一种理性 Agent 的 BDI 模型*

康小强　石纯一

(清华大学计算机科学与技术系　北京　100084)

摘要　该文通过引入假设信念,解释愿望和意图在 Agent 思维状态的认知方面的含义,进而定义愿望和意图,并引入规划,建立理性 Agent 的动态 BDI 模型.与 Cohen 和 Levesque,Rao 和 Georgeff,Konolige 和 Pollack 等人的工作相比,克服了对信念、愿望和意图的反直观解释问题,解决了关于愿望和意图的无为而治和副作用问题,强调了愿望的激发与维护作用,表达了信念、愿望和意图三者间的动态约束与激发关系.

关键词　理性 Agent,思维状态,信念,愿望,意图.

中图法分类号　TP18

Agent 模型是分布式人工智能和多 Agent 系统研究的重要组成部分.Hewitt 曾经提出基于 ACTOR 模型的开放信息系统语义,希望以此为分布式人工智能建立理论基础[1].随着研究的深入,面向细粒度并发计算的 ACTOR 模型难以适应对 Agent 的个体智能性和群体交互性的要求.特别是当自主性成为研究的基本出发点之后,Agent 必须适应环境的变化和群体交互的变化.于是,将Agent视为具有意图的智能系统,建立基于思维状态(mental state)的 Agent 模型就成为研究的主流.

借鉴心理学的研究成果,人类的思维状态属性有以下几个方面:(1) 认知,如信念、知识等;(2) 情感,如目标、愿望和偏好等;(3) 意动,如意图,承诺和规划等.相应地,当前的 Agent 模型研究侧重于形式描述信念(belief)、愿望(desire)和意图(intention),简称 BDI,进而向多 Agent 扩展,研究多 Agent 群体的 BDI 模型.

1　相关工作分析

BDI 模型的哲学基础是 Bratman 对理性和意图的分析,刻画了意图的客观性以及在理性平衡中的中心位置[2].分布式人工智能中的 BDI 模型研究着重于 BDI 的形式描述,主要有 Cohen 与 Levesque,Rao 和 Geogeff,Konolige 与 Pollack 以及 Shoham 等人的工作.这些工作一方面表达了 Bratman 对意图和理性的研究,另一方面,从智能系统的定义、构造和应用等角度对意图和理性的含义进行了阐述.

Cohen 和 Levesque 基于正规模态逻辑的可能世界模型,定义了 Agent 信念的基本性质,并将意图定义为一种持续目标,初步描述了意图在 Agent 行为中的作用[3].Rao 和 Georgeff 采用计算树逻辑,可能世界的时间结构由线性扩充为分支,进一步阐述了 BDI 的概念及相互关系[4].Konolige 和 Pollack 提出认知结构和意图关系图的概念,由此定义的意图不仅包含对目标世界的期待,也包含对不希望的世界状态的描述,在一定程度上解决了副作用问题[5].

在这些模型的基础上,Shoham 从系统实现角度讨论了 Agent 的结构和行为特性,提出了面向 Agent 的程序设计[6].Haddadi 分析了联合承诺,用以描述合作推理和协商[7].Dunin-Keplicz 和 Verbrugge 也就多 Agent 的承诺问题作了探讨,分析了 3 种社会承诺机制[8].Castelfrachi 则从社会行为角度研究了社会承诺[9].

尽管在实用性和社会性方面有了这些进展,但在 BDI 模型中的一些原有的问题并没有得到解决.

---

* 本文研究得到国家自然科学基金资助.作者康小强,1968 年生,博士生,主要研究领域为分布式人工智能,智能 Agent.石纯一,1935 年生,教授,博士生导师,主要研究领域为人工智能应用基础,知识工程.

本文通讯联系人:康小强,北京 100084,清华大学计算机科学与技术系

本文 1998-10-14 收到原稿,1998-12-28 收到修改稿

首先是逻辑全知问题：$\varphi \Rightarrow X\varphi$. X 是一个算子，如信念算子 BEL、愿望算子 DES 和意图算子 INT. 逻辑全知不仅对有限资源的 Agent 是不现实的，而且由此得到的性质，不能反映愿望和意图的直观含义. 例如，无为而治问题. 对一个必然总是为真的命题(例如，地球是圆的)或必然最终为真的命题(例如，太阳明天会升起)，Agent 根本无需把它作为愿望或意图. 又如，副作用问题 $((\varphi \Rightarrow \psi) \wedge X\varphi) \Rightarrow X\psi$. 将 φ 作为愿望或意图的 Agent 无需将 ψ 作为愿望或意图.

