

# 隐式愿望及其形式化\*

周熠<sup>†</sup>, 陈小平

(中国科学技术大学 计算机科学与技术系, 安徽 合肥 230027)

## Implicit Desire and Its Formalization

ZHOU Yi<sup>†</sup>, CHEN Xiao-Ping

(Department of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, China)

+ Corresponding author: Phn: +86-551-3606724, E-mail: zyz@mail.ustc.edu.cn, http://ai.ustc.edu.cn

Received 2003-12-30; Accepted 2004-11-03

Zhou Y, Chen XP. Implicit desire and its formalization. *Journal of Software*, 2005,16(5):771-778. DOI: 10.1360/jos160771

**Abstract:** The study of “Implicit Desire” is essential to agent theory. This paper first analyzes insufficiency and inadequacy of the existing theories of implicit desire. Then logical semantics is employed to define a new kind of implicit desire—pm-desire consequence, to investigate its main properties, to compare it with the relevant work, and thereby to argue its rationality such as its accordance with intuitions and ability to improve the autonomy intelligent-agents, etc.

**Key words:** intelligent Agent; implicit desire; formalization; BDI model; non-standard logic

**摘要:** “隐式愿望”的刻画是 Agent 理论研究中的一个重要课题. 首先分析现有工作存在的问题, 然后用逻辑语义学方法严格定义一种新的隐式愿望——pm-愿望后承, 研究其主要性质并与相关工作进行比较, 进而论证其合理性, 包括符合直觉、有利于提高智能主体的自主性等等.

**关键词:** 智能主体; 隐式愿望; 形式化; BDI 模型; 非标准逻辑

**中图法分类号:** TP18 **文献标识码:** A

关于理性主体(rational Agents)的研究是近年来人工智能领域的核心问题之一<sup>[1-7]</sup>. 愿望(desire)表达了主体的一种倾向性心态, 是一个理性主体的基本心理状态之一<sup>[8]</sup>. 在人工智能的规划(AI planning)领域<sup>[9]</sup>, 愿望表示主体的目标状态, 规划的目的是找出从初始条件达到目标状态的行动序列以实现主体的愿望. 与以往面向任务(task-oriented)的主体相比, 面向愿望(goal-oriented)的主体更能适应动态不可预测环境.

在 Agent 理论的 BDI 模型<sup>[2-6,8,10]</sup>中, 主体的心理状态(mental states)被区分为 3 种类型: “信念(belief)”、“愿望(desire)”和“意图(intention)”, 其中后两种又可以统称为倾向性心态(motivational attitude). BDI 建模方法通过研究这些状态本身的性质和它们之间的交互作用来把握主体的智能行为. 其中, 愿望用于刻画传统 AI 所称的“目标”, 但范围更广, 比如可以是非强制性的目标. 在基于 BDI 模型的规划中, 愿望被作为行动选择的基本条件:

\* Supported by the National Natural Science Foundation of China under Grant No.60275024 (国家自然科学基金)

**作者简介:** 周熠(1981—), 男, 湖南湘乡人, 博士生, 主要研究领域为人工智能, 多主体系统; 陈小平(1955—), 男, 博士, 教授, 博士生导师, 主要研究领域为人工智能, 多主体系统.

任意行动  $a$  是可选的,当且仅当通过  $a$  达到的状态符合主体的愿望。

马尔可夫决策论(MDP)是另外一种重要的 Agent 理论模型.在该模型中,愿望没有得到显式的表达,它的作用是通过数值型效用(utility)来体现的.由于数值型效用和数值型概率在应用中往往难以获得,而且相应的最优策略的计算效率难以达到实用要求,近年来又提出了将马尔可夫决策与 BDI 建模思想相融合的新方向——定性决策论(qualitative decision theory),愿望的显式表示是其必不可少的核心研究内容<sup>[11,12]</sup>.

在现有文献中,对愿望的研究比较薄弱<sup>[12]</sup>,一些基本问题还没有得到满意的解决.这些问题包括以下几类:第 1 类是倾向性心态的基本表示问题,包括:有哪几种基本的倾向性心态?如何表示它们?第 2 类是倾向性心态的变化问题.包括如何生成和更新/修订/修改等等.第 3 类涉及各种倾向性心态之间的关系,以及与主体信念和行动之间的关系等等.值得注意的是,这几类问题均涉及“隐式愿望”的表示和推理.

通常认为,主体的愿望可以表示为一个命题集合  $D=\{d_1,d_2,\dots\}$ ,其中  $d_i$  表达该主体的一项具体的愿望.由于  $D$  是(事先)显式陈述的,我们称其为“显式愿望”.考虑到智能主体的“有限性”和可能状态的无限性,显式愿望不可能是完备的,即不可能罗列出每一个合乎(或不合乎)主体愿望的状态,必然有一些(甚至大部分)愿望是隐式表达的,我们称其为“隐式愿望”.因此,基于定性决策论以及 BDI 模型的规划必须发挥隐式愿望的作用.

所谓“隐式愿望问题”就是隐式愿望的定义或标准问题,可表述如下:给定主体的显式愿望  $D=\{d_1,d_2,\dots\}$ ;对任何  $d \notin D$ , $d$  是不是该主体的愿望?换言之, $d$  是不是该主体的隐式愿望?

就我们掌握的情况来看,目前对隐式愿望问题有以下 3 种观点:

(1) 必要性定义:任何愿望  $d_i$  的逻辑后承都是愿望.经典的模态语义采取的就是这种观点<sup>[2,10]</sup>.因此, $d$  是该主体的隐式愿望当且仅当由  $D$  可逻辑地推出  $d$ .

(2) 等价性定义:只有与某个  $d_i$  逻辑等价的才是愿望.因此, $d$  是该主体的隐式愿望当且仅当  $d$  逻辑等价于某个  $d_i$ .

