

## 低阶数据映射和特征加权的线性 SVM

王瑞, 向新, 肖冰松

(空军工程大学航空工程学院, 西安, 710038)

**摘要** 针对传统线性支持向量机在训练数据集时均等对待每一维输入特征,以及在原始空间直接分类造成预测准确率低的问题,提出低阶多项式数据映射和特征加权相结合的方法,来提高线性支持向量机的分类性能。该方法首先将每个样本映射到多项式核对应的2-阶显式特征空间,从而增加样本的隐性信息,然后使用模糊熵特征加权算法计算每一维特征的权重,通过权重衡量特征对分类结果的贡献大小。从不同数据库选取7个数据集进行测试,在训练时间和预测准确率2个方面将该方法与核支持向量机、线性支持向量机的其他改进算法进行比较。结果显示,随着数据集规模的扩大,训练时间降低一个数量级,预测准确率在一些数据集上取得与核支持向量机相接近的效果。结果表明:所提方法可以有效提高线性支持向量机的整体性能。

**关键词** 线性支持向量机;核支持向量机;低阶多项式映射;隐性信息;特征加权;模糊熵

**DOI** 10.3969/j.issn.1009-3516.2019.04.011

**中图分类号** TP391 **文献标志码** A **文章编号** 1009-3516(2019)04-0072-06

## Low-Degree Data Mappings and Feature Weighted via Linear SVM

WANG Rui, XIANG Xin, XIAO Bingsong

(Aeronautics Engineering College, Air Force Engineering University, Xi'an 710038, China)

**Abstract:** Aimed at the problems that the traditional linear Support Vector Machine (SVM) is equal to each dimension of input features in training data sets, and the classification in the original space directly leads to the low prediction accuracy, a method for combining low-degree polynomial mappings and feature weighted is proposed to improve the classification performance of linear SVM. First of all, each sample is mapped into the two-degree explicit feature space by using the polynomial trick to increase the implicit information of samples. Then, the feature weights of each dimension are calculated by using the fuzzy entropy feature weighted algorithm. By feature weighting, the magnitude of contribution to the result of classification can be measured. To verify the robustness of the proposed method, the totally seven data sets from different databases are tested. Making a comparison between Kernel SVM and other improved Linear SVM algorithms in training time and prediction accuracy, the results show that the training time reduces an order of magnitude with the expansion of data set, and the prediction accuracy can keep up with few training samples even close to Kernel SVM. The running results show that the proposed method can effectively improve the overall performance of the linear vector machine.

**收稿日期:** 2017-07-16

**作者简介:** 王瑞(1990—),女,陕西榆林人,硕士,主要从事机器学习、数据挖掘研究。E-mail:15353724115@163.com

**引用格式:** 王瑞, 向新, 肖冰松. 低阶数据映射和特征加权的线性 SVM[J]. 空军工程大学学报(自然科学版), 2019, 20(4): 72-77. WANG Rui, XIANG Xin, XIAO Bingsong. Low-Degree Data Mappings and Feature Weighted via Linear SVM[J]. Journal of Air Force Engineering University (Natural Science Edition), 2019, 20(4): 72-77.

**Key words:** linear svm; kernel svm; low-degree polynomial mappings; hidden information; feature weighted; fuzzy entropy

支持向量机 (Support Vector Machine, SVM)<sup>[1]</sup>是解决分类、回归以及其他学习任务的主流机器学习方法之一。用于分类的有线性 SVM<sup>[2-4]</sup>和非线性 SVM。非线性 SVM<sup>[5-6]</sup>是利用核技术将原始数据映射到高维特征空间,从而解决非线性分类问题。但是,在训练大规模数据集时,内存和时间开销都很大。线性 SVM 直接在原始空间完成训练,训练时间短且容易扩展。但预测准确率往往次于非线性 SVM。最近研究显示对于一些数据集(如文本数据),二者的分类性能相似,其根本原因是文本数据的维度相对较高。SVM 分类器的核心是求解一个二次规划(Quadratic Programming, QP)问题,使得目标函数  $f(\mathbf{x}_i)$  最小:

$$f(\mathbf{x}_i) = \min_{\omega} \frac{1}{2} \omega^T \omega + C \sum_{i=1}^l \max(1 - y_i, \omega^T \phi(\mathbf{x}_i), 0) \quad (1)$$

