核零空间方法在乳腺癌异常检测中的应用

韩　笑[1],毕　波[1,2],唐锦萍[3],曹　莉[2]
(1. 东北石油大学 数学与统计学院,黑龙江 大庆 163318;
2. 海南医学院公共卫生学院,海南 海口 571101;
3. 黑龙江大学 数据科学与技术学院,黑龙江 哈尔滨 150080)

摘　要:当今时代,乳腺癌越来越成为了女性的高发病,因此尽早地排除异常因素,进行对症治疗,可以大大降低疾病风险。考虑到乳腺癌数据特征比较多,并且往往不仅存在线性特征还隐含着很多非线性特征,针对这一问题提出利用核零空间算法来进行乳腺癌的异常检测。首先利用核函数将所有的正常样本进行非线性映射变换到高维空间,再通过零空间变换将类内散度转换为 0,并且将零空间中整个类的数据用该类的平均值代替,最后通过计算测试样本到该值的距离判断测试样本的异常性。该算法大大降低了计算的复杂性,也提高了乳腺癌检测的速度。通过在 UCI 乳腺癌数据库上的仿真实验,并对不同核函数以及设定的不同异常阈值下得到的 F1-score 进行对比,发现在不同核函数以及不同异常阈值下的结果是不同的,且在选取高斯核作为核函数时,可使得 F1-score 结果达到 0.962 7。充分证明了将核零空间算法用于乳腺癌异常检测是有效的。

关键词:乳腺癌;异常检测;核零空间算法;核函数;异常阈值
中图分类号:TP39　　**文献标识码**:A　　**文章编号**:1673-629X(2022)01-0165-05
doi:10.3969/j.issn.1673-629X.2022.01.028

Application of Kernel Null Space Method in Breast Cancer Abnormal Detection

HAN Xiao[1], BI Bo[1,2], TANG Jin-ping[3], CAO Li[2]
(1. School of Mathematics and Statistics, Northeast Petroleum University, Daqing 163318, China;
2. School of Public Health, Hainan Medical College, Haikou 571101, China;
3. School of Data Science and Technology, Heilongjiang University, Harbin 150080, China)

Abstract: Nowadays, breast cancer has increasingly become a high incidence of women. Therefore, removing abnormal factors as soon as possible and conducting diagnosis and treatment can greatly reduce the risk of disease. Considering that there are many features in breast cancer data, and there are often not only linear features but also many non-linear features. To solve this problem, a kernel null space algorithm is proposed to detect abnormalities of breast cancer. Firstly, the kernel function is adopted to perform nonlinear mapping and transformation of all normal samples into high-dimensional space. Secondly, the intra-class divergence is converted to 0 through zero space transformation, and the data of the entire class in the zero space is replaced with the average value of the class. Finally, the abnormality of the test sample is judged by calculating the distance from the test sample to the value. The proposed algorithm greatly reduces the complexity of the calculation and also improves the speed of breast cancer detection. Through simulation experiments on the UCI breast cancer database, the F1-score obtained under different kernel functions and different set abnormal thresholds is compared. It is found that the results under different kernel functions and different abnormal thresholds are different, and when the Gaussian kernel is selected as the kernel function, the F1-score can reach 0.962 7, which fully proves that the kernel null space algorithm is effective in breast cancer abnormality detection.

Key words: breast cancer; abnormal detection; kernel null space method; kernel function; abnormal threshold

收稿日期:2021-02-02　　**修回日期**:2021-06-03
基金项目:国家自然科学基金(11701159);2020 年海南省基础与应用基础研究计划(自然科学领域)高层次人才项目基金(820RC649)
作者简介:韩　笑(1996-),女,硕士研究生,研究方向为人工智能、机器学习、应用数学;毕　波,副教授,研究方向为人工智能、机器学习、应用数学。

0 引 言

当今时代,乳腺癌已成为女性最为常见的恶性肿瘤,其发病率在全球范围内均持续增长,每年的确诊人数约高达 28 万,具有较高的死亡率,并且越来越倾向于年轻化[1-5]。临床研究表明,乳腺癌的演变过程大致可以概括为五个步骤,由一开始乳腺的良性病变,之后乳腺良性增生,乳腺不典型增生,再到后来的乳腺原位癌,到最后的浸润性的乳腺癌,但并不是所有的患者都一定会按照这样的规律逐渐演变,有时在临床当中也可能会发现跳跃式的演变。因此,要想预防乳腺癌或得到早期的治疗,就必须及早地进行检测,发现异常,采取相应的应对措施。

