第5章

不确定知识表示和推理

现有的知识表示和推理技术往往把研究和处理对象限定在特定的专业知识领域,对不具有规范表示的知识领域,现有技术的适用性就大为降低。近年来,由于实际应用的推动,知识处理研究发生了许多重大变化,已经从注重研究知识的形式转向研究知识的内容,从注重研究良构知识(well-structured knowledge)转向研究病构知识(ill-structured knowledge),从注重研究封闭性知识转向研究开放性知识,从研究内涵完整、协调和精确的知识转向研究内涵不完整、不协调和不精确的知识,这些趋势可以用"非规范知识处理"的概念概括。所谓知识的非规范性,是指知识内涵的难处理性,包括知识的不确定性(模糊知识、不确定、随机和不精确知识),或知识的不完整性(内容不完整的知识和结构不完整的知识),或知识的不协调性(含矛盾的知识、带噪声的知识和含冗余的知识),或知识非恒常性(时变知识和启发式知识)。

本章讨论处理数据的不精确和知识的不确定所需要的一些工具和方法,主要包括基于Bayes理论的概率推理、在经验基础上抽象得到的确定性因子方法、基于信任测度函数的证据理论、基于模糊集合论的模糊推理等技术。

5.1 概　述

诸如"鸟是会飞的"及"常在河边走,哪能不湿鞋"这样的常识和常识推理,如何形式化?

这里说的常识、常识推理与通常的逻辑推理不同。首先,常识具有不确定性。一个常识可能有众多的例外。一个常识可能是一种尚无理论依据或者缺乏充分验证的经验。其次,常识往往对环境有极强的依存性。由于常识的这种不确定性,决定了常识推理的所谓非单调性,即依据常识进行通常的逻辑推理,但保留对常识的不确定性及环境的变迁造成的推理失误的修正权。非单调推理技术试图解决不确定性推理问题。

既然人的信念常常是不确定的,就存在关于信念强度的问题,即确定性程度到底为多少。最常见的方法是把指示确定性程度的数据附加到推理规则,并由此研究不确定强度的表示和计算问题。

陆汝钤院士曾主持一项国家自然科学基金重大项目"非规范知识处理的基本理论和核心技术"。所谓知识的非规范性,是指知识内涵的难处理性,它包括知识的不确定性、知识的不完整性、知识的不协调性和知识的非恒常性。该项目的主要研究目标和研究内容是,从理论、技术和示范应用三个层面对非规范知识处理进行深入研究。在理论上,要研究非规范知识的数学理论、逻辑理论和认知理论;在技术上,要研究非规范知识的表示和建模、非规范知识的获取和融合,以及非规范知识的通信和传播;在示范应用上,要研究几个特定领域的非规范知识,开发海量非规范知识库、示范性语义网上知识获取和知识编辑器,以及通用网上的知识获取和知识编辑器。

5.1.1 什么是不确定性推理

不确定性是智能问题的本质特征，无论是人类智能还是人工智能，都离不开不确定性的处理。可以说，智能主要反映在求解不确定性问题的能力上。

推理是人类的思维过程，它是从已知事实出发，通过运用相关的知识逐步推出某个结论的过程。其中，已知事实和知识是构成推理的两个基本要素。已知事实又称为证据，用以指出推理的出发点及推理时应使用的知识；而知识是推理得以向前推进，并逐步达到最终目标的依据。第 2 章介绍的演绎推理是一种精确的推理，因为它处理的是精确事实和知识，并运用确定的推理方法得出精确的结论。

在客观世界中，由于事物发展的随机性和复杂性，人类认识的不完全、不可靠、不精确和不一致性，自然语言中存在的模糊性和歧义性，使得现实世界中的事物以及事物之间的关系极其复杂，带来了大量的不确定性。如果采用确定性的经典逻辑处理不确定性，就需要把知识或思维行为中原本具有的不确定性划归为确定性处理，这无疑会失去事物的某些重要属性，造成信息流失，妨碍人们做出最好的决定，甚至可能做出错误的决定。大多数要求智能行为的任务都具有某种程度的不确定性。不确定性可以理解为在缺少足够信息的情况下做出判断。

