International Journal of Approximate Reasoning 164 (2024) 109081

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

A novel information fusion method using improved entropy measure in multi-source incomplete interval-valued datasets

Weihua Xu [a,*], Ke Cai [a], Debby D. Wang [b]

[a] *College of Artificial Intelligence, Southwest University, Chongqing, 400715, PR China*
[b] *School of Science and Technology, Hong Kong Metropolitan University, Ho Man Tin, Kowloon, Hong Kong*

ARTICLE INFO

Keywords:
Entropy theory
Incomplete interval-valued
Multi-source information fusion
Granular computation

ABSTRACT

Multi-source data is a comprehensive data type that combines multiple sources of information or datasets. Compared to point-valued data, interval-valued data provides a more accurate representation of the uncertainty and variability associated with objects. In practical situations, data obtained from multiple sources may contain missing values for various reasons. Therefore, it is essential to develop multi-source information fusion technology in order to achieve information fusion or information extraction from multi-source incomplete data. This paper aims to explore the information fusion problem of multi-source incomplete interval-valued datasets. The primary contributions of this study involve utilizing the principle of statistical distribution and KL divergence to establish a metric for measuring the similarity between intervals. Firstly, this approach helps to reduce the problem of disregarding internal information within interval values, which can result in the loss of valuable information. Secondly, we establish an interval fuzzy similarity relation based on the mentioned concept of similarity among interval values. Moreover, we investigate the uncertainty measurement of incomplete interval-valued decision datasets and design an emerging information entropy fusion method. Finally, we comprehensively evaluate the effectiveness of the proposed method. Experimental results indicate that the proposed approach has advantage over the maximum, minimum, mean, and information entropy fusion method based on tolerance relationship. In addition, the distance metric used in this article can improve the fusion classification effect compared to several common interval-valued distance measures.

1. Introduction

In real life, data often carries a great deal of uncertainty. As an important research topic in terms of big data analysis, multi-source information fusion is a comprehensive processing process of multi-level and multi-aspect data or information to realize automatic detection, correlation, connection and evaluation of multi-source data, and obtain more essential and more precise information than a single information source [1]. At sometime, information fusion technology can integrate various information with spatial redundancy, time redundancy or complementary information constraints through specific rules to build a unified representation. This helps to reduce the ambiguity and uncertainty of the data and enhances characterization of the information [2,3]. Multi-source information fusion was successfully employed in military fields, such as military command automation system, multi-target tracking and recognition, precision guided weapons, remote sensing monitoring, wireless communication, medical diagnosis [4–7] and so on in civil fields.

* Corresponding author.
E-mail addresses: chxuwh@gmail.com (W. Xu), 2279732003@qq.com (K. Cai), dwang@hkmu.edu.hk (D.D. Wang).

https://doi.org/10.1016/j.ijar.2023.109081
Received 6 June 2023; Received in revised form 11 October 2023; Accepted 1 November 2023

Available online 10 November 2023

Information structure and uncertainty measurement are two important research directions in terms of granular computing. According to the equivalence or similarity relationship between samples, samples can be divided into corresponding equivalence classes or information granules. The set of all information granules constitutes an information structure within the system [8], and information structure is an essential method for studying information systems. Rough set theory, proposed by Pawlak [9], serves as an effective mathematical tool for coping with uncertainty, and has been successfully applied to fields of intelligent information processing, including decision analysis [11,12], machine learning [10,13,14], approximate inference [15], pattern identification and data mining [16–19]. From the perspective of data analysis, rough set theory has numerous advantages. Given these advantages, many scholars have combined it with multi-source information fusion and achieved excellent results. For example, Dong et al. [20] considered information fusion processing based on rough set theory. Wang et al. [21] studied information fusion based on multi-source sensors. Chen et al. [22] put forward six new kinds of double-quantitative multigranulation rough fuzzy set models in multi-source decision systems. Lin et al. [23] investigated an information fusion method based on a combination of multigranulation rough sets and evidence theory. Zhang et al. [24] researched a data-level fusion model, which uses the neighborhood rough set model to build up the domain granular structure, and uses the idea of granularity calculation to establish an uncertainty measurement methods. Yang et al. [25] introduced a novel fusion approach based multi-granulation approach which can reduce information loss during convergence. Uncertainty measure, as a robust evaluation tool, can effectively describe the uncertainty of decision system or datasets. Researchers can study the methods of uncertainty in datasets from different views to discuss the uncertainty of information system [30,31]. C.E. Shannon [32] proposed that information entropy is an uncertainty degree that is extensively applied in information fusion; In [26], Pawlak provided four uncertainty measures, which are respectively used to evaluate the accuracy and roughness of approximate classification in decision system. Qian et al. [27] offered a fuzzy information granularity measure based on a binary particle size structure. Wang et al. [28] presented uncertainty measure based on general fuzzy relationships for information systems. Yao Sheng et al. [29] introduced a method that combines tolerance information entropy and mixed approximate roughness to measure the information system indefinitely. Xu et al. [33] used conditional entropy to measure the importance of a source. However, in reality, there is a lost amount of scoped data, and isn't easy to analyze and mine the critical knowledge in the data.

It is worthy nothing that the above uncertainty measurement methods are based on symbolic data or fuzzy data in all single-valued information systems [34–36]. However, there are many types of data in practical situation, and it is difficult to analyze and mine essential knowledge from the data, due to limitations in both objective environment and people's subjective understanding, many things can't be accurately expressed, and they are described in the form of interval value pairs. Interval values are expressed as upper bound and lower bound, which is an uncertain representation of accurate data, mainly to preserve the information of things to the greatest extent, there are a lot of applications in real life. Compared with single-valued data, interval-valued data can effectively describe the randomness and uncertainty of information. For instance, the temperature sampling results over a period of time are often presented in the form of interval values, because the data at a certain moment is meaningless, and interval values can well reflect the temperature range in a period of time. Accurate values can also be expressed in the form of interval values, which reflects the accuracy of characterization. Thus, it is of great practical significance to measure the uncertainty of interval data. Many scholars have studied interval-valued data. Zhang et al. [37] studied the up-down approximation operator of information systems. Qian et al. [38] introduced the rough set method of dominant relationships to study interval-valued information systems. Liu et al. [39] put forward α-approximation equivalence relations to decrease the unsupervised attributes of interval-valued information systems. Xie et al. [40] investigated a definition of the probability similarity between interval-valued to measure the uncertainty of the interval-valued information system. Yager [41] studied a monotone set measures for multi-source mixed data. Xu et al. [42] used the DS evidence theory to offer a fault diagnosis method that fuses different diagnostic evidence with interval-valued data. Huang et al. [43] designed a new fusion model based on fuzzy information granulation, which converts multivariate interval-valued data into trapezoidal fuzzy numbers. Xu et al. [44] raised four incremental fusion mechanisms for dynamic interval-valued ordered data.

It can be seen that the above research is the knowledge discovery and rule extraction of complete information. However, there are much incomplete data in reality due to the aging of sensor failure led to data loss, measurement failure and storage loss. Missing of values in an information system may lead to the loss of a significant amount of useful data. So, it is essential to grasp the reliability of the whole data from these systems. In the last few years, lots of scholars have also studied incomplete information systems, Dai et al. [45] constructed similarity relationship on IIVIS, and presented uncertainty measures based on α-weak similarity. Luo et al. [46] proposed an incremental feature selection model from information-theoretic angle for dynamic defective data. Zhao et al. [47] established the feature selection of incomplete decision information system by using a new extended rough set model. Li et al. [48] investigated a novel interval set model for knowledge acquisition to summarize incomplete data. Han et al. [49] defined a generalized information entropy based on interval-valued similarity relation to fill incomplete information. However, most of the above research focus on single-source incomplete information systems, and it is clear that these methods cannot be straightly employed to the fusion of multi-source data. Some scholars have also studied the MsIIvIS. For example, Zhang et al. [50] studied the dynamic fusion mechanism of simultaneous changes of information sources and attributes for incomplete interval-valued data tables. Xu et al. [51] developed an information fusion model based on information entropy. There are few researchers on the fusion of incomplete interval-valued data at present. In addition, the distance measure is only considered the endpoint information of interval-valued. And the contribution of the inner interval to the information is ignored [44,50], which causes the loss of the practical information of the interval. Based on above analysis, it is necessary to introduce a new information fusion model. First of all, we treat the interval as a probability distribution and use the Kullback-Leibler divergence to construct the distance measurement of the incomplete interval-valued information systems. And the similarity is constructed using this distance. Furthermore, an interval similarity relationship is created according to the similarity, and the concept of information granularity is proposed based mentioned relation. Finally, from the perspective of the information granular structure, the infimum fusion function that we construct according

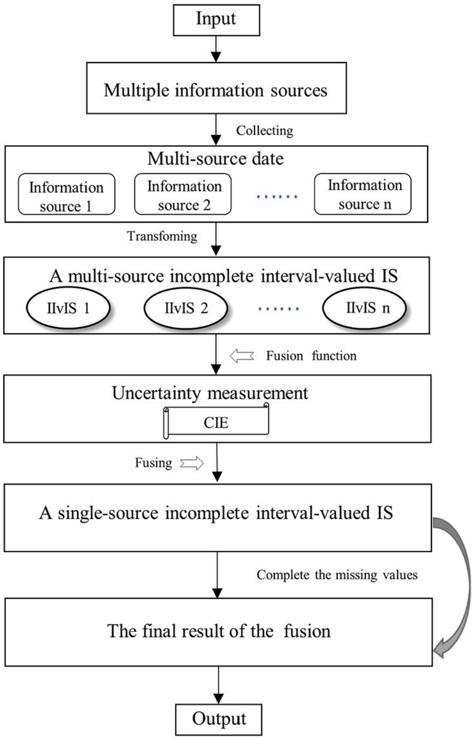


Fig. 1. The framework of this paper.

to the uncertainty measurements, and the multi-source information system is fused. The framework of this paper is shown in Fig. 1. Contributions of this work are summarized as below:

(1) To address the issue of losing valuable information due to the neglect of information within interval values, we employed the principle of statistical distribution. We transformed interval values into a probability distribution form and used KL divergence to quantify the dissimilarity between two interval distributions.
(2) Based on the distance, we defined a new similarity and present a new similarity relation. We established fuzzy similarity relations based on the distance measurement. According to the relation, we introduced the concept of information structure and established the entropy measurements.
(3) We constructed infimum fusion function based on the significantly measurements. Furthermore, experimental results have demonstrated that our proposed method enhances fusion performance.

The rest of this work consists of the following contents: Section 2 gives the basic definitions of distance measurement, similarity, and reviews the concepts of information tables. Section 3 offers uncertainty measures of multi-source incomplete interval-valued decision systems. An infimum-measure fusion function is provided in section 4. Section 5 analyzes the results of the experiment and the effectiveness and efficiency of fusion. Finally, in section 6 summarizes the work of this paper and future research.

2. Preliminary

In this section, we will formally review some mathematical concepts and definitions employed in this work.

2.1. Similarity between interval values

Many difference measurements of interval values are associated with distance. Numerous scholars have done a great deal of work to measure the difference of interval values. So far to, there are some familiar distances that can be used to reflect diverseness in interval values. For example:
City-block distance,

$$D_c = |y^- - x^-| + |y^+ - x^+|.$$

Euclid distance,

$$D_E = |y^- - x^-|^2 + |y^+ - x^+|^2.$$

Hausdorff distance etc.,

$$D_H = \max(|y^- - x^-|, |y^+ - x^+|),$$

where $x = [x^-, x^+]$ and $x = [y^-, y^+]$ are two interval values, where x^-, y^- is left endpoint and x^+, x^+ is right endpoint.

However, the above distance measure mentioned above only is considered from the endpoints of interval value. The internal contribution to information of interval-valued endpoints is ignored, leading to the loss of effective information with in interval-valued. Therefore, this paper uses the probability distribution principle to characterize an interval as a probability distribution, and utilizes Kullback-Leibler divergence to calculate the distance between distributions.

Definition 1. For continuous random variables, the Kullback-Leibler divergence of the two probability distributions P and Q is defined in the integral form as follows:

$$KL(P\|Q) = \int P(x)\ln\left(\frac{P(x)}{Q(x)}\right)dx, \tag{1}$$

where $P(x)$ and $Q(x)$ is the probability density function of P, Q. In the field of probability statistics, Kullback-Leibler divergence can be employed to measure the distance between two probability distributions.

Definition 2. Let $x_i = [x_i{}^-, x_i{}^+], y_j = [y_j{}^-, y_j{}^+]$ be two interval values and $a \in A$, where $x_i, y_j \in U$. The novel distance measurement between x_i and y_j is defined as follows:

$$x_i \sim N(\mu_1, \sigma_1^2) \quad , \quad y_j \sim N(\mu_2, \sigma_2^2), \tag{2}$$

where

$$\mu_1 = \frac{x_i{}^+ + x_i{}^-}{2} \quad , \quad \mu_2 = \frac{y_j{}^+ + y_j{}^-}{2}, \tag{3}$$

$$\sigma_1 = \frac{x_i{}^+ - x_i{}^-}{2} \quad , \quad \sigma_2 = \frac{y_j{}^+ - y_j{}^-}{2}. \tag{4}$$

Based on KL divergence, the interval-valued distance between x_i and y_j is defined as follows:

$$d_a(x_i, y_j) = \sqrt{dis_a(x_i, y_j)}, \tag{5}$$

where

$$dis_a(x_i, y_j) = \frac{KL_a(x_i\|y_j) + KL_a(y_j\|x_i)}{2}. \tag{6}$$

Let $x_i = [x_i{}^-, x_i{}^+], y_j = [y_j{}^-, y_j{}^+]$ be two interval values and $a \in A$, where $x_i{}^- < x_i{}^+, y_j{}^- < y_j{}^+$. The similarity of x_i and x_j can be defined as below:

$$Sim_a(x_i, y_j) = \frac{1}{1 + d_a(x_i, y_j)}. \tag{7}$$

Obviously, similarity $Sim_a(x, y)$ satisfies both reflexivity and symmetry.

But in fact, when we collect a great amount of data, we may acquire missing data. Therefore, it is indispensable to give a novel distance measure for incomplete interval-valued information datasets.

