Journal of the Korean Data & Information Science Society
2015, **26**(1), 255–260

http://dx.doi.org/10.7465/jkdi.2015.26.1.255
한국데이터정보과학회지

Two optimal threshold criteria for ROC analysis

Min Ho Cho[1] · Chong Sun Hong[2]

[12]Department of Statistics, Sungkyunkwan University

Received 8 December 2014, revised 2 January 2015, accepted 10 January 2015

Abstract

Among many optimal threshold criteria from ROC curve, the closest-to-(0,1) and amended closest-to-(0,1) criteria are considered. An ROC curve that passes close to the (0,1) point indicates that two models are well classified. In this case, the ROC curve is located far from the (1,0) point. Hence we propose two criteria: the farthest-to-(1,0) and amended farthest-to-(1,0) criteria. These criteria are found to have a relationship with the KolmogorovSmirnov statistic as well as some optimal threshold criteria. Moreover, we derive that a definition for the proposed criteria with more than two dimensions and with relations to multi-dimensional optimal threshold criteria.

Keywords: ROC, threshold, true negative rate, true positive rate.

1. Introduction

There are many criteria for determining the optimal threshold for two classified models. Some of them can be explained using the receiver operating characteristic (ROC) curve (Provost and Fawcett, 2001; Sobehart and Keenan, 2001; Engelmann *et al.*, 2003; Fawcett, 2003; Zho *et al.*, 2007; Hong, 2009; Hong *et al.*, 2013). Among them, there are the closest-to-(0,1) and amended closest-to-(0,1) criteria of Perkins and Schisterman (2006). These two criteria are based on the idea that models are well classified when the ROC curve is closer to the (0,1) point. Since the ROC curve that plots close to the (0,1) point is far from the (1,0) point, we propose two criteria in this paper: the farthest-to-(1,0) and amended farthest-to-(1,0) criteria.

Definitions of the farthest-to-(1,0) and amended farthest-to-(1,0) criteria, denoted as F and AF, respectively, are presented and explained in Section 2. The F and AF criteria could be extended to more than two dimensions, as explained in Section 3. Some relationships have been found between these criteria and others such as MVD (maximum vertical distance; Krzanowski and Hand, 2009), J (Youden index; Youden, 1950), SSS (sum of sensitivity and specificity; Connell and Koepsell, 1985), TR (true rate: Velez *et al.*, 2007; Hong and Joo, 2010), AA (accuracy area; Brasil, 2010) and $CCSR$ (correct classification simple rate; Hong and Wu, 2014).

[1] Graduate student, Department of Statistics, Sungkyunkwan University, Seoul 110-745, Korea.
[2] Corresponding author: Professor, Department of Statistics, Sungkyunkwan University, Seoul 110-745, Korea. E-mail: cshong@skku.edu

2. Farthest-to-(1,0) and amended farthest-to-(1,0) criteria

Perkins and Schisterman (2006) proposed the closest-to-(0,1) and amended closest-to-(0,1) criteria, C and AC, respectively. The closest-to-(0,1) criterion minimizes the distance between the (0,1) point and ROC curve, $C = \min\sqrt{(1-F_d(x))^2 + (F_n(x))^2}$, where $F_d(\cdot)$ and $F_n(\cdot)$ are the cumulative distribution functions of the default and non-default states, respectively, assuming $F_d(x) \geq F_n(x)$ for all x. The amended closest-to-(0,1) criterion obtains a threshold for minimizing the ratio of two distances: the distance from the (0,1) point to the ROC curve, and the distance from the (0,1) point to the straight line, $F_d(x) = F_n(x)$, so that $AC = \min\{(1-F_d(x)) + F_n(x)\}$.

With similar arguments, we consider a criterion maximizing the distance between the (1,0) point and the ROC curve, since we can assume that the ROC curve near the (0,1) point might be far away from the (1,0) point. This criterion is the square root of the sum of the square of the true positive rate (TPR) and squared of the true negative rate (TNR) (see Figure 2.1). This is called the farthest-to-(1,0) criterion.

