

## Agent 意图的双子集语义改进模型\*

胡山立<sup>1,3+</sup>, 石纯一<sup>2</sup>

<sup>1</sup>(福州大学 计算机科学与技术系,福建 福州 350002)

<sup>2</sup>(清华大学 计算机科学与技术系,北京 100084)

<sup>3</sup>(中国科学院 计算机科学重点实验室,北京 100080)

### An Improved Twin-Subset Semantic Model for Intention of Agent

HU Shan-Li<sup>1,3+</sup>, SHI Chun-Yi<sup>2</sup>

<sup>1</sup>(Department of Computer Science and Technology, Fuzhou University, Fuzhou 350002, China)

<sup>2</sup>(Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)

<sup>3</sup>(Laboratory for Computer Science, The Chinese Academy of Sciences, Beijing 100080, China)

+ Corresponding author: Phn: +86-591-87529249, E-mail: husl@fzu.edu.cn, <http://www.fzu.edu.cn>

**Hu SL, Shi CY. An improved twin-subset semantic model for intention of Agent. *Journal of Software*, 2006, 17(3):396-402. <http://www.jos.org.cn/1000-9825/17/396.htm>**

**Abstract:** Intention, a crucial part of the mental states of an agent, plays an important role in determining the behaviors of the rational agent. In order to eliminate the flaws with existing logic of intention and establish a suitable semantic representation for intention, this paper addresses the requests for intention semantics on formal frameworks of rational agents, points out the problems with existing logic of intention, introduces a novel possible world semantics for intention, called the improved twin-subset semantics, which was developed by us recently based on the previous work about the true-false subset semantics, and presents its application in the formalization of intention for agent. It is also proved that some desired properties could be obtained through some restrictions to the algebraic structure of the models. In the two-value logic, the false value is as important as the true value. Of course, to one proposition, once the true values are described in some possible worlds, the false values are just the rest. However, to a set of propositions (the agent intends to do), descriptions to the false values are as important as descriptions to the true values. The representation of intention should describe a set of propositions (the propositions agent intends to achieve), but in the possible world semantics of classical normal modal operator, it merely pays attention to the true-value, which is denoted as  $R_T(w)$ , and can be considered as single-subset semantics. In the improved twin-subset semantics, the false-value is considered as important as the true-value, and  $R_{IT}(w)$  is used to describe the true, while  $R_{IF}(w)$  the false. Thus this method can describe modal operator in two-value logic completely. In addition, the possible world semantics of the classical normal modal operator can be regarded as the degradation of the improved twin-subset semantics when  $R_{IF}(w)=\emptyset$ . It not only avoids the logical omniscience problem and other related problems (such as side-effect problem, and etc) but also overcomes the shortcomings of

\* Supported by the National Natural Science Foundation of China under Grant Nos.60373079, 60496323, 60573076 (国家自然科学基金)

Received 2004-02-13; Accepted 2005-07-11

the true-false subset semantics and twin-subset semantics. Compared with Konolige and Pollack's model of intention, this model not only is simpler and more natural, but also satisfies the K-axiom and the Joint Consistency. Actually, the improved twin-subset semantics provides a new method for semantic representation of non-normal modal operator and can be applied to establish a new proper agent's logic system.

**Key words:** Agent; intention; semantics; twin-subset semantics; modal operator; model

**摘要:** 意图是 Agent 的一个关键的意识属性,在决定理性 Agent 的行为中起着重要作用.为了克服现有意图逻辑中存在的缺陷,建立了适用于意图的语义表示.讨论了理性 Agent 性态的形式化中对意图语义的要求以及现有意图逻辑中存在的问题.介绍了在前期工作——真假子集语义基础上开发的双子集语义改进模型及其在 Agent 意图形式化中的应用,并且证明通过对模型的代数结构施加一定的约束,能获得许多希望得到的性质.在二值逻辑中,真和假是同等重要的.当然,对一个命题,描述了真值也就知道了假值,但对于一类命题却不是这样,对假值的刻画与对真值的刻画具有同等重要的意义.而对意图的描述是对一类命题(Agent 意图实现的命题)的刻画.经典的正规模态算子的可能世界语义只重视真,用  $R_I(w)$  来描述,可看成是单子集语义.而改进的双子集语义真假并重,用  $R_{IT}(w)$  来描述真,并用  $R_{IF}(w)$  来描述假,从而能更全面地描述二值逻辑中的模态算子.经典的正规模态算子的可能世界语义可以看成是改进的双子集语义当  $R_{IF}(w)=\emptyset$  时的退化情形.改进的双子集语义不仅避免了基于正规模态逻辑表示的“逻辑全知”问题以及由此带来的副作用等问题,与 Konolige 和 Pollack 的意图模型相比,比较简单、自然,且满足 K 公理和联合一致性原理,而且克服了前期工作真假子集语义和双子集语义表示的缺陷.实际上,改进的双子集语义为非正规模态算子的语义表示提供了一种新的方法,可应用于建立新的合适的 Agent 逻辑系统.

