

云理论在资源环境领域中的应用前景展望

陈明, 肖克炎, 唐菊兴, 张晓华

CHEN Ming, XIAO Ke-yan, TANG Ju-xing, ZHANG Xiao-hua

中国地质科学院矿产资源研究所, 北京 100037

Institute of Mineral Resources, Chinese Academy of Geological Sciences, Beijing 100037, China

摘要:由中国科学家李德毅提出的隶属云概念是对模糊隶属度的发展, 目前已经发展成为由云模型、云变换、云推理等组成的“云计算”理论与方法体系, 其在地学界的应用尚未开展起来。模糊隶属云现象在资源环境领域普遍存在, 例如环境污染等级的模糊评判、综合物化探异常含矿性的模糊识别、矿山安全评价、描述性地学数据的定量化处理、地学数据仓库的知识挖掘、地理信息系统中属性数据的不确定性处理、矿产资源评价预测等, 迫切要求形成一系列数学假设和处理方法均符合多重模糊属性的处理方法。云计算理论与方法满足了这种客观需要, 在地学信息处理方面具有十分广泛的应用前景。

关键词:隶属云; 模式识别; 矿产资源预测; 环境质量评价; 数据仓库; 知识挖掘

中图分类号: P628

文献标志码: A

文章编号: 1671-2552(2011)05-0670-07

Chen M, Xiao K Y, Tang J X, Zhang X H. The application prospect of the Membership Cloud Theory in geosciences. *Geological Bulletin of China*, 2011, 30(5):670-676

Abstract: The concept of membership cloud was proposed by Li Deyi to overcome the deficiency of Fuzzy theory. It is now developed into a series of theories, including Cloud Module, Cloud Transformation and Cloud Reasoning. Rare application of the Membership Cloud Theory (MCT, briefly) in geosciences has been found so far. Phenomenon of member cloud is common in mineral resources and environmental sciences, such as fuzzy assessment of environmental pollution grade, metallogenetic explanation of geochemical and/or geophysical anomalies, assessment of mining safety, quantitative processing of descriptive geodata, knowledge mining from data warehouse, uncertainty research of location and attribute data, and mineral resource prediction, etc. Reasonable methodologies with suitable mathematical hypothesis are strongly needed to process these problems. The creation of MCT provides edge tools for geo-data processing.

Key words: membership cloud; model recognition; ore resource prognosis; environment quality evaluation; data warehouse; knowledge mining

1 模糊数学在处理模糊现象中的局限性

在资源环境领域中普遍存在模糊现象。典型的例子包括: ①在大气、水、土壤环境的污染评价中, 常需要根据 m 个变量的实测数据来判定 n 个样品属于 p 个质量等级中的哪一个。此时要用到模糊数学中的隶属函数, 在得出每个样品的隶属度后, 再结合

样品的权重通过关联分析来判别各个样品属于哪个污染级别。②在矿山环境安全评价中, 也需要根据生产安全指标、地质环境、污染物含量水平、迁移转化等因素, 来综合判定矿工安全生产的工作环境、职业病发生概率和对周围居民生活的影响。③在地球化学找矿中, 为了准确识别异常的含矿性, 需要根据多个元素的多个含量级别来综合判定异常下面是否具有

收稿日期: 2010-12-14; 修订日期: 2011-03-07

资助项目: 中国地质调查局国土资源大调查项目(编号: 121201096036)和国家科技支撑项目(编号: 2006BAB01A01)

作者简介: 陈明(1967-), 男, 研究员, 博士生导师, 从事矿产资源评价和环境地球化学研究。E-mail: hcenming@vip.sina.com

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

矿体。④在矿产资源评价中,则更需要从众多的、具有模糊特性的地、物、化、遥数据中发掘矿产资源——特别是隐伏矿的信息。

类似的问题不胜枚举,它们共同特点在于:对于任意一个样品,要判定它的属性,需要依赖多个变量;并且,对于任意一个变量——不管它是定性的还是定量的,它与等级或属性之间的关系不是一一对应的,而是一些“区间”对应于某个等级(或属性)。例如在大气环境质量分级标准中,Ⅰ类大气中的 SO_2 含量为 $0.000\sim 0.050\text{mg}/\text{m}^3$,Ⅱ类大气中为 $0.050\sim 0.150\text{mg}/\text{m}^3$,Ⅲ类为 $0.150\sim 0.250\text{mg}/\text{m}^3$,等等。又例如,在黑龙江某金矿区水系沉积物测量中,以 1.34×10^{-9} 为下限圈定的异常被随后进行的普查工作证明为含矿异常;在华南地区,水系沉积物中Au元素的背景值就在 2.75×10^{-9} 以上。就全国范围而言,金的水系沉积物含矿异常下限值处于一个区间范围内。

