An Analysis of Reliable Classifiers through ROC Isometrics

Stijn Vanderlooy S.VANDERLOOY@CS.UNIMAAS.NL
Ida G. Sprinkhuizen-Kuyper KUYPER@CS.UNIMAAS.NL
Evgueni N. Smirnov SMIRNOV@CS.UNIMAAS.NL
MICC-IKAT, Universiteit Maastricht, PO Box 616, 6200 MD Maastricht, The Netherlands.

Abstract

Reliable classifiers abstain from uncertain instance classifications. In this paper we extend our previous approach to construct reliable classifiers which is based on isometrics in Receiver Operator Characteristic (ROC) space. We analyze the conditions to obtain a reliable classifier with higher performance than previously possible. Our results show that the approach is generally applicable to boost performance on each class simultaneously. Moreover, the approach is able to construct a classifier with at least a desired performance per class.

1. Introduction

Machine learning classifiers were applied to various classification problems. Nevertheless, only few classifiers were employed in domains with high misclassification costs, e.g., medical diagnosis and legal practice. In these domains it is desired to have classifiers that abstain from uncertain instance classifications such that a desired level of reliability is obtained. These classifiers are called reliable classifiers.

Recently, we proposed an easy-to-visualize approach to reliable instance classification (Vanderlooy et al., 2006). Classification performance is visualized as an ROC curve and a reliable classifier is constructed by skipping the part of the curve that represents instances difficult to classify. The transformation to the ROC curve of the reliable classifier was provided. An analysis showed when and where this new curve dominates the original one. If the underlying data of both curves have approximately equal class distributions, then dominance immediately results in performance increase. However, in case of different class distributions and a performance measure that is class-distribution dependent, dominance of an ROC curve does not always guarantee an increase in performance.

In this paper we analyze for which performance metrics the approach boosts performance on each class simultaneously. We restrict to widely used metrics characterized by rotating linear isometrics (Fürnkranz & Flach, 2005). Furthermore, skew sensitive metrics are used to generalize the approach to each possible scenario of error costs and class distributions.

This paper is organized as follows. Section 2 provides terminology and notation. Section 3 gives a brief background on ROC curves. Sections 4 and 5 introduce skew sensitive evaluation and isometrics, respectively. Section 7 defines reliable classifiers and their visualization in ROC space. In Section 8 we provide our main contribution. Section 9 concludes the paper.

2. Terminology and Notation

We consider classification problems with two classes: positive (p) and negative (n). A discrete classifier is a mapping from instances to classes. Counts of true positives, false positives, true negatives and false negatives are denoted by TP, FP, TN, and FN, respectively. The number of positive instances is $P = TP + FN$. Similarly, $N = TN + FP$ is the number of negative instances.

From these counts the following statistics are derived:

$$tpr = \frac{TP}{TP + FN} \qquad tnr = \frac{TN}{TN + FP}$$

$$fpr = \frac{FP}{FP + TN} \qquad fnr = \frac{FN}{TP + FN}$$

True positive rate is denoted by tpr and true negative rate by tnr. False positive rate and false negative rate are denoted by fpr and fnr, respectively. Note that $tnr = 1 - fpr$ and $fnr = 1 - tpr$.

Most classifiers are rankers or scoring classifiers. They

Appearing in *Proceedings of the ICML 2006 workshop on ROC Analysis in Machine Learning*, Pittsburgh, USA, 2006.

output two positive values $l(x|p)$ and $l(x|n)$ that indicate the likelihood that an instance x is positive and negative, respectively. The score of an instance combines these values as follows:

$$l(x) = \frac{l(x|p)}{l(x|n)} \quad (1)$$

and can be used to rank instances from most likely positive to most likely negative (Lachiche & Flach, 2003).

3. ROC Curves

The performance of a discrete classifier can be represented by a point (fpr, tpr) in ROC space. Optimal performance is obtained in $(0, 1)$. Points $(0, 0)$ and $(1, 1)$ represent classifiers that always predict the negative and positive class, respectively. The ascending diagonal connects these points and represents the strategy of random classification.

A threshold on the score $l(x)$ transforms a scoring classifier into a discrete one. Instances with a score higher than or equal to this threshold are classified as positive. The remaining instances are classified as negative. An ROC curve shows what happens with the corresponding confusion matrix for each possible threshold (Fawcett, 2003). The convex hull of the ROC curve (ROCCH) removes concavities.

Theorem 1 *For any point (fpr, tpr) on an ROCCH a classifier can be constructed that has the performance represented by that point.*

Provost and Fawcett (2001) prove this theorem. For simplicity of presentation, in the following we will assume that ROC curves are convex and all points can be obtained by a threshold.

