

# 多 Agent 动态影响图及其一种近似推理算法研究

姚宏亮 王 浩 张佑生 汪荣贵

(合肥工业大学计算机科学与技术系 合肥 230009)

**摘 要** 针对多 Agent 影响图不能建模动态环境和多 Agent 马尔可夫决策过程难以表示 Agents 之间结构关系的问题,提出一种新决策模型——多 Agent 动态影响图(MADIDs).为了能有效地对 MADIDs 进行推理,提出一种扩展的 BK(EBK)近似推理算法,其扩展体现在三个方面:在 BK 算法中加入效用结点的边际化操作,加入分割团来减小 BK 算法的推理误差,使用 MADIDs 分层分解所生成的联合树来降低推理的复杂性.在模型实例上的实验结果显示了 MADIDs 模型和 EBK 算法的有效性.

**关键词** 多 Agent 影响图;多 Agent 动态影响图;联合树;BK 算法

中图法分类号 TP181

## Research on Multi-Agent Dynamic Influence Diagrams and Its Approximate Inference Algorithm

YAO Hong-Liang WANG Hao ZHANG You-Sheng WANG Rong-Gui

(Department of Computer Science and Technology, Hefei University of Technology, Hefei 230009)

**Abstract** As MAIDs can not model dynamic environment and Multi-Agent MDPs are difficult to represent structural relations among Agents, Multi-Agent dynamic influences diagrams (MADIDs) are presented for modeling structural relations of Multi-Agent system in dynamic environment. For efficiently inferring in MADIDs, an extensional BK (EBK) algorithm is proposed, and the extensions are in three aspects: The marginalizing operation of utility nodes is added in BK algorithm; the separators are added for decreasing the error of inference; the junction tree that obtained by hierarchical decomposition of MADIDs is used for improving the efficiency of inference. Given model instances, the experiment results show the validity of MADIDs and EBK algorithm.

**Keywords** multi-Agent influence diagrams; multi-Agent dynamic influence diagrams; junction tree; BK algorithm

## 1 引 言

影响图(Influence Diagrams, IDs)是对单个 Agent 决策问题建模的一种工具<sup>[1-2]</sup>.为了使影响图能用于多 Agent 决策系统,Koller 和 Milch 对影

响图进行扩展提出多 Agent 影响图(Multi-Agent Influence Diagrams, MAIDs)<sup>[3]</sup>,它能建模多 Agent 之间非合作的结构依赖关系,用于处理 Agents 的对策(game)问题.但是,MAIDs 不能建模动态的决策环境;另外,由于 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 Agents 的结构

差异,因而 MAIDs 中的 Agent 不能对环境其它 Agent 建模<sup>[4]</sup>.

多 Agent 马尔可夫决策过程(MMDPs)可以处理动态环境中多 Agent 决策问题<sup>[5-6]</sup>. 由于每个 Agent 的决策需要考虑与 Agent 个数成指数关系的系统联合状态和联合行为<sup>[7]</sup>,这使得模型的联合概率分布在计算上难以处理;Carlos Guestrin 将 MMDPs 表示成多个 MDP 的因式来简化模型概率分析的计算<sup>[8]</sup>,但仍然难以处理较复杂的动态决策问题. Barto 指出<sup>[9]</sup>,MMDPs 的推理、Agent 之间结构关系的表示和协调等方面还面临重大挑战.

针对上述决策模型的不足,通过将动态影响图(DID)<sup>[10]</sup>推广到动态多 Agent 系统,提出一种新的决策模型——多 Agent 动态影响图(Mult-Agent Dynamic Influence Diagrams, MADIDs). MADIDs 用分层结构表示 Agents 之间的结构关系:外层是一个以 Agent 为结点的全局影响图,而每个 Agent 又对应一个影响图. 为了提高模型表示的计算效率,将系统的联合状态和行为空间分割成局部的小空间,并将这种结构关系抽象成一种局部概率因式. 由于 MADIDs 能对复杂动态决策问题中的结构关系建模,使得一个 Agent 可以通过推理对环境的变化进行预测,而其它 Agents 是环境的一部分,因而 Agents 能对环境中的其它 Agents 进行建模,从而避免 Agents 之间交互的通信问题.

概率图模型的推理在计算上是一个 NP 问题<sup>[11]</sup>,通常使用近似推理方法. BK 算法是动态贝叶斯网络的一种主要近似推理方法<sup>[12]</sup>,是根据变量间的弱相关性将系统分解成若干个子系统(称为团(cluster))以降低计算复杂性,但 BK 算法仅只适用于中等规模的系统,且会引入较大误差. 另外, BK 算法中的团在推理计算上仍是一个高维问题<sup>[13]</sup>.

基于 BK 算法思想,提出一种扩展的 BK(EBK)近似推理算法用于 MADIDs 的推理,其扩展体现在:(1)由于动态贝叶斯网络中没有效用结点,需要在 BK 算法中加入效用结点的边际化操作,使它能用于 MADIDs 的推理;(2)对 MADIDs 概率分布进行分层分解以降低推理的计算复杂性,分层分解是先将 MADIDs 的全局影响图分解成联合树(Junction Tree, JT),然后根据 MADIDs 的内层影响图结构和 Agents 之间的策略依赖性,对联合树中的团进行增量分解生成用于推理的联合树;(3)通过在 BK 算法中加入分割团来减小 BK 算法的推理误差.

本文第 2 节、第 3 节分别给出 MAIDs 的定义和 MADIDs 的形式化描述;第 4 节给出 MADIDs 的分层结构;第 5 节给出 MADIDs 的分层分解方法;第 6 节给出利用 MADIDs 的 EBK 算法描述;最后在两个模型实例上进行实验和算法比较.

## 2 多 Agent 影响图

多 Agent 影响图(MAIDs)用于表示多 Agent 之间静态的非合作关系,通过寻找最优策略模式  $\sigma$  来得到多 Agent 系统的纳什(Nash)平衡.

