正态分布真随机数云模型发生器①

吴立新，王 勇

(常州信息职业技术学院，常州 213164)

摘 要：云模型发生器多是基于伪随机数的发生器。在充分研究了随机数发生器、云模型的国内外发展现状的基础上，设计了一种产生高质量随机数的硬件实现方案，并使用软件的方法将均匀分布的真随机序列逐步转换成正态分布真随机数一维云模型，全面论述了一种基于真随机数的云模型发生器的实现方法，国内尚属首次，是对云模型研究的有益补充。

关键词：随机数发生器；均匀分布；正态分布；云模型；正向云发生器

Cloud Model Generator Based on the Normal True Random Number

WU Li-Xin, WANG Yong

(Changzhou College of Information Technology, Changzhou 213164, China)

Abstract: Most cloud model generators are based on pseudorandom number. With the research of the development situation of random-number generator and cloud model at home and abroad, this paper designs a high-quality-random number-producing hardware project and uses the software solution to gradually convert the uniformly distributed true random sequence into an one-dimensional cloud model of normally distributed true random numbers. It disscusses in depth about the implementation of cloud model generator based on true random numbers. It is the first time in our country and is helpful to the study of cloud model.

Key words: randomizer; uniform distribution; normal distribution; cloud model; normal could generator

1 概述

随着计算机技术、通信技术、网络技术的迅速发展，信息安全问题已受到人们的广泛关注。随机数在信息安全系统中扮演着重要的角色，在计算机通信、网络通信中有着广泛的应用。如何产生高质量的随机数已成为密码学乃至信息安全领域的一个重要研究方向[1,2]。

2 随机数发生器发展现状

随机数产生的最早方法为手工方法。随着随机模拟技术的广泛应用，需要使用大量的随机数。显然手工方法不能满足模拟计算的需要。

随着计算机和模拟方法的广泛应用，用计算机产生随机数成为新的课题。随机数发生器主要分成三类[3]即查表方法、物理方法、数学方法。

在信息安全领域的基本标准有美国发布的 ITSEC、FIPS 系列标准。对于真随机数发生器而言, 常用评价标准有德国的 AIS31 标准。高质量的随机数序列必须通过一系列的统计检验。常用的检测方法有比特分布检测、跟随特性检测、游程检测、碰撞检测、扑克检测等。

考查随机数发生器的好坏，就要对生成的随机数列进行一系列的统计检验: 比特分布检测，跟随特性检测，游程检测，扑克检测等。而每种检验要用各种方法进行，当各种检验均未得到否定结论时才能认为产生的数列是随机的。

① 收稿时间:2011-03-01;收到修改稿时间:2011-04-12

3　真随机序列的硬件实现方案

真随机数发生器是指用物理方法实现的随机数发生器。各种随机物理过程可用来产生随机物理信号。其中与 IC 工艺兼容的三种随机源: 放大电路噪声、振荡采样、混沌电路[4-6]。随机源的随机性能的好坏直接决定了真随机数发生器芯片的质量。通常实际芯片还要对随机源进行处理, 使其输出的随机序列能够更好的满足统计性检验。

本文中利用稳压二极管反向击穿时发生齐纳现象，可以产生良好的随机噪声。为了增加随机性，本文利用两个稳压管发生两路随机噪声源，经过比较器 U1 和 U2 产生两路随机的数字脉冲信号，并分别送入 RS 触发器的置’1’和置’0’端，最后输出性能更加优越的随机数字脉冲信号。

将利用放大的雪崩噪声产生的真随机序列通过采集板送入 NI 虚拟仪器，对真随机序列的统计特性进行分析。通过对随机序列的比特分布检测、跟随特性检测、游程检测、扑克检测的结果分析可以得出结论：本文中基于电阻热噪声，通过比较器来产生的随机数序列是具备检测特性的均匀分布的真随机序列。