**关键词:** Agent;意图;语义;双子集语义;模态算子;模型

中图法分类号: TP18 文献标识码: A

在多 Agent 系统(multi-Agent system)中如何抽象和模型化 Agent,已成为当前人工智能(AI)、软件工程等领域的一项重要研究课题.在 AI 领域,人们通常基于意向立场(intentional stance)来研究 Agent.从意向立场出发,一般把信念(belief)、愿望(desire)和意图(intention)当作 Agent 的基本意识属性(简称 BDI).信念、愿望和意图逻辑在有关理性 Agent 设计的 AI 文献中已受到充分重视<sup>[1-5]</sup>,其重要性在于:在 Agent 性态的形式化中作为逻辑框架.这样的逻辑需要一个建立在涉及知识和信念的认识逻辑上的形式基础,因此,大量的研究常以正规可能世界的方法为基础.人们通常把信念、愿望和意图模型化为可能世界语义下的正规模态算子<sup>[2-5]</sup>,从而不可避免地存在这种方法所固有的一些问题.而意图是抽象和模型化 Agent 的一个不可缺少的意识属性,在决定理性 Agent 的行为时起重要作用.意图决定了对未来行动的选择,从而使一个有限自治系统维持信念、目标、规划和行动的自我平衡.近年来,关于意图的形式化研究已成为哲学逻辑、人工智能、计算语言学和软件工程共同关心的重要课题.在实用推理、认知建模、自然语言处理以及多 Agent 系统的规划、协商、合作与对抗等方面均有着重要的应用.把意图模型化为可能世界语义下的正规模态算子存在以下几个主要问题(其中  $I$  是意图算子):

- (1) 逻辑全知问题:  $\models \varphi \Rightarrow \models I(\varphi)$ .
- (2) 重言隐含下的副作用问题:  $\models \varphi \rightarrow \psi \Rightarrow \models I(\varphi) \rightarrow I(\psi)$ .
- (3) 析取扩大化问题:  $\models I(\varphi) \rightarrow I(\varphi \vee \psi)$ .
- (4) 合取分离问题:  $\models I(\varphi \wedge \psi) \rightarrow I(\varphi) \wedge I(\psi)$ .

人们已经认识到这些问题在实际应用中是不可接受的.经典的正规模态逻辑不适宜作为 Agent 意图的形式化工具<sup>[3,6]</sup>.

针对上述问题, Konolige 和 Pollack<sup>[3]</sup>提出以非正规模态逻辑作为描述工具的意图表示理论(KP 模型),其语义模型是三元组  $M = \langle W, \Sigma, I \rangle$ .其中  $W$  是可能世界集,  $w \in W$  是状态的时序序列,假设所有解释均相对于一个共同的特定时刻 now(语句的求值点).对于每个世界  $w \in W$ ,存在一个求值函数,它确定了在语言  $L$  中语句的值,可能性模

态算子  $\diamond$  的语义是

$$W, w \models \diamond \varphi \text{ iff } \exists w' \in W \text{ 使 } W, w' \models \varphi.$$

必然性模态算子  $\Box \varphi =_{\text{def}} \neg \diamond (\neg \varphi)$ .

