doi:10.3969/j.issn.1005-3026.2015.01.025

一种动态博弈的多agent合作机制模型

范思遐,周奇才,熊肖磊,赵 炯
(同济大学 机械与能源工程学院,上海 201804)

摘 要:利用博弈学习思想,以多agent系统为平台,提出一种基于动态无限博弈的多agent合作机制模型,以多阶段邀请、考核模式形成无限次重复博弈结构体.提出信任基准测度评价控制agent博弈选取的盲目性,使其理性计划各阶段决策.通过博弈结果反馈信息,动态调整agent博弈主体的收益函数,控制各agent间的协同合作优先级,实现闭环调控.将多agent组合构成智能结构体的基本单元,利用九宫格实验对该基本单元进行试验测试,验证多工况下,基于多种信任基准条件agent单元体间的协同合作机制.实验表明,信任基准可有效调整agent间的信任等级,促进系统中agent合作频率的提高.

关 键 词:多智能体;合作机制;动态博弈;无限;信任基准
中图分类号:TP 273 **文献标志码**:A **文章编号**:1005-3026(2015)01-0114-06

Multi-agent Cooperation Mechanism Model Based on Dynamic Game

FAN Si-xia, *ZHOU Qi-cai*, *XIONG Xiao-lei*, *ZHAO Jiong*
(School of Mechanical and Engineering, Tongji University, Shanghai 201804, China. Corresponding author: FAN Si-xia. Email: dongxia1249@163.com)

Abstract: A multi-agent cooperation mechanism based on dynamic infinite game model was proposed in a multi-agent platform. The infinitely repeated game structure was formed with multi-stage inviting and evaluating actions. Meanwhile, the trust benchmark was proposed to control agent blindness in selection phase and to make sure it could do the rational stage decision. Through the feedback, the multi-agent cooperation mechanism could not only dynamically adjust the income function of each game-agent, but also control cooperation priority. Also, it could achieve a closed-loop control. Additionally, multi-agent were assembled as a basic component unit of intelligent structure. Nine grids experiment was used to validate cooperation mechanism between multi-agent under multiple loading conditions. Moreover, the collaboration status of the agent components unit was also tested based on different trust benchmark. Experiments show that, trust benchmark can effectively adjust the trust level between agents, and promote cooperation in the high frequency agent system.

Key words: multi-agent; cooperation mechanism; dynamic game; infinity; trust benchmark

博弈理论为研究多agent系统的协作奠定了坚实的基础,并逐步引入MAS(multi-agent systems)工程领域中[1-2],如王冠群等[3]以非合作模式建立了船舶电力系统重构博弈方案;Pendharkar[4]以合作和非合作模式分别建立了无线通信与制造业MAS的博弈选择;Feldman和Tamir[5]利用Nash均衡,验证了MAS中博弈平衡问题的稳定性;宋梅萍等[6]采用Pareto占优解论证agent博弈学习的有效性.上述内容采用静态博弈的研究方法,以参与者同时选择各自本次博弈最优决策为基础,对MAS各阶段选择进行了相应的优化处理.但由于多智能体系统中,请求处理、决策及任务分配等活动具有随机性强、时效性高、系统工作周期动态变化的特征;同时agent

收稿日期:2014-04-18
基金项目:国家自然科学基金资助项目(51375345);上海市科学技术委员会基金资助项目(08201202103).
作者简介:范思遐(1987-),女,辽宁沈阳人,同济大学博士研究生;周奇才(1962-),男,江苏无锡人,同济大学教授,博士生导师.

具有理性与自私性[7]两大特点，自私性致使 agent 具有一定的争先性[8]，即不遗余力地争取获得任何有利机会；理性的存在导致 agent 在决策时需考虑未来预期，在完成系统工作周期任务时，与其他 agent 的协作行为具有不确定性与无限性，并通过观测到其他 agent 的合作历史记录与选择动作序列抉择合作走向，因此静态博弈无法满足上述需求，需建立基于动态无限博弈的决策选择方法.

本文提出一种基于动态无限博弈的多 agent 合作机制，以多阶段邀请、考核模式形成无限次重复博弈. 以协作优先级主动选取协作智能体，并提出信任基准控制 agent 博弈选取的自私性，使其理性计划阶段决策，通过博弈结果反馈调整优先级，实现闭环调控.

