

文章编号: 1007-4619(2006)02-0191-06

基于 SVM 决策支持树的城市植被类型 遥感分类研究

张友静^{1,2}, 高云霄², 黄浩², 任立良¹

(1. 河海大学 水文水资源与水利工程国家重点实验室, 江苏 南京 210098;

2. 河海大学 遥感与地理信息系统研究所, 江苏 南京 210098)

摘 要: 城市植被类型不同, 生物量不同, 其生态功能与绿化效应也不同。在目前难直接获取城市“绿量”实测数据的情况下, 可以绿地面积和植被类型间接反映绿地的生物量和绿化效应。本文利用高分辨率卫星影像 KONOS 以实验区与验证区城市植被类型信息为对象, 在对常用的参数和非参数分类方法进行对比实验的基础上, 对 SVM 的核函数进行了分析, 构建了基于 SVM 决策树的城市植被类型分类模型。分类实验结果表明: 与其他传统方法分类结果比较, SVM 的决策树分类方法对植被类型的分类精度达到 83.5%, 绿化面积总精度接近 95%, 取得了良好的效果。

关键词: 城市植被类型; 高分辨率卫星影像; SVM 决策树; 遥感分类

中图分类号: TP79 文献标识码: A

Research on Remote Sensing Classification of Urban Vegetation Species Based on SVM Decision Making Tree

ZHANG You-jing^{1,2}, GAO Yun-xiao², HUANG Hao², REN Li-liang¹

(1. State Key Laboratory of Hydrology Water Resource and Hydraulic Engineering, Hohai Univ., Jiangsu Nanjing 210098, China;

2. Institute of RS & GIS, Hohai Univ., Jiangsu Nanjing 210098, China)

Abstract Different vegetation species have different biological quality and produce different ecological functions and greenery effect. Considering the “Vegetation Quality” is difficult to be obtained, the biological quality and ecological effect of urban greenery can be indirectly reflected using urban green-land area and vegetation species. Based on the comparison of the traditional statistic parameter and non-parameter classification methods and the analysis of kernel-function of SVM, SVM decision making tree model for urban vegetation classification is designed in this paper using the high resolution imagery data of KONOS. The classification results are compared to other traditional methods and have an average vegetation classification accuracy of about 83.5% and green-land area accuracy nearly 95%.

Key words urban vegetation species; high resolution imagery; SVM decision making tree; remote sensing classification model

1 引 言

绿地面积为指标的城市绿化评价标准不能确切

反映城市绿化水平^[1,2]。城市植被类型不同, 生物量不同, 其生态功能与绿化效应也不同^[3]。在目前难直接获取城市“绿量”实测数据的情况下, 可以绿地面积和植被类型间接反映绿地的生物量和绿化

收稿日期: 2004-09-03; 修订日期: 2005-01-20

基金项目: 国家自然科学基金项目(40171016)资助。

作者简介: 张友静(1955—), 男, 教授, 毕业于南京大学地理系, 现从事遥感、GIS 教学与科研工作, 已发表遥感方面论文 10 余篇。

效应。

城市植被类型的卫星遥感调查研究开展不多。低分辨率遥感影像进行城市植被类型调查难以取得理想结果。高分辨率传感器技术的长足发展,为城市植被类型遥感调查提供了位置更精确、地物识别率更高的信息,从而扩大了遥感信息应用的深度和广度。因此,高分辨率卫星影像的城市植被类型信息提取,越来越受到人们的高度重视。

相对于高分辨率传感器的飞速发展,基于高分辨率遥感影像的自动分类算法发展相对迟缓。目前城市植被遥感的自动分类算法中,传统的统计模式识别分类方法,是根据遥感数据的统计特征与训练样本数据之间的统计关系来进行地物分类的。由于城市植被类型分布方式本身的复杂性,以及城市绿地破碎、植被类型多样、林相不齐等特点,难以获取较为充分和完备的样本集和先验知识,同时因统计参数型分类方法自学习机制的缺乏,致使计算机分类面临着诸多模糊对象,导致其分类精度不甚理想。

