Information Sciences 608 (2022) 1–27

Contents lists available at ScienceDirect

Information Sciences

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

Dynamic information fusion in multi-source incomplete interval-valued information system with variation of information sources and attributes

Xiaoyan Zhang [a], Xiuwei Chen [a,*], Weihua Xu [a], Weiping Ding [b]

[a] College of Artificial Intelligence, Southwest University, Chongqing 400715, PR China
[b] School of Information Science and Technology, Nantong University, Nantong 226019, PR China

ARTICLE INFO

Article history:
Received 2 November 2021
Received in revised form 11 June 2022
Accepted 13 June 2022
Available online 15 June 2022

Keywords:
Dynamic fusion
Information entropy
Incomplete interval-valued
Multi-source information system

ABSTRACT

Interval-valued data describe the random phenomenon that abounds in the real world, a pivotal research orientation in uncertainty processing. With the rapid development of big data, we may gather information from multiple information sources. To effectively acquire knowledge from multiple information sources, information fusion is commonly used to get a unified representation. However, sometimes data gathered from multiple sources may be lost; it is meaningful and necessary to study the fusion of multi-source incomplete interval-valued data. We propose a novel information fusion method based on information entropy for multi-source incomplete interval-valued data and four incremental fusion mechanisms characterized by the change in information sources and attributes. The corresponding static and dynamic fusion algorithms are designed, and their time complexities are analyzed. Experimental results show that the proposed method outperforms the mean, max, and min fusion methods. Furthermore, the four incremental fusion mechanisms reduced the runtime compared with the static fusion mechanism.

1. Introduction

Nowadays, the amount, rate of growth, and frequency of alteration of data are increasing unprecedentedly. Interval-valued data can be effectively used to characterize random variation and imprecise information compared to the single-valued data, such as blood pressure [1], temperature [2], and unemployment [3]. Many studies have been conducted concerning classification [4], clustering [5], and mining of interval-valued data, aiming to extract knowledge from complete information tables. However, in the real world, owing to breakdown of sensors or errors on the part of data collectors, the collected data tables may have some missing value. In recent years, numerous studies on incomplete information systems, such as uncertainty measuring, feature selection, and knowledge acquisition have been conducted. In 2013, Dai et al. [6] established a novel uncertainty measuring approach for incomplete data, in 2017 [7] proposed a knowledge granularity measuring method for incomplete interval-valued data, and, Luo et al. [8] proposed an incremental feature selection mechanism for incomplete data. Zhao et al. [9] selected features from an incomplete decision table using a novel extended rough set model. Sun et al. [10] proposed a feature selection method for incomplete systems using Lebesgue and entropy measures. Regarding knowledge acquisition, Li et al. [11] proposed a novel interval set model to induce classification rules

* Corresponding author.
E-mail addresses: zxy19790915@163.com (X. Zhang), xiuweichen1998@163.com (X. Chen), chxuwh@gmail.com (W. Xu), dwp9988@163.com (W. Ding).

https://doi.org/10.1016/j.ins.2022.06.054
0020-0255/

from incomplete data. Hong et al. [12] established a new rules-deriving algorithm for incomplete data by estimating the missing value in the process of rule deriving. Leung et al. [13] efficiently acquired knowledge for incomplete information systems by defining a set of simpler discernibility functions. The above studies focus on single-source information systems; however, data may be available from multiple sensors or sources. For example, we can acquire patient data from different hospitals to predict diseases more accurately. Thus, it is essential to study the fusion of multi-source incomplete interval-valued information systems (MS-IIVIS).

Information fusion is a valid method to process MS-IIVIS to transform and fuse information from multiple and different data sources to integrate information representation. The rough set theory (RST) and information entropy are two effective tools for information fusion. RST [14] has been confirmed as an effective method to fuse information. Several studies have been conducted in the past few decades. Xu et al. [15] defined the importance of information sources for information fusion using internal and external confidence. Sang et al. [16] established three kinds of multi-granulation models for decision-making processes. On the other hand, information entropy proposed by Shannon [17] is a kind of measure of uncertainty, which can be widely used in information fusion. In recent years, many studies combining RST and information entropy have been reported. Xu et al. [18] utilized conditional entropy to define the degree of importance of the source for an attribute for fuzzy data. Xu et al. [19] proposed a novel fuzzy neighborhood conditional entropy to select gene features. Clearly, these methods cannot be directly employed to process MS-IIVIS. Therefore, we propose a fusion approach to effectively fuse MS-IIVIS based on conditional entropy. First, the distance between two interval-valued samples is determined. Then, the tolerance relation can be derived based on the defined distance. The uncertainty measure of the information system for an attribute, called conditional entropy, can be calculated. Finally, a new unified table can be obtained by searching the minimum conditional entropy.

There are two motivations for proposing the dynamic information fusion method in this study. On the one hand, many researchers focus on the static fusion mechanism such as[15,16,18,20–23]. These static fusion methods aim to fuse multi-source information systems without the variation of information systems. However, information systems change due to various reasons. With changes in multi-source information systems, there is no denying that the above methods cannot be directly applied. On the other hand, existing studies propose incremental updating methods only for single information sources [24–26]. For example, they establish incremental updating mechanisms with the change of objects or object set [27–29] and the addition or deletion of attributes [24,30]. These updating methods are intended to update approximations or knowledge of single information sources. However, in the era of big data, it is necessary to study dynamic updating methods in multi-source information systems (MS-ISs). Undoubtedly, for MS-IS dynamic updating, the above methods cannot be directly applied. Huang et al. [31] proposed an incremental fusion method for interval-valued data for dynamic information sources, and then in 2020, they established a dynamic maintenance mechanism [32] in multi-source hybrid information systems. However, when the number of sources change, the number of attributes may also change. For instance, to predict the weather conditions, several weather sensors acquire weather information such as temperature, humidity, and wind speed. However, some prediction models may not need wind speed for the prediction, so these weather sensors are set not to collect the wind speed, while other models may need more variables to predict the weather condition more accurately resulting in the addition of more sensors. Thus, it is necessary to study dynamic fusion with the variation of sources and conditional attributes. Nonetheless, the above approaches are not appropriate for fusing MS-IIVIS with the variation of information sources and attributes. We therefore propose four incremental fusion mechanisms for four scenarios: the addition of sources and deletion of attributes, the addition of sources and attributes, the deletion of sources and attributes, and the deletion of sources and the addition of attributes. The contributions of the study are summarized as follows:

(1) We define a tolerance relation in an incomplete interval-valued system. Based on the relation, we propose a novel conditional entropy to measure the importance of sources to attributes.
(2) A novel fusion framework for multi-source incomplete interval-valued data is presented, shown in Fig. 1. The approach selects the value of the attribute of the source corresponding to the minimum conditional entropy as the fusion result. Experimental results show that our fusion approach outperforms three common fusion methods in terms of approximation precision (AP), approximation quality (AQ), and classification accuracy.
(3) Four dynamic updating mechanisms are established with the variation of sources and attributes. The efficiency analysis manifests that the four dynamic fusion approaches can effectively reduce the computation time when the sources and attributes change simultaneously.

Although much effort has been dedicated to obtaining a unified representation of multi-source incomplete interval-valued data and promptly updating the fusion results, the noted algorithms suffer from the following limitations and challenges.

(1) The entropy-based fusion algorithm needs to compute the conditional entropy of each attributes under each information system, whose time complexity is $O(N \times |AT| \times (|U| \times (|U| + |U/D|)))$. It takes a lot of time for the proposed fusion algorithm to obtain the fusion results when the number of sources and attributes is large.
(2) The dynamic updating approaches can update the fusion results quickly when sources and attributes change simultaneously. A more general situation is that the sources, attributes, and samples change simultaneously. Nevertheless, our method cannot cope with multi-source data that simultaneously change sources, attributes, and samples.

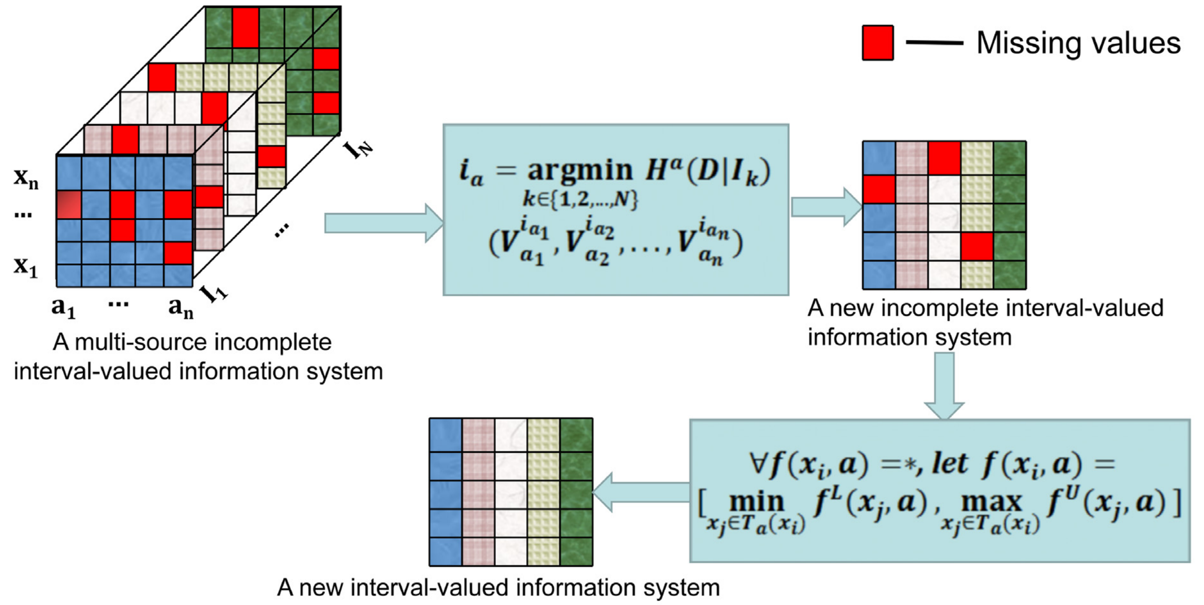

Fig. 1. The fusion framework of this paper.

The remainder of this paper is structured as follows. Section 2 gives some basic concepts, such as RST, incomplete information system (IIS), conditional entropy(CE), and multi-source incomplete interval-valued decision system (MS-IIVDS). In Section 3, a static fusion method for MS-IIVIS is proposed utilizing conditional entropy; the corresponding static algorithm is established and the time complexity is analyzed. In Section 4, four incremental fusion cases are discussed with the variation of sources and attributes. The corresponding dynamic fusion algorithms are established, and their time complexities are analyzed. Later, in Section 5, experimental analysis is conducted to evaluate the effectiveness of fusion and the efficiency of incremental fusion methods. Finally, we conclude the paper and propose future research direction in Section 6. All terminologies and abbreviations are presented in Table 1.

2. Preliminaries

In this section, some basic concepts, such as RST, IIS, CE, and MS-IIVDS are reviewed.

2.1. Rough Sets Theory [14]

Given $IS = (U, AT, V_{AT}, f_{AT})$ be an information system(IS), where U is set of all the research objects, AT is the finite set of attributes, V_{AT} is a domain of AT, and $f_{AT} : U \times AT \rightarrow V_{AT}$ is an information function. In particular, $DS = IS \cup \{DT, V_{DT}, f_{DT}\}$ be thought as decision system(DS), where IS is an information system, DT is the set of decision attributes, V_{DT} is the domain of DT, and $f_{DT} : U \times DT \rightarrow V_{DT}$ is an information function. For any $B \subseteq AT$, an equivalence relation R_B is defined by

Table 1
The abbreviations of the terminologies.

Terminologies	Abbreviations
information system	IS
decision system	DS
rough set theory	RST
conditional entropy	CE
incomplete information system	IIS
incomplete decision system	IDS
approximation classifier precision	AP
approximation classifier quality	AQ
multi-source information system	MS-IS
incomplete interval-valued information system	IIVIS
incomplete interval-valued decision system	IIVDS
multi-source incomplete interval-valued information system	MS-IIVIS
multi-source incomplete interval-valued decision system	MS-IIVDS

$$R_B = \{(x,y) \in U \times U | f(x,a) = f(y,a), \forall a \in B\}, \tag{1}$$

where $f(x,a)$ and $f(y,a)$ denote the value of x and y under a, respectively. For any $X \subseteq U$, the lower and upper approximations of X are defined by

$$\underline{R_B}(X) = \left\{x \in U \middle| [x]_{R_B} \subseteq X\right\}, \tag{2}$$

$$\overline{R_B}(X) = \left\{x \in U \middle| [x]_{R_B} \cap X \neq \varnothing\right\}, \tag{3}$$

where $[x]_{R_B} = \{y | (x,y) \in R_B\}$. Let $U/D = \{Y_1, Y_2, \ldots, Y_m\}$ be the decision partition of U based on DT. For $DS = (U, AT \cup DT, V, f)$, the AP and AQ of U/D with respect to R_B are defined as following:

$$AP_{R_B}(U/D) = \frac{\sum_{i=1}^{m} \left|\underline{R_B}(Y_i)\right|}{\sum_{i=1}^{m} \left|\overline{R_B}(Y_i)\right|}, \tag{4}$$

$$AQ_{R_B}(U/D) = \frac{\sum_{i=1}^{m} \left|\underline{R_B}(Y_i)\right|}{|U|}. \tag{5}$$

AP and AQ were first proposed by Pawlak [33,34], and can be used to measure the approximation classification precision and quality. There higher the AP and AQ, the better the approximation classification precision and quality.

2.2. Incomplete Information System

An incomplete information system can be denoted as $IIS = (U, AT, V_{AT}, f_{AT})$,where U is a nonempty finite set of samples; AT is the nonempty finite set of condition attributes; V_{AT} is the domain of AT;$f : U \times AT \to V_{AT}$ is an information function and there exist $a \in AT$ and $x \in U$ such that $f(x,a) = *(missing\,value)$. In particular, $IDS = IIS \cup \{DT, V_{DT}, f_{DT}\}$ be thought as an incomplete decision system(IDS), where IIS is an incomplete information system, DT is the set of decision attributes, V_{DT} is the domain of DT, and $f_{DT} : U \times DT \to V_{DT}$ is an information function.

The equivalence relation is not suitable for processing the missing value, so in [35,36], a new tolerance relation is defined to deal with the IIS. For an IIS, for any $B \subseteq AT$, let $T(B)$ denote the tolerance relation, which is defined as follows:

$$T(B) = \{(x,y) | f(x,a) = f(y,a) or f(x,a) = * or f(y,a) = *, \forall a \in B\}. \tag{6}$$

The tolerance class of sample x with respect to the attribute set B is defined as follows:

$$T_B(x) = \{y | (x,y) \in T(B)\}. \tag{7}$$

And for any $B \subseteq AT$, the tolerance relation $T(B)$ is defined as follows:

$$T(B) = \cap_{a \in B} T(a). \tag{8}$$

The tolerance class of sample x_i with respect to B is denoted as $T_B(x_i) = \left\{x_j | (x_i, x_j) \in T(B)\right\}$. Furthermore, for any $X \subseteq U$, the lower and upper approximations of X are defined as follows:

$$\underline{T_B}(X) = \{x_i \in U | T_B(x_i) \subseteq X\}, \tag{9}$$

$$\overline{T_B}(X) = \{x_i \in U | T_B(x_i) \cap X \neq \varnothing\}. \tag{10}$$

2.3. Conditional Entropy

In [6], a novel conditional entropy is proposed to measure the uncertainty of IDS. Let $IDS = (U, AT \cup DT, V, f)$ denote an IDS. For any $B \subseteq AT, U/UDD = \{Y_1, Y_2, \ldots, Y_m\}$, the conditional entropy of D with respect to B is computed as follows:

$$H(D|B) = -\sum_{i=1}^{|U|}\sum_{j=1}^{m} \frac{\left|T_B(x_i) \cap Y_j\right|}{|U|} \log \frac{\left|T_B(x_i) \cap Y_j\right|}{|T_B(x_i)|}. \tag{11}$$

Furthermore, the conditional entropy $H(D|B)$ satisfies the following propositions:

(1) $0 \leqslant H(D|B) \leqslant |U| \log |U|$,
(2) $H(D|B) \leqslant H(D|C)$, if $C \subseteq B$.