其次，信念、愿望和意图的区分. 尽管定义了不同的模态算子，但都依赖于可达关系，不能反映三者间本质的不同.

第 3，未能表达出 BDI 之间的动态约束与激发关系，特别是愿望在 Agent 动态执行过程中的约束与激发作用. 因此，无法建立 Agent 的动态 BDI 模型.

针对这些不足，本文首先引入假设信念、表达愿望和意图在思维状态认知方面的含义，进而重新定义愿望和意图，建立反映 BDI 三者间约束与激发关系的理性 Agent 的动态 BDI 模型.

2 BDI 的逻辑描述

使用基于计算树逻辑的语言 L 来描述. L 包括：原子命题公式；$\neg$，$\wedge$，$\vee$，$\Rightarrow$ 和 $\Leftrightarrow$ 算子；时态算子 U(直到)，G(总是)和 F(终将)，路径算子 A(全路径)，A_S(全假设路径)，E(存在路径)和 E_S(存在假设路径)以及 BEL，ASM，DES，GOAL 和 INT 等描述思维状态的算子. 其中，$\varphi_1 U \varphi_2$ 表示存在未来的某一时刻，φ_2 成立并且在此之前 φ_1 总是成立. $F\varphi = true U \varphi$，$G\varphi = \neg F \neg \varphi$. 设 Φ 为原子命题集，$FORM(L)$ 为全体公式集，$\mathscr{P}(\Phi)$ 为 Φ 的幂集，$\mathscr{P}(FORM(L))$ 为 $FORM(L)$ 的幂集.

定义 1. Agent 模型是元组 $M = \langle W, T, <, ACT, act, \pi, bel, asm, des, int \rangle$.

其中 W 是可能世界的集合，包括可能的现实世界和可能的假想世界. T 是时间点集合，$< \subseteq T \times T$ 是时间点间的二元关系，$\langle T, < \rangle$ 构成时间树. ACT 是原子动作集，动作函数 $act: < \to ACT$，为时间树的每条边标记一个原子动作. 真值分配函数 $\pi: W \times T \to \mathscr{P}(\Phi)$，返回某一可能世界中在某一时间点成立的原子命题的集合. 确信、假设、愿望和意图函数 $bel, asm, des, int: W \times T \to \mathscr{P}(FORM(L))$，分别返回某一可能世界中在某一时间点具有的确信信念、假设信念、愿望和意图的集合.

在此模型下，对原子命题 φ，$M, w, t \models \varphi$ iff $\varphi \in \pi(w, t)$. 对于算子 $\neg$，$\wedge$，$\vee$，$\Rightarrow$，$\Leftrightarrow$，U，G，F，A，E 的定义与常规定义相同，不再赘述. $<$ 具有反向线性，这样，对每个时间点来说都具有线性历史和分支未来. Agent 在某一可能世界中的某一时间点所具有的信念、愿望和意图由它的执行历史来确定.

2.1 信　念

信念是 Agent 具有的关于环境信息、其他 Agent 信息和自身信息的集合. 这些信息可以确切地知道，也可以假设知道，分别称为确信信念和假设信念. 由于采用信念集进行语义解释，我们可以放弃 Agent 的全知假设.

定义 2. 确信信念算子 BEL 和假设信念算子 ASM.

$$M, w, t \models BEL(\varphi) \text{ iff } \varphi \in bel(w, t); \quad M, w, t \models ASM(\varphi) \text{ iff } \varphi \in asm(w, t).$$

确信信念相当于已有工作中定义的信念，而假设信念是对确信信念的扩充. 假设信念描述假想世界，而确信信念不仅包括对现实世界的描述，还包括对假想世界的描述. 算子 A_S 和 E_S 的引入就是为了描述假想世界中的路径，称做假设路径，并限制其只在信念公式中使用. $A_S\psi = \neg E_S \neg \psi$. 在假设信念中，$ASM(A\psi) \Leftrightarrow ASM(A_S\psi)$，$ASM(E\psi) \Leftrightarrow ASM(E_S\psi)$，无需区分；在确信信念中，$BEL(A\psi) \Rightarrow BEL(A_S\psi)$，$BEL(E_S\psi) \Rightarrow BEL(E\psi)$，必须明确区分. 由确信信念集和假设信念集，可以定义确信信念与假设信念联合可达关系 $\mathscr{I}_{bu}$，进一步说明两种信念.

