

基于可拒识-双层支持向量分类器的 微钙化点检测

胡正平 张 晔

(哈尔滨工业大学信息工程系, 哈尔滨 150001)

摘 要 为克服医学图像微钙化点检测中假阳性高的缺点, 构造了一种带拒识能力的双层支持向量模型分类器, 用于钙化点检测。检测时, 首先利用基于最大间隔超平面的支持向量分类器(SVC)对输入模式进行分类判决; 然后通过求取真实钙化点样本特征空间最小的包含球形边界来得到钙化点样本的球形支持向量域表示(SVDD); 接着利用钙化点的支持向量域表示对输入模式进行拒识或接受处理; 最后利用 SVC 与 SVDD 两个分类器的结果来进行综合判决。仿真实验结果表明, 该算法在不影响微钙化点的检出率的情况下, 可部分解决假阳性高的问题。

关键词 支持向量分类器 微钙化点检测 支持向量域描述 拒识性能

中图分类号: TP391.41 R445 **文献标识码**: A **文章编号**: 1006-8961(2006)05-0652-04

Micro-calcifications Detection Based on Two-layer Support Vector Classifier with Rejection Feature

HU Zheng-ping, ZHANG Ye

(Department of Information Engineering, Harbin Institute of Technology, Harbin 150001)

Abstract To solve the problem of false positive in micro-calcification detection, a two-layer support vector classifier model with rejection feature (TLVCRF) is proposed. Firstly the first layer of support vector classifier (SVC) with maximum margin between two classes will be used for classifying the input pattern; then the sphere support vectors of true micro-calcification points to describe the distribution of the sample were obtained by searching all the sphere boundaries containing the samples of this class. Then the input pattern of no-object classes could be rejected by the second support vector domain description (SVDD). Lastly the results of SVC and SVDD classifier are integrated to obtain the right results. Experimental results demonstrate that the method of two-layer support vector classifier with rejection feature performs better in achieving lower false positive.

Keywords support vector classifier (SVC), micro-calcification detection, support vector data description (SVDD), rejection performance

1 引 言

不断提高乳腺 X 光透视图像微钙化点的正确检测率是解决乳腺癌诊断的关键, 为此这一问题已成为众多学者追求的目标。微钙化点检测与识别就

是从乳腺图像中抽取微钙化特征向量, 并在此基础上构建合适的分类器。常用的人工神经网络、贝叶斯分类器等都是基于传统统计学的方法, 但它们在样本数量有限的情况下表现不尽人意。基于统计学习理论的支持向量分类器 (support vector classifier, SVC) 因为其优越的性能而成为近年研究的热点, 它

基金项目: 国家自然科学基金资助项目(60272073); 河北省科技发展指导项目(Z2005310)

收稿日期: 2005-03-10; **改回日期**: 2005-06-10

第一作者简介: 胡正平(1970 ~), 男, 哈尔滨工业大学博士研究生, 燕山大学通信电子工程系教师。研究方向为统计学习理论与模式识别、医学图像处理。E-mail: tnpochw@263.net

在逼近与分类应用领域取得了极大的成功。SVC算法^[1]是Vapnik提出的一种有监督学习结构风险最小化的两类分类器。一些学者将SVC引入微钙化点检测中并改进了检测效果^[2,3]。在乳腺X光透视图象中,钙化点主要表现为一系列较周围正常组织亮的小区域,但由于其半径小,常常淹没于背景组织和噪声中,这是导致假阳性高的原因。经典的SVC分类器总是假定测试样本属于两类之一,然而在钙化点分类判决问题中,乳腺X光透视图象往往存在许多种类的非目标样本,这时由于经典SVC分类器缺乏拒识判决的能力,所以并不能很好解决假阳性高的问题。为了解决这个问题,本文提出一种带拒识能力的双层支持向量模型分类器,用于钙化点检测,它能够充分利用已知钙化点训练样本分布信息来提高拒识非目标样本的能力。David等针对单类数据提出了支持向量域描述(support vector domain description, SVDD)^[4],其主要用来进行数据描述和剔除奇异点。SVDD的基本思想就是通过计算包含一组数据的最小超球体边界来对数据的分布范围进行描述^[4]。该方法给人的启发是:如果能够知道一部分样本的分布范围,那么就可以利用它来构造带拒识能力的分类器。SVDD虽然具有良好的数据选择能力,但是缺乏对比鉴别分类能力(因为它仅仅针对一类数据本身);而SVC虽缺乏数据分布描述能力,却具有最优的对比鉴别能力。因此同时吸取二者的优点构造一种新的具有良好拒识能力的分类器,用来解决微钙化点检测假阳性问题是本文工作的出发点。

