Active Learning for Imbalanced Ordinal Regression

JIAMING GE¹, HAIYAN CHEN¹, DONGFANG ZHANG¹, XIAYE HOU¹, AND LIGANG YUAN²
¹College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China
²College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China
Corresponding author: Haiyan Chen (chenhaiyan@nuaa.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61501229, and in part by the Fundamental Research Funds for the Central Universities under Grant NS2019054 and Grant NS2020045.

ABSTRACT Ordinal regression (OR), also called ordinal classification, is a special multi-classification designed for problems with ordered classes. Imbalanced data hinders the performance of classification algorithms, especially for OR algorithms, as imbalanced class distributions often arise in OR problems. In this article, we address an active learning based solution for imbalanced OR problem. We propose an active learning algorithm for OR (AL-OR) to select the most informative samples from unlabeled samples, mark them and add them to the training set. Based on AL-OR, we put forward an improved active learning for imbalanced OR (IAL-IOR), which further adjust the sampling strategy of AL-OR dynamically to make the training data as valuable and balanced as possible. Recall rate for multi-classification and new mean absolute error are designed to measure the performance of the algorithms. To the best of our knowledge, our algorithm is the first algorithm for imbalanced OR in algorithm level. The experimental results show that the proposed algorithms have faster convergence and much better generalization ability than the classical methods and the state-of-the-art methods under the evaluation measurements for imbalance problems. In addition, we also proved the effectiveness of our algorithms by statistical analysis.

INDEX TERMS Ordinal regression, class imbalanced, active learning, evaluation method.

I. INTRODUCTION

Multi-classification is an important task in machine learning. As a special case of multi-classification, OR is designed to solve the problems with ordered labels. For example, a teacher always rates his/her students by giving grades (A, B, C, D, F) on their performance [1]. When we use common multi-classification algorithms to predict the grades, the order information of the grades is obviously ignored. OR algorithms are designed to make full use of the order among the labels. OR problems usually appear in many research areas, such as medical research, age predict, brain computer interface, credit rating, econometric model, face recognition, facial beauty assessment, image classification, wind speed prediction, social sciences, text classification, and more [2].

Class imbalanced problem is a big challenge for classification algorithms. In this problem, class with more samples is majority class and class with fewer samples is minority class. It is important to predict the minority classes correctly, for the minority classes often represent the unusual cases to which we should pay special attention, such as, high rate in credit rating, high speed in wind speed prediction and so on. Due to the designed principles, most machine learning algorithms optimize the overall classification accuracy, while sacrificing the accuracy on minority classes. Therefore it is necessary to design some methods to improve the classification accuracy on minority classes without jeopardizing the accuracy of the majority classes severely [3].

Imbalanced problem has been widely studied for standard classification problems (binary and multi-class classification). Data level methods balance the skew distributions on the dataset by using data pre-processing, such as under-sampling, over-sampling. In under-sampling algorithms, the sample size of majority classes will be dislodged to reach the desired rates of different classes. In over-sampling algorithms, such as SMOTE (Synthetic Minority Over Sampling Technique) [4], MDO (Mahalanobis Distance-based Over-sampling technique) [5], the sample size of minority classes will increase by generating new samples to reach the desired rates. When imbalanced data occurs in OR, new methods
should be designed to tackle the peculiarities. Currently, the Graph-Based Over-sampling can generate synthetic samples by considering the distribution of minority class data and the order of samples. During over-sampling it captures the structure of the data by constructing a sample graph and considers the paths which contain the ordinal constraints of the data. Besides, new samples are generated near the boundary of two adjacent classes to soften the ordinal structure of the samples [6]. The Cluster-Based Weighted Over-sampling clusters minority classes at first, and then oversamples them based on their distance, and finally sorts the classes [7]. Synthetic Minority oversampling technique to deal exclusively with imbalanced Ordinal Regression (SMOR) is a direction-aware oversampling algorithm [8], and it can effectively avoid wrong synthetic samples generation by considering the rank of the classes. SMOR computes the selection weight of being used to generate synthetic samples for each candidate generation direction. However, they are both designed in data level for OR. For under-sampling, it may dismiss valuable samples which are decisive for building classifiers. For over-sampling, it increases the probability of overfitting by replicating the minority classes samples.

Instead of focusing on modifying the training set in order to combat class skew, other methods aim at modifying the classifier learning procedure itself, such as cost-sensitive methods [9], ensemble learning algorithms [10], active learning [11], [11]–[14], and one-class learning [15]. Among them, active learning can counteract the harmful effects of learning under imbalanced classes by selecting the most useful samples for the classifier [11]. Since active learning is implemented on a random set of training populations rather than the entire training dataset, it can reduce the computational complexity of dealing with large imbalanced datasets [14]. Moreover, active learning provides a progressive sampling strategy which makes it is possible to adjust a sampling strategy dynamically by observing some indicators of the ease of the different classes [12], and gradually improve the performance of the classifier by selecting the samples that have the most learning value for the current classifier each time. The time complexity of active learning mainly comes from the time it takes to find the best samples from unlabeled samples and retrain. The time complexity of searching most informative sample from unlabeled data has been solved by small pools and early stopping, and the time complexity of retraining has been solved by incremental learning [12]. Active learning not only inherits the advantage of reducing computational complexity (by small pools, early stopping and incremental learning) of under-sampling, but also avoids the disadvantage that under-sampling may delete useful information. It’s noteworthy that learning difficulties of the different classes that make up a learning task may be different (in terms of the number of instances required). Therefore, we need sampling strategies that can be adjusted dynamically by observing different learnability indicators. Such sampling strategies can be achieved by active learning [13], which can select the most useful samples for OR classifier by these sampling strategies. Moreover, as shown in Figure 1, we can also reasonably assume the samples inside the different boundaries of OR are balanced [12].

In this article, we propose an active learning method to deal with the imbalanced OR in algorithm level. First, we transform OR to an extended binary classification [16], [17], so that ordinal regression can be achieved by SVM (Support Vector Machine). Then we design a sampling strategy of active learning for OR, and then we adjust the sampling strategy dynamically to get a more valuable and balanced training set from imbalanced data. Finally, we propose new evaluation methods specifically for imbalanced OR to prove the efficiency of our algorithm.

The main contributions of this article are summarized as follows.

1) To the best of our knowledge, our algorithm is the first algorithm for imbalanced OR in algorithm level.
2) We put forward a sampling strategy for OR (AL-OR), and an improved active learning method for imbalanced OR (IAL-IOR).
3) We propose improved evaluation methods to evaluate the performance of imbalanced ordinal regression.

