

基于k最近邻分类增强学习的除冰机器人抓线控制

魏书宁, 王耀南, 印峰, 杨易旻

(湖南大学 电气与信息工程学院, 湖南 长沙 410082)

摘要: 输电线柔性结构特性给除冰机器人越障抓线控制带来极大困难. 本文提出了一种结合k-最近邻分类算法(KNN)和增强学习算法的抓线控制方法. 利用基于KNN算法的状态感知机制选择机器人当前状态k个最邻近状态, 并根据加权结果选择当前最优动作. 该方法可以得到机器人连续状态的离散表达形式, 从而有效解决传统连续状态泛化方法带来的计算收敛性和维数灾难问题. 借助增强学习算法探测和学习环境的能力, 该方法能够克服机器人模型和姿态误差及环境干扰等因素对抓线控制的影响. 文中给出了算法具体实现步骤, 并给出了应用此方法控制除冰机器人抓线的仿真实验.

关键词: 除冰机器人; k-最近邻分类算法; 增强学习; 维数灾难

中图分类号: TP273 **文献标识码:** A

Line-grasping control of de-icing robot based on k-nearest neighbor reinforcement learning

Wei Shu-ning, Wang Yao-nan, Yin Feng, Yang Yi-min

(College of Electrical and Information Engineering, Hunan University, Changsha Hunan 410082, China)

Abstract: Flexible mechanical characteristic of power lines induces difficulties for line-grasping control of de-icing robot. A line-grasping control approach for de-icing robot combining with k-nearest neighbor (KNN) algorithm and Reinforcement learning (RL) is proposed in this paper. In each learning iteration this approach makes use of KNN algorithm based state perception mechanism to select k-nearest states, and then the weighted k states are used to choose an optimal action. Expressing a continuous state by k-nearest discrete states, this approach effectively overcomes the computational convergence and the curse of dimension problems in traditional continuous state space generalization methods. By means of RL algorithm's ability of perception and learning in environment, the approach can tolerate possible robot model errors, robot arm's attitude errors and environment interferences to line-grasping control. The approach is presented in detail and the simulation results of line-grasping control based on this approach are given in this paper.

Key words: de-icing robot; k-nearest neighbor; reinforcement learning; curse of dimension

1 引言(Introduction)

输电线路巡线和除冰机器人是目前机器人研究领域的热点问题^[1~4]. 由于作业环境的特殊性, 如何实现输电线路机器人的自主越障控制是研究难点之一. 目前已有学者在巡线机器人越障领域开展了相关研究工作并取得一定的进展: 文[5]采用分层规划的方式, 在行为规划和动作规划两个层次上进行机器人的越障规划, 并将离线规划和在线规划相结合, 实现机器人越障的自主控制; 文[6, 7]基于图像的视觉伺服控制, 较好地解决了越障过程中的自主抓线控制问题; 文[8]则采用电磁导航方式, 该方法处理方便并且不存在强磁场干扰问题. 与巡线机器人相比, 除冰机器人作业环境更为恶劣, 其控制难度更高, 在复杂线路条件下的输电线路机

器人自主越障控制问题(如跨越线塔)还未完全解决. 因此, 在现有研究方法基础上, 有必要对输电线路机器人越障控制问题作进一步的研究.

增强学习算法是一类重要的机器学习方法, 其特点是可以不需要系统模型而实现无导师的在线学习, 并且具备与环境交互获取评价性反馈信号进而实现行为决策优化的能力. 作为一种简单有效的在线学习算法, 在提高算法本身实时性能的基础上, 增强学习算法不失为解决除冰机器人抓线控制的一种可能的途径. 然而, 现有的理论研究结果主要针对小规模、离散状态空间问题, 对大规模和连续空间的优化决策问题通常难以保证算法的收敛性, 且存在学习效率不高等缺点. 而许多工程应用问题都具有大规模或连续的状态和决策空间,

因此如何实现在大规模或连续状态和决策空间中的泛化,提高求解复杂问题的学习效率,是决定增强学习方法是否得到广泛应用的关键.针对上述问题,本文提出一种可在线学习的KNN-SARSA(λ)算法.本文方法首先利用KNN算法选取当前状态的k个最邻近状态,然后以其加权结果选择最优动作.从理论上分析,通过加权计算的形式可以得到整个状态空间中的任意状态,而不仅仅局限于划分的状态表中的状态形式,因而较好的实现了对连续状态的泛化,有效解决了传统连续状态泛化方法带来的计算收敛性和维数灾难问题.模拟实验结果表明,采用KNN-SARSA(λ)算法实现的除冰机器人自主抓线控制方法由于具备探测和学习环境的能力,因此具有较强的环境适用性,可作为现有越障控制方法的补充.

