

最小化类内距离和分类算法

王晓初^{*①②} 王士同^① 包芳^② 蒋亦樟^①

^①(江南大学数字媒体学院 无锡 214122)

^②(江苏省信息融合软件工程技术研发中心 江阴 214405)

摘要: 支持向量机分类算法引入惩罚因子来调节过拟合和线性不可分时无解的问题,优点是可以通过调节参数取得最优解,但带来的问题是允许一部分样本错分。错分的样本在分类间隔之间失去了约束,导致两类交界处样本杂乱分布,并且增加了训练的负担。为了解决上述问题,该文根据大间隔分类思想,基于类内紧密类间松散的原则,提出一种新的分类算法,称之为最小化类内距离和(Intraclass-Distance-Sum-Minimization, IDSM)分类算法。该算法根据最小化类内距离和准则构造训练模型,通过解析法求解得到最佳的映射法则,进而利用该最佳映射法则对样本进行投影变换以达到类内间隔小类间间隔大的效果。相应地,为解决高维样本分类问题,进一步提出了该文算法的核化版本。在大量 UCI 数据集和 Yale 大学人脸数据库上的实验结果表明了该文算法的优越性。

关键词: 支持向量机; 惩罚因子; 大间隔分类思想; 类内距离和; 映射法则

中图分类号: TP391.41

文献标识码: A

文章编号: 1009-5896(2016)03-0532-09

DOI: 10.11999/JEIT150633

Intraclass-Distance-Sum-Minimization Based Classification Algorithm

WANG Xiaochu^{①②} WANG Shitong^① BAO Fang^② JIANG Yizhang^①

^①(School of Digital Media, Jiangnan University, Wuxi 214122, China)

^②(Information Fusion Software Engineering Research and Development Center of Jiangsu Province, Jiangyin 214405, China)

Abstract: Classification algorithm of Support Vector Machine (SVM) is introduced the penalty factor to adjust the overfit and nonlinear problem. The method is beneficial for seeking the optimal solution by allowing a part of samples error classified. But it also causes a problem that error classified samples distribute disorderedly and increase the burden of training. In order to solve the above problems, according to large margin classification thought, based on principles that the intraclass samples must be closer and the interclass samples must be looser, this research proposes a new classification algorithm called Intraclass-Distance-Sum-Minimization (IDSM) based classification algorithm. This algorithm constructs a training model by using principle of minimizing the sum of the intraclass distance and finds the optimal projection rule by analytical method. And then the optimal projection rule is used to samples' projection transformation to achieve the effect that intraclass intervals are small and the interclass intervals are large. Accordingly, this research offers a kernel version of the algorithm to solve high-dimensional classification problems. Experiment results on a large number of UCI datasets and the Yale face database indicate the superiority of the proposed algorithm.

Key words: Support Vector Machine (SVM); Penalty factor; Large margin classification thought; Sum of intraclass distance; Projection rule

1 引言

分类一直是人工智能和模式识别等领域研究的热点,分类算法广泛应用于数据挖掘、图像处理、视频跟踪与检测、智能检索、推荐系统、识别系统

等领域,所以分类算法的发展与进步改变着人们的生产和生活方式。目前常用的分类算法有决策树算法、贝叶斯分类器、K 近邻算法、fisher 判别分析法,神经网络算法和 SVM 等分类算法。决策树^[1,2]是通过一定的规则对样本已知属性进行选择达到分类目的的算法。主要的决策树算法有 ID3, C4.5 等。ID3 算法的目的在于减少树的深度,但不能处理具有连续值和具有缺失的数据属性。C4.5 算法在 ID3 算法的基础上对于预测变量的缺值等方面作了较大的改

收稿日期: 2015-05-04; 改回日期: 2016-01-05; 网络出版: 2015-11-19

*通信作者: 王晓初 icnice@yeah.net

基金项目: 国家自然科学基金(61170122, 61272210)

Foundation Items: The National Natural Science Foundation of China (61170122, 61272210)

进,使之适合于分类和回归问题。但决策树本身有容易过度拟合及忽略了数据集中属性之间的相关性等缺点。贝叶斯分类器^[3,4]的分类原理是对属性决定类别可能性大小用训练样本进行判断,然后对新样本进行预测。贝叶斯分类算法假设数据属性之间是相对独立的,而现实中的问题并不总是这样。K近邻算法^[5-7]原理简单易于实现,但它假设样本每个属性的作用都是相同的,当数据集包含有许多不相关属性时,会误导分类过程。fisher判别分析法^[8-10]是找到一个映射法则,使得样本投影到该空间后能在保证方差最小的情况下,将不同的样本很好分开,但该算法仅仅用方差来衡量样本的离散程度,具有一定局限性。神经网络由众多的神经元可调的连接权值连接而成,具有大规模并行处理、分布式信息存储、良好的自组织自学习能力等特点。BP(Back Propagation)网络^[11,12]是一种按误差逆向传播算法训练的多层前馈网络,是目前应用最广泛的神经网络模型之一。神经网络算法和SVM(Support Vector Machine)^[13-16]都支持非线性分类,因此都有很高的精确率,但神经网络算法的可解释性不强,并且需要大量的参数。而SVM以其良好的抗噪声性能在小样本的分类上优秀的表现成为了目前最为常用的分类算法,但SVM最求间隔最大化,并没有太多关注类内样本间的关系,并且引入惩罚因子后增加了训练的难度。

关于SVM的改进大多在时间方面的优化:文献[17~19]提出SMO(Sequential Minimal Optimization),SVMlight和libSVM等算法,其中SVMlight和libSVM已经成为目前最常用的方法,这些算法将SVM时间复杂度由 $O(N^3)$ 降低为 $O(N^2)$ 。其他改进如文献[20]提出的拉格朗日支持向量机(LSVNM)等,最好的情况下可以达到 $O(N)$ 。另外文献[21~23]将SVM推广到半监督领域。文献[24]提出了一种改进的支持向量机NN-SVM算法,它先对训练集进行修剪,根据每个样本与其最近邻类标的异同决定其取舍,然后再用SVM训练得到分类器,改善支持向量机的泛化能力。

上述方法都没有很好地解决引入惩罚因子后,错分的样本在分类间隔之间失去了约束,导致两类交界处样本杂乱分布的问题。本文提出的最小化类内距离和分类算法,利用样本通过映射法则投影到1维空间后可以在该1维空间重新分布的特性,对样本在该1维空间的分布进行训练。归一化投影后正负标签样本均值的相对间隔,且最小化投影后正负标签样本的类内距离和。训练后得到一个映射法则,样本通过该映射法则投影到1维空间后重新分

布为线性可分的形式,同时该方法可引入核函数用于线性不可分的情况。该方法通过最小化类内距离和,使得样本投影到1维空间后同类样本的分布向各自均值集中,非同类样本产生间隔,使得分类更加容易。总地来说本研究的贡献主要有以下几个方面。(1)从距离和大间隔的角度提供了一种新的实现类内差异最小化,类间差异最大化思想下的分类算法。提出了最小化类内距离和准则,用于构造大间隔思想下的分类模型。可以选择不同的距离公式用于不同的要求。(2)适当避免了支持向量机分类算法引入惩罚因子来调节过拟合和线性不可分时无解的问题时允许一部分样本错分,错分的样本在分类间隔之间失去了约束,导致两类交界处样本杂乱分布的问题。同时优化了线性训练时参数选优带来的额外开销。