定义 3. 思维状态等价关系 $\sim$. $\forall w_1, w_2 \in W, t \in T$，$w_1 \sim_t w_2$ iff $bel(w_1, t) = bel(w_2, t)$，$asm(w_1, t) = asm(w_2, t)$，$des(w_1, t) = des(w_2, t)$ 和 $int(w_1, t) = int(w_2, t)$ 同时成立.

定义 4. 确信信念与假设信念联合可达关系 $\mathscr{I}_{bu}$. $\mathscr{I}_{bu} \subseteq W \times T \times W \times W$，$\forall w_1, w_2, w_3 \in W, t \in T$，$(w_1, t, w_2, w_3) \in \mathscr{I}_{bu}$ iff

(1) $\forall\ \varphi\in asm(w_1,t), M,w_3,t\models\varphi$;并且,

(2) $\forall\ \varphi\in bel(w_1,t)$,$\varphi$为原子命题公式、原子BDI公式以及加否定$\neg$,$M,w_2,t\models\varphi$且$M,w_3,t\models\varphi$;并且,

(3) $\forall\ A\psi\in bel(w_1,t), M,w_2,t\models A\psi;\forall\ E\psi\in bel(w_1,t), M,w_2,t\models E\psi$;并且,

(4) $\forall\ A_S\psi\in bel(w_1,t), M,w_3,t\models A\psi;\forall\ E_S\psi\in bel(w_1,t), M,w_3,t\models E\psi$;并且,

(5) 设(2)~(4)中的公式为基本公式,$\forall\ \varphi\in bel(w_1,t)$,$\varphi$为基本公式$\varphi_1,\ldots,\varphi_n$按$\wedge$和$\vee$的组合,则$\varphi_1,\ldots,\varphi_n$应按相应的与或条件满足(2)~(4);并且,

(6) $w_1\sim_t w_2\sim_t w_3$.

对于$(w_1,t,w_2,w_3)\in\mathcal{I}_m$,$w_2$是确信信念可达的可能现实世界;$w_3$是确信信念和假设信念联合可达的可能假想世界,并受到可能现实世界的约束,反映在A_S,E_S与A,E之间的推导关系.在假设路径明确以A_S和E_S描述的前提下,两种信念满足以下约束:

约束 1. $\forall\ \varphi$,如果$\varphi\in asm(w,t)$,则$\varphi\notin bel(w,t)$.

可得确信与假设信念约束公理1:$ASM(\varphi)\Rightarrow\neg\ BEL(\varphi)$.

约束 2. $\forall\ \varphi$,如果$\varphi\in asm(w,t)$,则$\neg\ \varphi\notin bel(w,t)$.

可得确信与假设信念约束公理2:$ASM(\varphi)\Rightarrow\neg\ BEL(\neg\ \varphi)$.

定理 1. $ASM(A(\varphi_1\cup\varphi_2))\Rightarrow\neg\ BEL(A\neg\ (\varphi_1\cup\varphi_2)), ASM(AF\varphi)\Rightarrow\neg\ BEL(AG\neg\ \varphi), ASM(AG\varphi)\Rightarrow\neg\ BEL(AF\neg\ \varphi)$.

命题$ASM(A\varphi)\wedge BEL(\neg\ A\varphi)$可能成立.例如,对$\varphi=F\psi$,Agent确信在现实世界里存在一条路径使$\neg\ \psi$在未来一直保持成立,但可以排除演化到该路径的可能性,从而在假想世界里可以有$AF\psi$成立.对应地,命题$ASM(\neg\ E\varphi)\wedge BEL(E\varphi)$可能成立.假设信念与确信信念的这种约束关系对解释愿望和意图是重要的.

对确信信念算子BEL,仍采用KD45公理,对假设信念算子ASM,则需要相应变化.

对以下涉及假设信念的公理,要求公式中φ和ψ必须明确用E_S和A_S约束假设路径.