4　均匀真随机序列转换成匀分布的真随机数

(1) 浮点数在计算机中的存储格式

由于浮点数在计算机上是用类似±1.xxxx×2^{e-127}的格式表示的。一个浮点数由 4 个字节(32 bit)构成，浮点数在计算机中的存储格式[7]如下：

XXXXXXXX　XXXXXXXX　EXXXXXXX　SEEEEEEE

存储格式 X 表示尾数，E 表示阶数，S 是符号位。

我们看到，尾数 X 总共有 23 位，其实还有 1 位最高位没有出现(它总是 1，所以就被隐含了)，若所有尾数位都是 0，就表示 1.0000 的尾数。如果阶数 E 是 127，就表示数据的绝对值在[1.0,+2.0)的范围内。

(2) 随机序列转换成随机数的实现方法

由二进制随机数序列生成范围在[0.0，+1.0)区间的随机数的方法如下：

首先用 23 位随机的二进制数序列，填充一个浮点数中的 23 位尾数部分(符号位为 0)，并使阶数为 127，于是这个浮点数的数值就处于[1.0，+2.0)范围内。最后再把浮点数的数值减去 1.0 得到的随机数的范围就处于[0.0，+1.0)之间。

(3) 证明由均匀分布随机序列生成的随机数是均匀分布的

按照数学归纳法的思想，我们先判断只有一位尾数的情况。

在尾数是一位随机数的情况下，看看所表示的数值是否均匀分布：若只有 1 位尾数(其实还有一个隐含的最高位)，则该位的权为 2^{-1}，若尾数位是 1 表示 1.5，若尾数位是 0 表示 1.0。由于尾数位为 1 或 0 的概率相等，所以结果为 1.0 或 1.5 的概率都是 1/2。

证明：若有 n 位尾数时，数据分布的均匀的，则使尾数增加 1 位，数据分布仍是均匀的：设 n 位尾数时，数据共有 2^n 种数值，这些数值出现的概率都相等，都是 $1/2^n$。则再增加 1 位尾数时，由于该位为 0 或 1 的概率都是 1/2，且这个概率与其它尾数位的数值不相关，所以数据就共有 2^{n+1} 种数值，这些数值出现的概率都相等，都是 $1/2^{n+1}$。于是推断：由均匀分布随机序列生成的浮点数在[1.0,+2.0)范围内是服从均匀分布。也可以进一步推断：由在[1.0,+2.0)范围内是服从均匀分布的浮点数减去 1.0 后得到的随机数在[0.0,+1.0)范围内也是服从均匀分布。

5　匀分布的真随机数转换成正态分布的真随机数

由正态分布機率密度函数

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$ (σ 为标准差 μ 为期望值)

得

$$y=\int_{-\infty}^{x}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(u-\mu)^2}{2\sigma^2}}du$$ [8,9]

积分法产生正态分布随机数的方法如下：

假设有一组处于[0.0,+1.0)之间的、均匀分布的随机数 y 序列，对其中每一个元素 y 进行以下操作：使用正态分布的概率密度函数进行连续数值积分（从负无穷开始），假设积分到 x 时，积分值恰好等于 y，则 x 就是要找的正态分布的随机数，把它作为正态分布随机数序列 x_i 的元素。当 y 序列的均匀分布的随机数足够多时，x_i 分布规律就表现为正态分布的形态。

6 检验正态分布的真随机数转换的正确性

从真随机数硬件发生电路取一段连续的真随机序列($1.0*10^6$ B)并保存。采用 VB 编写软件将均匀分布的真随机序列转换成[0.0,+1.0)区间均匀分布的真随机数；再将[0.0,+1.0)区间均匀分布的真随机数采用积分法生成数学期望值 μ (默认为 0)、均方差 σ (默认为 1)的正态分布真随机数，对正态分布真随机数进行了保存和统计，描汇出概率分布图，如图 1 所示。

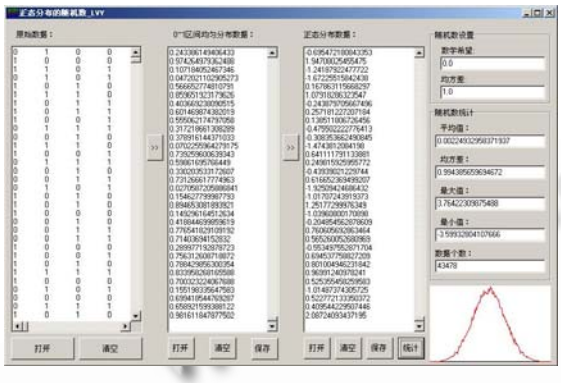

图 1

7 正向云发生器

(1) 云和云滴

定义. 设 U 是一个用精确数值表示的定量论域，C 是 U 上的定性概念，若定量值 $x \in U$，且 x 是定性概念 C 的一次随机实现，x 对 C 的确定度 $\mu(x) \in [0,1]$ 是有稳定倾向的随机数

