ISSN 1000-9825 Journal of Software 软件学报 2000,11(10):1353~1360

Agent-BDI 逻辑*

胡山立[1] 石纯一[2]

[1](福州大学计算机科学与技术系 福州 350002)

[2](清华大学计算机科学与技术系 北京 100084)

E-mail: husl@fzu.edu.cn

摘要 阐述了 Agent 的形式化描述应该采用含有正规和非正规模态算子的混合模态逻辑为逻辑工具的观点,建立了 Agent-BDI 逻辑的代表系统 A-BI,讨论了它的语法和语义.特别是给出了非正规模态算子基于 Kripke 标准可能世界的新的语义解释,证明了 A-BI 逻辑系统不但是可靠的,而且是完备的.A-BI 逻辑系统恰当地刻画了信念与意图的本质与内在联系,可作为 Agent 形式化研究的逻辑工具.

关键词 Agent,Agent-BDI 模型,模态逻辑,信念,意图.

中图法分类号 TP18

使用诸如相信、想要、希望等思维状态或意识属性来解释人类的行为一直是心理学界所采用的方法,并已取得了丰富的成果.许多学者认为把 Agent 系统作为意识系统来研究是合理和有用的[1~3].为了描述 Agent,应该采用哪些意识属性呢?对此,已有大量的研究工作[2~5].较有影响的是 Rao 和 Georgeff 的 BDI 模型[5]以及 Wooldridge 和 Jennings 的结果[3].Rao 和 Georgeff 的 BDI 模型用信念、愿望(或目标)和意图这 3 个意识属性来描述 Agent 结构.Wooldrige 和 Jennings 把意识属性分为 information attitudes 和 pro-attitudes 两类.他们认为,合理地表示 Agent 至少包含一个 information attitude 和一个 pro-attitude.对此,人们通常采用信念(belief)和意图(intention),即 BI 结构.为简明起见,本文采用 BI 结构而展开讨论,所得结果容易推广到 BDI 结构.

Agent 的 BDI 结构应采用什么样的形式化工具?事实证明,采用经典命题和一阶谓词逻辑是不适宜的.信念、目标和意图等意识概念实质上是一种模态算子,其表示和推理的合适的形式化工具是模态逻辑和可能世界语义[3].

在 AI 领域,Agent 的理论研究通常基于正规模态逻辑.正规模态逻辑存在"逻辑全知"(logical omniscience)问题.这一问题以及由此带来的其他一系列相关问题(如副作用问题、传递问题等)给 Agent 理论研究带来了困难[4,5].例如,正规模态逻辑在逻辑蕴涵下是封闭的:给定一个正规模态算子 $\Box$,α,β 是公式.如果 $\Box\alpha$ 是真的,且 $\vdash\alpha\rightarrow\beta$,那么 $\Box\beta$ 是真的.当 $\Box$ 代表信念时,由于 Agent 系统是有限系统,知识和推理能力不完备且资源有限,封闭性对 Agent 是不现实的,只能看成是理想情形.当 $\Box$ 代表意图时,意图不同于信念,逻辑蕴涵的封闭性即使在理想情形也不应对意图假设[6].我们认为,由于逻辑全知问题以及由此带来的副作用等问题是正规模态算子的固有性质,因此,意图不能是正规模态算子.基于 Kripke 标准可能世界语义的正规模态逻辑不适于描述 Agent,应该采用非正规模态逻辑或采用同时含有正规和非正规模态算子的混合模态逻辑,作为形式化工具并寻求合适的语义.由于 Agent 有多个意识属性,把信念和意图分别抽象为正规和非正规模态算子更具代表性.

下面,我们具体给出 Agent-BDI 逻辑系统,它是含有正规模态算子信念和非正规模态算子意图的混合模态逻辑系统.为了方便讨论,只限于命题逻辑情形.它是经典命题逻辑系统 P 在 Agent 的扩张,其形式语言记为 L_{PA}.

* 本文研究得到国家自然科学基金(Nos. 69733020,69973023)资助.作者胡山立,1944 年生,副教授,主要研究领域为人工智能应用基础,多 Agent 系统.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.

本文通讯联系人:胡山立,福州 350002,福州大学计算机科学与技术系

本文 1999-10-26 收到原稿,2000-02-17 收到修改稿

1 形式语言 L_{PA}

L_{PA}是经典命题逻辑语言 L_P 增加模态算子 B(信念)和 I(意图)的扩张.

L_{PA}-初始符号：非空的命题变元集 $\Phi=\{p,q,r,\ldots\}$；逻辑联结词 $\neg,\rightarrow$；模态算子 B,I；括号(,).

L_{PA}-公式形成规则：任意命题变元 p 是公式；若 α,β 是公式，则 $\neg\alpha,(\alpha\rightarrow\beta),B(\alpha),I(\alpha)$也都是公式.

L_{PA}-公式定义：α 是 L_{PA}-公式，当且仅当 α 是有限次使用 L_{PA}-公式形成规则得到的有限非空符号串. 规定 B 和 I 的优先级同 $\neg$，其次是 $\rightarrow$，公式最外层的括号可以省略.