2 大间隔分类算法

近年来机器学习领域,间隔成为代表性的特征评估策略之一,已成为研究的热点。间隔概念首次由Vapnik在构建SVM模型时提出,SVM也成为运用了大间隔思想最具代表算法。图1为2维空间中对样本用SVM分类算法分类原理图,其中横、纵坐标分别表示样本的两个维度。

如图1所示,SVM的目的就是通过间隔最大化,找到最优超平面 $w^T x + b = 0$,其中 $w \in R^n$ 且 $b \in R$ 。当训练样本线性可分时,则存在超平面使得训练样本中的样本 x_i 满足:

$$y_i (w^T x_i + b) \geq 1, \quad i = 1, 2, \dots, n \quad (1)$$

同时要求的分类间隔 $2/\|w\|$ 最大,那么最优超平面的求解问题可以写为

$$\left. \begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 \\ \text{s.t.} \quad & y_i (w^T x_i + b) \geq 1, \quad i = 1, 2, \dots, n \end{aligned} \right\} \quad (2)$$

引入惩罚参数 $C > 0$ 来控制对错分样本的惩罚程度,用于解决过拟合和线性不可分时无解的问

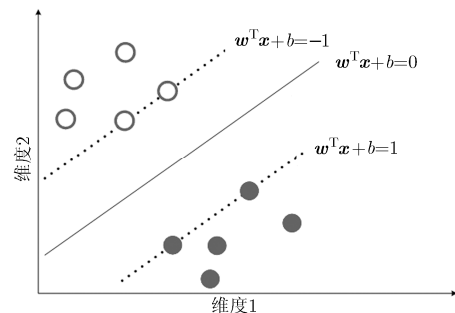


图1 SVM分类原理图

题, $\xi_i \geq 0$ 为松弛变量。式(2)问题变为

$$\left. \begin{aligned} \min & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \\ \text{s.t.} & y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, 2, \dots, n \end{aligned} \right\} \quad (3)$$

在式(3)中引入拉格朗日乘子 $\alpha_i > 0$ 后对应的对偶问题为

$$\left. \begin{aligned} \min & \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j \alpha_i \alpha_j (\mathbf{x}_i \cdot \mathbf{x}_j) - \sum_{i=1}^n \alpha_i \\ \text{s.t.} & \sum_{i=1}^n y_i \alpha_i = 0, 0 \leq \alpha_i \leq C, i = 1, 2, \dots, n \end{aligned} \right\} \quad (4)$$

求对偶问题, 得最优解 $\boldsymbol{\alpha}^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_n^*)^T$, 进而得到最优解 \mathbf{w}^* 和 b^* , 因此得到最优分类函数:

$$f(\mathbf{x}) = \text{sgn}(\mathbf{w}^* \cdot \mathbf{x} + b^*) = \text{sgn}\left(\sum_{i=1}^n \alpha_i^* y_i (\mathbf{x}_i \cdot \mathbf{x}) + b^*\right) \quad (5)$$

对于非线性问题, SVM 通过引入核函数, 其对应的对偶问题式(4)变为

$$\min \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i=1}^n \alpha_i \quad (6)$$

最优决策超平面式(4)变为

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^n \alpha_i^* y_i K(\mathbf{x}_i \cdot \mathbf{x}_j) + b^*\right) \quad (7)$$

支持向量机在分类过程中, 利用结构化风险最小原理, 求出最优决策超平面。并且引入惩罚因子来调节过拟合和线性不可分时无解的问题, 这样做的好处是可以通过调节参数取得最优解, 但同样带来的问题是训练时允许一部分样本错分, 而允许错分的样本并不一定是噪声, 错分的样本在分类间隔之间失去了约束, 导致错分的样本在两类交界处样本杂乱分布; 另外, 引入参数后, 增加了训练的成本。

3 最小化类内距离和分类算法

鉴于支持向量机的上述缺点, 本文依据大间隔思想提出了最小化类内距离和 (Intraclass-Distance-Sum-Minimization, IDSM) 分类算法, 该算法的原理是: 将样本通过映射法则做投影变换,

以最小化类内距离和为准则构造一个最大化类间间隔的训练模型, 对该模型求解, 寻找到最佳的映射法则。样本经过该最佳映射法则转化为线性可分问题。通过与之匹配的判别规则来判定未知样本的属性。由于本文采用最小化类内距离和准则, 使得同类样本聚集, 非同类样本相对远离, 从而两类之间形成间隔。这样所有训练样本都向同类样本聚集。一定程度上解决了两类交界处样本杂乱分布的问题。另外, 线性 IDSM 算法并不需要调节参数, 相比支持向量机训练更加容易。图 2 展示了 IDSM 算法的构造原理。

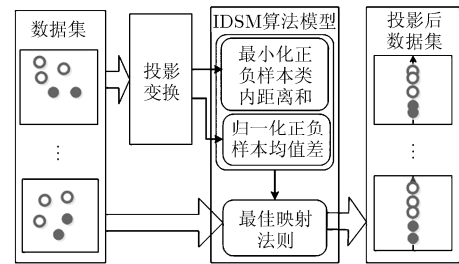


图 2 IDSM 算法构造原理

为了更加清晰地说明两者的区别, 本文在两类服从高斯分布的 2 维数据 (如图 3(a)) 上进行模拟实验, 图 3(b) 是惩罚因子 C 取 0.001 时的 SVM 的分类结果的 3 维图, 可清晰看到两类交界处附近点的分布很杂乱。图 3(c) 是欧氏距离和用于本文算法所得的结果, 可以看到两类样本在投影后的空间上均值相对间隔为 1, 并且两类样本聚集在同类均值附近, 类内更加紧凑, 类间相对形成间隔。

3.1 IDSM 算法框架的提出

样本间的距离和反映了样本在类内的聚集程度。距离和越小, 样本越集中。假设两样本 $\mathbf{x}_i, \mathbf{x}_j$ 间的距离用 $d(\mathbf{x}_i, \mathbf{x}_j)$ 表示。训练样本集 $B = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\}$ 包含 n 个特征为 m 的样本。其正样本子集 $B_1 = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p\}$ 由训练集 B 的前 p 个样本组成。

