Agent 逻辑和真假子集语义*

胡山立[1,2], 石纯一[3]

[1](中国科学院 软件研究所 计算机科学重点实验室,北京 100080);
[2](福州大学 计算机科学与技术系,福建 福州 350002);
[3](清华大学 计算机科学与技术系,北京 100084)
E-mail: husl@fzu.edu.cn

摘要: 理性 Agent 规约的形式框架通常基于信念、愿望和意图逻辑.为了克服现有的信念、愿望和意图逻辑中存在的问题,为非正规模态算子提供一种合适的语义表示.讨论了理性 Agent 性态的抽象规约中对语义表示的要求以及现有的信念、愿望和意图逻辑中存在的问题.介绍了作者开发的真假子集语义及其在 Agent 形式化中的应用.他们的框架使意图的有问题的性质无效.并且证明通过对模型的代数结构施加一定的约束,能获得许多希望的性质.最后对真假子集语义进行了分析.这一切表明真假子集语义为非正规模态算子提供了一种合适的语义表示,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 性态的逻辑规约的有力工具,可应用于建立新的合适的 Agent 逻辑系统.

关 键 词: Agent;意图;语义;真假子集语义;模型
中图法分类号: TP18 文献标识码: A

信念、愿望和意图(BDI)逻辑在有关理性 Agent 设计的 AI 文献中已得到充分的重视[1~5],其重要性在于在 Agent 性态的抽象规约中作为逻辑框架.这样的逻辑需要一个建立在涉及知识和信念的认识逻辑之上的形式基础,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[2~5],从而不可避免地存在这种方法所固有的一些问题.以意图为例主要存在以下几个问题,其中 INT 是意图算子:(1) 逻辑全知问题:$\models\phi\Rightarrow\models INT(\phi)$;(2) 重言隐含下的副作用问题:$\models\phi\rightarrow\gamma\Rightarrow\models INT(\phi)\rightarrow INT(\gamma)$;(3) 析取扩大化问题:$\models INT(\phi)\rightarrow INT(\phi\vee\gamma)$;(4) 合取分离问题:$\models INT(\phi\wedge\gamma)\rightarrow INT(\phi)\wedge INT(\gamma)$.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 的形式化工具[3,6].

Konolige 和 Pollack[3]认为正规模态逻辑不适用于意图.他们用方案(scenarios)来表示 Agent 的思维状态,公式ϕ的方案是 W 的使ϕ为真的子集.他们的模型中的方案集 I 实际上可看成是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 $INT(\gamma)$在该模型中可由 $INT(\phi)$推出,那么$\gamma\equiv\varnothing$,这样,意图推论从该模型中消失了.也就是说,该模型失去了非等价意图的推理能力,从而是不可取的.

因此,又有许多学者把同样是经典的但非正规的模态逻辑应用到 Agent 形式化中,并做了许多研究[6].但作为经典的非正规模态逻辑基础的非正规可能世界却是如此不可思议,在一个非正规可能世界中命题ϕ和它的否定$\neg\phi$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连接词的含义而复合得到,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.

为此,一种直观上可信,能适当地表征信念、愿望和意图的直观语义的形式语义的产生就是一种必然了.下

* 收稿日期: 2001-04-03; 修改日期: 2001-06-25
基金项目: 国家自然科学基金资助项目(69973023);福建省自然科学基金资助项目(F00012;F00013)
作者简介: 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,河北山海关人,教授,主要研究领域为人工智能应用基础.

面介绍我们开发的真假子集语义[7]及其在 Agent 形式化中的应用.

1 信念、愿望和意图逻辑应有的要求

信念、愿望和意图逻辑应有的性质[1~4,6,8]大体上可分为两类:一类是与两个模态有关,反映它们之间的关系的(A1~A5);另一类是只与一个模态有关,反映其特性的,这里只讨论意图(A6~A12).BEL,DEL 和 INT 分别是信念、愿望和意图算子.

(A1) $\forall M$ 有 $M\models INT(\phi)\rightarrow\neg BEL(\neg\phi)$.

(A2) $\forall M$ 有 $M\models INT(\phi)\rightarrow\neg BEL(\phi)$.

(A3) $\exists M$ 使 $M\models BEL(\phi)\wedge\neg INT(\phi)$.

