

# SVM 集成研究与应用

王晓丹<sup>1</sup>, 高晓峰<sup>2</sup>, 姚旭<sup>1</sup>, 雷蕾<sup>1</sup>

(1. 空军工程大学导弹学院, 陕西三原, 713800; 2. 武警陕西省总队后勤部, 陕西西安, 710054)

**摘要** 集成学习是机器学习的重要研究方向之一, SVM 集成近年来已经受到国内外很多从事机器学习、统计学习的研究者们的重视, 并使得该领域成为了一个相当活跃的研究热点。对近年来 SVM 集成的研究与应用进行了综述, 讨论了 SVM 集成需要解决的基本问题; 讨论分析了构造差异性大的集成成员 SVM 的方法、有效的集成结论生成方法、SVM 集成的典型应用; 指出了目前存在的问题、以及几个重要的研究方向。

**关键词** 支持向量机; 集成学习; 差异性

**DOI** 10.3969/j.issn.1009-3516.2012.02.018

**中图分类号** TP181 **文献标识码** A **文章编号** 1009-3516(2012)02-0084-06

支持向量机(Support Vector Machine, SVM)是在统计学习理论<sup>[1]</sup>基础上发展起来的一种新的机器学习方法, 是目前泛化能力最强的技术之一。统计学习理论对有限样本情况下模式识别中的一些根本性问题进行了系统的理论研究, 很大程度上解决了模型选择与过学习问题、非线性和维数灾难问题、局部极小点等问题, 引起了研究者的广泛关注。SVM 在解决有限样本、非线性及高维模式分类问题中表现出了许多特有的优势, 利用 SVM 设计泛化能力好的模式分类器是一种非常适合的途径<sup>[2]</sup>, 目前 SVM 在模式分类、回归分析、函数估计等领域得到了广泛的应用。近年来, 对 SVM 的研究主要集中在对 SVM 本身性质的研究和完善以及加大 SVM 应用研究的深度和广度两方面。但是, SVM 在用于解决分类预测等问题时, 仍存在以下问题需要解决: 核函数和参数的选择、核函数及参数的构造和选择缺乏理论指导; SVM 用于多类分类问题时的有效算法、多类支持向量分类器的优化设计; 有效的增量学习算法等<sup>[3-4]</sup>。SVM 自身存在的缺点降低了 SVM 的稳定性和泛化能力, 因此在实际应用中, SVM 的分类效果可能不如预期的好。一个自然的想法就是把 SVM 作为集成中的成员分类器, 即进行 SVM 集成以提高泛化能力<sup>[5-7]</sup>。

## 1 SVM 集成需要解决的基本问题

SVM 集成是集成学习的一种具体实现, 参与集成的个体分类器都是 SVM, 属同质集成。Schapire 针对弱学习算法与强学习算法二者之间的等价性问题给出了构造性证明<sup>[8]</sup>, 由此奠定了集成学习的理论基础, 一个学习方法可以提升为强可学习的充要条件是其为弱可学习。Kearns 和 Valiant 也证明只要有足够的数据, 弱学习算法就能通过集成的方式生成任意高精度的估计<sup>[9]</sup>。

对于 SVM 集成需要解决 2 个问题: ①如何生成集成中的各 SVM, 即构成集成成员个体的各弱分类器需要满足什么条件, 才能够有利于集成后系统性能的提高②如何生成集成的结论, 即怎样将多个 SVM 的输出进行结合, 才能成为性能良好的强分类器。所以, 对 SVM 集成的实现方法研究和理论分析都包含 2 个方面的内容: 构成集成的各成员 SVM 的生成和集成结论生成方法。

### 1.1 成员 SVM 的生成

参与集成的成员分类器应该满足一定的要求, 否则可能对集成结果没有贡献, 甚至可能降低集成效果。

\* 收稿日期: 2011-10-09

基金项目: 国家自然科学基金资助项目(60975026); 陕西省自然科学基金计划资助项目(2007F19)

