用 AO* 算法求解一个智力难题 *

朱福喜　余　亮　黄干平
武汉大学数学与计算机科学学院　(武汉 430072)

摘　要　文章提出了一种新的求解智力难题——假币与天平问题的方法,该方法利用了 AI 的搜索技术,将智力问题用一个适当的问题空间表达出来,然后将要求解的问题转化为与/或图中的搜索。研究表明 AO* 非常适合求解这个智力难题。

关键词　人工智能　状态空间　AO* 算法

Using AO* Algorithm to Solve a Difficult Intelligent Puzzle

Zhu Fuxi　Yu Liang　Huang Ganping
(College of Mathematics and Computer Science, Wuhan 430072)

Abstract: This paper presents a new method to solve an intelligent puzzle——the false coin and balance scale puzzle. The method exploits the searching technology of artificial intelligence, using an adequate state space to express the puzzle then translate the puzzle solution to the searching in and/or chart. Our research illustrate that the OA* algorithm is good for this puzzle.

Keywords: Artificial Intelligence, State Space, OA* algorithm

人工智能中的技术和方法来自于人们对智力的探索,反过来用它去求解一些智力难题,会有一种得心应手的效果。笔者在人工智能课程的教学中碰到这样一个智力难题,如果仅当作智力游戏去做,求解非常困难,找到全部解更是困难。该问题是:设有 12 枚硬币,凡轻于或重于真币者,即为假币(只有一枚假币),要设计一搜索算法来识别假币并指出它是轻于还是重于真币,且利用天平的次数不多于 3 次。

该问题的困难之处在于问题要求只称三次就要找到假币,否则就承认失败。然而有些称法不得当,使得留下的未知币太多,以至不可能在三次内称出假币。因此,每称一次,我们希望尽可能地得到关于假币的信息。

利用人工智能的求解方法解决这个问题首先必须解决下面两个问题:

(1)问题表示方法,记录和描述问题的状态。

(2)求解程序如何对某种称法进行评价。

下面就对使用 AO* 算法求解这个智力难题进行讨论。

1　问题的表示

文章先要分析构成该问题状态的因素有哪些。首先,每称一次后,新状态必须保留所剩的次数;其次,每称一次后,有关硬币的重量状态的变化也必须加以记载。为了对每称一次后得到的信息进行估计,将硬币的重量分为 4 种类型:

标准型(Standard) 标记为 S

轻标型(Light or Standard) 标记为 LS

重标型(Heavy or Standard) 标记为 HS

轻重标准型(Light or Heavy or Standard) 标记为 LHS

例如,一次称两个硬币,如果天平偏向左边,则天平左盘中的硬币属于中标型,而右盘中的硬币属于轻标型,其余属于标准型。每称一次,硬币的重量状态可能会从一种类型转变为另一种类型。问题处于初始状态时,由于不知每个硬币的重量类型,所以所有的硬币均属于 LHS 型。

综上所述,问题的状态空间可表示成一个五元组:

(LHS,LS,HS,S,T)

其中前四个元素表示当前这四种类型硬币的个数,T 表示所剩称硬币的次数。

在这样的状态空间表示下,有:

初始状态:(12,0,0,0,3)

目标状态:sg_1:(0,1,0,11,0) 和 sg_2:(0,0,1,11,0)

sg_1、sg_2 分别表示最后找到一个轻的或找到一个重的硬币,其余 11 个为标准硬币。

2　AO* 算法

AO* 算法是与/或图最佳优先搜索算法,用该算法求解问题需要三大要素:

(1)初始化问题描述;

(2)一组将问题变换成子问题的变换规则;

(3)一组本原问题描述。

该问题的初始问题前面已经表示出来了,本原问题就是两个目标状态。下面要定义一组转换规则。这里的转换规则就是每称一次,要考虑如何取硬币放到天平上,然后称完后,根据天平的状态,硬币的重量状态可能会从一种类型转变为另一种类型。

首先考虑如何取硬币的问题。设当前的状态为 (lhs,ls,hs,s,t),用函数 PICKUP([lhs_1,ls_1,hs_1,s_1],[lhs_2,ls_2,hs_2,s_2]) 表示本次分别从(lhs,ls,hs,s) 中取出了 lhs_1,ls_1,hs_1,s_1 个硬币放到天平的左边,取出了 lhs_2,ls_2,hs_2,s_2 个硬币放到天平的右边。

