

“知识论”基础研究

钟义信

(北京邮电大学,北京 100876)

摘要: 知识是人类所创造的宝贵财富,但至今没有形成系统的知识理论.本文旨在提出和建立知识论的必要基础,给出知识的概念、定义、表示、度量、推理和决策规则.知识论的建立将为信息论-知识论-智能论的统一理论奠定坚实的基础,促进人们在更高的水平上利用信息和知识,研究、设计和应用各种智能机器.

关键词: 知识;知识量;基于知识的推理;基于知识的决策;信息-知识-智能转化理论

中图分类号: TP273⁺.22 **文献标识码:** A **文章编号:** 0372-2112(2001)01-0096-07

Knowledge Theory: Fundamentals

ZHONG Yi-xin

(University of Posts and Telecommunications, Beijing 100876, China)

Abstract: Knowledge has been very important wealth to the mankind but there has not a Knowledge Theory existed yet till the present time. An attempt is thus made in the paper to present a foundation of Knowledge Theory that includes a series of basic issues such as the related concepts and definitions, the methods of representation, the measurements, the reasoning and decision rules. It is believed that the establishment of the Knowledge Theory will lay a solid foundation to the unified theory of information, knowledge, and intelligence and that will greatly facilitate the effective utilization of information and knowledge leading to the growth of the research in the field of intelligent machines.

Key words: knowledge; knowledge quantity; inference and decision-making based on knowledge; a unified theory of information-knowledge-intelligence

1 引言

苏联科学院院士哈尔凯维奇在他1955年出版的《通信论简述》中指出:“信息学存在一个如同物理学‘能量守恒与转换’那样的基本定律,可惜至今没有被发现.”

把信息提炼成知识并把知识激活成智能,是信息学的核心和灵魂.因此,“信息-知识-智能转化理论”当是信息学的“能量守恒与转换定律”.

从工程科学的视角来看,科学技术的主要任务在于启迪人们:怎样利用各种资源创制先进的生产工具,扩展人类认识世界和改造世界的能力,改善和优化人类生存和发展的环境.

整个工程科学进步的历史,就是一部利用资源、创制和改进社会生产工具的发展史:古代人类利用物质资源制造了第一代社会生产工具——只利用材料而没有动力和智能支持的人力工具,如锄头镰刀等等,扩展了人的体质功能;近代人类利用物质和能量两种资源创制了第二代社会生产工具——需要材料和动力而没有智能支持的动力工具,如机床机车等等,扩展了人的体力功能;进入现代社会以来,人类正在学习利用物质、能量和信息三种资源创制第三代社会生产工具——需要材料、动力和智能共同支持的智能工具,扩展人的智力功

能.

因此,现代科学技术应当特别关注创制智能工具的科学理论和技术,以便利用当今最现代化的工具——智能工具——来实现国家工业、农业、科学技术、国防、文化教育以及其他各行各业的现代化.

创制智能工具的核心问题是:怎样把信息加工成知识,怎样把知识激活成智能.这就是本文研究的主题——知识论,它是信息论—知识论—智能论统一理论的组成部分,是信息科学的灵魂.国外最近出现了知识获取和发现的研究,但还相当局限,远未形成体系.

信息论和智能论已经分别在40年代和50年代问世,但是,长期以来,信息论和智能论各自独立发展,互相没有联系.事实上,在信息论与智能论之间应当存在知识论作为桥梁.然而,知识论却长期成为一段空白.这种状况已经成为一个瓶颈,使信息论和智能论的发展也受到了严重的制约.因此,研究和建立知识论已经成为一项十分迫切的任务.知识论的建立,将沟通信息论与智能论之间的联系,形成信息、知识、智能的统一理论,为创制智能工具系统提供坚实的理论基础.

2 概念与定义

一切科学理论都建立在自己的科学概念基础上. 知识论也不例外. 因此, 本节首先将定义一组关于知识理论的基本概念: 信息, 知识, 智能. 考虑到信息和智能的理论虽然还不算完整, 但毕竟已经存在, 于是, 我们研究的重点将放在知识理论本身. 只是为了揭示信息、知识和智能之间的联系, 才对照地研究信息与智能的定义.

首先给出信息的定义. 所谓得到了关于某个事物的信息, 通常就是指: 知道了这个事物现在处在什么样的运动状态以及知道了这个状态会按照什么方式变化. 于是有

定义 2.1 从本体论意义上说, 信息, 是事物运动的状态和状态变化的方式; 从认识论意义上说, 信息是关于事物运动状态以及状态变化方式的描述^[1].

例: 对于某个随机型信息 X , 如果它有 N 种可能的运动状态, 即其状态空间为: $x_1, x_2, \dots, x_n, \dots, x_N$, 又如果这些状态的变化方式是按照某种概率分布的规则进行的: $p_1, p_2, \dots, p_n, \dots, p_N$. 那么, 由状态空间与概率分布结合而成的概率空间, 就充分地刻画了这个信息.