$$\mu : U \to [0,1] \quad \forall x \in U \quad x \to \mu(x)$$

则 x 在论域 U 上的分布成为云（Cloud），每一个 x 成为一个云滴。

云具有以下性质：

①论域 U 可以是一维的，也可以是多维的。

②定义提及的随机实现，是概率意义下的实现；定义中提及的确定度，是模糊集意义下的隶属度，同时又具有概率意义下的分布。所有这些都体现了模糊性和随机性的关联性。

③对于任意一个 $x \in U$，x 到区间[0，1]上的映射是一对多的变换，x 对 C 的确定度是一个概率分布，而不是一个固定的数值。

④云由云滴组成，云滴之间无次序性，一个云滴是定性概念在数量上的一次实现，云滴越多，越能反映这个定性概念的整体特征。

⑤云滴出现的概率大，云滴的确定度大，则云滴对概念的贡献大。

云是用语言值表示的某个定性概念与其定量表示之间的不确定转换模型，用以反映自然语言概念的不确定性，不但可以从经典的随机理论和模糊集合理论给出解释，而且反映了随机性和模糊性的关联性，构成定性和定量间的映射。

(2) 云的数字特性

云的数字特征用期望 Ex(Expected value)、熵 En(Entropy)和超熵 He(Hyper entropy) 三个数字特征来整体表征定性概念 $\widetilde{A}$，它们反映了定性概念 $\widetilde{A}$ 整体上的定量特征。

期望 Ex（Expectation）：云滴在论域空间分布的期望。通俗地说，就是在数域空间最能够代表定性概念的点，或者说是这个概念量化的最典型样本点。

熵 En（Entropy）：熵反映定性概念 $\widetilde{A}$ 的不确定性，这种不确定性表现在三个方面。一方面，熵反映了数域空间中可以被语言值 $\widetilde{A}$ 接受的云滴群范围大小，即模糊度, 是定性概念亦此亦彼性的度量；另一方面，熵还反映了数域空间中的云滴群能够代表这个语言值的概率，表示代表定性概念的云滴出现的随机性；此外，熵还揭示了模糊性和随机性的关联性。熵还可以用来代表一个定性概念的粒度。通常，熵越大，概念越宏观，模糊性和随机性也越大，确定性量化越难。

超熵 He（Hyper Entropy）：超熵是熵的不确定性的度量，即熵的熵。由熵的随机性和模糊性共同决定。

(3) 云的种类

云模型是云的具体实现方法，也是云运算、云聚类、云推理、云控制等方法的基础。由定性概念到定量表示的过程，也就是由云的数字 产生云滴的具体实现，称为正向云发生器；由定量表示到定性概念的过程，也就是由云滴群得到云的数字特征的具体实现，称为逆向云发生器。

云模型的具体实现方法可以有多种，构成了不同类型的云，如对称云模型、半云模型、组合云模型。

(4) 正向云发生器

正向云发生器是从定性到定量的映射。正向云发生器根据云的三个数字特征 Ex, En, He 产生云滴。

正向正态云定义[10,11]为：设 U 是一个用准确数值表示的定量论域，C 是 U 上的定性概念，若定量值 $x \in U$，且 x 是定性概念 C 的一次随机实现，若满足

