摘 要

随着机器人技术发展的愈发成熟，机器人在我们生活中扮演着越来越重要的角色。在机器人相关技术领域中，路径规划是非常重要的一项技术。在理论中，我们可以将地图数据传送给机器人进行分析，从而得到最佳路径，但是在实际应用中，机器人是不可能知道一个未知环境中的地理信息的，因此，对于在陌生环境中的，不具有先验知识的机器人，研究其探索和规划路径功能是具有非常重要的意义的。

本文以机器人在未知环境中的路径规划为背景，学习强化学习算法实现机器人的路径规划。在原有的强化学习算法Q-learning基础之上，提出改进，提高性能，完成规避障碍，求得最优路径功能。

**关键词：**Q-learning，移动机器人，路径规划，避障，强化学习，最优路径

Abstract

With the development of increasingly robotics, robots play an increasingly important role in our lives. In the related field of robotics, the path planning is a very important technology. In theory, we can transport map data to the robot for analysis to obtain the best path. However, in practice, it is impossible to know the map data in an unknown environment, geographic information. Therefore, the study which a robotic without prior knowledge explore and find the best way in an unfamiliar environment is a very important significance.

In this paper, robot path planning in unknown environment as the background, we learn reinforcement learning algorithm to achieve robot path planning. Based on the original reinforcement learning Q-learning algorithm, we propose improvements to enhance performance, complete obstacle avoidance, get the optimal path function.

**Keywords:** Q-learning, obstacle avoidance, reinforcement learning, the optimal path

目 录

[摘 要 I](#_Toc451329785)

[Abstract II](#_Toc451329786)

[目 录 III](#_Toc451329787)

[第一章 引言 1](#_Toc451329788)

[1.1 研究背景及意义 1](#_Toc451329789)

[1.2 论文结构及安排 1](#_Toc451329790)

[第二章 多载波系统传输方案 3](#_Toc451329791)

[2.1 多载波系统概述 3](#_Toc451329792)

[2.1.1 单载波传输系统 3](#_Toc451329793)

[2.1.2 多载波传输系统 3](#_Toc451329794)

[2.2 OFDM系统的介绍和分析 4](#_Toc451329795)

[2.2.1 OFDM技术介绍 4](#_Toc451329796)

[2.2.2 OFDM系统的原理和基本模型 4](#_Toc451329797)

[2.2.3 OFDM系统的优缺点 5](#_Toc451329798)

[2.2.4 OFDM系统中的同步要求 5](#_Toc451329799)

[2.3 SIM-OFDM系统的介绍和分析 5](#_Toc451329800)

[2.3.1 SIM-OFDM 技术介绍 5](#_Toc451329801)

[2.3.2 SIM-OFDM 原理和基本模型 5](#_Toc451329802)

[2.3.3 SIM-OFDM的优缺点 5](#_Toc451329803)

[2.4 通信系统性能指标 5](#_Toc451329804)

[2.4.1 有效性 5](#_Toc451329805)

[2.4.2 可靠性 6](#_Toc451329806)

[2.4.3 有效性和可靠性的关系 6](#_Toc451329807)

[2.4.4 误码率分析 6](#_Toc451329808)

[2.5 本章小结 6](#_Toc451329809)

[第三章SIM-OFDM系统中的误码率分析 7](#_Toc451329810)

[3.1 SIM-OFDM系统误码率分析 7](#_Toc451329811)

[3.1.1 能量判决方法 7](#_Toc451329812)

[3.1.2 联合界方法 7](#_Toc451329813)

[3.1.3 UPPER BOUND 方法 7](#_Toc451329814)

[3.1.4 基于最大似然检测的误码率分析 8](#_Toc451329815)

[3.2 仿真结果及分析 8](#_Toc451329816)

[3.2.1 能量判决方法仿真 8](#_Toc451329817)

[3.2.2 联合界方法仿真 9](#_Toc451329818)

[3.2.3 UPPER BOUND 方法仿真 10](#_Toc451329819)

[3.2.4 基于最大似然算法的误码率分析仿真 10](#_Toc451329820)

[3.3 本章小结 10](#_Toc451329821)

[第四章 SIM-OFDM系统的载波频率偏移的概念及分析 11](#_Toc451329822)

[4.1 载波频率偏移下的OFDM系统 11](#_Toc451329823)

[4.1.1 OFDM系统的频偏模型 11](#_Toc451329824)

[4.1.2 OFDM系统有频偏的误码率分析 11](#_Toc451329825)

[4.2 载波频率偏移下的SIM-OFDM系统 11](#_Toc451329826)

[4.2.1 SIM-OFDM系统的频偏模型 11](#_Toc451329827)

[4.2.2 频偏对SIM-OFDM系统的误码率分析 12](#_Toc451329828)

[4.3 仿真结果及分析 12](#_Toc451329829)

[4.3.1 频偏下的OFDM系统 12](#_Toc451329830)

[4.3.2 频偏下的SIM-OFDM系统 12](#_Toc451329831)

[4.3.3频偏下的SIM-OFDM与OFDM系统对比 12](#_Toc451329832)

[4.4 本章小结 12](#_Toc451329833)

[第五章 结束语 13](#_Toc451329834)

[5.1本文内容 13](#_Toc451329835)

[5.2下一步学习工作方向 13](#_Toc451329836)

[参考文献 14](#_Toc451329837)

[致谢 16](#_Toc451329838)

[外文资料原文 17](#_Toc451329839)

[外文资料译文 18](#_Toc451329840)

第一章 引言

伴随着社会的发展和进步，人类的生活越来越需要机器人的辅助，因此越来越多的国家都在机器人领域投入很大的精力进行研究。

1.1 研究背景及意义

移动机器人是一种在复杂环境下工作，具有自行组织、自主运行、自主规划的智能机器人，它集中了传感器技术、信息处理、电子工程、计算机工程、自动化控制工程以及人工智能等多学科的研究成果，代表机电一体化的最高成就，是目前科学技术发展最活跃的领域之一。随着机器人性能不断地完善，移动机器人的应用范围大为扩展，不仅在工业、农业、医疗、服务等行业中得到广泛的应用，而且在城市安全、国防和空间探测领域等有害与危险场合得到很好的应用。因此，移动机器人技术已经得到世界各国的普遍关注。