**定义 1.** 策略模式(strategy profile). 决策变量  $D$  的决策规则是一个函数,该函数将  $D$  的父结点集  $Pa(D)$  中的每个实例映射为  $D$  值域上的一个概率分布. 策略是指为某个 Agent  $A \in Ag$  ( $Ag$  是一个 Agents 的集合)的决策变量  $D_A$  分配决策规则. 策略模式  $\sigma$  是指为决策集中的每个决策变量分配决策规则的集合.

**定义 2.** MAIDs 的联合概率分布. 一个多 Agent 影响图  $M$  的联合分布和它的策略模式  $\sigma$  相关,因而该 MAID 的联合概率分布可表示为  $P_{M[\sigma]}$ .

设  $A$  是 MAID  $M$  中的一个 Agent,  $U_A = \{u_1, u_2, \dots, u_m\}$  为  $A$  的效用函数,  $dom(U_A)$  是  $U_A$  的值域,则对于给定策略模式  $\sigma$ ,  $A$  的期望效用为

$$EU_A(\sigma) = \sum_{(u_1, u_2, \dots, u_m) \in dom(U_A)} P_{M[\sigma]}(u_1, u_2, \dots, u_m) \sum_{i=1}^m u_i \quad (1)$$

**定义 3.** 若对所有的 Agent  $A \in Ag$ ,  $\sigma_{D_A}$  都是最优的,则  $\sigma$  是 MAID 的一个 Nash 平衡.

由定义 2 可知,MAIDs 没有引入时间参数,不能对动态环境进行建模,且 MAIDs 中的 Agents 不具有对环境变化和其它 Agents 行为进行推理的能力,因而 MAIDs 中的 Agents 不能建模其它 Agents, Agents 之间的交互要通过通信来实现.

## 3 多 Agent 动态影响图模型

多 Agent 动态影响图(MADIDs)定义为一个三元组  $(M, B_t, B_{ts})$ , 其中,  $M$  为 Agent 的个数,  $B_t$  为系统在  $t$  时的影响图,  $B_{ts} = (E_t, V_{ts})$  为转移模型,  $E_t$  为从  $B_{t-1}$  到  $B_t$  的有向边,  $V_{ts}$  为  $E_t$  所连接的结点集. MADIDs 中有 4 种结点:决策结点  $D$  表示 Agent 的行为选择,体现 Agent 的能力;随机结点  $X$  和观察结点  $Y$  是与决策相关的随机变量,分别表示 Agents

的信用和愿望;效用结点  $U$  用于揭示 Agents 的意图,每个效用结点有一个效用函数.

设  $t$  时, MADIDs 的状态变量集为  $X_t = \{X_t^1, \dots, X_t^M, X_t^{M+1}, \dots, X_t^N\}$ , 其对应的观察变量集为  $Y_t = \{Y_t^1, \dots, Y_t^M, Y_t^{M+1}, \dots, Y_t^N\}$ ; 决策变量集为  $D_t = \{D_t^1, \dots, D_t^M\}$ ; 效用变量集为  $U_t = \{U_t^1, \dots, U_t^M\}$ . 其中,  $X_t^i, Y_t^i, D_t^i, U_t^i (i \leq M)$  分别表示 Agent  $i$  的状态变量集、观察变量集、决策变量和效用变量; 每个 Agent 有一个决策变量和一个效用变量; 上标从  $M+1$  到  $N$  的变量是环境变量, 用于描述系统的环境信息.

MADIDs 由 3 个子模型组成: 结构策略模型、概率模型和效用模型.

(1) 结构策略模型. 设  $\Delta_t = (\delta_t^1, \dots, \delta_t^M)$  是  $t$  时所有 Agents 的决策规则集, 则结构策略模型为  $t$  时的每个决策变量  $D_t^i$  确定一个局部决策规则  $\delta_t^i$ ,  $\delta_t^i$  是一个从  $D_t^i$  的父结点集  $Pa(D_t^i)$  到  $D_t^i$  行为集的映射:

$$\delta_t^i: \Omega_{Pa(D_t^i)} \rightarrow \Omega_{D_t^i} \quad (2)$$

其中,  $\Omega_{Pa(D_t^i)}$  为  $D_t^i$  父结点集的取值空间,  $\Omega_{D_t^i}$  为决策变量  $D_t^i$  的行为空间. 策略模式  $\sigma_t$  为  $t$  时的决策变量集中的每个决策变量分配一个决策规则.

(2) 概率模型. 给定决策规则  $\delta_t^i$ , 下面的式(3)是将决策变量  $D_t^i$  转换成随机变量的一种方法, 并表示了  $D_t^i$  的条件概率:

$$P_{\delta_t^i}(D_t^i | Pa(D_t^i)) = \begin{cases} 1, & \delta_t^i(Pa(D_t^i)) = d_j^i \\ 0, & \text{否则} \end{cases} \quad (3)$$

其中,  $d_j^i$  是 Agent  $i$  行为集中的一个行为.

由于式(3)给出了决策变量的概率分布, 则决策结点成为 MADIDs 的概率模型的组成部分. 因而给定  $\sigma_t$ , 随机变量集、观察变量集和决策变量集在  $t$  时的联合概率分布为

$$P_{\sigma_t}(X_t, Y_t, D_t) = \prod_{X_t^k \in X_t} P(X_t^k | Y_{0:t}, D_{0:t}) \cdot \prod_{Y_t^k \in Y_t} P(Y_t^k | Pa(Y_t^k)) \prod_{D_t^i \in D_t} P_{\delta_t^i}(D_t^i | Pa(D_t^i)) \quad (4)$$

其中,  $D_t^i$  是第  $i$  个 Agent 的决策变量, 下标  $0:t$  表示从时间 0 到时间  $t$ . 式(4)所表示的概率模型是动态贝叶斯网络的一种扩展形式, 即将策略模式和决策结点引入到动态贝叶斯网络中.