核零空间算法作为一种单分类算法,经常用来进行异常检测。起初它是源于线性判别分析(LDA)的,利用最大化 Fisher 准则的思想,将所有的样本点通过某种线性变换(即 FST 变换),达到最小化类内散度,最大化类间散度的目的[6-8]。之后将类内距离变为 0,提出了零空间变换(即 NFST 变换)[9]。但是,这两种变换都是仅仅考虑了数据的线性特征,而数据往往还存在许多非线性特征,因此提出了该变换的核化方法,即 KNFST 变换[10]。首先利用核函数将数据进行非线性映射变换到高维空间,然后再利用 NFST 变换思想,提取使得类内散度为 0,且类间散度最大的特征方向,即提取零投影方向。乳腺癌样本往往具有多个显式的线性特征,但也具有很多观测不到的隐式的非线性特征,因此为了更好地提取样本的非线性特征,提高乳腺癌样本数据的异常识别率,利用核零空间算法对乳腺癌数据进行异常检测。

该文总结了零空间方法以及核零空间算法的计算步骤,有效提取了样本数据的非线性特征。将核零空间算法用于 UCI 数据库中的乳腺癌数据集进行仿真实验,通过对比不同核函数以及不同异常阈值下的异常识别率,充分证明了将核零空间算法用于乳腺癌异常检测的有效性。

1 核零空间方法

若 $X=\{x_{11},x_{12},\cdots,x_{1n_1},x_{21},x_{22},\cdots,x_{2n_2},\cdots,x_{c1},x_{c2},\cdots,x_{cn_c}\}$ 为一个含有 N 个样本的数据集,其中 $X\in R^n$,设 $X_1=\{x_{11},x_{12},\cdots,x_{1n_1}\}$,$X_2=\{x_{21},x_{22},\cdots,x_{2n_2}\}$,…,$X_c=\{x_{c1},x_{c2},\cdots,x_{cn_c}\}$ 为 C 个不同的类,其中第 i 类的类内均值为:

$$\overline{X_i}=\frac{1}{n_i}\sum_{j=1}^{n_i}x_j \tag{1}$$

总均值为:

$$\bar{X}=\frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{n_i}x_{ij} \tag{2}$$

1.1 Fisher 准则

假设 φ 为 LDA 算法的投影矩阵,则 Fisher 准则公式为[11]:

$$J(\varphi)=\frac{\varphi^{\mathrm{T}}\boldsymbol{S}_{\mathrm{b}}\varphi}{\varphi^{\mathrm{T}}\boldsymbol{S}_{\mathrm{w}}\varphi} \tag{3}$$

考虑到基于最小化类内散度,最大化类间散度的特点,需要计算类内方差矩阵与类间方差矩阵,分别将类内散度矩阵 $\boldsymbol{S}_{\mathrm{w}}$ 与类间散度矩阵 $\boldsymbol{S}_{\mathrm{b}}$ 定义如下:

$$\boldsymbol{S}_{\mathrm{w}}=\frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{n_i}(x_{ij}-\overline{X_i})(x_{ij}-\overline{X_i})^{\mathrm{T}} \tag{4}$$

$$\boldsymbol{S}_{\mathrm{b}}=\frac{1}{N}\sum_{i=1}^{c}n_i(\overline{X_i}-\bar{X})(\overline{X_i}-\bar{X})^{\mathrm{T}} \tag{5}$$

首先,基于最大化 Fisher 准则进行 FST 变换。设 φ 是通过计算 $\boldsymbol{S}_{\mathrm{b}}\boldsymbol{S}_{\mathrm{w}}^{-1}$ 的前 P 个最大特征值对应的特征向量所得到的投影矩阵,即 $\varphi=\{\varphi_1,\varphi_2,\cdots,\varphi_p\}$,其中 φ_p 为投影矩阵的第 P 个投影向量。