移动机器人根据移动方式来分，可分为：轮式移动机器人、步行移动机器人单腿式、双腿式和多腿式、履带式移动机器人、爬行机器人、蠕动式机器人和游动式机器人等类型；按工作环境来分，可分为：室内移动机器人和室外移动机器人；按控制体系结构来分，可分为：功能式水平式结构机器人、行为式垂直式结构机器人和混合式机器人；按功能和用途来分，可分为：医疗机器人、军用机器人、助残机器人、清洁机器人等

移动机器人的研究始于60年代末期。斯坦福研究院的和Charles Rosen等人，在1966年至1972年中研发出了取名为Shakey的自主移动机器人，目的是研究应用人工智能技术，在复杂环境下机器人系统的自主推理、规划和控制。我国机器人研究开始于世纪年代，至今已有多年。年，北京机械工业自动化研究所曹祥康在江苏嘉兴召开了全国性机械手技术交流大会，是我国第一个以机器人为主题的大型会议。比较有代表性的有：清华大学研制了智能移动THMR-III V型机器人；中科院自动化研究所研制出智能保安机器人，全方位移动式机器人视觉导航系统，智能移动机器人通用平台AIMR;香港城市大学研发了自动导航车及服务机器人；哈尔滨工业大学研制了导游机器人；国防科技大学研发成功了双足机器人等；“海龙2号”是我国自主研制的水下机器人，是我国目前仅有能在3500米水深、海底高温和复杂地形的特殊环境下幵展海洋调查和作业的最高精技术装备，它除了在潜水深度上的优势之外，还在国际上首次采用了一些自主研发的先进技术，包括虚拟控制系统和动力定位系统。

移动机器人导航移动机器人学的关键技术，是指按照预先给出的任务，根据已知的地图信息作出全局路径规划，并在行进过程中，不断感知周围的局部环境信息，自主地作出各种决策，随时调整位姿，引导自身安全行驶或跟踪已知路径，达到目标位姿，是移动机器人各项研究应用的基础和前提。自主导航是移动机器人导航的一项重要的基本能力，它主要解决的问题可以归纳为三个方面：“1.现在何处？2.要往哪去？3.如何去？”也就是移动机器人的地图构建、定位和路径规划题。地图构建是移动机器人在自主主导航的过程中，通过传感器感知环境并建立环境模型，这是述立环境模型最常用且有效的方法；定位移动机器人在工作环境中探索所处位置的过程，也就是机器人的全局坐标和姿态；路径规划技术就是移动机器人在具有障碍物的环境上按照一定的评价标准（如作代价最小、行走路径最短、行走时间最短等）、寻找一条从给起点到达目标终点的无碰路径

但是不管采取何种导航方式，移动机器人上主要是完成路校规划、定位和避障等任务。路径规划问题是移动机器人自助式导航技术研究中的关键技术之，它是指移动机器人按照某一性能指标，搜索一条从起始位置到目标位置的最优或次优的无碰路径。

随着计算机、传感器及控制技术的发展，特别是各种新算法的不断涌现，移动机器人路径规划技术已经取得了丰硕的研究成果，特别是在周围环境未知或部分未知的局部路径规划（人工势场法、模糊逻辑算法、神经网络法、进化算法、强化学习等），国内外学者已经作了大量的研究和改进，下面综述中将对与本文有关的一些方法进行讨论。

1986年，Khatib首次提出用人工势场法解决机器人避障问题，在此之后许多学者不仅发展了这一方法并将其应用于机器人的实时路径规划和足球机器人的运动控制。人工势场法可以实现快速的控制，所以人工势场法丨被广泛的应用于实时的运动控制。

E.Shi等进一步对传统的人工势能场方法进行改进，他提出的方法虽然能够使移动机器人避幵障碍物，但其实施性并不强，虽然克服了零势能场（在这种情况下机器人无法到达目标位置）的情况，但又出现了很多问题，比如障碍物的影响范围、引力增益和斥力增益会影响到机器人运动过程中振动的程度、运动路径的平滑度、路径的长短及其运动过程中是否会碰撞到障碍物等，所以必须同时选择这三个系数的最优值，而在实验过程中是很难选取的，必须通过反复的实验才能选取较为合适的值；王奇志等提出了一种改进的人工势场法，通过排除一个距离机器人最远的障碍物，同时加一个同等大小反方向力来消除零势能点，从而实现消除零势能域，达到多障碍物情况下机器人运动规划的快速、实时、避障的效果，结果表明，该算法对多个障碍物和非静态的障碍物同样适用。陈华志等采用一种具有速度负反馈的基于模糊控制的移动机器人避障算法进行研究和仿真对移动机器人在实际环境中进行了实验，并证明了方法的可行性；Lee等使用模糊控制来决定移动机器人可能的十三个移动方向，使机器人在这种导航下避免机器人内部碰撞；Pradhan等在完全未知的环境中使用模糊逻辑对多达1000个机器人进行了导航实验，在试验中，对四个输入变和两个输出变贤进行模糊化，最后证明了高斯隶属函数在移动机器人导航中最有效；Yang等设计了移动机器人的动态模型，根据多机器人的局部信息，建立了引力、斥力函数的模糊控制器，基于最优控制理论，来衡量系统的状态性能指标；刘利等针对海域不同的障碍环境，从运动学特性出发，提出了一种多主水航行器（AUV）编队避障策略。该策略是在领航跟随法的基础上，利用编队队形的几何关系，引入虚拟AUV，再采用模糊控制避障策略，使编队能顺利通过障碍物，到达目的地；Rigatos等研究了在包含固定障碍物和移动物体的部分未知的环境下，根据移动机器人的运动问题和控制规律，结合滑膜与模糊逻辑控制的基本原则，提出使机器人达到期望的运动效的控制器，即降低滑动模式模糊逻辑控制器的复杂度，具有稳定性和简易性；Montaner等为移动机器人导航设计了个模糊逻辑控制器，该控制器可以接收确定性和模糊的信息，并在一个由七个超声波传感器来获信息的环境中对移动机器人进行了实验研究了移动机器人在未知群集环境屮反应式导航的问题，中文把反应式导航定义为感知数据和命令的关系，建立一个反应式导航相当于对机器人提供了一个运动的地图，在此基础上提出了另一种“普遍”的逼近方法—模糊逻辑，并揭示了如何利用行为分解的方法来选择模糊规则；Tan等提出了一种可加速减速的模糊控制器模型，使移动机器人在动态环境中自主导航，包括寻找目标、躲避障碍和寻找最优路径。

