Improved Naive Possibilistic Classifiers for Imprecise Data

Jing Bai, Youlong Yang, and Jianming Xie

Abstract—**In real-world problems, input data may be uncertain or imprecise. Naïve possibilistic classifiers (NPC) based on possibility theory have been proposed for classification tasks in this situation. However, two strong assumptions: attributes independence and their equal importance, ignoring the dependent relationship among attributes, affect the classification performance. In this paper, three improved weighted NPC algorithms are presented for classifying imprecise data efficiently by relaxing these assumptions. We first present a weight based on the attribute contribution in order to take the effect of the different values of each attribute on classification into consideration. Then, another weight is presented on the basis of the non-specificity gain for each attribute, which measures the significance of each attribute. Whereafter, we introduce three improved algorithms, generated by two weights and the merged weight respectively. Finally, we make a comparison between three improved algorithms and the NPC. And a series of numeric studies are performed using a broad range of data sets for different traditional algorithms. Extensive experiments show that three proposed algorithms all have higher classification accuracy than NPC and other traditional algorithms in most data sets. Even the algorithm after merging two weights has better performance than others.**

Index Terms—**Classification, Possibility theory, Naïve possibilistic classifier, Specificity of gain, Imprecise data, Weight.**

I. Introduction

CLASSIFICATION, used to predict class value for an unknown instance, is one of the important tasks in machine learning and data mining applications [1]. The task is mainly divided into the training phase and testing phase. In the training phase, one produces a classifier from a set of training samples described by an attribute set with known class values. Once the classifier is constructed, it can assign the class value to new instances given the known values of their attributes, which constitutes the testing phase.

There have been various algorithms for classification tasks, including decision trees (TREE) [2], support vector machine (SVM) [3], [4], neural network [5], [6], bayesian network [7], [8], [9], and so on. Among these approaches, naïve bayesian classifier (NBC) [10], as a specific kind of bayesian network is one of the most efficient classifiers. And its classified performance is competitive with state-of-the-art classifiers. NBC based on probability theory, is very effective in the domains of uncertainty. However, it confronts problems when faced with imperfect data.

Manuscript received May 11, 2017; revised September 6, 2017. This work was supported by the National Natural Science Foundation of China (Grant No. 61573266).

Jing Bai is with the Department of Mathematics and Statistics, Xidian University, Xian, Shaanxi 710071, China e-mail:bj126726@126.com.

Youlong Yang is with the Department of Mathematics and Statistics, Xidian University, 266 Xinglong Section of Xifeng Road, Xian, Shaanxi 710126, China, e-mail: ylyang@mail.xidian.edu.cn.

Jianming Xie is with the Department of Mathematics and Statistics, Xidian University, Xian, Shaanxi 710071, China.

Almost all of classification techniques deal with the perfect data in which each instance is assigned to a certain class value. In fact, compared with an accurate label, it is more appropriate for an expert to give the possibility degree of the label to which an instance belongs. For example, a doctor cannot diagnose the exact disease of a patient sometimes, and a banker cannot determine whether or not to give a loan for a client sometimes. Hence, the inconsistency of information, especially the uncertainty among labels, may appear if several experts are consulted. As traditional classifiers ignore such data, there is no doubt that they are not suitable to classify in such situation.

In order to cope with imprecise data, several theories of uncertainty have been proposed, such as fuzzy set theory [11], [12], rough set theory [13], [14], evidence theory [15] and possibility theory [16]. In this paper, we are interested in the classification approach based on possibility theory. Possibility theory was introduced initially by Zadeh [17] and developed by Dubois and Prade [18]. Although both of them can deal with the uncertainty, one of the biggest differences between them is that possibility theory can represent not only uncertain but also imprecise information [19]. Naïve possibilistic classifiers (NPC) [20] based on possibility theory have been proposed for classification tasks. Subsequently, Myriam Bounhas et al. have pushed the research on NPC one step further from discrete attributes to numerical attributes [21]. And experiments demonstrated that the NPC has a robust behavior when coping with imprecise data. Then, decision trees as possibilistic classifiers was proposed by Jenhani [22]. In recent years, Baati et al. applied the NPC for diagnosis of lymphatic diseases and studied the modified NPC based on minimum algorithm [23]. In spite of the fact that possibility distributions are useful for representing imperfect knowledge, there is only little further research on the application of using possibility theory for classification tasks.

As a counterpart to bayesian classifier, the NPC also estimates conditional joint possibility of a set of attributes. In order to solve the reasoning calculation which is a NP-hard problem, algorithm is implemented under two strong assumptions: attributes independence and their equal importance. However, strong assumptions mentioned above are rarely true in reality. And the NPC sometimes shows poor classification performance in many data sets. In addition, with the number of imprecise instances in training data increasing, the classification accuracy gradually reduces [20]. Therefore, it is natural to improve the NPC by weakening these strong assumptions. Tree augmented naïve possibilistic classifier (TANPC) is proposed [24], which is a modified approach for structure extension.

To overcome the limitation, this paper focuses on feature weighting approaches for the NPC. As is known to all, many

(Advance online publication: 10 February 2018)

feature weighting algorithms have been especially designed for NBC, but there are only few further works for NPC. The paper tries to propose two types of weighting techniques. The main goal is to improve the existing NPC by relaxing strong assumptions, making it more suitable for many real data sets. Since the conditional possibility in NPC only considers the proportion of each attribute value given the class value to the whole number of this attribute value. Firstly, we consider the contribution degree of different attribute values for classification. Specifically, the proportion of each attribute value given the class value to the whole number of this class value, as the weight, is added to each conditional possibility. Then, in order to relax the strong assumption of equal importance, we use the specificity gain measure to calculate the importance of each attribute as the amount of information. Moreover, we incorporate two weights as a weight in classification task. As a result, the paper proposes three feature weighting algorithms for NPC. Compared to the traditional NPC, our feature weighting approaches show higher classification accuracy.

These algorithms not only consider the effect on classification by different attribute values, but also consider the significance of each attribute to the classification task. The main contributions stand out as follows:

- Take the contribution degree of different attribute values for classification into account, to reduce the influence of a small number of attribute values on classification tasks.
- Use the specificity gain measure to evaluate the importance of each attribute to the class system, to relax the equal importance assumption among attributes.
- Conduct experiments from two aspects according to the change of imprecise degrees, namely perfect and imprecise labeled data. Meanwhile, compare the improved NPCs with traditional classifiers. And the final results show the better performance of our algorithms.

The remainder of the paper is structured as follows. In section 2, we review the basics of possibility theory. Section 3 introduces the mechanism of NPC. Section 4 proposes feature weighting approaches for NPC. Section 5 presents in detail the experimental setup and results about improved classifiers for prefect and imprecise data, and shows the efficiency of weighted naïve possibilistic classifiers for handing uncertainty data. Finally, section 7 makes a conclusion and discusses some future research directions.

II. Basics of Possibility theory

In this section, we will provide a brief recalling on possibility theory. Then, the specificity gain as a counterpart of Shannon entropy, used to represent a measure of uncertainty information for possibility theory, is introduced.

A. Basic notions of possibility theory

Possibility theory, a non-classical theory, was initiated by Zadeh [17] and developed by Dubois and Prade [25]. The fundamental concept of possibility theory is possibility distribution which plays the same role with probability distribution in probability theory.

Definition 1. Let $\Omega = \{\omega_1, \omega_2, \cdots, \omega_n\}$ denote an universe of discourse. A possibility distribution, denoted as π, is defined to be a mapping from the domain of discourse Ω to a totally ordered scale $L \in [0, 1]$.

The value $\pi(\omega_i)$ is called possibility degree, which means to what extent ω_i is consistent with the truth value of the variable. As required, $\pi(\omega_i) = 1$ implicates that ω_i as the value of the variable is totally possible, $\pi(\omega_i) = 0$ denotes that it is impossible that ω_i is the value of the variable, and $\pi(\omega_i) > \pi(\omega_j)$ means that ω_i is more plausible than ω_j as the value of the variable.

