

# 知识表示中的不确定性

李德毅

(中国电子系统工程研究所, 北京 100840)

**[摘要]** 知识表示一直是人工智能研究中的一个瓶颈, 其难点在于知识中隐含有不确定性, 即模糊性和随机性。文章提出用云模型3个数字特征(期望值, 熵, 超熵)来描述一个定性概念, 用熵来关联模糊性和随机性。代表定性概念的云的某一次定量值, 被称为云滴, 可以用它对此概念的贡献度来衡量, 许许多多云滴构成云, 实现定性和定量之间的随时转换, 反映了知识表示中的不确定性。论文以此对我国农历24个节气进行了新的量化解释。云方法已经用于数据开采、智能控制、跳频电台和大系统效能评估中, 取得明显的效果。

**[关键词]** 知识表示; 定性概念; 不确定性; 云模型; 数字特征

## 1 从一个例子谈起

云是对不确定问题的一种描述方法。刚开始接触云时, 对云模型难以有深刻的认识, 人们往往会拿它和统计学方法或模糊学方法相比较, 将云模型简单地看作为随机加模糊、模糊加随机、二次模糊

或二次随机等。为了消除这些疑惑, 下面从一个射击的例子谈起。

三位学者: 统计学家、模糊学家和云理论研究者被邀请参加射击评判。射手甲、乙和丙参加射击比赛, 射击情况如图1。

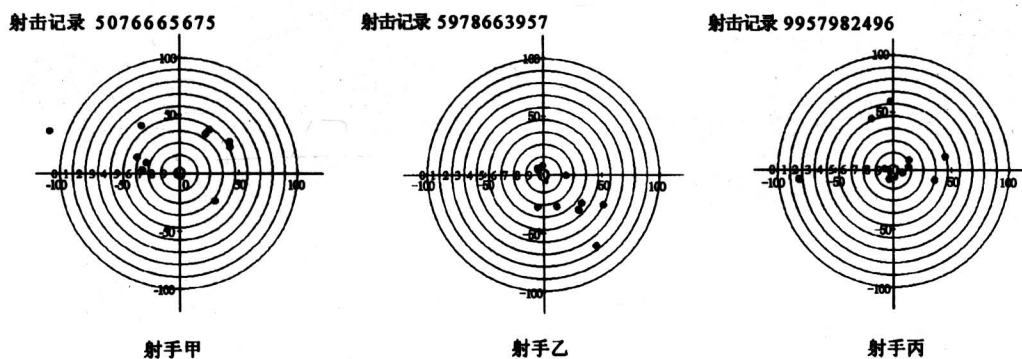


图1 射手甲、乙、丙的射击情况

Fig. 1 The records of three shooters

统计学方法认为, 射中与射不中有明确的定义, 是非此即彼的, 不存在亦此亦彼的中间状态。

但每次是不是射中, 包含有不确定性, 称为随机性, 试验的结果是随机的。令试验样本空间  $s =$

**[收稿日期]** 2000-05-19; 修回日期 2000-07-04

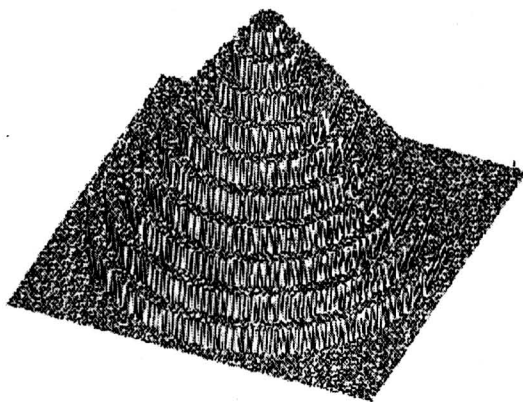
**[基金项目]** “八六三”高技术资助项目(863-309-ZT06-07-02)和“九七三”高技术资助项目(G1998030508-40)

