

基于博弈论的多机器人系统任务分配算法*

黎萍¹, 杨宜民²

(1. 电子科技大学中山学院, 广东 中山 528402; 2. 广东工业大学 自动化学院, 广州 510090)

摘要: 对多机器人系统任务分配策略进行了形式化描述, 为任务分配方案的求解提供了一种数学描述工具; 针对多机器人系统中机器人决策之间的相互依存性, 引入博弈论的思想分析了多机器人系统的任务分配问题, 提出了一种基于博弈论的多机器人系统任务分配算法(GT-MRTA)。实验结果表明, 算法复杂度较低, 计算量较小, 鲁棒性较好, 获得的任务分配方案质量较高。

关键词: 多机器人系统; 任务分配; 博弈论

中图分类号: TP24

文献标志码: A

文章编号: 1001-3695(2013)02-0392-04

doi:10.3969/j.issn.1001-3695.2013.02.020

Game theory based task allocation algorithm for multi-robot systems

LI Ping¹, YANG Yi-min²

(1. Zhongshan Institute, University of Electronic Science & Technology of China, Zhongshan Guangdong 528402, China; 2. Institute of Automation, Guangdong University of Technology, Guangzhou 510090, China)

Abstract: Considering to dependences between robots, this paper introduced main idea of game theory to depict multi-robot system task allocation which afforded a new mathematic description way for solving task allocation. To the relationships between strategies of robots, it analyzed multi-robot system task allocation based on game theory, and proposed a game theory based task allocation algorithm for multi-robot systems(GT-MRTA). Simulation experiment shows that GT-MRTA has low complexity, and needs less computation. It has good robustness for communication failure and can help to getting task allocation schemes with high quality.

Key words: multi-robot system (MRS); task allocation; game theory

0 引言

近年来,多机器人系统(MRS)以其突出的柔性、鲁棒性和自主性逐渐应用于工业生产的各个领域。MRS由多个功能、结构简单的机器人组成,通过机器人的合作完成单个机器人无法完成的复杂任务。多机器人的合作体现了多机器人系统存在的意义及其优越性,它是通过系统任务分配以及机器人的运动协调来实现的。任务分配作为MRS的重要研究内容之一,要将MRS中的任务分配给合适的机器人,以充分利用系统资源,提高系统的运作绩效^[1]。多机器人系统设计者所面临的一个主要问题是如何合理地在各个机器人之间分配任务,使它们协作完成全局任务。目前国内外学者主要采用组合优化法、基于阈值的方法、基于市场机制的方法等对多机器人系统的任务分配进行研究^[2]。

由于系统环境的动态性及机器人与任务组合的维数灾难,很难用传统的规划法来实现系统的任务分配。对于每个机器人而言,可以把其他机器人看成是系统环境的一部分,机器人所处的环境由于其他机器人的独立运行变得不确定。在这种不确定环境下,机器人根据已有的对环境状态的信念,从候选任务中选取合适的任务,机器人任务选择的效用不仅与机器人

本身的决策有关,也受其他机器人的决策影响。针对MRS环境的不确定性和机器人决策的相互依存性,本文引入博弈论的概念对分布式自主多机器人系统的任务分配进行了分析,并给出了基于博弈论的任务分配方法的基本工作过程。

1 多机器人系统描述

多机器人系统可以用以下要素来进行描述:

a) 机器人集合 $A = \{a_1, a_2, \dots, a_N\}$ 。在实际应用中,多机器人系统由多个同构或异构的机器人组成。

b) 系统所处的环境状态为 S' 。 $s'_t \in S'$ 表示在 t 时刻的环境状态。

c) 机器人 i 的可观测环境状态集合 S_i 。 $s_i = g(s'_t)$, $S = \prod_{i=1}^N S_i$ 为多机器人系统在 t 时刻对所处环境的联合观测,通过机器人感知和通信等途径获得观测环境状态。

d) 多机器人候选任务集合 B_i 。 B_i 表示机器人 i 的可能任务集合, $B = \bigcup_{i=1}^N B_i$ 表示所有机器人可能的任务集合,它是目标 M 的具体化。单个机器人能够根据对周围环境的观测选择一定的任务 $b_i \in B_i$ 来影响环境;从团队的角度看,各个机器人通过执行联合任务 $(b'_1, b'_2, \dots, b'_N) \in JB$ 来改变环境状态, $JB = \prod_{i=1}^N B_i$

