

基于特征值分解的中心支持向量机算法

陈素根^{①②} 吴小俊^{*①}

^①(江南大学物联网工程学院 无锡 214122)

^②(安庆师范学院数学与计算科学学院 安庆 246133)

摘要: 针对广义特征值中心支持向量机(GEPSVM)训练和决策过程不一致问题, 该文提出一类改进的基于特征值分解的中心支持向量机, 简称为 IGEPSVM。首先针对二分类问题提出了基于特征值分解的中心支持向量机, 然后基于“一类对余类”策略将其推广到多类分类问题。将 GEPSVM 求解广义特征值问题转化为求解标准特征值问题, 降低了计算复杂度。引入了一个新的参数, 可以调节模型的性能, 提高了 GEPSVM 的分类精度。提出了基于 IGEPSVM 的多类分类算法。实验结果表明, 与 GEPSVM 算法相比较, IGEPSVM 不仅提高了分类精度, 而且缩短了训练时间。

关键词: 支持向量机; 广义特征值中心支持向量机; 两类分类; 多类分类; 特征值分解

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2016)03-0557-08

DOI: 10.11999/JEIT150693

Eigenvalue Proximal Support Vector Machine Algorithm Based on Eigenvalue Decomposition

CHEN Sugeng^{①②} WU Xiaojun^①

^①(School of IoT Engineering, Jiangnan University, Wuxi 214122, China)

^②(School of Mathematics & Computational Science, Anqing Normal University, Anqing 246133, China)

Abstract: To deal with the consistency problem of training process and decision process in Generalized Eigenvalue Proximal Support Vector Machine (GEPSVM), an improved version of eigenvalue proximal support vector machine, called IGEPSVM for short is proposed. At first, IGEPSVM for binary classification problem is proposed, and then Multi-IGEPSVM is also presented for multi-class classification problem based on “one-versus-rest” strategy. The main contributions of this paper are as follows. The generalized eigenvalue decomposition problems are replaced by the standard eigenvalue decomposition problems, leading to simpler optimization problems. An extra parameter is introduced, which can adjust the performance of the model and improve the classification accuracy of GEPSVM. A corresponding multi-class classification algorithm is proposed, which is not studied in GEPSVM. Experimental results on several datasets illustrate that IGEPSVM is superior to GEPSVM in both classification accuracy and training speed.

Key words: Support Vector Machine (SVM); Generalized Eigenvalue Proximal SVM (GEPSVM); Binary classification; Multi-class classification; Eigenvalue decomposition

1 引言

支持向量机(SVM)算法是经典的分类算法^[1], 鉴于其坚实的理论基础和良好的泛化性能而被广泛使用于各个领域^[2,3]。SVM 在解决小样本、非线性和

高维模式问题中表现出了许多特有的优势, 标准的 SVM 算法问题可归结为求解一个受约束的二次规划(QP)问题, 对于小规模QP问题, 它表现出了非常优秀的学习能力, 但当训练集样本规模巨大时, 就会出现训练速度慢、算法复杂和效率低下等问题。为了解决这些问题, 一方面, 许多学者对 SVM 求解算法进行了广泛研究, 并取得了很多优秀成果, 如 Chunking 算法^[4]、分解算法^[5]、SMO 算法^[6]和 FD-SVM^[7]等; 另一方面, 新型的 SVM 算法被大量研究^[8-16], 其中非平行平面支持向量机算法是代表性成果之一。关于非平行平面支持向量机算法的研究

收稿日期: 2015-06-08; 改回日期: 2015-09-17; 网络出版: 2015-11-19

*通信作者: 吴小俊 wu_xiaojun@jiangnan.edu.cn

基金项目: 国家自然科学基金(61373055, 61103128), 111 引智计划项目(B12018), 江苏省工业支持计划项目(BE2012031)

Foundation Items: The National Natural Science Foundation of China (61373055, 61103128), 111 Project of Chinese Ministry of Education (B12018), Industrial Support Program of Jiangsu Province (BE2012031)

开始于2006年,文献[8]提出了广义特征值中心支持向量机算法(GEPSVM)来处理两类分类问题,考虑了最大化类间距离和最小化类内距离,对每一类样本训练一个分类超平面,开辟了新的决策思路。2007年,文献[9]在 GEPSSVM 算法的基础上提出了孪生支持向量机算法(TWSVM),与 GEPSSVM 求解两个规模较小的广义特征值问题有所不同, TWSVM 求解两个规模较小的 QP 问题,训练速度相当于传统 SVM 的 1/4。然而,许多实际问题都可归结为多类分类问题,传统 SVM 的多类分类算法大致分为“分解重构法”和“整体法”两大类,代表性研究成果有:“一类对余类”^[17]、“一类对一类”^[18]和 C&S (Crammer and Singer)方法^[19]等。近年来,非平行平面支持向量机多类分类算法也开始受到很多学者的关注,涌现了一些优秀成果^[20,21]。

GEPSVM 算法是最早被提出的非平行平面支持向量机算法,虽然它开辟了新的决策思路,但它依然有一些不足之处。首先,GEPSVM 在训练过程和决策过程中出现了不一致性,即:在训练过程中,是通过比较每一个超平面和两类训练样本之间的距离来寻求每一类的最优超平面的;在决策过程中,是通过比较测试样本和两个超平面的距离来实现分类的。因此,这一种不一致性,在一定程度上可能会影响 GEPSVM 的性能。其次,GEPSVM 算法是通过求解两个广义特征值问题来实现的,但求解广义特征值问题的算法复杂度是 $O(m^3)$,从而可能会影响模型的训练速度。最后,GEPSVM 算法本身是针对两类分类问题提出的,不能直接处理多类分类问题。针对以上问题,本文首先针对两类分类问题提出一类改进的基于特征分解的中心支持向量机算法(Improved GEPSVM, IGEPSVM);然后在 IGEPSVM 的基础上进一步研究多类分类问题,提出多类分类 IGEPSVM 算法。与 GEPSVM 算法相比较,IGEPSVM 有如下优点:(1)将求解广义特征值问题转化为求解标准特征值问题,降低了算法的复杂度;(2)引入一个新的参数,可以更好地调节模型的性能,提高了 GEPSVM 算法的分类精度。最后,在标准的 UCI 分类数据集上验证了本文算法的有效性。

