Knowledge-Based Systems 330 (2025) 114485

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

A supervised learning approach to dynamic weighted fusion in multi-source ordered decision systems

Xiaoyan Zhang *, Jiajia Lin

College of Artificial Intelligence, Southwest University, Chongqing, 400715, PR China

ARTICLE INFO

Keywords:
Dynamic fusion
Information entropy
Weight generation
Multi-source ordered decision systems

ABSTRACT

With the rapid advancement of new-generation artificial intelligence technologies, machines can process and analyze large-scale data more accurately and efficiently and for more complex tasks. Enhancing the usability and value of the information derived from various information systems across multiple dimensions is essential. However, traditional data dominance relationships cannot reflect people's different levels of attention to antithetic features, leading to higher complexity and lower classification accuracy. Therefore, it is necessary to consider the weight relationships between attributes in the data, which refers to the degree of correlation between each attribute and the decision in multi-source information systems. Based on these weights and dominance relationships, we consider an entropy-based weighted information fusion method for processing supervised data in multi-source ordered decision systems. We intend four incremental fusion mechanisms to adjust information sources and attribute changes to save running time. Furthermore, experiments are conducted on nine real datasets to demonstrate our method's effectiveness. The results show that the inevitable accuracy comparisons by the proposed method are superior to most fusion methods. In addition, the dynamic mechanisms, compared to static mechanisms, can significantly reduce running time.

1. Introduction

The supervised information fusion method is an information fusion technology that makes decisions based on known labels or categories of data, and it has wide-ranging applications in various fields [1–4].

Furthermore, rough set theory was proposed by Pawlak [5] in 1982 and is a mathematical tool for handling incomplete and fuzzy data. Fuzzy set theory, introduced by Zadeh [6] in 1965, deals with vague and uncertain concepts in the real world by using fuzzy sets and membership functions. Fuzzy rough set [7], formed by combining the two theories, which describe the fuzziness of attribute values using fuzzy membership functions and combine the approximation and equivalence relations of rough sets for data processing. Another concept of rough membership functions in pattern classification tasks has evolved and expanded into rough-fuzzy membership functions and ownership functions. Sarkar [8] introduced this evolution. Xu [9] constructed multi-granulation fuzzy rough sets on tolerance relations. After that, Xu [10] proposed a local multi-granulation neighborhood rough set model and explored dynamic approximate updating algorithms for data. By combining covering-based rough sets, fuzzy rough sets, and multi-granulation rough sets, covering-based multi-granulation fuzzy rough set models were introduced by Zhan [11] using fuzzy β-neighborhoods. Fuzzy set and rough set theory applied in feature selection [12,13], knowledge discovery [14–16], data mining [17], pattern recognition [18], decision analysis [19], machine learning [20], and other fields. Scholars have gradually refined fuzzy rough set theory. At the same time, based on the rough set theory, many researchers have proposed development models based on information fusion [21–24].

Initially proposed by Shannon [25], information entropy is a mathematical tool used to measure uncertainty and is now widely applied in information processing [26]. Entropy-based methods have flourished in different fields [27–30], such as when combined with information fusion. In order to extract useful information from incomplete multi-source data, a fusion method combining information entropy was proposed by Li [31]. Based on information entropy, a fusion method, developed by Xu [32] in fuzzy incomplete information system, and a novel information fusion method, proposed by Zhang [33] for multi-source incomplete interval-valued data. Additionally, Zhang [34] also introduced a two-way concept-cognitive learning method in the context of multi-source fuzzy data by integrating information fusion based on information entropy. Xu [35] established a fusion model for multi-source interval-valued ordered data based on the defined fuzzy dominance conditional entropy. The existing multi-source fusion methods always result in a single-source information system. In order to improve this situation,

* Corresponding author.
E-mail addresses: zhangxiaoyan@swu.edu.cn (X. Zhang), pnkaka8007@163.com (J. Lin).

https://doi.org/10.1016/j.knosys.2025.114485
Received 29 August 2023; Received in revised form 12 February 2024; Accepted 12 September 2025
Available online 18 September 2025

Table 1
Comparison of fusion methods based on information entropy.

Models	Main content
Conditional entropy [32]	Conditional entropy determines the importance of each information source for the entire system and facilitates multi-source data fusion by minimizing conditional entropy.
Generalized information entropy [37]	A novel information imputation technique based on generalized information entropy is proposed to address the issue of incompleteness in decision fusion.
Extension model based on information entropy [39]	For example, a new fault diagnosis method utilizes the Bayesian network to infer the fault severity of power components.

Yang [36] proposed a multi-granulation method for information fusion in multi-source decision information systems.

The primary approach for information fusion based on information entropy is to employ information entropy as a metric to quantify the amount of information contained in each source, effectively leveraging the crucial properties of information entropy to address the issue of information asymmetry in multi-source data. Consequently, this facilitates acquiring and fusing multi-source data's informational content or rules. As shown in Table 1, research on utilizing information entropy to achieve multi-source information fusion within multi-source systems primarily concentrates on three key aspects:

(1) The conditional entropy determines the importance of each information source to the whole system and the multi-source data fusion through the minimum conditional entropy method [32].
(2) The generalized information entropy is used for multi-source data fusion and applied to practical problems such as multi-sensor decision fusion [37], sensor strategy data processing, fusion index evaluation, and multi-label feature selection [38] in multi-source data.
(3) An expanded information entropy model can be the foundation for multi-source information fusion. For example, Zeng [39] proposes a new fault diagnosis method for power grids based on multi-source information fusion theory. Bayesian networks can help derive the degree of fault of power components to address the problem of misjudgment caused by traditional fault diagnosis methods. Following each circuit failure, the electric signal undergoes extraction of time-domain singular spectrum entropy, frequency-domain power spectral entropy, and wavelet packet energy spectral entropy. Multi-source information fusion can amalgamate these three feature quantities to construct the fault support component of the power source. The proposed method can improve the accuracy and reliability of fault diagnosis in power grids.

However, we propose a weight-based information fusion method with conditional entropy. In order to describe the typical relationship between samples more specifically, we introduce the conditional entropy of weighted dominance distance fuzzy relation to fuse the data of a multi-source ordered decision system, and the weight generation method [40] was combined. Based on this conditional entropy, a supervised information fusion method is proposed for multi-source ordered decision information systems based on weights. Firstly, the concepts of weighted distance fuzzy relation and dominance degree fuzzy relation between two ordered samples are introduced. Then, the weighted distance fuzzy class and the dominance degree fuzzy class are defined for each sample. On this basis, the conditional entropy of the weighted dominance distance fuzzy relation can be defined by combining it with the decision dominance relation. For each attribute, the lower the entropy, the more critical it is in the system. This implies that the conditional entropy needs to be calculated for each attribute in different systems separately, and an attempt is made to identify the system where each attribute is most important. Finally, the attribute values from these systems are synthesized into a new system as a result of the fusion.

Helpful information also needs to be updated accordingly. For dynamic data, using static methods can be time-consuming. In order to cope with data changes and effectively handle updated data, many researchers have conducted explorations [41–43]. A summary of multi-source information fusion was provided by Zhang [23], and future projections were made. The incremental learning still shows great potential. After research methods focusing on static multi-source environments were proposed [38,44,45], Zhu [46] presented an approach for incrementally fusing fuzzy and uncertain data. Subsequently, fusion methods were proposed by Huang [47] based on attribute and source dynamic updates. A matrix-based method was proposed by Zhang [48] for dynamically updating multigranulation fusion operators to cope with changes in samples and sources.

Based on the above analysis, while proposing the weight-based information fusion method, this paper also presents a fusion method based on attribute and source dynamic updates. The overall framework is shown in Fig. 1. Our main contributions are as follows:

(1) The focus of our study is on multi-source ordered decision systems. The attribute matrix and decision matrix are categorized and calculated. The partition coefficient of attributes is calculated as well. Then, the partition coefficients of attributes are processed to generate weight vectors for the attributes in the information system. Then, the weighted distance fuzzy relation for attributes is defined for multi-source ordered datasets.
(2) The dominance degree fuzzy relation for attributes is defined and combined with the weighted distance fuzzy relation to define the conditional entropy. The entropy of the weighted dominance distance fuzzy relation for each attribute in each information system is calculated using the information entropy method to describe the degree of uncertainty of attributes in each system. Once the most accurate data among multiple information systems is identified, it is fused into our fusion data, which represents the optimal data.
(3) Based on previous research, four incremental learning algorithms are designed to adapt to changes in sources and attributes at different times while maintaining and updating previously learned models.
(4) Comparative experiments are conducted on nine datasets, and the results show that our proposed static and dynamic fusion methods both have excellent performance.

This article is organized as follows. To facilitate understanding, Section 2 introduces the necessary and fundamental knowledge of multi-source ordered decision systems, fuzzy sets, and information entropy. Section 3 defines the weighted distance fuzzy relation and dominance degree fuzzy relation and presents the fusion method on multi-source ordered decision systems. Section 4 discusses four incremental fusion mechanisms and provides corresponding fusion algorithms. Section 5 presents the experimental data and plots the experimental results. Finally, in Section 6, the entire paper is recapped, and prospects are provided.

2. Related work

This section briefly reviews several fundamental concepts, specifically multi-source ordered decision systems (MS-ODS), fuzzy sets, and information entropy for our current work. These concepts are thoroughly discussed in the references [6,23,33].

2.1. Multi-source ordered decision systems

Consider an information system (IS) denoted as $\mathcal{IS} = (U, AT, F)$, where $U = \{x_1, x_2, \dots, x_n\}$ is finite object set, $AT = \{a_1, a_2, \dots, a_p\}$ is finite condition attribute set, $F = \{f \mid U \to V_a, a \in AT\}$ is relationship set between U and AT, V_a is the finite value domain of a. Therefore, a decison system (DS) can be denoted as $\mathcal{DS} = (U, AT \bigcup D, F, G)$, where $D = \{d_1, d_2, \dots, d_q\}$ is finite decision attribute set, $G = \{g \mid U \to V_d, d \in D\}$ is relationship set between U and D, V_d is the finite value domain of d.

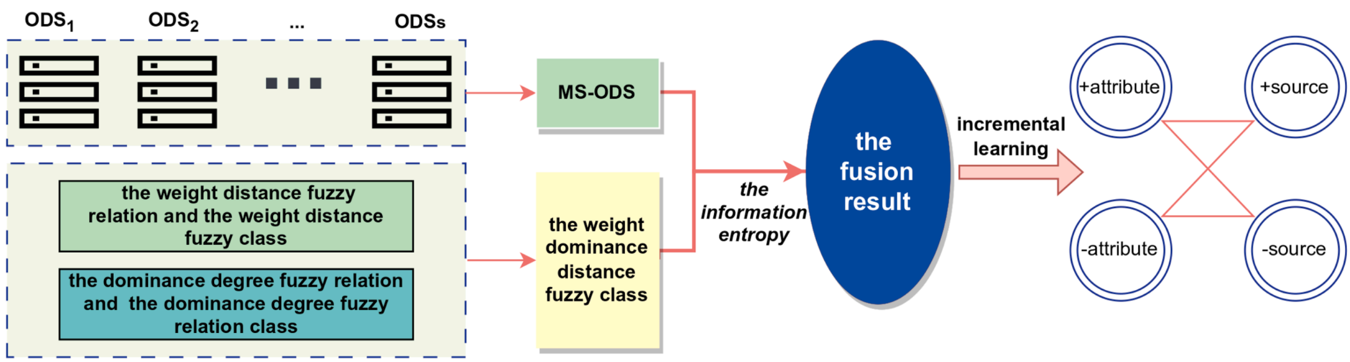

Fig. 1. Overview of our work.

For any $P \subseteq AT$, a relation $R_P \subseteq U \times U$ can be defined. And for any $a \in U$, the relation class of a is denoted as $[a]^{R_P} = \{b \in U | (a, b) \in R_P\}$. For any $A \subseteq U$, the lower and upper approximations of A are denoted as:

$$\underline{R_P}(A) = \{a \in U | [a]^{R_P} \subseteq A\}, \tag{1}$$

$$\overline{R_P}(A) = \{a \in U | [a]^{R_P} \cap A \neq \emptyset\}. \tag{2}$$

Let $U_D = \{D_1, D_2, \ldots, D_m\}$ be the subspace of U based on D, which corresponds to the classification of elements in U. the AP and AQ of U_D with respect to R_P are represented as follows:

$$AP_{R_P}(U_D) = \frac{\sum_{i=1}^{m} \left|\underline{R_P}(D_i)\right|}{\sum_{i=1}^{m} \left|\overline{R_P}(D_i)\right|}, \tag{3}$$

$$AQ_{R_P}(U_D) = \frac{\sum_{i=1}^{m} \left|\underline{R_P}(D_i)\right|}{|U|}. \tag{4}$$

AP and AQ were first proposed by Pawlak. The former is used to measure the proportion of correct classifications in the approximate classification result, while the latter is used to measure the consistency between the approximate classification result and the true class. Both of them indicate better results when the value is higher.

For any $p \in AT$, S_p is completely pre-ordered by the relation $\preceq_p$, where $a \preceq_p b \iff f(a,p) \leq f(b,p)$ or $a \preceq_p b \iff f(a,p) \geq f(b,p)$, then the decision system can be referred to as an ordered decision system (ODS). Essentially, this implies that there exists a preference that can be expressed as either increasing or decreasing.

$\mathcal{MS}-\mathcal{ODS} = \{\mathcal{ODS}_i | \mathcal{ODS}_i = (U, AT_i \bigcup D_i, F, G), i = 1, 2, \ldots, s\}$ is used to define an MS-ODS, where $\mathcal{ODS}_i$ represents ith ODS of MS-ODS.

2.2. Fuzzy set

Let $\widetilde{A}$ be a fuzzy set on the domain $U = \{u_1, u_2, \ldots, u_n\}$ with $\widetilde{A} : U \to [0,1]$. The membership function of $\widetilde{A}$, denoted as $\widetilde{A}(x)$, means the degree to which x belongs to $\widetilde{A}$, $x \in U$. Let $\widetilde{A} = \sum_{k=1}^{n} \frac{\widetilde{A}(u_k)}{u_k}$, $\widetilde{B} = \sum_{k=1}^{n} \frac{\widetilde{B}(u_k)}{u_k}$. The operations of intersection, union and complement for $\widetilde{A}$ and $\widetilde{B}$ are defined as follows:

$$\widetilde{A} \cap \widetilde{B} = \sum_{k=1}^{n} \frac{\widetilde{A}(u_k) \wedge \widetilde{B}(u_k)}{u_k}, \tag{5}$$

$$\widetilde{A} \cup \widetilde{B} = \sum_{k=1}^{n} \frac{\widetilde{A}(u_k) \vee \widetilde{B}(u_k)}{u_k}, \tag{6}$$

$$\widetilde{A}^C = \sum_{k=1}^{n} \frac{1 - \widetilde{A}(u_k)}{u_k}. \tag{7}$$

Table 2
The abbreviations of the terminologies.

Terminologies	Abbreviations
ODS	ordered decision system
$\widetilde{W}_p(a,b)$	the weighted distance fuzzy relation between sample a and sample b under attribute p
$\widetilde{D}_p(a,b)$	the dominance degree fuzzy relation between sample a and sample b under attribute p
$R_D(a,b)$	the decision dominance relation of D between sample a and sample b
$E_p(D\|\mathcal{ODS}_q)$	the information entropy of $\mathcal{ODS}_q$ for attribute p under decision attribute D
$i_p = \arg\min_{q \in 1,2,\ldots,s} E_p(D\|\mathcal{ODS}_q)$	The ith source which is the most essential for attribute p under the conditional entropy of weighted dominance distance fuzzy relation
$[a]_{\widetilde{W}_p}$	the weighted distance fuzzy class of sammple a under attribute p
$[a]_{\widetilde{D}_p}$	the dominance degree fuzzy class sammple a under attribute p
$[a]_{\widetilde{W \cap D}_p}$	the weighted dominance distance fuzzy class of a under p

2.3. Information entropy of ODS

To measure the degree of uncertainty, a novel form of information entropy is introduced in ODS. For any $P \subseteq AT$, $U_D = \{D_1, D_2, \ldots, D_m\}$, the information entropy of D with respect to P is computed as follows:

$$H(D|P) = -\sum_{i=1}^{|U|} \sum_{j=1}^{m} \frac{\left|[x_i]^{R_P} \cap D_j\right|}{|U|} \log \frac{\left|[x_i]^{R_P} \cap D_j\right|}{\left|[x_i]^{R_P}\right|}. \tag{8}$$

It is well known that $H(D|P)$ satisfies the following theorem:

$0 \leqslant H(D|P) \leqslant |U| \log |U|$,

$H(D|P_1) \leqslant H(D|P_2)$, if $P_2 \subseteq P_1$.

3. Weighted information fusion method based on MS-ODS

Fusing data from multiple sources can enhance knowledge discovery by generating a comprehensive and unified representation. The weight can indicate the relevance between attributes and decisions and can be combined with information entropy to measure the amount of information and its significance more accurately. The weight generation method can be found in [40].In this section, a weighted fusion method is introduced for MS-ODS. All Terminologies and abbreviations are presented in Table 2.

3.1. Weight distance fuzzy relation

The generation process of weights is as follows.

Given an ODS, $\mathcal{ODS} = (U, AT \bigcup D, F, G)$, for $\forall p \in AT$, $a \in U$, $f(a,p)$ represents the value of p for a, the coefficient matrix is showed as follows:

$$M_A = \begin{bmatrix} f(a_1,p_1) & f(a_1,p_2) & \cdots & f(a_1,p_m) \\ f(a_2,p_1) & f(a_2,p_2) & \cdots & f(a_2,p_m) \\ \vdots & \vdots & \ddots & \vdots \\ f(a_n,p_1) & f(a_n,p_2) & \cdots & f(a_n,p_m) \end{bmatrix}. \tag{9}$$

The vector of the decision attribute d be $M_d = (g(a_1,d), g(a_2,d), \ldots, g(a_n,d))^T$. Assuming $M_A\eta = M_d$, then $M_A^T M_A \eta = M_A^T M_d$, and the coefficient of attribute partition $\eta = (M_A^T M_A)^{-1} M_A^T M_d$ can be obtained. If the matrix $M_A^T M_A$ is not invertible, or a penalty term is in need in the optimum function, we can solve for $\eta = (\eta(p_1), \eta(p_2), \ldots .\eta(p_m))^T$ by using $(M_A^T M_A + E)\eta = M_A^T M_d$. The weight of p can be calculated as follows:

$$\omega(p) = \frac{|AT| \times |\eta(p)|}{\sum_{p_i \in AT} |\eta(p_i)|}. \tag{10}$$

Moreover, the weight vector of attributes $\omega = (\omega(p_1), \omega(p_2), \ldots .\omega(p_m))^T$ has the following properties:

$$\omega(p) \geqslant 0,$$

$$\sum_{p_i \in AT} \omega(p_i) = |AT|.$$

A definition of weighted distance can be expressed based on the weight generation method.

