

## Research Article

# A Fast Imbalanced Binary Classification Approach to NLOS Identification in UWB Positioning

**Bo Song** , **Sheng-Lin Li**, **Mian Tan**, and **Qing-Hui Ren**

*Department of Military Logistics, Army Logistics University of PLA, Chongqing 401311, China*

Correspondence should be addressed to Bo Song; 545641621@qq.com

Received 26 September 2018; Accepted 1 November 2018; Published 2 December 2018

Academic Editor: Xiangyu Meng

Copyright © 2018 Bo Song et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Non-line-of-sight (NLOS) propagation is an important factor affecting the positioning accuracy of ultra-wide band (UWB). In order to mitigate the NLOS ranging error caused by various obstacles in UWB ranging process, some scholars have applied machine learning methods such as support vector machine and support vector data description to the identification NLOS signals for mitigation NLOS error in recent years. Therefore, the identification of NLOS signals is of great significance in UWB positioning. The traditional machine learning method is based on the assumption that the number of samples of the line-of-sight (LOS) and NLOS signals are balanced. However, in reality, the number of LOS signals in UWB positioning is much larger than the NLOS signals. So the samples are characterized by class-imbalance. In response to this fact, we applied a fast imbalanced binary classification method based on moments (MIBC) to identify NLOS signals. The method uses the mean and covariance of the two first moments of the LOS signal samples to represent its probability distribution and then uses the probability distribution and all a small amount of NLOS signal samples to establish a model. This method does not depend on the number of LOS signals and is suitable for dealing with the problem of classification of the imbalance between the number of LOS and NLOS signals. Numerical simulations also verify that the method has better performance than LS-SVM and SVDD.

## 1. Introduction

UWB technology has the characteristics of extremely narrow pulse and extremely wide bandwidth, which satisfies the requirements of positioning accuracy in indoor complex environment. However, these characteristics are also a double-edged sword, which leads to many technical problems in the realization of technology [1–5]. Such as signal reception, multiuser interference, multipath effects, and NLOS propagation. The problem of NLOS propagation especially is crucial to the influence of high-resolution positioning systems [6, 7]. Firstly, direct signals will be mistakenly detected due to multipath effects caused by NLOS. Secondly, the propagation of UWB signals in the medium will cause additional delay, resulting in TOA estimate values being larger than the real values, that is, positively biased range estimates. It is worth noting that NLOS conditions often occur in many real environments, such as closed rooms, dense urban buildings, and complex environments under the canopy. UWB positioning system is to work in these

environments where GPS or cellular network signals cannot reach. Therefore, it is quite necessary to identify the NLOS signals so as to mitigate the NLOS error of the positioning system [8–10].

The purpose of identifying NLOS is to detect the existence of NLOS conditions between the transmitter and receiver and ultimately mitigate NLOS errors. Therefore, the traditional identification research on NLOS focuses on analyzing the statistical characteristics of the signals in UWB positioning. There are mainly two types of methods. The first type is to analyze the statistical characteristics of TOA estimate results. Literature [11] identifies NLOS by the variance of TOA estimate values and judges NLOS when the variance exceeds the set threshold. An approach combines the error distribution of RSS estimates results with the error distribution of TOA estimates results for NLOS identification and achieves the identification effect by adding identification information [12]. However, the shortcomings of this kind of methods are also very obvious. The statistical information of the obtained results needs to be measured many times, which wastes

more time. The randomness of the estimate error is relatively large at the same time. The second type of method is to directly analyze the statistical characteristics of the received signals [8–10], that is, the channel characteristics. Typical characteristics include signal energy, pulse rise time, kurtosis, RMS delay spread, and mean excess delay (MED). Under NLOS and LOS conditions, the statistical parameters of the above characteristics are different, so we can determine which type of sight distance conditions the above characteristics belong to by set thresholds. At the same time, it is also possible to combine all characteristics and give corresponding weights to classify the received signals.

What is worth our attention is that literature [9] uses a nonparametric identification method and only relies on UWB waveform measurement to realize the identification of NLOS signals. In the case where both LOS and NLOS exist, the nonparametric support vector machine is used for NLOS identification. This approach to identify NLOS can be realized without statistical characteristics of waveforms. In the literature [10] published later they proposed two nonparametric regression techniques which only estimated the NLOS error based on the waveform of the received signal and the estimated distance. The traditional machine learning technologies obtain high prediction accuracy on the assumption that the numbers of samples in different class are basically the same. The traditional techniques tend to ignore the small class data, because even if all test samples are mistakenly identified as large class, they can obtain a high recognition rate. These small class data are very few but more important than large class data in some cases and the result caused by false recognition is more serious. In the actual environment, most of the signals we get are LOS signals, while the NLOS signals are very few. So our goal is to identify the small amount of NLOS signals which are very important for localization in NLOS environment. This is a typical class-imbalance classification problem. Traditional machine learning methods such as LS-SVM will not effectively solve the problem of class-imbalance classification. The recent literature [8] uses the method of support vector data description (SVDD) to transform the LOS recognition problem into a one-class classification problem. The method establishes a spherically shaped boundary, with center  $u$  and radius  $r$ , which contains as many LOS signals as possible while keeping as much minimum volume as possible to realize the classification of the NLOS and LOS signals. On the one hand, the computational complexity of the algorithm is still high although this method saves the time to train NLOS signals samples compared to LS-SVM. On the other hand, the method also classifies other interference signals as NLOS signals, which is not appropriate, because only NLOS signals can be used for ranging and positioning in complete NLOS environment.