2 广义特征值中心支持向量机(GEPSVM)

对于两类分类问题,给定训练样本集:

$$T = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_m, y_m)\} \in (\mathbf{X} \times \mathbf{Y})^m \quad (1)$$

其中 $\mathbf{x}_i \in \mathbf{X} = R^n, y_i \in \{-1, 1\}$ 。线性广义特征值支持向量机算法^[8]的目标是寻找两个非平行的超平面:

$$\mathbf{w}_1^T \cdot \mathbf{x} + b_1 = 0 \text{ 和 } \mathbf{w}_2^T \cdot \mathbf{x} + b_2 = 0 \quad (2)$$

其中 $\mathbf{w}_i \in R^n, b_i \in R$, 每一个超平面使得该类的训练样本尽可能聚集在超平面附近,同时另一类训练样本尽可能地远离这个超平面。记 $\mathbf{A} \in R^{m_1 \times n}$ 表示所有正类训练样本, $\mathbf{B} \in R^{m_2 \times n}$ 表示所有负类训练样本, $\mathbf{e}_1, \mathbf{e}_2$ 分别表示相应维数的元素全部为 1 的列向量, $m_1 + m_2 = m$ 。因此, GEPSVM 算法满足如式(3)的优化模型:

$$\min_{(\mathbf{w}_1, b_1) \neq 0} \left\{ \frac{\|\mathbf{A}\mathbf{w}_1 + \mathbf{e}_1 b_1\|^2 + \delta \left\| \begin{bmatrix} \mathbf{w}_1 \\ b_1 \end{bmatrix} \right\|^2}{\|\mathbf{B}\mathbf{w}_1 + \mathbf{e}_2 b_1\|^2} \right\} \quad (3)$$

$$\min_{(\mathbf{w}_2, b_2) \neq 0} \left\{ \frac{\|\mathbf{B}\mathbf{w}_2 + \mathbf{e}_2 b_2\|^2 + \delta \left\| \begin{bmatrix} \mathbf{w}_2 \\ b_2 \end{bmatrix} \right\|^2}{\|\mathbf{A}\mathbf{w}_2 + \mathbf{e}_1 b_2\|^2} \right\}$$

记 $\mathbf{G} = [\mathbf{A} \ \mathbf{e}_1]^T [\mathbf{A} \ \mathbf{e}_1] + \delta \mathbf{I}, \mathbf{H} = [\mathbf{B} \ \mathbf{e}_2]^T [\mathbf{B} \ \mathbf{e}_2], \mathbf{L} = [\mathbf{B} \ \mathbf{e}_2]^T [\mathbf{B} \ \mathbf{e}_2] + \delta \mathbf{I}, \mathbf{M} = [\mathbf{A} \ \mathbf{e}_1]^T [\mathbf{A} \ \mathbf{e}_1], \mathbf{u}_1 = [\mathbf{w}_1^T \ b_1]^T, \mathbf{u}_2 = [\mathbf{w}_2^T \ b_2]^T$, 于是,模型式(3)可以转化为

$$\min_{\mathbf{u}_1 \neq 0} \frac{\mathbf{u}_1^T \mathbf{G} \mathbf{u}_1}{\mathbf{u}_1^T \mathbf{H} \mathbf{u}_1}, \min_{\mathbf{u}_2 \neq 0} \frac{\mathbf{u}_2^T \mathbf{L} \mathbf{u}_2}{\mathbf{u}_2^T \mathbf{M} \mathbf{u}_2} \quad (4)$$

显然,问题式(4)的目标函数是 Rayleigh 商。故,优化问题式(4)的最优解 $\mathbf{u}_i$ 就是广义特征值问题:

$$\mathbf{G} \mathbf{u}_1 = \lambda_1 \mathbf{H} \mathbf{u}_1, \mathbf{L} \mathbf{u}_2 = \lambda_2 \mathbf{M} \mathbf{u}_2 \quad (5)$$

的最小特征值对应的特征向量。于是,当 $\mathbf{u}_1, \mathbf{u}_2$ 被求出后,对于新的测试样本 $\mathbf{x} \in R^n$, 它将根据式(6)被分类到正类或负类。即

$$\mathbf{x} \in \text{Class } i = \arg \min_{i=1,2} \frac{|\mathbf{w}_i^T \mathbf{x} + b_i|}{\|\mathbf{w}_i\|} \quad (6)$$

其中, $|\cdot|$ 代表绝对值, $\|\cdot\|$ 代表 2-范数。对于非线性 GEPSVM 及 GEPSVM 的详细描述内容可参见文献[8]。

3 基于特征值分解的中心支持向量机(IGEPSVM)

仔细分析 GEPSVM 算法,我们不难发现, GEPSVM 算法在训练过程和决策过程具有一定的不一致性,这可能会影响 GEPSVM 算法的性能。为此,我们提出一种改进的基于特征值分解的中心支持向量机算法。

3.1 两类分类 IGEPSVM

首先考虑线性情形,对于训练样本集式(1),我们考察优化模型^[13]:

$$\min_{(\mathbf{w}_1, b_1) \neq 0, (\mathbf{w}_2, b_2) \neq 0} \left\{ \frac{\|\mathbf{A}\mathbf{w}_1 + \mathbf{e}_1 b_1\|^2}{\|\mathbf{u}_1\|^2} - \nu \frac{\|\mathbf{A}\mathbf{w}_2 + \mathbf{e}_1 b_2\|^2}{\|\mathbf{u}_2\|^2} \right. \\ \left. + \left(\frac{\|\mathbf{B}\mathbf{w}_2 + \mathbf{e}_2 b_2\|^2}{\|\mathbf{u}_2\|^2} - \nu \frac{\|\mathbf{B}\mathbf{w}_1 + \mathbf{e}_2 b_1\|^2}{\|\mathbf{u}_1\|^2} \right) \right\} \quad (7)$$

显然,直接观察式(7)是比较难以理解的。然而,式(7)有很好的几何解释。假设 IGESVM 寻求的两个非平行超平面也为式(2),那么在 $(\mathbf{w}, b)$ 空间中训练样本点 $(\mathbf{x}, 1)^T$ 到超平面 $\mathbf{w}^T \cdot \mathbf{x} + b = 0$ 的距离为

$$d(\mathbf{x}) = \frac{\|\mathbf{x}^T \cdot \mathbf{w} + 1 \cdot b\|}{\|(\mathbf{w}^T, b)\|} \quad (8)$$

