文章编号 : 1002-0446(2003) 06-0526-05

基于多 Agent 的混合智能学习算法及在足球机器人中的应用*

张淑军[1] , 孟庆春[1,2] , 宋长虹[1] , 李占斌[1] , 张 文[1]

(1 .中国海洋大学计算机科学系 , 山东 青岛 266071 ; 2 . 清华大学智能技术与系统国家重点实验室 , 北京 100084)

摘 要 : 本文提出了基于多 Agent 的混合智能学习算法 ,将个体学习和群体学习有效地结合起来 ,并给出了该算法在 RoboCup 足球机器人仿真系统中的具体应用 .实验结果表明了算法的可行性与有效性 .

关键词 : 多智能体系统 ;群体学习 ;增强式学习 ;足球机器人

中图分类号 : TP24 文献标识码 : B

HYBRID INTELLIGENT LEARNING ALGORITHM BASED ON MULTI-AGENT AND ITS APPLICATION TO ROBOT SOCCER

ZHANG Shu-jun[1] , MENG Qing-chun[1,2] , SONG Chang-hong[1] , LI Zhan-bin[1] , ZHANG Wen[1]

(1 . *Computer Science Department , Ocean University of China , 266071 ;*

2 . *State Key Lab of Intelligent Technology and Systems , Tsinghua University , Beijing 10004 , China*)

Abstract : This paper presents a hybrid intelligent learning algorithm based on multi-agent , and it combines individual learning with group learning effectively . The specific application and experimental results in RoboCup simulation system show the feasibility and effectiveness of this algorithm .

Keywords : MAS ; team learning ; reinforcement learning ; RoboCup

1 引言 (Introduction)

机器学习一直是多智能体系统(MAS) 中一个主要的研究方向 ,长期以来对于单一智能体(Agent) 的学习研究较多 .随着人工智能 、智能控制等学科的发展 ,目前已提出了很多比较成熟的 Agent 学习算法 ,如人工神经网络(ANN) 、遗传算法(GA) 、增强式学习(RL)[1]等 ,但这些算法主要用于单一 Agent 的学习 ,即使是在多 Agent 环境中每个个体的学习也是各自进行 、互不干扰 .为解决一些更为复杂的问题 ,产生了一个新的研究热点 ——MAS 中群体的协作学习[2] .在 MAS 中 ,常常需要多个 Agent 协调合作以共同完成由单一 Agent 难以胜任的复杂任务 ,多个 Agent 在合作完成任务时如何通过学习采取有效的行为 ,以及这种合作行为如何随时间逐步进化并适应动态变化的环境 ,成为多智能体系统研究的主要内容 .

机器人足球是人工智能领域中颇具挑战性的课题 ,特别是仿真机器人足球 ,对于多 Agent 系统和机器学习来说都是一个很好的研究平台[3] .在类似 RoboCup 足球机器人系统的协作性 、对抗性实时复杂 MAS 中 ,要求 Agent 既要自治地 ,又要作为团队的一分子进行有效的动作 ,所以每个 Agent 既要具备个人技术 ,又要有群体协作学习能力 .为此本文提出了一种基于多 Agent 的混合智能学习算法 ,将个体学习与群体学习有效结合起来 ,并给出了相应的 Agent 体系结构以及在 RoboCup 中的应用实例 .

2 个体学习与群体学习的不足 (Shortcomings of individual learning and group learning)

在多 Agent 系统中 ,每个 Agent 同时处于两个相互联系的层次 :个体层与群体层 ;并且均表现出认知

* 基金项目 : 本课题得到山东省自然科学基金资助(项目编号 Y2002G18) .
收稿日期 :2003 - 03 - 10

与行为.个体 Agent 是个体层角色与群体层角色的平衡结合体.据此,Agent 的学习可分为两类:

(1) 个体学习:即单一 Agent 的独立学习.Agent 学习单独与环境交互的能力,如基本动作、命令和通信能力.着眼于个体环境中的学习训练.

