Agent在多Agent系统中计算的意愿理论*

毛新军 王怀民 陈火旺 刘凤歧

(长沙工学院计算机科学系 长沙 410073)

E-mail: xjmao@nudt.edu.cn

摘要 提出了Agent在多Agent系统中计算的意愿理论,以支持Agent计算的理论研究.区分了两种意愿:实现型意愿和维护型意愿.基于多Agent系统计算的逻辑框架,给出了两种意愿新的语义定义,获取和描述了它们的一些重要逻辑属性.

关键词 Agent,多Agent系统,意向系统,信念,意愿.

中图法分类号 TP18

随着面向Agent程序设计范型[1]的提出,如何获取和分析多Agent系统的需求、规范和设计Agent已成为当前软件工程领域一项重要的研究课题.计算机科学研究的通常做法是寻求抽象的概念和工具来促进问题的解决,如过程抽象、进程、对象等等.在AI领域,人们通常基于意向观点(Intentional Stance)来研究Agent,即将Agent视为由一组认知部件所构成的意向系统.意向观点为我们研究Agent提供了一组高层的抽象认知概念,如信念、意愿等等.基于这些概念,我们可以独立于Agent的内部结构和具体实现细节来构造Agent体系结构,定义Agent状态,分析Agent行为的规律性特征.其中意愿是规范和设计Agent的一个重要抽象认知概念,原因是:(1) Agent的意愿将影响和约束Agent的行为,它是Agent计算的起因;(2) 意愿概念与其他概念(如信念等)是密切相关的,具有更强的约束,并体现了Agent的某些理性特征;(3) 意愿概念能被进一步用于规范Agent之间的交互和通信行为.为了使意愿概念能够有效地规范和描述Agent,指导Agent设计,推动多Agent系统的开发,必须深入地分析意愿概念与其他概念(如信念等)间的关系,形式化地给出其语义定义,获取和描述其逻辑属性.这就需要开展Agent计算的意愿理论研究.

1 意愿概念的非形式化讨论

为了指导意愿理论的研究,我们首先分析意愿与期望两个概念间的区别和联系,讨论意愿概念的性质和含义.意愿、期望这两个概念都刻画了Agent的某种选择特征.早期的一些研究工作试图把二者等同起来或者把意愿概念归结为信念和期望两个概念的组合,然而这些工作被认为是不可行的.[2,3]意愿与期望是两个不同的概念.Bartman指出,为了研究Agent的理性行为,必须在信念、期望的基础上引进一个新的概念——意愿.

与意愿概念相比较,期望概念更弱.主要表现在:(1) Agent的期望可能是不一致的;(2) Agent的期望可能是相互冲突的;(3) Agent的期望可能与其信念是不一致的;(4) Agent的期望可能缺乏持续性特征,即Agent可能随意地终止或放弃其期望;(5)具有某种期望的Agent可能仅仅拥有这种期望,而不会始发某些动作或采取某些手段和策略来实现其期望.期望概念的上述性质表明,这一概念不能很好地刻画Agent的理性特征,尤其是不一致的认知状态将使得Agent不知如何进行计算.

---

* 本文研究得到国家自然科学基金和国防预研基金资助.作者**毛新军**,1970年生,博士,主要研究领域为Agent理论,分布计算技术,软件重用.**王怀民**,1962年生,博士,副教授,主要研究领域为分布计算技术,面向Agent技术,智能软件环境.**陈火旺**,1936年生,教授,博士生导师,中国工程院院士,主要研究领域为软件自动化,计算机科学理论,人工智能.**刘凤歧**,1938年生,教授,主要研究领域为计算机软件,人工智能.