1.2 NFST 变换

由于很多时候,样本数往往远远小于数据的特征维数,这时会导致 $\boldsymbol{S}_{\mathrm{w}}$ 奇异。为了解决这个问题,将类内方差规定为 0,因此提出了零空间线性变换,即(NFST 变换)。则令:

$$\varphi^{\mathrm{T}}\boldsymbol{S}_{\mathrm{w}}\varphi=0 \tag{6}$$

$$\varphi^{\mathrm{T}}\boldsymbol{S}_{\mathrm{b}}\varphi>0 \tag{7}$$

为了同时满足上述两个条件,在计算时引入了总散度矩阵 $\boldsymbol{S}_{\mathrm{t}}$,令:

$$\boldsymbol{S}_{\mathrm{t}}=\boldsymbol{S}_{\mathrm{w}}+\boldsymbol{S}_{\mathrm{b}} \tag{8}$$

零投影方向计算步骤如下[12]:

第一步:计算 $\boldsymbol{S}_{\mathrm{t}}$ 的非零特征值对应的特征向量组成的投影矩阵,即存在一个由非零特征值对应的特征向量组成的投影矩阵 $\boldsymbol{P}$,使得,$\boldsymbol{S}'_t=\boldsymbol{P}^{\mathrm{T}}\boldsymbol{S}_{\mathrm{t}}\boldsymbol{P}$,则也有 $\boldsymbol{S}'_{\mathrm{w}}=\boldsymbol{P}^{\mathrm{T}}\boldsymbol{S}_{\mathrm{w}}\boldsymbol{P}$,$\boldsymbol{S}'_{\mathrm{b}}=\boldsymbol{P}^{\mathrm{T}}\boldsymbol{S}_{\mathrm{b}}\boldsymbol{P}$。

第二步:计算 $\boldsymbol{S}'_{\mathrm{w}}$ 的零特征值对应的特征向量组成的投影矩阵,即存在一个零投影矩阵 $\boldsymbol{U}$,使得 $\boldsymbol{S}''_{\mathrm{w}}=\boldsymbol{U}^{\mathrm{T}}\boldsymbol{S}'_{\mathrm{w}}\boldsymbol{U}=0$。

第三步:得到总的零投影矩阵 $\varphi=\boldsymbol{P}\boldsymbol{U}$。

1.3 KNFST 变换

通过零空间变换,得到的也仅仅是使得分类结果最优的线性特征组成的投影矩阵,但是很多数据集往往不仅具有显式的线性特征,而且具有隐式的非线性特征,因此,就提出了核零空间变换(KNFST 变换)。这时就需要先利用核函数将低维数据映射到高维非线性特征空间,然后再进行零投影矩阵的计算。设经过非线性映射后的特征空间为 F,非线性映射后的样本为 $\varphi(X)$,则此时有:

第 i 类的类内均值为:

$$\overline{\varphi(X_i)} = \frac{1}{n_i}\sum_{j=1}^{n_i}\varphi(x_j) \tag{9}$$

总均值为:

$$\overline{\varphi(X)} = \frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{n_i}\varphi(x_{ij}) \tag{10}$$

类内方差矩阵 $\boldsymbol{S}_{\mathrm{w}}^{\varphi}$ 为:

$$\boldsymbol{S}_{\mathrm{w}}^{\varphi} = \frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{n_i}(\varphi(x_{ij}) - \overline{\varphi(X_i)})(\varphi(x_{ij}) - \overline{\varphi(X_i)})^{\mathrm{T}} \tag{11}$$

类间方差矩阵 $\boldsymbol{S}_{\mathrm{b}}^{\varphi}$ 为:

$$\boldsymbol{S}_{\mathrm{b}}^{\varphi} = \frac{1}{N}\sum_{i=1}^{c} n_i(\overline{\varphi(X_i)} - \overline{\varphi(X)})(\overline{\varphi(X_i)} - \overline{\varphi(X)})^{\mathrm{T}} \tag{12}$$

总方差矩阵 $\boldsymbol{S}_{t}^{\varphi}$ 为:

$$\boldsymbol{S}_{\mathrm{t}}^{\varphi} = \boldsymbol{S}_{\mathrm{w}}^{\varphi} + \boldsymbol{S}_{\mathrm{b}}^{\varphi} \tag{13}$$

由于为了提取数据的非线性特征,将其利用核函数映射到了高维空间,因此,在计算时需要对映射后的数据进行零方向投影,现在需要计算核矩阵的类内方差、类间方差。其中核矩阵中的每一个元素都可以表示为样本间的内积形式:

$$K = (\varphi(x_{ij}),\varphi(x_{kl})) = <\varphi(x_{ij}),\varphi(x_{kl})> \tag{14}$$

