

文章编号: 1000-5862(2019)03-0309-05

# 基于分形理论的支持向量机核函数选择

梁礼明 陈明理 刘博文 吴 健

(江西理工大学电气工程与自动化学院 江西 赣州 341000)

**摘要:** 核方法是机器学习领域内的研究热点之一,在处理非线性和高维数据问题中表现出许多优势,已被广泛应用于分类、回归等领域。支持向量机是最具代表性的核方法,而不同的核函数具有各自的度量特征,故核函数的选择对支持向量机泛化能力有着重要的影响。而目前核函数的选择仍是一个开放性的问题,存在着一系列的偶然性和局限性。该文利用分形几何分析数据蕴含的特征信息来有指导性地选择核函数,以提高支持向量机的泛化能力,并通过实例仿真验证该方法是有效可行的。

**关键词:** 核方法; 支持向量机; 核函数; 分形几何; 分形维数

**中图分类号:** TP 181 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2019.03.15

## 0 引言

支持向量机(Support Vector Machine, SVM)是一种以VC维和结构风险最小原则为基础,借助最优化方法解决数据挖掘、模式识别等若干问题较有效的工具<sup>[1]</sup>。相对传统机器学习方法其优势是针对小样本问题能平衡学习能力和复杂性,有较好的泛化能力,在一定程度上避免了“维数灾难”和“过学习”等传统机器学习中难以解决的问题。在处理非线性分类问题时,以映射的方式将输入空间嵌入到高维特征空间,间接地实现复杂问题的分类<sup>[2]</sup>。该变换以核函数形式隐式定义,这类机器学习方法即为核方法。核方法不是针对样本空间进行复杂处理,而是在特征空间中高效地计算内积,因此核函数的引入显著地提高了计算能力<sup>[3]</sup>。SVM是最具代表性的一种核方法,合理地选择或构造核函数对SVM模型构建有着极为重要的作用。而各类核函数所表现的特征空间度量各异,即在不同核函数下由样本数据生成的核矩阵所隐含的空间分布信息特征各不相同,甚至差异程度较大。如何针对数据选择或构造合适的核函数便成为SVM研究的热点之一。但核函数的选择一直是SVM研究中开放性的问题,没有具体的理论指导,存在着偶然性和局限性。核函数的选择方法仍是支持向量机领域一个重要的研究课题。

## 1 相关工作

关于设计或构造核函数这类问题研究方面虽然取得一些进展,但仍有较大提升空间<sup>[4-8]</sup>。在实际问题中通常使用的方法有:(i)基于专家经验进行选择;(ii)Cross-Validation方法,对各核函数通过试用,然后选择准确率最高的核函数。如文献[4]结合各核函数对不同样本分类测试对比得出准确率更高的核函数;文献[5]预先选择RBF核函数,通过ESDR值对核函数参数优化获得更强的分类能力。这2种方法均未考虑样本数据的特征规律,没有充分利用数据隐含的先验信息。在此基础上也延伸出了一些根据实际问题来选择核函数的方法,如文献[7]以分类错误的样本为基础,计算其与决策界面的距离来进行秩和检验,并基于此提出了一种核函数选择方法,该方法在提高核函数分类精度和选择上有一定的指导意义。以分类错误的样本为基础是一种比较新颖的方式,为核函数选择提供一个较好的方向,但该方法在核函数选择方面对数据集有一定的限制而不适用于所有数据集。文献[8]提出并行计算模型的核函数选择方法,该方法利用COSMOS模型、遗传算法、3层次级联方法等分析模型误差来得到要选择的核函数,但该方法会使算法复杂度高,选择花费时间长,而且效果提高并不明显。本文以输入样

收稿日期: 2019-01-27

基金项目: 国家自然科学基金(51365017),江西省自然科学基金(20132BAB203020)和江西省教育厅科学技术研究(GJJ170491)资助项目。

作者简介: 梁礼明(1967-),男,江西赣州人,教授,主要从事机器学习、模式识别研究。E-mail: lianglm67@163.com

本的分形几何这一先验信息为前提,结合核函数蕴含的特征信息,提出了一种支持向量机核函数的选择方法.通过实例仿真可以看出,该方法不仅可以适用于各种数据集,同时也没有增加算法的复杂度,对 UCI 中数据集的测试准确率相较于其他方法也有着明显的优势.