另外,还可以按照以下定义来刻画隐式愿望.

(3) 充分性定义:只要逻辑蕴涵某个  $d_i$  的都是愿望.因此, $d$  是该主体的隐式愿望当且仅当由  $d$  可逻辑地推出某个  $d_i$ .

本文将说明,上述 3 种定义都是不充分或不恰当的.本文将对“隐式愿望”做进一步分析,给出一种新的隐式愿望定义,并用逻辑语义学方法加以严格刻画.我们定义 pm-愿望后承,记做  $\Gamma \vdash D(p) \rightarrow D(q)$ ,直观上是指在集合  $\Gamma$  下,命题  $q$  对命题  $p$  “有用无害”.对于隐式愿望问题,我们的回答是,设主体当前的信念(belief)集为  $\Gamma$ ,如果存在  $d_i \in D$ ,使得在本文所给出的系统中  $\Gamma \vdash D(d_i) \rightarrow D(d)$ ,那么  $d$  就是主体的隐式愿望.

本文第 1 节分析指出现有隐式愿望定义的缺陷,并非形式地分析隐式愿望的基本内涵.第 2 节用逻辑语义学的方法,严格定义一种新的隐式愿望并研究其主要性质.第 3 节与相关工作进行比较.最后给出结论.

## 1 理解继承反常现象

先看一个例子:“啤酒和酸奶问题”.假设“主人”让机器人去买啤酒和酸奶,并约定主人的指令就是机器人的愿望.再假定机器人在购买过程中发现,市场上酸奶无货而啤酒促销,买啤酒赠送开啤酒瓶的工具.那么,这个机器人该不该只买啤酒?

从直觉上看,在这种情况下,行动  $a$  = “只买啤酒”是机器人的一个合理的选择,是合乎愿望的.为了论证这一合理性,只需说明两点.第 1,尽管行动  $a$  并没有完全实现机器人的显式愿望(“买啤酒和酸奶”),只实现了其中的“一部分”(“买啤酒”),但在假定的条件下,机器人没有更好的选择了——在“全部实现”已不可能的情况下,“部分实现”就是最好的选择.事实上,一个“现实的”人类个体在同样情况下通常也是这样决策的.第 2,虽然在机器人的显式愿望中并没有“得到开啤酒的工具”这一项,但行动  $a$  产生的这一“副作用”相对于机器人的显式愿望而言,是有用无害的,是“可接受的”.因此,通过行动  $a$  而达到的状态  $p$  = “得到啤酒并且得到开啤酒的工具”是符合机器人的愿望的;换言之, $p$  也是机器人的愿望(严格地说是隐式愿望).

然而,现有 3 种愿望定义都不符合上述直觉判断.用逻辑语言来描述,机器人的显式愿望是  $\{x_1 \wedge x_2\}$ ,其中  $x_1$  表示“得到啤酒”, $x_2$  表示“得到酸奶”.通过行动  $a$  达到的状态(即  $a$  的“效果”)是  $x_1 \wedge y$ ,其中  $y$  表示“得到开啤酒的

工具”。显然,  $x_1 \wedge x_2$  与  $x_1 \wedge y$  之间的关系不满足 3 种现有定义中的任何一种。因此, 根据这 3 种定义,  $x_1 \wedge y$  不是机器人的愿望。所以, 这些愿望定义与我们的直觉是不相符合的。

不仅如此, 这 3 种愿望定义将对基于 BDI 模型的规划产生实际的不良后果。对上面的例子而言, 依照现有 3 种愿望定义, 行动  $a$  将不可能被机器人选择。显然, 这将导致行动机会的丧失和主体决策性能与自主性程度的下降。

从当前人工智能应用的实际需求来看, 现有愿望定义的上述缺陷是一般性的、严重的。“充分性定义”要求任何可行行动“完全”实现某个显式愿望; “必要性定义”要求每一个可行行动不能“突破”任何显式愿望; 而“等价性定义”则要求每一个可行行动只能不多不少地“恰好”实现某个显式愿望。不难看出, 这些愿望定义对智能主体的规划能力、行动能力和显式愿望的表达提出了过高的要求, 是不切实际的。另一方面, 这些定义又对智能主体的自主性强加了不必要的过多限制。因此, 进一步研究“隐式愿望问题”, 给出隐式愿望的恰当定义, 是完全必要的, 也是有普遍意义的。

显式愿望是主体有意识要达到的状态, 而隐式愿望则未必。比如在啤酒和酸奶问题中,  $x_1 \wedge y$  并不是机器人有意识要达到的状态, 只是一种可接受的后果, 而且这个后果对显式愿望的实现是“有帮助的”。我们认为, “可接受”和“有帮助”是隐式愿望的两个最基本的特性; 因此, 我们可以将隐式愿望的基本内涵非形式地概括为“有用无害”。这一概括与人们对愿望本质的通常看法——趋利避害——是一致的。

为了严格规定“有用无害”的内涵, 首先必须明确几个基本观点。第 1, 一个状态是否“有用无害”, 必定是相对于某个显式愿望而言的; 脱离了显式愿望, “有用无害”的判断就失去了根基。第 2, 这种“有用无害”必定是和“语境”密切相关的。这里的“语境”表示显式愿望之外的其他知识。所以, 任给语境  $\Gamma$  和显式愿望  $p$ , 判断一个命题  $q$  是不是相对于  $p$  的隐式愿望, 归结为判断在语境  $\Gamma$  下,  $q$  相对于  $p$  是不是“有用无害的”。我们把这种关系记为  $\Gamma \vdash D(p) \rightarrow D(q)$ , 并称  $q$  为  $p$  的“愿望后承”。第 3,  $p$  对  $q$  “有帮助”在直觉上可以理解为“ $p$  真蕴涵  $q$  部分真”。比如在“啤酒和酸奶问题”中,  $x_1 \wedge y$  真蕴涵  $x_1 \wedge x_2$  的“部分真”(即  $x_1$  的真)。第 4, 仅仅“有帮助”是不够的, 也就是说, “可接受”是一个必要条件。例如, 相对于显式愿望  $x \wedge y, x \wedge \neg y$  虽然“有帮助”, 但同时又有冲突, 因而不能作为隐式愿望。直觉上, “可接受”的基本涵义是“无负面影响”。

