Applied Soft Computing Journal 179 (2025) 113276

Contents lists available at ScienceDirect

Applied Soft Computing

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

Uncertainty measurement for hybrid data using KNN-neighbourhood rough set model: Application to attribute reduction based on overlap degree

Bozhan Li [a], Qin Huang [b,*], Zhaowen Li [c], Yonghua Lin [b]

[a] School of Mathematics and Information Science, Guangxi University, Nanning, Guangxi 530004, PR China
[b] Key Laboratory of Applied Mathematics in Fujian Province University, Fujian Key Laboratory of Financial Information Processing, Putian University, Putian, Fujian 351100, PR China
[c] School of Computer Science, Guangdong University of Science and Technology, Dongguan, Guangdong 523083, PR China

ARTICLE INFO

Keywords:
Hybrid information system
KNN-neighbourhood rough set model
N-rough set
Uncertainty measurement
Attribute reduction
Overlap degree

ABSTRACT

Neighbourhood rough set (N-rough set) model is extensively applied to work with numerical data. However, most N-rough set models cannot effectively classify categories when they are used to depict classification ability. k-nearest neighbour rule (KNN rule) is an important classification technique. This paper introduces KNN-neighbourhood rough set model by combining N-rough set model with KNN rule, investigates uncertainty measurement (UM) for hybrid data using this model, and considers its application to attribute reduction based on overlap degree. First, a specific distance function is constructed to describe the difference between two information values with respect to each attribute in a hybrid information system (HIS), the distance matrix is defined to establish the neighbourhood relation on the object set, and the neighbourhood classes where a variable parameter is selected to adjust the size of the neighbourhood classes are built. N-rough set model is introduced using the neighbourhood classes. The k-nearest neighbour in an HIS is then proposed on the basis of KNN rule, and KNN-neighbourhood rough set model is presented by combining N-rough set model with the k-nearest neighbour. On the basis of KNN-neighbourhood rough set model, the information granules in an HIS are constructed, and four UMs in an HIS are presented. Next, an experiment is carried out to select the UM with the optimal performance. Furthermore, an attribute reduction algorithm is designed using the selected UM. Finally, the designed algorithm is evaluated using a series of experiments implemented on real-world datasets. The experimental results demonstrate the statistical advantages of the designed algorithm over the other six reduction algorithms.

1. Introduction

1.1. Research background

Uncertainty, which encompasses fuzziness, vagueness, randomness, incompleteness, and inconsistency, is prevalent in the real world. Thus, uncertainty measurement (UM) is becoming increasingly important. UM plays a crucial role in gauging the uncertainty of an information system (IS), and numerous researchers have undertaken thorough studies on this subject. For example, Pawlak [1] suggested approximate accuracy and roughness by using upper and lower approximation sets. Miao [2] employed information entropy to quantify the uncertainty of an IS. In a related context, Liang et al. [3] conducted a study on UM in an incomplete IS. Wang et al. [4] delved into innovative UM by means of Dempster–Shafer evidence theory. Zeng et al. [5] discussed UM based on a Gaussian kernel in the context of a hybrid information system (HIS).

k-nearest neighbour rule (KNN rule) is recognised as one of the most influential classification algorithms in the field of machine learning [6]; its working principle is explained below. First, an appropriate number of neighbours, denoted as k, which represents the quantity of nearest neighbours considered during prediction, is selected. Then, for each given object, a predefined distance function is utilised to calculate the distance between that object and all other objects in the training dataset [7]. Next, KNN rule identifies the k-nearest neighbour based on the computed distances. Finally, for classification tasks, a majority voting approach is applied, wherein the predicted result is determined by the most frequently occurring class among the k-nearest neighbours. For regression tasks, the prediction is obtained by taking the average of the values from the k-nearest neighbour.

Pawlak proposed rough set theory (RST) [1], which aims to develop a granular view for interpreting and solving problems with uncertain

* Corresponding author.
E-mail addresses: libozhanpak@126.com (B. Li), qinhuang780@126.com (Q. Huang), lizhaowen8846@126.com (Z. Li), lyh040@163.com (Y. Lin).

https://doi.org/10.1016/j.asoc.2025.113276
Received 17 March 2024; Received in revised form 9 April 2025; Accepted 29 April 2025
Available online 29 May 2025

knowledge [8]. Rough set analysis requires no additional information, external parameters, models, functions, grades, or subjective interpretations to ascertain set membership. Instead, it relies solely on the information presented within the given data. Currently, RST has found applications in various fields, such as outlier detection [9], UM [10], and attribute reduction [11]. While the classical rough set model is well suited for symbolic or categorical data processing because of its dependence on an equivalence relation, a challenge arises when dealing with numerical data. In such cases, the method resorts to discretisation, resulting in the regrettable consequences of information loss and diminished efficiency.

To overcome the shortcomings of RST, N-rough set is proposed, which can work with numerical data. It focuses on the interactions among objects within their respective neighbourhoods, thereby conducting a more detailed analysis of the dataset's structure. Initially, N-rough set is based on attributes such as the similarity (or distance function) in attribute values and topological relationships to define the similarity measure between objects. Using this similarity measure, the model establishes neighbourhood relations. Neighbourhood relations play a critical role in N-rough set. However, the neighbourhood relation may not accurately capture the difference between objects. The reason is that objects with a neighbourhood relation to the target object still have differences, even if they are equidistant from the target point. Recently, many studies have been conducted to improve N-rough set in a variety of ways. These methods include the extended N-rough set [12], the variable radius N-rough set [13], the shared N-rough set [14], the pseudo-label N-rough set [15], and the adaptive N-rough set [16]. Additionally, KNN rule has been incorporated into N-rough set using various approaches, such as the k-nearest N-rough set [17] and the weighted k-nearest N-rough set [18].

Attribute reduction, also known as feature selection, is a vital data preprocessing technique. These methods can be classified into three categories: filter, wrapper, and embedded methods [19]. Filter methods score and rank features based on criteria such as distance, information, dependence, and consistency measures. Wrapper methods optimise feature subsets using classifiers, making them generally inefficient. Embedded methods integrate feature selection within the learning algorithm's training process. However, using rough set models for attribute reduction is a unique approach that typically does not fall into the categories of filter methods, wrapper methods, or embedded methods. Rough set models simplify datasets by identifying important attributes and reducing unnecessary redundancy. This method emphasises handling uncertainty and vagueness through UMs and decision tables. Attribute reduction methods based on rough set models use the dataset itself and do not rely on specific classifiers. These methods have many advantages over the three feature selection methods mentioned above, such as being very versatile, being easy to expand, and allowing autonomous selection of the number of retained attributes. There are numerous attribute reduction methods based on various rough set models, where each is designed to address the limitations of N-rough set and handle different datasets. For example, Wang et al. [20] designed an attribute reduction algorithm using local condition entropy based on N-rough set, which is capable of handling hybrid data. Additionally, Wang et al. [17] proposed the k-nearest N-rough set and developed an attribute reduction algorithm that handles numerical data using the positive region as the UM. For ease of presentation, Table 1 summarises other works on attribute reduction in recent years, along with the data processing scope and applicable UM of the attribute reduction algorithms in the corresponding articles.

1.2. Motivation

N-rough set model uses a distance function to measure the distance between two objects under each categorical and numerical attribute. It then establishes a neighbourhood for each object using a neighbourhood radius. The established neighbourhoods are treated as knowledge granules, and a corresponding attribute reduction algorithm is developed using granular computing. However, with a fixed neighbourhood radius, the classification ability of N-rough set is influenced by different information spaces [17]. Similarly, Hu et al. [30] observed that when the neighbourhood radius is too large, the granularity of neighbourhood information becomes excessively large, resulting in a reduction in the approximation ability of attribute subsets to decision classes. Conversely, if the neighbourhood radius is too small, then the granularity of neighbourhood information becomes too small, potentially enhancing the approximation ability of attribute subsets to decision classes but also increasing the risk of overfitting.

Based on the above analysis, we can conclude that (1) the same neighbourhood radius cannot make N-rough set model suitable for datasets with different information distributions and that (2) having a neighbourhood radius that is too large or too small can cause various negative impacts. To overcome these two issues, we consider introducing KNN rule into N-rough set model, integrating it to form a new rough set model named the KNN-neighbourhood rough set model. Before we further analyse why the KNN-neighbourhood rough set model can overcome these two issues, we introduce surrounded objects and isolated objects. In a certain decision table, objects with different labels are distributed in the information space M. For convenience in later descriptions, we define $[u]_d = [v]_d$ if u and v have the same label. Additionally, we define $d(u, v)$ as the distance between u and v in the information space M, $\delta(u, \lambda) = \{v \in M : d(u, v) \leq \lambda\}$, and $|\delta(u, \lambda)|$ is the cardinality of $\delta(u, \lambda)$. With these declarations, let us now formally define two important object types: surrounded objects and isolated objects. An object $u \in M$ is called a surrounded object if there exists a radius λ such that $\forall\ v \in \delta(u, \lambda)$, $[u]_d = [v]_d$. Moreover, $|\delta(u, \lambda)|$ should be greater than or equal to a positive number so that $\delta(u, \lambda)$ has not only one object u. An object $u \in M$ is called an isolated object if there exists a radius λ_1 such that $v \in \delta(u, \lambda_1) - \{u\}$ implies $[u]_d \neq [v]_d$ and $|\delta(u, \lambda_2)| = 1$ for $\lambda_2 < \lambda_1$. These two types of objects can induce three distinct types of patterns in the decision table. A label in the decision table is defined as surrounded category if all objects with this type of label are surrounded objects. A label in the decision table is an isolated category if all the objects with this type of label are isolated objects. A label in the decision table is defined as a mixed category if it is neither surrounded category nor an isolated category.

The motivation for KNN-neighbourhood rough set model is illustrated using three decision tables, each containing only "+" and "−" labels. In Fig. 1(a), all "+" objects (e.g., u_1) are surrounded, making "+" a surrounded category. In Fig. 1(b), all "−" objects (e.g., u_2) are isolated, so "−" is an isolated category. In Fig. 1(c), the "+" category is mixed because it includes both an isolated object (u_3) and a surrounded object (u_4).

The relevant concepts have been clarified in advance. Now, we formally analyse why KNN-neighbourhood rough set model can solve the above two issues. First, introducing KNN rule into N-rough set can better handle the scenario of surrounded categories in the information space than can the sole application of N-rough set. For example, consider the scenario of a surrounded category, as shown in Fig. 1(a), where the red circle illustrates the neighbourhood of u_1, and the black circle represents the k-nearest neighbours of u_1, where pick $k = 4$. Fig. 1(a) intuitively shows that N-rough set may classify u_1 as a boundary region. However, utilising k-nearest neighbour of u_1 can lead to better classification, correctly assigning it to "+". Second, N-rough set model can better handle the scenario of isolated categories in the information space than can the sole application of KNN rule. For example, in an isolated category scenario, as shown in Fig. 1(b), the red circle illustrates the neighbourhood of u_2, whereas the black circle denotes the k-nearest neighbour of u_2, where pick $k = 4$. In this situation, if we use KNN rule, then u_2 will be incorrectly classified as "+", whereas if we use N-rough set model, then u_2 is correctly classified as "−". Notably, according to the subsequent introduction of KNN-neighbourhood rough set, we can infer that when k is too large, it degenerates into N-rough set and can better identify isolated objects.

Table 1
Comparison of this paper with other works on attribute reduction.

Year	Literature	Data type	Model or Method	UM or criterion in attribute reduction
2019	[20]	Hybrid data	N-rough set	Local conditional entropy
2019	[17]	Numerical data	k-nearest N-rough set	Positive region
2020	[21]	Categorical data	Fuzzy rough computation model	Significance degree
2022	[22]	Dynamic ordered data	Fuzzy dominance N-rough set	Significance degree
2023	[23]	Categorical data	Pawlak rough set	Conditional information entropy
2023	[12]	Numerical data	The extended N-rough set	Dependence degree
2023	[24]	Numerical data	Class-specific feature selection	Dependence degree
2023	[25]	Partially labelled data	N-rough set	Significance degree
2024	[26]	Numerical data	Distance metric learning-based multi-granularity N-rough set	Dependence degree
2024	[27]	Numerical data	The granule-specific feature selection	Significance degree
2024	[16]	Numerical data	The adaptive N-rough set	Dependence degree
2024	[28]	Numerical data	N-rough set model based on neighbourhood equivalence relation	Dependence degree
2024	[18]	Numerical data	Weighted k-nearest N-rough set	Significance degree
2024	[14]	Numerical data	The shared N-rough set	Dependence degree
2024	[19]	Numerical data	Granular-ball N-rough set	Positive region samples
2024	[29]	Hybrid data	Fuzzy evidence theory	Fuzzy evidence function
Our		Hybrid data	KNN-neighbourhood rough set	Dependence degree

In other words, we can draw the following conclusions. First, when the information space is composed of surrounded categories, KNN-neighbourhood rough set model is more capable of correctly identifying surrounded objects than N-rough set model is. Second, when the information space is composed of isolated categories, by setting the parameter k large enough, KNN-neighbourhood rough set model will degrade into N-rough set model, thereby better identifying isolated objects. Third, when the information space is composed of mixed categories, KNN-neighbourhood rough set model can select an appropriate combination of parameters (k, λ) to identify as many surrounded objects and isolated objects as possible.

Furthermore, Our approach can capture and handle global outlier scenarios from noise by relying on strict definitions. Specifically, the definitions of surrounded and isolated objects require that a surrounded object must have a neighbourhood (with a prescribed radius and minimum cardinality) in which all objects share the same label, while an isolated object must have a neighbourhood (with a minimum cardinality) where, excluding itself, there exists at least one object that has a different label. If an object is merely a global outlier, then it will remain alone (within an suitable chosen neighbourhood radius). In this case, it cannot meet the definition of a surrounded object (it does not have a minimum cardinality) or isolated object (it does not have other objects, except for itself). Consequently, the label of such an outlier would not be recognised as a surrounded object or an isolated oject.

1.3. Contributions

This paper integrates the strengths of both N-rough set and KNN-rule and then combines the two methodologies by introducing KNN-neighbourhood rough set model. We investigate UM in an HIS using this model and apply it to attribute reduction based on overlap degree.

Drawing from the aforementioned research motivation, this paper makes the following contributions:

(1) This paper introduces KNN-rule into N-rough set model, resulting in the development of KNN-neighbourhood rough set model. This model is particularly effective in identifying isolated objects and surrounded objects within a dataset. By incorporating KNN-rule, the model can better handle local neighbourhood relationships, which traditional rough set models may overlook. This advancement improves the model's ability to capture subtle patterns in the data, making it more robust for real-world applications where isolated or surrounded objects are critical for accurate analysis.

(2) Utilising KNN-neighbourhood rough set model, four UMs are formulated: the dependence degree, conditional information entropy, conditional information amount and conditional discrimination index. Among these, the UM with the best performance is selected by dispersion analysis. By choosing the most stable UM, we ensure more reliable and consistent results in attribute reduction, thereby enhancing the overall effectiveness of the method.

(3) We design an attribute reduction algorithm based on the selected UM and overlap degree and then compare the designed algorithm with six existing algorithms in terms of classification accuracy and F1 score to demonstrate its statistically significant advantage.

1.4. Structure and organisation

The subsequent sections of this paper are structured as follows. Section 2 presents a comprehensive review of the background. Section 3 derives a KNN-neighbourhood rough set model. Section 4 introduces four UMs designed to quantify the uncertainty inherent in an HIS. Section 5 conducts an experimental analysis to evaluate the feasibility of these UMs and identifies the one that demonstrates optimal performance. The chosen UM is then utilised to formulate an attribute reduction algorithm. The performance of the designed algorithm is then compared with that of six alternative algorithms through a series of experiments, which demonstrate its effectiveness and superior performance. Section 6 provides a comprehensive summary of the key findings.

Fig. 2 illustrates the procedural flow of this paper.

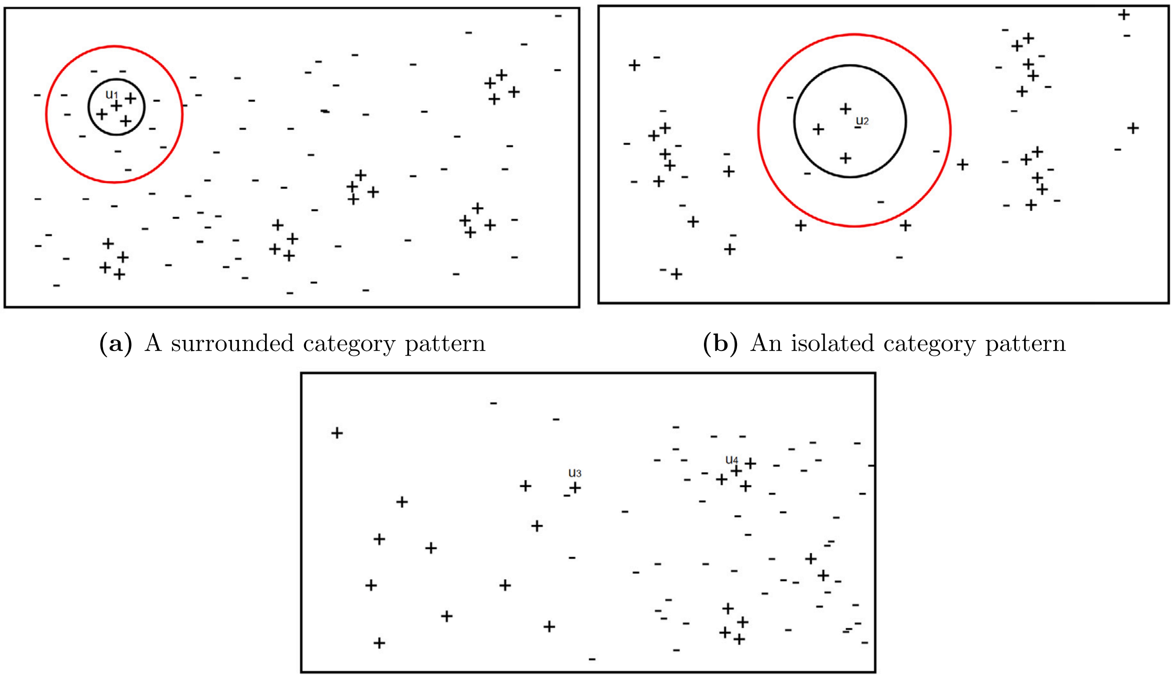
2. Background

2.1. Fundamentals of classification

Classification is a type of pattern analysis. When using a model to find patterns in data, feeding it raw data may not result in satisfactory performance; the reason is that the model has the challenging task of extracting useful information from the data. Therefore, the raw data should be preprocessed to retain the most valuable information. In this way, the model can achieve better performance and reduce the time required for training.

Classification is the process of categorising data into predefined classes or groups based on their features; it is widely used in various fields, such as image recognition, spam detection, and medical diagnosis. The process involves training a model on a labelled dataset, where the classes are already known, and then using this trained model to predict the class of new, unseen data. Effective classification relies on the quality of the features extracted from the raw data, which is why preprocessing steps such as attribute reduction, normalisation, and feature extraction are crucial. These steps help highlight the relevant information and reduce noise, making it easier for the model to identify the underlying patterns and relationships within the data.

By ensuring that the data fed into the classification model are clean and well structured, the overall accuracy and efficiency of the



(a) A surrounded category pattern

(b) An isolated category pattern

(c) A mixed category pattern

Fig. 1. Examples of category patterns: surrounded, isolated, and mixed.

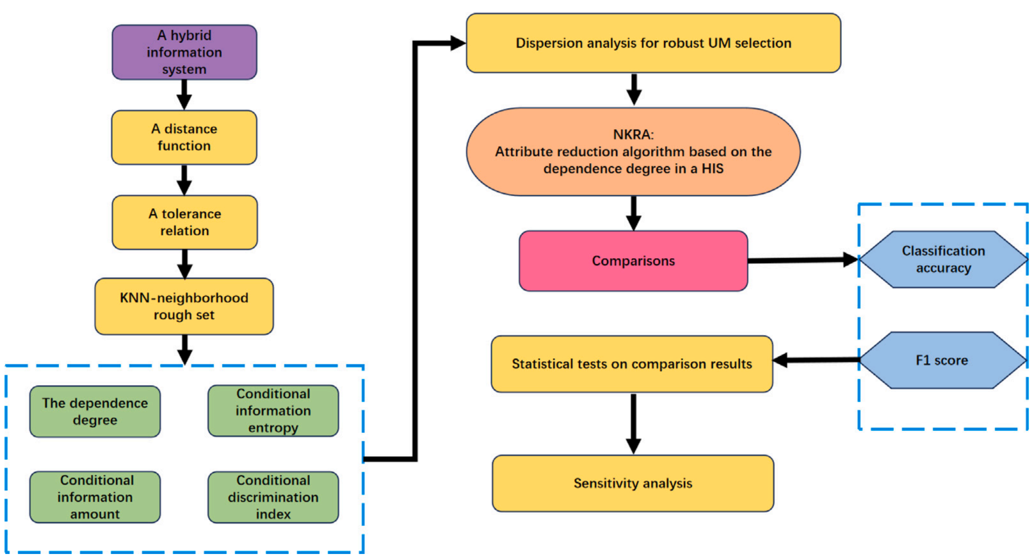


Fig. 2. Schematic diagram illustrating the process of this paper.

classification process can be significantly improved. This leads to more reliable and faster predictions, making the model more practical and useful in real-world applications.

2.2. Fundamentals of an HIS

In this subsection, an HIS and N-rough set model are reviewed.

Let U, 2^U, $|X|$, and $\mathbb{N}$ represent a finite set, the power set of U, the cardinality of $X \in 2^U$, and the set of natural numbers, respectively. Put

$$U = \{u_1, u_2, \dots, u_n\}; \tag{2.1}$$

$$\delta = U \times U,\ \triangle = \{(u,u) : u \in U\}. \tag{2.2}$$

Assume that U is a finite set of objects and that A represents a finite set of attributes. (U, A) denotes an information system (IS) where if $a : U \to V_a$ is a function of $a \in A$, then $V_a = \{a(u) : u \in U\}$.

(U, A, d) is characterised as an incomplete decision information system (IDIS) when $(U, A \cup \{d\})$ constitutes an IS and d represents a decision attribute. In this context, $a \in A$ such that $* \in V_a$, whereas $* \notin V_d$, where " $*$ " signifies a missing value.

$\forall\ a \in A$, denote

$$V_a^* = \{a(u) : u \in U,\ a(u) \neq *\}. \tag{2.3}$$

Definition 2.1 ([5]). Consider (U, A, d) as an IDIS. (U, A, d) is referred to as an HIS if $A = A^n \cup A^c$, where A^n and A^c represent numerical and categorical attributes, respectively.

Table 2
An HIS.

	Headache (a_1)	Runny nose (a_2)	Temperature (a_3)	Symptom (d)
u_1	Sick	Yes	37.8	Common cold
u_2	Middle	No	39.5	Influenza
u_3	No	No	37.2	Health
u_4	*	No	*	Health
u_5	No	Yes	36.4	Influenza
u_6	No	*	37.1	Health

Example 2.2. Table 2 presents an HIS (U, A, d), where $U = \{u_1, u_2, \ldots, u_7\}$, $A = A^c \cup A^n$, $A^c = \{a_1, a_2\}$, $A^n = \{a_3\}$ and d represent the decision attributes.