其中,$k = 1,2,\cdots,c$,$l = 1,2,\cdots,n_k$。

核类内方差矩阵为:

$$\boldsymbol{K}_{\mathrm{w}}^{\varphi} = \frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{n_i}(K(x_{ij}) - \overline{K(X_i)})(K(x_{ij}) - \overline{K(X_i)})^{\mathrm{T}} \tag{15}$$

核类间方差矩阵为:

$$\boldsymbol{K}_{\mathrm{b}}^{\varphi} = \frac{1}{N}\sum_{i=1}^{c} n_i(\overline{K(X_i)} - \overline{K(X)})(\overline{K(X_i)} - \overline{K(X)})^{\mathrm{T}} \tag{16}$$

核总方差矩阵为:

$$\boldsymbol{K}_{\mathrm{t}}^{\varphi} = \boldsymbol{K}_{\mathrm{w}}^{\varphi} + \boldsymbol{K}_{\mathrm{b}}^{\varphi} \tag{17}$$

零投影方向具体计算步骤如下:

第一步:非线性空间投影,$X \in R^n \rightarrow \varphi(X) \in F$。

第二步:计算总方差矩阵 $\boldsymbol{K}_{\mathrm{t}}^{\varphi}$ 的非零特征值对应的特征向量组成的投影矩阵,即存在一个由非零特征值对应的特征向量组成的投影矩阵 $\boldsymbol{P}^{\varphi}$,使得 $\boldsymbol{K}_{\mathrm{t}}^{\varphi'} = \boldsymbol{P}^{\varphi^{\mathrm{T}}}\boldsymbol{K}_{\mathrm{t}}^{\varphi}\boldsymbol{P}^{\varphi}$,则也有 $\boldsymbol{K}_{\mathrm{t}}^{\varphi'} = \boldsymbol{P}^{\varphi^{\mathrm{T}}}\boldsymbol{K}_{\mathrm{w}}^{\varphi}\boldsymbol{P}^{\varphi}$,$\boldsymbol{K}_{\mathrm{b}}^{\varphi'} = \boldsymbol{P}^{\varphi^{\mathrm{T}}}\boldsymbol{K}_{\mathrm{b}}^{\varphi}\boldsymbol{P}^{\varphi}$。

第三步:计算 $\boldsymbol{K}_{\mathrm{w}}^{\varphi'}$ 的零特征值对应的特征向量组成的投影矩阵,即存在一个零投影矩阵 $\boldsymbol{U}$,使得 $\boldsymbol{K}_{\mathrm{w}}^{\varphi''} = \boldsymbol{U}^{\varphi^{\mathrm{T}}}\boldsymbol{K}_{\mathrm{w}}^{\varphi'}\boldsymbol{U}^{\varphi} = 0$。

第四步:得到总的零投影矩阵 $\boldsymbol{\varphi}^{\varphi} = \boldsymbol{P}^{\varphi}\boldsymbol{U}^{\varphi}$。

在得到零投影矩阵以后,首先将训练集的核投影矩阵按照零投影矩阵的方向,将整个训练集投影为一个单点,之后将测试集中的每一个样本先按照非线性映射方向,得到测试集在非线性映射方向的核投影矩阵,然后再将其按照零投影矩阵的方向投影到零空间上的单个点,最后计算零空间上每个测试点到正常点样本的距离,并且通过判断该距离与事先设定的异常阈值的大小,来判断测试样本是否为异常样本。

2　实验应用分析

由于医疗行业的特殊性,时时刻刻都在产生海量的医疗数据,数据挖掘和机器学习技术为这些海量医疗数据的分析和应用提供了新的思路和手段[13]。通过读取乳房 X 光造影的测量指标,用机器学习算法来检测乳腺癌,是目前人工智能和医学领域交叉的研究热点[14-15]。

该文选取核零空间算法对乳腺细胞的各项测量数据进行异常检测,但是乳腺癌样本数据的特征维数很多,因此如何有效地进行非线性映射是取得满意结果的关键。运用核零空间算法进行乳腺癌异常检测主要依赖于核函数的选取以及核函数参数和异常阈值的设置。下面分别对乳腺癌数据在不同核函数、不同核参数和不同异常阈值下的 F1-score 做了对比,并且得出了结论。

