文章编号: 1000−8152(2012)12−1623−06

移动机器人路径规划强化学习的初始化

宋 勇[1,2], 李贻斌[1], 李彩虹[3]

(1. 山东大学 控制科学与工程学院, 山东 济南 250061;
2. 山东大学(威海) 机电与信息工程学院, 山东 威海 264209; 3. 山东理工大学 计算机科学与技术学院, 山东 淄博 255012)

摘要: 针对现有机器人路径规划强化学习算法收敛速度慢的问题, 提出了一种基于人工势能场的移动机器人强化学习初始化方法. 将机器人工作环境虚拟化为一个人工势能场, 利用先验知识确定场中每点的势能值, 它代表最优策略可获得的最大累积回报. 例如障碍物区域势能值为零, 目标点的势能值为全局最大. 然后定义Q初始值为当前点的立即回报加上后继点的最大折算累积回报. 改进算法通过Q值初始化, 使得学习过程收敛速度更快, 收敛过程更稳定. 最后利用机器人在栅格地图中的路径对所提出的改进算法进行验证, 结果表明该方法提高了初始阶段的学习效率, 改善了算法性能.

关键词: 移动机器人; 强化学习; 人工势能场; 路径规划; Q值初始化

中图分类号: TP242 **文献标识码**: A

Initialization in reinforcement learning for mobile robots path planning

SONG Yong[1,2], LI Yi-bin[1], LI Cai-hong[3]

(1. School of Control Science and Engineering, Shandong University, Jinan Shandong 250061, China;
2. School of Mechanical, Electrical & Information Engineering, Shandong University at Weihai, Weihai Shandong 264209, China;
3. School of Computer Science and Technology, Shandong University of Technology, Zibo Shandong 255012, China)

Abstract: To improve the convergence rate of the standard Q-learning algorithm, we propose an initialization method for the reinforcement learning of the mobile robot, based on the artificial potential field (APF) -a virtue field of the robot workspace. The potential energy of each point in the field is specified based on prior knowledge, which represents the maximum cumulative reward by following the optimal path policy. In APF, points corresponding to obstacles have null potential energy; the objective point has the global maximum potential energy in the workspace. The initial Q value is defined as the immediate reward at the current point plus the maximum cumulative reward at succeeding points by following the optimal path policy. By initializing the Q value, we find that the improved algorithm converges more rapidly and steadily than the original algorithm. The proposed algorithm is validated by the robot path in the grid workspace. Results of experiments show that the improved algorithm promotes the learning efficiency in the early stage of learning, and improves the performance.

Key words: mobile robots; reinforcement learning; artificial potential field; path planning; Q values initialization

1 引言(Introduction)

随着机器人应用领域的不断拓展, 机器人所面临的任务也越来越复杂, 尽管很多情况下研究人员可以对机器人可能执行的重复行为进行预编程, 但为实现整体的期望行为而进行行为设计变得越来越困难, 设计人员往往不可能事先对机器人的所有行为做出合理的预测[1]; 因此, 能够感知环境的自治机器人必须能够通过与环境的交互在线学习获得新的行为, 使得机器人能够根据特定的任务选择能达到目标的最优动作[2−3].

强化学习利用类似于人类思维中的试错(trial-and-error)的方法来发现最优行为策略, 目前已经在机器人行为学习方面展现出了良好的学习性能[4−5]. Q–学习算法是求解信息不完全Markov决策问题的一种强化学习方法, 根据环境状态和上一步学习获得的立即回报, 修改从状态到动作的映射策略, 以使行为从环境中获得的累积回报值最大, 从而获得最优行为策略. 标准Q–学习算法一般将Q值初始化为0或随机数, 机器人没有对环境的先验知识, 学习的初始阶段只能随机地选择动作, 因此, 在复杂环境中算法收敛速度较慢[6−7]. 为了提高算法收敛速度, 研究人员利用先验知识对Q值进行初始化, 提高初始阶段学习效率, 加快算法收敛速度. 目前, 对Q值进行初始化的方法主要包括神经网络法[8]、模糊规

收稿日期: 2011−10−17; 收修改稿日期: 2012−07−21.
基金项目: 国家自然科学基金资助项目(61075091, 61174054); 国家自然科学基金青年基金资助项目(61105100).

