

交互式动态影响图及其精确求解算法

李 波¹, 曹浪财¹, 庄进发^{2,3}

(1. 厦门大学 信息科学与技术学院, 福建 厦门 361005; 2. 厦门东南融通系统工程有限公司, 福建 厦门 361005;
3. 解放军信息工程大学 通信与信息学院, 河南 郑州 450002)

摘要: 为了表示部分可观察马尔可夫环境下, 多 Agent 决策中各 Agent 之间的动态结构关系, 对影响图 (IDs) 在结构和时间上进行扩展, 形成一种能够对其他 Agent 建模的决策模型: 交互式动态影响图 (I-DIDs)。I-DIDs 是不确定环境下多 Agent 进行序贯决策的图模型。该模型的解是在对其 Agent 行为概率分布的预测下提供给该 Agent 的最优决策, 能更有效地解决多 Agent 的决策问题。但 I-DIDs 状态空间太大, Agents 候选模型空间随着时间片的增加成指数级增长, 使计算变得复杂。因此, 提出了一种基于行为等价的最小化模型集合的方法, 通过限制模型增长来缓解模型空间不断扩大的趋势, 以达到简化计算的目的。在模型实例上的仿真实验结果显示了该算法的有效性。

关键词: 多 Agent 决策; 交互式动态影响图; 行为等价; 最小模型更新集

中图分类号: TP18 **文献标识码:** A **文章编号:** 1009-3443(2011)02-0119-06

Interactive dynamic influence diagrams and exact solution algorithm

LI Bo¹, CAO Lang-cai¹, ZHUANG Jin-fa^{2,3}

(1. Department of Automation, Xiamen University, Xiamen 361005, China;

2. South East Combination System Engineering Co. Ltd, Xiamen 361005, China;

3. Department of Communication and Information, PLA Information Engineering University, Zhengzhou 450002, China)

Abstract: To represent the dynamic relationship between agents in multi-agent Markov decision process with partially observable settings shared by other agents, the interactive dynamic influence diagrams (I-DIDs) were presented by extending influence diagrams (IDs) over time and structure. I-DIDs are graphical models for sequential decision making in partially observable setting shared by other agents. It may be used to compute the policy of an agent given its belief as the agent acts and observes in the setting. Exact algorithms for solving I-DIDs demand the solutions of possible models of the agents and then update all models at every time step. The space of other models grows exponentially with the number of time steps, increasing the computational complexity. Thus an exact solution of I-DIDs based on minimal sets was presented by reducing the space of other agents possible models and updating the selected models, thereby the computational complexity was simplified. Finally, model instances were given. The experimental results show the validity of the algorithm.

Key words: multi-Agent decision; interactive dynamic influence diagrams (I-DIDs); behaviorally equivalent; minimal updating sets

收稿日期: 2010-08-10.

基金项目: 国家自然科学基金资助项目(60975052)

作者简介: 李 波(1981-), 女, 博士生; 研究方向: 人工智能、

多 Agent 决策问题建模; E-mail: xiaopi_libo@126.com.

影响图^[1] (IDs) 已成为不确定环境下处理单个

Agent 决策问题的一种重要工具。为了使影响图能

处理多 Agent 决策问题, 文献[2]对 IDs 进行扩展,

提出多 Agent 影响图(MAIDs)模型, MAIDs 可以表示多 Agent 之间非合作的结构关系, 用来处理存在纳什均衡的对策问题十分有效。然而, MAIDs 中, 每个 Agent 的决策是假定环境信息是已知的, 且不考虑不同 Agent 之间的差异, 因而 Agent 不能对环境中的其他 Agent 进行建模。文献[3]在 MAIDs 的基础上, 提出影响图网(NIDs)模型, NIDs 模型中考虑了其他 Agent 模型的不确定性, 由多个 MAIDs 组合成 NIDs 模型, 应对不同环境下的对策问题。无论是 MAIDs 还是 NIDs, 都被局限在静态环境中解决单个时间片的决策问题。

