A study on cost behaviors of binary classification measures in class-imbalanced problems

Bao-Gang Hu, Senior Member, IEEE, Wei-Ming Dong, Member, IEEE

Abstract—This work investigates into cost behaviors of binary classification measures in a background of class-imbalanced problems. Twelve performance measures are studied, such as F measure, G-means in terms of accuracy rates, and of recall and precision, balance error rate (BER), Matthews correlation coefficient (MCC), Kappa coefficient (κ), etc. A new perspective is presented for those measures by revealing their cost functions with respect to the class imbalance ratio. Basically, they are described by four types of cost functions. The functions provides a theoretical understanding why some measures are suitable for dealing with class-imbalanced problems. Based on their cost functions, we are able to conclude that G-means of accuracy rates and BER are suitable measures because they show “proper” cost behaviors in terms of “a misclassification from a small class will cause a greater cost than that from a large class”. On the contrary, F1 measure, G-means of recall and precision, MCC and κ measures do not produce such behaviors so that they are unsuitable to serve our goal in dealing with the problems properly.

Index Terms—Binary classification, class imbalance, performance, measures, cost functions

I. INTRODUCTION

Class-imbalanced problems become more common and serious in the emergence of “Big Data” processing. The initial reason is due to a fact that useful information is generally represented by a minority class. Therefore, the class-imbalance (or skewness) ratio between a majority class over a minority one can be severely large [1]. The other reason can be appeared from utilizations of “one-versus-rest” binary classification scheme for a fast processing of multiple classes [2]. Generally, the greater the number of classes, the larger the class-imbalance ratio. When most investigations in the conventional classifications apply accuracy (or error) rate as a learning criterion, this performance measure is no more appropriate in dealing with highly-imbalanced datasets [3]. In addressing class-imbalanced problems properly, cost-sensitive learning is proposed in which users are required to specify the costs according to error types [4]. At the same time, the other investigations apply “proper” measures [5], or learning criteria, which do not require information about costs. Those measures, such as F-measures, AUC and G-means, are considered to be cost-free learning [6]. Significant progresses have been reported on using those measures [7], [8], [9], [10]. Within the classification studies, however, we consider that two important issues below are still unclear theoretically, that is:

I. Why some of measures are successful in dealing with highly-imbalanced datasets?

II. What are the function behaviors of binary classification measures when the class-imbalance ratio increases?

The questions above form the motivation of this work. In principle, we can view that any classification measure implies cost information even one does not specify it explicitly. Taking a measure of error rate for example. When this measure is set as a learning criterion in binary classifications, a “zero-one” cost function is given to the criterion [11]. This function assigns an equal cost to both errors from two classes. Therefore, a new perspective from the cost behaviors is proposed in this study in order to answer the questions. Twelve measures are selected in this study on binary classifications. The rest of this brief paper is organized as follows. In Section II, we discuss two levels of evaluations in the selection of measures. Twelve measures in binary classifications are presented in Section III. Their cost functions are derived in Section IV. We demonstrate numerical examples in Section V. The conclusions are given in Section VI.

II. FUNCTION-BASED VS PERFORMANCE-BASED EVALUATIONS

This section will discuss measure selection in classifications. Fig. 1 shows two levels of evaluations, namely, function-based and performance-based evaluations. From an application viewpoint, the performance-based evaluation seems more common because it can provide a fast and overall picture among the candidate measures. One of typical investigations is shown by Ferri et al [12] on eighteen performance measures over thirty datasets. However, this kind of investigations generally produce the performance responses, not only to the measures, but also to the data and associated learning algorithms. Therefore, conclusions from the performance-based evaluation may be changed accordingly with the different datasets. Due to the coupling feature in the performance responses, one may fail to obtain the intrinsic properties of the measures.

We consider that the function-based evaluation is more fundamental in the measure selection. This evaluation will reveal function (or property) differences among the measures. Without involving any learning algorithm and noisy data, one is able to gain the intrinsic properties of measures. The properties can be various depending on the specific concerns, such as, ROC isometrics [13], statistical properties of AUC measure [14], monotonicity and error-type differentiability.
According to a specific property, one is able to see why one measure is more “proper” than the others. The findings from the function-based evaluation will be independent of the learning algorithms and datasets.