2.1　数据准备

选取 UCI 数据库中的 breast-Cancer 数据集作为实验数据集,它一共包含 699 个样本数据,其中良性样本数据有 458 个,恶性样本数据有 241 个,其中每个样本都含有 9 个特征,分别为:Clump Thickness, Uniformity of Cell Size, Uniformity of Cell Shape, Marginal Adhesion, Single Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nucleoli, Mitoses。除此之外,还含有 16 处缺失值。

(1)数据缺失值处理:利用缺失值所在列的平均值填充缺失值。

(2)数据归一化处理:数据特征不同,对应的值可能存在的差异特别大,因此为了减小这种影响,对数据先进行归一化处理,将每个值都设定在[0,1]范围内。这里分别对每个样本的每个特征都做归一化处理,即:

$$\max\_\min = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{18}$$

其中,$x_{\min}$ 表示每一列的最小值,$x_{\max}$ 表示每一列的最大值。

(3)数据集划分:对归一化后的数据进行测试集与训练集的划分,将所有正常样本的 70% 作为训练集,剩下的 30% 作为测试集的一部分,然后再将所有的异常样本放入测试集中,共同组成完整的测试集。

2.2　核函数选取

分别利用多项式核函数与高斯核函数建立核矩阵。

多项式核函数为：

$$K(x,y) = (1 + x.y)^d \tag{19}$$

高斯核函数为：

$$K(x,y) = e^{-\gamma(x-y)^2} \tag{20}$$

利用训练集建立核矩阵，计算零投影方向，然后将测试集按照该方向投影在零空间上，通过计算测试集的 F1-score 来验证模型的有效性。

2.3 实验及结果分析

首先利用核零空间算法对乳腺癌训练集建立模型，然后再利用乳腺癌样本数据测试集对该模型进行测试，得到不同条件下的识别率。

分别取多项式核函数 $d=2$，$d=3$，$d=4$，$d=5$，以及高斯核函数的 ROC 曲线，如图 1 ~ 图 5 所示。

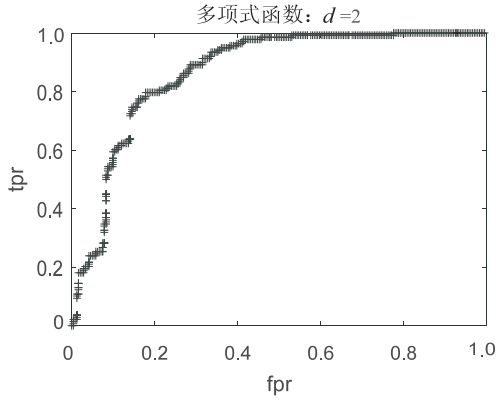

图 1 $d=2$ 的多项式函数的 ROC 曲线

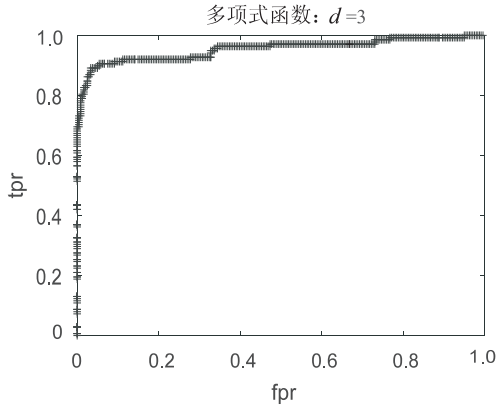

图 2 $d=3$ 的多项式函数的 ROC 曲线

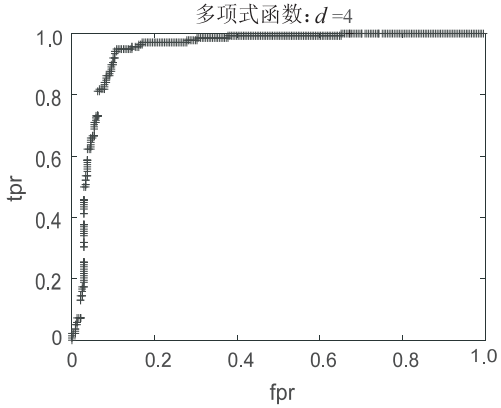

图 3 $d=4$ 的多项式函数的 ROC 曲线

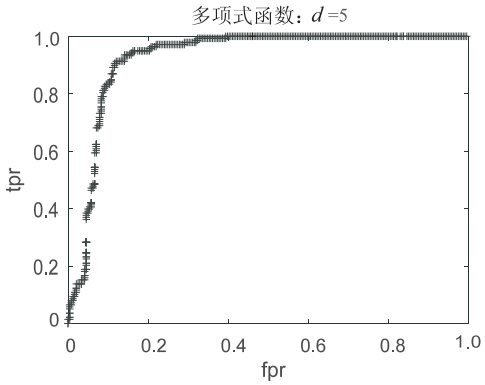

图 4 $d=5$ 的多项式函数的 ROC 曲线

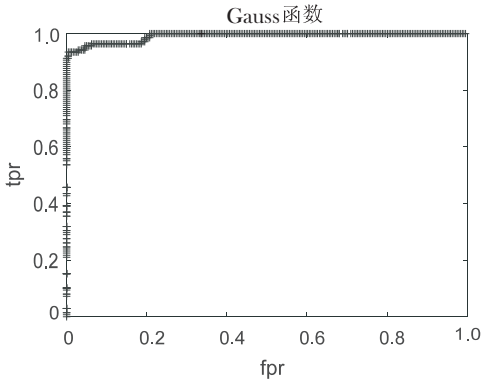

图 5 $\gamma=1$ 的高斯核函数的 ROC 曲线