本文通讯联系人:毛新军,长沙410073,长沙工学院计算机科学系

本文1997-09-09收到原稿,1998-02-11收到修改稿

意愿概念的基本内涵是对未来动作的合理选择.选择性是意愿概念的本质属性.我们将区分两种意愿:实现型意愿和维护型意愿,分别表示 Agent 的不同目的.Agent 的实现型意愿是指 Agent 意图实现某个命题;Agent 的维护性意愿是指 Agent 意图维持某个条件.如不作特别说明,下文的意愿是指上述两种意愿.直觉地讲,意愿概念具有下列属性:(1) 意愿体现了 Agent 对未来动作的合理选择.Agent 的意愿将影响和约束 Agent 的行为,其中实现型意愿是 Agent 动作的起因;(2) 可满足性.Agent 的意愿应是可满足的,或者说是可实现和可维护的;(3) 持续性.持续性是意愿概念的另一重要特征.意愿的持续性是指 Agent 将不会随意地放弃其已有意愿,它体现了 Agent 的某种承诺特征,即具有某种意愿的 Agent 将在不断变化的环境中持续性地拥有该意愿.持续性是 Agent 成功地实现其意愿的必要条件;(4) 内部一致性.Agent 的意愿之间必须是相互一致的.不一致的意愿将使得 Agent 不知道如何根据其意愿行事;(5) 非冲突性.Agent 的意愿之间不应是相互冲突的.Agent 的两个意愿是相互冲突的,是指 Agent 某个意愿的实现将阻止或妨碍另一个意愿的成功实现;(6) 与信念的一致性.Agent 的意愿与 Agent 的信念是相一致的,Agent 不应既有某种意愿,同时又认为该意愿是不可满足的.

2 逻辑框架

2.1 形式化语言 L

形式化语言 L 的公式集由状态公式集 L_s 和路径公式集 L_r 两部分组成.设 Φ 是原子命题符号集合,$Const_{Ag}$ 是 Agent 符号集合,$Const_{Ac}$ 是原子动作符号集合.为了简化说明,我们具有下列符号约定:$p,q,\ldots$ 表示原子命题符号;$\varphi,\psi,\ldots$ 表示公式符号;$i,j,\ldots$ 表示 Agent 符号;$a,b,\ldots$ 表示原子动作符号.

定义 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:

- 如果 $p\in\Phi$,则 $p\in L_s$;
- 如果 $\psi,\varphi\in L_s$ 且 $i\in Const_{Ag}$,则 $\neg\varphi,\psi\wedge\varphi,K_i\varphi,A\text{-}Intend_i\varphi,M\text{-}Intend_i\varphi,Intend_i(\varphi,\psi)\in L_s$;
- $L_s\subseteq L_r$;
- 如果 $\psi,\varphi\in L_r$,$i\in Const_{Ag}$ 且 $a\in Const_{Ac}$,则 $\neg\varphi,\psi\wedge\varphi,\psi U\varphi,\langle do_i(a)\rangle\varphi,[do_i(a)]\varphi\in L_r$;
- 如果 $\varphi\in L_r$,则 $A\varphi\in L_s$.

其中 K 是信念算子,$A\text{-}Intend$ 是实现型意愿算子,$M\text{-}Intend$ 是维护型意愿算子,$Intend$ 是意愿算子,$\langle\rangle$ 和 $[]$ 是动作算子,U 是"until"算子,A 是全称路径算子.

2.2 形式化模型和形式化语义

语言 L 的一个模型 M 是指元偶 $<T,<,U_{Ag},U_{Ac},\pi,Act,\blacksquare,B,I>$.$T$ 是时刻集,T 中的每一时刻对应于世界的一个状态(包括物理系统状态、系统中各个 Agent 的认知状态).物理系统状态由在该状态下为真的原子命题来表示.Agent 的认知状态是指 Agent 的信念状态和意愿状态.$<$ 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.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.

