

一种基于云模型的不确定性数据的建模与分类方法

秦丽^{1,2} 李兵^{1,3,4}

(武汉大学计算机学院软件工程国家重点实验室 武汉 430072)¹

(华中农业大学理学院 武汉 430070)² (武汉大学复杂网络研究中心 武汉 430072)³

(南京财经大学江苏省电子商务重点实验室 南京 210003)⁴

摘要 不确定性是数据的本质特征,它的产生可能来自于样本误差、更新延迟或重复测量等,对不确定性数据的分析在越来越多的领域得到了关注。传统的不确定性数据的数据项被表示成一个值域及其上的概率分布函数。由于不确定性数据存在模糊性与随机性,传统的概率分布函数难以准确定义不确定性数据的实际分布,因此利用云模型中云滴的分布提出一种不确定性数据的云建模过程,并通过云综合与云相似度计算来实现不确定性数据的分类。云模型能有效地将数据的随机性与模糊性融合在一起,能更真实地反映数据的实际分布,从而实现有效的数据分类。实验结果也证明了此种方法的有效性。

关键词 不确定性数据分类,高斯云,均匀云,云综合,云相似度计算

中图分类号 TP391.9 文献标识码 A DOI 10.11896/j.issn.1002-137X.2014.08.050

Novel Method of Uncertain Data Modeling and Classification Based on Cloud Model

QIN Li^{1,2} LI Bing^{1,3,4}

(State Key Laboratory of Software Engineering, School of Computer, Wuhan University, Wuhan 430072, China)¹

(School of Science, Huazhong Agricultural University, Wuhan 430070, China)²

(Research Center of Complex Network, Wuhan University, Wuhan 430072, China)³

(Jiangsu Provincial Key Laboratory of Electronic Business, Nanjing University of Finance & Economics, Nanjing 210003, China)⁴

Abstract Data contain inherent uncertainty. Sampling errors, data staling and repeated measurements are all the sources of uncertainty, so the analysis of uncertain data obtains more and more attention in many applications. The value of each traditional uncertain data is represented as domain over which a probability distribution function is defined. Because uncertainty data have fuzziness and randomness, and the traditional probability distribution function is difficult to define the actual distribution of the uncertainty data, this paper proposed a cloud modeling process of uncertainty data by the cloud drops distribution, and also designed a classify method by cloud union and similarity computing of cloud. Cloud model can effectively merge the randomness and fuzziness together, and can analyze uncertain data more effectively. For it's realistic reflection of actual distribution of the uncertain data, our experiments also prove the validity of this method.

Keywords Uncertain data classification, Gaussian cloud, Uniform cloud, Cloud union, Cloud similarity computing

1 引言

在许多需要进行数据分析的应用领域,如:无线传感器网络、无线射频识别(RFID)网络、移动对象跟踪、气象雷达网络和隐私保护等,逐渐涌现出一类特殊的数据,即不确定性数据。无论在科学研究领域、商业应用领域还是在军事、金融等应用中,数据的不确定性普遍存在^[1]。不确定性数据产生的原因可能来自于数据的反复测量与计算带来的误差,还有更新延迟或概念与数据的定性定量转换造成的不确定性等,比如:在传感器网络中,需要采集的温度、湿度、压力等特征值是

不断变化的;在移动对象跟踪时,对象的方位信息是间隔更新的,受监视器的精度及带宽的限制,所获得方位信息也往往是不精确的。因此对于不确定性数据的研究也得到了广泛的关注,包括不确定性数据的建模^[2]、不确定性数据的管理^[1]及不确定性数据的挖掘^[3-9],其中不确定性数据挖掘(uncertain data mining)在2005年人工智能研讨会上,作为一个新的研究方向被明确提出^[10]。而数据分类是机器学习与数据挖掘的经典问题,不确定性数据的分类自然成了学者研究的热点。近年来研究得比较多的方法主要有基于决策树的分类方法^[3-5]、基于规则的分类方法^[6,7]与基于朴素贝叶斯的分类方

到稿日期:2013-09-21 返修日期:2014-02-27 本文受国家自然科学基金(61273216,61272111,61202032,61202048),湖北省重大科技创新计划(2013AAA020),江苏省电子商务重点实验室开放基金(JSEB2012-02)资助。

秦丽(1978—),女,博士生,讲师,主要研究方向为人工智能、数据挖掘, E-mail: qinli@whu.edu.cn; 李兵(1969—),男,博士,教授,主要研究方向为软件工程、复杂网络。

法^[8,9],这些方法都是基于概率论提出的,计算复杂度高且未能充分考虑不确定性数据的模糊性,数据的不确定性包括随机性、模糊性、灰性(部分已知,部分未知的不确定性)、未知性(是纯主观上,认识上的不确定),其中随机性、模糊性是数据不确定性最重要的两个内容,它们将共同决定不确定性数据的实际分布,比如不确定性数据被研究最多的分布是高斯分布与均匀分布,但由于不确定数据的随机性与模糊性,使得它们的实际分布大多表现出近似高斯分布或近似均匀分布的特点,这使得传统的概率密度函数不能准确地定义不确定性数据的分布,而云模型正是同时考虑了随机性与模糊性的不确定性概念认识计算模型^[11],云滴可通过超熵的变化达到近似高斯分布或均匀分布的效果,所以本文着重研究如何运用云滴来对不确定性连续数据进行建模,并根据云模型的相关理论提出了一套不确定性连续数据的分类方法。

2 相关研究

云模型自提出后,经过十多年的研究发展已成功应用于智能决策与控制、系统评测与信任评估、云水印、自动摘要提取、机器学习算法改进、数据挖掘等领域。近年来也有学者研究了基于云模型的分类型器设计^[12-19],其中刘禹^[12]设计了适用于多属性数据集分类的云分类器;张国英、刘达^[13,14]先后研究了基于属性相似度的云分类器;石礼娟^[15,16]将云分类器用于大米的蛋白识别;皮二旭^[17,18]提出了粒子群优化云分类器与模拟退火半自动云分类器,并将之应用于山茶属植物数值分类中,部分茶组的分类正确率可达85%,秦昆^[19]将之用于图像分割的领域中。但上述研究都是针对确定性数据的分类方法,对于不确定性数据的分类方法尚未发现。而目前研究较多的不确定性数据的分类方法主要有基于决策树的分类方法、基于规则的分类方法与基于朴素贝叶斯的分类方法,其中Tsang^[3]与Qin^[4]相继提出了基于决策树的不确定性数据分类方法,决策树使用属性选择度量来将元组最好地划分成不同类的属性,而决策树的构造过程就是选择一个属性通过判定条件将元组划分到更小的类别子集中,传统的属性选择度量不能用于不确定性数据,Qin根据概率势、扩展熵、信息增益、信息增益率定义了概率熵、概率信息增益与概率信息增益率,从而扩展C4.5决策树用于分类处理不确定性数据。Tsang研究了不确定连续数据的决策树分类方法,即通过一组在不确定区域 $[a_i, b_i]$ 中的样本点 x 及其离散分布 $f(x)$ 来近似连续数据的概率密度函数,并基于此提出了平均方法与基于分布的方法。通过实验证实了基于分布的方法的分类准确率高于平均方法,然而,基于分布的方法的计算复杂度高,所以为降低不确定决策树算法的计算复杂度,周帅印^[5]在Qin的研究基础上,进一步提出不确定独立信息增益的概念,并据此提出快速不确定决策树算法。此外Qin^[6,7]还提出了基于规则的不确定性数据分类,其关键问题是生成的规则能根据概率信息增益有效地从训练集中选择最好的分裂属性与分裂点,并可对规则进行概率修剪。J. Ren^[8]与Qin^[9]还相继研究了不确定性连续数据的朴素贝叶斯分类,J. Ren提出了3种估计类条件密度的方法,分别是平均方法、基于样本的方法与基于公式的方法,其中前两种方法类似于Tsang的研究,而基于公式的方法是根据某些特别的核函数与概率密度推导出一个计算公式。而Qin在假设不确定性连续数据是高斯分布的前提下,提出两种计算条件概率的期望和方差参数的方法。

