

基于组合赋权的正态云模型形状相似性度量方法

官加俊, 张延飞*, 刘星星, 丁木华

东华理工大学理学院, 江西 南昌

收稿日期: 2024年7月13日; 录用日期: 2024年8月7日; 发布日期: 2024年8月15日

摘要

针对云模型相似性度量方法中存在的区分度不高、结果不稳定等问题, 提出一种基于组合赋权的正态云模型形状相似性度量方法。首先基于期望曲线的位置关系, 用正态模糊数的贴近度来表征云模型的形状相似度; 然后考虑云滴的离散程度, 在云滴方差基础上提出基于熵与云滴方差的形状相似度; 最后考虑云模型的三个数字特征, 基于偏好系数, 采用组合赋权, 将两种形状相似度进行组合度量云模型相似度。通过仿真实验及时间序列分类实验表明, 该方法是有用的, 并具有较好的区分度和稳定性。

关键词

云模型, 形状相似度, 期望曲线, 模糊贴近度, 组合赋权

Shape Similarity Measurement Method of Normal Cloud Model Based on Combination Weighting

Jiajun Guan, Yanfei Zhang*, Xingxing Liu, Muhua Ding

School of Science, East China University of Technology, Nanchang Jiangxi

Received: Jul. 13th, 2024; accepted: Aug. 7th, 2024; published: Aug. 15th, 2024

Abstract

To tackle the challenges of limited distinction and inconsistent outcomes in similarity measurement among cloud models, this paper proposed a method for measuring shape similarity of nor-

*通讯作者。

文章引用: 官加俊, 张延飞, 刘星星, 丁木华. 基于组合赋权的正态云模型形状相似性度量方法[J]. 应用数学进展, 2024, 13(8): 3803-3813. DOI: 10.12677/aam.2024.138362

mal cloud models based on combinatorial weighting. Firstly, the approximation degree of a normal fuzzy number is employed to characterize the shape similarity of the cloud model based on its positional relation with the expected curve. Then considering the dispersion degree of cloud droplets, the shape similarity based on entropy and cloud droplet variance is proposed on the basis of cloud droplet variance. Finally, considering the three digital features of the cloud model, based on the preference coefficient, the combination weighting is used to combine the two shape similarities to measure the similarity of the cloud model. The simulation results show that the method is effective and has good discrimination and stability.

Keywords

Cloud Model, Shape Similarity, Expectation Curve, Fuzzy Proximity, Combination Weighting

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

不确定性事件与现象普遍存在于人类社会之中，其中对概念认知的不确定性问题最为显著。云模型是李德毅院所提出的一种定性概念与定量描述之间的不确定性转换模型[1]。该模型将不确定性概念的随机性与模糊性紧密联系，通过云模型的数字特征(期望，熵，超熵)描述概念的不确定性。云模型属于定性定量不确定性转换模型，目前已在多个领域得到广泛应用，如决策及系统评估[2] [3]、风险评估[4]、数据挖掘[5]等。而如何准确地度量云模型间的相似性是其在各应用领域中面临的最为重要的问题。

现有的云模型相似度测算方法研究主要从以下几个角度出发：基于云滴的距离测算、基于数字特征的向量测算、基于形状的重叠面积测算以及兼顾云滴距离和形状面积的综合测算。张勇等[6]通过利用正向云发生器算法生成云滴，随机选取一定数量云滴并计算所有组合的累计距离的均值作为相似度，但该方法受云滴数量和实验次数的参数设置影响，计算量大且结果不稳定；张光卫等[7]基于云模型的数字特征，将云模型的相似度测算转化为向量的夹角余弦值计算，但当某个数字特征较大时，容易削弱其它数字特征的作用，产生较大的误差；李海林等[8]提出基于正态云模型期望曲线和最大边界曲线的相似性度量方法，通过计算特征曲线相交时的公共部分面积衡量云模型间的相似程度，但前者忽略了超熵的影响而后者扩大了超熵的影响，导致结果存在偏差；龚艳冰等[9]提出基于组合模糊贴近度的相似性度量方法，将修正的特征曲线作为模糊数的隶属函数，通过计算模糊贴近度来衡量云模型的相似程度，但当熵较大时，会削弱超熵的影响，并且未考虑模糊数之间存在两个交点的情况；汪军[10]、徐聪[11]等分别提出基于云滴方差和基于期望曲线与最大含熵期望曲线的云模型形状相似度，但都削弱了熵对云模型形状的影响。为了降低相似度测算方法的时间运行代价，许昌林[12]、Xu [13]等分别提出基于 Hellinger 距离和 Wasserstein 距离的云模型相似度方法，但当度量处于特殊情况下的云模型时其结果与实际存在误差。

针对上述问题，借鉴文献[14]中的相互隶属度概念，对文献[9]中所提及的模糊贴近度进行改进，构建基于改进的模糊贴近度和基于熵与云滴方差的正态云模型形状相似度。引入组合赋权思想[15]，提出一种组合赋权的正态云模型相似性度量方法，并通过仿真实验和时间序列分类实验验证该方法的有效性及稳定性。

