

# 一种新的基于 AUC 的多类分类评估方法

秦 锋, 罗 慧, 程泽凯, 任诗流

QIN Feng, LUO Hui, CHENG Ze-kai, REN Shi-liu

安徽工业大学 计算机学院, 安徽 马鞍山 243000

School of Computer Science, Anhui University of Technology, Ma'anshan, Anhui 243000, China

E-mail: lh821005@126.com

QIN Feng, LUO Hui, CHENG Ze-kai, et al. New multiple-classifier appraisal method based on AUC. *Computer Engineering and Applications*, 2008, 44(5): 194-196.

**Abstract:** The AUC method which based on ROC analysis is used to appraise two-classifier, cannot directly solve the appraisal of multiple-classifier. How effectively promoting it to the multiple-classifier appraisal is a hot topic in research area. This article first introduced a multi-class AUC appraisal method F-AUC which based on one against rest method. Then according to this foundation a new multi-class appraisal method B-AUC which based on the structure principle of binary tree was proposed. Making experiment in MBNC experiment platform, analyzing and comparing two methods, the results show that the new method not only is feasible, has the good performances in the appraisal effect and the speed effect.

**Key words:** classified appraisal; two-classifier; multiple-classifier; binary tree

**摘 要:** 基于 ROC 曲线分析的 AUC 方法最初用于评估二类分类, 不能直接解决多类别的评估, 如何有效地将其推广到多类别评估中是研究的热点问题。首先介绍一种基于一对多方法的多类别 AUC 评估方法 F-AUC, 在此基础上根据二叉树的构造思想提出一种新的多类别评估方法 B-AUC。该方法在 MBNC 实验平台上编程实现, 并和 F-AUC 方法进行比较和分析, 结果表明该方法是可行的, 在评估效果和速度效果上有良好的性能。

**关键词:** 分类评估; 二类分类; 多类分类; 二叉树

**文章编号:** 1002-8331(2008)05-0194-03 **文献标识码:** A **中图分类号:** TP18

## 1 引言

近几年, ROC (Receiver Operating Characteristics curve) 曲线分析广泛地应用到数据挖掘和机器学习领域, 基于 ROC 曲线分析的 AUC (the Area Under the roc Curve) 评估标准已经成功地应用在两类别的分类模型评估中, 但该标准不能直接解决多类别的评估。Srinivasan<sup>[1]</sup> 理论证明, ROC 曲线分析可以直接扩展到多类分类问题, 但由于数据空间维数高, 这种方法的精确定义难于表达和过于复杂。如何有效地将 AUC 评估方法推广到多类分类模型评估是研究的难点。

Fawcett 根据一对多“1-a-r” (one against rest) 的多类分类转成多个二类分类的思想, 提出一种多类别 AUC 评估方法 F-AUC (Fawcett-AUC)<sup>[2]</sup>。但该方法能加大类的不平衡性以及评估效率较低。根据 F-AUC 方法的思路, 提出一种新的基于二叉树构造思想的多类别评估方法 B-AUC (Binary-AUC)。

## 2 基于 ROC 曲线分析的 AUC 方法

ROC 曲线绘制在一个二维空间中如图 1, 表示正确的正例率 TPR (True Positive Rate) 和错误的正例率 FPR (False Positive Rate) 之间的变化关系<sup>[3]</sup>。

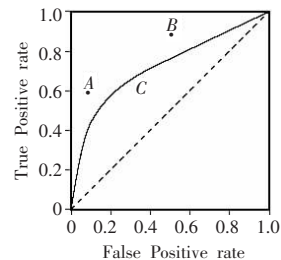


图 1 ROC 曲线

AUC 方法即计算 ROC 曲线下的面积, 用 AUC 值的大小评估分类器性能的好坏, AUC 值大的分类器性能优越, AUC 值小的分类器性能次之<sup>[4]</sup>。

**基金项目:** 安徽省教育厅自然科学研究重点项目 (the Grand Research Project of Department of Education of Anhui Province, China under Grant No. KJ2007A051); 安徽省教育厅自然科学研究项目 (the Research Project of Department of Education of Anhui Province, China under Grant No. 2006KJ061B)。

**作者简介:** 秦锋 (1962-), 教授, 硕士生导师, 研究方向为人工智能、数据挖掘、机器学习; 罗慧, 硕士, 研究方向为人工智能、数据挖掘、机器学习; 程泽凯, 副教授, 研究方向: 人工智能、数据挖掘、机器学习; 任诗流, 硕士, 研究方向为人工智能、数据挖掘、机器学习。