以上方法都未考虑当不确定性数据在分布上表现出近似高斯分布或近似均匀分布时的处理方法,所以本文下面的章节将讨论如何用高斯云与均匀云的云滴来建立不确定性数据的模型,并利用云综合与云相似度计算来进行不确定性数据的分类。

3 云模型

3.1 云模型

云模型用期望、熵和超熵这3个数字特征来定量描述一个不确定性概念,通过正向云发生器与逆向云发生器实现定性概念与定量数据之间的双向认知变换,揭示客观对象具有的模糊性与随机性,刻画人类的双向认知计算过程^[11]。

3.2 云数字特征与云分布的关系

云模型具体实现方法可以有多种,基于不同的概率分布可以构成不同的云,如基于均匀分布的均匀云、基于高斯分布的高斯云、基于幂律分布的幂律云等^[20]。因为不确定数据在其值域上被研究得最多的分布是高斯分布与均匀分布,所以本文主要讨论高斯云与均匀云,其中高斯云是研究与应用最为广泛的一种云模型,高斯分布的普适性与钟型隶属函数的普适性,共同奠定了高斯云的普适性^[21]。

3.2.1 高斯云的数字特征

高斯云是基于二阶高斯分布迭代来实现的云模型,它有3个数字特征,可表示为 (Ex, En, He) 。 Ex 是论域空间中最具代表性的数值,即期望。熵 En 反映了论域中可被接受的数域范围,是标准差的期望。超熵 He 是标准差的标准差,熵和超熵共同构成了高斯云偏离高斯分布的衡量指标^[11]。以下是通过3个数字特征生成高斯云的一维正向云发生器算法^[22]。

算法1 一维正向高斯云算法

输入:3个数字特征,生成云滴的个数 N

输出: N 个云滴 x 及其确定度(也可表示为 $\text{Drop}(x_i, y_i), i=1, \dots, N$)

算法步骤:

(1)以 En 为期望, He 为标准差,生成一个正态随机数熵 En' ;

(2)以 Ex 为期望, En' 为标准差,生成一个正态随机数 x_i ;

(3)计算确定度

$$y_i = e^{-\frac{(x_i - Ex)^2}{2(En')^2}}$$

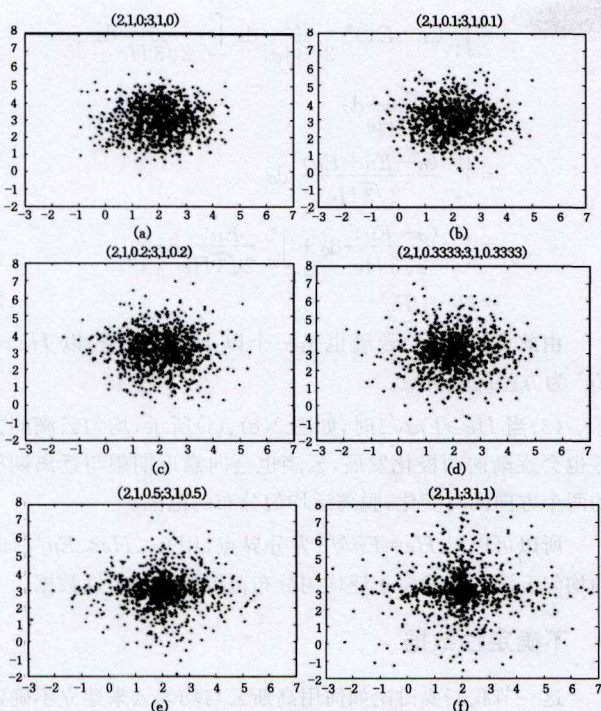
具有确定度 y_i 的 x_i 称为论域 U 上的一个云滴;

(4)重复步骤(1)–(3),直至产生 N 个云滴。

一维云算法可扩展到二维云或多维云算法^[23],多维云的数字特征可定义为期望 $(Ex_1, Ex_2, \dots, Ex_n)$,熵 $(En_1, En_2, \dots, En_n)$,超熵 $(He_1, He_2, \dots, He_n)$ 。

高斯云滴的分布与高斯分布相比,通过引入超熵来衡量标准差的不确定性,使高斯云滴逐渐偏离高斯分布,所以 He 的变化能反映云滴偏离高斯分布的程度,如图1所示。

图1(a)的 $He=0$,所以它的云滴呈高斯分布;图1(b)(c)的 $He>0$ 且 $He<En/3$,云滴的分布近似高斯分布,可称之为泛高斯分布;图1(d)的 $He=En/3$,此时高斯云滴表现出外部更加发散、内部更加集中的特点,开始脱离泛高斯分布的范围;图1(e)(f)的 He 进一步变大,云滴向两极化发展的特点越来越明显,也就是云滴向期望与周边同时发展,这时云滴完全脱离泛高斯分布的范围。经研究,这一变化与熵和超熵之比有对应关系^[24],下面给出熵与超熵之比对高斯云滴分布特征的影响。



(每张子图中的数字代表云滴的二维云数字特征($Ex_1, En_1, He_1; Ex_2, En_2, He_2$))

图1 二维高斯云滴分布图

3.2.2 高斯云分布

在论域上的高斯云分布的概率密度函数为^[11]:

$$f(x) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-Ex)^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi}He} e^{-\frac{(\sigma-En)^2}{2He^2}} d\sigma \quad (1)$$

其中, σ 是一个以 En 为期望, He 为标准差的高斯随机数, $f(x)$ 定义了 x 服从高斯云分布。

(1) 当 $He=0$ 时, x 的概率密度分布函数为 $f(x) = \frac{1}{\sqrt{2\pi}En} e^{-\frac{(x-Ex)^2}{2En^2}}$, 是一个高斯分布。