(A4) $\exists M$ 使 $M\models BEL(\phi\rightarrow\gamma)\wedge INT(\phi)\wedge\neg INT(\gamma)$.

(A5) $\forall M$ 有 $M\models INT(\phi)\rightarrow DES(\phi)$.

(A6) $\forall M$ 有 $M\models\neg INT(\phi\wedge\neg\phi)$.

(A7) $\forall M$ 有 $M\models INT(\phi)\rightarrow\neg INT(\neg\phi)$.

(A8) $\forall M$ 有 $M\models INT(\phi)\wedge INT(\gamma)\rightarrow INT(\phi\wedge\gamma)$.

(A9) $\exists M$ 使 $M\models INT(\phi\wedge\gamma)\wedge\neg INT(\phi)$.

(A10) $\exists M$ 使 $M\models INT(\phi\vee\gamma)\wedge\neg INT(\phi)\wedge\neg INT(\gamma)$.

(A11) $\exists M$ 使 $M\models INT(\phi)\wedge\neg INT(\phi\vee\gamma)$.

(A12) $\exists M$ 使 $M\models(\phi\rightarrow\gamma)\wedge INT(\phi)\wedge\neg INT(\gamma)$.

2 真假子集语义和 BDI 模型

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、愿望和意图引入相应的模态算子 BEL,DES 和 INT,把它扩充到可能世界框架.但模态算子 DES 和 INT 不是正规的,而且采用我们开发的真假子集语义[7].另外,这里不限定可能世界的结构.语义如下:

定义 **1**. 模型 M 是一个元组,$M=\langle W,T,\mathcal{B},\{\mathcal{DT},\mathcal{Df}\},\{\mathcal{IT},\mathcal{If}\},V\rangle$,其中 W 是可能世界集,T 是时间点的有序集,一个情景(situation)是处于某个时间点 t 的世界 w,记为 w_t;$\mathcal{B},\mathcal{DT},\mathcal{Df},\mathcal{IT},\mathcal{If}$ 分别将当前情景映射到信念、愿望(真)、愿望(假)、意图(真)、意图(假)可达世界集 $\mathcal{B}_t^w,\mathcal{DT}_t^w,\mathcal{Df}_t^w,\mathcal{IT}_t^w,\mathcal{If}_t^w$,它们都是 W 的子集.形式上 $\mathcal{B},\mathcal{DT},\mathcal{Df},\mathcal{IT},\mathcal{If}\subseteq W\times T\times W$.$V$ 是一个赋值,它规定世界 w 在时间点 t,各个公式的真值(BEL,DES 和 INT 算子的赋值规定如下所述).

对公式 ϕ,当且仅当 $V(w_t,\phi)=1$(真)时,称 ϕ 在 w_t 为真或有效,记为 $M,w_t\models\phi$.

在时间点 t,使公式 ϕ 为真的可能世界集记为 $WT(t,\phi)$,使公式 ϕ 为假的可能世界集记为 $WF(t,\phi)$.

定义 **2**. 模态算子 BEL,DES 和 INT 的形式语义定义如下,其中 M 是一个模型,ϕ 是一个公式.

(1) $M,w_t\models BEL(\phi)$ iff $\forall w'\in\mathcal{B}_t^w\quad M,w_t'\models\phi$.即 $M,w_t\models BEL(\phi)$ iff $\mathcal{B}_t^w\subseteq WT(t,\phi)$.

(2) $M,w_t\models DES(\phi)$ iff $\forall w'\in\mathcal{DT}_t^w\quad M,w_t'\models\phi$且$\forall w''\in\mathcal{Df}_t^w\quad M,w_t''\models\neg\phi$.即 $M,w_t\models DES(\phi)$ iff $\mathcal{DT}_t^w\subseteq WT(t,\phi)$且 $\mathcal{Df}_t^w\subseteq WF(t,\phi)$.

(3) $M,w_t\models INT(\phi)$ iff $\forall w'\in\mathcal{IT}_t^w\quad M,w_t'\models\phi$且$\forall w''\in\mathcal{If}_t^w\quad M,w_t''\models\neg\phi$;即 $M,w_t\models INT(\phi)$ iff $\mathcal{IT}_t^w\subseteq WT(t,\phi)$且 $\mathcal{If}_t^w\subseteq WF(t,\phi)$.