Possibility theory relies on two dual possibility and necessity measures.

Definition 2. Possibility and necessity measures are defined for an event $A \subseteq 2^\Omega$ by the following formulas, respectively:

$$\Pi(A) = \max_{\omega \in A} \pi(\omega), \tag{1}$$

$$N(A) = \min_{\omega \notin A} (1 - \pi(\omega)) = 1 - \Pi(\overline{A}), \tag{2}$$

where $\Pi(A)$ evaluates at which level A is consistent with our knowledge encoded by π, namely degrees of plausibility, while the necessity measure $N(A)$ evaluates at which level A is certainly implied by our knowledge represented by π, namely degrees of belief.

Indeed, it has been proved that a possibility measure is between probability measures and necessity measures, namely,

$$\forall P \in P(\Pi), \forall A, \Pi(A) \geq P(A) \geq N(A). \tag{3}$$

It is noteworthy that a possibility as an upper probability is the starting point for transforming a possibility distribution into a probability distribution and conversely [26], [27].

Conditioning is also an essential notion in possibility theory. In various publications, different types of conditional possibility have been introduced. Dubois and Prade [28] defined a conditional possibility according to a counterpart of Bayes rule.

Definition 3. Conditional possibility can be defined as follows:

$$\Pi(A \cap B) = \Pi(A \mid B) * \Pi(B), \tag{4}$$

where $*$ can be chosen as the minimum or the product operator [29], [30]. The min operator is applied to the qualitative setting, while the product is suitable in the quantitative situation. In the paper, we mainly focus on the quantitative setting.

B. Non-specificity

In various uncertainty theories, different uncertainty measures are applied to represent different types of uncertainty. As mentioned in the literature [31], [32], Higashi and Klir proposed that possibility theory deals with a source of uncertainty, namely, non-specificity. The non-specificity [33], as a counterpart of Shannon entropy in probability theory, represents a measure of uncertainty information for possibility theory.

Let π and π' represent two possibility distributions on Ω respectively, and π is more specific than π' if and only if, for each $\omega \in \Omega$, $\pi(\omega) \leq \pi'(\omega)$. Obviously, the more specific possibility distribution has, the more information it brings.

The first measure of non-specificity was proposed by Hartley for classical set theory [31]. Then the majority of non-specificity measures proposed for other uncertainty frameworks (e.g. evidence theory, fuzzy set theory, possibility theory, etc.) represent a generalization of Hartley function. In the possibilistic setting, the measure of non-specificity, called U-uncertainty, has the form: $U : R \rightarrow IR^+$, where R denotes the set of all finite and ordered possibility distributions.

Definition 4. Given an ordered possibility distribution $\pi =< \pi_1, \pi_2, \cdots, \pi_n >$ such that $1 = \pi_1 \geq \pi_2 \geq \cdots \geq \pi_n$. The U-uncertainty of π is given by the formula:

$$U(\pi) = \sum_{i=1}^{n} (\pi_i - \pi_{i+1}) \log_2 i, \quad (5)$$

where $\pi_{n+1} = 0$ by convention. Note that the range of U is $[0, \log_2 n]$. $U(\pi) = 0$ is obtained for the case of complete knowledge (no uncertainty) and $U(\pi) = \log_2 n$ is reached for the case of total ignorance.

In some areas of decision-making, one can measure the amount of uncertainty in order to decide which one is the most informative. In the paper, we use non-specificity gain other than the mutual information to measure the important degree of attributes for the classification system in possibility setting.

III. Naïve possibilistic classifier

In this section, the NPC will be presented in detail. NPC, applying possibility theory in classification problems, was derived from Bayesian classifiers. Therefore possibilistic approaches classify samples on the basis of the possibilistic version of the Bayes theorem.

A. Classification model of NPC

In classification problems, let us denote by $A = \{A_1, A_2, \cdots, A_n\}$ the set of attribute variables and $C = \{c_1, c_2, \cdots, c_d\}$ the set of classes in the training set. Given an observed vector $\{a_1, a_2, \cdots, a_n\}$ of X to be classified, possibilistic classifiers compute the posterior possibility $\pi(c_k|X)$ for each class c_k in C, and assign the instance X to the class c_k that maximizes the posterior possibility, ie

$$c = \arg \max_{c_k \in C} (\Pi(c_k|X)). \quad (6)$$

Use the possibilistic version of the Bayes rule in the quantitative setting:

$$\Pi(c_k|X) = \frac{\Pi(c_k, X)}{\Pi(X)}. \quad (7)$$

As a counterpart to bayesian classifier, NPCs estimate joint possibility of a set of attributes, i.e., $\Pi(c_k, X)$. Because the reasoning calculation is a NP-hard problem, algorithm is implemented according to the following two assumptions:

- Suppose that it works under the strong assumption of independence of attributes in the context of the class value. Its model is shown in [34].
- Assume that all attributes are equally important in assigning the class value for an unknown instance.

NPC is a simple form, once assuming the independence of attributes, Eq.(7) can be simplified:

$$\Pi(c_k|X) = \frac{\pi(c_k)\pi(a_1|c_k)\pi(a_2|c_k)\cdots\pi(a_n|c_k)}{\Pi(a_1 a_2 \cdots a_n)}, \quad (8)$$

where $\pi(c_k) = 1$ by convention. Each conditional possibility $\pi(a_i|c_k)$ in formula (8) can be estimated by using different methods for both categorical [20] and numerical attributes [21]. Eventually, Eq.(6) can be rewritten as:

$$c = \arg \max_{c_k \in C} \prod_{i=1}^{n} (\Pi(a_i|c_k)). \quad (9)$$

B. Conditional possibility

As mentioned above, $\pi(a_i|c_k)$ means the conditional possibility in the context of the class node. In this section, we mainly focus on the calculation of conditional possibility for categorical attributes proposed by Bakhta Haouari [20].

When classes are imprecise in the training set, they computed the conditional possibility by using geometric mean function as the following formula:

$$\pi(a_i|c_k) = mean_{Tr(a_i,c_k)} \pi(X_j|c_k), \quad (10)$$

where $Tr(a_i, c_k)$ means the number of training samples with the attribute value of a_i in the context of the class c_k, then $\pi(X_j|c_k)$ is the possibility degree of the sample X_j which belongs to the class c_k in $Tr(a_i, c_k)$.

Then, a simple classification example is followed to illustrate the computational process of the conditional possibility.

Example1. Table 1 shows a imprecise labeled training set, which is composed of fourteen samples characterized by four attributes. And these training samples belong to two classes (no, yes). Attributes are outlook, temperature, humidity and wind, respectively. π_1 denotes $\pi(X_j|Yes)$ and π_2 denotes $\pi(X_j|No)$. Then $\pi(X_1|Yes) = 0.2$ means that the possibility degree is 0.2 for instance X_1 which belongs to the Yes class.

Let us compute $\pi(rainy|Yes)$ by Eq.(10) as follow:

$$\pi(rainy|Yes) = \frac{(1 + 1 + 1 + 0.4 + 1 + 1 + 0)}{7} = 0.7714.$$

Once each conditional possibility is computed, we can predict the class value for an unknown labeled instance by Eq.(9).

IV. Feature weighting approaches

Two strong assumptions of NPC mentioned in section 1 do impact on its classification performance when it is violated in a learning task.

In practice, since the assumption of equally important attributes is rarely satisfied, several studies have tried to relax this assumption by assigning different weights to attribute variables. The underlying principle of weighted naïve possibilistic classification is that some attributes are more important than others in deciding the classification. This leads to the modification of Eq.(9) given as,

$$c = \arg \max_{c_k \in C} \prod_{i=1}^{n} (w_{ik} * \Pi(a_i|c_k)) \quad (11)$$

or

$$c = \arg\max_{c_k \in C} \prod_{i=1}^{n} \Pi(a_i|c_k)^{w_{ik}}, \tag{12}$$

where w_{ik} represents the weight of each attribute's conditional possibility.

Next, two types of attribute weighting approaches are proposed in this section to improve the NPC.