From the above analysis, we can see that the problem of NLOS identification is actually a classification problem of class-imbalance. Therefore, class-imbalance learning is an effective way to solve NLOS identification in UWB positioning. This paper applies a new approach to identify NLOS signals in UWB positioning system, which is called moment-based imbalanced binary classification (MIBC). Compared

with LS-SVM method, we think that it is more practical to adopt the classification method of class-imbalance, and our method can achieve better classification performance. Moreover, compared with SVDD, the MIBC reduces the computational burden further, but it can still maintain good classification performance while avoiding mistaking the interference signals as NLOS signals. To the best of authors' knowledge, this is the first time that NLOS signals identification is considered as a class-imbalance classification problem. We use the datasets used in literature [9] to verify the algorithm and compare it with LS-SVM method in literature [9] and SVDD method used in literature [8]. Experimental results show that when the number of LOS and NLOS signals is seriously unbalanced, the performance of MIBC method is better than LS-SVM and SVDD method.

The rest of this article is arranged as follows. Section 2 introduces the statement of the problem. The proposed technique for NLOS identification is described in Section 3. In the 4th part, we carried out simulation verification based on the measured data sets from the indoor positioning competition organized by MIT. Finally, see Section 5 for the conclusion.

## 2. Problem Statement

A typical UWB positioning system mainly consists of two types of nodes: one is the anchor node, and their location information is known; the other is the terminal equipment whose location information is unknown. We need to locate it through anchor nodes at known specific locations around it. In order to describe the problem conveniently, we assume that there are  $M$  terminal devices with unknown location information and  $S$  anchor nodes participating in location and with known location information. We set  $P_i = [x_i \ y_i \ z_i]$ ,  $i = 1, 2, \dots, M$  to indicate the location of terminal device  $i$  and  $S_j = [x_j \ y_j \ z_j]$ ,  $j = 1, 2, \dots, S$  to present the location of anchor node  $j$ .

Under the above assumptions, we can get the ranging error model between the device and anchor node:

$$r_{ij} = \begin{cases} \|P_i - S_j\| + e_{ij} + v_{ij} & j = 1, \dots, L \\ \|P_i - S_j\| + v_{ij} & j = L + 1, \dots, S \end{cases} \quad (1)$$

Between the device and the anchor node,  $\|\bullet\|$  Indicates the actual distance,  $e_{ij}$  indicates the NLOS error, and  $v_{ij}$  indicates the measurement error. What we should pay attention to is the number of paths with NLOS errors between the equipment and all surrounding anchor nodes and we use  $L$  to express it. Where  $L \in S$ , in particular, when  $L=0$ , it indicates when there is no positively biased range estimates caused by NLOS error between the equipment and all anchor nodes,  $r_{ij} = \|P_i - S_j\| + v_{ij}$ ,  $j = 1, \dots, S$ . When  $L = S$ , this indicates that all paths contain the positive biases caused by NLOS errors,  $r_{ij} = \|P_i - S_j\| + e_{ij} + v_{ij}$ ,  $j = 1, \dots, S$ .

The basic idea of this paper has two aspects. First is that the waveform features of the received signal are different under LOS and NLOS conditions. These features are extracted and used to identify whether the ranging

error is NLOS error. The typical waveform features in UWB positioning have been given in literature [9]: energy ( $\epsilon_r$ ) of the received signal; the maximum amplitude (max) of the received signal; rise time (rise time); mean excess delay (MED); root mean square delay spread (RMS-DS); kurtosis. Secondly, considering that the number of LOS signals in the actual environment is much larger than that of NLOS signals, it is a typical problem of imbalanced binary classification. The imbalanced binary classification method is more in line with the actual situation than the LS-SVM classification of sample equalization. Compared with SVDD, our method can not only reduce the misrecognition rate of very important NLOS signals and achieve better classification performance, but also has lower complexity.

### 3. Imbalanced Binary Classification for NLOS Identification

In this section, we will mathematically demonstrate the correctness and accuracy of our proposed method and describe how it can be used to identify NLOS signals using the characteristics of the received signal waveform. These features will be the input to the classifier using the new method. First, we introduce the six signal waveform features used, and then deduced the mathematical formula of the new classification method in detail. The theory proves the validity and accuracy of the method.