式中:  $\omega$  为分类超平面;  $C$  为惩罚因子;  $\phi(\mathbf{x}_i)$  为将  $\mathbf{x}_i$  ( $i=1, 2, \dots, l$ ) 映射到高维空间的核函数;  $l$  为训练样本集大小。对于两分类问题, 标签  $y_i \in \{+1, -1\}$ 。若  $\phi(\mathbf{x}) = \mathbf{x}$ , 则求解的是线性 SVM。

针对线性 SVM 预测准确率低的问题, 目前, 国内外主要有 3 种解决方法: ①分解法<sup>[7-8]</sup>: 文献[7]将特征空间分成几个区域, 在每个区域训练一个线性分类器, 多个线性分类器整合后的结果即为最终分类结果。该方法在划分区域时, 根据数据集的不同特征, 选择使用传统  $k$ -means 或者球面  $k$ -means。文献[8]采用分而治之的方式, CSVM(Clustered Support Vector Machine, CSVM)使用  $k$ -means 将数据划分成多簇, 每一簇训练一个局部线性 SVM 分类器。与文献[7]相比, 该方法为每个局部线性 SVM 分配一个全局归一化权重向量, 目的是平衡不同簇间的信息, 避免过拟合。②隐性信息学习方法<sup>[9-12]</sup>: 文献[9]提出隐性信息和特征相结合的方法提高分类器的学习能力, 隐性信息可作为第 2 特征或第 2 标签, 该方法主要用于图像处理。③特征加权法<sup>[13-15]</sup>: 文献[13]提出多特征加权算法。该算法使用梯度下降法为每个样本训练一个平衡向量, 使得样本最终的分值尽可能大, 即样本到超平面的距离达到最大。为了进一步提高分类性能, 同时为每一类样本训练一个权重向量, 目的是使得目标函数达到最小。文献[14]提出紧性分离系数 (Compactness and Separation Coefficient, CSC) 算法, 该

方法分别求解每一维特征的类内、类间距离均值, 用每一维的类间距离均值与类内距离均值作商, 得到样本的特征权重, 目的是拉大类间距离、缩小类内距离, 但是算法的训练时间较长。

以上线性 SVM 改进算法, 仅仅从求解线性 SVM 问题的本身出发, 并没有融入核 SVM 在预测准确率方面具有优势的思想。另外, 传统 SVM 每个特征分配相同的权重, 即没有区分不同特征对分类效果的贡献大小。其次, 在追求提高预测准确率的同时忽视训练/预测的时间开销<sup>[16]</sup>。为此, 提出低阶多项式数据映射和特征加权相结合的线性 SVM 算法。新方法的创新点在于: 采用低阶多项式数据映射实现融合核 SVM 的思想, 使用模糊熵 (Fuzzy Entropy, FE) 特征加权算法计算每一维特征的权重, 通过权重大小衡量特征对分类效果的贡献大小, 从而提高整体的分类性能。

## 1 低阶多项式映射

目前, 在实际应用中, 使用 RBF 核完成数据映射比多项式核更广泛。也许是由于在相同的训练/预测开销下, 多项式核的预测准确率次于 RBF 核。但是, 文献[17]通过对一些数据集进行测试发现, 使用低阶多项式核映射的预测准确率略低于 RBF 核。另外, 线性 SVM 具有训练/预测速度较快的优势。因此, 可以考虑采用多项式核对应的显式特征映射, 完成数据集从原始空间到低阶空间的数据映射, 再使用线性 SVM 完成训练/预测。这样既利用了线性 SVM 的优点, 又保留了一定程度的非线性性。多项式核的表达式为:  $K(\mathbf{x}_i^T, \mathbf{x}_j) = (\gamma \mathbf{x}_i^T \mathbf{x}_j + r)^d$ ,  $\gamma > 0$ 。其中,  $\gamma$  和  $r$  为核参数,  $d$  为阶数。多项式核是 2 个向量  $\mu(\mathbf{x}_i)$  和  $\mu(\mathbf{x}_j)$  的点积。假设阶数  $d=2$ 。样本  $\mathbf{x}$  到映射空间的函数  $\mu(\mathbf{x})$  可以表示为:

$$\mu(\mathbf{x}) = [r, \sqrt{2}\gamma r x_1, \dots, \sqrt{2}\gamma r x_1^2, \gamma x_2^2, \dots, \gamma x_n^2, \sqrt{2}\gamma x_1 x_2, \sqrt{2}\gamma x_2 x_3, \dots, \sqrt{2}\gamma x_{n-1} x_n]^T \quad (2)$$

