Information Sciences 667 (2024) 120479

Contents lists available at ScienceDirect

Information Sciences

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

Incremental information fusion in the presence of object variations for incomplete interval-valued data based on information entropy

Xiuwei Chen, Maokang Luo *

School of Mathematics, Sichuan University, Chengdu 610064, PR China

ARTICLE INFO

Keywords:
Incomplete interval-valued
Information fusion
Object variations
Rough set

ABSTRACT

Information fusion technology plays a crucial role in integrating data from multiple sources or sensors to generate comprehensive representation, which can eliminate uncertainty in multi-source information systems (Ms-IS). Incomplete interval-valued data, a generalized form of single-valued data, is commonly encountered in real-world scenarios and effectively represents uncertain information. This paper introduces a novel information entropy specifically designed to quantify the uncertainty in incomplete interval-valued data. Based on the proposed entropy, a new unsupervised fusion approach is developed. Additionally, two dynamic update mechanisms are established to obtain fusion results efficiently when collecting new objects and removing obsolete ones. The relevant static and dynamic fusion algorithms are provided, and a detailed analysis and comparison of their time complexities are conducted. Finally, the effectiveness analysis reveals that the proposed method achieves higher average classification accuracy (5% to 8.7% improvement) compared to three common fusion methods (MAXF, MEANF, and MINF), as well as the state-of-the-art entropy-based supervised fusion method (ESF). The efficiency analysis demonstrates that the average running time of dynamic fusion algorithms is significantly lower (66.9% to 85.6% reduction) compared to the static fusion algorithm, and this difference is statistically significant.

1. Introduction

In real-world scenarios, data from multiple sources or sensors often contains incompleteness and uncertainty. Interval-valued data serves as a suitable representation for data with unknown exact values, expressed in the form of intervals. The interval representation provides an estimate of the data range and effectively captures the inherent uncertainty. The applications of interval-valued data span various domains, including prediction [1,2], knowledge acquisition [3,4], and feature selection [5,6]. Information fusion technology is a process that involves integrating data, knowledge, or features from diverse information sources to generate comprehensive and reliable information or decision results [7–9]. By amalgamating information from different sources, information fusion technology can effectively handle uncertain and inaccurate information. Data fusion, a key aspect of mainstream information fusion technology, integrates data from multiple sources to provide a comprehensive perspective, eliminating redundancy, filling data gaps, and creating

* Corresponding author.
E-mail addresses: xiuweichen1998@163.com (X. Chen), makaluo@scu.edu.cn (M. Luo).

https://doi.org/10.1016/j.ins.2024.120479
Received 17 September 2023; Received in revised form 29 January 2024; Accepted 13 March 2024

Available online 19 March 2024
0020-0255/

reliable datasets for analysis and decision-making. In recent years, there has been significant research interest in data fusion for complex data collected from multiple information sources or sensors [10–14].

Rough set theory (RST), initially proposed by Pawlak [15], provides a framework to cope with incomplete [16], fuzzy [17], and uncertain information [18]. The fundamental principle of RST is to granulate the sample space through the establishment of sample relationships, serving as a mechanism to tackle intricate problems. Leveraging these information granules, the theory effectively simplifies the intrinsic complexity of problems, providing a valuable framework for the management and analysis of data characterized by incompleteness, fuzziness, and uncertainty. In recent years, an increasing number of fusion approaches based on RST and information entropy(IE) [19] have emerged. In these approaches, RST is employed to granulate complex data, while IE is used to measure the uncertainty of information sources. In 2017, Li et al. [20] proposed a novel fusion approach for incomplete decision systems (IDS) by defining the conditional entropy of incomplete data. An example in the context of medical examinations demonstrated the effectiveness of this method. Subsequently, Xu et al. [21] developed an entropy-based fusion method for incomplete fuzzy data, which was applied to ground object recognition problems. This method selects each attribute's most crucial information source by minimizing information entropy. In 2022, Zhang et al. [10] introduced a novel uncertainty measurement for incomplete interval-valued data and applied it to select the essential source with respect to attributes in multi-source incomplete interval-valued information systems (Ms-IIvIS). Similarly, in the same year, Xu et al. [11] proposed a novel fusion approach for multi-source interval-valued ordered information systems (Ms-IvOIS) by defining the uncertainty measurement of ordered data.

The motivation for this research stems from two key aspects. On the one hand, many existing fusion methods struggle to handle unlabeled incomplete interval-valued data [10,11,20,21]. These methods aim to define conditional entropy to select sources, thereby they are applicable only to labeled data. For unlabeled data, a novel Sup-Inf fusion framework is established in [22] for homogeneous data, capable of simultaneously handling categorical, numerical, and incomplete data. However, this method can not deal with Ms-IIvIS, which widely exists in real applications. For instance, in medical diagnostics, many diagnostic data are presented as interval values to reflect changes in a patient's body over a certain period. Moreover, certain indicator values may be missing due to machine malfunctions or improper operations. Consequently, there is a critical need to develop a method capable of effectively fusing Ms-IIvIS.

Another motivation behind this research is the observation that numerous established fusion methods often overlook the consideration of dynamic information, leading to low computational efficiency when dealing with dynamic data [22–27]. Dynamic fusion technology enables real-time analysis of sources, attributes, and objects, resulting in improved computational efficiency when handling dynamic data. Therefore, investigating dynamic fusion technology for Ms-IS is a crucial research direction. To address the variation of sources, Huang et al. [28] proposed two incremental update mechanisms for obtaining new sources and removing obsolete sources. In 2022, considering simultaneous changes in information sources and attributes, Xu et al. [11] and Zhang et al. [10] established two kinds of incremental fusion approaches for Ms-IvOIS and Ms-IIvIS, respectively. These dynamic fusion approaches mentioned above do not account for object changes. When new objects are collected or obsolete objects are removed from information systems, these methods will become ineffective. To address the variation of objects, Huang et al. [29] proposed an approximation updating method of multi-source hybrid data based on matrix operators. Later, Luo et al. [30] established an incremental fusing framework for dynamic incomplete data. However, the above methods can not deal with Ms-IIvIS directly. In Ms-IIvIS, the objects may undergo continuous dynamic changes. For instance, different detection reports from multiple hospitals or medical institutions are needed in medical examination scenarios to enhance the reliability of diagnosing the same group of patients. Every day, new patients undergo examinations, and certain patient samples may need to be excluded from the analysis due to issues like reagent contamination or equipment malfunctions to maintain the accuracy of judgments. Therefore, investigating dynamic fusion methods that specifically cope with object changes in Ms-IIvIS holds significant practical relevance.

Motivated by the limitations of existing fusion methods in handling Ms-IIvIS and addressing object variations, this paper proposes a novel unsupervised fusion method for Ms-IIvIS based on information entropy. The proposed method aims to select the most critical source with minimum entropy and utilizes the unions of tolerance classes (TC) of missing value objects to complete the fused results. Additionally, two dynamic update approaches are developed to handle object variations, thereby enhancing the fusion method's practical relevance and computational efficiency in dynamic data scenarios. The main innovations and contributions of this paper are summarized as follows: (1)A novel information entropy is defined to measure the uncertainty of incomplete interval-valued data. The properties of this entropy are studied and proven. (2)An unsupervised fusion framework for Ms-IIvIS is established based on the proposed information entropy, which is illustrated in Fig. 1. This framework employs information entropy for source selection, generates a fused information table, and utilizes TC of missing value objects to fill in the fusion results. (3)Two dynamic update mechanisms are introduced to handle object insertion or removal in Ms-IIvIS, with the corresponding algorithms provided in Algorithms 2-3. The time complexity analysis, presented in Table 18, demonstrates the superior performance of the dynamic fusion algorithms over the static fusion algorithm. (4)Experimental evaluations conducted on UCI datasets[1] reveal the superiority of the proposed fusion approach over the other four methods in terms of classification accuracy and also demonstrate that the dynamic update mechanisms can efficiently reduce the overall running time.

The structure of the paper is as follows: In Section 2, some concepts and definitions are listed and explained. Section 3 establishes an unsupervised fusion framework based on information entropy, and the relevant algorithm is provided. Section 4 proposes two dynamic fusion mechanisms for handling the insertion of new objects or the removal of obsolete objects from Ms-IIvIS, and the relevant algorithms are provided. Then, Section 5 conducts experiments on UCI datasets to evaluate the effectiveness and efficiency of the proposed fusion method. Conclusively, Section 6 summarizes the study and its main findings and discusses the identified

[1] http://archive.ics.uci.edu.

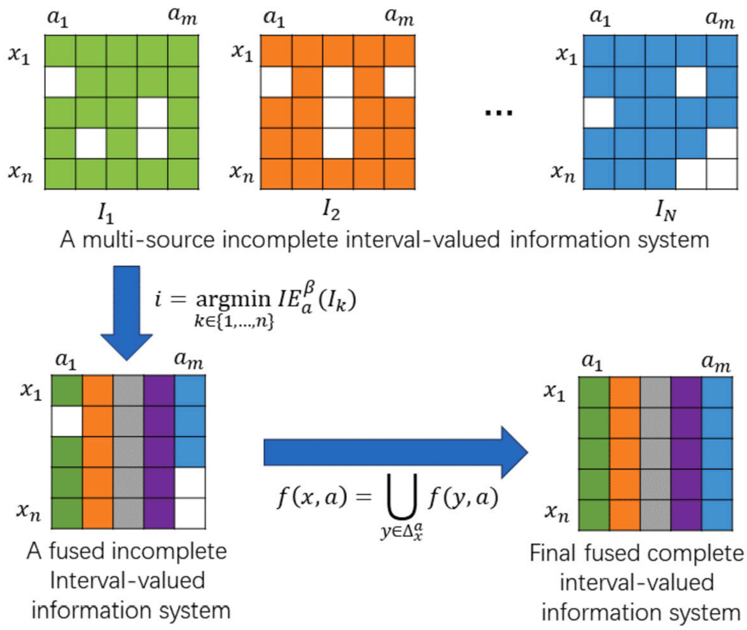

Fig. 1. The fusion framework of the proposed approach.

Table 1
Abbreviations of terms.

Index	Terms	Abbreviations
1	Approximation classification precision	AP
2	Information entropy	IE
3	Incomplete decision systems	IDS
4	Incomplete interval-valued information system	IIvIS
5	Multi-source information system	Ms-IS
6	Multi-source incomplete interval-valued information system	Ms-IIvIS
7	Multi-source interval-valued ordered information system	Ms-IvOIS
8	Rough set theory	RST
9	Speed-up ratio	SUR
10	Tolerance class	TC

limitations of the research while outlining future research directions. Additionally, Table 1 is provided to list and clarify all the terms and abbreviations used in this paper for ease of reading.

2. Preliminaries

This section will introduce some basic concepts about IDS, IIvIS, and Ms-IIvIS.

2.1. Incomplete decision system [31,32]

An IDS can be defined as $IDS = (U, A, V_A, f_A, d, V_d, f_d)$, where $U = \{x_1, \dots, x_n\}$ is an object set, $A = \{a_1, \dots, a_m\}$ is a conditional attributes set, V_A is the domain of A, and $f_A : U \times A \to V_A$ is an information function and there exist $a \in A$ such that $f_a(x, a) = *$ (* denotes missing values). Furthermore, d denotes decision attribute, V_d denotes the domain of d, and f_d denotes the information function of d.

Definition 2.1. Given an IDS $IDS = (U, A, V_A, f_A, d, V_d, f_d)$. For any $a \in A$, $x, y \in U$, the tolerance relation under a is denoted by [31,32]

$$Tr_a = \{(x, y) \in U \times U \mid f(x, a) = f(y, a) \text{ or } f(x, a) = * \text{ or } f(y, a) = * \}. \tag{1}$$

Furthermore, the toleration relation under $B \subseteq A$ is defined as $Tr_B = \bigcap_{b\in B} Tr_b$.

Definition 2.2. Given an IDS $IDS = (U, A, V_A, f_A, d, V_d, f_d)$. For any $B \subseteq A$, the lower and upper approximations under $X \subseteq U$ are defined by [31,32]

Table 2
The medical examination result of the first hospital.

U	a_1	a_2	a_3	a_4	a_5	a_6
x_1	[128.26,139.97]	[1.00,5.00]	[115.87,124.98]	*	*	[70.18,93.23]
x_2	[118.40,130.97]	[3.99,11.00]	[116.96,124.99]	[79.48,118.98]	*	*
x_3	[108.53,119.97]	[2.79,10.00]	[112.38,121.98]	[119.07,179.97]	[66.28,84.98]	[78.18,100.32]
x_4	[125.11,133.97]	[1.99,9.00]	[111.48,120.98]	[60.78,98.98]	*	[66.38,89.17]
x_5	[126.69,135.97]	[4.39,11.00]	[111.98,295.95]	[81.58,119.98]	[139.37,282.95]	[43.89,74.99]
x_6	[126.29,214.96]	[6.29,16.00]	*	*	[68.28,87.98]	[27.99,61.81]
x_7	[123.04,196.96]	[4.49,9.00]	[177.26,268.95]	[84.28,153.97]	[77.98,97.98]	[30.09,60.80]
x_8	[158.85,233.95]	[10.09,20.00]	*	[102.28,162.97]	[65.28,84.98]	*
x_9	[118.40,149.97]	[11.99,22.00]	[224.96,268.95]	[90.28,150.97]	*	[77.98,89.17]
x_{10}	[141.09,214.96]	[8.39,18.00]	[177.17,267.95]	[67.98,118.98]	[109.38,252.96]	[39.89,70.93]

Table 3
The medical examination result of the second hospital.

U	a_1	a_2	a_3	a_4	a_5	a_6
x_1	[129.07,139.99]	[3.00,7.00]	[115.89,124.99]	[68.50,107.99]	[43.40,186.99]	[70.20,92.66]
x_2	*	[4.00,11.00]	[116.99,124.99]	[79.50,118.99]	[80.40,223.99]	[67.20,88.63]
x_3	*	[2.80,10.00]	[112.39,121.99]	[119.09,179.99]	*	[78.20,99.71]
x_4	[125.89,133.99]	[2.00,9.00]	[111.49,120.99]	[60.80,98.99]	[99.99,259.99]	[66.40,88.63]
x_5	[127.48,135.99]	[4.40,11.00]	[111.99,295.99]	*	*	[43.90,74.53]
x_6	[127.08,214.99]	[6.30,16.00]	[204.29,295.99]	[81.00,162.99]	[68.30,88.00]	[28.00,61.44]
x_7	[123.81,196.99]	[4.50,9.00]	[177.29,268.99]	[84.30,153.99]	[78.00,98.00]	[30.10,60.43]
x_8	*	[10.10,20.00]	[224.09,314.98]	[102.30,162.99]	*	[25.80,56.40]
x_9	[119.14,149.99]	[12.00,22.00]	[224.99,268.99]	[90.30,150.99]	[71.30,91.00]	[78.00,88.63]
x_{10}	*	[8.40,18.00]	[177.19,267.99]	*	[109.39,252.99]	[39.90,70.50]

$$\underline{Tr_B}(X) = \left\{ x \left| Tr_B(x) \subseteq X \right. \right\}, \qquad \overline{Tr_B}(X) = \left\{ x \left| Tr_B(x) \cap X \neq \emptyset \right. \right\}, \tag{2}$$

where $Tr_B(x) = \left\{ y \left| (x,y) \in Tr_B \right. \right\}$.

