2010年 11月 北京航空航天大学学报 November 2010
第36卷 第11期 Journal of Beijing University of Aeronautics and Astronautics Vol. 36 No. 11

# 正态云模型雾化性质统计分析

刘 禹 李德毅

(北京航空航天大学 计算机学院，北京 100191) (中国电子系统工程研究所，北京 100039)

**摘 要**：对于超熵较大情况下的正态云模型，说明了云模型雾化过程. 通过统计分析云滴离散的整体趋势，说明超熵增大过程中，云滴整体趋于离散. 通过分析各论域区间内云滴离散趋势，说明靠近概念核心的云滴的离散速度相对缓慢. 归纳云模型雾化性质：在超熵取值持续增大的过程中($He > En/3$)，正态云表示的概念的论域范围持续增大，呈雾化状态，但靠近概念核心的论域区间内的云滴不失数量优势. 雾化性质适用于建模偏离正态分布、缺乏共识的定量数据，期望表示概念语义值核心，熵描述概念语义的离散程度，超熵表示各种语义的共识程度，扩展了云模型知识表示的应用范围.

**关 键 词**：知识表示；不确定性；统计分析；云模型

**中图分类号**：TP 18

**文献标识码**：A **文 章 编 号**：1001-5965(2010)11-1320-05

## Statistics on atomized feature of normal cloud model

Liu Yu

(School of Computer Science and Technology, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Li Deyi

(China Institute of Electronics Engineering, Beijing 100039, China)

**Abstract**: The cloud model atomization process was related to a larger hyper enctropy. Through statistical analysis of the overall trend of the cloud drops, the cloud drops dispersed over the course of hyper entropy increase. By analyzing the dispersion trend of cloud drops in each semantic ranges, it is indicated that the drops represent the core concept dispersed in a low speed. The atomization feature of the cloud model was summarized. The semantic range of the concept represented by the cloud model extended while the hyper entropy increased step by step. The cloud drops spread but the drops nearby the core semantic keep a high density. The atomization feature of the cloud model was used to model the data deviates from the normal distribution. A cloud with a large hyper entropy value represented the concept lack of consensus. For each parameter, the expectation stands for the core semantic value, the entropy represents the semantic range and the hyper entropy shows the degree of consensus of the different semantics ranges. The cloud model knowledge representation application range was extended.

**Key words**: knowledge representation; uncertainty; statistics; cloud model

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. 李德毅教授提出云模型[1]，用一个统一的模型实现定性概念与定量描述之间的不确定转换，已成功应用于数据挖掘[2]、系统评估[3]等领域. 传统应用中，超熵取值较熵小. 在超熵变大的过程中，云滴分布会呈现明显的离散趋势，本文使用统计分析的方法，对云模型超熵变大过程中云

**收稿日期**：2009-10-21

**基金项目**：国家基础研究重点计划资助项目(2007CB310803)

**作者简介**：刘 禹(1980－)，男，河北辛集人，博士生，liuyu8014@163.com.

滴分布规律进行分析,并将其定义为正态云模型的第 4 个数学性质.

## 1　云 模 型

云模型使用 3 个数字特征:期望 $Ex$、熵 $En$ 和超熵 $He$ 来表征定性概念,它们反映了定性概念 C 的整体特性[4].

期望 $Ex$:云滴在论域空间分布的期望,是最能够代表定性概念的点,反映这个概念的云滴群的重心.

熵 $En$:定性概念的不确定性度量,由概念的随机性和模糊性共同决定,反映了概念外延的离散程度和模糊程度.

超熵 $He$:超熵是熵的不确定性的度量,即熵的熵,反映了二阶不确定性,是对熵反映的不确定性的再描述.

可以计算求出任意一个云滴属于这个概念的隶属度,但是该隶属度不是一个确定的值,而是一个具有稳定倾向的随机数[4],正态云模型中,使用隶属度 $\mu$ 刻画云滴对概念的贡献.

云模型的示意图如图 1 所示.

