

文章编号: 1001-0920(2009)11-1723-06

一种 LDA 与 SVM 混合的多类分类方法

张宏达, 王晓丹, 徐海龙

(空军工程大学 导弹学院, 陕西 三原 713800)

摘要: 针对决策有向无环图支持向量机(DDAGSVM)需训练大量支持向量机(SVM)和误差积累的问题, 提出一种线性判别分析(LDA)与 SVM 混合的多类分类算法. 首先根据高维样本在低维空间中投影的特点, 给出一种优化 LDA 分类阈值; 然后以优化 LDA 对每个二类问题的分类误差作为类间线性可分度, 对线性可分度较低的问题采用非线性 SVM 加以解决, 并以分类误差作为对应二类问题的可分度; 最后将可分度作为混合 DDAG 分类器的决策依据. 实验表明, 与 DDAGSVM 相比, 所提出算法在确保泛化精度的条件下具有更高的训练和分类速度.

关键词: 决策有向无环图; 支持向量机; 线性判别分析; 分类阈值; 可分性

中图分类号: TP181

文献标识码: A

Hybrid multi-class classification approach based on LDA and SVM

ZHANG Hong-da, WANG Xiao-dan, XU Hai-long

(Missile Institute, Air Force Engineering University, Sanyuan 713800, China. Correspondent: ZHANG Hong-da, E-mail: zhdhonda@163.com)

Abstract: To the problems that decision directed acyclic graph support vector machines (DDAGSVMs) suffer from training a mass of SVMs and the error accumulation effect, a hybrid multi-class classification algorithm taking optimized linear discriminant analysis(LDA) and SVM as node classifiers is proposed. According to the characteristics of projecting high-dimension data to low-dimensional space, an optimized LDA classification threshold is derived. Then the linear separability of each pair of classes is defined with respect to the classification accuracy of the optimized LDA. The proposed algorithm trains SVMs for the dichotomies with relatively low linear separability, and then updates the separability matrix with classification accuracy of the SVMs. In the classification phase, the separability matrix is employed to decide the decision route of the DDAG. Finally, experiments show that, comparing with DDAGSVM, the proposed algorithm possesses higher training and classifying speed without the loss of generalization accuracy.

Key words: Decision directed acyclic graph; Support vector machines; Linear discriminant analysis; Classification threshold; Separability

1 引言

支持向量机(SVM)^[1]是一种二类分类方法, 各种分解重构算法可将其推广应用于多类分类问题^[2,3]. Hsu 等^[4]对多种多类 SVM(MSVM)算法进行了对比研究, 认为从训练速度、分类速度和分类精度上综合考虑, DDAGSVM 是一种实用的 MSVM 方法. 虽然 DDAGSVM 采用“一对一”(1-v-1)方法, 每次训练集规模降到最低, 但当类别数较大时, $K(K-1)/2$ 个二类 SVM 仍需要大量训练时间.

在多类分类问题中, 不同的二类问题所对应的类间可分性差异很大, 有些是线性可分的或线性可分性较高, 有些则属于线性不可分问题. 若对这些二类问题统一采用 SVM(若无特别说明, 下文中均指非线性 SVM), 则将造成训练和分类时间的浪费. 对于线性可分问题, 采用线性分类方法效率更高, 常用的线性方法包括线性判别分析(LDA)和线性 SVM 等. 文献[5,6]分别证明了 LDA 与线性 SVM 的内在关系, 并通过实验表明了 LDA 与线性 SVM 性能

收稿日期: 2008-12-14; 修回日期: 2009-04-14.

基金项目: 陕西省自然科学基金项目(2007F19); 空军工程大学导弹学院研究生学位论文创新基金项目(DY06102).

作者简介: 张宏达(1981—), 男, 安徽临泉人, 博士生, 从事智能信息处理、机器学习的研究; 王晓丹(1966—), 女, 陕西汉中, 教授, 博士, 从事智能信息处理、机器学习等研究.

相当,且 LDA 训练速度更快. 因此,对不同问题分别采用 LDA 和非线性 SVM,有望提高多类分类器的效率.

文献[2,7,8]在构造多类分类器时,均采用类间可分性作为参考指标,分别依据欧式距离^[2]、向量投影法^[7]和分类误差^[8]度量类间可分性. 依据类间可分性构造 DDAG,有望降低误差积累. 本文采用改进的 LDA 方法计算类别之间的线性可分性,对于线性可分性较差的二类问题采用 SVM 实现,根据 LDA 和 SVM 对于各二类问题的分类误差确定可分性矩阵,并将该可分性矩阵作为分类时节点分类器选择的依据. 下面将这种多类分类算法称为 Daglas (DDAG with LDA and SVM).