2. 云模型及模糊贴近度

定义 1 [16] 设 U 是一个定量论域, 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

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

则 x 在论域 U 上的分布称为云, 每一个 x 称为云滴。

定义 2 [16] 云模型用 $C(Ex, En, He)$ 整体表示一个概念。期望 Ex 表示云滴在论域空间上的期望, 即概念量化的最典型样本; 熵 En 表示定性概念的不确定性度量, 反映云滴的取值范围; 超熵 He 表示熵的不确定性度量, 即熵的熵, 如图 1 所示。

定义 3 [16] 若随机变量 x 满足: $x \sim N(Ex, En'^2)$, 其中 $En' \sim N(En, He^2)$ 且 $En \neq 0$, 则

$$y = \exp \left[-\frac{(x - Ex)^2}{2En^2} \right]$$

称为正态云模型的期望曲线。

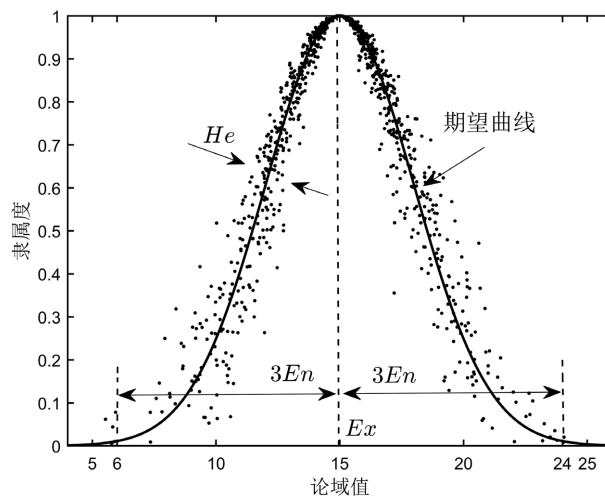


Figure 1. Normal cloud model $C(15, 3, 0.5)$ and expected curve diagram

图 1. 正态云模型 $C(15, 3, 0.5)$ 及期望曲线示意图

定义 4 [17] 设 $F(U)$ 为论域 U 的模糊幂集, 若映射

$$C: F(U) \times F(U) \rightarrow [0,1], (\tilde{A}, \tilde{B}) \mapsto C(\tilde{A}, \tilde{B}) \in [0,1]$$

满足:

- (1) $C(\tilde{A}, \tilde{A}) = 1, \forall \tilde{A} \in F(U)$
- (2) $C(\tilde{A}, \tilde{B}) = C(\tilde{B}, \tilde{A}), \forall \tilde{A}, \tilde{B} \in F(U)$
- (3) $\tilde{A} \subseteq \tilde{B} \subseteq \tilde{C} \Rightarrow C(\tilde{A}, \tilde{C}) \subseteq C(\tilde{A}, \tilde{B}) \wedge C(\tilde{B}, \tilde{C})$

称 $C(\tilde{A}, \tilde{B})$ 为 $\tilde{A}$ 与 $\tilde{B}$ 的贴近度。

3. 组合赋权的正态云模型相似性度量

3.1. 基于改进的模糊贴近度的正态云模型形状相似性度量

模糊贴近度概念是我国学者汪培庄教授首先提出来的, 用以描述模糊集之间的贴近程度。通过引入

相互隶属度概念[14], 定义如下模糊贴近度。

定义5 设论域 U 为实数域 R , $\tilde{A}$ 、 $\tilde{B}$ 为正态模糊数, 则有

$$C_1(\tilde{A}, \tilde{B}) = \frac{\int_{-\infty}^{+\infty} [\tilde{A}(x) \wedge \tilde{B}(x)] dx}{\int_{-\infty}^{+\infty} \tilde{A} dx}, \quad C_2(\tilde{B}, \tilde{A}) = \frac{\int_{-\infty}^{+\infty} [\tilde{A}(x) \wedge \tilde{B}(x)] dx}{\int_{-\infty}^{+\infty} \tilde{B} dx}$$

其中, $\int_{-\infty}^{+\infty} [\tilde{A}(x) \wedge \tilde{B}(x)] dx$ 表示两个模糊数的隶属函数相交重叠时公共部分与横轴围成的面积。假设模糊数的隶属函数为正态云模型的期望曲线, 那么云模型的相似度测算就可以转化为模糊数贴近度计算。

设正态模糊数 $\tilde{A}$ 与 $\tilde{B}$ 的隶属函数为

$$\mu_{\tilde{A}}(x) = \exp\left[-\left(\frac{x - Ex_1}{\sqrt{2}En_1}\right)^2\right], \quad \mu_{\tilde{B}}(x) = \exp\left[-\left(\frac{x - Ex_2}{\sqrt{2}En_2}\right)^2\right]$$

