基于改进 Q 学习算法的导航认知图构建

赵辰豪，吴德伟，何　晶，韩　昆，来　磊
（空军工程大学信息与导航学院，西安，710077）

摘要　针对导航认知图构建效率低，方向信息不准确等问题，提出了一种基于改进 Q 学习算法的导航认知图构建方法。首先，利用径向基(RBF)神经网络学习生成网格细胞到位置细胞的映射关系，并利用位置细胞对空间进行表征；其次，采用改进 Q 学习算法学习位置细胞面向目标的 Q 值大小；最后，根据重心估计原理计算面向目标的方向信息，并生成导航认知图。仿真结果表明：与传统 Q 学习算法相比，文中算法生成导航认知图的学习次数从 2 000 次缩减至 1 000 次，提高了导航认知图的构建效率；学习值(指面向目标的方向信息)的相对误差最大降低了 15%，提高了认知图的准确性。

关键词　类脑导航；网格细胞；位置细胞；改进 Q 学习算法；导航认知图
DOI　10.3969/j.issn.1009-3516.2020.02.008
中图分类号　TP183　　**文献标志码**　A　　**文章编号**　1009-3516(2020)02-0053-08

Navigation Cognitive Map Construction Based on Improved Q-Learning Algorithm

ZHAO Chenhao, WU Dewei, HE Jing, HAN Kun, LAI Lei
(Information and Navigation College, Air Force Engineering University, Xi'an 710077, China)

Abstract　In order to improve the efficiency in navigation cognitive mapping and reduce the error of direction information, a method of navigation cognitive mapping based on improved Q-learning algorithm is proposed in this paper. Firstly, Radial Basis Function (RBF) neural network is utilized for trainning the mapping relationship between grid cells and place cells, and the place cells are used to convey space information. Secondly, the improved Q-learning algorithm is used to learn the target-oriented Q value of the place cell to obtain direction information towards target. Finally, the center of gravity estimation principle is used to generate the target-oriented direction information, constructing the navigation cognition map. The simulation results show that learning rounds of the navigation cognitive mapping generated by this algorithm is reduced from 2 000 to 1 000 compared with the traditional Q-learning algorithm, improving the construction efficiency of navigation cognitive map. Meanwhile, a maximum reduction of 15% in relative error is achieved, improving the precision of navigation cognitive mapping.

Key words　brain-like navigation; grid cell; place cell; improved Q-learning algorithm; navigation cognition map

收稿日期：2019-10-15
基金项目：国家自然科学基金(61603409)
作者简介：赵辰豪(1996—)，男，山西侯马人，硕士生，主要从事智能导航和认知图构建研究。E-mail：14291024@bjtu.edu.cn
通信作者：吴德伟(1963—)，男，陕西西安人，教授，博士生导师，主要从事量子传感、智能导航等研究。E-mail：Wudewei74609@126.com

引用格式：*赵辰豪，吴德伟，何晶，等. 基于改进 Q 学习算法的导航认知图构建[J]. 空军工程大学学报(自然科学版)，2020，21(2)：53-60.*
ZHAO Chenhao, WU Dewei, HE Jing, et al. Navigation Cognitive Map Construction Based on Improved Q-Learning Algorithm[J]. Journal of Air Force Engineering University (Natural Science Edition), 2020, 21(2): 53-60.

实现运行体的自主“无人化”运行是研究人员不断追求的目标,也是实现无人作战系统的关键。运行体的“无人化”离不开自主导航技术,实现自主导航的方法有很多,目前主要分为 2 类:基于人工系统模型的方法与基于自然系统模型的方法。基于人工模型的方法在较为简单的环境中,能够一定程度上实现人工智能;当环境变得复杂,人工模型将变得十分复杂,不能够快速处理信息。而基于自然模型的方法主要通过模拟自然界中存在的智能系统模型。目前,经过大量试验证明动物大脑在认知导航中起到不可替代的作用,其自身具有最完美、最有效的信息处理机制,这使得导航领域的研究者对大脑在导航方面的应用产生了浓厚的兴趣[1]。近些年,经过大量验证与动物导航相关的细胞有:位置细胞、网格细胞、头朝向细胞、边界细胞等。导航工作人员结合对大脑的研究成果提出了脑神经科学启发下的自主导航方法,该方法通过模拟大脑处理导航信息的机制,使得智能体呈现出一种具有探索、记忆、学习以及选择等智能动作的导航行为。

实现脑神经科学启发下的导航需要解决 3 个问题:①实现模拟大脑的空间探索与表征;②采用类脑机制构建认知图;③在认知图上进行路径规划。目前模拟大脑进行空间探索与表征的研究已经相对成熟;文献[2]构建前馈网络,建立网格细胞到位置细胞的联系,并使用傅里叶分析网格细胞到位置细胞的权值,最终使用位置细胞对空间进行表征;文献[3~4]采用 Hebbian 学习方法建立位置细胞与网格细胞的权值,从而使位置细胞具有空间的放电野。