针对可拒识分类问题,不少学者也提出了一些有效的解决问题的思路,如文献[5]提出基于支持向量表示-鉴别机(support vector representation and discrimination machine, SVRDM)的具有拒识能力的分类方法,该方法的优点是可一步实现拒识分类,其缺点是控制拒绝能力的门限参数选择比较复杂,不具有自适应能力;文献[6]提出另外一种类似于本文的两步实现策略,即首先将两类归于一类,其第1步决定是否拒绝,第2步才进行分类,该方法缺点是第1步需要知道非目标样本,并需进行训练,这一方面计算量巨大,另一方面在实际应用中往往不容易收集到全部的非目标样本,而本文提出的方法则不需要全部非目标样本,其只需要知道部分两类目标样本本身即可。文献[7]提出了基于模板等预处理结合SVC的分类方法,其预处理的过程就是数据的粗选择过程,该方法对于可模板化描述的问题是

合理有效的,其缺陷是对于不能模板化的数据则无能为力,同时门限选择困难。

2 系统框图

本文构造的可拒识双层支持向量模型分类器如图1所示,该系统主要由基于最大间隔的超平面鉴别分类器和基于SVDD的数据选择器两个部分组成,其各自的支持向量可通过样本训练而得到。

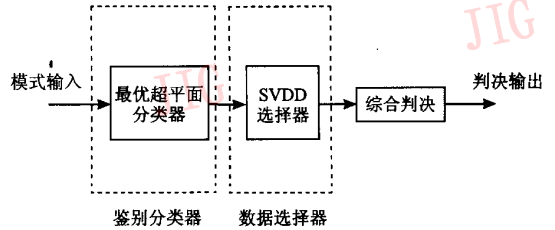


图1 带拒识能力的双层支持向量模型分类器框图
Fig. 1 The principle frame of two-layer support vector machine

3 带拒识能力的双层支持向量模型分类器

3.1 支持向量域数据选择器

SVDD是通过寻找包含一组数据的最小半径 R 的超球体来对其进行描述^[2],可见,它可对一组数据的分布进行描述,对于非本类的奇异数据点,它应该位于超球体的外面。为了减少奇异点数据的影响,本文通过引入松弛变量 ξ_i 和惩罚因子 C 来进行描述(即允许存在错误),以便使求最小超球体体积的问题可以转化为下面的二次规划问题:

$$\min \left[R^2 + C \sum_i \xi_i \right] \quad (1)$$

其约束条件为

$$(x_i - a)^T(x_i - a) \leq R^2 + \xi_i \quad (2)$$

其中, a 为球体中心,而 $\xi_i \geq 0$ 上面的优化问题,其通过引入拉格朗日系数其变为

$$\min L = R^2 + C \sum_i \xi_i - \beta_i \sum_i \xi_i \sum_i \alpha_i (R^2 + \xi_i - (x_i - a)^T(x_i - a)) \quad (3)$$

其中, $\alpha_i, \beta_i \geq 0$ 。

则求式(3)的最小值可转化为如下的对偶问题

$$\max F = \sum_i \alpha_i K(x_i, x_i) - \sum_i \sum_j \alpha_i \alpha_j K(x_i, x_j) \quad (4)$$

其中, $\sum_i \alpha_i = 1, 0 \leq \alpha_i \leq C, K(\cdot)$ 为核函数。通常情况下, 即使去除了奇异点样本, 钙化点样本依然不会呈现球状分布, 这时可以通过引入核函数将它映射到高维空间来进行讨论。

对于两类问题(也可扩展到多类问题中)的情况, 设有训练样本 $(x_i, y_i) (i=1, 2, \dots, n, y_i \in \{-1, +1\})$, 对于 $y_i = 1$ 的钙化点样本如何来建立超球体, 可通过求解式(4)分别计算其与拉格朗日系数对应的支持向量来实现。

如果一测试钙化点样本 z 满足如下条件, 则这个点被接受, 否则拒识判决。

$$K(z, z) - 2 \sum_i \alpha_i^{(1)} K(x_i, z) + \sum_{i,j} \alpha_i^{(1)} \alpha_j^{(1)} K(x_i, x_j) \leq R^2 \quad (5)$$