定义 1. (1) $\alpha\wedge\beta=_{df}\neg(\alpha\rightarrow\neg\beta);\alpha\vee\beta=_{df}\neg\alpha\rightarrow\beta;\alpha\leftrightarrow\beta=_{df}(\alpha\rightarrow\beta)\wedge(\beta\rightarrow\alpha)$；

(2) $Form(L_{PA})$是所有 L_{PA}-公式形成的公式集.

2 语　义

我们给出 Agent-BDI 逻辑系统的一种基于标准(正规)可能世界的非经典语义解释[7,8]. 与 Kripke 的非正规模态逻辑算子的可能世界语义不同的是，意图用可能世界集上的两个二元关系 R_i^I 和 R_i^f 来解释. 我们将说明这种语义解释对意图是合适的.

定义 2. $F=\langle W,R_b,R_i^I,R_i^f\rangle$是一个框架，当且仅当其中 W 是任一非空集(可看成是可能世界集)，R_b,R_i^I,R_i^f是 W 上的任意 3 个二元关系.

定义 3. 设 $F=\langle W,R_b,R_i^I,R_i^f\rangle$是任一框架. V 是框架 F 上对 L_{PA}-公式的一个 PA-赋值，当且仅当 V：$Form(L_{PA})\times W\rightarrow\{\text{true},\text{false}\}$，且对任意的 L_{PA}-公式 α,β 和 $w\in W$，满足(下面以 1 代表 true，以 0 代表 false)：

$[V_P]$,$[V_\neg]$和$[V_\rightarrow]$同经典命题逻辑中相应的定义.

$[V_B]$：$V(B(\alpha),w)=1$，当且仅当对任意 $w'\in W$，若 R_bww'，则 $V(\alpha,w')=1$，否则 $V(\alpha,w')=0$.

$[V_I]$

$$V(I(\alpha),w)=\begin{cases}1, & \text{对任意 } w'\in W,\text{若 } R_i^Iww',\text{则 } V(\alpha,w')=1,\text{并且对任意 } w''\in W,\text{若 } R_i^fww'',\text{则}\\ & V(\alpha,w'')=0;\\ 0, & \text{否则.}\end{cases}$$

根据以上定义和 $\wedge,\vee,\leftrightarrow$ 的语法定义，不难得到相应的赋值规则$[V_\wedge]$,$[V_\vee]$和$[V_\leftrightarrow]$.

定义 3 中(V_B)的直观意义是在可能世界 w 中 Agent 相信 α，当且仅当 Agent 认为在与 w 有关系的 R_b 的所有可能世界 w' 中 α 为真. $[V_I]$的直观意义是，在可能世界 w 中 Agent 意图 α，当且仅当 Agent 认为在与 w 有关系的 R_i^I 的所有可能世界 w' 中 α 为真，并且在与 w 有关系的 R_i^f 的所有可能世界 w''中 α 为假. 这是因为意图是对目标的承诺，表明 Agent 要为实现该目标而奋斗. 对理性 Agent 而言，意图 α 应该至少具备两个条件：一个是 Agent认为它是可能实现的，这由与 w 有关系的 R_i^I 的所有 w' 中 α 为真来刻画；另一个是 Agent 认为它是目前尚未实现的，这由与 w 有关系的 R_i^f 的所有 w''中 α 为假来刻画.

意图的这种语义解释比较自然、合理. 对非正规模态算子采用这种语义解释，正规模态算子可看成是非正规模态算子当 $R^f=\varnothing$ 时的退化情形，从而可以把两种算子的语义解释统一起来.

定义 4. $M=\langle W,R_b,R_i^I,R_i^f,V\rangle$是一个模型，当且仅当 $F=\langle W,R_b,R_i^I,R_i^f\rangle$是一个框架. V 是 F 上的一个 PA-赋值，也称 M 是 F 上的一个模型.

定义 5. 设 $M=\langle W,R_b,R_i^I,R_i^f,V\rangle$是任一模型，$\alpha$ 是任一 L_{PA}-公式.

(1) 对任意 $w\in W$，若 $V(\alpha,w)=1$，记为 $M\models_w\alpha$，称 α 在 w 上是真的；若 $V(\alpha,w)=0$，记为 $M\not\models_w\alpha$，称 α 在 w 上是假的.

(2) 若存在 $w\in W$，使得 $M\models_w\alpha$，称 α 在 M 上可满足.

(3) 若对任意 $w\in W$，有 $M\models_w\alpha$，则称 α 在模型 M 上有效，记为 $M\models\alpha$.

定义 6. 设 $F=\langle W,R_b,R_i^I,R_i^f\rangle$是任一框架，$\alpha$ 是任一 L_{PA}-公式.

(1) 若存在 F 上的一个模型 M,使得 $M\models\alpha$,则称 α 在 F 上可满足.

(2) 若对 F 上的任意模型 M 都有 $M\models\alpha$,则称 α 在框架 F 上有效,记为 $F=\alpha$.