Definition 3. In incomplete interval-valued information datasets, suppose $x_i =*$ or $y_j =*$, where $*$ is a missing value. The distance between two intervals can be expressed as below:
i.e.

$$d_a\left(x_i, y_j\right) = \begin{cases} 0 & x_i =* \ or \ y_j =* \\ \sqrt{dis_a\left(x_i, y_j\right)} & else, \end{cases} \tag{8}$$

$$Sim_a\left(x_i, y_j\right) = \begin{cases} 1 & x_i =* \ or \ y_j =* \\ \frac{1}{1+d_a\left(x_i, y_j\right)} & else. \end{cases} \tag{9}$$

2.2. Multi-source incomplete interval-valued information system

$IIvIS = (U, A, V, f)$ is an incomplete interval-valued information, where $U = \{x_1, x_2, ..., x_n\}$ represents a non-empty and finite object set, and $A = \{a_1, a_2, ..., a_p\}$ is non-empty and finite attribute set. Set V is called the range of attribute A. Function $f : U \times A \to V$ is information function, $\forall x \in U, a \in A, f(x, a) = [f^-(x, a), f^+(x, a)]$ or $f(x, a) =*$, where $*$ represents missing value).

Let $IIvIS_i = (U, A, V_i, f_i)$ be the i-th $IIvIS$, where the meanings of U, A, V_i and f_i as mentioned above. Generally, a multi-source incomplete interval-valued information system is defined as below:

Table 1
Physical examination report of the first hospital $IIvIS_1$.

U	a_1	a_2	a_3	a_4	a_5	a_6
x_1	[125.59, 136.29]	[1.00, 4.99]	[116.25, 125.97]	[65.98, 109.89]	[42.29, 187.98]	*
x_2	[124.58, 198.34]	[4.00, 11.00]	[114.96, 123.98]	[85.62, 121.99]	[82.60, 224.88]	*
x_3	[109.23, 120.02]	[2.89, 10.00]	*	[121.08, 176.99]	[67.00, 85.99]	[88.20, 96.73]
x_4	[125.19, 134.25]	*	[110.98, 120.45]	[48.93, 91.69]	[101.99, 260.88]	*
x_5	[119.69, 133.97]	[10.00, 20.00]	[110.45, 296.98]	*	[136.43, 279.25]	[45.99, 75.98]
x_6	[126.34, 215.07]	[8.40, 18.00]	*	[83.00, 163.98]	[68.29, 88.98]	[28.65, 62.98]
x_7	[117.97, 129.40]	[12.00, 21.00]	[169.34, 269.98]	[80.98, 156.99]	*	[32.69, 68.80]
x_8	[159.58, 232.69]	[6.30, 15.00]	[219.65, 314.68]	[103.25, 159.97]	*	[26.98, 66.42]

Table 2
Physical examination report of the second hospital $IIvIS_2$.

U	a_1	a_2	a_3	a_4	a_5	a_6
x_1	[125.60, 136.30]	[1.00, 5.00]	*	[65.99, 110.08]	*	[71.30, 96.08]
x_2	*	[4.00, 11.00]	[115.36, 124.00]	[86.26, 120.89]	[82.65, 224.90]	[69.25, 89.06]
x_3	[109.49, 120.00]	[3.00, 10.00]	[126.97, 185.34]	*	[67.03, 86.00]	[88.25, 96.79]
x_4	[125.29, 134.69]	[3.00, 8.00]	[111.00, 120.04]	[62.45, 97.99]	*	[68.43, 89.09]
x_5	*	[10.00, 20.00]	[110.40, 296.99]	[82.60, 121.92]	[136.45, 279.28]	[46.00, 75.99]
x_6	[125.99, 215.00]	[8.40, 18.00]	*	[83.02, 163.98]	[68.30, 89.00]	[28.68, 43.00]
x_7	[118.80, 129.68]	[12.00, 21.00]	[169.45, 269.99]	*	[79.65, 95.28]	[32.70, 68.82]
x_8	[159.98, 232.99]	[6.30, 15.00]	*	[103.30, 159.99]	[109.26, 260.39]	*

Table 3
Physical examination report of the third hospital $IIvIS_3$.

U	a_1	a_2	a_3	a_4	a_5	a_6
x_1	[200.15, 253.67]	[1.00, 5.00]	[116.30, 125.99]	[66.00, 109.92]	[42.29, 187.98]	[71.25, 96.10]
x_2	[124.58, 198.34]	[4.00, 11.00]	[114.98, 124.00]	[85.65, 121.94]	[82.60, 224.88]	[69.28, 89.08]
x_3	[109.23, 120.02]	[2.99, 10.00]	*	[166.35, 196.96]	*	[88.20, 96.73]
x_4	[125.19, 134.25]	[3.00, 8.00]	[110.98, 120.45]	[62.20, 97.76]	[101.99, 260.88]	[68.45, 89.10]
x_5	*	[10.00, 20.00]	[110.45, 296.98]	*	*	[45.95, 75.96]
x_6	*	[8.40, 18.00]	[176.36, 258.94]	[83.03, 163.96]	[68.29, 88.98]	[28.68, 62.96]
x_7	[117.97, 129.40]	[12.00, 21.00]	*	*	[79.62, 95.25]	[32.72, 68.80]
x_8	[160.00, 232.99]	*	[219.68, 314.70]	[103.28, 159.99]	[109.25, 260.40]	[26.95, 66.45]

$$MsIIvIS = \{IIvIS_i \,|\, IIvIS_i = (U, A, V_i, f_i), i = 1, 2, \cdots, N\}.$$

Similarly, $IIvDIS = (U, A, V_A, f_A, D, V_D, f_D)$ represents incomplete interval-valued decision information system, where the connotations of U, A, V_A and f_A are in agreement with those mentioned in the $IIvIS$, D represents the decision attribute set. V_D represents the range of the decision attribute value. Information function is expressed as $f_D : U \times D \to V_D$. $I = [0, 1]$, I^U is called as the family consisted of all fuzzy sets on U.

Let $IIvDIS = (U, A, V_{A_i}, f_{A_i}, D, V_{D_i}, f_{D_i})$ be the i-th $IIvDIS$, where the connotations of U, A, V_{A_i}, f_{A_i}, D, V_{D_i} and f_{D_i} as mentioned above. In general, a multi-source incomplete interval-valued decision information system (MsIIvDIS) is expressed as follows:

$$MsIIvDIS = \{IIvDIS_i \,|\, IIvDIS_i = (U, A, V_{A_i}, f_{A_i}, D, V_{D_i}, f_{D_i}), i = 1, 2, \cdots, N\}.$$

For convenience, this article abbreviated the above expression. We use $(U, A \bigcup D)_i$ to represent the decision information system and (U, A_i) to represent the information system.

Example 1. In order to better understand the definition of MsIIvDIS, we give the example as follows. With the awakening of people's health awareness, more and more friends began to develop the habit of regular physical examination. However, because the interval between medical examinations is long, and the specific time and place are not fixed. As a result, many people have several physical examinations, which are not carried out in the same hospital. After receiving the results of the physical examination, some people will find that the numerical results of the physical examination they did in several hospitals are very different. Tables 1–4 respectively represent the physical examination results of eight people in four hospitals. Attributes $a_1 - a_6$ indicate hemoglobin counts, leukocyte counts, blood fat, blood sugar, platelet counts, and Hb level, respectively. Where "$*$" is the missing value that represents a doctor cannot ensure the level of this project or people forget to check this project. Suppose that $V_D = \{Leukemia \quad patient, Non \quad leukemia \quad patient\}$, and $U/D = \{Y_1, Y_2\}$, where $Y_1 = \{x_1, x_2, x_6, x_8\}$, $Y_2 = \{x_3, x_4, x_5, x_7\}$.

Table 4
Physical examination report of the fourth hospital $IIvIS_4$.

U	a_1	a_2	a_3	a_4	a_5	a_6
x_1	*	*	[116.25, 125.97]	[65.95, 119.58]	[42.30, 187.96]	[34.30, 96.12]
x_2	[198.35, 216.56]	[4.00, 11.00]	[114.96, 123.98]	[85.65, 121.98]	[82.61, 224.90]	[69.30, 89.10]
x_3	[109.25, 120.06]	[2.99, 10.00]	[127.00, 185.35]	*	[67.08, 85.97]	[88.20, 96.73]
x_4	[65.89, 134.30]	[3.00, 8.00]	[110.99, 120.47]	[42.28, 67.80]	[101.98, 260.89]	*
x_5	[109.72, 123.99]	[10.00, 20.00]	[110.48, 296.96]	[82.60, 121.90]	*	[46.00, 75.96]
x_6	*	[8.39, 18.00]	[176.40, 258.96]	[83.00, 163.98]	[68.29, 88.98]	[28.70, 62.98]
x_7	*	[12.00, 21.00]	[169.34, 269.98]	[80.98, 156.99]	[79.65, 95.28]	[32.69, 68.82]
x_8	[98.24, 120.69]	[6.30, 15.00]	[219.65, 314.68]	*	[109.30, 260.44]	[26.98, 66.42]

2.3. *Fuzzy similarity relation in $IIvDIS$*

Fuzzy similarity relation is the basis of dividing decision table and the basis of information granulation in decision system. The product of this process is information granule, which is the basic component of information structure.

Definition 4. Suppose that $(U, A\bigcup D)_i$ is the i-th $IIvDIS_i$. For condition attribute subset $B \subseteq A$, the fuzzy similarity relation is expressed as follows:

$$R^i_B = (\mathop{\wedge}_{b\in B} Sim^i_b(x_j, x_k))_{n\times n} \; (\forall x_j \in U, (x_j, x_k) \in n \times n),$$

where $Sim^i_b(x_j, x_k)$ denotes the similarity of x_j and x_k under attribute b in the i-th $IIvDIS_i$.

$Sim^i_B(x_j)$ contains the similarity between the object and all other objects in the universe, and can be acknowledged to be a fuzzy information granule. The set of these information granules forms a fuzzy set vector, which is called fuzzy similarity class. And

$$Sim^i_B(x_j) = \frac{R^i_B(x_j, x_1)}{x_1} + \frac{R^i_B(x_j, x_2)}{x_2} + \cdots + \frac{R^i_B(x_j, x_n)}{x_n}.$$

Proposition 1. *Let the i-th $IIvDIS_i(U, A)_i$ $(i = 1, 2, \cdots, N)$ and the attribute subset $B \subseteq A$, for $\forall x_i \in U$. The following properties are true:*

(1) On the U, R^i_B is a fuzzy similarity relation.
(2) $Sim^i_B(x_j) = \bigcup_{b\in B} Sim^i_b(x_j)$, and $\bigcup_{x_i\in U} Sim^i_B(x_j) = \tilde{1}$.
(3) When $C \subseteq B$, have $Sim^i_B(x_j) \subseteq Sim^i_C(x_j)$.

Proof. (1) For $\forall x_j, x_k \in U$ and $b \in B$, since $Sim^i_b(x_j, x_j) = 1$ and $Sim^i_b(x_j, x_k) = Sim^i_b(x_k, x_j)$, then $R^i_B(x_j, x_j) = \mathop{\wedge}_{b\in B} Sim^i_b(x_j, x_j) = 1$, $R^i_B(x_j, x_k) = \mathop{\wedge}_{b\in B} Sim^i_b(x_j, x_k) = \mathop{\wedge}_{b\in B} Sim^i_b(x_k, x_j) = R^i_B(x_k, x_j)$. So R^i_B satisfies both reflexivity and symmetry. Thus, R^i_B is a fuzzy similarity relation.

(2) For $\forall x_k \in U$, we know $(Sim^i_B(x_j))(x_k) = R^i_B(x_j, x_k) = \mathop{\wedge}_{b\in B} Sim^i_b(x_j, x_k) = \mathop{\wedge}_{b\in B} Sim^i_b(x_j)(x_k) = (\bigcap_{b\in B} Sim^i_b(x_j))(x_k)$.

And $(\bigcup_{x_i\in U} Sim^i_B(x_j))(x_k) = \mathop{\vee}_{x_i\in U} Sim^i_B(x_j, x_k) = \mathop{\vee}_{x_i\in U} R^i_B(x_j, x_k) = R^i_B(x_k, x_k) = \tilde{1}$.

(3) For $\forall x_j, x_k \in U$, since $C \subseteq B$, then $\mathop{\wedge}_{b\in B} Sim^i_b(x_j, x_k) \le \mathop{\wedge}_{b\in C} Sim^i_b(x_j, x_k)$. So, for any $x_k \in U$, $(Sim^i_B(x_j))(x_k) = R^i_B(x_j, x_k) = \mathop{\wedge}_{b\in B} Sim^i_b(x_j, x_k) \le \mathop{\wedge}_{b\in C} Sim^i_b(x_j, x_k) = R^i_C(x_j, x_k) = (Sim_C(x_j))(x_k)$. Hence, $Sim^i_B(x_j) \subseteq Sim^i_B(x_k)$. □

Example 2 *(Continued from Example 1).* In accordance with to the above definition, we can figure up the fuzzy similarity class. Let's take attribute a_1 of the first information source as an example.

Firstly, according to Definition 2 we can calculate the distance between x_i and x_j $(i, j = 1, 2, ..., 8)$ w.r.t. a_1. The specific calculation process of distance $d_{a_1}(x_i, x_j)$ is as follows:

$$f(x_1, a_1) = [125.59, 136.29], f(x_2, a_2) = [124.58, 198.34],$$

$$\mu_1 = 130.94, \sigma_1 = 5.35,$$

$$\mu_2 = 161.46, \sigma_2 = 36.88.$$

Suppose that:

$$f(x_1, a_1) \sim N(130.94, 5.35^2), f(x_2, a_1) \sim N(111.46, 36.88^2)$$

so according to the Definition 2 we can calculation the KL divergence,

$KL(f(x_1,a_1)\,\|\,f(x_2,a_1))=0.11130,$

$KL(f(x_2,a_1)\,\|\,f(x_1\ ,a_1))=0.18375.$

Thus, $d_{a_1}(x_1,x_2)=4.4376$, the same we can also obtain a distance matrix,

$$d_{a_1}=\begin{pmatrix} 0 & 0.3841 & 0.1565 & 0.0152 & 0.0353 & 0.4722 & 0.0659 & 0.5791 \\ 0.3841 & 0 & 0.5290 & 0.4003 & 0.4031 & 0.0813 & 0.4412 & 0.1979 \\ 0.1565 & 0.5290 & 0 & 0.1434 & 0.1247 & 0.6223 & 0.0902 & 0.7496 \\ 0.0152 & 0.4003 & 0.1434 & 0 & 0.0291 & 0.4889 & 0.0544 & 0.5968 \\ 0.0353 & 0.4031 & 0.1247 & 0.0291 & 0 & 0.4921 & 0.0359 & 0.6055 \\ 0.4722 & 0.0813 & 0.6223 & 0.4889 & 4.3409 & 0 & 0.5315 & 0.1315 \\ 0.0659 & 0.4412 & 0.0902 & 0.0544 & 0.4178 & 0.5315 & 0 & 0.6477 \\ 0.5791 & 0.1979 & 0.7496 & 0.5968 & 5.5241 & 0.1315 & 0.6477 & 0 \end{pmatrix},$$

so the similarity between x_i and x_j w.r.t. a_1 in the first $IIvIS_1$ can be calculated as follows,

$$Sim^1_{a_1}=\begin{pmatrix} 1 & 0.7225 & 0.8647 & 0.9850 & 0.9659 & 0.6793 & 0.9382 & 0.6333 \\ 0.7225 & 1 & 0.6540 & 0.7142 & 0.7127 & 0.9248 & 0.6938 & 0.8348 \\ 0.8647 & 0.6540 & 1 & 0.8746 & 0.8891 & 0.6164 & 0.9172 & 0.5716 \\ 0.9850 & 0.7142 & 0.8746 & 1 & 0.9717 & 0.6716 & 0.9493 & 0.6263 \\ 0.9659 & 0.7127 & 0.8891 & 0.9717 & 1 & 0.6702 & 0.9653 & 0.6299 \\ 0.6793 & 0.9248 & 0.6164 & 0.6716 & 0.6702 & 1 & 0.6529 & 0.8838 \\ 0.9382 & 0.6938 & 0.9172 & 0.9493 & 0.9653 & 0.6529 & 1 & 0.6069 \\ 0.6333 & 0.8348 & 0.5716 & 0.6263 & 0.6299 & 0.8838 & 0.6069 & 1 \end{pmatrix}.$$