现在就来讨论知识的概念和定义. 不妨首先来看知识的辞典定义:

定义 2.2 知识是人们从实践经验中总结出来且为新的实践所证实的东西.

为了便于与信息的定义进行对照研究, 也可以把知识重新等效地定义为:

定义 2.3 知识是认识论范畴的概念, 是关于事物运动状态和状态变化规律的描述.

例如, 在牛顿力学中, $F = ma$ 是一个知识, 它告诉人们: 质量为 m 的物体, 受到大小为 F 的力作用后会产生加速度为 a 的加速运动. 在这里, 知识所告诉人们的, 正是一类受力作用的物体的运动状态以及状态变化的规律.

可见, 信息和知识概念一脉相通, 又相区别. 由具体的“状态变化方式”到抽象的“状态变化规律”, 其间所经历的变化就在于人们对信息所进行的加工和提炼. 这样就看出, 信息作为一种原材料, 经过加工提炼之后, 就可能形成相应的抽象化产物—知识. 由此可以得到:

推论 2.1 知识是对信息进行加工提炼所获得的抽象化产物.

进一步, 我们引入智能的概念.

定义 2.4 在给定问题、问题的环境、主体在问题求解过程中所追求的目的的条件下, 智能就是有针对性地获取问题与环境的信息、恰当地对这些信息进行处理以达到认知, 然后在此基础上结合主体的目的信息合理地产生解决问题的策略信息, 并利用所得到的策略信息在给定的环境下成功地解决问题达到主体的目的的能力^[2].

由定义 2.4 看出, 作为一种能力, 智能包含三个基本方面: 首先必须具有认知能力, 即获取有关信息和恰当处理这些信息并由此生成相应知识(达到认知)的能力; 同时, 必须具有再生策略信息的能力, 即根据所得到的知识结合主体的目的

生成解决问题的策略信息(决策)的能力; 最后, 还必须具有解决问题的能力, 即利用所生成的策略信息在给定环境下实际解决问题达到目的的能力. 三者缺一不可. 显见, 智能-信息-知识是三个紧密依存的概念.

总结起来, 信息、知识、智能之间的关系可以这样来表述: 信息是基本资源; 知识是对信息进行加工所得到的抽象化产物; 智能是利用信息资源加工生成知识、进而激活知识生成解决问题的策略信息并在策略信息引导下具体解决问题的能力. 信息、知识和智能在整个信息过程中的地位以及它们之间的相互关系也可以由图 1 形象地表示出来.

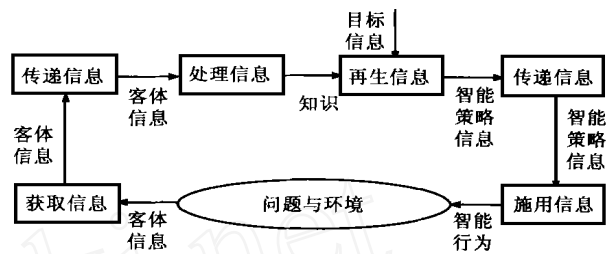


图 1 信息-知识-智能的关系

图 1 所示的关系, 正好符合人类自身认识世界和优化世界活动过程中由信息生成知识、由知识激活智能的过程: 其中, 获取信息的功能由人的感觉器官完成, 传递信息的功能由人的神经系统完成, 处理信息和再生信息的功能由人的思维器官完成, 施用信息的功能由人的效应器官完成. 由此可以知道信息-知识-智能统一理论的重要意义.

由此也可以得到:

推论 2.2 对问题和环境信息进行处理而生成知识, 知识被目的激活而生成智能.

3 知识的表示

上面已经阐明, 知识是认识论范畴的概念. 因此, 知识是相对于认识主体而存在的. 没有认识主体的存在, 就没有知识的存在. 正因为如此, 与本体论意义上的信息相比, 知识的概念更为复杂, 具有更丰富的内涵.

首先, 知识所表达的运动状态和状态变化的规律必然具有一定的外部形态, 与此相对应的知识可以称为“形态性知识”; 同时, 知识所表达的运动状态和状态变化的规律必然具有一定的逻辑内容, 与此相对应的知识可以称为“内容性知识”; 最后, 知识所表达的运动状态和状态变化的规律必然对认识主体呈现某种价值, 与此相对应的知识可以称为“效用性知识”. 需要强调的是, 形态、内容、效用三者的综合才构成知识的整体. 如图 2 所示.

由此可以引出:

公理 3.1 任何知识都由相应的形态性知识、内容性知识、效用性知识构成. 这种情形称为知识的三位一体.

为了对知识进行描述, 需要引进相应的描述参量.

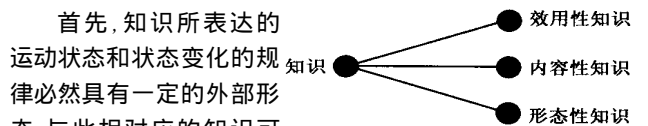


图 2 知识的三位一体

首先,我们来研究形态性知识的描述.