为了处理这类广泛存在的模糊问题,Zadeh^[1]提出了模糊集合的概念,并在各个学科中得到广泛的应用,国内外发表的相关文章不胜枚举。

随着理论研究和工程应用的不断深入,一部分科学家发现:模糊数学并不能完全胜任全部模糊现象的处理。主要理由是:①Zadeh本人也不能说清楚隶属函数——模糊集合理论的基石——概念的本质是什么。②模糊数学已产生近半个世纪了,但至今尚无确定隶属函数的普遍公认的方法。③现有的隶属函数都被强制地硬化为精确的数值表达,这实际上等于把模糊现象强行纳入精确数学的范畴,与模糊数学的本意背道而驰。

例如上面提到的金矿异常识别中,Au元素的异常下限并不遵循“元素含量越高,含矿性越好”的一般规律。当一位勘查地球化学家说“这个Au元素异常含金矿的概率大”时,指的是地表的土壤或水系沉积物或岩石等介质中的Au元素含量足够高。然而,到底Au元素含量多高才算高——指示“有金矿”,多低才算低——指示“无金矿”,没有哪位专家能够说得清楚。在黑龙江, 1.34×10^{-9} 的Au元素含量可能是金矿存在的良好指示;而在华南地区,即使Au元素含量达到 2.75×10^{-9} ,也不能说明金矿床的存在。虽然 2.75×10^{-9} 比 1.34×10^{-9} 的2倍还要高,却不能说明存在金矿的概率也提高了2倍。

由此可见,尽管“水系沉积物中Au元素含量”与“存在金矿”之间存在某种模糊的相关性——“Au

元素含量越高,存在金矿床的概率也越大”,但是却无法建立一个合适的隶属函数,使得对于任意一个样品,均可以根据该样品中的Au元素含量值来判定采样点下面是否含有金矿床。实践证明,想要在“精确的元素含量”与“模糊的含矿性”之间建立起一种确定性的、精确的判定关系是不可能实现的,这就是模糊隶属函数的局限性。

在科学实践中,只有当人脑的思维符合客观事物本身的规律时,才能得出正确的科学结论。事实上,在资源环境领域中很难找出一种现象(例如矿床)与另一种现象(例如地球物理、地球化学或遥感异常)之间存在“确定性的、精确的对应关系”,将地学中的模糊现象强制纳入精确数学的理想王国不可能取得实质性的成功。这就限制了“精确的”模糊数学在资源环境领域中的应用。

2 隶属云:一个地球化学找矿实例

鉴于模糊现象的模糊本质,为了克服Zadeh“精确的”模糊数学的缺陷并提供一种真正意义上的模糊数学工具来处理客观世界的模糊现象,中国学者李德毅等^[2]提出了“隶属云”的概念。

根据李德毅等^[2]的定义:设 X 是一个普通集合, $X=\{x_i|i=1,2,\cdots,n\}$ 称为论域。关于论域 X 中的模糊集合 \tilde{A} ,是指对于任意 x_i ,都存在一个有稳定倾向的随机数 $\mu_{\tilde{A}}(x_i)$,称为 x_i 在 \tilde{A} 的隶属度。如果 X 中的 x_i 都是简单有序的,或通过某种转换函数可以转化为简单有序的,则可把 X 看做是基础变量,隶属度在 X 上的分布称为隶属云。

上述抽象的概念对于许多并不精通数学的地质学家们来说是难以理解的。可以用下面某四级成矿区带内Au元素的区域化探异常的含矿性识别为例来说明。在该成矿带内的水系沉积物中,金矿的指示元素包括Au、As、Sb等。按照矿床的规模,可将28个已知的金矿区分成大型、中型、小型和矿化点4种类型,其分散流的异常规模(面金属量)与矿床规模之间的关系见表1,另有3个未知待判的综合异常。