令 $\left|\frac{x - Ex_1}{En_1}\right| = \left|\frac{x - Ex_2}{En_2}\right|$, 得到两个正态模糊数交点横坐标为

$$x_1 = \frac{Ex_1En_2 + Ex_2En_1}{En_1 + En_2}, \quad x_2 = \frac{Ex_1En_2 - Ex_2En_1}{En_2 - En_1} \quad (1)$$

(1) 存在单个交点

若 $Ex_1 < Ex_2$, $En_1 < En_2$, 两个正态模糊数的位置如图 2(a) 所示。依据云模型“3En”规则, 当交点 x_1 或交点 x_2 落在区间 $[Ex_2 - 3En_2, Ex_1 + 3En_1]$, 则存在单个交点。而交点 x_2 不在 Ex_1 与 Ex_2 之间, 故只需考虑交点 x_1 。根据上述定义的正态模糊数贴近度, 有

$$C_1(\tilde{A}, \tilde{B}) = \frac{\int_{-\infty}^{x_1} \exp\left[-\left(\frac{x - Ex_2}{\sqrt{2}En_2}\right)^2\right] dx + \int_{x_1}^{+\infty} \exp\left[-\left(\frac{x - Ex_1}{\sqrt{2}En_1}\right)^2\right] dx}{\int_{-\infty}^{+\infty} \exp\left[-\left(\frac{x - Ex_1}{\sqrt{2}En_1}\right)^2\right] dx} \quad (2)$$

$$C_2(\tilde{B}, \tilde{A}) = \frac{\int_{-\infty}^{x_1} \exp\left[-\left(\frac{x - Ex_2}{\sqrt{2}En_2}\right)^2\right] dx + \int_{x_1}^{+\infty} \exp\left[-\left(\frac{x - Ex_1}{\sqrt{2}En_1}\right)^2\right] dx}{\int_{-\infty}^{+\infty} \exp\left[-\left(\frac{x - Ex_2}{\sqrt{2}En_2}\right)^2\right] dx} \quad (3)$$

为简化计算, 令 $t = \frac{x - Ex_2}{\sqrt{2}En_2}$, $dx = \sqrt{2}En_2 dt$, 得到

$$\int_{-\infty}^{x_1} \exp\left[-\left(\frac{x - Ex_2}{\sqrt{2}En_2}\right)^2\right] dx = \sqrt{2}En_2 \int_{-\infty}^{-t_1} \exp(-t^2) dt$$

同理, 令 $t = \frac{x - Ex_1}{\sqrt{2}En_1}$, $dx = \sqrt{2}En_1 dt$, 得到

$$\int_{x_1}^{+\infty} \exp\left[-\left(\frac{x-Ex_1}{\sqrt{2En_1}}\right)^2\right] dx = \sqrt{2En_1} \int_{t_1}^{+\infty} \exp(-t^2) dt$$

$$\int_{-\infty}^{+\infty} \exp\left[-\left(\frac{x-Ex_1}{\sqrt{2En_1}}\right)^2\right] dx = \sqrt{2En_1} \int_{-\infty}^{+\infty} \exp(-t^2) dt$$

将上式代入式(2), 得到

$$C_1(\tilde{A}, \tilde{B}) = \frac{(En_1 + En_2) \cdot \int_{t_1}^{+\infty} \exp(-t^2) dt}{\sqrt{\pi En_1}} = \frac{(En_1 + En_2)}{En_1} [1 - \Phi(\sqrt{2}t_1)] \quad (4)$$

其中, $t_1 = \frac{Ex_2 - Ex_1}{\sqrt{2(En_1 + En_2)}}$, $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt$ 。同理得到

$$C_2(\tilde{B}, \tilde{A}) = \frac{(En_1 + En_2) \cdot \int_{t_1}^{+\infty} \exp(-t^2) dt}{\sqrt{\pi En_2}} = \frac{(En_1 + En_2)}{En_2} [1 - \Phi(\sqrt{2}t_1)] \quad (5)$$

若 $Ex_1 = Ex_2$, $En_1 \neq En_2$, 两个正态模糊数的位置如图 2(b)所示。正态模糊数贴近度为

$$C_1(\tilde{A}, \tilde{B}) = \begin{cases} 1 & En_1 < En_2 \\ \frac{En_2}{En_1} & En_1 > En_2 \end{cases}$$

$$C_2(\tilde{B}, \tilde{A}) = \begin{cases} \frac{En_1}{En_2} & En_1 < En_2 \\ 1 & En_1 > En_2 \end{cases} \quad (6)$$