2 除冰机器人本体机构(Structure design of de-icing robot)

如图1所示,机器人本体部分由三个灵巧机械臂和一个控制箱组成.手臂一和手臂二结构相同,分别由大臂、小臂和末端夹持器组成.大臂和箱体、大臂和小臂、小臂和夹持器连接处分别为具有仿人手自由度的肩关节、肘关节和腕关节.手臂三为伸缩手臂,主要起稳定重心的作用.如图示2所示,夹持器由结构左右对称的挂线滚轮、驱动轮和锁死装置组成.夹持器采用灵活的模块化结构,安装不同的工作头即可完成架空线视觉和红外检查、导线清污和除冰等各种带电作业.

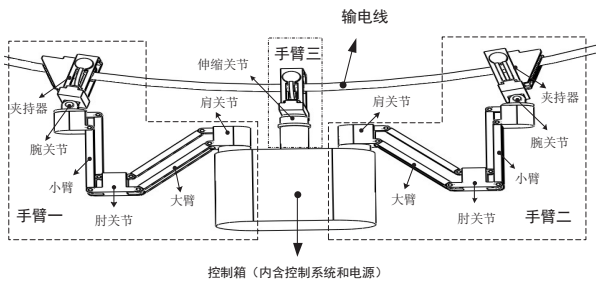


图1 除冰机器人本体结构图

Fig. 1 Structure design of de-icing robot

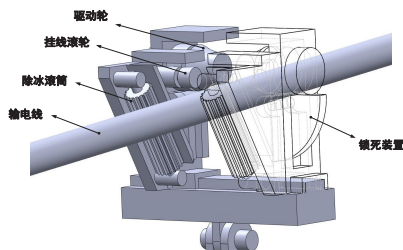


图2 夹持器内部结构图

Fig. 2 Internal structure of gripper

3 抓线控制问题描述(Line-grasping problem formulation)

除冰机器人在跨越输电线上障碍物时,总是一个手臂脱线,另外两个手臂悬挂在线上,三个手臂依次完成脱线→跨越障碍物→抓线的动作.悬挂手臂电机驱动机器人巡线前进,脱线手臂待越过障碍物后再抓线.由于手臂三只有一个升降自由度,其控制较为简单,因此,本文将重点研究手臂一和手臂二的抓线控制问题,由于二者结构相同,本文以手臂二为例进行分析.

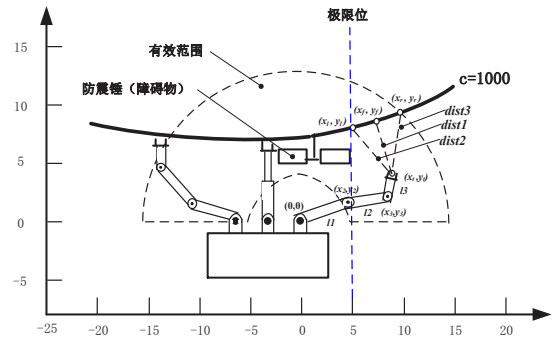


图3 除冰机器人手臂抓线模型图

Fig. 3 Model graph of line-grasping of de-icing robot

在越障过程中脱线手臂与机器人箱体,障碍物和输电线之间的位置相对静止,并且因重力作用脱线手臂与输电线始终保持在一个竖直平面内,为便于研究,以脱线手臂的基座为坐标原点建立如图3所示参考坐标系.其中大臂、小臂的长度为 l_1 和 l_2 ,考虑到实际的腕关节具有一定长度,为便于描述,可将其简化为长度为 l_3 的连杆结构.

如图3所示,虚线内区域为三关节手臂能够触及的范围,外侧虚线半圆与输电线交点坐标为 (x_r, y_r) .图中竖直虚线为机械臂不触及障碍物(以防振锤为例)的极限位置,该极限位与输电线的交点坐标为 (x_l, y_l) .在抓线控制过程中,将点 (x_r, y_r) 和 (x_l, y_l) 之间的线段定义为抓线有效线段,有效线段中心点(简称目标点)坐标为 (x_f, y_f) ,脱线手臂末端位置坐标为 (x_t, y_t) .定义相关距离: $dist1 = (x_f - x_t)^2 + (y_f - y_t)^2$, $dist2 = (x_t - x_l)^2 + (y_t - y_l)^2$, $dist3 = (x_t - x_r)^2 + (y_t - y_r)^2$.输电线可采用蔓叶线方程近似表示:

$$y^2 = \frac{x^3}{2c - x} \quad (1)$$

其中, c 为常数,对于不同弧度的输电线相应的 c 取值不同.在实际环境中,当安装在夹持器末端的接触传感器给出开关信号时,表明抓线成功.在仿真实验中,当手臂末端调整至抓线有效线段范围内时,可认为抓线成功.

