

基于对策论的 MAS-BDI 主体模型

Modeling MAS-BDI Agent in Game Theory

吴朝晖 忻 栋 潘云鹤

(浙江大学人工智能研究所 杭州 310027)

Abstract BDI Agent will choose the intention, which has minimal penalty and maximal income, to achieve its goal. While in multi-Agent system, it is hard for the Agent to find the best intention, not only because of the lack of information, but also the high complexity of computation. The factors, which will affect the Agents' decision-making, are the cooperation state of Agents, the Agents' reasoning process, etc. In this paper, we introduce a new model of MAS-BDI Agent based on game theory, and propose the intention selection strategy for a single BDI Agent, (cooperated) multi-Agent and (uncooperated) multi-Agent.

Keywords MAS, BDI, Intention, Game theory

1. 引言

1987年 Bratman^[1]首先从哲学上对行为意图进行研究, 此项研究对人工智能产生了深远的影响。Bratman 认为只有保持信念(Belief)、愿望(Desire)和意图(Intention)的理性平衡才能有效地解决问题。基于这个理论, Rao 和 Georgeff^[2]提出了 BDI 模型: 其中 B(信念)是 BDI 主体对自身和世界的认识; D(愿望)是主体要完成的任务; I(意图)是主体所选择的计划。该模型的形式化方法是基于时间的分支模型: 信念、愿望和意图分别是分支时间结构。如果将主体的愿望空间描述成一棵决策树, 每一节点表示一个选择, 一个主体的意图选择过程可以转化为主体在一棵决策树上选择最佳路径的问题^[3]。

本文将该模型扩展到 MAS 领域。在多主体系统中, 各个独立主体之间相互合作; 但是为了获得最大收益, 它们同时也相互竞争, 主体处于内外部信息动态变化的环境中。在这种情形下, 我们重新定义了主体的信念、愿望和意图; 并提出了基于对策论^[4]的对非合作主体和合作主体的意图选择策略。全文结构如下: 第 2 节提出模型假设; 第 3 节定义 MAS 环境下的 BDI 主体; 第 4 节描述基于对策论的非合作主体意图选择策略; 第 5 节讨论合作主体的意图选择策略; 最后是全文小结。

2. 模型假设

我们以多主体系统来模拟现实世界, 在 MAS 中每一主体都具有各自的信念、愿望和意图(BDI 模型)。

定义 1 系统状态的集合表示为 S 。在 t 时刻, 系统的状态以 s^t 表示。定义系统的初始状态为 s^0 。

定义 2 有限集 $N = \{1, 2, \dots, n\}$ 表示各个主体的位置。并且每一位置有且只有一个主体。 A_i 表示在位置 i 处的主体。 $G(N)$ 表示 N 的一个子集, 即一个主体组。

定义 3 有限集 $Action(A_i)$ 表示主体 A_i 的动作集合, 在每一特定时刻, 主体只能采取其中的一个动作, 给定当前的系统状态和主体的动作可以得出系统状态的转移方程:

$$\Psi: S \times \prod_{i \in N} Action(A_i) \rightarrow S$$

定义 4 对于有限主体的有限动作, 我们总可以将时间分割出来, 用自然数来表示。定义系统运行的整个时间过程以集合 $\{1, 2, \dots, T\}$ 表示, 其元素之间的关系是 $<$ (表示前后)。

我们假设时间是离散的并且系统保持同步, 即在任一时刻 $t \in \{1, 2, \dots, T\}$, 所有主体都采取一个行动(没有动作的以空动作表示); 而在任一时刻 $t \in \{1, 2, \dots, T\}$, 没有一个主体有所行动。例如, 在某一时刻 $t \in \{1, 2, \dots, T\}$, 各个主体的动作为 $a_i, i \in N$, 系统的状态为 s^t , 根据系统转移方程, 在下一时刻 $t+1$, 系统状态为 $s^{t+1} = \Psi(s^t, a_1, a_2, \dots, a_n)$ 。