**[作者简介]** 李德毅(1944-), 男, 江苏泰县人, 中国工程院院士, 北京电子系统工程研究所研究员, 博士生导师

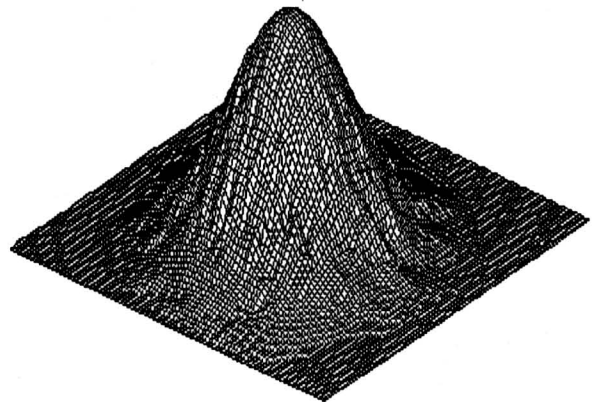
( $e$ ),  $e$  为样本空间的事件, 对于试验的结果, 引入变量  $x$ , 其值分别规定为 0 和 1。对应于样本空间的每个独立事件, 变量  $x$  取不同的值, 因而  $x$  可以作成是定义在样本空间上的函数。 $X(e)$  为随机变量。在大量重复试验时, 变量的取值具有统计规律性。概率论和数理统计就是研究和揭示随机现象的统计规律性的一门学科<sup>[1]</sup>。用中与不中来衡量每一次射击结果, 统计射手射击若干次后中靶的次数(频数)来反映射手的总体水平。例如, 射手甲经过 10 次射击, 9 次上靶, 一次跑靶, 则射手的击中概率为 0.9, 按照百分制计总成绩, 可为 90 分, 射手乙和丙的十次射击全部上靶, 成绩都为 100 分。因此, 射手乙和丙的水平相当, 都优于甲。

模糊学家认为, 中与不中的是相对的, 取决于弹着点离靶心的距离, 难以明确一个边界对中与不中进行精确的划分, 这种亦此亦彼的事件中所包含的不确定性, 称为模糊性<sup>[2,3]</sup>。如果样本空间  $s = (e)$  中的元素  $e$  代表不同的弹着点, 把“肯定射中”用数字 1 表示, “肯定不中”用数字 0 表示,

则对样本空间中的部分元素来说, 它们属于射中的程度可能不同, 用 0 和 1 之间的数值来反映这种中介过渡性<sup>[4]</sup>。模糊学就是研究和揭示模糊现象的规律性的一门学科。射中与射不中可以用弹着点对目标靶的隶属度表示。将目标从靶心开始分为十个等级表示击中目标的程度, 依次为 10 环、9 环、...、1 环, 跑靶为 0 环, 对应的隶属度分别为 1, 0.9, ..., 0.1, 0, 其对应的二维隶属函数见图 2a, 用弹着点在靶纸上所处环数作为射击的成绩。射手的总体水平, 还可以借助统计学, 采用公式  $SCORE = \sum_{i=1}^n w_i$  计算总成绩, 其中  $n$  为射击次数,  $w_i$  为第  $i$  次击中的环数。因为射手甲的射击记录为: 5/0/7/6/6/6/6/5/7/5, 乙的射击记录为: 5/9/7/8/6/6/3/9/5/7, 丙的射击记录为: 9/9/5/7/9/8/2/4/9/6。借助统计学的模糊学方法给出他们的总成绩分别为 53 分、65 分和 68 分, 射手丙的成绩最优, 射手乙的成绩优于甲。这里的 53 分、65 分以及 68 分与统计学家所给的 90 分、100 分是不同的概念。