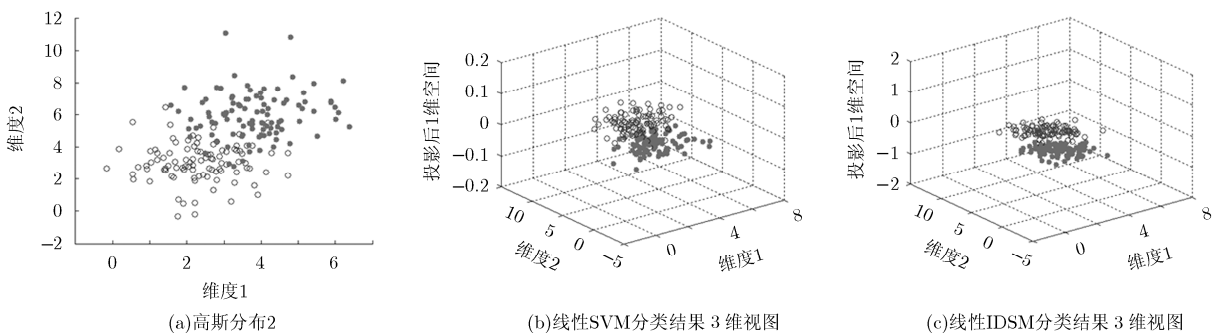


图 3 SVM 与 IDSM 算法区别示意图

其负样本子集 $B_2 = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_q\}$ 由训练集 B 的后 q 个样本组成。本文通过最小化类内距离和的方法来构造最大化类间间隔的训练模型，具体算法如下。

设投影变换的映射法则为 $\mathbf{w} = (w_1, w_2, \dots, w_m)$ ，样本通过映射法则 $f = \sum_{i=1}^m w_i x_i = \mathbf{w}^T \mathbf{z}$ 投影到 1 维空间后，正样本集合表示为 $\{f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_p)\}$ ，负样本集合表示为 $\{f(\mathbf{y}_1), f(\mathbf{y}_2), \dots, f(\mathbf{y}_q)\}$ 。正样本类内距离和与负样本类内距离和的表达式分别为

$$\left. \begin{aligned} D_1 &= \sum_{i=1}^p \sum_{j=1}^p d(f(\mathbf{x}_i), f(\mathbf{x}_j)) \\ D_2 &= \sum_{i=1}^q \sum_{j=1}^q d(f(\mathbf{y}_i), f(\mathbf{y}_j)) \end{aligned} \right\} \quad (8)$$

正样本和负样本的均值表达式为

$$\overline{f(\mathbf{x})} = \frac{1}{p} \sum_{i=1}^p f(\mathbf{x}_i), \quad \overline{f(\mathbf{y})} = \frac{1}{q} \sum_{i=1}^q f(\mathbf{y}_i) \quad (9)$$

归一化正负样本均值差且最小化类内距离和，得到 IDSM 算法训练模型：

$$\left. \begin{aligned} \min_{\mathbf{w}} \quad & \alpha + \beta \\ \text{s.t.} \quad & D_1 \geq \alpha, D_2 \geq \beta, \alpha \geq 0, \beta \geq 0 \\ & \overline{f(\mathbf{x})} - \overline{f(\mathbf{y})} = 1 \end{aligned} \right\} \quad (10)$$

式(10)中的训练模型中 $\min_{\mathbf{w}} \alpha, D_1 \geq \alpha, \alpha \geq 0 \Leftrightarrow \min_{\mathbf{w}} \sqrt{D_1}, \min_{\mathbf{w}} \beta, D_2 \geq \beta, \beta \geq 0 \Leftrightarrow \min_{\mathbf{w}} \sqrt{D_2}$ 。又由于现实情况中正负标签样本的样本个数并不总是一致，为了使正负标签样本平衡分布，用 λ 来统一正负标签样本的量纲。由于训练时正负标签样本比例并不能得知，一般估计为训练集中正负标签样本的个数的比值。IDSM 算法训练模型可以化简为

$$\min_{\mathbf{w}} \sqrt{D_1} + \lambda \sqrt{D_2}, \quad \text{s.t.} \quad \overline{f(\mathbf{x})} - \overline{f(\mathbf{y})} = 1 \quad (11)$$

得到的模型可以用递归最小二乘算法进行求解，进而得到最佳映射法则 \mathbf{w}_* 。对于测试样本 \mathbf{x}_t ，通过最优映射法则投影变换得到： $f(\mathbf{x}_t) = \mathbf{w}_*^T \mathbf{x}_t$ 。由于 IDSM 算法训练模型使得投影后的样本类内距离和达到最小，同类样本向各自均值聚集，从而类间产生间隔。那么，通过最优映射法则变换得到的测试样本 $f(\mathbf{x}_t)$ 便具有靠近同类样本均值，远离非同类样本的特性。因此，可以通过比较经过最佳映射法则投影变换后测试样本与正负训练样本均值之间的距离来判断样本的属性。假设测试样本 $f(\mathbf{x}_t)$ 到正训练样本均值的距离 $L_1 = |f(\mathbf{x}_t) - \overline{f(\mathbf{x})}| = |\mathbf{w}_*^T \mathbf{x}_t - \mathbf{w}_*^T \overline{\mathbf{x}}|$ ，到负训练样本均值的距离 $L_2 = |f(\mathbf{x}_t) - \overline{f(\mathbf{y})}| = |\mathbf{w}_*^T \mathbf{x}_t - \mathbf{w}_*^T \overline{\mathbf{y}}|$ 。IDSM 算法具体判别规则：

$$\text{若 } L_1 \leq L_2, \mathbf{x}_t \text{ 判为正类, 否则为负类} \quad (12)$$

上面介绍了 IDSM 算法的训练模型和判别方

法，IDSM 算法提供了一个框架，可以在不同的需求下选择不同的距离公式。本文选用欧氏距离公式来推理验证 IDSM 算法的有效性。欧氏距离公式下的 IDSM 算法简称为 EIDSM。3.2 节是对 EIDSM 算法的推导。

3.2 EIDSM 算法

欧氏距离作为 IDSM 算法中距离和的衡量标准，投影变换后两样本 $\mathbf{x}_i, \mathbf{x}_j$ 间的距离为

$$d(f(\mathbf{x}_i), f(\mathbf{x}_j)) = \|\mathbf{f}(\mathbf{x}_i) - \mathbf{f}(\mathbf{x}_j)\|^2 \quad (13)$$

那么，欧氏距离和表达式：

$$D_1 = \sum_{i=1}^p \sum_{j=1}^p \|\mathbf{f}(\mathbf{x}_i) - \mathbf{f}(\mathbf{x}_j)\|^2, D_2 = \sum_{i=1}^q \sum_{j=1}^q \|\mathbf{f}(\mathbf{y}_i) - \mathbf{f}(\mathbf{y}_j)\|^2 \quad (14)$$

