

硕士学位论文

小型足球机器人决策系统的设计和实现

Design and Implementation of the Decision System in Small Size Robot Soccer

作者姓名: 张 树 林
学科、专业: 计算机应用技术
学 号: 20409353
指导教师: 冯 林 教授
完 成 日 期: 2006 年 11 月 28 日

大连理工大学

Dalian University of Technology

独创性说明

作者郑重声明：本硕士学位论文是我个人在导师指导下进行的研究工作及取得研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得大连理工大学或者其他单位的学位或证书所使用过的材料。与我一同工作的同志对本研究所做的贡献均已在论文中做了明确的说明并表示了谢意。

作者签名：张树林 日期：2006.12.18

摘 要

机器人足球比赛是近年来迅速发展起来的一项科技竞赛，主要涉及精密机械、机器人技术、自动控制、感知与融合、通信、计算机视觉与图像处理、多 Agent、推理与决策以及机器学习等多个相关领域，是研究机器人技术和人工智能理论的良好实验平台。

本文以 RoboCup 小型组机器人足球比赛为研究平台，应用多 Agent 及强化学习理论，融合网络通信、串口通信和无线通信技术，设计并实现了一个小型足球机器人决策系统，解决了视觉信息传输、世界模型处理及防守决策等问题。具体内容包括：

首先，介绍了机器人足球的概况。将机器人足球队看作一个典型的多 Agent 系统，其中的协作和对抗问题是多 Agent 系统的研究热点，绪论部分对于 Agent 理论及多 Agent 系统中的强化学习做了简单介绍。

其次，详细描述了 RoboCup 小型组比赛的情况。包括小型足球机器人系统的组成结构，不同组成部分在整个系统中的作用，各个组成部分需要注意的问题以及相关的解决方案。

然后，本文设计和实现了一个小型足球机器人决策系统。在该系统中，本文提出并实现了一种双路实时收发网络消息的视觉传输模式和一种事件触发的高效决策机制，有效解决了两个摄像机并行工作所带来的效率和同步问题。同时，决策系统中还对世界模型和上层战术中需要注意的一些问题做了讨论。

最后，针对防守战术中的一对一盯人防守问题，本文应用基于 Markov 对策的强化学习，实现了一个防守策略，并在实验和实战中验证了该策略的有效性。

本文实现的系统作为小型组球队 DUT Fantasia SmallSize 的决策系统参加了 2006 年 RoboCup 中国公开赛，在比赛中，该决策系统取得了较好的实战效果。

关键词：RoboCup 小型组比赛；决策系统；事件触发；强化学习

Design and Implementation of the Decision System in Small Size Robot Soccer

Abstract

Robots' playing soccer games has been dramatically developed as a technical competition in recent years. It is a interdisciplinary field of different subjects, such as exact-mechanism, robotics, automated control, sensor and fusion, communication, computer vision and image processing, multi agent, reasoning and decision, machine learning, etc. It is an ideal test bed of robotics and artificial intelligence.

This thesis focuses on RoboCup Small Size League to design and implement the decision system of a small-size robot soccer team based on multi agent and reinforcement learning, which involves network communication, serial communication, wireless communication, etc. The decision system works out some problems in vision information transmission, world model processing and defense strategy. Detail content as follows:

First, it introduces the RoboCup soccer games. A robot soccer team can be treated as a multi agent system in which the cooperation and adversarial problems are hotspots and the first part introduces multi agent system (MAS) and reinforcement learning in MAS.

Second, it describes the parts that a small-size robot soccer team composed of in detail and the function of every part. Also, it discusses some solution of problems in the parts.

Then, it designs and implements the decision system of a small-size robot soccer team. In the implementation, it presents and implements a double-channel real time network communication model of vision information and an efficient decision mechanism with event trigger. Meanwhile, it discusses some questions in world model and strategy.

In the end, it presents a solution of resolving one-vs-one defense problem and implements a defense strategy using reinforcement learning based on Markov games. The strategy is proved effective in some experiments and games..

The system described in the thesis was integrated in the small-size robot soccer team DUT Fantasia SmallSize which participated RoboCup China Open 2006. In the games, the system worked well.

Key Words : RoboCup Small Size League ; Decision System ; Event Trigger ; Reinforcement Learning

目 录

摘 要.....	I
Abstract.....	III
1 绪论.....	1
1.1 研究背景.....	1
1.1.1 机器人足球比赛的起源和意义.....	1
1.1.2 机器人足球比赛的种类.....	2
1.1.3 小型组足球机器人系统.....	3
1.1.4 多 Agent 系统介绍.....	3
1.1.5 多 Agent 系统中的强化学习.....	5
1.2 本文主要工作.....	6
1.3 本文结构.....	6
2 RoboCup 小型组机器人足球比赛.....	7
2.1 RoboCup 小型组简介.....	7
2.2 RoboCup 小型组足球机器人系统的组成和结构.....	7
2.2.1 全局视觉系统和局部视觉系统.....	7
2.2.2 小型足球机器人系统的总体结构.....	9
2.2.3 视觉子系统.....	9
2.2.4 决策子系统.....	10
2.2.5 控制子系统.....	11
2.2.6 机械子系统.....	12
2.2.7 通讯子系统.....	13
2.3 RoboCup 小型组机器人足球比赛的现状.....	14
2.3.1 国际水平.....	14
2.3.2 国内水平.....	14
3 DUT Fantasia SmallSize 决策系统的设计与实现.....	16
3.1 决策子系统与其它子系统的交互.....	16
3.1.1 裁判盒.....	16
3.1.2 视觉子系统.....	17
3.1.3 无线收发装置.....	19
3.2 决策子系统的框架设计与实现.....	19
3.2.1 主决策触发模式.....	19
3.2.2 总体的程序流程.....	20

3.2.3 与视觉子系统的网络通信模块.....	22
3.2.4 与裁判盒的通信模块.....	24
3.2.5 无线通信模块.....	24
3.3 主决策模块详细设计与实现.....	24
3.3.1 场地信息的更新.....	24
3.3.2 世界模型更新.....	24
3.3.3 比赛状态判断.....	25
3.3.4 底层技术模块.....	26
3.3.5 上层策略模块.....	28
3.4 物理模型测定.....	28
3.4.1 球的运动模型测定.....	29
3.4.2 小车的运动模型测定.....	32
3.4.3 小车的带球和击球特性测定.....	35
3.5 战术设计与实现.....	35
3.5.1 进攻.....	35
3.5.2 防守.....	37
3.5.3 守门员.....	38
4 基于 Markov 对策的多 Agent 强化学习算法.....	40
4.1 多 Agent 强化学习.....	40
4.1.1 强化学习的基本原理.....	40
4.1.2 一种强化学习算法——Q 学习.....	42
4.2 Markov 对策学习框架.....	43
4.3 RoboCup 一对一盯人防守中的学习模型.....	44
4.3.1 一对一盯人防守的环境建模.....	44
4.3.2 基于 Markov 对策的学习算法.....	45
4.3.3 实验及实战效果.....	46
结 论.....	49
参 考 文 献.....	50
攻读硕士学位期间发表学术论文情况.....	53
致 谢.....	55
大连理工大学学位论文版权使用授权书.....	56

1 绪论

过去 50 年中,人工智能研究的主要问题是“单主体静态可预测环境中的问题求解”,其标准问题是国际象棋人——机对抗赛;未来 50 年中,人工智能的主要问题是“多主体动态不可预测环境中的问题求解”,其标准问题是足球的机——机对抗赛和人——机对抗赛。在这样的背景下,机器人足球比赛诞生并逐渐发展起来。

1.1 研究背景

1.1.1 机器人足球比赛的起源和意义

机器人足球的最初想法是由加拿大不列颠哥伦比亚大学的 Alan Mackworth 教授于 1992 年正式提出的^[1]。日本学者立即对这一想法进行了系统的调研和可行性分析。1993 年,Minoru Asada(浅田埴)、Hiroaki Kitano(北野宏明)和 Yasuo Kuniyoshi 等著名学者创办了 RoboCup 机器人足球世界杯比赛(Robot World Cup Soccer Games, 简称 RoboCup)。与此同时,一些研究人员开始将机器人足球作为研究课题。隶属于日本政府的电子技术实验室(ETL)的 Itsuki Noda(松原仁)以机器人足球为背景展开多主体系统的研究,日本大坂大学的浅田埴、美国卡内基—梅隆大学的 Veloso 等也相继开展了同类工作。

1997 年,在国际上权威的人工智能系列学术大会——第 15 届国际人工智能联合大会(The 15th International Joint Conference on Artificial Intelligence, 简称 IJCAI-97)上,机器人足球被正式列为人工智能的一项挑战。至此,机器人足球成为人工智能和机器人学新的标准问题^[2]。

目前,国际 RoboCup 联合会是世界上规模最大的、占主导地位的机器人足球国际组织,总部设在瑞士,现有成员国 40 多个。联合会现任主席是国际著名科学家、在 IJCAI-93 大会上获得国际人工智能最高奖——“计算机与思维”大奖的北野宏明。联合会负责世界范围的学术活动和竞赛,包括每年一届的世界杯赛和学术研讨会,并为相关的本科生和研究生教育提供支持(教材、教学软件等)。除国际 RoboCup 联合会之外,还有其他一些国际组织。其中较大的一个是 FIRA,该组织总部设在韩国大田,现有成员国 20 多个,每年举办一次国际性比赛。FIRA 与 RoboCup 的主要区别之一是采用不同的技术规范:FIRA 允许一支球队采用传统的集中控制方式,相当于一支球队中的全体队员受同一个大脑的控制;而 RoboCup 要求必须采用分布式控制方式,相当于每个队员有自己的大脑,因而是一个独立的“自主体”。

从科学研究的观点看,无论是现实世界中的智能机器人或机器人团队(如家用机器人和军用机器人团队),还是网络空间中的软件自主体(如用于网络计算和电子商务的各

种自主软件以及它们组成的“联盟”), 都可以抽象为具有自主性、社会性、反应性和能动性的“自主体”, 即 Agent。由这些 Agent 以及相关的人构成的多 Agent 系统(Multi-Agent Systems, 简称 MAS), 是未来物理和信息世界的一个缩影。其基本问题是 Agent(包括人)之间的协调和交互, 可细分为: Agent 设计、多 Agent 体系结构、Agent 合作和通讯、自动推理、规划、机器学习与知识获取、认识建模、系统生态和进化等一系列专题。这些专题有的是新提出的, 如合作, 有的是过去未能彻底解决并在新的条件下更加复杂化的, 如机器学习。这些问题不解决, 未来社会所需的一些关键性技术就无法得到。值得注意的是, 上述一系列问题中的大多数都在机器人足球中得到了集中的体现。在这个意义下, 将机器人足球作为未来人工智能和机器人学的标准问题是十分恰当的, 而这一研究意义之深远重大, 也是不言而喻的。

1.1.2 机器人足球比赛的种类

RoboCup 机器人比赛共有足球比赛(RoboCup Soccer)、救援比赛(RoboCup Rescue)和青少年比赛(RoboCup Junior)三大类, 其中足球比赛共分为以下几个类别:

(1) 仿真组(Simulation League)

RoboCup 仿真组比赛(RoboCup Soccer Simulation)的运行采用 Server/Client 模式, 通过 UDP/IP 端口通信。在一个仿真周期内, Server 将场上含有噪声的环境信息发送给 Client, Client 从 Server 得到当前的环境信息并进行加工处理建立自己的世界模型, 基于此世界模型, 做出自己的一系列决策, 然后把要执行的命令发送给 Server。比赛时双方各启动 11 个 Client 程序, 作为场上的 11 个队员。

(2) 小型组(Small Size Robot League)

RoboCup 小型组比赛时是双方各 5 个机器人之间的对抗, 属于半自主系统, 即集中视觉, 分布控制。每个小机器人的尺寸限制在 $0.18\text{m} \times 0.18\text{m}$, 高度根据视觉系统的不同限制在 0.15m 或者 0.225m , 场地长度为 $4.9\text{m} \times 3.4\text{m}$ 。

(3) 中型组(Middle Size Robot League)

RoboCup 中型组采用完全自主式控制, 每个机器人有自己的视觉系统, 机器人的尺寸限制为 $0.5\text{m} \times 0.5\text{m}$, 场地长度为 8m 到 12m , 宽度不小于 5m 。比赛有 2 对 2 对抗和 4 对 4 对抗两种赛制。

(4) 四腿组(Sony Legged Robot League)

RoboCup 四腿组采用日本 Sony 公司的机器狗产品 Aibo 作为比赛队员, 该比赛同样是自主式控制, 每个机器狗有自己独立的视觉系统。比赛时每队场上有 4 个机器狗。

(5) 类人组(Humanoid League)

RoboCup 类人组采用每个球队自主开发的人形机器人作为比赛队员，自主式控制，每个机器人有自己独立的视觉系统。比赛有 1 对 1 射门对抗和 2 对 2 射门对抗两种。

1.1.3 小型组足球机器人系统

在 RoboCup 的各个组别的比赛中，小型组比赛是开展比较早的组别之一。小型组比赛涉及到精密机械、自动控制、传感与感知融合、计算机视觉、图像处理、无线通讯以及人工智能等多个领域的知识。

小型组足球机器人系统一般由视觉子系统、决策子系统、控制子系统、机械子系统和通讯子系统五部分组成，其中决策子系统是整个系统的控制核心，是系统智能性的集中体现。决策子系统涉及到多 Agent 系统、实时推理以及强化学习等人工智能相关领域知识。本文将在第三章给出一个小型组足球机器人决策子系统的详细设计和实现方案，并在第四章介绍一个基于 Markov 对策的多 Agent 强化学习算法在决策系统中的应用。

1.1.4 多 Agent 系统介绍

首先，什么是 Agent？由于研究领域不同，对于这个问题，学术界一直没有统一的、严格的定义，对于人工智能领域的研究人员来说，Wooldridge 和 Jennings 的定义最为广泛接受：Agent 是一个计算机系统，有代表用户或其所有者独立动作或与交互的能力，它具有自治性、反应性、预动性及社会能力^[3]。Agent 技术的研究内容主要有以下几方面：

(1) Agent 及多 Agent 系统理论模型

Agent 的理论模型研究工作者从逻辑、行为、心理、社会等角度出发，试图对 Agent 的本质进行描述，从而为 Agent 系统创建和实现奠定理论基础。在 Agent 理论模型中，最具有代表性的 Bratman 提出的信念-愿望-意图(Belief-Desire-Intention, BDI)模型，其中“信念”是指 Agent 对环境的认知和已有知识的掌握，“愿望”指 Agent 期望或者可能达到的目标状态集合，“意图”指 Agent 选择的具体目标或者路线^[4]。

(2) Agent 的体系结构

Agent 的体系结构涉及的是如何用软件或硬件的方式实现所期望得到的 Agent 的特性，如何处理 Agent 理论模型中各元素时间的交互问题等。按照体系结构划分，Agent 可分为慎思式(Deliberative)、反应式(Reactive)和混合式(Hybrid)三类：

① 慎思式

慎思式 Agent 有特定的知识系统，该系统是对环境的描述以及常识性知识的集合，并表现为一个由 Agent 自己维护的符号表示的世界模型，Agent 的决策通过基于模式匹

配和符号操作的逻辑推理来实现。慎思式 Agent 的主要问题在于知识表示和行为的合理性以及模式匹配的复杂性^[5]。

IRMA(the Intelligent Resource-bounded Machine Architecture)是慎思式 Agent 研究的一个例子,它是一个基于 BDI 心智模型的结构。包括以下关键的部件:规划库以及知识的显式表示;用于推理的推理机;一个方法解释分析器用来确定哪些规划可以达到 Agent 的意图;一个时机分析器,监控环境并确定 Agent 将来的选择;一个过滤进程和审查进程,过滤进程用来确定与 Agent 意图一致的行为子集,审查进程从竞争的行为中选择最优的行为^[6]。

② 反应式

与慎思式 Agent 不同,反应式 Agent 没有描述环境和常识性知识的世界模型,它遵循“感知-动作”模型,不具备逻辑推理的功能。一些研究者认为,这种 Agent 的智能取决于感知和行动,Agent 的智能行为只有在与周围环境的交互中才能表现出来,这反应了人工智能中的行为主义思想。反应式 Agent 不会从以往经验中学习,不容易改进它的性能^[7]。

最著名的反应式 Agent 结构是 Rodney Brooks 给出的一种有争议的归类式结构,这种结构有两个明显的特点:第一个特点是 Agent 做决策是通过完成一个任务的行为集合实现的。在 Brooks 的实现方式中,行为模块都是有限态机,前提是假设这些完成任务的模块不包括复杂的符号表示方式,并假定不作任何符号推理,其形式就是“情景-动作”,简单的将感知的输入直接映射成动作;第二个特点是很多行为可以同时触发,Brooks 建议将不同的行为模块组织成归类式等级,把行为组织成层次结构^[8]。

③ 混合式

如果要求 Agent 具有反应行为能力和预动行为能力,要处理这些不同类型的行为,一个明显的分解要包括构造不同的子系统。考虑到慎思式 Agent 和反应式 Agent 各自的特点,构造一种基于这两种结构的混合式 Agent 成为一个可行的解决方案,这种混合式 Agent 既具有较强的灵活性,又有快速的响应性。

Touring 机是混合式 Agent 的一个例子,它包含三个动作的产生层:模型层、规划层、反应层,对 Agent 应该完成什么动作,每一层不断的产生“建议”。反应层或多或少对环境中发生的改变提供了迅速的反应。它用“情景-动作”规则集来实现,就像 Brooks 的归类式结构中的行为一样。这些规则将传感器输入直接映射到执行控制器的输出^[9]。

(3) Agent 的学习

学习是 Agent 的一项重要能力,是近年来 Agent 领域的研究热点之一。对于大多数实际应用,设计者都无法事先描述 Agent 的系统行为和环境所有的状态,也无法事先给

定所有可能发生事件的对策。因此具有较好的学习能力和自适应性,被认为是 Agent 及多 Agent 系统的重要特征之一。

在机器学习范畴,学习分为有监督学习(supervised learning)、无监督学习(unsupervised learning)和强化学习(reinforcement learning)三种类型。其中强化学习通过从非直接的、有延迟的回报中学习来获得一个最优策略(policy),从而选择能够达到目的动作。现在强化学习已经成为 Agent 及机器学习技术的研究热点之一^[10]。

