

一个刻画智能 Agent 理性信念的认知系统

高金胜¹, 周昌乐^{2,3}

(1 厦门大学 人文学院哲学系, 福建 厦门 361005; 2 厦门大学 信息科学与技术学院智能科学与技术系, 福建 厦门 361005; 3 福建省仿脑智能系统重点实验室, 福建 厦门 361005)

摘 要: 在信念的类别上进行细化研究, 研究智能 Agent 在持有理性信念下的认知系统及其所具有的认知性质. 在理性信念的语义分析的基础上, 我们建立了一个有关智能 Agent 具有理性信念的推理系统, 并证明了该推理系统的可靠性及完全性. 通过该认知系统, 找到理性信念的形式化方法, 并且得出了理性信念既是智能 Agent 获取知识的必要条件. 该条件也是以往对知识定义所疏忽的限定条件, 同时还是智能 Agent 与生物认知主体在认知推理上相异的内因之一.

关键词: 理性信念; 主观概率; 认知推理; 人工智能

中图分类号: TP182

文献标识码: A

文章编号: 1000-7180(2016)12-0139-08

A Cognitive System about Intelligent Agent with Rational Belief

GAO Jin-sheng¹, ZHOU Chang-le^{2,3}

(1 Department of Philosophy, College of Humanities, Xiamen University, Xiamen 361005, China;

2 Department of Cognitive Science and Technology, College of Information Science and Technology, Xiamen University, Xiamen 361005, China; 3 Fujian Key Lab of the Brain-Like

Intelligent Systems, Xiamen 361005, China)

Abstract: This paper mainly studies on the refining category of belief, and the cognitive properties about intelligent agent with rational belief. Based on the semantic analysis of the rational belief, we establish a cognitive inference system. After the proof of its soundness and completeness, we find rational belief is a necessary condition for intelligent agent to get knowledge, and also a limiting condition for knowledge definition. This condition is also a limiting condition for the definition of knowledge which was neglected in previous. What is more, it is one reason for the different between intelligent agent and biological cognitive agent in cognitive inference.

Key words: rational belief; subjective probability; cognitive inference; artificial intelligence

1 引言

人类经常希望计算机能像人一样思考及行动. 事实上, 我们也确实在这方面取得了许多成绩. 例如, Alan Turing 的测试、Herbert Simon 的普遍问题解答器、John McCarthy 的常识程序, 以及最近的 Google DeepMind 的 AlphaGo 等^[1]. 这些设计都需要 Agent 感知外部环境, 然后通过接收的数据进行认知计算, 最后再回应外部环境. 这个认知互动过程

大致可用图 1 来描述. 从图 1 可见, 认知计算在人工智能中扮演着最为核心的角色. 对此, 认知科学及计算机科学尝试建立起一个精确的模型去仿造人类的大脑. 如图 2 所示, 人类的认知是一个复杂的过程, 也具有丰富的特征, 例如一致性、同一性、内省、混沌……^[2]. 然而, 理性是最明显的一个特征. 在“理性”与“信念”的研究上, 两者都牵涉了许多学科, 例如, 心理学、人工智能、认知科学、经济学等. 可是, 遗憾的是两者之间一直处于独立状态. 其中, 经济学、心

理学倾向于在“理性”或“非理性”与“行为”的关系中寻找规律^[3-4]. 实际上,经济学上的“理性”假设,严格上应该是理性信念的假设. 例如,在博弈中,参与者在理性信念的驱使下,彼此都为了获得最大收益,最后实现纳什均衡而获得相对稳定的博弈状态. 而人工智能目前主要的研究进程是探究“信念”、“情感”、“意图”、“行为”等之间的关系^[5-6],同样忽略了对信念的细化研究. 而在认知逻辑中,一般是以归纳的视域去处理信念的认知行为,由此倾注于对主观概率论的研究,进一步产生了动态更新理论. 在“信念”的内容变化上,Alchourrón、Gärdenfors 及 Makinson 对信念的内容扩充与收缩而提出修正函数理论(简称 AGM)^[7],而 Grove 更进一步地提出了信念修正的量化函数^[8]. 而对于“信念”的类型,认识论更多的是一种概念分析,并未从认知技术上对信念进行细化. 理性信念也并非只是“理性”与“信念”的简单结合. 本文着手于“理性信念”的研究,既源自于人工智能理论所需,也是对以往在“理性”及“信念”上的模糊处理的一种修正. 由此可见,对于人工智能而言,在“信念”与“行为”的研究上,除了采用传统上经济学、认知逻辑、行为科学等通过增加规则而限制信念,或者通过比较和量化的方式模型化信念与认知主体的行为关系^[9],还需要在认知技术上对信念进行细化,再在细化的基础上模型化两者之间的关系.

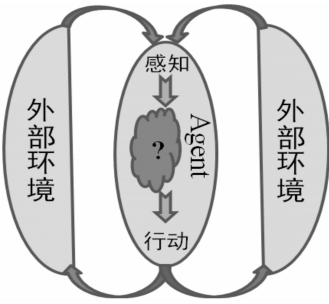


图 1 Agent 与外部环境的互动模型

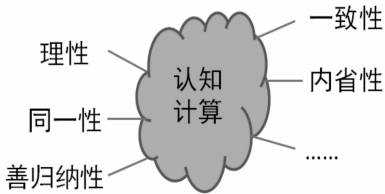


图 2 Agent 的认知计算特性

我们针对这个细化的方向而展开研究,主要在于填补人工智能中对理性信念刻画的空缺. 首先,根

据“理性信念”的语义解释,我们建立了一个具有理性信念的智能 Agent 的认知推理系统. 然后通过对该系统的可靠性及完全性的证明,验证了该推理系统是既可靠又完备的. 最后通过该系统的一些特征定理,分析具备理性信念的智能 Agent 所持有的认知特征,并探讨其在人工智能中的应用及未来的工作. 由此,本文在理性信念形式化研究的基础上,获得了智能 Agent 的理性信念的形式化方法及其认知特征,得出智能 Agent 获取知识的必要条件,以及智能 Agent 与生物认知主体在认知推理上相异的一个内因,并修正了传统知识定义的限定条件.

