22-26 认知硬度 默认分层理论 (4)

计算机科学1993Vol.20No.2

认知硬度与默认分层理论

TP18

白硕
（国家智能计算机研究开发中心 北京100080）
（北京大学数学系 北京100871）

摘 要

The concept of "epistemic stiffness" is introduced in this paper. By attaching epistemic stiffness labels to first order wffs, a natural deduction system that reasons about hierarchies of defauls is actually built. It can combine uncertain reasoning and nonmonotonic reasoning cohesively, can choose among conflicting believes through argumentation games, and thus can obtain nondeductive conclusions in a logic system with fine deductive properties.

一、引 言

信念是缺乏充分理性基础的判断，是知识不足、资源不足条件下适应环境、指导行为所不可缺少的认知手段。一个能对信念进行表示和推理的系统，应包括下述三个要素：(1) 信念的逻辑内容；(2) 对信念的评价及其随推理过程的传播；(3) 对相互冲突的信念的选择。根据对信念之组成的这种认识，我们来看一看传统上对信念进行逻辑刻画的几个主要途径。

1. 模态逻辑。模态逻辑更多地着眼于把"信念"定性地施用于自身，即处理"关于信念的信念"。然而，在涉及到评价和选择时，模态逻辑几乎没有什么贡献[1]。

2. 非单调逻辑。非单调逻辑把信念看成是一些相对于已有知识（包括已有信念）而言可错的逻辑公式，一旦遇到反驳，它们会被拒斥；但如果反驳不存在，它们就被认可。将来知识充分了，反驳出现了，它们再被拒斥。在这里，"知识"和"信念"的逻辑地位有明显的差别："知识"是"硬"的，能推出的就认可；"信念"是"软"的，即使能推出还要看有没有"硬"的对立面。就是说，"知识"和"信念"这两个标记就相当于两个分立的"评价值"，而非单调机制则在其中起"选择"作用[2][3][4]。

3. 不确定推理。不确定推理大都采用量化指标为信念内容指派其"不确定性"或"可信度"，并规定其随推理过程传播的具体办法。这一部分显然相当于"评价"。但到目前为止，对信念的评价（度量）方案大都受到多值逻辑和概率论的严重干扰。特别是，它们都无法使非单调逻辑处理信念的方案与之有机地结合为一体。这就使得处理信念的这两个主

则来构造容错系统，容错系统的扩张相当于一定的相容性维护，如极大相容子理论。

容错逻辑有许多直接的应用，例如，语义网中继承推理典型地是一种容错推理，例如，[Thomason, Horty 1988]给出的继承逻辑可作为一般容错逻辑的一种特形，再如，容错专家系统能克服专家系统的脆弱性瓶颈问题，建造具有常识的专家系统。容错推理还需要进行深入的研究，特别是，我们希望看到一般的容错数学基础。（参考文献略）

白硕 博士，1990-1992在北京大学数学系作博士后研究，主要研究方向为常识推理、机器学习、计算语言学、AI基础。

要阵营之间一直缺乏足够的沟通，难以取得使两个阵营都为之关注的突破和进展[5]。

作者主张用不确定推理的思想来扩展、改造非单调逻辑，使信念的内容和评价都能在一个演绎性质良好、逻辑结构清晰的形式系统内得到处理，而对相互冲突的信念的选择，则通过某种"辩论"的过程来完成。

要实现这样的目标，关键的一步是引入正确的评价体系。本文引入"认知硬度"（epistemic stiffness）作为对信念的评价指标。通俗地说，一个信念的"认知硬度"就是它在对立信念的驳斥下坚持自己的"顽固"程度。例如，如果公式p获得评价θ_1，$\neg p$获得评价θ_2，而θ_1比θ_2"硬"，那么二者比较而言，我们将乐于接受p而拒绝$\neg p$。如果p从另外的途径获得评价θ_3，而θ_3比θ_1"硬"，那么我们就会反过来倾向于接受$\neg p$。进而，如果p不可能从任何途径获得"硬"于θ_3的评价，它就被$\neg p$所驳倒，$\neg p$就被最终认可。由此可见，采用具有这样认知背景的评价体系，是与基于辩论的选择机制分不开的。正是在这一点上，我们把不确定推理和非单调逻辑有机地结合起来了。