We organize the rest of the paper as follows. In section II, we give a brief review on the related works. In section III, we present the transformation of OR to an extended binary classification model, and its SVM based solution. In section IV, we put forward a sampling strategy for OR and an active learning with balanced sampling process and novel evaluation methods for imbalanced OR. In section V, we carry out our experiments on a variety of datasets and the experimental results are discussed. Finally, in section VI, we give some conclusions.

II. RELATED WORK
In this section, we give a brief overview on ordinal regression and active learning.

A. ORDINAL REGRESSION
Many real world applications present the label ordinal structure and ordinal regression has increased the number of methods and algorithms developed over the last years [2].
Over the past decade, a number of noteworthy research advances have been made in supervised learning of ordinal regression [17]–[19]. Since support vector machines (SVMs) have gained profound interest because of good generalization performance [16], there are several support vector OR (SVOR) formulations proposed to tackle OR problems. Shashua and Levin [20] proposed fixed-margin-based formulation and sum-of-margin-based formulation by finding multiple parallel hyperplanes. Chu and keerthi [19] improved the fixed-margin-based formulation by explicitly and implicitly keeping ordinal inequalities on the thresholds. Cardoso and Costa [21] proposed a data replication method and mapped it into SVM by using the fixed-margin-based formulation implicitly. Li and Lin [16] proposed a reduction framework from ordinal regression to binary classification based on extended examples. This framework allows to design a ordinal regression model based on a binary classification and derive new generalization bounds for ordinal regression from known bounds for binary classification. Moreover, it unifies many existing ordinal regression algorithms. In our paper, we build the ordinal regression by using this reduction framework and SVM. Recently, [22] presented a new Kernel Extreme Learning Machine for ordinal regression (KELMOR) by exploiting a quadratic cost-sensitive encoding scheme to deal with the efficiency of OR in the big data scenario. Reference [23] proposed a novel ordinal regression model, which is named as nonparallel support vector ordinal regression (NPSVOR), a set of possible nonparallel hyperplanes are constructed independently. However, a small number of literatures have considered the imbalanced ordinal regression [24].

Reference [6] creates synthetic samples by considering the distribution and ordering of minority data. The main assumption of this method is that when resampling samples in ordinal regression problem, the ordering of the classes should be considered, and the ordering is generally represented by a latent manifold. To take advantage of this collector, it captures the structure of the data by constructing a pattern-based graph, and considers paths preserve data order constraints when oversampling. In addition, new samples are created at the boundaries between adjacent classes to smooth the ordinal nature of the dataset.

Reference [7] aims to address the imbalanced ordinal regression by clustering the minority classes and over-sampling them based on their distance firstly, and then ordering the relationship with the samples of other classes. The final size of an oversampling cluster depends on its complexity and initial size in order to generate more synthetic instances for more complex and smaller clusters and fewer instances for more complex and larger clusters. An improved agglomerative hierarchical clustering algorithm is proposed to reduce the occurrence of superimposed composite samples during oversampling. Moreover, a new measurement method is proposed to quantify the balance between the complexity of the cluster and the initial size.

B. ACTIVE LEARNING

Active learning as a standard machine learning problem, has been extensively studied in many research filed. Based on different sample strategies, they can be grouped into these categories [25]: 1) uncertainty sampling where an active learner queries the samples about which it is least certain how to label. 2) Query-By-Committee which involves maintaining a committee, and the most informative query is considered to be the samples about which they most disagree. 3) Expected Model Change where an active learner queries the samples that would impart the greatest change to the current model. Moreover, more and more studies focus on the active learning for imbalanced data [11]–[14], [26].

Reference [12] assumes that samples inside the boundaries are balanced, and active learning is used to choose samples in the boundaries so that the learner has a more balanced training set. Completely active learning is used to solve the imbalanced problem and the experimental results show that active learning implements fast solutions with competitive prediction performance in imbalanced classification. Meanwhile, it assumes that samples inside the boundaries are balanced, as shown in Figure 1, and active learning is used to choose samples in the boundaries so that the learner has a more balanced training set.

A co-selecting method is proposed in [26] which uses two-feature-subspace classifiers to choose the balanced samples by adjusting a sampling strategy dynamically from imbalanced sentiment data. Experiments of four domains demonstrate great potential and effectiveness of the approach for imbalanced sentiment classification.

Reference [27] analyses the effect of resampling techniques used in active learning for word sense disambiguation. It’s worth noting that the technique does not requires architecture or learning algorithms modification, which makes them very easy to use and extend to other areas. Experimental results show that under-sampling causes negative effects on active learning, but over-sampling is a relatively good choice.

Reference [14] proposes an ensemble-based active learning algorithm to tackle the medical diagnosis imbalance problem. The artificial data is created according to the distribution of the training set to make the ensemble diverse, and the random subspace re-sampling method is used to reduce the data dimension. In selecting member classifiers based on misclassification cost estimation, the minority class is assigned with higher weights for misclassification costs, while each testing sample has a variable penalty factor to induce the ensemble to correct current error. Experimental results show that compared with other ensemble methods, the proposed method has best performance, and needs less labeled samples.

III. SVM BASED OR SOLUTION

In this section, we first transform OR to an extended binary classification problem, and then give a SVM based solution.
A. OR AS AN EXTENDED BINARY CLASSIFICATION

The problems that OR handles can be described as follows: when an input vector $x$ is given, we can get a label $y$, where $x \in \mathcal{X} \subseteq \mathbb{R}^d$ and $y \in \mathcal{Y} = \{C_1, C_2, \ldots, C_K\}$, i.e., $x$ is a sample of a $d$-dimensional input space, and $y$ is one of $K$ different labels, where $C_1 < C_2 < \ldots < C_K$. Suppose that OR is a threshold model, then a $K$ ordinal classes problem has $K - 1$ ordered thresholds: $\theta_1 < \theta_2 < \ldots < \theta_{K-1}$ [28]. Thus, a sample $x$ is considered as class $C_i$ when the predictive function $h(x) = w^T x - b$ falls between $\theta_{i-1}$ and $\theta_i$, where $w \in \mathbb{R}^d$ and $b$ is an offset, and $\theta_0 = -\infty$ and $\theta_K = \infty$ are typically assumed. For example, the output of a class label $C_3$ should fall between $\theta_2$ and $\theta_3$.