作者简介: 王晓丹(1966-), 女, 陕西汉中, 教授, 博士生导师, 主要从事机器学习、模式识别、智能信息处理等研究。

E-mail: afeu\_w@yahoo.com.cn

通常,成员分类器应满足如下2个原则:①成员分类器的精确度要高;②成员分类器应具有差异(Diversity)或称多样性。

差异性是影响集成效果的另一个重要因素。因为用对相同样本犯同样分类错误的成员分类器进行集成,对集成分类器的精确度不能带来任何提高。通常,成员分类器越精确、差异性越大,那么集成效果越好。

Krogh 和 Vedelsby<sup>[10]</sup>以回归学习的集成推导出重要的集成学习的泛化误差公式,这个公式对于分类器的集成有着同样的意义。对于  $T$  个学习机,它们集成的误差由  $E = \bar{E} - \bar{A}$  决定,式中: $\bar{E}$  为  $T$  个学习机的绝对误差的加权平均; $\bar{A}$  为  $T$  个学习机相对于集成的误差的加权平均。 $\bar{E}$  指示出学习机固有的误差, $\bar{A}$  指示出这些学习机之间的差异。该式表明:要获得好的集成就需要降低成员学习机的误差并增加学习机间的差异。

在生成集成成员 SVM 方面,Bagging、Boosting 和 AdaBoost 等是研究和应用最多、影响最大的重要技术。不同的集成方法以不同的方式引入差异性,如 Bagging 方法通过在不同的训练集上训练分类器个体得到差异性,其集成成员间的差异性是通过 Bootstrap 重取样技术获得的,或者说它是通过训练样本的随机性及独立性来提供集成的差异性;Boosting 方法使用了确定性方法保证训练集中含有更难分类的样本以形成分类器之间的差异,它是通过不断修改样本分布产生差异性。

## 1.2 集成结论的生成

在获得了一组成员 SVM 后,SVM 集成需要解决的问题是如何生成集成结论,即成员 SVM 如何参与集成结论生成,以及如何对成员 SVM 的输出进行处理,以取得较理想的集成输出结果。

SVM 集成结论的生成方法可以分为2类:一类是组合法,其集成输出是各分类器输出的某种组合<sup>[11-12]</sup>,如投票法(简单投票,加权投票)、最小二乘加权法、叠加法或分级法(构建上层决策分类器)等;另一类是选择法,选择若干个成员分类器输出的组合作为集成决策输出<sup>[13-14]</sup>。

## 2 SVM 集成研究与应用

国内外的研究者们针对 SVM 集成进行了一系列的研究。以下将分为 SVM 集成理论与方法研究和应用研究分别进行阐述。

### 2.1 SVM 集成理论与方法研究

如何构建性能良好的 SVM 集成是 SVM 集成研究的核心内容。Oregon 州立大学计算机科学系的 Valentini 和 Dietterich 等对集成学习和 SVM 集成进行了一系列的研究<sup>[5-6,15-16]</sup>。Valentini 等对 SVM 的误差进行了偏差-方差分解分析,指出可以利用误差的偏差-方差分解设计 SVM 集成,一种途径是在 Bagging 集成中采用低偏差 SVM 作为成员分类器;另一种途径是利用偏差与方差对核参数的依赖性,通过偏差与方差分析构建异类、差异性大的低偏差 SVM 集成。理论及实验分析表明,Bagging 可以减少方差,为优化集成性能,文献[15]对 SVM 进行了偏差-方差分解,通过选择低偏差的 SVM 构成集成的成员分类器,给出了 SVM 集成的 Lobag 算法,实验结果表明,LobagSVM 与单个 SVM 及基于 bagging 的 SVM 集成相比,可以得到更好的泛化性能。Valentini 等将基于 Bagging 的 SVM 集成用于了具有高维和小样本特征的基因表达数据分析、识别,对有限样本和高维数据的分类实验表明,基于 Bagging 的 SVM 集成分类比单个 SVM 更稳定,而且具有比单个 SVM 相等或更好的分类精度,加上特征选择技术能进一步提高精度<sup>[6,15-16]</sup>。

通过构造差异性大的集成成员分类器是提高 SVM 集成泛化性能的有效途径。扰动训练样本集、扰动特征空间、扰动模型参数以及以上方法的混合等,是目前构造差异性大的集成成员 SVM、提高成员 SVM 间的差异性的常用方法。