由表1可知,水系沉积物中Au、As、Sb三元素的异常规模与金矿床的规模整体上呈现“异常规模越大,金矿床的规模也越大”的规律。但是,这种规模不是绝对的。大型、中型、小型和矿化点不同规模金矿所对应的异常,一方面其面金属量有一部分是重叠的(例如表1的①与②之间),另一方面,与具有不

表 1 某成矿带不同规模金矿 Au、As、Sb 元素的面金属量

Table 1 Areal productivity ranges of Au, As and Sb in 28 anomalies from different sizes of gold deposits in a metallogenic belt

金矿规模及异常	Au/ $10^{-9}\cdot\text{km}^{-2}$	As/ $10^{-6}\cdot\text{km}^{-2}$	Sb/ $10^{-6}\cdot\text{km}^{-2}$
大型	9.70~24.25 ^①	354.8~576.55	36.7~71.93
中型	7.34~14.86 ^②	208.05~463.13	26.69~41.14
小型	5.01~10.02	232.38~291.59 ^③	15.12~27.49
矿化点	2.90~5.82	69.99~114.32 ^④	5.12~13.47
未知异常 1	8.49	498.34	87.38
未知异常 2	11.34	299.03	14.38
未知异常 3	11.13	230.45	14.59

同规模、品质相对应的不同类别的异常之间,Au 元素和其它指示元素的面金属量的分级临界值之间可能是不连续的,或者说是“分裂的”,例如表 1 的分级范围③与④之间。

在上面这个例子中,所谓的论域 X 就是由 n 个 Au-As-Sb 综合异常的面金属量组成的数据集;而所谓的模糊集合 \tilde{A} 就是一个由不同规模的矿床所对应的各指示元素的面金属量构成的分类体系,相当于表 1 中双划线以上的部分。对于任意一个未知的综合异常,多大程度上属于大型或中型或小型或矿化点则由隶属度 $\mu_{\tilde{A}}(x_i)$ 来决定。 $\mu_{\tilde{A}}(x_i)$ 的具体计算公式由评判者通过一定的原则来确定。通常,可以将各个异常根据某个元素的面金属量的大小排列起来,形成一个有序的数据集,这样的数据集被称为“基础变量”。隶属度 $\mu_{\tilde{A}}(x_i)$ 表示 x_i 属于或不属于分类体系中某个类别(如金矿规模)的可能性的 大小,也就是日常所说的“找矿意义”。

上面的例子也明确地表达了一种现象:对于面金属量值,只有在某个数值区间才能明确地指示该异常属于某个亚类,而过大或过小都意味着它的隶属度是不确定的。也就是说,对于某个确定的 Au 元素的面金属量值,它指示大型、中型、小型金矿或矿化点的可能性不是唯一的,而具有模糊的特征。如图 1,对于任意一个面金属量值 x_i ,它的找矿意义(以 $\mu_{\tilde{A}}(x_i)$ 表示)并不总是确定的,而是有一定细微变化范围的,反之亦然。但是,这种变化一般不会过于剧烈而偏离隶属云的整体特征,并进一步导致勘查地球化学原理的颠覆。李德毅等^[2]将上述现象称之为“云”,形象地表达了“粗看有明显的形态,细

看则无明确的边界”的模糊现象。另外,“云”具有整体可移动性,这一点也符合地球化学数据的特性。例如前面提到的黑龙江和华南地区的 Au 元素背景值和异常下限问题中,如果脱离地球化学异常的整体形态和数字特征而去单独讨论某个采样点上 Au 元素含量的找矿指示意义将是没有意义的。这可以从根本上解释为什么单点异常的找矿意义不明确。