a 二维隶属函数



b 二维云模型

图 2 二维隶属函数和二维云模型比较

Fig.2 comparison of 2D membership function and 2D cloud model

不确定性有两种: 随机性和模糊性。统计学和模糊学用各自的方法认识客观世界, 形成不同的评价结果。

通常, 人们更习惯于用自然语言值而不是精确数值来评价射手水平。为此, 云理论研究者提出云方法。射手射中或射不中带有随机性, 射中的程度

又带有模糊性, 每次射击的弹着点可以看作是一个云滴, 射击若干次后形成的云团的整体特征反映了射手总体水平。用定性的语言来描述这些云团, 例如对上述三位射手的射击情况, 可认为“射手甲略偏右上且不够稳定, 射手乙略偏右下但较稳定, 丙的射点靠近靶心但不稳定”。云方法提出用 3 个数

字特征（期望值，熵，超熵）来描述整个云团，实现定性和定量之间的转换。由于多方面的随机因素（天气、心理等等）的影响，射手很难每一次都击中靶心，其多次射击的弹着点在靶纸上呈近似正态分布。因此，用二维正态云模型  $(E_{x1}, E_{x2}; E_{n1}, E_{n2}; H_{c1}, H_{c2})$ （如图 2b）来描述总的射击情况：期望值  $(E_{x1}, E_{x2})$  是所有云滴（弹着点）在靶纸上的平均点的坐标，反映了射手对准心的把握，是

最能代表射手水平的靶位置；熵  $(E_{n1}, E_{n2})$  一方面反映弹着点的随机性，即分别在水平和垂直方向上相对于期望值的离散程度，另一方面又体现了射中的模糊性——隶属度；超熵  $(H_{c1}, H_{c2})$  反映了熵的离散程度，可以称为二次熵（熵的熵），体现了隶属度的不确定性。表 1 给出了对射手射击情况的不同评价结果。

表 1 对射手射击情况的不同评价结果  
Table 1 Estimation results for the three shooters

专家	射手甲	射手乙	射手丙	总评
统计学家	90	100	100	乙同丙，甲最差
模糊学家	53	65	68	甲最差，丙最优
云理论研究者	期望 (0.08, 0.1)，偏右上；熵 (0.45, 0.3)，超熵 (0.05, 0.07)，较离散，不稳定	期望 (0.1, -0.12)，右偏下；熵 (0.25, 0.2)，超熵 (0.02, 0.02)，较集中，较稳定	期望 (0, 0.03)，靠近靶心；熵 (0.4, 0.3)，超熵 (0.05, 0.07)，较离散，不稳定	射手乙的射击水平略高。

## 2 可还原性——云滴的生成算法

云是用自然语言值表示的某个定性概念与其定量表示之间的不确定性转换模型。云由许多云滴组成，每一个云滴就是这个定性概念在数域空间中的一次具体实现，这种实现带有不确定性，设  $A$  是一个集合  $A = \{a\}$ ，称为语言域。关于语言域  $A$  中的语言值  $a$ ，是指其映射到数域空间  $X$  的任意点  $x$  都存在一个有稳定倾向的数  $\mu_{\tilde{A}}(x)$ ，叫做  $x$  对  $a$  的确定程度。云的数字特征用期望值  $E_x$ ，熵  $E_n$ ，超熵  $H_c$  三个数值表征，它把语言值中的模糊性和随机性关联到一起，构成定性和定量相互间的映射，作为知识表示的基础。其中  $E_x$  可以认为是所有云滴在数域中的重心位置，反映了最能够代表这个定性概念在数域的坐标。 $E_n$  是定性概念亦此亦彼性的度量，反映了在数域中可被语言值接受的数域范围，即模糊度；同时还反映了在数域中的这些点能够代表这个语言值的概率。 $H_c$  是熵  $E_n$  的离散程度，即熵的熵，反映了每个数值代表这个语言值确定度的凝聚性，也反映云滴的凝聚程度。

给定云的三个数字特征：期望值  $E_x$ 、熵  $E_n$  和超熵  $H_c$ ，可以通过一维正态云发生器的算法生

成云滴<sup>[5]</sup>。

给定二维正态云的数字特征：期望值  $(E_x, E_y)$ 、熵  $(E_{nx}, E_{ny})$  和超熵  $(H_{cx}, H_{cy})$ ，可以通过以下二维正态云发生器的算法生成云滴：