4 算法描述(Description of algorithm)

传统增强学习方法只能用于解决有限离散状态问题, 对于连续状态泛化问题目前主要有两类解决方法: 函数逼近法和离散划分法. 然而采用函数逼近法难以保证算法收敛性; 采用离散划分法则存在如何确定划分规则的问题: 如果对状态空间划分过粗, 可能导致算法不收敛^[9]; 反之, 状态总数将呈指数增长, 从而导致“维数灾难”问题. 本文采用结合了资格迹的传统增强学习SARSA算法作为主要控制方法, 并利用KNN算法解决连续状态的泛化问题. 下面将对SARSA(λ)算法和KNN算法作简要回顾, 并对本文提出的KNN-SARSA(λ)作出详细的描述.

4.1 SARSA(λ)学习算法

SARSA(0)算法是Rummery^[10]等于1994年提出的一种在策略的TD (temporal difference, 时间差分) 算法, 并逐渐成为一类重要的表格型增强学习算法. 在每个时间步, 观察当前状态 s_t , 选择和执行动作 a_t , 获得即时回报 r , 如果 r 满足一定条件则终止控制, 否则观察下一时间步状态 s_{t+1} , 选择和执行动作 a_{t+1} , 然后根据下式(2)调整 Q 值. 其中 Q 值定义为基于状态-动作(s, a)的评价函数 $Q(s, a)$, 表示在状态 s 下选择动作 a 所预期获得的累积奖赏.

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)] \quad (2)$$

在SARSA(0)算法中引入资格迹^[11]产生快速多步的在策略控制学习算法—SARSA(λ), 进一步提高了算法的收敛速度. SARSA(λ)算法的基本思想是将TD(λ)预测方法用于状态-动作对, 其 Q 值的更新如下公式:

$$Q(s, a) \leftarrow Q(s, a) + \alpha \delta e(s, a) \quad (3)$$

上式中:

$$\delta \leftarrow r + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t) \quad (4)$$

$$e(s, a) = \begin{cases} \gamma \lambda e(s, a) & s \neq s_t \\ 1 & s = s_t, a = a_t \\ 0 & s = s_t, a \neq a_t \end{cases} \quad (5)$$

具体算法见文献[12,13].

4.2 KNN算法

K最近邻 (K-Nearest Neighbor, KNN) 分类算法是一个理论上比较成熟的方法^[14]. 将训练的样本映射到一个划分好类域的特征空间, 如果一个样本在特征空间中的 k 个最相似(即特征空间中最邻

近)的样本中的大多数属于某一个类别, 则该样本也属于这个类别. 算法中的关键参数 k 决定了有多少样本需要被考虑成近邻, 计算距离时通常采用欧氏距离. 由于KNN方法主要靠周围有限的邻近的样本, 而不是靠判别类域的方法来确定所属类别的, 因此对于类域的交叉或重叠较多的待分样本集来说, KNN方法较其他方法更为适合. 具体算法见文献[15].

4.3 KNN-SARSA(λ)算法

为实现对除冰机器人手臂抓线的增强学习控制, 本文提出了KNN-SARSA(λ)算法, 该算法首先需要建立增强学习系统, 即建立相应的状态集、动作集、 Q 值查找表(简称 Q 表)以及回报函数, 相关概念定义如下:

定义一: 状态集

除冰机器人抓线控制的状态集 $S(X, Y)$ 定义为:

$$S(X, Y) = \{(X, Y) \mid X = -L + \frac{2L}{n-1} \cdot i, Y = -L + \frac{L}{n-1} \cdot j\}, \forall i, j \in [1, 2, \dots, n] \quad (6)$$

其中 X 、 Y 分别表示手臂末端和目标点横坐标和纵坐标的差值; 借助图像传感器, 通过图像标定技术或在仿真中直接求解运动学方程获取 X 和 Y 的值. L 为手臂三个连杆长度之和. 状态集元素表示坐标系中的手臂末端和目标点之间的相对位置. 本文中 $S(X, Y)$ 的状态变量 X 、 Y 的取值个数 n 为4, 因此该状态集合中有 $4 \times 4 = 16$ 个元素, 即可表示16种状态.