(2) 当 $0 < He < En/3$ 时(即 99.7% 的 σ 的取值为正), 高斯云的云滴逐渐偏离高斯分布, 呈泛高斯分布状态。高斯云分布的期望、一阶绝对中心距、方差分别为:

$$E\{X\} = Ex \quad (2)$$

$$E\{|X-Ex|\} = \sqrt{\frac{2}{\pi}} En \quad (3)$$

$$D(X) = He^2 + En^2 \quad (4)$$

所以, 这时的高斯云的云滴集合是一个以 Ex 为期望、以 $He^2 + En^2$ 为方差的随机变量集^[11]。

(3) 当 $He \geq 1/3 En$ 时, 高斯云滴的方差会逐渐向两极化发展, 云滴也会向靠近期望与远离期望的两个方向同时变化, 彻底脱离泛高斯分布的范围。

所以刘禹^[25]将 $3He/En$ 作为云模型是否属于泛高斯分布的一个拐点^[25]。

从上面的分析可知满足高斯云分布的云滴在 $He=0$ 时是一个高斯分布, 而在 $0 < He < En/3$ 时, 呈泛高斯分布状态, 用它来描述属于高斯分布范围的不确定性数据是十分合适的。

3.2.3 均匀云

均匀云是描述基于连续均匀分布的云模型。但是均匀云由于没有高斯云应用广泛, 因此尚未有人对其生成算法及分布特点进行研究, 所以本节重点讨论均匀云的生成算法及分

布特点。

均匀云也可表示成由 (Ex, En, He) 定义的三元数字特征, 各特征值的意义与高斯云类似, 只是熵和超熵变成共同构成均匀云偏离均匀分布的衡量指标, 下面讨论均匀云的实现方法。

设 x 服从 $[a, b]$ 上的均匀分布(uniform distribution)记作 $x \sim U[a, b]$, 而其概率密度函数为:

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{elsewhere} \end{cases} \quad (5)$$

因为均匀分布的期望 μ 与方差 σ^2 分别为:

$$\mu = \frac{a+b}{2} \quad (6)$$

$$\sigma^2 = \frac{(b-a)^2}{12} \quad (7)$$

联立式(6)、式(7), 可求出:

$$a = \mu - \sqrt{3}\sigma \quad (8)$$

$$b = \mu + \sqrt{3}\sigma \quad (9)$$

再结合式(5)概率密度函数可表示为:

$$f(x) = \frac{1}{2\sqrt{3}\sigma} \quad (10)$$

根据上面的推导, 算法2给出正向均匀云的生成过程。

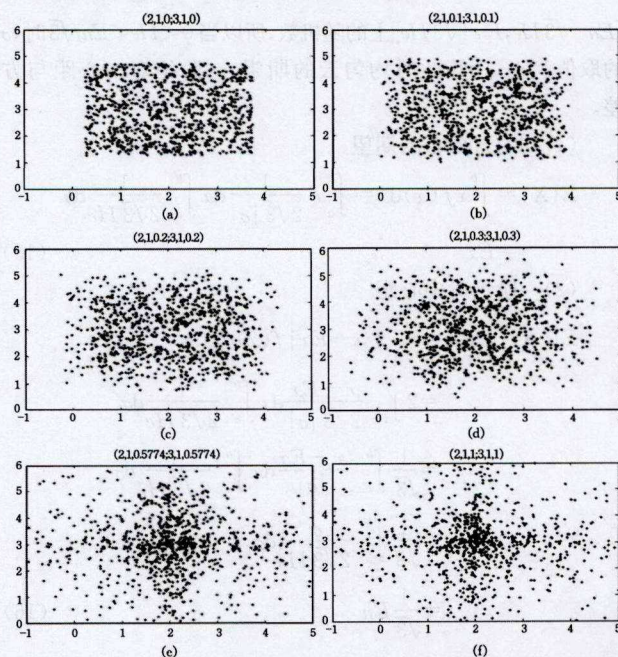
算法2 一维正向均匀云算法

输入: 3个数字特征(Ex, En, He), 生成云滴的个数 N

输出: N 个云滴(也可表示为 $\text{Drop}(x_i), i=1, \dots, N$)

算法步骤:

- (1) 根据式(10)与式(11), 以 En 为期望, He 为标准差, 计算 $a' = En - \sqrt{3}He$, $b' = En + \sqrt{3}He$
- (2) 生成 $[a', b']$ 上的一个均匀随机数 En' ;
- (3) 再以 Ex 为期望, $|En'|$ 为标准差, 计算 $a = Ex - \sqrt{3}|En'|$, $b = Ex + \sqrt{3}|En'|$
- (4) 生成 $[a, b]$ 上的一个均匀随机数 x_i , x_i 称为论域 U 上的一个均匀云滴;
- (5) 重复步骤(1)~(4), 直至产生 N 个云滴。



(每张子图中的数字代表云滴 x 的二维云数字特征($Ex_1, En_1, He_1; Ex_2, En_2, He_2$))

图2 二维均匀云滴分布图

算法 2 也可以扩展到二维,扩展方法类似高斯云。算法 2 中并未为云滴定义确定度,因为本文只是利用云滴来描述不确定性数据而非不确定性概念,所以暂不考虑云滴对概念的隶属程度。图 2 为二维均匀云滴的分布图。图 2(a)的 $He=0$,云滴服从均匀分布;图 2(b)、(c)、(d)的 $He>0$ 且 $He<En/\sqrt{3}$,从图中可以发现均匀云滴逐渐向边缘扩散,但在 $[Ex-\sqrt{3}En, Ex+\sqrt{3}En]$ 的区间内都满足近似均匀分布,也可称为泛均匀分布;直到图 2(e)的 $He=En/\sqrt{3}$ 时,均匀云滴也表现出两极化的特点,即向外围与期望处同时发展,脱离泛均匀分布的范围;图 2(f)的 He 进一步扩大,使两极化的特点更加明显,云滴也完全脱离了泛均匀分布的范围。

3.2.4 均匀云分布

下面讨论均匀云滴分布和熵与超熵的关系。根据式 (10),以 En 为期望、 He 为标准差的均匀随机数 S 的概率密度函数为:

$$f(s) = \frac{1}{2\sqrt{3}He} \quad (11)$$

根据算法 2,在 $S=\sigma$ 的条件下, x 的概率密度函数为:

$$f(x) = \int_{a'}^{b'} f(x|s=\sigma) f(\sigma) d\sigma \quad (12)$$

由算法 2 的步骤(3)、(4),可知

$$f(x|s=\sigma) = \frac{1}{2\sqrt{3}|\sigma|} \quad (13)$$

将式(11)、式(13)代入式(12),可得

$$f(x) = \int_{a'}^{b'} \frac{1}{2\sqrt{3}|\sigma|} \frac{1}{2\sqrt{3}He} d\sigma \quad (14)$$

