

一种基于理性原则的多机器人协调控制¹⁾

吴智政 席裕庚

(上海交通大学自动化研究所 上海 200030)

(E-mail: ygxi@mail. sjtu. edu. cn)

摘 要 自组织多机器人系统具有一定的柔性,且鲁棒性强,但系统建模复杂,优化协调控制困难.以多机器人系统的自组织自约束决策模型为基础,提出了一种多机器人系统的理性协调控制方法,并分析了这种方法与最优控制方法的关系.

关键词 多机器人系统,决策模型,理性协调控制.

COORDINATION CONTROL FOR MULTI-ROBOT SYSTEM BASED ON RATIONAL PRINCIPLE

WU Zhizheng XI Yugeng

(Institute of Automation, Shanghai Jiaotong University, Shanghai 200030)

Abstract The self-organizing robot system is flexible but difficult to coordinate the global system behavior with the optimal principle because of the ill-formed problem representation. Based on the self-organizing self-constraint model, a new rational strategy for controlling the behavior of multi-robot system is presented in this paper. The relationship between the optimal method and the rational method is also analysed.

Key words Multi-robot system, decision model, rational coordinated control.

1 引言

近年来,随着多机器人系统在自动制造、柔性生产和复杂装配等领域的应用,多机器人系统的研究已得到了广泛的重视.分布式多机器人系统因其没有中心规划体,各机器人之间的地位相同,协调行为是通过从周围环境获得的局部信息进行的,因而具有很强的鲁棒性和智能性.但系统中因没有中心协调器,要实现最优控制却非常困难.所以为多机器人系统寻找一种算法简单、实时性好的分布式控制方法非常必要.本文以理性决策原则为基础,结合多机器人系统的自组织自约束决策模型,提出了一种理性协调控制方法,这种

1) 国家自然科学基金(69889501)及国家 863 计划(9805-18)资助项目.

方法不但算法简单,而且与单个机器人系统的信息拥有量和计算能力相适应,具有广阔的应用前景.

2 决策系统的理性化准则^[1,2]

在工程应用及大系统理论中,大部分控制思想都基于最优原则,通常一个系统要实现最优控制需满足两个基本条件.首先,系统要存在最优解;其次,在寻找最优解的过程中,要能获得系统足够的信息.但在近年来发展的多智能体系统中,最优思想并不总能实现.因为这类系统往往没有确切的系统模型,而且也很难找到一个可以实现极值化的系统性能指标.在这种情况下人们越来越注重系统控制中的理性原则.

Simon 最早在 1955 年提出了理性选择的概念,建立了相应的行为模型,并在 1986 年成功地应用于经济和生物领域^[1],1996 年 Stirling 等人提出了决策控制系统中的理性概念^[2],现对决策系统的理性化准则进行简单描述.

在系统理性控制问题中,设计准则一般要考虑两个方面,一个是控制的最终结果(定义为准确度性能指标),另一个是控制所需的费用(定义为排斥度性能指标).根据具体需要建立决策系统性能指标后,不是对性能指标函数求极值,而是定义一个性能指标的最小标准或阈值,如果决策空间中的解不能满足这个标准,则这些解是无理的,而那些满足最小标准的解则认为是有理的.令 f_A 和 f_R 分别表示系统的准确度性能指标和排斥度性能指标.令 Θ 表示系统可能的自然状态.系统的每一个自然状态都可被表示成一个随机变量或状态矢量 θ ,它的概率度为 $P_c: B \rightarrow [0,1]$,其中 B 为 Θ 中的 Borel 域.令 F_C 为 θ 的累积分布函数,它的概率密度函数为 $f_c(\theta) = \frac{d}{d\theta} F_C(\theta)$,此式对于所有的 $\theta \in \Theta$ 都成立.现对指标函数 f_A 和 f_R 进行归一化处理,对于所有的自然状态变量 θ 都必须满足

$$1 \geq f_A(u; \theta) \geq 0, \forall u \in U, \quad 1 \geq f_R(u; \theta) \geq 0, \forall u \in U, \quad (1)$$