Example 2.3 (*Building Upon the Illustration in Example 2.2*).
$V^*_{a_1} = \{Sick, Middle, No\}$, $V^*_{a_2} = \{No, Yes\}$,
$V^*_{a_3} = \{37.8, 39.5, 37.2, 36.4, 37.1\}$,
$V_d = \{Common\ cold, Influenza, Health\}$.

For an HIS (U, A, d) with $|V_d| = s$. Given that $a \in A^c$ and u with $a(u) \neq *$. Denote

$$V_d = \{d_1, d_2, \ldots, d_s\}.$$

Put

$$N(a,u) = |\{v \in U : a(v) = a(u)\}|,$$

$$N_i(a,u) = |\{v \in U : a(v) = a(u),\ d(v) = d_i\}|.$$

Clearly,

$$N(a,u) = \sum_{i=1}^{s} N_i(a,u).$$

Example 2.4 (*Building Upon the Illustration in Example 2.3*).

$$V_d = \{d_1 = Common\ cold, d_2 = Influenza, d_3 = Health\};$$

$$\begin{aligned} N(a_1,u_3) &= |\{u \in U : a_1(u) = a_1(u_3) = No\}| \\ &= |\{u_3,u_5,u_6\}| = 3; \end{aligned}$$

$$\begin{aligned} N_2(a_1,u_3) = |\{&u \in U : a_1(u) = a_1(u_3) = No, \\ &d(u) = d_2 = Influenza\}| = |\{u_5\}| = 1; \end{aligned}$$

2.3. Fundamentals of the neighbourhood rough set model

In this subsection, Definition 2.5 introduces a distance between two information values in complete or incomplete ISs with categorical, numerical, or hybrid data. The neighbourhood relation is then induced by comparing the distance and a parameter λ. After this, a tolerance class can be computed by the neighbourhood relation. Finally, N-rough set model is induced by the tolerance class.

Definition 2.5 ([31]). In an HIS (U, A, d) where $|V_d|$ is denoted as s, consider $a \in A$ and $u, v \in U$. The distance between $a(u)$ and $a(v)$ is defined as follows:

$$\rho(a(u), a(v)) = \begin{cases} 0, & a \in A,\ a(u) = *\ \ or\ a(v) = *,\ d(u) = d(v); \\ 1 - \frac{1}{|V^*_a|^2}, & a \in A,\ a(u) = *,\ a(v) = *,\ d(u) \neq d(v); \\ 1 - \frac{1}{|V^*_a|}, & a \in A,\ a(u) \neq *,\ a(v) = *,\ d(u) \neq d(v); \\ 1 - \frac{1}{|V^*_a|}, & a \in A,\ a(u) = *,\ a(v) \neq *,\ d(u) \neq d(v); \\ 0, & a \in A,\ a(u) \neq *,\ a(v) \neq *,\ a(u) = a(v); \\ \frac{1}{2}\sum_{i=1}^{s} \left| \frac{N_i(a,u)}{N(a,u)} - \frac{N_i(a,v)}{N(a,v)} \right|, & a \in A^c,\ a(u) \neq *,\ a(v) \neq *,\ a(u) \neq a(v), \\ & d(u) = d(v); \\ \frac{1}{2}\sum_{i=1}^{s} \left| \frac{N_i(a,u)}{N(a,u)} - \frac{N_i(a,v)}{N(a,v)} \right|, & a \in A^c,\ a(u) \neq *,\ a(v) \neq *,\ a(u) \neq a(v), \\ & d(u) \neq d(v); \\ \frac{|a(u)-a(v)|}{max\{a(x):x\in U\}-min\{a(x):x\in U\}}, & a \in A^n,\ a(u) \neq *,\ a(v) \neq *,\ a(u) \neq a(v), \\ & d(u) = d(v); \\ \frac{|a(u)-a(v)|}{max\{a(x):x\in U\}-min\{a(x):x\in U\}}, & a \in A^n,\ a(u) \neq *,\ a(v) \neq *,\ a(u) \neq a(v), \\ & d(u) \neq d(v). \end{cases}$$

When computing the distance between two non-missing values in a categorical attribute, the calculation combines the information provided by the two non-missing values in that attribute and a decision attribute. When computing the distance between two non-missing values in a numerical attribute, the calculation involves the absolute difference between the two non-missing values in that attribute and the range of values in that attribute. Despite having greater computational complexity than other distance functions, this distance function can provide a more accurate description of the distance between two objects, offering significant theoretical advantages and broader applicability.

Example 2.6 (*Building Upon the Illustration in Example 2.4*).
(1) Since $a_1 \in A,\ a_1(u_3) = *\ or\ a_1(u_4) = *,\ d(u_3) = d(u_4)$, according to Definition 2.5, we have $\rho(a_1(u_3), a_1(u_4)) = 0$;
(2) Note that $a_1 \in A,\ a_1(u_2) \neq *,\ a_1(u_4) = *,\ d(u_2) \neq d(u_4)$. According to Definition 2.5, $\rho(a_1(u_2), a_1(u_4)) = 1 - \frac{1}{|V^*_{a_1}|} = \frac{2}{3} \approx 0.667$;
(3) Since $a_1 \in A,\ a_1(u_5) \neq *,\ a_1(u_6) \neq *,\ a_1(u_5) = a_1(u_6)$, according to Definition 2.5, we have $\rho(a_1(u_5), a_1(u_6)) = 0$;
(4) Note that $a_1 \in A^c,\ a_1(u_2) \neq *,\ a_1(u_3) \neq *,\ a_1(u_2) \neq a_1(u_3),\ d(u_2) \neq d(u_3)$. According to Definition 2.5, $\rho(a_1(u_2), a_1(u_3)) = (|\frac{0}{1} - \frac{0}{3}| + |\frac{1}{1} - \frac{1}{3}| + |\frac{0}{1} - \frac{2}{3}|)/2 = \frac{2}{3} \approx 0.667$;
(5) Since $a_3 \in A^n,\ a_3(u_1) \neq *,\ a_3(u_2) \neq *,\ a_3(u_1) \neq a_3(u_2),\ d(u_1) \neq d(u_2)$, according to Definition 2.5, we have $\rho(a_3(u_1), a_3(u_2)) = \frac{|37.8-39.5|}{39.5-36.4} \approx 0.548$.

Definition 2.7 ([32]). In the context of an HIS (U, A, d) with $|U| = n$, the distance matrix M_a for any $a \in A$ is introduced:

$$M_a = (\rho(a(u_i), a(u_j)))_{n\times n}. \tag{2.4}$$

Here, M_a signifies the distance matrix corresponding to attribute a.

Example 2.8 (*Building Upon the Illustration in Example 2.2*). Combining Definitions 2.5 and 2.7, we obtain

$$M_{a_1} = \begin{pmatrix} 0 & 1 & 1 & 0.667 & 1 & 1 \\ 1 & 0 & 0.667 & 0.667 & 0.667 & 0.667 \\ 1 & 0.667 & 0 & 0 & 0 & 0 \\ 0.667 & 0.667 & 0 & 0 & 0.667 & 0 \\ 1 & 0.667 & 0 & 0.667 & 0 & 0 \\ 1 & 0.667 & 0 & 0 & 0 & 0 \end{pmatrix},$$

Table 3
The tolerance classes of each object.

	$N^{\lambda}_{a_1}(u)$	$N^{\lambda}_{a_2}(u)$	$N^{\lambda}_{a_3}(u)$	$[u]_d$
u_1	$\{u_1\}$	$\{u_1,u_5\}$	$\{u_1,u_3,u_6\}$	$\{u_1\}$
u_2	$\{u_2\}$	$\{u_2,u_3,u_4\}$	$\{u_2\}$	$\{u_2,u_5\}$
u_3	$\{u_3,u_4,u_5,u_6\}$	$\{u_2,u_3,u_4,u_6\}$	$\{u_1,u_3,u_4,u_5,u_6\}$	$\{u_3,u_4,u_6\}$
u_4	$\{u_3,u_4,u_6\}$	$\{u_2,u_3,u_4,u_6\}$	$\{u_3,u_4,u_6\}$	$\{u_3,u_4,u_6\}$
u_5	$\{u_3,u_5,u_6\}$	$\{u_1,u_5\}$	$\{u_3,u_5,u_6\}$	$\{u_2,u_5\}$
u_6	$\{u_3,u_4,u_5,u_6\}$	$\{u_3,u_4,u_6\}$	$\{u_1,u_3,u_4,u_5,u_6\}$	$\{u_3,u_4,u_6\}$

$$M_{a_2}=\begin{pmatrix} 0 & 0.667 & 0.667 & 0.667 & 0 & 0.5\\ 0.667 & 0 & 0 & 0 & 0.667 & 0.5\\ 0.667 & 0 & 0 & 0 & 0.667 & 0\\ 0.667 & 0 & 0 & 0 & 0.667 & 0\\ 0 & 0.667 & 0.667 & 0.667 & 0 & 0.5\\ 0.5 & 0.5 & 0 & 0 & 0.5 & 0 \end{pmatrix},$$

$$M_{a_3}=\begin{pmatrix} 0 & 0.548 & 0.194 & 0.8 & 0.452 & 0.226\\ 0.548 & 0 & 0.742 & 0.8 & 1 & 0.774\\ 0.194 & 0.742 & 0 & 0 & 0.258 & 0.032\\ 0.8 & 0.8 & 0 & 0 & 0.8 & 0\\ 0.452 & 1 & 0.258 & 0.8 & 0 & 0.226\\ 0.226 & 0.774 & 0.032 & 0 & 0.226 & 0 \end{pmatrix}.$$

Definition 2.9 ([14]). Consider an HIS (U,A,d) with $|V_d|=s$, given $B\subseteq A$ and $\lambda\in[0,1]$. Define

$$N^{\lambda}_a=\{(u,v)\in U\times U : \rho(a(u),a(v))\le\lambda\},\ N^{\lambda}_B=\bigcap_{a\in B}N^{\lambda}_a,$$

$$N^{\lambda}_B(u)=\{v\in U : (u,v)\in N^{\lambda}_B\};$$

$$R_d=\{(u,v)\in U\times U : d(u)=d(v)\},$$

$$[u]_d=\{v\in U : (u,v)\in R_d\},$$

$$U/d=\{[u]_d : u\in U\}=\{D_1,D_2,\dots,D_s\}.$$

Then, N^{λ}_B, $N^{\lambda}_B(u)$ and $[u]_d$ are the neighbourhood relation on U with respect to B, the neighbourhood of u with respect to B and the decision class of u, respectively.

N^{λ}_B is reflexive and symmetric. $N^{\lambda}_B(u)$ can be seen as a neighbourhood information granule, referred to as the λ-neighbourhood information granule.

Example 2.10 (*Building Upon the Illustration in Example 2.8*). Choose $\lambda=0.4$. The tolerance classes for each object with respect to each attribute are displayed in Table 3.

For an HIS (U,A,d), given $B\subseteq A$, $X\in 2^U$, and $\lambda\in[0,1]$, denote

$$\underline{N^{\lambda}_B}(X)=\{u\in U : N^{\lambda}_B(u)\subseteq X\},$$

$$\overline{N^{\lambda}_B}(X)=\{u\in U : N^{\lambda}_B(u)\cap X\neq\emptyset\}.$$

Then, $\underline{N^{\lambda}_B}(X)$ and $\overline{N^{\lambda}_B}(X)$ are called neighbourhood-lower and neighbourhood-upper approximations of X with respect to B and λ, respectively. Moreover, X is known as a neighbourhood rough set (N-rough set) with respect to B and λ, if $\underline{N^{\lambda}_B}(X)\neq\overline{N^{\lambda}_B}(X)$.

3. KNN-neighbourhood rough set model in an HIS

In this part, KNN-neighbourhood rough set model in an HIS is built. After this, we propose four UMs for hybrid data using KNN-neighbourhood rough set model in the next section.

3.1. Overview

As stated in the motivation of the introduction, N-rough set approach struggles to identify surrounded objects effectively. In this section, we first introduce KNN rule. KNN-neighbourhood rough set model is subsequently proposed by incorporating KNN rule, which is controlled by the parameter k, and the neighbourhood tolerance relation, which is controlled by the parameter λ, to better handle surrounded objects and isolated objects in an HIS.

By adjusting the parameters k and λ, we can better classify surrounded objects and isolated objects. For example, when a dataset is composed of surrounded categories, selecting an appropriate k allows us to classify surrounded objects more accurately than N-rough set does, as shown in Fig. 1(a) of the motivation section. When a dataset is composed of isolated categories, choosing a sufficiently large k causes KNN-neighbourhood rough set model to degenerate into N-rough set, thereby inheriting its ability to correctly classify isolated objects, as illustrated in Fig. 1(b) of the motivation section. Furthermore, when a dataset consists of mixed categories, selecting appropriate k and λ values for the dataset enables us to correctly identify the objects as accurately as possible.

In other words, KNN-neighbourhood rough set model is based on the neighbourhood rough set, which incorporates KNN-rule to better handle surrounded objects and isolated objects in an HIS, as stated in the motivation in the introduction. By selecting different values of k and λ, we can accurately identify the two types of objects in the HIS-surrounded objects and isolated objects.

Next, we introduce KNN-neighbourhood rough set model starting from the specific mathematical theory.

3.2. Mathematical development of KNN-neighbourhood rough set model

Definition 3.1 ([30]). Consider an HIS (U,A,d), with $a\in A$ and $u\in U$. Let $k\in\mathbb{N}$ be a given parameter. Denote $u^0=u$. $\forall\ i\in\{1,2,\dots,k\}$, put

$$u^i=\underset{v}{argmin}\{\rho(a(u),a(v)) : v\in U-\{u^0,u^i,\dots,u^{i-1}\}\}$$

or

$$\rho(a(u),a(u^i))=min\{\rho(a(u),a(v)) : v\in U-\{u^0,u^1,\dots,u^{i-1}\}\}.$$

Define

$$a^k(u)=\{u^0,u^i,\dots,u^{k-1}\}.$$

Obviously, $\forall\ i\in\{1,2,\dots,k\}$,

$$\rho(a(u),a(u^{i-1}))\le\rho(a(u),a(u^i)).$$

Definition 3.2. Consider an HIS (U,A,d), where $B\subseteq A$, and $u\in U$. Given parameters $k\in\mathbb{N}$ and $\lambda\in[0,1]$, define or formulate the following:

$$B^k(u)=\bigcap_{a\in B}a^k(u),$$

$$B^{k,\lambda}_N(u)=B^k(u)\cap N^{\lambda}_B(u).$$

$B^k(u)$ is referred to as the k-nearest neighbour of u with respect to B, and $B^{k,\lambda}_N(u)$ is the KNN-neighbourhood of u with respect to B and λ.

Proposition 3.3. *Consider an HIS (U,A,d), given that $u\in U$.*

(1) *If $B\subseteq C\subseteq A$, then $\forall\ k\in\mathbb{N}$, $\forall\ \lambda\in[0,1]$, $C^{k,\lambda}_N(u)\subseteq B^{k,\lambda}_N(u)$.*
(2) *If $k_1\le k_2$, then $\forall\ B\subseteq A$, $\forall\ \lambda\in[0,1]$, $B^{k_1,\lambda}_N(u)\subseteq B^{k_2,\lambda}_N(u)$.*
(3) *If $\lambda_1\le\lambda_2$, then $\forall\ B\subseteq A$, $\forall\ k\in\mathbb{N}$, $B^{k,\lambda_1}_N(u)\subseteq B^{k,\lambda_2}_N(u)$.*

Proof. Please see "Appendix". □

Put

$B_N^{k,\lambda} = \{(u,v) \in U \times U : v \in B_N^{k,\lambda}(u)\}.$

While $B_N^{k,\lambda}$ exhibits reflexivity, it lacks symmetry and transitivity.

Definition 3.4. Consider an HIS (U,A,d); let $B \subseteq A$, $X \in 2^U$, given parameters for $k \in \mathbb{N}$ and $\lambda \in [0,1]$, define

$$\underline{B_N^{k,\lambda}}(X) = \{u \in U : B_N^{k,\lambda}(u) \subseteq X\},$$

$$\overline{B_N^{k,\lambda}}(X) = \{u \in U : B_N^{k,\lambda}(u) \cap X \neq \emptyset\}.$$

Then, $\underline{B_N^{k,\lambda}}(X)$ and $\overline{B_N^{k,\lambda}}(X)$ are called KNN-neighbourhood lower and upper approximations of X with respect to B and λ, respectively. Moreover, X is known as an KNN-neighbourhood rough set with respect to B and λ, if $\underline{B_N^{k,\lambda}}(X) \neq \overline{B_N^{k,\lambda}}(X)$.

Theorem 3.5. *Consider an HIS (U,A,d).*

(1) $\forall\ B \subseteq A,\ \forall\ X \in 2^U,\ \forall\ k \in \mathbb{N},\ \forall\ \lambda \in [0,1]$,

$$\underline{B_N^{k,\lambda}}(X) \subseteq X \subseteq \overline{B_N^{k,\lambda}}(X);$$

(2) $X \subseteq Y \subseteq U \ \Rightarrow \forall\ B \subseteq A,\ \forall\ k \in \mathbb{N},\ \forall\ \lambda \in [0,1]$,

$$\underline{B_N^{k,\lambda}}(X) \subseteq \underline{B_N^{k,\lambda}}(Y),\ \overline{B_N^{k,\lambda}}(X) \subseteq \overline{B_N^{k,\lambda}}(Y);$$

(3) *If* $B \subseteq C \subseteq A$, *then* $\forall\ X \in 2^U,\ \forall\ k \in \mathbb{N},\ \forall\ \lambda \in [0,1]$,

$$\underline{B_N^{k,\lambda}}(X) \subseteq \underline{C_N^{k,\lambda}}(X),\ \overline{C_N^{k,\lambda}}(X) \subseteq \overline{B_N^{k,\lambda}}(X);$$

(4) *If* $k_1 \leq k_2$, *then* $\forall\ B \subseteq A,\ \forall\ \lambda \in [0,1],\ \forall\ X \in 2^U$

$$\underline{B_N^{k_2,\lambda}}(X) \subseteq \underline{B_N^{k_1,\lambda}}(X),\ \overline{B_N^{k_1,\lambda}}(X) \subseteq \overline{B_N^{k_2,\lambda}}(X).$$

(5) *If* $\lambda_1 \leq \lambda_2$, *then* $\forall\ B \subseteq A,\ \forall\ k \in \mathbb{N},\ \forall\ X \in 2^U$

$$\underline{B_N^{k,\lambda_2}}(X) \subseteq \underline{B_N^{k,\lambda_1}}(X),\ \overline{B_N^{k,\lambda_1}}(X) \subseteq \overline{B_N^{k,\lambda_2}}(X).$$

Proof. Please see "Appendix". □

Based on Kryszkiewicz's ideal [33], for $B \subseteq A$, $\partial_{B,N}^{k,\lambda} : U \to 2^{V_d}$ is defined as

$$\partial_{B,N}^{k,\lambda}(u) = d(B_N^{k,\lambda}(u)). \tag{3.1}$$

Then, $\partial_{B,N}^{k,\lambda}$ is termed a generalised decision in (U,B,d).

Definition 3.6. In an HIS (U,A,d), if $\forall\ u \in U$, $|\partial_{A,N}^{k,\lambda}(u)| = 1$, it is termed (k,λ)-consistent.

Proposition 3.7. *Consider an HIS (U,A,d), given $B \subseteq A$, $k \in \mathbb{N}$ and $\lambda \in [0,1]$. Then,*

$$\forall\ u \in U,\ |\partial_{B,N}^{k,\lambda}(u)| = 1 \ \Leftrightarrow\ B_N^{k,\lambda} \subseteq R_d.$$

Proof. Please see "Appendix". □

Proposition 3.8. *Consider an HIS (U,A,d), given $k \in \mathbb{N}$ and $\lambda \in [0,1]$. Then, (U,A,d) is (k,λ)-consistent $\Leftrightarrow A_N^{k,\lambda} \subseteq R_d$.*

Proof. This is a consequence of Proposition 3.7. □

4. Uncertainty measurement in an HIS based on KNN-neighbourhood rough set model

In this section, we present four UMs designed to quantify the uncertainty inherent in an HIS using KNN-neighbourhood rough set model.

4.1. Overview

In the context of KNN-neighbourhood rough set model, four types of UMs can be induced: dependence degree, conditional information entropy, conditional information amount, and the conditional discrimination index. In this section, we rigorously prove the properties and relationships of these four UMs, thereby demonstrating their effectiveness.

Initially, each UM's definition and calculation method are detailed, followed by mathematical derivations to establish their properties. Specifically, the focus is on the following aspects:

(1) Dependence Degree: This measures the extent to which a subset of attributes depends on the decision. The monotonicity with respect to attribute subsets is proven, indicating that adding new features does not decrease the dependence degree;

(2) Conditional Information Entropy: This quantifies an HIS's uncertainty under given conditions. The monotonicity is demonstrated, showing that adding new features reduces or maintains the uncertainty level;

(3) Conditional Information Amount: This reflects the information contribution of conditional attributes to decision attributes. This measure decreases with an increasing attribute subset, ensuring more efficient feature selection;

(4) Conditional Discrimination Index: This assesses the ability of attribute subsets to discriminate between classes. The relationships among the aforementioned three UMs are explored, with an emphasis on their validity.

The monotonicity property of these induced UMs simplifies the process of adding or removing features in algorithms, enhancing computational efficiency and decision-making effectiveness. The proofs of these properties provide a solid theoretical foundation for the subsequent design and application of the attribute reduction algorithm.

4.2. Four UMs in an HIS using KNN-neighbourhood rough set model

4.2.1. The subsystem's dependence degree in an HIS

The dependency degree has values within the interval [0, 1]. A value closer to 1 indicates a higher level of certainty regarding the decision attribute d given a specific condition attribute set A, reflecting a reduced degree of uncertainty.