对于形态性知识的描述,就是对事物运动的状态及其变化方式的形式的描述,主要回答的问题是:“它有多少种可能的状态,这些状态变化方式的形式特征是什么”.

一般来说,从形式上来描述事物运动状态的特征就是直接对状态进行编号计数.而从形式上来描述状态变化方式的特征通常有三种不同的方法:如果事物是随机变量,它的状态变化方式的形式特征是概率分布 P ,如图 3 所示.

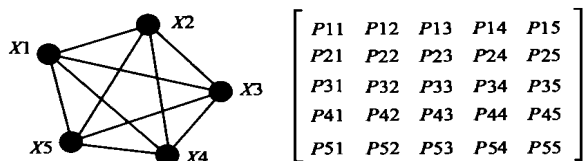


图 3 状态及其随机变化方式

如果事物是半随机变量(也称为“偶发变量”),它的状态变化方式的形式特征是可能度分布 Q .如果事物是确定性模糊变量,它的状态变化方式的形式特征则是隶属度分布 M .

根据莱布尼兹、贝努里和文献[1]的分析,概率分布、可能度分布、隶属度分布的统一概括是肯定度分布 C .因此,有

定义 3.1 设事物 X 具有 N 种可能的状态: $x_1, \dots, x_n, \dots, x_N$, 那么,状态 x_n 在形态上呈现的肯定程度称为状态 x_n 的肯定度,记为 $c_n, n=1, \dots, N$.由 X 的全部状态的肯定度所构成的集合,称为 X 的肯定度的(广义)分布,记为 C .

注意到概率和可能度的归一性、隶属度的不归一性^[1],肯定度 C 应当具有如下的性质:

$$0 \leq c_n \leq 1, \forall n \quad (1)$$

$$\sum_{n=1}^N c_n \geq 1$$

式(1)中的符号 \geq 表示全部状态的肯定度之和可以大于、等于、小于 1,不一定能够归一.具体来说,当给定的事物是随机型或偶发型变量时,和式归一;当给定的事物是模糊型变量时则不归一.

因此,可以用事物 X 的状态集合及其肯定度分布 $\{X, C\}$ 来描述事物 X 的形态性知识.

透过任何形态性知识,必有相应的逻辑内容.因此,应当进一步来讨论内容性知识的描述问题.显然,事物 X 的状态所代表的实际内容将随不同的具体事物而千差万别,不可能对它们一一做出具体的描述.关于内容性知识的比较合理的共性描述,是各个状态在逻辑上的真实性或真伪性.

定义 3.2 设事物 X 具有 N 种可能的状态: $x_1, \dots, x_n, \dots, x_N$, 那么,状态 x_n 在逻辑上真实的程度称为状态 x_n 的真实度,记为 $t_n, n=1, \dots, N$. X 的各个状态的真实度所构成的集合,称为 X 的真实度的(广义)分布,记为 T .

按照定义 3-2,显然有

$$0 \leq t_n \leq 1, \forall n \quad \text{和} \quad \sum_{n=1}^N t_n \geq 1 \quad (2)$$

因此,可以用事物 X 的状态集合及其真实度分布 $\{X, T\}$ 来描述事物 X 的内容性知识.

类似地,可以建立效用性知识的描述.很自然,可以根据

事物 X 各个状态 x_n 相对于主体目标所显示的价值来定义相应状态的效用度 $u_n, n=1, \dots, N$, 即

定义 3.3 设事物 X 具有 N 种可能的状态: $x_1, \dots, x_n, \dots, x_N$, 那么,状态 x_n 相对于主体目标所显示的价值称为状态 x_n 的效用度,记为 $u_n, n=1, \dots, N$. X 的各个状态的效用度所构成的集合,称为 X 的效用度的(广义)分布,记为 U .

按照定义 3.3, 也有

$$0 \leq u_n \leq 1, \forall n \quad \text{和} \quad \sum_{n=1}^N u_n \geq 1 \quad (3)$$

因此,可以用事物 X 的状态集合及其效用度分布 $\{X, U\}$ 来描述事物 X 的效用性知识.

注意到,从认识论的程序上考虑,在形式、内容和效用三要素之间,形式是最先被观察或感受到的要素,内容是要透过形式的分析才能进一步感受到的要素,效用则更是要针对一定的形式、内容和主体才能表现出来的要素.因此,除了如上所述分别给出单纯形态性知识、单纯内容性知识和单纯效用性知识的描述之外,还有必要讨论形态性与内容性知识的综合描述以及形态性、内容性和效用性知识的综合描述方法.为此,

定义 3.4 状态的肯定度与状态的真实度的结合称为状态的综合真实度,记为

$$\mathbf{T}_n = c_n \cdot t_n \Rightarrow c_n t_n, \forall n$$

$$\mathbf{T} = \{ \mathbf{T}_n \} \quad (4)$$

称 \mathbf{T} 为综合真实度分布.式中箭头符号表示“可简化为”的意思.显然有

$$0 \leq \mathbf{T}_n \leq 1, \forall n \quad \text{和} \quad \sum_{n=1}^N \mathbf{T}_n \geq 1 \quad (5)$$

与此相应的知识,称为综合内容性知识.

