

第5章

不确定性推理



许多人工智能系统具有复杂性、不完全性、模糊性或不确定性,当采用产生式系统或专家系统结构时,要求设计者建立某种不确定性问题的代数模型及其计算和推理过程。为此,本章讨论一些常用的不确定性推理方法。

5.1 不确定性推理概述

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。因此,不确定性推理模型是人工智能的一个核心研究课题。

5.1.1 什么是不确定性推理

不确定性推理(Uncertainty Reasoning),又称不精确推理(Inexact Reasoning),是相对于确定性推理提出来的。确定性推理的过程都是按照必然的因果关系或严格的逻辑推论进行的,是从已知事实出发,通过运用相关知识逐步推出结论的思维过程。其中,获得的推理结论也是严格按照一定的规则予以肯定或否定。一般来说,确定性推理有规可循、有据可依,能够且容易形成完备算法,往往有满足唯一解的特性,实现的难度较低。但是在运动规律的作用下,精确性往往是暂时的、局部的、相对的,而不精确性才是必然的、动态的、永恒的。可见,不精确性是科学认识中的重要规律,进行不确定性推理的研究是必需的,也是进行机器智能推理的主要工具之一。

所谓不确定性推理,是指推理中所使用的前提条件、判断是不确定的或者是模糊的,因而推理所得出的结论和判断也是不精确的、不确定或模糊的。一般来说,出现不确定性推理的原因和特征有如下几个。

- ① 证据的不确定性;

- ② 规则的不确定性;
- ③ 方法的不确定性。

以上“三性”的存在决定了推理的最后结果具有不确定但却近乎合理的特性,人们把这种性质的推理及其理论和方法总称为不确定性推理。

5.1.2 知识不确定性的来源

研究不确定性推理首先要研究知识的不确定性。知识的不确定性用相应的知识表示模式与之对应,以便进行推理与计算,还需用适当的方法描述其不确定性及程度。常见的知识不确定性主要有以下几个方面。

(1) 随机性。这是一种最为常见的知识不确定性,随机性使我们的生活充满了未知的魅力,是创造性不可缺少的因素,为我们提供了种种机遇。确定性可以告诉我们事物的普遍规律,这也许是群体的统计规律,也许仅是一个相对的真理。而个体的“机遇”是一种特殊的随机性。小概率的机遇一般不会出现,一旦出现,往往就会创造奇迹。

(2) 模糊性。模糊性能够用较少的代价传送足够的信息,并能对复杂事物做出高效率的判断和处理。也就是说,模糊性有助于提高效率。1965年,扎德(L. A. Zadeh)的论文 *Fuzzy Sets* 正式创立了模糊集合理论。扎德深入分析了模糊性、近似性以及随机性,主张用模糊性作为基本的研究对象,提出了隶属度、隶属函数、模糊集合等基本概念。模糊性使我们的生活简单而有效,借助模糊性可以对复杂事物做出高效率的判断和处理。例如,医生可以根据患者的模糊症状做出正确的判断,画家不用精确测量和计算就能画出栩栩如生的风景人物等。

(3) 粗糙性。知识粗糙性是由粗糙集理论通过不可区分关系和集合包含关系定义的。粗糙集理论最早是由波兰数学家帕夫拉克(Z. Pawlak)于1982年提出的一种不确定性数据分析理论,其基本思想是在保持分类能力不变的前提下,通过知识约简剔除数据中冗余的信息,从而导出问题的正确决策或分类。这一理论为处理具有不精确和不完全信息的分类问题提供了一种新的框架。粗糙集理论具有如下特点:①从新的视角对知识进行了定义,把知识看作对论域的划分,从而认为知识是有粒度的;②认为知识的不精确性是由知识粒度太大引起的;③为处理数据(特别是带噪声、不精确或不完全数据)分类问题提供了一套严密的数学工具,使得对知识能够进行严密的分析和操作。粗糙集理论将研究对象的全体称为论域,利用等价关系将论域划分为若干互斥的等价类,作为描述论域中任意集合的基本信息粒子。其利用两个可定义的集合——上近似集合和下近似集合来近似表达空间中的任意概念,这种方法自然地模拟了人类的学习和推理过程,学习到的知识采用产生式规则表示,容易被用户理解、接受和使用,因此得到了广泛的重视。粗糙集的一个显著特点是不需要用户提供数据之外的任何先验知识,比如统计学中的概率分布和模糊集中理论中的隶属函数,所以对问题的不确定性的描述和处理比较客观。

另外,知识的不确定性还来自知识的不完备性、不协调性和非恒常性。

知识的不完备性:包括知识内容的不完整、知识结构的不完备等。内容的不完整,可能来源于获取知识时观测不充分、设备不精确;知识结构的不完备可能因为人的认知能力、获取手段的限制等,造成对解决某个特定问题的背景和结构认识不全,忽略了一些重要因素。

知识的不协调性：是指知识内在的矛盾，不协调的程度可以依次为冗余、干扰、冲突等。不协调性是知识不确定性的重要体现，人们不可能也没必要在所有场合下都试图消除知识的不协调性，追求知识的一致性，要把不协调看成知识的一种常态，允许包容、并蓄、折中、调和。

知识的非恒常性：是指知识随时间的变化而变化的特性。人类对自然、社会乃至自身的认识都是一个由不知到知、由不深刻到深刻，不断更新的过程，是一个否定之否定的过程。人类的知识是无限发展的，不可能永远停留在某个水平上。

5.1.3 不确定性推理方法分类

目前，不确定性推理方法主要分为控制法和模型法。

控制方法：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来限制或减少不确定性对系统产生的影响。控制方法没有处理不确定性的统一模型，其效果极大地依赖于控制策略，主要包括相关性指导、机缘控制、启发式搜索、随机过程控制等。

模型方法：是把不确定证据和不确定知识分别与某种度量标准对应起来，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。模型方法具体又可分为数值模型方法和非数值模型方法两类。按其依据的理论不同，数值模型方法主要包括基于概率的方法和基于模糊理论的推理方法。

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。人们又在概率论的基础上提出了一些新的理论和方法，主要有主观 Bayes 方法、可信度方法、证据理论等，从而为不确定性的传递和合成提供了许多现成的公式，是最早成功应用于不确定性推理的重要方法之一。

5.2 不确定性推理的基本问题

推理是运用知识求解问题的过程，是证据和规则相结合得出结论的过程。知识的不确定性导致了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，通过不确定测度的传递计算，最终得到结果的不确定测度。在专家系统中，不确定性表现在证据、规则和推理三个方面，需要对专家系统中的证据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，实现对不确定性的处理应解决表示问题、计算问题和语义问题。

5.2.1 表示问题

表示问题指的是采用什么方法描述不确定性，这是解决不确定性推理的关键一步，通常有数值表示方法和非数值的语义表示方法。数值表示便于计算、比较；非数值表示是一种定性的描述，以便较好地解决不确定性问题。

在专家系统中，“不确定性”一般分为规则的不确定性以及证据的不确定性。

1) 规则不确定性的表示

规则的不确定性是指用相应的规则表示模式与之对应,以便于进行推理与计算,还须用适当的方法把规则的不确定性及其程度描述表达出来。一般用 $(E \rightarrow H, f(H, E))$ 表示规则的不确定性,它表示相应规则的不确定性程度,称为规则强度。

2) 证据不确定性的表示

证据的不确定性(命题 $E, C(E)$)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(通常由计算得到)。一般来说,证据不确定性的表示方法与知识不确定性的表示方法保持一致,通常也是一个数值,代表相应证据的不确定性程度,称为动态强度。

规则和证据不确定性的程度常用可信度来表示。例如,在专家系统 MYCIN 中,采用可信度表示规则及证据的不确定性,取值范围为 $[-1, 1]$ 。当可信度大于零时,其数值越大,表示相应的规则或证据越接近于“真”;当可信度小于零时,其数值越小,表示相应的规则或证据越接近于“假”。

5.2.2 计算问题

计算问题主要指不确定性的传播和更新,即获得新信息的过程。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,计算问题定义了一组函数,求出结论的不确定性度量,主要包括如下三个方面。

1) 不确定性的传递算法。

① 在每一步推理中,如何把证据和规则的不确定性传递给结论。

② 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:

$$C(H) = f_1(C(E), f(H, E)) \quad (5.1)$$

2) 结论不确定性合成

推理中有时会出现这样的一种情况,用不同的规则进行推理得到了相同的结论,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 求得的结论 H 的不确定性 $C_1(H)$ 和 $C_2(H)$,求证据 E_1 和 E_2 的组合导致的结论 H 的不确定性 $C(H)$,定义函数 f_2 使得:

$$C(H) = f_2(C_1(H), C_2(H)) \quad (5.2)$$

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 $C(E_1)$ 和 $C(E_2)$,求证据 E_1 和 E_2 的析取和合取的不确定性,定义函数 f_3 和 f_4 使得:

$$C(E_1 | E_2) = f_3(C(E_1), C(E_2)) \quad (5.3)$$

$$C(E_1 \uparrow E_2) = f_4(C(E_1), C(E_2)) \quad (5.4)$$

目前,关于组合证据不确定性的计算,常用的方法有如下三种。

① 最大最小法:

$$C(E_1 | E_2) = \min\{C(E_1), C(E_2)\} \quad (5.5)$$

$$C(E_1 \uparrow E_2) = \max\{C(E_1), C(E_2)\} \quad (5.6)$$

② 概率方法:

$$C(E_1|E_2)=C(E_1)\leftrightarrow C(E_2) \quad (5.7)$$

$$C(E_1\uparrow E_2)=C(E_1)+C(E_2)-C(E_1)\leftrightarrow C(E_2) \quad (5.8)$$

③ 有界方法:

$$C(E_1|E_2)=\max\{0,C(E_1)+C(E_2)-1\} \quad (5.9)$$

$$C(E_1\uparrow E_2)=\min\{1,C(E_1)+C(E_2)\} \quad (5.10)$$

5.2.3 语义问题

语义问题指上述表示和计算的含义是什么,即对它们进行解释。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,在人工智能中,处理不确定性问题的主要数学工具有概率论和模糊数学,但是它们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 $f(H,E)$,需要定义在下述三种典型情况下的取值。

(1) 若 E 为真,则 H 为真,这时 $f(H,E)=?$

(2) 若 E 为真,则 H 为假,这时 $f(H,E)=?$

(3) 若 E 对 H 没有影响,这时 $f(H,E)=?$

对于证据的不确定性度量 $C(E)$,需要定义在下述三种典型情况下的取值。

(1) E 为真, $C(E)=?$

(2) E 为假, $C(E)=?$

(3) 若对 E 一无所知, $C(E)=?$

对于一个专家系统,一旦给出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是证据和规则的不确定性的测度方法、不确定性的组合计算模式。

5.3 概率方法

长期以来,概率论的有关理论和方法都被用来度量不确定性的重要手段,因为它不仅有完善的理论,而且还为不确定性的合成与传递提供了现成的公式,因而它被最早用于不确定性知识的表示和处理,像这样纯粹用概率模型来表示和处理不确定性的方法称为纯概率方法或概率方法。

5.3.1 概率论基础

定义 5.1 全概率公式

设事件 $A_1, A_2, \dots, A_n$ 满足:

(1) 任意两个事件都互不相容,即当 $i \neq j$ 时,有 $A_i \cap A_j = \phi (i=1,2,\cdots,n; j=1,2,\cdots,n)$;

(2) $P(A_i) > 0 (i=1,2,\cdots,n)$;

(3) 样本空间 D 是各个 $A_i (i=1,2,\cdots,n)$ 集合的并集,即 $D = \bigcup_{i=1}^n A_i$ 。

则对任何事件 B 来说,有式(5.11)成立,即:

$$P(B) = \sum_{i=1}^n P(A_i)P(B | A_i) \quad (5.11)$$

该公式称为全概率公式,它提供了一种计算 $P(B)$ 的方法。

定义 5.2 Bayes 公式

设事件 $A_1, A_2, \cdots, A_n$ 满足:

(1) 任意两个事件都互不相容,即当 $i \neq j$ 时,有 $A_i \cap A_j = \phi (i=1,2,\cdots,n; j=1,2,\cdots,n)$;

(2) $P(A_i) > 0 (i=1,2,\cdots,n)$;

(3) 样本空间 D 是各个 $A_i (i=1,2,\cdots,n)$ 集合的并集,即 $D = \bigcup_{i=1}^n A_i$ 。则对任何事件 B 来说,有式(5.12)成立,即:

$$P(A_i | B) = \frac{P(A_i)P(B | A_i)}{P(B)}, \quad i=1,2,\cdots,n \quad (5.12)$$

该公式称为 Bayes 公式。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 的后验概率。

如果把全概率公式代入 Bayes 公式,则有:

$$P(A_i | B) = \frac{P(A_i)P(B | A_i)}{\sum_{j=1}^n P(A_j)P(B | A_j)}, \quad i=1,2,\cdots,n \quad (5.13)$$

这是 Bayes 公式的另一种形式。该公式给出了用逆概率 $P(B | A_i)$ 求原概率 $P(A_i | B)$ 的方法。

5.3.2 经典概率方法

设有如下产生规则:

$$\text{IF } E \text{ THEN } H_i, \quad i=1,2,\cdots,n$$

其中, E 为前提条件; H_i 为结论,具有随机性。

根据概率论中条件概率的含义,我们可以用条件概率 $P(H_i | E)$ 表示上述产生式规则的不确定性程度,即表示为在证据 E 出现的条件下,结论 H_i 成立的确定性程度。

对于复合条件:

$$E = E_1 \text{ AND } E_2 \text{ AND } \cdots \text{ AND } E_m$$

可以用条件概率 $P(H_i | E_1, E_2, \cdots, E_m)$ 作为证据 $E_1, E_2, \cdots, E_m$ 出现时结论 H 的确定性程度。

显然,这是一种很简单的方法,只能用于简单的不确定性推理。另外,由于它只考虑证据为“真”或“假”这两种极端情况,因而使其应用受到了限制。

5.3.3 逆概率方法

1. 逆概率方法的基本思想

经典概率方法要求给出在证据 E 出现情况下结论 H_i 的条件概率 $P(H_i|E)$ 。这在实际应用中是相当困难的。逆概率方法根据 Bayes 公式,用逆概率 $P(E|H_i)$ 来求原概率 $P(H_i|E)$ 。确定逆概率 $P(E|H_i)$ 比确定原概率 $P(H_i|E)$ 要容易些。例如,若以 E 代表咳嗽,以 H_i 代表支气管炎,若要得到条件概率 $P(H_i|E)$,就需要统计在咳嗽的人中有多少是患支气管炎的,统计工作量较大,而要得到逆概率 $P(E|H_i)$ 相对容易些,因为这时仅仅需要统计在患支气管炎的人中有多少人是咳嗽的,患支气管炎的人毕竟比咳嗽的人少得多。