1) 扰动训练样本集的方法。又称为基于数据划分的方法、基于样本的方法。通过划分训练集产生多个训练子集,分别在各个训练子集上进行训练,生成相应的个体 SVM。Bagging 和 Boosting 方法是目前此类方法中最为流行的2种方法。文献[17]提出了一种针对超大规模问题的 SVM 集成算法,由于采用非交叠划分训练集,很大程度地降低了个体分类器的训练时间,实验结果显示该方法的分类效果比单 SVM 好。Yan 等<sup>[18]</sup>针对类间样本数目不均的训练集,将大样本的负类样本分成  $L$  等份,每份与小样本的正类样本合成一个训练样本集,在其上训练生成一个个体 SVM,最后对得到的  $L$  个个体 SVM 采用多数投票法进行集成输出。He 等<sup>[19]</sup>在 Boosting 算法基础上提出了一种 DBoosting SVM 集成算法,实验结果表明,DBoosting 算法比 Bagging、Boosting 性能更好。Ying 等<sup>[20]</sup>提出基于距离估计的渐进思想,其方法是利用被当前基 SVM 错分的样本训练下一个成员分类器,以形成具有差异的基 SVM 集合,最后用基于距离估计的方法进行融合,此方法用于人名识别取得了很好的效果。

2) 扰动特征空间的方法。又称基于特征划分的方法。该类方法通过改变个体分类器训练数据的特征分布,以提高个体分类器间的差异性。Dimitrios 等<sup>[21]</sup>提出了一种监督和非监督学习相结合的 SVM 集成方法,首先采用模糊 C 均值法对输入特征空间进行聚类划分,然后把训练集投影到各个划分特征子空间上训练各成员分类器,分类时对所有成员分类器的结论进行合成,在 UCI 数据集上的实验表明该集成方法可提高精度 0.3% - 1.7%。Robert 等<sup>[22]</sup>提出特征 Bagging (Attribute Bagging, AB) 集成技术,指出基于特征选择的集成学习算法能进一步提升 SVM 的泛化性能。Hu 等<sup>[23]</sup>分别利用粗集理论把原始特征空间分成多个特征子空间,将训练集投影到各特征子空间上训练 SVM 成员分类器,然后进行集成。分别在医学诊断数据和 UCI 数据上的实验结果显示,该方法分类性能优于 Bagging、Boosting 及  $K$  等划分法等常用方法。文献[24]研究了 SVM 集成中的特征选择问题;文献[25]利用基于熵的方法对数据特征进行约简构造 SVM 集成,以消除不相关特征及冗余特征对集成精度的影响。

3) 扰动模型参数的方法。该类方法通过改变 SVM 模型参数或结构,以提高个体 SVM 间的差异性。Valentini 等<sup>[5]</sup>对 SVM 的误差进行了偏差 - 方差分解分析,通过分析误差的偏差和方差分解及各种核函数下各参数对 SVM 性能的影响,在低偏差区域内随机选择 SVM 的核参数、惩罚系数,构造具有低偏差的 SVM 作为成员分类器,提出了一种低偏差 Bagging SVM 集成学习算法。该方法将 SVM 的低偏差特性与 Bagging 的低无偏方差特性相结合,可以产生较低总体误差的集成。可以认为该方法是一种基于参数扰动的 SVM 集成学习。Li 等<sup>[26]</sup>将扰动径向基核函数宽度参数方式用于 AdaBoost 算法中,来产生大量个体分类器,最后从中选择出部分差异度较大的个体参与集成,实验结果表明该方法能显著提高分类精度。文献[27]在 SVM 集成中采用了文献[28]中的差异性度量,基于成员分类器间精度与差异性平衡准则,对 SVM 的径向基核参数  $\sigma$  进行扰动,通过自动调整 RBFSVM 的核参数  $\sigma$  获得具有不同差异性的 SVM,选择精度较好、差异性较大的 SVM 构成基 SVM 进行集成。实验比较表明,所提出的 SVM 集成方法与用 RBF 神经网络作成员分类器的集成方法相比,可以得到更好的泛化性能。

