International Journal of Approximate Reasoning 83 (2017) 371–384

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Tri-partition neighborhood covering reduction for robust classification ☆

Xiaodong Yue [a], Yufei Chen [b,*], Duoqian Miao [b], Jin Qian [c]

[a] School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China
[b] College of Electronic and Information Engineering, Tongji University, Shanghai 201804, China
[c] School of Computer Engineering, Jiangsu University of Technology, Changzhou 213001, China

ARTICLE INFO

Article history:
Received 20 June 2016
Received in revised form 4 October 2016
Accepted 10 November 2016
Available online 24 November 2016

Keywords:
Neighborhood covering reduction
Three-way decision
Rule extraction

ABSTRACT

Neighborhood Covering Reduction extracts rules for classification through formulating the covering of data space with neighborhoods. The covering of neighborhoods is constructed based on distance measure and strictly constrained to be homogeneous. However, this strategy over-focuses on individual samples and thus makes the neighborhood covering model sensitive to noise and outliers. To tackle this problem, we construct a flexible Tri-partition Neighborhood for robust classification. This novel neighborhood originates from Three-way Decision theory and is partitioned into the regions of certain neighborhood, neighborhood boundary and non-neighborhood. The neighborhood boundary consists of uncertain neighbors and is helpful to tolerate noise. Besides the neighborhood construction, we also proposed complete and partial strategies to reduce redundant neighborhoods to optimize the neighborhood covering for classification. The reduction process preserves lower and upper approximations of neighborhood covering and thereby provides a flexible way to handle uncertain samples and noise. Experiments verify the classification based on tri-partition neighborhood covering is robust and achieves precise and stable results on noisy data.

1. Introduction

In the area of soft computing, Rough Set theory has been proven to be an effective tool for uncertain data analysis [27–29]. Classic rough set model specializes in feature selection and rule extraction from the table-formed symbolic data and thus is widely applied in the tasks of structural data mining, concept learning and rule-based expert systems [7,24,31]. To handle the data of both symbolic and numerical features, rough set model was extended to neighborhood systems [15, 20,21,37]. Different from the equivalence classes defined for symbols in classic rough sets, the neighborhoods in numerical data space are basic granules to constitute Neighborhood Rough Sets [35,44]. For a sample, its neighborhood consists of the neighbors surrounding it [22,25,32]. Neighborhood rough sets actually provide us a way to formulate data space on neighborhood level [20,43]. From the view of topology, it has been proven that the concepts derived from neighborhood spaces are more general than from data-level spaces [21]. This indicates that transforming original data space into a neighborhood system will facilitate the data generalization [35,37].

☆ This paper is part of the virtual special issue on Tri-partition, edited by Davide Ciucci and Yiyu Yao.

* Corresponding author.
E-mail address: april337@163.com (Y. Chen).

http://dx.doi.org/10.1016/j.ijar.2016.11.010
0888-613X/

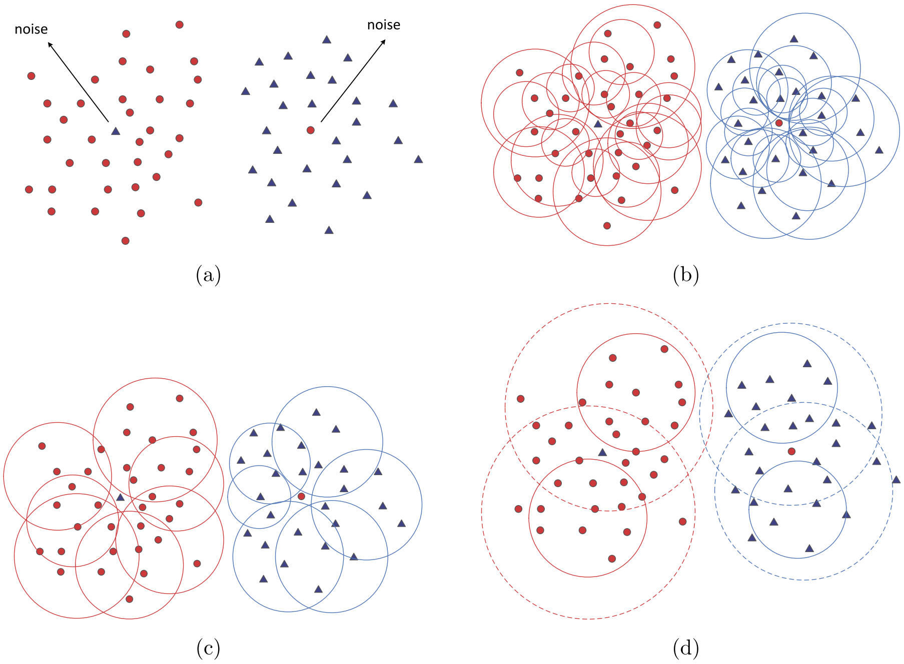

Fig. 1. Neighborhood covering reduction on noisy data, (a) original data with noise, (b) initial neighborhood covering, (c) reduced neighborhood covering, (d) flexible covering of tri-partition neighborhoods.

To formulate neighborhood systems of data space, neighborhood rough sets were proposed to form neighborhood coverings of data samples and the union of neighborhoods constitute an approximate representation of data space [16,26,33]. Based on this, the methods of Neighborhood Covering Reduction (NCR) were further proposed to approximate data space with neighborhoods for learning tasks [6,36]. Specifically, the samples are initially grouped into neighborhoods based on distance measure and then redundant neighborhoods are iteratively reduced to generate a concise covering of data space for classification [12,41] and feature selection [14,15]. In most existing neighborhood covering models, neighborhoods are constrained to be homogeneous, i.e. all the samples in a neighborhood should belong to the same class, and the boundary of neighborhood is decided by the nearest heterogeneous sample. However, this strategy of neighborhood construction over-focuses on individual samples and thus makes the model sensitive to noise and outliers. Fig. 1(a–c) shows the disturbance of noise to the formulation of neighborhood covering. Because of the embedded noisy data, the neighborhoods over-partition the data space and the distribution of samples belonging to the same class is broken up.

Aiming to handle the problem above, we expect to construct more flexible neighborhoods for robust classification. The methodology of neighborhood construction originates from the theory of Three-way Decision (3WD) [38–40]. Generally speaking, in the process of three-way decision, knowledge are extracted from the data with uncertainty through tri-partitioning data space into Positive, Negative and Boundary regions [1,10,13]. From the decision view, the three regions correspond to the cases of acceptance, rejection and non-commitment (uncertain case) respectively [5,19,23]. Similarly, in a neighborhood system, a flexible neighborhood should contain certain and uncertain neighbors to tolerate noise and thereby could be tri-partitioned into the regions of certain neighborhood, uncertain boundary and non-neighborhood. The certain neighborhood consists of the neighbors certainly belonging to the same class, the neighborhood boundary involves uncertain neighbors may belong to different classes and the non-neighborhood region denotes the samples outside the neighborhood.

As introduced above, motivated by three-way decision, we extend traditional neighborhood to Tri-partition Neighborhood, briefly denoted by T-neighborhood, to improve the robustness of neighborhood-based model for classification. T-neighborhood consists of inner and outer parts, which involve certain and uncertain neighbors respectively, see Fig. 1(d). For a sample, its inner and outer neighborhoods are decided according to the proportion of the heterogeneous samples surrounding it. As neighborhood coverings form an approximation of data space, the coverings of T-neighborhoods will lead to multilevel approximations of data space. Specifically, the union of inner neighborhoods constitutes the Lower Approximation of data space, in which the samples in neighborhoods certainly belong to a class. And the outer neighborhoods comprise the Upper Approximation of data space, in which the samples in neighborhoods belong to a class with uncertainty. With the lower and upper approximations of data space, the covering of T-neighborhoods is able to tolerate uncertain samples and thus weaken the disturbance of noise.