4. Skew Sensitive Evaluation

The metrics tpr, fpr, tnr, and fnr evaluate performance on a single class. This follows from the confusion matrix since values are used from a single column. In most cases a metric is desired that indicates performance on both classes simultaneously. Unfortunately, such metric assumes that the class distribution of the application domain is known and used in the test set.

Accuracy is a well-known example. Provost et al. (1998) showed that classifier selection with this metric has two severe shortcomings with regard to class and error costs distributions.

To overcome these problems, Flach (2003) considers class and error costs distributions as a parameter of performance metrics. Evaluation with these metrics is called skew sensitive evaluation. The parameter is called the skew ratio and expresses the relative importance of negative versus positive class:

$$c = \frac{c(p,n)}{c(n,p)} \frac{P(n)}{P(p)} \quad (2)$$

Here, $c(p, n)$ and $c(n, p)$ denote the costs of a false positive and false negative, respectively[1]. The probabilities of a positive and negative instance are denoted by $P(p) = \frac{P}{P+N}$ and $P(n) = \frac{N}{P+N}$, respectively. The class ratio is then $\frac{P(n)}{P(p)} = \frac{N}{P}$.

From Eq. 2 it is clear that we can cover all possible scenarios of class and cost distributions by a single value of c used as parameter in the performance metric. If $c < 1$ ($c > 1$), then the positive (negative) class is most important.

In the following we assume without restriction that c is the ratio of negative to positive instances in the test set, i.e., $c = \frac{N}{P}$. The reader should keep in mind that our results are also valid for $c = \frac{c(p,n)}{c(n,p)} \frac{N}{P}$.

5. ROC Isometrics

Classifier performance is evaluated on both classes. We define a positive (negative) performance metric as a metric that measures performance on the positive (negative) classifications. The skew sensitive metrics used in this paper are summarized in Table 1. An explanation of these metrics follows.

ROC isometrics are collections of points in ROC space with the same value for a performance metric. Flach (2003) and Fürnkranz and Flach (2005) investigate isometrics to understand metrics. However, isometrics can be used for the task of classifier selection and to construct reliable classifiers (see Section 6).

Table 1 also shows the isometrics for the performance metrics. They are obtained by fixing the performance metric and rewriting its equation to that of a line in ROC space. Varying the value of the metric results in linear lines that rotate around a single point in which the metric is undefined.

5.1. Precision

Positive precision, $prec_p^c$, is defined as the proportion of true positives to the total number of positive classifications. The isometrics are linear lines that rotate around the origin $(0, 0)$.

[1]Benefits of true positives and true negatives are incorporated by adding them to the corresponding errors. This operation normalizes the cost matrix such that the two values on the main diagonal are zero.

Table 1. Performance metrics and corresponding isometrics defined in terms of *fpr*, *tpr*, $c = \frac{N}{P}$, $\alpha \in \mathbb{R}^+$, and $\hat{m} = \frac{m}{P+N}$.

Metric	Indicator	Formula	Isometric
Pos. precision	$prec_p^c$	$\frac{tpr}{tpr+c\,fpr}$	$tpr = \frac{prec_p^c}{1-prec_p^c}\, c\, fpr$
Neg. precision	$prec_n^c$	$\frac{tnr}{tnr+\frac{1}{c}fnr}$	$tpr = \frac{1-prec_n^c}{prec_n^c}\, c\, fpr + 1 - \frac{1-prec_n^c}{prec_n^c}\, c$
Pos. F-measure	$F_p^{c,\alpha}$	$\frac{(1+\alpha^2)tpr}{\alpha^2+tpr+c\,fpr}$	$tpr = \frac{F_p^{c,\alpha}}{1+\alpha^2-F_p^{c,\alpha}} c\, fpr + \frac{\alpha^2\, F_p^{c,\alpha}}{1+\alpha^2-F_p^{c,\alpha}}$
Neg. F-measure	$F_n^{c,\alpha}$	$\frac{(1+\alpha^2)tnr}{\alpha^2+tnr+\frac{1}{c}fnr}$	$tpr = \frac{1+\alpha^2-F_n^{c,\alpha}}{F_n^{c,\alpha}} c\, fpr + 1 + \frac{(1+\alpha^2)(F_n^{c,\alpha}-1)}{F_n^{c,\alpha}} c$
Pos. gm-estimate	$gm_p^{c,\hat{m}}$	$\frac{tpr+\hat{m}}{tpr+c\,fpr+\hat{m}(1+c)}$	$tpr = \frac{gm_p^{c,\hat{m}}}{1-gm_p^{c,\hat{m}}} c\, fpr + \frac{\hat{m}\left(gm_p^{c,\hat{m}}(1+c)-1\right)}{1-gm_p^{c,\hat{m}}}$
Neg. gm-estimate	$gm_n^{c,\hat{m}}$	$\frac{tnr+\hat{m}}{tnr+\frac{1}{c}fnr+\hat{m}\frac{1+c}{c}}$	$tpr = \frac{1-gm_n^{c,\hat{m}}}{gm_n^{c,\hat{m}}} c\, fpr + 1 - \frac{1-gm_n^{c,\hat{m}}}{gm_n^{c,\hat{m}}} c + \frac{\hat{m}\left(gm_n^{c,\hat{m}}(1+c)-c\right)}{gm_n^{c,\hat{m}}}$