在对以上决策模型分析的基础上, 文献[4]将 MAIDs 在时间上进行扩展, 提出多 Agent 动态影响图(MADIDs), 用于表示动态环境中多 Agent 协作的结构关系。文献[5]提出交互式动态影响图(I-DIDs)模型, 用于描述交互式部分可观察马尔可夫决策过程(I-POMDPs)。该模型中, Agent 通过嵌套结构对其他 Agents 建模, 其模型解是在对其他 Agents 行为概率分布的预测下, 提供给该 Agent 的最优决策, 因此更能有效解决多 Agent 的决策问题。但是它的状态空间太大, 其他 Agents 候选模型的数量随着时间片的增加呈指数级增长, 且多 Agent 之间相互建模的嵌套层数使得模型状态空间更是随着层数的提高而急剧增加, 这使 I-DIDs 模型在计算上变得不可处理。因此本文提出基于行为等价的最小模型集合方法压缩模型状态空间, 以达到简化计算的目的, 提高 I-DIDs 模型求解能力。

1 交互式动态影响图数学描述

1.1 I-DIDs 的数学框架

I-DIDs 作为交互式部分可观察马尔可夫决策过程(I-POMDPs)的图形表示形式而被提出来, 因此成熟的 I-POMDPs 理论为 I-DIDs 提供了严密的数学框架。

在一个包含 2 个 Agent (Agent i 与 Agent j) 的有限嵌套的决策系统中, Agent i 的 I-POMDPs 模型可定义为一个七元组:

$$I\text{-POMDP}_{i,l} = \{IS_{i,l}, A, T_i, \Omega, O_i, R_i, \gamma\},$$

其中: $IS: S \times M_{j,l-1}$ 表示交互式状态空间, S 为 Agent 所处的物理环境状态, $M_{j,l-1}$ 为 Agent j 在 $l-1$ (l 为嵌套层数) 层的模型空间。对于每个模型 m_j M_j , $m_j = \theta = \langle b_j, \hat{\theta}_j \rangle$, 其中, $b_j \in B(IS_j)$, $\hat{\theta}_j = \langle A_j, \Omega, T_j, O_j, R_j, OC_j \rangle$ 分别为智能体 j 的信念状态和模

型框架。 OC_j 为智能体 j 的最优决策标准。 $A = A_i \times A_j$: 智能体的联合动作集合。 $T_i: S \times A \rightarrow S [0, 1]$, Agent i 的状态转移函数。 $\Omega: i$ 的观察值集合。 $O_i: A \times S \times \Omega [0, 1]$, 观察函数。 $R_i: S \times A \rightarrow \mathbb{R}$ 收益函数, 表示在状态 $s \in S$, 执行联合动作 $a_i \times a_j$ 获得的收益。 γ : 收益折扣因子, $\gamma \in (0, 1)$ 。

1.2 信度更新

在交互式部分可观察马尔可夫环境下, 信念状态是智能体根据观察及历史信息计算得到的一个当前状态对交互式状态空间(包括物理环境状态 S 及其他 Agent 候选模型 $m_j \in M_j$) 的一个概率分布, 分别记为 $b(s), b(m_j | s)$ 。满足 $0 < b(s), b(m_j | s) < 1$, 且 $\sum_{s \in S} b(s) = 1, \sum_{m_j \in M_j} b(m_j | s) = 1$ 。在 I-POMDPs 中, 每一步的信念状态都是 Agent 的一个主观概率, 而获得新的观察后, 得到信度更新公式^[6]:

$$b_i(is^t) = \beta_{is^{t-1}, m_j^{t-1} = \hat{\theta}_j} b_i(is^{t-1}) \times \prod_{d_j \in \mathcal{D}_j} Pr(a_j^{t-1} | \theta_j^{t-1}) O_i(o_i^t, s^t, a^{t-1}) \times \prod_{d_j \in \mathcal{D}_j} \pi(b_j^t, a_j^t, a_i^{t-1} \times a_i^{t-1}, b_j^{t-1}) O_j(s^t, o_j^t, a_i^{t-1} \times a_i^{t-1}) \times T_i(s^t, a_i^{t-1} \times a_j^{t-1}, s^{t-1}). \quad (1)$$

1.3 策略评价

I-DID 的目标是寻求一个可以最大化系统在 $0 \sim T$ 时间内期望收益值的策略 p , 即

$$V^T(b_{i,l}, \hat{\theta}_i) = \max_p \{ \sum_{is \in IS} R(is, a_i) b_{i,l}(is) + \gamma \text{(Expected Value in the Future)} \} = \max_p \{ \sum_{is \in IS} R(is, a_i) b_{i,l}(is) + \gamma \sum_{o_i} Pr(o_i | a_i, b_{i,l}) V^{n-1} \}. \quad (2)$$