基于上述分析, 我们的问题转化为“愿望后承”的形式化。为此, 我们在以往工作<sup>[3,4]</sup>的基础上, 通过推广“归约蕴涵”的概念, 给出“愿望后承”的一个基于非经典逻辑的形式刻画。

## 2 形式刻画

本文在命题语言的范畴内讨论问题。约定  $\Gamma$  为命题公式的集合, 称为前提;  $p, q$  等为命题公式; 所有原子的集合记为  $Atom = \{x_1, x_2, \dots\}$ , 有时用  $x, y$  等表示原子; 命题公式的定义如常;  $\vdash_{cl}$  为经典命题逻辑的语义后承关系。本文用文字的相容集表示赋值。

**定义 1(赋值)**. 一个赋值  $\Pi$  是一个文字的集合, 满足不存在原子  $x$  使  $x \in \Pi$  且  $\neg x \in \Pi$ 。特别地,  $\Pi$  为空集时记为  $\Pi_0$ 。

原子  $x$  在赋值  $\Pi$  中出现表示  $x$  在  $\Pi$  中被赋值为真;  $\neg x$  在  $\Pi$  中出现表示  $x$  被赋值为假; 如果  $x$  和  $\neg x$  都不在  $\Pi$  中出现, 则表示  $x$  未被赋值。因此, 一个赋值实际上是一个部分指派。为方便起见, 有时把赋值如  $\{x_1, \neg x_2, x_4, \dots\}$  表示为  $\{1, 0, *, 1, \dots\}$ , 其中 \* 表示对应文字(这里是  $x_3$ ) 不出现。

**定义 2( $\Gamma$ -模型)**. 任给前提  $\Gamma$  赋值  $\Pi$  是公式  $p$  的  $\Gamma$ -模型, 如果

(1)  $\Gamma \cup \Pi$  可满足;

(2)  $\Gamma \cup \Pi \vdash_{cl} p$ .

$p$  的所有  $\Gamma$ -模型的集合记为  $[p]_{\Gamma}$ 。一般来说,  $[p]_{\Gamma}$  为一无穷集。

$\Gamma$ -模型是经典逻辑语义中“成真指派”概念的推广。一个  $\Gamma$ -模型包含了  $\Gamma$  的作用, 而且可以对某些原子不赋予“真假”, 以此表示即使不考虑或不知道这些原子的真假, 命题  $p$  都成立。例如, 若  $\Gamma$  为空, 则  $x_1 \vee x_2$  的一个  $\Gamma$ -模型是  $\{1, *, *, \dots\}$ , 其中只有  $x_1$  被赋值为真, 其余原子均未被赋值, 这就足以使命题为真了。  $\{1, 0, *, \dots\}$  也是  $x_1 \vee x_2$  的一个  $\Gamma$ -模型, 但  $\{*, *, *, \dots\}$  就不是了。直观上,  $p$  的一个  $\Gamma$ -模型表示在前提  $\Gamma$  下使  $p$  为真的一种“简化”了的“情形”或“环境”。

状况”;这里的“简化”表现为:这些“情形”或“环境状况”中的某些“不必要的属性”被省略了。

集合包含关系直接构成 $\Gamma$ -模型之间的一个偏序关系.称 $\Pi_1$ “小于等于” $\Pi_2$ ,如果 $\Pi_1 \subseteq \Pi_2$ ;称 $\Pi_1$ “小于” $\Pi_2$ ,如果 $\Pi_1 \subset \Pi_2$ ,即 $\Pi_1 \subseteq \Pi_2$ 且 $\Pi_1 \neq \Pi_2$ .

**定义 3(极小 $\Gamma$ -模型).**  $\Pi$ 是公式 $p$ 的极小 $\Gamma$ -模型,如果

- (1)  $\Pi \in [p]_{\Gamma}$ ;
- (2) 不存在 $\Pi' \in [p]_{\Gamma}$ 使 $\Pi'$ 小于 $\Pi$ .

$p$ 的所有极小 $\Gamma$ -模型的集合记为 $(p)_{\Gamma}$ ,易知该集合为 $[p]_{\Gamma}$ 的有限子集.

极小 $\Gamma$ -模型是最大限度地省略了“不必要属性”的 $\Gamma$ -模型.若 $\Pi$ 是 $p$ 的一个极小 $\Gamma$ -模型,则 $\Pi$ 中的每一个元素 $u$ 相对于其他元素而言是 $\Pi$ 的一个“必要属性”——没有它的“帮助”, $\Pi$ 仅靠其他元素将无法使 $p$ 为真;而有了 $u$ 的帮助, $\Pi$ 将使 $p$ 为真.所以,假如 $\Pi$ 和 $\Pi'$ 分别是 $q$ 和 $p$ 的极小 $\Gamma$ -模型,且 $u \in \Pi \cap \Pi'$ ,则 $u$ 不仅对 $q$ ,而且对 $p$ 是“有帮助”的,因此 $\Pi$ 对 $p$ 是“有帮助”的——这意味着, $q$ 在某种“情况”(即 $\Pi$ )下,对 $p$ 是“有帮助”的.假如对 $q$ 的每一个极小 $\Gamma$ -模型 $\Pi$ ,都存在 $p$ 的一个极小 $\Gamma$ -模型 $\Pi'$ ,使得 $\Pi \cap \Pi' \neq \emptyset$ ,则表明 $q$ 在任何“情况”下对 $p$ 都是“有帮助”的,也就是说, $q$ 对 $p$ 是“有帮助”的.