定义二: 动作集

定义动作 $a = \langle a_1, a_2, a_3 \rangle$, 其中 $a_i (i = 1, 2, 3)$ 对应除冰机器人抓线手臂的第 i 个关节的关节转动量. 手臂关节的动作集设为: $A = \{-2, -1, 0, 1, 2\}$, 动作集单位为度, $a_i \in A (i = 1, 2, 3)$. 正号表示关节沿顺时针方向转动, 负号表示关节沿逆时针方向转动. 例如 $a_1 = -2$ 表示第一个手臂关节朝逆时针方向旋转 2° .

定义三: Q 值查找表

为三关节手臂的每一个关节建立 $C_S \times C_A$ 维 $Q_i (i=1,2,3)$ 矩阵, 记为 Q 值查找表(Q 表), 其中 C_S 为状态集元素个数, C_A 为动作集元素个数, Q_i 表中元素 $Q_i(s, a)$ 即为 Q 值.

定义四: 回报函数

通过安装在机器人末端夹持器上的接触传感器检测手臂是否抓线成功. 当手臂成功抓线时, 给出一个正的、较大的回报值, 否则给出负的回报

值.本文回报函数 r 定义如下:

$$r = \begin{cases} 35000/(1 + dist), \\ \text{当接触传感器给出开关信号时} \\ -(dist1 + dist2 + dist3)/3, \text{ 其它} \end{cases} \quad (7)$$

其中, $dist = dist1 + dist2 + dist3$.在仿真研究中, 由于没有传感器反馈信号, 相应回报函数 r 定义如下:

$$r = \begin{cases} 35000/(1 + dist), \\ \text{当} |y_t - \frac{x_t^3}{2c - x_t}|^{\frac{1}{2}} < 0.01, x_r \leq x_t \leq x_l \text{时} \\ -(dist1 + dist2 + dist3)/3, \text{ 其它} \end{cases} \quad (8)$$

当某个动作能够获得环境较高的回报, 那么以后产生这个动作的趋势就会加强, 否则产生这个动作的趋势就会减弱.值得注意的是, 回报函数的建立并不是唯一的, 合理地建立回报函数能够加快收敛速度.

建立了增强学习系统后, 下面将给出基于KNN-SARSA(λ)算法控制抓线的详细步骤:

Step1: 初始化

将 $Q_i(i=1,2,3)$ 表均初始化为0矩阵, 令资格迹 e 矩阵与 Q_i 矩阵维数相同, 并初始化为0矩阵.给下列参数赋值: $k, \alpha, \gamma, \lambda, \varepsilon$. k 表示选择近邻的个数.当 $k=1$,本文算法转化为SARSA(λ); 当 $k>1$,实验表明 k 在范围[3, 7]之间取整数值效果较好. $\alpha \in [0, 1]$ 表示学习率,它决定了学习过程中调整策略的程度. $\gamma \in [0, 1]$ 表示折扣因子,它决定了期望回报对当前的影响. $\lambda \in [0, 1]$ 为资格迹系数, $\gamma\lambda$ 构成资格迹衰减率, 使资格迹依照概率 $\gamma\lambda$ 衰减. $\varepsilon \in [0, 1]$ 表示随机动作的选择概率.

Step2: 获取当前状态

(1) 求出当前状态, 即计算手臂末端和目标点的相对距离: $s_t = [(x_f - x_t), (y_f - y_t)]$.

(2) 计算 s_t 与状态集中欧氏距离最小的 k 个状态 $\bar{s}_j(j = 1, 2, \dots, k)$.

(3) 计算状态 \bar{s}_j 的权值 ω_j :

$$\omega_j = \frac{1}{1 + d_j^2}, \quad \forall j \in [1, 2, \dots, k] \quad (9)$$

其中 d_j 为状态 s_t 和近邻状态 \bar{s}_j 的欧氏距离, 根据(9)式, 欧氏距离越大, 相应的权值越小. 在实际计算中, 权值 ω_j 采用经过归一化处理以概率分布的形式表示^[16]:

$$p(j) = \frac{\omega_j}{\sum \omega_j}, \quad \forall j \in [1, 2, \dots, k] \quad (10)$$

Step3: 计算每个动作与 k 个状态对应的 Q 值总

合

计算 Q 表中动作集每一个元素 $a_n(n = 1, 2, \dots, m)$ 与所有近邻状态 \bar{s}_j 对应的值 $Q(\bar{s}_j, a_j)$ 与相应权值概率分布值乘积之和^[17]:

$$\bar{Q}(a_n) = \sum_{j=1}^k Q(\bar{s}_j, a_n) \cdot p(j), \quad \forall n \in [1, 2, \dots, m] \quad (11)$$

其中 k 表示选择近邻的个数, m 表示动作集元素的个数.

