

# 基于 ROC 曲线的两类分类问题性能评估方法

韦修喜<sup>1</sup>, 周永权<sup>2</sup>

(1. 广西国际商务职业技术学院 信息工程系, 广西 南宁 530007;  
2. 广西民族大学 数学与计算机科学学院, 广西 南宁 530006)

**摘要:** 为了克服传统分类算法指标存在的不足, 即分类准确度、精确度和检测概率等指标对类别先验概率不具有稳健性, 采用基于雷达接收机工作特性曲线即 ROC 曲线的评估方法来评估两类分类算法, 通过参数法或非参数法建立 ROC 曲线, 使用几率点欧氏距离、曲线下面积、最佳阈值点等指标对两类算法进行性能评估。实验结果表明, 相对于传统算法, 基于 ROC 曲线的评估方法的性能得到了很大的提高, 在分类器识别算法性能评估中是一种有效的评估方法。

**关键词:** ROC 曲线; 两类分类算法; 性能评估

**中图分类号:** TP18

**文献标识码:** A

**文章编号:** 1673-629X(2010)11-0047-04

## A New Performance Categories Evaluation Method Based on ROC Curve

WEI Xiu-xi<sup>1</sup>, ZHOU Yong-quan<sup>2</sup>

(1. Information Engineering Department, Guangxi International Business  
Vocational College, Nanning 530007, China;  
2. College of Computer and Information Science, Guangxi University for  
Nationalities, Nanning 530006, China)

**Abstract:** To overcome the shortage of traditional methods, for example, decision analysis and pattern recognition, precision, the probability of detection and so on, they are not robust to the prior probability of classes, so the ROC curve evaluation method is used in this paper. In order to establish the ROC, uses the parameters method, non-parameters method, the area of the curve method and the best at the threshold as stands. The experimental results show that compared with traditional algorithms, based on ROC curve performance evaluation method has been greatly improved. So it is an effective evaluation measure of the classifier recognition algorithm performance evaluation.

**Key words:** ROC curve; categories algorithm; performance evaluation

## 0 引言

在电子、工程等诸多领域, 数据分类问题的性能评估是重要的研究方向。其中, 传统的评估两类分类算法性能的指标常用的有精确度、检测概率和分类准确度等。然而, 随着研究的深入和应用的广泛, 人们发现这些传统算法的性能指标存在某些局限性, 特别是对类别先验概率和阈值不具有稳健性<sup>[1,2]</sup>。为了克服传统算法的不足, 学者们采用雷达接收机工作特性曲线,

即 ROC 曲线来评估两类分类算法的性能。ROC 曲线描绘的是诊断中敏感性和特殊性之间的制约关系, 由于它能更全面客观地对评估系统进行分析, 所以在实验医学、临床流行病学、生物统计学、放射学、数据挖掘和模式识别等领域得到了广泛的研究和应用<sup>[3~8]</sup>。

文中首先介绍 ROC 曲线的建立方法和评估指标, 然后通过数值模拟实验证明基于 ROC 曲线的性能评估方法的有效性。

## 1 问题定义

在模式识别中, 最简单的两类分类问题是正常事件和非正常事件的分类情况<sup>[9]</sup>。当分类器训练出来的样本数据特征值大于某个事先预置的阈值  $h$  时称为正常事件, 小于该阈值时称为非正常事件。因此, 针对不同的输入数据样本, 分类器的判别结果不同, 会出现

收稿日期: 2010-01-29; 修回日期: 2010-05-28

基金项目: 国家自然科学基金(60461001); 广西自然科学基金(0542048, 0832082, 0991086); 国家民委科研项目(08GX01)

作者简介: 韦修喜(1980-), 男, 广西百色人, 硕士, 研究方向为神经网络、计算智能及应用; 周永权, 博士, 教授, 研究方向为计算智能、神经网络及应用。