**收稿日期:** 2007-06-07 **修回日期:** 2007-09-07

二类分类的 AUC 值计算有多种方法, 本文采用 Hand 和 Till 提出的一种简洁计算方法<sup>[4]</sup>: 先按计算得到各测试例的后验概率大小将数据集重排成一个新的排序表, 再按如下公式计算,  $AUC = \frac{S_0 - n_0(n_0 + 1)/2}{n_0 n_1}$ , 其中  $n_0$  和  $n_1$  分别是测试数据集中正例个数和负例个数,  $s_0 = \sum r_i$ ,  $r_i$  表示第  $i$  个正例在排序表中的序号。目前, 用 AUC 方法直接评估分类器的性能适用于二类分类模型。

### 3 一种多类别的 AUC 评估方法

将二类 AUC 方法直接扩展到多类分类评估中, 存在表述空间维数高、复杂性大的问题。一般采用将多类分类转成多个二类分类的思想, 用二类 AUC 方法来评估多类分类器的性能。Fawcett 根据这种思想提出了 F-AUC 方法<sup>[2]</sup>, 该评估模型如下:

$$F-AUC = \sum_c p(i) AUC(i, rest) \quad (1)$$

其中  $AUC(i, rest)$  是计算用“1-a-r”方法得到的每个二类分类器的 AUC 值, “1-a-r”方法思想是  $k$  类分类问题构造  $k$  个二类分类器, 第  $i$  个二类分类器中用第  $i$  类的训练样本作为正例, 其他所有样本作为负例。  $p(i)$  是计算每个类在所有样本中占有的比例, 文献[6]中指出公式中乘以权值  $p(i)$ , 能使 F-AUC 值较少地独立于类和有偏数据集。

F-AUC 方法需构造  $k$  个两类分类器, 该方法存在着缺点: 第一, 每个二类分类器的训练都将全部的样本作为训练样本, 分类器的训练速度受训练样本个数影响, 该评估方法的速度相对较慢。第二, “1-a-r”方法构造二类分类器, 分类时会出现某一类小类别样本远远小于其它所有类样本总和的情况, 加大了训练样本之间的不平衡。

### 4 一种新的评估方法

本文根据二叉树的构造思想提出一种 B-AUC 的多类分类评估方法, 同样采用“1-a-r”方法将多类问题转成二类问题, 但二类分类器在训练样本时不是将所有样本每次都参加训练, 其思想如下: 首先将各个类标签按所占概率大小排列, 并将各个类标签按概率大小标识为 1, 2, ...,  $n$  类, 分类时按此类标签顺序依次构造二类分类器。对  $n$  类的训练样本, 第 1 次训练以第 1 类为正例, 将第 2, 3, ...,  $n$  类的样本为负例, 计算  $AUC_1$  的值, 二类分类后将作为正例的样本去除, 剩余所有类的样本作为下一个二类分类器的训练样本, 第  $i$  次训练时以第  $i$  类样本为正例, 将第  $i+1, i+2, \dots, n$  样本为负例, 计算出  $AUC_i$  值, 一直到构造第  $n-1$  个分类器, 以第  $n-1$  类样本为正例, 第  $n$  类样本为负例, 最后将所有二类分类器的 AUC 值乘以其正例概率值并进行累加。流程如图 2。

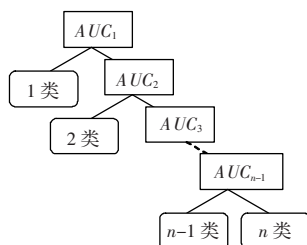


图 2 B-AUC 方法的流程

新的评估模型如下:

$$B-AUC = \sum_{i \in n} p(i) AUC(i, rest(i+1, n)) \quad (2)$$

其中  $AUC(i, rest(i+1, n))$  表示从第  $i$  类到第  $n$  类中, 以第  $i$  类为正例, 剩余所有类为负例建构二类分类模型中, 用 AUC 方法计算得到的 AUC 值。  $p(i)$  是计算每个二类分类器中的正例在所有样本中所占的概率大小。

B-AUC 方法较之 F-AUC 方法有如下优点:

(1) B-AUC 方法只需构造  $k-1$  个二类分类器, 而 F-AUC 方法需要构造  $k$  个二类分类器。

(2) B-AUC 方法构造的二类分类模型中, 每进行一次二类分类后, 将相应的二类分类模型中的正例样本去除, 除第一次外, 不需要所有样本都参加训练, 训练样本数依次减少, 提高了评估速度。而 F-AUC 方法在构造二类分类模型中每次都需要所有样本参加训练。

