ARGUMENTATION SYSTEMS: REASONING IN A CONTEXT OF DISAGREEMENT

论辩系统：不一致情境中的推理

廖备水 著

ZHEJIANG UNIVERSITY PRESS
浙江大学出版社

ARGUMENTATION SYSTEMS: REASONING IN A CONTEXT OF DISAGREEMENT

论辩系统:不一致情境中的推理

廖备水 著

ZHEJIANG UNIVERSITY PRESS
浙江大学出版社

图书在版编目（CIP）数据

论辩系统：不一致情境中的推理 / 廖备水著. —杭州：浙江大学出版社，2012.9
ISBN 978-7-308-10526-2

Ⅰ. ①论… Ⅱ. ①廖… Ⅲ. ①智能控制－自动控制系统－研究 Ⅳ. ①TP273

中国版本图书馆 CIP 数据核字（2012）第 207266 号

论辩系统：不一致情境中的推理

廖备水 著

责任编辑 李峰伟（lifwxy@zju.edu.cn）
封面设计 王波红
出版发行 浙江大学出版社
（杭州市天目山路 148 号 邮政编码 310007）
（网址：http://www.zjupress.com）
排　　版 杭州好友排版工作室
印　　刷 德清县第二印刷厂
开　　本 787mm×1092mm 1/16
印　　张 9.75
字　　数 238 千
版 印 次 2012 年 9 月第 1 版 2012 年 9 月第 1 次印刷
书　　号 ISBN 978-7-308-10526-2
定　　价 30.00 元

本书系国家自然科学基金项目“辩论推理系统的语义计算：一种基于划分的方法及其实现”(批准号：61175058)和国家社会科学基金重大项目“基于逻辑视域的认知研究”(批准号：11&ZD088)的阶段性成果，由浙江大学语言与认知研究国家创新基地资助出版。

前　言

在非理想的情况下，智能主体(Agent)需要在各类冲突的情境中进行推理，以便尽可能合理地认识世界，作出决策，或与其他 Agent 实现一定目的的交互(如说服、协商、对话等)。那么，由此产生的问题是：能否建立一个一般的形式理论来刻画 Agent 在冲突情境中的推理模式？

可喜的是，近年来从人工智能和逻辑学研究领域发展起来的论辩系统形式体系有望为这个问题提供一个肯定的回答。与经典逻辑不同，论辩系统形式体系的核心任务是实现对不完全的、不确定的和不一致的知识和信息的表示和推理，得到合理的(或可接受的)结论。现有研究表明，论辩理论不仅可以表达个体 Agent 的各种非单调推理机制，而且便于对多 Agent 交互时的推理模式进行形式刻画。

由于论辩系统是一个新兴的研究领域，国内相关学者对这一领域还不是很熟悉。因此，本书的目的是为了梳理论辩系统形式体系的产生背景、基本概念、核心理论及方法，以期为论辩系统这一研究领域在中国的发展起到抛砖引玉的作用。

为了达到该目的，本书首先分析在不同应用背景下，Agent 在进行各种推理时所面临的共同问题，并在此基础上引入论辩系统的基本概念(第 1 章)。接着，系统阐述论辩系统“静态方面”的形式理论，包括论辩系统的表示(第 2 章)、论辩系统的语义描述(第 3 章)和论辩系统的语义求解(第 4 章)。随后，简要介绍论辩系统“动态方面”的相关理论和方法(第 5 章)。此外，由于论辩系统的语义求解属于 NP 问题，如何高效计算论辩系统的语义是目前所面临的核心问题之一。本书第 6 章着重讨论论辩系统语义求解(包括静态方面和动态方面)的高效性问题，并给出一种高效求解论辩语义的基础理论：论辩语义的局部性与可组合性。本书的第 7 章则介绍论辩系统形式体系在 Agent 非单调推理中的应用情况。最后，本书的第 8 章给出了总结性阐述。

