

基于状态-动作图测地高斯基的策略迭代强化学习

程玉虎, 冯涣婷, 王雪松

1. 中国矿业大学信息与电气工程学院 徐州 221116

收稿日期 2010-7-5 修回日期 2010-10-13 网络版发布日期 接受日期

摘要

在策略迭代强化学习中, 基函数构造是影响动作值函数逼近精度的一个重要因素. 为了给动作值函数逼近提供合适的基函数, 提出一种基于状态-动作图测地高斯基的策略迭代强化学习方法. 首先, 根据离策略方法建立马尔可夫决策过程的状态-动作图论描述; 然后, 在状态-动作图上定义测地高斯核函数, 利用基于近似线性相关的核稀疏方法自动选择测地高斯核的中心; 最后, 在策略评估阶段利用基于状态-动作图的测地高斯核逼近动作值函数, 并基于估计的值函数进行策略改进. 10×10 格子世界的仿真结果表明, 与基于状态图普通高斯基和测地高斯基的策略迭代强化学习方法相比, 本文所提方法能以较少的基函数、高精度地逼近具有光滑且不连续特性的动作值函数, 从而有效地获得最优策略.

关键词 [状态-动作图](#) [测地高斯核](#) [基函数](#) [策略迭代](#) [强化学习](#)

分类号

Policy Iteration Reinforcement Learning Based on Geodesic Gaussian Basis Defined on State-action Graph

CHENG Yu-Hu, FENG Huan-Ting, WANG Xue-Song

1. School of Information and Electrical Engineering, China University of Mining and Technology, Xuzhou 221116

Abstract

For policy iteration reinforcement learning methods, the construction of basis functions is an important factor of influencing the accuracy of action-value function approximation. In order to construct appropriate basis functions for the action-value function approximation, a policy iteration reinforcement learning method based on geodesic Gaussian basis defined on state-action graph is proposed. At first, a state-action graph for a Markov decision process is constructed according to an off-policy method. Secondly, geodesic Gaussian kernel functions are defined on the state-action graph and a kernel sparsification approach based on approximate linear dependency is used to automatically select centers of the geodesic Gaussian kernels. At last, the geodesic Gaussian kernels based on the state-action graph is used to approximate the action-value function during the process of policy evaluation, and then the policy is improved based on the estimated action-value function. Simulation results concerning a 10×10 grid-world illustrate that the proposed method can accurately approximate the action-value function having smoothness and discontinuity properties with less basis functions as compared with the policy iteration reinforcement learning methods based on either ordinary Gaussian basis or geodesic Gaussian basis defined on a state graph, which is helpful for obtaining an optimal policy effectively.

Key words [State-action graph](#) [geodesic Gaussian kernel](#) [basis function](#) [policy iteration](#) [reinforcement learning](#)

DOI: 10.3724/SP.J.1004.2011.00044

通讯作者 程玉虎 chengyuhu@163.com

作者个人主页 程玉虎; 冯涣婷; 王雪松

扩展功能

本文信息

- ▶ [Supporting info](#)
- ▶ [PDF\(1488KB\)](#)
- ▶ [\[HTML全文\]\(OKB\)](#)
- ▶ [参考文献\[PDF\]](#)

参考文献

服务与反馈

- ▶ [把本文推荐给朋友](#)
- ▶ [加入我的书架](#)
- ▶ [加入引用管理器](#)
- ▶ [复制索引](#)
- ▶ [Email Alert](#)

相关信息

- ▶ [本刊中 包含“状态-动作图”的 相关文章](#)
- ▶ 本文作者相关文章

- [程玉虎](#)
- [冯涣婷](#)
- [王雪松](#)