Then the fuzzy similarity class in the 1-th $IIvIS_1$ can be calculated as follows,

$$Sim^1_{a_1}(x_1)=\frac{1}{x_1}+\frac{0.7225}{x_2}+\frac{0.8647}{x_3}+\frac{0.9850}{x_4}+\frac{0.9659}{x_5}+\frac{0.6793}{x_6}+\frac{0.9382}{x_7}+\frac{0.6333}{x_8},$$
$$Sim^1_{a_1}(x_2)=\frac{0.7225}{x_1}+\frac{1}{x_2}+\frac{0.6540}{x_3}+\frac{0.7142}{x_4}+\frac{0.7127}{x_5}+\frac{0.9248}{x_6}+\frac{0.6938}{x_7}+\frac{0.8348}{x_8},$$
$$Sim^1_{a_1}(x_3)=\frac{0.8647}{x_1}+\frac{0.6540}{x_2}+\frac{1}{x_3}+\frac{0.8746}{x_4}+\frac{0.8891}{x_5}+\frac{0.6164}{x_6}+\frac{0.9172}{x_7}+\frac{0.5716}{x_8},$$
$$Sim^1_{a_1}(x_4)=\frac{0.9850}{x_1}+\frac{0.7142}{x_2}+\frac{0.8746}{x_3}+\frac{1}{x_4}+\frac{0.9717}{x_5}+\frac{0.6716}{x_6}+\frac{0.9493}{x_7}+\frac{0.6263}{x_8},$$
$$Sim^1_{a_1}(x_5)=\frac{0.9659}{x_1}+\frac{0.7127}{x_2}+\frac{0.8891}{x_3}+\frac{0.9717}{x_4}+\frac{1}{x_5}+\frac{0.6702}{x_6}+\frac{0.9653}{x_7}+\frac{0.6299}{x_8},$$
$$Sim^1_{a_1}(x_6)=\frac{0.6793}{x_1}+\frac{0.9248}{x_2}+\frac{0.6164}{x_3}+\frac{0.6716}{x_4}+\frac{0.6702}{x_5}+\frac{1}{x_6}+\frac{0.6529}{x_7}+\frac{0.8838}{x_8},$$
$$Sim^1_{a_1}(x_7)=\frac{0.9382}{x_1}+\frac{0.6938}{x_2}+\frac{0.9172}{x_3}+\frac{0.9493}{x_4}+\frac{0.9653}{x_5}+\frac{0.6529}{x_6}+\frac{1}{x_7}+\frac{0.6069}{x_8},$$
$$Sim^1_{a_1}(x_8)=\frac{0.6333}{x_1}+\frac{0.8348}{x_2}+\frac{0.5716}{x_3}+\frac{0.6263}{x_4}+\frac{0.6299}{x_5}+\frac{0.8838}{x_6}+\frac{0.6069}{x_7}+\frac{1}{x_8}.$$

3. Uncertainty measurement of $MsIIvDIS$ based on information granularity

The procedure of information granularity is to separate a sample into various information granularities under the given rules, in which each granular is the set of samples gathered through indiscernible relations, similar relations, etc. In this paper, each granular is the set of samples gathered by δ-similarity equivalence relation, and δ-similarity equivalence relation is a similar relation. Then, all the granular forms an adjacent granular structure. Based on the δ-similarity equivalent granular structure, the uncertainty measure is defined. In this section, we introduce the definitions and properties of δ-similarity equivalence relation, and then define the uncertainty measures based on information structure.

3.1. δ-similarity equivalence relation and information structure

So far, many measures have described the distance between two fuzzy sets. Such as Chebyshev distance, Hamming distance, the Euclid distance, the Minkowski distance, etc. But in this subsection, we use the Euclidean distance. Expressed as follows:

Let $F(x), G(x)$ be two fuzzy sets. Assume $F,G\in I^U$, the Euclidean distance as follows:

$$d_E(F,G)=(\frac{1}{n}\sum_{j=1}^{n}(F(x_j)-G(x_j))^2)^{1/2}.$$

Definition 5. Assume that $(U, A \bigcup D)_i$ $(i = 1, 2, \cdots, N)$ is the i-th $IIvDIS_i$. For attribute subset $B \subseteq A$ and a given parameter $\delta \in [0, 1]$. Then the δ-similarity equivalence relation is presented as below:

$$AER^{\delta}_{B,i} = \{(x_j, x_k) \in U \times U : \forall b \in B, d^i_E(Sim^i_b(x_j), Sim^i_b(x_k)) < \delta\}. \tag{10}$$

For any $x_j \in U$, the δ-similarity equivalence class can be presented as below:

$$(x_j)^{\delta}_{B,i} = \{x_k \in U : (x_j, x_k) \in AER^{\delta}_{B,i}\}. \tag{11}$$

For any $B \subseteq A$, the δ-similarity equivalence class satisfies $(x_j)^{\delta}_{B,i} = \bigcap_{b\in B} (x_j)^{\delta}_{b,i}$. In other words, a family of δ-similarity equivalence relation forms covering of U, i.e., $\sum_{x\in U} (x)^{\delta}_B = U$. Then, for $\forall x_i$, $x_j \in (x)^{\delta}_B$, which cannot be distinguished under $(x)^{\delta}_B$. Thus, δ-similarity equivalence relation can be viewed as an information granularity.

Definition 6. For a given parameter $\delta \in [0, 1]$ and $B \subseteq A$, the δ-similarity equivalence granular structure is a information structure can be defined by:

$$AEGS^{\delta}_i(B) = ((x_1)^{\delta}_{B,i}, (x_2)^{\delta}_{B,i}, (x_3)^{\delta}_{B,i}, ..., (x_{|U|})^{\delta}_{B,i}).$$

Proposition 2. *Assume that $(U, A \bigcup D)_i$ $(i = 1, 2, \cdots, N)$ is the i-th $IIvDIS_i$. For $B, C \subseteq A$ and $\delta \in [0, 1]$, the following properties are true:*

(1) $(x_j)^{\delta}_{B,i} \neq \emptyset$ *and* $\bigcup_{x_j\in U} (x_j)^{\delta}_{B,i} = U$.

(2) If $C \subseteq B$, *then* $(x_j)^{\delta}_{B,i} \subseteq (x_j)^{\delta}_{C,i}$.

(3) $0 \le \delta_1 \le \delta_2 \le 1$, $(x_j)^{\delta_1}_{B,i} \subseteq (x_j)^{\delta_2}_{B,i}$.

Proof. (1) For $x_j \in U$ and $b \in B$, we have $d^i_E(Sim^i_b(x_j), Sim^i_b(x_j)) = 0 \le \delta$, so $x_j \in (x_j)^{\delta}_B$. Thus $(x_j)^{\delta}_{B,i} \neq \emptyset$ and $\bigcup_{x_j\in U} (x_j)^{\delta}_{B,i} = U$.

(2) For $\forall x_k \in (x_j)^{\delta}_B$, we know $\forall b \in B$, $d_E(Sim^i_b(x_j), Sim^i_b(x_k)) \le \delta$, since $C \subseteq B$, then $\forall b \in C$, $d_E(Sim^i_b(x_i), Sim^i_b(x_j)) \le \delta$, so according to the Definition 8, we know $x_k \in (x_j)^{\delta}_C$, thus $(x_j)^{\delta}_B \subseteq (x_j)^{\delta}_C$.

(3) For $\forall x_k \in (x_j)^{\delta_1}_{B,i}$, we have $\forall b \in B$, $d^i_E(Sim^i_b(x_j), Sim^i_b(x_k)) \le \delta_1$. Since $\delta_1 \le \delta_2$, then $\forall b \in B$, $d^i_E(Sim^i_b(x_j), Sim^i_b(x_k)) \le \delta_1 \le \delta_2$, so $x_k \in (x_j)^{\delta_2}_{B,i}$, thus $(x_j)^{\delta_1}_{B,i} \subseteq (x_j)^{\delta_2}_{B,i}$. □

On the U, AER^{δ}_B is also a similarity relation.

Assume that $(U, A \bigcup D)_i$ $(i = 1, 2, \cdots, N)$ is the i-th $IIvDIS_i$. For $B \subseteq A$, $(x_j)^{\delta}_{B,i}$ is the fuzzy δ-similarity equivalence relation induced by B, and a given parameter value $\delta \in [0, 1]$. Then the δ-equivalence relation is considered as follows:

$$ER^{\delta}_{B,i} = \{(x_j, x_k) \in U \times U : (x_j)^{\delta}_{B,i} = (x_k)^{\delta}_{B,i}\}. \tag{12}$$

For any $x_j \in U$, the $\delta-$ equivalence class can be defined as follows:

$$X^{B,i}_l = \{x_k \in U : (x_j)^{\delta}_{B,i} = (x_k)^{\delta}_{B,i}\}, x_j \in X^{B,i}_l, l = 1, 2, \cdots, r. \tag{13}$$

Additionally, the partition induced by ER^{δ}_B is defined as $U/ER^{\delta}_B = \{X^B_1, X^B_2, \cdots, X^B_r\}$. Obviously, for any $x_j \in X^B_l$ $(l = 1, 2, \cdots, r)$, $(X^B_l)^{\delta}_B = (x_j)^{\delta}_B$ and $X^B_l \subseteq (x_j)^{\delta}_B$.

Example 3 *(Continued from Example 2)*. In this case, we set the value of parameter δ to 0.3. For space reasons, we take attribute a_1 as an example. We have:

$(x_1)^{0.3}_{a_1} = \{x_1, x_2, x_4, x_5, x_6, x_7\}$; $(x_2)^{0.3}_{a_1} = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$; $(x_3)^{0.3}_{a_1} = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$; $(x_4)^{0.3}_{a_1} = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$; $(x_5)^{0.3}_{a_1} = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$; $(x_6)^{0.3}_{a_1} = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8)$; $(x_7)^{0.3}_{a_1} = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$; $(x_8)^{0.3}_{a_1} = \{x_2, x_6, x_8\}$.

3.2. Uncertainty measurement based on information granularity

In this subsection, we will use the information granular structure to study the uncertainty measurement for MsIIvDIS.

Definition 7. Given that $(U, A \bigcup D)_i$ $(i = 1, 2, \cdots, N)$ is i-th $IIvDIS_i$. For $B \subseteq A$ and a given parameter $\delta \in [0, 1]$, $AEGS^{\delta}_i(B)$ represents the information structure caused by B in i-th $IIvDIS_i$, $AEGS^{\delta}_i(B) = ((x_1)^{\delta}_{B,i}, (x_2)^{\delta}_{B,i}, ..., (x_{|U|})^{\delta}_{B,i})$. The lower approximation, upper approximation and boundary region of X according to B can be expressed as below:

$$\underline{R^{\delta}_{B,i}}(X) = \{x \in U : (x_j)^{\delta}_{B,i} \subseteq X\},$$

$$\overline{R^{\delta}_{B,i}}(X) = \{x \in U : (x_j)^{\delta}_{B,i} \bigcap X \neq \varnothing\},$$
$$BR^{\delta}_{B,i}(X) = \overline{R^{\delta}_{B,i}}(X) - \underline{R^{\delta}_{B,i}}(X).$$

Suppose that $U/D = \{Y_1, Y_2, \cdots, Y_m\}$ is decision partition of U according to the decision attribute D. For decision information system, the δ-approximation classified accuracy (AP) and δ-approximation classified quality (AQ) of U/D in relation to AER^{δ}_{B} are expressed as below:

$$AP_{AER^{\delta}_{B,i}}(U/D) = \frac{\sum_{r=1}^{m} \left|\underline{R^{\delta}_{B,i}}(Y_r)\right|}{\sum_{r=1}^{m} \left|\overline{R^{\delta}_{B,i}}(Y_r)\right|}, \tag{14}$$

$$AQ_{AER^{\delta}_{B,i}}(U/D) = \frac{\sum_{r=1}^{m} \left|\underline{R^{\delta}_{B,i}}(Y_r)\right|}{|U|}. \tag{15}$$

Definition 8. Given that $MsIIvDIS$ is multi-source incomplete interval-valued decision information system. For $B \subseteq A$, $X \subseteq U$ and a given threshold $0 \leq \delta \leq 1$. Under the condition attribute set B, the measurement accuracy and roughness for the degree of knowledge uncertainty for a given set X are expressed as follows:

$$Accuracy^{\delta}_{B,i}(X) = \frac{\underline{R^{\alpha}_{B,i}}(X)}{\overline{R^{\delta}_{B,i}}(X)}, \tag{16}$$

$$Roughness^{\delta}_{B,i}(X) = 1 - Accuracy^{\delta}_{B,i}(X). \tag{17}$$

Definition 9. Let $(U, A)_i$ $(i = 1, 2, \cdots, N)$ be the i-th $IIvIS_i$. For $B \subseteq A$ and a given threshold $\delta \in [0, 1]$. The partition induced by $ER^{\delta}_{B,i}$: $U/ER^{\delta}_{B,i} = \{X^{B,i}_1, X^{B,i}_2, \cdots, X^{B,i}_r\}$. $AEGS^{\delta}_i(B)$ is the information structure induced by B in i-th $IIvDIS_i$, $AEGS^{\delta}_i(B) = ((x_1)^{\delta}_{B,i}, (x_2)^{\delta}_{B,i}, ..., (x_{|U|})^{\delta}_{B,i})$. The δ-equivalence information entropy of i-th $IIvDIS_i$ can be defined as below:

$$EIE^{\delta}_i(B) = -\sum_{l=1}^{r} \frac{\left|X^{B,i}_l\right|}{|U|} \log_2 \frac{\left|\left(X^{B,i}_l\right)^{\delta}_{B,i}\right|}{|U|}. \tag{18}$$