4) 混合式方法。该类方法通过采用 2 种以上的上述方法产生差异性大的成员 SVM。Tao 等<sup>[29]</sup>提出了一种非均匀 Bagging 和随机子空间相结合的 SVM 集成算法,克服了样本数在类间分布严重失衡及高维数小样本分类问题对 SVM 性能的影响,在图像检索应用的实验验证中取得了较好的结果。文献[30]基于对 Bagging 方法产生的训练子集进行特征选择,分别提出 PRIFEB 和 MIFEB 两种集成算法,提高了集成的差异度和精度。文献[31]提出特征与参数二重扰动机制的 SVM 集成学习算法,通过让 2 种扰动机制同时产生作用和依次产生作用,分别得到了 RAB 算法和 ZD - RBagging 算法,实验结果表明,2 种新算法均提升了 SVM 集成的泛化性能。

5) 其它方法。文献[32]提出了基于负相关学习的 SVM 集成方法,利用负相关理论来生成差异个体分类器;通过递归删除法选择出一组泛化性能优良、相互间差异性大的 SVM 参与集成学习。文献[33]提出基于 RSBR 离散化方法的 SVM 集成学习算法,采用基于粗糙集和布尔推理的离散化算法处理训练样本集,利用不同断点集合数据离散化过程形成不同的训练集合,用来构造有差异的个体分类器,实验结果表明,所提算法的分类性能明显优于单 SVM、Bagging 和 AdaBoost 方法。

另一方面,为提高 SVM 集成泛化性能,对于有效的集成结论生成方法,研究者们也进行了大量的研究。除 1.2 投票法和选择法外,常用的方法可以划分为非线性融合、多层次融合等几类。

非线性融合。文献[34]首先利用 bootstrap 方法生成训练子集,在其上训练基 SVM 分类器,SVM 的核参数在设定范围内随机选取,然后利用离散二进制粒子群算法选择部分基 SVM 参与集成。文献[35]方法与它有共同之处,但在核参数设定上有所改进:使用网格搜索方法选取较好的核参数,选择部分则用遗传算法来实现。文献[36]提出了一种基于聚类的选择性 SVM 集成预测模型,采用自组织映射和  $K$  均值聚类算法结合的聚类组合算法,从每簇中选择出精度最高的基 SVM 进行集成,在保证基 SVM 有较高精度的同时提高了它们之间的差异度。文献[37]则对正负类训练样本进行聚类,然后从中选择最具代表性的若干样本构造具有一定差异的基 SVM 分类器集成。此外,还可以结合证据理论、模糊积分、Bayes 方法等进行非线性集成,如:为减少决策的不确定性,文献[38]将 DS 证据理论应用于多分类 SVM 度量层输出信息集成,针对  $1 - v - r$  和  $1 - v - 1$  这 2 种多类扩展策略,分别定义基本概率分配函数,并根据证据冲突程度采用不同的证据组合规则,实验结果表明,证据理论方法能有效地利用两类 SVM 的度量级输出信息,取得了满意的分类结果。文献[39]给出了基于模糊积分的 SVM 集成方法。

多层次集成。文献[40]提出了 2 层 SVM 集成方法,第 1 层以 SVM 作为各成员分类器,第 2 层用一个 SVM 进行决策做出判决取得了较好的效果。Kim 等<sup>[41]</sup>利用 Bagging 和 Boosting 产生成员 SVM 分类器,用多

数投票法、加权投票法和2层SVM 3种方式相结合,构建不同的SVM集成,比较了其性能。实验结果表明2层SVM具有最高的分类精度,3种集成方法都能获得比单个SVM更高的泛化能力。然而,在已有的研究结果中,仍然存在有不尽相同的结果,如Kim等<sup>[41]</sup>认为简单地Bagging SVM和Boosting SVM就能提高分类精度,SVM集成方法与用RBF神经网络、Bayes分类器等作成员分类器的集成方法相比,可以得到更好的泛化性能;然而Dong等<sup>[42]</sup>通过用文本分类数据作实验,对采用5种不同数据划分方法的SVM集成的性能比较后发现,采用K等分划分法加Stacking集成技术获得了较好的集成效果,Bagging SVM不能提高性能,Boosting SVM反而降低了性能。造成这种现象的原因在于,尽管SVM存在待解决的问题,但与神经网络和决策树相比,SVM是一种相对“稳定”和“高精度”的学习机,这增加了SVM集成的难度,常用于神经网络等不稳定算法的集成技术未必对SVM有效。