这样,结合式(8),我们分别定义正类和负类训练样本点到正类和负类两个超平面的4个距离:

$$\begin{aligned} d_{PP}^2 &= \frac{\|\mathbf{A}\mathbf{w}_1 + \mathbf{e}_1 b_1\|^2}{\|(\mathbf{w}_1^T, b_1)\|^2} = \frac{\|\mathbf{A}\mathbf{w}_1 + \mathbf{e}_1 b_1\|^2}{\|\mathbf{u}_1\|^2} \\ d_{NP}^2 &= \frac{\|\mathbf{B}\mathbf{w}_1 + \mathbf{e}_2 b_1\|^2}{\|(\mathbf{w}_1^T, b_1)\|^2} = \frac{\|\mathbf{B}\mathbf{w}_1 + \mathbf{e}_2 b_1\|^2}{\|\mathbf{u}_1\|^2} \\ d_{PN}^2 &= \frac{\|\mathbf{A}\mathbf{w}_2 + \mathbf{e}_1 b_2\|^2}{\|(\mathbf{w}_2^T, b_2)\|^2} = \frac{\|\mathbf{A}\mathbf{w}_2 + \mathbf{e}_1 b_2\|^2}{\|\mathbf{u}_2\|^2} \\ d_{NN}^2 &= \frac{\|\mathbf{B}\mathbf{w}_2 + \mathbf{e}_2 b_2\|^2}{\|(\mathbf{w}_2^T, b_2)\|^2} = \frac{\|\mathbf{B}\mathbf{w}_2 + \mathbf{e}_2 b_2\|^2}{\|\mathbf{u}_2\|^2} \end{aligned}$$

于是,优化模型式(7)转化为

$$\min_{\mathbf{u}_1 \neq 0, \mathbf{u}_2 \neq 0} (d_{PP}^2 - \mathbf{v} \cdot d_{PN}^2) + (d_{NN}^2 - \mathbf{v} \cdot d_{NP}^2) \quad (9)$$

优化模型式(9)中,最小化 $(d_{PP}^2 - \mathbf{v} \cdot d_{PN}^2)$ 意味着最小化正类训练样本点到正类超平面的距离的同时最大化正类训练样本点到负类超平面的距离;类似地,最小化 $(d_{NN}^2 - \mathbf{v} \cdot d_{NP}^2)$ 意味着最小化负类训练样本点到负类超平面的距离的同时最大化负类训练样本点到正类超平面的距离。这样寻找超平面,就可以保证训练过程与决策过程的一致性。将4个距离函数代入式(9),并化简可得如式(10)模型:

$$\begin{aligned} \min_{\|\mathbf{u}_1\|=1, \|\mathbf{u}_2\|=1} & (d_{PP}^2 - \mathbf{v} \cdot d_{PN}^2) + (d_{NN}^2 - \mathbf{v} \cdot d_{NP}^2) \\ &= \min_{\|\mathbf{u}_1\|=1, \|\mathbf{u}_2\|=1} (\mathbf{u}_1^T \mathbf{Q}_1 \mathbf{u}_1 + \mathbf{u}_2^T \mathbf{Q}_2 \mathbf{u}_2) \end{aligned} \quad (10)$$

其中, $\mathbf{Q}_1 = [\mathbf{A} \ \mathbf{e}_1]^T [\mathbf{A} \ \mathbf{e}_1] - \mathbf{v} \cdot [\mathbf{B} \ \mathbf{e}_2]^T [\mathbf{B} \ \mathbf{e}_2]$, $\mathbf{Q}_2 = [\mathbf{B} \ \mathbf{e}_2]^T [\mathbf{B} \ \mathbf{e}_2] - \mathbf{v} \cdot [\mathbf{A} \ \mathbf{e}_1]^T [\mathbf{A} \ \mathbf{e}_1]$, 且 $\mathbf{v} > 0$ 是新引入的参数。

为了克服模型式(10)中 $\mathbf{Q}_1, \mathbf{Q}_2$ 有可能会奇异的,进一步引入 Tikhonov 正则化参数 δ , 将模型式(10)转化为等价的模型式(11):

$$\min_{\|\mathbf{u}_1\|=1, \|\mathbf{u}_2\|=1} (\mathbf{u}_1^T (\mathbf{Q}_1 + \delta \cdot \mathbf{I}_1) \mathbf{u}_1 + \mathbf{u}_2^T (\mathbf{Q}_2 + \delta \cdot \mathbf{I}_2) \mathbf{u}_2) \quad (11)$$

其中 $\mathbf{Q}_1, \mathbf{Q}_2$ 均为 $(n+1) \times (n+1)$ 阶矩阵, $\mathbf{I}_1, \mathbf{I}_2$ 为相应阶的单位矩阵。同时,注意到模型式(11)中,变量 $\mathbf{u}_1, \mathbf{u}_2$ 在目标函数中是独立的,根据最优化理论,可以将优化问题式(11)转化为优化问题式(12):

$$\left. \begin{aligned} \min_{\mathbf{u}_1} & \mathbf{u}_1^T (\mathbf{Q}_1 + \delta \cdot \mathbf{I}_1) \mathbf{u}_1, \quad \text{s.t.} \quad \|\mathbf{u}_1\|^2 = 1 \\ \min_{\mathbf{u}_2} & \mathbf{u}_2^T (\mathbf{Q}_2 + \delta \cdot \mathbf{I}_2) \mathbf{u}_2, \quad \text{s.t.} \quad \|\mathbf{u}_2\|^2 = 1 \end{aligned} \right\} \quad (12)$$

通过构造 Lagrange 函数并结合 KKT 条件,求解模型式(12)可以转化为求解标准特征值问题式(13):

$$(\mathbf{Q}_1 + \delta \cdot \mathbf{I}_1) \mathbf{u}_1 = \lambda_1 \mathbf{u}_1, \quad (\mathbf{Q}_2 + \delta \cdot \mathbf{I}_2) \mathbf{u}_2 = \lambda_2 \mathbf{u}_2 \quad (13)$$

