

云映射和映射隶属云^{*}

杨志晓, 范艳峰

(河南工业大学 信息科学与工程学院, 郑州 450001)

摘 要: 为描述事物间关系的不确定性及其变化情况, 提出云映射概念。对定义域的每个元素, 通过随机变量将其非确定地变换到可能的取值空间, 各随机变量的分布特征在定义域上构成特征函数, 能够反映非确定变换随自变量的变化情况; 提出了映射的隶属云概念, 云映射在定义域全体元素的隶属云片构成映射的隶属云; 采用映射的期望(即精确函数)、熵函数和超熵函数描述定性概念在整个自变量取值空间的变化情况, 给出了正向映射云和逆向映射云算法。以人类记忆的遗忘过程为例, 说明了云映射和映射隶属云的有效性及其应用前景。

关键词: 云映射; 映射隶属云; 模糊性; 随机性; 遗忘规律

中图分类号: TP391 **文献标志码:** A **文章编号:** 1001-3695(2012)02-0553-04

doi: 10.3969/j.issn.1001-3695.2012.02.040

Cloud mapping and membership cloud of mapping

YANG Zhi-xiao, FAN Yan-feng

(College of Information Science & Engineering, Henan University of Technology, Zhengzhou 450001, China)

Abstract: In order to describe uncertainties and their variation between events, this paper proposed a concept of cloud mapping. For each element of definition domain, used a matched stochastic variable to convert it to a possible value space. Distribution features of the stochastic variable cluster in definition domain form feature functions, which indicated variation of uncertainties according to change of independent variable. This paper proposed a concept of membership cloud of mapping. Each element of definition domain has a cloud piece. All cloud pieces form membership cloud of mapping. Expected function, used entropy function, and hyper entropy function to describe variation of qualitative concept in the whole definition domain. This paper proposed algorithms for generating positive mapping cloud and reverse mapping cloud. Taking human memory law as an example, introduced effectiveness and possible application of proposed cloud mapping and mapping cloud.

Key words: cloud mapping; membership cloud of mapping; fuzziness; randomness; memory law

0 引言

自然界和人类社会各种现象存在大量不确定性, 包括模糊性、随机性、不稳定性、不完全性、不一致性等。其中模糊性和随机性是最重要的两种不确定性, 且常常连在一起难以区分^[1]。

精确数学是对各种不确定现象在期望状态下的抽象。它能够刻画事物在给定条件下的精确解, 但不能描述广泛存在的不确定现象。概率和统计理论以事件发生的概率、样本期望、方差等指标描述事件发生的随机特征, 能够处理各种随机现象。但无法描述定性概念的模糊性。Zadeh 提出的模糊集理论采用 $[0, 1]$ 区间的连续逻辑值, 相对经典的二值逻辑, 能够描述各种模糊现象, 但是其隶属函数是确定性函数, 对给定变量值, 模糊变量的隶属度唯一确定, 不能够刻画随机现象, 且隶属函数的确定依赖经验。

为统一刻画定性概念的模糊性和随机性, 李德毅等人^[2]提出了隶属云理论, 采用隶属云描述定量论域内随机实现的云滴可代表定性概念的确定程度。与模糊集理论不同, 论域内给定取值的云滴可代表定性概念的确定度不是一个唯一确定值, 而是具有稳定倾向的随机数。因此, 云模型能够将模糊性和随

机性这两种最重要的不确定性统一量化描述。且由于云模型的数字特征值来源于样本估计, 能够克服模糊性、随机性量化过程的主观性。由于云模型能够统一刻画定性概念量化过程的模糊性和随机性, 因此一经提出便迅速在不确定性数据挖掘^[36]、控制等领域得到了广泛的应用。