目前，有关论辩系统形式体系及其应用的各种理论和方法尚处于快速发展的过程中。本书介绍的内容不仅不能体现这一领域的全貌，而且所介绍的一些理论也不够成熟。不过，能够将这一生机勃勃的研究领域介绍给各位读者，还是让笔者感到非常兴奋。

该领域的发展态势可以从如下数据得到部分反映：(1)1995 年泰国学者 Phan Minh Dung 在人工智能顶级期刊 *Artificial Intelligence* 上发表了一篇题为“*On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games*”的文章，提出抽象论辩框架和基于外延的论辩语义的概念，标志着论辩系统形式体系作为一个新的研究领域的兴起；(2)近年来，一系列专题会议相继推出，包括“自然论证的计算模型国际研讨会(CMNA)”(始于 2001 年，每年一次)、“多 Agent 系统中

的论辩国际研讨会（ArgMAS）”（始于 2004 年，每年一次）、“论证的计算模型国际会议（COMMA）”（始于 2006 年，两年一次）；（3）人工智能领域的顶级国际会议（包括 IJCAI，AAAI，KR 等）均把论辩理论作为一个重要的主题；（4）人工智能国际顶级期刊 *Artificial Intelligence* 推出关于论辩理论的研究专辑（2007 年第 10 至第 15 期）；（5）以论辩理论为主题（或主题之一）的学术刊物开始出现，包括 *Argument & Computation* 国际期刊（2010 年创刊），*Journal of Logic and Computation* 国际期刊（开设形式论辩专题）等。

从上述数据中我们不难窥探出论辩系统这一新兴研究领域的发展轨迹。作为一种非单调推理形式体系，论辩系统之所以受到如此广泛的关注，其根本原因在于这种形式体系不仅可以刻画各种不同的推理模式（既包括个体 Agent 的内部推理，如认识推理和实践推理，又包括多 Agent 交互时的推理），而且它具有模块化的特点，贴近于人类的日常推理模式。鉴于这些原因，论辩系统形式体系有着广泛的应用领域，例如智能机器人、电子商务、语义 Web、机器学习、法律推理、医学推理等。

此外，论辩系统形式体系也属于逻辑学领域的研究范畴。因此，本书的读者至少包括人工智能（计算机科学、自动化）和逻辑学方向的学生和学者。

由于时间仓促，加上笔者水平有限，书中一定存在一些错误或问题，敬请专家和读者不吝赐教。

廖备水

2012 年 6 月

目　录

第1章 导 论

1.1 引 言

如何运用不完全、不一致、不确定的信息进行正确、高效的推理是人工智能领域的一个重要主题。近20年来，论辩(Argumentation)*作为一种新的非单调推理形式体系得到了快速发展。

论辩的概念最早起源于亚里士多德的《论题篇》和《辩谬篇》，意指实践的、应用的逻辑。20世纪80年代以来，作为非形式逻辑的论辩得到了充分发展。1995年，泰国学者 Phan Minh Dung 在 *Artificial Intelligence* 上发表了一篇关于论辩理论的重要论文，阐明了人们在论辩中使用的基本机制，并揭示了在计算机上实现这种机制的方法[1]。该理论表明，人工智能和逻辑程序设计中大多数非单调推理方法都是论辩理论的特殊形式。此后，关于论辩系统的语义、计算理论和算法，以及论辩理论在各个领域中的应用等研究如雨后春笋般出现。

论辩这一研究领域的产生和发展有着重要的应用背景和理论背景。一方面，从应用背景的角度看，处于现实世界和虚拟世界中的智能主体(Agent)在认识世界，作出决策，或与其他 Agent 实现一定目的的交互(如说服、协商、对话等)时，需要进行不同形式的推理(明确的或隐含的)。由于 Agent 认知能力的局限性、世界的动态性和不确定性，以及 Agent 间利益、价值观、偏好和立场等的差异性，在特定情境中，Agent 只能依据不完全的、不一致的、不确定的信息进行推理决策，或在相互冲突的交互中取得平衡(包括达成共识、实现说服、形成妥协等)。另一方面，从理论背景的角度看，经典一阶逻辑和传统非单调逻辑理论已经为 Agent 在上述应用中的推理进行了不同层面的形式刻画，因此可以在一定程度上阐明实际世界中 Agent 的推理模式。然而，由于经典一阶逻辑拒斥矛盾，而传统非单调逻辑主要处理认识推理，因此需要在它们的基础上，建立更一般的形式体系，使之能够在更大范围上满足 Agent 的各种推理需求。