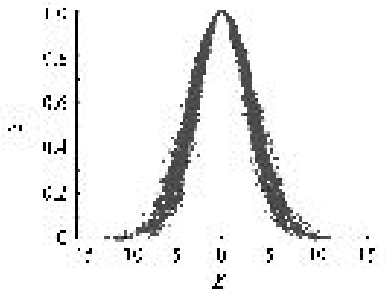


图 1　云模型示意图

($Ex=0$, $En=3$, $He=0.3$, 云滴数 $n=10\,000$.)

由云模型的定义及正态云发生器算法可知,从统计学角度,正态云模型具有 3 点数学性质[4],3 个数学特征说明:正态云模型 X 的分布可以退化为正态分布,由于正态分布的普适性[4],应用云表示不确定概念时,往往采用较小的超熵,此时云接近于正态分布;正态云模型确定度 Y 的分布与云的数字特征无关;正态云的几何形状特点明显,存在云心曲线.然而,性质 1 和性质 3 均是在超熵 $He$ 取值较小的情况下研究正态云模型特征.

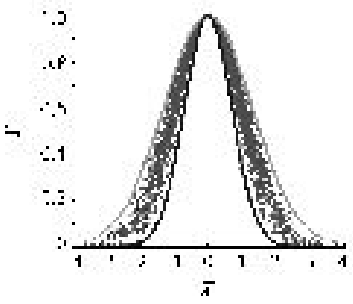
## 2　正态云模型的雾化性质

超熵的存在使正态云模型区别于正态分布,$He$ 取值为 0 时,离散的云滴勾勒出正态分布的形态;$He$ 取值较小时,云滴呈泛正态分布状态;$He$ 取值较大时,云滴所呈现的形状明显区别于正态分布:外围云滴更加分散,核心云滴出现明显的集中趋势,云的期望曲线不再明显,将超熵取值较大时的云称之为"雾".随着 $He$ 的变化,正态云由一个极端(正态分布)到另一个极端(充分离散)的变化过程称之为雾化.

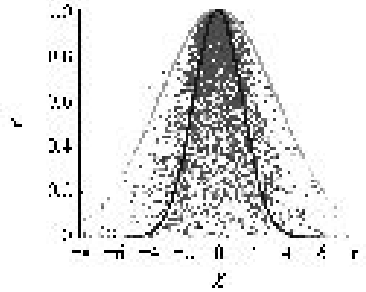
### 2.1　雾化的形成过程

随着 $He$ 逐渐增大,云滴开始离散,由于 $P\{En-3He<En'<En+3He\}=0.997$,由正态云发生器算法可知,当 $En-3He>0$,即 $He<En/3$ 时,99.7%的云滴落在曲线 $y_1=\exp\left[-\frac{(x-Ex)^2}{2(En+3He)^2}\right]$ 和 $y_2=\exp\left[-\frac{(x-Ex)^2}{2(En-3He)^2}\right]$ 所围的区域内[4],见图 2a.

$He>En/3$ 时,部分云滴均逃离了两曲线所夹范围,见图 2b. 正态云的形态在 $He=En/3$ 时出现分界,可以将 $En/3$ 称作正态云模型的雾化点,当 $He<En/3$,云滴呈现泛正态状态,$He>En/3$ 时,呈现雾化状态.



a　$He=0.1En$



b　$He=0.8En$

图 2　雾化的形成过程

### 2.2　云滴分布规律统计分析

研究雾化状态下正态云模型的云滴分布规律,可以考察正态云模型 X 的分布. 正态云模型中,所有云滴 $x$ 构成随机变量 X. $En'$ 服从以 $En$ 为期望值,$He^2$ 为方差的正态分布,X 的概率密度函数没有明确的解析形式[4],可采用统计分析方法研究正态云模型雾化状态下云滴的分布规律.

### 2.3　云滴离散的整体趋势

**定义 2.1**　云论域区间 $C_d$. 云论域区间分割了正态云所表示概念的论域范围,若记云 C(X)

中所有云滴的集合为 Drops=$\{x|x\in C(X)\}$，则论域区间 $C_d$ 为云滴在 $X$ 轴上的投影（投影点与 $Ex$ 的距离不大于 $d$）所构成的区域. 对于二维正态云模型，如图3所示，论域区间可以看作以概念核心为中心，以变量 $d$ 为半径的圆形.