Definition 4.1. Consider an HIS (U,A,d), given $B \subseteq A$, $k \in \mathbb{N}$ *and* $\lambda \in [0,1]$. The positive region of B with respect to the decision attribute d is defined as

$$POS_{B,N}^{k,\lambda}(d) = \bigcup_{i=1}^{s} \underline{B_N^{k,\lambda}}(D_i).$$

Definition 4.2. Consider an HIS (U,A,d), given $B \subseteq A$, $k \in \mathbb{N}$ *and* $\lambda \in [0,1]$. The dependence degree of B on d is described as

$$\Gamma_{B,N}^{k,\lambda}(d) = \frac{1}{n}|POS_{B,N}^{k,\lambda}(d)|.$$

Proposition 4.3. *Consider an HIS (U,A,d), the following properties can be delineated:*

(1) $\forall\ B \subseteq A,\ \forall\ k \in \mathbb{N},\ \forall\ \lambda \in [0,1],\ \Gamma_{B,N}^{k,\lambda}(d) = \frac{1}{n}\sum_{i=1}^{s}|\underline{B_N^{k,\lambda}}(D_i)|;$

(2) $\forall\ B \subseteq A,\ \forall\ k \in \mathbb{N},\ \forall\ \lambda \in [0,1],\ 0 \leq \Gamma_{B,N}^{k,\lambda}(d) \leq 1;$

(3) *If* $B \subseteq C \subseteq A$, *then* $\forall\ k \in \mathbb{N},\ \forall\ \lambda \in [0,1],\ \Gamma_{B,N}^{k,\lambda}(d) \leq \Gamma_{C,N}^{k,\lambda}(d);$

(4) *If* $k_1 \leq k_2$, *then* $\forall\ B \subseteq A,\ \forall\ \lambda \in [0,1],\ \Gamma_{B,N}^{k_2,\lambda}(d) \leq \Gamma_{B,N}^{k_1,\lambda}(d);$

(5) *If* $\lambda_1 \leq \lambda_2$, *then* $\forall\ B \subseteq A,\ \forall\ k \in \mathbb{N},\ \Gamma_{B,N}^{k,\lambda_2}(d) \leq \Gamma_{B,N}^{k,\lambda_1}(d);$

(6) $\Gamma_{B,N}^{k,\lambda}(d) = 1 \ \Leftrightarrow\ \forall\ i,\ \underline{B_N^{k,\lambda}}(D_i) = D_i.$

Proof. Please see "Appendix". □

4.2.2. The subsystem's conditional information entropy in an HIS

Conditional information entropy captures the uncertain information within an HIS and is capable of handling hybrid attributes without the need for a discretisation process. Importantly, we define $0\log_2 0 = 0$ for the following calculations.

Definition 4.4. Consider an HIS (U, A, d), given $B \subseteq A$, $k \in \mathbb{N}$ *and* $\lambda \in [0,1]$. The definition of the (k,λ)-neighbourhood information entropy of B is as follows:

$$H_N^{k,\lambda}(B) = -\sum_{i=1}^{n} \frac{|B_N^{k,\lambda}(u_i)|}{n} \log_2 \frac{|B_N^{k,\lambda}(u_i)|}{n}. \tag{4.1}$$

Proposition 4.5. *Consider an HIS (U, A, d), given $B \subseteq A$, $k \in \mathbb{N}$ and $\lambda \in [0,1]$. Then,*

$$0 \le H_N^{k,\lambda}(B) \le n\log_2 n. \tag{4.2}$$

Furthermore, if $B^k = \triangle$, then $H_N^{k,\lambda}(B) = \log_2 n$; if $B^k = \delta$, then $H_N^{k,\lambda}(B) = 0$.

Proof. Please see "Appendix". □

Definition 4.6. Consider an HIS (U, A, d), given $B \subseteq A$, $k \in \mathbb{N}$ *and* $\lambda \in [0,1]$. Put

$$H_N^{k,\lambda}(d|B) = -\sum_{i=1}^{n}\sum_{j=1}^{s} \frac{|B_N^{k,\lambda}(u_i) \cap D_j|}{n} \log_2 \frac{|B_N^{k,\lambda}(u_i) \cap D_j|}{|B_N^{k,\lambda}(u_i)|}. \tag{4.3}$$

Then, $H_N^{k,\lambda}(d|B)$ is termed the (k,λ)-conditional neighbourhood information entropy of d to B in U.

Proposition 4.7. *Consider an HIS (U, A, d), given $k \in \mathbb{N}$ and $\lambda \in [0,1]$. If $B \subseteq C \subseteq A$, then*

$$H_N^{k,\lambda}(d|C) \le H_N^{k,\lambda}(d|B). \tag{4.4}$$

Proof. Please see "Appendix". □

Definition 4.8. Consider an HIS (U, A, d), given $B \subseteq A$, $k \in \mathbb{N}$ *and* $\lambda \in [0,1]$. The (k,λ)-joint neighbourhood information entropy of B and d is defined as

$$H_N^{k,\lambda}(B \cup d) = -\sum_{i=1}^{n}\sum_{j=1}^{s} \frac{|B_N^{k,\lambda}(u_i) \cap D_j|}{n} \log_2 \frac{|B_N^{k,\lambda}(u_i) \cap D_j|}{n}. \tag{4.5}$$

Proposition 4.9. *Consider an HIS (U, A, d), given $B \subseteq A$, $k \in \mathbb{N}$ and $\lambda \in [0,1]$. Then,*

$$H_N^{k,\lambda}(d|B) = H_N^{k,\lambda}(B \cup d) - H_N^{k,\lambda}(B). \tag{4.6}$$

Proof. Please see "Appendix". □

Proposition 4.10. *Consider an HIS (U, A, d), given $B \subseteq A$, $k \in \mathbb{N}$ and $\lambda \in [0,1]$. Then, $H_N^{k,\lambda}(d|B) \ge 0$.*

Proof. Please see "Appendix". □

4.2.3. The subsystem's conditional information amount in an HIS

The procedure of information acquisition is essentially the procedure of reducing information uncertainty.

Definition 4.11. Consider an HIS (U, A, d), given $B \subseteq A$, $k \in \mathbb{N}$ *and* $\lambda \in [0,1]$. Then, the (k,λ)-neighbourhood information amount of B is defined as

$$E_N^{k,\lambda}(B) = \sum_{i=1}^{n} \frac{|B_N^{k,\lambda}(u_i)|}{n} \frac{|U - B_N^{k,\lambda}(u_i)|}{n}. \tag{4.7}$$

Clearly, $E_N^{k,\lambda}(B) = \sum_{i=1}^{n} \frac{|B_N^{k,\lambda}(u_i)|}{n}(1 - \frac{|B_N^{k,\lambda}(u_i)|}{n})$.

Proposition 4.12. *Consider an HIS (U, A, d), given $B \subseteq A$, $k \in \mathbb{N}$ and $\lambda \in [0,1]$. Then,*

$$0 \le E_N^{k,\lambda}(B) \le n - 1. \tag{4.8}$$

Furthermore, if $B^k = \triangle$, then $E_N^{k,\lambda}(B) = 1 - \frac{1}{n}$ can be obtained;

Proof. Please see "Appendix". □

Definition 4.13. Consider an HIS (U, A, d), given $B \subseteq A$, $k \in \mathbb{N}$ *and* $\lambda \in [0,1]$. Define

$$E_N^{k,\lambda}(d|B) = \sum_{i=1}^{n}\sum_{j=1}^{s} \frac{|B_N^{k,\lambda}(u_i) \cap D_j|}{n} \frac{|B_N^{k,\lambda}(u_i) - D_j|}{n}. \tag{4.9}$$

Then, $E_N^{k,\lambda}(d|B)$ is referred to as the (k,λ)-neighbourhood conditional information amount of d to B in U.

Obviously, $E_N^{k,\lambda}(d|B) \ge 0$.

Proposition 4.14. *Consider an HIS (U, A, d), given $k \in \mathbb{N}$ and $\lambda \in [0,1]$. If $B \subseteq C \subseteq A$, then*

$$E_N^{k,\lambda}(d|C) \le E_N^{k,\lambda}(d|B). \tag{4.10}$$

Proof. Please see "Appendix". □

Definition 4.15. Consider an HIS (U, A, d), given $B \subseteq A$, $k \in \mathbb{N}$ *and* $\lambda \in [0,1]$. Then, the (k,λ)-joint neighbourhood information amount of B and d is defined as

$$E_N^{k,\lambda}(B \cup d) = \sum_{i=1}^{n}\sum_{j=1}^{s} \frac{|B_N^{k,\lambda}(u_i) \cap D_j|}{n}(1 - \frac{|B_N^{k,\lambda}(u_i) \cap D_j|}{n}). \tag{4.11}$$

Proposition 4.16. *Consider an HIS (U, A, d), given $B \subseteq A$, $k \in \mathbb{N}$ and $\lambda \in [0,1]$. Then,*

$$E_N^{k,\lambda}(d|B) = E_N^{k,\lambda}(B \cup d) - E_N^{k,\lambda}(B). \tag{4.12}$$

Proof. Please see "Appendix". □

4.2.4. The subsystem's conditional discrimination index in an HIS

Conditional discrimination index quantifies the uncertainty related to the discriminative capability of an attribute subset.

Definition 4.17. Consider an HIS (U, A, d), given $B \subseteq A$ and $k \in \mathbb{N}$. The k-neighbourhood discrimination index of B is defined as

$$DI_N^k(B) = log_2 \frac{|U \times U|}{|B^k|}. \tag{4.13}$$

Clearly, $DI_N^k(B) = log_2 \frac{n^2}{|B^k|}$.

Proposition 4.18. *Consider an HIS (U, A, d), given $B \subseteq A$ and $k \in \mathbb{N}$. Then,*

$$0 \le DI_N^k(B) \le log_2 n. \tag{4.14}$$

Moreover, if $B^k = \triangle$, then $DI_N^k(B) = log_2 n$; if $B^k = \delta$, then $DI_N^k(B) = 0$.

Proof. Please see "Appendix". □

Definition 4.19. Consider an HIS (U, A, d), given $B \subseteq A$, $k \in \mathbb{N}$ *and* $\lambda \in [0,1]$. Put

$$DI_N^{k,\lambda}(d|B) = log_2 \frac{|B^k|}{|B_N^{k,\lambda} \cap R_d|}. \tag{4.15}$$

Then, $DI_N^{k,\lambda}(d|B)$ is termed the (k,λ)-conditional neighbourhood discrimination index of d to B in U.

A lower conditional discrimination index value indicates a higher (k,λ)-consistency degree in an HIS.

Proposition 4.20. *Consider an HIS (U,A,d), given $B \subseteq A$, $k \in \mathbb{N}$ and $\lambda \in [0,1]$. Then,*

$$0 \le DI_N^{k,\lambda}(d|B) \le log_2 n. \tag{4.16}$$

Moreover, if $B_N^{k,\lambda} = \triangle$, then $DI_N^{k,\lambda}(d|B) = 0$; If $B_N^{k,\lambda} = \delta$, then $DI_N^{k,\lambda}(d|B) = log_2 \frac{n^2}{|R_d|}$.

Proof. Please see "Appendix". □

Definition 4.21. Consider an HIS (U,A,d), given $B \subseteq A$, $k \in \mathbb{N}$ and $\lambda \in [0,1]$. The (k,λ)-joint neighbourhood discrimination index of B and d is defined as

$$DI_N^{k,\lambda}(B \cup d) = log_2 \frac{|U \times U|}{|B_N^{k,\lambda} \cap R_d|}. \tag{4.17}$$

Proposition 4.22. *Consider an HIS (U,A,d), given $B \subseteq A$, $k \in \mathbb{N}$ and $\lambda \in [0,1]$. Then,*

$$DI_N^{k,\lambda}(d|B) = DI_N^{k,\lambda}(B \cup d) - DI_N^{k,\lambda}(B). \tag{4.18}$$

Proof. Clearly. □

4.2.5. Properties of UMs

Some properties of the aforementioned four UMs are described below.

Lemma 4.23. *Consider an HIS (U,A,d), given $B \subseteq A$, $k \in \mathbb{N}$ and $\lambda \in [0,1]$. If $B_N^{k,\lambda} \subseteq R_d$, then $\forall\, u \in U$, $\forall\, j$,*

$$B_N^{k,\lambda}(u) \cap D_j = \begin{cases} B_N^{k,\lambda}(u), & u \in D_j; \\ \emptyset, & u \notin D_j. \end{cases} \tag{4.19}$$

Proof. Please see "Appendix". □

Lemma 4.24. *Consider an HIS (U,A,d), given $B \subseteq A$, $k \in \mathbb{N}$ and $\lambda \in [0,1]$. If $B_N^{k,\lambda} \subseteq R_d$, then $\forall\, u \in U$,*

$$\sum_{j=1}^{s} \frac{|B_N^{k,\lambda}(u) \cap D_j|}{n}\left(1 - \frac{|B_N^{k,\lambda}(u) \cap D_j|}{n}\right) = \frac{|B_N^{k,\lambda}(u)|}{n}\left(1 - \frac{|B_N^{k,\lambda}(u)|}{n}\right),$$

$$\sum_{j=1}^{s} \frac{|B_N^{k,\lambda}(u) \cap D_j|}{n} \log_2 \frac{|B_N^{k,\lambda}(u) \cap D_j|}{n} = \frac{|B_N^{k,\lambda}(u)|}{n} \log_2 \frac{|B_N^{k,\lambda}(u)|}{n}.$$

Proof. Please see "Appendix". □

The subsequent Proposition 4.25 and Theorem 4.26 delineate the relationships among the dependence degree, conditional information entropy, and conditional discrimination index.

Proposition 4.25. *Consider an HIS (U,A,d), given $B \subseteq A$, $k \in \mathbb{N}$ and $\lambda \in [0,1]$. The statements below are equivalent:*

(1) $B_N^{k,\lambda} \subseteq R_d$;
(2) $\Gamma_{B,N}^{k,\lambda}(d) = 1$;
(3) $H_N^{k,\lambda}(d|B) = 0$;
(4) $DI_N^{k,\lambda}(d|B) = 0$.

Proof. Please see "Appendix". □

Theorem 4.26. *Consider an HIS (U,A,d), given $k \in \mathbb{N}$ and $\lambda \in [0,1]$. The statements below are equivalent:*

(1) (U,A,d) is (k,λ)-consistent;
(2) $A_N^{k,\lambda} \subseteq R_d$;
(3) $\Gamma_{A,N}^{k,\lambda}(d) = 1$;
(4) $H_N^{k,\lambda}(d|A) = 0$;
(5) $DI_N^{k,\lambda}(d|A) = 0$.

Proof. Please see "Appendix". □

Proposition 4.27. *Consider an HIS (U,A,d), given $B \subseteq A$, $k \in \mathbb{N}$ and $\lambda \in [0,1]$. If $B_N^{k,\lambda} \subseteq R_d$, then $E_N^{k,\lambda}(d|B) = 0$.*

Proof. Please see "Appendix". □

Theorem 4.28. *For an HIS (U,A,d), let $k \in \mathbb{N}$ and $\lambda \in [0,1]$. If (U,A,d) is (k,λ)-consistent, then $E_N^{k,\lambda}(d|A) = 0$.*

Proof. This can be proven using Propositions 3.8 and 4.27. □

5. Experiments

In this section, we perform an experimental analysis on seventeen datasets obtained from the Machine Learning Repository (UCI). This experimental analysis aims to validate the feasibility and assess the effectiveness of the aforementioned UMs.

5.1. Datasets and parameter setup

The performance specifications of the experimental setup include the 12th Gen Intel (R) Core (TM) i7-12700H processor platform with a frequency of 2.30 GHz and 16 GB of memory. The operating system used is Windows 11, and the code of experiment execution is carried out via Python 3.10. To determine the UM that has the optimal performance for handling an HIS, we perform numerical experiments on seventeen meticulously chosen datasets. The selection criteria for these datasets are based on their ability to visually emphasise the differences among the four UMs when represented graphically. Table 4 presents crucial details about these seventeen datasets.

In "5.9. Sensitivity analysis of k and λ" subsections, following the approach of Hu et al. [32] and Hu et al. [30], we varied the neighbourhood radius parameter λ from 0.1 to 0.9 (step size 0.1) and the number of nearest neighbours k from $0.05|U|$ to $0.5|U|$ (step size $0.05|U|$), where $|U|$ is the number of objects. Instead of tuning parameters for each dataset, we selected a combination that performed consistently well across all 17 datasets to ensure generalisability and fairness. Notably, in the following experiments, except for the "5.3. Dispersion analysis for robust UM selection" and "5.9. Sensitivity analysis of k and λ" subsections, the parameter setup for the remaining parts is as follows: k is set to $0.4|U|$, where $|U|$ is the number of objects in a dataset, and λ is set to 0.2.

5.2. Overview

Before formally starting the experiments, we first provide a summary of the content of this section. The following subsections are presented in this section.

(1) Dispersion analysis for robust UM selection: This subsection conducts a numerical experiment to identify the most robust UM among the four UMs;

(2) Relationships between the four UMs and attribute reduction in an HIS: Following the selection of the robust UM, a rigorous mathematical analysis explores the relationships between the four UMs and reducts;

Table 4
The details of seventeen datasets selected from UCI.

No.	Dataset	Abbr.	Objects	Attributes	Class	Type
1	Period Changer	PC	90	1177	2	Hybrid
2	Burst Header Packet	BHP	1075	21	3	Hybrid
3	Horse Colic	HC	368	27	2	Hybrid
4	Statlog (Heart)	SH	270	13	2	Hybrid
5	Heart failure clinical records	HFC	299	12	2	Hybrid
6	TCGA InfoWithGrade	TCGA	839	23	2	Hybrid
7	Cryotherapy Dataset	CD	90	6	2	Hybrid
8	Credit Approval	CA	690	15	2	Hybrid
9	Immunotherapy Dataset	ID	90	7	2	Hybrid
10	Iris	Iris	150	4	3	Numerical
11	Lymphography	LYM	148	18	8	Numerical
12	Climate	Climate	540	18	2	Numerical
13	Audit data	Audit	776	17	3	Numerical
14	Breast Tissue	BT	106	9	6	Numerical
15	Cervical Cancer Behavior Risk	CBB	72	19	2	Categorical
16	Higher Education Students Performance Evaluation Dataset	HES	145	31	8	Categorical
17	Congressional voting records	CVR	435	16	2	Categorical

(3) An attribute reduction algorithm based on the overlap degree in an HIS: Utilising the chosen UM and incorporating the concept of the overlap degree, this subsection presents the design of an attribute reduction algorithm NKRA;

(4) Comparison against state-of-the-art methods: A comparative study of the proposed algorithm against six existing state-of-the-art methods is conducted. The performance metrics for comparison include the F1 score and classification accuracy;

(5) Statistical tests on comparison results: This subsection details the statistical tests on the comparison results, including the Friedman test, the Bonferroni–Dunn test, and the Wilcoxon signed-rank test, to rigorously evaluate the performance of the NKRA;

(6) Comparison of computation time: The computational efficiency of the NKRA algorithm is compared with that of the state-of-the-art methods, highlighting the practicality of the approach;

(7) Sensitivity analysis of k and λ: This final subsection discusses how different combinations of k and λ affect the classification metrics (F1 score and classification accuracy), providing insights into the robustness and adaptability of the algorithm NKRA.

5.3. Dispersion analysis for robust UM selection

In this subsection, the best UM is chosen after performing a dispersion analysis. Propositions 4.7 and 4.14 reveal that, with fixed parameters k and λ, both $H_N^{k,\lambda}$ and $E_N^{k,\lambda}$ monotonically decrease as the cardinality of the attribute subset increases. On the other hand, $\Gamma_N^{k,\lambda}$ monotonically increases with increasing cardinality of the attribute subset according to Proposition 4.3. Therefore, $\Gamma_N^{k,\lambda}$, $H_N^{k,\lambda}$, and $E_N^{k,\lambda}$ are suitable for quantifying the uncertainty of an HIS. The definition of UM $DI_N^{k,\lambda}$ is defined based on the uncertainty of an HIS. In summary, all four uncertainty measures can effectively gauge the uncertainty within an HIS.

The variability of the proposed UMs is assessed through the utilisation of standard deviation. Let $X = \{x_1, \ldots, x_n\}$ be a dataset. Then, the standard deviation coefficient of X is defined as

$$CV(X) = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \frac{1}{n}\sum_{i=1}^{n} x_i)^2}}{\frac{1}{n}\sum_{i=1}^{n} x_i} \tag{5.1}$$

First, for each dataset, let $B_j = \{a_{1\times unit}, \ldots, a_{j\times unit}\}$ $(j = 1, 2, \ldots, 5)$ be defined. "$Unit$" here refers to the result of dividing the total number of conditional attributes by five via integer division. The sets of measurements for each UM on each dataset can be expressed as follows:

$$X_{\Gamma_N^{k,\lambda}}(d|each\ dataset) = \{\Gamma_{B_1,N}^{k,\lambda}(d), \Gamma_{B_2,N}^{k,\lambda}(d), \Gamma_{B_3,N}^{k,\lambda}(d), \Gamma_{B_4,N}^{k,\lambda}(d), \Gamma_{B_5}^{k,\lambda}(d)\},$$

$$X_{H_N^{k,\lambda}}(d|each\ dataset) = \{H_N^{k,\lambda}(d|B_1), H_N^{k,\lambda}(d|B_2), H_N^{k,\lambda}(d|B_3), H_N^{k,\lambda}(d|B_4), H_N^{k,\lambda}(d|B_5)\},$$

$$X_{E_N^{k,\lambda}}(d|each\ dataset) = \{E_N^{k,\lambda}(d|B_1), E_N^{k,\lambda}(d|B_2), E_N^{k,\lambda}(d|B_3), E_N^{k,\lambda}(d|B_4), E_N^{k,\lambda}(d|B_5)\},$$

$$X_{DI_N^{k,\lambda}}(d|each\ dataset) = \{DI_N^{k,\lambda}(d|B_1), DI_N^{k,\lambda}(d|B_2), DI_N^{k,\lambda}(d|B_3), DI_N^{k,\lambda}(d|B_4), DI_N^{k,\lambda}(d|B_5)\}.$$

We subsequently compute the CV values for the measurement sets on each dataset, as illustrated in Fig. 3.

From Fig. 3, we can conclude that for two fixed parameters k and λ, $\Gamma_N^{k,\lambda}$ has the minimum CV value in most datasets. This leads to a reasonable inference that $\Gamma_N^{k,\lambda}$ performs optimally in quantifying the uncertainty of an HIS.

Building upon the analysis provided, $\Gamma_{B,N}^{k,\lambda}$ stands out as an optimal UM for capturing the uncertainty within an HIS. The primary objective of the next subsection is to design an attribute reduction algorithm for an HIS based on $\Gamma_{B,N}^{k,\lambda}$ and evaluate its effectiveness.