对公式  $\varphi$ , 记使  $\varphi$  为真的可能世界集为  $M_\varphi =_{\text{def}} \{w \in W \mid W, w \models \varphi\}$ .  $\Sigma \subseteq W$  用来解释信念, 信念模态算子  $B$  的语义是

$$\langle W, \Sigma, I \rangle \models B(\varphi) \text{ iff } \forall w \in \Sigma \text{ 有 } W, w \models \varphi, \text{ 即 } \Sigma \subseteq M_\varphi.$$

$I \subseteq P(W)$  ( $P(W)$  表记  $W$  的幂集) 用来解释意图, 意图模态算子  $I$  的语义是

$$\langle W, \Sigma, I \rangle \models I(\varphi) \text{ iff } \exists \psi \in I \text{ 使 } \psi = M_\varphi, \text{ 即 } M_\varphi \in I.$$

其实,  $I$  中的每个元素  $M_\varphi$  对应于一个意图, 其作用相当于一个模态算子. 与正规模态逻辑语义解释不同, 在上述解释规则中, 用“ $=$ ”而不是“ $\subseteq$ ”, 这使得 KP 模型避免了逻辑蕴涵副作用. 为了表达意图与信念之间的关系, KP 语义模型有两条约束:

- 1) 可达性:  $\exists w \in \Sigma, \forall M_\varphi \in I$  有  $w \in M_\varphi$ .
- 2) 非平凡性:  $\forall M_\varphi \in I, \exists w \in \Sigma$  使  $w \notin M_\varphi$ .

满足这两条约束的模型称为可采纳的. 有以下性质:

$$\neg I(\varphi \rightarrow \psi) \tag{KP1}$$

$$I(\varphi) \wedge I(\psi) \rightarrow \diamond(\varphi \wedge \psi) \tag{KP2}$$

$$I(\varphi) \rightarrow \neg B(\varphi) \wedge \neg B(\neg \varphi) \tag{KP3}$$

然而, 在 KP 模型中存在如下问题:

- 1) KP 模型中, 用来确定意图的  $I$  是  $W$  幂集的子集, 与信念的表示相比不够自然, 而且复杂度高. 例如, 如果有  $n$  个意图, 确定意图要用  $W$  的  $n$  个子集, 而确定信念只要用  $W$  的一个子集;
- 2) KP 模型不满足  $K$  公理和联合一致性原理(见第 3 节);
- 3) 特别是 KP 模型中用来确定意图的  $I$ , 实际上可看成由所有相互不等价的意图公式组成的集合. 只有与  $I$  中的某个公式等价的公式才是该 Agent 的意图. 因此, 如果  $I(\varphi)$  在该模型中可由  $I(\psi)$  推出, 那么  $\varphi = \psi$ . 这样, 意图推论从该模型中消失了, 也就是该模型失去了非等价意图的推理能力, 所以不可取.

因此, 又有许多学者把同样是经典但非正规的模态逻辑应用到 Agent 意图的形式化中, 并做了许多研究<sup>[6]</sup>. 但是, 作为经典非正规模态逻辑基础的非正规可能世界却是那样的不可思议: 在一个非正规可能世界中, 命题  $\varphi$  与它的否定  $\neg \varphi$  可以同时为真; 一个复合公式的真值却不能由原始命题的真值经联接词的含义复合得到, 它可能为真, 又可能为假. 在现实的可推理系统中, 人们难以接受这样一个不可置信的语义解释来形式化与人们生活密切相关的现实的系统. 文献[7,8]对意图的形式化研究采用多值逻辑的方法, 使用逻辑值  $t, f$  表示意向性真假, 0, 1 表示假定性真假;  $\Delta$  用于抽象意图内容之外的状态, 同样不够自然, 不够直观可信.

为此, 需要寻求一种直观上可信的、能适当地表征意图的直观语义的形式语义. 下面介绍我们开发的双子集语义改进模型及其在 Agent 形式化中的应用, 它仍然把意图形式化为二值逻辑中的模态算子, 仍然基于可能世界语义表示, 却具有我们所需要的性质.

### 1 双子集语义改进模型和直观解释

下面提出的意图的语义模型很好地解决了这个问题, 并满足 KP 模型的所有性质以及人们对意图语义的要求. 为了突出意图的语义表示便于与 KP 系统比较, 在本文中暂不涉及目标或愿望.

为简单计, 这里使用的形式化系统是对经典命题逻辑的扩充, 通过对信念和意图引入相应的模态算子  $B$  和  $I$ , 把它扩充到可能世界框架. 但意图算子  $I$  不是正规的, 且采用我们开发的改进的双子集语义. 从而是简单的混合模态逻辑系统.