在这一节中，我们将首先介绍论辩系统产生的应用背景(包括个体 Agent 的认识推理、实践推理，以及多 Agent 的对话与协商)和理论背景。在此基础上，引入论辩系统的基本概念。最后，介绍本书的主要内容与结构。

---

* 在本书中，为简单起见，所有用于明确词语含义的英文名词均采用单数、首字母大写的形式。

1.2 论辩系统产生的应用背景

1.2.1 个体 Agent 的认识推理

认识推理(又称为理论推理)是对命题态度(如知识、信念等)进行的推理。依据哲学认识论,一个根本问题是解释人类如何可能从外部世界获取知识。同样的,对于人工理性 Agent,也面临同样的问题。人工 Agent 必须能够获得关于世界的信息,并形成可靠的信念。为了达到该目的,一种简单的方法是在设计的时候把所有的知识内建到人工 Agent 的知识库中。然而,如果要让 Agent 可以在一个复杂的不断变化的环境中运行,这种方法显然是不可行的。因此,Agent 的设计者至少需要解决如下三个问题[2]:

第一,感知不一定是真实的。世界不总是与它所呈现的完全一样。由于这个原因,自动把感知信息转化为信念并不合适。

第二,感知实际上是一种形式的采样。由于认知的局限性,一个 Agent 不能在所有时间检测到世界的所有状态。换句话说,Agent 所感知到的信息只是世界的某个局部、某个片段。由于知识的不完全性,因此 Agent 的推理是试探性的。

例 1.1 考虑一个简单的例子。假设一个 Agent 有知识集合:"典型地,鸟会飞"和"企鹅不会飞"。如果该 Agent 得知"Tweety 是一只鸟",而没有其他信息,那么它只能进行试探性的推理,得出一个似真的结论"Tweety 会飞"。后来,当该 Agent 又得知"Tweety 是一只企鹅"时,它将得出一个与前面相冲突的结论"Tweety 不会飞"。

第三,世界在变化。Agent 的推理应该能够适应一个演化的世界。当世界发生变化时,Agent 的信念需要随之改变。

1.2.2 个体 Agent 的实践推理

实践推理(Practical Reasoning)是对非命题态度(如愿望、目标、动作等)的推理,它主要包含两个方面:慎思(Deliberation)和手段—目的推理。其中,慎思是 Agent 决定它想要达到的事件状态(即它的愿望),而手段—目的推理则是 Agent 寻找实现这些目标的途径,即动作或规划。

首先,当 Agent 在进行慎思时,同一时刻可能存在着不同的动机(内部愿望、外部义务等)。由于资源的有限性,Agent 无法实现所有的目标,从而引起动机的冲突。

例如,当 Agent 参与电子商务时,可能存在规范(Norm)冲突。在特性的虚拟社会中,规范是用来描述哪些行为是被鼓励的、禁止的或命令的。当一个 Agent 受到多个规范约束时,不同的规范之间以及规范与 Agent 的内部愿望之间可能存在冲突。

由于冲突的存在,Agent 需要依据特定的评判标准(如道德上的、利益上的、审美上的等)作出选择。

其次,对于手段—目的推理,需要考虑在任何给定的情境中所有可用的选项。由于可能存在相互对抗的选项,在作出最后的决策之前,需要进行仔细的考虑。请考虑如下的例子:

例 1.2 下面是一个实践三段论的例子:

- G 是 Agent X 的一个目标;

- 执行动作 A 是 Agent X 实现目标 G 的充分条件；
- 所以，Agent X 必须执行动作 A。

上述实践三段论所推出的结论并不意味着动作是可靠的或最好的。可能的原因包括：

(1)动作 A 不是可执行的；

(2)存在其他用于实现目标 G 的更合适的动作；

(3)执行动作 A 会带来其他效果。

1.2.3 多 Agent 交互中的推理

推理不仅出现于信念修正、慎思、决策和手段—目的推理等个体 Agent 的内部心智活动中，而且发生于多 Agent 之间的交互中，如辩论和协商等。

首先，日常辩论涉及特定观点的可辩护性问题。

例 1.3 让我们考虑《列子·汤问》中《两小儿辩日》的例子：

> 孔子东游，见两小儿辩斗，问其故。
> 一儿曰："我以日始出时去人近，而日中时远也。"
> 一儿以日初出远，而日中时近也。
> 一儿曰："日初出大如车盖，及日中则如盘盂，此不力远者小而近者大乎？"
> 一儿曰："日初出沧沧凉凉，及其日中如探汤，此不力近者热而远者凉乎？"
> 孔子不能决也。
> 两小儿笑曰："孰力知多乎？"

在这个例子中，两个孩子的观点都有相应的理由，但它们之间相互冲突。为了得出两个孩子的观点是否成立(可辩护)，需要依据特定的评价标准进行推理。如果采用怀疑的评价标准，那么两个孩子的观点都不采纳，因为它们互相攻击，但分不出胜负；而如果采用轻信的评价标准，那么两个孩子的观点都是可辩护的。

第二，在法庭中出现的推理中，参与论辩的各方所依据的前提集一般是不完全的、不一致的、动态的。对于一个证据，一旦它的反面是可变化的，该证据就会被推翻，而推理的结论也会因新证据的加入而被推翻。

第三，在多 Agent 系统中，协商是一种形式的交互，其中一组存在利益冲突但有意合作的 Agent 为了分割稀缺的资源(如商品、服务、时间、金钱等)而开展说服和讨价还价活动，以达成一个双方均可接受的协议。在协商交互的过程中，Agent 的知识、利益、偏好和目标等通常不是相互重叠的，因此冲突的情景经常发生。为了解决冲突，参与交互的 Agent 需要有相应的机制。

1.2.4 各种应用的共同特点

上述各种应用具有如下共同特点：

第一，需要在不一致的情景中进行理性推理，得到合理的(或可接受的)结论。尽管引起冲突的背后原因不同，但各种推理系统首要解决的问题都是遇到冲突时如何处理。

(1) 对于认识推理，信息冲突的原因主要在于知识的不完全性和不确定性；

(2) 对于实践推理，可能的冲突还来自于 Agent 的内部动机与外部动机(如由策略和规

范等带来的义务)之间的不一致性;

(3) 对于多 Agent 交互中的推理,冲突主要表现为不同 Agent 存在各自的利益、偏好、主张等的不一致。

第二,结论的可废止性与非单调性。说服性的论证与逻辑证明不同。前者是可废止的:当 Agent 知道了前面未知的信息后,先前得出的结论可以被推翻;当出现了更加重要的目标时,已经形成的目标可能被修改;当辩论或协商的对手提出更强的反面论证时,相关主张的可辩护性可以发生转化。推理结果的这种可废止性是不可避免的:论证可以被停止挑战,并且被接受,但受挑战的可能性仍然存在。

第三,系统的动态性。在上述各种应用中,Agent 置身于开放的环境,新信息的加入、推理知识的改变、各个 Agent 的观点和偏好的变化等均会引起系统状态的变化。这种变化在很多情况下是局部的。然而,由于推理的结果具有非单调性的特点,当新信息加入时,系统中已经形成的所有结论都有被推翻的可能。因此,如何局部地计算系统的动态性是一个非常重要的挑战性问题。