20世纪90年代中期以来,基于光谱特征空间分布的非参数型计算机分类方法已逐步成熟。近年来在有限样本情况下的机器学习理论研究逐渐兴起,形成了一个比较完善的理论体系,即统计学习理论(Statistical Learning Theory, SLT)。随着该理论的不断发展和成熟,也由于神经网络等学习方法在理论上缺乏实质性进展,SLT开始受到广泛的重视^[4-6]。在这一理论基础之上,发展了一种新的通用学习方法——支持向量机(Support Vector Machine, SVM),它已初步表现出很多优于已有方法的性能^[7]。Zhu Guobin & Dan G. Blumberg et al. 利用基于SVM的算法对ASTER传感器获得的数据进行了分类处理,分类结果表明,基于SVM的方法在收敛性、训练速度、分类精度等方面具有较高的性能^[8]。李侃等将SVM引入到空间数据库,通过构造二叉树对多类问题进行层次聚类分析^[9]。赵书河将SVM算法引入到决策级多源遥感影像的融合研究,取得了较好效果^[10]。但到目前为止,尚未见有将SVM决策树分类方法用于城市高分辨率卫星影像植被类型分类的系统报道。

本文利用高分辨率卫星影像,以实验区与验证区城市植被类型信息为对象,在对常用的参数和非参数分类方法进行对比实验的基础上,对SVM的核函数进行了分析,构建了基于SVM决策树的城市植被类型分类模型。实验表明该方法具有较高的分类精度和空间稳定性。

2 基于 SVM 决策树的城市植被类型分类模型的构建

2.1 SVM 决策树分类模型

SVM用于模式识别的基本思想是,构造一个超平面作为决策平面,使正负模式之间的空白最大^[10]。两类模式识别问题的最优分类超平面为:

$$\omega \cdot z + b = 0$$

式中, z_i 为一组样本:

$$\{(z_1, y_1), (z_2, y_2), \dots, (z_k, y_k)\}, z \in R^n, y \in \{-1, +1\}$$

k 为样本数, n 为输入的维数; ω 和 b 为非零常数。

2.1.1 线性可分模型

对于线形可分的情况下,求分类间隔最大的分类超平面即为最优分类超平面。也即在约束条件下求最佳 (ω, b)

$$y_i(\omega \cdot z_i + b) \geq 1 \quad i = 1, 2, \dots, k$$

的最小值

$$\frac{1}{2} \|\omega\|^2 \quad (1)$$

这是一个凸二次规划问题,因此存在全局惟一的最优解^[11]。引入 Lagrange 函数,并根据鞍点定理,由 Kuhn-Tucher 定理可得,最优解满足的条件为:

$$a_i(y_i(\omega \cdot x_i + b) - 1) = 0 \quad i = 1, 2, \dots, k \quad (2)$$

式中, a_i 为与每个样本对应的 Lagrange 乘子,且不为 0; 对应的样本 x_i 就是支持向量; w 为:

$$\omega^* = \sum_{i=1}^k a_i^* y_i z_i$$

则分类间隔最大化目标函数为:

$$w(a) = \sum_{i=1}^k a_i - \frac{1}{2} \sum_{i,j=1}^k a_i a_j y_i y_j (x_i \cdot x_j)$$

式中, a_i 非负, $\sum_{i=1}^k a_i y_i = 0$

根据式(2)得: $b^* = y_i - \omega^* \cdot z_i$

得到的最优分类面的函数表达式为:

$$f(z) = \text{sign}\{(\omega \cdot z) + b\} = \text{sign}\left\{\sum_{i=1}^k a_i^* y_i (z \cdot z_i) + b^*\right\}$$

式中的求和只对支持向量进行。 b^* 是分类阈值,可以用任意一个支持向量求得。

2.1.2 非线性分类面构建

为了将线性 SVM 推广到非线性分类面的情况, Vapnik 提出了核函数(Kernel Function)的概念,其基本思想是:通过事先确定的非线性映射将输入向量映射到一个高维的特征空间中,然后在高维特征空间中构建最优超平面^[12](Vapnik, 1995; 张学

工, 2000)。首先将输入向量 z 通过映射 φ 映射到高维特征空间 H 中:

$$\varphi: R^n \rightarrow H$$

定义核函数, 该核函数 k 满足:

$$k(z_i, z_j) = \varphi(z_i) \cdot \varphi(z_j)$$

则决策函数变为:

$$f(z) = \sum_{\text{sup vector}} a_i^* y_i \varphi(z_i) \varphi(z) + b^* \\ = \sum_{\text{sup vectors}} a_i^* y_i k(z_i, z) + b^*$$