确定性推理是建立在经典逻辑基础上的，经典逻辑的基础之一就是集合论，集合论中的隶属概念是一个非常精确和确定的概念，一个元素是否属于某个集合是非常明确的。这在很多实际情况中是很难做到的，如高、矮、胖、瘦就很难精确地区分。因此，经典逻辑不适合用来处理不确定性。针对不同的不确定性起因，人们提出了不同的理论和方法，以建立适合描述不确定和不精确的新的逻辑模型。因此可以说，不确定性推理是建立在非经典逻辑基础上的一种推理，它是对不确定性知识的运用与处理。严格地说，不确定性推理就是从不确定性初始证据出发，通过运用不确定性的知识，最终推出具有一定程度的不确定性，但却是合理或者近乎合理的结论的思维过程。

5.1.2 不确定性推理要解决的基本问题

证据和规则的不确定性，导致所产生的结论的不确定性。不确定性推理反映了知识不确定性的动态积累和传播过程，推理的每一步都需要综合证据和规则的不确定性因素，通过某种不确定性测度，寻找尽可能符合客观实际的计算模式，通过不确定性测度的传递计算，最终得到结果的不确定性测度。

因此，在基于规则的专家系统中，不确定性表现在证据、规则和推理三方面，需要对专家系统中的事实与规则给出不确定性描述，并在此基础上建立不确定性的传递计算方法。因此，要实现对不确定性知识的处理，必须解决不确定性知识的表示问题、不确定性信息的计算问题，以及不确定性表示和计算的语义解释问题。

1. 表示问题

表示问题指的是采用什么方法描述不确定性。通常有数值表示和非数值的语义表示两种方法。数值表示便于计算、比较；非数值表示是一种定性的描述。

专家系统中的“不确定性”一般分为两类：一是规则的不确定性，二是证据的不确定性。

(1) 规则的不确定性$(E\rightarrow H, f(H,E))$，表示相应知识的不确定性程度，称为知识或规则强度。

(2) 证据的不确定性$(E, C(E))$，表示证据 E 为真的程度。它有两种来源：初始证据(由用户给出)；前面推出的结论作为当前证据(通过计算得到)。

一般来说,证据不确定性的表示方法应与知识不确定性的表示方法保持一致,证据的不确定性通常也是一个数值表示,它代表相应证据的不确定性程度,称为动态强度。

2. 计算问题

计算问题主要指不确定性的传播与更新,即获得新信息的过程。它是在领域专家给出的规则强度和用户给出的原始证据的不确定性的基础上,定义一组函数,求出结论的不确定性度量。它主要包括如下三方面。

1) 不确定性的传递算法

在每一步推理中,如何把证据及规则的不确定性传递给结论?在多步推理中,如何把初始证据的不确定性传递给结论?

也就是说,已知规则的前提 E 的不确定性 $C(E)$ 和规则强度 $f(H,E)$,求假设 H 的不确定性 $C(H)$,即定义函数 f_1,使得

$$C(H)=f_1(C(E),f(H,E))$$

2) 结论不确定性合成

推理中有时会出现这样一种情况:用不同的知识进行推理,得到相同结论,但不确定性的程度却不相同。

即已知由两个独立的证据 E_1 和 E_2 求得的假设 H 的不确定性度量 $C_1(H)$ 和 $C_2(H)$,求证据 E_1 和 E_2 的组合导致的假设 H 的不确定性 $C(H)$,即定义函数 f_2,使得

$$C(H)=f_2(C_1(E),C_2(H))$$

3) 组合证据的不确定性算法

即已知证据 E_1 和 E_2 的不确定性度量 $C(E_1)$ 和 $C(E_2)$,求证据 E_1 和 E_2 的析取和合取的不确定性,即定义函数 f_3 和 f_4,使得

$$C(E_1 \land E_2)=f_3(C(E_1),C(E_2))$$
$$C(E_1 \lor E_2)=f_4(C(E_1),C(E_2))$$

目前关于组合证据的不确定性的计算已经提出了多种方法,用得最多的是如下三种。

(1) 最大最小法。

$$C(E_1 \land E_2)=\min(C(E_1),C(E_2))$$
$$C(E_1 \lor E_2)=\max(C(E_1),C(E_2))$$

(2) 概率方法。

$$C(E_1 \land E_2)=C(E_1)\times C(E_2)$$
$$C(E_1 \lor E_2)=C(E_1)+C(E_2)-C(E_1)\times C(E_2)$$