2 SVM 简介

设训练集

$$D = \{(\mathbf{x}_i, y_i)\}_{i=1}^l, \mathbf{x}_i \in R^d, y_i \in \{1, -1\},$$

则标准支持向量机(C-SVM)的优化形式为

$$\begin{aligned} \min J(\mathbf{w}, b, \xi) &= \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i; \\ \text{s. t. } & y_i[\mathbf{w}^T \phi(\mathbf{x}_i) + b] \geq 1 - \xi_i, \\ & \xi_i \geq 0, i = 1, \dots, l. \end{aligned} \quad (1)$$

其中: $\mathbf{w} \in R^d$ 为分类超平面的法向量, $\phi(\mathbf{x}_i)$ 为核函数 $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$ 定义的隐式映射. 对式(1)采用 Lagrange 函数方法,可得如下等价的对偶形式:

$$\begin{aligned} \max_{\alpha} L(\alpha) &= \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j), \\ \text{s. t. } & \sum_{i=1}^l y_i \alpha_i = 0, 0 \leq \alpha_i \leq C, i = 1, \dots, l. \end{aligned} \quad (2)$$

求解式(2),得支持向量集

$$SV_s = \{\mathbf{x}_i \mid \alpha_i > 0, \mathbf{x}_i \in D\},$$

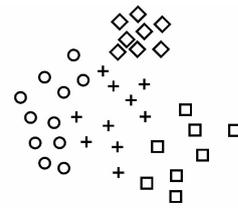
则分类函数为

$$\varphi(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^{n_s} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b\right),$$

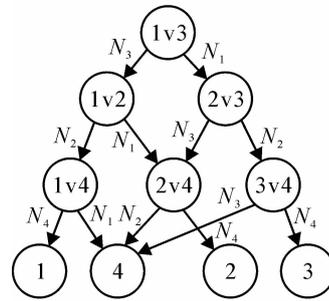
其中 $n_s = |SV_s|$.

3 DDAG 方法

对于 K 类分类问题,DDAG^[3] 的训练阶段与“一对一”方法一致,因此具有较高的训练速度. 在分类阶段,将训练得到的 $K(K-1)/2$ 个分类器按照一定的次序构成一个有根的有向无环图(DAG). DDAG 中共有 $K(K-1)/2$ 个由二类分类器构成的内部节点以及 K 个表示最终分类结果的叶子节点. 如图 1 所示,对于给定样本,DDAG 从根节点开始,根据每个内部节点的判决结果决定下一步的方向,经过 $K-1$ 个分类器到达叶子节点.



(a) 分类问题的样本分布



(b) 对应的 DDAG

图 1 一个小规模四类分类问题的 DDAG

4 线性判别分析

4.1 类间可分性

定义 1^[9] 给定两类样本集 D ,若存在(非)线性函数能将 D 正确划分,则称样本集 D (非)线性可分.

定义 2 任意线性分类器 φ 对 D 的分类误差称为 D 的 φ 线性可分度,任意非线性 φ 对 D 的分类误差称为 D 的 φ 非线性可分度,统称为类间可分度.

由上述定义可知,类间可分性具有如下性质:

性质 1 若 D 满足线性可分,则 D 一定非线性可分,反之则不成立.

性质 1 说明非线性分类器在某种条件下与线性分类器是等价的. 文献[10]认为线性核是 RBF 核函数的一种特例,在某种参数设置下,RBF-SVM 具有与线性 SVM 一样的性能. 但是,若参数选择不当,则 RBF-SVM 性能可能不如线性分类器. 这个事实为本文采用线性方法解决 DDAG 中的部分二类问题提供了依据.

定义 3 对于 K 类分类问题,二类分类模型 φ_{ij} 对类别 $i, j(1 \leq i, j \leq K)$ 的分类误差 s_{ij} 称为类别 i, j 之间的可分度,称 $\mathbf{M}_s = \{s_{ij}\}_{i,j=1}^K$ 为可分度矩阵.

