《现代应用数学方法》丛书(5)

科学出版社

21

数学天元基金

人工智能中的概率统计方法

张尧庭　杜劲松　著

《现代应用数学方法》丛书(5)

人工智能中的概率统计方法

张尧庭　杜劲松　著

科学出版社

1998

内 容 简 介

本书系统论述了人工智能领域中在推理方面的概率统计方法，其中包括作者的研究工作．主要内容包括：似然比推理、信任函数，规则与信任函数，多元信任函数与知识的表示和获取，似然比与统计推断．本书提出并处理了一些概率论和人工智能中不确定性推理的基础问题，对基础理论研究，开发各种类型的专家系统都有推动作用．

本书可供人工智能专家系统的工程技术人员、研究人员，概率统计专家阅读．

图书在版编目(CIP)数据

人工智能中的概率统计方法/张尧庭，杜劲松 著．—北京：科学出版社，1998

（现代应用数学方法丛书/方开泰主编）

ISBN 7-03-006154-3

Ⅰ．人… Ⅱ．①张… ②杜… Ⅲ．数理统计-应用-人工智能 Ⅳ．TP18

中国版本图书馆 CIP 数据核字（97）第 14504 号

科 学 出 版 社 出 版
北京东黄城根北街 16 号
邮政编码：100717

中国科学院印刷厂印刷

新华书店北京发行所发行　各地新华书店经售

*

1998 年 3 月第 一 版　开本：850×1168 1/32
1998 年 3 月第一次印刷　印张：3 3/8
印数：1—3 900　字数：81 000

定 价：7.00 元

总　序

应用数学的发展与自然科学和社会科学有着广泛的交叉和渗透.一方面,它为形形色色的物理、化学、生物、社会等现象提供描述和分析的数学工具.另一方面,这些实际问题的解决又为数学学科的发展提供了动力和永不枯竭的源泉.许多成功的应用数学方法,如解非线性方程的牛顿-高斯法、曲线拟合的最小二乘法、线性规划的单纯形法等,成了当今应用数学工作者手中不可缺少的工具.它们之所以有如此强大的生命力,原因在于方法本身有坚实的理论基础,同时又有鲜明的应用背景,能用于不同的领域.因此,成功的应用数学方法是理论联系实际的桥梁和纽带.

我国的数学要达到世界先进水平,要对人类有所贡献,一个重要的方面是要有一批独创的应用数学方法.《现代应用数学方法》丛书的出版,希望能为鼓励和促进我国的数学工作者创造或介绍更多的现代应用数学方法增加一个舞台.

这套丛书的宗旨是介绍现代应用数学方法. 这些方法应该是目前世界上最先进的,或者是我国独创的,或者是国外已经普遍使用但国内知之甚少的方法. 丛书着重阐明所介绍方法的应用背景和思想,避开深奥的数学论证,力求深入浅出、图文并茂. 有数值及应用性的例子,使读者易于理解和使用. 丛书要求短小精练,突出新的方法,不求齐全,一般 10 万字左右. 书末所附的文献将指出方法的理论背景以及最近的进展,以便读者进一步深入研究.

本丛书的出版得到国家"天元"项目的资助,得到科学出版社的大力支持,是全体编委努力的成果.我们要特别感谢许多作者在百忙中为丛书撰写文稿,付出了辛勤的劳动.我们希望这套丛书的

出版对我国应用数学的发展起到促进作用，衷心地希望丛书成为广大读者的良师益友.

胡国定(南开大学)
方开泰(中国科学院应用数学所)

前　　言

不确定性推理是人工智能中必须处理的一个专题. 这几年来，在戴汝为院士主持的“攀登计划”项目“思维与智能模拟”这一课题中，我们参加了这一方面的研究，本书可以看成是反映我们研究工作的一本小册子.

本书主要内容有两个：似然比推理和信任函数理论. 似然比推理是我们的工作，本意是企图克服概率方法的一些困难又与概率不是相距很远；信任函数在十几年前就已成形了，而且发展很快，它也是概率的一种延拓，本书不是介绍这一理论、方法的进展，而是重点介绍我们自己在这一方面的工作，提出一些新的看法和处理，为此不得不介绍一些信任函数的基本知识. 应该说，这两个内容是有联系的，书中说明了这一联系.