任给赋值 $\Pi$ ,令 $\neg \Pi$ 为由 $\Pi$ 中文字的否定组成的赋值.

**定义 4(愿望后承).**  $q$ 是前提 $\Gamma$ 下 $p$ 的一个愿望后承,记为 $\Gamma \vdash D(p) \rightarrow D(q)$ ,如果 $(q)_{\Gamma}$ 非空且对任何 $\Pi \in (q)_{\Gamma}$ ,存在 $\Pi' \in (p)_{\Gamma}$ ,使得:

- (1)  $\Pi \cap \Pi' \neq \emptyset$ ;
- (2)  $\Pi \cap \neg \Pi'$ 为空.

根据上述定义,“ $q$ 是 $p$ 的愿望后承”当且仅当下列4项条件同时得到满足:(i) $q$ 的极小 $\Gamma$ -模型集 $(q)_{\Gamma}$ 非空;这就保证了 $q$ 不是矛盾式.一般认为,尽管主体的不同愿望之间可以有矛盾,但一个自身矛盾的愿望对BDI规划是没有意义的.(ii) $p$ 的极小 $\Gamma$ -模型集 $(p)_{\Gamma}$ 非空,这是由“存在 $\Pi' \in (p)_{\Gamma}$ 使(1)、(2)成立”表达的;这就保证了 $p$ 也不是矛盾式.(iii) $\Pi \cap \Pi' \neq \emptyset$ ;其直观含义是, $q$ 的每一个 $\Gamma$ -模型 $\Pi$ 都至少包含 $p$ 的一个 $\Gamma$ -模型 $\Pi'$ 的“必要属性”,即 $\Pi'$ 对 $\Pi$ 是“有帮助”的.这是对“有帮助”的形式表述.(iv) $\Pi'$ 与 $\Pi$ 相容,即它们对任何原子的赋值都不相反(一个为真,另一个为假);这是由条款(2)表达的.易证此条件保证了 $\Gamma \cup \{q\} \vdash_{CL} \neg p$ 不成立,即 $q$ 真不可能导致 $p$ 假.这是对“可接受”的形式表述.

定义4使用极小 $\Gamma$ -模型的必要性来自下述事实:假如在定义中直接使用 $\Gamma$ -模型,则 $\Pi$ 和 $\Pi'$ 中可能包含的“不必要属性”将对“有帮助”和“可接受”的严格定义产生干扰.例如, $x_1$ 有一个 $\Gamma$ -模型 $\{x_1, x_2, \dots\}$ ,它同时也是 $x_2$ 一个 $\Gamma$ -模型;于是由上述条件(iii)(去掉“极小”之后), $x_2$ 将对 $x_1$ “有帮助”,这显然是不合理的,而且完全是虚假的.引入极小 $\Gamma$ -模型来定义愿望后承,就排除了虚假的“有帮助”和“可接受”,保证了直觉定义得到恰当的形式刻画.

上述愿望后承就是本文所定义的隐式愿望.为了与其他可能的隐式愿望定义相区别,称其为 pm-隐式愿望或 pm-愿望后承.下面研究它的基本性质.

**定理 1.** 如果 $\Gamma \vdash_{CL} p \rightarrow q$ ,那么 $[p]_{\Gamma} \subseteq [q]_{\Gamma}$ .

证明:设 $\Pi$ 是 $p$ 的 $\Gamma$ -模型,则 $\Gamma \cup \Pi \vdash_{CL} p$ ;又因为 $\Gamma \vdash_{CL} p \rightarrow q$ ,所以 $\Gamma \cup \Pi \vdash_{CL} q$ ;故 $\Pi$ 也是 $q$ 的 $\Gamma$ -模型.  $\square$

**定理 2.** 若 $\Gamma \vdash_{CL} (p \rightarrow q) \wedge (q \rightarrow p)$ ,则 $[p]_{\Gamma} = [q]_{\Gamma}$ ,并且 $(p)_{\Gamma} = (q)_{\Gamma}$ .

证明:由定理1,  $[p]_{\Gamma} = [q]_{\Gamma}$ 显然成立.对 $(p)_{\Gamma} = (q)_{\Gamma}$ ,只需证明 $(p)_{\Gamma} \subseteq (q)_{\Gamma}$ 即可.假设 $p$ 的某个极小 $\Gamma$ -模型 $\Pi$ 不是 $q$ 的极小 $\Gamma$ -模型,则存在小于 $\Pi$ 的 $\Pi'$ 是 $q$ 的 $\Gamma$ -模型.于是由 $[p]_{\Gamma} = [q]_{\Gamma}$ , $\Pi'$ 和 $\Pi$ 都是 $p$ 的 $\Gamma$ -模型,所以 $\Pi$ 不是 $p$ 的极小 $\Gamma$ -模型,矛盾.故必有 $(p)_{\Gamma} \subseteq (q)_{\Gamma}$ .  $\square$

**定理 3(等价).** 如果 $\Gamma \vdash_{CL} (p \rightarrow q) \wedge (q \rightarrow p)$ ,则 $\Gamma \vdash D(r) \rightarrow D(p)$ 蕴涵 $\Gamma \vdash D(r) \rightarrow D(q)$ , $\Gamma \vdash D(p) \rightarrow D(r)$ 蕴涵 $\Gamma \vdash D(q) \rightarrow D(r)$ .