不过, 隶属云模型面向单个定性概念。而在工程技术、自然科学和社会科学等领域, 经常需要刻画事物(y)随某个(些)因素(x)变化的规律性 $f: x \rightarrow y$ 。如仪器仪表输出的线性度, 网络流量随时间的变化关系, 网络热点事件的关注度变化情况, 交通流量, 产品销量与广告投入的关系, 等等。这些规律性具有不确定性。表现在: a) 因变量 y 在给定自变量 x 取值下不是确定值; b) 在不同自变量取值下, 因变量的不确定性规律可能存在不同; c) 在不同自变量 x 取值下, y 所代表的概念性质可能发生了改变。且对 y 的部分测度空间, 由于危险性、成本巨大、不可实施等原因, 难以获取用于逆向生成隶属云的足够样本。因此, 针对单个定性概念的云模型不能描述事物及其关系在整个测度空间的不确定性规律。

针对上述问题, 本文提出云映射及其隶属云的概念, 利用随机变量簇将自变量取值空间的元素变换到各自可能的取值

收稿日期: 2011-07-06; **修回日期:** 2011-08-12 **基金项目:** 河南省重点科技攻关计划资助项目(082102210096); 河南省高校青年骨干教师计划资助项目(2009GGJS-056); 郑州市科技局科技攻关计划资助项目(2010GYXM364)

作者简介: 杨志晓(1974-), 男, 河南郑州人, 副教授, 博士, 主要研究方向为人机交互、智能计算技术及应用(yzxbox@haut.edu.cn); 范艳峰(1973-), 女, 河南郑州人, 副教授, 博士, 主要研究方向为不确定性人工智能。

空间,根据各随机变量的分布特征所构成的熵函数和超熵函数,表示不确定性在整个测度空间的变化规律。

1 云映射和映射隶属云

1.1 云映射

定义 1 云映射。设 A 和 B 是两个非空集合,若按照某种对应关系 $F, \forall x \in A, \exists$ 随机变量 Y_x ,在集合 B 上取得值域 $U_x \subseteq B$,称 F 为 A 到 B 的一个云映射,记为 $F: x \xrightarrow{Y_x} U_x$, 或 $Y_x = F(x)$ 。

与精确数学中的映射关系不同,定义 1 中每一个确定值 $x \in A$ 都对应一个随机变量 Y_x 。 Y_x 的每一次取值在实现前都不是确定的,其可能取值的范围即值域是集合 B 的一部分。对集合 A 中的全体元素, F 定义了一个随机变量簇 $Y = \{Y_x | x \in A\}$ 。各个随机变量的分布可能相同,也可能不同。随机变量簇中各随机变量在不同次实现中具有不确定性,它们各自的多次实现聚集如同云团。 F 通过随机变量簇 Y 中相应的随机变量将确定值 $x \in A$ 非精确地变换到可能的取值空间, F 的特征通过全体随机变量簇 Y 来体现,因此称 F 为云映射。

隶属云模型针对单个定性概念,采用期望 Ex 表示定量论域内最能够代表定性概念的点,采用熵 En 刻画定性概念的模糊性和随机性,采用超熵 He 刻画云滴可代表定性概念的确定程度的随机性^[2,7]。借鉴其思想,将单个定性概念的隶属云模型拓展至定义 1 云映射 F 的整个自变量空间 A ,有下述定义。

定义 2 映射隶属云、隶属云片和云滴。设 $Y_x = F(x)$ 为非空定量值集合 A 到 B 的一个云映射,则 $\forall x \in A$, 随机变量 Y_x 在集合 B 上取得定量值域 $U_x \subseteq B$ 。 C_x 是定量论域 U_x 上的定性概念, $\forall y \in U_x$ 是定性概念 C_x 在 U_x 上的一次随机实现。 y 对 C_x 的确定度 $\mu(y) \in [0,1]$ 是有稳定倾向的随机数, $\mu: U_x \rightarrow [0,1], \forall y \in U_x, y \rightarrow \mu(y)$, 则 y 在论域 U_x 上的分布称为 F 在 x 处的一个隶属云片,记为 $C(Y_x)$ 。每一个 y 称为一个云滴,全体云片组成 F 的隶属云簇,称为云映射 F 的隶属云,简称映射云,记为 $C(F(x))$ 。