通过上面的推导,我们知道 $f(x)$ 就是均匀云产生云滴的概率密度分布函数,定义了 x 服从均匀云分布。

(1) 当 $He=0$ 时, x 的概率密度分布函数为 $f(x) = \frac{1}{2\sqrt{3}En}$,云滴服从均匀分布。

(2) 当 $0<He<En/\sqrt{3}$ 时,云滴逐渐偏离均匀分布,因为 σ 是 $[En-\sqrt{3}He, En+\sqrt{3}He]$ 上的随机数,所以当 $0<He<En/\sqrt{3}$ 时, σ 的取值为正,计算此时均匀云的期望、一阶绝对中心距与方差。

(1) 均匀云分布的期望

$$E\{X\} = \int_a^b x f(x) dx = \int_a^b x \frac{1}{2\sqrt{3}|\sigma|} dx \int_{a'}^{b'} \frac{1}{2\sqrt{3}He} d\sigma = Ex \quad (15)$$

(2) 一阶绝对中心距

$$\begin{aligned} E\{|X-Ex|\} &= \int_a^b |x-Ex| f(x) dx \\ &= 2 \int_{Ex}^b \frac{x-Ex}{2\sqrt{3}|\sigma|} dx \int_{a'}^{b'} \frac{1}{2\sqrt{3}He} d\sigma \\ &= \frac{1}{\sqrt{3}} \int_{Ex}^b \frac{x-Ex}{|\sigma|} dx \int_{a'}^{b'} \frac{1}{2\sqrt{3}He} d\sigma \\ &= \frac{1}{\sqrt{3}} \int_{a'}^{b'} \frac{\sigma}{2\sqrt{3}He} d\sigma \\ &= \frac{1}{\sqrt{3}} En \end{aligned} \quad (16)$$

(3) 方差

$$D(X) = \int_a^b (x-Ex)^2 f(x) dx$$

$$\begin{aligned} &= \int_a^b (x-Ex)^2 \frac{1}{2\sqrt{3}|\sigma|} dx \int_{a'}^{b'} \frac{1}{2\sqrt{3}He} d\sigma \\ &= \int_{a'}^{b'} \frac{\sigma^2}{2\sqrt{3}He} d\sigma \\ &= \int_{a'}^{b'} \frac{(\sigma-En+En)^2}{2\sqrt{3}He} d\sigma \\ &= \int_{a'}^{b'} \frac{(\sigma-En)^2}{2\sqrt{3}He} d\sigma + \int_{a'}^{b'} \frac{En^2}{2\sqrt{3}He} d\sigma + 0 \\ &= He^2 + En^2 \end{aligned} \quad (17)$$

由此可见,均匀云滴也是一个以 Ex 为期望、以 $He^2 + En^2$ 为方差的随机数。

(3) 当 $He \geq En/\sqrt{3}$ 时,如图 2(e)、(f) 所示,均匀云滴的方差也会逐渐向两极化发展,云滴也会向靠近期望与远离期望的两个方向同时变化,脱离泛均匀分布的范围。

所以可以以 $He=En/\sqrt{3}$ 为分界点,用 $0<He<En/\sqrt{3}$ 时的均匀云滴来描述属于泛均匀分布范围的不确定性数据。

4 不确定性数据

这一节我们要讨论如何用高斯云与均匀云来建立不确定性数据的模型。不确定性数据由不确定的属性值与不确定的类标签组成,如果不确定属性的取值是一个数字区间,则称其为不确定数值属性(UNA),如果是类别数据则称为不确定类别属性(UCA)^[6]。本文主要关注数值属性的不确定性,而类别属性的不确定性暂不考虑,由于云模型是定性定量转换模型,要求数据是连续的数字,因此下文的讨论与分析都是针对 UNA 的。

4.1 不确定性数值数据的模型

设不确定性数据 S 有 n 条属性,表示为 $A_i (i=1 \cdots n)$,而 A_{ij} 则表示第 j 个元组的第 i 个属性值。 A_{ij} 的取值不是一个精确值,而被表示成一个满足一定概率分布函数(PDF)的数字区间 $[a_{ij}, b_{ij}]$, A_{ij} 则是这个取值范围中的一个连续随机变量,每个 A_i 可以拥有相同的 PDF,也可能有不同的 PDF^[9],表 1 给出了一个不确定性数值数据的例子。

表 1 不确定性数据

ID	A_i	...	A_j	class
1	100~120	...	Married	NO
2	60~85	...	Single	NO
3	110~120	...	Divorced	YES
4	50~80	...	Married	NO
5	85~100	...	single	YES
6	80~100	...	Married	NO
7	120~145	...	single	YES
8	105~125	...	Divorced	YES

从表 1 中可以发现 A_i 是不确定性数据,它的取值是一个数字区间,而这个区间上的 PDF 可以是均匀分布或高斯分布。

4.2 基于高斯云的连续不确定性数据的模型

本节提出两种基于高斯云的不确定性连续数据的构建方法:一种是基于区间的方法;一种是基于样本的方法。

(1) 基于区间的方法是利用不确定性数据取值区间的最大值 $V_{\max}$ 与最小值 $V_{\min}$ 来计算云的期望、熵与超熵。首先由式(2)可知期望的计算可通过数据值区间的均值来表示,即

$$Ex = \frac{V_{\max} + V_{\min}}{2} \quad (18)$$

再则,由 3 倍标准差原则可设计算方差的公式为:

$$D(x) = \left(\frac{V_{\max} - V_{\min}}{6} \right)^2 \quad (19)$$

结合式(4)可以得出:

$$En^2 + He^2 = \left(\frac{V_{\max} - V_{\min}}{6} \right)^2 \quad (20)$$

由高斯云满足泛高斯分布的要求,可设 $He = En/n (n > 3)$,代入式(20) 可得:

$$En = \frac{n}{6} \sqrt{\frac{(V_{\max} - V_{\min})^2}{n^2 + 1}} \quad (21)$$

$$He = \frac{1}{6} \sqrt{\frac{(V_{\max} - V_{\min})^2}{n^2 + 1}} \quad (22)$$

当 $n > 3$ 时,数值概率呈泛高斯分布。将表 1 中的数据根据上述方法进行转换($n=10$),结果如表 2 所列。

表 2 用高斯云表示的不确定性数据

ID	A_i	...	A_j	class
1	(110, 3.32, 0.332) ...		Married	NO
2	(72.5, 4.15, 0.415) ...		Single	NO
3	(115, 1.66, 0.166) ...		Divorced	YES
4	(65, 4.98, 0.498) ...		Married	NO
5	(92.5, 2.49, 0.249) ...		single	YES
6	(90, 3.32, 0.332) ...		Married	NO
7	(132.5, 4.15, 0.415) ...		single	YES
8	(115, 3.32, 0.332) ...		Divoced	YES