3. MAS 主体的 BDI 描述

3.1 主体信念

主体的信念(以 Bel 表示)是主体对系统当前状态的认识和对系统将来状态的估计。在任一时刻, 主体观

察到的只是系统当前状态 s 的一个子集 σ , 主体的观察函数为 $Pception(s)$; 而它对将来状态的估计则来源于其他主体对它的承诺 ($Commit-to(A_i)$) 和主体对其他主体的了解 ($Know(A_i)$)。

3.1.1 主体观察信念 是主体对全局状态的认知和估计。在时刻 t , 系统状态为 s' , 设主体 A_i 所观察到的局部状态为 σ'_i , 由主体的观察函数得知, $Pception(s') = \sigma'_i$ 。因此主体 A_i 对系统全局状态的把握为:

$$s' = Pception_i^{-1}(\sigma'_i)$$

3.1.2 主体承诺信念 是其他主体对该主体的保证。设在时刻 t , 主体 A_i 向主体 A_j 作出的承诺为 $Commit_{j,i}^t$ 。定义:

$$Commit_i^t = \prod_{j \in N, j \neq i} Commit_{j,i}^t$$

表示在 t 时刻主体 A_i 接受到的所有其他主体的承诺。

在承诺还没被收回之前, 主体维护其接收到的所有承诺。主体的承诺信念是一个与时间有关的集合:

$$Commit-to^t(A_i) = \prod_{t=1,2,\dots} Commit_i^t$$

一旦承诺被收回, 主体将修改其信念。如在时刻 t , 主体 A_i 收回其在时刻 t' 时向主体 A_j 所作出的承诺。则主体 A_i 将其承诺意图修改为:

$$Commit-to^t(A_i) = Commit-to^t(A_i) - Commit_{j,i}^{t'} \quad t' < t$$

主体收回承诺的理由可分为:

1. 承诺完成 (以 $Goal-reached(C; Commitment)$ 表示)
2. 承诺无法完成 (以 $Goal-unreachable(C; Commitment)$ 表示)
3. 承诺不需要完成 (以 $Goal-undesired(C; Commitment)$ 表示)

根据不同情形, 主体收回承诺的策略可有以下三种:

If $Goal-Reached(C; Commitment)$ Then Drop-Commitment(C; Commitment)
 If $Goal-Reached(C; Commitment)$ and $Goal-Unachievable(C; Commitment)$ Then Drop-Commitment(C; Commitment)
 If $Goal-Reached(C; Commitment)$ and $Goal-Unachievable(C; Commitment)$ and $Goal-Undesired(C; Commitment)$ Then Drop-Commitment(C; Commitment)

3.1.3 主体知识信念 是主体对其他主体的了解。主体就是根据这个信念建立其他主体的可能运行模型, 进而进行推理采取行动。设在时刻 t , 主体 A_i 对主体 A_j 所具有的知识信念为 $Know_{i,j}^t$ 。可知, $Know_{i,j}^t \subset Bel(A_j)$ 。定义:

$$Know^t(A_i) = \prod_{j \in N, j \neq i} Know_{i,j}^t$$

主体的知识信念可以传递, 例如主体 A_i 可以了解主体 A_j 所具有的知识信念, 但是它们之间的信息量是

递减的。定义 $Know_{i,j,k}^t$ 为 t 时刻主体 A_i 了解的主体 A_j 关于主体 A_k 的知识信念。显然有 $Know_{i,j,k}^t \subset Know_{i,j}^t$ 。综上所述, 在时刻 t , 主体的信念表现为:

$$Bel^t(A_i) = Pception_i^{-1}(\sigma'_i) + Know^t(A_i) - Commit-to^t(A_i)$$

3.2 主体愿望

主体的愿望是主体希望达到的状态, 它是系统状态的一个子集。我们将主体的愿望映射成实数。通过愿望评价函数 μ 和阈值 α 来判断主体的愿望是否达到。设从系统开始运行到主体达到其愿望, 主体所经历的时间为 $[1, 2, \dots, t]$ 。主体对所经历的局部状态为 $\delta^t = [\sigma^1, \sigma^2, \dots, \sigma^t]$ 。设 Q^t 为所有可能存在的 δ^t 的集合, 定义:

$$\mu: Q^t \rightarrow R$$

主体的愿望为 $Desire(A_i) = \{\delta^t; \mu(\delta^t) > \alpha; \delta^t \in Q^t\}$ 。其中 T 为主体结束运行时间。