A. Attribute contribution-based feature weighting

NPCs mainly deal with imprecise data, which include the imprecise labels in the training set and the imprecise attributes both in the training and testing set. In our paper, we mainly focus on the imprecise label in the training set.

Since the NPC induces a classifier from the imprecise labeled data, each conditional possibility can be calculated from the training data by Eq.(10). It can be seen from the formula that the approach regards the possibilistic mean of the attribute value a_j given the class c_k as the conditional possibility $\pi(a_i|c_k)$. In other words, it only considers the proportion of possibility degree $\pi(X_j|c_k)$ to the number of this attribute value.

However, if the number of a attribute value becomes small, it will cause the classified result tending to the class to which this attribute value belongs. In our paper, we take the contribution degree of different attribute values for classification into account, reducing the influence of a small number of attribute values in classification task. Hence, we define the weight w_{jk} of each conditional possibility as

$$w_{jk} = \frac{\sum_{Tr(a_i,c_k)} \pi(X_j|c_k)}{\sum_{i=1}^{n} \pi(X_i|c_k)}, \tag{13}$$

where n denotes the total number of samples which belong to the class c_k.

Example2. For $\pi(rainy|Yes)$ in example 1, let us compute the weight and weighted conditional possibility by Eq.(13) and Eq.(11):

$w_{rainy,Yes} = (1 + 1 + 1 + 0.4 + 1 + 1 + 0)/(0.2 + 0.4 + 1 + 1 + 1 + 0.4 + 1 + 0.3 + 1 + 1 + 1 + 1 + 1 + 0) = 0.5242$,

$\pi'(rainy|Yes) = w_{rainy,Yes} * \pi(rainy|Yes) = 0.4043$.

Next let us compute $\pi(overcast|Yes)$ by Eq.(10):

$$\pi(over|Yes) = \frac{(1+1)}{2} = 1.$$

Following Eq.(13) and Eq.(11), the weight and weighted conditional possibility is obtained, respectively,

$w_{over,Yes} = (1 + 1)/(0.2 + 0.4 + 1 + 1 + 1 + 0.4 + 1 + 0.3 + 1 + 1 + 1 + 1 + 1 + 0) = 0.1942$,

$\pi'(over|Yes) = w_{over,Yes} * \pi(over|Yes) = 0.1942$.

We can see that $\pi(rainy|Yes) < \pi(over|Yes)$, but $\pi'(rainy|Yes) > \pi'(over|Yes)$. It means greater contribution for the attribute $rainy$.

Considering the influence of different values on the classification as a weight, not only reflects the influence degree of each attribute on the classification process, but also quantities the relationships between each attribute and class values. Therefore, we use the w_{jk} as the weighted coefficient is more reasonable and accurate, and it contributes to improve the classified accuracy.

TABLE I
A LABELED IMPRECISE TRAINING SET

Sample	Outlook	Temp	Humidity	Wind	π_1	π_2
X_1	sunny	hot	high	weak	0.2	1
X_2	sunny	hot	high	strong	0.4	1
X_3	rainy	hot	high	weak	1	0.7
X_4	rainy	mild	high	weak	1	0
X_5	rainy	cool	normal	weak	1	0.8
X_6	rainy	cool	normal	strong	0.4	1
X_7	rainy	cool	normal	strong	1	0.9
X_8	sunny	mild	high	weak	0.3	1
X_9	sunny	cool	normal	weak	1	0.3
X_{10}	rainy	mild	normal	weak	1	0
X_{11}	sunny	mild	normal	strong	1	0.2
X_{12}	overcast	mild	high	strong	1	0
X_{13}	overcast	hot	normal	weak	1	0.3
X_{14}	rainy	mild	high	strong	0	1

B. Non-specificity gain-based feature weighting

Actually, different attributes have different impacts on the classification tasks in most of data sets. To weaken the equal importance assumption, another feature weighting approach is proposed in this section. It applies a weighted coefficient based on the different important degree of each attribute to classify unknown instances, thus improving the classified accuracy of NPC.

In the paper, due to the presence of the imprecise data, we use specificity gain other than mutual information to measure the importance degree of each attribute to class variables.

Given a training set D, the non-specificity gain ratio $NSGR(a_i, C)$ is defined as follows:

$$NSGR(a_i, C) = \frac{NSG(a_i, C)}{U(a_i)}, \tag{14}$$

where C is the class variable, $NSG(a_i, C)$ is the non-specificity gain which assesses the amount of information of the attribute a_i to the classification system. $U(a_i)$ denotes the amount of information about this attribute itself. The specific calculation formula is shown as follows:

$$NSG(a_i, C) = U_C(\pi_{rep}) - U_{a_i}(\pi_{rep}), \tag{15}$$

where U_C is the non-specificity of the set of possibility distribution in a training set D. Similarly, U_{a_i} is the non-specificity of the set of conditional possibility distribution given the attribute a_i, which can be calculated by Eq.(5). In addition, the possibility distribution π_{rep} in $U_C(\pi_{rep})$ or $U_{a_i}(\pi_{rep})$ is defined the arithmetic mean of possibility distributions. One can see the specific calculation formula in [22].

Once the gain ratio $NSGR(a_i, C)$ of each attribute is acquired, we can calculate the sum of all attributes' information gain ratios and define the weight w_i of each attribute a_i $(i = 1, 2, \cdots, m)$ as

$$w_i = \frac{\sum_{i=1}^{m} NSGR(a_i, C)}{NSGR(a_i, C) \times m}. \tag{16}$$

After getting the weight value w_i of each attribute a_i $(i = 1, 2, \cdots, m)$ by Eq.(16), we apply these weights to Eq.(12) instead of Eq.(11) to improve the classification performance. Because, we note that $NSG(a_i, C) \in [-log2(n), log2(n)]$. Then the value w_i can be positive or negative. Note that each conditional possibility is between 0 and 1 in Eq.(12).

With the value of w_i increasing, the $\Pi(a_i|c_k)^{w_i}$ decreases. So it is more suitable to apply Eq.(12), in which we assign a smaller value w_i to a more important attribute. Its main principle is that a feature with higher gain ratio is acquired higher possibility degree.

C. Improve weighted naïve possibilistic classifier algorithms

Let W_1 denote the attribute contribution-based feature weight, and W_2 denote the non-specificity gain ratio-based weight. Three improve algorithms for NPC are proposed in this section.

Firstly, we only consider weight W_1 for classification, called attribute contribution-based weighted naïve possibilistic classifier (ACWNPC). The detailed learning process is described as Algorithm 1.

Algorithm 1 ACWNPC

Input: a training set D, a testing set d.
Output: the class value c of the testing set d.

1 For each attribute $a_i(i = 1, 2, \cdots, m)$ from D, calculate $\pi(a_i|c_k)$ using Eq.(10).
2 In each conditional possibility, calculate the attribute's contribution degree for class c_k as its weight w_{ik} using Eq.(13).
3 For the testing set d, calculate $w_{ik} * \pi(a_i|c_k)$ and predict its class value c using Eq.(11)
4 Return the class value c of d.

Then, we consider weight W_2 for classification, called Non-specificity gain-based weighted naïve possibilistic classifier (NSWNPC)(see Algorithm 2).

Algorithm 2 NSWNPC

Input: a training set D, a testing set d.
Output: the class value c of the testing set d.

1 For each attribute $a_i(i = 1, 2, \cdots, m)$ from D, calculate $\pi(a_i|c_k)$ using Eq.(10).
2 For each attribute $a_i(i = 1, 2, \cdots, m)$ from D, calculate $NSGR(a_i, C)$ using Eq.(14).
3 Let $NSGR(a_i, C) = 0$ if and only if $NSGR(a_i, C) <= p$, in which p is a parameter and the determination of p value depends on each data set.
4 Calculate the sum of all attributes' non-specificity gain ratios.
5 For each attribute $a_i(i = 1, 2, \cdots, m)$, calculate its weight w_i using Eq.(16).
6 For the testing set d, calculate $\pi(a_i|c_k)^{w_i}$ and predict its class value c using Eq.(12).
7 Return the class value c of d.

