Journal of System Simulation

Volume 31 | Issue 7 Article 3

12-12-2019

Credibility Evaluation Method of Small Sample Data Based on Cluster Cloud Model

Jianmin Wang
1. Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences, Beijing 100094, China; ;

Yunjie Wu
2. School of Automation Science and Electrical Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China;

Credibility Evaluation Method of Small Sample Data Based on Cluster Cloud Model

Abstract

Abstract: In practical engineering, the system credibility evaluation with small amount of data and no evaluation criteria has always been a difficult problem for engineers. Aiming at this problem, *a small sample data credibility evaluation method combining clustering method and cloud model is proposed. The clustering method is used to calculate the cluster center value for the small sample, and the cloud model is established based on this. The expanded value for small sample data is generated by the cloud model. The credibility of small sample data can be calculated according to the confidence distribution of cloud drops.* It can fully exploit the implied information in the small sample data by combining the clustering method with the cloud model, which can increase the effectiveness of the evaluation. The case analysis and simulation are carried out to prove the validity and rationality of the proposed method.

Keywords

clustering method, cloud model, credibility evaluation, small sample data

Recommended Citation

Wang Jianmin, Wu Yunjie. Credibility Evaluation Method of Small Sample Data Based on Cluster Cloud Model[J]. Journal of System Simulation, 2019, 31(7): 1263-1271.

第 31 卷第 7 期 系统仿真学报© Vol. 31 No. 7
2019 年 7 月 Journal of System Simulation Jul., 2019

基于聚类云模型的小样本数据可信度评估

王建敏[1]，吴云洁[2]
（1. 中国科学院 空间应用工程与技术中心，北京 100094；2. 北京航空航天大学 自动化科学与电气工程学院，北京 100083）

摘要：在实际工程中，数据量小、且无评估标准的系统可信度评价问题一直是困扰工程人员的难题。针对该问题，*提出了一种将聚类算法和云模型相结合的小样本数据可信度评估方法。利用聚类算法先确定小样本中的聚类中心值，基于此建立云模型。通过云模型产生小样本的扩充数据。根据云滴的置信度分布可进一步计算小样本数据的可信度。*将聚类算法与云模型相结合，可以充分挖掘小样本数据中的潜在信息，增加评估的有效性。通过算例分析及仿真证明了所设计方法的合理性和有效性。

关键词：聚类算法；云模型；可信度评估；小样本数据
中图分类号：TP391.9 文献标识码：A 文章编号：1004-731X (2019) 07-1263-09
DOI: 10.16182/j.issn1004731x.joss.19-0202

Credibility Evaluation Method of Small Sample Data Based on Cluster Cloud Model

Wang Jianmin[1], *Wu Yunjie*[2]
(1. Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences, Beijing 100094, China;
2. School of Automation Science and Electrical Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: In practical engineering, the system credibility evaluation with small amount of data and no evaluation criteria has always been a difficult problem for engineers. Aiming at this problem, *a small sample data credibility evaluation method combining clustering method and cloud model is proposed. The clustering method is used to calculate the cluster center value for the small sample, and the cloud model is established based on this. The expanded value for small sample data is generated by the cloud model. The credibility of small sample data can be calculated according to the confidence distribution of cloud drops.* It can fully exploit the implied information in the small sample data by combining the clustering method with the cloud model, which can increase the effectiveness of the evaluation. The case analysis and simulation are carried out to prove the validity and rationality of the proposed method.

Keywords: clustering method; cloud model; credibility evaluation; small sample data

引言

在日常生活中，人们经常需要对一个设备、一个系统甚至对一个人进行评价。在工程中，更多的是对设备的性能、系统的运行状况进行评价。随着人工智能及仿真技术的飞速发展，大量智能设备及仿真系统进入生产生活当中，这些仿真系统有多大的可信度或者说在多大程度上能代替真实系统，则需要进行评估。

针对可以产生海量数据的系统的评估，可以采用统计分析、深度学习等方法，因其数据量大、信息丰富，因此评估方法相对比较简单。张文一[1]

收稿日期：2019-05-10 修回日期：2019-06-24；
作者简介：王建敏(1986-)，男，河北石家庄，博士，助理研究员，研究方向为软件评测、系统测试及控制、系统仿真及评估；吴云洁(1969-)，女，河北保定，博士，教授，博导，研究方向为智能控制理论、半实物仿真设备、系统仿真及评估。

利用序贯蒙特卡洛方法和弦截法对光伏电站的发输电系统和容量可信度进行了评估，充分利用了大数据信息。方浩等[2]针对仿真图像可信度的评估问题，建立了基于自学习框架的仿真图像相似性评估方法，实现了从样本生成到评估过程的自动化。有数据证明，海量数据信息有助于实现仿真系统的高可信度，文献[3]在此方面进行了验证。然而，在航空航天工程中，遇到的往往是小样本数据系统的评估，如飞机飞行试验、火箭发射、导弹打靶试验等，由于试验周期长、成本巨大、试验过程复杂，所获取的试验数据极其有限。此外，该设备或系统没有可供参考的标准，即没有一个系统的性能是完全准确的，无法找到参考标准。因此，没有参考标准的小样本数据系统的可信度评估是工程人员面临的一大难题。