(2)基于样本的方法是根据 Tsang 的研究将不确定性连续数据通过一组在不确定区域 $[a_i, b_i]$ 中的样本点 x 及其离散分布 $g(x)$ 来近似连续数据的概率密度函数 $f(x)$ 进而计算高斯云的期、熵与超熵的方法。此前已有人提出了高斯云模型的逆向云发生器算法,它也是根据样本的分布来倒推云的数字特征,主要有吕辉军^[27]提出的有确定度的逆向云算法与刘长昱^[28]提出的无确定度逆向云算法,还有最近王立新根据高阶距提出的逆向云新算法及王国胤提出的基于分组降阶的逆向云新算法^[11],但这些算法都没有考虑利用样本点的离散分布 $g(x)$ 来计算云数字特征。所以本文提出一种利用样本离散分布来近似连续数据的概率密度的逆向云新算法。

设有样本点 x ,根据算法 1,设它的离散分布 $g(x)$ 等于云滴熵为 σ 下的条件概率,即

$$g(x) = f(x|En' = \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-Ex)^2}{2\sigma^2}}$$

由此可以推出

$$\sigma = \sqrt{\frac{-(x-Ex)^2}{2(\ln \sqrt{2\pi}g(x) + \ln \sigma)}} \quad (23)$$

所以新的逆向云算法具体如下所示。

算法 3 利用样本离散分布的高斯逆向云算法

输入:样本点 x_i 及其离散分布 y_i ,其中 $i=1,2,\dots,n$ 。

输出:反映定性概念的数字特征(Ex,En,He)。

算法步骤:

- (1)计算 x_i 的平均值 $Ex = \text{MEAN}(x_i)$;
- (2)计算 x_i 的标准差 $En = \text{STDEV}(x_i)$;
- (3)对每一数对 (x_i, y_i) ,计算

$$En_i' = \sqrt{\frac{-(x_i - Ex)^2}{2(\ln \sqrt{2\pi}y_i + \ln En_i')}};$$

- (4)计算 En_i' 的标准差 $He = \text{STDEV}(En_i')$ 。

其中,MEAN 和 STDEV 分别为求样本均值和样本标准差的

函数。

4.3 基于均匀云的连续不确定性数据模型

基于均匀云的连续不确定性数据模型的构建方法也可分为基于区间的方法与基于样本的方法。

(1)基于区间的方法

设不确定性数据的取值区间的最大值为 $V_{\max}$,最小值为 $V_{\min}$,因为均匀云滴满足泛均匀分布,所以根据式(6)、式(7)与式(17),可推导(Ex,En,He)的计算公式。

$$Ex = \frac{V_{\min} + V_{\max}}{2} \quad (24)$$

$$En^2 + He^2 = \frac{(V_{\max} - V_{\min})^2}{12} \quad (25)$$

根据泛均匀分布的要求,设 $He = En/n (n > \sqrt{3})$,则由式(25)可得:

$$En = n \sqrt{\frac{(V_{\max} - V_{\min})^2}{12(n^2 + 1)}} \quad (26)$$

$$He = \sqrt{\frac{(V_{\max} - V_{\min})^2}{12(n^2 + 1)}} \quad (27)$$

将表 1 中的数据根据式(24)、式(26)、式(27)进行转换,当 $n=5$ 时结果如表 3 所列。

表 3 用均匀云表示的不确定性数据

ID	A_i	...	A_j	class
1	(110, 5.661, 1.132) ...		Married	NO
2	(72.5, 7.077, 1.415) ...		Single	NO
3	(115, 2.831, 0.566) ...		Divorced	YES
4	(65, 8.492, 1.698) ...		Married	NO
5	(92.5, 4.246, 0.849) ...		single	YES
6	(90, 5.661, 1.132) ...		Married	NO
7	(132.5, 7.077, 1.415) ...		single	YES
8	(115, 5.661, 1.132) ...		Divoced	YES

(2)基于样本的方法

根据前文提出的正向均匀云算法,下面提出通过不确定区域 $[a_i, b_i]$ 中的样本点 x 及其离散分布 $g(x)$ 来倒推(Ex,En,He)的算法,即逆向均匀云算法。

设有样本点 x ,且设它的离散分布 $g(x)$ 等于云滴熵为 σ 下的条件概率,根据算法 2 与式(13)有下式成立。

$$g(x) = f(x|En' = \sigma) = \frac{1}{2\sqrt{3}|\sigma|}$$

那么可推出

$$\sigma = \frac{1}{2\sqrt{3}g(x)} \quad (28)$$

所以均匀云的逆向云算法描述如下。

算法 4 逆向均匀云算法

输入:样本点 x_i 及其离散分布 y_i ,其中 $i=1,2,\dots,n$ 。

输出:反映定性概念的数字特征(Ex,En,He)。

算法步骤:

- (1)计算 x_i 的平均值 $Ex = \text{MEAN}(x_i)$;
- (2)计算 x_i 的标准差 $En = \text{STDEV}(x_i)$;
- (3)对每一数对 (x_i, y_i) ,计算 $En_i' = \frac{1}{2\sqrt{3}y_i}$;
- (4)计算 En_i' 的标准差 $He = \text{STDEV}(En_i')$ 。

其中,MEAN 和 STDEV 分别为求样本均值和样本标准差的函数。

5 基于云的不确定性数据分类

完成基于云模型的不确定性数据建模后,就可使用它来进行不确定性数据的分类。本文提出一种新的基于云模型的不确定性分类方法,这种分类方法主要包括 3 个步骤:云综合、属性权重计算及类别判定。

(1)云综合是对训练集进行建模。设不确定性数据训练集 S 有 P 个类,每类表示为 $S_\alpha (\alpha=1, \dots, P)$,有 n 条属性,表示为 $A_i (i=1, \dots, n)$,而 A_{ij} 则表示 S 中第 $j (j=1, \dots, m)$ 条元组的第 i 个属性,根据上节提出的方法,若 A_{ij} 为不确定性数据,则可以将 A_{ij} 的数值区间 $[a_{ij}, b_{ij}]$ 转换成由 $(Ex_{ij}, En_{ij}, He_{ij})$ 表示的数值云,云综合是将同类数据 S_α 中所有 A_{ij} 的数值云综合成一个代表 A_i 的属性云 (Ex_i, En_i, He_i) ,可以用 $(Ex_i, En_i, He_i)_\alpha$ 表示经云综合后得到的 α 类的第 i 个属性云。