Proposition 3. *Assume that $(U, A)_i$ $(i = 1, 2, \cdots N)$ is the i-th $IIvIS_i$. For given $\delta \in I$ and $B \subseteq A$. $AEGS^{\delta}_i(B)$ is the information structure induced by B in i-th $IIvDIS_i$, $AEGS^{\delta}_i(B) = ((x_1)^{\delta}_{B,i}, (x_2)^{\delta}_{B,i}, ..., (x_{|U|})^{\delta}_{B,i})$. The δ-equivalence information entropy of i-th $IIvDIS_i$ also can be written as:*

$$EIE^{\delta}_i(B) = -\sum_{j=1}^{|U|} \frac{1}{|U|} \log_2 \frac{\left|(x_j)^{\delta}_{B,i}\right|}{|U|}. \tag{19}$$

Proof. $U/ER^{\delta}_{B,i} = \{X^{B,i}_1, X^{B,i}_2, \cdots, X^{B,i}_l, \cdots, X^{B,i}_r\}$. Then $\sum_{l=1}^{r} \left|X^{B,i}_l\right| = n$. So $\sum_{l=1}^{r} \frac{|X^{B,i}_l|}{|U|} \log_2 \frac{\left|\left(X^{B,i}_l\right)^{\alpha}_{B,i}\right|}{|U|} = \frac{n}{|U|} \log_2 \frac{\left|\left(X^{B,i}_l\right)^{\alpha}_{B,i}\right|}{|U|}$. Since $(X^{B,i}_l)^{\alpha}_{B,i} = (x_j)^{\delta}_{B,i}$, so $\frac{n}{|U|} \log_2 \frac{\left|\left(X^{B,i}_l\right)^{\delta}_{B,i}\right|}{|U|} = \frac{n}{|U|} \log_2 \frac{|(x_j)^{\delta}_{B,i}|}{|U|}$, i.e. $\sum_{l=1}^{r} \frac{|X^{B,i}_l|}{|U|} \log_2 \frac{\left|\left(X^{B,i}_l\right)^{\alpha}_{B,i}\right|}{|U|} = \frac{n}{|U|} \log_2 \frac{|(x_j)^{\delta}_{B,i}|}{|U|} = \sum_{j=1}^{n} \frac{1}{|U|} \log_2 \frac{|(x_j)^{\delta}_{B,i}|}{|U|}$. Thus $EIE^{\delta}_i(B) = -\sum_{l=1}^{r} \frac{|X^{B,i}_l|}{|U|} \log_2 \frac{\left|\left(X^{B,i}_l\right)^{\delta}_{B,i}\right|}{|U|} = -\sum_{j=1}^{|U|} \frac{1}{|U|} \log_2 \frac{|(x_j)^{\delta}_{B,i}|}{|U|}$. $\square$

For convenience, the expression of information entropy in Proposition 3 is uniformly used in this paper.

(1) Let $(x_j)^{\delta}_{B,i} = \{x_k\}$ $(k = 1, 2, \cdots, n)$, then $EIE^{\delta}_i(B)$ of i-th $IIvIS_i$ reaches its maximum value: $EIE^{\delta}_i(B) = \log_2 |U|$.
(2) Let $(x_j)^{\delta}_{B,i} = U$, then $EIE^{\delta}_i(B)$ of i-th $IIvIS_i$ reaches its minimum value: $EIE^{\delta}_i(B) = 0$. Thus, $0 \leq EIE^{\delta}_i(B) \leq \log_2 |U|$.

Theorem 1. *Suppose that $(U, A)_i$ $(i = 1, 2, \cdots, N)$ is the i-th $IIvIS_i$. Given $\delta_1, \delta_2 \in [0, 1]$ and $B, C \subseteq A$, the δ-equivalence information entropy of B and C in i-th $IIvIS_i$ stratifies:*

(1) If $C \subseteq B \subseteq A$, then for $\forall \delta \in [0, 1]$, $EIE^{\delta}_i(C) \leq EIE^{\delta}_i(B)$,
(2) $\forall B \subseteq A$, $EIE^{\delta_2}_i(B) \leq EIE^{\delta_1}_i(B)$, if $0 < \delta_1 \leq \delta_2 \leq 1$.

Proof. (1) For any $(x_j) \in U$ and $\delta \in [0,1]$, since $C \subseteq B \subseteq A$. According to the Proposition 2(2), we have $(x_j)^\delta_B \subseteq (x_j)^\delta_C$. Then $|(x_j)^\delta_B| \leq |(x_j)^\delta_C|$, so $EIE^\delta(B) = -\sum_{j=1}^{|U|} \frac{1}{|U|}\log_2 \frac{|(x_j)^\delta_B|}{|U|} \geq -\sum_{j=1}^{|U|} \frac{1}{|U|}\log_2 \frac{|(x_j)^\delta_C|}{|U|} = EIE^\delta(C)$.

(2) For $\forall B \subseteq A$ and $(x_j) \in U$. Due to $0 \leq \delta_1 \leq \delta_2 \leq 1$, by Proposition 2(3), we have $(x_j)^{\delta_1}_B \subseteq (x_j)^{\delta_2}_C$. Then $|(x_j)^{\delta_1}_B| \leq |(x_j)^{\delta_2}_C|$, thus $EIE^{\delta_1}(B) = -\sum_{j=1}^{|U|} \frac{1}{|U|}\log_2 \frac{|(x_j)^\delta_B|}{|U|} \geq -\sum_{j=1}^{|U|} \frac{1}{|U|}\log_2 \frac{|(x_j)^\delta_B|}{|U|} = EIE^{\delta_2}(B)$. □

Definition 10. Assume that $(U, A \bigcup D)_i$ $(i = 1, 2, \cdots, N)$ is the i-th $IIvDIS_i$. For condition attribute subset $B \subseteq A$ and a given parameter $\delta \in [0,1]$, $U/R = \{Y_1, Y_2, \cdots, Y_m\}$ is the partition of U on the decision attribute set. $AEGS^\delta_i(B)$ is the information structure induced by B in i-th $IIvDIS_i$, $AEGS^\delta_i(B) = ((x_1)^\delta_{B,i}, (x_2)^\delta_{B,i}, ..., (x_{|U|})^\delta_{B,i})$. The i-th δ-approximate conditional entropy of B is expressed as below:

$$CIE^\delta_i(D|B) = -\sum_{j=1}^{|U|}\sum_{k=1}^{m} \frac{\left|(x_j)^\alpha_B \bigcap Y_k\right|}{|U|} \log \frac{\left|(x_j)^\delta_B \bigcap Y_k\right|}{\left|(x_j)^\delta_B\right|}. \tag{20}$$

Additionally, the δ-approximate conditional entropy $CIE^\delta_i(D|B)$ has two propositions which can be expressed as follows:

(1) $0 \leq CIE^\delta_i(D|B) \leq |U| \log |U|$,
(2) If $C \subseteq B$, $CIE^\delta_i(D|B) \leq CIE^\delta_i(D|C)$.

4. A novel information entropy fusion approach in $MsIIvDIS$

According to the property of δ-similarity conditional entropy, we can find the smaller the $CIE^\delta_i(D|a)$ $(i = 1, 2, ..., N)$ is, the more significant the information source is. Hence, we can obtain the following fusion function, which can be employed to fuse the MsIIvDIS.

4.1. Information fusion function in $MsIIvDIS$

Definition 11. Assume that $MsIIvDIS = \{IIvIIS_i | IIvI - IS_i = (U, A, V_i, f_i), i = 1, 2, \cdots, N)$ is a multi-source incomplete interval-valued decision information table, where $A = \{a_1, a_2, \cdots a_p\}$ (for convenience, I_i can represent the i-th information source). Given parameter $\delta \in [0,1]$, for all $a_m \in A$ $(m = \{1, 2, \cdots, p\})$, the m-th attribute of new information table after fusion w.r.t. δ is defined as below:

$$\mathrm{Inf\ ER}^\delta(a_m) = \mathrm{Inf}_{m\in\{1,2,\cdots,p\}}(F(I_1(a_m), F(I_2(a_m), \cdots, F(I_n(a_m))), \tag{21}$$

where $F = CIE^\delta_i(D|\{a_m\})$ which can be regard as infimum-measure function.

For a $MsIIvDIS = \{IIvIIS_i \,|IIvIIS_i = (U, A, V_i, f_i), i = 1, 2, \cdots, N)$, the result of fusion is still incomplete. In this paper, we can use following approach to complete the missing values in the information system.

$$f\left(x_j, a\right) = \begin{cases} \left[\min_{x_k \in U} f^L\left(x_k, a\right), \max_{x_k \in U} f^U\left(x_k, a\right)\right] & iff\ \left(x_j, a\right) = * \\ f\left(x_j, a\right) & else. \end{cases} \tag{22}$$

In this paper, we use Fig. 2 to show the fusion process more intuitively. The different colors of lines indicate different information sources, and each square represents a corresponding attribute value, the solid squares are used to express the missing value. Then, according to the fusion function to select the value of attribute, and it is reconstituted a novel incomplete information system. Finally, the missing values are completed to obtain the final information system.

Example 4 *(Continued from Example 3)*. For the sake of convenience, I_i represent the i-th information system $IIvDIS_i$. Then, the δ-approximate condition entropy of information sources for diverse attributes are calculated in Table 5.

Then, we can compute δ-approximate conditional entropy as follows $(\delta = 0.3)$.

$CIE^{0.3}_1(D|a_1) = -(2 \times \frac{3}{8}\log\frac{3}{6} + 4 \times \frac{4}{8}\log\frac{4}{8} + 4 \times \frac{3}{8}\log\frac{3}{7} + 4 \times \frac{4}{8}\log\frac{4}{7} + 2 \times 0) = 6.198298,$
$CIE^{0.3}_1(D|a_2) = -(0 + 3 \times \frac{3}{8}\log\frac{3}{5} + \frac{3}{8}\log\frac{3}{4} + \frac{1}{8}\log\frac{1}{2} + 2 \times \frac{2}{8}\log\frac{2}{4} + \frac{3}{8}\log\frac{3}{7} + 0 + 3 \times \frac{2}{8}\log\frac{2}{5} + \frac{1}{8}\log\frac{1}{4} + \frac{1}{8}\log\frac{1}{2} + \frac{4}{8}\log\frac{4}{7}) = 3.711400,$
$CIE^{0.3}_1(D|a_3) = -(4 \times \frac{2}{8}\log\frac{2}{3} + 4 \times \frac{1}{8}\log\frac{1}{2} + 2 \times \frac{2}{8}\log\frac{2}{4} + 4 \times \frac{1}{8}\log\frac{1}{3} + \frac{1}{8}\log\frac{1}{4} + \frac{3}{8}\log\frac{3}{4}) = 3.283083,$
$CIE^{0.3}_1(D|a_4) = -(12 \times \frac{4}{8}\log\frac{4}{8} + 2 \times \frac{4}{8}\log\frac{4}{7} + 2 \times \frac{3}{8}\log\frac{3}{7}) = 8.075422,$
$CIE^{0.3}_1(D|a_5) = -(4 \times \frac{3}{8}\log\frac{3}{6} + 4 \times \frac{1}{8}\log\frac{1}{2} + 4 \times \frac{2}{8}\log\frac{2}{4} + 2 \times \frac{3}{8}\log\frac{3}{4} + 2 \times \frac{1}{8}\log\frac{1}{4}) = 3.811278,$
$CIE^{0.3}_1(D|a_6) = -(10 \times \frac{4}{8}\log\frac{4}{8} + 3 \times \frac{2}{8}\log\frac{2}{3} + 3 \times \frac{1}{8}\log\frac{1}{3}) = 5.937985.$

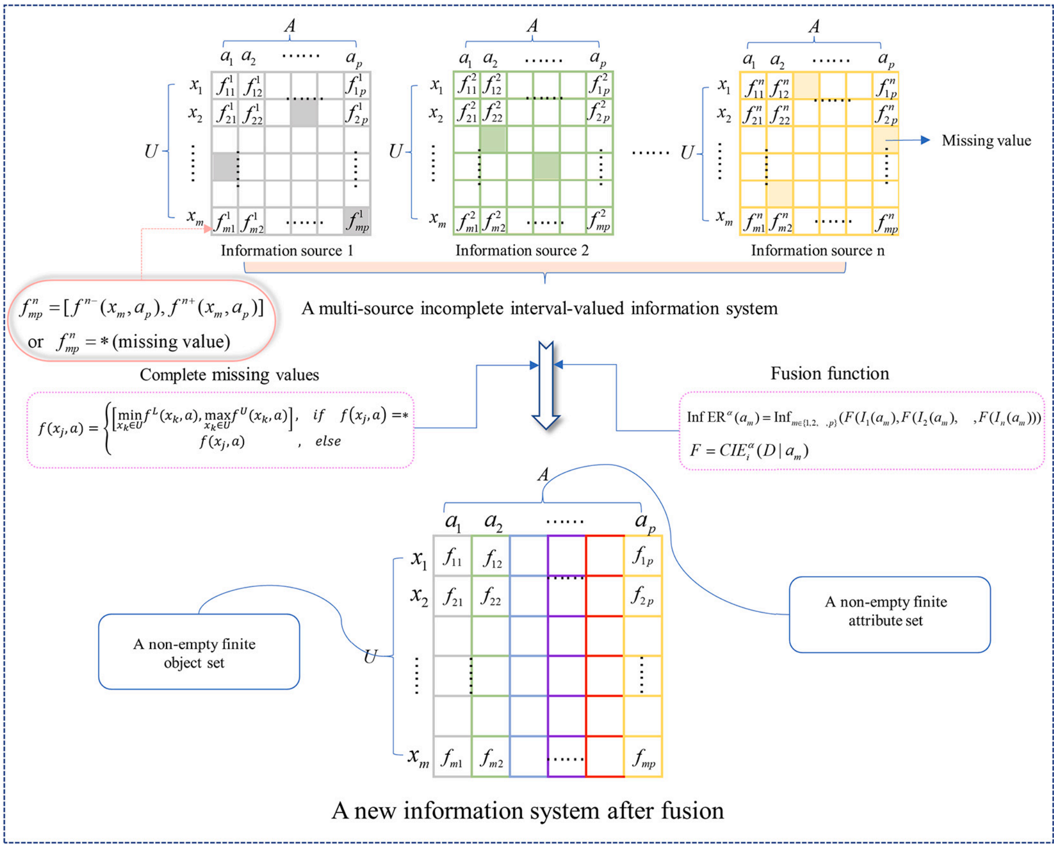


Fig. 2. The fusion process of multi-source incomplete interval-valued information system.

