

增量支持向量机核函数的优化^①

李村合, 马敏敏

(中国石油大学(华东) 计算机与通信工程学院, 青岛 266580)

摘要: 支持向量机的核函数类型分为两类: 局部核函数和全局核函数. 局部核函数的值只受到相距很近数据点的影响, 有很好的学习能力. 全局核函数的值会受到距离较远数据点的影响, 有很好的推广泛化能力. 针对局部核函数学习能力良好但泛化能力差的缺点, 提出一种结合局部核函数和全局核函数构造新联合函数的方法. 实验结果表明, 与局部核函数和全局核函数相比, 新联合核函数有更好的预测能力, 并且能够适应增量学习的过程.

关键词: 支持向量机; 增量学习; 全局核函数; 局部核函数; 联合核函数

引用格式: 李村合, 马敏敏. 增量支持向量机核函数的优化. 计算机系统应用, 2017, 26(8): 284-287. <http://www.c-s-a.org.cn/1003-3254/5957.html>

Optimization of Kernel Function in Incremental Support Vector Machine

LI Cun-He, MA Min-Min

(College of Computer and Communication Engineering, China University of Petroleum, Qingdao 266580, China)

Abstract: The Kernel function to support vector machines can be divided into two types: the local kernel function and the global kernel function. Because the local kernel function has excellent learning ability, but its generalization ability is limited, we structure a joint kernel function with two kinds of functions, so that it can combine the advantages of the two kinds of kernel functions. The experiment proves that the joint kernel function can adapt to the incremental learning process and it has better performance.

Key words: support vector machine; incremental learning; global kernel function; local kernel function; joint kernel function

引言

解决线性不可分问题是支持向量机的一个优势. 支持向量机引入非线性特征映射解决了非线性分类的问题, 即将待分类的数据从低维输入空间映射到高维特征空间. 在算法中, 这个非线性映射通过适当的核函数实现, 而且不需要计算非线性变换后的具体形式, 只要核函数满足 Mercer 条件.

2015 年 08 月重庆师范大学邬啸等人提出将混合核函数应用于支持向量机^[1], 用组合核函数的支持向量机分类器的整体性能得到了进一步的提高, 验证了相对于单独使用的核函数, 组合核函数很有优势. 但文中提出的组合核函数权值为变量, 需要对其做进一步的

定量分析. 本文提出一个依赖数据构建核函数的方法. 找到分类空间中的边界区域, 构造相应的核函数. 不需要对所有数据计算核函数.

1 核函数的主要形式

核函数有多种形式, 不同的核函数表现的特点也各有不同, 相应的由它们构建的支持向量机也会有不同的性能. 当前, 支持向量机的核函数类型分为两类: 局部核函数^[4]和全局核函数^[5]. 局部核函数的值只受到相距很近数据点的影响, 有很好的学习能力^[2]. 全局核函数的值会受到距离较远数据点的影响, 有很好的推广泛化能力^[3]. 核函数就是存在一个非线性的变换 $\phi(\cdot)$,

^① 收稿时间: 2016-12-02; 采用时间: 2017-01-20

使 $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$ 成立. 主要的核函数形式有如下几种.

(1) 多项式核函数^[6]:

$$K(x, x_i) = (x \cdot x_i + c)^d \quad (1)$$

其中, c 为多项式偏差值, d 为多项式阶数. c 值的一般经验取值为 1. 准确率会随着 d 的增加而提高, 但多项式阶数的增加会使特征空间的维数迅速增加, 导致支持向量机的数目增多, 从而使运算时间复杂度增加. 综合考虑, 多项式的阶数多取 2、3 阶. 多项式核函数具有较好的全局性质. 全局核函数允许远离测试的数据点对核函数值也有影响.

(2) 径向基核函数(RBF)^[7]:

$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \quad (2)$$

σ^2 的选择值对使用径向基核函数的支持向量机分类结果有显著影响, σ^2 取值不当会导致严重的“欠学习”或者“过学习”现象. 过学习即 σ^2 趋近于 0 时, 训练样本能够被正确分开, 但对测试样本的泛化能力趋近于无. 欠学习即当 σ^2 趋近于无穷大时, 训练样本会被划分到样本数较多的一类. 从上式中可以看出 σ^2 的大小相对 $\|x - x_i\|^2$ 来计算的. 在实际应用中, 只要 σ^2 的取值比训练样本之间的最小距离小得多就可以达到 σ^2 趋向于 0 的效果. 相同的, 只要 σ^2 的取值比训练样本之间的距离大得多就可以达到 σ^2 趋向于无穷的效果. 则 σ^2 的可行搜索范围如下:

$$[\min \|x - x_i\|^2 * 10^{-2}, \max \|x - x_i\|^2] \quad (3)$$

上式只是一种定性分析, 取值需要实验反复验证. 无论维数高低或者样本多少, 径向基核函数都有较好的收敛. 径向基函数具有很好的局部性, 而局部核函数仅对测试点附近小领域内对数据点有影响. 同时, RBF 核函数只对样本分布比较集中的数据点作用明显, 虽然它的学习能力很强, 但是推广能力不是很好.

