支持向量机理论与应用研究综述

张博洋
（北京交通大学 计算机与信息技术学院，北京 100044）

摘 要：文章研究支持向量机技术，分析支持向量机的运行基本原理，研究支持向量机技术中的多类问题和选择核函数，并且从人脸检测、文本分类、处理图像、识别手写字符等方面合理分析支持向量机，为进一步应用和发展支持向量机技术提供依据和保证。

关键词：支持向量机；理论；应用；综述

支持向量机（Support Vector Machine，SVM）是通过分析统计理论基础上形成的模式分类方法。上述方式在实际实施的时候，依据最小化风险的基本原则有效增加系统的泛化作用，也是一种为了得到最小误差实施的决策有限训练样中的独立测试集，能够适当分析和解决学习问题、选择模型问题、维数灾难问题等。研究SVM主要就是分析支持向量机自身性质，此外还分析提高应用支持向量机的广度和深度，在文本分类、模式分类、分析回归、基因分类、识别手写字符、处理图像等方面得到应用。

1 支持向量机的原理分析

1.1 结构风险最小化

依据能够应用的有限信息样本，不能合理计算分析期望风险，所以，传统方式应用主要是经验风险最小化（ERM）标准，利用样本对风险进行定义：

$$R_{\text{emp}}=\frac{1}{n}\sum_{i=1}^{n}L(y_i,f(x_i,w))$$

基于统计学理论分析函数集以及实际经验风险的关系，也就是推广性的界。总结分析上述问题，能够得到实际风险和经验风险之间概率1-符合以下条件关系：

$$R(\omega)\leqslant R_{\text{emp}}(\omega)+\sqrt{\frac{h(\ln(2l/h+1)-\ln(\eta/4))}{l}}$$

其中l是训练集样本数，h为函数集VC维，体现高低复杂性，从上述理论基础可以发现，通过两部分构成学习机实际风险：一是置信范围；二是经验风险也就是训练误差。机器学习的时候不仅需要经验风险，还要尽可能缩小VC维符合置信范围，保证能够获得实际比较小的风险，实际上就是结构风险最小化SRM（Structure Risk Minimization）原则[1]。

1.2 支持向量机

支持向量机实际上从最优化线性分析分类超平面形成技术，分析情况的时候，最基本理念就是2类线性。支持向量机学习的主要目的就是能够发现最优超平面，不仅需要正确分开2类样本，还能够具备最大的分类间隔。分类间隔就是说距离超平面最近的2类分类样本，并且可以与2类分类平面间距平行。分析线性分类问题，假设T是训练集：

$\{(x_1,y_2),\ldots,(x_1,y_1)\}\in(X\times Y)^1$，

其中$x_i\in x=R^n$，$yi\in y=\{-1,1\}$，i=1,2，…，l。

假设（ω x）+b=0是超平面，超平面和训练集之间的集合间距就是$1/\|\omega\|$。可以通过以下方式找到最大间隔超平面问题中的原始优化问题：

$$\min_{w\ b}\quad \tau(\omega)=1/2\|\omega\|^2,$$

$$\text{S.t.}\quad y_i((\omega x_i)+b)\geqslant 1,\ i=1,\ldots,l$$

利用Wolfe对偶定理，能够等价原始最优化问题得到相关对偶问题：

$$\min_{\alpha}\quad \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}y_iy_j\alpha_j\alpha_j\alpha_j(x_i\cdot x_j)-\sum_{j=1}^{l}\alpha_j$$

$$s.t.\quad \sum_{i=1}^{l}y_i\alpha_i=0,$$

$\alpha\geqslant 0$，i=1，…，l，

此时能够得到最优解就是 $\alpha^*=(\alpha_1^*,a_2^*,\cdots,\alpha_l^*)^T$；

引入松弛变量$\varepsilon=(\varepsilon_1,\cdots,\varepsilon_l)^T$以后能够得到等价对偶问题：

$$\min_{\alpha}\quad \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}y_iy_j\alpha_j\alpha_i\alpha_j(x_i\cdot x_j)-\sum_{j=1}^{l}\alpha_j$$

$$s.t.\quad \sum_{i=1}^{l}y_i\alpha_i=0,$$

$$0\leqslant\alpha\leqslant C,\ i=1,\cdots,l.$$

其中，C（C>0）是惩罚因子。

1.3 核函数

很多不可分线性问题，在某个高位特征空间中合理筛选符合分类样本情况的非线性变换映射，确保能够得到高维空间目标样本线性可分。依据上述方式进行计算的时候，仅仅只是计算训练样本内积，需要依据原空间来实现函数，不需要分析变换形式，依据泛函基本理论，一种核函数K（x，x'）需要充分符合Mercer，与某空间变化内积对应。

假设对应变化核函数是K（x，x'），K（x，x'）=（Φ（x），Φ（x'）），依据之前分析的原始对偶问题，得到相应的决策函数就是：

$f(x)=\text{sgn}(\sum_{i=1}^{l}\alpha i^*yK(x,xi)+b)$，

有3种常见的核函数，一是径向有机函数（RBF）：

$$K(x\cdot x')=\text{Exp}(-\frac{|x-x'|^2}{\sigma^2})$$

二是多项式核函数：

$$k(x,x')=((x\cdot x')+1)^q$$

作者简介：张博洋（1990-），男，天津，硕士研究生；研究方向：数据挖掘。

?1994-2018 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

三是内积为Sigmoid函数：

$K(x, x')=\tanh(v(x, x')+c)$，这里 $v>0$，$C<0$.