设 MADIDs 概率模型满足一阶马尔可夫假设, 有  $P(X_t | X_{0:t-1}, D_{0:t-1}) = P(X_t | X_{t-1}, D_{t-1})$ . 给定状态变量集的转移概率  $P(X_t | X_{t-1}, D_{t-1})$  和观察变量集的概率  $P(Y_t | X_t)$ , 状态变量集的概率分布通

过转移模型进行传播, 则  $t$  时状态变量的先验概率分布为

$$P(X_t | Y_{0:t-1}, D_{0:t-1}) = \sum_{X_{t-1}} P(X_{t-1} | Y_{0:t-2}, D_{0:t-2}) P(X_t | X_{t-1}, D_{t-1}) \quad (5)$$

给定观察变量集  $Y_t$  和决策变量集  $D_t$  的条件先验概率, 则状态变量集后验概率分布为

$$P(X_t | Y_{0:t}, D_{0:t}) = \frac{P(X_t | Y_{0:t-1}, D_{0:t-1}) P(Y_t | X_t)}{\sum_{X_t} P(X_t | Y_{0:t-1}, D_{0:t-1}) P(Y_t | X_t)} \quad (6)$$

式(5)和式(6)是一个迭代过程.

(3) 效用模型. 效用结点有一个关于决策变量和状态变量的函数. 在  $t$  时, 效用结点  $U_t^i$  的局部效用函数为  $U_t^i(Pa(U_t^i))$ , 其中  $Pa(U_t^i)$  为  $U_t^i$  的父结点集. MADIDs 的联合效用函数是各个局部效用函数的和:

$$U_t(X_t, D_t) = \sum_{i=1}^m U_t^i(Pa(U_t^i)) \quad (7)$$

一个效用结点  $U_t^i$  的效用函数可表示如下:

$$U_t^i(Pa(U_t^i)) = f_t^i(X_t^{i,1}, \dots, X_t^{i,N}) = w_{t,i,1} X_t^{i,1} + \dots + w_{t,i,N} X_t^{i,N} \quad (8)$$

其中,  $Pa(U_t^i) = \{X_t^{i,1}, \dots, X_t^{i,N}\}$  是  $U_t^i$  父结点集; 权重  $w_{t,i,j}$  表示  $X_t^{i,j}$  在  $U_t^i$  的效用函数中所起作用的大小.

综合上面 MADIDs 的三个子模型, 给定策略模式  $\sigma_t$ ,  $t$  时 MADIDs 的联合期望效用为

$$EU(\sigma_t) = \sum_{X_t^k} P_{\sigma_t}(X_t, Y_t, D_t) \sum_{U_t^i \in U_t} U_{\sigma_t}(Pa(U_t^i)) \quad (9)$$

**定理 1.** 给定一个 MADID 的初始策略模式  $\sigma_0$  和初始效用函数  $U_0(X_0, D_0)$ , 则  $t=1$  时的  $\sigma_1$  和  $U_1(X_1, D_1)$  的结构在时间上具有不变性.

证明. 由式(2)和式(7)可知,  $t=1$  时的  $\sigma_1$  和  $U_1(X_1, D_1)$  完全依赖于  $B_t$  和  $B_{ts}$  的结构; 又因为 MADID 的  $B_t$  和  $B_{ts}$  的结构具有时间不变性, 因而  $t=1$  时的  $\sigma_1$  和效用模型  $U_1(X_1, D_1)$  的结构在时间上也具有不变性. 而  $t=0$  时,  $B_t$  和  $B_{ts}$  的结构不能表示和前一时间点的关系, 且在下一时间点时  $B_t$  和  $B_{ts}$  的结构会产生变化, 因而  $\sigma_0$  和  $U_0(X_0, D_0)$  的结构需要事先给定. 证毕.

## 4 MADIDs 的分层结构

足球机器人是复杂动态环境中多 Agent 决策

问题的一个研究平台<sup>[14]</sup>, 在该平台上我们通过两个球员配合射门的一个实例来描述 MADIDs 的分层

结构, 如图 1 所示.

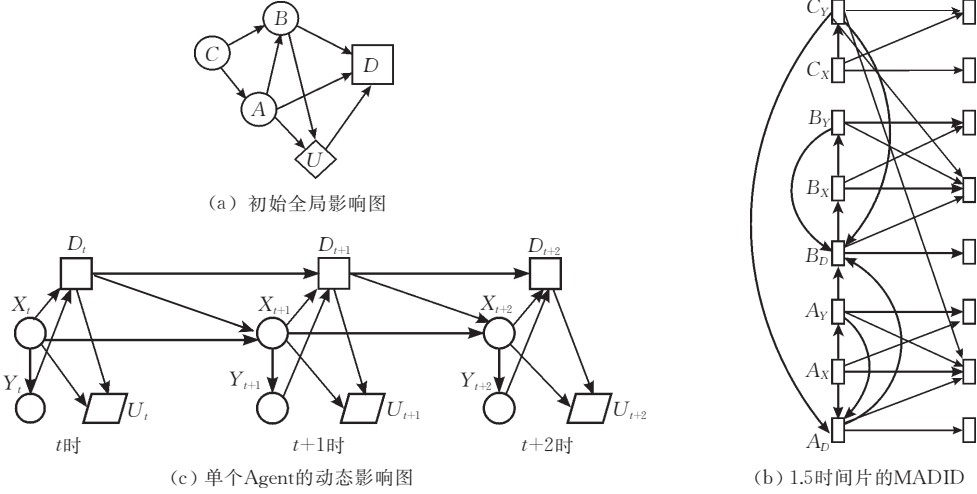


图 1

#### 4.1 MADIDs 的全局结构

MADID 的初始全局影响图的结构如图 1(a) 所示, 其中,  $A, B$  和  $C$  结点分别对应一个 Agent,  $D$  为  $A$  和  $B$  的联合决策结点,  $U$  为  $A$  和  $B$  的联合效用结点; 其中  $A$  和  $B$  为配合射门 Agent,  $C$  为守门 Agent, 当前球被  $A$  控制. 每个 Agent 都有多个变量组成,  $A$  有状态、观察和决策变量, 分别用  $A_X, A_Y, A_D$  表示;  $B$  有状态、观察和决策变量, 分别用  $B_X, B_Y, B_D$  表示; 由于设定  $C$  采取固定策略, 则  $C$  有状态变量和观察变量, 分别用  $C_X, C_Y$  表示.