Finally, it regards the incorporation of W_1 and W_2 as the final weight denoted as following, namely, merged weighted naïve possibilistic classifier (MWNPC). The detailed learning process is described as Algorithm 3. Merge the Eq.(12) and Eq.(13) as following:

$$c = \arg\max_{c_k \in C} \prod_{i=1}^{n} (w_{ik} * \Pi(a_i|c_k))^{w_i}. \tag{17}$$

Algorithm 3 MWNPC

Input: a training set D, a testing set d.
Output: the class value c of the testing set d.

1 For each attribute $a_i(i = 1, 2, \cdots, m)$ from D, calculate $\pi(a_i|c_k)$ using Eq.(10).
2 In each conditional possibility, calculate the attribute's contribution degree for class c_k as its weight w_{ik} using Eq.(13).
3 For each attribute $a_i(i = 1, 2, \cdots, m)$ from D, calculate $NSGR(a_i, C)$ using Eq.(14).
4 Let $NSGR(a_i, C) = 0$ if and only if $NSGR(a_i, C) <= p$, in which p is a parameter and the determination of p value depends on each data set.
5 Calculate the sum of all attributes' non-specificity gain ratios.
6 For each attribute $a_i(i = 1, 2, \cdots, m)$, calculate its weight w'_i using Eq.(16).
7 Calculate the fusion value of w_{ik} and w'_i, and predict its class value c using Eq.(17) for the testing set d.
8 Return the class value c of d.

D. Algorithm complexity

We mainly concerned about the time complexity of our algorithm in this section. In the training phase, let m and x denote the number of training instances and class variables, respectively. Each instance is described by y attributes, and the number of each attribute values is denoted by $n_1, n_2, \cdots, n_y$. First, the time complexity of conditional possibilities is $x * (n_1 + n_2 + \cdots + n_y)$. Then, in order to acquire the weighting based on attribute contribution, algorithm also need to be implemented $x*(n_1+n_2+\cdots+n_y)$ times. For the weighting based on non-specificity gain, it need $x*(n_1+n_2+\cdots+n_y)+y$ times. Thus, the algorithm ACWNPC performs $2 * x * (n_1 + n_2 + \cdots + n_y)$ times, NSWNPC performs $2 * x * (n_1 + n_2 + \cdots + n_y) + y$ times and MWNPC needs $3*x*(n_1+n_2+\cdots+n_y)+y$ times. In the worst situation, both of their time complexity are $O(n^3)$.

Similarly, in the testing phase, let m' denote the number of instances in the testing data. Classification process needs to be done $m' * y * x$ times. Hence, in the worst situation, the time complexity is also $O(n^3)$. From the above analysis, the total time complexity is $O(n^3)$, which is equal to the NPC's. Hence, the advantage of our algorithms which do not increase the time complexity is highlighted.

V. Experiments and results

In the section, we evaluate the NPC, ACWNPC, NSWNPC and MWNPC respectively and compare them against other state-of-the-art classifiers to validate the classification performance.

A. Data sets

We run our experiments on 8 widely used classification benchmark databases from UCI [35], whose detail descriptions are shown in Table 2. Instances, attributes and classes denote respectively the total number of instances, the number of attributes and the number of classes.

TABLE II
DATE SETS USED IN THE EXPERIMENTS

Data set	Instances	Attributes	Classes
Ionosphere	351	34	2
Voting	435	16	2
Balance scale	625	4	3
Breast cancer	699	9	2
Vehicle	846	18	4
Segment	2310	18	7
Wave	5000	40	3
Nursery	12960	8	5

TABLE III
EXPERIMENTS RESULTS FOR NPC AND OUR CLASSIFIERS($L = 0\%$)

Data sets	NPC	ACWNPC	NSWNPC	MWNPC
Ionosphere	94.60(2.9)	94.31(1.8)	**95.17(2.6)**	94.02(2.8)
Voting	89.63(5.5)	90.10(4.8)	90.31(3.7)	**90.55(3.8)**
Balance scale	90.87(3.6)	91.84(3.9)	91.52(3.8)	**91.84(3.9)**
Breast cancer	95.13(3.4)	95.85(3.0)	95.56(3.0)	**95.99(2.9)**
Vehicle	64.43(4.5)	64.32(5.2)	**64.91(0.4)**	64.55(4.1)
Segment	93.07(2.2)	**93.33(2.0)**	92.99(2.1)	93.20(1.9)
Wave	**82.30(1.2)**	81.62(1.4)	80.90(1.5)	80.60(1.4)
Nursery	87.94(1.4)	**89.66(0.9)**	87.33(0.9)	87.42 (0.8)
Average	87.24 (3.1)	**87.62(2.8)**	87.33(2.2)	87.27(2.7)

B. Experimental setting

Let us point out that all the continuous data are processed with the discrete way in the preprocessing stage. Then, since the NPC mainly deals with the imprecise data, we mainly obtain imprecise labeled data in our experiments.

In many real-world problems, imprecise data may be met for instance. However, there are no such data sets in the repository of machine learning databases which could be used for testing classifiers. Note that, imprecise data in our paper refer to the uncertainty associated with the class in the training set.

To artificially introduce imprecise labeling, the following procedure is set up. We generate different levels of uncertainty training sets from 0 to 50%. For each instance chosen $L\%$ from training sets randomly, we have associated a possibility degree equal to 1 with the original class and another possibility degree which obtained in an uniform way. For the remaining databases, we have defined a completely consistent possibility distribution with the initial classes. In our experiments, we generate imprecise data sets which vary the levels of uncertainty labels L from 0 to 50.

For example, let an instance belong to one of three classes, i.e., c_1, c_2, c_3. In the uncertainty case, we assign an instance three possibility degrees $\pi(X_j|c_1) = 0.2, \pi(X_j|c_2) = 0.6$, and $\pi(X_j|c_3) = 1$, respectively. It means that the degree to which the instance belongs c_3 is 1.

We use the percentage of correct classification as the criteria of accuracy, that is

$$Accuracy = \sum_{i=1}^{N} \delta(c_i', c_i)/N,$$

where c_i' denotes the class value obtained by the classifier for instance i, and c_i is its true value. $\delta(c_i', c_i) = 1$ if $c_i' = c_i$ and 0 otherwise, and N is the total number of classified instances.

TABLE IV
EXPERIMENTS RESULTS FOR OUR CLASSIFIERS AND TRADITIONAL CLASSIFIERS

Data	NBC	SVM	TREE	ACWNPC	NSWNPC	MWNPC
1	90.59	92.30	89.17	94.31	95.17	94.02
2	90.11	96.09	96.32	90.10	90.31	90.55
3	70.72	72.00	69.60	91.84	91.52	91.84
4	96.99	96.42	94.99	95.85	95.56	95.99
5	62.64	73.52	71.98	64.32	64.91	64.55
6	91.51	93.07	93.32	93.33	92.99	93.20
7	79.96	81.96	72.78	81.62	80.90	80.60
8	90.32	93.07	97.05	89.66	87.33	87.42
9	84.11	87.30	85.90	**87.62**	87.33	87.27

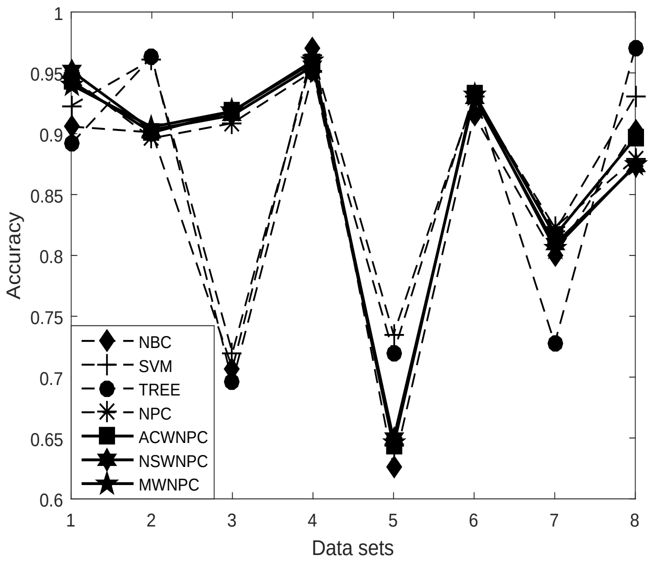


Fig. 1. Experimental comparison of NBC(diamond), SVM(plus), TREE(circle), NPC($*$), ACWNPC(square), NSWNPC(hexagram) and MWNPC(pentagram)

C. Experiments and conclusion

This section provides experimental results of proposed possibilistic classifiers for perfect and imperfect data. The experimental study is divided into two parts. First, we evaluate the ACWNPC, NSWNPC and MWNPC methods and compare our results to a classical NPC for perfect data. And on the basis of it, we compare other traditional methods with our methods. Traditional methods include NBC, SVM and TREE (J48). Second, we test the efficiency of the proposed classifiers to support uncertainty related to the classes.