(2) 群体学习:系统中的所有 Agent 一起学习,如学习与系统中其它 Agent 的交互、合作能力等.在多个 Agent 组成的复杂环境中进行训练.

基于单 Agent 的方法进行推理判断存在着较大的缺陷,因为任何单一 Agent 的独立求解能力是有限的,不可能同时考虑问题的所有因素,而且个体学习一般只处理与 Agent 自身相关的局部目标和本地信息以进行自主运动,彼此之间没有联系.只注重单一指标(Agent 个人技术)的提高,而不考虑 Agent 之间的交互,必然无法完成协作.而在 MAS 中,Agent 之间的交互是学习的必要条件.

另一方面,单纯的群体学习有时无法顾及 Agent 个体的需要,如当提高个人技术与提高整体性能相冲突时,由于群体学习侧重整个系统的工作效率,学习的结果就可能牺牲 Agent 个人技术的提高.

为了克服二者各自的不足,既能顾及 Agent 之间的相互作用,又不损失 Agent 自身的个性和特点,我们提出了基于多 Agent 的混合智能学习算法.该算法是一种多 Agent 交互式学习算法,可使每个 Agent 充分发挥其智能(学习能力)和自主行为来与系统中其它 Agent 进行合作,体现创现的智能行为.

3　基于多 Agent 的混合智能学习算法(Hybrid intelligent learning algorithm based on multi-agent)

3.1　算法解决的关键问题

(1) 区分个体行为与群体行为

当 Agent 个体的信念和目标都与其他 Agent 的意识有关,而该 Agent 个体的行为正是基于这样的信念和目标时,其行为才是群体行为.即一个 Agent 的行为是否属于群体行为,不是根据行为本身的外部描述,而是根据该行为是否涉及并影响到多个 Agent.除此以外,只关于某个 Agent 或涉及其他 Agent 但并不对其造成影响的行为均属个体行为.算法需要区分个体行为与群体行为,以有效地进行不同层次和复杂度的学习.根据行为的目标指向,群体行为还可分为合作行为与对抗行为.

(2) Agent 之间的差异和交互

在分布式开放环境中,涉及的多个 Agent 往往是异构的、动态的、不可预测的,这些特征使多 Agent 的合作变得十分困难和复杂.由于系统的总体性能和工作效率与各自治 Agent 之间的分工、协作有密切的关系,学习过程中必须考虑 Agent 之间的差异性和协调控制的分散性.并不是多个独立的 Agent 合在一起就能组成一个系统,系统必须是一个有机的整体,其间 Agent 的交互方式由环境条件和 Agent 自身行为模式来确定.

基于局部知识的分散控制是 MAS 中 Agent 之间的一种合作方法,这种方法使 Agent 获得一定的自主性,相对集中式控制而言,增加了灵活性,缓解了控制的瓶颈问题.但如果每个 Agent 的运作受限于局部的和不完整的信息(如局部目标、局部规划),则很难实现全局一致的行为.因此,在各 Agent 中嵌入必要的合作层知识是非常必要的.这些合作层知识包括其它 Agent 的能力、目标、方案、兴趣、行为以及相互依赖信息等.交互的 Agent 通过信息交换,改变它们所处的环境,可以显著地影响其它 Agent 的个体学习.特别当多个 Agent 试图以团队的形式去完成单个 Agent 不能完成的学习任务,即集体学习时,交互更是关键问题.一个有效的学习算法必须考虑 Agent 之间的差异和交互,以更好的分配学习任务,适应动态变化的复杂环境.

(3) 多 Agent 协商与自动协商

协商是 Agent 交互的一种具体形式.由于 MAS 中每个 Agent 都具有自主性,在执行系统任务时仅按照各自的目的、知识和能力进行活动往往会出现矛盾和冲突[2].因此,学习算法必须考虑 Agent 之间的协商和自动协商,使 Agent 能够根据其它 Agent 的信念进行推理,并在交互中进一步学习对方的行为方式,促进其相互了解及合作.体现在具体算法上,学习过程中某个 Agent 在某一时刻执行一次动作后,对环境的影响及从环境得到的反馈就不只与其自身动作有关,还与其它 Agent 的动作和状态有关.