(3) 有界方法。

$$C(E_1 \land E_2)=\max\{0,C(E_1)+C(E_2)-1\}$$
$$C(E_1 \lor E_2)=\min\{1,C(E_1)+C(E_2)\}$$

3. 语义问题

语义问题指上述表示和计算的含义是什么,如 $C(H,E)$ 可理解为当前提 E 为真时,对结论 H 为真的一种影响程度,$C(E)$ 可理解为 E 为真的程度。

目前,在人工智能领域,处理不确定性问题的主要数学工具有概率论和模糊数学。概率论与模糊数学研究和处理的是两种不同的不确定性。概率论研究和处理随机现象,事件本身有明确的含义,只是由于条件不充分,使得在条件和事件之间不能出现决定性的因果关系(随机性)。模糊数学研究和处理模糊现象,概念本身就没有明确的外延,一个对象是否符合这个概念是难以确定的(属于模糊的)。无论采用什么数学工具和模型,都需要对规则和证据的不确定性给出度量。

规则的不确定性度量 $f(H,E)$需要定义在下述三个典型情况下的取值：

(1) 若 E 为真，则 H 为真，这时 $f(H,E)$的值。

(2) 若 E 为真，则 H 为假，这时 $f(H,E)$的值。

(3) E 对 H 没有影响，这时 $f(H,E)$的值。

对于证据的不确定性度量 $C(E)$，需要定义在下述三个典型情况下的取值：

(1) E 为真，$C(E)$的值。

(2) E 为假，$C(E)$的值。

(3) 对 E 一无所知，$C(E)$的值。

对于一个专家系统，一旦给定了上述不确定性的表示、计算及其相关的解释，就可以从最初的观察证据出发，得出相应结论的不确定性程度。专家系统的不确定性推理模型指的就是证据和规则的不确定性的测度方法以及不确定性的组合计算模式。

5.1.3 不确定性推理方法分类

关于不确定性推理方法的研究，主要沿两条不同的路线发展。

(1) 在推理级扩展不确定性推理的方法：其特点是把不确定证据和不确定的知识分别与某种量度标准对应起来，并且给出更新结论不确定性算法，从而建立不确定性推理模式。通常把这一类方法统称为模型方法。

(2) 在控制策略级处理不确定性的方法：其特点是通过识别领域中引起不确定性的某些特征及相应的控制策略限制或减少不确定性对系统产生的影响，这类方法没有处理不确定性的统一模型，其效果极大地依赖控制策略，把这类方法统称为控制方法。

模型方法又分为数值方法及非数值方法两类。数值方法是对不确定性的一种定量表示和处理方法。非数值方法是指除数值方法外的其他各种处理不确定性的方法，如古典逻辑方法和非单调推理方法等。

在数值方法中，概率方法是重要的方法之一。概率论有着完善的理论和方法，而且具有现成的公式实现不确定性的合成和传递，因此可以用作度量不确定性的重要手段。

纯概率方法虽然有严格的理论依据，但通常要求给出事件的先验概率和条件概率，而这些数据又不易获得，因此使其应用受到限制。为了解决这个问题，人们在概率论的基础上发展起一些新的方法和理论，主要有主观概率论(又称主观 Bayes 方法)、可信度方法、证据理论等。

(1) 主观 Bayes 方法：它是 PROSPECTOR 专家系统中使用的不确定性推理模型，是对 Bayes 公式修正后形成的一种不确定性推理方法，为概率论在不确定性推理中的应用提供了一条途径。

(2) 可信度方法：它是 MYCIN 专家系统中使用的不确定性推理模型，它以确定性理论为基础，方法简单、易用。

(3) 证据理论：它通过定义信任函数、似然函数，把知道和不知道区别开。这些函数满足比概率函数的公理弱的公理，因此，概率函数是信任函数的一个子集。

基于概率的方法虽然可以表示和处理现实世界中存在的某些不确定性，在人工智能的不确定性推理方面占有重要地位，但它们没有把事物自身具有的模糊性反映出来，也不能对其客观存在的模糊性进行有效推理。Zadeh 等提出的模糊集理论及其在此基础上发展的可能性理论弥补了这一缺憾。概率论处理的是由随机性引起的不确定性，可能性理论处理的是由模糊性引起的不确定性。可能性理论对由模糊性引起的不确定性的表示及处理开辟了一种新的解决途径，并得到广泛的应用。