1　基于信任度的动态博弈合作机制

1.1　基于无限次重复博弈的 agent 合作

博弈本质体现的是参与者理性选择的冲突碰撞，而冲突结局可描述成参与者对利益追究的均衡态势. 重复博弈是一类特殊而又重要的动态博弈，由于博弈结构体多次重复出现，一些在一次性博弈中不可能出现的合作行为在重复博弈中却可能出现.

由于单一 agent 不具有足够的资源与能力完成指定任务，将启动请求计划，请求其他 agent 协助完成；主动性与交互性协助其他相关 agent 快速响应，接收请求，部署规划. 而 agent 自私性使其在响应交互时受两个重要因素影响：一是当系统中同构 agent 较多时，任一 agent 害怕完成系统分配任务总量较少，长时期处于空置状态，导致系统阶段更新后，遭到淘汰；二是在与其他 agent 的交互合作中，请求与被请求的次数过低，影响其在系统或层级模块间的熟人等级排列，而遭冷落. 由此 agent 在接收请求行为时有时对完成效果的思考欠缺，将导致对请求 agent 造成一定的损失，当请求 agent 察觉到接收 agent 具有盲目的行为决策后，将影响其之后的交互合作工作. 因此 MAS 中的合作活动可抽象为图 1 所示的多阶段博弈状态，通过多次邀请、考核形成博弈结构体，而将其扩展为 MAS 整体任务规划时，可演化为一种基于无限动态博弈的合作机制.

假设系统中存在 n 个 agent，则存在 2^n 种合作方式. 图 1 以系统中两个 agent 合作博弈为例. 设 agent i，agent $j \in \Gamma$；若 agent i 无法独立完成当前任务时，将请求 agent j 的帮助，同时 agent i 将给予 agent j 一定数额的奖励 r，agent j 若接收请求，将获得奖励 r，否则它可以选择拒绝而加入其他 agent 的请求，由此产生机会成本 C_{op}；当 agent j基于理性选择接收 agent i 的请求时，可与 agent i 合作共同完成系统分配任务，不仅使 agent j 得到奖励 r，也使 agent i 获得收益 e，其净利润为 $e-r$；若 agent j 接收请求时存在自私与盲目性，将导致合作项目无法进行，虽可继续获得收益 r，但 agent i 将造成 $-r$ 的经济损失. 因此对 agent 的合作期望与自私性加以鼓励与限制将对合作博弈达到何种纳什均衡产生重要的影响.

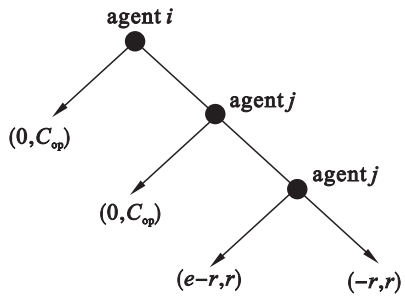

图 1　动态博弈结构体

Fig. 1　Dynamic game structure

1.2　基于信任基准的合作机制

在多智能体系统博弈中，其 agent 的个体收益不仅取决于一次博弈所带来的收益，同时也受博弈次数、长期合作历史记录及未来预期合作的影响. 因此提出信任基准(trust benchmark)评价指标，控制 agent 的理性程度，提高 agent 的合作期望. 信任基准评价 agents 间的相互信任程度，信任基准越高，数值越大，代表双方之间互信指数越高，对对方犯错的宽容度越大，且允许非合作或非理性选择的比率越高，越注重长远利益；反之则亦然. 假设两个 agents 间总共申请合作的次数为 AC(application cooperation)，未合作成功的次数为 NSC(not successful cooperation)，未成功的原因可能由另一个 agent 的拒绝或盲目性造成. 信任基准由式(1)所示.

$$TB = \frac{NSC}{AC}. \tag{1}$$

TB 描述了 agent 博弈合作中的互信信息，将信任基准赋予贴现特性，寻找博弈结构体纳什均衡时，TB 根据定义可表示为[0,1]区间，设被请求合作任务 agent j 在结构体各阶段博弈中，都出于理性选择，认为“执行合作”为最优决策，无限次重复博弈的纯收益现值为 V_e^j，则 V_e^j 可由式(2)表示，并化为式(3).

