

DOI:10.3969/j.issn.1673-4785.201112019

网络出版地址: <http://www.cnki.net/kcms/detail/23.1538.TP.20120712.1137.011.html>

相关向量机分类方法的研究进展与分析

赵春晖, 张毅

(哈尔滨工程大学 信息与通信工程学院, 黑龙江 哈尔滨 150001)

摘要: 相关向量机(RVM)是一种基于贝叶斯模型的监督机器学习算法,可用于处理回归以及分类问题.与支持向量机(SVM)相比,相关向量机的优点在于其输出结果是一种概率模型,其相关向量的个数远远小于支持向量的个数,并且测试时间短.总结了相关向量机的基本原理及主要应用领域,详细阐述了相关向量机的模型结构以及分类方法,重点介绍了在高光谱图像分类中的应用.并针对RVM算法在高光谱图像分类中的不足,给出了多种改进方案,并作了比较.希望对研究者今后的研究有所启发,以促进该领域的发展.

关键词: 相关向量机;改进型相关向量机;高光谱图像;分类算法

中图分类号: TP751.1 **文献标志码:** A **文章编号:** 1673-4785(2012)04-0294-08

Research progress and analysis on methods for classification of RVM

ZHAO Chunhui, ZHANG Yi

(College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China)

Abstract: The relevance vector machine (RVM) is a machine learning algorithm which is based on supervision of a Bayesian model. It can be used to deal with regression and classification problems. Compared with the support vector machine (SVM), the relevance vector machine has the advantage that its output is a probability model and the number of relevance vectors is far fewer than the number of support vectors. In this paper, the application was summarized with a relevance vector machine and the classification of a hyperspectral image with RVM was introduced; the RVM model and the method of classification were also explained. In light of the disadvantage of classification, some improved methods were summarized. Various methods were generalized and analyzed while attempting to find breakthroughs and promote further research.

Keywords: relevance vector machine; improved relevance vector machine; hyperspectral image; classification algorithm

机器学习在人工智能的研究中具有十分重要的地位.其应用已遍及人工智能的各个分支,如专家系统、自动推理、自然语言理解、模式识别、计算机视觉、智能机器人等领域.其中尤为典型的是专家系统中的知识获取瓶颈问题,人们一直试图采用机器学习的方法加以解决.

支持向量机(support vector machine, SVM)^[1]是由Vapnik于1995年提出的,建立在统计学理论和结构风险最小原理的基础上,在模型复杂性与学

习能力之间寻求最佳折衷,是监督分类的典型代表,已经成为模式识别以及机器学习研究领域比较领先的方法之一.但SVM仍存在着一定的局限性:1)支持向量的数量会随着训练样本数量的增加而线性增加,这样会导致分类效率降低,训练时间过长;2)预测结果没有后验概率,无法通过结果观测到分类的准确度;3)对于给定的误差参数,主观性过强,容易造成主观误差;4)核函数必须满足Mercer条件^[2-3].

相关向量机(relevance vector machine, RVM)^[1,4]是Tipping于2001年提出的一种与SVM类似的稀疏概率模型.它可以在保证分类精度与SVM相同的情况下,提供一个后验概率分布.与SVM相比,相关向量的个数远远小于支持向量的个

收稿日期:2011-12-30. 网络出版日期:2012-07-12.

基金项目:国家自然科学基金资助项目(61077079);教育部博士点基金资助项目(20102304110013).

通信作者:赵春晖. E-mail: zhaochunhui@hrbeu.edu.cn.

数,并且其核函数也不需要满足 Mercer 条件.但是 RVM 在解决以上 SVM 问题的同时也引入了新的问题.当 RVM 应用于高光谱图像分类时,由于高光谱图像数据样本大、维数高^[5],会使分类过程中的训练时间随着样本复杂度的增加而直线上升,严重影响了分类的效率.因此 Bishop 等提出一种改进型相关向量机方法^[6],该方法主要是通过引入一个新的概率分布,降低训练过程中的运算复杂度,提高分类效率,在小样本情况下有较好的分类精度.

1 相关向量机理论

1.1 相关向量机模型

如果输入数据组为 $\{x_n, t_n\}_{n=1}^N$, 并且考虑到目标函数只是一个标量,根据标准概率方程式,假设目标函数是模型的样本并且附加着噪声:

$$t_n = y(x_n; w) + \varepsilon_n. \quad (1)$$

式中: ε_n 是零均值的高斯噪声,并且相互独立,其方差为 σ^2 . 因此表达式 $p(t_n | x) = N(t_n | y(x_n), \sigma^2)$ 是服从高斯分布的表达式,其分布由 t_n 和 $y(x_n)$ 的值以及方差 σ^2 决定. 而其中 $y(x)$ 是一个由核函数决定的值,并且核函数由训练样本 $\phi_i(x) \equiv K(x, x_i)$ 决定. 假设 t_n 是相互独立的,则样本数据的完整概率表达式可以写为

$$p(t | w, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2} \|t - \Phi w\|^2\right\}. \quad (2)$$

式中: $t = [t_1 \ t_2 \ \cdots \ t_N]^T$, $w = [w_0 \ w_1 \ \cdots \ w_N]^T$ 和 Φ 都是预先设计好的大小为 $N \times (N+1)$ 的矩阵,并且 $\Phi = [\phi(x_1) \ \phi(x_2) \ \cdots \ \phi(x_N)]^T$ 的值为 $\phi(x_n) = [1 \ K(x_n, x_1) \ K(x_n, x_2) \ \cdots \ K(x_n, x_N)]^T$. 为了使表达式更加简单化,在式(2)中忽略一些隐含的输入数据 $\{x_n\}$.