* 湖北省自然科学基金项目资助,项目号 99J026。

作者简介:朱福喜,副教授,主要研究领域:人工智能,并行计算。余亮,硕士研究生,研究方向:人工智能。黄干平,教授,主要从事分布计算 CSCW 领域的研究。

对于 PICKUP 应默认有如下性质成立：

$0<lhs_1+ls_1+hs_1+s_1=lhs_2+ls_2+hs_2+s_2\leqslant 6$，即天平两边的硬币数相等且小于等于 6。

$lhs_1+lhs_2\leqslant lhs \wedge ls_1+ls_2\leqslant ls \wedge hs_1+hs_2\leqslant hs \wedge s_1+s_2\leqslant s$，即取出的硬币数小于等于相应类型原有的硬币数。

然后令 PICKUP()等于-1,0,1 分别表示天平左倾、平衡和右倾。在这个定义下，有如下转换规则：

L 规则：if PICKUP $([lhs_1,ls_1,hs_1,s_1],[lhs_2,ls_2,hs_2,s_2])=-1 \wedge (lhs,ls,hs,s,t)$

then (lhs',ls',hs',s',t-1)；

其中 $s'=s+ls-ls_2+hs-hs_1+lhs-(lhs_1+lhs_2)$；$ls'=ls_2+lhs_2$；$hs'=lhs_1+hs_1$；$lhs1=0$

这四个公式的含义分别是：若天平左倾，则在左天平的状态为 LS 的硬币，在右天平的状态为 HS 的硬币和未放到天平上的硬币都是标准的，即 $hs_2+ls_1+(lhs-lhs_1-lhs_2)+(ls-ls_1-ls_2)+(hs-hs_1-hs_2)$个硬币的状态都改变为标准型；右天平原有的 ls_2 个轻标准型的硬币仍然为轻标准型，lhs_2 个轻重标准型硬币改变为轻标准型；左天平原有的 hs_1 个重标准型的硬币仍然为重标准型，左天平的 lhs_2 个轻重标准型硬币改变为重标准型；只要不平衡，就不存在 lsh 型的硬币，所有的硬币都可以确定为 LS,HS 或 S 型，即 lhs'=0。

B 规则：if PICKUP $([lhs_1,ls_1,hs_1,s_1],[lhs_2,ls_2,hs_2,s_2])=0 \wedge (lhs,ls,hs,s,t)$

then (lhs',ls',hs',s',t-1)；

其中 $s'=s+ls_1+ls_2+hs_1+hs_2+lhs_1+lhs_2$；$ls'=ls-ls_1-ls_2$；$hs'=hs-hs_1-hs_2$；$lhs'=lhs-lhs_1-lhs_2$

这四个公式的含义分别是：若天平平衡，则所有在天平上的硬币都是标准的，即有 $ls_1+ls_2+hs_1+hs_2+lhs_1+lhs_2$ 个硬币的状态都改变为标准型；左、右天平原有的轻标准型的硬币改变为标准型，所以从 ls 中减去 ls_1 和 ls_2；左、右天平原有的重标准型的硬币改变为标准型，所以也要从 hs 中减去 hs1 和 hs2；在平衡情况下，lsh 型的硬币要减去左、右天平原有的轻重标准型的硬币，即 $lhs'=lhs-lhs_1-lhs_2$。

R 规则：if PICKUP $([lhs_1,ls_1,hs_1,s_1],[lhs_2,ls_2,hs_2,s_2])=1 \wedge (lhs,ls,hs,s,t)$

then (lhs',ls',hs',s',t-1)；

其中 $s'=s+ls-ls_1+hs-hs_2+lhs-(lhs_1+lhs_2)$；$ls'=ls_1+lhs_1$；$hs'=lhs_2+hs_2$；$lhs'=0$

R 规则的含义为：若天平右倾，则左天平上状态为 HS 的硬币和在右天平上状态为 LS 的硬币以及未放到天平上的硬币都是标准的，即 $hs_1+ls_2+(lhs-lhs_1-lhs_2)+(ls-ls_1-ls_2)+(hs-hs_1-hs_2)$ 个硬币的状态都改变为标准型；左天平原有的 ls_1 个轻标准型的硬币不变，lhs_1 个轻重标准型改变为轻标准型；右天平原有的 hs_2 个重标准型的硬币也不变，lhs_2 个轻重标准型的硬币改变为重标准型；因为不平衡，lhs'=0。