- 1) 产生一个期望值为  $(E_{nx}, E_{ny})$ ，均方差为  $(H_{cx}, H_{cy})$  的二维正态随机熵  $(E'_{nx}, E'_{ny})$ ；
- 2) 产生一个期望值为  $(E_x, E_y)$ ，均方差为  $(E'_{nx}, E'_{ny})$  的二维正态随机数  $(x, y)$ ；
- 3) 计算

$$z = \exp \left\{ - \left[ \frac{(x - E_x)^2}{2E'_{nx}} + \frac{(y - E_y)^2}{2E'_{ny}} \right] \right\}$$

- 4) 令  $(x, y, z)$  为一个云滴，它是该云表示的语言值在数量上的一次具体实现，其中  $(x, y)$  为定性概念在数域中这一次对应的点的位置， $z$  为  $(x, y)$  属于这个语言值的程度的量度；

- 5) 重复步骤 1 到步骤 4，直到产生满足要求数目的云滴数。

这样的二维云发生器称为正向云发生器（见图 3）。

根据正态云发生器算法中的第 2 步，由统计学知识知道，99.74 % 的云滴都将落在  $(E_x - 3E_n, E_x + 3E_n)$  范围内；根据算法中的第 1 步和第 3 步

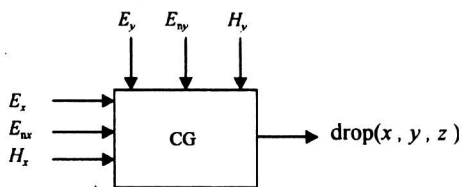


图3 二维正向正态云发生器

Fig.3 2D forward normalized cloud

计算，每一次的随机熵  $E'_n$  不同，导致云滴的离散性，包含云边缘的不分明和云厚度的不均匀；根据算法第3步计算，任何时候都会有  $0 < z \leq 1$ 。我们

可以认为函数：

$$z = \exp \left\{ - \left[ \frac{(x - E_x)^2}{2E_{n_x}^2} + \frac{(y - E_y)^2}{2E_{n_y}^2} \right] \right\}$$

是云的数学期望曲线。如果已经知道若干云滴，可以计算出它们所代表的正态云的三个数字特征：期望值  $E_x$ 、熵  $E_n$  和超熵  $H_e$ ，称之为反向云发生器，也可以构造带条件的正态云发生器。还可以利用类似方法构造其他分布的云发生器，如泊松云、 $\Gamma$  云等。

