

# 基于CD-型贴近度的云模型相似性度量方法

胡思雅<sup>1,2\*</sup>, 张延飞<sup>1,2#</sup>, 胡可<sup>2</sup>, 丁木华<sup>2</sup>

<sup>1</sup>铀资源探采与核遥感全国重点实验室(东华理工大学), 江西 南昌

<sup>2</sup>东华理工大学理学院, 江西 南昌

收稿日期: 2025年6月24日; 录用日期: 2025年7月18日; 发布日期: 2025年7月25日

## 摘要

在云模型的实际应用中, 云模型相似性度量是极为关键的环节。针对现有云模型相似性度量方法中存在的区分度欠佳、结果不稳定等问题, 本文以正态云的扩展模型三角云为研究对象, 在充分考虑期望曲线及最大边界曲线基础上, 将其视作三角模糊数, 通过计算三角模糊数的CD-型贴近度来度量云模型的相似性, 从而提出一种求两云模型相似度的CDTCM综合计算方法。由仿真实验可知, 提出的方法具有一定的区分度; 在Synthetic Control Chart Dataset数据集上的分类对比实验表明, 该方法的分类精度优于ECM、MCM、LICM、CFM、EMTCM等传统方法; 对于UCR数据库中的10个数据集表现出良好的分类效果, 验证了该方法具有一定的可行性及有效性。

## 关键词

三角云, CD-型贴近度, 期望曲线, 最大边界曲线, 相似性度量

# The Similarity Measurement Method of Cloud Models Based on CD-Type Closeness

Siya Hu<sup>1,2\*</sup>, Yanfei Zhang<sup>1,2#</sup>, Ke Hu<sup>2</sup>, Muhua Ding<sup>2</sup>

<sup>1</sup>State Key Laboratory of Uranium Resources Exploration, Mining and Nuclear Remote Sensing (East China University of Technology), Nanchang Jiangxi

<sup>2</sup>School of Science, East China University of Technology, Nanchang Jiangxi

Received: Jun. 24<sup>th</sup>, 2025; accepted: Jul. 18<sup>th</sup>, 2025; published: Jul. 25<sup>th</sup>, 2025

## Abstract

In the practical application of cloud models, similarity measurement is a critical component. Aiming

\*第一作者。

#通讯作者。

文章引用: 胡思雅, 张延飞, 胡可, 丁木华. 基于 CD-型贴近度的云模型相似性度量方法[J]. 应用数学进展, 2025, 14(7): 244-257. DOI: 10.12677/aam.2025.147361

at the issues of poor discrimination and unstable results in existing cloud model similarity measurement methods, this study focuses on the triangular cloud, an extended model of the normal cloud. By fully considering the expected curve and maximum boundary curve, the triangular cloud is treated as a triangular fuzzy number. The CD-type closeness degree of triangular fuzzy numbers is calculated to measure the similarity of cloud models, and a comprehensive calculation method named CDTCM for evaluating the similarity between two cloud models is proposed. Simulation experiments show that the proposed method has a certain degree of discrimination. Classification comparison experiments on the Synthetic Control Chart Dataset indicate that the classification accuracy of this method outperforms traditional methods, such as ECM, MCM, LICM, CFSM, and EMTCM, verifying that the method is stable and demonstrates certain feasibility and effectiveness.

## Keywords

Triangular Cloud, CD-Type Closeness, Expectation Curve, Maximum Boundary Curve, Similarity Measure

Copyright © 2025 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

云模型是李德毅院士于 20 世纪 90 年代初提出的一种研究定性概念量化模型[1]，现已广泛应用于人工智能[2]、数据挖掘[3]、数据处理[4]、推荐系统[5]和评价决策[6]等诸多领域。对不同的云模型进行相似度的测算是云模型应用过程的重要环节。相似性作为客观事物本质关联的直观反映，其度量方法构成了数据挖掘、推荐系统等领域的核心技术环节，且云模型相似性的计算精度直接影响到其在实际应用的效果。然而，当前现有云模型相似性度量方法普遍存在计算复杂、区分度不高、结果不稳定等问题。因此，研究一种更有效的相似性度量方法，不仅能够完善现有的云理论体系，提升云模型的运行效率，还能够对推动人工智能、数据挖掘、智能决策系统等领域发展有着重要的意义。

在云模型相似性度量领域，现有方法主要分为以下四类：一类是基于云滴度量方法。如，张勇[7]通过随机抽取云滴样本并计算所有组合的累计距离的均值作为相似度，但结果易受云滴个数和实验次数的影响，且计算复杂，不适合大规模数据；二类是基于数字特征的向量度量方法。如，张光卫等[8]将云模型的期望、熵、超熵等数字特征作为一组向量，将相似度测算转化为向量的夹角余弦值计算，但该方法未能充分考虑熵和超熵的影响，且在区分度方面存在局限；三类是基于形状的重叠面积度量方法。如，李海林等[9]通过求解两云模型期望曲线的交点及重叠面积进行度量。查翔等[10]提出一种云相似度概念跃升的间接计算相似性方法，通过截断熵获得综合云，计算综合云与原始云的曲线相交面积间接衡量相似性。汪军等[11]结合期望曲线相交面积和方差之比定义了云模型的距离相似性和形状相似性概念，构建一种云模型综合相似性度量方法。上述方法都涉及曲线交点的计算及其重叠面积的积分运算，时间复杂度较高；四类从模糊贴近度的角度来描述云模型相似性的度量方法，如，龚艳冰等[12]将两种模糊贴近度公式进行组合来度量正态云的相似性。黄琼桃等[13]基于 EW-型贴近度来度量三角云的相似性。此外，还有兼顾云滴距离和形状面积等多角度综合度量方法。如，Zhu 等[14]基于正态云的云滴数量分布和隶属度，提出一种双向加权的相似性度量方法。许昌林[15]、Xu [16]等分别提出基于 Hellinger 距离和 Wasserstein 距离的云模型相似度方法。以上方法在区分度、稳定性、时间代价大小方面各有优缺点，更好的云模型相似性度量方法有待提出。

针对上述问题, 本文以三角云模型作为研究对象, 将其期望曲线以及最大边界曲线视作三角模糊数 [13], 综合考虑云模型的位置和形状特征, 引入 CD-型模糊数贴近度公式, 提出一种基于 CD-型贴近度的云模型相似性度量方法(similarity measure method of CD type approximation formula based on triangular cloud model, CDTCM)。实验表明, 该方法能客观地度量云模型相似性, 较传统的 ECM、MCM、LICM、CFSM、EMTCM 等方法, 区分度更高, 并且在时间序列分类实验中能提高分类结果的精确度。