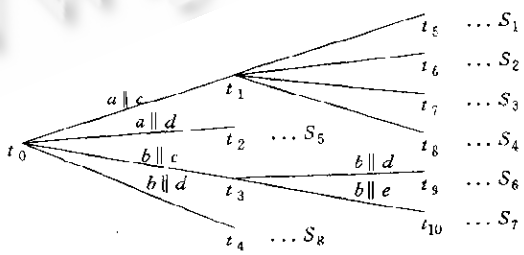

图1 多Agent系统计算的形式化模型

定义 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) $\forall t_1\in S:(t=t_1)\vee(t<t_1)$.

设 S_t 表示时刻 t 所有路径的集合.$S_\sum$是所有路径的集合,即 $S_\sum=\bigcup_{t\in T}S_t$.

定义 3.(路径子区间):设 $t\leqslant t'$,则$[t,t']=\{t''|t\leqslant t''\leqslant t'\}$为一路径子区间.

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,世界发展轨迹还受其他 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 所示的由两个 Agent 构成的形式化模型.图中的结点表示时刻,边表示多个 Agent 的动作并发地发生.我们假定"|"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,则世界沿 t_2 方向发展.在形式化模型中,不同路径对应于不同的 Agent 动作执行事件与环境事件的组合,反映了世界的不同发展轨迹.

U_{Ag}是 Agent 集合.U_{Ac}是原子动作集合.$\pi:\Phi\to\mathscr{P}(T)$,$\mathscr{P}$ 是幂集符号,$\pi(p)$定义了使原子命题 p 成立的时刻集.$Act:U_{Ag}\times U_{Ac}\to\mathscr{P}(T\times T)$定义了动作的发生.$[t,t']\in Act(i,a)$表示 $Agent_i$ 在$[t,t']$路径子区间中执行动作 a,t 是动作执行的起始时刻,t'是终止时刻.【】是对 Agent 符号和原子动作符号的解释,因而它是以下两种类型函数的联合:$(Const_{Ag}\to U_{Ag})$以及$(Const_{Ac}\to U_{Ac})$.$B:U_{Ag}\to\mathscr{P}(T\times T)$,$(t,t')\in B(i)$是指 $Agent_i$ 在 t 时刻认为 t'时刻是可能的.B 用于定义 Agent 的信念.$I:U_{Ag}\times T\to\mathscr{P}(S_\sum)$,$S\in I(i,t)$是指在 t 时刻 $Agent_i$ 选择路径 S,因而有 $I(i,t)\subseteq S_t$.I 用于定义 Agent 的意愿概念.L_t 中公式的可满足语义定义由模型 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 中公式的可满足语义由模型 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 4.(语言 L 的形式化语义)

- $M\models_t p$ iff $t\in\pi(p)$,其中 p 为原子命题;
- $M\models_t\psi\wedge\varphi$ iff $M\models_t\psi$ 且 $M\models_t\varphi$;
- $M\models_t\neg\varphi$ iff $M\not\models_t\varphi$;
- $M\models_t A\varphi$ iff $\forall S:S\in S_t\Rightarrow M\models_{s,t}\varphi$;
- $M\models_t K_i\varphi$ iff $\forall t':(t,t')\in B(【i】)\Rightarrow M\models_t\varphi$;
- $M\models_{s,t}\psi\wedge\varphi$ iff $M\models_{s,t}\psi$ 且 $M\models_{s,t}\varphi$;
- $M\models_{s,t}\neg\varphi$ iff $M\not\models_{s,t}\varphi$;
- $M\models_{s,t}\psi U\varphi$ iff $\exists t'\in S:(t\leqslant t')$且$(M\models_{s,t'}\varphi)$且$(\forall t'':t\leqslant t''<t'\Rightarrow M\models_{s,t''}\psi)$;
- $M\models_{s,t}\langle do_i(a)\rangle\varphi$ iff $\exists t'\in S:[t,t']\in Act(【i】,【a】)$且$(\exists t'':t<t''\leqslant t'$ 且 $M\models_{s,t''}\varphi)$;
- $M\models_{s,t}\langle do_i(a)\rangle\varphi$ iff $\forall t'\in S:[t,t']\in Act(【i】,【a】)\Rightarrow(\exists t'':t<t''\leqslant t'$ 且 $M\models_{s,t''}\varphi)$;
- $M\models_{s,t}\varphi$ iff $M\models_t\varphi$,其中 $\varphi\in L_t$.