对于给定的具体问题(样本数据集)而言,可用 VC 维来描述拟建模型函数集容量的表达能力,即用 VC 维来度量具体问题的复杂程度<sup>[9]</sup>.然而,遗憾的是 VC 维是一个抽象的概念,迄今为止尚无明确的统一计算方法<sup>[10]</sup>.而分形理论是基于自相似性的原理,将复杂现象看成是由简单现象迭代而成,由此反映出复杂现象中所蕴含的规律和特性,对解决复杂问题有着较为积极的影响.分形维数是分形中能够定量对分形集的复杂度进行衡量的定量指标.同时,无论是分析样本数据集先验信息还是核矩阵的特征信息,均需对数据本质特征探究并判断数据集中是否存在相互关联属性及属性的相关性.而本质维数是在尽可能保留数据结构特征的情况下,本质地表征数据所需要的最小维数,并能保留数据结构特征的固有属性.同时分形维数反映的是数据集的固有特性,故数据集的本质维数通常用分形维数定量的表示<sup>[11]</sup>.因此,本文利用分形理论给定输入样本的先验信息,并结合核特征空间隐含的特征信息来分析和指导选择核函数,这将会是一个新的尝试.

## 2 相关理论基础

**定义 1(核函数)** 称定义在  $\mathbf{R}^n \times \mathbf{R}^n$  上的函数  $K(x, x')$  是  $\mathbf{R}^n \times \mathbf{R}^n$  的核函数,若存在从  $\mathbf{R}^n$  到 Hilbert 空间  $H$  的变换  $\Phi: \mathbf{R}^n \rightarrow H, x \mapsto \Phi(x)$ ,使得  $K(x, x') = (\Phi(x) \cdot \Phi(x'))$ ,其中  $(\cdot)$  为空间  $H$  中的内积.所有满足 Mercer 定理的映射构成的函数都可以作为核函数<sup>[12]</sup>.常见的核函数有线性、高斯(RBF)、多项式、Sigmoid 等类型<sup>[13]</sup>.

核函数利用映射的方法将样本数据嵌入到特征空间中,特征空间中的样本分布决定了核方法性能.利用核特征空间的性质,对给定样本的先验信息进行分析将是解决核函数选择等问题的关键点.

核函数性质依赖于核度量,即有效的核度量标准对核函数选择有着积极的影响<sup>[14]</sup>.根据核函数的黎曼度量、距离度量和角度度量可以得出,核函数具有不同的分布特征,如高斯核函数呈现局部特性,多项式核函数呈现全局特性.无论是局部还是全局特性均表现在数据点之间的作用关系,可以看作是数

据点之间的相似性关系<sup>[15]</sup>.被选择的核函数应使相同属性输入样本间的相似性程度高而不同属性输入样本间的相似性程度低.这种物质本身相似性关系可以用自相似性进行分析.

核度量关系如何表征,这是解决核函数选择等问题的关键点.本文引入 Gram 矩阵的概念, Gram 矩阵是由核函数根据样本数据变换后产生的矩阵,因此,对核矩阵分析即可得到相应核函数自相似性.

**定义 2(核函数矩阵—Gram 矩阵)** 对于训练样本  $\{(x_i, y_j)\}_{1 \leq i, j \leq n}$ ,核函数  $k$  在样本点  $x_i$  和  $y_j$  的值通常可以用矩阵来表示,即  $K = (k_{ij})_{n \times n}$ ,其中  $k_{ij} = k(x_i, y_j)$ ,则称该矩阵为核函数在训练样本上的 Gram 矩阵.