收稿日期: 2012-05-27; 修回日期: 2012-08-15 基金项目: 电子科技大学中山学院博士启动基金资助项目(410YKQ01); 广东省中山市科技计划资助项目(20114A208); 广东省教育部产学研资助项目(2011B090400371)

作者简介: 黎萍(1981-),女,讲师,博士,主要研究方向为人工智能及智能机器人(lipingjxau@163.com); 杨宜民(1945-),男,教授,博导,主要研究方向为智能控制科学与技术。

表示多机器人系统可能执行的联合任务集。

e)信念。信念(bel)代表机器人所具有的所有知识,包括对周围环境、其他机器人的认识信息及自身的目标、策略、推理机制等信息。机器人*i*在*t*时刻的信念 bel_i^t 随着与周围环境的交互而不断变化。

f)策略。机器人的策略是从信念到任务的映射。机器人*i*的策略用 π_i 来表示, $\pi_i:bel_i \rightarrow B_i$ 。

g)学习。机器人执行任务后,根据从环境获得一个代表奖惩的再励信号来更新信念,获得到达最优策略的方向信息,最终获得最优的任务分配策略。

h)目标。在多机器人系统中,每个机器人都有各自的目标,整个系统可以有全局目标。各目标之间存在的关系一般有

(a)目标是一致的,一个机器人在实现自身目标的同时促进了其他机器人目标的实现,相互之间不存在冲突;

(b)目标是冲突的,各个机器人的利益或资源是冲突的,研究的主要内容是要减少或消除冲突。

2 博弈论介绍

博弈论主要研究智能体之间相互依存的理性行为,是研究智能体之间竞争和冲突关系的形式化表述方法,主要目的是确定智能体的理性决策以得到最大化的报酬或最小化惩罚。由于局中人的相互依存性,博弈中一个智能体的决策必定建立在预测其他局中人的反应之上,一个局中人将自己置身于其他局中人的位置并为人着想,从而预测其他局中人将选择的任务,在这个基础上该局中人选择自己最适合的任务。这就是博弈论方法的本质和精髓^[3]。

无论何种博弈,总存在以下三个要素:

a)局中人集合 $A = \{1, 2, \dots, N\}$ 。

b)每个局中人*i*有若干个策略可供选择,它构成的该局中人的纯策略空间 $\Lambda_i = \{\lambda_{i1}, \lambda_{i2}, \dots, \lambda_{ik_i}\}$ 。定义在 Λ_i 上的一个概率分布 $x_i = \{x_{i1}, x_{i2}, \dots, x_{ik_i}\}$ 被称为局中人*i*的一个混合策略,其中 x_{im} 表示选择纯策略 λ_{im} 的概率。

$$\sum_{m=1}^{k_i} x_{im} = 1 \quad x_{im} \geq 0, i=1, 2, \dots, N \quad (1)$$

一个局中人*i*的策略 π_i 是在博弈中其自身的一个任务计划,它是一个在博弈的每个阶段都分配一个任务或是分配选取一个任务的概率函数, $\pi_i \in \Pi_i$ 。 $\Pi_i = \{\pi_{i1}, \pi_{i2}, \dots, \pi_{ik_i}\}$ 为局中人*i*所有可能的策略集合。

c)每个局中人都有奖赏函数,记局中人*i*的奖赏函数为 $r_i(\pi)$,其中 $\pi = (\pi_1, \pi_2, \dots, \pi_N)$ 表示*N*个局中人的决策局势。

一个博弈可以用三维组 $\Gamma = \langle N, (\pi_i), (r_i) \rangle$ 表示。记 $\pi \parallel \rho_i = (\pi_1, \pi_2, \dots, \pi_{i-1}, \rho_i, \pi_{i+1}, \dots, \pi_N)$,表示在局势 $\pi = (\pi_1, \pi_2, \dots, \pi_N)$ 中,局中人*i*把它的策略从 π_i 换成 ρ_i ,其他局中人的策略不变,这样就得到新的局势 $\pi \parallel \rho_i$,显然 $\pi \parallel \pi_i = \pi$ 。