In this work, we will focus on a specific property which is not well studied in the function-based evaluation. Suppose that any binary classification measure produces cost functions in an implicit form. We consider a measure to be “proper” for processing class-imbalanced problems only when it holds a “desirable” property so that “a misclassification from a small class will cause a greater cost than that from a large class” [16]. We call this property to be a “meta measure” because it describes high-level or qualitative knowledge about a specific measure. If a binary classification measure satisfies (or does not satisfy) the meta measure, we call it “proper” (or “improper”). The examination in terms of the meta-measure enables clarification of the intrinsic causes of performance differences among classification measures.

III. TWO-CLASS MEASURES

A binary classification is considered in this work, and it is given by a confusion matrix $\mathbf{C}$ in a form of:

$$\mathbf{C} = \begin{bmatrix} TN & FP \\ FN & TP \end{bmatrix},$$

where “TN”, “TP”, “FN”, “FP”, represent “true negative”, “true positive”, “false negative”, “false positive”, respectively. Suppose $N = TN + TP + FN + FP$ to be the total number of samples in the classification. The confusion matrix can be shown in the other form:

$$\mathbf{C} = N \begin{bmatrix} CR_1 & E_1 \\ E_2 & CR_2 \end{bmatrix},$$

where $CR_1$, $CR_2$, $E_1$, and $E_2$ are the correct recognition rates and error rates [16] of Class 1 and Class 2, respectively. They are defined by:

$$CR_1 = \frac{TN}{N}, \quad CR_2 = \frac{TP}{N},$$

$$E_1 = \frac{FP}{N}, \quad E_2 = \frac{FN}{N},$$

and form the relations to the population rates by:

$$p_1 = CR_1 + E_1, \quad p_2 = CR_2 + E_2.$$  \hspace{1cm} (5)

From the non-negative terms in the confusion matrix, one can get the following constraints:

$$0 < p_1 < 1, \quad 0 < p_2 < 1, \quad p_1 + p_2 = 1$$

$$0 \leq E_1 \leq p_1, \quad 0 \leq E_2 \leq p_2.$$  \hspace{1cm} (6)

Twelve measures are investigated in this work. The first measure is the total accuracy rate:

$$A_T = \frac{TN + TP}{N} = 1 - E_1 - E_2.$$  \hspace{1cm} (7)

In this work, we will adopt the notions of four means (Fig. 2), namely, Arithmetic Mean, Geometric Mean, Quadratic Mean and Harmonic Mean, in constructions of performance measures.

Fig. 1. Schematic diagram of two levels of evaluations in the measure selections.

Fig. 2. Graphical interpretations of four means.

From the definitions of precision ($P$) and recall ($R$):

$$P = \frac{TP}{TP + FP} = \frac{CR_2}{CR_2 + E_1}, \quad R = \frac{CR_2}{p_2},$$  \hspace{1cm} (8)

one can obtain four precision-recall-based means:

$$A_{PR} = (P + R)/2.$$  \hspace{1cm} (9)

$$G_{PR} = \sqrt{PR}.$$  \hspace{1cm} (10)

$$Q_{PR} = \sqrt{\frac{P^2 + R^2}{2}}.$$  \hspace{1cm} (11)

$$H_{PR} = F_1 = \frac{2PR}{P + R}.$$  \hspace{1cm} (12)

Eq. (12) shows that $F_1$ measure is the harmonic mean of precision and recall. More definitions are given below

$$A_1 = TNR = \text{Specificity} = \frac{TN}{TN + FP} = \frac{CR_1}{p_1},$$

$$A_2 = TPR = \text{Sensitivity} = \frac{TP}{TP + FN} = \frac{CR_2}{p_2} = R,$$

where the accuracy rate of the first class ($A_1$) can also be called true negative rate (TNR) or specificity; the accuracy rate of the second class ($A_2$) called true positive rate (TPR), sensitivity or recall. In this work, we adopt the term of accuracy rate of the $i$th class ($A_i$) because it is extendable if multiple-class problems are considered. The relation between the total accuracy rate and the accuracy rate of the $i$th class is

$$A_T = p_1 \ast A_1 + p_2 \ast A_2.$$  \hspace{1cm} (14)