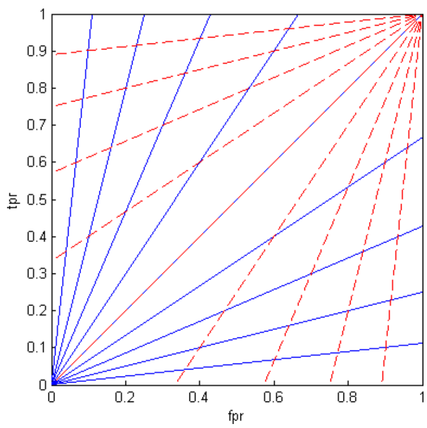


Figure 1. Precision isometrics in ROC space: solid lines are $prec_p^1$-isometrics and dashed lines are $prec_n^1$-isometrics.

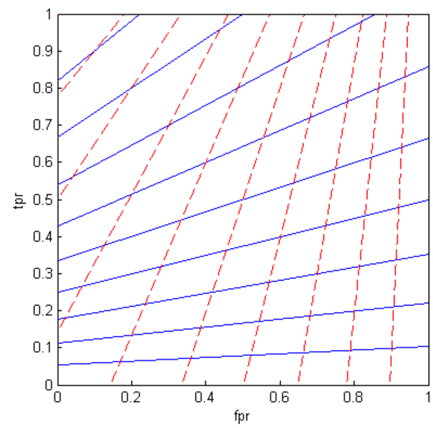


Figure 2. F-measure isometrics in ROC space: solid lines are $F_p^{1,1}$-isometrics and dashed lines are $F_n^{1,1}$-isometrics.

The case of negative precision, $prec_n^c$, is similar. Corresponding isometrics rotate around point $(1,1)$. Figure 1 shows $prec_p^c$-isometrics and $prec_n^c$-isometrics for $c = 1$. In this and subsequent figures the value of the performance metric is varied from 0.1 to 0.9.

5.2. F-measure

Positive precision is maximized when all positive classifications are correct. To know if $prec_p^c$ uses enough positive instances to be considered as reliable, it is combined with *tpr*. Note that $prec_p^c$ and *tpr* are antagonistic, i.e., if $prec_p^c$ goes up, then *tpr* usually goes down (and vice versa).

Rijsbergen (1979) introduced the positive F-measure for the trade-off between these metrics:

$$F_p^{c,\alpha} = \frac{(1+\alpha^2)\, prec_p^c\, tpr}{\alpha^2\, prec_p^c + tpr} = \frac{(1+\alpha^2)\, tpr}{\alpha^2 + tpr + c\, fpr} \quad (3)$$

where the parameter α indicates the importance given to $prec_p^c$ relative to *tpr*. If $\alpha < 1$ ($\alpha > 1$) then *tpr* is less (more) important than $prec_p^c$. If $\alpha = 1$, then both terms are equally important.

The isometrics of $F_p^{c,\alpha}$ are linear lines rotating around $(-\frac{\alpha^2}{c}, 0)$. Therefore, they can be seen as a shifted version of the positive precision isometrics. The larger c and/or the smaller α, the smaller the difference with $prec_p^c$-isometrics.

Similar to $F_p^{c,\alpha}$ the negative F-measure is a metric for the trade-off between $prec_n^c$ and *tnr*. Isometrics are a shifted version of the $prec_n^c$-isometrics and rotate around $(1, 1 + \alpha^2 c)$. Figure 2 shows $F_p^{c,\alpha}$-isometrics and $F_n^{c,\alpha}$-isometrics for $c = 1$ and $\alpha = 1$ in the relevant region $(0,1) \times (0,1)$ of ROC space.

5.3. Generalized m-estimate

The m-estimate computes a precision estimate assuming that m instances are a priori classified. One of the main reasons why it is favored over precision is that it is less sensitive to noise and more effective in avoiding overfitting (Fürnkranz & Flach, 2005; Lavrac & Dzeroski, 1994, Chapters 8-10). This is especially true if the metric is used for the minority class when the class distribution is very skewed.

The positive m-estimate assumes that m instances are a priori classified as positive. These instances are distributed according to the class distribution in the training set:

$$gm_p^{c,m} = \frac{TP + m\frac{P}{P+N}}{TP + FP + m} \tag{4}$$

or equivalently:

$$gm_p^{c,m} = \frac{tpr + \frac{m}{P+N}}{tpr + c\,fpr + \frac{m}{P}} \tag{5}$$

To eliminate absolute numbers P and N we define $\hat{m} = \frac{m}{P+N}$ and obtain the formula in Table 1. Fürnkranz and Flach (2005) call this metric the positive gm-estimate (generalized m-estimate) since $\hat{m}$ defines the rotation point of the isometrics (see below)[2].