2.4. *Multi-source Incomplete Interval-valued Decision System*

An incomplete interval-valued information system can be denoted as $IIVIS = (U, AT, V_{AT}, f_{AT})$,where U is a nonempty finite set of samples; AT is the nonempty finite set of condition attributes; V_{AT} is the domain of AT;$f_{AT} : U \times AT \rightarrow V_{AT}$ is an information function such that $f(x,a) = \left[f^L(x,a), f^U(x,a)\right]$ or $*$ (missing value), where $f^L(x,a)$ and $f^U(x,a)$ denote the lower and upper endpoints of the interval, respectively.

An MS-IIVIS can be denoted as

$$MS - IIVIS = \left\{IIVIS_i \middle| IIVIS_i = \left(U, AT_i, V_{AT_i}, f_{AT_i}\right), i = 1, 2, \ldots, n\right\},$$

where $IIVIS_i$ is the i-th incomplete interval-valued information subsystem.

Specially, an MS-IIVDS can be denoted as

$$MS - IIVDS = MS - IIVIS \cup \{DT, V_{DT}, f_{DT}\},$$

where $MS - IIVIS$ is the multi-source incomplete interval-valued information system, DT is the set of decision attributes, V_{DT} is the domain of DT, and $f_{DT} : U \times DT \rightarrow V_{DT}$ is an information function.

Example 2.1. To present the $MS - IIVDS$ intuitively, we give an Example 2.1 as follows. Table 2–5 are four incomplete interval-valued information tables, which denote the results of medical examinations of ten patients based on four hospitals. The attributes a_1-a_6 represent hemoglobin count, leukocyte count, blood fat, blood sugar, platelet count, and Hb level. The symbol $*$ means that a doctor can not determine the level of a project. Assume that $V_D = \{Leukemiapatient, Nonleukemiapatient\}$, and $U/D = \{Y_1, Y_2\}$, where $Y_1 = \{x_1, x_2, x_6, x_8, x_9\}$, $Y_2 = \{x_3, x_4, x_5, x_7, x_{10}\}$.

3. Information fusion based on information entropy in MS-IIVDS

In recent years, many incomplete information systems have been generated with the era of big data. So it is essential to fuse these data acquired from multiple sensors. Information fusion aims to produce a comprehensive and unified presentation of multiple information systems. Information entropy can represent the amount of information, which can be used to measure the importance of an information system. This section proposes a novel fusion method for MS-IIVDS with information entropy.

Definition 3.1. . Given an IIVIS $IIVIS = (U, AT, V_{AT}, f_{AT})$, where $U = \{x_1, x_2, \ldots, x_n\}$. $\forall x_i, x_j \in U$ and for any $a \in AT$, the distance of any two samples in U with respect to attribute a is defined as follows:

$$dis_a(x_i, x_j) = \begin{cases} 0, & if\, f(x_i, a) = * \, or \, f(x_j, a) = *, \\ \sqrt{\left(f^L(x_i,a) - f^L(x_j,a)\right)^2 + \left(f^U(x_i,a) - f^U(x_j,a)\right)^2}, & else. \end{cases} \tag{12}$$

Definition 3.2. Given an IIVIS $IIVIS = (U, AT, V_{AT}, f_{AT})$, for any $a \in AT$, the tolerance relation T_a is defined as follows:

$$T_a = \left\{ (x_i, x_j) \middle| \frac{dis_a(x_i, x_j)}{\max\limits_{y \in U}(dis_a(x_i, y))} \leqslant \alpha \;\; or \;\; f(x_i, a) = * or f(x_j, a) = * \right\}, \tag{13}$$

where α denotes the threshold. The tolerance class of x_i under a is denoted as $T_a(x_i) = \left\{x_j \middle| (x_i, x_j) \in T_a\right\}$.

Table 2
The first source of information.

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]	[176.96,124.99]	[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 second source of information.

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]	[176.99,124.99]	[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]

Table 4
The third source of information.

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]

Definition 3.3. Given a MS-IIVDS $MS-IIVDS = MS-IIVIS \cup \{DT, V_{DT}, f_{DT}\}$, where $\{I_1, I_2, \ldots, I_s\}$ denotes the set of information sources of it, $U = \{x_1, x_2, \ldots, x_n\}$ and $U/UDD = \{Y_1, Y_2, \ldots, Y_m\}$. For any $a \in AT$, the uncertainty measurement of I_q for a under D is defined as follows:

$$H^a(D|I_q) = -\sum_{i=1}^{|U|}\sum_{j=1}^{m} \frac{\left|T_a^q(x_i) \cap Y_j\right|}{|U|} \log \frac{\left|T_a^q(x_i) \cap Y_j\right|}{\left|T_a^q(x_i)\right|}, \tag{14}$$

where $T_a^q(x_i)$ denote the tolerance class of sample x_i of attribute a under I_q.

Proposition 3.1. Given a MS-IIVDS $MS-IIVDS = MS-IIVIS \cup \{DT, V_{DT}, f_{DT}\}$, the uncertainty measurement $H^a(D|I_q)$ satisfies

(1) $0 \leqslant H^a(D|I_q) \leqslant |U| \log |U|$,
(2) $H^{a_i}(D|I_{q_s}) \leqslant H^{a_j}(D|I_{q_m})$ if $T_{a_i}^{q_s}(x) \subseteq T_{a_j}^{q_m}(x)$.

Proof. (1) We first demonstrate that $H^a(D|I_q) \geqslant 0$. $\forall x_i \in U$, when $\exists Y_j$ satisfies $T_a^q(x_i) \subseteq Y_j$ and $\forall k \neq j, T_a^q(x_i) \cap Y_j = \varnothing$, then we can get $\log \frac{\left|T_a^q(x_i) \cap Y_j\right|}{\left|T_a^q(x_i)\right|} = 0$ and $\left|T_a^q(x_i) \cap Y_k\right| = 0$. In this situation, $H^a(D|I_q) = 0$. And we know $H^a(D|I_q)$ is nonnegative, so the minimum value of $H^a(D|I_q)$ is 0.

Now, let us demonstrate that $H^a(D|I_q) \leqslant |U| \log |U|$. We can get that $H^a(D|I_q) = -\sum_{i=1}^{|U|}\sum_{j=1}^{m} \frac{\left|T_a^q(x_i) \cap Y_j\right|}{|U|} \log \frac{\left|T_a^q(x_i) \cap Y_j\right|}{\left|T_a^q(x_i)\right|} = -\sum_{i=1}^{|U|}\sum_{j=1}^{m} \frac{\left|T_a^q(x_i) \cap Y_j\right|}{|U|} \log \frac{\left|T_a^q(x_i) \cap Y_j\right|}{\left|T_a^q(x_i) \cap Y_j\right| + \left|T_a^q(x_i) \cap Y_j^c\right|}$. Then $H^a(D|I_q) \leqslant -\sum_{i=1}^{|U|}\sum_{j=1}^{m} \frac{\left|T_a^q(x_i) \cap Y_j\right|}{|U|} \log \frac{\left|T_a^q(x_i) \cap Y_j\right|}{|U|}$ because $\left|T_a^q(x_i) \cap Y_j\right| + \left|T_a^q(x_i) \cap Y_j^c\right| \leqslant |U|$. And when $T_a^q(x_i) = U$, we have $-\sum_{i=1}^{|U|}\sum_{j=1}^{m} \frac{\left|T_a^q(x_i) \cap Y_j\right|}{|U|} \log \frac{\left|T_a^q(x_i) \cap Y_j\right|}{|U|} = -\sum_{i=1}^{|U|}\sum_{j=1}^{m} \frac{\left|Y_j\right|}{|U|} \log \frac{\left|Y_j\right|}{|U|} \leqslant -\sum_{i=1}^{n}\sum_{j=1}^{m} \frac{1}{|U|} \log \frac{1}{|U|} = |U| \log |U|$ because $\left|Y_j\right| \geqslant 1$. Thus the maximum value of $H^a(D|I_q)$ is $|U| \log |U|$.

(2) Given the function $f(x,y) = -x \log \frac{x}{x+y}$, we have $\frac{\partial f}{\partial x} = \log \frac{x+y}{x} - \frac{y}{x+y}$. When $x \geqslant 0$ and $y \geqslant 0$, let $t = \frac{y}{x}$, we have $\frac{\partial f}{\partial x} = \log(1+t) - \frac{t}{1+t}$. Let $g(t) = \log(1+t) - \frac{t}{1+t}$. When $t > 0$, we can get $\frac{\partial g}{\partial x} = \frac{t}{(1+t)^2} > 0$, so $g(t) > g(0) = 0$. Thus, we can get $\frac{\partial f}{\partial x} > 0$. Similarly, we can find $\frac{\partial f}{\partial y} = \frac{x}{x+y} > 0$ when $x \geqslant 0$ and $y \geqslant 0$. We all know $\left|T_{a_i}^{q_s}(x) \cap Y_j\right| \leqslant \left|T_{a_j}^{q_m}(x) \cap Y_j\right|$ and

$\left|T^{q_s}_{a_i}(x) \cap Y^c_j\right| \leqslant \left|T^{q_m}_{a_j}(x) \cap Y^c_j\right|$. Then based on the monotonicity of the function $f(x,y)$, we have $-\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{\left|T^{q_s}_{a_i}(x)\cap Y_j\right|}{|U|}\log\frac{\left|T^{q_s}_{a_i}(x)\cap Y_j\right|}{\left|T^{q_s}_{a_i}(x)\cap Y_j\right|+\left|T^{q_s}_{a_i}(x)\cap Y^c_j\right|} \leqslant -\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{\left|T^{q_m}_{a_j}(x)\cap Y_j\right|}{|U|}\log\frac{\left|T^{q_m}_{a_j}(x)\cap Y_j\right|}{\left|T^{q_m}_{a_j}(x)\cap Y_j\right|+\left|T^{q_m}_{a_j}(x)\cap Y^c_j\right|}$, which means $H^{a_i}(D|I_{q_s}) \leqslant H^{a_j}(D|I_{q_m})$.

We observe that the smaller the $H^a(D|I_q)$, the more important the I_q. Thus, we have the following Definition 3.4, which can be used to fuse MS-IIVDS.

Definition 3.4. Let $\{I_1, I_2, \ldots, I_N\}$ denotes the set of information sources of MS-IIVDS, $U = \{x_1, x_2, \ldots, x_n\}$. The i_a-th system which is the most important for attribute a, can be obtained by

$$i_a = \underset{k\in\{1,2,\ldots,N\}}{\arg\min} H^a(D|I_k). \tag{15}$$

Based on Definition 3.4, we can create a new information table through computing $H^a(D|I_q)$. Firstly, for any attribute $a \in AT$ and $I_q \in \{I_1, I_2, \ldots, I_N\}$, the value of $H^a(D|I_q)$ is computed. Secondly, for any attribute a, we find the smallest value of $H^a(D|I_q)$ and then we put the values of all samples under a with respect to I_q into the new table. Finally, a new incomplete information table is generated based on the above method. After that, we utilize the maximum fusion results of the samples in the tolerance classes of the missing value samples to complete missing values, which can be denoted as

$$\mathrm{f}(x_i, a) = \begin{cases} \left[\min\limits_{x_j\in T_a(x_i)} \mathrm{f}^L(x_j, a), \max\limits_{x_j\in T_a(x_i)} \mathrm{f}^U(x_j, a)\right], & \text{if } \mathrm{f}(x_i, a) = *, \\ \mathrm{f}(x_i, a), & \text{else}. \end{cases} \tag{16}$$

An intuitive diagram of fusion is shown in Fig. 1. The static Algorithm 3.1 is given in the following. In Algorithm 3.1, the time complexity of the step 5 is $O(|U|)$. So the time complexity of the step 1 to 15 is $O(\mathrm{N} \times |AT| \times |U| \times (|U| + |U/D|))$. Thus the time complexity of Algorithm 3.1 is $O(N \times |AT| \times |U| \times (|U| + |U/D|) + |AT|)$.

Algorithm 3.1: The static fusion algorithm of multi-source information systems based on information entropy

Data: $MS-IIVDS = \{(U, AT_i, V_{AT_i}, f_i, D, V_D, f_D), i = 1, 2, ..., N\}$; the decision partition $U/D = \{Y_1, Y_2, ..., Y_m\}$;
Result: A new fusion table

```
1  for q = 1 : N do
2      # N is the number of information sources.
3      for each a ∈ AT do
4          H^a(D|I_q) ← 0;
5          for i=1:|U| do
6              compute T_a^q(x_i);
7              for j=1:m do
8                  # m is the cardinality of |U/D|.
9                  if |T_a^q(x_i ∩ Y_j)| > 0 then
10                     H^a(D|I_q) ← H^a(D|I_q) − (|T_a^q(x_i)∩Y_j| / |U|) log (|T_a^q(x_i)∩Y_j| / |T_a^q(x_i)|)
11                 end
12             end
13         end
14     end
15 end
16 for each a ∈ AT do
17     compute i_a = arg min_{k∈{1,2,...,s}} H^a(D|I_k)
18 end
19 return (V_{a_1}^{i_{a_1}}, V_{a_2}^{i_{a_2}}, ..., V_{a_{|AT|}}^{i_{a_{|AT|}}})
```

Example 3.1 (*Continued from Example 2.1*). Based on Definition 3.3, we can calculate the conditional entropy of each source for each attribute. Take the attribute a_1 of the first source of information for example. Without loss of generality, let the value of α be 0.5. First, we can compute the distance between any two samples based on Eq. (12). For example, the distance between x_1 and x_2 under a_1 with respect to the first source is computed by $\mathrm{dis}_a(x_1,x_2)=\sqrt{(128.26-118.4)^2+(139.97-130.97)^2}=13.3499$. The distance matrix Dis is shown as follows.

$$Dis=\begin{pmatrix}0 & 13.3499 & 28.0941 & 6.77911 & 4.29961 & 75.0127 & 57.2293 & 98.8354 & 14.047 & 76.0758\\ 13.3499 & 0 & 14.7754 & 7.34939 & 9.67908 & 84.3552 & 66.151 & 110.643 & 18.9966 & 86.9971\\ 28.0941 & 14.7754 & 0 & 21.6957 & 24.1974 & 96.6293 & 78.3407 & 124.594 & 31.5759 & 100.409\\ 6.77911 & 7.34939 & 21.6957 & 0 & 2.54772 & 80.9943 & 63.0229 & 105.523 & 17.3472 & 82.548\\ 4.29961 & 9.67908 & 24.1974 & 2.54772 & 0 & 78.987 & 61.0984 & 103.127 & 16.2673 & 80.2889\\ 75.0127 & 84.3552 & 96.6293 & 80.9943 & 78.987 & 0 & 18.289 & 37.697 & 65.4661 & 14.8003\\ 57.2293 & 66.151 & 78.3407 & 63.0229 & 61.0984 & 18.289 & 0 & 51.4912 & 47.2199 & 25.4934\\ 98.8354 & 110.643 & 124.594 & 105.523 & 103.127 & 37.697 & 51.4912 & 0 & 93.2203 & 26.0058\\ 14.047 & 18.9966 & 31.5759 & 17.3472 & 16.2673 & 65.4661 & 47.2199 & 93.2203 & 0 & 68.8368\\ 76.0758 & 86.9971 & 100.409 & 82.548 & 80.2889 & 14.8003 & 25.4934 & 26.0058 & 68.8368 & 0\end{pmatrix}$$

Dividing each element of Dis by the maximum value of all the elements in the corresponding row, we can get the matrix Dis^*.

$$Dis^*=\begin{pmatrix}0 & 0.1351 & 0.2843 & 0.0686 & 0.0435 & 0.7590 & 0.5790 & 1 & 0.1421 & 0.7697\\ 0.1207 & 0 & 0.1335 & 0.06642 & 0.0875 & 0.7624 & 0.5979 & 1 & 0.1717 & 0.7863\\ 0.2255 & 0.1186 & 0 & 0.1741 & 0.1942 & 0.7756 & 0.6288 & 1 & 0.2534 & 0.8059\\ 0.0642 & 0.0696 & 0.2056 & 0 & 0.0241 & 0.7675 & 0.5972 & 1 & 0.1644 & 0.7823\\ 0.0417 & 0.0939 & 0.2346 & 0.0247 & 0 & 0.7659 & 0.5925 & 1 & 0.1577 & 0.7785\\ 0.7763 & 0.8730 & 1 & 0.8382 & 0.8174 & 0 & 0.1893 & 0.3901 & 0.6775 & 0.1532\\ 0.7305 & 0.8444 & 1 & 0.8045 & 0.7799 & 0.2335 & 0 & 0.6572 & 0.6028 & 0.3254\\ 0.7933 & 0.8880 & 1 & 0.8470 & 0.8277 & 0.3026 & 0.4132 & 0 & 0.7482 & 0.2087\\ 0.1507 & 0.2038 & 0.3387 & 0.1861 & 0.1745 & 0.7023 & 0.5065 & 1 & 0 & 0.7384\\ 0.7577 & 0.8664 & 1 & 0.8221 & 0.7997 & 0.1474 & 0.2539 & 0.2590 & 0.6856 & 0\end{pmatrix}$$

Then the tolerance classes of these ten samples are computed based on Eq. (13). For example, because 0, 0.1351, 0.2843, 0.0686, 0.0435, and 0.1421 are smaller than 0.5, so x_1,x_2,x_3,x_4,x_5, and x_9 are put in the tolerance class of x_1 under a_1.

$$\begin{aligned}&T^1_{a_1}(x_1)=\{x_1,x_2,x_3,x_4,x_5,x_9\}, T^1_{a_1}(x_2)=\{x_1,x_2,x_3,x_4,x_5,x_9\},\\&T^1_{a_1}(x_3)=\{x_1,x_2,x_3,x_4,x_5,x_9\}, T^1_{a_1}(x_4)=\{x_1,x_2,x_3,x_4,x_5,x_9\},\\&T^1_{a_1}(x_5)=\{x_1,x_2,x_3,x_4,x_5,x_9\}, T^1_{a_1}(x_6)=\{x_6,x_7,x_8,x_{10}\},\\&T^1_{a_1}(x_7)=\{x_6,x_7,x_{10}\}, T^1_{a_1}(x_8)=\{x_6,x_7,x_8,x_{10}\},\\&T^1_{a_1}(x_9)=\{x_1,x_2,x_3,x_4,x_5,x_9\}, T^1_{a_1}(x_{10})=\{x_6,x_7,x_8,x_{10}\}.\end{aligned}$$

We can then compute the conditional entropy based on Eq. (14).

$$\begin{aligned}&H^{a_1}(D|I_1)=-\left(6\times\tfrac{3}{10}\times\ln\tfrac{3}{6}+\tfrac{1}{10}\times\ln\tfrac{1}{3}+3\times\tfrac{2}{10}\times\ln\tfrac{2}{4}+6\times\tfrac{3}{10}\times\ln\tfrac{3}{6}+3\times\tfrac{2}{10}\times\ln\tfrac{2}{4}+\tfrac{2}{10}\times\ln\tfrac{2}{3}\right)\\&=3.5180607171761813\end{aligned}$$

Table 5
The fourth source of information.