其中 U 为系统的决策空间.换言之,通过自然状态变量 θ 的参数变换,指标函数 f_A 和 f_R 成为一个概率密度函数,如果决策空间 U 为离散空间,则为概率聚集函数.通过以上变换,对于任何的 Borel 子空间 $V \subset U$,系统的准确度指标和排斥度指标可表示为

$$P_A(V; \theta) = \int_V f_A(u; \theta) du, \quad P_R(V; \theta) = \int_V f_R(u; \theta) du. \quad (2)$$

对以上两个性能指标进行凸组合,则可获得一个系统决策性能指标,称之为系统认识度性能指标函数(epistemic utility function)

$$P_e(V; \theta) = \beta P_A(V; \theta) + (1 - \beta)(1 - P_R(V; \theta)), \quad (3)$$

其中 $0 \leq \beta \leq 1$. 对其进行正的线性变换后,可得一个相等的系统认识度性能指标函数为

$$P_e(V; \theta) = P_A(V; \theta) - b P_R(V; \theta) = \int_V [f_A(u; \theta) - b f_R(u; \theta)] du, \quad (4)$$

其中 $b = (1 - \beta) / \beta$. 显然系统认识度性能指标函数 P_e 是通过自然状态变量 θ 参数化后获得的,而自然状态变量又具有随机变量的特点,因而在 Borel 域 Θ 中,系统期望的认识度性能指标函数可表示为

$$\bar{P}_e(V) = \int_{\Theta} \int_V [f_A(u; \theta) - b f_R(u; \theta)] f_c(\theta) d\theta du = \int_V [\bar{f}_A(u) - b \bar{f}_R(u)] du, \quad (5)$$

其中 $\bar{f}_A(u) = \int_{\Theta} f_A(u; \theta) f_C(\theta) d\theta$; $\bar{f}_R(u) = \int_{\Theta} f_R(u; \theta) f_C(\theta) d\theta$.

针对获得的系统认识度性能指标函数 $\bar{P}_\epsilon(S)$, 作如下定义.

定义 1. 令 F 为在系统决策空间 U 中的 Borel 域, 那么构造系统的决策解集空间为

$$S = \arg \max_{V \in F} \bar{P}_\epsilon(V). \quad (6)$$

假如 $\bar{P}_\epsilon(S) > 0$, 那么

$$\bar{f}_A(u) \geq b \bar{f}_R(u) \quad \forall u \in S. \quad (7)$$

满足方程式(7)的 S 称为系统的理性决策解集.

3 多机器人系统的理性决策控制

分布式多机器人系统具有一定的柔性和鲁棒性, 但系统建模复杂且不确定, 进行优化协调控制非常困难, 所以建立系统恰当的动态模型非常必要. 多机器人系统的自组织自约束模型是一种非常有效的方法^[3,4]. 多机器人系统在执行任务的过程中, 每一机器人将与其它机器人和环境发生通讯或感知交互, 同时整个系统在机器人集体行为作用下也不断发生变化. 为简化计, 相对于机器人决策过程, 可认为系统变化缓慢, 这样, 针对每个机器人来说, 短期规划和决策所需的系统约束条件都是静态的. 总的看来, 机器人实际上在遵守着自己不断创造出的约束条件. 这种抽象的动态模型, 我们称之为自组织自约束动态模型.

根据 Jerne 的生物细胞网络模型^[4,5], 可以获得相应的自组织自约束多机器人系统的数学模型. 把机器人和环境都考虑为模型中的各元素, 且根据不同的策略集, 把机器人分解成相应的机器人群, 系统中各元素之间的交互作用也被分成两类: 一类是刺激交互 (stimulant interaction), 在这种交互作用下, 机器人将保持现有的策略不变; 另一类是压制交互 (suppressive interaction), 在这种交互作用下, 机器人将改变现有的策略到另一策略集.