$$V_{\mathrm{e}}^{j}=r+\mathrm{TB}\cdot V_{\mathrm{e}}^{j}, \tag{2}$$

$$V_{\mathrm{e}}^{j}=\frac{r}{1-\mathrm{TB}}. \tag{3}$$

设如果被请求 agent j 在接收任务时,存在盲目争先性,做出"不执行合作"的决定,不仅使请求 agent i 遭受损失,在以后的合作中做出不请求的决策,也会使其自身收益减少,信用等级有所降低. 如 agent j 在博弈阶段认为"不执行"为最优选择,则收益现值 V_{une}^{j} 即为式(5).

$$V_{\mathrm{une}}^{j}=r+C_{\mathrm{op}}\sum_{i=1}^{+\infty}\mathrm{TB}^{i}, \tag{4}$$

$$V_{\mathrm{une}}^{j}=r+\frac{C_{\mathrm{op}}\cdot\mathrm{TB}}{1-\mathrm{TB}}. \tag{5}$$

对于 agent j 而言,任务选择执行合作认为对其自身价值与收益都有利时,$V_{\mathrm{e}}^{j}\geqslant V_{\mathrm{une}}^{j}$,则

$$\frac{r}{1-\mathrm{TB}}\geqslant r+\frac{C_{\mathrm{op}}\cdot\mathrm{TB}}{1-\mathrm{TB}}, \tag{6}$$

$$\mathrm{TB}(r-C_{\mathrm{op}})\geqslant 0. \tag{7}$$

由式(7)可知,$\mathrm{TB}\geqslant 0$,当 $r-C_{\mathrm{op}}\geqslant 0$ 时,也即 agent i 给予 agent j 的奖励大于等于其选择其他 agent 获得机会成本时,agent j 认为在无限次重复博弈中,当合作收益满足其对合作的期望收益时,选择执行合作认为系统纳什均衡.

若请求者 agent i 在请求 agent j 合作时,都能得到 agent j 的帮助并顺利完成任务,则 V_{e}^{i} 为

$$V_{\mathrm{e}}^{i}=(e-r)+\mathrm{TB}\cdot V_{\mathrm{e}}^{i}, \tag{8}$$

$$V_{\mathrm{e}}^{i}=\frac{(e-r)}{1-\mathrm{TB}}. \tag{9}$$

若请求者 agent i 在请求合作完成任务,发现 agent j 具有盲目性,并对其造成损失时,将会屏蔽 agent i,停止合作任务,其收益现值 V_{une}^{i} 为

$$V_{\mathrm{une}}^{i}=-r+0\cdot\sum_{i=1}^{+\infty}\mathrm{TB}^{i}. \tag{10}$$

对 agent i 而言,选择和 agent j 共同完成任务认为是最优选择时,$V_{\mathrm{e}}^{i}\geqslant V_{\mathrm{une}}^{i}$,则

$$\frac{(e-r)}{1-\mathrm{TB}}\geqslant -r, \tag{11}$$

$$\mathrm{TB}\leqslant\frac{e}{r}. \tag{12}$$

当 $e>r$ 时,agent 间合作可顺利建立,因此 TB 的最低限值可设为 1,符合 TB 定义中数值的最高限制范围. 因此,当合作请求满足合作可建条件时,即 $e>r$,对于具有 $\mathrm{TB}\in[0,1]$ 限定的 agent i 和 agent j 双方,选择共同执行合作任务是系统博弈的纳什均衡.

