

支持向量机理论与应用研究综述

周宇 181619

Abstract: Statistical learning theory is a theory that studies the statistical laws and learning methods in the case of small samples. Support vector machine is a machine learning method based on statistical learning theory. Compared to neural networks, support vector machines have a solid theoretical foundation. Besides, support vector machine has good robustness and generalization ability, can effectively overcome the curse of dimensionality, widely used in handwriting recognition, text classification, face recognition and so on. This paper first expounds the theory of support vector machine and the mainstream training algorithm, then introduces the application of support vector machine. Finally, it analyzes and summarizes the existing problems of support vector machine, and then forecasts the future research direction of support vector machine.

Key words: statistical learning, support vector machines, training algorithm

摘要: 统计学习理论是一种研究小样本情况下统计规律和学习方法的理论。支持向量机是建立在统计学习理论上的一种机器学习方法。相比于神经网络,支持向量机具有坚实的理论基础。此外,支持向量机具有良好的鲁棒性和泛化能力,可以有效的克服维度灾难的问题,广泛应用于手写识别,文本分类,人脸识别等。本文首先阐述了支持向量机的理论和主流的训练算法,之后介绍了支持向量机的应用,最后对支持向量机的现存在的问题进行了分析总结,进而展望了支持向量机的未来研究方向。

关键词: 统计学习,支持向量机,训练算法

1. 支持向量机理论

1.1 支持向量机

支持向量机是一种二类分类模型,主要的学习目的是寻找一个间隔最大的分离超平面将训练样本数据正确得分为两类。与神经网络中的感知机不同,感知机得到的分离超平面只是能正确得将样本数据分为两类,并没有考虑间隔问题。根据训练样本的是否可以线分,可以将支持向量机分为两类。一类是线性支持向量机,线性支持向量机应用于线性可分样本集或者近似线性可分样本集。一类是非线性支持向量机。当训练数据线性不可分时,可以将原样本特征空间,利用非线性变换,映射到高维特征空间中。在高维特征空间中,利用线性支持向量机解决问题。经典的支持向量机只能解决二分类问题,可以将传统的支持向量机进行拓展,形成能够处理多分类问题的支持向量机。

1.2 线性支持向量机

给定一个训练样本 $T = \{ (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) \}$, (其中 $x_i \in R^n, y_i = \{+1, -1\}$

$i = 1, 2, \dots, n$ 。 x_i 为训练样本中第 i 个特征向量。 y_i 是 x_i 的类标记 +1 代表正类, -1 代表负类, 假设训练数据集是线性可分的, 线性支持向量机学习目标是寻找一个分类间隔最大的分离超平面, 将特征空间划分为两部分, 一部分为正类, 一部分为负类。分类间隔指的是距分离超平面最近的样本点到平面的距离。

分离超平面为 $w \cdot x_i + b = 0$, 计算出分类间隔为 $\frac{1}{\|w\|}$, 所以线性可分支持向量机的原始最优化问题为

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad (1)$$

$$s.t. \quad y_i(w \cdot x_i + b) - 1 \geq 0, i = 1, 2, \dots, n \quad (2)$$

其中，最大化 $\frac{1}{\|w\|}$ 和最小化 $\frac{1}{2} \|w\|^2$ 是等价的。约束条件表示任一训练样本点到分离超平面的距离要大于

于等于分类间隔

原始最优化问题是一个凸二次规划问题。为求解原问题。利用 Wolfe 对偶定理，能够得到等价对偶问题，如下，通过求解对偶问题可以得到原始问题的最优解

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^n \alpha_i \quad (3)$$

$$s.t. \quad \sum_{i=1}^n \alpha_i y_i = 0 \quad (4)$$

$$\alpha_i \geq 0, i = 1, 2, \dots, n$$

如果训练样本存在少量噪声数据，即训练样本不能完全线分，可以为每个样本引入松弛变量 $\xi_i \geq 0$ ，使训练样本点到分离超平面的距离加上松弛变量大于等于分类间隔，所以线性不可分的线性支持向量机学习问题变为