在该问题上，已有不少学者开展了大量研究。目前，处理小样本数据的一个主要思路就是将小样本数据转换为大样本数据，即扩充数据。在扩充数据方面，可以利用神经网络生成数据。先用小样本数据训练网络，然后利用训练后的网络生成新的数据[4-5]。戴志辉等[6]将 BP 神经网络与最小二乘法进行了结合，用 BP 神经网络扩充数据，然后利用最小二乘法估计参数，进而建立了可靠性评估的模型，实现了继电保护可信度评估。有学者分析指出最小二乘支持向量机对小样本数据的拟合特性优于神经网络，王硕[7]在此基础上提出了基于粒子群优化的最小二乘支持向量机，并用它扩充了工程造价数据，完成了对建筑工程造价的预测研究。当样本数据量小于 10 h，利用 Bootstrap 方法处理小样本数据的结果将优于上述方法[8]。Bootstrap 方法是有放回的重抽样技术，可以实现任意样本的扩充。但目前处理极小样本最成熟且应用最广泛的是 Bayes 理论。Bayes 理论通过补充其他信息(先验信息)来增加样本的信息量[9]，这是它能处理极小样本数据的一大特点。文献[10-11]分别采用了 2 种方法处理小样本数据，并将 Bayes 方法与另外一种方法进行了对比，验证了 Bayes 理论的优越性。Dai 等[12]利用 Bayes 方法建立了小样本失效数据模型，采用 Monte-Carlo 方法评估了继电保护系统的可靠性。王超[13]、胡利平等[14]则将 Bayes 方法、加权检验法、层次分析法等相结合，设计了小样本数据评估的综合评估方法。

在系统可信度评估中，除数据量少的情况外，还存在不确定性的问题。李德毅院士[15]提出的云模型可以用来处理系统的不确定性，在数据处理、人工智能等方面已有广泛应用。目前已有学者将云模型理论应用于系统可信度评估中。王玉珏[16]通过定性和定量 VV&A 结果云化，构建了指标权重云模型，解决了可信度评估结果的量化问题。郑宇[17]提出了基于均匀分布的一维逆向云算法，对仿真系统的可信度进行了评估。以上方法均是先建立评估模型，然后借助云模型的定性定量转换特性实现可信度评估，并没有完全利用云模型处理不确定性的特性。提出了一种聚类方法和云模型理论相结合的计算方法，充分利用了云模型能够处理不确定性的特点来挖掘小样本数据的信息，并利用云模型的确定度分布情况计算原始数据的可信度。

1 聚类算法及云模型

1.1 聚类算法

聚类是一种将数据集分成不同的类簇的数据分析工具。相似性较高的数据分在同一类簇中，相似度较低的数据分处于不同的类簇。聚类算法是数据挖掘方法的重要研究对象，其本质是对数据之间的内在联系进行搜索和量化。聚类算法最终是要获得数据集内部的分布规律，进而可以简化或易于挖掘数据信息[18-19]。聚类算法的定义可描述如下：

设有一数据集 $X=\{x_1,x_2,\cdots,x_n\}$，对其进行聚类分析后产生一个集合 $C=\{C_1,C_2,\cdots,C_k\}$，该集合中的任意元素也是数据 $x_i\,(i=1,2,\cdots,n)$ 的集合，且满足

$$\begin{aligned} &C_i(i=1,2,\cdots,k)\subset X\\ &C_1\cup C_2\cup\cdots\cup C_k=X\\ &C_i\cap C_j=\varnothing, i\neq j \end{aligned} \tag{1}$$

聚类算法根据数据处理方式的不同，可分为 6 大类：层次聚类、划分聚类、网格聚类、密度聚类、模型聚类和图聚类。在实际应用中，不同的场景需求的聚类算法各不相同，需要根据不同的应用需求选择最适合的聚类算法[20]。

1.2 云模型

云模型是一种实现定性概念与定量表示之间相互转换的双向认知模型，可以表示自然语言中概念的不确定性，且能反映随机性和模糊性之间的关联关系，构成了定性和定量之间的相互映射[15]。云模型的定义[21]表示为