则法[9]、势函数法[10]等. 神经网络法利用神经网络逼近最优值函数, 将先验知识映射成为Q函数表, 使机器人在整个状态空间的子集上进行学习, 从而能够加快算法收敛速度. 模糊规则法根据初始环境信息建立模糊规则库, 然后利用模糊逻辑对Q值进行初始化. 势函数法在整个状态空间定义相应的状态势函数, 每一点势能值对应于状态空间中相应离散状态值, 然后利用状态势函数对Q值进行初始化, 学习系统的Q值可以表示为初始值加上每次迭代的改变量. 在机器人的各种行为当中, 机器人必须遵守一系列的行为准则, 机器人通过认知与交互作用涌现出相应的行为与智能, 机器人强化学习Q值初始化就是要将先验知识映射成为相应的机器人行为. 因此, 如何获得先验知识的规则化表达形式, 特别是实现领域专家的经验与常识的机器推理, 将人的认知和智能转化为机器的计算和推理的人机智能融合技术是机器人行为学习急需解决的问题.

针对上述研究现状及不足, 本文提出了一种基于人工势能场的移动机器人路径规划强化学习初始化方法. 根据已知环境信息在机器人工作空间构建人工势能场, 使得障碍物区域势能值为零, 目标点具有全局最大的势能值, 整个势能场形成单调递增的曲面, 这时人工势能场中每个状态的势能值就代表该状态可获得的最大累积回报. 然后将所有状态–动作对的$Q(s,a)$初始值定义为当前状态执行选定的动作获得的立即回报加上后继状态遵循最优策略获得的最大折算累积回报(最大累积回报乘以折算因子). 通过Q值初始化能够将先验知识融入到学习系统中, 对机器人初始阶段的学习进行优化, 从而为机器人提供一个较好的学习基础. 改进算法通过Q值初始化, 使得算法收敛速度更快, 收敛过程更稳定. 最后, 笔者利用机器人在栅格地图中的路径规划问题对所提出的算法进行了仿真实验.

2 人工势能场模型(Model of artificial potential field)

人工势能场法是由Khatib最先提出来的一种虚拟方法[11], 最初只是为了解决机械手臂的避障问题, 目前已成为应用最为广泛的机器人实时路径规划方法之一. 其基本原理是将机器人的整个工作环境虚拟化为每一状态点都具有相应势能的空间, 目标点在全局环境产生引力势场, 障碍物在局部产生斥力场, 利用叠加原理将引力场与斥力场叠加产生势场中每个状态的总场强. 在人工势场中机器人依靠斥力场进行避障, 利用引力场趋向目标, 使得机器人能够从起始位置出发, 避开障碍物到达目标点. 目前, 大多采用库伦定理创建人工势能场计算模型, 某一状态的势能大小与该状态和目标之间的距离平方成正比, 与该状态和障碍物之间的距离成反比, 计算式如下:

$$U(s)=U_{\rm a}(s)+U_{\rm r}(s), \tag{1}$$

其中: $U(s)$为状态s点的势能, $U_{\rm a}(s)$为引力场在状态s点产生的势能, $U_{\rm r}(s)$为斥力场在状态s点产生的势能, $U_{\rm a}(s)$与$U_{\rm r}(s)$分别由式(2)和式(3)进行计算.

$$U_{\rm a}(s)=\frac{1}{2}k_{\rm a}\rho_g^2(s), \tag{2}$$

其中: $k_{\rm a}$为比例因子, $\rho_g(s)$为状态s点与目标点之间的最短距离.

$$U_{\rm r}(s)=\begin{cases}\dfrac{1}{2}k_{\rm r}\left(\dfrac{1}{\rho_{\rm ob}(s)}-\dfrac{1}{\rho_0}\right)^2, & \rho(s)<\rho_0,\\ 0, & \rho(s)\geqslant\rho_0,\end{cases} \tag{3}$$

其中: $k_{\rm r}$为比例因子, $\rho_{\rm ob}(s)$为状态s点与障碍物之间的最短距离, ρ_0为障碍物影响系数.