## 2.2 SVM集成应用研究

应用SVM集成解决实际问题也是SVM集成研究的一个重要方面。到目前为止,SVM集成被应用到回归分析、欺诈检测、基因数据分析,生物学数据分类、疾病诊断、故障诊断、场景分类、人脸检测、文本分类、图像检索、遥感数据分类等不同的应用领域,并显示出了优势。

针对基因表达数据具有高维和小样本的特征,Valentini<sup>[43]</sup>中提出了一种基于Bagging和特征选择的SVM集成方法,结果显示比单SVM具有更好的稳定性和更高的分类精度,在文献[6]中将该方法应用于肿瘤识别也取得了理想的效果。Loris Nanni等<sup>[44]</sup>在研究氨基酸的物理化学特性分类时,引入了SVM集成分类方法,个体SVM分类器在每一个氨基酸的物理化学特性上训练生成,采用多数投票法进行结论合成,实验取得了较好的效果。谷雨等<sup>[45]</sup>针对入侵检测中的入侵数据相对稳定及对检测速度的需求,提出分别以PCA和ICA方法对入侵数据进行特征提取,结合增量式SVM算法,在其上分别构造2个SVM子分类器,最后采用集成技术对2个子分类器的结果进行合成,实验表明,该集成分类系统具有理想的入侵检测性能,尤其具有较低的漏检率。

唐静远等<sup>[46]</sup>针对模拟电路故障诊断问题,提出了基于特征扰动的SVM集成算法,将采集来的信号先进行Haar小波变换,提取1-5层小波变换的每层第1个低频系数构成特征子集,然后在各特征子集上分别训练个体SVM,结果进行集成输出,实验结果表明该方法具有较好的分类准确率和泛化性。使用SVM集成可以构成有效的多类分类器<sup>[47]</sup>。文献[48]把SVM集成应用到增量学习中,并用于不稳定环境中数据的预测。Kang<sup>[49]</sup>利用基于遗传算法的覆盖优化技术解决基SVM分类器之间的多重共线性问题以降低成员分类器之间的相关性,该方法用于公司破产预测中。

## 3 存在的问题与几个重要的研究方向

相比于对神经网络、决策树等集成的研究,对于SVM集成学习技术的研究还有很多理论和应用问题需要解决。要进行有限训练样本下有效的集成学习,如何保证SVM集成的泛化性能比构成它的子SVM的性能好,如何进行有效的SVM集成增量学习,如何将SVM集成扩展到多类分类应用中,如何以较快的速度完成大规模数据的学习问题等等,都是需要进行进一步深入研究的问题。

因此,寻找有效的、泛化性能好的集成技术成为目前SVM集成研究中需要进一步深入研究的关键问题,而引入新的集成技术和结合多种集成技术的研究方法正成为SVM集成研究的主要方向。针对存在的问题以及各种新任务的需要,研究SVM集成的理论、方法、以及SVM集成在新的研究和应用领域的拓展,将是今后相当长的一段时间内SVM集成学习研究与应用面临的挑战。

## 参考文献(References):

- [1] Vapnik V N. The nature of statistical learning theory[M]. New York: Springer-verlag, 1995.
- [2] Jain A K, Duijn R P W, Mao J. Statistical pattern recognition: a review[J]. IEEE trans. on pattern analysis and machine intelligence, 2000, 22(1): 4-37.
- [3] 王晓丹,王积勤. 支持向量机研究与应用[J]. 空军工程大学学报:自然科学版, 2004, 5(3): 49-55.  
WANG Xiaodan, WANG Jiqin. Research and application of support vector machine[J]. Journal of air force engineering university: natural science edition, 2004, 5(3): 49-55. (in Chinese)
- [4] 许建华,张学工,李衍达. 支持向量机的新发展[J]. 控制与决策, 2004, 19(5): 481-484.  
XU Jianhua, ZHANG Xuegong, LI Yanda. Advances in support vector machines[J]. Control and decision, 2004, 19(5): 481-484. (in Chinese)