1.4 多分类问题

支持向量机仅仅只能够分类2种类别，大部分情况下可以扩展2类支持向量机，形成多类别分类器。现阶段，从2个方面分析和研究：一是，更改支持向量机中最原始分类和最优化问题，计算相应的分类决策函数。二是，把多类问题合理地变为2类问题，组合2类分类器，达到多类分类的目的[2]。

1.4.1 一类对余类

在每一个类别i都构造2类分类器ci，能够分析n类问题，此时需要支持向量机数量n个，对Ci进行训练的时候，正例样本是i类别样本，负例样本是n-1的样本。

1.4.2 成对分类

每2个类别能够形成2类分类器，可以适当解决n类问题，此时需要支持向量机数量是n（n-1）/2，通过n（n-1）/2分类器来决定测试样本类别。

1.4.3 层次支持向量机

层次分类方式可以对所有类别都进行分类，进一步规划和分析2个次级子类，反复循环，得到单独类别为止。

由于不断增多样本类别，急剧增加成对分类方式中的分类器数量，会在一定程度上提高计算复杂度。一类对余类分析方式属于不对称2类问题，并且能够提高样本数量，增加计算复杂度。层次法如果具备邻近正态树层次结构，存在比较理想的训练速度[3]。

2 支持向量机的应用研究

2.1 识别手写阿拉伯数字

在实际应用支持向量的时候，最重要的是手写数字识别问题。上述问题属于多类问题。相关专家学者在研究2类问题的SVM前提下，形成了能够处理多类问题的相关SVM，其中主要核函数就是sigmoid核函数、径向基核函数、多项式核函数。不但可以支持比较其他分类和支持向量机，还能够支持比较不同形式的SVM，经过大量实践可以发现，存在很大优势[4]。

2.2 检测人脸

相关学者和专家经过不断研究和分析以后形成以层次结构形式的支持向量机分类器，由一个非线性和线性支持向量机构成，上述方式不但具备比较低的误差率和比较高的检测率，还存在比较快的速度。此后，人们利用SVM方式来有效判断人脸姿态，并且合理分为6个类别，手工标定方式在多姿态人脸库中发现测试样本和训练样本集，在SVM基础上的训练集姿态分类器，可以降低到1.67%的错误率。在支持向量机和小波技术上形成的识别人脸技术，压缩提取人脸特征的时候应用小波技术，然后结合支持向量机技术和邻近分类器进行分类，确保具备比较好的鲁棒性能和分类性能[5]。

2.3 文本分类

文本分类主要就是在一定的分类体系中，依据文本相关类别和实际内容记性分析。本文分类是一个十分重要的自然语言处理，大部分都是应用在邮件分类、过滤信息、自动文摘、检索信息等中。相关学者和专家建立一种能够支持向量机进行在线学习的文本分类RBF，此方式可以在一定程度上提高学习增量。在分析了支持向量机文块组织的时候，可以把其当做分类问题，从而依据支持向量机解决问题[6]。

2.4 其他应用

支持向量机具备一定的优越性，已经得到大量应用。专家学者提出了支持向量机基础上的水印算法，在数字水印中合理应用支持向量机，存在十分良好的应用效果。并且入侵监测系统已经是十分重要的网络安全技术之一，在分析入侵检测系统的时候，主要应用的就是SVM基础上的主动学习算法，可以在一定程度上降低学习样本，能够增加入侵监测系统整体分类性能。在处理图像中迷糊噪音的时候，依据SVM模糊推理方式形成的一种噪音检测系统，上述方式能够合理除去检测中的噪音，适当保存图像相关信息。在分析混合气体定量和多维光谱定性的时候，不能应用同一种方式来定性和定量分析组合气体吸收普线重叠、输入光谱的维数，训练样本数目有限，在分析地混合气体多维光谱的时候应用支持向量机，依据核函数有效把重叠光谱数据变为支持向量机回归模型，此时可以定量分析混合气体的组分浓度以及定性分析种类[7]。

3 结语

本文主要分析了基于支持向量机理论与应用研究，着重分析了支持向量机基本原理，从人脸识别、文本分类、手写识别3方面来分析应用支持向量机。

[参考文献]

[1]姚潇，余乐安.模糊近似支持向量机模型及其在信用风险评估中的应用[J].系统工程理论与实践，2012(3):549-554.
[2]崔长春，刘文林，郑俊哲，等.支持向量机理论与应用[J].沈阳工程学院学报：自然科学版，2010(2):170-172.
[3]信晶，孙保民，肖海平，等.应用支持向量机监测电站锅炉受热面积灰研究[J].中国电机工程学报，2013(5):3，21-27.
[4]冯能山，廖志良，熊金志，等.支持向量机用于图像水印技术的研究综述[J].信息系统工程，2012(11):131-133.
[5]金焱，胡云安，黄隽，等.支持向量机回归在电子器件易损性评估中的应用[J].强激光与粒子束，2012(9):2145-2150.
[6]纪昌明，周婷，向腾飞，等.基于网格搜索和交叉验证的支持向量机在梯级水电系统隐随机调度中的应用[J].电力自动化设备，2014(3):125-131.
[7]崔万照，朱长纯，保文星，等.最小二乘小波支持向量机在非线性系统辨识中的应用[J].西安交通大学学报，2011(6):562-565，586.

Overview of Support Vector Machine Theory and Application Research

Zhang Boyang
（School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China）

Abstract: The paper studies support vector machine technology, analyzes the basic principle of support vector machine, studies the multi class problems and selection kernel function in support vector machine technology.And the paper reasonably analyzes support vector machine from the human face detection, text classification, image processing and recognition of handwritten characters and so on, provides the basis and guarantee for further application and development of support vector machine technology

Key words: support vector machine; theory; application; overview

?1994-2018 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net