

Research Article

Uncertain Distribution-Based Similarity Measure of Concepts

Shuai Li,^{1,2} Jie Yang ,³ Zhipeng Qi,¹ and Juanli Zeng³

¹School of Mathematics and Information Sciences, Nanchang Hangkong University, Nanchang 330063, Jiangxi, China

²Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

³School of Physics and Electronic Science, Zunyi Normal University, Zunyi 563002, China

Correspondence should be addressed to Jie Yang; 530966074@qq.com

Received 15 April 2020; Accepted 14 August 2020; Published 19 September 2020

Academic Editor: Eric Lefevre

Copyright © 2020 Shuai Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The similarity of concepts is a basic task in the field of artificial intelligence, e.g., image retrieval, collaborative filtering, and public opinion guidance. As a powerful tool to express the uncertain concepts, similarity measure based on cloud model (SMCM) is always utilized to measure the similarity between two concepts. However, current studies on SMCM have two main limitations: (1) the similarity measures based on conceptual intension lack interpretability for merging the numerical characteristics and cannot discriminate some different concepts. (2) The similarity measures based on conceptual extension are always unstable and inefficient. To address the above problems, an uncertain distribution-based similarity measure of cloud model (UDCM) is proposed in this paper. By analyzing the definition of the CM, we propose a new complete uncertainty including first-order and second-order uncertainty to calculate the uncertainty more accurately. Then, based on the difference between the complete uncertainty of two concepts, the computing process of UDCM and its some properties are introduced. Finally, we exhibit its advantages by comparing with other methods and verify its validity by experiments.

1. Introduction

The similarity of concepts is a basic sense for human cognition, which is also a fundamental task in artificial intelligence. It plays a crucial role in semantic information retrieval systems [1–3], sense disambiguation [1, 4], and information extraction [3, 5]. Due to the uncertainty of the concept, for different people, their cognitions for the same concept are also different [6–10]. To describe various forms of uncertainty, researchers proposed many models: probability models for randomness [11], fuzzy set models for vagueness [12], and rough set models for inconformity and incompleteness [13]. Cloud model (CM) [14] proposed by Li et al. described the uncertain concepts in human language and realized human bidirectional cognition between their extension and intension. Owing to the strong expression ability for uncertain concepts, similarity measure based on cloud model (SMCM) has been applied in many scenarios, such as image segmentation [15–18], collaborative filtering [19–21], and synthetic evaluation [22–26]. Although SMCM

has been successful in many scenarios, there are two urgent problems which need to be addressed.

From the perspective on conceptual intension, the similarity of the numerical characteristics can depict the similarity of concepts. To express the complex form of uncertainty, different numerical characteristics express different meaning for an uncertain concept. There lacks a reasonable method of merging numerical characteristics to measure similarity.

Example 1. Figure 1 shows the shooting results of three shooters. We are asked to evaluate the similarity of performances between A , B , and C . It is an intractable problem. As shown in Figure 1, the variance of C 's shooting results is close to A , which means the same psychological state is shared by them, but the mean value of C 's performance is far from A , which means a significant difference between their shooting levels. Although B 's shooting results are scattered, the mean value of B performance is close to A . For the psychological state, C is more close to A than B , but B is

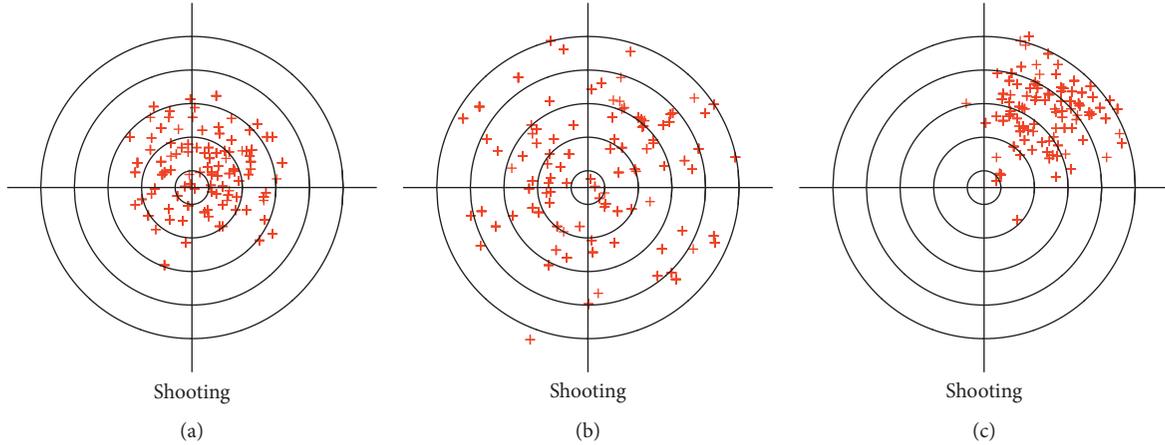


FIGURE 1: Shooting results of three shooters.

more close to A than C from the aspect of shooting level. That is, considering the similarity from different angles, different results can be obtained. Hence, we need a more comprehensive method to measure the similarity.

On the contrary, the similarity measure is always computed by random realization, which directly reflects the similarity from abundant samples of concepts. The more the samples are generated, the higher the accuracy of SMCM is. Therefore, we have to spend excessive computing time acquiring accurate SMCM. Besides, due to random realization, the results of SMCM are different each time, which can be illustrated in the following example:

Example 2. Supposing two cloud models $C_1(5, 3, 1)$ and $C_2(5, 2, 0.5)$, their similarity is computed by fuzzy distance-based similarity (FDCM) [27], wherein the estimated overall score \hat{s} is with $n = 50$. Table 1 shows results of their 10 similarity measures.

From the above analyses, it can be seen that SMCM still needs further study. To address the above problems, in this paper, we propose a new notion called complete uncertainty to depict the whole uncertainty in the process from numerical characteristics to the conceptual extension. Then, a new SMCM is presented based on completed uncertainty, which reflects the similarity of the uncertain distribution of two concepts. Compared with the SMCM based on extension, the new SMCM has an invariable result. Besides, it has a more reasonable method to merge numerical characteristics compared with the SMCM based on intension. Moreover, because that new SMCM reflects the complete uncertainty of the CM, it can acquire more accurate similarity results between two concepts.

The remainder of this paper is organized as follows. The related definitions of the CM and current SMCM are introduced in Section 2. We propose uncertain distribution-based SMCM and elaborate the calculation in detail in Section 3. Section 4 provides four experiments to show the effectiveness of the proposed method. Final conclusions are presented in Section 5.

2. Preliminaries

In this section, we review relative concepts of the CM and current methods for similarity of the cloud model.

2.1. Cloud Model

Definition 1 (see [14]). Let U be a nonempty infinite set and $C(Ex, En, He)$. If there is a number $x \in U$, which is a random realization of the concept C and satisfies $x \sim R_N(Ex, y)$, where $y \sim R_N(En, He)$ and the certainty degree of x on U is $\mu(x) = \exp\{-(x - Ex)^2/2y^2\}$, then the distribution of x on U is a Gaussian cloud or normal cloud, and each x is defined as a Gaussian cloud drop.