Definition 1: Let $\mathcal{ODS} = (U, AT \bigcup D, F, G)$ be an ODS. For any $p \in AT$, $\forall a, b \in U$ the weighted distance is defined as follows:

$$dis_p(a,b) = |\omega(p) \times (f(a,p) - f(b,p))|. \tag{11}$$

A definition of fuzzy relationship for weighted distance is provided based on the definition of weighted distance, in conjunction with the concept of normalization. The values of the fuzzy relationship are precise, but the concept itself is fuzzy. The weighted distance fuzzy relation between a and b under p is obtained as follows:

$$\widetilde{W}_p(a,b) = \frac{dis_p(a,b) - \min(dis_p)}{\max(dis_p) - \min(dis_p)}. \tag{12}$$

And, the weighted distance fuzzy class of a under p is denoted as $[a]_{\widetilde{W}_p} = \sum_{b \in U} \frac{\widetilde{W}_p(a,b)}{b}$.

3.2. Dominance degree fuzzy relation

Furthermore, a fuzzy relationship of dominance degree is considered, which can describe the ranking relationship between two samples in a certain attribute. Its values are also within the normalized range.

Definition 2: Given an ODS, $\mathcal{ODS} = (U, AT \bigcup D, F, G)$, $\forall a, b \in U$ and for any $p \in AT$, the dominance degree fuzzy relation between a and b under p is defined as:

$$\widetilde{D}_p(a,b) = \begin{cases} (\frac{1}{1+e^{k\times(f(b,p)-f(a,p))}} - 0.5) \times 2, & (\frac{1}{1+e^{k\times(f(b,p)-f(a,p))}} - 0.5) \geq 0 \\ 0, & (\frac{1}{1+e^{k\times(f(b,p)-f(a,p))}} - 0.5) < 0 \end{cases}, \tag{13}$$

where k is a positive integer. The dominance degree fuzzy relation function can be defined differently depending on the actual circumstances. The dominating and dominated degree fuzzy classes of a under p is denoted as $[a]^{\geqslant}_{\widetilde{D}_p} = \sum_{b \in U} \frac{\widetilde{D}_p(a,b)}{b}$, $[a]^{\leqslant}_{\widetilde{D}_p} = \sum_{b \in U} \frac{\widetilde{D}_p(b,a)}{b}$. Denoting the dominance degree fuzzy class by $[a]_{\widetilde{D}_p}$, if a higher value of p is preferred, then $[a]_{\widetilde{D}_p} = [a]^{\geqslant}_{\widetilde{D}_p}$. And, if a lower value of p is preferred, then $[a]_{\widetilde{D}_p} = [a]^{\leqslant}_{\widetilde{D}_p}$.

3.3. Weighted dominance distance fuzzy class

The fuzzy relationship of weighted dominance distance has been obtained by combining the fuzzy relationships of weighted distance and dominance degree.

Definition 3: Given an ODS, $\mathcal{ODS} = (U, AT \bigcup D, F, G)$, $\forall a \in U$ and for any $p \in AT$, the weighted dominance distance fuzzy class of a under p is defined as follows:

$$[a]_{\widetilde{W \cap D_p}} = [a]_{\widetilde{W}_p} \cap [a]_{\widetilde{D}_p}. \tag{14}$$

3.4. The conditional entropy of the weighted dominance distance fuzzy relation

The decision dominance relationship between two samples is provided, which pertains to the classification of labels for the two samples. Based on the values, the superiority or inferiority between the two samples can be determined.

Definition 4: Given an ODS, $\mathcal{ODS} = (U, AT \bigcup D, F, G)$, $\forall a, b \in U$, the decision dominance relation of D is defined as follows:

$$R_D(a,b) = \begin{cases} 1, & g(b,d) \geq g(a,d) \\ 0, & g(b,d) < g(a,d) \end{cases}. \tag{15}$$

The dominating and dominated classes of a under D is denoted as $[a]^{\geqslant}_{R_D} = \sum_{b \in U} \frac{R_D(a,b)}{b}$, $[a]^{\leqslant}_{R_D} = \sum_{b \in U} \frac{R_D(b,a)}{b}$. Denoting the decision dominance class by $[a]_{R_D}$, if a higher value of D is preferred, then $[a]_{R_D} = [a]^{\geqslant}_{R_D}$. And, if a lower value of D is preferred, then $[a]_{R_D} = [a]^{\leqslant}_{R_D}$.

In the next, the definition of entropy is presented.

Definition 5: Given a MS-ODS, $\mathcal{MS} - \mathcal{ODS} = \{\mathcal{ODS}_i | \mathcal{ODS}_i = (U, AT \bigcup D, F, G), i = 1, 2, \ldots, s\}$, where $U = \{x_1, x_2, \ldots, x_n\}$. For any $p \in AT$, the information entropy of $\mathcal{ODS}_q$ for p under D is defined as follows:

$$E_p(D|\mathcal{ODS}_q) = -\frac{1}{|U|} \sum_{i=1}^{n} \log \frac{\left|[x_i]_{\widetilde{W \cap D_p}} \cap [x_i]_{R_D}\right|}{\left|[x_i]_{\widetilde{W \cap D_p}}\right|}. \tag{16}$$

Proposition 1. *The conditional entropy of the weighted dominance distance fuzzy relation has the following properties.*

(1) $E_p(D|\mathcal{ODS}_q) \geq 0$;
(2) $E_p(D|\mathcal{ODS}_q) < \infty$;
(3) $E_{p_1}(D|\mathcal{ODS}_q) \leq E_{p_2}(D|\mathcal{ODS}_q)$, *if* $\widetilde{W \cap D}_{p_1} \subseteq \widetilde{W \cap D}_{p_2}$.

Proof.

(1) Since $\frac{\left|[x_i]_{\widetilde{W \cap D_p}} \cap [x_i]_{R_D}\right|}{\left|[x_i]_{\widetilde{W \cap D_p}}\right|} \leq 1$, it follows that $\log \frac{\left|[x_i]_{\widetilde{W \cap D_p}} \cap [x_i]_{R_D}\right|}{\left|[x_i]_{\widetilde{W \cap D_p}}\right|} \leq 0$.
And $E_p(D|\mathcal{ODS}_q) = -\frac{1}{|U|} \sum_{i=1}^{n} \log \frac{\left|[x_i]_{\widetilde{W \cap D_p}} \cap [x_i]_{R_D}\right|}{\left|[x_i]_{\widetilde{W \cap D_p}}\right|}$, so the inequality is satisfied.

(2) Since $\frac{\left|[x_i]_{\widetilde{W \cap D_p}} \cap [x_i]_{R_D}\right|}{\left|[x_i]_{\widetilde{W \cap D_p}}\right|} \geq 0$, it follows that $\log \frac{\left|[x_i]_{\widetilde{W \cap D_p}} \cap [x_i]_{R_D}\right|}{\left|[x_i]_{\widetilde{W \cap D_p}}\right|} \geq -\infty$.
And $E_p(D|\mathcal{ODS}_q) = -\frac{1}{|U|} \sum_{i=1}^{n} \log \frac{\left|[x_i]_{\widetilde{W \cap D_p}} \cap [x_i]_{R_D}\right|}{\left|[x_i]_{\widetilde{W \cap D_p}}\right|}$, then the inequality is true.

(3) Given $f(x,y) = -\log \frac{x}{x+y}$, then $\frac{\partial f}{\partial x} = -\frac{y}{x(x+y)}$ and $\frac{\partial f}{\partial y} = \frac{1}{x+y}$.
When $x, y > 0$, $\frac{\partial f}{\partial x} < 0$ and $\frac{\partial f}{\partial y} > 0$ can be hold.
For $\forall x_i \in U$, $[x_i]_{\widetilde{W \cap D_p}} \cap [x_i]_{R_D} = ([x_i]_{\widetilde{W \cap D_p}} \cap [x_i]_{R_D}) \cup ([x_i]_{\widetilde{W \cap D_p}} \cap ([x_i]_{R_D})^C)$,
then $\left|[x_i]_{\widetilde{W \cap D_p}} \cap [x_i]_{R_D}\right| = \left|[x_i]_{\widetilde{W \cap D_p}} \cap [x_i]_{R_D}\right| + \left|[x_i]_{\widetilde{W \cap D_p}} \cap ([x_i]_{R_D})^C\right|$.

$$-\log \frac{\left|[x_i]_{\widetilde{W \cap D_p}} \cap [x_i]_{R_D}\right|}{\left|[x_i]_{\widetilde{W \cap D_p}}\right|} = -\log \frac{\left|[x_i]_{\widetilde{W \cap D_p}} \cap [x_i]_{R_D}\right|}{\left|[x_i]_{\widetilde{W \cap D_p}} \cap [x_i]_{R_D}\right| + \left|[x_i]_{\widetilde{W \cap D_p}} \cap ([x_i]_{R_D})^C\right|}$$

can be get.

By substituting $\left|[x_i]_{\widetilde{W \cap D_p}} \cap [x_i]_{R_D}\right|$ with x and $\left|[x_i]_{\widetilde{W \cap D_p}} \cap ([x_i]_{R_D})^C\right|$ with y,

if $\widetilde{W \cap D}_{p_1} \subseteq \widetilde{W \cap D}_{p_2}$,

then $\left|[x_i]_{\widetilde{W \cap D_{p_1}}} \cap [x_i]_{R_D}\right| \le \left|[x_i]_{\widetilde{W \cap D_{p_2}}} \cap [x_i]_{R_D}\right|$ and $\left|[x_i]_{\widetilde{W \cap D_{p_1}}} \cap ([x_i]_{R_D})^C\right| \le \left|[x_i]_{\widetilde{W \cap D_{p_2}}} \cap ([x_i]_{R_D})^C\right|$.

Thus, $E_{p_1}(D|\mathcal{ODS}_q) \le E_{p_2}(D|\mathcal{ODS}_q)$.

□

3.5. Static fusion algorithm

Given a MS-ODS, $\mathcal{MS}-\mathcal{ODS} = \{\mathcal{ODS}_i | \mathcal{ODS}_i = (U, AT \bigcup D, F, G), i = 1, 2, \dots, s\}$, The ith source which is the most essential for p under the conditional entropy of weighted dominance distance fuzzy relation can be obtained by

$$i_p = \arg\min_{q \in 1,2,\dots,s} E_p(D|\mathcal{ODS}_q). \quad (17)$$

By above, the source with the lowest entropy for each attribute can be identified, and it can be used to generate a new information system. The fusion algorithm is shown in Algorithm 1, and the figure is shown in Fig. 2.

Algorithm 1: The conditional entropy-based fusion algorithm using the weighted dominance distance fuzzy relation.

Input : $\mathcal{MS}-\mathcal{ODS} = \{\mathcal{ODS}_i | \mathcal{ODS}_i = (U, AT \bigcup D, F, G), i = 1, 2, \dots, s\}$.
Output : A new fusion result.

begin
  for $q = 1 : s$ **do**
    for *each* $p \in AT$ **do**
      compute $\omega(p)$
    end
  end
  for $q = 1 : s$ **do**
    obtain $E_p(D|\mathcal{ODS}_q) \leftarrow$
    for *each* $p \in AT$ **do**
      for $i = 1 : |U|$ **do**
        get $[x_i]_{\widetilde{W_p}}$ and $[x_i]_{\widetilde{D_p}}$ and $[x_i]_{R_D}$,then compute $[x_i]_{\widetilde{W \cap D_p}}$
      end
    end
  end
  for *each* $p \in AT$ **do**
    find $i_p = \arg\min_{q \in 1,2,\dots,s} E_p(D|\mathcal{ODS}_q)$
  end
end
return : $(F_{p_1}^{i_{p_1}}, F_{p_2}^{i_{p_2}}, \dots, F_{|AT|}^{i_{|AT|}})$

The algorithm takes a MS-ODS as input and outputs a new fusion result. The main steps of the algorithm include calculating attribute weights, obtaining conditional entropy, and selecting the optimal fusion result. We analyze its time and space complexity, as shown in Table 3.

Example 1: In this example, a process to find the result of static fusion will be shown. A MS-ODS is given, as shown in Table 4. This MS-ODS represents the grades of seven students in four exams for four courses, with d indicating the classification of excellence level for each student. In this system, each attribute and d are considered better if its value is higher. Starting from $\mathcal{ODS}_1$, the weight vector under $\mathcal{ODS}_1$ is calculated as

$$\omega = (1.093872, 2.320957, 0.515018, 0.070153)^T,$$

then compute the weighted distance of each sample under the attribute, such as $dis_{a_1}(x_1, x_1) = 0.0$, $dis_{a_1}(x_1, x_2) = 10.93872, \dots$, $dis_{a_1}(x_1, x_7) = 13.126463$.

Therefore, the weighted distance fuzzy relation matrix of ODS_1 samples under a_1 can be calculated as

$$M_{\widetilde{W_{a_1}}} = \begin{bmatrix} 0.000000 & 0.222222 & 0.733333 & 0.022222 & 0.133333 & 0.244444 & 0.266667 \\ 0.222222 & 0.000000 & 0.955556 & 0.244444 & 0.088889 & 0.022222 & 0.044444 \\ 0.733333 & 0.955556 & 0.000000 & 0.711111 & 0.866667 & 0.977778 & 1.000000 \\ 0.022222 & 0.244444 & 0.711111 & 0.000000 & 0.155556 & 0.266667 & 0.288889 \\ 0.133333 & 0.088889 & 0.866667 & 0.155556 & 0.000000 & 0.111111 & 0.133333 \\ 0.244444 & 0.022222 & 0.977778 & 0.266667 & 0.111111 & 0.000000 & 0.022222 \\ 0.266667 & 0.044444 & 1.000000 & 0.288889 & 0.133333 & 0.022222 & 0.000000 \end{bmatrix}.$$

And from this matrix we get, $[x_1]_{\widetilde{W_{a_1}}} = \frac{0.222222}{x_2} + \frac{0.733333}{x_3} + \frac{0.022222}{x_4} + \frac{0.133333}{x_5} + \frac{0.244444}{x_6} + \frac{0.266667}{x_7}$, $[x_2]_{\widetilde{W_{a_1}}} = \frac{0.222222}{x_1} + \frac{0.955556}{x_3} + \frac{0.244444}{x_4} + \frac{0.088889}{x_5} + \frac{0.022222}{x_6} + \frac{0.044444}{x_7}$, …, $[x_7]_{\widetilde{W_{a_1}}} = \frac{0.266667}{x_1} + \frac{0.044444}{x_2} + \frac{1.000000}{x_3} + \frac{0.288889}{x_4} + \frac{0.133333}{x_5} + \frac{0.022222}{x_6}$.

The matrix of the dominance degree fuzzy relation under a_1 can be obtained by definition:

$$M_{\widetilde{D_{a_1}}} = \begin{bmatrix} 0.000000 & 0.000000 & 1.000000 & 0.462117 & 0.000000 & 0.000000 & 0.000000 \\ 0.999909 & 0.000000 & 1.000000 & 0.999967 & 0.964028 & 0.000000 & 0.000000 \\ 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 \\ 0.000000 & 0.000000 & 1.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 \\ 0.995055 & 0.000000 & 1.000000 & 0.998178 & 0.000000 & 0.000000 & 0.000000 \\ 0.999967 & 0.462117 & 1.000000 & 0.999988 & 0.986614 & 0.000000 & 0.000000 \\ 0.999988 & 0.761594 & 1.000000 & 0.999995 & 0.995055 & 0.462117 & 0.000000 \end{bmatrix}.$$

And by this matrix, we have $[x_1]_{\widetilde{D_{a_1}}} = \frac{1}{x_3} + \frac{0.462117}{x_4}$, $[x_2]_{\widetilde{D_{a_1}}} = \frac{0.999909}{x_1} + \frac{1}{x_3} + \frac{0.999967}{x_4} + \frac{0.964028}{x_5}$, …, $[x_7]_{\widetilde{D_{a_1}}} = \frac{0.999988}{x_1} + \frac{0.761594}{x_2} + \frac{1}{x_3} + \frac{0.999995}{x_4} + \frac{0.995055}{x_5} + \frac{0.462117}{x_6}$.

Then we can get, $[x_1]_{\widetilde{W \cap D_{a_1}}} = \frac{0.733333}{x_3} + \frac{0.022222}{x_4}$, $[x_1]_{\widetilde{W \cap D_{a_1}}} = \frac{0.222222}{x_1} + \frac{0.955556}{x_3} + \frac{0.244444}{x_4} + \frac{0.088889}{x_5}$, …, $[x_7]_{\widetilde{W \cap D_{a_1}}} = \frac{0.266667}{x_1} + \frac{0.044444}{x_2} + \frac{1}{x_3} + \frac{0.288889}{x_4} + \frac{0.133333}{x_5} + \frac{0.022222}{x_6}$.

According to decision matrix

$$M_{R_D} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix},$$

$H_{a_1}(D|\mathcal{ODS}_1) = 8.805261$ can be gain. Similarly, the information entropy of all sources is explored for each attribute as shown in Table 5. The result of the fusion is the underline attribute column in the Table 4.

4. The incremental learning mechanism in the fusion process of MS-ODS

This section studies four mechanisms and their corresponding algorithms for incremental fusion that adapt to changes in information sources and attributes. Four cases of data changes are investigated, namely:

(a) Inserting new significant information sources into the data while removing some dispensable attributes.
(b) Simultaneously inserting new significant information sources and pivotal attributes into the data.
(c) Removing dispensable information sources and redundant attributes from the data.