给定定量论域 W，设 G 是 W 上的定性概念，选择定量值 $x \in W$，且 x 是 G 的一次随机实现，设 $\mu(x) \in [0,1]$ 是 x 对 G 的确定度且满足

$$\mu: W \to [0,1] \quad \forall x \in W \quad x \to \mu(x) \tag{2}$$

则 x 在论域 W 上的分布称为云，每一个 x 称为一个云滴。

云模型是由大量的云滴组成的，云滴之间是随机的，无时序性，一个云滴就是定性概念的一次定量实现。云滴的数量越多，反映的定性概念的特征越多。

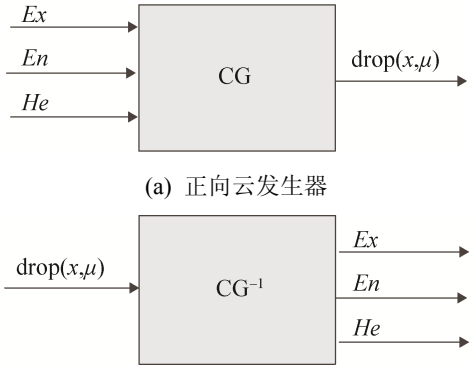
云模型用 3 个数字特征来表示：期望 Ex (Expected Value)、熵 En (Entropy)和超熵 He (Hyper Entropy)。Ex 是对定性概念的基本信息的度量，是云滴在论域空间分布的数学期望；En 反映的是定性概念的不确定性，由概念的随机性和模糊性共同决定；He 是熵的不确定性的度量，也称为熵的熵，或二阶熵[15]。

云滴可以用论域空间值和其确定度来表示，即 $\mathrm{drop}(x,\mu)$，其中 μ 是 x 对 G 的确定度，满足

$$\mu = e^{-\frac{(x-Ex)^2}{2(En')^2}}, \quad En' \sim N\left(En, He^2\right) \tag{3}$$

式中：En' 是以 En 为期望值，以 He^2 为方差的一个高斯随机数。

云模型是定性概念和定量表示的相互转换模型，因此其正向云发生器和逆向云发生器可表示为图 1 所示的形式。



(a) 正向云发生器

(b) 逆向云发生器

图 1 云发生器

Fig. 1 Cloud generator

云模型数据的分布形式如图 2 所示。

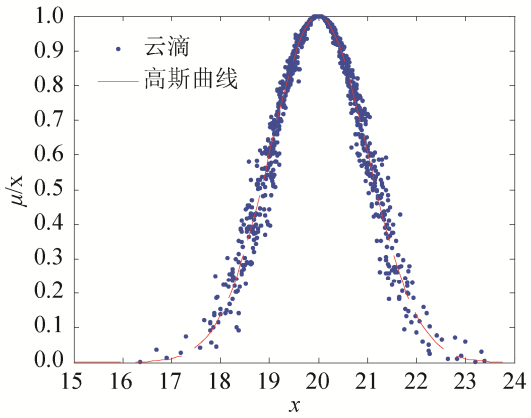


图 2 云模型分布

Fig. 2 Distribution of cloud model

从图 2 可见，云滴数据的分布与高斯曲线的趋势是一致的。靠近高斯曲线峰值的云滴数据的确定度高，云滴之间的密集度越大；靠近高斯曲线波谷的云滴数据的确定度低，且云滴之间分散度变大。

2 聚类云模型

将聚类算法与云模型理论相结合，挖掘小样本数据的分布信息并将其通过云模型放大，生成更多的样本数据，以便于可信度的评估。

定义 1 给定一个 R^n 维数据样本 $X = \{x_1, x_2, \cdots, x_n\}$，则任意两个样本元素之间的距离定义为

$$d_{ij}=\begin{cases}\left|x_i-x_j\right|, i\neq j\\ 0 \quad ,i=j\end{cases} \tag{4}$$

式中：$i,j=1,2,\cdots,n$。

定义 2 对于一个给定的数据样本 $X=\{x_1,x_2,\cdots,x_n\}$，且 $X\in R^n$，则任一样本元素 $x_i\in X$ ($i=1,2,\cdots,n$)在该样本域内的密度定义为

$$\rho_i=N\{x_j\in X\mid d_{ij}<\varepsilon\} \tag{5}$$

式中：$N\{\bullet\}$ 为计算满足条件的样本元素的个数的函数；ε 为任意小的实数。

定义 3 对于给定的数据样本 $X=\{x_1,x_2,\cdots,x_n\}\in R^n$，该样本的聚类中心值定义为

$$C_X=\sum_{i=1}^{n}(\omega_i\cdot x_i) \tag{6}$$

式中：ω_i 为样本 x_i 在样本域中的权重，定义为

$$\omega_i=\frac{\rho_i}{\sum_{j=1}^{n}\rho_j} \tag{7}$$

在聚类云模型中，将样本的聚类中心值作为云模型的期望值，即

$$Ex=C_X \tag{8}$$

定义 4 给定数据样本 $X=\{x_1,x_2,\cdots,x_n\}\in R^n$，则样本的类一阶绝对中心矩定义为

$$F=\frac{1}{n}\sum_{i=1}^{n}\left|x_i-C_X\right| \tag{9}$$

样本的类二阶中心矩定义为

$$S=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-C_X)^2 \tag{10}$$

定理 1 给定数据样本 $X=\{x_1,x_2,\cdots,x_n\}\in R^n$，根据其建立聚类云模型，则聚类云模型的熵可根据类一阶绝对中心矩计算，即

$$\hat{En}=\sqrt{\frac{\pi}{2}}\times F \tag{11}$$

证明[15]：根据高斯云模型的定义，所有的云滴服从高斯随机变量 X 的分布。假设存在一个中间随机变量 M，它与 X 构成条件概率关系。

根据高斯云算法，假设随机变量 M 服从均值为 En、标准差为 He 的高斯分布，则 M 的概率密度函数为

$$f(m)=\frac{1}{\sqrt{2\pi He^2}}e^{-(m-En)^2/(2He^2)},\forall m\in W \tag{12}$$

设 X 服从均值为 Ex、方差为 m^2 的高斯分布，则它是建立在随机变量 M 基础上的高斯分布，则它的概率密度函数为

$$f(x\mid m)=\frac{1}{\sqrt{2\pi m^2}}e^{-(x-Ex)^2/(2m^2)},\forall x\in W \tag{13}$$