式(2)中的系数  $\sqrt{2}$  是为了使表达式  $\mu(\mathbf{x}_i)^T \mu(\mathbf{x}_j)$  的形式更简单, 并没有使用核函数。如果移除  $\sqrt{2}$ , 令  $r=1, \gamma=1$ , 可得到一个更灵活、简洁的映射向量:

$$\mu'(\mathbf{x}) = [1, x_1, \dots, x_n, x_1^2, x_2^2, \dots, x_n^2, x_1 x_2, x_2 x_3, \dots, x_{n-1} x_n]^T \quad (3)$$

使用低阶多项式映射, 线性 SVM 的优化问题将转化为:

$$f(\mathbf{x}_i) = \min_{\omega} \frac{1}{2} \omega^T \omega + C \sum_{i=1}^l \max(1 - y_i \omega^T \mu(\mathbf{x}_i), 0) \quad (4)$$

预测时,对于任意样本  $\mathbf{x}_i$ ,需计算  $\omega^T \mu(\mathbf{x}_i)$ 。其时间复杂度为  $O(\hat{n})$ ,  $\hat{n}$  为  $\mu(\mathbf{x}_i)$  非零元的个数。

从训练时间和预测准确率 2 个方面比较线性、2-阶多项式和 RBF 核 3 种映射方法的性能。线性 SVM 使用 liblinear<sup>[18-19]</sup>, RBF 核使用 libsvm<sup>[20-21]</sup>。对于每个数据集,libsvm 经过 5 次交叉验证选择最好的参数  $C$  和  $\gamma$ 。2-阶多项式映射在 liblinear 的基础上进行改进,参数  $\gamma$ 、 $r$  设为 1。从内存占用的角度分析,线性 SVM 需要存储分类超平面的权重  $\omega$ ,

核 SVM 需要存储核矩阵  $\mathbf{Q}$ 。通常,后者的内存占用较大。表 1 中的数据集来自 libsvm tools<sup>[22]</sup>,realsim 的特征中含有文档数据,线性 SVM 和核 SVM 的预测准确率相近。对于其他数据集,线性 SVM 的性能明显次于 RBF SVM,但是 2-阶多项式映射在这些数据集上却表现出很好的效果,而且训练时间明显缩短。

从表 1 中不难发现,线性 SVM 的训练时间最短,RBF SVM 的训练时间最长,2-阶多项式映射的训练时间介于二者之间,但比 RBF SVM 明显缩短。2-阶多项式映射保留了一定的非线性性,所以预测准确率优于线性 SVM。另外,该算法不仅适用于两分类问题,对多分类问题也同样适用。

表 1 训练时间和预测准确率的性能比较

数据集名称	类别数	样本数 $l$	特征数 $n$	时间/ms			准确率/%		
				liblinear	libsvm	2-阶映射	liblinear	libsvm	2-阶映射
letter. scale	26	15 000	16	4 100	25 200	8 300	66.36	82.24	83.14
iris. scale	3	150	4	15	16	16	88.66	97.33	98.00
glass. scale	6	214	9	16	78	46	66.35	60.28	77.57
splice	2	1 000	60	546	968	648	82.52	90.20	89.24
german. numer	2	1 000	24	421	812	537	40.10	93.50	76.70
cod-rna	2	59 535	8	25 491	773 125	36 317	73.00	84.32	81.40
real-sim	2	72 309	20 958	179 400	3 661 500	206 700	98.49	98.10	98.51

## 2 模糊熵特征加权算法

无论使用线性 SVM,还是非线性核 SVM 完成训练,样本的每一维特征对分类效果的贡献大小都不一定相同<sup>[13-14]</sup>。另外,某些特征可能本来就是噪声,若按照传统 SVM 给所有输入特征赋予相同的权重(即,每个特征均等对待),然后直接完成训练,则必然影响最终分类效果。因此,采用模糊熵特征加权算法给对分类效果贡献较大的重要特征分配较高权重,同时减少对分类效果贡献不大的特征权重。