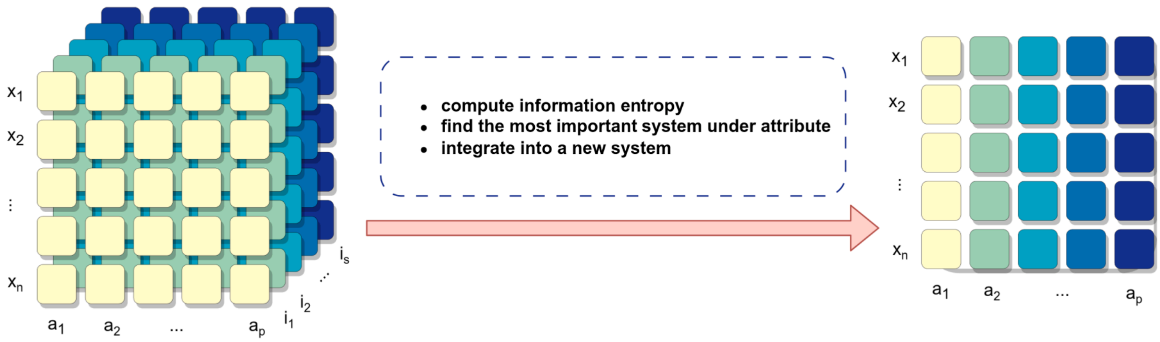

Fig. 2. The fusion process of the proposed method.

Table 3
Complexity analysis of the static fusion algorithm (Algorithm 1).

Time complexity	Cyclic analysis	In Algorithm 1, steps from 2 to 6 traverse each ODS (s in total) and each attribute ($\|AT\|$ in total) and calculate the weight $\omega(p)$, the time complexity of this part is $O(s \times \|AT\|)$; steps from 7 to 14 first traverse each ODS and then operate on each attribute and each object in U, which has a complexity of $O(s \times \|AT\| \times \|U\|)$; steps from 15 to 17 find the smallest $E_p(D\|\mathcal{O}DS_q)$ for each attribute, which is $O(\|AT\| \times s)$.
	Overall time complexity	Therefore, the time complexity of the whole algorithm is rough $O(s \times \|AT\| \times \|U\|)$, assuming that $\|U\|$ (the number of objects) is the dominant factor.
Space complexity	Static space requirement	It mainly includes the input MS-ODS and the variables used inside the algorithm and the space complexity is $O(1)$.
	Dynamic space requirement	To reduce the running time in practice, it mainly uses numpy for two-dimensional operations in code. The time complexity is calculated separately for each part. (1) The space complexity of ODS weights is $O(\|U\| \times \|AT\| + \|AT\|^2)$, depending on the input array size and linear transformation. (2) The space complexity of the weight distance fuzzy relation is $O(\|AT\| \times \|U\|^2)$. For each attribute, there is a distance matrix between samples. (3) The space complexity of the dominance degree fuzzy relation is $O(\|AT\| \times \|U\|^2)$. For each attribute, there is a dominant matrix between samples. (4) The space complexity of the weighted dominance distance fuzzy relation is $O(\|AT\| \times \|U\|^2)$, which mainly depends on the fusion of the dominant relationship and the distance relationship. (5) The space complexity of the conditional entropy of the weighted dominance distance fuzzy relation is $O(\|AT\| \times \|U\|^2)$, depending on the calculation of the dominant distance relationship.
	Overall space complexity	Thus, the space complexity is roughly $O(s \times \|AT\| \times \|U\|^2)$, which takes into account storing information about each object under the information source for each attribute.

Table 4
A example of MS-ODS.

	$\mathcal{O}DS_1$				$\mathcal{O}DS_2$				$\mathcal{O}DS_3$				$\mathcal{O}DS_4$				
	a_1	a_2	a_3	a_4	a_1	a_2	a_3	a_4	a_1	a_2	a_3	a_4	a_1	a_2	a_3	a_4	d
x_1	79	68	78	45	69	71	78	40	70	71	75	55	76	71	75	45	1
x_2	89	70	70	57	89	70	70	57	89	70	70	57	81	72	80	47	1
x_3	46	57	90	66	49	50	91	60	40	69	90	60	59	75	95	49	1
x_4	78	80	75	91	78	80	75	91	78	75	70	94	74	75	71	82	2
x_5	85	85	69	89	80	83	75	79	80	82	78	79	89	80	70	69	2
x_6	90	90	95	95	90	91	95	95	95	86	90	91	95	96	97	82	3
x_7	91	92	92	92	85	90	92	98	80	80	97	91	89	89	91	89	3

Table 5
The information entropy for Example 1.

	a_1	a_2	a_3	a_4
$\mathcal{O}DS_1$	8.805261	8.822688	5.844748	9.001521
$\mathcal{O}DS_2$	8.749065	8.985611	11.010392	8.806077
$\mathcal{O}DS_3$	8.671777	**8.473404**	8.410354	10.787628
$\mathcal{O}DS_4$	**8.510980**	8.534104	**5.664968**	**8.656720**

(d) Removing dispensable information sources from the data and inserting new crucial attributes.

The updating process of the data is shown in Fig. 3, and the fusion processes of four changing cases are shown in Fig. 4.

Case (a): Inserting new significant information sources into the data while removing some dispensable attributes.

Suppose there is a MS-DOS at time T_1, and the incremental fusion algorithm is introduced below to add some information sources and reduce some attributes at time T_2. Let $\{\mathcal{O}DS_j, j = m+1, m+2, \dots, m+\Delta m\}$ be the set of added information sources and $\{a_i, i = n+1, n+2, \dots, n+\Delta n\}$ be the set of reduced attributes. The algorithm is expressed in Algorithm 2, the complexity analysis of the algorithm is in Table 9, and has the following proposition:

Proposition 2. *For $\{a_1, a_2, \dots, a_n\}$, the following properties are true.*

(1) If $\min_{q\in m+1,m+2,\dots,m+\Delta m} E_p(D|\mathcal{O}DS_q) \geq \min_{q\in 1,2,\dots,m} E_p(D|\mathcal{O}DS_q)$, then $F_a^{T_2} = F_a^{T_1}$.

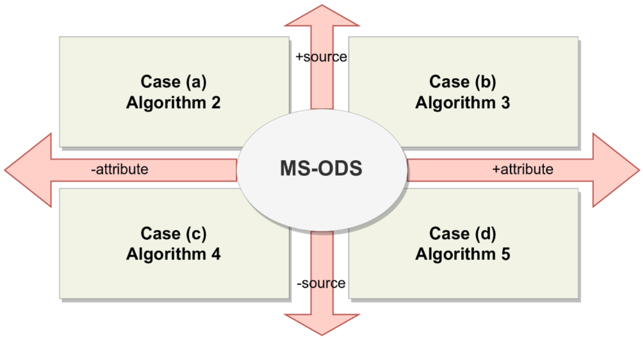

Fig. 3. The updating processes in four different scenarios.

Table 6
The updated entropy at time T_2 in Example 2.

	a_1	a_2
$\mathcal{ODS}_1$	8.805261	8.822688
$\mathcal{ODS}_2$	8.749065	8.985611
$\mathcal{ODS}_3$	8.671777	**8.473404**
$\mathcal{ODS}_4$	**8.510980**	8.534104

Table 7
The initial entropy for at time T_1 in Example 2.

	a_1	a_2	a_3	a_4
$\mathcal{ODS}_1$	8.805261	**8.822688**	**5.844748**	9.001521
$\mathcal{ODS}_2$	**8.749065**	8.985611	11.010392	**8.806077**

(2) If $\min_{q\in m+1,m+2,\dots,m+\Delta m} E_p(D|\mathcal{ODS}_q) < \min_{q\in 1,2,\dots,m} E_p(D|\mathcal{ODS}_q)$, *then* $F_a^{T_2} = F_a^{i_a}$,
where $i_a = \arg\min_{q\in m+1,m+2,\dots,m+\Delta m} E_p(D|\mathcal{ODS}_q)$.

Proof.

(1) If $\min_{q\in m+1,m+2,\dots,m+\Delta m} E_p(D|\mathcal{ODS}_q) \geq \min_{q\in 1,2,\dots,m} E_p(D|\mathcal{ODS}_q)$, then $\min_{q\in 1,2,\dots,m+\Delta m} E_p(D|\mathcal{ODS}_q) = \min_{q\in 1,2,\dots,m} E_p(D|\mathcal{ODS}_q)$ can be get, so we have $F_a^{T_2} = F_a^{T_1}$.
(2) If $\min_{q\in m+1,m+2,\dots,m+\Delta m} E_p(D|\mathcal{ODS}_q) < \min_{q\in 1,2,\dots,m} E_p(D|\mathcal{ODS}_q)$, then $\min_{q\in 1,2,\dots,m+\Delta m} E_p(D|\mathcal{ODS}_q) = \min_{q\in m+1,m+2,\dots,m+\Delta m} E_p(D|\mathcal{ODS}_q)$,
$i_a = \arg\min_{q\in m+1,m+2,\dots,m+\Delta m} E_p(D|\mathcal{ODS}_q)$.
Thus, we have $F_a^{T_2} = F_a^{i_a}$.

According to Proposition 2, the fusion table can be incrementally updated when new important information source data is added and there is a need to abandon unimportant conditional attributes. □

Example 1. (Continued from Example 1) Suppose that the data from the latter two ODSs in Table 4 are newly added important information sources, and only information on the first two important attributes is needed for all samples. This means that the scores of two additional exams need to be considered, and courses a_3 and a_4 should be excluded from the assessment to evaluate the students' excellence. At time T_1, we provide the first two ODSs with attributes a_1 to a_4, and the conditional entropy of each attribute for each information source at time T_1 is given in Table 7 as our initial fusion result. In Table 6, the updated entropy values for the dynamically updated data are provided.

Case (b): Simultaneously inserting new significant information sources and pivotal attributes into the data.

Suppose there is a MS-ODS at time T_1, and the incremental fusion algorithm is introduced below to add some attributes and information sources at time T_2. Let $\{\mathcal{ODS}_j, j = m+1, m+2, \dots, m+\Delta m\}$ be the set of added information sources and $\{a_i, i = n+1, n+2, \dots, n+\Delta n\}$ be the set of added attributes. The algorithm is expressed in Algorithm 3, the complexity analysis of the algorithm is in Table 10, and has the following proposition:

Algorithm 2: The dynamic fusion algorithm of inserting new information sources and removing existing attributes.

Input :
1. Original fusion result $(F_{a_1}^{T_1}, F_{a_2}^{T_1}, \dots, F_{a_{n+\Delta n}}^{T_1})$;
2. The information entropy set $\{H_{a_i}(D|\mathcal{ODS}_j), i = 1, 2, \dots, n+\Delta n, j = 1, 2, \dots, m\}$;
3. Deleted attribute set $\{a_i, i = n+1, n+2, \dots, n+\Delta n\}$;
4. Inserted source set $\{\mathcal{ODS}_j, j = m+1, m+2, \dots, m+\Delta m\}$.

Output : An updated result.

```
begin
    for q = m+1 : m+Δm do
        for each p ∈ {a_1, a_2, ..., a_n} do
            get E_p(D|ODS_q)
        end
    end
    for each p ∈ {a_1, a_2, ..., a_n} do
        if min_{q∈{m+1,m+2,...,m+Δm}} E_p(D|ODS_q) ≥
          min_{q∈{1,2,...,m}} E_p(D|ODS_q) then
            F_a^{T_2} = F_a^{T_1}
        else
            F_a^{T_2} = F_a^{i_a}(i_a =
              arg min_{q∈m+1,m+2,...,m+Δm} E_p(D|ODS_q))
        end
    end
end
return : (F_{a_1}^{T_2}, F_{a_2}^{T_2}, ..., F_{a_n}^{T_2})
```

Proposition 3. *For* $\{a_1, a_2, \dots, a_n\}$*, the following properties are true:*

(1) If $\min_{q\in m+1,m+2,\dots,m+\Delta m} E_p(D|\mathcal{ODS}_q) \geq \min_{q\in 1,2,\dots,m} E_p(D|\mathcal{ODS}_q)$, *then* $F_a^{T_2} = F_a^{T_1}$.
(2) If $\min_{q\in m+1,m+2,\dots,m+\Delta m} E_p(D|\mathcal{ODS}_q) < \min_{q\in 1,2,\dots,m} E_p(D|\mathcal{ODS}_q)$, *then* $F_a^{T_2} = F_a^{i_a}$,
where $i_a = \arg\min_{q\in m+1,m+2,\dots,m+\Delta m} E_p(D|\mathcal{ODS}_q)$.
(3) For $a \in \{a_i, i = n+1, n+2, \dots, n+\Delta n\}$,
we have $F_a^{T_2} = F_a^{i_a}$ *where* $i_a = \arg\min_{q\in 1,2,\dots,m+\Delta m} E_p(D|\mathcal{ODS}_q)$.

Proof:

(1) Proposition 3(1) is similar to Proposition 2(1), so its proof is similar to the proof of Proposition 2(1).
(2) Proposition 3(2) is similar to Proposition 2(2), so its proof is similar to the proof of Proposition 2(2).
(3) According to Definition 7, it can be obtained.

According to Proposition 3, the fusion table can be incrementally updated when new important information sources and attributes are added.

Example 2. (Continued from Example 1) It is assumed that the data from the latter two ODSs in Table 4 and the last two attributes are newly added important information sources and attributes. This implies that the scores of two additional exams need to be taken into account, and evaluations for all courses are needed to assess the students' excellence. At time T_1, the first two ODSs with attributes a_1 and a_2 are provided, and the initial fusion result is obtained by presenting the initial conditional entropy of each attribute for each information source at time T_1 in Table 8. It can be observed that the entropy values of the dynamically updated data remain the same as those in Table 5.

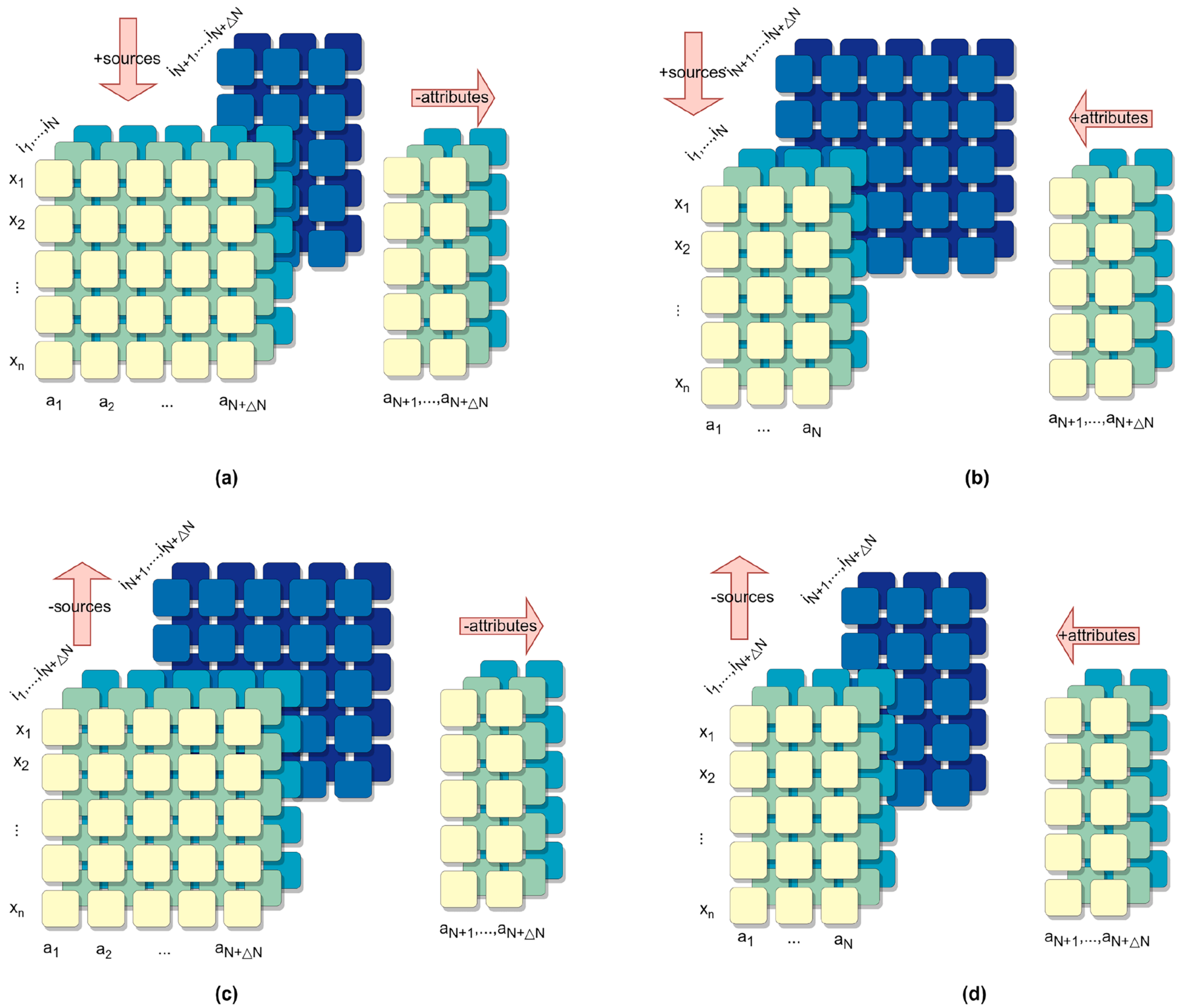

Fig. 4. The dynamic fusion processes of four changing cases.

Table 8
The initial entropy at time T_1 in Example 3.

	a_1	a_2
$\mathcal{O}DS_1$	8.805261	**8.822688**
$\mathcal{O}DS_2$	**8.749065**	8.985611

Case (c): Removing dispensable information sources and redundant attributes from the data.

Suppose there is a MS-ODS at time T_1, and the incremental fusion algorithm is introduced below to delete some attributes and information sources at time T_2. Let $\{\mathcal{O}DS_j, j = m+1, m+2, \ldots, m+\Delta m\}$ be the set of deleted information sources and $\{a_i, i = n+1, n+2, \ldots, n+\Delta n\}$ be the set of deleted attributes. The algorithm is expressed in Algorithm 4, the complexity analysis of the algorithm is in Table 11, and has the following proposition:

Proposition 4. *For* $\{a_1, a_2, \ldots, a_n\}$*, the following properties are true:*

(1) If $\min_{q\in m+1,m+2,\ldots,m+\Delta m} E_p(D|\mathcal{O}DS_q) \geq \min_{q\in 1,2,\ldots,m} E_p(D|\mathcal{O}DS_q)$*, then* $F_a^{T_2} = F_a^{T_1}$.