(2) 存在两个交点

若 $Ex_1 < Ex_2$, $En_1 < En_2$, 两个正态模糊数的位置如图 2(c)所示。当两个交点均落在区间 $[Ex_2 - 3En_2, Ex_1 + 3En_1]$, 则存在两个交点, 且交点 $x_1 > x_2$ 。正态模糊数贴近度为

$$C_1(\tilde{A}, \tilde{B}) = \frac{\int_{-\infty}^{x_2} \exp\left[-\left(\frac{x-Ex_1}{\sqrt{2En_1}}\right)^2\right] dx + \int_{x_2}^{x_1} \exp\left[-\left(\frac{x-Ex_2}{\sqrt{2En_2}}\right)^2\right] dx + \int_{x_1}^{+\infty} \exp\left[-\left(\frac{x-Ex_1}{\sqrt{2En_1}}\right)^2\right] dx}{\int_{-\infty}^{+\infty} \exp\left[-\left(\frac{x-Ex_1}{\sqrt{2En_1}}\right)^2\right] dx} \quad (7)$$

$$C_2(\tilde{B}, \tilde{A}) = \frac{\int_{-\infty}^{x_2} \exp\left[-\left(\frac{x-Ex_1}{\sqrt{2En_1}}\right)^2\right] dx + \int_{x_2}^{x_1} \exp\left[-\left(\frac{x-Ex_2}{\sqrt{2En_2}}\right)^2\right] dx + \int_{x_1}^{+\infty} \exp\left[-\left(\frac{x-Ex_1}{\sqrt{2En_1}}\right)^2\right] dx}{\int_{-\infty}^{+\infty} \exp\left[-\left(\frac{x-Ex_2}{\sqrt{2En_2}}\right)^2\right] dx} \quad (8)$$

同(1)的换元方式, 有

$$\int_{-\infty}^{x_2} \exp\left[-\left(\frac{x-Ex_1}{\sqrt{2En_1}}\right)^2\right] dx = \sqrt{2En_1} \int_{-\infty}^{t_2} \exp(-t^2) dt$$

$$\int_{x_2}^{x_1} \exp\left[-\left(\frac{x-Ex_2}{\sqrt{2En_2}}\right)^2\right] dx = \sqrt{2En_2} \int_{t_2}^{-t_1} \exp(-t^2) dt$$

$$\int_{x_1}^{+\infty} \exp\left[-\left(\frac{x-Ex_1}{\sqrt{2}En_1}\right)^2\right] dx = \sqrt{2}En_1 \int_{t_1}^{+\infty} \exp(-t^2) dt$$

将上式代入式(7), 得到

$$C_1(\tilde{A}, \tilde{B}) = \frac{En_1\Phi(\sqrt{2}t_2) + En_2[\Phi(-\sqrt{2}t_1 - \sqrt{2}t_2)] + En_1[1 - \Phi(\sqrt{2}t_1)]}{En_1} \quad (9)$$

其中, $t_1 = \frac{Ex_2 - Ex_1}{\sqrt{2}(En_1 + En_2)}$, $t_2 = \frac{Ex_1 - Ex_2}{\sqrt{2}(En_2 - En_1)}$, $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt$ 。同理得到

$$C_2(\tilde{B}, \tilde{A}) = \frac{En_1\Phi(\sqrt{2}t_1) + En_2[\Phi(-\sqrt{2}t_1 - \sqrt{2}t_2)] + En_1[1 - \Phi(\sqrt{2}t_1)]}{En_2} \quad (10)$$

定义6 设云模型 $C_1(Ex_1, En_1, He_1)$ 和 $C_2(Ex_2, En_2, He_2)$, $\tilde{A}$ 、 $\tilde{B}$ 为对应的正态模糊数, 则称

$$Sim_1(C_1, C_2) = \frac{1}{2}(C_1(\tilde{A}, \tilde{B}) + C_2(\tilde{B}, \tilde{A})) \quad (11)$$