通过构造核函数, 避免了直接在高维空间中对映射后的样本进行操作, 减少了计算量, 以线性的代价获得了非线性的优良特性。根据核函数的不同形式, 可以生成不同的支持向量机。在构造判别函数时, 先在输入空间比较向量, 再对结果作非线性变换。

2.1.3 SVM 决策函数计算过程

SVM 决策函数形式上类似于一个神经网络, 输出是中间节点的线性组合, 每个中间节点对应一个支持向量, 如图 1 所示。

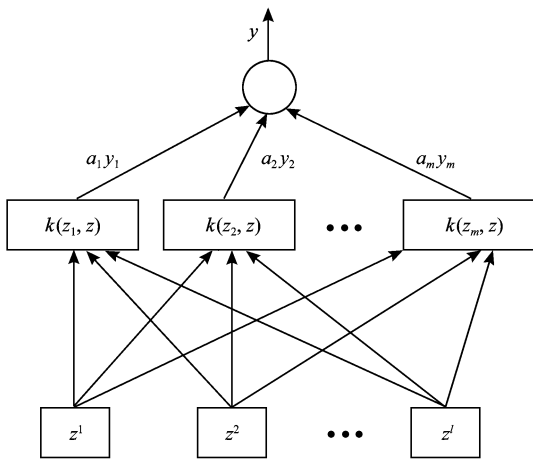


图 1 支持向量机示意图

Fig 1 SVM program

决策函数:

$$y = \text{sign} \left[\sum_{i=1}^m a_i y_i k(z_i, z) + b \right]$$

权值 $a_i y_i$

基于 m 个支持向量 z_1, z_2, \dots, z_m 的非线性变换 (内积)

输入向量 $z = (z^1, z^2, \dots, z^l)$

式中, 常用的核函数包括:

(1) 多项式核函数: $k(z, z_i) = (z \cdot z_i + 1)^d$

(2) 径向基核函数: $k(z, z_i) = \exp \left[-\frac{\|z - z_i\|^2}{\sigma^2} \right]$

(3) 神经网络核函数: $k(z, z_i) = k \cdot (z \cdot z_i) - \theta$

2.1.4 基于 SVM 决策树的分类

决策树是遥感图像分类中的一种分层处理结构。其基本思想是, 通过一些判断条件, 对原始数据集逐步进行二分和细化。其中, 每一个分叉点代表一个决策判断条件, 每个分叉点下有两个叶节点, 分别代表满足和不满足条件的类别。逐步从原始影像中分离每一种目标作为一个图层或树枝, 避免此目标对其他目标提取时造成干扰及影响, 最终复合所有的图层以实现图像的自动分类。这种方法不需要依赖任何先验的统计假设条件, 并可通过聚类等方法获得先验知识, 提高训练速度等, 这对城市绿地的遥感分类具有重要意义。

将支持向量机和决策树的基本思想结合起来可构成多类别的识别器。对于大类别分类问题, 可将其有效地转化为小类别识别问题的组合, 以期降低错误识别率。 $N-1$ 分类问题 ($N > 2$) 和两分类问题之间存在一定的对应关系。如果一个分类问题 N 类可分, 则这 N 类中的任何两类间一定可分。通过一定的组合法则, 通过两两可分实现 N 类可分。

2.2 基于 SVM 决策树的城市植被类型分类模型的构建与分析

2.2.1 核函数对城市植被分类的影响分析

构建城市植被类型分类模型的核心在于分析核函数对城市植被分类的影响。用三种核函数构造的不同分类器, 对实验区的植被进行了对比分类实验, 分类精度均在 97% 左右。本文选择利用多项式核函数进行分类。

对多项式核函数的参数 d 与分类精度之间的关系进行分析, 结果如图 2 所示。参数 d 的选取范围为 5-7 时分类精度可以达到最大值, 在其他情况下精度有较大幅度降低。

基于该核函数训练样本数目对分类结果精度的影响如图 3 所示。由图可见, 随着训练样本的增

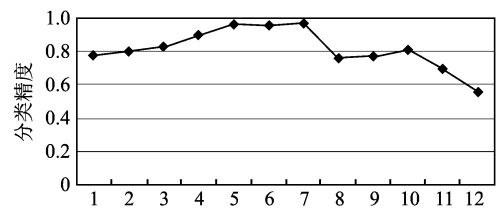


图 2 多项式核函数分类精度随参数 d 的变化

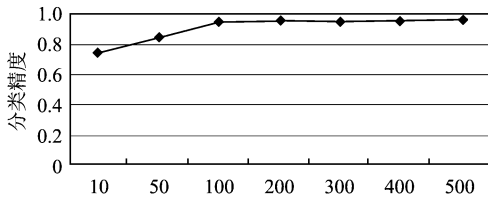


图 3 不同训练样本数量与分类精度关系图

Fig 3 The relation between the classification accuracy and the number of samples

加,分类精度提高较快,但增加到 100左右时趋于稳定。表明在一定的先验知识下,SVM分类可以少量样本得到现有信息下最优解。这在有限样本的城市植被类型分类中具有重要意义。