长期的找矿实践证明,地球化学异常的元素含量平均值、方差、线金属量、面金属量等统计参数的找矿意义具有整体的规律性,但对于某个特定的异常,其找矿意义又具有一定的模糊特性。以往我们试图通过精确数学中的正态分布或者对数正态分布来解决化探异常识别的问题,20 世纪 70~80 年代以后又开始试用模糊数学工具,但终究因为数学工具与实际工程问题之间先天的、不可逾越的障碍而收效甚微。从纯理论上讲,隶属云的理念更适合去解决勘查地球化学的实际问题。在基础地质、地球物理勘探、遥感影像处理、矿产资源评价、环境评价等领域,也存在与上述化探找矿相似的模糊现象,因此它们也同样应当可以在隶属云的思想的指导下,用云信息处理技术来解决那些“似是而非”的问题。

隶属云之于云信息处理技术就如同结构分析之于克里金。结构分析揭示了区域化变量的结构性和随机性,从而使地学曲面的重建、矿产资源的储量计算、多孔介质的连通性估计等复杂数值计算有了简单而科学的工具。隶属云概念的提出,则把隶属函数的结构性和随机模糊性进行统一的刻画;建立在隶属云函数基础上的云信息处理技术,也将深刻地影响地学领域的科学家处理科研与生产实践中常常遇

到的那些“似是而非”的问题的方式。

经过大约 15 年的发展,云理论已经发展成为云模型、云变换和云推理的体系,在空间广义知识和关联规则的挖掘、知识表达、不确定性推理、遥感影像解译和目标识别、智能控制等方面有了广泛的应用。

3 隶属云的重要特征参数及其地学意义

与地学领域专家非常熟悉的“随机函数服从正态分布函数”不同,隶属云内蕴了一个 3 次正态分布规律,即 $X \sim N^3(x_0, b^2, \sigma_{\max}^2)$ 。

(1)上式中 x_0 是隶属云的期望值,是隶属云函数曲线(图 1)下方的面积的重心,其坐标为 $(x=x_0, \mu=\sqrt{2}/4)$,它代表了某个隶属云所对应的论域(例如元素含量、面金属量、重磁测量值、图像灰度等)的中心值,实际上也是模糊集 \tilde{A} 中与该论域所对应的模糊子集 \tilde{A}_k (例如污染的等级、矿床的规模、断裂构造的深度等)的信息中心值。

(2) b 是隶属云的带宽。如果把正态分布看作是隶属云的一个特例,那么隶属云函数的带宽 b 相当于正态分布函数的均方差。

(3) σ_{\max} 是隶属云的方差,定义为隶属云期望曲线上点 $M(x=x_0+\sqrt{\ln 8b}, \mu=\sqrt{2}/4)$ 处所对应的隶属度随机分布的方差,反映了隶属云的离散程度,表征了事物的模糊性的 σ_{\max} 越大,模糊性越大,判定事物的属性的难度也越大; σ_{\max} 越小,隶属函数的确定性越好,判定事物属性的难度也越小。

对于地球化学找矿而言, σ_{\max} 越小,说明地球化学参数与矿产资源的对应性越好,化探数据的可解释性就越好。如此可类推到地球物理勘探、地球化学

找矿、遥感图像识别、污染评估、生态质量评价、地震预警、医疗诊断、生理学、生态学、心理学等领域。

在图 1 中,点 A 的坐标为 $(x=x_0, \mu=1)$ 。如果一个样品对于模糊子集 \tilde{A}_k 的隶属度正好处于 A 点,则说明该样品代表了此类样品的平均属性,并在分类体系中具有确定的分类位置。 B 点的坐标为 $(x=x_0+3b, \mu=0.0111)$,是隶属函数的右边界。在 A 点和 B 点,隶属函数的方差等于 0,说明在 A 和 B 处样品的属性是可以明确地判定的。在资源环境领域,可以将这些参数应用于描述性的、模糊性的问题进行定量化研究。

4 隶属云在资源环境领域中的应用前景

4.1 隶属云在数学地质中的应用前景

4.1.1 隶属云在单个模糊总体研究中的应用前景

不难理解,隶属云和云滴可以替代正态分布函数,借助蒙特卡洛模拟技术,广泛应用于地学领域。

蒙特卡洛模拟已经被广泛应用于矿产资源评价^[3-6]、地球物理场解译^[7-8]、石油勘探^[9]、遥感图像处理^[10]和地热勘探^[11]。对于资源环境领域的模糊隶属问题,只要用 3 次正态分布的隶属云发生器替代普通的正态分布随机数发生器,即将隶属云和云信息处理技术应用于那些“可靠性特征量已知的复杂系统”的模拟。

