RECEIVER OPERATING CHARACTERISTIC CURVES AND OPTIMAL BAYESIAN OPERATING POINTS

Tapas Kanungo and Robert M. Haralick

Intelligent Systems Laboratory
Department of Electrical Engineering
University of Washington
Seattle, WA 98195
Email: {tapas,haralick}@george.ee.washington.edu

ABSTRACT

Receiver operating characteristic curve is a standard method for reporting performance of a system. In this paper we show how to choose the optimal operating point when we are given a receiver operating curve, the prior probabilities, and the economic gain matrix. Unlike earlier methods, we make no assumptions regarding underlying distributions.

1. THEORY

Let us assume that the signal can either have only noise, N, or else it can have target plus noise, S. Let $f_N(x)$ be the density function of the noise strength, and let $f_S(x)$ be the density function of the noise with target. If a particular value, say t, is chosen as the decision threshold, then for any signal strength above t the classifier labels the input as S, whereas, if the signal strength is lower than t, the classifier labels the input as N.

The classifier is given a large number of signals with either N or S. Denote the known prior probability of S and N by $P(S)$ and $P(N)$, respectively. For a given threshold t, let $P(S|N,t)$ represent the probability of the classifier wrongly labeling the input as S when there is N in the input. $P(S|N,t)$ is referred to as the *probability of false alarm,* and is denoted by $P_F(t)$. Similarly, let $P(N|S,t)$ represent the probability of the classifier wrongly labeling the input as N when there is S in the input, for a particular threshold t. $P(S|N,t)$ is referred to as the *probability of mis-detection,* and is denoted by $P_M(t)$. The terms $P(S|S,t)$ and $P(N|N,t)$ are defined similarly.

We will need the following equations in the subsequent discussion.

$$P(S|N,t) = P_F(t) = \int_t^{\infty} f_N(x)dx \tag{1}$$

$$\frac{\partial P(S|N,t)}{\partial t} = -f_N(t) \tag{2}$$

$$P(N|S,t) = P_M(t) = \int_{-\infty}^{t} f_S(x)dx \tag{3}$$

$$\frac{\partial P(N|S,t)}{\partial t} = f_S(t) \tag{4}$$

The *receiver operating characteristic* curve, or ROC curve for short, is 2D parametric curve that is a plot of the points $(P_F(t), P_M(t))$ as threshold t is varied.

Let $C(S|S)$ be the economic gain associated with labeling the input as S when the input is S. $C(S|N)$, $C(N|N)$ and $C(N|S)$ are defined similarly. Now, we can write the equation for the gain, $G(t)$, as a function of t :

$$\begin{aligned} G(t) &= P(N)P(S|N,t)C(S|N) \\ &\quad +P(N)P(N|N,t)C(N|N) \\ &\quad +P(S)P(N|S,t)C(N|S) \\ &\quad +P(S)P(S|S,t)C(S|S). \end{aligned}$$

Substituting for conditional probabilities we get,

$$\begin{aligned} G(t) &= P(N)P(S|N,t)C(S|N) \\ &\quad +P(N)[1-P(S|N,t)]C(N|N) \\ &\quad +P(S)P(N|S,t)C(N|S) \\ &\quad +P(S)[1-P(N|S,t)]C(S|S) \\ &= P(N)P(S|N,t)[C(S|N)-C(N|N)] \\ &\quad +P(N)C(N|N) \\ &\quad +P(S)P(N|S,t)[C(N|S)-C(S|S)] \\ &\quad +P(S)C(S|S). \end{aligned} \tag{5}$$

To achieve the maximum gain, we should operate the classifier at the threshold t for which the derivative of the gain equals zero: $G'(t) = 0$. Thus,

$$\begin{aligned} \frac{\partial G(t)}{\partial t} &= P(N)[C(S|N)-C(N|N)]\frac{\partial P(S|N,t)}{\partial t} \\ &\quad +P(S)[C(N|S)-C(S|S)]\frac{\partial P(N|S,t)}{\partial t} \\ &= 0. \end{aligned}$$