定义1 在一个*N*人博弈 Γ 中, π_i 表示局中人*i*的策略, $r_i(\pi)$ 表示定义在所有局中人联合策略 π 上的局中人*i*的奖赏函数。设 $\pi^* = (\pi_1^*, \pi_2^*, \dots, \pi_N^*)$ 为博弈 Γ 的一个局势,若对于每个局中人*i* ∈ *A*及每个 $\pi_{im} \in \Pi_i$ ($m = 1, 2, \dots, k_i$),有 $r(\pi^* \parallel \pi_{im}) \leq r(\pi^*)$,则 π^* 是博弈 Γ 的一个Nash平衡点^[4]。

博弈可以根据局中人达成协议的强制性划分成合作博弈

和非合作博弈两大类。合作博弈是局中人就某些行为达成协议,局中人不得轻易违反。在非合作博弈中,局中人相互独立决策,不能和其他局中人达成强制协议。

3 多机器人系统的任务分配

3.1 多机器人系统任务分配的分析

多机器人系统中各个机器人之间的交互是动态变化的,每个机器人都面临着动态决策,为了自身效用最大化而采取理性行为。单个机器人就相当于博弈论中的局中人,单个机器人的任务选取可以看做是一定局势下的博弈,每个机器人在一定的环境状态下,根据各自的情况以及其他机器人的决策形成一个博弈局势。多机器人系统任务分配就是在这个博弈状态下按照一定的目标,由每个机器人选择出效用最大的任务。在单机机器人系统中,机器人的决策就是马尔可夫决策过程(MDP),而在多机器人系统中就是一个多机器人随机博弈问题^[3,5]。

定义2 多机器人系统任务分配可以用一个博弈表示: $\Gamma = \langle N, S_i, B_i, Q_i, Q \rangle$ 。其中*N*表示系统中机器人的个数; S_i 表示机器人*i*的可能状态集合; B_i 表示机器人*i*的可能任务集合,是机器人*i*的纯策略空间; Q_i 表示机器人*i*在联合观测状态 $s \in S$ ($S = \prod_{i=1}^N S_i$)下,系统选择执行联合任务 $j \in JB$ ($JB = \prod_{i=1}^N B_i$)时获得的效用, $Q_i: S \times JB \rightarrow \mathbb{R}$ 是机器人*i*的期望累计折扣奖赏函数; Q 表示多智能体团队在联合状态 $s \in S$ 下执行联合任务 $j \in JB$ 获得的团队效用。

任务分配的目的就是要寻求一个最优任务分配策略,即寻求博弈 Γ 的一个Nash平衡点 π^* 。一个策略 $\pi = (\pi_0, \pi_1, \dots, \pi_i)$ 是定义在任务分配的整个过程上,其中 π_i 称为在时刻*t*的决策法则。多机器人系统在各个机器人目标一致的情况下有如下定理。

定理1 若多机器人系统中各机器人目标一致,则系统的效用*Q*可以表示成单个机器人效用 Q_i 的单调增函数 $Q = f(Q_1, Q_2, \dots, Q_N)$ 。

证明 由于各机器人的目标一致,机器人*i*目标的实现意味着机器人*j* ($j \neq i, i, j \in N$)目标的实现,因此,机器人*i*执行任务的效用 Q_i 增加,意味着 Q_j 增加,所以:

a) $Q_i \uparrow \rightarrow Q_j \uparrow; j \neq i; j = \{1, 2, \dots, N\}$ 。

b) 整个系统的效用 $Q \uparrow$ 。

c) Q 可以表示成单个机器人效用 Q_i 的单调增函数 $Q = f(Q_1, Q_2, \dots, Q_N)$ 。证毕。

在实际应用中,各机器人目标一致通过设计合适的效用函数,使得机器人效用函数 Q_i 为系统中另外机器人执行任务的效用 Q_j ($j \neq i, i, j \in N$)的增函数来体现,从而系统的效用为单个机器人效用的增函数,单个机器人效用的最大化将使得整个多机器人系统的效用最大化,因此可以通过最大化个体机器人的效用使得整个系统的效用最大化。

如图1所示为两个足球运动员之间的博弈矩阵,行和列分别代表两个球员某一时刻的候选任务,矩阵的元素表示执行一定任务后两个运动员获得的效用。该博弈中,两个运动员需要在不知道队友策略的情况下执行截球任务或者跑到防守位置,