3.3 主体意图

主体意图表示主体准备实行的一个计划。其表现为: 一个主体要改变已有的意图必须要有理由; 一个主体不能无视环境的改变而坚持不符合实际的或已不重要的意图^[5]。

显然主体将寻找并实施代价最小、收益最大的意图。定义在时刻 t , 主体 A_i 的意图为 $Intention^t(A_i)$ 。

3.3.1 联合意图 是需要多个主体共同执行的意图。若主体意图只涉及该主体本身, 就不需要任何形式的承诺, 若涉及其他主体 (即联合意图), 承诺就显得必要了。

联合意图 $Intention^t(A_i, A_j, \dots, A_k)$ 的执行条件为:

$$\prod_{i,j \in \{1,2,\dots,t\}, i \neq j} Commit(i, j)$$

3.3.2 理想意图 是指在系统中, 每一主体都选择一个对于主体自身来说的最佳意图。在多主体系统中, 意图选择有以下三个因素^[5]:

1. 主体集合 N

2. 每一主体的可行有限意图集: $I-Set(A_i), i \in N$; 每一主体的选择意图构成系统的意图 SI , 定义:

$$SI = \prod_{i \in N} Intention(A_i), Intention(A_i) \in I-Set(A_i)$$

3. 对于每一个系统意图 SI , 每一主体都有一个评价函数: $P_i(SI), i \in N$

每一主体都希望自己的 $P_i(SI)$ 达到最大值。定义:

$$SI \parallel P_j = (I_1, \dots, I_{j-1}, P_j, I_{j+1}, \dots, I_n)$$

其中 $I_i \in I-Set(A_i), i \in N$ 并且 $P_j \in I-Set(A_j)$

若对任何 $P_j \in I-Set(j)$ 且 $j \in N$, 有: $P_i(SI \parallel P_j) \leq P_i(SI)$ 则 SI 就是理想意图。

之所以称之“理想”是因为: 1) 主体不可能知晓所

有其他主体的意图;2)主体很难获得全局的评价函数;3)其计算都毫无例外地牵涉到非线性方程组的求解,因而也无法导出一个切实有效的计算方法,

3.4 主体运行模型

主体的运行模型如图1所示。主体的意图通过动作来改变系统的状态;主体通过对系统的感知、相互间的交流和对承诺(收回承诺)的处理来维护自己的信念。

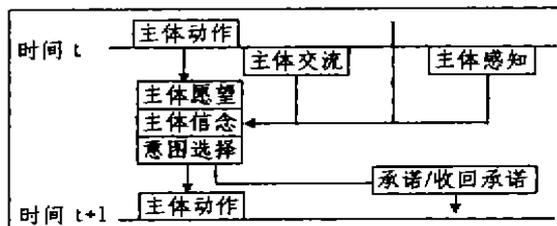


图1 主体运行模型

4. 基于对策论的非合作主体意图选择

对策论^[6]的研究是以 Von Neumann 的效用理论为基础的。本文借用对策论中决策树的一些知识讨论主体的意图选择。在任一时刻,我们将主体所具备的信念转化为一棵决策树,进而在这棵决策树上搜寻最佳路径,找出主体的意图。

4.1 从信念到决策树

在任一时刻 $t \in [1, 2, \dots, T]$, 主体根据自身的信念建立决策树。这是因为在多主体系统下, 主体处于内外部信息动态变化的环境中, 主体所具备的信念信息会随时修改更新。另外, 为保证主体能迅速地作出决定, 决策树不可能建得很深; 并且在系统刚开始运行时, 主体往往没法预见最后的情形。这就需要决策树能实时生成, 步步深入。

在第3节中我们已经定义主体的信念由观察信念、承诺信念和知识信念组成。从信念到决策树的转化遵循如下规则:

1. 每一个主体都有自己相应的决策树;
2. 决策树的树根是主体当前状态;
3. 决策树向下发展表示主体对将来状态的预计, 以时间为基准, 树叶所对应的时间与树根所对应的时间差为树的深度;
4. 每一节点(除树根外)表示在将来某一时刻, 主体可能会达到的状态;
5. 节点之间的联系表示主体的动作, 即前后两个状态的迁移需要主体自身采用什么动作或其他主体执行什么动作;
6. 每个动作都有相应的置信度, 对于主体自身的

动作, 置信度为1; 对于其他主体的动作, 置信度为(0, 1]之间的实数。经承诺的动作置信度为1, 收回承诺的动作置信度为0(该动作将从决策树上删除), 未经承诺的动作根据主体的知识信念赋以0到1之间的实数。

下面是某一主体 A_i 在时刻 t 的决策树, 主体在 t 时刻的状态为 s_0 , 采用一个动作(它自身的或别的主体的)后, 系统时间为 $t+1$, 其状态为 s_{11}, s_{12} 和 s_{13} 之一, 依次类推。动作 A_1 表示是主体自身的动作; 动作 A_2 表示是主体 A_i 的动作, 其置信度为0.7; 动作 A_3 表示是主体 A_j 的动作, A_j 已经向主体 A_i 作出肯定的承诺, 故其置信度为1。对于收回承诺的动作, 按规则将它们从决策树上去除。

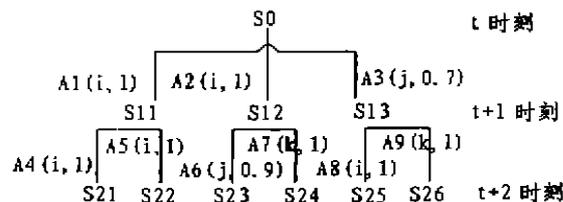


图2 一棵简单决策树

4.2 主体的推理行为

有很多策略^[6]可供主体推理时使用, 如最大最小原理, 最大期望原理等。我们定义一般形式的主体推理行为(Rational Behavior)如下:

$$Rb = (N, Action(N), \mu(a), p(a), x(a))$$

其中 N 为所有主体的集合, $Action(N)$ 为所有主体的所有动作集合, $a \in Action(N)$, $\mu(a) \in R, x(a) \in [0, 1]$ 表示采用动作 a 的获利 $\mu(a)$ 代价 $p(a)$ 的可能性为 $x(a)$ 。给定一个动作序列 (a_1, a_2, \dots, a_k) , 定义主体在动作 a_1 后的效用为:

$$v(a_1) = \sum_{i=1}^k (\mu(a_i) - p(a_i)) \times x(a_i)$$

我们以最大最小原理为例描述决策树的搜索, 它是从叶节点开始的回溯过程。

定义 T-Node 为终节点, 表示决策树中的叶子, 它不存在选择问题。对于中间节点, 我们作如下分类:

定义 D-Node 为决策节点, 表示主体可以支配该节点的信息, 选择以下的动作。在决策树中表现为由该点出发向下的动作都是该主体自身的。根据最大最小原理, 选择效用最大的动作(以函数 S-max 表示);

定义 C-Node 为机会节点, 表示主体无法支配该节点的信息, 其以下动作由其他主体或系统决定。在决定树中表现为由该节点出发向下的动作都不是该主体自身的。根据最大最小原理, 选择效用最小的动作(以函数 S-min 表示);

定义 M-Node 为混合节点,表示主体在这种状态下既可以采取自己的行动也可以等待特别的主体的动作。我们定义在这种情况下选择函数为 S-mix,说明如下:将 M-Node 分为两个子节点, D(M-Node) 和 C(M-Node), 分别表示决策子节点和机会子节点。在决策子节点中选出效用最大的动作 a_d , 其相应效用为 v_d ; 在机会子节点中选出效用最小的动作 a_c , 其相应效用为 v_c 。如果 $v_d < v_c$, 则选择 a_c , 否则选择 a_d 。