2. 单个证据的情况

如果用产生式规则

$$\text{IF } E \text{ THEN } H_i, \quad i=1,2,\dots,n$$

中前提条件 E 代替 Bayes 公式中 B ,用 H_i 替代公式中的 A_i ,就可得到

$$P(H_i|E) = \frac{P(H_i)P(E|H_i)}{\sum_{j=1}^n P(H_j)P(E|H_j)}, \quad i=1,2,\dots,n \quad (5.14)$$

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 $H_i (i=1,2,\dots,n)$ 成立时前提条件 E 所对应的证据出现条件概率 $P(E|H_i)$,就可使用式(5.14)求出相应证据出现时结论 H_i 的条件概率 $P(H_i|E)$ 。

例 5.1 设 H_1 、 H_2 、 H_3 分别是三个结论, E 是支持这些结论的证据,且已知:

$$P(H_1)=0.3, \quad P(H_2)=0.4, \quad P(H_3)=0.5$$

$$P(E|H_1)=0.5, \quad P(E|H_2)=0.3, \quad P(E|H_3)=0.4$$

求 $P(H_1|E)$ 、 $P(H_2|E)$ 及 $P(H_3|E)$ 的值各是多少。

解: 根据式(5.14)可得:

$$\begin{aligned} P(H_1|E) &= \frac{P(H_1)P(E|H_1)}{P(H_1)P(E|H_1) + P(H_2)P(E|H_2) + P(H_3)P(E|H_3)} \\ &= \frac{0.3 \times 0.5}{0.3 \times 0.5 + 0.4 \times 0.3 + 0.5 \times 0.4} \\ &= 0.32 \end{aligned}$$

同理可得:

$$P(H_2|E) = 0.26$$

$$P(H_3|E) = 0.43$$

由此例可以看出,由于证据 E 的出现, H_1 成立的可能性略有增加, H_2 、 H_3 成立的可能性有不同程度的下降。

3. 多个证据的情况

对于有多个证据 $E_1, E_2, \dots, E_m$ 和多个结论 $H_1, H_2, \dots, H_n$, 并且每个证据都以一定程度支持结论的情况, 上面的式(5.14)可进一步扩充为

$$P(H_i | E_1, E_2, \dots, E_m) = \frac{P(H_i)P(E_1 | H_i)P(E_2 | H_i) \cdots P(E_m | H_i)}{\sum_{j=1}^n P(H_j)P(E_1 | H_j)P(E_2 | H_j) \cdots P(E_m | H_j)}, \quad i = 1, 2, \dots, n \quad (5.15)$$

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 $E_1, E_2, \dots, E_m$ 出现的条件概率 $P(E_1 | H_i), P(E_2 | H_i), \dots, P(E_m | H_i)$, 就可利用上式计算出在 $E_1, E_2, \dots, E_m$ 出现情况下 H_i 的条件概率 $P(H_i | E_1, E_2, \dots, E_m)$ 。

例 5.2 设已知:

$$P(H_1) = 0.4, \quad P(H_2) = 0.3, \quad P(H_3) = 0.3$$

$$P(E_1 | H_1) = 0.5, \quad P(E_1 | H_2) = 0.6, \quad P(E_1 | H_3) = 0.3$$

$$P(E_2 | H_1) = 0.7, \quad P(E_2 | H_2) = 0.9, \quad P(E_2 | H_3) = 0.1$$

求 $P(H_1 | E_1, E_2)$ 、 $P(H_2 | E_1, E_2)$ 及 $P(H_3 | E_1, E_2)$ 的值各是多少。

解: 根据式(5.15)可得:

$$\begin{aligned} & P(H_1 | E_1, E_2) \\ &= \frac{P(H_1)P(E_1 | H_1)P(E_2 | H_1)}{P(H_1)P(E_1 | H_1)P(E_2 | H_1) + P(H_2)P(E_1 | H_2)P(E_2 | H_2) + P(H_3)P(E_1 | H_3)P(E_2 | H_3)} \\ &= \frac{0.5 \times 0.7 \times 0.4}{0.5 \times 0.7 \times 0.4 + 0.6 \times 0.9 \times 0.3 + 0.3 \times 0.1 \times 0.3} \\ &= 0.45 \end{aligned}$$

同理可得:

$$P(H_2 | E_1, E_2) = 0.52$$

$$P(H_3 | E_1, E_2) = 0.03$$

由此例可以看出, 由于证据 E_1 和 E_2 的出现, H_1 和 H_2 成立的可能性有不同程度的增加, H_3 成立的可能性下降了。

4. 逆概率方法的优缺点

在实际应用中, 这种方法有时是很有用的。例如, 如果把 $H_i (i=1, 2, \dots, n)$ 当作一组可能发生的疾病, 把 $E_j (j=1, 2, \dots, m)$ 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 发生时观察到的症状 E_j 的条件概率, 则当对某患者观察到有症状 $E_1, E_2, \dots, E_m$ 时, 应用 Bayes 公式就可以计算出 $P(H_i | E_1, E_2, \dots, E_m)$, 从而得知患者患疾病 H_i 的可能性。

逆概率方法的优点是它有较强的理论背景和良好的数学特征, 当证据及结论都彼此独立时计算的复杂度比较低。其缺点是要求给出结论 H_i 的先验概率 $P(H_i)$ 及证据 E_j 的条件概率 $P(E_j | H_i)$, 尽管有些时候 $P(E_j | H_i)$ 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件相对困难的工作。另外, Bayes 公式的应用条件是很严格的, 它要求各事件相互独立等。如若证据间存在依赖关系, 就不能直接使用这个方法。

5.4 主观 Bayes 方法

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。概率一般可以解释为对证据和知识的主观信任度。概率推理中起关键作用的是 Bayes 公式,它是主观 Bayes 方法的基础。

主观 Bayes 方法是杜达(R. O. Duda)等于 1976 年提出的一种不确定性推理模型,是最早用于处理不确定性推理的方法之一,已成功应用于地矿勘探专家系统 PROSPECTOR 中。

5.4.1 规则不确定性的表示

1. 规则不确定性表示方法

在主观 Bayes 方法中,规则的不确定性可表示为

$$\text{IF } E \text{ THEN } (LS, LN) \ H$$

其中, (LS, LN) 用来表示规则强度。LS 和 LN 的表示形式如下。

(1) 充分性度量(LS)定义为

$$LS = \frac{P(E | H)}{P(E | \neg H)}$$

其表示 E 对 H 的支持程度,取值范围为 $[0, +\infty]$,由专家给出。

(2) 必要性度量(LN)定义为

$$LN = \frac{P(\neg E | H)}{P(\neg E | \neg H)} = \frac{1 - P(E | H)}{1 - P(E | \neg H)}$$

其表示 $\neg E$ 对 H 的支持程度,即 E 对 H 为真的必要性程度,取值范围为 $[0, +\infty]$,也是由专家凭经验给出。

下面进一步讨论 LS 和 LN 的含义。由 Bayes 公式可知

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

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

两式相除得:

$$\frac{P(H | E)}{P(\neg H | E)} = \frac{P(H) \times P(E | H)}{P(\neg H) \times P(E | \neg H)} \quad (5.16)$$

为讨论方便,下面引入几率函数

$$O(X) = \frac{P(X)}{1 - P(X)} \quad \text{或} \quad O(X) = \frac{P(X)}{P(\neg X)} \quad (5.17)$$

可见, X 的几率表示 X 出现的概率与 X 不出现的概率之比, $P(X)$ 与 $O(X)$ 的变化一致,且当 $P(X)=0$ 时, $O(X)=0$; 当 $P(X)=1$ 时, $O(X)=+\infty$ 。这样就可以把取值为 $[0, 1]$ 的 $P(X)$ 放大为取值为 $[0, +\infty]$ 的 $O(X)$ 。

把式(5.17)式中几率和概率的关系代入式(5.16)得

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

再把 LS 代入此式,可得:

$$O(H | E) = LS \times O(H) \quad (5.18)$$

式(5.18)称为 Bayes 公式的充分似然性形式。LS 称为充分似然性,因为如果 $LS = \infty$,则证据 E 对于推出 H 为真是逻辑充分的。

同理,可得到关于 LN 的公式

$$O(H | \neg E) = LN \times O(H) \quad (5.19)$$

式(5.19)称为 Bayes 公式的必要似然性形式。LN 称为必要似然性,因为如果 $LN = 0$,则 $O(H | \neg E) = 0$,这说明当 $\neg E$ 为真时, H 必假,即 E 对 H 来说是必要的。

式(5.18)和式(5.19)就是修改的 Bayes 公式。可以看出,当 E 为真时,可以利用 LS 将 H 的先验概率 $O(H)$ 更新为其后验概率 $O(H | E)$; 当 E 为假时,可以利用 LN 将 H 的先验概率 $O(H)$ 更新为其后验概率 $O(H | \neg E)$ 。

2. LS 和 LN 的性质

1) LS 的性质

当 $LS > 1$ 时, $O(H | E) > O(H)$, 说明 E 支持 H , LS 越大, $O(H | E)$ 比 $O(H)$ 大得越多, E 对 H 的支持越充分。

当 $LS \rightarrow \infty$ 时, $O(H | E) \rightarrow \infty$, 即 $P(H | E) \rightarrow 1$, 表示 E 的存在将导致 H 为真。

当 $LS = 1$ 时, $O(H | E) = O(H)$, 说明 E 对 H 没有影响。

当 $LS < 1$ 时, $O(H | E) < O(H)$, 说明 E 不支持 H 。

当 $LS = 0$ 时, $O(H | E) = 0$, 说明 E 的存在使 H 为假。

由上述分析可以看出, LS 反映的是 E 的出现对 H 为真的影响程度, 因此 LS 称为知识的充分性度量。

2) LN 的性质

当 $LN > 1$ 时, $O(H | \neg E) > O(H)$, 说明 $\neg E$ 支持 H , 即由于 E 的不出现, 增大了 H 为真的概率。并且 LN 越大, $P(H | \neg E)$ 就越大, 即 $\neg E$ 对 H 为真的支持就越强。

当 $LN \rightarrow \infty$ 时, $O(H | \neg E) \rightarrow \infty$, 即 $P(H | \neg E) \rightarrow 1$, 表示 $\neg E$ 的存在将导致 H 为真。

当 $LN = 1$ 时, $O(H | \neg E) = O(H)$, 说明 $\neg E$ 对 H 没有影响。

当 $LN < 1$ 时, $O(H | \neg E) < O(H)$, 说明 $\neg E$ 不支持 H , 即 $\neg E$ 的存在使 H 为真的可能性下降, 或者说由于 E 不存在, 将反对 H 为真。

当 $LN = 0$ 时, $O(H | \neg E) = 0$, 说明 $\neg E$ 的存在(E 不存在)将导致 H 为假。

由上述分析可以看出, LN 反映的是当 E 不存在时对 H 为真的影响程度, 因此 LN 被称为知识的必要性度量。

3) LS 与 LN 的关系

由于 E 和 $\neg E$ 不会同时支持或同时排斥 H , 因此只有下述三种情况。

① $LS > 1$ 且 $LN < 1$ 。

② $LS < 1$ 且 $LN > 1$ 。

③ $LS = LN = 1$ 。

以上结论可以进行证明。

证明①:

$$\begin{aligned}
 LS > 1 &\Leftrightarrow \frac{P(E|H)}{P(E|\neg H)} > 1 \\
 &\Leftrightarrow P(E|H) > P(E|\neg H) \\
 &\Leftrightarrow 1 - P(E|H) < 1 - P(E|\neg H) \\
 &\Leftrightarrow P(\neg E|H) < P(\neg E|\neg H) \\
 &\Leftrightarrow \frac{P(\neg E|H)}{P(\neg E|\neg H)} < 1 \\
 &\Leftrightarrow LN < 1
 \end{aligned}$$

同理可证②③,证明略。

计算公式 LS 和 LN 除了在推理过程中使用,还可以作为领域专家为 LS 和 LN 赋值的依据。在实际系统中,LS 和 LN 的值均由领域专家根据经验给出,而不是计算得出。当证据 E 越支持 H 为真时,LS 的值越大;当证据 E 对 H 越重要时,相应的 LN 的值越小。

5.4.2 证据不确定性的表示

证据通常可以分为全证据(Complete Evidence)和部分证据(Partial Evidence)。全证据就是所有的证据,即所有可能的证据和假设,它们组成证据 E 。部分证据 S 就是 E 的一部分,这部分证据也可以称为观察。在主观 Bayes 方法中,证据的不确定性是用概率表示的。全证据的可信度依赖于部分证据,表示为 $P(E|S)$ 。如果知道所有的证据,则 $E=S$,且有 $P(E|S)=P(E)$ 。其中, $P(E)$ 就是证据 E 的先验似然性, $P(E|S)$ 是已知全证据 E 中部分知识 S 后对 E 的信任,为 E 的后验似然性。

例如,在 PROSPECTOR 中,根据观察 S 直接求出 $P(E|S)$ 非常困难,所以它采用了一种变通的方法,即引进了可信度 $C(E|S)$ 的概念把区间 $[0,1]$ 的概率转换为 $-5 \sim 5$ 的 11 个整数。用户可根据实际情况在 $[-5,5]$ 中选取一个整数作为初始证据的可信度。可信度 $C(E|S)$ 与概率 $P(E|S)$ 的对应关系可用下式表示:

$$P(E|S) = \begin{cases} \frac{C(E|S) + P(E) \times (5 - C(E|S))}{5}, & 0 \leq C(E|S) \leq 5 \\ \frac{P(E) \times (C(E|S) + 5)}{5}, & -5 \leq C(E|S) < 0 \end{cases} \quad (5.20)$$

特别地

$C(E|S) = -5$, 表示在观察 S 下证据 E 肯定不存在,即 $P(E|S) = 0$;

$C(E|S) = 0$, 表示在观察 S 下与证据 E 无关,即 $P(E|S) = P(E)$;

$C(E|S) = 5$, 表示在观察 S 下证据 E 肯定存在,即 $P(E|S) = 1$ 。

这样,用户只要对证据 E 给出在观察 S 下的可信度 $C(E|S)$,系统即可求出相应的 $P(E|S)$ 。

5.4.3 组合证据不确定性的计算

当组合证据是多个单一证据的合取时,即:

$$E = E_1 \text{ AND } E_2 \text{ AND} \cdots \text{AND } E_n \quad (5.21)$$

如果已知在当前观察 S 下,每个单一证据 E_i 都有概率 $P(E_i | S)$,则:

$$P(E | S) = \min\{P(E_1 | S), P(E_2 | S), \dots, P(E_n | S)\} \quad (5.22)$$

当组合证据是多个单一证据的析取时,即:

$$E = E_1 \text{ OR } E_2 \text{ OR} \cdots \text{OR } E_n \quad (5.23)$$

如果已知在当前观察 S 下,每个单一证据 E_i 都有概率 $P(E_i | S)$,则:

$$P(E | S) = \max\{P(E_1 | S), P(E_2 | S), \dots, P(E_n | S)\} \quad (5.24)$$

对于“非”运算,用下式计算,即:

$$P(\neg E | S) = 1 - P(E | S) \quad (5.25)$$

5.4.4 不确定性推理的计算

主观 Bayes 方法推理的任务就是根据 E 的概率 $P(E)$ 及 LS、LN 的值,把 H 的先验概率 $P(H)$ 更新为后验概率,或把 H 的先验概率 $O(H)$ 更新为后验概率。由于一条规则所对应的证据有可能肯定为真,也有可能肯定为假,还有可能既非真又非假,而且在不同情况下求解后验概率的方法也不相同,因此以下分别予以讨论。

