Journal of Nonlinear and Convex Analysis
Volume 26, Number 6, 2025, 1879–1886

Yokohama Publishers
ISSN 1880-5221 ONLINE JOURNAL

INTUITIONISTIC FUZZY ROBUST TWIN LEARNING FRAMEWORK FOR IMBALANCED DATA

FENG JI AND JUN MA*

Abstract. Addressing the formidable challenge of managing imbalanced datasets in machine learning, where conventional approaches like twin extreme learning machine (TELM) often suffer from bias towards the majority class and misclassify minority instances as noise, necessitates innovative solutions. Furthermore, the ubiquitous presence of noise and outliers in datasets exacerbates this issue, as TELM struggles to effectively deal with them. To overcome these limitations, we introduce the intuitionistic fuzzy twin extreme learning machine for imbalanced data (IFRTELM), a novel approach tailored to tackle the intricacies of imbalanced datasets, even amidst noise and outliers. Our method meticulously addresses the imbalance through a strategic weighting system that balances the influence of different classes, ensuring that minority classes are not overlooked. Additionally, IFRTELM incorporates a margin-based mechanism that skillfully dampens the disruptive effects of noise and outliers, enhancing the robustness and accuracy of the model. Experimental evaluations have underscored the remarkable performance of IFRTELM, demonstrating its superiority over comparable techniques in handling imbalanced datasets, even in noisy environments. This innovative approach paves the way for more effective and reliable machine learning solutions in real-world applications where data imbalance and contamination are prevalent.

1. Introduction

Imbalanced classification tasks frequently arise in numerous real-world applications, posing a formidable challenge in the realm of machine learning. This issue is particularly pressing as it necessitates urgent attention and prompt solutions due to the suboptimal performance exhibited by most classification algorithms when confronted with imbalanced datasets. Typically designed for balanced learning scenarios, these algorithms struggle to effectively handle the disparities in class distributions, leading to reduced accuracy and biased predictions [2, 4]. Thus, addressing the challenges associated with imbalanced classification has become a critical research focus in the field. The Extreme Learning Machine (ELM) [1, 5–7] has garnered substantial attention in recent years. ELM and its derivatives provide a robust alternative to traditional neural network algorithms, which often succumb

2020 *Mathematics Subject Classification.* 65F50, 68T10, 90C31.

Key words and phrases. Intuitionistic fuzzy number, quadratic programming problem, Twin extreme learning machine.

This research was supported in part by the Natural Science Foundation of Ningxia Provincial (No.2024AAC05055, No.2023AAC02053), in part by the Fundamental Research Funds for the Central Universities of North Minzu University (No.2023ZRLG01, No.2021JCYJ07), in part by the National Natural Science Foundation of China (No.62366001, No.12361062), in part by the Key Research and Development Program of Ningxia (Introduction of Talents Project)(No.2022BSB03046).

*Corresponding author.

to the issue of local minima. Consequently, ELM has been the subject of extensive research and has been successfully applied to tackle imbalanced classification problems, demonstrating its efficacy in overcoming the challenges posed by uneven class distributions. Even though ELM and its extensions have been deeply examined and debated within the context of imbalanced classification problems, their ability to discern classes in binary classification problems is restricted to learning a singular hyperplane. This major limitation considerably deters their overall application and evolution. Currently, Wan et al. [10] introduced the notion of twin extreme learning machine (TELM), it boasts two primary advantages: its user-friendliness and remarkable learning efficiency. Its inherent symmetry permits TELM to divide a complex Quadratic Programming Problem (QPP) into two Sub-QPP. Addressing the formidable task of handling imbalanced datasets in machine learning, traditional methods like TELM often show a bias towards the majority class, leading to misclassifying minority instances as noise. This highlights the need for creative and urgent new approaches. The pervasive issue of noise and outliers within datasets further complicates this scenario, as TELM encounters difficulties in effectively managing these elements. Thus, there is a pressing need for novel solutions that can adeptly handle both data imbalance and contamination. In this study, we introduce a new classification model called Intuitionistic Fuzzy Twin Extreme Learning Machine (IFRTELM) designed to tackle binary classification challenges, especially when dealing with imbalanced datasets that include noise and outliers. The IFRTELM integrates intuitionistic fuzzy numbers (IFNs) to assign membership and nonmembership functions to individual training samples. The membership function evaluates the closeness of samples to their class centroids, whereas the nonmembership function measures the discrepancy between conflicting samples and their surroundings. Furthermore, the IFRTELM minimizes structural risk and enhances classification accuracy, outperforming comparable techniques in handling complex, imbalanced datasets even under noisy conditions.