**定义 5(非平凡).**  $p$ 在前提 $\Gamma$ 下非平凡,如果 $\Gamma \vdash_{CL} p$ 和 $\Gamma \vdash_{CL} \neg p$ 都不成立;否则 $p$ 在 $\Gamma$ 下平凡.

**引理 1.** 所有矛盾式和重言式在任何 $\Gamma$ 下都是平凡的.

**引理 2.**  $p$ 在 $\Gamma$ 下平凡当且仅当 $(p)_{\Gamma}$ 是空集或者 $\Pi_0 \in [p]_{\Gamma}$ .

证明:设 $p$ 在 $\Gamma$ 下平凡.若 $\Gamma \vdash_{CL} p$ ,则易知 $\Pi_0 \in [p]_{\Gamma}$ ;若 $\Gamma \vdash_{CL} \neg p$ ,则易证 $[p]_{\Gamma}$ 为空集,因而 $(p)_{\Gamma}$ 也是空集.因此, $p$ 平

凡蕴含 $(p)_\Gamma$ 是空集或者 $\Pi_0 \in [p]_\Gamma$ .反之,设 $p$ 在 $\Gamma$ 下非平凡.于是 $p$ 不是重言式或矛盾式.令 $p^*$ 为 $p$ 的一个析取范式,则 $p^*$ 的每一个析取支 $p_i$ 对应于 $p$ 的一个赋值 $\Pi_i$ ,且 $\Pi_i \vdash_{cl} p$ .若所有 $\Pi_i$ 都与 $\Gamma$ 不相容,则对所有 $i$ 有 $\Gamma \vdash_{cl} \neg p_i$ ,于是必有 $\Gamma \vdash_{cl} \neg p$ ,与假设矛盾;故至少有一个 $\Pi_i$ 与 $\Gamma$ 相容,因而 $\Pi_i \in [p]_\Gamma$ ,即 $[p]_\Gamma$ 不是空集,所以 $(p)_\Gamma$ 也不是空集.此时必有 $\Pi_0 \notin [p]_\Gamma$ ;如若不然,则 $\Pi_0$ 也是 $p$ 的 $\Gamma$ -模型,由 $\Gamma$ -模型的定义可知 $\Gamma \cup \Pi_0 \vdash_{cl} p$ ,与 $p$ 在 $\Gamma$ 下非平凡矛盾.所以 $\Pi_0 \notin [p]_\Gamma$ 成立.综上,本引理成立.  $\square$

$(p)_\Gamma$ 是空集表示 $p$ 在前提 $\Gamma$ 下是不可满足的,即在 $\Gamma$ 描述的环境中不可能为真(不可能实现).由 $\Pi_0 \in [p]_\Gamma$ 立即推出 $(p)_\Gamma = \{\Pi_0\}$ .根据上面对待小 $\Gamma$ -模型意义的讨论, $(p)_\Gamma = \{\Pi_0\}$ 意味着 $p$ 在前提 $\Gamma$ 下无须任何“必要属性”的“帮助”而“自动”为真,即 $p$ 在前提 $\Gamma$ 描述的环境中已经实现.综合两方面的情况可知,若 $p$ 在 $\Gamma$ 下平凡,则 $p$ 在 $\Gamma$ 下不可能实现或已经实现,因此不应将 $p$ 作为隐式愿望,也就是说,任何时候都不应将 $p$ 作为愿望后承推出.但是,不能禁止这样的 $p$ 作为显式愿望(显式愿望通常是事先给出的),只应禁止从这样的显式愿望推出任何隐式愿望.

**定理 4(反推).** 如果 $\Gamma \vdash_{cl} p \rightarrow q$ ,且 $p, q$ 在 $\Gamma$ 下非平凡,则 $\Gamma \vdash D(q) \rightarrow D(p)$ .

证明:任取 $p$ 的极小 $\Gamma$ -模型 $\Pi$ .由定理 1,  $\Pi$ 也是 $q$ 的 $\Gamma$ -模型.如果 $\Pi$ 是 $q$ 的极小 $\Gamma$ -模型,由非平凡性,易知愿望后承的 4 项条件均得到满足,定理成立;否则,存在文字 $x \in \Pi$ ,删去 $x$ 后 $\Pi$ 仍为 $q$ 的 $\Gamma$ -模型.重复此过程有限次,最终将得到 $q$ 的一个极小 $\Gamma$ -模型 $\Pi_1$ ,它是 $\Pi$ 的真子集而且由 $q$ 非平凡性非空.易知此时愿望后承的 4 项条件仍得到满足,定理成立.  $\square$

**定理 5(非平凡性).** 若 $p$ 或 $q$ 在 $\Gamma$ 下平凡,则 $\Gamma \vdash D(p) \rightarrow D(q)$ 不成立.

**定理 6(非矛盾性).** 若 $p$ 或 $q$ 为矛盾式,则 $\Gamma \vdash D(p) \rightarrow D(q)$ 不成立.

定理 5、定理 6 表明,任何非偶然命题(重言式或矛盾式)都不是隐式愿望,也不可能派生出隐式愿望.这一性质恰好合乎 BDI 规划的需要.

**定理 7(有穷可判定性).** 对任何有穷前提 $\Gamma$ , pm-愿望后承是可判定的.

这表明, pm-愿望后承不仅给出了隐式愿望的定义,而且同时提供了隐式愿望的推理工具.

**性质 1(非单调性).** 存在 $p$ 的 $\Gamma$ -模型 $\Pi$ ,使 $\Pi$ 不是 $p$ 的 $\Gamma'$ -模型,其中 $\Gamma' = \Gamma \cup \{q\}$ .此结论对极小 $\Gamma$ -模型也成立.