(4) 多 Agent 系统的交互

在多 Agent 系统中存在着不同层次的交互,包括协作、通信以及对抗等。协作可以有部分协作、完全协作以及针对特定任务的协作;通信有完全不受限制的通信以及带宽和内容受限的通信等;对抗有系统之间的对抗以及单个 Agent 之间的对抗等。对于多个 Agent 共同完成一个任务的系统来说,往往还需要一些冲突协调机制来避免由于通信受限而产生的任务或行为冲突^[11]。

多 Agent 系统是计算机科学中比较新的一个分支,从 20 世纪 80 年代才开始研究,而直到 20 世纪 90 年代中才得到广泛的认可。从此以后国际上对这一领域的热情快速增加至少部分地是由于认识到 Agent 是一中合适的软件范例,这种范例为研究大规模分布式开放系统(如 Internet)提供了可能性。尽管多 Agent 系统在探索 Internet 的潜力方面能起到关键作用,但是多 Agent 系统的作用远不止于此。对于理解和构造各种所谓的人工社会系统来说,多 Agent 系统似乎是一个自然的比喻。多 Agent 系统的思想并不局限于某个特定的领域,像在此前出现的对象一样,多 Agent 系统会在许多不同的应用领域中广泛出现。包括工业、电子商务、Internet、软件工程以及仿真环境、娱乐业等领域都有 Agent 技术的成功应用。其中 NASA 在 1998 年 10 月 24 号发射的外层空间 1 号任务(DS1, Deep Space1)是第一个有自治能力、基于 Agent 的控制系统空间探测器^[12,13]。

1.1.5 多 Agent 系统中的强化学习

强化学习解决的是这样的问题:一个能够感知环境的自治 Agent,怎样通过学习选择能够达到其目标的最优动作。其中 Agent 的任务是从一个非直接的、有延迟的回报中学习,以便后续的动作产生最大的累积回报。

多 Agent 强化学习已经成为多 Agent 学习中一个活跃的分支,国内外的许多学者都对其展开了相应的研究。Sutton 研究了时序差分(Temporal Difference, TD)中的预测问题^[14], Barto 等人研究了利用实时动态规划的学习方法^[15], Riedmiller 用前馈神经网络对状态进行泛化^[16]。迄今为止,较有影响的强化学习算法有:TD 算法、Q 学习算法、Sarsa 算法、Actor-Critic 算法、Monte-Carlo 算法及 R 学习算法等^[14-16]。其中 Q 学习是目前应

用比较广泛的学习算法，本文将在第四章介绍的基于 Markov 对策的强化学习采用了一种叫做“极大极小 Q 学习”的算法。

1.2 本文主要工作

本文以小型组机器人足球比赛为研究平台，设计并实现了一个小型足球机器人决策系统，解决了该系统中的视觉信息传输、世界模型处理及防守决策等问题，并针对防守决策中的一对一盯人防守问题，应用基于 Markov 对策的强化学习算法给出解决方案，实现其防守策略。具体内容如下：

简要介绍机器人足球的背景和其中用到的多 Agent 及其强化学习技术，并详细介绍了 RoboCup 小型组比赛的情况。

构建小型足球机器人决策系统，设计并实现了一个融合了网络通信、串口通信、无线通信、视觉信息处理和多 Agent 协作及对抗的决策程序，以该程序作为 AI 决策模块的小型组球队 DUT Fantasia SmallSize 代表大连理工大学参加了 2006 年全国机器人大赛暨 RoboCup2006 中国公开赛。在比赛过程中，本文所实现的决策系统取得了较好的实战效果。

针对小型组比赛防守决策中的一对一防守问题，以小型组仿真平台为实验平台，应用基于 Markov 对策的强化学习算法解决该问题。通过将一对一防守问题看作 Markov 零和对策问题，用极大极小 Q 学习法给出其防守策略。

1.3 本文结构

本文共有四个章节，内容如下：

第一章 绪论。概括介绍了机器人足球的起源和意义，引入小型组机器人足球比赛的研究内容，并对其中的多 Agent 系统及强化学习作了简要介绍。

第二章 详细介绍 RoboCup 小型组机器人足球比赛的情况。包括小型组足球机器人系统的组成和结构，以及小型组机器人足球比赛的国内外研究现状。

第三章 设计并实现了 RoboCup 小型组足球机器人球队 DUT Fantasia SmallSize 的决策子系统。包括决策系统与其他子系统的接口、各模块的详细设计以及相关战术的实现等。

第四章 根据 DUT Fantasia SmallSize 的战术设计，针对防守战术中的一对一防守问题，应用基于 Markov 对策的强化学习给出其防守策略。并在实验和实战中验证了该防守策略的有效性。

结论部分总结了本文所做的工作以及取得的成果，并给出了本文的不足和对于今后工作的展望。

2 RoboCup 小型组机器人足球比赛

2.1 RoboCup 小型组简介

RoboCup 小型组机器人足球的比赛项目伴随着 RoboCup 组织的诞生而发展起来。1997 年, RoboCup 官方组委会在日本举办了第一届足球机器人的国际比赛, 共有 4 支队伍报名参加了小型组比赛。从这以后, 在短短的 10 年时间里, RoboCup 小型组机器人足球比赛得到迅猛的发展, 参赛队伍不断增加, 比赛水平逐渐提高。目前小型组比赛已经成为 RoboCup 足球比赛中最具观赏性的比赛之一。在 2006 年德国不来梅举行的第十届 RoboCup 世界杯中, 共有来自 11 个国家的 20 支球队参加了小型组的比赛。

RoboCup 小型组机器人足球的比赛场地是长宽分别为 4.9m 和 3.4m 的绿色的、平坦的毛毡地毯。地毯表面必须水平, 坚硬。场地上画有中线、中圈、禁区、任意球点和点球点等。比赛用球是高尔夫球。机器人的体积必须限制在横截面直径小于 0.18m, 如果球队采用全局视觉, 那么机器人小车的高度不得大于 0.15m, 如果采用局部视觉, 则机器人小车的高度不得大于 0.225m。机器人小车可以有带球装置和击球或挑球装置, 但是这些装置不能限制球的自由度。比赛规则与人类足球规则类似, 也有点球、任意球和点球等。机器人如果犯规, 会受到黄牌警告, 如果犯规情节比较严重(如故意冲撞对方机器人小车等), 会被直接罚出场外。比赛分为上下半场, 每个半场的时间是 10 分钟, 两个半场之间休息五分钟。如果比赛中双方的比分差距达到 10 个球, 比赛即自动结束, 领先一方获得胜利。

2.2 RoboCup 小型组足球机器人系统的组成和结构

2.2.1 全局视觉系统和局部视觉系统

根据视觉系统的不同, RoboCup 小型足球机器人系统可以分为全局视觉和局部视觉两种结构类型。

(1) 全局视觉

全局视觉系统将采集场地信息所用的摄像头安装在比赛场地上方的横梁上。每支球队可以有一个或者多个摄像头, 这些摄像头负责采集整个场地的信息并通过图像采集卡将图像信息传送到视觉处理的计算机。视觉计算机对图像进行处理, 提取出可用的视觉信息传送给决策系统, 决策系统做出决策并通过无线通信系统将指令传送给机器人小车, 机器人小车根据指令做出相应的动作。其比赛示意图如图 2.1 所示。

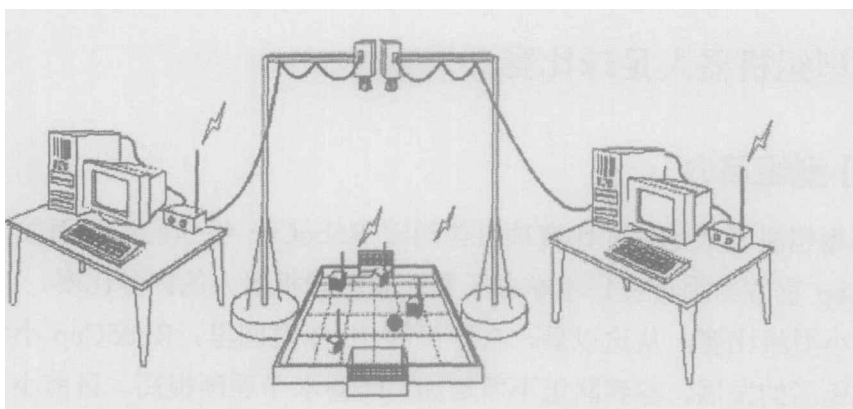


图 2.1 全局视觉小型足球机器人系统示意图

Fig. 2.1 Sketch map of small-size robot system with global vision

(2) 局部视觉

采用局部视觉的小型足球机器人系统不使用安装在场地上方的摄像头，而是使用每个机器人小车上自身配备的摄像头和传感器等装置作为每个机器人自己的视觉系统。机器人小车的视觉系统获得场地信息后通过无线通信系统与其它机器人进行交互达到共享场地信息的目的。同时，决策系统也是完全分布式的，每个机器人小车根据自己所掌握的环境信息以及与其它机器人小车的通信来做出自己的决策和动作。

(3) 两种视觉系统结构的区别：

① 全局视觉系统使用安装在场地上方的摄像头，因此可以快速获得整个场地信息；而局部视觉中，由于摄像头的角度和高度等原因，每个机器人小车的摄像头只能得到一部分场地的信息，不能获得实时的全场信息。

② 全局视觉系统采用的是集中式决策，决策计算机负责统一制定出所有机器人小车的策略，类似“教练”的角色；而局部视觉系统则只能采用分布式决策，因此对机器人小车的智能性和不同机器人小车之间的合作和协调要求很高。如果把每个小车看作一个 Agent，那么这些 Agent 应该是完全自治的并具有较好的学习能力和协作能力。

③ 全局视觉系统中的无线通信是半双工的，即数据传输方向是单一的，且通信量比较小；局部视觉的无线通信则是全双工的，机器人小车之间要互相发送位置及决策信息，通信量比较大。

④ 全局视觉的视觉处理和决策制定都是在上位机完成的，上位机的性能好，运行速度快，视觉处理和决策的效率也就比较高；而局部视觉的结构中，视觉处理和决策过程都是在机器人小车的处理器芯片中完成的，同时机器人小车还要处理大量无线通信，控制自己的运动和带球、击球等装置的动作，因此整个系统的运行速度将大大慢于全局视觉系统。

由于全局视觉系统较局部视觉系统实现简单，系统的运行速度快，整体性能好，因此，目前参加比赛的大多数队伍均采用全局视觉系统。但是由于分布视觉系统中机器人小车的自主性和智能性更好，而且不需要场外的辅助设备，因此是小型足球机器人系统的发展方向。

2.2.2 小型足球机器人系统的总体结构

由于目前大多数球队仍采用全局视觉系统，因此，本文主要介绍使用全局视觉的小型足球机器人系统，为提高系统性能，目前大多数球队采用的都是视觉系统与决策系统分离的结构，其组成结构如图 2.2 所示。

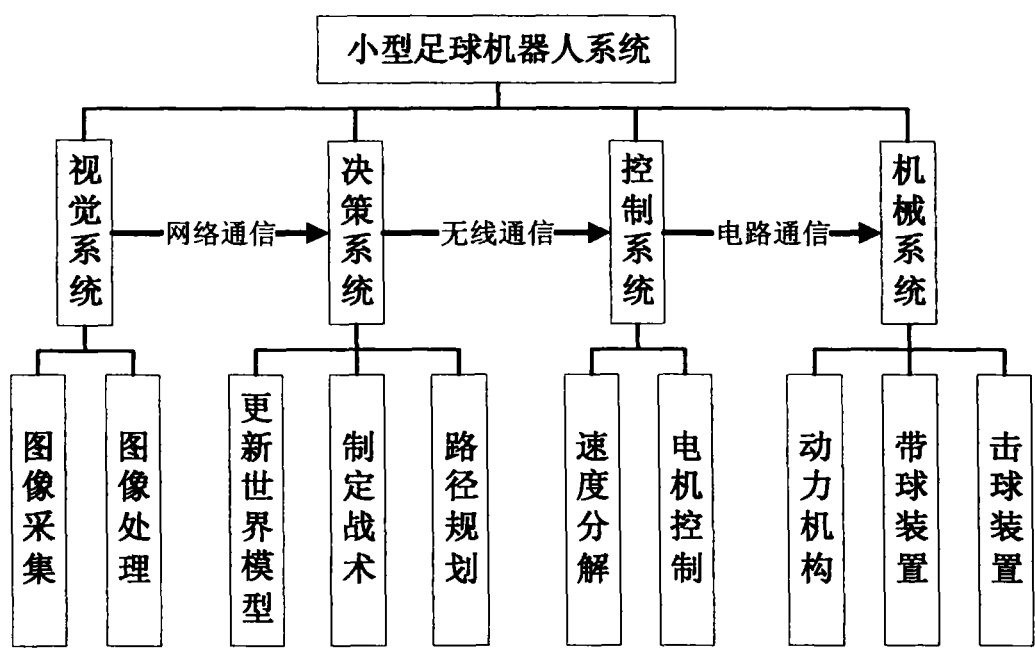


图 2.2 小型组足球机器人系统结构
Fig. 2.2 Structure of the small-size robot soccer system

2.2.3 视觉子系统

视觉子系统是整个系统的信息采集和反馈机构，其主要任务是识别场上的机器人和小球的位置信息，并将这些信息经过处理后提供给决策系统。

视觉是整个足球机器人系统中至关重要的一个组成部分，它是所有其他子系统正常运行的前提和保障，是决定系统总体性能的关键因素。整个小型组足球机器人系统的工作是从视觉子系统开始的，其他子系统都是根据视觉系统提供的视觉信息执行相关的决

策或者动作的,视觉子系统是否能够提供准确的场地信息,关系到整个系统是否能够正确地做出决策和执行相关动作。另外,视觉子系统是实现从图像中恢复真实环境信息,实现二维图像到三维场景变换的复杂的计算机应用技术,涉及到投影几何,物理光学,色度学,数字图像处理,模式识别和计算机图形学的相关学科^[17,18]。

场上位置信息的采集由安装在场地上方固定高度的摄像头完成,每支球队可以有一个或者多个摄像头。对于机器人小车和球的识别都是通过颜色识别和分割来实现的。由于比赛用球是标准的高尔夫球,所以其颜色是基本固定的。双方机器人小车的上方都贴有不同颜色块,称之为色标。正中央的色标颜色是固定的,只有黄色和蓝色两种,是用来区分球队的,其他色标可以在允许的颜色内随意组合。

视觉子系统的工作流程大致如下:

第一步,图像采集。摄像头将三维的场景转换为二维的图像,视觉处理的计算机通过图像采集卡抓取图像,并放入内存缓冲队列中。

第二步,图像处理。对缓冲队列中的每帧图像进行特征提取、颜色分割、重心提取等操作,计算出球的位置以及各个机器人小车的位置和朝向。

第三步,网络通信。如果视觉和决策分布在不同的计算机上,还需要通过网络将处理后的视觉信息传送给决策计算机。

2.2.4 决策子系统

决策子系统是整个系统的控制核心,相当于一个实时指挥场上队员行动的“教练”,他从视觉系统获取视觉信息,根据场上小球的位置、双方机器人的位置和方向等信息,判断场上形势,并做出进攻或者防守的决策,然后根据该决策制定每个机器人小车的动作,最后将该动作转换为机器人小车的运动、带球或者击球指令,通过无线通线传送给机器人小车。决策子系统涉及到多 Agent、实时推理及机器学习等理论,是人工智能在小型足球机器人系统中的集中体现^[19,20]。

一般来讲,决策子系统可以分为以下几个子模块:

① 世界模型。这个模块主要负责对视觉信息进行进一步的加工处理,如给定球的多个连续的位置数据,根据事先测定好的球的运动模型计算出球的速度和加速度等信息;给定本方机器人小车的多个连续的位置和方向数据,根据事先测定好的小车的运动模型计算出本方机器人小车的速度和加速度等信息。维护和更新世界模型的目标就是为决策系统提供尽量详尽和具体的环境信息。

② 裁判盒。这个模块负责接受裁判盒计算机通过串口发送过来的裁判指令,主要包括犯规的判罚和比赛状态的切换等信息。

③ 主决策模块。根据裁判盒指令判断比赛状态，通过世界模型提供的信息分析场上形势，制定出具体的战术决策并给出每个本方机器人小车的目标动作。

④ 路径规划。由于比赛场地中有 10 个实体的机器人小车，因此小车在朝向目标点运动的过程中很可能与其它小车发生碰撞。路径规划的作用就是找到一条通向目标点的无碰撞的最短路径。路径规划是机器人足球比赛中的重要组成部分，也是难点之一，其中许多子问题甚至是 NP-Hard 问题。针对路径规划，国内外的研究人员做了许多工作并取得了很多有价值的成果，例如 Decomposition 方法、Roadmap 方法和 Potential Field 方法等^[21-24]。

2.2.5 控制子系统

这里的控制子系统指的是机器人小车上的电路控制部分，它的功能是接收决策主机发送的无线指令，根据该指令将运动速度进行分解，转换为带动各个车轮转动的电机的转速，并将带球和击球指令转换为相应的电机转速或者螺线管放电动作。

根据控制子系统的任务要求可将其分为以下四个子功能模块：

① 电机及驱动模块

电机及驱动模块的任务包括直流电机的驱动和减速比的确定，确保电机的启动和停止的响应时间能够满足小车运动性能的需要。

② 通讯收发模块

采用无线通信协议，确保小车能够正确地接受到决策主机发来的动作指令。以往的无线通信大都是半双工运行的，即小车上的无线收发模块只负责接受决策主机的数据，而不发送任何数据。近年来，一些球队逐渐开始采用双工通信模式，即无线通信模块在接收数据的同时也向决策主机发送一些有价值的信息，如通过红外线检测到的球的具体位置或者小车当前的电池电量等。

③ 主控制器和数据处理模块

小车电路板上的微控制器需要不断接受和处理来自通讯模块的数据，获取传感器信号，进行控制算法的运算并得到输出给电机驱动器的控制信号。

④ 电源供给模块。

电源供给模块式机器人小车子系统的能源装置，为了保证机器人能够顺利地完比赛，要求电源供给模块能够提供足够的能量。同时为了保证电机能够正常地工作，还要保证电源能够保持尽量稳定的电压。

