联合意图的理论框架*)

毛新军 王怀民 吴 刚
(国防科学技术大学计算机系 长沙410073)

The Theoretical Framework of Joint Intention

MAO Xin-Jun WANG Huai-Min WU Gang
(Dept. of Computer Science, National University of Defense Technology, Changsha 410073)

Abstract The joint intention is an important concept to specify and examine the social behaviors of multi-agent system. A theoretical framework of joint intention is presented. It covers two different joint intentions to represent different social characteristics, including joint achievement intention and joint maintenance intention. The characteristics of joint intention are investigated, the new semantics of joint intention is given, and several important properties are specified and proved. The theoretical framework can be used to support the development of multi-agent system.

Keywords Multi-agent system, Joint intention

0 引言

agent是指在某一环境下能够持续自主运行,具有社会性、反应性等特征的计算实体[5,11]。多agent系统由一组具有一定资源和能力、相对独立且交互合作的agent组成。由于多agent系统提供了更高层次的抽象模型,能够自然、贴切、直观地表示现实世界中的计算实体及其问题求解方式,因而有关agent理论和技术的研究引起了人们的高度重视。

在多agent系统中,由于agent间动作的相关性、单个agent能力的有限性、资源的分布性,以及为了满足全局性约束、实现问题求解的自然性,agent间进行联合工作是必须的[10],agent间的联合工作揭示了多agent系统的社会性行为。为了支持多agent系统的描述、设计和实现,必须提供有效的机制来规范和分析agent之间的这种联合社会性行为。例如,什么是agent间的联合社会性行为?它如何影响agent的动作选择和执行、实现agent之间的合作?它与agent内部状态之间有什么样的关系等等。

为了刻画agent的自主性、社会性和反应性等特征,在人工智能领域,人们将agent视为由各种认知部件构成的认知系统,代表性的工作是agent的BDI体系结构。然而,单个agent的认知部件,如信念(Belief)、期望(Desire)、意图(Intention),不足以刻画和表征多个agent之间的联合社会性行为。一方面,多个agent的联合社会性行为并不是各个agent行为的简单组合;另一方面,单个agent的行为不同于由多个agent组成的社会性行为。因此,需要提供新的抽象概念模型,用来表示和分析多agent系统中的联合社会性行为[10]。

联合意图抽象概念可以有效地支持agent间联合社会性行为的描述和分析。本文介绍和分析了联合意图理论已有研究工作及其局限性和存在的问题;为了系统、全面地刻画多agent系统的社会性行为,提出了2种不同的联合意图:联合实现型意图和联合维护型意图,讨论了它们的内涵和性质;基于多agent系统的逻辑框架,分别给出了联合实现型意图和联合维护型意图新的、更为自然和直观的语义解释,描述和获取了它们的重要属性。

1 联合意图特征讨论

为了指导联合意图理论的研究,我们首先非形式化地分析和讨论一下联合意图概念的内涵和性质。

在多agent系统中,agent间的联合社会性行为不仅用于实现共同任务,而且可能用于维持系统的某种状态。例如,2个机器人agent负责联合将一个物体从一个场地移到另一个场地,为了满足系统的设计约束,要求在移动过程中必须保持物体处于水平状态。显然,系统的上述约束将对agent间的联合社会性行为产生影响,它们广泛存在于agent的合作过程之中,然而这种设计约束用传统的联合意图理论是无法刻画的。因此,我们提出并区分2种联合意图:联合实现型意图和联合维护型意图。agent间的联合实现型意图是指多个agent联合实现某个命题,agent的联合维护性意图是指多个agent通过联合社会性行为来维持某个命题,使之恒成立。如不做特别说明,下文的联合意图泛指这2种意图。

联合意图体现了多个agent的联合行为选择,因而选择性和联合性是联合意图概念的本质属性。直觉地,联合意图概念具有下列重要性质:

(1)选择性。联合意图体现了agent对未来行为的合理选择,它将影响和约束agent的行为,其中联合实现型意图是agent实施联合社会性行为的起因;

(2)联合性。联合意图体现了agent间的联合特征,包括互知和合作2部分内容;

(3)可满足性。agent的联合意图应是可满足的,或者说是可实现和可维护的;

(4)持续性。持续性是联合意图概念的另一重要特征,它是指agent在实施其社会性行为的过程中不会随意地放弃其意图,它体现了agent的某种承诺;

(5)一致性。agent的多个联合意图是相互一致的。不一致

*)本文得到国家自然科学基金60003002、60103009,"973"项目G1999032700的支持。毛新军 博士,副教授,研究方向:agent理论和技术,新型软件开发方法学等。

的联合意图将使得相关 agent 不知道如何根据其意图行事，agent 间的联合意图应与 agent 的内部意图相一致；

(6)非冲突性。agent 的多个联合意图不应是相互冲突的。agent 的2个联合意图是相互冲突的，是指某个联合意图的实现将阻止或障碍另一个联合意图的成功实现；

(7)与信念的一致性。agent 的联合意图与 agent 的信念是相一致的，如果 agent 具有某种联合意图，则 agent 应认为该意图是可实现的，agent 不应既有某种联合意图同时又认为该联合意图是不可满足的。