为两个正态云模型基于模糊贴近度的形状相似度。

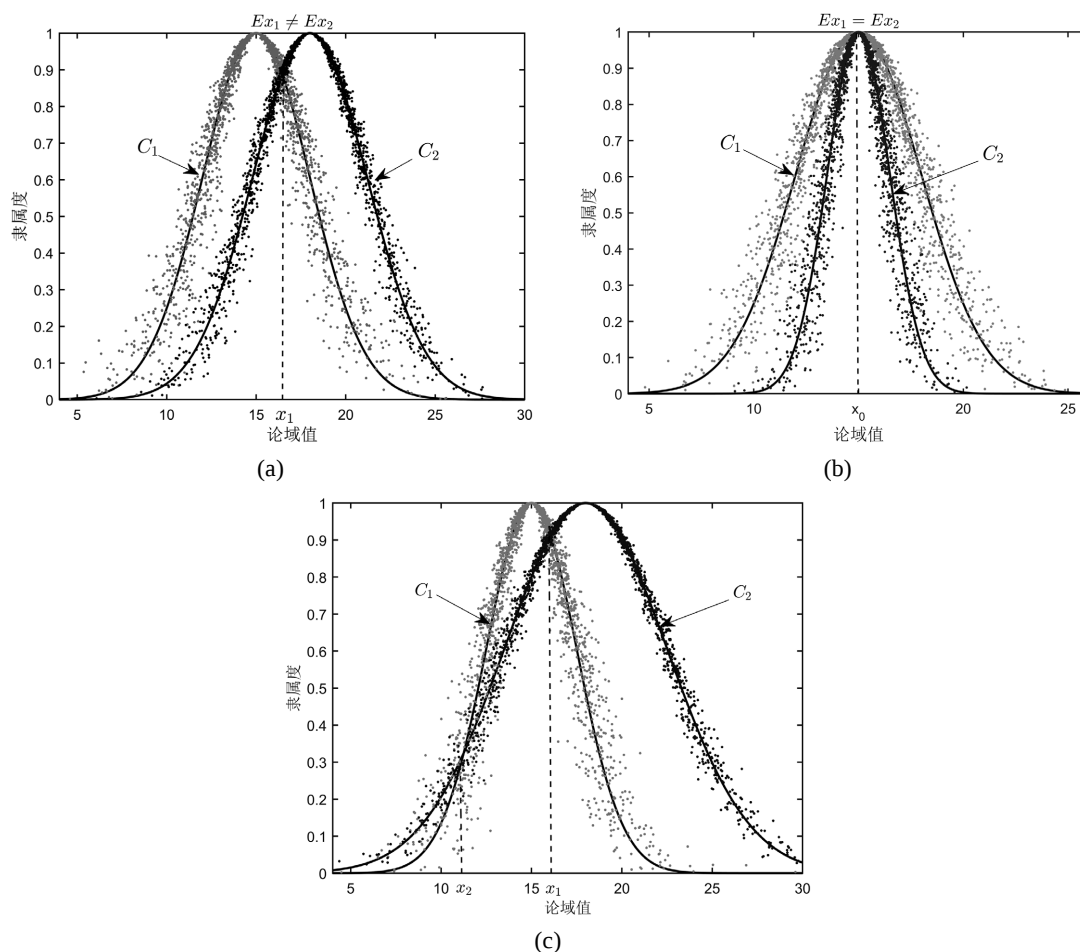


Figure 2. Different cases of intersection of normal fuzzy numbers
图2. 正态模糊数相交时的不同情况

3.2. 基于熵与云滴方差的正态云模型形状相似性度量

云模型的三个数字特征中, 熵 En 表示定性概念的不确定性度量, 可以描述云的宽度。正态云模型的云滴集合是一个期望为 Ex , 方差为 $En^2 + He^2$ 的随机变量, 其方差可以描述云滴整体的离散程度。当熵 En 相差较大时, 两朵云的位置就越趋近于“包含”关系。它们的差值越大, 其相似度就越低。故本文考虑在云的熵 En 下正态云模型云滴的整体分布特征, 采用云的熵 En 与方差的算术平方根 $\sqrt{En^2 + He^2}$ 的乘积能更加全面地描述云的形状, 凸显熵 En 对形状的影响。它们之比更能反映两个云模型之间的形状差异。

定义 7 设云模型 $C_1(Ex_1, En_1, He_1)$ 和 $C_2(Ex_2, En_2, He_2)$, 则称

$$Sim_2(C_1, C_2) = \frac{\min(En_1 \cdot \sqrt{En_1^2 + He_1^2}, En_2 \cdot \sqrt{En_2^2 + He_2^2})}{\max(En_1 \cdot \sqrt{En_1^2 + He_1^2}, En_2 \cdot \sqrt{En_2^2 + He_2^2})} \quad (12)$$

为两个正态云模型基于熵与云滴方差的形状相似度。

3.3. 基于组合赋权的正态云模型形状相似性度量

上述提出的两种云模型相似性度量方法基于不同的角度出发, 衡量云模型间的相似程度。基于改进的模糊贴近度的形状相似性度量方法仅考虑期望 Ex 与熵 En 的影响, 忽略了超熵 He 的作用。而基于熵与云滴方差的形状相似性度量方法并未利用到期望 Ex 。云模型是通过三个数字特征表征的, 度量云模型间的相似性需综合考虑它们的影响。本文参考文献[15]中组合赋权的方法, 引入偏好系数, 将以上两种正态云模型形状相似性度量方法进行组合, 得到一种组合赋权的正态云模型形状相似性度量方法(Combination Weighting based Cloud Model, CWCM)。

$$Sim(C_1, C_2) = \alpha \cdot Sim_1(C_1, C_2) + \beta \cdot Sim_2(C_1, C_2) \quad (13)$$

其中, α 、 β (≥ 0) 为两种相似度的偏好系数, 满足:

$$\begin{cases} d(Sim_1(C_1, C_2), Sim_2(C_1, C_2)) = \sqrt{(Sim_1(C_1, C_2) - Sim_2(C_1, C_2))^2} \\ d(Sim_1(C_1, C_2), Sim_2(C_1, C_2))^2 = (\alpha - \beta)^2 \\ \alpha + \beta = 1 \\ \alpha > \beta \end{cases} \quad (14)$$

具体算法如下:

算法 1 基于组合赋权的正态云模型形状相似性度量算法

输入: 两个云模型 $C_1(Ex_1, En_1, He_1)$ 和 $C_2(Ex_2, En_2, He_2)$;

输出: 两个云模型间的相似度 $Sim(C_1, C_2)$ 。

Step1: 设 $Ex_1 < Ex_2$, $En_1 < En_2$, 则交点 $x_1 > x_2$;