本书涉及到概率论、统计、数理逻辑、专家系统中的一些知识，我们认为读者已经熟悉有关知识，不作一一介绍. 我们使用的概念、符号都是通用的，在教科书上常见的，因此也不一一说明. 为了说明专家系统中常见的方法，在第五章 § 5.5 介绍了有关材料作为参考，我们用一个 * 表示可以不看.

本书的重点是阐明一些观点、引入一些概念，用来分析问题. 全书由我执笔，书中一些结果是杜劲松攻读博士时的工作. 本书的出版希望在概率统计界和计算技术专家系统方面引起兴趣，促进这两方面人士进一步合作. 也希望有关专业的人士感到有意思. 这一工作实际上还是迈了第一步，诚恳欢迎读者提出批评和建议.

本书的成果是在攀登计划项目支持下获得的；我还承担了“应用统计”的课题，所以也是应用统计的一个重要方面；本书的出版得到了国家“天元”基金的资助. 所有这些，我衷心地表示感谢. 全

书得到李国英同志的认真审阅，她提出了不少改进和建议，我们也都采纳了，并对她表示深深的感谢.

张尧庭

1997 年 3 月

目　　录

第一章 绪　论

§1.1 信息的不确定性

人工智能、专家系统的研究，使人们发现在问题求解过程中会遇到各种不确定的信息. 其实在社会活动、经济活动中，每人必须面对不确定的信息作出自己的决策. 在这一方面，人们积累了许多经验，例如直觉就是这一类经验的某种反映. 一般说来，信息的不确定性可能有以下的几种原因：

（1）信息不是完全可靠的. 因为现实世界中事件之间的因果性、关联性往往过于复杂，要使某一结论成立，需要很多前提. 这样专家自己对给出的规则缺乏自信，只是表达了一种对规则的信念，这就需要对信念给出一种测度，可以度量它的程度，并由此导出相应的规则. 这就引出了不确定的关系所反映的不确定的信息；

（2）信息本身不是完备的. 有时全局的信息是无法得到的，如天气预报，我们能有的只是局部的（历史的）信息. 这种不完备性也导致不确定性.

（3）信息是不精确的. 这种不精确往往来源于语言、文字表达的内容是含糊的，如“年青的”、“方形的”等概念. 人们在日常生活中就是用这些信息来作出判断，例如老朋友多年不见，尽管有容貌上的变化，但见面还能认出来，这也利用了不精确的信息.

（4）信息是矛盾的. 例如几个医生对同一个病人各自诊断的结论不同，就会提供矛盾的信息.

如何描述、反映这些信息的不确定性的程度？如何总结人们综合这些不确定信息的经验？这些就成了研究的课题. 概率论和数理统计实际上已给出了处理这一类信息的方法. 随着人工智能、专家系统的技术发展，发现传统的方法很难处理有些问题，

70 年代以来，有一些新的数值方法和非数值方法不断涌现，其目的都是想给出处理这类信息一个系统和合理的方法，例如从模糊数学的框架来处理不确定性的推理，从引入区间概率来处理，或从信任函数出发来考虑，这些可以从本书最后所附的参考文献中看到.

本书主要是介绍近年来我们自己的工作，想把概率方法加以适当改变，可以适应在人工智能、决策、不确定性推理方面的需要，主要是似然比推理和信任函数这两方面的工作，这两者是有联系而又有区别的. 在与别的方法作比较时，有时会对别的方法稍作介绍，作一简略的说明.

§1.2 概率方法面临的困难

概率论已有很长的历史了，它成功地处理了许多与不确定性有关的问题，有丰富的理论和系统的方法. 但是，面对不确定性推理中遇到的问题，确实有困难，这一节先谈谈这一方面的看法.

为什么概率方法在不确定信息的处理中会不适应呢? 这还要从概率论本身谈起.

1. 什么是概率? 是主观的还是客观的? 这一点历来就有不同的观点，从而形成不同的学派和分支. 长期以来，一种认为概率是频率的稳定性基础，它是客观的，不能随人的认识而改变的；一种认为概率是人的信念程度，是能随着信息的变化而改变的. 应该说这两类概念是不同的，但它们是否有一些共同的规则呢? 而且客观的概率规律应该说是人们形成信念的一个重要途径. 承认这两者有区别，又有一定的联系，这是一个正确的考虑. 实际上，统计方法就是从样本获得了信息改变了人们对事物认识的信念，是把客观的信息转化为主观信念的一个典型的例子.