根据隶属云的数字特征,一方面可以利用隶属云的 3 个特征参数 x_0, b 和 σ_{\max} 产生满足正态隶属云分布规律的“云滴”,即二维数据点 $\zeta(x, \mu)$,成千上万的云滴就构成了整个隶属云;另一方面可以利用一定数量的已知“云滴”反推正态隶属云的特征参数 x_0, b 和 σ_{\max} 。隶属云在地学变量数字特征与分布规律研究中的应用,关键在于根据有限大小的样本确定隶属云的特征值,并使用隶属云发生器产生隶属云。隶属云发生器的算法如下:

(1)确定希望产生的隶属云的 3 个特征参数 x_0, b 和 σ_{\max} 。

(2)利用正态分布随机数发生器产生期望值等于 x_0 、带宽等于 b 的随机数,共计 n 个点, n 为一足够大的整数: $x \sim N(x_0, b^2)$ 。

(3)计算这 n 个随机数据点在期望值等于 x_0 、带宽等于 b 的正态隶属云的期望曲线在 x 处的隶属度: $\bar{\mu}_x = \mu(x) = e^{-\frac{(x-x_0)^2}{2b^2}}$,共得到 n 个隶属度值。

(4)计算这 n 个随机数据点在 $\zeta_0(x, \bar{\mu}_x)$ 处的方差

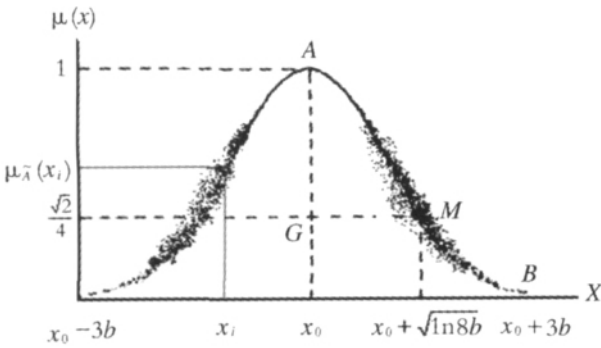


图 1 隶属云的图形与数字特征
(根据参考文献[2]修改,详细说明见正文)

Fig. 1 Graphic form and figure characteristics
of membership cloud

σ_x 。分 2 步计算:

①根据图 1 中的“3b 规则”,分别计算正态隶属云函数的左带宽 b_L 和右带宽 b_R :

$$b_L = \frac{1}{3} \int_{x_0}^{x_0 + \sqrt{\ln 8b}} \sqrt{1 + [\mu'(x)]^2} dx$$

$$b_R = \frac{1}{3} \int_{x_0 + \sqrt{\ln 8b}}^{x_0 + 3b} \sqrt{1 + [\mu'(x)]^2} dx$$

,其中 $\mu'(x) = e^{-\frac{(x-x_0)^2}{2b^2}}$

②计算这 n 个随机数据在点 $\zeta_0(x, \bar{\mu}_x)$ 处的方差:

$$\sigma_x = \begin{cases} \sigma_{\max} e^{-\frac{-(\Delta L)^2}{2b_L^2}} \\ \sigma_{\max} e^{-\frac{-(\Delta R)^2}{2b_R^2}} \end{cases}, \text{其中 } \Delta L = \int_{x_0 + \sqrt{\ln 8b}}^x \sqrt{1 + [\mu'(x)]^2} dx$$

得到 n 个方差值。

③根据 n 个随机数据所对应的 n 个隶属度值和 n 个方差值,分别计算数学期望为 $\bar{\mu}$ 、方差为 σ_x 的符合正态随机函数的云滴 $\zeta(x, \mu)$:

$$\mu \sim N(\bar{\mu}, \sigma_x)$$

如果把这 n 个云滴投影到 $x-\mu(x)$ 二维平面图上,则产生形如图 1 的隶属云。关于如何利用蒙特卡洛模拟进行矿产资源评价、地球物理场解译、石油勘探、遥感图像处理、地热勘探等,请参考有关文献。同样原理,也可以形成符合泊松分布和 Γ 分布的云,称为“泊松云”和“ Γ 云”。“泊松云”和“ Γ 云”同样在资源环境领域中有着广泛的应用。