## 2. 理论基础

**定义 1 [17]** 设  $U$  为数值型定量论域,  $C$  为  $U$  上的语言值概念, 对于任意  $x \in U$ , 若  $x$  是概念  $C$  的单个随机映射结果, 且  $x$  对  $C$  的隶属度  $\mu_c(x) \in [0, 1]$  是具有稳定倾向的随机数值。即  $\forall x \in U, x \rightarrow \mu(x)$ , 则  $x$  在论域  $U$  上的分布称为云, 每个  $(x, \mu_c(x))$  称为一个云滴。

**定义 2 [17]** 云模型  $C(Ex, En, He)$  通过三个数字特征来描述一个定性概念:

- 1) 期望值  $Ex$  (expectation): 体现了概念外延的理想中心位置, 对应云图的几何重心, 即随机实现一次定性概念时所有  $x$  的平均值;
- 2) 熵  $En$  (entropy): 反映定性概念外延的随机波动范围, 即云滴(样本点)围绕期望值( $Ex$ )的离散程度, 同时也代表某一定性概念的接受范围;
- 3) 超熵  $He$  (hyper entropy): 二阶熵, 是熵( $En$ )的不确定性度量。  $He$  越小, 云滴的  $\mu(x)$  分布越集中, 概念边界越清晰; 反之则云图越“厚”, 概念边界越模糊。下图 1 表示了三角云模型以及其三个数字特征的含义。

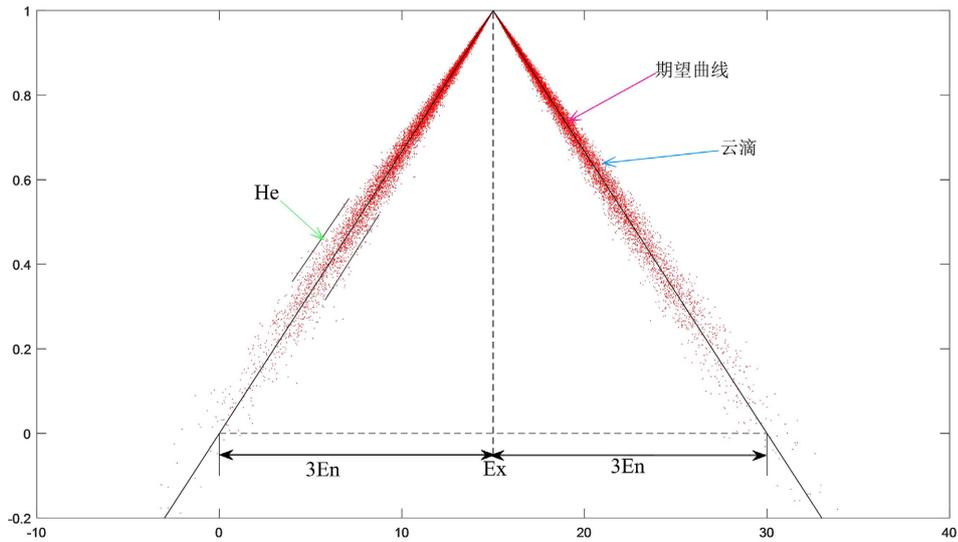


Figure 1. The schematic diagram of triangular cloud model (15, 5, 0.3)

图 1. 三角云模型(15, 5, 0.3)的示意图

**定义 3 [13]** 若云模型  $C(Ex, En, He)$  生成的云滴  $x$  满足:  $x \sim N(Ex, En'^2)$ , 其中  $En' \sim N(En, He^2)$ , 且  $x$  对定性概念  $C$  的隶属函数为:

$$\mu_c(x) = \begin{cases} \frac{x - (Ex - 3En)}{3En'}, & x < Ex \\ 1 - \frac{x - Ex}{3En'}, & x \geq Ex \end{cases}$$

则称  $x$  在论域  $U$  上的分布为三角云。

**定义 4 [13]** 若随机变量  $x$  满足:  $x \sim N(Ex, En^2)$  其中  $En' \sim N(En, He^2)$ , 且  $En \neq 0$ , 则称

$$y(x) = \begin{cases} \frac{x - (Ex - 3En)}{3En}, & x < Ex \\ 1 - \frac{x - Ex}{3En}, & x \geq Ex \end{cases}$$

为三角云的期望曲线, 如图 1 中线条所示。

**定义 5 [18]** 设  $R$  为实数集,  $u$  为模糊集, 若其满足正则性, 凸性上半连续性, 且支集是紧集, 则模糊集  $u: R \rightarrow [0,1]$  称为模糊数。模糊数空间是由全体模糊数构成的空间, 记为  $F_0$ 。

特别地, 对于模糊数  $u \in F_0$ , 若

$$u(x) = \begin{cases} \frac{x - r^l}{r^m - r^l}, & r^l \leq x \leq r^m \\ \frac{x - r^u}{r^m - r^u}, & r^m \leq x \leq r^u \\ 0, & \text{其他} \end{cases}$$

则称  $u$  为三角模糊数, 记为  $u = (r^l, r^m, r^u)$ 。

**定义 6 [19]** 对于  $\bar{a}, \underline{a} \in R$ , 且  $\underline{a} \leq \bar{a}$ , 则  $a = [\underline{a}, \bar{a}]$  称为区间数。区间的全体记为  $[R]$ 。对于  $a \in [R]$ , 存在公式:  $E(a) = \frac{\bar{a} + \underline{a}}{2}$ ,  $W(a) = \frac{\bar{a} - \underline{a}}{2}$ ,  $E(a)$  和  $W(a)$  分别是区间数的期望值和宽度。

**定义 7 [20]** 设  $u \in F_0$ , 对于给定的  $r \in [0,1]$ , 模糊数  $u$  的  $r$ -截集记为  $[u]^r = [\underline{u}(r), \bar{u}(r)]$ , 其中  $[u]^r = \{x | u(x) \geq r\}$ 。

**定义 8 [18]** 设  $u, v, w \in F_0$ , 若映射  $T: F_0 \times F_0 \rightarrow [0,1]; (u, v) \rightarrow N(u, v)$  满足:

- 1)  $T(u, v) = T(v, u)$ ;
- 2)  $T(u, v) = 1 \Leftrightarrow u = v$ ;
- 3)  $u \leq v \leq w \Rightarrow T(u, w) \leq T(u, v)$  且  $T(u, w) \leq T(v, w)$ 。