假设系统拥有的策略集共有 N 种, 定义相应的策略系数为 $\{x_1, x_2, \dots, x_N\}$, 而 n 种环境约束条件系数分别为 $\{y_1, y_2, \dots, y_n\}$, 方程(8)给出了机器人系统的策略系数变化率

$$\sum_{j=1}^N p_{ij} x_i x_j - \sum_{j=1}^N s_{ij} x_i x_j + \sum_{k=1}^n p_{ik} x_i y_k - \sum_{k=1}^n s_{ik} x_i y_k + \lambda x_i = \tau \frac{d}{dt} x_i. \quad (8)$$

这里 p_{ij} 为具有第 i 种策略类型的机器人与具有第 j 种策略类型的机器人之间的刺激交互度, 在刺激交互的作用下, 具有第 i 种策略类型的机器人数目将增多; s_{ij} 为具有第 i 种策略类型的机器人与具有第 j 种策略类型的机器人之间的压制交互度, 在压制交互的作用下, 具有第 i 种策略类型的机器人数目将减少; p_{ik} 和 s_{ik} 分别为具有第 i 种策略类型的机器人与第 k 种环境约束条件之间的刺激交互度和压制交互度; 方程式左边的最后一项表示具有第 i 种策略类型的机器人在无任何交互作用下的自然死亡率.

在以上的策略模型基础上, 本文将针对多机器人系统提出一种理性决策控制方法. 机器人在作出决策之前, 必须先与其它机器人和周围环境进行通信交互, 并把这种交互分为刺激性和压制性两类, 然后基于理性准则, 构造出系统的认识度性能指标函数.

3.1 单个机器人的理性决策解集构成

对系统中的单个机器人,令系统在 k 时刻的状态变量为决策系统的自然状态变量,即 $\theta = X(k)$. 为了处理简化,假设系统的自然状态变量的概率度为 1,因而可得

$$f_c(X) = \delta[X - X(k)], \quad (9)$$

即为 Dirac delta 函数. 因而有

$$\bar{f}_A(u) = f_A[u, X(k)], \quad \bar{f}_R(u) = f_R[u, X(k)]. \quad (10)$$

现定义机器人在某一轮决策过程中所有获得的刺激交互值为决策系统的准确度性能指标函数 $J_A(u)$, 即为

$$J_A(u) = f_1^T(X(k), u, k) P f_1(X(k), u, k), \quad (11)$$

其中 f_1 为包含刺激机器人现行策略的表达函数. 这里定义 f_1 为时间消耗函数.

相同的,定义机器人在这一轮决策过程中所有获得的压制交互值为决策系统的排斥度性能指标函数 $J_R(u)$, 即为

$$J_R(u) = f_2^T(X(k), u, k) Q f_2(X(k), u, k), \quad (12)$$

其中 f_2 为包含压制机器人现行策略的表达函数. 这里定义 f_2 为系统能量消耗函数.

为了获得系统最大的动态变化范围,把 J_A 与机器人在决策空间中所可能获得最大的 J_A 值进行相减,得偏差为

$$J'_A(u) = \max_{z \in u_k} J_A(z) - J_A(u), \quad (13)$$

其中 u_k 为机器人在 k 时刻可采用的策略. 作如下归一化处理,获系统新的准确度性能指标 \bar{f}'_A 为

$$\bar{f}'_A(u) = \frac{J'_A(u)}{\sum_{u \in u_k} J'_A(u)}. \quad (14)$$

相应把 J_R 与机器人在决策空间中所可能获得最小的 J_R 值进行相减,可得偏差为

$$J'_R = J_R(u) - \min_{z \in u_k} J_R(z). \quad (15)$$

作如下归一化处理,获得系统新的排斥度性能指标 \bar{f}'_R 为

$$\bar{f}'_R(u) = \frac{J'_R(u)}{\sum_{u \in u_k} J'_R(u)}. \quad (16)$$

然后利用式(7)获得单个机器人在 k 时刻可采用的理性决策解集 S_k 为

$$S_k = \{u \in u_k : \bar{f}'_A(u) - b \bar{f}'_R(u) \geq 0\}, \quad (17)$$

其中 b 为决策系统的意向调节参数,通过调整合适的参数值可保证 $S_k \neq \emptyset$.