As a crucial model of CM, the Gaussian cloud model is applied widely due to the universality of the Gaussian distribution, and we only discuss the Gaussian cloud model in this paper. Gaussian cloud model introduces three numerical characteristics including Ex , En , and He , which denote mathematical expectation, entropy, and hyperentropy. It accords with human thought [14, 28–32] and depicts a unified framework of randomness and vagueness in the human cognitive process. Herein, the expectation represents the basic determinate domain of the qualitative concept, the entropy represents the uncertainty for the qualitative concept, and the hyperentropy represents the uncertainty for entropy. Figure 2 shows the shape and the characteristic curves of CM $C(20, 6, 1)$, i.e., curve $y_1 = \exp\{-(x - Ex)^2/2(En - 3He)^2\}$, which is called the inner envelope, and $y_2 = \exp\{-(x - Ex)^2/2(En - 3He)^2\}$, which is called the outer envelope. We also call curve $y = \exp\{-(x - Ex)^2/2En^2\}$ the expectation curve.

In cognitive computing, cloud drops are called conceptual extension, i.e., samples of a concept. The numerical characteristics, expectation Ex , entropy En , and hyperentropy He , are called conceptual intension representing the essence of a concept. There are many algorithms to implement bidirectional transformation between extension and intension of the CM, which can be called forward cloud transformation (FCT) algorithm and backward cloud

TABLE 1: Results of 10 similarity measures of $C_1(5, 3, 1)$ and $C_2(5, 2, 0.5)$.

| No. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|--------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Similarity measure | 0.9858 | 0.9766 | 0.9522 | 0.9912 | 0.9832 | 0.9732 | 0.9655 | 0.9557 | 0.9321 | 0.9815 |

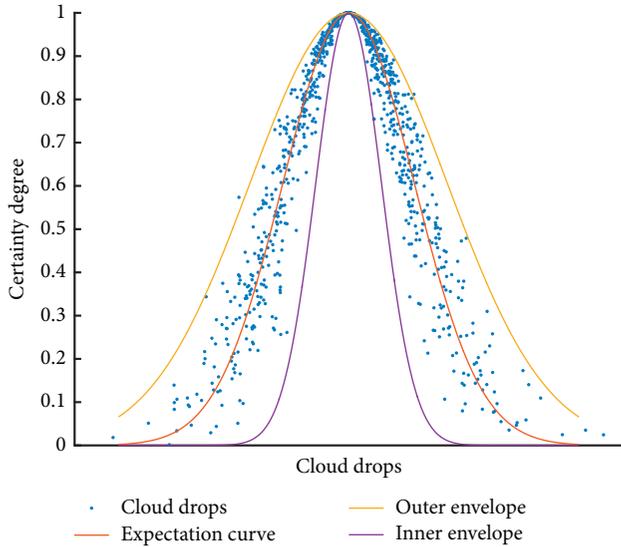


FIGURE 2: Characteristic curves of $C(20, 6, 1)$.

transformation (BCT) algorithm. Due to the limitation of space, we do not expatiate these algorithms in this paper, and relative methods can be found in [14, 32, 33].

2.2. *Similarity Measure of Cloud Model.* To describe the similarity between two cloud concepts, many SMCM are proposed currently. Generally speaking, a suitable similarity measure should assure the correct conclusions in specific situations and require discriminability, efficiency, stability, and interpretability. Zhang et al. [34] used the average distance of cloud drops generated by FCT to measure the distance between two cloud models. This method is called concept extension-based similarity measure (CS). It is understandable and accords with human cognition. However, calculation of average distance is highly complex and instable. Other researchers study SMCM based on concept intension. Likeness comparing method based on cloud model (LICM) [19] employs included angle cosine of vectors composed by numerical characteristics to measure similarity. It has high efficiency, but ignores the relationships among numerical characteristics leading to unreasonable results sometimes. Li et al. [35] proposed expectation based on cloud model (ECM) and max boundary based on cloud model (MCM), which define similarity measure by overlapping the area of characteristic curves. These methods employ characteristic curves to denote the certainty degrees of cloud drops and then use similarity of uncertainty to measure similarity of concepts. Other SMCM based on characteristic curves can be found in [16, 27, 36, 37]. Table 2 shows the comparison of

TABLE 2: A list of similarity measures of CM.

| | CS | LICM | MCM | ECM |
|------------------|------|------|--------|--------|
| Discriminability | High | Low | Low | Low |
| Efficiency | Low | High | High | High |
| Stability | Low | High | High | High |
| Interpretability | High | Low | Medium | Medium |

similarity measures mentioned above on discriminability, efficiency, stability, and interpretability perspectives. In reality, due to merits on high efficiency, stability, and interpretability, MCM and ECM are widely applied in many situations.

From Table 2, the current popular SMCM are inefficient to distinguish two different concepts except CS because they only use partial uncertain relationship to measure the similarity between concepts. For example, if two cloud models have same expectation and entropy, their similarity calculated by ECM equals to 1, although they have different hyperentropy. In other words, ECM only uses the uncertain relationship between expectation and entropy, which is called first-order uncertainty. Ignoring hyperentropy will cause incorrect results. In the next section, we define a new notion of CM called complete (first-order and second-order) uncertainty and propose a new SMCM based on complete uncertainty, which focuses on the difference between distributions of two concepts.

3. Uncertain Distribution-Based SMCM

3.1. *Complete Uncertainty of the Cloud Model.* From discussion in Section 2, it is a critical step for SMCM based on conceptual intension to compute the whole uncertainty of CM. Due to the 3σ principle, we only consider cloud drops of CM in the interval $[Ex - 3En, Ex + 3En]$. Let $\mu(x)$ be a certainty degree on cloud drop x ; the uncertainty of CM C is calculated by $U(C) = \int_{Ex-3En}^{Ex+3En} \mu(x) dx$. This integral is very complex because $\mu(x) = \exp\{-(x - Ex)^2/2y^2\}$ also varies depending on the random variable $y \sim R_N(En, He)$. So, there are two forms of uncertainty in Definition 1. First-order uncertainty is certainty degree of drops, and second-order uncertainty is the uncertainty of certainty degree. The complete uncertainty should include two uncertainties and reflect the relationships among three numerical characteristics. Figure 3 shows the two forms of uncertainty and their relationships. Hence, for each cloud drop, its uncertainty is a fuzzy set $U(x) = \{(\exp\{-(x - Ex)^2/2y^2\}, \mu_y)\}$, where $\mu_y = \exp\{-(y - En)^2/2He^2\}$ is the membership function depending on y . Then, the complete uncertainty of CM C is denoted by

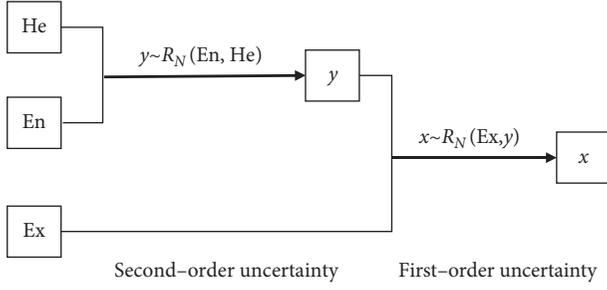


FIGURE 3: Two uncertainties and relationships among numerical characteristics.

$$U(x) = \left\{ \left(\exp \left\{ -\frac{(x - Ex)^2}{2y^2} \right\}, \exp \left\{ -\frac{(y - En)^2}{2He^2} \right\} \right) \mid y \in \mathbb{R}^+ \right\}. \quad (1)$$

3.2. *Uncertain Distribution-Based SMCM.* Based on formula (1), uncertain distribution-based similarity measure of cloud model (UDCM) can be defined as the following:

Definition 2. Let $C_1(Ex_1, En_1, He_1)$ and $C_2(Ex_2, En_2, He_2)$ be two cloud models; uncertain distribution-based similarity measure of cloud model between C_1 and C_2 is defined as the following:

$$\text{UDCM}(C_1, C_2) = \frac{1}{|L|} \int_L S(U_1(x), U_2(x)) dx, \quad (2)$$

where $L = [Ex_1 - 3En_1, Ex_1 + 3En_1] \cup [Ex_2 - 3En_2, Ex_2 + 3En_2]$, $|L|$ is the length of L , and $U_1(x)$ and $U_2(x)$ are uncertainties of x for C_1 and C_2 , respectively.