5.2 非单调逻辑

为了形式地表述常识，并在常识间进行有效的形式推理，20 世纪 70 年代末，人们提出了非单调逻辑。

传统逻辑系统都是单调的，因为由已知事实推出的逻辑结论绝不会在已知事实增加时反而丧失。更形式地，可定义逻辑系统的单调性如下。

定义 5.1 设 FS 为一逻辑系统，称 FS 是单调的，如果对于 FS 的任意公式集合 $\Gamma_1, \Gamma_2, \Gamma_1 \subseteq \Gamma_2$ 蕴涵 $\mathrm{Th}(\Gamma_1) \subseteq \mathrm{Th}(\Gamma_2)$。这里 $\mathrm{Th}(\Gamma)$表示公式集合$\{A \mid \Gamma \vdash_{\mathrm{FS}} A\}$，即 Γ 的演绎结果的集合。

已讨论的所有逻辑系统都是单调的。可是，常识推理却并不具有这种单调性。例如，当你告诉我“a 是一只鸟”时，我立即会根据常识“鸟是会飞的”进行推理，做出结论“a 是会飞的”。可当你又告诉我“a 是一只鸵鸟”时，我自然会立即撤回上述结论，相反会根据常识“鸵鸟不会飞”而做出结论“a 是不会飞的”。如果我足够机敏，还应对常识“鸟是会飞的”做出修正，例如，改为“鸟是会飞的，除非它是鸵鸟”。在上述推理过程中，第一个结论在已知事实增加时会自行撤销(而不是仍然接受它)，并修改推理的依据(而不是让互相矛盾的依据共存，因而被迫接受一切断言)。

常识推理的这种特性称为非单调性，具有非单调性的推理称为非单调推理，而使用非单调推理的逻辑系统称为非单调逻辑。和定义 5.1 相对，可形式地定义非单调性。

定义 5.2 逻辑系统 FS 称为非单调的，如果存在公式集合 Γ_1 和 $\Gamma_2, \Gamma_1 \subseteq \Gamma_2$。但 $\mathrm{Th}(\Gamma_1) \not\subseteq \mathrm{Th}(\Gamma_2)$。

要使机器具有智能，就应当使它具有进行常识推理的能力，具有依据“不完全的信息”和“不可靠的经验”进行推理及预测的能力，因此使机器具有这种非单调的逻辑推理机制是非常必要的。

5.2.1 非单调逻辑的产生

非单调逻辑这个名词的第一次出现，大致是在 20 世纪 70 年代中期，但在人工智能对推理机制的模拟的研究中，非单调推理的运用则更早些。

最早的 Prolog 版本中就已经有了“封闭系统假设”，即当系统推不出 A 时，便认为 $\neg A$ 成立。当系统的知识库扩充时，可能推出 A，那时 $\neg A$ 便不再为系统所接受。

PLANNER 系统则更进一步，其中设有运算 THNOT。THNOT(A)表示“试图证明 A，若不成功，则 THNOT(A)为真”。不仅如此，为了便于在运行中更新系统，PLANNER 还设有前提表和删除表，可随时删除那些系统已经导出而又在系统更改后不再成立的事实。

采用“封闭系统假设”或算符 THNOT 的方式都有一个明显的缺点，即必须保证“A 是否可证”是可判定的，而这并不总是可以办到的。一阶逻辑是不可判定的。此外，系统还可能遇到“循环论证”的情况。例如，系统已知

$$A(f(x)) \rightarrow B(x)$$
$$B(f(x)) \rightarrow C(x)$$
$$C(f(x)) \rightarrow A(x)$$

要证 $A(a)$。这时系统既无法确定“$A(a)$可证”，也无法确定“$A(a)$不可证”，因为要证 $A(a)$，须证 $C(f(a)), B(f(f(a))), A(f(f(f(a)))), \cdots$。