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]	[177.00,125.00]	[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]	*

Similarly, we can compute the conditional entropy of other sources for attributes. The results are shown in Table 6. Based on Definition 3.4 and Eq. (16), we can get the fused complete interval-valued system, which is shown in Table 7.

4. Dynamic fusion mechanisms and algorithms with the change of information sources and attributes in MS-IIVDS

In this section, four incremental fusion mechanisms and the corresponding algorithms are proposed with the change of information sources and conditional attributes. We study four dynamic cases (a) inserting information sources and removing attributes; (b) inserting sources and attributes; (c) removing sources and attributes; (d) removing sources and inserting attributes. We illustrate the four situations of variation intuitively, in Fig. 2.

Case (a). Addition of information sources and deletion of conditional attributes.

Let $MS-IIVDS^t$ denote the multi-source incomplete interval-valued decision system at time t. In the following section, an incremental fusion mechanism is discussed when some sources are added and some conditional attributes are deleted at time $t+1$. Suppose $\{I_{N+1}, I_{N+2}, \ldots, I_{N+\Delta N}\}$ are inserted into multi-source system and $\{a_{n-\Delta n+1}, \ldots, a_n\}$ are deleted from multi-source system at time $t+1$. Then we have the following properties.

Proposition 4.1. For $\{a_1, \ldots, a_{n-\Delta n}\}$, the following properties are true:

(1) If $\min\limits_{q\in\{N+1,N+2,\ldots,N+\Delta N\}} H^{a_i}(D|I_q) \geqslant \min\limits_{q\in\{1,2,\ldots,N\}} H^{a_i}(D|I_q)$, then $V_{a_i}^{t+1} = V_{a_i}^t$;

(2) If $\min\limits_{q\in\{N+1,N+2,\ldots,N+\Delta N\}} H^{a_i}(D|I_q) < \min\limits_{q\in\{1,2,\ldots,N\}} H^{a_i}(D|I_q)$, then $V_{a_i}^{t+1} = V_{I_q}(a_i)$, where $q = \arg\min\limits_{k\in\{N+1,N+2,\ldots,N+\Delta N\}} H^{a_i}(D|I_k)$ and $V_{I_q}(a_i)$ denotes the value of a_i under I_q.

Proof. 1) If $\min\limits_{q\in\{N+1,N+2,\ldots,N+\Delta N\}} H^{a_i}(D|I_q) \geqslant \min\limits_{q\in\{1,2,\ldots,N\}} H^{a_i}(D|I_q)$, then we can get $\min\limits_{q\in\{1,2,\ldots,N,N+1,N+2,\ldots,N+\Delta N\}} H^{a}(D|I_q) = \min\limits_{q\in\{1,2,\ldots,N\}} H^{a}(D|I_q)$, so the information source which is pivotal for a is unchanged. Thus, we have $V_{a_i}^{t+1} = V_{a_i}^t$.

2) If $\min\limits_{q\in\{N+1,N+2,\ldots,N+\Delta N\}} H^{a_i}(D|I_q) < \min\limits_{q\in\{1,2,\ldots,N\}} H^{a_i}(D|I_q)$, then we can get that after $\{I_{N+1}, I_{N+2}, \ldots, I_{N+\Delta N}\}$ are inserted, $\min\limits_{q\in\{1,2,\ldots,N,N+1,N+2,\ldots,N+\Delta N\}} H^{a}(D|I_q) = \min\limits_{q\in\{N+1,N+2,\ldots,N+\Delta N\}} H^{a}(D|I_q)$, so the information source which is pivotal for a is transformed to I_q, where $q = \arg\min\limits_{k\in\{N+1,N+2,\ldots,N+\Delta N\}} H^{a_i}(D|I_k)$.

Based on Proposition 4.2, we can incrementally update the fusion table when new sources are obtained and abandon invalid conditional attributes.

Example 4.1. (*Continued from Example 2.1*). Suppose that we collect the results of medical examinations for these ten patients from only two hospitals in the beginning. Next, through the discussion of experts, it is found that the attributes 'Hb

Table 6
The conditional entropy of information sources for attributes.

U	I_1	I_2	I_3	I_4
a_1	3.518060717	5.822436317	5.143121988	5.525041931
a_2	3.591140530	3.489391305	3.832629088	4.675589474
a_3	4.418716649	3.029712081	3.947781877	4.118612056
a_4	4.729955328	4.741077973	5.013015008	3.978180671
a_5	5.721367701	4.846218014	4.001406785	4.489755422
a_6	4.577481598	3.253486559	4.039229581	4.007040281

Table 7
The result of fusion based on information entropy.

U	a_1	a_2	a_3	a_4	a_5	a_6
x_1	[128.26,139.97]	[3.00,7.00]	[115.89,124.99]	[60.80,180]	[43.39,186.96]	[70.20,92.66]
x_2	[118.40,130.97]	[4.00,11.00]	[176.99,124.99]	[79.50,119.00]	[80.38,223.95]	[67.20,88.63]
x_3	[108.53,119.97]	[2.80,10.00]	[112.39,121.99]	[119.10,180.00]	[66.28,84.98]	[78.20,99.71]
x_4	[125.11,133.97]	[2.00,9.00]	[111.49,120.99]	[60.80,99.00]	[43.39,282.93]	[66.40,88.63]
x_5	[126.69,135.97]	[4.40,11.00]	[111.99,295.99]	[81.60,120.00]	[139.37,282.93]	[43.90,74.53]
x_6	[126.29,214.96]	[6.30,16.00]	[204.29,295.99]	[81.00,163.00]	[68.28,87.98]	[28.00,61.44]
x_7	[123.04,196.96]	[4.50,9.00]	[177.29,268.99]	[84.30,154.00]	[77.98,97.98]	[30.10,60.43]
x_8	[158.85,233.95]	[10.10,20.00]	[224.09,314.98]	[102.30,163.00]	[65.28,84.98]	[25.80,56.40]
x_9	[118.40,149.97]	[12.00,22.00]	[224.99,268.99]	[90.30,151.00]	[71.28,90.98]	[78.00,88.63]
x_{10}	[141.09,214.96]	[8.40,18.00]	[177.19,267.99]	[68.00,119.00]	[109.37,252.94]	[39.90,70.50]

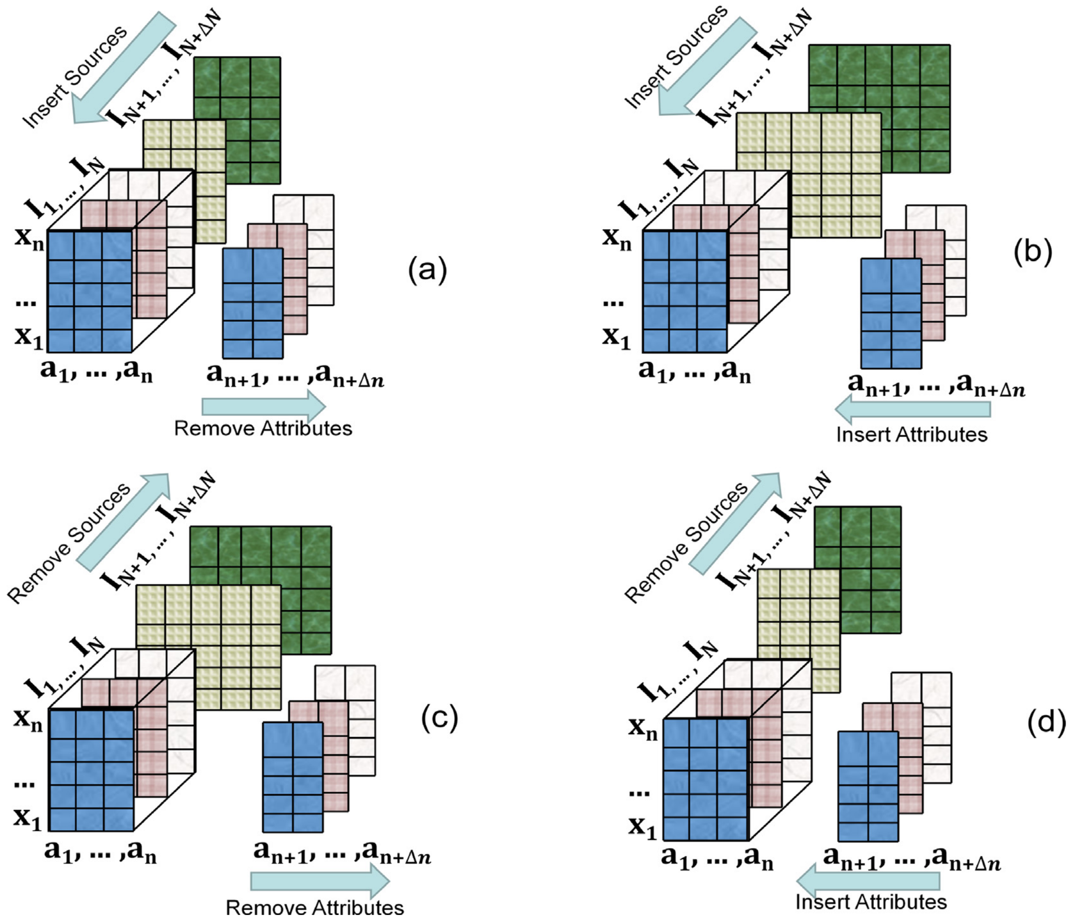

Fig. 2. Four changing cases of multi-source information system.

Table 8
The results of conditional entropy at time t.

U	I_1	I_2
a_1	3.518060717	5.822436317
a_2	3.591140530	3.489391305
a_3	4.418716649	3.029712081
a_4	4.729955328	4.741077973
a_5	5.721367701	4.846218014
a_6	4.577481598	3.253486559

Table 9
The results of conditional entropy at time $t + 1$.

U	I_1	I_2	I_3	I_4
a_1	3.518060717	5.822436317	5.143121988	5.525041931
a_2	3.591140530	3.489391305	3.832629088	4.675589474
a_3	4.418716649	3.029712081	3.947781877	4.118612056
a_4	4.729955328	4.741077973	5.013015008	3.978180671

level' and 'platelet count' are not necessary for determining whether a patient has leukemia, and these experts decide that the patients should go to more hospitals to get checked. The situation is the same as Case (a).

Let us abstract the above situation mathematically. Suppose that in time t, the MS-IIVDS has two sources I_1 and I_2, where the attributes are a_1-a_6. The results of conditional entropy at time t are shown in Table 8. Then in time $t + 1$, two sources I_3 and I_4

Table 10
The final fusion results at time $t+1$.

U	a_1	a_2	a_3	a_4
x_1	[128.26,139.97]	[3.00,7.00]	[115.89,124.99]	[60.80,180]
x_2	[118.40,130.97]	[4.00,11.00]	[176.99,124.99]	[79.50,119.00]
x_3	[108.53,119.97]	[2.80,10.00]	[112.39,121.99]	[119.10,180.00]
x_4	[125.11,133.97]	[2.00,9.00]	[111.49,120.99]	[60.80,99.00]
x_5	[126.69,135.97]	[4.40,11.00]	[111.99,295.99]	[81.60,120.00]
x_6	[126.29,214.96]	[6.30,16.00]	[204.29,295.99]	[81.00,163.00]
x_7	[123.04,196.96]	[4.50,9.00]	[177.29,268.99]	[84.30,154.00]
x_8	[158.85,233.95]	[10.10,20.00]	[224.09,314.98]	[102.30,163.00]
x_9	[118.40,149.97]	[12.00,22.00]	[224.99,268.99]	[90.30,151.00]
x_{10}	[141.09,214.96]	[8.40,18.00]	[177.19,267.99]	[68.00,119.00]

are inserted into MS-IIVDS, and a_5 and a_6 are removed from the original attributes set. The results of conditional entropy at time $t+1$ are shown in Table 9. Based on the Property 4.1, we only need to compute $H^{a_i}(D|I_j)$, where $i = 1,2,3,4$ and $j = 3,4$. For $i = 1,2,3,4$, the $\min\limits_{j\in\{1,2\}} H^{a_i}(D|I_j)$ represents the former information which does not need to be recalculated. The final fusion results at time $t+1$ are shown in Table 10.

The dynamic algorithm is given in Algorithm4.3; its time complexity of it is shown in Table 14.

Algorithm4.1: The incremental algorithm with the addition of sources and deletion of attributes

---

Data: the original fusion table $\left(V^t_{a_1}, V^t_{a_2}, ..., V^t_{a_n}, V^t_{a_{n+1}}, ..., V^t_{a_{n+\Delta n}}\right)$; the deleted attributes set $\{a_{n+1}, a_{n+2}, ..., a_{n+\Delta n}\}$; the inserted sources set $\{I_{N+1}, I_{N+2}, ..., I_{N+\Delta N}\}$; the decision partition $U/D = \{Y_1, Y_2, ..., Y_m\}$;

Result: An updated fusion table.

for $q = N+1 : N + \Delta N$ **do**
    for *each* $a \in \{a_1, a_2, ..., a_n\}$ **do**
        compute $H^a\left(D\left|I_q\right.\right)$; # n is the number of remained attributes.
    end
end
for *each* $a \in \{a_1, a_2, ..., a_n\}$ **do**
    if $\min\limits_{q\in\{N+1,N+2,...,N+\Delta N\}} H^a\left(D\left|I_q\right.\right) \geq \min\limits_{q\in\{1,2,...,N\}} H^a\left(D\left|I_q\right.\right)$ **then**
        $V^{t+1}_a = V^t_a$
    end
    if $\min\limits_{q\in\{N+1,N+2,...,N+\Delta N\}} H^a\left(D\left|I_q\right.\right) < \min\limits_{q\in\{1,2,...,N\}} H^a\left(D\left|I_q\right.\right)$ **then**
        $V^{t+1}_a = V_{I_q}(a), q = \mathop{\arg\min}\limits_{k\{N+1,N+2,...,N+\Delta N\}} H^a\left(D\left|I_k\right.\right)$ #$V_{I_q}(a)$ denotes the value of attribute a under the information source I_q
    end
end
return $\left(V^{t+1}_{a_1}, V^{t+1}_{a_2}, ..., V^{t+1}_{a_n}\right)$

Case (b). Addition of information sources and addition of conditional attributes.