0-8186-7310-9/95 $4.00

Rearranging we get,

$$\frac{P(N)[C(S|N) - C(N|N)]}{P(S)[C(N|S) - C(S|S)]} = -\frac{\partial P(N|S,t)/\partial t}{\partial P(S|N,t)/\partial t}.$$

Canceling the ∂t, we get,

$$\frac{P(N)[C(N|N) - C(S|N)]}{P(S)[C(S|S) - C(N|S)]} = -\frac{\partial P(N|S,t)}{\partial P(S|N,t)}. \quad (6)$$

Alternatively, the above equation can be written in terms of the density functions by using results from equations (2) and (4)

$$\frac{P(N)[C(N|N) - C(S|N)]}{P(S)[C(S|S) - C(N|S)]} = \frac{f_S(t)}{f_N(t)}. \quad (7)$$

Equation (6) implies that t should be chosen such that the derivative of the ROC curve at t should equal the quantity on the left hand side of the equation. For example, if the $P(N) = P(S) = 0.5$, $C(N|N) = C(S|S)$, and $C(N|S) = C(S|N)$, we see from equation (6) that the slope of the tangent at the point corresponding to t on the ROC curve should be equal to -1. That is, the tangent should be at 135 degrees counterclockwise with respect to the positive x-axis. In general, the operating point on the ROC curve should be where the slope of the tangent is equal to the ratio on the left hand side of equation (6).

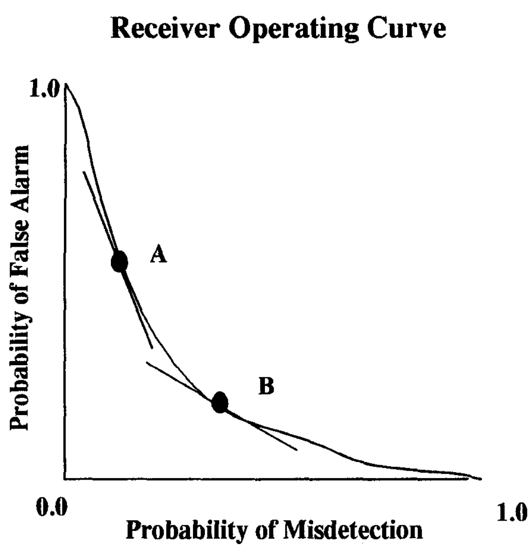

Figure 1: Point A is an optimal operating points when the slope of the tangent at A is equal to the ratio $P(N)(C(N|N) - C(S|N))/[P(S)(C(S|S) - C(N|S))]$. Similarly, the point B is an optimal point for a different value of the ratio.

Equation (7) implies that t should be chosen such that the ratio of the density function values $f_N(t)$ and

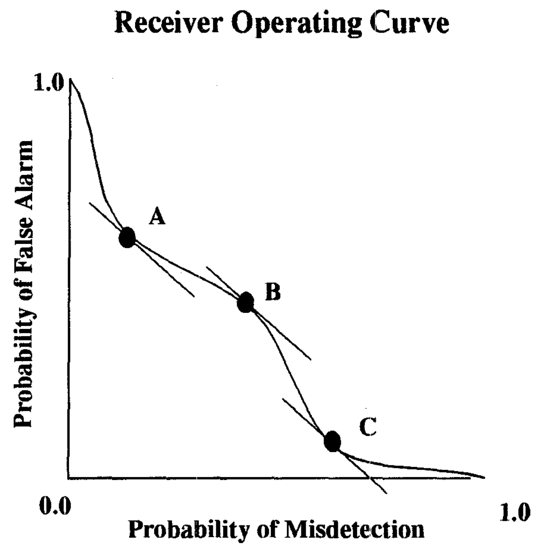

Figure 2: In this operating curve, the slopes at points A, B and C ar the same. The point at which the tangent line nearest to the origin (perpendicular distance), is the optimal operating point.