则称  $T(u, v)$  为模糊数  $u, v$  的贴近度。

### 3. CD-型贴近度公式

本文从模糊贴近度角度来描述云模型的相似性, 从 CD-型区间数距离出发, 构造了一种新的 CD-型模糊贴近度公式, 并基于该模糊贴近度公式来度量两三角云模型的相似性。

**引理 1 [18]** 对任意  $a, b \in [R]$ , 设函数  $N: [R] \times [R] \rightarrow [0,1]$  为:

$$N(a, b) = f(g(a, b))$$

其中函数  $f: [0, +\infty) \rightarrow (0, 1]$  严格单调递减且  $f(0) = 1, \lim_{x \rightarrow +\infty} f(x) = 0$ ; 距离函数

$$g: [R] \times [R] \rightarrow [0, +\infty)$$

满足如下条件:

- 1)  $g(a, b) = g(b, a)$ ;
- 2)  $g(a, b) = 0 \Leftrightarrow a = b$ ;
- 3)  $a \leq b \leq c \Rightarrow g(a, c) \geq g(a, b)$  且  $g(a, c) \geq g(b, c)$ 。

则  $N(a,b) = g(f(a,b))$  为区间数与的贴程度。

根据上述引理 1, 本文选取严格单调递减函数  $f(x) = \frac{1}{1+x^n}$ , (其中  $x \in [0, +\infty)$ , 且  $n \geq 1$ ) 显然该函数满足  $f(0) = 1, \lim_{x \rightarrow +\infty} f(x) = 0$ , 及文献[21]提出的如下式(1)的 CD-型区间数距离公式

$$g(a,b) = \left( 3|E(a) - E(b)|^2 + 2|W(a) - W(b)|^2 \right)^{\frac{1}{2}}, \forall a, b \in [\mathbf{R}], \tag{1}$$

可定义如式(2)所示的 CD-型区间数贴程度公式[21]

$$n_{CD}(a,b) = \frac{1}{1+(g(a,b))^n}, \tag{2}$$

受文献[18]的启发, 对 CD-型区间数贴程度进行积分, 构造 CD-型模糊数贴程度公式, 给出如下定理 1。

**定理 1** 对于  $\forall u, v \in F_0$ ,  $[u]^r, [v]^r$  为其  $r$ -截集, 设映射  $N_{CD}^2 : F_0 \times F_0 \rightarrow [0, 1]$  定义为:

$$N_{CD}^2(u, v) = \left( \int_0^1 \left( n_{CD}([u]^r, [v]^r) \right)^2 dr \right)^{\frac{1}{2}}, \tag{3}$$

则  $N_{CD}^2$  成为模糊数  $u$  和  $v$  的贴程度, 称为 CD-型贴程度。

证: 设  $\forall u, v, w \in F_0$ , 其中  $F_0$  是模糊数空间, 则显然有

$$0 \leq N_{CD}^2(u, v) = \left( \int_0^1 \left( n_{CD}([u]^r, [v]^r) \right)^2 dr \right)^{\frac{1}{2}} \leq 1,$$

这表明映射  $N_{CD}^2$  满足:  $F_0 \times F_0 \rightarrow [0, 1]$ 。

1)  $N(u, v) = N(v, u)$ :

$$N_{CD}^2(u, v) = \left( \int_0^1 \left( n_{CD}([u]^r, [v]^r) \right)^2 dr \right)^{\frac{1}{2}} = \left( \int_0^1 \left( n_{CD}([v]^r, [u]^r) \right)^2 dr \right)^{\frac{1}{2}} = N_{CD}^2(v, u).$$

2)  $N_{CD}^2(u, v) = 1 \Leftrightarrow u = v$ :

当  $u = v$  时, 对任意  $r \in [0, 1]$ , 均有  $\bar{u}(r) = \bar{v}(r)$  且  $\underline{u}(r) = \underline{v}(r)$ , 即  $[u]^r = [v]^r, r \in [0, 1]$ , 于是由式(1), (2)易知, 有  $n_{CD}([u]^r, [v]^r) = 1$ , 所以易推出

$$N_{CD}^2(u, v) = \left( \int_0^1 \left( n_{CD}([u]^r, [v]^r) \right)^2 dr \right)^{\frac{1}{2}} = 1.$$

下面只需证明当  $N_{CD}^2(u, v) = 1$  时, 有  $u = v$ 。

由  $N_{CD}^2(u, v) = 1$ , 有  $\left( \int_0^1 \left( n_{CD}([u]^r, [v]^r) \right)^2 dr \right)^{\frac{1}{2}} = 1$ , 即

$$\int_0^1 \left( n_{CD}([u]^r, [v]^r) \right)^2 dr = 1,$$

又由  $0 \leq n_{CD}([u]^r, [v]^r) \leq 1$  且  $\int_0^1 \left( n_{CD}([u]^r, [v]^r) \right)^2 dr = 1$ , 有

$$n_{CD}([u]^r, [v]^r) = 1.$$

对于  $r \in [0, 1]$  几乎处处成立, 从而  $[u]^r = [v]^r$  对于  $r \in [0, 1]$  几乎处处成立。即存在  $[0, 1]$  上的零测度集  $\Omega$ , 使得  $[u]^r = [v]^r$  关于  $r$  在  $[0, 1] \setminus \Omega$  上处处成立。

对于  $r \in \Omega$  且  $r \neq 0$  时, 由零测度集  $\Omega$  的性质可知, 存在非减数列  $r_k \in [0, 1] \setminus \Omega$  使得  $\lim_{k \rightarrow +\infty} r_k = r$ 。从而根据模糊数的表示定理[23], 有

$$[u]^r = \bigcap_{k=1}^{+\infty} [u]^{r_k} = \bigcap_{k=1}^{+\infty} [v]^{r_k} = [v]^r.$$

对  $r = 0$ , 可以取单调下降的  $r_k \in [0, 1] \setminus \Omega$  收敛于 0, 所以同样有

$$[u]^0 = \overline{\bigcap_{k=1}^{+\infty} [u]^{r_k}} = \overline{\bigcap_{k=1}^{+\infty} [v]^{r_k}} = [v]^0,$$