第四,推理结果的主观性。在真实生活的领域,Agent 的推理结果不仅与它们自己的偏好、态度和价值观等有关,而且有时还与听众或"法官"有关。例如,对于谨慎的 Agent,在证据不确凿的情况下,它不会接受相应的论证;而对于轻信的 Agent,它会依据不可靠的证据作出判断。在上述《两小儿辩日》的例子中,一个谨慎的 Agent 不会接受任一小孩的观点,而一个轻信的 Agent 可能会选择接受其中一个小孩的观点。

1.3 论辩系统产生的理论背景

1.3.1 经典一阶逻辑的缺陷

经典一阶逻辑具有如下特点:

(1)前提可以按照封闭的概念明确定义,且是一致的。在经典一阶逻辑中,如果系统是不一致的,那么由该系统可以推出任何结论。假定 q 和 $\neg q$ 都是系统的定理,那么对于任何 p,$q \wedge \neg q \rightarrow p$ 都是重言式。换句话说,在经典的一阶逻辑系统中,不允许出现系统不一致的情况。

(2)推理和分析是在一个封闭的、精确定义的上下文中进行,即不存在不完全的或不确定的信息。

(3)结论是最后的且确定的。经典一阶逻辑主要关心普遍"真"的形式化,它们没有异常情况,在所有情况下总是成立的。经典一阶逻辑是单调的,这意味着当新的公理加入时,一组公理集合的任何逻辑后承仍然是逻辑后承。

(4)推理和结论是完全客观的。

上述要求在日常的推理中往往难以达到。

首先,在 Agent 认识推理方面,由于置身于现实世界中的 Agent(自然 Agent 或人工 Agent)认知能力的局限性,它所获取的信息经常是不完全的和不一致的。随着 Agent 认识的深入,它的信念应该是可修改的。在例 1.1 中,若我们用 p 表示"Tweety 是一只鸟",用 q 表示"Tweety 会飞",用 $p \rightarrow q$ 表示"因为 Tweety 是一只鸟,所以 Tweety 会飞",用 t 表示

"Tweety是一只企鹅",用 $t\rightarrow\neg q$ 表示"因为Tweety是一只企鹅,所以Tweety不会飞"。我们有下列两个公式集合:

$$\Gamma=\{(p\rightarrow q),\ p\},\ B=\{t,\ (t\rightarrow\neg q)\}$$

因此 $\Gamma\vdash q$,但 $\Gamma\cup B\vdash\perp$,即 $\Gamma\cup B\nvdash q$。

上述推理的关键问题在于公式"$p\rightarrow q$"不是"鸟会飞"的准确表达。事实上"鸟会飞"这个知识存在不确定性,即它的含义应该是"在一般的情况下,鸟会飞",但允许存在例外情况,如企鹅、鸵鸟等不会飞。

依据经典一阶逻辑,一种貌似可行的方法来表达不确定知识"在一般情况下,鸟会飞"是在公式中考虑异常情况:

$$\forall x(bird(x)\wedge\neg exception(x)\rightarrow fly(x))$$
$$\forall x(exception(x)\leftrightarrow penguin(x)\vee ostrich(x)\vee canary(x)\vee\cdots)$$

为了证明"Tweety不是异常的",我们要证明"$\neg penguin$(Tweety)"、"$\neg ostrich$(Tweety)"等。然而,问题是我们无法预知所有的异常情况。换句话说,在实际的推理中,规则不是总能刚好标记所有的可能异常;相反,人们经常在缺少反面证据的情况下被迫应用"经验规则"或"缺省规则"。

其次,演绎逻辑不适合于实践推理。重新考虑与例1.2相关的例子:如果一个人想要去伦敦,那么他必须寻找一种途径来实现它,比如赶一辆去伦敦的火车[3]。然而,对于想要去伦敦这件事,有许多其他的途径来实现,比如赶一辆飞机、走路、坐船,等等。这些不同的途径之间存在冲突。