Besides the neighborhood construction, the reduction strategies of T-neighborhoods are also proposed to optimize neighborhood coverings for classification. The reduction objectives aim to filter out redundant neighborhoods while preserving the approximations of data space. In the reduction process, T-neighborhoods can be categorized into complete/partial reducible and irreducible. The classifier based on irreducible T-neighborhoods generalizes data space well and is robust for

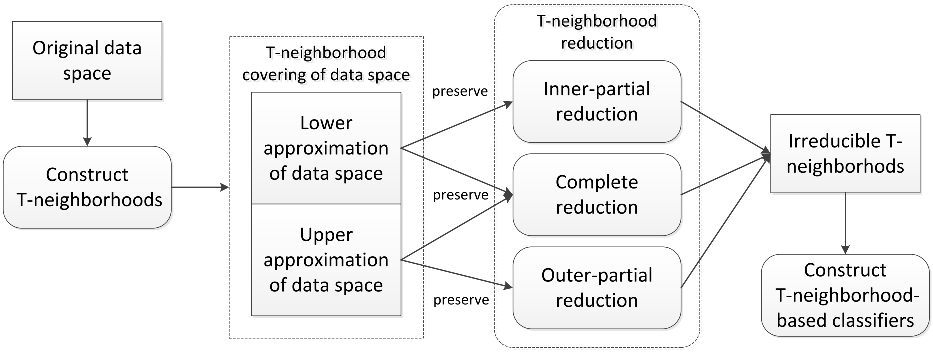

Fig. 2. Workflow of robust classification with T-neighborhoods.

noisy data classification. Fig. 2 illustrates the workflow of robust classification with T-neighborhoods. First, heterogeneous T-neighborhoods are constructed to tolerate noise and the coverings of T-neighborhoods form the lower and upper approximations of data space. Second, to simplify the coverings of T-neighborhoods, initial T-neighborhoods are reduced to preserve the lower and upper approximations of data space respectively. Finally, after the reduction of T-neighborhoods, the irreducible T-neighborhoods are united to constitute the classifiers for noisy data classification. The contributions of this paper are summarized as follows.

- *Construct Tri-partition Neighborhood (T-neighborhood).* T-neighborhood consists of inner and outer neighborhood regions which involve both certain and uncertain neighbors. The coverings of T-neighborhoods form lower and upper approximations of data space.
- *Propose complete and partial reduction strategies for T-neighborhoods.* The reduction of T-neighborhoods generates concise neighborhood coverings to preserve the approximations of data space.
- *Design a classifier based on T-neighborhood covering reduction.* The classifier consists of the irreducible T-neighborhoods and distinguishes samples according to their distances from multiple neighborhood regions.

The remainder of this paper is organized as follows. Section 2 introduces the related work. Section 3 describes the methodology of T-neighborhood construction. Section 4 presents several strategies of T-neighborhood covering reduction and a workflow of classification with irreducible T-neighborhoods. In Section 5, experimental results validate the robustness of the proposed T-neighborhood covering model for noisy data classification. The work conclusion is given in Section 6.

2. Related work

Neighborhood covering reduction

First we briefly introduce the preliminaries of neighborhood covering reduction. There are two kinds of neighborhood covering reduction methods. The first one aims to reduce redundant neighborhoods in a covering to find the minimal data description. These reduction methods are generally used for rule learning [6,12,36,41]. Another kind of reduction methods aims to reduce redundant coverings from a family of coverings and is usually related to feature selection [14,15]. In this paper, we just focus on the neighborhood covering reduction for rule learning. The covering of neighborhoods is formally defined as follows.

Definition 1. Neighborhood covering. Suppose U is the data space $\{x_1, x_2, ...x_n\}$ and $O(x)$ is the neighborhood of a sample x, $O(x) = \{x_i | \Delta(x, x_i) \leq \lambda\}$, $\Delta(\cdot)$ is a distance function and λ denotes a threshold. The universal set of neighborhoods is $O = \{O(x_1), O(x_2), ..., O(x_n)\}$ and the union $O = \cup_{i=1}^{n} O(x_i)$ forms a covering of data space. The neighborhood covering is denoted by $C =< U, O >$.

The neighborhoods in a covering overlap each other and some of them may be redundant to maintain the structure of data space. In order to extract the essential structure of data space, it is necessary to reduce redundant neighborhoods to generate concise neighborhood coverings.

Definition 2. Neighborhood covering reduction. Let $C =< U, O >$ be a neighborhood covering of data space, for a neighborhood $O(x) \in O$, remove it from the neighborhood set and obtain $O' = O - \{O(x)\}$, if the union $\cup_{O(x_i) \in O} O(x_i) = \cup_{O(x_j) \in O'} O(x_j)$, the neighborhood $O(x)$ is considered reducible, otherwise $O(x)$ is irreducible. In addition, if every element in the neighborhood set O is irreducible, the covering $C =< U, O >$ is irreducible, otherwise C is reducible.

Definition 3. Relative neighborhood covering reduction. Let $C = < U, O >$ be a neighborhood covering, $X \subseteq U$ is a set of samples and $O(x_i) \in O$ is a neighborhood, if $\exists O(x_j) \in O$, such that $O(x_i) \subseteq O(x_j) \subseteq X$, we consider $O(x_i)$ is a relatively reducible neighborhood with respect to X, otherwise $O(x_i)$ is relatively irreducible. If all the neighborhoods in C are relatively irreducible, the neighborhood covering C is relatively irreducible.

The relative reduction of neighborhoods will produce a concise neighborhood covering for a class of samples and thus can be used for data classification. However, it is required that all the samples in a reduced neighborhood certainly belong to the same class. This constraint of pure neighborhood homogeneity may lead to over complex partition of data space and make the neighborhood-based classification sensitive to noise and outliers. To tackle this problem, we expect to adopt the Tri-Partition Methodology to construct more flexible neighborhoods for robust classification.

Tri-partition methodology

The basic idea of tri-partition methodology is to divide a universal set into three pair-wise disjoint regions which denote the certain and uncertain parts in problem domain [4,40]. Tri-partition methodology is built on solid cognitive foundations and provides flexible ways for human-like problem solving and information processing [10]. As typical approaches, Three-Way Decisions (3WD) [38,39], Orthopairs [2] and Hexagon of Opposition [8] represent knowledge and perform reasoning through tri-partitioning the universe. These approaches have been applied to extend the design and implementation of intelligent systems and the investigations of tri-partition methodology are gaining interest [17,18,30,42].

Three-Way Decisions (3WD) is an extension of the commonly used binary-decision model through adding a third option [38]. The approach of Three-Way Decisions divides the universe into the Positive, Negative and Boundary regions which denote the regions of acceptance, rejection and non-commitment for ternary classifications [39]. Specifically, for the objects partially satisfy the classification criteria, it is difficult to directly identify them without uncertainty. Instead of making a binary decision, we use thresholds on the degrees of satisfiability to make one of three decisions: accept, reject, non-commitment. The third option may also be referred to as a deferment decision that requires further judgments. With the ordered evaluation of acceptance, the three regions are formally defined as

Definition 4. Three-way decision with ordered set. Suppose $(L, \preceq)$ is a totally ordered set, in which $\preceq$ is a total order. For two thresholds α, β with $\alpha \prec \beta$, suppose that the set of designated values for acceptance is given by $L^+ = \{t \in L | t \succeq \alpha\}$ and the set for rejection is $L^- = \{b \in L | b \preceq \beta\}$. For an evaluation function $v : U \to L$, its three regions are defined by:

$$\begin{aligned} POS_{\alpha,\beta}(v) &= \{x \in U | v(x) \succeq \alpha\}, \\ NEG_{\alpha,\beta}(v) &= \{x \in U | v(x) \preceq \beta\}, \\ BND_{\alpha,\beta}(v) &= \{x \in U | \alpha \prec v(x) \prec \beta\}. \end{aligned} \tag{1}$$

Many soft computing models for leaning uncertain concepts, such as Interval Sets, Many-valued Logic, Rough Sets, Fuzzy Sets and Shadowed Sets, have the tri-partitioning properties and can be reinvestigated within the framework of three-way decisions [40]. At present, the research issues of Three-Way Decisions focus on the strategies for trisecting a universe, the evaluation functions of acceptance/rejection, the optimization and interpretation of the thresholds etc.

Orthopair consists of a pair of disjoint sets $O = (P, N)$ which commonly exists in many tools for managing data uncertainty. The set P and N stand for the positive and negative regions and an orthopair tri-partitions the universe into three regions $O = (P, N, (P \cup N)^c)$, in which the last term denotes the boundary region Bnd. Combining the three regions to construct the orthopairs such as (P, Bnd) and (P, N^c), we can obtain multiple set approximations to abstract concepts at multiple levels [2]. Orthopair has strict links to Three-valued Logics and can be generalized to Atanassov Intuitionistic Fuzzy Sets, Possibility Theory and Three-way Decision [3]. It actually provides us a common representation to formulate the partial knowledge, positive/negative examples and trust/distrust for uncertain reasoning. The orthopairs and their hierarchical structures are also discussed in the light of Granular Computing [4].