本文把"认知硬度"标记引入作为扩展的逻辑公式的一部分，建立了一个能对"一阶公式＋硬度标记"型的扩展逻辑公式进行推理的自然演绎系统——SH，在其中自然而然地实现了标记（即评价）随推理过程的传播。在此基础上，本文建立了基于辩论过程的选择机制。特别地，如果硬度标记只取"S(Soft)"和"H(Hard)"两个分立值，那么上述系统实现的恰好是传统的非单调机制。因此，具有不同硬度标记的"信念"公式，实际上相当于不同层级的默认(Default)规则或由它们参与推出的结论。也就是说，本文提出的形式理论，从非单调逻辑的角度看可以认为是一种默认分层理论。

二、自然演绎系统SH

设$\langle\Sigma, \succ\rangle$为一个有穷的全序集，它的最大元为H，其余的依序为$S_1 \succ S_2 \cdots \succ S_n$。注意这里"$\succ$"是严格序，也就是说，$\theta_1 \succ \theta_2 \Rightarrow \theta_1 \neq \theta_2$。我们称这个全序集为一个**硬度结构**。

定义2.1 一个SH合式公式是一个形如$p:\theta$的公式，其中p为一阶公式，$\theta \in \Sigma$称为p的**硬度标记**，简称**硬度**。$p:\theta$读作："p，硬度为θ"。

设Γ，$\triangle$是SH合式公式的有穷序列，θ_1，$\theta_2 \in \Sigma$。下面是SH自然演绎规则：

(1) $p_1:\theta_1 \cdots p_n:\theta_n \vdash p_i:\theta_i (i=1,\cdots,n)$ [肯定前提律，简记$\in$]

(2) 如果$\Gamma \vdash \triangle \vdash p:\theta$（$\triangle$不空），则$\Gamma \vdash p:\theta$[传递律，简记$\tau$]

(3) 如果Γ，$\neg p:H \vdash q:H, \neg q:H$，则$\Gamma \vdash p:H$[反证律，简记$\neg$]

(4) $\neg\neg p:\theta \vdash p:\theta$，$p:\theta \vdash \neg\neg p:\theta$[双重否定律，简记$\neg\neg$]

(5) $p \to q:\theta_1$，$p:\theta_2 \vdash q:\min(\theta_1, \theta_2)$，

其中 $\min(\theta_1, \theta_2)=\begin{cases}\theta_2 & 若\theta_1 \succ \theta_2 \\ \theta_1 & 不然。\end{cases}$

[蕴涵词消去律，简记$\to_-$]

(6) 如果Γ，$p:\theta_1 \vdash q:\theta_2$，则$\Gamma \vdash p \to q:\min(\theta_1, \theta_2)$ [蕴涵词引入律，简记$\to_+$]

(7) $p \wedge q:\theta \vdash p:\theta$，$q:\theta$ [合取词消去律，简记$\wedge_-$]

(8) $p:\theta_1$，$q:\theta_2 \vdash p \wedge q:\min(\theta_1, \theta_2)$ [合取词引入律，简记$\wedge_+$]

(9) 如果 $p:\theta_1 \vdash r:\theta_3$；$q:\theta_2 \vdash r:\theta_4$，则 $p \vee q:\min(\theta_1, \theta_2) \vdash r:\max(\theta_3, \theta_4)$，

其中 $\max(\theta_3, \theta_4)=\begin{cases}\theta_3 & 若\theta_3 \succ \theta_4 \\ \theta_4 & 不然。\end{cases}$

[析取词消去律，简记$\vee_-$]

(10) $p:\theta \vdash p \vee q:\theta, q \vee p:\theta$ [析取词引入律，简记$\vee_+$]

(11) $p \longleftrightarrow q:\theta_1, p:\theta_2 \vdash q:\min(\theta_1, \theta_2)$

$p \longleftrightarrow q:\theta_1, \theta:\theta_2 \vdash p:\min(\theta_1, \theta_2)$

[等值词消去律，简记$\longleftrightarrow_-$]

(12) 如果 Γ，$p:\theta_1 \vdash q:\theta_2$；$\Gamma$，$q:\theta_3 \vdash p:\theta_4$，则 $\Gamma \vdash p \longleftrightarrow q:\min(\min(\theta_1, \theta_2), \min(\theta_3, \theta_4))$[等值词引入律，简记$\longleftrightarrow_+$]