Step4: 选择输出动作

本文根据 $\varepsilon - greedy$ 策略选择动作输出:

$$a_t = \begin{cases} \arg \max_{a_n} \bar{Q}(a_n), & \text{以概率}(1 - \varepsilon)\text{选择} \\ \text{动作集中任意动作}, & \text{以概率}\varepsilon\text{选择} \end{cases} \quad (12)$$

概率值 ε 在每个学习周期中以如下规则衰减: $\varepsilon = 0.99\varepsilon$ 即在学习初始阶段增加随机探索的概率, 避免算法陷入局部最优; 随着有效信息的增加, ε 逐渐降低, 同时保证了算法收敛性.

Step5: 执行动作

执行动作 a_t , 即控制手臂三个关节转动相应的角度.

Step6: 计算回报值 r

根据回报函数公式(7)或(8), 计算出回报值 r .判断回报值 r 符号, 如果为正, 表明抓线成功, 控制终止, 否则执行步骤Step7.

Step7: 重复执行step2到step4, 获取下一时间步 $t+1$ 时刻状态 $s_{t+1}, \bar{s}_j, a_{t+1}$.

Step8: 更新

(1) 资格迹赋值:

$$e(s, a) = \begin{cases} p(i), & s \neq \bar{s}_j \\ 1, & s = \bar{s}_j, a = a_t \\ 0, & s = \bar{s}_j, a \neq a_t \end{cases} \quad (13)$$

(2) 对每个关节的所有近邻状态对应的 $Q_i(\bar{s}_j, a_t)$ 进行更新:

$$\delta \leftarrow r + \gamma Q_i(\bar{s}_j, a_{t+1}) - Q_i(\bar{s}_j, a_t),$$

$$\forall i \in [1, 2, 3], \forall j \in [1, 2, \dots, k]$$

$$Q_i(s, a) \leftarrow Q_i(s, a) + \alpha \delta e(s, a), \forall i \in [1, 2, 3]$$

$$(3) s_t \leftarrow s_{t+1}, a_t \leftarrow a_{t+1}, e(s, a) \leftarrow \gamma \lambda e(s, a)$$

Step9: 转步骤step5.

整个机器人抓线控制系统结构如图4所示.

5 仿真结果(Simulation)

为验证KNN-SARSA(λ)算法的有效性, 对除冰机器人三关节手臂抓线控制问题进行仿真研究, 并将KNN-SARSA(λ)算法与传统的SARSA(λ)算法作

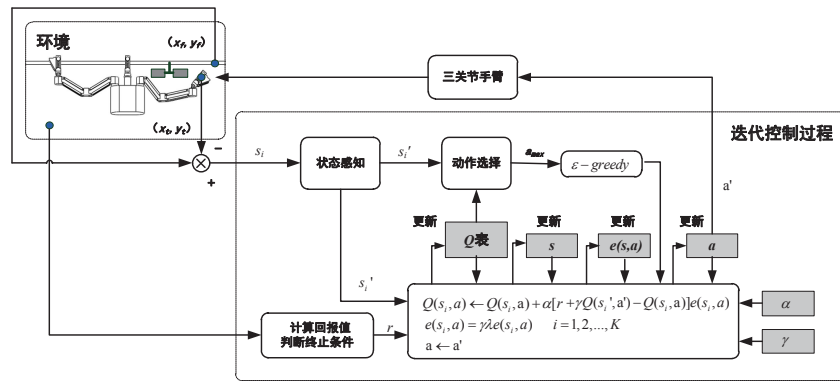


图4 基于KNN-SARSA(λ)的抓线控制结构图

Fig. 4 Structure of line-grasping control based on KNN-SARSA(λ)

了比较. 仿真算例选择300个学习周期, 每个学习周期完成一次完整的模拟抓线过程, 输电线曲线方程(1)中参数c在每次学习周期中从[100, 100000]中随机选取. 学习算法最大训练步数取400步, 当训练步数超过400步仍未成功抓线时, 表示抓线失败, 并进入下一个学习周期. 相关参数设定为: $k = 4, \alpha = 0.3, \gamma = 0.9, l_1 = 8.5cm, l_2 = 8.5cm, l_3 = 3cm, \varepsilon = 0.01$, 回报函数值r采用公式(8)计算.