按照上述方法构造的人工势能场使得机器人工作的环境被转换为一个矢量场, 目标点具有最低势能, 势能值为零, 障碍物区域具有最高势能. 为了缩小势能场量值的范围差, 并且使目标点具有全局最高势能, 利用式(4)对构造的矢量场进行归一化处理:

$$U'(s)=\frac{U_{\max}(s)-U(s)}{|U_{\max}(s)|}, \tag{4}$$

其中: $U'(s)$为势能场U'中状态s的势能, $U_{\max}(s)$为势能场中最高的势能, 其对应的势能值为$|U_{\max}(s)|$. 通过上式转换在机器人工作空间构造出一个新的势能场, 使得障碍物区域势能值为零, 目标点势能值为1, 并且整个势能场形成单调递增的曲面.

3 改进的Q–学习算法(Improved Q-learning algorithm)

3.1 Q值初始化(Q values initialization)

机器人在强化学习过程中, 通过传感器感知周围环境获知当前状态, 并选择当前要执行的动作, 环境响应该动作给出立即回报, 并产生新的状态. 机器人强化学习的任务就是获得一个最优策略使得机器人从当前状态出发获得最大的折算累积回报. 机器人从任意初始状态s_t出发获得的累积回报定义如下:

$$V^{\pi}(s_t)=r_t+\gamma r_{t+1}+\gamma^2 r_{t+2}+\cdots=\sum_{i=0}^{\infty}\gamma^i r_{t+i}, \tag{5}$$

其中: π为控制策略, r为获得的立即回报序列, γ为折算因子. 则机器人从状态s出发遵循最优策略所获得的最大累积回报V^*计算如下:

$$V^*=\arg\max_{\pi}V^{\pi}(s),\ \forall s. \tag{6}$$

在已知的初始环境中, 根据目标点和障碍物的位

置信息构建人工势能场, 这时人工势能场中每个状态的势能值就代表该状态的最大累积回报$V_0(s_i)$, 其关系式如下:

$$V_0(s_i) = |U'(s_i)|, \tag{7}$$

其中: $|U'(s_i)|$为根据已知环境信息构造的人工势能场中状态s_i的势能值, $V_0(s_i)$为状态s_i遵循最优策略所能获得的最大累积回报. 状态–动作对(s_i, a)的初始Q值定义为在状态s_i下选择动作a所获得的立即回报r加上后继状态的最大折算累积回报:

$$Q_0(s_i, a) = r + \gamma V_0(s_j), \tag{8}$$

其中: s_j为机器人在状态s_i下选择动作a所产生的后继状态, $Q_0(s_i, a)$为状态–动作对(s_i, a)的初始Q值, r为立即回报, 计算公式如下:

$$r = \begin{cases} 1, & s_j\text{为目标点}, \\ -0.2, & s_j\text{为障碍物区域点}, \\ -0.1, & \text{其他}. \end{cases} \tag{9}$$

为了说明机器人强化学习Q值初始化过程, 选择仿真实验中机器人目标点邻域进行演示. 在人工势能场中, 目标点邻域内状态的势能值如图1节点中数值所示, 每个节点对应于一个离散的状态, 每个状态的最大累积回报等于该状态的势能值, 深灰色节点表示目标状态, 浅灰色节点表示障碍物. 每个箭头代表一个动作, 若机器人导向目标状态G则获得的立即回报为1, 若与障碍物发生碰撞则获得的立即回报为–0.2, 若机器人在自由空间移动则获得的立即回报为–0.1. γ为折算因子, 选择$\gamma = 0.95$, 根据式(8)即可实现对Q值的初始化, 每个状态–动作对的初始化Q值如图1中箭头代表数值所示. 初始化完成以后, 机器人在初始环境的任意状态下都能够选择合适的动作, 当机器人面临较复杂的环境时, 在学习的初始阶段就具有一定的目的性, 而不是完全随机地选择动作, 从而加快算法收敛速度.