随着训练样本中参数被大量地使用,在进行 w 和 σ^2 的最大似然估计时会产生过适应现象. 为了避免这种现象的发生,可以对一些参数加入一定附加强制条件. 这种作法已经在 SVM 中大量使用,并且非常有效.

在 RVM 中采用贝叶斯透视方法作为强制条件. 定义一个 w 的比较简单函数,这个函数是一个应用广泛的零均值的高斯概率分布:

$$p(w | \alpha) = \prod_{i=0}^N N(w_i | 0, \alpha_i^{-1}). \quad (3)$$

对于每一个权重来说,式(3)中都存在独立分布的参数,可以极大地缓解之前分布的复杂度^[6].

定义一个超参数 α , 同样地为了匹配最后的函

数,也需要加入噪声函数 σ^2 . 这个函数属于一种刻度函数,其分布满足伽玛分布:

$$p(\alpha) = \prod_{i=0}^N \text{Gamma}(\alpha_i | a, b),$$

$$p(\beta) = \text{Gamma}(\beta | c, d).$$

式中: $\beta = \sigma^2$, 并且

$$\text{Gamma}(\alpha | a, b) = \Gamma(a)^{-1} b^a \alpha^{a-1} e^{-b\alpha}.$$

式中: $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt$ 为伽马函数. 为了确保这些参数是无先验知识的,规定它们的取值很小,如 $a = b = c = d = 10^{-4}$. 若将这些参数设为 0, 则会获得均匀的超参数. 在大多数的文献中一般取 $a = b = c = d = 0$.

1.2 相关向量机回归方法

在超参数估计的收敛过程中,预测部分是基于后验概率模型的加权,并且找到最大化的 α_{MP} 与 σ_{MP}^2 . 然后根据

$$p(t_i | t) = \int p(t_i | w, \alpha, \sigma^2) p(w, \alpha, \sigma^2 | t) dw d\alpha d\sigma^2$$

计算分布的预测. 对于一个新的参数 x_i , 利用

$$p(w | t, \alpha, \sigma^2) = (2\pi)^{-(N+1)/2} |\Sigma|^{-1/2} \cdot$$

$$\exp\left\{-\frac{1}{2}(w - \mu)^T \Sigma (w - \mu)\right\},$$

可以得到

$$p(t_i | t, \alpha_{MP}, \sigma_{MP}^2) =$$

$$\int p(t_i | w, \sigma_{MP}^2) p(w | t, \alpha_{MP}, \sigma_{MP}^2) dw.$$

可简化为

$$p(t_i | t, \alpha_{MP}, \sigma_{MP}^2) = N(t_i | y_i, \sigma_i^2).$$

式中:

$$y_i = \mu^T \phi(x_i), \sigma_i^2 = \sigma_{MP}^2 + \phi(x_i)^T \Sigma \phi(x_i).$$

所以预测的含义就是直观地得到 $y(x_i; \mu)$, 或者通过后验权均值预测基础函数的权值, 其中很多基础函数的权值都是 0. 数据中的噪声估计决定了预测权值的不确定性^[7].

1.3 相关向量机分类方法

相关向量机分类方法就是一种拉普拉斯逼近的回归算法^[8].

若需要预测输入 x 的部分后验概率, 可根据统计学原理, 利用函数 $\sigma(y) = 1/(1 + e^{-y})$ 对线性模型 $y(x)$ 进行归一化, 其中 $p(t | x)$ 是一个贝努利分布, 其分布表达式为

$$p(t | w) = \prod_{n=1}^N \sigma\{y(x_n; w)\}^{t_n} \cdot [1 - \sigma\{y(x_n; w)\}]^{1-t_n}. \quad (4)$$

式中:根据概率定义,目标函数 $t_n \in \{0, 1\}$, 需要注意式(4)中并没有附加的噪声函数。

分类过程中,不能利用卷积方式计算权重,所以并不能给出 $p(\mathbf{w}|\mathbf{t}, \alpha)$ 或边缘分布 $p(\mathbf{t}|\alpha)$ 的解析解。因此,需要利用拉普拉斯逼近的近似解求得,具体过程如下:

1) 首先固定 α 的值,给出模型的后验概率分布位置,得到权重 \mathbf{w}_{MP} 的最可能值。因为 $p(\mathbf{w}|\mathbf{t}, \alpha) \propto p(\mathbf{t}|\mathbf{w})p(\mathbf{w}|\alpha)$, 所以这一过程等价于找到式(5)的最小值

$$\lg\{p(\mathbf{t}, \mathbf{w})p(\mathbf{w}|\alpha)\} = \sum_{n=1}^N [t_n \lg y_n + (1 - t_n) \lg(1 - y_n)] - \frac{1}{2} \mathbf{w}^T \mathbf{A} \mathbf{w}. \quad (5)$$