“认知图”一词,最早由 Tolman[5]在研究大鼠如何探索路径的试验中提出。类脑机制下导航认知图构建的研究国外已有一定理论基础。文献[6]将生物大脑放电过程与传统 SLAM 结合成功实现水下智能体的定位与地图创建。文献[7]提出了基于多层目标的导向的导航模型,将导航模型分为两步:一是构建空间表征(构建认知图);二是基于表征信息进行导航。而国内的研究相对较少,吴德伟团队[8]提出了一种多尺度网格细胞的路径整合,完成了运行体自主位置推算。于乃功团队[9]提出位置细胞到网格细胞的竞争型神经网络模型,并生成位置细胞对空间进行表征;唐华锦团队[10]提出采用类脑神经机制进行定位,认知图构建以及情景记忆,并对该过程进行误差校正;吴德伟团队[11]提出视觉位置细胞模型并利用其放电机理对空间进行表征。目前,类脑机制的导航认知图构建的研究较少,国内还没有团队提出构建认知图的系统方法,而已有的研究构建的认知图存在以下问题:构建认知图的效率较低,认知图中面向目标的方向误差较大,图中的信息准确度不高。

针对上述问题本文提出了一种基于改进 Q 学习算法的导航认知图构建方法(本文中认知图特指动物出发觅食的过程,通过整合地理位置信息与面向目标的方向信息产生位置细胞,并对空间进行表征,最终生成基于目标导向的向量地图[12])。该方法明确了类脑机制下认知图构建过程,同时提高了构建认知图的效率以及认知图内面向目标信息的精确度。同时,本文优化了网格细胞到位置细胞的 RBF 映射网络,提高了位置细胞对空间的表征能力。

1 模型

1.1 基于网格细胞与位置细胞的位置表征

对于空间表征主要解决的核心问题是如何实现网格细胞到位置细胞的映射,目前实现网格细胞到位置细胞的转换模型主要有:基于竞争学习的转换模型[13],基于傅里叶分析的转换模型[14]与基于 ICA 编码的转换模型[15]。本文提出采用 RBF 神经网络,建立网格细胞与位置细胞的映射关系。通过运行体感知自身运动信息,输入到网格细胞模型中得到网格细胞的放电率,再将网格细胞放电率作为 RBF 网络的输入,对 RBF 网络进行训练,建立网格细胞到位置细胞的转换模型,最终使用位置细胞表征空间。

针对建立空间的位置细胞表征,给出具体方法,包括计算网格细胞的放电率,建立 RBF 神经网络映射,建立位置细胞的空间表征 3 个步骤。

步骤 1 计算网格细胞的放电率。

目前模拟网格细胞放电活动的模型主要有:吸引子网络模型[16-17]和振荡干涉模型[18-19]。本文采用振荡干涉模型对网格细胞进行模拟,其放电率的计算公式为:

$$R_{GC}^{i}(\boldsymbol{r})=\frac{2}{3}\left\{\frac{1}{3}\sum_{d=1}^{3}\cos\left[\frac{4\pi}{\sqrt{3}A_i}\boldsymbol{k}_{\mathrm{id}}(\boldsymbol{r}-\boldsymbol{\varphi}_i)\right]+\frac{1}{2}\right\},\quad i=1,2,\cdots,N_{AC} \tag{1}$$

式中:$\boldsymbol{r}=[x,y]^{\mathrm{T}}$ 为运行体的位置坐标;$R_{GC}^{i}(\boldsymbol{r})$为网格细胞的放电率;$N_{AC}$ 为网格细胞数;$\boldsymbol{\varphi}_i=[x_i,y_i]^{\mathrm{T}}$ 为网格细胞的相位;A_i 为网格细胞的网格间距;$\boldsymbol{k}_{\mathrm{id}}$ 设置为:

$$\boldsymbol{k}_{\mathrm{id}}=\left[\cos\left(\omega_i+\frac{2(d-1)}{3}\pi\right),\sin\left(\omega_i+\frac{2(d-1)}{3}\pi\right)\right],i=1,2,\cdots,N_{GC},d=1,2,3 \tag{2}$$

式中：ω_i 为网格细胞的网格方向。

步骤 2　建立 RBF 网络映射。

RBF 神经网络分为输入层，隐含层与输出层。其中输入层到隐含层的变换是非线性的，隐含层到输出层的变换为线性的。RBF 网络的输出可以表示为：

$$R_{\mathrm{RBF}}(X)=\sum_{N}\omega_i\exp\left(-\frac{\|X-U_i\|^2}{\sigma_i^2}\right) \tag{3}$$

式中：N 为隐藏节点的个数；ω_i 为第 i 个隐藏节点与输出节点之间的连接权值；U_i 和 σ_i^2 为隐藏节点的中心和方差。

由**步骤 1** 可得，运行体位于 $\boldsymbol{r}=[x,y]$ 时，网格细胞的放电率表示为：

$$R_{GC}^{i}(\boldsymbol{r})=[r_1,r_2,\cdots,r_i],i=1,2,\cdots,N_{AC} \tag{4}$$

将网格细胞的放电率作为 RBF 网络的输入，位置细胞的理论放电率作为输出，根据输入输出对 RBF 网络进行学习，最终确定 RBF 网络的参数，得到当前位置处网格细胞到位置细胞的映射关系。