通过观察图像，利用多项式核作为核函数时，在 $d=2$ 时，模型表现最差，之后随着参数 d 的增加，在 $d=3$ 时，表现最好，随后，$d=4$，$d=5$ 时，模型表现逐渐变差。这说明利用多项式作为核函数时，选取 $d=3$ 为最佳参数。但是，相比于多项式作为核函数，发现使用 $\gamma=1$ 的高斯核作为核函数时，模型的表现要比任何参数下的多项式函数的表现都好，当然高斯核函数也可通过调节参数得到不同的结果，但这里仅讨论 $\gamma=1$ 时的高斯核函数。因此，利用核零空间算法对乳腺癌数据进行检测时，利用高斯核函数要优于多项式核函数。

下面是具体地使用不同核函数时，取所有样本点到正常点的测试样本的平均值(Mean)作为异常阈值与取使得 F1-score 不等于 1 的最大值对应的测试点的距离作为异常阈值(Best)的 F1-score 结果对比。

表 1 不同核函数下取 Mean 与 Best 作为异常阈值的 F1-score 结果

	多项式核 ($d=2$)	多项式核 ($d=3$)	多项式核 ($d=4$)	多项式核 ($d=5$)	高斯核 ($\gamma=1$)
Mean	0.688 6	0.692 5	0.661 8	0.711 3	0.779 7
Best	0.756 0	0.914 5	0.888 1	0.860 1	0.962 7

通过观察图表发现，纵向来看，无论利用什么核函数，当选取测试集到正常点的平均距离(Mean)作为异常阈值进行判断时的识别率都比选取 Best 作为异常阈值的 F1-score 分数低。横向来看，仅看多项式函数

时,最佳 F1-score 为取参数 $d=3$ 时,获得最高的 F1-score 分数 91.45%,随后,随着参数的增加,F1-score 分数逐渐减少,但是,若取 $\gamma=1$ 的高斯核作为核函数,则模型的 F1-score 分数要远远超过任何参数下的多项式核函数的模型 F1-score 分数,达到了 96.27%。这表明使用 $\gamma=1$ 的高斯核作为核函数时的模型的性能比任何参数下的多项式核函数的模型的性能都好。

综上所述,在利用核零空间算法进行异常检测时,选取一个合适的核函数以及定义一个最佳的异常判别阈值,对模型的结果有很大的影响。在对乳腺癌数据集进行异常检测时,选取高斯核函数进行非线性映射无疑是要优于多项式核函数的。

3　结束语

该文基于最大化 Fisher 原则,利用核零空间算法在处理高维数据及有效提取数据非线性特征上的优势,将其运用于 UCI 数据集的乳腺癌数据集上,通过 MATLAB 仿真实验发现,使用 $\gamma=1$ 的高斯核作为核函数时的模型的 F1-score 分数比任何参数下的多项式核函数的模型的 F1-score 分数都高,并且不同异常阈值下的 F1-score 分数也不同,充分证明了运用核零空间算法进行乳腺癌异常检测的有效性。未来如何通过建立更加有效的核函数,选取更加合适的异常阈值从而实现更高的识别率,加快大数据集的运行速度仍然是一个值得深入研究的问题。

参考文献:

[1] SALATINO M, GIROTTI M R, RABINOVICH G A. Glycans pave the way for immunotherapy in triple-negative breast cancer[J]. Cancer Cell,2018,33(2):155-157.