定义 2 中,每一个确定值 $x \in A$ 都对应一个定性概念 C_x ,集合 A 中的全体元素定义了一个定性概念簇 $C_A = \{C_x | x \in A\}$,它们既可能是具有实际意义易于理解的概念,也可能是没有现实意义的抽象概念。对任何一个概念 C_x ,定量论域 $U_x \subseteq B$ 内最能代表它的点是随机变量 Y_x 的期望值的点,记为 $f(x)$ 。 C_x 在每次实现前都是不确定的,其每个实现称为云滴。 C_x 在定量论域 $U_x \subseteq B$ 内的模糊性和随机性用熵度量,记为 $e(x)$ 。云滴 y 能够代表定性概念 C_x 的确定度 $\mu(y)$ 是一个具有稳定倾向的随机数,其随机性用超熵度量,记为 $h(x)$ 。定性概念 C_x 的特征通过其在 x 处的大量云滴构成的隶属云片的整体来体现,全体隶属云片能够反映概念簇 C_A 随集合 A 中元素的变化情况。

在整个自变量取值空间 A , $f(x)$ 、 $e(x)$ 和 $h(x)$ 是与 x 有关的函数,分别称为云映射 F 的期望函数、映射隶属云 $C(F(x))$ 的熵函数和超熵函数,则 $\forall x \in A, C(x)$ 的云数字特征为 $C(f(x), e(x), h(x))$ 。 $f(x)$ 、 $e(x)$ 和 $h(x)$ 分别表示了云映射 F 的期望(即精确映射 f)、熵和超熵在自变量域的变化情况,因此, F 的隶属云也可称为精确函数 $y=f(x)$ 的云。它在本质上反映

了以函数 $y=f(x)$ 为期望的不确定性映射关系在值域空间和确定度空间的分布。

从前文定义和分析可以看出,与文献[2]的云模型相比,定义 2 的映射隶属云模型给出了定性概念在整个自变量取值空间的变化情况。文献[2]的云模型即为定义 2 的一个隶属云片。文献[8]论述了正态云的普适性,本文沿用正态分布描述随机性,即 $\forall x \in X, y \sim N(f(x), En'^2), En' \sim N(e(x), h(x)^2)$ 。 y 的确定度为

$$\mu = e^{-\frac{(y-f(x))^2}{2En'^2}} \quad (1)$$

1.2 几种特殊云映射

定义 3 同质云和同质云映射。 $\forall x \in A$, 若 $e(x)$ 和 $h(x)$ 在整个自变量空间 A 分别取得固定值,称云 $C(F(x))$ 同质。

$F: x \xrightarrow{Y_x} U_x$ 为 A 到 B 的一个同质云映射。

定义 4 异质云和异质云映射。 $\exists x_1 \in A, \exists x_2 \in A$, 且 $x_1 \neq x_2$, 若 $e(x_1) \neq e(x_2)$ 和 $h(x_1) \neq h(x_2)$ 至少有一个成立,则称

云 $C(F(x))$ 异质。 $F: x \xrightarrow{Y_x} U_x$ 为 A 到 B 的一个异质云映射。

根据上述定义,有:

定理 1 若 $C(F(x))$ 在 A 上同质, $\exists x_0 \in A$, 定性概念 C_{x_0} 的云模型数字特征为 $C(f(x_0), En_0, He_0)$, 则 $\forall x \in A$, 概念 C_x 的云模型数字特征为 $C(f(x), En_0, He_0)$ 。

证明 $\forall x \in A$, 由定义 3 知 $e(x) = e(x_0) = En_0, h(x) = h(x_0) = He_0$, 则定性概念 C_x 的云数字特征为 $C(f(x), En_0, He_0)$ 。得证。