式中: $y_n = \sigma\{\gamma(\mathbf{x}_n; \mathbf{w})\}$. 应用最小平方反复迭代方法得到 \mathbf{w}_{MP} 的值。

2) 拉普拉斯方法是利用一个简单的二次方程逼近 Log 分布的方法。其值是通过二次变形给出:

$$\nabla_{\mathbf{w}} \nabla_{\mathbf{w}} \lg p(\mathbf{w}|\mathbf{t}, \alpha)|_{\mathbf{w}_{\text{MP}}} = -(\Phi^T \mathbf{B} \Phi + \mathbf{A}). \quad (6)$$

式中: $\mathbf{B} = \text{diag}(\beta_1, \beta_2, \dots, \beta_N)$ 是一个对角线矩阵,其中 $\beta_n = \sigma(\gamma(\mathbf{x}_n)) [1 - \sigma(\gamma(\mathbf{x}_n))]$. 对于高斯逼近来说权重主要集中在 \mathbf{w}_{MP} , 并且通过式(6)可以得到协方差矩阵 Σ .

3) 利用 Σ 的统计学和 \mathbf{w}_{MP} 的高斯逼近(代替 μ), 超参数 α 可以利用 $\alpha_i^{\text{new}} = \frac{\gamma_i}{\mu_i^2}$ 来更新。

在模型 $p(\mathbf{w}|\mathbf{t}, \alpha)$ 中, 利用式(6)以及 $\nabla_{\mathbf{w}} \lg p(\mathbf{w}|\mathbf{t}, \alpha)|_{\mathbf{w}_{\text{MP}}} = 0$, 可以得到

$$\Sigma = (\Phi^T \mathbf{B} \Phi + \mathbf{A})^{-1}, \quad (7)$$

$$\mathbf{w}_{\text{MP}} = \Sigma \Phi^T \mathbf{B} \mathbf{t}. \quad (8)$$

式(7)和(8)可用于解决最小二乘问题。可以看出拉普拉斯逼近方法能够有效地将分类问题映射为回归问题。

2 相关向量机的应用

因为相关向量机可用于回归以及分类问题,所以其应用非常广泛。目前国内外主要应用领域有以下几方面。

2.1 医学影像处理

相关向量机在医学领域中的主要应用是对数字医学图片的分析。首先 Lukic 等在 2007 年将 RVM 应用在分析神经成像之中,所采用的 RVM 方法与处理一般图像的现有方法有着相近的效果,然而 RVM 的复杂度比其他方法要低很多,并且可以准确

地找到感兴趣的目标^[7]。

之后 Gholami 等于 2010 年将 RVM 应用在新生儿痛苦评价上,成功地对新生儿痛苦表情进行分类,正确区分了痛苦与非痛苦的婴儿,并且对痛苦程度进行区分^[8]。需要注意的是,该方法也是基于新生儿表情的数字图像处理,大多归结为回归或者分类的问题,因此相关向量机在该领域的应用是很有前景的。

2.2 数字图像处理

对于一般数字图像处理,相关向量机的应用也主要分为处理回归和处理分类问题两大方面。

在处理回归问题中,Agarwal 等^[9]在 2006 年、Guo 等^[10]在 2007 年、Sedai 等^[11]在 2009 年、Guo 等^[12]在 2009 年,将 RVM 应用于人类 3-D 图像从单眼图像恢复过程中,该算法主要应用了 RVM 的回归理论,通过真实数据的训练, RVM 回归机可以正确捕捉人体行动,算法可以将正确率提高到 90% 左右。Ouyang 等在 2009 年,将 RVM 应用于网络视频监控技术,结果表明,采用了 RVM 回归算法的网络视频监视系统可以较快速地找到人体框架,并且 RVM 的寻找速度要比其他方法更快,尤其在测试进行一段时间之后,其分辨的正确率也很高,可达到 80% 以上^[13]。Oliveri 等在 2011 年将 RVM 应用于图像恢复中。研究表明,将快速 RVM 算法与其他算法相结合,其图像恢复效果更好,运算速度也较快^[14]。

在分类处理方法中 X. Li 等^[15]和 X. Wang 等^[16]在 2009 年将 RVM 应用于图像分类比较中,实验结果表明 RVM 分类方法对于线性数据以及非线性的高维数据都具有较好的分类效果。

2.3 高光谱图像分类

由于高光谱图像样本数量大、维数较高等特点,相关向量机分类方法非常适合应用于高光谱图像分类。Demir 等于 2007 年在高光谱图像分类中应用了 RVM 方法,详细阐述了 RVM 方法与 SVM 方法的不同,并且比较了分类结果,利用分类精度、相关向量和支持向量的个数等标准,评价了各自优缺点^[17]。随后 Mianji 等在 2011 年又进一步比较了 RVM 方法与 SVM 方法在高光谱图像分类中的性能,比较的标准增加了训练时间以及测试时间^[18]。通过实验发现 RVM 方法与 SVM 方法在训练样本相同时,其分类精度基本一致,但相关向量的数量明显少于支持向量的数量。在测试时间方面, RVM 的测试时间非常短,几乎可以做到实时测试,而 SVM 的测试时间相对较长。然而 RVM 的缺点在于随着训练样本的增加,其训练时间也会直线上升。分析比较结果如表 1、2。