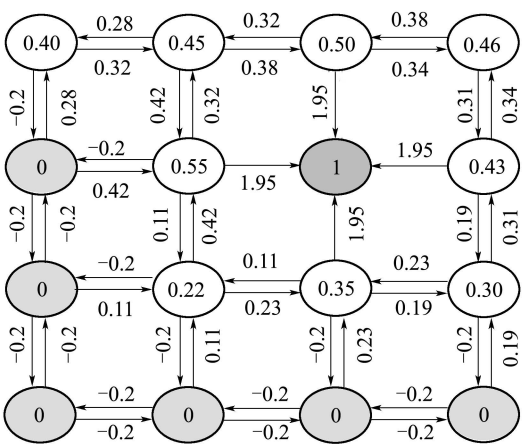


图 1 机器人目标点邻域内状态势能值及初始Q值$Q_0(s, a)$

Fig. 1 The potential energy of the neighbouring states of the target and the initial values of Q values

基于人工势能场的机器人路径规划强化学习算法步骤如下:

a) 根据已知环境信息构造人工势能场U'.

b) 根据下式将势能场中状态s_i的势能值定义为状态s_i可获得的最大累积回报:

$$V_0(s_i) = |U'(s_i)|.$$

c) 按照如下规则初始化Q值:

$$Q_0(s_i, a) = r + \gamma V_0(s_j).$$

d) 观察当前状态s.

e) 继续在复杂环境中探索, 在当前状态s下选择一个动作a并执行; 环境状态更新为新的状态s', 并接收立即回报r.

f) 观察新状态s'.

g) 根据下式更新表项$Q(s, a)$值:

$$Q(s,a) = (1-\alpha)Q(s,a) + \alpha(r + \gamma \underset{a'}{\text{argmax}} Q(s', a')).$$

h) 判断机器人是否已经到达目标或者学习系统已经达到设定的最大学习次数, 如果两者满足其一, 则学习结束, 否则返回到步骤d)继续学习.

3.2 收敛性(Convergence)

在对Q值进行初始化时, 根据先验知识可以构造出惟一的人工势能场, 因此, 模型参数一定时Q初始值是惟一确定的值, 算法初始化过程是收敛的. 为了证明学习过程的收敛性, 假设学习系统为确定性马尔可夫决策过程, 机器人在学习过程中获得的回报是有界的, 即$(\forall s, a)|r(s, a)| \leqslant c$, $\alpha(0 \leqslant \alpha < 1)$为学习率, $\gamma(0 \leqslant \gamma < 1)$ 为折算因子. 假设$Q'_n(s, a)$为第n次更新Q值时$Q(s, a)$的估计值, 定义Q'_n的最大误差为[12]

$$\delta_n = \max_a |Q'_n(s, a) - Q(s, a)|.$$

第$n+1$次迭代中更新Q值后, 估计值Q'_{n+1}的误差为

$$\begin{aligned}
&|Q'_{n+1}(s, a) - Q(s, a)| = \\
&|(1-\alpha)Q'_n(s, a) + \alpha(r + \gamma \max_{a'} Q'_n(s', a')) - \\
&(1-\alpha)Q(s, a) - \alpha(r + \gamma \max_{a'} Q(s', a'))| = \\
&|(1-\alpha)(Q'_n(s, a) - Q(s, a)) + \\
&\alpha\gamma(\max_{a'} Q'_n(s', a') - \max_{a'} Q(s', a'))| \leqslant \\
&|(1-\alpha)(Q'_n(s, a) - Q(s, a))| + \\
&\alpha\gamma \max_{a'} |Q'_n(s', a') - Q(s', a')| \leqslant \\
&(1-\alpha)\delta_n + \alpha\gamma\delta_n = \\
&(1 + (\gamma - 1)\alpha)\delta_n < \delta_n.
\end{aligned}$$

从以上推导过程可以看出, 对于任意s和a在第$n+1$次更新Q值后, 其误差小于第n次更新时的值,

随着学习过程的进行, Q值误差趋于零, 其值收敛于唯一确定的值. 学习过程中每个状态都被不断地访问, Q值会以概率1收敛于最优值, 因此, Q学习算法的收敛性与Q初始值无关, Q初始值只有可能影响算法的学习速度, 在初始化过程收敛的情况下, 改进算法必定是收敛的.

4 仿真实验分析(Analysis of simulation experiments)

为了验证本文提出的算法的有效性和先进性, 利用机器人在二维栅格地图环境中的路径规划问题进行仿真实验. 机器人工作空间由20×20个方格组成, 机器人能够在4个方向上移动, 在任意状态可以选择上下左右4个动作. 在学习过程中, 机器人根据当前状态选择动作, 如果该动作使机器人到达目标则获得的立即回报值为1, 如果机器人与障碍物发生碰撞则获得立即回报为−0.2, 如果机器人在自由空间移动则获得的立即回报为−0.1.