Then, four accuracy-rate-based means are formed as:

$$A_{A_i} = \frac{A_T - C_b}{A_1 + A_2},$$

$$G_{A_i} = \sqrt{A_1 \ast A_2}.$$  \hspace{1cm} (15)

(16)
In eq. (15), $AUC_b$ is the area under the curve (AUC) for a single classification point in the ROC curve. $AUC_b$ is also called balanced accuracy \[17\]. Three other measures are also received attentions. The balance error rate (BER) is given in a form of:

$$BER = \frac{1}{2} (\frac{E_1}{p_1} + \frac{E_2}{p_2}).$$

The Matthews correlation coefficient (MCC) is given by:

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{p_1p_2N^2(TN + FN)(TP + FP)}}$$

The Kappa coefficient ($\kappa$) is given by:

$$\kappa = \frac{Pr(a) - Pr(e)}{1 - Pr(e)},$$

$$Pr(a) = \frac{TP}{TN + TP},$$

$$Pr(e) = p_1 \frac{TN + FN}{N} + p_2 \frac{TP + FP}{N}.$$  

One needs to note that the first ten measures are given in a range of $[0, 1]$, and the last two measures, MCC and $\kappa$, are within a range of $[-1,1]$. When the four precision-recall-based measures do not take the true negative rate into account, all other measures do. Some measures above may not be well adopted in applications. We investigate them for the reason of a comparative study.

IV. COST FUNCTIONS OF MEASURES

The risk of binary classifications can be described by \[11\]:

$$Risk = \lambda_{i1} CR_1 + \lambda_{i2} E_1 + \lambda_{i2} CR_2 + \lambda_{i2} E_2.$$  

(22)

where $\lambda_{ij}$ is a cost term for the true class of a pattern to be $i$, but be misclassified as $j$. In the cost sensitive learning, the cost terms are generally assigned with constants \[4\]. However, we consider all costs in binary classifications can be described in a form of $\lambda_{ij}(\nu)$, where $\nu$ is a variable vector. The size of the vector will be discussed later. We call $\lambda_{ij}(\nu)$ “cost function”, or “equivalent cost” if it is not given explicitly. In the derivation of cost functions of the given measures, we make several assumptions below:

$A1$. The basic information to derive the cost functions is a confusion matrix in a binary classification problem without a reject option.

$A2$. The population rate of the second class $p_2$ corresponds to the minority class, that is, $p_2 < 0.5$. Hence, $p_1$ corresponds to the majority class.

$A3$. For simplifying analysis without losing generality, we assume $\lambda_{11} = \lambda_{22} = 0$. Therefore, only $\lambda_{12}(\nu)$ and $\lambda_{21}(\nu)$ are considered, but required to be non-negative ($\geq 0$) for $Risk \geq 0$.

$A4$. When the exact cost function cannot be obtained, the Taylor approximation will be applied by keeping the linear terms, and neglecting the remaining higher-order terms. The function is then denoted by $\lambda_{ij}(\nu)$.

When all the measures, except $BER$, are given in a maximum sense to the task of classifications, we need to transfer them into the minimum sense in the form of eq. (22). This transformation should not destroy the evaluation conclusions. For example, we can find an equivalent relation between the total accuracy rate and error rates:

$$\max A_T \iff \min Risk (A_T) = E_1 + E_2$$

(23)

where “$\max$” and “$\min$” are denoted “maximization” and “minimization” operators, respectively; the symbol “$\iff$” is for “equivalency”; and “$Risk$” is the transformation operator. Using the expression of eq. (22), one can immediately obtain the equivalent costs for the accuracy measure, $\lambda_{12} = \lambda_{21} = 1$.

The costs indicate constant values and no distinction between two types of errors.