- [5] Valentini G, Dietterich T G. Bias - variance analysis of support vector machines for the development of svm - based ensemble methods[J]. Journal of machine learning research, 2004, 5:725 - 775.
- [6] Valentini G, Muselli M, Ruffino F. Cancer recognition with bagged ensembles of support vector machines[J]. Neurocomputing, 2004, 56: 461 - 466.
- [7] Wang Shijin. Empirical analysis of support vector machine ensemble classifiers[J]. Expert systems with applications, 2009, 36(3): 6466 - 6476.
- [8] Schapire R E. The strength of weak learnability[J]. Machine learning, 1990, 5(2): 197 - 227.
- [9] Kearns M, Valiant L G. Cryptographic limitations on learning boolean formulae and finite automata[J]. Journal of the ACM, 1994, 41(1): 67 - 95.
- [10] Krogh A, Vedelsby J. Neural network ensembles, cross validation, and active learning[C]//Advances in neural information processing systems. Cambridge, MA: MIT Press, 1995:231 - 238.
- [11] Kuncheva L L, Whitaker C J, Shipp C A, et al. Limits on the majority vote accuracy in classifier fusion[J]. Pattern analysis and applications, 2003, 6(1): 22 - 31.
- [12] Kuncheva L. Combining pattern classifiers: methods and algorithms[M]. New Jersey: John Wiley & Sons Inc, 2004.
- [13] Zhou Z H, Wu J X, Tang W. Ensembling neural networks; many could be better than all[J]. Artificial intelligence, 2002, 137(1-2): 239 - 263.
- [14] 米爱中, 郝红卫, 郑雪峰, 等. 一种自整定权值的多分类器融合方法[J]. 电子学报, 2009, 37(11): 2604 - 2609.  
MI Aizhong, HAO Hongwei, ZHENG Xuefeng, et al. A method of multiple classifier fusion with self - adjusting weights[J]. Acta electronica sinica, 2009, 37(11): 2604 - 2609. (in Chinese)
- [15] Valentini G, Dietterich T G. Low bias bagged support vector machines[C]//Proceedings of the twentieth international conference on machine learning. Washington: AIAA Press, 2003: 752 - 759.
- [16] Giorgio Valentini. An experimental bias - variance analysis of SVM ensembles based on resampling techniques[J]. IEEE trans on systems man and cybernetics - part B: Cybernetics, 2005, 35(6): 1252 - 1271.
- [17] Collobert R, Bengio S, Bengio Y. A parallel mixture of SVM for very large scale problems[J]. Neural computation, 2002, 14: 1105 - 1114.
- [18] Yan R, Liu Y, Jin R, et al. On predicting rare classes with SVM ensembles in scene classification[C]//Proceedings of IEEE international conference on acoustics speech and signal processing. [S. l.]: IEEE press, 2003: 6 - 10.
- [19] Lingming He, Xiaobing Yang, Huijun Lu. A comparison of support vector machines ensemble for classification[C]//Proceedings of the sixth international conference on machine learning and cybernetics. Harbin: [s. n.], 2007: 3613 - 3617.