Step2: 若 $x_2 \leq \min(Ex_1 - 3En_1, Ex_2 - 3En_2)$ 且 $x_1 \geq \max(Ex_1 + 3En_1, Ex_2 + 3En_2)$, 则两个正态模糊数间不存在交点, $Sim(C_1, C_2) = 0$, 否则进行下一步;

Step3: 若 $x_2 \geq \max(Ex_1 - 3En_1, Ex_2 - 3En_2)$ 且 $x_1 \leq \min(Ex_1 + 3En_1, Ex_2 + 3En_2)$, 则两个正态模糊数间存在两个交点, 根据式(9)、式(10)和式(11)计算得到 $Sim_1(C_1, C_2)$;

Step4: 若 x_1 或 x_2 在区间 $[Ex_2 - 3En_2, Ex_1 + 3En_1]$ 内, 则两个正态模糊数间存在一个交点。若 $Ex_1 \neq Ex_2$, 根据(4)、式(5)和式(11)计算 $Sim_1(C_1, C_2)$, 否则根据式(6)和式(11)计算 $Sim_1(C_1, C_2)$;

Step5: 根据式(12)计算 $Sim_2(C_1, C_2)$;

Step6: 根据式(13)和式(14)计算 $Sim(C_1, C_2)$ 。

4. 实验与结果分析

4.1. 仿真实验

为了验证本文方法的有效性和区分度，选取文献[6]和文献[7]的云模型进行相似度计算，与现有方法进行比较，分析它们的相似度实验结果。

实例 1. 文献[6]给出 3 个云模型，分别为 $C_1(3,3.123,2.05)$, $C_2(2,3,1)$, $C_3(1.585,3.556,1.358)$ ，并与现有方法中的 ECM (expectation based cloud model) [8]、MCM (maximum boundary based cloud model) [8]、LICM (likeness comparing method based on cloud model) [7]和 CFSM (combined fuzzy similarity measure) [9] 进行比较，说明本文方法 CWCM 的有效性，结果如表 1 所示。

Table 1. Comparison of results from different similarity measures (Illustration of the validity of the methods in this paper)
表 1. 不同相似性度量方法的结果对比(本文方法的有效性说明)

	ECM	MCM	LICM	CFSM	CWCM
(C_1, C_2)	0.8728	0.7821	0.9717	0.8390	0.8506
(C_1, C_3)	0.8336	0.8983	0.9438	0.7961	0.8476
(C_2, C_3)	0.9138	0.8800	0.9850	0.9309	0.8620

从表 1 可知，通过本文方法 CWCM 计算得到的相似度中，云模型 C_2 与 C_3 相似度最大，其次是 C_1 与 C_2 ，最后是 C_1 与 C_3 。MCM 方法由于扩大了超熵对相似性的影响，导致云模型 C_1 与 C_3 的相似程度大于 C_2 与 C_3 ，其结果与其它四种方法存在偏差，从而说明本文方法是有效的。

实例 2. 文献[7]给出 4 个云模型，分别为 $C_1(1.5,0.62666,0.339)$, $C_2(4.6,0.60159,0.30862)$, $C_3(4.4,0.75199,0.27676)$, $C_4(1.6,0.60159,0.30862)$ 。同实例 1 中的对照方法进行比较，说明本文方法 CWCM 的区分度，结果如表 2 所示。

Table 2. Comparison of results from different similarity measures (Illustration of the discrimination of methods in this paper)
表 2. 不同相似性度量方法的结果对比(本文方法的区分度说明)

	ECM	MCM	LICM	CFSM	CWCM
(C_1, C_2)	0.01	0.33	0.96	0.02	0.04
(C_1, C_3)	0.04	0.37	0.97	0.04	0.09
(C_1, C_4)	0.94	0.96	0.99	0.92	0.93
(C_2, C_3)	0.86	0.95	0.99	0.85	0.82
(C_2, C_4)	0.01	0.38	0.97	0.02	0.02
(C_3, C_4)	0.04	0.37	0.98	0.04	0.10

从表 2 可知, 四个云模型中, 云模型 C_1 与 C_4 最相似, 其次是 C_2 与 C_3 , 其结果与其它四种云模型相似性度量方法一致。其中, (C_1, C_4) 与 (C_2, C_3) 相似度相差 0.11。其余方法中, 区分度最大的 ECM 方法计算得到的相似度仅相差 0.08, 最小的 LICM 方法区分度为 0。因此, 本文所提出的方法具有更好的区分度。

4.2. 时间序列分类实验

时间序列是将数据按时间顺序排列的具有高维特征的数值序列, 广泛运用于金融分析、经济预测、气象研究及数据挖掘等领域。时间序列的聚类与分类是数据挖掘中的重要内容, 而相似性度量方法是决定聚类与分类效果的关键因素。这里利用时间序列数据分类实验验证本文方法的稳定性。实验选取 UCI 数据库中的 Synthetic control chart 时间序列数据集, 该数据集共有 600 个序列样本, 样本长度为 60, 包含 6 类不同特征的数据, 每类由 100 个样本构成。在测试集与训练集的选取上, 将每类数据等份划分为 10 份, 每次实验选取其中的 1 份作为测试集, 剩下的 590 个序列样本作为训练集[18]。