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \quad (5)$$

$$s.t. \quad y_i(w \cdot x_i + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0, i = 1, 2, \dots, n \quad (6)$$

其中 $C > 0$ ，为惩罚参数， C 值大时，对误分类的惩罚增大， C 值小时，对误分类的惩罚减小。最优化目标函数包含两层含义，一使分类间隔尽量大，同时使误分类点的个数尽量少。相应的对偶问题为

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^n \alpha_i \quad (7)$$

$$s.t. \quad \sum_{i=1}^n \alpha_i y_i = 0 \quad (8)$$

$$0 \leq \alpha_i \leq C, i = 1, 2, \dots, n$$

1.3 非线性支持向量机

对于线性不可分的情况，可以用一个非线性变换 ϕ ，将原特征空间 R^n 映射到新的特征空间 H 。将在原来特征空间线性不可分的问题，转换为新特征空间中的线性可分的问题。

$$x \rightarrow \phi(x) = (\phi_1(x), \phi_2(x), \phi_3(x), \dots, \phi_l(x))^T \quad (9)$$

定义核函数 $K(x, z)$ ， $K(x, z) = \phi(x) \cdot \phi(z)$ ，其中 x, z 都是样本点的特征向量。相应的非线性支

持向量的对偶优化问题为

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \sum_{i=1}^n \alpha \quad (10)$$

$$\begin{aligned} s.t. \quad & \sum_{i=1}^n \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, n \end{aligned} \quad (11)$$

在求解最优化的问题的时候，只需要考虑核函数 $K(x, z)$ ，而不需要显式地定义特征空间和非线性变换函数，这样的技巧称为核技巧。在实际的应用中往往需要应用领域知识直接选择核函数，常用的核函数主要有 3 类

1. 多项式和函数 $K(x, z) = (x \cdot z + 1)^p$ ，所得到的 q 阶多项式分类器
2. 高斯核函数 $K(x, z) = \exp(-\frac{\|x - z\|^2}{2\sigma^2})$ ，所得到的分类器与传统的径向基神经网络分类器最大区别在于每个基函数中心对应一个支持向量，它们与输出权值都是由算法自动确定
3. *sigmoid* 核函数 $K(x, z) = \tanh(\alpha x^T z + c)$ ，这时支持向量机实现是包含一个隐含层的多层感知机，隐含层的节点个数由算法自动确定，而且不存在困扰神经网络的局部极小点问题。

1.4 支持向量机应用于多分类问题

传统的支持向量机是一个二分类模型，仅仅解决两类分类问题。可以对传统的支持向量机进行拓展，形成能够处理多分类问题的支持向量机。下面是目前存在的三种主流方法。

1. **直接法**，直接在目标函数上进行更改，将多个分类面的参数求解合并到一个最优化问题，通过求解该最优化问题一次性实现多类分类。表面上，方法很简单，但实际上计算复杂度高，实现起来比较困难。
2. **间接法**，通过组合多个二类分类器来构造多类分类器。常见的实现方法是有两种。一是一对一法，在任意两类样本之间构造一个二类分类器，所以 k 个类别就需要 $\frac{k(k-1)}{2}$ 个二类分类器。当对一个未标记的样本进行分类时，利用 $\frac{k(k-1)}{2}$ 个分类器进行分类，其中得票最多的类别为预测的样本标记。
3. **层次支持向量机**，层次分类法首先将所有类别分成两个子类，再将子类进一步划分成两个次级子类，如此循环下去，直到得到一个单独的类别为止，这样就得到一个倒立的二叉分类树。该方法将原有的多类问题同样分解成了一系列的两类分类问题，其中两个子类间的分类模型采用支持向量机模型。

2. 支持向量机训练算法

2.1 块算法

块算法是由 Boser^[1]等人提出来的，出发点是删除矩阵中对应拉格朗日乘子为零的行和式不会对最终的结果产生影响。其目标就是通过某种迭代方式逐步排除非支持向量，块算法将矩阵的规模从样本数的平方减少到具有非零拉格朗日乘子的样本数的平方。块算法减少了矩阵的规模，所以在很大程度上降低了训练过程中对存储容量的要求。但是块算法的目标是找出所有的支持向量，所以当支持向量的数量很大的时候，块算法会变得非常复杂。