2. Theoretical basis

2.1. **Twin Extreme Learning Machine.**

In the given training set $T = \{(x_i, y_i)|i = 1, 2 \ldots l\} \in (R^n, y)^l$, containing m_1 instances belonging to the positive category and m_2 instances belonging to the negative category, we can divide this set according to the category labels. Here, $x_i \in R^n$ denotes the feature vector of the sample, and $y_i \in y = \{-1, 1\}$ denotes the corresponding category labels, where -1 denotes the negative category and 1 denotes the positive category. The total number of samples l is the sum of the number of positive and negative instances, i.e. $l = m_1 + m_2$. To further elaborate the context of the neural network, we assume that the network processes the input samples through a hidden layer that produces different outputs for positive and negative class samples. We define H_1 and H_2 as the matrices representing the outputs of the hidden layer specifically for positive and negative class samples, respectively. H_1 is a matrix whose rows correspond to the hidden layer outputs for each of the m_1 positive class samples. The columns of H_1 would represent the activations of the hidden layer neurons for these positive samples. H_2 is similarly defined but for the m_2 negative class samples. Its rows correspond to the hidden layer outputs for each of the negative class instances, and its columns represent the hidden neuron activations for these samples.

$$H_1 = \begin{bmatrix} h_1(x_1) & \dots & h_l(x_1) \\ \vdots & \ddots & \vdots \\ h_1(x_{m_1}) & \dots & h_l(x_{m_1}) \end{bmatrix}, \ H_2 = \begin{bmatrix} h_1(x_1) & \dots & h_l(x_1) \\ \vdots & \ddots & \vdots \\ h_1(x_{m_2}) & \dots & h_l(x_{m_2}) \end{bmatrix},$$

where $h_i(x)$ is ith column vector of the hidden layer output matrices, $i = 1, 2, \dots N$. Wan et al. [10] introduced the notion of twin extreme learning machine (TELM), it boasts two primary advantages: its user-friendliness and remarkable learning efficiency. Its inherent symmetry permits TELM to divide a complex Quadratic Programming Problem (QPP) into two distinct, yet non-identical, QPPs involving hyperplanes, as exemplified in equations (2.1) and (2.2). This unique property endows TELM with the capability to operate at a significantly faster pace than traditional ELM, thereby enhancing its practical applicability and performance.

$$f_1(x) = \beta_1 h(x) = 0, \tag{2.1}$$

and

$$f_2(x) = \beta_2 h(x) = 0. \tag{2.2}$$

Thuly, the TELM can be written as

$$\begin{aligned} &\min_{\beta_1, \xi_1} \ \frac{1}{2}\|H_1\beta_1\|^2 + C_1 e_2^T \xi_1 \\ &\text{s.t.} \ - H_2\beta_1 + \xi_1 \geq e_2, \end{aligned} \tag{2.3}$$

and

$$\begin{aligned} &\min_{\beta_2, \xi_2} \ \frac{1}{2}\|H_2\beta_2\|^2 + C_2 e_1^T \xi_2 \\ &\text{s.t.} \ H_1\beta_2 + \xi_2 \geq e_1, \end{aligned} \tag{2.4}$$

where the slack vectors, denoted by the symbols $\xi_1 > 0$ and $\xi_2 > 0$, are assumed to be positive. The regularisation parameters, represented by the symbols $C_1 \geq 0, C_2 \geq 0$, are assumed to be positive. The vectorse$_1$ and e_2, which are of appropriate dimensions, are defined as vectors of ones.