Definition 2.1 The farthest-to-(1,0) criterion: F

$$F = \max\sqrt{(1-F_n(x))^2 + (F_d(x))^2}. \tag{2.1}$$

By extending the amended closest-to-(0,1) criteria (AC) of Perkins and Schisterman (2006), two distances are considered: the distance from the (1,0) point to the straight line, $F_d(x) = F_n(x)$, (r_2) and the distance from the (1,0) point to the ROC curve (r_1) (see Figure 2.2). We then propose an additional criterion that maximizes the ratio of two distances r_1 and r_2 such that

Definition 2.2 The amended farthest-to-(1,0) criterion: AF

$$\begin{aligned} AF = \max\left(\frac{r_1}{r_2}\right) &= \max\sqrt{\frac{(1-F_n(x))^2 + (F_d(x))^2}{\left(\frac{1-F_n(x)}{1-F_n(x)+F_d(x)}\right)^2 + \left(\frac{F_d(x)}{1-F_n(x)+F_d(x)}\right)^2}} \\ &= \max\{(1-F_n(x)) + F_d(x)\}. \end{aligned} \tag{2.2}$$

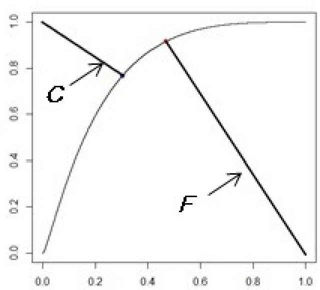

Figure 2.1 F and C in ROC curve

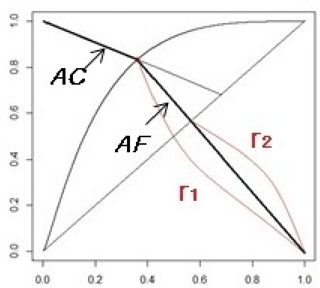

Figure 2.2 AF and AC in ROC curve

Note that the AF is the same measure as the SSS (Connell and Koepsell, 1985). It is found that F and AF are designed with TPR and TNR and the thresholds obtained by using F and AF are not identical, and also differ from those for C and AC.

The square of the AF criterion can be expressed as

$$(AF)^2 = \max\{(1 - F_n(x))^2 + (F_d(x))^2 + 2(1 - F_n(x))(F_d(x))\}. \quad (2.3)$$

The first two terms on the right side of (2.3) are the same as F in (2.1), and the last third cross product term is identical to AA (accuracy area: Brasil, 2010). This can be described geometrically, as shown in Figure 2.3. Since the thresholds obtained by F and AA are not always the same as those obtained by AF, and the square of AF is not equivalent to the sum of both the square of F and two times that of AA such as $(AF)^2 \neq (F)^2 + 2AA$, where $AA = \max\{(1 - F_n(x))F_d(x)\}$.

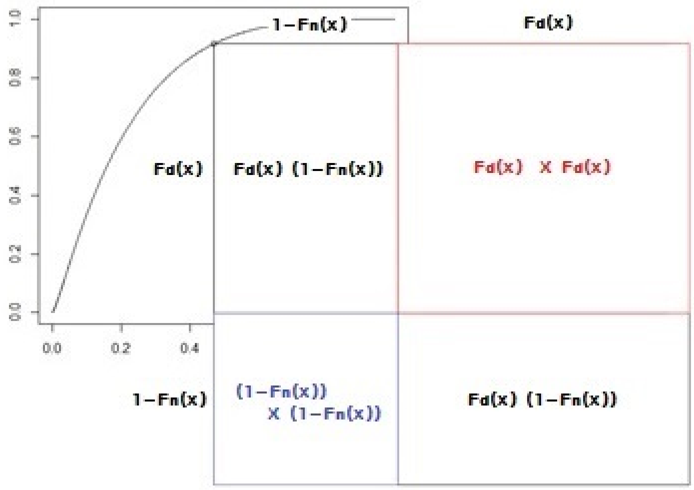

Figure 2.3 Geometric meaning of relationships among F, AF and AA

3. Extension to multi-dimensions

These criteria are applicable to the ROC surface in three dimensions: the farthest-to-(0,0,0) and amended farthest-to-(0,0,0) criteria. The farthest-to-(0,0,0) criterion is to find a threshold which maximizes the distance between the origin point (0,0,0) and the ROC surface (see Figure 3.1):

$$F^3 = \max\sqrt{(F_1(x))^2 + (F_2(y) - F_2(x))^2 + (1 - F_3(y))^2}. \quad (3.1)$$

The superscript over the F in (3.1) indicates that this criterion has three outcomes in three dimensions.