2.2.2 基于高分辨率遥感影像的城市植被分类模型构建

基于上述思想可构建遥感影像的城市植被分类模型。图 4是该模型的基本流程。

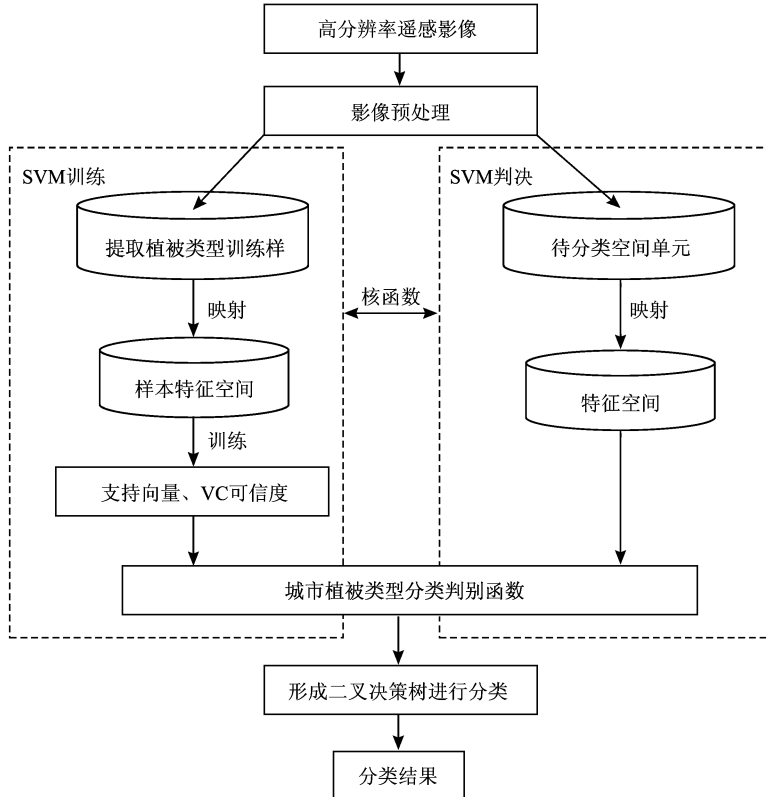


图 4 基于 SVM 决策树的城市植被类型分类模型

Fig 4 Model of urban vegetation species classification based on SVM decision making tree

图 4 的分类步骤如下:

(1)预处理过程:首先对影像进行几何精纠正和辐射校正。根据调查资料,将城市植被分成阔叶林、针叶林和草三类。抽取训练样本并对输入样本进行归一化和格式化处理。

(2)SVM 训练过程:选择 5 次方的多项式核函数,将训练样本映射到高维特征空间。利用 SVM 在样本特征空间中找出各类别特征样本与其他特征样本的最优分类超平面,得到代表各样本特征的支持向量集及其相应的 VC 可信度,形成判断各特征类别的判别函数。

(3)SVM 判决过程:将影像中待分类像元通过

核函数作用映射到特征空间中,作为判别函数的输入,利用分类判决函数得出二类可分的结果。

(4)二叉决策树分类过程:在 SVM 二类可分的基础上形成二叉决策树,利用分类判决函数得出多类可分的结果,最后通过竞争判断输入单元所属的样本特征类别。

(5)输出分类结果图。

3 基于 SVM 决策树的城市植被类型分类的实验与分析

本文以南京市主城区 IKONOS 影像为例,分别

选择了实验区和验证区。采用参数方法的最大似然、最小距离法和非参数的人工神经网络、SVM 决策树方法对针叶树、阔叶树和草地进行分类,并按针、阔、草对各区抽取 200 个像元点进行验证,结果见表 1。

表 1 常用分类方法与基于 SVM 的决策树方法分类精度比较
Table 1 Accuracy comparison of SVM decision making tree and other classification approach