Definition 2.3. Given an IDS $IDS = (U, A, V_A, f_A, d, V_d, f_d)$. Let $U/d = \left\{ Y_1, Y_2, ..., Y_m \right\}$ be the decision partition of U, where $Y_m = \left\{ x \in U \left| f_d(x,d) = y_m \right. \right\}$ denotes the collection of objects for the m-th decision label y_m. For any $B \subseteq A$, the approximation classification precision with respect to decision attribute d is defined by [33]

$$AP_{Tr_B}(U/d) = \frac{\sum_{m=1}^{|U/d|} \left| \underline{Tr_B}(Y_m) \right|}{\sum_{m=1}^{|U/d|} \left| \overline{Tr_B}(Y_m) \right|}, \tag{3}$$

where $|U/d|$ denotes the cardinality of the decision partition U/d.

In RST, the approximation classification precision (AP) can represent the approximation classification performance of models. The higher the AP value, the better the approximation classification performance.

2.2. Multi-source incomplete interval-valued information systems [10]

Based on the definition of IDS, an IIvIS can be denoted as $IIvIS = \left(U, A, V_A, f_A \right)$, where $U = \{x_1, ..., x_n\}$, $A = \{a_1, ..., a_m\}$, and $f(x,a) = [f^L(x,a), f^U(x,a)]$ or $*$ (missing values), where $f^L(x,a)$ and $f^U(x,a)$ represent the left and right endpoints of $f(x,a)$, respectively.
Furthermore, a Ms-IIvIS can be defined as

$$Ms - IIvIS = \left\{ I_q \left| I_q = (U, A_q, V_q, f_{A_q}), q = 1, ..., N \right. \right\},$$

where $I_q = IIvIS_q$ is the q-th subsystem of Ms-IIvIS.

Example 2.1. [10] Tables 2–5 are four IIVISs denoting examination results from 10 patients to 4 hospitals. The objects x_1 - x_{10} denote ten patients, and the attributes a_1 - a_6 denote six blood routine examination indicators, respectively. The decision attribute is Leukemia or not, and $V_d = \left\{ y_1, y_2 \right\}$, where y_1 and y_2 denote Leukemia patient and Non leukemia patient, respectively. The decision partition is denoted as $U/d = \left\{ Y_1, Y_2 \right\}$, where $Y_1 = \left\{ x_1, x_2, x_6, x_8, x_9 \right\}$ and $Y_2 = \left\{ x_3, x_4, x_5, x_7, x_{10} \right\}$.

Table 4
The medical examination result of the third hospital.

U	a_1	a_2	a_3	a_4	a_5	a_6
x_1	*	[3.00,7.00]	[115.87,124.97]	*	[43.39,186.96]	[70.18,93.42]
x_2	[118.15,130.97]	*	*	[79.48,118.97]	[80.38,223.95]	[67.18,89.36]
x_3	*	[2.80,10.00]	[112.37,121.97]	*	[66.28,84.98]	[78.18,100.53]
x_4	*	[2.00,9.00]	[111.47,120.97]	[60.79,98.98]	*	[66.38,89.36]
x_5	[126.42,135.97]	[4.40,11.00]	[111.97,295.93]	[81.58,119.97]	[139.37,282.93]	[43.89,74.14]
x_6	[126.02,214.9]	[6.30,16.00]	[204.25,295.93]	[80.98,162.96]	[68.28,87.98]	[27.99,61.94]
x_7	[122.77,196.95]	[4.50,9.00]	[177.26,268.94]	[84.28,153.96]	[77.98,97.98]	[30.09,60.93]
x_8	[158.51,233.94]	[10.10,20.00]	[224.05,314.93]	[102.28,162.96]	[65.28,84.98]	[25.79,56.87]
x_9	[118.15,149.96]	[12.00,21.99]	[224.95,268.94]	*	[71.28,90.98]	*
x_{10}	[141.79,214.95]	[8.40,18.00]	[177.16,267.94]	[67.98,118.97]	[109.37,252.94]	[39.89,71.08]

Table 5
The medical examination result of the fourth hospital.

U	a_1	a_2	a_3	a_4	a_5	a_6
x_1	*	[3.00,7.00]	[115.90,125.00]	*	*	[70.20,92.01]
x_2	[119.99,131.00]	[4.00,11.00]	[117.00,125.00]	[79.50,119.00]	*	[67.20,88.01]
x_3	[109.99,112.00]	*	[112.40,121.98]	[119.10,180.00]	[66.30,85.00]	[78.20,99.01]
x_4	[126.79,134.00]	[2.00,9.00]	[111.48,120.98]	[60.80,99.00]	[100.00,260.00]	[66.40,88.01]
x_5	[128.39,136.00]	[4.40,11.00]	*	[81.60,120.00]	[139.40,283.00]	[43.90,74.01]
x_6	[127.99,215.00]	[6.30,16.00]	[204.30,296.00]	[81.00,163.00]	[68.30,88.00]	[28.00,61.01]
x_7	*	[4.50,9.00]	[177.30,269.00]	[84.30,154.00]	[78.00,98.00]	[30.10,60.01]
x_8	*	[10.10,20.00]	[224.10,315.00]	[102.30,163.00]	[65.30,85.00]	[25.80,56.00]
x_9	[119.99,150.00]	[12.00,22.00]	[225.00,269.00]	[90.30,151.00]	[71.30,91.00]	[78.00,88.01]
x_{10}	[142.99,215.00]	*	[177.20,268.00]	[68.00,119.00]	[109.40,253.00]	*

3. Unsupervised information fusion in Ms-IIvIS based on information entropy

In the era of big data, it is common to encounter data that originates from multiple sources, contains uncertainties, and even has missing values. Moreover, obtaining sample labels can be costly in terms of manpower and resources under some circumstances. Therefore, it is of great practical significance to study how to fuse Ms-IIvIS effectively. This section focuses on the development of an unsupervised fusion approach for Ms-IIvIS, providing a solution to fuse the information from different sources without relying on sample labels. The relevant algorithm for the unsupervised fusion approach is provided, and the time complexity of it is analyzed.

Definition 3.1. Let $IIvIS = \left(U, A, V_A, f_A\right)$ be an IIvIS. $\forall a \in A$, the distance between any two objects $x,\ y \in U$ is calculated by:

$$d_a(x,\ y) = \begin{cases} 0, & \text{if } f(x,a) = * \ \text{ or } f(y,a) = *, \\ \sqrt{(f^L(x,a) - f^L(y,a))^2 + (f^U(x,a) - f^U(y,a))^2}, & \text{else}, \end{cases} \tag{4}$$

where $*$ denotes the missing value of the $IIvIS$.

Definition 3.2. Let $IIvIS = \left(U, A, V_A, f_A\right)$ be an IIvIS. $\forall a \in A$, the tolerance relation is defined by:

$$Tr_a^{\beta} = \left\{ (x, y) \in U \times U \,\middle|\, \frac{d_a(x,y)}{\max\limits_{z \in U} d_a(x,z)} \le \beta \ \text{ or } f(x,a) = * \ \text{ or } \ f(y,a) = * \right\}, \tag{5}$$

where $\beta \in [0,1]$ is an adjustable threshold. Furthermore, the TC of x with respect to a is defined as $Tr_a^{\beta}(x) = \left\{ y \,\middle|\, (x,y) \in Tr_a^{\beta} \right\}$.

In the context of data analysis, the establishment of the tolerance relationship plays a pivotal role in assessing the similarity between objects. For each pair of objects $(x,\ y)$, object x is considered to have a relationship with object y if the distance between object y and object x is less than or equal to β times the distance between the object furthest from object x and object x. Additionally, the tolerance relation is also established if either object x or object y is a missing value object. This tolerance relationship implies that the two objects, x and y, share similar characteristics. By harnessing this relationship, one can effectively identify and retrieve analogous samples from the object set.

Proposition 3.1. *Let $a \in A$ and $x,\ y,\ z \in U$, the tolerance relation Tr_a^{β} is subject to the following properties:*

(1) $(x,x) \in Tr_a^{\beta}$,

(2) $(x,y) \in Tr_a^{\beta} \Rightarrow (y,x) \in Tr_a^{\beta}$ *or* $(y,x) \notin Tr_a^{\beta}$,

(3) $(x,y) \in Tr_a^\beta, (y,z) \in Tr_a^\beta \Rightarrow (x,z) \in Tr_a^\beta$ *or* $(x,z) \notin Tr_a^\beta$,
(4) $\beta_1 \le \beta_2 \Rightarrow Tr_a^{\beta_1} \subseteq Tr_a^{\beta_2}$.

Proof. (1) The property is straightforward because for any object x, $d(x,x)=0$.
(2) When x or y is missing value, we can get that $(x,y) \in Tr_a^\beta \Rightarrow (y,x) \in Tr_a^\beta$ because the distance between missing values and any objects is zero. However, when x and y are not missing values, we can find that $\max\limits_{m\in U} d_a(x,m) \le \max\limits_{m\in U} d_a(y,m)$ or $\max\limits_{m\in U} d_a(x,m) > \max\limits_{m\in U} d_a(y,m)$, so $\frac{d_a(y,x)}{\max_{m\in U} d_a(y,m)} \le \beta$ or $\frac{d_a(y,x)}{\max_{m\in U} d_a(y,m)} > \beta$. Thus, we have $(y,x) \in Tr_a^\beta$ or $(y,x) \notin Tr_a^\beta$. For example, assume that $U=\{x,y,z\}$, where $x,y,z \in R$ and satisfy $f_a(x,a)=0$ and $f_a(z,a)=10$. Let $\beta=0.5$, when $f_a(y,a)=2$, we have $(x,y) \in Tr_a^{0.5}$ and $(y,x) \in Tr_a^{0.5}$. However, when $f_a(y,a)=4$, we only have $(x,y) \in Tr_a^{0.5}$.
(3)When x, y and z are missing values, we can get that $(x,y) \in Tr_a^\beta, (y,z) \in Tr_a^\beta \Rightarrow (x,z) \in Tr_a^\beta$. However, when x, y and z are not missing values, we can get that $d_a(x,z) \le d_a(x,y)+d_a(y,z) \le \beta \max\limits_{m\in U}\left\{d_a(x,m)+d_a(y,m)\right\}$ and we all know that $\beta \max\limits_{m\in U}\left\{d_a(x,m)+d_a(y,m)\right\} \ge \beta \max\limits_{m\in U} d_a(x,m)$. Thus, we can get that $d_a(x,z) \le \beta \max\limits_{m\in U} d_a(x,m)$ or $d_a(x,z) > \beta \max\limits_{m\in U} d_a(x,m)$, which means that $(x,z) \in Tr_a^\beta$ or $(x,z) \notin Tr_a^\beta$. For example, assume that $U=\{x,y,z,p\}$, where $x,y,z,p \in R$ and satisfy $f_a(x,a)=0$, $f_a(y,a)=5$ and $f_a(p,a)=20$. Let $\beta=0.5$, when $f_a(z,a)=8$, we have $(x,y) \in Tr_a^{0.5}$, $(y,z) \in Tr_a^{0.5}$ and $(x,z) \in Tr_a^{0.5}$. However, when $f_a(z,a)=12$, we have $(x,y) \in Tr_a^{0.5}$, $(y,z) \in Tr_a^{0.5}$ and $(x,z) \notin Tr_a^{0.5}$.
(4) $\forall (x,y) \in Tr_a^{\beta_1}$, we have $\frac{d_a(x,y)}{\max_{m\in U} d_a(x,m)} \le \beta_1$. And when $\beta_1 \le \beta_2$, we can get that $\frac{d_a(x,y)}{\max_{m\in U} d_a(x,m)} \le \beta_1 \le \beta_2$, so $(x,y) \in Tr_a^{\beta_2}$. Thus, we can get $Tr_a^{\beta_1} \subseteq Tr_a^{\beta_2}$. □

Definition 3.3. Let $Ms-IIvIS = \left\{ I_q \,\middle|\, I_q = (U, A_q, V_q, f_{A_q}), q = 1,...,N \right\}$ be a Ms-IIvIS. For any $a \in A$ and $\beta \in [0,1]$, the IE of the q-th information source I_q is defined as follows:

$$IE_a^\beta(I_q) = -\sum_{i=1}^{|U|} \frac{1}{|U|} \ln\left(\frac{\left|Tr_a^\beta(x_i)^q\right|}{|U|}\right), \tag{6}$$

where $Tr_a^\beta(x_i)^q$ denotes the tolerance class of x_i with respect to a under I_q.

Proposition 3.2. *Let* $Ms-IIvIS = \left\{ I_q \,\middle|\, I_q = (U, A_q, V_q, f_{A_q}), q = 1,...,N \right\}$ *be a Ms-IIvIS. For any* $a,b \in A$ *and* $\beta,\eta \in [0,1]$, *the information entropy of* I_q *and* I_m *are subject to the following properties:*
(1) $0 \le IE_a^\beta(I_q) \le \ln|U|$
(2) $IE_b^\eta(I_m) \le IE_a^\beta(I_q)$, *if* $Tr_a^\beta(x_i)^q \subseteq Tr_b^\eta(x_i)^m$, $\forall x_i \in U$

Proof. (1) Given the function $f(x) = -\ln x$, we can get $\frac{\partial f}{\partial x} = -\frac{1}{x}$. When $x>0$, we have $\frac{\partial f}{\partial x} < 0$. Thus, when $Tr_a^\beta(x_i)^q = \{x_i\}$, $\forall x_i \in U$, the maximum value of $IE_a^\beta(I_q)$ is $\ln|U|$. When $Tr_a^\beta(x_i)^q = U$, $\forall x_i \in U$, the minimum value of $IE_a^\beta(I_q)$ is 0.
(2) If $Tr_a^\beta(x_i)^q \subseteq Tr_b^\eta(x_i)^m$, then we can get $\left|Tr_a^\beta(x_i)^q\right| \le \left|Tr_b^\eta(x_i)^m\right|$. So according to the monotonicity of $f(x) = -\ln x$, we have $-\ln\left(\frac{\left|Tr_a^\beta(x_i)^q\right|}{|U|}\right) \ge -\ln\left(\frac{\left|Tr_b^\eta(x_i)^m\right|}{|U|}\right)$, $\forall x_i \in U$. Thus, we can get $-\sum_{i=1}^{|U|} \frac{1}{|U|} \ln\left(\frac{\left|Tr_a^\beta(x_i)^q\right|}{|U|}\right) \ge -\sum_{i=1}^{|U|} \frac{1}{|U|} \ln\left(\frac{\left|Tr_b^\eta(x_i)^m\right|}{|U|}\right)$. □

According to the monotonicity of $IE_a^\beta(I_i)$, we can know that the smaller the $IE_a^\beta(I_i)$ is, the bigger the amount of information of I_i for a is. Thus, in order to minimize the uncertainty of systems, we choose the information source with the lowest information entropy as the fusion result of the attribute, which is shown in Definition 3.4.