(1) 证据 E 肯定为真时,即 $P(E) = P(E | S) = 1$,由 Bayes 公式得:

$$\begin{aligned} P(H | E) &= P(E | H) \times P(H) / P(E) \\ P(\neg H | E) &= P(E | \neg H) \times P(\neg H) / P(E) \end{aligned}$$

由以上两式相除得:

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

由 LS 和概率函数的定义得:

$$P(H | E) = \frac{LS \times P(H)}{(LS - 1) \times P(H) + 1} \quad (5.26)$$

这就是把先验概率 $P(H)$ 更新为后验概率 $P(H | E)$ 的计算公式。

(2) 证据 E 肯定为假时,即 $P(E) = P(E | S) = 0$,采用和上述类似的方法可得:

$$P(H | \neg E) = \frac{LN \times P(H)}{(LN - 1) \times P(H) + 1} \quad (5.27)$$

这就是把先验概率 $P(H)$ 更新为后验概率 $P(H | \neg E)$ 的计算公式。

(3) 当证据 E 既非真又非假时,即 $0 < P(E | S) < 1$,此时就不能用上面的公式计算后验概率,可用杜达于 1976 年给出的公式:

$$P(H | S) = P(H | E) \times P(E | S) + P(H | \neg E) \leftrightarrow P(\neg E | S) \quad (5.28)$$

来计算后验概率。下面分四种情况来讨论这个公式。

① 当 $P(E | S) = 1$ 时, $P(\neg E | S) = 0$,则:

$$P(H | S) = P(H | E) = \frac{LS \times P(H)}{(LS - 1) \times P(H) + 1}$$

这是证据肯定存在的情况。

② 当 $P(E|S)=0$ 时, $P(\neg E|S)=1$, 则:

$$P(H|S) = P(H|\neg E) = \frac{LN \times P(H)}{(LN-1) \times P(H) + 1}$$

这是证据肯定不存在的情况。

③ 当 $P(E|S)=P(E)$ 时, E 与 S 无关, 利用全概率公式得:

$$\begin{aligned} P(H|S) &= P(H|E) \times P(E|S) + P(H|\neg E) \times P(\neg E|S) \\ &= P(H|E) \times P(E) + P(H|\neg E) \times P(\neg E) \\ &= P(H) \end{aligned}$$

通过分析, 可以得到 $P(E|S)$ 上的 3 个特殊值 0、 $P(E)$ 、1, 并分别取对应值 $P(H|\neg E)$ 、 $P(H)$ 、 $P(H|E)$, 从而构成 3 个特殊点的坐标 $A(0, P(H|\neg E))$ 、 $B(P(E), P(H))$ 、 $C(1, P(H|E))$ 。

④ 当 $P(E|S)$ 为其他值时, $P(E|S)$ 的值可通过上述 3 个特殊点的分段线性插值函数求得, 即:

$$P(H|S) = \begin{cases} P(H|\neg E) + \frac{P(H) - P(H|\neg E)}{P(E)} \times P(E|S), & 0 \leq P(E|S) < P(E) \\ P(H) + \frac{P(H|E) - P(H)}{1 - P(E)} \times (P(E|S) - P(E)), & P(E) \leq P(E|S) \leq 1 \end{cases} \quad (5.29)$$

该公式称为 EH 公式, 其函数图如图 5.1 所示。

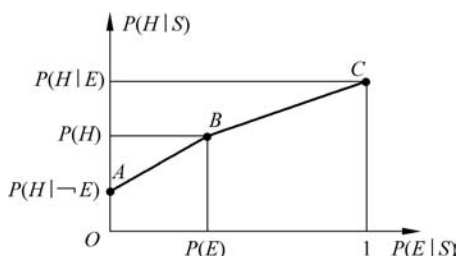


图 5.1 分段线性插值函数

对于初始证据, 由于其不确定性是由可信度 $C(E|S)$ 给出, 则此时只要把 $P(E|S)$ 与 $C(E|S)$ 的对应关系代入 EH 公式 (见式 (5.20)), 就可以得到用可信度 $C(E|S)$ 计算 $P(H|S)$ 的公式:

$$P(H|S) = \begin{cases} P(H|\neg E) + (P(H) - P(H|\neg E)) \times \left(\frac{C(E|S)}{5} + 1 \right), & C(E|S) \leq 0 \\ P(H) + (P(H|E) - P(H)) \times \frac{C(E|S)}{5}, & C(E|S) > 0 \end{cases} \quad (5.30)$$

该公式称为 CP 公式。

这样, 当用初始证据进行推理时, 根据用户告知的 $C(E|S)$, 运用 CP 公式可以求出 $P(H|S)$; 当用推理过程中得到的中间结论作为证据进行推理时, 运用 EH 公式可求出 $P(H|S)$ 。

5.4.5 结论不确定性的合成算法

若有 n 条规则都支持相同的结论,而且每条规则的前提条件所对应的证据 $E_i (i=1, 2, \dots, n)$ 都有相应的观察 S_i 与之对应,此时只要先对每条规则分别求出 H 的后验概率 $O(H|S_i)$,然后根据下述公式求出所在观察下 H 的后验概率:

$$O(H | S_1, S_2, \dots, S_n) = \frac{O(H | S_1)}{O(H)} \times \frac{O(H | S_2)}{O(H)} \times \dots \times \frac{O(H | S_n)}{O(H)} \times O(H) \quad (5.31)$$

为了进一步说明主观 Bayes 方法的推理过程,下面给出几个例子。

例 5.3 设有如下规则:

$$R_1: \text{IF } E_1 \text{ THEN } (400, 0.1) \ H$$

$$R_2: \text{IF } E_2 \text{ THEN } (60, 0.1) \ H$$

已知证据 E_1, E_2 必然发生,并且 $P(H)=0.04$,求 H 的后验概率。

解: 因为 $P(H)=0.04$,则:

$$O(H) = \frac{0.04}{1-0.04} = 0.0417$$

根据 R_1 有:

$$O(H | E_1) = \text{LS}_1 \times O(H) = 400 \times 0.0417 = 16.68$$

根据 R_2 有:

$$O(H | E_2) = \text{LS}_2 \times O(H) = 60 \times 0.0417 = 2.502$$

那么

$$\begin{aligned} O(H | E_1 E_2) &= \frac{O(H | E_1)}{O(H)} \times \frac{O(H | E_2)}{O(H)} \times O(H) \\ &= \frac{16.68}{0.0417} \times \frac{2.502}{0.0417} \times 0.0417 = 1000.8 \\ P(H | E_1 E_2) &= \frac{O(H | E_1 E_2)}{1 + O(H | E_1 E_2)} = \frac{1000.8}{1 + 1000.8} = 0.9990 \end{aligned}$$

例 5.4 设有如下规则:

$$R_1: \text{IF } E_1 \text{ THEN } (10, 0.1) \ H_1(0.03)$$

$$R_2: \text{IF } E_2 \text{ THEN } (20, 0.01) \ H_2(0.05)$$

$$R_3: \text{IF } E_3 \text{ THEN } (1, 1) \ H_3(0.3)$$

求:当证据 E_1, E_2, E_3 存在和不存在时, $P(H_i|E_i)$ 及 $P(H_i|\neg E_i)$ 的值各是多少?

解: (1) 当证据 E_1, E_2, E_3 都存在时:

$$P(H_1 | E_1) = \frac{\text{LS}_1 \times P(H_1)}{(\text{LS}_1 - 1) \times P(H_1) + 1} = \frac{10 \times 0.03}{(10 - 1) \times 0.03 + 1} = 0.2362$$

$$P(H_2 | E_2) = \frac{\text{LS}_2 \times P(H_2)}{(\text{LS}_2 - 1) \times P(H_2) + 1} = \frac{20 \times 0.05}{(20 - 1) \times 0.05 + 1} = 0.5128$$

对于 R_3 , 由于 $LS=1$, 所以 E_3 的存在对 H_3 无影响, 即 $P(H_3|E_3)=0.3$ 。

由此可以看出, E_1 的存在使 H_1 为真的可能性增加了 8 倍, E_2 的存在使 H_2 为真的可能性增加了 10 多倍。

(2) 当证据 E_1 、 E_2 、 E_3 都不存在时, R_1 和 R_2 中的 $LN=1$, 所以 E_1 与 E_2 不存在时对 H_1 和 H_2 不产生影响, 即:

$$P(H_1 | \neg E_1) = \frac{LN_1 \times P(H_1)}{(LN_1 - 1) \times P(H_1) + 1} = \frac{0.1 \times 0.03}{(0.1 - 1) \times 0.03 + 1} = 0.00308$$

$$P(H_2 | \neg E_2) = \frac{LN_2 \times P(H_2)}{(LN_2 - 1) \times P(H_2) + 1} = \frac{0.01 \times 0.05}{(0.01 - 1) \times 0.05 + 1} = 0.00053$$

对于 R_3 , 由于 $LN=1$, 所以 E_3 的不存在对 H_3 无影响, 即 $P(H_3|E_3)=0.3$ 。

由此可以看出, 由于 E_1 不存在使 H_1 为真的可能性削弱为原来的 $1/10$, E_2 不存在使 H_2 为真的可能性削弱为原来的 $1/100$ 。

例 5.5 设有规则:

R_1 : IF E_1 THEN (2,0.001) H_1

R_2 : IF E_1 AND E_2 THEN (100,0.001) H_1

R_3 : IF H_1 THEN (200,0.01) H_2

已知: $P(E_1)=P(E_2)=0.6$, $P(H_1)=0.091$, $P(H_2)=0.01$, $P(E_1|S_1)=0.76$, $P(E_2|S_2)=0.68$ 。

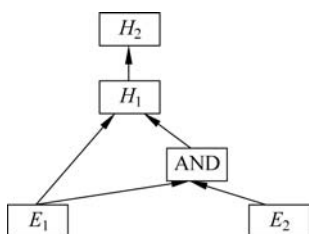


图 5.2 例 5.5 的推理网络

求: $P(H_1|S_1, S_2)$ 。

解: 由已知知识得到的推理网络如图 5.2 所示。

(1) 计算 $O(H_1|S_1)$ 。

首先将 $P(H_1)$ 更新为 E_1 下的后验概率 $P(H_1|E_1)$:

$$\begin{aligned} P(H_1 | E_1) &= \frac{LS_1 \times P(H_1)}{(LS_1 - 1) \times P(H_1) + 1} \\ &= \frac{2 \times 0.091}{(2 - 1) \times 0.091 + 1} = 0.167 \end{aligned}$$

由于 $P(E_1|S_1)=0.76 < P(E_1)$, 根据式(5.29)得:

$$\begin{aligned} P(H_1 | S_1) &= P(H_1) + \frac{P(H_1 | E_1) - P(H_1)}{1 - P(E_1)} \times (P(E_1 | S_1) - P(E_1)) \\ &= 0.091 + \frac{0.167 - 0.091}{1 - 0.6} \times (0.76 - 0.6) \\ &= 0.121 \end{aligned}$$

故:

$$O(H_1 | S_1) = \frac{P(H_1 | S_1)}{1 - P(H_1 | S_1)} = \frac{0.121}{1 - 0.121} = 0.138$$

(2) 计算 $O(H_1|S_2)$ 。

R_2 的前件是 E_1 、 E_2 的合取关系, 且 $P(E_1|S_1)=0.76$, $P(E_2|S_2)=0.68$, 则 $P(E_2|S_2) < P(E_1|S_1)$ 。按合取取最小的原则, 这里仅考虑 E_2 对 H_1 的影响, 即把计算 $P(H_1|S_1 \text{ AND } S_2)$ 的问题转化为计算 $O(H_1|S_2)$ 的问题。

首先将 H_1 的先验概率 $P(H_1)$ 更新为在 E_2 下的后验概率 $P(H_1|E_2)$:

$$P(H_1 | E_2) = \frac{LS_2 \times P(H_1)}{(LS_2 - 1) \times P(H_1) + 1} = \frac{100 \times 0.091}{(100 - 1) \times 0.091 + 1} = 0.909$$

又由于 $P(E_2|S_2) = 0.68 > P(E_2)$, 根据式(5.29)得:

$$\begin{aligned} P(H_1 | S_2) &= P(H_1) + \frac{P(H_1 | E_2) - P(H_1)}{1 - P(E_2)} \times (P(E_2 | S_2) - P(E_2)) \\ &= 0.091 + \frac{0.909 - 0.091}{1 - 0.6} \times (0.68 - 0.6) \\ &= 0.255 \end{aligned}$$

故:

$$O(H_1 | S_2) = \frac{P(H_1 | S_2)}{1 - P(H_1 | S_2)} = \frac{0.255}{1 - 0.255} = 0.342$$

(3) 计算 $O(H_1 | S_1, S_2)$ 。

首先将 H_1 的先验概率转换为先验概率:

$$O(H_1) = \frac{P(H_1)}{1 - P(H_1)} = \frac{0.091}{1 - 0.091} = 0.1$$

再根据合成公式计算 H_1 的后验概率:

$$\begin{aligned} O(H_1 | S_1, S_2) &= \frac{O(H_1 | S_1)}{O(H_1)} \times \frac{O(H_1 | S_2)}{O(H_1)} \times O(H_1) \\ &= \frac{0.138}{0.1} \times \frac{0.342}{0.1} \times 0.1 = 0.472 \end{aligned}$$

然后将后验概率转换为后验概率:

$$P(H_1 | S_1, S_2) = \frac{O(H_1 | S_1, S_2)}{1 + O(H_1 | S_1, S_2)} = \frac{0.472}{1 + 0.472} = 0.321$$

(4) 计算 $P(H_2 | S_1, S_2)$ 。

对 R_3, H_1 相当于已知事实, H_2 为结论。将 H_1 的先验概率 $P(H_2)$ 更新为在 H_1 下的后验概率 $P(H_2 | H_1)$, 由于 $P(H_1 | S_1, S_2) = 0.321 > P(H_1)$, 根据式(5.29)得到在当前观察 S_1, S_2 下 H_2 的后验概率:

$$\begin{aligned} P(H_2 | H_1) &= \frac{LS_3 \times P(H_2)}{(LS_3 - 1) \times P(H_2) + 1} = \frac{200 \times 0.01}{(200 - 1) \times 0.01 + 1} = 0.669 \\ P(H_2 | S_1, S_2) &= P(H_2) + \frac{P(H_2 | H_1) - P(H_2)}{1 - P(H_1)} \times (P(H_1 | S_1, S_2) - P(H_1)) \\ &= 0.1 + \frac{0.669 - 0.01}{1 - 0.091} \times (0.321 - 0.091) \\ &= 0.177 \end{aligned}$$

