

文章编号:1001-9081(2009)06-1559-04

一种新的最小二乘支持向量机稀疏化算法

吴宗亮, 窦 衡

(电子科技大学 电子工程学院, 成都 610054)

(nanyuanfeixue@sina.com)

摘 要:普通的最小二乘支持向量机(LS-SVM)稀疏化算法在处理有些常见的模式识别问题时,随着训练样本的删减,识别率下滑很快,往往达不到稀疏化的目的。针对这种情况,提出了一种新的 LS-SVM 稀疏化算法来弥补这种不足,从而使得 LS-SVM 稀疏化算法体系更加完善。将新算法应用到雷达一维距离像的识别中,实验结果证明了新算法的有效性。

关键词:最小二乘支持向量机;稀疏化;雷达一维距离像

中图分类号: TP391.4 **文献标志码:** A

New sparse least squares support vector machine algorithm

WU Zong-liang, DOU Heng

(College of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu Sichuan 610054, China)

Abstract: The recognition rate of Least Squares Support Vector Machine (LS-SVM) sparse algorithm rapidly decreases with the reduction of training samples in dealing with some pattern recognition issues, and the sparsification can not be achieved. To overcome such a shortage, a new sparse algorithm was proposed. The method was applied to radar range profile's recognition and the experimental results show its validity in recognition.

Key words: Least Squares Support Vector Machine (LS-SVM); sparsification; radar range profile

0 引言

最小二乘支持向量机^[1-2](LS-SVM)是处理不可分样本集情况下模式分类的有效工具,该算法最大的不足之处就是缺乏稀疏性。目前,比较常见的最小二乘支持向量机稀疏化算法主要是通过一定的策略,删减一些对分类影响不大的训练样本,以达到对训练样本稀疏化的目的。不过这些算法有一个共同的缺点,就是在处理有些常见的模式分类问题时,随着训练样本的删减,识别率下滑很快,往往达不到稀疏化的目的。为了克服这种困难,本文提出了一种新的最小二乘稀疏化算法,通过引入二阶 Renyi 熵以及对其所用参数进行事先设定,以达到对正负训练样本集合分别自适应聚类的目的。然后抽取每一个聚类的中心作为新的训练样本,从而达到对训练样本稀疏化的目的。在该算法中,并没有删减任何训练样本,对训练样本的稀疏化是通过聚类来完成的。

1 最小二乘支持向量机算法简介

设训练样本 $\mathbf{x}_i \in \mathbf{R}^n (i = 1, 2, \dots, l)$ 属于两类,用 $y_i \in \{-1, 1\}$ 作为标记,其中正样本集合和负样本集合的样本数分别为 l_1 和 l_2 , 并且样本总数 $l = l_1 + l_2$ 。通过一个非线性映射 Φ , 将输入的训练样本 \mathbf{x} 映射到一个高维特征空间 F 中, 即: $\Phi: \mathbf{R}^n \rightarrow F, \mathbf{x} \rightarrow \Phi(\mathbf{x})$ 。此时, 输入的训练样本由原来的 \mathbf{x} 变为 $\Phi(\mathbf{x})$ 。 n 维空间中的样本集 $\{\mathbf{x}_i\}_{1 \leq i \leq l}$, 由非线性映射到高维特征空间 F 后, 得到高维样本集 $\{\Phi(\mathbf{x}_i)\}_{1 \leq i \leq l}$ 。

支持向量机的目标就是寻找一个最优分类超平面^[3], 尽可能地使得不同类别的样本能够正确分开。其数学表述就是构造一个如下形式的分类器:

$$f(\mathbf{x}) = \text{sign}(\mathbf{w}^T \Phi(\mathbf{x}) + b) \quad (1)$$

使得样本 \mathbf{x} 能够被函数 $f(\mathbf{x})$ 正确分类。利用最小二乘支持向量机算法寻找最优分类超平面就是求解下面的优化问题:

$$\min_{\mathbf{w}, e} J(\mathbf{w}, e) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{r}{2} \sum_{i=1}^l e_i^2 \quad (2)$$

并且满足等式约束:

$$y_i [\mathbf{w}^T \Phi(\mathbf{x}_i) + b] = 1 - e_i; i = 1, \dots, l \quad (3)$$

其中: $e_i \in \mathbf{R}$ 是分类误差, $r > 0$ 为惩罚系数, 其目的主要是调整分类误差。如果训练数据中含有较大的噪声, 应该选择较小的 r 。