However, in most cases, one fails to obtain the exact expressions on $\lambda_{ij}$. One example is given on the general form of $F$ measure by a transformation \[13\]:

$$\max F_\beta = (1 + \beta^2) \frac{RB}{p_1p_2 + RB} \iff \min Risk (F_\beta) = \frac{E_1}{p_2 - E_2} + \frac{\beta^2}{p_2 - E_2},$$

(24)

from which we can only get so called “apparent cost functions” in a form of:

$$\lambda_{12}^A = \frac{1}{p_2 - E_2}, \quad \lambda_{21}^A = \frac{\beta^2}{p_2 - E_2}.$$  

(25)

The term of “apparent” is used because the exact functions without coupling with $E_i$ may never be obtained from the given measure. Hence, the apparent cost functions in binary classifications without a reject option can be described in a general form of:

$$\lambda_{ij}^A = \lambda_{ij}^A (E_1, E_2, p_2).$$  

(26)

From the relations of eqs. (2)-(6), only three independent variables are used in describing the functions. One can apply the “class imbalance (or skewness) ratio”, $S_r = p_1/p_2$, to replace the variable $p_2$ for the analysis. The apparent cost functions provide users an analytical power in terms of a complete set of independent variables.

However, one is unable to realize unique representations of costs, either exact or apparent, on all measures, such as on $G_{AI}$ or $G_{PR}$. For overcoming this difficulty, we adopt a strategy of the first-order approximation, $A4$. Therefore, one will get a general form of $\lambda_{ij}(p_2)$ with only a single variable for binary classifications. From the relation \[4\] of min Risk $\iff \min a * Risk + b$, the constants $a$ and $b$ will be removed in the derivation of $\lambda_{ij}(p_2)$, which will not destroy the classification conclusions.

Table I lists the all measures and their cost functions or values. Only three measures exist the exact solutions on the costs. The other measures, originally given in a form of maximization sense in classifications, need to be transformed into a minimization sense. Suppose $M$ to be one of those measures, we adopt the following transformation:

$$Risk (M) = \frac{1}{M - M_{\text{min}}},$$

(27)

where $M_{\text{min}}$ is the minimum value of $M$. The transformation above is meaningful on three aspects. First, it keeps classification conclusions invariant. Second, it satisfies the assumption
of Risk ≥ 0 because M − M_{min} ≥ 0. Third, it can describe an infinitesimal risk when M = M_{min}.

From Table I, one can observe that all measures investigated in this work can be classified by four types of cost functions. Fig. 3 depicts the functions with respect to a single independent variable $p_2$. We will discuss the cost behaviors according to the function types first, and then the specific measures.

**Type I:** $\lambda_{12} = \lambda_{21} = \lambda > 0$.

The costs are positive constants with equality. The classification solutions will be independent of the constant values of costs whenever their equality relation holds. According to the meta measure, this feature suggests that the total accuracy (or error) rate measure be “improper” for dealing with class-imbalanced problems.

**Type II:** $\lambda_{12} = \lambda_{21} = \frac{1}{p_2}$.

Within this type of cost functions, both types of errors show the same cost behaviors with respect to the $p_2$. It indicates no distinctions between two types of errors, which can be considered as an “improper” feature in class-imbalanced problems.

Four measures from the precision-recall-based means demonstrate the same approximation expressions of $\lambda_{12} = \lambda_{21} = \frac{1}{p_2}$.
as the lower bounds to the exact functions (Table I). However, their approximation rates are different and are not given for the reason of their tedious expressions. The feature of the lower bounds will support the conclusions about the cost behaviors of their exact functions on: \( \lambda_{12} \) and \( \lambda_{21} \rightarrow \infty \) when \( p_2 \rightarrow 0 \).

Another important feature is that this type of functions is asymmetric and imposes more costs on the positive class than on the negative class. For example, from eq. (25), \( F_1 \) measure shows smaller costs of \( \lambda_{12} = \lambda_{21} = \frac{1}{1-E_2} \) if \( p_1 = 0 \).