并产生相应的最优决策:

$$OPT(b_{i,l}, \hat{\theta}_i) = \arg(V^T(b_{i,l}, \hat{\theta}_i)).$$

2 I-DIDs 的图形表示

2.1 交互式影响图(I-IDs)

影响图(IDs)是由节点和有向弧组成的无环路有向图。其中, 机会节点、决策节点和效用节点代表所研究问题的主要变量, 有向弧表示变量之间的各种相互关系。单个时间片的 I-DIDs 称为交互式影响

图(I-Ds)^[7], I-Ds 在 IDs 的基础上加入一种新类型的节点: 模型节点, 用来对其他 Agent 建模。一个包含 2 个 Agent 的决策系统, 对其中一个 Agent_{*i*} 建立第 *l* 层的 I-Ds 模型, 如图 1(a) 所示, 代表其他 Agent 决策行为的变量 A_j 被转化为机会节点, 其取值由相应的概率分布所控制: 当 $a_j = \text{OPT}$ 时, $Pr(a_j = A_j^i) = 1/|\text{OPT}|$, 否则为 0。

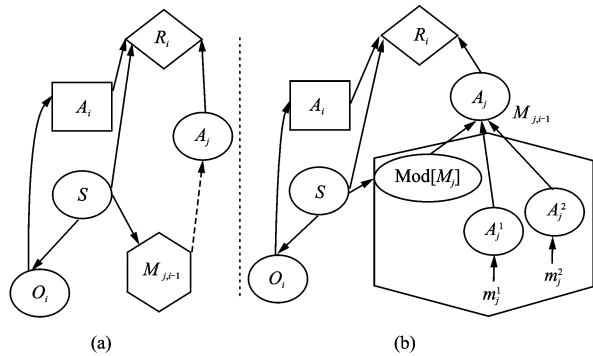


图 1 交互式影响图模型

Fig. 1 I-ID model

益。I-DID_s 模型中, 引入了新的结构关系: 模型更新连接(连接模型节点 $M_{j,l-1}^{t-1}$ 和 $M_{j,l-1}^t$ 的点划线)。

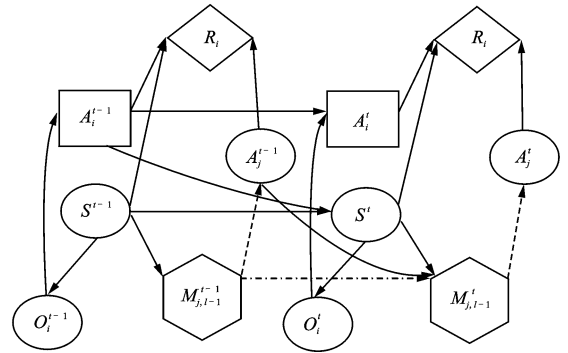


图 2 2 个时间片上的 I-DID 模型

Fig. 2 DID model at two-time slice

模型更新连接表达了 2 方面的内容: 第一, 它记录了候选模型空间不断增长的过程。如图 3 所示, 假设 $t-1$ 时刻模型节点 $M_{j,l-1}^{t-1}$ 包含 2 个模型 ($m_{j,l-1}^{t-1,1}$, $m_{j,l-1}^{t-1,2}$), 每个模型对应一个最优动作, 即 $|\text{OPT}(m_{j,l-1}^{t-1})| = 1, j$ 的观察集合 $\Omega_j = \{o_j^1, o_j^2\}$, 则下一时间片的模型数量 $|M_{j,l-1}^t| = |M_{j,l-1}^{t-1} \times \text{OPT}(m_{j,l-1}^{t-1}) \times \Omega_j| = 4$; 第二, 模型更新连接记录了信度更新过程。节点 $\text{Mod}[M_{j,l-1}^t]$ 的 CPD 取值是关于 $\pi(b_j^t, o_j^t, a_i^{t-1} \times a_j^{t-1})$ 的函数, 如果 j 在信度状态为 $b_j^{t-1} = m_{j,l-1}^{t-1}$ 时, 执行动作 $\text{OPT}(m_{j,l-1}^{t-1})$, 进入 t 时间片, 获得观察 o_j^t , 更新信度状态达 $b_j^{t-1} = m_{j,l-1}^t$, 此时函数 $\pi(b_j^t, o_j^t, a_i^{t-1} \times a_j^{t-1})$ 的取值为 1, 否则为 0。