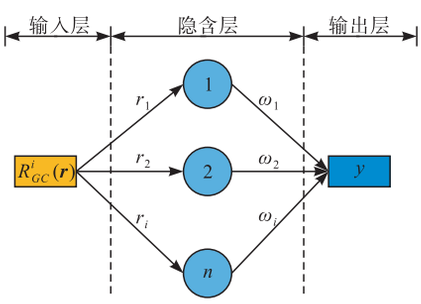


图 1　RBF-神经网络结构图

步骤 3　建立位置细胞的空间表征。

为了能够生成新的位置细胞，设置位置细胞的放电阈值；当已有位置细胞的放电率小于该阈值时，则执行生成新的位置细胞。建立位置细胞的表征过程即为生成 RBF 神经网络群的过程。当运行体探索某一位置时，把该位置处网格细胞的放电率输入 RBF 神经网络群，得到一组输出值，将输出值与放电阈值比较，当输出都小于放电阈值时，重复**步骤 2** 建立新的 RBF 神经网络。

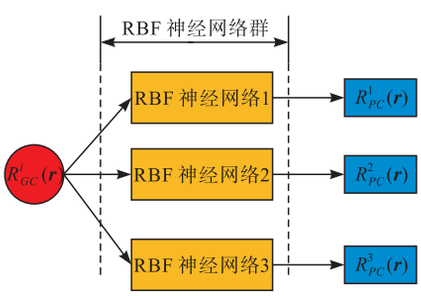


图 2　RBF 神经网络群示意图

通过上述 3 个步骤，运行体感知自身运动信息后，通过网格细胞到位置细胞的模型，得到位置细胞的放电情况，根据位置细胞放电情况确定是否生成新的 RBF 网络与位置细胞，最终得到位置细胞的空间表征。

1.2　Q 学习算法及改进

Q 学习算法[20]是一种经典的强化学习方法。强化学习不同于传统的监督学习机制，主要采用试错的方法寻找最优的行为策略。对于环境模型，Q 学习算法不会进行预判与估计，而是直接采用函数迭代的方法，通过对策略的选择，获得相应的奖惩值，对函数值进行更新，在学习过程中逼近最优的动作序列，最终达到全局最优。这个过程中使用的函数即为 Q 函数。Q 函数的迭代过程为：

$$Q(s,a)=Q(s,a)+\beta[R(s,a)+\gamma\max_{a'}Q(s',a')-Q(s,a)] \tag{5}$$

式中：$Q(s,a)$ 表示状态 s 时执行动作 a 的 Q 值；β 为学习率，$\beta\in[0,1]$；γ 为折扣率，$\gamma\in[0,1]$；$R(s,a)$ 为状态 s 时执行动作 a 的奖励值。

Q 学习算法的过程可以概括为：

步骤 1　初始化 Q 表；

步骤 2　得到 t 时刻的状态 s，采用 ε-贪婪策略选择当前状态下的动作即以概率选择该状态下最大 Q 值对应的动作，以概率 1-随机选择动作；

步骤 3　通过 Q 表以及奖惩策略得到当前状态的 Q 值与回报值，代入式(5)中更新 Q 表，直到学习结束。

Q 学习算法具有学习能力强以及效果较好的特点，但其还存在一定的问题。由于 Q 表的设置状态或动作需要的都是离散空间，Q 学习算法不能运用到连续空间的学习。同时当状态空间或动作空间中的元素过多时，Q 表将变得十分庞大，对于 Q 表的查询将变得十分困难。ε-贪婪策略是强化学习中比较普遍且有效的探索方法，但其缺点是在算法后期还会选择非最优策略为最优策略，造成一些不必要的学习与资源浪费。

针对上述的问题，本文对传统 Q 学习进行改进，引入 Boltzmann 分布，提出 Boltzmann-选择策略：

$$P(s,a)=\frac{\exp\left(\frac{Q(s,a)}{T}\right)}{\sum_{a}\exp\left(\frac{Q(s,a)}{T}\right)} \tag{6}$$

式中：$P(s,a)$ 表示在状态 s 时选择动作 a 的概率；$Q(s,a)$ 表示在状态 s 时选择动作 a 的 Q 值。

传统 Q 学习算法中使用贪婪策略，在整个学习过程中随着学习的深入选择动作的概率基本保持在不变。这一问题导致在学习初期存在次优动作的 Q 值大于最优动作的 Q 值的现象，使得 Q 值不能够准

确的表达动作的优劣性;当学习后期由于较大的概率选择非最优动作,因此导致学习值不准确,学习结果收敛较慢,学习效果不好。当概率值选择不恰当时,会使得整个探索过程不合理,造成资源浪费。

改进 Q 学习算法中使用的是 Boltzmann 选择策略。根据式(6)可得,当初始状态的 Q 值都为 0 时,运行体对各个动作选择的概率相等,从而使得前期的动作选择更加随机,探索更加合理,对 Q 值的更新更准确;当学习进行到一定阶段时,由于各个动作的回报值不同,因此 Q 值更新情况也不同;动作的回报值越大,Q 值越大;动作回报值越小的,Q 值越小。因此根据式(6)得,Q 值越大,被选择的概率就越大。因此随着学习的深入,运行体倾向选择回报值高的动作。