2 基础认知系统 S_B

定义 1(基础认知语言 L_B) 给定 P 为原子命题集, φ 为任一合式公式, B 为信念模态算子, K 为知道模态算子. 那么基础认知语言 L_B 定义如下:

$$\varphi ::= p \mid \neg \varphi \mid \varphi \wedge \varphi \mid B_i \varphi \mid K_i \varphi \quad (p \in P, i \in N)$$

(1)

对于信念,我们采用经典定义,即“Agent i 在当前状态 w 相信 φ ,当且仅当,对于任一状态 v ,Agent i 认为有可能 v 是真实状态,且 φ 在 v 为真”. 因此, $B_i \varphi$ 可视为“Agent i 相信 φ ”或“ φ 是 Agent i 的信念”. 然而,对于“知道”的语义定义,本文不再采用传统上类似信念一样的必然化定义,这是因为一方面考虑到避免在多模系统中与“信念”形成混淆,另一方面是根源于对“知识”的传统定义. 传统上,人们将“知识”的定义追溯到 Plato 的三分定义,即“知识是确证的真信念”^[10]. 事实上到目前为止,对知识的定义还是存在许多分歧,这些分歧主要是来自于知识论领域. Hintikka 认为知识是认知主体在认知空间中认为是真的命题,而 Williamson 认为若是认知主体知道了某个命题,那么意味着该命题在相应的特例中为真^[11-12]. 这些分歧的关键点在于对“确证的(justified)”的刻画,以及如何实现的问题上. 因此,Nozick 认为它应该在 Plato 的定义上增加一个限定“如果命题为假,则 Agent 不再相信它”,即 $K_i \varphi \leftrightarrow \varphi \wedge B_i \varphi \wedge (\neg \varphi \rightarrow \neg B_i \varphi)$. 然而这些分歧大都是围绕着 Plato 的定义而展开的,至少都会涉及“真”及“信念”. 所以在本文,我们将知识定义为“真信念”(至于是怎样的信念,通过下文的系统研究将可以得到明确的答案). 这样,基础认知系统 S_B 主要就是由以下几部分组成:(1) 所有命题逻辑的重言式;(2) 信念分配公理“ $B_i(\varphi \rightarrow \psi) \rightarrow (B_i \varphi \rightarrow B_i \psi)$ ”;(3) 正内省“ $B_i \varphi \rightarrow B_i B_i \varphi$ ”“ $K_i \varphi \rightarrow K_i K_i \varphi$ ”;(4) 负内

省“ $\neg B_i\varphi \rightarrow B_i\neg B_i\varphi$ ”“ $\neg K_i\varphi \rightarrow K_i\neg K_i\varphi$ ”; (5) 知识的信念定义“ $K_i\varphi \leftrightarrow \varphi \wedge B_i\varphi$ ”,以及分离规则.

3 理性信念认知系统的语法及语义

3.1 理性信念的语言

定义 2(理性信念语言 L_{RB}) 给定 P 为原子命题集, φ 为任意合式公式, B 为信念模态算子, K 为知道模态算子, B^R 为理性信念模态算子. 那么理性信念语言 L_{RB} 是由 BNF 定义如下:

$$\varphi ::= p \mid \neg \varphi \mid \varphi \wedge \varphi \mid K_i\varphi \mid B_i\varphi \mid B_i^R\varphi \quad (p \in P, i \in N) \quad (2)$$

式中, $K_i\varphi$ 表示“Agent i 知道 φ ”, $B_i\varphi$ 表示“Agent i 相信 φ ”, $B_i^R\varphi$ 表示“Agent i 对 φ 持有理性信念”或者“Agent i 理性地相信 φ ”. 另外, 其他命题符号缩写定义如常. 例如, $\varphi \vee \psi \stackrel{\text{def}}{=} \neg(\neg\varphi \wedge \neg\psi)$, $\varphi \rightarrow \psi \stackrel{\text{def}}{=} \neg(\varphi \wedge \neg\psi)$, $\varphi \leftrightarrow \psi \stackrel{\text{def}}{=} (\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi)$.

3.2 理性信念的语义

定义 3(理性信念框架 F_{RB} 与模型 M_{RB}): 给定一个二元组 $F_{RB} = \langle W, R_i \rangle$ 作为理性信念的框架, 在 F_{RB} 的基础上, 给定一个赋值函数 V , 那么三元组 $M_{RB} = \langle W, R_i, V \rangle$ 就是理性信念的模型, 其中 $(i \in N)$:

- (1) $W \neq \emptyset$, W 是非空的认知状态集,
- (2) $R_i \subseteq W \times W$ 是认知状态集上的二元等价关系,
- (3) $V: p \rightarrow \delta(W)$ 是每个原子命题对应到每个认知状态的映射函数.

定义 4(等价关系、等价框架及等价模型) 关系 R_i 是等价关系, 如果它满足以下条件:

- (1) 传递性: $\forall w \forall u \forall v (R_iwu \wedge R_iuv \rightarrow R_iwv)$;
- (2) 自反性: $\forall w (R_iww)$;
- (3) 对称性: $\forall w \forall u (R_iwu \rightarrow R_iuw)$.

若 R_i 是等价关系, 则 $F_{RB} = \langle W, R_i \rangle$ 是等价框架. 同理, 若 R_i 是等价关系, 则 $M_{RB} = \langle W, R_i, V \rangle$ 就是等价模型.

定义 5(真) 给定一个理性信念模型 $M_{RB} = \langle W, R_i, V \rangle$, 及任一状态 $w \in W$, 那么 (M_{RB}, w) 为理性信念的点模型. 若将“公式 φ 在 w 状态上为真”标记为 $M_{RB}, w \models \varphi$, 那么对于任一 L_{RB} 下公式 φ 在 (M_{RB}, w) 上的真可以定义如下:

- (1) $M_{RB}, w \models p$, 当且仅当, $w \in V(p)$;
- (2) $M_{RB}, w \models \neg \varphi$, 当且仅当, $M_{RB}, w_i \not\models \varphi$;
- (3) $M_{RB}, w \models \varphi \wedge \psi$, 当且仅当, $M_{RB}, w \models \varphi$ 且

$M_{RB}, w \models \psi$;

(4) $M_{RB}, w \models B_i\varphi$, 当且仅当, 对任意的 $w_m \in W$, 若 R_iww_m 则 $M_{RB}, w_m \models \varphi$;

(5) $M_{RB}, w \models K_i\varphi$, 当且仅当, 对任意的 $w_m \in W$, 若 R_iww_m 则 $M_{RB}, w \models \varphi$ 且 $M_{RB}, w_m \models \varphi$;

(6) $M_{RB}, w \models B_i^R\varphi$, 当且仅当, 若对任意的 $w_m \in W$ 如果 R_iww_m 那么 $M_{RB}, w_m \models \varphi$, 则对任意的 $w_n \in W$ 如果 R_iww_n 那么 $M_{RB}, w \models \varphi$ 且 $M_{RB}, w_n \models \varphi$, 或者对任意的 $w_u \in W$ 如果 R_iww_u 那么 $M_{RB}, w \models \varphi$ 且 $M_{RB}, w_u \not\models \varphi$.

其中, (1)、(2) 及 (3) 的语义解释如常; (4) 的语义是“Agent i 相信 φ 在状态 w 为真, 当且仅当, 对于任一 Agent i 认知可通达的状态 w_m , φ 在 w_m 上为真”. 而 (5) 的语义是“Agent i 知道 φ 在状态 w 为真, 当且仅当, 在状态 w 中, φ 为真, 且 Agent i 相信 φ 也为真”; 最后, (6) 的语义是“在状态 w 中, Agent i 理性地相信 φ 为真, 当且仅当, Agent i 在状态 w 中无法确定是否知道 φ 真假的情况下, 是不会相信 φ 的”.