总之,经典一阶逻辑中这种无主体、独白式、纯粹的逻辑运算,难以刻画现实世界中的Agent的各种推理。

1.3.2 传统非单调逻辑及其不足

为了解决经典一阶逻辑不能处理不完全的、不确定的、不一致的信息,20个世纪80年代以来,出现了许多"非单调逻辑",包括缺省逻辑[4]、限定推理[5]、自认识逻辑[6]等。

大多数非单调逻辑把缺省的概念或可废止条件式当作基本的"非标准"单元。在这里,条件式可以用短语"典型地"、"正常地"或"除非表明例外"等来刻画。缺省不能保证它们的结论在前提成立时总是成立的。相反,它们允许我们在这种情况下,可废止地推出它们的结论,即如果不知异常情况,就可以推出试探性的结论。下面介绍几种主要的非单调方法。

优先蕴含(Preferential Entailment)

优先蕴含是一种基于标准一阶逻辑的模型理论方法[7,8],其基本思想是:为了判断某个结论是否成立,不是通过检查前提的所有模型,而是只检查一些模型,即对缺省来说异常尽可能少的模型。为了实现这个目的,通常的方法是给缺省提供一个特殊的"正常条件",如:

(1) $\forall x.\ bird(x)\wedge\neg ab(x)\supset canfly(x)$

这个公式可以读作"鸟会飞,除非它们关于飞行是异常的"。现在我们假定:

(2) $bird$(Tweety)

我们希望从(1)和(2)推出 $canfly$(Tweety),因为没有理由相信 ab(Tweety)。这个推理是通过仅查看那些 ab 谓词的外延为最小的(关于集合包含关系)模型来实现的。因为依据(1)和(2),不能得知 Tweety 是否是一只异常的鸟,所以存在两个关于这些前提的一阶逻辑模型:Tweety 在一个模型里 ab(Tweety)满足,在另一个模型里 ab(Tweety)不满足。因此,这种方法的思想是我们可以放弃满足 ab(Tweety)的模型,而仅仅查看满足 $\neg ab$(Tweety)的模型。显然,在所有这些模型中,$canfly$(Tweety)成立。

这个推理的可废止性可以通过添加 ab(Tweety)到前提中来说明。这时,所有前提的模型满足 ab(Tweety)。因此现在优先模型是那些满足如下条件的模型:在这些模型中,ab 的外延是{Tweety}。这些模型中有些满足 $canfly$(Tweety),而其他的满足 $\neg canfly$(Tweety)。因此,我们不能再得出结论 $canfly$(Tweety)。

用于缺省的内涵语义(Intensional Semantics)

用于缺省的内涵逻辑把缺省在一个可能世界语义中解释。在这种方法中,限制的不是模型集合,而是可能世界集合。该方法的灵感来自于条件逻辑中反事实的相似性语义。在这些逻辑中,一个反事实的条件式被解释为:$\varphi \Rightarrow \psi$ 为真,仅当 ψ 在 φ 为真的可能世界的一个子集中为真,即假定在与实际世界尽可能相似的可能世界中,φ 成立。现在,关于可废止条件式,该方法的思想是以相同的方式为可废止条件式定义一个可能世界语义。一个可废止条件式 $\varphi \Rightarrow \psi$ 可以被粗略地解释为"在所有最正常的世界中若 φ 成立,那么 ψ 也成立"。显然,如果以这种方式解读,那么肯定前件式对于这种条件式不是有效的,因为即使 φ 在实际世界中成立,实际世界也不一定是一个正常的世界。这与反事实条件式不同。在反事实条件式中,实际世界总是与自己最相似的一个。因此,用可废止条件式进行推理包含一个第二阶段。它在反事实推理中不出现,同时它与在优先蕴含中选择优先模型相似:为了从可废止条件式推导出缺省结论,给定前提时,实际世界被假定尽可能正常。这个近似于在优先蕴含中,主要关注前提的"最正常的"模型。它意味着从可废止条件式推出的缺省结论也是可废止的。