如果用二维正向正态云发生器来生成不同数量的云滴，可以大致还原3位射手的水平。图4中分别给出了还原各射手10个和100个弹着点的情况。

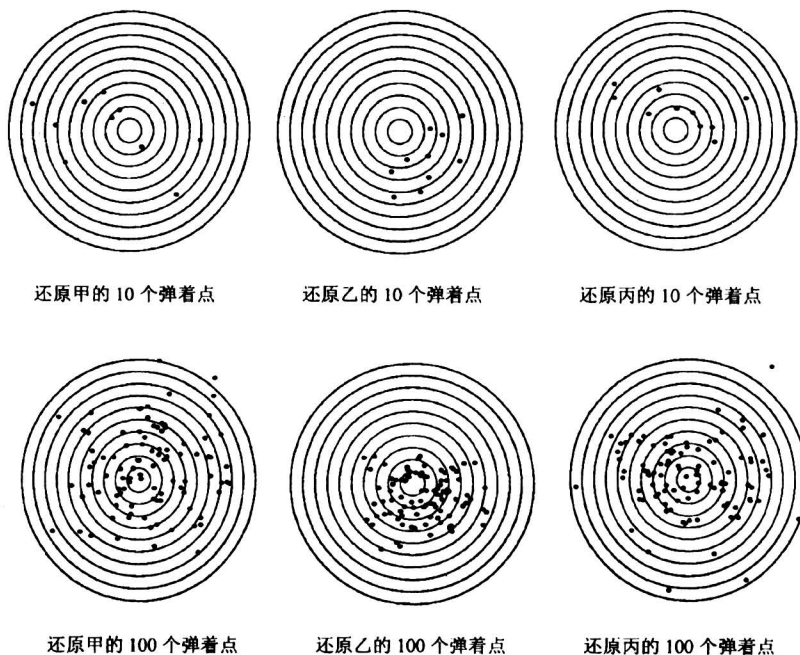


图4 用云模型还原射手的射击效果

Fig.4 Using cloud model to simulate the records

由此可见，三位学者对同一客观事物的认识分析方法不同：统计学方法从事件发生的频数出发，假设基本事件等可能地出现，通过概率来衡量其随机性；模糊学方法考虑概念的内涵和外延之间的关系，通过确定的隶属函数值来刻画事物的亦此亦彼性。云方法认为定性表述——自然语言具有不可替代性，基本事件等可能性的假设和隶属函数的单一确定值都具有局限性，通过云模型综合考虑随机性和模糊性，探索定量与定性的转换方式，用数字特征表示语言值。

### 3 定性概念的某一次定量值对此概念的贡献

为进一步深理解一维正态云模型，作一个重要的补充定义：

定义 基础变量  $X$  中的任一小区间上的元素  $\Delta X$  对定性概念  $\tilde{A}$  的贡献  $\Delta C$  为：

$$\Delta C \approx \mu_{\tilde{A}}(x) * \Delta x / \sqrt{2\pi} E_n$$

显然，论域上所有元素对概念  $\tilde{A}$  的总贡献  $C$  为：

$$C = \frac{\int_{-\infty}^{+\infty} \mu_{\tilde{A}}(x) dx}{\sqrt{2\pi}E_n} = \frac{\int_{-\infty}^{+\infty} \exp[-(x - E_x)^2/2E_n^2] dx}{\sqrt{2\pi}E_n} = 1$$

因为  $C = \frac{1}{2\pi E_n} \int_{E_x-3E_n}^{E_x+3E_n} \mu_{\tilde{A}}(x) dx = 99.74\%$ ,

所以对于论域  $X$  中的定性概念  $\tilde{A}$  有贡献的定量值，主要落在区间  $[E_x - 3E_n, E_x + 3E_n]$ ，甚至可以忽略  $[E_x - 3E_n, E_x + 3E_n]$  区间之外的定量值对定性概念  $\tilde{A}$  的贡献，这就是正态云的“ $3E_n$  规则”。根据定义，位于区间  $[E_x - 0.67E_n, E_x + 0.67E_n]$  的那些元素，占全部定量值的 22.33%，它们对定性概念的贡献占总贡献的 50%

，这部分元素称为“骨干元素”；位于区间  $[E_x - E_n, E_x + E_n]$  的那些元素，占全部元素的 33.33%，它们对定性概念的贡献占总贡献的 68.26%，这部分元素称为“基本元素”；位于区间  $[E_x - 2E_n, E_x - E_n]$  和  $[E_x + E_n, E_x + 2E_n]$  的那些元素，占全部元素的 33.33%，它们的贡献占总贡献的 27.18%，这部分元素称为“外围元素”；位于区间  $[E_x - 3E_n, E_x - 2E_n]$  和  $[E_x + 2E_n, E_x + 3E_n]$  的那些元素，占全部元素的 33.33%，它们的贡献占总贡献的 4.3%，这部分元素称为“弱外围元素”。如图 5 所示：

