Information Sciences 507 (2020) 795–808

Contents lists available at ScienceDirect

Information Sciences

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

Fuzzy neighborhood covering for three-way classification

X.D. Yue [a,b,*], Y.F. Chen [c,*], D.Q. Miao [c], H. Fujita [d]

[a] School of Computer Engineering and Science, Shanghai University, China
[b] Shanghai Institute for Advanced Communication and Data Science, Shanghai University, China
[c] College of Electronic and Information Engineering, Tongji University, China
[d] Iwate Prefectural University (IPU), Faculty of Software and Information Science, Iwate, Japan

ARTICLE INFO

Article history:
Received 15 October 2017
Revised 12 July 2018
Accepted 25 July 2018
Available online 27 July 2018

Keywords:
Three-way classification
Fuzzy neighborhood covering

ABSTRACT

Neighborhood Covering (NC) is the union of homogeneous neighborhoods and provides a set-level approximation of data distribution. Because of the nonparametric property and the robustness to complex data, neighborhood covering has been widely used for data classification. Most existing methods directly classify data samples according to the nearest neighborhoods. However, the certain classification methods strictly classify the uncertain data and may lead to serious classification mistakes. To tackle this problem, we extend traditional neighborhood coverings to fuzzy ones and thereby propose a Three-Way Classification method with Fuzzy Neighborhood Covering (3WC-FNC). Fuzzy neighborhood covering consists of membership functions and forms an approximate distribution of neighborhood belongingness. Based on the soft partition induced by the memberships of fuzzy neighborhood coverings of different classes, data samples are classified into Positive (certainly belonging to a class), Negative (certainly beyond classes) and Uncertain cases. Experiments verify that the proposed three-way classification method is effective to handle the uncertain data and in the meantime reduce the classification risk.

© 2018 Elsevier Inc. All rights reserved.

1. Introduction

Neighborhood Systems were proposed through extending the strategies of the nearest neighbors [26]. In a neighborhood system, an object is associated with its neighborhood rather than its nearest neighbors [21]. The classifications based on neighborhoods were proven to be more efficient than the classifications based on nearest-neighbor search [29]. The space of neighborhoods was also investigated to approximate global data distribution. From the view of topology, it has been demonstrated that the neighborhood spaces are more general than the data-level spaces [20,42]. This indicates that transforming original data into neighborhood systems will facilitate the data generalization [34].

Through extending Rough Sets [22,23] with neighborhoods, Neighborhood Rough Sets were proposed to construct approximations of data space [16,20,34]. Different from the equivalence classes defined by symbols in the classic rough sets, the basic granules in neighborhood rough sets are the neighborhoods in numerical/norminal data spaces, which makes the model represent the mixed-type data well [16,17]. Formulating data space with neighborhood rough sets, data distributions can be approximated by Neighborhood Covering (NC), which consists of a group of homogeneous neighborhoods, i.e. all the data samples in a neighborhood belonging to the same class. Neighborhood covering provides us an effective way to

* Corresponding authors.
E-mail addresses: yswantfly@shu.edu.cn (X.D. Yue), yufeichen@tongji.edu.cn (Y.F. Chen).

https://doi.org/10.1016/j.ins.2018.07.065
0020-0255/© 2018 Elsevier Inc. All rights reserved.

represent data distributions on neighborhood level [41,42]. Moreover, to obtain the concise representation of data distribution, Neighborhood Covering Reduction (NCR) methods were used to remove the redundant neighborhoods from the initial neighborhood coverings [8,32,38].

Based on the neighborhood coverings of data distributions, the learning methods can be implemented for classification [14,37] and feature selection [15,16,27]. Comparing with other kinds of learning methods, NC-based methods require no parameter setting and are robust to complex data. For the classification with neighborhood covering, the existing methods directly classify an unknown sample into a class according to its nearest neighborhood. However, this certain classification strategy strictly classifies the uncertain data and may lead to serious classification mistakes. Because of the unavoidable inconsistency between the training data and the unknown world, there generally exit uncertain cases in data classification. Thus it is required to design a cautious NC-based classification for uncertain data to reduce the classification risk.

To implement the uncertain classification with neighborhood covering, we expect to construct a possibilistic measure of the belongingness of neighborhood coverings and thereby design a three-way classification strategy. This solution originates from the methodologies of Three-Way Decisions (3WD) [35,36]. In the process of three-way decision making, decision rules are extracted from the data with uncertainty through tri-partitioning data space into Positive, Negative and Boundary regions [2,12]. From the view of classification, the three regions correspond to the cases of certainly belonging to a class, certainly beyond a class and non-commitment, i.e. uncertain case [6,19].

In the light of the superiority of fuzzy sets for the learning tasks on uncertain data [1,7,24], we adopt fuzzy membership functions to measure the possibilities of data samples belonging to neighborhoods. Based on the neighborhood memberships of data samples, we apply tri-partitioning methodology to reformulate the neighborhood-based classification and propose a Three-Way Classification method with Fuzzy Neighborhood Covering (3WC-FNC). The proposed method involves two parts: fuzzy extension of neighborhood coverings and three-way classification with fuzzy neighborhood coverings. Different from the traditional covering model formed by the union of neighborhoods, the fuzzy neighborhood covering consists of a group of neighborhood membership functions which are integrated to form the membership distribution of neighborhood coverings. The membership distribution of neighborhood coverings of different classes induces a soft partition of data space. According to the memberships of neighborhood coverings, data samples are classified into certain classes and uncertain case. The contributions of this paper are summarized as follows.

- *Extend neighborhood covering to Fuzzy Neighborhood Covering (FNC).* Fuzzy neighborhood covering consists of a group of neighborhood membership functions and forms an uncertain measure of neighborhood covering belongingness. In contrast to the set-level approximation of neighborhood coverings, the fuzzy neighborhood covering provides a membership-level approximation of data distributions.
- *Propose Three-Way Classification with Fuzzy Neighborhood Covering (3WC-FNC).* Based on the fuzzy neighborhood covering of a class, data samples are classified into Positive (certainly belonging to the class), Negative (certainly beyond the class) and Uncertain cases according to their memberships. The three-way strategy separates uncertain cases to reduce the classification risk.

The remainder of this paper is organized as follows. Section 2 introduces the related work. Section 3 describes the entire workflow of the proposed three-way classification method. Section 4 introduces the strategy of the fuzzy extension of neighborhood coverings and also presents the three-way classification algorithms with fuzzy neighborhood coverings. In Section 5, experimental results validate the effectiveness of the proposed three-way method for uncertain data classification. The work conclusion is given in Section 6.

2. Related work

2.1. Neighborhood covering model

Neighborhood Covering (NC) is the union of homogeneous neighborhoods and thereby provides a set-level approximation of data distribution. Because of the advantages of nonparametric property and model robustness to complex data, neighborhood covering models have been improved and widely used in data mining tasks. There are two kinds of NC-based learning methods. The first one aims to approximate the data distributions of different classes for data classification [8,14,32,37,39]. Another kind of methods aims to select the independent features through removing the coverings related to redundant features from the covering family [15,16,27]. In this paper, we focus on the NC-based classification. Next we briefly introduce the preliminaries of neighborhood covering model.

Definition 1 Neighborhood covering. Suppose $U = \{x_1, x_2, \ldots, x_n\}$ is the data space and $O(x) = \{y \in U | \Delta(x, y) \leq \eta\}$ be the neighborhood of $x \in U$, where $\Delta(\cdot)$ is a distance function and η is the threshold. The set of neighborhoods $O_U = \{O(x) | x \in U\}$ forms a covering of data space U and the pair $C = \langle U, O_U \rangle$ denotes the neighborhood covering approximation space.