图 1(a) 中  $A, B, C$  之间的边表示决策依赖关系, 如  $A$  和  $C$  的状态直接影响到  $B$  的决策, 因而  $A$  和  $C$  都有指  $B$  的边; 另外, 图 1(a) 中的边不是表示单个变量之间的关系, 而是多个变量之间的关系, 如图 1(b) 中左侧结点之间的边.

MADID 在 1.5 个时间片上的扩展结构如图 1(b) 所示, 其中左侧和右侧结点之间的边组成 MADID 的转移网络, 左侧结点之间的边对应图 1(a) 中去掉  $D$  和  $U$  结点后的内部结构; 右侧结点为左侧结点在时间上进行的一步扩展 (如左侧为  $t$  时 MADID, 则右侧为  $t+1$  时 MADID), 其中右侧结点之间的结构和左侧相同, 为了清晰地表示 MADID, 没有给出右侧结点之间的边, 只给出了 MADID 在下一时间片中的接口结点, 因而图 1(b) 表示的是 1.5 个时间片的 MADID.

效用结点与其父结点之间不是概率依赖关系, 而是一种函数关系, 且根据影响图的定义<sup>[15]</sup> 效用结点没有孩子结点, 因而效用结点容易从影响图结构

中分离; 这种分解不会破坏 MADID 内部的结构依赖关系, 并可使网络的结构表示变得简单而易于处理; 另外, 且由定理 1 可知 MADIDs 的效用结构具有时间不变性. 因而图 1(b) 没有给出 MADID 的效用结点.

#### 4.2 单个 Agent 的动态影响图结构

MADID 中的一个 Agent 所对应的内层影响图结构如图 1(c) 所示. Agent  $j$  在  $t$  时的最优行为  $d_t^j$  可表示为

$$d_t^j = \arg \max_D \sum_{X_t^j} P_{\delta_1} (X_t^j, Y_t^j, D_t^j) U_t (D_t^j, X_t^j, Y_t^j) \quad (10)$$

其中,  $D_t^j, X_t^j, Y_t^j, U_t^j$  分别为  $t$  时 Agent  $j$  的决策结点、状态结点、观察结点和效用结点;  $P_{\delta_1} (X_t^j, Y_t^j, D_t^j) = P (X_t^j | Y_{0:t}^j, D_{0:t}^j) P (Y_t^j | X_t^j) P_{\delta_t^j} (D_t^j | Pa(D_t^j))$ .

给定 Agent  $j$  的观察变量序列和决策变量序列, Agent  $j$  在  $t$  时状态变量的概率分布为

$$P \{ X_t^j | Y_{0:t}^j, D_{0:t}^j \} = \frac{\pi_0 P(Y_1^j | X_1^j) P(D_1^j | X_1^j, Y_1^j)}{P(Y^j, D^j)} \times \prod_{t=2}^T P(X_t^j | X_{t-1}^j) P(X_t^j | Y_t^j, D_t^j) \quad (11)$$

其中,  $\pi_0$  为 Agent  $j$  的初始状态变量的概率分布,  $Y^j$  为观察变量序列,  $D^j$  为决策变量序列.

## 5 MADIDs 的分层分解

由于概率图模型的推理是一个 NP 问题, 为了

提高推理效率需要对 MADIDs 进行分层分解, 分层分解的主要过程是: 先将 MADIDs 的全局影响图分解成联合树, 然后对该联合树中的团进行增量分解, 生成用于推理的联合树。

### 5.1 初始全局影响图分解成联合树

**定义 4.** 带弦图. 带弦图  $G$  是一个无向图, 图中任意一条长度大于 3 的环都是带弦的。

**定义 5.** 团(cluster). 团是一个最大无向完备图, 其中最大完备图是指图中每个结点和图中的其它结点都相连. 带弦图中的每个不再被其他完备子图包含的完备子图对应一个团。

在上面两个定义的基础之上, 给出概率图可分解性的定义<sup>[17]</sup>。

**定义 6.** 概率图的可分解性. 设图  $G(V, E)$  的分解树是一个元组  $(V_C, T)$ , 其中  $V_C = \{X_{C_i} \mid C_i \in C\}$  是  $G$  中结点子集的集合;  $T$  是一个树,  $T$  中结点是  $V_C$  的子集, 且满足以下 3 个条件: (1)  $\bigcup_{C_i \in C} X_{C_i} = V$ ; (2) 对于每条边  $(v, w) \in E$ , 有一个  $X_{C_i} \in V_C$  包含了  $v$  和  $w$ ; (3) 对任意的  $C_i, C_j, C_k \in C$ ,  $C_j$  为  $C_i$  和  $C_k$  的连接团, 则有  $X_{C_i} \cap X_{C_k} \subseteq X_{C_j}$ 。

由概率图的可分解性条件可知图  $G$  的分解树  $T$  的树宽为  $\max_{C_i \in C} (|X_{C_i}| - 1)$ 。

**定义 7.** 联合树(Junction Tree, JT). 联合树是一个二元组  $T = (\Gamma, \Delta)$ , 其中,  $\Gamma$  是联合树中团结点的集合,  $\Delta$  是分割团的集合. 对任意一对相邻团  $C_i$  和  $C_j$ ,  $C_i \in \Gamma$ ,  $C_j \in \Gamma$ ,  $S_k \in \Delta$ ,  $S_k$  为  $C_i$  和  $C_j$  之间的一个分割团, 则有  $V_{S_k} = V_{C_i} \cap V_{C_j}$ 。

MADID 的初始全局影响图分解联合树的过程可描述如下:

1. 对图  $G$  进行正规化(normalization)操作, 即在有共同子结点的两个结点之间添加一条无向边, 然后去掉所有边的方向, 从而生成正规化图;
2. 对正规化图进行三角化操作, 是指通过对正规化图进行结点消除操作生成一个带弦图, 得到一个三角图; 该三角图的团结点含有多个 Agent;
3. 将三角图变成联合树, 在直接相连的两团之间, 生成一个分割团, 通过所生成的分割团将它们相连。