一个自底向上的, 不同层次的交互式状态空间 $IS: S \times M_{j,l-1}$ 的迭代结构, 如式 (3) 所示:

$$\left. \begin{aligned} IS_{i,0} &= S, \\ \Theta_{i,0} &= \{ b_{j,0}, \hat{\theta} : b_{j,0} \in B(IS_{j,0}) \}, \\ IS_{i,l} &= S \times M_{j,l-1}, \\ \Theta_{i,l} &= \{ b_{j,l}, \hat{\theta} : b_{j,l} \in B(IS_{j,l}) \}. \end{aligned} \right\} \quad (3)$$

$\Theta_{i,0}$ 是 0 层上 Agent_{*j*} 模型的集合, 每个模型是一个 ID_s 模型。 $\Theta_{i,l}$ 是 *l* 层上 *j* 模型的集合。当 $l > 0$ 时, 每个模型又是 I-Ds 模型。图 1(a) 中连接模型节点 $M_{j,l-1}$ 和机会节点 A_j 的虚线称为策略连接^[8], 反映了模型 $m_{j,l-1}$ 与其最优决策 $\text{OPT}(m_{j,l-1})$ 的一一映射关系。图 1(b) 用只包含机会节点的贝叶斯网替代模型节点和策略连接, 把一个 I-D 模型转化成 ID 模型(这里假设 $|M_{j,l-1}| = 2$), 以便根据标准的期望效用最大方法求解 I-D 模型。

2.2 交互式动态影响图(I-DIDs)

将 I-Ds 在时间上进行扩展得到另一种决策模型: I-DIDs。图 2 给出了 2 个时间片上的包含 2 个 Agent 的 I-DIDs 模型。节点 A_i 和 A_j 的状态是 *i* 和 *j* 动作集合, 节点 O_i 的状态是 *i* 的观察集合。机会节点 S^t, O^t 上的条件概率分布 CPD(conditional probability distribution) $Pr(S^t | S^{t-1}, a_i^{t-1}, a_j^{t-1})$ 和 $Pr(O_i^t | S^t, a_i, a_j^{t-1})$ 分别对应 I-POMDPs 模型中的概率转移函数和观察函数。效用节点 R_i 代表 *i* 的收

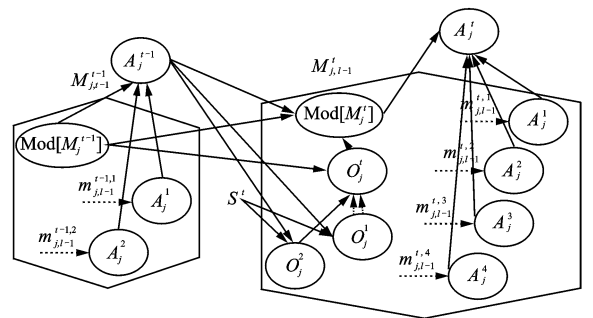


图 3 模型更新过程

Fig. 3 Process of the model updated

3 I-DIDs 的精确求解算法

I-DIDs 模型的解是在对其他 Agent 行为的概率分布的预测下提供给该 Agent 的最优决策, 因此首先要求解其他 Agent 底层模型, 这也决定了 I-DIDs 模型求解过程是自底向上的反复迭代的过程。得到底层模型解, 即获得 *j* 所有模型的策略树, 亦可获得节点 A_j 和 $\text{Mod}[M_j]$ 的条件概率表, 即可实现

推理。具体算法如下:

EXACT 算法

输入: I-DIDs($l > 1$) 或 DIDs($l = 0$) 模型, 时间片数: T , 初始模型($m_{j,0} \quad M_{j,0}$)

输出: 表示 Agent i 在 $1 \sim T$ 时间上决策策略图

1): For($t = 1; t < T; t++$),

2): If ($l > 0$), 首先处理其他 Agent 底层模型更新问题;

3): For ($le = 0; i < l; le++$);