(13) $\forall x P(x):\theta \vdash P(a):\theta$

[全称量词消去律，简记$\forall_-$]

(14) 如果$\Gamma \vdash P(a):\theta$，a不在$\Gamma$中出现，则 $\Gamma \vdash \forall x P(x):\theta$ [全称量词引入律，简记$\forall_+$]

(15) 如果$P(a):\theta_1\vdash q:\theta_2$，a不在q中出现，则 $\exists xP(x):\theta_1\vdash q:\theta_2$ [存在量词消去律，简记$\exists_-$]

(16)$P(a):\theta\vdash \exists xP(x):\theta$，其中$P(x)$为用x部分替换$P(a)$中的a而得。[存在量词引入律，简记$\exists_+$]

自然演绎系统SH_k是胡世华、陆钟万提出的F^*系统的一个扩充[6]。有关的推理规则在一阶公式部分基本上保持了原有的形式，增加的只有硬度标记部分。我们对其中的关键之处作一些必要的说明。

1. 规则(3)及(4)是SH系统处理信念间逻辑冲突的较有特色之处。规则(3)说明，只有"硬碰硬"的矛盾才真正具有破坏力，因而可用来进行"否定前提"式的反证法推理。而在矛盾的一方是"软"的即可错的情况下，SH系统采取了"容许矛盾存在，但不使矛盾有破坏力"的作法，不允许使用反证法来传播、扩散矛盾。这种作法，与所谓"paraconsistent logic"[7]是相通的，但这里的标记不是真值，其传播特性也不同于真值，故本质上不同于前人的系统。规则（4）干脆就是排中律，但它没有写成 $\vdash p\vee\neg p:\theta$ 的形式。这是由于在SH里，$p\vee q$与$\neg p\to q$不是可以无条件互换的，归根结底是因为反证律的使用是有条件的，它只适用于标记为H的公式。

2. SH的大部分自然演绎规则都涉及到硬度标记随推理过程的传播。一般来说，如果结论是直接由前提推出的，那么最"软"的前提将会把其硬度标记传给结论；如果前提与结论间的推论关系是由另外的推论关系所决定的，则要依具体情况对硬度关系进行传递、缩小及放大等不同处理。

由SH可以得出许多形式推论。我们在这里重点说明两个情况：

1. 如果$\Gamma=\{p_i:H,\ i=1,2,\cdots,N\}$，则q是$p_1\cdots,p_N$在传统意义上的演绎推论，当且仅当$\Gamma\vdash q:H$。换句话说，一阶谓词演算的传统演绎推理是SH推理的特例。

2. 我们可以证明：若$\theta_1\succ\theta_2$，则$p:\theta_1\vdash p:\theta_2$。用"斜形证明"的格式来写这个证明，就是：

(1)	$p:\theta_1$	(假设)
(2)	$p:\theta_2$	(假设)
(3)	$p:\theta_2$	($\in$)
(4)	$p\to p:\theta_2$	(2)(3)($\to_+$)
(5)	$p:\theta_2$	(1)(4)($\to_-$)

这一形式推论关系的意义在于：由于推理途径的不同，关于某一一阶公式p可能存在一系列的硬度标记$\theta_1\cdots,\theta_K$，使$\Gamma\vdash p:\theta_i(i=1,\cdots,k)$。然而我们最终只需关心其中最"硬"的一个就可以了，因为有了最大硬度标记，其它标记的合理性就自动得到了保证。也正因为如此，相当于"证据合并"的规则(9)中才需要引入max函数。

关于信念及其评价的表示与推理，我们已通过SH系统的自然演绎规则作了全面的描述。下面，我们将进入"选择"这一层次，讨论SH在相互冲突的信念间如何进行取舍的机制。

三、SH中的辩论与选择

定义3.1 设给定SH合式公式的有限集Γ。我们称θ是一阶公式p在Γ下的**最大硬度**当且仅当$\Gamma\vdash p:\theta$且若$\Gamma\vdash p:\theta'$，则$\theta\succ\theta'$。p在Γ下的最大硬度记做$St(p,\Gamma)$。若$\Gamma\nvdash p:\theta$，则$St(p,\Gamma)$无定义，记做$St(p,\Gamma)=\perp$。