由此可以看出, H_2 的先验概率是 0.01, 通过 R_1, R_2, R_3 及初始证据进行推理, 最后推出 H_2 的后验概率为 0.177, 相当于概率增加了 17 倍多。

主观 Bayes 方法是在概率论的基础上发展起来的, 具有较完善的理论基础, 且知识的输

入转化为对 LS 和 LN 的赋值,避免了大量的数据统计工作,是一种比较实用且较灵活的不确定性推理方法。但是它在要求专家给出 LS 和 LN 的同时,还要求给出先验概率 $P(H)$,而且要求事件间相互独立,这仍然比较困难,从而也就限制了它的应用。

5.5 可信度方法

可信度方法是由美国斯坦福大学肖特里菲(E. H. Shortliffe)等在考查了非概率的和非形式化的推理过程后,于 1975 年提出的一种不确定性推理方法,并于 1976 年首次在血液病诊断专家系统 MYCIN 中得到了成功应用。

5.5.1 可信度的定义和性质

可信度是指人们根据以往的经验对某个事物或现象为真的程度的一个判断,或者说是人们对某个事物或现象为真的相信程度。

1. 可信度的定义

可信度最初定义为信任与不信任的差,即 $CF(H, E)$ 定义为

$$CF(H, E) = MB(H, E) - MD(H, E) \quad (5.32)$$

其中,CF 是由证据 E 得到假设 H 的可信度,也称为确定性因子(Certainty Factor)。

MB(Measure Belief)称为信任增长度,表示因为与前提条件 E 匹配的证据的出现,使结论 H 为真的信任的增长程度。MB(H, E)定义为

$$MB(H, E) = \begin{cases} 1, & P(H) = 1 \\ \frac{\max\{P(H | E), P(H)\} - P(H)}{1 - P(H)}, & \text{其他} \end{cases} \quad (5.33)$$

MD(Measure Disbelief)称为不信任增长度,表示因为与前提条件 E 匹配的证据的出现,对结论 H 的不信任的增长程度。MD(H, E)定义为

$$MD(H, E) = \begin{cases} 1, & P(H) = 0 \\ \frac{\min\{P(H | E), P(H)\} - P(H)}{-P(H)}, & \text{其他} \end{cases} \quad (5.34)$$

其中, $P(H)$ 表示 H 的先验概率; $P(H | E)$ 表示在前提条件 E 所对应的证据出现的情况下,结论 H 的条件概率(后验概率)。

由 MB 与 MD 的定义可以得出如下结论。

当 $MB(H, E) > 0$ 时,有 $P(H | E) > P(H)$,这说明由于 E 所对应的证据的出现增加了 H 的信任程度,但不信任程度没有变化。

当 $MD(H, E) > 0$ 时,有 $P(H | E) < P(H)$,这说明由于 E 所对应的证据的出现增加了 H 的不信任程度,而不改变对其信任的程度。

根据前面对 $CF(H, E)$ 、 $MB(H, E)$ 、 $MD(H, E)$ 的定义,可得到 $CF(H, E)$ 的计算公式:

$$CF(H, E) = \begin{cases} MB(H, E) - 0 = \frac{P(H | E) - P(H)}{1 - P(H)}, & P(H | E) > P(H) \\ 0, & P(H | E) = P(H) \\ 0 - MD(H, E) = -\frac{P(H) - P(H | E)}{P(H)}, & P(H | E) < P(H) \end{cases} \quad (5.35)$$

从上式可以看出：

若 $CF(H, E) > 0$, 则 $P(H | E) > P(H)$ 。说明由于前提条件 E 所对应的证据的出现增加了 H 为真的概率, 即增加了 H 的可信度; $CF(H, E)$ 的值越大, 增加 H 为真的可信度越大。

若 $CF(H, E) < 0$, 则 $P(H | E) < P(H)$ 。这说明由于前提条件 E 所对应的证据的出现减少了 H 为真的概率, 即增加了 H 为假的可信度; $CF(H, E)$ 的值越小, 增加 H 为假的可信度越大。

2. 可信度的性质

根据以上 CF 、 MB 、 MD 的定义, 可得到它们的如下性质。

(1) 对同一证据不可能既增加对 H 的信任程度, 同时增加对 H 的不信任程度, 因此 MB 与 MD 是互斥的, 即有如下互斥性:

$$\text{当 } MB(H, E) > 0 \text{ 时, } MD(H, E) = 0$$

$$\text{当 } MD(H, E) > 0 \text{ 时, } MB(H, E) = 0$$

(2) MB 、 MD 、 CF 具有如下值域:

$$0 \leq MB(H, E) \leq 1$$

$$0 \leq MD(H, E) \leq 1$$

$$-1 \leq CF(H, E) \leq 1$$

(3) CF 、 MB 、 MD 包括如下三种典型值:

① 当 $CF(H, E) = 1$ 时, 有 $P(H | E) = 1$, 表明 E 所对应的证据的出现使 H 为真, 此时 $MB(H, E) = 1$, $MD(H, E) = 0$;

② 当 $CF(H, E) = -1$ 时, 有 $P(H | E) = 0$, 表明 E 所对应的证据的出现使 H 为假, 此时 $MB(H, E) = 0$, $MD(H, E) = 1$;

③ 当 $CF(H, E) = 0$ 时, 则 $P(H | E) = P(H)$, 表示 H 与 E 独立, 即 E 所对应的证据出现对 H 没有影响。

(4) 对 H 的信任增长长度等于非 H 的不信任增长长度, 即:

$$\begin{aligned} MD(\neg H, E) &= \frac{P(\neg H | E) - P(\neg H)}{-P(\neg H)} = \frac{(1 - P(H | E)) - (1 - P(H))}{-(1 - P(H))} \\ &= \frac{-P(H | E) + P(H)}{-(1 - P(H))} = MB(H, E) \end{aligned}$$

(5) 对 H 的可信度与对非 H 的可信度之和等于 0, 即:

$$\begin{aligned} CF(H, E) + CF(\neg H, E) &= (MB(H, E) - MD(H, E)) + (MB(\neg H, E) - MD(\neg H, E)) \\ &= (MB(H, E) - 0) + (0 - MD(\neg H, E)) \\ &= MB(H, E) - MD(\neg H, E) \\ &= 0 \end{aligned}$$

(6) 可信度不是概率。对于概率有 $P(H) + P(\neg H) = 1$ 且 $0 \leq P(H), P(\neg H) \leq 1$, 而可信度不满足此条件。

(7) 对同一前提 E , 若支持若干个不同的结论 $H_i (i=1, 2, \dots, n)$, 则

$$\sum_{i=1}^n CF(H_i, E) \leq 1$$

因此, 如果专家给出的知识为 $CF(H_1, E) = 0.7, CF(H_2, E) = 0.4$, 则因为 $0.7 + 0.4 = 1.1 > 1$ 为非法, 应进行调整或规范化。

在实际应用中, $P(H)$ 和 $P(H|E)$ 的值很难获取, 因此 $CF(H, E)$ 的值应由领域专家给出。原则为: 若相应的证据的出现会增加 H 为真的可信度, 则 $CF(H, E) > 0$, 证据的出现对 H 为真的支持程度越高, 则 $CF(H, E)$ 的值越大; 反之, 证据的出现减少 H 为真的可信度, 则 $CF(H, E) < 0$, 证据的出现对 H 为假的支持程度越高, 使 $CF(H, E)$ 的值越小; 若相应的证据的出现与 H 无关, 则使 $CF(H, E) = 0$ 。

5.5.2 C-F 模型

C-F 模型是基于可信度表示的不确定性推理的基本方法。下面讨论其知识表示和推理问题。

1. 规则不确定性的表示

在 C-F 模型中, 规则是用产生式规则表示的, 其一般形式为

$$\text{IF } E \text{ THEN } H \text{ (CF(H, E))} \quad (5.36)$$

其中, E 是规则的前提条件, H 是规则的结论, $CF(H, E)$ 是规则的可信度, 也称为规则强度, 它描述的是知识的静态强度。

这里, 前提和结论都可以是单个命题, 也可由复合命题组成, 对它们简单说明如下。

(1) 前提证据 E 可以是一个简单条件, 也可以是由合取和析取构成的复合条件, 例如

$$E = (E_1 \text{ OR } E_2) \text{ AND } E_3 \text{ AND } E_4$$

就是一个复合条件。

(2) 结论 H 可以是一个单一的结论, 也可以是多个结论。

(3) 可信度因子 CF 通常称为可信度, 或称规则强度, 它实际上是知识的静态强度。

$CF(H, E)$ 取值范围是 $[-1, 1]$, 其值表示当证据 E 为真时, 该证据对结论 H 为真的支持程度, $CF(H, E)$ 的值越大, 说明 E 对结论 H 为真的支持程度越大。

2. 证据不确定性的表示

在 C-F 模型中, 证据 E 的不确定性也是用可信度因子 $CF(E)$ 表示的, 其取值范围同样是 $[-1, 1]$, 其典型值为

- 当证据 E 肯定为真时, $CF(E) = 1$;
- 当证据 E 肯定为假时, $CF(E) = -1$;
- 当证据 E 一无所知时, $CF(E) = 0$ 。

证据可信度的来源有以下两种情况: 如果是初始证据, 其可信度是由提供证据的用户给出的; 如果是先前推出的中间结论又作为当前推理的证据, 则其可信度在推出该结论时

由不确定性的更新算法计算得到。

$CF(E)$ 所描述的是证据的动态强度,尽管它和知识的静态强度在表示方法上类似,但二者的含义完全不同。知识的静态强度 $CF(H, E)$ 表示的是规则的强度,即当 E 所对应的证据为真时对 H 的影响程度,而动态强度 $CF(E)$ 表示的是证据 E 当前的不确定性程度。

3. 组合证据不确定性的计算

对证据的组合形式可分为“合取”与“析取”两种基本情况。当组合证据是多个单一证据的合取时,即:

$$E = E_1 \text{ AND } E_2 \text{ AND } \cdots \text{ AND } E_n$$

若已知 $CF(E_1), CF(E_2), \dots, CF(E_n)$, 则:

$$CF(E) = \min\{CF(E_1), CF(E_2), \dots, CF(E_n)\} \quad (5.37)$$

当组合证据是多个单一证据的析取时,即:

$$E = E_1 \text{ OR } E_2 \text{ OR } \cdots \text{ OR } E_n$$

若已知 $CF(E_1), CF(E_2), \dots, CF(E_n)$, 则:

$$CF(E) = \max\{CF(E_1), CF(E_2), \dots, CF(E_n)\} \quad (5.38)$$

4. 否定证据不确定性的计算

设 E 为证据,则该证据的否定,记为 $\neg E$,若已知 E 的可信度为 $CF(E)$, 则:

$$CF(\neg E) = -CF(E) \quad (5.39)$$

5. 不确定性推理计算

C-F 模型中的不确定性推理实际上是从不确定性的初始证据出发,不断运用相关的不确定性规则,逐步推出最终结论和该结论的可信度的过程。而每次运用不确定性知识都需要由证据的不确定性和规则的不确定性去计算结论的不确定性。

① 证据肯定存在($CF(E)=1$)时,则:

$$CF(H) = CF(H, E)$$

这说明,规则强度 $CF(H, E)$ 实际上就是在前提条件对应的证据为真时结论 H 的可信度。

② 证据不是肯定存在($CF(E) \neq 1$)时,则:

$$CF(H) = CF(H, E) \times \max\{0, CF(E)\} \quad (5.40)$$

由此可以看出,若 $CF(E) < 0$, 即相应的证据以某种程度为假,则 $CF(H) = 0$ 。这说明在该模型中没有考虑证据为假时对结论 H 所产生的影响。

③ 证据是多个条件组合的情况。即如果有两条规则推出一个相同结论,并且这两条规则的前提相互独立,结论的可信度又不相同,则可用不确定性的合成算法求出该结论的综合可信度。

设有如下规则:

$$\text{IF } E_1 \text{ THEN } H \text{ (CF(H, E}_1\text{))}$$

$$\text{IF } E_2 \text{ THEN } H \text{ (CF(H, E}_2\text{))}$$

则结论 H 的综合可信度可分以下两步计算。

第一步: 分别对每条规则求出其 $CF(H)$, 即:

$$CF_1(H) = CF(H, E_1) \times \max\{0, CF(E_1)\}$$

$$CF_2(H) = CF(H, E_2) \times \max\{0, CF(E_2)\}$$

第二步: 求 E_1 与 E_2 对 H 的综合可信度, 即:

$$CF(H) = \begin{cases} CF_1(H) + CF_2(H) - CF_1(H) \times CF_2(H), & CF_1(H) \geq 0, CF_2(H) \geq 0 \\ CF_1(H) + CF_2(H) + CF_1(H) \times CF_2(H), & CF_1(H) < 0, CF_2(H) < 0 \\ \frac{CF_1(H) + CF_2(H)}{1 - \min\{|CF_1(H)|, |CF_2(H)|\}}, & CF_1(H) \text{ 与 } CF_2(H) \text{ 异号} \end{cases} \quad (5.41)$$

如果可由多条规则推出同一个结论, 并且这些规则的前提相互独立, 结论的可信度又不相同, 则可以将上述合成过程推广应用到多条规则支持同一条结论, 且规则前提可以包含多个证据的情况。这时合成过程是先把第一条与第二条合成, 再用该合成后的结论与第三条合成, 依次进行下去, 直到全部合成完为止。

例 5.6 已知有下列一组规则:

$$R_1: \text{IF } E_1 \text{ THEN } H_1 \quad (0.8)$$

$$R_2: \text{IF } E_2 \text{ THEN } H_1 \quad (0.5)$$

$$R_3: \text{IF } E_3 \text{ AND } H_1 \text{ THEN } H_2 \quad (0.8)$$

已知初始可信度: $CF(E_1) = CF(E_2) = CF(E_3) = 1$, 求: $CF(H_1), CF(H_2)$ 。

解: (1) 对知识 R_1, R_2 , 分别计算 $CF(H_1)$ 。

$$CF_1(H_1) = CF(H_1, E_1) \times \max\{0, CF(E_1)\} = 0.8 \times \max\{0, 1\} = 0.8$$

$$CF_2(H_1) = CF(H_1, E_2) \times \max\{0, CF(E_2)\} = 0.5 \times \max\{0, 1\} = 0.5$$

(2) 利用合成算法计算 H_1 的综合可信度。

$$\begin{aligned} CF_{1,2}(H_1) &= CF_1(H_1) + CF_2(H_1) - CF_1(H_1) \times CF_2(H_1) \\ &= 0.8 + 0.5 - 0.8 \times 0.5 \\ &= 0.9 \end{aligned}$$