容易验证,α 在 M(或 F)上有效,当且仅当 $\neg\alpha$ 在 M(或 F)上不是可满足的. 对 $w\in W$,定义 $R_b(w)=_{df}\{w'\in W|R_bww'\}$,$R_i^c(w)=_{df}\{w'\in W|R_i^cww'\}$,$R_i^f(w)=_{df}\{w'\in W|R_i^fww'\}$.

定理 1. 对任意 $n\geqslant 2$,任意 L_{PA}-公式 $\alpha,\beta,\alpha1,\alpha2,\ldots,\alpha n$,任意模型 M,有:

(1) 若 α 是命题重言式,则 $M\models\alpha$.

(2) 若 $M\models\alpha$ 且 $M\models\alpha\rightarrow\beta$,则 $M\models\beta$.

(3) $M\models B(\alpha\rightarrow\beta)\rightarrow(B(\alpha)\rightarrow B(\beta))$.

(4) 若 $M\vdash\alpha$,则 $M\models B(\alpha)$.

(5) $M\models I(\alpha)\wedge I(\beta)\rightarrow I(\alpha\wedge\beta)$.

(6) 若 $M\models\alpha1\wedge\alpha2\wedge\ldots\wedge\alpha n\rightarrow\beta$ 且 $M\models\neg\alpha1\wedge\neg\alpha2\wedge\ldots\wedge\neg\alpha n\rightarrow\neg\beta$,则 $M\models I(\alpha1)\wedge I(\alpha2)\wedge\ldots\wedge I(\alpha n)\rightarrow I(\beta)$.

证明从略.

所有模型 M 的集合称为模型类,记为 $M1$;所有框架 F 的集合称为框架类,记为 $F1$.

3 Agent-BDI 逻辑系统

本节从建立一个基本的 Agent-BDI 逻辑系统 A-BI1 开始,然后讨论为了恰当地反映信念和意图的本质以及它们之间的内在关系所需引入的公理,探求相应的语义约束,最后得到我们所需要的 Agent-BDI 逻辑系统 A-BI. 在本文中,Agent-BDI 逻辑系统泛指系统 A-BI1 的扩张.

A-BI1 包括如下公理模式和初始推理规则模式,是经典命题逻辑系统 P 的真扩张.

公理模式

(A1) 所有命题重言式.

(A2) $B(\alpha\rightarrow\beta)\rightarrow(B(\alpha)\rightarrow B(\beta))$.

初始推理规则模式

(R1) 若 $\alpha\rightarrow\beta$ 且 α,则 β.

(R2) 若 α,则 $B(\alpha)$.

(R3) 若 $\alpha1\wedge\alpha2\wedge\ldots\wedge\alpha n\rightarrow\beta$ 且 $\neg\alpha1\wedge\neg\alpha2\wedge\ldots\wedge\neg\alpha n\rightarrow\neg\beta$,则 $I(\alpha1)\wedge I(\alpha2)\wedge\ldots\wedge I(\alpha n)\rightarrow I(\beta)$.

由 (A1)(A2)(R1)(R2)可以证明,对于信念与(R3)类似但略强于它的推理规则:对任意 $n\geqslant 2$,若 $\alpha1\wedge\alpha2\wedge\ldots\wedge\alpha n\rightarrow\beta$,则 $B(\alpha1)\wedge B(\alpha2)\wedge\ldots\wedge B(\alpha n)\rightarrow B(\beta)$. 对于意图,我们采用略弱于(A2)(R2)的(R3)作为推理规则,这可数个推理规则(R3)刻画了意图语义的内涵. 我们当然希望能用有限个来取代它,但没能做到,不过容易证明,这可数个规则不是相互独立的,它们之间有这样的关系:当 $k>m$ 时,由 $n=k$ 的规则(R3)可导出 $n=m$ 的规则(R3). 由 $n=2$ 的规则(R3)不难证明,意图有定理 K:$I(\alpha\rightarrow\beta)\rightarrow(I(\alpha)\rightarrow I(\beta))$. 对于意图,没有必然化规则 N,从而不存在逻辑全知问题. 意图不是正规模态算子.

定理 2. 系统 A-BI1 关于框架类 $F1$ 是可靠的.

证明:由定理 1 可证,从略.

为了证明系统 A-BI1 的完备性,先做些准备.

定义 7. 设 S 是任一系统,α 是任一公式,$\{\alpha1,\alpha2,\ldots,\alpha n\}$ 是任一有穷公式集,Γ 是任一无穷公式集.

(1) α 是 S-相容的,当且仅当 $\neg\alpha$ 不是 S-可证的.

(2) $\{\alpha1,\alpha2,\ldots,\alpha n\}$ 是 S-相容的,当且仅当 $\alpha1\wedge\alpha2\wedge\ldots\wedge\alpha n$ 是 S-相容的.

(3) Γ 是 S-相容的,当且仅当 Γ 的任意有穷子集是 S-相容的.

