

HfutEngine2012 仿真 2D

球队说明文档

吴东飞，王世全，戴双陆，潘健雄，唐怀智，王辉，卢强，代永兴

指导老师：王浩，方宝富

摘要

本文档主要介绍了 HfutEngine2012 球队开发背景、代码架构、以及设计方法。我们添加了一个全新的方法来进行多 Agent 协作技术的系统开发。我们使用的主要方法是基于球队决策的开发。从全局来看，一个拥有全场信息的教练可以没有误差地取得当前周期所有球员的场上信息，这样也就可以及时的更换球员类型使其适应场上的遇到的各种情况。

首先，教练在有噪声的情况下收集数据来大致估计球员智能体类型。然后再根据精确化计算和当前周期的场上情况对比得出球员可能的类型。之后球员智能体经过分析教练信息以及自己看到 server 发送的信息，从其中选择出一个最可靠的信息，根据这些信息来得到执行某一些决策动作的可能性。最后，球员智能体根据决策评估作出最好的决策。

1 球队介绍

HfutEngine2012 成立于 2002 年并且在当年参加了 Robocup2002 中国机器人大赛。在随后几年里，HfutEngine 迅速发展并且参加了各项赛事。从 2005 年起，我们使用了 UVA BASE2003 作为我们的底层代码，并且伴随着 server 版本的更新在其中添加了我们自己的人工智能算法。我们获得了 2007 年 Robocup 中国公开赛仿真 2D 组亚军，2008 年 RoboCup 世界杯仿真 2D 组第 7 名，和 2009 年 RoboCup 中国公开赛仿真 2D 组第 4 名。但是由于球队建设的问题，在 2011 年 Robocup 中国公开赛上没有取得理想的成绩。2012 年是我们第 5 次参加 RoboCup 中国公开赛，我们希望能够在 RoboCup2012 中国公开赛上获得一个更好的成绩。我们希望同其他任何对多 agent 协作技术以及 RoboCup 的科研工作者一起进行探讨。

2 HfutEngine2012 球队框架

在我们的研究中，我们发现任何多 Agent 协作都是基于单个球员智能体对于多 Agent 协作系统的适应性。如果系统中每一个智能体都能够适应这个系统，那么这个系统就是稳定的。没有必要单独的给每一个球员智能体发送固定的指令。为了使每一个球员智能体动作决策都拥有它们自己的价值，我们决定使用价值理论对球员动作进行评估。这样每个球员智能体都会执行所有可执行动作决策中价值最高的动作。

球队框架建立于 2005 年。有效的评价系统的使用保证了球员智能体在可执行的一系列动作中执行价值更高的评估动作。从这个角度看，评估决策主要由经验学习的方法建立。现在我们的动作决策主要依赖于输入评价系统的可执行动作，得到评价系统根据场上情况得出的归一化的评估价值来执行动作。

球队基本框架如如图 1:

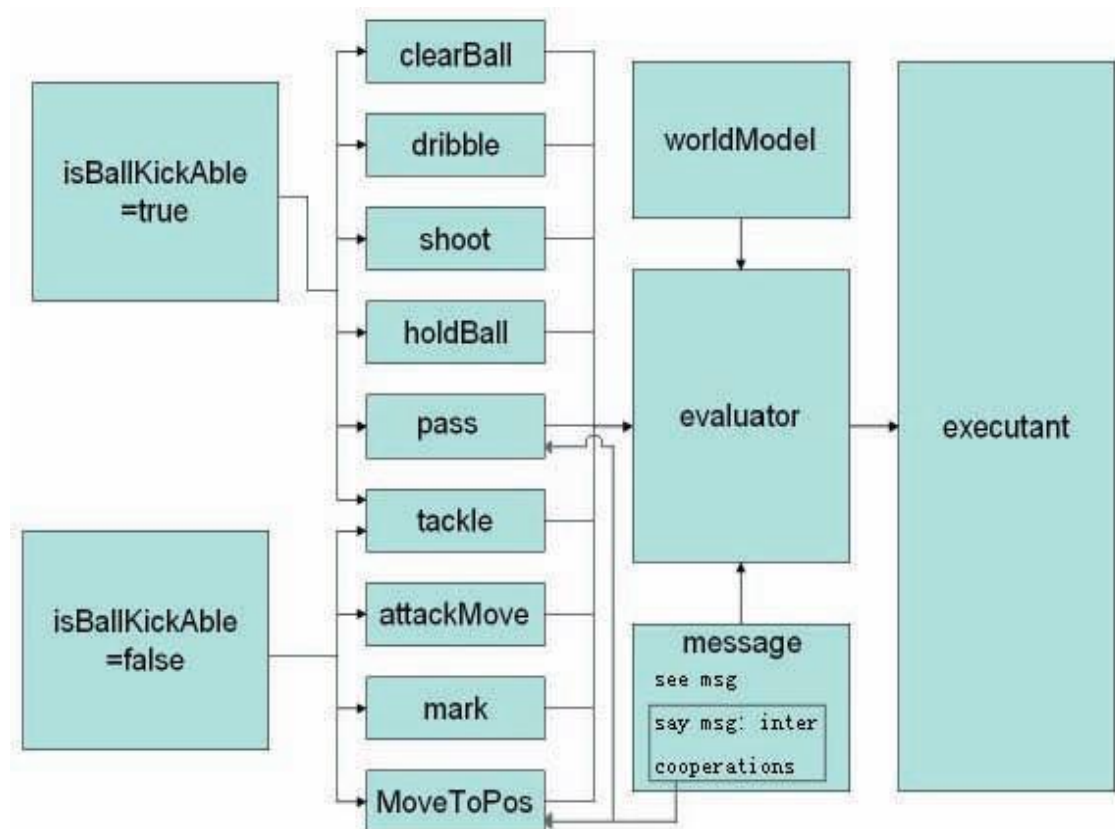


图 1 HfutEngine2012 球队框架

3 HfutEngine2012 高层决策

HfutEngine2012 高层决策共有两个部分，评估和执行。评估部分主要对选择出的可执行动作进行综合评价，从而得出一个最高价值的动作决策执行。执行部分主要负责如何组合完成评估部分所选择出来的动作。

首先，在线教练从 server 中获取信息，然后进行对我方球员智能体建模，我们使用 $\times 2$ 值来评定其球员类型。同时教练在比赛过程中发送新的球员类型给 server。球员从 server 中取得新的球员类型并使用 Q 学习的方法决定每一个个体的球员类型。最后，教练再把新的球场信息返回给球员，图 2 显示了整个球员异构的过程。

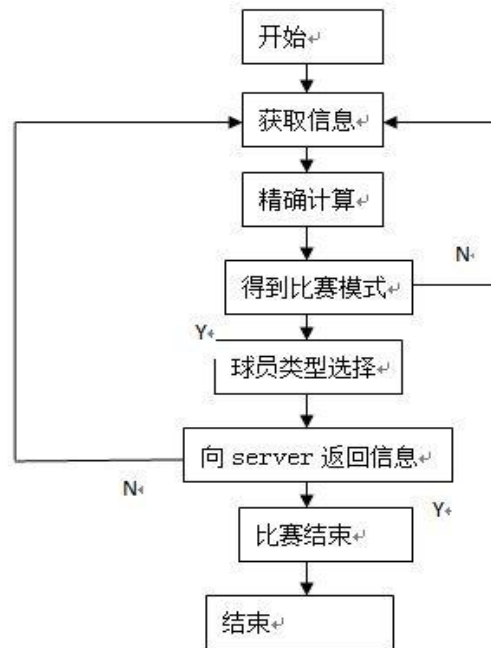


图 2 球员异构过程

4 强化学习

强化学习[Sutton & Barto 1998]是一种无指导的学习，它的主要思想是“与环境交互（Interaction with Environment）”和“试错（trial-and-error）”。这也是自然界中人类或动物学习的基本途径。当一个婴儿学习说话、学会走路时，并没有一个明确的老师，他的学习都是通过眼睛和其它的感知器官来观察不同的行动会有什么样的结果，观察什么样的行动可以得到自己满意的目标。先行动再观察是与环境交互的基本方式，在交互过程中，特别是开始，肯定会发现很多行动都无法完成目标，也就是在不断试错了。不仅婴儿是这样学习的，它会伴随我们一生，我们学习开车就是这样。学车人要时刻注意车子和环境的情况，要发现自己的不同行为到底会有什么样效果和影响。