4.2 LDA 原理

LDA 的目标是将数据投影到判别方向,使得数据集在该方向上的投影满足类间距离大而类内距离小,在该方向上实现类别之间的划分. 对于二类问题,有 $D_1 = \{\mathbf{x}_i \mid y_i = 1\}_{i=1}^l$ 和 $D_2 = \{\mathbf{x}_i \mid y_i = -1\}_{i=1}^l$,类别样本的均值为 $\mu_i, i = 1, 2$. 样本集的类内散布矩阵与类间散布矩阵为

$$S_i = \sum_{\mathbf{x} \in D_i} (\mathbf{x} - \mu_i)(\mathbf{x} - \mu_i)^T,$$

$$S_w = S_1 + S_2, S_b = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T.$$

LDA 目标为

$$\max_w J_F(w) = \frac{w^T S_b w}{w^T S_w w}. \quad (3)$$

用 Lagrange 函数法可得 $w = S_w^{-1}(\mu_1 - \mu_2)$, 采用 $f(x) = \text{sgn}(w^T x - b)$ 进行分类, b 为分类阈值.

4.3 LDA 的分类阈值

文献[12] 给出了如下 3 种分类阈值:

$$b_1 = \frac{1}{2} w^T (\mu_1 + \mu_2), b_2 = \frac{n_1 w^T \mu_1 + n_2 w^T \mu_2}{n_1 + n_2},$$

$$b_3 = \frac{w^T (\mu_1 + \mu_2)}{2} + \frac{\log(P_1/P_2)}{n_1 + n_2 - 2}.$$

其中 b_2 和 b_3 以不同方式引入了对类别先验概率的考虑, 但没有考虑类别之间分布的差异. 为得到更合理的分类阈值, 下面给出一种近似 Bayes 最优的 LDA 分类阈值.

对于二类问题, 假设两类别 ω_1 和 ω_2 的先验概率为 $P(\omega_i) = |D_i| / (|D_1| + |D_2|), i = 1, 2$, Bayes 分类准则形如

$$\begin{aligned} & \text{if } \frac{p(x | \omega_i)}{p(x | \omega_j)} > \frac{P(\omega_j)}{P(\omega_i)}, \text{ then } x \in \omega_i; \\ & \text{s. t. } i, j = 1, 2, i \neq j. \end{aligned} \quad (4)$$

可见, 分类判决不仅与类别先验概率有关, 同时与类别条件概率有关. LDA 方法将样本从原特征空间投影到 w 定义的一维空间, 根据高维数据在低维空间投影的统计特性^[11], 可近似认为每类样本在 w 上的投影是服从高斯分布的, 其中类别的方差为

$$\sigma_i = \sqrt{\frac{1}{l_i} \sum_{x \in D_i} (w^T(x - \mu_i))^2}, i = 1, 2;$$

一维空间中的类别条件概率为

$$p(x | \omega_i) = \frac{1}{\sqrt{2\pi} \sigma_i} \exp\left(-\frac{(x - \mu_i)^2}{2\sigma_i^2}\right).$$

因此, 依据 Bayes 准则, 分类判别需同时考虑类别的先验概率和类别的分布.

对于服从 Gauss 分布的一维分类问题, 若设定分类阈值为 t , 则 Bayes 分类误差(参考图 2) 近似为

$$\begin{aligned} El(t) = & 1 - \int_{-\infty}^t \frac{1}{\sqrt{2\pi} \sigma_1} \exp(-x^2/2\sigma_1^2) P(\omega_1) dx + \\ & \int_{-\infty}^t \frac{1}{\sqrt{2\pi} \sigma_2} \exp(-(x-a)^2/2\sigma_2^2) P(\omega_2) dx, \end{aligned} \quad (5)$$

其中 $a = w^T(\mu_2 - \mu_1)$. 令 $\frac{d}{dt} El(t) = 0$, 可得

$$\exp\left(-\frac{x^2}{2\sigma_1^2}\right) \frac{P(\omega_1)}{\sigma_1} = \exp\left(-\frac{(x-a)^2}{2\sigma_2^2}\right) \frac{P(\omega_2)}{\sigma_2}. \quad (6)$$

解式(6), 保留落入 $[0, a]$ ($a > 0$) 或 $[a, 0]$ ($a \leq 0$)

内部的解(图 2 中 t). 若该区间中不存在解, 则说明两个类别交叠严重, 此时仅考虑类别之间样本规模的不均衡性, 取 $t = b_2$; 若该方程无解, 也取 $t = b_2$. 前面假设 μ_1 的投影在一维特征空间的原点, 而实际上 μ_1 的投影可表示为 $w^T \mu_1$, 因此有 $b_0 = w^T \mu_1 + t$.