Yang和Meng提出了一种离散拓扑结构的祌经网络模型，移动机器人使这种并联的祌经网络模型在离散化的完全已知环境中进行路径规划，并假定机器人的速度恒定，这在真实的环境中对移动机器人进行控制是没意义的；Zhu等提出了种基于模糊神经网络将传感器信息与机器人的移动结合起来的导航方法，这种力法可以使移动机器人充分感受周围环境、自主避开静态和动态障物，并在不同情况下避开“死循环”产生到达目标的合理的轨迹，通过仿真证明了该方法的实用性和有效性；Nishida等基于神经元系统提出了种感知预测机器人伙伴的控制方法，该预测方法对于减少计算量、提取感知信息是非常重要的；Kala等在文中提出了遗传算法在移动机器人导航中的应川，使机器人可用于业界派调查，数据采集，移动机器人可以在动态环境中避免内部碰撞证明了这种算法的有效性。

1.2 论文结构及安排

本文的主要内容是研究和分析Q-learning强化学习算法使得机器人在未知环境中，通过自主学习环境信息，得出到达目标位置的最佳路径。并基于Q-learning算法，提出新的算法，使得在保证准确性的前提下性能得到优化。

本文的主要结构安排如下：

第一章主要是介绍了机器人路径规划意义，发展史，和在现实生活中的应用情况。

第二章主要是介绍了移动机器人路径规划算法以及强化学习算法的研究。阐述了目前常见的移动机器人算法和强化学习算法在其领域的实现

第三章主要是在分析了Q-learning强化学习算法基础之上，提出自己新的改进算法。

第四章内容为对经典算法和自己提出的算法进行比较，性能分析

第五章为结束语，对本次论文撰写工作进行总结

第二章 移动机器人路径规划及强化学习算法

2.1 移动机器人路径规划

机器人相关学习中的一个非常重要的研究领域就是机器人导航，所有类型的机器人，在工作的时候，一般都是根据特定标准来找出一条最优或者接近最优的路径行走。这些标准常见的有：距离短、时间少、性能高、能耗少等。 在机器人寻路过程中，也会有一些限制条件，比如必须经过某一点，躲避障碍等等，我们要在这些限制条件下，对算法进行优化。来算法，使得在保证准确性的前提下性能得到优化。

在机器人导航中，路径规划又是它的重要任务之一。路径规划可以分类为

1. 已知环境，静态障碍物路径规划
2. 未知环境，静态障碍物路径规划
3. 已知环境，动态障碍物路径规划
4. 未知环境，动态障碍物路径规划。

除了这种分类方式意外，还可以根据对寻路环境的掌握情况，分为两类

1. 环境信息完全掌握的全局路径规划
2. 基于传感器的局部路径规划

本小节主要介绍：全局路径规划技术，局部路径规划技术。

2.1.1 全局路径规划技术

全局路径规划，就是寻路周围的环境已知，机器人根据已知的环境制定一条从初始位置到目标位置的最优路径。其主要方法有：可视图法，自由空间法，最优控制法，栅格法，拓扑法等。

1. 可视图法

可视图法视移动机器人为一点，将机器人、目标点和多边形障碍物的各顶点进行组合连接，并保证这些直线均不与障碍物相交，这就形成了一张图，称为可视图。由于任意两直线的顶点都是可见的，从起点沿着这些直线到达目标点的所有路径均是运动物体的无碰路径。搜索最优路径的问题就转化为从起点到目标点经过这些可视直线的最短距离问题。运用优化算法，可删除一些不必要的连线以简化可视图，缩短搜索时间。该法能够求得最短路径，但假设忽略移动机器人的尺寸大小，使得机器人通过障碍物顶点时离障碍物太近甚至接触，并且搜索时间长。切线图法和Voronoi图法对可视图法进行了改造。切线图用障碍物的切线表示弧，因此是从起始点到目标点的最短路径的图，即移动机器人必须几乎接近障碍物行走。其缺点是如果控制过程中产生位置误差，移动机器人碰撞的可能性会很高。Voronoi图法用尽可能远离障碍物和墙壁的路径表示弧。由此，从起始节点到目标节点的路径将会增长，但采用这种控制方式时，即使产生位置误差，移动机器人也不会碰到障碍物。

1. 拓扑法

将规划空间分割成具有拓扑特征子空间，根据彼此连通性建立拓扑网络，在网络上寻找起始点到目标点的拓扑路径，最终由拓扑路径求出几何路径。拓扑法基本思想是降维法，即将在高维几何空间中求路径的问题转化为低维拓扑空间中判别连通性的问题。优点在于利用拓扑特征大大缩小了搜索空间。算法复杂性仅依赖于障碍物数目，理论上是完备的。而且拓扑法通常不需要机器人的准确位置，对于位置误差也就有了更好的鲁棒性；缺点是建立拓扑网络的过程相当复杂，特别在增加障碍物时如何有效地修正已经存在的拓扑网是有待解决的问题。

1. 栅格法

将移动机器人工作环境分解成一系列具有二值信息的网格单元，多采用四叉树或八叉树表示，并通过优化算法完成路径搜索。该法以栅格为单位记录环境信息，有障碍物的地方累积值比较高，移动机器人就会采用优化算法避开。环境被量化成具有一定分辨率的栅格，栅格大小直接影响环境信息存储量大小和规划时间长短。栅格划分大了，环境信息存储量小，规划时间短，但分辨率下降，在密集环境下发现路径的能力减弱；栅格划分小了，环境分辨率高，在密集环境下发现路径的能力强，但环境信息存储量大，规划时间长。