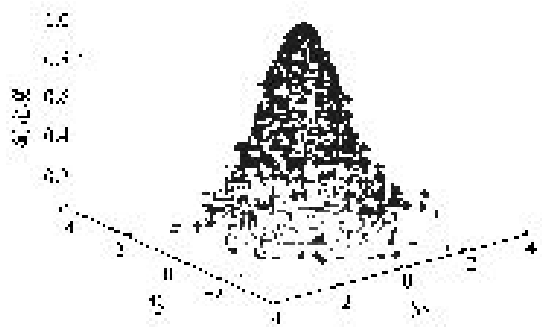


图3 二维正态云的论域区间

云图中拥有无穷多个论域区间，论域区间的中心是云的期望，由于云滴的离散特性，不存在精确的最大论域区间（$C_T$）. 在实验过程中为了计算的可行性，对于一维正态云模型，可以近似地将最大论域区间定义为云滴最小值 min(Drops)到最大值 max(Drops)之间的直线距离.

**定义 2.2** 云密度. 云密度用来表示投影在论域区间单位宽度或单位面积上的云滴数，一维正态云记为 $\rho=\Delta c/\Delta d$，二维正态云记为 $\rho=\Delta c/\Delta s$. $\rho$ 代表云密度，反映某个论域区间上云滴的密集程度；$\Delta d(\Delta s)$ 表示单位宽度（面积）；$\Delta c$ 代表在 $\Delta d(\Delta s)$ 之上的云滴个数.

由于云滴本身是离散的点，对一维正态云模型，统计 $X$ 的某一区间 $A(A\in C_d)$ 上的密度更有意义，可以近似地用投影在区间 A 上的云滴个数与区间长度的比值来表示区间云密度，记为 $\rho_A$，而将云滴整体的平均密度记为 $\rho_V$.

**实验 1**

1）取数字特征 $Ex=0$，$En=1$，$He=h$，通过正态云发生器生成云滴（$n=1\,000$，$h$ 初值为0）；

2）计算云近似最大论域区间 $C_{Ti}$ 与云平均密度 $\rho_{Vi}$，$i$ 初值为0；

3）重复步骤1）～2），$i=i+1$，5 000次，得到正态云模型在 $He$ 为 $h$ 时的平均密度 $\rho_{Vh}$；

4）变化参数 $h$，$h=h+0.01En$，重复步骤1）～3），分别计算出 $He$ 在区间$[0,10En]$上的近似最大论域区间和云密度.

实验1结果表明，随着超熵的变大，云最大论域区间宽度呈线性增长趋势，而云团的整体密度呈下降趋势，下降过程中，随 $He$ 变大，云密度下降趋势趋于缓和.

### 2.4 典型论域区间内云滴离散趋势

本部分研究各个典型论域区间内的云密度变化情况，论域区间的选取仍旧沿用已有区间范围和命名方式[4]：骨干区间 $A[Ex-0.67En,Ex+0.67En]$，基本区间 $B[Ex-En,Ex+En]$，外围区间 $C[Ex-2En,Ex+2En]$，弱外围区间 $D[Ex-3En,Ex+3En]$.

**实验 2**

1）取数字特征 $Ex=0$，$En=1$，$He=h$，通过正态云发生器生成云滴（$n=1\,000$，$h$ 初值为0）；

2）计算云滴 $X$ 投影于骨干区间、基本区间、外围区间、弱外围区间内的云滴数，从而得到各区间内的云密度 $\rho_{Ai}$，$\rho_{Bi}$，$\rho_{Ci}$ 和 $\rho_{Di}$，$i$ 初始0；

3）重复步骤1）～2），5 000次，得到正态云模型在 $He$ 为 $h$ 时各区间密度的均值 $\rho_A$，$\rho_B$，$\rho_C$，$\rho_D$；

4）变化参数 $h$，$h=h+0.01En$，重复步骤1）～3），分别计算出 $He$ 在区间$[0,10En]$上各区间的区间密度.

表1给出了实验2的具有典型代表性的部分结果.