In Definition 2, $S(\cdot, \cdot)$ is a similarity measure of the fuzzy set. Therefore, UDCM is a framework of SMCM. There are various UDCM depending on different definitions of $S(\cdot, \cdot)$. In this paper, we define $S(\cdot, \cdot)$ as the following:

$$S(\tilde{A}, \tilde{B}) = \frac{|\tilde{A} \cap \tilde{B}|}{|\tilde{A} \cup \tilde{B}|}, \quad (3)$$

where \tilde{A} and \tilde{B} are two fuzzy sets on universal set U and $|\cdot|$ is the cardinal number of the fuzzy set.

Figure 4 shows the illustration of UDCM. Two horizontal scatter diagrams represent the relationship between cloud drops and certain degree, i.e., first-order uncertainty. Two vertical curves represent the uncertainty of certain degree, where $x = x_0$, i.e., second-order uncertainty. Owing to formula (3), $S(U_1, U_2)$ is the ratio of the blue area to the green area. From Figure 4, we know that $0 \leq S(U_1(x), U_2(x)) \leq 1$, so $0 \leq \text{UDCM}(C_1, C_2) \leq 1$.

Next, we try to calculate $S(U_1(x), U_2(x))$ for each $x \in L$. Its uncertainty is a fuzzy set:

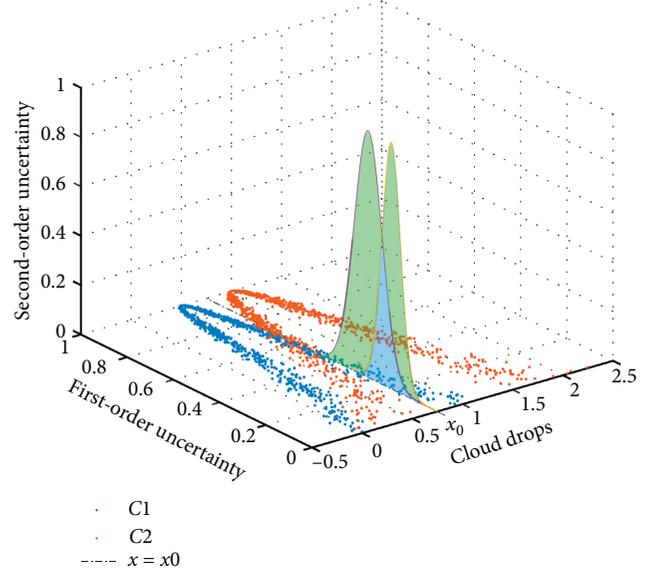


FIGURE 4: Illustration of Definition 2.

$$U_1(x) = \left\{ \left(\exp \left\{ -\frac{(x - Ex_1)^2}{2y^2} \right\}, \exp \left\{ -\frac{(y - En_1)^2}{2He_1^2} \right\} \right) \mid y \in \mathbb{R}^+ \right\},$$

$$U_2(x) = \left\{ \left(\exp \left\{ -\frac{(x - Ex_2)^2}{2y^2} \right\}, \exp \left\{ -\frac{(y - En_2)^2}{2He_2^2} \right\} \right) \mid y \in \mathbb{R}^+ \right\}. \quad (4)$$

Firstly, we should find the relationship between membership functions of two fuzzy sets when their elements are equal. For each $x \in L$, let y_1 and y_2 be variables for membership functions of $U_1(x)$ and $U_2(x)$, respectively. Since

$$\exp \left\{ -\frac{(x - Ex_1)^2}{2y_1^2} \right\} = \exp \left\{ -\frac{(x - Ex_2)^2}{2y_2^2} \right\}, \quad (5)$$

we have

$$y_2 = \frac{|x - Ex_2|}{|x - Ex_1|} y_1. \quad (6)$$

Then,

$$U_2(x) = \left\{ \left(\exp \left\{ -\frac{(x - Ex_2)^2}{2y_1^2} \right\}, \exp \left\{ -\frac{(y_1 - |x - Ex_1/x - Ex_2|En_2)^2}{2(|x - Ex_1/x - Ex_2|He_2)^2} \right\} \right) \mid y \in \mathbb{R}^+ \right\}. \quad (7)$$

So, (2) can be written as

$$\text{UDCM}(C_1, C_2) = \frac{1}{|L|} \int_L \left(\int_{\mathbb{R}^+} \frac{|U_1(x, y) \cap U_2(x, |x - \text{Ex}_2/x - \text{Ex}_1|y)|}{|U_1(x, y) \cup U_2(x, |x - \text{Ex}_2/x - \text{Ex}_1|y)|} dy \right) dx. \tag{8}$$

It is obvious that membership functions of both fuzzy sets U_1 and U_2 are exponential parts of the Gaussian distribution for each $x \in L$, respectively. Hence, we only calculate the intersection and the union of two exponential parts respective with $R_N(\text{En}_1, \text{He}_1)$ and $R_N(|x - \text{Ex}_1/x - \text{Ex}_2| \text{En}_2, |x - \text{Ex}_1/x - \text{Ex}_2| \text{He}_2)$. We do not introduce the process of calculation in this paper; more details can be found in [38].

The remainder is to calculate the integral in formula (8). Unfortunately, its integrand is not an elementary function, and the result of the integral is not an analytic expression. We have to calculate it by the numerical method. In definition of integration, it satisfies the following equation

$$\int_a^b f(x)dx = \lim_{\Delta x \rightarrow 0} \sum_{i=0}^{n-1} \frac{1}{2} \Delta x (f(a + i\Delta x) + f(a + (i + 1)\Delta x)),$$

$$n = \left\lfloor \frac{b - a}{\Delta x} \right\rfloor. \tag{9}$$

Function $[a]$ obtains the maximal integer less than a . For convenience, we denote function $S(U_1(x), U_2(x))$ as $\text{Sim}(x)$. There are two cases of integral interval L . If two intervals $[\text{Ex}_1 - 3\text{En}_1, \text{Ex}_1 + 3\text{En}_1]$ and $[\text{Ex}_2 - 3\text{En}_2, \text{Ex}_2 + 3\text{En}_2]$ are intersecting, let $a = \min\{\text{Ex}_1 - 3\text{En}_1, \text{Ex}_2 - 3\text{En}_2\}$ and $b = \max\{\text{Ex}_1 - 3\text{En}_1, \text{Ex}_2 - 3\text{En}_2\}$, and $\text{UDCM}(C_1, C_2)$ can be calculated by

$$\text{UDCM}(C_1, C_2) = \frac{1}{|L|} \lim_{\Delta x \rightarrow 0} \sum_{i=0}^{n-1} \frac{1}{2} \Delta x (\text{Sim}(a + i\Delta x) + \text{Sim}(a + (i + 1)\Delta x)), \quad n = \left\lfloor \frac{b - a}{\Delta x} \right\rfloor. \tag{10}$$