Definition 3.4. Let $Ms-IIvIS = \left\{ I_q \,\middle|\, I_q = (U, A_q, V_q, f_{A_q}), q = 1,...,N \right\}$ be a Ms-IIvIS. For any $a \in A$, the most essential source I_i can be selected by

$$i = \underset{q\in\{1,...,N\}}{\arg\min}\ IE_a^\beta(I_q). \tag{7}$$

According to Definition 3.4, the process of obtaining a final fused table involves selecting the information source with the minimum entropy per attribute and integrating them into a new information table. However, it is essential to acknowledge that the original information sources may contain missing values, which implies that the resulting fused table will also contain missing values. In the context of tolerance relation, missing objects can be treated as any object under the same attribute. Therefore, to account for all potential values of missing objects, the unions of objects in the TC of missing value objects can be regarded as virtual measurements or representations of these missing objects. By considering the unions of objects in the TC of missing value objects, the fused table encompasses a broader range of possible values for the missing objects, thereby providing a more comprehensive representation of the overall data. The complete rule can be defined as

$$f(x,a)=\bigcup_{y\in\Delta_x^a} f(y,a)=[\min_{y\in\Delta_x^a} f^L(y,a), \max_{y\in\Delta_x^a} f^U(y,a)], \tag{8}$$

where $\Delta_x^a = Tr_a^\beta(x)^{\underset{q=\{1,\dots,N\}}{\arg\min}\ IE_a^\beta(I_q)}$ denotes the TC set of object x under the selected information source for attribute a.

An intuitive fusion framework is presented in Fig. 1. The fusion process comprises three steps: firstly, calculating the entropy of each information source under each attribute; secondly, selecting the representative information source with the lowest entropy and combining these selected sources to form the fusion information table. Lastly, filling in missing values by TC to obtain a complete information table. The corresponding static fusion algorithm, which can be referred to as Algorithm 1, is established to implement the fusion process. The time complexity of step 6 is determined to be $O(n)$. Therefore, the overall time complexity of steps 2 to 10 is calculated as $O\left(N\times|A|\times n^2\right)$. Moreover, the time complexity of steps 11 to 16 is determined to be $O(|A|)$. Consequently, the total time complexity of the static algorithm is $O\left(N\times|A|\times n^2+|A|\right)$.

Algorithm 1: The static fusion algorithm in Ms-IIvIS.

Input : A Ms-IIvIS $Ms-IIvIS=\left\{I_q \,\middle|\, I_q=(U,A_q,V_q,f_{A_q}), q=1,\dots,N\right\}$, where $U=\{x_1,x_2,\dots,x_n\}$; parameter β
Output : A new fused information table

1 **begin**
2   **for** $q=1:N$ **do**
3   **for** *each* $a\in A$ **do**
4       $IE_a^\beta(I_q)\leftarrow 0$
5   **for** $i=1:n$ **do**
6   compute $Tr_a^\beta(x_i)^q$
7         $IE_a^\beta(I_q)\leftarrow IE_a^\beta(I_q)-\frac{1}{|U|}\ln\left(\frac{\left|Tr_a^\beta(x_i)^q\right|}{|U|}\right)$
8   **end**
9   **end**
10   **end**
11   **for** *each* $a\in A$ **do**
12   compute $i=\underset{q\in\{1,\dots,N\}}{\arg\min}\ IE_a^\beta(I_q)$
13   **if** *f(x,a) = ** **then**
14       $f(x,a)=\bigcup_{y\in\Delta_x^a} f(y,a)$, where $\Delta_x^a=Tr_a^\beta(x)^{\underset{q=\{1,\dots,N\}}{\arg\min}\ IE_a^\beta(I_q)}$
15   **end**
16   **end**
  return : $\left\{V_{\underset{q=\{1,\dots,N\}}{\arg\min}\ IE_{a_1}^\beta(I_q)},\dots,V_{\underset{q=\{1,\dots,N\}}{\arg\min}\ IE_{a_{|A|}}^\beta(I_q)}\right\}$
17 **end**

Example 3.1. (Continue with Example 2.1.) Based on Definition 3.3, the IE of any subsystem under any attribute can be computed. Take a_1 of the second subsystem as an example. Firstly, the distance between any two objects can be obtained according to Definition 3.1, as follows.

$$Dis=\begin{pmatrix}
0 & 0 & 0 & 6.79061 & 4.30443 & 75.0264 & 57.2422 & 0 & 14.0927 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
6.79061 & 0 & 0 & 0 & 2.55501 & 81.0087 & 63.0343 & 0 & 17.3656 & 0\\
4.30443 & 0 & 0 & 2.55501 & 0 & 79.001 & 61.1103 & 0 & 16.2959 & 0\\
75.0264 & 0 & 0 & 81.0087 & 79.001 & 0 & 18.2946 & 0 & 65.4832 & 0\\
57.2422 & 0 & 0 & 63.0343 & 61.1103 & 18.2946 & 0 & 0 & 47.2314 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
14.0927 & 0 & 0 & 17.3656 & 16.2959 & 65.4832 & 47.2314 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}$$

Secondly, according to Definition 3.2, the TC of objects can be calculated as $Tr_{a_1}^{0.5}(x_1)^2=Tr_{a_1}^{0.5}(x_9)^2=Tr_{a_1}^{0.5}(x_4)^2=Tr_{a_1}^{0.5}(x_5)^2=\{x_1,x_2,x_3,x_4,x_5,x_8,x_9,x_{10}\}$; $Tr_{a_1}^{0.5}(x_2)^2=Tr_{a_1}^{0.5}(x_3)^2=Tr_{a_1}^{0.5}(x_8)^2=Tr_{a_1}^{0.5}(x_{10})^2=U$, $Tr_{a_1}^{0.5}(x_6)^2=Tr_{a_1}^{0.5}(x_7)^2=\{x_2,x_3,x_6,x_7,x_8,x_{10}\}$. Then, according to Definition 3.3, the information entropy of the second subsystem with respect to a_1 can be computed as $IE_{a_1}^{0.5}(I_2)=-\frac{1}{10}\left(4\times\ln\frac{8}{10}+4\times 0+2\times\ln\frac{6}{10}\right)=0.191423$. Similarly, the IE of other subsystems with respect to other attributes can be calculated, shown in Table 6. Lastly, based on Definition 3.4, the fusion table can be obtained, shown in Table 7.

Table 6
The information entropy of sources with respect to attributes.

A	I_1	I_2	I_3	I_4
a_1	0.70178	0.191423	0.311333	0.263554
a_2	0.604452	0.680666	0.537509	0.397305
a_3	0.445125	0.778791	0.575684	0.532672
a_4	0.35882	0.39888	0.445125	0.486426
a_5	0.191423	0.382709	0.573622	0.386282
a_6	0.40866	0.715462	0.550904	0.554986

Table 7
The final fusion results.

U	a_1	a_2	a_3	a_4	a_5	a_6
x_1	[129.07, 139.99]	[3, 7]	[115.87, 124.98]	[60.78, 179.97]	[65.28, 282.95]	[70.18, 93.23]
x_2	[119.14, 214.99]	[4, 11]	[116.96, 124.99]	[79.48, 118.98]	[65.28, 282.95]	[27.99, 100.32]
x_3	[119.14, 214.99]	[2, 22]	[112.38, 121.98]	[119.07, 179.97]	[66.28, 84.98]	[78.18, 100.32]
x_4	[125.89, 133.99]	[2, 9]	[111.48, 120.98]	[60.78, 98.98]	[65.28, 282.95]	[66.38, 89.17]
x_5	[127.48, 135.99]	[4.4, 11]	[111.98, 295.95]	[81.58, 119.98]	[139.37, 282.95]	[43.89, 74.99]
x_6	[127.08, 214.99]	[6.3, 16]	[111.48, 295.95]	[60.78, 179.97]	[68.28, 87.98]	[27.99, 61.81]
x_7	[123.81, 196.99]	[4.5, 9]	[177.26, 268.95]	[84.28, 153.97]	[77.98, 97.98]	[30.09, 60.8]
x_8	[119.14, 214.99]	[10.1, 20]	[111.48, 295.95]	[102.28, 162.97]	[65.28, 84.98]	[27.99, 100.32]
x_9	[119.14, 149.99]	[12, 22]	[224.96, 268.95]	[90.28, 150.97]	[65.28, 282.95]	[77.98, 89.17]
x_{10}	[119.14, 214.99]	[2, 22]	[177.17, 267.95]	[67.98, 118.98]	[109.38, 252.96]	[39.89, 70.93]

After obtaining the final result, we can calculate the AP based on Definition 2.3. Firstly, the TC of each object with respect to each single attribute can be computed based on Definition 3.2. Then, the TC of each object with respect to the entire attribute set can be calculated. Take the object x_1 as an example. The TC of x_1 with respect to each attribute are computed as $Tr^{0.5}_{a_1}(x_1) = \{x_1, x_4, x_5, x_9\}$, $Tr^{0.5}_{a_2}(x_1) = \{x_1, x_2, x_4, x_5, x_7\}$, $Tr^{0.5}_{a_3}(x_1) = \{x_1, x_2, x_3, x_4\}$, $Tr^{0.5}_{a_4}(x_1) = \{x_1, x_6, x_7\}$, $Tr^{0.5}_{a_5}(x_1) = \{x_1, x_2, x_4, x_5, x_9, x_{10}\}$, and $Tr^{0.5}_{a_6}(x_1) = \{x_1, x_3, x_4, x_9\}$. Then, the TC of the entire attribute A can be computed as $Tr^{0.5}_{A}(x_1) = \bigcap_i Tr^{0.5}_{a_i}(x_1) = \{x_1\}$. Similarly, the TC of other five objects with respect to A can be computed as $Tr^{0.5}_{A}(x_2) = \{x_2\}$, $Tr^{0.5}_{A}(x_3) = \{x_3\}$, ..., $Tr^{0.5}_{A}(x_{10}) = \{x_{10}\}$. Later, for Y_1 of the decision partition U/d, the lower approximation $\underline{Tr_A}(Y_1)$ can be computed by $\underline{Tr_A}(Y_1) = \left\{ x \in U \,\middle|\, Tr^{0.5}_{A}(x) \subseteq Y_1 \right\} = \{x_1, x_2, x_6, x_8, x_9\}$. And the upper approximation can be calculated as $\overline{Tr_B}(Y_1) = \left\{ x \in U \,\middle|\, Tr^{0.5}_{A}(x) \cap Y_1 \neq \emptyset \right\} = \{x_1, x_2, x_6, x_8, x_9\}$. Likewise, for Y_2 of the decision partition U/d, the lower and upper approximations can be calculated as $\overline{Tr_A}(Y_1) = \underline{Tr_A}(Y_2) = \{x_3, x_4, x_5, x_7, x_{10}\}$. Thus, based on Definition 2.3, the AP can be calculated as $AP_{Tr_B}(U/d) = \frac{|\underline{Tr_B}(Y_1)| + |\underline{Tr_B}(Y_2)|}{|\overline{Tr_B}(Y_1)| + |\overline{Tr_B}(Y_2)|} = \frac{5+5}{5+5} = 1$.

4. Dynamic fusion approaches with the variation of objects

In real applications, new data is continuously collected, leading to the expansion of object sets in Ms-IIvIS. At the same time, it becomes necessary to remove obsolete objects from the object sets to ensure the accuracy of fusion results. Consequently, efficient dynamic fusion techniques are required to quickly obtain fusion results in the presence of object insertions or deletions. This section establishes two dynamic fusion approaches to handle object insertions and deletions, respectively. The two studied variation situations are illustrated in Fig. 2 and 3, respectively, providing visual representations of the dynamic changes in the object sets. The relevant algorithms for handling object insertions and deletions are provided, and the time complexities of these algorithms are analyzed to assess their computational efficiency.

4.1. *Dynamic fusion approach when inserting objects*

Given $Ms-IIvIS^t = \left\{ I_q \,\middle|\, I_q = (U, A_q, V_q, f_{A_q}), q = 1, ..., N \right\}$ be a Ms-IIvIS at time t, where $U = \{x_1, x_2, ..., x_n\}$. Assume that $\{x_{n+1}, x_{n+2}, ..., x_{n+\Delta n}\}$ are inserted into systems at time $t+1$. The following propositions are true.

Proposition 4.1. *For $\{x_1, x_2, ..., x_n\}$, the following properties are true:*
(1) If $\max\limits_{k \in \{n+1,...,n+\Delta n\}} dis_a(x_i, x_k) \geq \max\limits_{k \in \{1,...,n\}} dis_a(x_i, x_k)$, then

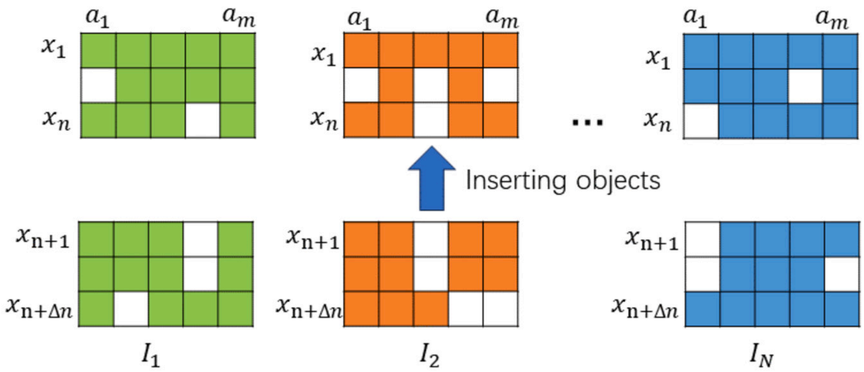

Fig. 2. The situation of inserting objects.

$$Tr_a^{\beta}(x_i)_{t+1}^q = Tr_a^{\beta}(x_i)_t^q \cup \left\{ y \in \left(Tr_a^{\beta}(x_i)_t^q\right)^C \,\middle|\, \frac{dis_a(x_i,y)}{\max\limits_{k\in\{n+1,\dots,n+\Delta n\}} dis_a(x_i,x_k)} \le \beta \right\}. \tag{9}$$

(2) If $\max\limits_{k\in\{n+1,\dots,n+\Delta n\}} dis_a(x_i,x_k) < \max\limits_{k\in\{1,\dots,n\}} dis_a(x_i,x_k)$*, then*

$$Tr_a^{\beta}(x_i)_{t+1}^q = Tr_a^{\beta}(x_i)_t^q \cup \left\{ y \in \left\{x_{n+1},\dots,x_{n+\Delta n}\right\} \,\middle|\, \frac{dis_a(x_i,y)}{\max\limits_{k\in\{1,\dots,n\}} dis_a(x_i,x_k)} \le \beta \right\}. \tag{10}$$

Proof. (1) For $\{x_1,x_2,\dots,x_n\}$, we know that if $\max\limits_{k\in\{n+1,\dots,n+\Delta n\}} dis_a(x_i,x_k) \ge \max\limits_{k\in\{1,\dots,n\}} dis_a(x_i,x_k)$, then $\forall z \in Tr_a^{\beta}(x_i)_t^q$, $\frac{dis_a(x_i,z)}{\max\limits_{k\in\{n+1,\dots,n+\Delta n\}} dis_a(x_i,x_k)} \le \frac{dis_a(x_i,z)}{\max\limits_{k\in\{1,\dots,n\}} dis_a(x_i,x_k)} \le \beta$. So we can get that $z \in Tr_a^{\beta}(x_i)_{t+1}^q$. Furthermore, for any $y \in \left(Tr_a^{\beta}(x_i)_t^q\right)^C$, we can deduce that $\frac{dis_a(x_i,y)}{\max\limits_{k\in\{1,\dots,n+\Delta n\}} dis_a(x_i,x_k)} = \frac{dis_a(x_i,y)}{\max\limits_{k\in\{n+1,\dots,n+\Delta n\}} dis_a(x_i,x_k)}$. So if $\frac{dis_a(x_i,y)}{\max\limits_{k\in\{n+1,\dots,n+\Delta n\}} dis_a(x_i,x_k)} \le \beta$, then $y \in Tr_a^{\beta}(x_i)_{t+1}^q$. Thus, we have $Tr_a^{\beta}(x_i)_{t+1}^q = Tr_a^{\beta}(x_i)_t^q \cup \left\{ y \in \left(Tr_a^{\beta}(x_i)_t^q\right)^C \,\middle|\, \frac{dis_a(x_i,y)}{\max\limits_{k\in\{n+1,\dots,n+\Delta n\}} dis_a(x_i,x_k)} \le \beta \right\}$.