栅格法是由w．E．Howden在1968年提出的。栅格法将机器人工作环境分解成一系列具有二值信息的网格单元，工作空间中障碍物的位置和大小一致，并且在机器人运动过程中，障碍物的位置和大小不发生变化。用尺寸相同的栅格对机器人的二维工作空间进行划分，栅格的大小以机器人自身的尺寸为准。若某个栅格范围内不含任何障碍物，则称此栅格为自由栅格；反之，称为障碍栅格。自由空间和障碍物均可表示为栅格块的集成。栅格的标识方法有两种：直角坐标法和序号法。多采用四叉树或八叉树表示工作环境，并通过优化算法完成路径搜索。该方法以栅格为单位记录环境信息，栅格粒度越小，障碍物的表示越精确，但同时会占用大量的存储空问，算法的搜索范围将按指数增加。栅格的粒度太大，规划的路径会很不精确。所以栅格粒度的大小的确定，是栅格法的主要问题。

1. 自由空间

自由空间应用于移动机器人路径规划，采用预先定义的如广义锥形和凸多边形等基本形状构造自由空间，并将自由空间表示为连通图，通过搜索连通图来进行路径规划。自由空间的构造方法是：从障碍物的一个顶点开始，依次作其它顶点的链接线，删除不必要的链接线，使得链接线与障碍物边界所围成的每一个自由空间都是面积最大的凸多边形：连接各链接线的中点形成的网络图即为机器人可自由运动的路线。其优点是比较灵活，起始点和目标点的改变不会造成连通图的重构，缺点是复杂程度与障碍物的多少成正比，且有时无法获得最短路径用栅格法建模受到了空间分辨率和内存容量的矛盾限制。而自由空间法建模，解决了这一矛盾。但自由空间法的分割需构造想象边界，想象边界本身具有任意性，于是导致路径的不确定性。

1. 最优控制法

在确定的空间里，二维平面上的一条边界可由方程来描述。那么，机器人在运动过程中，从起点到终点的众多路径里，有障碍物的路径是不允许机器人通过的。这些路径可以作为约束条件，由数学表达式表示。非完整移动机器人通过适当的变换，可将其转化为链式形式。因此，通过选择适当的控制量就可以驱使机器人从一个位置运动到另一个位置。

2.1.2 局部路径规划

(1) 人工势场法

人工势场法是Khatib提出的一种虚拟力法。其基本思想是将移动机器人在环境中的运动视为一种虚拟人工受力场中的运动。障碍物对移动机器人产生斥力，目标点产生引力，引力和斥力周围由一定的算法产生相应的势，机器人在势场中受到抽象力作用，抽象力使得机器人绕过障碍物。该法结构简单，便于低层的实时控制，在实时避障和平滑的轨迹控制方面，得到了广泛应用，其不足在于存在局部最优解，容易产生死锁现象，因而可能使移动机器人在到达目标点之前就停留在局部最优点。

主要的缺陷：

1)陷阱区域

2)在相近的障碍物之间不能发现路径。

3)在障碍物前振荡。

4)在狭窄通道中摆动。

针对这些缺陷，提出了一些改进办法。针对人工势场法存在“机器人在到达目标位置前由于陷入局部极小点而无法到达目标位置”的问题，解决的方法有：重新定义势函数，使之没有或有更少的局部极小点；利用搜索算法跳出局部极小点。还可以利用模拟退火算法使势函数跳出局部极小点，到达机器人的目标位置。

(2) 模糊逻辑算法

模糊逻辑算法基于对驾驶员的工作过程观察研究得出。驾驶员避碰动作并非对环境信息精确计算完成的，而是根据模糊的环境信息，通过查表得到规划出的信息，完成局部路径规划。优点是克服了势场法易产生的局部极小问题，对处理未知环境下的规划问题显示出很大优越性，对于解决用通常的定量方法来说是很复杂的问题或当外界只能提供定性近似的、不确定信息数据时非常有效。假设检测的是障碍物与机器人的距离和障碍物的运动信息，输出机器人速度变化和转角变化。

模糊控制算法有诸多优点，但也有固有缺陷：人的经验不一定完备；输入量增多时，推理规则或模糊表会急剧膨胀。由于模糊隶属度函数的设计、模糊控制规则的制定主要靠人的经验和试凑，总结模糊控制规则时比较困难，而且，控制规则一旦确定，在线调整困难，无法很好地适应情况的变化。因此，如何得到最优的隶属度函数、控制规则以及对控制规则进行在线调整是该方法最大的问题。

(3) 模拟退火算法

模拟退火算法由Kirkpatrick S于1983年提出，源于物理退火过程。基本思想是利用随机优化问题求解过程与统计力学中热平衡问题的相似性，通过设定初温、初态和降温率控制温度的不断下降，结合概率突跳特性，利用解空间的邻域结构进行随机搜索。模拟退火算法用于路径规划可避免局部极值，但其理论收敛条件过于苛刻，在实际应用中往往无法满足。在有限计算量条件下的收敛性能依赖于自身参数，这使得参数设定成为算法应用过程中的一个关键环节。

(4) 神经网络法

路径规划是感知空间到行为空间的一种映射。映射关系可用不同方法实现，很难用精确数学方程表示，但采用神经网络易于表示。将传感器数据作为网络输入，由人给定相应场合下期望运动方向角增量作为网络输出，由多个选定位姿下的一组数据构成原始样本集，经过剔除重复或冲突样本等加工处理，得到最终样本集。将神经网络和模糊数学结合也可实现移动机器人局部路径规划。先对机器人传感器信息进行模糊处理，总结人的经验形成模糊规则。再把模糊规则作用于样本，对神经网络进行训练。通过学习典型样本，把规则融会贯通，整体体现出一定智能。实际中允许输入值偏离学习样本，只要输入接近一个学习样本的输入模式，则输出也就接近该样本输出模式。该性质使得神经网络可以模仿人脑在丢失部分信息时仍具有对事物正确的辨识力。

(5) 遗传算法

遗传算法以自然遗传机制和自然选择等生物进化理论为基础，构造了一类随机化搜索算法。利用选择、交叉和变异编制控制机构的计算程序，在某种程度上对生物进化过程作数学方式的模拟。只要求适应度函数为正，不要求可导或连续，同时作为并行算法，其隐并行性适用于全局搜索。多数优化算法都是单点搜索，易于陷入局部最优，而遗传算法却是一种多点搜索算法，故更有可能搜索到全局最优解。遗传算法的整体搜索策略和优化计算不依赖于梯度信息，解决了一些其它优化算法无法解决的问题。遗传算法运算速度不快，进化众多的规划要占据较大存储空间和运算时间，优点是克服了势场法的局部极小值问题，计算量不大，易做到边规划边跟踪，适用于时变未知环境的路径规划，实时性较好。遗传算法运用于移动机器人路径规划的研究近来取得了许多成果，其基本思想：将路径个体表达为路径中一系列中途点，并转换为二进制串。首先初始化路径群体，然后进行遗传操作，如选择、交叉、复制、变异。经过若干代进化以后，停止进化，输出当前最优个体。

