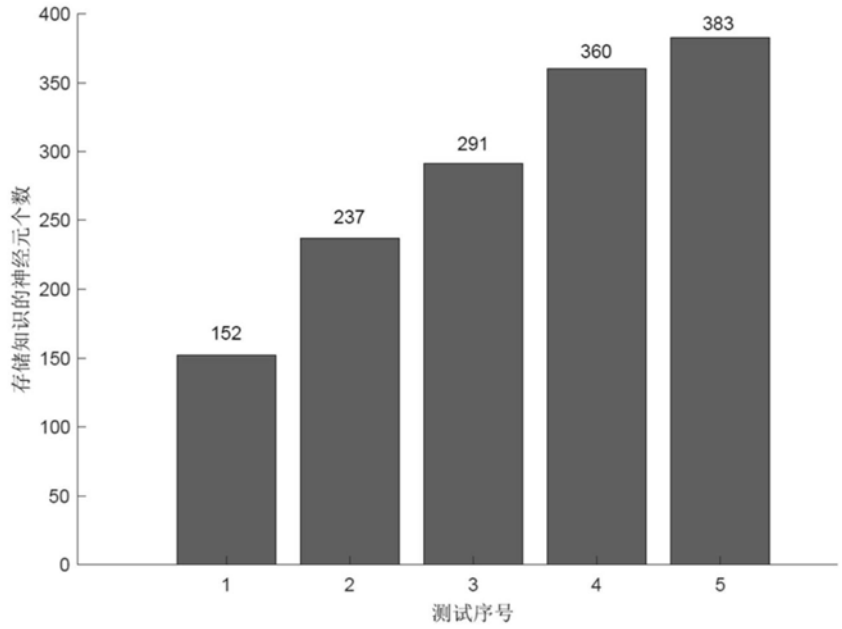

存储知识的神经元个数
测试序号
152
237
291
360
383
0
50
100
150
200
250
300
350
400
1
2
3
4
5

图12

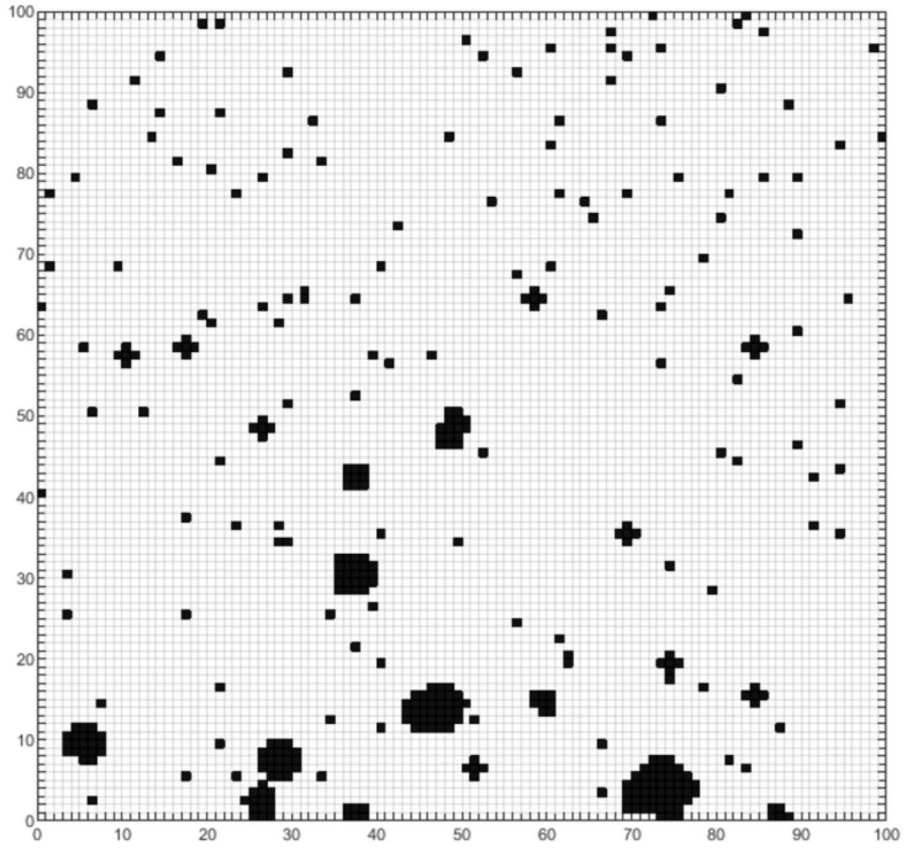

0
10
20
30
40
50
60
70
80
90
100

图13

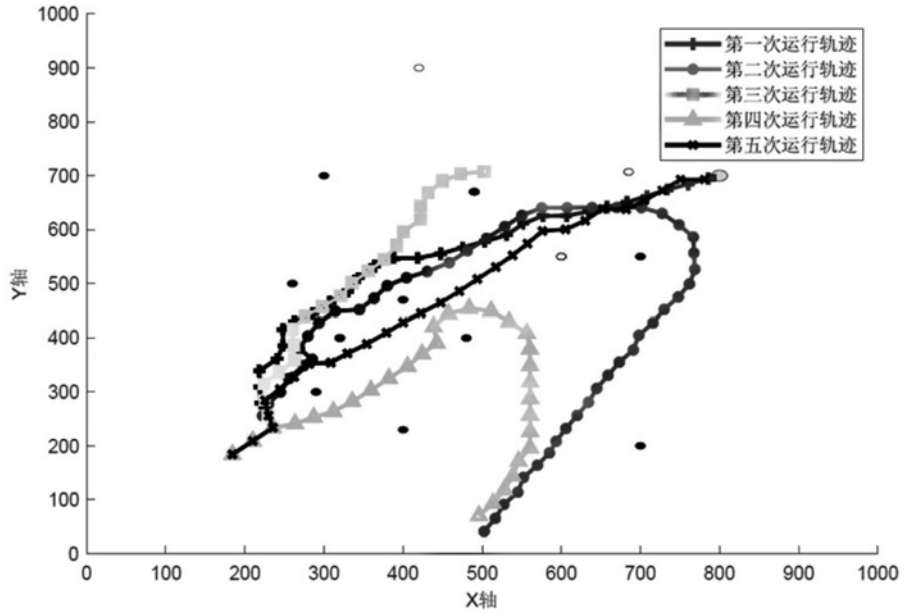

第一次运行轨迹