定义 3.5 状态的肯定度、真实度与效用度的结合称为状态的综合效用度,记为

$$\mathbf{U}_n = c_n \cdot t_n \cdot u_n \Rightarrow c_n t_n u_n, \forall n$$

$$\mathbf{U} = \{ \mathbf{U}_n \} \quad (6)$$

称为综合效用度分布.式中箭头符号表示“可简化为”.显然也有

$$0 \leq \mathbf{U}_n \leq 1, \forall n \quad \text{和} \quad \sum_{n=1}^N \mathbf{U}_n \geq 1 \quad (7)$$

与此相应的知识,称为综合效用性知识.

总之,不论何种描述知识的方法,都应当能够描述事物的运动状态以及状态的变化规律.

4 知识的处理

对于知识所进行的任何操作过程,都可以称为知识的处理.但是,实际上,知识处理的最基本和最重要的方面主要包括:各种知识的度量、基于综合内容性知识的推理、以及基于综合效用性知识的决策.下面就分别来讨论知识的度量、推理和决策问题.

4.1 知识的度量

定义 4.1 知识的数量称为知识量.

如同知识本身一样,知识量也可以进一步分为形态性知

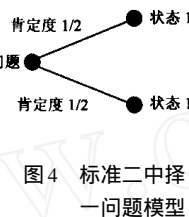
度量、内容性知识量、效用性知识量、综合内容性知识量和综合效用性知识量。其中,最基础的知识量是形态性知识量。我们的讨论就从形态性知识的度量问题开始。

研究知识量的一个直观而合理的思路是:用“所能解决的问题的多少”来度量相应的知识量的大小。因此,“知识量”的研究就转化为“问题量”的研究。

进一步,按照一般的度量理论,首先应当设计一种合理的问题标准,把它所包含的问题量作为问题量的单位,然后,任何一个实际问题的问题量就可以同这个单位相比较,从而得出这个实际问题的问题量。

注意到,一个形态性问题的量与两个因素有关:一是问题的可能状态数;二是问题各状态的肯定度分布。一方面,在同样的肯定度分布条件下,问题的可能状态数越大,问题量也越大;另一方面,在同样的可能状态数的条件下,肯定度分布越均匀,问题量也越大。

因此,最容易被接受的合理标准问题是“标准二中择一问题”,即一个问题只有两种可能的状态,且这两种状态的肯定度相等。这里所谓“标准的”二中择一问题,是指两种可能状态的肯定度分布为均匀分布。标准二中择一问题的模型如图 4 所示。



于是有

定义 4.2 把标准二中择一问题所包含的问题量定义为一个单位问题量,单位为奥特。如果一个知识正好解决了一个单位问题,那么,它的知识量也就等于 1 奥特知识量。

定义 4.2 是很自然的,因为按照这个定义,所谓 1 单位知识量就是解决一个标准的“是或非”问题、一个标准的“正或负”问题、一个标准的“男或女”问题、一个标准的“有或无”问题、一个标准的“好或坏”问题、一个标准的“输或赢”问题等等所需要的知识量。

其中单位“奥特”是“二中择一”问题的英文(Alternative)头三个字母 alt 的译音,也可以简称为“奥”。单位的中文“奥”字本身含有“深奥”和“奥妙”的意思,与“问题”的含义相通,因此用作问题量的单位,不无美感。在随机问题场合,“奥特”就等于“比特”。

有了问题量的单位,那么,一个具体问题的问题量就等于它所包含的标准二中择一问题的数量。例如,如果某个实际问题有 4 种可能的状态,且 4 种状态的肯定度都等于 1/4,那么,这个问题实际上包含了两个“标准二中择一问题”,它的问题量为两个单位,即 2 奥。如果一个问题有 8 种可能的状态,而且这 8 种状态的肯定度都等于 1/8,那么,这个问题所包含的问题量就是 3 奥,如此等等。

虽然利用单位问题量来测度实际问题量的方法在概念上非常直观自然,但是,实际使用却并不方便。一方面,小于 1 奥的问题量不好度量,另一方面,当问题所包含的可能状态数目不是正好等于 2 的 n 次方或肯定度不是均匀分布时,问题量也不好度量。因此,还是要寻求一般的问题量的度量方法。