如果不同的 Agent 基于各自的感知信息和推理,对于系统当前的最佳策略动作有不同的见解,就需要协商[2].图 1 显示了 Agent 协商过程及状态转换.开始,Agent 处于等待运行状态,当周围条件满足或接收到其它 Agent 的请求时,它将按照一定的规则与其它 Agent 进行协商,此时 Agent 处于协商状态.协商的结果有两种,一种是不满足执行的条件或规则,Agent 将拒绝参与协商的活动,进入拒绝状态,然后经过一定时间,恢复初始状态;另外一种是 Agent 满足运行条件、符合执行规则,此时它将按照一定的规

则,执行相应的活动,即处于运行状态.在运行过程中,如果遇到异常,Agent 将按照一定的规则,恢复执行:若恢复不成功,则回到初始状态;若恢复成功,则继续运行,直到完成任务.

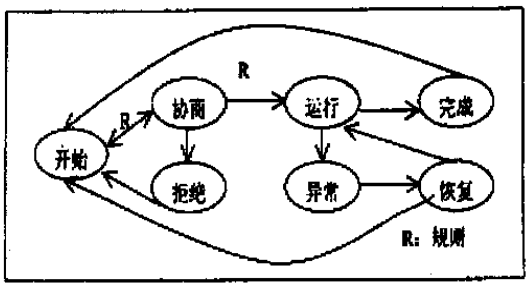

图 1 Agent 协商与状态转换

Fig.1 Agent negotiation and state transformation

3.2 算法实现

本文的混合智能学习算法以增强式学习(RL)为基础.RL 是一种通用学习方法,Agent 从环境感知到状态信息,根据自己的策略选择合适的动作,从而改变环境状态,得到回报或激励,以此表示新的状态对 Agent 而言的功效函数.Agent 的目标是在探索环境的状态空间的基础上,寻找一种策略用于动作的选择,使其能够得到最大可能的回报.

我们的学习系统基于分层思想,Agent 直接使用 RL 学习个体行为,因而以下主要讨论对于群体行为的学习.算法扩展了 RL 思想,在以反馈信息调整决策概率的基础上,考虑并利用了多 Agent 之间的交互关系及其变化,因而是一种动态协作模型.在此模型中,个体最优化概念失去其意义,因为每个 Agent 的回报不仅取决于自身,而且取决于其它 Agent 的动作决策.系统学习目标就是寻找一种策略,可以最大化系统将来获得的总回报.

设 MAS 系统由 n 个 Agent 构成,表示为 MAS = $\{Agent_1, Agent_2, \cdots, Agent_n\}$,系统外部环境状态集记为 E.Agent 用三元组 $<S, A, f>$ 表示,其中 S 为 Agent 状态集,A 为 Agent 控制行为集,f 为 Agent 策略规划函数.$P(s, a, s')$ 为状态转换概率,表示 Agent 在状态 $s \in S$ 执行动作 $a \in A$,转换到状态 $s' \in S$ 的概率.Agent 合作知识即 Agent 之间交互行为的相关性、Agent 个体与总体任务目标之间的相关性,分别用模糊矩阵 $H(t)$、$H'(t)$ 表示,由于二者是时间 t 的矩阵函数,因而可反映 Agent 间交互关系的动态变化以及各 Agent 与不断变化的任务目标间关系的动态变化.在 t 时刻,混合智能学习算法的流程如下:

(1) 随机设定外部环境在 t 时刻的状态 $E(t)$;

(2) $Agent_i$ $(i = 1, 2, \cdots, n)$ 观察环境状态 $E(t)$,并根据 $H(t)$、$H'(t)$ 相应地更新自身状态 $S_i(t)$;

(3) 所有 Agent 并发学习:$Agent_i$ 根据状态转换概率 $P(S_i(t), a_i, s_i')$ 及自身的策略规划函数 f_i 选择控制行为 a_i;