改进 Q 学习算法对动作的选择概率是变化的,学习初期能够随机探索各个动作,随着学习过程的深入,回报值越大的动作被选中的概率越大。这种动态概率既能够提高学习的合理性与有效性,又能够提升算法的收敛速度,使学习更效率、更精确。

1.3 基于改进 Q 学习算法的认知图构建

本文中构建的认知图的过程分为 3 个步骤:

步骤 1 创建状态空间与动作空间。

状态空间:运行体在空间探索过程中,生成一组位置细胞,将生成的位置细胞群设定为状态空间,其中每一个位置细胞表示一个状态。当运行体到达空间中的某一位置时,通过空间表征模型确定当前状态。

动作空间:运行体在探索空间的过程,能够向各个方向运动。因此设定动作空间时需要满足运行体运动的全方位性。将方向区间[0,2π]离散化,离散后的每个值表示相应的动作方向,组成动作空间。以 X 轴正半轴为基准,各个动作方向分别表示为:

$$\varphi_{AC}^{i}=(i-1)\frac{2\pi}{N_{AC}},i=1,2,\cdots,N_{AC} \tag{7}$$

式中:φ_{AC}^{i} 表示动作方向;N_{AC} 表示动作空间精度,N_{AC} 越大,精度越高。通过设置不同的精度值,可以得到不同分辨率的动作空间。

步骤 2 采用改进 Q 学习算法建立状态-动作关系。

结合改进 Q 学习算法运行体对学习状态-动作过程如下:

首先,根据状态空间与位置空间建立 Q 表,并对 Q 表初始化,Q 表中主要记录状态与动作的匹配程度;

其次,当运行体从起点出发后,根据位置细胞表征模型确定当前状态为 s_t,根据改进后 Boltzmann-选择策略选择当前状态下的动作方向 φ。由当前状态与动作得到下一状态,此时根据 Q 学习算法的原理对 Q 表进行更新如下:

$$Q(s_t,\varphi)=Q(s_t,\varphi)+\alpha[rewards+\gamma\max Q(s_{t+1},:)-Q(s_t,\varphi)] \tag{8}$$

式中:$Q(s_t,\varphi)$ 表示当处于状态 s_t 时,选择动作 φ 对应的 Q 值;$\max Q(s_{t+1},:)$ 为状态 s_{t+1} 时各个动作中的最大 Q 值;α 为学习率;γ 为折扣率;$rewards$ 为奖励值,奖励值的多少取决与下一状态,其设置为:

1)当下一个状态为目标所在状态时,$rewards>0$;

2)当下一个状态超出探索区域或者遇到障碍物时,$rewards<0$;

3)其他情况,$rewards=0$。

最后,当运行体运动状态数超过设置值或运行体发现目标时,该轮学习自动结束。启动下一轮学习,直到学习结束。

经过学习后,运行体能够获得最终的 Q 表。该表中 $Q(s_t,\varphi)$ 表示状态 s_t 与动作 φ 的匹配关系,该值越大代表状态 s_t 与动作 φ 越匹配。

步骤 3 构建认知图。

在构建认知图时,将**步骤 2** 中最终得到的 Q 表作为各个动作的权重,然后结合所对应的动作方向,直接使用重心估计原理计算各个状态相对于目标的方向信息,从而生成面向目标的认知图。方向的计算方法如下:

$$\theta(s)=\frac{\sum_{i=1}^{N_{AC}}R_{AC}^{i}(s)\varphi_{AC}^{i}}{\sum_{i=1}^{N_{AC}}R_{AC}^{i}(s)} \tag{9}$$

式中:$R_{AC}^{i}(s)$ 表示状态动作的权重,即 $R_{AC}^{i}(s)=Q(s,\varphi_{AC}^{i})$;$\varphi_{AC}^{i}$ 为动作方向;计算结果 $\theta(s)$ 为该位置细胞面向目标的方向角。

2 仿真分析及讨论

仿真环境设置如下:

1)空间大小为 50 m×50 m,运行体采用离散方式,以 5 m/s 的速度对空间进行探索;

2)在空间探索过程中,将目标探索生成的位置细胞设置为状态空间中的元素,在学习过程中,通过比较状态确定是否达到目标;

3)奖励策略设置为:当运行体运动位置超出探索空间时,奖励值为−1;当运行体发现目标时,奖励值为 10;其余情况,奖励值为 0;

4)网格细胞的参数设置:根据式(1)建立网格细

胞，网格细胞总数为 50，网格间距 A 为 4，网格方向为 30°；

5）位置细胞的放电野设置：当距离位置细胞中心距离在 0～5 m 之间时，位置细胞的放电率大于 0.1，当距离位置细胞中心距离大于 5 m 时，位置细胞的放电率小于 0.1（位置细胞不放电），从而位置细胞能够感应距离放电中心约 5 m 的范围；