4 认知系统 S_{RB} 及其定理

4.1 S_{RB} 中的公理及推理规则

定义 6(S_{RB}) 智能 Agent 具有理性信念的认知系统 (标记为 S_{RB}) 是由以下公理及推理规则所组成:

- (A1) 所有命题逻辑中的重言式
- (A2) $B_i(\varphi \rightarrow \psi) \rightarrow (B_i\varphi \rightarrow B_i\psi)$
- (A3) $B_i\varphi \rightarrow B_iB_i\varphi$
- (A4) $\neg B_i\varphi \rightarrow B_i\neg B_i\varphi$
- (A5) $\neg K_i\varphi \rightarrow K_i\neg K_i\varphi$
- (A6) $K_i\varphi \leftrightarrow \varphi \wedge B_i\varphi$
- (A7) $B_i^R\varphi \leftrightarrow \neg(K_i\varphi \vee K_i\neg\varphi) \rightarrow \neg B_i\varphi$

(R1) 替换规则 (Substitution): 若公式 $\varphi \in S_{RB}$, 原子命题 p , 及 $\psi \in L_{RB}$, 则 $\varphi(p/\psi) \in S_{RB}$;

(R2) 概括规则 (Generalization): 若 $\vdash \varphi$ 则 $\vdash B_i\varphi$;

(R3) 分离规则 (Modus Ponens): 若 $\vdash \varphi$ 且 $\vdash \varphi \rightarrow \psi$, 则 $\vdash \psi$.

4.2 S_{RB} 中的一些导出规则及定理

(1) 导出规则 (DR1): 若 $\vdash \varphi$ 则 $\vdash K_i\varphi$.

证明: 若 $\vdash \varphi$, 则由 (R2) 可得 $\vdash B_i\varphi$. 再由公理 (A1) 可得 $\vdash \varphi \rightarrow (B_i\varphi \rightarrow \varphi \wedge B_i\varphi)$. 再根据前提及分离规则 (R3) 可得 $\vdash \varphi \wedge B_i\varphi$. 最后由公理 (A6) 可得 $\vdash K_i\varphi$.

(2) 导出规则(DR2): 若 $\vdash \varphi \rightarrow \psi$ 则 $\vdash B_i \varphi \rightarrow B_i \psi$.

证明: 若 $\vdash \varphi \rightarrow \psi$, 则由(R2)可得 $\vdash B_i(\varphi \rightarrow \psi)$. 再由(A2)及(R3)可得 $\vdash B_i \varphi \rightarrow B_i \psi$.

(3) 导出规则(DR3): 若 $\vdash \varphi \rightarrow \psi$ 则 $\vdash K_i \varphi \rightarrow K_i \psi$.

证明: 若 $\vdash \varphi \rightarrow \psi$, 那么由(DR2)可得

(i) $\vdash B_i \varphi \rightarrow B_i \psi$

再由(A1)及(R1)可得

(ii) $\vdash (\varphi \rightarrow \psi) \rightarrow ((B_i \varphi \rightarrow B_i \psi) \rightarrow (\varphi \wedge B_i \varphi \rightarrow \psi \wedge B_i \psi))$

再由(R3)可得

(iii) $\vdash \varphi \wedge B_i \varphi \rightarrow \psi \wedge B_i \psi$

再由(A1)及(R1)可得

(iv) $\vdash (\varphi \wedge B_i \varphi \rightarrow K_i \varphi) \rightarrow ((\psi \wedge B_i \psi \rightarrow K_i \psi) \rightarrow ((\varphi \wedge B_i \varphi \rightarrow \psi \wedge B_i \psi) \rightarrow (K_i \varphi \rightarrow K_i \psi)))$

再由(A6)、(iii)及(R3)可得

(v) $\vdash K_i \varphi \rightarrow K_i \psi$

(4) 导出规则(DR4): 若 $\vdash \varphi \leftrightarrow \psi$, 则 $\vdash B_i \varphi \leftrightarrow B_i \psi$

证明: 若 $\vdash \varphi \leftrightarrow \psi$, 则有 $\vdash \varphi \rightarrow \psi$ 及 $\vdash \psi \rightarrow \varphi$. 再由(DR2)可得 $\vdash B_i \varphi \rightarrow B_i \psi$ 及 $\vdash B_i \psi \rightarrow B_i \varphi$, 即 $\vdash B_i \varphi \leftrightarrow B_i \psi$.

(5) 导出规则(DR5): 若 $\vdash \varphi \leftrightarrow \psi$, $\vdash K_i \varphi \leftrightarrow K_i \psi$

证明: 类似(DR4)的证明(略).

(6) 导出规则(DR6) 等值替换 (Equivalent Replacement): 任给公式 $\varphi, \psi \in L_{RB}$, 如果 $\vdash \varphi \leftrightarrow \psi$, 则 $\vdash S[\varphi] \leftrightarrow S[\psi]$. 其中, $S[\varphi]$ 是含有 φ 为子公式的公式, $S[\psi]$ 是用 ψ 对 $S[\varphi]$ 中出现 φ 的一处或多处进行替换而获得的公式.

证明: 对此采用归纳证明.

(i) 若 $\varphi = S[\varphi], \psi = S[\psi]$, 则显然有 $\vdash S[\varphi] \leftrightarrow S[\psi]$.

现在对 $S[\varphi]$ 中出现 $\neg, \wedge, B_i, K_i$ 及 B_i^R 的次数 n 进行归纳.

(ii) 若 $n = 0$, 则 $S[\varphi]$ 要么是 φ , 要么是原子命题 p . 而这两种情况都如证明如(i), 显然有 $\vdash S[\varphi] \leftrightarrow S[\psi]$.

(iii) 假设 $n = k$ 时, 等值替换规则成立. 那么以下是 $n = k + 1$ 的情况

(a) 若 $S[\varphi] = \neg S_1[\varphi]$, 则由归纳假设可得 $\vdash S_1[\varphi] \leftrightarrow S_1[\psi]$. 再根据(A1)可得 $\vdash \neg S_1[\varphi] \leftrightarrow \neg S_1[\psi]$, 即 $\vdash S[\varphi] \leftrightarrow S[\psi]$.

(b) 若 $S[\varphi] = S_1[\varphi] \wedge S_2[\dots]$, 则由归纳假设

可得 $\vdash S_1[\varphi] \leftrightarrow S_1[\psi]$. 再根据(A1)可得 $\vdash S_1[\varphi] \wedge S_2[\dots] \leftrightarrow S_1[\psi] \wedge S_2[\dots]$, 即 $\vdash S[\varphi] \leftrightarrow S[\psi]$.