只有两人选择的任务不同才分别获得效用2,否则都只能获得效用0。因此,两个球员需要各自选择效用最大化的任务使得整个系统的效用增加,实现系统目标。

用 b_{it}^* 表示机器人 i 在观测状态为 s_{it} 以及其他机器人执行联合任务 b_{-it} 下的最优任务选择,若此时的最优策略为 π_i^* ,则 $\pi_i^* \parallel b_{it}^* = \pi_i^*$ 。因此,每个机器人在其观测状态为 s_{it} 时,选择其最优任务 b_{it}^* :

$$b_{it}^* = \arg \max_{b_i \in B_i} Q(s_{it}, b_i) \quad (2)$$

使得个体机器人的效用最大化,进而使得整个系统在时刻 t 的效用最大化。由于在动态决策前获得如图1所示的决策矩阵,本文中单个机器人通过基于自适应模糊神经推理系统的效用评价网络,采用增强学习在动态环境下对执行不同任务的效用评价^[4],求解其任务选择策略。

3.2 基于博弈论的任务分配步骤

基于博弈论的多机器人系统任务分配方法的基本工作过程如图2所示。方法的关键是通过设计合适的效用函数,使得系统的效用为单个机器人效用的增函数。在此前提下,系统中的每个机器人按如下步骤进行决策,通过增强学习逐渐获得各自的最优策略,使整个系统的效用最大化:

- 观测当前环境状态 s_{it} ,提取出与任务分配相关的状态向量。
- 评价任务 $b_i \in B_i$ 的效用值,并选择一个任务 b_{it} 。
- 执行任务 b_{it} 。
- 观测新的状态 $s_{i(t+1)}$,从环境获得奖赏 r_{it} 。
- 更新任务的效用值。
- 重复步骤 a) ~ e),直至结束。

乙

	截球	防守
甲 截球	0/0	2/2
防守	2/2	0/0

图1 两人博弈示例

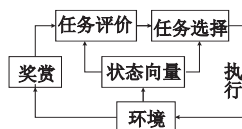


图2 基于博弈论的任务分配方法基本流程

4 实验与讨论

考虑多个机器人的排雷问题。假设有一片雷区,范围已知,但其中雷的具体位置未知,假设为随机分布。采用多个排雷机器人组成一个团队对该雷区进行清除,目标是尽快排除所有雷。排雷机器人能够对自身进行定位,也能够探测到距自己一定范围内的雷。为了将研究的重点放在任务分配问题上,不考虑机器人运行的底层动作,并假设机器人运动速度相同;同时也不考虑机器人在执行排雷任务时所消耗的时间,即假设当机器人移动到雷所在的位置时,就认为该雷被清除。机器人之间能够进行通信,但是不保证通信的可靠性,通信可以通过广播方式或者点对点方式进行。另外,假设每个机器人都能够独立完成任意一个排雷任务,场地为 $2\,000 \times 2\,000$,机器人的速度均为1单位/周期,观测半径为100单位。

为了验证基于博弈论的任务分配算法 GT-MRTA 的性能,将 GT-MRTA 与集中式任务分配算法^[6]、基于行为的任务分配算法^[6,7]、基于市场机制的任务分配算法^[8]实现扫雷任务的结果进行了比较研究。实验中,机器人和任务的初始位置随机产

生,GT-MRTA 实现扫雷任务时,采用基于自适应神经-模糊推理系统的效用评价算法^[4]进行效用评价。首先比较了具有不同系统规模(系统中机器人个数)的多机器人系统分别采用上述四种任务分配方法时单次完成10个任务的时间,结果如图3所示。从图中可以看出,集中式任务分配方法可以获得全局最优;基于博弈论的多机器人系统任务分配方法 GT-MRTA 次之,略优于基于市场机制的任务分配方法;基于行为的任务分配方法随着机器人数量的增多,完成相同数量任务的时间反而增加,这是由于各机器人各自为政,不能配合完成任务,执行任务的过程中产生较多的冲突、干扰。随着机器人数量的增多,四种任务分配方法中的单个机器人完成任务的平均时间均呈下降趋势(图4)。