4): 根据 t 时间片的最优决策 $a_j = \text{OPT}(m_j^t)$ 和观察值 o_j , 更新模型信度 $b_j^{t+1} = SE(b_j^t, a_j, o_j)$, 并获得以 b_j^{t+1} 作为初始信度的模型 m_j^{t+1} ;

5): 返回 a_i, o_i ;

6): 各个时间片上 a_i, o_i 的累计, 形成 $|M_{j,0}^l|$ 棵策略树;

7): 所有策略树简单累积合并就形成了代表 Agent j 决策的策略图;

8): 根据策略图建立 ($le+1$) 层上的机会节点 $\text{Mod}[M_j^{le+1}]$ 和 A_j^{le+1} 的条件概率表;

9): 添加 ($le+1$) 层上的节点及相应的条件概率表, 编译图形文件;

10): 输出 i 的策略树。

4 基于最小模型集的精确求解算法

针对一般的情形, I-DIDs 求解过程的复杂程度通常与其状态空间、行动空间及过程需要经历的时间长度有关。I-DIDs 模型的状态空间, 不仅包括 i 所处的物理环境状态, 还包括环境中其他 Agent 的候选模型空间, 且其他 Agent 模型随时间的增加成指数级增长, i 不仅要求解各时间片上的每个候选模型, 还要记录这些模型的演变过程。所以, 求解 I-DIDs 模型面临着“维数灾难”和“历史灾难”双重问题。本节提出基于行为等价的方法对候选模型聚类, 同时只更新将产生不同行为的候选模型, 以压缩模型空间, 减轻计算复杂度。

4.1 最小模型集合

直观上, 对于 i 来说, 只注重 j 最后执行的行为, 而 j 的观察及其对状态的信度分布等其他因素均不能直接影响 i 的决策。因此, i 不必区分处于同一行为等价类中 j 的 2 个不同模型。这样所得 I-DIDs 模型的解仍然保持着问题的最优性。下面介绍行为等价的定义。

定义 1 行为等价: 2 个模型 $m_{j,l-1}$ 和 $m_{j,l-1}$, 满足 $\text{OPT}(m_{j,l-1}) = \text{OPT}(m_{j,l-1})$, 称为行为等价模型。

在行为等价基础上给出最小模型集合的定义:

定义 2 最小模型集合: 模型集合 $\hat{M}_{j,l-1}$ 为 $M_{j,l-1}$ 的一个最大子集, 且不存在其他模型与 $m_{j,l-1}$ $M_{j,l-1}$ 同处于一个行为等价类中。此时 $\hat{M}_{j,l-1}$ 为 $M_{j,l-1}$ 的一个最小模型集合。

由于在各个行为等价类中选取的具有代表性的模型不同, 所以最小模型集合 $\hat{M}_{j,l-1}$ 不唯一, 但其模型基数是固定的。集合 $M_{j,l-1}$ 用最少数量的模型包括了 $M_{j,l-1}$ 中 j 的所有行为, 所以经最小模型集合简化后的候选模型空间, 仍保留了 i 决策时所需要考虑的全部情形。但是, j 的模型被删除后, 势必会影响 i 对 j 模型信度的概率分布情况, 为了弥补这一损失, 此时调整 i 对 $m_{j,l-1} \hat{M}_{j,l-1}$ 的条件概率分布, 使其等于所有同处于该行为等价类模型 $m_{j,l-1} M_{j,l-1}$ 的概率之和, 即

$$b_i(m_{j,l-1}^q | s) = \sum_{m_{j,l-1} \in M_{j,l-1}^q} b_i(m_{j,l-1} | s). \quad (4)$$

$M_{j,l-1}$ 是与 $\hat{M}_{j,l-1}$ 处在同一个行为等价类中的模型集合。

4.2 算法实现

经最小模型集简化的 I-DIDs 求解算法, 关键是获得基于行为等价的最小模型集合 $M_{j,l-1}$, 一个直接的获得 $\hat{M}_{j,l-1}$ 的方法是自底向上合并策略树。为区别策略树, 把合并后的图形称为策略图。