为此,我们还需要引入新的概念和定义。

注意到式(1),肯定度存在归一和不归一两种情形。这里首先研究归一的情形。

定义 4.3 均匀分布的肯定度和零一型分布的肯定度代表肯定度分布的两种极端情形,分别把它们记为

$$C_0 = \{c_n | c_n = 1/N, \forall n\} \quad (8)$$

和

$$C_s = \{c_n | c_n \in \{0, 1\}, \forall n\} \quad (9)$$

定义 4.4 定义在肯定度分布 C 上的平均肯定度由下式给出:

$$M_\phi(C) = \Phi^{-1} \left\{ \sum_{n=1}^N c_n \phi(c_n) \right\} \quad (10)$$

式(10)中的 ϕ 是待定的单调连续函数, Φ^{-1} 是它的逆函数,也单调连续。

定义 4.5 两个问题 X 和 Y 具有相同的状态数 N , 各自的肯定度分布为 C 和 D , 若满足条件

$$\Phi^{-1} \left\{ \sum_{n=1}^N c_n \phi(c_n d_n) \right\} = \Phi^{-1} \left\{ \sum_{n=1}^N c_n \phi(c_n) \right\} \cdot \Phi^{-1} \left\{ \sum_{n=1}^N c_n \phi(d_n) \right\} \quad (11)$$

则称它们互相 ϕ 无关。

于是有下面的

定理 4.1 满足定义 4.4 和 4.5 条件的待定函数 ϕ 必为对数形式。

这是一个很重要的结果。但它的证明可在文献[3]和[4]中找到,这里从略。由此可以得出

系 4.1 状态数为 N 且肯定度分布为 C 的事件 X 的平均肯定度为

$$M_\phi(C) = \sum_{n=1}^N (c_n)^{c_n} \quad (12)$$

系 4.2 这样定义的平均肯定度的值界于 $1/N$ 与 1 之间

$$1/N = M_\phi(C_0) \leq M_\phi(C) \leq M_\phi(C_s) = 1 \quad (13)$$

系 4.1 和 4.2 的证明是直截了当的。

系 4.2 的结果表明,肯定度为均匀分布时,平均肯定度最小;肯定度为零一分布时则平均肯定度最大。前者是最不肯定的情形,相当于无知识的情形;后者是完全肯定的情形,相当于拥有充分知识的情形。

由此,可以很自然地引进一个重要的概念:某个观察者对于某个事物是否拥有知识,或拥有多少知识,可以用这个观察者对于这个事物所具有的平均肯定度的大小来判断。平均肯定度越大,拥有的知识越充分。

如果我们把最小平均肯定度作为一个比较的基准,就可以有

定义 4.6 观察者 R 关于事物 (X, C) 的知识量,可以用下式测度:

$$K(C) = \log \frac{M_\phi(C)}{M_\phi(C_0)} = \log N + \sum_{n=1}^N c_n \log c_n \quad (14)$$

定义 4.7 观察者 R 在观察某个事物 X 之前所具有的关于 X 的肯定度分布称为他关于 X 的先验肯定度分布,通常记为 C ; 观察之后的肯定度分布则称为后验肯定度分布,记为

C^* .

于是,所谓观察者 R 通过观察获得了关于事物 X 的知识,就是指他在观察之后关于 X 的后验平均肯定度比观察之前的先验平均肯定度增大了.

定义 4.8 观察者 R 通过观察 X 所获得的知识量则可以用下式测度:

$$K(C, C^*; R) = K(C^*) - K(C) = \sum_{n=1}^N c_n^* \log c_n^* - \sum_{n=1}^N c_n \log c_n \quad (15)$$

这是在肯定度分布归一的情形下关于知识量的重要结果.只要知道了观察者在观察某一事物或实验的先验和后验肯定度分布,就总是可以利用式(15)计算出观察者在观察过程中所得到的知识量.

由式(15)可知,当先验肯定度为均匀分布而后验肯定度分布为零一分布时,观察者所获得的知识量达到最大值.一般而言,只要观察者的后验平均肯定度大于先验平均肯定度,就意味着他能够在观察过程中获得某种程度的知识.另一方面,不管先验和后验肯定度分布的形式如何,只要两者相同,观察者在观察过程中所获得的知识量就总是为零.反之,若观察者的平均后验肯定度小于平均先验肯定度,就意味着他在观察过程中丢失了知识量.若观察者的先验肯定度分布为零一形式而后验肯定度分布为均匀分布,那么,观察者在观察过程中所丢失的知识量达到最大值.这些都是与人们的直觉相一致的结果,因而是合理的结果.

不难看出,定义 4.2 规定的单位知识量与定义 4.8 的理论结果是完全一致的.为此,只要在式(15)中令状态数 $N=2$, $c_1 = c_2 = 1/2$, C^* 为零一分布,就可以得到:

$$K(C_0, C_s^*; R) = 0 - \log(1/2) = 1$$

其中对数的底取为 2,单位为奥(特).

在理想观察条件下,后验肯定度为零一型分布.此时,若假定先验分布为均匀分布,那么由式(15)可以得到

$$K(C_0, C_s^*; R) = \log N$$

这时,知识量与状态数目呈增函数关系.