6）探索过程中位置细胞的放电阈值设置为0.1，当已有位置细胞的放电率都小于 0.1 时，生成新的位置细胞；

7）Q 学习的相关参数设置：学习率为 0.1，折扣因子为 0.1，认知图精度为 8。

经过仿真实验得到以下结果：图 3 给出了不同间距网格细胞的放电情况（网格间距依次为：5，10，15）。仿真结果表明：改变网格间距，固定运行体位置能够得到不同的网格细胞放电情况。图 4 给出了在不同位置处网格细胞的放电情况（位置坐标依次为(5，15)，(15，25)，(25，35)）。仿真结果表明：当固定网格细胞间距，改变运行体的位置，网格细胞的放电情况将发生改变。图 3，图 4 证明在不同位置处网格细胞的放电情况不同，网格细胞的放电率能够作为 RBF 网络的有效输入。

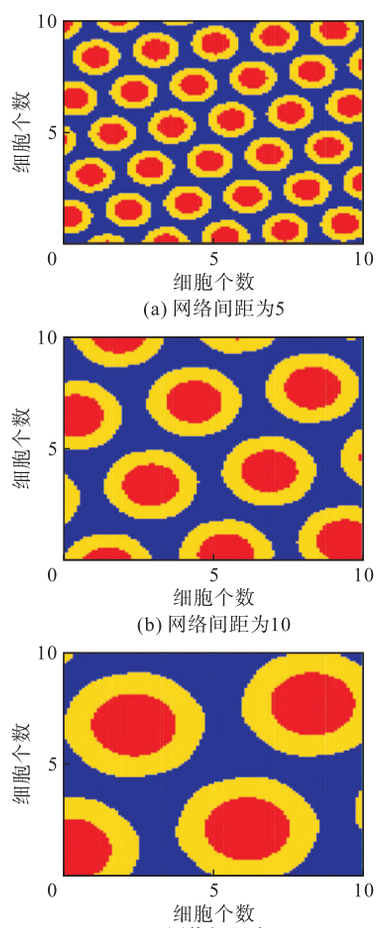


图 3　不同间距的网格细胞放电情况

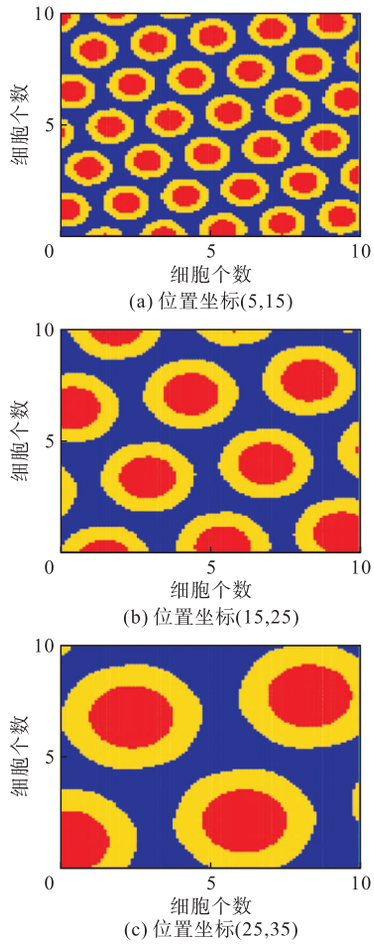


图 4　不同位置处网格细胞的放电情况

图 5 给出了 RBF 网络训练得到的位置细胞放电率与空间位置的函数关系。图中的函数关系表明，当运行体与位置细胞的中心点距离小于等于 5 m 时，位置细胞放电（当且仅当距离为 0 时，位置细胞的放电率达到峰值）；当距离大于 5 m 之后，位置细胞不放电。

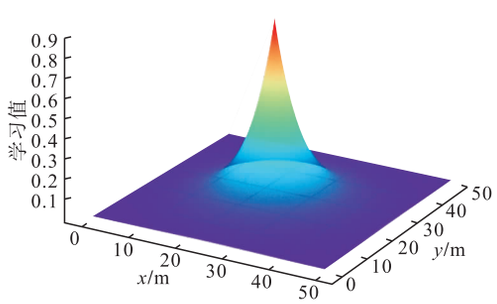


图 5　训练后 RBF 神经网络的函数图像

图 6 表示中心点为(35，40)的位置细胞的放电域。分析图像可得距离位置细胞的中心点距离在 5m 之内，位置细胞放电。该放电情况与图 5 的训练结果相符。

图 7 给出了运行体经过探索后生成的部分位置细胞的放电情况。位置细胞的放电野覆盖了整个区域，当运行体在空间运行时相应的位置细胞将正常

放电,因此能够通过观察位置细胞的放电率得到运行体的位置。位置细胞仿真结果表明,改进后的 RBF 网络能够训练得到位置细胞的放电函数,生成的位置细胞能够覆盖探索区域,并成功对空间进行表征。