- [20] Ying W, Xiaoling Wang, Bingquan Liu. A gradual combining method for multi - SVM classifiers based on distance estimation[C]//Proceedings of the third international conference on machine learning and cybernetics. Shanghai: [s. n.], 2004: 26 - 29.
- [21] Dimitrios S, Frossyniotis, Andreas S. A multi - SVM classification system[C]//Proceedings of the Second International workshop on multiple classifier system. London: Josef Kittler Fabio Roli Springer - Verlag, 2001: 198 - 207.
- [22] Robert B, Ricardo G O, Francis Q. Attribute bagging: improving accuracy of classifier ensembles by using random feature subsets[J]. Pattern recognition, 2003, 36: 1291 - 1302.
- [23] Hu Zhonghui, Cai Ylinze, He Xing, et al. Support vector machine ensemble using rough sets theory[J]. High technology letters, 2006, 12(1): 58 - 62.
- [24] Yu E Z, Cho S Z. Ensemble based on GA wrapper feature selection[J]. Computers & industrial engineering, 2006, 51: 111 - 116.
- [25] Lei R H, Kong X X, Wang X S. An ensemble SVM using entropy - based attribute selection[C]//Control and decision conference. Xuzhou: [s. n.], 2010: 802 - 805.
- [26] Li X C, Wang L, Eric Sung. A daBoost with SVM based component - classifiers[J]. Engineering application of artificial intelligence, 2008, 21: 785 - 795.
- [27] Li Xuchun, Wang Lei, Sung Eric. A study of adaBoost with SVM based weak learners[C]//Proceedings 2005 IJCNN. Montreal, Quebec, Canada: IEEE press, 2005: 196 - 201.
- [28] Kuncheva L I, Whitaker C J. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy[J]. Machine learning, 2003, 51(2): 181 - 207.
- [29] Tao D C, Tang X O. Asymmetric bagging and random subspace for support vector machines - based relevance feed back in image retrieval[J]. IEEE transactions on pattern analysis and machine intelligence, 2006, 28(7): 1088 - 1099.
- [30] 李国正, 李丹. 集成学习中特征选择技术[J]. 上海大学学报: 自然科学版, 2007, 13(5): 598 - 604.  
LI Guozheng, LI Dan. Feature selection for ensemble learning[J]. Journal of Shanghai University: natural science edition, 2007, 13(5): 598 - 604. (in Chinese)
- [31] 贾华丁, 游志胜, 王磊. 采用二重扰动机制的支持向量机的集成训练算法[J]. 控制与决策, 2008, 23(7): 525 - 532.  
JIA Huading, YOU Zhisheng, WANG Lei. Ensemble algorithms for training support vector machine based on the double disturbance mechanism[J]. Control and decision, 2008, 23(7): 525 - 532. (in Chinese)
- [32] 唐耀华, 高静怀, 包乾宗. 一种新的选择性支持向量机集成学习算法[J]. 西安交通大学学报, 2008, 42(10): 1221 - 1225.  
TANG Yaohua, GAO Jinghuai, BAO Qianzong. Novel selective support vector machine ensemble learning algorithm[J]. Journal of Xi'an Jiaotong University, 2008, 42(10): 1221 - 1225. (in Chinese)
- [33] 蔡铁, 伍星, 李焯. 集成学习中基于离散化方法的基分类器构造研究[J]. 计算机应用, 2008, 28(8): 2091 - 2093.  
CAI Tie, WU Xing, LI Ye. Research on construction of base classifiers based on discretization method for ensemble learning[J]. Computer ap-