在仿真试验中, 学习过程参数设置如下: 最大尝试次数maxtri = 300, 每次尝试最大迭代次数maxiter = 300, 如果机器人连续10次尝试迭代次数标准误差小于0.25, 则算法收敛. 人工势能场相关参数设置如下:

$$k_{\rm a}=1.5,\ k_{\rm r}=1.2,\ \rho_0=2.0.$$

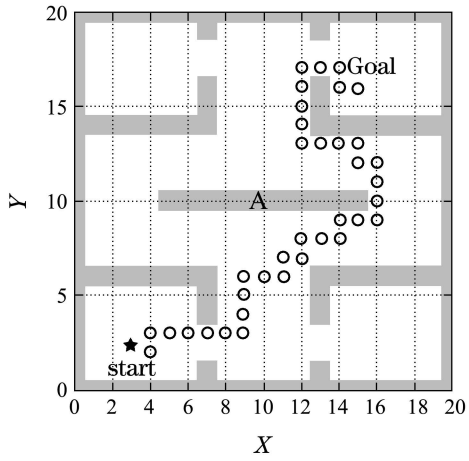
强化学习相关参数设置如下:

$$\alpha=0.3,\ \gamma=0.95,$$

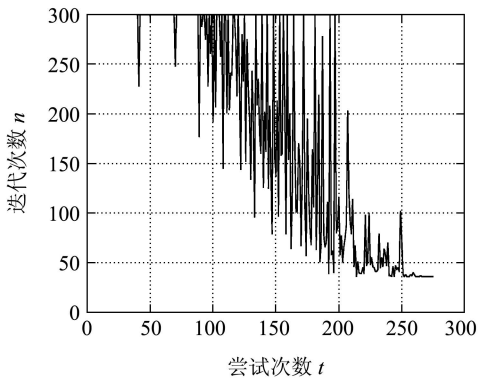
探索概率ε初始化为0.5, 并随尝试次数衰减. 为了对所提出的改进强化学习算法进行验证, 在相同的工作环境下分别对改进Q–学习算法和标准Q–学习算法进行了仿真实验.

标准Q–学习算法仿真结果如图2所示, Q值被初始化为0, 移动机器人从初始位置Start出发, 通过最短路径到达目标点Goal, 规划路径由连续的小圆圈表示(如图2(a)所示). 图2(b)为基于传统Q–学习的机器人强化学习收敛过程, 算法在经过265次尝试以后开始收敛, 在学习的初始阶段(如前80次尝试)机器人在最大迭代次数内基本都不能到达目标点. 这是由于Q值被初始化为0, 使得机器人没有任何先验知识, 只能随机地选择动作, 从而导致学习初始阶段效率较低, 算法收敛速度较慢.

为了加快机器人路径规划Q–学习算法的收敛速度, 笔者利用人工势能场对已知的环境信息进行描述, 目标位置在全局范围产生吸引势能场, 已知的障碍物产生局部排斥势能场, 两种势能场的叠加产生每个状态点的总场强, 并对构建的势能场进行归一化处理, 使得目标点具有全局最大势能, 障碍物区域具有最小势能(如图3所示). 假设建筑物尺寸及位置信息已知, 障碍物A 随机放置, 其相关信息未知, 根据目标点和已知的环境信息构建人工势能场, 工作环境中每一状态的势能值定义为该状态能够获得的最大累积回报. 然后根据当前状态的立即回报与后继状态的最大折算累积回报对Q值进行初始化, 利用这种方式就把关于环境的先验知识传递给了机器人, 使机器人具有较好的学习基础.