2 已有研究分析

计算机科学领域和哲学领域的许多学者开展了联合意图的研究工作。代表性工作为 Cohen，Levesque(C&L)等人的研究成果[2]。C&L 扩展了文[1]中有关意图理论，用于描述联合意图概念。C&L 的联合意图理论试图说明多个 agent 在实施联合的社会性行为过程中，它们的信念、期望、意图之间的相互关系。

首先，C&L 定义了弱目标(weak goal)概念。弱目标说明了在什么条件下 agent 保持其目标，以及当目标已经实现或者已经不可能实现时 agent 应采取的行为。

$$WG(x,y,p)=(\neg Bel(x,p)\wedge Goal(x,\Diamond p))\vee(Bel(x,p)\wedge Goal(x,\Diamond MB(x,y,p)))\vee(Bel(x,\Box\neg p)\wedge Goal(x,\Diamond MB(x,y,\Box\neg p)))$$

然后，定义了联合持续性目标概念。联合持续性目标说明了在什么条件下2个 agent 拥有联合目标，以及如何持续保持联合目标。

$$JPG(x,y,p,q)=MB(x,y,\neg p)\wedge MG(x,y,p)\wedge Until(MB(x,y,p)\vee MB(x,y,\Box\neg p)\vee MB(x,y,\neg q),MB(x,y,(MG(x,y,p)\wedge MG(y,x,p))))$$

基于弱目标和联合持续性目标概念，联合意图概念定义为：

$$JI(x,y,a,\mathbf{q})=JPG(x,y,DONE(x,y,Until(DONE(x,y,a),MB(x,y,DOING(x,y,a)))?;a),q)$$

上述定义较为完整地描述了联合意图概念的内涵。但该理论存在以下问题：首先，选择性是联合意图概念的本质属性，上述语义定义基于线性的时序逻辑和可能世界语义模型，不能很好地刻画和揭示联合意图概念的选择性特征；其次，上述语义定义将联合意图的修改策略融入到联合意图概念的语义定义之中，这种处理一方面使得理论不能更好地刻画联合意图概念本身的内涵，另一方面这种嵌套定义使得整个理论变得极为复杂；第三，联合意图的语义基于可能世界语义模型，而可能世界语义模型存在"逻辑无所不晓"问题；最后，agent 间的社会性行为具有多样性，不仅包括联合实现某个任务，还可能包括联合维持某种状态，根据联合意图概念的上述语义定义，它仅仅揭示前一种情况，而不能描述和分析后一种情况。

这一领域的研究还包括 Nunes，Raimo，Jennings 等人的工作[3,4,10]，它们的工作有些是对 C&L 工作的完善和扩充，有些是从哲学角度非形式化地探讨了联合意图概念的内涵和属性。由于篇幅限制，在此不对它们的工作一一作出分析和评论。

3 逻辑框架

我们将基于多 agent 系统的逻辑框架来定义和分析联合意图概念。逻辑框架包括语法、模型和语义解释3个部分。

形式化语言 L 是对命题分枝时序逻辑 CTL^*[7]的扩充。语言 L 由状态公式集 L_s 和路径公式集 L_r 组成。设 Φ 是原子命题符号集合，$Const_{ag}$ 是 agent 符号集合。为简化说明，有下列符号约定：p,q,…为原子命题；$\varphi,\psi,\cdots$ 为公式；x,y,…为 agent。

定义3.1(语言 L 的语法) 形式化语言 L 是由下列规则定义的最小封闭集合：

(1)如果 $p\in\Phi$，则 $p\in L_s$

(2)如果 $\psi,\varphi\in L_s$ 且 $x,y\in Const_{ag}$，则 $\neg\varphi,\psi\wedge\varphi,Bel(x,\varphi),MB(x,y,\varphi),AI(x,y,\varphi),MI(x,y,\varphi),MAI(x,y,\varphi),MAB(x,y,\varphi),WAC(x,y,\varphi),MAC(x,y,\varphi),JAI(x,y,\varphi),MMI(x,y,\varphi),MMB(x,y,\varphi),WMC(x,y,\varphi),MMC(x,y,\varphi),JMI(x,y,\varphi),JI(x,y,\varphi,\psi)\in L_s$

(3)$L_r\in L_s$

(4)如果 $\psi,\varphi\in L_s$，$x\in Const_{ag}$，则 $\neg\varphi,\psi\wedge\varphi,\psi\ Until\ \varphi,\psi\ Until_Y\ \varphi\in L_s$