在用逻辑演算刻画状态转换、动作规划时，非单调性显得尤为重要，因为状态、动作都不是一

成不变的。

在规划生成系统 STRIPS 中，用状态变换的规则模拟机器人的动作。这些规则均由以下三部分组成。

（1）前提：规则执行的前提。

（2）删除表：规则执行后状态描述中应当删除的事实表。

（3）添加表：规则执行后状态描述中应当添加的事实表。

例 5.1　表示机器人拾起一块积木的动作可用规则 Pickup(x)，它由以下三部分组成。

前提：ontable(x)　　（x 在桌子上）

clear(x)　　（x 上无他物）

handempty　　（机械手闲置）

删除表：ontable(x)，clear(x)，handempty

添加表：holding(x)（机械手持有 x）

如果图 5-1(a)的状态描述是

$$\{\text{ontable}(A),\text{ontable}(B),\text{handempty},\text{clear}(A),\text{clear}(B)\}$$

那么，经过动作 Pickup(A)后，其状态描述应为

$$\{\text{ontable}(B),\text{clear}(B),\text{holding}(A)\}$$

如图 5-1(b)所示。

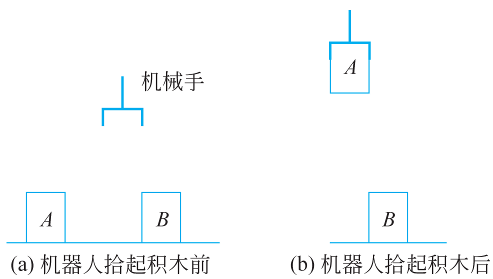

图 5-1　机器人拾起一块积木的动作前后

在实际应用中出现的这些处理非单调性的方法是很有启发意义的，它们为非单调逻辑的出现奠定了基础。到了 20 世纪 70 年代后期和 80 年代初，人们开始研究非单调推理，并提出多种非单调逻辑系统，较令人注目的是 Reiter 的默认推理逻辑，以及 McDermott 和 Doyle 的非单调逻辑系统。

5.2.2　非单调规则

在非单调规则系统中，即使所有前提已知，规则也可能不会被应用，因为还必须考虑是否同时存在与之相矛盾的推理。一般来说，下面考虑的规则都称为可废止的，因为其他规则可以废止它们。为了允许规则间的冲突，否定的原子公式可以出现在规则的头或体中。例如，可以有下面的规则：

$$p(X)\rightarrow q(X)$$
$$r(X)\rightarrow \neg q(X)$$

下面使用不同的箭头区别可废止规则和标准的单调规则：

$$p(X)=>q(X)$$
$$r(X)=>>q(X)$$

在这个例子中,如果同时还给出事实 $p(a)$、$r(a)$,则根据非单调规则,既推不出 $q(a)$,也推不出 $\neg q(a)$。这是一个两条规则彼此阻塞的典型例子。这种冲突可以通过使用规则间优先序解决。假设我们知道由于某些原因,第一条规则比第二条规则可靠,那么可以确定地推出 $q(a)$。

在实践中,优先序的出现是很自然的,可以基于各种不同的原则,例如:

(1) 一条规则的来源可能比另一条规则的来源更可靠或更权威。例如,在法律领域,联邦法就优先于州立法。同样,在商业经营中,高层管理部门比中层管理部门更权威。

(2) 一条规则可能比另一条规则更优先,因为它在时间上更近。

(3) 一条规则可能比另一条规则更优先,因为它更特殊。典型的例子是一条普遍的规则带有一些对例外情况的特殊规定;在出现这些例外情况时,特殊规则比一般规则本身更应当被遵守。

对于给定的一组规则,特殊性通常可以根据这些规则计算出来,但第一条和第二条原则无法由逻辑推理定义。所以,我们对具体的优先原则加以抽象,假定存在规则集上的一种外在优先关系,用它统一地刻画各种具体的优先原则。为了从语法上表达这种关系,拓展规则语法以包含一个唯一的标号,例如:

$$r_1: p(X) => q(X)$$
$$r_2: r(X) => \neg q(X)$$

于是可以用 $r_1 > r_2$ 表示 r_1 比 r_2 更优先。

这里不对>施加很多条件,甚至不要求它是规则间的全序关系。我们仅要求优先关系是无环的,也就是不能有如下形式的环 $r_1 > r_2 > \cdots > r_n > r_1$。注意,优先关系是为了解决竞争规则间的冲突而引进的。在简单情况下,仅当一条规则的头是另一条规则的头的否定时,这两条规则才出现竞争。但实际应用中未必总是如此,常见的情况是,当某个谓词 p 被推出时,不再允许另一些谓词成立。例如,一个投资顾问可能将他的建议建立在投资者可以接受的三种风险级别上:低、中等和高。显然,每个投资者在任何给定时刻只能选择一种风险级别。技术上,可以通过给每个文字 L 维护一个冲突集 $C(L)$ 刻画这种情形。$C(L)$ 总是含有 L 的否定,也可以包含更多文字。