定义 1(多模态语言  $L$ ). 设  $Prop = \{p, q, \dots\}$  是可数的原子命题集. 合式公式集  $FML$  递归地定义如下:

- (1) 如果  $p \in Prop$ , 那么  $p \in FML$ ;
- (2) 如果  $\varphi, \psi \in FML$ , 那么  $\neg \varphi \in FML$  且  $\varphi \vee \psi \in FML$ ;

(3) 如果  $\varphi \in FML$ , 那么  $B(\varphi) \in FML$  且  $I(\varphi) \in FML$ .

需要说明的是,算子“ $\neg$ ”(非)和“ $\vee$ ”(或)的语义同经典命题逻辑,其他的经典命题逻辑联结词的定义从略.在语言  $L$  中,引入模态算子  $B$  和  $I$  分别称为信念算子和意图算子.公式  $B(\varphi)$  读作“Agent 相信  $\varphi$ ”,而公式  $I(\varphi)$  读作“Agent 意图  $\varphi$ ”.

定义 2(语义模型  $M$ ). Agent 的语义模型  $M$  是一个元组,  $M = \langle W, R_B, R_{IT}, R_{IF}, V \rangle$ , 其中  $W$  是一个非空集, 可以看成是可能世界集,  $R_B, R_{IT}, R_{IF} \subseteq W \times W$ . 对任意  $w \in W$ ,  $R_B, R_{IT}, R_{IF}$  分别将  $w$  映射到  $w$  的信念、意图(真)、意图(假)可达世界集  $R_B(w), R_{IT}(w), R_{IF}(w)$ , 它们都是  $W$  的非空子集, 且  $R_{IT}(w) \cap R_{IF}(w) = \emptyset$ .  $V$  是赋值函数.

$$V: W \times Prop \rightarrow \{1, 0\}.$$

对公式  $\varphi$ , 当且仅当  $V(w, \varphi) = 1$  (真)时, 称  $\varphi$  在  $w$  为真或有效, 记为  $\langle M, w \rangle \models \varphi$ . 由于在讨论中  $V$  并不重要, 所以在讨论中略去. 增加的模态算子  $B$  和  $I$  的解释由以下定义给出.

定义 3. 对任意  $w \in W$ :

$$\langle M, w \rangle \models p \text{ where } p \in Prop, \text{ iff } V(w, p) = 1.$$

$$\langle M, w \rangle \models \neg \varphi \text{ iff } \langle M, w \rangle \not\models \varphi$$

$$\langle M, w \rangle \models \varphi \vee \psi \text{ iff } \langle M, w \rangle \models \varphi \text{ or } \langle M, w \rangle \models \psi$$

$$\langle M, w \rangle \models B(\varphi) \text{ iff } \forall w' \in W \text{ if } w' \in R_B(w) \text{ then } \langle M, w' \rangle \models \varphi.$$

$$\langle M, w \rangle \models I(\varphi) \text{ iff}$$

$$(1) \forall w' \in W, \text{ if } w' \in R_{IT}(w) \text{ then } \langle M, w' \rangle \models \varphi; \text{ and}$$

$$(2) \exists w_1'' \in W \text{ such that } w_1'' \in R_{IF}(w) \text{ and } \langle M, w_1'' \rangle \models \neg \varphi; \text{ and}$$

$$(3) \exists w_2'' \in W \text{ such that } w_2'' \neq w_1'', w_2'' \in R_{IF}(w) \text{ and } \langle M, w_2'' \rangle \models \neg \varphi.$$

信念和意图的语义的直观解释: 在世界  $w$ , 信念表示 Agent 对某些可能世界(即  $R_B(w)$ )的“偏爱”, 只有对这些可能世界上均为真的命题, Agent 才是相信的; 而意图也表示 Agent 对某些可能世界的“重视”. 不过, 与信念不同的是, 这些被重视的可能世界被分成两个不相交的部分  $R_{IT}(w)$  和  $R_{IF}(w)$ : 在  $R_{IT}(w)$  上为真的命题, Agent 认为是可能实现的或已实现的; 而在  $R_{IF}(w)$  上部分为假(至少两个:  $w_1, w_2$ )的命题, Agent 认为是当前尚未实现的且不是必然实现的. 显然, 只有尚未实现的不是必然实现而又可能实现的命题才是理性 Agent 值得去意图的.