(5)如果 $\varphi\in L_s$，则 $A_\varphi\in L_r$

语言 L 的一个模型是 $M=(T,<,U_{ag},\pi,[\,],B,C)$。T 是时刻集，T 中的每一时刻对应于世界的一个状态；$<$ 是 T 上的偏序关系，它描述了时刻间的先后次序。任一时刻的过去是确定和线性的，它的将来可能是分枝的，形式化模型呈图1所示的树形结构。U_{ag} 是 agent 集合。$\pi:\Phi\to\mathcal{G}(T)$，$\mathcal{G}$ 是幂集符号，$\pi(p)$ 定义了使原子公式 p 成立的时刻集。[]是对 agent 符号的赋值。$B:U_{ag}\to\mathcal{G}(T\times T)$，$(t,t')\in B(x)$ 是指 $agent_x$ 在 t 时刻认为 t′时刻是可能的，B 用于定义 agent 的信念。

时刻 t 的一条路径是指始于该时刻，由 t 的将来时刻构成的一条线性分枝，它刻画了世界的某种发展轨迹。

定义3.2(路径) 时刻 t 的一条路径是指集合 $S\subseteq T$ 且满足：(1)$t\in S$；(2)$\forall t_1,t_2\in S:(t_1<t_2)\vee(t_2<t_1)\vee(t_1=t_2)$；(3)$\forall t_1,t_2\in S;t_3\in T:(t_1<t_3<t_2)\Rightarrow(t_3\in S)$；(4)$\forall t_1\in S;t_2\in T:(t_1<t_2)\Rightarrow(\exists t_3\in S:(t_1<t_3)\wedge\neg(t_3<t_2))$；(5)$\forall t_1\in S:(t=t_1)\vee(t<t_1)$

其中，(1)表示时刻 t 的路径包含该时刻；(2)刻画了路径的线性特征；(3)指出了路径的稠密性；(4)描述了路径的相对最大性；(5)刻画了路径的初始性。设 S_t 表示时刻 t 的所有路径的集合，S_Σ 是所有路径的集合。

在多 agent 系统中，各个 agent 的动作并发、异步地发生。在任一时刻，agent 可能执行各种动作并通过动作的执行来影响和控制世界的发展。然而，这种影响和控制是有限的，世界发展的轨迹还受其它 agent 动作执行事件的影响，所有 agent 动作执行事件和环境事件共同确定世界的发展。考虑图1所示的由2个 agent 构成的多 agent 系统的形式化模型。图中的结点表示时刻，边表示多个 agent 的动作并发地发生。假设"‖"左侧符号表示 $agent_1$ 的动作，右侧符号表示 $agent_2$ 的动作。在 t_0 时刻 $agent_1$ 通过执行动作 a 使得世界沿 t_1 或 t_2 方向发展，但世界发展的将来时刻是 t_1 还是 t_2 还取决于 $agent_2$ 的动作。当 $agent_2$ 执行动作 c 时，则世界沿 t_1 方向发展，当 $agent_2$ 执行动作 d 时，世界沿 t_2 方向发展。

$C:U_{ag}\times T\to\mathcal{G}(S_\Sigma)$，$S\in C(x,t)$ 是指在 t 时刻 $agent_x$ 选择的路径，因而有 $C(x,t)\subseteq S_t$，C 用于定义联合意图概念。为简化研究，在模型中限定：$\forall t\in T,x\in U_{ag}:C(x,t)\neq\varnothing$。

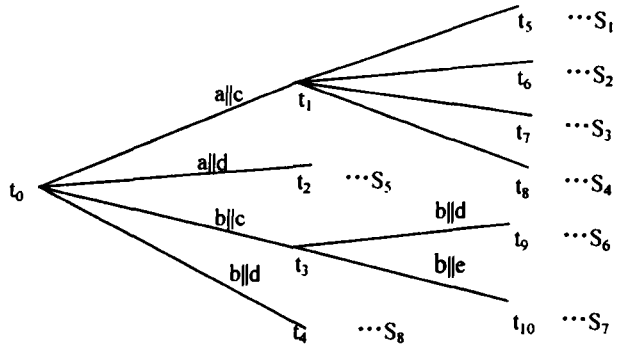

图1 多 agent 系统的形式化模型

L_t 中公式的语义定义由模型 M 和时刻 t 给出。$M\models_t\varphi$ 表示模型 M 在时刻 t 满足公式 φ。L_t 中公式的语义由模型 M、路径 S 和时刻 t 加以定义。$M\models_{s,t}\Psi$ 表示模型 M 在路径 S 的时刻 t 满足公式 Ψ。

定义3.3(语言 L 的语义)

(1)$M\models_t p$ iff $t\in\pi(p)$

(2)$M\models_t\psi\wedge\varphi$ iff $M\models_t\psi$ 且 $M\models_t\varphi$

(3)$M\models_t\neg\varphi$ iff $M\not\models\varphi$

(4)$M\models_t A\varphi$ iff $\forall S: S\in S_t\Rightarrow M\models_{s,t}\varphi$

(5)$M\models_t Bel(x,\varphi)$ iff $\forall t': (t,t')\in B([x])\Rightarrow M\models_{t'}\varphi$