(2) If $\min_{q\in m+1,m+2,\ldots,m+\Delta m} E_p(D|\mathcal{O}DS_q) < \min_{q\in 1,2,\ldots,m} E_p(D|\mathcal{O}DS_q)$*, then* $F_a^{T_2} = F_a^{i_a}$,
where $i_a = \arg\min_{q\in 1,2,\ldots,m+\Delta m} E_p(D|\mathcal{O}DS_q)$.

Proof.

(1) Proposition 4(1) is similar to Proposition 2(1), so its proof is similar to the proof of Proposition 2(1).
(2) Proposition 4(2) is similar to Proposition 2(2), so its proof is similar to the proof of Proposition 2(2).

According to Proposition 4, the fusion table can be incrementally updated when information sources are deleted and conditional attributes are deleted, too. □

Example 3. (Continued from Example 1) Suppose that the data from the latter two ODSs in Table 4 and the last two attributes are invalid information sources and attributes that do not need to be considered. This means that the scores of the last two exams should be excluded, and only the first two courses need to be evaluated to assess the students' excellence. At time T_1, we provide all of the ODSs, including attributes a_1 to a_4. In Table 5, we can observe the initial conditional entropy of

Table 9
Complexity analysis of case (a) (Algorithm 2).

Time complexity	Cyclic analysis	In Algorithm 2, Steps from 2 to 6 loop through the new sources (Δm in total) and each attribute (n in total) and calculate the corresponding information entropy $E_p(D\|\mathcal{ODS}_q)$, the time complexity of this part is $O(\Delta m \times n \times \|U\|)$. Steps from 7 to 14 loops through the original attribute from a_1 to a_n, and compares the minimum entropy of the newly added source with the minimum entropy of the original attribute, the time complexity is $O(n)$.
	Overall time complexity	Therefore, The time complexity of this part is $O(\Delta m \times n \times \|U\|)$.
Space complexity	Static space requirement	It mainly includes all inputs in Algorithm 2 and the variables used inside the algorithm, and the space complexity is $O(1)$.
	Dynamic space requirement	To reduce the running time in practice, it mainly uses a numpy array for two-dimensional operation. The time complexity is calculated separately for each part. The space complexity of steps from 2 to 6 is about $O(\Delta m \times n \times \|U\|^2)$, depending on the calculation of the dominant and distance relationships and sources.
	Overall space complexity	Thus, the space complexity is roughly $O(\Delta m \times n \times \|U\|^2)$.

Algorithm 3: The dynamic fusion algorithm of inserting both new sources and attributes.

Input :
1. Original fusion result $(F_{a_1}^{T_1}, F_{a_2}^{T_1}, \ldots, F_{a_n}^{T_1})$;
2. The information entropy set $\{H_{a_i}(D|\mathcal{ODS}_j), i = 1, 2, \ldots, n, j = 1, 2, \ldots, m\}$;
3. Inserted attribute set $\{a_i, i = n+1, n+2, \ldots, n+\Delta n\}$;
4. Inserted source set $\{\mathcal{ODS}_j, j = m+1, m+2, \ldots, m+\Delta m\}$.

Output : An updated result.

begin
  for $q = m+1 : m+\Delta m$ **do**
    for *each* $p \in \{a_1, a_2, \ldots, a_{n+\Delta n}\}$ **do**
      get $E_p(D|\mathcal{ODS}_q)$
    end
  end
  for $q = 1 : m$ **do**
    for *each* $p \in \{a_{n+1}, a_{n+2}, \ldots, a_{n+\Delta n}\}$ **do**
      get $E_p(D|\mathcal{ODS}_q)$
    end
  end
  for *each* $p \in \{a_1, a_2, \ldots, a_n\}$ **do**
    if $\min_{q\in\{m+1,m+2,\ldots,m+\Delta m\}} E_p(D|\mathcal{ODS}_q) \geq \min_{q\in\{1,2,\ldots,m\}} E_p(D|\mathcal{ODS}_q)$ **then**
      $F_a^{T_2} = F_a^{T_1}$
    else
      $F_a^{T_2} = F_a^{i_a}(i_a = \arg\min_{q\in m+1,m+2,\ldots,m+\Delta m} E_p(D|\mathcal{ODS}_q))$
    end
  end
  for *each* $p \in \{a_{n+1}, a_{n+2}, \ldots, a_{n+\Delta n}\}$ **do**
    $F_a^{T_2} = F_a^{i_a}(i_a = \arg\min_{q\in 1,2,\ldots,m+\Delta m} E_p(D|\mathcal{ODS}_q))$
  end
end
return : $(F_{a_1}^{T_2}, F_{a_2}^{T_2}, \ldots, F_{a_n}^{T_2}, F_{a_{n+1}}^{i_{a_{n+1}}}, F_{a_{n+2}}^{i_{a_{n+2}}}, \ldots, F_{a_{n+\Delta n}}^{i_{a_{n+\Delta n}}})$

each attribute for each information source at time T_1, which serves as our initial fusion result. It can be observed that the entropy values of the dynamically updated data are the same as those in Table 8.

Case (d): Removing dispensable information sources from the data and inserting new crucial attributes.

Suppose there is a MS-DOS at time T_1, and the incremental fusion algorithm is introduced below to reduce some information sources and add some attributes at time T_2. Let $\{\mathcal{ODS}_j, j = m+1, m+2, \ldots, m+\Delta m\}$ be the set of deleted information sources and

Algorithm 4: The dynamic fusion algorithm of removing both sources and attributes.

Input :
1. Original fusion result $(F_{a_1}^{T_1}, F_{a_2}^{T_1}, \ldots, F_{a_{n+\Delta n}}^{T_1})$;
2. The information entropy set $\{H_{a_i}(D|\mathcal{ODS}_j), i = 1, 2, \ldots, n+\Delta n, j = 1, 2, \ldots, m+\Delta m\}$;
3. Deleted attribute set $\{a_i, i = n+1, n+2, \ldots, n+\Delta n\}$;
4. Deleted source set $\{\mathcal{ODS}_j, j = m+1, m+2, \ldots, m+\Delta m\}$.

Output : An updated result.

begin
  for *each* $p \in \{a_1, a_2, \ldots, a_n\}$ **do**
    if $\min_{q\in\{m+1,m+2,\ldots,m+\Delta m\}} E_p(D|\mathcal{ODS}_q) \geq \min_{q\in\{1,2,\ldots,m\}} E_p(D|\mathcal{ODS}_q)$ **then**
      $F_a^{T_2} = F_a^{T_1}$
    else
      $F_a^{T_2} = F_a^{i_a}(i_a = \arg\min_{q\in 1,2,\ldots,m} E_p(D|\mathcal{ODS}_q))$
    end
  end
end
return : $(F_{a_1}^{T_2}, F_{a_2}^{T_2}, \ldots, F_{a_n}^{T_2})$

$\{a_i, i = n+1, n+2, \ldots, n+\Delta n\}$ be the set of added attributes. The algorithm is expressed in Algorithm 5, the complexity analysis of the algorithm is in Table 12, and has the following proposition:

Proposition 5. *For* $\{a_1, a_2, \ldots, a_n\}$*, the following properties are true:*

(1) If $\min_{q\in m+1,m+2,\ldots,m+\Delta m} E_p(D|\mathcal{ODS}_q) \geq \min_{q\in 1,2,\ldots,m} E_p(D|\mathcal{ODS}_q)$*, then* $F_a^{T_2} = F_a^{T_1}$.
(2) If $\min_{q\in m+1,m+2,\ldots,m+\Delta m} E_p(D|\mathcal{ODS}_q) < \min_{q\in 1,2,\ldots,m} E_p(D|\mathcal{ODS}_q)$*, then* $F_a^{T_2} = F_a^{i_a}$*,*
where $i_a = \arg\min_{q\in m+1,m+2,\ldots,m+\Delta m} E_p(D|\mathcal{ODS}_q)$.
(3) For $a \in \{a_i, i = n+1, n+2, \ldots, n+\Delta n\}$*,*
we have $F_a^{T_2} = F_a^{i_a}$ *where* $i_a = \arg\min_{q\in 1,2,\ldots,m} E_p(D|\mathcal{ODS}_q)$.

Proof.

(1) Proposition 5(1) is similar to Proposition 2(1), so its proof is similar to the proof of Proposition 2(1).
(2) Proposition 5(2) is similar to Proposition 2(2), so its proof is similar to the proof of Proposition 2(2).
(3) According to Definition 7, it can be obtained.

The fusion table can be incrementally updated when sources are deleted and valid conditional attributes need to be inserted, according to Proposition 5. □

Table 10
Complexity analysis of case (b) (Algorithm 3).

Time complexity	Cyclic analysis	In Algorithm 3, steps from 2 to 6 loop through the new source (Δ in total) and each attribute ($n+\Delta n$ in total) and compute the corresponding information entropy $E_p(D\|\mathcal{ODS}_q)$, the time complexity of this part is $O(\Delta m\times(n+\Delta n)\times\|U\|)$. Steps from 7 to 11 traverse the attribute from $a_{n+1}$to $a_{n+\Delta n}$, and calculate the information entropy $E_p(D\|\mathcal{ODS}_q)$ of the new attribute of the original source, the time complexity of this part is $O(m\times\Delta n\times\|U\|)$. The complexity of Steps from 12 to 18 is $O(n)$. The time complexity of steps from 19 to 21 is $O(\Delta n)$.
	Overall time complexity	Therefore, The time complexity of this part is $O(\Delta m\times(n+\Delta n)\times\|U\|+m\times\Delta n\times\|U\|)$.
Space complexity	Static space requirement	It mainly includes all inputs in Algorithm 3 and the variables used inside the algorithm, and the space complexity is $O(1)$.
	Dynamic space requirement	To reduce the running time in practice, it mainly uses a numpy array for two-dimensional operation. The time complexity is calculated separately for each part. The space complexity of steps from 2 to 6 is about $O(\Delta m\times(n+\Delta n)\times\|U\|^2)$, depending on the calculation of the dominant and distance relationships and sources. The space complexity of steps from 7 to 11 is about $O(m\times\Delta n\times\|U\|^2)$, depending on the calculation of the dominant and distance relationships and sources.
	Overall space complexity	Thus, the space complexity is roughly $O(\Delta m\times(n+\Delta n)\times\|U\|^2+m\times\Delta n\times\|U\|^2)$.

Table 11
Complexity analysis of case (c) (Algorithm 4).

Time complexity	Cyclic analysis	In Algorithm 4, steps from 2 to 8 the time complexity of this part is $O(n)$.
	Overall time complexity	Therefore, The time complexity of this part is $O(n)$.
Space complexity	Static space requirement	It mainly includes all inputs in Algorithm 4 and the variables used inside the algorithm, and the space complexity is $O(1)$.
	Dynamic space requirement	To reduce the running time in practice, it mainly uses a numpy array for two-dimensional operation. The time complexity is calculated separately for each part. The space complexity of steps from 2 to 8 is about $O(1)$.
	Overall space complexity	Thus, the space complexity is roughly $O(1)$.

Algorithm 5: The dynamic fusion algorithm of removing sources and inserting new attributes.

Input :
1. Original fusion result $(F_{a_1}^{T_1}, F_{a_2}^{T_1}, \dots, F_{a_n}^{T_1})$;
2. The information entropy set $\{H_{a_i}(D|\mathcal{ODS}_j), i=1,2,\dots,n, j=1,2,\dots,m+\Delta m\}$;
3. Inserted attribute set $\{a_i, i=n+1,n+2,\dots,n+\Delta n\}$;
4. Deleted source set $\{\mathcal{ODS}_j, j=m+1,m+2,\dots,m+\Delta m\}$.

Output : An updated result.

```
begin
    for q = 1 : m do
        for each p ∈ {a_{n+1}, a_{n+2}, ..., a_{n+Δn}} do
            compute E_p(D|ODS_q)
        end
    end
    for each p ∈ {a_1, a_2, ..., a_n} do
        if min_{q∈{m+1,m+2,...,m+Δm}} E_p(D|ODS_q) ≥ min_{q∈{1,2,...,m}} E_p(D|ODS_q) then
            F_a^{T_2} = F_a^{T_1}
        else
            F_a^{T_2} = F_a^{i_a} (i_a = arg min_{q∈1,2,...,m} E_p(D|ODS_q))
        end
    end
    for each p ∈ {a_{n+1}, a_{n+2}, ..., a_{n+Δn}} do
        compute i_a = arg min_{q∈1,2,...,m} E_p(D|ODS_q)
    end
end
return : (F_{a_1}^{T_2}, F_{a_2}^{T_2}, ..., F_{a_n}^{T_2}, F_{a_{n+1}}^{i_{a_{n+1}}}, F_{a_{n+2}}^{i_{a_{n+2}}}, ..., F_{a_{n+Δn}}^{i_{a_{n+Δn}}})
```

Example 4. (Continued from Example 1) Suppose that the data from the latter two ODSs in Table 4 are invalid information sources that do not need to be considered, and the last two attributes are important attributes that need to be taken into account. This means that the scores of the last two exams should be excluded, and evaluations for all four courses are required to assess the students' excellence. At time T_1, all of the ODSs are provided, including attributes a1 and a2. In Table 6, initial conditional entropy of each attribute for each information source at time T_1 can be observed, which serves as our initial fusion result. It can be noted that the entropy values of the dynamically updated data are the same as those in Table 7.

Moreover, Tables 13 and 14 compare time and space complexity between the static and four incremental algorithms. Based on the highest order term of the complexity, we can observe that the four incremental algorithms are more efficient than the static algorithm in both time and space.

5. Experimental analysis

To evaluate the efficiency of the four proposed method, a series of comparative experiments are performed on UCI and KEEL-related datasets. Nine datasets were selected for comparison, which is shown in Table 15. All experiments were conducted on a personal computer with Windows11 operating system, Python 3.10 environment, Intel (R) Core (TM) i7 13500U CPU and 32GB RAM. Before conducting experiments, 20 multi-source datasets were generated using the composite noise addition method. Specifically, gaussian white noise with a standard deviation of 0.05 was added to the attribute data to preserve the structure and features of the data. Additionally, 25 % of the data was randomly altered with salt-and-pepper noise, changing it to the maximum or minimum value of the attribute under a particular decision class.

5.1. Performance of static fusion method

The purpose of this section is to validate the effectiveness of the proposed fusion method, abbreviated as DDF. DDF was compared with three common fusion methods:

(1) $MaxF_p(x) = \max_{i\in\{1,2,\dots,m\}} f_i(x,p)$;
(2) $MinF_p(x) = \min_{i\in\{1,2,\dots,m\}} f_i(x,p)$;
(3) $MeanF_p(x) = \frac{1}{N}\sum_{i=1}^{N} f_i(x,p)$.

Table 12
Complexity analysis of case (c) (Algorithm 5).

Time complexity	Cyclic analysis	In Algorithm 5, Steps from 2 to 6 iterate over the new properties of the reserved source and calculate the corresponding information entropy $E_p(D\|\mathcal{O}DS_q)$, the time complexity of this part is $O(m\times\Delta n\times\|U\|)$. Steps from 7 to 13 compare the information entropy of the original properties, the time complexity of this part is $O(n)$. The complexity of steps from 14 to 16 is $O(n)$.
	Overall time complexity	Therefore, The time complexity of this part is $O(m\times\Delta n\times\|U\|)$.
Space complexity	Static space requirement	It mainly includes all inputs in Algorithm 5 and the variables used inside the algorithm, and the space complexity is $O(1)$.
	Dynamic space requirement	To reduce the running time in practice, it mainly uses a numpy array for two-dimensional operation. The time complexity is calculated separately for each part. The space complexity of steps from 2 to 6 is about $O(m\times\Delta n\times\|U\|^2)$, depending on the calculation of the dominant and distance relationships, sources and attributes.
	Overall space complexity	Thus, the space complexity is roughly $O(m\times\Delta n\times\|U\|^2)$.

Table 13
The time complexity of fusion algorithms.

Case	Static algorithm	Dynamic algorithm
(a)	$O((m+\Delta m)\times n\times\|U\|)$	$O(\Delta m\times n\times\|U\|)$
(b)	$O((m+\Delta m)\times(n+\Delta n)\times\|U\|)$	$O(\Delta m\times(n+\Delta n)\times\|U\|+m\times\Delta n\times\|U\|)$
(c)	$O(m\times n\times\|U\|)$	$O(n)$
(d)	$O(m\times(n+\Delta n)\times\|U\|)$	$O(m\times\Delta n\times\|U\|)$

Table 14
The space complexity of fusion algorithms.

Case	Static algorithm	Dynamic algorithm
(a)	$O((m+\Delta m)\times n\times\|U\|^2)$	$O(\Delta m\times n\times\|U\|^2)$
(b)	$O((m+\Delta m)\times(n+\Delta n)\times\|U\|^2)$	$O(\Delta m\times(n+\Delta n)\times\|U\|^2+m\times\Delta n\times\|U\|^2)$
(c)	$O(m\times n\times\|U\|^2)$	$O(1)$
(d)	$O(m\times(n+\Delta n)\times\|U\|^2)$	$O(m\times\Delta n\times\|U\|^2)$

Table 15
The detailed presentation of data sets.

No.	Data sets	Abbreviation	Instances	Attributes	Labels
1	Automobile	Automobile	159	18	4
2	Wine	Wine	178	13	3
3	Bands	Bands	365	19	2
4	Auto MPG	AM	398	7	3
5	Wisconsin Diagnostic Breast Cancer	WDBC	569	30	2
6	Abalone	Abalone	4177	8	3
7	Wine Quality-white	WQW	4898	11	7
8	Phoneme	Phoneme	5404	5	2
9	Apartments for Rent Classified	ARC	10000	6	3

AP and AQ were used as indicators of approximate classification precision and quality to evaluate the effectiveness of DDF. AP and AQ effectively represent the accuracy and quality of approximate classification. The higher the values of AP and AQ, the more accurate and higher quality the approximate classification is. A relation R_α is defined on the set U as $\{(x,y)\in U\times U\,|\,\frac{dis(x,y)}{\max_{z\in U}\{dis(x,z)\}}\leq\alpha\}$, where dis represents the Euclidean distance between two samples. The value of α ranges from 0.05 to 0.5 within the interval defining the sample relationship. A smaller value indicates a more minor difference between the samples. We compare the approximation precision and quality of the DDF method with three standard methods using the relationship R_α. The comparative results are illustrated in Figs. 5 and 6. From these two figures, it is clear that DDF performs better on the Automobile, Wine, AM, WDBC, WQR, Abalone and WQW datasets when α takes smaller values. For the Bands dataset, MinF methods exhibit equally good performance. On both Phoneme and ARC datasets, the MeanF method demonstrates superior performance. Therefore, in most cases, our method is a better choice for integrating MS-ODS.