The isometrics of the $gm_p^{c,\hat{m}}$-estimate rotate around $(-\hat{m}, -\hat{m})$. If $\hat{m} = 0$, then we obtain $prec_p^c$-isometrics. For $\hat{m} \to \infty$ the performance metric converges to $\frac{1}{1+c} = P(p)$ and the corresponding isometric is the ascending diagonal.

The case of the negative gm-estimate is similar. The rotation point of the isometrics is $(1+\hat{m}, 1+\hat{m})$. Figure 3 shows $gm_p^{c,\hat{m}}$-isometrics and $gm_n^{c,\hat{m}}$-isometrics for $c = 1$ and $\hat{m} = 0.1$.

For simplicity of presentation, in the following the isometric of a positive (negative) performance metric is simply called a positive (negative) isometric.

6. Classifier Design through Isometrics

In Vanderlooy et al. (2006) we used precision isometrics as a tool to design classifiers. We generalize this approach to include all isometrics defined in Section 5.

For specific skew ratio, a positive isometric is build with a desired positive performance. By definition, the intersection point (fpr_a, tpr_a) with an ROCCH represents a classifier with this performance. Similarly, the intersection point (fpr_b, tpr_b) of a negative isometric and the ROCCH represents a classifier with negative performance defined by that isometric. If we

[2]The gm-estimate of Fürnkranz and Flach (2005) is more general than ours since they also vary $a = \frac{1}{P+N}$ in Eq. 5.

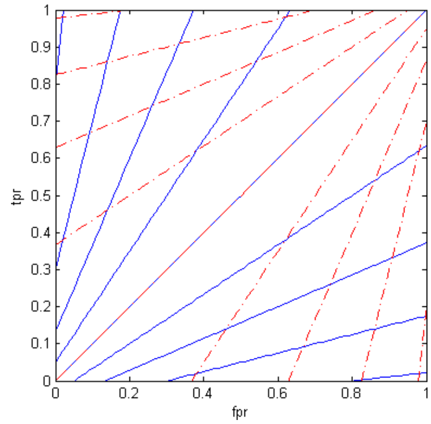
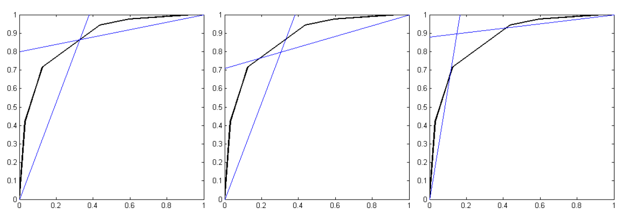


Figure 3. Generalized m-estimate isometrics in ROC space: solid lines are $gm_p^{1,0.1}$-isometrics and dashed lines are $gm_n^{1,0.1}$-isometrics.



(a) (b) (c)

Figure 4. Location of intersection between a positive and negative isometric: (a) Case 1, (b) Case 2, and (c) Case 3.

assume that the positive and negative isometrics intersect each other in the relevant region of ROC space, then three cases can be distinguished to construct the desired classifier (see Figure 4).

Case 1: the isometrics intersect on the ROCCH
The discrete classifier corresponding to this point has the performance defined by both isometrics. Theorem 1 guarantees that we can construct it. Therefore, the isometrics provide an approach to construct a classifier with a desired performance per class.

Case 2: the isometrics intersect below the ROCCH
This classifier can also be constructed. However, the classifiers corresponding to any point on the ROCCH between (fpr_b, tpr_b) and (fpr_a, tpr_a) have better performance.

Case 3: the isometrics intersect above the ROCCH
There is no classifier with the desired performances. To increase performance instances between (fpr_a, tpr_a) and (fpr_b, tpr_b) are not classified. In case

of more than one intersection point for the positive (negative) isometric and the ROCCH, the intersection point with highest tpr (lowest fpr) is chosen such that $fpr_a < fpr_b$. Then, the number of unclassified instances is minimized. The resulting classifier is called a reliable classifier.

7. Reliable Instance Classification

A scoring classifier is almost never optimal: there exists negative instances with higher score than some positive instances. A reliable classifier abstains from these uncertain instance classifications. It simulates the behavior of a human expert in fields with high error costs. For example, in medical diagnosis an expert does not state a possibly incorrect diagnosis but she says "I do not know" and performs more tests.