2.2 分解算法

分解算法^[2]是由 Osuna 等人提出来的,是解决大规模问题的有效方法。分解算法是将原二次规划问题分解成一系列规模较小的二次规划问题,进行迭代求解,在每次迭代的过程中,选择拉格朗日乘子分量的一个子集作为工作集,在工作集上,利用传统的优化算法,求解子问题。但是可以发现分解算法的工作集是随机选取的,选取最优的工作集可以提高算法的收敛性和收敛速度。针对这一点,Osuna 等人对原方法进行了改进,提出了一些选择工作集的策略与技术,并实现了这一算法,最终形成了软件 *SVM^{light}*。在选择工作集方面,其思想是从全部变量中挑选出 q 个变量,这些变量应满足下述条件:不等于零且与目标函数可行的最速下降方向对应。

序列最小最优化^[3] (sequential minimal optimization) 算法,是在 1998 年,由 Platt 提出。序列最小最优化算法是分解算法的一个特例。其工作集个数为 2,SMO 将原二次规划问题分解为只包含两个变量的子问题。其中两个子问题可以解析求解,工作集的选择是基于启发式的。首先从外层循环中选取违背 KKT 条件最严重的点,并将其对于的变量之后作为第一个变量,再在内层循环中,选择第二个变量。SMO 算法的工作集为 2,虽然可能会增加迭代次数,但是子问题规模很小,所需的计算量小,所以 SMO 算法通常表现出整体快速收敛的性质。另外 SMO 算法不需要存储核矩阵,没有矩阵运算,容易实现。自从 SMO 算法提出后,很多专家学者提出了对其改进的方法。Liu-ling Dai^[4]等人提出了 3SAO (Ternary Sequential Analytic Optimization) 算法。3SAO 算法突出的特点是工作集为三个拉格朗日乘子变量,三个变量也是可以解析求解的。通过实验证明,3SAO 算法相比于传统的 SMO 算法,训练速度更快。

2.3 增量算法

前面介绍的块算法,分解算法都不支持在线训练。而增量算法支持在线训练。增量学习是机器学习系统在处理新增样本时,能够只对原学习结果中与新样本有关的部分进行增加修改或删除操作,与之无关的部分则不被触及。增量训练算法的一个突出特点是支持向量机的学习不是一次离线进行的,而是一个数据逐一加入反复优化的过程。有关学者,提出了一种 SVM 增量学习淘汰算法^[5],通过对 SVM 增量学习算法的研究,发现若某样本为支持向量,则其在连续多次的学习中甚至在训练过程的全程,都违背 KKT 条件,反之如果某样本在连续多次学习中都满足 KKT,在以后的学习中该样本将以很高的概率满足 KKT 条件。基于上述分析将后者遗忘,则会对后续的学习影响较小。也有学者通过改进用于选择初始训练集的 K 近邻方法,和采用中心密度的方法选取有效的非支持向量集^[6],进一步优化了算法的训练时间和精度。

3. 支持向量机的应用

3.1 手写识别

手写识别是支持向量机的一个重要应用领域。从大量手写文字图片中提取出来的样本集作为训练集,用基于核函数的支持向量机进行训练,得到分类模型,用来预测分类输入的手写文字。GDTW (高斯动态时间规整) 核函数是联机手写识别分类器设计中应用最成功的核函数,有关学者针对 GDTW 核函数计算复杂度高,提出了一种优化 GDTW 核函数的方法^[7],应用于联机手写识别,取得了较好的效果。