**定理 8(前提非单调性).** 设 $\Gamma \vdash_{cl} r$ 不成立,则 $\Gamma \vdash D(p) \rightarrow D(q)$ 不蕴涵 $\Gamma \cup \{r\} \vdash D(p) \rightarrow D(q)$ .

**定理 9(前件非单调性).** 设 $\Gamma \vdash_{cl} p \rightarrow q$ ,则 $\Gamma \vdash D(q) \rightarrow D(r)$ 不蕴涵 $\Gamma \vdash D(p) \rightarrow D(r)$ .

**性质 2(非传递性).**  $\Gamma \vdash D(p) \rightarrow D(q)$ 且 $\Gamma \vdash D(q) \rightarrow D(r)$ 不蕴涵 $\Gamma \vdash D(p) \rightarrow D(r)$ .

第 3 节将讨论传递性不成立的原因和影响.下面列举一些基本愿望后承关系,以便更直观地表明 pm-愿望后承的特点.为简洁起见,当 $\Gamma$ 为空时予以省略.

**性质 3(基本后承式).** 对任何 $x, y, z \in Atom$ 有

$$(p1-1) \quad \vdash D(x \wedge y) \rightarrow D(x)$$

$$(p1-2) \quad \vdash D(x \vee y) \rightarrow D(x)$$

$$(p1-3) \quad \vdash D(x) \rightarrow D(x \wedge y)$$

$$(p1-4) \quad \vdash D(x) \rightarrow D(x \vee y) \text{不成立}$$

$$(p1-5) \quad \vdash D(x) \rightarrow D(x \wedge \neg x) \text{不成立}$$

$$(p1-6) \quad \vdash D(x \wedge (x \vee y)) \rightarrow D(x)$$

$$(p1-7) \quad \vdash D(x \wedge (x \vee y)) \rightarrow D(y) \text{不成立}$$

$$(p1-8) \quad \vdash D(x \wedge (\neg x \vee y)) \rightarrow D(y)$$

$$(p1-9) \quad \vdash D(x \wedge (\neg x \vee y)) \rightarrow D(\neg x) \text{不成立}$$

$$(p1-10) \quad \vdash D((x \wedge y) \vee (\neg x \wedge \neg y)) \rightarrow D(x)$$

$$(p1-11) \quad \vdash D((x \vee y) \wedge (\neg x \vee \neg y)) \rightarrow D(x)$$

$$(p1-12) \quad \{\neg y\} \vdash D(x) \rightarrow D(x \wedge y) \text{不成立}$$

$$(p1-13) \quad \{x\} \vdash D(x \wedge y) \rightarrow D(x) \text{不成立}$$

$$(p1-14) \quad \{y \rightarrow x\} \vdash D(x) \rightarrow D(y)$$

$$(p1-15) \quad \{y \rightarrow x, x \rightarrow z\} \vdash D(z) \rightarrow D(y)$$

由(p1-1)~(p1-4)可知,pm-愿望后承与经典逻辑后承是不同的,但也有相似之处.经典逻辑后承适合于描述与 Belief 相关的推理,但却不适合于描述与愿望相关的推理.例如,(p1-2)说明,在经典逻辑中不成立的蕴涵关系,在愿望后承推理中应该是成立的.就此例而言,后件对前件是“有帮助”且“可接受”的.再如,(p1-4)表明,pm-愿望后承要求后件一定对前件有帮助:由于  $x \vee y$  不一定对  $x$  有帮助(可能只是  $y$  成立而  $x$  不成立),因而不是愿望后承.(p1-5)表明矛盾式不能成为隐式愿望,这是“非矛盾性”的一个例子.(p1-6)和(p1-7)表明逻辑等价的公式,比如  $x \wedge (x \vee y)$  与  $x$ ,在 pm-愿望后承中也是等价的.(p1-8)和(p1-9)反映了 pm-愿望后承与经典逻辑的一个相似之处,对应的后承关系在经典逻辑中也成立.(p1-10)和(p1-11)表明,一个状态可以同时是一个显式愿望和其否定的隐式愿望;这是符合直觉的:在某些(但并非所有)情况下,同一个行动或状态可以对冲突的愿望都是有用无害的.(p1-1)与(p1-13)的对比、(p1-3)与(p1-12)的对比都表明,pm-愿望后承是非单调的:在前提增加了条件之后,原来成立的后承关系可能不再成立了.(p1-15)反映了 pm-愿望后承相对于前提的传递性.

### 3 讨论

回到第 1 节给出的例子“啤酒和酸奶问题”,用逻辑语言来描述,主体的信念集合为空,显式愿望是  $\{x_1 \wedge x_2\}$ ,其中  $x_1$  表示“得到啤酒”, $x_2$  表示“得到酸奶”.通过行动  $a$  达到的状态(即  $a$  的“效果”)是  $x_1 \wedge y$ ,其中  $y$  表示“得到开啤酒的工具”.在 pm-隐式愿望中有  $\vdash D(x_1 \wedge x_2) \rightarrow D(x_1 \wedge y)$ ,所以,此时  $x_1 \wedge y$  是主体的显式愿望.行动  $a$  也是可选行动.

下面我们进一步研究 pm-愿望后承的其他性质,着重与相关工作进行比较.

**推论 1.** 若  $\vdash_{cl} (p \rightarrow q) \wedge (q \rightarrow p)$ , 则  $\Gamma \vdash D(p) \rightarrow D(q)$ .

这表明,如果根据传统的等价性定义, $q$  是相对于  $p$  的隐式愿望,则根据 pm-愿望后承, $q$  也是相对于  $p$  的隐式愿望.这就是说,pm-愿望后承包含了传统的等价性定义为其特例.事实上,等价性定义并不错,只是不充分.另外,由推论 1 和定理 3 可知,逻辑等价的命题在 pm-愿望后承系统中是永远等价的.