根据上述语义定义,我们可以派生出其他命题连接词和算子.U 是"until"时序算子,$F\varphi=\text{true}\,U\varphi$.$G$ 是 F 的对偶算子,即 $G\varphi=\neg F(\neg\varphi)$.$A$ 是全称路径算子.E 是 A 的对偶算子,即 $E\varphi=\neg A(\neg\varphi)$.〈〉和[]是两个动作算子.直觉地讲,$\langle do_i(a)\rangle\varphi$表示 $Agent_i$ 完成动作 a 且具有结果 φ.$[do_i(a)]\varphi$表示,如果 $Agent_i$ 能够完成动作 a,则具有结果 φ.〈〉和[]的上述语义定义要求 φ 在动作执行过程中(而不是在动作完成之时)成立.值得注意的是,[]不是〈〉的对偶算子.$K_i\varphi$表示 Agent 具有信念 φ.我们假定对于任意 $Agent_i$,关系 $B(i)$满足自反性和传递性,因而算子 K_i 具有下列性质.

定理 1. 算子 K_i 具有以下性质:(1) $\models K_i\varphi\to\varphi$;(2) $\models K_i\varphi\wedge K_i(\varphi\to\psi)\to K_i\psi$;(3) $\models K_i\varphi\to K_iK_i\varphi$;(4) 如果$\models\varphi$,则$\vdash K_i\varphi$.

3　实现型意愿

Agent 的实现型意愿是指 Agent 意图实现某个命题,它对应于 Agent 的任务和目标.形式化模型 M 呈一树形结构.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的选择.直觉地讲,Agent 具有实现型意

愿 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,φ 将最终成立.

定义 5. $M\models_t A\text{-}Intend_i\varphi$ iff $M\models_t\neg\varphi$ 且 $(\forall S: S\in I(i,t)\Rightarrow M\models_{t,S}F\varphi)$.

上述语义定义揭示了实现型意愿最本质的特性即选择性.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意愿概念,而是将 Agent 意愿视为 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.Agent 通过执行动作,在一定程度上来影响和控制世界的发展.我们将 Agent 的意愿解释为 Agent 对世界发展轨迹的选择,不仅清晰地刻画了意愿概念的选择特征,揭示了在多 Agent 系统中 Agent 意愿与 Agent 的未来行为之间的关系,而且有助于我们获取意愿概念的一些重要属性,并可以帮助我们进一步定义维护型意愿的形式化语义.

定理 2. $\models A\text{-}Intend_i\varphi\rightarrow\neg\varphi$.

这一定理揭示了 Agent 接受和放弃实现型意愿的条件.Agent 在某一时刻具有实现型意愿 φ 仅当 φ 在该时刻不成立.理性 Agent 不会试图去实现那些已经成立的命题或条件.

定理 3. $\models\neg(A\text{-}Intend_i\varphi\wedge A\text{-}Intend_i(\neg\varphi))$.

上述定理表明 Agent 的实现型意愿是一致的.在任意时刻,Agent 不可能既有实现型意愿 φ 同时又有实现型意愿 $\neg\varphi$.可根据定理 2 的结论来证明该定理.

Agent 的实现型意愿应是可满足的,或者说是可实现的,因而实现型意愿具有以下公理:

公理 1.(实现型意愿的可满足公理)$A\text{-}Intend_i\varphi\rightarrow E F\varphi$.

公理 1 指出,如果 Agent 具有实现型意愿 φ,则存在某一世界发展轨迹,在该世界发展轨迹上,φ 必然成立.为了使上述公理是可靠的,我们对形式化模型作以下约束:

模型约束 1. $\forall t\in T; i\in U_{Ag}: I(i,t)\neq\varnothing$.