(3) 计算 H_2 的可信度。这时 H_1 作为 H_2 的证据, 其可信度由前面计算, 即 $CF(H_1) = 0.9$, 又 $CF(E_3) = 1$, 故:

$$\begin{aligned} CF(H_2) &= CF(H_2, E_3 \text{ AND } H_1) \times \max\{0, CF(E_3 \text{ AND } H_1)\} \\ &= 0.8 \times \max\{0, 0.9\} \\ &= 0.72 \end{aligned}$$

例 5.7 设有如下一组规则:

$$R_1: \text{IF } E_1 \text{ THEN } H \quad (0.9)$$

$$R_2: \text{IF } E_2 \text{ THEN } H \quad (0.6)$$

$$R_3: \text{IF } E_3 \text{ THEN } H \quad (-0.5)$$

$$R_4: \text{IF } E_4 \text{ AND } (E_5 \text{ OR } E_6) \text{ THEN } H \quad (0.8)$$

已知: $CF(E_2) = 0.8, CF(E_3) = 0.6, CF(E_4) = 0.5, CF(E_5) = 0.6, CF(E_6) = 0.8$ 。

求: H 的综合可信度 $CF(H)$ 。

解: 由 R_4 得到:

$$\begin{aligned}
CF(E_1) &= 0.8 \times \max\{0, CF(E_4 \text{ AND}(E_5 \text{ OR } E_6))\} \\
&= 0.8 \times \max\{0, \min\{CF(E_4), CF(E_5 \text{ OR } E_6)\}\} \\
&= 0.8 \times \max\{0, \min\{CF(E_4), \max\{CF(E_5), CF(E_6)\}\}\} \\
&= 0.8 \times \max\{0, \min\{0.5, 0.8\}\} \\
&= 0.8 \times \max\{0, 0.5\} \\
&= 0.4
\end{aligned}$$

由 R_1 得到:

$$\begin{aligned}
CF_1(H) &= CF(H, E_1) \times \max\{0, CF(E_1)\} \\
&= 0.9 \times \max\{0, 0.4\} \\
&= 0.36
\end{aligned}$$

由 R_2 得到:

$$\begin{aligned}
CF_2(H) &= CF(H, E_2) \times \max\{0, CF(E_2)\} \\
&= 0.6 \times \max\{0, 0.8\} \\
&= 0.48
\end{aligned}$$

由 R_3 得到:

$$\begin{aligned}
CF_3(H) &= CF(H, E_3) \times \max\{0, CF(E_3)\} \\
&= -0.5 \times \max\{0, 0.6\} \\
&= -0.3
\end{aligned}$$

根据结论不确定性的合成算法得到:

$$\begin{aligned}
CF_{1,2}(H) &= CF_1(H) + CF_2(H) - CF_1(H) \times CF_2(H) \\
&= 0.36 + 0.48 - 0.36 \times 0.48 \\
&= 0.67
\end{aligned}$$

$$\begin{aligned}
CF_{1,2,3}(H) &= \frac{CF_{1,2}(H) + CF_3(H)}{1 - \min\{|CF_{1,2}(H)|, |CF_3(H)|\}} \\
&= \frac{0.67 - 0.3}{1 - \min\{0.67, 0.3\}} = \frac{0.37}{0.7} \\
&= 0.53
\end{aligned}$$

这就是所求出的综合可信度,即 $CF(H)=0.53$ 。

5.5.3 可信度方法的说明

1. 可信度的计算问题

CF 的原始定义为

$$CF = MB - MD$$

该定义有一个困难之处。因为一个反面证据的影响可以抑制很多正面证据的影响,反之亦然。例如,如果 $MB=0.999, MD=0.799$, 则 $CF=0.2$ 。后来,MYCIN 中 CF 的定义修改为

$$CF = \frac{MB - MD}{1 - \min\{MB, MD\}}$$

这样可以削弱一个反面证据对多个正面证据的影响。例如对上面的 MB, MD 值, 有

$$CF = \frac{0.999 - 0.799}{1 - \min\{0.999, 0.799\}} = 0.995$$

另外, 在 MYCIN 中, 一个规则前件的 CF 值必须大于 0.2, 这样该规则的前件能认为真并激活该规则。在 CF 理论中, 阈值 0.2 不是作为一个基本公理, 而是作为一个处理方法来减少所激活的仅弱支持的规则数目。如果没有这个阈值, 许多 CF 值很小甚至没有值的规则将被激活, 这将大大降低系统的效率。

2. 可信度方法的特点:

可信度方法的优点如下。

(1) 可信度方法具有简洁、直观的优点。通过简单的计算, 不确定性就可以在系统中传播, 并且计算具有线性的复杂度, 推理的近似效果也比较理想。

(2) 可信度方法也很容易理解, 并且将不信任和信任清楚地区分开来。

可信度方法的缺点如下。

(1) CF 值可能与条件概率得出的值相反。例如:

$$P(H_1) = 0.8, \quad P(H_2) = 0.2, \quad P(H_1 | E) = 0.9, \quad P(H_2 | E) = 0.8$$

则:

$$CF(H_1, E) = 0.5, \quad CF(H_2, E) = 0.75$$

如果一种疾病具有很高的条件概率, 但却有很低的 CF 值, 则可能会产生矛盾。

(2) 通常:

$$P(H | E) \neq P(H | S) \times P(S | E)$$

其中, S 是基于证据 E 的某些中间假设。但在推理链中的两条规则的 CF 却是作为独立概率计算的, 即

$$CF(H, E) = CF(H, S) \times CF(S, E)$$

(3) MYCIN 一般应用于短推理链, 而且假设简单的问题。如果把该方法应用于不具备短推理链、简单假设的领域, 则可能会出问题。

(4) 由于可能导致计算的累计误差, 如果多个规则逻辑等价于一个规则, 则采用一个规则和多个规则计算的 CF 值可能就不相同。

(5) 组合规则使用的顺序不同, 可能得出不同的结果。

5.6 证据理论

证据理论(Evidential of Evidence), 也称为 D-S(Dempster-Shafer)理论, 最早是由德姆斯特(A. P. Dempster)提出的。他试图用一个概率范围而不是单个的概率值去模拟不确定性。莎弗(G. Shafer)进一步拓展了德姆斯特的工作, 称为证据推理(Evidential Reasoning), 用于处理不确定性、不精确以及间或不准确的信息。由于证据理论将概率论中的单点赋值扩展为集合赋值, 弱化了相应的公理系统, 满足了比概率更弱的要求, 因此可看作一种广义概率论。

证据理论中引入了信任函数来度量不确定性, 并引用似然函数来处理由于“不知道”引起的不确定性, 并且不必事先给出知识的先验概率, 与主观 Bayes 方法相比, 具有较大的灵

活性。因此,证据理论得到了广泛的应用。同时,可信度可以看作证据理论的一个特例,证据理论给了可信度一个理论性的基础。

5.6.1 证据理论的形式描述

在证据理论中,可以分别用信任函数、似然函数及类概率函数来描述知识的精确信任度、不可驳斥信任度及估计信任度,即可以从各种不同角度刻画命题的不确定性。

1. 概率分配函数

证据理论处理集合上的不确定性问题。为适应这一需要,首先应该建立命题与集合之间的一一对应关系,把命题的不确定性转化为集合的不确定性。

设 Ω 为变量 x 的所有可能取值的有限集合(亦称样本空间),且 Ω 中的每个元素都相互独立,则由 Ω 的所有子集构成的集合称为幂集,记为 2^Ω 。

当 Ω 中的元素个数为 N 时,则其幂集的元素个数为 2^N ,且其中的每个元素 A 都对应于一个关于 x 的命题,称该命题为“ x 的值在 A 中”。

例 5.8 设: $\Omega = \{\text{黑}, \text{白}, \text{蓝}\}$, 求: Ω 的幂集 2^Ω 。

解: Ω 的幂集可包括如下子集:

$$A_0 = \emptyset, \quad A_1 = \{\text{黑}\}, \quad A_2 = \{\text{白}\}, \quad A_3 = \{\text{蓝}\}$$

$$A_4 = \{\text{黑}, \text{白}\}, \quad A_5 = \{\text{黑}, \text{蓝}\}, \quad A_6 = \{\text{白}, \text{蓝}\}, \quad A_7 = \{\text{黑}, \text{白}, \text{蓝}\}$$

其中, $\emptyset$ 表示空集,上述子集的个数正好是 $2^3 = 8$, 所以: $2^\Omega = \{A_0, A_1, A_2, A_3, A_4, A_5, A_6, A_7\}$ 。

定义 5.3 设函数 $m: 2^\Omega \rightarrow [0, 1]$, 且满足:

$$m(\emptyset) = 0$$

$$\sum_{A \subseteq \Omega} m(A) = 1$$

则称 m 是 2^Ω 上的概率分配函数, $m(A)$ 称为 A 的基本概率数。 $m(A)$ 表示依据当前的环境对假设集 A 的信任程度。

对例 5.8 所给出的有限集 Ω , 若定义 2^Ω 上的一个基本函数 m :

$$m(A_0, A_1, A_2, A_3, A_4, A_5, A_6, A_7) = (0, 0.3, 0, 0.1, 0.2, 0.2, 0, 0.2)$$

其中, $(0, 0.3, 0, 0.1, 0.2, 0.2, 0, 0.2)$ 分别是幂集中各个子集的基本概率数。显然, m 满足概率分配函数的定义。

对概率分配函数的几点说明。

(1) 概率分配函数的作用是把 Ω 的任意一个子集都映射为 $[0, 1]$ 上的一个数 $m(A)$ 。

当 A 包含于 Ω 且 A 由单个元素组成时, $m(A)$ 表示对 A 的精确信任度;

当 A 包含于 Ω 、 $A \neq \Omega$, 且 A 由多个元素组成时, $m(A)$ 也表示对 A 的精确信任度, 但不知道这部分信任度该分给 A 中哪些元素;

当 $A = \Omega$ 时, 则 $m(A)$ 是对 Ω 的各个子集进行信任分配后剩下的部分, 表示不知道该如何对它进行分配。

例如, 对上例所给出的有限集 Ω 及基本函数 m :

当 $A_1 = \{\text{黑}\}$ 时, 有 $m(A_1) = 0.3$, 表示对命题“ x 是黑色”的精确信任度为 0.3。

当 $A_4 = \{\text{黑}, \text{白}\}$ 时, 有 $m(A_4) = 0.2$, 表示对命题“ x 或者是黑色, 或者是白色”的精确信任度为 0.2, 却不知道该怎么把这 0.2 是分给 $\{\text{黑}\}$ 还是分给 $\{\text{白}\}$ 。

当 $A_7 = \Omega = \{\text{黑}, \text{白}, \text{蓝}\}$ 时, 有 $m(A_7) = 0.2$, 表示不知道该对这 0.2 如何分配, 但它不属于 $\{\text{黑}\}$, 就一定属于 $\{\text{白}\}$ 或 $\{\text{蓝}\}$, 只是在现有认识下还不知道该如何分配而已。

(2) m 是 2^Ω 上而非 Ω 上的概率分布, 所以概率分配函数不是概率, 它们不必相等, 而且 $m(A) \neq 1 - m(\neg A)$ 。

例如, 在例 5.8 中 m 符合概率分配函数的定义, 但是:

$$m(A_1) + m(A_2) + m(A_3) = 0.3 + 0 + 0.1 = 0.4 < 1$$

而概率要求 $P(A_1) + P(A_2) + P(A_3) = 1$, 因此 m 不是概率。

2. 信任函数

定义 5.4 信任函数 (Belief Function)

$$\text{Bel}: 2^\Omega \rightarrow [0, 1]$$

对任意的 $A \subseteq \Omega$ 有:

$$\text{Bel}(A) = \sum_{B \subseteq A} m(B)$$

$\text{Bel}(A)$ 表示当前环境下, 对假设集 A 的信任程度, 其值为 A 的所有子集的基本概率之和, 表示对 A 的总的信任度。当 A 为单一元素组成的集合时, $\text{Bel}(A) = m(A)$, 因此 $\text{Bel}(A)$ 又称为下限函数。

例如, 对例 5.8 有:

$$\begin{aligned} \text{Bel}(A_4) &= m(A_1) + m(A_2) + m(A_4) \\ &= 0.3 + 0 + 0.2 \\ &= 0.5 \end{aligned}$$

3. 似然函数

定义 5.5 似然函数 (Plausibility Function)

$$\text{Pl}: 2^\Omega \rightarrow [0, 1]$$

对任意的 $A \subseteq \Omega$ 有:

$$\text{Pl}(A) = 1 - \text{Bel}(\neg A)$$

其中, $\neg A = \Omega - A$ 。

似然函数又称不可驳斥函数或上限函数。由于 $\text{Bel}(A)$ 表示对 A 为真的信任度, $\text{Bel}(\neg A)$ 表示对 $\neg A$ 的信任度, 即 A 为假的信任度, 因此, $\text{Pl}(A)$ 表示对 A 为非假的信任度。

例如, 对例 5.8 有:

$$\begin{aligned} \text{Pl}(A_3) &= 1 - \text{Bel}(\neg A_3) \\ &= 1 - (m(A_1) + m(A_2) + m(A_4)) \\ &= 1 - (0.3 + 0 + 0.2) \\ &= 0.5 \end{aligned}$$

这里的 0.5 是对“蓝”为非假的信任度。由于“蓝”为真的精确信任度为 0.1, 而剩下的 0.5—

0.1=0.4 则是知道非假但却不能肯定为真的那部分。

推论 5.1 设有信任函数 m , 似然函数 Pl , 则

$$Pl(A) = \sum_{A \cap B \neq \emptyset} m(B)$$

证明:

$$\begin{aligned} Pl(A) - \sum_{A \cap B \neq \emptyset} m(B) &= 1 - Bel(\neg A) - \sum_{A \cap B \neq \emptyset} m(B) \\ &= 1 - (Bel(\neg A) + \sum_{A \cap B \neq \emptyset} m(B)) \\ &= 1 - \left(\sum_{C \subseteq \neg A} m(C) + \sum_{A \cap B \neq \emptyset} m(B) \right) \\ &= 1 - \sum_{D \subseteq \Omega} m(D) \\ &= 0 \end{aligned}$$

所以可得

$$Pl(A) = \sum_{A \cap B \neq \emptyset} m(B)$$

因此命题“ x 在 A 中”的似然性由与命题“ x 在 B 中”有关的 m 值确定, 其中命题“ x 在 B 中”并不会使得命题“ x 不在 A 中”成立。所以, 一个事件的似然性是建立在对其相反事件不信任的基础上的。

4. 信任函数和似然函数的性质

信任函数和似然函数满足下列性质:

- (1) $Bel(\emptyset)=0, Bel(\Omega)=1, Pl(\emptyset)=0, Pl(\Omega)=1$;
- (2) 如果 $A \subseteq B$, 则 $Bel(A) \leq Bel(B), Pl(A) \leq Pl(B)$;
- (3) $\forall A \subseteq \Omega, Pl(A) \geq Bel(A)$;
- (4) $\forall A \subseteq \Omega, Bel(A) + Bel(\neg A) \leq 1, Pl(A) + Pl(\neg A) \geq 1$ 。