(c) 若 $S[\varphi] = S_2[\dots] \wedge S_1[\varphi]$, 同理于(b)可得 $\vdash S_2[\dots] \wedge S_1[\varphi] \leftrightarrow S_2[\dots] \wedge S_1[\psi]$, 即 $\vdash S[\varphi] \leftrightarrow S[\psi]$.

(d) 若 $S[\varphi] = B_i S_1[\varphi]$, 则由归纳假设可得 $\vdash S_1[\varphi] \leftrightarrow S_1[\psi]$. 再由(DR4)可得 $\vdash B_i S_1[\varphi] \leftrightarrow B_i S_1[\psi]$, 即 $\vdash S[\varphi] \leftrightarrow S[\psi]$.

(e) 若 $S[\varphi] = K_i S_1[\varphi]$, 则由归纳假设可得 $\vdash S_1[\varphi] \leftrightarrow S_1[\psi]$. 再由(DR5)可得 $\vdash K_i S_1[\varphi] \leftrightarrow K_i S_1[\psi]$, 即 $\vdash S[\varphi] \leftrightarrow S[\psi]$.

(f) 若 $S[\varphi] = B_i^R S_1[\varphi]$, 同理可得 $\vdash S_1[\varphi] \leftrightarrow S_1[\psi]$. 再根据(A1)可得

(f1) $\vdash \neg S_1[\varphi] \leftrightarrow \neg S_1[\psi]$.

再由归纳假设及(DR4)可得

(f2) $\vdash B_i S_1[\varphi] \leftrightarrow B_i S_1[\psi]$.

再根据(A1)可得

(f3) $\vdash \neg B_i S_1[\varphi] \leftrightarrow \neg B_i S_1[\psi]$.

再根据(DR5)、(f1)及归纳假设可得

(f4) $\vdash K_i S_1[\varphi] \leftrightarrow K_i S_1[\psi]$.

(f5) $\vdash K_i \neg S_1[\varphi] \leftrightarrow K_i \neg S_1[\psi]$.

由(A1)可得

(f6) $\vdash (K_i S_1[\varphi] \leftrightarrow K_i S_1[\psi]) \wedge (K_i \neg S_1[\varphi] \leftrightarrow K_i \neg S_1[\psi]) \wedge (\neg B_i S_1[\varphi] \leftrightarrow \neg B_i S_1[\psi]) \rightarrow ((K_i S_1[\varphi] \vee K_i \neg S_1[\varphi] \vee \neg B_i S_1[\varphi]) \leftrightarrow (K_i S_1[\psi] \vee K_i \neg S_1[\psi] \vee \neg B_i S_1[\psi]))$

再由(f3)、(f4)、(f5)及(R3)可得

(f7) $\vdash (K_i S_1[\varphi] \vee K_i \neg S_1[\varphi] \vee \neg B_i S_1[\varphi]) \leftrightarrow (K_i S_1[\psi] \vee K_i \neg S_1[\psi] \vee \neg B_i S_1[\psi])$

再由(A1)可得

(f8) $\vdash (\neg (K_i S_1[\varphi] \vee K_i \neg S_1[\varphi]) \rightarrow \neg B_i S_1[\varphi]) \leftrightarrow (\neg (K_i S_1[\psi] \vee K_i \neg S_1[\psi]) \rightarrow \neg B_i S_1[\psi])$

再由(A1)及(R1)可得

(f9) $\vdash ((\neg (K_i S_1[\varphi] \vee K_i \neg S_1[\varphi]) \rightarrow \neg B_i S_1[\varphi]) \leftrightarrow B_i^R S_1[\varphi]) \rightarrow (((\neg (K_i S_1[\psi] \vee K_i \neg S_1[\psi]) \rightarrow \neg B_i S_1[\psi]) \leftrightarrow B_i^R S_1[\psi]) \rightarrow (((\neg (K_i S_1[\varphi] \vee K_i \neg S_1[\varphi]) \rightarrow \neg B_i S_1[\varphi]) \leftrightarrow (\neg (K_i S_1[\psi] \vee K_i \neg S_1[\psi]) \rightarrow \neg B_i S_1[\psi])) \rightarrow (B_i^R S_1[\varphi] \leftrightarrow B_i^R S_1[\psi])))$

再由(A7)及(R1)可得

(f10) $\vdash (\neg (K_i S_1[\varphi] \vee K_i \neg S_1[\varphi]) \rightarrow \neg B_i S_1[\varphi]) \leftrightarrow B_i^R S_1[\varphi]$

$$(f11) \vdash (\neg (K_i S_1[\psi] \vee K_i \neg S_1[\psi]) \rightarrow \neg B_i S_1[\psi]) \leftrightarrow B_i^R S_1[\psi]$$

再由(f8)、(f9)、(f10)、(f11)及(R3)可得

$$(f12) \vdash B_i^R S_1[\varphi] \leftrightarrow B_i^R S_1[\psi], \text{即} \vdash S[\varphi] \leftrightarrow S[\psi].$$

定理 1(内定理) 以下公式都是 S_{RB} 的定理

$$(T1) \vdash K_i \varphi \rightarrow B_i \varphi$$

$$(T2) \vdash K_i \varphi \rightarrow \varphi$$

$$(T3) \vdash K_i \varphi \rightarrow B_i^R \varphi$$

$$(T4) \vdash B_i^R \varphi \leftrightarrow (B_i \varphi \rightarrow \varphi \vee B_i \neg \varphi)$$

$$(T5) \vdash K_i(\varphi \rightarrow \psi) \rightarrow (K_i \varphi \rightarrow K_i \psi)$$

$$(T6) \vdash B_i \varphi \rightarrow K_i B_i \varphi$$

$$(T7) \vdash B_i \varphi \rightarrow B_i K_i \varphi$$

$$(T8) \vdash K_i \varphi \rightarrow K_i K_i \varphi$$

$$(T9) \vdash K_i \varphi \rightarrow B_i K_i \varphi$$

$$(T10) \vdash (B_i \varphi \rightarrow \varphi) \rightarrow B_i^R \varphi$$

证明:为了简明起见,我们约定以下证明过程中以 PC 缩写命题演算 (propositional calculus), 以 RPC 表示命题演算规则.

(T1) 与 (T2)) 可容易地由 (A1)、(A6) 及 (R3) 获得,(证明略去).