根据条件概率公式，高斯云 X 分布的概率密度函数[22]为

$$f(x)=\int_{-\infty}^{+\infty}f(x\mid m)f(m)=\frac{1}{\sqrt{2\pi m^2}}e^{-(x-Ex)^2/(2m^2)}\cdot\frac{1}{\sqrt{2\pi He^2}}e^{-(m-En)^2/(2He^2)}\mathrm{d}m \tag{14}$$

根据概率密度函数，样本 X 的类一阶绝对中心矩也可计算为

$$E\{\left|X-Ex\right|\}=\int_{-\infty}^{+\infty}\left|x-Ex\right|f(x)\mathrm{d}x=\sqrt{\frac{2}{\pi}}\int_{-\infty}^{+\infty}\frac{|m|}{\sqrt{2\pi He^2}}e^{-(m-En)^2/(2He^2)}\mathrm{d}m \tag{15}$$

这里的积分都属于反常积分，反常积分Γ函数的积分公式[23]为

$$\int_0^{+\infty}e^{-u^2}u^t\mathrm{d}u=\frac{1}{2}\Gamma\left(\frac{1+t}{2}\right),(t>-1) \tag{16}$$

反常积分具有如下特性：

$$\Gamma(s+1)=\Gamma(s)$$
$$\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi} \tag{17}$$

根据公式(16)~(17)，公式(15)可以简化为

$$E\{\left|X-Ex\right|\}=\sqrt{\frac{2}{\pi}}En \tag{18}$$

此外，根据定义 4，样本 X 的类一阶绝对中心矩也可表达为

$$E\{\left|X-Ex\right|\}=F=\frac{1}{n}\sum_{i=1}^{n}\left|x_i-C_X\right| \tag{19}$$

结合公式(18)~(19)，可得

$$\hat{En}=\sqrt{\frac{\pi}{2}}\times F \tag{20}$$

定理得证。

定理 2 给定数据样本 $X=\{x_1,x_2,\cdots,x_n\}\in R^n$，根据该样本建立聚类云模型，则聚类云模型的超熵可根据定理 1 及样本的类二阶中心矩计算，即

DOI: 10.16182/j.issn1004731x.joss.19-0202

$$\hat{He}=\sqrt{S-\hat{En}^2} \tag{21}$$

证明：同定理 1，样本 X 服从高斯分布，其概率密度函数为式(14)所示的形式。则样本 X 的类二阶中心矩用概率密度函数计算为

$$D(x)=E[(X-Ex)^2]=\int_{-\infty}^{+\infty}(x-Ex)^2 f(x)\mathrm{d}x=$$

$$\int_{-\infty}^{+\infty}\frac{m^2}{\sqrt{2\pi He^2}}e^{-(m-En)^2/(2He^2)}\mathrm{d}m \tag{22}$$

该积分仍然是反常积分，根据公式(16)和(17)，可将上式化简为

$$D(x)=He^2+En^2 \tag{23}$$

此外，根据定义 4，样本 X 的类二阶中心矩为

$$D(x)=S=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-C_X)^2 \tag{24}$$

因此，根据公式(23)~(24)以及定理 1，可得

$$\hat{H}=\sqrt{S-\hat{En}^2} \tag{25}$$

定理得证。

至此，从公式(8)及定理 1 和定理 2 可见，云模型的 3 个参数 Ex，En，He 均利用样本 X 的数据得到了计算值。因此，聚类云模型建立完成，根据该模型可进行样本 X 数据的可信度评估。

3 可信度评估方法

根据上述建立的聚类云模型，可生成 N 个云滴数据。生成的云滴数量越多，可信度评估的准确度越高，同时计算量也会越大。通过聚类云模型生成的数据将覆盖原始样本数据，即原小样本中的数据将分散于 N 个云滴数据中，如图 3 所示。

从图 3 可见，每一个原始样本数据附近都有很多云滴包围，所有的云滴的确定度均已明确(即纵坐标值)。为了确保计算的可信度值更加准确，为了减小随机性对评估结果的影响，选取原始样本数据周围某一邻域范围内的云滴数据，作为进一步计算的样本点，如图 4 所示。