A series of experiments was conducted to demonstrate the improvement of the new multi-source fusion-based representation method for classification tasks. Classic classifiers, such as K-nearest neighbors (KNN), decision tree (DT), and support vector machine (SVM), were employed to handle the data after weighted distance fuzzy dominance class processing. To obtain more accurate results, 10-fold cross-validation was utilized during the classification process, and the mean and standard deviation of 10-fold classification accuracy were presented in Tables 16, 17, and 18. The parameters k and C, which can impact the performance of KNN and SVM classifiers, were also adjusted to achieve the optimal classification outcomes.

Table 16
Classification accuracy and variance of the fusion results of KNN.

Data sets	KNN			
	DDF	MaxF	MinF	MeanF
Automobile ($k=3$)	**34.67 ± 0.55**	21.25 ± 1.34	33.38 ± 0.89	32.04 ± 0.48
Wine ($k=7$)	**64.61 ± 0.86**	35.39 ± 1.92	62.39 ± 0.97	64.08 ± 0.91
Bands ($k=3$)	**61.63 ± 0.42**	49.30 ± 1.04	59.43 ± 0.22	55.83 ± 0.84
AM ($k=12$)	**66.06 ± 0.44**	60.17 ± 0.40	66.32 ± 0.34	65.56 ± 0.43
WDBC ($k=18$)	**87.00 ± 0.05**	51.49 ± 0.34	86.64 ± 0.14	84.71 ± 0.22
Abalone ($k=35$)	50.56 ± 0.04	34.00 ± 0.08	**50.68 ± 0.04**	50.49 ± 0.06
WQW($k=68$)	**45.40 ± 0.05**	43.56 ± 0.02	45.21 ± 0.01	44.48 ± 0.04
Phoneme ($k=3$)	**77.61 ± 0.01**	62.95 ± 0.02	74.46 ± 0.02	76.02 ± 0.02
ARC ($k=23$)	**88.58 ± 0.01**	50.80 ± 0.02	78.09 ± 0.02	84.62 ± 0.02

In around 80 % of the datasets, the proposed fusion method using weighted distance fuzzy dominance class is outperformed by the other three standard methods. In addition, we use the Wilcoxon signed-rank test to evaluate whether the proposed method's classification accuracy on the 10-fold cross-validation is statistically superior to the three commonly used methods. The null hypothesis (H_0) is defined as: DDF ≤ MeanF/MinF/MaxF, where DDF represents the fusion method's classification accuracy and MeanF, MinF and MaxF represent the classification accuracy of the other three fusion methods. The alternative hypothesis (H_1) is defined as: DDF > MeanF/MinF/MaxF, indicating that the fusion method's classification accuracy center is greater than that of the other three fusion methods. The comparison results' p-values are shown in Table 19. A smaller p-value indicates stronger statistical significance. In the case of a significance level of 20 %, most of the comparison results are statistically significant.

To enhance the reliability and diversity of our experiments, we referred to the CeF method proposed in [40] and compared it with our process. This method offers a fusion approach on multi-source interval-valued ordered datasets, where single values can be seen as a particular case of interval values. Therefore, we can compare them from a classification perspective. The classification results are shown in Table 20, and the p-values comparison results are shown in Table 21. It can be observed from the results that our proposed method outperforms the CeF method on most datasets.

5.2. Performance of incremental fusion method

In this section, the efficiency of incremental DDF is compared with static DDF. Four dynamic methods will be compared with static methods

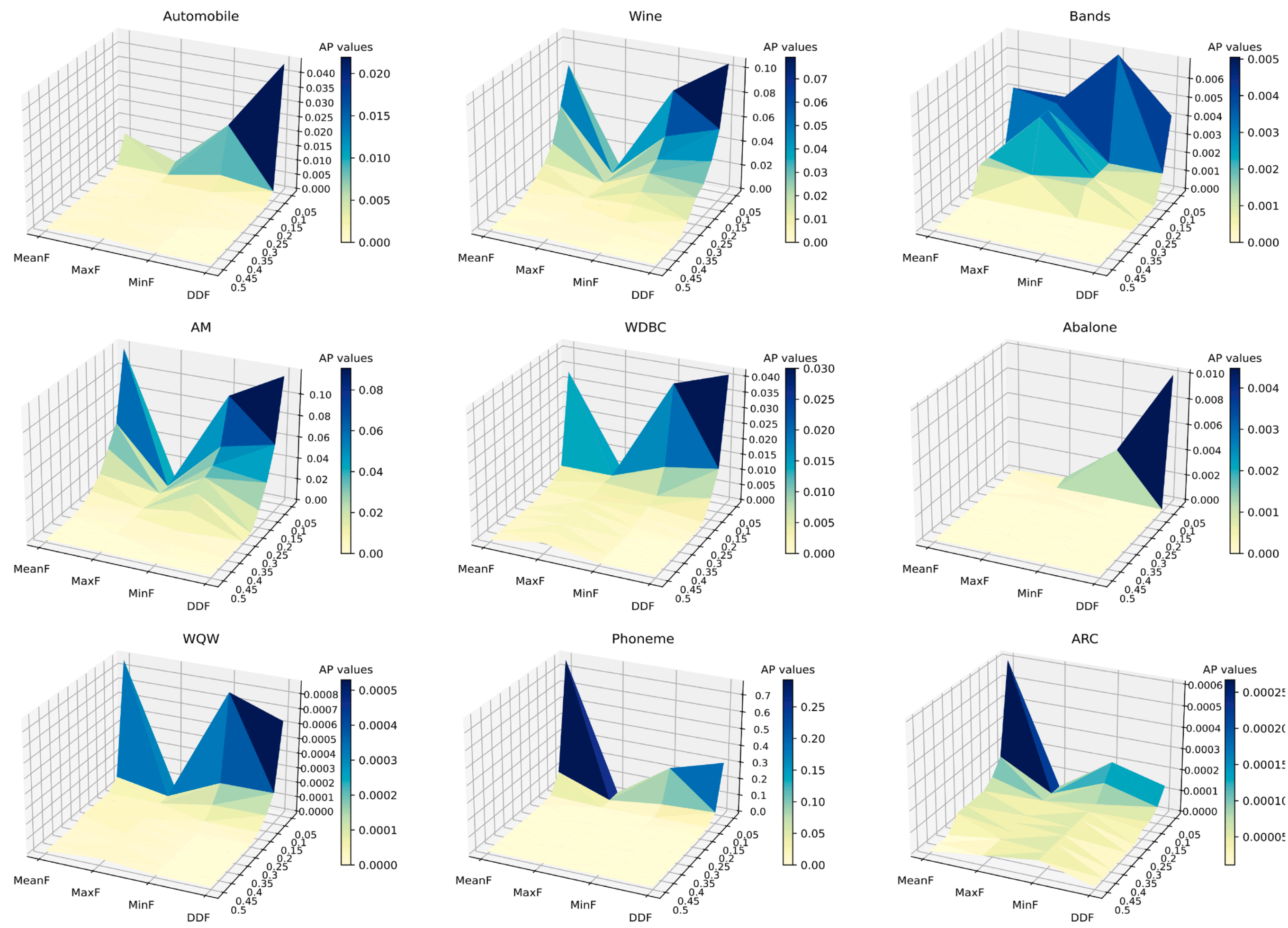

Fig. 5. Approximation precision contrast.

Table 17
Classification accuracy and variance of the fusion results of DT.

Data sets	DT			
	DDF	MaxF	MinF	MeanF
Automobile	**48.42 ± 2.25**	25.79 ± 1.21	45.79 ± 2.13	45.33 ± 2.06
Wine	83.69 ± 0.90	36.99 ± 1.26	61.70 ± 1.73	**88.17 ± 0.53**
Bands	**59.19 ± 0.46**	55.07 ± 0.79	56.16 ± 0.06	58.91 ± 0.59
AM	**67.10 ± 0.18**	48.46 ± 0.36	62.72 ± 0.75	65.54 ± 0.55
WDBC	87.34 ± 0.07	51.69 ± 0.35	**88.93 ± 0.15**	85.94 ± 0.14
Abalone	**44.72 ± 0.02**	33.56 ± 0.10	45.30 ± 0.04	44.60 ± 0.02
WQW	**37.82 ± 0.02**	32.39 ± 0.05	36.51 ± 0.03	37.78 ± 0.08
Phoneme	**74.50 ± 0.01**	59.14 ± 0.04	70.73 ± 0.02	72.45 ± 0.05
ARC	87.16 ± 0.01	39.94 ± 0.01	74.08 ± 0.01	**91.02 ± 0.01**

Table 18
Classification accuracy and variance of the fusion results of SVM.

Data sets	SVM			
	DDF	MaxF	MinF	MeanF
Automobile ($C = 8$)	**42.83 ± 0.98**	30.21 ± 0.07	39.08 ± 1.01	40.92 ± 0.59
Wine ($C = 30$)	**68.53 ± 1.05**	39.38 ± 0.10	63.95 ± 0.76	66.24 ± 0.34
Bands ($C = 2$)	**63.59 ± 0.07**	63.03 ± 0.01	63.03 ± 0.01	63.03 ± 0.01
AM ($C = 12$)	**66.84 ± 0.14**	62.76 ± 0.03	62.50 ± 0.01	66.34 ± 0.13
WDBC ($C = 60$)	**88.75 ± 0.21**	62.92 ± 0.01	88.05 ± 0.12	87.69 ± 0.19
Abalone ($C = 45$)	51.04 ± 0.04	36.68 ± 0.00	49.94 ± 0.05	**52.72 ± 0.04**
WQW ($C = 8$)	**45.15 ± 0.00**	44.86 ± 0.00	44.86 ± 0.01	45.03 ± 0.00
Phoneme ($C = 100$)	76.94 ± 0.02	70.82 ± 0.00	75.81 ± 0.01	**79.11 ± 0.04**
ARC ($C = 165$)	**86.51 ± 0.01**	54.09 ± 0.00	78.34 ± 0.01	54.10 ± 0.00

based on computation time. The updated simulation of the data set is as follows.

When there are changes in information sources, 50 % of the data sources are treated as the source, and the number of sources for the remaining 50 % increases by 10 %. When the source is reduced, the entire data set is treated as the base and is decreased sequentially by 10 %. The same approach is applied for attribute changes, except the range of changes is chosen based on the number of attributes in each dataset.

The dynamic fusion method time is compared with the static fusion method time shown in Figs. 7–10. The x axis represents the specific changes in data set attributes and information sources, and the y axis represents the fusion running time. In Tables 24–27, static time, incremental time, and the time ratio between them are provided for different increment scenarios. It can be observed that the fusion time of the incremental method is significantly less than that of the static method. Due to the application of previous knowledge and the comparison of time complexity, it can be seen that the dynamic mechanism can avoid double computation and significantly save calculation time.

5.3. Data quality analysis

In this section, we evaluated the data quality of all datasets. Using ten-fold cross-validation, we calculated the average multi-class cross-entropy loss for different levels of noise using the DDF, MaxF, MinF, and MeanF methods. Specifically, we started with Gaussian noise at a standard deviation of 0.1 and incrementally increased it by 0.1 for five

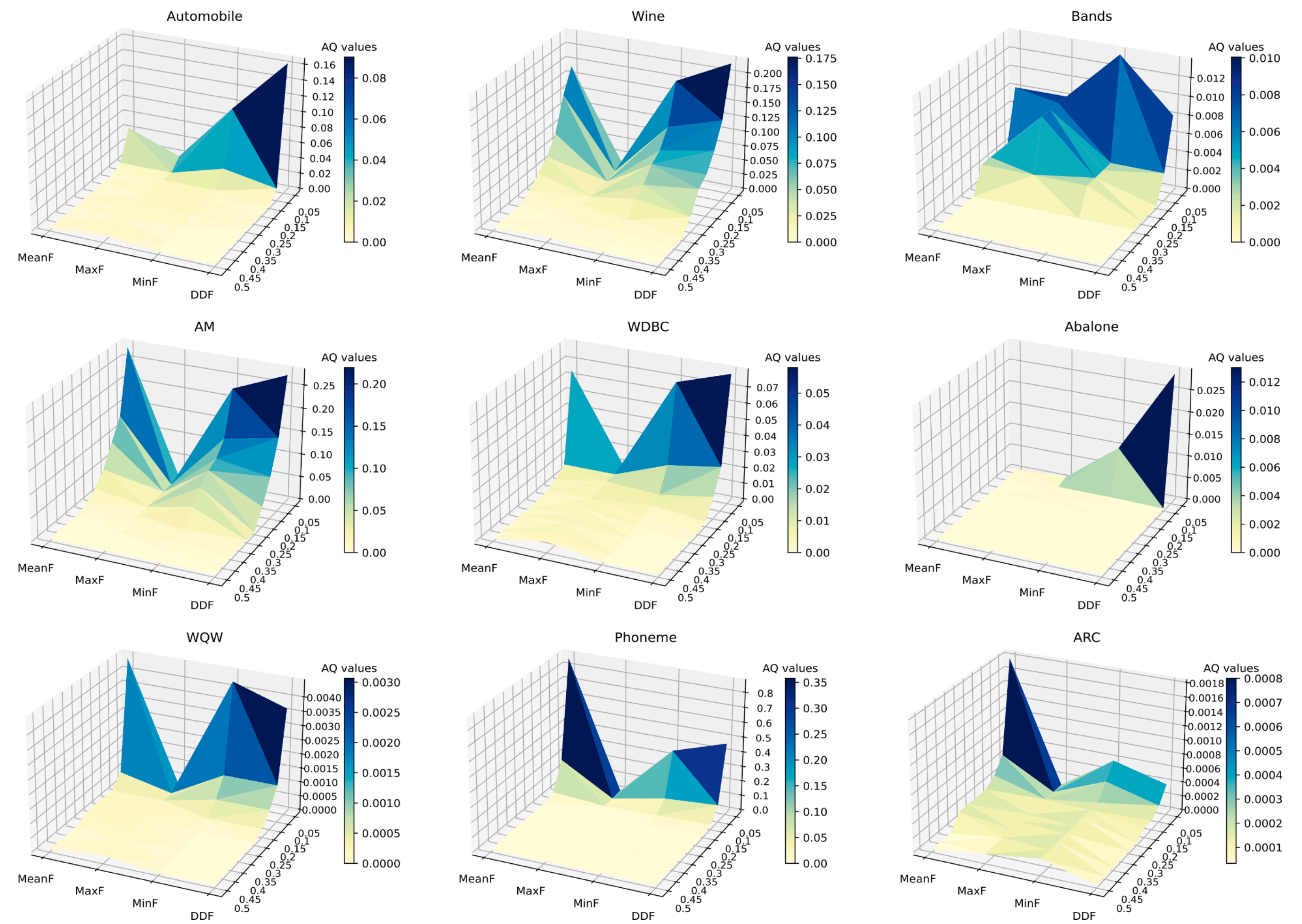

Fig. 6. Approximation quality contrast.

Table 19
P-values of the comparison results in KNN, DT and SVM.

Data sets	KNN			DT			SVM		
	$H_1 : DDF > MaxF$	$H_1 : DDF > MinF$	$H_1 : DDF > MeanF$	$H_1 : DDF > MaxF$	$H_1 : DDF > MinF$	$H_1 : DDF > MeanF$	$H_1 : DDF > MaxF$	$H_1 : DDF > MinF$	$H_1 : DDF > MeanF$
Automobile	**0.020709**	0.347656	0.304124	**0.002930**	0.347656	0.316140	**0.006836**	**0.070229**	0.213136
Wine	**0.004545**	**0.171441**	0.603770	**0.000977**	**0.007074**	0.784180	**0.000977**	**0.154814**	0.317047
Bands	**0.009766**	**0.186077**	**0.065430**	**0.117863**	**0.146218**	0.406190	**0.068351**	**0.068351**	**0.068351**
AM	**0.010353**	0.638496	0.317047	**0.000977**	**0.137695**	0.384766	**0.013672**	**0.007074**	0.389053
WDBC	**0.000977**	0.384766	**0.142099**	**0.000977**	0.946063	**0.199538**	**0.000977**	0.299587	0.287232
Abalone	**0.000977**	0.539063	0.500000	**0.000977**	0.761402	0.500000	**0.000977**	**0.048601**	0.997070
WQW	**0.004545**	0.384766	**0.065430**	**0.000977**	**0.065430**	0.500000	**0.020905**	**0.171202**	0.305726
Phoneme	**0.000977**	**0.000977**	**0.004883**	**0.000977**	**0.000977**	**0.006836**	**0.000977**	**0.021913**	0.995424
ARC	**0.000977**	**0.000977**	**0.000977**	**0.000977**	**0.000977**	1.000000	**0.000977**	**0.000977**	**0.000977**

Table 20
Classification accuracy and variance of DDF and CeF.