的迭代步数较多, 且都存在抓线失败的情况. 随着学习周期增长, 算法所需的迭代步数总体大幅减少. 从图中可以看出, KNN-SARSA(λ)学习系统在经过大约40个学习周期后已能够获得了有效的抓线控制策略, 迭代步数稳定在20步左右, 而SARSA(λ)学习系统在经过大约40个学习周期后虽也能够获得有效的抓线控制策略, 但迭代步数仍会出现较大波动, 因此, 其学习稳定性显然要低于KNN-SARSA(λ). 另外, SARSA(λ)和KNN-SARSA(λ)算法所需的平均迭代步数分别为39.61步和22.37步, 这说明结合KNN算法后SARSA(λ)算法的学习效率得到较大提高.

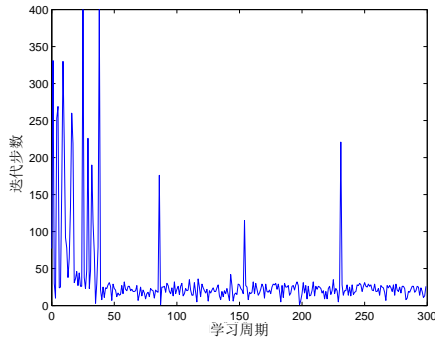


图5 SARSA(λ)学习曲线图

Fig. 5 Diagram of SARSA(λ) learning

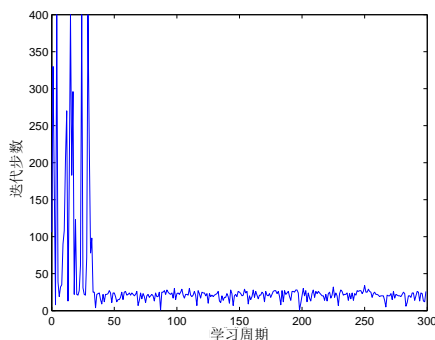


图6 KNN-SARSA(λ)学习曲线图

Fig. 6 Diagram of KNN-SARSA(λ) learning

图5、6分别给出了基于SARSA(λ)和KNN-SARSA(λ)算法收敛特性曲线, 在学习的初始阶段(约40个学习周期), 两种算法成功抓线所需

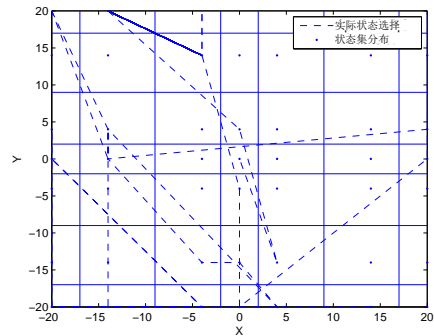


图7 离散状态轨迹图

Fig. 7 Diagram of discrete state track

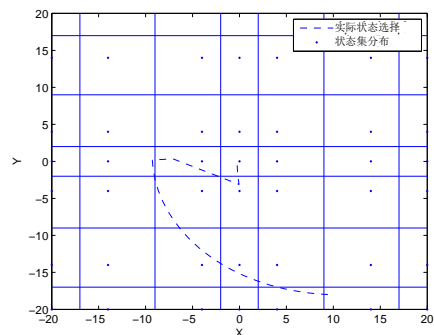


图8 连续状态轨迹图

Fig. 8 Diagram of continuous state track

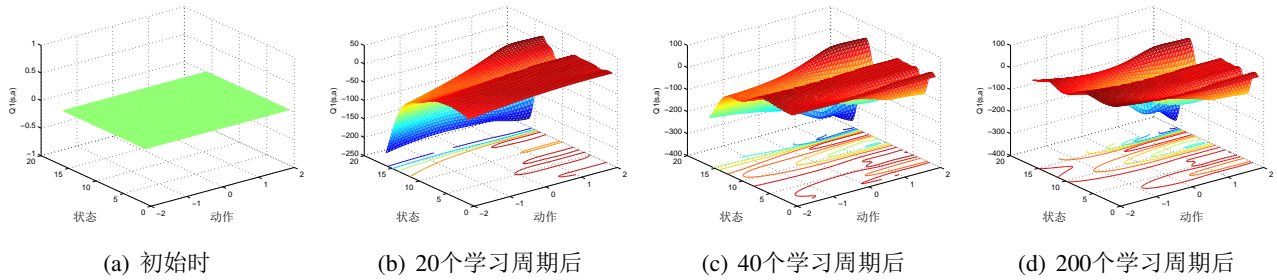


图9 不同学习周期后的 $Q_1(s,a)$ 曲面图
Fig. 9 Surface of $Q_1(s,a)$ after different episodes

图7、图8给出了分别采用SARSA(λ)和KNN-SARSA(λ)算法得到机器人手臂在收敛周期内从给定的初始状态到目标状态的状态转换轨迹(如图中虚线所示).网格区域表示以离散状态为中心所对应的离散状态的范围;离散点表示状态集中的离散状态,其中坐标(0,0)对应目标状态,该状态表示机器人抓线成功.由图7可知,SARSA(λ)算法得到的状态轨迹由离散状态集中的元素组成,由于划分的离散状态数目有限,造成的泛化误差使得算法很难平滑地收敛到目标状态.而由图8可知,KNN-SARSA(λ)算法得到的状态可以遍历状态空间的任意状态,而不仅局限于离散划分的几个状态,从而可以实现更快速平滑地收敛到目标状态.

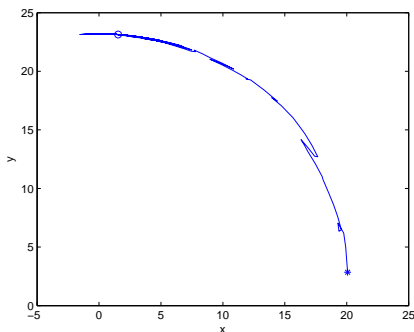


图10 第10个学习周期手臂末端运动轨迹