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

Definition 2 Neighborhood covering reduction. Let $C = \langle U, O_U \rangle$ be a neighborhood covering approximation space. For any $x \in U$, if $\cup_{y \in U-\{x\}} O(y) = U$, $O(x)$ is reducible, otherwise it is irreducible. In addition, C is irreducible iff for any $x \in U$, $O(x)$ is irreducible.

For the samples of the same class, the relative reduction of neighborhoods will produce a concise approximation of the data distribution of the class and thus can be used for data classification.

Definition 3 Relative neighborhood covering reduction. Let $C = \langle U, O_U \rangle$ be a neighborhood covering approximation space, $X \subseteq U$ and $x_i \in U$. If $\exists x_j \in U$ and $j \neq i$ such that $O(x_i) \subseteq O(x_j) \subseteq X$, $O(x_i)$ is a relatively reducible neighborhood with respect to X, otherwise $O(x_i)$ is relatively irreducible.

The existing NC-based classification methods classify data samples according to their nearest neighborhoods. The certain classification induces hard partition of data space and cannot handle the uncertain data well. To tackle this problem, we expect to construct a flexible neighborhood covering model for uncertain classification through fuzzy extension.

2.2. 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 [4,36]. Tri-partition methodology is built on solid cognitive foundations and provides flexible ways for human-like problem solving and information processing [12,28]. As typical approaches, Three-Way Decisions (3WD) [35,36], Orthopairs and Hexagon of Opposition represent knowledge and perform reasoning through tri-partitioning the universe [5]. These approaches have been applied to extend the design and implementation of intelligent systems and the investigations of tri-partition methodology are gaining interest [3,18,25,40].

Three-Way Decisions (3WD) is an extension of the commonly used binary-decision model through adding a third option [35]. 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 [36]. 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. 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 and etc. [36].

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. Orthopair has strict links to Three-valued Logics and can be generalized to Atanassov Intuitionistic Fuzzy Sets, Possibility Theory and Three-way Decision [4]. 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 [5,33].

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.

Although the tri-partition methodology has been investigated in many areas, its applications in neighborhood systems are still limited. In this paper, we expect to extend the neighborhood coverings to fuzzy ones and apply the tri-partition methodology to implement the three-way classification with fuzzy neighborhood covering, in which data samples are classified into certain classes and uncertain cases according to the neighborhood memberships.

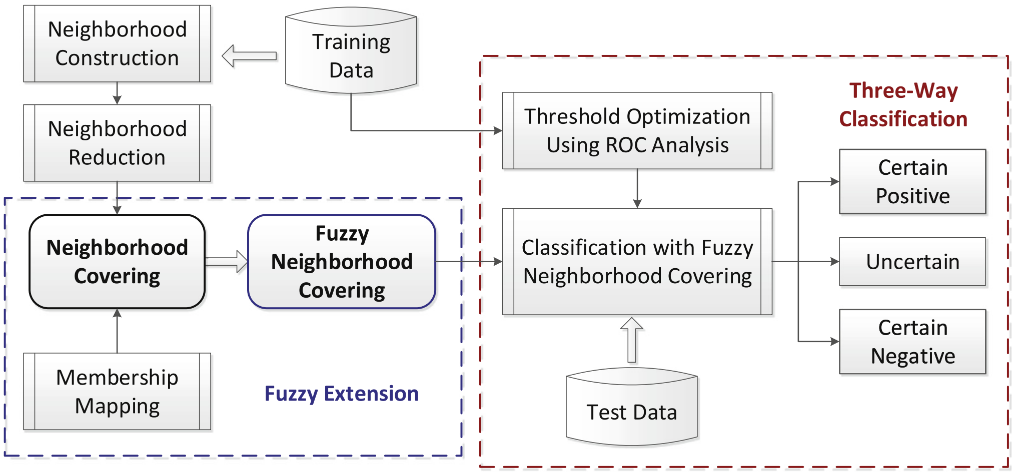

Fig. 1. Workflow of 3-way classification with fuzzy neighborhood covering.

3. Workflow

In this section, we present the entire workflow of the proposed three-way classification method with fuzzy neighborhood covering. Fig. 1 provides an overview of the workflow, which can be divided into three stages: neighborhood covering construction, fuzzy extension of neighborhood covering and three-way classification with fuzzy neighborhood covering. Different from the neighborhood union, fuzzy neighborhood covering consists of neighborhood membership functions which quantify the neighborhood belongingness into continuous membership values. Based on the fuzzy neighborhood membership, we can further formulate the membership distribution of different classes for data classification. Referring to the three-way decision making, for a specific class, the three-way classifiers of fuzzy neighborhood covering judge the data samples as Positive (certainly belonging to the class), Negative (certainly beyond the class) and Uncertain according to their memberships. The uncertain samples can be post-processed to reduce the classification risk.

The stage of neighborhood covering construction involves neighborhood construction and neighborhood reduction. As introduced in Section 2, the neighborhoods are constructed according the distance measure and the data samples distributed into a neighborhood have the same class label, i.e. the neighborhoods are homogeneous. The union of the neighborhoods forms a covering of data samples and the covering of the neighborhoods belonging to the same class actually form an approximation of the data distribution of the class. Moreover, to simply the approximation of data distribution, redundant neighborhoods should be reduced to generate the concise neighborhood covering. In the second stage, we extend the neighborhood coverings to fuzzy ones. Specifically, we map the possibilities of data samples belonging to neighborhoods into fuzzy memberships according to distances between samples and neighborhoods. Moreover, through integrating the neighborhood memberships, we formulate the membership distribution of data belongingness to neighborhood coverings.

Based on the membership distributions of neighborhood coverings of different classes, we implement the three-way classification in the final stage. The stage of three-way classification involves the modules of classification and optimization. The three-way classifier with fuzzy neighborhood coverings classifies unknown samples according to their memberships to different classes, and in the meantime adopts membership thresholds to divide the classified samples into certain and uncertain cases. The parameters of membership thresholds play an important role in three-way classification and determine the uncertain boundary between different classes. Classifying samples into uncertain cases is helpful to avoid serious misclassifications but over classification of uncertain samples will lead to bad effects. To optimize the uncertain boundary of three-way classification, in the algorithm implementation, we use Receiver Operating Characteristic curve (ROC) [9,11] to search for the optimum membership thresholds. ROC point consists of the True Positive (TP) rate and False Positive (FP) rate, and the curve of ROC points under varying thresholds reflect the influences of the threshold parameters to classification. In the ROC curve of data training, the membership thresholds corresponding to the points of high TP rate and low FP rate are chosen as the optimal parameters for uncertain classification.

4. Fuzzy neighborhood covering for 3-Way classification

4.1. Neighborhood covering construction

Neighborhood is constructed based on the similarity/distance between samples. To handle the ubiquitous mixed-type data with the attributes of nominal and numerical value domain, Heterogeneous Euclidean-Overlap Metric (HEOM) [30] is adopted to measure the sample distance according to the following formula.

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

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

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

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

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

To simplify the neighborhood construction, we set all the attribute weights $w_{a_i} = 1$ as default.

Based on the HEOM distance, we can construct neighborhoods through grouping nearby samples. Given a sample x, the neighborhood $O(x)$ consists of the samples surrounding x, $O(x) = \{y | \Delta(x,y) \leq \eta\}$, η is a distance threshold to denote the neighborhood radius. 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 [8,13]. Specifically, for a sample x, its Nearest Hit $NH(x)$ 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 with different class label and is named the Nearest Miss. The neighborhood radius is computed 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.