OR as an extended binary classification can be formed as:

$$
\begin{align*}
x_i^k &= (x_i, e_k) \in \mathbb{R}^{d+K-1} \\
y_i^k &= 1 - 2I[y_i \leq k], \quad \forall k \in [1, K-1]
\end{align*}
$$

(1)

where $e_k \in \mathbb{R}^{K-1}$ denotes a $(K - 1)$-dimensional vector whose $k$th element is 1 and the rest of the elements are 0, the function $I[\cdot]$ is an indicator function which will return 1 if the inside condition holds, otherwise zero will be returned. Essentially, each sample in original dataset is duplicated $K - 1$ times. For example, a sample $(x, y) = (4)$ in a 6-class classification problem can be extended into $(x, e_1)$, $(x, e_2)$, $(x, e_3)$, $(x, e_4)$, $(x, e_5)$, $(x, e_6)$ by using Equation (1), where $e_1 = [1, 0, 0, 0, 0]$. Then, OR becomes a binary classification, and the weight vector:

$$
\overline{w} = (w, -\theta)
$$

(2)

The weight vector can be used to predict $y_i^k$ such that $\overline{w}x_i^k = (w, -\theta)x_i - \theta_k$. Therefore, the $\theta_k$ can be calculated by feature extension. Finally, the label of each OR sample can be predicted as:

$$
f(x_i) = 1 + \sum_{k=1}^{K-1} I[g(x_i^k) > 0]
$$

(3)

where $g(x_i^k) = \overline{w}x_i^k - b = h(x_i) - \theta_k$.

B. SVM BASED SOLUTION FOR OR

Given the original OR dataset $L = \{(x_1, y_1), \ldots, (x_L, y_L)\}$, we can extend the datasets $L$ into the corresponding dataset of binary classification $L_b = \{(x_1^1, y_1^1), \ldots, (x_L^{K-1}, y_L^{K-1})\}$, where $x_i^k = (x_i, e_k)$ and $y_i^k \in 1, -1$.

Thus, we can minimize the structural risk function by [29] and [16] as the following primal problem:

$$
\begin{align*}
\min_{w, \theta, \xi, \alpha} & \frac{1}{2} \|w\|^2 + \frac{1}{2} \|\theta\|^2 + C \sum_{i=1}^{L} \sum_{k=1}^{K-1} \xi_i^k \\
\text{s.t.} & \quad y_i^k (\phi(x_i) - \theta_k - b) \geq 1 - \xi_i^k \\
& \quad \xi_i^k \geq 0, \quad \forall i \in [1, L], \forall k \in [1, K-1]
\end{align*}
$$

(4)

where $\phi$ denotes kernel function, $C$ is a positive number and the $\xi_i^k$ denotes slack variable that allow $x_i$ to have some error at the $k$th boundary. The kernel function in Equation (4), makes the decision function become nonlinear by virtue of the kernel trick [30].

By introducing lagrangian $\alpha_i^k$ and $\mu_i^k$, the dual form of the minimization problem in Equation (4) becomes:

$$
\begin{align*}
\max_{\alpha} & \quad -\frac{1}{2} \sum_{i=1}^{L} \sum_{k=1}^{K-1} \sum_{j=1}^{L} \sum_{l=1}^{K-1} \alpha_i^k \alpha_j^l (\phi(x_i^k)^T \phi(x_j^l)) \\
& + e_i^T e_k + \sum_{i=1}^{L} \sum_{k=1}^{K-1} \alpha_i^k \\
\text{s.t.} & \quad 0 \leq \alpha_i^k \leq C, \quad \forall i \in [1, L], \forall k \in [1, K-1] \\
& \quad \sum_{i=1}^{L} \sum_{k=1}^{K-1} y_i^k \alpha_i^k = 0
\end{align*}
$$

(5)

IV. IMPROVED ACTIVE LEARNING FOR IMBALANCED OR

In this section, we first put forward a sampling strategy of active learning for OR. Then we design a balanced active learning method for imbalanced OR. Finally, we introduce two improved evaluation indicators for imbalanced OR. The general flow of the algorithm is shown in Figure 2.

A. ACTIVE LEARNING FOR OR

It is easy to collect large number of unlabeled data in many real-world applications, so that effective pool-based active learning, as shown in Algorithm 1, becomes more and more important [33]. The most critical step of pool-based active learning is how to evaluate informative samples.

The most commonly used technique in active learning focuses on selecting samples from the area of uncertainty (the area closest to the decision prediction boundary of the current model), and many exiting popular techniques are specializations of uncertainty selecting, including query-by-committee-based approaches [34]–[36]. Therefore, the most frequently used active learning strategy in SVM is to check the distance of each unlabeled sample to the hyperplane, by which the most informative samples are decided for the

Algorithm 1 AL-OR

Input: Labeled data $L$ and Unlabeled data $U$
Output: The OR model
1: for $i < N$ do
2: Learn an OR classifier using current $L$
3: Use the classifier to predict the unlabeled data $U$
4: Use sampling strategy to Equation (8) shows to select informative samples for manual annotation
5: Move the informative samples from $U$ to $L$
6: end for
7: return An OR classifier

VOLUME 8, 2020
As Equation (4) shows, we can get the parameter of ordinal regression learner [37]. We can get the parameter of ordinal regression as Equation (4) shows:

\[
w = \sum_{i=1}^{L} \sum_{k=1}^{K-1} \alpha_i^k y_i^k \phi(x_i) \]

\[
\theta = - \sum_{i=1}^{L} \sum_{k=1}^{K-1} \alpha_i^k y_i^k e_k \quad (6)
\]

Sample \(x_i\) will be extended to \(K - 1\) samples by Equation (1), therefore a sample’s confidence can be calculated as:

\[
\text{Confidence}(x_i) = \min |(w, -\theta)x_i^k - b| \quad \text{where } k \in [1, K-1] \quad (7)
\]

As Equation (7) shows, we can calculate the distance from the extended samples \(x_i^k\) to the boundary by \(|(w, -\theta)x_i^k - b|\). The minimum of \(K - 1\) distances represents the distance between the sample \(x_i\) and the decision surface of the final category. Then, we can get the most informative sample from the unlabeled data:

\[
x_i = \arg\min_{x_i} \text{Confidence}(x_i) \quad x_i \in U \quad (8)
\]

According to Equation (8), we can get our AL-OR algorithm as shown in Algorithm 1.

