支持向量机理论综述

Survey of Support Vector Machine Theory

萧 嵘 王继成 张福炎
(南京大学软件新技术国家重点实验室 计算机科学与技术系 南京 210093)

Abstract Support Vector Machine(SVM)is an effective and general method for representing complex functions in high dimensional space. It has been used successfully in many traditional neural network dominated applications. This paper introduces the development status of SVM, describes the main theory and its applications.

Keywords SV, SVM, Categorization, Function simulation, Pattern recognition

1 引言

随着信息技术的发展,人类积累的数据量呈几何级数递增。如何从这些数据中发掘出有效的信息而不被信息海洋所淹没已经成为迫在眉睫的问题。对数据挖掘相关领域的研究,例如数据分类、聚类,函数模拟和规则抽取等技术,逐渐成为当前计算机基础技术研究的重心之一。支持向量机(Support Vector Machine, SVM)方法作为一种解决多维函数预测的通用工具,一经提出,便成功地应用于函数模拟、模式识别和数据分类等领域,并取得了极好的效果,成为当前国际上的一个研究热点。

SVM 理论源于 Vapnik 在 1963 年提出的用于解决模式识别问题的支持向量方法[1]。这种方法从训练集中选择一组特征子集,使得对特征子集的线性划分等价于对整个数据集的分割。这组特征子集称为支持向量 SV。在此后近 30 年中,对 SV 的研究主要集中在对分类函数的改进和函数预测上。在 1971 年,Kimeldorf 提出使用线性不等约束重新构造 SV 的核空间[2],解决了一部分线性不可分的问题,为以后的 SVM 研究开辟了道路。1990 年,Grace,Boser 和 Vapnik 等人开始对 SVM 技术进行研究,并取得突破性进展[3,4]。1995 年,Vapnik 提出了统计学习理论,较好地解决了线性不可分的问题,正式奠定了 SVM 的理论基础[5]。

2 支持向量机理论

支持向量机的理论最初来自对数据分类问题的处理。对于数据分类问题,如果采用通用的神经网络方法来实现,其机理可以简单地描述为:系统随机产生一个超平面并移动它,直到训练集中属于不同分类的点正好位于平面的不同侧面。这种处理机制决定了:用神经网络方法进行数据分类最终获得的分割平面将相当靠近训练集中的点,而在绝大多数情况下,并不是一个最优解。为此 SVM 考虑寻找一个满足分类要求的分割平面,并使训练集中的点距离该分割平面尽可能地远,即寻找一个分割平面,使其两侧的空白区域(margin)最大。如图 1 所示。

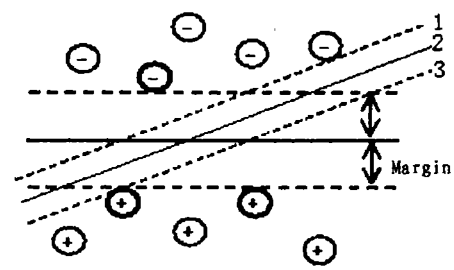

图 1 数据点集的超平面分割

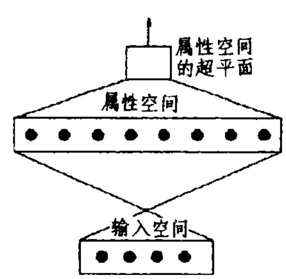

图 2 SVM 的工作原理

2.1 优化分割的超平面

对于图 1 中的数据分类问题，给定样本点：$(x_1, y_1), \cdots, (x_l, y_l), x_i \in R^n, y_i \in \{-1, +1\}$，SVM 用如下的方式定义优化分割的超平面：

构造分割平面：$(w \cdot x) + b = 0$ (1)

使得：$\begin{cases} (w \cdot x_i) + b \geqslant 1 & , y_i = 1 \\ (w \cdot x_i) + b \leqslant -1, & y_i = -1 \end{cases} \Leftrightarrow$

$$y_i[(w \cdot x_i) + b] \geqslant 1 \quad (i = 1, \cdots, l) \quad (2)$$

可以计算出，训练数据集到一给定的分割平面的最小距离为：

$$p(w,b) = \min_{\{x_i | y_i = 1\}} \frac{w \cdot x_i + b}{|w|} - \max_{\{x_i | y_i = -1\}} \frac{w \cdot x_i + b}{|w|} = \frac{2}{|w|} \quad (3)$$

根据 SVM 对优化分割平面的定义，可以看出对该平面的求解问题可以简化为：在满足条件式(2)的情况下，计算能最大化 $p(w,b)$ 的分割平面的法向量 w 和偏移量 b。Vapnik 等人证明：

(1)分割超平面的法向量 w_0 是所有训练集向量的线性组合。即 w_0 可以描述为：

$$w_0 = \sum\nolimits_{i=1}^{l} (a_i^0 y_i) x_i，其中，(a_i^0 \geqslant 0), i = 1, \cdots, l \quad (4)$$