在图 4 中，圆圈范围内的云滴数据为选定的新的样本点，利用聚类算法，求这些新的样本点的聚类中心值，该聚类中心值的确定度将作为所求样本元素的可信度值。

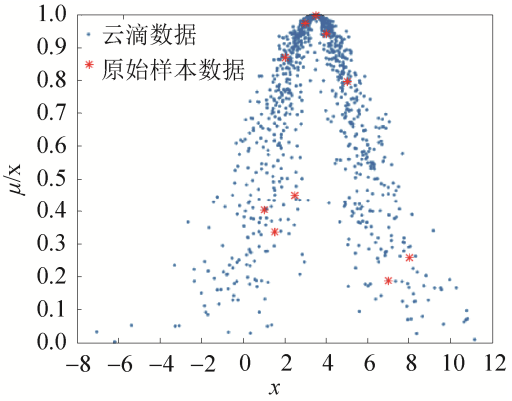


图 3 原始样本数据在云模型中的分布

Fig. 3 Distribution of raw sample data in cloud model

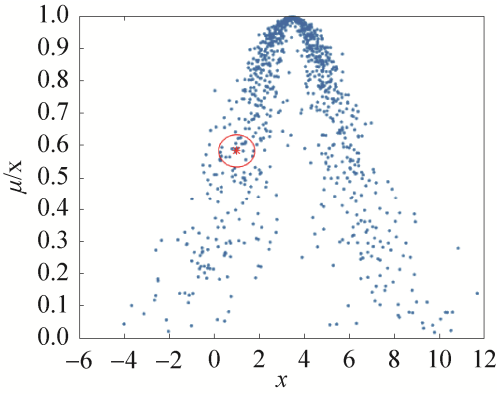


图 4 选择邻域云滴数据

Fig. 4 Selecting neighborhood cloud drop data

因此，所设计的计算样本 $X\in R^n$ 可信度的步骤如下：

步骤 1 根据定义 2 计算小样本中各个数据的密度值；

步骤 2 根据公式(7)及样本密度值，计算各个样本数据在样本域中的权重；

步骤 3 利用公式(6)计算给定的小样本的聚类中心值，并将其作为聚类云的期望值 E_x；

步骤 4 根据定义 4 计算给定的小样本的类一阶绝对中心矩 $\boldsymbol{F}$ 和类二阶中心矩 $\boldsymbol{S}$；

步骤 5 根据定理 1 和定理 2，计算聚类云的熵 $\hat{En}$ 和超熵 $\hat{He}$；

步骤 6 利用上述建立的聚类云模型生成 $N(N\geq 1\,000)$个云滴数据；

步骤 7 参照图 4 为每个原始样本原始选择 δ(δ 为任意小的实数)邻域内的云滴数据，组成新样本，

样本元素数量记为 k；

步骤 8 若 k<3，返回步骤 7，调整 δ 的值，重新选择邻域样本；否则继续执行；

步骤 9 计算每一个新样本的聚类中心值，结合云模型计算聚类中心值的确定度。

按照上述步骤计算得到的 n 个确定度的值，即为原始小样本中 n 个样本元素的可信度值。

4 算例分析

在本部分中将列举几个算例对上述方法进行仿真分析。

算例 1

选取一个小样本数据 $X=\{1,5,3,2,1.2,6,8,4,1.5,10\}$，该样本包含 10 个元素。

根据公式(7)可计算样本中各元素在样本域内的权重为

$$\omega=\{12.5\%,9.38\%,12.5\%,15.63\%,12.5\%,6.25\%,3.13\%,9.38\%,15.63\%,3.13\%\} \quad (26)$$

进而由公式(6)可得样本的聚类中心值为

$$C_X=\sum_{i=1}^{n}(\omega_i\times x_i)=2.978\ 1 \quad (27)$$

为了对比说明该聚类中心值与样本均值的区别，计算得到样本均值为

$$M_X=4.17 \quad (28)$$

将样本数据、聚类中心值及样本均值的分布综合到一起，如图 5 所示。

图 5 中选取了纵轴上的 1 这条直线，将样本数据分布在该条直线上，易于观察，同样也可以选取其他任意值。从图中可见，样本数据在[1,5]范围内分布比较密集，样本数据也居多；在该范围外的样本相对比较分散，样本间隔较大。根据公式计算得到的聚类中心值在[1,5]范围内，靠近样本密集的区域，符合聚类的特性。相对地，样本均值更靠近样本域的中心，它反映的是样本的整体“重心”，并没有体现出样本的聚集特点。因此，采用聚类算法做小样本数据评估更能反映样本的分布趋势。