Table 5
The δ-approximate conditional entropy of information sources for diverse attributes.

A	I_1	I_2	I_3	I_4
a_1	**6.198298**	8.000000	7.785427	7.193068
a_2	3.711400	4.294170	**3.283083**	3.631810
a_3	**3.283083**	6.877444	3.449243	4.177376
a_4	8.075422	8.000000	7.412871	**5.560376**
a_5	3.811278	5.523684	5.000000	**3.771737**
a_6	**5.937985**	7.172488	8.000000	7.250000

Table 6
The final fusion results.

U	a_1	a_2	a_3	a_4	a_5	a_6
x_1	[125.59, 136.29]	[1.00, 5.00]	[116.25, 125.97]	[65.95, 119.58]	[42.30, 187.96]	[26.98, 96.73]
x_2	[124.58, 198.34]	[4.00, 11.00]	[114.96, 123.98]	[85.65, 121.98]	[82.65, 224.90]	[26.98, 96.73]
x_3	[109.23, 120.02]	[2.99, 10.00]	[110.45, 296.98]	[42.28, 163.98]	[67.08, 85.97]	[88.20, 96.73]
x_4	[125.19, 134.25]	[3.00, 8.00]	[110.98, 120.45]	[42.28, 67.80]	[101.98, 260.89]	[26.98, 96.73]
x_5	[119.69, 133.97]	[10.00, 20.00]	[110.45, 314.98]	[82.60, 121.90]	[42.30, 260.59]	[45.99, 75.98]
x_6	[126.34, 215.40]	[8.40, 18.00]	[110.45, 314.68]	[83.00, 163.98]	[68.29, 88.98]	[28.65, 62.98]
x_7	[117.97, 129.40]	[12.00, 21.00]	[169.34, 269.98]	[80.98, 156.99]	[79.65, 95.28]	[32.69, 68.80]
x_8	[159.58, 232.69]	[1.00, 21.00]	[219.65, 314.68]	[42.28, 163.98]	[109.26, 260.44]	[26.98, 66.42]

Similarly, we can calculate the δ-approximate conditional entropy of different attributes of other sources. The results are displayed in follows. According to infimum-measure function, then based Definition 11 and Formula (22), we can complete the missing values in the information system. The results are displayed in Table 6.

Algorithm 1: The fusion algorithm of MsIIvDIS based on δ-approximate conditional entropy.

Input : $MsIIvDIS = \{(U, A, V_{A_i}, f_{A_i}, D, V_{D_i}, f_{D_i}), i = 1, 2, \cdots, N\}$; the decision partition $U/D = \{Y_1, Y_2, ..., Y_m\}$, parameter $\delta \in [0, 1]$;
Output : A new fusion table
1 **for** $k = 1 : s$ **do**
2 # s is the number of information sources.
3 **for** *each* $a \in A$ *and* $\forall x_i \in U$ **do**
4 Calculate the distance and similarity degree $Sim_a^k(x_i)$;
5 Given $\delta \in [0, 1]$; Compute $(x_i)_{a,k}^{\delta}$;
6 **end**
7 $CIE \leftarrow 0$
8 **while** $\left|(x_i)_{a,q}^{\delta} \cap Y_j\right| > 0$ **do**
9 **for** $j = 1{:}m$ **do**
10 # m is the cardinality of $|U/D|$
11 $CIE \leftarrow CIE - \frac{\left|(x_i)_{a,q}^{\delta} \cap Y_j\right|}{|U|} \log \frac{\left|(x_i)_{a,q}^{\delta} \cap Y_j\right|}{\left|(x_i)_{a,q}^{\delta}\right|}$
12 **end**
13 **end**
14 **end**
15 **for** *each* $a \in A$ **do**
16 $minCIE \leftarrow \infty$; **for** $k = 1 : s$ **do**
17 **if** $CIE_k^{\delta}(D|a) < minCIE$ **then**
18 $minCIE \leftarrow CIE_k^{\delta}(D|a)$
19 $i^a \leftarrow k$
20 **end**
21 **end**
22 **end**
return : $\left(V_{a_1}^{i_{a_1}}, V_{a_2}^{i_{a_2}}, ..., V_{a_{|AT|}}^{i_{a_{|AT|}}}\right)$

Table 7
The description of experimental data sets.

No.	Data set name	Abbreviation	Objects	Attributes	Decision classes
1	Wine	Wine	178	13	3
2	Auto MPG	AM	398	7	3
3	Breast Cancer Wisconsin	BCW	569	32	2
4	Hill-Valley	HV	606	100	2
5	South German Credit	SGC	1000	21	3
6	Maternal Health Risk	MHR	1014	7	3
7	Contraceptive Method Choice	CMC	1473	9	3
8	Car Evaluation	CE	1728	7	4
9	Wireless Indoor Localization	WIL	2000	7	4
10	Letter	Letter	3349	17	5
11	Abalone	Abalone	4170	8	3
12	Electrical Grid Stability Simulated Data	Ele	10000	14	2

4.2. Fusion algorithm based δ-similarity equivalence relation

From the above, we can obtain fusion Algorithm 1 based on δ-similarity equivalence relation. In Steps 3–6, the computation of the δ-similarity equivalence class for conditional attribute set can be completed in $O(\left|U^2\right| \times |A| \times s)$. Steps 7–12 are to compute the δ-approximate condition entropy, and its the complexity is $O(|U| \times m)$. The time complexity of Steps 1–14 are $O(|U| \times |A| \times s \times (|U| + m))$. The time complexity of Steps 15–23 are $O(|A| \times s)$. Therefore, the total time complexity of Algorithm 1 is $O(|U| \times |A| \times s \times (|U| + m) + |A| \times s)$.

5. Experiment and results

In this part, in the case of verifying the effectiveness and efficiency of the put forward approach, we conducted some comparative experiments based on twelve data sets from UCI database (https://archive.ics.uci.edu/ml/index.php). The details of these data sets are shown in the Table 7. All the experimental programs are run on personal computer. The hardware and software are depicted in Table 8.

As is known to all, the MsIIvDIS cannot be obtained directly from any common databases. So we can use the method in [43] to generate MsIIvDIS. The detailed steps are as below:

(1) Convert single-valued data in the original dataset to interval-valued data.
Let $V(x, a)$ represent the value of x under attribute a, $\forall x \in U, a \in A$, $f^-(x, a) = V(x, a)$-$2\sigma_a$, $f^-(x, a) = V(x, a) + 2\sigma_a$, where σ_a denotes the standard deviation of the attribute a in the same decision class.

Table 8
Description of the experimental environment.

Name	Model	Parameter
CPU	AMD Ryzen 7 R7-5800H	3.2 GHz
System	Windows11	64 bit
Platform	Python	3.9
Memory	DDR4	16 GB; 3200 MHz
Hard Disk	SKHynix_HFS512GDE9X084N	512 G

(2) Generate $MsIIvDIS$.
First of all, m random numbers $\{r_1, r_2, r_3, ..., r_m\}$ that obey Gaussian distribution $N(0, 0.1)$ are generated randomly. If $r_i > 0$, then $f_i^-(x,a) = f^-(x,a)(1-r)$ and $f_i^+(x,a) = f^+(x,a)(1+r)$, otherwise $f_i^-(x,a) = f^-(x,a)(1+r)$ and $f_i^+(x,a) = f^+(x,a)(1-r)$.
(3) Create the missing values.
Missing values are generated by randomly removing 10 percent of the data.

5.1. Analysis of fusion effectiveness (horizontal comparison)

In this subsection, we compare the put forward fusion model with other three relevant fusion approaches to confirm the effectiveness.

We generate $n = 20$ sources, then the other three fusion approaches are expressed as below:

(i) Max fusion approach can be written as MaxF: MaxF $f^-(x,a) = min\left\{f_1^-(x,a), f_2^-(x,a), ..., f_n^-(x,a)\right\}$, MaxF $f^+(x,a) = max\left\{f_1^+(x,a), ..., f_n^+(x,a)\right\}$, where $f^-(x,a)$ and $f^+(x,a)$ are the left and right endpoints of max fusion result, respectively.
(ii) Min fusion method can be written as MinF: MinF $f^-(x,a) = max\left\{f_1^-(x,a), f_2^-(x,a), ..., f_n^-(x,a)\right\}$, MinF $f^+(x,a) = min\left\{f_1^+(x,a), ..., f_n^+(x,a)\right\}$, where $f^-(x,a)$ and $f^+(x,a)$ are the left and right endpoints of min fusion result, respectively.
(iii) Mean fusion method can be written as MeanF: MeanF $f^-(x,a) = mean\left\{f_1^-(x,a), f_2^-(x,a), ..., f_n^-(x,a)\right\}$, MeanF $f^+(x,a) = mean\left\{f_1^+(x,a), ..., f_n^+(x,a)\right\}$, where $f^-(x,a)$ and $f^+(x,a)$ are the left and right endpoints of mean fusion result, respectively.
(iv) The fusion approach is introduced by Zhang et al. [50] (written as CF).

In this paper, we use the AP, AQ and classification accuracy to reflect the fusion effectiveness in this subsection.

We compared the newly proposed information entropy fusion method, based on information structure, with three other fusion approaches across twelve datasets. The Figs. 3 to 4 illustrate the variations in AP and AQ for fusion results under different δ intervals of 0.05. In the initial dataset, when the parameter δ ranges from 0.05 to 0.15, the AP and AQ values remain consistent across all four fusion methods. When the parameter value is 0.2, only the MinF is the same as the AP and AQ values of the proposed approach. However, for the remaining parameters, the AP and AQ values obtained by the fusion method in this paper are significantly superior to those of the other three models. For the data AM, the CieF method is significantly superior to other methods with thresholds between 0.05 and 0.25. “In the BCW dataset, our approach shows an advantage over MaxF and MeanF when the parameter is set to 0.1. Similarly, for the HV dataset, CieF outperforms MaxF and MinF when the parameter is 0.35. In the CMC and MHR datasets, the values of AP and AQ from our approach are higher than those of MaxF and MeanF. Across all parameter values in the SGC, WIL, CE, and Ele datasets, the CieF method consistently outperforms all other centralized fusion methods. Observing the subplots in Figs. 3 and 4, we can see that for the Letter dataset, the AP and AQ values of the proposed algorithm are the same as those of other methods in the parameter range of 0.25–0.35. In other parameter ranges, the AP and AQ values of the proposed method are significantly higher than those of MeanF and MaxF, and slightly higher than MinF’s AP and AQ. For the Abalone dataset, the fusion method proposed in this paper outperforms all other fusion methods when δ values range between 0.05 and 0.15. However, our approach surpasses MeanF when the parameter δ is set to 0.2. In summary, the experimental results demonstrate that our model is generally superior to other fusion methods across various parameter values of δ. Furthermore, it can be seen from Fig. 3 and Fig. 4 that for the same data set, the values of AP and AQ consistently change with variations in the parameter δ and decrease as δ increases. This occurs due to the reduction of the parameter delta, which leads to a decrease in the similarity equivalent class of the object. Consequently, the upper approximation of the set of samples becomes smaller, and the lower approximation becomes more prominent. Therefore, as δ decreases, the values of AP and AQ enhance. However, when the threshold is too small and the approximate equivalent class of delta of the sample only contains the object itself, the values of AP and AQ will reach a maximum of 1. This also means that there is no connection between each object. On the contrary, the approximate equivalent class of each object will increase. When the approximate equivalent class of a sample contains all objects, there is no difference between each sample. Therefore, in practical applications, selecting the appropriate threshold is essential to enhance the efficiency of data mining tasks.

Additionally, we compare paper fusion approach with the CF which is proposed in [50], which are shown in Figs. 5 and 6. In light of the experimental results, the raised new fusion approach based on information structure surpasses the CF approach in most cases under the variation of δ.

K-nearest neighbor (KNN) classifier and probabilistic neural network (PNN) classifier are used to verify the fusion effectiveness. Table 9, Table 10 and Table 11 show the mean of classification precision and standard deviation by ten-times ten-fold cross-validation. In these tables, thickened blackbody numbers represent the highest classification effectiveness among datasets. It is worth noting

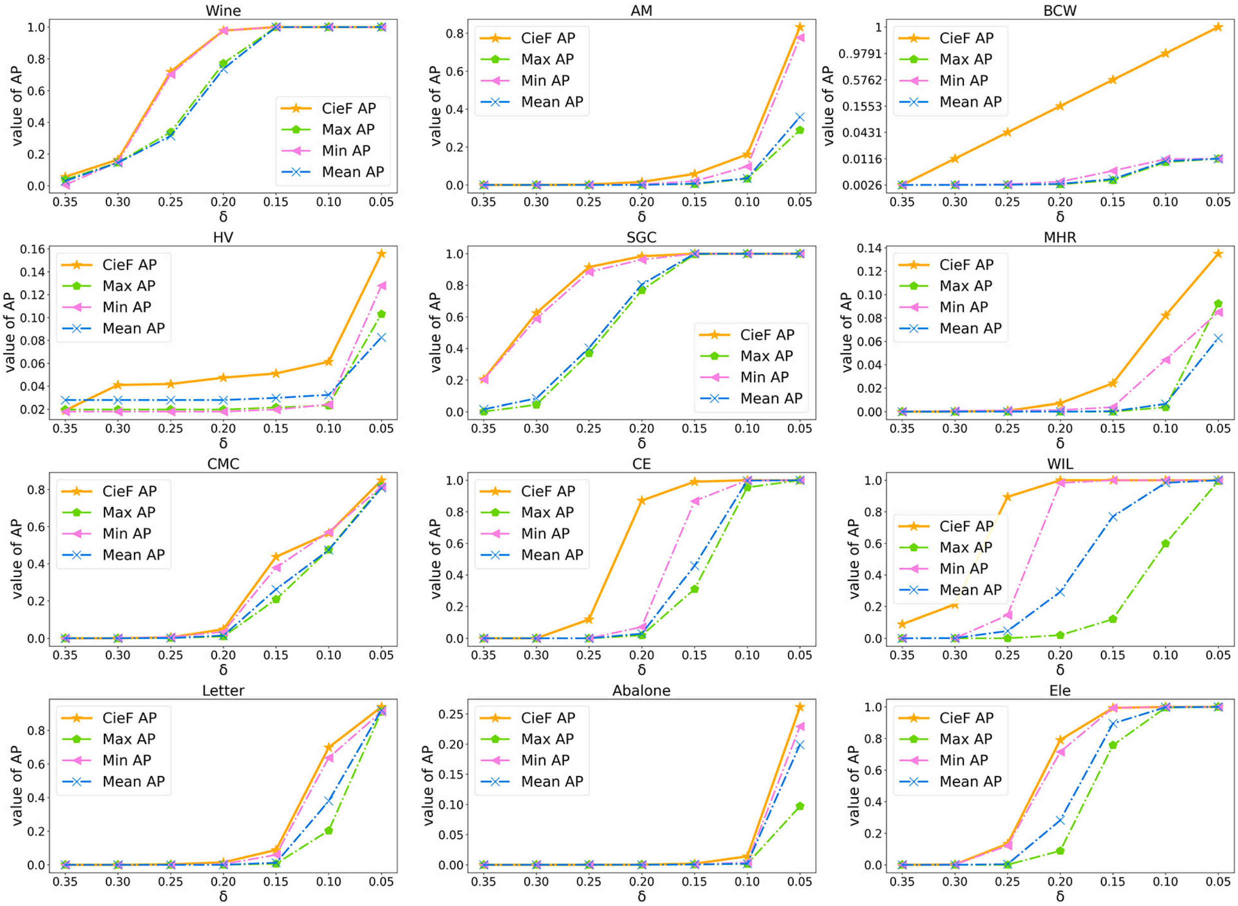


Fig. 3. The comparison results between CieF and other three methods AP.