其中,DES 和 INT 的形式语义分别用真假子集 $\mathcal{DT}_t^w,\mathcal{Df}_t^w$ 和 $\mathcal{IT}_t^w,\mathcal{If}_t^w$ 来刻画,我们称其为真假子集语义.

由于用真假子集 $\mathcal{IT}_t^w,\mathcal{If}_t^w$ 来刻画意图,从而当 $\mathcal{If}_t^w\neq\varnothing$ 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 A6~A12 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,有下面的定理 1.

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 A6~A12 成立.

为了反映意图和信念,以及意图和愿望之间的关系(A1~A5),应对模型的代数结构施加一定的约束.不难证

明相应的语义约束为

(CI1) $\forall w\in W,\forall t\in T$ 有 $I\mathcal{T}_t^w\cap\mathcal{B}_t^w\neq\varnothing$;

(CI2) $\forall w\in W,\forall t\in T$ 有 $If_t^w\cap\mathcal{B}_t^w\neq\varnothing$;

(CI3) $\forall w\in W,\forall t\in T$ 有 $I\mathcal{T}_t^w\supseteq\mathcal{DT}_t^w$ 且 $If_t^w\supseteq\mathcal{D}f_t^w$.

定理 **2**. 在满足约束 CI1,CI2 和 CI3 的模型 M 中,性质 A1~A5 成立.

3 意图维护的语义约束

下面考察满足约束 CI1,CI2 和 CI3 的模型,当环境变化从而引起 Agent 的信念或愿望发生变化时,意图的维护问题,即先前的意图中哪些是需要保留的.其目的是对上述模型给出动态的语义约束,使得从上一时间点到下一时间点模型的运转能反映意图的合理维护.

我们用 $VS_1(I\mathcal{T}_t^w)$表示在$I\mathcal{T}_t^w$ 中的每个世界 w'均为真的公式集,即定义 $VS_1(I\mathcal{T}_t^w)=\{\phi|\forall w'\in I\mathcal{T}_t^w,V(w_t',\phi)=1\}$,其中$\phi$是公式,$V$ 是赋值.$VS_1(\mathcal{B}_t^w)$和 $VS_1(\mathcal{DT}_t^w)$类似.用 $VS_0(If_t^w)$表示在If_t^w 中的每个世界 w'均为假的公式集,即定义 $VS_0(If_t^w)=\{\phi|\forall w'\in If_t^w,V(w'_t,\phi)=0\}$,$VS_0(\mathcal{D}f_t^w)$类似.用 $VDS_1(\mathcal{B}_t^w)$表示在 $\mathcal{B}_t^w$ 中至少一个世界 w'为真的公式集,即定义 $VDS_1(\mathcal{B}_t^w)=\{\phi|\exists w'\in\mathcal{B}_t^w,V(w_t',\phi)=1\}$.

直观上我们希望当信念、愿望改变时,意图和信念、愿望都能保持一致.也就是说,我们希望 Agent 仍然相信是可能实现的且仍然希望加以实现的意图应该保留下来.这称为信念愿望过滤约束,形式定义如下:

定义 **3**(**BDFC**). $VS_1(I\mathcal{T}_{t1}^w)\cap VS_0(If_{t1}^w)\cap VS_1(\mathcal{DT}_{t2}^w)\cap VS_0(\mathcal{D}f_{t2}^w)\cap VDS_1(\mathcal{B}_{t2}^w)\subseteq VS_1(I\mathcal{T}_{t2}^w)\cap VS_0(If_{t2}^w)$.

其中 $t1$,$t2$ 分别表示上下时间点.这里上一时间点 $t1$ 的意图集,用 $VS_1(I\mathcal{T}_{t1}^w)\cap VS_0(If_{t1}^w)$来表示.而下一时间点 $t2$ 相信是可能实现的公式集,由 $VDS_1(\mathcal{B}_{t2}^w)$来表示,下一时间点 $t2$ 仍然愿望的公式集由 $VS_1(I\mathcal{T}_{t2}^w)\cap VS_0(If_{t2}^w)$表示.从而约束 BDFC 反映了对上一时间点 $t1$ 意图的维护.

定理 **3**. 以下公式在满足约束 CI1~CI3,BDFC 的模型类中是有效的.

(A13) $INT(w_{t1},\phi)\wedge\neg BEL(w_{t2},\neg\phi)\wedge DES(w_{t2},\phi)\rightarrow INT(w_{t2},\phi)$.