Suppose $\{I_{N+1}, I_{N+2}, \ldots, I_{N+\Delta N}\}$ are inserted into multi-source system and $\{a_{n+1}, \ldots, a_{n+\Delta n}\}$ are inserted into multi-source system at time $t+1$. Then we have the following propositions.

Proposition 4.2. For $\{a_1, a_2, \ldots, a_n\}$, the following propositions are true:

(1) If $\min\limits_{q\in\{N+1,N+2,\ldots,N+\Delta N\}} H^{a_i}(D|I_q) \geqslant \min\limits_{q\in\{1,2,\ldots,N\}} H^{a_i}(D|I_q)$, then $V^{t+1}_{a_i} = V^t_{a_i}$;

(2) If $\min\limits_{q\in\{N+1,N+2,\ldots,N+\Delta N\}} H^{a_i}(D|I_q) < \min\limits_{q\in\{1,2,\ldots,N\}} H^{a_i}(D|I_q)$, then $V^{t+1}_{a_i} = V_{I_q}(a_i)$, where $q = \mathop{\arg\min}\limits_{k\in\{N+1,N+2,\ldots,N+\Delta N\}} H^{a_i}(D|I_k)$ and $V_{I_q}(a_i)$ denotes the value of a_i under I_q.

Proof. Proposition 4.3 is similar to Proposition 4.2, so its proof is similar to that of Proposition 4.2.

Proposition 4.3. For $\{a_{n+1}, a_{n+2}, \ldots, a_{n+\Delta n}\}$, we have $V_{a_i} = V_{I_q}(a_i)$, where $q = \underset{k\in\{1,2,\ldots,N+\Delta N\}}{\arg\min} H^{a_i}(D|I_k)$ and $V_{I_q}(a_i)$ denotes the the value of a_i under I_q.

Proof. It is easy to demonstrate Proposition 4.5 based on Definition 3.4.

Example 4.2 (*Continued from Example 2.1*). Suppose that in the beginning, we collect the results of medical examinations based on hemoglobin count, leukocyte count, blood fat, and blood sugar for these ten patients from only two hospitals. Next, the experts decide that Hb level and platelet count are necessary for determining whether a patient has leukemia, and direct that the patients go to more hospitals for checkup. The situation is the same as Case (b).

Let us abstract the above situation mathematically. Suppose that in time t, the MS-IIVDS has two sources I_1 and I_2, where the attributes are a_1-a_4. The results of entropy at time t are shown in Table 11. Then in time $t+1$, two sources I_3 and I_4 are inserted into $MS-IIVDS$, and a_5 and a_6 are inserted into the original attributes set. The results of entropy at time $t+1$ are the same as Table 6. Based on the Propositions 4.3 and 4.5, we only need to compute $H^{a_i}(D|I_j)$, where $i = 1,2,3,4$ and $j = 3,4$ and $H^{a_i}(D|I_j)$,where $i = 5,6$ and $j = 1,2,3,4$. For $i = 1,2,3,4$, the $\min_{j\in\{1,2\}} H^{a_i}(D|I_j)$ represents the former information which doed not need to be recalculated. The final fusion results at time $t+1$ are the same as Table 7.

The dynamic algorithm is given in Algorithm 4.2; its time complexity is shown in Table 14.

Algorithm 4.2: The incremental algorithm with the addition of sources and addition of attributes

---

Data: the original fusion table $\left(V^t_{a_1}, V^t_{a_2}, \ldots, V^t_{a_n}\right)$; the inserted sources set $\{I_{N+1}, I_{N+2}, \ldots, I_{N+\Delta N}\}$; the inserted attributes set $\{a_{n+1}, a_{n+2}, \ldots, a_{n+\Delta n}\}$; the decision partition $U/D = \{Y_1, Y_2, \ldots, Y_m\}$;
Result: An updated fusion table

for $q = N{+}1 : N + \Delta N$ **do**
  for *each* $a \in \{a_1, a_2, \ldots, a_n, .., a_{n+\Delta n}\}$ **do**
    compute $H^a\left(D\middle|I_q\right)$; # n is the number of original attributes.
  end
end
for $q = 1 : N$ **do**
  for *each* $a \in \{a_{n+1}, a_{n+2}, \ldots, a_{n+\Delta n}\}$ **do**
    compute $H^a\left(D\middle|I_q\right)$
  end
end
for *each* $a \in \{a_1, a_2, \ldots, a_n\}$ **do**
  if $\min_{q\in\{N+1,N+2,\ldots,N+\Delta N\}} H^a\left(D\middle|I_q\right) \geq \min_{q\in\{1,2,\ldots,N\}} H^a\left(D\middle|I_q\right)$ **then**
    $V^{t+1}_a = V^t_a$
  end
  if $\min_{q\in\{N+1,N+2,\ldots,N+\Delta N\}} H^a\left(D\middle|I_q\right) < \min_{q\in\{1,2,\ldots,N\}} H^a\left(D\middle|I_q\right)$ **then**
    $V^{t+1}_a = V_{I_q}(a), q = \underset{k\{N+1,N+2,\ldots,N+\Delta N\}}{\arg\min} H^a(D|I_k)$ #$V_{I_q}(a)$ denotes the value of attribute a under the information source I_q
  end
end
for *each* $a \in \{a_{n+1}, a_{n+2}, \ldots, a_{n+\Delta n}\}$ **do**
  compute $i_a = \underset{k\in\{1,2,\ldots,N+\Delta N\}}{\arg\min} H^a(D|I_k)$
  Let $V_a = V_{I_{i_a}}(a)$
end
return $\left(V^{t+1}_{a_1}, V^{t+1}_{a_2}, \ldots, V^{t+1}_{a_n}, V_{a_{n+1}}, \ldots, V_{a+\Delta n}\right)$.

Table 11
The results of conditional entropy at time t.

U	I_1	I_2
a_1	3.518060717	5.822436317
a_2	3.591140530	3.489391305
a_3	4.418716649	3.029712081
a_4	4.729955328	4.741077973

Case (c). Deletion of information sources and deletion of conditional attributes. Suppose $\{I_{N-\Delta N+1}, I_{N-\Delta N+2}, \ldots, I_N\}$ and $\{a_{n-\Delta n+1}, a_{n-\Delta n+2}, \ldots, a_n\}$ are removed from $\{I_1, I_2, \ldots, I_N, I_{N+1}, \ldots, I_{N+\Delta N}\}$ at time $t+1$. Then we have Proposition 4.6.

Proposition 4.4. labelspsproposition5 For $\{a_1, \ldots, a_{n-\Delta n}\}$, the following propositions are true:

(1) If $\min\limits_{q\in\{N-\Delta N+1, N-\Delta N+2,\ldots,N\}} H^{a_i}(D|I_q) \geqslant \min\limits_{q\in\{1,2,\ldots,N-\Delta N\}} H^{a_i}(D|I_q)$, then $V_{a_i}^{t+1} = V_{a_i}^{t}$;

(2) If $\min\limits_{q\in\{N-\Delta N+1, N-\Delta N+2,\ldots,N\}} H^{a_i}(D|I_q) < \min\limits_{q\in\{1,2,\ldots,N-\Delta N\}} H^{a_i}(D|I_q)$, then $V_{a_i}^{t+1} = V_{I_q}(a_i)$, where $q = \arg\min\limits_{k\in\{1,2,\ldots,N-\Delta N\}} H^{a_i}(D|I_k)$ and $V_{I_q}(a_i)$ denotes the value of a_i under I_q.

Proof. (1) If $\min\limits_{q\in\{N-\Delta N+1, N-\Delta N+2,\ldots,N\}} H^{a_i}(D|I_q) \geqslant \min\limits_{q\in\{1,2,\ldots,N-\Delta N\}} H^{a_i}(D|I_q)$, then we can get

$$\min_{q\in\{1,2,\ldots,N\}} H^{a}(D|I_q) = \min_{q\in\{1,2,\ldots,N-\Delta N\}} H^{a}(D|I_q),$$

so the information source which is pivotal for a is unchanged. Thus we have $V_{a_i}^{t+1} = V_{a_i}^{t}$.

(2) If $\min\limits_{q\in\{N-\Delta N+1, N-\Delta N+2,\ldots,N\}} H^{a_i}(D|I_q) < \min\limits_{q\in\{1,2,\ldots,N-\Delta N\}} H^{a_i}(D|I_q)$, then we can get that

$$\min_{q\in\{1,2,\ldots,N\}} H^{a}(D|I_q) = \min_{q\in\{N-\Delta N+1, N-\Delta N+2,\ldots,N\}} H^{a}(D|I_q).$$

Thus after $\{I_{N-\Delta N+1}, I_{N-\Delta N+2}, \ldots, I_N\}$ are removed, the information source which is pivotal for a is changed to I_q, where $q = \arg\min\limits_{k\in\{1,2,\ldots,N-\Delta N\}} H^{a_i}(D|I_k)$.

Example 4.3 (*Continued from Example 2.1*). Suppose that in the beginning, we collect the results of hemoglobin count, leukocyte count, blood fat, blood sugar, Hb level, and platelet count for these ten patients from four hospitals. Next, the experts decide that Hb level and platelet count are not necessary for determining whether a patient has leukemia. The experts note that there was equipment failure during medical examinations in two hospitals; therefore, the results of the two hospitals must be removed. The situation is the same as Case (c).

Let us abstract the above situation mathematically. Suppose that in time t, the MS-IIVDS has four sources I_1- I_4, where the attributes are a_1-a_6. The results of entropy at time t are the same as Table 6. Then in time $t+1$, two sources I_3 and I_4 are removed from MS-IIVDS, and, a_5 and a_6 are removed from the original attributes set. The results of entropy at time $t+1$ are shown in Table 11. Based on the Proposition 4.3, for the remaining attributes, we only need to use the former information $\min\limits_{q\in\{3,4\}} H^{a}(D|I_q)$ and $\min\limits_{q\in\{1,2\}} H^{a}(D|I_q)$ to update the fusion information table without any recalculation. The final fusion results at time $t+1$ are shown in Table 12.

The dynamic algorithm is given in Algorithm4.3; its complexity is shown in Table 14.

Table 12
The final fusion results at time $t+1$.

U	a_1	a_2	a_3	a_4
x_1	[128.26,139.97]	[3.00,7.00]	[115.89,124.99]	[60.78,179.97]
x_2	[118.40,130.97]	[4.00,11.00]	[176.99,124.99]	[79.48,118.98]
x_3	[108.53,119.97]	[2.80,10.00]	[112.39,121.99]	[119.07,179.97]
x_4	[125.11,133.97]	[2.00,9.00]	[111.49,120.99]	[60.78,98.98]
x_5	[126.69,135.97]	[4.40,11.00]	[111.99,295.99]	[81.58,119.98]
x_6	[126.29,214.96]	[6.30,16.00]	[204.29,295.99]	[60.78,179.97]
x_7	[123.04,196.96]	[4.50,9.00]	[177.29,268.99]	[84.28,153.97]
x_8	[158.85,233.95]	[10.10,20.00]	[224.09,314.98]	[102.28,162.97]
x_9	[118.40,149.97]	[12.00,22.00]	[224.99,268.99]	[90.28,150.97]
x_{10}	[141.09,214.96]	[8.40,18.00]	[177.19,267.99]	[67.98,118.98]

Algorithm 4.3: The incremental algorithm with the deletion of sources and attributes.

Data: the original fusion table $\left(V_{a_1}^t, V_{a_2}^t, ..., V_{a_n}^t, V_{a_{n+1}}^t, ..., V_{a_{n+\Delta n}}^t\right)$; the deleted attributes set $\{a_{n+1}, a_{n+2}, ..., a_{n+\Delta n}\}$; the deleted sources set $\{I_{N+1}, I_{N+2}, ..., I_{N+\Delta N}\}$; the decision partition $U/D = \{Y_1, Y_2, ..., Y_m\}$;
Result: An updated fusion table.

```
for each a ∈ {a_1, a_2, ..., a_n} do
    # n is the number of remained attributes.
    if min_{q∈{N+1,N+2,...,N+ΔN}} H^a(D|I_q) ≥ min_{q∈{1,2,...,N}} H^a(D|I_q) then
        V_a^{t+1} = V_a^t
    end
    if min_{q∈{N+1,N+2,...,N+ΔN}} H^a(D|I_q) < min_{q∈{1,2,...,N}} H^a(D|I_q) then
        V_a^{t+1} = V_{I_q}(a), q = arg min_{k{1,2,...,N}} H^a(D|I_k)  #V_{I_q}(a) denotes the value of attribute a under the information
          source I_q
    end
end
return (V_{a_1}^{t+1}, V_{a_2}^{t+1}, ..., V_{a_n}^{t+1}).
```

Case (d). Deletion of information sources and addition of conditional attributes. Suppose $\{I_{N-\Delta N+1}, I_{N-\Delta N+2}, \ldots, I_N\}$ are removed from $\{I_1, I_2, \ldots, I_N\}$ and $\{a_{n+1}, a_{n+2}, \ldots, a_{n+\Delta n}\}$ are inserted at time $t+1$. Then we have the following propositions.

Proposition 4.5. For $\{a_1, a_2, \ldots, a_n\}$, the following propositions are true:

(1) If $\min\limits_{q\in\{N-\Delta N+1, N-\Delta N+2,\ldots,N\}} H^{a_i}(D|I_q) \geqslant \min\limits_{q\in\{1,2,\ldots,N-\Delta N\}} H^{a_i}(D|I_q)$, then $V_{a_i}^{t+1} = V_{a_i}^t$;

(2) If $\min\limits_{q\in\{N-\Delta N+1, N-\Delta N+2,\ldots,N\}} H^{a_i}(D|I_q) < \min\limits_{q\in\{1,2,\ldots,N-\Delta N\}} H^{a_i}(D|I_q)$, then $V_{a_i}^{t+1} = V_{I_q}(a_i)$, where $q = \arg\min\limits_{k\in\{1,2,\ldots,N-\Delta N\}} H^{a_i}(D|I_k)$ and $V_{I_q}(a_i)$ denotes the value of a_i under I_q.

Proof. Proposition 4.2 is similar to Proposition 4.6, so its proof is similar to that of Proposition 4.6.

Proposition 4.6. For $\{a_{n+1}, a_{n+2}, \ldots, a_{n+\Delta n}\}$, we have $V_{a_i} = V_{I_q}(a_i)$, where $q = \arg\min\limits_{k\in\{1,2,\ldots,N-\Delta N\}} H^{a_i}(D|I_k)$ and $V_{I_q}(a_i)$ denotes the value of a_i under I_q.

Proof. It is easy to confirm based on the Definition 3.4.

Example 4.4 (*Continued from Example 2.1*). Suppose that in the beginning, we collect the results of hemoglobin count, leukocyte count, blood fat, and blood sugar for these ten patients from four hospitals. Next, the experts decide that Hb level and platelet count are necessary for determining whether a patient has leukemia. The experts further discover that the relevant equipment in two hospitals failed during medical examination, therefore, the results from the two hospitals must be removed. The situation is the same as Case (d).

Table 13
The final fusion results at time $t+1$.

U	a_1	a_2	a_3	a_4	a_5	a_6
x_1	[128.26,139.97]	[3.00,7.00]	[115.89,124.99]	[60.78,179.97]	[43.40,186.99]	[70.20,92.66]
x_2	[118.40,130.97]	[4.00,11.00]	[176.99,124.99]	[79.48,118.98]	[80.40,223.99]	[67.20,88.63]
x_3	[108.53,119.97]	[2.80,10.00]	[112.39,121.99]	[119.07,179.97]	[43.40,259.99]	[78.20,99.71]
x_4	[125.11,133.97]	[2.00,9.00]	[111.49,120.99]	[60.78,98.98]	[99.99,259.99]	[66.40,88.63]
x_5	[126.69,135.97]	[4.40,11.00]	[111.99,295.99]	[81.58,119.98]	[43.40,259.99]	[43.90,74.53]
x_6	[126.29,214.96]	[6.30,16.00]	[204.29,295.99]	[60.78,179.97]	[68.30,88.00]	[28.00,61.44]
x_7	[123.04,196.96]	[4.50,9.00]	[177.29,268.99]	[84.28,153.97]	[78.00,98.00]	[30.10,60.43]
x_8	[158.85,233.95]	[10.10,20.00]	[224.09,314.98]	[102.28,162.97]	[43.40,259.99]	[25.80,56.40]
x_9	[118.40,149.97]	[12.00,22.00]	[224.99,268.99]	[90.28,150.97]	[71.30,91.00]	[78.00,88.63]
x_{10}	[141.09,214.96]	[8.40,18.00]	[177.19,267.99]	[67.98,118.98]	[109.39,252.99]	[39.90,70.50]

Table 14
The comparison of the time complexity between the static and dynamic algorithm.