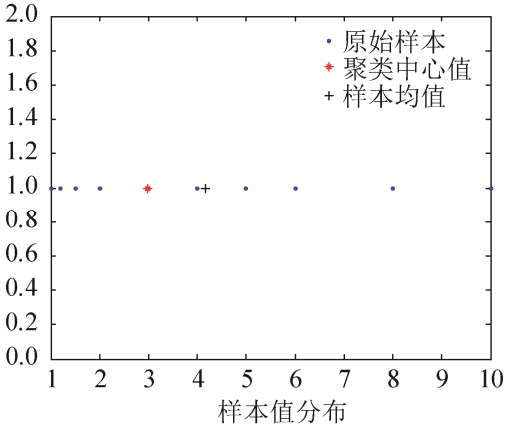


图 5 数据分布图

Fig. 5 Data distribution map

算例 2

选定同样样本量为 10 的样本数据 $X=\{1,5,3,2,1.2,6,8,4,1.5,10\}$。

根据式(8)和式(27)计算的聚类中心值，可得聚类云模型的期望值为

$$Ex=2.978\ 1 \quad (29)$$

此外，根据定义 4，可计算样本 X 的类一阶绝对中心矩和类二阶中心矩分别为

$$F=\frac{1}{n}\sum_{i=1}^{n}|x_i-C_X|=2.434\ 4 \quad (30)$$

$$S=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-C_X)^2=11.000\ 7 \quad (31)$$

由定理 1 和定理 2，可得聚类云模型的熵和超熵分别为

$$\hat{En}=\sqrt{\frac{\pi}{2}}\times F=3.051\ 0 \quad (32)$$

$$\hat{He}=\sqrt{S-\hat{En}^2}=1.300\ 7 \quad (33)$$

至此，聚类云模型的 3 个参数均已计算得到，模型建立完成。根据该模型生成 N=1 000 个云滴数据，如图 6 所示。

从图 6 可见，所建立的聚类云模型的波峰(可信度最大值)靠近原始样本分布密集的区域，即原始样本中样本密集的区域可信度大，而远离该区域的样本，分布稀疏，可信度降低。该云模型能够准确的反映样本的可信度分布趋势。

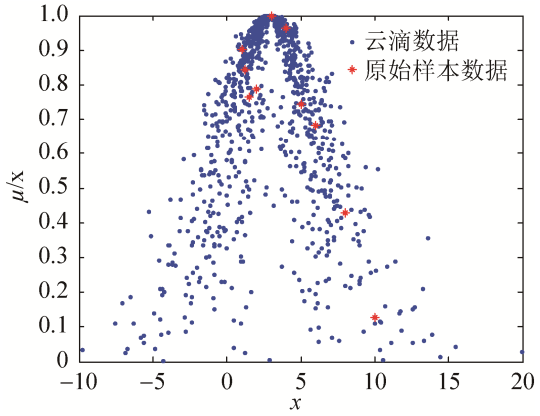


图 6 聚类云模型及原始样本分布
Fig. 6 Clustering cloud model and raw sample distribution

聚类云模型产生的 1 000 个云滴数据，相当于对原始小样本数据进行了扩充，增加了原始样本的信息量。为了进行对比说明，本文也采用 Bootstrap 方法[8]对该小样本数据进行了扩充，同样是生成了 1 000 个数据，其分布如图 7 所示。

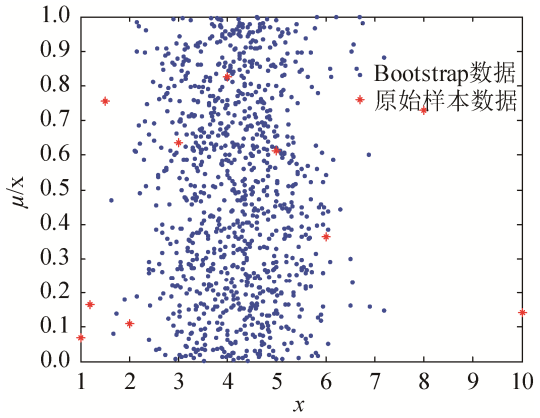


图 7 Bootstrap 数据及原始样本数据分布
Fig. 7 Bootstrap data and raw sample distribution

在图 7 中，横轴代表样本数据的值，纵轴是一个随机的可信度。由于 Bootstrap 重抽样获得的只是和原始样本相同的数据值，若绘制到一条直线上将无法辨识，因此为每个数据添加了一个随机纵坐标值，将生成的所有数据分散开来，便于观察。所以该图中只关注横坐标的分布即可。

从图 7 可见，扩充生成的数据相对均匀的分布在数据[2,6]范围内，基本是以原始样本的均值为中心进行均匀分布的。原始样本在数据[1,3]范围内的密集度比较大，而通过 Bootstrap 方法生成的数据并没有反映出该特性，通过该方法获得的数据只是抽样概率的一种体现。对比图 6 和图 7 的结果，本文设计的聚类云模型生成的数据更能代表原始样本的分布规律，在此数据基础上进行可信度评估也将更合理。

算例 3

选择小样本数据 $X=\{1,5,3,2,1.2,6,8,4,1.5,10\}$，计算各样本元素的可信度。

在算例 2 的基础上已经建立了聚类云模型，并生成了 1 000 个云滴数据。在此利用这些扩充后的数据计算原始样本的可信度。

选择 $\delta=0.5$，为每个原始样本数据选择一个 δ 邻域，统计处于该邻域内的云滴的数目 k，如表 1 所示。

表 1 原始样本邻域内的云滴数目
Tab. 1 Number of clouds in the neighborhood of the original sample