**推论 2.** 若  $\vdash_{cl} p \rightarrow q$ , 且  $p, q$  在  $\Gamma$  下非平凡, 则  $\Gamma \vdash D(q) \rightarrow D(p)$ .

因此,在一般情况下(即  $p, q$  非平凡时),根据充分性定义的隐式愿望也是 pm-隐式愿望,即充分性定义也是 pm-隐式愿望的特例.当  $p$  或  $q$  平凡时,即使依充分性定义  $q$  是相对于  $p$  的隐式愿望,却肯定不是 pm-隐式愿望,此时两个定义不一致.根据上一节的分析,任何平凡的命题在  $\Gamma$  下不可能实现或已经实现,因而不应作为隐式愿望或用来推出隐式愿望.对 BDI 规划而言,以平凡式(特别是矛盾式或重言式)为愿望是没有意义的,因为它们对行动选择不起正面的作用.所以,pm-隐式愿望将平凡式排除在外是更加合理的.总之,充分性定义的主要问题在于它要求太强,排斥了一些有意义的、重要的情况,如“啤酒和酸奶问题”所表明的那样;次要问题是包含了少量不恰当的情况.与此密切相关的一个性质是:pm-隐式愿望的后承关系虽然不是传递的,但保持前提中的传递性,如(p1-15)所反映的那样.

**推论 3.**  $\vdash_{cl} p \rightarrow q$  不蕴含  $\Gamma \vdash D(p) \rightarrow D(q)$ .

因此,必要性定义下的隐式愿望不一定是 pm-隐式愿望;(p1-4)是一个例子.事实上,必要性定义是不恰当的,它的条件太弱,既容忍不一定“有帮助”的状态作为隐式愿望,又容忍不一定“可接受”的状态作为隐式愿望.(p1-4)反映了前一种情况:如果采用必要性定义, $x \vee y$  将成为相对于  $x$  的隐式愿望.根据上面的分析, $x \vee y$  对  $x$  不一定“有帮助”.(p1-4)表明,pm-愿望后承拒绝这种情况.必要性定义容忍不一定“可接受”的状态作为隐式愿望的一个例子是: $x \vee \neg y$  在必要性定义之下将成为相对于  $x \wedge y$  的隐式愿望,尽管  $x \vee \neg y$  有可能与  $x \wedge y$  相矛盾.易知 pm-愿望后承也拒绝这种情况的出现.另外,值得注意的是,pm-愿望后承在排斥必要性定义不合理部分的同时,包含了必要性定义的合理部分,(p1-1)和(p1-8)都是这样的例子.

总之,pm-隐式愿望完整地包含了传统的等价性定义为其特例,同时包含了充分性定义和必要性定义的合理部分.因此,pm-隐式愿望定义对隐式愿望的要求是比较宽泛的.除具有直觉合理性之外,这种宽泛性还有一个更加重要的优点:向智能主体提供了更大的自主性,也就是说,智能主体在行动选择中拥有更大的“自主权”,可以在更大的范围内自主地选择自己的行动.显然,这将使得智能主体在复杂、不可预测的环境中能够独立地完成更多、更需要“智能”的任务.考虑到自主性是当前人工智能研究的核心目标,pm-隐式愿望的上述特点具有重要意义.

此外,pm-隐式愿望还有一个很大的优点:它是非单调的,而传统定义是单调的.显然,愿望(无论隐式还是显式)的表示和推理属于常识知识表示和推理的范围,因而应体现出非单调性.这使得 pm-隐式愿望更加接近实际应用的需要.

根据性质 2,在 pm-隐式愿望中传递性不成立.然而在适当的用法之下,这并不带来任何实际的损害.为此,要求在 BDI 规划中,pm-隐式愿望只用于“一次性”的行动选择.我们将在 pm-隐式愿望的基础上加以推进,得到适用于“连续性”行动选择的隐式愿望推理系统.

隐式愿望所刻画的是命题之间的“有用”关系,定性决策论<sup>[11,12]</sup>所刻画的是命题之间的“优先”关系.一个主体的行动是理性的满足两个条件,首先它是对主体的目标“有用”的;其次,它是众多对主体有用的行动中“优先”的.文献[10]中关于“有用”和“优先”关系有较详细的讨论.本文所刻画的是命题对命题的“有用”关系,下一步的工作将扩展到行动的表示和行动对命题的“有用”关系上.

## 4 结 论

隐式愿望的刻画在 Agent 理论研究中有着重要的作用.本文首先分析隐式愿望现有 3 种定义的不充分性或不当性,进而提出了对隐式愿望基本内涵的一种新的概括——“有用无害”.在此基础上,用推广的“归约蕴含”方法严格定义一种新的隐式愿望——pm-愿望后承,研究了它的主要性质以及与 3 种传统定义的关系,进而论证 pm-愿望后承的合理性,包括符合直觉、有利于提高智能主体的自主性、更接近 BDI 规划和实际应用的需要.

“隐式愿望问题”是愿望刻画的核心问题之一,和许多问题关系密切,其中之一是愿望的冲突问题.从隐式愿望的角度来看,愿望冲突具体体现在多项愿望之间发生直接或间接的矛盾,一个典型的案例是:设存在主体的两个显式愿望  $d_1, d_2$ , 使得  $\Gamma \vdash D(d_1) \rightarrow D(p)$  并且  $\Gamma \vdash D(d_2) \rightarrow D(\neg p)$ , 那么  $p, \neg p$  是不是主体的隐式愿望?对此,本文采取“相对性”的观点,我们称一个命题是主体的隐式愿望总是相对于主体的某一个给定的显式愿望来说的.在这个情形下,我们认为  $p$  是主体相对于  $d_1$  的隐式愿望,而  $\neg p$  是主体相对于  $d_2$  的隐式愿望.当然,这个答案并不能完全令人满意,我们将在未来的工作中进一步加以研究.