以上规则中 t-1 表示所剩称硬币的次数减少了一次。

3 问题搜索

人工智能问题的求解往往转化为问题搜索。在该问题搜索中，PICKUP 表示一种选取方法，不同的选取方法之间是或的关系，当选定一种 PICKUP 后，就必须考虑它的值为 -1,0,1 的情况下下层的节点都可解，则三种情况之间的关系为“与”的关系。这说明该问题的搜索图是与或图，所以采取的搜索算法是 AO* 算法。用于该算法的评价函数为：h((lhs,ls,hs,s,t))= ls + hs + lhs-1

显然有：h((0,1,0,11,0))=0 和 h((0,0,1,11,0))=0，即 $h(sg_1)=h(sg_2)=0$。

由于问题的本原问题对应的节点是可解节点，因此 sg_1、sg_2 为可解节点。不可解节点可定义为：如果节点 n =(lhs,ls,hs,s,t)中 t=0 且 (lhs,ls,hs,s)不属于{(0,1,0,11),(0,0,1,11)}，则 n 为不可解节点。做好上述准备工作后，就可用 AO* 算法进行求解，图 1 就是用 AO* 算法得到得一个解图。

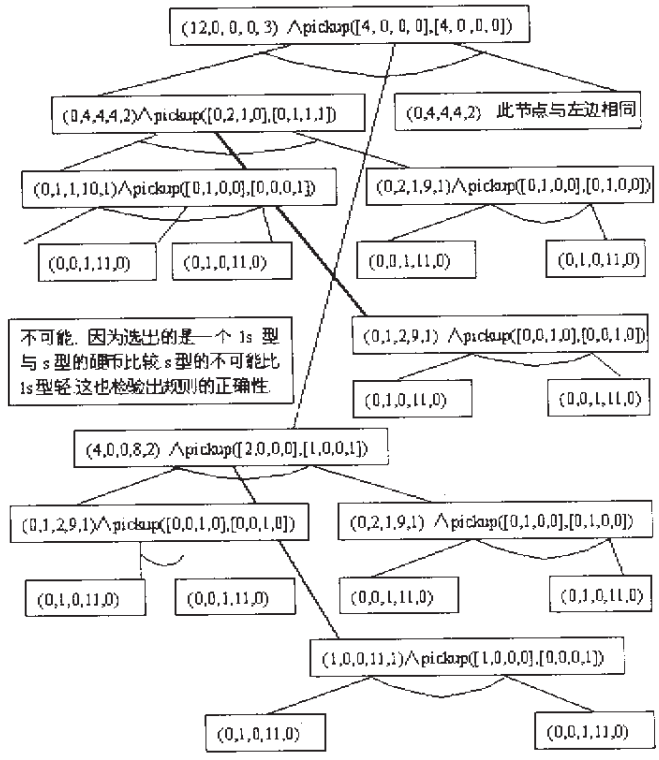

图 1

用 AO* 算法可以很快地接近目标，因为对于某种选法，它的下层节点层数不超过三层，且只要有一个下层节点为不可解节点，则马上推出它为不可解节点。例如：考虑 PICKUP([6,0,0,0],[6,0,0,0])，它的值只可能是 1 或-1，它的下层节点实际只有一个为(0,6,6,0,2)的节点。对于它的下两层的节点的搜索可很快导出都是不可解节点。因而导出 PICKUP([6,0,0,0],[6,0,0,0])不是一种正确的取法。同样的道理也可以很快导出 PICKUP ([1,0,0,0],[1,0,0,0]),PICKUP ([2,0,0,0],[2,0,0,0])，…等取法都不是正确的取法。因为抛去了大量的不可解的分支，所以，用 AO* 算法可以很快地找出所有解。

4 比较与结论

对于上述智力难题，L.Wos 等人采用的方法是用子句描述了 14 个转换公理，然后在推理系统上使用超归结推理规则求解。所用的机器是 IBM3033，运行时间 22 秒，求得 40 多种不同的解。该文所采用的 AO* 搜索方法，只用了三个转换规则，算法在 586 微机上实现，运行时间 1 秒，求得 410 种不同的解，这 410 种解已过滤了对称的情况。比较之下，用 AO* 算法求解这