(2) If $\max\limits_{k\in\{n+1,\dots,n+\Delta n\}} dis_a(x_i,x_k) < \max\limits_{k\in\{1,\dots,n\}} dis_a(x_i,x_k)$, then we can get that $\forall z \in Tr_a^{\beta}(x_i)_t^q$, $\frac{dis_a(x_i,z)}{\max\limits_{k\in\{1,\dots,n+\Delta n\}} dis_a(x_i,x_k)} = \frac{dis_a(x_i,z)}{\max\limits_{k\in\{1,\dots,n\}} dis_a(x_i,x_k)} \le \beta$. So we can deduce $z \in Tr_a^{\beta}(x_i)_{t+1}^q$. Furthermore, for inserting objects $\{x_{n+1},x_{n+2},\dots,x_{n+\Delta n}\}$, if $\frac{dis_a(x_i,y)}{\max\limits_{k\in\{1,\dots,n\}} dis_a(x_i,x_k)} \le \beta$, then $y \in Tr_a^{\beta}(x_i)_{t+1}^q$. Thus, we have $Tr_a^{\beta}(x_i)_{t+1}^q = Tr_a^{\beta}(x_i)_t^q \cup \left\{ y \in \left\{x_{n+1},\dots,x_{n+\Delta n}\right\} \,\middle|\, \frac{dis_a(x_i,y)}{\max\limits_{k\in\{1,\dots,n\}} dis_a(x_i,x_k)} \le \beta \right\}$. □

Proposition 4.2. *For* $\{x_{n+1},x_{n+2},\dots,x_{n+\Delta n}\}$*, we have:*

$$Tr_a^{\beta}(x_i)_{t+1}^q = \left\{ y \in \left\{x_1,\dots,x_{n+\Delta n}\right\} \,\middle|\, \frac{dis_a(x_i,y)}{\max\limits_{k\in\{1,\dots,n+\Delta n\}} dis_a(x_i,x_k)} \le \beta \right\} \tag{11}$$

Proof. The property is obviously according to Definition 3.4. □

Based on Propositions 4.1 and 4.2, we can quickly calculate the updated TC $Tr_a^{\beta}(x_i)_{t+1}^q$. And then, according to Definitions 3.3 and 3.4, the updated fusion results can be obtained fast. The incremental fusion algorithm for inserting objects is given in Algorithm 2. The time complexity of step 4 to 15 is $O(n+\Delta n\times(n+\Delta n))$. So the time complexity of step 2 to 17 is $O(N\times|A|\times(n+\Delta n\times(n+\Delta n)))$. Thus, the total time complexity of the Algorithm 2 is $O(N\times|A|\times(n+\Delta n\times(n+\Delta n))+|A|)$.

Example 4.1. (Continue with Example 3.1.) In real human life, a large number of different patients will go to hospitals for medical examinations day by day. Assume that in time t, only five patients visited the four hospitals for examinations, which means $U^t = \{x_1,x_2,x_3,x_4,x_5\}$. The Ms-IIvIS in time t is shown in Tables 8–11. We also take a_1 in the second hospital as an example. The TC of objects with respect to a_1 can be computed as follows. $Tr_{a_1}^{0.5}(x_1)_t^2 = \{x_1,x_2,x_3\}$; $Tr_{a_1}^{0.5}(x_2)_t^2 = Tr_{a_1}^{0.5}(x_3)_t^2 = \{x_1,x_2,x_3,x_4,x_5\}$;

Algorithm 2: The dynamic fusion algorithm when inserting objects.

Input : The original Ms-IIvIS $Ms-IIvIS^t = \left\{ I_q \middle| I_q = (U, A_q, V_q, f_{A_q}), q = 1, ..., N \right\}$; the inserting object set $\{x_{n+1}, ..., x_{n+\Delta n}\}$; parameters β
Output : An updated fusion table.
begin
  for $q = 1 : N$ **do**
    for *each* $a \in A$ **do**
      for $i = 1 : n$ **do**
        if $\max_{k\in\{n+1,...,n+\Delta n\}} dis_a(x_i, x_k) \geq \max_{k\in\{1,...,n\}} dis_a(x_i, x_k)$ **then**
          $Tr_a^\beta(x_i)_{t+1}^q = Tr_a^\beta(x_i)_t^q \cup \left\{ y \in \left(Tr_a^\beta(x_i)_t^q\right)^C \middle| \frac{dis_a(x_i,y)}{\max_{k\in\{n+1,...,n+\Delta n\}} dis_a(x_i,x_k)} \leq \beta \right\}$
        end
        if $\max_{k\in\{n+1,...,n+\Delta n\}} dis_a(x_i, x_k) < \max_{k\in\{1,...,n\}} dis_a(x_i, x_k)$ **then**
          $Tr_a^\beta(x_i)_{t+1}^q = Tr_a^\beta(x_i)_t^q \cup \left\{ y \in \{x_{n+1}, ..., x_{n+\Delta n}\} \middle| \frac{dis_a(x_i,y)}{\max_{k\in\{1,...,n\}} dis_a(x_i,x_k)} \leq \beta \right\}$
        end
      end
      for $i = n+1 : n+\Delta n$ **do**
        compute $Tr_a^\beta(x_i)_{t+1}^q = \left\{ y \in \{x_1, ..., x_{n+\Delta n}\} \middle| \frac{dis_a(x_i,y)}{\max_{k\in\{1,...,n+\Delta n\}} dis_a(x_i,x_k)} \leq \beta \right\}$
      end
      compute $IE_a^\beta(I_q)$
    end
  end
  for *each* $a \in A$ **do**
    compute $i = \arg\min_{q\in\{1,...,N\}} IE_a^\beta(I_q)$
    if *f(x,a)* = * **then**
      $f(x,a) = \bigcup_{y\in\Delta_x^a} f(y,a)$, where $\Delta_x^a = Tr_a^\beta(x)^{\arg\min_{q\in\{1,...,N\}} IE_a^\beta(I_q)}$
    end
  end
  return : $\left\{ V^{t+1}_{\arg\min_{q\in\{1,...,N\}} IE_{a_1}^\beta(I_q)}, ..., V^{t+1}_{\arg\min_{q\in\{1,...,N\}} IE_{a_{|A|}}^\beta(I_q)} \right\}$
end

Table 8
The medical examination results of the first hospital in time t.

U	a_1	a_2	a_3	a_4	a_5	a_6
x_1	[128.26,139.97]	[1.00,5.00]	[115.87,124.98]	*	*	[70.18,93.23]
x_2	[118.40,130.97]	[3.99,11.00]	[116.96,124.99]	[79.48,118.98]	*	*
x_3	[108.53,119.97]	[2.79,10.00]	[112.38,121.98]	[119.07,179.97]	[66.28,84.98]	[78.18,100.32]
x_4	[125.11,133.97]	[1.99,9.00]	[111.48,120.98]	[60.78,98.98]	*	[66.38,89.17]
x_5	[126.69,135.97]	[4.39,11.00]	[111.98,295.95]	[81.58,119.98]	[139.37,282.95]	[43.89,74.99]

$Tr_{a_1}^{0.5}(x_4)_t^2 = \{x_2, x_3, x_4, x_5\}$; $Tr_{a_1}^{0.5}(x_5)_t^2 = \{x_2, x_3, x_5\}$. Then, based on Definition 3.3, the IE of the second hospital with respect to a_1 can be obtained as $IE_{a_1}^{0.5}(I_2) = 0.248959$. Similarly, the IE of subsystems in time t can be calculated, shown in Table 12. Thus, the fusion results in time t can be obtained based on Definition 3.4, shown in Table 13.

However, in time $t+1$, five other patients visited the same four hospitals for the same medical examinations. The Ms-IIvIS in time $t+1$ can be seen in Tables 2–5. So according to Propositions 4.1 and 4.2, the TC of all objects with respect to a_1 under the second hospital can be fast computed as $Tr_{a_1}^{0.5}(x_1)_{t+1}^2 = Tr_{a_1}^{0.5}(x_1)_t^2 \cup \{x_4, x_5, x_8, x_9, x_{10}\}$; $Tr_{a_1}^{0.5}(x_2)_{t+1}^2 = Tr_{a_1}^{0.5}(x_2)_t^2 \cup \{x_6, x_7, x_8, x_9, x_{10}\}$; $Tr_{a_1}^{0.5}(x_3)_{t+1}^2 = Tr_{a_1}^{0.5}(x_3)_t^2 \cup \{x_6, x_7, x_8, x_9, x_{10}\}$; $Tr_{a_1}^{0.5}(x_4)_{t+1}^2 = Tr_{a_1}^{0.5}(x_4)_t^2 \cup \{x_1, x_8, x_9, x_{10}\}$; $Tr_{a_1}^{0.5}(x_5)_{t+1}^2 = Tr_{a_1}^{0.5}(x_5)_t^2 \cup \{x_1, x_4, x_8, x_9, x_{10}\}$; $Tr_{a_1}^{0.5}(x_6)_{t+1}^2 = Tr_{a_1}^{0.5}(x_7)_{t+1}^2 = \{x_2, x_3, x_6, x_7, x_8, x_{10}\}$; $Tr_{a_1}^{0.5}(x_8)_{t+1}^2 = Tr_{a_1}^{0.5}(x_{10})_{t+1}^2 = U$; $Tr_{a_1}^{0.5}(x_9)_{t+1}^2 = \{x_1, x_2, x_3, x_4, x_5, x_8, x_9, x_{10}\}$. Then, according to Definition 3.3, the IE in time $t+1$ can be quickly updated, shown in Table 6. Finally, based on Definition 3.4, the updated fusion results can be obtained, presented in Table 7.

4.2. Dynamic fusion approach when deleting objects

Given $Ms-IIvIS^t = \left\{ I_q \middle| I_q = (U, A_q, V_q, f_{A_q}), q = 1, ..., N \right\}$ be a Ms-IIvIS at time t, where $U = \{x_1, x_2, ..., x_n\}$. Assume that $\{x_{n-\Delta n+1}, x_{n-\Delta n+2}, ..., x_n\}$ are deleted from systems at time $t+1$. The following propositions are true.

Table 9
The medical examination results of the second hospital in time t.

U	a_1	a_2	a_3	a_4	a_5	a_6
x_1	[129.07,139.99]	[3.00,7.00]	[115.89,124.99]	[68.50,107.99]	[43.40,186.99]	[70.20,92.66]
x_2	*	[4.00,11.00]	[116.99,124.99]	[79.50,118.99]	[80.40,223.99]	[67.20,88.63]
x_3	*	[2.80,10.00]	[112.39,121.99]	[119.09,179.99]	*	[78.20,99.71]
x_4	[125.89,133.99]	[2.00,9.00]	[111.49,120.99]	[60.80,98.99]	[99.99,259.99]	[66.40,88.63]
x_5	[127.48,135.99]	[4.40,11.00]	[111.99,295.99]	*	*	[43.90,74.53]

Table 10
The medical examination results of the third hospital in time t.

U	a_1	a_2	a_3	a_4	a_5	a_6
x_1	*	[3.00,7.00]	[115.87,124.97]	*	[43.39,186.96]	[70.18,93.42]
x_2	[118.15,130.97]	*	*	[79.48,118.97]	[80.38,223.95]	[67.18,89.36]
x_3	*	[2.80,10.00]	[112.37,121.97]	*	[66.28,84.98]	[78.18,100.53]
x_4	*	[2.00,9.00]	[111.47,120.97]	[60.79,98.98]	*	[66.38,89.36]
x_5	[126.42,135.97]	[4.40,11.00]	[111.97,295.93]	[81.58,119.97]	[139.37,282.93]	[43.89,74.14]

Table 11
The medical examination results of the fourth hospital in time t.

U	a_1	a_2	a_3	a_4	a_5	a_6
x_1	*	[3.00,7.00]	[115.90,125.00]	*	*	[70.20,92.01]
x_2	[119.99,131.00]	[4.00,11.00]	[117.00,125.00]	[79.50,119.00]	*	[67.20,88.01]
x_3	[109.99,112.00]	*	[112.40,121.98]	[119.10,180.00]	[66.30,85.00]	[78.20,99.01]
x_4	[126.79,134.00]	[2.00,9.00]	[111.48,120.98]	[60.80,99.00]	[100.00,260.00]	[66.40,88.01]
x_5	[128.39,136.00]	[4.40,11.00]	*	[81.60,120.00]	[139.40,283.00]	[43.90,74.01]

Table 12
The information entropy of sources in time t.

A	I_1	I_2	I_3	I_4
a_1	0.639032	0.248959	0.089257	0.317144
a_2	0.639032	0.892734	0.489754	0.570847
a_3	0.500402	0.500402	0.317144	0.40866
a_4	0.317144	0.317144	0.191423	0.317144
a_5	0.089257	0.248959	0.40866	0.191423
a_6	0.374681	0.615475	0.615475	0.615475

Table 13
The fusion results in time t.