为了求目标函数 J (即式(2)) 的最小值, 现在构造 Lagrange 函数 L 如下:

$$L = J(\mathbf{w}, e) - \sum_{i=1}^l a_i \{y_i [\mathbf{w}^T \Phi(\mathbf{x}_i) + b] - 1 + e_i\} \quad (4)$$

其中 a_i 是 Lagrange 乘子 (也称支持值), 因为是等式约束, 所以其值可正可负。最优化的条件如下:

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^l a_i y_i \Phi(\mathbf{x}_i) \quad (5)$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^l a_i y_i = 0 \quad (6)$$

$$\frac{\partial L}{\partial e_i} = 0 \Rightarrow r e_i = a_i \Rightarrow e_i = \frac{a_i}{r} \quad (7)$$

$$\frac{\partial L}{\partial a_i} = 0 \Rightarrow y_i [\mathbf{w}^T \Phi(\mathbf{x}_i) + b] - 1 + e_i = 0 \quad (8)$$

其中 $i = 1, \dots, l$ 。把式(5)、(7)代入式(8), 消掉 \mathbf{w} 和 e_i , 并且利用满足 Mercer 条件的内积核函数^[4], 即:

收稿日期:2008-12-12;修回日期:2009-03-12。

作者简介:吴宗亮(1979-),男,陕西宝鸡人,硕士研究生,主要研究方向:信号处理、模式识别、雷达目标识别; 窦衡(1964-),男,四川乐山,副教授,主要研究方向:信号处理、模式识别、微波成像。

$$k(\mathbf{x}_i, \mathbf{x}_j) = (\Phi(\mathbf{x}_i) \times \Phi(\mathbf{x}_j)) = \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j) \quad (9)$$

其中 $1 \leq i \leq l, 1 \leq j \leq l$ 。式(8) 经过整理转变为:

$$\sum_{j=1}^l a_j y_j k(\mathbf{x}_i, \mathbf{x}_j) + (1/r) a_i + b y_i = 1 \quad (10)$$

其中 $i = 1, \dots, l$ 。把式(6)、(10) 写成如下矩阵形式:

$$\begin{bmatrix} 0 & Y^T \\ Y & A \end{bmatrix} \begin{bmatrix} b \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{1} \end{bmatrix} \quad (11)$$

$$Y = [y_1, y_2, \dots, y_l]^T \quad (12)$$

$$\mathbf{a} = [a_1, a_2, \dots, a_l]^T \quad (13)$$

$$\mathbf{1} = [1, 1, \dots, 1]^T \quad (14)$$

在这里重点交代一下 A 的组成:

$$A = \Omega + (1/r) I \quad (15)$$

其中 I 是 l 阶单位矩阵, $\Omega = (y_i y_j k(\mathbf{x}_i, \mathbf{x}_j))$, 其中 $1 \leq i \leq l, 1 \leq j \leq l$ 。求解方程组(11) 可以得到:

$$b = \frac{Y^T A^{-1} \mathbf{1}}{Y^T A^{-1} Y} \quad (16)$$

$$\mathbf{a} = A^{-1} (\mathbf{1} - Yb) \quad (17)$$

由于 A 是一个半正定对称矩阵, 一般情况下其逆是存在的。那么最小二乘支持向量机的分类函数(即最优分类面函数)为:

$$f(\mathbf{x}) = \text{sign} \left[\sum_{i=1}^l a_i y_i k(\mathbf{x}_i, \mathbf{x}) + b \right] \quad (18)$$

2 新的 LS-SVM 稀疏化算法的提出

2.1 现有 LS-SVM 稀疏化算法的不足

最小二乘支持向量机最大的不足之处就是缺乏稀疏性, 原因在于: 1) 其目标函数 J 引入了分类误差平方求和项 $\sum_{i=1}^l e_i^2$; 2) 支持值 a_i 的大小与训练点处的分类误差 e_i 成正比, 如式(7) 所示。所以往往会出现所有的训练样本都成了支持向量, 因为支持值 a_i 一般不等于零。所以如果要提高最小二乘支持向量机的执行效率, 势必要对训练样本进行稀疏化处理。但是一个合理的稀疏化算法必须保证分类识别率基本保持不变这个前提, 脱离了这个前提条件, 一切稀疏化算法都是无效的。

通过观察最优分类面函数(式(18)), 我们可以发现, 支持值 a_i 的绝对值越小, 那么它对分类面的贡献也就越小, 即绝对值小的支持值对分类影响不大, 既然这样, 那么就可以把绝对值小的支持值所对应的训练样本删除掉。目前, 比较常见的最小二乘支持向量机稀疏化算法, 一般都是基于上述思想的, 不管采取任何手段, 其目的都是先寻找绝对值小的支持值, 然后把其所对应的训练样本删除掉。这一类稀疏化算法只能解决训练样本支持值的绝对值相差很远的模式识别问题, 如果所有训练样本支持值的绝对值的和只决定于少量训练样本, 那么这一类算法就可以起到很好的稀疏化训练样本的目的, 而且稀疏化以后对识别率影响不大。