一致性和不可证明性方法

另一种方法使得一致性或不可证明性的表示成为可能。在缺省逻辑中,用如下结构来扩展一阶逻辑:该结构技术上起到了推理规则的作用,但表达了领域相关的泛化,而不是逻辑推理原则。在缺省逻辑中,Tweety 缺省可以表示如下:

$$bird(x) : canfly(x)/canfly(x)$$

这个缺省的中间部分可以用于表达一致性假设。它可以被读作"如果 Tweety 是一只鸟是可证明的,而 Tweety 不会飞是不可证明的,那么我们可以推断 Tweety 会飞"。为了看清这是如何工作的,假设我们有一个一阶理论:

$$W = \{bird(\text{Tweety}),\ \forall x.\ penguin(x) \supset \neg canfly(x)\}$$

以及上面的缺省规则。那么,由于 $canfly$(Tweety)与我们所知道的一致,我们可以把该缺省应用到 Tweety,从 W 可废止地推出 $canfly$(Tweety)。这个推理是可废止的,因为:如果我们把 $penguin$(Tweety)加入到 W 中,那么 $\neg canfly$(Tweety)被经典地推出,应用缺

省规则的一致性检查失败，因此不能从 $W \cup \{penguin(\text{Tweety})\}$ 推出 $canfly(\text{Tweety})$。

由上述分析可知，传统非单调逻辑可以在一定程度上解决经典一阶逻辑的问题。不过，它们一般只针对知识或信念。但不一致也经常出现于其他场合，包括实践推理以及 Agent 间交互时的推理等。

1.4 论辩系统的基本概念和主要特点

这一节我们简要介绍论辩系统的基本概念和主要特点。

论辩系统是一种可以在不一致的情境中进行非单调推理的形式体系。它通过产生和评估支持和反对特定主张的论证，以测试该主张是否是站得住脚的。

例如，对于“Tweety 会飞”这个主张，它有一个支持该主张的由底层知识构造而来的论证——“因为 Tweety 是鸟，所以它会飞”和一个反对该主张的论证——“因为 Tweety 是一只企鹅，所以它不会飞”，后一个论证具有较高的优先级，因此可击败前一个论证。依据论证之间的攻击关系，经过评估，可知前一个论证不成立，并进而得到“Tweety 会飞”这个主张是站不住脚的。

依据上述思路，在一个论辩系统中，首先需要有一组用于构造论证的推理知识和观察信息集合，它们由特定的逻辑语言来表达。推理知识经常是带变量的，它在运行中被实例化。一组实例化的推理知识集合和观察信息（事实）集合构成系统的一个“可废止理论”（Defeasible Theory）。依据可废止理论，论辩系统通过构造论证、比较论证（确定论证之间的冲突关系及其优越关系）和评估论证（决定论证的状态：可接受的、被拒绝的或未确定的）来实现推理，并通过识别结论，最终得到推理结果（图 1.1）。接下来，我们简要介绍论辩系统中一些基本概念。

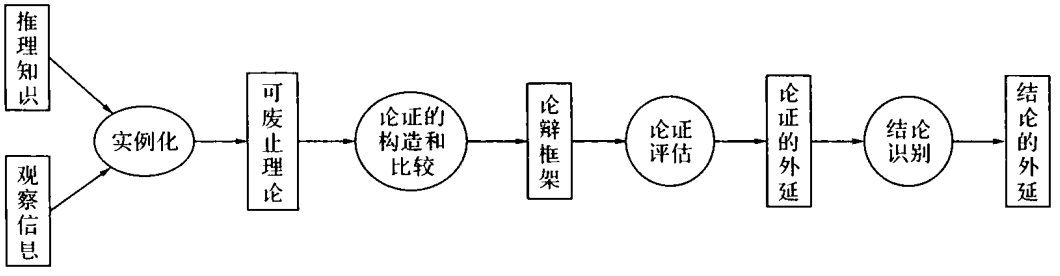

图 1.1 论辩系统的基本工作机制

1.4.1 论辩系统的基本概念

(1)可废止理论

从人工智能的角度看，我们可以把一个论辩系统当作是一个非单调推理形式体系。它的输入是一组不完全的、不确定的、不一致的推理知识集合和一组观察信息集合。在没有实例化之前，推理知识可能包含变量。这些推理知识可以是固定不变的，也可能是动态变化的。例如，在一个分布式系统中，系统管理者通过策略来管理各种服务的授权。在某个时

