

6-5-2020

Analysis and Optimization of the Action Chain Mechanism in Agent2D Underlying in RoboCup2D Soccer League

Chen Bing

1. School of Arts and Sciences, Information Engineering University, ZhengZhou City, 450001, China;;

Feifan Xu

2. School of Junior Commanding Officers, Information Engineering University, ZhengZhou City, 450001, China;;

Hanyan Xu

2. School of Junior Commanding Officers, Information Engineering University, ZhengZhou City, 450001, China;;

Zekai Cheng

3. School of Computer Science, AnHui University of Technology, Ma Anshan City, 243002, China;

See next page for additional authors

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research](#), [Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

Analysis and Optimization of the Action Chain Mechanism in Agent2D Underlying in RoboCup2D Soccer League

Abstract

Abstract: In the RoboCup2D soccer league, Agent2D is one of the most widely used underlying team in China. Data transmission noise and the incomplete action chain mechanism make the underlying teams using Agent2D be lack of flexibility. This paper introduces an action correcting parameter and optimizes the operation of the action chain by reinforcement learning mechanism. The performance of the Agent2D underlying team is improved in the game and the adaptability of the team is enhanced. Simulation experiment results show that this method has a certain effect.

Keywords

RoboCup2D, Simulation, action chain, action correcting parameter, reinforcement learning

Authors

Chen Bing, Feifan Xu, Hanyan Xu, Zekai Cheng, and Liu Cheng

Recommended Citation

Chen Bing, Xu Feifan, Xu Hanyan, Cheng Zekai, Liu Cheng. Analysis and Optimization of the Action Chain Mechanism in Agent2D Underlying in RoboCup2D Soccer League[J]. Journal of System Simulation, 2017, 29(11): 2782-2787.

Robocup2D 项目中 Agent2D 底层动作链机制的分析优化

陈冰¹, 许非凡², 徐涵延², 程泽凯³, 刘诚¹(1.信息工程大学理学院, 河南 郑州 450001; 2.信息工程大学指挥军官基础教育学院, 河南 郑州 450001;
3.安徽工业大学计算机学院, 安徽 马鞍山 243002)

摘要: 在 RoboCup2D 仿真足球项目中, Agent2D 是我国使用最为广泛的球队底层之一。仿真平台中数据传输的噪声干扰及代码自身动作链机制不完整等因素, 导致采用 Agent2D 底层的球队在应对不同的队伍时, 存在着适应能力不足的缺点, 影响了球队的整体能力。该论文引入了动作修正参数, 利用强化学习的手段对动作链机制进行优化, 使 Agent 底层球队在面对不同风格的对手时可以选择更加有效的动作执行, 以此来提升球队的适应性。仿真实验证明, 此法具有一定效果。

关键词: RoboCup2D; 仿真; 动作链; 动作修正参数; 强化学习

中图分类号: TP391.9 A 文献标识码: A 文章编号: 1004-731X (2017) 11-2782-06

DOI: 10.16182/j.issn1004731x.joss.201711026

Analysis and Optimization of the Action Chain Mechanism in Agent2D Underlying in RoboCup2D Soccer League

Chen Bing¹, Xu Feifan², Xu Hanyan², Cheng Zekai³, Liu Cheng¹(1. School of Arts and Sciences, Information Engineering University, ZhengZhou City, 450001, China;
2. School of Junior Commanding Officers, Information Engineering University, ZhengZhou City, 450001, China;
3. School of Computer Science, AnHui University of Technology, Ma Anshan City, 243002, China)

Abstract: In the RoboCup2D soccer league, Agent2D is one of the most widely used underlying team in China. Data transmission noise and the incomplete action chain mechanism make the underlying teams using Agent2D be lack of flexibility. This paper introduces an action correcting parameter and optimizes the operation of the action chain by reinforcement learning mechanism. The performance of the Agent2D underlying team is improved in the game and the adaptability of the team is enhanced. Simulation experiment results show that this method has a certain effect.