**3.1. Feature Selection.** As mentioned in [8–10], we have chosen the following signal waveform characteristics: the energy attenuation of NLOS signals is much more serious than that of LOS signals due to the shielding and blocking of obstacles, resulting in significant differences in amplitude and energy of the two kinds of signals. At the same time, the kurtosis also reflects the steepness of the distribution in the datasets. The kurtosis value of the LOS signals is larger but the kurtosis value of the NLOS signals is smaller. The excess delay and root mean square delay reflect the delay characteristics of multipath components; generally speaking, the average excess delay of NLOS signals is larger than LOS signal and RMS is more serious than LOS signals. Under the LOS condition, the rise time of the received signal is shorter, while, under the NLOS condition, the rise time is longer. Table 1 lists the waveform signal features we used, and the values of the parameters in the formula are the same as those in [8–10].

**3.2. Classification: Mathematical Framework.** In most classification tasks, the numbers of data under each class are basically impossible to be completely equal, but a little difference will not have any impact or problem. However, when the numbers of data under each class are seriously imbalanced, the traditional classification method will greatly fail to meet the classification requirements because of data imbalance [13]. Aiming at the NLOS signals identification of UWB positioning system is a typical class-imbalance classification problem, we adopt a new formula to solve the problem and achieve the purpose of NLOS identification. In

general, the number of LOS signals is much larger than that of NLOS signals. So, a large number of LOS signals samples need to be processed before the model is built. We use the two first moments of LOS signal samples, that is, the mean and covariance, to represent their probability distribution, and use this probability distribution and a small amount of all NLOS signal samples to build the model.

We use  $(x_i)_{i \in \{1, \dots, n\}}$  to represent a set of  $n$  NLOS of training samples and let  $\bar{x}$  and  $\Sigma$  represent the mean and covariance of the probability distribution of the LOS signal samples. It is worth noting that, in the following mathematical derivation, the covariance matrix  $\Sigma$  is assumed to be positive definite. Our goal is to find a hyperplane  $H(w, b)$ . All NLOS signal samples can be correctly classified:

$$w^T x_i - b \geq 0, \quad \forall i \in \{1, \dots, n\} \quad (2)$$

At the same time, it can maximize the probability of correctly classifying LOS signal samples with respect to the distribution with mean  $\bar{x}$  and covariance  $\Sigma$ :

$$\max_{\alpha, w \neq 0, b} \alpha \inf_{x \sim (\bar{x}, \Sigma)} \Pr \{w^T x - b \leq 0\} \geq \alpha \quad (3)$$

where  $x \sim (\bar{x}, \Sigma)$  refers to the class with a probability distribution of mean  $\bar{x}$ , covariance  $\Sigma$ .

By synthesizing formulas (2) and (3), our aim is to find an optimal hyperplane that can correctly classify all NLOS signal samples while also correctly classifying the LOS signal samples to the greatest extent possible. From Lemma 1 in literature [14] we can know that the constraint condition expressed by formula (3) has a geometric representation, which can be converted equivalently.

**Lemma 1.** *Using the notation above, with  $\bar{x} \in \mathbb{R}^d$ ,  $\Sigma \in \mathbb{R}^{d \times d}$  a positive definite matrix,  $w \neq 0, b$  given, such that  $w \in \mathbb{R}^d, b \in \mathbb{R}$  and  $\alpha \in [0, 1]$ . Then, the condition*

$$\inf_{x \sim (\bar{x}, \Sigma)} \Pr \{w^T x - b \leq 0\} \geq \alpha, \quad (4)$$

holds if and only if

$$b - \bar{x}^T w \geq k(\alpha) \sqrt{w^T \Sigma w}, \quad (5)$$

where  $k(\alpha) = \sqrt{\alpha/(1-\alpha)}$ .

Using the conclusion of Lemma 1, the optimal hyperplane  $H(w, b)$  we are looking for is the optimal solution of the following optimization problem:

$$\begin{aligned} \max_{\alpha, w \neq 0, b} \quad & \alpha \\ \text{s.t.} \quad & b - \bar{x}^T w \geq k(\alpha) \sqrt{w^T \Sigma w} \\ & \text{and } x_i^T w - b \geq 0 \end{aligned} \quad (6)$$

Considering that the function  $k(\alpha)$  is an increasing function on the domain  $\alpha \in [0, 1)$ , the above formula can be written as

TABLE 1: Extracted features.