(2)属性权重计算是计算各属性参与分类的权重。对属性 A_i ,通过计算各类别间属性云 $(Ex_i, En_i, He_i)_\alpha (\alpha=1, \dots, P)$ 的相似度 FS_i 来获取属性 A_i 的分类权重 W_i ,云相似度越大则权重越小,而属性 A_i 的权重计算公式为:

$$W_i = 1 - FS_i \quad (29)$$

(3)类别判定是对未知类别的不确定性数据进行分类。设有不确定性数据 D ,它也有 n 条属性,表示为 D_i ,将 D_i 转化成由 (Ex_i, En_i, He_i) 表示的数值云。分类时,将数值云分别与 S_α 对应的属性云 $(Ex_i, En_i, He_i)_\alpha$ 进行云相似度计算,从而得到 D_i 对 S_α 中 A_i 的相似度,表示为 $(sim_i)_\alpha$,再由式(30)计算 D 属于 S_α 的确定度 U_α 。

$$U_\alpha = \sum_{i=1}^n ((sim_i)_\alpha \times W_i) \quad (30)$$

(4)最后通过极大值判定法 $\text{MAX}(U_\alpha)$ 判定 D 所属的类别。

从上述步骤可以看出,这里用到了两个算法,1)云综合算法,2)云相似度计算方法。下面将详细介绍这两种算法。

5.1 云综合算法

云综合^[29]就是将两朵云合成一朵云的过程,云综合是概念跃升^[30]的方法,概念跃升就是指在泛概念树叶节点的基础上,直接将概念提升到所需的概念粒度或概念层次,“软或”^[31]是一种实现高斯云综合的方法,对于给定的两个相邻的高斯云 $C_1 (Ex_1, En_1, He_1)$ 、 $C_2 (Ex_2, En_2, He_2)$,如果 $Ex_1 \leq Ex_2$,那么合并后得到新的高斯云为 $C_3 = C_1 \cup C_2$,具体计算过程如下:

$$(1) Ex_3 = \frac{Ex_1 + Ex_2}{2} + \frac{En_2 - En_1}{4} \quad (31)$$

$$(2) En_3 = \frac{Ex_2 - Ex_1}{4} + \frac{En_1 + En_2}{2} \quad (32)$$

$$(3) He_3 = \max(He_1, He_2) \quad (33)$$

此外,王立新于 2011 年提出了云综合更加宽泛的计算方法,包括: $C_3 = C_1 \pm C_2$, $C_3 = C_1 \times C_2$, $C_3 = C_1 / C_2$, $C_3 = \max(C_1, C_2)$, $C_3 = \min(C_1, C_2)$, $C_3 = \sqrt{C_1^2 + C_2^2}$ 。

本文使用云综合将不确定属性 A_i 的若干数值云综合成能反映属性模型的属性云,相当于由实例上升到概念的过程,这也是一种概念提升的操作。至于云综合的方法,本文使用“软或”作为高斯云综合的主要手段。而均匀云的云综合方法尚未提出,根据高斯云“软或”的原理,本文给出了均匀云“软

或”的公式。设有均匀云 $C_1 (Ex_1, En_1, He_1)$ 与 $C_2 (Ex_2, En_2, He_2)$,综合云 $C_3 = C_1 \cup C_2$,则有下式成立。

$$\begin{cases} Ex_3 - \sqrt{3} En_3 = Ex_1 - \sqrt{3} En_1 \\ Ex_3 + \sqrt{3} En_3 = Ex_2 + \sqrt{3} En_2 \end{cases}$$

求解方程组可得出 Ex_3 与 En_3 的计算公式,具体如下:

$$(1) Ex_3 = \frac{Ex_1 + Ex_2}{2} + \frac{En_2 - En_1}{\sqrt{3}} \quad (34)$$

$$(2) En_3 = \frac{Ex_2 - Ex_1}{\sqrt{12}} + \frac{En_1 + En_2}{2} \quad (35)$$

而 He_3 的求解方法同式(33)。

5.2 云相似度计算

云相似度计算是为了计算由云描述的两个不确定性概念之间的相似程度,而高斯云与均匀云都是由 (Ex, En, He) 来定义的,因为云的相似度计算主要是通过 3 个数字特征的比较计算来实现的。云相似度计算在本文提出的不确定性数据分类过程中用到了两次,一次在计算属性权重时,通过不同类别间属性云的相似度来判定该属性在分类过程的权重,云相似度越大,权重就越小^[12]。另一次是在分类时,通过一条数据各属性的属性云与各类别对应属性云的相似度来判定该数据的类别。这两次使用虽然都是进行云相似度的比较,但计算权重时需要在多朵云之间进行比较,获取总体相似度,而分类时只需要在两两之间计算即可,所以下文首先讨论两朵云之间的相似度的计算方法。刘禹在文献[12]提出了一种高斯云的相似度计算方法,具体内容如下。

高斯云的相似度由内涵相似度与外延相似度组成,其中内涵相似度取决于云核间的距离,而外延相似度取决于云边沿的交叠度。其中云核的距离公式为:

$$\begin{aligned} Dis(C_1, C_2) &= \begin{cases} 1, & |Ex_1 - Ex_2| > 3(En_1 + En_2) \\ \frac{|Ex_1 - Ex_2|}{3(En_1 + En_2)}, & |Ex_1 - Ex_2| \leq 3(En_1 + En_2) \end{cases} \\ \alpha &= 1 - Dis(C_1, C_2) \end{aligned} \quad (36)$$

其中, $Dis(C_1, C_2)$ 代表云核距离,距离越大,相似度越小。云边沿的交叠度的定义为:

$$\beta = e^{-|He_1/En_1 - He_2/En_2|} \quad (37)$$

其中, $\frac{He}{En}$ 代表云滴偏离高斯分布的程度,通过计算两朵云偏离高斯分布程度之差,可以得出它们在不确定性上的差距,不确定性差距越大,云边缘交叠度就越小。

由于云核距离的作用大于云边沿交叠度,因此高斯云相似度的合成公式为:

$$\text{SIM}(C_1, C_2) = \frac{a\beta + b\alpha}{1 + b} (b > 0) \quad (38)$$

下面再来讨论均匀云的相似度计算公式,我们认为均匀云的相似度应由云滴交叠度与偏离均匀分布的相似度组成。

设有两朵均匀云 C_1 与 C_2 ,它们的覆盖范围为 $[a_1, b_1]$ 与 $[a_2, b_2]$,其中 $a_2 \geq a_1$,如果 $b_1 \leq a_2$,则两朵云形如图 3(a)所示,它们没有交集,完全独立,所以距离为 1;如果 $b_1 \geq a_2$,则如图 3(b)所示,两朵云有交集,距离小于 1;如果 $b_2 \leq b_1$,则两朵云完全交叠,如图 3(c)所示,距离为 0。

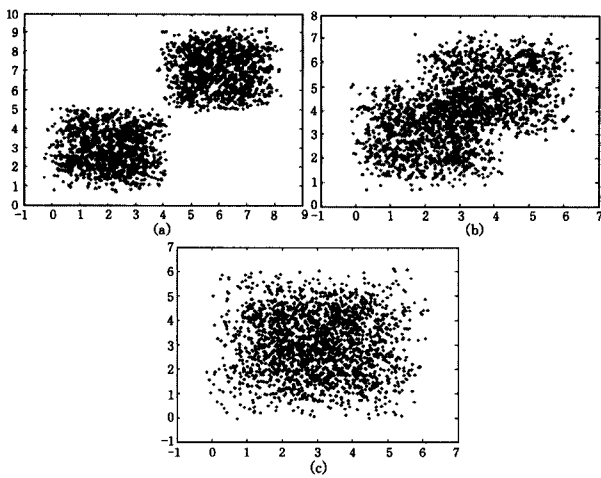


图3 二维均匀云滴重叠度图

根据上面的分析及式(8)、式(9),可推出云滴的距离为:

$$Dis(C_1, C_2) = \begin{cases} 1, & |Ex_2 - Ex_1| > \sqrt{3}(En_1 + En_2) \\ \frac{|Ex_1 - Ex_2|}{\sqrt{3}(En_1 + En_2)}, & |Ex_2 - Ex_1| \leq \sqrt{3}(En_1 + En_2) \\ 0, & |Ex_2 - Ex_1| \leq \sqrt{3}(En_1 - En_2) \end{cases} \quad (39)$$