在世界  $w$ , 定义中信念仍用  $W$  的一个子集  $R_B(w)$  来表示, 意图用  $W$  的两个不相交的子集  $R_{IT}(w)$  和  $R_{IF}(w)$  来表示, 而不是像 KP 系统那样用  $W$  的幂集的子集来表示(或当有  $n$  个意图时, 用  $W$  的  $n$  个子集来表示). 由于意图要满足的性质与信念不同, 容易理解意图不可能用  $W$  的一个子集来表示. 因此, 上述表示应是意图基于可能世界的语义表示的最好结果. 当然, 还需要证明这种表示具有所期望的性质.

## 2 模型的合理性

Bratman<sup>[1]</sup> 提出有限自治系统必须满足合理性的必要条件: “反对称论题”和“无副作用原理”. 反对称论题是说, 允许一个 Agent 以某一行动为意图, 同时又相信它不能实现是不合理的, 即信念-意图一致性; 另一方面, 一个 Agent 以某一行动为意图, 但不相信它能实现是合理的, 即信念-意图不完全性. 这种一致性与不完全性的不对称性称为反对称论题. 形式表示为

$$\forall M \text{ 有 } M \models I(\varphi) \wedge B(\neg \varphi), \text{ 一致性} \tag{A11}$$

$$\exists M \text{ 使 } M \models I(\varphi) \wedge \neg B(\varphi), \text{ 不完全性} \tag{A12}$$

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

$$\exists M \text{ 有 } M \models (\varphi \rightarrow \psi) \wedge I(\varphi) \wedge \neg I(\psi) \tag{A13}$$

Rao 和 Georgeff<sup>[9]</sup> 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

$$\exists M \text{ 有 } M \models B(\varphi) \wedge \neg I(\varphi) \tag{A14}$$

当然, 意图还应满足其他一些性质. 意图是对将来可能实现的目标的承诺, 意图除了要满足无副作用原理外, 还应满足:

$K$  公理:  $\forall M$  有  $M \models I(\varphi \rightarrow \psi) \rightarrow (I(\varphi) \rightarrow I(\psi))$  (A15)

一致性原理:  $\forall M$  有  $M \models \neg I(\varphi \wedge \neg \varphi)$  (A16)

联合一致性原理:  $\forall M$  有  $M \models I(\varphi) \wedge I(\psi) \rightarrow I(\varphi \wedge \psi)$  (A17)

联合不完全性原理:  $\exists M$  使  $M \models I(\varphi \vee \psi) \wedge \neg I(\varphi \wedge \psi)$  (A18)

$\exists M$  使  $M \models I(\varphi \wedge \psi) \wedge \neg I(\varphi) \wedge \neg I(\psi)$  (A19)

信念与意图除了满足反对称论题和非迁移原理外,还应满足

非平凡性原理:  $\forall M$  有  $M \models I(\varphi) \rightarrow \neg B(\varphi)$  (A110)

正规模态逻辑系统和一般非正规模态逻辑系统都有  $K$  公理,认识论模态逻辑系统也有  $K$  公理.意图应该满足  $K$  公理也是容易理解的:假设理性 Agent 意图  $\varphi$ ,且意图  $\varphi \rightarrow \psi$ ,当然它意图  $\psi$ ,否则它就不应该意图  $\varphi \rightarrow \psi$ 了.

意图要满足一致性原理 A16 不言而喻,理性 Agent 当然不应有互相对立的意图.意图要满足联合一致性原理 A17 是因为 Agent 意图  $\varphi$ 且意图  $\psi$ ,当然意图  $\varphi \wedge \psi$ .注意,这和“意图  $\varphi$  或  $\psi$ ,不一定要意图  $\varphi \wedge \psi$ ”完全不同,后者的形式化是 A18.与“意图  $\varphi \wedge \psi$ ,不一定要意图  $\varphi$ ,也不一定要意图  $\psi$ ”不同,后者的形式化是 A19.应该注意它们的区别.我们把 A18 和 A19 称为联合不完全性原理.信念与意图还应满足非平凡性原理 A110.这是因为理性 Agent 意图  $\varphi$ ,那么它不相信没有它的行动, $\varphi$  能自然实现,否则就不必意图  $\varphi$ 了.