其中 $\mathbf{Q}_1 + \delta \cdot \mathbf{I}_1$ 和 $\mathbf{Q}_2 + \delta \cdot \mathbf{I}_2$ 均为 $(n+1) \times (n+1)$ 阶的矩阵。经上述分析,可得到定理1。

定理1 最优化问题式(12)的最优解 $\mathbf{u}_1 = [\mathbf{w}_1^T \ b_1]^T$ 和 $\mathbf{u}_2 = [\mathbf{w}_2^T \ b_2]^T$ 分别是矩阵 $\mathbf{Q}_1 + \delta \cdot \mathbf{I}_1$ 和 $\mathbf{Q}_2 + \delta \cdot \mathbf{I}_2$ 的最小特征值所对应的单位特征向量。

证明 对于优化问题式(12)的第1部分,其目标函数为: $f_{\text{obj}}(\mathbf{u}_1) = \mathbf{u}_1^T (\mathbf{Q}_1 + \delta \cdot \mathbf{I}_1) \mathbf{u}_1$ 。设矩阵 $\mathbf{Q}_1 + \delta \cdot \mathbf{I}_1$ 的特征值分别为 $\lambda_1, \lambda_2, \dots, \lambda_{n+1}$, 且 $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{n+1}$, 其对应的单位正交特征向量为 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{n+1}$ 。于是,优化问题式(12)的第1部分的任意可行解为其特征向量的线性组合,即

$$\mathbf{u}_1 = \sum_{i=1}^{n+1} \alpha_i \mathbf{v}_i, \quad \boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_{n+1}) \geq 0$$

因此,优化问题式(12)的第1部分可表示为

$$\left. \begin{aligned} \min_{\boldsymbol{\alpha}} & f_{\text{obj}}(\boldsymbol{\alpha}) = \sum_{i=1}^{n+1} \alpha_i^2 \lambda_i \\ \text{s.t.} & \sum_{i=1}^{n+1} \alpha_i^2 = 1 \end{aligned} \right\} \quad (14)$$

显然, $\tilde{\boldsymbol{\alpha}} = (1, 0, \dots, 0)$ 是满足 $\sum_{i=1}^{n+1} \tilde{\alpha}_i^2 = 1$ 的, 故 $\tilde{\boldsymbol{\alpha}} = (1, 0, \dots, 0)$ 是式(14)的一个可行解。下面针对式(14), 证明对任意其它的可行解 $\boldsymbol{\alpha}$ 有 $f_{\text{obj}}(\boldsymbol{\alpha}) \geq f_{\text{obj}}(\tilde{\boldsymbol{\alpha}})$ 。事实上, 如果 $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_{n+1}) \geq 0$ 是可行解, 则有

$$\begin{aligned} f_{\text{obj}}(\boldsymbol{\alpha}) &= \sum_{i=1}^{n+1} \alpha_i^2 \lambda_i \geq \sum_{i=1}^{n+1} \alpha_i^2 \lambda_1 = \lambda_1 \\ &= \sum_{i=1}^{n+1} \tilde{\alpha}_i^2 \lambda_i = f_{\text{obj}}(\tilde{\boldsymbol{\alpha}}) \end{aligned} \quad (15)$$

由 Rayleigh 定理^[22]可知, 当且仅当 λ 是矩阵 $\mathbf{Q}_1 + \delta \cdot \mathbf{I}_1$ 最小特征值时, 优化问题式(12)的第1部分的目标函数 $f_{\text{obj}}(\mathbf{u}_1)$ 取得全局极小值。此时, 最小特征值 λ_1 对应的单位特征向量 $\mathbf{u}_1 = \sum_{i=1}^{n+1} \tilde{\alpha}_i \mathbf{v}_i = \mathbf{v}_1$ 是优化问题式(12)的第1部分的最优解。同理可证, 优化问题式(12)的第2部分的最优解 $\mathbf{u}_2 = [\mathbf{w}_2^T \ b_2]^T$ 是矩阵 $\mathbf{Q}_2 + \delta \cdot \mathbf{I}_2$ 的最小特征值所对应的单位特征向量。

证毕

于是, 当 $\mathbf{u}_1, \mathbf{u}_2$ 被求出后, 对于新的测试样本 $\mathbf{x} \in R^n$, 它将根据式(16), 被分类到正类或负类。即

$$\mathbf{x} \in \text{Class } i = \arg \min_{i=1,2} |\mathbf{w}_i^T \mathbf{x} + b_i| \quad (16)$$

其中, $|\cdot|$ 代表绝对值。

其次考虑非线性情形, 记 $C^T = [A \ B]^T$, K 是选择一个适当的核函数, 寻找两个非线性超曲面式(17):

$$K(x, C^T)w_1 + b_1 = 0, \quad K(x, C^T)w_2 + b_2 = 0 \quad (17)$$

类似地, 我们也可以分别定义正类和负类训练样本点到正类和负类两个非线性超曲面的4个距离:

$$\begin{aligned} d_{PP}^2 &= \frac{\|K(A^T, C^T)w_1 + e_1b_1\|^2}{\|(w_1^T, b_1)\|^2} \\ &= \frac{\|K(A^T, C^T)w_1 + e_1b_1\|^2}{\|u_1\|^2} \\ d_{NP}^2 &= \frac{\|K(B^T, C^T)w_1 + e_2b_1\|^2}{\|(w_1^T, b_1)\|^2} \\ &= \frac{\|K(B^T, C^T)w_1 + e_2b_1\|^2}{\|u_1\|^2} \\ d_{PN}^2 &= \frac{\|K(A^T, C^T)w_2 + e_1b_2\|^2}{\|(w_2^T, b_2)\|^2} \\ &= \frac{\|K(A^T, C^T)w_2 + e_1b_2\|^2}{\|u_2\|^2} \\ d_{NN}^2 &= \frac{\|K(B^T, C^T)w_2 + e_2b_2\|^2}{\|(w_2^T, b_2)\|^2} \\ &= \frac{\|K(B^T, C^T)w_2 + e_2b_2\|^2}{\|u_2\|^2} \end{aligned}$$