Type III: \( \lambda_{12} = \frac{1}{1-p_2}, \lambda_{21} = \frac{1}{p_2} \)

This type of cost functions shows a “proper” feature in processing class-imbalanced problems, because it satisfies the meta measure. One can observe that in Fig. 3, when \( p_2 \) decreases, Type II error will receive a higher cost than Type I error. Only when two classes are equal (also called “balanced”), two types of errors will share the same values of costs. Note that the meta measure implies such requirement. Four measures from the accuracy-rate-based means and BER measure are within this type of the functions. In a study of the cost-sensitive learning, this type of the functions can be viewed a “rebalance” approach \[4, 5, 2\]. The exact solutions of the cost functions inform that \( BER \) and \( A_{A_i} (= \text{AUC}_i) \) are fully equivalent in classifications. Their equivalency can also be gained from a relation of \( BER = 1 - A_{A_i} \). The other three measures, \( G_{A_i}, Q_{A_i} \) and \( H_{A_i} \), present only approximations to the exact cost functions. Their lower bound features guarantee the cost behaviors of their exact functions on \( \lambda_{12} = 1 \) and \( \lambda_{21} \rightarrow \infty \) when \( p_2 \rightarrow 0 \). This type of functions shows symmetric cost behaviors for any class to be a minority.

Type IV: \( \lambda_{12} = \lambda_{21} = \frac{1}{p_2[1-p_2]} \)

Both \( MCC \) and \( \kappa \) measures approximate this type of cost functions. Because the same functions are given for the two types of errors, any measure within this category will be “improper” for processing class-imbalanced problems. The functions are symmetric to either class being a minority.

From the context of class-imbalanced problems, one can further aggregate the four types of cost functions within two categories, namely, “proper cost type” and “improper cost type”. We consider only Type III cost function falls in the proper cost type, and all others belong to the improper cost type. Hence, one can reach the most important finding from the category discussions about each measure. For example, when the two geometric mean measures, \( G_{A_i} \) and \( G_{PR} \), are applied in the class-imbalanced problems \[7, 8\], respectively, their intrinsic differences are not well disclosed. The present cost function study reveals their property differences about the cost response to the skewness ratio. When \( G_{A_i} \) satisfies the desirable feature on the costs, \( G_{PR} \) does not hold such feature. To our best knowledge, this theoretical finding has not been reported before.

Further finding is gained on \( F \) measure. This measure is initially proposed in the area of information retrieval \[21\] for an overall balance between precision and recall. Recently, \( F \) measure is adopted increasingly in the study of class-imbalanced learning \[27, 28, 29\]. When \( F \) measure is designed by concerning a positive (minority) class correctly without taking the negative (majority) class into account directly, it does not mean suitability in processing highly-imbalanced problems.

The cost function analysis above confirms that \( F \) measure is “improper” in either class to be a minority when its population approximates zero.

V. Numerical examples

For a better understanding of the investigated measures, we present numerical examples within two specific scenarios below.

Scenario I: Class populations are given.

Within this scenario, only two measures, \( BER \) and \( F_1 \), are considered in the investigation for the following reasons. First, we need to demonstrate the exact cost functions graphically. When \( BER \) is qualified to this aspect, \( F_1 \) can also present the exact cost values when \( E_2 \) is known in eq. (25). Second, \( BER \) and \( F_1 \) measures are representative to be “proper cost type” and “improper cost type” respectively in cost functions. They form the baselines for understanding the other measures.

In the numerical examples, we assume the following data:

\[ N = 10000, E_1 = 0.1, E_2 = \frac{p_2}{2}, \]

where \( p_2 \) is given in a vector form to present classification changes, such as from the “balanced” to the “minority” and “rare” stages, respectively.

Table II shows the solutions to the given data in (28) for both \( BER \) and \( F_1 \) measures. The data of \( BER \) and \( F_1 \) are calculated directly from the equations defined. The data of \( \lambda_{ij} \) are the exact values to each measure, respectively. One is able to confirm the correctness of \( \lambda_{ij} \) data through the following relations:

\[ BER = \frac{1}{2}(\lambda_{12} \times E_1 + \lambda_{21} \times E_2). \]

From the data in Table II, we can depict the plots of “\( \lambda_{ij} \) vs. \( p_2 \)” for \( BER \) and \( F_1 \) measures (Fig. 4). One can observe that \( F_1 \) measure is unable to distinct the costs, but produces the same costs on the given data when \( p_2 \) decreases. Although \( F_3 \) can generate different cost functions shown in (25) when \( \beta \neq 1 \), the infinity feature still remains in the both cost functions if \( p_2 = 0 \). This numerical example is sufficient to conclude that \( F_1 \), or the other measures having the similar feature, is not suitable for processing class-imbalanced problems. On the contrary, the cost plots of \( BER \) measure confirm the theoretical findings in the previous section. Among the twelve measures investigated, the measures within Type III cost functions will exhibit the “proper” cost behaviors in compatible with our intuitions for solving class-imbalanced problems.