(3) B-AUC 方法中, 进行二类分类时, 训练样本根据类的概率大小顺序依次减少, 较之 F-AUC 方法, 小类样本的不平衡程度有所减轻。

### 5 实验结果及分析

MBNC (Bayesian Networks Classifier using Matlab)<sup>[7]</sup> 实验平台集成了多种贝叶斯分类器, 本文选用其中的朴素贝叶斯分类器 NBC (Naive Bayes Classifier) 和树扩展朴素贝叶斯分类器 TANC (Tree Augmented Naive Bayes Classifier)。在 MBNC 实验平台上扩充了 F-AUC 和 B-AUC 两种评估模块, 用它们分别评估 NBC 和 TANC。数据集是从 UCI (University of California in Irvine) 上下载, 网址是 <http://www.ics.uci.edu/~mllearn/ML-Repository.html>, 经过预处理后的数据属性如表 1。

表 1 数据集属性

| 数据集           | 属性 | 类别数 | 训练例   |
|---------------|----|-----|-------|
| Iris          | 4  | 3   | 150   |
| Waveform21    | 18 | 3   | 300   |
| Car           | 6  | 4   | 1 880 |
| Lymphograph   | 18 | 4   | 148   |
| Vehicle       | 18 | 4   | 846   |
| Glass         | 9  | 7   | 214   |
| Segment       | 18 | 7   | 2 310 |
| Shuttle_small | 8  | 7   | 5 800 |

实验结果如表 2 所示: 其中第 2 列到第 5 列表示用 F-AUC 方法得到的 NBC 和 TANC 的 AUC 值和运行时间(单位是秒); 第 6 列到第 9 列表示用 B-AUC 方法得到的 NBC 和 TANC 的 AUC 值和运行时间(单位是秒)。

表 2 实验结果

| 数据集           | 用 F-AUC 方法评估分类器 |            |            |             | 用 B-AUC 方法评估分类器 |            |            |             |
|---------------|-----------------|------------|------------|-------------|-----------------|------------|------------|-------------|
|               | NBC-F-AUC       | NBC-F-time | TANC-F-AUC | TANC-F-time | NBC-B-AUC       | NBC-B-time | TANC-B-AUC | TANC-B-time |
| Iris          | 0.992           | 4.7        | 0.988      | 5.0         | 0.991           | 3.2        | 0.986      | 4.0         |
| Waveform21    | 0.923           | 405.3      | 0.938      | 453.9       | 0.972           | 278.3      | 0.971      | 311.1       |
| Car           | 0.966           | 74.7       | 0.991      | 75.5        | 0.987           | 28.0       | 0.996      | 30.0        |
| Lymphograph   | 0.935           | 19.0       | 0.913      | 52.6        | 0.963           | 12.7       | 0.954      | 45.8        |
| Vehicle       | 0.851           | 94.7       | 0.895      | 130.7       | 0.902           | 60.6       | 0.939      | 94.5        |
| Glass         | 0.886           | 17.0       | 0.887      | 21.4        | 0.930           | 7.6        | 0.914      | 13.1        |
| Segment       | 0.981           | 419.0      | 0.991      | 469.7       | 0.987           | 244.0      | 0.992      | 297.1       |
| Shuttle_small | 0.998           | 517.0      | 1.000      | 515.9       | 0.999           | 126.7      | 1.000      | 132.8       |
| 平均            | 0.942           | 194.0      | 0.950      | 215.59      | 0.966           | 95.14      | 0.969      | 116.05      |

表3 F-AUC方法和B-AUC方法的AUC值比较表

| 分类器        | NBC-F-AUC | NBC-B-AUC |
|------------|-----------|-----------|
| TANC-F-AUC | 5-3       |           |
| TANC-B-AUC |           | 4-4       |

由表3可知,F-AUC方法中TANC的AUC值相对NBC的AUC值5高3低,B-AUC方法中TANC的AUC值相对NBC的AUC值4高4低,B-AUC方法与F-AUC方法评估分类器的结果基本一致;由表2可知,两种方法得到的TANC的平均AUC值均比NBC的平均AUC值高,可知TANC的性能要比NBC性能好,与理论上的一致,以上表明用B-AUC方法评估多类分类问题是可行的。

由表2可知,B-AUC方法评估NBC和TANC得到的平均AUC值均比F-AUC方法得到的平均值要高,可知用B-AUC方法评估多类分类问题不仅是可行的,并具有良好的分类效果。另外,通过评估运行时间的比较可知,分类速度有明显的提高。

综上可知,本文提出的B-AUC评估方法在评估多类分类问题上可行的,在评估效果和速度效果上都优于F-AUC方法,具有良好的性能。

## 6 结束语

本文根据多类分类转二类分类,提出一种基于二叉树构造思想的新评估方法B-AUC,分析该方法并将它与F-AUC方法进行比较,在MBNC实验平台上编程实现,实验结果表明新方法不仅可行,在分类评估效果和评估速度上都有良好性能。进一步的工作如下:首先,式(2)中的权值 $p(i)$ 有待进一步的优化;其次,二叉树的构造思想对分类器性能影响很大,怎样用启发式的二叉树思想构造AUC多类评估,还需进一步研究。