(T4) 由 (A1) 可得

$$(i) \vdash K_i \varphi \rightarrow K_i \varphi$$

由 (i) 及 RPC 可得

$$(ii) \vdash K_i \varphi \rightarrow K_i \varphi \vee K_i \neg \varphi$$

同理,由 (ii) 及 RPC 可得

$$(iii) \vdash K_i \varphi \wedge B_i \varphi \rightarrow K_i \varphi \vee K_i \neg \varphi$$

再由 (A1) 可得

$$(iv) \vdash (K_i \varphi \wedge B_i \varphi \rightarrow K_i \varphi \vee K_i \neg \varphi) \rightarrow (K_i \varphi \rightarrow (B_i \varphi \rightarrow K_i \varphi \vee K_i \neg \varphi))$$

再由 (iii)、(iv) 及 (R3) 可得

$$(v) \vdash K_i \varphi \rightarrow (B_i \varphi \rightarrow K_i \varphi \vee K_i \neg \varphi)$$

再根据 (A7) 及 (DR6) 可得

$$(vi) \vdash K_i \varphi \rightarrow B_i^R \varphi$$

(T4) 根据 (A7)、(A6) 及 (DR6) 可得

$$(i) \vdash B_i^R \varphi \leftrightarrow (B_i \varphi \rightarrow \varphi \wedge B_i \varphi \vee \neg \varphi \wedge B_i \neg \varphi)$$

再由 RPC 可得

$$(ii) \vdash B_i^R \varphi \leftrightarrow (\neg B_i \varphi \vee \varphi \wedge B_i \varphi \vee \neg \varphi \wedge B_i \neg \varphi)$$

再由德摩根律可得

$$(iii) \vdash B_i^R \varphi \leftrightarrow (\neg B_i \varphi \vee \varphi) \wedge (\neg B_i \varphi \vee B_i \varphi) \vee (\neg \varphi \wedge B_i \neg \varphi)$$

由于 $(\neg B_i \varphi \vee B_i \varphi)$ 是重言式,所以有

$$(iv) \vdash B_i^R \varphi \leftrightarrow (\neg B_i \varphi \vee \varphi) \vee (\neg \varphi \wedge B_i \neg \varphi)$$

同理,由德摩根律可得

$$(v) \vdash B_i^R \varphi \leftrightarrow (\neg B_i \varphi \vee \varphi \vee \neg \varphi) \wedge (\neg B_i \varphi \vee \varphi \vee B_i \neg \varphi)$$

同理, $(\neg B_i \varphi \vee \varphi \vee \neg \varphi)$ 是重言式,因此有

$$(iv) \vdash B_i^R \varphi \leftrightarrow (B_i \varphi \rightarrow \varphi \vee B_i \neg \varphi)$$

(T5) 由 (A1) 可得

$$(i) \vdash (\varphi \rightarrow \psi) \wedge \varphi \rightarrow \psi$$

$$(ii) \vdash (B_i \varphi \rightarrow B_i \psi) \wedge B_i \varphi \rightarrow B_i \psi$$

$$(iii) \vdash (B_i(\varphi \rightarrow \psi) \rightarrow (B_i \varphi \rightarrow B_i \psi)) \rightarrow (B_i(\varphi \rightarrow \psi) \wedge B_i \varphi \rightarrow (B_i \varphi \rightarrow B_i \psi) \wedge B_i \varphi)$$

由 (iii)、(A2) 及 (R3) 可得

$$(iv) \vdash B_i(\varphi \rightarrow \psi) \wedge B_i \varphi \rightarrow (B_i \varphi \rightarrow B_i \psi) \wedge B_i \varphi$$

再由 (ii)、(iv) 及 RPC 可得

$$(v) \vdash B_i(\varphi \rightarrow \psi) \wedge B_i \varphi \rightarrow B_i \psi$$

再由 (i)、(v) 及 RPC 可得

$$(vi) \vdash ((\varphi \rightarrow \psi) \wedge \varphi \rightarrow \psi) \wedge (B_i(\varphi \rightarrow \psi) \wedge B_i \varphi \rightarrow B_i \psi)$$

再由 (A1) 可得

$$(vii) \vdash ((\varphi \rightarrow \psi) \wedge \varphi \rightarrow \psi) \wedge (B_i(\varphi \rightarrow \psi) \wedge B_i \varphi \rightarrow B_i \psi) \rightarrow ((\varphi \rightarrow \psi) \wedge \varphi \wedge B_i(\varphi \rightarrow \psi) \wedge B_i \varphi \rightarrow \psi \wedge B_i \psi)$$

再由 (vi)、(vii) 及 (R3) 可得

$$(viii) \vdash (\varphi \rightarrow \psi) \wedge \varphi \wedge B_i(\varphi \rightarrow \psi) \wedge B_i \varphi \rightarrow \psi \wedge B_i \psi$$

再由 (A1) 可得

$$(ix) \vdash ((\varphi \rightarrow \psi) \wedge \varphi \wedge B_i(\varphi \rightarrow \psi) \wedge B_i \varphi \rightarrow \psi \wedge B_i \psi) \rightarrow ((\varphi \rightarrow \psi) \wedge B_i(\varphi \rightarrow \psi) \rightarrow (\varphi \wedge B_i \varphi \rightarrow \psi \wedge B_i \psi))$$

再根据 (ix)、(viii) 及 (R3) 可得

$$(x) \vdash (\varphi \rightarrow \psi) \wedge B_i(\varphi \rightarrow \psi) \rightarrow (\varphi \wedge B_i \varphi \rightarrow \psi \wedge B_i \psi)$$

再根据 (x)、(A6) 及 (DR6) 可得

$$(xi) \vdash K_i(\varphi \rightarrow \psi) \rightarrow (K_i \varphi \rightarrow K_i \psi)$$

(T6) 由 (A1) 可得

$$(i) \vdash B_i \varphi \rightarrow B_i \varphi$$

由 (A3) 及 RPC 可得

$$(ii) \vdash B_i \varphi \rightarrow B_i \varphi \wedge B_i B_i \varphi$$

再由 (DR6) 及 (A3) 可得

$$(iii) \vdash (B_i \varphi \rightarrow B_i \varphi \wedge B_i B_i \varphi) \leftrightarrow (B_i \varphi \rightarrow K_i B_i \varphi)$$

再由 (R3) 可得 $\vdash B_i \varphi \rightarrow K_i B_i \varphi$

(T7) 由 (A1) 及 (R1) 可得

$$(i) \vdash B_i \varphi \rightarrow B_i \varphi$$

再由 (A3) 及 RPC 可得

$$(ii) \vdash B_i \varphi \rightarrow B_i \varphi \wedge B_i B_i \varphi$$

再由 (DR6) 及 (T10) 可得

(iii) $\vdash (B_i\varphi \rightarrow B_i\varphi \wedge B_i B_i\varphi) \leftrightarrow (B_i\varphi \rightarrow B_i(\varphi \wedge B_i\varphi))$