(6)$M\models_{s,t}\psi\wedge\varphi$ iff $M\models_{s,t}\psi$ 且 $M\models S_{s,t}\varphi$

(7)$M\models_{s,t}\neg\varphi$ iff $M\not\models_{s,t}\varphi$

(8)$M\models_{s,t}\psi\ Until\varphi$ iff $\exists t'\in S: (t\leqslant t')$且$(M\models_{s,t'}\varphi)$且$(\forall t'': t\leqslant t''<t'\Rightarrow M\models_{s,t''}\psi)$

(9)$M\models_{s,t}\psi Until_V\varphi$ iff $\forall t'\in S: (\forall t'': t\leqslant t''\leqslant t'\Rightarrow M\models_{s,t''}\neg\varphi)\Rightarrow M\models_{s,t'}\psi$

(10)$M\models_{s,t}\varphi$ iff $M\models_t\varphi$, 其中 $\varphi\in L_t$

(11)$M\models_t MB(x,y,\varphi)$iff $M\models_t Bel(x,\varphi)$且 $M\models_t Bel(y,\varphi)$

根据上述语义定义,我们可以派生出其它命题连接词和算子。*Until* 是"until"算子,$Until_V$ 是弱"until"算子。$F\varphi=$ true *Until* φ, F 是存在时序量词。G 是 F 的对偶算子,即 $G\varphi=\neg F(\neg\varphi)$,因而 G 是全称时序量词。A 是全称路径算子,$A\varphi$ 在某一时刻 t 成立,当且仅当 φ 在该时刻的所有路径上均成立。E 是 A 的对偶算子,即 $E\varphi=\neg A(\neg\varphi)$,因而 E 是存在路径算子。$Bel(x,\varphi)$表示 agent 具有信念 φ,我们假定 B(x)满足自反性和传递性,因而 *Bel* 对应于 S4系统中的正规模态算子。

定理3.1 算子 *Bel* 有以下性质:(1)$\models Bel(x,\varphi)\rightarrow\varphi$;(2)$\models Bel(x,\varphi)\rightarrow Bel(x,Bel(x,\varphi))$;(3)$\models Bel(x,\varphi)\wedge Bel(x,\varphi\rightarrow\psi)\rightarrow Bel(x,\psi)$;(4)如果$\models\varphi$,则$\models Bel(x,\varphi)$。

4 联合意图理论框架

我们将通过扩展文[1]中的意图理论建立多 agent 系统联合意图的理论框架,说明在多 agent 系统中 agent 间的联合意图与个体 agent 的信念和意图之间的相互关系。

4.1 基本概念的定义

agent 的实现型意图是指 agent 意图实现某个命题,它对应于 agent 的任务和目标。多 agent 系统的形式化模型是一个树形的分枝结构,模型中任一时刻的分枝表示 agent 在该时刻所具有的各种可能选择。agent 具有实现型意图 φ 是指 agent 对世界发展轨迹(即路径)的选择,在这些被选择的世界发展轨迹中,agent 知道 φ 将最终成立。

定义4.1.1 (实现型意图的语义) $M\models_t AI(x,\varphi)$iff $M\models_t Bel(x,\neg\varphi)$;且$(\forall S: S\in C(x,t)\Rightarrow M\models_{s,t}FBel(x,\varphi))$。

上述语义定义揭示了实现型意图的最本质特性即选择性。不同于已有方法,我们没有基于可能世界间的可达关系来定义实现型意图概念,而是将 agent 意图视为 agent 对世界发展轨迹的选择。在形式化模型中,世界发展轨迹与 agent 的动作是密切相关的,任一世界发展轨迹对应于多 agent 系统中各个 agent 动作执行序列的组合。我们将 agent 的意图解释为 agent 对世界发展轨迹的选择,不仅清晰地刻画了意图概念的选择特征,揭示了在多 agent 系统中 agent 意图与 agent 的行为之间的关系,反映了 agent 意图对 agent 未来动作的影响和限制,而且有助于我们进一步定义维护型意图概念的语义。

agent 的维护型意图是指 agent 意图维持某个命题使之恒成立。不同于实现型意图,我们将 agent 的维护型意图解释为 agent 在实施社会性行为过程中所具有的系统约束。这些约束显式或隐式地存在于 agent 之中,它们将影响和约束 agent 的行为。

定义4.1.2(维护型意图的语义) $M\models_t MI(x,\varphi)$ iff $(\forall S: S\in C(x,t)\Rightarrow M\models_{s,t}GBel(x,\varphi))$

4.2 联合实现型意图

联合意图体现多个 agent 的联合行为选择,选择性和联合性是联合意图概念的本质属性。下面我们将基于这2个本质属性给出联合实现型意图的语义定义,描述和获取联合实现型意图的重要属性。

$agent_x$ 和 $agent_y$ 具有联合实现型意图 φ 的前提条件之一是它们具有共同的选择,即 $agent_x$ 和 $agent_y$ 同时具有实现型意图 φ。