(2)测试集的分类函数可以描述为：

$$f(x) = sign(\sum_{SVS} y_i a_i^0 (x_i \cdot x) - b_0) \quad (5)$$

其中，$b_0 = [(w_0 \cdot x^*(1)) + (w_0 \cdot x^*(-1))]/2$

在多数情况下，(4)式 w_0 的展开式中，系数 a_i^0 为零值，而非零值的 a_i^0 对应的 x_i 就称为支持向量 SV。这些向量充分描述了整个训练数据集数据的特征，使得对 SV 集的线性划分等价于对整个数据集的分割。

2.2 支持向量机

在很多情况下，训练数据集中的数据是线性不可分的，这使得 SV 的应用受到了很大的限制。为了解决这个问题，Vapnik 等人提出使用 SVM 作为超平面分割方法的扩展。使用 SVM 进行数据集分类工作的典型流程如图 2 所示。首先，通过预先选定的一些非线性映射将输入空间映射到高维属性空间，使得在高维属性空间中有可能对训练数据实现超平面的分割，避免了在原输入空间中进行非线性曲面分割计算。SVM 数据集形成的分类函数具有这样的性质：它是一组以 SV 为参数的非线性函数的线性组合，因此分类函数的表达式仅和 SV 的数量相关，而独立于空间的维度。在处理高维输入空间的分类时，这种方式尤其有效。

2.3 维数爆炸的控制

SVM 在处理非线性可分的问题时需要将低维的输入空间映射到高维的属性空间去，当使用不同的映射函数，生成的属性空间的维度变化很大，例如，使用 m 次多项式映射，将原有的 n 维输入空间映射到 $O(n^m)$ 维空间，如果对每一输入向量在高维属性空间的像进行计算的话，计算量将很大，SVM 通过定义核心函数来解决这一问题。

可以注意到，在分类函数式(5)中，空间映射函数 $k(x_i)$ 的具体形式不需要知道，分类函数系数的计算只涉及到像空间向量的内积。为了避免对每一输入向量在高维属性空间的像进行计算，SVM 定义属性空间向量内积为以下形式：

$$K(x_i, x_j) = k(x_i) \cdot k(x_j) = \sum_{k=1}^{\infty} a_k \Psi_k(x_i) \Psi_k(x_j) \quad (6)$$

$K(x_i, x_j)$ 刻画了所用的映射函数的特征，被称为核心函数。当 $K(x_i, x_j)$ 函数已知时，就可以避免计算每一个 $k(x)$，从而使得计算量和像空间的维度无关。

2.4 多维函数模拟

SVM 除了用于数据分类领域外，在函数模拟领域的应用也非常成功。假定根据某种概率分布 $P(x,y)$，$(x \in R^n, y \in R)$ 生成采样点集：$X = \{(x_1, y_1), (x_2, y_2), \cdots, (x_l, y_l)\}$，$x$ 和 y 存在函数依赖关系：$F = \{f | f: R^n \mapsto R\}$，函数模拟问题可以描述为如何寻找一个函数 $f \in F$ 使得下面代价函数最小：

$$R[f] = \int l(y - f(\vec{x}), \vec{x}) dP(\vec{x}, y) \quad (7)$$

其中，$l(y - f(\vec{x}), \vec{x})$ 是误差函数。

SVM 中定义：

$$l(y - f(\vec{x}), \vec{x}) = |y - f(\vec{x}, w)|_\varepsilon = \begin{cases} 0 & if |y - f(\vec{x}, \vec{w})| \leqslant \varepsilon \\ |y - f(\vec{x}, \vec{w})| - \varepsilon & 其它 \end{cases} \quad (8)$$

并考虑映射函数为一组基函数的线性组合，即：$f(x, a, a^*) = \sum_{i=1}^{l} (a_i^* - a_i)(\Phi(x_i) \cdot \Phi(x)) + b$(其中 $a_i^* \geqslant 0$, $a_i \geqslant 0$, $a_i^* \cdot a_i = 0$)。使用核心函数 $K(x_i, x) = \Phi(x_i) \cdot \Phi(x)$，可以通过条件极小化 $R[f]$ 确定 a_i 和 a_i^*。

3 SVM 的研究与应用

在数据分类和函数模拟领域，SVM 方法和常见的神经网络算法比较，具有训练时间短、分类模拟效果好的优点。目前对 SVM 的研究和应用主要包括以下几个方面。

3.1 算法的改进

对算法的改进是目前 SVM 研究的主要内容。这些研究主要集中在以下方面：

(1)降低算法的运算量：如 S. S. Keerthi[6] 将 SVM-NV 分类问题转化为计算两训练集凸包的最近点，并成功地将之运用到其他类型 SVM 分类问题。