Scenario II: Gaussian distributions are given.

This scenario is designed for a class-imbalance learning. A specific set of Gaussian distributions is exactly known,

\[ \mu_1 = -1, \mu_2 = 1, \sigma_1 = \sigma_2 = 1, \]

where \( \mu_i \) and \( \sigma_i \) are the mean and standard deviation to the \( i \)th class. Five measures, \( A_T, BER, F_1, G_{A_i} \) and \( G_{PR}, \]

\[ p_2 = [0.5, 0.1, 0.01, 0.001, 0.0001, 0.00001], \]

\[ (31) \]
are considered for a comparative study. Table III shows the optimum solutions to the given data in (31) from using the five measures, respectively. Based on the data in Table III, Fig. 5 depicts the plots of “$\lambda_{ij}$ vs. $p_2$” for the measures. When the class-imbalance ratio $p_2$ increases, the minority class (or Class 2) is mostly misclassified for measures $A_T$, $F_1$ and $G_{PR}$. The value of $\frac{E_2}{p_2} = 1.0$ suggests a complete misclassification on all samples in Class 2. In comparison, $BER$ and $G_{Ai}$ measures show a small constant value of $\frac{E_2}{p_2}$ ($= 0.1587$), which implies a good protection on the minority class. The two measures share the same solutions for the given distribution data in eq. (31). One can show that, when $\sigma_1 \neq \sigma_2$, $BER$ and $G_{Ai}$ will present the different constant values. It can be further proved that all measures in Type III will produce a constant behavior shown in Fig. 5, because their decision boundaries, $x_b$, will be independent with the population variables.

The numerical study in this scenario provides a counterexample to confirm a general conclusion that $A_T$, $F_1$ and $G_{PR}$ are “improper” measures. If “improper” measures are set as “learning targets” (or “criteria”) in highly-imbalanced problems, one may have a detrimental impact on classification quality. The numerical solutions of $BER$ and $G_{Ai}$ support the measures to be “proper” only for the given datasets. However, one is unable to reach a general conclusion on the two measures via numerical studies. This scenario study is also a function-based evaluation. If using real datasets for a performance-based evaluation, inconsistency findings may be introduced by population changes from sampling.

VI. Conclusions

This work aims at developing a theoretical insight into why some performance measures are appropriate, and some are not, for solving class-imbalanced problems. Before reviewing the existing approaches, we discuss the two levels of measure evaluations, that is, function-based evaluation and performance-based evaluation. For revealing the intrinsic properties of the measures, we consider the function-based evaluation to be necessary, and investigate one important aspect which is not well studied. This aspect is defined to be the cost behaviors of binary classification measures in terms of class-imbalance skewness ratio. We adopt a meta measure in [16] to examine each measure to be “proper” or “improper” in applications.

Twelve measures are studied and their cost functions, either exact or approximate, are derived. When four types of the cost functions are formed from the given measures, they are basically two kinds according to the meta measure. The “proper” kind includes the four means on accuracy rates and $BER$ (equivalently including $AUC_b$). The other measures, i.e. $A_T$, the four means on precision and recall (including $F_1$), $MCC$ and $\kappa$, belong to “improper” kind. Through the cost function analysis, one can observe their intrinsic equivalences or differences among the measures.

In apart from the measures investigated in this work, one can add other performance or meta measures for a systematic
study. From an application viewpoint, we understand that a final selection of measures (or learning criteria) may need to be based on an overall consideration regarding to each aspect in function-based evaluation and performance-based evaluation. The main point raised in this work confirms that “what to learn (or learning-target selection)” is the most imperative and primary issue in the study of machine learning.

ACKNOWLEDGMENT

This work is supported in part by NSFC (No. 61273196) for B.-G. Hu, and NSFC (No. 61172104) for W.-M. Dong.