U	a_1	a_2	a_3	a_4	a_5	a_6
x_1	[118.15, 135.97]	[3, 7]	[115.87, 124.97]	[60.79, 119.97]	[66.28, 282.95]	[70.18, 93.23]
x_2	[118.15, 130.97]	[2, 11]	[111.47, 295.93]	[79.48, 118.97]	[66.28, 282.95]	[43.89, 100.32]
x_3	[118.15, 135.97]	[2.8, 10]	[112.37, 121.97]	[60.79, 119.97]	[66.28, 84.98]	[78.18, 100.32]
x_4	[118.15, 135.97]	[2, 9]	[111.47, 120.97]	[60.79, 98.98]	[66.28, 282.95]	[66.38, 89.17]
x_5	[126.42, 135.97]	[4.4, 11]	[111.97, 295.93]	[81.58, 119.97]	[139.37, 282.95]	[43.89, 74.99]

Proposition 4.3. *For $\{x_1, x_2, ..., x_{n-\Delta n}\}$, we have:*
(1)If $\max\limits_{k\in\{n-\Delta n+1,...,n\}} dis_a(x_i,x_k) \geq \max\limits_{k\in\{1,...,n-\Delta n\}} dis_a(x_i,x_k)$, then

$$Tr_a^{\beta}(x_i)_{t+1}^q = Tr_a^{\beta}(x_i)_t^q \cap \{x_1, x_2, ..., x_{n-\Delta n}\} - \left\{ y \in Tr_a^{\beta}(x_i)_t^q \cap \{x_1, x_2, ..., x_{n-\Delta n}\} \,\middle|\, \frac{dis_a(x_i,y)}{\max\limits_{k\in\{1,...,n-\Delta n\}} dis_a(x_i,x_k)} > \beta \right\}. \tag{12}$$

(2)If $\max\limits_{k\in\{n-\Delta n+1,...,n\}} dis_a(x_i,x_k) < \max\limits_{k\in\{1,...,n-\Delta n\}} dis_a(x_i,x_k)$, then

$$Tr_a^{\beta}(x_i)_{t+1}^q = Tr_a^{\beta}(x_i)_t^q \cap \{x_1, x_2, ..., x_{n-\Delta n}\}. \tag{13}$$

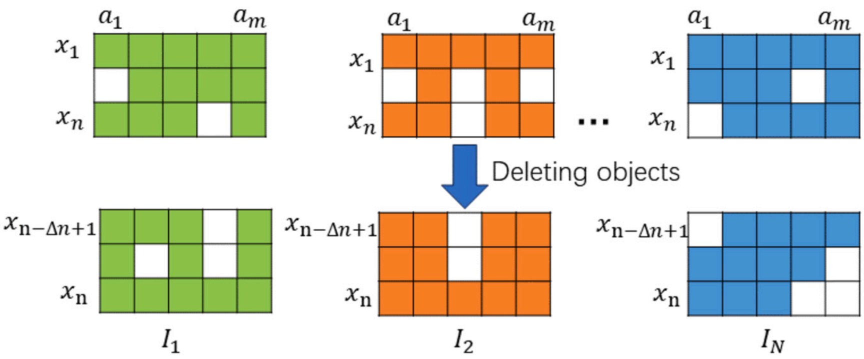

Fig. 3. The situation of deleting objects.

Proof. (1) For $\{x_1, x_2, ..., x_{n-\Delta n}\}$, if $\max\limits_{k\in\{n-\Delta n+1,...,n\}} dis_a(x_i,x_k) \geq \max\limits_{k\in\{1,...,n-\Delta n\}} dis_a(x_i,x_k)$, we can get that for any $z \in \left(Tr_a^{\beta}(x_i)_t^q\right)^C \cap \{x_1, x_2, ..., x_{n-\Delta n}\}$, $\frac{dis_a(x_i,z)}{\max\limits_{k\in\{1,...,n-\Delta n\}} dis_a(x_i,x_k)} > \frac{dis_a(x_i,z)}{\max\limits_{k\in\{n-\Delta n+1,...,n\}} dis_a(x_i,x_k)} > \beta$. So we can deduce $z \notin Tr_a^{\beta}(x_i)_{t+1}^q$. Furthermore, we all know that for any $y \in Tr_a^{\beta}(x_i)_t^q \cap \{x_1, x_2, ..., x_{n-\Delta n}\}$, if $\frac{dis_a(x_i,y)}{\max\limits_{k\in\{1,...,n-\Delta n\}} dis_a(x_i,x_k)} \leq \beta$, then $y \in Tr_a^{\beta}(x_i)_{t+1}^q$. Thus, we can get $Tr_a^{\beta}(x_i)_{t+1}^q = \left\{ y \in Tr_a^{\beta}(x_i)_t^q \cap \{x_1, x_2, ..., x_{n-\Delta n}\} \,\middle|\, \frac{dis_a(x_i,y)}{\max\limits_{k\in\{1,...,n-\Delta n\}} dis_a(x_i,x_k)} \leq \beta \right\} = Tr_a^{\beta}(x_i)_t^q \cap \{x_1, x_2, ..., x_{n-\Delta n}\} - \left\{ y \in Tr_a^{\beta}(x_i)_t^q \cap \{x_1, x_2, ..., x_{n-\Delta n}\} \,\middle|\, \frac{dis_a(x_i,y)}{\max\limits_{k\in\{1,...,n-\Delta n\}} dis_a(x_i,x_k)} > \beta \right\}$.

(2) If $\max\limits_{k\in\{n-\Delta n+1,...,n\}} dis_a(x_i,x_k) < \max\limits_{k\in\{1,...,n-\Delta n\}} dis_a(x_i,x_k)$, then we can get for any $z \in \left(Tr_a^{\beta}(x_i)_t^q\right)^C \cap \{x_1, x_2, ..., x_{n-\Delta n}\}$, $\frac{dis_a(x_i,z)}{\max\limits_{k\in\{1,...,n-\Delta n\}} dis_a(x_i,x_k)} = \frac{dis_a(x_i,z)}{\max\limits_{k\in\{1,...,n\}} dis_a(x_i,x_k)} > \beta$. So we can deduce that $z \notin Tr_a^{\beta}(x_i)_{t+1}^q$. Furthermore, for any $z \in Tr_a^{\beta}(x_i)_t^q \cap \{x_1, x_2, ..., x_{n-\Delta n}\}$, we have $\frac{dis_a(x_i,z)}{\max\limits_{k\in\{1,...,n-\Delta n\}} dis_a(x_i,x_k)} = \frac{dis_a(x_i,z)}{\max\limits_{k\in\{1,...,n\}} dis_a(x_i,x_k)} \leq \beta$. So we can get $z \in Tr_a^{\beta}(x_i)_{t+1}^q$. Thus, $Tr_a^{\beta}(x_i)_{t+1}^q = Tr_a^{\beta}(x_i)_t^q \cap \{x_1, x_2, ..., x_{n-\Delta n}\}$. □

Based on Proposition 4.3, we can quickly calculate the updated TC $Tr_a^{\beta}(x_i)_{t+1}^q$. And then, according to Definitions 3.3 and 3.4, the updated fusion results can be obtained fast. The incremental fusion algorithm for deleting objects is given in Algorithm 3. The time complexity of step 2 to 14 is $O(N \times |A| \times (n - \Delta n))$. Moreover, the time complexity of step 15 to 20 is $O(|A|)$. Thus, the total time complexity of the Algorithm 3 is $O(N \times |A| \times (n - \Delta n) + |A|)$.

Example 4.2. (Continue with Example 3.1.) In real applications, samples from some patients may need to be removed because of contamination. Assume that in time t, ten patients visited the four hospitals for examinations, which are shown in Tables 2–5. The fusion results are shown in Table 7.

However, in time $t+1$, samples from $\{x_6, x_7, x_8, x_9, x_{10}\}$ were found to be contaminated. So the last five objects need to be deleted from systems. The Ms-IIvIS in time $t+1$ can be seen in Tables 8–11. Take a_1 of the second hospital as an example. According to Proposition 4.3, the tolerance classes can be fast computed as $Tr_{a_1}^{0.5}(x_1)_{t+1}^2 = Tr_{a_1}^{0.5}(x_1)_t^2 - \{x_4, x_5, x_8, x_9, x_{10}\}$; $Tr_{a_1}^{0.5}(x_2)_{t+1}^2 = Tr_{a_1}^{0.5}(x_2)_t^2 - \{x_6, x_7, x_8, x_9, x_{10}\}$; $Tr_{a_1}^{0.5}(x_3)_{t+1}^2 = Tr_{a_1}^{0.5}(x_3)_t^2 - \{x_6, x_7, x_8, x_9, x_{10}\}$; $Tr_{a_1}^{0.5}(x_4)_{t+1}^2 = Tr_{a_1}^{0.5}(x_4)_t^2 - \{x_1, x_8, x_9, x_{10}\}$. Thus, according to Definition 3.3, the entropies in time $t+1$ can be quickly updated, shown in Table 12. After that, the final fusion results in time $t+1$ can be obtained, shown in Table 13.

5. Experimental analysis

In this section, numerous experiments are conducted to evaluate the effectiveness of the static fusion approach and the efficiency of two dynamic fusion mechanisms. The datasets used in these experiments are obtained from the UC Irvine Machine Learning Repository,[2] and the details of these datasets are provided in Table 14. Table 15 details the running environment used for conducting the experiments. It is widely recognized that obtaining multi-source incomplete interval-valued datasets directly from public datasets is not feasible. To overcome this challenge, the simulation method described in references [10] and [28] is utilized to create Ms-IIvIS from public datasets. This simulation method involves several steps. First, the original single-valued data is transformed into

[2] http://archive.ics.uci.edu.

Algorithm 3: The dynamic fusion algorithm when deleting objects.

Input : The original Ms-IIvIS $Ms - IIvIS^t = \left\{ I_q \,\middle|\, I_q = (U, A_q, V_q, f_{A_q}), q = 1, ..., N \right\}$; the deleting object set $\{x_{n-\Delta n+1}, x_{n-\Delta n+2}, ..., x_n\}$; parameters β

Output : An updated fusion table.

begin
  for $q = 1 : N$ **do**
    for *each* $a \in A$ **do**
      for $i = 1 : n - \Delta n$ **do**
        if $\max_{k \in \{n-\Delta n+1, ..., n\}} dis_a(x_i, x_k) \geq \max_{k \in \{1, ..., n-\Delta n\}} dis_a(x_i, x_k)$ **then**
          $Tr_a^\beta(x_i)_{t+1}^q = Tr_a^\beta(x_i)_t^q \cap \{x_1, x_2, ..., x_{n-\Delta n}\} - \left\{ y \in Tr_a^\beta(x_i)_t^q \cap \{x_1, x_2, ..., x_{n-\Delta n}\} \,\middle|\, \frac{dis_a(x_i, y)}{\max_{k \in \{1, ..., n-\Delta n\}} dis_a(x_i, x_k)} > \beta \right\}$
        end
        if $\max_{k \in \{n-\Delta n+1, ..., n\}} dis_a(x_i, x_k) < \max_{k \in \{1, ..., n-\Delta n\}} dis_a(x_i, x_k)$ **then**
          $Tr_a^\beta(x_i)_{t+1}^q = Tr_a^\beta(x_i)_t^q \cap \{x_1, x_2, ..., x_{n-\Delta n}\}$
        end
      end
      compute $IE_a^\beta(I_q)$
    end
  end
  for *each* $a \in A$ **do**
    compute $i = \arg\min_{q \in \{1, ..., N\}} IE_a^\beta(I_q)$
    if *f(x,a)* = * **then**
      $f(x, a) = \cup_{y \in \Delta_x^a} f(y, a)$, where $\Delta_x^a = Tr_a^\beta(x)^{\arg\min_{q=\{1,...,N\}} IE_a^\beta(I_q)}$
    end
  end
  return : $\left\{ V^{t+1}_{\arg\min_{q=\{1,...,N\}} IE_{a_1}^\beta(I_q)}, ..., V^{t+1}_{\arg\min_{q=\{1,...,N\}} IE_{a_{|A|}}^\beta(I_q)} \right\}$
end

Table 14
The description of data sets.

N0.	Datasets	Abbreviation	Samples	Attributes	Classes
1	Wholesale customers	WC	440	8	2
2	seeds	seeds	210	7	3
3	Breast Cancer Coimbra	BCC	116	10	2
4	User Knowledge Modeling	UKM	403	5	4
5	Parkinsons	Parkinsons	197	23	2
6	Mammographic Mass	MM	961	6	2
7	Sports articles for objectivity analysis	SAOA	1000	59	2
8	Wine Quality White	WQW	4898	12	7
9	Mushroom	Mushroom	8124	22	7

Table 15
The running environments of experiments.

Name	Model	Parameter
CPU	Intel(R) Xeon(R) W-2123	3.60 GHz
Platform	Python	3.10
System	Windows10	64bit
Memory	DDR4	64 GB

interval-valued data by $[f^L(x,a), f^U(x,a)] = [f(x,a) - 2\sigma(x,a), f(x,a) + 2\sigma(x,a)]$, where $f(x,a)$ denotes the original value of sample x under attribute a, $\sigma(x,a)$ denotes the standard deviation of samples with the same class of x with respect to attribute a. Then, multiple information sources are generated by adding different white noise, that is, $f_i^L(x,a) = f^L(x,a) - |r_i| f^L(x,a)$ and $f_i^U(x,a) = f^U(x,a) + |r_i| f^U(x,a)$, where $r_i \overset{iid}{\sim} N(0, 0.1)$. Finally, missing values are created by randomly removing 10% of the data from each subsystem. Using the simulation method mentioned above, 20 subsystems are generated for each dataset listed in Table 14 for the experiments conducted in this section. This ensures a consistent and comparable basis for evaluating the fusion approaches and mechanisms.

Table 16
Comparison of classification accuracy of the fusion results(%).

Datasets	ESF	MAXF	MINF	MEANF	EUF	β
WC(k = 2)	63.4 ± 4.7**	64.1 ± 6.6*	63.6 ± 7.3*	64.3 ± 6.9	**66.1 ± 5.3**	0.5
seeds(k = 3)	89.5 ± 5.6*	91.4 ± 6.0*	91.9 ± 4.8**	91.0 ± 5.4**	**94.3 ± 5.6**	0.5
BCC(k = 2)	72.3 ± 9.4***	81.0 ± 10.8**	80.1 ± 10.0***	76.5 ± 10.7**	**89.5 ± 10.6**	0.2
UKM(k = 3)	24.3 ± 4.4**	26.3 ± 8.6	25.0 ± 4.8*	24.6 ± 7.1**	**28.3 ± 7.0**	0.1
Parkinsons(k = 14)	73.8 ± 7.8***	72.8 ± 9.5***	77.4 ± 12.2**	77.9 ± 11.3**	**80.5 ± 11.4**	0.5
MM(k = 8)	48.0 ± 4.4***	49.7 ± 6.2*	49.3 ± 5.3**	51.3 ± 4.4	**51.9 ± 5.2**	0.5
SAOA(k = 9)	82.6 ± 2.6***	80.5 ± 3.0***	85.6 ± 1.7***	82.6 ± 2.5***	**89.7 ± 2.3**	0.1
WQW(k = 20)	40.8 ± 1.5	39.7 ± 2.2**	40.1 ± 1.9	39.8 ± 1.4*	**41.2 ± 2.7**	0.5
Mushroom(k = 7)	45.8 ± 1.0**	45.8 ± 1.8*	46.5 ± 1.5	45.9 ± 1.7**	**46.6 ± 1.4**	0.5
Average	60.1	61.3	62.2	61.5	**65.3**	-
* (p < 0.1)		** (p < 0.05)		*** (p < 0.01)		

5.1. The analysis of effectiveness

To evaluate the effectiveness of the proposed method (denoted as EUF) in enhancing classification tasks, a comparison is made between EUF and three commonly used fusion methods known as MAXF, MINF, and MEANF. Additionally, the state-of-the-art entropy-based supervised fusion method called ESF [10] is included in the comparison. The detailed description of these comparative methods is provided as follows:

(1) Entropy supervised fusion (ESF): The ESF method is aimed at Ms-IIvIS with decision attributes. It utilizes the decision attributes to define the conditional entropy of the information sources and integrates the raw systems by minimizing the conditional entropy. The ESF method has been proven to be an efficient fusion method for Ms-IIvIS.