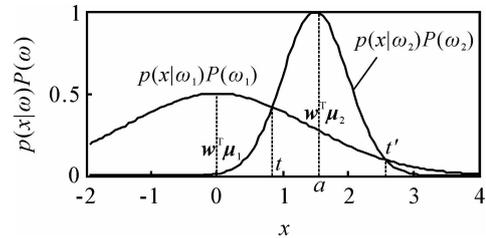


图 2 服从正态分布的两类一维问题的最优分类阈值

采用本文方法得出的判决结果的误差与 Bayes 分类误差的差别为图 2 中 t' 右侧两条曲线中间的部分(线性可分性较好时采用本文方法, 显然 $x > t'$ 部分对 ω_1 的分类误差很小).

5 Daglas 算法

5.1 算法描述

根据 1-v-1 方法, K 类问题可分解为 $K(K-1)/2$ 个二类问题. 首先对每个二类问题采用优化 LDA 得出 LDA 分类判别式, 并根据优化 LDA 得出训练集的线性可分性矩阵. 根据不同二类问题的线性可分性确定采用的分类器方式. 分类阶段每次采用最佳可分性对应的二类问题作为 DDAG 的下一步决策方向. 具体算法如下:

算法 1(Daglas 算法)

输入: 训练样本集合 D , 验证集合 D_v , 类别数 K , 线性容许误差 δ ;

输出: 分类器集合 $\{\varphi_{ij}\}$, 可分性矩阵 $M_s = \{s_{ij}\}_{i,j=1}^K$.

Step1: 将 D 按类别分为 $D = \{D_i\}_{i=1}^K, D_v$ 分为 $D_v = \{Dv_i\}_{i=1}^K$.

Step2: 若 $i \leq K-1$, 则取 $j = i+1$; 否则结束.

Step3: 若 $j \leq K$, 则转 Step4; 否则 $i = i+1$, 返回 Step2.

Step4: $\varphi_{ij} = \text{LDA}(D_i, D_j)$, 并计算

$$s_{ij} = \sum_{x_k \in D_i, D_j} I(\varphi_{ij}(x_k) = y_k).$$

Step5: 若 $s_{ij} < 1 - \delta$, $\varphi_{ij} = \text{SVM}(D_i, D_j)$, 则

$$s_{ij} = \sum_{x_k \in Dv_i, Dv_j} I(\varphi_{ij}(x_k) = y_k), j = j+1,$$

返回 Step3; 否则, $j = j+1$, 直接返回 Step3.

δ 是算法引入的唯一可调参数, 对于特定问题, δ 越大, 需训练的 SVM 越少, 整个多类分类器的训练时间越低. 当 $1 - \delta \leq \min_{i,j} (M_s)$ 时, 整个分类器退

化为线性分类器,这对于非线性可分问题,将造成分类精度降低.若 δ 取值过低,则可能增加分类器的训练和分类时间.基于对线性可分性较差的问题,以及非线性SVM性能更优的假设, δ 参考取值为 $\delta \in [0, 0.1)$.

5.2 算法有效性分析

5.2.1 算法精度

多类分类器的精度由两个条件决定:一是每个二类分类器的精度;二是分类器的组合形式或判别方法.对于前者,算法采用优化LDA解决线性可分性较高的二类问题,利用SVM解决非线性分类问题的优势来解决线性可分性较差的二类问题;对于后者,文献[3,4]已证明DDAG是一种实用的多类分类方法,而且本文依据由分类误差定义的类间可分度确定DDAG,可保证在分类误差较高的分类器远离根节点,从而有效降低误差积累.

5.2.2 算法的时间效率

训练LDA需计算类别的类内和类间散度,并需要对类内散度矩阵求逆,因此训练时间复杂度为 $O(ld^2) + O(d^3)$.采用快速算法,通常有 $\lambda < 3$,当样本的维数与样本规模满足 $d \ll l$ 时,LDA的训练时间复杂度为 $O(ld^2)$,比SVM训练时间复杂度低得多.对于每个二类问题,LDA的分类复杂度为 $O(d)$,而SVM的分类复杂度为 $O(dn_s)$ (对于非线性SVM,通常有 $n_s \gg 1$).分类时,DDAG只需判断 $K-1$ 个分类器,因为仅对分类误差较低的二类问题应用LDA,故采用上述算法,这些分类器将比较靠近根节点.若对每个样本平均采用的LDA与SVM的个数分别为 k_l 和 k_s ,则相比完全采用SVM,采用本文方法大约可节省 $k_l/(k_l + k_s)$ 的分类时间.

