计算机研究与发展
Journal of Computer Research and Development
ISSN 1000-1239/CN 11-1777/TP
45(9): 1455-1462, 2008

基于偏向信息学习的双层强化学习算法

林 芬[1,2] 石 川[1,3] 罗杰文[1,2] 史忠植[1]
[1](中国科学院计算技术研究所智能信息处理重点实验室 北京 100190)
[2](中国科学院研究生院 北京 100049)
[3](北京邮电大学北京市智能软件与多媒体重点实验室 北京 100876)
(linf@ics.ict.ac.cn)

Dual Reinforcement Learning Based on Bias Learning

Lin Fen[1,2], Shi Chuan[1,3], Luo Jiewen[1,2], and Shi Zhongzhi[1]
[1](*Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Beijing* 100190)
[2](*Graduate University of Chinese Academy of Sciences, Beijing* 100049)
[3](*Smart Software and Multimedia of Beijing Key Laboratory, Beijing University of Posts and Telecommunications, Beijing* 100876)

Abstract Reinforcement learning has received much attention in the past decade. Its incremental nature and adaptive capabilities make it suitable for use in various domains, such as automatic control, mobile robotics and multi-agent system. A critical problem in conventional reinforcement learning is the slow convergence of the learning process. To accelerate the learning speed, bias information is incorporated to boost learning process with priori knowledge. Current methods use bias information for the action selection strategies in reinforcement learning. They may suffer from the non-convergence problem when priori knowledge is incorrect. A dual reinforcement learning model based on bias learning is proposed, which integrates reinforcement learning process and bias learning process. Bias information is used for action selection strategies in reinforcement learning and reinforcement learning is used to guide bias learning process. Thus the dual reinforcement learning model could make effective use of priori knowledge, and eliminate the negative effects of incorrect priori knowledge. Finally, the proposed dual model is validated by experiment on maze problem including simple environment and complex environment. The experimental results demonstrate that the model could converge to the optimal strategy steadily. Moreover, the model could improve the learning performance and speed up the convergence of the learning process.

Key words reinforcement learning; *Q*-learning; bias; bias learning; priori knowledge

摘 要 传统的强化学习存在收敛速度慢等问题,结合先验知识预置某些偏向可以加快学习速度.但是当先验知识不正确时又可能导致学习过程不收敛.对此,提出基于偏向信息学习的双层强化学习模型.该模型将强化学习过程和偏向信息学习过程结合起来:偏向信息指导强化学习的行为选择策略,同时强化学习指导偏向信息学习过程.该方法在有效利用先验知识的同时能够消除不正确先验知识的影响.针对迷宫问题的实验表明,该方法能够稳定收敛到最优策略;并且能够有效利用先验知识提高学习效率,加快学习过程的收敛.

收稿日期:2007-05-17;**修回日期**:2008-05-22
基金项目:国家"八六三"高技术研究发展计划基金项目(2007AA01Z132);国家"九七三"重点基础研究发展规划基金项目(2003CB317004);国家自然科学基金项目(60775035,90604017)

关键词 强化学习;Q-学习算法;偏向信息;偏向信息学习;先验知识

中图法分类号 TP18

强化学习方法[1-2]是一种无导师机器学习方法,广泛应用于智能控制、机器人和多智能体等领域[3-4].强化学习是指从环境状态到行为映射的学习,以使系统行为从环境中获得的累积奖赏值最大,通过采用不断的试错(trial-and-error)方法来发现最优的行为策略.由于外部环境提供信息较少,强化学习的学习效率通常较低.其根本原因在于学习过程仅仅从经验获得的奖赏中进行策略的改进,而忽略了大量其他有用的领域信息,因此如何结合其他机器学习技术来帮助系统加快学习速度是强化学习研究和应用的重要方向[3].

对于实际应用问题,研究认为结合先验知识预置某些偏向(bias)可以有效减小搜索空间,加快学习速度[5-10].这些方法中偏向信息用来指导强化学习的行为选择策略,能够有效利用先验知识中的偏向信息提高学习速度.但是当先验知识不正确时又可能导致学习不收敛.对此,本文提出基于偏向信息学习的双层强化学习模型.该模型将强化学习过程和偏向信息学习过程结合起来:偏向信息指导强化学习的行为选择策略,同时强化学习指导偏向信息学习过程.该模型根据先验知识初始化偏向信息;在学习的过程中,强化学习根据偏向信息调整搜索空间,避免不必要的搜索,偏向信息根据强化学习调整偏向信息,避免不正确偏向误导.针对迷宫问题的实验表明,该方法加快了学习过程的收敛.