As introduced in Section 2, for a set of data samples $\{x_1, x_2, \ldots x_n\}$, the union of all the neighborhoods $O = \cup_{i=1}^{n} O(x_i)$ forms a covering and also a set-level approximation of global data distribution. Specially, the union of the homogeneous neighborhoods belonging to the same class d, $O_d = \cup\{O(x_i) | \forall x \in O(x_i), class(x) = d\}$ forms an approximation of the data distribution of d. The integration of neighborhoods can approximate data distribution but the neighborhood covering may contain redundant neighborhoods which lead to high model complexity. Therefore, the redundant neighborhoods in initial neighborhood coverings should be further removed to simplify the approximation of data distribution. Referring to Definition 2 and 3, $\forall O(x_i)$, $O(x_j) \in O_d$, if $O(x_i) \subseteq O(x_j)$, $O(x_i)$ is relatively reducible with respect to class d and considered to be redundant. Through neighborhood covering reduction [8], the reducible neighborhoods will be filtered out to generate concise neighborhood coverings of data distribution.

The neighborhood covering introduced above actually provides a set-level approximation of data distribution. Data samples are certainly distributed into neighborhoods, which lead to hard partitions of data space. This strategy is risky to distinguish uncertain data. To handle this problem, we expect to extend the set-level neighborhood coverings of different classes to a membership mapping.

4.2. Fuzzy extension of neighborhood covering

Distinguishing uncertain samples requires to form a soft partition of data space. Thus we expect to construct a membership mapping of different classes for uncertain data classification. To achieve this, we extend the traditional neighborhood coverings to Fuzzy Neighborhood Coverings (FNC). The basic idea of this extension is to quantify the discrete neighborhood belongingness {0, 1} to fuzzy memberships. The data samples will be distributed into neighborhoods according to their memberships with uncertainty. Comparing with the traditional neighborhood covering model, a fuzzy neighborhood covering consists of a group of neighborhood membership functions rather than the neighborhoods of sample sets.

Definition 5 Fuzzy neighborhood covering. Suppose $U = \{x_1, x_2, \ldots x_n\}$ is a data set and $O_U = \{O(x_1), O(x_2), \ldots, O(x_n)\}$ is the set of neighborhoods of data samples. Comparing with the neighborhood covering $\langle U, O_U \rangle$, fuzzy neighborhood covering consists of fuzzy membership functions of neighborhoods $P_{O_U} = \{P_{O(x_1)}, P_{O(x_2)}, \ldots P_{O(x_n)}\}$, in which $P_{O(x_i)}$ denotes the membership function of neighborhood $O(x_i)$ and is briefly denoted as P_{O_i}.

The neighborhood membership functions are used to measure the possibilities of samples belonging to neighborhoods and computed based on the distances between samples and neighborhoods. It is intuitive that the samples far from the neighborhoods should have low memberships and the near ones should have higher possibilities. For the data samples within neighborhoods, their memberships should be close to 1.

Definition 6 Neighborhood membership. Given a data sample x and a neighborhood $O(x_i)$, x_i is the neighborhood center, the possibility of x belonging to $O(x_i)$ is defined based on the distance between x and x_i,

$$P_{O(x_i)}(x) = P_{O_i}(x) = 1 - \frac{1}{1 + e^{-\lambda[d(x,x_i) - \eta - r]}} = \frac{e^{-\lambda[d(x,x_i) - \eta - r]}}{1 + e^{-\lambda[d(x,x_i) - \eta - r]}} \tag{4}$$

in which $d(x, x_i)$ is the distance between x and x_i, $\eta > 0$ is the radius of neighborhood, $\lambda \geq 1$ denotes the distance order and $r \geq 0$ denotes the distance bias. The formula of neighborhood membership is similar to the sigmoid function and $\forall x_i, x, P_{O(x_i)}(x) \in [0, 1]$.

Through investigating the possibilistic measure of neighborhood membership, we obtain the correlation between the sample-neighborhood distance and the neighborhood membership, see Theorem 1.

Theorem 1. *Suppose $O(x_i)$ is a neighborhood, d is the distance between the neighborhood center x_i and a sample x, η is the radius of neighborhood, setting the distance order $\lambda \geq 1$ as a positive integer and the distance bias as a ratio of neighborhood radius $r = \tau \cdot \eta$, $0 \leq \tau < 1$, we infer the following results about the neighborhood membership $P_{O(x_i)}(x)$.*

(1) *If the distance $d(x_i, x) = (1+\tau) \cdot \eta$, $P_{O(x_i)}(x) = 0.5$.*
(2) *If the distance $d(x_i, x) = \eta$, $P_{O(x_i)}(x) = \frac{e^{\lambda \cdot \tau \cdot \eta}}{1+e^{\lambda \cdot \tau \cdot \eta}}$.*
(3) *$P_{O(x_i)}(x) \to 1$ at the distance $\eta + C$, $C \in [-\eta, \tau \cdot \eta)$.*

Proof. (1) and (2) can be directly obtained according to the formula (4), we prove (3) in the following way. Suppose d is the distance between a sample and the neighborhood center, $\varepsilon \to 0$ is a small positive constant, we have

$$
\begin{aligned}
P_{O(x_i)}(x) \to 1 &\Rightarrow \frac{e^{-\lambda \cdot (d-\eta-\tau \cdot \eta)}}{1+e^{-\lambda \cdot (d-\eta-\tau \cdot \eta)}} = 1 - \varepsilon \\
&\Rightarrow e^{-\lambda \cdot (d-(1+\tau) \cdot \eta)} = (1-\varepsilon) \cdot (1 + e^{-\lambda \cdot (d-(1+\tau) \cdot \eta)}) \\
&\Rightarrow \varepsilon \cdot [1 + e^{-\lambda \cdot (d-(1+\tau) \cdot \eta)}] = 1 \\
&\Rightarrow e^{-\lambda \cdot (d-(1+\tau) \cdot \eta)} = \frac{1-\varepsilon}{\varepsilon} \\
&\Rightarrow -\lambda \cdot (d - (1+\tau) \cdot \eta) = \ln\left(\frac{1-\varepsilon}{\varepsilon}\right) \\
&\Rightarrow d = -\frac{1}{\lambda} \ln\left(\frac{1-\varepsilon}{\varepsilon}\right) + (1+\tau) \cdot \eta \\
&\Rightarrow d = \eta + \left[\tau \cdot \eta - \frac{1}{\lambda} \ln\left(\frac{1-\varepsilon}{\varepsilon}\right)\right]
\end{aligned}
$$

Because $\varepsilon \to 0$ is a small positive constant and $\lambda \geq 1$, we have $\frac{1}{\lambda} \ln(\frac{1-\varepsilon}{\varepsilon}) > 0$, thus $d < \eta + \tau \cdot \eta$. It is required that the distance $d \geq 0$, for any given ε, $\exists \lambda$, $\frac{1}{\lambda} \ln(\frac{1-\varepsilon}{\varepsilon}) \leq (1+\tau) \cdot \eta$, thus we have $[\tau \cdot \eta - \frac{1}{\lambda} \ln(\frac{1-\varepsilon}{\varepsilon})] \geq -\eta$ and $d \geq \eta + (-\eta)$. □

From Theorem 1, we know that the neighborhood membership decreases as the sample-neighborhood distance increasing. The distance bias r determines the position of neighborhood membership 0.5. For the samples beyond neighborhoods, if the distance between the samples and the neighborhood boundary is equal to r, the neighborhood membership will be 0.5. When the samples are more distant, their belongingness to the neighborhoods will become less. The distance order λ controls the change rate of membership against distance. Setting a proper λ, we can make the samples within (or nearly within) neighborhoods have the high membership close to 1. Fig. 2 illustrates the variation of neighborhood membership against distance under multiple orders. In this paper work, we set the distance order $\lambda = 1$ and the bias $r = \eta/3$.