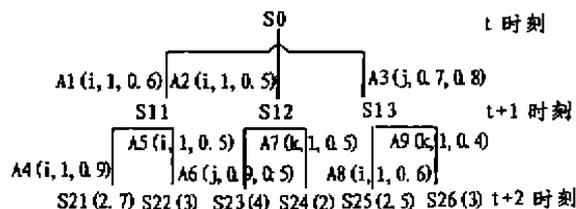


图 3 赋以利益和代价的决策树

再看上节提到的例子,我们赋以相应的效用(利益和代价的结合)。利益体现在状态,如 S26(3)表示如果主体能到达状态 S26,则可获利 3;代价体现在动作,如 A3(j, 0.7, 0.8)表示若采用动作 A3,代价是 0.8,可能性是 0.7。

根据上述规则,我们从叶节点开始回溯:

S11 是 D-Node,采用动作 A4,效用为 1.8,采用动作 A5,效用为 2.5,故取动作 A5,而 S11 的利益为 2.5;

S12 是 C-Node,采用动作 A6,效用为 3.15,采用动作 A7,效用为 1.5,故取动作 A7,S12 的利益是 1.5;

S13 是 M-Node, D(M-Node)的效用是 1.5, C(M-Node)的效用是 1.2,取动作 A8,S13 的利益是 1.5。

S0 是 M-Node, D(M-Node)的效用是 1.9, C(M-Node)的效用是 0.49,取动作 A1。

5. 合作主体的意图选择

为了便于研究,我们限于考虑两种极端的情形。如果各主体互不沟通,都单独行动,那么所得的就是主体非合作模式的意图选择,我们已经在上面讨论过了。另一种极端情形是,主体都在寻求能使自己获利更大的与他人合作的机会^[6],这就是主体合作模式的意图选择。

5.1 主体组

在合作对策中,主体组表示一部分主体所形成的集合。为了便于研究,我们假定:一旦一个主体组形成,其在意图选择过程中保持稳定;主体组内的利益分配

随意。

5.2 合作主体的意图选择

合作的多主体的意图选择使主体组 G(N)的获利最大。

如果将主体组看作一个新的主体,参与全局的意图选择,则基于主体组的意图选择策略与非合作主体意图选择策略是一样的。为简单起见,我们假设每一主体只能参加一个主体组,并且必须参加一个主体组(一个主体组可以只有一个主体)。

在主体组内部,主体的意图选择从属于整个主体组的利益,一般情形下,一个主体组有上层的监控主体(如果一主体组的组长主体)和下层的受控主体(组内其他主体),整个系统组成层次性的结构。在主体组内部有三种控制关系:1)监控主体产生一个完整的规划并发给受控主体;2)监控主体没有(或不能)生成一个完整的规划,只生成部分规划(抽象规划),并发给受控主体;3)监控主体无法生成或改善规划(由于知识或通信的限制),只发给受控主体一个高层目标。

由于监控主体对具体环境的知识的缺乏,他的完整规划可能有错误,在 2,3 两情况下,受控主体必须依靠自己更为精确的知识造成完整规划,提交监控主体通过并执行。

小结 本文就多主体环境下 BDI 主体的定义和意图选择作了一些有益的尝试。非合作和合作的多主体意图选择都是基于对策略作出的。非合作的多主体意图选择使每一主体都选择对自己最有利的意图。合作的多主体系统涉及到主体组的概念,在一个主体组中,有一个监控主体和若干个受控主体,合作的多主体以主体组的整体利益最大为目标。

参考文献

- 1 Bratman M. Intentions, Plans and Practical Reasoning. Harvard Univ. Press, Cambridge, MA, 1987
- 2 Rao A S, Georgeff M P. BDI Agents, From Theory to Practice. In: Proc. of the First Intl. Conf. on Multi-Agent Systems, San Francisco, USA, 1995
- 3 Rao A S, Georgeff M P. Modeling rational agents within a BDI-architecture. In: Proc. of Knowledge Representation and Reasoning (KR&R-91), 1991
- 4 Rapoport A. Game Theory As A Theory Of Conflict Resolution. Reidel Publishing Company, Holland, 1974. 1 ~ 103
- 5 史忠植 高级人工智能. 科学出版社, 1998
- 6 刘德铭, 黄振高. 对策论及其应用. 国防科技大学出版社, 1994