由 TB 也可分析出 agent 间的信任程度与合作期望值. TB 越小,允许非成功合作的次数越少,agent 两者间更注重眼前利益,对对方的信任程度较低,当双方实际未完成合作次数大于 TB 预设值时,合作将不复存在;$\mathrm{TB}=0$ 是双方合作的极限值,其描述的是"冷酷战略"情形[9],即只要双方有一次合作未成功,从此将互不信任,取消合作,对 agent 间的理性选择提出极高的要求,不仅要求被请求 agent 有求必应,而且限定其按合作计划履行职责,完成任务,否则再无合作任务. TB 越大,描述 agent 双方更注重长远利益,互信等级也将提高,只要双方拒绝合作或未执行合作的比率小于 TB 预设值时,请求合作 agent 将既往不咎,重新发起合作请求,被请求 agent 也将进行新一轮的博弈选择;$\mathrm{TB}=1$ 是双方合作的另一个极限值,描述双方在不考虑合作历史记录下,进行无限次重复博弈合作. 但由于极限值的存在将系统抽象成静态博弈环境,agent 间博弈的选择具有极强的随机性与盲目性,导致系统完成任务效率降低,由此通过 TB 赋值可有效调控 agent 间的合作期望. 因此在 MAS 中引入基于信任基准的无限重复博弈理念,将对 agent 间的合作起到促进作用,通过协作优先级与信任基准的闭环互相调整,增强 agent 做出理性选择的能力.

2 算法描述

1) MAS 根据 agent 的工作特性,初始化 TB,r,e,C_{op}数值,及 agent 协作的等级排序;

2) 接到分配任务后,分解任务,给予各 agent 相应的战略选择空间 S_i 与收益函数 $u_i(s_1,\cdots,s_i,\cdots,s_n)$;

3) 以两个 agent 为合作基本单元发起合作;

4) 进入博弈阶段,选择博弈策略,执行博弈,计算博弈结果,得到相应收入,如$(0,C_{\mathrm{op}})$,$(e-r,r)$,$(-r,r)$等.

5) 判断是否还需和其他 agent 合作,如果是将回到第 3)步,重新组成博弈体,进入 4)步,以 3)步博弈结果产生的合作结构体(结构体中 agent 数量小于等于 2)与新进 agent 进行合作博弈,否则进入第 6)步;

6) 判断系统中是否还有合作任务,如果有,将进入第 3)步,否则转入第 7)步;

7) 计算各 agent 执行合作行为所得收益;

8) 根据第 7)步输出的收益值,适当调整更新系统 TB,r,e,C_{op}数值;

9) 结束 MAS 阶段任务分配.

3　实验分析

实验采用九宫格游戏法，测试基于信任基准 agent 无限博弈合作机制的效用性. 九宫格游戏如图 2 所示可描述无限重复动态博弈的环境条件. 以 agent 体的合作博弈作为九宫格测试基本单元，agent 在格中的初始化位置设为 MAS 阶段任务分配的初始点，初始点为随机设置，agent 的目标终点为 MAS 阶段任务分配的终止点；在 agent 的自由移动中，agent 的相遇描述为 MAS 分发给 agent 系统任务，需要进行协商合作完成. Agent 的合作并执行任务表示接受完成任务，agent 的合作却不执行表示 agent 具有盲目性的选择特征，agent 的不合作代表无法完成任务，由于系统中 agent 相遇的偶然随机性，表达了系统任务分配中合作博弈的动态与无限重复性；同时根据信任基准(TB)赋值的不同控制了 agent 博弈选择执行合作的理性程度.

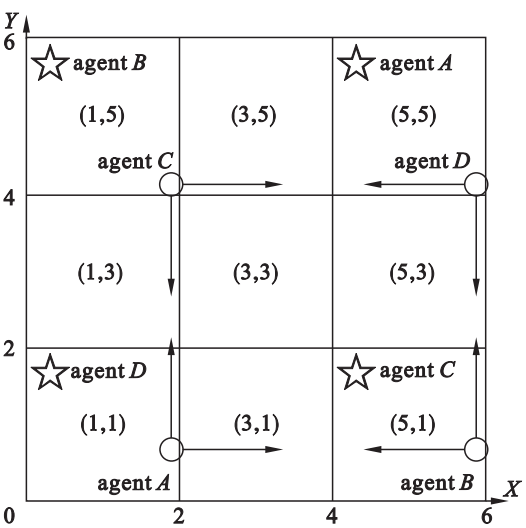

图 2　九宫格 agent 移动状态及移动坐标(☆为 agent 终点)

Fig. 2　Moving state and coordinates of agents in grids(☆ is agent terminal position)