Table 9
Comparison of classification accuracy based on KNN.

DataSets	KNN			
	CieF	MaxF	MinF	MeanF
Wine (k = 3)	**100.0±0.0**	98.3±2.5	98.3±2.5	98.9±2.2
AM (k = 14)	**63.3±5.1**	60.0±4.3	61.3±5.1	60.5±4.2
BCW (k = 5)	**62.2±5.3**	60.5±4.1	58.4±4.2	59.9±3.3
HV (k = 9)	**51.7±6.0**	50.2±6.1	49.2±6.7	51.5±6.3
SGC (k = 9)	73.0±2.7	**74.0±4.0**	71.7±3.2	73.8±4.3
MHR (k = 17)	**41.5±3.2**	40.8±3.6	39.4±3.5	38.3±2.9
CMC (k = 12)	**55.5±3.6**	54.1±3.7	53.2±4.2	49.0±4.0
CE (k = 16)	**70.8±2.5**	68.2±2.5	69.4±2.9	70.1±2.4
WIL (k = 3)	**55.3±3.7**	53.2±1.6	51.5±4.2	49.6±3.9
Letter (k = 6)	**99.3± 0.6**	98.8±0.6	98.5±0.5	98.9±0.6
Abalone (k = 16)	**35.1±1.7**	33.7±2.7	34.8±1.5	34.8±1.5
Ele (k = 4)	**61.3±3.4**	59.8±2.8	60.0±2.1	60.4±3.1
Avg.	**64.1±3.2**	62.6±3.2	62.1±3.2	62.2±3.2

that the parameters k and σ can influent the classification capability of KNN and PNN classifiers. This flexibility is also an advantage of these two classifiers. Therefore we can adjust the corresponding parameters to achieve the optimal result. The results in Tables 9, 10, 11 indicate clearly that in most situations, the classification accuracies computed by CieF have advantage over those calculated by other three fusion approaches like MaxF, MinF, MeanF and CF.

Furthermore, Wilcoxon signed-rank test in Python is employed to check whether the raised fusion approach (CieF) has an remarkably advantage over other three fusion ways. At the 10% level of significance, assume that the null hypothesis be $H_0 : \mu_{CeF} \leq \mu_{MeanF}/\mu_{MinF}/\mu_{MaxF}/\mu_{CF}$ and the alternative hypothesis be $H_1 : \mu_{CeF} > \mu_{MeanF}/\mu_{MinF}/\mu_{MaxF}/\mu_{CF}$, where $\mu_{CeF}, \mu_{MeanF}, \mu_{MinF}$, μ_{MaxF} and μ_{CF} represent the average of classification accuracy w.r.t. the fusion approach CieF and other fusion approaches, respec-

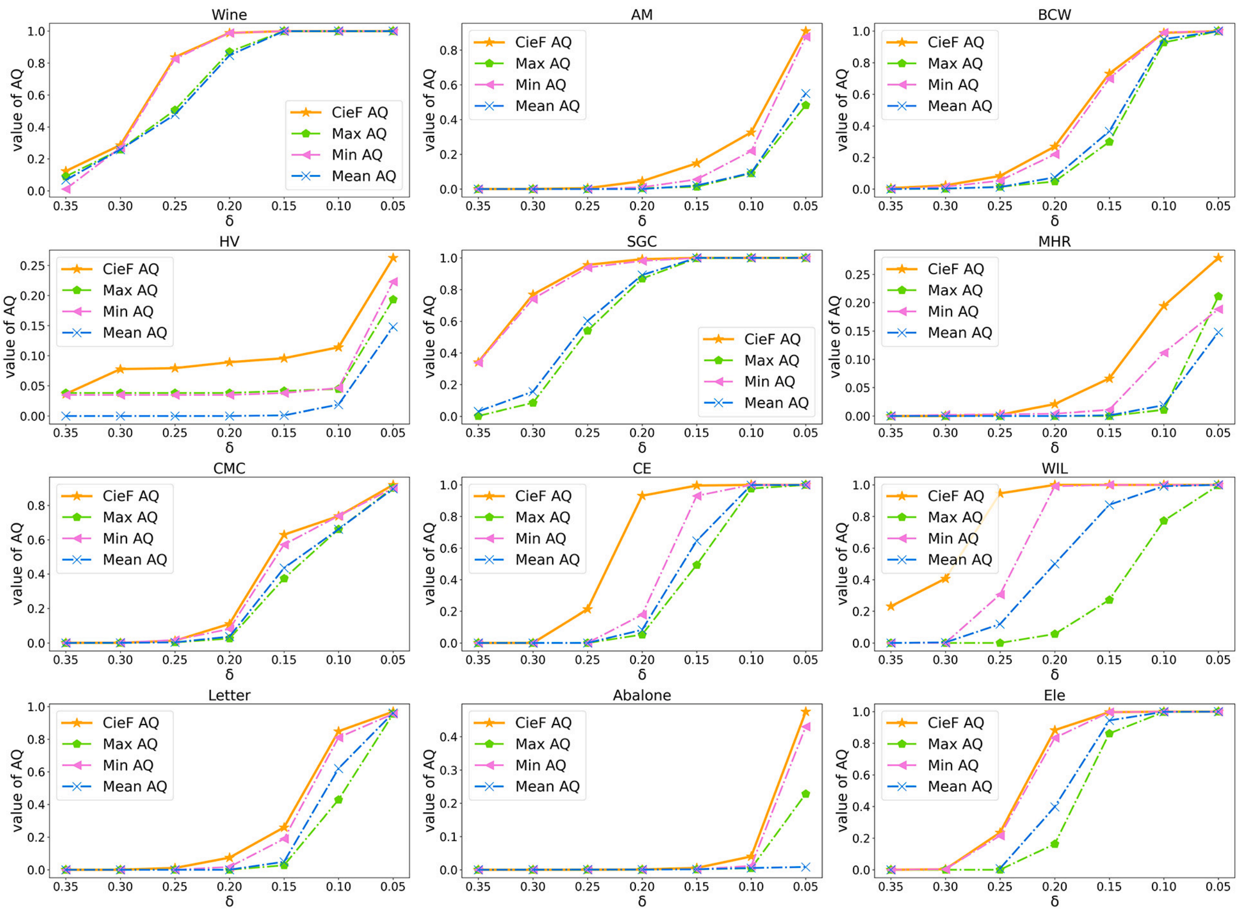


Fig. 4. The comparison results between CieF and other three methods AQ.

Table 10
Comparison of classification accuracy of based on PNN.

DataSets	PNN			
	CieF	MaxF	MinF	MeanF
Wine ($\sigma = 0.68$)	**94.4±5.0**	91.0±8.7	91.0±5.6	91.6±7.5
AM ($\sigma = 0.01$)	50.7±9.8	49.0±8.9	**53.5±8.0**	46.5±7.5
BCW ($\sigma = 0.35$)	62.2±4.8	61.1±5.1	**62.9±4.4**	61.9±4.7
HV ($\sigma = 0.54$)	**52.0±5.5**	50.5±5.0	51.2±5.2	50.9±5.1
SGC ($\sigma = 0.31$)	70.6±3.8	70.0±4.0	**71.4±4.3**	70.2±4.3
MHR ($\sigma = 0.39$)	**40.1±5.5**	40.0±5.5	39.6±5.3	**40.1±5.6**
CMC ($\sigma = 0.35$)	**55.5±4.9**	47.3±5.5	52.5±4.3	47.9±3.2
CE ($\sigma = 0.21$)	**70.4±1.9**	70.2±2.1	70.1±2.0	70.2± 2.1
WIL ($\sigma = 0.3$)	**56.5±4.8**	50.2±3.1	53.8±4.4	49.1±3.4
Letter ($\sigma = 0.3$)	93.9±0.8	87.2±1.6	**94.7±0.6**	90.7±1.0
Abalone ($\sigma = 0.36$)	**36.9±2.4**	36.8±2.5	36.6±2.2	36.8±2.4
Ele ($\sigma = 0.21$)	**62.7±3.3**	53.9±3.7	62.5±2.5	54.9±4.0
Avg.	**62.2±4.4**	58.9±4.6	61.7± 4.1	59.2±4.2

tively. The P-values of the verification results are displayed in Table 12, Table 13 and Table 14. We can know that the classification accuracies of CieF on most datasets have an statistically advantage over other fusion methods.

The above Wilcoxon signed-rank test is employed solely to examine the effectiveness of the model within individual datasets. Moving forward, we will conduct a statistical significance analysis over the whole of the collection. In each dataset, we rank the classification results of different models, assigning a rank of 1 to the top performer, and so on. Subsequently, we calculate the average rank across all datasets. Refer to Table 10 for details. With a P-value significantly less than 0.1 according to the Friedman test, we can reject the null hypothesis, indicating that the fusion method proposed in this paper is significantly superior to other fusion algorithms (Table 15).

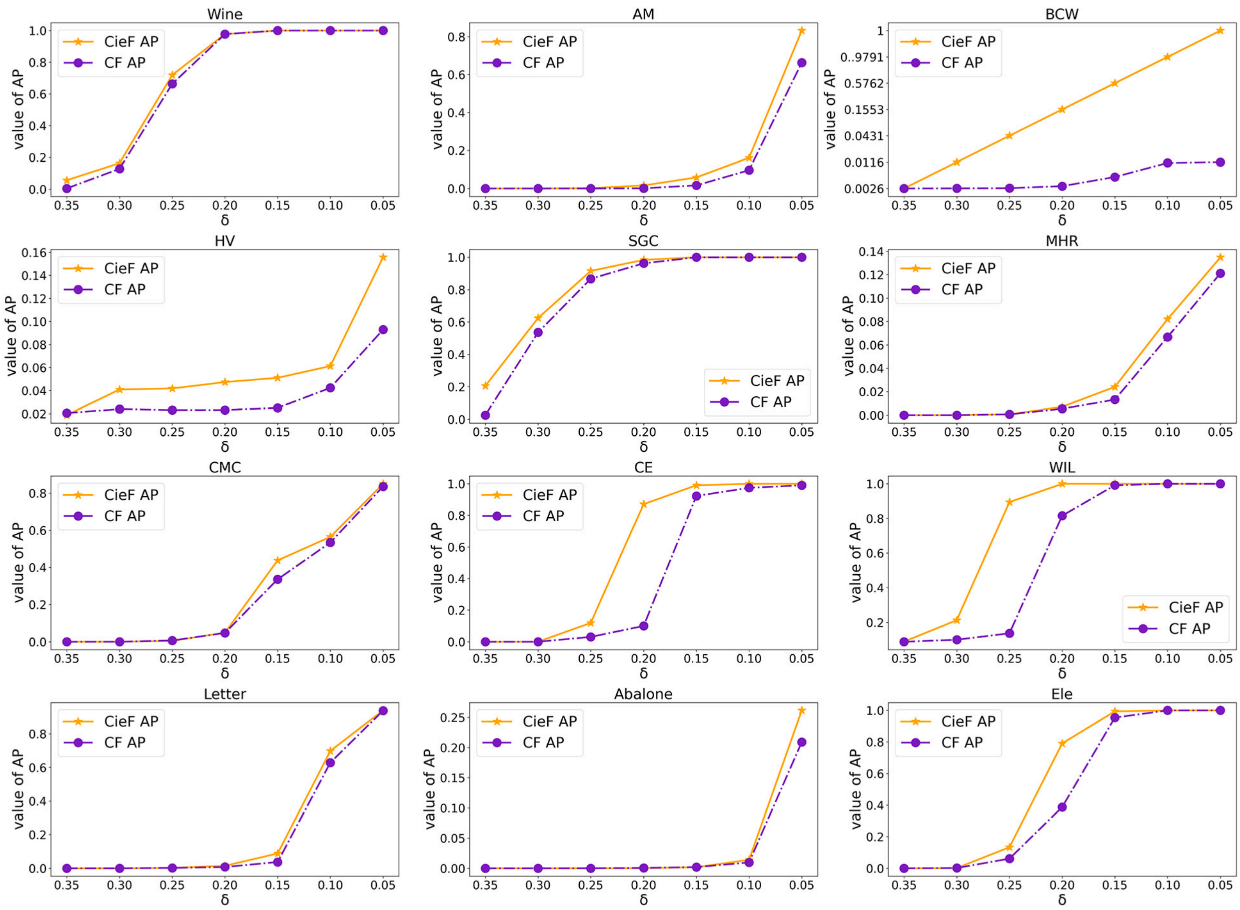


Fig. 5. The comparison results between CieF and CF AP.

Table 11
The classification accuracy of the fusion results of CF and CieF.

Datasets	KNN		Datasets	PNN	
	CF	CieF		CF	CieF
Wine (k = 3)	98.3±2.6	**100.0±0.0**	Wine (σ = 0.68)	89.9±6.5	**94.4±5.0**
AM (k = 14)	62.0±5.6	**63.3±5.1**	AM (σ = 0.01)	**50.8±7.7**	50.7±9.8
BCW (k = 5)	58.7±4.5	**62.2±5.3**	BCW (σ = 0.35)	61.7±2.8	**62.4±4.8**
HV (k = 9)	48.9±6.7	**51.7±6.0**	HV (σ = 0.54)	51.2±4.9	**52.0±6.0**
SGC (k = 9)	71.4±4.5	**73.0±2.7**	SGC (σ = 0.31)	**70.8±4.3**	70.6±3.8
MHR (k = 17)	37.3±2.8	**41.5±3.2**	MHR (σ = 0.39)	39.5±4.9	**40.1±5.5**
CMC (k = 4)	44.9±3.2	**55.5±3.6**	CMC (σ = 0.35)	47.5±2.7	**55.5±4.9**
CE (k = 16)	69.0±2.8	**70.8±2.5**	CE (σ = 0.21)	69.7±2.3	**70.4±1.9**
WIL (k = 3)	42.8±3.3	**55.3±3.7**	WIL (σ = 0.3)	47.3±3.0	**56.5±4.8**
Letter (k = 6)	96.7±0.7	**99.3±0.6**	Letter (σ = 0.3)	92.8±0.8	**93.9±0.8**
Abalone (k = 16)	**35.3±1.8**	35.1±1.7	Abalone (σ = 0.36)	36.6±2.4	**36.9±2.4**
Ele (k = 4)	59.9±2.8	**61.3±3.4**	Ele (σ = 0.02)	62.0± 2.8	**62.7±3.4**
Avg.	60.4±3.4	**64.1±3.2**	Avg.	60.0±3.8	**62.2±4.4**

Thus, compared with other four fusion ways, these verification outcomes indicate that the put forward fusion approach based on information structure is a better selection for the fusion of MsIIvDIS.