在信息论中,熵是随机变量不确定性程度的度量方法。可以反映一个事件所包含的信息量。对于训练样本数据集,如果样本特征的不确定性越高,那么,模糊熵越大。这样的特征对分类效果贡献较小,应分配较小的权重。反之,应分配较大的权重。

### 2.1 算法原理

假设所有样本可分为  $M$  类,  $\mathbf{X} = \{\bar{x}^1, \bar{x}^2, \dots, \bar{x}^m\}$ , 向量  $\mathbf{v}_m$  是质心(同一类的均值向量),表达式为:

$$\mathbf{v}_m = \frac{1}{a_m} \sum_{j=1}^{a_m} \bar{x}_j^m, m=1, 2, \dots, M \quad (5)$$

样本  $\mathbf{x}_i$  与均值向量  $\mathbf{v}_m$  之间的相似性,用  $\phi_m(i, \cdot)$  表示,某一维特征用  $k$  表示,计算公式如下:

$$\phi_m(i, k) = \left| \frac{1}{M-1} \sum_{\substack{j=1 \\ j \neq m}}^M |x_i^j(k) - v_j^2(k)| - |x_i^2(k) - v_m^2(k)| \right|^{1/2} \quad (6)$$

若  $l$  为训练样本集大小;  $d$  为特征维度。可得相似矩阵  $\Phi$ 。  $\Phi = [\Phi_1, \Phi_2, \dots, \Phi_l]^T$ , 第  $i$  个样本的相似矩阵  $\phi_i$  的表达式为:

$$\phi_i = \begin{bmatrix} \phi_1(i, 1) & \phi_1(i, 2) & \dots & \phi_1(i, d) \\ \phi_2(i, 1) & \phi_2(i, 2) & \dots & \phi_2(i, d) \\ \vdots & \vdots & \dots & \vdots \\ \phi_M(i, 1) & \phi_M(i, 2) & \dots & \phi_M(i, d) \end{bmatrix}_{M \times d} \quad (7)$$

使用相似矩阵  $\Phi$ , 计算每一维特征的模糊熵:

$$H(k) = - \sum_{l=1}^{N \times M} (\phi(l, k) \log(\phi(l, k)) + (1 - \phi(l, k)) \log(1 - \phi(l, k))) \quad (8)$$

$d$  维特征的模糊熵可用  $\mathbf{H} = [H(1), H(2), \dots, H(d)]$  表示。因此,第  $k$  个特征的权重:

$$\omega_{\text{feature}}(k) = \frac{1/H(k)}{\max(1/H)}, k=1, 2, \dots, d \quad (9)$$

式中:  $1/H = [1/H(1), 1/H(2), \dots, 1/H(d)]$ ;  $\max(1/H)$  为  $1/H$  中的最大元素。加权后线性 SVM 的优化问题将转化为:

$$f(\mathbf{x}_i) = \min_{\omega} \frac{1}{2} \omega^T \omega +$$

$$C \sum_{i=1}^l \max(1 - y_i \omega^T(x_j \cdot \omega_{\text{feature}}), 0) \quad (10)$$

预测时,对于任意样本  $x_i$ ,需计算  $\omega^T(x \cdot \omega_{\text{feature}})$ 。其时间复杂度为  $O(\bar{n})$ , $\bar{n}$  为样本  $x_i$  非零元的个数。

### 2.2 线性、模糊熵特征加权和 CSC 特征加权 SVM 的性能比较

关于特征加权思想对最终分类性能的影响,这部分做了两方面工作。其一,通过给特征添加不同权重,观察预测准确率的变化,从而从实验的角度证明特征加权对分类效果产生影响。其二,从训练时间和预测准确率 2 个角度,将模糊熵特征加权算法与 CSC 特征加权算法进行比较,证明采用前者实现特征加权的优越性。

实验验证特征加权对分类效果的影响。首先,将特征权重按由大到小的顺序进行排序,排在前面的特征权重较大,对分类效果的贡献也越大。因此,考虑给排在前 40% 的特征权重再乘以一个动态变化因子,这样可以更加突出这些特征的重要性。该动态变化因子也是按照由大到小的顺序依次分配给已排序的特征权重,并完成相乘运算。动态变化的目的是使得重要的特征分配较大的权重。

