

容错推理

(国家智能机研究中心, 北京100080)

林作铨

(汕头大学计算机系, 汕头515063)

TP18

摘 要

Fault-tolerant is a property of commonsense. Formalizing commonsense reasoning can be described by the problem of reasoning with fault-tolerant knowledge. Based on the concept of the maximal consistency, we define the maximally consistent subtheories as a special approach to formalizing fault-tolerant reasoning. We furthermore propose the preferential semantics for the general approach of fault-tolerant logics.

1. 引 论

人在日常生活中进行推理的能力包含诸多方面, 是相当复杂的过程, 但常识推理的最一般基础是在容错知识情形下的推理。

一般地, 容错知识包括两种最基本情形: 一种是容“错”知识, 即含矛盾的知识, 这是在不相容知识情形下的推理问题; 另一种是容“误”知识, 即含未知的知识, 这是在不完全知识情形下的推理问题。特别地, 两种容错知识情形是相互联系的, 构成常识推理的一般性质。

我们说知识是完全的, 是指对任一命题 A , 都有 A 或 $\neg A$ 可推出。完全知识通常是可望而不可得的, 在不完全(含未知)知识下推理通常需要先验地做出一些假设, 这些假设可能随进一步获取的新知识被废除。不完全推理是非单调的, 标准逻辑却具有单调性。AI中非单调逻辑研究这种容“误”或不完全知识推理问题。一般说来, 非单调推理是一种把不完全知识转换为某种完全知识的过程, 这个完全知识状态是相对的, 随进一步增加知识需要作出修正, 而这种修正通过发现矛盾维护知识的相容性达到的。

我们说知识是相容的, 是指对任一命题 A , A 与 $\neg A$ 不能同时推出。同样, 相容知识也是可望而不可得的, 在不相容(含矛盾)知识下推理通常需

要使矛盾孤立或局部化, 不影响推出与矛盾无干的结论。这种推理是非平凡的, 标准逻辑却具有平凡性。非协调推理研究这种容“错”或不相容知识推理问题, 哲学中的超协调逻辑与相干逻辑可以看成非协调推理的形式化。一般说来, 非协调推理是一种在不相容知识下寻找某种相容知识的过程, 这种相容知识状态排除矛盾的影响, 从而维护某种知识的相容性。

我们认为容错推理研究一方面是划分各种“错误”知识情形, 而另一方面, 亦更重要的是把各种“错误”知识情形容纳在一起, 构成常识推理的容错基础, 我们将在后一方面意义下研究容错推理的形式化, 并定义容错逻辑作为不完全与不相容知识情形下容错推理的形式化。

我们发现, 把非单调推理与非协调推理结合在一起可以克服各自的困难问题, 例如, 非单调逻辑都有平凡性, 因而存在反常模型问题; 非协调逻辑都是单调的, 因而存在矛盾扩散问题。容错逻辑是非单调与非平凡的, 不存在这些问题。从矛盾观点看, 容错推理所容纳的各种“错误”知识的处理都能通过矛盾起作用, 例如, 非单调推理通过发现矛盾维护知识, 非协调推理通过限制矛盾维护知识, 其结果都是维护知识状态的相容性, 因此, 通过矛盾维护知识相容性可以作容错的一般技术途径。

本文基于极大相容性概念, 定义极大相容子理

林作铨 副教授, 从事非单调推理研究。

论作为容错逻辑的一种特殊形式。我们给出优先语义途径作为容错逻辑的一般形式,最后,我们讨论一些有关容错推理的进一步研究问题。

2. 极大相容子理论

我们拟定义容错逻辑 $\vdash'$ 同时具有如下性质:

(1) 对任意 T, A , 若 T 是相容与完全的, 则 $T \vdash' A$ 当且仅当 $T \vdash A$, 这里 $\vdash$ 表示标准逻辑;

(2) 对任意 A , 存在 B , 使得 $\{A, \neg A\} \vdash' B$

(3) 对任意 A, C , 存在 B , 使得 $A \vdash' C$, 但 $\{A, B\} \vdash' C$

条件(1)表明容错逻辑包括标准逻辑作为特殊情形, 即在相容与完全知识下容错逻辑化归为标准逻辑, 这是考虑到: 一则我们希望在标准逻辑基础上构造容错逻辑, 二则AI研究经验表明在相容与完全知识下标准逻辑是好用而且必要的, 条件(2)是非平凡性, 这是不相容知识推理的一般性质, 使得容错逻辑是一种含矛盾而非平凡的理论; 条件(3)是非单调性, 这是不完全知识下推理的一般性质, 使得容错逻辑是一种含未知而非单调的理论。据此, 我们将在标准逻辑基础上定义容错逻辑。