(下转 87 页)

(上接 70 页)

个智能问题的解法比较简洁,且效率比较高。AO* 算法之所以可以很快地找出所有解,首先是因为该问题用与或图表示比较恰当。然后是 AO* 算法可以利用与或图的特点和启发式方法避免许多无用路径的搜索。(收稿日期:2000 年 2 月)

参考文献

1.E.丽奇著,李卫华,汤怡群,文中坚编译.人工智能引论.广东科技出版社,1986,6

2.周祥和,戴大为,麦卓文编译.自动推理引论及其应用.武汉大学出版社,1987,12

3.L.Wos,L.Henschen.Automated theorem proving.In Jorg Sickmann, Graham Wrightson,The Automation of Reasoning,Springer-Verlag, New York,1983;11.11:1965-1970

用AO算法求解一个智力难题

作者： 朱福喜， 余亮， 黄干平
作者单位： 武汉大学数学与计算机科学学院
刊名： 计算机工程与应用 ISTIC PKU
英文刊名： COMPUTER ENGINEERING AND APPLICATIONS
年，卷(期)： 2001，37(3)
被引用次数： 1次

参考文献(3条)

1. L Wos;L Henschen Automated theorem proving.In Jorg Sickmann,Graham Wrightson,The Automation of Reasoning 1983

2. 周祥和;戴大为;麦卓文 自动推理引论及其应用 1987

3. E丽奇;李卫华;汤怡群;文中坚 人工智能引论 1986

相似文献(10条)

1. 学位论文 姜贵栋 带有回溯和有用动作排序的FF规划算法研究 2006

智能规划是人工智能的一个重要领域。近年来，有关智能规划的研究在问题描述和问题求解两方面得到了新的突破，使得智能规划已成为一个热门的人工智能研究领域。对智能规划算法的研究主要集中在规划算法性能的提高和规划算法求解问题的范围的扩展。基于规划图原理和启发式状态空间搜索的智能规划算法在众多的智能规划算法中表现十分突出。FF规划器将规划图和启发式状态空间搜索原理进行了很好的结合，在两届IPC(InternationalPlanningCompetition)中取得了优异的成绩。但是FF规划算法所采用的搜索方法和规划图的使用上都存在很大不足，使得FF规划算法在某些规划问题域中，特别是含有死点状态的问题域中表现不够理想。

本文通过对规划图原理、启发式状态空间搜索原理和FF规划算法的研究，提出了一种更适合FF规划器使用的搜索方法——带有回溯的加强爬山搜索算法，同时在规划图的使用中加入了有用动作排序并且探索了多种排序标准。带有回溯的加强爬山算法在加强爬山算法失败时，通过回溯使规划器可以利用前一步的搜索结果来继续进行加强爬山搜索。有用动作排序使规划图和状态空间搜索更好地结合在一起。以上两种方法可以有效地解决死点问题。基于以上两点，本文提出了一种新的规划算法，即带有回溯和有用动作排序的FF规划算法。

本文将回溯应用在规划解的提取过程中，同时将规划图理论和状态空间理论有机地结合在一起，很好地解决了死点状态问题。基于新算法的BTFF规划器相对于FF规划器极大地提高了求解的效率和解决问题的范围。因此本文具有很重要的学术意义和研究价值。

2. 期刊论文 潘必超.张瑾 有限状态空间搜索技术在紧急逃生行为仿真中的应用 -计算机时代2006(6)

平面建筑紧急逃生性能是一组非常重要但又难以测量的参数,通过真人演练进行测试会导致人力物力的巨大耗费.文章通过结合人工智能中有限状态空间搜索技术,以紧急状态下人类行为趋势的统计结果为基础,使用Visual C#实现了紧急逃生行为智能体的开发,并以此实现对平面建筑逃生性能的动态通用测试,具有较高的可行性和实用价值.

