Information Fusion 68 (2021) 85–117

Contents lists available at ScienceDirect

Information Fusion

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

Multi-source information fusion based on rough set theory: A review

Pengfei Zhang [a,b,c], Tianrui Li [a,b,c,*], Guoqiang Wang [a,b,c], Chuan Luo [d], Hongmei Chen [a,b,c], Junbo Zhang [b,e,f], Dexian Wang [a,b,c], Zeng Yu [a,b,c]

[a] School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, PR China
[b] Institute of Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, PR China
[c] National Engineering Laboratory of Integrated Transportation Big Data Application Technology, Southwest Jiaotong University, Chengdu 611756, PR China
[d] College of Computer Science, Sichuan University, Chengdu 610065, PR China
[e] JD Intelligent Cities Business Unit, JD Digits, Beijing, PR China
[f] JD Intelligent Cities Research, Beijing, PR China

ARTICLE INFO

Keywords:
Multi-source information fusion
Rough set theory
Homogeneous and heterogeneous MSIF
Multi-view rough sets
Parallel computing
Incremental learning
Cluster ensembles

ABSTRACT

Multi-Source Information Fusion (MSIF) is a comprehensive and interdisciplinary subject, and is referred to as, multi-sensor information fusion which was originated in the 1970s. Nowadays, the types and updates of data are becoming more multifarious and frequent, which bring new challenges for information fusion to deal with the multi-source data. Consequently, the construction of MSIF models suitable for different scenarios and the application of different fusion technologies are the core problems that need to be solved urgently. Rough set theory (RST) provides a computing paradigm for uncertain data modeling and reasoning, especially for classification issues with noisy, inaccurate or incomplete data. Furthermore, due to the rapid development of MSIF in recent years, the methodologies of learning under RST are becoming increasingly mature and systematic, unveiling a framework which has not been mentioned in the literature. In order to better clarify the approaches and application of MSIF in RST research community, this paper reviews the existing models and technologies from the perspectives of MSIF model (i.e., homogeneous and heterogeneous MSIF model), multi-view rough sets information fusion model (i.e., multi-granulation, multi-scale and multi-view decisions information fusion models), parallel computing information fusion model, incremental learning fusion technology and cluster ensembles fusion technology. Finally, RST based MSIF related research directions and challenges are also covered and discussed. By providing state-of-the-art understanding in specialized literature, this survey will directly help researchers understand the research developments of MSIF under RST.

1. Introduction

The methodology of Multi-Source Information Fusion (MSIF) is based on multi-sensor information fusion, which can achieve more significant information with higher accuracy than a single source (or a single sensor). In the late 1970s, the word of fusion based on the comprehensive meaning of multi-source information began to appear in various publications. Since then, the theories and technologies of MSIF have developed rapidly, as an independent discipline, which has been successfully applied in military command automation system, strategic warning and defense system, multi-target tracking and identification. Moreover, MSIF is also gradually radiating to remote sensing monitoring, medical diagnosis, electronic commerce, wireless communication and fault diagnosis and other civilian fields [1–4].

The amount of data produced in the world each year is rising at a rate of thirty percent [5]. Data is produced by everything around us via social media exchange, and transmitted by all kinds of networks, sensors, and mobile devices. Data acquiring is no longer limited to a single data source with the full apperceive of information in a Big Data environment. The storage and description of data appears in the form of multiple sources. Various information of knowledge structure are implied in the relationships among data samples from different data sources, which express information among data samples from multiple perspectives. Fortunately, the fundamental principle of MSIF is to make full use of multiple information sources. According to specific criteria, we can combine multiple sources of information with constraints of spatial redundancy, temporal redundancy or complementary information. MSIF is studied widely in different real-life applications and diverse theories and methods have been used in MSIF. For instance, Zhang et al. proposed a multiple-metric learning algorithm to learn jointly

* Corresponding author.
E-mail addresses: feifeihappy55@163.com (P. Zhang), trli@swjtu.edu.cn (T. Li), gqwang18@163.com (G. Wang), cluo@scu.edu.cn (C. Luo), hmchen@swjtu.edu.cn (H. Chen), msjunbozhang@outlook.com (J. Zhang), wangdexian2019@outlook.com (D. Wang), zyu@swjtu.edu.cn (Z. Yu).

https://doi.org/10.1016/j.inffus.2020.11.004
Received 25 August 2020; Received in revised form 5 November 2020; Accepted 10 November 2020
Available online 14 November 2020

a set of optimal homogeneous or heterogeneous metrics in order to fuse the data collected from multiple sensors for joint classification [6]. Dasarathy described a panoramic overview of MSIF in the field of multi-sensor from three complementary perspectives [7]. Yang et al. presented a mixed structure multi-mode data fusion based on D–S evidence theory and subjective Bayesian algorithm [8]. Cai et al. proposed a data fusion method for fault detection by utilizing Bayes network model, and applied the method to the parallel simulation to improve the diagnostic accuracy of ground source heat pump system [9]. Gravina et al. presented the motivations and advantages of multi-sensor data fusion and parameters affecting data fusion, and further discussed the design choices of data fusion affected by parameters at different levels [10]. Yager introduced a general framework for the multi-source data fusion processing and algorithm development [11], and investigated a novel monotonic set measure as a means of representing the multi-source fusion imperative [12]. Li et al. addressed a method of multi-source data clustering on account of homogeneous observations which are applied for multi-target detection from cluttered background with misdetection [13]. Saadi et al. presented a framework to allow intelligent merging of multiple data sources and can be applied to the urban transportation [14].

Rough Set Theory (RST), put forward by Pawlak [15,16], provides an effective mathematical tool for dealing with uncertainty. RST plays a critical role in extracting useful features, simplifying information processing, studying expression learning, and finding imprecise and uncertain information. At present, RST has been successfully applied to machine learning, decision analysis, process control, approximate reasoning, pattern recognition, data mining and other intelligent information processing fields [17–21]. From the perspective of data analysis, the main advantages of RST can be summarized as five points [15]. *(i)* It does not require any prior knowledge of the data. *(ii)* It has a certain ability to search the smallest collection of the data. *(iii)* It can evaluate the significance of the data. *(iv)* It allows the use of both qualitative and quantitative data, and *(v)* it can generate the set of decision rules from the data.

Due to the superiority of RST, increasing effort has been directed to the study of RST based data analysis, especially the field of data fusion. The fusion process of MSIF refers to deal with the data based on different models and approaches, which is essentially information fusion. The interest in the field of MSIF based on RST has significantly increased, leading to a growing number of techniques and methods. For example, Khan and Banerjee proposed a concept of multiple-source approximation systems based on Pawlak approximation spaces, which is a precursor for researchers to study MSIF in RST [22]. Li and Fei discussed a method of information fusion in wireless sensor network [23]. Liu et al. addressed a framework for performance test and evaluation of the multi-sensor data fusion with C^3I applications background [24]. Li et al. proposed a weighted fusion approach based on Granular Computing (GrC) and RST, which has been applied to road safety indicator analysis [25]. Yao et al. came up with a multi-source alert data understanding scheme for security semantic discovery [26]. In order to solve the fusion problem of multi-source information system (MsIS), Xu et al. presented the internal-confidence and external-confidence degrees to estimate the reliability of each information source [27], then they also considered the information fusion issue based on information entropy in fuzzy incomplete information systems [28]. Che et al. employed three approaches to address the information fusion and numerical characterization of the uncertain data [29]. Yang et al. developed a multi-granulation method for information fusion [30]. Sang et al. discussed the three kinds of multi-source decision methods based on the uncertainty of decision-making process [31]. In addition, Huang et al. addressed a new fusion method based on fuzzy information granulation, which can translate multi-source interval-valued data into trapezoidal fuzzy granules [32].

In literature, a detailed information fusion survey paper on RST [33] was published in 2019, which gives a general overview of the state-of-the-art approaches. It provides a hybrid-view about information fusion from five primary perspectives, objects, attributes, rough approximations, attribute reduction and decision making. The survey paper is comprehensive and can be a good introduction to information fusion. However, information fusion is essentially a process of integrating multi-source information from multilevel and multifaceted. In recent years, the number of proposals in the area of MSIF have significantly increased. There is therefore a gap in the current literature that requires a fuller picture of established on MSIF models and technologies. It is essential to review the past research focuses and give the most recent research trends about MSIF. Hence, this paper aims to present a comprehensive survey of the five major aspects of MSIF base on RST: MSIF fusion models (including homogeneous and heterogeneous MSIF model), multi-view rough sets information fusion model, parallel computing information fusion model, incremental learning and cluster ensembles fusion technologies, as shown in Fig. 1, and a discussion about the new trend of MSIF research.

The main contributions of this review can be summarized as follows.

(1) It perceptively summarizes MSIF research achievements and clusters the research into five categories: MSIF fusion models, multi-view rough sets information fusion model, parallel computing information fusion model, incremental learning and clustering ensembles fusion technologies (Fig. 1);

(2) It considers the MSIF models according to two perspectives: homogeneous and heterogeneous models;

(3) It combines several rough set models to study MSIF from different perspectives, and collectively refers to these models as a multi-view rough set model;

(4) It introduces the parallel computing model, which can accelerate the data fusion and be good at dealing with large-scale data via MapReduce framework;

(6) It uncovers an incremental learning fusion technique, such as Incremental Learning Information Fusion (ILIF) under new immigrating multiple objects, attributes and attributes values. Moreover, it identifies related research involving incremental learning;

(7) It reviews cluster ensembles fusion technologies based RST, as well as the methods of roughness and rough k-means for clustering. In addition, it emphasizes the concepts and applications of three-way clustering;

(8) It suggests several emerging research topics and potential research directions in this area.

2. Problem description

This section first gives the basic definition of MSIF in Section 2.1. Then, in Section 2.2, the general application mechanism of RST in MSIF is introduced.

2.1. The definition of MSIF

The subject of MSIF has been studied for decades, but there is no generally accepted definition. Pan and Han et al. summarized the research of JDL (Joint Directors of Laboratories) and gave the following definition.

Definition 2.1 (*[1,34]*)**.** Multi-Source Information Fusion (MSIF), also known as multi-sensor fusion, is a multi-level and multi-faceted process, including the detection, correlation, combination and estimation of multi-source data, so as to improve the accuracy of state and identity estimation, as well as timely and complete evaluation of the ultimate degree of target situation and threat.

It should be noted that the sensor in this definition is generalized, not only refers to a variety of sensor systems, but also a variety of information acquisition systems, and even including human or animal perception systems. In addition, the research of fusion method in this paper is aimed at data processing, so it can also be called data fusion.

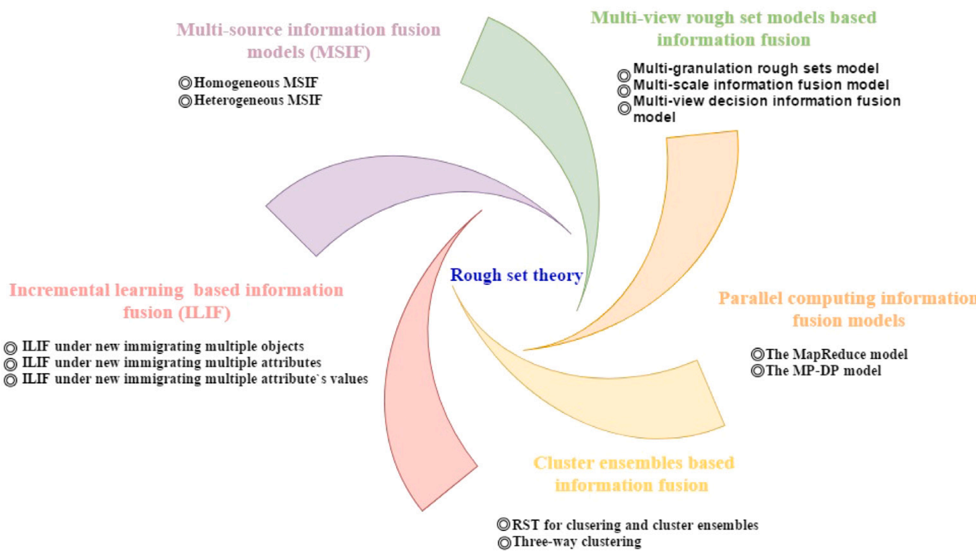

Fig. 1. The framework of MSIF in rough set theory.

Han et al. suspected that there might be three factors restricting the further development of information fusion [34]:

(1) The highly heterogeneous information types and vague attributes of content.

(2) The inherent complexity of multi-source information and multi-task.

(3) There is currently no unified mathematical tool to uniformly describe and deal with such complex problems.

2.2. Some applications of rough sets in MSIF

RST presents a novel paradigm of intelligent computing and been applied to the area of MSIF. At present, it becomes one of the important theories to promote the development of information fusion technology. RST can regard knowledge as the partition of the universe, so that knowledge is granularity, and the granularity of knowledge is the reason why certain concepts cannot be accurately expressed using existing knowledge. Then, by introducing the indiscernibility relation as the theoretical basis of RST, the lower and upper approximations are defined to describe the uncertain target concept.

Based on RST and researches of MSIF, the corresponding applications of software support systems are developed. For instance: LERS[1] system [35], KDD-R system [36], Rough DAS and Rough Class [37] and Rough Enough [38]. In addition to the above-mentioned software systems, many successful examples emerge in the practical applications of RST, such as stock market analysis [39], fault diagnosis [40], bankruptcy assessment [41], activity-based modeling [42], medical diagnosis [43–45], etc. To sum up, RST has its unique advantages in information fusion from various data models (i.e., information systems). Because of this, it has been paid more attention by many researchers in the last ten years and achieved some successes. Nowadays, rough sets have made breakthroughs in data fusion models and technologies. It is mainly reflected in the generalization of rough set models, the establishment of data models and the application of fusion technologies. What is more, RST is used as the theoretical basis for applying to MSIF, via discussing different MSIF models and related fusion technologies as the motivation for research in this paper.

The remaining part of this paper is organized as follows. In Section 3, the basic concepts of RST are given. Section 4 presents MSIF model in RST. Section 5 offers multi-view rough sets model based information fusion. Parallel computing model based information fusion is reported in Section 6. Section 7 presents incremental learning based information fusion. Section 8 summaries related research based RST concerning the cluster ensembles fusion. Section 9 is devoted to a comprehensive analysis of main findings and future research directions.

3. Preliminary concepts on RST

In this paper, suppose that $U = \{x_1, x_2, \dots, x_n\}$ is an universe (non-empty finite set), 2^U is recorded as the collection formed by all subsets of U and $|X|$ indicates the cardinality of $X \in 2^U$.

3.1. Information systems and Pawlak rough sets

Definition 3.1 (*[15]*). Consider that U is an object set and A is an attribute set. Suppose that U and A are finite sets. Then the pair (U, A) is called an information system (IS), if each attribute $a \in A$ determines an information function $a : U \to V_a$, where $V_a = \{a(x) : x \in U\}$ is the set of information function values of the attribute a. If (U, A) is an IS and $A = C \cup \{d\}$ where C is a conditional attribute set and d is a decision attribute set. Then $(U, C \cup \{d\})$ is referred to as a decision information system (DIS).

Suppose that (U, C) is an IS and $P \subseteq C$. Then a binary relation on U can be defined as

$$R_P = \{(x, y) \in U \times U : \forall\, a \in P,\ a(x) = a(y)\}. \tag{3.1}$$

Clearly, R_P is an equivalence relation on U and $R_P = \bigcap_{a \in P} R_{(\{a\})}$. Denote

$$[x]_P = \{y \in U : (x, y) \in R_P\}.$$

Then $[x]_P$ is the equivalence class of the object x under the equivalence relation R_P. Therefore, an IS is essentially a database and each attribute determines a knowledge.

Assume that R_P (for convenience, denoted as R) is an equivalence relation on U. Then R partitions the universe U into a family of disjoint subsets called equivalence classes. For an equivalence relation R, the equivalence class including x is denoted by:

$$[x]_R = \{y \in U : xRy\}. \tag{3.2}$$

$\mathbf{R}$ is called the family of all equivalence classes, which is denoted as follows:

$$U/\mathbf{R} = \{[x]_R : x \in U\}, \tag{3.3}$$

[1] https://people.eecs.ku.edu/~jerzygb/LERS.html.

where $R \in \mathbf{R}$, the pair $(U, \mathbf{R})$ can be referred to an IS (or knowledge base).

Definition 3.2 (*[16]*). Let $(U, \mathbf{R})$ be an IS, $R \in \mathbf{R}$. Then, the lower and upper approximations of $X \in 2^U$, denoted by $\underline{R}(X)$ and $\overline{R}(X)$, respectively, are defined as

$$\underline{R}(X) = \{x \in U : [x]_R \subseteq X\}, \quad \overline{R}(X) = \{x \in U : [x]_R \cap X \neq \emptyset\}. \tag{3.4}$$

According to Pawlak's rough set model. For any $X \in 2^U$, the universe U can be divided into three regions (positive, boundary and negative regions) as follows:

$$\begin{cases} POS(X) = \underline{R}(X), \\ BND(X) = \overline{R}(X) - \underline{R}(X), \\ NEG(X) = U - \overline{R}(X). \end{cases} \tag{3.5}$$

Definition 3.3 (*[16]*). Let $(U, \mathbf{R})$ be an IS, $P \subseteq R$ $(R \in \mathbf{R})$. For any $X \in 2^U, x \in U$, the rough membership of x in X is defined as

$$\mu_X^P(x) = \frac{|[x]_P \cap X|}{|[x]_P|} \tag{3.6}$$

3.2. Accuracy of approximation and rough accuracy

Definition 3.4 (*[16]*). Let $(U, \mathbf{R})$ be an IS, $R(R \in \mathbf{R})$ be an equivalence relation on U. For any $X \in 2^U$, the accuracy of the approximation is defined by

$$\alpha(X) = \frac{|\underline{R}(X)|}{|\overline{R}(X)|} \tag{3.7}$$

Definition 3.5 (*[46]*). Let $(U, \mathbf{R})$ be an IS, $R(R \in \mathbf{R})$ be an equivalence relation on U. For any $X \in 2^U$, the rough accuracy of the approximation is defined by

$$\begin{aligned} \rho_R(X) &= 1 - \frac{|\underline{R}(X) \cap \overline{R}(X)|}{|\underline{R}(X) \cup \overline{R}(X)|} \\ &= 1 - \frac{|\underline{R}(X)|}{|\overline{R}(X)|} = 1 - \alpha(X) \end{aligned} \tag{3.8}$$

3.3. Reduct of indiscernibility relation

Definition 3.6 (*[47]*). Let $(U, \mathbf{R})$ be an IS. For any $R \in \mathbf{R}$,

(i) $ind(\mathbf{R}) = ind(\mathbf{R} - R)$, then R is called dispensable in $\mathbf{R}$, otherwise R is indispensable in $\mathbf{R}$;

(ii) if each $R \in \mathbf{R}$ is indispensable in $\mathbf{R}$, then the family of the equivalence relation $\mathbf{R}$ is called independent, otherwise $\mathbf{R}$ is dependent.

Definition 3.7 (*[47]*). Let $(U, \mathbf{R})$ be an IS, $Q \subseteq P \in \mathbf{R}$,

(i) $ind(Q) = ind(P)$ and Q is independent, then Q is called a reduct of P;

(ii) $CORE(P)$ is called the core of P, which means the set of all indispensable relations in P.

Theorem 3.8 (*[47]*). *Let $(U, \mathbf{R})$ be an IS, $P \subseteq \mathbf{R}$,*

$$CORE(P) = \cap RED(P) \tag{3.9}$$

where $\cap RED(P)$ is the intersection of all reducts of P.

4. MSIF models in rough set theory

Nowadays, the high fusion of Cyber–Physical Human System (CPHS) has triggered the explosive growth of data scale and the high complexity of data models. The world has entered the era of networked Big Data [48,49], which has the characteristics of five Vs, i.e., Volume, Velocity, Variety, Value and Variety [50]. Big data is a fast growing field both from an application and from a research point of view. It is worth noting that the most significant thing is to extract useful knowledge from Big data. Aiming at this problem, Li et al. introduced a useful solution that named "PICKT" on Big data analysis based on the theories of GrC and rough sets [50]. In addition, due to the types of data are diverse in Big data, then attributes in information systems will generally present different types, such as qualitative value, quantitative value, discrete value, continuous value, missing value and so on. More specifically, it can also be classified into categorical attribute, numerical attribute, set-valued attribute, interval-valued attribute, etc. These data types make up a variety of data models in the form of information tables, which can be seen as a data-driven research method for the purpose of information fusion. Therefore, the study of various MSIF models will be a meaningful topic.

The characteristics of information are processed from different types of data, which are generally divided into two kinds of types: one is the data type of homogeneous, and the other is the data type of heterogeneous. In reality, information fusion is a formal framework. The process of information fusion is to employ intelligent computing methods and technical tools to synthesize information from different sources, which can obtain high-quality and useful information. In RST, MSIF models of complex data have been studied, such as homogeneous MSIF models (i.e., set-valued information systems (CISs), multiset-valued information systems (MvISs) and probabilistic set-valued information systems (PSvISs)) and heterogeneous MSIF models (i.e., composite information systems (CISs), multi-source information systems (MsISs) and multi-source heterogeneous information systems (MsHISs)). Typically, an information table (i.e., a data table) is described as the forms of labeling objects by columns and attributes by rows [15]. Undoubtedly, information is the basis for acquiring knowledge. One can acquire information through different channels, and discover and acquire knowledge from information systems, which is a specific knowledge discovery process. An information system is a database system with the relationship between objects and attributes. This kind of database system implies the relationship between objects and attributes through data. The final knowledge pattern is described by attributes, which has a clear and intuitive meaning. In this section, we interpret the models of MSIF in RST in terms of two different perspectives, i.e., homogeneous MSIF and heterogeneous MSIF. In addition, relevant research work in the literature is summarized in the Table 1.

4.1. Homogeneous MSIF

In an information system, the same attribute is described by multiple values, and expressed in the form of a set in an information system. For example, the information of a person who has multiple titles, it can be described by a set, namely, title $=\{CEO, Chair, Dr.\}$. Then, we say that "title" with the homogeneous data, which is described via set-valued feature. In practical application, objects are often described as having multiple values with the homogeneous data. Therefore, in this subsection, we mainly discuss the MSIF models with homogeneous data from the perspective of RST, as shown in Fig. 2.

From Fig. 2, the homogeneous MSIF in RST is divided into three types of models, i.e., set-valued information systems, multiset-valued information systems and probabilistic set-valued information systems.

4.1.1. Set-valued information systems (SvISs)

In practical applications, e.g. risk evaluation [78,79], multi-source information fusion [12] and movie ratings [80], due to insufficient information and diversity of data sources, the set-valued data often appears in information systems.

A SvIS is a significant generalized model of a single-valued information system. Yao and Noroozi first analyzed a framework for set-based computations [81]. This framework is particularly useful in situations where it is difficult to obtain a precise value of certain parameter, or where set-valued attributes play an important role. Meanwhile, the attribute value of an object is the subset of the domains of attributes in an incomplete information system, which can induce a SvIS [82].

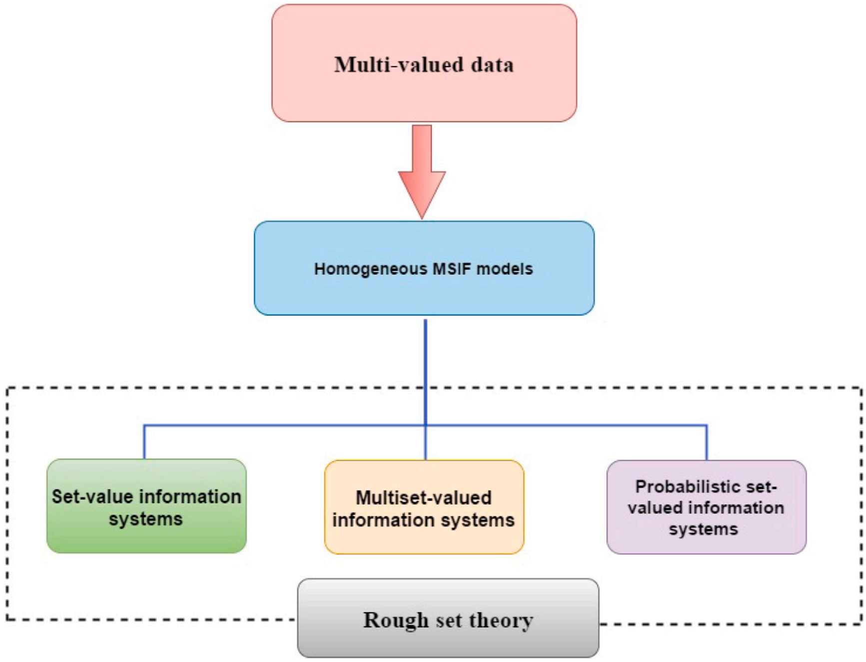

Fig. 2. The models of homogeneous MSIF in RST.

Table 1
Summary of MSIF models and its applications.

MSIF models.	Information fusion models based on RST.	Application of the fusion strategy.	References
Homogeneous MSIF.	SvISs.	Optimization decision rules.	[51,52].
	SvISs.	Uncertainty measurement.	[53,54].
	SvISs.	Incremental fusion.	[55–58].
	SvISs.	Attribute reduction.	[59–62].
	PSvISs.	Incremental fusion.	[63].
	PSvISs.	Uncertainty measurement.	[64].
	MvISs.	Three-way decisions.	[65].
Heterogeneous MSIF.	CISs.	Incremental fusion.	[66] [67–70].
	HISs.	Uncertainty measurement.	[71,72].
	HISs.	Three-way decisions.	[45,73].
	MsISs.	Fusion of approximation space.	[30,74,75].
	MsISs.	Uncertainty measurement.	[27,28].
	MsISs.	Optimization decision rules.	[31].
	MsHISs.	Group decision making.	[76].
	MsHISs.	Agent evaluation.	[77].

Table 2
A SvDIS.

U	Price(a_1)	Mileage(a_2)	Size(a_3)	Max-Speed(a_4)	d
x_1	{mid}	{high}	{full}	{high, mid}	Good
x_2	{high, low}	{mid, low}	{compact}	{high, mid, low}	Poor
x_3	{high, low}	{high, mid}	{full}	{high}	Good
x_4	{high, mid}	{high}	{compact}	{high, mid}	Excel
x_5	{high}	{high, mid}	{full}	{high, mid}	Good
x_6	{high, low}	{mid}	{compact}	{high}	Good

Definition 4.1 ([51,81]). Suppose that U denotes a set of objects, C denotes a set of condition attributes and d is called decision attribute with $C \cap \{d\} = \emptyset$. If for any $a \in C$ and $x \in U$, $a(x)$ is a set, then $(U, C \cup \{d\})$ is called a set-valued decision information system (SvDIS).

Example 4.2. Table 2 represents a SvDIS.

In 2004, Zhang et al. introduced the tolerance relation in a SvIS is defined as follows [83].

Definition 4.3 ([83]). Let $(U, C \cup \{d\})$ be a SvDIS in which $B \subseteq C$. Then the tolerance relation T_B is denoted as

$$T_B = \{(x, y) \in U^2 | \forall\ a \in B, f(x, a) \cap f(y, a) \neq \emptyset\} \tag{4.1}$$

Definition 4.4 ([51]). Let $(U, C \cup \{d\})$ be a SvDIS. If for any $X \subseteq U$ and $B \subseteq C$, then the lower and upper approximations of X in terms of the tolerance relation T_B is denoted as

$$\overline{T_B}(X) = \{x \in U | T_B(x) \cap X \neq \emptyset\}, \quad \underline{T_B}(X) = \{x \in U | T_B(x) \subseteq X\}, \tag{4.2}$$

where $T_B(x)$ is referred to as tolerance class of x under tolerance relation T_B, which is denoted by $T_B(x) = \{y \in U | (x, y) \in T_B$.

Definition 4.5 ([68]). Let $(U, C \cup \{d\})$ be a SvDIS. $\forall\ x, y \in U$, $\forall\ a \in C$, the distance between $a(x)$ and $a(y)$ is defined by

$$d(a(x), a(y)) = 1 - \frac{|a(x) \cap a(y)|}{M_a}, \tag{4.3}$$

where $M_a = max\{|a(x)| : x \in U\}$.

In recent years, there have been many studies on SvISs. For example, William proposed the feature vector to represent the values of a feature with a set of strings, then showed many real-word application problems can be efficiently represented by set-valued features [84]. Guan and Wang defined tolerance relation and used the maximal tolerance classes to classify the universe, and adopted the optimal decision rules from a set-valued decision information system [51]. Qian et al. presented a dominance relation based on RST for two types of set-valued ordered information systems [52]. Zhang et al. introduced the updating of rough set approximations with incremental approaches for the relation matrix in a set-valued decision information system [55]. Chen and Li studied principles for incrementally updating approximations in SvISs while attributes and objects are added [56]. Dai and Tian put forward an approach for attribute reduction in SvISs based on discernibility relation [59], and defined four types of knowledge measures in a SvIS [53]. To demonstrate the attribute reduction of a dynamic SvIS, Lang et al. addressed an incremental approach to attribute reduction in SvIS [85]. Bao and Yang investigated a method of attribute reduction in set-valued ordered fuzzy decision system that considers attributes with preference ordered domains and fuzzy decision attributes [60]. According to the dynamic variation of criteria values in the set-valued decision system which depends on the knowledge updating, Luo et al. presented the updating properties for dynamic maintenance of approximations in SvISs [58]. Zhuang et al. proposed a multi-granulation dominance relation model based on SvISs, and showed the proposed model to apply multi-source information systems [86]. To solve the issue of fuzzy rough approximations for set-valued data, Wei et al. proposed two types of fuzzy rough approximation which are applied to the corresponding relative positive region reduction [61]. To integrate quantitative rough sets and dominance-based rough sets, Zhang and Yang came up with a general framework of feature selection and approximate reasoning for large-scale set-valued information systems [62].