Keywords: RoboCup2D; Simulation; action chain; action correcting parameter; reinforcement learning

引言

RoboCup 机器人足球世界杯比赛是当今规模最大、水平最高的智能体机器人赛事之一^[1]。其中 RoboCup 2D 仿真项目作为机器人足球世界杯最经

典的项目之一, 是由开发者指派 Agent 球员组成的球队在一个标准比赛软件平台上进行足球比赛。在 RoboCup2D 仿真项目中, 开发者可以把研究重点放在球队的高级功能上, 包括动态不确定环境中的多 Agent 合作、实时推理规划决策、机器学习和策略获取等当前人工智能的热点问题^[2]。正由于不受硬件工程实现的限制, 仿真比赛是机器人世界杯赛中历史最久、参赛队伍最多的一个项目^[3], 其研究步伐不断加快, 为当今机器人智能体技术的发展做出了巨大的贡献^[4]。



收稿日期: 2016-09-24 修回日期: 2016-09-05;
作者简介: 陈冰(1982-), 男, 湖北孝感, 硕士, 讲师, 研究方向为系统分析与集成; 许非凡(1993-), 男, 北京, 本科, 研究方向为仿真系统及实现; 徐涵延(1994-), 男, 贵州, 本科, 研究方向为仿真系统及实现。

<http://www.china-simulation.com>

• 2782 •

1 球队的实现途径与底层局限性

在 RoboCup2D 项目中, 大部分的参赛球队都是由开发者基于已公开的底层代码进行二次开发和优化而得以实现。二次开发的形式主要为针对底层代码的各部分功能进行优化^[5], 就是试图从球员感知、球员动作和球队策略三个方面提升球队整体能力。提升球员的决策能力, 是当今提升球队整体能力的主流思路, 相较于其他两种思路, 通过对球员间协作、自身跑位、罚球战术等策略的设计可以更有有效的提升球队整体的能力。

目前国内应用较广的底层代码主要为以下三种, 即由日本 Helios 队提供的 Agent2D 底层代码^[6]、UVA_Trilean 队提供的 UVA 代码和中国科学技术大学 WrightEagle 队提供的 WrightEagle 底层代码。

国内和国际上的大部分球队都是在这三种底层的基础上做进一步开发、封装, 形成了具有自身特点的底层代码。如基于 Agent2D 底层的 Yushan 底层, 基于 UVA 底层的 GDUT_TiJi 底层等。在 2015 年的 RoboCup 世界杯国际比赛中, 使用自己底层的 WrightEagle 队排名位居第一, Agent2D 底层的 Helios、Yushan 等队分别位列第二和第六, 而使用 UVA 底层的 GDUT 则是排名第十二。

WrightEagle 队伍的底层由于其基本功能简单, 机器学习和策略生成模块并未公开, 导致上手难度大, 除了中国科学技术大学之外很少有球队可以充分发挥 WrightEagle 底层的威力。UVA 底层由于其结构简单, 被认为是最容易上手的底层^[7]。Agent2D 底层由于其基本功能完整, 对球队策略的设计相较其他代码更容易实现, 所以应用范围十分广泛, 本文也是针对 Agent2D 底层代码来进行讨论的^[8]。

2 Agent2D 底层球员及动作链分析

作为使用最为广泛的底层, Agent2D 底层的实现程序包含约 10 万行代码。本文对相关代码进行了研读和分析。

2.1 Agent2D 底层中的球员智能体结构

在 Agent2D 底层设计中, 球员 Agent 具有球场世界模型、感知模型、决策模型三个主要部分^[9]。其中, 球场模型用于保存球员通过感知模型从服务器获取到的信息, 方便决策模型对其进行处理; 感知模型主要从服务器端获取球场的信息, 由于服务器端传递给球员信息的途中会产生噪声干扰, 使得球员所建立的球场世界模型与服务器中实际的球场模型存在误差; 球员的决策模型通过对球员自身世界模型的分析模拟来对将要执行的动作进行决策。如图 1 所示。

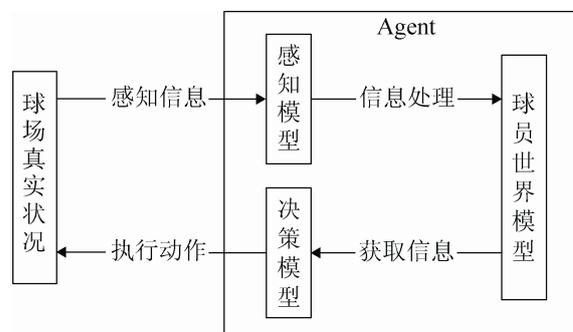


图 1 球员 Agent 结构示意图
Fig. 1 Structure of Player Agent

在决策模型中可以按照球员是否持球分为有球策略和无球策略。

(1) 有球策略

当球员认为自己可以在本周期内踢到球时, 球员就会执行有球策略。该策略主要依据动作链系统进行, 通过动作链系统的分析, 球员可以在事先制定好的动作中选择最符合球队利益的动作执行。