$x \sim N\left(En, He^2\right)$，且 x 对 C 的确定度满足

$$\mu = e^{\frac{-(x-Ex)^2}{2\left(En'\right)^2}}$$

则 x 在论域 U 上的分布称为正态云。

当正向云发生器对应的数域为一维数域时，正向正态云发生器的算法可以描述为：

数据输入：正向云三个数字特征值 Ex, En, He，云滴数 N

数据输出：N 个云滴的定量值及每一个云滴的确定度

算法步骤[12]：

①获取一个以 En 为期望值、He 为标准差的正态随机数 En'；

②获取一个以 Ex 为期望值、En' 为标准差的正态随机数 x；

③假设 x 为某定性概念 $\widetilde{A}$ 的一次具体量化，称之为云滴；

④计算

$$y = e^{\frac{-(x-Ex)^2}{2\left(En'\right)^2}}$$

⑤假设 y 为 x 是某定性概念 $\widetilde{A}$ 的确定度；

⑥$\{x, y\}$是定性概念 $\widetilde{A}$ 的一次完整的定性定量转换；

⑦重复步骤①-⑥，直至产生 N 个云滴。

8　正态分布真随机数转换成一维云模型的算法

假设要生成当特征值 Ex=16, En=4, He=1 的一维正态云，根据正向正态云发生器描述的算法，采用 VB 编写软件计算云滴的算法如下：

（1）将采用积分法产生的数学期望值 μ=0、均方差 σ=1 的正态分布真随机数作线形变换，变换成 μ=4、均方差 σ=1 的正态分布真随机数，记为集合 U

（2）从集合 U 依次取出一个数，计为 En’

（3）将数学期望值 μ=0、均方差 σ=1 的正态分布真随机数作线形变换，变换成 μ=16、均方差 σ=En’的正态分布真随机数，记为集合 V

（4）并从集合 V 任取一数，记为 x；

（5）假设 x 为某定性概念 $\widetilde{A}$ 的一次具体量化，称之为云滴；

（6）计算

$$y = e^{\frac{-(x-Ex)^2}{2\left(En'\right)^2}}$$

（7）假设 y 为 x 是某定性概念 $\widetilde{A}$ 的确定度；

（8）$\{x, y\}$是定性概念 $\widetilde{A}$ 的一次完整的定性定量转换；

（9）重复步骤 2-8，直至产生 N 个云滴（N =43478）。

9　正态分布真随机数云模型数发生器云图分析

在数学期望值 μ 等于 0、均方差 σ 等于 1 的正态分布真随机数基础上，采用一维正态云的算法产生了特征值 Ex=16, En=4, He=1 的一维正向正态云（如图 2 所示），并对云图的分布特性进行了统计。绝大部分云滴的横坐标都在[4,28]的范围之内，是较理想的正态正向云图。

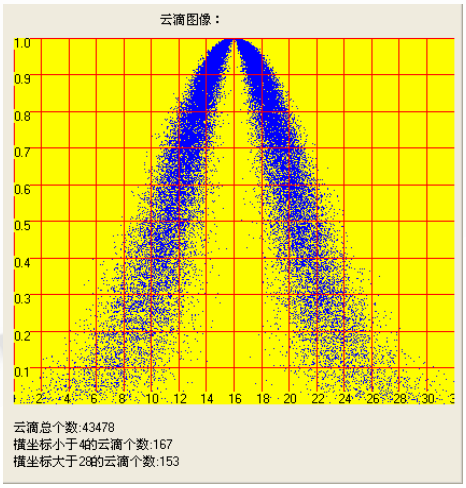

图 2

10　正态分布真随机数云模型数发生器的应用

在目前云发生器应用中，通常采用伪随机序列实现云发生器。伪随机序列虽然在一些统计特性方面接近随机序列，但是它有规律性和重复性，伪随机序列是可预测的，很难满足大型复杂控制系统的控制要求。真随机数是来源于真实的随机物理过程，因此彻底地消除了伪随机数的周期性问题，这对于云发生器的应

用有着重要的意义。

考虑真随机数实现的云发生器，根据对控制任务和实际模型的描述可以产生类似于比例函数、指数函数、正弦函数和抛物线曲线等映射曲线。下面分别根据不同的知识库和推理机制介绍基于真随机数的云发生器。

考虑基于真随机数的云发生器，设输入参数的云模型集合 A 的论域为 U＝[-1,1]，输出参数的云模型集合 B 的论域为 V＝[-1,1]，云模型集合 A＝{A1，A2，A3，A4，A5}，云模型集合 B＝{B1，B2，B3，B4，B5}，云模型集合采用一维云模型数字特征进行定义：