2.2.6 机械子系统

机械子系统是整个机器人足球系统中的执行机构，是实现各种动作的主体。机械子系统的结构是否合理、性能是否稳定直接决定了整个系统的性能。足球机器人要求反应迅速、运转灵活、运动平稳，因此机器人需要有坚固的车身及合理的结构。

机械子系统一般包括以下三个组成部分：

① 运动机构

运动机构包括电机的安装以及电机与车轮之间的传动等部分。从小型组机器人足球比赛开始到现在，机器人小车经历了 2 轮驱动、3 轮驱动和 4 轮驱动等几个发展历程，其中 4 轮驱动是近年来被广泛采用的方式。车轮大都采用万向轮，有单排和双排两种，其中单排轮的运动性能和精度明显好于双排轮。目前大多数球队采用的都是单排轮 4 轮驱动的运动机构，这种机构的优点是运动性能和运动精度非常好，缺点是控制比较复杂，加工成本高。

② 带球机构

带球机构的任务是保持球与机器人小车的连续接触以达到把球控制在小车周围的目的。由于规则限制，带球机构不能限制球的自由度，因此目前各球队所采用的带球机构大都如图 2.3 所示：带球装置高速向下旋转，与球接触时产生的摩擦力可以使球由于旋转而产生向着车身方向的滚动趋势，这种趋势可以保持球与带球装置一直接触而实现带球动作。

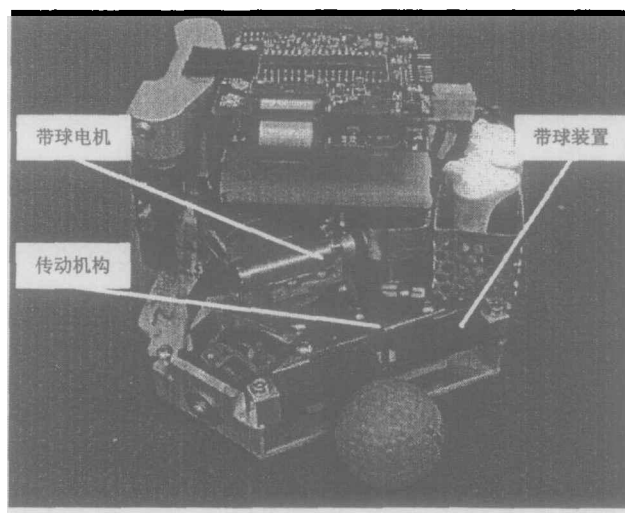


图 2.3 带球机构图
Fig. 2.3 Picture of dribble structure

③ 击球机构

目前大多数球队的击球机构有两种：一种是用来踢低平球的弹射机构，一种是用来踢高球的挑射机构。弹射机构是利用一个可以做直线运动的推杆在向前运动时撞击小球从而使小球产生一个较大的初速度将球踢出，而挑射机构则是利用一个可以自下向上运动一定角度的弹片将球拨向空中。有些球队只有弹射机构而没有挑射机构。

2.2.7 通讯子系统

采用全局视觉集中控制的小型足球机器人系统需要使用无线通信将决策指令发送给小车或者从小车的无线模块接收信息。如果视觉和决策系统分布在两台不同的计算机上，则还需要网络通信来完成视觉与决策子系统的连接。这些都要靠通讯系统来完成。无线通信和网络通信示意图分别如图 2.4 和图 2.5 所示。

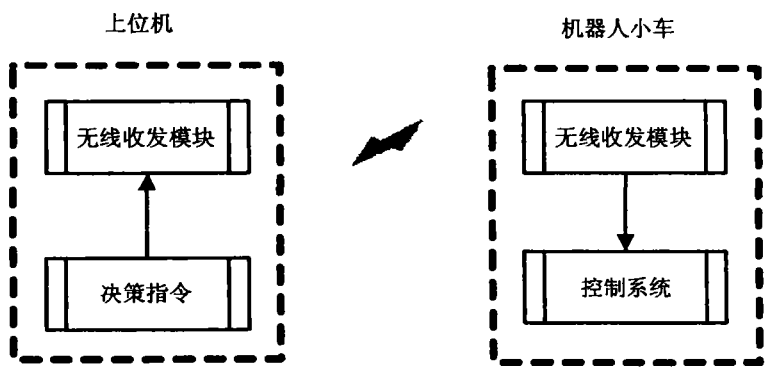


图 2.4 无线通信示意图
Fig. 2.4 Sketch map of wireless communication

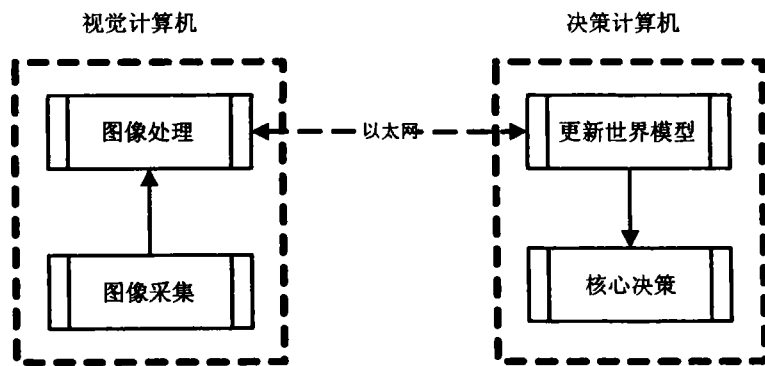


图 2.5 网络通信示意图
Fig. 2.5 Sketch map of network communication

2.3 RoboCup 小型组机器人足球比赛的现状

从 1997 年 8 月在日本名古屋举办的第一届 RoboCup 国际比赛, 到 2006 年 6 月在德国不来梅的 RoboCup 2006 世界杯, RoboCup 机器人足球比赛已经经历了 10 个年头。在这 10 年中, 通过这些不断举行的国际比赛, 各国的参赛选手得到了广泛的交流, 各队的参赛水平得到了不断的提高, 机器人足球比赛也变得越来越精彩和激烈。小型组机器人足球比赛与 RoboCup 一同诞生, 经历了 10 年的发展和进步, 水平得到了很大的提高, 现在已经成为 RoboCup 比赛中最具观赏性的比赛之一。

2.3.1 国际水平

在第一届国际比赛中, 当时参加 RoboCup 小型组比赛的共有四支队伍, 最后获得冠军的是美国 Carnegie Mellon 大学的 CMUnited。CMUnited 采用的是全局视觉系统, 凭借着非常稳定和先进的图像处理算法, CMUnited 以全胜战绩夺冠。值得一提的是, 来自日本的 NAIST 在这次国际比赛中首次使用了局部视觉^[25,26]。

此后, CMUnited 蝉联了 1998 年的世界冠军。而接下来的几年中, 美国 Cornell 大学的 Big Red 表现出众, 分别于 1999、2000、2002、2003 年四次获得世界冠军, 而 2001 年的世界冠军被来自新加坡南洋理工的 LuckyStar-2 获得。2004 和 2005 年, 来自德国 Freie 大学的 FU-Fighters 连续两年获得世界冠军。2006 年, RoboCup 世界杯创建 10 周年之际, 参加小型组比赛的球队达到了 20 支, 而老牌强队 CMDragon(前 CMUnited)再次以超强实力夺冠。此外, 来自澳大利亚 Queensland 大学的 RoboRoos 曾经多次获得世界亚军, 也是一支传统强队。在 10 年的发展历程中, 这些球队的研究和开发人员做了许多研究工作, 在分布式人工智能、模式识别、图像处理、电路控制及机械结构等各个领域取得了很多研究成果^[27]。

2.3.2 国内水平

与国际上这些强队相比, 国内在 RoboCup 小型组机器人足球方面的研究尚处于起步阶段。中科大、浙江大学和上海大学等学校在进行机器人足球相关研究的同时, 都开发了自己的小型足球机器人队伍。其中中科大的蓝鹰小型机器人足球队是国内的传统强队, 曾经多次获得全国比赛的冠军并多次参加了机器人足球世界杯的比赛。近两年, 浙江大学的小型组机器人足球队浙大求是队取得了长足的进步, 该队参加了 2006 年 6 月份在德国举行的机器人足球世界杯并进入了前 8 名, 在 2006 年 10 月结束的 RoboCup 中国公开赛上, 浙大求是队在决赛中输给了世界冠军 CMDragon, 获得亚军。上海大学的上大自强队也参加了 2006 年世界杯并进入了 16 强。近几年, 以中科大、浙大为代表

的国内球队在 RoboCup 小型组机器人足球比赛方面取得了可喜的进步，但与国际水平相比还有很大差距，还有很多方面需要改进和提高。

3 DUT Fantasia SmallSize 决策系统的设计与实现

决策子系统是小型足球机器人系统的核心，也是相关的人工智能理论在整个系统中的集中体现。决策子系统担负着比赛过程中分析场上形势，制定具体决策的任务。本章中，我们将详细介绍小型组足球机器人球队 DUT Fantasia SmallSize 的决策系统的设计与实现。

3.1 决策子系统与其它子系统的交互

在比赛过程中，决策子系统要和视觉子系统、裁判盒以及无线收发装置进行信息交互，考虑到性能问题，我们将视觉子系统部署在一台独立的计算机上。决策计算机与其他设备的信息交互示意图如图 3.1 所示。

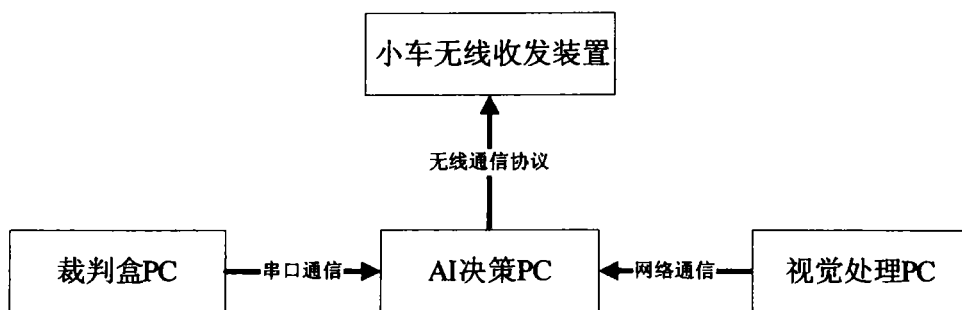


图 3.1 决策子系统信息交互示意图

Fig. 3.1 Sketch map of communication of the decision system

3.1.1 裁判盒

裁判盒是比赛过程中运行在一台独立的计算机上的一个应用程序，这个应用程序可以通过串口向正在进行比赛的每个球队的决策计算机发送裁判指令消息。这些消息的格式和意义是比赛规则中事先定义好的，每支球队按照比赛规则中的约定解析裁判指令消息。小型组机器人足球比赛过程中有个两个人类裁判，主裁判负责裁定场上的进球、犯规和发球等事件，而副裁判则负责操作裁判盒程序，将主裁判的判罚以裁判指令消息的形式发送到每个球队的决策程序。

这个裁判盒程序是比赛组织者提供的，参赛球队必须能够正确解析从裁判盒发送过来的消息，通过决策程序让场上的机器人小车做出正确的反应，否则将被判罚犯规或者被判负。图 3.2 为 2006 年中国公开赛使用的裁判盒程序。

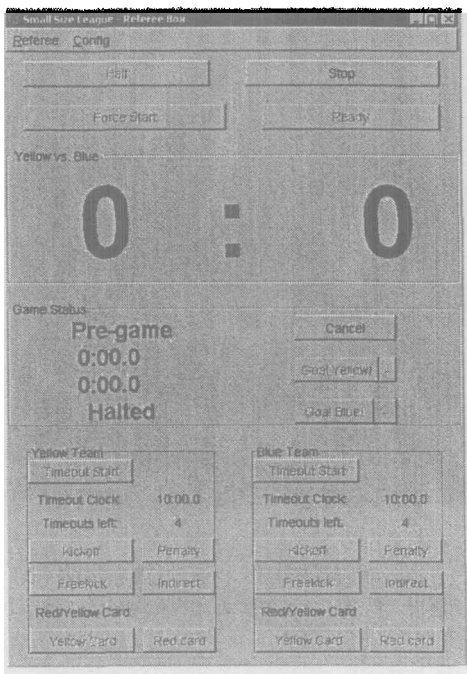


图 3.2 裁判盒的程序界面
Fig. 3.2 Interface of the referee box

3.1.2 视觉子系统

DUT Fantasia SmallSize 的视觉子系统采用两个高速工业摄像头作为比赛时的视觉采集设备，每个摄像头连接一个独立的采集卡。摄像头采用的美国 UNIQ 公司的 UC680 彩色摄像机，分辨率为 659*494，每秒可以拍摄 60 帧图片。采集卡采用加拿大 MATROX 公司的 Meteor II digital 采集卡，采样率可达 40MHZ。考虑到效率问题，我们为每个摄像机配备一块独立的视频采集卡。视觉子系统工作时，每个安装在场地上方的摄像机通过数据线连接到视觉处理计算机的采集卡。每个摄像机负责采集半个场地的信息，然后将图像传到采集卡。视觉处理程序从采集卡缓冲队列中提取场上物体的位置信息，通过网络通信传到决策计算机。图 3.3 为一个摄像机采集的半个场地的画面。

需要注意的是，对于两个图像采集卡的图像提取和处理是在视觉处理程序中的两个独立的线程中进行的，而处理之后的两个半场的视觉信息并未进行任何加工就直接通过网络发送到了决策计算机。也就是说决策子系统每次收到的视觉信息只是某个半场的信息，如何从这些只包含半个场地信息的视觉信息中提取出整个场地的完整信息是决策子系统需要解决的问题之一。

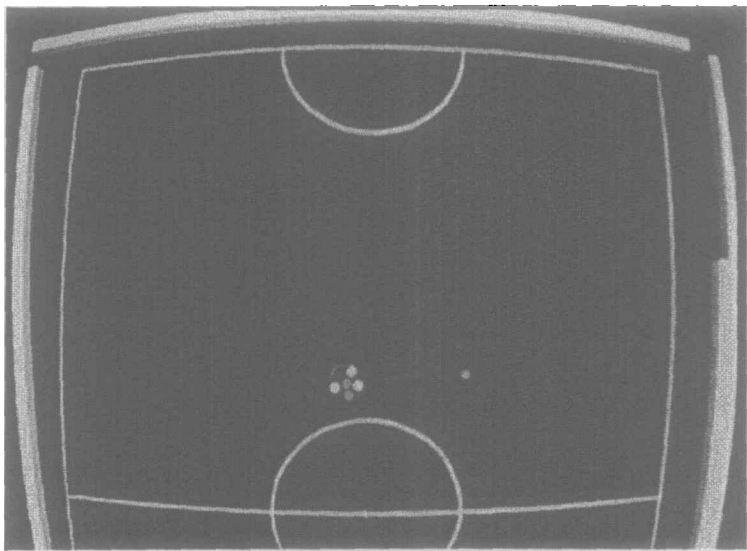


图 3.3 摄像机采集画面
Fig. 3.3 Picture of the single camera

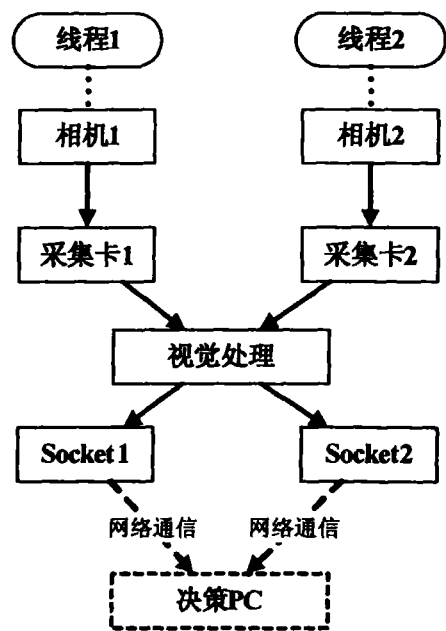


图 3.4 视觉子系统流程图
Fig. 3.4 Flow of vision sub-system

视觉系统每次只发送半个场地信息的原因是：进行视觉处理的计算机要同时采集两路视频信号并进行图像处理和信息提取，这已经给进行视觉处理的计算机的带来了很大

的运行负荷，如果再将半场信息的整合工作放在视觉计算机上执行，势必带来更大的系统开销。而视觉子系统的效率是影响整个系统效率的关键环节，因此应尽可能的将与图像采集无关的工作放到决策子系统的计算机上执行。图 3.4 是视觉子系统的工作流程图，两个独立的线程并行工作，分别处理每个半场的视觉信息。

3.1.3 无线收发装置

决策计算机上通过 USB 接口连接有一个无线收发装置，决策程序最终做出的决策都要以指定格式的消息形式通过无线收发装置发送给机器人小车。每个机器人小车上也有一个相应的无线收发装置，这些装置可以接受决策计算机发送过来的消息，通过解析该消息提取出自己的动作指令。无线收发装置采用的是 NORDIC 的 nRF2401 无线收发器，单芯片，工作频率在 2400MHZ——2524MHZ，最高工作频率可达 1Mbps。

3.2 决策子系统的框架设计与实现

3.2.1 主决策触发模式

比赛过程中，主决策程序需要不断的循环执行，做出决策，直到比赛结束。那么每一次决策何时开始执行，每次执行多少时间是需要考虑的问题，如果不断的连续执行，势必要占用过多的 CPU 资源，不但会做许多无用功，而其还会影响视觉消息和裁判盒消息的接受，从而严重影响整体系统的性能。

考虑到所有的决策都应该根据场上的最新形势做出，最理想的决策执行方式应该是每一次决策都是根据最新一次视觉消息做出，并且该次决策在下一个视觉消息到来之前就已经结束，而在下一个视觉消息到来之前，不再进行第二次决策。这样的决策触发方式既可以保证决策的实时性(根据最新视觉消息立即做出决策)，又可以有效地节省 CPU 的运行时间(每个视觉消息只会触发一次决策)。其视觉消息的到来和决策的执行序列示意图如图 3.5 所示。