对机器人来说,如果现行的策略在理性决策解集 S_k 中,则保持现在的控制策略不变;如果现行的策略不在理性决策解集 S_k 中,则改变策略到理性决策解集 S_k 中.

3.2 协调算法

为了进一步阐述多机器人系统的理性协调控制问题,设想有一多移动机器人系统,它们需把许多随机分散的物体搬运到另一指定地点,如货运码头等. 系统中所有机器人都完全独立,且地位平等,每个机器人都有自己的感知系统和通讯系统,能感知周围环境的变化及与其它机器人之间进行协调通讯. 系统的一些具体规定如下:

1) 期望约束. 所有的搬运任务希望能在尽快的时间内结束,同时又要尽量节约整个

系统的能量消耗.

2) 系统规定. ①所有机器人和被搬运的物体随机分布在工作环境中. ②系统中每个机器人都根据自己所选择的策略去搬运相应的物体, 每个机器人都只有两种可选择策略: 高能量消耗策略和低能量消耗策略. 如果机器人采用第一种策略, 那么其完成这次任务所需的时间较少, 而所消耗的能量较多; 如果机器人采用第二种策略, 那么其完成这次任务所需的时间较多, 而所消耗的能量较少. 如果机器人通过理性决策判断后, 策略解集为空, 则选择第二种策略方案. 当然在实际应用中, 可根据需要增加机器人可选择的策略空间. ③每个机器人都选择在自己感知空间内离自己距离最近的物体为搬运对象, 且一次只能搬运一个物体. ④机器人一旦开始搬运某个物体后, 在完成这次任务之前不能改变当前策略. ⑤设系统初始参数为: 机器人数量 $N_r=10$, 被搬物体数 $N_c=100$.

3) 算法. 为简化起见, 机器人与周围环境条件的交互作用不予考虑. 每个机器人与其它机器人之间的通讯都是有序的, 且采用应答的方式. 在交互中每个机器人从其他机器人获取两类信息: 刺激交互信息和压制交互信息(在本文假设下分别为其他机器人按现行策略完成任务所消耗的时间和能量), 其中刺激交互信息构成了该机器人决策系统的准确度性能指标 $J_A(u)$, 而压制交互信息构成了该机器人决策系统的排斥度性能指标 $J_R(u)$. 算法以机器人 i 为例.

步骤 1. 参数初始化(在 $t_i=T_i$ 时刻): 机器人 i 初始化参数 $J_A(u)=J_R(u)=0$.

步骤 2. 机器人 i 更新自己的交互值($T_i \leq t_i \leq T_i + \Delta T$).

1) 更新刺激交互值. 当机器人 i 与机器人 m 发生相互作用(即通信)时, 机器人 i 便在自己原有的刺激交互值 $J_A(u)$ 中增加一个与机器人 m 按现行策略完成任务所消耗时间相对应的 $\Delta J_A(u)_{(i/m)}$, 方程式(18)为机器人 i 在一轮交互作用后, 系统总共具有的刺激交互值. 其中 N_r 为系统的总机器人数量; $K(\cdot)$ 项表示在 T_i 时刻机器人 i 能否与相应的机器人发生正常的交互作用.

$$J_A(u) = \sum_{t_i=T_i}^{T_i+\Delta T} \sum_{m=1}^{N_r-1} \{K(t_{(i/m)}) \times \Delta J_{A(i/m)}\}; \quad K(t_{(i/m)}) = \begin{cases} 1, & \text{交互成功,} \\ 0, & \text{交互失败.} \end{cases} \quad (18)$$

2) 更新压制交互值. 当机器人 i 与机器人 m 发生相互作用时, 机器人 i 同时也在原有的压制交互值 $J_R(u)$ 中增加一个与机器人 m 按现行策略完成任务所消耗能量相对应的 $\Delta J_R(u)_{(i/m)}$, 方程式(19)为机器人 i 在一轮交互作用后, 系统总共具有的压制交互值.