实验结果如图 1 所示,训练/预测的数据集分别是 dna. scale/dna. scale.<sup>[22]</sup>(类别数为 3,样本总数

为(2 000/1 186),特征数为 180),图 1 中的“初始值”是分配给特征权重排在前 X % 的最后一个特征权重的因子。图 1(a)是在相同特征总数比例因子的条件下,改变权重初始值和动态步长的预测准确率示意图。算法思路是改变一次权重初始值,动态步长也相应重新设置为它的初始值(数据集不同初始值也不同,这里取 0.2),然后按照当前步长是上一个特征权重步长的 2 倍,实现动态增长。由图可知,数据集的预测准确率呈现出先上升后下降的趋势;图 1(b)是在选取图 1(a)中最优权重初始值的条件下,改变权重初始值和动态步长的预测准确率示意图。算法思路是改变一次权重初始值,动态步长也相应重新设置为它的初始值(数据集不同初始值也不同,这里取 0.2),然后按照当前步长是上一个特征权重步长的 2 倍,实现动态增长。由图可知,数据集的预测准确率呈现出先上升后下降的趋势;图 1(b)是在选取图 1(a)中最优权重初始值的条件下,改变加权特征数比例因子,发现预测准确率的趋势与图 1(a) 相同,而且在比例因子为 40% 时达到峰值。因此,特征加权会影响最终的分类性能。

从训练时间、预测准确率 2 个方面将线性条件下的模糊熵特征加权算法与线性 SVM 和 CSC 特征加权算法进行比较,实验结果见表 2。

表 2 训练时间和预测准确率的性能比较

数据集名称	类别数	样本数 $l$	特征数 $n$	时间/ms			准确率/%		
				liblinear	CSC 特征加权	模糊熵 特征加权	liblinear	CSC 特征加权	模糊熵 特征加权
letter. scale	26	15 000	16	4 100	272 300	8 300	66.36	66.56	66.98
iris. scale	3	150	4	15	125	16	88.66	90.00	89.33
glass. scale	6	214	9	16	63	16	66.35	68.23	66.35
splice	2	1 000	60	546	4 633	652	82.52	84.23	83.90
german. numer	2	1 000	24	421	2 060	563	40.10	73.60	75.00
cod-rna	2	59 535	8	25 491	1 994 270	35 273	73.00	85.74	86.88
real-sim	2	72 309	20 958	179 400	3 661 500	185 300	98.49	99.01	99.12

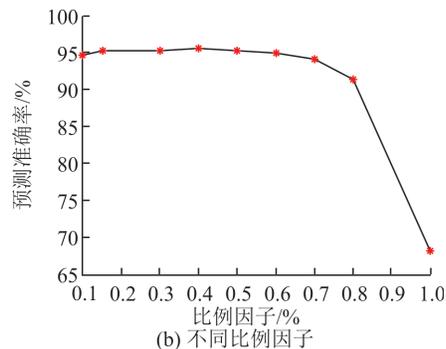
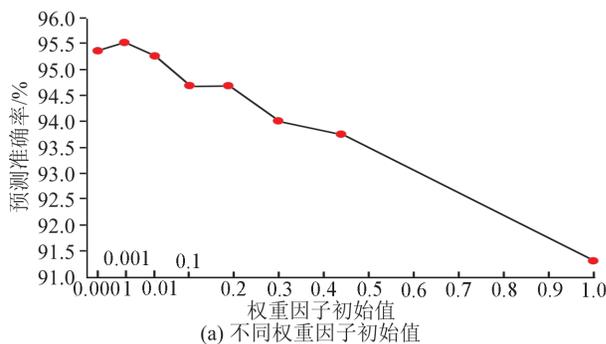


图 1 不同特征权重对分类效果的影响

由表 2 实验数据可知,CSC 特征加权算法在—些数据集上的预测准确率优于模糊熵特征加权算

法,但训练时间较长,时间开销主要用于计算每一维特征的类内、类间距离均值。所以,从时间开销和预

测准确率 2 个角度综合考虑,模糊熵特征加权算法更具有优势。

### 3 低阶特征映射和模糊熵特征加权融合算法

低阶特征映射能够保留一定程度的非线性性,模糊熵特征加权算法可以区分不同特征对分类效果的贡献大小。该部分尝试将二者融合,实现提高线性 SVM 的预测准确率。融合后,线性 SVM 的优化