4.1.2. Probabilistic set-valued information systems (PSvISs)

From the aforementioned subsection, the SvIS has been investigated extensively. In practice, a generalized information system based on a SvIS has emerged, which is called a probabilistic set-valued information system (PSvIS) [63]. There are many probabilistic data models proposed by Barbará et al. from the early 1990s [87,88]. However, in this subsection, we discuss an extended type of SvIS in [52], i.e., interpreted conjunctively. For example, the attribute a is showed as "speaking a language". A set $a(x)=\{English, Polish, French\}$ displays the attribute value. It can be interpreted as: One speaks English, Polish, and French. In order to better express this phenomenon, namely, the ability of languages by describing a set with a discrete probability distribution can be distinguished. For example, $a(x)=\{\frac{English, Polish, French}{0.85, 0.14, 0.01}\}$. It shows that the object x can speak English fluently, a little Polish, but hardly French.

Huang et al. first proposed the definition of a PSvIS as follows [63].

Definition 4.6 (*[63]*). A PSvIS can be denoted as a sextuple $=\{U, AT=A\cup D, V=V_A\cup V_D, f, \sigma, P\}$, where U is universe and AT is the set of attribute, and A and D are the set of condition attribute and decision attribute values with $(A\cup D=\emptyset)$, respectively. $V=V_A\cup V_D$ is the domain of attributes set AT. $f: U\times A\to 2^{V_A}$ is a set-valued mapping, and for each $x\in U$ and $d\in D$, $f: U\times D\to V_D$ is an information function such that $f(x,d)\in V_d$. σ is a sigma-algebra in V_a. P is called the probability distribution on σ and it satisfies $P(f_i(x,a))\geq 0$ and $\sum_{i=1}(f_i(x,a))=1$, where $f_i(x,a)\in f(x,a)$ for each $x\in U$ and $a\in A$.

For the sake of simplicity, the PSvIS=$\{U, AT=A\cup D, V=V_A\cup V_D, f, \sigma, P\}$ can be denoted (U,A). Then, they also presented the λ-tolerance relation based on Bhattacharyya distance (it can measure the similarity of two discrete probability distributions) in a PSvIS.

Definition 4.7 (*[63]*). Let (U,A) be a PSvIS. Given a non-negative threshold λ_a, then the λ-tolerance relation w.r.t. $a\in A$ is defined by

$$BD_a^{\lambda}=\{(x,y)\in U^2 | BD_a(x,y)\leq\lambda_a,\tag{4.4}$$

where $BD_a(x,y) = -ln(\sum_{k=1}^{K}\sqrt{P(f_k(x,a)P(f_k(y,a)))})$ is the Bhattacharyya distance and $P(f_k(x,a))$ displays the probability distribution of x under the attribute a. If for any $P\subseteq A$, the λ-tolerance relation can be defined as follows.

$$BD_P^{\lambda}=\{(x,y)\in U^2 | BD_p(x,y)\leq\lambda_p \ for\ each\ p\in P\}=\bigcap_{p\in P}BD_p^{\lambda}.\tag{4.5}$$

Similar distribution-based the Bhattacharyya distance method, Xie et al. proposed the Hellinger distance for measuring the similarity degree between two probability distribution [64].

Definition 4.8 (*[64]*). Let $A=\{a_1,a_2,\ldots,a_m\}$. For any $a\in A$, let $P_{V_a}=\left\{\frac{u_1,u_2,\ldots,u_n}{p_1,p_2,\ldots,p_n}\right\}$. If for any i, $0\leq p_i\leq 1$ and $\sum_{i=1}^{n}p_i=1$, then P_{V_a} is called a probability distribution set over V_a.

Definition 4.9 (*[64]*). Let P_{V_a} and Q_{V_a} be two probability distributions sets over V_a. Then, the Hellinger distance between P_{V_a} and Q_{V_a} is defined by

$$HD(P_{V_a},Q_{V_a})=\frac{1}{\sqrt{2}}\sqrt{\sum_{i=1}^{n}(\sqrt{p_i}-\sqrt{q_i})^2},\tag{4.6}$$

where $P_{V_a}=\left\{\frac{u_1,u_2,\ldots,u_n}{p_1,p_2,\ldots,p_n}\right\}$, $Q_{V_a}=\left\{\frac{v_1,v_2,\ldots,v_n}{q_1,q_2,\ldots,q_n}\right\}$. Apparently, $HD(P_{V_a}, Q_{V_a})=\sqrt{1-\sum_{i=1}^{n}\sqrt{p_iq_i}}$.

Definition 4.10 (*[64]*). Let P_{V_a} and Q_{V_a} be two probability distributions sets over V_a. Then, the Hellinger similarity degree between P_{V_a} and Q_{V_a} is defined by

$$HS(P_{V_a},Q_{V_a})=\begin{cases}1, & P_{V_a}=Q_{V_a};\\ 1-HD(P_{V_a},Q_{V_a}), & P_{V_a}\neq Q_{V_a},\end{cases}\tag{4.7}$$

where $P_{V_a}=\left\{\frac{u_1,u_2,\ldots,u_n}{p_1,p_2,\ldots,p_n}\right\}$, $Q_{V_a}=\left\{\frac{v_1,v_2,\ldots,v_n}{q_1,q_2,\ldots,q_n}\right\}$.

Definition 4.11 (*[64]*). Let (U,A) be a SvIS. Then (U,A) is called a PSvIS, if for any $a\in A$,

$$a(x_1)\simeq a(x_2)\simeq\cdots\cdots\simeq a(x_n),\tag{4.8}$$

where "$\simeq$" means that different probabilities distributed under the same attribute value, in other words, the numerator is the same but the denominator is different.

Therefore, the above definition can be viewed as "approximately equal". The following example shows the features of this definition (see Table 3).

Example 4.12. Table 3 shows a PSvIS, where $U=\{x_1,x_2,\ldots,x_6\}$ represents object set, $A=\{a_1,a_2,a_3,a_4,a_5\}$ represents attribute set, and $V_{a_1}=\{0,1,2\}, V_{a_2}=\{1,3\}, V_{a_3}=\{1,2,3\}, V_{a_4}=\{2,3\}, V_{a_5}=\{0,2\}$.

In practical applications, there are various of forms for multi-valued homogeneous data. Nevertheless, the SvIS is the most common form. The researches on SvISs have made some progress in RST society (see Section 4.1.1). However, there are a few researches on PSvISs. For instance, Huang et al. first introduced the notion of a PSvIS, and presented the variable precision rough set model based on tolerance relation [63]. To establish the framework of granular computing and understand the essence of uncertainty, Xie et al. explored information structure and uncertainty measurement in PSvISs [64]. Furthermore, there is still much work to be expanded for PSvISs. In view of this, it is suggested to consider from the following three aspects:

Table 3
A PSvIS.

U	a_1	a_2	a_3	a_4	a_5
x_1	$\{\frac{0,1,2}{0.24,0.32,0.44}\}$	$\{\frac{1,3}{0.45,0.55}\}$	$\{\frac{1,2,3}{0.10,0.40,0.50}\}$	$\{\frac{2,3}{0.14,0.86}\}$	$\{\frac{0,2}{0.44,0.56}\}$
x_2	$\{\frac{0,1,2}{0.48,0.22,0.30}\}$	$\{\frac{1,3}{0.40,0.60}\}$	$\{\frac{1,2,3}{0.50,0.20,0.30}\}$	$\{\frac{2,3}{0.27,0.73}\}$	$\{\frac{0,2}{0.58,0.42}\}$
x_3	$\{\frac{0,1,2}{0.10,0.60,0.30}\}$	$\{\frac{1,3}{0.45,0.55}\}$	$\{\frac{1,2,3}{0.25,0.45,0.30}\}$	$\{\frac{2,3}{0.14,0.86}\}$	$\{\frac{0,2}{0.25,0.75}\}$
x_4	$\{\frac{0,1,2}{0.24,0.44,0.32}\}$	$\{\frac{1,3}{0.65,0.35}\}$	$\{\frac{1,2,3}{0.24,0.44,0.32}\}$	$\{\frac{2,3}{0.14,0.86}\}$	$\{\frac{0,2}{0.58,0.42}\}$
x_5	$\{\frac{0,1,2}{0.53,0.27,0.20}\}$	$\{\frac{1,3}{0.82,0.18}\}$	$\{\frac{1,2,3}{0.25,0.45,0.30}\}$	$\{\frac{2,3}{0.58,0.42}\}$	$\{\frac{0,2}{0.36,0.64}\}$
x_6	$\{\frac{0,1,2}{0.29,0.41,0.30}\}$	$\{\frac{1,3}{0.29,0.61}\}$	$\{\frac{1,2,3}{0.10,0.40,0.50}\}$	$\{\frac{2,3}{0.77,0.23}\}$	$\{\frac{0,2}{0.50,0.50}\}$

(1) From the standpoint of MSIF, how to construct different rough sets models based on PSvISs.

(2) From the standpoint of Big data, how to find application of scenarios in PSvISs according to the practicality.

(3) From the standpoint of irrelevant probability distributions, i.e., the first type of semantics of [52], how to analyze such data table is also an issue to be discussed in the future.

4.1.3. Multiset-valued information systems (MvISs)

In classical set theory, a set refers to a collective made up of various individuals with certain common characteristics. These individuals are referred to as elements, and a collection of elements is called a set. In general, there are no repeating elements in a set. However, in practical applications, it is possible to have some of the same elements in a set, such as, computer programming [89,90] and feature fusion [91]. In addition, the following example can also be illustrated. Seven patients who might have the flu are needed to be diagnosed by five experts. Each patient has varying degrees of symptoms, such as, $temperature = \{normal, high, very\ high\}$, $cough = \{normal, dry, spasmodic, bad\}$, $headache = \{yes, no\}$ and $musclepain = \{nopain, mild, moderate\}$. Then, a patient is diagnosed according to the four symptoms by five experts, and they may diagnose the same result for some symptoms. For the symptoms of cough, five experts diagnosed patient A as dry, spasmodic, dry, bad and dry. Therefore, these diagnose results by five experts can be expressed as a set S, i.e., $S = \{dry, spasmodic, dry, bad, dry\}$. We find that the "dry" appears three times in set A. In order to reasonably explain this phenomenon, the term a multiset (or multiple membership set, bag, heap, bunch, etc.) is proposed which shows a set is a collection of elements in which elements may occur more than once [92–94]. Moreover, $S = \{dry, spasmodic, dry, bad, dry\}$ can be written as $S = \{3/dry, 1/spasmodic, 1/bad\}$ or $S = \{0/normal, 3/dry, 1/spasmodic, 1/bad\}$. Formally, the multiset is defined as follows.

Definition 4.13 (*[92]*)**.** Given the universe X. A multiset or bag M drawn from X is represented by a function count C_M defined as $C_M : X \to \mathbb{N}$.

For convenience, $C_M(x)$ can be denoted by $M(x)$ $(x \in X)$. If $M(x) = m$, then x appears m times in M. We denoted it by $m/x \in M$ or $x \in^m M$. Given $X = \{x_1, x_2, \ldots, x_n\}$. If $M(x_i) = m_i$ $(i = 1, 2, \ldots, n)$, then M is denoted by $\{m_1/x_1, m_2/x_2, \ldots, m_n/x_n\}$, i.e.,

$$M = \{m_1/x_1, m_2/x_2, \ldots, m_n/x_n\}.$$

Definition 4.14 (*[65]*)**.** A multiset-valued information system (MvIS) can be represented as a quadruple

$$MvIS = (U, A, \{V_a | a \in A\}, \{F_a | a \in A\}), \tag{4.9}$$

where U is referred to as the universe, A is called a set of attributes, V_a is the set of all finite normal multiset values of attribute, and $F_a : U \to V_a$ is an information function which maps each object to exactly one multiset in V_a. For the sake of simplicity, an MvIS is written as (U, A).

Table 4
An MvIS.

U	a_1	a_2	a_3	a_4
x_1	{2/N,2/H,1/Vh}	{0/N,1/D,3/S,1/B}	{4/Yes,1/No}	{1/Np,1/Mi,3/Mo}
x_2	{2/N,3/H,0/Vh}	{0/N,3/D,1/S,1/B}	{5/Yes,0/No}	{0/Np,4/Mi,1/Mo}
x_3	{0/N,4/H,1/Vh}	{1/N,1/D,1/S,2/B}	{3/Yes,2/No}	{0/Np,3/Mi,2/Mo}
x_4	{0/N,2/H,3/Vh}	{2/N,1/D,1/S,1/B}	{5/Yes,0/No}	{1/Np,4/Mi,0/Mo}
x_5	{1/N,1/H,3/Vh}	{1/N,1/D,3/S,0/B}	{3/Yes,2/No}	{0/Np,4/Mi,1/Mo}
x_6	{0/N,5/H,0/Vh}	{1/N,1/D,3/S,0/B}	{3/Yes,2/No}	{1/Np,3/Mi,2/Mo}
x_7	{1/N,3/H,1/Vh}	{2/N,2/D,0/S,1/B}	{4/Yes,1/No}	{2/Np,1/Mi,1/Mo}

Definition 4.15 (*[65]*)**.** Let (U, A) be an MvIS where $U = \{x_1, x_2, \ldots, x_n\}$ and $A = \{a_1, a_2, \ldots, a_m\}$. $\forall\ x_i, x_j \in U$ and $a_k \in A$, the similarity degree between x_i and x_j under attribute a_k is defined by

$$\mathcal{R}_{a_k} = 1 - \frac{|\, \| F_{a_k}(x_i) \|_p - \| F_{a_k}(x_j) \|_p \,|}{\sqrt[p]{c}}, \tag{4.10}$$

where $c = \max_{i,k}\{|F_{a_k}(x_i)|\}$.

Example 4.16. Table 4 shows an MvIS, where $U = \{x_1, x_2, \ldots, x_7\}$ and $A = \{a_1, a_2, a_3, a_4\}$ express seven patients and four different symptoms, namely, temperature(a_1) = {N=normal, H=high, Vh=very high}, cough(a_2) = {N=normal, D=dry, S=spasmodic, B=bad}, headache(a_3) = {yes, no} and muscle pain(a_4) = {Np=no pain, Mi=mild, Mo=moderate}. Moreover, $a_1(x_1) = \{2/N, 2/H, 1/Vh\}$, $a_2(x_1) = \{0/N, 1/D, 3/S, 1/B\}$, $a_3(x_1) = \{4/Yes, 1/No\}$ and $a_4(x_1) = \{1/Np, 1/Mi, 3/Mo\}$ indicate the diagnostic results of patient x_1, where form four multisets, i.e., $\{N, N, H, H, Vh\}$, $\{D, S, S, S, B\}$, $\{Yes, Yes, Yes, Yes, No\}$ and $\{Np, Mi, Mo, Mo, Mo\}$, respectively.

The model of multiset-based information systems refer to information systems which information values are multiple sets. This is a kind of database describing the relationship between objects (samples) and attributes (features) and also a very broad concept. Some general information systems are special cases of MvISs. In fact, an MvIS is an information system formed by the fusion of multiple information systems with the participation of multiple experts, which reflects the idea of MSIF. Up to now, most of the literature of multiset are dealing with the mathematical properties [95–99]. In [65], Zhao and Hu proposed a 3WD method with decision-theoretic rough sets in MvISs. In future research, it is suggested to study the decision-making problems in MvISs.

There are three types of multi-source homogeneous models based-rough sets are introduced, i.e., SvISs, MvISs and PSvISs. The common feature of the three models is the feature of multi-source set-value, which is presented in the form of data table. It needs to be emphasized that the SvISs are the most common form of set-valued data, and the other two are its extensions. For example, suppose that a student takes five tests in each of three subjects, and the examination subjects are Physics, Math and French. Then, $a_1(x) = \{Physics, Math, French\}$ refers to the set of examination subjects, $a_2(x) = \{\frac{Physics, Math, French}{0.4, 0.3, 0.3}\}$ refers to the set of the probability of Physics, Math and French of passing the exam are 0.4, 0.3 and 0.3, respectively. And $a_3(x) = \{4/Physics, 2/Math, 3/French\}$ refers to the set of passing the exam of Physics, Math and French are four times, two times and three times, respectively. Based on the previous Definitions 4.1, 4.11 and 4.14, $a_1(x)$, $a_2(x)$ and $a_3(x)$ can be viewed separately as a function value of SvISs, MvISs and PSvISs. Therefore, the application scenarios of three models are different, however, we can analyze them using the same fusion processing techniques. For instance, in [86], it was based on granulation fusion idea to deal with SvISs. Zhang and Chen et al. used incremental learning fusion techniques to study the SvISs [55,56]. Cao et al. introduced a clustering method for categorical data with set-valued features [80]. As for PSvISs, Huang et al. and Xie et al. adopted incremental learning and granularity respectively to study their fusion mechanism [63] and [64]. And finally for MvISs, it was only studied in [65] as a form of data table. Namely, there is still much work to be investigated on MvISs, such as, the incremental learning mechanism and cluster ensemble fusion methods can be used to combine with MvISs.

4.2. Heterogeneous MSIF

In this section, heterogeneous MSIF in RST is introduced. In an information system, the attribute values can come from different sources and the type of data may not be the same. Information fusion in the data with different types of feature values, i.e., the heterogeneous (or mixed data), is especially of practical importance because such types of data sets widely exist in real life. For example, the video as a multimedia medium can be decomposed into multiple single-modal modes such as dynamic images, dynamic voices and dynamic texts, which contains three different types of data [100]. In addition, the fusion of multiple sensors also involves various types of data [101]. How to deal with and fuse different types of data is a hot topic. In general, the researchers initially discussed heterogeneous characteristics by describing the distance functions to express data types [102]. However, many researches define heterogeneous data as two types of data attributes, i.e., numerical and categorical attributes [103,104]. In what follows, we introduce two models of multi-source heterogeneous data based on RST from the perspective of MSIF.

4.2.1. Composite information systems (CISs)

In order to make the information system more diversified and specific, not just expressed as numerical attributes and categorical attributes, Zhang et al. first proposed a unified data model called a composite information system which contains multiple kinds of data, such as categorical, numerical, set-valued, interval-valued and so on [66]. Such rough set model can be better describe the distribution and characteristics of heterogeneous data in an information system.

Definition 4.17 (*[66]*). Let $CIS = (U, A, V, f)$ be a CIS, where

$$\begin{cases} U, \ \textit{the set of objects (universe)}. \\ A = \bigcup A_k, \ \textit{a union of attribute sets, where } A_k \textit{ is an attribute set} \\ \qquad \textit{with the same data type.} \\ V = \bigcup_{A_k \in A} V_{A_k}, \ \ V_{A_k} = \bigcup_{a \in A} V_a. \\ f : U \times A \to V_a, \ \textit{is an information function which maps each object} \\ \qquad \textit{to an attribute value in } V_a. \end{cases}$$

Definition 4.18 (*[66]*). Let (U, A, V, f) be a CIS. For any $x, y \in U$, and $P = \bigcup P_k \subseteq A, P_k \subseteq A_k$, then the composite relation CR_P can be defined by

$$CR_P = \{(x, y) | (x, y) \in \bigcap_{P_k \in P} R_{P_k}\}, \tag{4.11}$$

where $R_{P_k} \subseteq U \times U$ is an indiscernibility relation defined by an attribute set P_k on U.

Definition 4.19 (*[66]*). Let $CIS = (U, A, V, f)$ be a CIS. For any $X \subseteq U, P \subseteq A$, then the lower and upper approximations of X based on composite relation CR_P are defined by

$$\underline{CR_P}(X) = \{x \in U | CR_P(x) \subseteq U\}, \ \overline{CR_P}(X) = \{x \in U | CR_P(x) \cap X \neq \emptyset\}. \tag{4.12}$$

The Definition 4.17 can also be defined as follows.

Definition 4.20. Given information (U, A), if $A = A^c \cup A^n \cup A^s \cup A^{in}$ (A^c, A^n, A^s and A^{in} are categorical, numerical, set-valued and interval-valued attribute sets, respectively), then (U, A) is called a composite information system (CIS) or hybrid information system (HIS).

Table 5
A CIS.

U	a_1	a_2	a_3	a_4
x_1	y	18.5	{0,1}	[1,5]
x_2	y	24.8	{0}	[2,5]
x_3	n	17.4	{0,2}	[3,6]
x_4	n	18.5	{0,1}	[1,5]
x_5	y	12.2	{0,1,2}	[1,5]
x_6	y	18.5	{2}	[4,5]
x_7	n	22.3	{1}	[1,3]

Example 4.21. Table 5 is a CIS with a categorical attribute "a_1", a numerical attribute "a_2", a set-valued attribute "a_3" an interval-valued attribute "a_4" and a decision attribute d.

From Table 5, such data table is a reduced heterogeneous database. Originally, to deal with heterogeneous data, Hu et al. introduced an efficient hybrid attribute reduction algorithm based on a generalized fuzzy-rough model, however, they considered only two types of data characteristics, i.e., numerical and categorical features [105]. Then, some scholars extended rough set model to handle the data with mixed categorical and numerical features [106–110]. Actually, the composite information system is an extension of the general information system, which corresponds to the composite rough set model [66]. After several years of development, it has become the important research content in the many application areas. For example, to combine classical rough set and fuzzy rough set models, Chen et al. proposed a method of attribute reduction for a decision information system with symbolic and real-valued condition attributes [111]. Zhang et al. defined the composite rough set which contains four different types of data for dynamic data mining [67]. In addition, they first presented the boolean matrix representation of the lower and upper approximations in a CIS, and designed a parallel approach for computing approximations operators [112]. Since the fuzzy RST can be defined for different types of attributes to measure the similarity between objects, Zhang et al. presented a method of information entropy for feature selection in a mixed data set [113]. Zeng et al. developed incremental algorithms to deal with HISs [68,69]. Meanwhile, some scholars have studied the properties and structures of hybrid information systems by extending the CISs [71,72], and applied to three-way decisions [45,73]. Recently, Huang et al. extended the CIS model via integrating various types of attributes and fusing multi-composite relations derived from various data sources, which fully reflects the importance of integrated CISs [70].

Compared to CISs, the following model is more in line with the characteristics of multiple sources and heterogeneous.

4.2.2. Multi-source information systems (MsISs)

In this subsection, before introducing multi-source heterogeneous information systems, we first introduce an information system named a multi-source information system (MsIS). In order to reflect the situation where information arrives from multiple sources and focus on families of Pawlak approximation spaces, Khan and Banerjee first proposed a system called MsISs [22,74]. According to their points of view, we should restrict our attention to a fixed domain, i.e., the discourse (U). The following three conditions must be considered.

(1) With time, the partition on U changes because of the inflow of varied information.

(2) At the same time, information arrives from different sources and changes the partition on U.

(3) A combination of (1) and (2).

Namely, at each time point there is information from different sources (multiple sources), and along with it, there is change of information with time. A family of the single information systems can be obtained with the same universe of discourse. Such system is then called MsISs which is defined as follows.

Table 6
An MsDIS.

U	IS_1				IS_2				IS_3				IS_4				d
	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	
x_1	b	c	c	b	b	c	c	b	b	c	b	b	b	c	c	b	1
x_2	b	c	a	b	b	c	c	b	b	c	b	b	b	c	b	b	1
x_3	b	b	c	b	b	b	b	b	b	c	b	b	b	a	c	b	0
x_4	a	b	a	b	b	b	b	b	a	b	c	b	a	a	c	a	1
x_5	c	b	a	c	a	b	b	b	b	b	b	b	b	c	b	b	0
x_6	a	b	a	a	a	b	b	b	a	b	c	b	b	a	c	a	1

Table 7
An MsHIS.

U	d_1				d_2				$\cdots$				d_l			
	c_1	c_2	$\cdots$	c_m	c_1	c_2	$\cdots$	c_m	c_1	c_2	$\cdots$	c_m	c_1	c_2	$\cdots$	c_m
x_1	1	2	$\cdots$	5	0.1	0.4	$\cdots$	0.4	[1,2]	[2.3]	$\cdots$	[3,4]	N	Y	$\cdots$	N
x_2	2	3	$\cdots$	3	0.1	0.8	$\cdots$	0.7	[0,2]	[1,3]	$\cdots$	[1,4]	N	N	$\cdots$	N
x_3	3	4	$\cdots$	2	0.7	0.3	$\cdots$	0.3	[1,3]	[2,4]	$\cdots$	[2,4]	Y	Y	$\cdots$	Y
x_4	5	5	$\cdots$	5	0.6	0.2	$\cdots$	0.9	[1,3]	[3,5]	$\cdots$	[3,5]	Y	N	$\cdots$	Y
x_5	2	2	$\cdots$	3	0.9	0.5	$\cdots$	0.6	[0,2]	[1,5]	$\cdots$	[1,4]	N	Y	$\cdots$	N
x_6	4	3	$\cdots$	2	0.4	0.8	$\cdots$	0.2	[0,3]	[4,5]	$\cdots$	[3,5]	N	Y	$\cdots$	N

Definition 4.22 (*[75]*). Let $MsIS = \{IS_i | IS_i = (U, A_i, \{(V_a)_{a\in A_i}\}), f_i\}$, where (i) U is a set of objects; (ii) A_i is a set of attributes of each subsystem; (ii) V_a is the value of the attribute $a \in A_i$; and (iv) $f_i : U \times A_i \to \{(V_a)_{a\in A_i}\}$ for each $x \in U$ and $a \in A_i$, $f_i(x, a) \in V_a$.

In addition, Yang et al. gave more visual definition of MsISs.

Definition 4.23 (*[30]*). Let (U, IS_i) be an information system ($i = 1, 2, \dots, m$). Then MsIS=$\{(U, IS_1), (U, IS_2), \dots, (U, IS_m)\}$ is called an MsIS.

Example 4.24. Table 6 shows an MsIS with decisions which is called a multi-source decision information system (MsDIS). Here $MsDIS = \{(U, IS_1), (U, IS_2), (U, IS_3), (U, IS_4)\}$, where $U = \{x_1, x_2, x_3, x_4, x_5, x_6\}$.

In [30], Yang et al. proposed the information fusion mechanism that regarded an MsDIS as an information box, and then fused the information box into an information table by different strategies, such as fuzzy granules [28,32,114], information entropy [115] and so on. Eventually, the fused information table is applied for rule extraction. The fusion process is shown in Fig. 3.

From Fig. 3, there is an MsIS $(I_1, I_2, \dots, I_s)$ with same structure. For each I_i has n objects and m attributes. The s pieces of information systems overlapping together can construct an information box (i.e., an MsIS), which has s levels. Every block of shadow displays the attribute value after fusion.

As can be seen from Table 6, the so-called MsDIS is formed based on the aggregation of multiple single information systems. Therefore, the domain of attribute values are the same. However, if the types of attribute value is various, a multi-source heterogeneous information system (MsHIS) will be formed.

4.2.3. Multi-source heterogeneous information systems (MsHISs)

Definition 4.25 (*[76]*). Let $(U, C, D, \mathcal{F})$ be an MsHIS, where (i) $U = \{x_1, x_2, \dots, x_n\}$ is a set of objects; (ii) $C = \{c_1, c_2, \dots, c_m\}$ is a set of attributes of each subsystem; (iii) $D = \{d_1, d_2, \dots, d_l\}$ is the set of source of information collection; (iv) $\mathcal{F} = \{f_k(x_i, c_j) | i = 1, 2, \dots, n; j = 1, 2, \dots, m; k = 1, 2, \dots, l\}$ is a family of mapping sets, where $f_i : U \times C \to V_k$ (V_k is the value of attribute C with the available information $d_k (d_k \in D)$ for each $x_i \in U, c_j \in C$, $d_k \in D$ and $f_k(x_i, c_j) \in V_k$.

From Definition 4.25, it is obvious that $f_k(x_i, c_j)$ can be seen as the evaluation of the alternative x_i with respect to the attribute c_j in the light of the available information d_k. The following example vividly describes the MsHIS.

Definition 4.26 (*[76]*). Let $(U, C, D, \mathcal{F})$ be an MsHIS, $\forall\ B \subseteq C$. Given a weight vector $\omega = \{\omega_1, \omega_2, \dots, \omega_m\}^T$, then the fuzzy equivalence (or similarity) relation with available information $(d_k(d_k \in D), k = 1, 2, \dots, l)$ on attribute B is defined as

$$\begin{aligned} R_k^F(B)(x_i, x_j) = \{\alpha \mid (x_i, x_j) \in U \times U, \alpha = \min_{c_t \in B} \omega_t c_t(x_i, x_j), \\ c_t \in C, i, j = 1, \dots, n; 1 < t < m; k = 1, 2, \dots, l\}, \end{aligned} \tag{4.13}$$

where $c_t(x_i, x_j)$ ($t \in (1, m)$) is the similarity degree between objects x_i and x_j. Therefore, α can also be regarded as the similarity degree on B, denoted as $r_{ij} = \alpha = R_k^F(B)(x_i, x_j) (r_{ij} \in [0, 1])$.

In general, the similarity degree in [109] can be defined as

$$T_k(x_i, x_j) = \begin{cases} 1, & i = j; \\ \min_{c_t \in C}\{\omega_t(1 - |f_k^F(x_i, c_t) - f_k^F(x_j, c_t)|)\}, & i \neq j. \end{cases} \tag{4.14}$$

Example 4.27. Assume that $U = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ is a set of six applicants. And $c_1, c_2, \dots, c_m$ display different types of attribute, such as c_1(Education), c_2(Salary), c_3(Age), …, c_m(Credit record). $D = \{d_1, d_2, \dots, d_l\}$ shows that experts evaluate and rank according to the attributes that are provided by the applicants.

From Table 7, there are multiple different types of attributes for evaluating the credit card applicant which forms an MsHIS. To some extent, an MsIS is a special case of MsHIS. Some researchers have studied the fusion technologies of an MsHIS in different fields. For example, Wilson et al. presented three new heterogeneous distance functions to deal with nominal attributes and continuous attributes [102]. Liu et al. discussed the collection of multi-source heterogeneous data and spatial data, which is an important problem in the smart city data integration and big data analysis for the public information platform [116]. Zhang et al. introduced big data fusion and methods for heterogeneous data and focused on the application of deep learning methods in multi-source heterogeneous data fusion [100]. Liu proposed multi-source heterogeneous data fusion based on perceptual semantics in narrow-band Internet of Things [117]. Liu et al. summarized the feature representation methods of multi-source big data fusion on account of deep learning of three categories [118]. In addition to the above mentioned, researchers have addressed many methods of data fusion about multi-source heterogeneous data [119–125]. At present, the technology of multi-source heterogeneous information fusion is not mature in RST. Yager [11] put forward a general view for multi-source data fusion process, and emphasized the importance of developing data fusion algorithm. Xu and his team studied the methods of MSIF from the perspectives GrC [27], information entropy [28] and decision-theoretic rough set [31], respectively. In 2018, to fusion multiple attribute group decision-making issue, Sun et al. [76] proposed a multigranulation fuzzy rough set approach based on MsHISs which provided us with a good form of data table to study heterogeneous data. Zhang et al. considered evaluation issues in an MvIS, and adopted a TOPSIS-based evaluation fusion approach to evaluate agents [77].