表1 分类精度比较

Table 1 The accuracy of classification

训练 样本	测试 样本	SVM		RVM	
		分类 精度/%	支持 向量数	分类 精度/%	支持 向量数
590	4 588	84.39	937	80.95	192
1 320	4 588	87.94	1 510	84.45	254
2 376	4 588	90.56	2 217	87.05	353
4 757	4 588	92.67	3 393	90.32	592

表2 分类时间

Table 2 The time of classification

样本数量	SVM		RVM	
	训练时间	测试时间	训练时间	测试时间
230	2.2	2.6	13.5	0.39
693	3	3.6	49.2	0.40
2 967	19.5	25.6	1 718	3.40
3 500	21.4	27.6	1 806	3.90

为了克服 RVM 在高光谱图像分类中存在的缺点,根据 RVM 自身特点,下面给出了多种改进方案。

3 相关向量机分类方法的改进

3.1 改进型相关向量机

3.1.1 改进型相关向量机模型

传统的概率模型可以将随机变量分为观测数据 D 和非观测变量 θ ^[6, 19], 观测数据 D 的边缘概率密度为

$$P(D) = \int P(D, \theta) d\theta. \quad (9)$$

对于式(9)的积分计算,通常来说较为复杂.因此改进思想是通过引入近似的分布 $Q(\theta)$, 其中对于任意选择的非观测变量 θ , 可以将边缘概率密度的积分形式分解为 2 个对数之和,即将式(9)近似分解为

$$\ln P(D) = L(Q) + K_L(Q|P). \quad (10)$$

式中:

$$L(Q) = \int Q(\theta) \ln \frac{P(D, \theta)}{Q(\theta)} d\theta. \quad (11)$$

引入的分布 $Q(\theta)$ 是 $P(D, \theta)$ 和后验分布 $K_L(Q|P)$ 之间的 Kullback-Leibler 距离差,表达式如式(12)

$$K_L(Q|P) = - \int Q(\theta) \ln \frac{P(\theta|D)}{Q(\theta)} d\theta. \quad (12)$$

式中: $K_L(Q|P) \geq 0$. 式(10)的右半部分是 Q 的独立分布,因此最大化 $L(Q)$ 等同于最小化 $K_L(Q|P)$.

综合式(11)和式(12),则 $Q(\theta)$ 就是 $P(\theta|D)$ 的近似边缘概率分布. 通过该近似的方法,只要选择一个合适的分布 Q , $L(Q)$ 的值即便在传统证据模型函数未知的情况下也可以通过简单的计算得到. 因此改进型 RVM 的方法就是选择一个形式较为简单的 $Q(\theta)$, 使得 $L(Q)$ 的计算较为简单,并且同时保证 $L(Q)$ 的形式足够灵活.

根据式(9)且考虑到参数 θ 的变形形式 $\{\theta_i\}$, 则 $Q(\theta)$ 可以定义为

$$Q(\theta) = \prod_i Q_i(\theta_i).$$

式中:

$$Q_i(\theta_i) = \frac{\exp\langle \ln P(D, \theta) \rangle_{k \neq i}}{\int \exp\langle \ln P(D, \theta) \rangle_{k \neq i} d\theta_i}. \quad (13)$$

式中: $\langle \cdot \rangle_{k \neq i}$ 定义为对于 $Q_k(\theta_k)$ 分布 (其中 $k \neq i$) 的期望. 可以证明,如果概率模型中的每个因素 $Q_i(\theta_i)$ 可以用一种带有节点的有向循环图来表示,那么 $Q_i(\theta_i)$ 的解就仅仅取决于 Q 分布的变量形式^[20].

式(13)的右半部分取决于 $Q_{k \neq i}$ 的任意时刻,所以式(13)是 $Q_i(\theta_i)$ 的一种特殊的求解表达式. 又因为 $Q_i(\theta_i)$ 是共轭的条件分布,因此可以通过计算得到在所需时刻的标准条件分布. 然后通过初始化,循环使用式(13),迭代更新变量,最终得到 Q 的所有分布.

3.1.2 改进型相关向量机分类算法

输入数据边缘概率的对数形式为

$$\ln P(T|X) = \ln \int P(T|X, \mathbf{w}) P(\mathbf{w}|\alpha) P(\alpha) d\mathbf{w} d\alpha.$$

利用之前处理结论和近似的分解变量形式 $Q_w(\mathbf{w})$ $Q_\alpha(\alpha)$, 并结合式(11), 将上式近似为

$$\ln P(T|X) \geq \int \int Q_w(\mathbf{w}) Q_\alpha(\alpha) \ln \left\{ \frac{P(T|X, \mathbf{w}) P(\mathbf{w}|\alpha) P(\alpha)}{Q_w(\mathbf{w}) Q_\alpha(\alpha)} \right\} d\mathbf{w} d\alpha. \quad (14)$$