**表1 论域区间密度变化表**

| $He$ | $\rho_A$ | $\rho_B$ | $\rho_C$ | $\rho_D$ |
|---|---|---|---|---|
| $0.1En$ | 746.618 | 684.703 | 476.426 | 332.178 |
| $0.2En$ | 759.619 | 689.826 | 473.600 | 331.422 |
| $0.4En$ | 805.643 | 700.814 | 464.924 | 328.619 |
| $0.6En$ | 835.836 | 700.282 | 454.92 | 324.718 |
| $0.8En$ | 839.702 | 690.704 | 445.051 | 320.309 |
| $1En$ | 822.084 | 673.664 | 435.083 | 315.629 |
| $2En$ | 664.636 | 555.901 | 379.425 | 287.734 |
| $3En$ | 539.564 | 461.064 | 329.959 | 259.050 |
| $5En$ | 394.822 | 344.777 | 260.374 | 213.061 |
| $8En$ | 285.349 | 253.797 | 199.379 | 168.661 |

从表1实验结果中可以看出：当 $He<En$ 时，随着 $He$ 的增大，骨干区间密度 $\rho_A$ 没有减小反而增大. 当 $He>En$ 时，$\rho_A$ 呈现出减小的趋势，此时才与云滴整体密度的趋势相符合. 基本区间与骨干区间有类似表现. 外围区间和弱外围区间上的密度变化，在 $He$ 很小时达到峰值，在表1中无法得到密度变大的过程，总体呈现出减小的趋势，与云整体密度的变化趋势相符合. 结论如下：

1）考虑超熵变大整体过程，各区间内的云滴都呈现离散趋势，但是不同区间内的云滴的离散速率不同；

2）越靠近概念核心（$Ex$），云滴密度越大，且在 $He$ 增大过程中，一直保持较外围区间的密度优势；

3）随着 $He$ 增大，骨干区间和基本区间内的云滴会出现一个密度增大的过程，到达极大值后，

呈下降状态.

在雾化过程中,云滴呈现抱团特性,靠近概念核心区域内的云滴密度明显高于外围区间的云滴密度.

设云 Cloud($Ex$,$En$,$He$),包括 N 个云滴,在 X 轴的投影落在区间[$Ex-\delta$,$Ex+\delta$]范围的云滴个数为 $m$.有

$$m = \mathrm{N} \times P\{|\ X - Ex\ | < \delta\} \tag{1}$$

据 $3\delta$ 规则[4],得

$$m \approx \mathrm{N} \times P\{3\ |\ En\ | \leqslant \delta\} \tag{2}$$

据正态云发生器算法,得

$$m \approx \mathrm{N} \times \left(\Phi\left(\frac{\delta - 3En}{3He}\right) - \Phi\left(\frac{-\delta - 3En}{3He}\right)\right) \tag{3}$$

显然,$m$ 与投影区间范围 $\delta$ 相关,在使用云模型表示定性概念时,希望核心云滴数目最多,亦即取 $\delta = En$,设 $He = kEn$,此时

$$m \approx \mathrm{N} \times \left(\Phi\left(\frac{-2}{3k}\right) - \Phi\left(\frac{-4}{3k}\right)\right) \tag{4}$$

通过数据拟合可以得到,$k = 0.98$ 时,$m$ 取最大值.所以,在雾化过程中,当 $He = 0.98En$ 时[$Ex-En$,$Ex+En$]区间内的核心云滴数量达到最大值.此时云模型适用于表示“难于形成共识的概念”:所谓“共识”,指的是观察值存在明显的多数核心;而“难于形成共识”是指在概念核心 $Ex$ 附近的区间之内,定量数据已经不再呈明显的阶梯型分布,彼此的数量十分接近,难分伯仲.虽然数据的核心不能由一个确定的取值来表示,但可以使用一系列接近的数值来表示,故仍旧称之为“概念”.极端情况下,当样本取值充分离散后,有限的定量数据已经不能够形成定性概念.

## 3　雾化性质应用举例

已有的云模型应用,大多采用较小的 $He$,雾化特征的提出,丰富了云模型在知识表示领域的应用范围.