3. 学位论文 李响 动态不确定性环境下的实时规划系统研究 2004

作为一种非常重要而且常见的智能行为和能力，规划(Planning)就成为人工智能研究的一个重要领域，很早就受到关注的主要问题之一。而在动态不确定性环境下的规划就因其更加贴近现实环境，具有更高的实用价值而成为目前规划问题研究的重点和热点。

本文首先分析动态不确定性环境的主要特点，包括：

■动态性：环境的状态无时无刻不在变化。它不仅仅受智能体自身的影响而变化，还受环境中其他智能体和其他因素的影响而变化。

■智能体知识的局限性：一般来说，智能体不可能掌握环境中所有的知识，不可能了解可以引起环境变化的所有因素，不可能了解其他智能体的所有情况。智能体只可能部分的掌握这些知识，甚至对一些方面一无所知。

■智能体行动的不确定性：智能体在环境中执行一定的行为，其结果是不确定的，事先无法对这个结果作准确的预测。

■智能体观察的局部性：一般来说，智能体对环境的观察是不全面的。在同一时刻，智能体只能观察到环境中一部分的情况。

■智能体观察的不确定性：智能体从环境中得到的观察一般来说是不准确的，有时甚至是错误的。

然后，对现有的规划系统在适应上述动态不确定性环境的能力进行了概述。分析了这些系统在适应动态不确定性环境方面各自的优点和不足。

本文的主要工作是基于以上的分析和认识，提出了基于PRS和决策论规划的面向动态不确定性环境的规划系统POMDPRS。并讨论了两种提高决策效率的改进方法。具体工作主要有：

1)提出了面向动态不确定性环境的规划系统POMDPRS。描述了其基本模型，并给出了形式化描述。POMDPRS通过保持PRS系统的持续规划机制来适应环境的动态性，通过使用环境状态空间上的概率分布作为智能体的信念来适应环境的不确定性，从而兼顾了两个大方面的要求。

2)阐述了状态因子化表示在POMDPRS中的应用，并给出了因子化的POMDPRS——FPOMDPRS的形式化描述。POMDPRS使用环境状态空间上的概率分布作为智能体的信念，并根据智能体输出的行为和接收到的观察来对其进行更新。但是在很多情况下，状态空间往往十分巨大，从而使得信念更新的时间消耗非常高，难以适应系统反应实时性的需要。因子化方法通过将状态表示中涉及到的环境属性根据其互相依赖关系来对它们进行划分。将一个状态表示为几个子状态的集合，从而将未因子化时的一个大状态空间变成几个较小的状态空间。从而信念也就变成几个子状态空间上的概率分布的集合。在信念更新的时候，对这几个子状态空间上的概率分布分别处理，从而达到削减信念分布时间消耗的作用。

3)阐述了Monte Carlo滤波表示在POMDPRS中的应用，并给出了应用MonteCarlo滤波的POMDPRS——MCPOMDPRS的形式化描述。削减信念更新的时间消耗的另一个方法是Monte Carlo滤波。它通过使用概率分布上有限的一些具体数值(样本)来代表整个分布，并根据行动和观察，使用SIR方法来对这个样本集进行更新。这使得信念更新的时间消耗依赖于样本集的大小。从而可以通过控制样本集的大小来控制信念更新的时间消耗。

因子化和Monte Carlo滤波可以在POMDPRS中结合起来使用。即先对状态进行因子化，然后再对一些仍然很大的子状态集使用Monte Carlo方法，从而达到进一步提高信念分布更新效率的目的。本文在最后具体描述了一个FPOMDPRS和MCPOMDPRS相结合的，在实体机器人上运行的机器人决策控制系统P-DOG并给出了实验结果，验证了POMDPRS及其变种的可行性。

4. 期刊论文 黄文生 A*算法的证明及其在人工智能领域的应用 -江苏电器2002(5)

A*算法在计算机和人工智能领域当中都是非常典型的算法，有着非常广泛的应用，本文对A*算法及其可采纳性进行了详细的论述和证明，并通过一个求最短路径的实例说明了A*算法在实际中的具体应用。

5. 学位论文 殷苌茗 激励学习的若干新算法及其理论研究 2006

本博士论文大体上可以分成两大部分，第一部分我们给出了激励学习的一些新算法，其目的是为了改进现有算法所面临的诸于维数灾难与计算速度等问题。第二部分是我们在基于风险敏感度概念的基础上，研究了与激励学习有关的最优方程与最优解的理论问题。