因此  $[u]^r = [v]^r$  对任何  $r \in [0, 1]$  成立, 即  $u = v$ 。

3)  $u \leq v \leq w \Rightarrow N(u, w) \leq N(u, v)$  且  $N(u, w) \leq N(v, w)$ :

设  $u \leq v \leq w$ , 则对任意  $r \in [0, 1]$ , 有

$$\underline{u}(r) \leq \underline{v}(r) \leq \underline{w}(r), \quad \bar{u}(r) \leq \bar{v}(r) \leq \bar{w}(r),$$

即对任意  $r \in [0, 1]$ , 有  $[u]^r \leq [v]^r \leq [w]^r$ , 由区间数贴近度性质[18]可知

$$n_{CD}([u]^r, [w]^r) \leq n_{CD}([u]^r, [v]^r), \quad n_{CD}([u]^r, [w]^r) \leq n_{CD}([v]^r, [w]^r),$$

从而有

$$\int_0^1 (n_{CD}([u]^r, [w]^r))^2 dr \leq \int_0^1 (n_{CD}([u]^r, [v]^r))^2 dr,$$

$$\int_0^1 (n_{CD}([u]^r, [w]^r))^2 dr \leq \int_0^1 (n_{CD}([v]^r, [w]^r))^2 dr,$$

即

$$N_{CD}^2(u, w) \leq N_{CD}^2(u, v), \quad N_{CD}^2(u, w) \leq N_{CD}^2(v, w).$$

于是由定义 8,  $N_{CD}(u, v)$  为模糊数  $u$  与  $v$  的贴近度。

#### 4. 基于 CD-型贴近度的三角云相似性度量

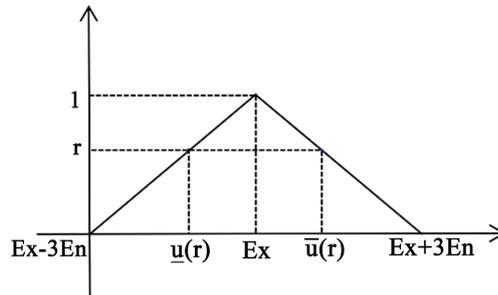
三角云模型  $C(Ex, En, He)$  作为一种拓展的云模型, 它以期望曲线作为骨架, 反映总体几何特征; 而根据三角云的雾化特性, 当  $0 < He < \frac{En}{3}$  时, 最大边界曲线可使 99.7% 的云滴落在其内部, 这表明最大边界曲线可反映云模型的局部特征。因此, 为融合云模型整体和局部特征, 本文先通过 CD-型贴近度来刻画期望曲线和最大边界曲线的贴近度, 再将其进行加权提出两三角云综合相似性度量方法。

##### 4.1. 三角云期望曲线的相似性度量

三角云模型的“ $3En$ ”覆盖原则表明 90% 以上的云滴落在  $[Ex - 3En, Ex + 3En]$  范围内, 所以在计算云模型的相似性时, 也只需考虑在该范围之内的云滴和期望曲线, 因此将“ $3En$ ”原则引入到期望曲线中, 可定义限制其定义域范围后的期望曲线表达式[13]:

$$y'(x) = \begin{cases} \frac{x - (Ex - 3En)}{Ex - (Ex - 3En)}, & Ex - 3En \leq x \leq Ex \\ 1, & x = Ex \\ \frac{(Ex + 3En) - x}{(Ex + 3En) - Ex}, & Ex \leq x \leq Ex + 3En \\ 0, & \text{其他} \end{cases} \quad (4)$$

显然上式(4)符合三角模糊数的概念, 即  $y'$  为三角形模糊数, 记为  $y' = \langle Ex - 3En, Ex, Ex + 3En \rangle$ 。该三角模糊数的  $r$ -截集如下图 2 所示。



**Figure 2.** The  $r$ -cut interval number of triangular numbers  
**图 2.** 三角模糊数的  $r$ -截集

易计算得  $y'$  的  $r$ -截集的上下界为:

$$\begin{cases} \underline{u}(r) = Ex - 3En(1 - r) \\ \bar{u}(r) = Ex + 3En(1 - r) \end{cases} \quad (5)$$

该  $r$ -截集的期望和宽度分别为

$$\begin{cases} E([y']^r) = \frac{\bar{u}(r) + \underline{u}(r)}{2} = Ex \\ W([y']^r) = \frac{\bar{u}(r) - \underline{u}(r)}{2} = 3En(1 - r) \end{cases} \quad (6)$$

现对于两云模型分别为  $C_1(Ex_1, En_1, He_1)$  和  $C_2(Ex_2, En_2, He_2)$ , 将其期望曲线可看成如下所示的两个三角模糊数:

$$\begin{cases} u = \langle Ex_1 - 3En_1, Ex_1, Ex_1 + 3En_1 \rangle \\ v = \langle Ex_2 - 3En_2, Ex_2, Ex_2 + 3En_2 \rangle \end{cases} \quad (7)$$

现利用模糊数  $u, v$  的 CD-型贴进度来刻画两三角云的相似性, 提出三角云期望曲线的相似性度量方法(similarity measure method of expectation curve for triangular cloud model, ETCM), 记为

$$Sim(\text{CDETCM}) = N_{CD}^2(u, v) = \left( \int_0^1 (n_{CD}([u]^r, [v]^r))^2 dr \right)^{\frac{1}{2}} \quad (8)$$

其中  $r \in [0, 1]$ , 且  $Sim(\text{CDETCM})$  越大, 说明两云模型的相似度就越高。

具体的基于期望曲线的云模型相似性度量算法如下所示:

### 算法 2.1. CDETCM 算法

输入：云模型  $C_1(E_{x_1}, En_1, He_1)$  和  $C_2(E_{x_2}, En_2, He_2)$

输出：两云模型的相似性  $Sim(CDETCM)$

Step 1 依据式(7)，将两云模型表示成三角模糊数形式；

Step 2 由式(5)、(6)分别计算两三角模糊数  $r$ -截集的期望值  $E([u]^r)$  和  $E([v]^r)$  及宽度  $W([u]^r)$  和  $W([v]^r)$ ；