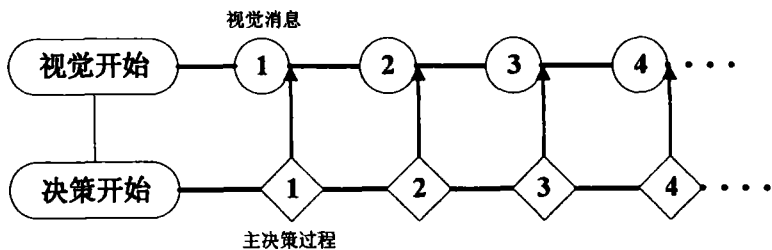


图 3.5 视觉消息和决策执行序列
Fig. 3.5 Sequence of vision message and decision

3.2.2 总体的程序流程

根据 3.2.1 中所示的决策触发模式，似乎可以将视觉消息的接收和主决策放在一个线程里顺序执行。但是 3.1.2 中所介绍的视觉系统发送的消息是两路并行的，而且每个视觉消息只包含半个场地的信息，用于决策的视觉信息则必须是整个场地的完整信息，所以这种用半个场地的信息触发主决策的办法是不可取的。

那么是否可以让半个场地的信息只触发半个场地的决策呢？技术上可以实现这种决策触发方式，但是这种方式同时会带来很多负面效果。首先是主决策循环的执行次数增加了一倍。考虑到每个摄像机每秒可以采集 60 帧图像，每采集两帧图像的时间间隔大概是 16ms。如果视觉处理算法的效率可以达到采集速度的要求且比较稳定，同时不考虑网络延迟，那么主决策模块从每个摄像机获得的两个视觉消息的时间间隔也是 16ms。如果采用这种“半场信息——半场决策”的方式，则在 16ms 的时间内可能要执行两次主决策循环，这对决策算法的效率要求是非常高的。其次这种“半场信息——半场决策”的每一次决策都是基于半个场地的信息做出的，虽然这个决策只会影响到这半个场地的机器人，但是这种信息量的严重缺失将使整个球队表现得像两只互不相关的球队在踢一场足球比赛，这是非常不合理的。因此，“半场信息——半场决策”方式是不可取的。

于是将视觉消息的接收放在两个独立的网络线程中是保持信息完整性的可行方案，但是这种方式下如何触发主决策是需要解决的问题。由于我们的决策子系统运行在 Windows 系统上，因此可以采用 Windows 底层的事件机制来解决这个问题。为每个接收视觉消息的线程创建一个事件，这个事件在接收到视觉消息时变为开启状态，在每一次主决策结束后变为关闭状态。则主决策的任务就是等待这两个事件，当两个事件都变为开启状态时，就开始一次决策过程，这一次决策完成后将两个事件都置为关闭状态。对事件的等待、开启和关闭可以通过 Windows 的 API 函数来实现。这种方案下，如果主决策模块的效率足够高的话，决策过程不会超时，每两次决策之间的时间间隔刚好是摄像机的采样时间间隔，即大约 16ms，我们可以把这个时间看作整个主决策模块的执行周期，它等于摄像机的采样周期。

这种基于事件触发的决策机制较好的解决了两个摄像机同时工作所带来的效率和同步问题。主决策的执行频率与摄像机的采样频率相等，对于摄像机采样频率的利用达到了最优化。基于双路实时收发的视觉消息的事件触发机制有效解决了两个摄像机的同步问题，对于偶发性的网络故障，该机制也能够提供一定鲁棒性。

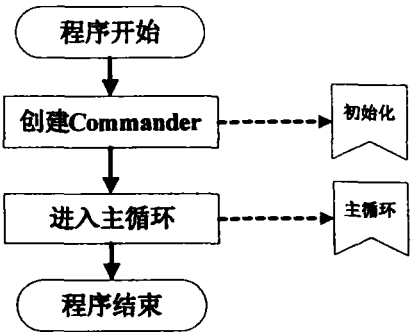


图 3.6 决策子系统总体流程
Fig. 3.6 Whole flow of the decision system

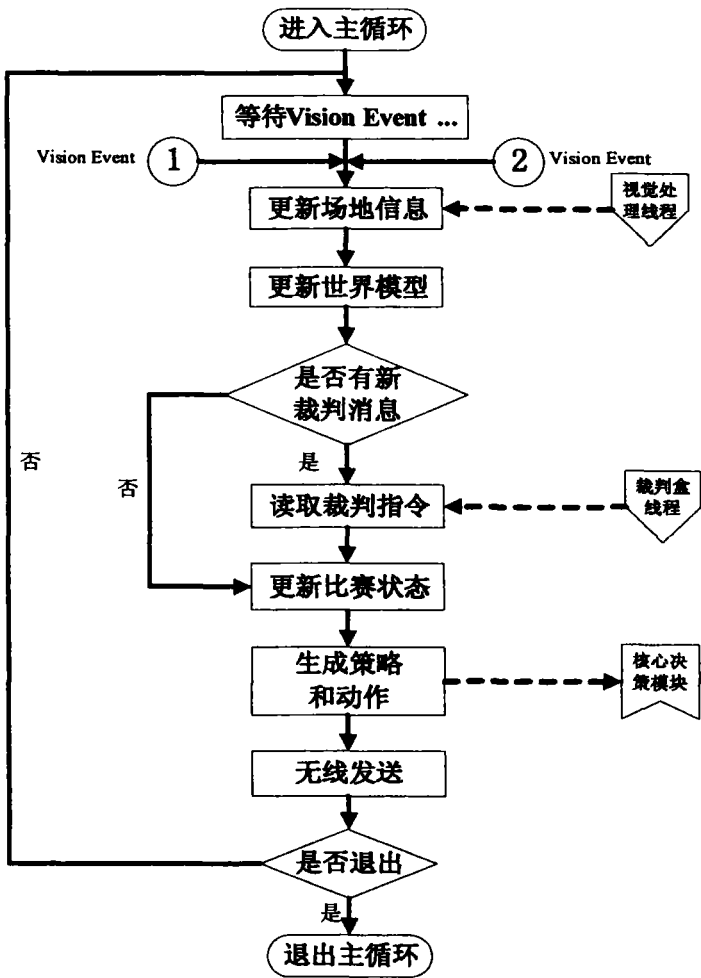


图 3.7 主决策循环流程图
Fig. 3.7 Flow of the main decision loop

除了视觉消息的接收，主决策模块还要根据裁判盒的信息做出正确响应，因此需要一个独立的裁判盒线程来监视与裁判盒计算机连接的串口，每次进行决策之前检查该串口是否收到了裁判盒的消息，如果收到消息，则解析该消息并对比赛状态进行相应的更新，否则正常进行决策。

决策的目的是指挥机器人小车做出正确的动作，因此，每次决策的结果还需要转换成相应的动作指令并将这些指令通过无线模块发送给机器人小车。这个工作可以放到决策主线程中完成。

综上所述，每次决策过程需要做的工作就构成了决策子系统大致的工作流程，如图 3.6 和 3.7 所示。

3.2.3 与视觉子系统的网络通信模块

3.2.2 中讨论的接收视觉信息的方案是采用双线程接收网络消息，在这种网络通信架构中，要求视觉系统在每次视觉处理后都要立即发送一个网络数据包，这个数据包的内容是经过处理的视觉信息。决策子系统在接收到这个数据包后，立即对其进行解析，从中提取视觉信息。我们可以把这种要求称作数据发送和接收的“即时”性要求，这种“即时”性要求是由 3.2.1 中设计的主决策触发模式所决定的，同时也是选择适当的网络通信协议的决定因素之一。

常用的网络通信协议有 TCP 和 UDP 等，TCP 是面向连接的网络通信协议，其特点是可以保证连接和数据传输的可靠性。UDP 是面向数据报的传输协议，不保证数据传输的可靠性，但无需建立连接。对于本系统中的视觉信息传输，UDP 是唯一可行的选择，这是由两种协议的特点决定。

因为 TCP 协议对数据包的组装方式会破坏数据发送和接收的“即时”性。在视觉系统计算机上，负责发送视觉消息的 Socket 如果是基于 TCP 协议的，由于我们每条视觉消息的字节数比较少，Socket 在发送视觉消息时会几个视觉消息累积后组装成一个数据包一起发送出去，而不是收到一个数据消息的发送请求就马上发出去。这种情况导致的后果就是决策子系统的决策周期不等于摄像机的采样周期，而是摄像机采样周期的若干倍。即决策计算机要等待若干个摄像机采样周期才能等到一次视觉消息，而这个视觉消息中包含了摄像机采集的若干次场地信息，这些场地信息中只有最新的信息是有意义的。很明显，这样的视觉消息传递方式非常不合理，它浪费了摄像机的高采样频率。

另外，本系统对信息传输的实时性要求很高，对安全性要求不高。也就是说，经过处理的视觉信息应该立即被传送到决策子系统，否则它就失去了时效性，也就失去了意义，不应该再被传送。由于摄像机的采样频率非常高，丢弃一个视觉消息不会造成过

多损失，但是重新传送一个过时的视觉消息则会导致决策子系统做出完全错误的决策。TCP 协议在一条消息无法成功传送后会不断尝试重新传送此条消息，这种行为不但会使该条消息成为决策子系统错误的视觉信息，还有可能影响后续视觉消息的传送；而 UDP 则只是将视觉消息传送到决策计算机的地址，并不会对决策计算机是否成功接受该消息做任何处理。因此，从这方面考虑，UDP 协议也应该是合理的视觉通信协议。

负责接收视觉消息的网络通信模块收到的每个消息都只包含半个场地的视觉信息，为了得到全场的视觉信息，需要把每个摄像机的视觉消息分别存储到一定的数据结构，然后对这些视觉信息进行整合，提取出全场信息。为了计算球或者机器人小车的速度及加速度等信息，不仅需要这些物体最新的一条位置信息，还需要他们最近一段时间的历史位置信息。如果可以为这些运动物体建立合理的运动模型，不仅可以通过一段时间的位置信息计算出它们当前的速度和加速度，还可以预测它们未来一段时间内的运动趋势，这对主决策模块的策略制定是很有帮助的。

为了接收双路并行传输的视觉消息，决策子系统中需要两个独立的网络线程与相应的视觉子系统视觉发送线程协同工作。每个线程接收到一条新的视觉消息后即将该消息放入一个固定大小的循环队列中。两个视觉消息队列中的每个元素都是一条视觉消息，并且记录了该消息的更新时间，队头总是最新的视觉消息。

以上就是负责接收视觉消息的网络通信模块所需要完成的工作，视觉消息的接收和存储过程如图 3.8 所示。

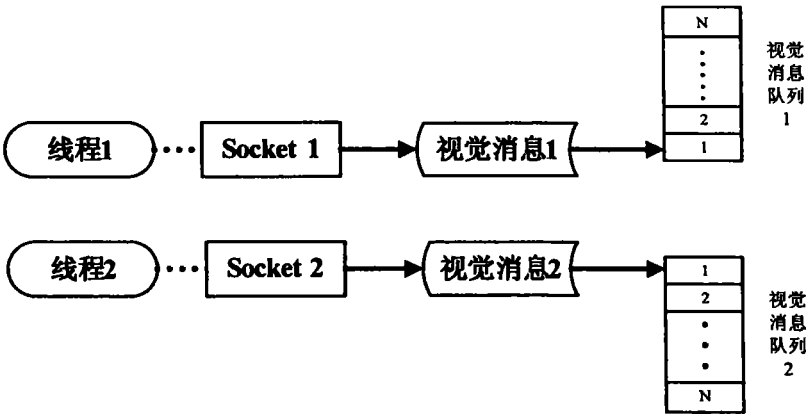


图 3.8 视觉消息接收和存储
Fig. 3.8 Acceptance and store of vision message

3.2.4 与裁判盒的通信模块

决策子系统与裁判盒的通信通过串口完成,决策程序需要一个独立的裁判盒线程监听与裁判盒计算机连接的串口。比赛时,每个裁判指令都会持续发送 10ms 的时间,裁判盒线程将接收到的裁判盒指令发到裁判指令队列中。决策程序每次进入主决策之前都要检查裁判指令队列中是否有新的指令,如果有,则根据裁判指令判断比赛状态,根据比赛状态制定具体策略并转化为机器人小车的动作指令发送出去。

3.2.5 无线通信模块

决策子系统与机器人小车的无线通信通过一个 USB 连接的无线收发设备完成的。主决策模块做出决策后,将最终策略转换为每个机器人小车的动作指令,然后将相应指令按照事先定义好的通信协议组装数据包。数据包的大小为 33 个字节,其中每个机器人的控制指令占 6 个字节。每个机器人小车收到决策计算机发出的指令后,将属于自己的 6 个字节的控制指令解析出来,其他部分的控制指令忽略。这样决策计算机可以只发送一条无线指令来控制 5 个机器人小车。

3.3 主决策模块详细设计与实现

主决策模块从视觉网络通信模块读取视觉信息,然后更新世界模型,读取裁判指令并更新比赛状态,基于更新的世界模型和比赛状态制定战术并发送控制指令。其流程如图 3.7 所示。

3.3.1 场地信息的更新

场地信息的保存与更新采用与视觉消息一样的策略,即使用固定长度的循环队列,所不同的是保存场地信息只需要一个队列,这个队列中的每个元素是整个场地上本方机器人、对方机器人以及球的位置信息,队头是最新的场地信息。场地信息的更新就是将两个视觉消息队列中队头元素提取出来,然后分析其中包含的机器人位置和球的位置信息,将这些信息合并到一个场地信息元素中,再把这个元素插入到场地信息队列中。

3.3.2 世界模型更新

在比赛过程中,有可能出现球或者机器人的位置超出本方摄像机的可视范围的情况(如球出界或者机器人离场等),我们称这种状态为“视觉丢失”。此时场地信息中就不会包含处于该状态的物体的位置信息。世界模型部分应该能够识别这种情况,并对世界模型信息做相应的处理

在我们的程序实现中,针对“视觉丢失”问题的解决方案是赋予每个物体视觉置信度的属性,该属性表示该物体的位置、速度及加速度等信息的准确程度。对于每个周期

都能被摄像机“看到”的物体，其置信度一直为 1；对于处于“视觉丢失”状态的物体，其置信度则按照一定比例衰减。在这个衰减的过程中，物体的位置信息由历史位置和速度预测得到。当某物体的置信度衰减到一定阈值时，该物体被视为完全丢失。造成这种情况的原因可能是物体被移出场外很长时间或者视觉系统出现了故障，此时针对该物体的决策应该停止。置信度的衰减如式(3.1)所示，其中 $Conf_n$ 是当前置信度， $Conf_{n-1}$ 是上一周期的置信度， $decay$ 是每周期的衰减常量。每次更新世界模型时，如果场地信息中包含某物体信息，则该物体的 $Conf_n$ 被置位 1，否则按照式(3.1)更新 $Conf_n$ 。

$$Conf_n = Conf_{n-1} - decay \quad (3.1)$$

对球进行更新时，从场地信息队列中提取球的位置信息，利用连续的位置信息差分得到连续的速度序列，由于在没有外力干扰的情况下，球在水平方向的投影是直线，所以可以对速度进行中值率波以得到一个唯一的速度方向，然后利用二乘法对速度大小进行率波并求得加速度。式(3.2)和(3.3)是计算速度和加速度的基本公式。

$$\bar{v}_n = \bar{p}_n - \bar{p}_{n-1} \quad (3.2)$$

$$\bar{a}_n = \bar{v}_n - \bar{v}_{n-1} \quad (3.3)$$

如果在连续记录球的位置信息的过程中，球由于受到外力作用而导致速度方向发生了变化，则对球的位置信息序列进行中值率波时会得到错误的速度方向，因此需要给定速度序列中速度方向变化量的阈值，超过此阈值则意味着球受到了外力。对于计算速度来讲，受力之前的场地信息已经失效，需要从受力点开始重新记录场地信息。

对机器人进行更新时，主要更新机器人的置信度以及计算机器人的当前速度。由于每周期机器人都会收到来自决策子系统的指令并对运动速度做出改变，因此加速度信息无需计算。本方机器人当前的速度决定了机器人达到目标速度所需要的加速度大小，所以有必要计算出来。

3.3.3 比赛状态判断

比赛状态的判断不仅需要读取裁判指令，还需要结合历史状态和场上形势综合判断。因为并不是所有的比赛状态转变都会发送裁判指令，一般来讲，比赛从 Play-On 状态到 Stop 状态的转变会发送裁判指令，而从 Stop 到 Play-On 的状态则需要决策系统自己判断。这里比赛的 Play-On 和 Stop 状态类似足球比赛中的“活球”和“死球”状态。例如从蓝队任意球(Free-Kick-Blue)到 Play-On 状态的转变，需要决策子系统判断任意球是否被正确发出，如果是，比赛进入 Play-On 状态，否则比赛仍处于 Free-Kick-Blue 状态并等待蓝队发任意球。

3.3.4 底层技术模块

底层技术是机器人小车的基本技能，包括跑位、截球和踢球三部分。优秀的底层技术需要对小车和球的物理模型的准确掌握，本章 3.4 节将详细讨论小车和球的物理模型的测定。其中小车的物理模型是否稳定可测与小车的运动和机械性能有关，同时良好的运动和机械性能也是良好的底层技术的前提和保障。

(1) 跑位

跑位是机器人足球比赛中的基本技术，也是其他技术能够发挥作用的基本保障。小车的跑位策略决定机器人小车的运动能力，关系到机器人小车是否能够完成决策子系统下达的任务，跑位正确与否的衡量尺度是机器人小车是否能够到达某一指定点，并且将车体朝向调整为指定方向。

跑位策略能否被正确的被执行决绝与机器人小车的运动控制和运动规划。这些问题是当前实体机器人研究的热点之一，学者们做了大量的研究，对各种控制任务提出了许多控制算法。Brockett 应用微分流形的工具研究了无定向系统(Driftless System)的控制问题^[28]，指出了哪类控制器适用于无定向的非完整系统的控制。Bloch、Sarkar 等人阐述了非完整机械系统的运动特性和控制特性，并介绍了非完整机械系统的基本控制方法^[29,30]。Campion 等人深入系统的研究了轮式移动机器人的结构特性和运动学以及动力学模型^[31]。

现在比较成熟的小型足球机器人系统的运动控制大都在控制子系统里实现，并且能够实现平稳的平动和转动相结合的运动，也就是运动和转向是同时进行的。由于我们的机器人小车刚刚开始研制，运动控制程序还很不完善，虽然运动方向可以达到 $0\sim 360^\circ$ ，但是对于同时进行平动和转动的控制还无法实现。因此为了满足正常比赛的需要，上层决策系统在制定跑位策略的时候不得不将运动和转动分开考虑，采取先平动再转动的跑位策略。