Similar to Ferri and Hernández-Orallo (2004), we define a reliable classifier as a filtering mechanism with two thresholds $a > b$. An instance x is classified as positive if $l(x) \geq a$. If $l(x) \leq b$, then x is classified as negative. Otherwise, the instance is left unclassified. Unclassified instances can be rejected, passed to a human, or to another classifier (Ferri et al., 2004). Pietraszek (2005) chooses a and b to minimize expected cost, also considering the abstention costs. Here, we focus on performance on the classified instances.

Counts of unclassified positives and unclassified negatives are denoted by UP and UN, respectively. Unclassified positive rate and unclassified negative rate are then defined as follows:

$$upr = \frac{UP}{TP+FN+UP} \quad (6)$$

$$unr = \frac{UN}{FP+TN+UN} \quad (7)$$

We define thresholds a and b to correspond with points (fpr_a, tpr_a) and (fpr_b, tpr_b), respectively. The ROC curve of the reliable classifier is obtained by skipping the part between (fpr_a, tpr_a) and (fpr_b, tpr_b). By definition we have:

$$upr = tpr_b - tpr_a \quad (8)$$

$$unr = fpr_b - fpr_a \quad (9)$$

The transformation from the original ROC curve to that of the reliable classifier is given in Theorem 2.

Theorem 2 *If the part between points (fpr_a, tpr_a) and (fpr_b, tpr_b) of an ROC curve is skipped with $0 < upr < 1$ and $0 < unr < 1$, then points (fpr_x, tpr_x) on this curve between $(0, 0)$ and (fpr_a, tpr_a) are transformed into points (fpr'_x, tpr'_x) such that:*

$$fpr'_x = \frac{fpr_x}{1 - unr}\ ,\ tpr'_x = \frac{tpr_x}{1 - upr} \quad (10)$$

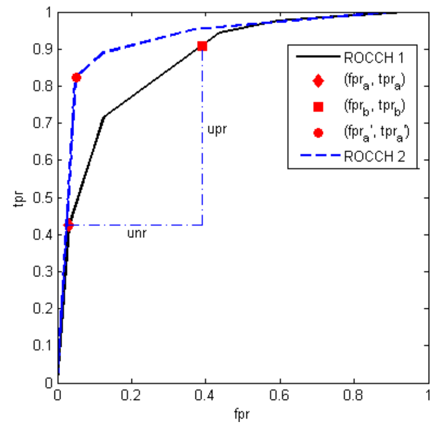


Figure 5. ROCCH 2 is obtained by not covering the part between (fpr_a, tpr_a) and (fpr_b, tpr_b) of ROCCH 1. The length of the horizontal (vertical) line below ROCCH 1 equals unr (upr).

Points (fpr_x, tpr_x) between (fpr_b, tpr_b) and $(1, 1)$ are transformed into points (fpr'_x, tpr'_x) such that:

$$fpr'_x = 1 - \frac{1 - fpr_x}{1 - unr}\ ,\ tpr'_x = 1 - \frac{1 - tpr_x}{1 - upr} \quad (11)$$

The proof is in Vanderlooy et al. (2006). Note that the transformations of (fpr_a, tpr_a) and (fpr_b, tpr_b) are the same point on the new ROC curve. Figure 5 shows an example of a transformation. The intersection points are obtained with precision isometrics for $c = 1$, $prec_p^c = 0.93$, and $prec_n^c = 0.87$.

Theorem 3 *If the original ROC curve is convex, then the ROC curve obtained by not considering the points between (fpr_a, tpr_a) and (fpr_b, tpr_b) is also convex.*

We proved this theorem in Vanderlooy et al. (2006). There, we also analyzed when and where the original ROCCH is dominated by that of the reliable classifier. Note that the underlying data of both ROCCHs can have a different class distribution when $upr \neq unr$. For skew insensitive metrics or when $upr \approx unr$, dominance of a ROCCH will immediately result in performance increase. In the next Section 8 we analyze when the skew sensitive performance metrics in Table 1 can be boosted by abstention.

8. Effect on Performance

We defined (fpr_a, tpr_a) and (fpr_b, tpr_b) as intersection points of an ROCCH and positive and negative isometric, respectively. The type of isometrics defines the effect on the performance of the reliable classifier corresponding to (fpr'_a, tpr'_a) as defined in Theorem 2.

8.1. Precision

Theorem 4 provides an easy and computationally efficient approach to construct a classifier with a desired precision per class.

Theorem 4 *If points (fpr_a, tpr_a) and (fpr_b, tpr_b) are defined by an $prec_p^c$-isometric and $prec_n^c$-isometric respectively, then the point (fpr'_a, tpr'_a) has the precisions of both isometrics.*

The proof of this theorem and also of following theorems are included in the appendix. Since isometrics of skew sensitive performance metrics are used, the approach does not commit to costs and class distributions[3]. Thus, when the application domain changes a new reliable classifier can be constructed from the original ROC curve only.