$f_S(t)$ should equal the quantity on the left hand side of the equation. For example, if the $P(N) = P(S) = 0.5$, $C(N|N) = C(S|S)$, and $C(N|S) = C(S|N)$, we see from equation (7) that the ratio $f_S(t)/f_N(t)$ should equal 1. That is, The density functions should be equal: $f_S(t) = f_N(t)$.

Note that it is possible to have multiple solutions to equations (6) and (7). This is because the operating curve might happen to be tangential to 135 degree lines at multiple places in the case of equation (6) and the density functions might happen to intersect at multiple points in the case of equation (7). The operating point at which the gain $G(t)$ is maximum can be found out by substituting the t values into the gain equations (5). There still is a possibility of existence of multiple operating points having the same gain. If that is the case, then any of the solutions where the gain function attains its maximum can be used as an operating point.

Furthermore, it can so turn out that the solution is actually a minimum instead of a maximum. This happens when the mean of the noise signal is higher than the mean of the signal with noise. In this case the operating point will be above the cross diagonal of the unit square.

In the case that the prior probabilities are deliberately chosen to make the performance as bad as possible, the optimal operating point is given by the Maximin decision rule, and the above analysis does not apply. In brief, if $C(S|S) = C(N|N) = 0$, the optimal operating point is the intersection of ROC and the line defined by the equation $P_M \cdot C(N|S) = (P_F) \cdot C(S|N)$, where P_M and P_F are variables and $P_F = 1 - P_M$ (see Van Trees [1] and Haralick and Shapiro [2] for details).

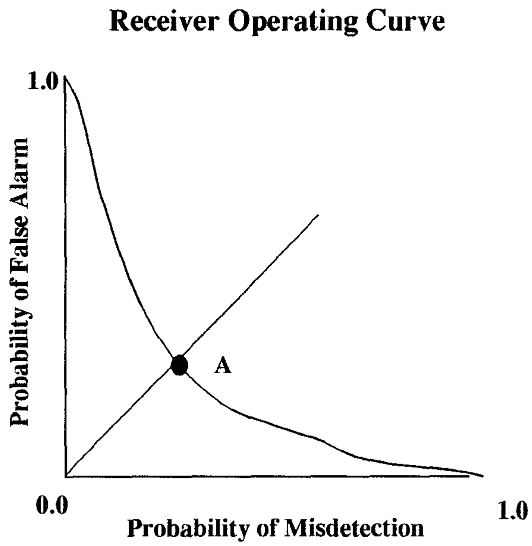

Figure 3: In this operating curve, the maximin operating point is at the point A, which is the intersection point of the ROC curve with a line passing through the origin. The slope of the line is computed from the economic gain matrix (see text).

Radiology community (see [3], and literature cited therein) and automatic target recognition (ATR) community make use of ROC methodology. For a discussion of quantitative performance evaluation for line detection algorithms where optimal operating points can be used see [4, 5].

2. PRACTICAL ISSUES

Although the theory described in the previous section has existed in the statistical signal detection literature for decades [1], and has been used by the radiology and ATR community quite extensively, there is a practical problem that still remains:

Since the ROC data is discrete and noisy, the derivatives computed directly from the data are noisy. Thus researchers using ROC methods assume that the density functions are Gaussian and then estimate parameters for the ROC curve. Unfortunately, Gaussian assumptions need not be valid in many cases. However, due to lack of better tools for computing reliable derivatives along the ROC curves, researchers still make the Gaussian assumptions.

Our approach does not make any Gaussian assumptions regarding the density functions. The crucial point for finding the optimal operating point is that the ROC should be represented in a way such that reliable derivatives can be computed anywhere along the curve. We accomplish this by computing a spline representation of the ROC data that satisfies the monotonicity and end-point constraints. The derivatives are then computed from this spline representation. Finally, the derivatives computed from the spline representation are used in the computation of the optimal operating point (according to equation 6). In the case one is looking for the Maximin operating point, the intersection point of any line through the origin and the spline representation of the ROC can be computed easily (derivatives are not required). Details of our monotone spline regression algorithm can be found in our papers [6, 7]. In figure 4 we show a spline fit using our algorithm to ROC data. Maximin operating points and operating points when priors are known are computed easily from the fitted spline.