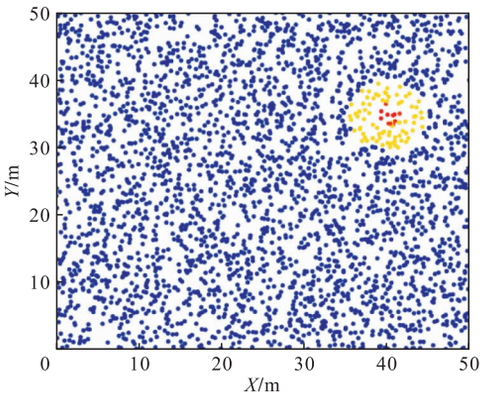


图 6 RBF 神经网络映射后位置细胞的放电域

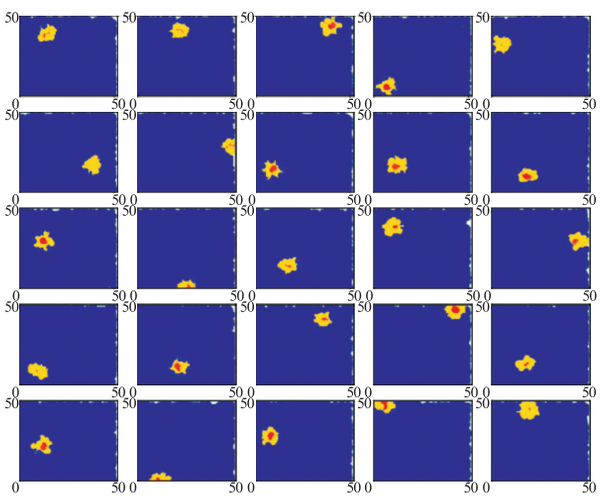
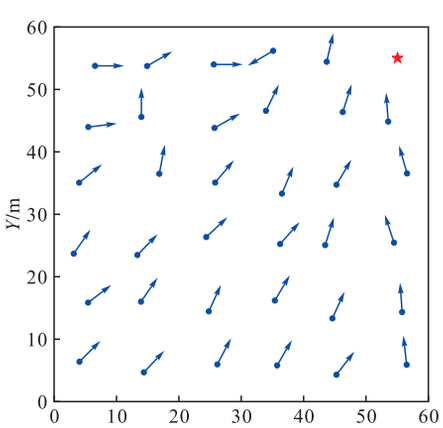
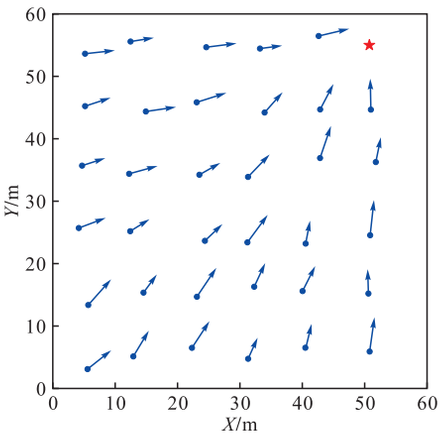


图 7 探索结束后生成位置细胞的放电情况

图 8 给出了采用传统 Q 学习算法(单个回合学习 5 000 次)构建认知图的情况与采用改进 Q 学习算法(单个回合学习 1 000 次)构建认知图的情况。仿真结果表明:改进后 Q 学习方法的效率更高,各个状态的角度值更加接近真实值,构建的认知图中的方向信息更加准确。



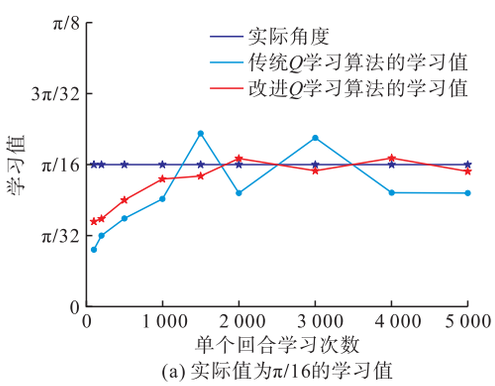
(a) 传统 Q 学习算法



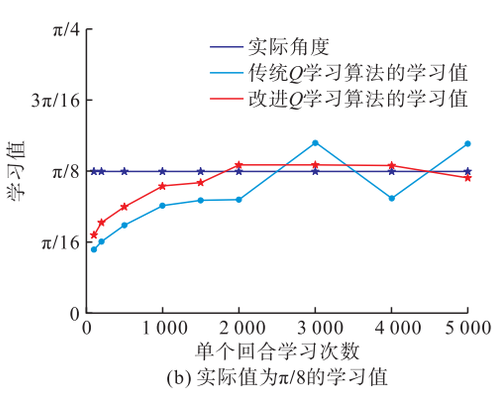
(b) 改进 Q 学习算法

图 8 不同算法构建的认知图

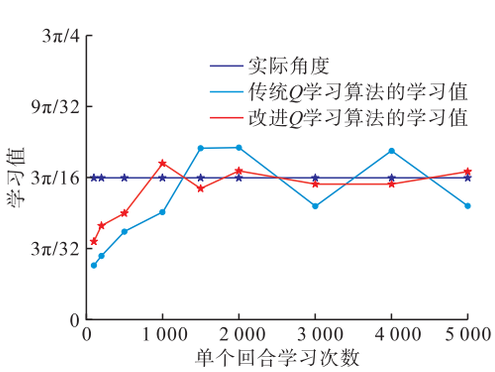
图 9 给出传统 Q 学习算法与改进 Q 学习算法在学习过程中部分状态的学习值。图 9 表明改进 Q 学习算法比传统 Q 学习算法的实际角度值收敛更快速;且在学习次数大于 2 000 之后,改进 Q 学习算法的学习值更加接近实际值。