**B. BALANCED ACTIVE LEARNING FOR IMBALANCED OR**

The sampling strategy for OR proposed before just find the informative samples from the entire unlabeled data. This may aggravate the imbalance rate in labeled data if the informative samples are always in majority classes. To balance the samples in different classes, once a majority classes sample is chosen by Equation (8), we will choose an informative sample from the least class. Algorithm 2 illustrates the improved active learning for imbalanced OR (IAL-IOR) in detail.

The main difference of our algorithm from the original active learning is step 6-8. If the most informative sample in current unlabeled dataset \(U\) belongs to the majority classes in current labeled dataset \(L\), a sample of the minority classes in current \(L\) will be added by manually annotating the most informative sample of the same class in \(U\). After many iterations, the samples in labeled dataset \(L\) will be more balanced.

![Flow chart of IAL-IOR algorithm](image)

**Algorithm 2 IAL-IOR**

**Input:** Labeled data \(L\) and Unlabeled data \(U\)  
**Output:** The OR model

1:  \(i < N\) do  
2:  Learn an OR classifier using current \(L\)  
3:  Use the classifier to predict the unlabeled data \(U\)  
4:  Use sampling strategy as Equation (8) shows to select the most informative sample \(x_{int}\) \(\ast\)  
5:  Put the \(x_{int}\) into set \(A\)  
6:  if \(x_{int}\) belongs to a majority class then  
7:  Choose the most informative sample from the least class \(\ast\)  
8:  end if \(\ast\)  
9:  Manually annotate the samples in \(A\)  
10: Move the informative samples in \(A\) from \(U\) to \(L\) \(\ast\)  
11: end for \(\ast\)  
12: return A balanced OR classifier

**C. EVALUATION METHODS FOR IMBALANCED OR**

Making proper comparisons between classification models is a complex and unsolved challenge. This task depends not only on the understanding of errors, but also the nature of the problem itself. To avoid erroneous biases in the assessment, evaluation indicator error rate is designed to evaluate the accuracy of classification algorithms, while for OR, the mean absolute error (MAE) may be a more reasonable metric [38]:

\[
\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |f(x_i) - y_i| \quad (9)
\]

where \(N\) is the number of samples. Small errors are not as important as large errors in OR. For example, when a student’s true grade is A, being predicted as B is more acceptable than being predicted as F.

However, error rate and MAE may be deceptive in imbalanced situation. For example, for a given dataset with 10 percent of the samples belong to the minority class and 90 percent of the samples belong to the majority class, if a classifier predicts every sample to be the majority class, it will be evaluated to have an accuracy of 90 percent by error rate. It is obviously that the classifier may be ineffective to the minority class. In research area, other evaluation indicators
are used to provide a comprehensive assessment of imbalanced problems, such as Recall. It is defined as the following formula in binary classification:

\[
\text{Recall} = \frac{TP}{TP + FN} \tag{10}
\]

where \( TP \) is the number of positive samples correctly predicted and \( FN \) is the number of positive samples predicted as negative. Equation (10) can be improved for OR (also can be used in multi-classification) as:

\[
\text{Recall}_k = \frac{\sum_{i=1}^{K} \text{Recall}_i}{K} \tag{11}
\]

where \( \text{Recall}_k \) denotes the recall rate in class \( k \). There are also some well-accepted measures for imbalanced OR, the average mean absolute error (AMAE) [39] and the maximum mean absolute error (MMAE) [40]:

\[
\text{AMAE} = \frac{1}{K} \sum_{k=1}^{K} \text{MAE}^k \tag{12}
\]

\[
\text{MMAE} = \max\{\text{MAE}^k; k = 1, 2, \ldots, K\} \tag{13}
\]

where \( \text{MAE}^k \) is the MAE of class \( k \). The MAE can also be improved for OR as:

\[
\text{MAE}_{im} = \frac{\sum_{k=1}^{K} \frac{N}{N_k} \sum_{j=1}^{N_k} (f(x_i) - y_i)}{N \cdot K} \tag{14}
\]

where \( N \) is the size of dataset and \( N_k \) is the size of class \( k \). \( \frac{N}{N_k} \) denotes the weight of different classes. It is easily to prove that Equation (14) is equivalent to \( \text{MAE} \) when the dataset is balanced.

V. EXPERIMENTS

In this section, we first introduce the experimental setups, and then present our experimental results and discuss.

A. EXPERIMENTAL SETUP

1) DESIGN OF EXPERIMENTS

As an active learning algorithm, we should verify the availability of our sampling strategy, and as an algorithm for imbalance OR, we should show the generalization of our algorithm.

To show the availability of our sampling strategy, we compare the performance of our algorithms (including AL-OR and IAL-IOR) with the random sampling (randomly select query samples). To show the generalization of our algorithm, we compare the performance of our algorithms with the state-of-the-art imbalance methods (including under-sampling (US) and an over-sampling methods (SMOTE [4])) and recent proposed imbalanced methods (including SMOR [8] and SMOM [41]). The performance of all algorithms will be evaluated by general measures such as accuracy, MAE, and measures especially for imbalanced OR, such as AMAE, MMAE, Recall\(_{im} \), and MAE\(_{im} \).

2) IMPLEMENTATION DETAILS

We implement our algorithms and the other comparison experiments in MATLAB. Experiments are run on a 2.4-GHz Intel Xeon machine with 128-GB RAM. The active learning algorithms we designed are based on the OR classifier mentioned in Section 3. To be fair, all the base classifiers of the next comparison algorithms are this OR classifier. This classifier mainly involves two parameters: kernel functions and \( C \). The kernel is Gaussian kernel \( K(x_1, x_2) = \exp(-\|x_1 - x_2\|^2) \) with \( k = 0.1 \), and the \( C \) is fixed to 10 respectively.

The values for the parameters of SMOR are: \( k = 2, w = 0.25 \), and \( r = 1/4 \). The nearest neighbor parameter \( k \) in SMOTE is set to 5. The parameters of SMOM is set as follows: \( k_1 = 12, k_2 = 8, r_Th = 5/8, n_Th = 10, w_1 = 0.2, w_2 = 1/2, r_1 = 1/3 \) and \( r_2 = 0.2 \). The code of SMOR [8] and SMOM [41] is available at the website https://github.com/zhutuanfei/SMOR, and the code of our algorithms is available at the website https://github.com/gjmrookie/active-learning-for-imbalanced-OR.

For the experiments to verify the availability of our sampling strategy, we divide the original dataset into a test set and a training set at a ratio of 4:1, and the samples proportions of different classes in the training and test sets are the same as in the original dataset. In the training set, we set different size of \( L \) to ensure only a small amount of data at the beginning of training. Random sampling and AL-OR take 10 samples per generation. For the experiment to verify the generalization of our algorithm, the original dataset is also divided into a test set and a training set as above.