(2) Max fusion (MAXF): The MAXF method takes the minimum value of the left endpoint and the maximum value of the right endpoint of the interval value of Ms-IS as the fusion result, which can be represented as $MAXF(x,a) = \left[\min\left\{f_1^L(x,a),...,f_N^L(x,a)\right\}, \max\left\{f_1^U(x,a),...,f_N^U(x,a)\right\}\right]$, where $f_i^L(x,a)$ and $f_i^U(x,a)$ are the left and right endpoints of a given object x from the i-th information source for a given attribute a.

(3) Min fusion (MINF): The MINF method takes the maximum value of the left endpoint and the minimum value of the right endpoint of the interval value of Ms-IS as the fusion result, which can be represented as $MINF(x,a) = \left[\max\left\{f_1^L(x,a),...,f_N^L(x,a)\right\}, \min\left\{f_1^U(x,a),...,f_N^U(x,a)\right\}\right]$, where $f_i^L(x,a)$ and $f_i^U(x,a)$ are the left and right endpoints of a given object x from the i-th information source for a given attribute a.

(4) Mean fusion (MEANF): The MEANF method takes the mean value of the left and right endpoints of the interval value of Ms-IS as the fusion result, which can be represented as $MEANF(x,a) = \left[\frac{f_1^L(x,a)+...+f_N^L(x,a)}{N}, \frac{f_1^U(x,a)+...+f_N^U(x,a)}{N}\right]$, where $f_i^L(x,a)$ and $f_i^U(x,a)$ are the left and right endpoints of a given object x from the i-th information source for a given attribute a.

This section performs the classification task using the extended k-nearest neighbor classifier developed in reference [34]. This is because many existing classifiers are not directly applicable to interval-valued data. The parameter β can affect the fusion results of the proposed method, so we adjust the value of β to obtain the optimal classification results. To evaluate the performance of the proposed method, ten-fold cross-validation is conducted, and the average and standard deviation of the classification results are presented in Table 16. From the results, it is evident that the EUF method outperforms the other four methods in nine datasets. Overall, the proposed method achieves an average accuracy of 65.3%, surpassing the other methods (MAXF, MINF, MEANF, and ESF) by 5.0% to 8.7%. However, since cross-validation involves a certain degree of randomness, relying solely on the average classification accuracy may not be enough to demonstrate the method's superiority. Therefore, we employ hypothesis testing to prove the superiority of our proposed approach statistically.

The hypothesis testing, specifically the Wilcoxon signed-rank test, provides a nonparametric statistical framework to assess the significance of the observed differences in classification accuracy between the proposed method EUF and the other comparison algorithms, which can be used to validate the superiority of the proposed method further. In the context of the Wilcoxon signed-rank test, the null hypothesis, denoted as $H_0: Me_{EUF} \le Me_*$, assumes that the median of the classification accuracy distribution for EUF methods, represented by Me_{EUF}, is less than or equal to the median of the classification accuracy distribution for other * methods, represented by Me_*. Conversely, the alternative hypothesis, $H_1: Me_{EUF} > Me_*$, states that the median of the EUF method's accuracy distribution is greater than the median of the other * methods' distribution. By conducting the Wilcoxon signed-rank test and obtaining the associated p-values, we can evaluate the significance of the observed differences. If the p-value is below a specified significance level (e.g., 0.1, 0.05, or 0.01), we reject the null hypothesis and conclude that there is a statistically significant difference in favor of the EUF method's classification accuracy over the other methods' classification accuracy. In the test results, the significance level is indicated by *, **, or ***, corresponding to P-values less than 0.1, 0.05, or 0.01, respectively. In the WC and MM datasets, the proposed EUF method demonstrates statistically significant superiority over the ESF, MAXF, and MINF methods. In the seeds, BCC, Parkinsons, and SAOA datasets, the EUF method outperforms the other four methods with statistical significance. In the UKM dataset, the EUF method is statistically better than the ESF, MINF, and MEANF methods. In the WQW dataset, the proposed method performs better than the MAXF and MEANF methods. In the Mushroom dataset, the EUF approach is statistically significantly

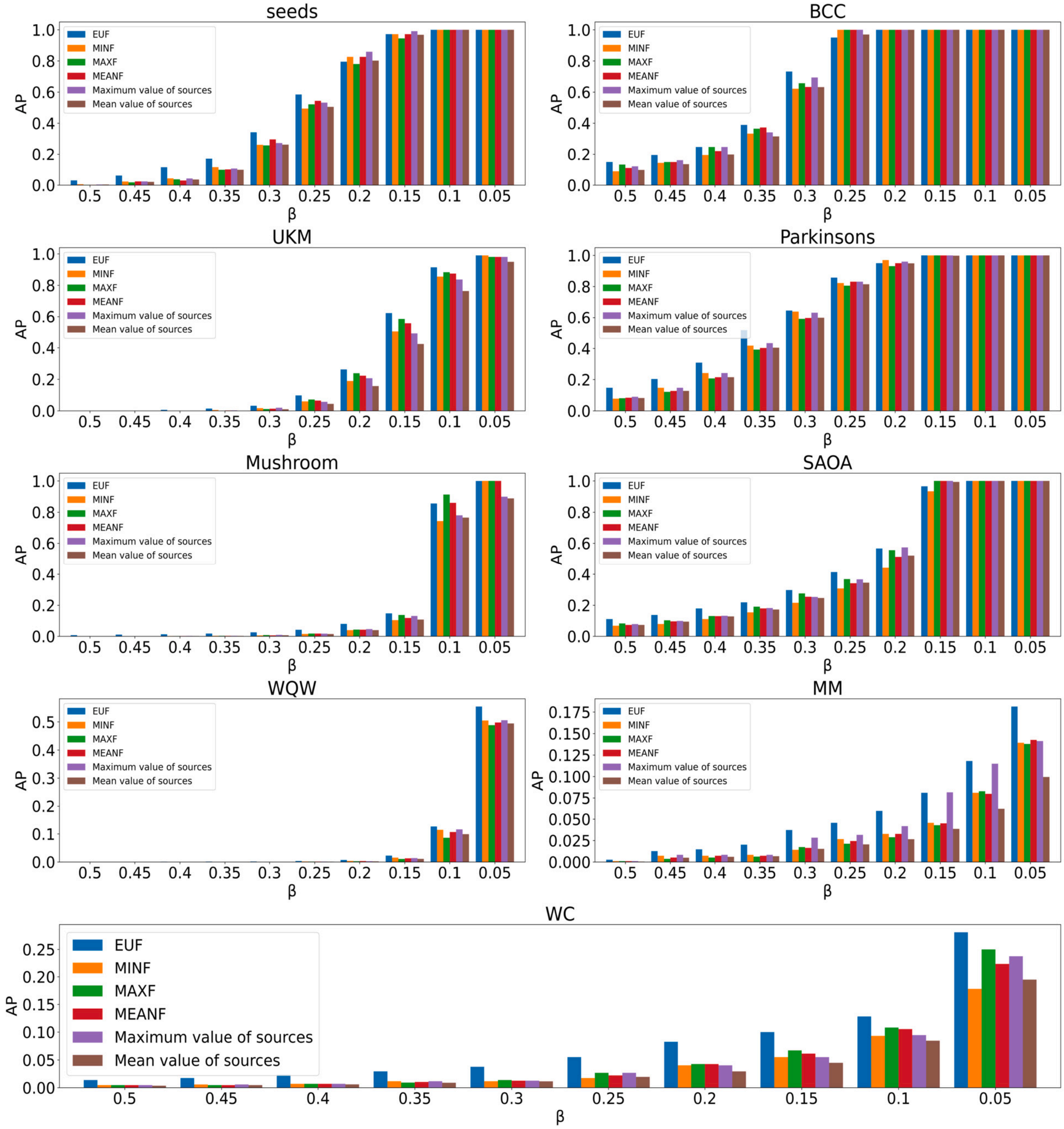

Fig. 4. Comparison of AP of the fusion results.

better than the ESF, MAXF, and MEANF methods. To summarize, the Wilcoxon signed-rank test reveals that 83.3% of the p-values are significant, indicating that the proposed method is statistically superior to the other methods in most situations.

As mentioned earlier, the AP can be utilized to evaluate the approximation classification performance of fusion results, as it reflects the approximation classification performance of the models [28]. In this context, we compare the proposed method (EUF) with MAXF, MINF, MEANF, Maximum value of raw sources, and Mean value of raw sources based on AP. The results are presented in Fig. 4, where the x-axis represents the value of β, and the y-axis represents the AP value. From the results, it can be observed that in the seeds and SAOA datasets, the EUF method outperforms the other methods when β is set to 0.5, 0.45, 0.4, 0.35, 0.3, and 0.25. In the BCC dataset, the EUF method performs better than the five methods when β is set to 0.5, 0.45, 0.35, and 0.3. Similarly, in the UKM dataset, the proposed method outperforms MAXF, MEANF, MINF, Maximum AP of raw sources, and Mean AP of raw sources when β is set to 0.3, 0.25, 0.2, 0.15, and 0.1. In the Parkinsons dataset, the EUF method surpasses the other five methods when β is set from 0.5 to 0.25. In the Mushroom dataset, the EUF method is better than other methods when the range of β is 0.5

Table 17
Comparison of AP of the fusion results(%).

Datasets	EUF	MINF	MAXF	MEANF	Maximum value of sources	Mean value of sources
WC	**7.7 ± 8.1**	4.2 ± 5.6***	5.3 ± 7.7***	4.9 ± 6.9***	5.0 ± 7.2***	4.1 ± 6.0***
seeds	**50.7 ± 41.0**	47.4 ± 43.5**	46.6 ± 43.2***	48.0 ± 43.8**	48.3 ± 44.2*	47.0 ± 43.6**
BCC	**66.6 ± 37.6**	63.8 ± 40.7**	65.5 ± 39.1	64.9 ± 39.7	65.6 ± 39.3	63.5 ± 40.5**
UKM	**29.4 ± 39.7**	26.2 ± 38.3***	27.7 ± 39.1***	27.1 ± 38.8***	25.9 ± 37.7***	23.5 ± 35.6***
Parkinsons	**66.3 ± 34.7**	63.2 ± 38.0**	61.3 ± 38.4**	62.1 ± 38.4**	63.3 ± 37.6**	61.9 ± 38.3***
MM	**5.7 ± 5.6**	3.6 ± 4.3***	3.5 ± 4.4***	3.6 ± 4.4***	4.7 ± 4.9***	2.8 ± 3.1***
SAOA	**48.9 ± 37.0**	43.1 ± 39.4***	47.1 ± 39.1**	45.9 ± 39.5**	46.9 ± 39.4**	45.7 ± 39.5**
WQW	**7.2 ± 17.4**	6.4 ± 15.9***	5.9 ± 15.3***	6.2 ± 15.6***	6.4 ± 15.9***	6.1 ± 15.5***
Mushroom	**22.0 ± 37.7**	19.2 ± 36.5***	21.3 ± 39.5*	20.5 ± 38.5***	18.9 ± 34.6***	18.3 ± 34.2***
Average	**33.8**	30.8	31.6	31.5	31.7	30.3
	* ($p < 0.1$)		** ($p < 0.05$)		*** ($p < 0.01$)	

to 0.15. In the WQW dataset, the proposed method outperforms the five methods when β is set to 0.2 and 0.05. In the MM dataset, the EUF method performs better than the other methods except for β of 0.15. In the WC dataset, the EUF method outperforms the other five methods consistently across all situations. Furthermore, it can be observed that in the seeds, BCC, UKM, Parkinsons, SAOA, and Mushroom datasets, the AP values equal 1 or are very close to 1 when β is set to 0.1 or 0.05. This is because smaller values of β result in smaller tolerance classes, leading to smaller upper approximations and larger lower approximations, yielding larger AP values. In this situation, although AP becomes high, the tolerance classes of the objects at this time will become small, making the model's generalization ability poor. Therefore, it is generally best to avoid using too small a β value. In summary, these comparisons demonstrate the superior performance of the EUF method across various datasets when considering AP. They indicate that in most situations, the EUF method consistently outperforms or performs on par with the other five fusion methods in terms of AP.

To enhance the persuasiveness of comparisons, Table 17 presents the means and standard deviations of AP for all parameter settings. The average AP of the proposed method is 33.8%, which is higher by 7.0% to 9.7% compared to MAXF, MEANF, and MEANF. This indicates that the proposed method outperforms these three common fusion methods in terms of AP. Furthermore, the average AP of the proposed method is higher by 6.6% to 11.6% compared to the maximum and mean value of the raw information sources. This suggests that the proposed method effectively promotes the AP of the raw sources. Moreover, we can employ the Wilcoxon signed-rank test to establish the superiority of the proposed method. The null hypothesis, denoted as $H_0 : Me_{EUF} \leq Me_*$, assumes that the median of the AP distribution for EUF methods, represented by Me_{EUF}, is less than or equal to the median of the AP distribution for other * methods, represented by Me_*. Conversely, the alternative hypothesis, $H_1 : Me_{EUF} > Me_*$, states that the median of the EUF method's AP distribution is greater than the median of the other * methods' AP distribution. When the obtained p-value falls below a predetermined significance level, we reject the null hypothesis and infer that a statistically significant difference exists, favoring the EUF method's AP over the other methods' AP. From the results of the Wilcoxon signed-rank test, we can find that in the WC, seeds, UKM, Parkinsons, MM, SAOA, WQW, and Mushroom datasets, the EUF method outperforms the other five methods with statistically significant. And in the BCC dataset, the EUF method is statistically significantly better than the MINF and Mean value of sources. Overall, the utilization of the Wilcoxon signed-rank test demonstrates that the proposed method exhibits statistically significant superiority over MAXF, MINF, MEANF, the Maximum value of sources, and the Mean value of sources across 93% of situations.

Furthermore, the comparison of the EUF with the state-of-the-art fusion method based on AP is illustrated in Fig. 5. From the results, we can find that in the Parkinsons dataset, the proposed method performs better than ESF when the value of β is set to 0.5, 0.45, 0.4, 0.35, 0.3 and 0.25. In the WC dataset, when the value of β is set to 0.4, 0.35, 0.3, 0.25 and 0.2, the AP of the EUF method surpasses that of the ESF method. However, there is no denying that in most datasets, our method exhibits no significant difference compared to ESF in terms of AP. This can be attributed to the fact that the ESF method is a supervised fusion approach that leverages information from sample labels during the fusion process, which proves beneficial in improving the AP of the fusion results. In contrast, this study focuses on unsupervised samples; therefore, it is reasonable that our approach does not surpass ESF in terms of AP. Furthermore, it should be emphasized that smaller values of β result in higher AP. This is because smaller values of β lead to smaller tolerance classes, resulting in higher lower approximation and smaller upper approximation. As a result, the value of AP increases. However, setting β too small may reduce the model's generalization ability, as the tolerance classes become excessively small. Therefore, in real applications, it is essential to consider both the model's generalization ability and approximation classification performance to select an appropriate parameter value for β.