Based on the neighborhood memberships, we can further define the possibilistic measure of sample belongingness to neighborhood coverings.

Definition 7 Neighborhood covering membership. Given a neighborhood covering $C = \langle U, O_U \rangle$, $P_{O_U} = \{P_{O(x_1)}, P_{O(x_2)}, \ldots, P_{O(x_n)}\}$ is the corresponding fuzzy neighborhood covering, the membership of a sample x belonging to C is defined by the maximum neighborhood membership,

$$P_C(x) = \max_{O(x_i) \in O_U} \{P_{O(x_i)}(x)\} = \max_{O_i \in O_U} \{P_{O_i}(x)\} \tag{5}$$

Based on the membership of neighborhood covering, we can directly represent the possibilities of data samples belonging to a specified class. For all the samples with class label d, $U_d = \{x | x \in U \wedge class(x) = d\}$, we construct the neighborhood covering C_d and further extend C_d to the fuzzy neighborhood covering $P_{C_d}(x) = \max_{O_i \in O_{U_d}} \{P_{O_i}(x)\}$, $P_{C_d}(x)$ denotes the possibility of a data sample x belonging to the class d. Fig. 3 shows the membership distribution of the neighborhood covering of one class.

4.3. Three-way classification with fuzzy neighborhood covering

Based on the fuzzy neighborhood covering, we can construct the possibilistic representation of different classes and thereby implement data classification. The neighborhood memberships provide an uncertain measure of sample belongingness to different classes and facilitate uncertain classification. Referring to the methodologies of three-way decisions, we formulate the Three-Way Classification with Fuzzy Neighborhood Covering (3WC-FNC). Similar to the three-way decision

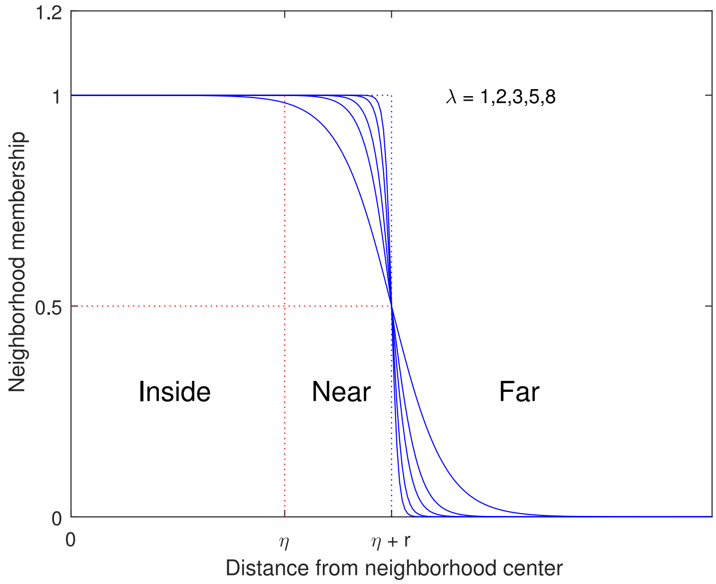

Fig. 2. Neighborhood membership.

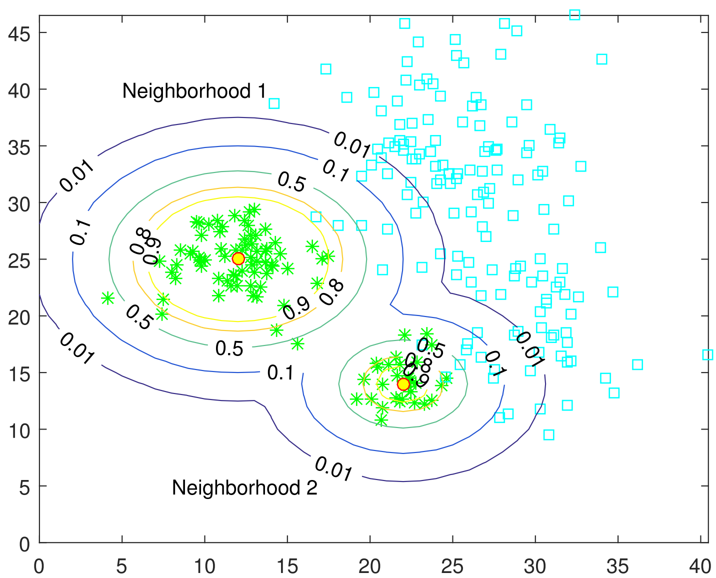

Fig. 3. Membership distribution of neighborhood covering.

making, through thresholding the memberships of neighborhood coverings, the three-way classifiers based on FNC classify the samples into three cases: Positive (certainly belonging to a class), Negative (certainly beyond a class) and Uncertain (uncertain belongingness with respect to a class).

Definition 8 Three-way classification with fuzzy neighborhood covering. For a class d, suppose $C_d = \langle U_d, O_{U_d} \rangle$ is the neighborhood covering of d and $P_{C_d}(x)$ is the fuzzy membership representation of C_d. Given a couple of membership thresholds α and β, $0 \le \beta < \alpha \le 1$, the three-way classification of a sample x for class d is defined as follows.

$$C_{\alpha,\beta}(x, d) = \begin{cases} Positive \quad (certainly\ belonging\ to\ d), & P_{C_d}(x) \ge \alpha \\ Uncertain \quad (uncertain\ for\ d), & \beta < P_{C_d}(x) < \alpha \\ Negative \quad (certainly\ beyond\ d), & P_{C_d}(x) \le \beta \end{cases} \tag{6}$$

If we just focus on the classification of one specific class d, we can directly perform three-way classification according to the formula (6). If the membership of x belonging the neighborhood covering C_d is no less than the upper threshold $P_{C_d}(x) \ge \alpha$, x is certainly classified into the class d. If $P_{C_d}(x) \le \beta$, x is certainly classified into the negative class of d. If

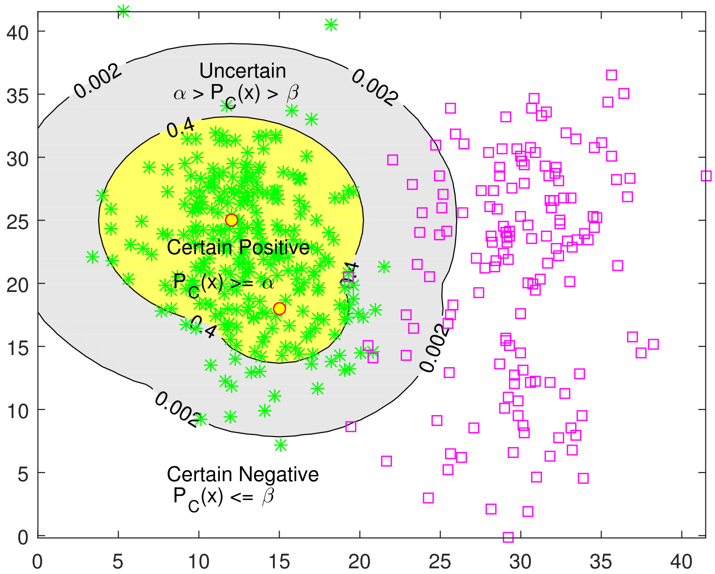

Fig. 4. Three-way classification with FNC.