**定理 1** 未知系统的数据蕴藏着近似或统计学意义上的自相似性.

自相似性是分形几何的基本特征<sup>[16]</sup>.自然界中部分的某些特征与整体相似被称为分形,具有该特性的几何被称为分形几何.分形具有比较明显的不规则性,并不会随着数据集的增大或减小而发生本质性的改变<sup>[17]</sup>.该性质也是自相似性的一种体现.

分形维数是描述分形的一个重要的特征指标,是一个不变的内在特征量,能有效地描述数据集并反映复杂数据集中隐藏的规律性.分形维数主要的表现形式为 Hausdorff 维数,而 Hausdorff 维数是在 Hausdorff 测度的基础上衍变出来的,无法直接计算,只能求近似值.故在此基础上又有许多种计算分形维数的方法先后被提出<sup>[18]</sup>.

**定义 3( Hausdorff 维数)** 对于任意给定的集合  $F$  和  $\delta < 1$ ,  $H_\delta^p(F)$  对于  $p$  来说是非增的,因为当  $p = 0$  时,只要  $F$  非空,就必有  $H_\delta^0(F) = \infty$ .若  $t > p$  且  $\{U_i\}$  是  $F$  的一个  $\delta$  覆盖,则有

$$\sum_i |U_i|^t = \sum_i |U_i|^p |U_i|^{t-p} \leq \delta^{t-p} \sum_i |U_i|^p,$$

从而有  $H_\delta^t(F) \leq \delta^{t-p} H_\delta^p(F)$ .令  $\delta \rightarrow 0$ ,若  $0 < H^p(F) = \lim_{\delta \rightarrow 0} H_\delta^p(F) < \infty$ ,则有  $H^t(F) = \lim_{\delta \rightarrow 0} H_\delta^t(F) = 0$  ( $t > p$ ).同理,可证当  $t < p$  时,若  $\delta \rightarrow 0$ ,  $0 < H^p(F) < \infty$ ,则有  $H^t(F) = \infty$ .这说明存在  $p$  的一个临界点使得  $H^p(F)$  从  $\infty$  突降为 0.这个临界值被称为  $F$  的 Hausdorff 维数,记为  $\dim_H F$ .

自然界事物普遍存在局部、整体在某种方面具有统计学意义上的自相似性,自相似性可以由分形维数定量的表现.核同样是样本相似度的一种度量,核相似性的特点体现在它可以看作样本空间映射到高维欧氏空间的内积.因此,本文对样本数据集的分形维数与各核函数的分形维数进行分析,从中找出

2 组数据集间的自相似性关系,并以此为依据及结合样本数据蕴含的先验信息来选择核函数。

### 3 模型框架

由于各核函数生成的核矩阵呈现差异,所以建立基于分形几何的核函数选择模型需要如下过程:首先分析各类核函数特征空间的分形维数并设定其取值范围;然后,将样本数据分形维数与核函数的特征信息进行有机结合,以此指导核函数的选择;最后将选出的核函数应用于样本数据并进行分类测试,从而判断选择核函数的有效性。本文核函数选择模型主要包括数据预处理、核矩阵构造、分形维数计算以及核函数选择等4个部分。

#### 3.1 数据预处理

输入样本数据  $X = \{x_i\}_{i=1}^l \in \mathbf{R}^n$ , 其中  $\mathbf{R}^n$  为  $n$  维数据空间,  $l$  为样本数,对  $x$  经过处理使其范数小于1。

#### 3.2 核矩阵构造

利用预处理后的样本数据及结合核函数的数学表达式,以内积的形式计算出样本数据的核矩阵。构造方法如下:

用  $G$  表示对应于核函数  $k$  的核矩阵,即

$$G_{ij} = \langle \varphi(x_i), \varphi(x_j) \rangle = k(x_i, x_j),$$

其中  $k(x_i, x_j) = k(x_j, x_i)$ ,  $\varphi$  为特征映射。

#### 3.3 分形维数计算

因本文主要是分析数据蕴含的特征信息,数据集间的每个数据点与数据特征之间是相互联系的,故利用关联维数的方法对数据集的分形维数进行计算。具体关联维数计算方法如下:

若在空间中某一集合是由  $N$  个点组成的,各点空间坐标是  $X_i (i = 1, 2, \dots, N)$ 。凡空间距离小于  $\delta$  的点认为有关联,对这些点计数并计算在一切可能的  $N^2$  配对中所占比例,称之为关联函数,其公式为

$$C(\delta) = \frac{1}{N^2} \sum_{i,j=1}^N \theta(\delta - |x_i - x_j|), \quad (1)$$

其中  $\theta(r)$  是 Heaviside 函数,即

$$\theta(r) = \begin{cases} 1 & r > 0, \\ 0 & r \leq 0, \end{cases} \quad (2)$$

P. Grassberger 等<sup>[19]</sup>在1983年应用关联函数  $C(\delta)$  给出了关联维数的定义:

$$D = - \lim_{\delta \rightarrow 0} \ln C(\delta) / \ln \delta. \quad (3)$$

由此,计算出样本数据与核矩阵的分形维数。根据(1)~(3)式可知,在相同  $\delta$  下若样本数据呈全局分布,其  $N^2$  较小,而  $C(\delta)$  与  $N^2$  成反比,即  $D$  与  $N^2$  成

正比,这表明具有全局分布特性的数据集具有较小的  $D$  值,即全局分布特性的数据关联程度较低;而样本数据具有局部分布特性,其  $N^2$  较大,具有较大的  $D$  值,即局部分布特性的数据关联程度较高。

#### 3.4 核函数选择

根据样本数据与核矩阵的分形维数值分析并通过大量实验验证,确定本文选取的分形维数的阈值。根据计算的分形维数值,以及给定的阈值选择相应的 SVM 核函数类型。

#### 3.5 模型分析

本文算法框架是以分形几何为基础,这能够更好地分析出不规则数据集的自相似等潜在信息,并且结合了对核矩阵的分析结果,直接通过数据集的特征信息进行核函数选择。充分地利用了核函数的几何度量和几何性质,避免了核函数选择的局限性和偶然性。

## 4 实验仿真与分析

本文实验平台为 Win7 系统下的 Matlab2014a,使用 LIBSVM 工具箱对 UCI 数据库中常用的且带标签的数据进行分类。结合核函数所蕴藏的度量特性,以典型的全局分布的多项式核函数和局部分布的径向基核函数为例进行仿真实验。随机选出样本数据的80%和20%分别作为SVM模型的训练集和测试集。

#### 4.1 实验仿真

4.1.1 样本数据分形维数估计 对 UCI 数据库中随机选出的8组样本数据进行预处理,并根据(1)~(3)式计算出样本数据的分形维数。由于核可看作样本空间映射到高维欧氏空间的内积,该内积不会改变样本本身蕴涵固有特征,即样本与核函数生成的核矩阵之间存在着自相似性的特点。本文以典型的全局分布特征的多项式核函数和局部分布特征的径向基核函数(RBF)为例进行数值仿真实验。由实验分析得出,具有全局特征的多项式核函数分形维数值均小于2,具有局部特征的RBF核函数分形维数值均大于2,且与由(1)~(3)式理论分析得出全局特征和局部特征数据分形维数的特性相吻合。因此,确定核函数矩阵分形维数的阈值为  $D^* = 2$ ,即若  $D \leq D^*$ ,则选择多项式核函数;若  $D > D^*$ ,则选择径向基核函数。8组样本数据核矩阵分形维数如表1所示,每组样本数据的分形维数估计值和拟选核函数类型如表2所示。

表 1 各核矩阵分形维数

数据集及其规模	核函数类型	分形维数
Wine Data Set	RBF	2.674 3
(178 × 13)	多项式	1.434 2
Balance Scale	RBF	3.142 2
Data Set(625 × 4)	多项式	1.700 5
Statlog (Heart)	RBF	3.156 2
Data Set(270 × 13)	多项式	1.511 6
User Knowledge Modeling Data	RBF	2.095 8
Set(403 × 5)	多项式	1.543 6
Australian Credit Approval Data	RBF	3.125 8
Set(690 × 14)	多项式	0.969 2
Tic-Tac-Toe Endgame Data Set	RBF	2.420 3
(958 × 9)	多项式	1.359 0
Liver Disorders Data Set(345 × 7)	RBF	2.459 9
	多项式	1.951 5
Haberman's Survival Data Set	RBF	2.233 7
(306 × 4)	多项式	1.616 6