If two intervals $[\text{Ex}_1 - 3\text{En}_1, \text{Ex}_1 + 3\text{En}_1]$ and $[\text{Ex}_2 - 3\text{En}_2, \text{Ex}_2 + 3\text{En}_2]$ are separated, let $a = \text{Ex}_1 - 3\text{En}_1$ and $b = \text{Ex}_2 - 3\text{En}_2$, and $\text{UDCM}(C_1, C_2)$ can be calculated by

$$\begin{aligned} \text{UDCM}(C_1, C_2) &= \frac{1}{|L|} \lim_{\Delta x \rightarrow 0} \sum_{i=0}^{n_1-1} \frac{1}{2} \Delta x (\text{Sim}(a + i\Delta x) \\ &\quad + \text{Sim}(a + (i + 1)\Delta x)) \\ &\quad + \frac{1}{|L|} \lim_{\Delta x \rightarrow 0} \sum_{i=0}^{n_2-1} \frac{1}{2} \Delta x (\text{Sim}(b + i\Delta x) \\ &\quad + \text{Sim}(b + (i + 1)\Delta x)), \\ n_1 &= \left\lfloor \frac{12\text{En}_1}{\Delta x} \right\rfloor, n_2 = \left\lfloor \frac{12\text{En}_2}{\Delta x} \right\rfloor. \end{aligned} \tag{11}$$

In order to explain calculation of Definition 2 clearly, we compute UDCM of $C_1(0.5, 0.2, 0.04)$ and $C_2(1, 0.3, 0.06)$. Since $[\text{Ex}_1 - 3\text{En}_1, \text{Ex}_1 + 3\text{En}_1] = [-0.1, 1.1]$ and $[\text{Ex}_2 - 3\text{En}_2, \text{Ex}_2 + 3\text{En}_2] = [0.1, 1.9]$, we have $L = [-0.1, 1.9]$.

We divide L into 9 intervals, while $\Delta x = 0.2222$ and $n = 10$. For each point x_i , $S(U_1(x), U_2(x))$ is calculated which is given in Table 3.

Then, we calculate $\text{UDCM}(C_1, C_2) = 1/|L|\Delta x \sum_{i=0}^9 1/2 (\text{Sim}(x_i) + \text{Sim}(x_{i+1})) = 0.0523$. Although this result is imprecise while $n = 10$, we can calculate a more precise value with n increasing, and the result converges to a stable value ultimately. Figure 5 shows that $\text{UDCM}(C_1, C_2)$ converges to 0.0724 with n increasing, which illustrates that we can obtain a stable result by Definition 2.

There is a special situation of Definition 2 which is as follows.

Theorem 1. Let $C_1(\text{Ex}_1, \text{En}_1, \text{He}_1)$ and $C_2(\text{Ex}_2, \text{En}_2, \text{He}_2)$ be two cloud models; if they satisfy $\text{Ex}_1 = \text{Ex}_2$ and $\text{En}_1 = \text{En}_2$, the UDCM of them is

$$\text{UDCM}(C_1, C_2) = \frac{1}{|L|} \int_L S(U_1(x), U_2(x)) dx = \frac{\min(\text{He}_1, \text{He}_2)}{\max(\text{He}_1, \text{He}_2)}. \tag{12}$$

Proof. Since $\text{Ex}_1 = \text{Ex}_2$ and $\text{En}_1 = \text{En}_2$, according to Definition 2, $L = [\text{Ex}_1 - 3\text{En}_1, \text{Ex}_1 + 3\text{En}_1]$, and $|L| = 6\text{En}_1$, and $|x - \text{Ex}_2/x - \text{Ex}_1| = 1$. For each $x \in L$, the value of $S(U_1(x), U_2(x))$ is invariable. Without loss of generality, we assume $\text{He}_1 \leq \text{He}_2$. Due to $\text{En}_1 = \text{En}_2$, there are

$$L_{\min}(y) = U_1(x) \cap U_2(x) = U_1(x) = \exp\left\{-\frac{(y - \text{En}_1)^2}{2\text{He}_1^2}\right\},$$

$$L_{\max}(y) = U_1(x) \cup U_2(x) = U_2(x) = \exp\left\{-\frac{(y - \text{En}_2)^2}{2\text{He}_2^2}\right\}. \tag{13}$$

Hence, the similarity of U_1 and U_2 is

$$S(U_1(x), U_2(x)) = \frac{\int_{\mathbb{R}^+} L_{\min}(y) dy}{\int_{\mathbb{R}^+} L_{\max}(y) dy} = \frac{\text{He}_1}{\text{He}_2}. \tag{14}$$

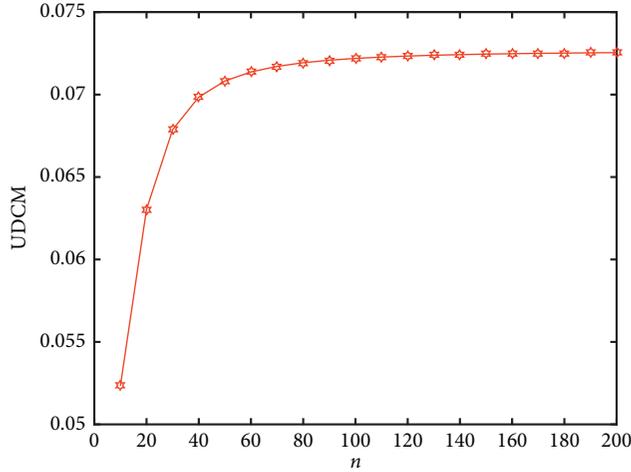
The UDCM of C_1 and C_2 is

$$\text{UDCM}(C_1, C_2) = \frac{1}{|L|} \int_L S(U_1(x), U_2(x)) dx = \frac{\text{He}_1}{\text{He}_2}. \tag{15}$$

In order to illustrate UDCM has a high ability to distinguish two different concepts, we have the following theorem. \square

TABLE 3: Results of $S(U_1(x), U_2(x))$ for each point x_i .

| x_i | -0.1000 | 0.1222 | 0.3444 | 0.5667 | 0.7889 | 1.0111 | 1.2333 | 1.4556 | 1.6778 | 1.9000 |
|--------------|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Sim(x_i) | 0.4462 | 0.1638 | 0.0089 | 8.9e-4 | 0.0442 | 3.1e-7 | 5.8e-4 | 0.0048 | 0.0131 | 0.0233 |

FIGURE 5: UDCM with variation of n .

Lemma 1. Let $f(x) \in [0, 1]$ be an integrable function and L be an interval. $f(x)$ satisfies $\int_L f(x)dx = |L|$, where $|L|$ is the length of interval L if and only if $f(x) = 1$ holds on L , almost everywhere.

Proof. To prove this lemma, we must employ measure theory. Sufficiency: suppose that $f(x) = 1$ holds on L , almost everywhere. Due to countable additivity of integration,

$$\int_L 1 - f(x)dx = \int_{E(f=1)} 1 - f(x)dm + \int_{E(f \neq 1)} 1 - f(x)dm = 0, \quad (16)$$

where $E(f = 1)$ means the set of x satisfying $f(x) = 1$.