Fig. 10 The trajectory of the arm's terminal in the 10th episode

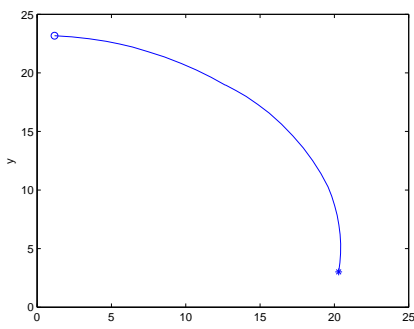


图11 第40个学习周期手臂末端运动轨迹

Fig. 11 The trajectory of the arm's terminal in the 40th episode

以关节一为例,图9给出了关节一初始时和经

过20个周期、40个周期和200个学习周期后的Q表曲面图.其中,图9(b)与图9(c)曲面形状相差较大,这表明经过20个学习周期后Q表还未学习充分.而比较图9(c)和图9(d)可知,经过40个学习周期和200个学习周期后的Q曲面没有显著的变化,说明经过40个学习周期后Q表趋于稳定,此时Q表已包含足够的知识以实现除冰机器人自主抓线控制.

图10、图11给出了KNN-SARSA(λ)算法分别在第10个和第40个学习周期得到的机器人手臂末端的运动轨迹图,为了便于比较,给定了相同的初始位置和目标位置.图中“o”点表示起始位置,*点表示目标位置,即抓线成功的位置.比较两图可知,在第10个学习周期虽然手臂末端最终能到达目标位置,但是需要较多的迭代步数(72步),并且运动轨迹不平滑,这说明机器手臂仍处于“试探”的学习过程中.到第40个学习周期由于Q表已经得到充分学习,因此机器手臂能够很快到达目标状态,并且轨迹光滑,迭代步数少(6步).

6 结论(Conclusions)

本文将KNN算法与SARSA(λ)算法相结合,提出一种可用于在线学习的KNN-SARSA(λ)控制方法.与传统增强学习算法相比较,本文方法较好地解决了对连续状态空间的泛化问题,提高了增强学习算法的学习效率.仿真实验表明,采用KNN-SARSA(λ)算法可以实现除冰机器人自主抓线控制,并且具有较强的鲁棒性,在输电线路机器人自主越障控制领域具有一定的应用前景.

参考文献(References):

- [1] TOUSSAINT K, NPOULIOT S, MONTAMBAULT. Transmission line maintenance robots capable of crossing obstacles: State-of-the-art review and challenges ahead[J]. *Journal of Field Robotics*, 2009, 26(5):477-499.
- [2] POULIOT N, MONTAMBAULT S. Geometric design of the LineScout: a teleoperated robot for power line inspection and maintenance[C]. // *Proceedings of IEEE International Conference on Robotics and Automation*. Pasadena: IEEE, 2008: 3970-3977.