定理 4. $\models\neg(A\text{-}Intend_i\varphi\wedge K_i(\neg F F\varphi))$.

上述定理指出 Agent 的实现型意愿与 Agent 的信念是一致的.Agent 不可能既有实现型意愿 φ,同时又认为 φ 是不可能实现的.可根据公理 1 以及定理 1 来证明该定理.为了指导 Agent 计算,Agent 必须知道其实现型意愿,即实现型意愿具有以下公理:

公理 2.(实现型意愿的自省公理)$A\text{-}Intend_i\varphi\rightarrow K_i(A\text{-}Intend_i\varphi)$.

为了使上述公理是可靠的,我们对形式化模型作了以下约束:

模型约束 2. $\forall t\in T; i\in U_{Ag}: M\models_t A\text{-}Intend_i\varphi\Rightarrow(\forall t': (t,t')\in B(i)\Rightarrow M\models_{t'}A\text{-}Intend_i\varphi)$.

定理 5. $\models A\text{-}Intend_i\varphi\wedge A\text{-}Intend_i\psi\rightarrow E(F\varphi\wedge F\psi)$.

上述定理揭示了 Agent 多个实现型意愿间的非冲突性.如果 Agent 在某一时刻具有多个实现型意愿,则存在某一世界发展轨迹,在该世界发展轨迹上,Agent 以某种时序关系实现这些意愿.可根据实现型意愿的语义定义来证明该定理.在 AI 领域,逻辑"无所不晓"问题以及由此而带来的其他相关问题(如副作用问题、传递问题等)给认知概念的理论研究带来了困难,因为在形式化认知概念的过程中,正规模态逻辑能够产生一些人们所不期望的属性.考虑下列在正规模态逻辑系统中为真的命题.

定义 6. 设 φ 和 ψ 为命题,X 为表示 Agent 某种认知概念(如信念、意愿等)的模态词.

(LO1) $X\varphi\wedge X(\varphi\rightarrow\psi)\rightarrow X\psi$;

(LO2) 如果 $\models\varphi$,则 $\models X\varphi$;

(LO3) 如果 $\models\varphi\rightarrow\psi$,则 $\models X\varphi\rightarrow X\psi$;

(LO4) 如果 $\models\varphi\leftrightarrow\psi$,则 $\models X\varphi\leftrightarrow X\psi$;

(LO5) $(X\varphi\wedge X\psi)\rightarrow X(\varphi\wedge\psi)$;

(LO6) $X\varphi\rightarrow X(\varphi\vee\psi)$;

(LO7) $\neg(X\varphi\wedge X\neg\varphi)$.

(LO1)和(LO3)构成了正规模态逻辑系统的副作用问题,即 Agent 认知状态封闭于逻辑结果.(LO2)构成了正规模态逻辑系统的传递问题.这些属性是对理想化 Agent 的一种刻画.实际的 Agent 是资源受限的计算实体.上述许多属性是反直觉的,例如,(LO2)指出,如果 φ 为一永真式,则 Agent 具有意愿 φ,这与意愿的直觉性认识

是不一致的.直觉地讲,Agent 具有某种意愿 φ 是指 Agent 意图通过自身的努力来实现 φ.如果 φ 是不可避免的,或者 φ 已成立,则 Agent 就不应具有该意愿.(LO5)并不是我们所需要的属性,它表示一种不受限制的组合.Agent 意图实现 φ 且 Agent 意图实现 ψ 并不意味着 Agent 意图同时实现 φ 和 ψ.(LO6)表示不受限制的变弱,同样亦不是我们所期望的.Agent 意图实现 φ 并不意味着 Agent 意图实现 φ 或 ψ.属性(LO4)对于我们的研究并无任何坏处,而且(LO7)是我们所期望的,因为这一属性能确保 Agent 的意愿是一致的.