An opposition is a relation between two logical statements expressing an opposite point of view. Square of Opposition is a diagram representing the relations between four propositions or four concepts. The origin of the square can be traced back to Aristotle making the distinction between two oppositions: contradiction and contrariety. The traditional square of opposition has been generalized to the Hexagon of Opposition through adding new kinds of oppositions into the relationship diagram. As explained by Dubois and Prade, a hexagon of opposition can be obtained by any tri-partition of a universe, hence by any orthopair. Given an orthopair (P, N), the six vertices of the hexagon are $(P, N, Bnd, Upp, P \cup N, P^c)$. The different links between the vertices represent different kinds of oppositions. Hexagon of opposition has been used to discover new paradigm in formal concept analysis [8].

Although the tri-partition methodology has been investigated in many areas, its application in neighborhood systems is still limited. In this paper, we expect to apply the tri-partition methodology to construct a flexible neighborhood, in which the inner neighborhood denotes the positive region, and the gap between inner and outer neighborhoods denotes the boundary region. Comparing with traditional certain neighborhoods, in the tri-partition neighborhoods, the boundary regions consist of uncertain neighbors and facilitate handling the noise.

3. Tri-partition neighborhood

To construct a neighborhood, we first measure the distance between samples. For the ubiquitous mixed-type data with the attributes of symbolic and numerical value domain, we adopt Heterogeneous Euclidean-Overlap Metric (HEOM) [34] to measure the sample similarity according to the following formula.

$$\Delta(x,y)=\sqrt{\sum_{i=1}^{m} w_{a_i}\times d^2_{a_i}(x_{a_i},y_{a_i})} \quad (2)$$

in which m is the number of attributes, w_{a_i} is the weight to present the significance of attribute a_i, d_{a_i} denotes the distance between samples x and y with respect to attribute a_i and is defined as

$$d_{a_i}(x,y)=\begin{cases} overlap_{a_i}(x,y), & \text{if } a_i \text{ is a symbolic attribute} \\ rn\_diff_{a_i}(x,y), & \text{if } a_i \text{ is a numerical attribute} \end{cases} \quad (3)$$

$$overlap_{a_i}(x,y)=\begin{cases} 0, & \textit{if } a_i(x)=a_i(y) \\ 1, & \textit{otherwise} \end{cases} \quad (4)$$

$$rn\_diff_{a_i}(x,y)=\frac{|a_i(x)-a_i(y)|}{\max_{a_i}-\min_{a_i}} \quad (5)$$

Focusing on the methodology of neighborhood construction, in this paper, we do not take the attribute significance into consideration and set all the attribute weights $w_{a_i}=1$. Based on the defined distance measure, we construct a neighborhood in the following way.

Definition 5. Neighborhood. Given a sample $x\in U$, the neighborhood $O(x)$ consists of the samples surrounding x.

$$O(x)=\{y|\Delta(x,y)\le\eta, y\in U\} \quad (6)$$

where Δ is HEOM distance function and η is the distance threshold to present the radius of neighborhood.

To guarantee the neighborhood homogeneity, the radius of neighborhood $O(x)$ is computed according to the distances between x and its nearest homogeneous and heterogeneous samples [6,11]. Given a sample $x\in U$, its Nearest Hit $NH(x)\in U-\{x\}$ is defined as the nearest sample belonging to the same class. For the class of only one sample, we set $NH(x)=x$. On the contrary, $NM(x)$ denotes the nearest sample to x of different class and is named the Nearest Miss. The radius of neighborhood $O(x)$ can be directly obtained by $\eta=\Delta(x,NM(x))-\text{constant}\times\Delta(x,NH(x))$. Obviously, all the samples located within the neighborhood of radius η belong to the same class as x.

This strategy of neighborhood construction depends too much on local samples and the constraint of pure homogeneity makes neighborhoods sensitive to noise. Therefore we try to extend the strict homogeneous neighborhood to a flexible one. The extended neighborhood consists of certain and uncertain neighbors and partitions data space into three regions. We construct the Tri-Partition Neighborhood (T-neighborhood) as follows.

Definition 6. Tri-partition neighborhood (T-neighborhood). Given a sample $x\in U$, the tri-partition neighborhood $O_\beta(x)$ of x is defined as

$$O_\beta(x)=\{y|\Delta(x,y)\le\eta^\beta, y\in U\} \quad (7)$$

where Δ is distance function, $\beta\in[0,1)$ is a parameter to control the heterogeneity of neighborhood. Specifically, β is the proportion of the samples in neighborhood belonging to different classes respect to x. η^β denotes the radius of neighborhood of heterogeneity degree β.

Set the parameter $\beta=r$ $(0<r<1)$, according to the heterogeneity degree, the data space surrounding x is partitioned into inner-neighborhood, outer-neighborhood and non-neighborhood regions. The inner-neighborhood corresponds to the homogeneous region in neighborhood and is formulated as follows.

Definition 7. Inner neighborhood. The inner part of tri-partition neighborhood of x is defined as

$$\underline{O}(x)=O_{\beta=0}(x)=\{y|\Delta(x,y)\le\eta^{\beta=0}, y\in U\} \quad (8)$$

The heterogeneity degree $\beta=0$ means all the neighboring samples within the radius $\eta^{\beta=0}$ belong to the same class. The region of inner neighborhood $O_{\beta=0}(x)$ is homogeneous. We use $\underline{O}(x)$ to denote the inner neighborhood for short, $\underline{O}(x)$ forms an inner hyper sphere centered at x.

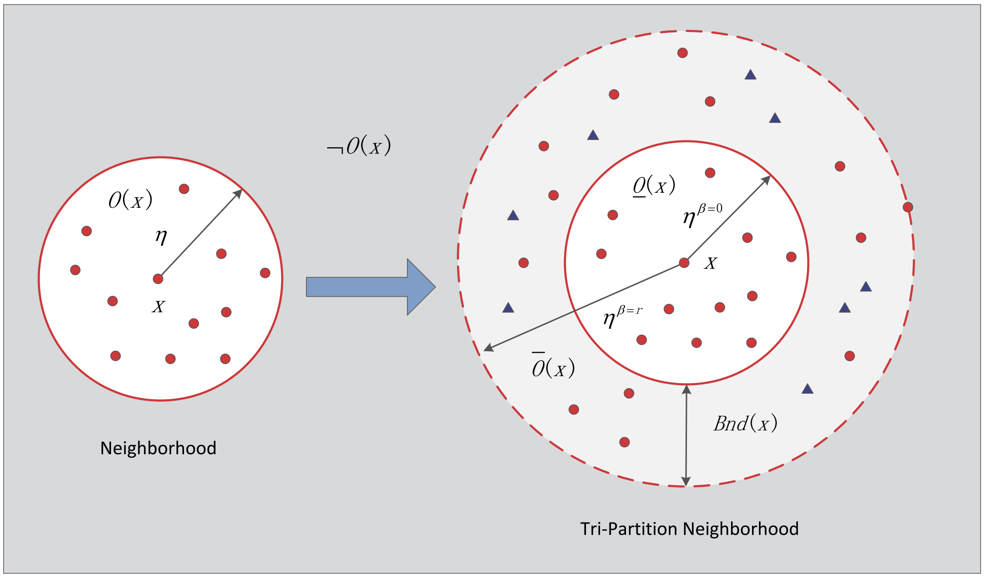

Fig. 3. From neighborhood to tri-partition neighborhood.

The samples within inner-neighborhood certainly belong to the same class. Further increasing the radius of inner-neighborhood, heterogeneous samples may be included. When the heterogeneity degree of neighborhood reaches the predefined threshold, the inner-neighborhood is extended to the outer one.

Definition 8. Outer neighborhood. Given a heterogeneity degree value r, $0 < r < 1$, the outer neighborhood of x is defined as

$$\overline{O}(x) = O_{\beta=r}(x) = \{y | \Delta(x, y) \le \eta^{\beta=r}, y \in U\} \tag{9}$$

The heterogeneity degree $\beta = r$ is the proportion of the samples within outer neighborhood belong to different classes. The region of outer neighborhood $O_{\beta=r}(x)$ is heterogeneous. We use $\overline{O}(x)$ to denote the outer neighborhood for short, $\overline{O}(x)$ forms an outer hyper sphere centered at x. We set $r = 0.1$ by default to construct the outer neighborhoods.