Step 3 将期望值和宽度代入公式(1)，计算 CD-型区间数距离  $f([u]^r, [v]^r)$ ；

Step 4 将  $f([u]^r, [v]^r)$  带入式(2)转化为区间数贴近度  $n_{CD}([u]^r, [v]^r)$ ；

Step 5 将  $n_{CD}([u]^r, [v]^r)$  代入式(3)，计算  $Sim(CDETCM)$ 。

注：计算过程中，令  $n=1$  以减少时间复杂度。

### 4.2. 三角云最大边界曲线的相似性度量

三角云最大边界曲线通过将超熵参数引入期望曲线实现动态扩展。即将原熵值  $En$  修正为复合参数  $en = En + 3He$ ，所以两个云模型的三角形模糊数分别变为

$$\begin{cases} u = \langle Ex_1 - 3en_1, Ex_1, Ex_1 + 3en_1 \rangle \\ v = \langle Ex_2 - 3en_2, Ex_2, Ex_2 + 3en_2 \rangle \end{cases}, \quad (9)$$

其中， $\begin{cases} en_1 = En_1 + 3He_1 \\ en_2 = En_2 + 3He_2 \end{cases}$ 。

该方法沿用与期望曲线相似的计算步骤，仅将原来的熵值( $En$ )替换成熵值( $En$ )加上超熵( $He$ )的 3 倍。三角云最大边界曲线的相似性度量方法(similarity measure method of maximum boundary curve for triangular cloud model, MTCM)记为  $Sim(CDMTCM)$ 。

具体的三角云最大边界曲线的相似性度量算法如下所示：

### 算法 2.2. CDMTCM 算法

输入：云模型  $C_1(E_{x_1}, En_1, He_1)$  和  $C_2(E_{x_2}, En_2, He_2)$

输出：两云模型的相似性  $Sim(CDMTCM)$

Step 1 依据式(9)，将两云模型表示成三角模糊数形式；

Step 2 将(5)式中的  $En$  替换成  $En + 3He$ ，再由式(5)、(6)分别计算两三角模糊数  $r$ -截集的期望值  $E([u]^r)$  和  $E([v]^r)$  及宽度  $W([u]^r)$  和  $W([v]^r)$ ；

Step 3 将期望值和宽度代入公式(1)，计算区间数距离  $f([u]^r, [v]^r)$

Step 4 将  $f([u]^r, [v]^r)$  带入式(2)转化为区间数贴近度  $n_{CD}([u]^r, [v]^r)$ ；

Step 5 将  $n_{CD}([u]^r, [v]^r)$  代入式(3)，计算  $Sim(CDMTCM)$ 。

注：计算过程中，令  $n=1$  以减少时间复杂度。

### 4.3. 综合的三角云模型相似性度量

云模型相似性度量方法需合理考虑三个特征值。基于三角云的期望曲线的相似性度量方法计算仅涉

及期望  $Ex$  和熵  $En$  的值, 没有考虑超熵  $He$  会影响的云模型的厚度, 这将导致一定的计算误差; 而基于三角云的最大边界曲线的相似性度量方法, 采用  $3He$  扩张边界, 虽捕捉了局部的波动特征, 但过度放大了超熵( $He$ )的影响。因此, 本文融合云模型的三个特征值, 定义一种加权计算三角云相似性的综合方法, 如下式(10)所示:

$$Sim(CDTCM) = \alpha Sim(CDETCM) + \beta Sim(CDMTCM) \tag{10}$$

其中,  $\alpha$  和  $\beta$  代表基于三角云的期望曲线及最大边界曲线的相似性权重, 且满足  $\alpha + \beta = 1$ 。另外, 受文献[24]的启发,  $\alpha$  与  $\beta$  的确定方法如下:

$$\begin{cases} d = \sqrt{(Sim(CDETCM) - Sim(CDMTCM))^2} \\ d = (\alpha - \beta)^2 \\ \alpha + \beta = 1 \\ \alpha \geq \beta \end{cases} \tag{11}$$

综合的三角云相似性度量算法如下所示:

**算法 2.3. CDTCM 算法**

- 
- 输入: 云模型  $C_1(Ex_1, En_1, He_1)$  和  $C_2(Ex_2, En_2, He_2)$
  - 输出: 两云模型的综合相似性  $Sim(CDTCM)$
  - Step 1 利用算法 2.1 计算出  $Sim(CDETCM)$ ;
  - Step 2 利用算法 2.2 计算出  $Sim(CDMTCM)$ ;
  - Step 3 将  $Sim(CDETCM)$  和  $Sim(CDMTCM)$  代入公式(11)解出其中的  $\alpha, \beta$ ;
  - Step 4 将  $\alpha, \beta$  代入公式(10), 计算综合相似度  $Sim(CDTCM)$ 。
- 

注: 计算过程中, 令  $n=1$  以减少时间复杂度。

## 5. 实验对比分析

为了说明本文所提算法的可行性及有效性, 首先通过数值仿真实验验证 CDETCM、CDMTCM、CDTCM 等方法的可行性; 其次利用该方法对 UCI 常用数据集(synthetic control chart dataset)和 UCR 中心心电图数据(ECG200)和牛肉光谱数据(Beef)等 10 个常用的数据集进行分类实验, 根据其分类精确度验证算法的分类性能以进一步说明方法的可行性和有效性。

### 5.1. 仿真实验

本文利用文献[8]中给出的 4 个云模型进行数值仿真实验, 它们分别为:

$$\begin{cases} C_1(1.5, 0.62666, 0.33900) \\ C_2(4.6, 0.60159, 0.30862) \\ C_3(4.4, 0.75199, 0.27676) \\ C_4(1.6, 0.60159, 0.30862) \end{cases}$$

其对应云图如图 3 所示, 根据本文所提算法上述 4 个云模型相似性度量结果如表 1 所示。

由下表 1 可知, 4 个云模型两两之间相似度大小为:

$(C_1, C_4) > (C_2, C_3) > (C_3, C_4) > (C_1, C_3) > (C_2, C_4) > (C_1, C_2)$ , 该相似性大小的排序与图 3 的直观印象相契合,

即该方法可行。再将本文所提 CDTCM 方法和现有的比较经典的相似性度量方法 ECM [9]、MCM [9]、LICM [8]、EMTCM 进行比较，不同相似性度量算法的计算结果对比如表 2 所示。