定义4.2.1(共同实现型选择) $MAI(x,y,\varphi)=AI(x,\varphi)\wedge AI(y,\varphi)$

2个 agent 具有共同的选择来实现 φ,并不意味着它们就具有联合实现型意图 φ。在具备共同选择的基础上,$agent_x$ 和 $agent_y$ 要形成联合实现型意图,还必须满足联合特征。agent 间的联合特征包括2方面的内容:意图互知和合作。

定义4.2.2(互知实现型意图) $MAB(x,y,\varphi)=MB(x,y,AI(x,\varphi)\wedge AI(y,\varphi))$

上述概念定义表明,$agent_x$ 和 $agent_y$ 互知意图是指它们都知道双方都具有实现型意图 φ。

为了定义实现型合作概念的语义,我们首先引入"弱实现型合作"概念。

定义4.2.3(弱实现型合作) $WAC(x,y,\varphi)=$
$(Bel(x,\varphi)\wedge\neg(Bel(x,Bel(y,\varphi))\rightarrow AI(x,MB(x,y,\varphi)))\wedge(Bel(x,AG\rightarrow\varphi)\wedge\neg Bel(x,Bel(y,AG\rightarrow\varphi))\rightarrow AI(x,MB(x,y,AG\rightarrow\varphi)))$

$AG\rightarrow\varphi$ 表示对于所有的路径,φ 在这些路径上的任何时刻均不成立。$agent_x$ 关于 φ 与 $agent_y$ 具有弱实现型合作是指:如果 $agent_x$ 知道 φ 已经成立,并且不知道 y 知道 φ 已成立,则 $agent_x$ 意图让双方都互知 φ 已成立;并且,如果 $agent_x$ 知道 φ 已经不可能成立,并且不知道 y 知道 φ 已不可能成立,则 $agent_x$ 意图让双方都互知 φ 已不可能成立。基于弱实现型合作概念,下面我们给出实现型合作概念的语义定义。

定义4.2.4(实现型合作) $M\models_t MAC(x,y,\varphi)$ iff
$(\forall S: S\in C(x,t)\Rightarrow M\models_{s,t}(MB(x,y,WAC(x,y,\varphi)\wedge WAC(y,x,\varphi)))Until\rightarrow AI(x,\varphi))$且$(\forall S: S\in C(y,t)\Rightarrow M\models_{s,t}(MB(x,y,WAC(x,y,\varphi)\wedge WAC(y,x,\varphi)))Until\rightarrow AI(y,\varphi))$

agent$_x$ 和 agent$_y$ 具有实现型合作 φ 是指在联合实现型意图的实施过程中，它们都知道自己会和对方合作，对方也会和自己合作，及至它们放弃实现型意图 φ。

定义4.2.5(联合实现型意图) $JAI(x,y,\varphi)=MAI(x,y,\varphi)\wedge MAB(x,y,\varphi)\wedge MAC(x,y,\varphi)$

上述概念定义表明，agent$_x$ 和 agent$_y$ 具有联合实现型意图 φ 当且仅当它们具有共同的选择实现 φ，它们相互知道双方都有这样的选择，并且在实施联合实现型意图的过程中，双方都互信会有必要的合作。联合实现型意图的上述语义定义清晰、简洁、准确地描述了联合实现型意图的本质特征：选择性和联合特征。基于该语义定义，我们可以获取联合实现型意图的重要属性。

定理4.2.1 $\models JAI(x,y,\varphi)\rightarrow MB(x,y,\neg\varphi)$

这一定理揭示了 agent 接受和放弃联合实现型意图的条件之一。agent$_x$ 和 agent$_y$ 具有联合实现型意图 φ 仅当 agent$_x$ 和 agent$_y$ 在该时刻认为 φ 不成立。理性 agent 不会试图去联合实现那些已经成立的命题。可根据 JAI、MB 和 AI 的语义定义证明该定理。

定理4.2.2(联合实现型意图的一致性) $\models\neg(JAI(x,y,\varphi)\wedge JAI(x,y,\neg\varphi))$

上述定理表明，agent 的联合实现型意图是一致的。在任意时刻，agent 不可能既有联合实现型意图 φ 同时又有联合实现型意图 $\neg\varphi$。可采用反证法，根据定理4.2.1和定理3.1的结论来证明该定理。

定理4.2.3(与 agent 实现型意图的一致性) $\models\neg(JAI(x,y,\varphi)\wedge(AI(x,\neg\varphi)\vee AI(y,\neg\varphi)))$

上述定理表明，agent 的联合实现型意图与 agent 的实现型意图是一致的。在任意时刻，agent 不可能既有联合实现型意图 φ 同时又有实现型意图 $\neg\varphi$。可根据定理4.2.1的结论，结合 JAI、MB 和 AI 的语义定义来证明该定理。