Theorem 4 together with the next Theorem 5 provides an approach to construct a classifier with desired accuracy. This approach overcomes the problems with accuracy explained in Section 4. From the proof it follows that if the precisions are not equal, then the accuracy is bounded by the smallest and largest precision.

Theorem 5 *If point (fpr'_a, tpr'_a) has $prec_p^c = prec_n^c$, then the accuracy in this point equals the precisions.*

8.2. F-measure

Theorem 6 shows that also the F-measure can be boosted on both classes if a part of an ROC curve is not covered. In this case, the resulting classifier has higher performance than defined by both isometrics. Figure 6 gives an example where positive (negative) performance is increased with approximately 5% (10%).

Theorem 6 *If points (fpr_a, tpr_a) and (fpr_b, tpr_b) are defined by an $F_p^{c,\alpha}$-isometric and $F_n^{c,\alpha}$-isometric respectively, then the point (fpr'_a, tpr'_a) has higher performance than defined by both isometrics.*

8.3. Generalized m-estimate

To analyze the effect of abstention on the gm-estimate, we can consider the number of a priori classified instances m to be fixed or the parameter $\hat{m}$ to be fixed.

Consider the case when m is not changed after transformation. In this case upr and unr can change the distribution of a priori instances over the classes. If $upr < unr$, then the distribution of these instances in

[3]Remember that, although our proofs use the simplest case $c = \frac{N}{P}$, the results are also valid for $c = \frac{c(p,n)}{c(n,p)} \frac{N}{P}$.

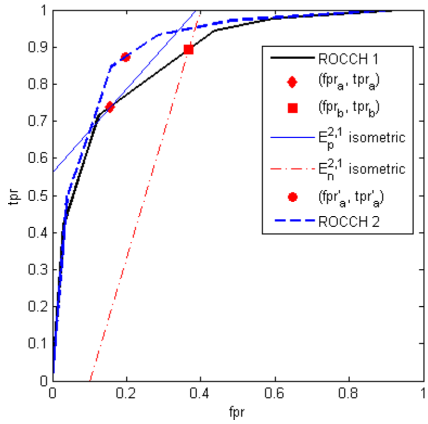


Figure 6. Designing with F-measure isometrics: $F_p^{2,1} = 0.72$ in (fpr_a, tpr_a) and $F_n^{2,1} = 0.75$ in (fpr_b, tpr_b). The reliable classifier represented by (fpr'_a, tpr'_a) has $F_p^{1.84,1} = 0.7693$ and $F_n^{1.84,1} = 0.8597$. The abstention is represented by $upr = 0.1541$ and $unr = 0.2116$.

the positive gm-estimate moves to the true positives resulting in higher performance. For the negative gm-estimate, the distribution moves to the false negatives resulting in lower performance. The case of $upr > unr$ is the other way around. Therefore, an increase in performance in both classes is only possible iff $upr = unr$.

For the case when $\hat{m}$ is not changed after transformation, a similar reasoning results in improvement of the positive gm-estimate if $upr \leq unr$ and $tpr_a \geq fpr_a$. The latter condition holds for all points on the ROCCH. Similarly, improvement in the negative gm-estimate occurs if $upr \geq unr$ and $tpr_b \geq fpr_b$. Thus, we find the following theorems for the gm-estimate.

Theorem 7 *If point (fpr_a, tpr_a) is defined by an $gm_p^{c,\hat{m}}$-estimate isometric with $m > 0$ and if $upr \leq unr$, then the point (fpr'_a, tpr'_a) has at least the positive performance defined by that isometric.*

Theorem 8 *If point (fpr_b, tpr_b) is defined by an $gm_n^{c,\hat{m}}$-estimate isometric with $m > 0$ and if $upr \geq unr$, then the point (fpr'_a, tpr'_a) has at least the negative performance defined by that isometric.*

Corollary 1 *If points (fpr_a, tpr_a) and (fpr_b, tpr_b) are defined by an $gm_p^{c,\hat{m}}$-estimate isometric and $gm_n^{c,\hat{m}}$-estimate isometric respectively with $m > 0$ and if $upr = unr$, then the point (fpr'_a, tpr'_a) has at least the gm-estimates of both isometrics.*

We suggest to use the gm-estimate for the minority class only and to use a normal precision for the majority class. From Theorems 7 and 8, if the minority