图 4 给出了最后 3 个时间片上 2 个策略树合并的过程。其中, * 代表通配符。需要指出的是, 图 4(b) 中, 合并 $t = T - 1$ 时间片上 2 个粗线圆的节点, 因为在 $(T - 1)$ 和 T 时间片上, 以此节点为根节点的 2 棵决策树相同。合并后所得到的策略图, 节点数量明显减少。图 4(c) 策略图中, 节点表示 Agent 的最优动作, 与模型有一一对应关系, 所以模型数量亦得到相应的减少。

本文所提算法(Exact MS), 在 Exact 算法的基础上, 得到底层模型的策略树后, 依行为等价原理自底向上合并策略树形成策略图, 用来指导模型是否进行更新, 并依式(4)转移被舍弃模型的信度。

4.3 算法复杂度分析

I-DIDs 模型中, t 时间片上, Agent j 的候选模型数量为 $|M_j| = |M_j^0| (|A_j \cup W_j|)^t$, 其中, $|M_j^0|$ 为初始模型数量。一个包含 $(N + 1)$ 个 Agents 的 I-

DIDs 模型, 若每一层上每个 Agent 的模型个数为 $|M|$, 则得到 l 层上问题的解需处理模型个数是 $O((N|M|)^l)$ 数量级的。基于行为等价原理合并策略树形成策略图, 在最坏情况下(不存在可供合并的叶子节点)的复杂度为 $O((|\Omega|^{T-1})^{|\hat{M}_j|})$, 主要由合并策略树时的比较次数决定。在 l 层上, 需处理的模型个数为 $O((N|\hat{M}_j^*|)^l)$, $|\hat{M}_j^*|$ 是最小模型集中包含最多模型的数量, 而 $|\hat{M}_j^*| \ll |M_j|$ 。所以, 经最小模型集合简化的候选模型空间, 不仅缓解了模型呈指数级增长的趋势, 同时, 也降低了求解更高层次上的 I-DIDs 模型的复杂度。

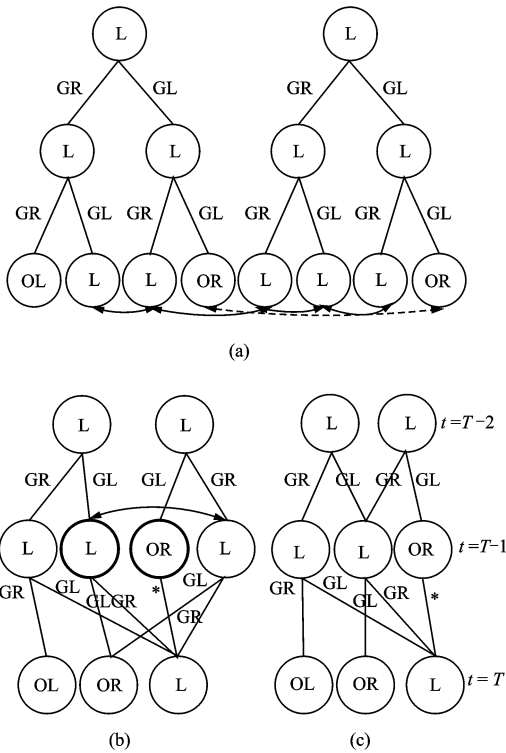


图 4 依行为等价原理合并策略树

Fig. 4 Merge policy trees based on the theory of behavioral equivalence

5 实验与分析

以 Piotr 等介绍的多 Agent 老虎问题^[6]为例进行实验。该问题包含左门和右门 2 扇门。状态集 $S = \{TL, TR\}$, 指示老虎在左门后或右门后; 动作集 $A_i = A_j = \{OL, OR, L\}$ 分别指示开左门、开右门和倾听, 即收集信息, 提供关于老虎位置和其他 Agent 动作的观察信息。这些动作可以任意组合成团队的联合动作。联合动作 $\langle L, L \rangle$ 保持环境状态不变。为了使其成为无限时间问题, 如果任一 Agent 打开一扇门, 环境随机地、一致地复原为一个新状态。观察

集 $\Omega_i = \{GL, GR\}$ 指示老虎的可能位置, 观察集 $\Omega = \{CL, CR, S\}$ 指示其他 Agent 的可能行为, 2 个观察集中元素可以任意组合成联合观察集合。对该问题建立 I-DIDs 模型, 如图 2 所示。为了能够用传统的 DIDs 求解方法(贝叶斯推理)求解 I-DIDs 模型, 该问题 I-DIDs 模型转化成 DIDs 模型如图 5 所示。