1 相关工作

1.1 强化学习

强化学习是一种无导师机器学习方法,其基本原理是如果 Agent 的某个行为策略导致环境正的奖赏(强化信号),那么 Agent 以后产生这个行为策略的趋势便会加强.强化学习基本模型如图 1 所示.大部分强化学习过程都可以表示成 Markov 型决策过程(Markov decision process, MDP). Markov 决策过程是由四元组$\langle S,A,R,P\rangle$定义,其中包含一个环境状态集 S、系统行为集合 A、奖励函数 $R:S\times A\to\mathcal{R}$ 和状态转移函数 $P:S\times A\to PD(S)$. Markov 决策过程的本质是,当前状态向下一个状态转移的概率和奖励值只取决于当前状态和选择的动作,而与历史状态和动作无关. Agent 的目标是在每个离散状态发现最优策略以使期望折扣奖赏和最大.

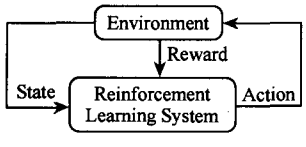

Fig. 1 Reinforcement learning model.

图 1 强化学习基本模型

为求解最优策略,学者们引入"价值函数"的概念.目前常用的价值函数是基于无限时域累积折扣回报的 Q 价值函数.Q 学习算法是一种常用的与模型无关的强化学习方法.Q 学习迭代时采用状态-动作奖赏和 Q^* 作为衡量标准,其更新规则是:

$$V(s)=\max_a Q(s,a),$$
$$Q_{t+1}(s,a)=(1-c)\times Q_t(s,a)+c\times[r+\gamma V(s')], \quad (1)$$

其中 c 是学习率,γ 是折扣因子,r 是当前回报,s' 是在当前状态 s 执行动作 a 后转移到的下一个状态.根据 Markov 理论和随机逼近理论可以证明,在一定条件下 Q 算法收敛到 MDP 的最优解[1].

1.2 相关研究

传统的强化学习过程仅仅从经验获得的奖赏中进行策略的改进,存在收敛速度慢、训练时间长等问题. Singer 等人[5]提出了一种通过归纳原问题中的局部特征来指导对新问题的求解方法. Hailu 等人[6]将环境信息置入,讨论了不同程度的偏向信息对学习速度的影响. Iglesias 等人[7-8]提出了在强化学习方法中融入先验知识的监督强化学习方法. Lin 等人[9]建立了基于隐偏向信息学习的强化学习模型. Fern′andez 等人[10]重用从过去任务中学到的偏向机制来指导新的相似任务的求解.这些方法中偏向信息用来指导强化学习的行为选择策略,能够有效利用环境或者任务等先验知识中的偏向信息,提高学习速度.但是当先验知识预置的偏向机制不正确时,偏向机制会误导 Agent,可能会导致学习不收敛.

2 基于偏向信息学习的双层强化学习模型

2.1 双层强化学习模型

双层强化学习主要目的是在强化学习中增加一个学习偏向信息的学习器,将强化学习过程和偏向

信息学习过程结合起来.强化学习器根据偏向信息调整搜索空间,避免不必要的搜索;偏向信息学习器根据强化学习自学习策略调整偏向信息,避免不正确的偏向误导.双层强化学习模型结构如图2所示,主要包括两个学习器和一个策略控制器(fusion block).两个学习器分别是强化学习器(RL controller)和偏向信息学习器(bias controller).对于某个环境状态,偏向信息学习器根据学习到的偏向机制得到一个决策建议,同时强化学习器根据自学习的策略得到一个自学习决策策略.然后策略控制器根据二者的重要性确定最终的动作对学习器进行训练.最后强化学习器根据从环境得到的反馈信号来调整内部状态,偏向信息学习器根据强化学习结果调整偏向信息.相对于传统的基于偏向信息的强化学习方法,该模型不仅能够有效地利用问题中包含的偏向信息,而且能够根据强化学习自学习过程来学习系统的偏向机制,从而有效消除了不正确先验知识的影响,加快了学习过程的收敛.

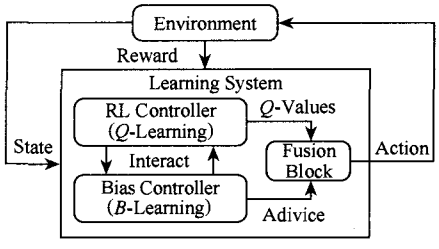

Fig. 2 Dual reinforcement learning model.