4.1.2 数据仿真 根据设定的阈值选择核函数,利用选择的核函数对以上 8 组数据进行分类预测.为

避免核函数的参数对本文实验带来影响,在对数据进行仿真时,核函数的参数均采用 LIBSVM 工具箱中设定的默认值,随机进行 3 次实验取平均值.各组数据仿真的准确率与程序运行时间如表 3 所示.

表 2 各样本数据的分形维数和拟选核函数类型

数据集及其规模	分形维数	拟选核函数类型
Wine Data Set(178 × 13)	2.900 5	RBF
Balance Scale Data Set(625 × 4)	1.545 8	多项式
Statlog (Heart) Data Set(270 × 13)	2.643 9	RBF
User Knowledge Modeling Data Set(403 × 5)	1.922 6	多项式
Australian Credit Approval Data Set(690 × 14)	4.235 2	RBF
Tic-Tac-Toe Endgame Data Set(958 × 9)	1.752 7	多项式
Liver Disorders Data Set(345 × 7)	2.753 6	RBF
Haberman's Survival Data Set(306 × 4)	1.456 0	多项式

表 3 样本数据仿真测试

数据集及其规模	拟选核函数类型	核函数类型	样本分类 1/%	样本分类 2/%	样本分类 3/%	平均值/%	实验时间/s
Wine Data Set(178 × 13)	RBF	RBF	100.00	100.00	100.00	100.00	1.12
	多项式	多项式	88.89	91.67	97.22	92.59	1.33
Balance Scale Data Set(625 × 4)	RBF	RBF	100.00	98.40	100.00	99.47	2.66
	多项式	多项式	100.00	100.00	100.00	100.00	2.79
Statlog (Heart) Data Set(270 × 13)	RBF	RBF	96.30	98.15	98.15	97.53	1.20
	多项式	多项式	88.89	94.44	98.15	93.83	1.32
User Knowledge Modeling Data Set(403 × 5)	RBF	RBF	92.59	93.83	96.30	94.24	2.68
	多项式	多项式	100.00	98.77	95.06	97.94	2.82
Australian Credit Approval Data Set(690 × 14)	RBF	RBF	100.00	100.00	100.00	100.00	2.69
	多项式	多项式	98.55	100.00	99.28	99.28	2.77
Tic-Tac-Toe Endgame Data Set(958 × 9)	RBF	RBF	86.46	84.38	84.90	85.25	2.71
	多项式	多项式	98.96	98.44	85.94	94.45	2.85
Liver Disorders Data Set(345 × 7)	RBF	RBF	100.00	100.00	100.00	100.00	2.67
	多项式	多项式	100.00	66.67	55.07	73.91	2.72
Haberman's Survival Data Set(306 × 4)	RBF	RBF	87.97	78.69	83.61	83.42	2.74
	多项式	多项式	100.00	96.72	90.16	95.48	2.80

从表 3 可以看出,本文方法选择出来的核函数对样本数据分类的正确率均比其它核函数的分类的正确率高,其中 Haberman 数据集利用本文方法选取的多项式核函数对样本数据分类准确率高于其他核函数分类的准确率(超过 10%); Liver 数据集利用本文方法选取的 RBF 核函数对样本数据分类准确率高于其他核函数分类的准确率(超过 20%).

由数据仿真结果可以得出:基于分形维数大小的核函数选择模型,能够针对样本数据先验信息充分利用各核函数的特点,并且选取的核函数处理数

据能够得到理想的结果.同时每组数据在不同核函数下运行时间相差无几,故根据本文方法选择的核函数对 SVM 模型的运行不会产生不利影响,这表明本文提出算法是有效和可行的.