以下 4 种结果:第一,当样本特征值大于阈值时,判别结果为正常事件;第二,当样本特征值小于阈值时,判别结果为非正常事件;第三,非正常事件的特征值大于阈值,判为正常事件;第四,当非正常事件特征值小于阈值时,判别结果称为非正常事件。这 4 种结果通常用一个二维混淆矩阵来表示,每个结果对应矩阵的一个元素,而发生这些结果的概率分别记作:TPR, TNR, FPR, FNR。如果样本特征值概率密度分布未知,通常用实验的方法算出它的估计值,具体操作如下:

$$(\text{敏感性})\text{TPR} = \frac{\text{正确识别正常事件的数量}}{\text{正常事件的总数量}}$$

$$(1 - \text{敏感性})\text{TNR} =$$

$$\frac{\text{正常事件判为非正常事件的数量}}{\text{正常事件总数量}}$$

$$(\text{特殊性})\text{FNR} = \frac{\text{正确识别非正常事件的数量}}{\text{非正常事件总数量}}$$

$$(1 - \text{特殊性})\text{FPR} =$$

$$\frac{\text{非正常事件判为正常事件的数量}}{\text{非正常事件总数量}}$$

由上述定义可知:

$$\text{TPR} + \text{TNR} = 1$$

$$\text{FNR} + \text{FPR} = 1$$

在传统评估算法中,通常把 TPR, TNR, FPR 和 FNR 作为评估识别算法的指标,这些指标评估的结果依赖于阈值,因此,这些指标对阈值不具有稳健性,其应用范围也受到了很大的限制。为了克服传统评估算法存在的这一缺陷,人们尝试采用 ROC 曲线评估方法,该曲线揭示了当阈值变化时 TPR 和 FPR 之间的关系,使评估结果更加全面和客观可信。

## 2 ROC 曲线的建立

一般地,ROC 曲线的建立主要有有参数和非参数两种方法<sup>[10]</sup>。

### 2.1 参数法

参数法适用于正常事件和非正常事件的特征值分布的概率密度函数为已知的情况。假设它们的特征值都服从于正态分布的情况:

$$g(x) \sim N(\mu_1, \sigma_1^2), f(x) \sim N(\mu_0, \sigma_0^2) \quad (1)$$

$g(x), f(x)$  分别表示正常事件和非正常事件特征值的概率密度函数。

由 TPR 和 FPR 的定义可知:

$$\begin{aligned} \text{TPR} &= \frac{1}{\sqrt{2\pi\sigma_1}} \int_0^{+\infty} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} dx \\ \text{FPR} &= \frac{1}{\sqrt{2\pi\sigma_0}} \int_0^{+\infty} e^{-\frac{(x-\mu_0)^2}{2\sigma_0^2}} dx \end{aligned} \quad (2)$$

由以上两式可以求得 ROC 曲线的函数解析式

为<sup>[7]</sup>:

$$\text{TPR} = \phi(a + b\phi^{-1}(\text{FPR})) \quad (3)$$

其中  $\phi$  为标准正态分布的累计分布函数:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt, a = \frac{\mu_1 - \mu_0}{\sigma_1}, b = \frac{\sigma_0}{\sigma_1} \quad (4)$$

由实验数据  $a$  和  $b$  给出极大似然估计值  $\hat{a}$  和  $\hat{b}$ ,代入式(3),然后画出 ROC 曲线。

式(3)又可以表示为

$$\phi^{-1}(\text{TPR}) = a + b\phi^{-1}(\text{FPR}) \quad (5)$$

显然,这是一条斜率为  $b$ ,截距为  $a$  的直线。这样,由测试数据得到 TPR 和 FPR 的估计值,再以  $\phi^{-1}(\text{TPR})$  和  $\phi^{-1}(\text{FPR})$  为坐标轴画曲线,即为标准偏差曲线(Normal Deviation Curve, NDC)<sup>[11]</sup>。

### 2.2 非参数法

在正常事件和非正常事件的特征值分布概率密度函数未知或者很难求出其解析式的情况,通常采用非参数法。其主要思想是通过改变阈值,得到分类算法的不同的敏感性和特殊性的点,然后以  $(1 - \text{特殊性})$  为横坐标,以敏感性为纵坐标,用平滑曲线连接各点从而作出 ROC 曲线。