分类方法	分类总精度 /%	Kappa 系数统计值
1. 最大似然分类	65.5	0.53
2. 最小距离分类	66.5	0.55
3. 神经网络分类	74.5	0.62
4. SVM 决策树分类	83.5	0.71

由表 1 可见,参数方法的分类精度较为稳定,其精度与同类研究相似;神经网络方法因其参数(如权重)设定等因素,对精度影响较大,不能轻易下结论。相比之下,SVM 决策树方法精度相对较高,表 2

和表 3 给出了实验区该方法的分类混淆矩阵和精度评价结果,分类结果见图 5。

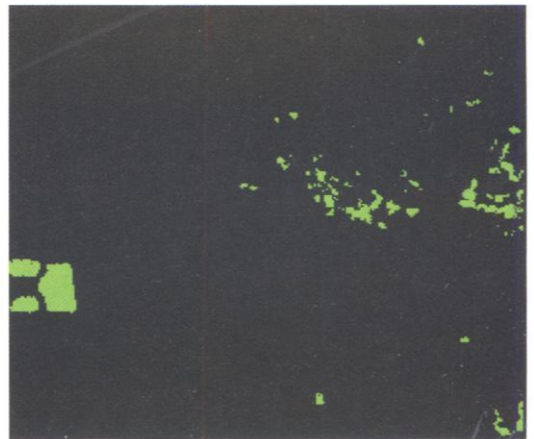
由表 2 和图 5 可见,林和草两类的分类精度均在 90% 以上;林草合一,即绿地总精度近 95%;而针、阔的精度相对不高,主要是由于居民区绿地破碎、植被类型多样以及行道树针、阔混种等情况导致针、阔错分类所致(图 5)。对于南京市的绿地类型,采用多时相遥感数据,可望有较好的效果。

表 2 实验区基于 SVM 的决策树分类混淆矩阵
Table 2 Classification confusion matrix of experiment section using SVM decision making tree

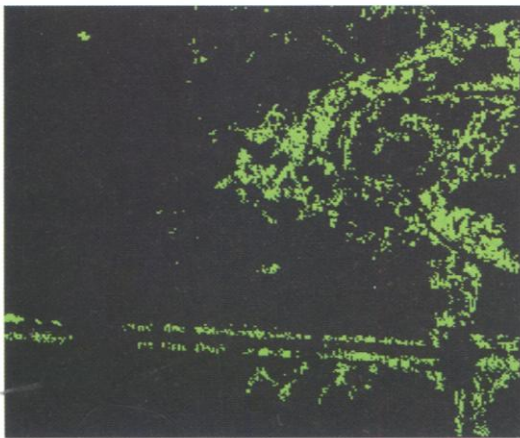
类别	阔叶林	针叶林	草	总量	精度 /%
阔叶林	58	10	1	69	84.06
针叶林	15	62	2	79	78.48
草	1	6	45	52	86.54
总量	74	78	48	200	
精度 /%	林: 95.39	草: 93.75	林草总精度: 95.0		



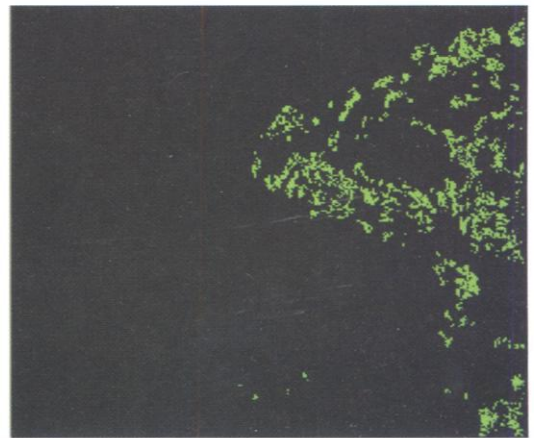
(a) 分类实验区影像



(b) 草的分类结果图



(c) 针叶林的分类结果图



(d) 阔叶林的分类结果图

图 5 实验区的植被分类结果图

Fig 5 Vegetation classification result in experiment section

表 3 验证区基于 SVM 的决策树分类混淆矩阵

Table 3 Classification confusion matrix of validating section using SVM decision making tree

类别	阔叶林	针叶林	草	总量	精度 /%
阔叶林	59	10	3	72	81.94
针叶林	13	60	5	78	76.92
草	2	2	46	50	92.00
总量	74	72	54	200	
精度 /%	林: 94.67	草: 92.00	林草总精度: 94.0		