于是, 建立非线性情形的优化模型为

$$\min_{u_1 \neq 0, u_2 \neq 0} (d_{PP}^2 - v \cdot d_{PN}^2) + (d_{NN}^2 - v \cdot d_{NP}^2) \quad (18)$$

类似于线性情形, 求解模型式(18)可以转化为求解模型式(19):

$$\left. \begin{aligned} \min_{u_1} u_1^T (\text{Ker } Q_1 + \delta \cdot I_1) u_1, \quad \text{s.t.} \quad \|u_1\|^2 = 1 \\ \min_{u_2} u_2^T (\text{Ker } Q_2 + \delta \cdot I_2) u_2, \quad \text{s.t.} \quad \|u_2\|^2 = 1 \end{aligned} \right\} \quad (19)$$

其中, $v > 0$ 是新引入了参数, δ 为 Tikhonov 正则化参数, 且

$$\begin{aligned} \text{Ker } Q_1 &= [K(A^T, C^T) \ e_1]^T [K(A^T, C^T) \ e_1] \\ &\quad - v \cdot [K(B^T, C^T) \ e_2]^T [K(B^T, C^T) \ e_2] \\ \text{Ker } Q_2 &= [K(B^T, C^T) \ e_2]^T [K(B^T, C^T) \ e_2] \\ &\quad - v \cdot [K(A^T, C^T) \ e_1]^T [K(A^T, C^T) \ e_1] \end{aligned}$$

通过构造 Lagrange 函数并结合 KKT 条件, 求解模型式(19)可以转化为求解标准特征值问题式(20):

$$\left. \begin{aligned} (\text{Ker } Q_1 + \delta \cdot I_1) u_1 &= \lambda_1 u_1 \\ (\text{Ker } Q_2 + \delta \cdot I_2) u_2 &= \lambda_2 u_2 \end{aligned} \right\} \quad (20)$$

类似于线性情形, 关于非线性模型, 我们可以得定理2。

定理2 最优化问题式(19)的最优解 $u_1 = [w_1^T \ b_1]^T$ 和 $u_2 = [w_2^T \ b_2]^T$ 分别是矩阵 $\text{Ker } Q_1 + \delta \cdot I_1$ 和 $\text{Ker } Q_2 + \delta \cdot I_2$ 的最小特征值所对应的单位特征向量。

定理2的证明过程类似于定理1, 简单起见, 此处省略。于是, 当 u_1, u_2 被求出后, 对于新的测试样本 $x \in R^n$, 它将根据式(21), 被分类到正类或负类。即:

$$x \in \text{Class } i = \arg \min_{i=1,2} |K(x, C^T)w_i + b_i| \quad (21)$$

其中, $|\cdot|$ 代表绝对值。

总结上述过程, 可以得到两类分类 IGEP SVM 算法如下:

算法1 两类分类 IGEP SVM 算法

步骤1 给定训练集式(1);

步骤2 通过交叉验证选择合适的参数 v 和 δ ;

步骤3 利用模型式(12)或式(19)训练, 求解特征值问题式(13)或式(20)得到最小特征值对应的单位特征向量 u_1, u_2 ;

步骤4 利用决策函数式(16)或式(21)进行分类。

3.2 多类分类 IGEP SVM

对于多类分类问题, 给定训练样本集 $T = \{A_1, \dots, A_L\}$, 其中 $A_i \in R^{m_i \times n}, i = 1, 2, \dots, L; m = m_1 + \dots + m_L$, 分别代表第 i 类训练样本集, m_i 为每一类训练样本的个数, m 为所有训练样本的总数。

为了处理多分类问题, 我们将第 i 类训练样本集 $A_i \in R^{m_i \times n}$ 视为正类, 其余样本放在一起视为负类, 记为

$B_i = T \setminus A_i = \{A_1, \dots, A_{i-1}, A_{i+1}, \dots, A_L\} \in R^{(m-m_i) \times n}$ 这样, 对于线性多类分类问题, 基于线性 IGEP SVM 就可以构造 L 个相互不平行的超平面:

$$w_i^T \cdot x + b_i = 0, \quad i = 1, 2, \dots, L \quad (22)$$

并构造决策函数如下:

$$x \in \text{Class } i = \arg \min_{i=1,2,\dots,L} |w_i^T x + b_i| \quad (23)$$

为了寻找第 i 个超平面, 只需求解优化问题如式(24):

$$\left. \begin{aligned} \min_{u_i} u_i^T (Q_i + \delta_i \cdot I_i) u_i \\ \text{s.t.} \quad \|u_i\|^2 = 1 \end{aligned} \right\} \quad (24)$$

其中 $Q_i = [A_i \ e_{i1}]^T [A_i \ e_{i1}] - v_i \cdot [B_i \ e_{i2}]^T [B_i \ e_{i2}], I_i$ 为相应阶的单位矩阵, $v_i > 0$ 是引入的参数, $\delta_i > 0$

是 Tikhonov 正则化参数。于是, 优化问题式(24)的最优解 $\mathbf{u}_i$ 就是特征值问题:

$$(\mathbf{Q}_i + \delta_i \cdot \mathbf{I}_i) \mathbf{u}_i = \lambda_i \mathbf{u}_i \quad (25)$$

的最小特征值对应的单位特征向量。

对于非线性多类分类问题, 基于非线性 IGEP SVM 可以构造 L 个超曲面:

$$K(\mathbf{x}, \mathbf{C}^T) \mathbf{w}_i + b_i = 0, \quad i = 1, 2, \dots, L \quad (26)$$

并构造决策函数:

$$\mathbf{x} \in \text{Class } i = \arg \min_{i=1,2,\dots,L} |K(\mathbf{x}, \mathbf{C}^T) \mathbf{w}_i + b_i| \quad (27)$$

为了寻找第 i 个超曲面, 只需求解优化问题:

$$\left. \begin{array}{l} \min_{\mathbf{u}_i} \mathbf{u}_i^T (\text{Ker } \mathbf{Q}_i + \delta_i \cdot \mathbf{I}_i) \mathbf{u}_i \\ \text{s.t.} \quad \|\mathbf{u}_i\|^2 = 1 \end{array} \right\} \quad (28)$$