It is obvious that $\overline{O}(x) \supseteq \underline{O}(x)$. The inner-neighborhood $\underline{O}(x)$ is homogeneous and consists of the neighbors certainly belonging to the same class. The outer-neighborhood $\overline{O}(x)$ is heterogeneous and considered as the neighborhood extension with uncertainty. $Bnd(x) = \overline{O}(x) - \underline{O}(x)$ is the boundary to represent the uncertain part of neighborhood. Moreover, the samples outside outer-neighborhood constitute the non-neighborhood region, i.e. $\neg O(x) = O_{\beta>r}(x) = \{y | \Delta(x, y) > \eta^{\beta=r}, y \in U\}$. Fig. 3 shows all the regions in T-neighborhood. The novel neighborhood is formally represented by the triple $O(x) = \{\underline{O}(x), \overline{O}(x), \neg O(x)\}$ and briefly denoted as $O(x) = \{\underline{O}(x), \overline{O}(x)\}$.

Similar to neighborhood construction, the radii of T-neighborhood are obtained from the distances of the nearest homogeneous and heterogeneous samples. Given a sample $x \in U$, the radius of its inner-neighborhood is computed based on the Nearest Hit and Nearest Miss as follows.

$$\eta^{\beta=0} = \Delta(x, NM(x)) - \Delta(x, NH(x))/10^3 \tag{10}$$

The inner-neighborhood can be extended through involving the samples belonging to the same class as x. Adding a homogeneous sample into the neighborhood, if the radius increases and the heterogeneity degree of neighborhood β reaches the threshold r, this sample is named $FM(x)$, i.e. the Farthest Miss of x with respect to r. $FM(x)$ is the farthest homogeneous sample from x to form the r-proportion heterogeneous neighborhood. The radius of outer neighborhood is computed as

$$\eta^{\beta=r} = \Delta(x, FM(x)) \tag{11}$$

It should be noticed that the T-neighborhood is a generalization of the traditional neighborhood. When the neighborhood heterogeneity degree $\beta = 0$ and the radius $\eta = \eta^{\beta=0}$, the tri-partition neighborhood will degrade to the traditional one. Fig. 4 illustrates the construction of T-neighborhood.

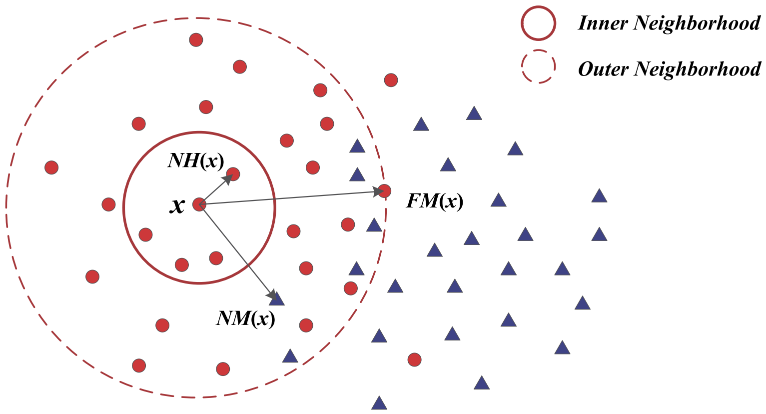

Fig. 4. Tri-partition neighborhood construction.

4. Tri-partition neighborhood covering reduction for classification

4.1. Tri-partition neighborhood covering reduction

As the union of neighborhoods forms a covering of data space, the union of T-neighborhoods induces multilevel approximations of data space covering. The approximations of covering include Lower Covering Approximation consisting of the inner parts of neighborhoods and Upper Covering Approximation of the outer neighborhoods. For the samples of a specific class, the lower covering approximation of their T-neighborhoods presents the data space certainly belonging to the class while the upper covering approximation presents the data space of the same class with uncertainty. We formally define the approximations of T-neighborhood covering as follows.

Definition 9. Neighborhood covering approximation. In a data space, for a set of samples with class d, $X_d \subseteq U$, the unions of inner-neighborhoods and outer-neighborhoods of X_d constitute the lower and upper approximations of the neighborhood covering of class d.

$$\begin{aligned} &\textit{Lower covering approximation}: \underline{N}(X_d) = \cup_{x\in X_d}\underline{O}(x)\\ &\textit{Upper covering approximation}: \overline{N}(X_d) = \cup_{x\in X_d}\overline{O}(x) \end{aligned} \tag{12}$$

$\underline{N}(X_d)$ and $\overline{N}(X_d)$ consist of the certain and uncertain neighborhoods respectively and provide multilevel approximations of data space of class d.

Not all the neighborhoods are necessary to constitute the covering of specified data space. The redundant neighborhoods should be reduced to generate the concise covering of data space. The reduction process of neighborhoods aims to filter out reducible neighborhoods. Different from the reduction of traditional neighborhoods, in a T-neighborhood system, the reducible T-neighborhoods are categorized into complete and partial reducible neighborhoods.

Definition 10. Complete reducible neighborhood. Given two T-neighborhoods of class d, $O(x_i) = \{\underline{O}(x_i), \overline{O}(x_i)\}$, $O(x_j) = \{\underline{O}(x_j), \overline{O}(x_j)\}$, $x_i, x_j \in X_d$, the neighborhood $O(x_j)$ is considered to be completely reducible with respect to $O(x_i)$ iff $O(x_j)$ is completely included in $O(x_i)$. The complete neighborhood inclusion is formulated as

$$O(x_j) \subseteq O(x_i) \Leftrightarrow \underline{O}(x_j) \subseteq \underline{O}(x_i) \wedge \overline{O}(x_j) \subseteq \overline{O}(x_i) \tag{13}$$

Under the complete reduction strategy, a neighborhood is reducible if both its inner and outer neighborhoods are correspondingly included in another one. Fig. 5(a) illustrates a complete reducible T-neighborhood.

Definition 11. Partial reducible neighborhood. Given two T-neighborhoods $O(x_i) = \{\underline{O}(x_i), \overline{O}(x_i)\}$ and $O(x_j) = \{\underline{O}(x_j), \overline{O}(x_j)\}$, $x_i, x_j \in X_d$, the neighborhood $O(x_j)$ is considered to be partially reducible with respect to $O(x_i)$ iff $O(x_j)$ is partially included in $O(x_i)$. The partial inclusion of T-neighborhood involves outer-partial inclusion and inner-partial inclusion.

$$\begin{aligned} &\textit{Outer-partial inclusion}: O(x_j) \overset{out}{\subseteq} O(x_i) \Leftrightarrow \overline{O}(x_j) \subseteq \overline{O}(x_i)\\ &\textit{Inner-partial inclusion}: O(x_j) \overset{in}{\subseteq} O(x_i) \Leftrightarrow \underline{O}(x_j) \subseteq \underline{O}(x_i) \end{aligned} \tag{14}$$

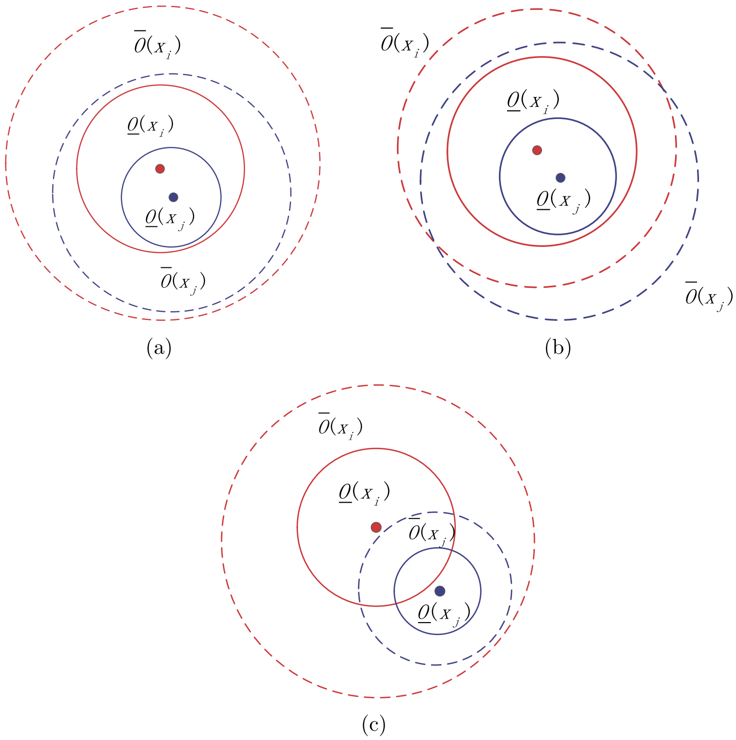

Fig. 5. Reduction strategies of tri-partition neighborhoods (a) complete reducible neighborhood (b) inner-partial reducible neighborhood (c) outer-partial reducible neighborhood.