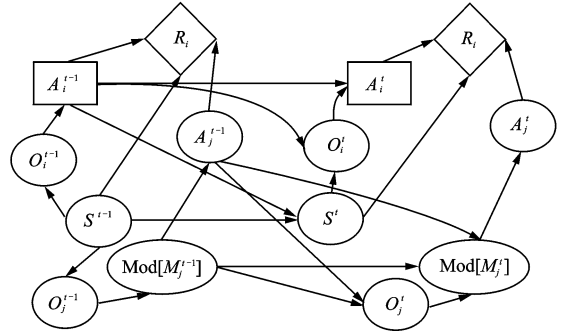


图 5 I-DID 转化成 DID 模型

Fig. 5 Transform I-DID model into DID model

实验中取其他 Agent 初始模型个数 $|M_{i,0}^1| = 50$, 取时间步分别为 3, 6, 8, 10 进行实验。以 Hugin 为基础, 在 Java 环境下运行程序(系统配置: WinXP, dual processor 1.73 GHz, 2 GB memory), 对两种算法进行比较, 结果如表 1 所示。从表中运行时间数据可以看出, 随着时间片的增加, 其他 Agent 模型数量越多, Exact MS 所体现的有效性就越明显。

表 1 Exact 与 Exact MS 算法运行时间比较

Tab. 1 Run time comparison between Exact and Exact MS

l	Algorithm	T			
		3	6	8	10
1	Exact	3.49	7.03	50.96	270.6
	Exact MS	3.01	5.58	12.63	20.36
2	Exact	28.82	68.74	205.5	—
	Exact MS	19.68	40.52	96.36	163.4

6 结 语

利用 I-DIDs 建立问题模型时, 可以清楚地了解 Agent 之间的相互关系、变量之间的函数关系、决策中的时序关系、信息流向和整个问题的结构。I-DIDs 模型作为 I-POMDPs 的图形表达形式, 它结合了贝叶斯网中变量独立关系, 使推理计算得到简化。但随着时间片的增加, I-DIDs 状态空间太大, 使模型求解备受计算量的困扰。因此提出一种基于行为等价的最小模型集合方法, 来缓解其他 Agent 底层模型不断增长的趋势, 减少计算复杂度。此外, 多 Agent 之间相互建模的嵌套层数, 也使得倍受计算量困扰

的 I-DIDs 求解更加束手无措,因此下一步的工作是从嵌套层数入手,寻求提高模型求解能力的方法。

参考文献:

- [1] HOWARD R A, MATHESON J E. Influence diagrams[J]. Readings on the Principle and Application of Decision Analysis, 1984, 11(2): 719-762.
- [2] KOLLER D, MILCH B. Multi-agent influence diagrams for representation and solving games [J]. Games and Economic Behavior, 2003, 45(1): 181-121.
- [3] GAL Y, PFEFFER A. A language for modeling agent's decision-making processes in games[C]. Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS 03, 2003.
- [4] 姚宏亮,王浩,张佑生,等.多 Agent 动态影响图及其概率分布的近似方法[J].模式识别与人工智能,2007, 4(20): 525-532.
YAO Hong-liang, WANG Hao, ZHANG You-sheng, et al. Multi-agent dynamic influence diagrams and its approximation of probability distribution[J]. Pattern Recognition and Artificial Intelligence, 2007, 4(20): 525-532. (in Chinese).
- [5] POLICH K, GMYTR S P. Interactive dynamic influence diagrams [C]. International Conference on Autonomous Agents Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS 07, 2007.
- [6] GMYTRASIEWICZ P, DOSHI P. A framework for sequential planning in multi-agent settings[J]. Journal of Artificial Intelligence Research, 2005(24): 49-79.
- [7] DOSHI P, ZENG Yi-feng, CHEN Qiong-yu. Graphical models for online solutions to interactive POMDPs [C]. International Conference on Autonomous Agents Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems. New York: ACM, 2007.
- [8] DOSHI P, ZENG Yi-feng, CHENG Qiong-yu. Graphical models for interactive POMDPs: representation and solutions [J]. Autonomous Agents and Multi-Agent Systems, 2009, 18(3): 376-416.