2 核函数的构造与证明

Mercer 定理^[8]描述了函数 $K(x, z)$ 为核函数的性质. 定理如下:

令 X 是 R^n 的紧子集, 设 K 是连续对称函数, 存在积分算子 $T_k: L_2(X) \rightarrow L_2(X)$, 使得: $(T_k f)(\cdot) = \int_X K(\cdot, x) f(x) dx$ 是正的, 也就是:

$$\int_{X \times X} K(x, z) f(x) f(z) dx dz \geq 0 \quad \forall f \in L_2(X) \quad (4)$$

对于所有的 $T_k: L_2(X) \rightarrow L_2(X)$ 成立, 然后扩展 $K(x, z)$ 到一个一致收敛的序列(在 $X \times X$ 上), 这个序列由 TK 的特征函数 $\phi_j \in L_2(X)$ 构成, 归一化使得 $\|\phi_j\|_{L_2} = 1$, 并且 $\lambda_j \geq 0$, 则有:

$$K(x, z) = \sum_{j=1}^{\infty} \lambda_j \phi_j(x) \phi_j(z) \quad (5)$$

Mercer 定理的条件等价于对 X 中任意有限子集合, 对应的矩阵是半正定的. 即确定一个对称函数是否为核函数的关键条件是该函数在任意有限点集上定义的矩阵是半正定的.

多项式核函数对分布比较分散的数据(即距离测试点比较远的点有一定的影响)可以有较好的影响(即拟合效果)即多项式核函数 kd 能够获取与待测点相距较远的全局信息, 推广能力强. RBF 核函数只对样本分布比较集中的数据点作用明显, 虽然它的学习能力很强, 但是推广能力不是很好. 结合两类核函数特点设定联合核函数的组合形式为:

$$f(x, z) = \eta K_d + (1 - \eta) K_{RBF}, (0 < \eta < 1)$$

两者加权和为 1, 保证 f 为凸函数.

证明: 固定一个有限点集 $\{x_1, \dots, x_n\}$, 考虑任意向量 $\alpha \in R^n$. 则对所有 $\alpha \in R^n$, 有 $\alpha' K_d \alpha \geq 0$; $\alpha' K_{RBF} \alpha \geq 0$. 则有:

$$\alpha' (\eta K_d + (1 - \eta) K_{RBF}) \alpha = \eta \alpha' K_d \alpha + (1 - \eta) \alpha' K_{RBF} \alpha \geq 0 \quad (6)$$

因此 $K_{new}(x, z) = \eta K_d + (1 - \eta) K_{RBF}$ 是半正定的, K_{new} 是核函数^[9,10].

3 加权值的确定

依据上节描述的多项式核函数和径向基核函数的分析, 考虑将两类核函数组合起来构成联合核函数, 利用两类核函数的优点, 减少缺陷的影响程度.

针对径向基核函数的特点引入权值 $e^{-\sigma^2}$, 即:

$$K_{new}(x, z) = e^{-\sigma^2} K_d + (1 - e^{-\sigma^2}) K_{RBF} \quad (7)$$

由上文知, 当 σ^2 趋近于 0 时, 会影响 KRBF 学习能力增强而泛化能力趋近于无. 引入该权值后, 当 σ^2 趋近于 0 时, $e^{-\sigma^2}$ 趋近于 1, 增大了 Kd 的权重值并降低了 KRBF 的权重值, 因此可以提高其泛化推广能力. 解决了 σ^2 取值过小引起的泛化能力差的问题.

针对支持向量机的增量学习, 在多次增量训练过程中可以根据学习效果的好坏来调节加权值. 进行支

持向量训练时,平衡数据训练得到的训练样本能够正确区分两类样本;对正负类数量相差较大的样本,训练得到的分类超平面会向数量较少一类偏移,数量较少类的样本到超平面的平均距离要小于数量较多类样本到超平面的平均距离,引起分类准确率下降.再次学习时可以调节核函数的权重来改善这种情况^[11].

设 d_+ 是正类样本到超平面距离的平均值, d_- 是负类样本到超平面距离的平均值.利用文献[12]中提供的公式可以求出这两个值.当 $|d_+ - d_-|$ 的值较大时,说明分类超平面不在两类样本中间,发生了偏移.在公式(7)中引入算子 $\Delta = \frac{|d_+ - d_-|}{d_+ + d_-}$, 构成联合核函数:

$$K_{new}(x, z) = \Delta \cdot e^{-\sigma^2} K_d + (1 - \Delta \cdot e^{-\sigma^2}) K_{RBF} \quad (8)$$

当 $|d_+ - d_-|$ 较大时, Δ 算子增大,因此增强了算法的泛化能力,减少了分类误差.