若考虑计算时间(图5),则当机器人数量大于8时,GT-MRTA 耗时最少,其余三种方法在机器人数量增大到一定规模时,完成任务的时间都有所增加,而基于博弈论的多机器人系统任务分配算法计算量变化不大。每种规模的系统分别采用四种任务分配算法运行100次,从完成任务的平均时间比较(图6)以及包含计算时间在内的平均时间比较(图7)中可以得到类似的结论。由于机器人个数大于5时,集中式任务分配算法需要大量的时间进行任务分配,因此本实验没能给出系统采用集中式任务分配方法的完整结果。

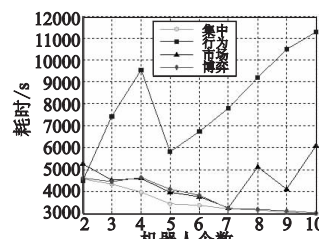


图3 单次运行耗时比较

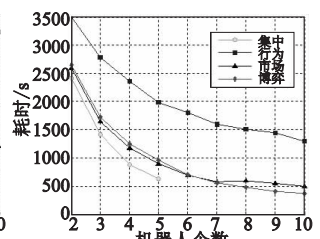


图4 单机器人平均耗时比较

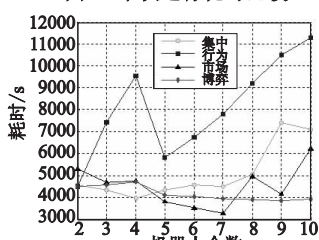


图5 单次运行耗时(包括计算时间)比较

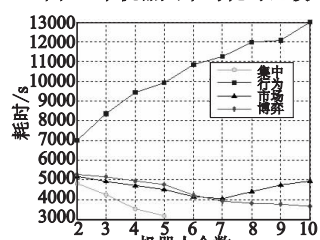


图6 100次运行平均耗时比较

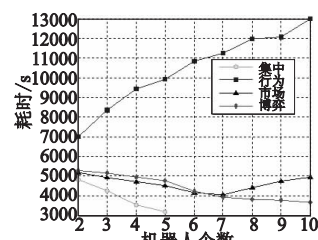


图7 100次运行平均耗时(包括计算时间)比较

在系统通信失效率分别为10%、20%时,将集中式任务分配算法、基于市场机制的任务分配算法和基于博弈论的多机器人系统任务分配算法完成任务的时间与各方法在无通信失效情况完成任务的时间进行了对比,结果如表1.2所示。

从表中可以看出,通信失效使得系统完成任务的时间均有所增加。集中的任务分配算法和基于市场机制的任务分配算法受通信失效的影响较大,通信失效10%时,完成任务

的时间分别增加约 20% 和 10%, 而 GT-MRTA 的性能所受影响较少(完成任务的时间约增加 3%)。随着通信失效率的增加,完成任务的时间相应增加,但总体来看,通信失效对 GT-MRTA 的性能影响较少,因此,其对通信失效具有很好的鲁棒性。

表 1 通信失效 10% 与无失效情况下 100 次完成任务的平均时间比

机器人	方法		
	集中	市场	博弈
2	1.139004	1.086873	1.045455
3	1.159251	1.079268	1.017442
4	1.290141	1.095137	1.024194
5	1.326019	1.095344	0.958246
6	/	1.088942	1.007092
7	/	1.08889	1.032995
8	/	1.065463	1.026178
9	/	1.088608	1.018519
10	/	1.092742	1.024523

表 2 通信失效 20% 与无失效情况下 100 次完成任务的平均时间之比

机器人	方法		
	集中	市场	博弈
2	1.168050	1.125483	1.098485
3	1.222482	1.142276	1.034884
4	1.363380	1.156448	1.052419
5	1.407524	1.137472	1.029228
6	/	1.168269	1.092199
7	/	1.116049	1.109137
8	/	1.119639	1.070681
9	/	1.111814	1.047619
10	/	1.139113	1.043597

(上接第 379 页)