由于用双子集  $R_{IT}(w), R_{IF}(w)$  来刻画意图,从而当  $R_{IF}(w) \neq \emptyset$  时,恒真的命题(或公式)不被意图,即对意图不存在引言中提到的逻辑全知问题.并且容易证明:在模型  $M$  中,性质 A12, A13, A15~A19 成立.对意图不存在引言中提到的重言隐含下的副作用等其他 3 个问题,也就是说,我们有:

定理 1. 在模型  $M$  中,对意图不存在引言中提到的逻辑全知等 4 个问题,且性质 A12, A13, A15~A19 成立.

证明:

(1) 如果  $\langle M, w \rangle \models I(\varphi)$ ,那么  $\exists w_1'' \in W$  such that  $w_1'' \in R_{IF}(w)$  and  $\langle M, w_1'' \rangle \models \neg \varphi$ ,从而恒真的命题(或公式)不被意图,即对意图不存在逻辑全知问题.

(2) 设  $W = \{w_1, w_2, w_3\}$ ,对任意  $w \in W$ ,取  $R_{IT}(w) = \{w_1\}, R_{IF}(w) = \{w_2, w_3\}, V(w_1, \varphi) = V(w_1, \psi) = 1, V(w_2, \varphi) = V(w_2, \psi) = 0, V(w_3, \varphi) = 0$ ,而  $V(w_3, \psi) = 1$ .容易验证对意图不存在重言隐含下的副作用等其他 3 个问题,且性质 A13 成立.

(3) 设  $W = \{w_1, w_2, w_3, w_4\}$ ,对任意  $w \in W$ ,取  $R_B(w) = \{w_1, w_3\}, R_{IT}(w) = \{w_1, w_2\}, R_{IF}(w) = \{w_3, w_4\}, V(w_1, \varphi) = V(w_2, \varphi) = 1, V(w_3, \varphi) = V(w_4, \varphi) = 0$ .容易验证性质 A12 成立.

(4) 设  $W = \{w_1, w_2, w_3\}$ ,对任意  $w \in W$ ,取  $R_{IT}(w) = \{w_1\}, R_{IF}(w) = \{w_2, w_3\}, V(w_1, \varphi) = 1, V(w_1, \psi) = 0, V(w_2, \varphi) = V(w_2, \psi) = 0, V(w_3, \varphi) = V(w_3, \psi) = 0$ .容易验证性质 A18 成立.

(5) 设  $W = \{w_1, w_2, w_3\}$ ,对任意  $w \in W$ ,取  $R_{IT}(w) = \{w_1\}, R_{IF}(w) = \{w_2, w_3\}, V(w_1, \varphi) = V(w_1, \psi) = 1, V(w_2, \varphi) = 1, V(w_2, \psi) = 0, V(w_3, \varphi) = 0, V(w_3, \psi) = 1$ .容易验证性质 A19 成立.

为了反映意图和信念之间的关系,使意图满足与信念有关的性质,应对模型的代数结构施加一定的约束.Rao 和 Georgeff 的系统<sup>[4]</sup>靠约束意图世界是信念世界的子世界以及每一个信念世界必存在一个意图子世界来部分地实现.Konolige 和 Pollack 的 KP 系统<sup>[3]</sup>对其语义模型要求满足可达性与非平凡性.对此,我们提出以下更简明的约束条件,称为可实现性与非平凡性.

定义 4. 模型  $M = \langle W, R_B, R_{IT}, R_{IF}, V \rangle$  称为可采纳的,当且仅当它是可实现的: $\forall w \in W, R_B(w) \cap R_{IT}(w) \neq \emptyset$  且是非平凡的: $R_{IF}(w) \subseteq R_B(w)$ .

定理 2. 在可采纳的模型  $M$  中,性质 A11, A110 成立.

证明:

(1) 在可采纳的模型  $M$  中, $\forall w \in W, R_B(w) \cap R_{IT}(w) \neq \emptyset$ ,从而  $\forall M$  有  $M \models I(\varphi) \wedge B(\neg \varphi)$ ,性质 A11 成立.

(2) 在可采纳的模型  $M$  中, $R_{IF}(w) \subseteq R_B(w)$ ,从而  $\forall M$  有  $M \models I(\varphi) \rightarrow \neg B(\varphi)$ ,性质 A110 成立.