- [3] 周风余,吴爱国,李贻斌,王吉岱,梁自泽.高压架空输电线路自动巡线机器人的研制[J].电力系统自动化,2004,28(23):89-91.
(ZHOU Fengyu,WU Aiguo,LI Yibin,WANG Jidai,LIANG Zize.Development of a mobile robot for inspection of high voltage overhead power transmission lines[J].*Automation of Electric Power System*,2004,28(23):89-81.)
- [4] 印峰,王耀南,余洪山.基于蒙特卡罗方法的除冰机器人作业空间边界提取[J].控制理论与应用,2010,27(7):891-896.
(YIN Feng,WANG Yaonan,YU Hongshan.Workspace boundary extraction of de-icing robot based on Monte Carlo method[J].*Control Theory & Applications*,2010,27(7):891-896.)
- [5] 李恩,梁自泽,谭民.基于规则库的巡线机器人自主越障动作规划[J].机器人,2005,27(5):400-405.
(LI En,LIANG Zize,TAN Min.Rule base based motion planning method for inspection robot to cross obstacles autonomously[J].*Robot*,2005,27(5):400-405.)
- [6] 王鲁单,王洪光,房立金,赵明扬.基于视觉伺服的输电线巡检机器人抓线控制[J].机器人,2007,29(5):451-455.
(WANG Ludan,WANG Hongguang,FANG Lijin,ZHAO Mingyang.Visual-servo-based line-grasping control for power transmission line inspection robot[J].*Robot*,2007,29(5):451-455.)
- [7] 张运楚,梁自泽,谭民,叶文波,练波.架空输电线路巡线机器人越障视觉伺服控制[J].机器人,2007,29(2):111-116.
(ZHANG Yunchu,LIANG Zize,TAN Min,YE Wenbo,LIAN Bo.Visual Servo Control of Obstacle Negotiation for Overhead Power Line Inspection Robot[J].*Robot*,2007,29(2):111-116.)
- [8] 陈中伟,肖华,吴功平.高压巡线机器人电磁传感器导航方法[J].传感器与微系统,2006,25(9):33-35.
(CHEN Zhongwei,XIAO Hua,WU Gongping.Electromagnetic sensor navigation system of robot for high-voltage transmission line inspection[J].*Transducer and Microsystem Technologies*,2006,25(9):33-35.)
- [9] Gordon G J.*Chattering in SARSA (λ)*[R].USA:CMU Learning Lab Technical Report,1996.
- [10] RUMMERY G A,NIRANJAN M.*On-line Q-learning using connectionist systems*[R].England:Engineering Department, Cambridge University,1994.
- [11] SINGH S P,SUTTON R S. Reinforcement learning with replacing eligibility traces[J].*Machine Learning*,1996,22(1):123-158.
- [12] BARTO A G,SUTTON R S. *Reinforcement Learning: an introduction*[M]. Cambridge,MA: MIT Press,1998.
- [13] KAEHLING L P,LITTMAN M L,MOORE A W. Reinforcement learning: A survey[J]. *Journal of Artificial Intelligence*,1996,4(1):237-285.
- [14] DUDANI S A. The distance-weighted k-nearest-neighbor rule[J]. *IEEE Transactions on Systems, Man and Cybernetics*,1978,6(4):325-327.
- [15] COVER T,HART P.Nearest neighbor pattern classification[J]. *IEEE Transactions on Information Theory*,2002,13(1):21-27.
- [16] MARTIN H,DE LOPE J.Ex $\langle \alpha \rangle$ An effective algorithm for continuous actions Reinforcement Learning problems[C]. //*Proceedings of 35th Annual Conference of the IEEE Industrial Electronics Society*.Spain: IEEE,2010:2063-2068.
- [17] MARTIN H,DE LOPE J,MARAVALLD.The kNN-TD Reinforcement Learning Algorithm[C]. //*Proceedings of the 3rd International Work-Conference on The Interplay Between Natural and Artificial Computation*. Spain: Springer,2009:305-314.

作者简介:

魏书宁 (1979—),女,博士研究生,目前研究方向为机器人智能控制方法,增强学习,E-mail: weishuning@sina.com;

王耀南 (1957—),男,教授,博士生导师,目前研究方向为智能机器人、智能信息处理和智能控制,E-mail: yaonan@hnu.cn;

印峰 (1983—),男,博士研究生,目前研究方向为智能机器人系统、智能控制理论,E-mail:yinfeng83@126.com;

杨易旻 (1982—),男,博士研究生,目前研究方向为智能机器人控制,自适应神经网络,E-mail:yangyimin@hnu.cn.