然而, 在现实生活中还存在其他一些模式识别问题, 即训练样本集合中各个训练样本所对应的支持值的绝对值相差不大, 这时, 各个训练样本对分类面(式(18)) 的贡献相当, 如果这个时候还采用把绝对值小的支持值所对应的训练样本删除掉这一类稀疏化算法的话, 势必会对最优分类面产生很大影响。训练样本删除少了达不到稀疏化的目的, 训练样本删除多了, 正确识别率迅速下滑。针对这种情况, 本文提出了一种新

的最小二乘支持向量机稀疏化算法。

2.2 新的 LS-SVM 稀疏化算法的数学指导思想

任意给定两个属于同一类别的样本, 如果这两个样本相似程度很高, 一般来说, 倘若其中一个样本能被某个分类器正确识别, 那么另外一个样本能被正确识别的概率就非常高。该文提出的 LS-SVM 稀疏化算法正是基于上述思想的。

设有一个训练样本集合 $X = \{\mathbf{x}_i | i = 1, 2, \dots, N\}$, X 中的所有样本都属于同一个类别。我们可以通过某种处理手段, 把集合 X 划分成若干个小的子集, 使得每个子集中的所有样本之间不但相似程度高, 而且在分布上也很集中, 不妨把这样的小的子集称之为聚类, 即 $X = \{\Omega_1, \Omega_2, \dots, \Omega_M\}$, Ω_i ($1 \leq i \leq M$) 为某一个聚类, M 为聚类的数目, 并且肯定 $M < N$ 。我们现在抽取每一个聚类的类中心向量组成一个新的训练样本集合 $\Omega = \{\bar{\Omega}_1, \bar{\Omega}_2, \dots, \bar{\Omega}_M\}$, 即 $\bar{\Omega}_i$ 为聚类 Ω_i 的类中心向量。因为集合 Ω 和集合 X 都属于同一个类别, 那么能不能把基于集合 X 的模式识别转换为基于集合 Ω 的模式识别呢? 现在以集合 Ω 作为某一个类别的训练样本集合, 利用最小二乘支持向量机算法求得最优分类面函数(如式(18)), 既然是以 Ω_i ($1 \leq i \leq M$) 作为训练样本的, 那么一般情况下 Ω_i 肯定能被式(18) 正确识别。对于聚类 Ω_i 而言, 它里面的所有样本不但属于同一个类别而且样本之间的相似程度很高, 那么这些样本肯定和类中心向量 $\bar{\Omega}_i$ 之间有着很高的相似程度, 既然 $\bar{\Omega}_i$ 能被式(18) 正确识别, 那么一般情况下 Ω_i 里面的样本也能被式(18) 正确识别。所以以集合 Ω 作为训练样本是合理的, 因为 $M < N$, 故可以在不降低识别率的基础之上, 达到稀疏化训练样本的目的。

通过对样本集合 X 进行上述聚类, 有两个好处。一是因为类中心向量中包含有丰富的特征信息, 所以用它作为新训练样本, 信息损失小; 二是集合 Ω 作为某一个类别的训练样本, 并没有把绝对值小的支持值所对应的训练样本删除掉, 从而避免了现有 LS-SVM 稀疏化算法的不足。

通过上面的讨论, 我们可以得知, 新的 LS-SVM 稀疏化算法的关键是如何对训练样本集合进行合理的划分, 才能使每个小的子集(聚类) 中的样本之间具有很高的相似程度。由此我们引入了二阶 Renyi 熵的概念。

3 二阶 Renyi 熵

在信息论中, 熵是定义用来描述平均信息量以及度量不确定性的。消息是携带信息的载体, 消息由信息源输出。设某个信息源输出若干个相互独立的消息 \mathbf{x}_i ($i = 1, 2, \dots, N$), 而每个消息出现的概率为 p_i ($i = 1, 2, \dots, N$), 则可以用 $H_i = -p_i \log p_i$ 来度量一个消息所给出的平均信息量, 那么整个信息源输出的消息所携带的平均信息量为:

$$H = - \sum_{i=1}^N p_i \log p_i \quad (19)$$

其中 H 称为信息熵或香农熵。它是对信息源输出信息的随机性的度量, 香农熵越大, 信息源输出的平均信息量也就越大, 其所对应系统的随机程度也就越高。Renyi 进一步扩展了香农熵的概念, 针对概率密度函数, 提出了更加灵活地度量不确定性的定义^[5]。概率密度函数为 $p(\mathbf{x})$ 的 α ($\alpha \geq 0, \alpha \neq 1$) 阶 Renyi 熵的定义如下:

$$H_{R\alpha} = \frac{1}{1-\alpha} \log \int p^\alpha(\mathbf{x}) d\mathbf{x} \quad (20)$$