Compared with homogeneous MSIF, the heterogeneous MSIF is very different in defining structure, which is mainly reflected in the data types, such as CISs [66] and MsHISs [76]. It is necessary to be emphasized that these two models are themselves multi-source heterogeneous, which can also be studied in accordance with previous several fusion techniques. The granulation fusion model is used to explore the uncertainty and information structure in CISs, e.g. [71] and [72]. In order to better solve the decision problem, Wang et al. combined the 3WD method with medical diagnosis by adopting the granulation structure, e.g. [45,126]. Moreover, the literature on the application of incremental learning fusion to study CISs is also increasing, e.g. [68] and [69]. More importantly, for large scale data sets, using parallel fusion model will greatly reduce the cost of operation, e.g. [127,128] and [129]. The MsHIS visually displays heterogeneous data representation from multiple sources, which is a useful fusion model, e.g. [30,77] and [130].

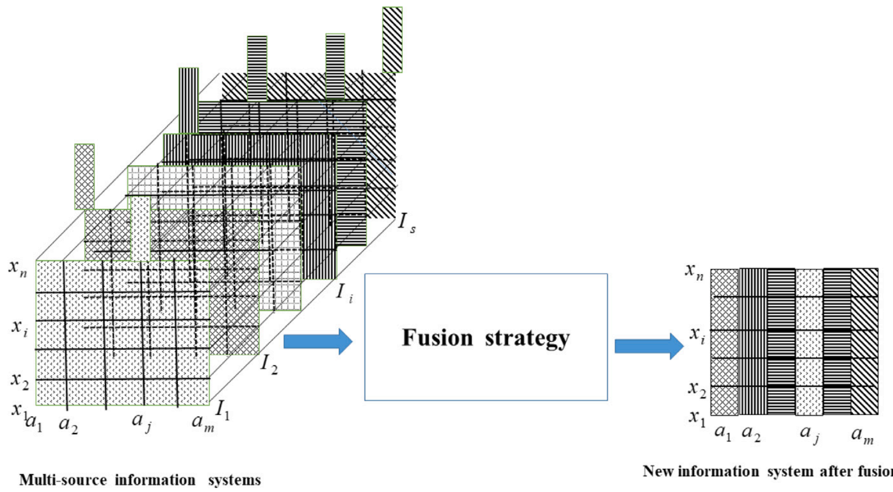

Fig. 3. The fusion process of an MsIS [30].

4.3. The summary of MSIF models

The relationship between information and data will be more complex in an MSIF model. It is necessary to mine potential, novel and useful knowledge from various of MSIF models. From these MSIF models, which are the databases with a relationship between objects and attributes (condition attributes and decision attributes). This kind of MSIF model implies the relationship between objects and attributes through data, and its biggest advantage is that it has a clear and intuitive meaning. However, these models have certain limitations and are not universal. This is due to the scale and diversity of the data table. It should be emphasized that this article provides existing solutions to solve these complex MSIF models.

Table 8
MvRSIF models and its applications.

MvRSIF models	Applications of the fusion strategy	References
MgIF.	Clustering ensemble.	[140].
MgIF.	Cost-sensitive.	[141] .
MgIF.	Attribute reduction.	[142,143].
MgIF.	Incremental fusion.	[144,145].
MgIF.	Medical diagnosis.	[13,64,146].
MgIF.	Rough approximation.	[33,147–151].
MsIF.	Granulation fusion.	[152,153].
MsIF.	Rough approximation.	[154].
MsIF.	Three-way decisions.	[155,156].
MvDIF.	Decision-making.	[157].
MvDIF.	Medical diagnosis.	[44,45,158].

5. Multi-view rough sets models based information fusion (MvRSIF)

The data of the same object obtained from different ways or different levels is called multi-view data, which presents the characteristics of polymorphism, multi-source, multi-descriptive and high-dimensional heterogeneity. Multi-view data exists widely in real life. For example, in the field of information technology, a web page can be described either by the text information in this web page or by the text information on the anchor link pointing to this web page. In the task of classifying the advertising pictures on the web page, one can judge whether it belongs to the advertising picture based on the content and color of the picture, or identify it based on the text information such as the logo title of the picture. These two methods can be seen as a feature is considered from two perspectives. Furthermore, in the classification of news reports, which from different sources constitute multiple perspectives of the news. Analogously, in the field of biometrics, person's identity can be recognized by fingerprints, sounds, faces etc. which constitute multiple views on the same object. In short, these features of different views describe the same object, but are distributed in different feature spaces. Simultaneously, these different features reveal different attributes of things from different views, so as to comprehensively and accurately describe the same object.

Multi-view data, which can be explained as combining multiple sources or views of datasets [131]. Namely, it can be captured from heterogeneous sources. However, in RST, the multi-view data is often expressed as the views of different levels of granularity or scales. When dealing with a large amount of complex information, due to the limited cognitive ability, people tend to divide them into many relatively simple blocks according to their respective characteristics and performance. Then, each block can be seen as a grain. In fact, grain refers to the block formed by some individuals through indistinguishability relation, similarity relation, neighborhood relation, and so on. This process of handling information is called information granulation. This idea was first called the term "granular computing (GrC)" by Lin [132]. Granulation is an operation in GrC and is the process of partitioning the data into groups representing granules. Granule is the basic components of GrC. Nguyen et al. presented several methods for synthesis of complex information granules in distributed environment, which are significant due to the need for approximate fusion of information from different sources of information [133]. To learn more about the principles and methods of GrC, one can refer to these papers [134–139].

In this section, the MvRSIF is discussed via introducing three kinds of fusion models to discuss, i.e., multi-granulation information fusion model (MgIF), multi-scale information fusion model (MsIF) and multi-view decision information fusion model (MvDIF). Furthermore, Table 8 lists these three models of MvRSIF and the existing fusion applications are summarized.

5.1. Multi-granulation information fusion model (MgIF)

In this subsection, we focus on the research processing of MSIF according to the idea of granulation. In recent years, how to use the advantage of GrC to deal with complex information to study the MSIF has been widely concerned. For example, decision making [130,159], medical diagnosis [13], person–job fit [160], fusion strategy [161,162], steam turbine fault diagnosis [163], etc. The diversity of information sources may lead to inconsistency in the process of decision-making, Qian et al. proposed a multi-granulation rough set model that extend Pawlak's single-granulation rough set model [147]. They applied two useful fusion rules, i.e., disjunctive and conjunctive combination rules. We first give the definition of a multigranulation rough set as follows.

Definition 5.1 (*[147]*). Let (U, A) be an information system, and R be a family of equivalence relations. Given $X \sqsubseteq U$ $(X \neq \emptyset)$, $P_i \in R$ $(i = 1, 2, \ldots, m)$ be the m partitions, then a pair of optimistic lower and upper approximation of X w.r.t. P_i based on disjunction can be defined by

$$(\underline{\sum_{i=1}^{m} P_i(x))^O} = \{x \mid \vee P_i(x) \subseteq X\}, \tag{5.1}$$

$$(\overline{\sum_{i=1}^{m} P_i(x))^O} = \sim (\sum_{i=1}^{m} P_i(\sim X))^O. \tag{5.2}$$

Furthermore, a multi-granulation decision fusion model based on pessimistic multi-granulation rough set model is presented as follows.

Definition 5.2 (*[162]*). Let (U, A) be an information system, and R be a family of equivalence relations. Given $X \sqsubseteq U$ $(X \neq \emptyset)$, $P_i \in R$ $(i = 1, 2, \ldots, m)$ be the m partitions, then a pair of pessimistic lower and upper approximation of X w.r.t. P_i based on conjunctive can be defined by

$$(\underline{\sum_{i=1}^{m} P_i(x))^P} = \{x \mid \wedge P_i(x) \subseteq X\}, \tag{5.3}$$

$$(\overline{\sum_{i=1}^{m} P_i(x))^P} = \sim (\sum_{i=1}^{m} P_i(\sim X))^O. \tag{5.4}$$

Meanwhile, they combined a multi-granulation method and a decision-theoretic rough set model to propose a multigranulation decision-theoretic rough set model, and explored the relationship between different multi-granulation models [148]. In addition, many researchers also have studied the multi-granulation rough set model from the angle of MSIF. For example, Xu et al. proposed an optimistic and pessimistic multi-granulation rough set model for multi-granulation fusion of partial order data, and built uncertainty measures such as the roughness and approximate classification quality of multi-granulation rough sets in an ordered information system [164]. And they presented generalized multi-granulation rough sets for selecting optimal granularity [149]. Lin et al. combined the granular structures with both reliability and conflict from multiple sources, and first addressed the connection between multi-granulation rough set theory and the evidence theory [75]. Li et al. put forward a multi-granulation information fusion method that in combination with evidence theory based on clustering ensemble algorithm [140]. Ju et al. addressed a cost-sensitive multi-granulation rough set model in view of two different costs simultaneously [141]. To obtain required knowledge from neighborhood information systems effectively, Hu et al. studied matrix based incremental approaches to update knowledge in neighborhood multi-granulation rough set [145]. Xu considered that attributes and attributes values may have different granulation, then proposed multi-granulation rough set based on attributes and attributes values [165]. Sun et al. presented a fuzzy neighborhood multigranulation rough sets model to deal with heterogeneous data sets containing numerical and symbolic feature values, and used fuzzy neighborhood entropy-based uncertainty measures for feature selection [143]. Moreover, in order to provide a unified framework for representing multi-granulation knowledge, multi-granulation space is researched by some scholars [166–168]. In literature [33], more than 10 research methods of multi-granulation rough sets are summarized, which shows the effectiveness and applicability of this multi-granulation model for MSIF.

5.2. Multi-scale information fusion model (MsIF)

In the aforementioned subsection, the main idea of multi-granulation rough set model, which is to perform intersection and union operation through attribute reduction. In practical application, people may need to observe, represent, analyze and make decisions on the different scales of data in an information system under the same attribute or variable. Namely, for an attribute corresponding to the same object, the values of different levels of labels can be taken according to the needs of different levels of granularity. Wu and Leung first proposed a model to GrC with multi-scale data which can be measured by different levels of granulations [152]. In addition, the problem of selecting the optimal granulation under this model is further discussed [153]. In [152], they defined a new rough set model named block-labeled rough set model (or called Wu–Leung model), which used the labeled partitions to define the rough approximations. This model shows that different scales of attribute values may lead to multiple granulations, so people need to observe and analyze the same object from the view of different scales of granulation. At present, this model is applied to many fields, for example, optimal scale selection [153,169–171], rule acquisition [172–174], decision analysis [155,156,175], formal concept analysis [176], etc. The notion of multi-scale information system is defined as follows.

Definition 5.3 (*[152]*). Let (U, A) be an information system, where U is the universe, $A = (a_1, a_2, \ldots, a_m)$ is the set of attributes. If each attribute $a_j \in A$ has I scales, then the (U, A) is called a multi-scale information system (MsIS), which can be defined as $(U, a_j^k | k = 1, 2, \ldots, I_j; j = 1, 2, \ldots, m)$.

If (U, A) is an information system and $A = C \cup \{d\}$, then $(U, a_j^k | k = 1, 2, \ldots, I; j = 1, 2, \ldots, m) \cup \{d\}$ is called a multi-scale decision information system (MsDIS).

From Definition 5.3, for the attribute $a_j \in A$ of the same object in U, which can take different values at different scales, and it is restricted on its l_jth $(j = 1, 2, \ldots, m)$ scale. In addition, the index set $(l_1, l_2, \ldots, l_m)$ is regarded as a scale combination of condition attributes, and $L = \{(l_1, l_2, \ldots, l_m) | l_j \in \{1, 2, \ldots, I\}, j = 1, \ldots, m\}$ is represented as the family of all scales combinations. Each scale combination $K = (l_1, l_2, \ldots, l_m)$ can form a single-scale decision system $S^K = (U, C^K \cup \{d\})$, where $C^K = \{a_1^{l_1}, a_2^{l_2}, \ldots, a_m^{l_m}\}$.

To introduce the multi-scale decision information into the decision table, Huang et al. first presented a generalized MsDIS as follows.

Definition 5.4 (*[177]*). Suppose that $(U, C \cup \{d\})$ is a generalized MsDIS, d is decision attributes with n scales $\{d^t | t = 1, 2, \ldots, n\}$. Then, the generalized MsDIS is described by a data table

$$\begin{aligned}&(U, C \cup \{d\})\\ &= (U, \{a_j^k | k = 1, 2, \ldots, l_j; j = 1, 2, \ldots, m\} \cup \{d^t | t = 1, 2, \ldots, n\}),\end{aligned} \tag{5.5}$$

where $(U, C) = (U, a_j^k | k = 1, 2, \ldots, I_j; j = 1, 2, \ldots, m)$ is an MsIS.

From Definition 5.4, let $K = (l_1, l_2, \ldots, l_m) \in L$. It is obvious that the generalized MsDIS can be decomposed into n single decision information systems, i.e.,

$$S^{Kt} = (U, \{a_1^{l_1} \cup a_2^{l_2}, \ldots, a_m^{l_m}\} \cup \{d^t\}, l_j = 1, 2, \ldots, l_j; t = 1, 2, \ldots, n). \tag{5.6}$$

In addition, if d is restricted on its tth scale, it will be degraded into Definition 5.3.

Example 5.5. Table 9 shows a generalized MsDIS $= (U, \{a_j^k | k = 1, 2; j = 1, 2, 3\} \cup \{d^t | t = 1, 2\})$, where $U = \{x_1, x_2, \ldots, x_6\}$ expresses six students, the attribute set A_1, A_2 and A_3 express the scores of three courses. Each attribute set is described as two different views, i.e., two scales. The scales a_j^1 and $a_j^2 (j = 1, 2, 3)$ are denoted as a ten-score and five-score systems, respectively. For decision attribute d also has two scales form two different views, which represent scholarship availability. The decision attribute d^1 expresses the scholarship grade, and “Nl”, “Ul”, “Cl”, and “Ns” express “national level”, “university level”, “college level”, and “no scholarship”, respectively. The decision attribute d^2 expresses whether the graduation clause has been met. Hence, “1” and “0” represent “Yes” and “N”, respectively.

Table 9
A generalized MsDIS.

U	A_1		A_2		A_3		D	
	a_1^1	a_1^2	a_2^1	a_2^2	a_3^1	a_3^2	d^1	d^2
x_1	9	E	6	M	9	E	Nl	1
x_2	7	G	7	G	9	E	Ul	1
x_3	5	M	3	B	6	M	Ns	0
x_4	7	G	7	G	4	B	Ns	0
x_5	8	G	9	E	7	G	Cl	1
x_6	4	B	8	G	6	G	Ns	0

So far, the research on MsIS is gradually increasing and expanding, which provides a solid theoretical model for the study of multi-view or multi-scale data. Gu et al. introduced a new multi-scale fusion model, i.e., multi-scale interval information systems, and defined lower and upper approximations in this system [154]. Wu et al. discussed the knowledge discovery with RST method in incomplete MsDIS [172]. For the optimal scale selection problem, Hao et al. investigated a sequential 3WD model a dynamic MsDIS [155]. In addition, Luo et al. also studied the update problem of 3WD from the view of dynamic variation of scales in an incomplete MsIS [156]. Chen et al. defined the scale significance to two kinds of scales from the perspective of the relation matrix in an MsIS [178].

From Sections 5.1 and 5.2, we can find that the multi-view data in RST is studied from the granule and scales of different levels, which reflects the view point of GrC. Namely, a granule can be interpreted as lots of small particles forming a larger unit. It is worth emphasizing that there are diverse granules at different levels of scales in datasets having hierarchical structures. Therefore, Fig. 4 is given to show the basic idea of fusion strategy for multi-view data in RST.

It is necessary to be emphasized that the construction of information structures which is the key step in the fusion strategy. There are many papers study the problem about information structure in an information system, such as [71,72,179–182] and [183].

In the next subsection, we introduce another approach to information fusion from multiple views for decision-making.

5.3. Multi-view decision information fusion model (MvDIF)

Due to the cost of misclassification on account of different decision such as optimistic and pessimistic decisions, Zhou and Li proposed a multi-view decision model based on Decision-Theoretic Rough Set models (DTRS) [157]. First, we review the concept of the DTRS model.

• **DTRS model** [184]

Assume that the set $\triangle = \{\omega_1, \omega_2, \dots, \omega_m\}$ is m states, and $\mathbb{A} = \{a_1, a_2, \dots, a_s\}$ is s possible actions. The $P(\omega_j|\mathbf{x})$ expresses the conditional probability of an object x, and $\lambda(a_i|\omega_j)$ expresses the cost, or loss, which can be calculated by taking action a_i as follows.

$$\mathcal{R} = \sum_{j=1}^{m} \lambda(a_i|\omega_j)P(\omega_j|\mathbf{x}) \tag{5.7}$$

Let U be the universe and $X \subseteq U$. Given a set of state is denote by $\Omega = \{X, -X\}$, which indicates an element is in X and not in X. The $\mathcal{A} = \{a_P, a_B, a_N\}$ represents three actions taken against an object, i.e., deciding $POS(X)$, $BND(X)$ and $NEG(X)$. Moreover, λ_{PP}, λ_{BP} and λ_{NP} express the losses of taking actions a_P, a_B and a_N when an object belongs to X. $\lambda_{PN}, \lambda_{BN}$ and λ_{NN} express the losses of taking the same three actions when an object does not belongs to X [184]. Then the expected loss $R(a_i|[x]R)$ can be described as:

$$\begin{aligned}
R(a_P|[x]_R) &= \lambda_{PP}P(X|[x]_R) + \lambda_{PN}P(-X|[x]_R);\\
R(a_B|[x]_R) &= \lambda_{BP}P(X|[x]_R) + \lambda_{BN}P(-X|[x]_R);\\
R(a_N|[x]_R) &= \lambda_{NP}P(X|[x]_R) + \lambda_{NN}P(-X|[x]_R).
\end{aligned} \tag{5.8}$$

Then obtain the minimum-risk decision rules based on Bayesian decision can be expressed as follows:

$$\begin{aligned}
&If\ R(a_P|[x]_R) \le R(a_N|[x]_R)\ and\ R(a_P|[x]_R) \le R(a_B|[x]_R),\\
&\quad then\ deciding\ POS(X);\\
&If\ R(a_B|[x]_R) \le R(a_P|[x]_R)\ and\ R(a_B|[x]_R) \le R(a_N|[x]_R),\\
&\quad then\ deciding\ BND(X);\\
&If\ R(a_N|[x]_R) \le R(a_P|[x]_R)\ and\ R(a_N|[x]_R) \le R(a_B|[x]_R),\\
&\quad then\ deciding\ NEG(X).
\end{aligned} \tag{5.9}$$

In general, the loss functions satisfied $\lambda_{PP} \le \lambda_{BP} \le \lambda_{NP}$ and $\lambda_{NN} \le \lambda_{BN} \le \lambda_{PN}$ and $P(X|[x]_R)+P(-X|[x]_R) = 1$, then the following decision rules are obtained.

$$\begin{aligned}
&If\ P(X|[x]_R) \ge \gamma\ and\ P(X|[x]_R) \ge \alpha,\ then\ deciding\ POS(X);\\
&If\ P(X|[x]_R) \le \beta\ and\ P(X|[x]_R) \ge \gamma,\ then\ deciding\ NEG(X);\\
&If\ P(X|[x]_R) \ge \beta\ and\ P(X|[x]_R) \le \alpha,\ then\ deciding\ BND(X),
\end{aligned} \tag{5.10}$$

where

$$\begin{aligned}
\alpha &= \frac{\lambda_{PN} - \lambda_{BN}}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})},\\
\beta &= \frac{\lambda_{BN} - \lambda_{NN}}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})},\\
\gamma &= \frac{\lambda_{PN} - \lambda_{NN}}{(\lambda_{PN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{PP})}.
\end{aligned} \tag{5.11}$$

When $(\lambda_{PN} - \lambda_{BN})(\lambda_{NP} - \lambda_{BP}) > (\lambda_{BP} - \lambda_{PP})(\lambda_{BN} - \lambda_{NN})$, let $\alpha > \beta$, thus $\alpha > \gamma > \beta$, so it can induce the decision rules as follows.

$$\begin{aligned}
&If\ P(X|[x]_R) \ge \alpha,\ then\ deciding\ POS(X);\\
&If\ P(X|[x]_R) \le \beta,\ then\ deciding\ NEG(X);\\
&If\ \beta < P(X|[x]_R) < \alpha,\ then\ deciding\ BND(X).
\end{aligned} \tag{5.12}$$

• **A multi-view decision model based on DTRS** [157]

When people make decisions in a practical application, they may find different characteristics between the diverse types of decisions. Different decisions will show different attitudes among groups of people. For the same thing, some people may take positive and optimistic decisions, while others may take negative and pessimistic decisions. For example, in the prediction of stock fluctuations, an optimistic person will think that the stock price will not fall, may spend money to buy, while a pessimist who fears a loss will not spend money to buy or even sell own shares. It follows that people always tend to make diverse types of decisions based on their cognitive abilities. Therefore, to consider the difference among different people, Zhou and Li developed a flexible decision model named multi-view decision model based on DTRS, in which different types of decisions can solve practical problem [157].

According to the values of loss $\lambda_{PP}, \lambda_{BP}, \lambda_{NP}, \lambda_{PN}, \lambda_{BN}$, and λ_{NN}, they provide three types of decision, namely, optimistic, pessimistic and equable decisions. However, these decisions are on account of the following three assumptions.

$$(i)\ \alpha > \beta; (ii)\ \lambda_{PP} = \lambda_{NN} = 0; (iii)\ \lambda_{BP} = \sigma\lambda_{NP}, \lambda_{BN} = \sigma\lambda_{PN}\quad (0 < \sigma < 1). \tag{5.13}$$

According to formula (5.12) and assumptions (5.13), we have

$$\alpha = \frac{(1-\sigma)\lambda_{PN}}{(1-\sigma)\lambda_{PN} + \sigma\lambda_{NP}}; \qquad \beta = \frac{\lambda_{PN}}{\sigma\lambda_{NP} + (1-\sigma)\lambda_{NP}}. \tag{5.14}$$

Then three decisions are described as follows:

(1) Optimistic decision (select a low λ_{NP} and high λ_{PN}):

$$\begin{aligned}
&If\ P(X|[x]_R) \ge \alpha_{Opt},\ then\ deciding\ POS(X);\\
&If\ P(X|[x]_R) \le \beta_{Opt},\ then\ deciding\ NEG(X);\\
&If\ \alpha_{Opt} < P(X|[x]_R) < \alpha_{Opt},\ then\ deciding\ BND(X).
\end{aligned} \tag{5.15}$$

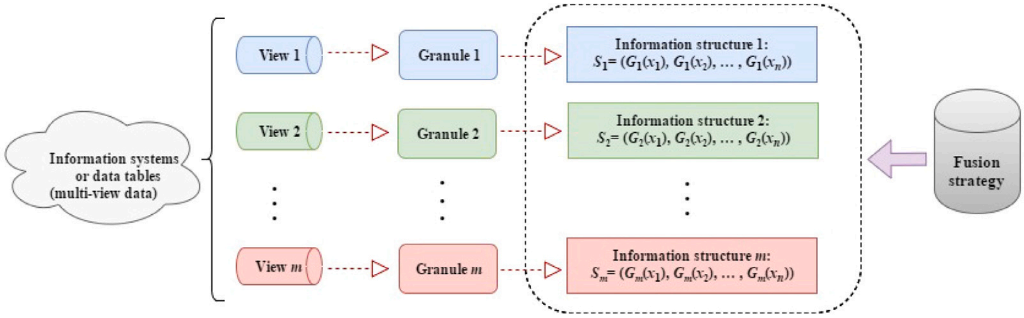

Fig. 4. The fusion strategy of multi-view data in RST.

(2) Pessimistic decision (select a high λ_{NP} and low λ_{PN}):

$$\begin{aligned} &If\ P(X|[x]_R) \geq \alpha_{Pes},\ then\ deciding\ POS(X);\\ &If\ P(X|[x]_R) \leq \beta_{Pes},\ then\ deciding\ NEG(X);\\ &If\ \alpha_{Pes} < P(X|[x]_R) < \alpha_{Pes},\ then\ deciding\ BND(X). \end{aligned} \tag{5.16}$$

(3) Equable decision (select a medium λ_{NP} and medium λ_{PN}):

$$\begin{aligned} &If\ P(X|[x]_R) \geq \alpha_{Equ},\ then\ deciding\ POS(X);\\ &If\ P(X|[x]_R) \leq \beta_{Equ},\ then\ deciding\ NEG(X);\\ &If\ \alpha_{Equ} < P(X|[x]_R) < \alpha_{Equ},\ then\ deciding\ BND(X). \end{aligned} \tag{5.17}$$

• **A multi-view decision model based on ranking method** [158]

We introduce another multi-view decision-making method that can fuse multiple decisions through different risk preferences. The decision makers can calculate the conditional probability measure and three thresholds α, β and γ to determine the decision. In general, people evaluate a thing in a certain range, and so does the loss function. Hence, the following contents define the loss function in a certain range to be considered.

At first, the loss function regarding the cost of actions is given by a 3×2 matrix as follows [158]:

	X	$-X$
a_P	$\lambda_{PP} = [\lambda^-_{PP}, \lambda^+_{PP}]$	$\lambda_{PN} = [\lambda^-_{PN}, \lambda^+_{PN}]$
a_B	$\lambda_{BP} = [\lambda^-_{BP}, \lambda^+_{BP}]$	$\lambda_{BN} = [\lambda^-_{BN}, \lambda^+_{BN}]$
a_N	$\lambda_{NP} = [\lambda^-_{NP}, \lambda^+_{NP}]$	$\lambda_{NN} = [\lambda^-_{NN}, \lambda^+_{NN}]$

Then, $\lambda^-_{\bullet\bullet}$ and $\lambda^+_{\bullet\bullet}$ are the lower and the upper bound of $\lambda_{\bullet\bullet}$ ($\bullet = P, B, N$), respectively.

(i) For each $x \in X$, $\lambda_{PP} = [\lambda^-_{PP}, \lambda^+_{PP}]$, $\lambda_{BP} = [\lambda^-_{BP}, \lambda^+_{BP}]$, $\lambda_{NP} = [\lambda^-_{NP}, \lambda^+_{NP}]$ display the losses incurred for taking actions of a_P, a_B and a_N, respectively.

(ii) For each $x \in -X$, $\lambda_{PN} = [\lambda^-_{PN}, \lambda^+_{PN}]$, $\lambda_{BN} = [\lambda^-_{BN}, \lambda^+_{BN}]$, $\lambda_{NN} = [\lambda^-_{NN}, \lambda^+_{NN}]$ display the losses incurred for taking actions of a_P, a_B and a_N, respectively.

Finally, based on the basic conditions of DTRS model [184], the loss functions hold the following conditions:

$$\lambda^-_{PP} \leq \lambda^-_{BP} < \lambda^-_{NP};\ \lambda^+_{PP} \leq \lambda^+_{BP} < \lambda^+_{NP}. \tag{5.18}$$

Analogously,

$$\lambda^-_{NN} \leq \lambda^-_{BN} < \lambda^-_{PN};\ \lambda^+_{NN} \leq \lambda^+_{BN} < \lambda^+_{PN}. \tag{5.19}$$

With above limitations, the loss function can be controlled in a certain interval-valued number, which is fused by the following definition.

Definition 5.6 (*[44]*). Let $\lambda \in [\Re]$($\Re$ *is real number*) and $\mu \in [0,1]$. The formula of fusion loss function is defined as $\phi_\mu(\lambda) = (1-\mu)\lambda^- + \mu\lambda^+$, where $\lambda = [\lambda^-, \lambda^+]$. Then, $\phi_\mu(\lambda)$ is called the transformed outcome of λ with respect to μ, and the parameter μ reflects the risk attitude of decision makers.

Risk decision makers can set different μ according to their different risk attitudes. In general, there are three basic risk attitudes that reflect the decision makers as shown below.

(1) Conservative decision: Select $\mu = 0$, $\phi_0(\lambda) = \lambda^-$;
(2) Neutral decision: Select $\mu = 0.5$, $\phi_{0.5}(\lambda) = \frac{\lambda^+ + \lambda^-}{2}$;
(3) Risky decision: Select $\mu = 1$, $\phi_1(\lambda) = \lambda^+$.

By Definition 5.6, and through steps similar to Formulas (5.8)–(5.10), the three thresholds α, β and γ are also can be calculated as follows:

$$\begin{aligned} \alpha &= \frac{\phi_\mu(\lambda_{PN}) - \phi_\mu(\lambda_{BN})}{(\phi_\mu(\lambda_{PN}) - \phi_\mu(\lambda_{BN})) + (\phi_\mu(\lambda_{BP}) - \phi_\mu(\lambda_{PP}))},\\ \beta &= \frac{\phi_\mu(\lambda_{BN}) - \phi_\mu(\lambda_{NN})}{(\phi_\mu(\lambda_{BN}) - \phi_\mu(\lambda_{NN})) + (\phi_\mu(\lambda_{NP}) - \phi_\mu(\lambda_{BP}))},\\ \gamma &= \frac{\phi_\mu(\lambda_{PN}) - \phi_\mu(\lambda_{NN})}{(\phi_\mu(\lambda_{PN}) - \phi_\mu(\lambda_{NN})) + (\phi_\mu(\lambda_{NP}) - \phi_\mu(\lambda_{PP}))}. \end{aligned} \tag{5.20}$$

• **Three-way decisions** [185]

Based on DTRS model, another very popular model has been studied in the last 10 years, called three-way decisions (3WD) [185]. The 3WD model is a trisecting–acting-outcome (TAO) mode proposed by Yao [139,186], which focus on structured thinking, problem solving, and information processing from the view point of GrC [187]. Namely, given a finite objects set U and conditional set C. The main work of the 3WD is to divide the set of objects (U) into three mutually disjoint parts based on the conditional set (C), which are denoted as L region, M region and R region, respectively. These three regions satisfy the following conditions:

(1) $L \cup M \cup R = U$.
(2) $L \cup M = \emptyset, L \cap R = \emptyset, R \cap M = \emptyset$.