$\beta < P_{C_d}(x) < \alpha$, x is judged as an uncertain sample with respect to the class d. Fig. 4 illustrates the one-class three-way classification with fuzzy neighborhood covering.

For multi-class problems, to classify an unknown sample x, we compute its memberships belonging to the neighborhood coverings of different classes $\{P_{C_{d_1}}(x), P_{C_{d_2}}(x), \dots, P_{C_{d_m}}(x)\}$ and perform three-way classification according to the memberships. It is intuitive to certainly classify the sample x into a class iff there exists at least one neighborhood covering to judge x as certain positive. If there is only one neighborhood covering P_{C_d} to judge x as certain positive, we certainly classify x into the class d. For the cases of multiple certain positive judgments by heterogeneous neighborhood coverings, we assign x to the class d that has the maximum membership $d = \arg\max_{d_i}\{P_{C_{d_i}} | P_{C_{d_i}}(x) >= \alpha_{d_i}\}$. For the non-positive cases, the classification criteria will be relaxed, if the sample x is judged as uncertain for one class d, and in the meantime judged negative for all the other classes, x can be classified into class d. The sample x is judged as negative for all the classes iff all the neighborhood coverings judge x as certain negative, i.e. $\forall C_{d_i}, P_{C_{d_i}}(x) < \beta_{d_i}$. Otherwise, x is classified as an uncertain sample. To avoid the blind area of classification, the negative samples that are rejected by all the classes can also be considered as uncertain in the algorithm implementation. The process of multi-class Three-Way Classification with Fuzzy Neighborhood Coverings (3WC-FNC) is formally presented in Algorithm 1.

5. Experimental results

Fuzzy Neighborhood Covering (FNC) is the fuzzy extension of set-level neighborhoods and provides a more flexible way to approximate data distributions. Based on the fuzzy neighborhood coverings of different classes, we propose a three-way classification method (3WC-FNC) . In contrast to traditional certain classifiers, the proposed three-way classifier separates the uncertain data and leads to robust classification results. To validate this, we implement two experiments. The first experiment aims to test the ability of 3WC-FNC for uncertain data classification. In the second experiment, we overall evaluate the performances of 3WC-FNC through comparing with other typical classification methods. To demonstrate the three-way strategy is effective to reduce the risk of classification, we adopt multiple medical and economic data sets in the experiments. The data sets are collected 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 adopted UCI data sets are given in Table 1.

To overall evaluate the performance of the proposed three-way classification method, we adopt the measures of Accuracy, Precision, Recall, F1 Score, Uncertain Ratio (UR) and Classification Cost as the evaluation criteria. Given a data set, suppose the number of the positive-class samples is P and the number of the negative-class samples is N, classifying the data samples with classifiers, TP and FP denote the numbers of true positive classified samples and false positive samples, TN and FN denote the numbers of true negative samples and false negative samples respectively. Based on the statistics above, the accuracy, precision, recall rate and F1 score are computed as follows to evaluate the quality of classification results.

$$accuracy = (TP + TN)/(P + N)$$

$$precision = TP/(TP + FP)$$

Algorithm 1 3WC-FNC.

Input: Fuzzy neighborhood covering of m classes $P_C, C = C_{d_1} \cup \ldots \cup C_{d_m}$;
Membership thresholds of m classes, $< \alpha_{d_1}, \beta_{d_1} >, \ldots, < \alpha_{d_m}, \beta_{d_m} >$;
Unknown sample x;
Output: Three-way classification result of x, $Cls(x)$;
Separate P_C into m fuzzy neighborhood coverings of different classes, $P_C = \{P_{C_{d_1}}, \ldots, P_{C_{d_m}}\}$;
Compute the memberships of x for neighborhood coverings $P_{C_{d_1}}(x), \ldots, P_{C_{d_m}}(x)$;
Initialize classification results of x for m classes, $Cls(x, d_1), \ldots, Cls(x, d_m) \leftarrow \phi$;
for each class d_i, $i = 1, \ldots, m$ **do**
  if $P_{C_{d_i}}(x) \geq \alpha_{d_i}$ **then**
    $Cls(x, d_i) =$ positive;
  else
    if $P_{C_{d_i}}(x) \leq \beta_{d_i}$ **then**
      $Cls(x, d_i) =$ negative;
    else
      $Cls(x, d_i) =$ uncertain;
    end if
  end if
end for
$PD = \{d_i | Cls(x, d_i) = \text{positive}\}$;
$ND = \{d_j | Cls(x, d_j) = \text{negative}\}$;
$UD = \{d_j | Cls(x, d_j) = \text{uncertain}\}$;
if $|PD = \{d\}| = 1$ **then**
  $Cls(x) =$ positive d;
else
  if $|PD| > 1$ **then**
    $d = \arg\max_{d_i \in PD} \{P_{C_{d_i}}(x)\}$, $Cls(x) =$ positive d;
  else
    if $|UD = \{d\}| = 1$ and $|ND| = m - 1$ **then**
      $Cls(x) =$ positive d;
    else
      $Cls(x) =$ uncertain;
    end if
  end if
end if
Return $Cls(x)$.

Table 1
Experimental data sets.

Data sets	*Feature*	*Instance*	*Class ratio*	*Attribute type*
Australian Credit	14	690	45% vs. 55%	Mixed
Banknote Authentication	4	1372	44% vs. 56%	Numerical
Breast Cancer (Diagnostic)	32	569	37% vs. 63%	Numerical
Breast Cancer (Original)	10	699	35% vs. 65%	Mixed
Diabetes	8	768	35% vs. 65%	Mixed
Mammographic Mass	6	961	49% vs. 51%	Numerical
Sonar	60	208	47% vs. 53%	Numerical
Vertebral Column	6	310	32% vs. 68%	Numerical

$$recall = TP/(TP + FN)$$

$$F1 = 2 \cdot precision \cdot recall/(precision + recall)$$

Besides the measures of classification precision, we also adopt the uncertain ratio, i.e. the ratio of classified uncertain samples, to evaluate the ability of classifiers to distinguish uncertain cases.

$$ur = |Uncertain(U)|/|U|$$

Assuming correct classifications cause no cost, λ_{NP}, λ_{PN}, λ_U denote the costs of false-positive classification, false-negative classification and uncertain case classification respectively, the total classification costs are defined in the following formula to measure the classification risk.

$$\cos t = \lambda_{NP} \cdot \frac{FP}{P + N} + \lambda_{PN} \cdot \frac{FN}{P + N} + \lambda_U \cdot ur$$

In the medical and economic data sets, the minimum class generally denotes the class of high classification risk, such as the class 'malignant' in Breast Cancer data. Thus we assume the minimum class as positive class and set $\lambda_{PN}/\lambda_{NP}/\lambda_U = 5/1/0.5$ in the following experiments.