5.2. The analysis of fusion effectiveness (longitudinal comparison)

In order to verify the validity of the distance measurement used in the fusion method in this paper, we have employed various distance formulas for similarity calculation. And two classifiers are used to calculate the classification accuracy of the fusion results

Table 12
P-value test of classification accuracy based on KNN.

DataSets	KNN		
	CieF > MaxF	CieF > MinF	CieF > MeanF
Wine (k = 3)	**0.074457337**	**0.074457337**	0.172889293
AM (k = 14)	**0.061437694**	**0.020077856**	**0.048160338**
BCW (k = 5)	**0.094156345**	**0.037523868**	**0.087761675**
HV (k = 9)	0.186076915	**0.061602456**	0.539062500
SGC (k = 9)	**0.061272843**	0.883789063	0.735946937
MHR (k = 17)	0.312500000	**0.017143980**	**0.041992188**
CMC (k = 12)	**0.037523868**	**0.004882813**	**0.000976562**
CE (k = 16)	**0.007073702**	**0.014886438**	0.102951605
WIL (k = 3)	**0.080078125**	**0.032265625**	**0.006835938**
Letter (k = 6)	0.061602246	**0.052378746**	0.069518469
Abalone (k = 16)	**0.096679688**	0.296815296	0.361141481
Ele (k = 4)	**0.065429688**	0.116210938	0.142312144

Table 13
P-value test of classification accuracy based on PNN.

DataSets	PNN		
	CieF > MaxF	CieF > MinF	CieF > MeanF
Wine ($\sigma = 0.68$)	**0.0462957978**	**0.028953633**	0.100821461
AM ($\sigma = 0.01$)	0.312512659	0.96089559618	**0.042280712**
BCW ($\sigma = 0.35$)	**0.023896270**	0.900820236	0.425053370
HV ($\sigma = 0.39$)	**0.070558069**	0.170592481	**0.088764926**
SGC ($\sigma = 0.31$)	**0.016007410**	0.989931624	0.148347447
MHR ($\sigma = 0.31$)	0.573038350	**0.090724604**	0.823419823
CMC ($\sigma = 0.29$)	**0.000976563**	**0.004519554**	**0.000976563**
CE ($\sigma = 0.21$)	0.13647068	**0.067208287**	0.13647068
WIL ($\sigma = 0.2$)	**0.007073702**	**0.000976563**	**0.000976563**
Letter ($\sigma = 0.3$)	**0.004002155**	0.983821723	**0.036569900**
Abalone ($\sigma = 0.2$)	0.178636280	**0.023742347**	**0.028953633**
Ele ($\sigma = 0.02$)	**0.000976563**	0.240570078	**0.000976563**

Table 14
P-value test of the comparison results in classification accuracy between CieF and CF.

Datasets	KNN	Datasets	PNN
	H_1: CieF > CF		H_1: CieF > CF
Wine (k = 3)	**0.086784083**	Wine ($\sigma = 0.68$)	**0.017003201**
AM (k = 14)	0.399215982	AM ($\sigma = 0.01$)	0.593972318
BCW (k = 5)	**0.074662305**	BCW ($\sigma = 0.35$)	0.593809818
HV (k = 9)	0.146588729	HV ($\sigma = 0.54$)	**0.044943261**
SGC (k = 9)	0.117037300	SGC ($\sigma = 0.31$)	0.908788780
MHR (k = 17)	**0.018554688**	MHR ($\sigma = 0.2$)	0.111400496
CMC (k = 12)	**0.000976563**	CMC ($\sigma = 0.35$)	**0.000976563**
CE (k = 16)	**0.011247136**	CE ($\sigma = 0.21$)	**0.020653615**
WIL (k = 3)	**0.000976563**	WIL ($\sigma = 0.3$)	**0.000976563**
Letter (k = 6)	**0.032491572**	Letter ($\sigma = 0.3$)	**0.052801743**
Abalone (k = 16)	0.903320313	Abalone ($\sigma = 0.36$)	**0.025121460**
Ele (k = 4)	**0.008634689**	ELe ($\sigma = 0.02$)	0.141885979

Table 15
Friedman test.

Classifiers	Mean ranking					χ_F^2	F_F	P-value
	CieF	MaxF	MinF	MeanF	CF			
KNN	1.250	3.208	3.583	2.875	4.084	32.7025	23.5155	1.00×10^{-3}
PNN	1.500	4.125	2.333	3.292	3.75	31.6346	21.2632	2.81×10^{-4}

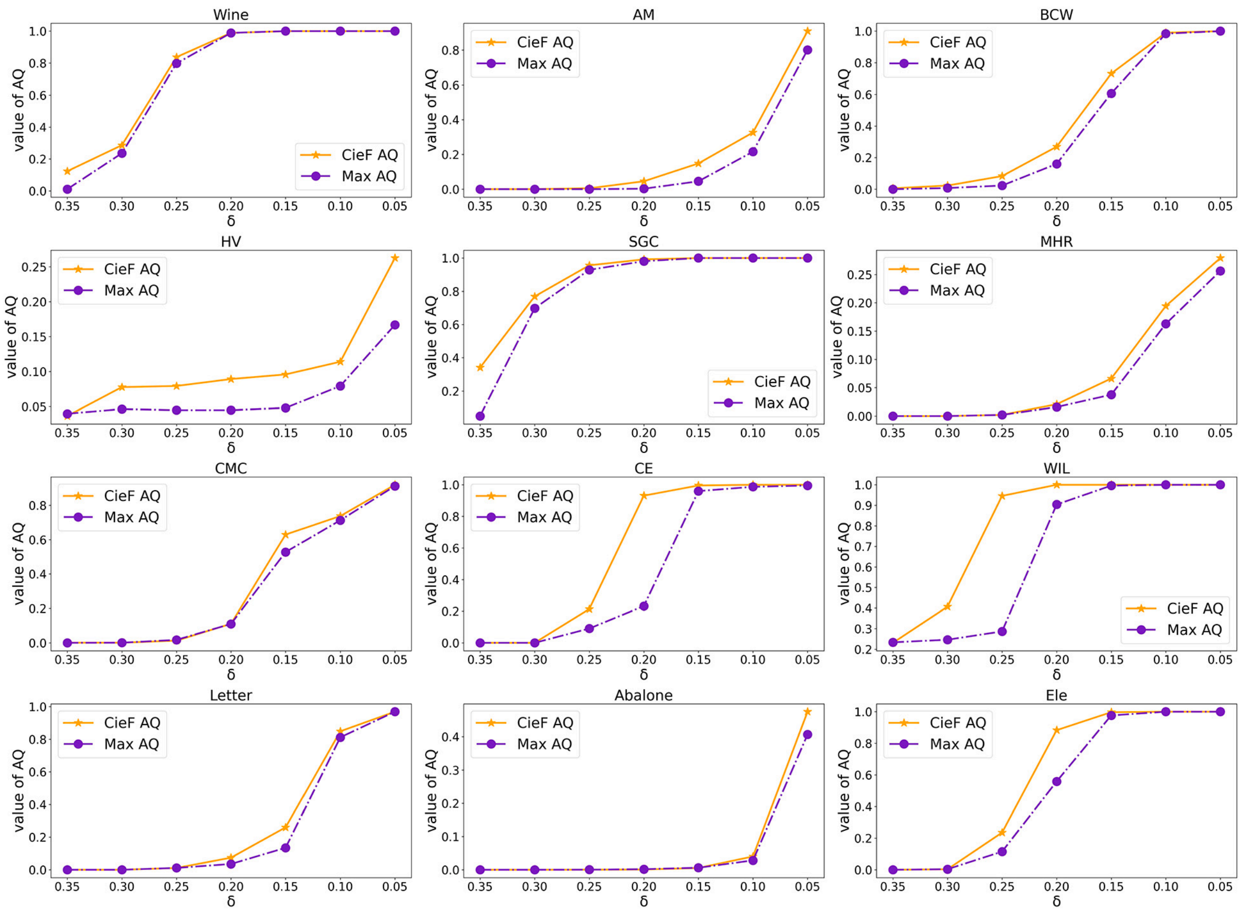


Fig. 6. The comparison results between CeF and CF AQ.

obtained by other distance measures. Obviously, we can find that the resulting fusion effect using other distance measurements is also different. Below, we present three typical distance metrics.

$$dis_1 = \begin{cases} 0 & x = * \ \ or \ \ y = * \\ \sqrt{(x^- - y^-)^2 + (x^+ - y^+)^2} & else, \end{cases}$$

$$dis_2 = \begin{cases} 0 & x = * \ \ or \ \ y = * \\ \sqrt{|x^- - y^-| + |x^+ - y^+|} & else, \end{cases}$$

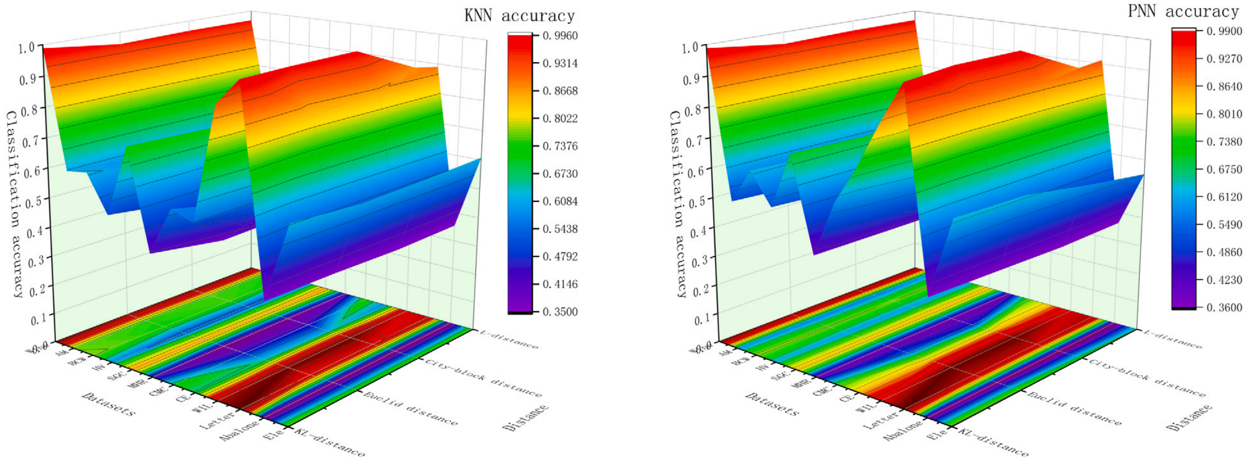
$$dis_3 = \begin{cases} 0 & x = * \ \ or \ \ y = * \\ \dfrac{\sqrt{|x^- - y^-|^2 + \left|\frac{x^- + x^+}{2} - \frac{y^- + y^+}{2}\right| + |x^+ - y^+|^2}}{3} & else, \end{cases}$$

where dis_1 is the Euclid distance, dis_2 is the City-block distance and dis_3 provided in [42] which can be written as L-distance. And the distance measurement used in this paper can be recorded as KL-distance.

5.2.1. Comparison of classification results based on different distance measurements

In this subsection, we primarily focus on validating whether the distance measurement used in this paper can reduce the loss of effective information caused by neglecting the contribution of interval values, thereby improving fusion effectiveness. For generality, we set the parameter to 0.35. Then, using different distance metrics to calculate the similarity of various interval values, we proceed with fusion. The classification accuracy of the fusion results is depicted in Fig. 7.

Clearly, the fusion classification results vary with different distance metrics for different datasets. Under the K-nearest neighbor classifier, for six datasets (Wine, AM, BCW, HV, MHR, and Abalone), the highest fusion accuracy is achieved when using the KL distance metric. For the SGC and CMC datasets, the fusion performance with KL distance surpasses that achieved with Euclidean distance and City-block distance. Concerning the WIL dataset, although the classification accuracy using KL divergence is not as superior as the other two distance measures, it still outperforms that of the L distance. In Fig. 7(b), employing the PNN classifier, KL-distance achieves the best classification results for the Wine, AM, BCW, HV, and Abalone datasets. However, concerning the MHR



(a) KNN (b) PNN

Fig. 7. The classification accuracy various with different distance measures in two classifiers.

Table 16
Friedman test.

Classifiers	Mean ranking				F_F	P-value
	KL-distance	Euclid distance	City-block distance	L-distance		
KNN	1.25	3.083	2.875	2.792	15.1842	0.0017
PNN	1.5	2.708	3.167	2.625	11.3947	0.0098

and CMC datasets, the fusion effect with KL-distance significantly outperforms that of Euclidean distance and City-block distance. For the SGC dataset, in comparison with the fusion effects of Euclidean distance and L distance, the fusion accuracy with KL distance is the highest. Overall, our model outperforms the other three distance measures in the fusion process.

5.2.2. Statistical analysis

In this subsection, we systematically explore the statistical performance of different distance measure fusion results in classification accuracy, and carry out Friedman test and corresponding post hoc test. The Friedman statistic is described [52] as:

$$\chi_F^2 = \frac{12N}{k(k+1)}\left(\sum_{j=1}^{k} R_j^2 - \frac{k(k+1)^2}{4}\right),$$

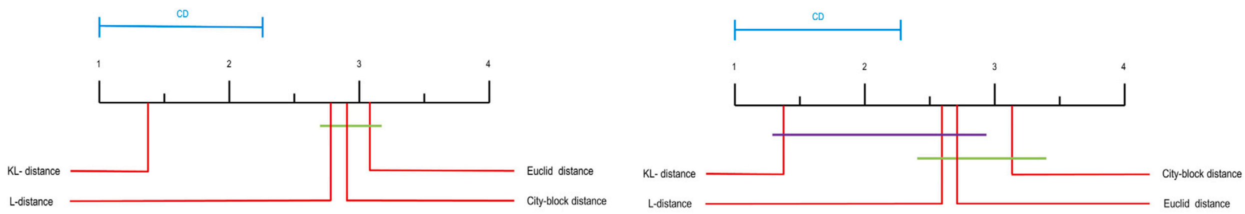
$$F_F = \frac{(N-1)\chi_F^2}{N(k-1)-\chi_F^2}.$$

Where N represents the number of data sets, while k represents the number of methods; R_j $(j = 1, 2, ..., k)$ represents the Average ranking of a certain approach on all data sets and F_F represents an F-distribution with $(k-1)$ and $(k-1)(N-1)$ degrees of freedom. Then the critical difference is expressed [53,54] as:

$$CD_\alpha = q_\alpha \sqrt{\frac{k(k+1)}{6N}},$$

here α expresses the significance level and q_α represents a critical value [53].