再由(R3)可得

(iv) $\vdash B_i\varphi \rightarrow B_i(\varphi \wedge B_i\varphi)$

再由(DR6)及(A6)可得

(v) $\vdash B_i\varphi \rightarrow B_i K_i\varphi$

(T8) 由(A1)及(R1)可得

(i) $\vdash \varphi \wedge B_i\varphi \rightarrow \varphi \wedge B_i\varphi$

(ii) $\vdash \varphi \wedge B_i\varphi \rightarrow B_i\varphi$

再由(ii)、(A3)及RPC可得

(iii) $\vdash \varphi \wedge B_i\varphi \rightarrow B_i B_i\varphi$

再由RPC可得

(iv) $\vdash \varphi \wedge B_i\varphi \rightarrow (\varphi \wedge B_i\varphi) \wedge B_i\varphi \wedge B_i B_i\varphi$

再由(DR6)及(T10)可得

(v) $\vdash \varphi \wedge B_i\varphi \rightarrow (\varphi \wedge B_i\varphi) \wedge B_i(\varphi \wedge B_i\varphi)$

再根据(DR6)及(A6)可得

(vi) $\vdash K_i\varphi \rightarrow K_i K_i\varphi$

(T9) 根据(T1)及(R1)可得

(i) $\vdash K_i K_i\varphi \rightarrow B_i K_i\varphi$

再由(T10)及RPC可得

(ii) $\vdash K_i\varphi \rightarrow B_i K_i\varphi$

(T10) 由(A1)及(R1)可得

(i) $\vdash \neg B_i\varphi \rightarrow \neg B_i\varphi$

(ii) $\vdash \varphi \rightarrow \varphi$

再由(i)、(ii)及RPC可得

(iii) $\vdash \neg B_i\varphi \rightarrow \neg B_i\varphi \vee \varphi \vee B_i\neg\varphi$

(iv) $\vdash \varphi \rightarrow \neg B_i\varphi \vee \varphi \vee B_i\neg\varphi$

再根据(A1)及(R1)可得

(v) $\vdash (\neg B_i\varphi \rightarrow \neg B_i\varphi \vee \varphi \vee B_i\neg\varphi) \rightarrow ((\varphi \rightarrow \neg B_i\varphi \vee \varphi \vee B_i\neg\varphi) \rightarrow (\neg B_i\varphi \vee \varphi \rightarrow \neg B_i\varphi \vee \varphi \vee B_i\neg\varphi))$

再根据(iii)、(iv)、(v)及(R3)可得

(vi) $\vdash \neg B_i\varphi \vee \varphi \rightarrow \neg B_i\varphi \vee \varphi \vee B_i\neg\varphi$

再根据RPC、(DR6)及(T4)可得

(vii) $\vdash (B_i\varphi \rightarrow \varphi) \rightarrow B_i^R\varphi$

4.3 S_{RB} 的可靠性及完全性

定理 2 任给 S_{RB} 的模型 $M_{RB} = \langle W, R_i, V \rangle$ 及公式 $\varphi, \psi \in L_{RB}$, 那么以下公式在 M_{RB} 上有效:

(1) $B_i(\varphi \rightarrow \psi) \rightarrow (B_i\varphi \rightarrow B_i\psi)$

(2) $B_i\varphi \rightarrow B_i B_i\varphi$

(3) $\neg B_i\varphi \rightarrow B_i\neg B_i\varphi$

(4) $\neg K_i\varphi \rightarrow K_i\neg K_i\varphi$

(5) $K_i\varphi \leftrightarrow \varphi \wedge B_i\varphi$

(6) $B_i^R\varphi \leftrightarrow (\neg(K_i\varphi \vee K_i\neg\varphi) \rightarrow \neg B_i\varphi)$

证明: 其中(1)、(2)、(3)及(5)根据赋值定义可容易得到证明, 在此略去.

(4) 假设存在 $(M_{RB}, \omega) \models \neg K_i\varphi \rightarrow K_i\neg K_i\varphi$, 那么由(A6)可得

(i) $(M_{RB}, \omega) \models \neg(\varphi \wedge B_i\varphi) \rightarrow \neg(\varphi \wedge B_i\varphi) \wedge B_i\neg(\varphi \wedge B_i\varphi)$

再由“定义 5”可得

(ii) $(M_{RB}, \omega) \models \neg(\varphi \wedge B_i\varphi)$, 及

(iii) $(M_{RB}, \omega) \models \neg(\varphi \wedge B_i\varphi) \wedge B_i\neg(\varphi \wedge B_i\varphi)$

再由(iii)及“定义 5”可得

(iv) $(M_{RB}, \omega) \models \neg(\varphi \wedge B_i\varphi)$, or

(v) $(M_{RB}, \omega) \models B_i\neg(\varphi \wedge B_i\varphi)$

再由(v)及“定义 5”可得

(vi) 存在 v , 使得 $R_i\omega v$ 且 $(M_{RB}, v) \models \neg(\varphi \wedge B_i\varphi)$

由“定义 3”可知 R_i 是等价关系, 再结合“定义 4”, 可得: 当 $R_i\omega\omega$, 则存在

(vii) $(M_{RB}, \omega) \models \neg(\varphi \wedge B_i\varphi)$

由于(iv)及(vii)矛盾于(ii), 所以假设不成立, 即应该有

(vii) $(M_{RB}, \omega) \models \neg K_i\varphi \rightarrow K_i\neg K_i\varphi$

(6) $(M_{RB}, \omega) \models BR_i\varphi$

当且仅当, 若 $(M_{RB}, \omega) \models \neg(K_i\varphi \vee K_i\neg\varphi)$ 则 $(M_{RB}, \omega) \models \neg B_i\varphi$

当且仅当, $(M_{RB}, \omega) \models \neg(K_i\varphi \vee K_i\neg\varphi) \rightarrow \neg B_i\varphi$

定理 3(可靠性) 任给 S_{RB} 的模型 $M_{RB} = \langle W, R_i, V \rangle$ 及公式 $\varphi \in L_{RB}$, 那么若 $\vdash \varphi$ 则 $M_{RB} \models \varphi$.

证明: 仅需证明每一个公理及推理规则在 M_{RB} 上保有效. 而由“定理 2”即可得到证明.

定义 7(一致性) 任给公式集 $\Gamma \subseteq \text{Form}(L_{RB})$, Γ 是一致的, 当且仅当, $\Gamma \not\vdash \perp$.

定义 8(极大一致集) 任给公式集 $\Theta \subseteq \text{Form}(L_{RB})$, Θ 是极大一致集, 当且仅当

(i) Θ 是一致的;

(ii) 对于任意 $\varphi \notin \Theta$, $\{\varphi\} \cup \Theta$ 是不一致的.

定理 4(一致性性质) 任给极大一致集 Θ , 及任一公式 $\varphi, \psi \in L_{RB}$, 则存在

(i) $\neg\varphi \in \Theta$, 当且仅当, $\varphi \notin \Theta$;

(ii) $\varphi \rightarrow \psi \in \Theta$, 当且仅当, 若 $\varphi \in \Theta$ 则 $\psi \in \Theta$;

(iii) 若 $\varphi \rightarrow \psi \in \Theta$ 且 $\varphi \in \Theta$, 则 $\psi \in \Theta$;

(iv) 若 $\varphi \in \Theta$ 且 $\varphi \in \Theta$, 则 $\varphi \wedge \psi \in \Theta$;

(v) 若 $\varphi \in \Theta$ 或 $\varphi \in \Theta$, 则 $\varphi \vee \psi \in \Theta$.