After then, using the positive region, boundary region and negative region in RST, the rules of 3WD is described as follows. The positive rule obtained from the positive region is used to accept something, the negative rule obtained from the negative region is used to reject something, and the rule that fall on the boundary region represents delayed decisions. For example, if a total order $([0,1], \leq)$ of U, where L is the unit interval [0,1], the total order $\leq$ is greater than less than. Then the evaluation function v can be viewed as a mapping from universe U to total order set L, i.e., $U \to L$. Given a pair of threshold (α, β), according to formula (3.5), then the universe U is divided into the following three parts.

(1) Positive region:

$$POS_{(\alpha,\beta)}(v) = \{x \in U | v(x) \geq \alpha\} \tag{5.21}$$

(2) Negative region:

$$NEG_{(\alpha,\beta)}(v) = \{x \in U | v(x) \leq \beta\} \tag{5.22}$$

(3) Boundary region:

$$BND_{(\alpha,\beta)}(v) = \{x \in U | \beta < v(x) < \alpha\} \tag{5.23}$$

Therefore, the 3WD model, which divides the discourse domain into three parts, describes the thinking pattern of human beings when solving practical decision-making problems. It provides a reliable theoretical basis for the application of rough set method to data-driven decision classification. In addition, it is a new concept and calculation method of information processing, which is mainly used to describe and process the uncertain, vague, incomplete and massive information, as well as provides a problem solving method based on the relationship between grains and grains. Until now, the 3WD model has been integrated into many practical fields, for example, medical diagnosis [44,45,73,188,189], credit card evaluation [190], SQL system [191], face recognition [192], project resource allocation [193], malware analysis [194], etc. Therefore, the cross application of 3WD model and MSIF will be a promising direction in future research.

5.4. The summary of MvRSIF

The so-called multi-view rough set model not only refers to investigating the traditional multi-view data, but also emphasizes the multiple views to study structure and model of the data. For example, multi-granulation rough set model is to extend the model of rough sets in terms of the structure of different granularity, which is from the view of granulation idea, e.g. [33]. A multi-scale information system is described as multi-view data which means the values of different levels of labels can be taken according to the needs of different levels of granularity, e.g. [152]. And as for multi-view decision model, it is researched according to the view of different decisions, e.g. [157]. The common feature of these models is the use of GrC model. It should also be emphasized that the 3WD model [139,186], as a research hotspot in recent years, can also be used as one of the multi-view decision models. In this paper, the multi-view rough sets model does not refer to single model, but are summarized from views of granularity and decisions in terms of different GrC models. Therefore, the current disadvantage is that there is no uniform model to describe it.

6. Parallel computing model based information fusion (PCIF)

In the age of Big data, the method of parallel computing can save a lot of running time and improve its efficiency. Parallel computing refers to the process of using multiple computing resources to solve computing problems at the same time, and an effective approach to improve the computing speed and processing power of computer systems. The fundamental idea is to use multiple processors to solve the same problem collaboratively. Namely, the problem to be solved is decomposed into many parts, and each part is calculated in parallel by an independent processor. Moreover, the process elements are diverse, such as a single computer with multiple processors, several networked computers, special hardware and so on [195,196].

6.1. The MapReduce fusion model

It is well known that most of the current algorithms on account of rough set are the sequential algorithms and the corresponding methods only run on a single computer to handle small datasets. As early as in 1998, Susmaga first proposed a method of parallel computation for reducts in RST [197]. In 2012, Zhang et al. found the parallel computing and GrC are perfectly compatible to some extent with the idea of granulation, because it converts an information system (i.e. original data table) into a granularity representation that reduces the amount of space. Therefore, it can be quickly cached in the distributed file system (DFS). Based on this point of view, they introduced the framework of MapReduce for computing approximations of rough set [198]. It is obvious that the MapReduce is an effective and useful model to accelerate data fusion. Therefore, according to the idea of literature [199], Fig. 5 indicates the data flow and different phases of the MapReduce framework.

In [200–202], scholars from Google introduced the principles and concepts of MapReduce in detail. The MapReduce model is a distributed parallel computing architecture, which is used for parallel computing of large-scale data. This model reduces the difficulty of parallel programming, and has become the mainstream parallel programming model of cloud computing platform. It can easily write applications and run on large clusters, and it has the characters of simplicity, high reliability and fault-tolerance. Based on the idea of “divide and conquer”, MapReduce highly abstracts complex parallel computing processes into two functions, i.e., Map and Reduce. In short, it can be seen as “the decomposition of tasks and the aggregation of results”. Generally speaking, users only need to implement Map and Reduce functions, without paying attention to how to divide the data and how to allocate the task.

In RST, the equivalence relation divides the universe U to obtain the equivalence class, and it can approximately describe any subset of U. According to the angle of granulation, the equivalence relation determines the basic knowledge of an information system. The finer the granulation is, the more accurate the approximate description. Conversely, the coarser the granulation, the fuzzier the approximation. In practice, the type of data is often extremely large and multi-source. And few researches are currently researching on parallel approaches in multi-source data environments. Therefore, we give an accelerated model for computing equivalence classes of multi-source data based on MapReduce framework, shown in Fig. 6.

From Fig. 6, an MsIS can be split by multiple blocks, and it can also divide different granulation based on the corresponding relation. In this MapReduce fusion model, there are also two fusion functions, i.e., Map and Reduce.

- Map step: each Map worker reads data split (U, A) from MsIS, and then maps its elements into key–value. The main function of this processing is dividing (U, A) into equivalence classes (or tolerance classes).
- Reduce step: each Reduce worker receives a group of data, and the key–value is $E_i (i = 1, 2, 3, 4)$. Then, the value of function $S_i (i = 1, 2, 3, 4)$ can be computed.
- Sum: the main function collects the each S_i from all Reduce workers, and finally outputs all results.

Note that the function can be set according to the actual situation. Moreover, the MSIF based on parallel model can be studied from the following several suggestions in future research.

(i) Adopt parallel method to optimize the GPU implementation and design the corresponding MSIF system based feature selection on GPUs.

(ii) Extend the rough sets model, such as composite-based rough set, neighborhood-based rough set and dominance-based rough sets and so on, and apply them to compute the rough approximations in MsIS by parallel fusion model.

(iii) Use parallel-based rough set methods to attribute reduction, and construct different MapReduce models on large-scale multi-source datasets.

6.2. The MP–DP fusion model for attribute reduction

Zhang et al. studied approximation space of rough sets for data mining and knowledge discovery based on the parallel computing methods [203–205]. Moreover, they proposed model-and-data parallelism (MDP) framework for attribute reduction [206], which is the combination of Model-Parallelism (MP) and Data-Parallelism (DP) when designing method for attribute selection. In MDP framework, two stages were considered in each loop, namely, generating candidates using heuristic search and calculating importance of each attribute in candidates. (1) In MP, a multi-processing method is employed to parallelize the processing of evaluation of multiple attributes generating candidates. (2) In DP, the processing of the computation of an attributes importance is parallelized by using mechanism of Spark. The framework of MP–DP model is depicted in Fig. 7. The granules are coarsened or refined during the process of MP and DP.

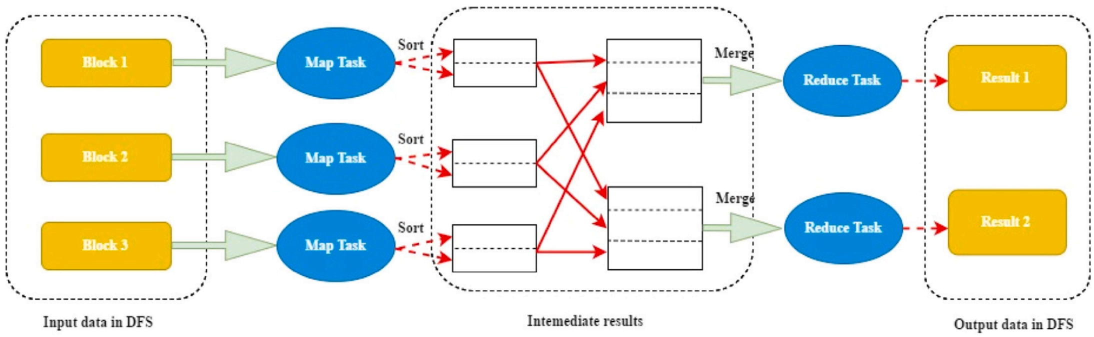

Fig. 5. The fusion model of MapReduce programming.

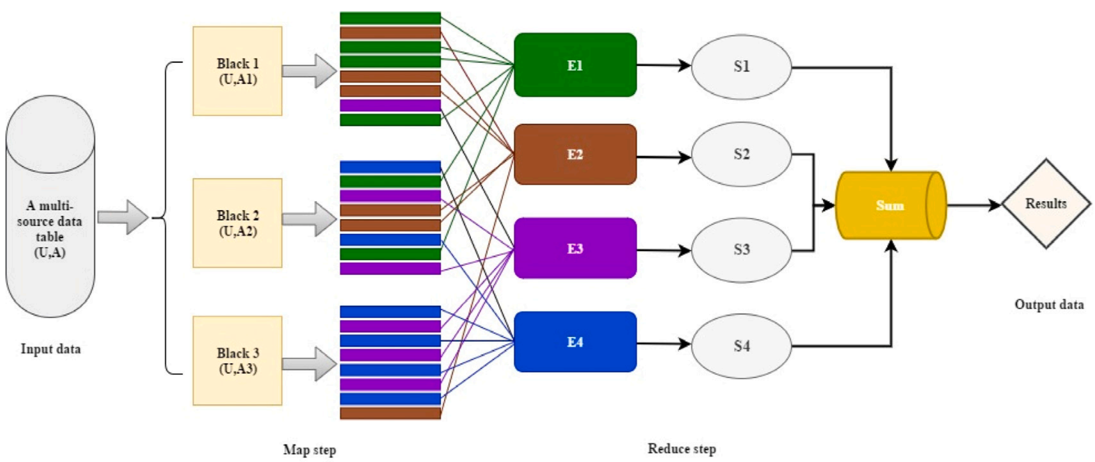

Fig. 6. The accelerated fusion model based MapReduce for computing equivalence classes of multi-source data.

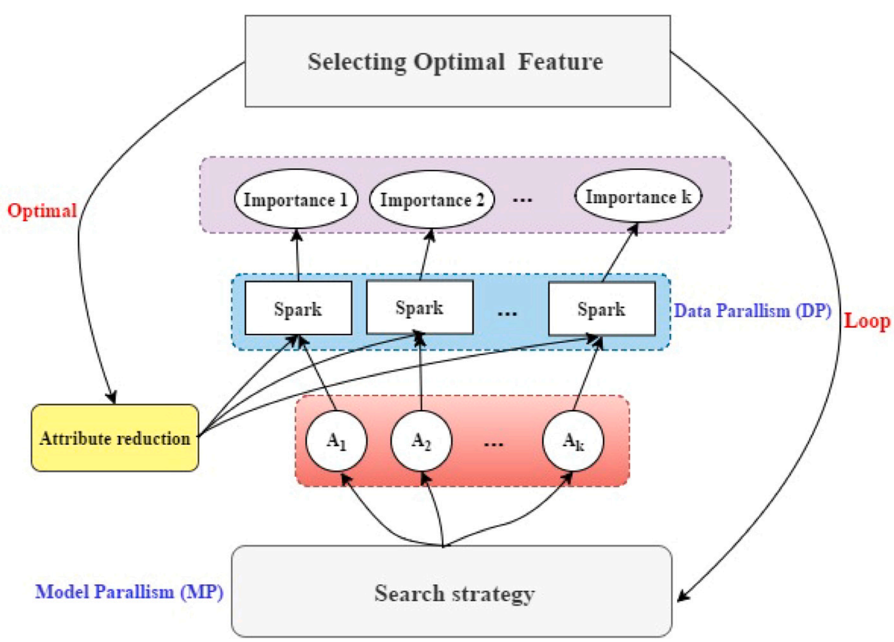

Fig. 7. The fusion mechanism of MP–DP model for attribute reduction [206].

The method of granularity representation is employed to accelerate the processing procedure of feature selection in MP–DP model. The granularity representation of a decision table (i.e., an information system) $S = (U, A)$ w.r.t. A is denoted as $G^A = \{(\vec{EA}, |EA|) | EA \in U/A\}$ such that $EA = (v_{a_1}, \ldots, v_{a|A|})$. The data stored in S is transformed to granules $G_i (i = 1, 2, 3, 4)$. The mechanism of coarsening and refining of granules is then easily expressed by these granules. These granules are used as basic elements in MP–DP model. The first stage of MP–DP model is to construct the granularity representation of the decision table, i.e., GrC-based initialization. The results of granulation is expressed in spark framework. Next, there are three main steps of obtaining attribute reduct, i.e., computing attribute core, computing the attribute reduct, and computing evaluation functions value. In each step, it is employed as elementary element which accelerates the algorithm. The granules are mapped and reduced in the course of computing evaluation functions value. And a reliable speed mechanism is given by granulation of the data. The detail of acceleration by granulating is given in [206].

According to Zhang's work, Li and Chen et al. proposed several fusion strategies of decomposition and composition of granulation in dominance rough sets, and designed the corresponding parallel algorithm for computing approximations and attribute reduction [207–209]. Qian et al. proposed the parallel attribute reduction algorithm based on the technology of MapReduce which can be accelerated the computation of equivalence classes and attribute significance [210]. To speed up the process of gene selection, Meng et al. came up with a parallel computation method to compute approximations of neighborhood rough set [211]. In RST, lots of methods for computing positive region based dependency measure that is a extremely time-consuming task, Raza and Qamar put forward a method named parallel dependency calculation which can directly find the positive region based objects [212]. Due to the parallel strategy of entire big data in a single machine is limited, Muthusamy and Subramani presented a parallel RST based the approach of attribute reduction in big data [213]. In addition, some scholars studied the parallel matrix method and parallelize it on GPUs, which can accelerate the computation of rough set approximations and have high-performance computing [127,214].

6.3. The summary of PCIF

Parallel computing refers to the process of fusing multiple computing resource to solve problems at the same time, which can save plenty of time and improve the efficiency. Therefore, it is useful for processing the large-scale data. It is worth emphasizing that the GrC can be associated with the idea of parallel computing. The calculation speed of approximations is greatly accelerated, e.g. [198]. Although parallel computing has been greatly improved in computing efficiency, how to build the scenarios of practical applications is an urgent problem. In addition, the collection of real and public large data sets for parallel computing is also a challenge.

7. Incremental learning based information fusion (ILIF)

At present, with the continuous development and wide application of storage technology, all walks of life not only accumulate a huge amount of various data, but also a large number of real-time data will be added at any time. As one of the important data types, the multi-source data widely exists in practical applications. In fact, the multi-source data probably vary rapidly with time. Therefore, dynamics has become one of the important characteristics of multi-source data, which needs accurate and high-efficient knowledge discovering and updating for the data. So, how to design a real-time computing model based on multi-source dynamic data and an efficient fusion algorithm is a hot topic in the research field of data processing. A method can effectively improve knowledge named incremental learning, which can make full use of the existing knowledge. In this section, the elementary concepts of information fusion methods based on incremental learning is introduced firstly, and then the major related researches are comprehensively summarized.

7.1. The fusion concepts of incremental learning in RST

In an information system, if the object set, attribute set and attribute's value change dynamically, it may cause the change of knowledge granulation. Furthermore, it also may lead to the dynamic variation of the approximations nature of the concept. Fig. 8 vividly depicts the lower and upper approximations updating process under the variation of attributes, where the block in the each sub-figure denotes an equivalence class.

From Fig. 8, the sub-figure (a) shows the approximate set is the key concept of RST, and its main function is the ability to describe knowledge. Whereas the remaining sub-figures show how the knowledge changes when attributes are added or removed. (1) The processes of approximations updating when Q is added into P are shown in sub-figures (b) and (c): In sub-figure (b), the purple shadows parts represent the lower approximations under P, and the green shadows represent the variation areas when Q is added to P; in sub-figure (c), the purple parts represent the upper approximations under P, and the green parts represent the variation areas when Q is added to P. (2) The processes of approximations updating when Q' is removed from P' are shown in sub-figures (d) and (e): In sub-figure (d), the purple shadows parts represent the lower approximations under P' and $P'-Q'$, respectively, and the green parts represent the variation areas when Q' is removed from P'; in sub-figure (e), the purple areas represent the upper approximations under P', and the green parts represent the variation areas when Q' is removed from P'.

Accordingly, it is necessary to introduce a technology that can fuse such changes. Incremental learning technology simulates the cognitive mechanism of human beings, and it is a gradual process with the data collected increase. The knowledge is revised, enhanced, updated, and maintained gradually according to the data added or deleted in a dynamic data environment by employing incremental learning. On one hand, the new information and old information can be fused effectively which may provide support for decision making; On the other hand, the requirement for storage and computation time may reduce greatly. Incremental learning is indispensable for knowledge acquirement in a dynamic data environment [216–218]. Therefore, incremental learning is an useful method for studying dynamic changes in information systems. Next, the mechanism incremental fusion will be discussed and its relevant definitions are also introduced.

- **The mechanism dynamic updating of knowledge granulation**

GrC provides a paradigm for analyzing and processing data from different views and hierarchies. In particular, it is necessary for big data to be analyzed via the approach of multi-granulation, which can improve efficiency and meet application requirements. In classical RST, a partition of the universe U is constituted by equivalence classes, which approximately describes any subset of objects in the universe. The finer the granularity of the equivalent class, the stronger its resolving power; otherwise, the coarser the granularity, the weaker its resolving power. For instance, the granulation measure is proposed for a partition by Yao [219], it can be defined by

$$G(U/P) = \sum_{i=1}^{n} \frac{|E_i|}{|U|} log|E_i|, \tag{7.1}$$

where $P \subseteq U$ with respect to a partition $U/P = \{E_1, E_2, \dots, E_n\}$.

For two partitions P_1 and P_2 with $P_1 \preceq P_2$, we have $G(P_1) \leq G(P_2)$. The coarsest partition $\{U\}$ has the maximum granularity value $log|U|$, and the finest partition $\{\{x\}|x \in U\}$ has the minimum granularity value 0.

Suppose that $S^t = (U^t, A^t, V^t, f^t)$ is an information system at time t where A is the attribute. $U^t/A^t = (E_1^t, E_2^t, \dots, E_n^t)$ is a partition on U^t. Let $S^{t+1} = (U^{t+1}, A^{t+1}, V^{t+1}, f^{t+1})$ is an information system at time $t+1$ and U^{t+1}/A^{t+1} is a partition on U^{t+1}. For any $E_i^{t+1} \in U^{t+1}/A^{t+1}(1 \leq i \leq m)$ and $\exists\ E_j^t \in U^t/A^t(1 \leq j \leq n)$, satisfies $E_i^{t+1} \subseteq E_i^t$, then $U^t/A^t \preceq U^{t+1}/A^{t+1}$. Hence, $U^t/A^t \preceq U^{t+1}/A^{t+1}$ is a finer partition than U^t/A^t. Below, when an object or an attribute is added to an information system, then the following properties hold [219–221].

Property 1. *When the object set $\triangle U$ is added to the universe U, i.e., $A^{t+1} = A^t, U^{t+1} = U^t \cup \triangle U$, then neither $G(U^t/A^t) > G(U^{t+1}/A^{t+1})$ nor $G(U^t/A^t) < G(U^{t+1}/A^{t+1})$ holds.*

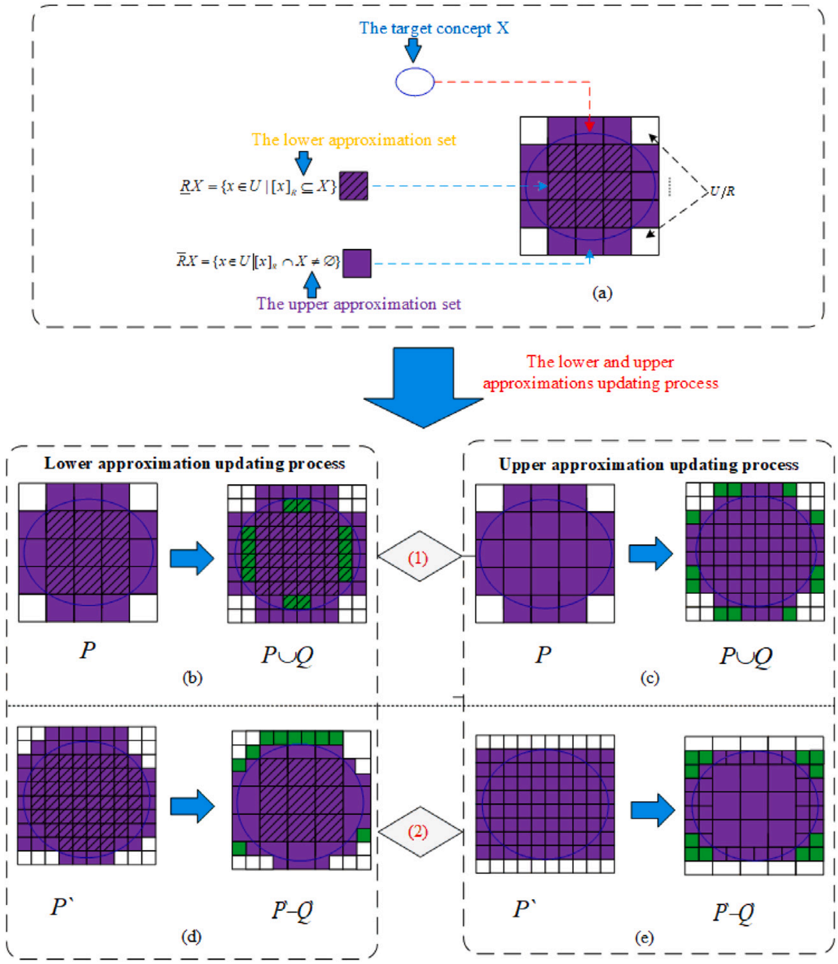

Fig. 8. The lower and upper approximations updating process when the attributes are added or removed [215]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Property 2. *When the attribute set* $\triangle A$ *is added to the universe* U, *i.e.,* $U^t = U^{t+1}, A^{t+1} = A^t \cup \triangle A$, *then* $G(U^t/A^t) \geq G(U^{t+1}/A^{t+1})$.

Property 3. *When the object set* $\triangle U$ *and the attribute set* $\triangle A$ *are simultaneously added to the universe* U, *i.e.,* $A^{t+1} = A^t \cup \triangle A, U^{t+1} = U^t \cup \triangle U$, *then neither* $G(U^t/A^t) > G(U^{t+1}/A^{t+1})$ *nor* $G(U^t/A^t) < G(U^{t+1}/A^{t+1})$ *holds.*

- **The mechanism incremental updating of approximations based equivalence class**

Let $S = (U, A = C \cup \{d\}, V, f)$ be a decision information system. $U/C = \{E_1, E_2, \ldots, E_l\}$ is a partition on U. The feature vector of the equivalence class E_i can be denoted as $\vec{E}_i = \{index_i, obj_i, reg_i\}$, where $index_i = x_k(\exists\, x_k \in E_i)$ is feature index, $obj_i = \{x_k | x_k \in E_i\}$.

(1) If $E_i \subseteq POS(X)$, then $reg_i = P$;

(2) If $E_i \subseteq BND(X)$, then $reg_i = B$;

(3) If $E_i \subseteq NEG(X)$, then $reg_i = N$.

$\vec{E}_{ic} = (e_{i1}, e_{i2}, \ldots, e_{ij}, \ldots, e_{im})$ is called the feature vector of the equivalence class E_i, where $e_{kj} = f(x_k, a_j)(\forall\, x_k \in E_i, a_j \in C)$, then the equivalence class feature matrix M_E and characteristic value matrix M_{EC} of S are denoted as follows, respectively [222].

$$M_E = \begin{pmatrix} \vec{E}_1 \\ \vdots \\ \vec{E}_j \\ \vdots \\ \vec{E}_l \end{pmatrix} = \begin{pmatrix} Index_1 & obj_1 & reg_1 \\ \vdots & \vdots & \vdots \\ Index_j & obj_j & reg_j \\ \vdots & \vdots & \vdots \\ Index_l & obj_l & reg_l \end{pmatrix}; \qquad M_{EC} = \begin{pmatrix} \vec{E}_{1c} \\ \vdots \\ \vec{E}_{jc} \\ \vdots \\ \vec{E}_{lc} \end{pmatrix} = \begin{pmatrix} e_{11} & \cdots & e_{1m} \\ \vdots & \vdots & \vdots \\ e_{j1} & \cdots & e_{jm} \\ \vdots & \vdots & \vdots \\ e_{l1} & \cdots & e_{lm} \end{pmatrix} \tag{7.2}$$

With the number of attributes increases, the cardinality of each equivalence class may decrease, yet the number of equivalence classes may increase. In addition, with the number of objects increases, the cardinality of each equivalence class may increase or the new equivalence classes may be generated. Suppose that $S^t = (U^t, A^t = C^t \cup \{d\}, V^t, f^t)$ is a decision information system at time t, where C^t is the conditional attribute and d is decision attribute. Let $S^{t+1} = (U^{t+1}, A^{t+1} = C^{t+1} \cup \{d\}, V^{t+1}, f^{t+1})$ be a decision information system at time $t+1$, where $U^{t+1} = U^t \cup \triangle U, C^{t+1} = C^t \cup \triangle C$, $\triangle U$ and $\triangle C$ represent the object set and attribute set added to the decision information system, respectively. In the dynamic variation of a decision information systems, the change of attribute set or object set finally affects the feature matrix and characteristic value matrix of the equivalence class. Therefore, how to maintain equivalence class matrix in dynamic environment is the key to incremental updating knowledge. The following theorems reveal the principle of incremental updating equivalence class feature matrix in different decision information systems.

Theorem 7.1 ([223]). *Suppose that M_E^t and $M_E^{\Delta U}$ are two the equivalence class feature matrices. For $M_E^{U^{t+1}}$, we have:*

(1) *If $\vec{E}_{ic}^t = \vec{E}_{kc}^{\Delta U}$, then $obj_i^{U^{t+1}} = obj_i^t \cup obj_k^{\Delta U}, Index_i^{U^{t+1}} = Index_i^t$, and $reg_i^{U^{t+1}}$ depends on the following two conditions: (1) If $reg_i^t = reg_k^{\Delta U}$, then $reg_k^{U^{t+1}} = reg_i^t$; (2) otherwise, (i) If $\alpha = 1, \beta = 0$, then $reg_i = B$. (ii) If $P(X^{t+1}|obj_i^{U^{t+1}}) \geq \alpha$, then $reg_i^{U^{t+1}} = P$; If $\beta < P(X^{t+1}|obj_i^{U^{t+1}}) < \alpha$, then $reg_i^{U^{t+1}} = B$; If $P(X^{t+1}|obj_i^{U^{t+1}}) \leq \beta$, then $reg_i^{U^{t+1}} = N$.*

(2) *If there is no $\vec{E}_{kc}^{\Delta U}$ that satisfies $\vec{E}_{ic}^t = \vec{E}_{kc}^{\Delta U}$, then $\vec{E}_i^{t+1} = \vec{E}_i^t$.*

(3) *Otherwise, if there is no $\vec{E}_{ic}^t$ that satisfies $\vec{E}_{kc}^{\Delta U} = \vec{E}_{ic}^t$, then $j = l^t + 1$, $\vec{E}_j^{t+1} = \vec{E}_k^{\Delta U}$.*

From Theorem 7.1, we find that when the object set is added to decision information system, there are two situations may occur in the basic knowledge granulation of the decision information system. (1) If the feature vectors of the grain in $S^{\Delta U}$ and S^t are the same, then these grains merge into one of $S^{U^{t+1}}$; (2) If the feature vectors of the grain of $S^{\Delta U}(S^t)$ is different from any grain in $S^t(S^{\Delta U})$, then this grain of $S^{\Delta U}(S^t)$ will become a new grain in $S^{\Delta U+1}$.

Theorem 7.2 ([223]). *Given $M_E^{U^{t+1}}$, we have $\vec{E}_{jc}^{t+1} = \{E_{j1}^{U^{t+1}}, E_{j2}^{U^{t+1}}, \dots, E_{j|C'|}^{U^{t+1}}, E_{i1}^{\Delta A}, \dots, E_{i|\Delta C|}^{\Delta A}\}$, and $obj_j^{t+1} \in obj_j^{U^{t+1}} / \vartriangle C$; For reg_j^{t+1}, the following conclusions holds:*

(1) *If $|obj_j^{t+1}| \neq 1$ or $obj_j^{t+1} \neq obj_j^{U^{t+1}} / \vartriangle C$, then (1) If $\alpha = 1, \beta = 0$, then (i) $reg_j^{U^{t+1}} = B$. If $(obj_i^{U^{t+1}} \cap obj_i^{\Delta U}) \subseteq X^{t+1}$, then $reg_j^{t+1} = L$; If $(obj_i^{U^{t+1}} \cap obj_i^{\Delta U}) \cap X^{t+1} = \emptyset$, then $reg_j^{t+1} = N$. (ii) Otherwise, $reg_j^{t+1} = reg_j^{U^{t+1}}$. (2) Otherwise, (i) if $P(X^{t+1}|(obj_i^{U^{t+1}} \cap obj_i^{\Delta U})) \geq \alpha$, then $reg_j^{t+1} = P$; (ii) If $\beta < P(X^{t+1}|(obj_i^{U^{t+1}} \cap obj_i^{\Delta U})) < \alpha$, then $reg_j^{t+1} = B$; (iii) If $P(X^{t+1}|(obj_i^{U^{t+1}} \cap obj_i^{\Delta U})) \leq \beta$, then $reg_j^{t+1} = N$.*

(2) *Otherwise, $reg_j^{t+1} = reg_j^{U^{t+1}}$.*

- **The mechanism incremental updating of rule extraction**

Rule extraction is one of the important applications of RST, which mainly includes attribute reduction and the calculation of approximation set. The function of attribute reduction is to remove redundant attributes. It is the expression form before simplifying rules. In addition, through the calculation of approximation sets, we can extract certainty rules from the positive region, uncertainty rules from the boundary region, and reject rules from the negative region. Therefore, when the information system changes dynamically, on the one hand, the approximate set will change, and on the other hand, the reduction will also change.