K:(k1) $(BEL(\varphi\Rightarrow\psi)\wedge(BEL(\varphi))\Rightarrow BEL(\psi)$;

(k2) $(\neg\ BEL(\psi)\wedge ASM(\varphi\Rightarrow\psi)\wedge ASM(\varphi))\Rightarrow ASM(\psi)$;

(k3) $(\neg\ BEL(\psi)\wedge ASM(\varphi\Rightarrow\psi)\wedge BEL(\varphi))\Rightarrow ASM(\psi)$;

(k4) $(\neg\ BEL(\psi)\wedge BEL(\varphi\Rightarrow\psi)\wedge ASM(\varphi))\Rightarrow ASM(\psi)$;

D:(d1) $BEL(\varphi)\Rightarrow\neg\ BEL(\neg\ \varphi)$;

(d2) $ASM(\varphi)\Rightarrow\neg\ ASM(\neg\ \varphi)$;

(d3) $ASM(\varphi)\Rightarrow\neg\ BEL(\varphi)$;

(d4) $ASM(\varphi)\Rightarrow\neg\ BEL(\neg\ \varphi)$;

4:对$X=BEL,ASM,DES,INT, X(\varphi)\Rightarrow BEL(X(\varphi))$;

5:对$X=BEL,ASM,DES,INT,\neg\ X(\varphi)\Rightarrow BEL(\neg\ (X(\varphi))$.

2.2 愿 望

定义 5. 愿望算子DES. $M,w,t\models DES(\varphi)$ iff $\varphi\in des(w,t)$.

愿望是Agent希望达到的状态或者希望保持的状态.分别称做实现型愿望和维护型愿望.直观上,在谈论愿望时,就隐含地对未来路径作了限制,将未来路径划分为满足所有愿望的路径和不能满足所有愿望的路径.只有由满足所有愿望的路径构成的世界才是愿望可达的世界.为此,愿望的表示应形如$DES(A\varphi)$,φ为路径公式.

$DES(AF\varphi)$是纯实现型愿望.$DES(A(\varphi_1\cup\varphi_2))$为包含维护条件的实现型愿望,即Agent不仅希望实现φ_2,还希望在实现过程中保持φ_1成立.$DES(AG\varphi)$是维护型愿望,希望在未来一直保持φ成立.

在已有研究中所定义的愿望不能采用这样的表示，因此不能充分表达愿望的直观含义. 例如在 Rao 和 Georgeff 的工作[4]中，只能表示形如 $GOAL(E\psi)$ 的目标(相当于 DES)，以便保持信念目标的一致关系 $GOAL(\varphi)\Rightarrow BEL(\varphi)$，但由此不能充分反映直观上愿望对未来路径的限制. 即使对目标能有不同的解释，对意图则更需要以 $INT(A\psi)$ 表示，由目标意图一致公理：$INT(\varphi)\Rightarrow GOAL(\varphi)$，还是使 $GOAL(A\psi)$ 无法回避. 解决以上问题的关键在于，用假设信念，而不是确信信念，来解释愿望在思维状态认知方面的含义，即愿望可达的世界只是假想世界.

约束 3. $\forall\ \varphi$，如果 $\varphi\in des(w,t)$，则 $\varphi\in asm(w,t)$.

可得愿望与假设信念一致公理：$DES(\varphi)\Rightarrow ASM(\varphi)$.

定理 2. $DES(A\varphi)\Rightarrow\neg\ BEL(A\varphi)$，$DES(A(\varphi_1\cup\varphi_2))\Rightarrow\neg\ BEL(A\neg(\varphi_1\cup\varphi_2))$，$DES(AF\varphi)\Rightarrow\neg\ BEL(AG\neg\varphi)$，$DES(AG\varphi)\Rightarrow\neg\ BEL(AF\neg\varphi)$.

以纯实现型愿望 $DES(AF\varphi)$ 为例，定理说明，Agent 不会愿望实现一个确信一定能达到的状态 φ，也不会愿望实现一个确信一定不能达到的状态 φ.

约束 4. $\forall\ A(\varphi_1\cup\varphi_2)$，如果 $A(\varphi_1\cup\varphi_2)\in des(w,t)$，则 $\varphi_1\in bel(w,t)$ 且 $\varphi_2\notin bel(w,t)$. $\forall\ AG\varphi$，如果 $AG\varphi\in des(w,t)$，则 $\varphi\in bel(w,t)$.