5.4. Relationships between the four UMs and attribute reduction in an HIS

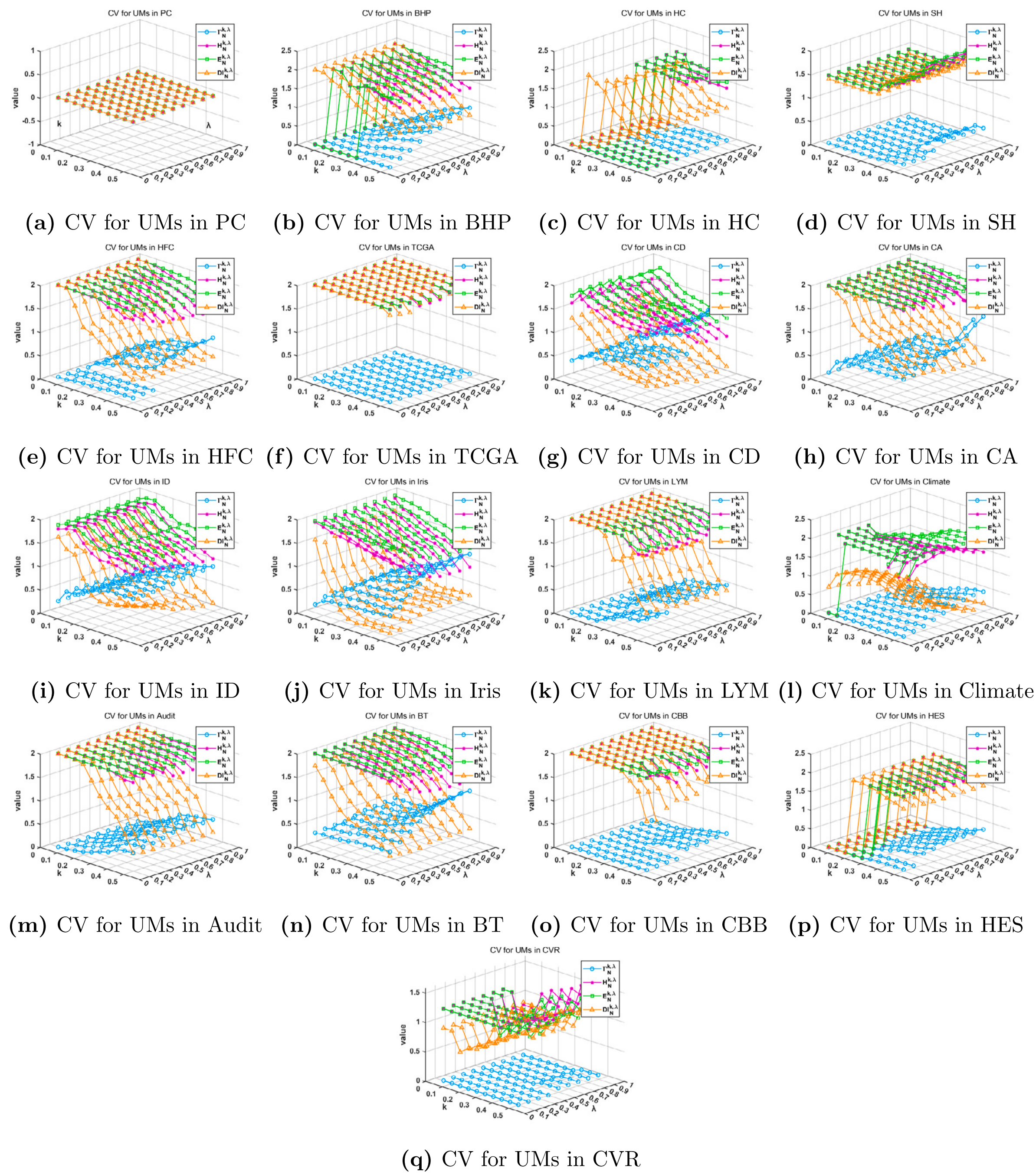
Consider an HIS (U, A, d), given $B \subseteq A$, $k \in \mathbb{N}$ *and* $\lambda \in [0,1]$. B is considered a coordinate subset of A, if $\Gamma_{B,N}^{k,\lambda}(d) = \Gamma_{A,N}^{k,\lambda}(d)$. $co_d^{k,\lambda}(A)$ represents the set of all coordinate subsets. B is a reduct of A; if $B \in co_d^{k,\lambda}(A)$ and $\forall\, a \in B$, $B - \{a\} \notin co_d^{k,\lambda}(A)$. $red_d^{k,\lambda}(A)$ represents the set of all reducts of A.

Theorem 5.1. *Consider an HIS (U, A, d), given $B \subseteq A$, $k \in \mathbb{N}$ and $\lambda \in [0,1]$. The statements below are equivalent:*
(1) $B \in co_d^{k,\lambda}(A)$;
(2) $POS_{B,N}^{k,\lambda}(d) = POS_{A,N}^{k,\lambda}(d)$;
(3) $\Gamma_{B,N}^{k,\lambda}(d) = \Gamma_{A,N}^{k,\lambda}(d)$.

Proof. Clearly. □

Corollary 5.2. *Consider an HIS (U, A, d), given $B \subseteq A$, $k \in \mathbb{N}$ and $\lambda \in [0,1]$. The statements below are equivalent:*
(1) $B \in red_d^{k,\lambda}(A)$;
(2) $POS_{B,N}^{k,\lambda}(d) = POS_{A,N}^{k,\lambda}(d)$ *and* $\forall\, a \in B$, $POS_{B-\{a\},N}^{k,\lambda}(d) \neq POS_{A,N}^{k,\lambda}(d)$;
(3) $\Gamma_{B,N}^{k,\lambda}(d) = \Gamma_{A,N}^{k,\lambda}(d)$ *and* $\forall\, a \in B$, $\Gamma_{B-\{a\},N}^{k,\lambda}(d) \neq \Gamma_{A,N}^{k,\lambda}(d)$.

Proof. It can be proven according to Theorem 5.1. □



(a) CV for UMs in PC (b) CV for UMs in BHP (c) CV for UMs in HC (d) CV for UMs in SH

(e) CV for UMs in HFC (f) CV for UMs in TCGA (g) CV for UMs in CD (h) CV for UMs in CA

(i) CV for UMs in ID (j) CV for UMs in Iris (k) CV for UMs in LYM (l) CV for UMs in Climate

(m) CV for UMs in Audit (n) CV for UMs in BT (o) CV for UMs in CBB (p) CV for UMs in HES

(q) CV for UMs in CVR

Fig. 3. Coefficient of variation (CV) for UMs across different datasets.

The connections between the conditional information entropy and conditional discrimination index are illustrated in the following Theorem 5.3 and Corollary 5.4.

Theorem 5.3. *Assuming that (U, A, d) is a (k, λ)-consistent HIS, with $k \in \mathbb{N}$ and $\lambda \in [0, 1]$, the following statements are equivalent:*
(1) $B \in co_d^{k,\lambda}(A)$;
(2) $H_N^{k,\lambda}(d|B) = H_N^{k,\lambda}(d|A)$;
(3) $DI_N^{k,\lambda}(d|B) = DI_N^{k,\lambda}(d|A)$.

Proof. Please see "Appendix". □

Corollary 5.4. *Given $k \in \mathbb{N}$ and $\lambda \in [0, 1]$. Suppose that (U, A, d) is an (k, λ)-consistent HIS. The statements below are equivalent:*
(1) $B \in red_d^{k,\lambda}(A)$;
(2) $H_N^{k,\lambda}(d|B) = H_N^{k,\lambda}(d|A)$ *and* $\forall\, a \in B$, $H_N^{k,\lambda}(d|B - \{a\}) \neq H_N^{k,\lambda}(d|A)$;
(3) $DI_N^{k,\lambda}(d|B) = DI_N^{k,\lambda}(d|A)$ *and* $\forall\, a \in B$, $DI_N^{k,\lambda}(d|B - \{a\}) \neq DI_N^{k,\lambda}(d|A)$.

Proof. This can be demonstrated via Theorem 5.3. □

Theorem 5.5. *Given $k \in \mathbb{N}$ and $\lambda \in [0, 1]$. Assume that (U, A, d) is an (k, λ)-consistent HIS. If $B \in co_d^{k,\lambda}(A)$, then $E_N^{k,\lambda}(d|B) = E_N^{k,\lambda}(d|A)$.*

Proof. Assume that $B \in co_d^{k,\lambda}(A)$. According to Theorem 5.1, $\Gamma_{B,N}^{k,\lambda}(d) = \Gamma_{A,N}^{k,\lambda}(d)$. Following a similar proof strategy as in Theorem 5.3, it can be established that $B_N^{k,\lambda} \subseteq R_d$. According to Proposition 4.27,

$E_N^{k,\lambda}(d|B) = 0$. Considering that (U, A, d) is (k, λ)-consistent, and according to Theorem 4.28, $E_N^{k,\lambda}(d|A) = 0$. Therefore, $E_N^{k,\lambda}(d|B) = E_N^{k,\lambda}(d|A)$. □

Corollary 5.6. *Let (U, A, d) be an (k, λ)-consistent HIS. Given that $k \in \mathbb{N}$ and $\lambda \in [0, 1]$. Then,*

$$B \in red_d^{k,\lambda}(A) \Leftrightarrow E_N^{k,\lambda}(d|B) = E_N^{k,\lambda}(d|A) \text{ and } \forall\, a \in B,$$
$$E_N^{k,\lambda}(d|B - \{a\}) \neq E_N^{k,\lambda}(d|A).$$

Proof. It can be derived from Theorem 5.5. □

5.5. An attribute reduction algorithm based on the overlap degree in an HIS

In this subsection, an algorithm is designed to search for an $\Gamma_N^{k,\lambda}$-based reduction set. According to Proposition 4.3, removing a conditional attribute from a subset of attributes results in a monotonic decrease in the value of $\Gamma_N^{k,\lambda}$. This implies that the reduction in this value can be utilised to gauge the redundancy of the removed attribute. In other words, we can devise a reduction algorithm via $\Gamma_N^{k,\lambda}$. Since $\Gamma_N^{k,\lambda}$ have the property in Corollary 5.2, we can design the $\Gamma_N^{k,\lambda}$-based attribute reduction algorithm (denoted as NKRA).

Computing KNN-information granules is a very time-consuming task, and the current heuristic search strategy has low efficiency. To address these computational challenges, we aim to increase the reduction efficiency by introducing a novel perspective on the search strategy. We plan to define measures for pre-evaluating the importance of individual attributes and pre-sort the attributes based on their importance levels. Starting with the attribute of least importance, we employ the dependency degree to assess iteratively whether each attribute can be removed, thereby optimising the reduction process.

Definition 5.7. For an HIS (U, A, d) with $a \in A$. $\forall\, i$, denote

$$a \in A^c,\ a(D_i) = \{a(u) : u \in D_i\};$$

$$a \in A^n,\ m_{D_i}^a = min\{a(u) : u \in D_i\},\ M_{D_i}^a = max\{a(u) : u \in D_i\}.$$

Then, the coincidence degree between $a(D_i)$ and $a(D_j)$ $(i \neq j)$ is represented as

$$CD_a(D_i, D_j) = CD(a(D_i), a(D_j)) =$$

$$\begin{cases} \frac{|a(D_i)\cap a(D_j)|}{|a(D_i)\cup a(D_j)|}, & a \in A^c \\ \frac{|[m_{D_i}^a, M_{D_i}^a]\cap[m_{D_j}^a, M_{D_j}^a]|}{|[m_{D_i}^a, M_{D_i}^a]\cup[m_{D_j}^a, M_{D_j}^a]|}, & a \in A^n \end{cases} \tag{5.2}$$

Definition 5.8. Consider an HIS (U, A, d) with $a \in A$. The overlap degree of the information values under a is represented as

$$OD_a = \sum_{1\le i,j\le s,\ i\neq j} CD_a(D_i, D_j). \tag{5.3}$$

Definition 5.9. Suppose that (U, A, d) is an HIS. Put $A = \{a_1, a_2, \dots, a_m\}$. If $A = \{a_1', a_2', \dots, a_m'\}$ and

$$OD_{a_1'} \ge OD_{a_2'} \ge \cdots \ge OD_{a_m'}, \tag{5.4}$$

then $\{a_1', a_2', \dots, a_m'\}$ is termed $SORT(A)$, i.e.,

$$SORT(A) = \{a_1', a_2', \dots, a_m'\}. \tag{5.5}$$

Example 5.10 (*Building Upon the Illustration in Example 2.2*).

As per Definitions 5.7–5.9, we have

$$OD_{a_1} = 1.3333,\ OD_{a_2} = 2,\ OD_{a_3} = 0,\ OD_{a_4} = 1.$$

Then,

$$SORT(A) = \{a_2, a_1, a_4, a_3\}.$$

Algorithm 1: Attribute reduction algorithm based on $\Gamma_N^{k,\lambda}$ in an HIS (NKRA).

Input: An HIS (U, A, d), and two parameters $\lambda \in [0, 1], k \in [1, |U|]$;
Output: One reduct B;
1: **for** any $a \in A$ **do**
2: Calculate the distance between two information values $\rho(a(u), a(v))$ under the attribute a by Definition 2.5;
3: Compute the tolerance class $N_{\{a\}}^{\lambda}(u)$ of each object u under a single attribute a according to Definition 2.9;
4: Compute the class $a^k(u)$ of each object u under a single attribute a according to Definition 3.1;
5: Compute OD_a according to Definition 5.8;
6: **end for**
7: Attributes are arranged in descending order based on the value of OD_a, and the outcome is labelled as SORT (A).
8: Initialise: $B \leftarrow SORT(A), start \leftarrow 1$; /* where attributes in SORT(A) are ordered. */
9: Calculate $\Gamma_{A,N}^{k,\lambda}(d)$ according to Definition 4.2;
10: **while** $start = 1$ **do**
11: $start \leftarrow 0$;
12: **for** each $a \in B$ **do**
13: Calculate $\Gamma_{B-\{a\},N}^{k,\lambda}(d)$according to Definition 4.2;
14: **if** $\Gamma_{B-\{a\},N}^{k,\lambda}(d) = \Gamma_{A,N}^{k,\lambda}(d)$ **then**
15: $start \leftarrow 1$;
16: $B \leftarrow B - \{a\}$;
17: **end if**
18: **if** $\Gamma_{B-\{a\},N}^{k,\lambda}(d) \neq \Gamma_{A,N}^{k,\lambda}(d)$ **then**
19: break
20: **end if**
21: **end for**
22: **end while**
23: **return** B.

The NKRA algorithm employs a heuristic approach in the attribute reduction process. The attributes are systematically removed from the complete set based on their $\Gamma_N^{k,\lambda}$ values in the attribute reduction process of the algorithm NKRA. The attributes in SORT(A) can be removed one by one until the value of the calculated function value changes; at this point, the algorithm NKRA concludes, and a reduct is obtained.

The algorithm procedure is detailed as follows. For an HIS with s decision classes, n objects, and m conditional attributes, where s is at most equal to n. In Steps 1–7, we calculate the distances between any two objects within each attribute using Definition 2.5. This enables the determination of the tolerance class $N_{\{a\}}^{\lambda}(u)$. With Definition 3.1, we obtain the class $a^k(u)$ for each attribute. Importantly, the calculation of the distance between any two objects in Definition 2.5 involves a summation operation. With Definition 5.8, we calculate the overlap degree of the information values within each attribute. The attributes are then arranged in descending order based on the overlap degree of the information values, and the outcome is labelled SORT (A). In Step 8, we set the initial values of B as $SORT(A)$ and $start$ as 1. In Step 9, we compute $\Gamma_{A,N}^{k,\lambda}$. In Steps 10–22, we enter a while loop when "start = 1". The while loop concludes when there is a change in the dependence degree value; otherwise, if no update is made to B, then the complexity of Steps 10–22 is $O(1)$ time. It is crucial to emphasise that the computational complexity of the dependence degree, as per Definition 4.2, is $O(ns)$ time. If B undergoes an update, then in the slowest scenario, the while loop iterates m times, so the computational complexity of Steps 10–22 is $O(mns)$ time. The algorithm NKRA outputs a reduct B in Step 23.

Following the analysis of the algorithm NKRA provided above, Steps 1–6 require $O(mn^2s + mn^2 + mkn^2 + ms^2)$ time. Step 7 takes $O(m)$ time.

Table 5
A detailed overview of the computational complexity of the algorithm NKRA.

Steps	Computational time
1–6	$O(mn^2s + mn^2 + mkn^2 + ms^2)$
7	$O(m)$
8	$O(1)$
9	$O(ns)$
10–22	$O(mns)$
Total	$O(mn^2s + mn^2 + mkn^2 + ms^2 + m + ns + mns)$

Table 6
Time complexity comparison of algorithms.

Algorithm	Time complexity
NKRA	$O(mn^2s + mn^2 + mkn^2 + ms^2 + m + ns + mns)$
Imp_2^D	$O(mn^2 + m^2n + mns)$
Imp_3^D	$O(mn^2 + mn + ms)$
BFB	$O(mn^2 + m^3 + m^2s)$
BFP	$O(mn^2 + m^3 + m^2s)$
ES	$O(mn^2s + mns + mn + n^2 + ns)$
ULGR	$O(mn^2 + m^2 + mn)$

The time complexity of Step 8 is $O(1)$ time. Step 9 requires $O(ns)$ time. Steps 10–22 require $O(mns)$ time. Therefore, the NKRA algorithm has a computational complexity of $O(mn^2s + mn^2 + mkn^2 + ms^2 + m + ns + mns)$ time.

To better illustrate, Table 5 presents a detailed overview of the computational complexity of the algorithm NKRA.

5.6. Comparison against state-of-the-art methods

BFB [29], BFP [29], Imp_2^D [25], Imp_3^D [25], ES [34] and ULGR [35] are six existing advanced attribute reduction algorithms that are compared with the algorithm NKRA to validate its effectiveness. A succinct presentation of the six advanced attribute reduction algorithms is provided below.

(1) BFB: The attribute reduction algorithm employs a novel distance function designed for handling incomplete and hybrid data, and integrates fuzzy belief functions with Dempster–Shafer evidence theory to establish a new theoretical framework. This approach demonstrates robust noise resistance without the need for parameter determination.

(2) BFP: An attribute reduction algorithm for hybrid data utilising fuzzy plausibility functions. This algorithm exhibits strong noise resistance without requiring parameter determination.

(3) Imp_2^D: A heuristic algorithm for partially labelled heterogeneous data based on the information-theoretic neighbourhood rough set model. This algorithm is designed to handle both complete and incomplete hybrid datasets.

(4) Imp_3^D: This algorithm is also a heuristic algorithm for partially labelled heterogeneous data based on the information-theoretic neighbourhood rough set model. This algorithm is designed to handle both complete and incomplete hybrid datasets.

(5) ES: The attribute reduction algorithm follows an ensemble selector strategy based on N-rough set model. This algorithm is designed to handle only complete hybrid datasets. In the case of incomplete datasets, where values are missing for categorical and numerical attributes, we employ a replacement strategy using patterns and means from the available data, as suggested by [36].

(6) ULGR: This algorithm is also a heuristic algorithm for heterogeneous data based on gain ratio in information entropy. This algorithm is designed to handle only complete hybrid datasets. In the case of incomplete datasets, where values are missing for categorical and numerical attributes, we employ a replacement strategy using patterns and means from the available data, as suggested by [36].

The distinctions among the above six algorithms and the algorithm NKRA are elucidated below:

(a) Diverse data processing methods are employed.

(b) Various UMs are utilised in these algorithms.

(c) All these algorithms can handle hybrid data.

In terms of the computational complexity, Table 6 shows that the computational complexity of the proposed algorithm in this paper is close to n^3, which corresponds to the worst-case scenario (i.e., when the corresponding for-loops run to completion). However, based on experimental results, the algorithm performs comparably to other algorithms in most cases in terms of execution time.

In terms of theory, this paper employs KNN-neighbourhood rough set model and constructs four corresponding UMs based on this model: the dependence degree, conditional information entropy, conditional information amount, and conditional discrimination index. Through dispersion analysis, the most stable UM, the dependence degree, was selected, which was then used to construct the attribute reduction algorithm. The UM selected for this algorithm is consistent with KNN-neighbourhood rough set model and also carries the key theoretical advantage of KNN-neighbourhood rough set model: the ability to freely select the corresponding k and λ to effectively identify isolated objects and surrounded objects within a dataset.

To validate the effectiveness of the algorithm NKRA, we employ AdaBoost classifier (n estimators = 50, learning rate = 1.0), DT classifier (max depth = 3) and KNN classifier (neighbours = 4) to assess the performance of these six algorithms. We adapt seventeen different datasets, their basic information is shown in Table 4. The criteria for selecting the datasets in this paper include the following: (a) datasets are sourced from UCI; (b) datasets are advisable for classification tasks; (c) The datasets include categorical, numerical, and hybrid datasets to assess the classification performance of the algorithm NKRA on different types of datasets.

Then the algorithm NKRA is utilised on the selected seventeen datasets. Three kinds of classifiers (AdaBoost, DT and KNN) are used to calculate the F1 score and classification accuracy of the reduced set of the algorithm NKRA for every dataset. The six state-of-the-art algorithms use their optimal parameters (if available). The algorithm NKRA maintains fixed parameters throughout the experiments: k is set to $0.4|U|$, where $|U|$ is the number of objects in a dataset, and λ is set to 0.2. (The reason why we choose this parameter combination is in the "Datasets and parameter setup")

Tables 7–12 present the following results: (1) The classification metrics (classification accuracy and F1 score) of the reduction sets produced by the algorithm NKRA is greater than or equal to its original dataset in a total of 76 cases. (2) In the majority of original HISs, the algorithm NKRA outperforms or equals four or more of the other six algorithms. (3) The average F1 score of the algorithm NKRA across all datasets is higher than five or more of the other six algorithms. (4) The average classification accuracy of the algorithm NKRA across all datasets is higher than (or equal to) the other six algorithms.

Table 13 presents the win/draw/loss results of the NKRA algorithm compared with those of the other six algorithms under the F1 score. In 51 comparison cases, the NKRA algorithm outperforms BFP, Imp_2^D, Imp_3^D, BFP, ES and ULGR by 28, 49, 51, 28, 12 and 16 times, respectively. Additionally, the algorithm NKRA draws 16, 0, 0, 16, 25 and 19 times against BFP, Imp_2^D, Imp_3^D, BFP, ES and ULGR, respectively.

Table 14 presents the win/draw/loss results of the NKRA algorithm compared with those of the other six algorithms in terms of classification accuracy. In 51 comparison cases, the algorithm NKRA performs 29, 44, 43, 29, 18 and 20 times better than BFP, Imp_2^D, Imp_3^D, BFP, ES and ULGR, respectively. Additionally, the algorithm NKRA draws 13, 2, 0, 13, 25 and 15 times against BFP, Imp_2^D, Imp_3^D, BFP, ES and ULGR, respectively.

These results indicate that the NKRA algorithm is superior to or equal to the other six advanced algorithms in the majority of cases, demonstrating its significant advantages and strong practicality.

Table 7
F1 score of raw data and reduced data with AdaBoost.

Data sets	Raw data	NKRA	BFP	Imp_2^D	Imp_3^D	BFB	ES	ULGR
PC	**1.0**	**1.0**	**1.0**	0.81	0.71	**1.0**	0.8	0.87
BHP	**1.0**	**1.0**	**1.0**	0.93	0.57	**1.0**	**1.0**	**1.0**
HC	**0.94**	0.93	0.82	0.58	0.83	0.82	0.8	0.81
SH	**0.88**	0.86	0.82	0.77	0.65	0.82	0.87	**0.88**
HFC	**0.9**	**0.9**	0.88	0.68	0.73	0.88	**0.9**	0.89
TCGA	**0.87**	0.76	0.74	0.46	0.74	0.74	**0.87**	**0.87**
CD	**1.0**	**1.0**	**1.0**	0.87	0.78	**1.0**	**1.0**	**1.0**
CA	**0.88**	**0.88**	0.79	0.6	0.74	0.79	**0.88**	**0.88**
ID	**0.97**	**0.97**	**0.97**	0.79	0.82	**0.97**	**0.97**	**0.97**
Iris	**0.97**	**0.97**	**0.97**	0.96	0.59	**0.97**	**0.97**	**0.97**
LYM	0.39	0.39	0.38	0.25	0.23	0.38	0.39	**0.4**
Climate	**0.98**	0.96	**0.98**	0.88	0.88	**0.98**	0.97	0.96
Audit	**1.0**	**1.0**	**1.0**	0.85	0.71	**1.0**	**1.0**	0.84
BT	0.66	0.66	**0.74**	0.27	0.36	**0.74**	0.66	0.65
CBB	**1.0**	0.75	0.74	0.74	0.74	0.74	0.94	0.93
HES	**0.18**	**0.18**	0.14	0.14	0.14	0.14	**0.18**	**0.18**
CVR	**0.96**	**0.96**	0.92	0.68	0.68	0.92	**0.96**	**0.96**
Average	**0.86**	0.83	0.82	0.66	0.64	0.82	0.83	0.83

Table 8
F1 score of raw data and reduced data with DT (max depth=3).