如果对这一关系作进一步的人为简化,把 $\log N$ 简化为 N ,就可以直接用状态数目来近似计算知识量.这就是为什么情报界和文化界通常都用字数来估计情报量的道理.显然,这只是一种非常粗糙的估计.

现在再来考虑肯定度不归一的情形,即模糊试验的情形.显然,由于肯定度不归一,不能直接应用前面的结果.但是,对于肯定度集合的任意元素,总可以构造新的分布:

$$\{c_n, (1 - c_n)\}, \forall n \quad (16)$$

显见,式(16)永远是归一的集合,因此,可以应用上面的结果.于是由式(12)有

$$M_\phi(C_n) = (c_n)^{c_n} (1 - c_n)^{(1 - c_n)} \quad (17)$$

由式(14)可以写出第 n 分量的先验形态性知识量:

$$K(C_n) = c_n \log c_n + (1 - c_n) \log(1 - c_n) + \log 2 \quad (18)$$

根据式(15)可以进一步写出第 n 分量的形态性知识量公式:

$$K(C_n, C_n^*; R) = c_n^* \log c_n^* + (1 - c_n^*) \log(1 - c_n^*)$$

$$- [c_n \log c_n + (1 - c_n) \log(1 - c_n)] \quad (19)$$

对于确定性的模糊试验来说,显然可以直接写出相应的平均知识量:

$$K(C, C^*; R) = \frac{1}{N} \sum_{n=1}^N K(C_n, C_n^*; R) \quad (20)$$

这样,我们就建立了形态性知识量的计算或测度的方法.

注意到逻辑真实度 T (综合逻辑真实度 T) 和效用度 U (综合效用度) 都具有模糊集合的性质,因此,式(16)~(20)的演算过程可以直接应用.只要把公式中的模糊肯定度参量换成相应的逻辑真实度(综合逻辑真实度)和效用度(综合效用度),同样也可以建立内容性知识(综合内容性知识)和效用性知识(综合效用性知识)的度量方法.只是由于本文的篇幅已经很长,这里就不一一写出了.

4.2 知识的推理

通常是对于“综合内容性知识”的处理.它的具体方法可以归结为三类:相对粗糙的经验推理、比较严谨的逻辑推理、更为高级的辩证逻辑推理.不过,由于辩证逻辑推理目前还不成熟,这里暂不涉及.

在一般情况下有

推理规则 4.1 推理的基本规则形式是

若 / 前提事物的运动状态和状态变化方式为真 / ,
则 / 结果事物的运动状态和状态变化方式为真 / .

(21 a)

或

If {前提} Then {结论} (21 b)

也可用符号表示为:

$$P \Rightarrow C \quad (21 c)$$

其中符号 P 表示推理的前提,可以是单一前提或多个前提; C 表示推理的结论,可以是单一结论或多个结论;双箭头表示推理的方向和过程.

容易看出,在式(21 a)的表达式中,无论是推理前提还是推理结论都是某种综合的内容性知识.因此,确实是基于综合内容性知识的推理.

类比推理是最常用的一种经验性推理,它的基础是“相似性原理”,其含义如下:

若事物 A 有 N 个状态,事物 B 也有 N 个状态;且事物 A 和 B 的前 $N-1$ 个状态都相同或相似,于是,如果 A 和 B 的状态变化方式也是相同的或基本相似的,那么,结论“ B 的第 N 个状态也与 A 的第 N 个状态相同或相似”应当为真.于是有

推理规则 4.2 类比推理规则的符号表达式是

If $\{A = \{A_n\}, B = \{B_n\}, n = 1, \dots, N; A_n = B_n, n = 1, \dots, N-1, C(A) \sim C(B)\}$

THEN $A_N = B_N$.

式中符号 C 表示事物运动状态的变化方式.

经验类比推理一般不能保证推理结果必然正确,但在经验和常识范围经常有效.经验类比推理容易理解,容易操作,容易被接受,因此,仍然是一种很有用的初级逻辑推理方法.