Name	Equation
Energy	$\epsilon_r = \int_T r(t)^2 dt$
Maximum amplitude	$r_{\max} = \max_t  r(t) $
Kurtosis <sup>a</sup>	$k = \sigma_{ r }^4 T \int_T (r(t) - \mu_{ r })^4 dt, \mu_{ r } = \frac{1}{T} \int_T r(t) dt, \sigma_{ r }^2 = \frac{1}{T} \int_T  r(t) - \mu_{ r } ^2 dt$
Mean excess delay <sup>b</sup>	$\tau_{MED} = \int_T t\psi(t)dt, \psi(t) = \frac{ r(t) ^2}{\epsilon_r}$
RMS delay spread <sup>c</sup>	$\tau_{RMS} = \int_T (t - \tau_{MED})^2 \psi(t)dt, \psi(t) = \frac{ r(t) ^2}{\epsilon_r}$
Rise time <sup>d</sup>	$t_{rise} = t_H - t_L, t_L = \min \{t :  r(t)  \geq \alpha \sigma_n\}, t_H = \min \{t :  r(t)  \geq \beta r_{\max}\}$

Where  $\sigma_n$  is the standard deviation of the thermal noise, the values of  $\alpha > 0$  and  $0 < \beta \leq 1$  are chosen empirically. We also used  $\alpha = 6, \beta = 0.6$  as in literature [8].

$$\begin{aligned} & \max_{k, w \neq 0, b} k \\ & \text{s.t. } b - \bar{x}^T w \geq k \sqrt{w^T \Sigma w} \\ & \text{and } x_i^T w - b \geq 0 \end{aligned} \quad (7)$$

By combining the two constraints to mitigate b, we finally get the formula to solve the imbalanced binary classification problem:

$$\begin{aligned} & \max_{k, w \neq 0} k \\ & \text{s.t. } x_i^T w - \bar{x}^T w \geq k \sqrt{w^T \Sigma w} \end{aligned} \quad (8)$$

The optimization problem represented by (8) belongs to a convex optimization problem, which has an interesting geometric explanation. Figure 1 shows this geometric explanation.

The ellipse with  $\bar{x}$  as its center and covariance matrix  $\Sigma$  as its shape has a unique intersection point  $x_i$  with the convex hull of positive class  $p^*$  as  $k$  increases. The intersection point  $p^*$  is the projection of  $\bar{x}$  on the convex set. The tangent line that passes through the intersection point  $p^*$  and is tangent to the ellipse is the hyperplane we are looking for.

Next, we will give the solving process of the hyperplane ( $w, b$ ) corresponding to formula (8):

$$\begin{aligned} & \max_{k, w \neq 0} k \\ & \text{s.t. } x_i^T w - \bar{x}^T w \geq k \sqrt{w^T \Sigma w} \end{aligned} \quad (9)$$

Using the homogeneity of  $w$ , we impose a constraint condition,  $\sqrt{w^T \Sigma w} = 1$ ; then (8) can be written as follows:

$$\begin{aligned} & \max_{k, w \neq 0} k \\ & \text{s.t. } x_i^T w - \bar{x}^T w \geq 1 \\ & \text{and } k = \frac{1}{\sqrt{w^T \Sigma w}} \end{aligned} \quad (10)$$

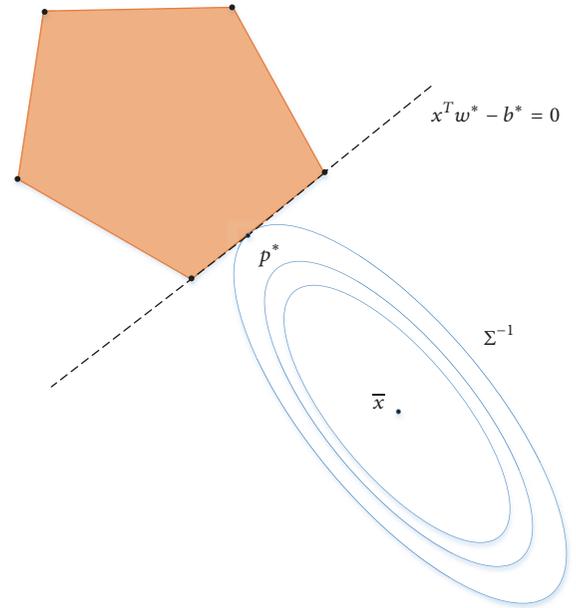


FIGURE 1: The geometric explanation.

From formula (10), we can see that the function of variable  $k$  is monotonically decreasing function in domain  $r$ , so we only need to satisfy the following formula conditions in order to obtain the maximum value of  $k$ :

$$\begin{aligned} & \min_{w \neq 0} w^T \Sigma w \\ & \text{s.t. } (x_i - \bar{x})^T w \geq 1 \end{aligned} \quad (11)$$

We can see that formula (11) is very similar to the formula of hard-spaced classification support vector machine in essence and also is an extension of the one-class support vector machine formula [15, 16]. However, there are two important differences between the two formula: first, we minimize the Mahalanobis distance between  $\bar{x}$  of the LOS signal samples and its projection on the convex hull of the NLOS signal

samples instead of the  $l_2$ -norms of  $W$ ; Secondly, we separate the NLOS signal samples from the mean of the LOS signal sample distribution, not from the origin of coordinates.