定义前件云模型为：

A1=[-1，0.3，0.01]; A2=[-0.5，0.3，0.01]; A3=[0，0.3，0.01]; A4=[0.5，0.3，0.01] ; A5=[1，0.3，0.01]。

定义后件云模型为：

B1=[-1，0.3，0.01]; B2=[-0.5，0.3，0.01];B3=[0，0.3，0.01]; B4=[0.5，0.3，0.01]; B5=[1，0.3，0.01]。

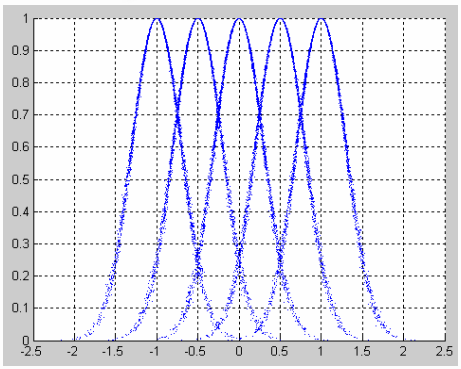

图 3 规则前件集合 A 的定性概念表示

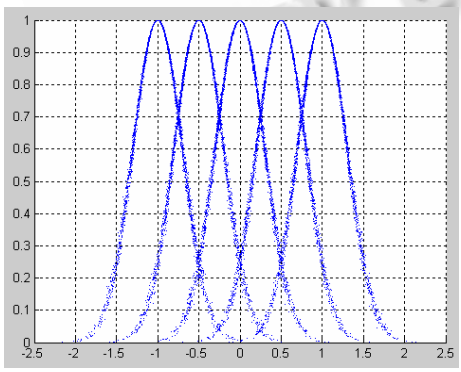

图 4 规则前件集合 B 的定性概念表示

定义推理规则为：

if A＝Ai ，then B＝Bi；i=1～5

得到的效果图如下图：

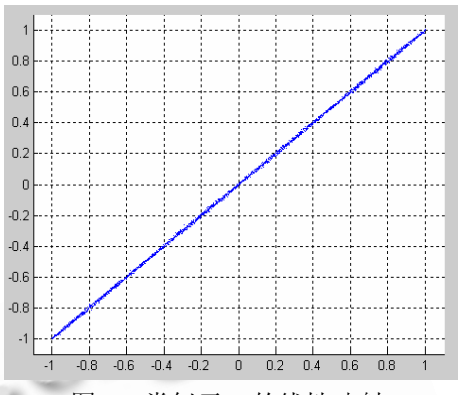

图 5 类似于 P 的线性映射

显然，基于真随机数云发生器线性映射已经接近一条清晰的直线，类似于 P 比例控制关系。相比传统的 P 比例控制关系，从图 3-图 5 可以很明显的看出基于真随机数的线性映射很好解决了独立于概念的隶属度的概率密度分布规律，揭示了云发生器 P 比例控制中随机性和模糊性的内在联系。

基于真随机数云发生器不但可以实现线性映射关系之外，还可以实现非线性映射关系，我们只要改变规则库和推理规则，就可以实现多种非线性映射关系。例如：

图 6 为在前件云（1）,后件云（2），规则库 R 中 5 条规则： if A=AK, then B=BK，K＝1~5，得到类似于抛物线的映射；

图 7 为在前件云（3）,后件云（4），规则库 R 中 13 条规则： if A=AK, then B=BK，K＝1~13，得到类似于 $y = x^3$ 的非线性映射；

图 8 为在前件云（5）,后件云（6），规则库 R 中 7 条规则：if A=AK, then B=BK，K＝1~7，得到类似于指数的非线性映射；

图 9 为在前件云（7）,后件云（8），规则库 R 中 5 条规则： if A=AK, then B=BK，K＝1~5，得到类似于正弦曲线的非线性映射。

根据这些仿真图形，同样可以很好的看到基于真随机数云发生器能很好的达到期望的效果，并且很好的进行定性概念和定量表示之间的不确定转换。

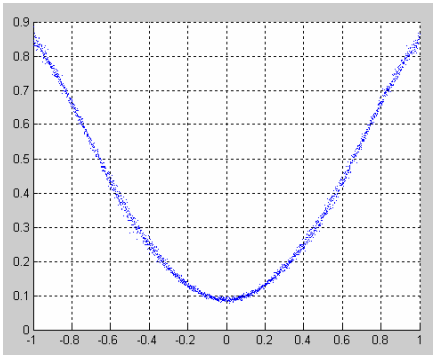

图 6　类似于抛物线的非线性映射

定义前件云模型为：

A1=[-1，0.3，0.01]; A2=[-0.5，0.3，0.01];