由于 $\mathbf{f}(\mathbf{x}_i) = \mathbf{w}^T \mathbf{x}_i$ ，可以将正类样本欧氏距离和表达式化为矩阵的形式：

$$\begin{aligned} D_1 &= \sum_{i=1}^p \sum_{j=1}^p \|\mathbf{w}^T \mathbf{x}_i - \mathbf{w}^T \mathbf{x}_j\|^2 \\ &= \sum_{i=1}^p \sum_{j=1}^p \mathbf{w}^T (\mathbf{x}_i - \mathbf{x}_j) (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{w} \\ &= \sum_{i=1}^p \sum_{j=1}^p \mathbf{w}^T (\mathbf{x}_i \mathbf{x}_i^T + \mathbf{x}_j \mathbf{x}_j^T - \mathbf{x}_i \mathbf{x}_j^T - \mathbf{x}_j \mathbf{x}_i^T) \mathbf{w} \\ &= 2\mathbf{w}^T (\mathbf{X} \mathbf{A}_1 \mathbf{X}^T - \mathbf{X} \mathbf{E}_1 \mathbf{X}^T) \mathbf{w} = 2\mathbf{w}^T (\mathbf{X} \mathbf{H}_1 \mathbf{X}^T) \mathbf{w} \end{aligned} \quad (15)$$

同理，由于 $\mathbf{f}(\mathbf{y}_i) = \mathbf{w}^T \mathbf{y}_i$ ，可以将负类样本欧氏距离和表达式化为矩阵的形式：

$$\begin{aligned} D_2 &= 2\mathbf{w}^T (\mathbf{Y} \mathbf{A}_2 \mathbf{Y}^T - \mathbf{Y} \mathbf{E}_2 \mathbf{Y}^T) \mathbf{w} \\ &= 2\mathbf{w}^T (\mathbf{Y} \mathbf{H}_2 \mathbf{Y}^T) \mathbf{w} \end{aligned} \quad (16)$$

其中 \mathbf{X}, \mathbf{Y} 分别为正训练样本和负训练样本组成的矩阵， $\mathbf{X} \in \mathbb{R}^{m \times p}, \mathbf{Y} \in \mathbb{R}^{m \times q}$ ； \mathbf{A}_1 为 $p \times p$ 的对角矩阵，对角元素都为 p ， \mathbf{A}_2 为 $q \times q$ 的对角矩阵，对角元素都为 q ； \mathbf{E}_1 和 \mathbf{E}_2 分别为 $p \times p$ 和 $q \times q$ 的全 1 矩阵； $\mathbf{H}_1 = \mathbf{A}_1 - \mathbf{E}_1$ ， $\mathbf{H}_2 = \mathbf{A}_2 - \mathbf{E}_2$ 。

那么欧氏距离和式(15)，式(16)衡量的最小化类内距离和分类算法(EIDSM 算法)，去掉常数 2，不影响结果，得目标函数：

$$\left. \begin{aligned} \min_{\mathbf{w}} \quad & \sqrt{\mathbf{w}^T (\mathbf{X} \mathbf{H}_1 \mathbf{X}^T) \mathbf{w}} + \lambda \sqrt{\mathbf{w}^T (\mathbf{Y} \mathbf{H}_2 \mathbf{Y}^T) \mathbf{w}} \\ \text{s.t.} \quad & \mathbf{w}^T (\overline{\mathbf{x}} - \overline{\mathbf{y}}) = 1 \end{aligned} \right\} \quad (17)$$

此处 $\lambda = p/q$ ，很显然该目标函数优化问题为凸优化问题^[25,26]，可以解出最佳映射法则为 \mathbf{w}_* ，具体的算法见第 4 节，判别方法见 3.1 节式(12)。

3.3 非线性 EIDSM 算法

将样本映射到非线性，数学表达为： $\mathbf{x} \mapsto \varphi(\mathbf{x})$ ， $\mathbf{y} \mapsto \varphi(\mathbf{y})$ 。这里 $\varphi: \mathbb{R}^m \mapsto \mathbb{R}^n$ ，核函数^[27]可写为

$K(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}_j)$ 。引入核函数后, 样本投影到 1 维空间的分布表示为 $f(\varphi(\mathbf{z})) = \mathbf{w}^T \varphi(\mathbf{z})$, ($\mathbf{w} \in \mathbb{R}^n$)。此时非线性欧氏距离和可以表达为

$$\left. \begin{aligned} D_1^k &= \sum_{i=1}^p \sum_{j=1}^p \left\| f(\varphi(\mathbf{x}_i)) - f(\varphi(\mathbf{x}_j)) \right\|^2 \\ D_1^k &= \sum_{i=1}^q \sum_{j=1}^q \left\| f(\varphi(\mathbf{y}_i)) - f(\varphi(\mathbf{y}_j)) \right\|^2 \end{aligned} \right\} \quad (18)$$

假设 \mathbf{w} 中有一分量正交于 $\varphi(\mathbf{x}_i)$, $i = 1, 2, \dots, p$ 和 $\varphi(\mathbf{y}_i)$, $i = 1, 2, \dots, q$ 的子空间, 那么这个分量不会影响到目标函数的结果, 因此我们可以将 \mathbf{w} 写成^[28]:

$$\mathbf{w} = \sum_{i=1}^p \alpha_i \varphi(\mathbf{x}_i) + \sum_{j=1}^q \beta_j \varphi(\mathbf{y}_j) \quad (19)$$

设

$$\boldsymbol{\gamma} = [\alpha_1 \ \alpha_2 \ \dots \ \alpha_p \ \beta_1 \ \beta_2 \ \dots \ \beta_q] \quad (20)$$

$$[\tilde{\mathbf{k}}_x]_l = \frac{1}{p} \sum_{i=1}^p K(\mathbf{x}_i, \mathbf{z}_l), \quad [\tilde{\mathbf{k}}_y]_l = \frac{1}{q} \sum_{j=1}^q K(\mathbf{y}_j, \mathbf{z}_l) \quad (21)$$

$$\left. \begin{aligned} \mathbf{K}_1 &= K(\mathbf{x}_i, \mathbf{z}_i), \quad \mathbf{K}_2 = K(\mathbf{y}_j, \mathbf{z}_i) \\ \tilde{\mathbf{K}}_1 &= \mathbf{K}_1 - \mathbf{1}_p \tilde{\mathbf{k}}_x^T, \quad \tilde{\mathbf{K}}_2 = \mathbf{K}_2 - \mathbf{1}_q \tilde{\mathbf{k}}_y^T \end{aligned} \right\} \quad (22)$$

其中, $i = 1, 2, \dots, p$, $j = 1, 2, \dots, q$, $l = 1, 2, \dots, n$, $\mathbf{1}_p$ 和 $\mathbf{1}_q$ 为 p 维和 q 维全 1 列向量。

用 $\boldsymbol{\gamma}$ 来表示 \mathbf{w} , 此时投影到 1 维空间的样本均值 $\overline{f(\varphi(\mathbf{x}))}$ 的非线性形式为

$$\begin{aligned} \overline{f(\varphi(\mathbf{x}))} &= \mathbf{w}^T \overline{\varphi(\mathbf{x})} = \boldsymbol{\gamma}^T \tilde{\mathbf{k}}_x, \overline{f(\varphi(\mathbf{y}))} \\ &= \mathbf{w}^T \overline{\varphi(\mathbf{y})} = \boldsymbol{\gamma}^T \tilde{\mathbf{k}}_y \end{aligned} \quad (23)$$