由于在实际应用中很难算出正常事件和非正常事件特征值分布的概率分布函数,因此在实际应用中通常不采用参数法,而是通过非参数法来建立 ROC 曲线。

## 3 ROC 曲线的评估指标

传统的评估指标只是对分类器单一阈值的性能进行评估,而参数法和非参数法得到的 ROC 曲线都是基于阈值变化而得到的,它反映了分类器在所有可能阈值下的性能,具有一般评估方法没有的特性,即 ROC 曲线的形状不会随先验概率的改变而改变,而且它的建立和错误分类代价无关。

鉴于 ROC 曲线评估方法的优良性能,ROC 曲线的评估方法得到了广泛的应用,并且在实际应用中逐渐形成了一系列评估指标和评估准则,以下介绍几种常用的 ROC 曲线评估指标。

### 3.1 几率点欧氏距离

所谓的几率点欧氏距离,就是在 ROC 曲线所在的坐标轴上,连接坐标点  $(0, 1)$  和坐标点  $(1, 0)$  画一条直线,设该直线与 ROC 曲线的交点为 A,则 A 点到点  $(0.5, 0.5)$  的欧氏距离就叫做几率点欧氏距离。由于几率点欧氏距离对 ROC 曲线的整体性不能给出合理的评价,现在已经很少使用这个指标。

### 3.2 曲线下面积

由于分类器的 4 种结果通常用一个二维混淆矩阵

来表示,所以常常通过计算 ROC 曲线下面积(Area Under the Curve, AUC)来比较不同的分类算法。AUC 反映了分类算法正确区分正常事件和非正常事件能力的大小<sup>[12]</sup>。图 1 显示的是两条 ROC 曲线对应的 AUC 值,从图 1 可以看出,AUC 的取值范围为 0.5 至 1.0 之间,ROC 曲线越快速逼近左上角,AUC 就越大,说明分类算法的性能就越好。由于 AUC 对 ROC 曲线的描述非常直观,所以 AUC 在很多领域的分类器评估中得到了应用。

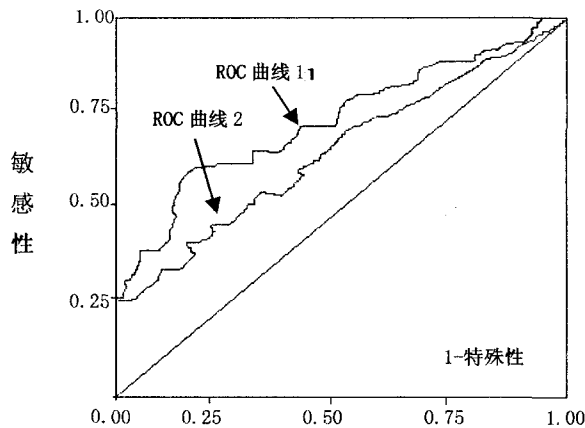


图 1 两条 ROC 曲线对应的 AUC 值

### 3.3 最佳阈值点

根据分类器结果进行评估,总是要付出一定的代价的。一般代价函数用贝叶斯代价函数,即为:

$$R = \lambda_{aa}P(A)s(\Delta) + \lambda_{an}P(A)(1 - s(\Delta)) + \lambda_{na}P(N)f(\Delta) + \lambda_{nn}P(N)(1 - f(\Delta)) \quad (6)$$

当设定一个阈值  $\Delta$  后,用  $s(\Delta)$  和  $f(\Delta)$  分别表示这个方法中的敏感性和特殊性概念。为了得到最好的阈值,可以最小化风险  $R$  来实现,具体做法就是对  $R$  求微分并令微分结果为 0,这样就得到:

$$\frac{ds(\Delta)}{df(\Delta)} = \frac{(\lambda_{nn} - \lambda_{na})P(N)}{(\lambda_{aa} - \lambda_{an})P(A)} \quad (7)$$

在一个两类分类问题中,ROC 曲线上斜率等于公式(7)计算出来的结果就称为一个最优运行点,换句话说,它对应着最好的阈值。注意到,对于一个两类分类问题,这是一种“模型—无关”的特征阈值选择方法,它对应于样本的分布形状并没有任何的特殊假设。ROC 曲线斜率越大,对应阈值的代价就越大;否则对应阈值的代价就越小。