比较高级、比较规范、也比较理论化的逻辑推理是数理逻辑推理.它包括命题逻辑和谓词逻辑两种规范.数理逻辑推理

的基本特征是：基于事物之间存在的某种因果性联系或者某种默认的关系，建立了一套严格的推理公式和推理程式。

下面是熟知的命题逻辑和谓词逻辑推理规则。

推理规则 4.3 典型的命题逻辑和谓词逻辑推理规则包括：

$$P_1, P_2, \dots, P_n \Rightarrow \prod_{n=1}^N P_n \quad (22)$$

$$P, P \rightarrow Q \Rightarrow Q \quad (23)$$

$$(\forall x) P(x), A \Rightarrow P(A) \quad (24)$$

运用这些基本推理规则和其它有关规则，可以相当有效地进行许多重要的推理。例如，人工智能理论中著名的“三段论”逻辑推理就可以很容易表达出来

大前提：人都是要死的。

小前提：苏格拉底是人。

结论：苏格拉底也是要死的。

或写成数理逻辑表达式

$$(\forall x) \{ MAN(x) \rightarrow MORTAL(x) \}$$

$$MAN(SOCRATE)$$

$$MORTAL(SOCRATE)$$

而且可以很容易运用推理规则和归解原理证明：结论逻辑为真。

同样可以看出，在数理逻辑推理公式中，前提和结论也都是综合的内容性知识，因此，也是基于综合内容性知识的推理。经典逻辑虽然已经自成体系，但是，还有许多问题未能充分考虑。于是就导致了一系列新的逻辑系统（即所谓非标准逻辑）的陆续问世。

首先，知识通常都通过语言表达，而语言学常常包含大量的模糊现象，因此，作为对于经典标准逻辑的补充，模糊逻辑得到了越来越多的关注。模糊逻辑是通过在经典逻辑的基础上引入模糊因素而形成的。模糊因素可以表现为或者前提模糊、或者结论模糊、或者推理模糊、或前提结论和推理均为模糊的情况。例如下面的推理就是一个典型的模糊推理：

“如果（能说一口流利的普通话），那么（在中国住了相当长的时间。）：十有八九”

在这里，推理的前提“能说一口流利的普通话”（有多流利 ϑ ）和结论“在中国住了相当的长的时间”（有多长 ϑ ）都是模糊的，而且，由前提到达结论的推理本身也是模糊的，因此才会有“十有八九”（有多大的可能性 ϑ ）这样的模糊判断。一般地，有，

推理规则 4.4 模糊推理的公式可以表示为：

$$IF(\text{模糊条件}) \rightarrow THEN(\text{模糊结论}) \quad \text{置信度}(b) \quad (25)$$

对照前例，式中的含义一目了然，而 $0 < b < 1$ 是推理置信度（模糊度）的表示。

其次，经典的标准逻辑不能表示语言的情态，因此，作为补充，人们又提出了所谓的模态逻辑。它是通过在一阶谓词逻辑的基础上引入“必然”算符和“可能”算符而形成的。此外，为了表达逻辑推理中的时间概念，又引入了时序逻辑等等，就不一一讨论了。

还要特别指出，有时，一个推理规则可能会得出多个不同的推理结论，这时就需要计算由前提到达各个不同结论的综

合内容性知识量，选择其中最大者作为优选的结论，即：

$$\text{若 } K(T_{k0}) = \max_k \{ K(T_k) \} \quad (26)$$

则选择第 $k0$ 个作为优选的推理结论。

可见，正确的知识度量方法是研究知识推理的必不可少的理论基础。

4.3 基于知识的决策

基于知识的决策，就是面对给定的问题、环境和目标，利用所得到的知识选择适当的策略来求解问题，以期达到最大的得益或最小的损失。在这里，直观地看，直接发挥作用的是效用性知识，不过，由于效用性知识总是建筑在形态性知识和内容性知识的基础之上，因此，实际上发挥作用的应当是综合效用性知识。

如果把问题的“综合效用性知识量”定义为求解问题的目标函数，那么，最有利的求解策略就是能够使综合效用性知识量达到最大值的策略。我们把这种策略称为“很大综合效用性知识量”决策准则。这就是一种基于知识的决策。

值得指出的是，在这里，作为目标函数的“综合效用性知识量”不但直接体现了决策问题的目标——效用性知识量，而且也把环境的约束条件——问题状态的肯定度分布和逻辑真实度分布——同时表达出来了，实际上集目标函数和约束条件于一身。

具体来说，假设决策问题 X 有 L 种可能的状态 $\{x_l\}_{l=1}^L$ 。又假定这些状态的肯定度分布和逻辑真实度分布分别为：肯定度分布 $\{c_l\}_{l=1}^L$ ，真实度分布 $\{t_l\}_{l=1}^L$ 。进一步假定存在 K 种不同的求解策略 $\{a_k\}_{k=1}^K$ 。显然，不同的策略不会改变状态肯定度和真实度这些客观参数，因此应当有： $c_{kl} = c_l, t_{kl} = t_l$ ，对所有 k 和 l ；但是，状态的效用度却必然因策略的不同而不同，即 u_l 应与策略有关，因而必须改写为 $u_{kl}, \forall l, \forall k$ 。这样就可以得到一个决策矩阵：

	x_1	\dots	x_l	\dots	x_L
	c_1	\dots	c_l	\dots	c_L
	t_1	\dots	t_l	\dots	t_L
a_1	u_{11}	\dots	u_{1l}	\dots	u_{1L}
a_k	\dots	\dots	\dots	\dots	\dots
	u_{k1}	\dots	u_{kl}	\dots	u_{kL}
	\dots	\dots	\dots	\dots	\dots
a_K	u_{K1}	\dots	u_{Kl}	\dots	u_{KL}