(2) 无球策略

当球员认为自己不能在本周期内踢到球时, 球员就会采取无球策略。该策略主要是针对球员的跑位进行的策略, 在整体的决策中起着辅助球队战术实施的作用。比赛中球员在大部分时间都在执行无球策略。

2.2 球员 Agent 的动作链机制分析

本文所研究的动作链机制即为有球策略的生

成原理, 当球员持有球时, 通过对当前状态的分析和对未来状态的模拟来对动作的执行效果进行评估打分, 最终将最符合球队整体策略的动作遴选出来, 传递至服务器^[10]。动作链在球员动作中的执行流程如图 2 所示。

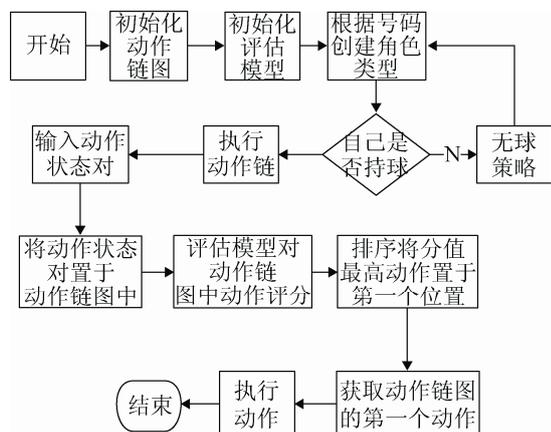


图 2 球员 Agent 决策流程

Fig. 2 Player Agent's Decision Process

动作链的运作主要依靠其中的 3 个模块: 动作推理机、动作链图以及动作链保存器。动作链机制的本质是对球员可执行动作的评估与筛选。

在初始化的过程中, 动作推理机会将每个协作动作的推理机注册进推理机存储容器中; 在动作链的动作生成模块中, 每个推理机都会将当前的状态作为初始状态进行动作的推理, 推理的结果是 1 个已经确定的可执行动作和执行该动作的预期结果。

在评估模块中, 评估模型会依据每个推理机生成动作的预期结果进行评分, 在最基础的评估中通常认为球的位置越靠近地方球门则分数越高, 当产生的执行结果为射门时分数为最高, 这种简单的评估模型可以确保球员在持球时具有基本的进攻意识, 但是较为缺乏灵活性。评估模块最终确定的分数最高动作将会被动作链最终执行。

2.2.1 动作推理机

动作链机制中为每个子动作都设计了推理机 (Generator), 并使用 1 个总的推理机来管理它们, 动作链的总推理机可以认为是 1 个依次执行每个动作推理机的迭代器, 通过虚函数来调用不同动作

的推理机, 动作推理机分为外部的接口部分与内部的功能实现两部分, 外部接口部分主要是提供与动作链图的联接, 从动作链图中获取当前的球场状况作为参数传递到内部, 通过内部的代码对情况进行分析和运算, 最终生成一个确定的动作来返回给动作链图。

2.2.2 动作链图

动作链图(Graph)的用处主要为通过搜索算法找出最终的结果, 起着联接动作链推理机和评估模型的作用。是动作链机制中的主要流程, 其中通过不断的对动作链推理机与评估模型的循环来确定一个进球策略, 然而这个策略是极其简略的, 其主要原因为球员对于球场情况的预测模型精确度不足, 当动作链图计算出的进球周期数距离当前周期越远则精确度越差, 随着周期数不断逼近预计的进球周期, 预测的结果也会随着进行修正, 最终直到我方无人持有球或者是成功进球。

2.2.3 动作链保存器

动作链保存器(Holder)的结构是动作链 3 个部分中最为简单的, 它与 Agent 智能体一起被初始化, 并且对动作链图进行初始化操作, 动作链保存器的主要功能为存储动作链的实例, 它会将每周期生成的动作链作为自身的成员进行保存, 并且进行更新, 以确保动作链在每个周期的正常运行。

2.2.4 动作状态对

动作状态对是动作链中的一种数据结构, 他由 1 个动作和 1 个球场状态两部分组成, 其中动作是球员可执行的协作动作中的一种, 球场状态为指定时刻的球场世界模型, 可以是当前的球场状态也可以是预测的未来状态。在动作链中动作状态对主要用于支持评估模型的计算与分析。

Agent2D 底层中的动作链机制并不完整。其设计者在底层代码中仅留下了 1 个基本的范例以供球队的正常运行。对动作链的优化可以使球队整体水平具有很大的提升空间。