第二次运行轨迹
第三次运行轨迹
第四次运行轨迹
第五次运行轨迹
Y轴
X轴
0
100
200
300
400
500
600
700
800
900
1000

图14

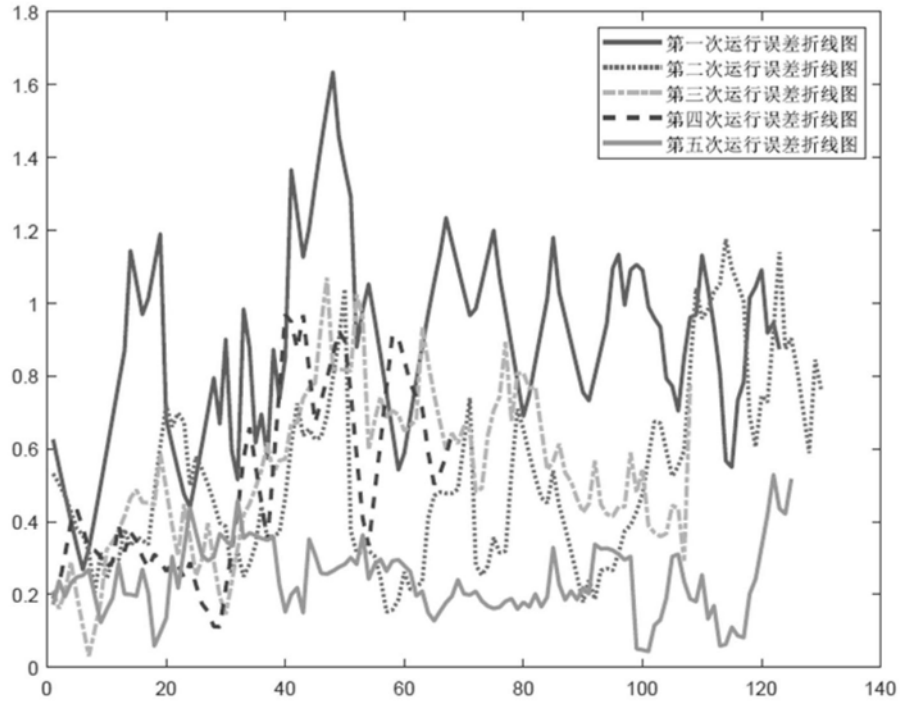

第一次运行误差折线图
第二次运行误差折线图
第三次运行误差折线图
第四次运行误差折线图
第五次运行误差折线图
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
20
40
60
80
100
120
140

图15