X	1	5	3	2	1.2
k	64	67	114	104	61
X	6	8	4	1.5	10
k	32	17	66	65	4

由表 1 可见，k 值均大于 3，满足步骤 8 的要求，可以计算可信度。

将每一个 δ 邻域内的云滴数据的确定度作为一个样本量为 k 值的样本，如第一个样本 δ 邻域内的云滴数为 64，将这些云滴数据的确定度值(纵坐标值)作为一个样本量为 64 的样本，利用定义 1~3 的内容，计算各新样本的聚类中心值，作为原始样本各数据的可信度值，如表 2 所示。

表 2 可信度评估结果
Tab. 2 Credibility assessment results

X	可信度/%	X	可信度/%
1	76.33	6	78.03
5	76.41	8	72.28
3	83.34	4	74.20
2	83.56	1.5	76.35
1.2	82.08	10	70.46

由表 2 可见，值为 2 的样本可信度最高，它靠

近样本域的聚类中心；值为 10 的样本可信度最低，它处于样本域的边缘，离样本域的聚类中心也最远。该可信度结果与原始样本的分布趋势相吻合，证明了该结果的合理性和正确性。

5 结论

针对小样本数据信息量少且缺乏参考标准的系统可信度评估问题，提出了一种聚类算法和云模型理论相结合的聚类云模型可信度评估方法。利用聚类算法计算小样本数据的聚类中心，将该中心值作为云模型的期望值，再根据小样本数据的一阶绝对中心矩和二阶中心矩计算云模型的输入参数，进而建立聚类云模型。利用建立的聚类云模型扩充小样本数据。通过聚类云模型中云滴的确定度信息，可以计算获得原始小样本数据的可信度值。通过仿真算例分析，聚类云模型生成的云滴数据的分布趋势与原始小样本的分布趋势相吻合，因此该可信度计算结果是合理和有效的。

参考文献：

[1] 张文一. 大规模光伏电站发电容量可信度研究[D]. 哈尔滨: 哈尔滨工程大学, 2017.
Zhang Wenyi. Research on Capacity Credit of Large-scale Photovoltaic Station [D]. Harbin: Harbin Engineering University, 2017.

[2] 方浩, 李艾华, 潘玉龙, 等. 基于自学习框架的红外场景仿真效果评价[J]. 系统工程与电子技术, 2019(2): 266-272.
Fang Hao, Li Aihua, Pan Yulong, et al. Evaluation for infrared scene simulation based on self-learning framework[J]. Systems Engineering and Electronics, 2019(2): 266-272.

[3] 薛俊杰, 周军华, 施国强, 等. 面向高可信度仿真的海量 CAD 模型实时绘制[J]. 系统仿真学报, 2018, 30(12): 4659-4667.
Xue Junjie, Zhou Junhua, Shi Guoqiang, et al. Real-time Rendering of Massive CAD Models for High-credibility Simulation[J]. Journal of System Simulation, 2018, 30(12): 4659-4667.

[4] 贾志新, 张宏斌, 郁安民. 利用神经网络扩充数控机床可靠性数据[J]. 吉林大学学报(工学版), 2011, 41(2): 403-407.
Jia Zhixin, Zhang Hongbin, Xi Anmin. Expanding reliability data of NC machine tool based on neural network[J]. Journal of Jilin University (Engineering and Technology Edition), 2011, 41(2): 403-407.

[5] Wei Li, Zhangqi Hu, Yulin Yang, et al. Credibility in evaluating on-line wear-depth detection of self-lubricating spherical plain bearings[J]. Advances in Mechanical Engineering (S1687-8140), 2016, 8(9): 1-12.

[6] 戴志辉, 李芷筠, 焦彦军, 等. 基于 BP 神经网络的小样本失效数据下继电保护可靠性评估[J]. 电力自动化设备, 2014, 34(11): 129-134.
Dai Zhihui, Li Zhijun, Jiao Yanjun, et al. Reliability assessment based on BP neural network for relay protection system with a few failure data samples[J]. Electric Power Automation Equipment, 2014, 34(11): 129-134.

[7] 王硕. 基于粒子群优化最小二乘支持向量机的建筑工程造价预测研究[D]. 青岛: 青岛理工大学, 2017.
Wang Shuo. Study of Prediction of Construction Cost based on Particle Swarm Optimization Least Squares Support Vector Machines[D]. Qingdao: Qingdao University of Technology, 2017.

[8] Wang Jianmin, Wang Jinbo, Zhang Tao, et al. Probability estimation based on grey system theory for simulation evaluation[J]. Journal of Systems Engineering and Electronics (S1004-4132), 2016, 27(4): 871-877.

[9] 王学姚. 基于 RBF 神经网络及非线性约束的数控机床 Bayes 可靠性评估研究[D]. 秦皇岛: 燕山大学, 2017.
Wang Xueyao. Research on Bayesian Reliability Assessment for NC Machine Tools based on RBF Neural Network and Nonlinear Constraint[D]. Qinhuangdao: Yanshan University, 2017.