本论文首先提出了一种新的激励学习算法，即我们所说的激励学习的遗忘算法。这种算法的基本思想是基于下面的考虑：以前的有关激励学习算法都只是通过

对状态被访问次数的短时记忆来确定状态值函数空间的更新长度。对于这些算法来说，无论是用lookup表的形式，还是用函数逼近器的形式，对所有状态的值函数都必须要全部记忆。这些方法对大状态空间问题会导致需要指数增长的记忆容量，从而呈指数地减慢计算速度。Sutton等人考虑过这个问题，但始终没有得到满意的解决方案。基于上述考虑，将记忆心理学中有关遗忘的基本原理引入值函数的激励学习算法的研究之中，特别是对于著名的SARSA(λ)算法，形成了一类适合于值函数激励学习的遗忘算法。

我们提出了基于效用聚类的激励学习算法。这种算法的模型使用了POMDP的某些概念。在激励学习的诸多算法中，把非常多的精力集中在如何对系统的大状态空间进行有效的分解，其中U-Tree算法是其之一。但是，由于U-Tree算法的一个最大的问题是边缘节点生成的随意性和统计检验的计算负担，各种扩展方法都没有解决这个问题。本文提出了一种新的效用聚类激励学习算法，即我们称之为U-Clustering算法。该算法完全不用进行边缘节点的生成和测试，克服了上述提及的U-Tree算法的致命弱点。我们的新算法首先根据实例链的观测动作值对实例进行聚类，然后对每个聚类进行特征选择，最后再进行特征压缩，经过压缩后的新特征就成为新的叶节点。实验与仿真结果表明，我们的新算法比一般的U-Tree算法更有效。

针对具有大状态空间的环境系统以及系统状态不完全可观测所面临的问题，本论文提出了求解部分可观测Markov决策过程的动态合并算法。这个方法利用区域这个概念，在环境状态空间上建立一个区域系统，而Agent在区域系统的每个区域上独自实现其最优目标，就好像有若干个Agent在并行工作一样。然后把各组成部分的最优值函数按一定的方式整合，最后得出POMDP的最优解。另外还对提出的算法进行了复杂度分析和仿真实验。通过对New York Driving的仿真和算法的实验分析，结果表明动态合并算法对于大状态空间上的求解问题是一个非常有效的算法。

本文提出了风险敏感度的激励学习广义平均算法。这个算法通过潜在地牺牲解的最优性来获取鲁棒性(Robustness)。提出这种算法的主要原因是因为，如果在理论模型与实际的物理系统之间存在不匹配，或者实际系统是非静态的，或者控制动作的“可使用性”随时间的变化而变化时，那么鲁棒性就可能成为一个十分重要的问题。我们利用广义平均算子来替代最大算子max(或sup)，对激励学习问题中的动态规划算法进行了研究，并讨论了它们的收敛性，目的就是为了提高激励学习算法的鲁棒性。我们提出了风险敏感度渐进策略递归激励学习算法并对策略的最优性进行了讨论。当系统的计算出现维数灾难时，比如在Markov决策过程的求解问题中，如果系统的动作空间非常之大，那么利用一般的策略递归(PI)算法或值递归(VI)算法，来进行策略的改进计算是不实际的。我们这个算法所关注的问题是，当状态空间相对较小而动作空间非常之大时如何得到最优策略或好的策略。在本算法的策略改进过程中，不需在整个动作空间上对值函数进行最大化运算，而是通过策略转换的方法来直接处理策略问题的。

本文的另一个主要内容是，我们对多时间尺度风险敏感度Markov决策过程的最优方程与解的最优性问题进行了初步研究。由于需要使智能体能够适应更加复杂的环境，因此大规模控制问题在实际应用中显得越来越重要。在本章中采用了一种更加符合实际情况的复杂环境，即多时间尺度下的Markov决策过程模型，并利用风险敏感度的概念，第一次提出了多时间尺度风险敏感度Markov决策过程的概念。这是一个全新的问题。我们利用效用函数和风险敏感度等概念，讨论了二时间尺度风险敏感度Markov决策问题，然后给出了二时间尺度风险敏感度Markov决策过程的Bellman最优控制方程的一般形式，并证明了最优值函数满足这个Bellman最优控制方程，同时还得到了一些相关的结论。

本博士论文所讨论的都是现阶段关于激励学习的热点和难点问题。激励学习方法已经成为控制理论与人工智能研究的最重要的分支之一，还有许多问题亟待解决。在本文的最末，我们给出了对这些问题的研究方向和未来的工作，希望能起到抛砖引玉的作用。