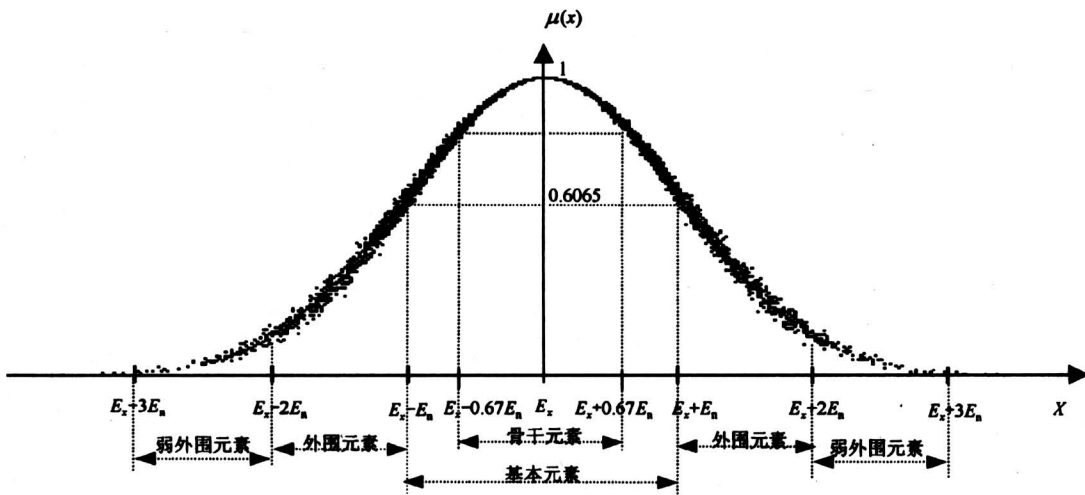


图 5 论域中的元素对定性概念的贡献

Fig.5 The contributions of elements in universe to qualitative

#### 4 用云方法表示并解释定性概念“春、夏、秋、冬”

“春天”、“夏天”、“秋天”、“冬天”是四个定性概念。我国古代劳动人民在长期的农业生产实践中，积累和掌握了农事季节与气候变化的丰富经验——定性的知识。农历 24 节气便是认识“春、夏、秋、冬”四个定性知识的一个重要经验结晶。“春、夏、秋、冬”这四个语言值，可以用如下的正态云表征，其正态云见图 6。图 6 中  $X$  坐标表示地球每年绕太阳公转 1 圈为  $360^\circ$ ，被 24 节气等分。它们所对应的数学期望曲线可以认为是模糊学中的隶属函数，呈钟形，峰值为 1。这和正态概率密度分布函数不相关。至于某年某月某日特定一天，属于某季节的程度，可以通过云滴来反映其不确定性。

据记载，早在春秋时代，便有“二分”（春分、秋分）“二至”（夏至、冬至）四个节气。“分”和“至”表示转折或极值，因此，分别对应“春夏秋冬”四个定性概念的期望值，而在这 4 个时刻，它们分别属于春夏秋冬的隶属度为 1，云的期望曲线——隶属函数在该点的一阶导数为 0，其左导数大于 0，右导数小于 0，反映了季节的转折。

经过 500 年左右，到了战国末期，又增加了四立（立春、立夏、立秋、立冬），“立”表示新季节的开始，也表示上一个季节的结束。又经过 100 多年的逐步补充，到秦、汉时期，就完备起来，在二分、二至和四立这八个节气之间又各增加了两个节气，分别用天气的炎热严寒（小暑、大暑、处暑、小寒、大寒）、水气凝结（白露、寒露、霜降）、雨雪多少（雨水、谷雨、小雪、大雪）和生物发育

(惊蛰、清明、小满、芒种)来反映它们分别属于春夏秋冬的程度。以定性概念“春”为例,雨水和谷雨属于“春天”的期望隶属度皆为0.1353;惊蛰和清明皆为0.6065;惊蛰到清明这一段时期对“春天”的贡献为68.26%,雨水到谷雨对“春天”