### 5.2 MADIDs 的联合树中团的增量分解

#### 5.2.1 联合树增量分解的相关定义

**定义 8.** 势(potential). 势是一个非负实函数(简称为函数), 该函数将变量或变量集的每个实例映射到一个非负的实数上, 用  $\phi$  或  $\psi$  表示. 设  $C$  为 MADIDs 全局影响图所对应的联合树中的任意一个团, 则  $C$  的概率函数  $\phi_C$  和效用函数  $\psi_C$  可表示为

$$\phi_C = \prod_{X \in V_C} P(X \mid Pa(X)), \quad \psi_C = \sum_{X_i \in V_C} \psi_{X_i}.$$

**定义 9.** 边际化(marginalization, marg). 设  $R \subseteq C$ , 团  $C$  的概率和效用函数分别为  $\phi_C$  和  $\psi_C$ , 则  $R$  的概率和效用边际化为

$$\text{marg}(C, R, \phi) = \sum_{V_C \setminus R} \phi_C; \quad \text{marg}(C, R, \psi) = \frac{\sum_{V_C \setminus R} \phi_C \psi_C}{\sum_{V_C \setminus R} \phi_C}.$$

**定义 10.** MADIDs 的策略依赖性. 决策变量集  $D$  将状态变量集  $X$  和观察变量集  $Y$  分解成集合序列  $I_0, I_1, \dots, I_n$ , 决策变量和集合序列之间存在一种偏序关系:  $I_0 < D_1 < I_1 < \dots < D_n < I_n$ ; 其中, 决策变量  $D_{k+1}$  依赖于变量集合  $I_k$  ( $0 < k < n$ ), 决策变量  $D_k$  依赖于变量集合  $I_{k-1}$ 。

#### 5.2.2 联合树中团的增量分解

设 MADID 初始全局影响图的联合树为  $T = (\Gamma, \Delta)$ , 设  $T$  的一个可能分割集是  $SP(T)$ , 其中,  $SP(T)$  中的元素是一个有序的二元组, 且该二元组中的每个元素完全属于一个团.  $SP(T)$  可以表示为  $SP(T) = \{\{u, v\} \mid \exists C \in \Gamma, \{u, v\} \subseteq V_C\}$ , 其中,  $u$  和  $v$  都是变量的集合,  $u$  是一个 Agent 自身所含有的变量和它的决策结点的父结点集,  $v$  是去掉  $u$  中的 Agent 自身所含有的变量后剩余的变量集。

MADIDs 的联合树中的团结点含有多个 Agent, 联合树中团结点的分解是以 Agent 为主体, 以 MADIDs 的策略为指导. 在  $T = (\Gamma, \Delta)$  上的一个分解  $\{u, v\} \in SP(T)$  的过程描述如下:

1. 移去一个  $\{u, v\} \subseteq V_C$  的团  $C$ , 添加团  $C_u$  和  $C_v$ , 其中,  $V_{C_u} = V_C \setminus \{v\}$  和  $V_{C_v} = V_C \setminus \{u\}$ 。
2. 添加一个新的分割团  $S$ , 且有  $V_S = V_C \setminus \{u, v\}$ ; 根据策略依赖性使  $C_u$  和  $C_v$  在联合树  $T$  中的位置满足偏序关系, 并通过  $S$  相连接。
3. 若  $u \in V_K$  ( $K$  为  $C$  的邻近团), 将  $K$  的分割团中的变量移到  $C_u$ ; 对  $C_v$  也同样操作。
4. 概率和效用函数的调整. 新团的概率和效用函数分别是  $\phi_{C_u} = \text{marg}(C, v, \phi)$ ,  $\psi_{C_u} = \text{marg}(C, v, \psi)$  和  $\phi_{C_v} = \text{marg}(C, u, \phi)$ ,  $\psi_{C_v} = \text{marg}(C, u, \psi)$ . 新分割团的概率和效用函数分别是  $\phi_S = \text{marg}(C, (v, u), \phi)$ ,  $\psi_S = \text{marg}(C, (u, v), \psi)$ 。

MADIDs 的联合树中的团都按照以上方式分解, 最后得到用于推理的联合树. Kjaerulff 提出贝叶斯网络的一种近似方案<sup>[16]</sup>, 通过移去边来对联合树中的团结点进行分割, 以提高联合树的推理效率. 在 Kjaerulff 方案的基础上, 上面给出一种 MADIDs 的联合树中团的增量分解方法, 该方法在分解中增加了效用结点的边际化操作. 其中二元组  $\{u, v\}$  在

策略依赖性作用下变得有序,且这种在有指导下的分解可以提高分解质量。

## 6 MADIDs的扩展BK(EBK)推理算法

### 6.1 BK 算法

为了提高动态贝叶斯网络推理的计算效率,BK算法根据系统不同部分之间的弱交互性,人工地将系统  $W$  划分成相对独立的子系统,即  $W = \{W_1, W_2, \dots, W_k\}$ ,其中  $W_i \in W$  是一个子系统;但对系统进行人工地分解会引入较大的误差.系统在  $t$  时的联合概率分布近似为子系统概率分布的因式:

$$P(X_t | Y_{0:t}) \approx \prod_{i=1}^k P(X_t^{w_i} | Y_{0:t}) \quad (12)$$

基于BK算法的分解思想,则在  $t$  时 MADIDs 中状态变量的联合概率分布为

$$P(X_t | Y_{0:t}, D_{0:t}) \approx \prod_W P(X_t^W | Y_{0:t}, D_{0:t}) \quad (13)$$

其中,  $X_t^W$  是  $X_t$  的一个子集。

### 6.2 MADIDs 的 EBK 推理算法

设图  $G$  的联合树  $T = (\Gamma, \Delta)$ ,  $C_i$  和  $C_j$  是  $T$  的两个邻接结点,当  $S_k = C_i \cap C_j$  时,有  $X_{S_k} = X_{C_i \cap C_j}$ ,则根据条件独立性, $G$  的联合概率分布可以表示为<sup>[17]</sup>

$$P(X) = \frac{\prod_{C_i \in \Gamma} P(X_{C_i})}{\prod_{S_k \in \Delta} P(X_{S_k})} \quad (14)$$