为了验证该方法的可靠性与空间稳定性,选择了该城市另一区域作为验证区。表 3 是验证区的分类混淆矩阵。

对比实验区和验证区各类植被分类情况(表 2 与表 3)可知,在实验区与验证区,该方法分类精度变化不大,说明了该算法具有一定的空间稳定性。

4 结 论

绿地面积为指标的城市绿化评价标准,不能确切反映城市绿化水平,城市植被类型不同,生物量不同,其生态功能与绿化效应也不同。但在目前难直接获取城市“绿量”实测数据的情况下,可以绿地面积和植被类型间接反映绿地的生物量和绿化效应。

高分辨率遥感影像为城市植被信息提取提供了有利条件。但因城市植被类型分布方式本身的复杂性,以及城市绿地破碎、植被类型多样、林相不齐等特点,传统的遥感分类算法不能满足要求。本文在分析 SVM 决策树分类方法的基础上,对 SVM 函数表达式的核函数进行了分析,构建了基于 SVM 决策树的植被类型分类模型。通过实验区与验证区的调查与分类实验,该方法的分类精度达到 83.50%,绿地总精度接近 95%,取得了很好的效果。若进一步利用多时相、多平台高分辨率卫星影像数据,其分类精度有望进一步提高。

参 考 文 献 (References)

[1] Lu M E. Application of Remote Sensing on Urban Green [J].

China Landscape Arch, 2000, (5): 41-44 [吕妙儿. 城市绿地监测遥感应用 [J]. 中国园林, 2000, (5): 41-44.]

[2] Zhou J H. Theory and Practice on Database of Three-dimension Vegetation Quality [J]. Acta Geographica Sinica, 2001, 56(1): 14-23 [周坚华. 城市绿量测算模式及信息系统 [J]. 地理学报, 2001, 56(1): 14-23.]

[3] Zhu W Q. et al. Quantitative Analysis of Urban Forest's Structure [J]. Chinese Journal of Applied Ecology, 2003, (14): 2090-2094 [朱文泉等. 城市森林结构的量化研究 [J]. 应用生态学报, 2003, (14): 2090-2094.]

[4] Cherkassky V, Mulier F. Learning from Data: Concepts, Theory and Methods [M]. NY: John Wiley & Sons, 1997.

[5] Vapnik V N. The Nature of Statistical Learning Theory [M]. Springer-Verlag, 1995 [张学工译. 统计学习理论的本质 [M]. 北京:清华大学出版社, 1999.]

[6] Christoph B Puhf et al. Modeling to Map Canopy Closure in Conifer Plantations: a Case Study Using Landsat TM Data from S W. Scotland [A]. Remote Sensing Society Observation & Internations [C]. New York: The Remote Sensing Society, 1997.

[7] Cortes C, Vapnik V. Support Vector Networks [J]. Machine Learning, 1995, 273-297.

[8] Zhu G B, Dan G, Blumberg. Classification Using ASTER Data and SVM Algorithms: The Case Study of Beer Sheva, Israel. Remote Sensing of Environment, Volume 80, Issue 2, May 2002, 233-240.

[9] Li K. An Approach for Attribute Reduction Based on Rough Set Theory [J]. Computer Engineering and Applications, 2002, (5): 15-16 [李侃. 一种粗糙集属性约简算法 [J]. 计算机工程与应用, 2002, (5): 15-16.]

[10] Zhao S H. Multi-source Remote Sensing Image Fusion Based on Decision-level and Its Application [D]. Nanjing: Nanjing University, 2003 [赵书河. 多源遥感影像决策级融合及其应用研究 [D]. 南京:南京大学图书馆, 博士论文, 2003.]

[11] Wang J F, Cao D Y. The Application of Support Vector Machine Classifying Large Number of Catalogs [J]. Journal of Beijing Institute of Technology, 2001, 21(2): 225-228 [王建芬, 曹大元. 支持向量机在大类别数分类中的应用 [J]. 北京理工大学学报, 2001, 21(2): 225-228.]

[12] Zhang X G. Introduction to Statistical Learning and Support Vector Machines [J]. Acta Automatica Sinica, 2000, 26(1): 37-38 [张学工. 关于统计学习理论与支持向量机 [J]. 自动化学报, 2000, 26(1): 37-38.]