余文中, 我们令 L 是任一标准逻辑, 如一阶逻辑或模态逻辑等等, 下面, 我们设定 L 为一阶逻辑, 有关定义如常。

定义2.1 令 T 是一个理论, A 为任一公式。称 T 是不相容的, 若 $T \vdash A$ 且 $T \vdash \neg A$; 一个理论是相容的, 若它不是不相容的, 称 $\{A, \neg A\}$ 是一个矛盾。

定义2.2 令 T_1, T_2 是两个理论。称 T_2 是 T_1 的子理论, 记如 $T_2 \subseteq T_1$, 若对任一公式 A , $T_2 \vdash A$ 则 $T_1 \vdash A$ 。

我们记 $T_1 = T_2$ 若 $T_2 \subseteq T_1$ 且 $T_1 \subseteq T_2$, 这时称 T_1 与 T_2 是相同的。注意到, 子理论定义相当于子集。

定义2.3 令 T, S 是理论。 S 称为 T 的极大相容子理论, 当且仅当: 1. S 是相容的;

2. $S \subseteq T$; 3. 不存在相容 $S' \subseteq T$ 使得 $S \subsetneq S'$ 。

我们记 $S = \text{Mcs}(T)$, T 的极大相容子理论就是 T 的子理论中相对于(偏序)关系 $\subseteq$ 的极大元。记 $\text{MCS}[T] = \{S \mid S = \text{Mcs}(T)\}$ 表示 T 的所有极大相容子理论集。

容易看出, 极大相容子理论 MCS 是标准逻辑中极大相容性概念的直接推广, 当 T 表示一阶逻辑的全部公式集时, 则极大相容(子)理论就是一阶逻辑的极大相容集。

命题2.1 令 T 是一个非空理论, 则存在 T 的极大相容子理论。

命题2.2 令 T 是一个相容理论, 则 $\text{Mcs}(T) = T$, 即相容理论的极大相容子理论就是自身。

基于极大相容子理论, 我们现在定义一种简单容错逻辑, 其可证关系 $\vdash_\subseteq$ 定义了 MCS 推理, 可以充分说明容错推理形式。

定义2.4 令 T 是一个理论, A 为任一公式, 容错可证关系 $\vdash_\subseteq$ 定义如下:

1. 强可证关系: $T \vdash_\subseteq A$ 当且仅当 对所有 $\text{MCS}(T)$, $\text{MCS}(T) \vdash A$

2. 弱可证关系: $T \vdash_\subseteq A$ 当且仅当 存在一个 $\text{MCS}(T)$, $\text{MCS}(T) \vdash A$

以下我们亦记 $\vdash_\subseteq$ 表示强定义形式。现在, 论证 $\vdash_\subseteq$ 满足容错逻辑的三个基本性质。

命题2.3 若 T 是相容的, 则 $T \vdash_\subseteq A$ 当且仅当 $T \vdash A$ 。

由命题2.2容易直接推出, 它说明 $\vdash_\subseteq$ 是标准逻辑的推广, 注意到, 在一阶逻辑中 $\vdash$ 是完全的。

命题2.4 $\vdash_\subseteq$ 是非平凡的。

容易看出, 存在命题 B , $\{A, \neg A\} \vdash_\subseteq B$ 。

命题2.5 $\vdash_\subseteq$ 是非单调的。

容易知道, 若 $A \vdash_\subseteq C$, 那么存在 $B = \neg A$, 可使 $\{A, B\} \vdash_\subseteq C$ 。

值得一提的是, 因为 $\text{MCS}(T)$ 在一阶逻辑

辑中不是递归可枚举的,所以 $\neg$ 一般是不可判定的,对命题逻辑MCS(T)可判定,但其计算一般是NP性的。

极大相容子理论的想法决不是新的,它是极大相容集的推广,[Rescher 1964]第一个定义极大相容子理论来形式化假言推理,Rescher基于命题逻辑并在语言中确定优先性以形成假设理论,在AI中[McCarthy, Hayes 1969]第一次看到它的作用,这一想法也体现在各种非单调推理形式化中,如TMS系统是一种维护相容性技术,默认逻辑通过相容性确定扩张首先直接用极大相容子理论定义默认推理,[Brewka 1989]明确指出这一点并推广到优先默认理论,另外,如[Lin 1987]指出极大相容子理论可作为非协调推理,[Morris 1988]说明了非单调推理与非协调推理的共性,但据我们所知,提出极大相容子理论作为形式化非协调与非单调推理的容错逻辑却是新的想法。在下节中,我们将基于MCS理论发展一种容错逻辑的一般基础。