Necessary: suppose n is a natural number.

$$\int_L 1 - f(x)dx \geq \int_{E(1-f \geq 1/n)} 1 - f(x)dm \geq \frac{1}{n}mE\left(1 - f \geq \frac{1}{n}\right), \quad (17)$$

where $mE(\cdot)$ is the Lebesgue measure. Due to $\int_L 1 - f(x)dx = 0$, for each n , there is $mE(1 - f \geq 1/n) = 0$. The set $E(f \neq 1)$ can be denoted by $E(f \neq 1) = \cup_{n=1}^{\infty} E(1 - f \geq 1/n)$. Hence, $E(f \neq 1) \leq \sum_{n=1}^{\infty} E(1 - f \geq 1/n) = 0$. \square

Theorem 2. Let $C_1(Ex_1, En_1, He_1)$ and $C_2(Ex_2, En_2, He_2)$ be two cloud models. $UDCM(C_1, C_2) = 1$ if and only if $Ex_1 = Ex_2$, $En_1 = En_2$, and $He_1 = He_2$.

Proof. Obviously, sufficiency can be proved by Theorem 1. Necessary: in Definition 2, $UDCM(C_1, C_2) = 1/|L| \int_L S(U_1(x), U_2(x))dx$. Because of Lemma 1, we know that $S(U_1(x), U_2(x)) = 1$ holds on L , almost everywhere. It induces that, for any $x \in L$ and $y \in \mathbb{R}^+$, the equation

$$\exp\left\{-\frac{(y - En_1)^2}{2He_1^2}\right\} = \exp\left\{-\frac{(y - (|x - Ex_1/x - Ex_2|)En_2)^2}{2(|x - Ex_1/x - Ex_2|)He_2^2}\right\}, \quad (18)$$

is satisfied. As we know, function $f(x) = \exp\{-x^2/2\sigma^2\}$ is determined uniquely by parameters μ and σ . Hence, $Ex_1 = Ex_2$, $En_1 = En_2$, and $He_1 = He_2$ hold. \square

4. Experimental Analysis

4.1. Comparison with Other SMCM. In order to demonstrate high discriminability of UDCM clearly as Theorem 2 claimed, we calculate similarity among four cloud models, $C_1(10, 5, 1)$, $C_2(5, 2.5, 0.5)$, $C_3(10, 5, 0.4)$, and $C_4(10, 2, 2)$, using different similarity measures. It is obvious that C_1, C_2, C_3 , and C_4 are four different CMs. Table 4 shows the similarity values of different SMCM. For LICM, due to proportional numerical characteristics of two cloud models, the angle cosine of two vectors composed by numerical characteristics of C_1 and C_2 is equal to 1. It only distinguishes the difference in the shape of CM, which is unrelated to expectation. C_1 and C_3 have the same expectation and entropy. Their difference cannot be found using ECM. Analogously, the difference between C_1 and C_4 cannot be found using MCM because they have the same outer envelope. UDCM can distinguish the difference between any two different CMs from the perspective of uncertain distribution. It equips high discriminability which is significant to artificial intelligence domains. We will analyze its merits and demonstrate its performance in experiments as the following.

4.2. Classification of Time Series. An appropriate SMCM is important for the time series classification. BCT reserves uncertain features in processing reduction of the dimensions. CM has been applied in many relative domains of time series [39]. Besides, similarity measure is the other critical factor to classifying after dimension reduction. In order to test the performance, we conduct the experiment to compare these measures based on the standard dataset. We download the synthetic

TABLE 4: Similarity values of LICM, ECM, MCM, and UDCM.

| | LICM | | | | ECM | | | | MCM | | | | UDCM | | | |
|-------|--------|---------------|--------|--------|--------|--------|---------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | C_1 | C_2 | C_3 | C_4 | C_1 | C_2 | C_3 | C_4 | C_1 | C_2 | C_3 | C_4 | C_1 | C_2 | C_3 | C_4 |
| C_1 | 1.0000 | 1.0000 | 0.9986 | 0.9601 | 1.0000 | 0.4898 | 1.0000 | 0.5714 | 1.0000 | 0.6065 | 0.8732 | 1.0000 | 1.0000 | 0.1634 | 0.4000 | 0.2149 |
| C_2 | — | 1.0000 | 0.9986 | 0.9601 | — | 1.0000 | 0.4898 | 0.2665 | — | 1.0000 | 0.6190 | 0.6065 | — | 1.0000 | 0.0851 | 0.1230 |
| C_3 | — | — | 1.0000 | 0.9530 | — | — | 1.0000 | 0.5714 | — | — | 1.0000 | 0.8732 | — | — | 1.0000 | 0.1091 |
| C_4 | — | — | — | 1.0000 | — | — | — | 1.0000 | — | — | — | 1.0000 | — | — | — | 1.0000 |

control chart dataset from the machine learning data repository, University of California at Irvine [40]. There are 600 time series and 6 classes in this dataset. We randomly divide the dataset into six equal portions and successively adopt cross-validation based on these portions. For each database, the training set contains 500 data, and the testing set contains 100 data. We employ misclassification rate to evaluate performances of LICM, ECM, MCM, and UDCM as follows:

$$\text{error rate} = \frac{1}{N} \sum_{i=1}^N \mathbb{1}(f_i \neq L_i). \quad (19)$$

As shown in Figure 6, the performance of UDCM surpasses other methods. LICM only captures tendency information, and the loss of distribution information results in high misclassification rate. Compared with ECM and MCM, UDCM utilizes more accurate uncertainty to measure similarity and can capture the second-order uncertain relationship to describe similarity more appropriately.

4.3. *Shooting Experiment.* In order to verify similarity results in accordance with the uncertainty distribution of concepts, we measure the similarity of four shooters' performance. We suppose cloud models of four shooters' performance $A(1, 2, 1)$, $B(3.2, 6.2, 3.2)$, $C(1.2, 2.1, 10)$, and $D(2, 3, 1)$. Cloud drops represent the deviation of hitting from bull's eye. Results of shooting are denoted as the 10-point system (off-target denoted as 0 point). Table 5 exhibits the score statistics of 100 times shooting simulated by FCT with respect to each shooter. The results of shooting and histograms of score statistics are shown in Figure 7. Similarities of score distributions are calculated by Jensen–Shannon divergence as

$$\text{JSD}(P(x), Q(x)) = \frac{1}{2} \sum_x \left(P(x) \log \frac{P(x)}{M(x)} + Q(x) \log \frac{Q(x)}{M(x)} \right), \quad (20)$$

where $M = 1/2(P(x) + Q(x))$. For $X, Y \in U = \{A, B, C, D\}$, similarities are defined as

$$S(X, Y) = 1 - \frac{\text{JSD}(X, Y)}{\max_{N, M \in U} \text{JSD}(N, M)} \quad (21)$$

Table 6 shows the similarity between A and other shooters by different similarity measures. After 100 times shooting, the sort of uncertain distribution similarities is $S(A, D) > S(A, C) > S(A, B)$. In LICM, the sort is $S(A, B) > S(A, D) > S(A, C)$. In ECM, the sort is $S(A, C) > S(A, D) > S(A, B)$. Compared with other SMCM,

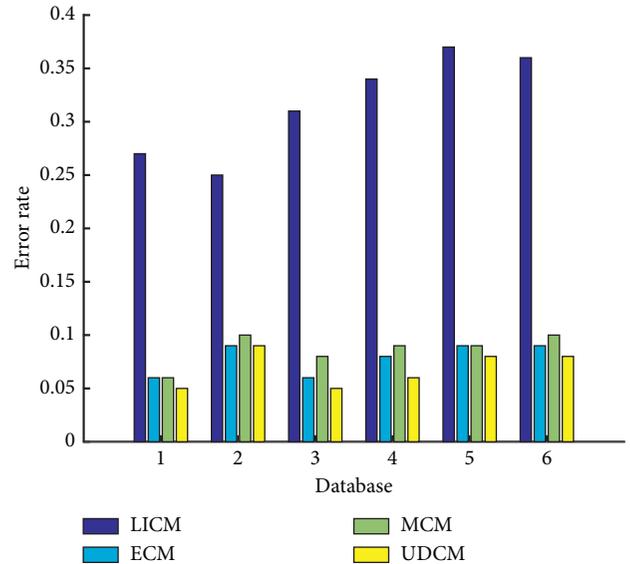


FIGURE 6: Comparison among SMCM.

the similarity by UDCM is in accordance with the similarity of shooting result distribution. It can be seen that UDCM takes into account more comprehensive similarity than other conceptual intension-based SMCM.