Cases	Static algorithm	Dynamic algorithm
(a)	$O((N+\Delta N)\times n\times \lvert U\rvert\times(\lvert U\rvert+m)+n)$	$O(\Delta N\times n\times \lvert U\rvert\times(\lvert U\rvert+m)+n)$
(b)	$O(((N+\Delta N)\times(n+\Delta n))\times \lvert U\rvert\times(\lvert U\rvert+m)+n+\Delta n)$	$O((N\times\Delta n+\Delta N\times(n+\Delta n))\times \lvert U\rvert\times(\lvert U\rvert+m)+n+\Delta n)$
(c)	$O(N\times n\times \lvert U\rvert\times(\lvert U\rvert+m)+n)$	$O(n)$
(d)	$O(N\times(n+\Delta n)\times \lvert U\rvert\times(\lvert U\rvert+m)+n+\Delta n)$	$O(N\times\Delta n\times \lvert U\rvert\times(\lvert U\rvert+m)+n+\Delta n)$

Let us abstract the above situation mathematically. Suppose that in time t, the MS-IIVDS has four sources I_1- I_4, where the attributes are a_1-a_4. The results of entropy at time t are the same as Table 9. Then in time $t+1$, two sources I_3 and I_4 are removed from MS-IIVDS, and a_5 and a_6 are inserted into the original attributes set. The results of entropy at time $t+1$ are the same as Table 8. Based on Propositions 4.6 and 4.2, we only need to compute $H^{a_i}(D|I_j)$, where $i=5,6$ and $j=1,2$. For $i=1,2,3,4$, the $\min\limits_{j\in\{1,2\}} H^{a_i}(D|I_j)$ and $\min\limits_{j\in\{3,4\}} H^{a_i}(D|I_j)$ represents the former information which does not need to be recalculated. The final fusion results are shown in Table 13.

The dynamic algorithm is given in Algorithm 4.4; its time complexity is shown in Table 14.

Algorithm 4.4: The incremental algorithm with the deletion of sources and addition of attributes

---

Data: the original fusion table $\left(V_{a_1}^t, V_{a_2}^t, \ldots, V_{a_n}^t\right)$; the deleted sources set $\{I_{N+1}, I_{N+2}, \ldots, I_{N+\Delta N}\}$; the inserted attributes set $\{a_{n+1}, a_{n+2}, \ldots, a_{n+\Delta n}\}$; the decision partition $U/D=\{Y_1, Y_2, \ldots, Y_m\}$;
Result: An updated fusion table.

for $q = 1 : N$ **do**
  for *each* $a \in \{a_{n+1}, a_{n+2}, \ldots, a_{n+\Delta n}\}$ **do**
    compute $H^a\left(D\middle|I_q\right)$;
  end
end
for *each* $a \in \{a_1, a_2, \ldots, a_n\}$ **do**
  if $\min\limits_{q\in\{N+1,N+2,\ldots,N+\Delta N\}} H^a\left(D\middle|I_q\right) \geq \min\limits_{q\in\{1,2,\ldots,N\}} H^a\left(D\middle|I_q\right)$ **then**
    $V_a^{t+1} = V_a^t$
  end
  if $\min\limits_{q\in\{N+1,N+2,\ldots,N+\Delta N\}} H^a\left(D\middle|I_q\right) < \min\limits_{q\in\{1,2,\ldots,N\}} H^a\left(D\middle|I_q\right)$ **then**
    $V_a^{t+1} = V_{I_q}(a), q = \arg\min\limits_{k\{1,2,\ldots,N\}} H^a(D|I_k)$ #$V_{I_q}(a)$ denotes the value of attribute a under the information source I_q
  end
end
for *each* $a \in \{a_{n+1}, a_{n+2}, \ldots, a_{n+\Delta n}\}$ **do**
  compute $i_a = \arg\min\limits_{k\in\{1,2,\ldots,N\}} H^a(D|I_k)$
  Let $V_a = V_{I_{i_a}}(a)$
end
return $\left(V_{a_1}^{t+1}, V_{a_2}^{t+1}, \ldots, V_{a_n}^{t+1}, V_{a_{n+1}}, \ldots, V_{a+\Delta n}\right)$.

Furthermore, the comparison of the time complexity between dynamic algorithms and the static fusion Algorithm 3.1 is given in Table 14. From the perspective of time complexity, we observe that the four dynamic fusion algorithms can reduce the runtime effectively.

5. Experimental analysis

This section reports on experiments conducted to illustrate the effectiveness of our fusion method and the efficiency of the incremental mechanisms based on nine data sets from UCI [37], shown in Table 15. All experiments were performed using a personal computer, and the operating environment is summarized in Table 16. It is well known that multi-source incomplete interval-valued data sets are not directly obtained from any public databases; therefore, we used the method

Table 15
The description of data sets.

No.	Data sets	Abbreviation	Samples	Attributes	Classes
1	Wine	Wine	178	13	3
2	Connectionist Bench (Sonar, Mines vs. Rocks)	CB	208	60	2
3	Speaker Accent Recognition	SAR	329	12	6
4	Ecoli	Ecoli	336	8	8
5	Breast Cancer Wisconsin(Diagnostic)	Breast	569	30	2
6	Hill-Valley	HV	606	101	2
7	South German Credit	SGC	1000	21	2
8	Wall-Following Robot Navigation Data	Wall	5456	24	4
9	Shill Bidding Dataset	Shill	6321	13	2

Table 16
Operating Ambient.

Name	Model	Parameter
CPU	Intel(R) Core(TM) i5-6300HQ	2.30 GHz
Platform	Python	3.7
System	Windows7	64bit
Memory	DDR3	8 GB;1600Mhz
Hard Disk	HTS545050A7E680	500 GB

in [31] to generate a multi-source interval-valued data set. The first step is to convert the original single-valued data into interval-valued data by $f^L(x,a) = v(x,a) - 2\sigma_a$ and $f^U(x,a) = v(x,a) + 2\sigma_a$, where $v(x,a)$ denotes the value of x under a and σ_a denotes the standard deviation of the a with the same class of x. The second step is to generate m random numbers $\{r_1, r_2, \ldots, r_m\}$ which satisfy the normal distribution $N(0, 0.1)$. And then if $r_i > 0$, let $f_i^L(x,a) = f^L(x,a)(1-r_i)$ and $f_i^U(x,a) = f^U(x,a)(1+r_i)$. Otherwise, let $f_i^L(x,a) = f^L(x,a)(1+r_i)$ and $f_i^U(x,a) = f^U(x,a)(1-r_i)$. Let $f_i(x,a) = \left[f_i^L(x,a), f_i^U(x,a)\right]$ be the value of x under a in i-th subsystem. Thus, m subsystems are generated. And the missing values are created by randomly deleting 10% data. In the following experiments, for each data set, we generated 20 sources to constitute an MS-IIVDS.

5.1. The analysis of fusion effectiveness

As in [31], AP and AQ reflect the precision and quality of approximation classification, respectively, confirming that they can be used as fusion performance metrics. We compare the fusion effectiveness between the proposed fusion method and three common fusion methods based on AP, AQ. The three common fusion methods are

(1) $MaxF_a(x) = \left[\min_{i\in\{1,2,\ldots N\}} f_i^L(x,a), \max_{i\in\{1,2,\ldots N\}} f_i^U(x,a)\right]$,

(2) $MinF_a(x) = \left[\max_{i\in\{1,2,\ldots N\}} f_i^L(x,a), \min_{i\in\{1,2,\ldots N\}} f_i^U(x,a)\right]$,

(3) $MeanF_a(x) = \left[\frac{1}{N}\sum_{i=1}^{N} f_i^L(x,a), \frac{1}{N}\sum_{i=1}^{N} f_i^U(x,a)\right]$,

where $f_i^L(x,a)$ and $f_i^U(x,a)$ denote the left and right endpoints of $i-th$ information source with respect to attribute a.

Table 17
AP and AQ of Wine with the variation of α.

	AP				AQ			
alpha	CeF	MaxF	MinF	MeanF	CeF	MaxF	MinF	MeanF
0.5	**0.4150**	0.3716	0.3370	0.3769	**0.5899**	0.5449	0.5169	0.5506
0.45	**0.6792**	0.5822	0.5683	0.5964	**0.8090**	0.7360	0.7247	0.7472
0.4	**0.8639**	0.7980	0.8163	0.7889	**0.9270**	0.8876	0.8989	0.8820
0.35	**0.9669**	0.9560	**0.9669**	**0.9669**	**0.9831**	0.9775	**0.9831**	**0.9831**
0.3	**1.0000**	0.9888	**1.0000**	**1.0000**	**1.0000**	0.9944	**1.0000**	**1.0000**
0.25	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
0.2	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
0.15	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
0.1	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
0.05	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**

Table 18
AP and AQ of CB with the variation of α

	AP				AQ			
α	CeF	MaxF	MinF	MeanF	CeF	MaxF	MinF	MeanF
0.5	0.7049	**0.7261**	0.6062	0.6574	0.8269	**0.8413**	0.7548	0.7933
0.45	0.8009	**0.8326**	0.7119	0.7702	0.8894	**0.9087**	0.8317	0.8702
0.4	0.8909	**0.8995**	0.8326	0.8489	0.9423	**0.9471**	0.9087	0.9183
0.35	0.9531	**0.9716**	0.8909	0.9531	0.9760	**0.9856**	0.9423	0.9760
0.3	**1.0000**	0.9904	0.9716	0.9904	**1.0000**	0.9952	0.9856	0.9952
0.25	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
0.2	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
0.15	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
0.1	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
0.05	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**

Table 19
AP and AQ of SAR with the variation of α.

	AP				AQ			
α	CeF	MaxF	MinF	MeanF	CeF	MaxF	MinF	MeanF
0.5	**0.0996**	0.0677	0.0810	0.0471	**0.3131**	0.2462	0.2736	0.1793
0.45	**0.1899**	0.1421	0.1851	0.1242	**0.4681**	0.4073	0.4742	0.3647
0.4	**0.2612**	0.2194	0.2493	0.2243	**0.5471**	0.5076	0.5471	0.5106
0.35	**0.3389**	0.3008	0.2912	0.2989	**0.6140**	0.5714	0.5805	0.5805
0.3	**0.4387**	0.4135	0.3842	0.4034	**0.6748**	0.6535	0.6505	0.6474
0.25	**0.6190**	0.5856	0.5316	0.5921	**0.7903**	0.7690	0.7416	0.7720
0.2	**0.8563**	0.8487	0.7919	0.8408	**0.9240**	0.9210	0.8906	0.9149
0.15	**0.9819**	**0.9819**	0.9701	**0.9819**	**0.9909**	**0.9909**	0.9848	**0.9909**
0.1	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
0.05	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**

Table 20
AP and AQ of Ecoli with the variation of α.

	AP				AQ			
α	CeF	MaxF	MinF	MeanF	CeF	MaxF	MinF	MeanF
0.5	**0.0325**	0.0026	0.0085	0.0048	**0.0893**	0.0089	0.0238	0.0149
0.45	**0.0442**	0.0061	0.0224	0.0088	**0.1042**	0.0179	0.0536	0.0238
0.4	**0.2047**	0.0681	0.1875	0.0873	**0.3601**	0.1577	0.3304	0.1815
0.35	0.3926	0.1848	**0.4386**	0.3137	0.5714	0.3333	**0.6161**	0.4911
0.3	0.6271	0.4483	**0.7731**	0.5520	0.7708	0.6190	**0.8720**	0.7113
0.25	0.9037	0.6430	**0.9941**	0.7827	0.9494	0.7968	**0.9970**	0.8869
0.2	**1.0000**	0.7968	**1.0000**	0.9941	**1.0000**	0.8869	**1.0000**	0.9970
0.15	**1.0000**	0.9649	**1.0000**	**1.0000**	**1.0000**	0.9821	**1.0000**	**1.0000**
0.1	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
0.05	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**

Table 21
AP and AQ of Breast with the variation of α.

	AP				AQ			
α	CeF	MaxF	MinF	MeanF	CeF	MaxF	MinF	MeanF
0.5	**0.0252**	0.0107	0.0143	0.0116	**0.0492**	0.0211	0.0281	0.0228
0.45	**0.0431**	0.0197	0.0215	0.0206	**0.0826**	0.0387	0.0422	0.0404
0.4	**0.0636**	0.0289	0.0336	0.0317	**0.1195**	0.0562	0.0650	0.0615
0.35	**0.0859**	0.0498	0.0557	0.0498	**0.1582**	0.0949	0.1054	0.0949
0.3	**0.1223**	0.0911	0.0974	0.0921	**0.2179**	0.1670	0.1775	0.1687
0.25	**0.2004**	0.1732	0.1805	0.1793	**0.3339**	0.2953	0.3058	0.3040
0.2	**0.3694**	0.3202	0.3357	0.3357	**0.5395**	0.4851	0.5026	0.5026
0.15	**0.6398**	0.6051	0.5983	0.6028	**0.7803**	0.7540	0.7487	0.7522
0.1	**0.9621**	**0.9621**	0.9420	0.9520	**0.9807**	**0.9807**	0.9701	0.9754
0.05	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**

Table 22
AP and AQ of HV with the variation of α.

	AP				AQ			
α	CeF	MaxF	MinF	MeanF	CeF	MaxF	MinF	MeanF
0.5	**0.0075**	0.0000	0.0000	0.0000	**0.0149**	0.0000	0.0000	0.0000
0.45	**0.0100**	0.0000	0.0008	0.0000	**0.0198**	0.0000	0.0017	0.0000
0.4	**0.0066**	0.0008	0.0008	0.0008	**0.0132**	0.0017	0.0017	0.0017
0.35	**0.0092**	0.0008	0.0008	0.0008	**0.0182**	0.0017	0.0017	0.0017
0.3	**0.0125**	0.0017	0.0033	0.0025	**0.0248**	0.0033	0.0066	0.0050
0.25	**0.0151**	0.0058	0.0100	0.0066	**0.0297**	0.0116	0.0198	0.0132
0.2	**0.0211**	0.0159	0.0202	0.0159	**0.0413**	0.0314	0.0396	0.0314
0.15	**0.0306**	0.0280	**0.0306**	0.0289	**0.0594**	0.0545	**0.0594**	0.0561
0.1	**0.0448**	0.0412	0.0368	0.0377	**0.0858**	0.0792	0.0710	0.0726
0.05	**0.0812**	0.0793	0.0678	0.0754	**0.1502**	0.1469	0.1271	0.1403

Table 23
AP and AQ of SGC with the variation of α.

	AP				AQ			
α	CeF	MaxF	MinF	MeanF	CeF	MaxF	MinF	MeanF
0.5	0.6849	0.5662	**0.6949**	0.4482	0.8130	0.7230	**0.8200**	0.6190
0.45	**0.9512**	0.9048	0.9212	0.9399	**0.9750**	0.9500	0.9590	0.9690
0.4	**0.9646**	0.9493	0.9531	0.9531	**0.9820**	0.9740	0.9760	0.9760
0.35	0.9685	0.9724	**0.9763**	0.9589	0.9840	0.9860	**0.9880**	0.9790
0.3	**0.9960**	0.9920	0.9900	0.9940	**0.9980**	0.9960	0.9950	0.9970
0.25	0.9960	**0.9980**	0.9940	0.9940	0.9980	**0.9990**	0.9970	0.9970
0.2	0.9980	**1.0000**	0.9980	0.9980	0.9990	**1.0000**	0.9990	0.9990
0.15	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
0.1	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
0.05	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**

Table 24
AP and AQ of Wall with the variation of α.

	AP				AQ			
α	CeF	MaxF	MinF	MeanF	CeF	MaxF	MinF	MeanF
0.5	**0.1414**	0.0720	0.1101	0.0871	**0.3185**	0.1880	0.2626	0.2192
0.45	**0.1858**	0.1150	0.1419	0.1357	**0.3842**	0.2674	0.3127	0.3037
0.4	**0.2449**	0.1712	0.1899	0.1768	**0.4606**	0.3572	0.3831	0.3640
0.35	**0.2920**	0.2393	0.2271	0.2408	**0.5148**	0.4489	0.4322	0.4511
0.3	**0.3511**	0.2922	0.2826	0.2907	**0.5753**	0.5119	0.4984	0.5092
0.25	**0.4086**	0.3591	0.3411	0.3563	**0.6276**	0.5795	0.5629	0.5761
0.2	**0.5008**	0.4459	0.4256	0.4475	**0.7025**	0.6562	0.6380	0.6565
0.15	**0.5979**	0.5517	0.5232	0.5514	**0.7729**	0.7392	0.7170	0.7383
0.1	**0.6945**	0.6616	0.6405	0.6523	**0.8363**	0.8160	0.8024	0.8098
0.05	**0.8023**	0.7773	0.7592	0.7750	**0.9008**	0.8860	0.8759	0.8845

Table 25
AP and AQ of Shill with the variation of α.