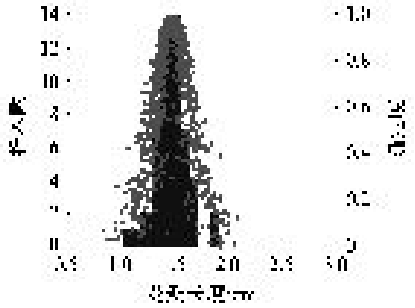
### 3.1　雾化性质表示控制知识

在基于云模型的进化算法中,使用云模型的雾化性质,通过超熵可以控制期望附近云滴的比例和远离期望的云滴的比例,从而达到定性控制进化方向的目的,而这种控制方法可以有效地保持基因(概念)的遗传特性并体现变异特性,而不是片面地强调一方,使得进化算法可以达到大范围、高精度的执行效果.实验结果表明,应用云模型的雾化性质指导进化过程,可得高精度进化算法[5-6].

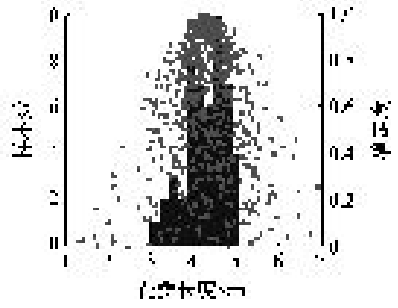
### 3.2　雾化性质与定量-定性转换

对于给定的定量数据集合,如果数据本身符合泛正态分布,使用逆向云发生器[4],可得到这些数据所代表的定性概念云描述.如果定量数据来源于随机实验结果,且数据量不够大,即使客观上符合正态分布,传统的逆向云发生器算法不能有效地得到数据所表示概念的定性描述.基于云模型的分类算法[7]中,将云模型雾化性质与逆向云发生器相结合,对缺乏共识的定量数据可以采用扩大超熵的办法进行表示.雾化性质的提出扩展了云模型知识表示的范围.

例如:采用 Iris 数据集[8],图 4 中给出了花瓣长度数据的云表示.图 4a 表示 setosa 类型的鸢尾花的花瓣长数据,图 4b 表示 versicolor 类型的鸢尾花的花瓣长数据.由柱状图可看出,setosa 的花瓣长度统计接近正态,表示其分布接近于正态分布,可以使用云 Cloud(1.464,0.164,0.056)表示其花瓣长:花瓣长度分布在 1.464 周围,熵为 0.164,且熵较为稳定($He$=0.056);针对 versicolor 鸢尾花,其花瓣长度比较接近,[3.8,5]区间内的花瓣长取值统计数量没有明显差异,样本统计结果表示缺乏概念共识.此时可以使用 Cloud(4.26,0.47,0.465)表示 versicolor 的花瓣长,$He$ 取值较大且 $He = 0.98En$,落在[4.26−0.47,4.26+0.47]区间内云滴数最多,最大限度地保证了靠近概念核心的样本数量,使用雾化性质实现了对不同统计样本的统一建模.



a　setosa



b　versicolor

图 4　Iris 数据集花瓣长度的云表示

## 4　结　论

本文使用统计学方法论述了正态云模型在超

熵增大过程中的云滴分布特征，拓展了正态云模型的数学性质，可定义正态云模型的数学性质4：在超熵取值持续增大的过程中（$He > En/3$），正态云表示的概念的论域范围持续增大，呈雾化状态，但靠近概念核心的论域区间内的云滴不失数量优势. 雾化状态下的超熵反映了定量数据对定性概念的共识程度. 正态云模型雾化性质的提出，为云模型在知识表示与定性-定量转换方面的应用提供了新的思路和依据. 下一步研究工作可将雾化性质应用于其他领域，充分验证其在知识表示领域的有效性.

## 参考文献（References）


[1] 李德毅，刘常昱，杜鹢，等. 不确定性人工智能[J]. 软件学报，2004，15(11)：1－13
Li Deyi，Liu Changyu，Du Yi，et al. Artificial intelligence with uncertainty[J]. Journal of Software，2004，15(11)：1－13(in Chinese)

[2] Wang Shuliang，Li Deren，Shi Wenzhong，et al. Cloud model-based spatial data mining[J]. Geographical Information Science，2003，9(2)：67－78

[3] 吕辉军，王晔，李德毅. 逆向云在定性评价中的应用[J]. 计算机学报，2003，26(8)：1009－1014
Lü Huijun，Wang Ye，Li Deyi. The application of backward cloud in qualitative evaluation[J]. Chinese Journal of Computers，2003，26(8)：1009－1014 (in Chinese)