图2 双层强化学习模型

2.2 偏向信息学习

偏向信息是为进行某种偏向性学习而预置的信息[6].领域知识表示方法各不相同,为了能够直接简单反映系统的偏向机制,这里采用状态-动作对的偏向概率来表示.

定义1. 偏向信息.偏向信息为一个三元组$\langle s, a, \mu(s, a)\rangle$,表示在状态$s$执行动作$a$的概率为$\mu(s, a)$,其中:

1) $\mu(s, a)\in[0,1]$;

2) $\Sigma_a\mu(s, a)=1$,即状态s下的所有动作a偏向概率之和为1.

使用偏向概率能够简单直接表示领域知识,比如,根据以往经验,有一条规则$s\rightarrow a$,那么将$\mu(s, a)$设为1,该状态下的其他动作的偏向概率都为0.

偏向信息学习(bias learning)指根据自学习经验、环境特征等信息学习偏向信息的过程,使得最终学到的偏向信息能够准确反映系统的偏向机制.

定义2. 偏向信息学习.偏向信息学习指学习系统偏向信息的过程,学习过程用迭代公式表示为

$$\mu_{t+1}(s, a)=\mu_t(s, a)+\eta\times f(s, a). \quad (2)$$

1) η为学习常数;

2) f为偏向信息学习函数;

3) $\mu_t(s, a)$表示第n次迭代更新的状态-动作对(s, a)偏向概率.

偏向信息学习的过程是通过学习函数f来实现的,依赖具体的问题设置不同的学习函数.通过f学到的偏向信息必须合理反映系统的偏向机制.也就是说,在学习的过程中,正确的偏向概率得到加强,不正确的会被减弱.一般来说,f满足以下性质:

性质1. 任意迭代步t时,在学习函数f调整后,任意状态动作对的偏向概率都应该大于等于0且小于等于1,即$0\leqslant\mu_{t+1}(s, a)\leqslant 1$;

性质2. 任意迭代步t时,在学习函数f调整后,状态s下的所有动作a偏向概率之和为1,即$\Sigma_a\mu_{t+1}(s, a)=1$;

性质3. 学习函数f对偏向概率的调整最终能到达一个稳定的值,即当$t\rightarrow\infty$,$f\rightarrow 0$;

性质4. 最终稳定的偏向信息能够准确反映系统的偏向机制.

其中性质1和性质2可以通过对调整后的偏向概率归一化处理得以满足.根据这些性质,针对不同的问题可以设计不同的偏向信息学习函数,使得最终学到的偏向信息能够准确反映系统的偏向机制.

2.3 行为选择策略

行为选择策略是指在状态s下选择执行哪个动作的过程,也就是根据当前Q价值和偏向信息如何确定最终动作.一般通过概率选择方法对两个决策进行权衡.行为选择可以取决于以下概率分布:

$$P(a\mid s)=\omega\times\mu(s, a)+(1-\omega)\times P_Q(a\mid s), \quad (3)$$

其中$P(a\mid s)$为Agent在状态s时选择a的概率;$P_Q(a\mid s)$为根据当前Q值在状态s时选择a的概率,常用的有Boltzmann搜索策略[8]、ξ-greedy策略[9].概率分布ω为一常量且$\omega\in[0,1]$,表示当前偏向信息对最终决策的重要性.通常情况下,ω是随着迭代次数而变化的,Agent在学习的初期可以更多使用偏向信息策略,然后逐步转换到自学习的策略,可以设计ω如下:

$$\omega=\omega_0\times e^{-k\times t/T}, \quad (4)$$

其中ω_0为初始概率分布,k为调节常数,t为当前步数,T为总的迭代步数.

2.4 双层强化学习算法

双层强化学习算法是在强化学习过程中增加偏向信息学习过程,根据策略控制器选择行为.其中强化学习算法可以选择各种算法,偏向信息学习过程根据设计不同偏向信息学习函数来实现.下面给出基于偏向信息学习的 Q 学习算法(bias-Q-learning),如图 3 所示.对于某个环境状态 s,策略控制器根据式(4)选择要执行的动作,得到反馈信号 r,观察到新状态 s'.Q 学习器根据式(1)更新 $Q(s, a)$ 表项,偏向信息学习器根据式(2)更新 $\mu(s, a)$ 表项.然后将当前状态 s 设置为观察到的新状态 s',即 $s \leftarrow s'$.

```
Bias-Q-Learning
Begin
  Initialize Q(s, a) for all s, a;
  Initialize μ(s, a) for all s, a;
  Observe the current state s;
  Repeat( for each step of trial)
    Select an action a according to Formula(3);
    Execute the action a;
    Receive immediate reward r;
    Observe the new state s';
      Update the table entry for Q(s, a) according to Formula(1);
      Update the table entry for μ(s, a) according to Formula(2);
    Set the current state to the new state s←s';
    Until( termination condition is met)
End
```

Fig. 3 Bias-Q-learning algorithm.