	AP				AQ			
α	CeF	MaxF	MinF	MeanF	CeF	MaxF	MinF	MeanF
0.5	**0.0984**	0.0596	0.0603	0.0586	**0.1792**	0.1125	0.1137	0.1107
0.45	**0.1351**	0.0968	0.0903	0.0885	**0.2381**	0.1766	0.1656	0.1626
0.4	**0.1875**	0.1477	0.1256	0.1388	**0.3158**	0.2574	0.2232	0.2438
0.35	**0.2945**	0.2418	0.2264	0.2367	**0.4550**	0.3895	0.3692	0.3829
0.3	**0.4275**	0.4042	0.3509	0.3616	**0.5990**	0.5757	0.5195	0.5311
0.25	**0.5860**	0.5396	0.4784	0.5255	**0.7390**	0.7010	0.6472	0.6890
0.2	**0.7500**	0.7221	0.6707	0.7072	**0.8571**	0.8386	0.8029	0.8285
0.15	**0.8635**	0.8431	0.8306	0.8370	**0.9268**	0.9149	0.9075	0.9112
0.1	**0.9485**	0.9443	0.9274	0.9401	**0.9736**	0.9714	0.9623	0.9692
0.05	**0.9852**	0.9843	0.9818	0.9831	**0.9926**	0.9921	0.9908	0.9915

We compare the four methods with the threshold α variation in the nine data sets. Tables 17–25 show AP and AQ of fusion results with the variation of α, where CeF denotes the proposed conditional entropy fusion method. The value of α is set from 0.05 to 0.5 with a step of 0.05. From the result, in the Wine data set, when the threshold changes from 0.35 to 0.5, the CeF method clearly outperforms the other three fusion methods and the CeF method is better than the Max method when the value of α is 0.3 and when the threshold changes from 0.05 to 0.25, the values of AP and AQ of the four fusion methods are equal. For the data set CB, our method is significantly better than MinF and MeanF methods when the value of α is set from 0.3 to 0.5. For the SAR data set, when the value of α is varied from 0.2 to 0.5, the CeF method outperforms all the comparison methods For the Ecoli data set, the CeF method is better than MaxF and MeanF methods when α is set from 0.2 and 0.5 and when the value of α is 0.4 and 0.5, our method is better than MinF. In the Breast data set, the fusion performance of the CeF method is the best when the value of the threshold is set from 0.15 to 0.5. The proposed method outperforms the other methods in most situations in Diabetic data set. For HV, Wall, and Shill data sets, the CeF method performs best in almost all situations. All in all, from the perspective of AP, AQ, the CeF method outperforms these common fusion methods on most occasions. Thus, our method present a better choice to fuse multi-source incomplete interval-valued decision information systems in most situations.

Furthermore, for the four methods, the smaller the threshold value, the larger the value of AP and AQ. It stems from the fact that the decreasing of α results in the decreasing of tolerance class of each sample, so the lower approximation of the object set becomes more extensive, and the upper approximation of the object set becomes smaller. Thus, the values of AP and AQ increase when the value of α decreases. We can obtain a high value of AP and AQ by setting a small threshold value. However, while a small value of α increases the values of AP and AQ high, it also results in small tolerance class of each sample, making the connection between samples small. Suppose in an extreme situation that the tolerance classes of all samples barely consist of themselves, the samples will not relate to each other, making the task of data mining hard. Therefore, we are supposed to select an appropriate threshold to extract the tolerance class of samples.

We compare the proposed method with the method using the single information source with maximum AP and AQ, which are shown in Figs. 3 and 4. The results show that the proposed method, CeF, can improve AP and AQ by fusing multiple information sources.

We conducted experiments to demonstrate that the new representation of multi-source data can improve the classification tasks. Many existing classifiers cannot directly cope with interval-valued data; therefore, in [21], the extended classical

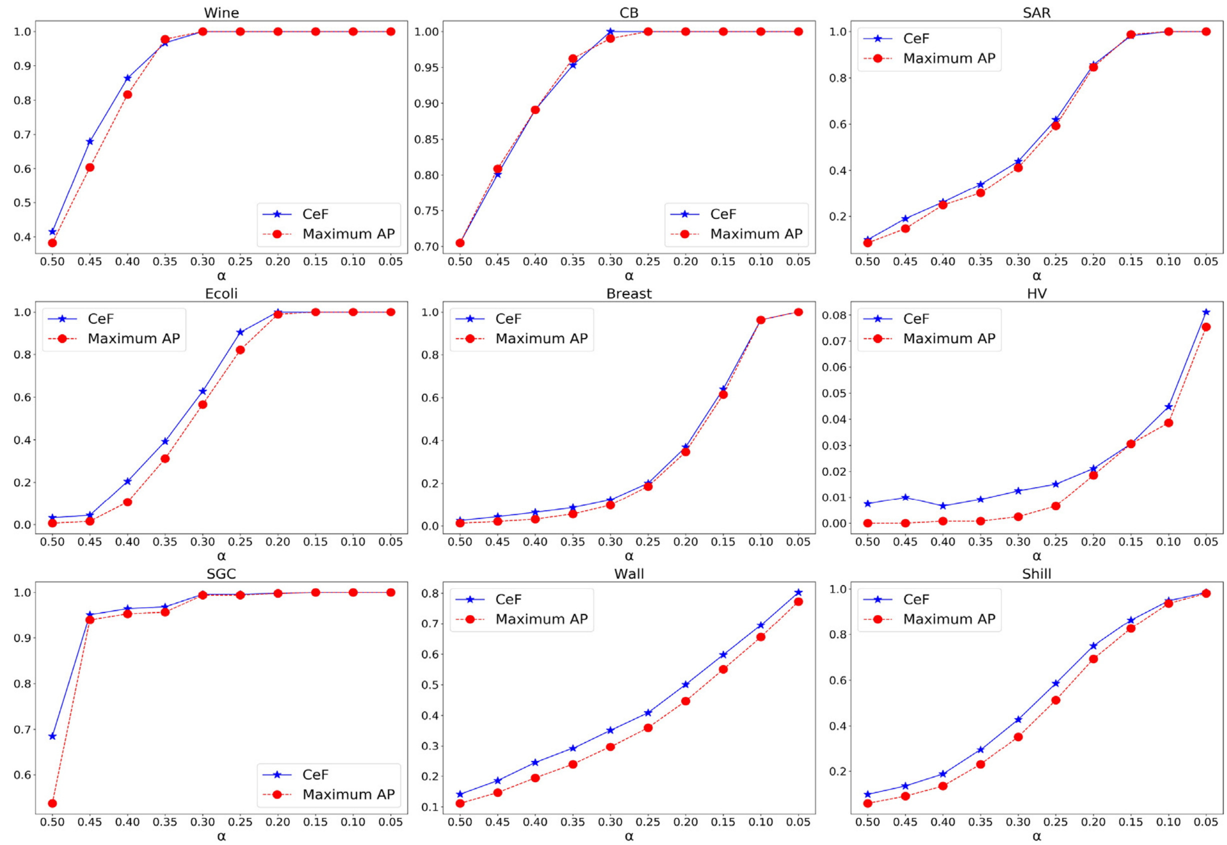

Fig. 3. The comparison results between CeF and single information source with maximum AP.

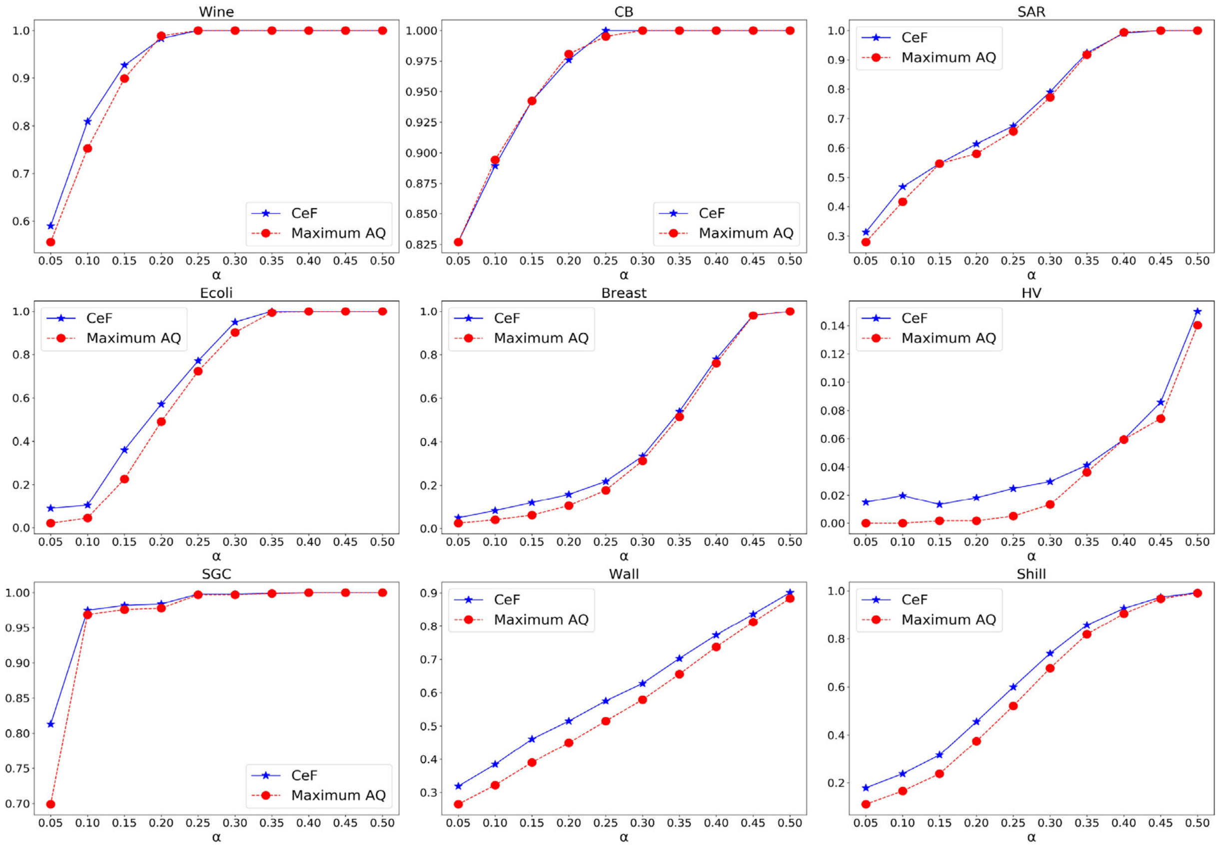

Fig. 4. The comparison results between CeF and single information source with maximum AQ.

Table 26
The classification accuracy of fusion results based on KNN.

DataSets	KNN			
	CeF	MaxF	MinF	MeanF
Wine(k = 15)	**99.4**±**1.7**	98.9±2.3	97.2±2.8	97.7±2.8
CB(k = 3)	**86.6**±**7.7**	80.3±7.2	80.8±8.2	79.4±7.7
SAR(k = 31)	**56.5**±**7.9**	54.7±6.0	40.2±8.6	35.6±7.8
Ecoli(k = 8)	**95.8**±**2.4**	94.0±3.4	**95.8**±**2.8**	94.9±4.3
Breast(k = 6)	**63.5**±**7.2**	62.6±6.0	62.6±4.7	62.6±3.9
HV(k = 5)	**49.9**±**8.8**	46.4±6.7	45.1±7.5	45.4±3.8
SGC(k = 12)	**70.5**±**3.0**	68.9±4.1	70.1±4.8	69.6±3.9
Wall(k = 6)	59.5±1.4	**59.8**±**0.7**	58.4±1.6	58.9±1.2
Shill(k = 6)	**89.3**±**1.2**	89.2±1.2	89.1±1.2	89.2±1.2

Table 27
The classification accuracy of fusion results based on PNN.

DataSets	PNN			
	CeF	MaxF	MinF	MeanF
Wine(σ=0.3)	**99.4**±**1.7**	97.7±2.8	97.7±2.8	98.9±2.2
CB(σ=0.2)	**86.1**±**6.6**	82.7±7.8	80.8±7.8	81.3±7.6
SAR(σ=0.41)	**64.7**±**6.3**	62.0±6.2	63.5±5.8	56.5±7.1
Ecoli(σ=0.2)	**96.7**±**3.6**	94.6±3.8	96.4±3.7	94.6±4.4
Breast(σ=0.2)	62.0±6.6	**62.4**±**6.6**	59.6±8.3	61.7±6.8
HV(σ=0.31)	**52.0**±**3.4**	50.2±2.5	48.8±3.0	49.8±2.2
SGC(σ=0.29)	70.7±3.6	70.0±3.4	**71.6**±**3.7**	70.5±3.7
Wall(σ=0.2)	63.7±1.7	**63.8**±**1.7**	63.5±1.8	63.3±1.7
Shill(σ=0.2)	89.0±1.3	**89.1**±**1.4**	88.8±1.3	89.0±1.3

k-nearest neighbor (KNN) classifier and probabilistic neural network (PNN) classifier are proposed to address interval-valued data. We adopted ten-fold cross-validation in classification, and the mean and standard deviation of classification accuracy are shown in Tables 26 and 27. The parameters k and σ can affect the classification performance of KNN and PNN classifiers, so we achieved the optimal result by adjusting the parameters.

We observe that for the KNN and PNN classifiers, the proposed fusion method performs better than the other three common fusion methods in most datasets. Furthermore, Wilcoxon signed-rank test was employed to determine whether the proposed method is statistically better than the three common methods. Let the null hypothesis be $H_0 : CeF \leqslant MeanF/MinF/MaxF$, which means that the distribution center of classification accuracy of the CeF method is less than or equal to the distribution center of classification accuracy of the four common methods. So the alternative hypothesis is $H_1 : CeF > MeanF/MinF/MaxF$, which means that the distribution center of classification accuracy of the CeF method is larger than the distribution center of classification accuracy of the three common methods. The P-values of the comparison results are shown in Table 28 and Table 29. Given a significance level of 10%, most comparison results are statistically significant.

5.2. The analysis of efficiency

We demonstrate the efficiency of the proposed incremental fusion methods. We compare the running time between incremental fusion algorithms and the static algorithm with the variation of information sources and attributes in nine data sets. Without loss of generality, the value of α is set to 0.5. The results are displayed in Figs. 5–8. The number of attributes and information sources is set from one value to another to compare the runtime of the incremental algorithms with the static algorithm in the above four cases. The x-coordinate (a, b) denotes the number of attributes and the number of sources, where a is the number of attributes and b is the number of sources, and the y-coordinate denotes the runtime of incremental and static algorithms.

The runtime of incremental algorithms is significantly lower than the static algorithm for the four scenarios for the Wine data set. Furthermore, for the other eight data sets, it is clear that in the four cases (a)-(d), the incremental fusion algorithms outperform the static fusion algorithm in terms of running time. We employed the Wilcoxon signed-rank test to determine whether the performance difference between the static and dynamic algorithms is statistically significant. Let the null hypothesis be $H_0 : T_{Static} \leqslant T_{Incremental}$, which means that the distribution center of the runtime of the static algorithm is less than or equal to the distribution center of the runtime of the dynamic algorithm. The alternative hypothesis is $H_1 : T_{Static} > T_{Incremental}$, which means that the distribution center of the runtime of the static algorithm is larger than the dis-

Table 28
P-values of the comparison results in classification accuracy based on KNN.

DataSets	KNN		
	CeF>MaxF	CeF>MinF	CeF>MeanF
Wine(k = 15)	0.5	**0.04733536**	**0.086784083**
CB(k = 3)	**0.006838343**	**0.004305063**	**0.004246019**
SAR(k = 31)	0.362981813	**0.002944635**	**0.002960769**
Ecoli(k = 8)	**0.027237018**	0.5	**0.068351292**
Breast(k = 6)	0.418709626	0.36249459	0.322585841
HV(k = 5)	**0.037701781**	**0.022005492**	**0.096258786**
SGC(k = 12)	**0.080200405**	0.3674303	**0.094690922**
Wall(k = 6)	0.682355341	**0.021912792**	**0.077565784**
Shill(k = 6)	**0.060667625**	**0.032750607**	**0.068351292**

Table 29
P-values of the comparison results in classification accuracy based on PNN.