本文采用 4 个 agent 组成的 MAS 系统作为测试基本单元，该基本单元较好地描述了 agent 间的合作关系，将合作形式根据发起者不同形成 64 种组合. 并根据完成任务先后、合作与执行任务状态、博弈选择的理性程度，将 agent 间的合作细化为 312 种形式，图 2 所示 agent A，agent B，agent C 与 agent D 的动作选择空间为{up，down，left，right}四种动作，行为选择为随机选择；战略选择空间表示各决策点的位置坐标，将九宫格转换为 $X-Y$ 坐标系，以每个格子的中点坐标作为决策选择坐标位置(如图 2 所示)，则 $S_i=\{(1,1),(1,3),\cdots,(5,5)\}$，单一战略 S_i 表示 agent A，agent B，agent C 与 agent D 在九宫格中位置状态. agent A，agent B，agent C 与 agent D 起点分别为(1，1)，(5，1)，(1，5)，(5，5)；终点分别为(5，5)，(1，5)，(5，1)，(1，1).

为测试信任基准的有效性，实验假设 agent A 为合作的发起人，在 agent A 与其他 agent 相遇时，agent A 向其他 agent 发起合作，若两者选择合作，将保留在同一个格子中，若合作者理性执行完成合作任务，各获得$(e-r,r)$的纳什均衡收益；若合作者具有选择争先盲目性，并没有按合作决定执行任务时，agent 将得收益为$(-r,r)$，对合作请求者造成一定的损失；反之，合作者选择不进行合作，各 agent 将返回上一步选择的格子中，获得$(0,C_{op})$的纳什均衡收益；当 4 个 agent 均完成系统分配的所有任务时，MAS 将结束系统阶段任务分配活动，此实验结束节点设为 4 个 agent 均到达终点.

实验中初始信任基准数值设为 1，描述静态博弈环境，agent 以均等概率理性选择，具有理性与自私性；在评价系统协作等级排序中，根据 agent 的收益情况，采用式(13)作为评价标准. 其中 α 为权重比例，R_{se}为实际执行应答合作任务的收益，R_{ac}为请求合作任务均被执行时的收益，R_{pc}为应答合作任务均被执行时的收益.

$$\alpha\frac{R_{se}}{R_{ac}}+(1-\alpha)\frac{R_{se}}{R_{pc}}. \tag{13}$$

式(13)可化简为式(14)，

$$\alpha\frac{se}{ac}+(1-\alpha)\frac{se}{pc}. \tag{14}$$

se/ac 为合作任务中实际执行的次数与申请合作次数的比值，描述了信任基准不同时对 agent 间合作信任等级的评估；se/pc 为合作任务中实际执行的次数与同意合作次数的比率，刻画了信任基准对 agent 博弈选择中理性程度的限制. 采用比率组合方式作为系统协作等级排序标准，增加 agent 间的可比性，同时消除系统中组合概率的随机性. 当 agent A 为合作申请者时，系统考虑 AB，AC，AD，ABC，ABD，ACD，$ABCD$ 七种合作形式，对信任基准赋值工况的不同，进行 100 次测试，将平均值计入测试结果(由于实验中 $ABCD$ 相遇次数较低，实验分析将不予以考虑)，α 为 0.5.

表 1 为静态系统中 AB，AC，AD，ABC，ABD，ACD 六种合作方式的合作情况，表 2 为根据合作情况的不同，产生系统阶段协作等级排序评估. 静态博弈系统中，理性选择即完成合作任务的实际

执行率不受信任基准限制,按等概率执行选择. 当系统阶段任务分配完成时,根据系统阶段协作排序等级,MAS 中合作申请者将对合作者进行信任基准重置,重置结果如表 3 所示,同时 agent 为提高自身协作等级,增强合作收益与合作次数,防止其他 agent 的冷漠对待及系统的定期淘汰,将对理性选择加以限定,其中理性选择比率为选择执行合作任务的权重,见表 3.