Data sets	Raw data	NKRA	BFP	Imp_2^D	Imp_3^D	BFB	ES	ULGR
PC	0.87	**0.89**	0.87	0.79	0.69	0.87	0.75	0.75
BHP	**1.0**	**1.0**	**1.0**	0.94	0.47	**1.0**	**1.0**	**1.0**
HC	**0.89**	**0.89**	0.8	0.57	0.78	0.8	0.63	0.63
SH	**0.86**	0.83	0.79	0.77	0.61	0.79	**0.86**	**0.86**
HFC	0.87	0.87	0.87	0.69	0.72	0.87	0.87	**0.88**
TCGA	**0.87**	0.76	0.74	0.46	0.74	0.74	**0.87**	**0.87**
CD	**0.94**	**0.94**	**0.94**	0.86	0.66	**0.94**	**0.94**	**0.94**
CA	**0.86**	**0.86**	0.78	0.58	0.73	0.78	**0.86**	**0.86**
ID	**0.94**	**0.94**	**0.94**	0.72	0.74	**0.94**	**0.94**	**0.94**
Iris	**0.97**	**0.97**	**0.97**	**0.97**	0.56	**0.97**	**0.97**	**0.97**
LYM	0.42	0.42	**0.43**	0.35	0.24	**0.43**	0.42	0.42
Climate	**0.94**	**0.94**	**0.94**	0.88	0.88	**0.94**	**0.94**	**0.94**
Audit	**1.0**	**1.0**	0.97	0.85	0.72	0.97	0.97	0.84
BT	**0.6**	**0.6**	**0.6**	0.55	0.53	**0.6**	**0.6**	**0.6**
CBB	**0.99**	0.75	0.72	0.72	0.72	0.72	0.88	0.9
HES	**0.42**	0.26	0.14	0.14	0.14	0.14	**0.42**	0.35
CVR	**0.96**	**0.96**	0.9	0.68	0.68	0.9	**0.96**	**0.96**
Average	**0.85**	0.82	0.79	0.68	0.62	0.79	0.82	0.81

Table 9
F1 score of raw data and reduced data with KNN.

Data sets	Raw data	NKRA	BFP	Imp_2^D	Imp_3^D	BFB	ES	ULGR
PC	0.75	0.76	0.74	**0.79**	0.68	0.74	0.78	0.78
BHP	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
HC	0.86	**0.87**	0.79	0.7	0.83	0.79	0.8	0.81
SH	**0.77**	**0.77**	**0.77**	0.76	0.71	**0.77**	0.74	**0.77**
HFC	0.66	0.66	**0.7**	0.65	0.63	**0.7**	0.66	0.66
TCGA	0.86	0.78	0.77	0.45	0.77	0.77	0.86	**0.87**
CD	**0.92**	**0.92**	**0.92**	0.84	0.78	**0.92**	**0.92**	**0.92**
CA	**0.79**	**0.79**	**0.79**	0.72	0.28	**0.79**	**0.79**	**0.79**
ID	**0.8**	**0.8**	**0.8**	0.78	0.78	**0.8**	**0.8**	**0.8**
Iris	0.96	0.96	0.96	**0.97**	0.44	0.96	0.96	**0.97**
LYM	**0.55**	**0.55**	0.5	0.33	0.14	0.5	**0.55**	0.53
Climate	0.94	0.94	0.94	0.91	0.91	0.94	**0.96**	**0.96**
Audit	**0.99**	**0.99**	**0.99**	0.85	0.67	**0.99**	**0.99**	0.84
BT	**0.78**	**0.78**	**0.78**	0.7	0.6	**0.78**	**0.78**	0.76
CBB	**0.91**	0.8	0.13	0.13	0.13	0.13	0.81	0.9
HES	0.5	0.36	0.14	0.14	0.14	0.14	**0.52**	0.48
CVR	**0.96**	**0.96**	0.94	0.47	0.47	0.94	0.95	**0.96**
Average	**0.82**	0.81	0.75	0.66	0.58	0.75	0.82	0.81

5.7. Statistical tests on comparison results

To systematically assess the differences in classification accuracy among the seven algorithms, we employ the Friedman test [37]. This test is well suited for determining whether there exists a significant difference exists in the performance of various comparison algorithms. Prior to conducting the Friedman test, we organised and assigned numerical ranks to the classification accuracies of all the datasets in descending order. The numbering schemes used are 1, 2, ⋯, and so on. For example, if, in a specific dataset, the algorithm NKRA achieves the highest performance among the seven algorithms, it is assigned the number 1. In cases where two algorithms exhibit the same classification accuracy, they share the same ranking, which is the average of their positions in the order. The Friedman test can be formulated as follows:

Table 10
Classification accuracy of raw data and reduced data with AdaBoost.

Data sets	Raw data	NKRA	BFP	Imp_2^D	Imp_3^D	BFB	ES	ULGR
PC	57.78	70.0	**73.33**	61.11	71.11	**73.33**	66.67	66.67
BHP	**100.0**	**100.0**	**100.0**	93.49	53.3	**100.0**	**100.0**	**100.0**
HC	**88.85**	88.84	79.88	65.5	83.14	79.88	67.36	68.47
SH	**81.85**	77.78	73.33	77.78	52.22	73.33	80.74	**81.85**
HFC	68.9	71.23	63.9	68.89	**71.24**	63.9	68.9	69.56
TCGA	**86.77**	74.73	70.92	59.24	70.92	70.92	**86.77**	**86.77**
CD	88.89	88.89	**90.0**	80.0	64.44	**90.0**	88.89	**90.0**
CA	**84.49**	**84.49**	76.81	57.97	72.17	76.81	**84.49**	**84.49**
ID	**82.22**	**82.22**	**82.22**	77.78	68.89	**82.22**	**82.22**	**82.22**
Iris	**94.67**	**94.67**	**94.67**	94.0	54.0	**94.67**	**94.67**	**94.67**
LYM	35.71	35.71	36.95	25.05	**38.57**	36.95	35.71	35.05
Climate	94.44	91.85	94.44	91.3	91.11	94.44	**95.56**	92.22
Audit	**99.87**	**99.87**	99.61	78.72	70.58	99.61	93.03	84.24
BT	58.36	58.36	**63.27**	35.64	46.73	**63.27**	58.36	58.64
CBB	**84.82**	73.39	70.71	70.71	70.71	70.71	73.39	70.89
HES	**29.86**	27.71	22.0	22.0	22.0	22.0	29.19	27.71
CVR	**95.62**	**95.62**	90.1	67.37	67.37	90.1	95.4	95.4
Average	**78.42**	77.37	75.42	66.27	62.85	75.42	76.55	75.81

Table 11
Classification accuracy of raw data and reduced data with DT (max depth=3).

Data sets	Raw data	NKRA	BFP	Imp_2^D	Imp_3^D	BFB	ES	ULGR
PC	65.56	67.78	66.67	65.56	**71.11**	66.67	67.78	64.44
BHP	**100.0**	**100.0**	**100.0**	93.95	50.8	**100.0**	**100.0**	**100.0**
HC	**87.5**	**87.5**	77.43	65.22	78.78	77.43	66.86	65.24
SH	77.78	77.04	65.93	77.78	52.22	65.93	**80.0**	77.78
HFC	73.23	73.23	73.23	67.55	73.24	73.23	73.23	**74.9**
TCGA	86.89	74.74	70.92	59.24	70.92	70.92	**87.01**	86.89
CD	**92.22**	**92.22**	**92.22**	82.22	63.33	**92.22**	**92.22**	**92.22**
CA	**79.71**	**79.71**	73.62	56.81	71.74	73.62	**79.71**	**79.71**
ID	**83.33**	82.22	82.22	76.67	77.78	82.22	82.22	**83.33**
Iris	**96.0**	**96.0**	**96.0**	95.33	55.33	**96.0**	**96.0**	**96.0**
LYM	38.57	38.57	39.95	**40.48**	39.24	39.95	38.57	39.24
Climate	91.11	92.22	91.11	90.74	91.11	91.11	92.04	**92.96**
Audit	**99.87**	**99.87**	97.3	73.6	67.5	97.3	97.17	84.24
BT	55.73	55.73	55.73	**59.45**	43.73	55.73	54.82	54.82
CBB	**81.96**	73.39	70.71	70.71	70.71	70.71	70.54	69.11
HES	29.71	**34.33**	22.0	22.0	22.0	22.0	29.71	24.81
CVR	**95.85**	**95.85**	86.63	67.37	67.37	86.63	**95.85**	**95.85**
Average	**78.53**	77.67	74.22	68.51	62.76	74.22	76.69	75.38

Table 12
Classification accuracy of raw data and reduced data with KNN.

Data sets	Raw data	NKRA	BFP	Imp_2^D	Imp_3^D	BFB	ES	ULGR
PC	56.67	62.22	57.78	67.78	**72.22**	57.78	60.0	61.11
BHP	**100.0**	**100.0**	**100.0**	99.35	99.35	**100.0**	**100.0**	**100.0**
HC	72.24	75.24	64.12	51.39	**82.05**	64.12	66.06	68.78
SH	63.7	63.33	62.22	**74.07**	62.59	62.22	62.22	63.7
HFC	65.53	65.53	**67.53**	54.44	56.18	**67.53**	65.53	65.53
TCGA	**78.2**	71.99	70.08	58.16	70.08	70.08	**78.2**	**78.2**
CD	82.22	82.22	82.22	82.22	65.56	82.22	82.22	**86.67**
CA	65.07	65.07	**65.36**	53.19	43.77	**65.36**	65.07	65.07
ID	**71.11**	**71.11**	70.0	68.89	63.33	70.0	**71.11**	**71.11**
Iris	**96.67**	**96.67**	**96.67**	95.33	54.67	**96.67**	**96.67**	96.0
LYM	**40.67**	**40.67**	38.67	33.86	24.95	38.67	**40.67**	**40.67**
Climate	92.41	90.56	92.41	88.52	88.33	92.41	**93.7**	92.41
Audit	94.82	**99.22**	95.07	66.71	65.17	95.07	94.82	82.06
BT	**58.91**	**58.91**	**58.91**	56.55	45.64	**58.91**	**58.91**	58.0
CBB	**84.46**	75.89	26.25	26.25	26.25	26.25	72.14	81.61
HES	31.1	**34.48**	24.05	24.05	24.05	24.05	28.86	22.81
CVR	**94.24**	**94.24**	90.32	63.42	63.42	90.32	93.54	93.77
Average	**73.41**	73.37	68.33	62.6	59.27	68.33	72.34	72.21

Table 13
The WIN/DRAW/LOSS results of NKRA against six algorithms under the F1 score.

Classifier	BFP	Imp_2^D	Imp_3^D	BFB	ES	ULGR
AdaBoost	9/6/2	17/0/0	17/0/0	9/6/2	4/8/5	6/6/5
DT	10/5/2	17/0/0	17/0/0	10/5/2	4/9/4	4/8/5
KNN	9/5/3	15/0/2	17/0/0	9/5/3	4/8/5	6/5/6
Total	28/16/7	49/0/2	51/0/0	28/16/7	12/25/14	16/19/16

Table 14
The WIN/DRAW/LOSS results of NKRA against six algorithms under the classification accuracy.

Classifier	BFP	Imp_2^D	Imp_3^D	BFB	ES	ULGR
AdaBoost	9/3/5	16/1/0	14/0/3	9/3/5	5/8/4	7/5/5
DT	10/6/1	14/0/3	14/0/3	10/6/1	6/9/2	6/5/6
KNN	10/4/3	14/1/2	15/0/2	10/4/3	7/8/2	7/5/5
Total	29/13/9	44/2/5	43/0/8	29/13/9	18/25/8	20/15/16

Table 15
Statistical test of five algorithms within two classifiers.

Classifiers	Classification metrics	Average ranking							χ_F^2	F_F
		BFP	Imp_2^D	Imp_3^D	BFB	ES	ULGR	NKRA		
AdaBoost	F1 score	3.59	6.26	6.09	3.59	2.88	3.00	2.59	51.26	16.16
DT	F1 score	3.53	6.12	6.29	3.53	2.97	3.00	2.56	52.19	16.77
KNN	F1 score	3.82	5.47	6.35	3.82	2.91	2.82	2.79	42.93	11.63
AdaBoost	Classification accuracy	3.76	5.82	5.03	3.76	3.26	3.35	3.00	23.51	4.79
DT	Classification accuracy	4.03	5.12	4.94	4.03	3.50	3.56	2.82	14.45	2.64
KNN	Classification accuracy	4.06	5.18	5.35	4.06	3.26	3.32	2.76	20.93	4.13

$$F_F = \frac{(N-1)\chi_F^2}{N(k-1)-\chi_F^2} \tag{5.6}$$

and

$$\chi_F^2 = \frac{12N}{k(k+1)}(\sum_{i=1}^{k} r_i^2 - \frac{k(k+1)^2}{4}), \tag{5.7}$$

Here, N and k denote the number of datasets and algorithms, respectively, and r_i represents the average ranking of the ith algorithm.

The variable χ_F^2 follows a χ^2 distribution with $\chi^2(k-1)$ degrees of freedom. Additionally, F_F adheres to the F distribution with $F(k-1,(k-1)(N-1))$ degrees of freedom.

By examining the classification metrics (F1 score and classification accuracy) across the seventeen UCI datasets, as detailed in Table 4, we ranked the seven algorithms for each dataset and computed the average ranking for each algorithm. The resulting average rankings, along with the corresponding χ_F^2 and F_F values, are presented in Table 15. If the F_F value surpasses the critical value of $F(k-1,(k-1)(N-1))$, we reject the null hypothesis that "there is no significant difference in the classification capabilities among all algorithms", thus indicating noteworthy performance distinctions among the algorithms.

With a significance level of $\alpha = 0.1$, F_F conforms to the F distribution with $F(6,96)$ degrees of freedom, and the critical value is 1.84. An examination of the data in Table 15 reveals that the F_F values surpass 1.84. This signifies significant disparities in the performance of the seven algorithms across each classifier under both classification metrics. Therefore, to delve deeper into the significant differences among algorithms, a post hoc test, such as the Bonferroni-Dunn test [38], becomes essential. The critical distance [39] is defined as:

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

Here, α represents the significance level, and q_α is the critical value for the Bonferroni-Dunn test.

This test assesses the average ranking difference between any two algorithms and compares it to the critical distance CD_α. If the ranking difference between two algorithms exceeds the critical distance CD_α, then it indicates a significant difference in their classification performance. For a significance level of $\alpha = 0.1$, the corresponding critical value $CD_\alpha = 1.7739$ can be calculated according to the above formula.

Fig. 4 illustrates the results of the Bonferroni-Dunn test conducted on the seven algorithms for different classifiers under the F1 score. Similarly, Fig. 5 illustrates the results of the Bonferroni-Dunn test conducted on the seven algorithms for the different classifiers under the classification accuracy. The outcomes from Figs. 4–5 reveal that, at a significance level of 0.1, three of the classifiers, NKRA significantly outperforms algorithms Imp_2^D, Imp_3^D, but there is no significant difference due to BFB, BFP, ES and ULGR.

Table 16
The Wilcoxon signed-rank test results in two classifiers across different algorithms.

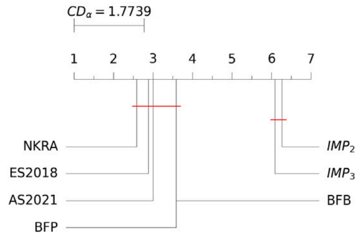
Classifier	Evaluation index	Algorithm	Wilcoxon statistic	P-value
AdaBoost	F1 Score	BFB	13	0.061
AdaBoost	Classification Accuracy	BFB	24	0.073
DT	F1 Score	BFB	0	0.007
DT	Classification Accuracy	BFB	14	0.007
KNN	F1 Score	BFB	0	0.049
KNN	Classification Accuracy	BFB	14	0.017
AdaBoost	F1 Score	BFP	13	0.061
AdaBoost	Classification Accuracy	BFP	24	0.073
DT	F1 Score	BFP	0	0.007
DT	Classification Accuracy	BFP	14	0.007
KNN	F1 Score	BFP	0	0.049
KNN	Classification Accuracy	BFP	14	0.017
AdaBoost	F1 Score	ES	36	0.767
AdaBoost	Classification Accuracy	ES	41	0.767
DT	F1 Score	ES	33	1.000
DT	Classification Accuracy	ES	38	0.400
KNN	F1 Score	ES	33	0.441
KNN	Classification Accuracy	ES	38	0.213
AdaBoost	F1 Score	ULGR	31	0.722
AdaBoost	Classification Accuracy	ULGR	38	0.432
DT	F1 Score	ULGR	39	0.858
DT	Classification Accuracy	ULGR	42	0.346
KNN	F1 Score	ULGR	39	0.8139
KNN	Classification Accuracy	ULGR	42	0.530

Post hoc tests based on mean ranks, as indicated in [40], may result in comparison outcomes between any two reduction algorithms being influenced by other algorithms. Therefore, it is essential for us to conduct Wilcoxon signed-rank tests [41] for pairwise comparisons to assess the differences between NKRA and BFB, BFP, ES, and ULGR.

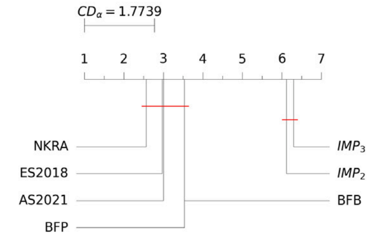
The detailed Wilcoxon signed-rank test results are presented separately in Table 16. For significance level $\alpha = 0.1$, regardless of the choice of F1 score or classification accuracy, and irrespective of whether AdaBoost or DT or KNN classifiers are used, the results indicate that there are significant differences between the BFB and BFP algorithms and the algorithm NKRA. However, the outcomes from Table 16 reveal that, at a significance level of 0.1, three of the classifiers, NKRA is not significantly better than ES and ULGR.

5.8. Comparison of computation times

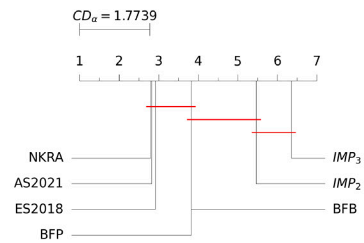
In this subsection, to test the efficiency of the algorithm NKRA, we compare the computation time of the algorithm NKRA with those of six other feature selection algorithms. Based on the results from Fig.



(a) Bonferroni-Dunn test results for AdaBoost under the F1 score

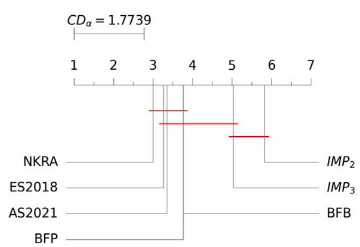


(b) Bonferroni-Dunn test results for DT under the F1 score

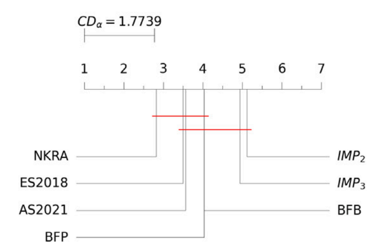


(c) Bonferroni-Dunn test results for KNN under the F1 score

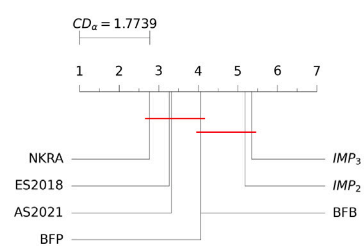
Fig. 4. Bonferroni-Dunn test results for different classifiers under the F1 score.



(a) Bonferroni-Dunn test results for AdaBoost under the classification accuracy



(b) Bonferroni-Dunn test results for DT under the classification accuracy



(c) Bonferroni-Dunn test results for KNN under the classification accuracy

Fig. 5. Bonferroni-Dunn test results for different classifiers under the classification.

6 and Table 17, the NKRA generally shorter computation time across the seventeen datasets than at least four of the other six algorithms do. This validates the effectiveness of the NKRA algorithm.

As seen in Table 17, our algorithm, which is based on KNN rule, occasionally takes longer than simpler methods, primarily because it carries out repeated distance metrics on heterogeneous data (including both numeric and categorical attributes, with possible missing values) and iteratively refines neighbourhoods. Even on moderately sized datasets, such meticulous checks – which improve reliability for isolated or surrounded objects – introduce additional overhead.

Meanwhile, the four other algorithms (IMP_2, IMP_3, BFB, and BFP) can also spend considerable time in practice despite promising theoretical complexity. This is not due to shortcomings in their design; rather, they each employ nested loops and multiple global updates, often re-checking attributes or object pairs to maintain high accuracy in mixed-attribute scenarios. When data distribution or missingness

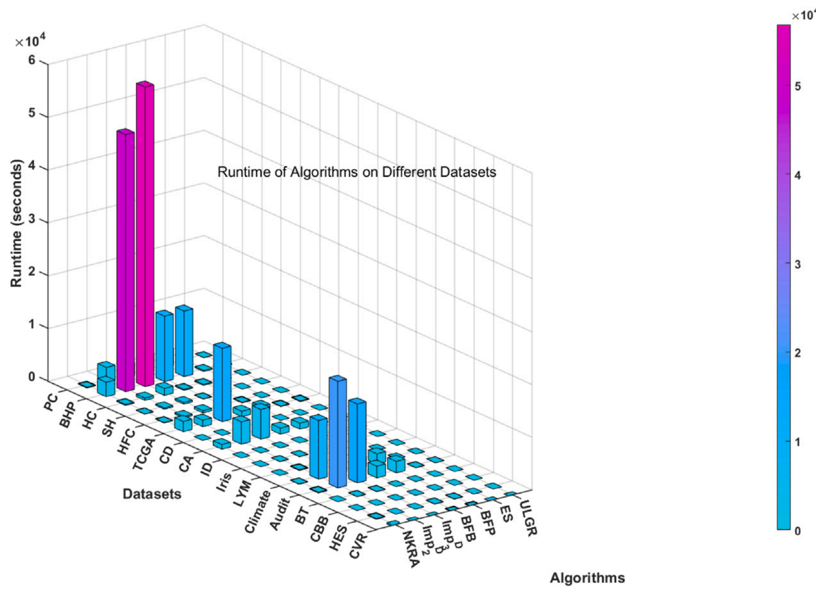


Fig. 6. Comparison of computation time.