Two distances r_1 and r_2 in Figure 2.2 are extended to the ROC surface, so that the amended farthest-to-(0,0,0) criterion could be defined to maximize the ratio of the distance

(r_1) from the origin to the curve and that of (r_2) from the origin to the plane, $F_1(x) + F_2(x) + F_3(x) = 1$, (see Figure 3.1).

$$AF^3 = \max(r_1/r_2),$$

where $r_1^2 = F_1(x)^2 + (F_2(y) - F_2(x))^2 + (1 - F_3(y))^2$, and $r_2^2 = [F_1(x)^2 + (F_2(y) - F_2(x))^2 + (1 - F_3(y))^2]/[F_1(x) + (F_2(y) - F_2(x)) + (1 - F_3(y))]^2$. Hence the AF in three dimensions is expressed as

$$AF^3 = \max\{(F_1(x)) + (F_2(y) - F_2(x)) + (1 - F_3(y))\}. \tag{3.2}$$

The three terms on the right side of (3.2) represent all true classification rates for a 3×3 confusion matrix. Since both sensitivity and specificity are true classification rates, SSS might be regarded as the STR (sum of true rates). Then it can be said that the AF in three dimensions is identical to STR.

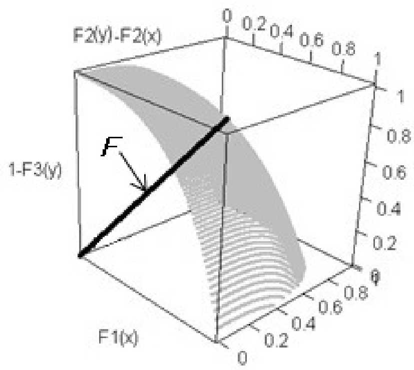

Figure 3.1 F in ROC surface

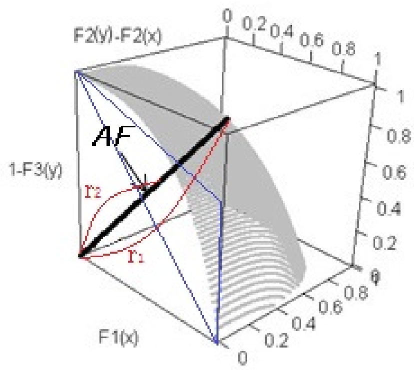

Figure 3.2 AF in ROC surface

The farthest-to-origin and amended farthest-to-origin criteria can be generalized to the ROC manifold for k dimensions.

Definition 3.1 The farthest-to-origin and amended farthest-to-origin criteria for k dimensions

$$\begin{aligned} F^k &= \max \sqrt{(F_1(x_1))^2 + (F_2(x_2) - F_2(x_1))^2 + \cdots + (1 - F_k(x_{k-1}))^2}, \\ AF^k &= \max\{(F_1(x_1)) + (F_2(x_2) - F_2(x_1)) + \cdots + (1 - F_k(x_{k-1}))\}. \end{aligned} \tag{3.3}$$

We can also say that the AF is the same as STR for multi-dimensions, as the preceding that AF is the same as STR for two and three dimensions.

For k dimensions, the square of AF is not identical to the sum of both the square of F and two times that of AA's such as

$$(AF^k)^2 \neq (F^k)^2 + 2(AA_{1,2} + AA_{2,3} + \cdots + AA_{k,1}),$$

since thresholds obtained by F and AA are not the same as those obtained by AF.

Hong and Yoo (2011) and Hong and Jung (2013) showed relationships among J, AC, TR,

MVD, SSS ($=STR$) criteria and the Kolmogorov-Smirnov statistic for a three dimensional ROC surface. Furthermore, Hong and Wu (2014) proposed the $CCSR$ (correct classification simple rate), and described relationships with $CCSR$ and all of these measures in multi-dimensions. Since the AF has a linear relationship with STR, AF can also be expressed in terms of relationships with KS, MVD, J, AC, STR, TR and $CCSR$ for k dimensions as follows.