Like the one-class support vector machine, the constraint condition that all NLOS signals are correctly classified in formula (11) is not in conformity with reality; therefore, we add a relaxation factor  $\xi_i \geq 0$  to the constraint condition, so that we can obtain the following convex optimization problem:

$$\begin{aligned} \min_{w \neq 0} \quad & \frac{1}{2} w^T \Sigma w + C \sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & (x_i - \bar{x})^T w \geq 1 - \xi_i, \\ & \xi_i \geq 0 \end{aligned} \quad (12)$$

where  $C \in \mathbb{R}_+$  is the penalty factor, which is used to balance the complexity of the model and the size of the training samples that are allowed to go wrong. Formula (12) is our method to deal with the classification of imbalanced LOS signals and NLOS signals; we call it moment-based imbalanced binary classifier or MIBC. Introduce Lagrange multiplier  $\alpha \in \mathbb{R}^n$ ,  $\nu \in \mathbb{R}^n$ ,  $\alpha_i \geq 0$ ,  $\nu_i \in \mathbb{R}$  and define  $\tilde{x}_i = x_i - \bar{x}$ . The Lagrangian function constructed by (12) is

$$\begin{aligned} L(w, \xi_i, \alpha, \nu) = & \frac{1}{2} w^T \Sigma w + C \sum_{i=1}^n \xi_i \\ & - \sum_{i=1}^n \alpha_i [\tilde{x}_i^T w - 1 + \xi_i] - \sum_{i=1}^n \nu_i \xi_i \end{aligned} \quad (13)$$

In order to obtain the minimum value of the constructed Lagrangian function, the partial derivative of  $w$ ,  $\xi_i$  is zero:

$$\begin{aligned} \frac{\partial L}{\partial w} = 0 : w = & \Sigma^{-1} \sum_{i=1}^n \alpha_i \tilde{x}_i, \\ \frac{\partial L}{\partial \xi_i} = 0 : C - \alpha_i - \nu_i = & 0. \end{aligned} \quad (14)$$

By substituting the result of the above formula and the constraint condition  $C \geq \alpha_i \geq 0$  into formula (13), we can obtain the dual problem of maximizing  $\alpha$ :

$$\begin{aligned} \max_{\alpha} \quad & -\frac{1}{2} \sum_{i=1, j=1}^n \alpha_i \alpha_j \tilde{x}_i^T \Sigma^{-1} \tilde{x}_j + \sum_{i=1}^n \alpha_i \\ \text{s.t.} \quad & C \geq \alpha_i \geq 0 \end{aligned} \quad (15)$$

The above formula is a quadratic function extremum problem constrained by inequality; according to Karush-Kuhn-Tucker theorem, formula (15) must satisfy the following optimization conditions (KKT conditions):

$$\alpha_i [\tilde{x}_i^T w - 1 + \xi_i] = 0, \quad \forall i \quad (16)$$

Through formulas (15) and (16), the SMO algorithm can be used to obtain the optimal solution of the daily coefficient  $\alpha$ .

Substituting formula (14) can obtain the weight vector  $W$  of the optimal hyperplane and the support vector  $x_i = \tilde{x}_i + \bar{x}$  and then obtain the hyperplane threshold  $b = w^T x_i$ . After finding the optimal hyperplane ( $w, b$ ), the discriminant function is used:

$$f(z) = I(w^T z - b \geq 0) \quad (17)$$

If the condition is satisfied, the output is 1, indicating that the test sample signal is a NLOS signal, and if the condition is not satisfied, the output is 0, indicating that the test sample signal is a LOS signal. When LOS signals and NLOS signals are not linearly separable in the original space, we can solve the problem of nonlinear classification by introducing a modified kernel function  $\widehat{K}(x_i, x_j)$  to map the original data to a high-dimensional space. The modified Gaussian kernel function is used in this paper:

$$\widehat{K}(x_i, x_j) = \exp\left(-\frac{(x_i - x_j)^T \Sigma^{-1} (x_i - x_j)}{\sigma^2}\right) \quad (18)$$

As can be seen from (15), the MIBC algorithm first needs to preprocess the LOS signal samples. The calculation amount in calculating the covariance matrix is  $O(d^3)$ , and the calculation amount in finding the optimal solution to the dual problem is  $O(n^3)$ . Therefore, the overall complexity of the algorithm is  $O(d^3) + O(n^3)$ . However, the overall complexity of SVDD proposed in literature [8] is  $O(N^3)$ . Since the number of LOS signal samples  $N$  is much larger than the number of NLOS signals  $n$ , the MIBC method proposed by us has lower complexity than the SVDD algorithm. The identification of NLOS signals by UWB indoor positioning system is that the receiver classifies a large number of received LOS signals and a small amount of NLOS signals. In essence, this is an imbalanced classification problem, although the number of NLOS signals is very small; it is very important; misreading of NLOS signals will lead to serious ranging errors. MIBC is a typical class-imbalance classification method, which separates LOS signals from NLOS signals by finding the optimal hyperplane. In order to reduce the error rate of identifying NLOS signals, we use the mean and covariance matrix of the LOS signal samples and a small amount of all NLOS signal samples to train the classifier. After training, we can determine whether the test signal is a LOS signal or a NLOS signal.