证明: 由于此处一致性性质定理无异于其他数

理逻辑上的一致性性质定理,因此在此略去其证明.

定理 5(Lindenbaum) 任给一致公式集 $\Gamma \in \text{Form}(L_{RB})$, Γ 可以扩充为极大一致集 Θ .

证明:同理,该性质定理的证明可从其他数理逻辑教材上获得,在此略去证明过程.

定义 9 给定任一模型 $M_{RB}^C = \langle W^C, R_i^C, V^C \rangle$ 为 S_{RB} 典范模型,当且仅当满足:

- (i) $W^C = \{\Theta \mid \Theta \text{ 是极大一致的}\}$;
- (ii) $R_i^C \Theta \Theta' := B_i \Theta \subseteq \Theta'$;
- (iii) $V^C(p) := \{\Theta \in W^C \mid p \in \Theta\}$.

定理 6 给定任一模型 $M_{RB}^C = \langle W^C, R_i^C, V^C \rangle$ 为 S_{RB} 的典范模型,及任一公式 $\varphi \in L_{RB}$,则对于任一 $\Theta \in W^C$,有

$$(M_{RB}^C, \Theta) \models \varphi, \text{ 当且仅当, } \varphi \in \Theta \quad (3)$$

证明:使用归纳法证明,施归纳于 φ .

(i) 归纳基始:若 φ 是原子命题 p ,则由定义显然有

$$(M_{RB}^C, \Theta) \models p, \text{ 当且仅当, } p \in \Theta.$$

(ii) 归纳假设:对于任一公式 $\varphi \in L_{RB}$ 及任一极大一致集 Θ ,假设有

$$(M_{RB}^C, \Theta) \models \varphi, \text{ 当且仅当, } \varphi \in \Theta$$

(iii) 施归纳于 φ 的结构:

$$(a) \text{ 若 } \varphi = \neg \psi, \text{ 则 } (M_{RB}^C, \Theta) \models \neg \psi$$

$$\text{当且仅当, } (M_{RB}^C, \Theta) \models \psi \text{ (根据“定义 5”)}$$

$$\text{当且仅当, } \psi \notin \Theta \text{ (根据归纳假设)}$$

$$\text{当且仅当, } \neg \psi \in \Theta \text{ (根据“定理 4”)}$$

$$(b) \text{ 若 } \varphi = \psi_1 \wedge \psi_2, \text{ 则 } (M_{RB}^C, \Theta) \models \psi_1 \wedge \psi_2$$

$$\text{当且仅当, } (M_{RB}^C, \Theta) \models \psi_1 \text{ 且 } (M_{RB}^C, \Theta) \models \psi_2 \text{ (根据“定义 5”)}$$

$$\text{当且仅当, } \psi_1 \in \Theta \text{ 且 } \psi_2 \in \Theta \text{ (根据归纳假设)}$$

$$(c) \text{ 若 } \varphi = B_i \varphi, \text{ 则分两种情况}$$

$$\text{如果 } B_i \varphi \in \Theta,$$

那么,对于任一 Θ' ,若 $R_i^C \Theta \Theta'$ 则 $\varphi \in \Theta'$ (根据“定义 9”)

那么,对于任一 Θ' ,若 $R_i^C \Theta \Theta'$ 则 $(M_{RB}^C, \Theta') \models \varphi$ (根据归纳假设)

$$\text{那么, } (M_{RB}^C, \Theta) \models B_i \varphi \text{ (根据“定义 5”)}$$

$$\text{如果 } B_i \varphi \notin \Theta,$$

$$\text{那么, } \neg B_i \varphi \in \Theta \text{ (根据“定理 4”)}$$

那么,存在 Θ' ,使得 $R_i^C \Theta \Theta'$ 且 $\neg \varphi \in \Theta'$ (根据“定义 5”)

那么,存在 Θ' ,使得 $R_i^C \Theta \Theta'$ 且 $\varphi \notin \Theta'$ (根据“定理 4”)

$$\text{那么,存在 } \Theta', \text{ 使得 } R_i^C \Theta \Theta' \text{ 且 } (M_{RB}^C, \Theta') \models \varphi \text{ (根$$

据归纳假设)

$$\text{那么, } (M_{RB}^C, \Theta) \models B_i \varphi \text{ (根据“定义 5”)}$$

$$(d) \text{ 若 } \varphi = K_i \varphi, \text{ 则 } (M_{RB}^C, \Theta) \models K_i \varphi$$

$$\text{当且仅当, } (M_{RB}^C, \Theta) \models \varphi \wedge B_i \varphi \text{ (根据“定理 2(5)”)}$$

$$\text{当且仅当, } (M_{RB}^C, \Theta) \models \varphi \text{ 且 } (M_{RB}^C, \Theta) \models B_i \varphi \text{ (根据“定义 5”)}$$

$$\text{当且仅当, } \varphi \in \Theta \text{ 且 } \varphi \wedge B_i \varphi \in \Theta \text{ (根据归纳假设及(c))}$$

$$\text{当且仅当, } \varphi \wedge B_i \varphi \in \Theta \text{ (根据“定理 4”)}$$

$$\text{当且仅当, } K_i \varphi \in \Theta \text{ (由 (A6) 及“定理 4”)}$$

$$(5) \text{ 若 } \varphi = B^R_i \varphi, \text{ 则 } (M_{RB}^C, \Theta) \models B^R_i \varphi$$

$$\text{当且仅当, } (M_{RB}^C, \Theta) \models \neg (K_i \varphi \vee K_i \neg \varphi) \rightarrow \neg B_i \varphi \text{ (根据“定理 2(6)”)}$$

$$\text{当且仅当, } (M_{RB}^C, \Theta) \models K_i \varphi, \text{ or } (M_{RB}^C, \Theta) \models K_i \neg \varphi, \text{ or } (M_{RB}^C, \Theta) \models \neg B_i \varphi \text{ (由“定义 5”)}$$

$$\text{当且仅当, } K_i \varphi \in \Theta, \text{ or } K_i \neg \varphi \in \Theta, \text{ or } \neg B_i \varphi \in \Theta \text{ (由(a) 及(c))}$$

$$\text{当且仅当, } K_i \varphi \vee K_i \neg \varphi \vee \neg B_i \varphi \in \Theta \text{ (根据“定理 4”)}$$

$$\text{当且仅当, } (\neg (K_i \varphi \vee K_i \neg \varphi) \rightarrow \neg B_i \varphi) \in \Theta$$

$$\text{当且仅当, } B^R_i \varphi \in \Theta \text{ (由(A7) 及“定理 4”)}$$

定理 7(完全性): 给定任一 S_{RB} 模型 $M_{RB} = \langle W, R_i, V \rangle$, 以及任一公式 $\varphi \in L_{RB}$, 那么若 $M_{RB} \models \varphi$ 则 $\vdash \varphi$.