Table 3 shows the detailed classification accuracy of NPC, ACWNPC, NSWNPC and MWNPC on each database obtained via ten runs of ten-fold cross-validation, respectively. And the values in brackets are the standard deviation. The averages are summarized at the bottom of the tables, which provide a gross indication of relative performance in addition to other statistics.

By comparing Table 3, we can see the results of our classifiers are better than the NPC in most data sets: (a)the ACWNPC, NSWNPC and MWNPC methods all obtain five wins and have competitive performance with NPC in other three data sets; (b)if compared with the standard deviation, we note that our methods perform better stability than the NPC in most of the data sets; (c)in average, the ACWNPC, NSWNPC and MWNPC methods obtain better results.

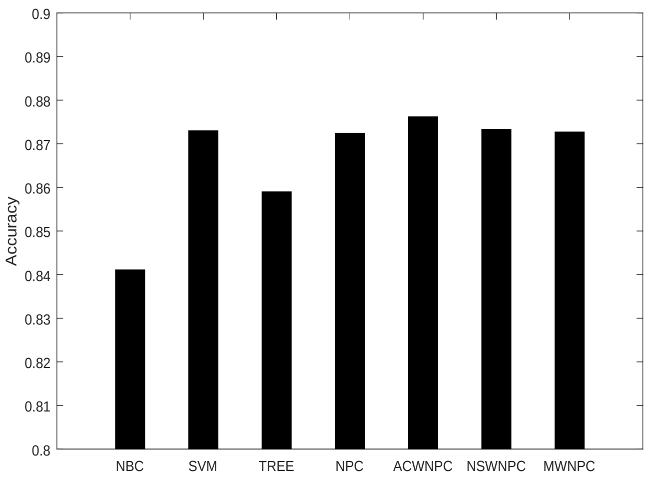


Fig. 2. Average accuracy results for the comparison between improved algorithms and traditional algorithms

TABLE V
EXPERIMENTS RESULTS FOR NPC AND OUR CLASSIFIERS($L = 10\%$)

Data sets	NPC	ACWNPC	NSWNPC	MWNPC
Ionosphere	88.30(6.9)	90.60(5.0)	91.16(4.9)	**91.73(4.5)**
Voting	88.95(5.6)	89.64(4.3)	88.74(4.5)	**89.89(3.4)**
Balance scale	88.31(4.2)	89.11(3.7)	88.47(4.0)	**89.43(4.1)**
Breast cancer	95.71(1.3)	96.71(1.9)	96.42(1.5)	**96.71(1.6)**
Vehicle	62.89(5.1)	62.89(6.1)	62.30(5.2)	**63.01(6.2)**
Segment	88.18(1.5)	**88.18(1.5)**	87.92(1.5)	87.92(1.5)
Wave	**81.88(2.7)**	81.06(2.3)	80.24(2.3)	79.52(2.2)
Nursery	88.48(1.0)	**89.54(0.8)**	85.93(1.1)	85.74(1.1)
Average	85.33(3.5)	**85.96(3.2)**	85.14(3.1)	85.49(3.0)

TABLE VI
EXPERIMENTS RESULTS FOR NPC AND OUR CLASSIFIERS($L = 20\%$)

Data sets	NPC	ACWNPC	NSWNPC	MWNPC
Ionosphere	87.75(4.2)	90.31(4.3)	90.60(5.0)	**91.16(4.3)**
Voting	89.20(4.2)	89.19(3.7)	89.64(4.3)	**91.01(4.2)**
Balance scale	81.78(6.1)	86.41(5.5)	80.98(7.3)	**87.52(3.3)**
Breast cancer	95.56(2.3)	97.14(1.7)	96.71(2.1)	**97.28(1.5)**
Vehicle	60.05(5.1)	60.87(6.0)	61.11(5.1)	**61.70(5.2)**
Segment	**86.62(2.6)**	86.41(2.9)	86.36(2.3)	86.36(2.3)
Wave	**81.04(2.5)**	80.02(2.4)	79.42(2.2)	78.90(2.0)
Nursery	77.33(2.3)	**84.98(0.8)**	81.66(1.8)	83.16(1.1)
Average	82.41(3.6)	84.41(3.4)	83.31(3.7	**84.63(2.9)**

Compared with the traditional NBC, SVM and TREE(J48) methods, the experimental results are given in Table 4 (from 1 to 8, they represent data sets: Ionosphere, Voting, Balance scale, Breast cancer, Vehicle, Segment, Wave and Nursery in turn. And 9 represents the average). The results show that: (a)the ACWNPC method outperforms the NBC in five data sets and have competitive performance with NPC in other three data sets; NSWNPC and MWNPC both obtain six wins vs. two losses; (b)our proposed methods obtain three wins and have competitive performance with SVM in other data sets; (c)improved algorithms significatively outperform the TREE(J48) in four data sets and significatively beaten by it in three data sets. And in the Segment, our classifiers and the TREE(J48) have competitive performance; (d)in average, our improved algorithms are superior to the traditional algorithms.

TABLE VII
EXPERIMENTS RESULTS FOR NPC AND OUR CLASSIFIERS($L = 30\%$)

Data sets	NPC	ACWNPC	NSWNPC	MWNPC
Ionosphere	87.78(8.5)	90.32(6.4)	90.05(7.1)	**91.17(5.9)**
Voting	88.53(4.7)	89.19(3.1)	89.67(3.7)	**89.89(5.0)**
Balance scale	79.05(5.9)	84.17(5.4)	79.37(6.8)	**84.65(4.9)**
Breast cancer	95.28(2.1)	97.14(1.1)	96.28(1.8)	**97.14(1.3)**
Vehicle	61.72(7.0)	62.31(5.6)	61.72(6.0)	**62.90(4.8)**
Segment	84.76(4.0)	**84.81(3.6)**	84.33(4.1)	84.64(3.9)
Wave	**79.30(2.7)**	78.98(2.6)	78.40(2.4)	78.20(2.2)
Nursery	79.29(2.3)	**85.81(1.4)**	81.06(0.9)	84.38(1.4)
Average	81.96(4.6)	84.09(3.6)	82.61(4.1)	**84.12(3.6)**

TABLE VIII
EXPERIMENTS RESULTS FOR NPC AND OUR CLASSIFIERS($L = 40\%$)

Data sets	NPC	ACWNPC	NSWNPC	MWNPC
Ionosphere	87.19(6.0)	90.32(3.8)	89.46(5.2)	**91.45(3.5)**
Voting	88.28(5.8)	89.20(4.4)	89.19(4.5)	**89.55(5.1)**
Balance scale	76.43(7.4)	83.33(5.1)	77.40(7.1)	**84.60(7.4)**
Breast cancer	95.28(2.5)	96.85(2.2)	95.70(2.8)	**97.28(1.7)**
Vehicle	60.77(5.4)	61.95(5.2)	61.83(4.6)	**62.66(4.2)**
Segment	84.03(3.7)	83.98(3.5)	84.07(3.5)	**84.11(3.4)**
Wave	**79.52(2.0)**	79.00(2.0)	78.36(2.0)	78.18(2.0)
Nursery	73.88(2.4)	82.04(1.9)	79.10(2.2)	**82.27(1.1)**
Average	80.67(4.4)	83.33(3.5)	81.88(3.9)	**83.76(3.5)**