DataSets	PNN		
	CeF>MaxF	CeF>MinF	CeF>MeanF
Wine(σ=0.3)	**0.086784083**	**0.086784083**	0.5
CB(σ=0.2)	**0.030784922**	**0.004246019**	**0.00689685**
SAR(σ=0.41)	**0.052585976**	0.101545894	**0.004545349**
Ecoli(σ=0.2)	**0.02470581**	0.5	**0.028379723**
Breast(σ=0.2)	0.706031627	**0.036917158**	0.307633063
HV(σ=0.31)	**0.018015843**	**0.011247136**	**0.029529115**
SGC(σ=0.29)	**0.01551626**	0.994588677	0.328360643
Wall(σ=0.2)	0.459380048	0.110335681	**0.077384603**
Shill(σ=0.2)	0.993718499	**0.028953633**	0.58537577

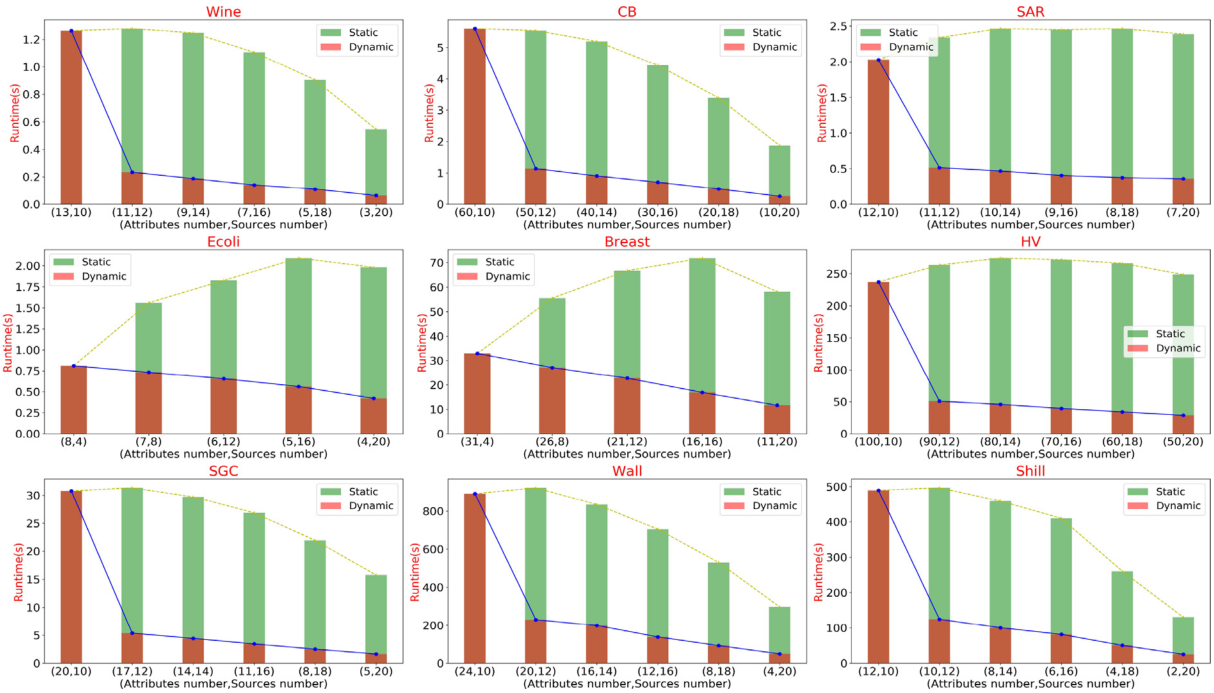

Fig. 5. Runtime times of incremental and static algorithms in Case(a).

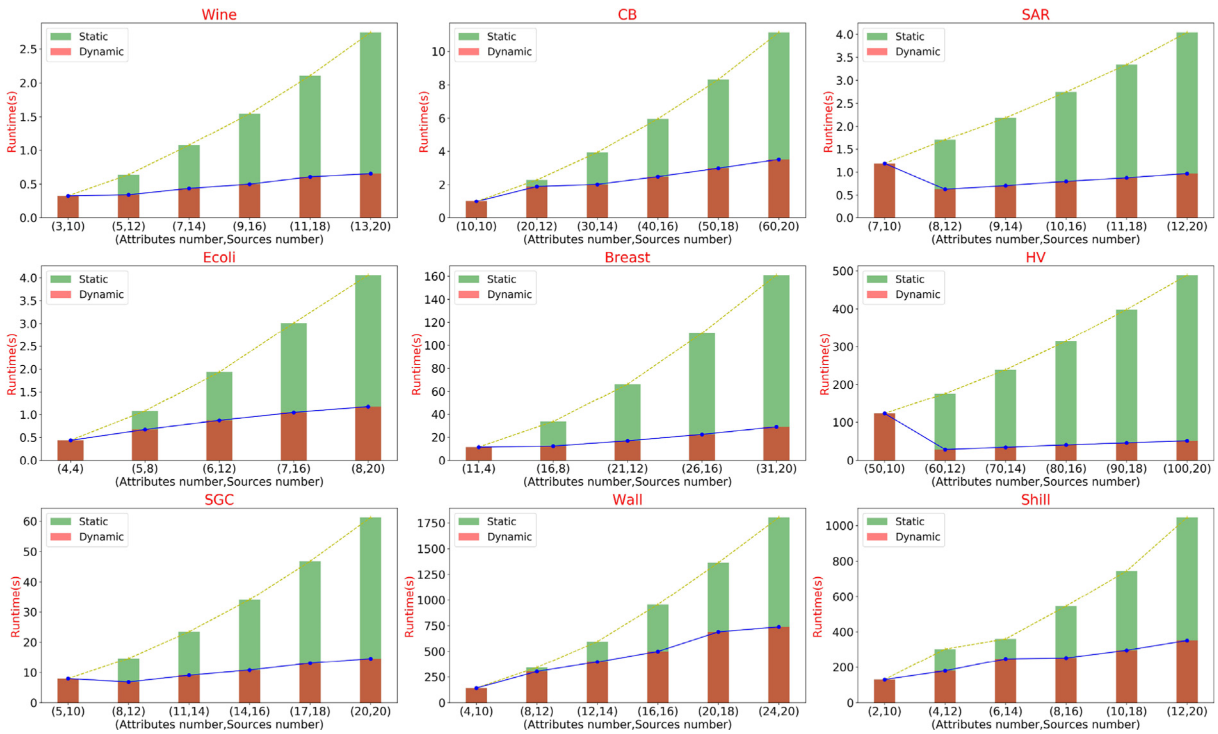

Fig. 6. Runtime times of incremental and static algorithms in Case(b).

tribution center of the runtime of the dynamic algorithm. The P-values of the four situations are all 8.417908123323452e-10. Given a significance level of 5%, all the results are statistically significant. Furthermore, we computed the speed-up ratio of the four incremental algorithms, which is calculated as $R = T_{Staticalgorithm}/T_{Incrementalalgorithm}$, T_* denotes the computational time of

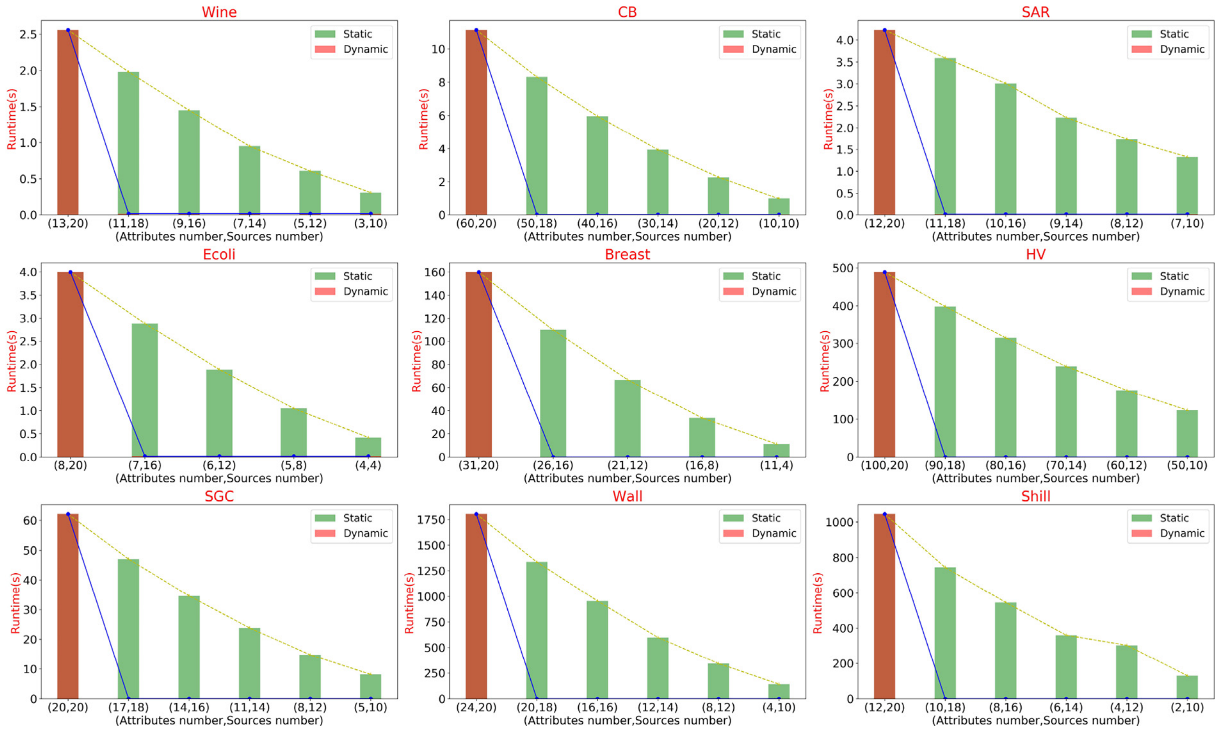

Fig. 7. Runtime times of incremental and static algorithms in Case(c).

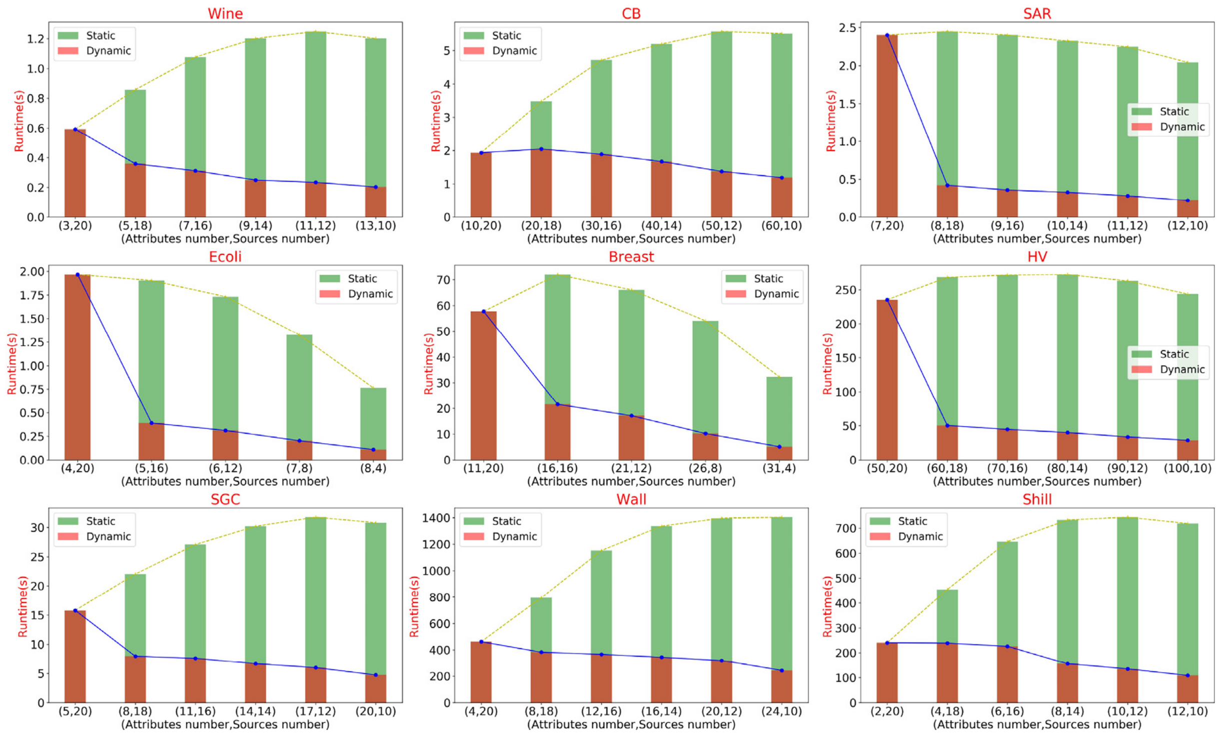

Fig. 8. Runtime times of incremental and static algorithms in Case(d).

Table 30
The speed-up ratio of four incremental algorithms in nine data sets.

Ecoli	a	2.13	2.79	3.72	4.70	
	b	1.60	2.21	2.88	3.47	
	c	185.00	121.00	67.00	27.00	
	d	4.88	5.55	6.54	7.00	
Breast	a	2.05	2.93	4.25	5.00	
	b	2.72	3.91	4.93	5.55	
	c	7040.00	4276.00	2162.00	720.00	
	d	3.34	3.86	5.28	6.31	
CB	a	4.86	5.74	6.31	7.00	7.50
	b	1.21	1.95	2.40	2.79	3.17
	c	714.00	533.00	382.00	252.00	146.00
	d	1.70	2.50	3.11	4.06	4.64
SAR	a	4.55	5.27	6.04	6.58	6.65
	b	2.72	3.11	3.45	3.82	4.18
	c	259.00	214.00	176.00	140.00	109.00
	d	5.81	6.70	7.10	8.00	9.36
HV	a	5.18	5.93	6.89	7.81	8.64
	b	6.13	6.93	7.75	8.66	9.47
	c	6381.25	6720.33	7664.00	5647.50	3952.00
	d	5.32	6.07	6.78	7.80	8.48
Wine	a	5.47	6.67	7.89	8.29	8.75
	b	1.86	2.46	3.09	3.46	4.19
	c	127.00	93.00	61.00	39.00	20.00
	d	2.39	3.45	4.81	5.33	5.92
SGC	a	5.86	6.73	7.82	8.78	9.71
	b	2.14	2.59	3.16	3.58	4.19
	c	3015.00	2220.00	1527.00	949.00	523.00
	d	2.77	3.56	4.53	5.31	6.51
Wall	a	4.02	4.21	5.10	5.74	6.13
	b	1.13	1.50	1.91	1.98	2.44
	c	42835.50	20420.00	19194.50	11014.50	9113.00
	d	2.08	3.15	3.88	4.38	5.77
Shill	a	3.98	4.58	5.04	5.18	5.26
	b	1.69	1.45	2.16	2.51	2.97
	c	47689.00	34950.00	23082.00	19418.00	8348.00
	d	1.90	2.85	4.69	5.51	6.60

Table 31
The comparison results between static algorithm and incremental algorithm for Case (a).