4.4. *Application in Multicriteria Group Decision-Making.* In multicriteria group decision-making (MCDM), the linguistic variable is a good choice to express personal sense. There are various linguistic variables applied widely in different fields, e.g., 2-tuple linguistic model [41, 42], probabilistic uncertain linguistic model [43, 44], intuitionistic fuzzy linguistic model [45], and hesitant fuzzy linguistic model [46, 47]. All of the aforementioned linguistic models can only describe the fuzziness but not randomness. Wang et al. [22] proposed conversion between the linguistic variable and the cloud model, which has the ability to describe fuzziness and randomness in human cognition. Based on their work, other researchers utilized cloud model in group decision-making [27, 48, 49]. As shown in Example 1 and 2, we have to cost more running time to acquire correct results. In this section, we will demonstrate the merit of UDCM in MCDM.

Example 3 (continued from Example 2). As shown in Figure 8, the similarity between C_1 and C_2 is calculated by FDCM with different times of Monte Carlo simulation. The results shock around 1 with n increasing and exceed 1 sometimes. Therefore, FDCM is not a real similarity measure of cloud model. Figure 9

TABLE 5: Score statistics of 100 times shooting simulated by FCT.

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|----|---|---|----|----|----|----|----|----|----|----|
| A | 0 | 0 | 2 | 2 | 4 | 3 | 12 | 17 | 19 | 29 | 12 |
| B | 42 | 6 | 7 | 5 | 9 | 9 | 6 | 5 | 5 | 5 | 1 |
| C | 37 | 2 | 5 | 4 | 2 | 9 | 9 | 13 | 10 | 9 | 0 |
| D | 2 | 7 | 4 | 10 | 11 | 11 | 11 | 18 | 10 | 10 | 6 |

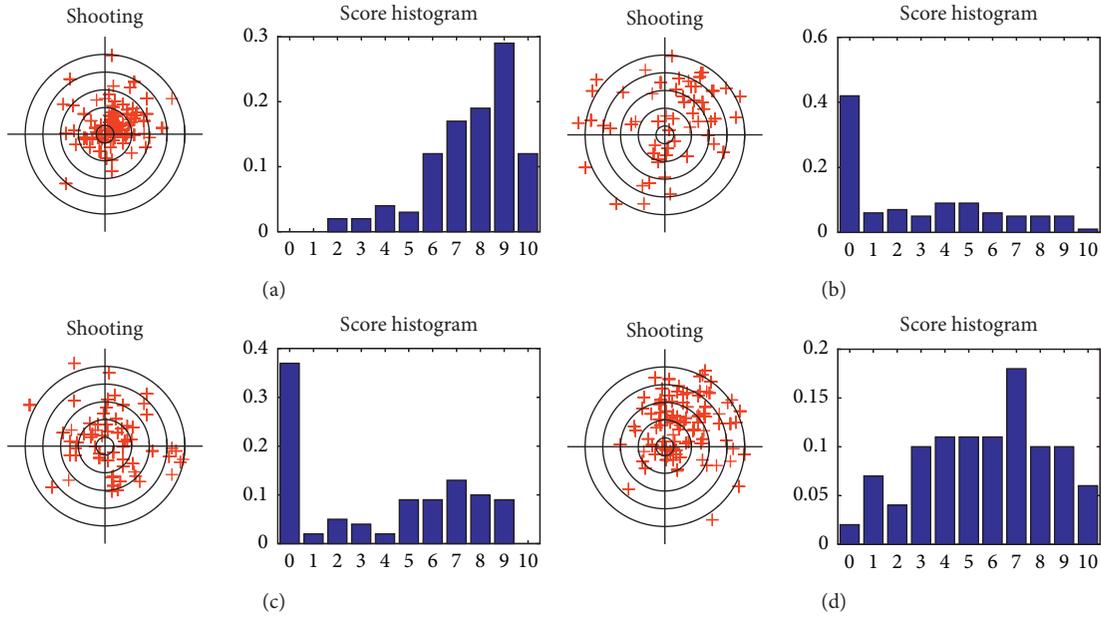


FIGURE 7: Results of shooting and histograms of score statistics.

TABLE 6: Comparison among different similarity measures.

| Similarity measure | $S(A, B)$ | $S(A, C)$ | $S(A, D)$ |
|------------------------|-----------------------|---------------|---------------|
| Extension distribution | 2.2×10^{-16} | 0.2586 | 0.6606 |
| LICM | 0.9999 | 0.6111 | 0.9820 |
| ECM | 0.4809 | 0.9600 | 0.7780 |
| UDCM | 0.1614 | 0.1714 | 0.3862 |

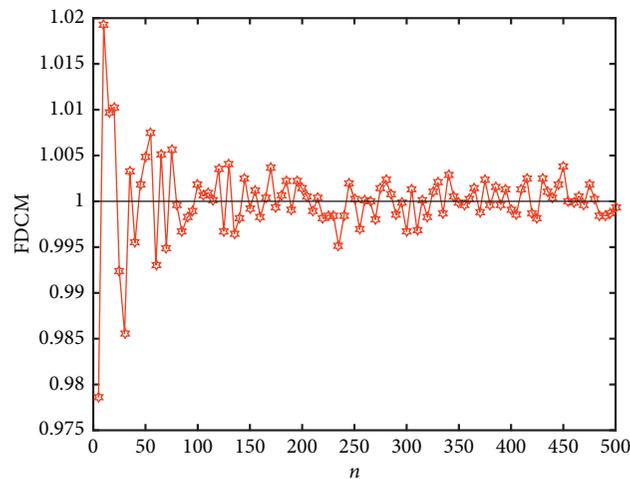


FIGURE 8: FDCM with variation of n .

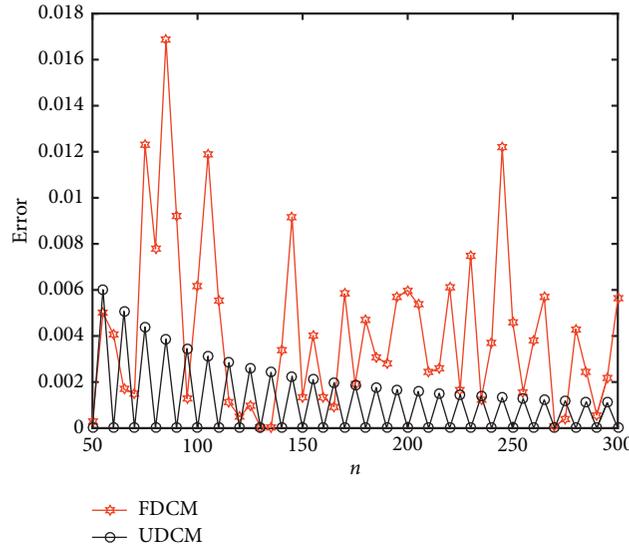


FIGURE 9: Stability of FDCM and UDCM.