定理4.2.4(联合实现型意图的可满足性) $\models JAI(x,y,\varphi)\rightarrow MB(x,y,EF\varphi)$

$EF\varphi$ 表示存在一条路径，在该路径上 φ 将最终成立。agent 的联合实现型意图应是可满足的，或者说是可实现的。如果 agent$_x$ 和 agent$_y$ 具有联合实现型意图 φ，则 agent$_x$ 和 agent$_y$ 相互认为 φ 是可实现的。可根据 JAI、MAB、AI 的语义定义以及第3节中的模型约束 $\forall t\in T, x\in U_{ag}: C(x,t)\neq\varnothing$，结合定理3.1的结论来证明该定理。

定理4.2.5(与信念的一致性) $\models\neg(JAI(x,y,\varphi)\wedge(Bel(x,\neg EF\varphi)\vee Bel(y,\neg EF\varphi)))$

上述定理指出，agent 的联合实现型意图与 agent 信念是相一致的。agent 不可能既有联合实现型意图 φ 同时又认为 φ 是不可能实现的，可根据定理4.2.4和定理3.1的结论证明该定理。

定理4.2.6(非冲突性) $\models JAI(x,y,\varphi)\wedge JAI(x,y,\psi)\rightarrow MB(x,y,E(F\varphi\wedge F\psi))$

上述定理揭示了 agent 多个联合实现型意图之间的非冲突性。如果 agent$_x$ 和 agent$_y$ 具有联合实现型意图 φ，同时具有联合实现型意图 ψ，则双方都知道存在某一世界发展轨迹，在该世界发展轨迹上 φ 和 ψ 以某种时序关系得以实现。可根据 JAI、MAB、AI 概念定义以及定理3.1的结论来证明该定理。

持续性是联合实现型意图的一个重要特征，联合实现型意图具有以下持续性公理：

公理4.2.1(联合实现型意图的持续性公理) $A(JAI(x,y,\varphi)\rightarrow JAI(x,y,\varphi)\ Until_y\ (MB(x,y,\varphi)\vee MB(x,y,\rightarrow EF\varphi)))$

上述公理指出，如果 agent$_x$ 和 agent$_y$ 具有联合实现型意图 φ，则 agent 将持续性地拥有该联合实现型意图及至它们知道 φ 已成立或已经不可能成立。为了使上述公理是可靠的，我们对形式化模型做以下约束：$\forall t\in T; S\in S_2; x,y\in U_{ag}: M\models_t JAI(x,y,\varphi)\Rightarrow(\forall t'\in S:(\forall t'': t\leqslant t''\leqslant t'\Rightarrow M\models_{t''}\neg(MB(x,y,\varphi)\vee MB(x,y,\neg EF\varphi)))\Rightarrow M\models_{t'} JAI(x,y,\varphi))$

4.3 联合维护型意图

agent$_x$ 和 agent$_y$ 具有联合维护型意图 φ 的前提条件之一是它们具有共同的选择，即 agent$_x$ 和 agent$_y$ 同时具有维护型意图 φ。

定义4.3.1(共同维护型选择) $MMI(x,y,\varphi)=MI(x,\varphi)\wedge MI(y,\varphi)$

2个 agent 具有共同的选择来维护 φ，并不意味着它们具有联合维护型意图 φ。在具备共同选择的基础上，agent$_x$ 和 agent$_y$ 要形成联合维护型意图，还必须满足联合特征，即意图互知和合作。

定义4.3.2(互知维护型意图) $MMB(x,y,\varphi)=MB(x,y,MI(x,\varphi)\wedge MI(y,\varphi))$

上述概念定义表明，agent$_x$ 和 agent$_y$ 互知意图是指它们都知道双方都具有维护型意图 φ。

为了定义维护型合作概念的语义，我们首先引入“弱维护型合作”概念。

定义4.3.3(弱维护型合作) $WMC(x,y,\varphi)=(Bel(x,\neg\varphi)\wedge\neg Bel(x,Bel(y,\neg\varphi))\rightarrow AI(x,MB(x,y,\neg\varphi)))\wedge(Bel(x,AF\neg\varphi)\wedge\neg Bel(x,Bel(y,AF\neg\varphi))\rightarrow AI(x,MB(x,y,AF\neg\varphi)))$

$AF\neg\varphi$ 表示在所有路径上，$\neg\varphi$ 将最终成立。agent$_x$ 关于 φ 与 agent$_y$ 具有弱维护型合作是指，如果 agent$_x$ 知道 $\neg\varphi$ 已经成立，并且不知道 y 知道 $\neg\varphi$ 已成立，则 agent$_x$ 意图让双方都互知 $\neg\varphi$ 已成立；并且，如果 agent$_x$ 知道 φ 已经不可能始终维护其成立，并且不知道 y 知道 φ 已不可能始终维护其成立，则 agent$_x$ 意图让双方都互知 φ 已不可能始终维护其成立。基于弱维护型合作概念，下面我们给出维护型合作概念的语义定义。