为方便起见，我们把$\perp$也引进硬度结构，并规定对任何θ，有$\theta\succ\perp$。

定义3.2 若$St(p,\Gamma)\succ St(\neg p,\Gamma)$，我们称p可以在$\Gamma$下**驳倒**(defeats)$\neg p$，若$St(p,\Gamma)=St(\neg p,\Gamma)$，称p在$\Gamma$下与$\neg p$持平(draw)。

定义3.3 若p在Γ下可以驳倒$\neg p$，则称p在Γ下是**被认可的**(accepted)，称$\neg p$在Γ下是**被拒斥的**(rejected)。

定义3.4 p在Γ下是**被强认可的**，如果下列条件满足：

(1) p在Γ下是被认可的；

(2) 设R是在Γ下被认可的任何一组一阶公式的集合。若R协调，则R∪{p}也是协调的。

有了上述定义，我们实际上为SH在选择层次上建立了一套基于辩论的决策机制。可以设想，给定前提集Γ和一阶公式p，SH的推理机制就分解为辩论着的正反双方。正方要证出具有尽可能"硬"的硬度标记θ的SH公式p:θ，反方也要尽量提高¬p的硬度标记以驳倒p。这样一种辩论的对局一般要经过一个图灵不可判定的过程才能对最终的潜在胜方有准确的判断。然而，在实际运行中，时间资源总是有限的。我们可把在时间资源用尽时p与¬p双方获得的最大实际硬度作为各自最大（理想）硬度的近似，从而实现作者在文[8]中所定义的T－非单调性。

设$\Sigma=\langle\{H,S\},\succ\rangle$，$\theta_1\succ\theta_2$当且仅当$\theta_1=H,\theta_2=S$。我们通过例子来看SH如何实现文[8]中定义的K－非单调性，也就是传统的基于不完全知识资源的非单调性。

例1 设$\Gamma=\{p\vee q:H,\ \neg p:S,\ \neg q:S,\ \neg r:S\}$，则$p\vee q$，$\neg p$，$\neg q$，$\neg r$是Γ下被认可的；$\neg p\wedge\neg q$是Γ下被拒斥的；$p\vee q$，$\neg r$是Γ下被强认可的。

例2：设$\Gamma=\{\forall x(B(x)\rightarrow F(x)):S,\ B(a):H,\ \neg F(a):H,\ B(b):H\}$，则$\forall x(B(x)\rightarrow F(x))$是Γ下被拒斥的，F(b)是Γ下被认可的，而且是被强认可的。

例1说明，两个被认可的一阶公式的合取有可能是被拒斥的，这就是著名的"彩票悖论"(Lottery paradox)[8]。只有引进"强认可"的概念，悖论才会消除。在承认Γ的前提下，¬p被认可，因为没有理由否认¬p；¬q被认可，因为没有理由否认¬q。然而，却不能说没有理由否认$\neg p\wedge\neg q$。事实上，Γ决定了$p\vee q$的无条件成立，这就使得¬p和¬q不可能在同一个模型中为真。用默认逻辑的话说，¬p和¬q可以在不同的扩张中分别找到各自成立的理由，但却不能在同一扩张中成立。而被强认可的¬r，在所有扩张中都是成立的。

例2揭示了另一个有趣的现象。要想推出F(b):S，必须用到$\forall x(B(x)\rightarrow F(x)):S$。而最后选择的结果，$\forall x(B(x)\rightarrow F(x)):S$被拒斥了，F(b)却被认可，而且是被强认可。就是说，反例可以驳倒一个全称陈述，但这不等于驳倒了该全称陈述的所有演绎推论。从另一个角度说，一个全称陈述如果自认是"可错的"(即"有例外的")，那么它在反例前面就有了一定程度的"弹性"，至少使它的大部分演绎推论不受影响。这种典型的非单调现象，在SH中得到了准确的刻划。

以上两例都是$\Sigma=\langle\{H,S\}\rangle$ ($\theta_1\succ\theta_2$当且仅当$\theta_1=H$，$\theta_2=S$)这种特殊的硬度结构下SH系统的推理实例。实际上，对于这类硬度结构而言，SH的推理能力基本上不超过传统的非单调逻辑（全面分析SH与现存各非单调逻辑的关系已超出本文的范围）。我们只想指出，对于更复杂的硬度结构来说，SH的推理能力可以大大超出传统非单调逻辑的范围。