TABLE 7: Comparison of FDCM and UDCM in MCDM.

| Method | Final group decision matrix | | | Overall score of different n | | | | Ranking | Time cost |
|--------|-----------------------------|---------------------|---------------------|--------------------------------|----------------|----------------|----------------|-------------------|-----------|
| | | | | n | $\hat{s}(x_2)$ | $\hat{s}(x_2)$ | $\hat{s}(x_2)$ | | |
| FDCM | (4.864 0.352 0.024) | (3.922 0.369 0.025) | (4.536 0.343 0.023) | 5000 | 3.152 | 3.782 | 4.272 | $x_2 > x_2 > x_2$ | 103.117s |
| | (5.420 0.386 0.026) | (5.465 0.344 0.023) | (4.165 0.333 0.022) | 50000 | 3.169 | 3.780 | 4.289 | | |
| | (4.725 0.373 0.025) | (6.295 0.378 0.025) | (5.519 0.352 0.023) | 500000 | 3.175 | 3.782 | 4.285 | | |
| UDCM | (5.582 0.356 0.024) | (3.437 0.380 0.025) | (5.469 0.347 0.023) | 5000 | 3.294 | 3.488 | 3.888 | $x_2 > x_2 > x_2$ | 17.423s |
| | (4.411 0.381 0.025) | (5.073 0.353 0.024) | (4.238 0.312 0.021) | 50000 | 3.286 | 3.497 | 3.862 | | |
| | (4.045 0.366 0.025) | (5.925 0.361 0.024) | (4.693 0.323 0.021) | 500000 | 3.286 | 3.497 | 3.871 | | |

shows the convergence of FDCM and UDCM. UDCM is tending towards stability with n increasing, while it holds on for FDCM.

To acquire a more accurate similarity, Wang et al. [27] claimed that FDCM has to execute no less than 50,000 Monte Carlo simulations to obtain a stable estimated overall score \hat{s} . In the remainder, we calculate the stable FDCM with $n = 2000$. To verify our method validity in MCDM, we make a decision for the application example in [27] by UDCM. The experiment is run on a personal computer with Windows 10 and Inter (R) Core (TM) i7-7700 CPU 3.6 GHz, and DDR3, 16 GB memory. Matlab R2015b software is used. Let the consensus threshold $\delta = 0.9$; we make the decision using FDCM and UDCM, respectively, in MCDM. Their final group decision matrices, ranking result, and cost time are listed in Table 7, where the parameter n is the number of times for executing the Monte Carlo simulation to obtain the overall score of the cloud model. Although the decision results are consistent, UDCM only takes one-sixth of the time by FDCM. Hence, UDCM is an efficient method for FDCM.

5. Conclusion

Similarity of concepts is a fundamental study in uncertain artificial intelligence. By utilizing FCT and BCT,

bidirectional cognitive transformation between intension and extension of a concept is realized by the CM. Furthermore, CM reflects the uncertainty of qualitative concept itself and, meanwhile, reveals the objective relationship between probability and fuzziness in the uncertain concept. As a significant expression, distribution is always utilized to describe the uncertain phenomenon. Based on this, we propose a new similarity measure UDCM and introduce its calculation in detail. Due to employing complete uncertainty, it acquires similarity results in accordance with the uncertain distribution and then gives some valuable consultations in synthetic evaluation. Besides, UDCM has merits of discriminability and stability and is an effective tool for cognitive computing. Finally, UDCM is also a framework of SMCM. Employing different forms of second-order uncertainty will result in different results. In the future, selection of uncertain forms for different situations also deserves to be studied.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by Innovation and Exploration Project of Guizhou Province (QKHPTRC [2017]5727-06), PhD Initiation Fund of Zunyi Normal University (ZSBS [2019]04), and PhD Training Program of Chongqing University of Posts and Telecommunication (no. BYJS201902).