$$J_R(u) = \sum_{t_i=T_i}^{T_i+\Delta T} \sum_{m=1}^{N_r-1} \{K(t_{(i/m)}) \times \Delta J_{R(i/m)}\}; \quad K(t_{(i/m)}) = \begin{cases} 1, & \text{交互成功,} \\ 0, & \text{交互失败.} \end{cases} \quad (19)$$

步骤 3. 计算决策系统的认识度性能指标. 在 $t_i=T_i+\Delta T$ 时刻, 机器人 i 通过一轮交互作用获得的 $J_A(u)$ 和 $J_R(u)$, 实际上反映了它对整个系统按目前策略运行所消耗时间与能量的认识, 在此基础上, 该机器人便开始计算自己的准确度性能指标 $\bar{f}_A(u)$ 和排斥度性能指标 $\bar{f}_R(u)$. 最后计算认识度性能指标函数 \bar{P}_e .

步骤 4. 作出决策. 如果认识度性能指标 \bar{P}_e 满足期望的最小标准, 表明机器人 i 认为整体系统按现行策略运行的时间与能量消耗尚属合理, 便保持现行策略不变, 如果不满足期望的最小标准, 则更换为另一策略.

步骤 5. 回到步骤 1, 开始新一轮的决策选择.

3.3 计算机仿真

对以上的多机器人系统进行计算机仿真,结果如图 1 和图 2 所示.当选择参数 b 为 0.8 时,图 1(a)为每个机器人完成所有任务所花费的时间,图 1(b)为在相同条件下每个机器人所消耗总的能量;当调节参数 b 从 0.8 变化到 1 时,图 2(a)和图 2(b)分别表示了机器人系统总共所花费的时间值和能量消耗值.仿真结果表明,当调节参数值 b 从小到大变化时,机器人系统总共所花费的时间将减少,而相应所消耗的系统能量将增多,因而可通过选择合适的参数值 b ,来达到期望的时间和能量消耗的平衡点,以便获得理想的系统输出性能.