- plications, 2008, 28(8): 2091 – 2093. (in Chinese)
- [34] Zhang H D, Wang X D, Wu C M, et al. Selective SVM ensembles based on modified BPSO [C]//Proceedings of computational intelligence and industrial application. Wuhan: [s. n.], 2008: 243 – 246.
- [35] Liu Y, Liao Z J, Yin Y F, et al. Selective and heterogeneous svm ensemble for demand forecasting [C]//Computer and information technology. Bradford: IEEE press, 2010: 1519 – 1524.
- [36] 蔡俊伟, 胡寿松, 陶洪峰. 基于选择性支持向量机集成的混沌时间序列预测[J]. 物理学报, 2007, 56(12): 6820 – 6828  
CAI Junwei, HU Shousong, TAO Hongfeng. Prediction of chaotic time series based on selective support vector machine ensemble [J]. ACTA physica sinica, 2007, 56(12): 6820 – 6828. (in Chinese)
- [37] Yuan Hejin, Zhang Yanning, Yang Fuzeng, et al. A novel svm ensemble approach using clustering analysis [J]. Journal of electronics, 2008, 25(2): 2341 – 2346.
- [38] 李烨, 蔡云泽, 尹汝泼, 等. 基于证据理论的多类分类支持向量机集成[J]. 计算机研究与发展, 2008, 45(4): 571 – 578.  
LI Ye, CAI Yunze, YIN Rupo, et al. Support vector machine ensemble based on evidence theory for multi – class classification [J]. Journal of computer research and development, 2008, 45(4): 571 – 578. (in Chinese)
- [39] 颜根廷, 李传江, 马广富. 支持向量分类器的模糊积分集成方法[J]. 哈尔滨工业大学学报, 2008, 40(7): 1017 – 1020.  
YAN Genting, LI Chuanjiang, MA Guangfu. Support vector classifiers ensemble based on fuzzy integral [J]. Journal of harbin institute of technology, 2008, 40(7): 1017 – 1020. (in Chinese)
- [40] Pang S N, Kim Daijin, Bang S Y. Fraud detection using support vector machine ensemble [C]//Proceedings of 8th ICONIP 2001. Shanghai: Fudan university press, 2001: 1344 – 1349.
- [41] Hyun Chul Kim. Constructing support vector machine ensemble [J]. Pattern recognition, 2003, 36(12): 2757 – 2767.
- [42] Dong Yan Shi, Han Ke Song. Boosting SVM classifiers by ensemble [C]//Proceedings of the 14th IEEE international conference on world wide web. [S. l.]: IEEE press, 2005: 1072 – 1073.
- [43] Valentini G, Muselli M, Ruffino F. Bagged ensembles of support vector machines for gene expression data analysis [C]//Proceedings of the international joint conference on neural networks. Portland: IEEE computer society, 2003: 110 – 115.
- [44] Loris Nanni, Alessandra Lumini. An ensemble of support vector machine based on multiple physicochemical properties of amino acids [J]. Neurocomputing, 2006, 69: 1688 – 1690.
- [45] 谷雨, 徐宗本, 孙剑, 等. 基于 PCA 与 ICA 特征提取的入侵检测分类系统 [J]. 计算机研究与发展, 2006, 43(4): 633 – 638.  
GU Yu, XU Zongben, SUN Jian, et al. An intrusion detection ensemble system based on the features extracted by PCA and ICA [J]. Journal of computer research and development, 2006, 43(4): 633 – 638. (in Chinese)
- [46] 唐静远, 师奕兵, 张伟. 基于支持向量机集成的模拟电路故障诊断 [J]. 仪器仪表学报, 2008, 29(6): 1216 – 1220.  
TANG Jingyuan, SHI Yibing, ZHANG Wei. Analog circuit fault diagnosis based on support vector machine ensemble [J]. Chinese journal of scientific instrument, 2008, 29(6): 1216 – 1220. (in Chinese)
- [47] Jin Deng Zhou, Xiao Dan Wang, Heng Song. Research on the unbiased probability estimation of error – correcting output coding [J]. Pattern recognition, 2011, 44(7): 1552 – 1565.
- [48] Yalcin A, Erdem Z, Gurgen F. Ensemble based incremental SVM classifiers for changing environments [C]//Computer and information sciences. Ankara, Turkey: [s. n.], 2007: 1 – 5.
- [49] Kang Daeki Kim, Jong Myoung. Performance enhancement of SVM ensembles using genetic algorithms in bankruptcy prediction [C]//Advanced computer theory and engineering. Chengdu: [s. n.], 2010: 154 – 158.

(编辑:田新华)

## Research and Application of SVM Ensemble

WANG Xiao – dan<sup>1</sup>, GAO Xiao – feng<sup>2</sup>, YAO Xu<sup>1</sup>, LEI Lei<sup>1</sup>

(1. Missile Institute, Air Force Engineering University, Sanyuan 713800, Shaanxi, China; 2. The Logistics Department, Shaanxi Armed Police Forces, Xi'an 710054, China)

**Abstract:** Ensemble learning is an important research direction in machine learning, in recent years, support vector machine (SVM) ensemble has attracted the attentions of researchers in the field of machine learning and statistical learning, and has become an active research hotspot. The research and application of SVM ensemble in recent years are surveyed in this paper and the fundamental problems in SVM ensemble are discussed. Methods of constructing SVM base classifiers with higher diversity, effective ways for combining base classifiers, and the typical applications of SVM ensemble are discussed and analyzed. The problems those exist in the research and application of SVM ensemble are pointed out and the important research directions are given.

**Key Words:** support vector machine; ensemble learning; diversity