定义4.3.4(维护型合作) $M\models_t MMC(x,y,\varphi)$ iff $(\forall S: S\in C(x,t)\Rightarrow M\models_{s,t}(MB(x,y,WMC(x,y,\varphi)\wedge WMC(y,x,\varphi)))Until\ \neg MI(x,\varphi))$ 且 $(\forall S: S\in C(y,t)\Rightarrow M\models_{s,t}(MB(x,y,WMC(x,y,\varphi)\wedge WMC(y,x,\varphi)))Until\rightarrow MI(y,\varphi))$

agent$_x$ 和 agent$_y$ 具有维护型合作 φ 是指在联合维护型意图的实施过程中，它们都知道自己会和对方合作，对方也会和自己合作，及至它们放弃维护型意图 φ。

定义4.3.5(联合维护型意图) $JMI(x,y,\varphi)=MMI(x,y,\varphi)\wedge MMB(x,y,\varphi)\wedge MMC(x,y,\varphi)$

上述语义定义表明，agent$_x$ 和 agent$_y$ 具有联合维护型意图 φ 当且仅当它们具有共同的选择维护 φ，它们相互知道双方都有这样的选择，并且在实施联合维护型意图的过程中，双方都互信会有必要的合作。联合维护型意图的上述语义定义清晰、简洁、准确地描述了联合维护型意图的本质特征。基于该语义定义，我们可以获取联合维护型意图的重要属性。

定理4.3.1 $\models JMI(x,y,\varphi)\rightarrow MB(x,y,\varphi)$

这一定理揭示了 agent 接受和放弃联合维护型意图的条

件之一。$agent_x$ 和 $agent_y$ 具有联合维护型意图 φ 仅当 $agent_x$ 和 $agent_y$ 在该时刻都认为 φ 已成立。可根据 *JMI*、*MMB* 和 *MI* 概念的定义来证明该定理。

定理4.3.2(联合维护型意图的一致性) $\models\neg(JMI(x,y,\varphi)\wedge JMI(x,y,\neg\varphi))$

上述定理表明,agent 间的联合维护型意图是一致的。在任意时刻,agent 间不可能既有联合维护型意图 φ 同时又有联合维护型意图 $\neg\varphi$。可采用反证法,根据定理4.3.1和定理3.1的结论来证明该定理。

定理4.3.3 (与 agent 维护型意图的一致性) $\models\neg(JMI(x,y,\varphi)\wedge(MI(x,\neg\varphi)\vee MI(y,\neg\varphi)))$

上述定理表明,agent 的联合维护型意图与 agent 的维护型意图是一致的。在任意时刻,agent 不可能既有联合维护型意图 φ 同时又有维护型意图 $\neg\varphi$。可根据定理4.3.1和定理3.1的结论,结合 MI 的语义定义来证明该定理。

定理4.3.4 (联合维护型意图的可满足性) $\models JMI(x,y,\varphi)\rightarrow MB(x,y,EG\ \varphi)$

EG φ 表示存在一条路径 φ 恒成立。agent 的联合维护型意图应是可满足的,或者说是可维护的。如果 $agent_x$ 和 $agent_y$ 具有联合维护型意图 φ,则 $agent_x$ 和 $agent_y$ 都认为 φ 是可维护的。可根据 *JMI*、*MMB*、*MI* 概念的定义以及第3节中的模型约束$\forall\ t\in T, x\in U_{ag}: C(x,t)\neq\varnothing$来证明该定理。

定理4.3.5(与信念的一致性) $\models\neg(JMI(x,y,\varphi)\wedge(Bel(x,\neg EG\varphi)\vee Bel(y,\neg EG\varphi)))$

$\neg EG\varphi$ 表示不存在一条路径,在该路径上 φ 恒成立。上述定理指出 agent 的联合维护型意图与 agent 信念是相一致的。agent 不可能既有联合维护型意图 φ 同时又认为 φ 是不可维护的,可根据定理4.3.4和定理3.1的结论来证明该定理。

定理4.3.6(非冲突性) $\models JMI(x,y,\varphi)\wedge JMI(x,y,\psi)\rightarrow EG(\varphi\wedge\psi)$

上述定理揭示了 agent 间多个联合维护型意图间的非冲突性。如果 $agent_x$ 和 $agent_y$ 具有联合维护型意图 φ,同时具有联合维护型意图 ψ,则存在某一世界发展轨迹,在该世界发展轨迹上 φ 和 ψ 恒成立。可根据 *JMI*、*MMB*、*MI* 概念的定义,结合定理3.1来证明该定理。

公理4.3.1(联合维护型意图的持续性公理) $A(JMI(x,y,\varphi)\rightarrow JMI(x,y,\varphi)Until_y(MB(x,y,\neg\varphi)\vee MB(x,y,\neg EG\varphi)))$