强化学习模型

强化学习试图从交互过程中找到一些可以用计算机程序实现学习的方法，并不直接考虑人类或动物是怎么通过交互来学习的。同时强化学习也只考虑有具体目标的学习。为了得到可计算的学习方法，首先就要明确学习过程中的交互方式，图3给出了强化学习的标准模型。

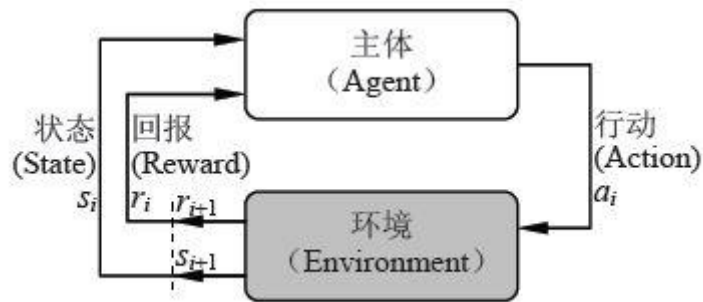


图3 强化学习的标准模型

从模型中可以看到,主体与环境的交互接口包括行动(Action)、回报(Reward)和状态(State)。交互过程可以表述为如下模式:这里的回报,和指导学习(Supervised Learning)中的教师是不同的。在指导学习中,教师要给出很多例子,告诉主体什么情况下,执行什么行动效果最好;在强化学习中,回报只告诉主体当前行动的执行效果,主体要在与环境交互过程中,不断测试每个行动的效果,在长时间的收集回报后,判断出每个行动的长远回报,完成主体的学习。图3 的强化学习模型事实上刻画了一类问题,只要一个问题可以描述成强化学习模型,那么解决这个问题所有方法,都称为强化学习方法。

5 将价值学习理论运用于视觉决策系统

AE 是 Advisor-Evaluator 的简称,早在 2005 年 HfutEngine2D 就提出了 Advisor-Evaluator 的概念,并将其运用于进攻和防守决策中。在实际足球对抗中,以某种模式从大量已知信息中经过分析得出应对方案是机器人足球高层策略设计的关键。

基于 IF-THEN-ELSE 规则的高层策略可以迅速将巨大的状态空间分化,以求得决策方案,再调用对应方案模块以完成细节决策,这种从上而下的决策,输入与输出对应关系明确,速度快,实现简单,曾被很多 RoboCup 队伍使用。但其缺点也很明显,一是决策模块与下层模块耦合性强,一旦下层模块性能改变将会使上层做出较大改动;二是很难找到合适的条件临界值,适应性不强,并不适合复杂多变的 RoboCup 仿真环境。为此,我们提出 Advisor-Evaluator 模型,Advisor 用于提出所有可行策略, Evaluator 则用于评价这些策略并选出最优策略并交付 Method Switch 模块执行。

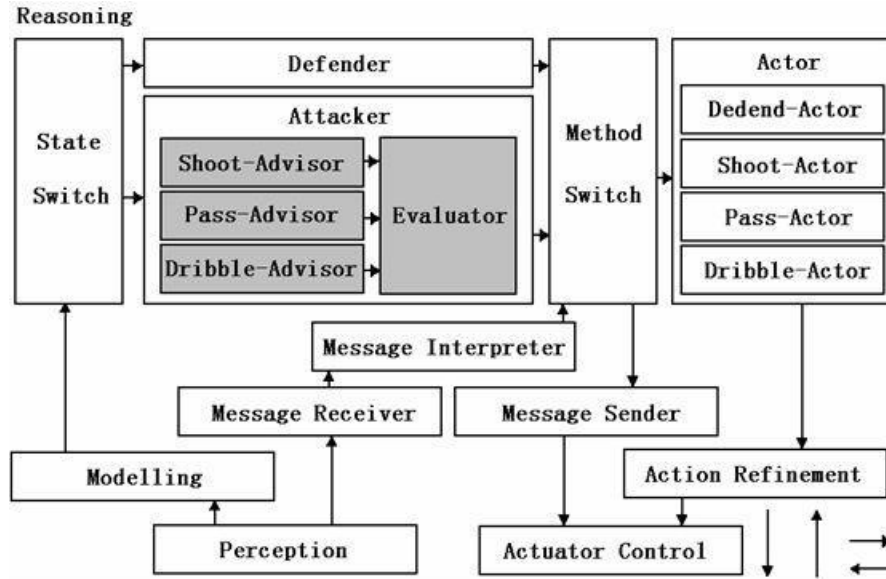


图4 HfutEngine2D 的整体进攻架构

通过研究，我们发现，AE系统在视觉决策方面也有很大的利用价值。参考前面提到的Attack-Advisor和Defend-Advisor，我们在视觉决策系统中也结合了AE系统，建立了AE视觉决策系统。