3) DATASETS

The datasets used in our experiments are summarized in TABLE 1. The first three benchmark datasets are used for metric regression problems. Equal-length merging are used to discrete target values into ordinal numbers by dividing the range of target values into a given number of equal-length intervals [42]. The last twelve benchmark datasets are OR datasets from the real world.

B. RESULTS AND DISCUSSIONS

Figure 3 and Figure 4 demonstrate the averages of Recall\(_{im} \) and MAE\(_{im} \) of random sampling, AL-OR and IAL-IOR on

| Table 1. The datasets used in the experiments. |
|-----------------------------|
| **Dataset**                | **Dimensionality** | **Samples** | **class distribution** | **L** |
| Housing                    | 13               | 506         | 101,100,105,81,68,33   | 50   |
| Stocks                     | 9                | 950         | 154,227,272,207,86     | 20   |
| Machine                    | 6                | 200         | 107,35,216,5,3,4,6,4,6 | 20   |
| Balance-scale              | 4                | 625         | 288,49,288             | 100  |
| Car                        | 21               | 1728        | 1210,38,69,65          | 200  |
| Cement-strength            | 8                | 998         | 196,310,244,152,96     | 100  |
| LEV                        | 4                | 1000        | 93,280,403,197,27      | 100  |
| New-thyroid                | 5                | 215         | 30,150,35              | 100  |
| SWD                        | 10               | 1000        | 32,352,359,317         | 100  |
| autism-attack              | 15               | 150         | 3,20,48,47,29,13       | 20   |
| glass                      | 9                | 213         | 69,76,17,13,9,29      | 100  |
| automobile                 | 71               | 205         | 3,22,67,54,32,27      | 20   |
| contact-lenses             | 6                | 24          | 15,5,4                | 6    |
| squash-mustard             | 52               | 52          | 24,24,4               | 9    |
| toy                        | 2                | 300         | 35,87,79,68,31        | 50   |
the test data by running 20 trials. The results show that our algorithms (AL-OR and IAL-IOR) are much better than random sampling, and our algorithms converge early. The performance of AL-OR demonstrates that active learning can deal with the class imbalanced problems. The performance of IAL-IOR is even better, however it is less effective because the samples added per generation is limited.

TABLE 2, TABLE 3 and TABLE 4 display the averages with standard deviation of accuracy, Recall, MAE, MAE_im, AMAE and MMAE of different methods for imbalance OR by running 20 trails, where the OR denotes the OR algorithm without any design for imbalanced data. The accuracy and MAE are designed to evaluate the performance of general machine learning algorithms, and the AMAE, MAME, Recall_im and MAE_im are designed to evaluate the performance of machine learning algorithms in the case of imbalanced data. The bold parts in different datasets is the best results.

The OR algorithm has a similar performance in the accuracy and MAE with the standard imbalanced algorithm (US and SMOTE) and our imbalanced algorithms. These results show

| Dataset         | OR | US | SM-MOR | SM-MOM | SM-MOR | AL-OR | LAL-IOR | OR | US | SM-MOR | SM-MOM | SM-MOR | AL-OR | LAL-IOR |
|-----------------|----|----|--------|--------|--------|-------|--------|----|----|--------|--------|--------|-------|--------|
| Balance-scale   | 0.96±0.02 | 0.91±0.02 | 0.96±0.01 | 0.94±0.04 | 0.98±0.01 | 0.97±0.01 | 0.96±0.01 | 0.94±0.04 | 0.96±0.01 | 0.94±0.04 | 0.96±0.01 | 0.94±0.04 |
| Cement-strength | 0.64±0.03 | 0.51±0.02 | 0.60±0.03 | 0.58±0.02 | 0.63±0.02 | 0.64±0.03 | 0.63±0.02 | 0.64±0.03 | 0.63±0.02 | 0.64±0.03 | 0.63±0.02 |
| LEV             | 0.58±0.03 | 0.49±0.04 | 0.57±0.02 | 0.57±0.02 | 0.58±0.03 | 0.57±0.02 | 0.57±0.02 | 0.58±0.03 | 0.57±0.02 | 0.58±0.03 | 0.57±0.02 |
| Newhyroid       | 0.63±0.05 | 0.46±0.05 | 0.63±0.05 | 0.63±0.05 | 0.64±0.05 | 0.64±0.05 | 0.64±0.05 | 0.64±0.05 | 0.64±0.05 | 0.64±0.05 | 0.64±0.05 |
| SWD             | 0.60±0.03 | 0.47±0.02 | 0.60±0.03 | 0.60±0.03 | 0.61±0.03 | 0.61±0.03 | 0.61±0.03 | 0.61±0.03 | 0.61±0.03 | 0.61±0.03 | 0.61±0.03 |
| Housing         | 0.97±0.01 | 0.88±0.01 | 0.96±0.01 | 0.96±0.01 | 0.97±0.01 | 0.97±0.01 | 0.97±0.01 | 0.97±0.01 | 0.97±0.01 | 0.97±0.01 | 0.97±0.01 |
| Stock           | 0.94±0.03 | 0.85±0.03 | 0.94±0.03 | 0.94±0.03 | 0.95±0.03 | 0.95±0.03 | 0.95±0.03 | 0.95±0.03 | 0.95±0.03 | 0.95±0.03 | 0.95±0.03 |
| Car             | 0.95±0.01 | 0.86±0.01 | 0.96±0.01 | 0.96±0.01 | 0.97±0.01 | 0.97±0.01 | 0.97±0.01 | 0.97±0.01 | 0.97±0.01 | 0.97±0.01 | 0.97±0.01 |
| auto-riskness    | 0.59±0.04 | 0.48±0.04 | 0.60±0.04 | 0.60±0.04 | 0.61±0.04 | 0.61±0.04 | 0.61±0.04 | 0.61±0.04 | 0.61±0.04 | 0.61±0.04 | 0.61±0.04 |
| glass           | 0.45±0.06 | 0.38±0.06 | 0.50±0.07 | 0.50±0.07 | 0.51±0.07 | 0.51±0.07 | 0.51±0.07 | 0.51±0.07 | 0.51±0.07 | 0.51±0.07 | 0.51±0.07 |
| automobile      | 0.65±0.18 | 0.58±0.10 | 0.70±0.20 | 0.70±0.16 | 0.73±0.21 | 0.68±0.31 | 0.70±0.23 | 0.68±0.31 | 0.70±0.23 | 0.68±0.31 | 0.70±0.23 |
| contact-lenses  | 0.92±0.03 | 0.89±0.03 | 0.92±0.03 | 0.92±0.03 | 0.93±0.02 | 0.93±0.02 | 0.93±0.02 | 0.93±0.02 | 0.93±0.02 | 0.93±0.02 | 0.93±0.02 |
| squash-unstored | 0.22±0.01 | 0.19±0.01 | 0.22±0.01 | 0.22±0.01 | 0.23±0.01 | 0.23±0.01 | 0.23±0.01 | 0.23±0.01 | 0.23±0.01 | 0.23±0.01 | 0.23±0.01 |