A3=[0，0.3，0.01]; A4=[0.5，0.3，0.01];

A5=[1，0.3，0.01]。　　(1)

定义后件云模型为：

B1=[-1，0.3，0.01]; B2=[-0.25，0.3，0.01];

B3=[0，0.3，0.01]; B4=[0.25，0.3，0.01];

B5=[1，0.3，0.01]。　　(2)

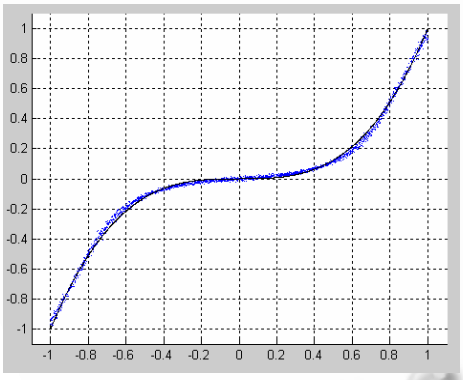

图 7　类似于的非线性映射

定义前件云模型为：

A1=[-1，0.2，0.01];A2=[-0.9，0.3，0.03];

A3=[-0.7，0.4，0.03]; A4=[-0.5，0.3，0.03] ;

A5=[-0.3，0.5，0.03]; A6=[-0.1,0.5,0.03];

A7=[0, 0.5,0.03]; A8=[0.1,0.5,0.03];

A9=[0.3,0.5,0.03];A10=[0.5,0.3,0.03];

A11=[0.7,0.4,0.03]; A12=[0.9,0.3,0.03];

A13=[1,0.2,0.01]　　(3)

定义后件云模型为：

B1=[-2.6,0.3,0.01];　B2=[-0.1,0.3,0.01];

B3=[-0.27,0.3,0.01];　B4=[0.1，0.3，0.03];

B5=[0.15，0.5，0.03]; B6=[0.2，0.5，0.03];

B7=[0，0.5，0.03];　　B8=[-0.2，0.5，0.];

B9=[-0.15，0.3，0.01];B10=[-0.1，0.3，0.01];

B11=[0.27，0.3，0.01];B12=[0.1，0.3，0.01];

B13=[2.6，0.3，0.01]　　(4)

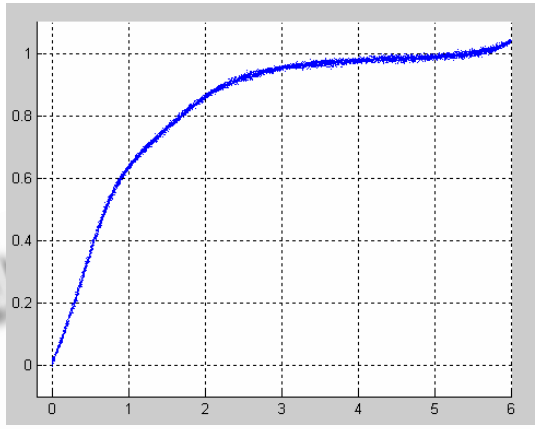

图 8　类似于指数的非线性映射

定义前件云模型为：

A1＝[0，0.4，0.01]；A2＝[1，0.7，0.01]；

A3＝[2，0.6，0.01]；A4＝[3，0.6，0.01]；

A5＝[4，0.6，0.01]；A6＝[5，0.6，0.01]；

A7＝[6，0.6，0.01];　　(5)

定义后件云模型为：

B1＝[-0.15，0.15，0.01]；B2＝[0.65，0.15，0.01]；

B3＝[9，0.15, 0.01]；　B4＝[0.96，0.15，0.01]；

B5＝[0.98，0.15，0.01]；B6＝[0.99, 0.15, 0.01]；

B7＝[0.99，0.15，0.01]。　　(6)