这里 R 表示与包含真实钙化点的空间对应的特征空间分布的最小球形半径, 当输入样本包含在球形空间内部时, 则为真实钙化点, 否则认为是非钙化点。

3.2 最大间隔超平面分类鉴别器

设 $x_i (i=1, \dots, p)$ 为训练样本, 这里 x_i 为 N 维矢量, p 为训练样本数, $y_i (+1, -1)$ 为类标记。如果输入数据线性可分, 则其决策函数为

$$D(x) = w^T g(x) + b \quad (6)$$

$g(x)$ 为一非线性映射函数, 用其将 $x (N$ 维) 映射到 l 维空间, w 为 l 维权值矢量, b 为偏移量。

对于线性可分的数据, 决策函数满足

$$y_i (w^T g(x_i) + b) \geq 1 \quad i=1, \dots, N \quad (7)$$

如果训练数据线性不可分, 且同时引入松弛变量 $\xi_i (>0)$, 则决策函数变为

$$y_i (w^T g(x_i) + b) \geq 1 - \xi_i \quad i=1, \dots, N \quad (8)$$

$$\xi_i \geq 0$$

与一阶函数软间隔优化问题对应的拉格朗日函数为

$$\max L = \frac{1}{2} \langle w, w \rangle + C \sum_i \xi_i - \sum_i \beta_i \xi_i - \sum_i \alpha_i [y_i (\langle g(x_i), w \rangle + b) - 1 + \xi_i] \quad (9)$$

式(9)中第 1 项用于控制分类器的复杂度, 第 2 项用于控制错误分类的经验误差, C 为常数, $\alpha_i, \beta_i \geq 0$ 。对偶表示可以通过求对应于 w, ξ, b 的偏导、置零, 并带入式(9)来得到

$$\max L = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} y_i y_j \alpha_i \alpha_j \langle g(x_i), g(x_j) \rangle \quad (10)$$

若通过引入核函数, 则上式变为

$$\max L = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} y_i y_j \alpha_i \alpha_j K(x_i, x_j) \quad (11)$$

其中, $\sum_i \alpha_i y_i = 1$, 且 $0 \leq \alpha_i \leq C$, 而判决输出函数为

$$f(x) = \text{sgn} \left(\sum_i y_i \alpha_i K(x_i, x) + b \right) \quad (12)$$

4 可拒识分类器及其判决规则

本文构造的双层可拒识支持向量分类器最终规则如下: 对于输入检测模式 x , 可首先利用 SVC 分类器进行分类, 如果判决为钙化点; 然后利用钙化点 SVDD 数据选择器进行判决, 如果输入的钙化点样本通过两个判决器, 则判决为真正的钙化点; 否则判决为非钙化点。

5 仿真实验

为了验证本文提出的方法在乳腺钙化点自动检测的效果, 本文选择 MIAS (mammographic image analysis society) 数据库的钙化点进行了仿真实验, 该数据库钙化区域由专家肉眼标定, 通过有经验的专家标定, 确定了 20 幅乳腺图像的 107 个钙化点, 同时由“自举法”收集到 201 个非钙化点非钙化点作为训练样本。同时选取 150 个钙化点可疑样本作为测试样本。本文首先利用迭代中值滤波器结合自适应门限来对钙化点进行粗检测, 然后抽取这些可能的钙化点的相对灰度、对比度、面积、方差、邻域可疑钙化点平均对比度、距离分布方差、面积方差等 7 个特征参数组成特征矢量。

检测时, 首先通过对鉴别分类器 SVC 进行训练来得到其支持向量(高斯核函数), 然后对真实的钙化点利用 SVDD 训练来构造钙化点数据选择器。钙化点分类器训练以及检测算法框图如图 2、图 3 所示。为了验证本文构造算法的效果, 还分别将它与直接 SVC 检测方法、神经网络 (artificial neural network, ANN) 检测方法进行了对比实验, 对比结果如图 4 和表 1 所示。

实验结果表明, 本文提出的基于可拒识-双层支持向量模型分类器, 用于钙化点检测能够较好地克服假阳性的影响, 相对于直接利用 SVM (support vector machine) 与神经网络进行分类具有更好的效果。其原因主要在于本文提出的可拒识-双层支持向量模型分类器不但充分利用了钙化点-非钙化点的区别信息, 而且充分利用了钙化点特征自身的分布信息。