FIGURE 3. Comparison on Recall_im.
two findings: 1) our algorithms have a similar or even better generalization than OR; 2) the accuracy and MAE cannot reflect the accuracy on the minority classes. Our algorithms outperform the OR and the standard imbalanced algorithm on the AMAE, MMAE, Recall and MAE_{im}, which means our algorithms are better than under-sampling and SMOTE on most datasets. It also means that the completely active learning is an efficient method for imbalanced OR.

In addition, we quoted two evaluation methods: Bayes SignTest and Bayes Signrank Test [43], and applied them to our three experimental indexes (accuracy, AMAE and MAE_{im}) for verification. The Bayes Signrank results are shown in TABLE 5 and the Bayes Signtest results are shown in TABLE 6. The comparison results of each pair of algorithms have three indicators: left, rope and right, where left represents the probability that Classifier 1 is superior to Classifier 2, rope represents the probability that the algorithms are equivalent, and right is the opposite of the first case. The experimental results come from 20 trails on 15 data sets. According to the experimental

![FIGURE 4. Comparison on MAE_{im}.

TABLE 3. Average results with deviation of MAE and the new mae of US, SMOTE, SMOM, SMOR, AL-OR and IAL-IOR (keep two decimals).](image)

| Dataset          | OR   | US   | SMOTEM | SMOM  | SMOR  | AL-OR | IAL-IOR |
|------------------|------|------|--------|-------|-------|-------|---------|
| balance-scale    | 0.04 | 0.02 | 0.09   | 0.03  | 0.04  | 0.01  | 0.05    |
| Cement-strength  | 0.04 | 0.02 | 0.56   | 0.04  | 0.45  | 0.03  | 0.52    |
| LEV              | 0.04 | 0.02 | 0.56   | 0.04  | 0.45  | 0.03  | 0.52    |
| Newhybrid        | 0.04 | 0.02 | 0.56   | 0.04  | 0.45  | 0.03  | 0.52    |
| SWD              | 0.04 | 0.02 | 0.56   | 0.04  | 0.45  | 0.03  | 0.52    |
| Housing          | 0.04 | 0.02 | 0.56   | 0.04  | 0.45  | 0.03  | 0.52    |
| Machine          | 0.04 | 0.02 | 0.56   | 0.04  | 0.45  | 0.03  | 0.52    |
| Stock            | 0.04 | 0.02 | 0.56   | 0.04  | 0.45  | 0.03  | 0.52    |
| Car              | 0.04 | 0.02 | 0.56   | 0.04  | 0.45  | 0.03  | 0.52    |
| auto-riskness     | 0.04 | 0.02 | 0.56   | 0.04  | 0.45  | 0.03  | 0.52    |
| glass            | 0.04 | 0.02 | 0.56   | 0.04  | 0.45  | 0.03  | 0.52    |
| automobile       | 0.04 | 0.02 | 0.56   | 0.04  | 0.45  | 0.03  | 0.52    |
| contact-lenses   | 0.04 | 0.02 | 0.56   | 0.04  | 0.45  | 0.03  | 0.52    |
| squash-structured toy | 0.04 | 0.02 | 0.56   | 0.04  | 0.45  | 0.03  | 0.52    |
TABLE 4. Average results with deviation of AMAE and MMAE of US, SMOTE, SMOM, SMOR, AL-OR and IAL-IOR (keep two decimals).