For all datasets, we conducted the following statistical test. We calculated the average ranking by taking the mean of the rankings based on classification accuracy. The top-performing result in terms of accuracy measurement was assigned a rank of 1, the second-best received a rank of 2, and so forth. Fig. 7 shows the changes of fusion classification accuracy of nine data sets under four different distance measures, the Friedman tests are accomplished by the comparison of this paper's distance with Euclid distance, City-block distance and L-distance. When all algorithms are equal in measures of classification accuracy, the null hypothesis of Friedman's test can be established. Then, the rankings of the four models can be lightly computed and their average order is acquired under the KNN and PNN. Thus, the values of χ_F^2 and F_F can be calculated. Table 16 shows the average sort results of the four distance models and the values of χ_F^2 and F_F under the classifier KNN and PNN. When the significance level is equal 10%. It follows from [52], by calculation, one has the critical point of $F(4-1,(4-1)*(12-1))$ in the F-distribution calculated to be 2.2577, and the critical point $q_{0.1}$ in the Nemenyi test is 2.291, the critical difference is 1.2075, that is, $CD = 1.2075$. So, all null hypotheses are refused, and the four distance measures are various under KNN and PNN.



(a) KNN (b) PNN

Fig. 8. Accuracy comparison with four distance measures on classifiers KNN and PNN.

In order to compare the differences between the fusion results under different measures more intuitively, we use CD critical charts [53] to connect methods that do not differ significantly from each other, and then in these graphs the critical values between all models can be clearly illustrated. Fig. 8 is the CD critical diagram which shows the comparison of the fusion result under KL-distance with the other three distance measures. As can be seen from Fig. 8, we can know the significant differences in fusion results under four different distance measures are prominent. In Fig. 8(a), under the classifier KNN, the average ranking of fusion effects using KL distance is the lowest, and the fusion result using KL-distance is clearly better than the other three distance measures. Likewise, as shown in Fig. 8(b) on the PNN, the fusion result using KL-distance outperforms the fusion result using City-block distance, and is similar to L-distance and Euclid distance. In conclusion, the fusion effect using KL-distance really outperforms the other three compared approach under the outcomes of the Friedman statistic test.

6. Conclusion

In this paper, a novel information entropy fusion approach based on information structure was put forward to improve the capability of classification for incomplete interval-valued decision systems. The definition of similarity was investigated based on new distance measurement making use of KL divergence. Then some uncertainty measures by using similarity were explored in incomplete interval-valued decision table. After that, we establish an infimum-measure function based new information entropy to fusion the MsIIvDIS. We gave the fusion algorithm and analyzed its time complexity of it. All the designed experiments demonstrated that our fusion approach can improve the capability of classification for low-dimensional small-sample data sets. In the meantime, we compared the fusion results using other three distance measures and carried out the statistical analysis, the experiments exhibit that The distance metric used in this article can improve the fusion effect. Nevertheless, our fusion approach is time-consuming for numerous large-scale and high-dimensional data sets. Thus, in our future discussed, the more effective fusion approaches and uncertainty measurements for the MsIIvDIS will be further discussed to enhance the computing efficiency of our model on numerous large-sample and high-dimensional data sets.

CRediT authorship contribution statement

Weihua Xu: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation. **Ke Cai:** Data curation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Debby D. Wang:** Investigation, Methodology, Software, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

This paper is supported by the National Natural Science Foundation of China (NO. 61976245) and Southwest University Post-graduate Research and Innovation Project (No. SWUS23130).

References

[1] Q. Pan, Multi-source Information Fusion Theory and Its Applications, Tsinghua University Press, 2013.
[2] D. Dubois, W.R. Liu, J.B. Ma, The basic principles of uncertain information fusion, Inf. Fusion 32 (2016) 12–39.
[3] X.Y. Zhang, D.D. Guo, W.H. Xu, Two-way concept-cognitive learning with multi-source fuzzy context, Cogn. Comput. 15 (2023) 1526?1548.
[4] C.Z. Han, H.Y. Zhu, Z.S. Duan, Multi-source Information Fusion, Tsinghua University Press, 2016.

[5] J. Llinas, E. Waltz, Multisensor Data Fusion, Artech House Publisher, 1992.
[6] Y. Wei, D. Wu, J. Terpenny, Decision-level data fusion in quality control and predictive maintenance, IEEE Trans. Autom. Sci. Eng. 18 (1) (2021) 184–194.
[7] X.K. Wang, L.T. Yang, X.Y. Chen, L.Z. Wang, R. Ranjan, X.D. Chen, M.J. Deen, A multi-order distributed HOSVD with its incremental computing for big services in cyber-physical-social systems, IEEE Trans. Big Data 6 (4) (2020) 666–678.
[8] W.H. Xu, J.H. Yu, Novel approach to information fusion in multi-source datasets: a granular computing viewpoint, Inf. Sci. 378 (2017) 410–423.
[9] Z. Pawlak, Rough sets and intelligent data analysis, Inf. Sci. 147 (2002) 1–12.
[10] W.H. Xu, D.D. Guo, J.S. Mi, et al., Two-way concept-cognitive learning via concept movement viewpoint, IEEE Trans. Neural Netw. Learn. Syst. 34 (10) (2023) 6798–6812, https://doi.org/10.1109/TNNLS.2023.3235800.
[11] D.D. Guo, C.M. Jiang, R.X. Sheng, S.S. Liu, A novel outcome evaluation model of tree-way decision: a change viewpoint, Inf. Sci. 607 (2022) 1089–1110.
[12] D.D. Guo, W.H. Xu, Fuzzy-based concept-cognitive learning: an investigation of novel approach to tumor diagnosis analysis, Inf. Sci. 639 (2023) 118998.
[13] W.H. Xu, D.D. Guo, Y.H. Qian, W.P. Ding, Two-way concept-cognitive learning method: a fuzzy-based progressive learning, IEEE Trans. Fuzzy Syst. (2022), https://doi.org/10.1109/TFUZZ.2022.3216110.
[14] D.D. Guo, W.H. Xu, Y.H. Qian, et al., Memory-based concept-cognitive learning for dynamic fuzzy data classification and knowledge fusion, Inf. Fusion 100 (2023) 101962.
[15] D.D. Guo, J.M. Jiang, P. Wu, Three-way decision based on confidence level change in rough set, Int. J. Approx. Reason. 143 (2022) 57–77.
[16] J. Han, Y. Cai, N. Cercone, Data-driven discovery of quantitative rules in relational databases, IEEE Trans. Knowl. Data Eng. 5 (1) (1993) 29–40.
[17] C.C. Chan, A rough set approach to attribute generalization in data mining, Inf. Sci. 107 (1) (1998) 169–176.
[18] X. Hu, H. Zhang, C.M. Yang, X.J. Zhao, B. Li, Regularized spectral clustering with entropy perturbation, IEEE Trans. Big Data 7 (6) (2021) 967–972.
[19] Q.C. Zhang, L.T. Yang, Z.K. Chen, P. Li, Incremental deep computation model for wireless big data feature learning, IEEE Trans. Big Data 6 (2) (2020) 248–257.
[20] G.J. Dong, Y.S. Zhang, C.G. Dai, Y.H. Fan, The processing of information fusion based on rough set theory, J. Sci. Instrum. 26 (2005) 570–571.
[21] X.J. Lv, Y. Zhao, G.S. Yao, Q.C. Lv, N. Wang, Multi-sensor information fusion based on rough set theory, Comput. Eng. Syst. Appl. 1 (1) (2006) 28–30.
[22] X.W. Chen, W.H. Xu, Double-quantitative multi-granulation rough fuzzy set based on logical operations in multi-source decision systems, Int. J. Mach. Learn. Cybern. 13 (4) (2021) 1021–1048.
[23] G.P. Lin, J.Y. Liang, Y.H. Qian, An information fusion approach by combining multigranulation rough sets and evidence theory, Inf. Sci. 314 (2015) 184–199.
[24] P.F. Zhang, T.R. Li, Y. Zhong, L. Chuan, A data-level fusion model for unsupervised attribute selection in multi-source homogeneous data, Inf. Fusion 80 (2022) 87–103.
[25] L. Yang, W.H. Xu, X.Y. Zhang, B.B. Sang, Multi-granulation method for information fusion in multi-source decision information system, Int. J. Approx. Reason. 122 (2020) 47–65.
[26] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning About Data, Springer, Heidelberg, Germany, 1991.
[27] Y.H. Qian, J.Y. Liang, C.Y. Dang, Fuzzy information granularity in binary granular structure, IEEE Trans. Fuzzy Syst. 19 (2) (2021) 253–260.
[28] C.Z. Wang, H. Yang, M.W. Shao, Uncertainty measures for general fuzzy relations, Fuzzy Sets Syst. 360 (2018) 82–96.
[29] S. Yao, J. Chen, Z.Y. Wu, Combination measurement method based on neighborhood tolerance information entropy, J. Chin. Comput. Syst. 40 (1) (2020) 46–50.
[30] K.H. Yuan, W.H. Xu, W.T. Li, W.P. Ding, An incremental learning mechanism for object classification based on progressive fuzzy three-way concept, Inf. Sci. 584 (1) (2022) 127–147.
[31] W.H. Wei, K.H. Yuan, W.T. Li, Dynamic updating approximations of local generalized multigranulation neighborhood rough set, Appl. Intell. 52 (8) (2022) 9148?9173.
[32] C.E. Shannon, W. Weaver, The mathematical theory of communication, Bell Syst. Tech. J. 27 (1948) 373–423.
[33] W.H. Xu, M.M. Li, X.Z. W, Information fusion based on information entropy in fuzzy multi-source incomplete information system, Int. J. Fuzzy Syst. 19 (2017) 1200–1216.
[34] P.F. Zhang, T.R. Li, G.Q. Wang, C. Luo, et al., Multi-source information fusion based on rough set theory: a review, Inf. Fusion 68 (2021) 85–117.
[35] P.F. Zhang, T.R. Li, Z. Yuan, Z.X. Deng, et al., A possibilistic information fusion-based unsupervised feature selection method using information quality measures, IEEE Trans. Fuzzy Syst. (2023) 1–14.
[36] W.H. Xu, K.H. Yuan, W.T. Li, W.P. Ding, An emerging fuzzy feature selection method using composite entropy-based uncertainty measure and data distribution, IEEE Trans. Emerg. Top. Comput. Intell. 7 (1) (2023) 76–88.
[37] G.Q. Zhang, W.L. Zhao, W.Z. Wu, Information structures and uncertainty measures in a fully fuzzy information system, Int. J. Approx. Reason. 101 (2018) 119–149.
[38] Y.H. Qian, J.Y. Liang, C.Y. Dang, Interval ordered information system, Comput. Math. Appl. 56 (8) (2008) 1994–2009.
[39] X.F. Liu, J.H. Dai, J.L. Chen, C.C. Zhang, Unsupervised attribute reduction based on α-approximate equal relation in interval-valued information systems, Int. J. Mach. Learn. Cybern. 11 (2020) 2021–2038.
[40] N.X. Xie, M. Liu, Z.W. Li, G.Q. Zhang, New measures of uncertainty for an interval-valued information system, Inf. Sci. 470 (2019) 156–174.
[41] R.R. Yager, A framework for multi-source data fusion, Inf. Sci. 163 (13) (2004) 175–200.
[42] X.B. Xu, Z. Zhang, D.L. Xu, Y.W. Chen, Interval-valued evidence updating with reliability and sensitivity analysis for fault diagnosis, Int. J. Comput. Intell. Syst. 9 (3) (2016) 396–415.
[43] Y.Y. Huang, T.R. Li, C. Luo, H. Fujita, S.J. Horng, Dynamic fusion of multisource interval-valued data by fuzzy granulation, IEEE Trans. Fuzzy Syst. 26 (6) (2018) 3403–3417.
[44] W.H. Xu, Y.Z. Pan, X.W. Chen, Y.H. Qian, W.P. Ding, A novel dynamic fusion approach using information entropy for interval-valued ordered datasets, IEEE Trans. Big Data (2022), https://doi.org/10.1109/TBDATA.2022.321549.
[45] J.H. Dai, B.J. Wei, X.H. Zhang, Q.L. Zhang, Uncertainty measurement for incomplete interval-valued information systems based on α-weak similarity, Knowl.-Based Syst. 136 (2017) 159–171.
[46] C. Luo, T.R. Li, Z. Yi, An incremental feature selection approach based on information entropy for incomplete data, in: 2019 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress, DASC/PiCom/CBDCom/CyberSciTech, Fukuoka, Japan, 2019, pp. 483–488.
[47] H. Zhao, K.Y. Qin, Mixed feature selection in incomplete decision table, Knowl.-Based Syst. 57 (2014) 181–190.
[48] H.X. Li, M.H. Wang, X.Z. Zhou, B. Zhao, An interval set model for learning rules from incomplete information table, Int. J. Approx. Reason. 53 (1) (2012) 24–37.
[49] S. Han, L. Chen, J.X. Li, A new information filling technique based on generalized information entropy, Int. J. Comput. Commun. 9 (2) (2014) 172–186.
[50] X.Y. Zhang, X.W. Chen, W.H. Xu, W.P. Ding, Dynamic information fusion in multi-source incomplete interval-valued information system with variation of information sources and attributes, Inf. Sci. 608 (2002) 1–27.
[51] M.M. Li, X.Y. Zhang, Information fusion in a multi-source incomplete information system based on information entropy, Entropy 19 (11) (2017) 570–586.
[52] M. Friedman, A comparison of alternative tests of significance for the problem of m rankings, Ann. Math. Stat. 11 (1) (1940) 86–92.
[53] J. Demsar, D. Schuurmans, Statistical comparison of classifiers over multiple data sets, J. Mach. Learn. Res. 7 (2006) 1–30.
[54] L. Sun, L.Y. Wang, W.P. Ding, Y.H. Qian, J.C. Xu, Feature selection using fuzzy neighborhood entropy-based uncertainty measures for fuzzy neighborhood multigranulation rough sets, IEEE Trans. Fuzzy Syst. 29 (1) (2021) 19–33.