Data sets	KNN		DT		SVM	
	DDF	CeF	DDF	CeF	DDF	CeF
Automobile	**34.67 ± 0.55**	32.63 ± 1.59	**48.42 ± 2.25**	43.42 ± 0.44	**42.83 ± 0.98**	39.67 ± 1.21
Wine	64.61 ± 0.86	**70.29 ± 1.73**	**83.69 ± 0.90**	83.1 ± 0.77	**68.53 ± 1.05**	59.41 ± 1.97
Bands	**61.63 ± 0.42**	61.08 ± 0.48	**59.19 ± 0.46**	58.89 ± 0.53	**63.59 ± 0.07**	62.48 ± 0.02
AM	66.06 ± 0.44	**68.85 ± 0.35**	**67.10 ± 0.18**	64.78 ± 0.44	66.84 ± 0.14	**67.62 ± 0.1**
WDBC	**87.00 ± 0.05**	86.82 ± 0.12	**87.34 ± 0.07**	86.29 ± 0.15	88.75 ± 0.21	**89.63 ± 0.19**
Abalone	**50.56 ± 0.04**	49.1 ± 0.03	44.72 ± 0.02	**46.23 ± 0.03**	**51.04 ± 0.04**	50.75 ± 0.05
WQW	**45.40 ± 0.05**	44.07 ± 0.03	37.82 ± 0.02	**38.21 ± 0.04**	**45.15 ± 0.00**	44.88 ± 0.0
Phoneme	**77.61 ± 0.01**	77.44 ± 0.02	**74.50 ± 0.01**	72.89 ± 0.04	76.94 ± 0.02	**77.54 ± 0.03**
ARC	**88.58 ± 0.01**	87.11 ± 0.01	**87.16 ± 0.01**	86.41 ± 0.01	**86.51 ± 0.01**	85.35 ± 0.01

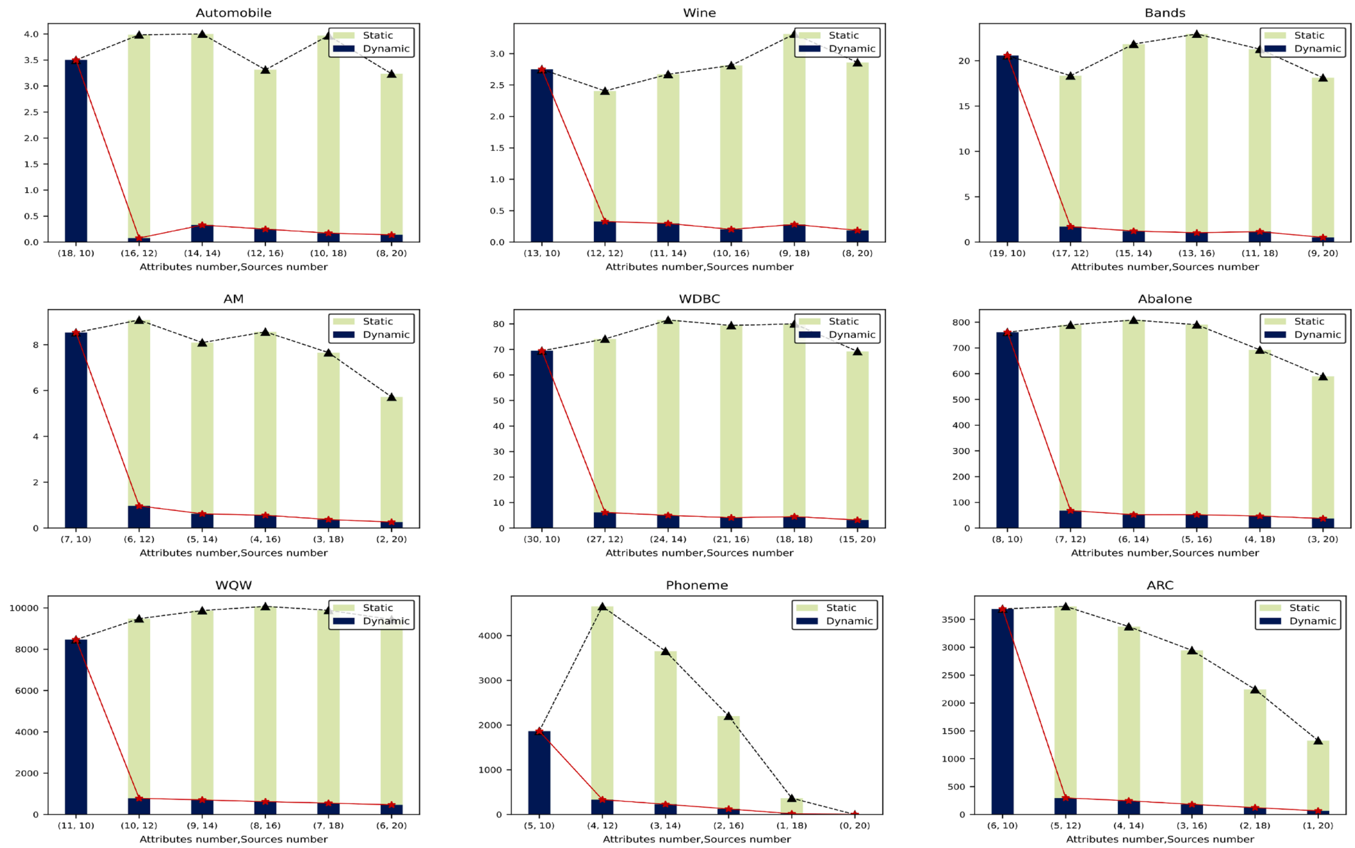

Fig. 7. Runtime time of case (a).

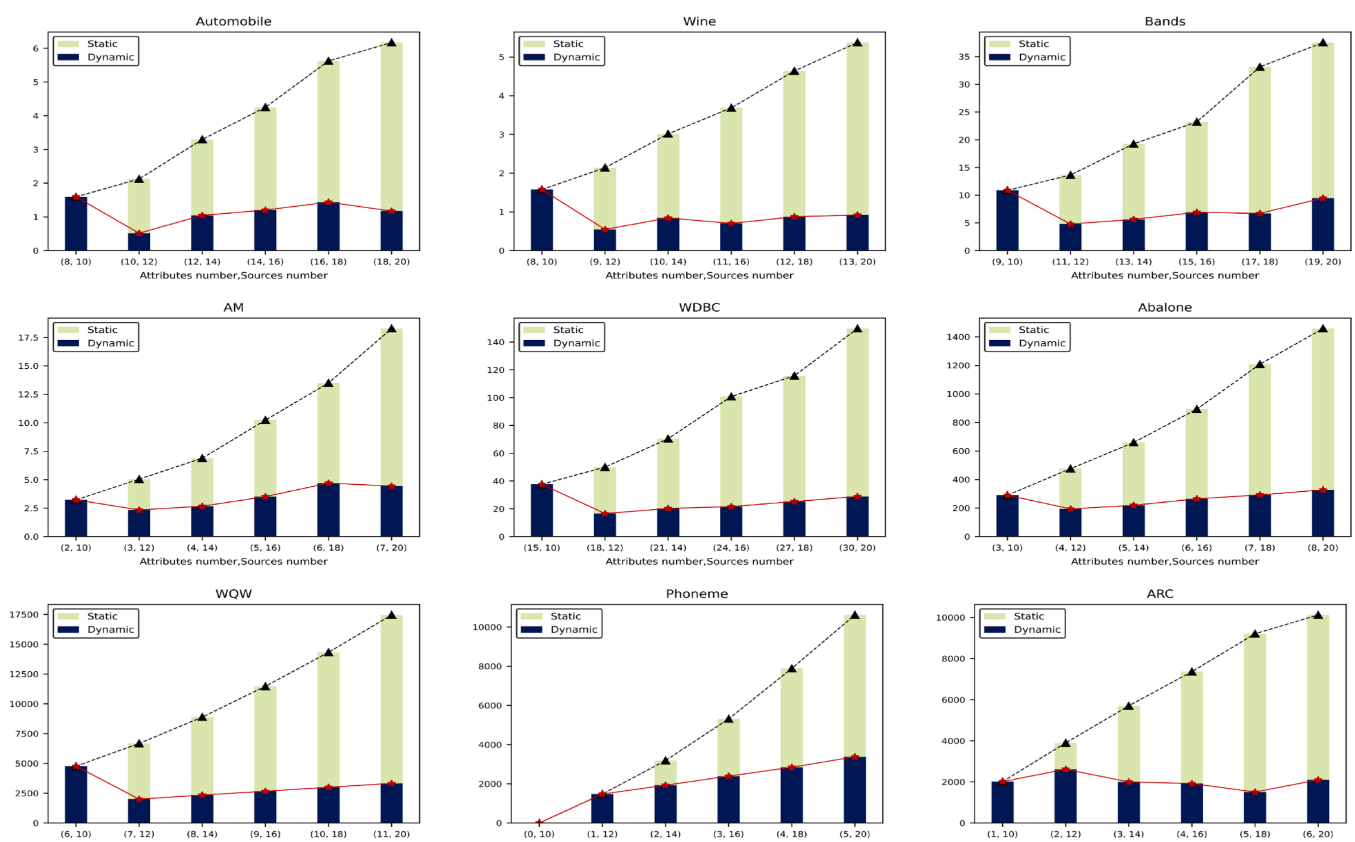

Fig. 8. Runtime time of case (b).

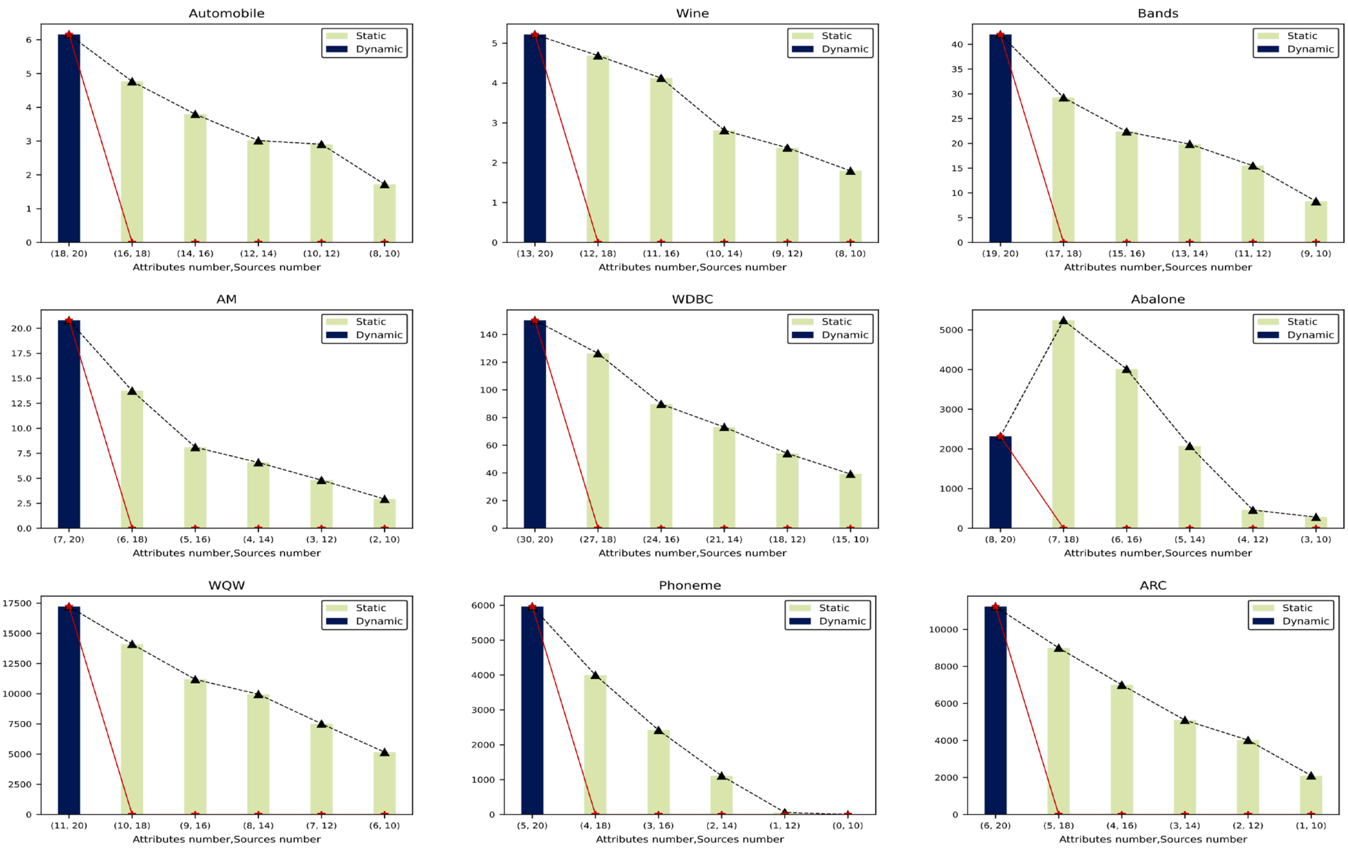

Fig. 9. Runtime time of case (c).

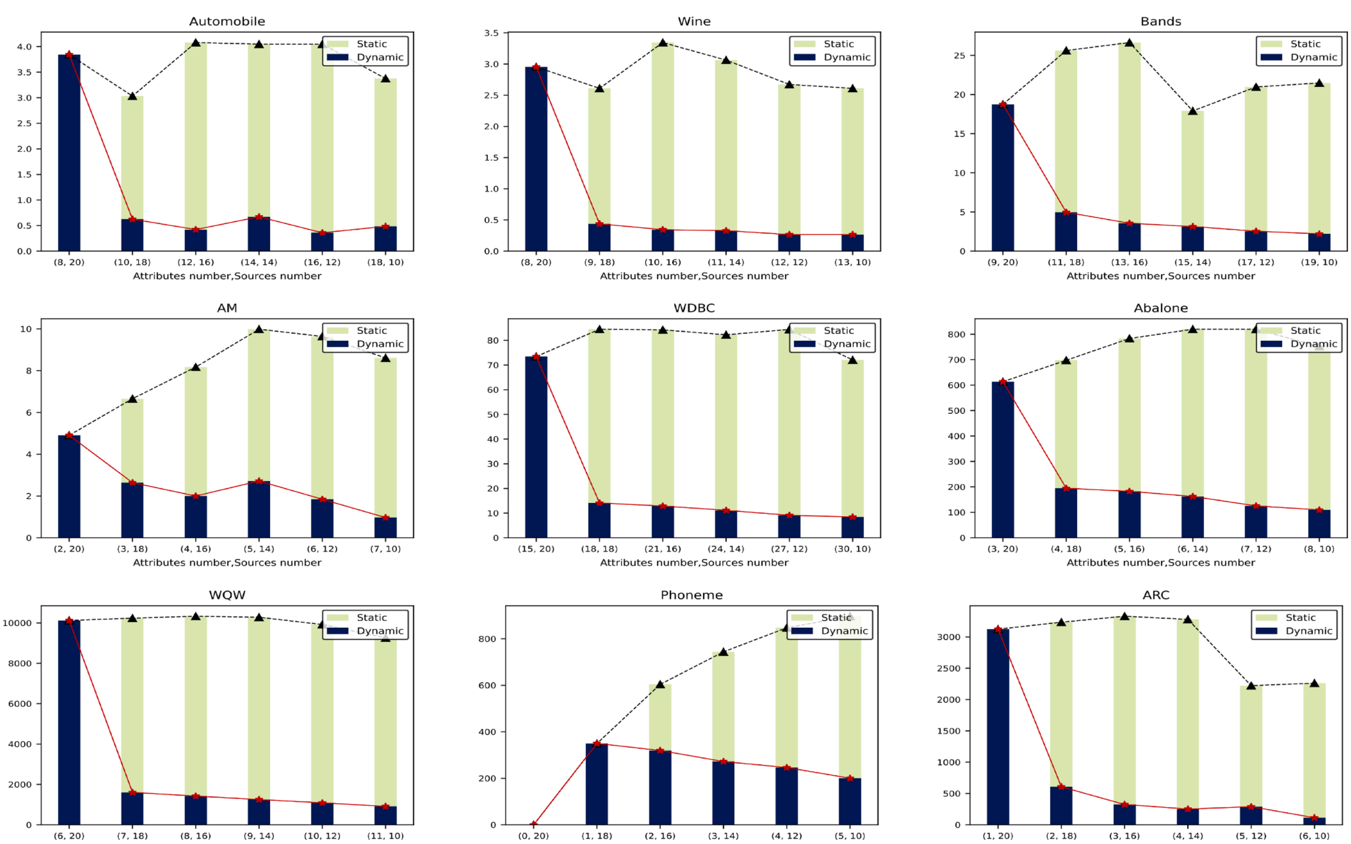

Fig. 10. Runtime time of case (d).

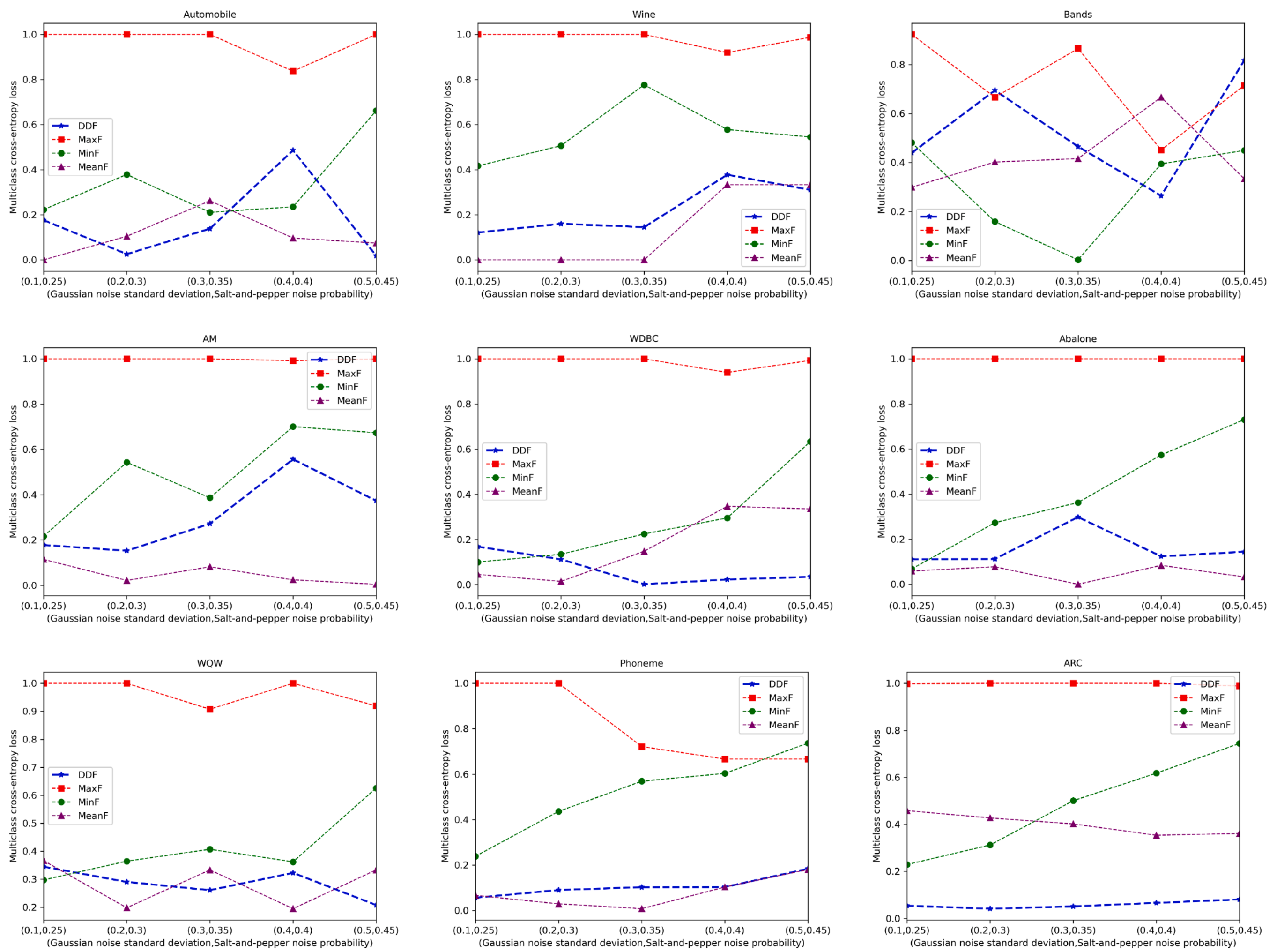

Fig. 11. Comparison of multiclass cross-entropy loss in different dataset.