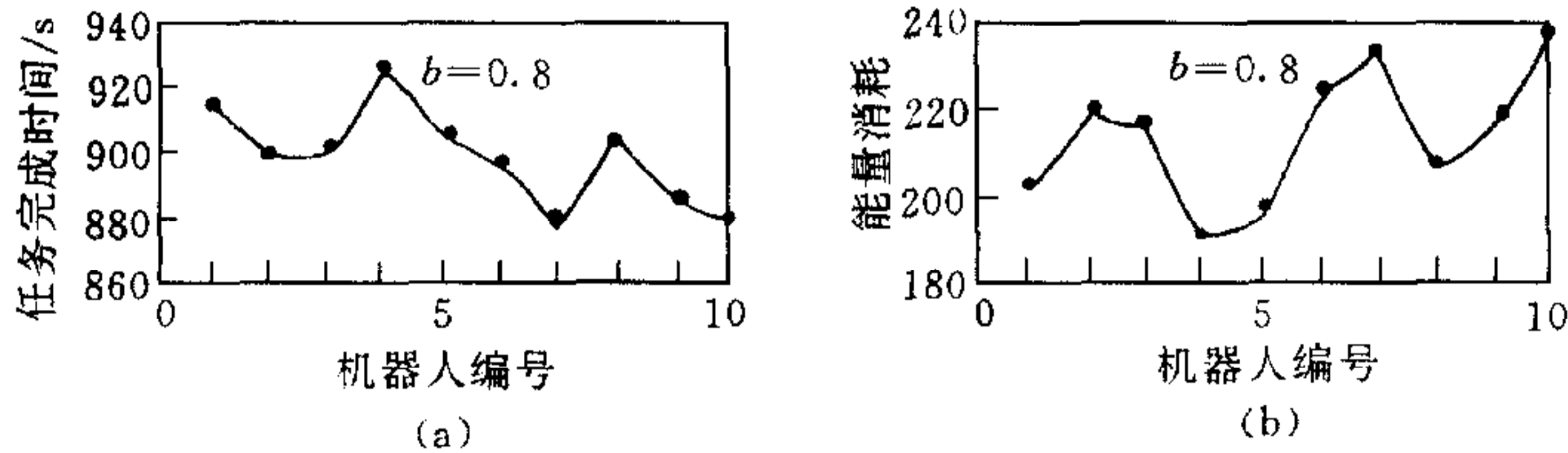


图 1 当参数 b 为 0.8 时各机器人的输出情况

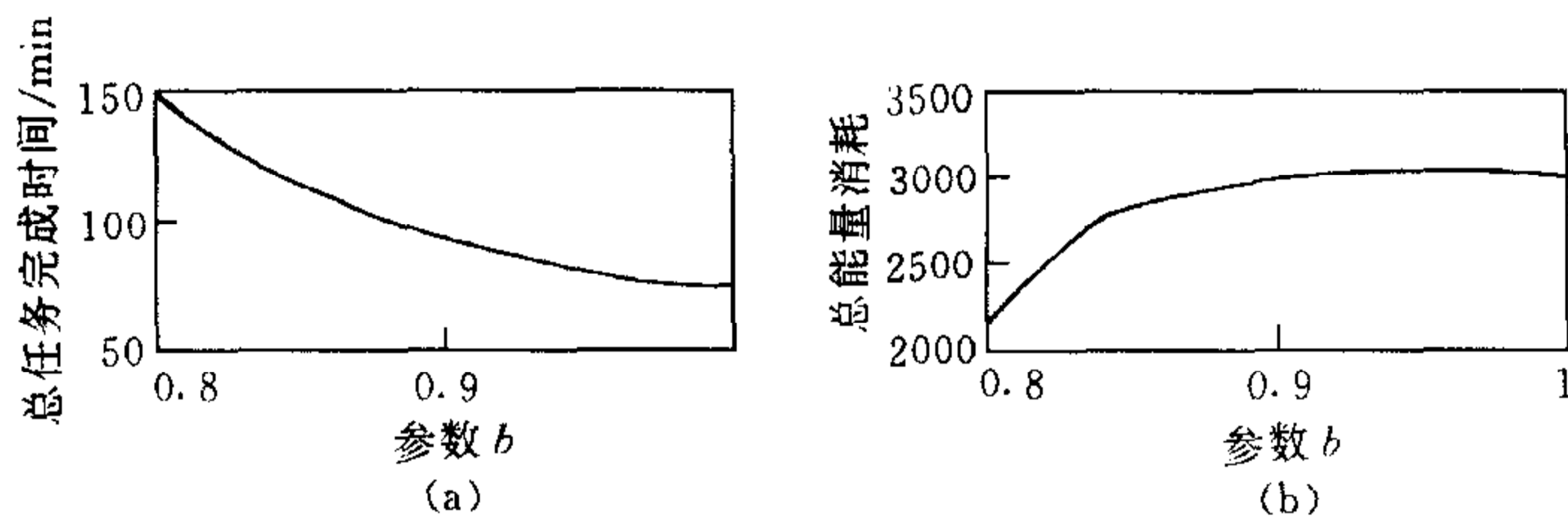


图 2 当参数 b 从 0.8 变到 1 时系统的输出情况

4 多机器人系统的理性决策控制与最优控制的关系

对于上述多机器人系统,如果系统中有一中心规划器,并能获得系统足够信息,那么也可对它实施最优控制.现把多机器人系统的状态方程式表示为

$$T(k+1) = f(T(k), u(k), k), \quad T(0) = T_0, \quad k = 0, 1, \dots, \quad (20)$$