#### 4.2 样本数据对比试验分析

将对文献[7,13]的数据运用本文方法进行仿真实验分析,使用的数据有 Statlog(Heart) Data Set、Car Evaluation Database、Glass Identification Database、Iris Plants Database 4 组.经过对这 4 组数据的

分形维数估计,并根据本文设定的阈值选择相应的核函数,然后进行分类测试.取 3 次分类测试结果的平均值,本文方法与文献 [7,13] 中的分类准确率对比如表 4 所示.

表 4 与相关文献分类准确率对比分析

数据集	分形维数	本文方法		文献 [7,13] 的方法	
		拟选核函数	准确率/%	选用核函数	准确率/%
Statlog(Heart) Data Set	2.643 9	RBF	92.59	多项式	89.62
Car Evaluation Database	2.391 0	RBF	99.13	多项式	93.33
Glass Identification Database	2.789 8	RBF	99.22	多项式	99.07
Iris Plants Database	0.962 6	多项式	98.89	RBF	96.67

由表 4 可以看出,上述 4 组数据集用本文方法指定的核函数类型的分类准确率均优于文献 [7,13] 所指定的核函数类型的分类准确率.

5 结束语

本文利用分形几何中的分形维数并结合核函数蕴含的度量特征,提出一种 SVM 核函数选择机制.这是一种有理论基础且对 SVM 核函数具有积极意义的选择方法,能够针对具体问题有效地选择合适的核函数,有效地避免核函数选择中存在的偶然性与局限性,克服传统 SVM 模型选择方法中不能使模型性能达到最优的不足.通过对 UCI 数据集的仿真以及对比试验,验证了本文方法是可行且有效的,丰富了核函数选择的方法.此外,对大多数复杂问题而言单一核函数难以反映出其分布特征,有必要优化核函数选择形式,如混合核函数.因此,如何根据分形理论进行混合核函数选择将是本文下一步研究的延伸方向.

6 参考文献

[1] 郝云霄,闫楚良,刘克格.基于支持向量机的机翼载荷模型研究[J].科学技术与工程,2013,13(25):7432-7437.

[2] 陈鹏,胡啸峰,陈建国.基于模糊信息粒化的支持向量机在犯罪时序预测中的应用[J].科学技术与工程,2015,15(35):54-57,63.

[3] 刘高辉,杨星.一种混合核函数的支持向量机[J].微型机与应用,2017,36(11):19-22.

[4] 宋晖,薛云,张良均.基于 SVM 分类问题的核函数选择仿真研究[J].计算机与现代化,2011(8):133-136.

[5] 尹嘉鹏.支持向量机核函数及关键参数选择研究[D].哈尔滨:哈尔滨工业大学,2016.

[6] 梁礼明,冯新刚,陈云嫩,等.基于样本分布特征的核函数选择方法研究[J].计算机仿真,2013,30(1):323-

328.

[7] 王振武,何关瑶.核函数选择方法研究[J].湖南大学学报:自然科学版,2018,45(10):155-160.

[8] Mehmani A, Chowdhury S, Meinrenken C, et al. Concurrent surrogate model selection (COSMOS): optimizing model type, kernel function, and hyper-parameters[J]. Structural and Multidisciplinary Optimization, 2018, 57(3): 1093-1114.

[9] Jayadeva. Learning a hyperplane classifier by minimizing an exact bound on the VC dimension[J]. Neurocomputing, 2015, 149(3): 683-689.

[10] 刘学艺,宋春跃,李平.基于 Vapnik-Chervonenkis 泛化界的极限学习机模型复杂性控制[J].控制理论与应用,2014,31(5):644-653.

[11] 倪志伟,肖宏旺,伍章俊,等.基于改进离散型萤火虫群优化算法和分形维数的属性选择方法[J].模式识别与人工智能,2013,26(12):1169-1178.

[12] 邬啸,魏延,吴瑕.基于混合核函数的支持向量机[J].重庆理工大学学报:自然科学,2011,25(10):66-70.