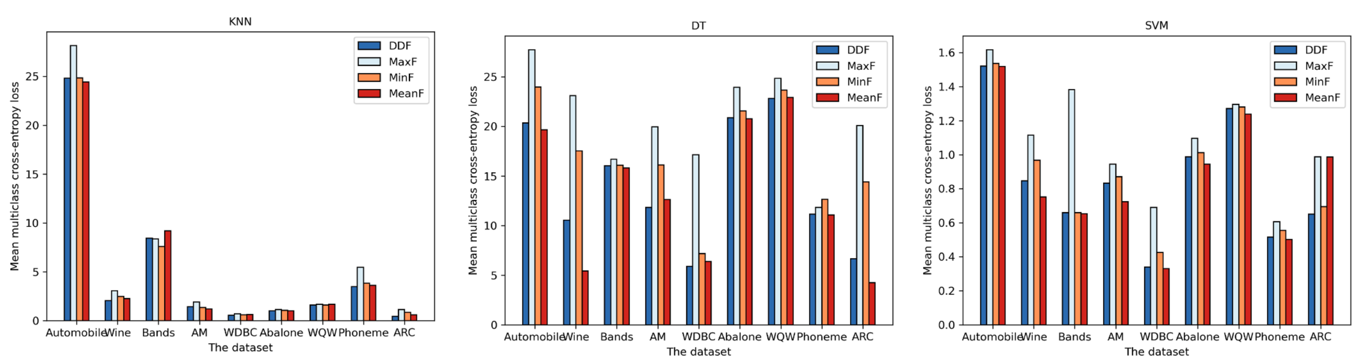

Fig. 12. Comparison of multiclass cross-entropy loss in different classifier.

sets. We started at 0.25 and increased it by 0.05 increments for salt-and-pepper noise. The average multi-class cross-entropy loss was computed as Algorithm 6.

These average values served as evaluation metrics, allowing us to measure the classification performance across different datasets. A minor multi-class cross-entropy loss indicates a close relationship between the classification predictions and the actual labels, demonstrating superior classification performance. Our algorithm performs better as the dataset size increases, as shown in Fig. 11. The average multi-class cross-entropy loss under different noise levels is also consistently similar across the three classifiers, as demonstrated in Fig. 12.

5.4. K-parameter analysis

In this paper, the dominance degree fuzzy relation in the proposed model has a single parameter variable, denoted as k, which is set to 1 for all datasets analyzed. Therefore, this analysis investigates the impact of different parameter k values on the model's fusion results.

Firstly, the definition of dominance degree fuzzy relation can be traced back to the logsig sigmoid transfer function $f(x) = \frac{1}{1+e^{-kx}}$. The logsig sigmoid transfer function introduces nonlinearity, constrains the neural network's output within the range of $[0, 1]$, and facilitates the learning of nonlinear relationships in the network. Here, the parameter

Table 21
P-values of the comparison results of DDF and CeF.

Data sets	KNN (H_1 : $DDF > CeF$)	DT (H_1 : $DDF > CeF$)	SVM (H_1 : $DDF > CeF$)
Automobile	**0.086986**	**0.137695**	**0.084782**
Wine	0.952133	0.261739	**0.046045**
Bands	0.347656	0.577148	**0.028119**
AM	0.953015	**0.096064**	0.778615
WDBC	0.335189	0.287232	0.800172
Abalone	**0.028572**	0.870013	**0.137695**
WQW	**0.018555**	0.500000	**0.024555**
Phoneme	0.406028	**0.024414**	0.883789
ARC	**0.007489**	**0.024414**	**0.032227**

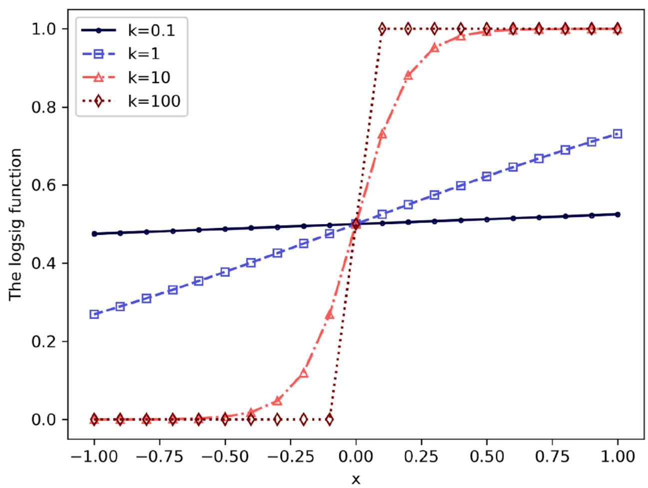

Fig. 13. Curve of function logsig in interval [−1,1].

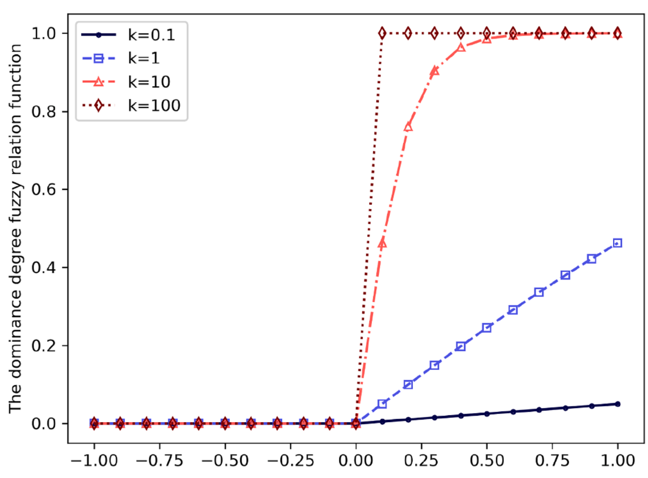

Fig. 14. Curve of function the dominance degree fuzzy relation in interval [−1,1].

k represents the user's preference degree, and its graph is illustrated in Fig. 13. Building upon this function, we apply it to the dominance degree fuzzy relation and further evolve it, ensuring that the value range of the entire dominance degree unclear relation function remains within $[0, 1]$. The graph of this function is depicted in Fig. 14. Subsequently, If the number of elements in result sets A and B are denoted as $|A|$ and $|B|$ respectively, and the number of elements in their intersection is $|A \bigcap B|$, then their overlap can be represented as: $O(A, B) = \frac{|A \bigcap B|}{\min(|A|,|B|)}$. Here, $O(A, B)$ represents the overlap between result sets A and B. It indicates the proportion of common elements between the two result sets relative to the total number of elements in both sets. The value of $O(A, B)$ ranges from 0 to 1, where $O(A, B) = 1$ indicates complete overlap between result sets A and B, while $O(A, B) = 0$ implies no common elements between the sets. We conducted experiments on nine datasets, comparing the fusion results of the model with k values of 0.1, 10, and 100 to those with k value of 1. The experimental results demonstrate that different values of k do not significantly affect the fusion results, as shown in Table 22. These experimental findings suggest that the k value of 1 in the proposed model is already adequate for model fusion. Different values of k may not yield a significant impact.

5.5. *Conflict analysis*

We employ a voting strategy for comparative analysis to conduct a more in-depth comparison between the fusion results and the results from most data sources for assessing the existence of conflicting attributes. When comparing the conflict rates of a specific attribute, we extract the results of the sample for that attribute from different data sources, forming a comparison domain. For instance, we extract the result of sample x_1 under attribute a_1 from the fusion model and compare it with the results in the comparison domain, which refers to the val-

Table 22
Comparison results of overlap for k parameter (%).

Data sets	$O(result_{(k=1)}, result_{k=k_0})$		
	$k_0 = 0.1$	$k_0 = 10$	$k_0 = 100$
Automobile	94.55	100.0	100.0
Wine	100.0	100.0	100.0
Bands	100.0	100.0	100.0
AM	100.0	100.0	100.0
WDBC	100.0	100.0	100.0
Abalone	87.71	87.71	87.71
WQW	100.0	100.0	100.0
Phoneme	100.0	100.0	100.0
ARC	100.0	100.0	100.0

Table 23
Comparison results of overlap for conflict rate (%).

Data sets	overall attribute conflict rate	average conflict rate
Automobile	5.66, 3.77, 4.4, 3.14, 1.89, 3.77, 3.14, 5.03, 5.03, 5.03, 10.06, 1.26, 3.14, 5.66, 6.92, 4.4, 3.77, 1.89	4.33
Wine	7.3, 4.49, 4.49, 5.06, 3.37, 7.87, 8.99, 2.25, 5.06, 11.24, 2.81, 2.81, 2.81	5.27
Bands	3.56, 4.66, 2.74, 2.47, 5.48, 3.84, 3.84, 6.03, 2.74, 10.41, 6.58, 6.3, 6.03, 3.56, 9.86, 7.4, 7.12, 3.56, 2.47	5.19
AM	4.08, 5.1, 6.38, 5.87, 5.61, 4.85, 4.34	5.18
WDBC	5.45, 2.99, 4.04, 4.92, 1.76, 3.69, 3.16, 4.39, 3.51, 4.57, 3.69, 9.84, 5.1, 5.62, 2.11, 1.76, 1.58, 2.11, 8.61, 2.46, 4.39, 6.33, 3.16, 5.98, 5.1, 5.45, 3.34, 1.76, 4.22, 4.22	4.18
Abalone	3.85, 2.87, 3.47, 7.71, 5.53, 6.18, 7.61, 6.85	5.51
WQW	4.59, 3.78, 5.0, 4.55, 2.84, 5.6, 4.68, 3.9, 7.64, 4.04, 6.58	4.84
Phoneme	6.99, 4.42, 5.29, 4.29, 6.37	5.47
ARC	6.94, 5.6, 7.24, 4.36, 5.25, 5.69	5.85

Table 24
Time and time ratio in case (a).

Data sets	Attribute decreases by 10 %, and source increases by 10 %.						
Automobile	Static time	3.5	3.98	4.0	3.31	3.97	3.23
	Dynamic time	3.5	0.08	0.33	0.25	0.17	0.14
	Ratio	1.0	49.75	12.12	13.24	23.35	23.07
Wine	Static time	2.75	2.41	2.67	2.81	3.31	2.86
	Dynamic time	2.75	0.33	0.3	0.2	0.28	0.19
	Ratio	1.0	7.3	8.9	14.05	11.82	15.05
Bands	Static time	20.58	18.34	21.84	22.95	21.28	18.12
	Dynamic time	20.58	1.7	1.22	1.03	1.16	0.52
	Ratio	1.0	10.79	17.9	22.28	18.34	34.85
AM	Static time	8.53	9.08	8.09	8.56	7.66	5.72
	Dynamic time	8.53	0.97	0.62	0.56	0.38	0.27
	Ratio	1.0	9.36	13.05	15.29	20.16	21.19
WDBC	Static time	69.45	74.11	81.52	79.41	80.02	69.22
	Dynamic time	69.45	6.14	5.02	4.08	4.52	3.19
	Ratio	1.0	12.07	16.24	19.46	17.7	21.7
Abalone	Static time	761.05	789.47	808.48	790.22	692.45	589.28
	Dynamic time	761.05	68.12	52.17	52.3	47.09	37.22
	Ratio	1.0	11.59	15.5	15.11	14.7	15.83
WQW	Static time	8469.02	9483.11	9876.53	10075.55	9888.86	9441.3
	Dynamic time	8469.02	784.16	708.83	624.36	551.75	470.73
	Ratio	1.0	12.09	13.93	16.14	17.92	20.06
Phoneme	Static time	1863.28	4652.78	3650.61	2202.08	365.72	0.0
	Dynamic time	1863.28	333.08	229.33	124.67	18.42	0.0
	Ratio	1.0	13.97	15.92	17.66	19.85	/
ARC	Static time	3688.94	3734.36	3371.78	2946.0	2245.2	1325.53
	Dynamic time	3688.94	298.55	246.48	182.17	123.66	65.86
	Ratio	1.0	12.51	13.68	16.17	18.16	20.13

Table 25
Time and time ratio in case (b).

Data sets	Attribute increases by 10 %, and the source increases by 10 %.						
Automobile	Static time	1.59	2.12	3.3	4.25	5.62	6.17
	Dynamic time	1.59	0.52	1.05	1.2	1.44	1.17
	Ratio	1.0	4.08	3.14	3.54	3.9	5.27
Wine	Static time	1.58	2.14	3.02	3.69	4.64	5.38
	Dynamic time	1.58	0.55	0.84	0.7	0.88	0.92
	Ratio	1.0	3.89	3.6	5.27	5.27	5.85
Bands	Static time	10.88	13.64	19.27	23.22	33.14	37.55
	Dynamic time	10.88	4.8	5.64	6.92	6.72	9.47
	Ratio	1.0	2.84	3.42	3.36	4.93	3.97
AM	Static time	3.23	5.05	6.89	10.23	13.5	18.28
	Dynamic time	3.23	2.34	2.67	3.52	4.7	4.44
	Ratio	1.0	2.16	2.58	2.91	2.87	4.12
WDBC	Static time	37.64	50.03	70.47	100.81	115.8	149.75
	Dynamic time	37.64	16.58	20.27	21.61	25.27	28.89
	Ratio	1.0	3.02	3.48	4.66	4.58	5.18
Abalone	Static time	290.28	476.31	659.84	893.28	1208.23	1458.45
	Dynamic time	290.28	194.3	219.56	265.23	292.44	327.95
	Ratio	1.0	2.45	3.01	3.37	4.13	4.45
WQW	Static time	4768.88	6677.09	8889.94	11471.5	14324.67	17461.91
	Dynamic time	4768.88	2005.55	2346.84	2668.27	2999.53	3325.59
	Ratio	1.0	3.33	3.79	4.3	4.78	5.25
Phoneme	Static time	0.0	1479.5	3189.36	5313.33	7889.12	10620.11
	Dynamic time	0.0	1479.5	1928.62	2392.89	2845.77	3379.02
	Ratio	/	1.0	1.65	2.22	2.77	3.14
ARC	Static time	2013.99	3897.67	5701.68	7367.67	9208.88	10124.65
	Dynamic time	2013.99	2613.99	1998.99	1920.99	1507.99	2100.99
	Ratio	1.0	1.49	2.85	3.84	6.11	4.82

Algorithm 6: Calculation of average classification cross-entropy loss.

```
for each noise value do
    Divide the dataset into ten folds for ten-fold
      cross-validation;
    for each fold do
        Train and predict using KNN, DT, and SVM classifiers,
          and calculate the classification cross-entropy loss for
          that fold;
    end
    Taking the average of all folds as the multi-class
      cross-entropy loss of the dataset under that noise
      condition;
    for each classifier do
        Normalize the classification cross-entropy losses of the
          four fusion methods (DDF, MaxF, MinF, MeanF)
          using min-max normalization to scale the loss values
          to the range of [0, 1];
        Assign equal weights of 1/3 to the classifiers (KNN, DT,
          and SVM);
        Calculate the weighted average classification
          cross-entropy loss by multiplying the normalized
          classification cross-entropy losses by the respective
          weights of the classifiers and summing them;
    end
end
Take the average of the weighted average classification
  cross-entropy losses for all noise values to obtain the final
  average classification cross-entropy loss;
```

ues of attribute a_1 for sample x_1 across all data sources. The simplified algorithmic flow is outlined as Algorithm 7.

Algorithm 7: Voting strategy for comparative analysis.

```
Data: Sample result,comparison domain;
Calculate the mean and standard deviation of all values in
  comparison domain;
Compute the difference between sample result and the mean;
if difference> 2×standard deviation then
    return False;
    // Conflict exists
end
else
    return True;
    // No conflict
end
```

Considering the voting results of all samples under attribute a_1 in the fusion results, we can obtain the conflict rate for attribute a_1. By calculating the average of the conflict rates for all attributes, we can determine the sample conflict rate for this fusion result. Experimental results demonstrate that all conflict rates are below 10 %, as presented in Tables 22 and 23.

Table 26
Time and time ratio in case (c).

Data sets	Attribute decreases by 10 %, and the source decreases by 10 %.						
Automobile	Static time	6.16	4.77	3.8	3.02	2.91	1.72
	Dynamic time	6.16	0.0	0.0	0.0	0.0	0.0
	Ratio	1.0	/	/	/	/	/
Wine	Static time	5.22	4.69	4.12	2.81	2.38	1.8
	Dynamic time	5.22	0.0	0.0	0.0	0.0	0.0
	Ratio	1.0	/	/	/	/	/
Bands	Static time	41.98	29.25	22.38	19.84	15.5	8.3
	Dynamic time	41.98	0.0	0.0	0.0	0.0	0.0
	Ratio	1.0	/	/	/	/	/
AM	Static time	20.8	13.75	8.11	6.58	4.81	2.92
	Dynamic time	20.8	0.0	0.0	0.0	0.0	0.0
	Ratio	1.0	/	/	/	/	/
WDBC	Static time	150.23	126.31	89.72	73.12	54.08	39.17
	Dynamic time	150.23	0.0	0.0	0.0	0.0	0.0
	Ratio	1.0	/	/	/	/	/
Abalone	Static time	2317.08	5240.92	4012.53	2071.45	459.86	283.0
	Dynamic time	2317.08	0.0	0.0	0.0	0.0	0.0
	Ratio	1.0	/	/	/	/	/
WQW	Static time	17221.36	14116.91	11188.8	9961.44	7517.0	5165.86
	Dynamic time	17221.36	0.0	0.0	0.0	0.0	0.0
	Ratio	1.0	/	/	/	/	/
Phoneme	Static time	5964.41	3999.17	2421.02	1110.38	62.12	0.0
	Dynamic time	5964.41	0.0	0.0	0.0	0.0	0.0
	Ratio	1.0	/	/	/	/	
ARC	Static time	11235.98	9001.23	7005.24	5100.5	4021.39	2103.48
	Dynamic time	11235.98	0.0	0.0	0.0	0.0	0.0
	Ratio	1.0	/	/	/	/	/

Table 27
Time and time ratio in case (d).