定义 5.3 可废止规则有如下形式:

$$r: L_1, L_2, \cdots, L_n => L$$

其中,r 是标号,$\{L_1, L_2, \cdots, L_n\}$ 是体(或前提),L 是规则的头。L、$L_1, L_2, \cdots, L_n$ 是正或负文字(一个文字是一个原子公式 $p(t_1, t_2, \cdots, t_m)$ 或它的否定 $\neg p(t_1, t_2, \cdots, t_m)$)。在规则中没有函数词出现。有时用 head($r$)表示规则的头,body($r$)表示体。有时用标号 r 指代整个规则,虽然这有些不严格。

可废止逻辑程序是一个三元组(F, R, >),包括事实集 F,可废止规则的有限集 R,以及 R 上的无环二元关系>(严格地说,是 $r > r'$ 的集合,其中,r、r' 是 R 中规则的标号)。

5.2.3 案例:有经纪人的交易

本例说明在电子商务领域怎样使用规则。有经纪人的交易通过独立的第三方——经纪人实现。经纪人匹配买家的需求和卖家的能力,当双方都满意时,提议进行交易。

作为一个具体应用,下面讨论公寓租赁这种常见但通常乏味耗时的活动。适当的网络服务可以相当大地减少工作量。下面给出一个潜在租赁者的需求,具体解答扫二维码。

5.2.3 案例解答

颜炯正在找一个至少 45m^2 且至少有两个卧室的公寓。如果是在三楼或三楼以上,楼必须有电梯。而且可以养宠物。

颜炯愿意为市中心的 $45m^2$ 大小的公寓付 900 元，为在市郊的类似公寓付 750 元。并且，他愿意为公寓超出 $45m^2$ 的部分每平方米支付 15 元，为花园每平方米支付 6 元。

他的付款总额不会超过 1200 元。在给定的可选项中，他将选择最便宜的，第二优先的是有花园的，最后才是有额外空间的。

5.3　主观 Bayes 方法

概率论被广泛用于处理随机性以及人类知识的不可靠性，如随机事件 A 的概率 $P(A)$ 可表示 A 发生的可能性大小，因而可用概率方法表示和处理事件 A 的确定性程度。主观 Bayes 方法是由 R.O.Duda 等于 1976 年在概率论的基础上，通过对 Bayes 公式的修正而形成的一种不确定性推理模型，并成功地应用在他们自己开发的地矿勘探专家系统 PROSPECTOR 中。

5.3.1　全概率公式和 Bayes 公式

在概率论中，一个事件或命题的概率是在大量统计数据的基础上计算出来的，并且要处理条件概率中复杂的证据之间的内在关系。在使用概率进行不确定性推理中，需要收集大量的样本事件进行统计，以便获得事件发生的概率用来表示命题的确定性程度。然而，在许多情况下，同类事件发生的频率不高，甚至很低，无法做概率统计，这时一般是根据观测到的数据，凭领域专家的经验给出一些主观上的判断，称为主观概率。因此，概率一般可以解释为对证据和规则的主观信任度。概率推理中起关键作用的就是所谓的 Bayes 公式，它也是主观 Bayes 方法的基础。

1. Bayes 公式

定义 5.4（全概率公式）　设有事件 $A_1, A_2, \cdots, A_n$ 满足：

（1）任意两个事件互不相容；

（2）$P(A_i) > 0 (i=1,2,\cdots,n)$；

（3）样本空间 D 是所有 $A_i (i=1,2,\cdots,n)$ 构成的集合。

则对任何事件 B 来说，有下式成立：

$$P(B) = P(A_1) \cdot P(B \mid A_1) + P(A_2) \cdot P(B \mid A_2) + \cdots + P(A_n) \cdot P(B \mid A_n)$$