#### 4. Identification Performance

In order to verify the effectiveness of the proposed MIBC method, we designed three experimental schemes based on the datasets of the received signal waveform on the platform of MATLAB 2013 for verification. During the simulation, in order to prove the performance of the MIBC method in this paper more intuitively, we compared the LS-SVM method in document [9] with the SVDD method in document [8]. At the same time, we verified the performance of LS-SVM and MIBC when the ratio of the two kinds of datasets is different. Finally, it is also verified that when the ratio of the number

TABLE 2: Extracted features and AUC values.

Subset of features	Combination of features	AUC value		
		LS-SVM	SVDD	MIBC
Set A	$\varepsilon_r$	0.9013	0.8902	0.9006
Set B	$r_{\max}$	0.9311	0.9031	0.9284
Set C	$r_{\max}, t_{\text{rise}}$	0.9548	0.9384	0.9461
Set D	$\varepsilon_r, t_{\text{rise}}, k$	0.9572	0.9406	0.9593
Set E	$r_{\max}, t_{\text{rise}}, k, \tau_{\text{MED}}$	0.9582	0.9438	0.9605
Set F	$\varepsilon_r, r_{\max}, t_{\text{rise}}, k$	0.9646	0.9500	0.9634
Set G	$t_{\text{rise}}, k, \tau_{\text{MED}}, \tau_{\text{RMS}}$	0.9617	0.9483	0.9610
Set H	$\varepsilon_r, r_{\max}, t_{\text{rise}}, \tau_{\text{MED}}, k$	0.9589	0.9457	0.9553
Set I	$\varepsilon_r, r_{\max}, t_{\text{rise}}, \tau_{\text{MED}}, \tau_{\text{RMS}}, k$	0.9624	0.9465	0.9618

of training samples for the LOS signal and the NLOS signal is 1: 0.06, the influence of the capacity of the two types of datasets on the classification performance is correspondingly increased according to this ratio. In addition, in order to evaluate the classification performance of the three algorithms, we used k-fold cross-validation. LS-SVM needs k-1 groups of samples from both LOS signals and NLOS signals to train when using k-fold cross-validation. SVDD only needs k-1 groups of samples in the LOS signal data to train, and then the remaining LOS group and NLOS signals with the same capacity as the randomly selected samples are used as verification data. MIBC requires k-1 groups of samples in the NLOS signal datasets and the mean and covariance matrix of k-1 groups of samples in the LOS signal datasets to train.

Datasets and their processing: we adopted the datasets used in [8, 9] to carry out experiments and compared three methods with their results. The datasets include 1024 sets of measurement data, including 512 sets of LOS signals and 512 sets of NLOS signals. Log function is used to convert the values of the features. At the same time, it should be noted that before using the MIBC method, we must first calculate the mean value and covariance matrix of the LOS signal training samples' features.

Experimental design: the kernel functions involved in the experimental algorithm all adopt Gaussian radial basis function kernels, and the kernel parameter size is 5. Because the recognition accuracy in traditional machine learning cannot be used to evaluate the performance of the algorithm in class-imbalance learning, like literature [8], receiver operating characteristic curve (ROC) and area under roc curve (AUC) values are used as evaluation criteria in this paper. ROC curve reflects the balance between the recognition rate of large class and the false recognition rate of small class. If one ROC curve is at the upper left of the other curve, the performance of the classifier corresponding to the former is better than that of the latter. The AUC value refers to the area under the ROC curve. The higher the AUC value, the better the performance of the corresponding classifier.

Performance: in this experiment, nine kinds of feature combinations were extracted. The datasets adopted by the three methods include 500 LOS signals and 500 NLOS

TABLE 3: The AUC values of the two methods for the number of training samples of LOS and NLOS in different proportions.

$n_{\text{LOS}} : n_{\text{NLOS}}$	LS-SVM	MIBC
1 : 0.4	0.9491	0.9628
1 : 0.1	0.8623	0.9624
1 : 0.08	0.8216	0.9623
1 : 0.06	0.7840	0.9621

signals. The feature combinations and corresponding AUC values are shown in Table 2. As can be seen from Table 2, under the feature combination of group F, the classification performance of the three methods is the best. Figure 2 shows the roc curves of the three methods using 10-fold cross-validation under the feature combination of group F. It can be seen that the MIBC method is superior to the LS-SVM method in areas where the NLOS signal acceptance rate is less than 0.07, and vice versa. Compared with SVDD, MIBC has slightly better classification performance.