此时正样本欧氏距离和的非线性表达式可化简为矩阵形式:

$$\begin{aligned} D_1^k &= \sum_{i=1}^p \sum_{j=1}^p \left\| f(\varphi(\mathbf{x}_i)) - f(\varphi(\mathbf{x}_j)) \right\|^2 \\ &= \sum_{i=1}^p \sum_{j=1}^p \boldsymbol{\gamma}^T \left\| \tilde{\mathbf{k}}_{x_i} - \tilde{\mathbf{k}}_{x_j} \right\|^2 \boldsymbol{\gamma} \\ &= 2\boldsymbol{\gamma}^T (\mathbf{K}_1^T \mathbf{A}_1 \mathbf{K}_1 - \mathbf{K}_1^T \mathbf{E}_1 \mathbf{K}_1) \boldsymbol{\gamma} \\ &= 2\boldsymbol{\gamma}^T (\mathbf{K}_1^T \mathbf{H}_1 \mathbf{K}_1) \boldsymbol{\gamma} \end{aligned} \quad (24)$$

同理, 负样本欧氏距离和的非线性表达式可化简为矩阵形式:

$$\begin{aligned} D_2^k &= 2\boldsymbol{\gamma}^T (\mathbf{K}_2^T \mathbf{A}_2 \mathbf{K}_2 - \mathbf{K}_2^T \mathbf{E}_2 \mathbf{K}_2) \boldsymbol{\gamma} \\ &= 2\boldsymbol{\gamma}^T (\mathbf{K}_2^T \mathbf{H}_2 \mathbf{K}_2) \boldsymbol{\gamma} \end{aligned} \quad (25)$$

那么非线性欧氏距离和式(24), 式(25)衡量的最小化类内距离和分类算法(非线性 EIDSM), 去掉常数 2, 不影响结果, 可以写成如式(26)目标函数:

$$\left. \begin{aligned} \min_{\boldsymbol{\gamma}} & \sqrt{\boldsymbol{\gamma}^T (\mathbf{K}_1^T \mathbf{H}_1 \mathbf{K}_1) \boldsymbol{\gamma}} + \lambda \sqrt{\boldsymbol{\gamma}^T (\mathbf{K}_2^T \mathbf{H}_2 \mathbf{K}_2) \boldsymbol{\gamma}} \\ \text{s.t.} & \boldsymbol{\gamma}^T (\tilde{\mathbf{k}}_x - \tilde{\mathbf{k}}_y) = 1 \end{aligned} \right\} \quad (26)$$

此处 $\lambda = p/q$, 对目标函数式(26)进行求解, 算法同

线性部分, 得最佳映射法则 $\boldsymbol{\gamma}^*$ 。

对于一个测试的样本 \mathbf{x}_t , \mathbf{x}_t 与正样本中心之间的距离为 L_1^k :

$$\begin{aligned} L_1^k &= \left| f(\varphi(\mathbf{x}_t)) - \overline{f(\varphi(\mathbf{x}))} \right| = \left| \mathbf{w}_*^T \varphi(\mathbf{x}_t) - \mathbf{w}_*^T \overline{\varphi(\mathbf{x})} \right| \\ &= \left| \sum_{i=1}^n [\boldsymbol{\gamma}_*]_i K(\mathbf{z}_i, \mathbf{x}_t) - \sum_{i=1}^n [\boldsymbol{\gamma}_*]_i K(\mathbf{z}_i, \bar{\mathbf{x}}) \right| \end{aligned} \quad (27)$$

\mathbf{x}_t 与负样本中心之间的距离为 L_2^k :

$$\begin{aligned} L_2^k &= \left| f(\varphi(\mathbf{x}_t)) - \overline{f(\varphi(\mathbf{y}))} \right| = \left| \mathbf{w}_*^T \varphi(\mathbf{x}_t) - \mathbf{w}_*^T \overline{\varphi(\mathbf{y})} \right| \\ &= \left| \sum_{i=1}^n [\boldsymbol{\gamma}_*]_i K(\mathbf{z}_i, \mathbf{x}_t) - \sum_{i=1}^n [\boldsymbol{\gamma}_*]_i K(\mathbf{z}_i, \bar{\mathbf{y}}) \right| \end{aligned} \quad (28)$$

判别方法见 3.1 节式(12)。

4 目标函数求解算法

目标函数式(17), 式(26)是非线性有约束优化问题, 并且是凸优化问题。有多种方法^[29-31]可以选择, 比如拟牛顿法, Rosen 梯度投影法等, 线性和非线性目标函数可以化简到相同的形式, 我们只解析线性目标函数算法过程。这里我们用解析法对该优化问题进行求解, 算法步骤如下:

步骤 1 令 $\mathbf{M}_1 = \mathbf{X} \mathbf{H}_1 \mathbf{X}^T$, $\mathbf{M}_2 = \lambda \mathbf{Y} \mathbf{H}_2 \mathbf{Y}^T$, $\mathbf{w} = \mathbf{w}_0 + \mathbf{F} \mathbf{v}$, \mathbf{F} 的列正交于 $\bar{\mathbf{x}} - \bar{\mathbf{y}}$;

步骤 2 目标函数式(17)可写为

$$\begin{aligned} \min_{\boldsymbol{\varepsilon}, \boldsymbol{\rho}, \mathbf{v}} & \boldsymbol{\varepsilon}_k + \frac{1}{\boldsymbol{\varepsilon}_k} \left\| \mathbf{M}_1^{1/2} (\mathbf{w}_k + \mathbf{F} \mathbf{v}_k) \right\|_2^2 + \boldsymbol{\rho}_k \\ & + \frac{\lambda}{\boldsymbol{\rho}_k} \left\| \mathbf{M}_2^{1/2} (\mathbf{w}_k + \mathbf{F} \mathbf{v}_k) \right\|_2^2 \end{aligned} \quad (29)$$

步骤 3 初始化: $\mathbf{w}_0 = (\bar{\mathbf{x}} - \bar{\mathbf{y}}) / \|\bar{\mathbf{x}} - \bar{\mathbf{y}}\|_2$, $\boldsymbol{\varepsilon}_0 = 1$, $\boldsymbol{\rho}_0 = 1$ 。并代入式(29)得到最小二乘问题, 最细小二乘法求解得 \mathbf{v}_0 ;

步骤 4 $k = 1$;

步骤 5 $\mathbf{w}_k = \mathbf{w}_{k-1} + \mathbf{F} \mathbf{v}_{k-1}$, $\boldsymbol{\varepsilon}_k = \left\| \mathbf{H}_1^{1/2} \mathbf{w}_{k-1} \right\|_2^2$, $\boldsymbol{\rho}_k = \lambda \left\| \mathbf{H}_2^{1/2} \mathbf{w}_{k-1} \right\|_2^2$ 并代入式(29), 最小二乘法求解得 \mathbf{v}_k ;

步骤 6 $k = k + 1$;

步骤 7 重复步骤 5, 步骤 6 直到收敛或满足停止条件;

步骤 8 $\mathbf{w}_* = \mathbf{w}_k$;