### 3 比较与分析

本文讨论了意图的语义要求,并提出了一种新的意图模型,这种模型有直观的基于可能世界的语义解释,能

够满足对意图语义的要求(AI1~AI10).与 Cohen 和 Levesque 的工作<sup>[2]</sup>以及 Rao 和 Georgeff 的工作<sup>[4]</sup>相比,避免了“逻辑全知”问题以及由此带来的副作用等问题.与 Konolige 和 Pollack 的工作<sup>[3]</sup>相比,意图的表示简单、自然,有基于可能世界的语义解释,且满足  $K$  公理(AI5)和联合一致性原理(AI7).

Konolige 和 Pollack<sup>[3]</sup>的 KP 模型不满足  $K$  公理 AI5.因为当  $M_\phi \in I$  且  $M_{\phi \rightarrow \psi} \in I$  时,不一定有  $M_\psi \in I$ .KP 模型也不满足联合一致性原理 AI7.因为当  $M_\phi \in I$  且  $M_\psi \in I$  时,不一定有  $M_{\phi \wedge \psi} \in I$ .

与文献<sup>[7,8]</sup>采用的多值逻辑方法相比,我们的模型不仅简单、自然,而且仍然采用经典的二值逻辑和可能世界语义方法,从而具有可信的理论基础.

本文提出的这种改进的双子集语义不仅克服了真假子集语义<sup>[10,11]</sup>不能同时意图  $\phi \rightarrow \psi$  和  $\phi$  的缺陷,而且克服了双子集语义<sup>[12]</sup>不满足联合不完全性原理中的  $A19: \exists M$  使  $M \models I(\phi \wedge \psi) \wedge \neg I(\phi) \wedge \neg I(\psi)$  的不足.具体讨论从略.

我们提出的这一方法还具有一般性,不仅适用于 Agent BDI 结构中的信念、愿望或目标等的形式表示,而且为非正规模态算子找到了一种基于正规可能世界的语义表示.

从 Agent 的意图属性到它的形式语义,这是个抽象.传统的 Kripke 可能世界语义和正规模态逻辑方法用一个子集  $I$  来描述意图(正像我们用一个子集  $R_B(w)$  来描述信念一样).由于只描述了真值,存在这种方法所固有的一些问题.我们的模型用两个子集  $R_{IT}(w)$  和  $R_{IF}(w)$  来描述意图,既描述了真值,又描述了假值,从而解决了问题.

特别地,我们认为在二值逻辑中真和假是同等重要的.当然,对于一个命题,描述了真值也就知道了假值.但对一类命题却不是这样,对假值的刻画与对真值的刻画具有同等重要的意义.而对意图的描述是对一类命题(Agent 意图实现的命题)的刻画.经典的正规模态算子的可能世界语义只重视真,用  $I$  来描述,可看成是单子集语义.而改进的双子集语义真假并重,用  $R_{IT}(w)$  来描述真,并用  $R_{IF}(w)$  来描述假,从而能更全面地描述二值逻辑中的模态算子.上述改进的双子集语义对意图算子的应用就是很好的例子.另外,经典的正规模态算子的可能世界语义可以看成是改进的双子集语义当  $R_{IF}(w) = \emptyset$  时的退化情形.

上述讨论表明,改进的双子集语义为非正规模态算子提供了一个新的合适的语义表示,是对经典的正规模态算子可能世界语义的一个重要发展,是理性 Agent 性态的形式化的有力工具,可应用于建立新的合适的 Agent 逻辑系统.