(6) 蚁群算法

蚁群优化算法是由意大利学者Dorigo等人在2O世纪9O年代从蚁群觅食行为受到启发，通过模拟自然界蚂蚁寻径的行为，提出的一种全新的模拟进化算法。蚁群优化算法在并行运行环境(如网格环境)下可以同步寻优，加快了寻优速度。另外，它是一种通用性强的算法，稍加修改便可用于其他优化问题。但计算量较大，搜索时间较长，易于陷入局部最优解。

(7) 粒子群算法

粒子群算法是由Kennedy博士和Eberhart博士于1995年从鸟类的捕食行为中受到启发提出的一种基于群体的智能随机优化算法。粒子群算法具有收敛速度快、算法简单、容易编程实现和鲁棒性强等特点，但是，粒子群算法也有一些缺陷，一是容易陷入局部极值点，导致得不到全局最优解；二是粒子群算法本身的参数设置，若参数选择不当，会导致寻优过程中粒子的多样性迅速消失，造成算法“早熟收敛”。

(8) 启发式搜索方法

启发式方法的最初代表是A\*算法 ，其新发展是D\*和Focussed D\*。这2种由Stentz A提出的增量式图搜索算法的产生。D\*算法可以理解为动态的最短路径算法，而Focussed D\*算法则利用A\*算法的主要优点即使用启发式估价函数，2种方法都能根据机器人在移动中探测到的环境信息快速修正和规划出最优路径，减少了局部规划的时间，对于在线的实时路径规划有很好的效果。此外，还出现了一些基于A\*的改进算法，它们一般都是通过修改A\*算法中的估价函数和图搜索方向来实现的，可以较大地提高路径规划的速度，具有一定的复杂环境自适应能力。

2.2 强化学习算法

在机器学习领域内，随着反馈的不同，可以分为：监督学习（supervised learning），非监督学习（Unsupervised learning）和强化学习（Reinforcement）三类。本文要讲的强化学习，是以环境反馈作为输入信息，不断适应环境的方法。

2.2.1 强化学习基本原理

对于强化学习，我们可以把他看是一个不断“试探”环境的过程，并且实时“评价”自己的试探:首先，智能体（也就是学习系统）感受周围的环境“状态”，然后对环境施加一个“动作”，在动作之后，环境会产生改变，同时生成一个奖励值(可能是正，也可能是负)反馈给智能体（即学习系统），最后智能体凭借强化信号以及当前的环境状态作出评估，再进行下一个动作，选择动作的基本原则就是使得即将收到的奖励概率变大。

强化学习系统应该具有的特点：

1. 适应性，智能体不仅仅适用于某种特定环境，而是可以根据环境的反馈而改善性能
2. 反应性，智能体可以根据以往的经验来判断对当前状态应采取什么行为
3. 对教师信号依赖少，智能体从内置的强化规则生成强化信号，再根据强化信号进行学习

强化学习系统的基本结构，由“环境”和“智能体(Agent)”组成。智能体通过接收外部信号（比如通过传感器）来感知所处环境，然后对环境采取相应行为。这里的环境，指的是，所有与该智能体产生交互行为的物体。

在与环境交互过程中，智能体通过反复“试探”环境来学习合适的行为，这就是强化学习。它不同于监督式学习和非监督式学习，而是介于他们之间，是和策略紧密联系的学习，通过不断“试探”，从而学习到最优的策略。

强化信号（可能为奖励或者惩罚）是环境在被智能体的某个行为影响了之后反馈给智能体的。当奖赏值为正，那么智能体在之后会更加倾向于采取此行为；反之，如果奖赏值为负，那么智能体在此之后会不倾向于采取此行为



智能体处在环境中，并与之交互，每一个交互周期都会发生如下事件：

1. 智能体在*t*时刻接受环境信息，得到环境状态*s(t);*
2. 根据当前状态和回报*r(t)*选择动作*a(t)*，并执行
3. 智能体做出的动作，对环境产生影响，环境状态发生变化；
   1. 环境状态更新到新的状态*s(t+1)*
   2. 给出回报*r(t)*,有可能为正，也有可能为负
4. 回报*r(t)*传递给智能体，*t←t+1*
5. 如果新的状态为结束状态，则停止循环，否则转第二步

其中，环境状态*s(t)*和智能体的行为*a(t)*共同决定了回报*r(t)*

用一个四元数组来描述强化学习系统

*RL=<S,A,W,R>*

其中，*S=(,…,)*为智能体遇到的环境状态的集合；*A=(,…,)*为智能体的动作的合集；W表示环境状态转移:*W:S×A→S*; *R=(,…,)*表示环境因不同行为对智能体产生的强化信号（也就是奖赏值），在每一个交互周期内，环境和智能体交互都会获得环境状态，智能体根据获得的环境状态来选择一个行为，在*t+1*时刻，环境状态更新为，智能体会根据更新的环境状态获得一个奖赏值，强化学习最终会找到一个最优策略，使得智能体得到的奖励值总和最大。

上述内容可以表示为：四个元素来表示一个强化学习系统：环境模型，奖赏函数，状态值函数以及控制策略

1. 控制策略（Policy，又称决策函数）：是强化学习的核心,指的是映射（状态到动作），规定在状态*s*下智能体应该选择动作*a*。策略的好坏决定了只能提的行为以及系统的整体性能。
2. 奖赏函数（Reward Function，即强化信号）:是智能体将行为作用在环境上，环境对智能体的该行为产生的反馈信号，也就是奖赏信号R，奖赏信号可正可负，如果为正，那么表示动作为好；如果为负，那么表示动作并不好，并且负值越大，说明此行为越差
3. 状态值函数（Value Function，又称评价函数）:是评价“状态-动作”对的指标
4. 环境模型（Environment Model）：外界环境经过一定处理后，一种能被智能体识别的形式，就建立了环境模型。环境模型会因为智能体选择的行为而改变，环境模型也会给出下一状态和奖赏信号的预测值。这样，只能提就可以根据环境模型对行为的选择进行预测，从而选出比较好的行为。