表 1 TB 为 1 时测试结果
Table 1 Test results with TB = 1

序号	信任基准	理性选择	执行合作次数	申请合作次数	合作次数
AB	1	0.5	0.41	2.96	0.75
AC	1	0.5	0.45	3.30	0.75
AD	1	0.5	0.34	3.41	0.72
ABC	1	0.5	0.04	0.1	0.07
ABD	1	0.5	0.01	0.13	0.03
ACD	1	0.5	0.03	0.14	0.08

表 2 TB 为 1 时合作情况测试结果
Table 2 Test results cooperation with TB = 1

序号	信任基准	理性选择	执行申请成功率	执行合作成功率	综合评价	协作排序
AB	1	0.5	0.138 5	0.546 7	0.342 6	3
AC	1	0.5	0.136 4	0.600 0	0.368 2	2
AD	1	0.5	0.099 7	0.472 2	0.286 0	5
ABC	1	0.5	0.400 0	0.571 4	0.485 7	1
ABD	1	0.5	0.076 9	0.333 3	0.205 1	6
ACD	1	0.5	0.214 3	0.375 0	0.294 6	4

表 3 系统阶段信任基准更新
Table 3 Updating TB in stage of system

序号	协作排序	信任基准	理性选择
ABC	1	1.00	0.5
AC	2	0.75	0.6
AB	3	0.75	0.6
ACD	4	0.50	0.8
AD	5	0.50	0.8
ABD	6	0.25	1.0

表 3 所示,系统根据 agent 协作等级排序对信任基准赋值各异. 协作排序越高,信任基准越大,agent 间允许非合作或非执行的宽容度越大,体现 agent 间更注重长期合作的意愿;当信任基准较低时,agent 在 MAS 中公信度较低,因此为提高自身协作等级,理性选择将给予较高数值,当理性选择为 1 时,描述 agent 将全部完成做答应的合作任务,属于完全理性博弈者,不存在任何自私或盲目的思想. 当理性选择为 0.5 时,表示 agent 具有较高的协作等级,对自身各阶段博弈具有较强自主选择性与自私性,根据收益与机会成本较为灵活地选择是否执行合作任务. 根据表 3 系统阶段更新赋值,产生 MAS 新阶段合作情况,如表 4 所示. 表 5 为系统新阶段协作测试结果.

表 4 系统新阶段合作情况测试结果
Table 4 Test results of cooperation in the new stage

序号	信任基准	理性选择	执行合作次数	申请合作次数	合作次数
ABC	1	0.5	0.03	0.10	0.13
AC	0.75	0.6	0.49	3.42	0.79
AB	0.75	0.6	0.42	3.14	0.67
ACD	0.5	0.8	0.03	0.05	0.04
AD	0.5	0.8	1.27	3.84	1.50
ABD	0.25	1.0	0.06	0.13	0.06

表 5 系统新阶段协作测试结果
Table 5 Test results of collaboration in the new stage

序号	协作排序	执行申请成功率	执行合作成功率	综合评价	新协作排序
ABC	1	0.400 0	0.571 4	0.485 7	3
AC	2	0.143 3	0.620 3	0.381 8	5
AB	3	0.133 8	0.626 9	0.380 3	6
ACD	4	0.600 0	0.750 0	0.675 0	1
AD	5	0.330 7	0.846 7	0.588 7	2
ABD	6	0.461 5	1.000 0	0.461 5	4

分析表 4 和表 5 可得,由于系统阶段信任基准的更新,致使 MAS 中各 agent 对理性选择进行相应调整. 以 *ACD*,*AD*,*ABD* 三种合作方式为例,理性选择的提高有助于系统以较高的比率执行并完成合作任务,执行合作任务频率的上升,促使申请合作任务中应答合作的概率提高,从而提高合作次数与执行申请成功比率,使 agent 向理性博弈者模式转化,并提高协作等级排序席位. 而 *ABC*,*AC*,*AB* 三种组合方式具有较高的信任基准,在合作选择中具有较强的灵活性与争先性,使其在系统阶段更新后,排名有所变动,也反映出该实验可较好地描述 agent 在 MAS 运行中具有较高的自主性. 因此,通过变换系统中信任基准数值,可调整 MAS 中智能体的合作趋势与选择能力,使系统中各 agent 具有较强的合作性与竞争性,保证系统的高效稳定运行.

(下转第 147 页)