References

- [1] A. Budanitsky and G. Hirst, "Evaluating wordnet-based measures of lexical semantic relatedness," *Computational Linguistics*, vol. 32, no. 1, pp. 13–47, 2006.
- [2] H. A. Nguyen and H. Al-Mubaid, "New ontology-based semantic similarity measure for the biomedical domain," in *Proceedings of the 2006 IEEE International Conference on Granular Computing*, pp. 623–628, Atlanta, USA, 2006.
- [3] K. M. Sim and P. T. Wong, "Web-based information retrieval using agent and ontology," in *Proceedings of the Asia-Pacific Conference on Web Intelligence*, pp. 384–388, Maebashi, Japan, 2001.
- [4] V. Rus, N. Niraula, and R. Banjade, "Similarity measures based on latent dirichlet allocation," in *Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics*, pp. 459–470, Karlovasi, Samos, Greece, 2013.
- [5] R. Laza, R. Pavn, M. Reboiro-Jato, and F. Fdez-Riverola, "Assessing the suitability of mesh ontology for classifying medline documents," in *Proceedings of the 5th International Conference on Practical Applications of Computational Biology & Bioinformatics (PACBB 2011)*, pp. 337–344, Salamanca, Spain, 2011.
- [6] G. Chen and Y. Xu, "Special issue on dealing with uncertainty in data mining and information extraction," *Information Sciences*, vol. 173, no. 4, pp. 277–279, 2005.
- [7] M. Hsu, M. Bhatt, R. Adolphs, D. Tranel, and C. F. Camerer, "Neural systems responding to degrees of uncertainty in human decision-making," *Science*, vol. 310, no. 5754, pp. 1680–1683, 2005.
- [8] J. Montero and D. Ruan, "Modelling uncertainty," *Information Sciences*, vol. 180, no. 6, pp. 799–802, 2010.
- [9] F. Sabahi and M.-R. Akbarzadeh-T, "A qualified description of extended fuzzy logic," *Information Sciences*, vol. 244, pp. 60–74, 2013.
- [10] R. R. Yager, "Modeling uncertainty using partial information," *Information Sciences*, vol. 121, no. 3-4, pp. 271–294, 1999.
- [11] D. Y. Li, C. Y. Liu, Y. Du, and X. Han, "Artificial intelligence with uncertainty," *Journal of Software*, vol. 15, no. 11, pp. 1583–1594, 2004.
- [12] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.
- [13] Z. a. Pawlak, "Rough sets," *International Journal of Computer & Information Sciences*, vol. 11, no. 5, pp. 341–356, 1982.
- [14] D. Y. Li and Y. Du, *Artificial Intelligence with Uncertainty*, IEEE, Piscataway, NJ, USA, 2017.
- [15] W. Li, Y. Wang, J. Du, and J. Lai, "Synergistic integration of graph-cut and cloud model strategies for image segmentation," *Neuro computing*, vol. 257, pp. 37–46, 2017.
- [16] C.-L. Xu and G.-Y. Wang, "A novel cognitive transformation algorithm based on Gaussian cloud model and its application in image segmentation," *Numerical Algorithms*, vol. 76, no. 4, pp. 1039–1070, 2017.
- [17] W. Li, F. Li, and J. Du, "A level set image segmentation method based on a cloud model as the priori contour," *Signal, Image and Video Processing*, vol. 13, no. 1, pp. 103–110, 2018.
- [18] B. Vahid, A. Jahangiri, and S. G. Machiani, "Multi-class us traffic signs 3d recognition and localization via image-based point cloud model using color candidate extraction and texture-based recognition," *Advanced Engineering Informatics*, vol. 32, pp. 263–274, 2017.
- [19] G. W. Zhang, D. Y. Li, and P. Li, "A collaborative filtering recommendation algorithm based on cloud model," *Journal of Software*, vol. 18, no. 10, pp. 2403–2411, 2007.
- [20] Y. P. Xiao, H. C. Sun, and T. J. Dai, "A rating prediction method based on cloud model in social recommendation system," *ACTA Electronica Sinica*, vol. 46, no. 7, pp. 1762–1776, 2018.
- [21] J. G. Sun and L. R. Ai, "Collaborative filtering recommendation algorithm based on item attribute and cloud model filling," *Journal of Computer Applications*, vol. 32, no. 3, pp. 658–660, 2012.
- [22] J.-q. Wang, L. Peng, H.-y. Zhang, and X.-h. Chen, "Method of multi-criteria group decision-making based on cloud aggregation operators with linguistic information," *Information Sciences*, vol. 274, pp. 177–191, 2014.
- [23] J.-q. Wang, J.-j. Peng, H.-y. Zhang, T. Liu, and X.-h. Chen, "An uncertain linguistic multi-criteria group decision-making method based on a cloud model," *Group Decision and Negotiation*, vol. 24, no. 1, pp. 171–192, 2015.
- [24] H.-y. Zhang, P. Ji, J.-q. Wang, and X.-h. Chen, "A neutrosophic normal cloud and its application in decision-making," *Cognitive Computation*, vol. 8, no. 4, pp. 649–669, 2016.
- [25] H. B. Gao, X. Y. Zhang, and T. L. Zhang, "Research of intelligent vehicle variable granularity evaluation based on cloud model," *ACTA Electronica Sinica*, vol. 44, no. 2, pp. 365–373, 2016.
- [26] A. Wu, Z. Ma, and G. Zeng, "Set pair fuzzy decision method based on cloud model," *Chinese Journal of Electronics*, vol. 25, no. 2, pp. 215–219, 2016.
- [27] P. Wang, X. Xu, S. Huang, and C. Cai, "A linguistic large group decision making method based on the cloud model," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 6, pp. 3314–3326, 2018.
- [28] D. Li, C. Liu, and W. Gan, "A new cognitive model: cloud model," *International Journal of Intelligent Systems*, vol. 24, no. 3, pp. 357–375, 2009.
- [29] C. Liu and D. Y. Li, "Study on the universality of the normal cloud model," *Engineering Science*, vol. 6, no. 8, pp. 28–34, 2004.
- [30] D. Y. Li, H. J. Meng, and X. M. Shi, "Membership clouds and membership cloud generators," *Journal of Computer Research and Development*, vol. 6, pp. 15–20, 1995.
- [31] X. Yang, L. Yan, and L. Zeng, "How to handle uncertainties in ahp: the cloud delphi hierarchical analysis," *Information Sciences*, vol. 222, pp. 384–404, 2013.
- [32] G. Wang, C. Xu, and D. Li, "Generic normal cloud model," *Information Sciences*, vol. 280, pp. 1–15, 2014.
- [33] J.-Q. Wang, P. Wang, J. Wang, H.-Y. Zhang, and X.-H. Chen, "Atanassov's interval-valued intuitionistic linguistic multi-criteria group decision-making method based on the

- trapezium cloud model,” *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 3, pp. 542–554, 2015.
- [34] Y. Zhang, D. N. Zhao, and D. Y. Li, “The similar cloud and measurement method,” *Information and Control*, vol. 33, no. 2, pp. 129–132, 2004.
- [35] H. Li, C. H. Guo, and W. R. Qiu, “Similarity measurement between normal cloud models,” *ACTA Electronica Sinica*, vol. 39, no. 11, pp. 2561–2567, 2011.
- [36] C. B. Cai, W. Fang, and J. Zhao, “Research of interval-based cloud similarity comparison algorithm,” *Journal of Chinese Computer System*, vol. 32, no. 12, pp. 2456–2460, 2011.
- [37] J. Yang, G. Y. Wang, Q. H. Zhang, and L. Feng, “Similarity measure of multi-granularity cloud model,” *Pattern Recognition and Artificial Intelligence*, vol. 31, no. 8, pp. 677–692, 2018.
- [38] X. Zha and S. H. Ni, “Indirect computation approach of cloud model similarity based on conception skipping,” *System Engineering and Electronics*, vol. 37, no. 7, pp. 1676–1682, 2015.
- [39] W. Deng, G. Wang, and J. Xu, “Piecewise two-dimensional normal cloud representation for time-series data mining,” *Information Sciences*, vol. 374, pp. 32–50, 2016.
- [40] UCI Machine Learning Repository, 2020, <https://archive.ics.uci.edu/ml/datasets>.
- [41] T. He, G. Wei, C. Wei, and J. Wang, *Codas Method for Pythagorean 2-tuple Linguistic Multiple Attribute Group Decision Making*, IEEE Access, Piscataway, NJ, USA, 2019.
- [42] Y. Liang, J. Liu, J. Qin, and Y. Tu, “An improved multi-granularity interval 2-tuple todim approach and its application to green supplier selection,” *International Journal of Fuzzy Systems*, vol. 21, no. 1, pp. 129–144, 2019.
- [43] M. Lin, Z. Xu, Y. Zhai, and Z. Yao, “Multi-attribute group decision-making under probabilistic uncertain linguistic environment,” *Journal of the Operational Research Society*, vol. 69, no. 2, pp. 157–170, 2018.
- [44] Z. Wu, J. Xu, X. Jiang, and L. Zhong, “Two magdm models based on hesitant fuzzy linguistic term sets with possibility distributions: vikor and topsis,” *Information Sciences*, vol. 473, pp. 101–120, 2019.
- [45] L. Zhang, J. Zhan, and Z. Xu, “Covering-based generalized if rough sets with applications to multi-attribute decision-making,” *Information Sciences*, vol. 478, pp. 275–302, 2019.
- [46] L. Wang, H. Wang, Z. Xu, and Z. Ren, “The interval-valued hesitant Pythagorean fuzzy set and its applications with extended TOPSIS and Choquet integral-based method,” *International Journal of Intelligent Systems*, vol. 34, no. 6, pp. 1063–1085, 2019.
- [47] J. Montserrat-Adell, Z. Xu, X. Gou, and N. Agell, “Free double hierarchy hesitant fuzzy linguistic term sets: an application on ranking alternatives in gdm,” *Information Fusion*, vol. 47, pp. 45–59, 2019.
- [48] H. C. Liu, L. E. Wang, and Z. W. Li, “Improving risk evaluation in fmea with cloud model and hierarchical topsis method,” *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 1, pp. 84–95, 2018.
- [49] J. Zhou, C. Han, L. Sun, and F. Xiao, “Linguistic multi-criteria group decision-making method combining cloud model and evidence theory,” *IEICE TRANSACTIONS on Information and Systems*, vol. E102.D, no. 4, pp. 845–855, 2019.