6.学位论文 戴帅 基于因素化表示的强化学习方法研究 2009

强化学习是随机环境中解决决策问题一种有效的方法。然而，在大状态空间，特别是在复杂随机状态下的应用领域，它仍然没有解决“维数灾难”的问题。目前，因素化强化学习作为强化学习在时间和空间上的扩展，已经被证明比强化学习更适合解决大状态随机控制问题，在机器人导航等方面有着广阔的应用前景。但是，目前的研究工作集中在学习前状态空间的前期处理，对学习过程缺乏深入研究。本文围绕强化学习前的状态空间的前期处理以及学习过程中值函数的值的存储和表示，对以下方面进行了研究和探讨：

1.介绍了因素化学习的基本学习理论和研究进展，并对四种典型的强化学习算法作了分析比较，分析了它们的各自特点和适用情况，为后面的工作中算法的选择提供了基础。

2.提出了改进的基于因素化表示的动态规划方法，针对动态规划方法中求解精确的Vπ值计算量复杂的问题，提出了改进的使用生成Vπ的线性近似值以获取算法的加速的方法；针对传统强化学习算法使用值函数Look-up表存储和表示值函数的值存在着的冗余度过高的问题，提出了决策树方法，并在后面的仿真实验中验证算法效果。

3.提出了一种新的基于因素法方法的TD(λ)算法。其基本思想是状态因素化表示，通过动态贝叶斯网络(Dynamic Bayes Networks，DBNs)表示Markov决策过程(Markov decision Process，MDP)中的状态概率转移函数，结合决策树(decision tree)表示TD(λ)算法中的状态值函数的值，大大降低了状态空间的搜索与计算复杂度、以及数据的冗余度，因而适用于求解大状态空间的MDPs问题，对照实验证明了该表示方法是有效的。

7.期刊论文 许精明 状态空间的启发式搜索方法研究 -微机发展2002,12(4)

对人工智能中用于状态空间问题求解的启发式搜索方法--A算法和A*算法进行了详细分析,并指出了影响搜索算法启发能力的主要因素和提高搜索效率的措施.

8.学位论文 王骄 中国象棋计算机博弈关键技术研究 2006

人工智能的先驱者们曾认真地表明：如果能够掌握下棋的本质，也许就掌握了人类智能行为的核心；那些能够存在于下棋活动中的重大原则，或许就存在于其它任何需要人类智能的活动中。

计算机博弈是人工智能领域中一个重要的课题，国际象棋计算机博弈已经取得了巨大的成功，而中国象棋计算机博弈却远远落后。“棋天大圣”是东北大学人工智能与机器人研究所自主开发的中国象棋计算机博弈软件，取得了第11届世界电脑象棋奥林匹克竞赛中国象棋组的冠军。本文通过结合“棋天大圣”的研究成果，阐述了一个可以达到人类特级大师水平的中国象棋程序的设计和实现原理。

本文对中国象棋计算机博弈的关键技术进行了分析和介绍，并提出了一些新的思想和算法，以及基于数据测试集的分析和检验。其中包括以下几点。

第一，介绍了中国象棋状态空间的数据表示方法，介绍了对状态空间的数字表示和布尔表示，提出了用棋盘编码数组、棋子编码数组、映射数组数字表示状态空间，用新型数据结构路向行向比特向量与比特棋盘相结合布尔表示状态空间的新方法。在这些描述方法共同作用下，程序具有很高的运行效率，大大节省了计算时间。

第二，详细介绍了当今流行的各种着法生成算法，提出了路向行向比特向量与模板法相结合，并且辅以预置表作为辅助进行着法生成的新方法。着法生成的速度获得很大的提升。

第三，介绍了棋类搜索领域的搜索算法及其分类，给出了多种搜索算法的融合方式，以及在中国象棋计算机博弈领域应用的创新。

第四，结合“棋天大圣”的审局函数，介绍了审局函数的概念、组成和计算的方法。提出了小子同形表这种新的计算方法，可以对引擎起到良好的加速作用。

第五，本文就开局库的原理做出了介绍，提出了理想开局库并做了深入的探讨。对开局库自学习算法原理和成果进行了总结。

第六，提出了残局处理系统的新概念，将残局处理系统划分为残局知识库与残局数据库。对自行设计的残局知识库的原理、结构、实现方法做出了详细的介绍。

第七，创造性地将自适应遗传算法、神经元网络结合TD(λ)算法两种机器学习算法引入审局函数中，详细的介绍了与审局函数的结合、测试的方法以及取得的成果。本文的研究在中国象棋计算机博弈领域处于前沿，结合本文的研究成果可以创建高水平的博弈软件，而且在一系列电脑之间的比赛和人机挑战赛中，也得到了印证。本文的成果和结论，对于其它中国象棋计算机博弈程序，具有一定的参考价值。