## 4 模拟实验

下面举两个实例来说明 ROC 曲线如何应用于具体实际应用中。

例 1:一个信号噪声数据集包含 100 个附加了噪声的信号样本  $s(n)$ ,信号里的噪声加入方法是随机

的,幅度也是随机的,发生时间也是随机的服从一个泊松分布。掺入了噪声的信号幅度通常比平均噪声幅度大,出现这个现象的地方就反映了信号脉冲的存在。假定选择一个幅度阈值  $\Delta$ ,根据这个阈值去判别一个信号是纯信号脉冲还是噪声。也就是说,如果  $s(n) > \Delta$  就判断信号为脉冲(这里对应为非正常事件),否则就判断信号为噪声(这里对应着正常事件)。只要给定一个阈值,就可以得到相应的纯信号和噪声的分类矩阵,并计算出此时的敏感性和特殊性。通过改变阈值的大小,就可以得到对应的敏感性和特殊性的数值,如表 1 所示。

表 1 脉冲检测中的敏感性和特殊性  
(一共 100 个信号值)

阈值	敏感性	特殊性
1	0.90	0.66
2	0.80	0.80
3	0.70	0.87
4	0.70	0.93

从表 1 中可看出,在敏感性和特殊性之间有着一个“折中”的问题。目标是要有高的敏感性和高的特殊性,为此以  $(1 - \text{特殊性})$  为横坐标,敏感性为纵坐标,画出对应的 ROC 曲线,当所给的信号样本越多,得到的阈值也越多,得到的 ROC 曲线就会越光滑,如图 2 所示。

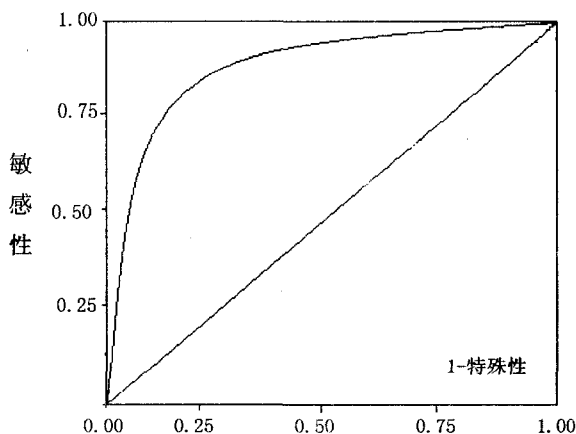


图 2 对应的 ROC 曲线图

用(7)式就可以求出最好的阈值,求出最佳阈值后就可以进行分类了。

例 2:应用 ROC 曲线来比较各个分类方法的优劣。假设用 3 种方法对同一个数据集进行分类,所得的对应的 ROC 曲线如图 3 所示。

用曲线下面积法可知,一个从多个分类方法中选择最好方法的途径就是看哪种方法对应的 ROC 曲线下包围的面积大。

由此可看出,在对这个数据集进行分类中,ROC

曲线 1 对应的分类方法的分类效果最好。

## 5 结束语

利用参数法或非参数法建立 ROC 曲线,使用几率点欧氏距离、曲线下面积、最佳阈值点等指标对两类算法进行性能评估,克服了传统算法评估结果依赖阈值的局限性。通过数值模拟实验证明,利用 ROC 曲线进行两类分类算法的评估是高效可行的。目前,ROC 曲线主要应用于两类分类方法的评估问题,把 ROC 曲线应用于多目标分类算法的评估问题,是下一步的研究工作。

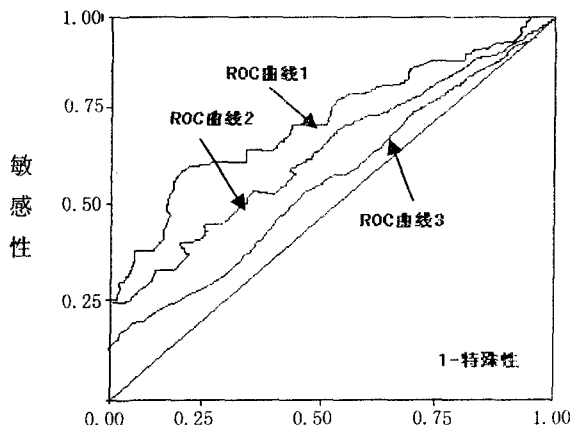


图 3 三条 ROC 曲线图

## 参考文献:

- [1] Alsing S G. The Evaluation of Competing Classification[D].