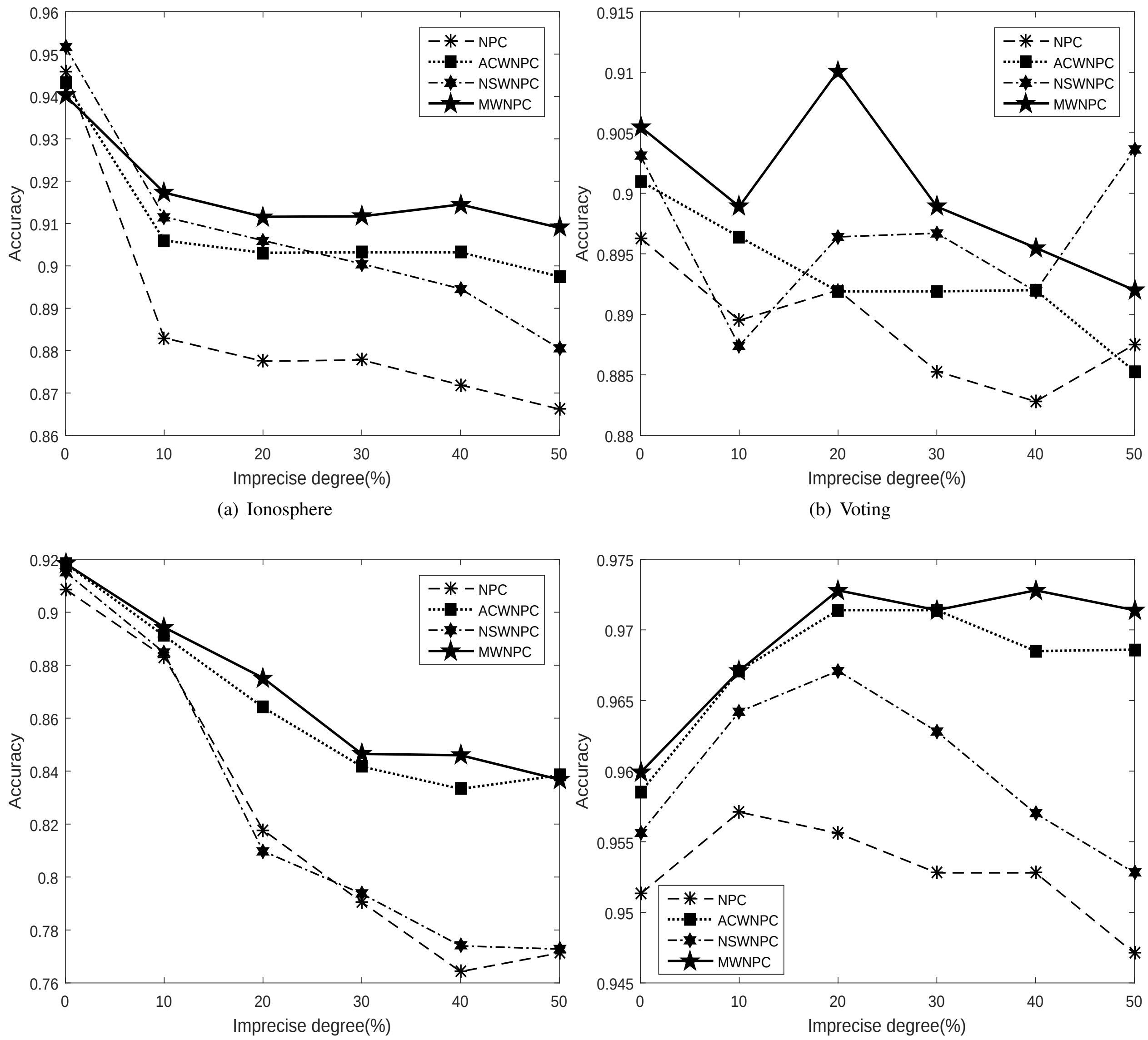
TABLE IX
EXPERIMENTS RESULTS FOR NPC AND OUR CLASSIFIERS($L = 50\%$)

Data sets	NPC	ACWNPC	NSWNPC	MWNPC
Ionosphere	86.62(5.3)	89.75(5.2)	88.05(4.5)	**90.90(5.1)**
Voting	88.75(3.4)	88.53(3.8)	90.36(4.9)	**89.20(5.1)**
Balance scale	77.13(7.9)	**83.84(6.3)**	77.28(8.1)	83.68(6.9)
Breast cancer	94.71(2.7)	96.86(2.1)	95.28(2.7)	**97.14(2.4)**
Vehicle	60.63(3.7)	60.87(3.9)	61.58(4.5)	**62.56(4.3)**
Segment	83.46(3.4)	83.51(3.2)	83.51(3.4)	**83.51(3.4)**
Wave	**79.40(2.3)**	78.84(1.8)	78.46(1.9)	77.90(1.8)
Nursery	66.37(1.2)	75.52(1.0)	75.64(1.5)	**79.77(1.2)**
Average	79.63(3.7)	82.31(3.4)	81.27(3.9)	**83.08(3.7)**

1) Experiments for the perfect data: In Fig.1, the improved classifiers are represented by the solid lines and the traditional classifiers are represented by the dotted lines. And we can see that improved classifiers obtained better results than the traditional classifiers because the solid lines are above the dotted lines on most data sets. In order to make the results intuitive and clear, the average accuracy results for the comparison between improved algorithms and traditional algorithms are displayed in Fig.2.

2) Experiments for the imperfect data: We preform the experiment six times for each method, since it generates imprecise data which varies the levels of uncertainty labels L from 0 to 50 for each database. Thus, Table 5-9 report the results after varying the training sets' level of uncertainty L from 10 to 50 for each database, respectively. Note that there is no comparison with the traditional classification method mentioned above, such as NBC, TREE and SVM. It is because these methods are only used in certain environments while the NPC approach and improved weighted algorithms deal with both certain and uncertain environments.

In order to make the results intuitive, the accuracy results for each data are displayed in Figs.3 and Figs.4. Figs.3(a) shows four lines of the accuracy for different algorithms over imprecise degree and is the same with other figures.



(a) Ionosphere

(b) Voting

(c) Balance scale

(d) Breast cancer

Fig. 3. The classified accuracy on different datasets from the $L = 0\%$ to $L = 50\%$

As we can see from Figs.3 and Figs.4, almost all proposed algorithms in our paper, i.e., ACWNPC, NSWNPC and MWNPC, have better performance than the NPC: (a)the lines denoted MWNPC are above the other lines in six data sets; (b)these lines denoted ACWNPC, NSWNPC and MWNPC are above the NPC line in seven data sets; (c)with the increase of imprecise degree, the accuracy increases more remarkably than other imprecise groups.

Fig.5 mainly shows that the variations with respect to standard deviation over algorithms on the average values of all databases. It reflects that the improved algorithms perform better than the traditional NPC not only in accuracy but also in stability.

We use the Wilcoxon Matched-Pairs Signed-Ranks Test [36] to compare four classifiers. It is a non-parametric alternative to the paired t-test which has the advantages of not having to assume the data distribution. And it enables us to compare two classifiers over multiple data sets. Comparison results given in Table 10 show that the ACWNPC and MWNPC have competitive performance and they are always significantly better than the NPC ($p - value < 0.05$).