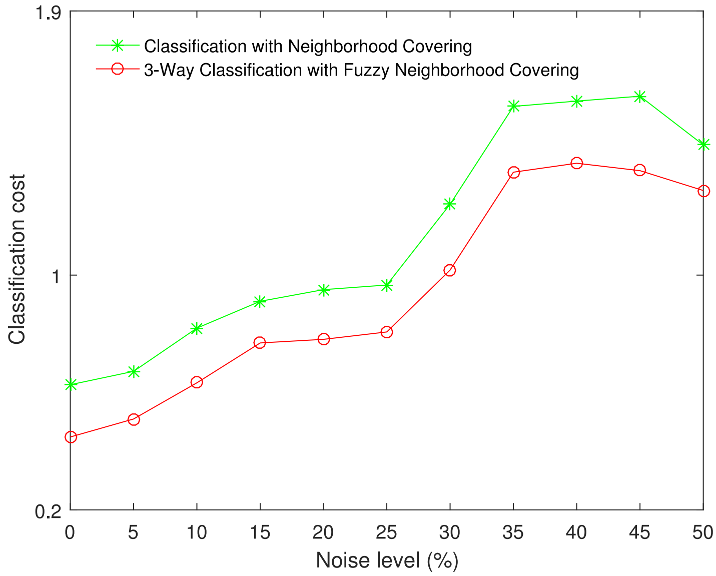

Fig. 5. Classification costs of 3WC-FNC and NC on multilevel noisy data.

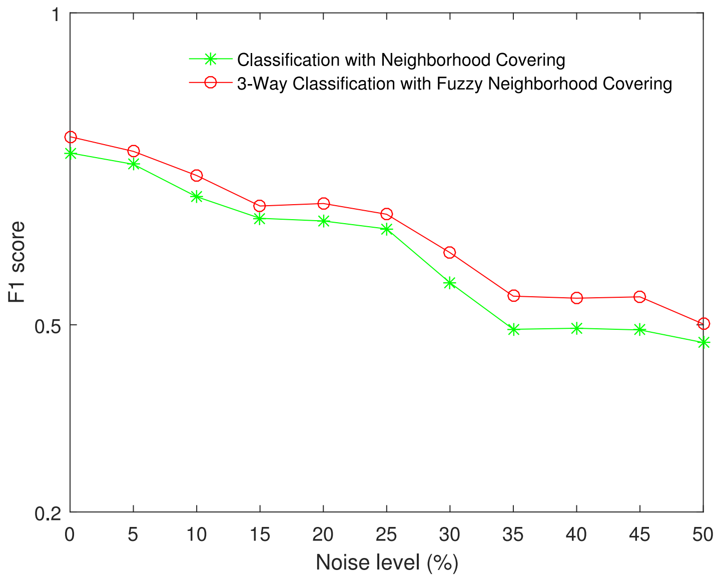

Fig. 6. F1 scores of 3WC-FNC and NC on multilevel noisy data.

5.1. Evaluation on uncertain classification

This experiment involves two tests to validate the ability of the proposed 3WC-FNC for uncertain classification. First we expect to demonstrate the fuzzy extension of neighborhood covering is helpful to improve the classification of uncertain cases. Because the inconsistency between training data and test data will bring about uncertain cases for classification, we produce the uncertain cases for classification through adding noise to test data. Adding multilevel noise to test data to produce the multi-grade uncertain data, we compare the classification with traditional neighborhood covering (NC) and the three-way classification with fuzzy neighborhood covering (3WC-FNC). Figs. 5–8 show the classification cost, F1 score, precision and accuracy of NC and 3WC-FNC methods against the noise level from 0% to 50% and Table 2 presents the corresponding classification results.

It can be found that the three-way classification with FNC achieves better performance than the certain classification with NC on multi-grade uncertain data. The fuzzy neighborhood membership and the three-way classification strategy facilitate to distinguish the uncertain cases and thereby greatly reduce the classification cost. Moreover, 3WC-FNC produces the higher precision and recall rate, which means the more precise classification for the positive class (risky class). Separating uncertain

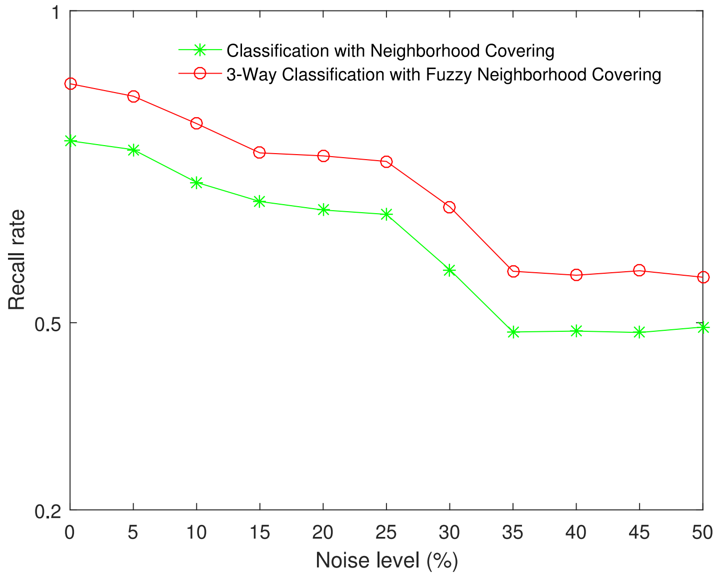

Fig. 7. Recall rates of 3WC-FNC and NC on multilevel noisy data.

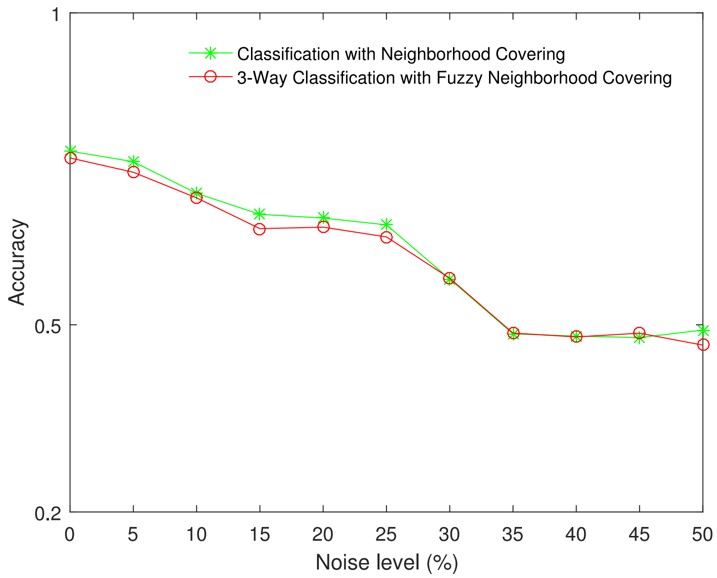

Fig. 8. Classification accuracies of 3WC-FNC and NC on multilevel noisy data.

cases without assigning classes, 3WC-FNC produces lower accuracy on low-level noisy data than the certain classification method. But for the classification on high-level noisy data, as the certain misclassification increases, 3WC-FNC achieves the similar accuracy as the certain NC-based classification.

In the second test, we further compare the proposed three-way classification method (3WC-FNC) with a typical Three-Way Decision (3WD) model based on Probabilistic Attribute Reduction [35]. Probabilistic attribute reduction formulates three-way decision rules through constructing the probabilistic attribute reducts, which partition samples into positive, negative and boundary cases for a given class. Different from the neighborhood covering handling mixed-type data, probabilistic attribute reduction is used to extract decision rules from discrete data. For the tests on mixed/numerical data, we apply the supervised Multi-interval Discretization methods (MDL) and the unsupervised Equal-width Discretization methods (5 bins and 3 bins) [10,31] to discretize the numerical attribute values into discrete ones. Fig. 9 illustrates the classification results of 3WC-FNC and 3WD with different discretization strategies and Table 3 presents the details. We find that the 3WD based on attribute reduction is sensitive to discretization methods. The preprocessing of discretization may bring about the information loss and thus leads to low-quality decision rules. Depending on the superiority of neighborhood covering on mixed-type data, 3WC-FNC achieves better performances.