| Dataset         | OK      | US      | SMOTES   | SMOM     | SMOR     | AL-OR    | IAL-IOR   |
|-----------------|---------|---------|----------|----------|----------|----------|-----------|
| Balance-sacle   | 0.05±0.02 | 0.08±0.04 | 0.08±0.01 | 0.03±0.02 | 0.05±0.02 | 0.04±0.03 | 0.09±0.03 | 0.07±0.03 | 0.13±0.06 | 0.04±0.02 | 0.06±0.02 | 0.07±0.03 | 0.07±0.03 |
| Census-strength | 0.01±0.03 | 0.01±0.02 | 0.02±0.01 | 0.01±0.03 | 0.01±0.02 | 0.01±0.02 | 0.01±0.01 | 0.01±0.01 | 0.02±0.02 | 0.01±0.01 | 0.02±0.01 | 0.01±0.01 | 0.01±0.01 |
| LSV             | 0.05±0.06 | 0.05±0.03 | 0.06±0.03 | 0.05±0.04 | 0.06±0.03 | 0.05±0.02 | 0.06±0.03 | 0.05±0.04 | 0.05±0.02 | 0.05±0.03 | 0.05±0.02 | 0.05±0.02 | 0.05±0.02 |
| Newthordford    | 0.07±0.02 | 0.07±0.01 | 0.07±0.02 | 0.06±0.01 | 0.07±0.02 | 0.07±0.02 | 0.07±0.02 | 0.07±0.02 | 0.07±0.02 | 0.07±0.02 | 0.07±0.02 | 0.07±0.02 | 0.07±0.02 |
| DWD             | 0.00±0.00 | 0.00±0.00 | 0.00±0.00 | 0.00±0.00 | 0.00±0.00 | 0.00±0.00 | 0.00±0.00 | 0.00±0.00 | 0.00±0.00 | 0.00±0.00 | 0.00±0.00 | 0.00±0.00 | 0.00±0.00 |
| Housing         | 0.41±0.05 | 1.31±0.28 | 0.42±0.06 | 0.42±0.05 | 0.42±0.04 | 0.46±0.03 | 0.65±0.16 | 0.99±0.16 | 0.70±0.19 | 0.41±0.16 | 0.48±0.15 | 0.69±0.12 | 0.72±0.14 |
| Machine         | 1.07±0.34 | 1.31±0.33 | 0.99±0.30 | 1.00±0.37 | 1.06±0.31 | 1.50±0.38 | 2.75±0.88 | 2.98±0.82 | 2.57±0.80 | 2.63±0.95 | 2.91±1.06 | 2.48±0.99 | 2.22±0.78 |
| Stock           | 0.11±0.03 | 0.53±0.04 | 0.10±0.02 | 0.10±0.02 | 0.11±0.02 | 0.12±0.02 | 0.21±0.08 | 0.21±0.08 | 0.22±0.08 | 0.23±0.06 | 0.24±0.07 | 0.29±0.09 | 0.33±0.04 |
| Car             | 0.12±0.03 | 0.11±0.02 | 0.12±0.04 | 0.12±0.04 | 0.10±0.03 | 0.10±0.03 | 0.09±0.02 | 0.11±0.02 | 0.12±0.09 | 0.16±0.16 | 0.36±0.17 | 0.31±0.14 | 0.31±0.06 |
| auto-stickness  | 0.70±0.14 | 1.05±0.11 | 0.76±0.11 | 0.54±0.11 | 0.55±0.11 | 0.66±0.13 | 0.72±0.15 | 1.49±0.26 | 1.18±0.39 | 1.23±0.53 | 1.35±0.32 | 1.23±0.38 | 1.46±0.28 |
| glass           | 0.78±0.16 | 0.84±0.18 | 0.57±0.16 | 0.56±0.17 | 0.62±0.19 | 0.36±0.12 | 0.73±0.14 | 0.72±0.12 | 0.47±0.10 | 0.36±0.12 | 0.36±0.12 | 0.36±0.12 | 0.36±0.12 |
| smartphone      | 0.70±0.11 | 1.32±0.09 | 0.44±0.12 | 0.44±0.11 | 0.65±0.11 | 0.71±0.08 | 0.71±0.09 | 1.44±0.34 | 2.83±0.31 | 1.36±0.31 | 0.24±0.08 | 3.95±0.35 | 1.94±0.37 |
| contact-lenses  | 0.47±0.26 | 0.64±0.33 | 0.45±0.30 | 0.32±0.36 | 0.42±0.33 | 0.53±0.12 | 0.32±0.14 | 0.83±0.41 | 0.81±0.34 | 0.71±0.40 | 0.81±0.33 | 0.65±0.42 | 0.92±0.21 | 0.99±0.23 |
| drug             | 0.01±0.02 | 0.03±0.02 | 0.08±0.02 | 0.06±0.02 | 0.07±0.02 | 0.06±0.02 | 0.04±0.02 | 0.02±0.01 | 0.04±0.02 | 0.02±0.01 | 0.02±0.01 | 0.02±0.01 | 0.02±0.01 |
| toy              | 0.00±0.00 | 0.00±0.00 | 0.00±0.00 | 0.00±0.00 | 0.00±0.00 | 0.00±0.00 | 0.00±0.00 | 0.00±0.00 | 0.00±0.00 | 0.00±0.00 | 0.00±0.00 | 0.00±0.00 | 0.00±0.00 |

VI. CONCLUSION

In this article, we put forward a sampling strategy for ordinal regression (AL-OR) and design a improved active learning for imbalanced ordinal regression (IAL-IOR). Firstly, we convert the ordinal regression problem to a binary classification problem. Secondly, we design a sampling strategy for ordinal regression and a balanced active learning based on the sampling strategy. In order to get more reasonable evaluations, we design improved recall and mean absolute error for imbalance problems. Moreover, we also proved the effectiveness of our algorithms through statistical analysis.

REFERENCES

[1] W. Chu and Z. Ghahramani, “Gaussian processes for ordinal regression,” J. Mach. Learn. Res., vol. 6, pp. 1019–1041, Jul. 2005.
[2] P. A. Gutiérrez, M. Perez-Ortiz, J. Sanchez-Monedero, F. Fernandez-Navarro, and C. Hervás-Martínez, “Ordinal regression methods: Survey and experimental study,” IEEE Trans. Knowl. Data Eng., vol. 28, no. 1, pp. 127–146, Jan. 2016.

[3] H. He and E. A. Garcia, “Learning from imbalanced data,” IEEE Trans. Knowl. Data Eng., vol. 21, no. 9, pp. 1263–1284, Jun. 2009.

[4] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic minority over-sampling technique,” J. Artif. Intell. Res., vol. 16, pp. 321–357, Jun. 2002.

[5] L. Abdi and S. Hashemi, “To combat multi-class imbalanced problems by means of over-sampling techniques,” IEEE Trans. Knowl. Data Eng., vol. 28, no. 1, pp. 238–251, Jan. 2016.

[6] M. Pérez-Ortiz, P. A. Gutiérrez, C. Hervás-Martínez, and X. Yao, “Graph-based approaches for over-sampling in the context of ordinal regression,” IEEE Trans. Knowl. Data Eng., vol. 27, no. 5, pp. 1233–1245, May 2015.

[7] I. Nekoeimehr and S. K. Lai-Yuen, “Cluster-based weighted oversampling for ordinal regression (CWOS-Ord),” Neurocomputing, vol. 218, pp. 51–60, Dec. 2016.

[8] T. Zhu, Y. Lin, Y. Liu, W. Zhang, and J. Zhang, “Minority oversampling for imbalanced ordinal regression,” Knowl.-Based Syst., vol. 166, pp. 140–155, Feb. 2019.

[9] A. Riccardi, F. Fernández-Navarro, and S. Carloni, “Cost-sensitive AdaBoost algorithm for ordinal regression based on extreme learning machine,” IEEE Trans. Cybern., vol. 44, no. 10, pp. 1908–1909, Oct. 2014.

[10] N. V. Chawla, A. Lazarevic, L. O. Hall, and K. W. Bowyer, “SMOTEBoost: Improving prediction of the minority class in boosting,” in Proc. Eur. Conf. Princ. Data Mining Knowl. Discovery, Berlin, Germany: Springer, 2003, pp. 107–119.

[11] J. Attenberg and S. Ertekin, “Class imbalance and active learning,” in Imbalanced Learning: Foundations, Algorithms, and Applications, H. He and Y. Ma, Eds. Hoboken, NJ, USA: Wiley, 2013, pp. 101–149.

[12] S. Ertekin, J. Huang, L. Bottou, and L. Giles, “Learning on the border: Active learning in imbalanced data classification,” in Proc. 16th ACM Conf. Conf. Inf. Knowl. Manage., 2007, pp. 127–136.