由式(14)可知将 MADIDs 分解成联合树不会引入误差;对 MADIDs 进行分层分解,可以得到在分割团的条件下 MADIDs 中状态变量在  $t$  时的联合概率分布:

$$\hat{P}(X_t | Y_{0:t}, D_{0:t}) \approx \frac{\prod_{C_i \in \Gamma_t} P(C_i | Y_{0:t}, D_{0:t})}{\prod_{S_i \in \Delta_t} P(S_i | Y_{0:t}, D_{0:t})} \quad (15)$$

#### 6.2.1 EBK 算法的信息传播

由式(15)可知,在分割团的条件下,MADIDs 在  $t$  时状态变量的联合概率分布为团的概率分布的因式;在加入新的证据后,需要对这些团的概率分布进行更新,并作为下一时间片的先验信息,从而实现信息的传播;信息的传播过程如图 2 所示。

由图 2 可知,信息的传播是先从 MADIDs 因式(factored)形式的先验联合概率分布  $\tilde{\alpha}_{t-1}$  (即式(15)右侧形式)开始,加入新的证据对概率分布进行精确更新,得到 MADIDs 耦合(coupled)形式的联合概率分布  $\tilde{\alpha}_t$  (即式(15)左侧形式),即图 2 中的  $U$  步操

作;然后通过执行概率函数和效用函数的边际化操作,将耦合形式的后验联合概率分布  $\tilde{\alpha}_t$  投影到因式形式的联合概率分布  $\tilde{\alpha}_{t+1}$ ,投影是对概率分布的一种近似操作,即图 2 中的  $P$  步操作。

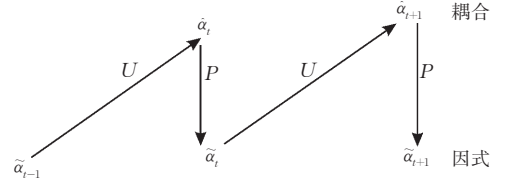


图 2 EBK 算法的信息传播过程

#### 6.2.2 EBK 算法的描述

EBK 算法的主要步骤是:先对 MADIDs 分层分解生成用于推理的联合树,然后由该联合树生成原型联合树(Prototype Junction Tree, PJT),最后通过 PJT 上的信息传播进行推理.由于 EBK 算法的推理是在原型联合树进行的,下面先给出原型联合树的定义及生成过程。

**定义 11.** 原型联合树. 原型联合树  $G^T$  中的结点是由  $\Gamma_t, \Gamma_{t+1}$  和  $F_{ts}$  组成。

变量  $x$  的家庭(family)  $f_{tx}$  是由该变量和其父结点组成  $f_{tx} = \{x \cup pa(x)\}$ ,转移网络  $V_{ts}$  中所有结点的家庭集为  $F_{ts} = \{f_{tx} | x \in V_{ts}\}$ ,  $f_{tx}$  的概率函数为  $\phi_{tx}$ ,  $F_{ts}$  的概率函数为  $\Phi_{ts} = \prod_{x \in V_{ts}} \phi_{tx}$ 。

MADIDs 原型联合树的生成和初始化过程如下:

1. 将 MADIDs 的初始外层影响图  $B_0$  分解成一个联合树,然后通过对该联合树中的团进行增量分解得到用于推理的联合树  $J_0$ ,  $J_0 = \{\Gamma_0, \Delta_0\}$ ;

2. 生成一个空  $G^T$ ,将团集  $\Gamma$  中所有团和  $F_{ts}$  中的所有结点合并到  $G^T$  中;

3. 对每个  $C_0 \in \Gamma_0$ , 边际化:  $\phi_{C_0} = \text{marg}(J_0, V_{C_0}, \phi)$ ,  $\phi_{C_0} = \text{marg}(J_0, V_{C_0}, \psi)$ ;

同样对每个  $S_0 \in \Delta_0$ , 边际化:  $\phi_{S_0} = \text{marg}(J_0, V_{S_0}, \phi)$ ,  $\phi_{S_0} = \text{marg}(J_0, V_{S_0}, \psi)$ ;

4. 原型联合树  $G^T$  的初始概率函数为  $\Phi_{G^T} = \Phi_{ts}$ 。

基于原型联合树, MADIDs 的 EBK 算法可描述如下:

1. 构建一个原型联合树  $G^T$ , 将  $F_{ts}$  中的概率函数  $\Phi_{ts} = \prod_{x \in V_{ts}} \phi_{tx}$  和  $G^T$  相乘;

2. 对于推理所用的联合树  $J_0$ ,  $t$  从 0 开始;

3. 将  $G^T$  复制给一个过渡的原型联合树  $G$ ;

4. 对于每个  $C \in \Gamma_t, S \in \Delta_t$ , 有  $G = \frac{\prod_{C \in \Gamma} \phi_{C_t}}{\prod_{S \in \Delta} \phi_{S_t}} \Phi_{ts}$ ;

5. 将  $Y_{t+1}$  引入到  $G$  中, 执行局部信息传递, 并调整  $G$



使其达到新的一致状态；

6. 对每个  $C \in \Gamma_{t+1}$ ,  $\phi_{C_{t+1}} = \text{marg}(G, V_{C_{t+1}}, \phi)$ ,  $\psi_{C_{t+1}} = \text{marg}(G, V_{C_{t+1}}, \psi)$ ;
- 对每个  $S \in \Delta_{t+1}$ ,  $\phi_{S_{t+1}} = \text{marg}(J_0, V_{S_{t+1}}, \phi)$ ,  $\psi_{S_{t+1}} = \text{marg}(G, V_{S_{t+1}}, \psi)$ ;
7.  $t = t + 1$  返回到步 3.