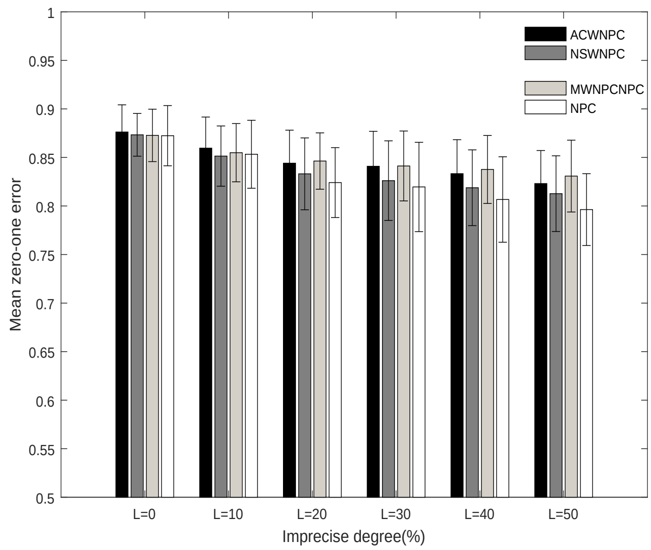
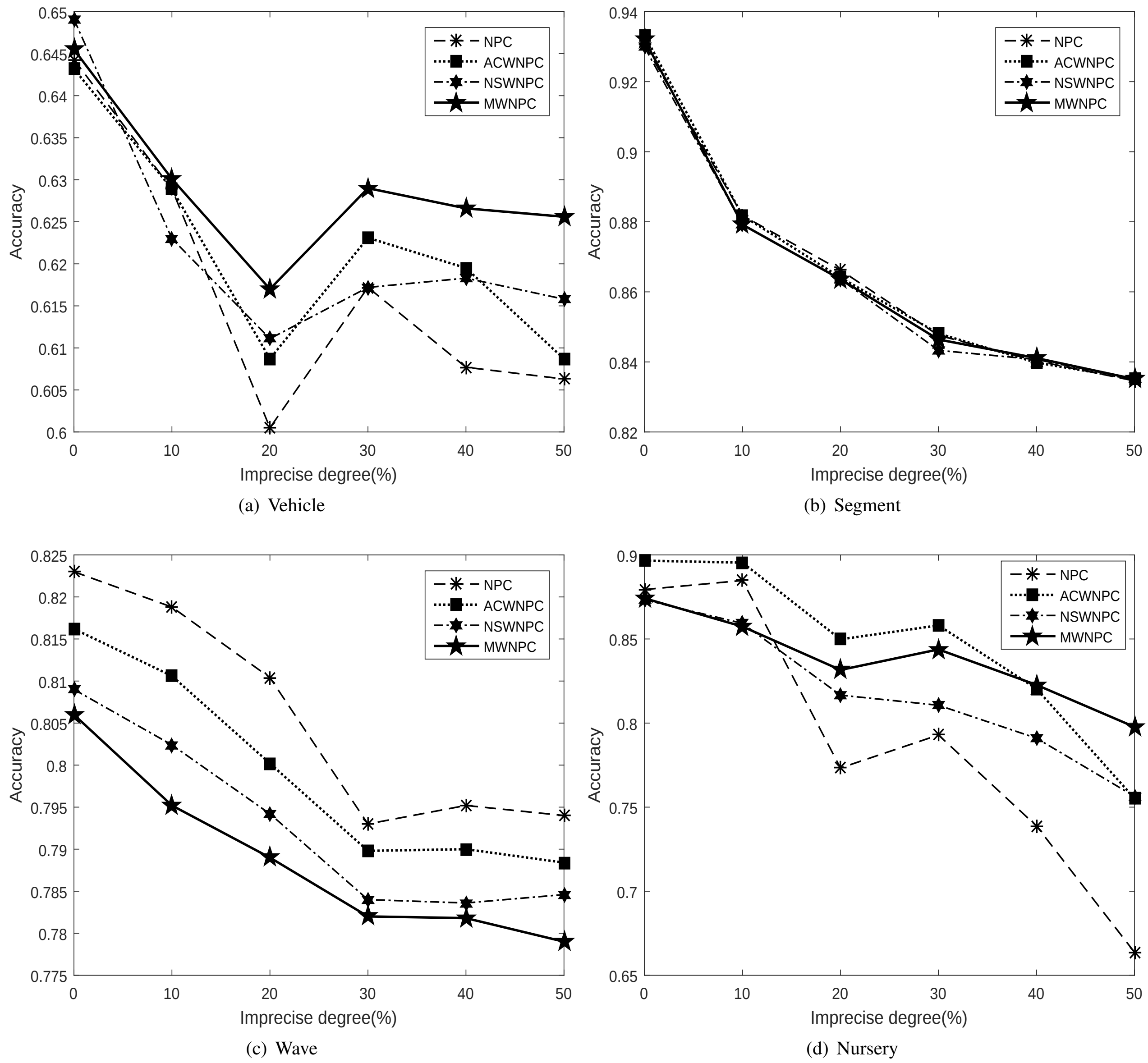


Fig. 5. Average accuracy and standard deviation of NPC and our classifiers



(a) Vehicle (b) Segment

(c) Wave (d) Nursery

Fig. 4. The classified accuracy on different datasets from the $L = 0\%$ to $L = 50\%$

TABLE X
RESULTS FOR THE WILCOXON MATCHED-PAIRS SIGNED-RANKS TEST

ACWNPC vs NPC	ACWNPC vs NSWNPC	ACWNPC vs MWNPC
$p \leq 0.028$	$p \leq 0.028$	$p \leq 0.600$
NSWNPC vs NPC	MWNPC vs NSWNPC	MWNPC vs NPC
$p \leq 0.075$	$p \leq 0.046$	$p \leq 0.028$

From these experimental results, we can see that our proposed feature weighting approaches rarely degrade the quality of original naïve possibilistic classifiers and, in many cases, improve them remarkably.

Now, we summarize the highlights as follows:

- Our algorithms improve the accuracy and do not increase its computation complexity, as is shown in section 4.
- As the level of uncertainty increases, classification accuracies of the NPC, ACWNPC, NSWNPC and MWNPC decrease in most of the data sets. It can be explained by the fact that the higher the level of uncertainty is, the less information it leads.
- Almost all proposed algorithms, i.e., ACWNPC, N-SWNPC and MWNPC, have better performance than the NPC, NBC, SVM and TREE in most of the data sets. In the other cases, our classifiers have competitive performance with the SVM and TREE.
- The difference between the merging weighted algorithm MWNPC and NPC is statistically significant in most of the data sets.
- Even, with the increase of imprecise degree, the accuracy increases more remarkably than other imprecise groups.
- Comparing with the standard deviation, we note that improved algorithms perform better in stability than the NPC for a majority of data sets.

VI. CONCLUSION

The main purpose of the paper is to overcome the limitation of the strong assumptions for the traditional NPC. We study two feature weighting approaches to improve standard naïve possibilistic classifiers. In the paper, we adapt two simple and effective feature weighting approaches to the NPC. One is the attribute contribution-based feature weighting approach, and the other is the non-specificity gain ratio-based feature weighting approach. These algorithms not only consider the effect on classification by the different attribute values, but also consider the significance of each attribute to the classification task.

In the paper, three improved algorithms are proposed, namely, ACWNPC, NSWNPC and MWNPC. The experimental study is divided into two parts. First, we evaluate the ACWNPC, NSWNPC and MWNPC methods and compare our results to a classical NPC for perfect data. And on the basis of it, we compare other traditional methods with our methods. Second, we test the efficiency of the proposed classifiers to support uncertainty related to the classes.

Experimental results show that our proposed feature weighting approaches rarely degrade the quality of original NPC and, in many cases, improve them remarkably. First, proposed methods improve the accuracy and do not increase its computation complexity. Second, the ACWNPC, NSWNPC and MWNPC have better performance than the NPC and traditional classifiers in most of the data sets. Third, compared with the standard deviation, improved algorithms perform better in stability than the NPC for a majority of data sets. Finally, the experimental results on a large number of data validate their efficiency in terms of classified accuracy.

It is well known that the feature selection can be regarded as a special feature weighting method, and studying feature selection is significative in classification tasks. In the future, we will try to do more works following the idea of feature selection for imprecise data.

ACKNOWLEDGMENTS

The authors thank the editors and the anonymous reviewers for helpful comments and suggestions. The research was supported by the National Natural Science Foundation of China (Grant No. 61573266).