(4) 环境对所有 Agent 的动作产生外部增强信号 $R(t)$;

(5) $Agent_i$ 根据 $H(t)$、$H'(t)$ 得到 $W_i(t)$,它表示 $Agent_i$ 在 t 时刻对于整个系统的权值,且满足

$$\sum_{i=1}^{n} W_i(t) = 1$$

(6) $Agent_i$ 计算自身的有效增强信号 $r_i(t) = W_i(t) * R(t)$,其中 $r_i(t)$ 表示 $Agent_i$ 从状态 $S_i(t)$ 出发,按照策略规划函数 f_i 执行所选动作,在 t 时刻得到的回报;

(7) 更新 $H(t)$、$H'(t)$;

(8) 寻找一种系统策略,使得在此策略下各 Agent 的子策略决策可以使系统获得最大的折扣奖赏和(总回报),即最大化 $\sum_{i=1}^{n} W_i(t) \sum_{j=0}^{\infty} \gamma^j r_i(t)$,其中 γ 是折扣因子,反映了时间的远近对回报的影响程度,一般取略小于 1 的值.

算法采用分层思想,将个体行为与群体行为的学习分开,利用系统中各 Agent 之间的合作知识,通过加权法协调其交互及与环境的相互作用,使 Agent 根据不同的学习重点进行学习,既发挥了它们各自的特长,又使其作为一个整体进行训练,因而同时提高了 Agent 的个体技能和整个系统的性能.

3.3 采用混合智能学习算法的 Agent 体系结构

根据以上分析,把 Agent 体系结构分为感知、通信管理、学习、执行四部分(图 2).Agent 在外部环境当中,进行感知、学习、决策、处理,最终产生控制行为输出.各部分功能如下:

- 感知模块:Agent 对外部环境 E 的感知和建模.
- 通信管理模块:负责命令和消息的发送和接收.
- 学习模块:采用混合智能学习算法(个体学习与群体学习相结合的分层结构),完成主要的学习决策功能.其中,状态信息即感知模块的输出:Agent 对外部环境及自身状态的模型信息;合作知识即 Agent 之间的交互、协商信息、系统的协同工作进展情况等,体现为算法中的 H、H' 矩阵;学习过程使用状态信息及合作知识又不断对其进行更新,产生决策行为.
- 执行模块:执行控制行为,完成最终任务,并

把获得的信息反馈给学习模块 .

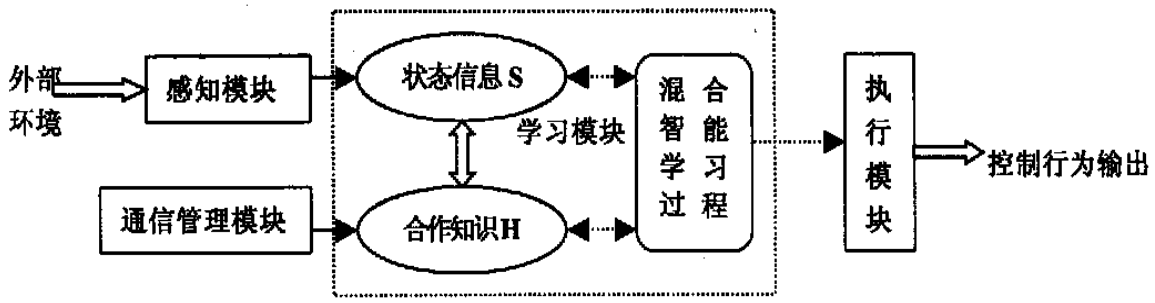

图 2　采用混合智能学习算法的 Agent 体系结构

Fig .2　Agent architecture using hybrid intelligent learning algorithm

4　在足球机器人仿真比赛中的应用 (Application in RoboCup)

机器人足球比赛已成为研究多智能体系统的一个标准的实验平台 , MAS 中 Agent 的学习同样可以运用到该领域 .将基于多 Agent 的混合智能学习算法引入 RoboCup 仿真系统 ,并予以具体实现 .