本文将采用二阶 Renyi 熵, 因为对于 Gaussian 核函数, 二

阶 Renyi 熵的计算可以简化。二阶 Renyi 熵的表达式如下:

$$H_{R2} = -\log \int p^2(\mathbf{x}) d\mathbf{x} \quad (21)$$

其中 $\int p^2(\mathbf{x}) d\mathbf{x}$ 可以用式(22) 进行估计^[6]:

$$\int p^2(\mathbf{x}) d\mathbf{x} \approx \hat{\int p^2(\mathbf{x}) d\mathbf{x}} = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N k(\mathbf{x}_i, \mathbf{x}_j) \quad (22)$$

其中 $\{\mathbf{x}_i | i = 1, \dots, N\}$ 是参与计算的全体样本数据。所以,二阶 Renyi 熵可以用下面的公式近似计算:

$$H_{R2} = -\log \left[\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N k(\mathbf{x}_i, \mathbf{x}_j) \right] \quad (23)$$

在这里需要说明的一点,无论是计算最小二乘支持向量机的分类函数还是计算二阶 Renyi 熵,在该文中统一采取高斯径向基核函数:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2) \quad (24)$$

4 新的 LS-SVM 稀疏化算法的实现

设有一个训练样本集合 $X = \{\mathbf{x}_i | i = 1, 2, \dots, N\}$, X 中的所有样本都属于同一个类别。现在从中随机选取 l ($1 < l < N$) 个样本构成一个子集 Y , 并计算它的二阶 Renyi 熵 H_{R2} , 可以用 H_{R2} 来度量该子集的信息冗余度。如果 H_{R2} 越大, 说明子集 Y 所包含的平均信息量越大, 那么其样本之间的信息冗余度也就越小, 它里面样本之间的相似程度也就越低, 子集 Y 也就相对较分散; 如果 H_{R2} 越小, 说明子集 Y 所包含的平均信息量越小, 那么其样本之间的信息冗余度也就越大, 它里面样本之间的相似程度也就越高, 子集 Y 也就相对较集中。既然如此, 那么就可以根据样本之间二阶 Renyi 熵的大小对集合 X 中的样本进行聚类, 把二阶 Renyi 熵较小的若干个样本聚成一类。通过聚类, 就可以把样本集合 X 划分成若干个小的子集, 每个子集中的若干个样本具有较小的二阶 Renyi 熵。

为了达到自适应聚类的目的, 本文引进了一个参数 β , 称其为二阶 Renyi 熵聚类精度, 简称聚类熵精度。 β 的值事先进行设定, 那么就可以把样本集合 X 中二阶 Renyi 熵小于 β 的若干个样本聚成一类。并且 β 越大, 相应的聚类数目也就越小, 而每一类中所包含的样本也就越多; β 越小, 相应的聚类数目也就越大, 而每一类中所包含的样本也就越少。当然了, 目前在模式识别领域中, 聚类的算法很多, 但是基于二阶 Renyi 熵的聚类方法操作简捷, 便于理解, 不失为一种可取的方法。

下面给出基于二阶 Renyi 熵的新的 LS-SVM 稀疏化算法的实现步骤:

1) 首先从样本集合 X 中随机抽取一个样本, 不妨把该样本设为 \mathbf{x}_r , $1 \leq r \leq N$ 。接着把 X 中的其他样本分别(不是依次)和 \mathbf{x}_r 相结合, 从而构成若干个新的子集, 此时每个子集中包含两个样本, 即为: $\Omega_j = \{\mathbf{x}_i, \mathbf{x}_r | i = 1, 2, \dots, N; i \neq r\}$, $j = 1, 2, \dots, N-1$ 。然后计算每个子集 Ω_j 的二阶 Renyi 熵, 并把其最小值 H_{R2min} 找出来。如果 H_{R2min} 小于事先设定的聚类熵精度 β , 那么就把 H_{R2min} 所对应的子集找出来, 在这里不妨把该子集设为 Ω_r , 此时 Ω_r 含有两个样本, Ω_r 就是要寻找的某个聚类, 但 Ω_r 是一个并未完成聚类的集合; 如果 H_{R2min} 大于或等于 β , 那么 \mathbf{x}_r 就是一个独立的已经完成的新的聚类, 但该聚类中只包含一个样本 \mathbf{x}_r 。