REFERENCES

[1] F. Gullo, "From patterns in data to knowledge discovery: What data mining can do," *Physics Procedia*, vol. 62, pp. 18–22, 2015.

[2] L. V. N. Prasad and M. M. Naidu, "Cc-sliq: Performance enhancement with 2k split points in sliq decision tree algorithm," IAENG International Journal of Computer Science, vol. 41, no. 3, pp 163–173, 2014.

[3] L. Demidova, E. Nikulchev, and Y. Sokolova, "The svm classifier based on the modified particle swarm optimization," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 2, pp. 16–24, 2016.

[4] A. Hambarde, M. F. Hashmi, and A. Keskar, "Robust image authentication based on hmm and svm classifiers," Engineering Letters, vol. 22, no. 4, pp 183–193, 2014.

[5] M. Mohamadian, H. Afarideh, and F. Babapour, "New 2d matrix-based neural network for image processing applications," IAENG International Journal of Computer Science, vol. 42, no. 3, pp 265–274, 2015.

[6] J. Schmidhuber, "Deep learning in neural networks: an overview," *Neural Networks*, vol. 61, pp. 85–117, 2014.

[7] Y. Yang, Y. Li, W. Wang, and W. Yang, "Fast and efficient evolutionary algorithms based on bayesian networks," IAENG International Journal of Computer Science, vol. 42, no. 3, pp 174–182, 2015.

[8] C. F. Wang and K. Liu, "Learning bayesian network classifier based on artificial fish swarm algorithm," IAENG International Journal of Computer Science, vol. 42, no. 4, pp 355–360, 2015.

[9] N. Friedman, G. Dan, and M. Goldszmidt, "Bayesian network classifiers," *Machine Learning*, vol. 29, no. 2-3, pp. 131–163, 1997.

[10] D. M. Diab and K. M. E. Hindi, "Using differential evolution for fine tuning naive bayesian classifiers and its application for text classification," *Applied Soft Computing*, vol. 54, pp. 183–199, 2017.

[11] D. Dubois and H. Prade, "Fuzzy set and possibility theory-based methods in artificial intelligence," *Artificial Intelligence*, vol. 148, no. 1-2, pp. 1–9, 2003.

[12] D. Yan, Y. Yang, and B. Li, "An improved fuzzy classifier for imbalanced data," *Journal of Intelligent and Fuzzy Systems*, vol. 32, no. 3, pp. 2315–2325, 2017.

[13] R. S. Kumar and G. T. Arasu, "Rough set theory and fuzzy logic based warehousing of heterogeneous clinical databases," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 25, no. 3, pp. 385–408, 2017.

[14] Q. Duan, Y. L. Yang, and Y. Li, "Rough k-modes clustering algorithm based on entropy," IAENG International Journal of Computer Science, vol. 44, no. 1, pp 13–18, 2017.

[15] W. Quan, J. Wang, L. Lei, and M. Gao, "Target recognition method based on multi-class svm and evidence theory," *International Conference on Emerging Internetworking, Data and Web Technologies 2017*, pp. 262–272.

[16] Q. Gu and Z. Xuan, "A new approach for ranking fuzzy numbers based on possibility theory," *Journal of Computational and Applied Mathematics*, vol. 309, pp. 674–682, 2017.

[17] L. A. Zadeh, "Fuzzy sets as a basis for a theory of possibility," *Fuzzy Sets and Systems*, vol. 1, no. 1, pp. 3–28, 1978.

[18] D. Dubois and H. Prade, "Possibility theory and its applications: a retrospective and prospective view," *The IEEE International Conference on Fuzzy Systems 2003*, pp. 5–11.

[19] G. Mauris, "A review of relationships between possibility and probability representations of uncertainty in measurement," *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 3, pp. 622–632, 2013.

[20] B. Haouari, N. B. Amor, Z. Elouedi, and K. Mellouli, "Naive possibilistic network classifiers," *Fuzzy Sets and Systems*, vol. 160, no. 22, pp. 3224–3238, 2009.

[21] M. Bounhas, M. G. Hamed, H. Prade, M. Serrurier, and K. Mellouli, "Naive possibilistic classifiers for imprecise or uncertain numerical data," *Fuzzy Sets and Systems*, vol. 239, no. 1, pp. 137–156, 2014.

[22] I. Jenhani, N. B. Amor, and Z. Elouedi, "Decision trees as possibilistic classifiers," *International Journal of Approximate Reasoning*, vol. 48, no. 3, pp. 784–807, 2008.

[23] K. Baati, T. M. Hamdani, A. M. Alimi, and A. Abraham, "A new classifier for categorical data based on a possibilistic estimation and a novel generalized minimum-based algorithm," *Journal of Intelligent and Fuzzy Systems*, vol. 33, no. 3, pp. 1723–1731, 2017.

[24] J. Zhao, J. Liu, Y. Sun, and Z. Sun, "Tree augmented naive possibilistic network classifier," *Eighth International Conference on Fuzzy Systems and Knowledge Discovery 2011*, pp. 1065–1069.

[25] D. Dubois and H. Prade, "Possibility theory and formal concept analysis in information systems," *Joint 2009 International Fuzzy Systems Association World Congress and 2009 European Society of Fuzzy Logic and Technology Conference 2009*, pp. 1021–1026.

[26] D. Dubois, "Possibility theory and statistical reasoning," *Computational Statistics and Data Analysis*, vol. 51, no. 1, pp. 47–69, 2006.

[27] A. G. Bronevich, "The logical inferences in possibility theory based on the concept of upper (lower) probability," *The IEEE International Conference on Fuzzy Systems 2001*, pp. 192–194.

[28] D. Dubois and H. Prade, "Bayesian conditioning in possibility theory," *Fuzzy Sets and Systems*, vol. 92, no. 2, pp. 223–240, 1997.

[29] K. Boutouhami, S. Benferhat, F. Khellaf, and F. Nouioua, "Uncertain lightweight ontologies in a product-based possibility theory framework," *International Journal of Approximate Reasoning*, vol. 88, pp. 237–258, 2017.

[30] G. Coletti and R. Scozzafava, "Conditional probability, fuzzy sets, and possibility: a unifying view," *Fuzzy Sets and Systems*, vol. 144, no. 1, pp. 227–249, 2004.

[31] M. Higashi and G. J. Klir, "Measures of uncertainty and information based on possibility distributions," *International Journal of General Systems*, vol. 9, no. 1, pp. 43–58, 2007.

[32] G. J. Klir and M. J. Wierman, *Uncertainty-Based Information*. Physica-Verlag HD, 1999.

[33] M. Caetano, "Evaluation of soft possibilistic classifications with non-specificity uncertainty measures," *International Journal of Remote Sensing*, vol. 31, no. 19, pp. 5199–5219, 2010.

[34] C. Borgelt and J. Gebhardt, "A naive bayes style possibilistic classifier," *Proceedings of 7th European Congress on Intelligent Techniques and Soft Computing (EUFIT'99)*, 1999.

[35] "Ucivine machine learning repository," http://archive.ics.uci.edu/ml/.

[36] J. Demsar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol. 7, no. 1, pp. 1–30, 2006.

Jing Bai was born in Lvliang, Shanxi Province, China in 1991. She received her B. S. degree in the Department of Mathematics and Applied Mathematics from Shanxi Normal University in 2014. Now she is a master of mathematics and statistics at Xidian University. She is major in Probabilistic graphical models, data analysis and its application.

Youlong Yang received his B. S. and M. S. in Mathematics from Shaanxi Normal University, Xian, China in 1990 and 1993 respectively, and Ph.D. in System Engineering from Northwester Polytechnical University, Xian,China in 2003. Since 2004, he has been with the faculty at Xidian University, Xi'an, China. His research interests include Machine learning, Statistical data analysis and Probabilistic graphical models.

Jianming Xie was born in 1990. He received his B. S. degree in the Department of information and computing science from Shihezi University in 2015. And he began work for a master at Xidian University in 2015. He is interested in Data Analysis and Machine Learning.