## 7 MADID 模型及实验

实验以贝叶斯网络工具包<sup>[18]</sup> (FullBNT)为基础,以 Matlab 为程序运行环境;实验的时间步为 100. BK 算法采用人工方式对 MADIDs 进行分解,但人工分解难以得到最优划分;EBK 算法的分层分解是自动进行的,可以得到最优划分.

对于图 1 模型,设球员  $A, B$  的决策空间为(传球,跑位,射门),状态空间为(速度改变,方向改变),观察变量空间为(速度,方向,位置); $C$  球员的状态和观察变量空间和  $A, B$  相同.在禁区内,根据射手的位置是否有利于射门,将连续的观察空间离散为图 3 所示的网格区域,其中,白圆、黑圆、空白和叉号分别对应有利、较有利、一般、不利的射门区域;状态空间离散成速度和方向的改变范围,其中方向可离散成 Agent 周围 8 个网格对应的方向.图 4 是图 3 场景进一步离散的形式(表示守门员的观察空间),设守门员位于图 4 中第三行的中间网格中,当射手处在不同网格时,守门员对射手的威胁权重对应为该网格中的数值,其值越大表示守门员对射手的威胁越大,而守门员采用移向最强威胁网格的固定策略.其中图 4 中的权值是根据经验设定的,可通过学习进行修改.

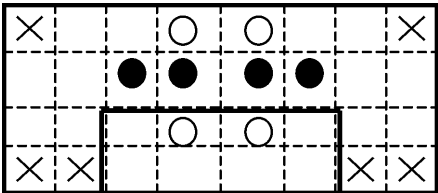


图 3 模型的场景

|     |     |     |
|-----|-----|-----|
| 0.2 | 0.3 | 0.2 |
| 0.3 | 0.6 | 0.3 |
| 0.2 | 0.3 | 0.2 |

图 4 威胁权重

对于 MADID 的不同 PJT 边界  $s = 1100, \dots, 3500$ ,表 1 给出 BK 算法和 EBK 算法所生成的 PJT 规模、平均推理误差  $\delta$  和运行时间.

表 1 对于不同 PJT 规模下的算法的平均误差和时间

| 算法        | PJT 规模 | 平均误差( $\delta$ ) | 时间/s  |
|-----------|--------|------------------|-------|
| BK(1100)  | 998    | 0.9E-3           | 16.04 |
| BK(2200)  | 1755   | 0.76E-3          | 19.06 |
| EBK(1100) | 998    | 0.81E-3          | 16.07 |
| EBK(1700) | 1556   | 1.6E-4           | 18.06 |
| EBK(2200) | 1755   | 1.2E-4           | 19.12 |
| EBK(3500) | 2793   | 0.96E-4          | 21.63 |

表 1 显示:对于 BK 算法,当 PJT 边界  $s = 2200$  时,算法的运行时间为 19.06s;而 EBK 算法可以对  $s = 2200$  作进一步分解,将减 PJT 边界  $s$  减小到 1700,此时 EBK(1700)的运行时间为 18.06s,其运行时间和平均误差都比 BK(2200)要小;当 BK 算法和 EBK 都使用最好的分解  $s = 1100$  时,EBK(1100)的精度要明显高于 BK(1100),而它们的时间性能基本相等,另外在实际中 BK 算法难以通过人工分解方法获得最优的分解.

在 100 个时间步上,BK(1100)和 EBK(2200)在每个时间步上误差的传播变化如图 5 所示.与 BK 算法相比,EBK 算法的误差峰值基本上都较低,误差波动不明显;而 BK 算法的误差的峰值很高,误差的波动很大;另外,从平均误差来看 EBK 算法的平均误差要明显优于 BK 算法的平均误差.

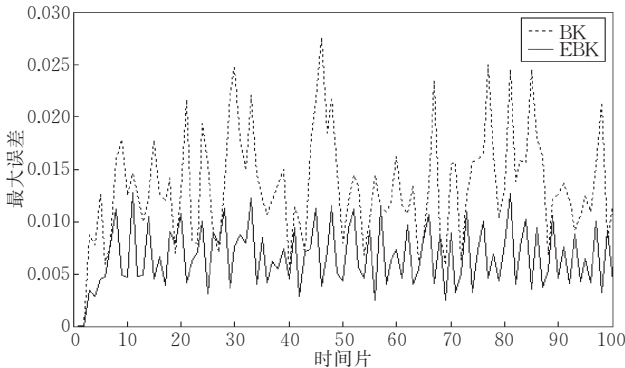


图 5 EBK 和 BK 算法的最大误差

表 2 对于不同 PJT 规模下的算法的平均误差和时间

| 算法         | PJT 规模 | 平均误差( $\delta$ ) | 时间/s  |
|------------|--------|------------------|-------|
| BK(3500)   | 3376   | 1.22E-3          | 22.15 |
| EBK(3500)  | 3376   | 1.03E-3          | 22.28 |
| EBK(4000)  | 3992   | 6.27E-4          | 23.07 |
| EBK(7000)  | 6552   | 2.08E-4          | 25.40 |
| EBK(10000) | 8212   | 1.76E-4          | 26.62 |
| EBK(15500) | 14394  | 1.64E-4          | 31.96 |

另外,我们还对  $10\text{m} \times 10\text{m}$  固定场景中的 4 个 Agents 之间传接球问题进行仿真实验.设  $A, B$  球员为传接球配合 Agent, $C, D$  为干扰 Agent,其中  $A$  是控球 Agent.将 Agent 连续的状态空间映射到所处区域的离散观察空间,根据他们之间的相互关系

对区域进行划分：强干扰、中干扰和小干扰。

对于上面的传接球问题,在 100 个时间步上,取不同 PJT 边界  $s=3500, \cdots, 15500$ ,表 2 给出 BK 算法和 EBK 算法所生成的 PJT 规模、平均推理误差  $\delta$  和运行时间;BK(3500)和 EBK(4000)在每个时间步上误差的传播变化如图 6 所示.表 2 和图 6 的实验结果也证明了 EBK 算法的有效性.