(27)

由式(26)所示的决策矩阵可以建立相应于各个策略的状态综合效用度：

$$K_k = \sum_{l=1}^L c_l t_l u_{kl}, \forall l, \forall k \quad (28)$$

至此，就可以利用式(18)分别构造它们的先验综合效用性知识量。这里要根据 $\{K_k\}$ 是否归一区分两种情况：如果归一，则有

$$K(\text{归一}) = \sum_{k=1}^K K_k \log K_k + \log L \quad \forall k \quad (29)$$

否则有

$$K(\text{不归一}) = \sum_{k=1}^K K_k \log K_k + (1 - \sum_{k=1}^K K_k) \log(1 - \sum_{k=1}^K K_k) + \log L \quad \forall l \quad (30)$$

$$K(k) = \frac{1}{L} \sum_{l=1}^L K(kl) \quad \forall k \quad (31)$$

决策规则 4.1 根据“最大综合效用性知识”准则,基于综合效用性知识的决策规则是

$$\text{若 } K(k_0) = \max_{k, k_0, \forall k} \{K(k)\} \quad (32)$$

则 选择策略 k_0

由公式(28)~(32)所描述的基于综合效用性知识的决策规则是具有普遍意义的决策规则。可以看出,有了综合效用性知识,就有可能做出科学合理的决策。也要看到,公式中一些参数(如肯定度和真实度)是通过统计或其它方法提供的,具有客观性;但是,另一些参数(特别是效用度)的赋值也存在很强的经验色彩。

因此,一个好的决策者不但要具有决策理论的科学知识,也要具有比较丰富的决策经验和艺术水平,这样才能真正做出好的决策。这就是为什么人们经常说“决策是一门科学,同时又是一门艺术”的道理。

颇有意思的是,如果把这个规则作一些简化就可以发现,历史上已有的许多重要决策规则都可以在一定意义上看作“基于综合效用性知识的决策规则”这个普遍规则的简化规则。

先来简化基于知识的决策规则。例如,把式(28)简化为

$$K(k) = \frac{1}{L} \sum_{l=1}^L \sum_{kl=1}^L c_l t_l u_{kl} \quad (33)$$

进一步,若令 $t_l = 1, \forall l$,即认为所有状态都为真,同时假设决策问题是随机性事件,那么,肯定度就退化为概率,式(33)就化为:

$$K(k) = \frac{1}{L} \sum_{l=1}^L p_l u_{kl} \quad \forall k \quad (34)$$

于是,决策规则(32)就化为:

决策规则 4.2

$$\text{若有 } \frac{1}{L} \sum_{l=1}^L p_l u_{k_0 l} \geq \frac{1}{L} \sum_{l=1}^L p_l u_{kl} \quad (35)$$

则选择策略 k_0 。否则就选择策略 k 。

这恰好就是 Bayes 决策规则。它显然是“基于综合效用性知识决策规则”的一个特例。同样,容易证明,著名的极大极小决策规则也是基于知识的决策规则的简化规则。

决策规则 4.3 极大极小决策规则是:

$$u_{opt} = \max_k \{ \min_l [u_{kl}] \} \quad (36)$$

它的基本操作过程是:在决策矩阵 $\{u_{kl}\}, k=1, \dots, K, l=1, \dots, L$ 中,首先找出每一行 $u_{kl}, l=1, \dots, L, \forall k$ 的最小值(即每个策略的最小得益),然后,在所有各行的最小值之间挑选最大者(即挑选“最佳”策略)。因此,极大极小决策规则连每个策略的平均得益都不计算,只关心每个策略的最小得益(因此是最保守的规则)。它的最佳策略实质是最大的最小得益,因而有“极大极小”决策规则的名称。

可见,极大极小决策规则是基于知识的决策规则的进一步退化结果。

5 结语

本文以上给出的知识的概念定义、表示方法、定量计算、推理规则和决策准则都是知识理论的必要基础。有了这些基础,就可以进一步来研究知识论本身。作者认为,知识论的核心问题是揭示知识与信息、知识与智能之间的关系,阐明“如何把信息提炼成为知识;如何把知识激活成为智能”。

有关这方面的研究结果,将在另文讨论。

参考文献:

- [1] 钟义信. 信息科学原理 [M]. 福州市:福建人民出版社第一版,1988,北京邮电大学出版社再版,1996.
- [2] 钟义信等. 智能理论与技术:人工智能与神经网络 [M]. 北京市:人民邮电出版社,1992.
- [3] J. Aczel. Lectures on Functional Equations and Their Applications, Academic Press [M], New York, 1966.
- [4] G. H. Hardy, J. E. Littlewood, G. Pólya. Inequalities, Cambridge University Press [M]. London, 1973.

作者简介:



钟义信 北京邮电大学副校长,博士生导师,长期从事信息科学与人工智能的教学与研究。