Definition 7.3 ([222]). Let $(U, A = C \cup \{d\}, V, f)$ be a decision system and $U/C = \{E_1, E_2, \dots, E_l\}$, $U/D = \{D_1, D_2, \dots, D_k\}$. $\mathbb{E}_i = \{index_i, obj_i, reg_i, \delta_i\}$ and $\vec{E}_{ic} = \{e_{i1}, e_{i2}, \dots, e_{im}\}$ are called the decision feature vector and characteristic value vector, respectively. Where $e_{kj} = f(x_k, a_j)$ for any $x_k \in E_i$ and $a_i \in C$, $\exists\, x_k \in E_i$, $index_i = x_k$. $obj_i = \{x_k | x_k \in E_i\}$. If $E_i \subseteq POS_C(U/\{d\})$, then $reg_i = P$; If $E_i \subseteq BND_C(U/\{d\})$, then $reg_i = B$, $\delta_i = \delta_C(E_i) = \{d_j | E_i \cap d_j \neq \emptyset\}$. Then, the decision feature matrix d_E and characteristic value matrix M_{EC} are defines as follows, respectively.

$$d_E = \begin{pmatrix} \vec{E}_1 \\ \vdots \\ \vec{E}_j \\ \vdots \\ \vec{E}_l \end{pmatrix} = \begin{pmatrix} Index_1 & obj_1 & reg_1 & \delta_1 \\ \vdots & \vdots & \vdots & \vdots \\ Index_j & obj_j & reg_j & \delta_j \\ \vdots & \vdots & \vdots & \vdots \\ Index_l & obj_l & reg_l & \delta_l \end{pmatrix}; \quad M_{EC} = \begin{pmatrix} \vec{E}_{1c} \\ \vdots \\ \vec{E}_{jc} \\ \vdots \\ \vec{E}_{lc} \end{pmatrix} = \begin{pmatrix} e_{11} & \cdots & e_{1m} \\ \vdots & \vdots & \vdots \\ e_{j1} & \cdots & e_{jm} \\ \vdots & \vdots & \vdots \\ e_{l1} & \cdots & e_{lm} \end{pmatrix} \tag{7.3}$$

where $1 \leq j \leq l$ and $m = |C|$.

The Definition 7.3 uses d_E and M_{EC} to describe the decision information system through the level of granularity, and the d_E contains the generalization decision information of the granule. In the attribute reduction, the discernibility matrix is constructed based on decision feature and characteristic value vectors. Therefore, for the assignment reduction, the value of generalized decision of the decision feature vector determines whether the value of discernibility attribute of the corresponding decision feature vector should be calculated. Furthermore, the generalization decision may also change with the granularity changes, and then the value of attribute. The following properties give the relationship between the generalization decision of equivalence class and two regions (i.e., positive region and boundary region), and the relationship between generalization decision and dynamic granularity, respectively [222].

Property 4. *For each $E_i \in U/C$, we have: (1) If $|\delta_C(E_i)| = 1$, then $E_i \in POS_C(U/\{d\})$, otherwise, $E_i \in BND_C(U/\{d\})$. (2) If $E_i \in POS_C(U/\{d\})$, then $|\delta_C(E_i)| = 1$, otherwise, $|\delta_C(E_i)| \neq 1$.*

Property 5. *For each $E' \subseteq E_i$ and $E_i \in U/C$, if $|\delta_C(E_i)| = 1$, then $|\delta_C(E')| = 1$.*

Definition 7.4 ([222]). Suppose that I_i represents the importance of attribute a_i in the assignment reduction, then the attribute importance matrix can be denoted as

$$AI_M = \begin{pmatrix} a_1 & I_1 \\ \vdots & \vdots \\ a_i & I_I \\ \vdots & \vdots \\ a_m & I_m \end{pmatrix}, m = |C|; \; A_M = \begin{pmatrix} a_1 \\ \vdots \\ a_i \\ \vdots \\ a_m \end{pmatrix} \; and \; I_M = \begin{pmatrix} I_1 \\ \vdots \\ I_I \\ \vdots \\ I_m \end{pmatrix}, \tag{7.4}$$

where A_M expresses as the attribute vector of AI_M, and I_M expresses as the importance vector of AI_M. The order matrix of the attribute importance can be denoted by

$$AI_M^{\geqslant} = \begin{pmatrix} a_1 & I_1 \\ \vdots & \vdots \\ a_i & I_I \\ \vdots & \vdots \\ a_m & I_m \end{pmatrix}, I_i \geqslant I_j, i, j \in \{1, 2, \dots, m\} \; and \; m = |C|, \tag{7.5}$$

where $I_i = \sum_{i=1}^{l} \sum_{j=1}^{l} N_i$, if $a_i \in D(E_i, E_j)$, then $N_i = 1$; Otherwise, $N_i = 0, l = |U/C|$.

- **The mechanism incremental updating of attribute's values**

From Fig. 9, the classification is coarser when the value domain of attribute at a higher level, and finer when the value domain of attribute is at a lower level. Therefore, the values of attribute are located in different classifications depending on the requirements of the application. In addition, the values of attribute values also likely to change when the data changes. The following definitions reveal the process of changing of attribute values.

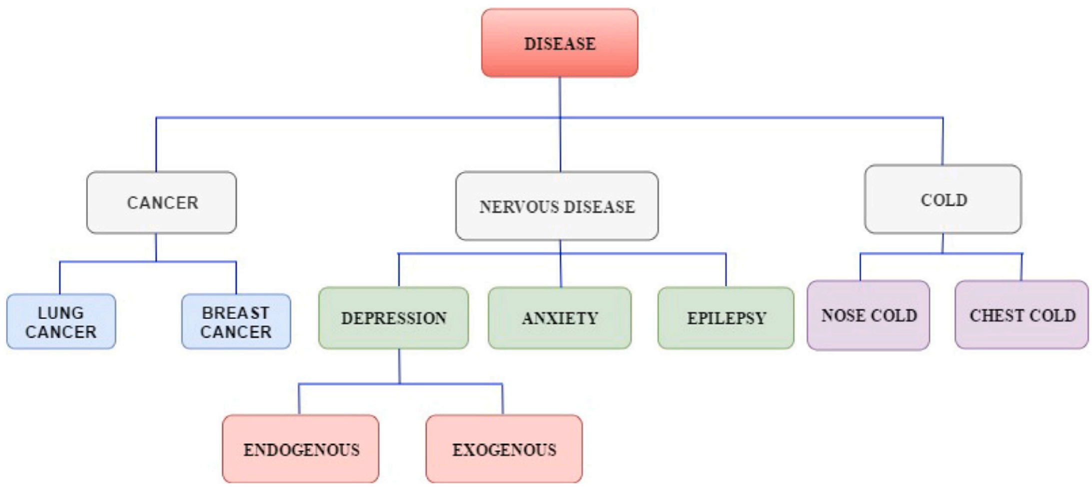

Fig. 9. The hierarchical diagram of disease classification [224].

Definition 7.5 (*[222]*)**.** Suppose that $S = (U, A = C \cap \{d\}, V, f)$ is a decision system. $\forall\ B \subseteq C$, $f(x_i, a_j)$ is the value of attribute of object x_i on attribute $a_j (a_j \in B)$. $f(x_k, a_j)$ is the value of attribute of object $x_k (i \neq k)$ on attribute a_j, where $f(x_i, a_j) \neq f(x_k, a_j)$. Pick $U_{a_j} = \{x_{i'} \in U | f(x_{i'}, a_j) = f(x_i, a_j)\}$, then the attribute value $f(x_i, a_j)$ is called coarsening to $f(x_k, a_j)$ for any $x_{i'} \in U_{a_j}$.

Definition 7.6 (*[222]*)**.** Suppose that $S = (U, A = C \cap \{d\}, V, f)$ is a decision system. $\forall\ B \subseteq C$, $f(x_i, a_j)$ is the value of attribute of object x_i on attribute $a_j (a_j \in B)$. Pick $U_{a_j} = \{x_{i'} \in U | f(x_{i'}, a_j) = f(x_i, a_j)\}$, if $\exists\ v \notin V_j, \exists\ x_{i'} \in U_{a_j}$, let $f(x_{i'}, a_j) = v$, then the attribute value $f(x_{i'}, a_j)$ of $x_{i'}$ is called refining to v.

In what follows, according to the fusion mechanisms discussed above, we will introduce the specific application of the variation of different dimensions (single and multi-dimensional) and summarize the corresponding literature.

7.2. Information fusion approach based on the variation of single-dimensional

As we all know, the timeliness is an obvious feature of dynamic data. It is specifically reflected in three different dimensions, i.e., the number of data objects (samples), the number of attributes (features), and the attribute' values (values of feature) that change with time. Generally speaking, the mechanism of ILIF is embodied in an information systems from three aspects, namely, (i) ILIF under new immigrating objects; (ii) ILIF under new immigrating attributes; and (iii) ILIF under new immigrating attribute' values. Simultaneously, it can also be said to observe the dynamic update of data from these three different variation of dimensions. The variation of single dimensional refers to the multiple objects, attributes and attribute values of an information system have changed individually and separately. It means that the variation among them do not occur at the same time. Meanwhile, there is no need to consider the interplay of these three dimensions (see Fig. 10).

• **ILIF under new immigrating multiple objects (samples):** Just as its name implies, "the variation of multiple objects" expresses the number of objects' added or deleted over time in an information system. Suppose that (U, A, V, f) is an information system where U is the universe. In this (U, A, V, f), the new objects are added can be denoted as U^+, and the obsoleted objects are deleted can be denoted as U^-. Then, according to different varies of objects, the universe is updated as $U^{t+1} = U \cup U^+$ or $U^{t+1} = U - U^-$. The approximations and reducts of an information system need to be updated. Therefore, maintaining approximations and reducts by information fusion is an efficacious way. The original information in U is fused with the information provided by U^+ (or U^-), which can avoid computing from the scratch. Consequently, it is efficient to update knowledge in time. For example, many scholars have discussed the fusion mechanism when multiple objects varying [145,218,225–235] (see Table 10).

• **ILIF under new immigrating multiple attributes (features):** "The variation of multiple attributes" expresses the number of attributes' added or deleted over time in an information system. Suppose that (U, A, V, f) is an information system where A is the attribute. In this information system (U, A, V, f), the new attributes are added can be denoted as A^+, and the obsoleted attributes are deleted can be denoted as A^-. The granularity of the information granules, the approximations of the concepts, and the attribute reducts of an information system may be obsolete. The knowledge acquired from the information system needed to be updated. Therefore, fusing the real time information into former information is an efficient way of knowledge maintenance. For instance, some scholars have studied the fusion mechanism when multiple attributes varying [20,55,68,215,217,236,237] (see Table 10).

• **ILIF under new immigrating multiple attribute' values (feature values):** "The variation of multiple attribute' values" expresses the number of attribute' values are added or deleted over time in an information system. Suppose that (U, A, V, f) is an information system where $V = \bigcup_{a \in A}$ and V_a is the domain of attribute a, f is an information system such that $f(x, a) \in V_a$ for each $a \in A, x \in U$. In this (U, A, V, f), the new attribute' values are added can be denoted as V_A^+, and the obsoleted attribute' values are deleted can be denoted as V_A^-. For any objects $x \in U$, and any attribute $a \in A$, the attribute value of a at time t and $t+1$ can be defined as $f^t(x, a)$ and $f^{t+1}(x, a)$. In fact, the attribute' values will be varied when objects and attributes change. Therefore, the knowledge acquired from an information system needed to be updated. Some articles have studied the dynamic updating fusion under the variation of attribute' values, such as [58,222,238–241] and [242] (see Table 10).

7.3. Information fusion approach based on the variation of multi-dimensional

In an information system, "the variation of multi-dimensional" means any two dimensions simultaneous variation of the objects, attributes and attribute's values (or the objects, attributes and attribute's values are varying at the same time). In the process of varying, the three dimensions may interact and influence each other (see Fig. 11).

In fact, it is often random and multi-source in the process of data generation. Particularly, it is obvious that the variation of multi-dimensional in big data and massive data environments. At present, the

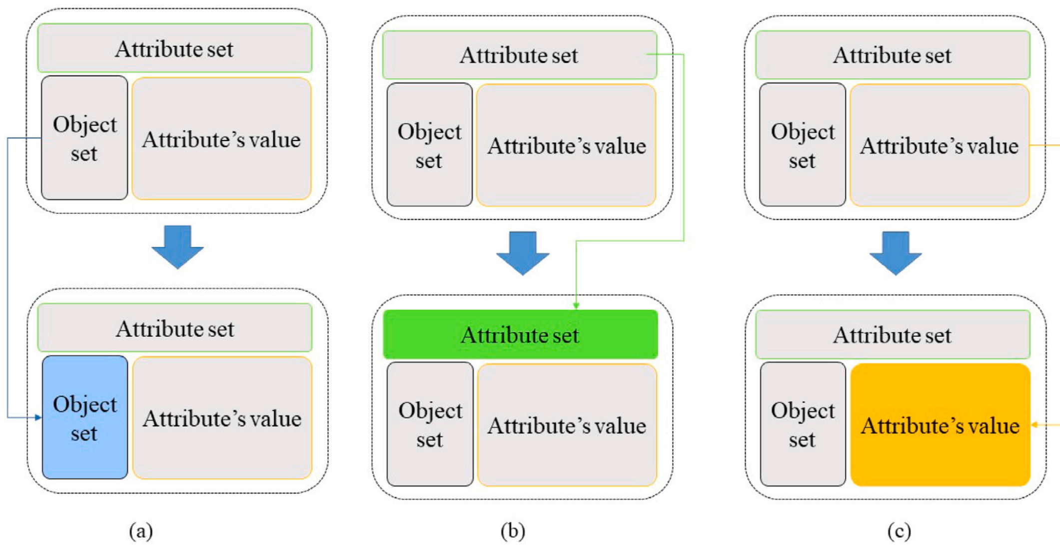

Fig. 10. The variation of single dimensional in an information system.

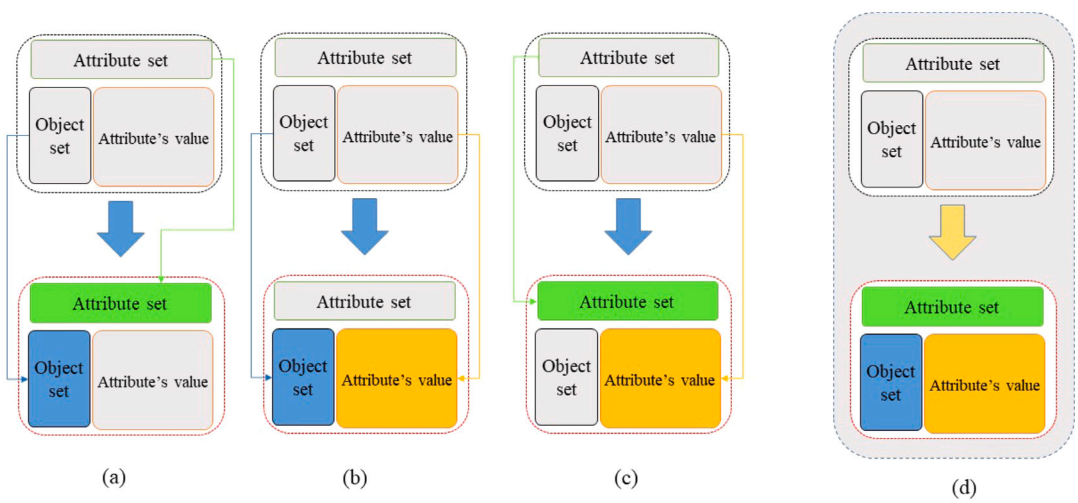

Fig. 11. The variation of multi-dimensional in an information system.

researches on incremental fusion of multi-dimensional varying have a late start with limited results. Some of the published results are shown in Table 11. In the future research, it is suggested that researchers can refer to the ideas of one-dimensional' approaches to expand the fusion model and find a better fusion mechanism.

7.4. The summary of ILIF

Since the data in actual applications changes continuously over time, how to deal with such dynamic data becomes more and more important. In RST, according to the variation of objects, attributes and attribute' values to study incremental fusion, it develops gradually a popular research topic. At present, many scholars have published a large number of books and papers, see Tables 10 and 11. However, most incremental algorithms do not filter dynamic data (e.g. the newly added data), which may interfuse low-quality information, thereby reducing the usefulness of the model. It is suggested that future research can link incremental technology with open-world learning to solve more practical problems.

8. Cluster ensembles based information fusion (CEIF)

In the era of big data, it is quite easy to collect unlabeled samples. However, it is hard to obtain samples with labels, since it may lead to plenty of manpower and resources. Therefore, cluster analysis, which is a technology how to analyze unlabeled samples to obtain the distribution characteristics of data, has become the important research contents of machine learning, pattern recognition and data mining [253–255]. Aiming at the phenomenon of unclear, vague and overlapping structure of data cluster, researchers extend the relationship between data and class cluster from traditional hard partition (i.e., there is no overlap between class clusters) to soft partition (i.e., overlap regions are allowed between class clusters) [256].

In recent ten years, cluster ensembles have emerged as an useful solution that can overcome the limitation, and improve the robustness as well as the quality of clustering results. Cluster ensemble is an emerging approach which has been extensively studied both in classification and clustering. The target of cluster ensemble is to improve the performance of measure metric by combining different classifier or cluster [257]. On

Table 10
ILIF with the variation of single-dimensional.

The variation of dimensionality.	The model of fusion	The mechanisms of incremental fusion	References
The variation of multiple objects.	Dominance-based rough set model (DRS).	Incremental updating algorithm generating satisfactory decision rules with the variation of objects	Błaszczyński and Słowiński [225]
The variation of multiple objects.	Set-valued ordered information systems (SOIS).	Updating method for computing approximations with the variation of the objects	Luo et al. [226]
The variation of multiple objects.	Variable precision rough set model (VPRS).	Incremental updating approximations when objects change dynamically in an information system.	Chen et al. [218]
The variation of multiple objects.	DRS.	Dynamic maintenance of approximations in DRS under the variation of the objects.	Li et al. [227]
The variation of multiple objects.	Decision theoretic rough set model (DTRS).	Incremental algorithms for computing approximations in DTRS with the addition or deletion of objects.	Luo et al. [243]
The variation of multiple objects.	Rough fuzzy set model (RFS).	Incremental algorithm for updating approximations of RFS under the variation of the objects.	Zeng et al. [228]
The variation of multiple objects.	Traditional RST model.	Incremental rule-extraction algorithm based on the previous rule-extraction algorithm when objects change.	Fan and huang et al. [229,230]
The variation of multiple objects.	Traditional RST model.	A group incremental rough feature selection algorithm based on information entropy when multiple objects are added in an information system.	Liang et al. [231]
The variation of multiple objects.	Traditional RST model.	Attribute reduction based on incremental learning strategies in a decision table.	Shu and Qian et al. [232,233]
The variation of multiple objects.	Traditional RST model.	An active sample selection based incremental method for attribute reduction.	Yang et al. [234]
The variation of multiple objects.	Soft set theory and RST.	Incremental feature selection approach for large-scale and high dimensional data.	Gong et al. [244]
The variation of multiple objects.	Multi-source interval-valued information systems and fuzzy set theory.	Dynamic updating of data sources with the addition or deletion of objects.	Huang et al. [32]
The variation of multiple objects	Fuzzy rough set (FRS).	Incremental method for FRS based feature selection when sample subsets are added or deleted.	Yang et al. [245]
The variation of multiple objects.	Neighborhood multi-granulation rough set (NMGRS).	Incremental methods to update knowledge in NMGRS with the variation of granular structures (objects).	Hu and Li et al. [145,235]
The variation of multiple attributes.	Probabilistic rough sets model (PRS).	Dynamically updating approximations in PRS when attributes vary.	Liu et al. [215]
The variation of multiple attributes.	Set-valued information systems (SIS).	Matrix approaches with dynamic attribute variation in SIS.	Zhang et al. [55]
The variation of multiple attributes.	Traditional RST model.	Updating approximations and attribute generalization.	Chan. [20]
The variation of multiple attributes.	Traditional RST model.	Incrementally updating the approximations when the variation of attributes.	Li et al. [217]
The variation of multiple attributes.	Traditional RST model.	Incremental strategy for reduction when the attribute set is added.	Wang et al. [238]
The variation of multiple attributes.	Incomplete decision information systems (IDISs).	Updating attribute reduction in IDISs with the variation of attributes.	Shu et al. [236]
The variation of multiple attributes.	Hybrid information systems (HIS)	Updating mechanisms for attribute reduction with the variation of the attributes.	Zeng et al. [68]
The variation of multiple attributes.	DRS	Incremental updating approximations in DRS approach with the variation of the attributes.	Li et al. [237]
The variation of multiple attribute' values.	HIS and FRS.	Dynamical updating approximations for HIS when the variation of attribute values.	Zeng et al. [69]
The variation of multiple attribute' values.	Traditional RST model.	Updating decision rules when a tribute values' coarsening and refining.	Chen et al. [222]
The variation of multiple attribute' values.	Traditional RST model.	Attribute reduction for dynamic data sets with the variation of the attribute values.	Wang et al. [238]
The variation of multiple attribute' values.	Incomplete information systems and RST	Incremental method for feature selection when the variation of the attribute values.	Shu et al. [239]
The variation of multiple attribute' values.	Traditional RST model.	Incremental reduction algorithm with varying attribute values.	Jing et al. [240]
The variation of multiple attribute' values.	Traditional RST model.	Discernibility matrix based incremental attribute reduction algorithm with varying attribute values.	Wei et al. [241]
The variation of multiple attribute' values.	Set-valued decision information systems (SDISs).	Incremental algorithms for computing rough approximations with the addition and removal of attribute values.	Luo et al. [58]
The variation of multiple attribute' values.	Covering decision information systems (CDISs).	Updating mechanisms of related families in CDISs with the variation of attribute values.	Cai et al. [242]

the account of ensemble learning theory, which fuses multiple different clustering results either from the same clustering algorithm with different initial parameters, or from the different clustering algorithms to generate a final result which has better accuracy and robust than any single clustering result. With the rapid development of information technology, it is becoming easier and easier for people to access data. Moreover, due to the roughness, vagueness and uncertainty of the data itself and the differences in human cognition levels, it becomes more difficult to obtain useful knowledge information in the massive data with complex and high-dimensional. Therefore, to solve this issue, the technologies of cluster ensembles are introduced and discussed from

Table 11
ILIF with the variation of multi-dimensional.

The variation of dimensionality	The model of fusion	The mechanisms of incremental fusion	References
The variation of multiple objects and attributes.	DTRS.	Dynamic approach of approximations when objects and attributes added simultaneously.	Chen et al. [246]
The variation of multiple objects and attributes.	RFS.	Incremental mechanisms for updating approximations when adding the objects and attributes concurrently.	Huang et al. [247]
The variation of multiple objects and attributes.	Three-way decision model (3WD) and DTRS.	Incrementally updating three-way probabilistic region.	Yang et al. [248]
The variation of multiple objects and attributes.	Decision information systems (DISs).	Incremental approach for computing reduction when objects and attributes of the DISs change dynamically.	Jing et al. [249]
The variation of multiple objects and attributes.	OIS and DRS.	Dynamic DRSA for the multi-dimensional variation of an OIS.	Wang et al. [250]
The variation of multiple objects and attribute' values.	OIS and DRS.	Updating approximations with variation of objects and attributes.	Wang et al. [251]
The variation of multiple attribute and attribute' values.	OIS and DRS.	Updating approximations with variation of attribute and attribute' values.	Wang et al. [252]

multiple angles of information fusion, which combined with the main theoretical models such as RST, fuzzy set theory and so on.

8.1. Cluster ensembles

Cluster ensemble, also known as cluster fusion, was first proposed by Strehl et al. Its specific expression is described as follows [258]:

Given a data set $X = \{x_1, x_2, \ldots, x_n\}$ that contains N objects. If run H times clustering algorithms on the data set X, then obtain H clustering results that can be represent as $P = \{p_1, p_2, \ldots, p_H\}$, where the ith clustering result is expressed as $P_i = \{C_i^1, C_i^2, \ldots, C_i^{k_i}\}, i = 1, 2, \ldots, H$, and k_i shows the number of class cluster with respect to the member of cluster P_i. Next, a function Γ is designed to fuse all cluster members $(P_1, P_2, \ldots, P_H)$ to obtain the final clustering result P^f. Γ is called consensus function or fusion function, which is the key issue of cluster ensembles algorithm research. It integrates multiple different cluster members to obtain a unified clustering result can be expressed as [259]:

$$\Gamma : \{P_i | i \in \{1, 2, \ldots, H\}\} \to P^f \tag{8.1}$$

Fig. 12 displays the basic process of cluster ensembles.

Since the concept of cluster ensembles was proposed, a large number of improved algorithms for cluster ensembles have emerged for more than ten years [257]. In [257], it reviewed different applications and extensions of cluster ensembles. However, there are relatively few methods to deal with uncertain, inaccurate, and overlapping information in the integration process. Whereas the RST, as a core part of GrC, forms the granules via the equivalence relation determined on objects of the universe as well as approximately express information granulation based on the lower and upper approximation operators. Accordingly, using RST to study cluster ensembles will be a very promising research direction. For example, the expressing methods of overlap regions of a class cluster which can be expressed as two regions. One is the lower approximation region (i.e., data points definitely belong to this cluster), and the other is boundary region (i.e., data points may belong to this cluster). It follows that applying RST to clustering algorithms can reveal the uncertain structure of data sets [260–264]. In addition, the traditional "hard clustering" is transformed into "soft clustering", which enriches the ability of clustering analysis algorithm to dig deeply into the potential information of data, and can more clearly show the internal correlations of the actual data set.

It can be seen from the current research that many results have been achieved in the research of cluster ensembles algorithms, and the related algorithms of clustering analysis by combining specific theoretical models with cluster ensembles algorithms have also been studied [257]. However, the ability to explore the potential structure of data in an uncertain environment is still relatively lacking, and the methods need to be improved. Therefore, there are still facing various problems for the effective processing of cluster ensembles algorithms on imprecise and inconsistent data sets as follows.

(1) In cluster ensembles algorithms, the diversity of base clustering results may produce good final results. Nevertheless, the traditional cluster ensembles algorithms fusion of all base cluster members will not only increase the time and space overhead, but also low-quality cluster members will cause undesirable interference to the final fusion result [265–269]. Therefore, it may affect the accuracy of the final result. To address this problem, some cluster ensemble selection algorithms have been proposed by scholars [259,270,271]. The core idea of solving this problem is to merge by selecting part of the ensemble members, which may get better clustering results than fusing all cluster members.

(2) The traditional cluster ensembles selection algorithms mostly regard the base clustering results as a whole, and it is generally considered that all clusters in the same base clustering result have the same stability. Therefore, when the quality of base clustering results is measured, a global weight is assigned to each base clustering result and the differences between clusters and clusters contained in the results are ignored.

(3) In real data sets, due to the inherent complexity of the data itself as well as the deviation of application scenario of a specific clustering algorithm, even if the division results obtained by the same clustering algorithm, the reliability between clusters may not be necessarily the same.

(4) The evaluation index is applied to select the base clustering results, which contain one or more highly stable clusters may be abandoned due to the weights obtained by the whole member individuals are not high. This will lead to the loss of valuable information during the fusion process, and may not be conducive to yielding high-quality fusion results.

(5) Fuzzy clustering is a common soft clustering algorithm, which quantifies the possibility of data samples belonging to each cluster by constructing a membership matrix [272–278]. However, the fuzzy clustering algorithm is inherently uncertain, and is affected by the random initial cluster center and fuzzy factor, then each base clustering result obtained may not be necessarily valid. Namely, if the results with too much fuzziness are involved in the final result fusion, not only the time complexity will be increased, but also the final clustering result will be affected. Therefore, it is one of the methods to improve the performance of fuzzy cluster ensembles algorithms to select the fuzzy clustering algorithms before the fusion.

Since RST has certain advantages in dealing with uncertain problems, in response to the above problems, in what follows, we introduce the related work of RST in clustering and cluster ensembles. These works reveal the role of RST in cluster analysis and also indicate its contribution to information fusion.

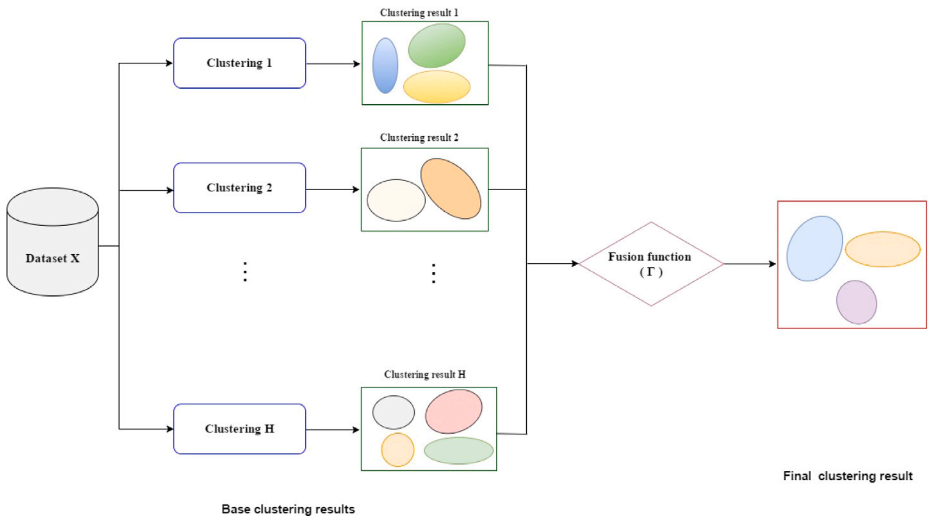

Fig. 12. The basic process of cluster ensembles.