步骤 9 根据式(12)对新样本进行归类。

分析最小二乘的递归求解过程, 本文算法在线性和非线性的时间复杂度分别为 $O(km^3)$ 和 $O(kn^3)$, k 为最小二乘算法收敛时循环的次数。

5 实验与分析

设计了两组试验来验证本文方法的有效性。第

1 组用 UCI 公共有效数据集中的 10 个数据集进行随机抽取 75% 的样本用于训练, 剩余的 25% 用于测试, 记录正确率和每次运行所用时间(s)。表 1 列出了 10 个 UCI 数据集的信息。第 2 组用 Yale 大学人脸数据库, 在小样本下, 不断增加训练样本个数, 记录正确率和每次运行所用时间(s)。这样设计可验证其小样本上的表现。鉴于 IDSM 算法和 fisher 用于分类的算法^[32]都具有使类内样本聚集的特点, 相同之处在于他们都利用了最小化类内差异的思想, 区别在于 IDSM 算法是用距离来度量类内和类间的差异, 而 fisher 用于分类算法是用方差来度量类内差异; 其次 IDSM 算法由于引入了 λ , 对不平衡数据

适应性相对较好。所以用 SVM 和 fisher 作为 EIDSM 算法的对比试验。表 2 包含了对实验平台和部分算法英文缩写的说明。

5.1 UCI 数据集

实验中非线性算法都选用高斯核函数 $\exp\left(\frac{-\|x-y\|^2}{\sigma^2}\right)$, σ 在 $2^{-15} \sim 2^{15}$ 中以 2 为等比间隔选优, 对比试验 SVM(SMO 算法)的惩罚参数 C 在 $10^{-6} \sim 10^6$ 中以 10 为等比间隔选优。表 3 记录了 10 次随机独立重复所得的正确率的均值与方差, 表 4 记录了 10 次独立重复实验所用时间的均值与方差。

表 1 UCI 数据集信息

数据集	breast	cleve	German	haberman	heart	ionosphere	liver	pima	sonar	WBC
正样本个数	81	160	300	225	150	126	145	268	111	444
负样本个数	196	136	700	81	120	225	200	500	97	239
属性个数	9	13	24	3	13	34	6	8	60	9

表 2 实验说明

实验平台	CPU: Intel(R)Core(TM)2Duo 3.00 GHz RAM: 4.00 GB 系统: 64 位 Win8.1 操作系统 工具: 32 位 MatlabR2012a
实验中缩写算法的全称	LSVM: 线性支持向量机 KSVM: 非线性支持向量机 LFISHER: fisher 线性判别分类算法 KFISHER: 核 fisher 判别分类算法 LEIDSM: 线性欧氏距离和衡量的最小化类内距离和分类算法 KEIDSM: 非线性欧氏距离和衡量的最小化类内距离和分类算法

表 3 UCI 数据集的正确率测试结果(%)

UCI	LSVM	KSVM	LFISHER	KFISHER	LEIDSM	KEIDSM
breast	75.36±3.92	75.71±4.74	66.81±3.58	72.17±2.38	72.17±5.19	74.88±2.11
cleve	51.30±4.38	68.15±5.62	81.43±2.31	81.84±3.50	82.59±3.52	83.01±2.33
German	32.16±1.71	73.32±1.88	71.60±3.35	71.68±2.20	73.84±2.01	69.20±3.86
haberman	71.84±3.34	77.37±4.96	72.76±3.40	76.58±2.35	75.63±0.72	79.39±2.47
heart	51.34±3.89	68.36±2.45	83.13±4.39	80.30±4.34	87.16±3.27	80.63±3.11
ionosphere	88.41±8.22	91.82±3.25	87.16±2.99	87.27±3.15	90.91±2.41	89.39±5.37
liver	52.56±9.60	64.19±3.03	61.63±4.87	62.09±2.11	64.88±7.04	67.83±2.69
pima	46.06±17.68	69.06±4.37	73.83±3.09	69.08±3.13	77.92±2.46	76.83±2.27
sonar	79.52±4.17	86.92±2.85	73.00±6.14	73.63±8.64	75.00±3.85	80.13±4.84
WBC	96.37±1.12	97.89±0.89	96.19±0.88	97.54±0.64	96.61±0.96	97.96±0.89
均值	64.49±5.80	77.28±3.40	76.75±3.50	77.22±3.24	79.67±3.14	79.93±2.99

表 4 UCI 数据集的时间测试结果(s)

UCI	LSVM	K SVM	LFISHER	KFISHER	LEIDSM	KEIDSM
breast	0.1719±0.0082	0.0780±0.0061	0.0035±0.0004	0.1156±0.0301	0.0040±0.0004	0.1780±0.0060
cleve	0.4265±0.2948	0.0961±0.0039	0.0031±0.0003	0.0906±0.0007	0.0043±0.0003	0.0746±0.0028
German	3.7270±0.5067	0.6778±0.0139	0.0094±0.0037	1.3044±0.2950	0.0181±0.0012	2.6074±0.3863
haberman	0.2881±0.2186	0.0762±0.0110	0.0041±0.0035	0.1088±0.0381	0.0035±0.0005	0.2166±0.0130
heart	0.4100±0.2112	0.0839±0.0048	0.0047±0.0005	0.0781±0.0053	0.0043±0.0005	0.0682±0.0032
ionosphere	0.4992±0.2488	0.1208±0.0049	0.0063±0.0004	0.1469±0.0075	0.0081±0.0004	0.3057±0.0075
liver	0.2979±0.1338	0.0764±0.0057	0.0054±0.0011	0.1031±0.0013	0.0044±0.0009	0.1621±0.0080
pima	1.0151±0.3822	0.2824±0.0079	0.0070±0.0005	0.3719±0.0104	0.0096±0.0007	0.5191±0.0202
sonar	0.6587±0.1144	0.0679±0.0046	0.0062±0.0002	0.0500±0.0020	0.0100±0.0004	0.0654±0.0045
WBC	0.1167±0.0442	0.1290±0.0029	0.0031±0.0017	0.5125±0.0036	0.0089±0.0009	1.4975±0.3204
均值	0.7611±0.2163	0.1689±0.0066	0.0043±0.0012	0.2882±0.039	0.0075±0.0006	0.5695±0.0752

根据表 3 所示, 通过比较本文 LEIDSM 算法与其他对比算法在各 UCI 数据集上的实验结果可知:

(1)LEIDSM 和 KEIDSM 算法在大部分数据集上取得了比 LSVM 和 K SVM 更高的正确率。说明本文算法的通过类内距离和最小化准则构造的大间隔分类模型在一定程度上优于 SVM 结构风险化最小和引入惩罚因子构造的标准 SVM 大间隔分类模型。

(2)LEIDSM 和 KEIDSM 算法在大部分数据集上取得了比 LFISHER 和 KFISHER 更高的正确率。分析数据知, 对于正负样本个数差别不大的数据集, 如 cleve, liver, sonar 等上的表现, fisher 用于分类的算法与本文算法正确率差距相对较小, 而在正负样本个数差别很大的数据集, 如 breast 等上的表现, fisher 算法与本文算法正确率差距相对较大。这是由于本文算法在提出时便考虑到样本不平衡时算法的适应性问题, 加入了参数 λ , 一定程度上解决了数据样本不平衡带来的不适性。而 fisher 算法认为两类均衡, 在不平衡数据集上表现相对较差。