信息论是用概率作为工具描述了信息的度量、讨论了不确定性，但信息的改变如何具体影响到概率的变化，这在信息论中是不考虑的. 如果把概率认为是客观的，这样的问题都是不能接受的.

如果把概率看作是信念，这样的问题是非常自然的.

客观的概率由频率的稳定性表现出来. 它的度量自然可以从频率去估计；而信念却是难以度量它的值，用一个[0,1]区间内的数值去表达一个信念，本身就很不精确. 因此如何用概率去反映、描述、确定人们对于某一事物的信念，这就是一个不易处理的难点.

2. 怎样综合不同信息的信念？即使我们已有办法把人们的信念用概率表示出来，不同信息提供的概率如何给以综合，这在原来的概率理论框架内是无法处理的. 用数学的语言来叙述，问题是这样的：

给定了一个可测空间$(\Omega,\mathscr{F})$，把一种信念看成是$(\Omega,\mathscr{F})$上的一个概率测度，如果有 K 个信息，就给出了 K 个测度 $P_i,i=1,2,\cdots,K$. 所谓综合，就是要从 $P_1,\cdots,P_K$ 去获得一个概率测度 P，P 还是$(\Omega,\mathscr{F})$上的测度. 常用的一种办法是对 P_i 赋以权 $\omega_i,\omega_i\geqslant 0$，$\sum_{i=1}^{K}\omega_i=1$，用$\sum_{i=1}^{K}\omega_iP_i$ 作为综合后的 P. 且不说如何去定 ω_i，这就有困难. 从直觉上看，这一类 P 是不合要求的. 因为不论怎样选择 ω_i，总有

$$\min_{1\leqslant i\leqslant K}P_i\leqslant P=\sum_{i=1}^{K}\omega_iP_i\leqslant\max_{1\leqslant i\leqslant K}P_i$$

人们的直觉是，如果不同的信息都肯定了某一事件出现的可能性较大，那么人的信念就会增大，它比每个信息提供的信念会更大，而上述不等式是与这个直觉不一致的.

人工智能和专家系统的技术处理中，有不少内容都是与这个问题有关的. 综合的过程是一个动态的过程，它有一个顺序. 从概率的观点来看，事件 A 发生后，又发生了事件 B，其条件就是 A 与 B 都已发生，这与先发生 B、后发生 A 并无什么不同，对 AB 考虑条件概率和对 BA 考虑是一样的，就是把一个动态的过程，给以静态化了. 然而，在人的信念变化中，A 先发生与 B 先发生，人的感受是不同的，俗语“先入为主”正是反映了这一种现象. 人的信念不

断变化的过程，反映了人的认识在不断变化的过程，所以动态地反映人的信念的变化，也是一个新的问题.

尽管有这些困难，不可否认，概率论的一些结果和处理方法在不确定性的研究中仍然有着很大的影响，它是一个基本的工具. 为了使词意更为明确，本书中概率一般都是指客观的、呈现频率稳定的可能性的度量，而把人的主观的对某一事物认识确信的程度称为信念. 信念是可以随认识而改变的.

在本书中假定读者已具有概率论的知识，对事件 $A,B,C\cdots$ 的运算，对 $P(A)$ 表示 A 的概率等符号都已熟悉. $\bar{A}$ 表示 A 的逆，$A\cup B$ 表示集合的加法，$A\cap B$（或 AB）表示集合的乘法，$\varnothing$ 表示空集等就不一一加以说明了，一些公式，如贝叶斯公式、乘法公式等就不给证明，直接引用. 我们认为概率遵从的一些规则对信念也是适用的.