4.1.2 隶属云在多模糊总体研究中的应用前景

有 3 类地质数据:定量数据、模糊数据和定性数据。目前,与定量数据有关的统计分析方法有线性回归、判别分析、因子分析、对应分析、典型相关分析等,与定性数据有关的统计分析方法包括数量化理论 ~ 。而模糊数据则处于定量数据和定性数据之间,目前除了模糊数学以外,尚未形成相应的多元统计分析方法,是多元统计分析研究领域未开垦的处女地。可以把上述将隶属云应用于单模糊总体研究的思路直接推广到多模糊总体的研究,并对那些基于样本相关系数矩阵的统计计算(例如 Q 型聚类分析、Q 型因子分析、对应分析等)产生直接影响,形成新的计算方法体系。

张勇等^[12]讨论了相似云及其度量分析方法,建议用相似云的欧氏距离来表征云的相似性,通过计算机模拟说明了该相似性度量的可行性,并应用于数字水印领域。很显然,这也可推广到相似系数和相关系数。但是鉴于隶属云包含三次正态分布函数,

是否可以用相似云的相似系数或相关系数直接代替多元统计分析中定量数据的相关系数矩阵,还有待于进一步研究。但可以预见,基于隶属云和云计算方法体系的“多维云统计分析”系列方法乃至“多维云统计学”终将形成。

4.2 隶属云在描述性地质学信息定量处理中的应用前景

描述性地质学信息的定量处理的实质就是知识挖掘。有一个笑话,说有一个学生数学学得很差。其父对学生说:“你数学这么差,以后也找不到工作。”学生回答:“那我就去当地质学家。”其父诧异:“地质学家就可以不懂数学吗?”答曰:“是啊,老师们都说,地质学家只会用语言描述,根本不需要数学。”这个笑话说明了一个问题,即传统的地质学的确是描述性科学,主要研究包括:野外现象的观察与记录、有限样品的分析测试与模拟实验、分类归纳、异同对比,并通过推理得出结论。

传统地质学使用的都是自然语言,所使用的概念和形成的知识具有 2 种不确定性:地质现象的模糊性(概念的亦此亦彼性)和随机性(地质事件发生的概率具有随机性)。而隶属云最擅长的就是处理既有模糊性又有随机性的问题,可以用于研究自然语言中的“基本语言值(语言原子)所蕴含的不确定性的普遍规律,使得有可能从语言表达的定性信息中获得定量数据的范围和分布规律,也有可能把精确的数值有效地转换为恰当的定性语言值”^[12]。

我们可以这么理解:每一个用自然描述的地质现象相当于一个云滴,这个云滴是某个定量模型的一次实现,由于定量模型是未知的,地质学家只能用一些定性概念进行描述。某一次这样的描述可能不足以表明定量模型的整体面貌,而且不同的地质学家在不同的时空条件下对同类现象会有不同的描述。但是,如果将数十年、上百年来成千上万不同的地质学家们的论著中的文字全部分解成“地质描述原子”,必将得到对应的云的整体形态并充分反映相关定性概念的基本特征。

不仅如此,还可以根据云的特征参数(x_0 、 b 和 σ_{\max})对该定性概念的确定性和随机性进行评价,甚至对某个地质学家的理论水平和观念的正确与否做出定量评价。这是因为云的“厚度(σ)”反映了科学家们的意见离散程度。不同的科学家对于某一地质现象的认识可以定量地归置于隶属云的某个确定的

点上,而隶属云的厚度是不均匀的,腰部(图 1 中的 M 点附近)最分散,而顶部和底部的汇聚性相对较好。

除了纯自然语言的传统地质学的定量化转化与处理外,隶属云也可以对地学中的定量数据做定性变换。例如,在资源与环境评价领域,人们关心的往往不是某个元素或指标的绝对值,而是这个数值所代表的定性化概念。例如,对于重磁测量值,虽然也关注重力值和磁场的强度,但更关注重力场或磁场的变化——梯度带的空间位置及其变化趋势;对于化探原生晕测量数据,指示元素的绝对含量往往需要转化成定性的前晕、后晕等概念;在遥感图像处理中,图像的明暗、色差等才是解译环线构造的关键依据;对于重砂测量数据,重砂矿物对矿产资源的指示意义关键要看矿物“是否存在”和“矿物组合”,而比较少地关注矿物颗粒的数量。