图 3 偏向信息-Q 学习算法

3 偏向信息设计

3.1 迷宫问题

迷宫问题是一类经典的人工智能问题,常被用来研究强化学习算法,比如推箱问题[8]、机器人导航问题[9]等.一般用二维网格世界来描述,每个网格代表 Agent 的一种状态.环境中可能存在障碍物.学习的目的是寻找一条从起始状态到目标状态的最短路径.Agent 共有 4 种控制:向上、下、左和右移动.

3.2 偏向信息初始化

偏向信息初始化指根据先验知识预置偏向概率的过程.以迷宫问题为例,预置的信息可分为两类:环境偏向信息和目标偏向信息[5].

1) 环境偏向信息(environment bias):Agent 根据环境特征避免碰撞设置的信息.Agent 为避免碰到边界或者障碍物,将那些会导致碰到边界或者障碍物的状态-动作对的偏向概率设置为 0.

2) 目标偏向信息(goal bias):Agent 根据目标特征预置的信息.目标偏向信息一般很难精确表示,这里采用以下规则来描述.

规则 1. 若目标在上方,则向上方移动的概率比其他方向大.若目标在下方,则向下方移动的概率比其他方向大.若目标在左方,则向左方移动的概率比其他方向大.若目标在右方,则向右方移动的概率比其他方向大.

规则 2. 若目标在障碍物的后面,则向通过障碍物方向移动的概率比其他方向大.

规则 3. 若目标不在障碍物的后面,则向避开障碍物方向移动的概率比其他方向大.

规则 1 是 Agent 不了解环境中障碍物而设置的靠近目标的规则,规则 2,3 是 Agent 对环境中障碍物有一定了解而设置避开障碍物靠近目标的规则.使用这些规则初始化偏向概率,将优先动作的偏向概率按照一定程度设置比其他动作大.

3.3 偏向信息学习

根据先验知识初始化的偏向概率可能存在误导,偏向信息学习目的是根据自学习过程来调整状态-动作对的偏向概率,使得正确的偏向概率得到加强,不正确的能够减弱.偏向信息学习是通过学习函数 f 来调整的,根据第 2.2 节中学习函数的性质,定义 Q 偏向学习函数 f_Q 如下.

定义 3. Q 偏向学习函数.Q 偏向信息学习函数 f_Q 指根据 Q 学习过程中 Q 值变化来调整状态-动作对的偏向概率:

$$f_Q = ((Q - Q^-) \times \Delta Q^{1/\lambda} \times \mu) / Q^-. \quad (5)$$

1) Q^- 为该状态所有状态-动作对 Q 平均值;

2) ΔQ 为当前 Q 值变化;

3) λ 为更新率,且大于 0,控制根据 Q 值变化调整的程度.

如果 Q 回报非负并且所有的 Q 值初始化为 0 时,$\Delta Q \geqslant 0$,f_Q 有下列性质:

性质 5. 当 $t \to \infty$,$\Delta Q \to 0$,有 $f_Q \to 0$.偏向概率最终会收敛到一个稳定值.

性质 6. 若 $Q \geqslant Q^-$,有 $f_Q \geqslant 0$,若 $Q \leqslant Q^-$,有 $f_Q \leqslant 0$,当 Q 值大于平均值时偏向概率得到加强,当 Q 值小于平均值时偏向概率会被减弱.Q 值是 Agent 获得的累积回报估计值,较大的 Q 值对应较优的策略,因此该学习函数能够朝着优化的方向调整偏向概率,准确反映系统的偏向机制.

4　实验研究

4.1　实验环境

本文实验采用10×10的二维网格世界,如图4所示.图中每个网格代表智能体的一种状态,其中S为初始状态,G为目标状态,白色表示Agent的活动区域,黑色表示障碍物,环境中障碍物和目标都是静止的.实验分别在两种迷宫环境进行,图4(a)为无障碍物环境,而图4(b)为有障碍物环境.无障碍物环境是指除四周墙壁外,内部没有不可越区域,通常代表简单环境;有障碍物环境是指在四周和内部均存在不可越区域,代表复杂环境.