定义 8. 设 S 是任一系统,Γ 是任一公式集,称 Γ 是 S-极大相容集,当且仅当 Γ 是 S-相容的,并且对任意公式 α,若 $\Gamma\cup\{\alpha\}$ 是 S-相容的,则 $\alpha\in\Gamma$.

引理 1. 设 S 是包含(A1)和(R1)的任一系统.

(1) 设 θ 是任一 S-相容集,α 是任一公式,有:

(a) $\theta\cup\{\alpha\}$ 和 $\theta\cup\{\neg\alpha\}$ 中必有一个是 S-相容的;

(b) θ 可被扩充为一个 S-极大相容集 Γ,$\theta\subseteq\Gamma$.

(2) 设 Γ 是任一 S-极大相容集,α,β 是任意公式,有:

(a) $\alpha\in\Gamma$,当且仅当 $\neg\alpha\notin\Gamma$;

(b) $(\alpha\rightarrow\beta)\in\Gamma$,当且仅当 $\neg\alpha\in\Gamma$ 或 $\beta\in\Gamma$;

(c) 若 $\alpha\in\Gamma$ 且 $(\alpha\rightarrow\beta)\in\Gamma$,则 $\beta\in\Gamma$;

(d) 若 α 是 S-可证的,则 $\alpha\in\Gamma$.

证明从略.

设 θ 是任一公式集,定义 $B(\theta)=_{df}\{\alpha|B(\alpha)\in\theta\}$,$I^-(\theta)=_{df}\{\alpha|I(\alpha)\in\theta\}$,$\neg I^-(\theta)=_{df}\{\neg\alpha|I(\alpha)\in\theta\}$.

定义 9. 设 S 是任一混合模态逻辑系统 Agent-BDI.

(1) 四元组 $F_s=\langle W_s,R_{bs},R_{is},R_{is}^f\rangle$,其中 W_s 是所有 S-极大相容集的集合,R_{bs},R_{is},R_{is}^f 是 W_s 上的二元关系.F_s 是 S 的正则框架当且仅当满足:对任意 $\Gamma,\Gamma'\in W_s$,$R_{bs}\Gamma\Gamma'$ 当且仅当 $B^-(\Gamma)\subseteq\Gamma'$,$R_{is}\Gamma\Gamma'$ 当且仅当 $I^-(\Gamma)\subseteq\Gamma'$,且 $R_{is}^f\Gamma\Gamma'$ 当且仅当 $\neg I^-(\Gamma)\subseteq\Gamma'$.

(2) 五元组 $M_s=\langle W_s,R_{bs},R_{is},R_{is}^f,V_s\rangle$ 是 S 的正则模型,当且仅当 $\langle W_s,R_{bs},R_{is},R_{is}^f\rangle$ 是 S 的正则框架,V_s 是 PA-赋值并且满足:对任意命题变元 p,任意 $\Gamma\in W_s$,有 $V_s(p,\Gamma)=1$ 当且仅当 $p\in\Gamma$.

引理 2. 设 M_s 是正则模型,若 $\Gamma\in W_s$ 且 $\neg B(\alpha)\in\Gamma$,则存在 $\Gamma'\in W_s$,使得 $R_{bs}\Gamma\Gamma'$ 且 $\neg\alpha\in\Gamma'$.

证明从略.

引理 3. 设 M_s 是正则模型,若 $\Gamma\in W_s$ 且 $\neg I(\alpha)\in\Gamma$,则存在 $\Gamma'\in W_s$,使得 $R_{is}^f\Gamma\Gamma'$ 且 $\neg\alpha\in\Gamma'$,或存在 $\Gamma''\in W_s$,使得 $R_{is}^f\Gamma\Gamma''$ 且 $\alpha\in\Gamma''$.

证明:因为 $\neg I(\alpha)\in\Gamma$,$\Gamma\in W_s$,由引理 1 的(2)(a),有 $I(\alpha)\notin\Gamma$,从而 $\alpha\notin I^-(\Gamma)$.假如我们能证明 $I^-(\Gamma)\cup\{\neg\alpha\}$ 是 S-相容的,或 $\neg I^-(\Gamma)\cup\{\alpha\}$ 是 S-相容的,则由引理 1 的(1)可将 $I^-(\Gamma)\cup\{\neg\alpha\}$ 或 $\neg I^-(\Gamma)\cup\{\alpha\}$ 扩大为 S-极大相容集.显然,这就是所求的 Γ' 或 Γ''.下面,我们证明 $I^-(\Gamma)\cup\{\neg\alpha\}$ 是 S-相容的,或 $\neg I^-(\Gamma)\cup\{\alpha\}$ 是 S-相容的.

用反证法,假如 $I^-(\Gamma)\cup\{\neg\alpha\}$ 和 $\neg I^-(\Gamma)\cup\{\alpha\}$ 都不是 S-相容的,那么存在 $I^-(\Gamma)$ 的有穷子集 $\{\beta1,\ldots,\beta n\}$,$n\geqslant0$ 和 $\neg I^-(\Gamma)$ 的有穷子集 $\{\neg\beta1,\ldots,\neg\beta n\}$,使得 $\vdash\neg(\beta1\wedge\ldots\wedge\beta n\wedge\neg\alpha)$ 且 $\vdash\neg(\neg\beta1\wedge\ldots\wedge\beta n\wedge\alpha)$.