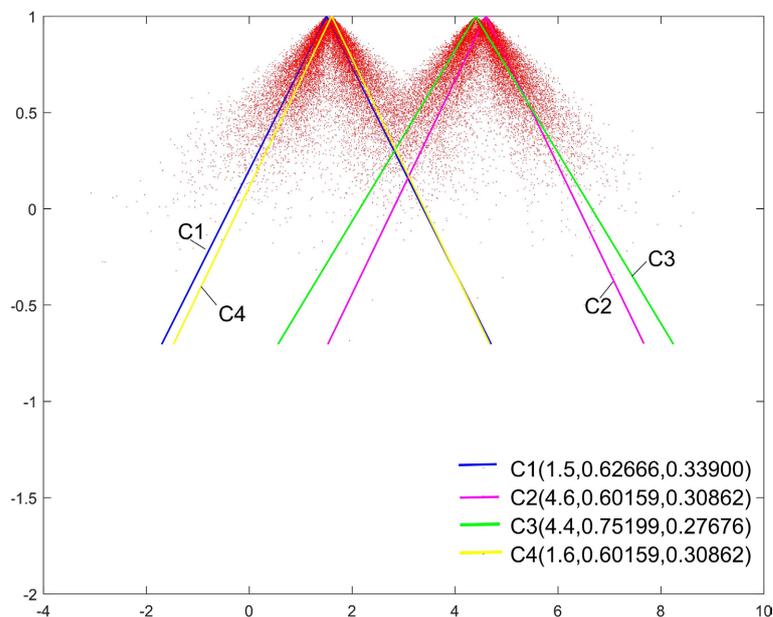


Figure 3. Cloud droplet distribution and the expected curve of four cloud models  
图 3. 四个三角云的云滴分布及期望曲线

Table 1. Three different similarity measurement methods for cloud models

表 1. 3 种不同云模型相似性度量方法

相似度	CDETCM	CDMTCM	CDTCM
$(C_1, C_2)$	0.1570	0.1568	0.1596
$(C_1, C_3)$	0.1658	0.1660	0.1659
$(C_1, C_4)$	0.8499	0.7680	<b>0.8172</b>
$(C_2, C_3)$	0.6763	0.7297	0.6968
$(C_2, C_4)$	0.1614	0.1614	0.1614
$(C_3, C_4)$	0.1705	0.1709	0.1707

表 2 的实验数据表明，CDTCM 与 ECM、MCM、LICM 及 EMTCM 方法在相似性大小判断上具有一致性，均识别出云模型  $(C_1, C_4)$  与  $(C_2, C_3)$  的相似度最高。然而，在差异性区分能力方面，CDTCM 展现出明显优势：同样是比较云模型  $(C_1, C_4)$  和  $(C_2, C_3)$  的相似度差值，ECM (差异度 0.077)、MCM (差异度 0.0028)、LICM (差异度 0.0002) 和 EMTCM (差异度 0.0874)，CDTCM 对  $(C_1, C_4)$  与  $(C_2, C_3)$  的相似性差异度达到 0.1204。特别地，LICM 因计算结果都很接近，导致区分能力不足。此外，CDTCM 方法还具有一定的稳定性，由于其实验结果仅由云模型的 3 个数字特征决定，不受云滴数量和实验重复次数的影响。

### 5.2. 时间序列分类实验

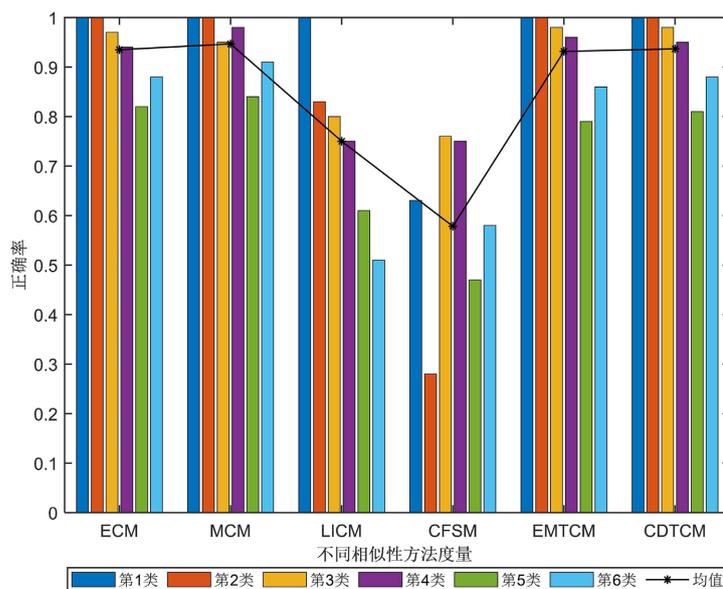
在数据挖掘[25]领域，时序数据分类方法[26]始终是研究焦点之一，而分类模型的好坏关键在于其相似性度量方法的选取。因此，为验证本文所提方法的有效性，本节进行时间序列分类实验。本实验首先

采用 UCI 中的常用时间序列数据集(synthetic control chart dataset) [27], 该数据集包含 6 类不同的时序数据, 每类 100 条, 单条长度为 60。实验中, 对每类数据采用 10 折交叉验证, 即将每类的 100 条数据均分为 10 组, 每次测试取其中的 1 组(10 条)为测试集, 剩下的 9 组(90 条)数据为训练集。

**Table 2.** Comparison of different similarity measurement methods for cloud models  
**表 2.** 不同的云模型相似性度量方法比较

相似度	ECM	MCM	LICM	EMTCM	CDTCM
$(C_1, C_2)$	0.0252	0.4545	0.9561	0.2419	0.1596
$(C_1, C_3)$	0.0893	0.4905	0.9648	0.2543	0.1659
$(C_1, C_4)$	<b>0.9465</b>	<b>0.9604</b>	<b>0.9990</b>	<b>0.8848</b>	<b>0.8172</b>
$(C_2, C_3)$	<b>0.8695</b>	<b>0.9576</b>	<b>0.9992</b>	<b>0.7974</b>	<b>0.6968</b>
$(C_2, C_4)$	0.0285	0.4525	0.9679	0.2500	0.1614
$(C_3, C_4)$	0.0964	0.4898	0.9755	0.2594	0.1707