运动的目标是快速平稳的达到某一目标点，由于小车可以向 $0\sim 360^\circ$ 范围内的任意方向移动，所以机器人小车从当前位置点到目标点的运动是直线运动，则两点间的距离决定了小车的运动时间。另外到达目标点的目标末速度也会影响运动时间。由于具体的跑位算法与小车的运动模型关系紧密，因此我们将在讨论小车运动模型时给出具体的跑位算法。

(2) 截球

截球是机器人足球比赛中的关键技术，对球的抢断和控制都要依靠先进的截球技术来实现，快速的截球可以使得机器人在比赛中能有更多的机会控球。在仿真比赛中，由于环境模型是确定的，对于截球策略和算法的研究已经比较成型。但是在小型组比赛中，

由于硬件性能的影响和环境模型的不确定性，基于精确模型和公式推导的解析法并不适用，因此大多数球队都采用模拟法求得截球策略的近似解，再通过闭环反馈的方式逐步调整截球策略，实现截球技术。

模拟法的基本思路是根据球的运动模型和当前速度，模拟球的运行轨迹，并判断球的运行轨迹上是否存在某一点，使得机器人运动到该点的时间小于球的运行时间。如果存在，那么所有满足条件的点都是可以截球的点，其中机器人运动时间最短的点就是该机器人的最快截球点。模拟法示意图如图 3.9 所示，其中 \overline{v}_b 是球的初速度， \overline{v}_r 是机器人的初速度， t_b 是球的运动时间， t_r 是机器人运动到截球点 P 的时间，最快的截球时间是满足 $t_r < t_b$ 的最小的 t_r 。从图中可以看出，最优的截球策略是：机器人从初始位置 R_0 点开始运动时就可以确定最终的截球点 P 的准确位置，在向截球点的运动过程中，没有因为截球点位置的变化而导致的速度改变。这样的截球运动分为两个阶段，第一阶段是加速阶段，这个阶段可能包括机器人速度大小和方向的改变，它的运动轨迹可能是一条弧线，如图中 R_0 到 R_1 的阶段；第二阶段是匀速运动阶段，当机器人加速到朝向截球点的最大速度时，开始匀速运动，直到到达截球点，这一段的运动轨迹是一条直线，如图中的 R_1 到 P 的阶段。

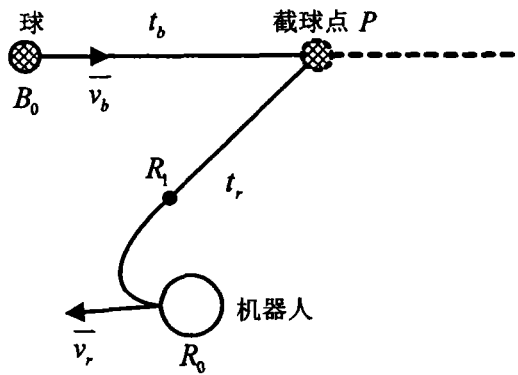


图 3.9 模拟法截球示意图
Fig. 3.9 Sketch map of intercept simulation

由于环境噪声和运动模型的误差影响，实际比赛中的截球策略不可能达到最优，因此只是一个基于最优理论的相对较优的实现，这个优化程度取决于视觉信息的以及球和机器人小车运动模型的精确度。

(3) 踢球

机器人小车的踢球技术主要取决于击球(或挑球)机构的性能。在击球(或挑球)机构性能确定的情况下,击球(或挑球)距离和力量的不同可能导致出球速度和方向的巨大差别,因此决策系统通过控制击球(或挑球)距离和力量可以达到控制出球速度的目的。由于我们球队的机器人小车的击球机构性能还不够稳定和理想,无法对其进行详细的性能测试,因此我们对于踢球技术未做过多工作,只是按照最大的允许力量踢球。通过实验得知,我们的机器人小车在使用最大力量击球时,可以使小球产生大约 $1\sim 2\text{m/s}$ 的初速度,而目前国内外一些强队的踢球速度可以达到 $10\sim 20\text{m/s}$ 。

3.3.5 上层策略模块

上层策略模块包括攻防选择、具体战术制定、任务分配以及任务实现等内容。

根据世界模型信息和比赛状态,上策策略模块首先要决定当前应该进攻还是防守或是执行定位球策略等,即进行攻防选择。在 Play-On 的比赛状态中,攻防选择的主要依据是对球的控制,首先要运用模拟法计算本方机器人是否能够比对方更早截球,如果本方能够截球,就执行进攻战术,否则执行防守战术。这一步涉及到对手的运动建模,因为对手的运动特性和截球策略都是未知的。目前无论是仿真还是小型组,在计算对手截球时间时都是计算理论最优时间,而自己的截球时间则使用模拟法计算实际时间。在小型组中,理论最优时间是很难确定的,因为无法确切知道对方机器人的最大速度。因此需要根据本方和对方的实力对比以及攻防策略的侧重给对方机器人一个假定的最大速度,然后根据此速度估算对方的最优截球时间。

对于 DUT Fantasia SmallSize 来说,机器人小车的性能与其他球队有着比较大的差距,因此在估算对手截球时间时通常给对手一个较大的速度,采取相对保守的策略。这种策略导致的后果就是本队大部分时间执行的都是防守策略,只有在机器人距离球非常近时才会执行进攻策略。处于对这种保守的攻防策略的补偿,我们在进攻和防守战中都安排一个本队截球最快的机器人按照自己的截球位置执行截球策略,这个补偿措施可以避免由于过高估计对手而导致的机会丧失。

具体的战术制定包括进攻战术、防守战术、定位球战术以及守门员等,由于大部分的战术需要多个机器人参与并配合,因此还要涉及具体某个战术中的任务分配,将具体任务分配到每个机器人之后,还需要制定每个机器人的任务实现策略,例如带球、传球或者射门等。这部分内容将在 3.5 节中做进一步讨论。

3.4 物理模型测定

比赛环境以及运动物体的物理模型是进行 AI 决策的基础,精确的物理模型是对场上形势进行正确判断和估计的前提,因此对于物理模型的测定工作至关重要。物理模型

的测定主要包括球的运动模型测定、机器人小车的运动模型测定以及机器人的带球和击球机构特性的测定。

3.4.1 球的运动模型测定

球运动模型的测定是为了了解小球在比赛过程中的运动规律，以便可以通过其运动规律对球在未来某段时间内的运动情况进行模拟或预测。准确掌握球的运动模型对于决策系统的性能至关重要，截球、传球及射门等关键技术都要以准确的小球运动模型为基础。

比赛中小球的运动大致分为两类：在地毯上的运动和在空中的运动。在空中运动时，球所受的空气阻力可以忽略不计，那么球只受竖直方向的重力作用，水平方向相当于匀速直线运动，其模型基本不需要测定。

对于球在地毯上运动的情况，从受力方面进行精确分析比较复杂，可能同时存在滚动和滑动摩擦。因此对于球的运动模型测定，主要从分析其大致的运动规律入手。

为了使小球产生一个稳定的在地毯上运动初速度，在进行测试实验时，我们让球从一个斜面上平滑滚落到地毯上，通过调整斜面的高度来得到不同的初速度。当球在地毯上运动时，我们利用视觉系统记录球的位置信息，并将这些位置信息和记录信息时的时间保存到文件中，然后对文件中的数据进行分析。在对不同的初始球速进行了几百次实验后，我们得到了球的大致运动规律。以下图片是其中两组数据的分析结果：

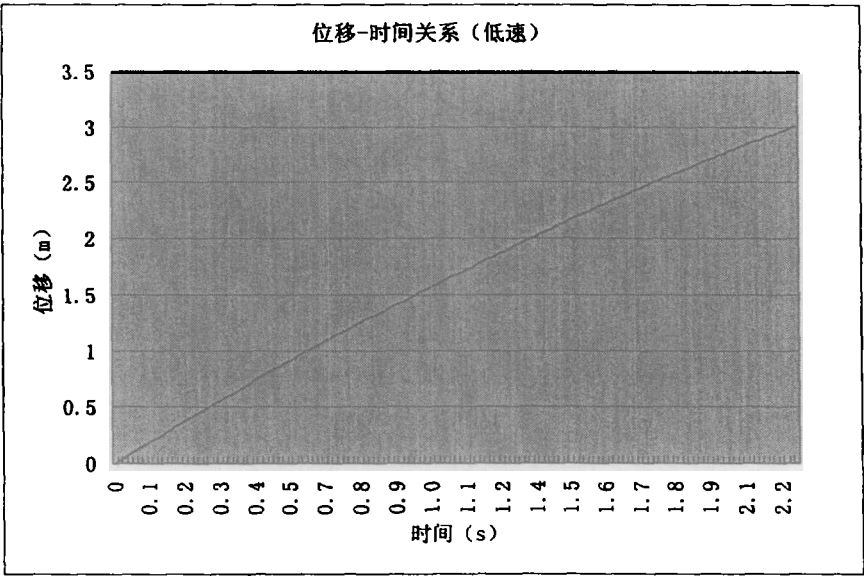


图 3.10 小球运动的位移-时间关系(低速)
Fig. 3.10 Relation between displacement and time of a moving ball (low speed)

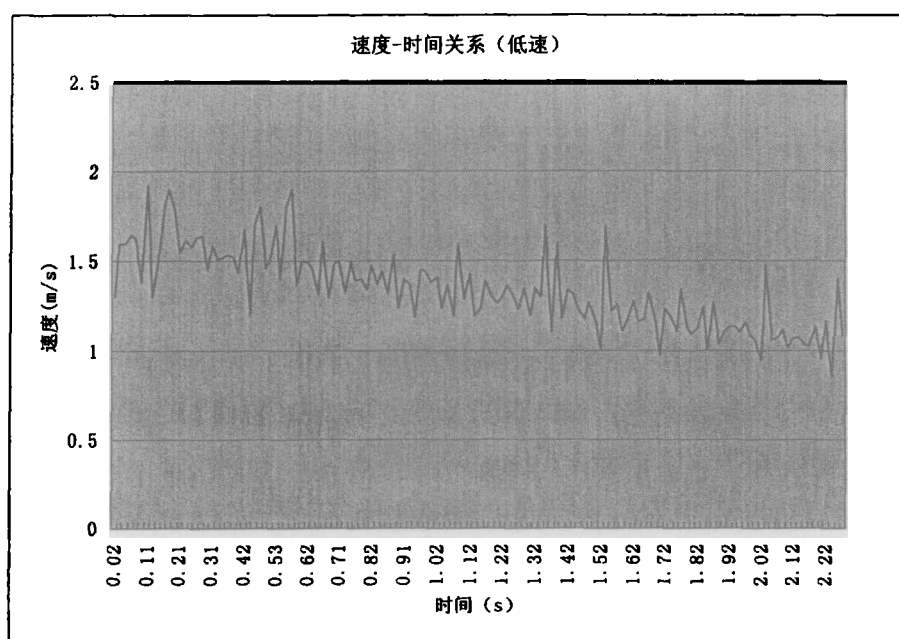


图 3.11 小球的速度-时间关系曲线(低速)
Fig. 3.11 Relation between velocity and time of a moving ball (low speed)

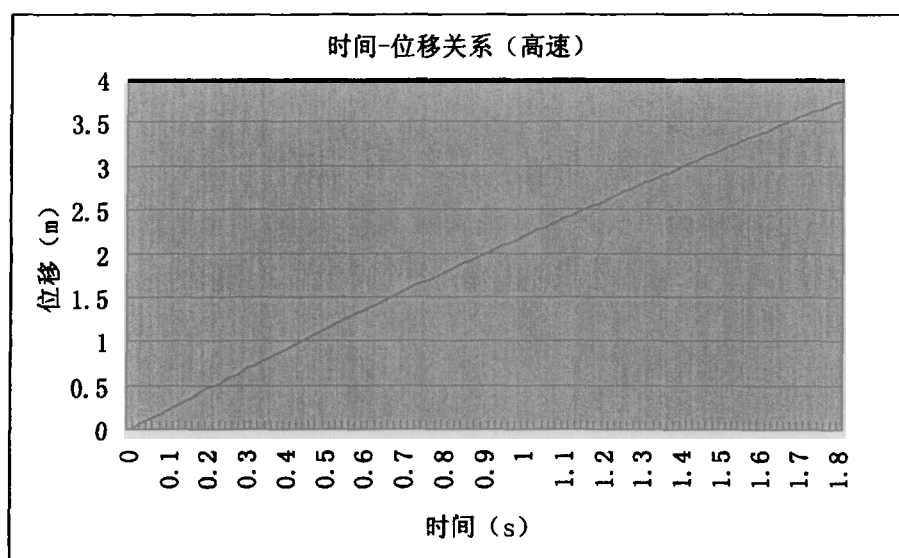


图 3.12 小球运动的位移-时间关系(高速)
Fig. 3.12 Relation between displacement and time of a moving ball (high speed)

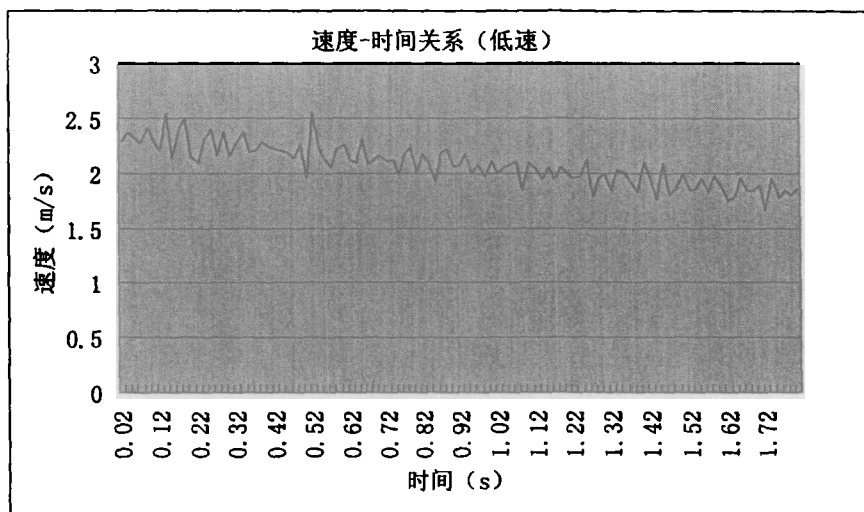


图 3.13 小球的速度-时间关系(高速)

Fig. 3.13 Relation between velocity and time of a moving ball (high speed)

图 3.10 和图 3.11 是在小球初速较低的情况的测得的“位移-时间”关系和“速度-时间”关系，图 3.12 和图 3.13 是在小球初速相对较高的情况下测得的“位移-时间”关系和“速度-时间”关系。从图 3.10 和 3.12 种可以看出，小球在地毯上运动是加速度较小的减速运动，而从图 3.11 和图 3.13 种可以看出，小球的运动基本上遵循匀减速的规律。

在采集实验数据时，视觉子系统的采样频率是 16ms，即每 16ms 记录一次位置和时间信息。在 16ms 内小球运动的距离很短，如果球速是 2.5m/s 的话，16ms 运动的距离是 0.04m。同时视觉系统对位置信息的采集又存在一定误差，这个误差一方面是由于对颜色块进行重心提取时所产生的位置信息误差，另一方面是采集数据的时间误差。之所以会产生时间误差，是因为在记录采集时间时，实际上记录的是图像处理之后的系统时间，由于每帧图像的传送和处理时间不可能完全相同，因此每两帧图像之间的时间间隔也就不会完全相同。

视觉系统的误差和采集数据的高频率，造成了速度曲线的波动比较大，而且速度曲线的波动幅度会随着球速的减小而增大，如图 3.11 和 3.13 所示，低速情况下的速度-时间曲线比高速情况下的曲线波动幅度更大一些，这是因为低速情况下每两个数据之间位置信息差别更小，对于误差更敏感。

我们对于其他实验数据进行同样的分析，发现小球的运动规律与这两组数据中基本一致，都可以近似看作匀减速直线运动，因此可以球在地毯上的运动可以用基本的牛顿动力学公式表示：

$$v_t = v_0 + at \quad (3.4)$$

$$s_t = v_0 t + \frac{1}{2} at^2 \quad (3.5)$$

需要注意的是以上小球的运动模型测定只是验证其运动规律，在实际比赛过程中并未应用上述曲线来解决小球的运动相关问题，而是根据小球的历史视觉数据对球的运动特性进行实时计算，其中采用了中值率波、最小二乘法拟合等方法。

3.4.2 小车的运动模型测定

小车的运动模型测定分为运动特性测定和转动特性测定，其中包括最大运动和转动速度，加速和减速特性等。

对于 DUT Fantasia SmallSize 的小车运动机构，其运动和转动的控制输入为代表电机分级转速的标量，如运动的输入为 0~150，转动的输入为 0~192。运动模型测定的任务就是确定这些标量所对应的实际速度。测定方法为让小车在实际的比赛环境中按照不同输入执行动作，记录视觉信息中相对应的位置信息。然后对这些位置信息进行数据拟合，得到小车近似的运动模型。由于我们的小车采用的车轮是双排万向轮，其运动精度比较低，因此小车的运动性能很不稳定，只能通过数据大致了解其运动特性。以下是对一些测试数据的分析。

