控制与决策 2007年 第2期：159

ISSN 1001-0920
CN 21-1124/TP

论文

SMDP 基于Actor网络的统一NDP方法

唐昊 陈栋 周雷 吴玉华

(合肥工业大学计算机与信息学院 230009)

Abstract 研究半马尔可夫决策过程(SMDP)基于性能势学习和策略逼近的神经元动态规划(NDP)方法.通过SMDP 的一致马尔可夫链的单个样本轨道,给出了折扣和平均准则下统一的性能势TD(λ)学习算法,进行逼近策略评估.利用一个神经元网络逼近结构作为行动器(Actor)表示策略,并根据性能势的学习值给出策略参数改进的两种方法.最后通过数值例子说明了有关算法的有效性.

Keywords Markov决策过程 性能势 TD(λ)学习 神经元动态规划

收稿日期 2005-10-23 修回日期 2006-1-12

通讯作者 唐昊 htang@hfut.edu.cn

DOI 分类号 TP202