REFERENCES

[1] N.V. Chawla, N. Japkowicz, and A. Kotcz, “Editorial to the special issue on learning from imbalanced data sets,” ACM SIGKDD Explorations Newsletter, vol. 6, no. 1, pp. 1-6, 2004.
[2] Z. Akata, P. Perronnin, Z. Harchaoui, and C. Schmid, “Good practice in large-scale learning for image classification,” IEEE Trans. Pattern Anal. Mach. Intell. vol. 36, no. 3, p. 570-520, 2014.
[3] H. He, and E.A. Garcia, “Learning from imbalanced data,” IEEE Trans. Knowl. Data Eng. vol. 21, no. 9, pp. 1263-1284, 2009.
[4] C. Elkan, “The foundations of cost-sensitive learning,” IJCAI, pp. 973-978, 2001.
[5] G.M. Weiss, “Mining with rarity: A unifying framework,” SIGKDD Explorations, vol. 6, no. 1, pp. 7-19, 2004.
[6] X.-W. Zhang, and B.-G. Hu, “A new strategy of cost-free learning in the class imbalance problem,” IEEE Trans. Knowl. Data Eng. accepted, 2014.
[7] M. Kubat, and S. Matwin, “Addressing the curse of imbalanced training sets: one-sided selection,” ICML, pp. 179-186, 1997.
[8] S. Daskalaki, I. Kopanas, and N. Avouris, “Evaluation of classifiers for an uneven class distribution problem,” Applied Artificial Intelligence, vol. 20, no. 5, pp. 381-417, 2006.
[9] J. Huang, and C.X. Ling, “Using AUC and accuracy in evaluating learning algorithms,” IEEE Trans. Knowl. Data Eng., vol. 17, no. 3, pp. 299-310, 2005.
[10] A. K. Menon, H. Narasimhan, S. Agarwal, and S. Chawla, “On the statistical consistence of algorithms for binary classification under class imbalance,” ICML, 2013.
[11] R.O. Duda, P.E. Hart, and D. Stork, Pattern Classification, 2nd eds., John Wiley, New York, 2001.
[12] C. Ferri, J. Hernández-Orallo, and R. Modroiu, “An experimental comparison of performance measures for classification,” Pattern Recognition Letters, vol. 30, no. 1, pp. 27-38, 2009.

TABLE III

| \( p_2 \) | \( (A_T)_{max} \) | \( x_0 \) | \( E_1/p_1 \) | \( E_2/p_2 \) | \( (EHE)_{min} \) | \( (P_1)_{max} \) | \( x_0 \) | \( E_1/p_1 \) | \( E_2/p_2 \) | \( (G_{IP})_{max} \) | \( x_0 \) | \( E_1/p_1 \) | \( E_2/p_2 \) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.050000 | 0.100000 | 0.010000 | 0.001000 | 0.000100 | 0.000010 | 0.000010 |
| 0.8413 | 0.9299 | 0.9905 | 0.9999 | 0.9999 | 0.9999 | 0.9999 |
| 0.0 | 1.0986 | 2.2976 | 3.4534 | 4.6051 | 5.7564 |
| 1.587/e-1 | 1.792/e-2 | 4.876/e-4 | 4.226/e-6 | 1.916/e-8 | 7.00/e-12 |
| 0.1587 | 0.1587 | 0.1587 | 0.1587 | 0.1587 | 0.1587 | 0.1587 |
| 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 1.587/e-1 | 1.792/e-2 | 4.876/e-4 | 4.226/e-6 | 1.916/e-8 | 7.00/e-12 |
| 0.1587 | 0.1587 | 0.1587 | 0.1587 | 0.1587 | 0.1587 | 0.1587 |
| 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 1.587/e-1 | 1.792/e-2 | 4.876/e-4 | 4.226/e-6 | 1.916/e-8 | 7.00/e-12 |
| 0.1587 | 0.1587 | 0.1587 | 0.1587 | 0.1587 | 0.1587 | 0.1587 |
| 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 1.587/e-1 | 1.792/e-2 | 4.876/e-4 | 4.226/e-6 | 1.916/e-8 | 7.00/e-12 |
| 0.016l | 0.0376 | 0.0853 | 0.8527 | 0.9004 | 0.9076 | 0.9076 |