由于 $Bel(A)$ 和 $Pl(A)$ 分别表示 A 为真的信任度和 A 为非假的信任度, 因此, 可分别称 $Bel(A)$ 和 $Pl(A)$ 为对 A 信任程度的下限和上限, 记为

$$A[Bel(A), Pl(A)]$$

$Pl(A) - Bel(A)$ 表示既不信任 A , 也不信任 $\neg A$ 的程度, 即对于 A 是真是假不知道的程度。

5. 概率分配函数的正交和

在实际问题中, 对于相同的证据, 由于来源不同, 可能会得到不同的概率分配函数。例如, 考虑 $\Omega = \{\text{黑}, \text{白}\}$, 假设从不同知识源得到的概率分配函数分别为

$$m_1(\emptyset, \{\text{黑}\}, \{\text{白}\}, \{\text{黑}, \text{白}\}) = (0, 0.4, 0.5, 0.1)$$

$$m_2(\emptyset, \{\text{黑}\}, \{\text{白}\}, \{\text{黑}, \text{白}\}) = (0, 0.6, 0.2, 0.2)$$

在这种情况下, 需要对它们进行组合。

定义 5.6 设 m_1 和 m_2 是两个不同的概率分配函数, 则其正交和 $m = m_1 \oplus m_2$ 满足

$$m(\emptyset) = 0$$

$$m(A) = K^{-1} \sum_{x \cap y = A} m_1(x)m_2(y) \quad (5.42)$$

其中

$$K = 1 - \sum_{x \cap y = \emptyset} m_1(x)m_2(y) = \sum_{x \cap y \neq \emptyset} m_1(x)m_2(y) \quad (5.43)$$

如果 $K \neq 0$, 则正交和 m 也是一个概率分配函数; 如果 $K = 0$, 则不存在正交和 m , 称 m_1 与 m_2 矛盾。

例 5.9 设 $\Omega = \{a, b\}$, 且从不同知识源得到的概率分配函数分别为

$$m_1(\emptyset, \{a\}, \{b\}, \{a, b\}) = (0, 0.5, 0.3, 0.2)$$

$$m_2(\emptyset, \{a\}, \{b\}, \{a, b\}) = (0, 0.3, 0.6, 0.1)$$

求: 正交和 $m = m_1 \oplus m_2$ 。

解: 先求 K :

$$\begin{aligned} K &= 1 - \sum_{x \cap y = \emptyset} m_1(x)m_2(y) \\ &= 1 - (m_1(\{a\})m_2(\{b\})) + (m_1(\{b\})m_2(\{a\})) \\ &= 1 - (0.5 \times 0.6 + 0.3 \times 0.3) \\ &= 0.61 \end{aligned}$$

然后求 $m(\emptyset, \{a\}, \{b\}, \{a, b\})$, 由于

$$\begin{aligned} m(\{a\}) &= \frac{1}{0.61} \times \sum_{x \cap y = \{a\}} m_1(x)m_2(y) \\ &= \frac{1}{0.61} \times (m_1(\{a\})m_2(\{a\}) + m_1(\{a\})m_2(\{a, b\}) + m_1(\{a, b\})m_2(\{a\})) \\ &= \frac{1}{0.61} \times (0.5 \times 0.3 + 0.5 \times 0.1 + 0.2 \times 0.3) \\ &= 0.43 \end{aligned}$$

同理可得

$$m(\{b\}) = 0.54$$

$$m(\{a, b\}) = 0.03$$

组合后得到概率分配函数

$$m(\emptyset, \{a\}, \{b\}, \{a, b\}) = (0, 0.43, 0.54, 0.03)$$

5.6.2 证据理论的推理模型

在上述证据理论的形式描述中, 信任函数 $\text{Bel}(A)$ 和似然函数 $\text{Pl}(A)$ 分别表示命题 A 的信任度的下限和上限, 同样可用它来表述知识强度的下限和上限。这样就可在此表示的基础上建立相应的不确定性推理模型。

另外, 从信任函数和似然函数的定义可以看出, 它们都是建立在概率分配函数的基础上的。那么, 当概率分配函数的定义不同时, 将会得到不同的推理模型。下面给出一个特殊的概率分配函数, 并在该函数的基础上建立一个具体的不确定性推理模型。

1. 一个特殊的概率分配函数

设 $\Omega = \{s_1, s_2, \dots, s_n\}$, m 为定义在 2^Ω 上的概率分配函数, 且 m 满足

(1) $m(\{s_i\}) \geq 0$, 对任意 $s_i \in \Omega$;

(2) $\sum_{i=1}^n m(\{s_i\}) \leq 1$;

(3) $m(\Omega) = 1 - \sum_{i=1}^n m(\{s_i\})$;

(4) 当 $A \subset \Omega$, 且 $|A| > 1$ 或 $|A| = 0$ 时, $m(A) = 0$ 。其中, $|A|$ 表示命题 A 对应的集合中元素的个数。

这里定义的是一个特殊的概率分配函数, 只有当子集中的元素个数为 1 时, 其概率分配函数才有可能大于 0; 当子集中有多个或空集且不等于全集时, 其概率分配函数均为 0; 全集 Ω 的概率分配函数第(3)式计算。

例 5.10 设 $\Omega = \{\text{红}, \text{黄}, \text{白}\}$, 有如下的概率分配函数:

$$m_1(\emptyset, \{\text{红}\}, \{\text{黄}\}, \{\text{白}\}, \Omega) = (0, 0.6, 0.2, 0.1, 0.1)$$

其中, $m(\{\text{红}, \text{黄}\}) = m(\{\text{红}, \text{白}\}) = m(\{\text{黄}, \text{白}\}) = 0$ 符合上述概率分配函数的定义。

下面讨论满足上述特殊概率分配函数的信任函数、似然函数, 以及它们的正交和。

定义 5.7 设 m 为上述定义的一个特殊概率分配函数, 对任意命题 $A \subseteq \Omega$, 则

(1) 信任函数为:

$$\text{Bel}(A) = \sum_{s_i \in A} m(\{s_i\})$$

$$\text{Bel}(\Omega) = \sum_{B \subseteq \Omega} m(B) = \sum_{i=1}^n m(\{s_i\}) + m(\Omega) = 1$$

(2) 似然函数为:

$$\begin{aligned} \text{Pl}(A) &= 1 - \text{Bel}(\neg A) = 1 - \sum_{s_i \in \neg A} m(\{s_i\}) \\ &= 1 - \sum_{s_i \in \neg A} m(\{s_i\}) = 1 - \left(\sum_{i=1}^n m(\{s_i\}) - \sum_{s_i \in A} m(\{s_i\}) \right) \\ &= 1 - (1 - m(\Omega) - \text{Bel}(A)) \\ &= m(\Omega) + \text{Bel}(A) \end{aligned}$$

$$\text{Pl}(\Omega) = 1 - \text{Bel}(\neg \Omega) = 1 - \text{Bel}(\emptyset) = 1$$

从上面的定义可以看出, 对任何命题 $A \subseteq \Omega$ 和 $B \subseteq \Omega$, 均有

$$\text{Pl}(A) - \text{Bel}(A) = \text{Pl}(B) - \text{Bel}(B) = m(\Omega)$$

例 5.11 设 $\Omega = \{\text{红}, \text{黄}, \text{绿}\}$, 如下的概率分配函数:

$$m_1(\emptyset, \{\text{红}\}, \{\text{黄}\}, \{\text{绿}\}, \Omega) = (0, 0.6, 0.2, 0.1, 0.1)$$

设 $A = \{\text{红}, \text{黄}\}$, 求 $m(\Omega)$, $\text{Bel}(A)$, $\text{Pl}(A)$ 的值。

解: $m(\Omega) = 1 - (m(\{\text{红}\}) + m(\{\text{黄}\}) + m(\{\text{绿}\})) = 1 - (0.6 + 0.2 + 0.1) = 0.1$

$$\text{Bel}(A) = m(\{\text{红}\}) + m(\{\text{黄}\}) = 0.6 + 0.2 = 0.8$$

$$\text{Pl}(A) = m(\Omega) + \text{Bel}(\{\text{红}, \text{黄}\}) = 0.1 + 0.8 = 0.9$$

或 $Pl(A) = 1 - Bel(\neg \{\text{红, 黄}\}) = 1 - Bel(\{\text{绿}\}) = 1 - 0.1 = 0.9$

2. 类概率函数

利用信任函数 $Bel(A)$ 和似然函数 $Pl(A)$, 可以定义 A 的类概率函数, 并把它作为 A 的不确定性度量。

定义 5.8 设 Ω 为有限域, 对任意命题 $A \subseteq \Omega$, 命题 A 的类概率函数为

$$f(A) = Bel(A) + \frac{|A|}{|\Omega|} (Pl(A) - Bel(A)) \quad (5.44)$$

其中, $|A|$ 、 $|\Omega|$ 分别表示 A 和 Ω 中包含元素的个数。

类概率函数 $f(A)$ 具有以下性质。

- (1) $f(\emptyset) = 0, f(\Omega) = 1$;
- (2) $0 \leq f(A) \leq 1, \forall A \subseteq \Omega$;
- (3) $Bel(A) \leq f(A) \leq Pl(A), \forall A \subseteq \Omega$;
- (4) $f(\neg A) = 1 - f(A), \forall A \subseteq \Omega$ 。

5.6.3 证据不确定性的表示

在证据理论中, 所有输入的已知数据、规则前提条件及结论部分的命题都称为证据。证据 E 的不确定性可以用类概率函数 $f(E)$ 表示。

在实际系统中, 如果是初始证据, 其不确定性是由用户给出, 如果是推理过程中得到的中间结论, 则其不确定性由推理得到。

5.6.4 规则不确定性的表示

在证据理论中, 规则的不确定性可表示为

$$\text{IF } E \text{ THEN } H, \text{ CF}$$

其中, H 为假设, E 为支持 H 成立的假设集, 它们是命题的逻辑组合。CF 为可信度因子。

$H = \{a_1, a_2, \dots, a_m\}, a_i \in \Omega (i=1, 2, \dots, m), H$ 为假设集合 Ω 的子集。

$CF = \{c_1, c_2, \dots, c_m\}, c_i$ 用来描述前提 E 成立时 a_i 的可信度。CF 应满足如下条件:

$$c_i \geq 0, \quad 1 \leq i \leq m$$

$$\sum_{i=1}^m c_i \leq 1$$

5.6.5 不确定性推理计算

定义 5.9 对于不确定性规则

$$\text{IF } E \text{ THEN } H, \text{ CF}$$

定义

$$m(\{a_i\}) = f(E)c_i, \quad i = 1, 2, \dots, m$$

或表示为

$$m(\{a_1\}, \{a_2\}, \dots, \{a_m\}) = (f(E)c_1, f(E)c_2, \dots, f(E)c_m)$$

规定

$$m(\Omega) = 1 - \sum_{i=1}^m m(\{a_i\})$$

而对于 Ω 的所有其他子集 H , 均有 $m(H) = 0$ 。

当 H 为 Ω 的真子集时, 有

$$\text{Bel}(H) = \sum_{B \subseteq H} m(B) = \sum_{i=1}^m m(\{a_i\}) \quad (5.45)$$

进一步地, 可以计算 $\text{Pl}(H)$ 和 $f(H)$ 。

5.6.6 组合证据不确定性的计算

当规则的前提(证据) E 是多个命题的合取或析取时, 定义

$$f(E_1 \wedge E_2 \wedge \dots \wedge E_n) = \min\{f(E_1), f(E_2), \dots, f(E_n)\}$$

$$f(E_1 \vee E_2 \vee \dots \vee E_n) = \max\{f(E_1), f(E_2), \dots, f(E_n)\}$$

当有多条规则支持同一结论时, 如果 $H = \{a_1, a_2, \dots, a_n\}$, 则

$$\text{IF } E_1 \text{ THEN } H, \text{ CF}_1 \quad (\text{CF}_1 = \{c_{11}, c_{12}, \dots, c_{1n}\})$$

$$\text{IF } E_2 \text{ THEN } H, \text{ CF}_2 \quad (\text{CF}_2 = \{c_{21}, c_{22}, \dots, c_{2n}\})$$

⋮

$$\text{IF } E_m \text{ THEN } H, \text{ CF}_m \quad (\text{CF}_m = \{c_{m1}, c_{m2}, \dots, c_{mn}\})$$

如果这些规则相互独立地支持结论 H 的成立, 可以先计算

$$m_i(\{a_1\}, \{a_2\}, \dots, \{a_m\}) = (f(E_i)c_{i1}, f(E_i)c_{i2}, \dots, f(E_i)c_{im}), \quad i = 1, 2, \dots, m$$

然后根据前面介绍的求正交和的方法, 对这些 m_i 求正交和, 以组合所有规则对结论 H 的支持。一旦累加的正交和 $m(H)$ 计算出来, 就可以计算 $\text{Bel}(H)$ 、 $\text{Pl}(H)$ 、 $f(H)$ 。

例 5.12 有如下的推理规则:

$$R_1: \text{IF } E_1 \vee (E_2 \wedge E_3) \text{ THEN } A_1 = \{a_{11}, a_{12}, a_{13}\} \quad \text{CF} = \{0.4, 0.3, 0.2\}$$

$$R_2: \text{IF } E_4 \wedge (E_5 \wedge E_6) \text{ THEN } A_2 = \{a_{21}\} \quad \text{CF}_2 = \{0.7\}$$

$$R_3: \text{IF } A_1 \text{ THEN } A = \{a_1, a_2\} \quad \text{CF}_3 = \{0.5, 0.4\}$$

$$R_4: \text{IF } A_2 \text{ THEN } A = \{a_1, a_2\} \quad \text{CF}_4 = \{0.4, 0.4\}$$

这些规则形成如图 5.3 所示的推理网络, 原始数

据的概率在系统中已经给出:

$$f(E_1) = 0.5, f(E_2) = 0.9, f(E_3) = 0.7, f(E_4) = 0.9,$$

$$f(E_5) = 0.7, f(E_6) = 0.8$$

假设 $|\Omega| = 10$, 现在要求出 A 的确定性 $f(A)$ 。

解: 第一步, 求 A_1 的确定性。

$$f(E_1 \vee (E_2 \wedge E_3)) = \max\{0.5, \min(0.9, 0.7)\} = 0.7$$

$$m_1(\{a_{11}\}, \{a_{12}\}, \{a_{13}\}) = (0.7 \times 0.4, 0.7 \times 0.3, 0.7 \times 0.2) = (0.28, 0.21, 0.14)$$