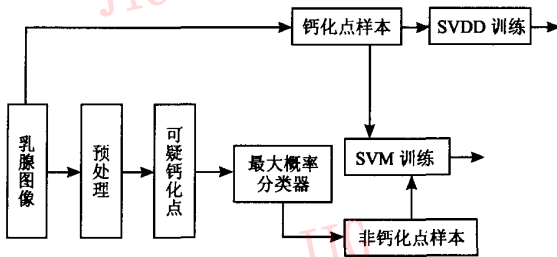


图 2 钙化点检测可拒识双层分类器训练原理框图
 Fig.2 The principle frame of micro-calcification detection training model based on two-layer support vector classifier with rejection feature

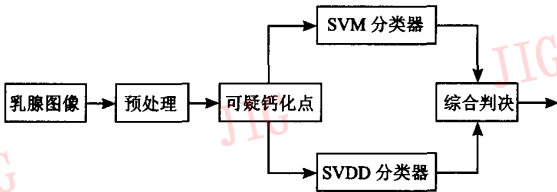


图 3 钙化点识别原理框图
 Fig.3 The principle frame of micro-calcification recognition

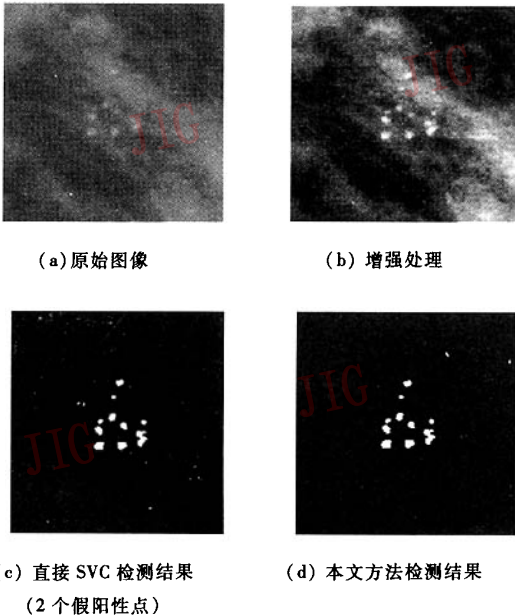


图 4 检测结果显示
 Fig.4 Detection results display

表 1 几种钙化点检测方法性能对比

Tab.1 Performance comparison based on different methods

性能指标	方法		
	直接 SVM 法	ANN 法	本文方法
检出率 (%)	96.3	92.7	96.1
假阳性率 (FPR) (%)	24.9	28.5	10.6

注: FPR (false position rate) 表示将非钙化点判决为钙化点的比率。

6 结 论

由于基于 SVDD 的球形支持向量具有很好的数据分布描述能力,其能够较好地提供每一类数据分布的紧凑描述;而 SVC 则能够以最大间隔提供两类数据的分类鉴别能力,其缺点是不能提供数据的分布范围描述,为此本文将二者的优点相结合,提出了带拒识能力的两层支持向量模型,用于乳腺 X 射线透视图像钙化点的自动检测,实验结果的对比表明,本文提出的方法在不影响微钙化点的检出率的情况下,由于充分利用了真实钙化点的特征分布信息,从而部分解决了假阳性高的问题。

参考文献 (References)

- 1 Vapnik Vladimir N. The nature of statistical learning theory [M]. New York, NY: Wiley, 1998, chapter 5.
- 2 WAN Bai-kun, WANG Rui-ping, ZHU Xin, et al. Principles of SVM and its application in micro-calcifications detection in mammogram [J]. Acta Electronic Sinica, 2004, 32(4): 587 ~ 590. [万柏坤,王瑞平,朱欣等. SVM 算法及其在乳腺 X 片微钙化点自动检测中的应用 [J]. 电子学报, 2004, 32(4):587 ~ 590.]
- 3 EI-Naqa Issam, YANG Yong-yi, Wernick Miles N, et al. A support vector machine approach for detection of microcalcifications [J]. IEEE Transactions on Medical Imaging, 2002, 21(12): 1552 ~ 1563.
- 4 Tax David M J, Duin Robert P W. Support vector data description [J]. Machine Learning, 2004, 54(1):45 ~ 66.
- 5 CHAO Yuan, David Casasent. A novel support vector classifier with better rejection performance [A]. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition [C], Madison, WI, United States, 2003:1063 ~ 1069.
- 6 Rizvi S A, Saasawi T N, Nasrabadi N M. A clutter rejection technique for FLIR imagery using region based principal component analysis [J]. Pattern Recognition, 2000, 33(11):1931 ~ 1933.
- 7 YANG Guang-zheng, Huang Thomas S. Human face detection in a complex background [J]. Pattern Recognition, 1994, 27(1): 58 ~ 63.