上述公理指出如果 $agent_x$ 和 $agent_y$ 具有联合维护型意图 φ,则 agent 将持续性地拥有该联合维护型意图及至它们知道 φ 已不成立或已经不可能恒成立。为了使上述公理是可靠的,我们对形式化模型做以下约束:$\forall\ t\in T;\ S\in S_\Sigma; x,y\in U_{ag}: M\models_t JMI(x,y,\varphi)\Rightarrow(\forall\ t'\in S:(\forall\ t'': t\leqslant t''\leqslant t'\Rightarrow M\models_{t''}\neg(MB(x,y,\neg\varphi)\vee MB(x,y,\neg EG\varphi)))\Rightarrow M\models_{t'}(JMI(x,y,\varphi))$

4.4 联合意图

下面我们给出联合意图概念的形式化语义定义。

定义4.4.1(agent 的联合意图) $JI(x,y,\varphi,\psi)=JAI(x,y,\varphi)\wedge JMI(x,y,\psi)$

根据联合意图的上述语义定义,我们可以获取联合意图的下列属性。

定理4.4.1(联合意图的一致性) $\models\neg JI(x,y,\varphi,\neg\varphi)\wedge\neg JI(x,y,\varphi,\varphi)$

上述定理表明 agent 的联合实现型意图与维护型意图间是一致的。可根据 *JAI* 和 *JMI* 的语义定义来证明该定理。

定理4.4.2(联合意图的可满足性) $\models JI(x,y,\varphi,\psi)\rightarrow MB(x,y,E(\psi\ Until\ \varphi))$

E(ψ Until φ)表示存在一条路径 φ 将最终成立,并且在此过程中 ψ 将得到维护。agent 的联合意图应是可满足的。如果 $agent_x$ 和 $agent_y$ 具有联合意图 φ 和 ψ,则 $agent_x$ 和 $agent_y$ 相互认为 φ 是可实现的并且 ψ 将得到维护。可根据 *JMI*、*JAI*、*MAB*、*AI*、*JMI*、*MMB*、*MI* 概念的定义以及第3节中的模型约束$\forall\ t\in T, x\in U_{ag}: C(x,t)\neq\varnothing$来证明该定理。

定理4.4.3 (与信念的一致性) $\models\neg JI(x,y,\varphi,\psi)\wedge(Bel(x,\neg E(\psi\ Until\ \varphi))\vee Bel(y,\neg E(\psi\ Until\ \varphi))))$

上述定理指出 agent 的联合意图与 agent 信念是相一致的。agent 不可能既有联合意图 φ 和 ψ,同时又认为 φ 是不可能实现并且 ψ 得不到维护,可根据定理4.4.2来证明该定理。

结论 本文阐述了开展联合意图研究的重要性和意义,分析了联合意图理论已有研究工作及其局限性和存在的问题。为了系统、全面地刻画多 agent 系统的社会性行为,提出了2种不同的联合意图:联合实现型意图和联合维护型意图,分析了它们的性质。联合维护型意图是对多 agent 系统设计约束的抽象表示,因而理论框架具有更强的表达能力。基于多 agent 系统的逻辑框架,分别给出了联合实现型意图和联合维护型意图新的、更为自然和直观的语义解释。不同于已有研究,我们没有基于可能世界间的可达关系来定义联合意图概念,而是将 agent 的联合意图视为 agent 对世界发展轨迹的共同选择。联合意图概念的上述语义定义清晰、简洁、准确地描述了联合实现型意图的选择性和联合特征,获取了联合意图的重要属性。联合意图理论可以有效地支持多 agent 系统的分析和设计,尤其是用于研究多 agent 系统的社会性行为。我们将利用该理论框架,描述和验证多 agent 系统的合作模型及其交互行为。

参考文献

1 Cohen P R, Levesque H J. Intention is choice with commitment. Artificial Intelligence, 1990, 42(2-3): 213~261

2 Chen P R, Levesque H J. Teamwork. Nous, 1991, 21

3 Tuomela R. Collective and joint intention. In: Proc. of cognitive theory of social action, 1998

4 Tuomela R. Philosophy and distributed artificial intelligence: The case of joint intention. Foundation of distributed artificial intelligence, John Wiley&Sons, 1996. 487~504

5 毛新军,王怀民. Agent 在多 Agent 系统中计算的意愿理论. 软件学报, 1999, 10(1): 43~48

6 Singh M P. Multiagent System: A Theoretical framework for Intentions, Know-how, and Communications Berlin, Heidelberg: Springer-Verlag, 1994

7 Chellas B. Modal logic: an introduction. Cambridge University Press, Cambridge, MA, 1980

8 Dunin-Keplicz B, Verbrugge R. Collective Commitment. In: Proc. of the second intl. conf. on multi-agent system, Menlo Park, CA, 1996

9 Panzarasa P, Jennings N R. Social mental shaping: Modelling the impact of sociality on the mental state of autonomous agents. Computational Intelligence, 2001

10 Jennings N R. Specification and implementation of belief-desire-joint intention architecture for collaborative problem solving. Journal of Intelligent and Cooperative Information Systems, 1993, 2 (3): 289~318

11 Jennings N R. Building complex software system: the case for an agent-based approach. Communication of ACM, 2001