[13] Wang Fuguang, He Ketai, Liu Ying, et al. Research on the selection of kernel function in SVM based facial expression recognition [EB/OL]. [2018-03-02]. <http://ieeexplore.ieee.org/document/6566586/>.

[14] 汪廷华,陈峻婷.核函数的度量研究进展[J].计算机应用研究,2011,28(1):25-28.

[15] 梁礼明,钟震,陈召阳.支持向量机核函数选择研究与仿真[J].计算机工程与科学,2015,37(6):1135-1141.

[16] 李倩倩,李春,杨阳.自相似性和分形维数在风场分析中的应用[J].动力工程学报,2016,36(11):914-919,926.

[17] 朱志宝,白永强.分形几何及其应用[J].价值工程,2012,31(35):5-7.

[18] 申文栋,梁静.基于分形维数的 GIS 局部放电信号发展过程分析[J].机电信息,2011(3):62-63,65.

[19] Grassberger P, Procaccia I. Measuring the strangeness of strange attractors[J]. Physica D: Nonlinear Phenomena, 1983, 9(1/2):189-208.

- [31] 江西省鄱阳湖水利枢纽建设办公室. 为了“一湖清水”:鄱阳湖水利枢纽工程介绍 [J]. 江西水利科技, 2013, 39(2): 83-91.
- [32] Vlietstra L S, Parga J A. Long-term changes in the type, but not amount of ingested plastic particles in short-tailed shearwaters in the southeastern Bering Sea [J]. Marine Pollution Bulletin 2002, 44(9): 945-955.
- [33] 黄晓平, 龚雁. 鄱阳湖渔业资源现状与养护对策研究 [J]. 江西水产科技 2007(4): 2-6.

## The Analysis on Composition and Source of Typical Coastal Litter Belt in Poyang Lake

LI Na<sup>1</sup>, ZHOU Hui<sup>1</sup>, FU Wenchang<sup>1</sup>, NI Caiying<sup>1\*</sup>, JIAN Minfei<sup>2</sup>

(1. College of Geography and Environment, Jiangxi Normal University, Nanchang Jiangxi 330022, China; 2. College of Life Science, Jiangxi Provincial Key Lab of Protection and Utilization of Subtropical Plant Resources, Jiangxi Normal University, Nanchang Jiangxi 330022, China)

**Abstract:** In order to understand the sources and composition of the litter in Poyang Lake, rubbish belts on the embankments of Poyang Lake are surveyed. Through the investigation on 20 garbage belts sample plot, the result finds that plastic garbage is the maximum, then followed by woven garbage, wood garbage, glass supplies, metal garbage, smoking articles and others. By composition analysis, the main source of sampled rubbish is fishing activities, the second sources of sampled rubbish are shoreline activities and other disposals of resident. These results can provide a scientific basis for the prevention and control of garbage in Poyang Lake, as well as decision-making basis for the water resources protection and regional socio-economic development planning of Poyang Lake.

**Key words:** Poyang Lake; coastal litter belt; plastic; source analysis

(责任编辑: 曾剑锋)

(上接第313页)

## The Selection of SVM Kernel Function Based on Fractal Theory

LIANG Liming, CHEN Mingli, LIU Bowen, WU Jian

(School of Electrical Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou Jiangxi 341000, China)

**Abstract:** As one of the research focuses in machine learning, Kernel method has been widely applied in classification, regression and some other fields due to its a great deal of advantages in dealing with nonlinear and high-dimensional data problems. Support Vector Machine (SVM) is the most representative method, and different kernel functions have different measurement features. Therefore, the selection of kernel function has an important influence on the generalization capability of SVM. However, kernel function selection is still an open problem at present, and there exists the contingency and limitations which the process reveals. In order to improve the generalization ability of SVM, the kernel function is selected by using the adjustment information of fractal geometry analysis data and the simulation results show that the method is effective and feasible.

**Key words:** kernel method; SVM; kernel function; fractal geometry; fractal dimension

(责任编辑: 冉小晓)