2) 把在样本集合 X 中的, 但并不属于未完成聚类的集合 Ω_r 中的那些样本, 分别(不是依次)加入到 Ω_r 中, 从而又构成若干个新的子集。然后计算每个子集的二阶 Renyi 熵, 并把其最小值 H_{R2min} 找出来。如果 H_{R2min} 小于事先设定的聚类熵精

度 β , 那么就把 H_{R2min} 所对应的子集找出来, 该子集是在原来 Ω_r 的基础上扩充了一个样本得到的, 在这里用扩充以后的 Ω_r (即此时 H_{R2min} 所对应的子集) 来替代原来未扩充的 Ω_r , 但此时的 Ω_r 仍旧是一个未完成聚类的集合; 如果 H_{R2min} 大于或等于 β , 那么原来未扩充的 Ω_r 就是一个独立的已经完成的新的聚类。

3) 重复第2) 步的操作, 直到由未完成聚类的集合 Ω_r 所构成的若干个新子集的 H_{R2min} 大于或等于事先设定的聚类熵精度 β , 此时退出循环。这个时候, 样本集合 Ω_r 才是一个独立的已经完成的新的聚类, 并且把 Ω_r 作为一个新的聚类保存起来。

4) 在训练样本集合 X 中, 把属于已经完成的新的聚类 Ω_r 中的样本全部去掉, 对样本集合 X 进行缩减, 把缩减以后的样本集合仍旧用 X 表示。设 Ω_r 中有 n 个样本, 那么缩减以后的样本集合 X 的样本数目变为: $N = N - n$ 。

5) 重复上面4个步骤, 直到样本集合 X 中的样本数目为零为止, 此时也就完成了所有的聚类。这时, 最初的由 N 个训练样本组成的集合 X 就被划分成为若干个小的聚类, 即表示为: $X = \{\Omega_1, \Omega_2, \dots, \Omega_M\}$, Ω_i ($1 \leq i \leq M$) 为某一个聚类, M 为聚类的数目, 一般情况下, 肯定 $M < N$ 。

6) 抽取每个聚类的类中心向量即 $\bar{\Omega}_i$ ($1 \leq i \leq M$, $\bar{\Omega}_i$ 为聚类 Ω_i 的类中心向量) 作为最小二乘支持向量机某一类别的新的训练样本, 此时, 最初的训练样本集 X 也就缩减成了集合 $\Omega = \{\bar{\Omega}_1, \bar{\Omega}_2, \dots, \bar{\Omega}_M\}$ 。由于新的训练样本的数目等于聚类的数目, 由此就达到了不删减训练样本而稀疏化的目的。

在这里需要说明以下几个问题:

1) 聚类熵精度 β 的设置非常重要, β 的值与训练样本集合 X 的二阶 Renyi 熵密切相关。所以可以在 X 中, 事先随机抽取若干个样本, 这样的抽取不妨可以进行多次, 计算每次所抽取的若干个样本的二阶 Renyi 熵, 再取所计算出的二阶 Renyi 熵的最小值作为对 β 的初步估计。具体在实验的时候, 可以根据估计值对 β 由小到大进行调节, 以达到对训练样本集合最佳的聚类效果。

2) 设某个聚类 $\Omega_r = \{\mathbf{x}_{r1}, \mathbf{x}_{r2}, \dots, \mathbf{x}_{rn}\}$, 可以采取用 Ω_r 中的样本均值作为该聚类的中心, 即:

$$\bar{\Omega}_r = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_{ri} \quad (25)$$

其中 $\bar{\Omega}_r$ 用来表示 Ω_r 的聚类中心。

3) 具体在用上述稀疏化策略处理最小二乘支持向量机的分类问题时, 我们事先分别对正负两类训练样本集合进行各自独立的聚类。设正样本集合的聚类熵精度为 β_+ , 而负样本集合的聚类熵精度为 β_- 。实验前, 分别从正负两类样本集合中随机选出相同数目的若干个样本, 构成正样本和负样本子集, 分别计算各自的二阶 Renyi 熵。在这里, 不妨假设所计算出的正样本子集的二阶 Renyi 熵连续好几次都大于负样本子集的二阶 Renyi 熵。这说明, 和负样本集合比较起来, 正样本集合中样本之间的信息冗余度相对较小以及样本之间相对分散的可能性更大, 所以在聚类的时候, 可以选择 β_+ 适当大于 β_- 。这样处理的好处是, 正负两类训练样本集合各自分别聚类以后, 两者的聚类数目不至于相差很远。换句话说, 正负样本集合稀疏化以后, 各自的训练样本数不至于相差很远。毕竟稀疏化以后的训练样本的数目等于正负样本集合各自的聚类数目。