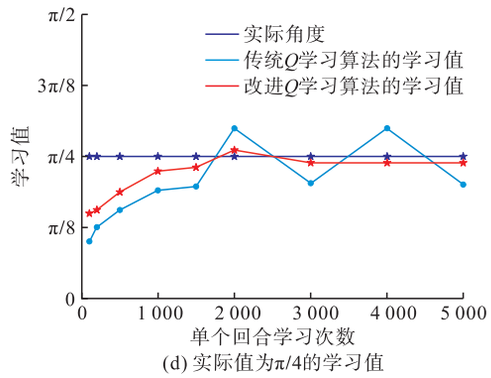
(a) 实际值为π/16的学习值



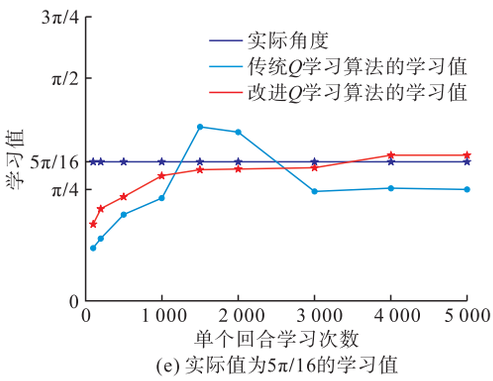
(b) 实际值为π/8的学习值



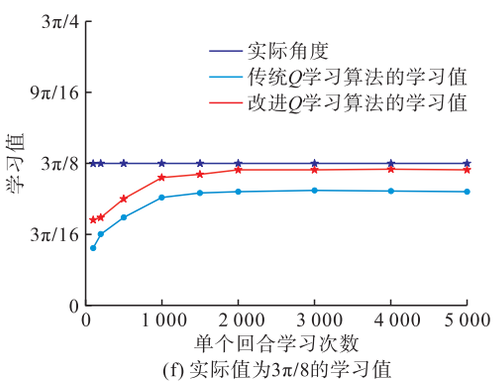
(c) 实际值为3π/16的学习值



(d) 实际值为π/4的学习值



(e) 实际值为5π/16的学习值



(f) 实际值为3π/8的学习值

图 9　不同算法的学习值

图 10 给出了单个回合中学习次数与平均相对误差的关系(平均相对误差计算方法如下：以单个回合学习 1 000 次为例，运行体进行 10 个回合的学习，计算各个状态平均相对误差后，对各个状态的平均误差求和取平均，最后得到探索 1 000 次的平均相对误差。)改进 Q 学习算法学习的相对误差一直小于传统 Q 学习算法的相对误差，且当学习次数大于 2 000 次以后，传统 Q 学习算法的相对误差稳定在 20%，改进 Q 学习算法的相对误差基本稳定在 4%。仿真结果表明：引入 Boltzmann 分布对贪婪策略进行改进能够提高 Q 学习的收敛速度，提升学习值的精确性。

图 11 和图 12 给出了传统 Q 学习算法与改进 Q 学习算法每回合的平均相对误差，仿真实验中一共学习了 100 个回合(图 11 单个回合学习次数为 1 000 次，图 12 单个回合学习次数为 2 000 次)。图 11，图 12 表明改进 Q 学习算法学习值的平均相对误差普遍小于传统 Q 学习算法。

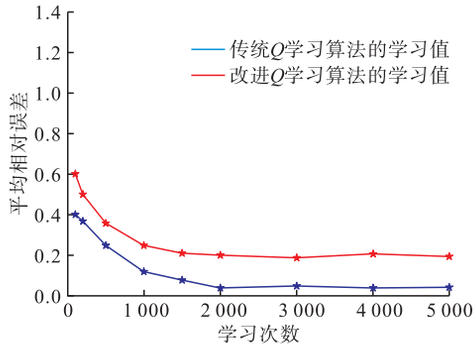


图 10　传统 Q 学习算法与改进 Q 学习算法的平均相对误差

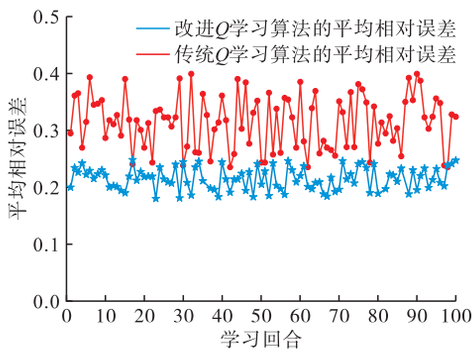


图 11　传统 Q 学习算法与改进 Q 学习算法每回合的平均相对误差(单回合次数 1 000 次)