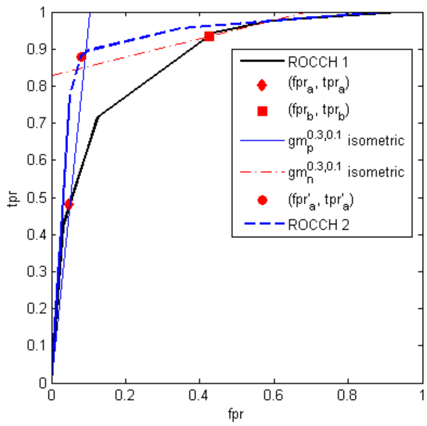


Figure 7. Designing with precision and gm-estimate isometrics: $prec_p^{0.3} = 0.97$ in (fpr_a, tpr_a) and $gm_n^{0.3,0.1} = 0.55$ in (fpr_b, tpr_b). The reliable classifier represented by (fpr'_a, tpr'_a) has $prec_p^{0.3} = 0.97$ and $gm_n^{0.34,0.18} = 0.5584$. The abstention is represented by $upr = 0.4549$ and $unr = 0.3763$.

class is the positive (negative) class, then we need an abstention characterized by $upr \leq unr$ ($upr \geq unr$).

Figure 7 shows an example with fixed m and the negative class as minority class. Therefore, we want that the $gm_n^{c,\hat{m}}$-estimate isometric covers a large part in ROC space and consequently the condition $upr \geq unr$ is easily satisfied.

9. Conclusions

A reliable classifier abstains from uncertain instance classifications. Benefits are significant in application domains with high error costs, e.g., medical diagnosis and legal practice. A classifier is transformed into a reliable one by not covering a part of its ROC curve. This part is defined by two isometrics indicating performance on a different class.

In case of a classifier and corresponding reliable classifier, dominance of an ROC curve immediately represents an increase in performance if the underlying data of both curves have approximately equal class distributions. Since this assumption is too strong, we analyzed when performance can be boosted by abstention.

We showed how to construct a (reliable) classifier with a desired precision per class. We did the same for accuracy. For the F-measure a classifier is obtained with at least the desired performance per class. To prevent a possible performance decrease with the gm-estimate, we propose to use it for the minority class and to use a normal precision for the majority class.

We may conclude that the proposed approach is able to boost performance on each class simultaneously. Benefits of the approach are numerous: it guarantees a classifier with an acceptable performance in domains with high error costs, it is efficient in terms of time and space, classifier independent, and it incorporates changing error costs and class distributions easily.

Acknowledgments

We thank the reviewers for useful comments and suggestions. This work was partially supported by the Dutch Organization for Scientific Research (NWO), grant nr: 634.000.435.

References

Fawcett, T. (2003). *ROC graphs: Notes and practical considerations for researchers* (Technical Report HPL-2003-4). HP Laboratories.

Ferri, C., Flach, P., & Hernández-Orallo, J. (2004). Delegating classifiers. *Proceedings of the 1st International Workshop on ROC Analysis in Artificial Intelligence (ROCAI-2004)* (pp. 37–44).

Ferri, C., & Hernández-Orallo, J. (2004). Cautious classifiers. *Proceedings of the 1st International Workshop on ROC Analysis in Artificial Intelligence (ROCAI-2004)* (pp. 27–36).

Flach, P. (2003). The geometry of ROC space: Understanding machine learning metrics through ROC isometrics. *Proceedings of the 20th International Conference on Machine Learning (ICML-2003)* (pp. 194–201).

Fürnkranz, J., & Flach, P. (2005). Roc 'n' rule learning – towards a better understanding of covering algorithms. *Machine Learning*, *58*, 39–77.

Lachiche, N., & Flach, P. (2003). Improving accuracy and cost of two-class and multi-class probabilistic classifiers using ROC curves. *Proceedings of the 20th International Conference on Machine Learning (ICML-2003)* (pp. 416–423).

Lavrac, N., & Dzeroski, S. (1994). *Inductive logic programming: Techniques and applications.* Ellis Horwood, New York.

Pietraszek, T. (2005). Optimizing abstaining classifiers using ROC analysis. *Proceedings of the 22th International Conference on Machine Learning (ICML-2005)* (pp. 665–672).

Provost, F., & Fawcett, T. (2001). Robust classification for imprecise environments. *Machine Learning*, *42*, 203–231.

Provost, F., Fawcett, T., & Kohavi, R. (1998). The case against accuracy estimation for comparing induction algorithms. *Proceedings of the 15th International Conference on Machine Learning (ICML-1998)* (pp. 43–48).

Rijsbergen, C. V. (1979). *Information retrieval.* Department of Computer Science, University of Glasgow. 2nd edition.

Vanderlooy, S., Sprinkhuizen-Kuyper, I., & Smirnov, E. (2006). Reliable classifiers in ROC space. *Proceedings of the 15th Annual Machine Learning Conference of Belgium and the Netherlands (BENELEARN-2006)* (pp. 113–120).