证明:给定任一 S_{RB} 的典范模型 $M_{RB}^C = \langle W^C, R_i^C, V^C \rangle$, 及任一公式集 $\varphi \in \text{Form}(L_{RB})$. 假设 $\not\models \varphi$, 则有

$$(i) \{\neg \varphi\} \text{ 是一致集;}$$

$$\text{再由“定理 5”可得}$$

$$(ii) \{\neg \varphi\} \text{ 可扩充为极大一致集 } \Theta, \text{ 且 } \neg \varphi \in \Theta;$$

$$\text{再由“定理 6”可得}$$

$$(iii) (M_{RB}^C, \Theta) \models \varphi$$

5 结束语

在建立智能 Agent 理性信念认知系统 S_{RB} 的基础上,本文证明了 S_{RB} 同时具有可靠性及完全性. 而在 S_{RB} 中,本文也证明了一些定理,从而得出相应的认知性质. 其一,定理(T1) $\vdash K_i \varphi \rightarrow B_i \varphi$ 及(T2) $\vdash K_i \varphi \rightarrow \varphi$ 体现了“知识”定义的必要条件,即相信一个真实的命题是知道一个命题的必要条件;而(T3) $\vdash K_i \varphi \rightarrow B^R_i \varphi$ 表明了智能 Agent 保持理性信念是其获得知识的必要条件. 事实上,我们还可以证明出 $\vdash \neg B_i \varphi \rightarrow B^R_i \varphi$, 再结合(T1) 及(T3),就可得

出: 认知主体获得知识的必要条件不是简单的相信某个事实命题就行, 而是对于认知主体必须持有一种理性的信念. 这就揭示了传统知识定义上的不足, 这种不足在于以往并没有对知识定义中的信念做出细分研究, 而事实上, 许多生物认知主体对外部世界感应都会或多或少地带有情感因素, 这些情感因素无疑干扰了认知主体的知识推理. 比如, 生物认知主体往往会由于无知而迷信, 或由于从众而盲信, 在这些不具备理性信念下的知识获取显然不是真正意义上的“知识”, 进一步也无助于心智中的认知计算, 相反只会出现错误的推理结论. 与此不同的是, 对人工智能上的认知主体, 智能 Agent 往往不受情感因素的干扰, 也就是说理性信念是成为它的认知推理的必要条件, 也是其与生物认知主体推理的区别之一. 这也是本文研究智能 Agent 认知推理的信念条件的原因. 其二, 进一步地, 由 (T4) $\vdash B^R_i \varphi \leftrightarrow (B_i \varphi \rightarrow \varphi \vee B_i \neg \varphi)$ 可知, 在智能 Agent 持有理性信念进行认知计算时, 如果它相信某个命题 φ , 则要么该命题就是事实命题, 要么它只相信该命题的反命题. 需要注意的是, (T4) 对于生物认知主体一般难以做到, 这样的认知能力有点过强, 而生物认知主体一般认知能力具有明显的局限性. 其三, 由 (T5)、(T6)、(T7)、(T8) 及 (T9) 可知, 在理性信念认知系统 S_{RB} 中, 认知活动中的“知道”与“相信”与一般认知系统中的“知道”与“相信”无异, 即由 (T5) 可知“知识可分配性”同样成立, (T6) 与 (T9) 说明“知道”与“相信”具有互知性, 而 (T8) 正是传统上的“正内省性”. 其四, (T10) $\vdash (B_i \varphi \rightarrow \varphi) \rightarrow B^R_i \varphi$ 表达了“智能 Agent 如果相信了某个命题则该命题一定是真的, 那么智能 Agent 一定是持有理性信念去相信该命题的”, 换句话说, 如果智能 Agent 不是理性地相信某个命题, 那么它相信的命题则是假的.

需要说明的是, 在人工智能中, 智能 Agent 与外部环境互动的认知活动往往是呈现出动态的过程, 这就需要我们采用动态化的理性信念来计算认知. 更进一步地, 若是智能 Agent 具有理性信念, 而生物认知主体不具有理性信念, 当智能 Agent 与生物认知主体两者进行互动时 (例如人机对话), 则情况会进一步复杂化, 此时需要考虑的不再只是理性信念的认知, 也不只是非理性的认知, 而是需要计算在理性与非理性之间的认知以及博弈的策略等因素. 这些问题都将是我们将接下来将要进一步研究的内容.

参考文献:

- [1] Müller V C, Bostrom N. Future progress in artificial intelligence: A survey of expert opinion[C]//Fundamental issues of artificial intelligence. Berling: Springer International Publishing, 2016: 553-570.
- [2] Russell S J, Norvig P, Canny J F, et al. Artificial intelligence: a modern approach[M]. England: Pearson Education Limited, 2014: 34-61.
- [3] Mandler M. Rational agents are the quickest[J]. Journal of Economic Theory, 2015, 155: 206-233.
- [4] Kahneman D, Tversky A. Can irrationality be intelligently discussed? [J]. Behavioral and Brain Sciences, 1983, 6(3): 509-510.
- [5] Boukhris I, Elouedi Z, Benferhat S. Dealing with external actions in belief causal networks[J]. International Journal of Approximate Reasoning, 2013, 54(8): 978-999.
- [6] Visser S, Thangarajah J, Harland J, et al. Preference-based reasoning in BDI agent systems[J]. Autonomous Agents and Multi-Agent Systems, 2016, 30(2): 291-330.
- [7] Alchourrón C E, Gärdenfors P, Makinson D. On the logic of theory change: partial meet contraction and revision functions [J]. Journal of Symbolic Logic, 1985, 50: 510-530.
- [8] Grove A. Two modellings for theorychange[J]. Journal of philosophical logic, 1988, 17(2): 157-170.
- [9] Ghahramani Z. Probabilistic machine learning and artificial intelligence [J]. Nature, 2015, 521 (7553): 452-459.
- [10] Hamilton E, Cairns H. Plato: The collected dialogues [M]. Princeton: Princeton University Press, 1961: 853.
- [11] Hintikka J. Knowledge and belief: An introduction to the logic of the two notions[M]. Ithaca: Cornell University Press, 1962: 3-5.
- [12] Williamson T. Knowledge and its Limits[M]. Oxford: Oxford University Press on Demand, 2002: 2-20.

作者简介:

高金胜 男, (1983-), 博士研究生. 研究方向为认知逻辑、知识推理. E-mail: Hi. Jackson@163.com.

周昌乐 男, (1959-), 博士, 教授. 研究方向为人工智能、计算语言学、理论脑科学和认知逻辑.