Via Lagrange function and Karush-Kuhn-Kucher condition, we can get

$$\begin{aligned} &\min_{\alpha_1} \ \frac{1}{2}\alpha_1^T H_2 (H_1^T H_1)^{-1} H_2^T \alpha_1 - e_2^T \alpha_1 \\ &\text{s.t.} \ 0 \leq \alpha_1 \leq C_1 e_2, \end{aligned} \tag{2.5}$$

and

$$\begin{aligned} &\min_{\alpha_2} \ \frac{1}{2}\alpha_2^T H_1 (H_2^T H_2)^{-1} H_1^T \alpha_2 - e_1^T \alpha_2 \\ &\text{s.t.} \ 0 \leq \alpha_2 \leq C_2 e_1, \end{aligned} \tag{2.6}$$

In this paper, we may consider the vectors $\alpha_1 \geq 0$ and $\alpha_2 \geq 0$ of Lagrange multipliers to be non-negative, specifically equal to zero or greater than zero, and may also define the regularised term as a quantity of the form $I \times \epsilon$.

2.2. **Intuitionistic Fuzzy Set.** We assume that a non-empty set X can consist of a fuzzy set A in the universe X as follows [3,9]:

$$A = \{(x, \mu_A(x)) \mid x \in X\} \tag{2.7}$$

where $\mu_A : X \to [0,1]$ denotes the degree of membership of $xinX$ by $\mu_A(x)$. Then an intuitionistic fuzzy set can be defined as [3,9]

$$\tilde{A} = \left\{ \left(x, \mu_{\tilde{A}}(x), \nu_{\tilde{A}}(x)\right) \mid x \in X \right\}. \tag{2.8}$$

The membership and non-membership degrees of $xinX$ are given by $\mu_{\tilde{A}}(x)$ and $\nu_{\tilde{A}}(x)$, respectively, where $\mu_{\tilde{A}} : X \to [0,1]$ and $\nu_{\tilde{A}} : X \to [0,1]$. These functions satisfy $0 \le \mu_{\tilde{A}}(x) + \nu_{\tilde{A}}(x) \le 1$ and the degree of hesitancy of $xinX$ can be denoted as [3,9].

$$\pi_{\tilde{A}}(x) = 1 - \mu_{\tilde{A}}(x) - \nu_{\tilde{A}}(x). \tag{2.9}$$

For an IFN [3,9], we can express this as $\alpha = (\mu_\alpha, \nu_\alpha)$, where $\mu_\alpha \in [0,1]$ and $\nu_\alpha \in [0,1]$, and $0 \le \mu_\alpha + \nu_\alpha \le 1$. The largest IFN is $\alpha^+ = (1,0)$ and the smallest IFN is $\alpha^- = (0,1)$. The IFN for a given α can be calculated as follows:

$$s(\alpha) = \mu_\alpha - \nu_\alpha \tag{2.10}$$

where $s(\alpha)$ denotes the score of the IFN and $\alpha = (\mu_\alpha, \nu_\alpha)$. However, the scores of some IFNs cannot be determined. To solve this problem, the following function can be used instead

$$h(\alpha) = \mu_\alpha + \nu_\alpha. \tag{2.11}$$

Based on (2.9) and (2.11), we can get

$$h(\alpha) + \pi(\alpha) = 1. \tag{2.12}$$

If $s(\alpha_1) = s(\alpha_2)$ and $h(\alpha_1) < h(\alpha_2)$, then $\alpha_1 < \alpha_2$. Using Equation (2.10), we can derive the score function for other cases as follows:

$$H(\alpha) = \frac{1 - \nu(\alpha)}{2 - \mu(\alpha) - \nu(\alpha)}. \tag{2.13}$$

Thus, the connections between membership and nonmembership functions can be described as follows: 1) $s(\alpha_1) < s(\alpha_2) \Rightarrow H(\alpha_1) < H(\alpha_2)$; 2) $s(\alpha_1) = s(\alpha_2), h(\alpha_1) < h(\alpha_2) \Rightarrow H(\alpha_1) < H(\alpha_2)$ [3,9].