5 实验结果

本文实验所采取的实测数据是 ISAR 雷达对空中的 2 种飞机(a, b)所成的距离像。ISAR 交替发射窄带(其脉冲时宽为 $1\ \mu\text{s}$)和宽带两种波形,窄带系统主要用于跟踪目标和产生宽带本振定时信号。宽带信号的带宽为 400 MHz(其理论上距离分辨率为 $0.375\ \text{m}$),采样点数为 256(其经过 FFT 后所得一维距离像的像点数也为 256)。每种飞机录取了 7 段数据,每段数据含 26000 个宽、窄带信号(相邻间隔 $2.5\ \text{ms}$)。宽带信号为全去斜后的正交双通道信号(其 FFT 即为一维距离像),每段数据含 260 个宽带正交双通道信号,即每段数据有 260 幅距离像。实验数据为 2 种飞机各取一段的 130 幅距离像(总数为 260 幅)作训练样本,另外 130 幅为测试样本(总数为 260 幅)。在对实测数据训练前,作如下两步预处理:

1)归一化。将每一幅像用其总能量归一。

2)距离对准。利用 Fourier 变换的平移不变性,将目标一维距离像作 Fourier 变换即可对齐,同时根据实数 Fourier 变换的共轭对称性,可取一维距离像 Fourier 变换的一半(128 维)作为识别输入矢量进行实验。

在这里需要说明的是,雷达一维距离像和其他类型的图

像相比(例如人脸图像)并没有什么特别之处。本文所采取的雷达一维距离像,其一幅图像从数学角度来说,只不过是一个维数为 128 的向量而已。在下面实验中,参数设置为: $r = 100$;高斯径向基核函数的参数 $\sigma^2 = 0.2$ 。

实验一 采用普通的最小二乘支持向量机的稀疏化策略。利用所给出的 a 类和 b 类飞机的训练样本集合,首先采用 LS-SVM 算法求出每个训练样本的支持值,接着把绝对值较小的支持值所对应的训练样本删除掉,然后利用剩余的训练样本求出 LS-SVM 的最优分类函数,最后再采用该分类函数对测试样本进行检测。之所以采取上述手段,目的是为了比较和说明问题。表 1 给出了这种稀疏化策略的识别效果。在表 1 中,用 N_a 表示稀疏化以后 a 类飞机的训练样本数; N_b 表示稀疏化以后 b 类飞机的训练样本数; N_{a+b} 表示稀疏化以后总的训练样本数; η_a 表示 N_a 与稀疏化前 a 类飞机的训练样本数的百分比; η_b 表示 N_b 与稀疏化前 b 类飞机的训练样本数的百分比; η_{a+b} 表示 N_{a+b} 与稀疏化前总的训练样本数的百分比。

实验二 采用本文所提出的最小二乘支持向量机的稀疏化策略,表 2 给出了这种新的稀疏化策略的识别效果。在表 2 中,用 β_a 表示 a 类飞机的聚类熵精度; β_b 表示 b 类飞机的聚类熵精度,其他变量的定义同实验一。

表 1 普通 LS-SVM 稀疏化算法对两类飞机的识别结果

序号	N_a	N_b	N_{a+b}	a 类识别率/%	b 类识别率/%	平均识别率/%	η_a /%	η_b /%	η_{a+b} /%
1	130	130	260	96.15	97.69	96.92	100.00	100.00	100.00
2	78	87	165	93.85	98.46	96.15	60.00	66.92	63.46
3	45	58	103	87.69	98.46	93.08	34.62	44.62	39.62
4	24	32	56	64.62	90.00	77.31	18.46	24.62	21.54

表 2 新的 LS-SVM 稀疏化算法对两类飞机的识别结果

序号	β_a	β_b	N_a	N_b	N_{a+b}	a 类识别率/%	b 类识别率/%	平均识别率/%	η_a /%	η_b /%	η_{a+b} /%
1	0.01	0.04	130	130	260	96.15	97.69	96.92	100.00	100.00	100.00
2	0.04	0.04	126	130	256	96.15	97.69	96.92	96.92	100.00	98.46
3	0.08	0.08	85	80	165	96.92	97.69	97.30	65.38	61.54	63.46
4	0.12	0.12	53	50	103	96.92	96.92	96.92	40.77	38.46	39.62
5	0.18	0.16	30	26	56	93.85	95.39	94.62	23.08	20.00	21.54
6	0.22	0.19	20	17	37	93.85	94.62	94.23	15.38	13.08	14.23
7	0.23	0.19	18	17	35	92.31	93.85	93.08	13.85	13.08	13.46
8	0.25	0.21	14	13	27	90.77	94.62	92.69	10.77	10.00	10.38
9	0.27	0.23	10	9	19	90.00	93.08	91.54	7.69	6.92	7.31