Under the partial reduction strategy, a neighborhood is partially reducible if its inner or outer neighborhood is included in the corresponding part of another neighborhood. Fig. 5(b) and Fig. 5(c) illustrate the inner and outer partial reducible neighborhoods.

The formulation of complete reducible neighborhood is strict and requires the complete inclusion of both inner and outer parts of neighborhood. In contrast, the partial reducible neighborhoods are more flexible and require either inner or outer neighborhood inclusion. Through removing the reducible neighborhoods, we maintain the key neighborhoods, i.e. irreducible neighborhoods to form concise coverings of data space. Moreover, the reduced neighborhood coverings formulate lower and upper approximations of data space of different classes and the irreducible neighborhoods are collected as rules for classification. The reduction processes of complete and partial reducible neighborhoods are summarized in Algorithm 1.

Algorithm 1 T-neighborhood covering reduction.

Input: Training samples $\{(x_1, d_1), ...(x_i, d_i), ...(x_n, d_n)\}$;
Output: Rule set R of irreducible neighborhoods
For each sample $x_i, i = 1, 2, \ldots, n$, construct its tri-partition neighborhood $O(x_i) = \{\underline{O}(x_i), \overline{O}(x_i)\}$ and compute the inner and outer neighborhood radii $\eta^{\beta=0}, \eta^{\beta=r}$ according to formula (10), (11);
Compute the number of samples contained in each neighborhood $O(x_i)$ as $num(x_i)$ and sort all the neighborhoods in descending order;
Initialize rule set $R \leftarrow \phi$ and neighborhood set $O(X) = \cup O(x_i)$;
while $O(X) \neq \phi$ **do**
  Select the neighborhood $O(x)$ covering the maximum samples;
  Move $O(x)$ from $O(X)$ to R in the form of $(x, \eta^{\beta=0}, \eta^{\beta=r}, d)$;
  For any neighborhood $O(x') \in O(X)$, remove $O(x')$ in $O(X)$ if $O(x')$ is complete/partial reducible with respect to $O(x)$;
end while
Sort the rules in R according to the outer radii $\eta^{\beta=r}$ and select the top h rules.

The reductions of different kinds of reducible T-neighborhoods will lead to different neighborhood coverings of data space. Removing complete reducible neighborhoods preserves both lower and upper approximations of neighborhood covering. On the other side, the inner-partial neighborhood reduction keeps only the lower approximation and the outer-partial reduction just preserves the upper approximation of neighborhood covering. The properties of T-neighborhood covering reduction are further analyzed in the following theorems.

Theorem 1. *Inner-partial T-neighborhood reduction preserves the lower approximation of neighborhood covering.*

Proof. Suppose a set of initial T-neighborhoods of class d is $O(X_d) = \{O(x)|x \in X_d\} = \{(\underline{O}(x), \overline{O}(x))|x \in X_d\}$, the lower approximation of the initial neighborhood covering is $\underline{N}(X_d) = \cup_{x \in X_d} \underline{O}(x) = \cup_{O(x) \in O(X_d)} \underline{O}(x)$. Removing the inner-partial reducible neighborhoods from $O(X_d)$, we have a neighborhood set $O^{ir}(X_d)$ after inner-partial reduction, which leads to the lower and upper approximations of the reduced neighborhood covering $\underline{N}^{ir}(X_d) = \cup_{O(x) \in O^{ir}(X_d)} \underline{O}(x)$, $\overline{N}^{ir}(X_d) = \cup_{O(x) \in O^{ir}(X_d)} \overline{O}(x)$.

Obviously, $O^{ir}(X_d)$ is produced through removing the neighborhoods from $O(X_d)$, thus $O(X_d) \supseteq O^{ir}(X_d)$ and we have

$$\begin{aligned}
& O(X_d) \supseteq O^{ir}(X_d) \\
\Rightarrow & \cup_{O(x) \in O(X_d)} \underline{O}(x) \supseteq \cup_{O(x) \in O^{ir}(X_d)} \underline{O}(x) \\
\Rightarrow & \underline{N}(X_d) = \cup_{O(x) \in O(X_d)} \underline{O}(x) \supseteq \cup_{O(x) \in O^{ir}(X_d)} \underline{O}(x) = \underline{N}^{ir}(X_d) \\
\Rightarrow & \underline{N}(X_d) \supseteq \underline{N}^{ir}(X_d)
\end{aligned}$$

On the other side, because every T-neighborhood in $O^{ir}(X_d)$ is inner-partial irreducible, according to Definition 11, we also have

$$\begin{aligned}
& \forall O(x) \in O(X_d), \exists O(x') \in O^{ir}(X_d), O(x) \overset{in}{\subseteq} O(x') \\
\Rightarrow & \forall \underline{O}(x) \in O(X_d), \exists \underline{O}(x') \in O^{ir}(X_d), \underline{O}(x) \subseteq \underline{O}(x') \\
\Rightarrow & \cup_{O(x) \in O(X_d)} \underline{O}(x) \subseteq \cup_{O(x') \in O^{ir}(X_d)} \underline{O}(x') \\
\Rightarrow & \underline{N}(X_d) = \{\cup_{O(x) \in O(X_d)} \underline{O}(x)\} \subseteq \{\cup_{O(x') \in O^{ir}(X_d)} \underline{O}(x')\} = \underline{N}^{ir}(X_d) \\
\Rightarrow & \underline{N}(X_d) \subseteq \underline{N}^{ir}(X_d)
\end{aligned}$$

Thus we obtain $\underline{N}(X_d) = \underline{N}^{ir}(X_d)$ and prove the strategy of inner-partial reduction preserves the lower approximation of initial neighborhood covering. □

Theorem 2. *Outer-partial T-neighborhood reduction preserves the upper approximation of neighborhood covering.*

Proof. For a set of initial T-neighborhoods of class d, $O(X_d) = \{O(x)|x \in X_d\} = \{(\underline{O}(x), \overline{O}(x))|x \in X_d\}$, the upper approximation of the initial neighborhood covering is $\overline{N}(X_d) = \cup_{x \in X_d} \overline{O}(x) = \cup_{O(x) \in O(X_d)} \overline{O}(x)$. Reducing the outer-partial reducible neighborhoods from $O(X_d)$, we obtain a neighborhood set $O^{or}(X_d)$, which leads to the approximations of neighborhood covering $\underline{N}^{or}(X_d) = \cup_{O(x) \in O^{or}(X_d)} \underline{O}(x)$, $\overline{N}^{or}(X_d) = \cup_{O(x) \in O^{or}(X_d)} \overline{O}(x)$.

$O^{or}(X_d)$ is produced through removing the neighborhoods from $O(X_d)$, thus $O(X_d) \supseteq O^{or}(X_d)$ and we infer that

$$\begin{aligned}
& O(X_d) \supseteq O^{or}(X_d) \\
\Rightarrow & \cup_{O(x) \in O(X_d)} \overline{O}(x) \supseteq \cup_{O(x) \in O^{or}(X_d)} \overline{O}(x) \\
\Rightarrow & \overline{N}(X_d) \supseteq \overline{N}^{or}(X_d)
\end{aligned}$$

Because every T-neighborhood in $O^{or}(X_d)$ is outer-partial irreducible, referring to the definition of partial reduction, we have

$$\begin{aligned}
& \forall O(x) \in O(X_d), \exists O(x') \in O^{or}(X_d), O(x) \overset{out}{\subseteq} O(x') \\
\Rightarrow & \forall \overline{O}(x) \in O(X_d), \exists \overline{O}(x') \in O^{or}(X_d), \overline{O}(x) \subseteq \overline{O}(x') \\
\Rightarrow & \cup_{O(x) \in O(X_d)} \overline{O}(x) \subseteq \cup_{O(x') \in O^{or}(X_d)} \overline{O}(x') \\
\Rightarrow & \overline{N}(X_d) = \{\cup_{O(x) \in O(X_d)} \overline{O}(x)\} \subseteq \{\cup_{O(x') \in O^{or}(X_d)} \overline{O}(x')\} = \overline{N}^{or}(X_d) \\
\Rightarrow & \overline{N}(X_d) \subseteq \overline{N}^{or}(X_d)
\end{aligned}$$