实验采用最近邻(K-Nearest Neighbor, KNN)算法进行时间序列分类。首先对每个序列样本数据进行云模型表示预处理, 即通过基于云 X 信息的逆向云发生器算法[19], 将所有数据样本转化为用数字特征表征的云模型, 然后计算每类测试集与训练集的相似度矩阵, 最后依据相似度矩阵得到分类结果。根据各分类结果, 可以计算得到各分类正确率及其均值。各云模型相似性度量方法的分类正确率如图 3 所示。其中, 选取 ECM、MCM、LICM 和 CFSM 方法作为实验对照。由于 KNN 算法中的参数 K 会影响分类效果, 故又采用不同的参数 K 进行分类实验, 分析本文方法 CWCM 分类正确率的变化趋势, 结果如图 4 所示。

由图 3 可知, 本文所提出的相似性度量方法的分类正确率与其余 4 种方法中的 MCM 方法一致, 与最高的 ECM 方法仅略低 0.01, 而 LICM 方法仅在第 1 类的分类实验中效果较好, 其它分类实验的效果并不稳定, CFSM 方法的分类正确率是最低的。由图 4 可知, 对于不同的近邻数 K , 第 1 类与第 2 类的分类正确率并无影响, 而其余各类的分类正确率均有不同程度的变化。其中, 第 3、4 类的分类正确率呈现波动趋势, 而第 5、6 类整体上呈现下降趋势。结果表明, 本文方法相比于 LICM 和 CFSM 方法在分类正确率上有更好的提升, 且通过实验验证了本文方法在时间序列分类中有不错的效果, 而总体上近邻数 K 对分类正确率的影响并不大。该实验进一步说明了本文方法的稳定性。

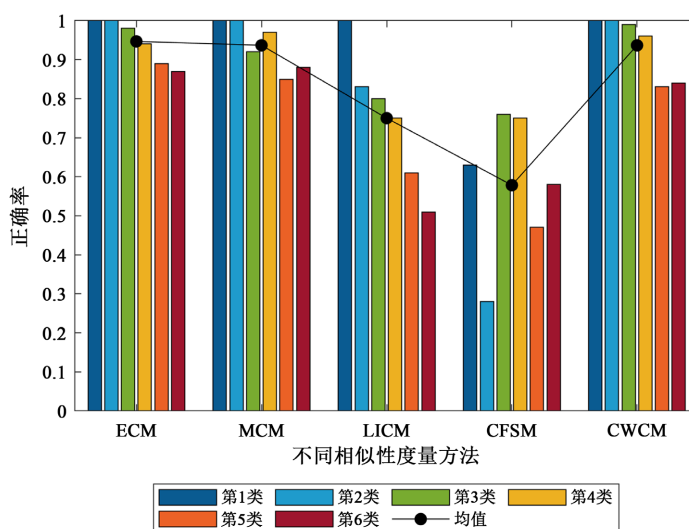


Figure 3. Comparison of classification accuracy results of different similarity measurement methods

图 3. 不同相似性度量方法分类正确率结果对比

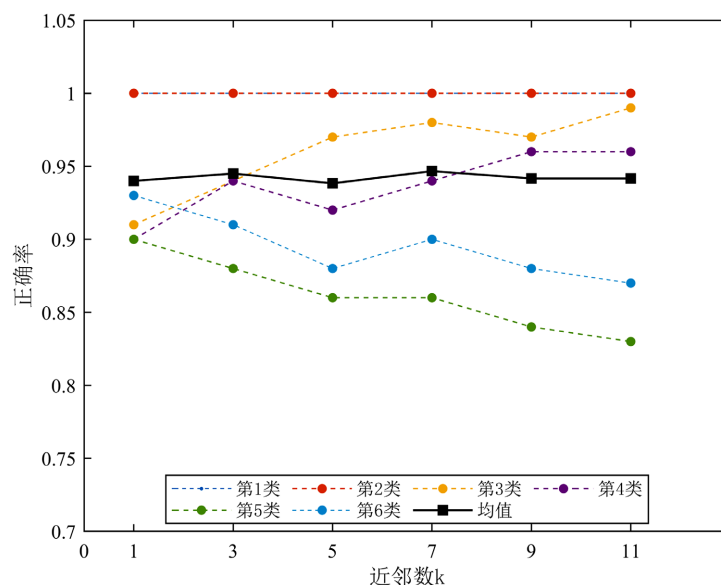


Figure 4. Comparison of classification accuracy results under different neighbor numbers K

图 4. 不同近邻数 K 下的分类正确率结果对比

5. 结论

正态云模型作为最为重要的一种云模型，已在多个领域得到广泛应用。而正态云模型相似性度量方法在云模型应用中极为重要，将直接影响结果的准确性。本文针对现有方法中存在的问题，提出了一种基于组合赋权的正态云模型形状相似性度量方法，并通过数值仿真和时间序列分类实验验证了该方法的有效性、区分度和稳定性，得到以下结论：