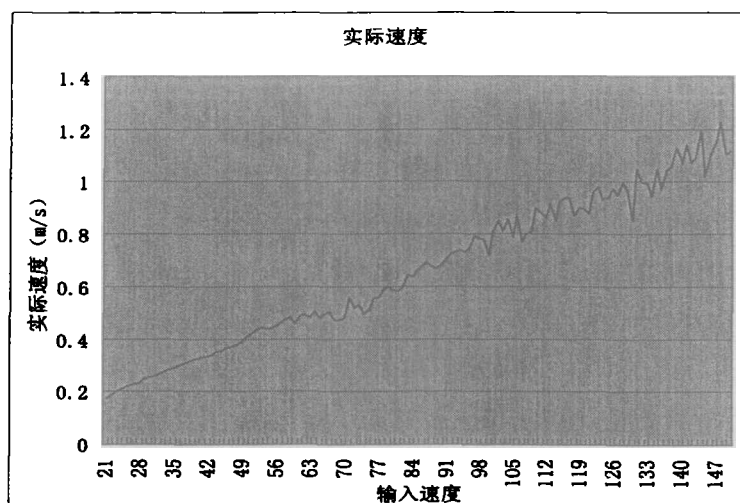


图 3.14 实际速度与输入速度的关系

Fig. 3.14 Relation between real velocity and input velocity

图 3.14 为测得的实际运动速度与输入速度的数据关系曲线图。在实验时，当速度达到稳定值之后，取一段时间内的平均速度作为其输出速度，由图可知，其关系相对比较

稳定，曲线比较平滑，大致符合线性关系。使用曲线拟合工具可以获得实际速度与输入速度的函数表达式为式(3.6)所示，其中， $p_1 = 0.018329$ ， $p_2 = 0.007498$ 。

$$y = p_1 + p_2 x \tag{3.6}$$

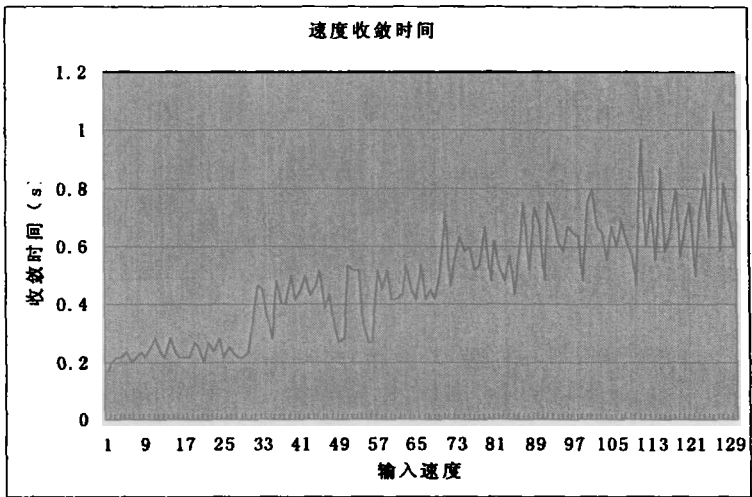


图 3.15 速度收敛时间与输入速度的关系曲线
Fig. 3.15 Relation between converge time of real velocity and input velocity

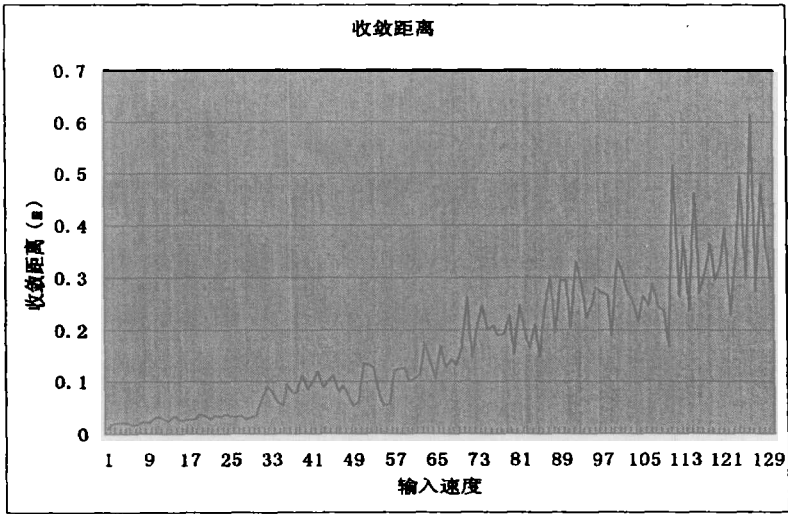


图 3.16 速度收敛距离与输入速度的关系曲线
Fig. 3.16 Relation between converge distance of real velocity and input velocity

图 3.15 和图 3.16 分别为测得的小车从启动到速度到达稳定速度所经过的时间和距离与输入速度的数据关系曲线图。实验时，当实际运动速度连续几个周期变化量小于一定范围时，即认为实际速度已经收敛，此时小车的运动时间就是其收敛时间，此时小车的运动距离为其收敛距离。由图可知，由于小车的运动性能不稳地和视觉系统存在误差，因此其收敛时间和收敛距离非常不稳定，用曲线拟合工具拟合后得到公式(3.7)和(3.8)，式(3.7)中， $p_1=0.105722$ ， $p_2=0.004378$ ；式(3.8)中 $p_1=0.000065$ ， $p_2=1.737006$ 。

$$y = p_1 + p_2 x \quad (3.7)$$

$$y = p_1 + p_2 x^2 \quad (3.8)$$

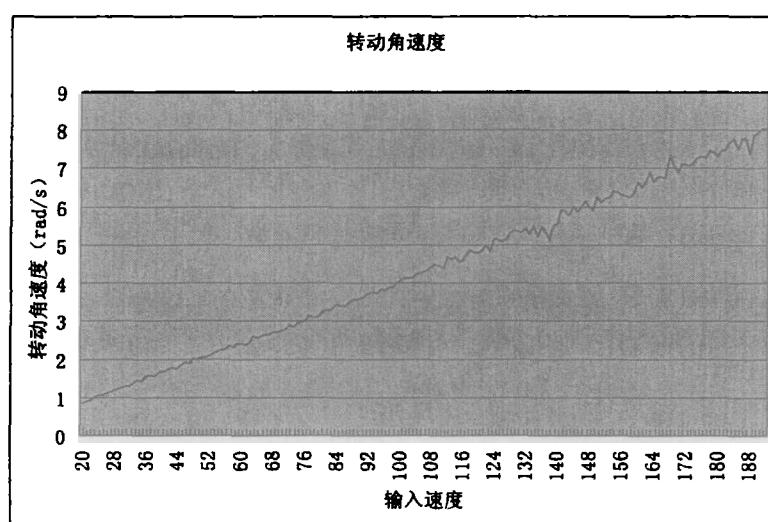


图 3.17 转动角速度与输入速度的关系曲线
Fig. 3.17 Relation between angle velocity and input velocity

图 3.17 是测得的小车转动速度与实际输入转速的数据关系曲线图。实验时，当转动速度稳定后，取一段时间内的平均转动速度作为其输出转速，由图可知，其关系相对比较稳定，曲线比较平滑。通过曲线拟合工具可得实际转速和输入转速的函数表达式为式(3.9)所示，其中， $p_1=0.007728$ ， $p_2=0.007421$ 。

$$y = p_1 + p_2 x \quad (3.9)$$

式(3.6)~(3.9)的运动模型公式是在测试数据的基础上通过曲线拟合得到的，其中实际运动速度和实际转动速度与输入速度的关系相对比较准确，这是因为实际的运动速度和转动速度在计算时取得是统计平均值，小车在这两项指标上的性能相对比较稳定；而在运动收敛时间和收敛距离两项指标方面，上面给出的模型公式只是一个能够代表大致

趋势的估计值；而对于转动速度的收敛时间和收敛角度数据，则因为其随机性过大，无法从中分析出统一的规律。

在具体的比赛过程中，实际运动速度和转动速度与输入速度的关系要经常用到，式(3.6)和式(3.9)所表示的模型基本能够满足对精度的要求，可以直接使用；运动收敛时间和收敛距离可以结合式(3.7)和(3.8)在粗略计算时使用；而转动速度的收敛时间和收敛角度则要靠经验值来判断。

3.4.3 小车的带球和击球特性测定

带球和射门机构的特性决定小车在截球成功之后对球的处理能力。带球机构的特性主要包括横向带球特性和纵向带球特性，即小车前后移动时的带球速度和横向移动时的带球速度。击球机构的特性包括击球距离及击球力量与出球速度的关系等。由于机械结构方面的缺陷，我们的机器人小车的带球和击球机构性能非常不稳定，因此无法对具体参数进行测定。在比赛过程中，带球和击球的输入值都是允许范围内的最大值。

3.5 战术设计与实现

在小型组这个高度动态和竞争的环境中，如何控制好一个具有五个机器人的队伍，是一项富有挑战性的研究工作。一支成功的队伍不仅需要很好的单个机器人的能力，也需要一个很好的团队协作，即一个整体的战术体系。在这个体系中，要使每个机器人能够执行对整体有用的任务，因此战术体系的设计是小型足球机器人决策系统的重要组成部分。根据一场真正的足球比赛需要解决的问题，可将战术分为三部分：进攻、防守和守门员。

3.5.1 进攻

我们的进攻战术是基于一个基本的 1-1-3 进攻阵型设计的。进攻时，除控球队员和守门员外，其他队员按照阵型跑位。每个队员的跑位点获取策略是：按照球的位置计算出自己的阵型本位点，并在本位点周围一定范围内(如半径 0.3m 的圆周内)选择距离对手或者自己的队友距离较远的点作为自己的阵型跑位点。每个队员的本位点与球的位置关系根据经验公式计算，而寻找距离队友和对手较远点则通过搜索离散点来完成。图 3.18 是 2006 年中国公开赛比赛过程中的开场阵型。

进攻战术还包括传球和射门等关键技术。由于我们机器人小车的带球和击球机构性能还不够理想，击球距离短且击球准确性不够，无法实现精确传球和射门。因此，我们只实现了选择最大空档传球和射门的简单策略。一个参与进攻的队员的决策流程如图 3.19 所示：

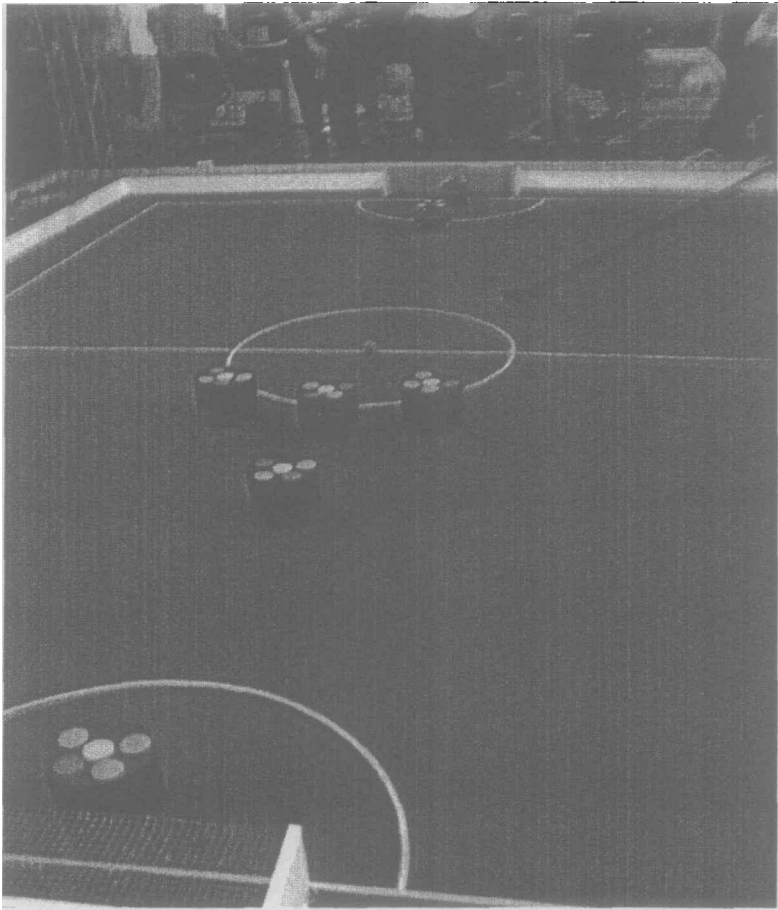


图 3.18 开场阵型
Fig. 3.18 Kick-off-formation

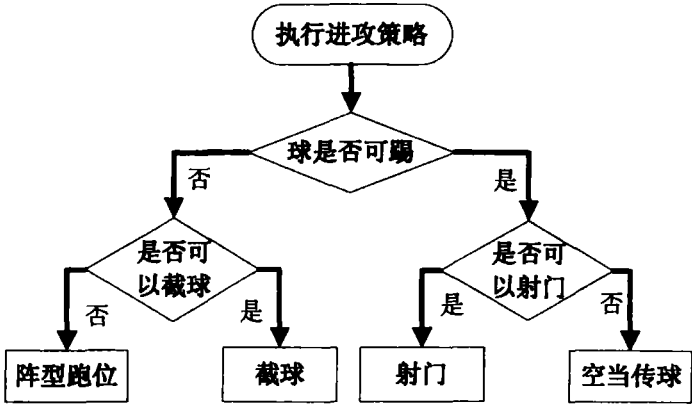


图 3.19 进攻策略流程
Fig. 3.19 Flow of attack strategy

3.5.2 防守

防守战术是整个战术系统中的重要环节，需要守门员和其他参与防守的队员互相协作，共同完成防守任务。由于小型组比赛中，场地尺寸小，场上比赛队员少，球的运动速度大大快于小车的运动速度，因此现在大多数球队采用的防守战术都是盯人防守。盯人防守战术中，首先要根据场上形势，进行防守任务分配。小型组比赛采用的是集中决策、分布控制的工作模式，因此只要合理的进行任务分配，就不会出现任务冲突的情况。在我们的防守策略中，只对位于本方半场内的对方队员采取盯人防守，具体的半场盯人防守策略实现过程如下：

- (1) 对方控球时，半场防守策略被触发
 - (2) 确定对方进攻队员的威胁度，与本方球门距离越近，威胁度越大
 - (3) 按照威胁度从大到小，依次为每个进攻队员分配防守队员，进行任务匹配。
- 防守任务分配后一个典型的半场盯人防守场景如图 3.20 所示。

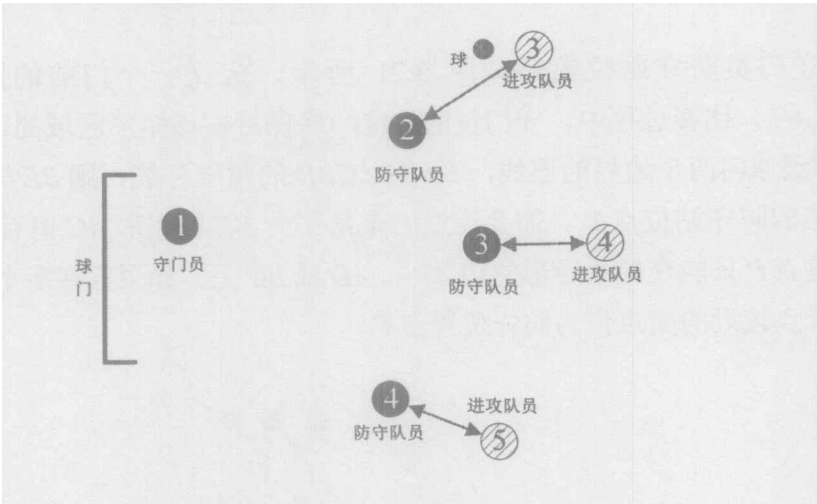


图 3.20 半场防守场景

Fig. 3.20 Episode of half-field defense

在进行任务分配时，根据进攻队员与球门的距离和角度确定其危险度。根据防守队员与进攻队员的距离和角度，结合球队的整体实力以及比赛的实际情况，选择适合的匹配策略对防守队员与进攻队员进行匹配。一种简单的匹配策略可以如下所示：

- (1) 从待匹配的所有未被匹配的进攻队员中选择最危险的一个，如果所有进攻队员都已经被匹配，则匹配结束。

(2) 从所有未被匹配的防守队员中选择一个防守队员与(1)中选出的进攻队员进行匹配, 这个防守 Agent 应该满足如下条件: 位置在进攻 Agent 与球门之间; 距离防守 Agent 的距离最近。

(3) 记录(2)中的匹配结果并标记被匹配的 Agent 为已匹配, 转到(1)。

防守任务分配将半场防守问题分解为多一个一对一盯人防守的子问题。小型组比赛中, 场上的环境比较复杂, 机器人动作的状态空间比较大, 而且由于不同队伍的机器人小车性能相差很大, 无法精确建立机器人小车的动作行为模型, 因此依靠手工编码制定一个有效的一对一防守策略比较困难。针对此问题, 我们将在第四章给出了一个应用基于 Markov 对策的强化学习解决该问题的方案。

3.5.3 守门员

在比赛过程中的大部分时间里, 守门员是比赛场上距离本方球门最近的本方队员, 是防止对方射门得分的最主要的队员, 因此防守站位点的选择是守门员战术的关键内容。

我们的守门员防守站位策略如图 3.21 所示, 定义一个门前的矩形防守区域 Defend-Box AC , 比赛过程中, 守门员的站位始终保持在该矩形区域的边界上。 BC 和 BD 分别是球到球门两个边界的连线, BE 为 $\angle CBD$ 的角平分线, 则 BE 与矩形 AC 的交点即为守门员的防守站位点 P 。需要注意的是角平分 BE 与矩形 AC 恒有交点, 但是合理的防守位置点 P 只能在处于矩形的边 AB 、 AD 或 BC 上, 如果存在多个交点, 则取距离球的位置 B 点较近的交点作为防守位置点 P 。

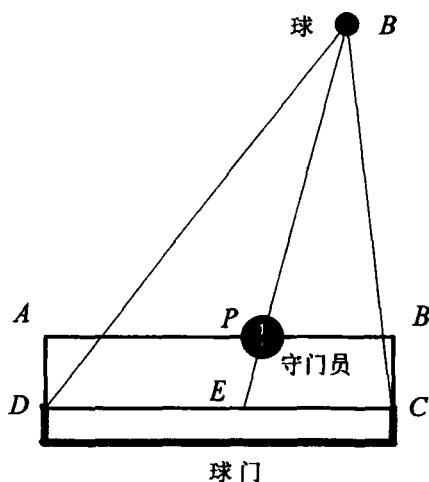


图 3.21 守门员防守站位
Fig. 3.21 Defense position of the goalie

防守站位除了防守点位置选择之外,还包括机器人小车的方向调整。我们的机器人小车的击球装置放在小车前方,因此比赛过程中,将小车的前方向调整为始终面向小球有助于及时将来球踢出门前的危险区域。

守门员策略中,除了防守站位,还包括截球和破坏球。其中守门员的截球策略相对于其他球员的截球策略要保守一些(比如截球区域要尽量限制在门前区域),因为截球存在失败的风险,如果守门员到距离球门很远的地方截球并且失败了,可能会导致对方的进攻队员直接面对空门。破坏球则是守门员在第一时间将球踢出门前的危险区域,而无需考虑传球目标等。守门员的具体策略流程如图 3.22 所示。