定理 6. 在定义 6 所列的各种属性中,(LO2),(LO3),(LO5),(LO6)对于实现型意愿算子 *A-Intend* 并不成立.

上述定理的证明比较简单.属性(LO1)所以成立是由于前提不可能成立,即公式 $A\text{-}Intend_i\varphi$ 和 $A\text{-}Intend_i(\varphi\rightarrow\psi)$ 不可能同时成立.

4　维护型意愿

Agent 的维护型意愿是指 Agent 意图维持某个条件,使之恒成立.与实现型意愿不同,我们将 Agent 的维护型意愿解释为在多 Agent 系统中 Agent 必须遵循的社会规则(Social Rule).在多 Agent 系统中,每个 Agent 并非是完全自主和独立的,且其资源和能力都是有限的,它们构成了一个社会.为了使多 Agent 系统是有序的,能够有效地促进 Agent 间的协同与合作、减少冲突,多 Agent 系统可能具有相应的社会规则.Agent 处于这样的社会环境中,因而必定受社会规则的约束和限制.[4]社会规则对 Agent 而言是外在、具体的,为了指导 Agent 计算以及规范和描述 Agent,必须对它们进行抽象和内部化.例如,在某一环境中存在一组机器人 Agent.为了减少冲突,这些机器人的设计人员在开发它们的过程中提出了一组它们必须遵循的社会规则:每个机器人在移动过程中必须靠左行走.

定义 7. $M\models_t M\text{-}Intend_i(\varphi)$ iff $(\forall\ S:S\in I(i,t)\Rightarrow M\models_{t,t}G\varphi)$.

根据维护型意愿的上述语义定义,我们可以获取维护型意愿的一系列逻辑属性.Agent 的维护型意愿应是可满足的或者说是可维护的,因而维护型意愿具有下列公理:

公理 3. (维护型意愿的可满足公理)$M\text{-}Intend_i(\varphi)\rightarrow E G\varphi$.

公理 3 指出,如果 Agent 具有维护型意愿 φ,则存在某一世界发展轨迹,在该世界发展轨迹 φ 上恒成立.基于模型约束 1,上述公理是可靠的.

定理 7. $\models\neg(M\text{-}Intend_i(\varphi)\wedge M\text{-}Intend_i(\neg\varphi))$.

上述定理刻画了 Agent 维护型意愿的一致性.在任意时刻,Agent 不可能既有维护型意愿 φ 同时又有维护型意愿 $\neg\varphi$.可根据维护型意愿的语义定义来证明该定理.

定理 8. $\models\neg(M\text{-}Intend_i(\varphi)\wedge K_i(\neg E G\varphi))$.

上述定理表明,Agent 的维护型意愿与 Agent 的信念是一致的.Agent 不可能具有维护型意愿 φ,同时又认为 φ 不可能恒成立.可根据公理 3 和定理 1 来证明该定理.为了指导 Agent 计算,Agent 必须知道其维护型意愿,即维护型意愿具有以下公理:

公理 4. (维护型意愿的自省公理)$M\text{-}Intend_i\varphi\rightarrow K_i(M\text{-}Intend_i\varphi)$.

为了使上述公理是可靠的,我们对形式化模型作了以下约束:

模型约束 3. $\forall\ t\in T;i\in U_{Ag}:M\models_t M\text{-}Intend_i\varphi\Rightarrow(\forall\ t':(t,t')\in B(i)\Rightarrow M\models_{t'}M\text{-}Intend_i\varphi)$.

5　Agent 的意愿

定义 8. $Intend_i(\varphi,\psi)\xlongequal{\text{def}}A\text{-}Intend_i(\varphi)\wedge M\text{-}Intend_i(\psi)$.

定理 9. $\models\neg Intend_i(\varphi,\neg\varphi)\wedge\neg Intend_i(\varphi,\varphi)$.