2.3. **Intuitionistic Fuzzy Membership Assignment.** In the following sections, we discuss the customized levels of membership and non-membership functions, uniquely crafted for individual training samples in the complex feature space.

- 1) The membership function is defined in the high-dimensional feature space, based on the distance between a training sample and the corresponding class centre. For each individual training sample, the degree of membership can be formulated as follows:

$$\mu(x_i) = \begin{cases} 1 - \frac{\left\|\phi(x_i) - C^+\right\|}{r^+ + \delta} y_i = +1 \\ 1 - \frac{\left\|\phi(x_i) - C^-\right\|}{r^- + \delta} y_i = -1 \end{cases} \tag{2.14}$$

  In the case where the variable parameter $\delta > 0$ is greater than or equal to zero, the notation $r^+ (r^-)$ and $C^+ (C^-)$ represent the radius and class centre of the positive (negative) class, respectively. The symbol $\|\cdot\|$ denotes the distance between the input sample and the corresponding class centre.

$$D\left(\phi\left(x_i\right), \phi\left(x_j\right)\right)=\left\|\phi\left(x_i\right)-\phi\left(x_j\right)\right\| \tag{2.15}$$

where the symbol ϕ is used to denote the input sample in the high feature space. Accordingly, the centroid of each class can be calculated using the following formula:

$$C^{\pm}=\frac{1}{l_{\pm}} \sum_{y_i= \pm 1} \phi\left(x_i\right) \tag{2.16}$$

where the variables l_{+} (l_{-}) represent the total number of samples of a positive and negative nature, in that order. The radius of each class can be calculated as follows:

$$r^{\pm}=\max_{y_i= \pm 1}\left\|\phi\left(x_i\right)-C^{\pm}\right\| . \tag{2.17}$$

- 2) The function of non-membership is as follows, as evidenced by references [3, 8, 9]: The non-membership function employs the relationship between all inharmonic points and the total number of training samples in their neighbourhood (i.e., $\rho\left(x_i\right)$)) in the following manner:

$$\nu\left(x_i\right)=\left(1-\mu\left(x_i\right)\right) \rho\left(x_i\right) \tag{2.18}$$

where $0 \leq \mu\left(x_i\right)+\nu\left(x_i\right) \leq 1$, and $\rho\left(x_i\right)$ is defined as

$$\rho\left(x_i\right)=\frac{\left|\left\{x_j \mid\left\|\phi\left(x_i\right)-\phi\left(x_j\right)\right\| \leq \alpha, y_j \neq y_i\right\}\right|}{\left|\left\{x_j \mid\left\|\phi\left(x_i\right)-\phi\left(x_j\right)\right\| \leq \alpha\right\}\right|} \tag{2.19}$$

where $\alpha>0$ is an adjustable parameter and $|\cdot|$ denotes the cardinality.

The degrees of membership and non-membership within the Intuitionistic Fuzzy Numbers (IFNs) are formulated by leveraging the inner product distance in the feature space. Consequently, kernel functions are employed as a means of constructing these IFNs.

Theorem 2.1 ([3, 8, 9])**.** *Suppose $K\left(x, x^{\prime}\right)$ is a kernel function. Thus, the inner product distance is expressed as*

$$\left\|\phi(x)-\phi\left(x^{\prime}\right)\right\|=\sqrt{K(x, x)+K\left(x^{\prime}, x^{\prime}\right)-2 K\left(x, x^{\prime}\right)} . \tag{2.20}$$

Theorem 2.2. *Building upon Theorem 2.1, the radii of the two classes can be expressed as follows:*

$$\text{1) } r^{+}=\max_{y_i=+1} \sqrt{K\left(x_i, x_i\right)+\frac{1}{l_{+}^2} \sum_{y_m=+1} \sum_{y_n=+1} K\left(x_m, x_n\right)-\frac{2}{l_{+}} \sum_{y_j=+1} K\left(x_i, x_j\right)} \tag{2.21}$$

$$\text{2) } r^{-}=\max_{y_i=-1} \sqrt{K\left(x_i, x_i\right)+\frac{1}{l_{-}^2} \sum_{y_m=-1} \sum_{y_n=-1} K\left(x_m, x_n\right)-\frac{2}{l_{-}} \sum_{y_{j=-1}} K\left(x_i, x_j\right)} . \tag{2.22}$$