在早期，强化学习可以称为“试错（Trial-and-Error）”的方式，自从将环境模型和规划引入之后，强化学习就更倾向于动态规划。试错和规划都是强化学习中智能体获得经验的一个过程。

2.2.2 Markov决策过程

在路径规划中，强化学习算法的智能体的每一步都是离散的，因此，我们可以对强化学习中的智能体进行数学建模。马尔科夫决策过程（Markov Decision Process，MDPs）是很多强化学习算法的基础，因为马尔科夫决策过程就是有限状态的，离散的。常见的马尔科夫模型有：

|  |  |  |  |
| --- | --- | --- | --- |
| 马氏模型 | | 是否智能系统行为控制环境状态转移 | |
| 否 | 是 |
| 是否环境为部分可感知 | 否 | 马尔科夫链 | 马氏决策过程 |
| 是 | 隐马尔科夫模型 | 部分感知马氏决策过程 |

在本文讨论范围内，环境模型对于智能体来说是未知的，并且是可以被感知的，所以我们可以认为环境模型是马尔科夫型的。所以我们可以用马氏决策过程（Markov Decision Process，MDP）来给强化学习问题建模。

我们用四元组*<S,A,R,W>*来定义马氏决策过程。*S*是环境状态集合，*A*是系统行为集合，状态转移函数*W: S×A→S*和奖赏函数*R:S×A→R*。我们把系统在状态通过使用行为来转移到状态获得的瞬时奖赏值记为，系统在状态采用使环境状态转移到的概率定义为。

马氏决策过程的实质就是：智能体在当前状态下，转移到下一个状态的概率和因此产生的奖励值，仅仅由目前状态和选择的行为决定，和以前的状态和行为无关。所以在指导W和R的基础上，可以采用动态规划技术来得到最好的策略。而强化学习则是侧重在W和R未知的情况下，只能提如何自己学习并得到最好的策略。

怎么定义环境模型是学习系统面临的一个重要问题，因为事先并不知道W和R，智能体只能够依赖在每一个“试探”周期获得的奖励值来选择合适的策略。在选择行为的时候，为了解决环境模型的波动性和目标的长远性，要在奖励值和行为策略之间构造函数，也就是评价值函数，用评价值函数来进行策略的选择

利用一式我们可以得到一个返回函数，这个返回函数是用来得出在一次学习周期中，在策略下，从状态之后获得的所有奖赏值得累计折扣和。因为环境是未知的，所以每一次学习循环中，在策略下，系统的奖赏值累计和有可能是不一样的。所以利用第二个式子，来考虑不同学习周期中所有返回函数的期望:在某策略下，只能提所获得的期望的累计奖赏折扣。

根据Bellman工时，在我们选定的最优策略下，并且在状态s下，值函数的定义为：

如果P和R已知的情况下，从随意设置的策略出发，为了逼近最优的和，我们可以采用策略迭代的方案：

因为在强化学习中，转移函数P和奖赏函数R是不知道的，所以学习系统不能通过上两个式子进行计算值函数。实际中，如果要进行值函数的估计，常用的方法是采用逼近的方法，比如Monte Carlo采样。

2.2.3 动态规划方法

动态规划方法，又称之为分阶段决策。也就是说能够在做出下一个决策之前，预测这个决策产生的结果(并不能完全预测，有误差)。动态规划方法就是强化学习中最重要的技术。动态规划有两个常用的主要方法，值迭代和策略迭代。虽然是两种方法，但是都是利用动态规划求解，不过一个求最优策略，另一个求最优值。目前比较常用的方法就是值迭代。

2.2.3.1 值迭代

研究值迭代的目的并不是为了找到最优策略，而是找出在当前学习周期内，最好的一个值函数所采取的行为。定义回报折扣和为

这个回报折扣和表示：在当前状态下，采取行为和后续策略得到的值。时，可以得到与之相应得最优评价函数，则Bellman方程可以写作：

值迭代算法具体流程如下：

1. 对于所有的，选择随意的初始值；
2. 对于所有的，计算公式为
3. 一直进行前一步操作，直到

我们这里设为一个特别小的值，这样就可以使得充分接近也就是最优解。

2.2.3.2 策略迭代

策略迭代的目的就是找到最优策略，也就是通过不断迭代从而不断提高策略。

1. 从任意策略开始
2. 对所有的状态和行动
3. 对所有状态，使用下列式子来计算当前状态的策略
4. 一直重复2、3步，直到策略值不再发生改变

2.5 本章小结

本章主要介绍了一些基础模型和背景知识。……

第三章强化学习算法在机器人路径规划中的应用

本章节主要研究经典Q-learning强化学习算法分析，以及在此基础上提出改进算法

3.1 Q-learning算法

Q-learning算法是强化学习中比较常见的一种无模型的迭代算法，1989年，第一次被Watkins提出。该算法第一步就是设置一个Q值表，在智能体与环境交互过程中，通过环境信息的反馈，对智能体的“状态-动作”对形成或正或负的奖励，一个接近最佳路径的行为，对应的Q值会不断增加，偏离最佳路径的行为，对应的Q值会不断减小，通过对Q值的选择，使得智能体的选择不断接近最优行为。在马尔科夫尼科夫规则下，该算法的Q值是可以收敛的。

3.1.1 Q-learning算法的基本原理

Q-learning算法常常采用数值迭代计算来逼近最优值，若移动机器人所处的复杂环境是一个有限状态的Markov过程，当机器人在动作集合中选择某动作后，环境接收该动作并发出状态转移，同时给出奖赏*r*。如果其动作导致环境正奖赏（*+r*），则移动机器人此后产该动作的趋势就会增强；反之，如果导致环境负奖赏（*-r*），则会减弱。设在*t*时刻，机器人选取动作，环境从状态转移到时，所给出的奖赏为，显然