5.2. The analysis of efficiency

In Section 4, two dynamic update approaches are introduced, aiming to compute the TC fast based on the former results, whose efficiencies are independent of the number of information sources. The comparison of time complexities of static fusion algorithm and dynamic fusion algorithms can be seen in Table 18. Theoretical analysis shows that the dynamic fusion algorithms have lower time complexity than the static fusion method in both cases.

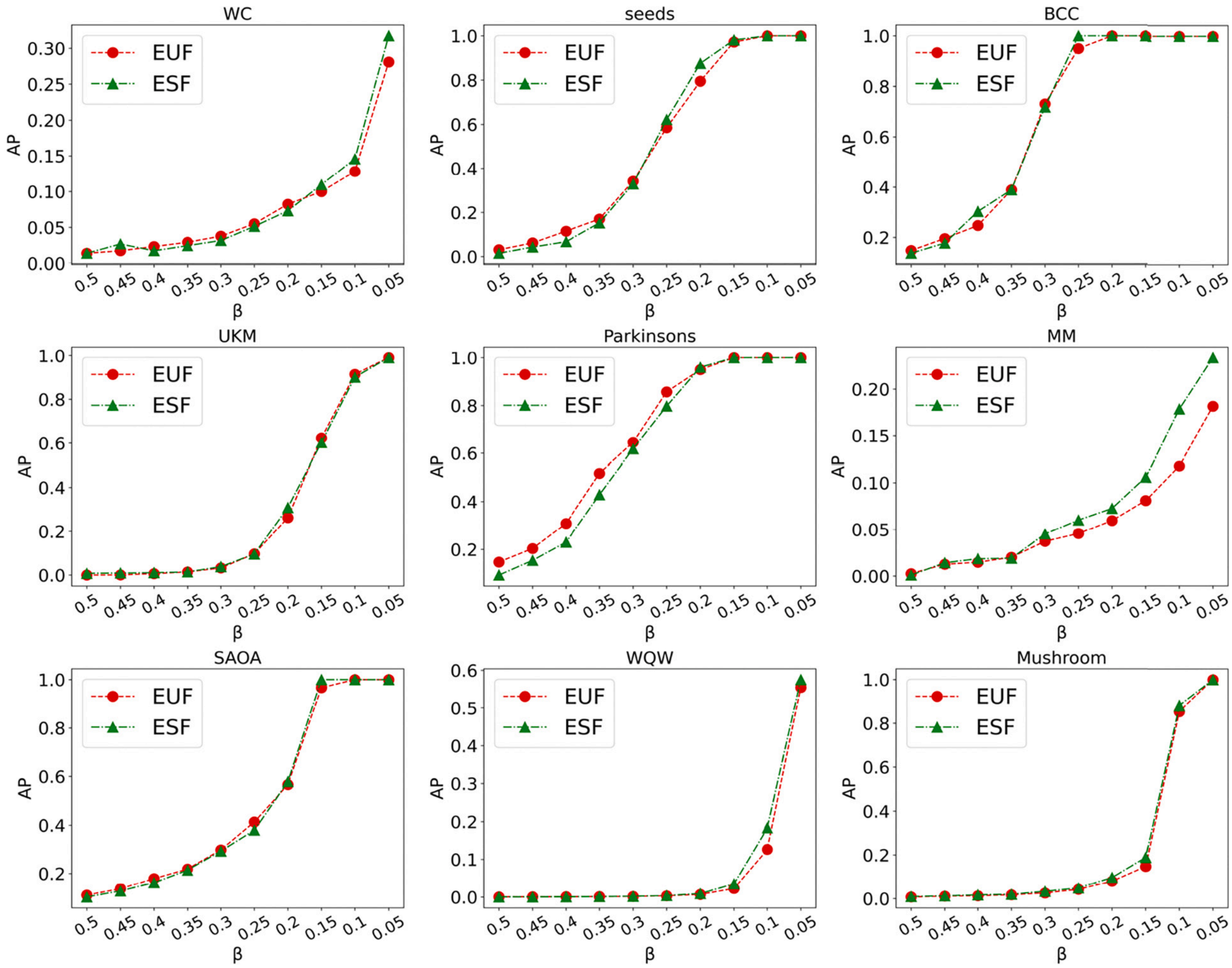

Fig. 5. Comparison of AP of the proposed method EUF and the state-of-the-art fusion method ESF.

Table 18
Comparison of time complexities of static fusion algorithm and dynamic fusion algorithms(%).

Cases	Static algorithm	Dynamic algorithm
Inserting objects	$O\left(N \times \|A\| \times (\text{n}+\Delta\text{n})^2 + \|A\|\right)$	$O(N \times \|A\| \times (n + \Delta\text{n} \times (n + \Delta\text{n})) + \|A\|)$
Deleting objects	$O\left(N \times \|A\| \times (\text{n}-\Delta\text{n})^2 + \|A\|\right)$	$O(N \times \|A\| \times (\text{n}-\Delta\text{n}) + \|A\|)$

To conduct empirical analysis, we compared the running time of the static and dynamic algorithms in three scenarios: when the number of sources is one, three, and five. To simulate scenarios when object sets increase, we initially set aside half of the objects and gradually add 10% of the remaining data each time. Conversely, to simulate scenarios with object sets decreasing, we start with all the objects and gradually reduce them by 10% at each time until we have half of the original objects remaining. It is worth noting that the object inserting and deleting ratios range from 10% to 50%, with the step of 10%. For simplicity and without sacrificing generality, we set the value of β to 0.01. The mean value and standard deviation of the running time are presented in Tables 19–21. For the case of one source, it can be observed that in all nine datasets, the two dynamic fusion algorithms outperform the static fusion algorithm. The average running time of the dynamic fusion algorithms is lower by 68.1% to 85.6% compared to the static fusion algorithm. Similarly, for the cases of three and five sources, it can be observed that in all datasets, the two dynamic fusion algorithms outperform the static fusion algorithm. The average running time of the dynamic fusion algorithms is lower by 67.6% to 85.2% and 66.9% to 85.0% compared to the static fusion algorithm, respectively.

Next, the Wilcoxon signed-rank test is utilized to evaluate whether the running time of the static algorithm is statistically larger than that of the dynamic algorithms. The null hypothesis can be denoted as $H_0 : Me_{Static} \leq Me_{Dynamic}$, assumes that the median of the runtime distribution for Static algorithm, represented by Me_{Static}, is less than or equal to the median of the runtime distribution for dynamic algorithm, represented by $Me_{Dynamic}$. Conversely, the alternative hypothesis, $H_1 : Me_{Static} > Me_{Dynamic}$, states that the median of the Static algorithm's runtime distribution is greater than the median of the Dynamic algorithm's runtime distribution.

Table 19
Comparison of running time of one information source(s).

Datasets	Inserting objects		Deleting objects	
	Static	Dynamic	Static	Dynamic
WC	2.2 ± 1.0	**0.4 ± 0.3****	2.2 ± 1.1	**0.7 ± 1.5****
seeds	0.5 ± 0.2	**0.1 ± 0.1****	0.5 ± 0.3	**0.2 ± 0.4****
BCC	0.2 ± 0.1	**0.1 ± 0.0****	0.2 ± 0.1	**0.1 ± 0.1****
UKM	1.4 ± 0.6	**0.2 ± 0.2****	1.3 ± 0.6	**0.4 ± 0.9****
Parkinsons	1.4 ± 0.6	**0.3 ± 0.2****	1.4 ± 0.6	**0.5 ± 0.9****
MM	7.4 ± 3.5	**1.2 ± 1.0****	7.0 ± 3.3	**2.6 ± 4.6****
SAOA	92.0 ± 42.7	**13.4 ± 13.5****	89.0 ± 42.3	**28.8 ± 59.8****
WQW	406.0 ± 191.7	**58.3 ± 55.2****	391.6 ± 192.9	**124.5 ± 267.7****
Mushroom	2134.5 ± 1065.8	**305.7 ± 285.2****	2099.1 ± 1035.7	**670.0 ± 1440.8****
Average	293.9555556	**42.18888889**	288.0333333	**91.97777778**
* ($p<0.1$)		** ($p<0.05$)		*** ($p<0.01$)

Table 20
Comparison of running time of three information source(s).

Datasets	Inserting objects		Deleting objects	
	Static	Dynamic	Static	Dynamic
WC	6.7 ± 3.1	**1.1 ± 0.9****	6.8 ± 3.2	**2.2 ± 4.5****
seeds	1.6 ± 0.7	**0.4 ± 0.2****	1.6 ± 0.7	**0.6 ± 1.1****
BCC	0.7 ± 0.3	**0.2 ± 0.1****	0.7 ± 0.3	**0.3 ± 0.4****
UKM	3.9 ± 1.9	**0.7 ± 0.5****	4.0 ± 1.9	**1.3 ± 2.6****
Parkinsons	4.2 ± 1.9	**0.9 ± 0.5****	4.3 ± 1.9	**1.6 ± 2.7****
MM	21.9 ± 10.4	**3.6 ± 2.9****	21.8 ± 10.4	**8.1 ± 14.3****
SAOA	270.5 ± 127.7	**38.6 ± 38.2****	271.6 ± 129.4	**87.5 ± 182.1****
WQW	1196.6 ± 545.7	**177.2 ± 175.9****	1236.8 ± 589.4	**391.6 ± 833.0****
Mushroom	6467.1 ± 3254.8	**955.9 ± 902.0****	6376.3 ± 3133.2	**2070.2 ± 4358.7****
Average	885.9111111	**130.9555556**	880.4333333	**284.8222222**
* ($p<0.1$)		** ($p<0.05$)		*** ($p<0.01$)

Table 21
Comparison of running time of five information source(s).

Datasets	Inserting objects		Deleting objects	
	Static	Dynamic	Static	Dynamic
WC	10.9 ± 5.0	**1.8 ± 1.5****	10.7 ± 5.0	**3.5 ± 7.1****
seeds	2.5 ± 1.1	**0.6 ± 0.3****	2.6 ± 1.2	**0.9 ± 1.7****
BCC	1.1 ± 0.4	**0.3 ± 0.2****	1.2 ± 0.5	**0.4 ± 0.7****
UKM	6.5 ± 3.0	**1.2 ± 0.9****	6.5 ± 3.0	**2.1 ± 4.3****
Parkinsons	6.8 ± 3.0	**1.5 ± 0.9****	7.0 ± 3.2	**2.6 ± 4.5****
MM	35.8 ± 16.9	**5.9 ± 4.8****	35.2 ± 16.5	**12.8 ± 22.7****
SAOA	436.0 ± 204.4	**63.0 ± 61.7****	441.6 ± 208.6	**142.1 ± 294.2****
WQW	1963.2 ± 925.3	**290.3 ± 276.5****	1978.3 ± 956.8	**632.0 ± 1341.7****
Mushroom	11120.9 ± 5433.4	**1672.8 ± 1405.7****	10460.9 ± 5277.3	**3483.1 ± 7213.8****
Average	1509.3	**226.3777778**	1438.222222	**475.5**
* ($p<0.1$)		** ($p<0.05$)		*** ($p<0.01$)

When the p-value is smaller than a predetermined significance level, we can reject the null hypothesis. This rejection implies that there is statistically solid evidence supporting a significant difference, favoring the dynamic methods over the static method. The results of the Wilcoxon signed-rank test demonstrate that the dynamic algorithms exhibit statistically significant superiority over the static algorithm across 100% situations. These results indicate that the running time of the dynamic fusion algorithms is statistically significantly less than that of the static fusion algorithm in all nine datasets when inserting or deleting objects.

Moreover, the speed-up ratio (SUR) is utilized as a metric for comparison, which is computed as

$$SUR = \frac{t(static\ fusion\ algorithm)}{t(dynamic\ fusion\ algorithm)}, \tag{14}$$

where $t(*)$ denotes the running time of the respective algorithm. Figs. 6-7 illustrate the relationship between the inserting and deleting ratios of objects (x-axis) and the SUR of the dynamic fusion algorithms (y-axis). When objects are inserted, significant observations can be made across different datasets. In the WC dataset, the SUR values for one, three, and five sources consistently exceed 6. Similarly, in the seeds and BCC datasets, the SUR values for all three cases are consistently greater than 3. The UKM and

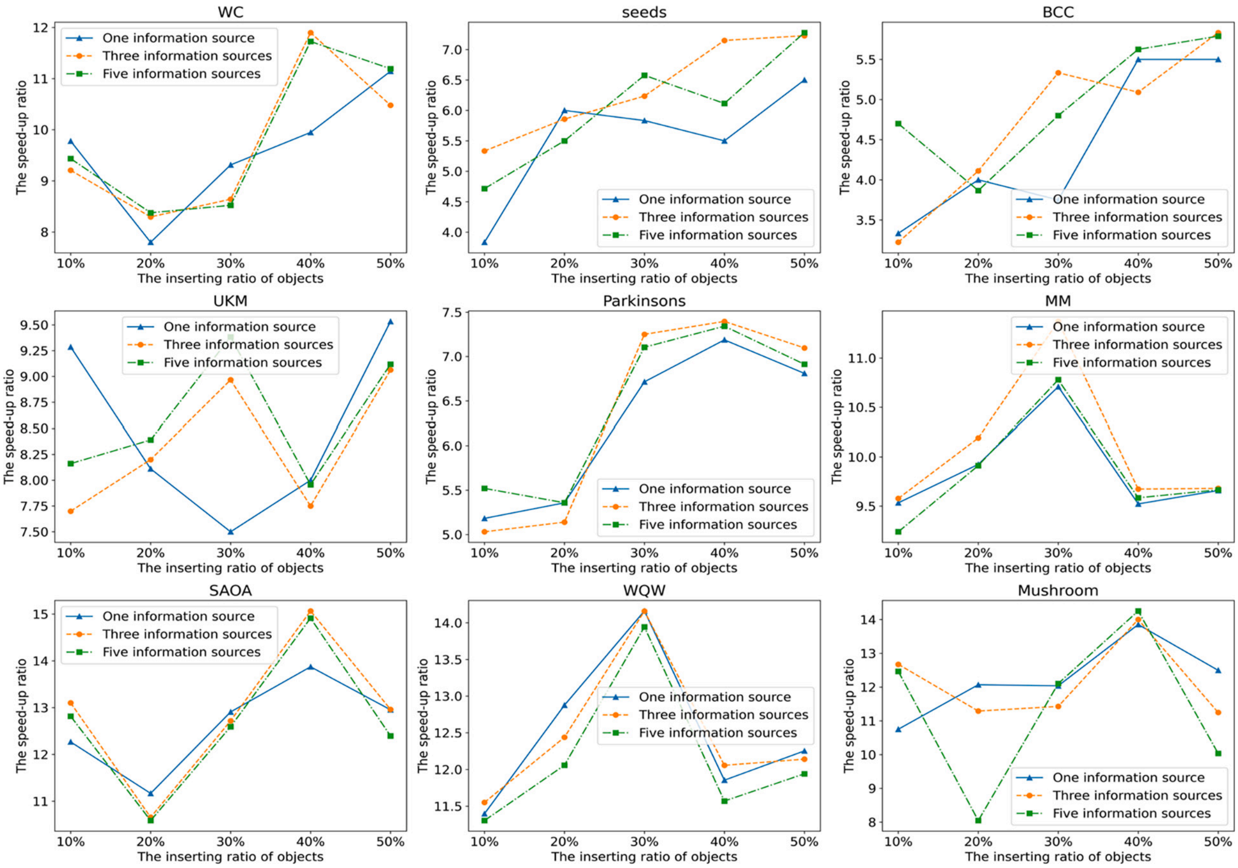

Fig. 6. Comparison results of SUR of the static and dynamic fusion methods when inserting objects.