定理 1 的意义在于,对同质云映射 F ,若某个定性概念的云片数字特征已知,则可将其应用于全体概念簇 C_A 中的任意云片。

定义 5 线性云和线性云映射。对云映射 $F, \forall x \in A$, 若 $e(x)$ 和 $h(x)$ 分别与 $f(x)$ 呈线性关系,即满足

$$e(x) = af(x) + b \quad (2)$$

$$h(x) = cf(x) + d \quad (3)$$

式中: a, b, c, d 为常数。则称 $C(F(x))$ 为线性云, F 为线性云映射。

定理 2 若 F 为线性云映射, $\exists x_0 \in A$, 定性概念 C_{x_0} 的云模型数字特征为 $C(f(x_0), En_0, He_0)$, 则 $\forall x \in A$, 概念 C_x 的云模型数字特征为 $C(f(x), En_0 + a(f(x) - f(x_0)), He_0 + c(f(x) - f(x_0)))$ 。

证明 由定义 5 知

$$e(x_0) = af(x_0) + b = En_0 \quad (4)$$

$$h(x_0) = cf(x_0) + d = He_0 \quad (5)$$

$\forall x \in A$, 有

$$e(x) = af(x) + b \quad (6)$$

$$h(x) = cf(x) + d \quad (7)$$

式(6)减式(4)可得

$$e(x) - e(x_0) = af(x) + b - af(x_0) - b = a(f(x) - f(x_0)) \quad (8)$$

则有 $e(x) = e(x_0) + a(f(x) - f(x_0)) = En_0 + a(f(x) - f(x_0))$ 。

同理,式(8)减式(5)可得 $h(x) = He_0 + c(f(x) - f(x_0))$ 。得证。

定理 2 的意义在于对线性云映射 F ,若 x_0 处的云数字特征熵和超熵已知,则可将测度较难获得的定性概念的不确定性

问题转换为测度较易获得的定性概念的不确定性问题。

2 初等函数的隶属云

在第 1 章中,若云映射 F 的期望 $y=f(x)$ 为基本初等函数,则很容易获得基本初等函数的云。图 1 给出了若干基本初等函数的云,其中(a)~(h)为同质云映射,它们的云数字特征同为 $\forall x \in A, e(x)=h(x)=0.1$ 。可以看出,在各个基本初等函数的云中,对任意的确定值 x ,函数值 $f(x)$ 不是一个确定值,而是一个服从正态分布的随机变量。

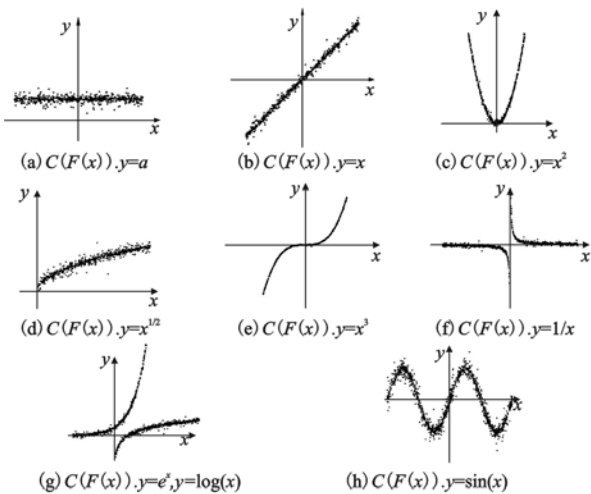


图1 基本初等函数的同质云. $e(x)=h(x)=0.1$

图 2 分别给出了两个基本初等函数 ($y=x$ 和 $y=\sin(x)$) 的线性云。熵函数都取为 $e(x)=0.1f(x)$, 超熵函数都取为 $h(x)=0.1e(x)=0.01f(x)$ 。可以看出,随着绝对函数值 $|y|$ 的增加, y 的不确定性加剧。

用垂直 x 轴的平面截取映射云,即得到在该处的云片。图 3 给出了截取的某个云片。横轴为 $y=F(x)$,表示云映射 $F(x)$ 在该处的取值;纵轴为确定度 μ ,表示云滴可代表定性概念 $C(x)$ 的程度。可以看出,云片即为文献[2]的隶属云。因此,采用图 1 或图 2 与图 3 的组合形式,可以表示映射云在整个定义域上的形态。