余下的计算过程同式(36)。此外根据算法2可知均匀云偏离均匀分布的程度也由 $\frac{He}{En}$ 决定,因此 C_1 与 C_2 偏离相似度也可以由式(37)定义。而最终的均匀云相似度合成公式同式(38)。但式(38)只实现了两朵云之间的相似度比较,多个云之间的总体相似度由式(40)来计算。

$$FS = \frac{\sum_{i=1}^{n-1} \sum_{j=2}^n SIM(C_i, C_j)}{n(n-1)} \quad (40)$$

6 实验

我们用 matlab 实现了上述模型及分类算法,并通过 8 组 UCI 数据集¹⁾对基于高斯云与均匀云的不确定数据进行了分类实验,实验所用的机器配置是 Genuine Interl(R) 1.3GHZ CPU 及 2.0GB 内存。

表4 实验中使用的 UCI 数据集

编号	数据集	属性条数	类别数	记录数
1	diabetes	8	2	768
2	glass	9	6	214
3	heart-s	13	2	270
4	ionosphere	34	2	351
5	iris	4	3	150
6	sonar	60	2	208
7	vehicle	18	4	846
8	wine	13	3	178

实验中使用的数据集如表4所列,为方便表达,给每个数据集一个编号,后面的实验结果均采用这个编号来对应其数据集。由于实验中需要用到不确定性数据,因此在实验之前,首先需要将所有数据中的精确数据转成不确定性数据,转换的方法是对数据中的每一个属性 A_i 搜索它所有的取值,分别找到它的最大值 x_{Max} 与最小值 x_{Min} ,并计算 $Length = x_{Max} - x_{Min}$,而后将每个属性值 A_{ij} 转成数值区间 $[A_{ij} - Length \times$

$0.05, A_{ij} + Length \times 0.05]$,这种方法使数据有 10% 的不确定性,称为 U10^[6],当然还可以扩大或缩小不确定度。

6.1 正确率

本次实验采用了 4 种不确定度 U1、U5、U10、U20,并利用基于区间的方法将不确定性数据分别用高斯云与均匀云表示,再对两种云数据进行不确定性分类实验,实验结果如表5所列,加下划线的数据为最大正确率,实验结果均在十折交叉验证法下完成。

表5 不确定分类的正确率表(%)

编号	U1		U5		U10		U20	
	高斯云	均匀云	高斯云	均匀云	高斯云	均匀云	高斯云	均匀云
1	<u>80.90</u>	78.95	80.75	78.64	80.44	78.90	77.00	76.14
2	45.63	46.21	48.14	<u>48.58</u>	46.78	46.90	43.56	43.82
3	88.91	85.60	<u>89.15</u>	86.07	85.75	84.53	84.85	84.33
4	<u>73.74</u>	73.02	72.54	73.04	72.45	72.86	72.31	72.54
5	96.80	96.80	96.80	96.67	96.71	95.33	<u>96.97</u>	95.33
6	74.81	74.56	74.90	74.90	74.91	<u>74.92</u>	74.33	74.35
7	48.67	49.26	48.78	<u>49.35</u>	48.06	48.67	47.90	48.50
8	<u>97.13</u>	95.40	96.25	94.32	95.63	93.72	95.63	93.71

从表5中我们发现除了 glass 与 vehicle 以外,大多数数据集的分类结果是均匀云略低于高斯云,再则随着不确定度的加大,大多数数据的分类效果呈现缓慢下降趋势,但部分数据的分类正确率峰值出现在 U5 到 U20,比如 glass、heart-s 与 vehicle 3 个数据集的正确率峰值在 U5,sonar 的正确率峰值在 U10,iris 的正确率峰值在 U20。

为了对比,我们查阅了近年来有关不确定性数据分类实验的结果,并统计了与本实验使用了相同数据集的实验结果,挑选出了其中最优的正确率列在表6中,表中的编号对应的是前文给出的数据集编号,对比表5与表6中的结果可以发现,我们的实验结果除了 ionosphere 的结果略差以外,其它的结果都超过或接近最优正确率,所以从实验的结果来看,我们提出的方法还是很有效的。

表6 其它文献中对应数据集的不确定分类最高正确率表(%)

编号	1	2	4	5	6	7	8
正确率(%)	75.7	47.84	87.52	96	74	45.04	97.46
文献编号	[9]	[9]	[9]	[9]	[6]	[9]	[9]
不确定性	U1, U5	U1	U1	U20	U20	U1	U1

6.2 计算时间

关于不确定性分类实验的执行时间,我们也做了统计,结果如表7所列。由于表6中其它文献不确定分类的最高正确率都来自于文献[6,9],因此本文也统计了这两篇文献中不确定分类的构造时间,结果如表8与表9所列。

表7 不确定分类的计算时间表(秒)

编号	U1		U5		U10		U20	
	高斯分布	均匀分布	高斯分布	均匀分布	高斯分布	均匀分布	高斯分布	均匀分布
1	4.34	1.61	4.47	1.71	4.21	1.62	4.42	1.70
2	2.43	0.88	2.51	0.98	2.44	0.97	2.60	1.06
3	2.24	0.89	2.26	0.91	2.38	0.92	2.44	0.99
4	3.72	1.54	3.7	1.54	3.75	1.56	3.76	1.52
5	1.08	0.55	1.15	0.54	1.21	0.50	1.33	0.51
6	4.57	1.68	4.53	1.80	4.61	1.82	4.65	1.84
7	6.01	2.36	6.1	2.38	5.95	2.36	6.1	2.42
8	2.19	0.97	2.18	0.95	2.21	0.95	2.27	0.95

¹⁾ UCI Machine Learning Repository <http://archive.ics.uci.edu/ml/>

表8 文献[6]中相应数据的不确定分类的计算时间表(秒)

编号	U5		U10		U20	
	高斯分布	均匀分布	高斯分布	均匀分布	高斯分布	均匀分布
1	2.85	0.47	4.75	0.49	5.04	0.59
2	0.6	0.12	0.78	0.18	1.12	0.16
5	0.11	0.08	0.16	0.09	0.21	0.06
6	1.1	0.52	1.73	0.49	2.26	0.75