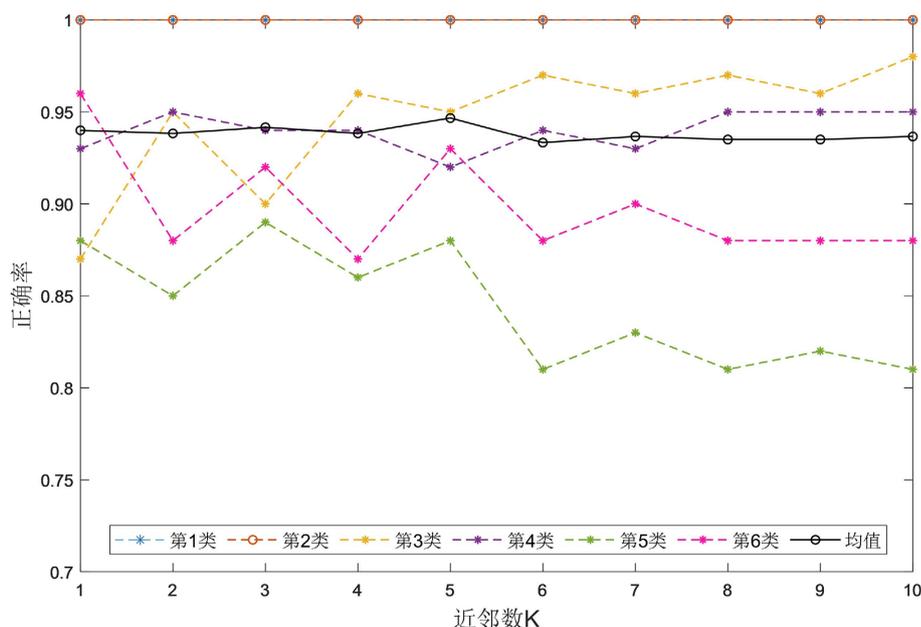
为了验证本文 CDTCM 方法的分类性能, 采用的是最近邻分类(K-Nearest Neighbors, KNN)算法[28]进行分类实验( $K=10$ )。仿真实验对比了云模型在不同相似性度量方法下的时间序列分类正确率。首先, 将每条时序数据通过逆向云变换算法表示为云模型, 然后利用不同的云模型相似度量方法分别计算每类测试集与其他数据(包含目标类训练样本与全量非目标类数据的混合数据)的相似度矩阵, 根据该相似度矩阵, 利用 KNN 算法对分类结果进行计算, 并通过 10 折交叉验证的方式, 以得到每类测试集的分类正确率, 最后对六类数据集的分类正确率取算术平均值, 可得到如图 4 所示的各方法的分类正确率。其中, 对比方法为 ECM [9]、MCM [9]、LICM [8]、CFSM [12]、EMTCM [13]等方法。其次, 对于 KNN 算法中不同的  $K$  值会导致不同的分类效果, 因此为了突显 CDTCM 方法的优点, 采用不同的参数  $K$  进行分类实验, 分析 CDTCM 方法分类正确率的变化趋势, 结果如图 5 所示。



**Figure 4.** Accuracy of different metric methods in classification  
**图 4.** 不同度量方法的分类正确率

由图 4 可清楚看到, 本文提出的相似度量方法在前四类数据中表现良好, 在前四类数据的平均分类正确率排序为  $CDTCM > MCM > EMTCM > ECM > LICM > CFSM$ 。在六类数据中的平均分类正确率该方法与 MCM 基本一致, 但高于 ECM、LICM、CFSM 及 EMTCM 方法。在分类正确率的稳定性方面, CDTCM 方法在每一类数据中的分类正确率均在 0.8 以上, 稳定性与传统的 ECM、MCM、EMTCM 方法相当, 但明显优于 LICM 及 CFSM 方法。该实验结果表明本文方法是有效的, 且充分展示了其在时间序列数据分类中良好性能。

由图 5 可知, 对于不同的近邻数  $K$ , CDTCM 方法对第 1 类与第 2 类数据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 类数据的分类正确率虽呈现波动趋势, 但波动幅度较小, 且随着  $K$  值增加, 分类正确率略有提升; 第 5、6 类数据的分类正确率也呈波动趋势, 但当  $K \geq 4$  时, 波动幅度较小且趋于稳定; 而对于在不同  $K$  值下的平均分类率波动幅度很小。该仿真实验结果表明, 总体上近邻数  $K$  对平均分类正确率的影响并不大, 说明本文所提方法性能良好。



**Figure 5.** Comparison of the classification accuracy results of CDTCM under different values of the number of neighbors ( $K$ )

**图 5.** CDTCM 在不同近邻数  $K$  下的分类正确率结果对比

为进一步更加全面地验证本文所提 CDTCM 方法的分类性能, 从 UCR 数据库中选取了心电图数据 (ECG200) 和牛肉光谱数据 (Beef) 等 10 个时间序列数据集进行分类实验, 这些数据集中的各个数据序列样本带有分类标签, 且均已被划分为不同规模大小的训练集和测试集。本实验分析的是各云模型相似性度量算法在时间序列分类中的分类错误率, 采用最近邻算法 1-NN 作为时间序列分类器。为提高分类效率, 将每个时间序列降维分段处理, 并将分段后的时间序列利用逆向云算法转换为云模型, 在同一维数段上云模型, 分别利用 LICM、ECM、MCM、CFSM 和 CDTCM 算法来计算相似度。本实验的降维强度  $w = \{2, 3, 4, 5, 6, 8, 10, 15, 20\}$ 。各方法在 10 个不同数据集下的分类错误率如表 3 所示。

由上表 3 可知, 本文所提出的 CDTCM 方法在上述 10 个数据集上进行分类实验的分类错误率较低, 除 Adiac、Beef 数据集外, 其他 9 个数据集的分类错误率都在 0.1 以下, 这表明该方法针对不同的数据集都表现出良好的分类效果, 且与传统的 ECM、MCM 及 CFSM 等方法不相上下。

**Table 3.** Classification error rates of different cloud model similarity measures  
**表 3.** 不同的云模型相似性度量方法分类错误率比较

Name	ECM	MCM	LICM	CFSM	CDTCM
Gunpoint	0.060	0.053	0.067	0.120	0.080
CBF	0.003	0.015	0.037	0.015	0.010
Trace	0.080	0.000	0.110	0.170	0.090
Wafer	0.010	0.016	0.004	0.021	0.013
Facefour	0.057	0.057	0.057	0.090	0.045
ECG200	0.130	0.120	0.160	0.150	0.090
Adiac	0.376	0.384	0.439	0.491	0.366
Beef	0.433	0.300	0.300	0.366	0.366
Coffee	0.250	0.036	0.036	0.143	0.036
OliveOil	0.167	0.100	0.066	0.133	0.100

## 6. 结束语

云模型相似性度量是当前研究的热点，其相似性度量方法的好坏至关重要。因此，为解决现有方法存在的区分度不高、结果不稳定等问题，本文提出了一种基于 CD-型贴近度的云模型相似性度量方法，并通过数值模拟和时间序列分类实验，对该方法的可行性及有效性进行验证，得到以下结论：