8.2. Rough sets for clustering and cluster ensembles

8.2.1. Rough sets for clustering

Currently, a series of methods based on rough set clustering have emerged and formed an important research branch of cluster analysis, which greatly enrich the clustering fusion. For example, Lingras and West first proposed the adaptation of the k-means algorithm based on RST [279]. Zhang et al. defined a hybrid imbalanced measure for c-means clustering which is modified [280]. To deal with uncertainty in clustering process, Herawan et al. presented a method based on RST which is called maximum dependency attribute [281]. According to DTRS model, Li et al. proposed an extended rough c-means clustering algorithm [282]. Cao et al. studied a novel initialization method for clustering based on neighborhood-based rough set model [283]. Yanto et al. applied the variable precision rough set model for clustering groups student in each study's anxiety [284]. To avoid a retraining process, Pacheco et al. addressed a unsupervised algorithm with incremental learning property for feature selection based on attribute clustering and RST [285]. Peters et al. proposed a dynamic rough clustering algorithm which is applied to synthetic and real-world data sets [262]. For the problems of categorical data, Parmar et al. researched a new algorithm (named Min–Min-Roughness) for clustering categorical data based on RST, which has the ability to handle the uncertainty in the clustering process [286]. Li et al. came up with a new clustering algorithm named total mean distribution precision (TMDP) for selecting the partitioning attribute based on PRS [287].

Combined with the above description, we briefly introduce the clustering fusion technologies commonly based on RST.

• **Roughness for clustering:** Mazlack et al. designed a RST approach for clustering by choosing partitioning attributes [288]. In their paper, which applied a measure named total roughness (TR) to control the crispness of the partition. Nevertheless, there is a handicap between binary valued attributes and multi-valued attributes for partitioning. Namely, even if the partition of a binary attribute is over, the total roughness is also lower for a multi-valued attribute. To overcome the issue, Parmar et al. [286] proposed the MMR algorithm for clustering objects on all attributes. In what follows, some techniques of calculating the roughness (accuracy of roughness, mean roughness, total roughness and min–Min-Roughness) for clustering are described.

(1) Accuracy of roughness [289]: Given an information system (U, A), where U is object set and A is attribute set. For any $X \subseteq U, B \subseteq A$, the accuracy of approximation of X w.r.t. B is defined by

$$\alpha_B = \frac{|\underline{B}(X)|}{|\overline{B}(X)|}. \tag{8.2}$$

The formula (8.2) shows that the value of higher of α_B, the more precise of itself.

(2) Mean roughness(MR) [288]: Given an information system (U, A), where U is object set and A is attribute set. Given $a_i \in A$ and $X \subseteq U$. For attribute a_i has k-different value, denoted $\gamma_k (1 \le k \le n)$. Suppose that $X(a_i = \gamma_k)$ is a subset of the objects which have k-different values of attribute a_i. Then the mean roughness of attribute $a_i \in A$ w.r.t. $a_j \in A$ is defined as

$$MR_{a_j}(a_i) = \frac{\sum_{k=1}^{|V(a_i)|} R_{a_j}(X|a_i = \gamma_k)}{|V(a_i)|}, \tag{8.3}$$

where $R_{a_j}(X|a_i = \gamma_k) = \frac{|\underline{X_{a_j}}(a_i=\gamma_k)|}{|\overline{X_{a_j}}(a_i=\gamma_k)|}$, $R_{a_j}(X)$ is the roughness of X w.r.t $a_j (i \neq j)$.

(3) Total roughness (TR) [288]:

$$TR(a_i) = \frac{\sum_{k=1}^{|V(a_i)|} TR_{a_j}(a_i)}{|A| - 1}, \tag{8.4}$$

where attributes $a_i, a_j \in A, i \neq j$. If the total roughness is the highest value, then the attributes of clustering will be the best selection.

(4) Min–Min-Roughness (MMR) [286]:

$$MMR_{a_j}(X|a_i = \gamma_k) = 1 - R_{a_j}(X|a_i = \gamma_k) = \frac{|\underline{X_{a_j}}(a_i = \gamma_k)|}{|\overline{X_{a_j}}(a_i = \gamma_k)|}. \tag{8.5}$$

It is obvious that MMR method is for measuring the roughness of the set $X(a_i = \gamma_k) 1 \le k \le n$ w.r.t. $a_j (i \neq j)$.

• **Rough k-means clustering:** Lingras and West [279] first proposed the concepts of rough k-means clustering, however, it is derived from the interval interpretation of rough sets in contrast to the original set based on RST. In [290], Yao discussed these two kind of views on RST in detail. Namely, (i) if an object belongs to lower approximation of a cluster k, then it does not belong to any other cluster; (ii) the lower approximation of kth cluster is contained in the upper approximation of kth cluster; and (iii) if an object is not a member of any lower approximation, then it is member of at least two upper approximations.

In [291], Peters gave the computing paradigm of $mean_k$, it is defined as follows:

$$mean_k = \begin{cases} \underline{\omega} \sum\limits_{x_n \in \underline{C_k}} \frac{x_n}{|\underline{C_k}|} + \overline{\omega} \sum\limits_{x_n \in |\widehat{C_k}|} \frac{x_n}{|\widehat{C_k}|}, & for\ \underline{C_k} \neq \emptyset \wedge \widehat{C_k} \neq \emptyset, \\ \sum\limits_{x_n \in \underline{C_k}} \frac{x_n}{|\underline{C_k}|}, & for\ \underline{C_k} \neq \emptyset \wedge \widehat{C_k} \neq \emptyset, \\ \sum\limits_{x_n \in \widehat{C_k}} \frac{x_n}{|\widehat{C_k}|}, & for\ \underline{C_k} \neq \emptyset \wedge \widehat{C_k} \neq \emptyset, \end{cases} \tag{8.6}$$

where $\underline{C_k}$, $\overline{C_k}$ and $\widehat{C_k}$ are denoted as belonging to the lower approximation of C_k, belonging to the upper approximation of C_k, and the boundary of C_k, respectively. Moreover, x_n is object, $\underline{\omega}$ and $\overline{\omega} = 1 - \underline{\omega}$ are the weights.

It is worth mentioning that there are also other approaches to calculate the $mean_k$, such as [292,293] and [294]. Additionally, by combining RST and fuzzy sets, there are also a number of papers are published. For instance, Mitra et al. [295] introduced a new clustering architecture, which can be processed together with an objective of finding a common structure. Peter et al. compared servals soft clustering methods, such as k-means, fuzzy c-means and rough k-means, and gave some examples where these methods are used in studies [256]. Saha et al. first put forward a clustering method by fusing RST and FST [296]. Moreover, they proposed an integrated clustering technique applying multi-phase learning based on rough fuzzy k-modes [292]. Simiński presented rough fuzzy subspace clustering algorithm which uses marginalization, imputation and rough sets for dealing with missing values [297]. In order to deal with the automated brain tumor segmentation of MR image, Bal et al. presented a clustering method using rough fuzzy c-means and shape based topological properties [298].

8.2.2. Rough sets for cluster ensembles

Several works in the clustering literature suggest that the RST-based method is crucial for a successful cluster ensembles. For example, Hu et al. proposed an incremental fuzzy cluster ensemble method based on RST [299]. They first used the basic of methods of soft clustering such as FCM and rough-k-means to obtain the positive region, boundary region and negative region of cluster ensembles. Then, the supervised learning method is introduced, i.e., random forest. The hierarchical clustering on the samples is applied in the positive region, and the result is used as the initial training set of the random forest classifier to predict the sample category in the boundary region. Finally, an incremental strategy is employed to train the samples in the boundary region to obtain the sample category in the negative region. The proposed algorithm is not sensitive to the size of the base cluster members. And experiments have proved that the algorithm is better than a single fuzzy clustering algorithm and most cluster ensembles algorithms. In [300], Hu et al. presented objective function for cluster ensembles in the framework of GrC, knowledge distance based on roughness knowledge granulation, and agglomeration degree based on total knowledge granulation by employing basic concepts in RST. Furthermore, they further studied performance metric rough distance as one of measurements, and showed that the RST based cluster ensembles is an effective method. In general, this is a top-down hierarchical cluster ensembles algorithm which provides a new idea for cluster ensembles from the perspective of knowledge granulation. To solve the problem of redundant unrelated attributes in high-dimensional small sample data, Gao et al. put forward a cluster ensembles algorithm based on rough subspaces. The attribute reduction algorithm in RST is used to delete noise attributes and generate multiple related attribute subspaces, and the high-quality subspace is selected to produce multiple different clustering results [301]. This not only preserves the granular structure of the original data, but also increases the diversity of base clustering results. And an internal metric is called cluster cardinality index (CCI) based on a set operator is designed to evaluate the quality of the classification data base clustering results, and finally effective clustering results are obtained via the fusion function. Lingras et al. analyzed the partial order relationship between clusters in clustering results, and used the intersection operation of clusters to fuse the hard clustering and rough clustering results generated in different attribute subspaces [302]. Wang et al. came up with a reduction method of rough set based on the significance of attribute, which was an unsupervised feature selection method based on ranking and forward selection strategies as well as applied it in the clustering ensembles algorithm [303].

• **Clustering ensemble selection based on granulation distance (CESGD):** In an information system, the knowledge granulation is a main method to measure the uncertainty. It is defined as follows:

Definition 8.1 (*[304,305]*). Given an information system $S = (U, A)$ and $U/A = \{X_1, X_2, \dots, X_m\}$. Then $GK(A)$ is a knowledge granulation in (U, A), if it satisfies (i) Non-negativity; (ii) Invariability; and (iii) monotonicity. The knowledge granulation of A can be defined as

$$GK(A) = \frac{1}{|U^2|}\sum_{i=1}^{m}|X_i|^2, \tag{8.7}$$

where $\frac{1}{|U|} \le GK(A) \le 1$.

If the knowledge granulation of one partition is the same as that of another partition, we say that these two partitions have the same uncertainty. However, it does not mean that these two partitions are equal, i.e., knowledge granulation cannot quantitatively describe the difference between the two partitions. Therefore, the concept of granulation distance is given as follows.

Definition 8.2 (*[183,306]*). Given a knowledge base $(U, \mathbf{R})$, $P, Q \in \mathbf{R}$. $K(P) = \{[x_i]_P | x_i \in U\}$ and $K(Q) = \{[x_i]_Q | x_i \in U\}$ are two partitions induced by two equivalence relations P and Q. The granulation distance between $K(P)$ and $K(Q)$ is defined by

$$dis(K(P), K(Q)) = \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|[x_i]_P \oplus [x_i]_Q|}{|U|}, \tag{8.8}$$

where $(U, \mathbf{R})$ is a finite and non-empty set and $\mathbf{R}$ is a family of equivalence relations, and $|[x_i]_P \oplus [x_i]_Q| = |[x_i]_P \cup [x_i]_Q| - |[x_i]_P \cap [x_i]_Q|$, $0 \le dis(K(P), K(Q)) \le 1 - \frac{1}{|U|}$.

Therefore, based on the idea of granularity, the whole original data set can be regarded as the universe of an information system, and each base clustering result is regarded as a newly added attribute of the original data. It is similar to the idea of partition the fields by attributes in information systems. The diversity of base clustering results is measured by granulation distance based on knowledge granulation, and base clustering results with large differences and high quality are selected to participate in the final fusion. This fusion method is called base clustering selection based on granulation distance. According to the definition of an information system in RST, the cluster ensemble problem is converted to be solved in the corresponding cluster ensemble information system. The specific description is as follows:

Definition 8.3 (*[300]*). The 4-tuple $S = (U, P, V, f)$ is called a cluster ensemble information system, where

(i) $U = \{x_1, x_2, \dots, x_n\}$ represents the universe of the cluster ensemble information system which includes n objects.

(ii) $P = \{p_1, p_2, \dots, p_H\}$ represents a set of H partitions on U, which represents H clustering results generated by H base clusterings. $\forall\ p_i \in P$, it divides U into $|V_{p_i}|$ subsets, namely, $U/p_i = \{X_{i1}, X_{i2}, \dots, X_{i|V_{p_i}|}\}$, where $|V_{p_i}|$ shows the number of granules determined by partition p_i;

(iii) $V = \bigcup_{p_i \in P} V_{p_i}$, where V_{p_i} represents the label domain of a partition p_i; and

(iv) f represents an information function $f : U \times P \to V$ such that $f(x_i, p_j) \in V_{p_j}$, $\forall\ x_i \in U$, $p_j \in P$. $f(x_i, p_j)$ reflects the category label generated by any element x_i in the universe of discourse by the jth partition p_j.

Definition 8.4 (*[300]*). Suppose that $P = \{p_1, p_2, \dots, p_H\}$ is H base clustering results on U. For each $p_i, p_j \in P$, then the dissimilarity between partition p_i and a set of partitions P is defined by

$$Dis(p_i, P) = \sum_{i=1}^{H} dis(p_i, p_j), \tag{8.9}$$

where p_j means unselected partition on U.

Moreover, the cluster member P_c with the minimum granulation distance is selected in P, and add it to the set S_j, where

$$S_j = \arg\min_{p_c} Dis(p_c, P) \tag{8.10}$$

Then, the remaining cluster members are selected iteratively, the criterion is to share the minimum distance with the candidate cluster member set and the maximum distance from the selected cluster member set. The partition P_λ meets the condition that is added to the selection set until the specified number of selected members H' is satisfied. In other words, P_λ is the objective function which can be defined by

$$p_\lambda = \arg\max_{p_i} \left(\frac{\mathrm{Dis}\left(p_i, S_j\right)}{|j|} - \frac{\sum_{q=1}^{\left|P\setminus S_j\right|-1} \mathrm{Dis}\left(p_i, p_q\right)}{|H-1|-|j|} \right), p_i, p_q \in P\setminus S_j, i \neq q,$$
$$\lambda = 2, 3, \dots, H', \tag{8.11}$$

where S_j is the set of currently selected cluster members, $|j|$ is the number of selected member sets, $\left|P\setminus S_j\right|$ is the number of candidate cluster member sets.

• **Method for generating elements of co-association matrix based on GrC (MGECMG):** The method based on co-association matrix is a commonly used clustering fusion algorithm, whose main idea is to regard each base clustering result as a new data reorganization pattern [307–310]. On this basis, the traditional method of directly calculating the similarity between data points with various distance measures is replaced by the similarity measurement between the new modes. Then, a co-association matrix A_{ij} based on granular partition is defined as follows.

$$A_{ij} = \frac{\sum_{\lambda=1}^{H} \delta\left(x_i, x_j\right)}{\sum_{i=1}^{H} \left|V_{p\lambda}\right|}, A_{ij} \in [0,1];$$
$$\delta\left(x_i, x_j\right) = \begin{cases} \left|V_{p\lambda}\right|, & \text{if } C^{p\lambda}\left(x_i\right) = C^{p^\lambda}\left(x_j\right) \\ 0, & \text{if } C^{p\lambda}\left(x_i\right) \neq C^{p\lambda}\left(x_j\right) \end{cases}, \tag{8.12}$$

where $\left|V_{p\lambda}\right|$ represents the number of class clusters in partition result $p\lambda$, $C^{p\lambda}\left(x_i\right)$ represents the label corresponding to the class cluster in the partition result $p\lambda$ of the sample x_i.

The co-association matrix A_{ij} can be regarded as the similarity matrix of a sample, which is applied as the input clustering algorithm, and the final clustering result is obtained via the fusion function. Obviously, $A_{ij} \in [0,1]$. According to the value of A_{ij}, the probability of belonging to the same class cluster among samples can be inferred. For example, the following three special cases are considered to explain:

(1) If $A_{ij} \approx 1$, it means that the two samples are divided into one class in most of the base clustering results, then it is more likely to belong to the same class in the final result.

(2) If $A_{ij} \approx 0$, it means that the two samples are not divided into the same class in most of the base clustering results, then it is less likely to belong to the same class in the final results.

(3) If $A_{ij} \approx 0.5$, it means that the probability of two samples being divided into the same class and different classes are almost equal to half in all clustering results, then it is more difficult to intuitively judge the category of the two samples.

• **Incremental fuzzy cluster ensemble learning based on RST (IFCERS):** In 2017, Hu et al. first used the soft clustering method to obtain the positive, boundary and negative regions of cluster ensembles [299], and proposed an incremental fuzzy cluster ensemble learning based on RST (IFCERS). The overview of the IFCERS is showed in Fig. 13.

From Fig. 13, the fusion steps of clustering ensemble are described as follows.

Step 1: According to soft clustering algorithms (such as FCM, rough k-means and rough-fuzzy k-means), the number of clusters are employed to obtain multiple clustering solutions.

Step 2: Each clustering solution is converted into fuzzy membership matrix. Then, a fuzzy cluster ensemble is formed via selecting the clustering solutions with lower fuzzy partition entropy.

Step 3: Based on the principle of rough approximation of RST, the positive, boundary and negative regions of fuzzy clustering ensemble are obtained.

Step 4: A fuzzy cluster ensemble technology is applied to acquire the group structure of data points in the positive region.

Step 5: The method of random forests is used incrementally to points in the boundary and negative regions with the group structure of the positive region.

The IFCERS provides a novel research approach to deal with overlapping clusters, outliers or uncertain cluster memberships in data sets. The core idea is to integrate RST into the noise filtering of fuzzy clustering results.

8.3. An extended clustering fusion technology: Three-way clustering

In [311], Ma emphasized that cluster analysis is a central topic in GrC due to a cluster can be conveniently seen as a granule. Moreover, clustering algorithms provide a number of methods to construct a useful and interpretable information granules. For example, [300] and [312] produced multilevel granular structures by handling hierarchical clustering algorithms. Based on this, the two clustering models (rough clustering and interval clustering) can be unified a framework of three-way cluster analysis, in which a cluster is expressed as a nested pair of sets instead of one. Therefore, Yu introduced a novel framework for three-way clustering [313]. Namely, based on Three-Way Decisions (3WD), there are also having three relationships between an object and a cluster: (i) The object certainly belongs to this cluster, (ii) the object certainly does not belong to the cluster, and (iii) the object might not belong to the cluster. Hence, the relationship between objects and clusters can be formed a 3WD model, which is introduced into the cluster analysis problem.

• **The basic concepts of three-way clustering** [313]

Suppose that $U = \{x_1, x_2, \dots, x_n\}$ is a object set, and $C = \{c_1, c_2, \dots, c_K\}$ is a family of clusters. Then, for each cluster C_K, which can be represented as a pair of sets $C_K = \{In(c_K), Pt(c_K)\}$, where $In(c_K), Pt(c_K) \subseteq U$. The pair of sets are used to create the three regions of a cluster as follows.

$$\begin{aligned} Inside(c_K) &= In(c_K), \\ Partial(c_K) &= Pt(c_K), \\ Outside(c_K) &= U - In(c_K) - Pt(c_K), \end{aligned} \tag{8.13}$$

where $Inside(c_K)$ means the objects that belong to the cluster c_K, $Partial(c_K)$ means the objects that might not belong to the cluster c_K, and $Outside(c_K)$ means the objects that does not belong to the cluster c_K. In Yao's work [139], he introduced a trisecting–acting-outcome (TAO) model and provided an architectural framework for the 3WD. In Fig. 14, a three-way clustering framework based on the TAO model is compared. Apparently, the three-way clustering framework can also be interpreted based on the general TAO framework of 3WD. In the trisection step, the set of objects can be partition three disjoint regions, which corresponding to a particular cluster.

• **Application of 3WD in overlapping clustering**

In this era of increasing complexity and diversity of information, the classification of information becomes more and more significance. However, many applications in real life, the disadvantages of general clustering methods are that the objects can only belong to one class. For example, social network structure analysis, genetic data and biological information processing which generally exist overlap between classes,

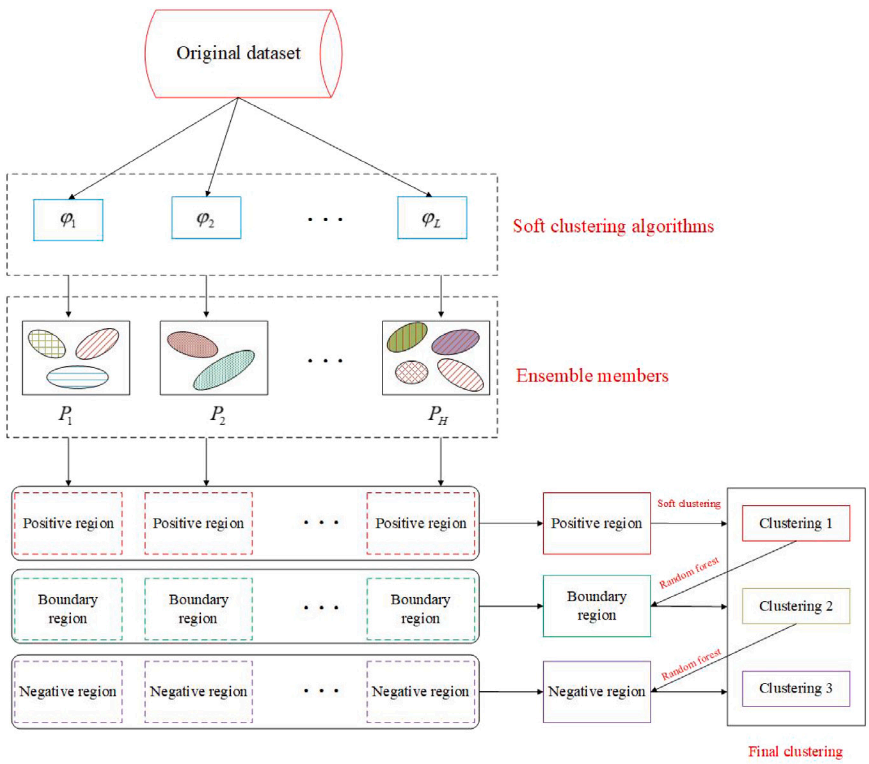

Fig. 13. An overview of the IFCERS [299].

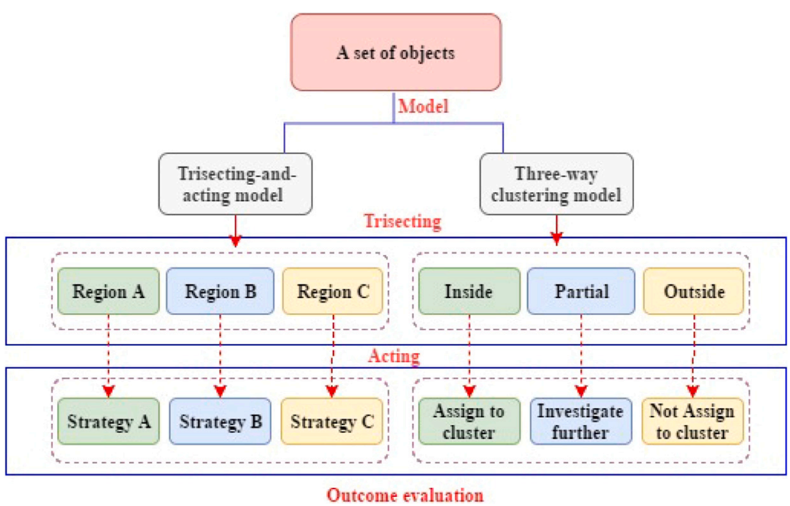

Fig. 14. A three-way clustering framework based on the TAO model [139].

they cannot be classified according to the conventional clustering algorithms. It is said that the soft partition (i.e., soft clustering) means the overlap regions are allowed between class clusters. Therefore, the 3WD can be used to cluster the data with class overlap. For instance, consider any two classes C_i and C_j, then the positive region and boundary region of class C_i can be represented as $POS(C_i)$ and $BND(C_i)$. Similarly, for the class C_j can be expressed as $POS(C_j)$ and $BND(C_j)$. In addition, there are two scenarios that need to be considered: (i) When the positive region of a class overlaps with the positive and boundary regions of the other classes, only the positive region part is considered; (ii) When there is no overlap between the positive region of a class and other classes, the overlap between the boundary region of the class and other classes is considered. Therefore, there are three kinds of overlaps between classes, i.e., positive region overlaps with positive region, positive region overlaps with boundary region, and boundary region overlaps with boundary region. Fig. 15 shows classes overlap with each other.

In the case of above, three types of overlap degrees are defined as:

(1) Positive region overlaps with positive region

$$ODPP(C_i, C_j) = \frac{|POS(C_i) \cap POS(C_j)|}{|POS(C_i) \cup POS(C_j)|} \tag{8.14}$$

(2) Positive region overlaps with boundary region

$$ODPB(C_i, C_j) = \frac{|POS(C_i) \cap BND(C_j)|}{|POS(C_i) \cup POS(C_j) \cup BND(C_j)|} \tag{8.15}$$

(3) Boundary region overlaps with boundary region

$$ODBB(C_i, C_j) = \frac{|BND(C_i) \cap BND(C_j)|}{|BND(C_i) \cup POS(C_j) \cup BND(C_j)|} \tag{8.16}$$

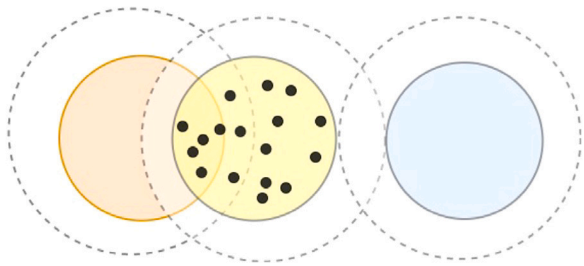

Fig. 15. Classes overlap with each other [314].

where $|C|$ expresses the number of elements in set C, $i, j \in N^*$.

According to the definitions of the types of overlap and overlap degree, the merging strategy of classes and classes is defined by combining with the initial clustering results. Additionally, this merge strategy also needs to introduce a pair of thresholds (α, β). For example, three cases are also given for discussion:

(1) When $ODPP(C_i, C_j) \geq \alpha$, the positive regions of C_i and C_j are merged into the positive region of a new class C_k, and the boundary regions of two types of C_i and C_j are merged into the boundary region of a new class C_k. However, when $\beta \leq ODPP(C_i, C_j) < \alpha$, the positive regions of classes C_i and C_j will be divided into each other's boundary region.

(2) When $ODPB(C_i, C_j) \geq \alpha$, the positive and boundary regions of class C_j are added to the positive region of C_i. However, when $\beta \leq ODPB(C_i, C_j) < \alpha$, the positive and boundary regions of class C_j are added to the boundary region of C_i.

(3) When $ODBB(C_i, C_j) \geq \alpha$, the positive and boundary regions of class C_j are added to the boundary region of C_i. However, when $\beta \leq ODBB(C_i, C_j) < \alpha$, one can do not to take any action to avoid oversize the boundary region.

This section by no means attempts to cover every research field in which clustering is used. To sum up, the 3WD has an important application in overlapping clustering, especially the definition of overlapping degree and the strategy of class–class merger fusion, which provides a good solution for overlapping clustering. Nevertheless, some researches on three-way clustering are increasing gradually in recent years. For example, by combining mathematical morphology and 3WD, Wang and Yao proposed a framework of contraction-and-expansion based three-way clustering named CE3 [315]. To overcome the low utilization, and improve energy-efficient, Jiang et al. put forward a clustering weight algorithm which is called three-way clustering weight based on 3WD [316]. Afridi et al. explored the game-theoretic rough sets, and applied the three-way clustering method for dealing with missing datasets and overlapping clustering [317,318]. To better deal with the relationship between elements and clusters, Yu et al. presented three-way clustering method based on an improved DBSCAN algorithm [319]. In addition, in order to obtain the clustering threshold value in a decision-making environment, Liu and Zhang proposed a three-way gray incidence clustering method [320]. Yu et al. presented a three-way density peak clustering method based on evidence theory (3WDPET), which forms clusters as interval sets using three-way clustering representation including three disjoint regions called positive, boundary, and negative regions [321]. For multiple sources and high dimensionality of datasets, Yu et al. investigated an active three-way clustering method to improve clustering accuracy [131].

8.4. The summary of CEIF

Currently, we have introduced some cluster ensembles methods, such as CESGD, MGECMG and IFCERS. Compared with the traditional cluster ensembles methods [257], the cluster ensembles fusion technologies of soft clustering mentioned in this paper are more suitable for the study of the phenomenon of uncertainty. So far, the cluster ensembles algorithms for data processing are mainly based on traditional machine learning methods. For example, Bai et al. proposed a semi-supervised clustering method to fuse different types constraints which from multiple information sources [322]. In the previous subsection, many methods based on RST for cluster analysis have been introduced. In addition, a new fusion model is also discussed to carry out clustering, i.e., three-way clustering. The representation of three-way clustering is different from traditional representation which divides the clustering result C_i into three disjoint parts: $POS(C_i)$, $BND(C_i)$ and $NEG(C_i)$. This will be the focus of future research, e.g. [131] and [321]. However, the disadvantage of the some algorithms of cluster ensembles are sensitive to initial parameter settings and highly dependent on the sample of data. In addition, the construction of the objective function is also a difficulty. From the perspective of information fusion, we can combine the existing MSIF models for research. For example, the multi-granulation can construct approximations regarding multiple granulations and gradually receives widespread attention in GrC. Therefore, how to develop three-way clustering based on multi-granulations will be a direction worthy of research.

9. Conclusion: Findings and future directions

In this paper, our aim is to introduce the research progress of MSIF based on RST including conventional models and techniques, which are MSIF models (homogeneous and heterogeneous MSIF models), MvRSIF (MgIF, MsIF and MvDIF models), PCIF (MapReduce and MP–DP models), ILIF (ILIF under new immigrating multiple objects, attributes and attribute' values) and CEIF (rough sets for cluster ensembles, rough sets for clustering and three-way clustering). In this section, we further discuss this paper from findings and future directions.

• **Findings**

We summarize the recent developments of MSIF research in RST, and the following significant findings can be extracted:

(1) MSIF models, MvRSIF and ILIF technologies are still playing a dominant role in RST research, while PCIF models and CEIF technology emerge in recent years;

(2) Regarding to MSIF understanding, all models and technologies are aimed at data processing.

(3) Heterogeneous MSIF models and incremental fusion technique have played an increasingly important role in recent MSIF developments. In contrast, research of PCIF models with rough sets has slowed;

(4) Most existing MSIF models and technologies assume the label is available after classification, and very few research has been conducted to handle the multi-source data with unsupervised or semi-supervised method.

(5) Some techniques, such as parallel computing, incremental learning, have been applied in MSIF.

(6) There is no comprehensive analysis on real-world data streams from the MSIF models.

(7) An increasing number of other research areas have recognized the importance of dealing with multi-source data, particularly in big data community.

• **Future directions**

According to these findings and discussions, we suggest five new directions in future research of MSIF:

(1) MSIF research should not only focus on extension of the models, but also need to improve the basic theory of MSIF.

(2) In real practice, the cost to acquire true label could be expensive, namely, unsupervised or semi-supervised method of MSIF could still be promising in the future.