(1) 本文通过改进模糊贴近度，考虑云滴的离散程度与熵 En 对云模型形状的影响，并基于偏好系数，将两种不同的正态云模型形状相似度组合，提出了一种新的形状相似性度量方法。该方法综合利用了云模型的三个数字特征，能够较好地描述云模型间的差异。

(2) 在数值仿真实验中，该方法能够准确地衡量云模型间的相似程度，其结果具有更高的区分度，而在时间序列分类实验中，相比 LICM 和 CFSM 方法具有更好的稳定性，有效地解决了现有方法中区分度不高，结果不稳定等问题。

(3) 本文方法是对现有方法的进一步补充完善，如何将不确定性理论中的模糊贴近度概念运用到云模型相似性度量中，是今后研究的一个重要方向。

基金项目

国家自然科学基金项目(71961001)。

参考文献

- [1] 王国胤, 李德毅, 姚一豫, 等. 云模型与粒计算[M]. 北京: 科学出版社, 2012: 6-10.
- [2] Liu, Z., Wang, X., Wang, W., Wang, D. and Liu, P. (2022) An Integrated Topsis-Oreste-Based Decision-Making Framework for New Energy Investment Assessment with Cloud Model. *Computational and Applied Mathematics*, **41**, Article No. 42. <https://doi.org/10.1007/s40314-021-01751-9>
- [3] 陈萍, 张延飞, 刘星星, 等. 基于正态云组合赋权的城市生态环境质量综合评价[J]. 科技管理研究, 2023, 43(4): 68-74.
- [4] Yu, J.X., Chen, H.C., Wu, S.B. and Fan, H.Z. (2021) A Novel Risk Matrix Approach Based on Cloud Model for Risk

- Assessment under Uncertainty. *IEEE Access*, **9**, 27884-27896. <https://doi.org/10.1109/access.2021.3058392>
- [5] Gu, B., Zhang, T., Meng, H. and Zhang, J. (2021) Short-Term Forecasting and Uncertainty Analysis of Wind Power Based on Long Short-Term Memory, Cloud Model and Non-Parametric Kernel Density Estimation. *Renewable Energy*, **164**, 687-708. <https://doi.org/10.1016/j.renene.2020.09.087>
 - [6] 张勇, 赵东宁, 李德毅. 相似云及其度量分析方法[J]. 信息与控制, 2004, 33(2): 129-132.
 - [7] 张光卫, 李德毅, 李鹏, 等. 基于云模型的协同过滤推荐算法[J]. 软件学报, 2007, 18(10): 2403-2411.
 - [8] 李海林, 郭崇慧, 邱望仁. 正态云模型相似度计算方法[J]. 电子学报, 2011, 39(11): 2561-2567.
 - [9] 龚艳冰, 蒋亚东, 梁雪春. 基于模糊贴近度的正态云模型相似度度量[J]. 系统工程, 2015, 33(9): 133-137.
 - [10] 汪军, 朱建军, 刘小弟. 兼顾形状-距离的正态云模型综合相似度测算[J]. 系统工程理论与实践, 2017, 37(3): 742-751.
 - [11] 徐聪, 潘小东. 基于正态云相似度的语言型多属性群决策方法[J]. 计算机科学, 2019, 46(6): 218-223.
 - [12] 许昌林, 徐浩. 基于 Hellinger 距离的正态云相似性度量方法及应用研究[J]. 智能系统学报, 2023, 18(6): 1312-1321.
 - [13] Xu, C. and Yang, L. (2023) Research on Linguistic Multi-Attribute Decision Making Method for Normal Cloud Similarity. *Heliyon*, **9**, e20961. <https://doi.org/10.1016/j.heliyon.2023.e20961>
 - [14] 付凯, 夏靖波, 韦泽鲲, 等. 基于相互隶属度的云模型相似性度量方法[J]. 北京理工大学学报, 2018, 38(4): 405-411.
 - [15] 裴启涛, 李海波, 刘亚群, 等. 基于组合赋权的岩爆倾向性预测灰评估模型及应用[J]. 岩土力学, 2014, 35(z1): 49-56.
 - [16] 李德毅, 杜鹁. 不确定性人工智能[M]. 北京: 国防工业出版社, 2014: 143-177.
 - [17] Xuecheng, L. (1992) Entropy, Distance Measure and Similarity Measure of Fuzzy Sets and Their Relations. *Fuzzy Sets and Systems*, **52**, 305-318. [https://doi.org/10.1016/0165-0114\(92\)90239-z](https://doi.org/10.1016/0165-0114(92)90239-z)
 - [18] 代劲, 胡彪, 王国胤, 等. 分布轮廓与局部特征融合的云模型不确定性相似度量[J]. 电子与信息学报, 2022, 44(4): 1429-1439.
 - [19] 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.