A. Proofs

Proof of Theorem 4
The positive precisions in (fpr_a, tpr_a) and (fpr'_a, tpr'_a) are defined as follows:

$$prec_p^c(fpr_a, tpr_a) = \frac{tpr_a}{tpr_a + c\, fpr_a} \quad (12)$$

$$prec_p^{c'}(fpr'_a, tpr'_a) = \frac{tpr'_a}{tpr'_a + c'\, fpr'_a} \quad (13)$$

with $c' = c\frac{1-unr}{1-upr}$. Substitution of Eq. 10 in Eq. 13 results in Eq 12. In a similar way, Eq. 11 is used to show that the negative precisions in (fpr_b, tpr_b) and (fpr'_b, tpr'_b) are the same. The theorem follows since $(fpr'_b, tpr'_b) = (fpr'_a, tpr'_a)$. □

Proof of Theorem 5
Since the positive precision and negative precision in (fpr'_a, tpr'_a) are equal, we can write:

$$tpr'_a = a\left(tpr'_a + c'\, fpr'_a\right) \quad (14)$$

$$tnr'_a = a\left(tnr'_a + \frac{1}{c'} fnr'_a\right) \quad (15)$$

with $a = prec_p^{c'} = prec_n^{c'}$. It follows that:

$$tpr'_a + c'\, tnr'_a = a\left(tpr'_a + c'\, fpr'_a + c'\, tnr'_a + fnr'_a\right) \quad (16)$$

or equivalently:

$$a = \frac{tpr'_a + c'\, tnr'_a}{tpr'_a + c'\, fpr'_a + c'\, tnr'_a + fnr'_a} \quad (17)$$

and this is the accuracy with skew ratio c'. □

Proof of Theorem 6
The positive F-measures in (fpr_a, tpr_a) and (fpr'_a, tpr'_a) are defined as follows:

$$F_p^{c,\alpha}(fpr_a, tpr_a) = \frac{\left(1+\alpha^2\right) tpr_a}{\alpha^2 + tpr_a + c\, fpr_a} \quad (18)$$

$$F_p^{c',\alpha}\left(fpr'_a, tpr'_a\right) = \frac{\left(1+\alpha^2\right) tpr'_a}{\alpha^2 + tpr'_a + c'\, fpr'_a} \quad (19)$$

Using Eq. 10 and $c' = c\frac{1-unr}{1-upr}$, the right-hand side of Eq. 19 becomes:

$$\frac{\left(1+\alpha^2\right) tpr_a}{\alpha^2(1-upr) + tpr_a + c\, fpr_a} \quad (20)$$

It follows that $F_p^{c',\alpha}\left(fpr'_a, tpr'_a\right) > F_p^{c,\alpha}(fpr_a, tpr_a)$ since $0 < upr < 1$. The case of the negative F-measure is similar. □

Proof of Theorem 7
The positive gm-estimates in (fpr_a, tpr_a) and (fpr'_a, tpr'_a) are defined as follows:

$$gm_p^{c,\hat{m}}(fpr_a, tpr_a) = \frac{tpr+\hat{m}}{tpr+c\, fpr+\hat{m}(1+c)} \quad (21)$$

$$gm_p^{c',\hat{m}'}\left(fpr'_a, tpr'_a\right) = \frac{tpr'+\hat{m}'}{tpr'+c'\, fpr'+\hat{m}'(1+c')} \quad (22)$$

with $\hat{m} = \frac{m}{P+N}$, and $c' = c\frac{1-unr}{1-upr}$.

Case 1: m is not changed after transformation
In this case we can write $\hat{m}' = \frac{m}{P(1-upr)+N(1-unr)}$. Substitution of Eq. 10 in Eq. 22 results in the following right-hand side:

$$\frac{tpr + m\frac{1-upr}{P(1-upr)+N(1-unr)}}{tpr + c\, fpr + \hat{m}(1+c)} \quad (23)$$

Clearly, $gm_p^{c',\hat{m}'}\left(fpr'_a, tpr'_a\right) \geq gm_p^{c,\hat{m}}(fpr_a, tpr_a)$ iff:

$$\frac{1-upr}{P(1-upr)+N(1-unr)} \geq \frac{1}{P+N} \quad (24)$$

This holds iff $upr \leq unr$.

Case 2: $\hat{m}$ is not changed after transformation
Substitution of Eq. 10 in Eq. 22 with fixed $\hat{m}$ results in the following right-hand side:

$$\frac{tpr + \hat{m}(1-upr)}{tpr + c\, fpr + \hat{m}(1-upr+c(1-unr))} \quad (25)$$

Straightforward computation results in $gm_p^{c',\hat{m}}\left(fpr'_a, tpr'_a\right) \geq gm_p^{c,\hat{m}}(fpr_a, tpr_a)$ iff:

$$\hat{m}(unr - upr) + (tpr_a\, unr - fpr_a\, upr) \geq 0 \quad (26)$$

This holds if $upr \leq unr$ and $tpr_a \geq fpr_a$.□

Proof of Theorem 8
The proof is similar to that of Theorem 7. □