Table 2
Classification results on multilevel noisy data.

Noise level	*Methods*	*Cost*(10^{-2})	*Accuracy* (%)	*Precision* (%)	*Recall* (%)	*F1 score* (%)
No noise	3WC-FNC	44.8	76.7	73.5	88.3	80.1
	NC	62.6	77.8	76.3	79.2	77.5
5%	3WC-FNC	51	74.5	71.1	86.3	77.8
	NC	67.2	76.1	74.3	77.7	75.7
10%	3WC-FNC	63.3	70.3	67.6	81.9	73.9
	NC	81.9	71.1	69	72.5	70.5
15%	3WC-FNC	76.9	65.4	62.7	77.2	69
	NC	91.1	67.7	65.5	69.4	67.1
20%	3WC-FNC	78.1	65.6	63.7	76.7	69.4
	NC	95	67.1	65.6	68.1	66.6
25%	3WC-FNC	80.6	64.1	61.6	75.8	67.7
	NC	96.6	66	63.9	67.4	65.4
30%	3WC-FNC	101.7	57.5	56.4	68.6	61.6
	NC	124.1	57.3	55.8	58.5	56.7
35%	3WC-FNC	135.1	48.7	51.6	58.2	54.6
	NC	157.5	48.6	50.2	48.5	49.3
40%	3WC-FNC	138.1	48.1	51.7	57.6	54.3
	NC	159.3	48.2	50.6	48.7	49.5
45%	3WC-FNC	135.6	48.7	52	58.3	54.5
	NC	161	48	50.9	48.4	49.2
50%	3WC-FNC	128.8	46.7	44.8	57.3	50.1
	NC	144.4	49.2	45.4	49.3	47.2

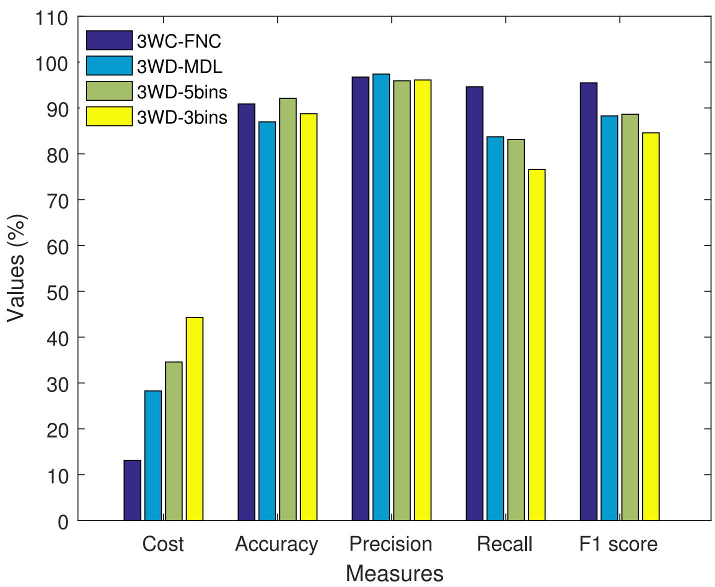

Fig. 9. Classification results of fuzzy neighborhood covering (3WC-FNC) and probabilistic attribute reduction (3WD- discretization).

Table 3
Classification results of 3WC-FNC and 3WD with discretization.

Methods	*TP* (%)	*FN* (%)	*ur*(%)	*Cost* (10^{-2})	*Acc* (%)	*Prec* (%)	*Recall* (%)	*F1* (%)
3WC-FNC	85	4.72	6.12	13.11	90.88	96.74	94.63	95.48
3WD-MDL	76.36	13.39	7.42	28.28	86.97	97.39	83.71	88.27
3WD-5bins	83.13	16.87	0	34.58	92.1	95.93	83.13	88.62
3WD-3bins	74.6	22.28	1.4	44.29	88.75	96.11	76.59	84.58

5.2. Overall evaluation

The second experiment overall evaluates the proposed 3WC-FNC method. Focusing on the evaluation of classification risk, we compare 3WC-FNC with three elegant classification methods: Naive Bayes, Support Vector Machine (SVM) and Decision Trees (J48), and other three typical cost-sensitive classification methods: Cost-sensitive Bayes, Cost-sensitive Decision Trees

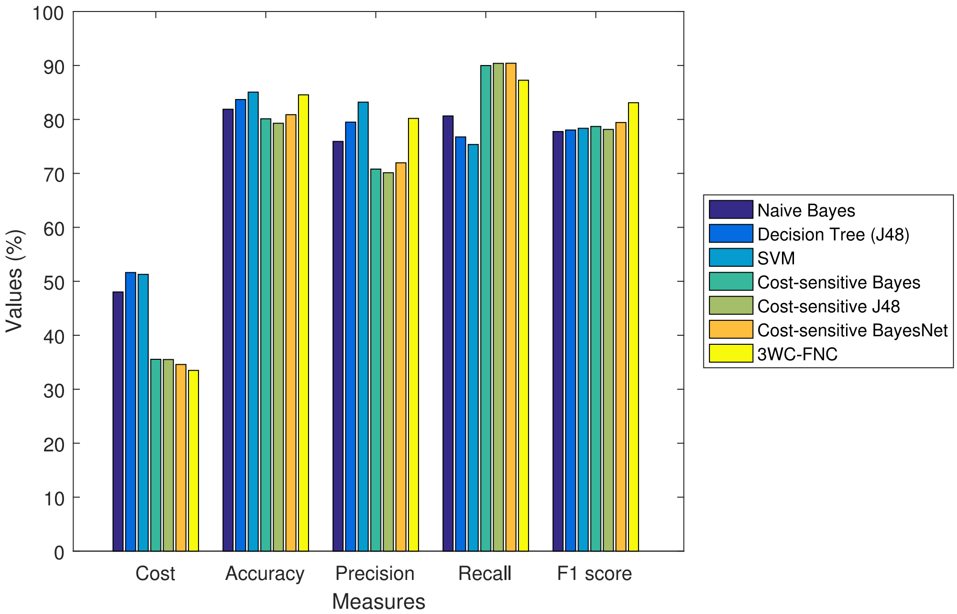

Fig. 10. Classification results of different classification methods.

Table 4
Overall evaluation of different classification methods.

Methods	*TP* (%)	*TN* (%)	*Cost* (10^{-2})	*Acc* (%)	*Prec* (%)	*Recall* (%)	*F1* (%)
Naive Bayes	80.65	82.01	48.03	81.89	75.93	80.65	77.76
Decision-Tree (J48)	76.76	87.59	51.63	83.69	79.5	76.76	78.04
SVM	75.36	89.79	51.29	85.06	83.2	75.36	78.36
Cost-sensitive Bayes	89.99	73.29	35.55	80.12	70.8	89.99	78.7
Cost-sensitive J48	90.39	72.45	35.51	79.29	70.12.	90.39	78.15
Cost-sensitive Bayes Net	90.41	74.94	34.59	80.88	71.96	90.41	79.43
3WC-FNC	85.75	82.96	33.49	84.56	80.19	87.26	83.10

and Cost-sensitive Bayes Net. We perform the classification methods on all the test data sets, Fig. 10 illustrates the average classification results for each method and the details are presented in Table 4.