另外,泛化精度较高的SVM通常有着较简单的分类面,更少支持向量,因此按照由错分率定义的可分性矩阵确定DDAG,不但能够降低分类误差,而且相比随机决策的DDAG,分类时LDA和支持向量较少的SVM所占比例更大,因而该算法有望提高分类速度.

6 实验与讨论

为验证算法的有效性,首先通过实验分析所提出的LDA分类阈值的优劣,并在此基础上检验Daglas算法的有效性.

采用分离测试方法,即将样本集分为训练集、验证集和测试集.其中训练集和测试集是不相交的,验证集为测试集的一部分.实验中共采用6个UCI数据集,实验数据描述见表1,其中“样本划分”列表示训练集/验证集/测试集占整个数据集的百分比.

为使实验结果更加可靠,每次的训练集、验证集

表1 实验数据设置

数据集	样本数	类别数	特征数	样本划分/%
Iris	150	3	4	80/20/20
Wine	178	3	13	80/20/20
Glass	214	6	13	70/20/30
Balance	627	3	4	50/30/50
Image	2310	7	5	20/20/80
Yeast	1485	7	8	50/30/50

与测试集都随机产生,其中对Iris, Wine和Glass数据集实验50次,对其余数据集实验20次,取实验均值列于表中.实验机器配置为512M内存,1.80G CPU,采用Steve Gunn的SVM工具箱,整个算法用Matlab7.0实现.

6.1 近似Bayes最优分类阈值的有效性实验

表2给出了采用不同阈值对训练集进行分类时得出的平均类别间两两线性分离度,黑体表示从基于几种阈值的LDA得出的最高线性可分度,表中显示数据为10次实验均值.对于Daglas算法,给定 δ ,则由LDA定义的线性可分性越高,需要训练的SVM越少,整个分类器的训练和分类速度越高.同时,LDA定义的线性可分度越高,表示对应的线性分类器分类精度越高.统计数据显示,以 b_0 作为分类阈值时,LDA得出的线性可分度最高,这说明了近似Bayes最优分类阈值的有效性.

表2 不同LDA分类阈值对应的平均类间可分性

数据集	采用不同阈值的LDA				文献[7]方法
	b_1	b_2	b_3	b_0	
Iris	0.989	0.989	0.989	0.988	0.838
Wine	1	0.998	1	0.999	0.987
Glass	0.958	0.951	0.943	0.963	0.867
Balance	0.817	0.933	0.814	0.933	0.793
Image	0.992	0.992	0.992	0.993	0.912
Yeast	0.900	0.910	0.899	0.914	0.701

6.2 Daglas有效性验证

文献[4]已经对DDAGSVM与其他多类分类方法进行过详细的实验比较,下面实验仅限于将Daglas与采用不同节点分类器和不同决策方法的DDAG分类器比较.实验中4种算法描述见表3.

表3 对比实验中各种DDAG算法描述

算法标记	DAGLDA	DAGSVM	Daglas	RDaglas
二类分类器	LDA	SVM	LDA&SVM	LDA&SVM
决策方法	依据 M_s	依据 M_s	依据 M_s	随机选择

参考类别间平均线性可分度,对于前5个数据集取 $\delta=0.01$,对于Yeast数据集取 $\delta=0.05$.实验中,

表 4 4 种 DDAG 分类器性能对比

数据集及 参数设置(C, σ)	DAGLDA			DAGSVM			Daglas			RDaglas		
	trn T/s	tst T/s	Acc/%	trn T/s	tst T/s	Acc/%	trn T/s	tst T/s	Acc/%	trn T/s	tst T/s	Acc/%
Iris (1000, 16)	0.035	0.004	3.00	1.08	0.024	3.47	0.201	0.015	3.40	0.196	0.017	3.40
Wine (100, 2)	0.008	0.005	1.62	2.11	0.153	2.35	0.109	0.006	2.21	0.107	0.006	1.62
Balance (100, 3)	0.048	0.024	11.07	13.63	0.815	3.90	13.63	0.813	3.90	13.32	0.869	4.07
Image (100, 3)	0.212	0.475	7.16	24.03	11.59	5.65	6.55	4.43	6.04	5.79	5.23	6.22
Glass (100, 2)	0.117	0.015	36.66	3.88	0.333	29.12	2.45	0.191	28.82	2.24	0.255	30.10
Yeast (100, 10)	0.145	0.184	40.43	185.4	16.7	39.49	123.9	13.45	39.43	116.4	15.56	39.57
平均值 / %	1.32	5.43	16.66	100	100	14.00	46.83	57.05	13.97	44.26	66.04	14.16