(2)自调整算法的研究：如 Cristianini[10] 引入带自由参数的核心函数，使得核心函数可以根据训练数据集进行自适应调整，提高了分类算法的有效性；又如 Koji Tsuda[7] 通过使用带自由加权参数的欧几里德距

离对 SVM 超平面分类器进行优化。

(3)噪音数据的处理:如 Pontil[9]提出了一个噪音模型来解决 SVM 回归分析中的噪音处理问题。

3.2 多值分类

在分类问题上,SVM 方法的基本理论只考虑了二值分类这一最简单的方法,因而在多值分类情况下,系统需要组合多个 SVM 进行分类。一般有以下两种方案:

(1)一一区分模式(one against one model):在这种模式下,为对 n 个分类的训练集进行两两区分,分别为之构造分类函数(共 $n(n-1)/2$ 个)。在测试时,使用成对的 SVM 进行鉴别比较,每次产生一个优胜者,然后优胜者间继续进行竞争淘汰,直至最后仅剩唯一一个优胜者。测试数据的类别由这个 SVM 决定。

(2)逐一鉴别模式(one against all model):仅构建 n 个 SVM,每一个 SVM 分别将某一分类的数据从其他分类数据集数据中鉴别出来。这种方式的缺点在于,一些测试数据的分类将有可能出现多义性,即被同时分入多个类。另外,在这种分类模式下,每个 SVM 的分类难度增加,将不得不使用高阶核心函数映射到更高维空间进行超平面划分,使得分类效率大大下降。

3.3 手写体识别和 3D 对象识别

手写体识别和三维对象识别是目前计算机研究的热点之一。通常采用统计决策、结构分析和神经网络算法等方法实现,这些算法对特定领域的识别比较有效,很难作为一种通用的识别算法得到应用,另外识别率普遍不高。

早在七十年代初,SV 分类算法便被应用于模式识别领域,但由于其处理线性不可分问题的局限性,效果并不很好。进入九十年代后,SVM 技术逐渐成熟,成为一个性能良好的通用分类算法。这样,如何将其成功地应用到模式识别中去,便成为新的研究方向。Boser,Guyon 等人[5]使用多项式核心函数对 16 * 16 的手写体阿拉伯数字进行识别取得了很大成功。在实验中,他们采用了 600 个数字位图进行训练,然后用另外 600 个位图进行测试。实验结果显示,当采用二阶多项式核心函数时,错误率下降到 1.5%,稍好于其他神经网络算法(具有复杂结构的五层神经网络)的结果。进一步研究显示,采用更为光滑的核心函数(如高斯函数),误差将进一步减少到 0.3%。

三维对象识别是新兴的模式识别技术,是机器人视觉、三维重建等应用领域的核心技术,Danny Roobaert[8]等人将 SVM-VL 方法应用到三维对象识别技术上,取得很好的效果。实验结果显示,如果不考虑对识别对象的先验知识的话,其识别准确率远优于其他方法。

结束语 SVM 方法拥有完备的数学理论基础,是一种在高维空间表示复杂的函数依赖关系的高效通用手段,适用于函数预测、模式识别和数据分类领域。和其他同类方法相比,SVM 具有适应性强、效率高的特点。作为一种新兴技术,SVM 还有很多应用领域有待探索。目前对 SVM 方法的研究主要集中在 GMD/MPIK、Lucent、AT&T、RHUL 和 MIT 等五家研究机构,国内的研究尚处于起步阶段。

参 考 文 献

1 Vapnik V, Lerner A. Pattern Recognition using Generalized Portrait. Automation and Remote Control, 1963,24:6

2 Kimeldorf G, Wahba G. Some results on Tchebycheffian spline functions. J. Math. Anal. Applic., 1971,33(1):82~95

3 Wahba G. Spline Models for Observational Data(book). SIAM, CBMS-NSF Regional Conference Series, V59. 1990

4 Boser B, et al. A training algorithm for optimal margin classifiers. Fifth Annual Workshop on Computational Learning Theory. ACM Press, Pittsburgh. 1992

5 Vapnik V. The Nature of Statistical Learning Theory. Springer-Verlag, New York. 1995

6 Keerthi S S, et al. A fast iterative nearest point algorithm for support vector machine classifier design:[Technical Report No. TR-ISL-99-03]. Dept of CSA, IISC, Bangalore, INDIA. 1999

7 Tsuda K. Optimal Hyperplane Classifier with Adaptive Norm:[ETL technical report TR-99-9]. 1999

8 Roobaert D, et al. View-based 3D object recognition with Support Vector Machines. In: Proc. IEEE Neural Networks for Signal Processing Workshop. 1999

9 Pontil M, et al. On the Noise Model of Support Vector Machine Regression, CBCL Paper #168, AI Memo #1651, Massachusetts Institute of Technology, Cambridge, MA, October 1998

10 Cristianini N, et al. Dynamically Adapting Kernels in Support Vector Machines Neuro COLT:[Technical Report TR-1998-017]. Royal Holloway College