的贡献为95.44%,立春到立夏对“春天”的贡献为99.74%,夏、秋、冬天以此类推。这24个节气,乃是正态云期望曲线上特定点,都落在 $E_n$ 点的整数倍上。

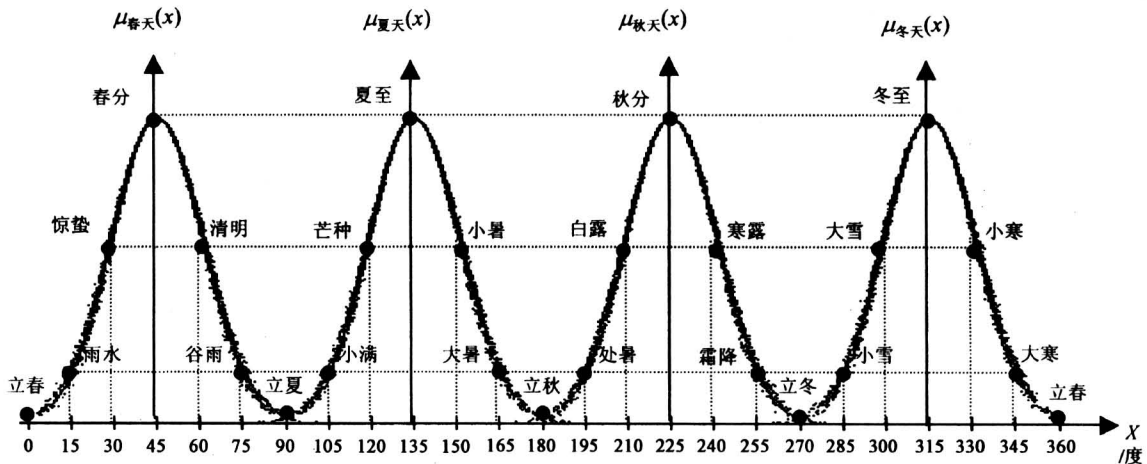


图6 24节气图

Fig.6 24 solar terms in lunar calendar

## 5 结语

科学需要重复,不能够重复的一次性现象,科学中一般不予研究。原则上说,多次重复表现出来的规律性——概率,通常由实验的条件决定。但是,实验条件常常难以确切地叙述,有时过于苛刻,如要求独立的随机事件。通过修改条件把不可重复的试验化为可重复的试验有时非常困难。这是概率论的局限性。科学需要精确,不能够定量描述的现象,科学中一般也不予研究。模糊学用隶属函数来定量描述客观事物的亦此亦比性,然而它忽视了隶属函数自身的不确定性,这是模糊学的局限性。客观世界中许多问题,特别是复杂系统和人文社会,其最有效的知识表示方法还只能是自然语言。因此,我们企图用云方法来表示知识中的不确定性。云模型不是简单地随机加模糊,也不是模糊加随机,更不是二次随机或二次模糊。云就是云,它很难把模糊性和随机性人为地分开,而是通过每个数字特征的双重性有机地关联在一起,实现定性语言值与定量数值之间的自然转换。目前,云理论已经运用于智能控制,成功地实现了对三级倒立摆的实时动平衡转换<sup>[6,7]</sup>;用于数据挖掘,实现了对关联规则和预测知识的发现<sup>[8~10]</sup>;用于跳频电台