SVM 均采用 RBF 核函数实现, 采用的超参数参考文献[4] 中的最优参数或由多次实验择优。

为确保实验结果的可靠性, 每次实验中, 4 种方法采用的训练集、验证集和测试集完全一致, 4 种方法的平均训练时间、分类时间 (trn T , tst T) 以及测试精度 (Acc, 用分类误差表示) 见表 4。每种算法的平均分类误差为该算法对各数据集分类误差的均值。算法训练和分类时间平均值的统计方法为: 对于每个数据集, 以 DAGSVM 的训练和分类时间为标准, 计算各种算法的训练 (分类) 时间占 DAGSVM 算法耗时的百分比, 最后对这些比值求平均。这是由于对于不同的数据集, 算法时间耗费可能不在同一量级上。上述方法可避免某样本集的实验结果掩盖其余实验结果, 因而统计值更可信。

根据表 4 中的统计数据, DAGLDA 比其余 3 种算法的训练速度快 33~75 倍, 分类速度快 10~18 倍, 数据集越大, DAGLDA 的速度优势越明显。对于 Iris 和 Wine, 其实验结果说明了 LDA 算法在解决线性可分度较高的问题时有优势 (若对每个二类问题选用最优参数组合, 则 DAGSVM 可能得出更高的分类精度, 但考虑到时间代价, 采用线性分类器更适合)。但是, 另外 3 种算法平均泛化能力的明显优势说明, 对有些二类问题采用非线性 SVM 的必要性。对 Balance 数据的实验结果表明, Daglas 与 DAGSVM 算法各种统计量均相当, 说明对于该数据集, Daglas 算法未能体现出优势。但从平均值看, Daglas 算法与 DAGSVM 分类精度相当, 而训练与测试速度均约比 DAGSVM 快一倍。RDaglas 与 Daglas 算法中每个二类分类器完全一致, 但 RDaglas 不需要对每个分类器计算验证误差, 因而训练速度略高于 Daglas。Daglas 相比 RDaglas 分类误差降低了 1.34%, 且分类速度比后者快 15.8%, 即依据可分性矩阵确定 DDAG, 能进一步降低分类误差。同时, 由于低复杂度的分类器引用概率增加, Daglas 分类速度更快。

7 结 论

在多元分类问题中, 不同二类问题的可分性差距很大。基于这个事实, 本文提出了 Daglas 算法。该算法针对线性可分性不同的二类问题, 分别采用了优化的 LDA 和 SVM, 结合了 LDA 的速度优势与 SVM 对于非线性可分问题的精度优势。依据类别之间的可分性构造 DDAG, 能有效降低分类误差, 且分类速度更快。当多元问题中线性可分性较高的二类问题占有一定比例时, 相比 DDAGSVM, Daglas 算法精度相当且具有明显速度优势, 是一种有效的多元分类方法。

参考文献 (References)

- [1] Cristianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods[M]. Beijing: China Machine Press, 2005.
- [2] Takahashi F, Abe S. Decision-tree-based multi-class support vector machines [C]. Proc of ICONIP. Singapore, 2002, 12: 547-553.
- [3] Platt J C, Cristianini N, Shawe-Taylor J. Large margin DAGs for multiclass classification [C]. Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2000, 12: 547-553.
- [4] Hsu C W, Lin C J. A comparison of methods for multi-class support vector machines [J]. IEEE Trans on Neural Networks, 2002, 13(2): 415-425.
- [5] Tao Q, Wu G W, Wang J. The theoretical analysis of FDA and applications [J]. Pattern Recognition, 2006, 39(5): 1199-1204.
- [6] 宋枫溪, 程科, 杨静宇, 等. 最大散度差和大间距线性投影与支持向量机 [J]. 自动化学报, 2004, 30(6): 890-896.
(Song F X, Cheng K, Yang J Y, et al. Maximum scatter difference, large margin linear projection and support vector machines [J]. Acta Automatica Sinica, 2004, 30(6): 890-896.)
- [7] 厉小润, 赵光宙, 赵辽英. 决策树支持向量机多分类器设计的向量投影法 [J]. 控制与决策, 2008, 23(7): 745-