经过分析,最终得到核函数的形式为公式(8),其中仍有 $0 < \Delta \cdot e^{-\sigma^2} < 1$ 且两类核函数权重和为 1. 初始训练时,设定 Δ 算子为 1. 在下节中进行实验验证.

4 实验结果与分析

4.1 实验设计

用 Matlab7.0 进行仿真实验,引入 LibSVM-mat-2.83 算法包,从 <http://www.csie.ntu.edu.tw/~cjlin/libsvm-tools/datasets/> 中下载 libsvm 格式的 uci 数据集: liver-disorders 和 mushrooms 数据集.

Libsvm 工具箱中自带核函数及其计算形式主要有:

- (1) 线性核函数, 计算公式: $u \cdot v$
- (2) 多项式核函数, 计算公式: $(\text{gamma} \cdot u \cdot v + \text{coef0})^{\text{degree}}$
- (3) 径向基核函数, 计算公式: $\exp(-\text{gamma} \cdot |u - v|^2)$
- (4) 感知机核函数, 计算公式: $\tanh(\text{gamma} \cdot u \cdot v + \text{coef0})$

本文中构造的联合核函数为公式(8),在 libsvm 中应用自定义核函数需要构造核矩阵^[12].

$$H(i, j) = \Delta \cdot e^{-\sigma^2} * (\text{gamma} * x_i * x_j' + c) \wedge d_+ + (1 - \Delta \cdot e^{-\sigma^2}) * \exp(-\text{gamma} * |x_i - x_j| \wedge 2) \quad (9)$$

$H(i, j)$ 为联合核函数对应的核矩阵.在进行训练时 x_i 和 x_j 为训练样本集的样本点,测试过程中, x_i 为测试样本点, x_j 为训练样本点.分别利用 libsvm 自带核函数 1 和自带核函数 2 以及自定义核函数进行实验.

4.2 实验结果分析

分析特征维数小且数量较少的样本集 liver-disorders. 将前 100 个样本作为初始训练集,依次将后面 50 个作为增量训练集 1, 100 个作为增量训练集 2, 50 个作为增量训练集 3, 最后 45 个作为测试集. 训练集中的正负类数量如表 1 所示.

表 1 liver-disorders 样本构成

类别	初始训练集	增量训练集1	增量训练集2	增量训练集3	测试训练集
正类	46	19	45	13	22
负类	54	31	55	37	23

在训练过程中,使用启发式算法 GA 进行参数寻优得到最优惩罚函数值 $c=66.5730$, 参数值 $g=0.9756$. 进行初始训练和增量训练后得到分类准确率如表 2 所示.

表 2 不同核函数得到的不同分类准确率

	核函数多项式	径向基核函数	联合核函数
初始训练集	57.8	55.6	60.1
增量训练集1	64.4	73.3	77.8
增量训练集2	66.7	64.4	78
增量训练集3	70.1	65	79

根据表 2 的数据得到图 1, 能够更直观的看出,对 liver-disorders 样本,在初始训练中和所有增量训练中,应用联合核函数训练的准确率都要高于应用其他两类核函数的准确率.

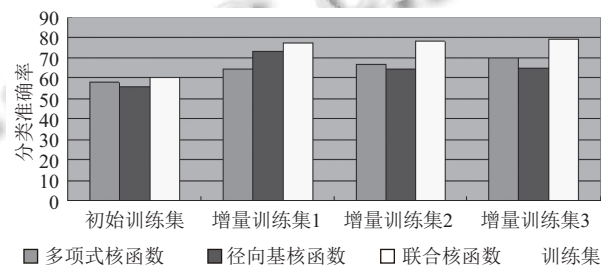


图 1 应用不同核函数的分类准确率

分析特征维度较多且样本数量很大的 mushrooms 样本集. 将前 3000 个样本作为初始训练集,中间 2000 个样本作为增量样本集,最后 2000 个样本作为测试集. 其中正负类样本数量如表 3 所示.

表 3 mushrooms 样本构成

类别	初始训练集	增量训练集1	增量训练集2	增量训练集3	测试训练集
正类	237	54	1266	763	1539
负类	1763	946	734	237	461

在训练过程中,使用启发式算法 GA 进行参数寻优得到最优惩罚函数值 $C=170$, 最优参数 $g=0.0001$. 进行初始训练和增量训练后得到分类准确率如表 4 所示.