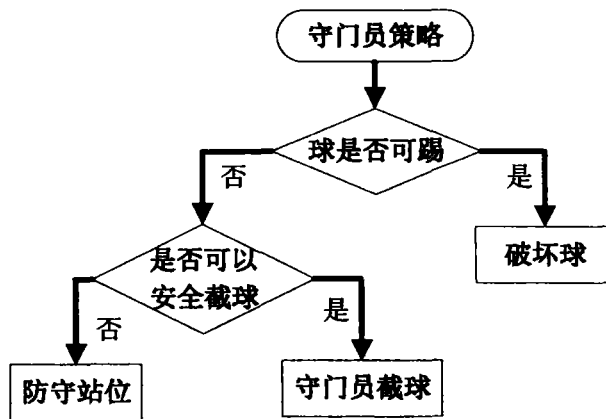


图 3.22 守门员策略流程
Fig. 3.22 Flow of goalie strategy

4 基于 Markov 对策的多 Agent 强化学习算法

在 3.5.1 中讨论的半场防守战术中，经过任务分解和匹配，最终需要解决的是一对一的盯人防守问题。在制定一对一盯人防守策略时，用简单的手工编码实现起来比较困难，而且经常会因为出现边界情况而造成防守失位并导致对方射门得分，因此需要考虑一种柔性更强的制定防守策略的算法。考虑到小型足球机器人系统可以看作是一个典型的多 Agent 系统，可以尝试利用多 Agent 系统中的强化学习来实现一个防守策略。

强化学习是学习如何把状态应设到动作使回报达到最大的学习算法。Agent 通过在环境中不断的感知和动作，来学习选择最优的动作以实现目标任务。强化学习坚实的理论基础和诱人的应用前景正逐渐受到个研究领域学者的广泛重视。

4.1 多 Agent 强化学习

4.1.1 强化学习的基本原理

强化学习的基本框图如图 4.1 所示，Agent 与环境的接口包括状态(state)、回报(reward)和动作(action)，其基本原理是：在某个环境状态 s_t 下，如果 Agent 采取某个动作 a_t 所导致出现的后续状态 s_{t+1} 会给 Agent 带来正的回报 r_{t+1} ，则当 Agent 再次面临相同的环境状态 s_t 时，采取动作 a_t 的趋势就会增强，反之趋势就会减弱。需要说明的一点是，这里的回报和有监督性学习中的监督或者教师是不同的。在监督性学习中，教师给出很多例子，告诉 Agent 在什么情况下，执行什么动作效果最好；而在强化学习中，回报只是告诉 Agent 当前动作的执行效果，Agent 要在与环境的交互过程中，不端测试每个动作的执行效果，在长时间的收集回报后，判断出每个行动的长期回报，完成 Agent 的学习过程。Agent 在环境中的学习过程如图 4.2 所示^[32]。

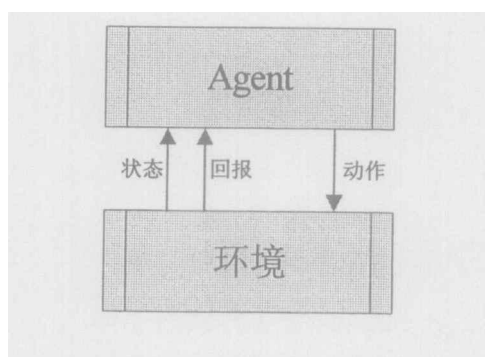


图 4.1 强化学习基本框图

Fig. 4.1 Framework of reinforcement learning

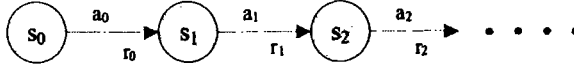


图 4.2 强化学习的學習过程
Fig. 4.2 Process of reinforcement learning

Agent 为了完成任务，必须知道采取某个策略而导致出现的某个状态对该 Agent 所产生的长期回报，而不是其立即回报。长期回报必须经过一定时间的延迟才能获得，Agent 在 t 时刻获得的长期回报可以用式(4.1)来表示：

$$V^{\pi}(s_t) = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \dots = \sum_{i=0}^{\infty} \gamma^i r_{t+i} \quad (4.1)$$

$V^{\pi}(s_t)$ 是 s_t 状态下采取策略 π 而产生的长期回报， γ 是延迟折算因子。其中回报函数和状态转移函数如式(4.2)和(4.3)所示：

$$r_t = r(s_t, a_t) \quad (4.2)$$

$$s_{t+1} = \delta(s_t, a_t) \quad (4.3)$$

强化学习的任务就是要学习一个策略 π ，使得对于所有状态 s ， $V^{\pi}(s)$ 最大。此策略被称为最优策略(Optimal Policy)，并用 π^* 表示。如式(4.4)所示，为简化表示，我们将此最优策略的值函数 $V^{\pi^*}(s)$ 用 $V^*(s)$ 表示， $V^*(s)$ 给出了当 Agent 从状态 s 开始可获得的最大折算累计回报，即从状态 s 开始遵循最优策略时获得的折算累计回报。

$$\pi^* = \arg \max_{\pi} V^{\pi}(s), (\forall s) \quad (4.4)$$

如果 t 时刻之后的环境状态只取决于 t 时刻的状态 s_t 和 t 时刻之后的动作，而与 t 时刻之前的状态和动作无关，则该学习过程可看作是一个 Markov 决策过程(Markov Decision Processes, MDP)。Markov 过程的定义如下：

$$MDP = \langle S, A, R, P \rangle \quad (4.5)$$

Markov 决策过程是如式 4.5 所示的一个四元组。其中 S 是有限的离散状态空间， A 是有限的离散动作空间； $R: S \times A \rightarrow \text{Real}$ 是回报函数； $P: S \times A \rightarrow \Delta$ 是状态转移函数， Δ 是 S 上的概率分布集合。Markov 决策过程中的 Agent，在每一个时刻 t ，可在 A 中选择某一个动作 a_t ， a_t 使环境状态 s_t 转移到 s_{t+1} ，同时给出一个回报值 r_t 。MDP 的本质是：时刻 t 之后的状态只与当前状态 s_t 和所选择的动作 a_t 有关，而与 t 时刻之前的历史状态和动作无关。强化学习着重研究 MDP 中的 P 函数和 R 函数未知的情况下，Agent 如何获得最优策略。

4.1.2 一种强化学习算法——Q 学习

本文 1.1.5 小节中提到了一些比较成形的强化学习算法，其中 Q 学习算法是应用比较广泛的一个。

考虑到让一个 Agent 直接学习函数 $\pi^*: S \rightarrow A$ 很困难，因此可以让 Agent 对评估函数 V^* 进行学习，如式(4.6)，根据式(4.6)，Agent 可以通过学习 V^* 获得最优策略的条件是：它具有立即回报函数 r 和状态转移函数 δ 的完美知识。当 Agent 得知了外界环境用来相应动作的函数 r 和 δ 的完美知识，它就可以用式(4.6)来计算任意状态下的最优动作。

$$\pi^*(s) = \arg_a \max [r(s, a) + \gamma V^*(\delta(s, a))] \quad (4.6)$$

然而遗憾的是，许多情况下，Agent 并不能够获得函数 r 和 δ 的完美知识，此时通过学习这样一个叫做 Q 函数的评估函数可以获得最优策略。评估函数 $Q(s, a)$ 定义为：它的值是从状态 s 开始并使用 a 作为第一个动作时的最大折算累计回报。换言之， Q 的值为从状态 s 执行动作 a 的立即回报加上以后遵循最优策略的值，如式(4.7)和(4.8)所示。

$$V(s) = \max_{a \in A} Q(s, a) \quad (4.7)$$

$$Q(s, a) \equiv r(s, a) + \gamma V(\delta(s, a)) \quad (4.8)$$

式(4.7)和(4.8)就是 Q 学习的一般形式，Q 学习的目标就是通过不断更新 Q 值来学习一个最优策略^[33,34]。

在具体的学习过程中，Agent 从环境中能够得到的反馈只有立即回报函数 $r(s, a)$ 的值，因此需要找到一个可靠的方法，只在时间轴上展开的立即回报序列的基础上估计训练值，这可以通过迭代逼近的方法实现。注意到式(4.7)和(4.8)可以重写为式(4.9)。

$$Q(s, a) = r(s, a) + \gamma \max_{a'} Q(\delta(s, a), a') \quad (4.9)$$

这个 Q 函数的递归定义提供了迭代逼近 Q 算法的基础^[21]。在此算法中学习器通过一个大表表示其对 Q 函数的估计值，其中每个状态-动作对有一个表项。状态-动作对 $\langle s, a \rangle$ 的表项存储的实际是学习器对实际 Q 函数值的当前估计，此表可被初始化为任意随机值。Agent 重复地观察其当前的状态 s ，选择某动作 a ，执行此动作，然后观察结果回报 $r = r(s, a)$ 以及新状态 $s' = \delta(s, a)$ 。然后 Agent 遵循每个这样的转换按照式(4.10)所示的规则更新 $Q(s, a)$ 的表项，其中 s' 是当前状态 s 下采取动作 a 所得到的后继状态。

$$Q(s, a) = r(s, a) + \gamma \max_{a'} Q(s', a') \quad (4.10)$$

如果状态转移函数 δ 的值不是确定性的，而是服从一定概率分布，则这种 MDP 叫做非确定性 MDP。假设用 A 来表示 Agent A 的动作集， S 表示所有可能的状态集，状态转移函数 δ 是服从概率分布 T ，则 Q 学习可以用式(4.7)和(4.11)来表示。

$$Q(s, a) = r(s, a) + \gamma \sum_{s' \in S} T(s, a, s') V(s') \quad (4.11)$$

4.2 Markov 对策学习框架

从机器学习的角度讲，多 Agent 系统中的强化学习大致有 2 种类型：一个是将整个多 Agent 系统作为一个可计算的学习 Agent，另一个是让多 Agent 系统中的每个 Agent 都有自己的强化学习机制，通过与其它 Agent 适当交互加快学习过程。这些交互可能包括通信、协作或者对抗等。其中对策模型是一种常用的数学模型。

在对策模型中，每个 Agent 获得的立即回报不仅仅取决于 Agent 自身的动作，同时还依赖于环境中其他 Agent 的动作。因此可以将多 Agent 系统中的每个离散状态 s 形式化为一个对策 g ，那么强化学习的 Markov 决策过程可以扩展为多 Agent 系统的 Markov 对策模型^[35]。

Markov 对策也就随机对策，可以用如下多元组表示：

S ：状态集

A_i ：Agent 的状态集， $i=1,2,\dots,n$ ， n 是 Agent 的个数

$T: S \times A_1 \times \dots \times A_n \rightarrow PD(S)$ ：状态转移函数，其中 $PD(S)$ 为状态集 S 上的状态分布

$R_i: S \times A_1 \times \dots \times A_n \rightarrow \mathbb{R}$ ：联合回报函数，其中 R_i 是 Agent i 在状态 s 下，所有 Agent 采取动作后获得的回报

如果环境中只存在两个互相竞争的 Agent A 与 O，他们动作集分别为 A 和 O ，则 A 所获得的折算长期回报 $Q(s, a)$ 可以扩展为 $Q(s, a, o)$ ，即当前状态 s 下，A 采取动作 a 而 O 采取动作 o 时所获得的长期回报。这种简单的一对一竞争交互，可以看作是二人零和对策问题，则 Q 学习的一般形式可扩展为式(4.12)和(4.13)。

$$V(s) = \max_{\pi \in PD(A)} \min_{o \in O} \sum_{a \in A} Q(s, a, o) \pi_a \quad (4.12)$$

$$Q(s, a, o) = r(s, a, o) + \gamma \sum_{s' \in S} T(s, a, o, s') V(s') \quad (4.13)$$

式(4.11)中，对手 Agent O 选择的动作是对 A 最不利的动作 o ， π_a 为动作集 A 上的分布，该式的含义是在对手 Agent 选择对自己最不利的动作的情况下，自己选择最有利的动作所获得的回报，这种方法叫做“极大极小 Q 学习”法^[36]。

在 Q 函数的学习过程中，为了保证算法可以收敛，需要假定 Agent 选择动作的方式为它可以无限频繁的访问所有可能的状态-动作对，即对于某一个动作-状态对 $\langle s, a \rangle$ ，可能重复多次访问到。因此对于同一个状态-动作对，考虑到原有的 Q 值对每次迭代中

新 Q 值的影响, 引入学习率 α_n , 其定义如式(4.13)所示, 则对于式(4.12)和(4.13)所表示的非确定性 MDP, Q 学习的迭代更新法则可以修改为式(4.14)和式(4.15)

$$\alpha_n = \frac{1}{1 + n(s, a, o)} \quad (4.14)$$

$$Q(s, a, o) = (1 - \alpha_n)Q(s, a, o) + \alpha_n(r(s, a, o) + \gamma \sum_{s' \in S} T(s, a, o, s')V(s')) \quad (4.15)$$

其中, s 和 a 为第 n 次循环中更新的状态和动作, 而 $n(s, a, o)$ 是状态-动作对 $\langle s, a \rangle$ 在这 n 次循环中被访问到的总次数, 式(4.14)和(4.15)的修改的关键思想是让 Q 值的更新更为平缓, 对于具有概率性输出的回报函数 r 的非确定性 MDP, 此迭代方法也可以确保学习算法最终会收敛。在式(4.15)中, 随着 $n(s, a, o)$ 增大, α_n 逐渐减小, 因此当训练进行时更新程度逐渐变小。在训练中以一定速率减小 α_n , 可以达到收敛到正确的 Q 函数。

4.3 RoboCup 一对一盯人防守中的学习模型

机器人足球比赛可以作为研究多 Agent 系统强化学习理论的实验平台。对于 RoboCup 比赛中的强化学习, 国内外的研究人员已经做了很多研究工作, 并且大多数几集中在仿真领域。其中 Stone 将分层学习应用到了 RoboCup 仿真球队 CMUnited 中^[37], 清华大学用 Q 学习训练 Agent 的踢球和带球等个人技术^[38], 德国卡尔斯鲁厄大学利用基于神经网络的学习算法训练仿真球队 BrainStormer 的战术^[39], 这些研究都在实际的比赛中取得了较好的效果。但是这些研究较少针对多 Agent 系统中的对抗性问题做过多研究, 而且对于小型组中的强化学习讨论比较少, 因此本章以 RoboCup 小型组比赛为应用背景, 以小型组比赛的仿真平台实验平台, 研究一种基于 Markov 对策的强化学习算法在解决一对一盯人防守问题中的应用。以该算法作为基本防守策略的小型组足球比赛球队 DUT Fantasia SmallSize 参加了 2006 年苏州举行的 RoboCup 中国公开赛, 在三场小组赛中取得了一胜一平一负, 只丢一球的成绩。

4.3.1 一对一盯人防守的环境建模

在防守任务分配结束后, 半场防守问题被分解为多个一对一盯人防守问题, 一个典型的一对一盯人防守问题如图 4.5 所示。在这一特定场景中, 环境可用如下变量描述:

$\overrightarrow{d_{og}}$: 进攻 Agent 到球门中点的距离矢量, 包括大小和方向

$\overrightarrow{d_{oa}}$: 进攻 Agent 与防守 Agent 的距离矢量, 包括大小和方向

$\overrightarrow{v_o}$: 进攻 Agent 的速度矢量, 包括大小和方向

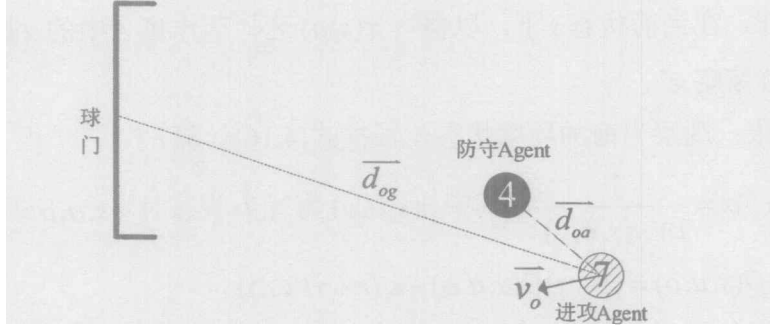


图 4.5 一对一防守场景
Fig. 4.5 Episode of the one-vs-one defense

在图 4.5 所示的场景中，我们主要考虑的是一对一盯人防守。根据以往的比赛经验，比赛中绝大多数的失球是因为对方进攻队员面前没有任何有效防守队员造成的。根据仿真比赛的平台规则，如果进攻队员的决策足够合理，任何已经被进攻队员甩在身后的防守队员都无法再次追上进攻队员并实施有效的防守，因此，一对一防守时，防守 Agent 的主要目标是保持一个有效的防守位置，与进攻 Agent 形成一种“对峙”的平衡状态，在这种状态中，进攻 Agent 迫于防守压力，将无法继续向球门方向推进，也无法找到较好的射门角度。在这种情况下，球的位置对防守效果的影响不大。因此我们用三个矢量来描述这一防守场景， $\overline{d_{og}}$ 代表了进攻 Agent 对球门的威胁程度， $\overline{v_o}$ 代表了进攻 Agent 的运动趋势， $\overline{d_{oa}}$ 则代表了防守 Agent 对进攻 Agent 施加防守压力的程度。

在任意时刻，防守 Agent 从环境获得的立即回报遵循式(4.16)所示的函数

$$r = f(\overline{d_{og}}, \overline{d_{oa}}, \overline{v_o}) \quad (4.16)$$

防守 Agent 和进攻 Agent 的动作集均为 $Move(dir, power)$ ，其中 dir 的取值范围是 $(0 \leq 360^\circ)$ ， $power$ 的取值范围是 $(0 \leq MaxPower)$ ， $MaxPower$ 可根据具体情况给定一个上限值。在具体学习过程中，可在权衡计算精度和效率的前提下按照一定步长取 dir 和 $power$ 的一些离散值点，构成 Agent 的动作集 A 和 O 。

4.3.2 基于 Markov 对策的学习算法

对于 4.3.1 中讨论的一对一盯人防守问题，具体的学习算法如下：

- (1) 初始化参数，对所有 $s \in S, a \in A, o \in O$ ， $Q(s, a, o) = 1$ ， $V(s) = 1$ ， $\alpha_0 = 1.0$ ， $\pi(s, a) = 1/|A(s)|$ ；
- (2) 选择一些特定场景，初始化实际环境；