(a) 基于标准Q–学习的机器人规划路径



(b) 标准Q–学习算法收敛过程

图 2 机器人路径规划标准Q–学习仿真结果

Fig. 2 The results of mobile robots path planning based on conventional Q-learning

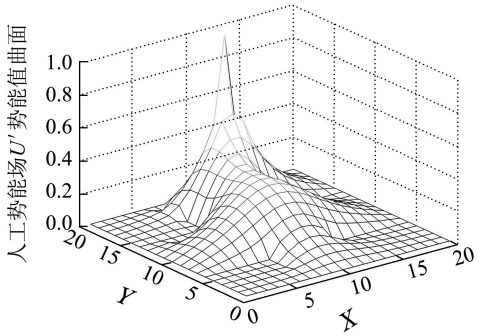
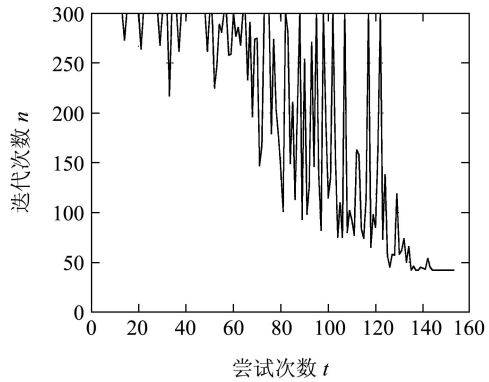


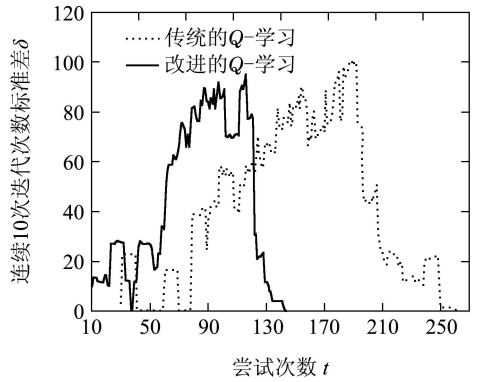
图 3 基于初始环境确定的势能值曲面

Fig. 3 The potential values based on the initial environment

为了说明基于人工势能场的机器人强化学习初始化方法的优越性, 利用与图2完全相同的环境对改进算法进行仿真实验. 首先构建如图3所示的人工势能场, 环境中每一状态的势能值代表该状态遵循最优策略获得的最大累积回报, 根据式(7)−(9) 对Q值进行初始化, 然后使机器人在新的环境中进行学习. 利用改进Q–学习算法获得的机器人强化学习收敛过程如图4(a)所示, 改进算法明显改善了学习过程的收敛速度, 算法在经过143次尝试以后开始收敛, 而且机器人在经过十几次尝试以后, 基本上都能够在最大迭代次数之内到达目标点, 与图3(b)比较可以发现, 这种启发式Q值初始化方法有效提高了算法初始阶段的学习效率, 明显地改善了机器人路径规划强化学习算法的性能.



(a) 改进Q–学习算法收敛过程



(b) 两种Q–学习算法迭代次数标准差

图 4 Q–学习算法性能分析

Fig. 4 The performance of different Q-learning algorithms

机器人每次尝试的迭代次数直接反映了本次学习的效果, 为了更加直观地比较两种学习算法的性能, 笔者计算学习过程中每连续10次尝试迭代次数的标准差, 标准差计算式如下:

$$\delta_j=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(n_{i+j-1}-\bar{n})^2},\ j=1,2,3\cdots, \quad (10)$$

其中: δ_j为第j个连续10次迭代次数的标准差, $N=10$, n_{i+j-1}为第$i+j-1$次迭代次数, $\bar{n}$为连续10次迭代次数的平均值. 根据标准差的演化过程从另一个角度比较两种算法的学习速度以及算法的稳定性, 两种Q–学习算法标准差的演化过程如图4(b)所示, 在学习的初始阶段基于标准Q–学习的机器人在最大迭代次数范围内无法到达目标点, 每次学习的迭代次数都等于最大迭代次数, 标准差为零. 随着学习过程的进行, 经过几十次尝试机器人偶尔能够到达目标点, 标准差逐渐增大. 经过一定时间的学习, 机器人每次尝试的迭代次数逐渐收敛于最短路径步数, 标准差逐渐减小直至为零. 与标准Q–学习算法相比, 改进的Q–学习算法在经过十几次尝试, 标准差就大于零, 说明机器人已经开始能够到达目标点. 在学习的最后阶段, 改进的Q–学习算法的标准差曲线也较平滑, 而且收敛速度更快, 这一现象表明基于人工势能场的Q值初始化方法能够明显加快算法的收敛速度, 并且能够使得算法收敛过程更加稳定.