全概率公式提供了计算 $P(B)$ 的方法。

定义 5.5（Bayes 公式）　设有事件 $A_1, A_2, \cdots, A_n$ 满足：

（1）任意两个事件互不相容；

（2）$P(A) > 0 (i=1,2,\cdots,n)$；

（3）样本空间 D 是所有 $A_i (i=1,2,\cdots,n)$ 构成的集合。

则对任何事件 B 来说，有下式成立：

$$P(B) \cdot P(A_i \mid B) = P(A_i) \cdot P(B \mid A_i) \quad (i = 1,2,\cdots,n)$$

$$P(A_i \mid B) = \frac{P(A_i) \times P(B \mid A_i)}{P(B)} \quad (i = 1,2,\cdots,n)$$

由全概率公式得到

$$P(A_i \mid B) = \frac{P(A_i) \times P(B \mid A_i)}{\sum_{j=1}^{n} P(A_j) \times P(B \mid A_j)} \quad (i = 1,2,\cdots,n)$$

其中，$P(A_i)$ 是事件 A_i 的先验概率；$P(B|A_i)$ 是在事件 A_i 发生条件下事件 B 的条件概率；

$P(A_i|B)$是在事件B发生条件下事件A_i的条件概率,称为后验概率。

2. 利用 Bayes 公式进行推理

在专家系统中,假设有如下规则:

$$\text{If} \quad E \quad \text{Then} \quad H$$

其中,E为前提条件,H为结论。那么条件概率$P(H|E)$就表示在E发生时,H的概率,可以用它作为证据E出现时结论H的确定性程度。

同样,对于复合条件$E=E_1 \wedge E_2 \wedge \cdots \wedge E_n$,也可以用条件概率$P(H|E_1,E_2,\cdots,E_n)$作为证据$E_1,E_2,\cdots,E_n$出现时,结论$H$的确定性程度。

对于产生式规则 If E Then H_i,用条件概率$P(H_i|E)$作为证据E出现时,结论H_i的确定性程度。根据 Bayes 公式,可以得到

$$P(H_i \mid E) = \frac{P(H_i) \times P(E \mid H_i)}{\sum_{j=1}^{n} P(H_j) \times P(E \mid H_j)} \quad (i = 1,2,\cdots,n)$$

这就是说,当已知结论H_i的先验概率$P(H_i)$,并且已知结论$H_i(i=1,2,\cdots,n)$成立时,前提条件E对应的证据出现的条件概率$P(E|H_i)$就可以用上式求出相应证据出现时结论H_i的条件概率$P(H_i|E)$。

当有多个证据$E_1,E_2,\cdots,E_m$和多个结论$H_1,H_2,\cdots,H_n$,并且每个证据都以一定程度支持每个结论时,根据独立事件的概率公式和全概率公式,Bayes 公式可变为

$$P(H_i \mid E_1E_2\cdots E_m) = \frac{P(E_1 \mid H_i) \times P(E_2 \mid H_i) \times \cdots \times P(E_m \mid H_i) \times P(H_i)}{\sum_{j=1}^{n} P(E_1 \mid H_j) \times P(E_2 \mid H_j) \times \cdots \times P(E_m \mid H_j) \times P(H_j)}$$
$$(i = 1,2,\cdots,n)$$

此时,只要已知H_i的先验概率$P(H_i)$以及H_i成立时证据$E_1,E_2,\cdots,E_m$出现的条件概率$P(E_1|H_i),P(E_2|H_i),\cdots,P(E_m|H_i)$,就可利用上式计算出在$E_1,E_2,\cdots,E_m$出现情况下$H_i$的条件概率$P(H_i|E_1,E_2,\cdots,E_m)$。

在实际应用中,有时这种方法是很有用的。例如,如果把$H_i(i=1,2,\cdots,n)$当作一组可能发生的疾病,把$E_j(j=1,2,\cdots,m)$当作相应的症状,$P(H_i)$是从大量实践中经统计得到的疾病H_i发生的先验概率,$P(E_j|H_i)$是疾病H_i发生时观察到症状E_j的条件概率,则当观察到患者有症状$E_1,E_2,\cdots,E_m$时,应用上述 Bayes 公式就可计算出$P(H_i|E_1,E_2,\cdots,E_m)$,从而得知患者患疾病$H_i$的可能性。