(1) 若 $n=0$,则 $\vdash\alpha$ 且 $\vdash\neg\alpha$,与系统 S 的可靠性矛盾.

(2) 若 $n>0$,则 $\vdash\neg(\beta1\wedge\ldots\wedge\beta n\wedge\neg\alpha)$ 且 $\vdash\neg(\neg\beta1\wedge\ldots\wedge\neg\beta n\wedge\alpha)$,从而 $\vdash\beta1\wedge\ldots\wedge\beta n\rightarrow\alpha$ 且 $\vdash\neg\beta1\wedge\ldots\wedge\neg\beta n\rightarrow\neg\alpha$.据此由(R3)知,$\vdash\neg(I(\beta1)\wedge\ldots\wedge I(\beta n)\wedge\neg I(\alpha))$.这表明 $\{I(\beta1),\ldots,I(\beta n),\neg I(\alpha)\}$ 不是 S-相容的,与 Γ 是 S-极大相容集相矛盾.

定理 3. 设 M_s 是正则模型,对任意公式 α,任意 $\Gamma\in W_s$,有 $\alpha\in\Gamma$ 当且仅当 $V_s(\alpha,\Gamma)=1$.

证明:对公式 α 的结构施归纳.

当 α 是命题变元时,由 V_s 的定义可直接推得.

当 $\alpha=\neg\beta$ 时,由引理 1 中(2)(a),若 $\neg\beta\in\Gamma$,有 $\beta\notin\Gamma$,从而 $V_s(\beta,\Gamma)=0$,于是 $V_s(\neg\beta,\Gamma)=1$;若 $\neg\beta\notin\Gamma$,有 $\beta\in\Gamma$,从而 $V_s(\beta,\Gamma)=1$,于是 $V_s(\neg\beta,\Gamma)=0$.

当 $\alpha=\beta\rightarrow\gamma$ 时,由引理 1 中(2)(b),若 $(\beta\rightarrow\gamma)\in\Gamma$,有 $\beta\notin\Gamma$,或 $\gamma\in\Gamma$,从而 $V_s(\beta,\Gamma)=0$ 或 $V_s(\gamma,\Gamma)=1$,于是

$V_s(\beta\to\gamma,\Gamma)=1$;若$(\beta\to\gamma)\notin\Gamma$,有 $\beta\in\Gamma$ 且 $\gamma\notin\Gamma$,从而 $V_s(\beta,\Gamma)=1$ 且 $V_s(\gamma,\Gamma)=0$,于是 $V_s(\beta\to\gamma,\Gamma)=0$.

当 $\alpha=B(\beta)$时,若 $B(\beta)\in\Gamma$,有对任意 $\Gamma'\in W_s$,若 $R_b\Gamma\Gamma'$,则 $\beta\in\Gamma'$,从而 $V_s(\beta,\Gamma')=1$,于是 $V_s(B(\beta),\Gamma)=1$;若 $B(\beta)\notin\Gamma$,有 $\neg B(\beta)\in\Gamma$,由引理 2 推得,存在 $\Gamma'\in W_s$,使得 $R_b\Gamma\Gamma'$ 且 $\neg\beta\in\Gamma'$(从而 $\beta\notin\Gamma'$),于是 $V_s(\beta,\Gamma')=0$,所以 $V_s(B(\beta),\Gamma)=0$.

当 $\alpha=I(\beta)$时,若 $I(\beta)\in\Gamma$,有对任意 $\Gamma'\in W_s$,若 $R_i^c\Gamma\Gamma'$,则 $\beta\in\Gamma'$,且对任意 $\Gamma''\in W_s$,若 $R_i^f\Gamma\Gamma''$,则 $\neg\beta\in\Gamma''$(从而 $\beta\notin\Gamma''$),于是 $V_s(\beta,\Gamma')=1$ 且 $V_s(\beta,\Gamma'')=0$,所以 $V_s(I(\beta),\Gamma)=1$;若 $I(\beta)\notin\Gamma$,有 $\neg I(\beta)\in\Gamma$,由引理 3 推得,存在 $\Gamma'\in W_s$,使得 $R_i^c\Gamma\Gamma'$ 且 $\neg\beta\in\Gamma'$(从而 $\beta\notin\Gamma'$),或存在 $\Gamma''\in W_s$,使得 $R_i^f\Gamma\Gamma''$ 且 $\beta\in\Gamma''$,于是 $V_s(\beta,\Gamma')=0$ 或 $V_s(\beta,\Gamma'')=1$,所以 $V_s(I(\beta),\Gamma)=0$. □

定理 4. 系统 A-BI1(证明中简称系统 S)关于框架类 $F1$ 是完备的.