可得愿望与确信信念约束公理：$DES(A(\varphi_1\cup\varphi_2))\Rightarrow BEL(\varphi_1)\wedge\neg\ BEL(\varphi_2)$，$DES(AG\varphi)\Rightarrow BEL(\varphi)$.

定理 3. $DES(AF\varphi)\Rightarrow\neg\ BEL(\varphi)$.

约束 3，4 给出了愿望存在的必要条件，并且由相应的公理将维护愿望所需的推理转化到信念中进行.

实现型愿望持续公理：$DES(A(\varphi_1\cup\varphi_2))\Rightarrow A(DES(A(\varphi_1\cup\varphi_2))\cup(\neg\ ASM(A(\varphi_1\cup\varphi_2))\vee\neg\ (BEL(\varphi_1)\vee BEL(\varphi_2))))$.

维护型愿望持续公理：$DES(AG\varphi)\Rightarrow A(G(DES(AG\varphi))\vee(DES(AG\varphi)\cup(\neg\ ASM(AG\varphi)\vee\neg\ BEL(\varphi))))$.

定理 4. $DES(AF\varphi)\Rightarrow A(DES(AF\varphi)\cup(\neg\ ASM(AF\varphi)\vee BEL(\varphi)))$.

愿望具有持续性. 以纯实现型愿望 $DES(AF\varphi)$ 为例，Agent 希望保持一个愿望，直到不再假设该愿望一定能实现或者确信该愿望已实现. 由前面对假设信念的有关约束可知，放弃假设信念 $ASM(AF\varphi)$ 的主要理由是确信愿望必将成立，即 $BEL(AF\varphi)$，或者确信愿望一定不成立，即 $BEL(AG\neg\varphi)$.

愿望之间的关系结合意图给出，经典的 KD 公理不再采用. 可以得出的结论是，本文的模型避免了关于愿望的无为而治和副作用问题. Agent 不会将确信的状态作为愿望，也不会自然地把愿望的逻辑推论作为自己的愿望.

在已有工作中使用的目标算子 $GOAL$，可依据 DES 和 BEL 定义为 Agent 确信可能满足的愿望. 相对于信念、愿望和意图，目标只是中间概念.

定义 6. 目标算子 $GOAL$. $M,w,t\models GOAL(A\varphi)$ iff $M,w,t\models DES(A\varphi)\wedge BEL(E\varphi)$.

定义 7. 确信信念与愿望联合可达关系 $\mathcal{J}_{wd}$. 参照确信信念与假设信念联合可达关系 $\mathcal{J}_{bd}$ 的定义，将函数 asm 替换为 des，即可得到 $\mathcal{J}_{wd}$ 的定义.

由约束 3，$des(w,t)\subseteq asm(w,t)$，有 $\mathcal{J}_{bd}\subseteq\mathcal{J}_{wd}$.

2.3 意　图

定义 8. 意图算子 INT. $M,w,t\models INT(\varphi)$ iff $\varphi\in int(w,t)$.

意图是承诺的愿望. 从实现型意图中，Agent 在不违反意图约束的前提下选择下一个动作，并确信由这种选择可能产生的动作序列能够保证所有意图的满足. 意图可达世界中的每条路径都应满足所有意图，由此，意图应该形如 $INT(A\varphi)$，φ 为路径公式. 在 Rao 和 Georgeff 的工作中[4]，φ 只能形如 $F\varphi$，不能充分体现承诺的含义.

约束 5. $\forall\ \varphi$，如果 $\varphi\in int(w,t)$，则 $\varphi\in des(w,t)$.

可得意图与愿望一致公理：$INT(\varphi)\Rightarrow DES(\varphi)$.

定理 5. $INT(AF\varphi)\Rightarrow\neg\ BEL(AF\varphi)\wedge\neg\ BEL(AG\neg\varphi)\wedge\neg\ BEL(\varphi)$.

定义 8. 确信信念与意图联合可达关系 $\mathcal{F}_{bi}$. 参照确信信念与假设信念联合可达关系 $\mathcal{F}_{bd}$的定义,将函数 asm 替换为 int,即可得到 $\mathcal{F}_{bi}$的定义.

由约束 5,$int(w,t)\subseteq des(w,t)$,有 $\mathcal{F}_{bd}\subseteq\mathcal{F}_{bd}\subseteq\mathcal{F}_{bi}$.

与愿望相比,意图与确信信念有更多的一致,通过定义意图全路径算子 A_I 和意图存在路径算子 E_I 进行说明.