问题将转化为:

$$f(\mathbf{x}_i) = \min_{\omega} \frac{1}{2} \omega^T + C \sum_{i=1}^l \max$$

$$(1 - y_i \omega^T (\mu(\mathbf{x}_i) \cdot \omega_{\text{feature}}), 0) \quad (11)$$

融合算法的处理过程都是在使用线性分类器训练前完成,因此并没有引入新的正则化问题。将融合算法与 libsvm、基于  $k$ -means 聚簇的多线性 SVM 进行性能比较,结果如表 3 所示。融合算法通过参数调优选择最佳的参数  $r$ 。

表 3 训练时间和预测准确率的性能比较

数据集名称	类别数	样本数 $l$	特征数 $n$	时间/ms			准确率/%		
				CSVM	libsvm	融合算法	CSVM	libsvm	融合算法
letter. scale	26	15 000	16	24 876	25 200	16 530	82.98	82.24	83.94
iris. scale	3	150	4	16	16	15	98.00	97.33	98.01
glass. scale	6	214	9	63	78	52	65.32	60.28	78.12
splice	2	1 000	60	848	968	750	89.33	90.20	90.20
german. numer	2	1 000	24	936	812	626	72.66	93.50	70.00
cod-rna	2	59 535	8	152 309	773 125	39 343	82.78	84.32	83.35
real-sim	2	72 309	20 958	962 875	3 661 500	225 373	98.02	98.10	98.62

表 3 实验数据表明融合算法有利于提高线性 SVM 的分类性能,而且在一些数据集上的预测准确率接近于 libsvm。CSVM 同样提高了分类性能,但时间开销明显高于融合算法。因为  $k$ -means 算法的时间复杂度是  $O(nkt)$ ,其中  $n$  为样本总数, $k$  为聚类簇数, $t$  为迭代次数。融合算法的低阶特征映射的时间复杂度是线性的,模糊熵特征加权算法仅需要近似于  $k$ -means 算法的 1 次迭代。因此,可以有效减少时间开销。

文献[14]通过大量实验测试证明 2-阶多项式映射的参数  $\gamma$  对分类效果影响不大。因此,为了便于计算,取  $\gamma=1$ 。但是,参数  $r$  对分类性能敏感。实验结果见图 2。

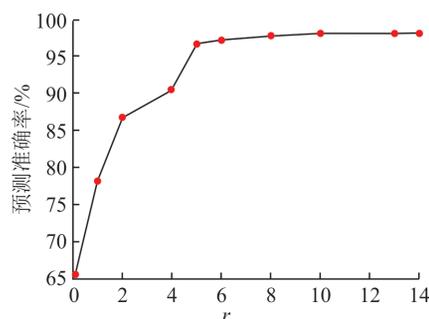


图 2 参数  $r$  对分类性能的影响

### 4 结语

线性 SVM 具有训练/预测速度快的优势,但预测准确率较低。因此,笔者借鉴核 SVM 的思想,引

入多项式核对应的低阶特征映射,该方法既保留了一定程度的非线性性,又增加了特征的一些隐性信息。另外,传统 SVM 对所有输入特征都均等对待(即分配相同的权重),并没有考虑不同特征对分类效果的贡献大小。因此,提出使用模糊熵特征加权算法区分不同特征的重要性,进而改善最终的性能。最后,融合 2 种算法思想,进一步提高整体预测准确率。而且所提算法对两分类和多分类问题均适用,实验结果也验证了这一结论。

### 参考文献(References):

- [1] 崔萌,张春雷. LIBSVM, LIBLINEAR, SVM 比较研究[J]. 数码世界, 2016, 1(7): 85-87.  
CUI M, ZHANG C L. LIBSVM, LIBLINEAR, SVM Comparative Study[J]. Digital Space, 2016, 1(7): 85-87. (in Chinese)
- [2] CHIANG W L, LEE M C, LIN C J. Parallel Dual Coordinate Descent Method for Large-Scale Linear Classification in Multi-Core Environments[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016: 1485-1494.
- [3] TANG J J, TIAN Y J, LIU D L. Connected Bit Min-wise Hashing for Large-Scale Linear SVM[C]//International Conference on Fuzzy Systems and Knowledge Discovery. Zhangjiajie, China: IEEE, 2016: 995-1002.
- [4] TORR P H S. Locally Linear Support Vector Machines[C]//International Conference on International