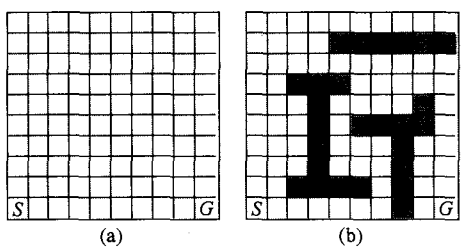

Fig. 4　Experimental environment. (a) Simple Environnent and (b) Complex Environnent.

图4　实验环境.(a)无障碍物环境;(b)有障碍物环境

每次实验在两种实验环境中独立运行20次,并记录下Agent在每次尝试中的平均步数.每一次实验,学习者从初始状态出发,选择行为并执行,接受所获得的即时报酬,只有当到达目标时有回报值$r=10$,如果遇到边界或者障碍物Agent停在原地.若到达目标,则立即重新从起始状态开始进行新的实验,每次实验迭代200步.Q初始化为0,μ根据各种先验知识来初始化.Q学习算法中参数设置如下:学习率$c=0.1$,折扣因子$\gamma=0.9$,采用Boltzmann搜索策略.偏向信息-Q学习采用式(5)来学习偏向信息,其中更新率$\lambda=4$,行为选择策略采用式(4),其中$\omega_0=1$,$k=20$.

4.2　实验结果

实验主要关注3个方面:①没有先验知识时偏向信息-Q学习性能;②不同程度先验知识时偏向信息-Q学习性能;③先验知识不正确或者不充分时偏向信息-Q学习性能,进一步与已有相关工作比较,分析了偏向信息-Q学习优势.

1) 没有先验知识时偏向信息-Q学习性能

偏向概率初始化为$1/m$,m为该状态下的动作数.在两种实验环境中运行结果如图5所示.没有先验知识时,偏向信息-Q学习收敛速度与Q学习基本一样.由此可见,没有先验知识时偏向信息-Q学习能够稳定收敛到最优解.也就是说没有先验知识的偏向信息-Q学习退化为标准的Q学习.

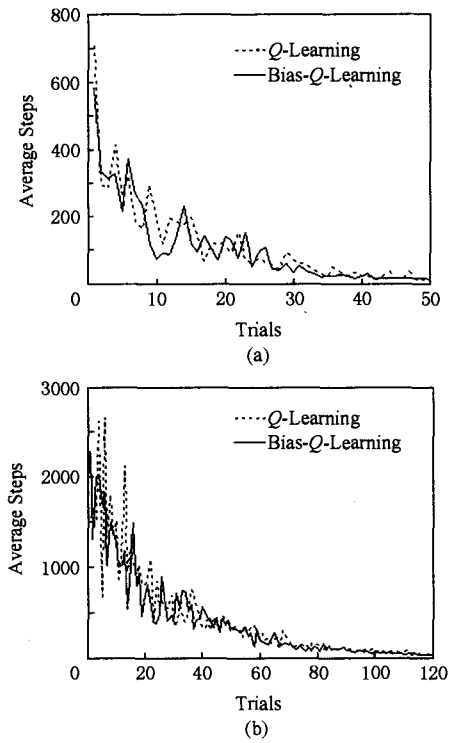

Fig. 5　Performance of bias-Q-learning without priori knowledge. (a) Simple Environment and (b) Complex Environnent.

图5　无先验知识时偏向信息-Q学习性能.(a)无障碍物环境;(b)有障碍物环境

2) 不同程度先验知识时偏向信息-Q学习性能

根据先验知识预置偏向信息主要包括两类:环境偏向信息和目标偏向信息.对于两种实验环境下,环境偏向信息的初始化是一样的,将导致碰撞的状态-动作对的偏向概率设置为0.但是目标偏向信息很难精确表达,采用第3.2节中的3个规则来初始化偏向概率.对无障碍物环境,采用规则1.对于有障碍物环境,采用规则1～3.