为了进行比较,实验一和实验二中稀疏化以后总的训练样本数(即 N_{a+b})设置为相同。

对实验结果的分析:

1)通过分别计算 a 类和 b 类飞机训练样本集的二阶 Renyi 熵,可以发现 a 类训练样本分布相对比较分散,而 b 类训练样本分布相对比较集中。所以在稀疏化以后,应该让 a 类训练样本数 N_a 略大于 b 类训练样本数 N_b 。但是如果采用实验一的稀疏化策略,则完全没有考虑 a 类和 b 类训练样本的分布状态。所以在稀疏化以后,有可能会使 a 类样本的数目 N_a 小于 b 类样本的数目 N_b 。最终将导致 a 类的识别率迅速下滑,如表 1 所示。

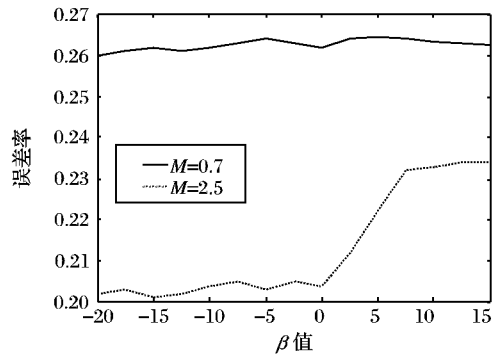
2)如果各训练样本支持值的绝对值相差不是很远,若采用实验一的稀疏化策略,训练样本删除少了,达不到稀疏化的目的,训练样本删除多了,可能会导致分类识别率大幅度下滑。实验一的结果(表 1)对这一现象做出了有力的证明。采用实验一的稀疏化策略,当总的训练样本由 260 个稀疏到 165 和

103 个的时候,平均识别率尚且可以,但是当稀疏到 56 个的时候, a 类飞机的识别率迅速下滑,平均识别率下滑也很大。

3)采用本文所提出的稀疏化策略,不但分别考虑了 a 类训练样本集与 b 类训练样本集各自内部的散布状态,而且也考虑了这两个集合之间的相对散布状态。通过调整各自的聚类熵精度 β_a 与 β_b ,使得稀疏化以后的 a 类样本的数目 N_a 略大于 b 类样本的数目 N_b ,从而避免了稀疏化以后 a 类识别率迅速下滑,如表 2 所示。

4)采用本文所提出的稀疏化策略,把 a 类训练样本集与 b 类训练样本集分别合理地划分成若干个小聚类集合,然后抽取每个聚类的中心作为稀疏化以后的训练样本。由于每个聚类中的训练样本相似程度高,而且类中心向量中包含有丰富的特征信息,所以避免了稀疏化以后分类识别率迅速下滑。如表 2 中,当总的训练样本由 260 个稀疏到 19 个的时候,仍旧能保持较高的平均识别率。而且,如果适当地调整聚

(下转第 1581 页)

图4 β 对 StreamME 算法分类误差的影响

采用传感器数据集和车辆数据集对 StreamME 算法和 CB 算法进行了对比实验,表1给出了 StreamME 方法与 CB 方法在分类性能方面的比较,误差率 P 定义为: $\Delta p_i = \frac{(v_i^c - v_i^a)}{v_i^a} \times 100\%$, 其中 v_i^c 表示量化的预测值, v_i^a 表示量化的实际值, $P = \left[1 - \sqrt{\frac{1}{n} \sum_{i=1}^n \Delta p_i^2} \right] \times 100\%$, 物理含义为平均预测精度。

表1 StreamME 方法与启发式 CB 方法的分类性能比较

数据集	M	StreamME		heuristic CB	
		误差率/%	条件熵	误差率/%	条件熵
传感器	0.95	52	0.87	58	0.09
	2.70	50	0.76	60	0.11
	missing	48	0.73	62	0.08
车辆	0.89	27	0.34	32	1E-3
	2.28	19	0.25	34	1E-6
	missing	17	0.21	34	1E-6