对应于约束 BDFC,不难证明定理 3,从而实现了我们对意图维护的要求.

4 真假子集语义分析与结论

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 R_t^w 来描述意图(正像我们用子集 $\mathcal{B}_t^w$ 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(真假子集)$I\mathcal{T}_t^w$ 和If_t^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.

实际上,信念和意图的语义有不同的直观解释.信念表示 Agent 对某些可能世界 $\mathcal{B}_t^w$ 的“偏爱”.只有在这些可能世界上均为真的命题,Agent 才相信.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部分$I\mathcal{T}_t^w$ 和If_t^w,在$I\mathcal{T}_t^w$ 上为真的命题 Agent 认为是可能实现的或已实现的,而在If_t^w 上为假的命题 Agent 则认为是当前尚未实现的且不是必然会实现的.显然,只有尚未实现,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.因此,在我们的模型中对信念和意图采用不同的形式语义.

特别是,我们认为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.而对意图的描述应是对一类命题(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_t^w 来描述 $WT(t,\phi)$可看成是真子集语义.而真假子集语义真假并重,用 $R\mathcal{T}_t^w$ 来描述 $WT(t,\phi)$,并用 Rf_t^w 来描述 $WF(t,\phi)$,从而能更全面地描述二值逻辑中的模态算子.上述真假子集语义对意图算子的应用就是个很好的例子.另外,.经典的正规模态算子的可能世界语义可以看成是真假子集语义当 $Rf_t^w=\varnothing$时的退化情形.

上述讨论表明,真假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 性态的逻辑规约的有力工具.它可以应用于建立新的合适的 Agent 逻辑系统[9].

References:

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

[2] Cohen, P.R., Levesque, H.J. Intention is Choice with Commitment. Artificial Intelligence, 1990,42(2-3):213~261.

[3] 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.

[4] 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.

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

[6] Cavedon, L., Padgham, L., Rao, A., *et al*. Revisiting rationality for Agents with intentions. In: Xin, Yao, ed. Proceedings of the 8th Australian Joint Conference on Artificial Intelligence. Singapore: World Scientific Publishing Co. Pte. Ltd., 1995.131~138.

[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 (in Chinese).

[8] Hu, Shan-li, Shi, Chun-yi. An intention model for agent. Journal of Software, 2000,11(7):965~970 (in Chinese).

[9] Hu, Shan-li, Shi, Chun-yi. Agent-BDI logic. Journal of Software, 2000,11(10):1353~1360 (in Chinese).

附中文参考文献:

[7] 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] 胡山立,石纯一.Agent 的意图模型.软件学报,2000,11(7):965~970.

[9] 胡山立,石纯一.Agen-BDI 逻辑.软件学报,2000,11(10):1353~1360.

Agent Logic and the True-False Subset Semantics*

HU Shan-li[1,2], SHI Chun-yi[3]

[1](*Key Laboratory for Computer Science, Institute of Software, The Chinese Academy of Sciences, Beijing* 100080, *China*);

[2](*Department of Computer Science and Technology, Fuzhou University, Fuzhou* 350002, *China*);

[3](*Department of Computer Science and Technology, Tsinghua University, Beijing* 100084, *China*)

E-mail: husl@fzu.edu.cn

Abstract: Formal frameworks for the specification of rational agents are commonly based on logic of belief, desire and intention. In order to eliminate the problems with existing logic of belief, desire and intention, and to provide a proper semantic representation for non-normal modal operator, the problems with existing logic of belief, desire and intention are addressed, the true-false subset semantics, which is developed by the authors, and its application in the formalization of agent are introduced. The proposed framework invalidates the problematic properties of intention, and by imposing certain constraints on the algebraic structure of the models, it is showed that many desirable properties can be obtained. Finally the true-false subset semantics is analyzed. Thus the true-false subset semantics provides a proper semantic representation for non-normal modal operator. It is an important development of classical possible worlds semantics for normal modal operators, and is proved to be a powerful tool for the logical specification of rational agent behavior. It can be applied to establish a new proper agent logic system.

Key words: agent; intention; semantics; true-false subset semantics; model

* Received April 3, 2001; accepted June 25, 2001

Supported by the National Natural Science Foundation of China under Grant No.69973023; the Natural Science Foundation of Fujian Province of China under Grant Nos.F00012, F00013