定义 9. 子世界. $\forall w\in W$,w 中所有路径的集合 P_w 与 w 一一对应,$P_w=\{p\mid p$ 为 w 中的路径$\}$. 称 w'为 w 的子世界,记做 $w'=subworld(w)$,当且仅当 $P_{w'}\subseteq P_w$.

这里,可不考虑子世界 w'是否属于 W.

约束 6. $\forall w_1,w_2,w_3\in W,t\in T,(w_1,t,w_2,w_3)\in\mathcal{F}_{bi}$,存在 $w_2'=subworld(w_2)$,满足所有意图,即$\forall\varphi\in int(w_1,t)$,有 $M,w_2',t\models\varphi$.

定义 10. 意图全路径算子 A_I 和意图存在路径算子 E_I. 基于约束 6,设 w_2^* 为 w_2 中满足所有意图的 w_2'中的最大子世界,即 $P_{w_2'}\subseteq P_{w_2^*}$,则 $M,w_2,t\models A_I\varphi$ iff $M,w_2^*,t\models A\varphi$;$M,w_2,t\models E_I\varphi$ iff $M,w_2^*,t\models E\varphi$.

意图与确信信念一致公理:$INT(A\varphi)\Rightarrow BEL(A_I\varphi)$.

A_I 和 E_I 是对偶算子,即 $A_I\psi=\neg E_I\neg\psi$. $A\psi\Rightarrow A_I\psi$,$E_I\psi\Rightarrow E\psi$. A_I,E_I 与 A_S,E_S 无推导关系.

定理 6. $INT(A\varphi)\Rightarrow GOAL(A\varphi)$.

实现型意图持续公理:$INT(A(\varphi_1\cup\varphi_2))\Rightarrow A(INT(A(\varphi_1\cup\varphi_2))\cup(\neg DES(A(\varphi_1\cup\varphi_2))\vee\neg BEL(A_I(\varphi_1\cup\varphi_2))))$.

维护型意图持续公理:$INT(AG\varphi)\Rightarrow A(G(INT(AG\varphi))\vee(INT(AG\varphi)\cup(\neg DES(AG\varphi)\vee\neg BEL(A_IG\varphi))))$.

定理 7. $INT(AF\varphi)\Rightarrow A(INT(AF\varphi)\cup(\neg DES(AF\varphi)\vee\neg BEL(A_IF\varphi)))$.

意图同样具有持续性. Agent 将保持一个意图,直到不再具有相应的愿望或者不再确信该意图一定能在承诺的前提下满足. 当一个实现型意图无法实现时,还可能有其他手段实现相应的愿望. Agent 会在该愿望激发下,产生新的意图.

与愿望一样,不再采用 KD 公理对意图进行推导,而是将有关的推导转化到信念中进行. 本文的模型避免了关于意图的无为而治和副作用问题.

3 动态 BDI 模型

定义 11. Agent 的执行模型为元组$\langle MS^0,next_act,next_ms\rangle$. 其中 MS^0 是 Agent 的初始思维状态,由初始确信信念、假设信念、愿望和意图的集合组成;$next\_act:MS\times ACT\rightarrow MS$,是从思维状态集 MS 到原子动作集 ACT 的选择函数;$next\_ms:MS\times ACT\rightarrow MS$,是在一个动作执行后,产生新的思维状态的函数.

在此执行模型下,Agent 的执行流程如下. Agent 的退出操作可作为基本的计算动作,放入动作集.

(S1) $MS^0=InitMentalState()$,$n=0$; //初始化思维状态;

(S2) $a=next\_act(MS^n)$; //选择动作;

(S3) Do (a); //执行选择的动作;

(S4) $MS^{n+1}=next\_ms(MS^n,a)$,$n=n+1$,转(S2). //修正思维状态.

3.1 从愿望到动作

Agent 选择执行的动作必须是为了实现愿望而承诺执行的动作,是意图的一部分. 对此,需要引入规划,对意图的表达进行扩展.

定义 12. 规划表达式. 设 a 为原子动作个体词,p 为规划表达式当且仅当 $p=a$ 或 $p=p_1;p_2$ 或 $p=p_1|p_2$. p_1,p_2 为规划表达式,$p_1;p_2$ 表示一个由 p_1 和 p_2 顺序执行构成的规划,$p_1|p_2$ 表示一个由 p_1 和 p_2 选择执行构成的规划.