- Conference on Machine Learning. New York: Omni-Press, 2011:985-992.
- [5] NDEZ-DELGADO M, CERNADAS E, BARRO S, et al. Do We Need Hundreds of Classifiers to Solve Real World Classification Problems? [J]. *Journal of Machine Learning Research*, 2014, 15(1):3133-3181.
- [6] 王玉冰,程嗣怡,周一鹏,等. 参数寻优 LSSVM 算法在机载火控雷达工作模式判定中的应用[J]. *空军工程大学学报(自然科学版)*, 2017, 18(3):49-53.
- WANG Y B, CHENG S Y, ZHOU Y P, et al. A Parameter-Optimized LSSVM Method for Operation Modes Recognition of Airborne Fire Control Radar[J]. *Journal of Air Force Engineering University (Natural Science Edition)*, 2017, 18(3):49-53. (in Chinese)
- [7] HUANG H Y, LIN C J. Linear and Kernel Classification: When to Use Which? [J]. *SIAM International Conference on Data Mining*, 2016, 16(3):302-311.
- [8] GU Q Q, HAN J W. Clustered Support Vector Machines [J]. *Journal of Machine Learning Research*, 2013, 31(16):307-315.
- [9] WANG Z H, JI Q. Classifier Learning with Hidden Information[C]//*IEEE Conference on Computer Vision and Pattern Recognition*. New York: IEEE, 2015: 4969-4977.
- [10] LI W, DAI D X, TAN M K, et al. Fast Algorithms for Linear and Kernel SVM+[C]//*IEEE Conference on Computer Vision and Pattern Recognition*. New York: IEEE, 2016:225-2266.
- [11] FEYEREISL J, KWAK S, SON J, et al. Object Localization Based on Structural SVM Using Privileged Information[J]. *Advances in Neural Information Processing Systems*, 2015, 1(2):208-216.
- [12] 孙广玲,董勇,刘志. 伪特权信息和 SVM+[J]. *西安电子科技大学学报(自然科学版)*, 2016, 43(6):103-108.
- SUN G L, DONG Y, LIU Z. Pseudo-Privileged Information and SVM+[J]. *Journal of Xidian University (Nature Science)*, 2016, 43(6):103-108. (in Chinese)
- [13] YAN C, BAI X, REN P, et al. Band Weighting via Maximizing Interclass Distance for Hyperspectral Image Classification [J]. *IEEE Geoscience & Remote Sensing Letters*, 2016, 13(7):922-925.
- [14] QI B, ZHAO C H, YIN G S. Feature Weighting Algorithms for Classification of Hyperspectral Images Using A Support Vector Machine [J]. *Applied Optics*, 2014, 53(13):2839-2846.
- [15] QI B, ZHAO C, YOUN E, et al. Use of Weighting Algorithms to Improve Traditional Support Vector Machine Based Classifications of Reflectance Data[J]. *Optics Express*, 2011, 19(27):26816-26826.
- [16] MAO X, WU O, HU W M, et al. Nonlinear Classification via Linear SVMs and Multi-Task Learning[J]. *ACM International Conference on Conference on Information & Knowledge Management*, 2014, 14(11): 1955-1958.
- [17] CHANG Y W, HSIEH C J, CHANG K W, et al. Training and Testing Low-Degree Polynomial Data Mappings via Linear SVM [J]. *Journal of Machine Learning Research*, 2010, 11(11):1471-1490.
- [18] FAN R E, CHANG K W, HSIEH C J, et al. LIBLINEAR: A Library for Large Linear Classification [J]. *Journal of Machine Learning Research*, 2008, 9(9):1871-1874.
- [19] HELLEPUTTE T, GRAMME P. Liblinear: Linear Predictive Models based on the LIBLINEAR C/C++ Library [J]. *Journal of Machine Learning Research*, 2013, 33(9):372-387.
- [20] CHANG C C, LIN C J. LIBSVM: A Library for Support Vector Machines [J]. *ACM Transactions on Intelligent Systems & Technology*, 2011, 2(3): 389-396.
- [21] AJAY A, DIXON K D M, SOWMYA V, et al. Aerial Image Classification using GURLS and LIBSVM[C]//*International Conference on Communication and Signal Processing*. India: IEEE, 2016:396-401.
- [22] LIN C J. LIBSVM Data Sets [EB/OL]. [2017-02-12]. <http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/DOI>.

(编辑:徐楠楠)