Thus we obtain $\overline{N}(X_d) = \overline{N}^{or}(X_d)$ and prove that outer-partial reduction maintains the upper approximation of initial neighborhood covering. □

Theorem 3. *Complete T-neighborhood reduction preserves both the lower and upper approximations of neighborhood covering.*

Proof. The lower and upper approximations of the covering of initial neighborhoods $O(X_d)$ are $\underline{N}(X_d) = \cup_{x \in X_d} \underline{O}(x) = \cup_{O(x) \in O(X_d)} \underline{O}(x)$ and $\overline{N}(X_d) = \cup_{x \in X_d} \overline{O}(x) = \cup_{O(x) \in O(X_d)} \overline{O}(x)$. Reducing the complete reducible neighborhoods from $O(X_d)$, we obtain a neighborhood set $O^r(X_d)$, which leads to the approximations $\underline{N}^r(X_d) = \cup_{O(x) \in O^r(X_d)} \underline{O}(x)$ and $\overline{N}^r(X_d) = \cup_{O(x) \in O^r(X_d)} \overline{O}(x)$. We have

$$\begin{aligned}
& O(X_d) \supseteq O^r(X_d) \\
\Rightarrow & \{\cup_{O(x) \in O(X_d)} \underline{O}(x)\} \supseteq \{\cup_{O(x) \in O^r(X_d)} \underline{O}(x)\} \wedge \{\cup_{O(x) \in O(X_d)} \overline{O}(x)\} \supseteq \{\cup_{O(x) \in O^r(X_d)} \overline{O}(x)\} \\
\Rightarrow & \{\underline{N}(X_d) \supseteq \underline{N}^r(X_d)\} \wedge \{\overline{N}(X_d) \supseteq \overline{N}^r(X_d)\}
\end{aligned}$$

Because every T-neighborhood in $O^r(X_d)$ is complete irreducible, according to Definition 10, we have

$$\begin{aligned} &\forall O(x) \in O(X_d), \exists O(x') \in O^r(X_d), O(x) \subseteq O(x') \\ \Rightarrow &\forall(\underline{O}(x), \overline{O}(x)) \in O(X_d), \exists(\underline{O}(x'), \overline{O}(x')) \in O^r(X_d), \underline{O}(x) \subseteq \underline{O}(x') \wedge \overline{O}(x) \subseteq \overline{O}(x') \\ \Rightarrow &\{\cup_{O(x)\in O(X_d)} \underline{O}(x)\} \subseteq \{\cup_{O(x')\in O^r(X_d)} \underline{O}(x')\} \wedge \{\cup_{O(x)\in O(X_d)} \overline{O}(x)\} \subseteq \{\cup_{O(x')\in O^r(X_d)} \overline{O}(x')\} \\ \Rightarrow &\{\underline{N}(X_d) \subseteq \underline{N}^r(X_d)\} \wedge \{\overline{N}(X_d) \subseteq \overline{N}^r(X_d)\} \end{aligned}$$

Thus we obtain $\underline{N}(X_d) = \underline{N}^r(X_d)$ and $\overline{N}(X_d) = \overline{N}^r(X_d)$ and prove the complete reduction preserves both the lower and upper approximations of initial neighborhood covering. □

The theorems above provide us the following enlightenments. The process of compete T-neighborhood reduction preserves both the certain and uncertain regions of neighborhood covering and thereby maintains the complete structure of data space. The inner-partial reduction aims to preserve only the certain parts of neighborhood covering and its objective is same as the traditional neighborhood reduction. On the contrary, the outer-partial reduction just focuses on the preservation of uncertain parts of neighborhoods. This strategy may lose the certain neighborhoods in reduction process, but in the meantime, the preservation of uncertain region will extend the neighborhood boundaries, which facilitates model to tolerate heavy noise. With complete and partial reduction strategies, T-neighborhood covering generalizes the traditional neighborhood model and the preserved covering approximations are flexible to represent noisy data. The robustness of T-neighborhood covering will be further validated in the experiments.

4.2. Classification with tri-partition neighborhoods

The covering of irreducible T-neighborhoods approximates the data space of different classes. Based on the rules of irreducible T-neighborhoods, we can design a classifier on neighborhood level. Because of the uncertain region in the upper approximation of T-neighborhood covering, we cannot directly classify a sample according to its distance from neighborhoods. The classification with T-neighborhoods depends on the location of samples in the partitioned neighborhood regions.

Algorithm 2 Classification with T-neighborhoods.

Input:
Test samples $\{x_1, ..., x_i, ..., x_m\}$,
Rule set R of irreducible T-neighborhoods, $|R| = h$;
Output: Classification results
1: For each sample x_i, compute the distances $\Delta(x_i, x_j)$ between x_i and the center of every irreducible neighborhood $(x_j, \eta^{\beta=0}_{O(x_j)}, \eta^{\beta=r}_{O(x_j)}, d)$ in rule set R, $j = 1, 2, ...h$;
2: If $\exists k$, $\Delta(x_i, x_k) \le \eta^{\beta=0}_{O(x_k)}$, $1 \le k \le h$;
3: Classify x_i according to the kth neighborhood rule;
4: Else if $\overline{\Delta}(x_i) = min_{1\le j\le h}\{\Delta(x_i, x_j) - \eta^{\beta=r}_{O(x_j)}\} \le 0$, find all k, s.t. $\Delta(x_i, x_k) - \eta^{\beta=r}_{O(x_k)} = \overline{\Delta}(x_i)$ and add kth neighborhood into set $O^\wedge$;
5: $\forall O(x_k) \in O^\wedge$, $\underline{\Delta}(x_i) = min_{O(x_k)\in O^\wedge}\{\Delta(x_i, x_k) - \eta^{\beta=0}_{O(x_k)}\}$ and $l = argmin_{O(x_k)\in O^\wedge}\{\Delta(x_i, x_k) - \eta^{\beta=0}_{O(x_k)}\}$;
6: Classify x_i according to the lth neighborhood rule;
7: Else if $\overline{\Delta}(x_i) > 0$, $k = argmin_{1\le j\le h}\{\Delta(x_i, x_j) - \eta^{\beta=r}_{O(x_j)}\}$;
8: Classify x_i according to the kth neighborhood rule.

For the sample located in the inner region of a neighborhood, i.e. lower approximation of neighborhood covering, we can consider the sample certainly belongs to the neighborhood and directly classify it according to the class label of neighborhood. If a sample x locates in the boundary of neighborhoods, i.e. uncertain region between lower and upper covering approximations, the sample may belong to multiple neighborhoods. In this case, we choose the neighborhood with the maximum distance from its outer border to the sample x and have

$$O^\wedge = argmax_{O(x'):x\in\overline{O}(x')}\{\eta^{\beta=r}_{O(x')} - \Delta(x, x')\} \tag{15}$$

If the neighborhood with maximum outer-border distance is not unique, we further check the distance between the sample and the inner-border of neighborhood to label the class of sample and obtain

$$O^\star = argmin_{O(x'):O(x')\in O^\wedge}\{\Delta(x, x') - \eta^{\beta=0}_{O(x')}\} \tag{16}$$

For the sample beyond any neighborhood, it will be classified through searching for the nearest neighborhood. The distances between the outside sample and neighborhoods are computed as the distances from the sample to the outer-borders of neighborhoods. The workflow of the classification with irreducible T-neighborhoods is presented in Algorithm 2.

Table 1
Experimental data sets.