从 StreamME 与启发式 CB 方法的分类性能比较表可以看出,由于 StreamME 采用了寻找最大熵解的思想,StreamME 可以获得更好的分类误差率,得到期望中更高的熵。

4 结语

本文提出一种新颖的适用于数据流组合分类的直推学习方法,在多个分类器之间没有公共标签数据的情况下,如何选择决策聚集规则。本方法解决了约束不可行、解不唯一等问题,在训练集上测算约束概率,通过对训练集的增强保证了约束的可行。实验证明,在给定测度约束前提下,本文提出的

StreamME 算法的精度比基于约束的启发式方法要高,获得了较好的性能,可广泛应用于传感器数据分类等领域。

参考文献:

- [1] WANG H, FAN W, YU P S. Mining concept-drifting data streams using ensemble classifiers [C]// Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2003: 226-235.
- [2] KOLTER J Z, MALOOF M A. Dynamic weighted majority: a new ensemble method for tracking concept drift [C]// Proceedings of the 3rd IEEE Conference on Data Mining. Los Alamitos, USA: IEEE Computer Society Press, 2003: 123-130.
- [3] KOLTER J Z, MALOOF M A. Using additive expert ensembles to cope with concept drift [C]// Proceedings of the 22nd International Conference on Machine Learning. Bonn, Germany: ACM Press, 2005: 449-456.
- [4] RUSHING J, GRAVES S, CRISWELL E, et al. coverage based ensemble algorithm (CBEA) for streaming data [C]// Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence. Huntsville, USA: IEEE Computer Society Press, 2004: 106-112.
- [5] CHU F, ZANIOLO C. Fast and light boosting for adaptive mining of data streams [C]// Proceedings of the 5th Pacific-Asia Conference on Knowledge Discovery and Data Mining. Sydney, Australia: Springer Press, 2004: 282-292.
- [6] MILLER D J, PAL S. Transductive methods for the distributed ensemble classification problem [J]. Neural Computation, 2007, 19(3): 856-884.
- [7] 孙岳, 毛国君, 刘旭, 等. 基于多分类器的数据流中的概念漂移挖掘[J]. 自动化学报, 2008, 34(1): 94-96.
- [8] KRISHNAPURAM B, WILLIAMS D, XUE YA, et al. On semi-supervised classification [EB/OL]. [2008-10-10]. http://www.lx.it.pt/~mtf/NIPS_2005.pdf.
- [9] JAYNES E T. Papers on probability, statistics, and statistical physics [M]. Dordrecht, Holland: Kluwer Academic Publishers, 1989.
- [10] PIETRA S D, BERGER A L, PIETRA V J. A maximum entropy approach to natural language processing [J]. Computational Linguistics, 1996, 22(1): 39-71.
- [11] MUTHUKRISHNAN S. Data streams: Algorithms and applications [EB/OL]. [2008-10-10]. <http://www.cs.rutgers.edu/~muthu/stream-1-1.ps>.

(上接第1562页)

类熵精度 β_a 和 β_b , 对 a 类和 b 类训练样本集分别进行合理的聚类,甚至可以保证在 b 类识别率基本保持不变的情况下, a 类的识别率有所提高,如表2中序号为3和4的数据所示。

通过对实验结果的分析,可以发现本文所提出的新的稀疏化算法确实具有很大的优势。

6 结语

本文提出了一种新的最小二乘支持向量机的稀疏化算法,并对其进行了详尽的阐述。本文所提出的稀疏化算法完全可以应用到计算机或雷达图像处理、DNA 检测、故障诊断以及其他模式识别等领域。

参考文献:

- [1] 陈爱军. 最小二乘支持向量机及其在工业过程建模中的应用 [D]. 杭州: 浙江大学, 2006.

- [2] 朱家元, 陈开陶, 张恒喜. 最小二乘支持向量机算法研究 [J]. 计算机科学, 2003, 30(7): 157-159.
- [3] 边肇祺, 张学工. 模式识别 [M]. 2版. 北京: 清华大学出版社, 2000: 296-301.
- [4] 田盛丰. 基于核函数的学习算法 [J]. 北方交通大学学报, 2003, 27(2): 1-8.
- [5] RENYI A. On measures of entropy and information [EB/OL]. [2008-10-10]. http://digitalassets.lib.berkeley.edu/math/ucb/text/math_s4_v1_article-27.pdf.
- [6] GIROLAMI M. Orthogonal series density estimation and the kernel eigenvalue problem [J]. Neural Computation, 2002, 14(3): 669-688.
- [7] 吴宗亮, 窦衡. 一种广义最小二乘支持向量机算法及其应用 [J]. 计算机应用, 2009, 29(3): 877-879.