式(14)的积分计算会增加运算的复杂度,因此根据 Jaakkola 等^[21]的分析方法,引入

$$\sigma(y)^t [1 - \sigma(y)]^{1-t} = \sigma(z) \geq \sigma(\xi) \exp\left(\frac{z - \xi}{2} - \lambda(\xi)(z^2 - \xi^2)\right). \quad (15)$$

式中: $z = (2t - 1)y$, $\lambda(\xi) = (1/4\xi) \tanh(\xi/2)$. ξ 是一个可变的参数,当等号成立时,可以得到 $\xi = z$. 将式(15)代入相关向量机模型中,可以得到

$$P(T|X, \mathbf{w}) \geq F(T, X, \mathbf{w}, \xi) = \prod_{n=1}^N \sigma(\xi_n) \exp\left(\frac{z_n - \xi_n}{2} - \lambda(\xi_n)(z_n^2 - \xi_n^2)\right). \quad (16)$$

式中: $z_n = (2t_n - 1)^T \phi_n$, 且 $P(T|X, \mathbf{w})/F(T, X, \mathbf{w}, \xi) \geq 1$ 等同于 $\ln P(T|X, \mathbf{w})/F(T, X, \mathbf{w}, \xi) \geq 0$. 将式(16)代入式(14)中, 可得

$$\ln P(T|X) \geq L = \iint d\mathbf{w} d\boldsymbol{\alpha} Q_{\mathbf{w}}(\mathbf{w}) Q_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}) \cdot \ln \left(\frac{F(T, X, \mathbf{w}) P(\mathbf{w} | \boldsymbol{\alpha}) P(\boldsymbol{\alpha})}{Q_{\mathbf{w}}(\mathbf{w}) Q_{\boldsymbol{\alpha}}(\boldsymbol{\alpha})} \right). \quad (17)$$

优化式(17)右边的函数 $Q_{\mathbf{w}}(\mathbf{w})$ 和 $Q_{\boldsymbol{\alpha}}(\boldsymbol{\alpha})$ 以及参数 $\xi = \xi_n$. 优化的结果为 $Q_{\mathbf{w}}(\mathbf{w})$ 产生了一个一般的正态分布形式:

$$Q_{\mathbf{w}}(\mathbf{w}) = N(\mathbf{w} | \mathbf{m}, \mathbf{S}),$$

$$\mathbf{S} = (\mathbf{A} + 2 \sum_{n=1}^N \lambda(\xi_n) \phi_n \phi_n^T)^{-1},$$

$$\mathbf{m} = \frac{1}{2} \mathbf{S} (\sum_{n=1}^N (2t_n - 1) \phi_n).$$

式中: $\mathbf{A} = \text{diag}(a_m)$. 相对而言, 优化的结果使得 $Q_{\boldsymbol{\alpha}}(\boldsymbol{\alpha})$ 是服从伽玛分布的.

$$Q_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}) = \prod_{m=0}^N \Gamma(a_m | \tilde{a}, \tilde{b}_m),$$

$$\tilde{a} = a + \frac{1}{2}, \tilde{b}_m = b + \frac{1}{2} \langle \omega_m^2 \rangle.$$

最后, 对于变形参数 ξ_n 进行最小化, 该参数的值可通过 $\xi_n^2 = \phi_n^T \langle \mathbf{w} \mathbf{w}^T \rangle \phi_n$ 求得.

同样, 通过式(17)的右半部分, 对 L 的值可以通过下式计算而得

$$L = \langle \ln F \rangle + \langle \ln P(\mathbf{w} | \boldsymbol{\alpha}) \rangle + \langle \ln P(\boldsymbol{\alpha}) \rangle -$$

$$\langle \ln Q_{\mathbf{w}}(\mathbf{w}) \rangle - \langle Q_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}) \rangle.$$

式中:

$$\langle \ln F \rangle = \sum_{n=1}^N \{ \ln \sigma(\xi_n) + \frac{1}{2} (2t_n - 1) \langle \mathbf{w}^T \rangle \phi_n - \frac{1}{2} \xi_n^2 - \lambda(\xi_n) \langle \phi_n^T \langle \mathbf{w} \mathbf{w}^T \rangle \phi_n - \xi_n^2 \} ,$$

$$\langle \ln P(\mathbf{w} | \boldsymbol{\alpha}) \rangle = -\frac{1}{2} \sum_{m=0}^N \langle \alpha_m \rangle \langle \omega_m^2 \rangle + \frac{1}{2} \sum_{m=0}^N \langle \ln \alpha_m \rangle - \frac{(N+1)}{2} \ln(2\pi),$$

$$\langle \ln P(\boldsymbol{\alpha}) \rangle = \sum_{m=0}^N \{ -\tilde{b} \tilde{a} + (\alpha - 1) (\psi(\tilde{a}) - \ln \tilde{b}) + \alpha \ln \tilde{b} - \ln \Gamma(\tilde{a}) \} ,$$

$$- \langle \ln Q_{\mathbf{w}}(\mathbf{w}) \rangle = \frac{N+1}{2} (1 + \ln 2\pi) + \frac{1}{2} \ln |\mathbf{S}| ,$$

$$- \langle \ln Q_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}) \rangle = \sum_{m=0}^N \{ -(\tilde{a}_m - 1) \psi(\tilde{a}_m) - \ln \tilde{b}_m + \tilde{a}_m + \ln \Gamma(\tilde{a}_m) \} .$$

对于新的输入的预测, 可以通过现有的训练好的权重模型计算求得 $P(t|x, \langle \mathbf{w} \rangle)$, 即可进行分类.