Table 17
Actual running time for seven algorithms on seventeen datasets.

Dataset	NKRA	IMP_2	IMP_3	BFB	BFP	ES	ULGR
PC	139.149	2925.918	2578.976	3361.623	3370.372	21.307	34.555
BHP	2741.393	48 740.550	56 801.550	12 413.606	12 425.203	217.424	153.810
HC	181.169	454.936	1323.208	181.07	183.16	4.277	13.846
SH	53.074	237.822	316.933	59.462	60.043	2.609	4.805
HFC	75.839	277.166	425.549	113.719	108.093	4.574	7.072
TCGA	1944.466	1234.316	13 916.335	1020.016	1011.37	97.644	116.542
CD	1.592	4.651	6.785	1.731	1.748	0.182	0.26
CA	889.096	4248.499	5635.147	1157.268	1126.483	29.797	44.685
ID	1.585	9.435	10.269	2.245	2.215	0.223	0.285
Iris	0.883	36.063	35.504	7.302	7.837	0.388	0.394
LYM	4.66	139.366	119.482	28.094	26.932	8.706	2.933
Climate	56.926	11 130.66	11 166.845	1927.53	1956.62	16.501	19.733
Audit	150.531	20 215.834	14 968.349	2232.73	2187.82	57.293	28.78
BT	1.117	50.8	55.173	9.979	9.93	0.583	0.445
CBB	4.578	0.238	0.173	2.018	2.013	0.117	0.408
HES	109.852	1.174	1.052	53.839	55.301	7.577	3.541
CVR	280.762	4.344	4.421	133.334	130.893	7.866	16.74

hinders early pruning, these steps must be fully executed, amplifying constant factors and occasionally resulting in higher runtime.

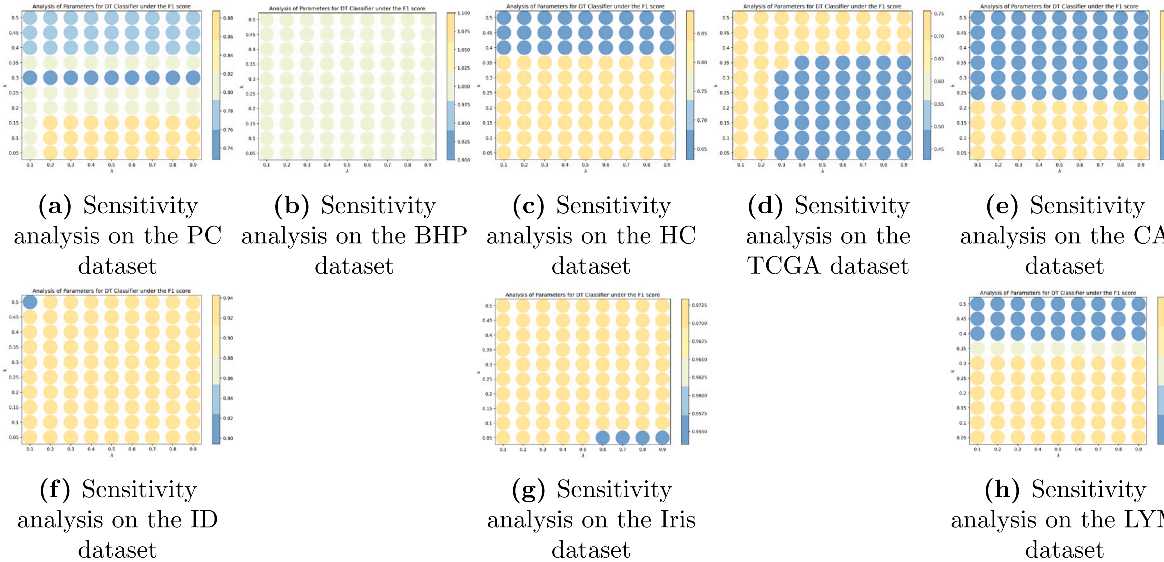
By contrast, ULGR and ES typically utilise fewer nested iterations or narrower searches, so they often finish faster under most distributions. Their lighter implementations are well-optimised for typical feature selection tasks, whereas our algorithm (and the aforementioned four methods) emphasise exhaustive checks that facilitate robust handling of more complex or partially-labelled data. Hence, these observed timing differences reflect varied trade-offs between thoroughness and efficiency, rather than any intrinsic flaw or failing in a given technique.

5.9. Sensitivity analysis of k and λ

In theory, the effectiveness of the algorithm NKRA is influenced by the parameters k and λ. To assess the influence of the parameters k and λ on the performance of the algorithm NKRA, we evaluate the F1 score and the classification accuracy of the reduced sets obtained from each dataset listed in Table 4. The parameter λ is varied from 0.1 to 0.9 with an increment of 0.1. If a smaller value of k is chosen, only the closest samples to the test object will have an impact on the prediction result, while the less close objects will not have any effect, leading to a larger estimation error. On the other hand, if a larger value of k is chosen, some samples that are farther away from the test sample will also have an impact on the prediction, resulting in prediction errors. Referring to the practices of Hu et al. [32] and Hu et al. [30], the parameter k is varied from $0.05|U|$ to $0.5|U|$ with an increment of $0.05|U|$, where $|U|$ is the number of objects in an HIS. We analyse how the algorithm NKRA behaves with different parameter values k and λ when employed with the DT classifier under both the F1 score and classification accuracy. The experimental results are shown in Figs. 7–8.

For simplicity, we focus solely on examining the impact of variations in both the F1 score and classification accuracy of the algorithm NKRA within the DT classifier (max depth = 3) concerning the parameters k and λ. The experimental outcomes are depicted in Figs. 7–8. Figs. 7–8 comprises a set of bubble charts. In each bubble chart, the horizontal axis represents the λ parameter, the vertical axis represents the k parameter, and the size of each bubble corresponds to the F1 score (or the classification accuracy). The legend positioned on the right delineates the ranges of F1 score (or classification accuracy) values. The more consistent the sizes of the bubbles in the chart are, the lower the sensitivity of the algorithm to the parameters. If multiple datasets show similar sensitivity analysis results, only one figure is presented in Figs. 7–8.

(1) When k is fixed, the algorithm NKRA has a notable influence on classification performance with respect to λ, as illustrated in Figs. 7–8.



(a) Sensitivity analysis on the PC dataset

(b) Sensitivity analysis on the BHP dataset

(c) Sensitivity analysis on the HC dataset

(d) Sensitivity analysis on the TCGA dataset

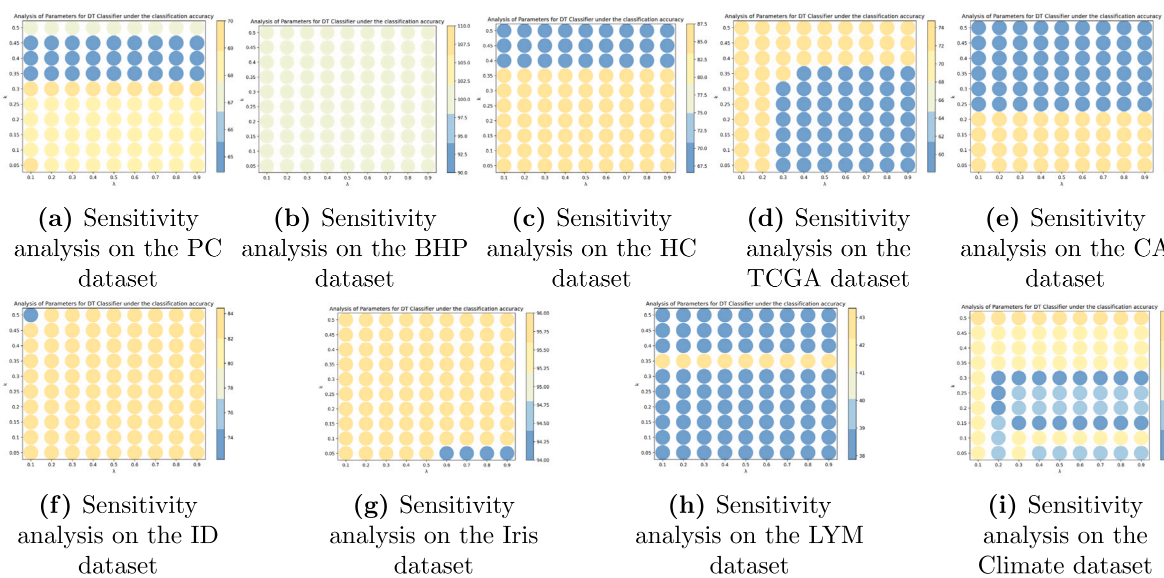
(e) Sensitivity analysis on the CA dataset

(f) Sensitivity analysis on the ID dataset

(g) Sensitivity analysis on the Iris dataset

(h) Sensitivity analysis on the LYM dataset

Fig. 7. Sensitivity analysis for the DT classifier under the F1 score across different datasets.



(a) Sensitivity analysis on the PC dataset

(b) Sensitivity analysis on the BHP dataset

(c) Sensitivity analysis on the HC dataset

(d) Sensitivity analysis on the TCGA dataset

(e) Sensitivity analysis on the CA dataset

(f) Sensitivity analysis on the ID dataset

(g) Sensitivity analysis on the Iris dataset

(h) Sensitivity analysis on the LYM dataset

(i) Sensitivity analysis on the Climate dataset

Fig. 8. Sensitivity analysis for the DT classifier under the classification accuracy across different datasets.

(2) When λ is fixed, the algorithm NKRA has a significant effect on classification performance with respect to k across the majority of datasets, as illustrated in Figs. 7–8.

In summary, the classification performance of the algorithm NKRA is significantly influenced by k and λ. Therefore, effective adjustment of k and λ can enhance the classification performance of the NKRA algorithm.

6. Conclusions

In this paper, a distance function in an HIS is introduced. Based on this function, a tolerance relation is defined, and an N-rough set model is built. Considering the limitation of N-rough set model under a surrounded category pattern, KNN-neighbourhood rough set model, is established through the combination of N-rough set model and KNN-rule. Four UMs based on KNN-neighbourhood rough set model are investigated. Their properties are validated through mathematical proofs, and the UM with optimal performance is selected through dispersion analysis. An attribute reduction algorithm (denoted as NKRA) based on the selected UM and overlap degree is designed. By using the AdaBoost, DT and KNN classifiers, the algorithm NKRA under the fixed parameters is compared with six other attribute reduction algorithms. A statistical test on comparison results is conducted. The statistical test has shown that the NKRA algorithm exhibits significant advantages in both F1 score and classification accuracy at a confidence level of 0.1 compared to BFB, BFP, Imp_2^D, and Imp_3^D and there is no significant difference between NKRA and ES or ULGR. In summary, the NKRA algorithm is superior to BFB, BFP, Imp_2^D, and Imp_3^D, and is on par with ES and ULGR. Following this, a parameter analysis and a time comparison are conducted to show the effect of two parameters, and that the NKRA algorithms time efficiency is at least on par with four of the other six algorithms in most cases.

Based on the theoretical analysis and experimental results, it is reasonable to believe that the KNN-neighbourhood rough set model has effectively utilised the strengths of both the N-rough set and KNN-rule to overcome their individual limitations. Future work will explore how to extend the application of KNN-neighbourhood rough set to explore attribute reduction in different information systems.

CRediT authorship contribution statement

Bozhan Li: Writing – original draft, Methodology, Investigation. **Qin Huang:** Writing – original draft, Investigation, Formal analysis. **Zhaowen Li:** Writing – review & editing, Methodology, Formal analysis. **Yonghua Lin:** Software, Investigation, Data curation.

Declaration of competing interest

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

Acknowledgements

The authors would like to thank the editors and the anonymous reviewers for their valuable comments and suggestions, which have helped immensely in improving the quality of the paper. This work is supported by Doctoral Research of Guangdong University of Science Technology (GKY-2024BSQDK-11) and Science Foundation in Guangdong University of Science and Technology (GKY-023KYZDK-1).

Appendix

Proposition 3.3. Consider an HIS (U, A, d), and given $u \in U$.

(1) If $B \subseteq C \subseteq A$, then $\forall\ k \in \mathbb{N},\ \forall\ \lambda \in [0,1]$, $C_N^{k,\lambda}(u) \subseteq B_N^{k,\lambda}(u)$.

(2) If $k_1 \leq k_2$, then $\forall\ B \subseteq A,\ \forall\ \lambda \in [0,1]$, $B_N^{k_1,\lambda}(u) \subseteq B_N^{k_2,\lambda}(u)$.

(3) If $\lambda_1 \leq \lambda_2$, then $\forall\ B \subseteq A,\ \forall\ k \in \mathbb{N}$, $B_N^{k,\lambda_1}(u) \subseteq B_N^{k,\lambda_2}(u)$.

Proof. (1) Suppose $B \subseteq C \subseteq A$. Obviously,

$$C^k(u) \subseteq B^k(u),\ \ N_C^{\lambda}(u) \subseteq N_B^{\lambda}(u).$$

This indicates that

$$C^k(u) \cap N_C^{\lambda}(u) \subseteq B^k(u) \cap N_B^{\lambda}(u).$$

This shows that

$$C_N^{k,\lambda}(u) \subseteq B_N^{k,\lambda}(u).$$

(2) Suppose $k_1 \leq k_2$. Then $B^{k_1}(u) \subseteq B^{k_2}(u)$. Hence

$$B^{k_1}(u) \cap N_B^{\lambda}(u) \subseteq B^{k_2}(u) \cap N_B^{\lambda}(u).$$

Therefore $B_N^{k_1,\lambda}(u) \subseteq B_N^{k_2,\lambda}(u)$.

(3) Suppose $\lambda_1 \leq \lambda_2$. Then $N_B^{\lambda_1}(u) \subseteq N_B^{\lambda_2}(u)$. Hence

$$B^k(u) \cap N_B^{\lambda_1}(u) \subseteq B^k(u) \cap N_B^{\lambda_2}(u).$$

Therefore $B_N^{k,\lambda_1}(u) \subseteq B_N^{k,\lambda_2}(u)$. □

Theorem 3.5. Consider an HIS (U, A, d).

(1) $\forall\ B \subseteq A,\ \forall\ X \in 2^U,\ \forall\ k \in \mathbb{N},\ \forall\ \lambda \in [0,1]$,

$$\underline{B_N^{k,\lambda}}(X) \subseteq X \subseteq \overline{B_N^{k,\lambda}}(X);$$

(2) $X \subseteq Y \subseteq U\ \ \Rightarrow \forall\ B \subseteq A,\ \forall\ k \in \mathbb{N},\ \forall\ \lambda \in [0,1]$,

$$\underline{B_N^{k,\lambda}}(X) \subseteq \underline{B_N^{k,\lambda}}(Y),\ \ \overline{B_N^{k,\lambda}}(X) \subseteq \overline{B_N^{k,\lambda}}(Y);$$

(3) If $B \subseteq C \subseteq A$, then $\forall\ X \in 2^U,\ \forall\ k \in \mathbb{N},\ \forall\ \lambda \in [0,1]$,

$$\underline{B_N^{k,\lambda}}(X) \subseteq \underline{C_N^{k,\lambda}}(X),\ \ \overline{C_N^{k,\lambda}}(X) \subseteq \overline{B_N^{k,\lambda}}(X);$$

(4) If $k_1 \leq k_2$, then $\forall\ B \subseteq A,\ \forall\ \lambda \in [0,1],\ \forall\ X \in 2^U$

$$\underline{B_N^{k_2,\lambda}}(X) \subseteq \underline{B_N^{k_1,\lambda}}(X),\ \ \overline{B_N^{k_1,\lambda}}(X) \subseteq \overline{B_N^{k_2,\lambda}}(X).$$

(5) If $\lambda_1 \leq \lambda_2$, then $\forall\ B \subseteq A,\ \forall\ k \in \mathbb{N},\ \forall\ X \in 2^U$

$$\underline{B_N^{k,\lambda_2}}(X) \subseteq \underline{B_N^{k,\lambda_1}}(X),\ \ \overline{B_N^{k,\lambda_1}}(X) \subseteq \overline{B_N^{k,\lambda_2}}(X).$$

Proof. (1) and (2) are obvious.

(3) (i) $\forall\ u \in \underline{B_N^{k,\lambda}}(X)$, this implies that $B_N^{k,\lambda}(u) \subseteq X$.

It can be seen that $C_N^{k,\lambda}(u) \subseteq B_N^{k,\lambda}(u) \subseteq X$ according to Proposition 3.3.

It follows that $u \in \underline{C_N^{k,\lambda}}(X)$, thus $\underline{B_N^{k,\lambda}}(X) \subseteq \underline{C_N^{k,\lambda}}(X)$.

(ii) $\forall\ u \in \overline{C_N^{k,\lambda}}(X)$, this shows that $C_N^{k,\lambda}(u) \cap X \neq \emptyset$.

It can be seen that $C_N^{k,\lambda}(u) \subseteq B_N^{k,\lambda}(u)$ according to Proposition 3.3.

Then $B_N^{k,\lambda}(u) \cap X \neq \emptyset$. Therefore $u \in \overline{B_N^{k,\lambda}}(X)$.

This implies that $\overline{C_N^{k,\lambda}}(X) \subseteq \overline{B_N^{k,\lambda}}(X)$.

(4) (i)$\forall\ u \in \underline{B_N^{k_2,\lambda}}(X)$, this shows that $B_N^{k_2,\lambda}(u) \subseteq X$.

The fact that $B_N^{k_1,\lambda}(u) \subseteq B_N^{k_2,\lambda}(u)$ follows from Proposition 3.3.

Then $B_N^{k_1,\lambda}(u) \subseteq X$. This indicates that $u \in \underline{B_N^{k_1,\lambda}}(X)$.

Therefore $\underline{B_N^{k_2,\lambda}}(X) \subseteq \underline{B_N^{k_1,\lambda}}(X)$.

(ii) $\forall\ u \in \overline{B_N^{k_1,\lambda}}(X)$, this indicates that $B_N^{k_1,\lambda}(u) \cap X \neq \emptyset$.

According to Proposition 3.3, we have $B_N^{k_1,\lambda}(u) \subseteq B_N^{k_2,\lambda}(u)$.

Therefore $B_N^{k_2,\lambda}(u) \cap X \neq \emptyset$. It shows that $u \in \overline{B_N^{k_2,\lambda}}(X)$.

Thus $\overline{B_N^{k_1,\lambda}}(X) \subseteq \overline{B_N^{k_2,\lambda}}(X)$.

(5) (i) $\forall\ u \in \underline{B_N^{k,\lambda_2}}(X)$, it shows that $B_N^{k,\lambda_2}(u) \subseteq X$.

The fact that $B_N^{k,\lambda_1}(u) \subseteq B_N^{k,\lambda_2}(u)$ follows from Proposition 3.3.

Therefore $B_N^{k,\lambda_1}(u) \subseteq B_N^{k,\lambda_2}(u) \subseteq X$. It shows that $u \in \underline{B_N^{k,\lambda_1}}(X)$.

Thus $\underline{B_N^{k,\lambda_2}}(X) \subseteq \underline{B_N^{k,\lambda_1}}(X)$.

(ii) $\forall\ u \in \overline{B_N^{k,\lambda_1}}(X)$, it indicates that $B_N^{k,\lambda_1}(u) \cap X \neq \emptyset$.

It can be seen that $B_N^{k,\lambda_1}(u) \subseteq B_N^{k,\lambda_2}(u)$ according to Proposition 3.3.

Then $B_N^{k,\lambda_2}(u) \cap X \neq \emptyset$. It shows that $u \in \overline{B_N^{k,\lambda_2}}(X)$.

Thus $\overline{B_N^{k,\lambda_1}}(X) \subseteq \overline{B_N^{k,\lambda_2}}(X)$. □

Proposition 3.7. Consider an HIS (U, A, d), given $B \subseteq A$, $k \in \mathbb{N}$ *and* $\lambda \in [0,1]$. Then

$$\forall\ u \in U,\ |\partial_{B,N}^{k,\lambda}(u)| = 1\quad \Leftrightarrow\quad B_N^{k,\lambda} \subseteq R_d.$$

Proof. " $\Leftarrow$ ". Let $B_N^{k,\lambda} \subseteq R_d$ and $u \in U$. $\forall\ w \in \partial_{B,N}^{k,\lambda}(u)$, there exists $v \in B_N^{k,\lambda}(u)$ such that $w = d(v)$. As $v \in B_N^{k,\lambda}(u) \subseteq [u]_d$, we conclude that $v \in [u]_d$. Consequently, $w = d(v) = d(u)$. Thus, $|\partial_{B,N}^{k,\lambda}(u)| = 1$.

" $\Rightarrow$ ". In view of the fact that $\forall\ u \in U$, $|\partial_{B,N}^{k,\lambda}(u)| = 1$. $\forall\ v \in B_N^{k,\lambda}(u)$, it is evident that $d(v) \in \partial_{B,N}^{k,\lambda}(u)$. It is worth noting that $d(u) \in \partial_{B,N}^{k,\lambda}(u)$ and $|\partial_{B,N}^{k,\lambda}(u)| = 1$. Consequently, $d(u) = d(v)$. This indicates that $v \in [u]_d$. Thus $B_N^{k,\lambda}(u) \subseteq [u]_d$. Therefore $B_N^{k,\lambda} \subseteq R_d$. □

Proposition 4.3. Consider an HIS (U, A, d), the ensuing properties can be delineated:

(1) $\forall\ B \subseteq A,\ \forall\ k \in \mathbb{N},\ \forall\ \lambda \in [0,1]$,

$$\Gamma_{B,N}^{k,\lambda}(d) = \frac{1}{n}\sum_{i=1}^{s} |\underline{B_N^{k,\lambda}}(D_i)|;$$

(2) $\forall\ B \subseteq A,\ \forall\ k \in \mathbb{N},\ \forall\ \lambda \in [0,1]$,

$$0 \leq \Gamma_{B,N}^{k,\lambda}(d) \leq 1;$$

(3) If $B \subseteq C \subseteq A$, then $\forall\ k \in \mathbb{N},\ \forall\ \lambda \in [0,1]$,

$$\Gamma_{B,N}^{k,\lambda}(d) \leq \Gamma_{C,N}^{k,\lambda}(d);$$

(4) If $k_1 \leq k_2$, then $\forall\ B \subseteq A,\ \forall\ \lambda \in [0,1]$,

$$\Gamma_{B,N}^{k_2,\lambda}(d) \leq \Gamma_{B,N}^{k_1,\lambda}(d);$$

(5) If $\lambda_1 \leq \lambda_2$, then $\forall\ B \subseteq A,\ \forall\ k \in \mathbb{N}$,

$$\Gamma_{B,N}^{k,\lambda_2}(d) \leq \Gamma_{B,N}^{k,\lambda_1}(d);$$

(6) $\Gamma_{B,N}^{k,\lambda}(d) = 1\ \Leftrightarrow\ \forall\ i,\ \underline{B_N^{k,\lambda}}(D_i) = D_i$.