(上接124页)

由于计算机性能和工作量的限制,实验所分割的体数据是原始分辨率的一半,尺寸为1024×1024×2133分辨率的体数据,先通过对256×256×534分辨率数据浏览,粗定包含肝脏器官的最小外接立方体,然后从1024×1024×2133分辨率的体数据中确定肝脏所在的精确位置: $y$ 方向从243到746, $x$ 方向从237到777,层片从741到1219,利用裁剪功能分割出包含整个肝脏器官的所有层片,获得一组541×504×479子数据集。图2是通过ITK工具包中的NeighborhoodConnected算法对该组中的一层图片进行分割的结果。

图4显示的是通过绘图工具对分割结果进行的修改,交互工具有画笔,橡皮,形态学膨胀,腐蚀,开操作,填充,选择等。用户可以对图像进行放大,在像素级对轮廓边缘进行修改。

图5是得到的肝脏表面分割结果。在处理相邻的图片时,通过继承上一层的表面轮廓或者分割参数可以减少工作量,提高分割速度。最后将轮廓边缘点的坐标存盘,为后面数据人数据库的建立和可视化打基础。

## 4 结论

为了对数字人器官进行三维重建,需要对彩色图像进行分割,获得其表面轮廓。本文采用一个交互式分割方法,通过使用

## 参考文献:

- [1] Srinivasan A. Note on the location of optimal classifiers in  $N$ -dimensional ROC space, PRG-TR-2-99 [R]. Oxford University Computing Laboratory, Wolfson Building, Parks Road, Oxford, 1999.
- [2] Fawcett T. Using rule sets to maximize ROC performance [C]//2001 IEEE International Conference on Data Mining (ICDM-01), San Jose, California, USA, 2001: 131-138.
- [3] Fawcett T. ROC graphs: notes and practical considerations for data mining researchers, HPL-2003-4 [R]. Palo Alto: HP Laboratories, 2003.
- [4] Hand D J, Till R J. A simple generalization of the area under the ROC curve for multiple class classification problems [J]. Machine Learning, 2001(45): 171-186.
- [5] Lachiche N, Flach P. Improving accuracy and cost of two-class and multi-class probabilistic classifiers using ROC curves [C]//The Twentieth International Conference on Machine Learning (ICML-2003), Washington DC, 2003: 416-423.
- [6] Ferri C, Hernandez-Orallo J, Salido M. Volume under ROC surface for multi-class problems [C]//European Conference on Machine Learning, Dubrovnik, Croatia, 2003: 108-120.
- [7] 程泽凯, 林士敏, 陆玉昌, 等. 基于 Matlab 的贝叶斯分类器平台 MB-NC [J]. 复旦学报, 2004, 43(5): 729-732.
- [8] Lane T. Extensions of ROC Analysis to multi-class domains [C]//ICML-2000 Workshop on Cost-sensitive Learning, 2000.
- [9] 安金龙, 王正欧, 马振平. 一种新的支持向量机多类分类方法 [J]. 信息与控制, 2004(3).

ITK 工具包和利用 MATLAB 提供的图像处理工具, 来对图像进行分割, 然后从中选取感兴趣器官的轮廓, 获得所需人体器官的表面。该平台具有很好的开放性, 能方便地将 ITK 工具包中最新的分割算法融入平台, 实现一定程度的自动分割, 又能让用户在非常友好的界面中在像素级对自动分割的结果进行修改、存盘。该交互分割平台为数字人的深入应用和其它类似图像分割应用提供了一个高效的工具。

## 参考文献:

- [1] 钟世镇. 数字人和数字解剖学 [M]. 济南: 山东科学技术出版社, 2004.
- [2] Schiemann T, Bomans M, Tiede U, et al. Interactive 3D-segmentation [C]//Robb R A. Visualization in Biomedical Computing II, Proc SPIE 1808, Chapel Hill, NC, 1992: 376-383.
- [3] Bomans M, Höhne K H, Tiede U, et al. 3D-segmentation of MR-images of the head for 3D-display [J]. IEEE Trans Med Imaging, 1990, 9(2): 177-183.
- [4] Haralick R M, Sternberg S R, Zhuang X. Image analysis using mathematical morphology [J]. IEEE Trans Pattern Anal Machine Intell, 1987, 9: 532-550.
- [5] 唐雷, 原林, 洪辉文, 等. 中国数字人女婴 1 号数据集构建报告 [J]. 中国临床解剖学杂志, 2004, 22(1): 98-100.