3.1.3 实验数据

这是一种对于 RVM 算法本身的改进, 将改进 RVM 算法应用于高光谱图像分类中, 当训练样本较小时, 其分类精度可以高于 RVM 以及 SVM 分类. 当训练样本较大时, 其训练时间以及测试时间会有所减低, 结果如表 3、4.

表 3 各类别分类精度

Table 3 The accuracy of classification %			
类别	SVM	RVM	VRVM
玉米 1	77.61	77.40	77.68
玉米 2	75.54	77.10	77.70
大豆 1	83.67	83.57	83.78
大豆 2	58.75	59.40	60.37
乔木	91.19	92.19	92.35
总体	74.06	74.61	75.14

表 4 分类效率

Table 4 The efficiency of classification s						
样本数 (所占比例)	SVM		RVM		VRVM	
	训练 时间	测试 时间	训练 时间	测试 时间	训练 时间	测试 时间
350(5%)	2.70	3.2	14	0.3	12	0.3
700(10%)	3.00	3.6	43	0.4	40	0.4
1050(15%)	5.30	5.1	124	1.2	114	1.2
1400(20%)	8.40	12.3	809	2.4	609	2.3
3499(50%)	21.47	27.6	1806	3.9	1430	3.8

从表 3、4 中可以看出对于 5 种地物的分类精度和总体分类精度, VRVM 最高, RVM 次之, 而 SVM 相对较低; VRVM 和 RVM 的测试时间短, 训练时间较长, 而 SVM 测试时间长, 训练时间相对较短.

3.2 相关向量机的多类分类技术

类似于 SVM, RVM 最初的设计也是针对 1-a-1 分类问题, 然而高光谱图像地物类别较多, 品种复杂, 因此需要发展多类分类技术. Wu 等在 2009 年提出的 RVM 多类分类技术, 有效地改善了 RVM 在多类分类中的效果^[22]. 算法流程如图 1 所示.

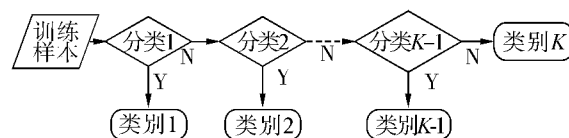


图 1 算法流程

Fig. 1 The algorithm flow chart

该算法主要是利用逐层分级的思想进行, 首先

判断所有的数据是否属于第 1 类,其次判断是否属于第 2 类,以此类推,最后得出结果.利用高光谱数据进行多类地物分类,分析比较下列分类方法:One-against-rest (OAR)、One-against-one (OAO)、directed acyclic graph method(DAG)、binary tree based on class distance(CDBT)、binary tree based on class distance and class distribution (CDDBT)等在分类精度和计算效率方面的性能^[23-24],实验结果如表 5 和表 6 所示.

表 5 多类分类效果

Table 5 The effect of multi-classification

数据源	样本数量		RVM			
	训练样本	测试样本	分类精度/%	相关向量个数	训练时间/s	测试时间/s
Pima	100	668	79.1	3	0.64	0.001
	300	468	80.1	11	4.40	0.001
	500	168	80.7	31	14.70	0.002
	250	1 000	90.3	5	2.07	0.003
Ripley	500	750	90.8	7	11.10	0.001
	1 000	250	91.6	9	68.70	0.001

表 6 分类效果比较

Table 6 The comparison of effect of multi-classification

数据源	分类方法	训练时间/s	测试时间/s	分类准确度/%
Iris	OAO	1.31	0.013	96.7
	OAR	1.96	0.015	95.8
	DAG	1.32	0.010	97.8
	CDBT	1.29	0.012	98
	CDDBT	1.28	0.012	99
Shuttle	OAO	11.20	0.021	97.5
	OAR	24.68	0.015	95.5
	DAG	11.50	0.016	98.5
	CDBT	16.50	0.014	99.5
	CDDBT	16.40	0.013	99.8
Satimage	OAO	43.54	0.026	90.2
	OAR	54.36	0.014	89.6
	DAG	43.61	0.02	90.3
	CDBT	37.44	0.024	90.4
	CDDBT	37.64	0.021	90.6

从表 5 中可以看出,随着地物训练样本的增加,

RVM 的分类精度会有所提高,但相关向量个数明显增多,训练时间也变长.表 6 说明采用 5 种方法对 3 类地物进行分类,CDDBT 方法在训练时间和测试时间相当情况下,分类精度最高.

3.3 应用改进型核函数

RVM 也是一种基于核方法的机器学习理论,因此核函数的效果好坏也在一定程度上影响了 RVM 分类效果,尤其对于数据混叠较为严重的高光谱数据,如果核函数能将混叠数据较为合理地投影出来,那么对于分类精度以及分类效率都有着至关重要的作用,因此应用改进型核函数也是提高 RVM 分类效果的一种途径.