Figure 3 shows that, under the feature combination of set F, LS-SVM and MIBC use 500 LOS samples and 30 NLOS samples and the ratio of the two datasets is 1: 0.06, while, for SVDD, only 500 LOS samples datasets are still utilized. We utilize tenfold cross-validation to verify the ROC curves of the three methods finally. It can be seen that the MIBC method is obviously better than the LS-SVM method and slightly better than the SVDD method in classification performance. Table 3 shows AUC values of LS-SVM and MIBC under different ratios for the number of LOS and NLOS samples.

Set the ratio of LOS and NLOS samples datasets to be 1: 0.06, randomly select 500 groups, 400 groups, 300 groups, and 200 groups of LOS samples, and then randomly select 30 groups, 24 groups, 18 groups, and 12 groups of NLOS samples according to the ratio. Using 10-fold, 8-fold, 6-fold, and 4-fold cross-verification, respectively, the experimental results are shown in Figure 4. Under the condition that the proportion of LOS and NLOS samples remains unchanged, the classification accuracy of MIBC can be improved by increasing the number of NLOS and LOS samples proportionally.

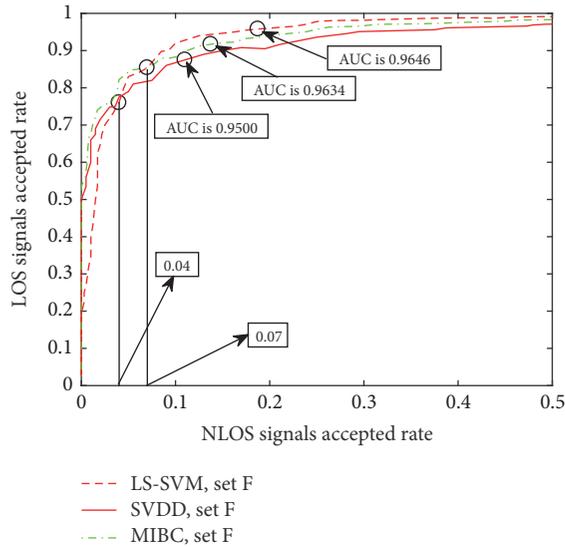


FIGURE 2: The ROC curves for the best cases in the LS-SVM, SVDD and MIBC algorithms. For LS-SVM, both LOS and NLOS signals are needed for training, and the number of each of them is 500. For SVDD, only 500 LOS signals are needed for training. For MIBC, the mean value and covariance matrix of the features of 500 LOS signals are calculated firstly, and then 500 NLOS signals are needed for training.

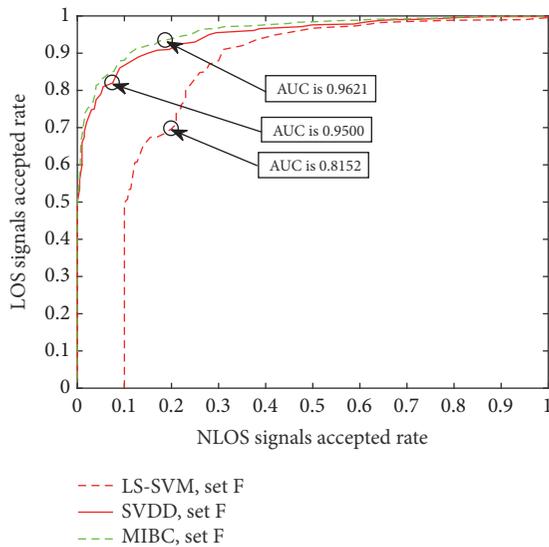


FIGURE 3: The ROC curves for the best cases in the LS-SVM, SVDD and MIBC algorithms when the quantity ratio of the two classification data is 1:0.06.

### 5. Conclusion

UWB positioning system faces the problem of serious positioning error caused by NLOS signals in practical application. Although the number of NLOS signals in the actual environment is far less than that of LOS signals, it is very important for accurate positioning. For the first time, this paper transforms the identification of NLOS signals into

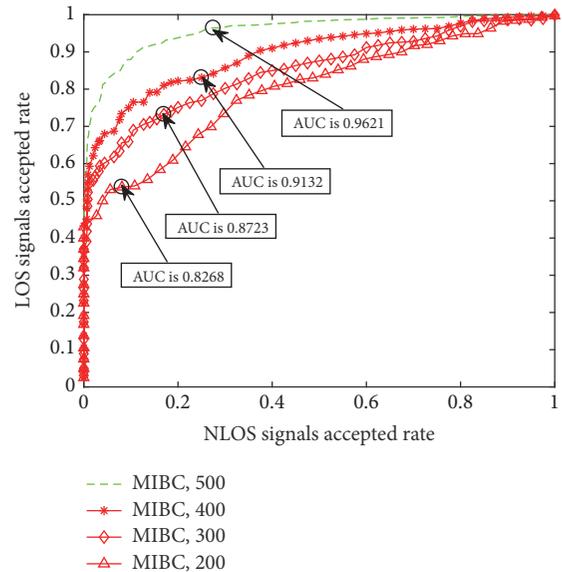


FIGURE 4: Illustration of how the performance of the MIBC detectors varies with the number of the NLOS and LOS samples. Set F is utilized and the ratio of the LOS and NLOS samples datasets is 1:0.06.