Proof. (1) Clearly, $\underline{B_N^{k,\lambda}}(D_i) \subseteq D_i$.

Because $U/d = \{D_1, D_2, \ldots, D_s\}$, it follows that

$$|POS_{B,N}^{k,\lambda}(d)| = |\bigcup_{i=1}^{s} \underline{B_N^{k,\lambda}}(D_i)| = \sum_{i=1}^{s} |\underline{B_N^{k,\lambda}}(D_i)|.$$

Hence $\Gamma_{B,N}^{k,\lambda}(d) = \frac{1}{n}\sum_{i=1}^{s} |\underline{B_N^{k,\lambda}}(D_i)|$.

(2) It can be evidenced by Definition 4.2.

(3) Let $B \subseteq C \subseteq A$, by Definition 4.2 and (1), it follows that

$$\Gamma_{B,N}^{k,\lambda}(d) = \frac{1}{n}|POS_{B,N}^{k,\lambda}(d)| = \frac{1}{n}\sum_{i=1}^{s} |\underline{B_N^{k,\lambda}}(D_i)|.$$

$$\Gamma_{C,N}^{k,\lambda}(d) = \frac{1}{n}|POS_{C,N}^{k,\lambda}(d)| = \frac{1}{n}\sum_{i=1}^{s} |\underline{C_N^{k,\lambda}}(D_i)|.$$

By Theorem 3.5, $\forall\ i$, $\underline{B_N^{k,\lambda}}(D_i) \subseteq \underline{C_N^{k,\lambda}}(D_i)$. Thus $\Gamma_{B,N}^{k,\lambda}(d) \leq \Gamma_{C,N}^{k,\lambda}(d)$.

(4) By Theorem 3.5, $\forall\ i, \underline{B_N^{k_2,\lambda}}(D_i) \subseteq \underline{B_N^{k_1,\lambda}}(D_i)$

Hence $\forall\ i, |\underline{B_N^{k_2,\lambda}}(D_i)| \leq |\underline{B_N^{k_1,\lambda}}(D_i)|$. Thus

$$\Gamma_{B,N}^{k_2,\lambda}(d) \leq \Gamma_{B,N}^{k_1,\lambda}(d).$$

(5) It can be evidenced that $\forall\ i, \underline{B_N^{k,\lambda_2}}(D_i) \subseteq \underline{B_N^{k,\lambda_1}}(D_i)$ by Theorem 3.5.

Hence, $\forall\ i, |\underline{B_N^{k_2,\lambda}}(D_i)| \leq |\underline{B_N^{k_1,\lambda}}(D_i)|$. Thus

$$\Gamma_{B,N}^{k,\lambda_2}(d) \leq \Gamma_{B,N}^{k,\lambda_1}(d).$$

(6) "Sufficiency". This is obvious.

"Necessity". Assume that $\Gamma_{B,N}^{k,\lambda}(d) = 1$. According to (1),

$$\sum_{i=1}^{s} |\underline{B_N^{k,\lambda}}(D_i)| = \sum_{i=1}^{s} |D_i|.$$

This suggest that $\sum_{i=1}^{s}(|D_i| - |\underline{B_N^{k,\lambda}}(D_i)|) = 0$.

By Theorem 3.5, $\forall\ i$, $\underline{B_N^{k,\lambda}}(D_i) \subseteq D_i$.

It follows that $\forall\ i$, $|D_i| - |\underline{B_N^{k,\lambda}}(D_i)| \geq 0$. Then

$$\forall\ i,\quad |D_i| - |\underline{B_N^{k,\lambda}}(D_i)| = 0.$$

Thus $\forall\ i$, $\underline{B_N^{k,\lambda}}(D_i) = D_i$. □

Proposition 4.5. Consider an HIS (U, A, d), given $B \subseteq A$, $k \in \mathbb{N}$ *and* $\lambda \in [0,1]$. Then

$$0 \leq H_N^{k,\lambda}(B) \leq n\log_2 n. \tag{A.1}$$

Furthermore, if $B^k = \triangle$, then $H_N^{k,\lambda}(B) = \log_2 n$; if $B^k = \delta$, then $H_N^{k,\lambda}(B) = 0$.

Proof. Note that $\forall\ i$, $1 \leq |B_N^{k,\lambda}(u_i)| \leq n$. It follows that

$$\frac{1}{n} \leq \frac{|B_N^{k,\lambda}(u_i)|}{n} \leq 1,$$

$$0 \leq -\log_2 \frac{|B_N^{k,\lambda}(u_i)|}{n} \leq \log_2 n.$$

Then

$$0 \leq -\frac{|B_N^{k,\lambda}(u_i)|}{n}\log_2 \frac{|B_N^{k,\lambda}(u_i)|}{n} \leq \log_2 n.$$

Therefore

$$0 \leq H_N^{k,\lambda}(B) \leq n\log_2 n.$$

Let $B^k = \triangle$. Then $\forall\ i$, $|B_N^{k,\lambda}(u_i)| = 1$. Thus $H_N^{k,\lambda}(B) = \log_2 n$.

Let $B^k = \delta$. Then $\forall\ i$, $|B_N^{k,\lambda}(u_i)| = n$. This shows that $H_N^{k,\lambda}(B) = 0$. □

Proposition 4.7. Consider an HIS (U, A, d), given $k \in \mathbb{N}$ *and* $\lambda \in [0,1]$. If $B \subseteq C \subseteq A$, then

$$H_N^{k,\lambda}(d|C) \leq H_N^{k,\lambda}(d|B). \tag{A.2}$$

Proof.

$$p_{ij}^{(1)} = |B_N^{k,\lambda}(u_i) \cap D_j|,\quad p_{ij}^{(2)} = |B_N^{k,\lambda}(u_i) \cap (U - D_j)|;$$

$$q_{ij}^{(1)} = |C_N^{k,\lambda}(u_i) \cap D_j|,\quad q_{ij}^{(2)} = |C_N^{k,\lambda}(u_i) \cap (U - D_j)|.$$

Then

$$\forall\ i,j,\ |B_N^{k,\lambda}(u_i)| = p_{ij}^{(1)} + p_{ij}^{(2)},\ |C_N^{k,\lambda}(u_i)| = q_{ij}^{(1)} + q_{ij}^{(2)}.$$

By Proposition 3.3, $\forall\ i$, $C_N^{k,\lambda}(u_i) \subseteq B_N^{k,\lambda}(u_i)$.

Then

$$\forall\ i,j,\quad 0 \leq q_{ij}^{(1)} \leq p_{ij}^{(1)},\quad 0 \leq q_{ij}^{(2)} \leq p_{ij}^{(2)}.$$

$$\begin{aligned}
H_N^{k,\lambda}(d|B) &= -\sum_{i=1}^{n}\sum_{j=1}^{s} \frac{|B_N^{k,\lambda}(u_i) \cap D_j|}{n}\log_2 \frac{|B_N^{k,\lambda}(u_i) \cap D_j|}{|B_N^{k,\lambda}(u_i)|} \\
&= -\sum_{i=1}^{n}\sum_{j=1}^{s} \frac{p_{ij}^{(1)}}{n}\log_2 \frac{p_{ij}^{(1)}}{p_{ij}^{(1)} + p_{ij}^{(2)}} \\
&\triangleq \sum_{i=1}^{n}\sum_{j=1}^{s} \frac{1}{n} f(p_{ij}^{(1)}, p_{ij}^{(2)});
\end{aligned}$$

$$\begin{aligned}
H_N^{k,\lambda}(d|C) &= -\sum_{i=1}^{n}\sum_{j=1}^{s} \frac{|C_N^{k,\lambda}(u_i) \cap D_j|}{n}\log_2 \frac{|C_N^{k,\lambda}(u_i) \cap D_j|}{|C_N^{k,\lambda}(u_i)|} \\
&= -\sum_{i=1}^{n}\sum_{j=1}^{s} \frac{q_{ij}^{(1)}}{n}\log_2 \frac{q_{ij}^{(1)}}{q_{ij}^{(1)} + q_{ij}^{(2)}} \\
&\triangleq \sum_{i=1}^{n}\sum_{j=1}^{s} \frac{1}{n} f(q_{ij}^{(1)}, q_{ij}^{(2)}).
\end{aligned}$$

Put $f(u,v) = -u\log_2 \frac{u}{u+v} (u > 0, v \geq 0)$. If u or v is held constant, the function $f(u,v)$ becomes a univariate function and exhibits monotonic increase.

Since $q_{ij}^{(1)} \leq p_{ij}^{(1)}, q_{ij}^{(2)} \leq p_{ij}^{(2)}$, it can be seen that

$$f(q_{ij}^{(1)}, q_{ij}^{(2)}) \leq f(p_{ij}^{(1)}, q_{ij}^{(2)}) \leq f(p_{ij}^{(1)}, p_{ij}^{(2)}).$$

Thus

$$H_N^{k,\lambda}(d|C) \leq H_N^{k,\lambda}(d|B).\quad □$$

Proposition 4.9. Consider an HIS (U, A, d), given $B \subseteq A$, $k \in \mathbb{N}$ *and* $\lambda \in [0,1]$. Then

$$H_N^{k,\lambda}(d|B) = H_N^{k,\lambda}(B \cup d) - H_N^{k,\lambda}(B). \tag{A.3}$$

Proof. As $\{D_1, D_2 \cdots, D_s\}$ forms a partition of U, it can be observed that $\forall\ i$,

$$\sum_{j=1}^{s} |B_N^{k,\lambda}(u_i) \cap D_j| = |B_N^{k,\lambda}(u_i)|;$$

$$H_N^{k,\lambda}(d|B) = -\sum_{i=1}^{n}\sum_{j=1}^{s}\frac{|B_N^{k,\lambda}(u_i)\cap D_j|}{n}\log_2\frac{|B_N^{k,\lambda}(u_i)\cap D_j|}{|B_N^{k,\lambda}(u_i)|}$$
$$= -\sum_{i=1}^{n}\sum_{j=1}^{s}\frac{|B_N^{k,\lambda}(u_i)\cap D_j|}{n}(\log_2\frac{|B_N^{k,\lambda}(u_i)\cap D_j|}{n}$$
$$-\log_2\frac{|B_N^{k,\lambda}(u_i)|}{n})$$
$$= -\sum_{i=1}^{n}\sum_{j=1}^{s}\frac{|B_N^{k,\lambda}(u_i)\cap D_j|}{n}\log_2\frac{|B_N^{k,\lambda}(u_i)\cap D_j|}{n}$$
$$+\sum_{i=1}^{n}\sum_{j=1}^{s}\frac{|B_N^{k,\lambda}(u_i)\cap D_j|}{n}\log_2\frac{|B_N^{k,\lambda}(u_i)|}{n}$$
$$= H_N^{k,\lambda}(B\cup d)+\sum_{i=1}^{n}\frac{|B_N^{k,\lambda}(u_i)|}{n}\log_2\frac{|B_N^{k,\lambda}(u_i)|}{n}$$
$$= H_N^{k,\lambda}(B\cup d)-H_N^{k,\lambda}(B). \quad \square$$

Proposition 4.10. Consider an HIS (U,A,d), given $B\subseteq A$, $k\in\mathbb{N}$ *and* $\lambda\in[0,1]$. Then $H_N^{k,\lambda}(d|B)\geq 0$.

Proof. According to Definition 4.4,

$$H_N^{k,\lambda}(B) = -\sum_{i=1}^{n}\frac{|B_N^{k,\lambda}(u_i)|}{n}\log_2\frac{|B_N^{k,\lambda}(u_i)|}{n}.$$

As $\{D_1, D_2\cdots, D_s\}$ forms a partition of U, it can be observed that $\forall\ i$,

$$\sum_{j=1}^{s}|B_N^{k,\lambda}(u_i)\cap D_j| = |B_N^{k,\lambda}(u_i)|.$$

Then

$$H_N^{k,\lambda}(B) = -\sum_{i=1}^{n}\frac{\sum_{j=1}^{s}|B_N^{k,\lambda}(u_i)\cap D_j|}{n}\log_2\frac{|B_N^{k,\lambda}(u_i)|}{n}$$
$$= -\sum_{i=1}^{n}\sum_{j=1}^{s}\frac{|B_N^{k,\lambda}(u_i)\cap D_j|}{n}\log_2\frac{|B_N^{k,\lambda}(u_i)|}{n}.$$

By Definition 4.8,

$$H_N^{k,\lambda}(B\cup d) = -\sum_{i=1}^{n}\sum_{j=1}^{s}\frac{|B_N^{k,\lambda}(u_i)\cap D_j|}{n}\log_2\frac{|B_N^{k,\lambda}(u_i)\cap D_j|}{n}.$$

$\forall\ i,j$,

$$log_2\frac{|B_N^{k,\lambda}(u_i)\cap D_j|}{n}\leq\ log_2\frac{|B_N^{k,\lambda}(u_i)|}{n}.$$

Then

$$H_N^{k,\lambda}(B)\leq H_N^{k,\lambda}(B\cup d).$$

The fact that $H_N^{k,\lambda}(d|B)\geq 0$ follows from Proposition 4.9. $\square$

Proposition 4.12. Consider an HIS (U,A,d), given $B\subseteq A$, $k\in\mathbb{N}$ *and* $\lambda\in[0,1]$. Then

$$0\leq E_N^{k,\lambda}(B)\leq n-1. \tag{A.4}$$

Furtherover, if $B^k=\triangle$, it can be obtained that $E_N^{k,\lambda}(B)=1-\frac{1}{n}$;

Proof. Clearly, $\forall\ i$, $1\leq|B_N^{k,\lambda}(u_i)|\leq n$. This indicates that

$$\frac{1}{n}\leq\frac{|B_N^{k,\lambda}(u_i)|}{n}\leq 1,$$

$$0\leq 1-\frac{|B_N^{k,\lambda}(u_i)|}{n}\leq 1-\frac{1}{n}.$$

So

$$0\leq\frac{|B_N^{k,\lambda}(u_i)|}{n}(1-\frac{|B_N^{k,\lambda}(u_i)|}{n})\leq 1-\frac{1}{n}$$

It shows that

$$0\leq\sum_{i=1}^{n}\frac{|B_N^{k,\lambda}(u_i)|}{n}(1-\frac{|B_N^{k,\lambda}(u_i)|}{n})\leq n-1$$

Thus

$$0\leq E_N^{k,\lambda}(B)\leq n-1.$$

Let $B^k=\triangle$. Then $\forall\ i$, $|B_N^{k,\lambda}(u_i)|=1$. Therefore $E_N^{k,\lambda}(B)=1-\frac{1}{n}$. $\square$

Proposition 4.14. Consider an HIS (U,A,d), given $k\in\mathbb{N}$ *and* $\lambda\in[0,1]$. If $B\subseteq C\subseteq A$, then

$$E_N^{k,\lambda}(d|C)\leq E_N^{k,\lambda}(d|B). \tag{A.5}$$

Proof. Assume that $B\subseteq C\subseteq A$. By Proposition 3.3, $\forall\ i$, $C_N^{k,\lambda}(u_i)\subseteq B_N^{k,\lambda}(u_i)$. It follows that

$$\forall\ i,j,\ C_N^{k,\lambda}(u_i)\cap D_j\subseteq B_N^{k,\lambda}(u_i)\cap D_j,\ C_N^{k,\lambda}(u_i)-D_j\subseteq B_N^{k,\lambda}(u_i)-D_j.$$

This implies that

$$\forall\ i,j,\ |C_N^{k,\lambda}(u_i)\cap D_j|\leq|B_N^{k,\lambda}(u_i)\cap D_j|,\ |C_N^{k,\lambda}(u_i)-D_j|\leq|B_N^{k,\lambda}(u_i)-D_j|.$$

Thus

$$E_N^{k,\lambda}(d|C)\leq E_N^{k,\lambda}(d|B). \quad \square$$

Proposition 4.16. Consider an HIS (U,A,d), given $B\subseteq A$, $k\in\mathbb{N}$ *and* $\lambda\in[0,1]$. Then

$$E_N^{k,\lambda}(d|B)=E_N^{k,\lambda}(B\cup d)-E_N^{k,\lambda}(B). \tag{A.6}$$

Proof. As $\{D_1,D_2,\dots,D_s\}$ forms a partition of U, it can be observed that $\forall\ i$,

$$\sum_{j=1}^{s}|B_N^{k,\lambda}(u_i)\cap D_j|=|B_N^{k,\lambda}(u_i)|;$$

$$E_N^{k,\lambda}(d|B)=\sum_{i=1}^{n}\sum_{j=1}^{s}\frac{|B_N^{k,\lambda}(u_i)\cap D_j|}{n}\frac{|B_N^{k,\lambda}(u_i)-B_N^{k,\lambda}(u_i)\cap D_j|}{n}$$
$$=\sum_{i=1}^{n}\sum_{j=1}^{s}\frac{|B_N^{k,\lambda}(u_i)\cap D_j|}{n}\frac{|B_N^{k,\lambda}(u_i)|-|B_N^{k,\lambda}(u_i)\cap D_j|}{n}$$
$$=\sum_{i=1}^{n}\sum_{j=1}^{s}\frac{|B_N^{k,\lambda}(u_i)\cap D_j|}{n}(\frac{|B_N^{k,\lambda}(u_i)|}{n}-\frac{|B_N^{k,\lambda}(u_i)\cap D_j|}{n})$$
$$=\sum_{i=1}^{n}\sum_{j=1}^{s}\frac{|B_N^{k,\lambda}(u_i)\cap D_j|}{n}((1-\frac{|B_N^{k,\lambda}(u_i)\cap D_j|}{n})-(1-\frac{|B_N^{k,\lambda}(u_i)|}{n}))$$
$$=\sum_{i=1}^{n}\sum_{j=1}^{s}\frac{|B_N^{k,\lambda}(u_i)\cap D_j|}{n}(1-\frac{|B_N^{k,\lambda}(u_i)\cap D_j|}{n})-$$
$$\sum_{i=1}^{n}\sum_{j=1}^{s}\frac{|B_N^{k,\lambda}(u_i)\cap D_j|}{n}(1-\frac{|B_N^{k,\lambda}(u_i)|}{n})$$
$$=E_N^{k,\lambda}(B\cup d)-\sum_{i=1}^{n}\frac{|B_N^{k,\lambda}(u_i)|}{n}(1-\frac{|B_N^{k,\lambda}(u_i)|}{n})$$
$$=E_N^{k,\lambda}(B\cup d)-E_N^{k,\lambda}(B). \quad \square$$

Proposition 4.18. Consider an HIS (U,A,d), given $B\subseteq A$ and $k\in\mathbb{N}$. Then

$$0\leq DI_N^k(B)\leq log_2 n. \tag{A.7}$$

Moreover, if $B^k=\triangle$, then $DI_N^k(B)=log_2 n$; if $B^k=\delta$, then $DI_N^k(B)=0$.

Proof. Since $n\leq|B^k|\leq n^2$, it follows that $\frac{1}{n^2}\leq\frac{1}{|B^k|}\leq\frac{1}{n}$.

Then

$$1 \le \frac{n^2}{|B^k|} \le n.$$

Thus

$$0 \le DI_N^k(B) \le log_2 n.$$

Let $B^k = \triangle$. Then $|B^k| = n$. This implies that $DI_N^k(B) = log_2 n$.
Let $B^k = \delta$. Then $|B^k| = n^2$. This indicates that $DI_N^k(B) = 0$. □

Proposition 4.20. Consider an HIS (U, A, d), given $B \subseteq A$, $k \in \mathbb{N}$ *and* $\lambda \in [0, 1]$. Then

$$0 \le DI_N^{k,\lambda}(d|B) \le log_2 n. \tag{A.8}$$

Moreover, if $B_N^{k,\lambda} = \triangle$, then $DI_N^{k,\lambda}(d|B) = 0$; if $B_N^{k,\lambda} = \delta$, then $DI_N^{k,\lambda}(d|B) = log_2 \frac{n^2}{|R_d|}$.

Proof. Since $\triangle \subseteq B_N^{k,\lambda}$ and $\triangle \subseteq R_d$, it can be seen that $\triangle \subseteq B_N^{k,\lambda} \cap R_d$. Then $n \le |B_N^{k,\lambda} \cap R_d| \le n^2$. This implies that

$$\frac{1}{n^2} \le \frac{1}{|B_N^{k,\lambda} \cap R_d|} \le \frac{1}{n}.$$

Clearly, $n \le |B^k| \le n^2$, $|B_N^{k,\lambda} \cap R_d| \le |B^k|$. Then

$$1 \le \frac{|B^k|}{|B_N^{k,\lambda} \cap R_d|} \le n.$$

Thus

$$0 \le DI_N^{k,\lambda}(d|B) \le log_2 n.$$

Let $B^k = \triangle$. Then $B_N^{k,\lambda} \cap R_d = \triangle = B^k$. Therefore $DI_N^{k,\lambda}(d|B) = 0$.
Let $B_N^{k,\lambda} = \delta$. Then $B_N^{k,\lambda} \cap R_d = R_d$. It shows that $DI_N^{k,\lambda}(d|B) = log_2 \frac{n^2}{|R_d|}$. □

Lemma 4.23. Consider an HIS (U, A, d), given $B \subseteq A$, $k \in \mathbb{N}$ *and* $\lambda \in [0, 1]$. If $B_N^{k,\lambda} \subseteq R_d$, then $\forall\, u \in U$, $\forall\, j$,

$$B_N^{k,\lambda}(u) \cap D_j = \begin{cases} B_N^{k,\lambda}(u), & u \in D_j; \\ \emptyset, & u \notin D_j. \end{cases}$$

Proof. Assume that $u \in D_j$. This indicates that $D_j = [u]_d$. $B_N^{k,\lambda} \subseteq R_d$ implies that $B_N^{k,\lambda}(u) \subseteq [u]_d$. Therefore $B_N^{k,\lambda}(u) \cap D_j = B_N^{k,\lambda}(u)$.

Assume that $u \notin D_j$. This shows that $[u]_d \cap D_j = \emptyset$. $B_N^{k,\lambda} \subseteq R_d$ implies that $B_N^{k,\lambda}(u) \subseteq [u]_d$. Thus $B_N^{k,\lambda}(u) \cap D_j = \emptyset$. □

Lemma 4.24. Consider an HIS (U, A, d), given $B \subseteq A$, $k \in \mathbb{N}$ *and* $\lambda \in [0, 1]$. If $B_N^{k,\lambda} \subseteq R_d$, then $\forall\, u \in U$,

$$\sum_{j=1}^{s} \frac{|B_N^{k,\lambda}(u) \cap D_j|}{n}\left(1 - \frac{|B_N^{k,\lambda}(u) \cap D_j|}{n}\right) = \frac{|B_N^{k,\lambda}(u)|}{n}\left(1 - \frac{|B_N^{k,\lambda}(u)|}{n}\right),$$

$$\sum_{j=1}^{s} \frac{|B_N^{k,\lambda}(u) \cap D_j|}{n} \log_2 \frac{|B_N^{k,\lambda}(u) \cap D_j|}{n} = \frac{|B_N^{k,\lambda}(u)|}{n} \log_2 \frac{|B_N^{k,\lambda}(u)|}{n}.$$

Proof. As $\{D_1, D_2, \ldots, D_s\}$ forms a partition of U, it can be seen that $u \in D_{j^*}$ for some j^*.