Data sets	Attribute increases by 10 %, and the source decreases by 10 %.						
Automobile	Static time	3.84	3.03	4.08	4.05	4.05	3.38
	Dynamic time	3.84	0.62	0.42	0.67	0.36	0.48
	Ratio	1.0	4.89	9.71	6.04	11.25	7.04
Wine	Static time	2.95	2.61	3.34	3.06	2.67	2.61
	Dynamic time	2.95	0.44	0.34	0.33	0.27	0.27
	Ratio	1.0	5.93	9.82	9.27	9.89	9.67
Bands	Static time	18.75	25.62	26.64	17.91	20.97	21.48
	Dynamic time	18.75	4.97	3.56	3.14	2.55	2.2
	Ratio	1.0	5.15	7.48	5.7	8.22	9.76
AM	Static time	4.91	6.66	8.17	9.98	9.64	8.61
	Dynamic time	4.91	2.64	2.0	2.72	1.84	0.97
	Ratio	1.0	2.52	4.08	3.67	5.24	8.88
WDBC	Static time	73.45	84.52	84.25	82.27	84.44	72.06
	Dynamic time	73.45	14.11	12.88	11.14	9.12	8.45
	Ratio	1.0	5.99	6.54	7.39	9.26	8.53
Abalone	Static time	613.47	697.75	783.33	819.75	819.55	751.95
	Dynamic time	613.47	194.69	182.73	162.42	125.47	109.64
	Ratio	1.0	3.58	4.29	5.05	6.53	6.86
WQW	Static time	10123.69	10235.64	10330.34	10280.84	9922.28	9262.11
	Dynamic time	10123.69	1607.3	1424.86	1252.05	1089.77	913.77
	Ratio	1.0	6.37	7.25	8.21	9.1	10.14
Phoneme	Static time	0.0	349.58	604.05	743.88	847.3	896.94
	Dynamic time	0.0	349.98	319.83	272.16	246.41	200.41
	Ratio	/	1.0	1.89	2.73	3.44	4.48
ARC	Static time	3123.69	3235.64	3330.34	3280.84	2222.28	2262.11
	Dynamic time	3123.69	607.3	324.86	252.05	289.77	113.77
	Ratio	1.0	5.33	10.25	13.02	7.67	19.88

6. Conclusions

This paper proposes a new fusion method that uses weighted information entropy to fuse ordered data with decisions. Firstly, the weighted distance between two samples is defined and used to determine weighted distance fuzzy relations. Then, dominance degree fuzzy relations between samples are characterized, and these two relationships are combined to define information entropy. Essential sources of information are selected to fuse into new data. A complete static algorithm for the fusion process is provided, and its time complexity is analyzed. Four incremental fusion methods are proposed to address changes in information sources and conditional properties, and their time complexity is analyzed. Finally, comparative experiments on nine datasets were conducted, including experiments on approximate precision and approximate quality and classification comparison experiments. Our DDF fusion method demonstrated excellent performance. As fundamental data cannot be wholly ordered, our fusion method's effectiveness would be enhanced if applied to an information system with ultimately requested attributes. Additionally, time analysis comparisons show that the incremental fusion methods effectively reduce the runtime when the data sources and attributes change dynamically.

In the age of big data, classifying each object or fusing other information systems sometimes requires allocating substantial time and financial resources. Therefore, extending this research to handling different information systems and decision-less data is essential. Integrating heterogeneous data using fusion methods is also a significant challenge researchers face and an important direction for future research.

CRediT authorship contribution statement

Xiaoyan Zhang: Supervision, Investigation, Validation, Funding acquisition, Writing – review & editing, Methodology, Conceptualization; **Jiajia Lin:** Writing – review & editing, Writing – original draft, Data curation, Visualization, Methodology, Software.

Data availability

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

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.

Acknowledgment

The authors would like to thank the Associate Editor and the reviewers for their insightful comments and suggestions. This work was supported by the National Natural Science Foundation of China (Grant NOs.12371465, 62376229) and the Chongqing Natural Science Foundation of China (Grant NO.CSTB2023NSCQ-MSX1063).

References

[1] R. Zhang, F. Nie, X. Li, X. Wei, Feature selection with multi-view data: a survey, Inf. Fusion 50 (2019) 158–167. https://doi.org/10.1016/j.inffus.2018.11.019
[2] O. Basir, F. Karray, H. Zhu, Connectionist-based dempster-shafer evidential reasoning for data fusion, Trans. Neural Netw. 16 (6) (2005) 1513–1530. https://doi.org/10.1109/TNN.2005.853337. https://doi.org/10.1109/TNN.2005.853337
[3] L. Xie, P. Pan, Y. Lu, Markov random field based fusion for supervised and semi-supervised multi-modal image classification, Multimed. Tools Appl. 74 (2) (2015) 613–634. https://doi.org/10.1007/s11042-014-2018-y. https://doi.org/10.1007/s11042-014-2018-y
[4] K. Rafiezadeh Shahi, P. Ghamisi, B. Rasti, R. Jackisch, P. Scheunders, R. Gloaguen, Data fusion using a multi-sensor sparse-based clustering algorithm, Remote Sens. 12 (23) (2020) 4007. https://www.mdpi.com/2072-4292/12/23/4007. https://doi.org/10.3390/rs12234007
[5] Z. Pawlak, Rough sets, Int. J. Comput. Inf. Sci. 11 (5) (1982) 341–356. https://doi.org/10.1007/BF01001956. https://doi.org/10.1007/BF01001956
[6] L.A. Zadeh, Fuzzy sets, Inf. Contr. 8 (3) (1965) 338–353. https://www.sciencedirect.com/science/article/pii/S001999586590241X.
[7] M.S. Raza, U. Qamar, Fuzzy Rough Sets, Springer Singapore, Singapore, 2019, pp. 179–188. https://doi.org/10.1007/978-981-32-9166-9_9. https://doi.org/10.1007/978-981-32-9166-9_9
[8] M. Sarkar, Rough-fuzzy functions in classification, Fuzzy Sets Syst. 132 (3) (2002) 353–369. https://www.sciencedirect.com/science/article/pii/S0165011402001197. https://doi.org/10.1016/S0165-0114(02)00119-7
[9] W. Xu, Q. Wang, X. Zhang, Multi-granulation fuzzy rough set model on tolerance relations, The Fourth Int. Workshop Adv. Comput. Intell. (2011) 357–364. https://doi.org/10.1109/IWACI.2011.6160031
[10] W. Xu, K. Yuan, W. Li, Dynamic updating approximations of local generalized multigranulation neighborhood rough set, Appl. Intell. 52 (8) (2022) 9148–9173. https://doi.org/10.1007/s10489-021-02861-x. https://doi.org/10.1007/s10489-021-02861-x
[11] J. Zhan, X. Zhang, Y. Yao, Covering based multigranulation fuzzy rough sets and corresponding applications, Artif. Intell. Rev. 53 (2) (2020) 1093–1126. https://doi.org/10.1007/s10462-019-09690-y. https://doi.org/10.1007/s10462-019-09690-y
[12] W. Li, H. Zhou, W. Xu, X.-Z. Wang, W. Pedrycz, Interval dominance-based feature selection for interval-valued ordered data, IEEE Trans. Neural Netw. Learn. Syst. (2022) 1–15. https://doi.org/10.1109/TNNLS.2022.3184120. https://doi.org/10.1109/TNNLS.2022.3184120
[13] B. Sang, H. Chen, L. Yang, T. Li, W. Xu, Incremental feature selection using a conditional entropy based on fuzzy dominance neighborhood rough sets, IEEE Trans. Fuzzy Syst. 30 (6) (2022) 1683–1697. https://doi.org/10.1109/TFUZZ.2021.3064686. https://doi.org/10.1109/TFUZZ.2021.3064686
[14] C. Shu, Z. Mo, Rough fuzzy concept lattice and its properties, in: B.-Y. Cao, Z.-L. Liu, Y.-B. Zhong, H.-H. Mi (Eds.), Fuzzy Systems & Operations Research and Management, Springer International Publishing, Cham, 2016, pp. 163–170.
[15] X. Zhang, J. Hou, A novel rough set method based on adjustable-perspective dominance relations in intuitionistic fuzzy ordered decision tables, Int. J. Approximate Reasoning 154 (2023) 218–241. https://www.sciencedirect.com/science/article/pii/S0888613X23000014.
[16] D. Guo, W. Xu, Y. Qian, W. Ding, M-FCCL: memory-based concept-cognitive learning for dynamic fuzzy data classification and knowledge fusion, Inf. Fusion 100 (2023) 101962. https://www.sciencedirect.com/science/article/pii/S1566253523002786.
[17] Y. Huang, T. Li, C. Luo, H. Fujita, S.-j. Horng, Matrix-based dynamic updating rough fuzzy approximations for data mining, Knowl. Based Syst. 119 (2017) 273–283. https://www.sciencedirect.com/science/article/pii/S0950705116305196.
[18] S. Roy, P. Shivakumara, N. Jain, V. Khare, A. Dutta, U. Pal, T. Lu, Rough-fuzzy based scene categorization for text detection and recognition in video, Pattern Recognit. 80 (2018) 64–82. https://www.sciencedirect.com/science/article/pii/S0031320318300669.
[19] F. Zafar, M. Akram, A novel decision-making method based on rough fuzzy information, Int. J. Fuzzy Syst. 20 (3) (2018) 1000–1014. https://doi.org/10.1007/s40815-017-0368-0. https://doi.org/10.1007/s40815-017-0368-0
[20] S. Vluymans, L. D'eer, Y. Saeys, C. Cornelis, C. Cornelis, H.S. Nguyen, S.K. Pal, A. Skowron, W. Wu, Applications of fuzzy rough set theory in machine learning: a survey, Fundam. Inform. 142 (1–4) (2015) 53–86. https://doi.org/10.3233/FI-2015-1284. https://doi.org/10.3233/FI-2015-1284
[21] D. Liu, A study on information fusion methodology based on rough set and neural network, in: The 5th International Conference on New Trends in Information Science and Service Science, 1 2011, pp. 107–110.
[22] W. Wei, J. Liang, Information fusion in rough set theory : an overview, Inf. Fusion 48 (2019) 107–118. https://www.sciencedirect.com/science/article/pii/S1566253518300071.
[23] P. Zhang, T. Li, G. Wang, C. Luo, H. Chen, J. Zhang, D. Wang, Z. Yu, Multi-source information fusion based on rough set theory: a review, Inf. Fusion 68 (2021) 85–117. https://www.sciencedirect.com/science/article/pii/S1566253520304085.
[24] J. Uddin, R. Ghazali, H.A. Jemal, H. Shah, N.A. Husaini, A. Zeb, Rough set based information theoretic approach for clustering uncertain categorical data, PLOS ONE 17 (5) (2022) e0265190. https://doi.org/10.1371/journal.pone.0265190. https://doi.org/10.1371/journal.pone.0265190
[25] C.E. Shannon, A mathematical theory of communication, The Bell Syst. Tech. J. 27 (3) (1948) 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
[26] T. Masters, Information and Entropy, Apress, Berkeley, CA, 2018, pp. 417–507. https://doi.org/10.1007/978-1-4842-3336-8_9. https://doi.org/10.1007/978-1-4842-3336-8_9
[27] Z. Zou, Y. Yun, J. Sun, Entropy method for determination of weight of evaluating indicators in fuzzy synthetic evaluation for water quality assessment, J. Environ. Sci. 18 (5) (2006) 1020–1023. https://www.sciencedirect.com/science/article/pii/S1001074206600326.
[28] Q. Hu, M. Guo, D. Yu, J. Liu, Information entropy for ordinal classification, Sci. China Inf. Sci. 53 (6) (2010) 1188–1200. https://doi.org/10.1007/s11432-010-3117-7. https://doi.org/10.1007/s11432-010-3117-7
[29] Q. Zhang, Z. Zhang, X. Jin, W. Zeng, S. Lou, X. Jiang, Z. Zhang, Entropy-based method for evaluating spatial distribution of form errors for precision assembly, Precis. Eng. 60 (2019) 374–382. https://www.sciencedirect.com/science/article/pii/S0141635919300881.
[30] M. Frank, J. Kusch, J. Wolters, Entropy-based methods for uncertainty quantification of hyperbolic conservation laws, in: M.L. Muñoz-Ruiz, C. Parés, G. Russo (Eds.),

Recent Advances in Numerical Methods for Hyperbolic PDE Systems, Springer International Publishing, Cham, 2021, pp. 29–56.
[31] M. Li, X. Zhang, Information fusion in a multi-source incomplete information system based on information entropy, Entropy 19 (11) (2017) 570. https://doi.org/10.3390/e19110570. https://doi.org/10.3390/e19110570
[32] W. Xu, M. Li, X. Wang, Information fusion based on information entropy in fuzzy multi-source incomplete information system, Int. J. Fuzzy Syst. 19 (4) (2017) 1200–1216. https://doi.org/10.1007/s40815-016-0230-9. https://doi.org/10.1007/s40815-016-0230-9
[33] X. Zhang, X. Chen, W. Xu, W. Ding, Dynamic information fusion in multi-source incomplete interval-valued information system with variation of information sources and attributes, Inf. Sci. 608 (2022) 1–27. https://www.sciencedirect.com/science/article/pii/S0020025522006478.
[34] X. Zhang, D. Guo, W. Xu, Two-way concept-cognitive learning with multisource fuzzy context, Cognit. Comput. (2023). https://doi.org/10.1007/s12559-023-10107-w. https://doi.org/10.1007/s12559-023-10107-w
[35] W. Xu, Y. Pan, X. Chen, W. Ding, Y. Qian, A novel dynamic fusion approach using information entropy for interval-valued ordered datasets, IEEE Trans. Big Data 9 (3) (2023) 845–859. https://doi.org/10.1109/TBDATA.2022.3215494. https://doi.org/10.1109/TBDATA.2022.3215494
[36] L. Yang, W. Xu, X. Zhang, B. Sang, Multi-granulation method for information fusion in multi-source decision information system, Int. J. Approximate Reasoning 122 (2020) 47–65. https://www.sciencedirect.com/science/article/pii/S0888613X20301523.
[37] S. Han, L. Chen, Z. Zhang, J. Li, A new information filling technique based on generalized information entropy, Int. J. Comput. Commun. & Contr. 9 (2014) 172–186. https://api.semanticscholar.org/CorpusID:56387559.
[38] W. Qian, S. Yu, J. Yang, Y. Wang, J. Zhang, Multi-label feature selection based on information entropy fusion in multi-source decision system, Evol. Intell. 13 (2) (2020) 255–268. https://doi.org/10.1007/s12065-019-00349-9. https://doi.org/10.1007/s12065-019-00349-9
[39] X. Zeng, X. Xiong, Z. Luo, Grid fault diagnosis based on information entropy and multi-source information fusion, Int. J. Electron. Telecommun. 67 (2) (2021) 143–148. https://api.semanticscholar.org/CorpusID:245149275.
[40] Y. Pan, W. Xu, Q. Ran, An incremental approach to feature selection using the weighted dominance-based neighborhood rough sets, Int. J. Mach. Learn. Cybern. 14 (4) (2023) 1217–1233. https://doi.org/10.1007/s13042-022-01695-4. https://doi.org/10.1007/s13042-022-01695-4
[41] J. Hu, C. Yan, X. Liu, Z. Li, C. Ren, J. Zhang, D. Peng, Y. Yang, An integrated classification model for incremental learning, Multimed. Tools Appl. 80 (11) (2021) 17275–17290. https://doi.org/10.1007/s11042-020-10070-w. https://doi.org/10.1007/s11042-020-10070-w
[42] G.M. van de Ven, T. Tuytelaars, A.S. Tolias, Three types of incremental learning, Nature Mach. Intell. 4 (12) (2022) 1185–1197. https://doi.org/10.1038/s42256-022-00568-3. https://doi.org/10.1038/s42256-022-00568-3
[43] M. De Lange, A. Rahaf, M. Marc, P. Sarah, J. Xu, L. Ales, S. Gregory, T. Tinne, A continual learning survey: defying forgetting in classification tasks, IEEE Trans. Pattern Anal. Mach. Intell. 44 (7) (2022) 3366–3385. https://doi.org/10.1109/TPAMI.2021.3057446. https://doi.org/10.1109/TPAMI.2021.3057446
[44] I. Saadi, B. Farooq, A. Mustafa, J. Teller, M. Cools, An efficient hierarchical model for multi-source information fusion, Expert Syst. Appl. 110 (2018) 352–362. https://www.sciencedirect.com/science/article/pii/S0957417418303646.
[45] X. Che, J. Mi, D. Chen, Information fusion and numerical characterization of a multi-source information system, Knowl. Based Syst. 145 (2018) 121–133. https://www.sciencedirect.com/science/article/pii/S095070511830008X.
[46] Y. Zhu, D. Liu, Y. Li, X. Wang, Selective and incremental fusion for fuzzy and uncertain data based on probabilistic graphical model, J. Intell. Fuzzy Syst. 29 (6) (2015) 2397–2403. https://doi.org/10.3233/IFS-151939. https://doi.org/10.3233/IFS-151939
[47] Y. Huang, T. Li, C. Luo, H. Fujita, S.-J. Horng, Dynamic fusion of multisource interval-valued data by fuzzy granulation, IEEE Trans. Fuzzy Syst. 26 (6) (2018) 3403–3417. https://doi.org/10.1109/TFUZZ.2018.2832608. https://doi.org/10.1109/TFUZZ.2018.2832608
[48] X. Zhang, X. Huang, W. Xu, Matrix-based multi-granulation fusion approach for dynamic updating of knowledge in multi-source information, Knowl. Based Syst. 262 (2023) 110257. https://www.sciencedirect.com/science/article/pii/S0950705123000072.