例3 [Tversky悖论]以70%:30%的比例选一批工程师和律师构成一个样本，从中随机抽取一人，向被试描述其有善于交际、衣着讲究、关心政治等特点。现在问被试：这个人是工程师还是律师？被试普遍回答是律师[10]。现在我们把被试的推理过程用SH加以形式化的表达。

设P(x)表"x符合关于律师的身份描述特征"，Q(x)表"x符合关于工程师的身份描述特征"，L(x)表"x是律师"，E(x)表"x是工程师"，S(x)表"x是从题述样本中随机抽取的"。取硬度结构$\langle\{H,S_1,S_2\},\succ\rangle$，$H\succ S_1\succ S_2$则已知条件可形式化为

$$\Gamma=\{\forall x(P(x)\rightarrow L(x)):S_1,\ \forall x(Q(x)\rightarrow E(x)):S_1,\ \forall x(S(x)\rightarrow E(x)):S_2,\ \forall x(E(x)\rightarrow\neg L(x)):H,\ P(a):H,\ S(a):H\}$$

可求得

$$St(E(a))=St(\neg L(a))=S_2$$

$$St(L(a))=St(\neg E(a))=S_1$$

就是说，L(a)，¬E(a)是被（强）认可的，¬L(a)，E(a)是被拒斥的。

这说明，SH的选择机制倾向于认为这个人是律师。

在这个例子里，关键之处在于“先验”硬度分配。即把身份特征描述置于比样本中的比例更“硬”的地位。一旦身份特征描述符合了律师这一身份的要求，在样本中的比例这一因素就下降到次要地位，这与我们的直观是非常符合的。顺便指出，用现有非单调逻辑来处理这一问题，无法摆脱“多扩张”问题的困扰。而用现有不确定推理方法来处理这一问题，也因缺乏连贯的理论解释而陷于悖论。在SH中，有确定因素与非单调因素通过带硬度标记的推理统一起来了，这一问题也因之获得了简单而圆满的解答。

四、结束语

本文是一项很初步的工作。基于辩论的选择机制是在证明论的范围内叙述的，它的模型论语义基础的工作尚未完成。此外，这里处理的主要是表示和推理，关于信念及其评价的获取问题是极其重要的，但本文已无法涉及。本文中Γ里各公式的评价部分显得似乎带有“先验”性，实际上，它们都应是学习的结果，反馈的结果。把整个问题放到一个大循环当中来看，硬度是可以动态升降的，而受到“例外”攻击的历史记录，是决定硬度升降的关键。作者觉得，如能把获取、表示和推理完整地连成一体，必将在对信念的认识上产生一个很大的飞跃。

最后，作者想说明，认知硬度不是也不应该是对信念的唯一评价尺度。非单调逻辑界最近十分关注的“Specificity”[11]很可能是另一个重要的评价指标。如果能在SH中加进Specificity方面的因素，必将会覆盖更多、更有意义的一类非单调现象。作者已在这个方向上有初步的进展，具体结果将在另文中给出。

感谢李未教授、林作铨先生、王献昌博士的启发和帮助。感谢曲非同志为本文所提的建设性意见。（参考文献略）

（接第60页）

(2)用公式(8)计算μ_{ij}($i=1,\cdots,m$, $j=1\cdots,n$)；

(3) 置k为0；

(4) 若$p_g \geq \alpha$，停止规划；

(5) 用公式(9)计算SEU(a_i)($i=1,\cdots,m$)，选取对应于SEU(a_i)($i=1,\cdots,m$)的最大值的动作作为a^k；

(6)用公式(11)计算新的状态概率分布$p(s_j^{k+1})$，$j=1,\cdots,n$；

(7)k增加1；

(8)转向步骤(4)。

上述步骤产生的规划为$a^0a^1a^2\cdots a^k\cdots$，且$a^k \in A$($k=0,1,\cdots$)。若开始时$p_g \geq \alpha$，则不需规划。

五、结束语

本文提出了一种传统人工智能和决策论相结合的非确定性规划方法。在这种方法中，两种传统的规划方法步骤是相互交叉的。要使这种方法实际可行，必须在以下两方面进一步开展研究。第一，构造好的效用函数以便加快规划进程。第二，构造符合实际的动作-状态因果关系模型。这两项工作都密切依赖于具体的应用领域。在非确性规划中，规划终止问题非常重要。好的效用函数有助于但不能保证规划终止。（参考文献略）