这些都是定性推理的典型实例,目前一般采用“人机对话”的方式,将专家的即时判断作为运算变量,再进行判断和统计计算。这种方法当然有许多优点,但其缺陷也不可忽略。因为专家的情绪是随时间变化的,而且在不断地进行某些修正,对于同一个现象,同一个专家在不同时间或不同专家在同一个时间点上判断会有一定的偏差,甚至得出相反的结论。也就是说,人机交互的主观随意性比较大。如果使用隶属云对地学中的定量数据做定性变换,则可消除这种主观随意性,同时避免专家的判断受其所在单位或行业的利益、所持学术观点等方面的不利影响。从本质上讲,云推理过程可以是确定性的,也可以是不确定性的。确定性的云推理是不确定性云推理的特例。推理的模式可以是“一对一”、“多对一”和“多对多”。这是应用云理论从地学数据库或数据仓库中进行知识挖掘的技术手段的基础。

知识挖掘的核心问题是知识的归纳和不确定性推理,其中归纳过程主要实现知识(规则)的获取,可以理解为一种建模过程,称为“云化过程”,形成的结果为“云规则”;而不确定推理过程则是将归纳出来的知识应用于未知对象的属性判别,也可以理解为一种预测过程,称为“云化计算”。尹国定等^[13]讨论了利用云计算实现概念计算的具体方法,给出了云计算逻辑的描述、云计算逻辑转换和云计算过程。

单条件(C_i)多规则(R_i)可形式化为 If C_1, C_2, \dots, C_m then R_1, R_2, \dots, R_m ; 多条件(C_{ij})多规则(R_i)可形式化为:

$$\left\{ \begin{array}{ll} C_{11} \cdots C_{1n} & R_1 \\ C_{21} \cdots C_{2n} & R_2 \\ \vdots & \vdots \\ C_{m1} \cdots C_{mn} & R_m \end{array} \right. \text{ If } \quad \text{ then } \quad$$

在多规则推理中,每一条规则都可视为一条单规则,推理的关键在于如何协调各条规则之间的相互关系。以往将每个条件都看作为“硬条件”并用“与”或者“或”来处理,而在云推理过程中,每个条件都是“软”的。每个条件都将转化成相应的“云团”,然后用几何云技术将若干个云团转化成理想条件下的“几何云”作为推理结论。

尹国定等^[13]将云计算过程分成云化过程和云计算两大步骤。其中,云化过程包括:①样本计算规则集、②数值变量的云化、③数值的云化、④计算的云化、⑤计算规则的云化、⑥计算的云规则的生成,并用“规则支持度 Sup”来定量估算云化过程和云推理过程的精度,这也是云计算优于模糊数学之处的一个方面。

4.3 隶属云在处理 GIS 的位置与属性不确定性中的应用前景

正确认识和处理 GIS 中的位置不确定性与属性不确定性具有重要意义^[14-16]。属性的不确定性是 GIS 数据精度的重要内容,主要反映在对矢量数据(点、线、面、体)或栅格数据的属性值的具体描述的不确定性,来源于数据采集、数据计算、数据传递等过程中的数据取舍、离散和抽象。目前处理 GIS 不确定性的理论与方法主要有概率论、证据理论、区域化变量理论、模糊集合、粗糙集理论、灵敏度分析和它们对应的方法。

云理论为处理这类问题提供了新的有效工具。根据云模型,可以构建定量数据和定性数据之间的相互映射关系,在充分考虑了定性概念的模糊性和随机性的基础上,进行空间关联规则的挖掘和空间数据库的不确定性查询。

4.4 云综合评判模型

云综合评判可以广泛应用于任何关于模糊评判的科技问题,例如大气、水、土壤、太空气候环境的模糊质量评判和地球化学、地球物理、遥感异常属性的模糊模式识别等。

表 2 模糊识别和隶属云模型对表 1 中 3 个异常的属性判定的差异

Table 2 Difference in ore-bearing potential recognition produced by fuzzy and membership cloud