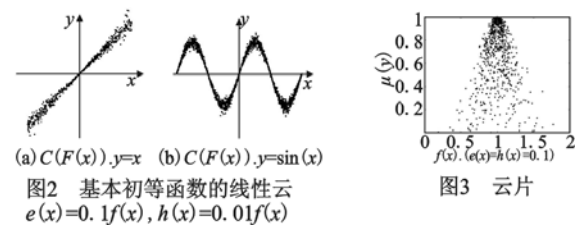


图2 基本初等函数的线性云
 $e(x)=0.1f(x), h(x)=0.01f(x)$

图3 云片

与单个定性概念的隶属云片相比,映射云多了一个自变量维,因此其云具有三维立体形态。直线 $y=x$ 和抛物线 $y=x^2$ 映射云的三维效果如图 4 所示。图中不仅给出了云映射 $F(x)$ 在自变量各取值下的取值分布,还给出了云滴可代表定性概念 C_x 的程度 μ_0 。图中的两个映射云在三维空间呈立体形态,具有自然界中云的真实效果。

对云映射 $F(x)$,若熵函数和超熵函数在整个自变量取值空间都同为 0,即 $\forall x \in A, e(x)=h(x)=0$,则 $F(x)$ 蜕变为精确映射,即其期望函数 $y=f(x)$ 。

通过上述分析可以看出,云映射和映射的隶属云具有三个重要功能:a)它可以模糊现象和随机现象统一刻画描述;b)映射云不仅能够刻画单个定性概念的模糊性与随机性,而且能

够表示这种不确定性在自变量取值空间的变化情况,从而描述定性概念的动态变化情况;c)云映射将精确数学与不确定现象统一起来,既能将精确值非精确地变换到不确定空间,又能从不确定性的期望抽象出精确映射。

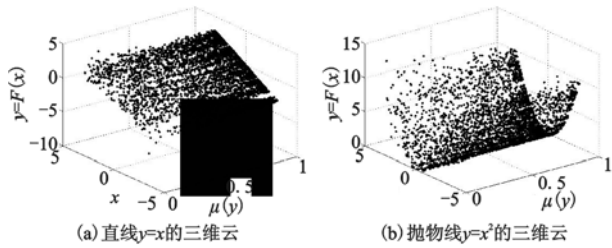


图4 两个映射云的三维效果图

3 云映射算法

3.1 逆向映射云生成算法

对云映射 $Y_x=F(x)$,选取定义域的关键点,分别获得 $F(x)$ 的多次实现,从中估计出在各关键点 $F(x)$ 的期望、熵和超熵,分别对它们在各个关键点的取值进行拟合,估计出期望函数 $f(x)$ 、熵函数 $e(x)$ 和超熵函数 $h(x)$ 的过程,称为逆向生成映射云。算法描述如下:

- a)确定定义域 A ;
- b)在区间 $[\min\{A\}, \max\{A\}]$ 上随机产生 M 个均匀分布的值 $X=\{x_i|i=1,2,\cdots,M\}$;
- c) $\forall x_i \in X$,测量 $F(x_i)$ 的 N 个实现 $Y_i=\{y_{ij}|j=1,2,\cdots,N\}$;
- d)按照式(9)估计 $F(x)$ 在 x_i 处的期望:

$$Ex_i=\frac{1}{N}\sum_{j=1}^Ny_{ij}$$
 (9)

按照式(10)估计 $F(x)$ 在 x_i 处的熵:

$$En_i=\sqrt{\frac{\pi}{2}}\frac{1}{N}\sum_{j=1}^N|y_{ij}-Ex_i|$$
 (10)

按照式(11)估计 $F(x)$ 在 x_i 处的超熵:

$$He_i=\sqrt{S_i^2-En_i^2}$$
 (11)

其中: S_i^2 为样本方差,即

$$S_i^2=\frac{1}{N-1}\sum_{j=1}^N(y_{ij}-Ex_i)^2$$
 (12)

e)分别对 M 对 $\{(x_i, Ex_i)|i=1,2,\cdots,M\}$, $\{(x_i, En_i)|i=1,2,\cdots,M\}$, $\{(x_i, He_i)|i=1,2,\cdots,M\}$ 数据进行拟合,获得期望函数 $f(x)$ 、熵函数 $e(x)$ 和超熵函数 $h(x)$ 的估计。