4 .1　RoboCup 仿真比赛环境的特点

(1) 动态实时性 .要求每个 Agent 在每个仿真周期内根据从 server 接收的各种消息完成所有计算 、决策 ,并将所要执行的命令发送给 server ,否则将失去本次动作执行的机会 ;

(2) 有噪声干扰 .每个 Agent 不能准确地感知和改变环境[3];

(3) 合作与协调 .同一个团队中的所有 Agent 有共同的总体目标 ,又有不同的子目标 ,需要用有效的方法使各 Agent 进行合作 ,解决局部/ 全局目标 、个体/ 总体目标的冲突 ;

(4) 对抗性 .Agent 必须考虑如何与对手争夺球的控制权 ,如何赢得比赛 ;

(5) 通讯受限制 .同一队各 Agent 之间不能直接通信 ,必须通过 server 进行 .

4 .2　算法的具体应用

首先 ,基于分层思想 ,把 Agent 的任务分解成不同级别(图 3) ,以便于应用混合智能学习算法 ,从低到高 :(1) 为个体学习 ,从世界模型习得个人技术 ; (2) (3) 为群体学习 :根据世界模型及已习得的个体技能学习群体行为 ,层次越来越高 .具体来说 ,个体行为主要包括到指定点 、带球 、踢球等 ;合作行为(即团队行为 ,同队 Agent 之间) 包括传球 、接球 ;对抗行为(与对方 Agent 之间) 包括射门 、守门 、攻防 、抢断等 .算法的应用实例 :

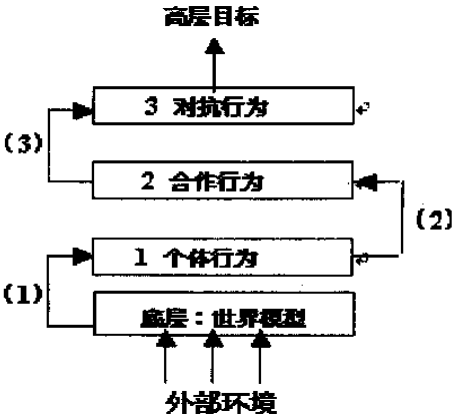

图 3　分层学习任务分解[4]

Fig .3　Task decomposition in layered learning[4]

• 到指定点(Position) :对所有队员

需要考虑避障 、速度和体力的权衡 、是稳定在该点(速度为 0) 还是到该点后仍保持一定速度等很多因素 .由于队员得到的所有视觉信息都是相对的 ,在运动过程中队员对自身以及对目标点的持续定位至关重要 .这是一个动态多障碍环境中的路径规划问题 ,采用个体增强式学习算法 ,首先把 Agent 的二维运动平面离散化 ,Agent 每一个可能出现的位置定义为一个环境状态 .从起始状态开始 ,根据控制策略决定选择哪一个位置作为下一状态 ,通过增强信号调整决策概率 ,最终得到一个到达目标状态的最优(这里为时间最短) 策略 .

• 传球(Pass) :对所有队员

需要考虑传球的对象 、传球路线 、传球时机的选择 .对足球比赛来说 ,传球是最主要的队员合作方式 ,通过传 、接球 ,每个队都想控球在前场 ,以利于该方进球 .传球涉及到场上的整个局势 ,一次传球是否成功 ,很难有一个具体的判断标准 ,因此要综合多种

因素考虑.采用混合智能学习算法,用模糊矩阵 $H(t)$ 表示控球队员和其他队员之间传接球的概率,$H'(t)$ 表示 Agent 个体与总体任务目标(进球)之间的相关性,用 22 个队员和球的位置及速度构成环境状态向量(连续状态空间离散化),所有基本命令组成控制行为集,进行多 Agent 群体行为学习.若传接球成功,则外部增强信号 $R(t)=1$,控球队员及接球队友 Agenti 相应的权值就比较大,其有效增强信号对于系统总回报所占比重也较大,是主要学习者.