From the experimental results, we find that 3WC-FNC and SVM obtain the top precise results, and 3WC-FNC achieves the lowest classification cost. Because of the abstaining uncertain samples in three-way classification, 3WC-FNC produces slightly lower accuracy and precision than SVM. However, without considering the classification risks, SVM suffers much classification costs. Considering the classification risks of different classes, all the cost-sensitive methods achieve lower classification costs. To reduce the classification costs, the cost-sensitive methods tend to over classify the negative samples into the positive class (risky class) and thus lead to the low precision. Through separating uncertain samples according to memberships, 3WC-FNC can balance the classification precision and recall rate and therefore induce the highest F1 score. In general, the three-way classifier based on fuzzy neighborhood coverings achieves better performances than the certain classifiers. The three-way strategy and the fuzzy representation of neighborhood belongingness are effective to reduce the classification cost, and in the meantime guarantee the precise classification results.

6. Conclusion

To improve the classification with neighborhood covering for uncertain data, in this paper, we propose a three-way classification method with fuzzy neighborhood covering. The research work involves two parts: fuzzy extension of neighborhood covering and three-way classification with fuzzy neighborhood covering. Rather than neighborhoods, fuzzy neighborhood covering consists of neighborhood membership functions and the neighborhood belongingness is quantified into continuous membership values. Based on the fuzzy neighborhood membership, we further formulate the three-way classification. Data samples are classified into three cases of Positive (certainly belonging to a class), Negative (certainly beyond classes) and Uncertain through thresholding memberships. The separation of uncertain samples facilitates to reduce the classification risk. Experiments verify the effectiveness of the proposed three-way method for uncertain data classification. Our future work will focus on the following issues. The first issue is to further investigate the parameter optimization strategy which involves the uncertain classification results. Second, the memberships are computed based on Euclidean distances which is not effective in the extremely high-dimensional data space, therefore we try to utilize the kernel methods to construct neighborhood coverings to implement uncertain classification on high-dimensional data.

Acknowledgment

This work was supported by National Natural Science Foundation of China (Serial Nos. 61573235, 61673301).

References

[1] R.A. Aliev, W. Pedrycz, B.G. Guirimov, Type-2 fuzzy neural networks with fuzzy clustering and differential evolution optimization, Inf. Sci. 181 (9) (2011) 1591–1608.
[2] 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.
[3] Y.F. Chen, X.D. Yue, H. Fujita, S.Y. Fu, Three-way decision support for diagnosis on focal liver lesions, Knowl. Based Syst. 127 (2017) 85–99.
[4] D. Ciucci, D. Dubois, J. Lawry, Borderline vs. unknown: comparing three-valued representations of imperfect information, Int. J. Approximate Reasoning 55 (2014) 1866–1889.
[5] D. Ciucci, Orthopairs and granular computing, Granular Comput. 1 (2016) 1–12.
[6] X.F. Deng, Y.Y. Yao, Decision-theoretic three-way approximations of fuzzy sets, Inf. Sci. 279 (2014) 702–715.
[7] Z.H. Deng, K.S. Choi, L.B. Cao, S.T. Wang, T2FELA: Type-2 fuzzy extreme learning algorithm for fast training of interval type-2 TSK fuzzy logic system, IEEE Trans. Neural Netw. Learn. Syst. 25 (4) (2014) 664–676.
[8] 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.
[9] C. Elkan, The foundations of cost-sensitive learning, in: Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJ-CAI'01), Kaufmann, Seattle, 2001, pp. 973–978.
[10] M.U. Fayyad, B.I. Keki, Multi-interval discretization of continuousvalued attributes for classification learning, in: Proceedings of the Thirteenth International Joint Conference on Articial Intelligence (IJCAI'93), 1993, pp. 1022–1027.
[11] P.A. Flach, S. Wu, Repairing concavities in ROC curves, in: Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI'05), Edinburgh, Scotland, 2005, pp. 702–707.
[12] H. Fujita, T.R. Li, Y.Y. Yao, Advances in three-way decisions and granular computing, Knowl. Based Syst. 91 (2016) 1–3.
[13] 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.
[14] 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.
[15] 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.
[16] 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.
[17] Q.H. Hu, D.R. Yu, Z.X. Xie, Neighborhood classifier, Expert Syst. Appl. 34 (2008) 866–876.
[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 - application to qualitative fuzzy and rough sets, Advances in machine intelligence and soft computing, Duke University, NC, 1997.
[21] A. Owen, A neighborhood-based classifier for LANDSAT data, Can. J. Stat. 12 (1984) 191–200.
[22] Z. Pawlak, Rough sets, Int. J. Inf. Comput. Sci. 11 (5) (1982) 314–356.
[23] Z. Pawlak, Some issues on rough sets, Trans. Rough Sets I 3100 (2004) 1–58.
[24] W. Pedrycz, Collaborative fuzzy clustering, Pattern Recognit. Lett. 23 (14) (2002) 1675–1686.
[25] J.F. Peters, S. Ramanna, Proximal three-way decisions: theory and applications in social networks, Knowl. Based Syst. 91 (2016) 4–15.
[26] R.D. Short, K. Fukunaga, Optimal distance measure for nearest neighbor classification, IEEE Trans. Inf. Theory 27 (1981) 622–627.
[27] C.Z. Wang, Q.H. Hu, X.Z. Wang, Y.H. Qian, Feature selection based on neighborhood discrimination index, IEEE Trans. Neural Netw. Learn. Syst. 99 (2017) 1–14.
[28] M. Wang, F. Min, Z.H. Zhang, Y.X. Wu, Active learning through density clustering, Expert Syst. Appl. 85 (2017) 305–317.
[29] D. Wettschereck, T.G. Dieterich, An experimental comparison of the nearest neighbor and nearest-hyper-rectangle algorithms, Mach. Learn. 19 (1995) 5–27.
[30] D.R. Wilson, T.R. Martinez, Improved heterogeneous distance functions, J. Artif. Intell. Res. (1997) 1–34.
[31] I.H. Witten, E. Frank, M.A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann Publisher, 2005.
[32] T. Yang, Q. Li, Reduction about approximation spaces of covering generalized rough sets, Int. J. Approximate Reasoning 51 (2010) 335–345.
[33] J. Yao, A.V. Vasilakos, W. Pedrycz, Granular computing: perspectives and challenges, IEEE Trans. Cybern. 43 (6) (2013) 1977–1989.
[34] Y.Y. Yao, Relational interpretations of neighborhood operators and rough set approximation operators, Inf. Sci. 111 (1998) 239–259.
[35] Y.Y. Yao, Three-way decisions with probabilistic rough sets, Inf. Sci. 180 (2010) 341–353.
[36] Y.Y. Yao, The superiority of three-way decisions in probabilistic rough set models, Inf. Sci. 181 (2011) 1080–1096.
[37] R. Younsi, A. Bagnall, An efficient randomised sphere cover classifier, Int. J. Data Mining Model. Manage. 4 (2) (2012) 156–171.
[38] X.D. Yue, Y.F. Chen, D.Q. Miao, J. Qian, Tri-partition neighborhood covering reduction for robust classification, Int. J. Approximate Reasoning 83 (2017) 371–384.
[39] B.W. Zhang, F. Min, D. Ciucci, Representative-based classification through covering-based neighborhood rough sets, Appl. Intell. 43 (4) (2015) 840–854.
[40] H.R. Zhang, F. Min, Three-way recommender systems based on random forests, Knowl. Based Syst. 91 (2016) 275–286.
[41] W. Zhu, F. Wang, Reduction and axiomization of covering generalized rough sets, Inf. Sci. 152 (2003) 217–230.
[42] W. Zhu, F. Wang, On three types of covering-based rough sets, IEEE Trans. Knowl. Data Eng. 19 (2007) 1131–1144.