[13] F. Provost, “Machine learning from imbalanced data sets 101,” in Proc. AAAI Workshop Imbalanced Data Sets, vol. 68, 2000, pp. 1–3.

[14] Y. Yang and G. Ma, “Ensemble-based active learning for class imbalance problem,” J. Biomed. Sci. Eng., vol. 3, no. 10, p. 1022, 2010.

[15] N. V. Chawla, N. Japkowicz, and A. Kotcz, “Special issue on learning from imbalanced data sets,” ACM SIGKDD Explor. Newsletter, vol. 6, no. 1, pp. 1–6, 2004.

[16] L. Li and H.-T. Lin, “Ordinal regression by extended binary classification,” in Proc. Adv. Neural Inf. Process. Syst., 2007, pp. 865–872.

[17] H.-T. Lin and L. Li, “Reduction from cost-sensitive ordinal ranking to weighted binary classification,” Neural Comput., vol. 24, no. 5, pp. 1329–1367, May 2012.

[18] W. Chu and S. S. Keerthi, “New approaches to support vector ordinal regression,” in Proc. 22nd Int. Conf. Mach. Learn., 2005, pp. 145–152.

[19] W. Chu and S. S. Keerthi, “Support vector ordinal regression,” Neural Comput., vol. 19, no. 3, pp. 792–815, Mar. 2007.

[20] A. Shashua and A. Levin, “Taxonomy of large margin principle algorithms for ordinal regression problems,” in Proc. Adv. Neural Inf. Process. Syst., vol. 15, 2002, pp. 937–944.

[21] J. S. Cardoso and J. F. Costa, “Learning to classify ordinal data: The data replication method,” J. Mach. Learn. Res., vol. 8, pp. 1393–1429, Jul. 2007.

[22] Y. Shi, P. Li, H. Yuan, J. Miao, and L. Niu, “Fast kernel extreme learning machine for ordinal regression,” Knowl.-Based Syst., vol. 177, pp. 44–54, Aug. 2019.

[23] H. Wang, Y. Shi, L. Niu, and Y. Tian, “Nonparallel support vector ordinal regression,” IEEE Trans. Cybern., vol. 47, no. 10, pp. 3306–3317, Oct. 2017.

[24] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, Learning From Imbalanced Data Sets, Berlin, Germany: Springer, 2018.

[25] B. Settles, “Active learning literature survey,” Dept. Comput. Sci., Univ. Wisconsin, Madison, WI, USA, Tech. Rep. 1648, 2009.

[26] S. Li, S. Ju, G. Zhou, and X. Li, “Active learning for imbalanced sentiment classification,” in Proc. Joint Conf. Empirical Methods Natural Lang. Process. Comput. Natural Lang. Learn., 2012, pp. 139–148.

[27] J. Zhu and E. Hovy, “Active learning for word sense disambiguation with methods for addressing the class imbalance problem,” in Proc. Joint Conf. Empirical Methods Natural Lang. Process. Comput. Natural Lang. Learn. (EMNLP-CoNLL), 2007, pp. 783–790.

[28] C.-W. Seah, I. W. Tsang, and Y.-S. Ong, “Transductive ordinal regression,” IEEE Trans. Neural Netw. Learn. Syst., vol. 23, no. 7, pp. 1074–1086, Jul. 2012.

[29] C. Cortes and V. Vapnik, “Support-vector networks,” Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995.

[30] T. Hofmann, B. Schölkopf, and A. J. Smola, “Kernel methods in machine learning,” Ann. Statist., pp. 1171–1220, 2008.

[31] Z.-Q. Zeng, H.-B. Yu, H.-R. Xu, Y.-Q. Xie, and J. Gao, “Fast training support vector machines using parallel sequential minimal optimization,” in Proc. 3rd Int. Conf. Intell. Syst. Knowl. Eng., vol. 1, Nov. 2008, pp. 997–1001.

[32] P.-H. Chen, R.-E. Fan, and C.-J. Lin, “A study on SMO-type decompostion methods for support vector machines,” IEEE Trans. Neural Netw., vol. 17, no. 4, pp. 893–908, Jul. 2006.

[33] D. D. Lewis and W. A. Gale, “A sequential algorithm for training text classifiers,” in Proc. SIGIR. Berlin, Germany: Springer, 1994, pp. 3–12.

[34] Y. Freund, “Sifting informative examples from a random source,” in Proc. Adv. Neural Inf. Process. Syst., 1994, pp. 85–89.

[35] I. Dagan and S. P. Engelson, “Committee-based sampling for training probabilistic classifiers,” in Machine Learning Proceedings. Amsterdam, The Netherlands: Elsevier, 1995, pp. 150–157.

[36] Y. Freund, H. S. Seung, E. Shamir, and N. Tishby, “Information, prediction, and query by committee,” in Proc. Adv. Neural Inf. Process. Syst., 1993, pp. 483–490.

[37] S. Tong and D. Koller, “Support vector machine active learning with applications to text classification,” J. Mach. Learn. Res., vol. 2, pp. 45–66, Nov. 2001.

[38] L. Gaudette and N. Japkowicz, “Evaluation methods for ordinal classification,” in Proc. Can. Conf. Artif. Intell. Berlin, Germany: Springer, 2009, pp. 207–210.

[39] S. Baccianella, A. Esuli, and F. Sebastiani, “Evaluation measures for ordinal regression,” in Proc. 9th Int. Conf. Intell. Syst. Design Appl., 2009, pp. 283–287.

[40] M. Cruz-Ramírez, C. Hervás-Martínez, J. Sánchez-Monedero, and P. A. Gutiérrez, “Metrics to guide a multi-objective evolutionary algorithm for ordinal classification,” Neurocomputing, vol. 135, pp. 21–31, 2014.

[41] T. Zhu, Y. Lin, and Y. Liu, “Synthetic minority oversampling technique for multiclass imbalance problems,” Pattern Recognit., vol. 72, pp. 327–340, Dec. 2017.

[42] J. Sánchez-Monedero, P. A. Gutiérrez, and M. Pérez-Ortiz, “ORCA: A MATLAB/octave toolbox for ordinal regression,” J. Mach. Learn. Res., vol. 20, no. 125, pp. 1–5, 2019. [Online]. Available: http://jmlr.org/papers/v20/18-349.html

[43] A. Benavoli, G. Corani, J. Demsar, and M. Zaffalon, “Time for a change: A tutorial for comparing multiple classifiers through Bayesian analysis,” J. Mach. Learn. Res., vol. 18, no. 1, pp. 2653–2688, 2017.

***