(3) A knowledge base or database for selecting real-world data of information fusion should be established for evaluating learning algorithm dealing with large-scale multi-source data.

(4) The MSIF models and techniques discussed in this paper can be extended, of which the parallel computing model and the three-way clustering fusion technique are worthy of further study.

(5) Research on effectively integrating MSIF handling techniques with different theories or methodologies for data-driven applications is highly desired, such as RST, fuzzy set theory, evidence theory, neural network, machine learning and so on. It may open up a glorious prospect for MSIF.

In this paper, we hope that we can provide researchers with state-of-the-art knowledge on MSIF research developments based on rough sets and provide guidelines about how to apply techniques of information fusion in different scenarios to support users in various data processing and decision-making.

CRediT authorship contribution statement

Pengfei Zhang: Conceptualization, Writing - original draft, Read and contributed to the manuscript. **Tianrui Li:** Supervision, Project administration, Read and contributed to the manuscript. **Guoqiang Wang:** Investigation, Read and contributed to the manuscript. **Chuan Luo:** Writing - review & editing, Read and contributed to the manuscript. **Hongmei Chen:** Visualization, Read and contributed to the manuscript. **Junbo Zhang:** Methodology, Read and contributed to the manuscript. **Dexian Wang:** Structure fabrication, Read and contributed to the manuscript. **Zeng Yu:** Structure fabrication, Read and contributed to the manuscript.

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.

Acknowledgments

The authors thank the anonymous reviewers and the editors for their valuable comments and suggestions in improving this paper.

This work was partially supported by the National Natural Science Foundation of China (Nos. 61573292, 62076171, 61976182), Applied Fundamental Research Program of Science & Technology Department of Sichuan Province (No. 2019YJ0084), and Fundamental Research Funds for the Central Universities (No. 2682020CX89).

References

[1] Q. Pan, Multi-Soure Information Fusion Theory and its Applications, Tsinghua university press, 2013.
[2] J. Llinas, E. Waltz, Multisensor Data Fusion, Artech Housse publisher, 1990.
[3] D. Hall, Mathematical Techniques in Multisenor Data Fusion, Artech Housse publisher, 1992.
[4] D. Hall, Handbook of Multisenor Data Fusion, CRC Press, 2001.
[5] X. Dong, F. Naumann, Data fusion - Resolving data conflicts for integration, 2 (2) (2009) 1654–1655.
[6] Y. Zhang, H. Zhang, N.M. Nasrabadi, T.S. Huang, Multi-metric learning for multi-sensor fusion based classification, Inf. Fusion 14 (4) (2013) 431–440.
[7] B.V. Dasarathy, Multi-sensor, multi-source information fusion: architecture, algorithms, and applications-a panoramic overview, in: Second IEEE International Conference on Computational Cybernetics, 2004. ICCC 2004., IEEE, 2004, p. 5.
[8] Y. Yang, Z. Jing, T. Gao, H. Wang, Multi-sources information fusion algorithm in airborne detection systems, J. Syst. Eng. Electron. 18 (1) (2007) 171–176.
[9] B. Cai, Y. Liu, Q. Fan, Y. Zhang, Z. Liu, S. Yu, R. Ji, Multi-source information fusion based fault diagnosis of ground-source heat pump using Bayesian network, Appl. Energy 114 (2014) 1–9.
[10] R. Gravina, P. Alinia, H. Ghasemzadeh, G. Fortino, Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges, Inf. Fusion 35 (2017) 68–80.
[11] R.R. Yager, A framework for multi-source data fusion, Inform. Sci. 163 (1–3) (2004) 175–200.
[12] R.R. Yager, Set measure directed multi-source information fusion, IEEE Trans. Fuzzy Syst. 19 (6) (2011) 1031–1039.
[13] T. Li, F. De la Prieta Pintado, J.M. Corchado, J. Bajo, Multi-source homogeneous data clustering for multi-target detection from cluttered background with misdetection, Appl. Soft Comput. 60 (2017) 436–446.
[14] 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.
[15] Z. Pawlak, Rough sets, Int. J. Comput. Inf. Sci. 11 (5) (1982) 341–356.
[16] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, Springer Netherlands, 1991.
[17] Z. Pawlak, R. Soŵinski, Rough set approach to multi-attribute decision analysis, European J. Oper. Res. 72 (3) 443–459.
[18] Z. Pawlak, Drawing conclusions from datałthe rough set way, Int. J. Intell. Syst. 16 (1) (2001) 3–11.
[19] J. Han, Y. Cai, N. Cercone, Data-driven discovery of quantitative rules in relational databases, IEEE Trans. Knowl. Data Eng. 5 (1) (1993) 29–40.
[20] C.-C. Chan, A rough set approach to attribute generalization in data mining, Inform. Sci. 107 (1–4) (1998) 169–176.
[21] T.-F. Fan, D.-R. Liu, G.-H. Tzeng, Rough set-based logics for multicriteria decision analysis, European J. Oper. Res. 182 (1) (2007) 340–355.
[22] M.A. Khan, M. Banerjee, Formal reasoning with rough sets in multiple-source approximation systems, Internat. J. Approx. Reason. 49 (2) (2008) 466–477.
[23] T. Li, M. Fei, Information fusion in wireless sensor network based on rough set, in: 2009 IEEE International Conference on Network Infrastructure and Digital Content, IEEE, 2009, pp. 129–134.
[24] Y. Liu, Y. Chen, X. Yuan, Study of information fusion evaluation methodology based on multiple approximation space, in: 2010 2nd International Symposium on Information Engineering and Electronic Commerce, IEEE, 2010, pp. 1–4.
[25] T. Li, D. Ruan, Y. Shen, E. Hermans, G. Wets, A new weighting approach based on rough set theory and granular computing for road safety indicator analysis, Comput. Intell. 32 (4) (2016) 517–534.
[26] Y. Yao, Z. Wang, C. Gan, Q. Kang, X. Liu, Y. Xia, L. Zhang, Multi-source alert data understanding for security semantic discovery based on rough set theory, Neurocomputing 208 (2016) 39–45.
[27] W. Xu, J. Yu, A novel approach to information fusion in multi-source datasets: a granular computing viewpoint, Inf. Sci. 378 (2017) 410–423.
[28] 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.
[29] X. Che, J. Mi, D. Chen, Information fusion and numerical characterization of a multi-source information system, Knowl.-Based Syst. 145 (2018) 121–133.
[30] L. Yang, W. Xu, X. Zhang, B. Sang, Multi-granulation method for information fusion in multi-source decision information system, Int. J. Approx. Reason. (2020) http://dx.doi.org/10.1016/j.ijar.2020.04.003.
[31] B. Sang, Y. Guo, D. Shi, W. Xu, Decision-theoretic rough set model of multi-source decision systems, Int. J. Mach. Learn. Cybern. 9 (11) (2018) 1941–1954.
[32] 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.
[33] W. Wei, J. Liang, Information fusion in rough set theory: An overview, Inf. Fusion 48 (2019) 107–118.
[34] C. Han, H. Zhu, Z. Duan, Multi-Soure Information Fusion (Second Edition), Tsinghua university press, 2010.
[35] J.W. Grzymala-Busse, LERS-a system for learning from examples based on rough sets, in: Intelligent Decision Support, Springer, 1992, pp. 3–18.
[36] A. An, N. Shan, C. Chan, N. Cercone, W. Ziarko, Discovering rules for water demand prediction: an enhanced rough-set approach, Eng. Appl. Artif. Intell. 9 (6) (1996) 645–653.
[37] R. Słowiński, J. Stefanowski, Roughdas and roughclass software implementations of the rough sets approach, in: Intelligent Decision Support, Springer, 1992, pp. 445–456.
[38] D. Ohrn, Rosetta Technical Reference Manual, Knowledge Systems Group, Dept of Computer and Information Science, Norwegian University of Science and Technology, 2000.
[39] R.H. Golan, W. Ziarko, A methodology for stock market analysis utilizing rough set theory, in: Proceedings of 1995 Conference on Computational Intelligence for Financial Engineering (CIFEr), IEEE, 1995, pp. 32–40.
[40] S. Wengang, W. Rixin, H. Wenhu, Application of rough set theory to fault diagnosis of check valves in reciprocating pumps, in: Canadian Conference on Electrical and Computer Engineering 2001. Conference Proceedings (Cat. No. 01TH8555), Vol. 2, IEEE, 2001, pp. 1247–1250.
[41] R. Slowinski, C. Zopounidis, Application of the rough set approach to evaluation of bankruptcy risk, Intell. Syst. Account. Finance Manage. 4 (1) (1995) 27–41.
[42] F. Witlox, H. Tindemans, The application of rough sets analysis in activity-based modelling. opportunities and constraints, Expert Syst. Appl. 27 (4) (2004) 585–592.
[43] J. Czerniak, H. Zarzycki, Application of rough sets in the presumptive diagnosis of urinary system diseases, in: Artificial Intelligence and Security in Computing Systems, Springer, 2003, pp. 41–51.

[44] Z. Li, P. Zhang, N. Xie, G. Zhang, C.-F. Wen, A novel three-way decision method in a hybrid information system with images and its application in medical diagnosis, Eng. Appl. Artif. Intell. 92 (2020) 103651.
[45] P. Wang, P. Zhang, Z. Li, A three-way decision method based on Gaussian kernel in a hybrid information system with images: An application in medical diagnosis, Appl. Soft Comput. 77 (2019) 734–749.
[46] Y. Yao, Information granulation and rough set approximation, Int. J. Intell. Syst. 16 (1) (2001) 87–104.
[47] W. Zhang, W. Wu, J. Liang, D. Li, Rough Set Theory and Methods, Chinese Scientific Publishers, Beijing, 2001.
[48] C. Dobre, F. Xhafa, Intelligent services for big data science, Future Gener. Comput. Syst. 37 (2014) 267–281.
[49] J. Camacho, Visualizing big data with compressed score plots: approach and research challenges, Chemometr. Intell. Lab. Syst. 135 (2014) 110–125.
[50] T. Li, C. Luo, H. Chen, J. Zhang, PICKT: a solution for big data analysis, in: International Conference on Rough Sets and Knowledge Technology, Springer, 2015, pp. 15–25.
[51] Y.-Y. Guan, H.-K. Wang, Set-valued information systems, Inform. Sci. 176 (17) (2006) 2507–2525.
[52] Y. Qian, C. Dang, J. Liang, D. Tang, Set-valued ordered information systems, Inform. Sci. 179 (16) (2009) 2809–2832.
[53] J. Dai, H. Tian, Entropy measures and granularity measures for set-valued information systems, Inform. Sci. 240 (2013) 72–82.
[54] J. He, P. Wang, Z. Li, Uncertainty measurement for a set-valued information system: Gaussian kernel method, Symmetry 11 (2) (2019).
[55] J. Zhang, T. Li, D. Ruan, D. Liu, Rough sets based matrix approaches with dynamic attribute variation in set-valued information systems, Internat. J. Approx. Reason. 53 (4) (2012) 620–635.
[56] H. Chen, T. Li, H. Tian, Approaches for updating approximations in set-valued information systems while objects and attributes vary with time, in: Rough Sets and Intelligent Systems-Professor Zdzisław Pawlak in Memoriam, Springer, 2013, pp. 229–248.
[57] C. Luo, T. Li, H. Chen, Dynamic maintenance of approximations in set-valued ordered decision systems under the attribute generalization, Inform. Sci. 257 (2014) 210–228.
[58] C. Luo, T. Li, H. Chen, L. Lu, Fast algorithms for computing rough approximations in set-valued decision systems while updating criteria values, Inform. Sci. 299 (2015) 221–242.
[59] J. Dai, H. Tian, Fuzzy rough set model for set-valued data, Fuzzy Sets and Systems 229 (2013) 54–68.
[60] Z. Bao, S. Yang, Attribute reduction for set-valued ordered fuzzy decision system, in: 2014 Sixth International Conference on Intelligent Human-Machine Systems and Cybernetics, Vol. 2, IEEE, 2014, pp. 96–99.
[61] W. Wei, J. Cui, J. Liang, J. Wang, Fuzzy rough approximations for set-valued data, Inform. Sci. 360 (2016) 181–201.
[62] H.-Y. Zhang, S.-Y. Yang, Feature selection and approximate reasoning of large-scale set-valued decision tables based on α-dominance-based quantitative rough sets, Inform. Sci. 378 (2017) 328–347.
[63] Y. Huang, T. Li, C. Luo, H. Fujita, S.-j. Horng, Dynamic variable precision rough set approach for probabilistic set-valued information systems, Knowl.-Based Syst. 122 (2017) 131–147.
[64] X. Xie, Z. Li, P. Zhang, G. Zhang, Information structures and uncertainty measures in an incomplete probabilistic set-valued information system, IEEE Access 7 (2019) 27501–27514.
[65] X.R. Zhao, B.Q. Hu, Three-way decisions with decision-theoretic rough sets in multiset-valued information tables, Inform. Sci. 507 (2020) 684–699.
[66] J. Zhang, T. Li, H. Chen, Composite rough sets, in: International Conference on Artificial Intelligence and Computational Intelligence, Springer, 2012, pp. 150–159.
[67] J. Zhang, T. Li, H. Chen, Composite rough sets for dynamic data mining, Inform. Sci. 257 (2014) 81–100.
[68] A. Zeng, T. Li, D. Liu, J. Zhang, H. Chen, A fuzzy rough set approach for incremental feature selection on hybrid information systems, Fuzzy Sets and Systems 258 (2015) 39–60.
[69] A. Zeng, T. Li, J. Hu, H. Chen, C. Luo, Dynamical updating fuzzy rough approximations for hybrid data under the variation of attribute values, Inform. Sci. 378 (2017) 363–388.
[70] Y. Huang, T. Li, C. Luo, H. Fujita, S.-j. Horng, B. Wang, Dynamic maintenance of rough approximations in multi-source hybrid information systems, Inf. Sci. (2020) http://dx.doi.org/10.1016/j.ins.2020.03.097.
[71] G. Yu, Information structures and uncertainty measures in a hybrid information system with images, Soft Comput. 23 (24) (2019) 12961–12979.
[72] J. Zeng, Z. Li, P. Zhang, P. Wang, Information structures and uncertainty measures in a hybrid information system: Gaussian kernel method, Int. J. Fuzzy Syst. 22 (1) (2020) 212–231.
[73] Z. Li, N. Xie, D. Huang, G. Zhang, A three-way decision method in a hybrid decision information system and its application in medical diagnosis, Artif. Intell. Rev. (2020) 1–30.
[74] M.A. Khan, M. Banerjee, Multiple-source approximation systems: membership functions and indiscernibility, in: International Conference on Rough Sets and Knowledge Technology, Springer, 2008, pp. 80–87.
[75] G. Lin, J. Liang, Y. Qian, An information fusion approach by combining multigranulation rough sets and evidence theory, Inform. Sci. 314 (2015) 184–199.
[76] B. Sun, W. Ma, X. Chen, X. Li, Heterogeneous multigranulation fuzzy rough set-based multiple attribute group decision making with heterogeneous preference information, Comput. Ind. Eng. 122 (2018) 24–38.
[77] L. Zhang, T. Wang, H. Li, B. Huang, X. Zhou, Agent evaluation based on multi-source heterogeneous information table using TOPSIS, Adv. Eng. Inform. 42 (2019) 100971.
[78] R. Ma, W. Liu, Multi-granulation rough set model based on set-valued information system, J. Syst. Eng. Electron. 36 (5) (2014) 921–925.
[79] Y. Zhuang, W. Liu, M. Fan, J. Li, Multi-granulation dominance Łelation and information fusion based on set-valued information system, PR&AI 28 (8) (2015) 742–749.
[80] F. Cao, J.Z. Huang, J. Liang, X. Zhao, Y. Meng, K. Feng, Y. Qian, An algorithm for clustering categorical data with set-valued features, IEEE Trans. Neural Netw. Learn. Syst. 29 (10) (2017) 4593–4606.
[81] Y. Yao, N. Noroozi, A unified model for set-based computations, in: Soft Computing: 3rd International Workshop on Rough Sets and Soft Computing, Citeseer, 1994, pp. 252–255.
[82] W. Zou, T. Li, H. Chen, X. Ji, Approaches for incrementally updating approximations based on set-valued information systems while attribute values' coarsening and refining, in: 2009 IEEE International Conference on Granular Computing, IEEE, 2009, pp. 824–829.
[83] W.-X. Zhang, J.-S. Mi, Incomplete information system andits optimal selections, Comput. Math. Appl. 48 (5–6) (2004) 691–698.
[84] W.W. Cohen, Learning trees and rules with set-valued features, in: AAAI/IAAI, Vol. 1, 1996, pp. 709–716.
[85] G. Lang, Q. Li, T. Yang, An incremental approach to attribute reduction of dynamic set-valued information systems, Int. J. Mach. Learn. Cybern. 5 (5) (2014) 775–788.
[86] Y. Zhuang, W. Liu, M. Fan, J. Li, Multi-granulation dominance relation and information fusion based on set-valued information system, Pattern Recognit. Artif. Intell. (8) (2015) 10.
[87] D. Barbara, H. Garcia-Molina, D. Porter, A probabilistic relational data model, 1990.
[88] D. Barbará, H. Garcia-Molina, D. Porter, The management of probabilistic data, IEEE Trans. Knowl. Data Eng. 4 (5) (1992) 487–502.
[89] G. Hurlbert, T. Johnson, J. Zahl, On universal cycles for multisets, Discrete Math. 309 (17) (2009) 5321–5327.
[90] S. Eliahou, M.J. Erickson, Mutually describing multisets and integer partitions, Discrete Math. 313 (4) (2013) 422–433.
[91] Y.-H. Yuan, Q.-S. Sun, Q. Zhou, D.-S. Xia, A novel multiset integrated canonical correlation analysis framework and its application in feature fusion, Pattern Recognit. 44 (5) (2011) 1031–1040.
[92] R.R. Yager, On the theory of bags, Int. J. Gen. Syst. 13 (1) (1986) 23–37.
[93] W.D. Blizard, Multiset theory., Notre Dame J. Form. Log. 30 (1) (1989) 36–66.
[94] D. Singh, A. Ibrahim, T. Yohanna, J. Singh, An overview of the applications of multisets, Novi Sad J. Math. 37 (3) (2007) 73–92.
[95] A. Syropoulos, Mathematics of multisets, in: Workshop on Membrane Computing, Springer, 2000, pp. 347–358.
[96] S. Miyamoto, Generalizations of multisets and rough approximations, Int. J. Intell. Syst. 19 (7) (2004) 639–652.
[97] D. Singh, A. Ibrahim, T. Yohana, J. Singh, Complementation in multiset theory, in: International Mathematical Forum, Vol. 6, 2011, pp. 1877–1884.
[98] K. Girish, S.J. John, Multiset topologies induced by multiset relations, Inform. Sci. 188 (2012) 298–313.
[99] Á. Riesgo, P. Alonso, I. Díaz, S. Montes, Basic operations for fuzzy multisets, Internat. J. Approx. Reason. 101 (2018) 107–118.
[100] L. Zhang, Y. Xie, L. Xidao, X. Zhang, Multi-source heterogeneous data fusion, in: 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD), IEEE, 2018, pp. 47–51.
[101] Z. Duan, T. Wu, S. Guo, T. Shao, R. Malekian, Z. Li, Development and trend of condition monitoring and fault diagnosis of multi-sensors information fusion for rolling bearings: a review, Int. J. Adv. Manuf. Technol. 96 (1–4) (2018) 803–819.
[102] D.R. Wilson, T.R. Martinez, Improved heterogeneous distance functions, J. Artif. Intell. Res. 6 (1997) 1–34.
[103] H. Wang, Nearest neighbors by neighborhood counting, IEEE Trans. Pattern Anal. Mach. Intell. 28 (6) (2006) 942–953.
[104] Q. Hu, D. Yu, Z. Xie, Information-preserving hybrid data reduction based on fuzzy-rough techniques, Pattern Recognit. Lett. 27 (5) (2006) 414–423.
[105] Q. Hu, Z. Xie, D. Yu, Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation, Pattern Recognit. 40 (12) (2007) 3509–3521.
[106] Q. Hu, J. Liu, D. Yu, Mixed feature selection based on granulation and approximation, Knowl.-Based Syst. 21 (4) (2008) 294–304.
[107] Q. Hu, D. Yu, J. Liu, C. Wu, Neighborhood rough set based heterogeneous feature subset selection, Inf. Sci. 178 (18) (2008) 3577–3594.

[108] H. Zhao, K. Qin, Mixed feature selection in incomplete decision table, Knowl.-Based Syst. 57 (2014) 181–190.
[109] B. Sun, W. Ma, D. Chen, Rough approximation of a fuzzy concept on a hybrid attribute information system and its uncertainty measure, Inform. Sci. 284 (2014) 60–80.
[110] Y. Yang, S. Song, D. Chen, X. Zhang, Discernible neighborhood counting based incremental feature selection for heterogeneous data, Int. J. Mach. Learn. Cybern. (2019) 1–13.
[111] D. Chen, Y. Yang, Attribute reduction for heterogeneous data based on the combination of classical and fuzzy rough set models, IEEE Trans. Fuzzy Syst. 22 (5) (2013) 1325–1334.
[112] J. Zhang, Y. Zhu, Y. Pan, T. Li, Efficient parallel boolean matrix based algorithms for computing composite rough set approximations, Inform. Sci. 329 (2016) 287–302.
[113] X. Zhang, C. Mei, D. Chen, J. Li, Feature selection in mixed data: A method using a novel fuzzy rough set-based information entropy, Pattern Recognit. 56 (2016) 1–15.
[114] J. Yu, W. Xu, Information fusion for multi-source fuzzy information system with the same structure, in: 2015 International Conference on Machine Learning and Cybernetics (ICMLC), Vol. 1, IEEE, 2015, pp. 170–175.
[115] M. Li, X. Zhang, Information fusion in a multi-source incomplete information system based on information entropy, Entropy 19 (11) (2017) 570.
[116] S. Liu, L. Peng, T. Chi, X. Wang, Research on multi-source heterogeneous data collection for the smart city public information platform, in: 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), IEEE, 2016, pp. 623–626.
[117] Y. Liu, Multi-source heterogeneous data fusion based on perceptual semantics in narrow-band internet of things, Pers. Ubiquitous Comput. 23 (3–4) (2019) 413–420.
[118] J. Liu, T. Li, P. Xie, S. Du, F. Teng, X. Yang, Urban big data fusion based on deep learning: An overview, Inf. Fusion 53 (2020) 123–133.
[119] H. Aliakbarpour, V.S. Prasath, K. Palaniappan, G. Seetharaman, J. Dias, Heterogeneous multi-view information fusion: Review of 3-D reconstruction methods and a new registration with uncertainty modeling, IEEE Access 4 (2016) 8264–8285.
[120] G. Hu, J. Shao, F. Liu, Y. Wang, H.T. Shen, If-matching: Towards accurate map-matching with information fusion, IEEE Trans. Knowl. Data Eng. 29 (1) (2016) 114–127.
[121] H.-C. Huang, An evolutionary optimal fuzzy system with information fusion of heterogeneous distributed computing and polar-space dynamic model for online motion control of swedish redundant robots, IEEE Trans. Ind. Electron. 64 (2) (2016) 1743–1750.
[122] Y. Li, D.K. Jha, A. Ray, T.A. Wettergren, Information fusion of passive sensors for detection of moving targets in dynamic environments, IEEE Trans. Cybern. 47 (1) (2016) 93–104.
[123] L. Piras, G. Giacinto, Information fusion in content based image retrieval: A comprehensive overview, Inf. Fusion 37 (2017) 50–60.
[124] G. Quellec, M. Lamard, G. Cazuguel, C. Roux, B. Cochener, Case retrieval in medical databases by fusing heterogeneous information, IEEE Trans. Med. Imaging 30 (1) (2010) 108–118.
[125] Y.-R. Yeh, T.-C. Lin, Y.-Y. Chung, Y.-C.F. Wang, A novel multiple kernel learning framework for heterogeneous feature fusion and variable selection, IEEE Trans. Multimed. 14 (3) (2012) 563–574.
[126] S. Luo, Three-way decision in a multi-source information system and its applications, IEEE Access 7 (2019) 108343–108359.
[127] J. Zhang, Y. Zhu, Y. Pan, T. Li, A parallel implementation of computing composite rough set approximations on GPUs, in: International Conference on Rough Sets and Knowledge Technology, Springer, 2013, pp. 240–250.
[128] J. Yuan, M. Chen, T. Jiang, T. Li, Complete tolerance relation based parallel filling for incomplete energy big data, Knowl.-Based Syst. 132 (2017) 215–225.
[129] H. Asfoor, R. Srinivasan, G. Vasudevan, N. Verbiest, C. Cornells, M. Tolentino, A. Teredesai, M. De Cock, Computing fuzzy rough approximations in large scale information systems, in: 2014 IEEE International Conference on Big Data (Big Data), IEEE, 2014, pp. 9–16.
[130] B. Sun, X. Zhou, N. Lin, Diversified binary relation-based fuzzy multigranulation rough set over two universes and application to multiple attribute group decision making, Inf. Fusion 55 (2020) 91–104.
[131] H. Yu, X. Wang, G. Wang, X. Zeng, An active three-way clustering method via low-rank matrices for multi-view data, Inform. Sci. 507 (2020) 823–839.
[132] L.A. Zadeh, Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems, Soft Comput. 2 (1) 23–25.
[133] H.S. Nguyen, A. Skowron, J. Stepaniuk, Granular computing: A rough set approach, Comput. Intell. 17 (3) (2001) 514–544.
[134] T.Y. Lin, Granular computing: examples, intuitions and modeling, in: 2005 IEEE International Conference on Granular Computing, Vol. 1, IEEE, 2005, pp. 40–44.
[135] Y. Chen, Y. Yao, Multiview intelligent data analysis based on granular computing., in: GrC, 2006, pp. 281–286.
[136] A. Skowron, J. Stepaniuk, R. Swiniarski, Approximation spaces in rough–granular computing, Fund. Inform. 100 (1–4) (2010) 141–157.
[137] W. Pedrycz, S.-M. Chen, Granular Computing and Intelligent Systems: Design with Information Granules of Higher Order and Higher Type, Vol. 13, Springer Science & Business Media, 2011.
[138] J.T. Yao, A.V. Vasilakos, W. Pedrycz, Granular computing: perspectives and challenges, IEEE Trans. Cybern. 43 (6) (2013) 1977–1989.
[139] Y. Yao, Three-way decision and granular computing, Internat. J. Approx. Reason. 103 (2018) 107–123.
[140] F. Li, Y. Qian, J. Wang, J. Liang, Multigranulation information fusion: A Dempster-Shafer evidence theory-based clustering ensemble method, Inf. Sci. 378 (2017) 389–409.
[141] H. Ju, H. Li, X. Yang, X. Zhou, B. Huang, Cost-sensitive rough set: a multi-granulation approach, Knowl.-Based Syst. 123 (2017) 137–153.
[142] A. Tan, W.-Z. Wu, J. Li, T. Li, Reduction foundation with multigranulation rough sets using discernibility, Artif. Intell. Rev. (2019) 1–28.
[143] L. Sun, L. Wang, Y. Qian, J. Xu, Feature selection using fuzzy neighborhood entropy-based uncertainty measures for fuzzy neighborhood multigranulation rough sets, IEEE Trans. Fuzzy Syst. (2020).
[144] X. Yang, Y. Qi, H. Yu, X. Song, J. Yang, Updating multigranulation rough approximations with increasing of granular structures, Knowl.-Based Syst. 64 (2014) 59–69.
[145] C. Hu, L. Zhang, B. Wang, Z. Zhang, F. Li, Incremental updating knowledge in neighborhood multigranulation rough sets under dynamic granular structures, Knowl.-Based Syst. 163 (2019) 811–829.
[146] S.S. Kumar, H.H. Inbarani, Optimistic multi-granulation rough set based classification for medical diagnosis, Procedia Comput. Sci. 47 (2015) 374–382.
[147] Y. Qian, J. Liang, Y. Yao, C. Dang, MGRS: A multi-granulation rough set, Inf. Sci. 180 (6) (2010) 949–970.
[148] Y. Qian, H. Zhang, Y. Sang, J. Liang, Multigranulation decision-theoretic rough sets., Internat. J. Approx. Reason. 55 (1) (2014) 225–237.
[149] W. Xu, W. Li, X. Zhang, Generalized multigranulation rough sets and optimal granularity selection, Granul. Comput. 2 (4) (2017) 271–288.
[150] C. Liu, D. Miao, J. Qian, On multi-granulation covering rough sets, Int. J. Approx. Reason. 55 (6) (2014) 1404–1418.
[151] C. Liu, W. Pedrycz, Covering-based multi-granulation fuzzy rough sets, J. Intell. Fuzzy Syst. 30 (1) (2016) 303–318.
[152] W.-Z. Wu, Y. Leung, Theory and applications of granular labelled partitions in multi-scale decision tables, Inform. Sci. 181 (18) (2011) 3878–3897.
[153] W.-Z. Wu, Y. Leung, Optimal scale selection for multi-scale decision tables, Internat. J. Approx. Reason. 54 (8) (2013) 1107–1129.
[154] S.-M. Gu, Y.-H. Wan, W.-Z. Wu, T.-J. Li, Rough set approximations in multi-scale interval information systems, in: Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, Springer, 2015, pp. 73–81.
[155] C. Hao, J. Li, M. Fan, W. Liu, E.C. Tsang, Optimal scale selection in dynamic multi-scale decision tables based on sequential three-way decisions, Inform. Sci. 415 (2017) 213–232.
[156] C. Luo, T. Li, Y. Huang, H. Fujita, Updating three-way decisions in incomplete multi-scale information systems, Inform. Sci. 476 (2019) 274–289.
[157] X. Zhou, H. Li, A multi-view decision model based on decision-theoretic rough set, in: International Conference on Rough Sets and Knowledge Technology, Springer, 2009, pp. 650–657.
[158] D. Liu, D. Liang, C. Wang, A novel three-way decision model based on incomplete information system, Knowl.-Based Syst. 91 (2016) 32–45.
[159] B. Sun, W. Ma, Y. Qian, Multigranulation fuzzy rough set over two universes and its application to decision making, Knowl.-Based Syst. 123 (2017) 61–74.
[160] C. Zhang, D. Li, Y. Zhai, Y. Yang, Multigranulation rough set model in hesitant fuzzy information systems and its application in person-job fit, Int. J. Mach. Learn. Cybern. 10 (4) (2019) 717–729.
[161] P. Mandal, A. Ranadive, Multi-granulation pythagorean fuzzy decision-theoretic rough sets based on inclusion measure and their application in incomplete multi-source information systems, Complex Intell. Syst. 5 (2) (2019) 145–163.
[162] Y. Qian, S. Li, J. Liang, Z. Shi, F. Wang, Pessimistic rough set based decisions: a multigranulation fusion strategy, Inform. Sci. 264 (2014) 196–210.
[163] C. Zhang, D. Li, Y. Mu, D. Song, An interval-valued hesitant fuzzy multigranulation rough set over two universes model for steam turbine fault diagnosis, Appl. Math. Model. 42 (2017) 693–704.
[164] W. Xu, W. Sun, X. Zhang, W. Zhang, Multiple granulation rough set approach to ordered information systems, Int. J. Gen. Syst. 41 (5) (2012) 475–501.
[165] Y. Xu, Multigranulation rough set model based on granulation of attributes and granulation of attribute values, Inform. Sci. 484 (2019) 1–13.
[166] Y. Yao, Y. She, Rough set models in multigranulation spaces, Inform. Sci. 327 (2016) 40–56.
[167] Q. Zhang, Q. Zhang, G. Wang, The uncertainty of probabilistic rough sets in multi-granulation spaces, Int. J. Approx. Reason. 77 (2016) 38–54.
[168] A. Tan, W.-Z. Wu, Y. Tao, On the belief structures and reductions of multigranulation spaces with decisions, Internat. J. Approx. Reason. 88 (2017) 39–52.
[169] F. Li, B.Q. Hu, J. Wang, Stepwise optimal scale selection for multi-scale decision tables via attribute significance, Knowl.-Based Syst. 129 (2017) 4–16.
[170] W.-Z. Wu, Y. Leung, A comparison study of optimal scale combination selection in generalized multi-scale decision tables, Int. J. Mach. Learn. Cybern. (2019) 1–12.