a typical class-imbalance classification problem. The traditional machine learning methods are all aimed at balanced datasets, so the traditional classification methods are not suitable for the real scenario. In addition, there are many differences between the waveform features of NLOS and LOS signals. Based on the above two points, we propose a new imbalanced binary classification method for NLOS identification of UWB positioning system, which we call MIBC. We compare the LS-SVM, SVDD and MIBC methods. The experimental results show that the performance of the MIBC method in imbalanced datasets classification is better, and its complexity is lower than that of the previous two methods. In the next step, we will study how to mitigate the error caused by the NLOS signal, because identifying the NLOS is only the first step to mitigate the error caused by it.

### Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

### Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

### Acknowledgments

The authors would like to thank Shiwei Tian for providing the dataset of UWB waveforms. The authors also would like to thank Edouard Grave for the development and free use of the MIBC. This work is supported by the Department of Military Logistics, Army Logistics University of PLA.

## References

- [1] F. Sun, H. Yin, and W. Wang, "Finite-Resolution Digital Receiver for UWB TOA Estimation," *IEEE Communications Letters*, vol. 16, no. 1, pp. 76–79, 2012.
- [2] E. Baccarelli, M. Biagi, C. Pelizzoni, and N. Cordeschi, "Non-coherent synchronization of UWB-IR multiple-antenna multipath channels," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 6, pp. 2713–2726, 2009.
- [3] Y. Salih-Alj, C. Despins, and S. Affes, "Design considerations for an UWB computationally-efficient fast acquisition system for indoor line-of-sight ranging applications," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2776–2784, 2011.
- [4] H. Shao and N. C. Beaulieu, "Direct sequence and time-hopping sequence designs for narrowband interference mitigation in impulse radio UWB systems," *IEEE Transactions on Communications*, vol. 59, no. 7, pp. 1957–1965, 2011.
- [5] S. Zhao, P. Orlik, A. F. Molisch, H. Liu, and J. Zhang, "Hybrid ultrawideband modulations compatible for both coherent and transmit-reference receivers," *IEEE Transactions on Wireless Communications*, vol. 6, no. 7, pp. 2551–2558, 2007.
- [6] D. Dardari, A. Conti, J. Lien, and M. Z. Win, "The effect of cooperation on uwb-based positioning systems using experimental data," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, no. 1, 2008.
- [7] İ. Güvenç, C. Chong, F. Watanabe, and H. Inamura, "NLOS Identification and Weighted Least-Squares Localization for UWB Systems Using Multipath Channel Statistics," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, no. 1, 2007.
- [8] S. Tian, L. Zhao, and G. Li, "A support vector data description approach to NLOS identification in UWB positioning," *Mathematical Problems in Engineering*, vol. 2014, 2014.
- [9] S. Maranò, W. M. Gifford, H. Wymeersch, and M. Z. Win, "NLOS identification and mitigation for localization based on UWB experimental data," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 7, pp. 1026–1035, 2010.
- [10] H. Wymeersch, S. Maranò, W. M. Gifford, and M. Z. Win, "A machine learning approach to ranging error mitigation for UWB localization," *IEEE Transactions on Communications*, vol. 60, no. 6, pp. 1719–1728, 2012.
- [11] J. Borras, P. Hatrack, and N. B. Mandayam, "Decision theoretic framework for NLOS identification," in *Proceedings of the 48th IEEE Vehicular Technology Conference (VTC '98)*, pp. 1583–1587, May 1998.
- [12] N. Alsindi, C. Duan, J. Zhang, and T. Tsuboi, "NLOS channel identification and mitigation in ultra wideband ToA-based wireless sensor networks," in *Proceedings of the 6th Workshop on Positioning, Navigation and Communication (WPNC '09)*, pp. 59–66, March 2009.
- [13] E. Grave and L. E. Ghaoui, "Fast imbalanced binary classification: a moment-based approach," in *Machine Learning*, pp. 1–11, Fast imbalanced binary classification, 2014.
- [14] G. R. Lanckriet, L. El Ghaoui, C. Bhattacharyya, and M. I. Jordan, "A robust minimax approach to classification," *Journal of Machine Learning Research*, vol. 3, no. 3, pp. 555–582, 2003.
- [15] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.
- [16] Y.-S. Choi, "Least squares one-class support vector machine," *Pattern Recognition Letters*, vol. 30, no. 13, pp. 1236–1240, 2009.



**Hindawi**

Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