Data sets	Feature	Instance	Class	Attribute type
glass	10	214	7	Numerical
vote	17	435	2	Categorical
soybean	36	683	19	Categorical
credit-rating	16	690	2	Mixed
breast-w	10	699	2	Numerical
diabetes	9	768	2	Mixed
segment	20	2310	7	Numerical
car	6	1728	4	Categorical
banknote	4	1372	2	Numerical
page	10	5473	5	Numerical

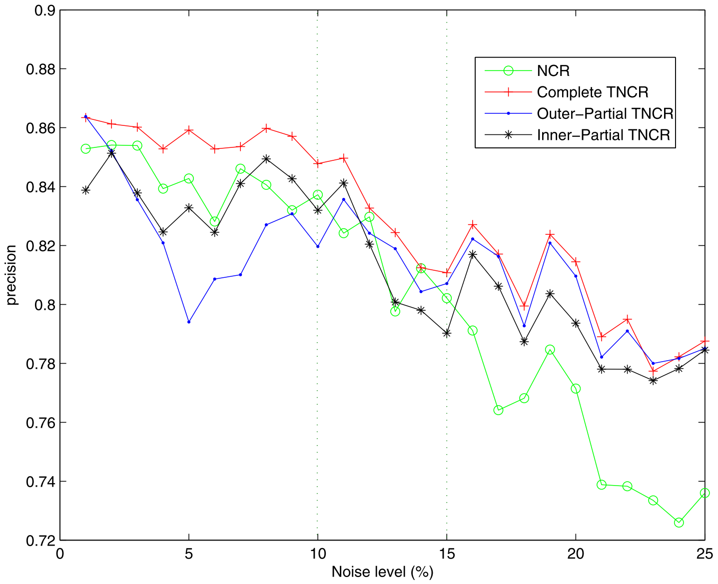

Fig. 6. Classification results of neighborhood reduction algorithms on noisy data.

5. Experimental results

Partitioning the neighborhood into certain and uncertain regions, the covering of T-neighborhoods provides a flexible way to represent the data space. The classifiers induced by the T-neighborhood covering reduction handle the noise well and are robust for data classification. We implement two tests to validate this. The first one aims to verify the robustness of T-neighborhoods to noise. And in the second test, the classification based on T-neighborhood covering reduction will be overall evaluated through comparing it with other kinds of classification methods, which include the neighborhood-based methods, such as improved KNN and NCR [6], and other state-of-the-art algorithms, such as SVM, Naive Bayes and Decision Trees (C4.5) [9]. The noisy data in experiments are generated through randomly changing the items of data sets from the machine learning data repository, University of California at Irvine (UCI). For all the tests of classification, 10-fold cross validation is performed on each data set. The descriptions of the UCI data sets are given in Table 1.

To validate the robustness of T-neighborhoods to noise, we perform the classifications based on complete and partial T-neighborhood covering reductions (TNCR) on multi-grade noisy data respectively. We also compare the T-neighborhood-based classifications with the classification based on traditional neighborhood covering reduction (NCR). Fig. 6 presents the average classification precision of different neighborhood methods against the noise level from 1% to 25%. It can be found that the performance of the classification based on traditional neighborhoods deteriorates rapidly as noise level increases. In contrast to the strict constraint of certain neighbors, with the boundary between inner and outer neighborhoods, T-neighborhoods tolerate the uncertain neighbors and thus handle the noise well. Under both complete and partial reduction strategies, the classifications based on T-neighborhood covering model achieves more stable results even the data is polluted by serious noise.

Among the different reduction strategies of T-neighborhoods, the reduction of the complete reducible neighborhoods achieves the best performance. Fig. 6 indicates that the rule set of complete irreducible neighborhoods obtains the most precise classification results on all the noisy data sets. This is because the complete reduction strategy preserves both the lower and upper approximations of neighborhood covering and thereby provides a precise data representation for classifica-

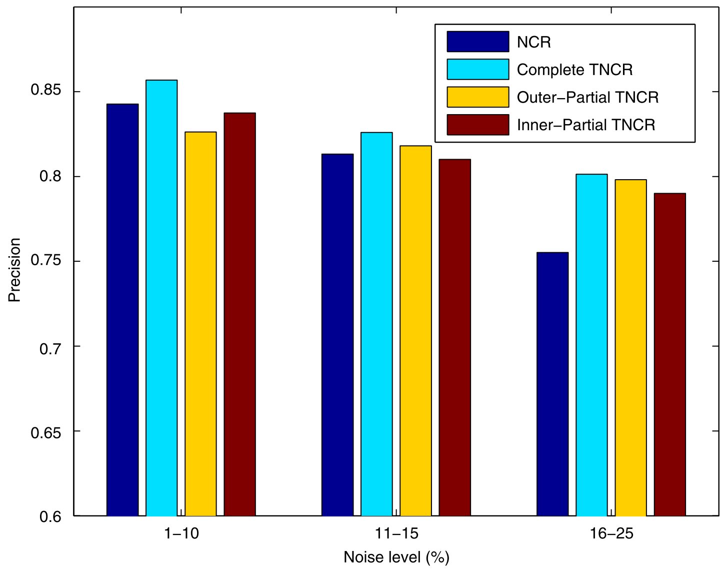

Fig. 7. Multi-stage classification results of neighborhood reduction algorithms.

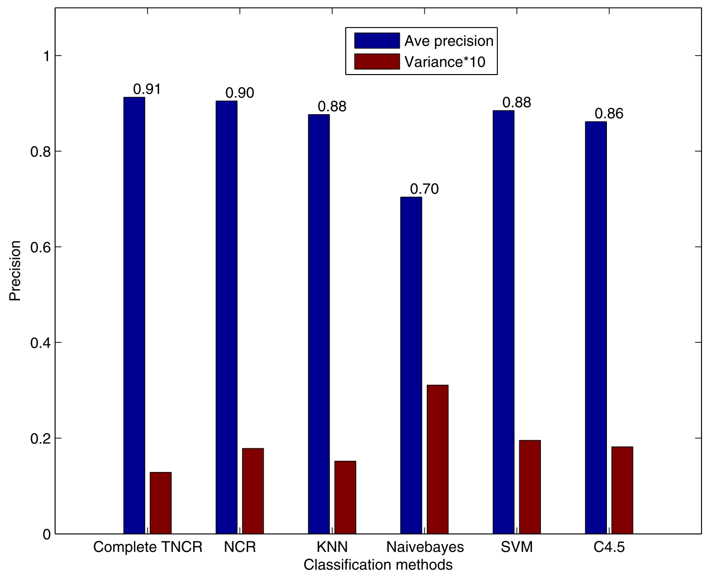

Fig. 8. Average precision and variance of different classification methods.

tion. The inner-partial reduction preserves the lower approximation of neighborhood covering and keeps the distribution of neighbors certainly belonging to the same class. Therefore, on the low-level noisy data, the classification based on inner-partial reduction achieves similar performance as the traditional neighborhood method. As noise increasing, with the flexible boundary, the T-neighborhood reduction gradually out-performs the traditional neighborhood methods. In contrast, the outer-partial reduction preserves the upper approximation of neighborhood covering and aims to maintain the distribution of uncertain neighbors. Under this reduction strategy, when noise is mild, the distortion of inner neighborhoods may lead to the misclassification of certain neighbors. However, preserving the upper approximation of neighborhood covering will keep all the neighborhood boundaries that facilitate tolerating the heavy noise. As shown in Fig. 7, for the data with high-level noise, the outer-partial neighborhood reduction achieves precise and stable classification results.

The second test overall evaluates the classification based on T-neighborhood covering reduction. Specifically, we compare the classification based on Complete T-Neighborhood Covering Reduction (Complete TNCR) with two typical neighbor-based methods: Neighborhood Covering Reduction (NCR) and improved K-Nearest Neighbors (KNN), and other three elegant classification methods: Naive Bayes, Support Vector Machine (SVM) and Decision Trees (C4.5). We randomly select data sets to perform the classification tasks and add 15% noise to each data set. Fig. 8 illustrates the average classification precision and the corresponding variance for each method.

We find that neighborhood-based methods and SVM obtain the top precise results, and in the meantime, Complete TNCR achieves the minimum precision variance. This indicates the classification based on the complete reduction of T-neighborhoods is much more stable on noisy data. It is also observed that the classifications based on T-neighborhoods achieve better performances than the other methods when adding more noise ($>15\%$) to the testing data sets. Besides the resistance to noise, the set approximations adopted in T-neighborhood systems are also helpful to handle the imbalanced data distribution and thereby leads to robust classification results.

6. Conclusion

Based on the tri-partition methodology, in this paper, we extend the traditional neighborhoods to tri-partition neighborhoods for robust classification. The tri-partition neighborhood consists of inner and outer parts, which involve certain and uncertain neighbors respectively. The region between the inner and outer neighborhoods denotes the neighborhood boundary to tolerate noise. Moreover, the covering consists of tri-partition neighborhoods leads to multilevel representations of data space. The union of inner neighborhoods constitutes the certain lower approximation of data space and the outer neighborhoods comprise the upper approximation with uncertainty. Besides the neighborhood construction, we also proposed complete and partial reduction strategies to filter out the redundant neighborhoods to simplify the covering for classification. It is proven that complete reduction preserves both the lower and upper approximations of neighborhood covering and the inner/outer neighborhood reduction just partially preserves them. Through preserving the approximations of neighborhood covering, the reduction of tri-partition neighborhoods provides us a flexible way to handle uncertain samples and noise. Experiments verify the robustness of the proposed tri-partition neighborhood covering reduction method.