Bayes 推理的优点是它有较强的理论背景和良好的数学特性,当证据和结论都彼此独立时,计算的复杂度比较低,但是它也有其局限性。

(1) 因为需要$\sum_{j=1}^{n} P(H_j) = 1$,如果又增加一个新的假设,则对所有的$1 \leqslant j \leqslant n+1$,$P(H_j)$都需要重新定义。

(2) Bayes 公式的应用条件是很严格的,它要求各事件互相独立,如证据间存在依赖关系,就不能直接使用此方法。

(3) 在概率论中,一个事件或命题的概率是在大量统计数据的基础上计算出来的,因此尽管有时$P(E_j|H_i)$比$P(H_i|E_j)$相对容易得到,但总的来说,要想得到这些数据,仍然是一项相当困难的工作。

5.3.2 主观 Bayes 方法

主观 Bayes 方法是在对 Bayes 公式修正的基础上形成的一种不确定性推理模型。

1. 知识不确定性的表示

1）信任机率

我们知道，概率论考虑的是可重复性的事件，但是对于许多不可重复事件的概率，如医疗上的诊断和矿产的探测，每个患者或矿产的位置是不同的，这时必须扩大事件的范围，以便能够处理类似的命题。例如，一个可能的事件是：

“一个患者浑身长满了红斑点”

命题是：

“患者出麻疹”

设 A 是一个命题，条件概率为 $P(A|B)$。

如果事件或命题不可重复或没有数学依据，通常概率 $P(A|B)$ 是没有必要的。这时可以把 $P(A|B)$ 解释为在 B 成立时 A 为真的可信度。

如果 $P(A|B)=1$，则可以相信 A 为真；如果 $P(A|B)=0$，则可以相信 A 为假。而对于其他值 $0<P(A|B)<1$，则表示不能完全确定 A 是真还是假。在统计学上，一般认为假设就是依据某些证据还不能确定其真假的命题，这样可以使用条件概率表示似然性，如 $P(A|B)$ 表示在证据 B 的基础上，假设 A 的似然性。

概率适用于重复事件，而似然性适用于表示非重复事件中信任的程度。一般在专家系统中，$P(H|E)$ 表示在有证据 E 的情况下，专家对某种假设 H 为真的信任度。但是，如果事件是可重复的，则 $P(H|E)$ 就表示概率。表达这种似然性的方法可以采用赌博中的机率(ODDS)方法。

定义 5.6　机率定义如下。

在某事件 C 的前提下，A 相对于 B 的机率可以表示为：

$$\text{odds}=P(A\mid C)/P(B\mid C)$$

如果 $B=\neg A$，则有

$$\text{odds}=\frac{P(A\mid C)}{P(\neg A\mid C)}=\frac{P(A\mid C)}{1-P(A\mid C)}$$

用 P 表示 $P(A|C)$，则有

$$\text{odds}=\frac{P}{1-P}\quad 并且\quad P=\frac{\text{odds}}{1+\text{odds}}$$

即已知机率可以计算似然性，反之亦然。如果把 P 解释为证据 X 出现的可能性，而 $1-P$ 表示证据 X 不出现的可能性，可见，X 的机率等于 X 出现的可能性与 X 不出现的可能性之比。用 $P(X)$ 表示 X 出现的可能性，$O(X)$ 表示 X 的机率。显然，随着 $P(X)$ 的增大，$O(X)$ 也在增大，并且 $P(X)=0$ 时有 $O(X)=0$，$P(X)=1$ 时有 $O(X)=\infty$。这样，就可以把取值为 $[0,1]$ 的 $P(X)$ 放大到取值为 $[0,+\infty)$ 的 $O(X)$。

概率通常和演绎问题一起使用，即处理在相同的假设下，一系列不同事件 E_i 均可能发生的问题。概率本质上是正向链或演绎的，而似然性则是反向链或归纳的。虽然对概率和似然性使用同样的符号，但应用却不同，通常我们说：一种假设下的似然性，或一个事件的概率。

2）充分性和必然性

由 Bayes 公式可知：

$$P(H\mid E)=\frac{P(E\mid H)\times P(H)}{P(E)}$$

$$P(\neg H\mid E)=\frac{P(E\mid\neg H)\times P(\neg H)}{P(E)}$$