其中 $\text{Ker } \mathbf{Q}_i = [K(\mathbf{A}_i^T, \mathbf{C}^T) \mathbf{e}_{i1}]^T [K(\mathbf{A}_i^T, \mathbf{C}^T) \mathbf{e}_{i1}] - \mathbf{v}_i \cdot [K(\mathbf{B}_i^T, \mathbf{C}^T) \mathbf{e}_{i2}]^T [K(\mathbf{B}_i^T, \mathbf{C}^T) \mathbf{e}_{i2}]$, $\mathbf{C}^T = [\mathbf{A}_1 \dots \mathbf{A}_L]^T$, $\mathbf{v}_i > 0$ 是引入的参数, $\delta_i > 0$ 是 Tikhonov 正则化参数。于是, 优化问题式(28)的最优解 $\mathbf{u}_i$ 就是特征值问题:

$$(\text{Ker } \mathbf{Q}_i + \delta_i \cdot \mathbf{I}_i) \mathbf{u}_i = \lambda_i \mathbf{u}_i \quad (29)$$

的最小特征值对应的单位特征向量。

总结上述过程, 可以得到多类分类 IGEP SVM 算法如下:

算法2 多类分类 IGEP SVM 算法

步骤1 给定训练集 $T = \{\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_L\}$;

步骤2 For $i \leftarrow 1$ to L

(1)通过交叉验证选择合适的参数 $\mathbf{v}_i$ 和 δ_i ;

(2)利用模型式(24)或式(28)训练, 求解特征值问题式(25)或式(29)得到最小特征值对应的单位特征向量 $\mathbf{u}_i$;

End

步骤3 利用决策函数式(23)或式(27)进行分类。

3.3 计算复杂度分析

综上所述, 对于两类分类问题, 线性 IGEP SVM 仅需要求解两个标准的特征值问题, 其计算复杂度为 $O(2n^2)$, 其中 n 为训练样本点的特征维数, 而线性 GEP SVM 则需要求解两个广义特征值问题, 其计算复杂度为 $O(2n^3)$ 。对于非线性情形, 我们的非线性 IGEP SVM 的计算复杂度为 $O(2m^2)$, 其中 m 为训练样本点的规模, 而非线性 GEP SVM 的计算复杂度为 $O(2m^3)$ 。对于 L 类多分类问题, IGEP SVM 算法无论对线性还是非线性都需要求解 L 个特征值问题, 它们的计算复杂度分别为 $O(L \cdot n^2)$ 和 $O(L \cdot m^2)$ 。因此, 由理论分析可知, IGEP SVM 训练速度要比 GEP SVM 快。

4 实验结果与分析

为了验证 IGEP SVM 算法的有效性, 分别在 8 个两类分类和 6 个多类分类 UCI 数据集上^[23]进行实验, 并分别与传统 SVM, GEP SVM 和 TWSVM 方法进行对比。所有实验均以 Matlab R2013a 为实验环境, 以 Intel P4(3.40 GHz)处理器、4 GB 内存的 PC 为硬件平台。在实验过程中, 使用 LIBSVM 工具包^[24]实现传统 SVM, 使用逐次超松弛技术 (SOR)^[10]来求解 TWSVM 中的 QPP 问题, 使用 Matlab 的 “eig” 函数求解 GEP SVM 和 IGEP SVM 分别对应的广义特征值问题和标准特征值问题。对于非线性情形, 选择 RBF 核 $K(x, y) = e^{-\mu \|x - y\|^2}$ 为核函数, 其中 μ 为核参数。参数选择对于各算法性能有一定的影响, 因此, 对于不同的数据集, 使用 10-折交叉验证方法为算法选择最优参数, 所有算法中的参数选取范围均为 $\{2^i \mid i = -8, -7, \dots, 7, 8\}$ 。

针对两类分类问题, 我们选择了 8 个 UCI 数据集来验证本文算法的有效性, 表 1 和表 2 分别给出了 4 种算法在线性和非线性情形下的 10-折交叉验证结果和训练时间。针对多类分类问题, 我们选择了 6 个 UCI 数据集来验证本文算法的有效性, 表 3 和表 4 分别给出了 4 种算法在线性和非线性情形下的 10-折交叉验证结果和训练时间。在表 1 至表 4 中, 为了更好地体现算法的优越性, 用黑体数字代表特定数据集下获得的最高分类精度。同时, 我们在表的倒数第 2 行给出了所有算法的平均分类精度和训练时间, 并在表的最后一行用 “W-T-L” (win-tie-loss) 概括了 IGEP SVM 算法与其它算法的性能。

对于两类分类问题, 从表 1 和表 2 中可以发现, 从分类精度来说, IGEP SVM 算法比 GEP SVM 和传统的 SVM 表现出了更好的性能, 同时也获得了与 TWSVM 算法非常类似的性能; 但从训练时间上来说, 本文算法表现出了更好的优势, 虽然在非线性情形下 SVM 也有较好的结果, 那可能是由于 LIBSVM 工具包实现 SVM 利用了 SMO 算法的原因。容易看出, 本文方法 IGEP SVM 在绝大多数的数据集上训练时间相对较少, 尤其与 GEP SVM 方法相比较, 训练时间明显减少, 进一步验证了本文方法是对 GEP SVM 方法的一种有效改进。

对于多类分类问题而言, 观察表 3 和表 4 不难发现, 一方面, 无论在分类精度还是训练时间, 本文 IGEP SVM 算法比 GEP SVM 算法都有所提高。另一方面, 本文 IGEP SVM 算法与 SVM 和 TWSVM 相比较, 从表 3 和表 4 的最后两行看出, 本文方法在分类精度上要稍微低一点, 但也取得了较好的效果; 然而, 本文方法在训练时间上总体上是有一定