3 基于强化学习的动作链优化

3.1 动作链机制的问题分析与优化设计

通过分析球队底层代码的比赛录像等方式,在代码的动作链机制上可以发现以下 2 个问题:

(1) 目前 Agent2D 底层代码在选择执行动作时考虑的因素较少,导致最后选择的不是最优动作。

(2) 在 RoboCup2D 的比赛中,球队的进攻及防守方式较为单一,对于不同的对手不能及时调整进攻模式。

在当前的动作链模型中,球员选择执行动作主要依据的是动作链的评估结果,而在动作链的评估模型中主要以足球位置为评估参数。一方面,由于服务器中存在噪声干扰,使得动作链评估模型的预测结果和动作的实际执行结果差距较大,导致最终选择的不是最优动作。另一方面,该评估模型中不存在反馈机制,不能根据比赛中各动作的执行效率调整选择执行动作的策略,出现进攻模式单一的问题。

本文拟通过在评估模型中对 Shoot、Pass、dribble、Hold 四个动作添加可变的权值“动作修正参数”,利用增强学习的方法不断调整动作修正参数^[11],增加有效动作的执行频率,从而提升球队整体能力,评估流程如图 3 所示。

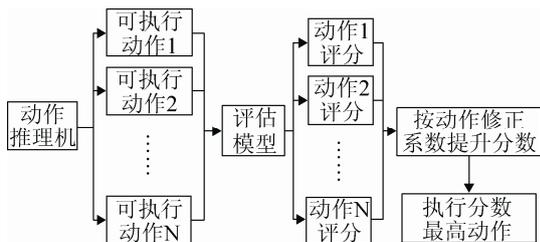


图 3 优化后的动作链评估执行流程

Fig. 3 Optimized Process of Action Chain Evaluation and Implement

动作修正参数是依据动作的执行效率进行调整的。球员的一个“预测—执行”周期结束之后,将动作的执行结果与其预测结果进行比较,如果两者之间的差值在误差的允许范围之内,则认为该动作的执行效率高,我们把这类动作称为有效动作。在下一个“预测—执行”周期中,增大有效动作的

动作修正系数,提高有效动作的最终评分,以此提升有效动作的执行频率。因为对动作链优化的最终目的是提升球队进攻时的传球效率,所以本文以足球位置作为衡量动作是否有效的指标,通过对比动作执行后球的预测位置与实际位置来判断动作的执行是否有效。

3.2 动作链机制优化实现

优化的实现过程主要分为计算动作修正系数和限制动作修正系数增长两部分。传球在 RoboCup2D 仿真环境中是一个非常重要的高级动作,优秀的传球策略在进攻方面意味着能够快速地将球推向前场,形成射门威胁^[12]。因此,下文将以传球动作中的 Pass 动作为例来具体阐述优化的实现过程。

3.2.1 计算动作修正系数

初始状态下,Pass 动作的动作修正系数默认值为 1。首先以动作的执行为起点,如果动作开始执行后,学习过程的持续周期数过长,则说明动作的执行效果与动作链预测系统的预计误差较大,认为此动作在本次对局中效率较低,不对其动作系数进行修正。当在学习过程中存在某一时刻,使得球的位置与动作链系统预测模型所预测的位置相比较,在一个可以允许的误差范围内时,认为在本次比赛中球员自身执行该动作会有较好的效率,对其动作系数进行上升调整。具体计算流程如图 4 所示。

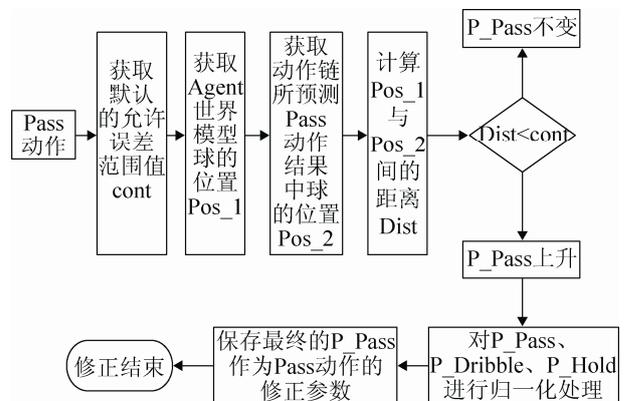


图 4 动作系数修正流程

Fig. 4 Correcting Process of Action Parameter

3.2.2 限制动作修正系数增长

当使用动作修正系数对动作链评估结果进行修正时，为了避免由于某一动作的修正系数过大，导致球队在进攻时仅执行一个动作的情况，对于动作修正系数的增长还需要进行如下限制：