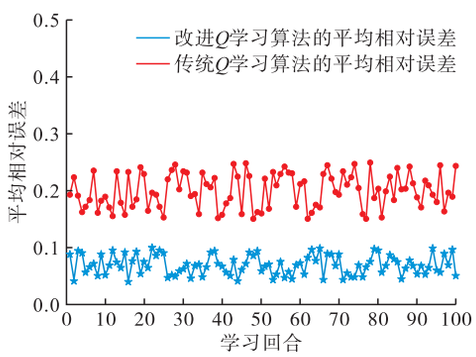


图 12　传统 Q 学习算法与改进 Q 学习算法每回合的平均相对误差(单回合次数 2 000 次)

仿真结果表明：改进后的 Q 学习算法相比于传统 Q 学习算法学习效率更高，学习值更贴近真实值，学习结果更加准确，构建的认知导航图更加准确。

3　结语

本文提出了类脑机制的导航认知图构建的系统方法，并对传统 Q 学习算法进行改进。仿真结果表明，改进后的 Q 学习算法能够提升学习效率，从而提高了导航认知图构建的效率与精度。但最终生成的导航认知图仅包含面向目标的信息，缺少该位置

处的环境信息以及状态之间的连接关系。如何整合环境信息与位置信息生成位置细胞对空间进行表征以及如何增加各个状态之间的连接关系构建认知图等问题还有待继续研究。

参考文献

[1] 于佳丽. 递归神经网络的连续吸引子与模糊控制[D]. 成都：电子科技大学，2013.

[2] SOLSTAD T，MOSER E I，EINEVOLL G T. From Grid Cells to Place Cells：A Mathematical Model[J]. Hippocampus，2006，16(12)：1026-1031.

[3] FRANZIU M，VOLLGRAF R，WISKOTT L. From Grids to Places [J]. Computational Neuroscience，2007，22(3)：297-299.

[4] SAVELLI F，KNIERIM J J. Hebbian Analysis of the Transformation of Medial Entorhinal Grid-Cell Inputs to Hippocampal Place Fields [J]. Journal of Neurophysiology，2010，103(6)：3167-3183.

[5] TOLMAN E C. Cognitive Maps in Rats and Man [J]. Psychological Review，1948，55(4)：189-208.

[6] GUTH F A，SILVEIRA L，AMARAL M，et al. Underwater Visual 3D SLAM Using a Bio-Inspired System [C]//2013 Symposium on Computing and Automation for Offshore Shipbuilding. Rio Grande，Brazil：IEEE，2013：87-92.

[7] ERDEM U M，HASSELMO M E. A Biologically Inspired Hierarchical Goal Directed Navigation Model [J]. Journal of Physiology，2014，108(1)：28-37.

[8] 郜能建，吴德伟，戚君宜，等. 多尺度网格细胞路径综合方法[J]. 北京航空航天大学学报，2013，39(6)：756-760.

[9] 于乃功，王琳，李倜，等. 网格细胞到位置细胞的竞争型神经网络模型[J]. 控制与决策，2015，30(8)：1372-1378.

[10] TANG H J，YAN R，CHEN K. Corrections to Cognitive Navigation by Neuro-Inspired Localization，Mapping and Episodic Memory.[J]. IEEE Transactions on Cognitive and Developmental Systems，2018，10(3)：751-761.

[11] 李伟龙，吴德伟，周阳，等. 基于生物位置细胞放电机理的空间位置表征方法[J]. 电子与信息学报，2016，38(8)：2040-2046.

[12] ZHU Q，WANG R B，WANG Z Y. A Cognitive Map Model Based on Spatial and Goal-Oriented Mental Exploration in Rodents [J]. Behavioural Brain Research，2013，256(11)：128-139.

[13] SOLSTAD T，MOSER E I，Einevoll G T. From Grid Cells to Place Cells：A Mathematical Model[J]. Hippocampus，2006，16(12)：1026-1031.

[14] ROLLS E，STRINGER S T，ELLIOT T. Entorhinal Cortex Grid Cells Can Map to Hippocampal Place Cells by Competitive Learning[J] Network：Computation in Neural Systems，2006，17(4)：447-465.

[15] FRANZIU M，VOLLGRAF R，WISKOTT L. From Grids to Places[J]. Journal of Computational Neuroscience，2007，22(3)：297-299.

[16] WELINDER P E，BURAK Y，FIETE I R. Grid Cells：the Position Code，Neural Networks Models of Activity，and the Problem of Learning [J]. Hippocampus，2008，18(12)：1283-1300.

[17] SI B，ROMANI S，TSODYKS M. Continuous Attractor Network Model for Conjunctive Position-by-Velocity Tuning of Grid Cells [J]. Plos Computational Biology，2014，10(4)：1-18.

[18] FUHS M C，TOURETZKY D S. A Spin Glass Model of Path Integration in Rat Medial Entorhinal Cortex [J]. The Journal of Neuroscience：the Official Journal of the Society for Neuroscience，2006，26 (16)：4266-4276.

[19] HASSELMO M E. Grid Cell Mechanisms and Function：Contributions of Entorhinal Persistent Spiking and Phase Resetting [J]. Hippocampus，2008，18(12)：1213-1229.

[20] WATKINCE C，DAYAN P. Q-Learning [J]. Machine Learning，1989，8：279-292.

(编辑：徐楠楠)