证明:若公式 α 关于 $F1$ 是有效的,要证明 α 是 S-可证的.用反证法,若 α 不是 S-可证的,那么 $\neg\alpha$ 是 S-相容的,由引理 1 的(1),$\neg\alpha$ 包含在某个极大相容集 Γ 中,于是对正则模型 M_s,存在 $\Gamma\in W_s$,使得 $\neg\alpha\in\Gamma$.由定理 3,$V_s(\neg\alpha,\Gamma)=1$,即 $V_s(\alpha,\Gamma)=0$,α 在模型 M_s 上不是有效的,从而在框架 F_s 上不是有效的,与公式 α 关于 $F1$ 是有效的相矛盾.

下面考虑 A-BI 逻辑中需要引入的其他公理,对所得到的系统讨论关于框架类的可靠性和完备性问题.

关于信念,可以考虑增加公理 D、公理 4 和公理 5:

(A3) 关于信念的公理 D 模式:$B(\alpha)\to\neg B(\neg\alpha)$.

(A4) 关于信念的公理 4 模式:$B(\alpha)\to B(B(\alpha))$.

(A5) 关于信念的公理 5 模式:$\neg B(\alpha)\to B(\neg B(\alpha))$.

关于意图,可以考虑增加公理 D:

(A6) 关于意图的公理 D 模式:$I(\alpha)\to\neg I(\neg\alpha)$.

定义 10. 设 W 是任一非空集,R,T 均是 W 上的二元关系.

(1) 称 R 在 W 上是持续的,若对任意 $w\in W$,存在 $w'\in W$,使得 Rww'.

(2) 称 $R\cup T$ 在 W 上是持续的,若对任意 $w\in W$ 有:存在 $w'\in W$,使得 Rww';或存在 $w''\in W$,使得 Tww''.

(3) 称 R 在 W 上是传递的,若对任意 $w,w',w''\in W$,如果 Rww' 且 $Rw'w''$,则 Rww''.

(4) 称 R 在 W 上是欧几里德的,若对任意 $w,w',w''\in W$,如果 Rww' 且 Rww'',则 $Rw'w''$.

引理 4. 设 $F=(W,R_b,R_i^c,R_i^f)$是任一框架.

(1) (A3)关于 F 是有效的,当且仅当 R_b 在 W 上是持续的.

(2) (A4)关于 F 是有效的,当且仅当 R_b 在 W 上是传递的.

(3) (A5)关于 F 是有效的,当且仅当 R_b 在 W 上是欧几里德的.

(4) (A6)关于 F 是有效的,当且仅当 $R_i^c\cup R_i^f$ 在 W 上是持续的.

证明:下面只证明(4);对(1)~(3)的证明从略.

若(A6)关于 F 是有效的,即 $F\models I(\alpha)\to\neg I(\neg\alpha)$,而 $R_i^c\cup R_i^f$ 在 W 上不是持续的,则存在 $w\in W$,使得对任一 $w'\in W$,R_i^cww' 和 R_i^fww' 均不成立,从而对任一 L_{PA}-赋值 V,$V(I(\alpha),w)=1$,$V(I(\neg\alpha),w)=1$,于是 $V(I(\alpha)\to\neg I(\neg\alpha),w)=0$,与 $F\models I(\alpha)\to\neg I(\neg\alpha)$相矛盾.

若 $R_i^c\cup R_i^f$ 在 W 上是持续的,而(A6)关于 F 不是有效的,即 $F\not\models I(\alpha)\to\neg I(\neg\alpha)$,那么存在 L_{PA}-赋值 V 和 $w\in W$,使得 $V(I(\alpha),w)=1$ 且 $V(\neg I(\neg\alpha),w)=0$.由 $V(I(\alpha),w)=1$,有

(*) 对任一 $w'\in W$,若 R_i^cww',则 $V(\alpha,w')=1$,且对任一 $w''\in W$,若 R_i^fww'',则 $V(\alpha,w'')=0$.

由 $V(\neg I(\neg\alpha),w)=0$,即 $V(I(\neg\alpha),w)=1$,有

(* *)对任一 $w'\in W$,若 $R_i^f ww'$,则 $V(\neg\alpha,w')=1$,且对任一 $w''\in W$,若 $R_i^f ww''$,则 $V(\neg\alpha,w'')=0$.

因为 $R_i^f\cup R_i^f$ 在 W 上是持续的,所以(*)与(* *)矛盾. □

将 R_b 在 W 上满足持续性、传递性和欧几里德性,$R_i^f\cup R_i^f$ 在 W 上满足持续性的框架类记为 $F2$. 将公理(A3)、(A4)、(A5)和(A6)加入系统 A-BI1 中,记为 A-BI2. 那么由定理 2、定理 4 和引理 4,不难证明如下定理:

定理 5. 系统 A-BI2 关于框架类 $F2$ 是可靠和完备的.