3. 优先语义

从模型论看,在标准逻辑中,若 $A \models C$,由于满足 $A \wedge B$ 的模型也是 A 的模型,则有 $A \wedge B \models C$,由于 $A \wedge \neg A$ 没有模型,故 $A \wedge \neg A \models C$,这是推理的随意性;容错逻辑不仅具有标准逻辑推理能力,而且描述在不完全与不相容前提下作出容错推理,这是推理的先验性,这样,在容错逻辑中,若 $A \models' C$,由于对 A 的模型有一定的优先选择,如 A 的模型的某些子集,满足 $A \wedge B$ 的模型子集不一定是满足 A 的那些模型子集,因此有 $A \wedge B \models' C$,同理,由于对 $A \wedge \neg A$ 可以选择一些有模型的子集,如 A (或 $\neg A$)的模型,因此有若 $A \models' C$ 则 $A \wedge \neg A \models' C$ 。这种选择先验模型的途径称为优先语义,[Shoham 1987]用优先语义作为非单调推理的一般基础,[林 1990]给出一种先验逻辑形式,同样想法在决策论中亦可见([Wellman, Doyle 1991]),我们将推广优先语义到更一般情形

作为各种容错推理的统一基础。

令 L 是一个标准逻辑, L 的作用是提供标准模型论,如Tarski-Kripke型语义,记 $MOD(T)$ 表示理论 T 的所有子集的模型集, $Mod(T)$ 表示理论 T 自身的模型集,显然,在 L 中若 T 是相容的,则 $MOD(T) = Mod(T)$ 。记 $\models$ 表示 L 的可满足关系。一般容错逻辑 $L_{\sqsubseteq}$ 是基于标准逻辑 L 之上,根据一定优先选择准则确定模型之间优先关系 $\sqsubseteq$,定义相应优先模型获得的。

定义3.1 容错逻辑 $L_{\sqsubseteq}$ 定义如下:

1. 语法:如同标准逻辑 L
2. 语义:定义容错语义后承 $\models_{\sqsubseteq}$ 如下: $T \models_{\sqsubseteq} A$ 当且仅当 A 在所有 T 的 $\sqsubseteq$ -模型(亦称容错模型)中都可满足。

定义3.2 M 称为理论 T 的容错模型,若

1. $M \in MOD(T)$
2. 不存在 $\sqsubseteq$ -序列(优先序列)中其它模型 M' 使得 $M \sqsubseteq M'$

定义3.3 令 W 是任意理论,任一 $MOD(W)$ 中非空序列 $s = \langle M_1, M_2, \dots \rangle$ 构成一个 $\sqsubseteq$ -序列,若对任意 $M_i, M_j \in s, M_i \sqsubseteq M_j$,并且不存在 $M \notin s$ 使得 $M_k \sqsubseteq M, M_k \in s$ 。为简单起见,令 $\sqsubseteq$ -序列是有界的,即不存在无穷 $\sqsubseteq$ -序列 $M_1 \sqsubseteq M_2 \sqsubseteq \dots$ 。

定义3.4 令 W 是任意理论, $\sqsubseteq$ 是 $MOD(W)$ 上的任意二元关系,即 $\sqsubseteq: MOD(W) \times MOD(W)$, $\sqsubseteq$ 称为优先关系。

注意,我们对优先关系不作任何限制,特别地, $\sqsubseteq$ 关系不一定是有序关系。容易看出,当 $\sqsubseteq$ 是等价关系(特别是空关系)时, $L_{\sqsubseteq}$ 化归为标准逻辑 L 。注意到, $L_{\sqsubseteq}$ 中的 $\sqsubseteq$ -模型, $\sqsubseteq$ -可满足性与 $\sqsubseteq$ -有效性一般地不同于标准逻辑情形。

定义3.5 称 $\sqsubseteq$ 是完全的,若对任意理论 S ,

1. $MOD(S) \subseteq MOD(0)$ 即 L 的所有模型。
2. $\sqsubseteq$ -模型满足 S 。

定理3.1 令 $L_{\sqsubseteq}$ 是一个容错逻辑具有

完全优先关系 $\sqsubseteq$ ，则对任一理论 S ， S 在 $L_{\sqsubseteq}^*$ 中 $\sqsubseteq$ -可满足（或有效）当且仅当 S 在 L 中可满足（或有效）。

定理3.1说明了在相容与完全情形下 $L_{\sqsubseteq}$ 化归为标准逻辑 L ，即容错逻辑包括标准逻辑作为特殊情形，等价关系（空关系）是完全 $\sqsubseteq$ 的特殊形式。