750.
(Li X R, Zhao G Z, Zhao L Y. Design of decision-tree-based support vector machines multi-class classifier based on vector projection[J]. Control and Decision, 2008, 23(7): 745-750.)
- [8] Faraoun K M, Boukelif A. Genetic programming approach for multi-category pattern classification applied to network intrusions detection [J]. Int J of Computational Intelligence, 2007, 3(1): 79-90.
- [9] 任双桥, 魏玺章, 黎湘, 等. 基于特征可分性的核函数自适应构造[J]. 计算机学报, 2008, 31(5): 803-808.
(Ren S Q, Wei X Z, Li X, et al. Adaptive construction for kernel function based on the feature discriminability [J]. Chinese J of Computers, 2008, 31(5): 803-808.)
- [10] Keerthi S S, Lin C J. Asymptotic behaviors of support vector machines with Gaussian kernel [J]. Neural Computation, 2003, 15(7): 1667-1689.
- [11] 黄睿, 何明一, 杨少军. 一种适用于小样本问题的基于边界的特征提取算法[J]. 计算机学报, 2007, 30(7): 1173-1178.
(Huang R, He M Y, Yang S J. A margin based feature extraction algorithm for the small sample size problem [J]. Chinese J of Computers, 2007, 30(7): 1173-1178.)
- [12] 边肇祺, 张学工. 模式识别[M]. 第 2 版. 北京: 清华大学出版社. 2000: 178-185.
(Bian Z Q, Zhang X G. Pattern recognition[M]. 2nd ed. Beijing: Qinghua University Press, 2000: 178-185.)

(上接第 1718 页)

- [12] Solteiro Pires E J, Moura Oliveira P B, Tenreiro Machado J A. Multi-objective maximin sorting scheme [J]. IEEE Trans on Evolutionary Computation, 2000, 8(2): 173-195.
- [13] 徐鸣. 基于直流变频技术的压缩机驱动系统的研究 [D]. 杭州: 浙江工业大学, 2006.
(Xu M. The research of compressor drive system based on DC converter technology [D]. Hangzhou: Zhejiang University of Technology, 2006.)
- [14] 金欣磊. 基于 PSO 的多目标优化算法研究及应用 [D]. 杭州: 浙江大学, 2006.
(Jin X L. PSO-based multi-objective optimization algorithm research and its applications [D]. Hangzhou: Zhejiang University, 2006.)

(上接第 1722 页)

- [2] Edwards C, Spurgeon S K. Sliding mode control: Theory and applications of systems and control [M]. London: Taylor & Francis, 1998.
- [3] 冯勇, 安澄全, 李涛. 采用双滑模平面减小一类非线性系统稳态误差[J]. 控制与决策, 2000, 15(3): 361-364.
(Feng Y, An C Q, Li T. Steady state error reduction using two sliding surfaces in a class of nonlinear systems [J]. Control and Decision, 2000, 15(3): 361-364.)
- [4] Filippov A. Differential equations with discontinuous right-hand side [J]. American Mathematical Society Translations, 1964, 42(2): 199-231.
- [5] Filippov A. Differential equations with discontinuous right-hand sides [M]. Netherlands: Kluwer Academic Publishers, 1988.
- [6] Clarke F H. Generalized gradients and applications [J]. Trans of the American Mathematical Society, 1975, 205: 247-262.
- [7] Clarke F H. Optimization and nonsmooth analysis [M]. New York: Wiley-Interscience, 1983.
- [8] Clarke F H, Ledyaev Y S, Stern R J, et al. Nonsmooth analysis and control theory [M]. New York: Springer-Verlag, 1998.
- [9] Paden B, Sastry S. A calculus for computing Filippov differential inclusion with application to the variable structure control of robot manipulators [J]. IEEE Trans on Circuits and Systems, 1987, 34(1): 73-82.
- [10] Shevitz D, Parden B. Lyapunov stability theory of nonsmooth systems [J]. IEEE Trans on Automatic Control, 1994, 39(9): 1910-1914.
- [11] Bacciotti A, Ceragioli F. Stability and stabilization of discontinuous systems and nonsmooth Lyapunov functions [J]. Esaim-Cocv, 1999, 4: 361-376.