Our future work will focus on the following issues. The first issue is to optimize the parameters to construct tri-partition neighborhoods. The optimized parameters will improve the neighborhood boundaries to fit various noise and outliers. Second, we try to utilize the kernel methods to construct more flexible neighborhoods. The kernel-based transformation will map the neighbors into different feature spaces to reveal the local homogeneity. Finally, aiming to handle the mining tasks for massive data, we expect to improve the efficiency of tri-partition neighborhood covering reduction. The feasible solutions include the methodologies of neighborhood sampling and the parallelized strategies of neighborhood reduction.

Acknowledgements

This work was supported by National Natural Science Foundation of China (Serial Nos. 61573235, 61175054, 61403329), Natural Science Foundation of Jiangsu Province (No. BK20141152), The Humanity and Social Science Youth Foundation of Ministry of Education of China (No. 15YJCZH129), and The Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University (No. ESSCKF201303).

References

[1] J. Chen, Y.P. Zhang, S. Zhao, Multi-granular mining for boundary regions in three-way decision theory, Knowl.-Based Syst. 91 (2016) 287–292.
[2] D. Ciucci, Orthopairs: a simple and widely used way to model uncertainty, Fundam. Inform. 108 (3–4) (2011) 287–304.
[3] D. Ciucci, D. Dubois, J. Lawry, Borderline vs. unknown: comparing three-valued representations of imperfect information, Int. J. Approx. Reason. 55 (2014) 1866–1889.
[4] D. Ciucci, Orthopairs and granular computing, in: Granular Computing, vol. 1, 2016, pp. 1–12.
[5] X.F. Deng, Y.Y. Yao, Decision-theoretic three-way approximations of fuzzy sets, Inf. Sci. 279 (2014) 702–715.
[6] Y. Du, Q.H. Hu, P.F. Zhu, P.J. Ma, Rule learning for classification based on neighborhood covering reduction, Inf. Sci. 181 (2011) 5457–5467.
[7] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, Int. J. Gen. Syst. 17 (1990) 191–209.
[8] D. Dubois, H. Prade From, Blanché's hexagonal organization of concepts to formal concept analysis and possibility theory, Log. Univers. 6 (2012) 149–169.
[9] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, John Wiley & Sons, 2012.
[10] H. Fujita, T.R. Li, Y.Y. Yao, Advances in three-way decisions and granular computing, Knowl.-Based Syst. 91 (2016) 1–3.
[11] R. Gilad-Bachrach, A. Navot, N. Tishby, Margin based feature selection-theory and algorithms, in: ACM Proceedings of the Twenty-first International Conference on Machine Learning, 2004, pp. 43–50.
[12] T.P. Hong, T.T. Wang, S.L. Wang, B.C. Chien, Learning a coverage set of maximally general fuzzy rules by rough sets, Expert Syst. Appl. 19 (2000) 97–103.
[13] B.Q. Hu, Three-way decisions space and three-way decisions, Inf. Sci. 281 (2014) 21–52.
[14] Q.H. Hu, W. Pedrycz, D. Yu, J. Lang, Selecting discrete and continuous features based on neighborhood decision error minimization, IEEE Trans. Syst. Man Cybern., Part B 40 (2010) 137–150.
[15] Q.H. Hu, D.R. Yu, J.F. Liu, C.X. Wu, Neighborhood rough set based heterogeneous feature subset selection, Inf. Sci. 178 (2008) 3577–3594.
[16] Q.H. Hu, D.R. Yu, Z.X. Xie, Neighborhood classifier, Expert Syst. Appl. 34 (2008) 866–876.
[17] X. Jia, W. Liao, Z. Tang, L. Shang, Minimum cost attribute reduction in decision-theoretic rough set models, Inf. Sci. 219 (2013) 151–167.
[18] H. Li, L. Zhang, B. Huang, X. Zhou, Sequential three-way decision and granulation for cost-sensitive face recognition, Knowl.-Based Syst. 91 (2016) 241–251.
[19] D.C. Liang, D. Liu, Deriving three-way decisions from intuitionistic fuzzy decision-theoretic rough sets, Inf. Sci. 300 (2015) 28–48.
[20] T.Y. Lin, Neighborhood systems and relational database, in: Proceedings of ACM Sixteenth Annual Computer Science Conference, 1988, pp. 23–25.
[21] T.Y. Lin, Neighborhood systems – application to qualitative fuzzy and rough sets, in: Advances in Machine Intelligence and Soft Computing, Duke University, NC, 1997.
[22] M. Lindenbaum, S. Markovitch, D. Rusakov, Selective sampling for nearest neighbor classifiers, Mach. Learn. 54 (2004) 125–152.
[23] D. Liu, D.C. Liang, C.C. Wang, A novel three-way decision model based on incomplete information system, Knowl.-Based Syst. 91 (2016) 32–45.
[24] D.Q. Miao, Y. Zhao, Y.Y. Yao, Relative reducts in consistent and inconsistent decision tables of the Pawlak rough set model, Inf. Sci. 179 (2009) 4140–4150.

[25] M. Muja, D.G. Lowe, Scalable nearest neighbor algorithms for high dimensional data, IEEE Trans. Pattern Anal. Mach. Intell. 36 (11) (2014) 2227–2240.
[26] A. Owen, A neighborhood-based classifier for LANDSAT data, Can. J. Stat. 12 (1984) 191–200.
[27] Z. Pawlak, Rough sets, Int. J. Inf. Comput. Sci. 11 (5) (1982) 314–356.
[28] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning About Data, Kluwer Academic Publishers, Boston, 1991.
[29] Z. Pawlak, Some issues on rough sets, in: Transactions on Rough Sets I, vol. 3100, 2004, pp. 1–58.
[30] J.F. Peters, S. Ramanna, Proximal three-way decisions: theory and applications in social networks, Knowl.-Based Syst. 91 (2016) 4–15.
[31] Y.H. Qian, J.Y. Liang, W. Pedrycz, C.Y. Dang, Positive approximation: an accelerator for attribute reduction in rough set theory, Artif. Intell. 174 (9–10) (2010) 597–618.
[32] H. Wang, Nearest neighbors by neighborhood counting, IEEE Trans. Pattern Anal. Mach. Intell. 28 (2006) 942–953.
[33] D. Wettschereck, T.G. Dieterich, An experimental comparison of the nearest neighbor and nearest-hyper-rectangle algorithms, Mach. Learn. 19 (1995) 5–27.
[34] D.R. Wilson, T.R. Martinez, Improved heterogeneous distance functions, J. Artif. Intell. Res. 1 (34) (1997).
[35] W.Z. Wu, W.X. Zhang, Neighborhood operator systems and approximations, Inf. Sci. 144 (2002) 201–217.
[36] T. Yang, Q. Li, Reduction about approximation spaces of covering generalized rough sets, Int. J. Approx. Reason. 51 (2010) 335–345.
[37] Y.Y. Yao, Relational interpretations of neighborhood operators and rough set approximation operators, Inf. Sci. 111 (1998) 239–259.
[38] Y.Y. Yao, Three-way decisions with probabilistic rough sets, Inf. Sci. 180 (2010) 341–353.
[39] Y.Y. Yao, The superiority of three-way decisions in probabilistic rough set models, Inf. Sci. 181 (2011) 1080–1096.
[40] Y.Y. Yao, An outline of a theory of three-way decisions, in: RSCTC 2012, in: LNCS (LNAI), vol. 7413, 2012, pp. 1–17.
[41] R. Younsi, A. Bagnall, An efficient randomised sphere cover classifier, Int. J. Data Min. Model. Manag. 4 (2) (2012) 156–171.
[42] H. Yu, C. Zhang, G.Y. Wang, A tree-based incremental overlapping clustering method using the three-way decision theory, Knowl.-Based Syst. 91 (2016) 189–203.
[43] W. Zhu, F. Wang, Reduction and axiomization of covering generalized rough sets, Inf. Sci. 152 (2003) 217–230.
[44] W. Zhu, F. Wang, On three types of covering-based rough sets, IEEE Trans. Knowl. Data Eng. 19 (2007) 1131–1144.