Since $B_N^{k,\lambda} \subseteq R_d$. Then by Lemma 4.23,

$$B_N^{k,\lambda}(u) \cap D_j = \begin{cases} B_N^{k,\lambda}(u), & j = j^*; \\ \emptyset, & j \ne j^*. \end{cases}$$

It follows that

$$\sum_{j=1}^{s} \frac{|B_N^{k,\lambda}(u) \cap D_j|}{n}\left(1 - \frac{|B_N^{k,\lambda}(u) \cap D_j|}{n}\right) = \frac{|B_N^{k,\lambda}(u)|}{n}\left(1 - \frac{|B_N^{k,\lambda}(u)|}{n}\right),$$

$$\sum_{j=1}^{s} \frac{|B_N^{k,\lambda}(u) \cap D_j|}{n} \log_2 \frac{|B_N^{k,\lambda}(u) \cap D_j|}{n} = \frac{|B_N^{k,\lambda}(u)|}{n} \log_2 \frac{|B_N^{k,\lambda}(u)|}{n}. \quad □$$

Proposition 4.25. Consider an HIS (U, A, d), given $B \subseteq A$, $k \in \mathbb{N}$ *and* $\lambda \in [0, 1]$. The statements below are equivalent:

(1) $B_N^{k,\lambda} \subseteq R_d$;
(2) $\Gamma_{B,N}^{k,\lambda}(d) = 1$;
(3) $H_N^{k,\lambda}(d|B) = 0$;
(4) $DI_N^{k,\lambda}(d|B) = 0$.

Proof. "(1) ⇒ (2)". Let $B_N^{k,\lambda} \subseteq R_d$. Suppose there exists j^* such that $D_{j^*} \not\subseteq \underline{B_N^{k,\lambda}}(D_{j^*})$. Then $D_{j^*} - \underline{B_N^{k,\lambda}}(D_{j^*}) \ne \emptyset$. Pick

$$u^* \in D_{j^*} - \underline{B_N^{k,\lambda}}(D_{j^*}).$$

This indicates that $u^* \in D_{j^*}$, $u^* \notin \underline{B_N^{k,\lambda}}(D_{j^*})$.
$u^* \in D_{j^*}$ implies that $D_{j^*} = [u^*]_d$.
$u^* \notin \underline{B_N^{k,\lambda}}(D_{j^*})$ implies that $B_N^{k,\lambda}(u^*) \not\subseteq D_{j^*}$.
Then $B_N^{k,\lambda}(u^*) \not\subseteq [u^*]_d$. So $B_N^{k,\lambda} \not\subseteq R_d$. This leads to a contradiction.
It follows that $\forall\, j$, $D_j \subseteq \underline{B_N^{k,\lambda}}(D_j)$. Then by Theorem 3.5, $\forall\, j$, $D_j = \underline{B_N^{k,\lambda}}(D_j)$. By Proposition 4.3, $\Gamma_{B,N}^{k,\lambda}(d) = 1$.

"(2) ⇒ (1)". Let $\Gamma_{B,N}^{k,\lambda}(d) = 1$. The fact that $\forall\, j, D_j = \underline{B_N^{k,\lambda}}(D_j)$ follows from Proposition 4.3. $\forall\, u \in U, \exists j^*$, $u \in D_{j^*}$. Since $D_{j^*} = \underline{B_N^{k,\lambda}}(D_{j^*})$, we have $u \in \underline{B_N^{k,\lambda}}(D_{j^*})$, Then $u \in B_N^{k,\lambda}(u) \subseteq D_{j^*}$. $u \in D_{j^*}$ implies that $D_{j^*} = [u]_d$. Hence $\forall\, u \in U$, $B_N^{k,\lambda}(u) \subseteq [u]_d$. Therefore $B_N^{k,\lambda} \subseteq R_d$.

"(1) ⇒ (3)". Let $B_N^{k,\lambda} \subseteq R_d$, then according to Lemma 4.24,

$$\forall\, i, \ \sum_{j=1}^{s} \frac{|B_N^{k,\lambda}(u_i) \cap D_j|}{n} log_2 \frac{|B_N^{k,\lambda}(u_i) \cap D_j|}{n} = \frac{|B_N^{k,\lambda}(u_i)|}{n} log_2 \frac{|B_N^{k,\lambda}(u_i)|}{n}.$$

This indicates that $H_N^{k,\lambda}(B \cup d) = H_N^{k,\lambda}(B)$. Then the fact that $H_N^{k,\lambda}(d|B) = 0$ follows from Proposition 4.9.

"(3) ⇒ (1)". Let $H_N^{k,\lambda}(d|B) = 0$. This shows that

$$\sum_{i=1}^{n} \sum_{j=1}^{s} \frac{|B_N^{k,\lambda}(u_i) \cap D_j|}{n} \log_2 \frac{|B_N^{k,\lambda}(u_i)|}{|B_N^{k,\lambda}(u_i) \cap D_j|} = 0.$$

Assume that $B_N^{k,\lambda} \not\subseteq R_d$. Then $\exists\, i^*$, $B_N^{k,\lambda}(u_{i^*}) \not\subseteq [u_{i^*}]_d$. Denote

$$[u_{i^*}]_d = D_{j^*}.$$

We have

$$|B_N^{k,\lambda}(u_{i^*})| > |B_N^{k,\lambda}(u_{i^*}) \cap D_{j^*}|.$$

It follows that

$$\frac{|B_N^{k,\lambda}(u_{i^*}) \cap D_{j^*}|}{n} \log_2 \frac{|B_N^{k,\lambda}(u_{i^*})|}{|B_N^{k,\lambda}(u_{i^*}) \cap D_{j^*}|} > 0.$$

Note that

$$\forall\, i, j, \ \ |B_N^{k,\lambda}(u_i)| \ge |B_N^{k,\lambda}(u_i) \cap D_j|.$$

Then

$$\forall\, i, j, \ \ \frac{|B_N^{k,\lambda}(u_i) \cap D_j|}{n} \log_2 \frac{|B_N^{k,\lambda}(u_i)|}{|B_N^{k,\lambda}(u_i) \cap D_j|} \ge 0.$$

It shows that

$$\sum_{i=1}^{n} \sum_{j=1}^{s} \frac{|B_N^{k,\lambda}(u_i) \cap D_j|}{n} \log_2 \frac{|B_N^{k,\lambda}(u_i)|}{|B_N^{k,\lambda}(u_i) \cap D_j|} > 0.$$

Hence $H_N^{k,\lambda}(d|B) > 0$. This is a contradiction.
Thus $B_N^{k,\lambda} \subseteq R_d$.
"(1) ⇒ (4)". Let $B_N^{k,\lambda} \subseteq R_d$.
Hence $B^k \subseteq R_d, N_B^{\lambda} \subseteq R_d$.
Since $R_d \subseteq N_B^{\lambda}$.
It follows that $N_B^{\lambda} = R_d, B_N^{k,\lambda} = B^k \cap N_B^{\lambda} = B^k \cap R_d = B^k$.
Clearly, $B_N^{k,\lambda} \cap R_d = B_N^{k,\lambda}$.

Thus $|B_N^{k,\lambda} \cap R_d| = |B_N^{k,\lambda}| = |B^k|$.
Therefore $DI_N^{k,\lambda}(d|B) = 0$.
"(4) ⇒ (1)". Suppose $DI_N^{k,\lambda}(d|B) = 0$.
It can be seen that $B_N^{k,\lambda} \cap R_d \subseteq B^k$ according to $B_N^{k,\lambda} \subseteq B^k$.
Since $DI_N^{k,\lambda}(d|B) = 0$ indicates $|B_N^{k,\lambda} \cap R_d| = |B^k|$.
It follows that $B^k = B_N^{k,\lambda} \cap R_d$.
Hence $B^k \subseteq N_B^\lambda \cap R_d$.
Thus $B^k \cap N_B^\lambda \subseteq N_B^\lambda \cap R_d \subseteq R_d$.
This shows that $B_N^{k,\lambda} \subseteq R_d$. □

Theorem 4.26. Consider an HIS (U, A, d), given $k \in \mathbb{N}$ *and* $\lambda \in [0, 1]$. The statements below are equivalent:
(1) (U, A, d) is (k, λ)-consistent;
(2) $A_N^{k,\lambda} \subseteq R_d$;
(3) $\Gamma_{A,N}^{k,\lambda}(d) = 1$;
(4) $H_N^{k,\lambda}(d|A) = 0$;
(5) $DI_N^{k,\lambda}(d|A) = 0$.

Proof. It can be evidenced using Propositions 3.8 and 4.25. □

Proposition 4.27. Consider an HIS (U, A, d), given $B \subseteq A$, $k \in \mathbb{N}$ *and* $\lambda \in [0, 1]$. If $B_N^{k,\lambda} \subseteq R_d$, then $E_N^{k,\lambda}(d|B) = 0$.

Proof. Let $B_N^{k,\lambda} \subseteq R_d$. Then according to Lemma 4.24,

$$\forall\ i,\ \sum_{j=1}^{s} \frac{|B_N^{k,\lambda}(u_i) \cap D_j|}{n}(1 - \frac{|B_N^{k,\lambda}(u_i) \cap D_j|}{n}) = \frac{|B_N^{k,\lambda}(u_i)|}{n}(1 - \frac{|B_N^{k,\lambda}(u_i)|}{n}).$$

Thus $E_N^{k,\lambda}(B \cup d) = E_N^{k,\lambda}(B)$. The fact that $E_N^{k,\lambda}(d|B) = 0$ follows from Proposition 4.16. □

Theorem 5.3. Assuming (U, A, d) is a (k, λ)-consistent HIS, with $k \in \mathbb{N}$ and $\lambda \in [0, 1]$, the following statements are equivalent:
(1) $B \in co_d^{k,\lambda}(A)$;
(2) $H_N^{k,\lambda}(d|B) = H_N^{k,\lambda}(d|A)$;
(3) $DI_N^{k,\lambda}(d|B) = DI_N^{k,\lambda}(d|A)$.

Proof. "(1) ⇒ (2)". Let $B \in co_d^{k,\lambda}(A)$. The fact that $\Gamma_{B,N}^{k,\lambda}(d) = \Gamma_{A,N}^{k,\lambda}(d)$ follows from Theorem 5.1. By Proposition 4.3,

$$\sum_{j=1}^{s}(|\underline{A_N^{k,\lambda}}(D_j)| - |\underline{B_N^{k,\lambda}}(D_j)|) = 0.$$

According to Theorem 3.5, $\forall\ j$, $\underline{A_N^{k,\lambda}}(D_j) \supseteq \underline{B_N^{k,\lambda}}(D_j)$. Then $\forall\ j$,

$$|\underline{A_N^{k,\lambda}}(D_j)| - |\underline{B_N^{k,\lambda}}(D_j)| \geq 0.$$

So $\forall\ j$,

$$|\underline{A_N^{k,\lambda}}(D_j)| - |\underline{B_N^{k,\lambda}}(D_j)| = 0.$$

It follows that $\forall\ j$,

$$\underline{A_N^{k,\lambda}}(D_j) = \underline{B_N^{k,\lambda}}(D_j).$$

Thus $\forall\ j$,

$$A_N^{k,\lambda}(u) \subseteq D_j \ \Leftrightarrow\ B_N^{k,\lambda}(u) \subseteq D_j.$$

Note that (U, A, d) is (k, λ)-consistent, as per Proposition 3.8, it implies that $A_N^{k,\lambda} \subseteq R_d$. Consequently, $\forall\ u \in U$, $A_N^{k,\lambda}(u) \subseteq [u]_d$. Assume that $[u]_d = D^u \in \{D_1, D_2, \ldots, D_s\}$. This leads to the conclusion that $\forall\ u \in U$, $B_N^{k,\lambda}(u) \subseteq D^u = [u]_d$. Therefore, $B_N^{k,\lambda} \subseteq R_d$. According to Proposition 4.25, $H_N^{k,\lambda}(d|B) = 0$.

Considering (U, A, d) being (k, λ)-consistent, based on Theorem 4.26, it follows that $H_N^{k,\lambda}(d|A) = 0$. Consequently, $H_N^{k,\lambda}(d|B) = H_N^{k,\lambda}(d|A)$.

"(2) ⇒ (3)". Let $H_N^{k,\lambda}(d|B) = H_N^{k,\lambda}(d|A)$. Notably, (U, A, d) being (k, λ)-consistent implies, by Theorem 4.26, that $H_N^{k,\lambda}(d|A) = DI_N^{k,\lambda}(d|A) = 0$. This leads to $H_N^{k,\lambda}(d|B) = 0$. Therefore

$$DI_N^{k,\lambda}(d|B) = DI_N^{k,\lambda}(d|A).$$

"(3) ⇒ (1)". Let $DI_N^{k,\lambda}(d|B) = DI_N^{k,\lambda}(d|A)$. Note that (U, A, d) is (k, λ)-consistent. Then $DI_N^{k,\lambda}(d|A) = 0$ follows from Theorem 4.26. Therefore $DI_N^{k,\lambda}(d|B) = 0$. By Proposition 4.25, $B_N^{k,\lambda} \subseteq R_d$. Suppose that $\exists\ j^*$,

$$\underline{A_N^{k,\lambda}}(D_{j^*}) \nsubseteq \underline{B_N^{k,\lambda}}(D_{j^*}).$$

Then $\underline{A_N^{k,\lambda}}(D_{j^*}) - \underline{B_N^{k,\lambda}}(D_{j^*}) \neq \emptyset$. Pick

$$u^* \in \underline{A_N^{k,\lambda}}(D_{j^*}) - \underline{B_N^{k,\lambda}}(D_{j^*}).$$

This indicates that

$$u^* \in \underline{A_N^{k,\lambda}}(D_{j^*}),\ u^* \notin \underline{B_N^{k,\lambda}}(D_{j^*}).$$

The fact that $u^* \in A_N^{k,\lambda}(u^*) \subseteq D_{j^*}$ follows from $u^* \in \underline{A_N^{k,\lambda}}(D_{j^*})$. Hence $D_{j^*} = [u^*]_d$. Then $B_N^{k,\lambda}(u^*) \nsubseteq D_{j^*}$ follows from $u^* \notin \underline{B_N^{k,\lambda}}(D_{j^*})$. Thus $B_N^{k,\lambda}(u^*) \nsubseteq [u^*]_d$. It follows that $B_N^{k,\lambda} \nsubseteq R_d$. This leads to a contradiction. Therefore $\forall\ j$,

$$\underline{A_N^{k,\lambda}}(D_j) \subseteq \underline{B_N^{k,\lambda}}(D_j).$$

According to Theorem 3.5, $\forall\ j$, $\underline{A_N^{k,\lambda}}(D_j) \supseteq \underline{B_N^{k,\lambda}}(D_j)$. Hence $\forall\ j$,

$$\underline{A_N^{k,\lambda}}(D_j) = \underline{B_N^{k,\lambda}}(D_j).$$

Thus

$$POS_{A,N}^{k,\lambda}(d) = \bigcup_{j=1}^{s} \underline{A_N^{k,\lambda}}(D_j) = \bigcup_{j=1}^{s} \underline{B_N^{k,\lambda}}(D_j) = POS_{B,N}^{k,\lambda}(d).$$

Hence $B \in co_d^{k,\lambda}(A)$. □

Data availability

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

References

[1] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning About Data, Kluwer Academic Publishers, Dordrecht, 1991.
[2] D. Miao, Information representation of the concepts and operations in rough set theory, J. Softw. 22 (1999) 113–116.
[3] J.Y. Liang, Z.Z. Shi, D. Li, M.J. Wierman, Information entropy, rough entropy and knowledge granulation in incomplete information systems, Int. J. Gen. Syst. 35 (2006) 641–654.
[4] X.D. Wang, Y.F. Song, Uncertainty measure in evidence theory with its applications, Appl. Intell. 48 (2018) 1672–1688.
[5] J.S. Zeng, Z.W. Li, P.F. Zhang, P. Wang, Information structures and uncertainty measures in a hybrid information system: gaussian kernel method, Int. J. Fuzzy Syst. 22 (2020) 212–231.
[6] S.C. Zhang, Cost-sensitive KNN classification, Neurocomputing 391 (2020) 234–242.
[7] W. Li, Y.M. Chen, Y.P. Song, Boosted K-nearest neighbor classifiers based on fuzzy granules, Knowl.-Based Syst. 195 (2020) 105606.
[8] D. Huang, Y.Y. Chen, F. Liu, Z.W. Li, Feature selection for multiset-valued data based on fuzzy conditional information entropy using iterative model and matrix operation, Appl. Soft Comput. 142 (2023) 110345.
[9] F. Jiang, Y.M. Chen, Outlier detection based on granular computing and rough set theory, Appl. Intell. 42 (2015) 303–322.
[10] J.H. Dai, H.W. Tian, Entropy measures and granularity measures for set-valued information systems, Inform. Sci. 240 (2013) 72–82.
[11] H.R. Su, J.K. Chen, Y.J. Lin, A four-stage branch local search algorithm for minimal test cost attribute reduction based on the set covering, Appl. Soft Comput. J. 153 (2024) 111303.
[12] X. Ji, J.H. Peng, P. Zhao, S. Yao, Extended rough sets model based on fuzzy granular ball and its attribute reduction, Inform. Sci. 640 (2023) 119071.
[13] D. Zhang, P. Zhu, Variable radius neighborhood rough sets and attribute reduction, Internat. J. Approx. Reason. 150 (2022) 98–121.

[14] F. Xu, M.J. Cai, Q.G. Li, H.C. Wang, H. Fujita, Shared neighbors rough set model and neighborhood classifiers, Expert Syst. Appl. 244 (2024) 122965.
[15] X.B. Yang, S.C. Liang, H.L. Yu, S. Gao, Y.H. Qian, Pseudo-label neighborhood rough set: Measures and attribute reductions, Internat. J. Approx. Reason. 105 (2019) 112–129.
[16] T.X. Shu, Y.J. Lin, L. Guo, Online hierarchical streaming feature selection based on adaptive neighborhood rough set, Appl. Soft Comput. J. 152 (2024) 111276.
[17] C.Z. Wang, Y.P. Shi, X.D. Fan, M.W. Shao, Attribute reduction based on k-nearest neighborhood rough sets, Internat. J. Approx. Reason. 106 (2019) 18–31.
[18] N. Wang, E.H. Zhao, A new method for feature selection based on weighted k-nearest neighborhood rough set, Expert Syst. Appl. 238 (2024) 122324.
[19] S.Y. Xia, C. Wang, G.Y. Wang, X.B. Gao, W.P. Ding, J.H. Yu, Y.J. Zhai, Z.Z. Chen, GBRS: A unified granular-ball learning model of pawlak rough set and neighborhood rough set, IEEE Trans. Neural Netw. Learn. Syst. (2024) 1–15.
[20] Y.B. Wang, X.J. Chen, K. Dong, Attribute reduction via local conditional entropy, Int. J. Mach. Learn. Cybern. 10 (2019) 3619–3634.
[21] C.Z. Wang, Y. Wang, M.W. Shao, Y.H. Qian, D.G. Chen, Fuzzy rough attribute reduction for categorical data, IEEE Trans. Fuzzy Syst. 28 (2020) 818–830.
[22] B.B. Sang, H.M. Chen, L. Yang, T.R. Li, Incremental feature selection using a conditional entropy based on fuzzy dominance neighborhood rough sets, IEEE Trans. Fuzzy Syst. 30 (2022) 1683–1697.
[23] D. Huang, Q.L. Zhang, Z.W. Li, Semi-supervised attribute reduction for partially labeled categorical data based on predicted label, Internat. J. Approx. Reason. 154 (2023) 242–261.
[24] M.A.N.D. Sewwandi, Y.F. Li, J.L. Zhang, A class-specific feature selection and classification approach using neighborhood rough set and K-nearest neighbor theories, Appl. Soft Comput. 143 (2023) 110366.
[25] H.Y. Zhang, Q.Q. Sun, K.Z. Dong, Information-theoretic partially labeled heterogeneous feature selection based on neighborhood rough sets, Internat. J. Approx. Reason. 154 (2023) 200–217.
[26] S.G. Cui, G.S. Li, B.B. Sang, W.H. Xu, H.M. Chen, Distance metric learning-based multi-granularity neighborhood rough sets for attribute reduction, Appl. Soft Comput. J. 159 (2024) 111656.
[27] M.A.N.D. Sewwandi, Y.F. Li, J.L. Zhang, Granule-specific feature selection for continuous data classification using neighborhood rough sets, Expert Syst. Appl. 238 (2024) 121765.
[28] S.Z. Wu, L.T. Wang, S.Y. Ge, Z.W. Hao, Y.L. Liu, Neighborhood rough set with neighborhood equivalence relation for feature selection, Knowl. Inf. Syst. 66 (2024) 1833–1859.
[29] Q.L. Zhang, Y. Song, Y.C. Peng, Z.W. Li, Attribute reduction algorithms with an anti-noise mechanism for hybrid data based on fuzzy evidence theory, Eng. Appl. Artif. Intell. 129 (2024) 107659.
[30] M. Hu, E.C.C. Tsang, Y.T. Guo, D.G. Chen, W.H. Xu, Attribute reduction based on overlap degree and k-nearest-neighbor rough sets in decision information systems, Inform. Sci. 584 (2022) 301–324.
[31] P. Wang, J.L. He, Z.W. Li, Attribute reduction for hybrid data based on fuzzy rough iterative computation model, Inform. Sci. 632 (2023) 555–575.
[32] Q.H. Hu, J.F. Liu, D.R. Yu, Mixed feature selection based on granulation and approximation, Knowl.-Based Syst. 21 (2008) 294–304.
[33] M. Kryszkiewicz, Rules in incomplete information systems, Inform. Sci. 113 (1999) 271–292.
[34] X.B. Yang, Y.Y. Yao, Ensemble selector for attribute reduction, Appl. Soft Comput. 70 (2018) 1–11.
[35] Z.W. Li, L.D. Qu, G.Q. Zhang, N.X. Xie, Attribute selection for heterogeneous data based on information entropy, Int. J. Gen. Syst. 50 (2021) 548–566.
[36] J.W. Grzymala-Busse, M. Hu, A comparison of several approaches tomissing attribute values in data mining, in: International Conferenceon Rough Sets and Current Trends in Computing, Springer, Berlin, Heidelberg, 2001, pp. 378–385.
[37] M. Friedman, A comparison of alternative tests of significance for the problem of m rankings, Ann. Math. Stat. 11 (1940) 86–92.
[38] O.J. Dunn, Multiple comparisons among means, J. Amer. Statist. Assoc. 56 (1961) 52–64.
[39] J. Demsar, Statistical comparisons of classifiers over multiple datasets, J. Mach. Learn. Res. 7 (2006) 1–30.
[40] A. Benavoli, C. Giorgio, M. Francesca, Should we really use post-hoc tests based on mean-ranks? J. Mach. Learn. Res. 17 (2016) 152–161.
[41] F. Wilcoxon, Individual comparisons by ranking methods, Biom. 1 (1945) 80–83.