3.2 正向映射云生成算法

已知云映射 $Y_x=F(x)$ 及其云特征函数 $C(f(x), e(x), h(x))$,产生 $F(x)$ 在定义域某取值下的一个实现的过程,称为正向生成映射云。算法描述如下:

- a)输入 $x_0 \in A$,分别根据 $f(x)$ 、 $e(x)$ 、 $h(x)$ 计算 $F(x_0)$ 的期望 $f(x_0)$ 、熵 $En_0=e(x_0)$, 超熵 $He_0=h(x_0)$;
- b)根据 $En_0' \sim N(e(x_0), h(x_0)^2)$ 产生 En_0' ;
- c)根据 $y \sim N(f(x_0), En_0'^2)$ 产生 y ;
- d)根据式(13)计算确定度。

$$\mu=e^{-\frac{(y-f(x_0))^2}{2En_0'^2}}$$
 (13)

4 云映射应用例

云映射及其算法具有第 1 章中所描述的三个重要功能,可

以广泛应用于刻画定性概念的不确定性及其动态变化。本文以人类记忆的遗忘过程为例,说明云映射描述定性概念不确定性变化的有效性。

德国心理学家 Ebbinghaus 研究发现,人类记忆的遗忘过程大致符合指数规律。记忆保持量 r (无单位量纲,1 表示没有遗忘,0 表示完全遗忘)与时间 t (单位:h)的经验公式为

$$r = \frac{a}{(\log t)^b + a} \tag{14}$$

其中; a 和 b 为常数,经验值分别为 $a = 1.84$, $b = 1.25$ 。

人的记忆能力存在个体差异,即使同一个人在不同时期、对不同事物的记忆能力也不相同。式(14)是从时域到记忆保留量的一个精确变换,是对人类遗忘过程的抽象,其精确特征不能描述个体记忆能力和被记忆事物的差异性,因此可以采用云映射来刻画遗忘过程。

图 5 分别给出了特征函数为 $C(r(t), e(t) = 0.02, h(t) = 0.02)$ 和 $C(r(t), e(t) = 0.1r(t), h(t) = 0.1r(t))$ 的映射云。

在图 5 的两个遗忘映射云指导下,图 6 所示的原始人脸在人的记忆中的遗忘过程分别如图 7 和 8 所示。两组图分别给出了经过 2 h、5 h、20 h、10 d 后对原始人脸保留的记忆。其中每行图的左 1 为该时刻对原始人脸保留记忆的期望,其他三个为在映射云指导下实现的三个保留记忆。可以看出,期望保留记忆相对清晰,而实现的保留记忆较为模糊,且同一时刻各实现的保留记忆存在差异。因此遗忘映射云能够描述遗忘过程的不确定性。

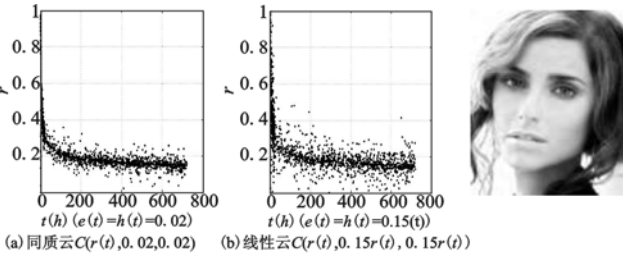


图5 遗忘映射云

图6 原始人脸

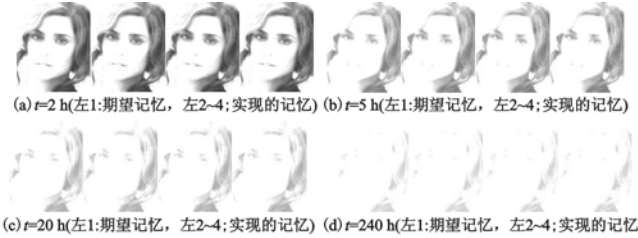


图7 人脸在记忆中的衰退过程 $C(r(t), 0.02, 0.02)$

表 1 给出了 5 h 后两个遗忘映射云各自实现的一个记忆保留量云滴。可以看出,对于同质遗忘映射云 $C(r(t), 0.02, 0.02)$,实现的一个记忆保留量为 0.493 4,且其确定度为 0.971 0。线性遗忘映射云 $C(r(t), 0.15r(t), 0.15r(t))$ 此时