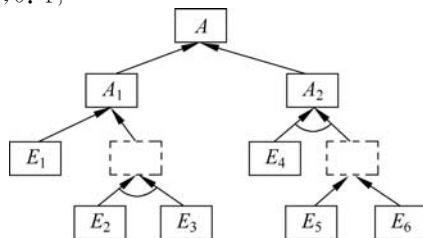


图 5.3 例 5.12 推理网络

$$\text{Bel}(A_1) = m_1(\{a_{11}\}) + m_1(\{a_{12}\}) + m_1(\{a_{13}\}) = 0.28 + 0.21 + 0.14 = 0.63$$

$$\text{Pl}(A_1) = 1 - \text{Bel}(\neg A_1) = 1 - 0 = 1$$

$$f(A_1) = \text{Bel}(A_1) + \frac{|A_1|}{\Omega} \times (\text{Pl}(A_1) - \text{Bel}(A_1)) = 0.63 + \frac{3}{10} \times (1 - 0.63) = 0.74$$

第二步,求 A_2 的确定性。

$$f(E_4 \wedge (E_5 \vee E_6)) = \min\{0.9, \max(0.7, 0.8)\} = 0.8$$

$$m_2(\{a_{21}\}) = 0.8 \times 0.7 = 0.56$$

$$\text{Bel}(A_2) = m_2(\{a_{21}\}) = 0.56$$

$$\text{Pl}(A_2) = 1 - \text{Bel}(\neg A_2) = 1 - 0 = 1$$

$$f(A_2) = \text{Bel}(A_2) + \frac{|A_2|}{\Omega} \times (\text{Pl}(A_2) - \text{Bel}(A_2)) = 0.56 + \frac{1}{10} \times (1 - 0.56) = 0.60$$

第三步,求 A 的确定性。

根据 R_3 和 R_4 ,有:

$$m_3(\{a_1\}, \{a_2\}) = (0.74 \times 0.5, 0.74 \times 0.4) = (0.37, 0.296)$$

$$m_4(\{a_1\}, \{a_2\}) = (0.6 \times 0.4, 0.6 \times 0.4) = (0.24, 0.24)$$

$$m_3(\Omega) = 1 - (m_3(\{a_1\}) + m_3(\{a_2\})) = 1 - (0.37 + 0.296) = 0.334$$

$$m_4(\Omega) = 1 - (m_4(\{a_1\}) + m_4(\{a_2\})) = 1 - (0.24 + 0.24) = 0.52$$

由正交和公式得到:

$$\begin{aligned} K &= 1 - \sum_{x \cap y = \emptyset} m_3(x)m_4(y) \\ &= 1 - (m_3(\{a_1\})m_4(\{a_2\}) + m_3(\{a_2\})m_4(\{a_1\})) \\ &= 1 - (0.37 \times 0.24 + 0.296 \times 0.24) \\ &= 0.84 \end{aligned}$$

则:

$$\begin{aligned} m(\{a_1\}) &= K^{-1} \sum_{x \cap y = \{a_1\}} m_1(x)m_2(y) \\ &= \frac{1}{0.84} \times (m_3(\Omega)m_4(\{a_1\}) + m_3(\{a_1\})m_4(\Omega) + m_3(\{a_1\})m_4(\{a_1\})) \\ &= \frac{1}{0.84} \times (0.334 \times 0.24 + 0.37 \times 0.52 + 0.37 \times 0.24) = 0.43 \end{aligned}$$

$$\begin{aligned} m(\{a_2\}) &= K^{-1} \sum_{x \cap y = \{a_2\}} m_1(x)m_2(y) \\ &= \frac{1}{0.84} \times (m_3(\Omega)m_4(\{a_2\}) + m_3(\{a_2\})m_4(\Omega) + m_3(\{a_2\})m_4(\{a_2\})) \\ &= \frac{1}{0.84} \times (0.334 \times 0.24 + 0.296 \times 0.52 + 0.296 \times 0.24) \\ &= 0.36 \end{aligned}$$

于是:

$$\text{Bel}(A) = m(\{a_1\}) + m(\{a_2\}) = 0.43 + 0.36 = 0.79$$

$$Pl(A) = 1 - Bel(\neg A) = 1 - 0 = 1$$

$$f(A) = Bel(A) + \frac{|A|}{\Omega} \times (Pl(A) - Bel(A)) = 0.79 + \frac{2}{10} \times (1 - 0.79) = 0.832$$

证据理论的优点在于能够满足比概率论更弱的公理系统,可以区分不知道和不确定的情况,可以依赖证据的积累,不断缩小假设的集合。

证据理论最早是作为经典概率理论的扩展而引入的,所以受到很多的批评;在证据理论中,证据的独立性不易得到保证,基本概率分配函数要求给的值太多,计算传递关系复杂,随着诊断问题可能答案的增加,证据理论的计算呈指数级增长,传递关系复杂,比较难以实现。

5.7 模糊推理

模糊推理的理论基础是模糊集理论以及在此基础上发展起来的模糊逻辑,起源于1965年美国 California 大学的扎德(L. A. Zadeh)在 *Information and Control* 上发表的论文 *Fuzzy Sets*。模糊逻辑所处理的事物自身是模糊的,概念本身没有明确的外延,一个对象是否符合这个概念难以明确地确定,模糊推理是对这种不确定性,即模糊性的表示与处理。

模糊推理是利用模糊性知识进行的一种不确定性推理,与前面几节讨论的不确定性推理有着实质性的区别。不确定性推理的理论基础是概率论,所研究的事件本身有明确且确定的含义,只是由于发生的条件不充分,使得在条件与事件之间不能出现确定的因果关系,从而在事件的出现与否上表现出不确定性。

在人工智能的应用领域中,知识及信息的不确定性大多是由模糊性引起的,这就使得对模糊推理的研究显得尤为重要。本节以模糊数学为基础,讨论模糊假言推理。

5.7.1 模糊数学的基本知识

1. 模糊集合

1) 隶属度

集合元素对集合的隶属程度称为隶属度,用 μ 表示。设 A 是论域 U 上的模糊集合, U 中完全属于 A 的元素,其 μ 值为1,完全不属于 A 的元素其值为0。对于 $(0,1)$ 内的 μ 值,其值越大,隶属程度越高,当 μ 值为1时,就是经典集合的“属于”,当 μ 值为0时,就是经典集合的“不属于”。

模糊集合用“隶属度/元素”的形式来记,例如:

$$A = \mu_1/x_1 + \mu_2/x_2 + \cdots + \mu_n/x_n$$

注意: 这里的“+”号并不是求和,“/”号也不是求商,仅仅是一种记法,是模糊数学创始人扎德给出的记法。当某一项的 μ 值为0时,可以省略不写。由于这种记法中的“+”号的原意是求和,扎德又用记号

$$A = \int_{u \in U} \mu_A(u)/u$$

作为模糊集合 A 的一般表示形式。当然,这里的积分符号也不是求和,只是一种记法。

模糊集合中,论域的概念十分重要,论域是一个经典集合。任何一个模糊集合都是建立在一个论域之上的,模糊集合中的元素 x_i 取自其论域,因此空谈模糊集合是没有意义的。谈到某一模糊集合,必须声明它是哪一论域上的模糊集合。

2) 模糊集合相等

两个模糊集合相等,当且仅当它们的隶属函数在论域 U 上恒等,即 $A=B$,当且仅当 $\forall x \in U, \mu_A(x) = \mu_B(x)$ 。

3) 模糊集合的包含

模糊集合 A 包含于模糊集合 B 中,当且仅当对于论域 U 上所有元素 x ,恒有 $\mu_A(x) \leq \mu_B(x)$ 。

4) 模糊集合的并、交、补

$$\mu_{(A \cup B)}(x) = \max\{\mu_A(x), \mu_B(x)\}, \quad \forall x \in U$$

$$\mu_{(A \cap B)}(x) = \min\{\mu_A(x), \mu_B(x)\}, \quad \forall x \in U$$

$$\mu_{\neg A}(x) = 1 - \mu_A(x), \quad \forall x \in U$$

5) 模糊集合的积

设 A, B 分别是论域 U 和论域 V 上的模糊集合,那么:

$$A \times B = \int_{U \times V} (\mu_A(u_i) \wedge \mu_B(v_j)) / (u_i, v_j)$$

特别地,当 A 或 B 有一个是论域时,上面表达式可以简化为

$$A \times V = \int_{U \times V} \mu_A(u_i) / (u_i, v_j)$$

$$U \times B = \int_{U \times V} \mu_B(v_j) / (u_i, v_j)$$

2. 模糊关系及其运算

1) 模糊关系

设 U, V 是论域,从 U 到 V 上的模糊关系 R 是指 $U \times V$ 上的一个模糊集合,由隶属函数 $\mu_R(x)$ 刻画, $\mu_R(x, y)$ 代表有序对 $\langle x, y \rangle$ 具有关系 R 的程度。

例 5.13 设某地区的身高论域 $U = \{150, 160, 170, 180\}$ (单位: cm), 体重论域 $V = \{45, 55, 65, 75\}$ (单位: kg)。身高和体重两个集合的元素之间没有确定的关系,只有一定程度的关联。 $\mu(x, y)$ 表示 x 和 y 的关联程度,如表 5.1 所示。

表 5.1 身高与体重的模糊关系

x	y			
	45	55	65	75
150	1	0.2	0.1	0
160	0.2	1	0.8	0.1
170	0.1	0.8	1	0.2
180	0	0.1	0.2	1

模糊关系 R 通常用矩阵表示,将上面表格表示转化成模糊矩阵表示,即:

$$R = \begin{bmatrix} 1 & 0.2 & 0.1 & 0 \\ 0.2 & 1 & 0.8 & 0.1 \\ 0.1 & 0.8 & 1 & 0.2 \\ 1 & 0.1 & 0.2 & 1 \end{bmatrix}$$

2) 模糊关系的合成

设 R 是 $U \times V$ 上的模糊关系, S 是 $V \times W$ 上的模糊关系,则 R, S 的复合是 $U \times W$ 上的模糊关系 T ,记为

$$T = R \circ S$$

其隶属函数为

$$T(x, y) = R(x, y) \circ S(y, z) = \sup_{y \in V} \min(\mu_R(x, y), \mu_S(y, z)) = \bigcup_{y \in V} (\mu_R(x, y) \wedge \mu_S(y, z))$$

其中, $\sup_{y \in V}$ 表示对所有 $y \in V$ 取最小上界。

当论域为有限集时,模糊关系的合成运算可转化为模糊关系矩阵的乘法运算,该乘法运算类似于普通矩阵的乘法运算,区别是:将普通矩阵乘法中的“ $\times$ ”换为取极小值“ $\wedge$ ”,将普通矩阵乘法中的“ $+$ ”换为取极大值“ $\vee$ ”。

设 R 为 $n \times m$ 阶矩阵, S 为 $m \times p$ 阶矩阵,则 $R \circ S = T$ 是 $n \times p$ 阶矩阵, T 的元素 T_{ij} 计算如下。

$$T_{ij} = \bigcup_{k=1}^m (r_{ik} \wedge s_{kj}), \quad i=1,2,\dots,n; j=1,2,\dots,p$$

两个模糊关系能够进行合成运算的条件:第一个模糊关系矩阵的列数=第二个模糊关系矩阵的行数。这与两个普通矩阵的乘法运算的条件相同。

例 5.14 设有如下两个模糊关系:

$$R_1 = \begin{bmatrix} 0.5 & 0.6 & 0.3 \\ 0.7 & 0.4 & 1 \\ 0 & 0.8 & 0 \\ 1 & 0.2 & 0.9 \end{bmatrix}, \quad R_2 = \begin{bmatrix} 0.2 & 1 \\ 0.8 & 0.4 \\ 0.5 & 0.3 \end{bmatrix}$$

求: $R_1 \circ R_2$ 。

解: R_1 是 4×3 模糊关系矩阵, R_2 是 3×2 模糊关系矩阵,因此 $R_1 \circ R_2$ 是 4×2 的模糊关系矩阵,令 $T = R_1 \circ R_2$,则:

$$T(1,1) = (0.5 \wedge 0.2) \vee (0.6 \wedge 0.8) \vee (0.3 \wedge 0.5) = 0.6$$

$$T(1,2) = (0.5 \wedge 1) \vee (0.6 \wedge 0.4) \vee (0.3 \wedge 0.3) = 0.5$$

$$T(2,1) = (0.7 \wedge 0.2) \vee (0.4 \wedge 0.8) \vee (1 \wedge 0.5) = 0.5$$

$$T(2,2) = (0.7 \wedge 1) \vee (0.4 \wedge 0.4) \vee (1 \wedge 0.3) = 0.7$$

$$T(3,1) = (0 \wedge 0.2) \vee (0.8 \wedge 0.8) \vee (0 \wedge 0.5) = 0.8$$

$$T(3,2) = (0 \wedge 1) \vee (0.8 \wedge 0.4) \vee (0 \wedge 0.3) = 0.4$$

$$T(4,1) = (1 \wedge 0.2) \vee (0.2 \wedge 0.8) \vee (0.9 \wedge 0.5) = 0.5$$

$$T(4,2) = (1 \wedge 1) \vee (0.2 \wedge 0.4) \vee (0.9 \wedge 0.3) = 1$$

所以,得到模糊关系矩阵为

$$T = \begin{bmatrix} 0.6 & 0.5 \\ 0.5 & 0.7 \\ 0.8 & 0.4 \\ 0.5 & 1 \end{bmatrix}$$

5.7.2 模糊假言推理

1. 模糊规则的表达

模糊产生式规则的一般形式为

$$\text{IF } E \text{ THEN } R(\text{CF}, \lambda)$$

其中, E 是用模糊命题表示的模糊条件,既可以由单个模糊命题表示的简单条件,也可以是由多个模糊命题构成的复合条件; R 是用模糊命题表示的模糊结论; CF 是该产生式规则所表示的知识的可信度因子,既可以是一个确定的实数,又可以是一个模糊数或模糊语言值, CF 的值由领域专家在给出规则时同时给出; λ 是阈值,用于指出相应知识在什么情况下可被应用。

例如,各种形式的规则:

$$(1) \text{ IF } x \text{ is } A \text{ THEN } y \text{ is } B(\lambda)$$

$$(2) \text{ IF } x \text{ is } A \text{ THEN } y \text{ is } B(\text{CF}, \lambda)$$

$$(3) \text{ IF } x_1 \text{ is } A_1 \text{ AND } x_2 \text{ is } A_2 \text{ THEN } y \text{ is } B(\lambda)$$

$$(4) \text{ IF } x_1 \text{ is } A_1 \text{ AND } x_2 \text{ is } A_2 \text{ AND } x_3 \text{ is } A_3 \text{ THEN } y \text{ is } B(\text{CF}, \lambda)$$

推理中所用的证据也是用模糊命题表示的,一般形式为

$$x \text{ is } A' \text{ 或 } x \text{ is } A'(\text{CF})$$

2. 证据的模糊匹配

在模糊推理中,规则的前提条件中的 A 与证据中的 A' 不一定完全相同,因此在决定选用哪条规则进行推理时必须首先考虑哪条规则的 A 可与 A' 近似匹配的问题,即它们的相似程度是否大于某个预先设定的阈值。例如,设有如下规则及证据:

$$\text{IF } x \text{ is 小 THEN } y \text{ is 大}(0.6)$$

$$x \text{ is 较小}$$

那么,是否有“ y is 大”这个结论呢? 这决定于 λ 值,若“ x is 较小”与“ x is 小”的接近程度大于等于 λ 值,则有“ y is 大”的模糊结论(其模糊值需计算),否则没有这一结论。