§1.3 度量理论的重要性

从上面的议论可以看出，无论是概率、信念、它都是一种度量. 所以首要的问题是我们所考察的对象能否赋以一种度量. 例如对于某一事件在一定条件下发生的可能性，能否用[0,1]中的一个值来度量，有人认为只能给出大小的序，无法给出确切的值. 这就表示是不能用一个数值函数来度量可能性. 在人们作出决策时，对各种不同的决策排出一个优先的次序就行了，并不在于要定量地给出它的“价值”有多少. 那么在人们不断获得信息时，判断某一事物的信念也是变大或变小，而不一定能给出变化的“量”的大小. 因此对度量理论给以足够的重视，在不少学科中具有很明显的学术意义. 尽管在数学方面，早已有这一方面比较系统的研究，但并未被其他学科，特别是社会科学、决策、管理等专家所重视，在我国更是缺少这一类的教材. 在这里我们对这一内容略作一些介绍，使得我们对信念的度量会有一个更明确的观点.

给定了一个由对象组成的集合 A，对 A 中的元素 $a,b,c,\cdots$ 之

间给出了一种关系，通常是一个二元的关系，记为 aRb，即 a 与 b 有关系 R. 从度量的观点来看，有两个基本问题要回答：

(i) 是否存在一个函数 $f: A \to (-\infty, \infty)$，使得

$$aRb \Rightarrow f(a) > f(b)$$

这称为表示问题.

(ii) 是否这种表示是唯一的，在什么意义下是唯一的.

例如给定了一个集合 Ω，对 Ω 的一切子集组成的 σ 代数，可以定义一个包含关系，若 A 与 B 都是 Ω 的子集，当 $A \subset B$ 时，就认为有关系 ARB. 容易看出，概率 $P(A)$ 就具有这种性质，即

$$ARB \Rightarrow P(A) \leqslant P(B).$$

概率就是一种度量，当集合看作事件时，$P(A)$ 就是事件发生可能性的一种度量.

同样的，信念也是一种度量，它反映了人们对某一事物确信的程度. 因此它也存在上述两个基本问题，能否表示，表示的度量是唯一的吗？

从数学上看，$(-\infty, \infty)$ 上的实数有几种不同的结构，一种度量就是把一个抽象集合上定义的关系 R，保持某种性质映射到实轴上，与实轴上的某一种结构呈现同态或同构的映射. 因此，从同态、同构的性质来看，关系 R 必须具备某种性质是很自然的事了.

对于有限集，已有一些明确的结论，我们列举一些定理，有关证明可以参见参考文献[91]中有关的章节.

定理 3.1 设 A 是一个有限集，R 是一个 A 上的二元关系，则存在一个实值函数，定义在 A 上，满足

$$aRb \Rightarrow f(a) \geqslant f(b)$$

的充分必要条件是 (A, R) 是一个弱序结构.

这条定理告诉我们，对 A 上的关系 R 能给出一个序的度量（大小，优劣），R 必须是一个弱序结构. 因此弱序结构是什么就很

重要.

A 上的一个关系 R，如果具有传递性(transitive)和强完备性(strong complete)，则称 R 是一个弱序结构.

传递性是：$aRb, bRc \Rightarrow aRc$.

强完备性是：对 A 中的任何两个 a 与 b，或者 a 与 b 相同，或者 aRb，或者 bRa.

在$(-\infty, \infty)$上的实数就具有这种性质，我们把

$$a \geqslant b \text{ 定义为 } aRb$$

易见它有上述两条性质. 因此一个能用实数的大小顺序(注意不是实数的确切的值是多少)来度量的关系，它必须具有弱序的结构. 很明显，定理中的 f 并不是唯一的.

在信念中，信念程度的强弱并不具有强完备性，通常这种强弱是一个优先关系. 我们考察的对象中 $A>B$ 表示对 A 的信念强于 B. 如果$\sim(A>B)$[即 $A>B$ 不成立]，并且$\sim(B>A)$，则 A 与 B 认为是相同的，记为 $A \approx B$(这是指在信念程度上的相同). 于是就可问，对给定的优先关系，能否给出一个信念的度量 $f(\cdot)$，使得

(1) 若 $A>B$，就有 $f(A)>f(B)$ （弱度量）；

(2) $A>B$ 必须而只须 $f(A)>f(B)$ （强度量）.

这两个问题的答案是 Wong 等人的工作[72]，他们的结论是

定理 3.2 设考虑的对象是 Ω，在 Ω 的一切子集上定义了一个优先关系. 则

(1) 弱度量有解的充要条件是下面 $C1$—$C3$；

(2) 强度量有解的充要条件是下面 $C1$—$C5$.