(1) 限制修正系数的增长周期

将动作修正系数的增长限定在某一周期内，防止某一动作修正系数无止境的增长。在具体实现中，对周期数的设定不宜过短，否则会导致动作修正系数的增长不足，导致最终效果不明显。

(2) 限制修正系数的增长速度

在三个动作中，若某一动作的修正系数在初期增长较快，则会最终造成球队在持球时只执行一个动作的现象产生，所以为了抑制这种现象，本文采用了如下公式来抑制动作修正系数的增长速度。

$$sum = pass + dribble + hold$$

$$result = \frac{action}{sum}$$

式中： $action$ 为目标动作的动作修正系数值， $pass$ 、 $dribble$ 、 $hold$ 分别为对应动作的修正系数值， $result$ 为最终传入动作链评估模型的动作修正系数。通过以上公式的归一化处理，可以在一定程度上解决修正系数增长过快的问题。

4 仿真试验及分析

本文在指定的 Ubuntu 系统环境中搭建平台进行仿真试验运行，以检验动作链优化的效果。由于试验的目的是检验动作链优化对底层性能的影响，故引入对抗的都是排除了策略开发的纯底层代码。试验根据胜负情况来判断本文优化方法对球队自身能力提升的有效性。

第一轮试验中采用的球队分别为未经优化的 Agent2D-3.1.0 纯底层与优化过后的 Agent2D-3.1.0 纯底层代码，进行 100 场的比赛对抗测试，这样更有助于对比区分双方动作执行次数上的变化；第二轮试验则是采用 YuShan、UVA、WrightEagle 等球队的无策略纯底层代码作为测试队伍，与优化前后

的 Agent2D-3.1.0 纯底层代码分别进行 20 场的比赛对抗测试，目的是控制变量，防止球队的策略对试验的结果产生影响。

第一轮试验的结果统计如表 1 所示。

表 1 优化前后 Agent 底层球队对抗结果统计
Tab. 1 Comparison between the Optimized and Original Agent Underlying

| | 未经优化的 Agent2D-3.1.0 | 优化后的 Agent2D-3.1.0 |
|------|------------------------|-----------------------|
| 获胜场次 | 37 | 41 |

从上述 100 场测试中抽取了部分比赛录像进行分析，发现优化后球队在与战术策略相同的对手进行比赛时每个球员各个动作的权值增长较为平均，导致了双方最终在行动方式上并没有发生太大变化。但是最终优化后的球队在胜率上要略高于优化前球队，说明本次优化具有一定的有效性。

第二轮试验的结果统计如表 2 所示：

表 2 优化前后 Agent 底层与各球队底层对抗结果统计
Tab. 2 Comparison of Optimized and Original Results with Different Teams' Underlying

| 测试队伍 | YuShan | UVA | WrightEagle |
|-----------------------|--------|------|-------------|
| 未经优化的 Agent2D 获胜场次 | 2 | 3 | 3 |
| 优化后的 Agent2D 获胜场次 | 4 | 4 | 3 |
| 获胜场数增加率/% | 100 | 33.3 | 0 |

从第二轮实验结果可以看出优化后的 Agent 底层在胜场上与 YuShan 和 UVA 队伍的底层代码进行对抗时胜率都比优化前有所提升，但是在与 WrightEagle 对抗时出现了胜率无明显变化的情况。

优化后队伍的胜利场次情况比优化前有所提升，但是并不明显，可能产生这种现象的原因在于 Agent 底层代码的基本策略较为薄弱，在对手球队的协作十分紧密的情况下很难执行出有效的动作，导致无法触发动作修正系数的修改条件。在第二轮实验出现的 WrightEagle 队底层代码就具有相对更强的应变能力，因此前后两次的测试结果在胜场上并没有产生变化；相比较之下，YuShan 队的强大

在于其策略的丰富, 在不存在策略的 YuShan 队底层代码的默认策略就会变得十分单调; UVA 队的底层可以说强度正好处于 Yushan 与 WrightEagle 之间, 相对 YuShan 底层会采取更丰富的进攻防守模式, 但也无法像 WrightEagle 具有的学习应变能力。