系统 A-BI2 还没有反映出意图和信念的关系,为了反映意图和信念的关系,应增加哪些公理呢? 我们认为,意图是对实现目标的承诺,理性 Agent 在 $w\in W$ 意图 α,应该具备两个条件. 一个条件是 Agent 认为 α 是可能实现的,即 α 在 $R_b(w)$ 的某些可能世界为真,从而 $\neg\alpha$ 在这些可能世界为假,即 $B(\neg\alpha)$ 为假,$\neg B(\neg\alpha)$ 为真,因此 $I(\alpha)\rightarrow\neg B(\neg\alpha)$ 为真. 另一个条件是 Agent 认为 α 不是已经实现或必然为真的,也就是说要经过它的努力才可能实现,即 α 在 $R_b(w)$ 的某些可能世界为假,从而 $B(\alpha)$ 为假,$\neg B(\alpha)$ 为真,因此 $I(\alpha)\rightarrow\neg B(\alpha)$ 为真. 显然,只有可能实现又不是必然为真的 α,才是理性 Agent 值得去意图的. 可以考虑增加以下公理[6]:

(A7) 意图可实现性公理:$I(\alpha)\rightarrow\neg B(\neg\alpha)$.

(A8) 意图非平凡性公理:$I(\alpha)\rightarrow\neg B(\alpha)$.

引理 5. 设 $F=\langle W,R_b,R_i^e,R_i^f\rangle$ 是任一框架.

(1) (A7)关于 F 是有效的,当且仅当对任意 $w\in W$,若 $R_i^e(w)\cap R_i^f(w)=\varnothing$,则有 $R_b(w)\cap R_i^e(w)\neq\varnothing$.

(2) (A8)关于 F 是有效的,当且仅当对任意 $w\in W$,若 $R_i^e(w)\cap R_i^f(w)=\varnothing$,则有 $R_b(w)\cap R_i^f(w)\neq\varnothing$.

证明:(1) 若 $F\models I(\alpha)\rightarrow\neg B(\neg\alpha)$,而 $\exists\, w1\in W$,使得 $R_i^e(w1)\cap R_i^f(w1)=\varnothing$ 且 $R_b(w1)\cap R_i^e(w1)=\varnothing$. 对任一命题 p,定义一个 L_{PA}-赋值 $V1$,使得当 $w'\in R_i^e(w1)$ 时,$V1(p,w')=1$,否则 $V1(p,w')=0$;于是 $V1(I(p),w1)=1$. 另一方面,因为 $R_b(w1)\cap R_i^e(w1)=\varnothing$,对任一 $w''\in R_b(w1)$,有 $V1(p,w'')=0$,$V1(\neg p,w'')=1$,从而 $V1(B(\neg p),w1)=1$,$V1(\neg B(\neg p),w1)=0$. 因此,$V1(I(p)\rightarrow\neg B(\neg p),w1)=0$,与对任意公式 α,$F\models I(\alpha)\rightarrow\neg B(\neg\alpha)$ 相矛盾. 当在 $R_b(w1)$,$R_i^e(w1)$,$R_i^f(w1)$ 中有空集时不影响证明.

对任意 $w\in W$,若 $R_i^e(w)\cap R_i^f(w)=\varnothing$,有 $R_b(w)\cap R_i^e(w)\neq\varnothing$,那么存在两种情形:

(a) 对任意 $w\in W$,$R_i^e(w)\cap R_i^f(w)\neq\varnothing$,那么对任一 $w\in W$,任一 L_{PA}-赋值 V,有 $V(I(\alpha),w)=0$,从而 $F\models I(\alpha)\rightarrow\neg B(\neg\alpha)$.

(b) 若 $F\not\models I(\alpha)\rightarrow\neg B(\neg\alpha)$,则存在一个 L_{PA}-赋值 $V1$,和 $w1\in W$,使得 $V1(I(\alpha),w1)=1$ 且 $V1(\neg B(\neg\alpha),w1)=0$. 由 $V1(I(\alpha),w1)=1$,有 $R_i^e(w1)\cap R_i^f(w1)=\varnothing$,从而 $R_b(w1)\cap R_i^e(w1)\neq\varnothing$. 设 $w'\in R_b(w1)\cap R_i^e(w1)$,于是由 $V1(I(\alpha),w1)=1$,有 $V1(\alpha,w')=1$,而由 $V1(\neg B(\neg\alpha),w1)=0$,有 $V1(B(\neg\alpha),w1)=1$,从而 $V1(\neg\alpha,w')=1$,$V1(\alpha,w')=0$;矛盾.

对(2)的证明类似,从略. □

将(A7)和(A8)加入系统 A-BI2,得到系统 A-BI. 将对任意 $w\in W$,若 $R_i^e(w)\cap R_i^f(w)=\varnothing$,则有 $R_b(w)\cap R_i^e(w)\neq\varnothing$ 且 $R_b(w)\cap R_i^f(w)\neq\varnothing$ 的约束加入框架类 $F2$ 中,得到框架类 $F3$. 由定理 5 和引理 5 不难证明如下定理:

定理 6. 系统 A-BI 关于框架类 $F3$ 是可靠和完备的.

系统 A-BI 就是我们所需要的 Agent-BDI 逻辑系统.