定义 13. 动作算子 ACHIEVED 和 DONE. 设 a 为原子动作个体词，p 为规划表达式，$i(a)$将 a 解释为 ACT 中的一个原子动作，

$M,w,t\models ACHIEVED(\psi,a,\varphi)$ iff 对 t 在 w 中的直接前驱 t'，$M,w,t'\models\psi$ 且 $M,w,t\models\varphi$ 且 $act(t',t)=i(a)$；

$M,w,t\models ACHIEVED(\psi,p_1;p_2,\varphi)$ iff $\exists t'$，t' 为 t 在 w 中的前驱，$M,w,t'\models ACHIEVED(\psi,p_1,\varphi_1)$ 且 $M,w,t\models ACHIEVED(\varphi_1,p_2,\varphi)$；

$M,w,t\models ACHIEVED(\psi,p_1|p_2,\varphi)$ iff $M,w,t\models ACHIEVED(\psi,p_1,\varphi)$或 $M,w,t\models ACHIEVED(\psi,p_2,\varphi)$；

$M,w,t\models DONE(\psi,a)$ iff 对 t 在 w 中的直接前驱 t'，$M,w,t'\models\psi$ 且 $act(t',t)=i(a)$；

$M,w,t\models DONE(\psi,p_1;p_2)$ iff $\exists t'$，t' 为 t 在 w 中的前驱，$M,w,t'\models ACHIEVED(\psi,p_1,\varphi_1)$ 且 $M,w,t\models DONE(\varphi_1,p_2)$；

$M,w,t\models DONE(\psi,p_1|p_2)$ iff $M,w,t\models DONE(\psi,p_1)$或 $M,w,t\models DONE(\psi,p_2)$.

$ACHIEVED(\psi,p,\varphi)$说明，在 ψ 成立的前提下，完成了规划 p 的执行，并且使 φ 成立. $DONE(\psi,p)$说明，在 ψ 成立的前提下，完成了规划 p 的执行.

意图包含对动作或规划的承诺. 对实现型意图 $INT(A(\varphi\cup\psi))$，ψ 的完整表达应是 $ACHIEVED(\psi,p,\psi')$. 当只关心意图结果时，简写成 ψ'；当只关心规划 p 是否完成执行时，简写成 $DONE(\psi,p)$.

实现型意图分解公理：$INT(A(\varphi\cup(ACHIEVED(\psi,p,\psi'))))\Rightarrow INT(A(\varphi\cup(DONE(\psi,p))))\wedge INT(A(\varphi\cup\psi'))$.

定理 8. $INT(A(\varphi\cup(ACHIEVED(\psi,p,\psi'))))\Rightarrow DES(A(\varphi\cup(DONE(\psi,p))))\wedge DES(A(\varphi\cup\psi'))$.

在从愿望产生意图的过程中，Agent 需要为愿望作部分规划，以明确愿望是否可能实现，即得到目标，进而判断是否存在部分规划，满足对意图的约束条件，最后选择一个部分规划进行承诺，产生意图. 可以将这一过程抽象为一个 Agent 的计算动作，称为 GetInt. 愿望的激发作用就表现在促使 Agent 选择执行 GetInt.

愿望意图激发公理：$DES(A(\varphi\cup\psi))\wedge\neg INT(A(\varphi\cup\psi))\Rightarrow INT(AF(DONE(DES(A(\varphi\cup\psi))\wedge\neg INT(A(\varphi\cup\psi)),GetInt)))$；$DES(AG\varphi)\wedge\neg INT(AG\varphi)\Rightarrow INT(AF(DONE(DES(AG\varphi)\wedge\neg INT(AG\varphi),GetInt)))$.

愿望意图激发公理说明，对每个尚未形成意图的愿望，存在一个意图，承诺执行可能使该愿望成为意图的计算动作 GetInt.

从意图到动作执行需要在意图间进行选择. 这依赖于效用分析，在意图间建立序关系，由此形成基于各个意图部分规划的 Agent 的全局规划. 序关系的改变，导致全局规划的改变.

3.2　子意图和子愿望

对实现型意图，可以结合承诺的规划给出子意图.