表 1 线性 IGEPSVM 与几种算法在两类分类 UCI 数据集上性能比较

数据集	IGEPSVM	GEPSVM	SVM	TWSVM
	精度±方差(%)时间(s)	精度±方差(%)时间(s)	精度±方差(%)时间(s)	精度±方差(%)时间(s)
Australian(690×14)	86.52±3.21 , 3.8563e-04	85.80±4.67, 5.1659e-04	85.51±3.86, 0.0185	86.10±4.55, 0.0381
House votes(435×16)	94.26±2.68, 2.6548e-04	94.03±2.88, 4.4901e-04	94.73±3.77, 0.0033	95.87±3.21 , 0.0063
Heart-c(303×14)	84.87±5.85, 2.2177e-04	84.49±7.40, 5.4313e-04	84.83±5.41, 0.0029	85.13±5.07 , 0.0277
Heart-Statlog(270×13)	84.70±4.68, 2.2044e-04	84.59±8.20, 3.5033e-04	84.44±4.88, 0.0023	84.81±4.77 , 0.0333
Monk3(432×7)	83.47±5.66, 1.3137e-04	80.76±7.00, 2.2355e-04	82.59±6.03, 0.0027	85.91±3.96 , 0.0464
Sonar(208×60)	81.98±7.96 , 0.0032	79.31±5.59, 0.0052	81.69±6.88, 0.0035	81.93±7.28, 0.0371
Spect(267×44)	80.51±5.78 , 0.0016	79.46±7.54, 0.0025	80.45±7.46, 0.0030	80.24±6.21, 0.0691
Wpbc(198×34)	80.32±2.26 , 0.0010	78.18±6.88, 0.0015	80.14±6.90, 0.0022	79.89±5.23, 0.0431
平均精度/平均时间	84.58/0.0009	83.33/0.0014	84.30/0.0048	84.99 /0.0376
W-T-L(胜-平-负)	8-0-0		7-0-1	4-0-4

表 2 非线性 IGEPSVM 与几种算法在两类分类 UCI 数据集上性能比较

数据集	IGEPSVM	GEPSVM	SVM	TWSVM
	精度±方差(%)时间(s)	精度±方差(%)时间(s)	精度±方差(%)时间(s)	精度±方差(%)时间(s)
Australian(690×14)	86.67±5.19, 0.7215	86.10±2.51, 5.2741	86.23±4.70, 0.5162	86.96±4.27 , 0.7381
House votes(435×16)	95.42±2.85, 0.1356	94.12±5.05, 0.8937	95.24±4.07, 0.1112	96.32±2.49 , 0.1641
Heart-c(303×14)	84.86±5.20, 0.0810	82.52±7.49, 0.2952	83.19±8.27, 0.0652	85.13±2.47 , 0.0834
Heart-Statlog(270×13)	84.07±4.64, 0.0165	83.33±7.25, 0.2023	83.59±7.21, 0.0140	84.81±7.08 , 0.0184
Monk3(432×7)	98.72±4.80 , 0.2135	95.84±3.29, 0.8769	97.25±3.18, 0.1045	98.26±1.32, 0.2450
Sonar(208×60)	87.24±6.33, 0.0153	86.55±8.08, 0.1533	86.10±8.51, 0.0257	89.26±7.19 , 0.0289
Spect(267×44)	82.71±8.96 , 0.0675	80.48±8.09, 0.3178	82.46±8.54, 0.0549	82.07±5.91, 0.0709
Wpbc(198×34)	82.76±6.33 , 0.0353	81.26±8.24, 0.1137	82.56±5.38, 0.0534	82.74±5.23, 0.0504
平均精度/平均时间	87.81/0.1608	86.28/1.0159	87.08/0.1181	88.19 /0.1749
W-T-L(胜-平-负)	8-0-0		8-0-0	3-0-5

表 3 线性 IGEPSVM 与几种算法在多类分类 UCI 数据集上性能比较

数据集	IGEPSVM	GEPSVM	SVM	TWSVM
	精度±方差(%)时间(s)	精度±方差(%)时间(s)	精度±方差(%)时间(s)	精度±方差(%)时间(s)
Iris(150×4×3)	96.67±5.67 , 0.0012	95.33±4.66, 0.0028	95.33±4.50, 0.0034	96.53±5.84, 0.0209
Wine(178×13×3)	96.76±5.17, 0.0017	94.71±7.92, 0.0042	97.78±3.88, 0.0028	98.33±2.68 , 0.0669
Seeds(210×7×3)	93.71±1.23, 0.0015	89.05±5.96, 0.0018	92.86±5.14, 0.0027	95.71±4.74 , 0.0474
Dermatology(358×34×6)	95.03±6.79, 0.0080	94.14±4.97, 0.0099	95.83±3.00 , 0.0121	95.25±2.63, 0.2177
Balance(625×4×3)	88.48±3.77, 0.0017	89.69±3.72 , 0.0020	87.52±3.54, 0.0031	87.66±3.18, 0.2026
Car(1278×6×4)	77.78±2.85, 0.0045	73.59±2.84, 0.0068	84.78±4.13 , 0.0329	77.26±4.27, 2.1232
平均精度/平均时间	91.41/0.0031	89.42/0.0046	92.35 /0.0095	91.79/0.4465
W-T-L(胜-平-负)	5-0-1		3-0-3	3-0-3

表 4 非线性 IGESVM 与几种算法在 multi-class UCI 数据集上性能比较

数据集	IGESVM	GESVM	SVM	TWSVM
	精度±方差(%)时间(s)	精度±方差(%)时间(s)	精度±方差(%)时间(s)	精度±方差(%)时间(s)
Iris(150×4×3)	97.33±3.51 , 0.0100	96.67±4.71, 0.0399	96.67±3.51, 0.0121	97.33±3.44 , 0.0162
Wine(178×13×3)	97.78±2.87, 0.0444	95.49±5.20, 0.0766	98.30±2.74, 0.0463	99.44±1.76 , 0.1237
Seeds(210×7×3)	94.33±6.02, 0.0546	92.38±7.17, 0.1377	93.86±4.63, 0.0510	96.67±3.21 , 0.1508
Dermatology(358×34×6)	96.23±4.22, 0.3165	96.39±2.64, 1.0836	96.01±5.19, 0.2205	96.65±4.11 , 0.6842
Balance(625×4×3)	94.25±2.26, 0.5918	96.65±2.74, 3.7493	93.97±4.11, 0.3067	99.20±1.13 , 1.2449
Car(1278×6×4)	96.20±1.47, 3.8730	94.04±1.07, 18.4988	98.67±1.22 , 0.8437	98.44±0.67, 7.8442
平均精度/平均时间	96.02/0.8150	95.27/3.9310	96.25/0.2467	97.96/1.6773
W-T-L(胜-平-负)		4-0-2	4-0-2	0-1-5