9.期刊论文 段国林.查建中.徐安平.张满囤 启发式算法及其在工程中的应用 -机械设计2000,17(6)

启发式算法利用与所求问题有关的某些特殊信息来控制搜索状态空间的过程.对于某些难于用理论方法解决的问题,启发式算法可以起到独到的作用.分析了启发式算法在人工智能和运筹学领域内的研究现状,指出了各自研究的特点,提出应加强各学科在启发式算法问题研究上的合作.充分利用启发式算法已有研究成果,是解决工程实际问题中某些难题的有效方法.

10.学位论文 张科 基于强化学习的多智能体协作与应用的研究 2007

多Agent系统(Multi-Agent System，MAS)是分布式人工智能(Distributed Artificial Intelligence，DAI)的一个主要领域，而多个Agent之间如何进行组织协调和协作以实现共同目标是MAS研究的核心问题。解决MAS的协作问题有许多方法，Agent的学习方法是其中很重要的一种。通过Agent的学习实现MAS的协调与协作是一个非常值得研究、具有挑战性的课题。

本文将研究如何通过强化Q-学习方法来实现多Agent之间学习与协作，主要工作包括：

●单Agent行为搜索方案的优化多Agent系统的构成单元是一个个单独的Agent，很多个Agent的独立学习构成了多Agent系统的学习过程，那么要使用学习方法来实现多Agent之间的协作学习，首先要强化单个Agent的学习能力。原始的Q学习中采用的是非直接搜索的行为选择方法(如ε-greedy、Boltzmann 策略)，Wiering在其论文中又提出了直接搜索的方法。本文在以上工作的基础上探索了一种能够平衡Agent行为选择中探索与利用关系的方法，利用遗忘函数作为加权系数，使

Agent在刚开始搜索环境的时候能够按照人为制定的搜索方法对环境进行充分的探索，而在学习一段时间后则能使行为选择逐渐趋向于贪婪策略。实验证明，这种方法较一般行为选择策略，能更好的加速单Agent的学习进程。

●基于知识共享的多Agerlt学习方法传统的多Agent系统中，每个Agent在完成自己独立的学习过程之后，并不能够将自己学到的知识与其他Agent共享。本文研究了一种多Agent知识共享的方法(Q Table Sharing，Q表共享法)来提高整个MAS系统学习能力和性能。在状态空间较小时利用Q表来共享知识；在状态空间较大时，利用Q表进行知识的暂存与共享空间，利用小脑关节模型来完成最终Q值的存储。实验证明，当有多Agent同时学习时，这种方法不仅能强化每个组成Agent的学习能力，还可以提升多Agent系统的整体性能，学习效果也较单Agent的学习要稳定很多。

●基于叠加法的小脑关节模型(Cerebellar Model Articulation，CMAC) 强化学习算法传统的CMAC都是利用哈希方法来解决输入空间到记忆空间的映射冲突，减少输入维数增加带来的记忆空间急剧增大的问题，并且网络的学习误差只被分担到了泛化参数(C)个单元上。本文采用了一种新的基于叠加法的状态空间映射方法，可以使CMAC网络在输入向量维数很大的时候不仅可以避免映射冲突、减少网络的存储空间，而且学习误差也被分担到n*C个单元上。实验证明，这种方法能有效的与Q学习相结合解决大状态空间的Q值存储问题。

●方法的验证与应用本文通过Agent的路径寻优问题来检验以上方法的有效性，将其应用在RoboCup3D(Robot Soccer Cup Three Dimension)的3v2局部协作问题上，取得了很好的效果。此外，本文还讨论了方法在其他协作、学习问题上的应用前景。

引证文献(1条)

1.朱福喜.卓识 十二硬币问题的进一步求解[期刊论文]-计算机工程与应用 2001(21)