子意图公理：设 p_i 为规划 p 的任意一个子规划，ψ_i 和 ψ_i' 为执行 p_i 的前提和结果，φ_i 为 φ 中与 p_i 有关的维护条件，有 $INT(A(\varphi\cup(ACHIEVED(\psi,p,\psi'))))\Rightarrow INT(A(\varphi_i\cup(ACHIEVED(\psi_i,p_i,\psi_i'))))$.

称 $INT(A(\varphi_i\cup(ACHIEVED(\psi_i,p_i,\psi_i'))))$为 $INT(A(\varphi\cup(ACHIEVED(\psi,p,\psi'))))$的子意图，相应地，称 $DES(A(\varphi_i\cup(ACHIEVED(\psi_i,p_i,\psi_i'))))$，$DES(A(\varphi_i\cup(DONE(\varphi_i,p_i))))$和 $DES(A(\varphi_i\cup\psi_i'))$为 $DES(A(\varphi\cup(ACHIEVED(\psi,p,\psi'))))$的子愿望.

由于子愿望 $DES(A(\varphi_i\cup\psi_i'))$不是限定在某个子规划，当子意图 $INT(A(\varphi_i\cup(ACHIEVED(\psi_i,p_i,\psi_i'))))$无法实现时，$DES(A(\varphi_i\cup\psi_i'))$可能依然成立. 这样，Agent 会由愿望激发，产生新的子意图来实现 $\varphi_i\cup\psi_i'$，而不是完全放弃整个意图规划 p.

4　结　语

本文针对已有 BDI 模型中的概念不明确，逻辑全知以及缺乏 BDI 之间动态关系的表达等问题展开讨论，通过引入假设信念丰富了信念的表达，并以此解释愿望和意图在 Agent 思维状态认知方面的含义，符合直观的理解. 在此基础上，重新定义了愿望和意图的概念，给出了 BDI 三者之间的静态约束关系，建立了静态 BDI 模型，

随后,分析了 Agent 的执行过程,说明了愿望对意图的激发作用,并通过引入规划,定义子意图和子愿望,从而更充分地表达了 BDI 三者之间的动态约束与激发关系,改进了理性 Agent 的已有 BDI 模型.

参考文献

1 Hewitt C. Open information systems semantics for distributed artificial intelligence. Artificial Intelligence, 1991,47(1): 79～106

2 Bratman M E. Intentions, Plans, and Practical Reason. Cambridge, MA: Hardvard University Press, 1987

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

4 Rao A S, Georgeff M P. The semantics of intention maintenance for rational agents. In: Mellish C S ed. Proceedings of the 14th International Joint Conference on Artificial Intelligence. San Mateo, CA: Morgan Kaufmann Publishers, 1995. 704～710

5 Konolige K, Pollack M E. A respresentationalist theory of intention. In: Bajcsy R ed. Proceedings of the 13th International Joint Conference on Artificial Intelligence. San Mateo, CA: Morgan Kaufmann Publishers, 1993. 390～395

6 Shoham Y. An overview of agent-oriented programming. In: Bradshaw M ed. Software Agents. Menlo Park, CA: AAAI Press, 1997. 271～289

7 Haddadi A S. Communication and Cooperation in Agent Systems. Berlin: Springer-Verlag, 1996. 1～134

8 Dunin-Keplicz B, Verbrugge R. Collective Commitments. In: Durfee F ed. Proceedings of the 2nd International Conference on Multi-Agent Systems. Menlo Park, CA: AAAI Press, 1996. 56～63

9 Castelfrachi C. Commitments: from individual intentions to groups and organizations. In: Lesser V R ed. Proceedings of the 1st International Conference on Multi-Agent Systems. Menlo Park, CA: AAAI Press/The MIT Press, 1995. 41～48

A BDI Model for Rational Agents

KANG Xiao-qiang SHI Chun-yi

(*Department of Computer Science and Technology Tsinghua University Beijing* 100084)

Abstract In this paper, a new BDI model for rational agents is presented by introducing assumptive belief with traditional belief in order to express the intuitive meaning of desire and intention on the cognitive aspect of the metal state of rational agents. Comparing with the BDI models from Cohen & Levesque, Rao & Georgeff, and Konolige & Pollack, this model overcomes the misunderstanding of the concepts of BDI, solves the transference problem and the side-effect problem for desire and intention, and shows both of the static and the dynamic relations between BDI, especially the maintaining and triggering role of desire.

Key words Rational agent, mental state, belief, desire, intention.