实验结果如图6所示,根据先验知识初始化偏向概率后,较大程度地提高了学习速度.在无障碍物环境下,环境偏向信息只包含四周边界环境信息.而在有障碍物环境下,环境偏向信息不仅包含了四周边界环境信息,而且包含了障碍物环境信息.因此加入环境偏向信息后,有障碍物环境的改进程度比无障碍物环境大.加入目标偏向信息后,对于无障碍物环境,在规则1的作用下,算法有完全的先验知识,

曲线基本上为一条直线;对于有障碍物环境,规则1~3也在很大程度提高了收敛速度.由此可见,偏向信息-Q学习能够有效利用先验知识,加快学习速度,信息越多收敛速度越快.

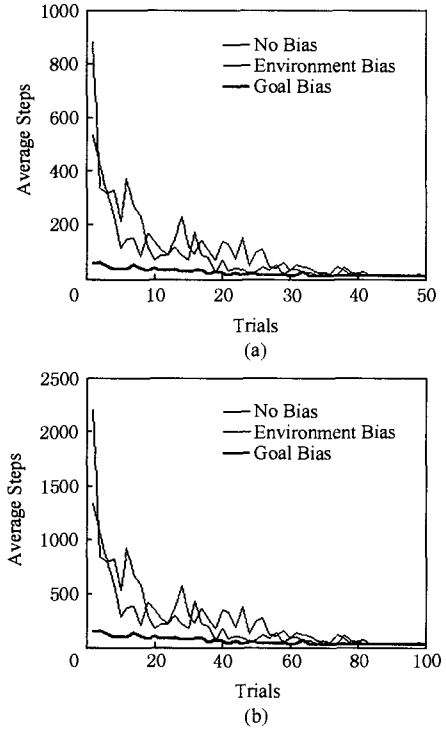

Fig. 6 Performance of bias-Q-learning with different amounts of priori knowledge. (a) Simple Environment and (b) Complex Environnent

图6 不同程度先验知识时偏向信息-Q学习性能.(a)无障碍物环境;(b)有障碍物环境

3) 先验知识不正确或者不充分时,偏向信息-Q学习性能

对于动态的环境先验知识很难准确获得.假设两种环境下没有环境偏向信息,Agent也不知道障碍物的情况.采用不同方式设置目标偏向信息.对无障碍物环境,将远离目标方向的概率设置得比靠近目标方向大.对有障碍物环境,根据规则1来初始化偏向信息.

实验结果如图7所示.对于无障碍物环境,初始化的偏向信息是错误的,使得学习初期效果较差,随着自学习的过程,Agent能够朝着正确的方向调整偏向概率,从而保证算法收敛.对于有障碍物环境,规则1没有描述障碍物的情况,可能会对Agent产生误导,使得初期效果较差,但是随着自学习的过程,Agent能够朝着正确的方向调整偏向概率.学习初期之所以出现误导是因为规则1没有描述障碍物情况,比如目标在Agent的右方,根据规则Agent总是往右走,总会遇到障碍物,导致Agent来回折返,因此出现初期的震荡情况.如果不调整偏向概率,算法可能会一直像初期这样震荡,从而不收敛.尽管这些规则描述得不充分,但是对于Agent仍然有一定的指导意义,因此在学习的后期,算法的性能仍然得到了一定的提高.由此可见,偏向信息-Q学习能够有效避免先验知识的误导,对于描述不准确的领域知识,算法仍然能够有效利用先验知识中包含的偏向信息,加快学习速度.

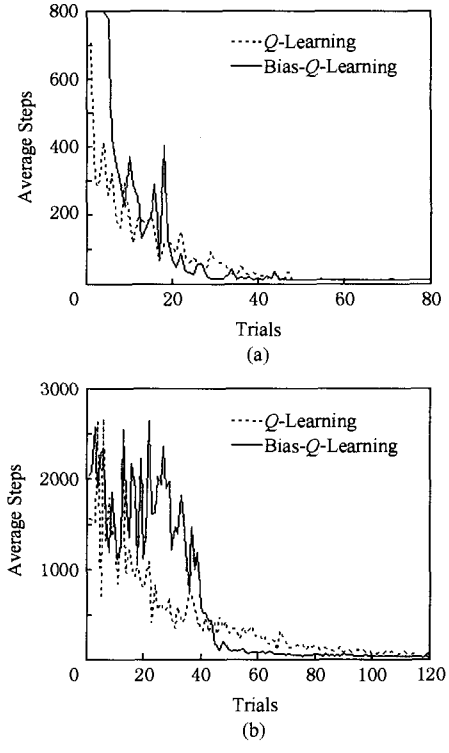

Fig. 7 Performance of bias-Q-learning with incorrect priori knowledge. (a) Simple Environment and (b) Complex Environnent.