3. Main Contributions

To overcome these limitations of TELM [10], we introduce the intuitionistic fuzzy twin extreme learning machine for imbalanced data (IFRTELM), as follow:

$$\begin{aligned} \min_{\beta_1,\xi_1} \ & \frac{1}{2}\|H_1\beta_1\|^2 + \|\beta_1\|^2 + C_1 s_2^T \xi_1 \\ \text{s.t.} \ & -H_2\beta_1 + \xi_1 \geq e_2, \end{aligned} \tag{3.1}$$

and

$$\begin{aligned} \min_{\beta_2,\xi_2} \ & \frac{1}{2}\|H_2\beta_2\|^2 + \|\beta_2\|^2 + C_2 s_1^T \xi_2 \\ \text{s.t.} \ & H_1\beta_2 + \xi_2 \geq e_1, \end{aligned} \tag{3.2}$$

where $\xi_1 > 0$ and $\xi_2 > 0$ are slack vector, $C_1 \geq 0, C_2 \geq 0$ are regularization parameters, e_1 and e_2 are column vectors of ones with desirable length, and $s_1 \in \mathcal{R}^{l_+}$and $s_2 \in \mathcal{R}^{l_-}$are the score values of class + and − , respectively.

Using the Lagrangian function and the Karush-Kuhn-Tucker (KKT) condition, we can conclude that

$$\beta_1 = -(H_1^T H_1 + I)^{-1} H_2^T \alpha_1, \tag{3.3}$$

and

$$\beta_2 = (H_2^T H_2 + I)^{-1} H_1^T \alpha_2. \tag{3.4}$$

where ϵI is a regularized term that overcomes the singular of $H_1^T H_1$ and $H_2^T H_2$.

Likewise, the Wolfe dual for (3.1) and (3.2) can be written as:

$$\begin{aligned} \min_{\alpha_1} \ & \frac{1}{2}\alpha_1^T H_2 (H_1^T H_1 + I)^{-1} H_2^T \alpha_1 - e_2^T \alpha_1 \\ \text{s.t.} \ & 0 \leq \alpha_1 \leq C_1 s_2, \end{aligned} \tag{3.5}$$

and

$$\begin{aligned} \min_{\alpha_2} \ & \frac{1}{2}\alpha_2^T H_1 (H_2^T H_2 + I)^{-1} H_1^T \alpha_2 - e_1^T \alpha_2 \\ \text{s.t.} \ & 0 \leq \alpha_2 \leq C_2 s_1, \end{aligned} \tag{3.6}$$

where $\alpha_1 \geq 0$ and $\alpha_2 \geq 0$ are the vectors of Lagrange multipliers.

After obtaining β_1 and β_2, we classify new sample points x utilizing the decision function outlined below:

$$f(x) = \arg\min_{k=1,2} d_k(x) = \arg\min_{k=1,2} |\beta_k^T h(x)|. \tag{3.7}$$

where $|\cdot|$ signifies the orthogonal distance of the data point x from the hyperplane defined by β_k. This function determines the class label by selecting the hyperplane ($k = 1$ or $k = 2$) that minimizes the perpendicular distance of x from it.

Furthermore, we delve into the computational intricacies of our method. The comprehensive computational cost of our algorithm is encapsulated by the complexity of $\mathcal{O}(\frac{m^3}{4} + 2L^3 + 2n^3)$, where L represents the number of nodes in the hidden layer, m signifies the total count of training samples, and n denotes the dimensionality of each sample. This breakdown offers insights into the scalability and efficiency of our approach.

4. Experimental comparative analysis

To assess the efficacy and generalization prowess of the IFRTELM, we embarked on a series of experiments utilizing extensive datasets crafted by David Musicant's NDC Data Generator[1]. Our objective was to gain a comprehensive understanding of how the computational demands of various algorithms escalate in tandem with the proliferation of data points. TABLE 1 outlines the specifics of the NDC datasets employed. Within our experimental framework, we randomly allocated 30% of each NDC dataset for training purposes, reserving the remaining 70% for testing. This process was iterated tenfold to ensure a reliable average classification accuracy. Additionally, we examined the performance under label noise conditions, introducing simulated noise by randomly selecting samples and inverting their labels. We experimented with varying degrees of label inversion, specifically at 0%, 5%, and 15% noise levels.