[10] 朱孟兵. 基于小样本数据的数控机床 MTBF 评估[D]. 沈阳: 东北大学, 2012.
Zhu Mengbing. MTBF Evaluation for CNC System based on Small Sample Data[D]. Shenyang: Northeastern University, 2012.

[11] 张鑫鹏, 黄勇, 阎杰. 基于随机加权法的小子样数据评估研究[J]. 计算机仿真, 2010, 27(9): 9-11, 19.
Zhang Xinpeng, Huang Yong, Yan Jie. Study of Small Sample Test Data Estimation base on Bayes Bootstrap Method[J]. Computer Simulation, 2010, 27(9): 9-11, 19.

[12] Dai Zhihui, Wang Zengping, Jiao Yanjun. Bayes Monte-Carlo Assessment Method of Protection Systems Reliability Based on Small Failure Sample Data[J]. IEEE Transactions on Power Delivery (S0885-8977), 2014, 29(4): 1841-1848.

https://dc-china-simulation.researchcommons.org/journal/vol31/iss7/3
DOI: 10.16182/j.issn1004731x.joss.19-0202

[13] 王超. 虚实结合的测试性试验与综合评估技术[D]. 长沙: 国防科学技术大学, 2014.
Wang Chao. Testability Test and Integrated Evaluation Technology with Virtual-Physical Test[D]. Changsha: National University of Defense Technology, 2014.

[14] 胡利平, 刘锦帆, 王洪叶, 等. 基于模糊综合评判的车辆目标 SAR 仿真图像评估方法[J]. 系统工程与电子技术, 2019(3): 534-540.
Hu Liping, Liu Jinfan, Wang Hongye, et al. Vehicle SAR simulation images validation method based on fuzzy comprehensive evaluation [J]. Systems Engineering and Electronics, 2019(3): 534-540.

[15] 李德毅, 杜鹢. 不确定性人工智能 [M]. (2 版)北京: 国防工业出版社, 2014.
Li Deyi, Du yi. Artificial Intelligence with Uncertainty [M]. (2ed.)Beijing: National Defense Industry Press, 2014.

[16] 王玉珏, 杨继坤, 卢道伟, 等. 基于云模型的复杂仿真系统可信度评估方法[J]. 舰船电子工程, 2014, 34(9): 86-91.
Wang Yujue, Yang Jikun, Lu Daowei, et al. Method for Complex Simulation System Credibility Evaluation Based on Cloudy Model[J]. Ship Electronic Engineering, 2014, 34(9): 86-91.

[17] 郑宇, 方洋旺, 魏贤智, 等. 基于云模型的仿真可信度评估方法[J]. 计算机应用, 2018, 38(6): 1535-1541.
Zheng Yu, Fang Yangwang, Wei Xianzhi, et al. Evaluation method for simulation credibility based on cloud model[J]. Journal of Computer Applications, 2018, 38(6): 1535-1541.

[18] 蒋君妍. 面向高维数据的聚类算法改进研究[D]. 南京: 南京邮电大学, 2018.
Jiang Junyan. Improvement Research of Clustering Algorithm based on High-dimensional Data[D]. Nanjing: Nanjing University of Posts and Telecommunications, 2018.

[19] 何宇凡. 基于三支决策理论的时间序列聚类算法研究[D]. 南昌: 南昌大学, 2018.
He Yufan. Research of time series clustering algorithm based on three-way decisions theory [D]. Nanchang: Nanchang University, 2018.

[20] 刘磊. 基于 Spark 平台的大数据聚类算法研究及其应用[D]. 南京: 南京邮电大学, 2018.
Liu Lei. Research and Application of Big Data Clustering Algorithm Based on Spark Platform [D]. Nanjing: Nanjing University of Posts and Telecommunications, 2018.

[21] Li Deyi, Shi X M, Gupta M M. Soft Inference Mechanism Based on Cloud Models[C]. Proceedings of the 1st International Workshop on Logic Programming and Soft Computing: Theory and Applications. Reach Studies Press: Bonn, Germany, 1996.

[22] 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.
Liu Changyu, Feng Mang, Dai Xiaojun, et al. A New Algorithm of Backward Cloud[J]. Journal of System Simulation, 2004, 16(11): 2417-2420.

[23] 同济大学应用数学系. 高等数学 [M]. (五版)北京: 高等教育出版社, 2006.
Department of Applied Mathematics, Tongji University. Advanced Mathematics [M]. (5ed.)Beijing: Higher Education Press, 2006.

[24] 张守玉, 封伟书. 基于 Bootstrap 方法的正态分布样本数据生成研究[J]. 装备指挥技术学院学报, 2009, 20(2): 97-100.
Zhang Shouyu, Feng Weishu. Study of Sampled Data Creation for Norm Distribution on Bootstrap Method [J]. Journal of the Academy of Equipment Command & Technology, 2009, 20(2): 97-100.