在本文基础上,还有其他很多工作需要完成,其中包括 pm-愿望后承的快速判定算法和 pm-愿望后承的一阶推广等等.另外,pm-愿望后承还引入一些“冗余”的东西,一个显式愿望的愿望后承是无限多个的,如何使得 pm-愿望后承的推理更加具有可控制性,也是下一步工作的重点.

**致谢** 朱紫汝、诸世卓及中国科学技术大学计算机科学与技术系 AI 逻辑小组的其他成员对本文提出了很多宝贵的建议,在此一并表示感谢.

## References:

- [1] Rao AS, Georgeff MP. Modeling rational agents within a BDI-architecture. In: Allen J, *et al*, eds. Proc. of the 2nd Int'l Conf. (KR'91). San Mateo: Morgan Kaufmann Publishers, Inc., 1991. 473–484.
- [2] Herzig A, Longin D. C & L intention revised. In: Dubois D, *et al*. eds. Proc. of the 9th Int'l Conf (KR 2004). Menlo Park: AAAI Press, 2004. 527–535.
- [3] Chen XP, Liu GQ. A logic of intention. In: Dean T, ed. Proc. of the IJCAI'99. San Mateo: Morgan Kaufmann Publishers, Inc., 1999. 172–177.
- [4] Zhu ZR, Chen XP, Zhou Y. A new formal system of intention consequences  $L_{msc}$ . Journal of Software, 2002,13(7):1271–1277 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/13/1271.pdf>
- [5] Mao XJ, Wang HM, Chen HW, Liu FQ. The intention theory of agent computing in multi-agent system. Journal of software, 1999, 10(1):43–48 (in Chinese with English abstract).
- [6] Hu SL, Shi CY. An intention model for agent. Journal of Software, 2000,11(10):965–970 (in Chinese with English abstract).
- [7] Wang WJ, Tian QJ, Shi ZZ. Reasoning about others in multi-agent system. Journal of Computer Research and Development, 1998, 35(11):971–974 (in Chinese with English abstract).

- [8] Georgeff M, Pell B, Pollack M, Tambe M, Wooldridge M. The belief desire intention model of agency. In: Muller J, *et al*, eds. Proc. of the Agents, Theories, Architectures and Languages (ATAL'98). Heidelberg: Springer-Verlag, 1998. 1–10.
- [9] Weld. DS. Recent advances in AI planning. AI Magazine, 1999,20(2):93–123.
- [10] Zhou Y, Chen XP. Partial implications semantics for desirable propositions. In: Dubois D, *et al*. eds. Proc. of the 9th Int'l Conf. (KR 2004). Menlo Park: AAAI Press, 2004. 606–611.
- [11] Doyle J, Thomason R. Background to qualitative decision theory. AI-Magazine, 1999,20(2):55–68.
- [12] Lang J, van der Torre L, Weydert E. Hidden uncertainty in the logical representation of desires. In: Gottlob T, Walsh T, eds. Proc. of the IJCAI 2003. Acapulco: Morgan Kaufmann Publishers, 2003. 685–690.

#### 附中文参考文献:

- [4] 朱紫汝,陈小平,周熠.一个新的意图后承形式系统  $L_{msc}$ . 软件学报,2002,13(7):1271–1277. <http://www.jos.org.cn/1000-9825/13/1271.pdf>
- [5] 毛新军,王怀民,陈火旺,刘凤歧.Agent 在多 Agent 系统中计算的意愿理论.软件学报,1999,10(1):43–48.
- [6] 胡山立,石纯一.Agent 的意图模型.软件学报,2000,11(10):965–970.
- [7] 王文杰,田启家,史忠植.多主体系统中对其他主体的研究.计算机研究与发展,1998,35(11):971–974.

## 第 10 次全国 Petri 网学术年会暨形式化方法学术讨论会

### 征文通知

由中国计算机学会 Petri 网专业委员会主办的第 10 次全国 Petri 网学术年会暨形式化方法学术讨论会将于 2005 年 10 月在镇江召开 (江苏大学承办), 会议将对 Petri 网理论及应用, 以及并行处理的形式化方法开展广泛、深入的讨论。现发出征文通知。

#### 一、征文范围

Petri 网理论研究; Petri 网在各领域的应用研究; Petri 网工具开发与 Petri 网运行仿真技术; 相关的并发模型及仿真技术研究; 并行计算与并行处理的形式化方法; 其它相关问题的研究。

#### 二、征文要求

1、凡在正式刊物或其他学术会议上发表 (或录用) 过的论文不再征用; 2、投送论文无论录用与否概不退稿, 请作者自留底稿; 3、会议论文集以《系统仿真学报》增刊出版 (部分优秀的论文可推荐到正刊发表), 寄送的论文要求按《系统仿真学报》要求的论文格式 (可在网上查阅) 编辑打印; 4、被录用的论文, 按要求修改后寄出电子版 (软盘), 并按规定寄出版面费。5、投送论文请写清联系人姓名、地址、(包括邮政编码), 以便联系。

#### 三、重要日期 (邮寄以邮戳日期为准)

征文截稿日期: 2005 年 5 月 31 日

发出录用或 (和) 修改通知日期: 2005 年 7 月 10 日

编辑打印好的稿件返回日期: 2005 年 7 月 31 日

#### 四、联系方式

论文请寄: 北京航空航天大学软件学院 (邮编 100083) 联系人: 武晓乐

电话: 010-82316479 86354705 (小灵通) E-mail: wuxiaole@buaa.edu.cn