Wine					CB				
Attributes	Sources	Static	Incremental	Ratio	Attributes	Sources	Static	Incremental	Ratio
650	50	341.42	341.42		1,200	20	308.73	308.73	
600	100	626.34	313.37	**2.00**	1100	40	566.44	283.69	**2.00**
550	150	866.55	287.35	3.02	1,000	60	769.51	257.67	2.99
500	200	1046.84	262.22	3.99	900	80	926.32	231.76	4.00
450	250	1180.38	234.95	5.02	800	100	1028.12	206.65	4.98
400	300	1263.11	208.67	6.05	700	120	1081.40	179.81	6.01
350	350	1281.69	183.02	7.00	600	140	1075.85	154.32	6.97
300	400	1260.04	157.56	8.00	500	160	1041.76	128.83	8.09
250	450	1158.18	131.01	8.84	400	180	924.99	102.65	9.01
200	500	1045.88	104.40	10.02	300	200	769.29	77.33	9.95
SAR					Ecoli				
Attributes	Sources	Static	Incremental	Ratio	Attributes	Sources	Static	Incremental	Ratio
360	30	265.92	265.92		320	40	339.96	339.96	
330	60	485.15	241.51	2.01	290	80	616.25	307.18	2.01
300	90	669.15	223.11	3.00	260	120	833.19	276.20	3.02
270	120	802.89	191.10	4.20	230	160	977.39	244.22	4.00
240	150	893.89	179.74	4.97	200	200	1067.90	211.96	5.04
210	180	934.91	154.18	6.06	170	240	1088.87	180.27	6.04
180	210	923.34	132.16	6.99	140	280	1047.23	148.26	7.06
150	240	887.16	108.47	8.18	110	320	922.37	116.70	7.90
120	270	836.07	89.34	9.36	80	360	765.22	85.38	8.96
90	300	707.12	75.96	9.31	50	400	541.87	52.51	**10.32**

the $*$ algorithm. The results of the speed-up ratio are shown in Table 30. From the results, the incremental algorithms yielded 2.05–9.71, 1.13–9.47, 20.00–47689.00, and 1.70–9.36$\times$ speed-up over the static algorithm in Case (a), (b), (c), and (d), respectively. For Case (c), as shown in Example 4.3, the incremental fusion method uses the former information to update the fusion table without any new information. In contrast, the static fusion algorithm needs to recalculate the information entropy for each retained attribute with respect to each retained source. Thus, the speed-up ratio is very high for big data sets, such as Wall and Shill. All in all, the proposed four incremental fusion mechanisms significantly reduce the runtime of fusion with the variation of information sources and conditional attributes.

Table 32
The comparison results between static algorithm and incremental algorithm for Case (b).

Wine					CB				
Attributes	Sources	Static	Incremental	Ratio	Attributes	Sources	Static	Incremental	Ratio
200	50	103.01	103.01		300	20	76.99	76.99	
250	100	249.99	151.21	1.65	400	40	201.48	125.42	1.61
300	150	453.35	203.44	2.23	500	60	390.03	177.20	2.20
350	200	734.17	257.64	2.85	600	80	624.24	231.91	2.69
400	250	1076.45	319.07	3.37	700	100	917.66	290.99	3.15
450	300	1450.23	373.51	3.88	800	120	1260.43	340.94	3.70
500	350	1875.09	433.73	4.32	900	140	1650.21	390.25	4.23
550	400	2350.14	483.73	4.86	1000	160	2083.11	442.59	4.71
600	450	2930.46	539.73	5.43	1,100	180	2620.99	498.44	5.26
650	500	3478.88	594.32	**5.85**	1200	200	3244.66	588.90	5.51
SAR					Ecoli				
Attributes	Sources	Static	Incremental	Ratio	Attributes	Sources	Static	Incremental	Ratio
90	30	70.31	70.31		50	40	55.47	55.47	
120	60	189.35	116.95	1.62	80	80	174.58	123.76	**1.41**
150	90	352.98	161.13	2.19	110	120	365.85	189.00	1.94
180	120	571.99	213.71	2.68	140	160	604.55	260.12	2.32
210	150	824.28	257.18	3.21	170	200	928.85	312.59	2.97
240	180	1147.67	303.94	3.78	200	240	1303.98	378.52	3.44
270	210	1495.46	354.01	4.22	230	280	1762.70	447.30	3.94
300	240	1894.52	397.55	4.77	260	320	2274.21	507.50	4.48
330	270	2369.06	445.98	5.31	290	360	2881.45	590.09	4.88
360	300	2967.62	520.43	5.70	320	400	3403.13	644.85	5.28

Table 33
The comparison results between static algorithm and incremental algorithm for Case (c).

Wine					CB				
Attributes	Sources	Static	Incremental	Ratio	Attributes	Sources	Static	Incremental	Ratio
650	500	3360.82	3360.82		1,200	200	3114.14	3114.14	
600	450	2893.80	0.78	3709.98	1100	180	2593.24	9.44	274.76
550	400	2368.33	0.72	3300.33	1,000	160	2096.62	9.11	230.13
500	350	1897.63	0.58	3287.62	900	140	1674.03	7.58	220.80
450	300	1445.30	0.55	2647.06	800	120	1256.01	7.27	172.77
400	250	1071.38	0.41	2641.46	700	100	910.14	5.46	166.69
350	200	746.92	0.37	1994.96	600	80	625.81	4.74	131.96
300	150	473.84	0.39	1214.96	500	60	387.10	3.76	102.96
250	100	263.05	0.20	1297.08	400	40	205.33	5.66	36.26
200	50	105.33	0.19	562.67	300	20	77.08	4.12	**18.72**
SAR					Ecoli				
Attributes	Sources	Static	Incremental	Ratio	Attributes	Sources	Static	Incremental	Ratio
360	300	2791.22	2791.22		320	400	3525.45	3525.45	
330	270	2193.11	0.51	4260.09	290	360	2860.22	0.42	**6790.59**
300	240	1839.44	0.37	4913.00	260	320	2287.41	0.44	5236.71
270	210	1449.08	0.41	3572.65	230	280	1758.43	0.36	4900.83
240	180	1021.09	0.27	3850.24	200	240	1292.98	0.33	3946.81
210	150	748.82	0.27	2823.59	170	200	910.16	0.27	3431.94
180	120	514.32	0.17	2997.18	140	160	612.60	0.23	2617.93
150	90	335.48	0.17	1955.00	110	120	352.80	0.17	2055.91
120	60	176.86	0.12	1417.13	80	80	174.13	0.12	1395.25
90	30	66.69	0.09	712.50	50	40	54.02	0.08	692.60

Table 34
The comparison results between static algorithm and incremental algorithm for Case (d).

Wine					CB				
Attributes	Sources	Static	Incremental	Ratio	Attributes	Sources	Static	Incremental	Ratio
200	500	991.64	991.64		300	200	757.34	757.34	
250	450	1141.26	217.15	5.26	400	180	918.36	228.93	4.01
300	400	1263.58	204.08	6.19	500	160	1026.53	223.46	4.59
350	350	1308.22	185.13	7.07	600	140	1089.48	186.19	5.85
400	300	1271.89	160.42	7.93	700	120	1093.27	161.79	6.76
450	250	1204.78	134.75	8.94	800	100	1038.90	136.11	7.63
500	200	1067.23	107.66	9.91	900	80	932.81	108.55	8.59
550	150	877.80	80.56	10.90	1000	60	778.96	82.57	9.43
600	100	635.41	53.90	11.79	1,100	40	570.50	55.43	10.29
650	50	344.47	26.93	**12.79**	1200	20	312.74	28.11	11.12
SAR					Ecoli				
Attributes	Sources	Static	Incremental	Ratio	Attributes	Sources	Static	Incremental	Ratio
90	300	719.16	719.16		50	400	542.65	542.65	
120	270	861.73	212.89	4.05	80	360	790.47	294.16	**2.69**
150	240	958.17	190.96	5.02	110	320	981.09	258.28	3.80
180	210	1002.07	165.53	6.05	140	280	1063.38	229.18	4.64
210	180	999.42	141.54	7.06	170	240	1106.55	198.34	5.58
240	150	952.71	117.87	8.08	200	200	1081.73	163.13	6.63
270	120	838.22	95.35	8.79	230	160	915.41	130.09	7.04
300	90	697.18	69.58	10.02	260	120	752.69	85.60	8.79
330	60	508.28	46.91	10.84	290	80	563.68	62.77	8.98
360	30	264.81	22.28	11.89	320	40	306.26	29.13	10.52

To demonstrate the efficiency of the proposed dynamic fusion mechanisms for Ms-IIVDSs which consist of large sources and attributes, we chose the datasets CB, SAR, Ecoli, and Wine to artificially construct four datasets by copying the original conditional attributes 20, 30, 40, and 50 times respectively. Furthermore, we create 200, 300, 400, and 500 information sources for the four datasets, respectively. The comparison results between static and incremental algorithms are shown in Tables 31–34. From the results, the proposed four dynamic fusion approaches reduce runtime effectively for the datasets that consist of large sources and attributes.

6. Conclusion and future work

In this paper, a novel fusion method is established to fuse multiple incomplete interval-valued information tables using conditional entropy. We first define the distance between any two samples and use the distance to define the tolerance class. Then we define the conditional entropy using the tolerance classes of samples and selecting information sources that are pivotal for attributes to compose new information tables. After that, we complete the missing values using the maximum fusion results of the samples in the tolerance classes of the missing value samples. We present the static algorithm to show the fusion process and analyze its time complexity. Four incremental fusion approaches are proposed for the variation of information sources and conditional attributes. The four incremental algorithms are established together with their time complexities. In the end, experimental results on nine data sets show that, in the aspect of fusion effectiveness, the proposed fusion method is better than the three common fusion methods in terms of AP and AQ in most scenarios. Moreover, the efficiency analysis results show that, for the four scenarios, these incremental fusion approaches effectively save the running time of fusion when the number of sources and attributes vary continuously.

In the era of big data, it sometimes takes a lot of time and money to label the category of each object. Therefore, it is of great significance to extend this study to cope with unlabeled information systems, that is, unsupervised or semi-supervised fusion methods. Establishing effective fusion mechanisms for this kind of problem is a challenge. Furthermore, we also face some complicated classification problems in real life, such as multilabel classification, fuzzy classification, and hierarchical classification. So it is also a profound direction to study how to improve the accuracy of these problems in multi-source information systems.

CRediT authorship contribution statement

Xiaoyan Zhang: Investigation, Methodology, Project administration, Supervision, Validation. **Xiuwei Chen:** Data curation, Methodology, Software, Visualization, Writing - original draft, Writing - review & editing. **Weihua Xu:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation. **Weiping Ding:** Funding acquisition, Investigation, Software.

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.

Acknowledgements

This paper is supported by the National Natural Science Foundation of China (Nos. 61976245, 61976120).

References

[1] A. Blanco-Fernández, N. Corral, G. Gonzá lez-Rodríguez, Estimation of a flexible simple linear model for interval data based on set arithmetic, Computational Statistics & Data Analysis, 55(9) (2011), pp. 2568-2578..
[2] P. Durso, J.M. Leski, Fuzzy c-ordered medoids clustering for interval-valued data, Pattern Recognition 58 (2016) 49–67.
[3] S. Dias, P. Brito, Off the beaten track: a new linear model for interval data, European Journal of Operational Research 258 (3) (2017) 1118–1130.
[4] Y. Leung, M.M. Fischer, W.Z. Wu, J.S. Mi, A rough set approach for the discovery of classification rules in interval-valued information systems, International Journal of Approximate Reasoning 47 (2) (2008) 233–246.
[5] D.S. Guru, B.B. Kiranagi, P. Nagabhushan, Multivalued type proximity measure and concept of mutual similarity value useful for clustering symbolic patterns, Pattern Recognition Letters 25 (10) (2004) 1203–1213.
[6] J.H. Dai, W.T. Wang, Q. Xu, An Uncertainty Measure for Incomplete Decision Tables and Its Applications, IEEE Transactions on Cybernetics 43 (4) (2013) 1277–1289.
[7] J.H. Dai, Y.J. Yan, H. Shi, Knowledge granularity measures for incomplete interval-valued information, in: 2017 3rd International Conference on Information Management (ICIM), Chengdu, 2017, pp. 227–231.
[8] C. Luo, T.R. Li and Z. Yi, An Incremental Feature Selection Approach Based on Information Entropy for Incomplete Data, 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..
[9] H. Zhao, K.Y. Qin, Mixed feature selection in incomplete decision table, Knowledge-Based Systems 57 (2014) 181–190.
[10] L. Sun, L.Y. Wang, Y.H. Qian, J.C. Xu, S.G. Zhang, Feature selection using Lebesgue and entropy measures for incomplete neighborhood decision systems, Knowledge-Based Systems 186 (2019) 104962.
[11] H.X. Li, M.H. Wang, X.Z. Zhou, J.B. Zhao, An interval set model for learning rules from incomplete information table, International Journal of Approximate Reasoning 53 (1) (2012) 24–37.
[12] T.P. Hong, L.H. Tseng, S.L. Wang, Learning rules from incomplete training examples by rough sets, Expert Systems with Applications 22 (4) (2002) 285–293.
[13] Y. Leung, D.Y. Li, Maximal consistent block technique for rule acquisition in incomplete information systems, Information Sciences 153 (2003) 85–106.
[14] Z. Pawlak, Rough sets, International Journal of Computer & Information Sciences 11 (5) (1982) 341–365.
[15] W.H. Xu, J.H. Yu, A novel approach to information fusion in multi-source datasets: a granular computing viewpoint, Information Sciences 378 (2017) 410–423.
[16] B.B. Sang, Y.T. Guo, D.R. Shi, W.H. Xu, Decision-theoretic rough set model of multi-source decision systems, International Journal of Machine Learning & Cybernetics 9 (11) (2018) 1941–1954.
[17] C.E. Shannon, W. Weaver, The mathematical theory of communication, The Bell System Technical Journal 27 (3/4) (1948) 373–423.
[18] W.H. Xu, M.M. Li, and X.Z. W, Information Fusion Based on Information Entropy in Fuzzy Multi-source Incomplete Information System. International Journal of Fuzzy Systems, 19, (2017), pp. 1200-1216..
[19] J.C. Xu, Y. Wang, H.Y. Mu, F.Z. Huang, Feature Genes Selection Based on Fuzzy Neighborhood Conditional Entropy, Journal of Intelligent and Fuzzy Systems (2018) 117–126.
[20] G.P. Lin, J.Y. Liang, Y.H. Qian, An information fusion approach by combining multigranulation rough sets and evidence theory, Information Sciences 314 (2015) 184–199.
[21] L. Piras, G. Giacinto, Information fusion in content based image retrieval: A comprehensive overview, Information Fusion 37 (2017) 50–60.
[22] X.W. Chen, W.H. Xu, Double-quantitative multigranulation rough fuzzy set based on logical operations in multi-source decision systems, International Journal of Machine Learning and Cybernetics, doi: 10.1007/s13042-021-01433-2..
[23] L. Yang, W.H. Xu, X.Y. Zhang, B.B. Sang, Multi-granulation method for information fusion in multi-source decision information system, International Journal of Approximate Reasoning 122 (2020) 47–65.
[24] Y.Y. Zhang, T.R. Li, C. Luo, J.B. Zhang, H.M. Chen, Incremental updating of rough approximations in interval-valued information systems under attribute generalization, Information Sciences 373 (2016) 461–475.
[25] H.M. Chen, T.R. Li, C. Luo, S.J. Horng, G.Y. Wang, A decision-theoretic rough set approach for dynamic data mining, IEEE Transactions on Fuzzy Systems 23 (6) (2015) 1958–1970.
[26] D. Liu, T.R. Li, J.B. Zhang, Incremental updating approximations in probabilistic rough sets under the variation of attributes, Knowledge-Based Systems 73 (1) (2015) 81–96.
[27] J.H. Yu, W.H. Xu, Incremental knowledge discovering in interval-valued decision information system with the dynamic data, International Journal of Machine Learning and Cybernetics 8 (3) (2017) 849–864.
[28] 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, Information Sciences 584 (1) (2022) 127–147.
[29] W.T. Li, W.H. Xu, X.Y. Zhang, J. Zhang, Updating approximations with dynamic objects based on local multigranulation rough sets in ordered information systems, Artificial Intelligence Review (2021), https://doi.org/10.1007/s10462-021-10053-9.
[30] W.H. Xu, K.H. Yuan, W.T. Li, et al, An emerging fuzzy feature selection method using composite entropy-based uncertainty measure and data distribution, IEEE Transactions on Emerging Topics in computational Inteligence. (2022), https://doi.org/10.1109/TETCI.2022.3171784.
[31] Y.Y. Huang, T.R. Li, C. Luo, H. Fujita, S. Horng, Dynamic Fusion of Multisource Interval-Valued Data by Fuzzy Granulation, IEEE Transactions on Fuzzy Systems 26 (6) (2018) 3403–3417.
[32] Y.Y. Huang, T.R. Li, C. Luo, H. Fujita, S-. Horng, B. Wang, Dynamic maintenance of rough approximations in multi-source hybrid information systems, Information Sciences 530 (2020) 108–127.
[33] Z. Pawlak, Rough sets: Theoretical Aspects of Reasoning About Data, Kluwer, Norwell, MA, USA, 1991.
[34] Z. Pawlak, A. Skowron, Rudiments of rough sets, Information Sciences 177 (1) (2007) 3–27.
[35] M. Kryszkiewicz, Rules in incomplete information systems, Information Sciences 113 (3/4) (1999) 271–292.
[36] M. Kryszkiewicz, Rough set approach to incomplete information systems, Information Sciences 112 (1/4) (1998) 39–49.
[37] http://archive.ics.uci.edu/ml/datasets.html..