Properties of the AF

(1) $$AF^k = \sum_{j=2}^{k} KS_{j-1,j} + 1,$$

(2) $$AF^k = MVD^k + 1,$$

(3) $$AF^k = STR^k,$$

(4) $$AF^k = \frac{k}{2}(J^k + 1),$$

(5) $$AF^k = k \times TR^k,$$

(6) $$AF^k = 2 - AC^k,$$

(7) $$AF^k = \sum_{i}^{k} CCR_i = CCSR^k.$$

Therefore, the AF has a linear relationship with seven other criteria for k (≥ 3) dimensions. The optimal thresholds obtained via AF are identical to those obtained via the seven criteria: KS, MVD, STR, J, TR, AC and $CCSR$.

4. Conclusion

In this paper, two optimal threshold criteria, the farthest-to-(1,0) and amended farthest-to-(1,0) criteria (F and AF) are proposed based on the closest-to-(0,1) and amended closest-to-(0,1) criteria (C and AC) of Perkins and Schisterman (2006).

It is found that the threshold obtained by using F may be different from that by using C. The AF is known to be the same measure as STR. The F and AF can also be extended to more than two dimensions. Mathematical and geometrical relationships among F, AF and AA are discussed.

Furthermore, we derive seven properties of the AF; AF has a linear relationship with the summation of the Kolmogorov-Smirnov statistics and some optimal threshold criteria such as the Youden index, the maximum vertical distance, the amended closet-to-(0,1) criterion, the sum of sensitivity and specificity, the true rate and the correct classification simple rate.

References

Brasil, P. (2010). DiagnosisMed: Diagnostic test accuracy evaluation for medical professionals. R package version 0.2.3.

Connell, F. A. and Koepsell, T. D. (1985). Measures of gain in certainty from a diagnostic test. *American Journal of Epidiology*, **121**, 744-753.

Engelmann, B., Hayden, E. and Tasche, D. (2003). Measuring the discriminative power of rating systems. *Risk*, 82-86.

Fawcett, T. (2003). *ROC graphs: notes and practical considerations for data mining researchers*, HP Labs Tech Report HPL-2003-4, CA, USA.

Hong, C. S. (2009). Optimal threshold from ROC and CAP curves. *Communications in Statistics - Simulation and Computation*, **38**, 2060-2072.

Hong, C. S. and Joo, J. S. (2010). Optimal thresholds from non-normal mixture. *The Korean Journal of Applied Statistics*, **23**, 943-953.

Hong, C. S., Joo, J. S. and Choi, J. S. (2010). Optimal thresholds from mixture distributions. *The Korean Journal of Applied Statistics*, **23**, 13-28.

Hong, C. S. and Jung, E. S. (2013). Optimal thresholds criteria for ROC surfaces. *Journal of the Korean Data & Information Science Society*, **24**, 1489-1496.

Hong, C. S. and Wu, Z. Q. (2014). Alternative accuracy for multiple ROC analysis. *Journal of the Korean Data & Information Science Society*, **25**, 1521-1530.

Krzanowshi, W. J. and Hand, D. J. (2009). *ROC curves for continuous data*, Chapman and Hall, London.

Li, J. and Fine, J. P. (2008). ROC analysis with multiple classes and multiple tests: Methodology and its application in microarray studies. *Biostatistics*, **9**, 566-576.

Perkins, N. J. and Schisterman, E. F. (2006). The inconsistency of optimal cutpoints obtained using two criteria based on based on the receiver operating characteristic curve. *American Journal of Epidiology*, **163**, 670-675.

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

Sobehart, J. R. and Keenan, S. C. (2001). Measuring default accurately. *Credit Risk Special Report, Risk*, **14**, 31-33.

Velez, D. R., White, B. C., Motsinger, A. A., Bush, W. S., Ritchie, M. D., Williams, S. M. and Moore, J. H. (2007). A balanced accuracy function for epistasis modeling in imbalanced datasets using multifactor dimensionality reduction. *Genetic Epidiology*, **31**, 306-315.

Youden, W. J. (1950). Index for rating diagnostic test. *Cancer*, **3**, 32-35.

Zou, K. H., O'Malley, A. J. and Mauri, L. (2007). Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation*, **115**, 654-657.