Table 1. Description of NDC datasets

Dataset	Samples	Features
NDC-11	100000	32
NDC-31	300000	32
NDC-51	500000	32

As evident from TABLE 2 , which presents a comparative analysis of classification accuracy, our methodology consistently outperforms other algorithms across all three noise level scenarios, demonstrating a superior ability to accurately classify data points in the presence of varying degrees of label noise.

Table 2. Experimental results.

Datasets	Label noise	TELM Accuracy (%)	IFRTELM Accuracy (%)
NDC-11	0%	83.90	86.77
	5%	74.35	81.09
	15%	68.63	73.96
NDC-3l	0%	75.66	82.02
	5%	69.17	78.56
	15%	63.79	77.01
NDC-51	0%	75.64	82.04
	5%	70.22	76.89
	15%	62.86	71.33

[1]Accessible at http://www.cs.wisc.edu/musicant/data/ndc

5. Conclusion

This paper presents a novel classification model, designated IFRTELM, which is proposed as a means of addressing the challenges inherent to binary classification. The IFRTELM employs intuitionistic fuzzy numbers (IFNs) to assign unique membership and non-membership functions to each training sample. The membership function quantifies the degree of proximity of a given sample to its class centre, whereas the non-membership function assesses the inter-relationship between discordant samples, thereby reducing the impact of noise. Moreover, the model minimises structural risk and enhances classification accuracy, thereby providing a robust solution for real-world applications that are confronted with data imbalance and contamination.

References

[1] Z. Bai, G. B. Huang, D. Wang, H. Wang and M. B. Westover,*Sparse extreme learning machine for classification*, IEEE Transactions on Cybernetics **44** (2014), 1858–1870.
[2] H. Guo, Y. Li, J. Shang, M. Gu, Y. Huang and B. Gong, *Learning from class-imbalanced data: Review of methods and applications*, Expert systems with applications **73** (2017), 220–239.
[3] M. Ha, C. Wang and J. Chen, *The support vector machine based on intuitionistic fuzzy number and kernel function*, Soft Computing **17** (2013), 635–641.
[4] H. He and E. A. Garcia, *Learning from imbalanced data*, IEEE Transactions on Knowledge and Data Engineering **21** (2009), 1263–1284.
[5] G. Huang, S. Song, J. N. Gupt and C. Wu, *Semi-supervised and unsupervised extreme learning machines*, IEEE Transactions on Cybernetics **44** (2014), 2405–2417.
[6] G. B. Huang, H. Zhou, X. Ding and R. Zhang, *Extreme learning machine for regression and multiclass classification*, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) **42** (2011), 513–529.
[7] G. B. Huang, Q. Y. Zhu and C. K. Siew, *Extreme learning machine: theory and applications*, Neurocomputing **70** (2006), 489–501.
[8] S. Rezvani, X. Wang and F. Pourpanah, *Intuitionistic fuzzy twin support vector machines*, IEEE Transactions on Fuzzy Systems **27**(2019),2140–2151.
[9] M. Tanveer, M. A. Ganaie, A. Bhattacharjee and C. T. Lin, *Intuitionistic fuzzy weighted least squares twin SVMs*, IEEE Transactions on Cybernetics **53** (2022), 4400-4409.
[10] Y. Wan, S. Song, G. Huang and S. Li, *Twin extreme learning machines for pattern classification*, Neurocomputing **260** (2017), 235–244.

Manuscript received July 10, 2024
revised November 11, 2024

F. Ji
School of Mathematics and Information Science, North Minzu University, Yinchuan 750021, China
E-mail address: nxjifeng@163.com

J. Ma
School of Mathematics and Information Science, North Minzu University, Yinchuan 750021, China
E-mail address: jun_ma1990@126.com