图7 不正确或者不充分先验知识时偏向信息-Q学习性能.(a)无障碍物环境;(b)有障碍物环境

已有的相关工作利用先验知识来指导强化学习的行为选择策略方法.预置的偏向机制通过初始化偏向概率来指导强化学习的行为选择策略,没有偏向信息学习的过程.当预置的偏向机制不正确或者不充分时,偏向机制会误导Agent,从而可能导致学习不收敛.例如图7(b)中初期的震荡情况,当没有偏向信息学习过程中,会一直持续下去.本文提出的双层强化学习方法加入了偏向信息学习过程,能够根据强化学习结果朝着正确的方向调整偏向概率.

从而避免了当先验知识不正确或者不充分时算法不收敛情况.

5 总 结

本文提出了基于偏向信息学习的双层强化学习模型.该模型将强化学习的过程和偏向信息的学习过程结合起来,相互促进共同学习.以迷宫问题为例,详细讨论了偏向信息根据先验知识初始化方法和根据 Q 学习调整的学习方法.实验结果表明,该模型能够稳定收敛到最优策略.并且该模型能够有效利用先验知识加快学习速度,甚至在先验知识描述不正确时能够避免先验知识误导.

对于动态的环境,先验知识往往难以准确获得,因此该模型具有一定普遍意义.今后的工作主要围绕容错的偏向信息学习展开,将基于偏向信息学习的强化学习应用于软件自恢复、多智能体等实际领域.

参 考 文 献

[1] Mitchell Tom M. Machine Learning [M]. New York: McGraw Hill, 1997

[2] Sutton R S, Barto S. Reinforcement Learning: An Introduction [M]. Cambridge, MA: MIT Press, 1998

[3] Gao Yang, Chen Shifu, Lu Xin. Research on reinforcement learning technology: A review [J]. Acta Automatica Sinica, 2004, 30(1): 86-100 (in Chinese)
(高阳,陈世福,陆鑫.强化学习研究综述[J].自动化学报,2004,30(1):86-100)

[4] Li Ning, Gao Yang. A learning agent based on reinforcement learning [J]. Journal of Computer Research and Development, 2001, 38(9): 1051-1056 (in Chinese)
(李宁,高阳.一种基于强化学习的学习 Agent [J].计算机研究与发展,2001,38(9):1051-1056)

[5] Singer B, Veloso M. Learning state features from policies to bias exploration in reinforcement learning, CMU-CS-99-122 [R]. Pittsburgh: School of Computer Science Carnegie Mellon University, 999

[6] Hailu G, Sommer G. On amount and quality of bias in reinforcement learning [C] //Proc of IEEE SMC '99. Piscataway, NJ: IEEE, 1999: 1491-1495

[7] Iglesias Roberto, Regueiro Carlos V, José Correa, *et al*. Supervised reinforcement learning: Application to a wall following behaviour in a mobile robot [C] //Proc of IEA/AIE. Berlin: Springer, 1998: 300-309

[8] Moreno D L, Regueiro C V, Iglesias R, *et al*. Using priori knowledge to improve reinforcement learning in mobile robotics [C] //Proc of TAROS 2004. Berlin: Springer, 2004: 1744-8050

[9] Lin Yaping, Li Xueyong. Reinforcement learning based on local feature learning and policy adjustment [J]. Information Sciences, 2003, 154(1-2): 59-70

[10] Fern′andez F, Veloso M. Probabilistic policy reuse in a reinforcement learning agent [C] //Proc of AAMAS. New York: ACM, 2006: 720-727

Lin Fen, born in 1982. is a Ph. D. candidate the Key Lab of Intelligent Information Processing, the Institute of Computing Technology, CAS. Her main research interests include machine learning, information retrieval, and data mining.

林 芬,1982 年生,博士研究生,主要研究方向为机器学习、信息抽取、数据挖掘.

Shi Chuan, born in 1978. Ph. D. and lectuer. His main research interests include evolutionary computation, machine learning and data mining.

石 川,1978 年生,博士,讲师,主要研究方向为进化计算、机器学习、数据挖掘.

Luo Jiewen, born in 1980. He is a Ph. D. candidate in the Key Lab of Intelligent Information Processing, the Institute of Computing Technology, CAS. His research interests include distributed artificial intelligence, multi-agent system and decision support system.

罗杰文,1980 年生,博士研究生,主要研究方向为分布式人工智能、多主体系统、决策支持系统.