5 结论(Conclusion)

在利用标准强化学习算法进行机器人路径规划时, 由于机器人没有对环境的先验知识, 使得算法收敛速度较慢. 为了将先验知识融入到学习系统之中, 使机器人获得较好的学习基础, 提高学习效率, 本文提出了一种基于人工势能场的机器人路径规划强化学习算法. 根据已知的环境信息构建人工势能场, 目标点产生全局吸引势场, 已知的障碍物产生局部的排斥势场, 两种势场叠加产生新的势能场, 使得目标点具有全局最大势能, 障碍物区域具有最小势能, 这时每个状态的势能值就代表相应状态遵循最优策略可获得的最大累积回报. 利用当前状态的立即回报及后继状态的最大折算回报即可实现Q值的初始化. 改进的Q–学习算法利用人工势能场将初始环境信息映射成为Q函数的初始值, 使得机器人获得较好的学习基础, 提高了机器人的学习效率. 最后, 通过仿真实验, 验证了所提出的改进算法能够有效提高初始阶段的学习效率, 加快算法收敛速度.

参考文献(References):

[1] 甘亚辉, 戴先中. 基于遗传算法的多机器人系统最优轨迹规划 [J]. 控制理论与应用, 2010, 27(9): 1145 – 1152.
(GAN Yahui, DAI Xianzhong. Optimal trajectory-planning based on genetic algorithm for multi-robot system [J]. *Control Theory & Applications*, 2010, 27(9): 1145 – 1152.)

[2] LEE D W, SEO S W, SIM K B. Online evolution for cooperative behavior in group robot systems [J]. *International Journal of Control, Automation and Systems*, 2008, 6(2): 282 – 287.

[3] SCHAAL S, ATKESON C. Learning control in robotics [J]. *IEEE Robotics & Automation Magazine*, 2010, 17(3): 20 – 29.

[4] ANDERSEN K T, ZENG Y, CHRISTENSEN D D, et al. Experiments with online reinforcement learning in real-time strategy games [J]. *Applied Artificial Intelligence*, 2009, 23(9): 855 – 871.

[5] 段勇, 崔宝侠, 徐心和. 多智能体强化学习及其在足球机器人角色分配中的应用 [J]. 控制理论与应用, 2009, 26(04): 371 – 376.
(DUAN Yong, CUI Baoxia, XU Xinhe. Multi-agent reinforcement learning and its application to role assignment of robot soccer [J]. *Control Theory & Applications*, 2009, 26(4): 371 – 376.)

[6] GASKETT C. *Q–Learning for Robot Control* [D]. Canberra: The Australian National University, 2002.

[7] DUNG L T, KOMEDA T, TAKAGI M. Reinforcement learning for pomdp using state classification [J]. *Applied Artificial Intelligence*, 2008, 22(7): 761 – 779.

[8] LIN L, XIE H, ZHANG D. Supervised neural Q-learning based motion control for bionic underwater robots [J]. *Journal of Bionic Engineering*, 2010, 7(Sup): 177 – 184.

[9] OH C H, NAKASHIMA T, ISHIBUCHI H. Initialization of qvalues by fuzzy rules for accelerating qlearning [C] //*Proceedings of the IEEE World Congress on Computational Intelligence*. Anchorage: IEEE, 2002: 2051 – 2056.

[10] WIEWIORA E. Potential based shaping and qvalue initialization are equivalent [J]. *Journal of Artificial Intelligence Research*, 2003, 19(1): 208 – 208.

[11] KHATIB O. Real-time obstacle avoidance for manipulators and mobile robots [C] //*Proceedings of the International Conference on Robotics and Automation*. St. Louis: IEEE, 1985: 500 – 505.

[12] MITCHELL T M. *Machine Learning* [M]. New York: McGraw Hill Science, 1997.

作者简介:

宋 勇 (1978–), 男, 博士研究生, 高级工程师, 研究领域为机器学习、移动机器人导航, E-mail: songyong@sdu.edu.cn;

李贻斌 (1960–), 男, 博士, 教授, 研究领域为机器人控制、仿生机器人, E-mail: liyb@sdu.edu.cn, 通信作者;

李彩虹 (1970–), 女, 博士, 副教授, 研究领域为机器学习、移动机器人导航, E-mail: lich@sdut.edu.cn.