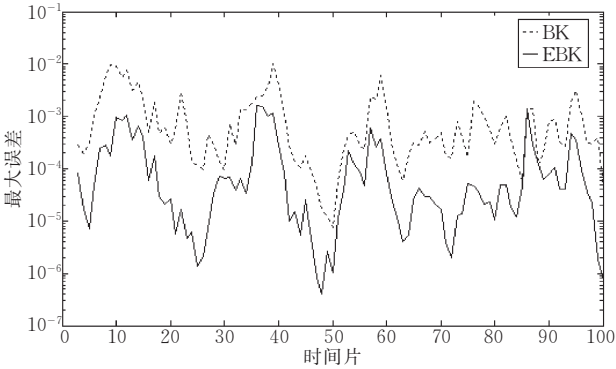


图 6 EBK 和 BK 算法的最大误差

8 结束语

针对 MAIDs 不能建模动态环境和 MMDPs 难以表示 Agents 之间的结构关系等问题,提出能表示动态环境中多 Agent 之间结构关系的 MADIDs 模型.基于分层分解的概率分布近似方法,提出 MADIDs 的 EBK 算法近似推理算法;仿真实验结果表明 EBK 算法的有效性.

然而,EBK 和 BK 算法都面临一个两难问题:复杂性降低会引起精度损失,而精度的提高又会导致复杂性的增加,因而推理精度和复杂性的均衡将是一个值得研究的问题.

参 考 文 献

[1] Howard R A, Matheson J E. Influence diagrams//Howard R A, Matheson J E eds. Readings on the Principles and Applications of Decision Analysis, Menlo Park; Strategic Decisions Group, 1984, 2: 719-792

[2] ZHANG Wei-Hong, Ji Qiang. A factorization approach to evaluating simultaneous influence diagrams. IEEE Transactions on Systems, Man and Cybernetics, Part A, 2006, 36 (4): 746-757

[3] Koller D, Milch B. Multi-agent influence diagrams for representing and solving games//Proceedings of the IJCAI. Seattle, USA, 2001; 1024-1034

[4] Gal Y, Pfeffer A. A language for modeling Agents decision making processes in games//Proceedings of the AAMAS-2nd. Melbourne, Australia, 2003; 265-272

[5] Marengoni Mauricio. Decision making and uncertainty management in a 3D reconstruction system. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(7): 852-858

[6] Wang Hong-Wei, Li Chen, Liu Hui-Xin. Entropic measurements of complexity for Markov decision processes. Control and Decision, 2004, 19(9): 983-987(in Chinese)  
(王红卫,李琛,刘会新.马尔可夫决策过程复杂性的熵测度.控制与决策,2004,19(9):983-987)

[7] Koller D, Parr R. Computing factored value functions for policies in structured MDPs//Proceedings of the IJCAI-99. Morgan Kaufmann, 1999; 1332-1339

[8] Carlos Guestrin, Daphne Koller. Efficient solution algorithms for factored MDPs. Journal of Artificial Intelligence Research, 2003, 19(10): 399-468

[9] Barto A G, Mahadevan. Recent advances in hierarchical reinforcement learning. Discrete Event Dynamic Systems; Theory and Applications, 2003, 13(1/2): 41-77

[10] Yao Hong-Liang, Wang Hao, Zhang You-Sheng. Simplifying computation of dynamic influence diagrams//Proceedings of the IEEE the 3rd International Conference on Machine Learning and Cybenetics. Shanghai, China, 2004, 8: 59-62

[11] Dagum P, Luby M. Approximating probabilistic inference using Bayesian networks is NP-hard. Artificial Intelligence, 1993, 60(1): 141-153

[12] Boyen X, Koller D. Tractable inference for complex stochastic processes//Proceedings of the UAI-98. San Francisco; Morgan Kaufmanns, 1998; 33-42

[13] Draper D. Clustering without (thinking about) triangulation//Proceedings of the UAI-95. CA; Morgan Kaufmann, 1995; 125-133

[14] Merkle A, Riedmiller M. A reinforcement learning approach to robotic soccer//Proceedings of the Robocup 2001. Springer, 2001; 435-440

[15] Zhang Run-Mei, Wang Hao, Yao Hong-Liang. Influence diagrams and its application in RoboCup. Journal of System Simulation, 2005, 17(1): 134-137(in Chinese)  
(张润梅,王浩,姚宏亮.影响图及其在 RoboCup 中的应用.系统仿真学报,2005,17(1):134-137)

[16] Kjaerulff U. Reduction of computational complexity in Bayesian networks through removal of weak dependences//Proceedings of the UAI-94. Seattle, USA, 1994; 374-382

[17] Paskin M A. Thin junction tree filters frontier for simultaneous localization and mapping//Proceedings of the IJCAI-03. Acapulco, Mexico, 2003; 1157-1164

[18] Murphy K. The bayes net toolbox for Matlab. Computing Science and Statistics, 2001, 33(2): 1024-1034





**YAO Hong-Liang**, born in 1972, Ph. D. candidate. His main research interests include artificial intelligence and knowledge engineer.

**WANG Hao**, born in 1962, Ph.D. , professor. His main research interests focus on artificial intelligence.

**ZHANG You-Sheng**, born in 1941, Ph. D. , professor. His main research interests include artificial intelligence and image recognition, and understanding.

**WANG Rong-Gui**, born in 1966, Ph. D. , associate professor. His main research interests focus on Bayesian technology.

**Background**

Due to high dimension and uncertainty of the complex system, the complexity system is often hard to be represented and processed. The knowledge representation and computation methods of complex system are open hard problems in complex system research.

The complex intelligent system is an important branch of the complexity system. At present, many the researchers use probability models to model and process the complexity intel-

ligent system, such as Multi-Agent influence diagrams (MAIDs) and Multi-Agent Markov decision processes (MM-DPs). In order to better represent and process complex dynamic intelligent system, the authors of the paper present a new probability model, which is Multi-Agent Dynamic Influence Diagrams (MADIDs). Another, an approximate inference algorithm of MADIDs is proposed.