如何计算接近程度? 有多种方法,这里举其中一种——贴进度。

设 A 、 B 分别是论域 $U = \{u_1, u_2, \dots, u_n\}$ 上的模糊集合,它们的贴进度定义为

$$(A, B) = \frac{1}{2} [A \cdot B + (1 - A \odot B)]$$

其中: $A \cdot B = \bigvee_U (\mu_A(u_i) \wedge \mu_B(u_i))$, $A \odot B = \bigwedge_U (\mu_A(u_i) \vee \mu_B(u_i))$ 。“ $\wedge$ ”表示取极小,“ $\vee$ ”表示取极大。

例 5.15 设 $U = \{a, b, c, d, e\}$, 有

$$A = 0.6/a + 0.8/b + 1/c + 0.8/d + 0.6/e + 0.4/f$$

$$B = 0.4/a + 0.6/b + 0.8/c + 1/d + 0.8/e + 0.6/f$$

求: (A, B) 。

解:

$$A \cdot B = 0.4 \vee 0.6 \vee 0.8 \vee 0.8 \vee 0.6 \vee 0.4 = 0.8$$

$$A \odot B = 0.6 \wedge 0.8 \wedge 1 \wedge 1 \wedge 0.8 \wedge 0.6 = 0.6$$

那么

$$(A, B) = \frac{1}{2} [0.8 + (1 - 0.6)] = 0.6$$

3. 简单模糊推理

简单模糊推理是指规则的前提 E 是单一条件, 结论 R 不含 CF, 即

$$\text{IF } x \text{ is } A \text{ THEN } y \text{ is } B(\lambda)$$

首先构造 A 、 B 之间的模糊关系 R , 然后通过 R 与前提的合成求出结论。如果已知证据是

$$x \text{ is } A'$$

且 $(A, A') \geq \lambda$, 那么有结论

$$y \text{ is } B'$$

其中, $B' = A' \circ R$ 。

所以, 在这种推理方法中, 关键是如何构造模糊关系 R 。构造模糊关系有多种方法, 这里只介绍扎德方法。扎德提出两种方法——条件命题的极大极小规则和条件命题的算术规则, 得到的模糊关系分别记为 R_m 和 R_a 。

设 A 、 B 分别表示为

$$A = \int_U \mu_A(u)/u$$

$$B = \int_V \mu_B(v)/v$$

则

$$R_m = (A \times B) \cup (\neg A \times V) = \int_{U \times V} (\mu_A(u) \wedge \mu_B(v)) \vee (1 - \mu_A(u)) / (u, v)$$

$$R_a = (\neg A \times V) \oplus (U \times B) = \int_{U \times V} 1 \wedge (1 - \mu_A(u) + \mu_B(v)) / (u, v)$$

其中, $\oplus$ 表示界和, 定义为

$$A \oplus B = \min\{1, \mu_A(u) + \mu_B(v)\}$$

对于模糊假言推理, 已知证据为“ $x \text{ is } A'$ ”, 且 $(A, A') \geq \lambda$, 则由 R_m 和 R_a 求得 B'_m 和 B'_a , 分别为

$$B'_m = A' \circ R_m = A' \circ [(A \times B) \cup (\neg A \times V)]$$

$$B'_a = A' \circ R_a = A' \circ [(\neg A \times V) \cup (U \times B)]$$

它们的隶属函数分别为

$$\mu_{B'_m}(v) = \bigvee_{u \in U} \{\mu_{A'}(u) \wedge [(\mu_A(u) \wedge \mu_B(v)) \vee (1 - \mu_A(u))]\}$$

$$\mu_{B'_a}(v) = \bigvee_{u \in U} \{\mu_{A'}(u) \wedge [1 \wedge (1 - \mu_A(u) + \mu_B(v))]\}$$

例 5.16 设 $U=V=\{1,2,3,4,5\}$, 有

$$A = 1/1 + 0.5/2$$

$$B = 0.4/3 + 0.6/4 + 1/5$$

模糊规则为

$$\text{IF } x \text{ is } A \text{ THEN } y \text{ is } B(\lambda)$$

证据为

$$x \text{ is } A'$$

其中, A' 的模糊集为 $A' = 1/1 + 0.4/2 + 0.2/3$, 且有 $(A, A') \geq \lambda$, 求 B'_m, B'_a 。

解: 先求 R_m, R_a 。由前面 R_m 和 R_a 定义, 知 $R_m(i, j)$ 与 $R_a(i, j)$ 分别为

$$R_m(i, j) = (\mu_A(u_i) \wedge \mu_B(v_j)) \vee (1 - \mu_A(u_i))$$

$$R_a(i, j) = 1 \wedge (1 - \mu_A(u_i) + \mu_B(v_j))$$

$R_m(i, j)$ 与 $R_a(i, j)$ 分别是 R_m 和 R_a 的第 i 行第 j 列元素。例如:

$$R_m(1, 3) = (\mu_A(u_1) \wedge \mu_B(v_3)) \vee (1 - \mu_A(u_1)) = (1 \wedge 0.4) \vee (1 - 1) = 0.4$$

$$R_a(1, 3) = 1 \wedge (1 - \mu_A(u_1) + \mu_B(v_3)) = 1 \wedge (1 - 1 + 0.4) = 0.4$$

由此求出 R_m, R_a , 即

$$R_m = \begin{bmatrix} 0 & 0 & 0.4 & 0.6 & 1 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}, \quad R_a = \begin{bmatrix} 0 & 0 & 0.4 & 0.6 & 1 \\ 0.5 & 0.5 & 0.9 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

下面求 B'_m 和 B'_a 。

$$B'_m = A' \circ R_m = \{1, 0.4, 0.2, 0, 0\} \circ \begin{bmatrix} 0 & 0 & 0.4 & 0.6 & 1 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} = \{0.4, 0.4, 0.4, 0.6, 1\}$$

$$B'_a = A' \circ R_a = \{1, 0.4, 0.2, 0, 0\} \circ \begin{bmatrix} 0 & 0 & 0.4 & 0.6 & 1 \\ 0.5 & 0.5 & 0.9 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} = \{0.4, 0.4, 0.4, 0.6, 1\}$$

这里 $B'_m = B'_a$ 只是一个巧合, 一般来说它们不一定相同。

5.8 小结

本章首先讨论了不确定性推理的基本概念、不确定性研究的基本问题和主要研究方法。“不确定性”是针对已知事实和推理中所用到的知识而言的, 应用这种不确定的事实和知识的推理称为不确定性推理。

目前, 关于不确定性处理方法的研究主要沿着两条路线发展。一是在推理级扩展确定

性推理,建立各种不确定性推理的模型,又分为数值方法和非数值方法。本章主要讨论的是数值方法,如主观 Bayes 方法、可信度方法、证据理论、模糊方法等。二是在控制级上处理不确定性,称为控制方法。对于处理不确定的最优方法,现在还没有一个统一的意见。

主观 Bayes 方法通过使用专家的主观概率,避免了所需的大量统计计算工作。主观 Bayes 方法讨论了信任与概率的关系、似然性问题,介绍了主观 Bayes 方法知识表示和推理方法。

可信度方法比较简单、直观,易于掌握和使用,并且已成功地应用于如 MYCIN 这样的推理链较短、概率计算精度要求不高的专家系统中。但是当推理长度较长时,由可信度的不精确估计而产生的累计误差会很大,所以它不适合长推理链的情况。

证据理论是用集合表示命题的一种处理不确定性的理论,引入信任函数而非概率来度量不确定性,并引入似然函数来处理不知道所引起的不确定性问题,只需要满足比概率论更弱的公理系统。证据理论基础严密,专门针对专家系统,是一种很有吸引力的不确定性推理模型。但如何把它普遍应用于专家系统,目前还没有一个统一的意见。

模糊推理建立在传统的假言推理之上,涉及两方面:一是前提是否匹配。传统的假言推理要求严格的匹配,而模糊假言推理是模糊匹配,引入了贴近度的概念,只有前提的模糊集与证据的模糊集的贴近度超过专家给定的阈值,才认为是匹配的;二是当前提与证据模糊匹配后,结论的模糊性如何计算。本章的方法是按照扎德给出的条件命题的极大极小规则和条件命题的算术规则,得到模糊关系 R_m 和 R_a ,然后经过模糊关系的合成,计算结论的模糊性。

习题

- 5.1 不确定推理的概念是什么?为什么要采用不确定推理?
- 5.2 不确定推理中需要解决的基本问题是什么?
- 5.3 主观 Bayes 方法的优点是什么?有什么问题?试说明 LS 和 LN 的意义。
- 5.4 为什么要在 MYCIN 中提出可信度方法?可信度方法还有什么问题?
- 5.5 何谓可信度?说明规则强度 $CF(H, E)$ 的含义。

5.6 设有三个独立的结论 H_1, H_2, H_3 及两个独立的证据 E_1, E_2 , 它们的先验概率和条件概率分别为

$$P(H_1) = 0.4, \quad P(H_2) = 0.3, \quad P(H_3) = 0.3$$

$$P(E_1 | H_1) = 0.5, \quad P(E_1 | H_2) = 0.3, \quad P(E_1 | H_3) = 0.5$$

$$P(E_2 | H_1) = 0.7, \quad P(E_2 | H_2) = 0.9, \quad P(E_2 | H_3) = 0.1$$

利用概率方法求出:

(1) 当只有证据 E_1 出现时, $P(H_1 | E_1)$ 、 $P(H_2 | E_1)$ 及 $P(H_3 | E_1)$ 的值;并说明 E_1 的出现对 H_1, H_2, H_3 的影响。

(2) 当 E_1 和 E_2 同时出现时, $P(H_1 | E_1, E_2)$ 、 $P(H_2 | E_1, E_2)$ 及 $P(H_3 | E_1, E_2)$ 的值;并说明 E_1 和 E_2 同时出现对 H_1, H_2, H_3 的影响。

5.7 设有如下规则:

$$R_1: \text{IF } E_1 \text{ THEN } (20, 0.01) \ H_1(0.06)$$

$R_2: \text{IF } E_2 \text{ THEN } (10, 0.1) \quad H_2(0.05)$

$R_3: \text{IF } E_3 \text{ THEN } (1, 1) \quad H_3(0.4)$

求: 当证据 E_1, E_2, E_3 存在时, $P(H_i | E_i)$ 的值各是多少?

5.8 设有如下规则:

$R_1: \text{IF } E_1 \text{ THEN } (20, 0.1) \quad H$

$R_2: \text{IF } E_2 \text{ THEN } (300, 0.1) \quad H$

已知: 证据 E_1 和 E_2 必然发生, 并且 $P(H) = 0.03$, 求: H 的后验概率。

5.9 设有规则:

$R_1: \text{IF } E_1 \text{ THEN } (65, 0.01) \quad H$

$R_2: \text{IF } E_2 \text{ THEN } (300, 0.0001) \quad H$

已知: $P(E_1 | S_1) = 0.5, P(E_2 | S_2) = 0.2, P(E_1) = 0.1, P(E_2) = 0.03, P(H) = 0.01$ 。

求: $P(H | S_1, S_2)$

5.10 设有如下规则:

$R_1: \text{IF } E_1 \text{ THEN } H(0.8)$

$R_2: \text{IF } E_2 \text{ THEN } H(0.6)$

$R_3: \text{IF } E_3 \text{ THEN } H(-0.5)$

$R_4: \text{IF } E_4 \text{ AND } (E_5 \text{ OR } E_6) \text{ THEN } E_1(0.7)$

$R_5: \text{IF } E_5 \text{ AND } E_8 \text{ THEN } E_3(0.7)$

且已知: $CF(E_2) = 0.8, CF(E_4) = 0.5, CF(E_5) = 0.6, CF(E_6) = 0.7, CF(E_7) = 0.6, CF(E_8) = 0.9$ 。

求: H 的综合可信度 $CF(H)$ 。

5.11 请说明证据理论中概率分配函数、信任函数、似然函数及类概率函数的含义。

5.12 设 $\Omega = \{\text{红}, \text{黄}, \text{绿}\}, A = \{\text{红}, \text{黄}\}$, 有如下的概率分配函数:

$m(\emptyset, \{\text{红}\}, \{\text{黄}\}, \{\text{绿}\}, \{\text{红}, \text{黄}, \text{绿}\}) = (0, 0.6, 0.2, 0.1, 0.1)$

求: $m(\Omega), \text{Bel}(A), \text{Pl}(A), f(A)$ 的值。

5.13 已知 $f(E_1) = 0.6, f(E_2) = 0.7, |\Omega| = 20, E_1 \wedge E_2 \rightarrow H, H = \{h_1, h_2\}, (c_1, c_2) = (0.5, 0.3)$ 。计算 $f(H)$ 。

5.14 设有如下规则:

$R_1: \text{IF } E_1 \text{ AND } E_2 \text{ THEN } A = \{a_1, a_2\} \quad CF = \{0.3, 0.5\}$

$R_2: \text{IF } E_3 \text{ AND } (E_4 \text{ OR } E_5) \text{ THEN } B = \{b_1\} \quad CF = \{0.7\}$

$R_3: \text{IF } A \text{ THEN } H = \{h_1, h_2, h_3\} \quad CF = \{0.1, 0.5, 0.3\}$

$R_4: \text{IF } B \text{ THEN } H = \{h_1, h_2, h_3\} \quad CF = \{0.4, 0.2, 0.1\}$

已知用户对初始证据给出的确定性为 $f(E_1) = 0.8, f(E_2) = 0.6, f(E_3) = 0.9, f(E_4) = 0.5, f(E_5) = 0.7$ 。并假定 Ω 中的元素个数 $|\Omega| = 10$ 。求: $f(H)$ 。

5.15 设有如下两个模糊关系：

$$\mathbf{R}_1 = \begin{bmatrix} 0.3 & 0.7 & 0.2 \\ 1 & 0 & 0.4 \\ 0 & 0.5 & 1 \\ 0.6 & 0.7 & 0.8 \end{bmatrix}, \quad \mathbf{R}_2 = \begin{bmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \\ 0.9 & 0.1 \end{bmatrix}$$

求： $\mathbf{R}_1 \circ \mathbf{R}_2$ 。

5.16 设 $U=V=\{1,2,3,4\}$, 有

$$A = 0.8/1 + 0.5/2 + 0.2/3$$

$$B = 0.3/2 + 0.7/3 + 0.9/4$$

模糊规则为

$$\text{IF } x \text{ is } A \text{ THEN } y \text{ is } B(\lambda)$$

证据为

$$x \text{ is } A'$$

其中, A' 的模糊集为 $A' = 0.8/1 + 0.5/2 + 0.2/3$, 且有 $(A, A') \geq \lambda$ 。求： B'_m, B'_a 。