• 守门员(Goalie)

观察 统计对方前锋的射门习惯,预测其射门趋势,调整截球点.给出了守门员的四种几何防守策略,并使用混合智能学习算法进行学习和选择,在不同的环境状态下训练守门员.开始使用一个前锋,进行多次练习,每次由系统环境给与增强信号:守门员扑到球表示成功,进球则表示失败.以后逐渐增加进攻的人数,提高防守难度.守门员策略的学习步骤如下:

step 1. 前锋随机放在场地任意位置,守门员放在其本位位置(home-position);

step 2. 前锋带球,伺机准备射门(在其已习得射门技术的基础上);

step 3. 守门员根据当前概率选择一种策略来防守;

step 4. 如果扑到球,则成功,给与正的奖赏;若进球,则给与负的惩罚信号;否则(如球根本未进球门)增强信号为 0;

step 5. 根据反馈信息调整决策概率,重复学习过程,直至找到一个具有最大状态评价函数的策略.

4.3 实验结果

RoboCup 仿真机器人足球赛是在标准软件平台上进行的.针对所提供的 SoccerServer,以 Visual C++ 为工具开发了相应的 Client 程序,每个 Client 就是一个 agent.通过应用基于多 Agent 的混合智能学习算法,使 Agent 学会在多障碍、合作性、对抗性动态环境中有策略地进行比赛,形成系统的合作规划.其在线学习的特点使决策系统具有自适应性,随着学习的进行,控制策略被不断优化.如在训练守门员时,为考察控制策略的优化情况,每经过 100 个学习周期,对防守成功的次数进行统计,根据实验数据得到图 4(横坐标表示在学习中进行统计的时间序号),表明守门员防守成功率随学习时间的增长不断提高.(图中曲线的震荡起伏是由于环境的不确定性及算法的随机性引起的.)

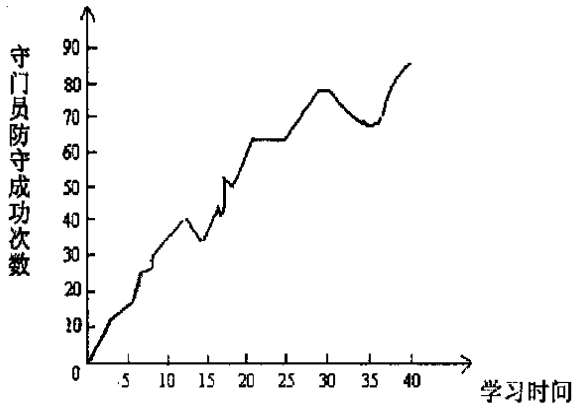

图 4 守门员防守成功次数与学习时间的关系

Fig.4 Relationship between the times of success in goaltending and the learning time

为考察整个系统性能的提高,用未使用此算法的仿真球队 A 与使用算法之后的仿真球队 B 进行了多次比赛,随机选取的 10 场比赛的实验结果如下:

表 1 10 场仿真比赛的实验结果

Table 1 Results of 10 simulated games

比分 \ 场次	1	2	3	4	5	6	7	8	9	10
A: B	3:11	4:12	5:13	4:12	3:14	3:12	2:14	3:13	4:13	3:15

实验数据表明,采用上述学习算法的足球机器人 Agent,其个人技术和团队协作水平都明显提高,整个多 Agent 系统的性能也大大改善,能更好地适应不断变化的足球比赛环境.

5 结论(Conclusion)

机器学习是多 Agent 系统研究的核心问题之一,是复杂智能系统(如足球机器人系统等)的研究热点.本文针对个体学习和群体学习各自的特点和不足,提出基于多 Agent 的混合智能学习算法,提高了 Agent 的个体性能及系统整体的工作效率和智能水平.在 RoboCup 足球机器人仿真系统中的应用和实验结果证明了本学习算法的可行性和有效性.进一步的研究包括算法与 Agent 意图预测、Agent 相似度的结合及细化等.

(下转第 535 页)