表 4 不同核函数得到的分类准确率

	多项式核函数	径向基核函数	联合核函数
初始训练集	75.4	80.7	80.5
增量训练集1	75.5	83	85.5
增量训练集2	80	89.1	90
增量训练集3	90.8	92.8	98.4

由表 4 的数据得到图 2, 能够直观看出, 对 mushrooms 数据, 在初始训练和所有增量训练中, 应用联合核函数训练的准确率要高于应用其他两种核函数的分类准确率. 而且应用三种核函数进行增量学习后, 分类准确率都有明显提高.

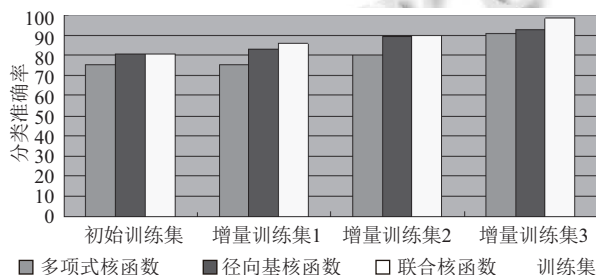


图 2 不同核函数的分类准确率

根据上述得到的分类准确率进行分析, 针对特征维度较少的样本, 应用联合核函数得到的分类准确率大幅度提高, 增量训练后比原初始训练的准确率也有了提高. 针对特征维度较高的样本, 应用联合核函数取得分类效果也高于应用前两种核函数的分类效果. 由图 1 和图 2 可以看出, 在增量训练过程中, 多项式核函数和联合核函数的分类准确率是依次增加的, 而径向基核函数的分类准确率不稳定, 可以看出多项式核函数有很好的泛化能力, 联合核函数继承了它的优点, 准确率能够随着增量学习稳定增长.

5 结束语

通过对支持向量机中核函数的分析, 得出不同核函数对应支持向量机算法的分类效果不同. 全局核函数的值会受到相距很远的数据点的影响, 具有较好的外推能力. 局部核函数的值只受到相距很近数据点的影响, 有很好的学习能力. 分析两类函数的缺陷和优点, 对分类过程中出现的问题综合利用两类核函数的优点

来解决. 使联合核函数能够结合两者的优点, 提高分类准确率. 将构造的联合核函数应用于信息的分类, 并选取两种类型的数据集, 同多项式核函数的实验结果和径向基核函数的实验结果进行了对比, 取得了良好的效果.

参考文献

- Vapnik V. The nature of statistical learning theory. 2nd ed. New York, NY: Springer, 2000.
- Cristianini N, Shawe-Taylor J. An introduction to support vector machines: And other kernel-based learning methods. New York, NY: Cambridge University Press, 2000.
- Vapnik V. The nature of statistical learning theory. New York, NY: Springer, 1995.
- Zhang Y, Liu WM. A novel pedestrian detection method based on cost-sensitive support vector machine and chaotic particle swarm optimization with T mutation. *Przeład Elektrotechniczny*, 2012, 88(1B): 22–25.
- Wang LY, Liu Z. RVM with wavelet kernel combined with PSO for short-term load forecasting in electric power systems. *Przeład Elektrotechniczny*, 2012, 88(1B): 146–149.
- Sun JC, Li XH, Yang Y, *et al.* Scaling the kernel function based on the separating boundary in input space: A data-dependent way for improving the performance of kernel methods. *Information Sciences*, 2012, 184(1): 140–154. [doi: 10.1016/j.ins.2011.08.028]
- Zanaty EA, Afifi A. Support Vector Machines (SVMs) with universal kernels. *Applied Artificial Intelligence*, 2011, 25(7): 575–589. [doi: 10.1080/08839514.2011.595280]
- Dioşan L, Rogozan A, Pecuchet JP. Improving classification performance of Support Vector Machine by genetically optimising kernel shape and hyper-parameters. *Applied Intelligence*, 2012, 36(2): 280–294. [doi: 10.1007/s10489-010-0260-1]
- Christmann A, Hable R. Consistency of support vector machines using additive kernels for additive models. *Computational Statistics & Data Analysis*, 2012, 56(4): 854–873.
- Dudek G. Optimization of the kernel regression model to short-term load forecasting. *Przeład Elektrotechniczny*, 2011, 87(9A): 222–225.
- Li ZY, Zhang JF, Hu SS. Incremental support vector machine algorithm based on multi-kernel learning. *Journal of Systems Engineering and Electronics*, 2011, 22(4): 702–706. [doi: 10.3969/j.issn.1004-4132.2011.04.021]
- Li CH, Liu KW, Wang HX. The incremental learning algorithm with support vector machine based on hyperplane-distance. *Applied Intelligence*, 2011, 34(1): 19–27. [doi: 10.1007/s10489-009-0176-9]