根据表 4 所示, 线性方面, LFISHER 平均用时最短, LSVM 用时最长, LEIDSM 介于他们之间。非线性方面, K SVM 用时最短。由于本文采用递归最小二乘法对算法进行求解, 在线性时时间复杂度为 $O(km^3)$, 非线性时时间复杂度为 $O(kn^3)$ 。而 fisher 算法线性和非线性时间复杂度分别为 $O(m^3)$ 和 $O(n^3)$ 。由于 SVM 用经典的改进算法 SMO 作对比, SMO 在线性和非线性平均时间复杂度为 $O(n^2)$ 。分析实验中 UCI 数据集知, 几乎所有数据集的特征个数远小于训练样本数, 因此在线性时有 $O(n^2) > O(km^3) \geq O(m^3)$, 所以 fisher 算法平均用时最短。而非线性时, 在训练样本个数相同的情况下, 有 $O(kn^3) \geq O(n^3) > O(n^2)$, 所以 K SVM 用时最短。

5.2 Yale 大学人脸数据库

Yale 大学人脸数据库由 Yale 大学计算视觉与扼制中心创立, 包括 15 位志愿者的 165 张图片, 每人 11 张照片, 主要包括光照条件和表情等的变化, 每幅图片重新采样为 10×10 的大小。实验参数的设置与 5.1 节相同, 实验中采取两两配对的方法, 共有 105 种配对方式, 在每个训练样本级别都取这 105 种准确率结果的平均值。每个图像分别用 2, 3, 4, 5, 6, 7, 8, 9, 10 个进行训练, 剩下的样本作为测试样本。表 5 记录了 5 次独立重复实验的正确率均值, 表 6 记录了平均每次运行所用的时间均值。

依据表 5 所示, 样本变化时, 随着训练样本的增多, 各种算法下分类的正确率几乎都有提升。

(1)比较各算法线性情况在样本变时的表现: LEIDSM 大部分情况下比 LSVM 和 LFISHER 正确率高。这是由于人脸本身是非线性的, SVM 引入惩罚因子来解决非线性无解的情况, 允许部分样本错分, 错分后的样本在两类交界处杂乱分布。而本文算法, 使得投影变换后的样本向各自均值聚集, 并且本文算法训练和判别都用距离最小准则, 训练和判别更加匹配, 所以更有利于分类。

(2)比较各算法非线性情况在样本变化时的表现: 在样本个数较少时, K SVM 比 KEIDSM 和 KFISHER 表现要好, 而当样本个数增大到一定程度时, KEIDSM 表现更好。这是因为 SVM 只需关键的支持向量, 所以对小样本有很强的学习能力。而样本增加时, 本文算法的类内距离和利用了更多的样本信息, 所以取得了更好的结果。另外, 由于实验 105 个配对分类, 每次实验对不同的人脸图像分类选用的是同一个参数, 若是算法对参数较为敏感, 实验结果会偏差。

表 5 Yale 人脸数据库实验准确率测试结果(%)

Yale	LSVM	K SVM	LFISHER	KFISHER	LEIDSM	KEIDSM
2	81.48±17.41	88.35±13.34	88.41±24.52	84.11±14.21	91.80±8.89	86.98±8.95
3	85.60±14.17	86.42±15.70	93.21±9.33	81.79±15.19	92.62±9.45	87.68±9.96
4	88.44±10.02	89.09±10.3	91.16±10.46	82.04±15.09	92.04±10.21	88.78±8.63
5	91.43±7.88	91.75±7.96	94.05±7.55	80.57±16.74	94.29±8.36	86.67±10.80
6	89.43±12.62	92.00±9.24	93.81±8.01	85.43±17.37	92.48±10.72	90.67±8.69
7	91.19±10.67	91.43±1.30	94.16±5.23	86.50±18.70	91.79±11.73	91.67±9.44
8	88.73±13.30	89.05±13.24	95.24±8.24	89.31±18.12	95.87±7.59	93.02±9.18
9	93.57±11.51	94.76±6.67	94.76±10.79	90.71±17.63	96.67±8.54	96.90±8.66
10	93.81±16.55	94.47±9.76	92.38±11.03	86.67±16.51	96.67±12.53	97.14±10.06

表 6 Yale 人脸数据库实验时间测试结果(s)

Yale	LSVM	K SVM	LFISHER	KFISHER	LEIDSM	KEIDSM
2	0.0059±0.0008	0.0049±0.0003	0.0060±0.0004	0.0025±0.0003	0.1378±0.2362	0.0303±0.0554
3	0.0069±0.0009	0.0053±0.0009	0.0071±0.0003	0.0058±0.0007	0.0085±0.0026	0.0069±0.0005
4	0.0070±0.0014	0.0055±0.0009	0.0071±0.0017	0.0060±0.0007	0.0171±0.0083	0.0076±0.0003
5	0.0073±0.0019	0.0066±0.0016	0.0079±0.0006	0.0042±0.0002	0.0077±0.0003	0.0084±0.0005
6	0.0087±0.0015	0.0071±0.0009	0.0077±0.0006	0.0075±0.0013	0.0073±0.0005	0.0092±0.0010
7	0.0093±0.0017	0.0088±0.0016	0.0109±0.0012	0.0057±0.0005	0.0110±0.0004	0.0095±0.0006
8	0.0105±0.0025	0.0090±0.0014	0.0116±0.0009	0.0126±0.0111	0.0145±0.0011	0.0103±0.0007
9	0.0125±0.0039	0.0098±0.0009	0.0058±0.0018	0.0135±0.0056	0.0078±0.0008	0.0113±0.0009
10	0.0104±0.0030	0.0101±0.0011	0.0094±0.0002	0.0138±0.0019	0.0082±0.0009	0.0116±0.0004

依据表 6 所示, 样本增加时各算法用时都变长。大部分情况下 LSVM 和 K SVM 用时更短, 这是由于图片本身维度为 100 维, 所以有 $O(km^3) \geq O(m^3) > O(n^2)$, 此时 LEIDSM, LFISHER 失去了线性方面时间复杂度的优势。

6 结束语

本文依据大间隔思想, 提出了最小化类内距离和分类算法。并且在欧式距离标准下对本文算法进行推理实验, 通过与 SVM 和 fisher 判别分类算法的比较, 证明了该分类算法的有效性。另外, 本文只在欧氏距离下验证了算法的有效性, 欧氏距离要求数据样本在模式空间中呈球状分布, 接下来的工作可以针对不同的数据选择不同的距离公式来构建距离矩阵拓展本文算法。该分类算法时间复杂度方面还有待改进, 且该分类算法是一种二分类算法, 对于多类问题, 需要设计多分类器。