其中 f 为系统的时间消耗函数,它反映了在 k 时刻,机器人系统的预计消耗时间矢量 $T(k+1)$ 和消耗时间矢量 $T(k)$ 及控制策略 $u(k)$ 之间的关系.考虑二次调节器问题,即寻找一个控制序列 $[u(0), u(1), \dots, u(k_t)]$,使得控制系统最小化如下的性能指标

$$J = T^T(k_t)PT(k_t) + \sum_{k=0}^{k_t} [u^T(k)R(k)u(k)], \quad (21)$$

其中 $P \geq 0, R(k) \geq 0, k_t$ 为终止时间.根据理性决策控制方法的定义,把上述性能指标的第一项作为系统的准确度性能指标 $J_A = T^T(k_t)PT(k_t)$;把最优性能指标的第二项作为系统的

排斥度性能指标 $J_R = \sum_{k=0}^{k_t} [u^T(k)R(k)u(k)]$.因而传统的最优化性能指标就可重新表示为

$$J = J_A + J_R, \quad (22)$$

而如下的定理正好反映了多机器人系统的理性决策控制与最优控制的关系.

定理 1. 对于二次调节器问题(20)及最优性能指标(21),其中 $U = (-\infty, +\infty)$,如果

系统在时间域 $d=k_t$ 中,能求解出最优控制解,那么理性决策控制系统也必存在一个意向调节参数 $b \in (0, \infty)$,使得系统有如下的一个理性决策解与最优解等同:

$$u_D = \arg \max_{z \in U^d} \{\bar{f}_A(z) - b\bar{f}_R(z)\}. \quad (23)$$

证明. 根据式(14)和(16),可得

$$\arg \max_{z \in U^d} \{\bar{f}_A(z) - b\bar{f}_R(z)\} = \arg \max_{z \in U^d} \{J'_A(z) - b'J'_R(z)\}, \quad (24)$$

其中 $b' = b \frac{\sum_{u \in U^d} J'_A(u)}{\sum_{u \in U^d} J'_R(u)}$. 根据式(13)和(15),可得

$$\begin{aligned} \arg \max_{z \in U^d} \{J'_A(z) - b'J'_R(z)\} &= \arg \max_{z \in U^d} \{-J_A(z) - b'J_R(z)\} = \\ &= \arg \min_{z \in U^d} \{J_A(z) + b'J_R(z)\}. \end{aligned} \quad (25)$$

利用式(22),取终止时刻 k_t ,得如下方程

$$u_D = \min_{z \in U^{k_t}} \{T^T(k_t)PT(k_t) + b' \sum_{j=0}^{k_t-1} Z^T(j)R(j)Z(j)\}. \quad (26)$$

如果选择 $b' = 1$,即 $b = \frac{\sum_{u \in U^d} J'_R(u)}{\sum_{u \in U^d} J'_A(u)}$,则式(26)正好为二次调节器的最优解. 证毕.

5 结论

本文以理性决策原则为基准,以多机器人系统的自组织自约束决策模型为基础,提出了一种多机器人系统的理性协调控制方法,且可通过选择合适的系统意向调节参数 b ,来使系统达到期望的控制性能. 同时本文还证明了多机器人系统理性决策控制方法得到的解集包含了一般的最优控制解. 并通过计算机仿真. 表明了所提出的控制方法的有效性.

参 考 文 献

- 1 Simon H A. Rationality in Psychology and Economics. In: Rational Choice (R. M. Hogarth and M. W. Reder), University of Chicago Press, 1986
- 2 Stirling W C, Goodrich M A *et al.* Procedurally rational decision-making and control. *IEEE Control Systems*, October, 1996: 66~75
- 3 Kokaji S, Murata S, Self-Organization of Mechanical System. In: Distributed Autonomous Robotic System, Berlin: Springer-Verlag, 1994: 237~242
- 4 Mitsumoto N, Fukuda T *et al.* Self-organizing multiple robotic system. IEEE International Conference on Robotics and Automation, 1996: 1614~1619
- 5 Jerne N. K. The Immune system. *Scientific American* 229, 1973: 52~60

吴智政 男, 1971年11月15日出生, 1998年毕业于上海交通大学自动控制理论与应用专业, 获博士学位. 主要研究方向为机器人控制, 非线性系统理论及微机电系统控制.

席裕庚 简介见《自动化学报》1999年第25卷第5期.