Shi Zhongzhi, born in 1941. He is a professor and Ph. D. supervisor in the Key Laboratory of Intelligent Information Processing, the Institute of Computing Technology, CAS. Senior member of CCF. His main research interests include intelligence science, multi-agent systems, semantic Web, machine learning and neural computing. He is a senior member of IEEE, a member of AAAI and ACM. He serves as vice president for the Chinese Association of Artificial Intelligence.

史忠植,1941 年生,研究员,博士生导师,计算机学会高级会员,主要研究方向为智能科学、多主体系统、语义 Web、机器学习和神经计算等.

Research Background

This work is supported by the 863 National High-Tech Program (No. 2007AA01Z132), the National Basic Research Priorities Programme (No. 2003CB317004), and the National Natural Science Foundation of China (No. 60775035, 90604017). Reinforcement learning has received much attention in the past decade. Its incremental nature and adaptive capabilities make it suitable for use in various domains, such as automatic control, mobile robotics and multi-agent system. A critical problem in conventional reinforcement learning is the slow convergence of the learning process. However, in most learning systems there usually exists priori knowledge in the form of human expertise or previously learned experience. Thus we propose a dual reinforcement learning model based on bias learning which integrates reinforcement learning process and bias learning process. The dual model makes effective use of the priori knowledge, and eliminates negative effects of incorrect priori knowledge. We believe that the model will greatly advance the use of reinforcement learning in reality, especially in complex and dynamic environment.

第3届中国数据挖掘会议(CCDM 2009)

征文通知

中国计算机学会人工智能与模式识别专委会分别在北京和郑州成功主办了第1届和第2届中国分类技术与应用研讨会(CSCA 2005与CCTA 2007),得到了国内相关领域学者的热烈响应.经过人工智能与模式识别专委会专门研究,决定将中国分类技术与应用研讨会扩展为中国数据挖掘会议(CCDM,China Conference on Data Mining),由本专委会与中国人工智能学会机器学习专委会联合主办.第3届中国数据挖掘会议(CCDM'09)将于2009年8月18日在烟台大学举行.本次会议旨在为学术界和工业界的广大研究人员提供一个交流、合作平台,使得研究人员之间分享数据挖掘与知识发现领域的初创性研究成果、创新思想、最新研究进展以及系统开发经验.本次会议录用的论文将在《计算机研究与发展》、《模式识别与人工智能》、《广西师范大学学报》、《烟台大学学报》、《北京交通大学学报》等期刊的正刊发表.

征文范围(包括但不限于以下主题)

A. 数据挖掘理论与算法
A1. 分类
A2. 聚类
A3. 回归
A4. 排序(Ranking)
A5. 关联分析
A6. 连接分析(Link Analysis)
A7. 频繁模式挖掘
A8. 异常与孤立点检测
A9. 概率与统计模型
A10. 软计算
A11. 数据预处理
A12. 降维
A13. 动态数据挖掘
A14. 并行与分布式挖掘
A15. 交互式与联机挖掘
A16. 可伸缩与高性能数据挖掘
A17. 基于约束的挖掘
A18. 基于隐私保护的数据挖掘
B. 特定数据类型的挖掘
B1. 关系数据挖掘
B2. 图模式挖掘
B3. 空间与时序数据挖掘
B4. 趋势与序列分析
B5. 数据流与增量挖掘
B6. 多媒体数据挖掘
B7. 文本挖掘
B8. Web与Internet挖掘
B9. 人机交互与可视化数据挖掘
B10. 数据仓库、OLAP与数据挖掘的集成
C. 数据挖掘技术应用
C1. 市场营销
C2. 风险管理
C3. 供应链管理
C4. 客户关系管理
C5. 电子商务
C6. 金融分析
C7. 电信
C8. 生物医药
C9. 基因分析
C10. 生物信息学
C11. 入侵与欺诈检测
C12. 社会网络分析

投稿要求

① 论文应是未发表的研究成果.

② 论文语言要求中文,采用Word格式排版.请参照《计算机研究与发展》网站(http://crad.ict.ac.cn)"作者须知"中的"最终修改稿要求"书写论文.

③ 论文通过电子邮件提交(ccdm09@ytu.edu.cn).

重要日期

截稿日期:2009年3月1日

录用通知日期:2009年4月30日

联系方式

联系人:童向荣,赵凯　　通信地址:山东烟台大学计算机学院　264005

联系电话:(0535)6902209-8621; 6902601

电子邮件:ccdm09@ytu.edu.cn　　会议网址:http://ccdm09.ytu.edu.cn