参考文献

- [1] QUINLAN J R. Induction of decision trees[J]. *Machine Learning*, 1986, 1(1): 81-106.
- [2] QUINLAN J R. Improved use of continuous attributes in C4.5[J]. *Journal of Artificial Intelligence Research*, 1996, 4(1): 77-90.
- [3] PENG F, SCHUURMANS D, and WANG S. Augmenting naive Bayes classifiers with statistical language models[J]. *Information Retrieval*, 2004, 7(3/4): 317-345.
- [4] CHENG J and GREINER R. Comparing Bayesian network classifiers[C]. *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, San Francisco, USA, 1999: 101-108.
- [5] COVER T and HART P. Nearest neighbor pattern classification [J]. *IEEE Transactions on Information Theory*, 1967, 13(1): 21-27.
- [6] BIJALWAN V, KUMAR V, KUMARI P, et al. KNN based machine learning approach for text and document mining[J]. *International Journal of Database Theory and Application*, 2014, 7(1): 61-70.
- [7] 黄剑华, 丁建睿, 刘家锋, 等. 基于局部加权的 Citation-kNN 算法[J]. *电子与信息学报*, 2013, 35(3): 627-632.

- HUANG Jianhua, DING Jianrui, LIU Jiafeng, et al. Citation-kNN algorithm based on locally-weighting[J]. *Journal of Electronics & Information Technology*, 2013, 35(3): 627-632.

- [8] WELLING M. Fisher linear discriminant analysis[J]. *Department of Computer Science*, 2008, 16(94): 237-280.
- [9] FUIN N, PEDEMONTE S, ARRIDGE S, *et al.* Efficient determination of the uncertainty for the optimization of SPECT system design: a subsampled fisher information matrix[J]. *IEEE Transactions on Medical Imaging*, 2014, 33(3): 618-635.
- [10] DUFRENOIS F. A one-class kernel fisher criterion for outlier detection[J]. *IEEE Transactions on Neural Networks & Learning Systems*, 2014, 26(5): 982-994.
- [11] VAN Ooyen A and NIENHUIS B. Improving the convergence of the back-propagation algorithm[J]. *Neural Networks*, 1992, 5(3): 465-471.
- [12] 潘舟浩, 李道京, 刘波, 等. 基于 BP 算法和时变基线的机载 InSAR 数据处理方法研究[J]. *电子与信息学报*, 2014, 36(7): 1585-1591.
PAN Zhouhao, LI Daojing, LIU Bo, *et al.* Processing of the airborne InSAR data based on the BP algorithm and the time-varying baseline[J]. *Journal of Electronics & Information Technology*, 2014, 36(7): 1585-1591.
- [13] SUYKENS J A K and VANDEWALLE J. Least squares support vector machine classifiers[J]. *Neural Processing Letters*, 1999, 9(3): 293-300.
- [14] 胡文军, 王士同, 王娟, 等. 非线性分类的分割超平面快速集成方法[J]. *电子与信息学报*, 2012, 34(3): 535-542.
HU Wenjun, WANG Shitong, WANG Juan, *et al.* Fast ensemble of separating hyperplanes for nonlinear classification[J]. *Journal of Electronics & Information Technology*, 2012, 34(3): 535-542.
- [15] GAO X, LU T, LIU P, *et al.* A soil moisture classification model based on SVM used in agricultural WSN[C]. 2014 IEEE 7th Joint International, Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, 2014: 432-436.
- [16] RIES C X, RICHTER F, ROMBERG S, *et al.* Automatic object annotation from weakly labeled data with latent structured SVM[C]. 2014 12th IEEE International Workshop on Content-based Multimedia Indexing (CBMI), Klagenfurt, Austria, 2014: 1-4.
- [17] PLATT J. Fast training of support vector machines using sequential minimal optimization[J]. *Advances in Kernel Methods: Support Vector Learning*, 1999(3): 185-208.
- [18] JOACHIMS T. Making large scale SVM learning practical[R]. Universität Dortmund, 1999.
- [19] CHANG C C and LIN C J. LIBSVM: A library for support vector machines[J]. *ACM Transactions on Intelligent Systems & Technology*, 2011, 2(3): 389-396.
- [20] MANGASARIAN O L and MUSICANT D R. Lagrangian support vector machines[J]. *The Journal of Machine Learning Research*, 2001, 1(3): 161-177.
- [21] SEOK K. Semi-supervised regression based on support vector machine[J]. *Computer Engineering & Applications*, 2014, 25(2): 447-454.
- [22] LENG Y, XU X, and QI G. Combining active learning and semi-supervised learning to construct SVM classifier[J]. *Knowledge-Based Systems*, 2013, 44(1): 121-131.
- [23] CHEN W J, SHAO Y H, XU D K, *et al.* Manifold proximal support vector machine for semi-supervised classification[J]. *Applied Intelligence*, 2014, 40(4): 623-638.
- [24] 李红莲, 王春花, 袁保宗. 一种改进的支持向量机 NN-SVM[J]. *计算机学报*, 2003, 26(8): 1015-1020.
LI Honglian, WANG Chunhua, and YUAN Baozong. An improved SVM: NN-SVM[J]. *Chinese Journal of Computers*, 2003, 26(8): 1015-1020.
- [25] 陈宝林. 最优化理论与算法[M]. 北京: 清华大学出版社, 2005: 281-322.
CHEN Baolin. *Optimization Theory and Algorithm*[M]. Beijing, Tsinghua University Press, 2005: 281-322.
- [26] YOSHIYAMA K and SAKURAI A. Laplacian minimax probability machine[J]. *Pattern Recognition Letters*, 2014, 37: 192-200.
- [27] MIGLIORATI G. Adaptive polynomial approximation by means of random discrete least squares[J]. *Lecture Notes in Computational Science & Engineering*, 2013, 103: 547-554.
- [28] HUANG K, YANG H, KING I, *et al.* The minimum error minimax probability machine[J]. *The Journal of Machine Learning Research*, 2004(5): 1253-1286.
- [29] PLAN Y and VERSHYNIN R. Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach[J]. *IEEE Transactions on Information Theory*, 2013, 59(1): 482-494.
- [30] ALIZADEN F. Interior point methods in semidefinite programming with applications to combinatorial optimization[J]. *SIAM Journal on Optimization*, 1995, 5(1): 13-51.
- [31] BOYD S and VANDENBERGHE L. *Convex Optimization*[M]. Cambridge University Press, 2009: 127-214.
- [32] 边肇祺, 张学工. 模式识别[M]. 北京: 清华大学出版社, 2001: 83-90.
BIAN Zhaoqi and ZHANG Xuegong. *Pattern Recognition*[M]. Beijing, Tsinghua University Press, 2001: 83-90.
- 王晓初: 男, 1987年生, 硕博连读生, 研究方向为模式识别、数字图像处理。
- 王士同: 男, 1964年生, 教授, 研究方向为人工智能、模式识别、生物信息。
- 包芳: 女, 1970年生, 教授, 研究方向为人工智能、模式识别。
- 蒋亦樟: 男, 1988年生, 博士, 研究方向为模式识别、智能计算。