本文将云模型扩展后的三角云的期望曲线及最大边界曲线看作三角模糊数，结合 CD-型区间数距离，构造 CD-型贴近度，求解两个云模型的相似度，该方法综合考虑了云模型的三个数字特征，计算过程简单，结果稳定且不受实验次数和云滴个数影响。

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，且结果较传统的 ECM、MCM、LICM、EMTCM 方法有更高的区分度。而在时间序列分类实验中，对于 synthetic control chart dataset 数据集的不同类型的数据 CDTCM 方法的分类准确率均在 0.8 以上，明显优于传统的 LICM 和 CFSM 方法，展现了算法的稳定性。而对于 UCR 数据库中 10 个数据集也展现出良好的分类效果。

本文提出的相似度量方法是对云模型理论的重要补充和完善，下一步的研究方向是将模糊数学中的相关理论与云模型相似性度量进行更深层次的有机融合。

## 基金项目

铀资源探采与核遥感全国重点实验室自主部署项目(2024QZ-TD-25)。

## 参考文献

- [1] 刘常昱, 李德毅, 潘莉莉. 基于云模型的不确定性知识表示[J]. 计算机工程与应用, 2004(2): 32-35.
- [2] 苗夺谦, 张清华, 钱宇华, 等. 从人类智能到机器实现模型-粒计算理论与方法[J]. 智能系统学报, 2016, 11(6): 743-757.
- [3] Kang, Y., Li, Q. and Liu, Y. (2024) Research on IoT Edge Data Mining Techniques Based on Cloud Computing Models. *Applied Mathematics and Nonlinear Sciences*, **9**, 1-17. <https://doi.org/10.2478/amns-2024-0903>
- [4] 郭秋燕, 胡磊, 代劲. 基于云模型的变分自编码器数据压缩方法[J]. 电子技术应用, 2023, 49(10): 96-99.
- [5] Wang, S., Xie, Y. and Fang, M. (2011) A Collaborative Filtering Recommendation Algorithm Based on Item and Cloud Model. *Wuhan University Journal of Natural Sciences*, **16**, 16-20. <https://doi.org/10.1007/s11859-011-0704-4>
- [6] Zhu, C., Liu, X., Ding, W. and Zhang, S. (2024) Cloud Model-Based Multi-Stage Multi-Attribute Decision-Making

- Method under Probabilistic Interval-Valued Hesitant Fuzzy Environment. *Expert Systems with Applications*, **255**, Article ID: 124595. <https://doi.org/10.1016/j.eswa.2024.124595>
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 查翔, 倪世宏, 谢川, 等. 云相似度的概念跃升间接计算方法[J]. 系统工程与电子技术, 2015, 37(7): 1676-1682.
- [11] 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.
- [12] 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.
- [13] 黄琼桃, 刘瑞敏. 基于 EW-型贴近度的云模型相似性度量方法[J]. 贵州大学学报(自然科学版), 2021, 38(2): 44-49.
- [14] Zhu, G., Yang, Y., Sun, R., Wu, E.Q. and Law, R. (2024) A Similarity Measurement Method of Normal Cloud Models for the Operational Status Perception and Computing of Urban Rail Transit. *IEEE Transactions on Computational Social Systems*, **11**, 746-755. <https://doi.org/10.1109/tcss.2022.3215893>
- [15] 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.
- [16] Xu, X., Guo, H., Zhang, Z., Yu, S., Chang, L., Steyskal, F., et al. (2024) A Cloud Model-Based Interval-Valued Evidence Fusion Method and Its Application in Fault Diagnosis. *Information Sciences*, **658**, Article ID: 119995. <https://doi.org/10.1016/j.ins.2023.119995>
- [17] 孙妮妮, 陈泽华, 牛昱光, 等. 基于云模型重叠度的相似性度量[J]. 计算机应用, 2015, 35(7): 1955-1964.
- [18] 雷一鸣, 代兵, 包玉娥. 关于模糊数贴近度问题的研究[J]. 纯粹数学与应用数学, 2015, 31(6): 611-619.
- [19] Yang, J., Han, J., Wan, Q., Xing, S. and Chen, F. (2023) A Novel Similarity Measurement for Triangular Cloud Models Based on Dual Consideration of Shape and Distance. *PeerJ Computer Science*, **9**, e1506. <https://doi.org/10.7717/peerj-cs.1506>
- [20] 包玉娥, 雷一鸣. 模糊数空间上的积分度量及其在模糊聚类中的应用[J]. 模糊系统与数学, 2017, 31(4): 69-77.
- [21] Coppi, R. and D'urso, P. (2003) Regression Analysis with Fuzzy Informational Paradigm: A Least-Squares Approach Using Membership Function Information. *International Journal of Pure and Applied Mathematics*, **8**, 279-306.
- [22] 彭晓芹, 包玉娥, 赵博. 关于区间数贴近度的一般表示形式的研究[J]. 数学的实践与认识, 2014, 44(22): 258-262.
- [23] 吴从焮, 赵志涛, 任雪昆. 模糊分析与特殊泛函空间[M]. 哈尔滨: 哈尔滨工业大学出版社, 2013.
- [24] 阮永芬, 高春钦, 李志伟, 等. 基于改进 AHP 与熵权法的膨胀土胀缩等级云模型评价[J]. 江苏大学学报(自然科学版), 2017, 38(2): 218-223.
- [25] Li, H. and Chen, M. (2023) Time Series Clustering Based on Normal Cloud Model and Complex Network. *Applied Soft Computing*, **148**, Article ID: 110876. <https://doi.org/10.1016/j.asoc.2023.110876>
- [26] Liu, Y., Liu, Z., Li, S., Guo, Y., Liu, Q. and Wang, G. (2023) Cloud-Cluster: An Uncertainty Clustering Algorithm Based on Cloud Model. *Knowledge-Based Systems*, **263**, Article ID: 110261. <https://doi.org/10.1016/j.knosys.2023.110261>
- [27] Pham, D.T. and Chan, A.B. (1998) Control Chart Pattern Recognition Using a New Type of Self-Organizing Neural Network. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, **212**, 115-127. <https://doi.org/10.1243/0959651981539343>
- [28] 崔昊阳, 张晖, 周雷, 等. 有序规范实数对多相似度 K 最近邻分类算法[J]. 计算机应用, 2023, 43(9): 2673-2678.