Mushroom datasets exhibit SUR values greater than 6 for dynamic algorithms under one, three, and five sources. In the Parkinsons dataset, the SUR values for all three cases surpass 4. The MM and SAOA datasets demonstrate SUR values exceeding 8 for one, three, and five sources. Lastly, the WQW dataset showcases SUR values higher than 10 for all three cases. Likewise, when objects are deleted, additional insights can be gleaned. Notably, in the WC, SAOA, WQW, and Mushroom datasets, the SUR values for one, three, and five sources consistently exceed 10. In the seeds and Parkinsons datasets, the SUR values for the dynamic algorithm under three cases are consistently higher than 4. The BCC dataset reports SUR values exceeding 2 for all three cases. The UKM dataset exhibits SUR values greater than 5 under all three situations. Finally, in the MM dataset, the SUR values for the dynamic algorithm under one, three, and five sources are all higher than 2.5. In summary, the SUR values for both dynamic fusion algorithms consistently surpass 2, indicating their effectiveness. Notably, the highest observed SUR reaches an impressive value of 33.38852459, highlighting the remarkable efficiency of the proposed dynamic fusion approaches in reducing computational time. These findings emphasize the potential benefits of the proposed dynamic fusion algorithms for updating fusion results in dynamic datasets.

Moreover, we can observe that when objects are inserted, the SUR curves of the three scenarios show significant overlap in the Parkinsons, MM, SAOA, WQW, and Mushroom datasets. Similarly, when objects are deleted, the SUR curves under one, three, and five sources almost coincide in the WC, seeds, BCC, UKM, Parkinsons, MM, SAOA, and WQW datasets. Excluding the random effects during program operation, the SUR curves of the three cases overlap in almost all scenarios. This indicates that the number of information sources does not significantly influence the efficiencies of the dynamic fusion methods. This observation provides evidence for the robustness of our proposed dynamic fusion approaches.

6. Conclusions

Studying efficient methodologies for the fusion of Ms-IIvIS is of great scholarly importance as it can accurately represent uncertain phenomena inherent in human existence. This paper proposed an unsupervised fusion framework for Ms-IIvIS by utilizing IE. The proposed framework consists of several steps. First, a novel IE is defined to evaluate the uncertainty of IIvIS with respect to attributes. This allows for the selection of the most essential source for each attribute by minimizing the entropy. Next, the fusion results are completed by employing the union of objects in the TC of missing value objects. Additionally, two dynamic fusion approaches are introduced to handle the insertion of new objects or the removal of obsolete objects from the systems. Experiments conducted on common datasets demonstrate the effectiveness and efficiency of the proposed fusion method. The effectiveness analysis reveals that

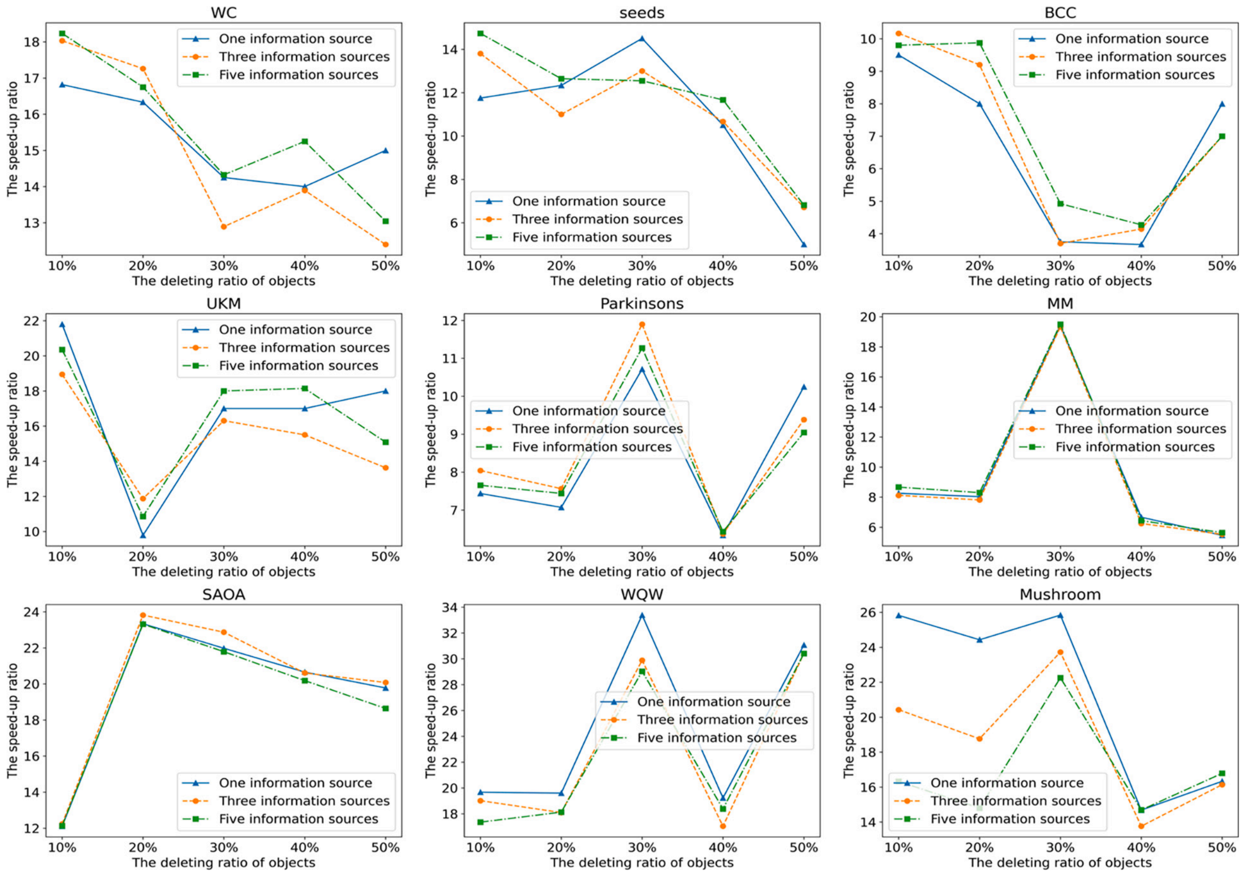

Fig. 7. Comparison results of SUR of the static and dynamic fusion methods when deleting objects.

the average classification accuracy of the proposed method is 5% to 8.7% higher compared to other fusion methods such as MAXF, MEANF, MINF, and ESF. The efficiency analysis shows that the dynamic fusion algorithms have an average running time that is 66.9% to 85.6% lower compared to the static fusion algorithm.

Significant efforts have been made to achieve a unified representation of Ms-IIvIS and promptly update fusion results. However, this research identifies limitations and challenges, including: (1)The proposed incremental fusion method is limited by its reliance on serial computing, leading to potential inefficiency with large-scale datasets. To enhance the algorithm's performance, future research can explore the utilization of distributed and parallel computing frameworks for simultaneous calculation and updating entropy in MS-IS to improve fusion efficiency [35,36]. (2)Information entropy has successfully assessed uncertainty in linguistic information for linguistic decision-making [37]. Expanding the proposed entropy-based fusion model to include multi-source linguistic data can enhance its effectiveness and applicability in this field.

CRediT authorship contribution statement

Xiuwei Chen: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Data curation. **Maokang Luo:** Validation, Supervision, Methodology, Investigation.

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

The data that has been used is common.

References

[1] M. Xu, Z. Qin, A bivariate Bayesian method for interval-valued regression models, Knowl.-Based Syst. 235 (2022) 107396.

[2] L. Kong, X. Song, X. Wang, Nonparametric regression for interval-valued data based on local linear smoothing approach, Neurocomputing 501 (2022) 834–843.
[3] J. Shi, Z. Suo, Three-way decisions method based on matrices approaches oriented dynamic interval-valued information system, Int. J. Approx. Reason. 149 (2022) 116–130.
[4] X.-H. Pan, S.-F. He, Y.-M. Wang, L. Martínez, A novel interval-valued three-way decision theory under multiple criteria environment, Knowl.-Based Syst. 253 (2022) 109522.
[5] W. Li, H. Zhou, W. Xu, X.-Z. Wang, W. Pedrycz, Interval dominance-based feature selection for interval-valued ordered data, IEEE Trans. Neural Netw. Learn. Syst. 34 (10) (2023) 6898–6912.
[6] J. Dai, Z. Wang, W. Huang, Interval-valued fuzzy discernibility pair approach for attribute reduction in incomplete interval-valued information systems, Inf. Sci. 642 (2023) 119215.
[7] Q. Pan, Multi-Soure Information Fusion Theory and Its Applications, Tsinghua University Press, Beijing, 2013.
[8] P. Zhang, T. Li, G. Wang, C. Luo, H. Chen, J. Zhang, D. Wang, Z. Yu, Multi-source information fusion based on rough set theory: a review, Inf. Fusion 68 (2021) 85–117.
[9] W. Wei, J. Liang, Information fusion in rough set theory: an overview, Inf. Fusion 48 (2019) 107–118.
[10] X. Zhang, X. Chen, W. Xu, W. Ding, Dynamic information fusion in multi-source incomplete interval-valued information system with variation of information sources and attributes, Inf. Sci. 608 (2022) 1–27.
[11] W. Xu, Y. Pan, X. Chen, W. Ding, Y. Qian, A novel dynamic fusion approach using information entropy for interval-valued ordered datasets, IEEE Trans. Big Data 9 (3) (2023) 845–859.
[12] H. Zhang, E. Liu, B. Zhang, Q. Miao, Rul prediction and uncertainty management for multisensor system using an integrated data-level fusion and upf approach, IEEE Trans. Ind. Inform. 17 (7) (2021) 4692–4701.
[13] A. Mirzaei, H. Bagheri, M. Sattari, Data level and decision level fusion of satellite multi-sensor aod retrievals for improving pm2.5 estimations, a study on Tehran, Earth Sci. Inform. 16 (7) (2023) 753–771.
[14] D. Song, T. Ma, Y. Li, F. Xu, Data and decision level fusion-based crack detection for compressor blade using acoustic and vibration signal, IEEE Sens. J. 22 (12) (2022) 12209–12218.
[15] Z. Pawlak, Rough set theory and its applications to data analysis, Cybern. Syst. 29 (7) (1998) 661–688.
[16] D. Li, H. Zhang, T. Li, A. Bouras, X. Yu, T. Wang, Hybrid missing value imputation algorithms using fuzzy c-means and vaguely quantified rough set, IEEE Trans. Fuzzy Syst. 30 (5) (2022) 1396–1408.
[17] B. Cao, J. Zhao, Z. Lv, Y. Gu, P. Yang, S.K. Halgamuge, Multiobjective evolution of fuzzy rough neural network via distributed parallelism for stock prediction, IEEE Trans. Fuzzy Syst. 28 (5) (2020) 939–952.
[18] B. Sang, H. Chen, L. Yang, T. Li, W. Xu, C. Luo, Feature selection for dynamic interval-valued ordered data based on fuzzy dominance neighborhood rough set, Knowl.-Based Syst. 227 (2021) 107223.
[19] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (3) (1948) 379–423.
[20] M. Li, X. Zhang, Information fusion in a multi-source incomplete information system based on information entropy, Entropy 19 (11) (2017).
[21] W. Xu, M. Li, X. Wang, Information fusion based on information entropy in fuzzy multi-source incomplete information system, Int. J. Fuzzy Syst. 19 (2017) 1200–1216.
[22] P. Zhang, T. Li, Z. Yuan, C. Luo, G. Wang, J. Liu, S. Du, A data-level fusion model for unsupervised attribute selection in multi-source homogeneous data, Inf. Fusion 80 (2022) 87–103.
[23] W. Xu, J. Yu, A novel approach to information fusion in multi-source datasets: a granular computing viewpoint, Inf. Sci. 378 (2017) 410–423.
[24] B. Sang, Y. Guo, D. Shi, W. Xu, Decision-theoretic rough set model of multi-source decision systems, Int. J. Mach. Learn. Cybern. 9 (2018) 1941–1954.
[25] G. Lin, J. Liang, Y. Qian, An information fusion approach by combining multigranulation rough sets and evidence theory, Inf. Sci. 314 (2015) 184–199.
[26] L. Yang, W. Xu, X. Zhang, B. Sang, Multi-granulation method for information fusion in multi-source decision information system, Int. J. Approx. Reason. 122 (2020) 47–65.
[27] X. Chen, W. Xu, Double-quantitative multigranulation rough fuzzy set based on logical operations in multi-source decision systems, Int. J. Mach. Learn. Cybern. (2022) 1–28.
[28] Y. Huang, T. 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.
[29] Y. Huang, T. Li, C. Luo, H. Fujita, S.-j. Horng, B. Wang, Dynamic maintenance of rough approximations in multi-source hybrid information systems, Inf. Sci. 530 (2020) 108–127.
[30] C. Luo, T. Li, H. Chen, J. Lv, Z. Yi, Fusing entropy measures for dynamic feature selection in incomplete approximation spaces, Knowl.-Based Syst. 252 (2022) 109329.
[31] M. Kryszkiewicz, Rules in incomplete information systems, Inf. Sci. 113 (3–4) (1999) 271–292.
[32] M. Kryszkiewicz, Rough set approach to incomplete information systems, Inf. Sci. 112 (1–4) (1998) 39–49.
[33] Z. Pawlak, A. Skowron, Rudiments of rough sets, Inf. Sci. 177 (1) (2007) 3–27.
[34] J. Dai, W. Wang, J.-S. Mi, Uncertainty measurement for interval-valued information systems, Inf. Sci. 251 (2013) 63–78.
[35] C. Luo, S. Wang, T. Li, H. Chen, J. Lv, Z. Yi, Large-scale meta-heuristic feature selection based on bpso assisted rough hypercuboid approach, IEEE Trans. Neural Netw. Learn. Syst. 34 (12) (2023) 10889–10903.
[36] C. Luo, S. Wang, T. Li, H. Chen, J. Lv, Z. Yi, Rhdofs: a distributed online algorithm towards scalable streaming feature selection, IEEE Trans. Parallel Distrib. Syst. 34 (6) (2023) 1830–1847.
[37] Y. Liu, L. Zhu, R.M. Rodríguez, L. Martínez, Personalized fuzzy semantic model of PHFLTS: application to linguistic group decision making, Inf. Fusion 103 (2024) 102118.