上述定理表明,Agent 的实现型意愿与维护型意愿是一致的.Agent 不可能既有维护型意愿 φ,又有实现型意愿 φ.同时,Agent 也不可能既有维护型意愿 $\neg\varphi$,又有实现型意愿 φ.可根据实现型意愿和维护型意愿的语义定义来证明该定理.

定义 9. (弱 until 算子的语义定义)$M\models_{s,t}\psi U_{\forall}\varphi$ iff $\forall\, t'\in S:(\forall\, t'':t\leqslant t''\leqslant t'\Rightarrow M\models_{s,t''}\neg\varphi)\Rightarrow M\models_{s,t'}\psi$.

持续性是意愿区别于其他概念的一个重要特征,亦是 Agent 成功地实现其意愿的基础和前提.意愿的持续性,是指在不断变化的环境中 Agent 将保持其意愿.为了指导 Agent 计算,确保 Agent 成功地实现和维护其意愿,Agent 的意愿具有以下持续性公理:

公理 5. $A(Intend_i(\varphi,\psi)\rightarrow Intend_i(\varphi,\psi)U_{\forall}(\neg\psi\vee\varphi\vee\neg E(F\varphi\wedge G\psi)))$.

公理 5 指出,如果 Agent 具有意愿 φ 和 ψ,则 Agent 将持续性地拥有该意愿及至 ψ 已不成立,或 φ 已成立或不存在路径使得 φ 在该路径上成立且 ψ 在该路径上恒成立.为了使上述公理是可靠的,我们对形式化模型作了以下约束:

模型约束 4. $\forall\, t\in T;S\in S_t;i\in U_{Ag}:M\models_t Intend_i(\varphi,\psi)\Rightarrow$

$$(\forall\, t'\in S:(\forall\, t'':t\leqslant t''\leqslant t'\Rightarrow M\models_{s,t''}\neg(\neg\psi\vee\varphi\vee\neg E(F\varphi\wedge G\psi)))$$

$$\Rightarrow(\forall\, S':S'\in I(i,t')\Rightarrow M\models_{s,t'}F\varphi\wedge G\psi)).$$

持续性的意愿是 Agent 成功地实现某个命题、维护某些条件的必要条件,但不是充分条件.Agent 成功地实现其意愿还必须满足其他条件,如 Agent 必须具有足够的知识和能力;Agent 必须根据其意愿行事等等.

6 结 论

本文提出了 Agent 在多 Agent 系统计算的意愿理论以规范和描述 Agent、支持 Agent 计算的理论研究.我们区分两种意愿:实现型意愿和维护型意愿,分别表示 Agent 的不同目的.实现型意愿对应于 Agent 的任务和目标,维护型意愿则被解释为 Agent 必须遵循的社会规则.本文分析了意愿与期望两个概念的区别和联系,讨论了意愿概念的性质和含义,给出了意愿概念新的语义定义,获取和描述了它们的一些重要逻辑属性.

参考文献

1 Shoham Y. Agent-oriented programming. Artificial Intelligence, 1993,60(1):51～92

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

3 Singh M P. Multi-agent System: a Thecretical Framework for Intentions, Know-how, and Communications. Berlin, Heidelberg: Springer-Verlag, 1994

4 Shoham Y, Tennenholtz M. On social laws for artificial agent societies: off-line design. Artificial Intelligence, 1995,73(1—2):231～252

The Intention Theory of Agent Computing in Multi-agent System

MAO Xin-jun WANG Huai-min CHEN Huo-wang LIU Feng-qi

(Department of Computer Science Changsha Institute of Technology Changsha 410073)

Abstract Intention is an important abstract concept to specify and design agent. In this paper, an intention theory of agent computing in multi-agent system is presented to support the research on the theory of agent computing. Two intentions: achievement intention and maintenance intention, respectively denoting the different aims of agent are differentiated. The formal semantics of two intentions are defined and some important properties are obtained based on the logic framework of multi-agent system computing.

Key words Agent, multi-agent system, intentional system, belief, intention.