优势的，只有在与非线性 SVM 比较时稍微差一些，主要原因可能是 SVM 采取了 LIBSVM 工具包和“一类对一类”的策略实现多类分类。总而言之，本文 IGESVM 算法是对 GESVM 算法的一种改进，无论在分类精度还是在训练时间上都取得了比 GESVM 较好的实验效果。

5 结束语

本文针对两类分类问题提出了一类改进的基于特征值分解的中心支持向量机算法(IGESVM)，并将其推广到多类分类问题。与 GESVM 算法相比较，本文的主要贡献在于：首先，解决了 GESVM 模型训练过程和决策过程不一致的问题；其次，将 GESVM 求解广义特征值问题转化为求解标准的特征值问题，降低了计算复杂度；最后，通过引入新的参数，可以方便调节 IGESVM 模型的性能，提高了分类精度。在 UCI 数据集上的实验结果表明，本文算法比 GESVM 算法更优越，但与 SVM 或 TWSVM 算法相比较，它们各有优势。近几年来，非平行平面支持向量机已经逐渐成为模式识别和机器学习领域新的研究热点之一，将已有的针对两类分类问题所提出的非平行平面支持向量机算法推广到多类分类问题以及在不同领域的应用是今后研究的重点。

参 考 文 献

- [1] CORTES C and VAPNIK V N. Support vector machine[J]. *Machine Learning*, 1995, 20(3): 273-297.
- [2] OSUNA E, FREUND R, and GIROSI F. Training support vector machines: an application to face detection[C]. *Proceedings of Computer Vision and Pattern Recognition*, San Juan, 1997: 130-136.
- [3] ISA D, LEE L H, KALLIMANI V P, *et al.* Text document preprocessing with the Bayes formula for classification using the support vector machine[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2008, 20(9): 1264-1272.
- [4] BOSER B, GUYON I, and VAPNIK V N. A training algorithm for optimal margin classifiers[C]. *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, New York, 1992: 144-152.
- [5] OSUNA E, FREUND R, and GIROSI F. An improved training algorithm for support vector machines[C]. *Proceedings of IEEE Workshop on Neural Networks for Signal Processing*, New York, USA, 1997: 276-285.
- [6] PLATT J. Fast Training of Support Vector Machines Using Sequential Minimal Optimization[M]. *Advances in Kernel Methods: Support Vector Machines*, Cambridge, MA, MIT Press, 1998: 41-65.
- [7] 张战成, 王士同, 邓赵红, 等. 支持向量机的一种快速分类算法[J]. *电子与信息学报*, 2011, 33(9): 2181-2186.
ZHANG Zhancheng, WANG Shitong, DENG Zhaozhong, *et al.* Fast decision using SVM for incoming samples[J]. *Journal of Electronics & Information Technology*, 2011, 33(9): 2181-2186.
- [8] MANGASARIAN O L and WILD E W. Multisurface proximal support vector machine classification via generalized eigenvalues[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(1): 69-74.
- [9] JAYADEVA, KHEMCHANDAI R, and CHANDRA S. Twin support vector machine classification for pattern classification [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(5): 905-910.
- [10] SHAO Y H, ZHANG C H, WANGX B, *et al.* Improvements on twin support vector machines[J]. *IEEE Transactions on Neural Networks*, 2011, 22(6): 962-968.
- [11] PENG X J. TPMSVM: A novel twin parametric-margin support vector machine for pattern recognition[J]. *Pattern Recognition*, 2011, 44(10): 2678-2692.
- [12] QI Z Q, TIAN Y J, and SHI Y. Robust twin support vector

- machine for pattern classification[J]. *Pattern Recognition*, 2013, 46(1): 305–316.
- [13] SHAO Y H, DENG N Y, and CHEN W J. A proximal classifier with consistency[J]. *Knowledge-Based Systems*, 2013, 49: 171–178.
- [14] Tian Y J, Qi Z Q, Ju X C, *et al.* Nonparallel support vector machines for pattern classification[J]. *IEEE Transactions on Cybernetics*, 2014, 44(7): 1067–1079.
- [15] DING S F, HUA X P, and YU J Z. An overview on nonparallel hyperplane support vector machine algorithms[J]. *Neural Computing and Applications*, 2014, 25(5): 975–982.
- [16] 王娜, 李霞. 基于类加权的双 ν 支持向量机[J]. 电子与信息学报, 2007, 29(4): 859–862.
- WANG Na and LI Xia. A new dual ν support vector machine based on class-weighted[J]. *Journal of Electronics & Information Technology*, 2007, 29(4): 859–862.
- [17] BOTTOU L, CORTES C, DENKER J S, *et al.* Comparison of classifier methods: a case study in handwritten digit recognition[C]. Proceedings of IEEE International Conference on Pattern Recognition, Paris, 1994: 77–82.
- [18] KREßEL U. Pairwise Classification and Support Vector Machines[M]. Advances in Kernel Methods–Support Vector Learning, Cambridge, MA, MIT Press, 1999: 255–268.
- [19] CRAMMER K and SINGER Y. On the learn ability and design of output codes for multi-class problems[J]. *Machine Learning*, 2002, 47(2/3): 201–233.
- [20] XU Y T, GUO R, and WANG L S. A twin multi-class classification support vector machine[J]. *Cognitive Computation*, 2013, 5(4): 580–588.
- [21] NASIRI J A, CHARKARI N M, and JALILI S. Least squares twin multi-class classification support vector machine[J]. *Pattern Recognition*, 2015, 48(3): 984–992.
- [22] PARLETT B. The Symmetric Eigenvalue Problem[M]. Upper Saddle River, NJ, USA, SIAM Press, 1998: 61–80.
- [23] BLAKE C L and MERZ C J. UCI repository of machine learning databases[R]. Irvine, CA: Department of Information and Computers Science, University of California, 1998.
- [24] CHANG C and LIN C. LIBSVM: A library for support vector machine[J]. *ACM Transactions on Intelligent Systems and Technology*, 2011, 2(3): 1–27.
- 陈素根: 男, 1982 年生, 副教授, 博士生, 研究方向为模式识别与智能系统.
- 吴小俊: 男, 1967 年生, 教授, 博士生导师, 研究方向为模式识别、人工智能与计算机视觉.