刻,系统管理者制订策略"对于内部用户,允许访问 A 类服务"。后来,由于用户的大量增加,对 A 类服务的访问出现拥堵现象。为了解决该问题,系统管理者制订另一条策略"在用户高峰期,只有金卡用户允许访问 A 类服务"。对于每一个用户,上述知识被实例化,并在随后被用于推理决策。我们把一组实例化的不确定的知识集合称为系统的一个可废止理论。

例 1.4 依据例 1.3,我们可以得到如下知识:

A. 太阳刚升起来时离人近,因为它看上去大得像一个车盖。

B. 太阳到中午时离人远,因为它看上去小得像一个盘盂。

C. 太阳刚升起来时离人远,因为它让人感觉清凉。

D. 太阳到中午时离人近,因为它让人感觉像把手伸进热水里一样热。

若用 p_1 表示命题"太阳刚升起时看上去大",p_2 表示"太阳到中午时看上去小",p_3 表示"太阳刚升起时让人感觉凉",p_4 表示"太阳到中午时让人感觉热",p_5 表示"太阳刚升起时离人近",p_6 表示"太阳到中午时离人远",上述知识可以形式化地表示为一个可废止理论,记作 $D_1=\{p_1, p_2, p_3, p_4, p_1\Rightarrow p_5, p_2\Rightarrow p_6, p_3\Rightarrow\neg p_5, p_4\Rightarrow\neg p_6\}$(其中,"$\Rightarrow$"用于指示一条可废止规则,详见定义 2.3)。

(2)论证

依据给定的可废止理论,论辩系统通过构造论证、比较论证和评估论证来实现推理。在这里,论证可以被看作是一组支持某个结论的知识集合。依据不同的抽象级别,我们可以粗略地把论证分为如下三种:

第一种论证具有详细的结构,即包含了由特定语言表示的知识和推导关系。目前已有许多阐述论证结构的方法,包括列表[9]、树[10]或假设集合[11]。

第二种论证表示为一个"前提—结论"对,隐去了底层逻辑以及如何证实从前提到结论的证明。

第三种论证最为抽象,其内部结构完全未指定。这种论证最初由 Dung 提出。他把论证的概念当作一个原子概念,仅仅关注论证之间交互的方式。

例 1.5 依据例 1.4,我们构造一组论证集合$\{\alpha, \beta, \gamma, \eta\}$,其中每个论证都是一棵证明树(Proof Tree)[10](图 1.2)。对于这些论证,如果我们在评估它们的状态时不关心它们的内部结构,那么我们可以把它们看作是原子概念(详见第 3 章"论辩系统的语义描述")。

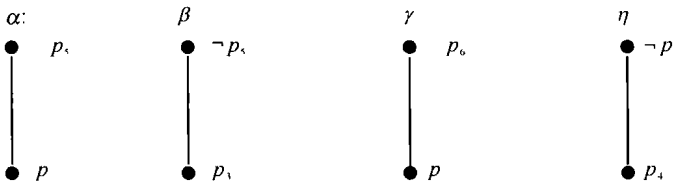

图 1.2 以证明树的形式建构的论证

(3)论证之间的攻击关系

论证之间的攻击关系可以分为三类:一个论证攻击另一个论证的前提(即假设),一个论证攻击另一个论证的结论,一个论证攻击另一个论证的推理关系。一般来说,所有论证的前