与的概率分布由和所决定，且为*t*时刻幵始时状态-动作对的佔计值，在Q学习过程中，以通过优化迭代计算的*Q(s,a)*函数值来逼近最优值函数，基本更新规则如下：

所以Q-learning算法的学习过程可以总结如下，智能体在t时刻的学习过程可以描述为：

1. 观察当前状态；
2. 选取并执行动作；
3. 观察下一个状态；
4. 生成一个强化信号；
5. 根据强化信号更新Q值函数

其中：

为折扣因子；

A为动作集合，元素个数n；

S状态集合，也是环境单元集合，元素个数为m；

Q，q值表集合，元素个数是m\*n，表示t时刻，由和确定的q值。表示t+1时刻，由 和确定的q值，也就是q的更新值。表示执行动作后，下一个状态的所对应的Q值，表示这些Q值表里面的最大值

为学习率，他控制着学习的速度，越大，则收敛速度越快，但是过大的可能会引起不成熟的收敛，所以选择合适的学习率，对于学习速度起着重要的作用。下面将会证明:当时，经过无数次迭代，Q-learning算法一定收敛于最优解。所以，Q-learning算法不是去估计环境模型，而是直接优化一个可以迭代计算的Q值函数。

3.1.2 Q-learning算法的流程

根据Q-learning算法的原理，总结其算法流程如下表

Initialize *Q(s,a)* arbitrarily

Repeat (for each episode)

Initialize *s*

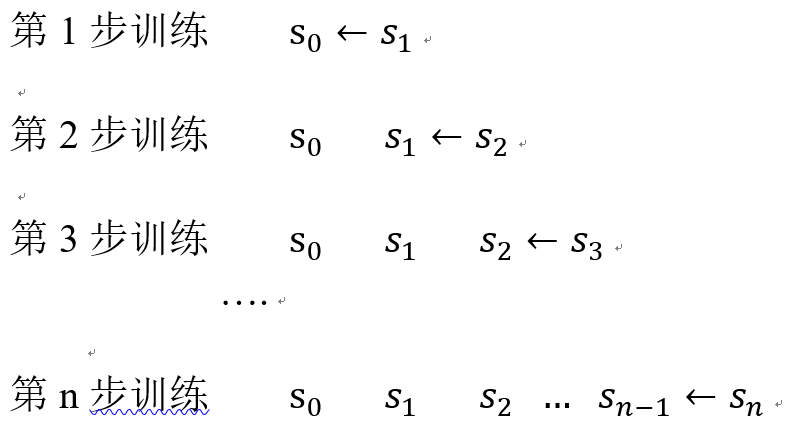
Repeat (for each step of episode)

Choose *a* from *s* using policy derived from *Q*

Take action *a,* observer *r , s*

Until *s* is terminal

在上表中，机器人的每一次尝试都是完成由起始状态到达目标状态的整个任务。

 机器人的每一次尝试中的迭代过程可以用下图说明，图中表示机器人的起始状态，表示机器人经历的第一个状态,表示机器人经历的第二个状态，表示机器人经历的第n个状态，箭头表示状态数据传递的方向，所传递的数据包括奖励值r和某状态占用的所有Q值。

由于该算法每一步训练只传递一次数据，所以常常被称为一步迭代Q-learning算法。该算法的特点是：所用的内存空间大，并且学习实践长，收敛速度慢。

3.1.2 Q-learning算法的流程

Q学习算法通常在一张搜索表中存储与每个状态-动作对相关的Q值，由上一节公式我们知道状态对应的Q值

3.1.2 联合界方法

我们用联合界[10]的方法来对SIM-OFDM系统的误码率进行分析。我们假设频域的信道系数可以表示为：,其中～N(0，1)。为离散傅里叶变换（DFT）矩阵，N（0,1）表示服从标准正态分布，即均值为零，方差为1的正态分布。的相关矩阵为：

 （3-4）

其中。

……

3.1.3 UPPER BOUND 方法

由于联合界方法在高阶下没有很紧的界，因此我们提出了UPPER BOUND[12]的推导方法。……

……

3.1.4 基于最大似然检测的误码率分析

上述方法或多或少的存在一些局限性和缺点，因此我们希望将上述方法进行一些改进，以便获得更准确的误码率表达式。……

……

3.2 仿真结果及分析

在本章和第四章我们使用的SIM-OFDM系统模型都是一样的，均为2选1的SIM-OFDM系统，选择的子载波数目为1024，循环前缀长度为256。我们使用的是EVA车载信道模型，车速为56km/h，信道的最大时延为2.51ms。

3.2.1 能量判决方法仿真

我们使用能量再分配策略，没有加交织，在瑞利衰落信道下将对SIM-OFDM系统和OFDM系统在不同的调制方式下的性能进行比较。这里我们使用能量判决方法，因此需要进行均衡。如图3-1，图3-2和图3-3所示，SIM-OFDM系统在误码率性能方面相较于传统的OFDM系统有一定的优势。……



图3-1 能量判决下BPSK的OFDM与SIM的误码率性能比较



图3-2 能量判决下QPSK的OFDM与SIM的误码率性能比较



图3-3 能量判决下16-QAM的OFDM与SIM的误码率性能比较

3.2.2 联合界方法仿真

联合界方法可以根据信道的相关性直接改变信道相关矩阵以获得相应信道条件下的误码率表达式，因此较为简单灵活。……

……

3.2.3 UPPER BOUND 方法仿真

我们使用能量再分配策略，没有加交织，在瑞利衰落信道下将对SIM-OFDM系统在不同的调制方式下进行仿真，使用最大似然判决，因此不需要进行均衡。……

……

3.2.4 基于最大似然算法的误码率分析仿真

我们使用能量再分配策略，没有加交织，在加性高斯白噪声信道下和瑞利衰落信道下将对SIM-OFDM系统在BPSK调制方式下进行仿真，使用最大似然判决，因此不需要进行均衡。……

……

3.3 本章小结

本章通过介绍通信系统性能指标引出对误码率的理论分析，之后又通过对能量判决的方法证明了其和最大似然（ML）方法的等效性。……

第四章 SIM-OFDM系统的载波频率偏移的概念及分析

4.1 载波频率偏移下的OFDM系统

我们都知道，OFDM系统将一串数据比特流通过串并转换之后将其分别调制到N个正交的子载波上，……因此我们希望通过对ICI进行研究、分析和处理，并在此基础上研究频率偏移对通信系统的误码率性能的影响。