4 A-BI 系统的合理性

Agent 系统是有限自治系统. Bratman[9]提出,合理的有限自治系统必须满足以下必要条件:"反对称论题"和"无副作用原理". 反对称论题是说,理性 Agent 以某一行动为意图,同时又相信它不能实现是不合理的,即

Agent应该相信其意图是可行的,称为信念-意图相容性;另一方面,一个 Agent 以某一行动为意图,但不必相信它最终一定会实现是合理的,即信念-意图不完全性.这种相容性与不完全性的不对称性称为反对称论题.形式表示为:

(A9) $\forall M$ 有 $M\models I(\alpha)\wedge B(\neg\alpha)$, 信念-意图相容性

(A10) $\exists M$ 使 $M\models I(\alpha)\wedge\neg B(\alpha)$, 信念 意图不完全性

无副作用原理是说,以 α 为意图的 Agent 不应被迫以 α 的逻辑结论为意图.形式表示为:

(A11) $\exists M$ 有 $M\models(\alpha\rightarrow\beta)\wedge I(\alpha)\wedge\neg I(\beta)$.

Rao 和 Georgeff[10]建议增加与副作用相关的非迁移原理作为合理性约束之一,形式表示为:

(A12) $\exists M$ 有 $M\models B(\alpha)\wedge\neg I(\alpha)$.

不难验证 A-BI 系统满足上述各合理性必要条件,即有以下定理:

定理 7. 在 A-BI 系统中,(A9),(A10),(A11),(A12)成立.

当然,在 A-BI 系统中信念、意图还具有其他一些性质,讨论从略.

5 结论和进一步的工作

本文建立了 Agent-BDI 逻辑的代表系统 A-BI,它是含有正规模态算子信念和非正规模态算子意图的混合模态逻辑系统,不存在逻辑全知问题以及由此带来的副作用等问题.特别是给出了非正规模态算子基于标准(正规)可能世界的语义解释,解决了可靠性和完备性问题.与 Kripke 的基于标准可能世界的语义解释恰当地刻画了含有公理 K(A2)和规则 N(R2)的正规模态算子相类似,定理 2 和定理 4 表明,本文给出的语义解释恰当地刻画了含有规则(R3)的非正规模态算子.通过给出相应的公理和语义约束,A-BI 系统恰当地表征了信念与意图的本质与内在联系,可作为 Agent 形式化研究的逻辑工具.

可将 A-BI 系统推广到多 Agent 的情形,为形式化描述多 Agent 系统提供逻辑工具.特别是 A-BI 系统是一种示例系统,为了描述 Agent 的其他意识属性和动态特征,不难引入其他正规或非正规模态算子和时态算子,也不难引入其他必要的公理和相应的语义约束.

参考文献

1 Bradshaw M. An introduction to software agents. In: Bradshaw M ed. Software Agents. Menlo Park, CA: AAAI Press, 1997. 3~46

2 Singh M P. Multiagent Systems: a Theoretical Framework for Intention, Know-How, and Communication. Berlin: Springer-Verlag, 1994

3 Wooldridge M, Jennings N R. Intelligent agents: theory and practice. The Knowledge Engineering Review, 1995,10(2): 115~152

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

5 Rao A S, Georgeff M P. Modeling rational agents within a BDI architecture. In: Allen J, Fikes R, Sandewall E eds. Principles of Knowledge Representation and Reasoning: Proceedings of the 2nd International Conference (KR-91). San Mateo, CA: Morgan Kaufmann Publishers, Inc., 1991. 473~484

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

7 Hu Shan-li, Shi Chun-yi. A semantic interpretation for agent's non normal modal operators. Journal of Computer Research and Development, 1999,36(10):1153~1157
(胡山立,石纯一.适用于 Agent 非正规模态算子的一种语义解释.计算机研究与发展,1999,36(10):1153~1157)

8 Hu Shan-li, Shi Chun-yi. An intention model for agent. Journal of Software, 2000,11(7):965~970
(胡山立,石纯一.Agent 的意图模型.软件学报,2000,11(7):965~970)

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

10 Rao A S, Georgeff M P. Asymmetry thesis and side-effect Problems in linear- time and branching-time intention logic. In: Sridharan N S ed. Proceedings of the 12th International Joint Conference on Artificial Intelligence. San Mateo, CA: Morgan Kaufmann Publishers, Inc., 1991. 498～504

Agent-BDI Logic

HU Shan-li[1] SHI Chun-yi[2]

[1](*Department of Computer Science and Technology Fuzhou University Fuzhou* 350002)

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

Abstract In this paper, it is demonstrated that the logic tool used in agent formalized depiction should be the mixed modal logic which has both normal and non-normal modal operators. Then a logic system A-BI is built for Agent-BDI logic and its semantics and axiom system are discussed. Especially for non-normal modal operators a new semantic interpretation based on Kripke's normal possible worlds is presented. It is proved that A-BI logic system is sound and complete. A-BI logic appropriately depicts the essence and relation of belief and intension, and can be used as logic tool in formalized research on agent.

Key words Agent, Agent-BDI model, modal logic, belief, intention.