[171] X. Zhang, Q. Zhang, Y. Cheng, G. Wang, Optimal scale selection by integrating uncertainty and cost-sensitive learning in multi-scale decision tables, Int. J. Mach. Learn. Cybern. (2020) 1–20.
[172] W.-Z. Wu, Y. Qian, T.-J. Li, S.-M. Gu, On rule acquisition in incomplete multi-scale decision tables, Inf. Sci. 378 (2017) 282–302.
[173] Y. She, J. Li, H. Yang, A local approach to rule induction in multi-scale decision tables, Knowl.-Based Syst. 89 (2015) 398–410.
[174] S.-M. Gu, W.-Z. Wu, On knowledge acquisition in multi-scale decision systems, Int. J. Mach. Learn. Cybern. 4 (5) (2013) 477–486.
[175] Y. Cheng, Q. Zhang, G. Wang, B.Q. Hu, Optimal scale selection and attribute reduction in multi-scale decision tables based on three-way decision, Inform. Sci. (2020).
[176] Y. She, X. He, T. Qian, Q. Wang, W. Zeng, A theoretical study on object-oriented and property-oriented multi-scale formal concept analysis, Int. J. Mach. Learn. Cybern. 10 (11) (2019) 3263–3271.
[177] Z. Huang, J. Li, W. Dai, R. Lin, Generalized multi-scale decision tables with multi-scale decision attributes, Internat. J. Approx. Reason. 115 (2019) 194–208.
[178] Y.S. Chen, J.J. Li, J.X. Huang, Matrix method for the optimal scale selection of multi-scale information decision systems, Mathematics 7 (3) (2019) 290.
[179] G. Zhang, Z. Li, W.-Z. Wu, X. Liu, N. Xie, Information structures and uncertainty measures in a fully fuzzy information system, Internat. J. Approx. Reason. 101 (2018) 119–149.
[180] N. Xie, Z. Li, W.-Z. Wu, G. Zhang, Fuzzy information granular structures: A further investigation, Internat. J. Approx. Reason. 114 (2019) 127–150.
[181] N. Chen, J. He, Information structures in a set-valued information system based on granular computing, J. Intell. Fuzzy Syst. (Preprint) (2019) 1–13.
[182] Z. Li, P. Zhang, X. Ge, N. Xie, G. Zhang, C.-F. Wen, Uncertainty measurement for a fuzzy relation information system, IEEE Trans. Fuzzy Syst. 27 (12) (2019) 2338–2352.
[183] Z. Li, D. Huang, X. Liu, N. Xie, G. Zhang, Information structures in a covering information system, Inform. Sci. 507 (2020) 449–471.
[184] Y. Yao, Y. Zhao, Attribute reduction in decision-theoretic rough set models, Inf. Sci. 178 (17) (2008) 3356–3373.
[185] Y. Yao, Three-way decisions with probabilistic rough sets, Inf. Sci. 180 (3) (2010) 341–353.
[186] Y. Yao, Three-way decisions and cognitive computing, Cogn. Comput. 8 (4) (2016) 543–554.
[187] A. Bargiela, W. Pedrycz, Toward a theory of granular computing for human-centered information processing, IEEE Trans. Fuzzy Syst. 16 (2) (2008) 320–330.
[188] Y. Yang, J. Hu, Y. Liu, X. Chen, A multiperiod hybrid decision support model for medical diagnosis and treatment based on similarities and three-way decision theory, Expert Syst. 36 (3) (2019) e12377.
[189] J. Hu, Y. Yang, X. Chen, A novel TODIM method-based three-way decision model for medical treatment selection, Int. J. Fuzzy Syst. 20 (4) (2018) 1240–1255.
[190] Z. Li, D. Huang, A three-way decision method in a fuzzy condition decision information system and its application in credit card evaluation, Granul. Comput. (2019) 1–14.
[191] H. Sakai, M. Nakata, J. Watada, NIS-apriori-based rule generation with three-way decisions and its application system in SQL, Inform. Sci. 507 (2020) 755–771.
[192] H. Li, L. Zhang, B. Huang, X. Zhou, Sequential three-way decision and granulation for cost-sensitive face recognition, Knowl.-Based Syst. 91 (2016) 241–251.
[193] D. Liang, M. Wang, Z. Xu, X. Chen, Risk interval-valued three-way decisions model with regret theory and its application to project resource allocation, J. Oper. Res. Soc. (2019) 1–20.
[194] M. Nauman, N. Azam, J. Yao, A three-way decision making approach to malware analysis using probabilistic rough sets, Inform. Sci. 374 (2016) 193–209.
[195] B. Barney, et al., Introduction to parallel computing, Lawrence Livermore Natl. Lab. 6 (13) (2010) 10.
[196] A. Grama, V. Kumar, A. Gupta, G. Karypis, Introduction to Parallel Computing, Pearson Education, 2003.
[197] R. Susmaga, Parallel computation of reducts, in: International Conference on Rough Sets and Current Trends in Computing, Springer, 1998, pp. 450–458.
[198] J. Zhang, T. Li, D. Ruan, Z. Gao, C. Zhao, A parallel method for computing rough set approximations, Inform. Sci. 194 (2012) 209–223.
[199] T. White, Hadoop: The Definitive Guide, O'Reilly Media, Inc., 2012.
[200] J. Dean, S. Ghemawat, Mapreduce: simplified data processing on large clusters, Commun. ACM 51 (1) (2008) 107–113.
[201] C. Ranger, R. Raghuraman, A. Penmetsa, G. Bradski, C. Kozyrakis, Evaluating mapreduce for multi-core and multiprocessor systems, in: 2007 IEEE 13th International Symposium on High Performance Computer Architecture, Ieee, 2007, pp. 13–24.
[202] R. Lämmel, Googles mapreduce programming model revisited, Sci. Comput. Program. 70 (1) (2008) 1–30.
[203] J. Zhang, T. Li, Y. Pan, Plar: Parallel large-scale attribute reduction on cloud systems, in: 2013 International Conference on Parallel and Distributed Computing, Applications and Technologies, IEEE, 2013, pp. 184–191.
[204] J. Zhang, J.-S. Wong, Y. Pan, T. Li, A parallel matrix-based method for computing approximations in incomplete information systems, IEEE Trans. Knowl. Data Eng. 27 (2) (2014) 326–339.
[205] J. Zhang, J.-S. Wong, T. Li, Y. Pan, A comparison of parallel large-scale knowledge acquisition using rough set theory on different mapreduce runtime systems, Internat. J. Approx. Reason. 55 (3) (2014) 896–907.
[206] J. Zhang, T. Li, Y. Pan, Parallel large-scale attribute reduction on cloud systems, 2016, arXiv preprint arXiv:1610.01807.
[207] S. Li, T. Li, A parallel matrix-based approach for computing approximations in dominance-based rough sets approach, in: International Conference on Rough Sets and Knowledge Technology, Springer, 2014, pp. 173–183.
[208] S. Li, T. Li, Z. Zhang, H. Chen, J. Zhang, Parallel computing of approximations in dominance-based rough sets approach, Knowl.-Based Syst. 87 (2015) 102–111.
[209] H. Chen, T. Li, Y. Cai, C. Luo, H. Fujita, Parallel attribute reduction in dominance-based neighborhood rough set, Inform. Sci. 373 (2016) 351–368.
[210] J. Qian, D. Miao, Z. Zhang, X. Yue, Parallel attribute reduction algorithms using mapreduce, Inform. Sci. 279 (2014) 671–690.
[211] J. Meng, J. Zhang, Y.-S. Luan, X.-Y. He, L.-S. Li, Y.-F. Zhu, Parallel gene selection and dynamic ensemble pruning based on affinity propagation, Comput. Biol. Med. 87 (2017) 8–21.
[212] M.S. Raza, U. Qamar, A parallel rough set based dependency calculation method for efficient feature selection, Appl. Soft Comput. 71 (2018) 1020–1034.
[213] A. Muthusamy, D. Subramani, A parallel rough set theory for nonlinear dimension-reduction in big data analysis.
[214] S.-Y. Jing, G.-L. Li, K. Zeng, W. Pan, C.-M. Liu, Efficient parallel algorithm for computing rough set approximation on GPU, Soft Comput. 22 (22) (2018) 7553–7569.
[215] D. Liu, T. Li, J. Zhang, Incremental updating approximations in probabilistic rough sets under the variation of attributes, Knowl.-Based Syst. 73 (2015) 81–96.
[216] W.-C. Bang, Z. Bien, Incremental inductive learning algorithm in the framework of rough set theory and its application, Int. J. Fuzzy Syst. 1 (1) (1999) 25–36.
[217] T. Li, D. Ruan, W. Geert, J. Song, Y. Xu, A rough sets based characteristic relation approach for dynamic attribute generalization in data mining, Knowl.-Based Syst. 20 (5) (2007) 485–494.
[218] H. Chen, T. Li, D. Ruan, J. Lin, C. Hu, A rough-set-based incremental approach for updating approximations under dynamic maintenance environments, IEEE Trans. Knowl. Data Eng. 25 (2) (2013) 274–284.
[219] Y. Yao, Probabilistic approaches to rough sets, Expert Syst. 20 (5) (2003) 287–297.
[220] J. Liang, Z. Shi, The information entropy, rough entropy and knowledge granulation in rough set theory, Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 12 (01) (2004) 37–46.
[221] J. Liang, D. Li, Uncertainty and Knowledge Acquisition in Information Systems, Beijing Science Press, 2005.
[222] H. Chen, T. Li, C. Luo, S.-J. Horng, G. Wang, A rough set-based method for updating decision rules on attribute values' coarsening and refining, IEEE Trans. Knowl. Data Eng. 26 (12) (2014) 2886–2899.
[223] H. Chen, T. Li, D. Ruan, J. Lin, C. Hu, A rough-set-based incremental approach for updating approximations under dynamic maintenance environments, IEEE Trans. Knowl. Data Eng. 25 (2) (2011) 274–284.
[224] Y.-L. Chen, H.-W. Hu, K. Tang, Constructing a decision tree from data with hierarchical class labels, Expert Syst. Appl. 36 (3) (2009) 4838–4847.
[225] J. Błaszczyński, R. Słowiński, Incremental induction of decision rules from dominance-based rough approximations, Electron. Notes Theor. Comput. Sci. 82 (4) (2003) 40–51.
[226] C. Luo, T. Li, H. Chen, D. Liu, Incremental approaches for updating approximations in set-valued ordered information systems, Knowl.-Based Syst. 50 (2013) 218–233.
[227] S. Li, T. Li, D. Liu, Dynamic maintenance of approximations in dominance-based rough set approach under the variation of the object set, Int. J. Intell. Syst. 28 (8) (2013) 729–751.
[228] A. Zeng, T. Li, J. Zhang, H. Chen, Incremental maintenance of rough fuzzy set approximations under the variation of object set, Fund. Inform. 132 (3) (2014) 401–422.
[229] Y.-N. Fan, T.-L.B. Tseng, C.-C. Chern, C.-C. Huang, Rule induction based on an incremental rough set, Expert Syst. Appl. 36 (9) (2009) 11439–11450.
[230] C.-C. Huang, T.-L.B. Tseng, Y.-N. Fan, C.-H. Hsu, Alternative rule induction methods based on incremental object using rough set theory, Appl. Soft Comput. 13 (1) (2013) 372–389.
[231] J. Liang, F. Wang, C. Dang, Y. Qian, A group incremental approach to feature selection applying rough set technique, IEEE Trans. Knowl. Data Eng. 26 (2) (2012) 294–308.
[232] W. Shu, W. Qian, An incremental approach to attribute reduction from dynamic incomplete decision systems in rough set theory, Data Knowl. Eng. 100 (2015) 116–132.

[233] W. Shu, W. Qian, Y. Xie, Incremental approaches for feature selection from dynamic data with the variation of multiple objects, Knowl.-Based Syst. 163 (2019) 320–331.
[234] Y. Yang, D. Chen, H. Wang, Active sample selection based incremental algorithm for attribute reduction with rough sets, IEEE Trans. Fuzzy Syst. 25 (4) (2016) 825–838.
[235] C. Hu, L. Zhang, A dynamic framework for updating neighborhood multigranulation approximations with the variation of objects, Inform. Sci. 519 (2020) 382–406.
[236] W. Shu, H. Shen, Updating attribute reduction in incomplete decision systems with the variation of attribute set, Internat. J. Approx. Reason. 55 (3) (2014) 867–884.
[237] S. Li, T. Li, D. Liu, Incremental updating approximations in dominance-based rough sets approach under the variation of the attribute set, Knowl.-Based Syst. 40 (2013) 17–26.
[238] F. Wang, J. Liang, Y. Qian, Attribute reduction: A dimension incremental strategy, Knowl.-Based Syst. 39 (2013) 95–108.
[239] W. Shu, H. Shen, Incremental feature selection based on rough set in dynamic incomplete data, Pattern Recognit. 47 (12) (2014) 3890–3906.
[240] Y. Jing, T. Li, J. Huang, H. Chen, S.-J. Horng, A group incremental reduction algorithm with varying data values, Int. J. Intell. Syst. 32 (9) (2017) 900–925.
[241] W. Wei, X. Wu, J. Liang, J. Cui, Y. Sun, Discernibility matrix based incremental attribute reduction for dynamic data, Knowl.-Based Syst. 140 (2018) 142–157.
[242] M. Cai, G. Lang, H. Fujita, Z. Li, T. Yang, Incremental approaches to updating reducts under dynamic covering granularity, Knowl.-Based Syst. 172 (2019) 130–140.
[243] C. Luo, T. Li, Z. Yi, H. Fujita, Matrix approach to decision-theoretic rough sets for evolving data, Knowl.-Based Syst. 99 (2016) 123–134.
[244] K. Gong, Y. Wang, M. Xu, Z. Xiao, Bssreduce an $O(|U|)$ incremental feature selection approach for large-scale and high-dimensional data, IEEE Trans. Fuzzy Syst. 26 (6) (2018) 3356–3367.
[245] Y. Yang, D. Chen, H. Wang, X. Wang, Incremental perspective for feature selection based on fuzzy rough sets, IEEE Trans. Fuzzy Syst. 26 (3) (2017) 1257–1273.
[246] H. Chen, T. Li, C. Luo, S.-J. Horng, G. Wang, A decision-theoretic rough set approach for dynamic data mining, IEEE Trans. Fuzzy Syst. 23 (6) (2015) 1958–1970.
[247] 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.
[248] X. Yang, T. Li, D. Liu, H. Chen, C. Luo, A unified framework of dynamic three-way probabilistic rough sets, Inform. Sci. 420 (2017) 126–147.
[249] Y. Jing, T. Li, H. Fujita, B. Wang, N. Cheng, An incremental attribute reduction method for dynamic data mining, Inform. Sci. 465 (2018) 202–218.
[250] S. Wang, T. Li, C. Luo, H. Fujita, Efficient updating rough approximations with multi-dimensional variation of ordered data, Inform. Sci. 372 (2016) 690–708.
[251] S. Wang, T. Li, C. Luo, H. Chen, H. Fujita, Domain-wise approaches for updating approximations with multi-dimensional variation of ordered information systems, Inform. Sci. 478 (2019) 100–124.
[252] S. Wang, T. Li, C. Luo, J. Hu, H. Fujita, T. Huang, A novel approach for efficient updating approximations in dynamic ordered information systems, Inform. Sci. 507 (2020) 197–219.
[253] A.K. Jain, R.P.W. Duin, J. Mao, Statistical pattern recognition: A review, IEEE Trans. Pattern Anal. Mach. Intell. 22 (1) (2000) 4–37.
[254] A.K. Jain, Data clustering: 50 years beyond K-means, Pattern Recognit. Lett. 31 (8) (2010) 651–666.
[255] L. De Angelis, J.G. Dias, Mining categorical sequences from data using a hybrid clustering method, European J. Oper. Res. 234 (3) (2014) 720–730.
[256] G. Peters, F. Crespo, P. Lingras, R. Weber, Soft clustering–fuzzy and rough approaches and their extensions and derivatives, Internat. J. Approx. Reason. 54 (2) (2013) 307–322.
[257] T. Boongoen, N. Iam-On, Cluster ensembles: A survey of approaches with recent extensions and applications, Comp. Sci. Rev. 28 (2018) 1–25.
[258] A. Strehl, J. Ghosh, Cluster ensembles—a knowledge reuse framework for combining multiple partitions, J. Mach. Learn. Res. 3 (Dec) (2002) 583–617.
[259] M.C. Naldi, A. Carvalho, R.J. Campello, Cluster ensemble selection based on relative validity indexes, Data Min. Knowl. Discov. 27 (2) (2013) 259–289.
[260] W. Chimphlee, A.H. Abdullah, M.N.M. Sap, S. Srinoy, S. Chimphlee, Anomaly-based intrusion detection using fuzzy rough clustering, in: 2006 International Conference on Hybrid Information Technology, Vol. 1, IEEE, 2006, pp. 329–334.
[261] P. Lingras, G. Peters, Rough clustering, Wiley Interdiscip. Rev.: Data Min. Knowl. Discov. 1 (1) (2011) 64–72.
[262] G. Peters, R. Weber, R. Nowatzke, Dynamic rough clustering and its applications, Appl. Soft Comput. 12 (10) (2012) 3193–3207.
[263] B. Karlik, The positive effects of fuzzy c-means clustering on supervised learning classifiers, Int. J. Artif. Intell. Expert Syst. 7 (2016) 1–8.
[264] N. Chen, Z.-s. Xu, M.-m. Xia, Hierarchical hesitant fuzzy K-means clustering algorithm, Appl. Math. A 29 (1) (2014) 1–17.
[265] K. Faceli, A.C. De Carvalho, M.C. De Souto, Multi-objective clustering ensemble, Int. J. Hybrid Intell. Syst. 4 (3) (2007) 145–156.
[266] S. Mimaroglu, E. Erdil, An efficient and scalable family of algorithms for combining clusterings, Eng. Appl. Artif. Intell. 26 (10) (2013) 2525–2539.
[267] R. Avogadri, G. Valentini, Fuzzy ensemble clustering based on random projections for dna microarray data analysis, Artif. Intell. Med. 45 (2–3) (2009) 173–183.
[268] K. Faceli, M.C. de Souto, D.S. de Araújo, A.C. de Carvalho, Multi-objective clustering ensemble for gene expression data analysis, Neurocomputing 72 (13–15) (2009) 2763–2774.
[269] A. Azadeh, M. Saberi, M. Anvari, A. Azaron, M. Mohammadi, An adaptive network based fuzzy inference system–genetic algorithm clustering ensemble algorithm for performance assessment and improvement of conventional power plants, Expert Syst. Appl. 38 (3) (2011) 2224–2234.
[270] H. Alizadeh, B. Minaei-Bidgoli, H. Parvin, Cluster ensemble selection based on a new cluster stability measure, Intell. Data Anal. 18 (3) (2014) 389–408.
[271] S.T. Hadjitodorov, L.I. Kuncheva, L.P. Todorova, Moderate diversity for better cluster ensembles, Inf. Fusion 7 (3) (2006) 264–275.
[272] X.L. Xie, G. Beni, A validity measure for fuzzy clustering, IEEE Trans. Pattern Anal. Mach. Intell. 13 (8) (1991) 841–847.
[273] H. Parvin, B. Minaei-Bidgoli, A clustering ensemble framework based on selection of fuzzy weighted clusters in a locally adaptive clustering algorithm, Pattern Anal. Appl. 18 (1) (2015) 87–112.
[274] P. D'Urso, E.A. Maharaj, A.M. Alonso, Fuzzy clustering of time series using extremes, Fuzzy Sets and Systems 318 (2017) 56–79.
[275] Y. Jiang, F.-L. Chung, S. Wang, Z. Deng, J. Wang, P. Qian, Collaborative fuzzy clustering from multiple weighted views, IEEE Trans. Cybern. 45 (4) (2014) 688–701.
[276] P. Rathore, J.C. Bezdek, S.M. Erfani, S. Rajasegarar, M. Palaniswami, Ensemble fuzzy clustering using cumulative aggregation on random projections, IEEE Trans. Fuzzy Syst. 26 (3) (2017) 1510–1524.
[277] L. Bai, J. Liang, Y. Guo, An ensemble clusterer of multiple fuzzy k-means clusterings to recognize arbitrarily shaped clusters, IEEE Trans. Fuzzy Syst. 26 (6) (2018) 3524–3533.
[278] A. Bagherinia, B. Minaei-Bidgoli, M. Hossinzadeh, H. Parvin, Reliability-based fuzzy clustering ensemble, Fuzzy Sets and Systems (2020).
[279] P. Lingras, C. West, Interval set clustering of web users with rough k-means, J. Intell. Inf. Syst. 23 (1) (2004) 5–16.
[280] T. Zhang, L. Chen, F. Ma, A modified rough c-means clustering algorithm based on hybrid imbalanced measure of distance and density, Int. J. Approx. Reason. 55 (8) (2014) 1805–1818.
[281] T. Herawan, M.M. Deris, J.H. Abawajy, A rough set approach for selecting clustering attribute, Knowl.-Based Syst. 23 (3) (2010) 220–231.
[282] F. Li, M. Ye, X. Chen, An extension to rough c-means clustering based on decision-theoretic rough sets model, Internat. J. Approx. Reason. 55 (1) (2014) 116–129.
[283] F. Cao, J. Liang, G. Jiang, An initialization method for the K-means algorithm using neighborhood model, Comput. Math. Appl. 58 (3) (2009) 474–483.
[284] I.T.R. Yanto, P. Vitasari, T. Herawan, M.M. Deris, Applying variable precision rough set model for clustering student suffering studys anxiety, Expert Syst. Appl. 39 (1) (2012) 452–459.
[285] F. Pacheco, M. Cerrada, R.-V. Sánchez, D. Cabrera, C. Li, J.V. de Oliveira, Attribute clustering using rough set theory for feature selection in fault severity classification of rotating machinery, Expert Syst. Appl. 71 (2017) 69–86.
[286] D. Parmar, T. Wu, J. Blackhurst, Mmr: an algorithm for clustering categorical data using rough set theory, Data Knowl. Eng. 63 (3) (2007) 879–893.
[287] M. Li, S. Deng, L. Wang, S. Feng, J. Fan, Hierarchical clustering algorithm for categorical data using a probabilistic rough set model, Knowl.-Based Syst. 65 (2014) 60–71.
[288] L.J. Mazlack, A. He, Y. Zhu, A rough set approach in choosing partitioning attributes, in: Proceedings of the ISCA 13th International Conference (CAINE-2000), Citeseer, 2000.
[289] Z. Pawlak, A. Skowron, Rudiments of rough sets, Inform. Sci. 177 (1) (2007) 3–27.
[290] Y. Yao, Two views of the theory of rough sets in finite universes, Int. J. Approx. Reason. 15 (4) (1996) 291–317.
[291] G. Peters, Rough clustering utilizing the principle of indifference, Inform. Sci. 277 (2014) 358–374.
[292] I. Saha, J.P. Sarkar, U. Maulik, Integrated rough fuzzy clustering for categorical data analysis, Fuzzy Sets and Systems 361 (2019) 1–32.
[293] G. Peters, Some refinements of rough k-means clustering, Pattern Recognit. 39 (8) (2006) 1481–1491.
[294] H. Wang, M. Zhou, A refined rough k-means clustering with hybrid threshold, in: International Conference on Rough Sets and Current Trends in Computing, Springer, 2012, pp. 26–35.
[295] S. Mitra, H. Banka, W. Pedrycz, Rough–fuzzy collaborative clustering, IEEE Trans. Syst. Man Cybern. B 36 (4) (2006) 795–805.
[296] I. Saha, J.P. Sarkar, U. Maulik, Ensemble based rough fuzzy clustering for categorical data, Knowl.-Based Syst. 77 (2015) 114–127.
[297] K. Simiński, Rough fuzzy subspace clustering for data with missing values, Comput. Inform. 33 (1) (2014) 131–153.

[298] A. Bal, M. Banerjee, A. Chakrabarti, P. Sharma, Mri brain tumor segmentation and analysis using rough-fuzzy C-means and shape based properties, J. King Saud Univ.-Comput. Inf. Sci. (2018).
[299] J. Hu, T. Li, C. Luo, H. Fujita, Y. Yang, Incremental fuzzy cluster ensemble learning based on rough set theory, Knowl.-Based Syst. 132 (2017) 144–155.
[300] J. Hu, T. Li, H. Wang, H. Fujita, Hierarchical cluster ensemble model based on knowledge granulation, Knowl.-Based Syst. 91 (2016) 179–188.
[301] C. Gao, W. Pedrycz, D. Miao, Rough subspace-based clustering ensemble for categorical data, Soft Comput. 17 (9) (2013) 1643–1658.
[302] P. Lingras, M. Chen, D. Miao, Qualitative and quantitative combinations of crisp and rough clustering schemes using dominance relations, Int. J. Approx. Reason. 55 (1) (2014) 238–258.
[303] X. Wang, D. Han, C. Han, Rough set based cluster ensemble selection, in: Proceedings of the 16th International Conference on Information Fusion, IEEE, 2013, pp. 438–444.
[304] D.-q. Miao, S.-d. Fan, The calculation of knowledge granulation and its application, Syst. Eng.-Theory Pract. 1 (2002) 7–14.
[305] J. Liang, J. Wang, Y. Qian, A new measure of uncertainty based on knowledge granulation for rough sets, Inform. Sci. 179 (4) (2009) 458–470.
[306] J. Liang, R. Li, Y. Qian, Distance: A more comprehensible perspective for measures in rough set theory, Knowl.-Based Syst. 27 (2012) 126–136.
[307] A. Fred, Finding consistent clusters in data partitions, in: International Workshop on Multiple Classifier Systems, Springer, 2001, pp. 309–318.
[308] Z. Lin, F. Yang, Y. Lai, X. Gao, T. Wang, A scalable approach of co-association cluster ensemble using representative points, in: 2017 32nd Youth Academic Annual Conference of Chinese Association of Automation (YAC), IEEE, 2017, pp. 1194–1199.
[309] V. Berikov, I. Pestunov, Ensemble clustering based on weighted co-association matrices: Error bound and convergence properties, Pattern Recognit. 63 (2017) 427–436.
[310] C. Zhong, X. Yue, Z. Zhang, J. Lei, A clustering ensemble: Two-level-refined co-association matrix with path-based transformation, Pattern Recognit. 48 (8) (2015) 2699–2709.
[311] M. Ma, Advances in three-way decisions and granular computing, Knowl.-Based Syst. 91 (2016) 1–3.
[312] H. Yu, C. Zhang, G. Wang, A tree-based incremental overlapping clustering method using the three-way decision theory, Knowl.-Based Syst. 91 (2016) 189–203.
[313] H. Yu, A framework of three-way cluster analysis, in: International Joint Conference on Rough Sets, Springer, 2017, pp. 300–312.
[314] H. Yu, C. Zhang, G. Wang, A tree-based incremental overlapping clustering method using the three-way decision theory, Knowl. Based Syst. 91.
[315] P. Wang, Y. Yao, Ce3: A three-way clustering method based on mathematical morphology, Knowl.-Based Syst. 155 (2018) 54–65.
[316] C. Jiang, Y. Duan, J. Yao, Resource-utilization-aware task scheduling in cloud platform using three-way clustering, J. Intell. Fuzzy Syst. 37 (4) (2019) 5297–5305.
[317] M.K. Afridi, N. Azam, J. Yao, E. Alanazi, A three-way clustering approach for handling missing data using gtrs, Internat. J. Approx. Reason. 98 (2018) 11–24.
[318] M.K. Afridi, N. Azam, J. Yao, Variance based three-way clustering approaches for handling overlapping clustering, Internat. J. Approx. Reason. 118 (2020) 47–63.
[319] H. Yu, L. Chen, J. Yao, X. Wang, A three-way clustering method based on an improved dbscan algorithm, Physica A 535 (2019) 122289.
[320] Y. Liu, R.-S. Zhang, A three-way grey incidence clustering approach with changing decision objects, Comput. Ind. Eng. 137 (2019) 106087.
[321] H. Yu, L. Chen, J. Yao, A three-way density peak clustering method based on evidence theory, Knowl.-Based Syst. 211 (2021) 106532.
[322] L. Bai, J. Liang, F. Cao, Semi-supervised clustering with constraints of different types from multiple information sources, IEEE Trans. Pattern Anal. Mach. Intell. (2020) http://dx.doi.org/10.1109/TPAMI.2020.2979699.