异常编号	Au/ $10^{-9}\cdot\text{km}^{-2}$	As/ $10^{-6}\cdot\text{km}^{-2}$	Sb/ $10^{-6}\cdot\text{km}^{-2}$	模糊识别 的结论	隶属云法 的结论
未知异常 1	8.49	498.34	87.38	大型	大型
未知异常 2	11.34	299.03	14.38	大型	中型
未知异常 3	11.13	230.45	14.59	中型	中型

作为一个典型例子,利用隶属云模型,表 2 中 3 个异常分别被判定为大型、中型和中型,而用模糊理论则判定为大型、大型和中型,两者对第二个异常的判别结果相差一个等级,对矿床经济价值的判断具有重大差异。这对于进一步提高“5P 地段”(成矿可能、找矿可行、找矿有利、潜在资源、远景矿体的地段)的找矿效率也将产生一定的影响。

4.5 云网格数据计算平台

这方面的工作比较多地涉及到国家层面的、多部门之间的、多层次发展水平的异构计算机网络和数据库,涉及到比“网格计算”更复杂的计算。其研究内容更多地涉及到计算机领域,在此暂不作讨论。

5 结论与建议

综上所述,云理论作为中国科学家自主创新的理论研究成果,已经在空间广义知识和关联规则的挖掘、知识表达、不确定性推理、模式识别、故障诊断、决策、综合评判、聚类分析、自动控制和遥感影像解译、目标识别、智能控制、军事指挥等方面取得了良好的应用效果,在资源环境领域中也具有广阔的应用前景。

目前,云理论在资源环境领域中的适应性研究和应用研究刚刚开始,尽管如此,初步的应用成果已经说明,云理论与传统方法相比,既具有一定的理论优势,其应用成果也会有区别。因此笔者建议:①在资源环境领域逐步加强云模型、云变换和云推理的理论体系研究;②加强在资源环境领域对云模型、云变换和云推理的应用研究;③国家的

有关管理部门给予相关理论研究与应用研究以积极的支持。

参考文献

[1]Zadeh L A. Fuzzy Sets[J].Inform. Control., 1965, 8(3): 338-353.

[2]李德毅,孟海军,史雪梅. 隶属云与隶属云发生器[J]. 计算机研究与发展,1995,32(6):15-20.

[3]赵玉琛. 成矿预测中非参数蒙特卡洛模拟方法研究[J]. 矿产与地质, 1991,5(3): 191-199.

[4]赵玉琛. 矿产资源预测中非参数自动化蒙特卡洛模拟程序[J].物探与化探计算技术, 1991,13(2): 173-179.

[5]彭媛媛.区域矿产资源储量模拟方法研究[D].吉林大学硕士学位论文,2007.

[6]彭省临,刘亮明,赖健清,等. 大型矿山接替资源勘查技术与示范研究[M].北京:地质出版社,2004.

[7]姚姚.地球物理非线性反演模拟退火法的改进[J].地球物理学报, 1995,5:643-650.

[8]王家映.地球物理反演理论[M].北京:高等教育出版社,2002.

[9]李军,郝天兆.油气储层随机模拟方法综述[J].地球物理学进展, 2006,2:458-464.

[10]林殷,王宏崎,洪文. 基于本地图像块模型的人造目标提取方法[J]. 电子测量技术,2008,31(9):117-120.

[11]韩征. 河北省雄县地热信息系统的设计及地热资源评价[D].首都师范大学硕士学位论文,2009.

[12]张勇,赵东宁,李德毅.相似云及其度量分析方法[J]. 信息与控制, 2004, 33(2):129-132.

[13]尹国定,卫红. 云计算——实现概念计算的方法[J]. 东南大学学报, 2003,33(4): 502-502.

[14]史文中. 空间误差处理的理论与方法[M]. 北京: 科学出版社, 1998.

[15]刘文宝, 邓敏, 夏宗国. 矢量 GIS 中属性数据的不确定分析[J]. 测绘学报,2000,28(1):76-81.

[16]史文中, 王树良. GIS 数据之属性不确定性的研究[J]. 中国图像图形学报,2001,6(A,9):918-914.