4.1.1 OFDM系统的频偏模型

发射机和接收机之间的载波频率偏移（CFO）将会导致接收信号在频域上发生偏移，从而使得各个子载波之间正交性被破坏，系统的误码率性能将会急剧变差。

……

4.1.2 OFDM系统有频偏的误码率分析

在对有载波频率偏移（CFO）的OFDM系统的模型有了基本的数学分析之后，我们来对OFDM系统进行误码率的性能分析。……

……

4.2 载波频率偏移下的SIM-OFDM系统

SIM-OFDM系统作为OFDM系统的一种改进形式，虽然传输形式和OFDM系统有一些不同，然而载波频率偏移（CFO）的出现依然会导致原本互相正交的子载波不再正交，……

4.2.1 SIM-OFDM系统的频偏模型

这里我们仅仅考虑BPSK调制下的AWGN信道的误码率分析，使用的系统依然是2选1的SIM-OFDM系统。……

……

4.2.2 频偏对SIM-OFDM系统的误码率分析

……

4.3 仿真结果及分析

4.3.1 频偏下的OFDM系统

……

4.3.2 频偏下的SIM-OFDM系统

……

4.3.3频偏下的SIM-OFDM与OFDM系统对比

……

4.4 本章小结

……

第五章 结束语

5.1本文内容

本文主要讨论了OFDM系统的一种改进形式SIM-OFDM系统。我们先介绍了OFDM系统，并在此基础上引出了SIM-OFDM系统，介绍了SIM-OFDM系统的基本原理和模型。……

5.2下一步学习工作方向

我们这里仅仅是在讨论有频率偏移的情况下仅仅考虑使用BPSK调制方式，二选一的SIM-OFDM系统的误码率进行了分析，接下来我们可以考虑调制方式为更高阶，并且子载波为N选k的情况下继续研究SIM-OFDM系统的误码率性能。

参考文献

[1] 啜钢，王文博，常永宇等.移动通信原理与应用[M].北京：北京邮电大学出版社，2002.1-13

[2] 佟学俭，罗涛等.OFDM移动通信技术原理与应用[M].北京：人民邮电出版社,2003年6月.1-21

[3] B. Stantchev, G. Fettweis. Time-variant distortions in OFDM[J]. *IEEE Communications Letters*, Vol. 4, No. 10, October 2000, pp. 312-314

[4] P.H. Moose. A technique for orthogonal frequency division multiplexing frequency offset correction[J].*IEEE Trans.Communication*,Vol.42,No.10,October 1994, pp. 2908-2914

[5] J.-J. Van de Beek, M. Sandell, P.O. Brjesson. ML Estimation of Timing and Frequency Offset in OFDM Systems[J]. *IEEE Trans. Signal Processing*, Vol. 45,No.7, July 1997, pp. 1800-1805

[6] R. Abualhiga and H. Haas.Subcarrier-index modulation OFDM[C]. in *Proc. IEEE Int. Symp. Pers., Indoor Mobile Radio Commun*, Tokyo, Japan, 2009, pp. 13-16.

[7] Mesleh R Y, Haas H, Sinanovic S. Spatial modulation[J]. *IEEE Transactions on Vehicular Technology*, 2008,Vol.57,No.4, pp. 2228-2241.

[8] D. Tsonev, S. Sinanovic, and H. Haas.Enhanced subcarrier index modulation (SIM) OFDM[C].in *Proc. IEEE GLOBECOM Workshops*, Houston,Tx2011, pp. 728-732.

[9] 李晓峰，周宁.周亮.通信原理[M]。北京：清华大学出版社，1973.22-25

[10] E. Basar, U. Aygolu, E. Panayirci, and H. V. Poor. Orthogonal frequency division multiplexing with index modulation[C]. in *Proc. IEEE GLOBECOM*, 2012,pp. 4741-4746.

[11] E. Basar, U. Aygolu, E. Panayirci, and H. V. Poor. Orthogonal frequency division multiplexing with index modulation[J].*IEEE Trans. Signal Process*., Vol. 61, No. 22, Nov. 2012,pp. 5536–5549.

[12] Youngwook Ko.A Tight Upper Bound on Bit Error Rate of Joint OFDM and Multi-Carrier Index Keying[J]. *IEEE Communications Letters,* Vol. 18, N0.10,Oct. 2014,pp. 1763 - 1766 .

[13] A.Goldsmith.无线通信[M]（杨鸿文，李卫东，郭文彬等译）.北京：人民邮电出版社，2006.143~161

[14] Y.Zhao and S.G.Haggman.BER analysis of OFDM communication systems with intercarrier interference [C].in *IEEE Int. Conf. Communication Technology*,Beijing,China,Oct.1998,pp.1-5.

[15] L.Rugini and P.Banelli.BER of OFDM Systems Impaired by Carrier Frequency Offset in Multipath Fading Channels[J].*IEEE Transactions on Wireless Communications,*Vol.4,No.5,Sep. 2005 ,pp. 2279-2288.

[16] X.Ma,H.Kobayashi and S.C.Schwartz.EFFECT OF FREQUENCY OFFSET ON BER OF OFDM AND SINGLE CARRIER SYSTEMSR[C].*IEEE International Symposium on Personal,Indoor and Mobile Radio Communication Proceedings*,Sep.2003,pp.2239-2243.

致谢

本论文的工作是在我的导师XX老师悉心指导下完成的，……

外文资料原文

A Tight Upper Bound on Bit Error Rate of Joint OFDM and Multi-Carrier Index Keying

II.Joint MCIK-OFDM System Model

We consider a peer-to-peer M-QAM OFDM transmission with Nc sub-carriers

that consists of n clusters of N sub—carriers(i.eNc=nN).A stream of M-QAM

symbols is first serial-to-parallel converted,where every n()symbols are grouped into a vector  and  are used to modulate sub-carriers,as in the

classical OFDM,but it differs from that the modulated sub-carriers are only those of n activated indices,similar to [4],[5].

……

外文资料译文

基于多载波索引键控的正交多路复用系统的误码率上界

二．基于多载波索引键控的正交频分多路复用系统模型

我们考虑一个端到端的M-QAM，Nc子载波的基于多载波索引键控的正交频分多路复用系统有n个簇，每个簇有N个子载波（Nc=nN）。M-QAM的符号流经过串并转换之后每n个符号组成一个相量，是和传统正交频分多路复用一样是用来调制子载波的，但是不同的是只有这n个活跃子载波进行了调制。……

……