其中

$C1$. 反对称性：$A>B \Rightarrow \sim(B>A)$，

$C2$. 负传递性：$\sim(A>B), \sim(B>C) \Rightarrow \sim(A>C)$，

$C3$. 支配性：$A \supseteq B \Rightarrow \sim(B>A)$，

$C4$. 部分单调性：$A>B, A \cap C=\varnothing$

$$\Rightarrow (A>B \Rightarrow A \cup C>B \cup C),$$

$C5$. 非平凡性：$\Omega > \varnothing$.

其实这些性质在文献[91]中早已有较为详尽的讨论，文献[91]所列的书是70年代出版的，文献[72]Wong等人的工作是1991年发表的. 这就可以明显感到，在处理不确定性的度量时，度量理论对我们的认识是起着重要的作用的.

§1.4 效用函数

在决策理论中，经常可以看到引入一个效用函数，然后根据平均效益的好坏来决定哪个决策是好的，应该采用的，哪个是不好的，不宜采用的.

采用平均效益的大小来衡量好坏，这与不确定性的程度有密切的联系，实际上就与决策引起的风险有关. 不同的人对风险的观点是差别很大的. 本书不考虑引入效用函数的观点来处理，主要是由于以下这些看法.

在不确定性的推理中，我们希望利用现有的信息，获得一个比较可信的结论，这个结论对个人、社会的后果如何，那是另一个问题. 而引入效用函数后，问题的重点就转变为与后果相联的效用了. 自然也可以这样设想，我们对命题赋一个效用函数，它就是对命题的信念程度，这样信念与效用不也就一致了吗?使用效用的观点，在选择合理的推断上，它本身就有困难. 法国的经济学家提出了下列问题：

考虑下面的四种彩票(见图1.1). 在彩票$l^{(1)}$与$l^{(2)}$中选择哪一个? 在$l^{(3)}$与$l^{(4)}$中选择哪一个? 线上的数字表示发生的概率. 从平均效益看$l^{(2)}$比$l^{(1)}$好，但不少人会考虑$l^{(1)}$，我有100%的把握得10^6元，为什么要去冒风险，反而可能什么都得不到呢? 这样他就会选择$l^{(1)}$而不是$l^{(2)}$. 对$l^{(3)}$和$l^{(4)}$就会这样考虑，既然得5×10^6元与10^6元机会差不多，而我最坏是什么都没有，为什么不选$l^{(3)}$呢. 从这种考虑来看，人们对风险、损失的考虑不同，选择也会不同.

在人们经验中，不同的信息对调整人们对某一事物的信念是

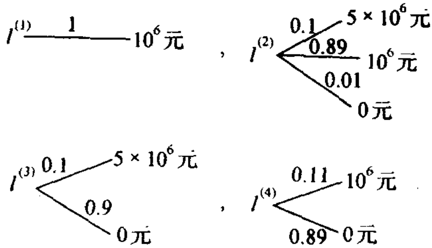

图 1.1

会随不同人的想法而有明显的差别. 然而在考虑不确定性的推理中,这一类因素也需加以考虑的话,问题就复杂多了.

另一方面,也是基于我们能作出什么解答,应该提供什么解答来考虑的.

不确定性推理的结果在一定意义上与天气预报是类似的. 根据目前掌握的信息,对于明天某一地区的天气状况,不可能完全肯定会是天晴,阴天,还是下雨,应该是这几种情况都可能,但各种可能性的程度是不同的,天晴的可能性最大,于是作出预报是明天天晴. 这个预报是可能不对,但是当我们要求预报的结论必须是明确的结论时,就会有报错的情形. 例如明天天气可能的情况是:

天晴 0.68　阴天 0.25　下雨 0.07

或　天晴 0.40　阴天 0.30　下雨 0.30

都会预报明天是天晴,但出错的可能性就大不一样. 在第二种情况下,与其报天晴,还不如把几种可能的情况及相应的概率告诉大家. 这就是能作什么预报和应该作什么预报. 能作的是带有不确定性的预报,应该作的也是附着不确定性的预报.

从这个观点来看,不确定性推理规则应放在如何去作出比较可靠的不确定性的推断,至于如何使用这个信息去作出决策,各人