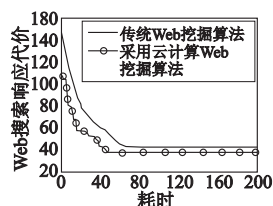


图3 Web搜索响应代价

4 结束语

在 Internet 网络日益发达的今天,更加强调数据挖掘的重要性。在数据容量越来越大的今天,如何能够更快更好地从 Web 数据挖掘中发现更多有价值的数据是现在研究的方向,特别是在云计算概念提出之后,可以将 Web 数据挖掘更好地与云计算结合起来。本文提出的三次选择算法可以更好地缩短 Web 挖掘的时间,提高挖掘效率,能够更好地面向商业应用。

参考文献:

- [1] BARROSO L A, DEAN J, HOLZLE U. Web search for a planet: the Google cluster architecture[J]. IEEE Micro, 2003, 23(2): 22-28.
- [2] DEAN J, GHEMAWAT S. Map/Reduce advantages over parallel databases include storage-system independence and fine-grain fault tolerance for large jobs[J]. Communications of the ACM, 2010,

5 结束语

本文对多机器人系统任务分配策略进行了形式化描述,为任务分配方案的求解提供了一种新的数学描述工具;针对多机器人系统中机器人决策之间的相互依存性,引入博弈论的思想分析了多机器人系统的任务分配问题,提出了一种基于博弈论的多机器人系统任务分配算法。实验表明该算法复杂度较低,计算量较小,鲁棒性较好,获得的任务分配方案质量较高。

参考文献:

- [1] 张芳. 面向协作多移动机器人系统的再学习方法研究[D]. 上海:上海交通大学, 2002.
- [2] 黎萍, 杨宜民. 多机器人系统任务分配的研究进展[J]. 计算机工程与应用, 2008, 44(17): 201-205.
- [3] 范如国, 韩民春. 博弈论[M]. 武汉: 武汉大学出版社, 2006.
- [4] LI Ping, YANG Yi-min, LIAN Jia-le. Layered task allocation in multi-robot systems[C]//Proc of Global Congress on Intelligent Systems. 2009: 62-67.
- [5] SHAPLEY L. Stochastic games[C]//Proc of National Academy of Sciences of the United States of America. 1953: 1095-1100.
- [6] VELOSO M, STONE P. Individual and collaborative behaviors in a team of homogeneous robotic soccer agents[C]//Proc of ICMAS. 1998: 309-316.
- [7] DIAS M B. TraderBots: a new paradigm for robust and efficient multi-robot coordination in dynamic environment[D]. Pittsburgh, Pennsylvania: Robotics Institute, Carnegie Mellon University, 2004.
- [8] DIAS M B, STENTZ A. A market approach to multirobot coordination, CMU-RI-TR-01-26[R]. Pittsburgh, Pennsylvania: Robotics Institute, Carnegie Mellon University, 2001.

53(1): 72-77.

- [3] CANNATARO M, TALIA D, TRUNFIO P. Knowledge grid: high performance knowledge discovery service on the grid[C]//Proc of the 2nd International Workshop on Grid Computing. London: Springer-Verlag, 2001: 38-50.
- [4] GILLICK D, FARIA A, De NERO J. Map/Reduce: distributed computing for machine learning[R]. 2006.
- [5] JEFFREY D, SANJAY G. Map/Reduce: simplified data processing on large clusters[J]. Communications of the ACM, 2008, 51(1): 107-113.
- [6] 吴宝贵, 丁振国. 基于 Map/Reduce 的分布式搜索引擎研究[J]. 现代图书情报技术, 2007, 2(8): 52-55.
- [7] 孙瑞锋, 赵政文. 基于云计算的资源调度策略[J]. 航空计算技术, 2010, 40(3): 103-105.
- [8] 席景科, 闫大顺. Web 数据挖掘中数据集成问题的研究[J]. 计算机工程与设计, 2006, 27(8): 1366-1368.
- [9] 纪俊. 一种基于云计算的数据挖掘平台架构设计与实现[D]. 青岛: 青岛大学, 2009.
- [10] 万至臻. 基于 Map/Reduce 模型的并行计算平台的设计与实现[D]. 杭州: 浙江大学, 2008.
- [11] 陆小丽, 何加铭. 基于 Map/Reduce 的索引数据云存储模型研究[J]. 宁波大学学报: 理工版, 2011, 24(3): 29-33.
- [12] 王鹏. 云计算的关键技术与应用实例[M]. 北京: 人民邮电出版社, 2010.