(3) 选择动作，在当前状态 s 下，以概率 $\pi(s, a)$ 选择动作集 A 中的 a 执行；

(4) 学习最优策略 π^* 。

计算立即回报：观察当前的环境状态，根据式(4.16)计算 r ；

修正学习率： $\alpha = \frac{1}{1 + n(s, a, o)}$ ，其中 $n(s, a, o)$ 为行为状态对 $\langle s, a, o \rangle$ 出现的次数；

更新 Q 值： $Q(s, a, o) = (1 - \alpha)Q(s, a, o) + \alpha(r + \gamma V(s'))$ ；

选择最优策略 $\pi^*(s, \dots)$ ：使得 $\min_{o' \in O} \sum_{a' \in A} \pi'(s, a') Q(s, a', o')$ 取得最大值，令 $V(s)$ 等于次最大

大值。

(5) 如果当前状态是终止状态，则转向步骤(2)，初始化场景；否则转向步骤(3)，继续选择动作。

注意到更新 Q 值时没有考虑状态转移概率，这是因为上述学习算法在学习过程中，每次迭代都是先在状态 s 下执行动作 a ，然后观察执行动作之后的新状态 s' ，根据 s' 可以计算出立即回报 r 和延迟回报 $V(s')$ ，而在整个学习过程中，每个 s' 出现的频率恰好等于状态转移概率 $T(s, a, o, s')$ ，因此上述学习算法中 Q 值的更新是合理的。

4.3.3 实验及实战效果

(1) 实验效果：一对一防守进攻 Agent 的带球推进

将 4.3.2 中的学习算法应用在 Robocup 小型组比赛的仿真平台上，让一个进行学习的防守 Agent 与一支仿真训练球队的前锋进行一对一的攻防训练。其中防守 Agent 的执行策略为：如果可以比进攻 Agent 更快截球，则截球，否则执行 Q 学习的防守策略；而进攻 Agent 则执行前锋的带球推进分支策略。

场景初始化为防守 Agent 位于禁区线上的随机位置，进攻 Agent 位于中线上的随机位置，进攻 Agent 的目标为带球向禁区推进，当出现如下状态时，该场景结束：

- ① 进攻 Agent 成功带球推进到禁区，防守 Agent 获得回报-1；
- ② 球出界，防守 Agent 获得回报 0；
- ③ 防守 Agent 断球，防守 Agent 获得回报 1；
- ④ 迭代次数到达 300 个仿真周期；

回报函数 r 的值满足： $|f(\overline{d_{og}}, \overline{d_{oa}}, \overline{v_o})| < 0.005$ ，动作集 $Move(dir, power)$ 离散为 24×5 个值。如果在 300 个仿真周期内，进攻 Agent 没有成功带球推进到禁区或者防守 Agent 成功断球，则认为防守成功，否则防守失败。经过 10000 个场景的训练，得到了基本稳定的防守策略。防守成功率与训练次数的曲线如图 4.6 所示。

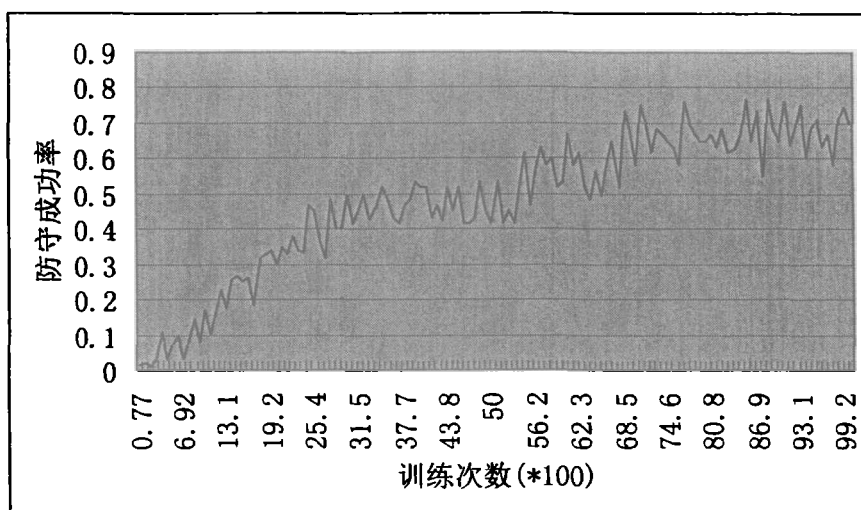


图 4.6 实验的防守成功率曲线

Fig. 4.6 Successful defense rate of the experiment

如图 4.6 所示, 经过学习的防守 Agent 最终的防守成功率稳定在 60~70%。观察实际的实验过程可知, 防守 Agent 的成功防守大多数是因为进攻 Agent 未在 300 个周期之内推进到本方禁区。由于极大极小 Q 学习算法是针对最坏情况的最优策略, 因此应用该策略的防守 Agent 在采取动作时相对比较保守。这种保守的策略经常会导致场上局面的僵持, 即进攻 Agent 没有足够的推进空间向禁区推进, 但是也不会因为防守 Agent 的防守压力后退, 因为防守 Agent 并不会主动抢球, 而是更多的采取防守卡位策略。

(2) 实战效果: 在实际比赛中检验防守效果

将(1)中学习后的防守策略应用到我们的小型组比赛球队 DUT Fantasia SmallSize 的防守策略中。在 2006 年苏州举办的 RoboCup 中国公开赛中, 我们的球队与科大蓝鹰、上大自强以及辽宁科技大学三支球队进行了三场比赛。

科大蓝鹰是国内的小型组机器人足球比赛领域的传统强队, 曾多次获得全国比赛的冠军并多次代表中国参加机器人足球世界杯比赛。在与科大蓝鹰队的比赛过程中, 我们球队的进攻机会虽然不多, 但在我们后卫队员的防守下, 科大蓝鹰也没有获得好的进攻机会, 并且多次出现失误将球带出了边界, 最终比赛以 0: 0 结束。

建队多年的上海大学自强队也是国内实力突出的一支球队, 该队曾经参加了 2006 年在德国举行的机器人世界杯比赛并进入了 16 强。在与上大自强队的比赛中, 我们的后卫队员较好的完成了防守任务, 但由于场上担任守门员的机器人小车电池电量耗光导致无法运动, 最终在没有守门员的情况下以 0: 1 惜败于上海大学队。

辽宁科技大学今年是第二次参加 RoboCup 小型组的比赛。在与该队的比赛过程中，我们基本占据了场上的主动，并打入 1 球，最终以 1: 0 获得比赛的胜利。

从以上结果可以看出，我们决策系统中的防守策略在实验和实战中都取得了较好的效果，由此证明，在解决一对一盯人防守这种对抗性问题时，基于 Markov 对策的强化学习是一种可行的方法。

结 论

机器人足球比赛作为研究机器人学和人工智能的实验平台,近年来得到了迅速的发展。小型组机器人足球比赛作为其中的一个分支,将机械、控制、传感、无线通信、多 Agent、机器学习等多个领域的知识融合到一起,构成一个软件和硬件相结合的系统。本文根据实际的比赛任务,完成了一个小型足球机器人决策系统的设计与实现,其中主要工作如下:

(1) 设计并实现了一个决策系统框架,包括决策系统与视觉系统的通信模式,主决策模块的触发方式以及决策系统与其他各个子系统的通讯模块。

(2) 通过大量的实验,测定了比赛用球和机器人小车在比赛环境中的相关物理模型,为决策系统中的世界模型提供了理论依据。

(3) 实现了决策子系统的核心决策模块程序,包括底层技术和上层策略的实现以及相关战术的制定。

(4) 针对小型组比赛防守决策中的一对一盯人防守问题,应用基于 Markov 对策的强化学习制定了相应的防守策略,并在实际比赛中验证了该策略的有效性。

以本文实现的系统为决策子系统的小型组球队 DUT Fantasia SmallSize 代表大连理工大学参加了 2006 年 10 月举行的 RoboCup 中国公开赛,在比赛过程中,该决策系统取得了较好的实战效果。

由于机器人小车的整体性能还很不完善以及准备时间方面的限制,本文所实现的系统还存在一些不足,还可以在以下方面做进一步研究和改进:

(1) 更加详细多样以及有针对性的战术体系。

(2) 机器人小车运动过程中的路径规划。

(3) 应用基于 Markov 对策的强化学习的防守策略在比赛中在线学习的可行性。

本文实现的系统是应用在大连理工大学小型组机器人足球队 DUT Fantasia SmallSize 的第一代机器人小车上的。由于硬件设施及其它条件的限制,还有很多想法没有实现。希望以后能够做更深入细致的研究,为机器人技术和人工智能理论提供更好的实验平台。

参 考 文 献

- [1] Mackworth A. On Seeing Robots. In: Basu A, Li X. Computer vision: systems, theory, and applications. Singapore: World Scientific Press, 1993:1-13.
- [2] Kitano H, Tambe M, Stone P, et al. The robocup synthetic agent challenge 97. In: Proceedings of IJCAI-97, Nagoya, Japan, 1997:342-355.
- [3] Wooldridge M, Jennings N R. Intelligent agents: theory and practice. Knowledge Engineering Review. 1997, 10(2):115-152.
- [4] Rao A, Georgeff M. BDI agent: from theory to practice. In: Proceedings of the First International Conference on Multi-Agent System, Cambridge, England, 1995:45-52.
- [5] Jennings N R. Specification and implementation of a belief desire joint-intention architecture for collaborative problem solving. Journal of Intelligent and Cooperative Information Systems. 1993, 2(3):289-318.
- [6] Bratman M E, Jsrael D J, Pollack M E. Plans and resource-bounded practical reasoning. Computational Intelligence. 1998, 4(1):349-355.
- [7] Wooldridge M, Jennings N R. Intelligent agents---theories, architectures, and languages. Lecture Notes in Artificial Intelligence, Springer-Verlag, 1995, 890:82-110.
- [8] Brooks R. A robust layered control system for a mobile robot. IEEE Journal of Robotics and Automation. 1986, 2(1):14-23.
- [9] Ferguson I. Towards an architecture for adaptive, rational, mobile agents. In: Proceeding of the Third European Workshop on Modeling Autonomous Agents and Multi-Agent World(MAAMAW-91), Kaiserslautern, Germany, 1992:249-262.
- [10] Kaelbling L, Littman M, Moore A. Reinforcement learning: a survey. Journal of Artificial Intelligence Research. 1996, 4(2):237-285.
- [11] Fisher M, Wooldridge M. Specifying and executing protocols for cooperative action. In: Proceedings of the Second International Working Conference on Cooperating Knowledge-Based Systems, Heidelberg, Germany, 1994:143-156.
- [12] Wooldridge M. An introduction to multiagent systems. Chichester, England : John Wiley and Sons Ltd. 2002.
- [13] Meo P D, Quattrone G, Terracina G, et al. A multi-agent system for the management of e-government services. Intelligent Agent Technology. 2005, 12:78-84.
- [14] Sutton R. Learning to predict by the method of temporal difference. Machine Learning. 1988, 3:9-44.
- [15] Barto A G, Bradtke S J, Singh S P. Learning to act using real-time dynamic programming. Artificial Intelligence. 1995, 72:81-138.

- [16] Riedmiller M. Concepts and facilities of a neural reinforcement learning in cooperative multi-agent systems. *Journal of Neural Computing and Application*. 2000, 8:323-338.
- [17] Bruce J, Balch T, Veloso M. Fast and Inexpensive Color Image Segmentation for Interactive Robot. In: *Proceedings of the 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Kagawa University, Takamatsu, Japan, 2000:2061-2066.
- [18] Simon M, Wiesel F, Egorova A, et al. Plug & play:fast automatic geometry and color calibration for tracking mobile robots. In: *Nardi D., Riedmiller M., Sammut C., et al. RoboCup 2004-Robot Soccer World Cup VIII*, Springer-Verlag, 2004:1354-1368.
- [19] Kim J H, Shim H S, Kim H S, et al. Action selection and strategies in robot soccer system. In: *Proceeding of the 40th Midwest Symposium on Circuits and Systems*, University of California, Davis, 1997:324-347.
- [20] Shim H S, Jung M J, Kim H S, et al. A hybrid control structure for vision based soccer robot system. *Intelligent Automation and Soft Computing*. 2000, 6(1):89-101.
- [21] Latombe J C. *Robot Motion Planning*. Boston: Kluwer Academic Publisher. 1996.
- [22] Khatib O. Real time obstacle avoidance for manipulators and mobile robots. *International Journal of Robotics Reaserch*. 1986, 5(1):90-98.
- [23] Kavraki L, Svestka P, Latombe J C, et al. Probabilistic roadmaps for path planning in high-dimensional configuration spaces. *IEEE Transactions on Robotics and Automation*. 1996, 5:66-88.
- [24] Lavalley S M, Hutchinson S A. Optimal motion planning for multiple robots having independent goals. In: *Proc. IEEE International Conference of Robot and Automation*, Minneapolis, Minnesota, USA, 1996:2847-2852.
- [25] Kitano H, Siekmann J, Carbonell J G. *Robocup-97: Robot Soccer World Cup I*, Springer Verlag, 1998.
- [26] Sargent R, Bailey B, Witty C, et al. Importance of fast vision in winning the First Micro-Robot World Cup Soccer Tournament. *Robotics and Autonomous Systems*. 1997, 21(2): 139-147.
- [27] 周科. RoboCup 小型组(F-180)足球机器人的运动控制和路径规划:(硕士论文). 浙江大学, 2004.
- [28] Brockett R W. Asymptotic stability and feedback stabilization. In: *Brockett R W, Millman R S, Sussmann H J, et al. Differential Geometric Control Theory*, Boston, MA:Birkhuser, 1983.
- [29] Sarkar N, Yun X, Kumar V. Control of mechanical systems with rolling constraints: Application to dynamic control of mobile robots. *International Journal of Robotics Research*. 1994, 13:55-69.
- [30] Bloch A, McClamroch N H, Reyhanoglu M. Controllability and stability properties of a nonholonomic control system. In: *Proceedings of 29th IEEE Conference on Decision and Control*, Honolulu, Hawaii, USA, 1990:587-596.

- [31] Gampion G, Bastin G. Structural properties and classification of kinematic and dynamic models of wheeled mobile robots. In: IEEE Transactions on Robotics and Automation, 1996, 12:47-61.
- [32] 张汝波. 强化学习理论及应用. 哈尔滨: 哈尔滨工程大学出版社. 2001.
- [33] Watkins C J, Doya K. Technical note: Q-learning. Machine Learning. 1992, 8(3):279-292.
- [34] Tsitsiklis, John N. Asynchronous stochastic approximation and Q-learning. Machine Learning. 1994, 16(3):185-202.
- [35] 高阳, 周志华, 何佳周等. 基于 Markov 对策的多 Agent 强化学习模型及算法研究. 计算机研究与发展. 2000, 37(3):257-263.
- [36] Littman M. Markov games as a framework for multi-agent reinforcement learning. In: Proceedings of the 11th International Conference on Machine Learning, San Francisco, CA, 1994:157-163.
- [37] Stone P. Layered learning in multi-agent systems: (PhD Thesis). Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, USA., 1998.
- [38] Yao J Y, Chen J, Cai Y P, et al. Architecture of tsinghua aeolus. In: Birk A., Coradeschi S., Tadokoro S. Robocup-2001: Robot Soccer World Cup V, LNAI2377, Springer, 2002:469-473.
- [39] Riedmiller M, Braun H. A direct adaptive method for faster back-propagation learning: The RPROP algorithm. In: Proceedings of the IEEE International Conference on Neural Networks(ICNN), San Francisco, 1993:586-591.

攻读硕士学位期间发表学术论文情况

[1] 张树林, 孙焘, 冯林. Robocup 半场防守中的一种强化学习算法. 辽宁师范大学学报(自然科学版), 已录用. 论文相关内容为第四章.

[2] Dingming Guo, Shulin Zhang, Lin Feng. Fantasia 2006 Team Description. Proceedings CD RoboCup 2006, Springer-Verlag. 论文相关内容为第三章和第四章.

致 谢

衷心感谢我的导师创新院院长冯林教授，本文的大部分研究和实践工作是在他的悉心指导下完成的。在我的研究生学习和生活中，冯林教授给予了我无微不至的关怀和无私的帮助指导，使我受益匪浅。他严谨的学风、渊博的知识、开阔的视野和敏锐的洞察力以及兢兢业业、一丝不苟的治学态度，给了我深深的影响和启迪。

衷心感谢创新院的孙焘老师。在研究生期间的项目和比赛过程中，孙焘老师给予了我细致入微的指导。孙老师在研究上开阔的思路、广博的知识和独到的见解以及严谨的治学理念，都给我留下了深刻的印象。

衷心感谢创新院的吴振宇老师。在 RoboCup 小型组球队开发和参加比赛的过程中，吴老师在设备条件上的坚定支持，在技术上的悉心指导，在生活上无微不至的关怀，使我能够顺利完成小型组比赛球队的开发。

衷心感谢创新院的同学郭定明、范成涛、刘军。在共同学习和生活的两年多时间里，他们给予了我无私的帮助和支持，使我度过了愉快的研究生学习生活。

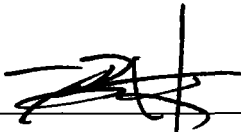
衷心感谢参与到 RoboCup 小型组项目中的李璞、刘崇杰、黄甫毅杰、崔超和黄泉生等同学，他们所作的工作为我的研究和论文提供了大力的支持。

最后，要感谢我的家人在读研期间对我一如既往的关心和支持，他们使我在面对困难时没有退缩，给了我前进的动力和勇气，使我顺利地完成了学业。

大连理工大学学位论文版权使用授权书

本学位论文作者及指导教师完全了解“大连理工大学硕士、博士学位论文版权使用规定”，同意大连理工大学保留并向国家有关部门或机构送交学位论文的复印件和电子版，允许论文被查阅和借阅。本人授权大连理工大学可以将本学位论文的全部内容编入有关数据库进行检索，也可采用影印、缩印或扫描等复制手段保存和汇编学位论文。

作者签名： 张树林

导师签名： 

2006 年 12 月 18 日