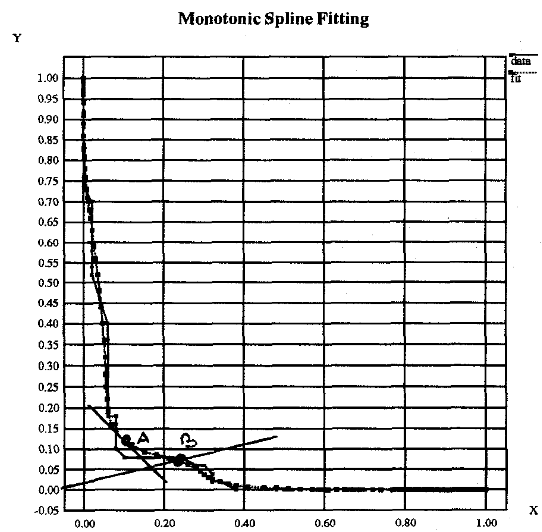

Figure 4: ROC curve fitting with splines. The number of data points was 88. The point A is an operating point when the priors are known and the point B is a maximin operating point.

3. DISCUSSION

In the previous sections we outlined the theory of finding optimal Bayesian operating points in binary hypothesis testing problems. Unlike other methods in the literature, our method does not make any assumptions regarding density functions. We fit the ROC curve with differentiable, monotonically decreasing spline functions that have the end-points at $(0, 1)$ and $(0, 1)$. Derivatives of splines, which are required for finding optimal operating points, are easily computed. For maximin operating points, intersections with straight lines with splines are required, which can again be computed easily.

There are few points that still need to be addressed. First, we have not provided a means of computing confidence intervals associated with the operating points. This can be approached via bootstrap methods. Second, currently the user has to provide the knot locations and order of splines to be fit to the ROC data; this eventually needs to be done automatically. The model selection problem can be solved using cross-validation techniques.

Another approach to the problem of fitting parametric curves to ROC data is through fitting functions to the density functions. That is, one can fit splines (that are non-negative and integrate to 1) to histograms and then compute the ROC fit using the smooth density functions. This is possible only when histograms are available (this need not always be the case, e.g., in psychophysics one does not usually have access to the underlying histograms). In our papers [6, 7] we have also provided a way of computing spline fits to histograms while satisfying non-negativity and integral constraints.

Acknowledgement: Authors would like to thank Henry Baird and David Gay from Bell Labs, Murray Hill, for discussions on the topic. Part of this work was done while Kanungo was visiting Bell Labs, and he would like to thank Henry Baird for inviting him there.

4. REFERENCES

[1] H. L. van Trees. *Detection, Estimation, and Modulation Theory: Part I.* John Wiley & Sons, New York, 1968.

[2] R. M. Haralick and L. G. Shapiro. *Computer and Robot Vision*, volume 1. Addison-Wesley, 1992.

[3] C. E. Metz. ROC methodology in radiologic imaging. *Investigative Radiology*, 21(9), 1986.

[4] T. Kanungo, M.Y. Jaisimha, J. Palmer, and R.M. Haralick. A quantitative methodology for analyzing the performance of detection algorithms. In *Proc. of IEEE Int. Conf. on Computer Vision*, pages 247–252, Berlin, May 1993.

[5] T. Kanungo, M.Y. Jaisimha, J. Palmer, and R.M. Haralick. A methodology for quantitative performance evaluation of detection algorithms. *IEEE Trans. on Image Processing (to appear)*, December 1994.

[6] T. Kanungo, D. M. Gay, and R.M. Haralick. Constrained monotone regression of roc curves and histograms using splines and polynomials. In *Proc. of IEEE Int. Conf. on Image Processing*, Washington, D.C., October 1995.

[7] T. Kanungo. Constrained monotone regression with splines and polynomials. Technical Report ISL-TR-94-06, University of Washington, Seattle, WA, 1994.

[8] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis.* Wiley, New York, 1973.

[9] K. Fukunaga. *Introduction to statistical pattern recognition.* Academic Press, Boston, 1990.