[2] 钱小霞,司　芩,钱晓莉,等. 乳腺癌超声造影特征分析[J]. 中华超声影像学杂志,2012,21(3):217-219.

[3] DESANTIS C,MA J,BRYAN L,et al. Breast cancer statistics, 2013[J]. CA:A Cancer Journal for Clinicians,2014,64(1):52-62.

[4] 任玉琳,贾勇圣,佟仲生. 乳腺癌化学预防的研究进展[J]. 肿瘤,2018,38(6):617-622.

[5] 耿　怡,马富成,热西达·加帕尔,等. 双模态超声在评估乳腺癌新辅助化疗疗效中的应用价值[J]. 新疆医科大学学报,2017,40(3):275-278.

[6] BELHUMEUR P N, HESPANHA J P, KRIEGMAN D J. Eigenfaces vs. fisherfaces: recognition using class specific linear projection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,1997,19(7):711-720.

[7] 郑宇杰,杨静宇,徐　勇,等. 一种基于 Fisher 鉴别极小准则的特征提取方法[J]. 计算机研究与发展,2006,43(7):1201-1206.

[8] FOLEY D H,SAMMON J W. An optimal set of discriminant vectors[J]. IEEE Transactions on Computers,1975,C-24(3):281-289.

[9] GUO Y F,WU L,LU H,et al. Null Foley – Sammon transform[J]. Pattern Recognition,2006,39(11):2248-2251.

[10] BODESHEIM P,FREYTAG A,RODNER E,et al. Kernel null space methods for novelty detection[C]//IEEE conference on computer vision & pattern recognition. Portland,Oregon,USA:IEEE,2013:3374-3381.

[11] 郭跃飞,杨静宇. 求解广义最佳鉴别矢量集的一种迭代算法及人脸识别[J]. 计算机学报,2000,23(11):1189-1195.

[12] 甘俊英,何国辉,何思斌. 核零空间线性鉴别分析及其在人脸识别中的应用[J]. 计算机学报,2014,37(11):2374-2379.

[13] 舒影岚,陈艳萍,吉臻宇,等. 健康医疗大数据研究进展[J]. 中国医学装备,2019,16(1):143-147.

[14] LEHMAN C D,WELLMAN R D,BUIST D S M,et al. Diagnostic accuracy of digital screening mammography with and without computer-aided detection[J]. JAMA Internal Medicine,2015,175(11):1828-1837.

[15] SU H,LIU F,XIE Y,et al. Region segmentation in histopathological breast cancer images using deep convolutional neural network[C]//2015 IEEE international symposium on biomedical imaging. Boston:IEEE,2015:55-58.

(上接第 164 页)

测方法[J]. 计算机与数字工程,2014,42(1):122-128.

[17] 梁肇峻,钟　俊. 基于 Otsu 算法与直方图分析的自适应 Canny 算法的改进[J]. 现代电子技术,2019,42(11):54-58.

[18] 韩慧妍,韩　燮. 形态学和 Otsu 方法在 Canny 边缘检测算子中的应用[J]. 微电子学与计算机,2012,29(2):146-149.

[19] 中国气象局. 地面气象观测规范[M]. 北京:气象出版社,2003:48-51.

[20] 中华人民共和国国家质量监督检验检疫总局. GB/T 35227-2017 地面气象观测规范-风向和风速[S]. 北京:中国标准出版社,2017.

[21] HALL G,TERRELL T J,SENIOR J M,et al. A new fast discrete radon transform for enhancing linear features in noisy images[J]. Electronics Letters,1988,24(14):876-877.

[22] ZHANG Yuhua,WANG Xin. Study of finite radon transform in face recognition[J]. Image Processing and Its Applications,2010(2):29-32.

[23] 赵晓莉,苑　跃,黄晓龙,等. EL 型电接风向风速自记纸数字化风向识别方法研究[J]. 中低纬山地气象,2019,43(4):83-86.