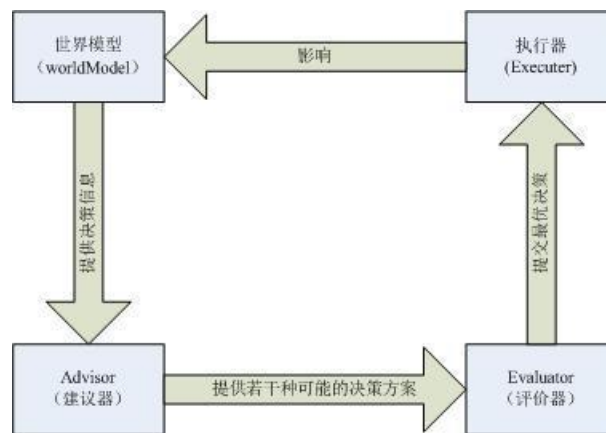


图5 AE系统决策流程图

6 Helios 对抗

HfutEngine2012 使用的仍然是 UVA BASE 2003 开发，UVA BASE 2003 相对 WrightEagleBase 和 Helios Base 来说智能算法较少，没有专门的高层决策提供，如价值理论、人工神经网络函数、以及独立的对手反算机制。对于世界上比较流行的 Helios Base，HfutEngine2012 根据多年积累的经验使用人工方法对球队评价函数进行修正，使其对阵 Helios Base 的时候决策针对性更强，增强 evaluator 对于 Helios Base 决策的针对性。这里给出简单的评价值手工修正方法。

```

if ( 情况 1 )
{
    value +=5;
}
if ( 情况 2 )
{
    value +=15;
}
if (情况 3)
{
    value +=10;
}

```

对于不同的情况 value 变化的大小不同,即体现当前状况下,各种因素对于当前评估值的影响程度大小,HfutEngine2012 通过多次模拟与测试,得到一种针对性更强的、趋于稳定的修正的评价系统。由该系统选择出来的动作将对于 Helios Base 更具有针对性。

5 结论以及展望

HfutEngine2012 在球队评估针对性方向上进行了修改,并且在以后的比赛中还会逐步完善针对 Helios Base 的经验价值修正工作,使评估得出的价值对于 UVA BASE 2003 以及 HeliosBase 的同时具有较强的适应性和稳定性。

6 球队使用底层

UVA BASE 2003

下载地址 <http://staff.science.uva.nl/~jellekok/robocup/>

参考文献

1. Zhulin AN, Jingjing YU, Hao WANG. Robocup Simulation League Goalie Design. Proceedings of 1 Austria Open of Robocup. 2003.
2. Xiaoping Chen, et al, Challenges in Research on Autonomous Robots, Communications of CCF, Vol. 3, No. 12, Dec 2007.
3. Kitano H, Tambe M, Stion P, et al. The Robocup synthetic agent challenge 97[A]. Proceedings of Fifteenth International Joint Conference on Artificial Intelligence (IJCA-97) [C]. Berlin:Springer, 1997. 24-29.

4. Yang LIU, Hao WANG, Baofu FANG, Hongliang Yao. Application of the method of support vector regression in RoboCup. Journal of Hefei University of Technology Vol.30(10). Oct, 2007.
6. Grefenstette, J., Schultz, A. (1994). An evolutionary approach to learning in robots. In Proceedings of the Machine Learning Workshop on Robot Learning, Eleventh International Conference on Machine Learning, New Brunswick, NJ.
7. P. Stone, R. S. Sutton, and G. Kuhlmann Reinforcement learning for RoboCupsoccer keep away Adaptive Behavior, 13(3): 165-188, 2005. __