## References:

- [1] Bratman ME. Intentions, Plans and Practical Reason. Cambridge: Harvard University Press, 1987.
- [2] Cohen PR, Levesque HJ. Intention is choice with commitment. *Artificial Intelligence*, 1990,42(2-3):213-261.
- [3] Konolige K, Pollack ME. A representationalist theory of intention. In: Bajcsy R, ed. *Proc. of the 13th Int'l Joint Conf. on Artificial Intelligence*. San Mateo: Morgan Kaufmann Publishers, 1993. 390-395.
- [4] Rao AS, Georgeff MP. Modeling rational agents within a BDI architecture. In: Allen J, Fikes R, Sandewall E, eds. *Principles of Knowledge Representation and Reasoning: Proc. of the 2nd Int'l Conf. (KR-91)*. San Mateo: Morgan Kaufmann Publishers, 1991. 473-484.
- [5] Rao AS, Georgeff MP. The semantics of intention maintenance for rational agents. In: Mellish SC, ed. *Proc. of the 14th Int'l Joint Conf. on Artificial Intelligence*. San Mateo: Morgan Kaufmann Publishers, 1995. 704-710.
- [6] Cavedon L, Padgham L, Rao A, Sonenberg E. Revisiting rationality for agents with intentions. In: Yao X, ed. *Proc. of the 8th Australian Joint Conf. on Artificial Intelligence*. Singapore: World Scientific Publishing Co. Pet. Ltd., 1995. 131-138.
- [7] Chen X, Liu GA. Logic of intention. In: Dean T, ed. *Proc. of the IJCAI'99*. San Mateo: Morgan Kaufmann Publishers, 1999. 172-177.
- [8] Zhu ZR, Chen XP, Zhou Y. A new formal system of intention consequences  $L_{m5c}$ . *Journal of Software*, 2002,13(7):1271-1277 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/13/1271.htm>
- [9] Rao AS, Georgeff MP. Asymmetry thesis and side-effect problems in linear-time and branching-time intention logic. In: *Proc. of the 12th Int'l Joint Conf. on Artificial Intelligence*. San Mateo: Morgan Kaufmann Publishers, 1991. 498-504.
- [10] Hu SL, Shi CY. Agent-BDI logic. *Journal of Software*, 2000,11(10):1353-1360 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/11/1353.pdf>

- [11] Hu SL, Shi CY. Agent logic and the true-false subset semantics. Journal of Software, 2002,13(11):2112-2115 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/13/2112.htm>
- [12] Hu SL, Shi CY. Twin-Subset semantic model for intention. In: Proc. of the 2nd Int'l Conf. on Machine Learning and Cybernetics. IEEE, 2003. 2004-2008.

#### 附中文参考文献:

- [8] 朱紫汝,陈小平,周熠. 一个新的意图后承形式系统  $L_{m5c}$ . 软件学报,2002,13(7):1271-1277. <http://www.jos.org.cn/1000-9825/13/1271.htm>
- [10] 胡山立,石纯一. Agen-BDI 逻辑. 软件学报,2000,11(10):1353-1360. <http://www.jos.org.cn/1000-9825/11/1353.htm>
- [11] 胡山立,石纯一. Agent 逻辑和真假子集语义. 软件学报,2002,13(11):2112-2115. <http://www.jos.org.cn/1000-9825/13/2112.pdf>



胡山立(1944 - ),男,福建福州人,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 - ),男,教授,博士生导师,CCF 高级会员,主要研究领域为人工智能应用基础.



## 第 6 届中国 Rough 集与软计算学术研讨会(CRSSC2006)

### 征文通知

由中国人工智能学会粗糙集与软计算专业委员会和中国计算机学会人工智能与模式识别专业委员会主办、浙江师范大学承办的“第六届中国 Rough 集与软计算学术研讨会”(CRSSC2006)拟定于 2006 年 10 月 30 日至 11 月 3 日在浙江金华召开。

#### 一、征文范围

|              |           |           |            |
|--------------|-----------|-----------|------------|
| Rough 集理论及应用 | 计算智能      | 机器学习      | 文字计算       |
| Fuzzy 集理论及应用 | 粒计算       | 软计算及其应用   | 演化计算       |
| Petri 网      | 软计算的逻辑基础  | 非经典逻辑     | 神经网络       |
| 计算复杂性        | 空间推理      | 统计与概率推理   | 智能 Agent   |
| 多准则决策分析      | 决策支持系统    | 知识发现与数据挖掘 | 多 Agent 技术 |
| 近似推理与不确定性推理  | 网络智能      | 集成智能系统    | 数据仓库       |
| 模式识别与图像处理    | 生物信息与生物计算 | 认知信息学     | 其他有关领域     |

#### 二、征文要求

详见会议网站：<http://cs.cqupt.edu.cn/crssc/crssc2006>

#### 三、重要日期

截稿日期(收到)：2006 年 4 月 31 日

录用日期(发出)：2006 年 6 月 10 日

论文清样付印和论文注册截止日期(收到)：2006 年 7 月 10 日

#### 四、联系方式

- a) 联系人(联系电话)：梁久祯(0579-2298258)；王基一(0579-2283436)；吴小红(0579-2298903)
- b) 电子信箱：liangjz@zjnu.cn(梁久祯), xx51@zjnu.cn(王基一)