(上接第 46 页)

使基于统一接口的查询成为可能;开发语言基于 Java,易于实现程序的移植;用 Web Service 技术发布包装器和中介器,提高了软件的复用性;采用 B/S 架构,可以对全局数据源、局部数据源及它们之间的映射关系实现方便管理;对所有的局部数据源采用一个包装器进行封装,使增加新的数据源更加方便,增强了系统的扩展性。

## 参考文献:

- [1] 史万江,吴健.一种异构数据集成框架及关键技术研究[J].微处理机,2008(4):167-173.  
[2] El-Khatib H T, Williams M L, MacKinnon L M. Using a distributed approach to retrieve and integrate information from heterogeneous distributed databases[J]. Computer Journal, 2002, 45(4): 381-394.  
[3] 方长江.异构数据源集成技术在军事中的应用研究[D].济南:山东大学,2007.  
[4] 王蕊,陈庆奎.异构数据库集成中间件的研究与实现

US:Air Force Institute of Technology,2002.

- [2] Provost F, Fawcett T. Robust Classification for Imprecise Environments[J]. Machine Learning, 2001, 42(3): 203-231.  
[3] 宇传华.ROC 分析方法及其在医学中的应用[D].西安:第四军医大学,2000.  
[4] Swets J A. ROC Analysis Applied to the Evaluation of Medical Imaging Techniques[J]. Investigative Radiology, 1997, 14(2):109-121.  
[5] 田俊.两个诊断指标的 ROC 曲线下面积的非参数检验方法[J].数理医学杂志,2002,15(3):201-204.  
[6] 邹莉玲,沈其君,陈峰,等.ROC 曲线下面积的 ML 估计与假设检验[J].中国公共卫生,2003,19(1):127-128.  
[7] 王昌元,谢晋东,李月卿.ROC 曲线中 AZ 的物理意义及数学表达式[J].泰山医学院学报,2003,24(2):102-105.  
[8] 邹洪侠,秦锋,程泽凯.二类分类器的 ROC 曲线生成算法[J].计算机技术与发展,2009,19(6):115-118.  
[9] de sa J P M. 模式识别—原理、方法及应用[M].北京:清华大学出版社,2002.  
[10] 孙长亮,何峻,肖怀铁.基于 ROC 曲线的目标识别性能评估方法[J].雷达科学与技术,2007,5(1):22-25.  
[11] Marzban C. A Comment on the ROC Curve and the Area Under it as Performance Measures[EB/OL]. 2004. <http://www.nhn.ou.edu/marzban>.  
[12] Hanley J A, McNeil B J. The Meaning and Use of the Area Under a Receiver Operating Characteristic(ROC) Curve[J]. Radiology, 1982, 143(1):29-36.

[J]. 计算机工程与设计,2008,29:5738-5744.

- [5] May W. An integrated architecture for exploring, wrapping, mediating and restructuring information from the web[C]//In Australasian Database Conference (ADC 2000). Australia: Austr. Comput. Sci. Commun,2000:82-89.  
[6] 李军怀,周明全,耿国华,等.XML 在异构数据集成中的应用研究[J]. 计算机应用,2002,22(9):18-24.  
[7] 孙长俊,周晓峰.基于 Web Services 的企业应用集成模型[J]. 计算机技术与发展,2006,16(5):209-210.  
[8] 余名高,贾秀峰,林坤江,等.基于 Web 服务的企业应用集成[J]. 计算机技术与发展,2007,17(5):55-58.  
[9] 熊玉庆,唐新怀.XCouple:一个新型异构信息集成平台[J]. 微电子学与计算机,2007,24:171-173.  
[10] W3C Recommendation, XQuery 1.0: an XML query language[EB/OL].2007. <http://www.w3.org/TR/xquery/>.  
[11] 李烨,冯志勇.基于 XQuery 的数据集成研究[J].微处理机,2008(3):120-122.  
[12] 尚蕾,孙志辉.基于 XML 的异构数据集成系统的查询处理[J]. 计算机工程,2005(3):121-123.