Camps 等在 2007 年提出了利用 Mahalanobis 核函数代替 RBF 核函数,应用于高光谱图像分类^[25]. Mahalanobis 核函数表达式如下:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{1}{2\sigma^2}(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{Q}^{-1}(\mathbf{x}_i - \mathbf{x}_j)\right).$$

Mahalanobis 核函数的优点在于对于输入空间的每一个轴空间都有其独特的特征参数,这样在分类时,其边界更加合理.在高光谱图像分类中,图像数据为 DAIS7915 PAVIA IMAGE, RBF-RVM 分类精度为 80.33%,而 MK-RVM 的分类精度为 81.67%.可以看出 Mahalanobis 核函数的分类精度更高,是一种较为理想的改进方案.

詹环等在 2008 年提出利用一种小波核函数的相关向量机分类方法^[26].该方法选取墨西哥草帽小波函数来构造平移不变小波核函数,即

$$\psi(x) = (1 - x^2) \exp\left(-\frac{1}{2}x^2\right),$$

并且满足平移不变核函数的墨西哥草帽小波核函数为

$$K(x, z) = \prod_{i=1}^d \left(1 - \left(\frac{x_i - z_i}{a_i}\right)^2\right) \exp\left(-\frac{1}{2}\left(\frac{x_i - z_i}{a_i}\right)^2\right).$$

将小波核函数应用于 SINC(X) 函数回归问题中,实验结果表明,其准确率高于 RBF-RVM 回归方法,训练时间也相对较短.

本文作者于 2011 年将小波核函数相关向量机应用于高光谱图像分类^[27],该方法应用的小波核函数与文献[26]中的略有不同,其表达式如下:

$$K(x, z) = \prod_{i=1}^d \left(\cos\left(1.75 \times \frac{(x_i - z_i)}{a_i}\right) \exp\left(-\frac{1}{2}\left(\frac{x_i - z_i}{a_i}\right)^2\right)\right).$$

实验结果表明其分类精度比 RVM 高 2.73%,获得了较好的效果.

4 结 论

本文对国内外相关向量机研究现状进行了全面

的综述. 通过研究相关向量机的模型、回归算法以及分类算法, 有效地将相关向量机应用到多个领域当中. 尤其着重介绍了相关向量机分类方法在高光谱图像分类中的应用.

高光谱数据的特点决定了相关向量机非常适合其应用, 无论是检测还是分类问题, 相关向量机都有着较好的表现. 高光谱图像分类的评价标准有许多种, 但是最主要的还是分类精度和分类效率. 分类精度主要是体现了该方法对高光谱数据的分类准确度, 在这方面, 相关向量机可以达到较高的分类精度, 满足分类的要求; 分类效率主要是体现该方法在分类过程中的空间复杂度以及时间复杂度, 相关向量机的空间复杂度非常低, 是支持向量机的 15% 左右, 但其时间复杂度却比较高, 尤其体现在大训练样本的训练时间上. 因此对于如何减少训练时间, 提高训练效率也是今后进一步研究相关向量机的重点方向之一.

在现有的改进方法中, 主要分为两大类, 一是对算法本身的改进, 也叫改进型相关向量机; 另外一类是对算法中应用到的核函数进行改进. 2 种改进方法都可以提高相关向量机分类的效率, 并且保证了较高的分类精度. 对于整个高光谱图像分类研究领域, 如何进一步挖掘各种方法间的优点, 拓展其应用范围是该领域未来发展的重点.

目前相关向量机的研究已经取得了一定进展, 但其方法还是存在着一些不足, 如何改进这些算法, 突破制约技术发展瓶颈, 是未来其理论研究的目标. 本文仅对其进行综合分析论述, 以达到抛砖引玉之功效.

参考文献:

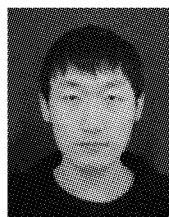
- [1] TIPPING M. Sparse Bayesian learning and the relevance vector machine[J]. *Machine Learning Research*, 2001, 6(1): 211-244.
- [2] DUAN Qing, ZHAO Jianguo, MA Yan. RVM and SVM for classification in transient stability assessment[C]//2010 Asia-Pacific Power and Energy Engineering Conference (APPEEC). Chengdu, China, 2010: 1-4.
- [3] BILGIN G, ERTURK S, YILDIRIM T. Segmentation of hyperspectral images via subtractive clustering and cluster validation using one-class support vector machines[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2011, 49(8): 2936-2944.
- [4] CAMPS V G, MARTINEZ R M. Nonlinear system identification with composite relevance vector machines[J]. *IEEE Signal Processing Letters*, 2007, 14(4): 279-282.
- [5] DEMIR B, ERTURK S. Hyperspectral data classification using RVM with pre-segmentation and RANSAC[C]//2007 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Barcelona, Spain, 2007: 1763-1766.
- [6] BISHOP C M, TIPPING M E. Variational relevance vector machines[C]//Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence. Catalina Island, USA, 2010: 46-53.
- [7] LUKIC A S, WERNICK M N, TZIKAS D G, et al. Bayesian kernel methods for analysis of functional neuroimages[J]. *Medical Imaging*, 2007, 26(12): 1613-1624.
- [8] GHOLAMI B, HADDAD W M, TANNENBAUM A R. Relevance vector machine learning for neonate pain intensity assessment using digital imaging[J]. *Biomedical Engineering*, 2010, 57(6): 1457-1466.
- [9] AGARWAL A, TRIGGS B. Recovering 3D human pose from monocular images[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(1): 44-58.
- [10] GUO Feng, QIAN Gang. 3D human motion tracking using manifold learning[C]//2007 IEEE International Conference on Image Processing. San Antonio, USA, 2007: 357-360.
- [11] SEDAI S, BENNAMOUN M, HUYNH D. Context-based appearance descriptor for 3D human pose estimation from monocular images[C]//Proceedings of the 2009 Digital Image Computing: Techniques and Applications. Melbourne, Australia, 2009: 484-491.
- [12] GUO Weiwei, PATRAS I. Discriminative 3D human pose estimation from monocular images via topological preserving hierarchical affinity clustering[C]//2009 IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops). Kyoto, Japan, 2009: 9-15.
- [13] OUYANG Yi, ZHANG Sanyuan. RVM-based classification of the network video surveillance system[C]//WASE International Conference on Information Engineering. Taiyuan, China, 2009: 144-148.
- [14] OLIVERI G, ROCCA P, MASSA A. A Bayesian compressive sampling based inversion for imaging sparse scatters[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2011, 49(10): 3993-4006.
- [15] LI Xiangru, LI Xiaoming, CHENG Xuezheng. Rejecting mismatches between fish-eye camera images by RVM[C]//5th International Conference on Natural Computation. Tianjin, China, 2009: 293-295.
- [16] WANG Xiaofei, ZHANG Junping, ZHANG Ye. Registration of remote sensing images based on the relevance vector machine[C]//5th International Conference on Natural Computation. Tianjin, China, 2009: 547-551.
- [17] DEMIR B, ERTURK S. Hyperspectral image classification using relevance vector machines[J]. *IEEE Geoscience and Remote Sensing Letters*, 2007, 4(4): 586-590.

- [18] MIANJI F A, ZHANG Y. Robust hyperspectral classification using relevance vector machine [J]. IEEE Transactions on Geoscience and Remote Sensing, 2011, 49(6): 2100-2112.
- [19] CHYKEYUK K, CLIFTON D A, NOBLE J A. Feature extraction and wall motion classification of 2D stress echocardiography with relevance vector machines [C]//2011 IEEE International Symposium on Biomedical Imaging. Chicago, USA, 2011: 677-680.
- [20] RAM R, CHETTY M. A Markov-blanket-based model for gene regulatory network inference [J]. IEEE Transactions on Computational Biology and Bioinformatics, 2011, 8(2): 353-367.
- [21] JAAKKOLA T S, JORDAN M I. Bayesian parameter estimation through variational methods [J]. Statistics and Computing, 2000, 10(1): 25-37.
- [22] WU Huilan, LIU Guodong, PU Zhaobang. Multi-class image recognition based on relevance vector machine [C]//2009 International Workshop on Intelligent Systems and Applications. Wuhan, China, 2009: 1-4.
- [23] 陈世哲. 微电子产品视觉检测中关键技术研究 [D]. 哈尔滨: 哈尔滨工业大学, 2006: 5-24.
CHENG Shizhe. Research on the key techniques of micro-electronic products vision inspection [D]. Harbin: Harbin Institute of Technology, 2006: 5-24.
- [24] 唐发明. 基于统计学习理论的支持向量机算法研究 [D]. 武汉: 华中科技大学, 2005: 67-81.
TANG Faming. Study of support vector machines algorithm based on statistical learning theory [D]. Wuhuan: Huazhong University of Science and Technology, 2005: 67-81.
- [25] CAMPS V G, RODRIGOG A, MUOZ M J. Hyperspectral image classification with Mahalanobis relevance vector machines [C]//2007 IEEE Geoscience and Remote Sensing Symposium. Barcelona, Spain, 2007: 3802-3805.
- [26] 詹环, 王雪亭. 小波相关向量机 [J]. 五邑大学学报: 自然科学版, 2008, 22(1): 58-62.
ZHAN Huan, WANG Xueting. The wavelet relevance vector machine [J]. Journal of Wuyi University: Natural Science Edition, 2008, 22(1): 58-62.
- [27] ZHAO Chunhui, ZHANG Yi, WANG Ying. RVM classification of hyperspectral image based on wavelet kernel function [C]//2011 International Conference on Multimedia Technology. Hangzhou, China, 2011: 3647-3650.

作者简介:



赵春晖,男,1965年生,教授、博士生导师。主要研究方向为智能信息处理、图像处理。获省部级科技奖5项,出版著作3部,发表学术论文430余篇。



张焱,男,1987年生,硕士研究生。主要研究方向为高光谱图像分类技术,发表学术论文5篇。