[4] 李德毅，杜鹢. 不确定性人工智能[M]. 北京：国防工业出版社，2004
Li Deyi，Du Yi. Artificial intelligence with uncertainty[M]. Beijing：National Defence Industry Press，2004(in Chinese)

[5] 张光卫，李德毅，刘禹. 基于正态云模型的进化算法[J]. 计算机学报，2008，7(7)：1082－1091
Zhang Guangwei，Li Deyi，Liu Yu. An evolutionary algorithm based on cloud model[J]. Chinese Journal of Computers，2008，7(7)：1082－1091(in Chinese)

[6] 张光卫，康建初，李鹤松，等. 基于云模型的全局最优化算法[J]. 北京航空航天大学学报，2007，33(4)：486－490
Zhang Guangwei，Kang Jianchu，Li Hesong，et al. Cloud model based algorithm for global optimization of functions[J]. Journal of Beijing University of Aeronautics and Astronautics，2007，33(4)：486－490(in Chinese)

[7] Liu Yu，Chen Guisheng. Cloud model based classifier[C]// Luo Qi，Tan Honghua. 2009 Internal Conference on Test and Measurement. Hong Kong：IEEE，2009：427－430

[8] Blake C L，Merz C J. UCI repository of machine learning databases[DB/OL]. Irvine，CA：University of California，1998. http://www.ics.uci.edu/～mlearn/MLRepository.html


（编　辑：文丽芳）

（上接第1310页）

## 4 结　论

针对双转子航空发动机机匣拾取振动信号的特征，基于提高航空发动机振动状态监测和故障诊断准确性的目的，采用基于数学形态滤波前处理的盲源分离识别框架，对实测发动机振动数据的分析表明：

1）开—闭和闭—开组合数学形态滤波器不仅能够很好地滤除掉原始信号中的噪声分量，而且能够较好保留原信号的基本形状；

2）盲源分离技术虽然存在幅值和顺序上的不确定性，但分离出的信号能够显著削弱彼此间的混叠，源信号的波形得到了较好的恢复，因此能够进一步提高故障定位的精度.

## 参考文献（References）


[1] Shi D F，Qu L D，Gindy N N. General interpolated fast fourier transform：a new tool for diagnosing large rotating machinery[J]. Journal of Vibration and Acoustics，2005，127(4)：351－361

[2] 胡爱军，唐贵基，安连锁. 基于数学形态学的旋转机械振动信号降噪方法[J]. 机械工程学报，2006，42(4)：127－130
Hu Aijun，Tang Guiji，An Liansuo. De-noising technique for vibration signals of rotating machinery based on mathematical morphology filter[J]. Chinese Journal of Mechanical Engineering，2006，42(4)：127－130(in Chinese)

[3] 郭景峰，申光宪，郑绳楦，等. 数学形态学在数字滤波中的应用研究[J]. 机械工程学报，2002，38(10)：144－147
Guo Jingfeng，Shen Guangxian，Zhen Shengxuan，et al. Application of mathematical morphology in digital filter[J]. Chinese Journal of Mechanical Engineering，2002，38(10)：144－147(in Chinese)

[4] 张发启. 盲信号处理及应用[M]. 西安：西安电子科技大学出版社，2006
Zhang Faqi. Operation and application of blind signal[M]. Xi'an：Xi'an Electronical Science and Technology University Publishing Company，2006( in Chinese)

[5] Hyvärinen A. Fast and robust fixed-point algorithm for independent component analysis [J]. IEEE Transactions on Neural Networks，1999，10 (3)：626－634

[6] Hyvärinen A，Oja E. A fast fixed-point algorithm for independent component analysis[J]. Neural Computation，1997，9(7)：1483－1492

[7] 秦海勤，徐可君，江龙平. 某型航空发动机机载振动监测点选择[J]. 推进技术，2007，28(6)：697－702
Qin Haiqin，Xu Kejun，Jiang Longping. Research on distribution of airborne vibration monitoring measuring points for the aeroengine[J]. Journal of Propulsion Technology，2007，28(6)：697－702( in Chinese)


（编　辑：张　嵘）