表9 文献[9]中相应数据的不确定分类的计算时间表(秒)

编号	U1	U5	U10	U20
	高斯分布	高斯分布	高斯分布	高斯分布
1	0.03	0.02	0.02	0.02
2	0.04	0.02	0.02	0.02
4	0.03	0.03	0.03	0.03
5	0.01	0.01	0.01	0.01
6	0.05	0.03	0.03	0.03
7	0.14	0.14	0.14	0.14
8	0.02	0.02	0.02	0.02

从表7—表9的对比情况来看,本文的分类器计算时间较长,但考虑到机器的配置及开发工具,算法的效率还是可以接受的。

结束语 云模型是不确定人工智能的重要理论,它对于表达不确定知识与不确定性数据有同样优秀的效果。本文重点研究了不确定性数据中两种常见的分布,即高斯分布与均匀分布的云建模方法,并在此基础上提出了利用云综合与云相似度计算来实现不确定性数据的分类过程。均匀云在此前的云模型研究中少有提及,所以本文也着重讨论了均匀云的生成算法及均匀云分布的特点,并根据均匀云的特点给出了它的云综合公式及相似度计算公式。最后实验结果证明云模型是一种优秀的不确定性数据的分析工具。

参考文献

- [1] 王意洁,李小勇,祁亚斐,等. 不确定数据查询技术研究[J]. 计算机研究与发展, 2012, 49(7): 1460-1466
- [2] Parag A, Jennifer W. Generalized Uncertain Databases: First Steps[C]//Proceedings of the 4' International VLDB Workshop on Management of Uncertain Data (MUD 2010) in conjunction with VLDB 2010. Twente, Netherlands; Singapore, 2010: 99-111
- [3] Smith T, Kao Ben, Yip K Y, et al. Decision Trees for Uncertain Data[C]// Proceedings of the 25th IEEE International Conference on Data Engineering (ICDE 2009). Shanghai, China; IEEE, 2009: 441-444
- [4] Qin Biao, Xia Yu-ni, Li Fang. DTU: A Decision Tree for Uncertain Data[C]//Proceedings of the 13th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2009). Bangkok, Thailand; Springer, 2009: 4-15
- [5] 周帅印,李晨,王勇,等. FDTU: 针对不确定数据的快速决策树生成算法[J]. 计算机研究与发展, 2010, 47(zl): 363-369
- [6] Qin Biao, Xia Yu-ni, Sunil P, et al. A Rule-Based Classification Algorithm for Uncertain Data[C]// Proceedings of the 25th IEEE International Conference on Data Engineering (ICDE 2009). Shanghai, China; IEEE, 2009: 1633-1640
- [7] Qin Biao, Xia Yu-ni, Sunil P, et al. Rule induction for uncertain data[J]. Knowledge Information Systems, 2011, 29(1): 103-130
- [8] Ren Jiang-tao, Lee S D, Chen Xian-lu, et al. Naive Bayes Classification of Uncertain Data[C]//Proceedings of the 9th IEEE International Conference on Data Mining (ICDM2009). Miami,

USA; IEEE, 2009: 944-949

- [9] Qin Biao, Xia Yu-ni, Wang Shan, et al. A novel Bayesian classification for uncertain data[J]. Knowledge-Based Systems, 2011, 24(8): 1151-1158
- [10] Chau M, Cheng R, Kao Ben. Uncertain Data Mining: A New Research Direction[C]//Proceedings of the Workshop on the Sciences of the Artificial. Hualien, Taiwan; 2005: 89-102
- [11] 王国胤,李德毅,等. 云模型与粒计算[M]. 北京: 科学出版社, 2012: 1-39
- [12] Liu Yu, Chen Gui-sheng. Cloud Model Based Classifier[C]// 2009 International Conference on Test and Measurement (ICTM2009). HongKong; IEEE, 2009: 427-430
- [13] 张国英,沙芸,余有明,等. 基于属性相似度的云分类器[J]. 北京理工大学学报, 2005, 25(6): 499-503
- [14] 刘达,张国英,刘冠洲,等. 基于特征筛选的云分类器[J]. 北京石油化学工业学院学报, 2011, 19(1): 10-14
- [15] 石礼娟,文友先,牟同敏,等. 逆向云在空白识别中的应用[J]. 农业机械学, 2009, 40(12): 196-199
- [16] Shi Li-juan, Wen You-xian, Xie Xin-gang. Classifier Based on Cloud Model and Its Application[C]// Computational Intelligence and Software Engineering. WuHan, China; IEEE, 2009: 1-4
- [17] Lu Hong-fei, Pi Er-xu, Peng Qiu-fa, et al. A particle swarm optimization-aided fuzzy cloud classifier applied for plant numerical taxonomy based on attribute similarity[J]. Expert Systems with Applications, 2009, 36: 9388-9397
- [18] Pi Er-xu, Lu Hong-fei, Jiang Bo, et al. Precise plant classification within genus level based on simulated annealing aided cloud classifier[J]. Expert Systems with Applications, 2011, 38: 3009-3014
- [19] Qin Kun, Xu Kai, Liu Fei-long, et al. Image segmentation based on histogram analysis utilizing the cloud model[J]. Computers & Mathematics with Applications, 2011, 62(7): 2824-2833
- [20] 李德毅,杜鹤. 不确定性人工智能[M]. 北京: 国防工业出版社, 2005: 255-257
- [21] 李德毅,刘常昱. 论正态云模型的普适性[J]. 中国工程科学, 2004, 6(8): 28-33
- [22] 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, 2(6): 16-21
- [23] 杨朝辉,李德毅. 二维云模型及其在预测中的应用[J]. 计算机学报, 1998, 21(11): 962-968
- [24] 刘禹,李德毅,张光卫,等. 云模型雾化特性及在进化算法中的应用[J]. 电子学报, 2009, 37(8): 1651-1658
- [25] 刘禹,李德毅. 正态云模型雾化性质统计分析[J]. 北京航空航天大学学报, 2010, 36(11): 1320-1324
- [26] 罗自强,张光卫,李德毅. 一维正态云的概率统计分析[J]. 信息与控制, 2007, 36(4): 471-475
- [27] 吕辉军,王晔,李德毅,等. 逆向云在定性评价中的应用[J]. 计算机学报, 2003, 26(8): 1009-1014
- [28] 刘常昱,冯芒,戴晓军,等. 基于云 X 信息的逆向云新算法[J]. 系统仿真学报, 2004, 16(11): 2417-2420
- [29] 邸凯昌,李德毅,李德仁. 云理论及其在空间数据挖掘和知识发现中的应用[J]. 中国图象图形学报, 1999, 4(11): 930-935
- [30] 蒋嵘,李德毅,范建华. 数值型数据的泛概念树的自动生成方法[J]. 计算机学报, 2000, 23(5): 470-476
- [31] 秦昆,李德毅,许凯. 基于云模型的图像分割方法研究[J]. 测绘信息与工程, 2006, 31(5): 3-5