产生跳频序列码,还用于大系统效能评估等多个领域。相信云方法作为知识表示和定性定量转换的有力工具,必将在更多的领域得到广泛应用。

### 参考文献

- [1] 王梓坤. 概率论基础及其应用 [M]. 北京: 北京师范大学出版社, 1996
- [2] 伯特兰·罗素. 论模糊性 [J]. 模糊系统与数学, 1989, 3 (1): 16~22
- [3] Zadeh L A. Fuzzy sets [J]. Information and Control, 1965. (8): 338~353
- [4] 李中夫. 隶属度含义剖析 [J]. 模糊系统与数学, 1967, 1 (1): 1~6
- [5] 李德毅; 孟海军, 史雪梅. 隶属云和隶属云发生器 [J]. 计算机研究与发展, 1995, 32 (6): 15~20
- [6] 李德毅. 三级倒立摆的云控制方法及动平衡模式 [J]. 中国工程科学, 1999, 1 (2): 41~45
- [7] Li Deyi, Cheung David, Ng V, et al. Uncertainty reasoning based on cloud models in controllers [J]. Computers and Mathematics with Application Elsevier Science, 1998, 35 (3): 99~123
- [8] Li Deyi, Di Caichang, Li Deren, et al. Mining association rules with linguistic cloud models [J]. Journal of Software, 2000, 11 (2): 143~158
- [9] Jiang R, Li D, Chen H. Time-series prediction with

cloud models in DMKD [A]. In: Zhong Ning, Zhou Lizhu. Lecture Notes in Artificial Intelligence 1574, Subseries of Lecture Notes in Computer Science [C], Methodologies for Knowledge Discovery and Data Mining, The Third Pacific-Asia Conference, PAKDD '99, Beijing, China, April 1999. 525~529

[10] Li Deyi, Di Kaichang, Li Deren, et al. Mining association rules with linguistic cloud models [R]. PAKDD '98, The Second Pacific-Asia Conference on Knowledge Discovery & Data Mining, Melbourne, Australia, April, 1998

## Uncertainty in Knowledge Representation

Li Deyi

(China Electronic System Engineering Corporation, Beijing 100840, China)

[Abstract] Knowledge representation in AI has been a bottleneck for years. And the difficulty is uncertainty hidden in qualitative concepts, that is the randomness and fuzziness. At this junction, this paper presents a new concept of cloud models with three digital characteristics: expected value  $E_x$ , entropy  $E_n$ , and hyper entropy  $H_e$ . This methodology has effectively made mapping between quantitative and qualitative knowledge much easier at any time. A cloud drop, that is a quantitative value, representing the qualitative concept can be measured by contributions. A new explanation for the 24 solar terms in lunar calendar is given as well. The cloud models have been used in data mining, intelligent control, hopping frequency technique, system evaluation, and so on.

[Key words] knowledge representation; qualitative concept; uncertainty; cloud model; digital characteristics

## 国际大坝委员会第 68 届年会及第 20 届大会 在北京隆重召开

为世界坝工界瞩目的国际大坝会议——第 68 届年会和第 20 届大会，分别于 9 月 14~18 日和 9 月 19~22 日在北京国际会议中心召开。这次会议由水利部、国家电力公司、中国长江三峡工程开发总公司、中国水利学会、中国水力发电工程学会和中国大坝委员会共同主办，并得到了北京市政府、国家自然科学基金委员会等有关单位的大力支持。大会组委会名誉主席、国务院副总理温家宝出席了 19 日的开幕式并发表了讲话。中国工程院副院长、中国大坝委员会名誉主席潘家铮出席并主持会议。

我国水利部部长汪恕诚任大会主席，副主席分别由长江三峡开发总公司陆佑楣和国家电力公司副总经理周大兵担任。

会议期间，组织了“中国长江三峡工程专题讨论会和混凝土面板堆石坝国际研讨会”。着重讨论了风

险分析在大坝决策与管理中的作用，大坝与坝基监测，有闸门溢洪道，其它控制的泄洪设施与大坝安全，大坝的效益等坝工界关注的问题。

大会收到论文数百篇并汇编成论文集。

国际大坝委员会既是一个交流工程科技和管理的机构，也是一个进入国际水利水电市场的窗口，具有相当高的权威性。大会每三年一届，每届会议一般都有三四百篇论文。

我国于 1973 年 12 月申请加入国际大坝委员会，并于次年 4 月在雅典举行的第 42 届执行会议上批准通过，成淡国家会员。国际大坝委员会目前共有 22 个专业委员会，我国已参加了其中 14 个专业委员会的工作。

(柳 宏)