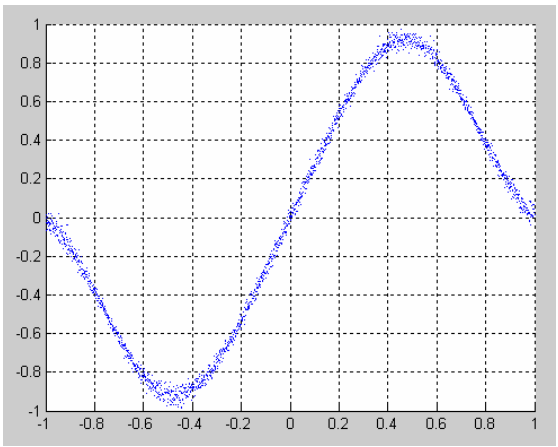

图 9　类似于正弦函数的非线性映射

定义前件云模型为：

A1＝[-1，0.3，0.01]；A2＝[-0.5，0.3，0.01]；

A3＝[0，0.3，0.01]； A4＝[0.5，0.3，0.01]；

A5＝[1，0.3，0.01]。 (7)

定义后件云模型为：

B1＝[0.5，0.3，0.01]； B2＝[-1.5，0.3，0.01]；

B3＝[0，0.3，0.01]； B4＝[1.5，0.3，0.01]；

B5＝[-0.5，0.3，0.01]。 (8)

11 结语

本文全面论述了一种正态分布真随机数云模型发生器的实现方法，是对云模型研究的有益补充，为进一步使用真随机数云模型发生器做更深入的科学研究打下了坚实基础。

参考文献

1 郭弘,刘钰,党安红,韦韦.物理真随机数发生器.中国科学(科学通报),2009,54(23):3651–3657.

2 周丽娜,沈海斌,潘洋洋,董文箫.一种无记忆的真随机数发生器.电子器件,2008,31(3):945–947.

3 沈华韵,张鹏,王侃.改进线性同余法随机数发生器.清华大学学报(自然科学版),2009,49(2):191–193.

4 王欣,周童,王永生,喻明艳.一种基于混沌原理的真随机数发生器.微电子学与计算机,2009,26(2):135–139.

5 刘晓旭,曹林,董秀成.ATme ga128 单片机的真随机数发生器.单片机与嵌入式系统应用,2009,(11):71–73.

6 周童,周志波,喻明艳,叶以正.一种基于混沌的鲁棒低功耗真随机数发生器(英文).半导体学报,2008,29(1):69–73.

7 龙银东,敬岚,方正,乔卫民.用 VHDL 实现的 23 位快速浮点数加减法器.微计算机信息,2009,25(1-2):290–291.

8 王兆红,肖孟强,李燕,刘昕.类正态分布数据云模型的预测算法.计算机应用与软件,2009,26(9):78–79.

9 詹惠琴,古军,习友宝.正态分布随机 Petri 网的串并行结构化简.电子科技大学学报,2008,37(3):424–427.

10 徐燕娟,李众,张日勋.一维多规则正态云模型映射器的算法研究.科学技术与工程,2010,10(1):244–247.

11 杨金牛,李众,杨真荣.基于遗传算法的云模型控制器设计.计算机仿真,2009,26(3):175–178.

12 王飞燕,李峰,陈松贵.基于一维正态云模型的半脆弱文本水印.计算机工程与设计,2008,29(17):4578–4580.

（上接第 70 页）

计算机系统应用,2007,16(4):57–60.

6 Mobasher B, Cooley R, Srivastava J. Automatic personalization based on WEB usage mining. Communications of the ACM, 2000,43(8):142–151.

7 Coleho B, Martins C, Almeida A. Web Intelligence in Tourism: User Modeling and Recommender System. Web Intelligence and Intelligent Agent Technology(WI-IAT), 2010,236(1):619–622.

8 Herlocker JL, Konstan JA, Terveen LG, Riedl JT. Evaluating collaborative filtering recommender systems. ACM Trans. on Information Systems, 2004, 22(1):5–53.

9 Angulo C, Ruiz F J, Gonzalez L, et a1. Multi-classification by using tri-class SVM. Neural Processing Letters, 2006, 23(1): 89–101.

10 Adomavicius MG, Tuzhilin MA. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. IEEE Trans. on Knowl and Data Eng, 2005,17(6):734–749.