(上接第 552 页)

[5] 梁炎夏,李会勇,黄秋欣,等. 基于 PASTd 的圆阵 ESPRIT 算法[J]. 系统仿真技术, 2010,6(3):187-191.

[6] DOUKOPOULOS X G, MOUSTAKIDES G V. Fast and stable subspace tracking[J]. IEEE Trans on Signal Processing, 2008,56(4):1452-1465.

[7] 魏志强,李艳玲,陈招迪. 基于特征向量跟踪的数据压缩 MMSE 多用户检测[J]. 华北水利水电学院学报, 2010,31(5):110-112.

[8] LIAO B, ZHANG Z G, CHAN S C. A new robust kalman filter-based

的熵和超熵为 0.075 6,实现的一个记忆保留量为 0.508 7,且其确定度为 0.891 6。因此,遗忘映射云不仅能够描述遗忘过程的不确定性,且能够回答在任意时刻实现的记忆保留量的确定程度。改变熵函数和超熵函数,还能够设计不同的符合需要的遗忘过程。



图8 人脸在记忆中的衰退过程 $C(r(t), 0.15r(t), 0.15r(t))$

表 1 5 h 后遗忘映射云实现的记忆保留量云滴

遗忘映射云	$r(5)$ 期望	$e(5)$	$h(5)$	实现的 $r(5)$	实现的 μ
$C(r(t), 0.02, 0.02)$	0.5037	0.0200	0.0200	0.4934	0.9710
$C(r(t), 0.15r(t), 0.15r(t))$	0.5037	0.0756	0.0756	0.5087	0.8916

5 结束语

本文提出了云映射和映射的隶属云概念。采用映射的期望(即精确映射 $y = f(x)$)、熵函数和超熵函数构造映射隶属云,刻画定性概念的模糊性和随机性在自变量取值空间的变化情况,给出了映射云的逆向生成算法和正向生成算法。应用实例说明了映射隶属云的有效性。

云映射和映射云概念为不确定现象的动态变化提供研究工具。未来的研究包括复合型云映射、云映射性质及其应用的进一步发掘等。

参考文献:

[1] 李德毅,刘常昱,杜鹃,等. 不确定性人工智能[J]. 软件学报, 2004,15(11):1583-1594.

[2] 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.

[3] 郭戎潇,夏靖波,董淑福,等. 一种基于多维云模型的多属性综合评价方法[J]. 计算机科学,2010,37(11):75-77.

[4] 夏建明,杨俊安,张琼. 基于数据符号化表示和云模型的时序数据生成方法[J]. 计算机应用研究,2010,27(10):3691-3693,3697.

[5] 宋晶,戚建淮. 基于云变换和特性关系下粗糙集的决策树构造[J]. 西南交通大学学报,2010,45(2):312-316.

[6] 代劲,何中市. 基于云模型的决策表规则约简[J]. 计算机科学, 2010,37(6):265-267,277.

[7] 刘常昱,李德毅,杜鹃,等. 正态云模型的统计分析[J]. 信息与控制,2005,34(2):236-239,248.

[8] 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):28-34.

subspace tracking algorithm in an impulsive noise environment[J]. IEEE Trans on Circuits and Systems- II, 2010,57(9):740-744.

[9] SCHMIDT R O. Multiple emitter location and signal parameter estimation[J]. IEEE Trans on Antennas Propagation,1986,34(3):276-280.

[10] SANTOS P L, RAMOS J A, JORGE L M. Identification of bilinear systems with white noise inputs: an iterative deterministic-stochastic subspace approach[J]. IEEE Trans on Control Systems Technology,2009,17(5):1145-1153.