值得指出，在 $L_{\sqsubseteq}^*$ 定义中，若 $\text{MOD}(S) = \text{Mod}(S) \subseteq \text{MOD}(0)$ ， $L_{\sqsubseteq}$ 相当于[Shoham 1987]中非单调优先逻辑，其中 $\sqsubseteq$ 是有序关系，如（严格）偏序关系（[林 1990]）或半序关系（[Brown, Shoham 1988]），若只限定 $\text{MOD}(S) = \text{Mod}(S)$ ， $L_{\sqsubseteq}$ 相当于[Rychlik 1990]中无序关系。[Lin 1991]结果表明，非单调优先逻辑能捕捉各种非单调逻辑作为特殊形式；优先语义同样可捕捉各种非协调逻辑作为特殊形式，例如，基于[Dunn 1976]与[Lin 1987]中给出的非协调逻辑证明论，可以完全地被基于 $\text{MOD}(T)$ 模型中的优先模型所刻画。

现在，我们证明一般容错逻辑 $L_{\sqsubseteq}$ 能捕捉极大相容子理论 $L \sqsubseteq$ 作为特殊情形，这只要定义优先关系 $\sqsubseteq$ 表达相容子理论关系。下面，我们简记： $\text{Wff}(M)$ 表示满足模型 M 的所有公式集。

定义3.6 令 T 是一个理论， $M_1, M_2 \in \text{MOD}(T)$ ，定义 $M_2 \sqsubseteq M_1$ 若 $\text{Wff}(M_1) \subseteq \text{Wff}(M_2)$ 。

以上定义是相当简化的，实际上，定义优先关系基于一定的优先准则条件，优先准则可以进一步形式化（[林 1990]），例如，对优先准则 $\text{Wff}(M)$ 我们可以通过赋值函项确定优先规则，以选择一个理论中相容的子理论。容易看到，极大相容子理论是 $L_{\sqsubseteq}$ 的一种简单情况，即定义优先关系的理论 W 就是定义优先模型的理论 T 。可以指出，[Priest 1990]提出的非单调超协调逻辑其实是极大相容子理论的一种转化形式。

定理3.2 令 T 是一个理论， A 为任一公式，则 $T \sqsubseteq A$ 当且仅当 $T \vdash A$ 。

证：由定义可看出等价性。

优先语义可以进一步推广，我们能在一个较简单容错逻辑 $L_{\sqsubseteq 1}$ 基础上定义一个新的更丰富容错逻辑 $L_{\sqsubseteq 2}$ 。例如，我们可以在标准逻辑 L 基础上定义一个非协调逻辑 $L_{\sqsubseteq 1}$ ，然后在 $L_{\sqsubseteq 1}$ 基础上定义一个非单调逻辑 $L_{\sqsubseteq 2}$ ；反之亦然。这样，例如，我们可以把极大相容子理论加进默认逻辑抑或相干逻辑定义中，从而获得各种相干默认逻辑。

4. 讨论

容错是常识的特征，容错推理容纳各种错误知识而能作出合理推论，特别是，在不完全与不相容知识情形下推理是容错推理的一般基础，在容错框架中可以描述各类常识本体论（如时间空间）与各种常识推理能力（如行动规划）。我们尝试用容错推理作为形式化常识推理的途径，为此必须进一步研究常识中各种容错特征，以致建立足够一般的统一的内在独立的容错基础，我们希望容错逻辑可以解决这一问题。

我们基于优先语义给出一种容错推理的统一逻辑基础，它具有相当丰富的表达能力，能够刻画各种容错特征，不仅能够定义容纳的各种“错误”知识推理，如 $L_{\sqsubseteq 1} \dots, L_{\sqsubseteq n}$ ，而且可以把各种“错误”知识推理容纳在一起，构成统一内在的容错系统 $L_{\sqsubseteq}$ ，使得 $\sqsubseteq$ 集成各种容错的优先关系作为统一的优先关系，即 $\sqsubseteq = \sqsubseteq_1 \dots \sqsubseteq_n$ 。

容错逻辑实质上基于一定的相容性维护机制，我们捕捉极大相容子理论作为容错逻辑的一种特殊情形，从而说明了通过形式化相容性维护可作为常识推理的一般技术，而且容易通过维护相容性技术实现容错逻辑。当然，容错逻辑一般地是不可计算的，但我们可以寻找它的可计算部分，例如，基于Horn理论的容错逻辑相当于计算Herbrand域的优先模型，可纳入一种容错逻辑程序设计途径。其实，计算优先模型相当于一定的相容性维护，因此，基于优先语义的容错逻辑也能用证明论给出，通过各种容错论证规

认知硬度与默认分层理论

(国家智能计算机研究开发中心 北京100080)

TP18

白 硕

(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 defaults 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。