在测试中也充分体现出了, 进行了优化的队伍可以更好的应付策略较为单调的球队。

5 结论

在 RoboCup2D 项目中, 球队的整体实力是由球队的进攻及防守战术、球队的阵型、球员执行动作的效率等诸多要素所综合决定的。多要素之间的关系错综复杂。本文所提出的基于增强学习的参数修正方法, 能够对底层代码的动作链起到一定优化作用, 可以作为一种解决 Agent2D 底层对不同风格对手适应能力不足的思路。后续研究中, 可以借鉴 YuShan 等强队采用的对 .rcl 与 .rcg 文件做数据解析的方法^[13], 结合统计学手段, 优化有效动作的衡量指标选取, 结合球队的综合评价体系^[14], 从而更精确的对动作修正参数进行建模和计算。

参考文献:

- [1] 赵发军. RoboCup 仿真 2D 系统的研究[D]. 安徽大学, 2013.
ZHAO Fajun. The research of RoboCup simulation 2D system[D]. Anhui University. 2013.
- [2] 李实, 徐旭明, 叶棒, 等. 国际机器人足球比赛及其相关技术[J]. 机器人 ROBOT, 2000, 9(22): 420-426.
LI Shi, XU Xu-ming, YE Bang, et al. International Robot Soccer Tournament and correlative technique [J]. ROBOT. 2000, 9(22): 420-426.
- [3] 龚剑. 关于高校开展机器人足球比赛的思考[J]. 安徽建筑工业学院学报(自然科学版), 2005(4): 94-96.
GONG Jian. On robotic soccer game in college[J]. Journal of Anhui Institute of Architecture & Industry. 2005(4): 94-96.
- [4] 王婕. 机器人足球仿真比赛的综合评价研究[D]. 广东工业大学, 2011.
Jie Wang. Research of Comprehensive Evaluation Based on Robot Simulation Soccer[D]. Faculty of Automation Guangdong University of Technology, 2011.
- [5] 肖圳杰. RoboCup_2D 仿真比赛中多智能体协作问题的研究[D]. 西安理工大学, 2011.
Zhenjie XIAO. Cooperation Strategy Research of Multi-agents System based on Robocup-2D Soccer Simulation[D]. Xi'an University Of Technology, 2011.
- [6] 李学俊. RoboCup 仿真 2D 实验平台[J]. 实验室研究与探索, 2014:58-61.
LI Xue-jun. An Experimental Platform for RoboCup 2D Simulation[J]. RESEARCH AND EXPLORATION IN LABORATORY, 2014: 58-61.
- [7] 金文文. 基于球员协作关系的 RoboCup 仿真 2D 阵型学习[D]. 安徽工业大学, 2015.
Jin Wenwen. Formation Learning of RoboCup Simulation2D Bases on Player Cooperative[D]. Anhui University Of Technology, 2015.
- [8] 章惠龙, 李龙澍. Q 学习在 RoboCup 前场进攻动作决策中的应用[J]. 计算机工程与应用, 2013(7): 240-242.
ZHANG Huilong, LI Longshu. Application of Q-Learning in local attacking decision[J]. Computer Engineering and Applications, 2013, 49(7): 240-242.
- [9] 秋山英久. ロボカップサッカーシミュレーション 2D リーグ必勝ガイド[M]. 2010.
- [10] 王军. 基于分层强化学习的 RoboCup 3D 球队的设计与实现[D]. 江苏大学, 2006.
Jun Wang. Design and Implement of RoboCup 3D Simulation Team Bases on Layered Learning[D]. Jiangsu University, 2003.
- [11] 孙清. 基于强化学习的多智能体协同机制研究[D]. 浙江工业大学, 2015.
Qing Sun. Research of Multi-agent Cooperation Mechanism Based on Reinforcement Learning[D]. College of Information Engineering Zhejiang University of Technology, 2015.
- [12] 章惠龙. RoboCup 仿真 2D 中的 Agent 智能决策系统[D]. 安徽大学, 2012.
ZHANG Huilong. The Agent Intelligent Decision System in RoboCup 2D Simulation League[D]. Anhui University. 2012.
- [13] 宋园, 刘乾, 王灿, 等. RoboCup2D 日志文件数据挖掘研究及应用[J]. 大庆师范学院学报, 2015, 6: 31-34.
SONG Yuan, LIU Qian, WANG Can, et al Research and Application on RoboCup2D Log Data Digging. JOURNAL OF DAQINGN ORMAL UNIVERSITY, 2015, 6: 31-34.
- [14] Bing C, Kaixuan L, Zijian Z, et al. The establishment and implement of evaluation index system in RoboCup 2D Simulation Soccer Platform[C]//35th Chinese Control Conference (CCC), 2016.