3.2 文本分类

另一个支持向量机的重要应用领域就是文本分类。通常来说,文本分类的主要任务就是将未被标记的文档自动分类到预定义类别。文本分类是一个十分重要的自然语言处理,大部分都是应用在邮件分类,过滤信息,检索信息中。常见的文本分类方法有支持向量机, K 最近邻算法和朴素贝叶斯等。与其他的文本分类方法相比,支持向量机有坚实的统计学理论基础和优异的泛化能力。有学者针对支持向量机在文本分类中不平衡文本数据集的局限,提出 PSO-SMOTE 混合算法优化支持向量机的文本分类能力^[8]。针对不平衡文本数据集问题,利用 SMOTE 算法生成插值样本均衡数据集,并通过 PSO 算法迭代进化得到最佳的插值样本,对支持向量机的文本分类能力进行优化。同时为了解决支持向量机存在训练和测试时候过长的

问题, 有关学者提出了一种改进的支持向量机文本分类方法^[9], 利用 Mercer 核函数把实际输入样本向量空间中呈非线性可分分布的样本向量映射到一个高维的特征向量空间, 以实现线性可分, 之后采用二叉树来创建高维特征空间中的 SVM 分类器, 实现文本分类。

3.3 人脸识别

支持向量机在人脸识别中的应用, 也是学者的一个研究热点。人脸识别的特征提取通用的算法是主成分分析算法 (PCA)。PCA 算法的原理是利用 K-L 变换抽取人脸的主要成分, 构造特征脸空间, 识别时将测试图像投影到此空间, 得到一组投影系数, 然后再和各个人脸进行比较识别。进行人脸识别的时候可以采用基于核函数的支持向量机模型, 可以达到较高的准确率。针对传统主成分分析法提取人脸特征时效率低下的问题, 有关学者提出了一种基于粒子群优化的支持向量机人脸识别方法^[10], 利用交叉验证方法, 将支持向量机模型训练时的识别准确率作为粒子群的适应度值, 利用粒子群算法对 SVM 惩罚参数的核函数进行全局寻优, 得到参数的最优解, 用于训练最终的分类器模型进行人脸识别。此外, 也有学者提出了蚁群算法与支持向量机模型进行结合的人脸识别算法^[11], 采用支持向量机进行人脸多重特征检测提取, 然后对提取到的特征信息采用蚁群算法进行训练分类, 实现对面脸特征的准确检测和分类识别。

4. 总结

支持向量机是建立在统计学习理论基础上, 以结构风险最小化的原则建立起来的一种机器学习方法, 具有坚实的理论基础。在解决小样本, 高维问题和非线性问题, 表现出了良好泛化能力, 广泛用于手写识别, 文本分类和人脸识别。但是支持向量机也存在一些缺陷。

1. 在支持向量机模型中, 将非线性问题转换为线性问题的手段是将原特征空间映射到高维空间中, 虽然利用核技巧可以不用显式地定义高维空间和非线性变换函数, 只要给出核函数即可。核函数的好坏决定支持向量机模型的泛化能力。但是现在还没有给出一种合适的方法去确定核函数。
2. 支持向量机模型是借助二次规划来求解最优化决策函数。当面临大规模的样本数据时, 支持向量机的训练算法收敛速度较慢并且占用较多的计算机内存。未来可以考虑如何优化训练算法, 提高支持向量机的训练速度和泛化能力。
3. 传统的支持向量机是一个二分类模型。前文也介绍了目前几种将传统支持向量机拓展处理多分类问题的方法。未来可以考虑结合其他算法的优势, 来提高多类问题的分类准确率。比如与粗集理论结合, 形成一种优势互补的多类问题的组合分类器。
4. 支持向量机具有坚实的理论基础和良好的泛化能力, 并有效得克服了维度灾难问题, 广泛应用于文本分类, 人脸识别, 手写识别。未来可以研究支持向量机和其他优势算法结合, 进一步拓宽其应用领域。

References

- [1] BOSER B E, GUYON I M, VAPNIK V N. A training algorithm for optimal margin classifiers[C]//Proceedings of The Fifth Annual Workshop on Computational Learning Theory. New York: ACM Press, 1992: 144-152.
- [2] OSUNA E, FRENUD R, GIROSI F. An improved training algorithm for support vector machines[C]//Proceedings of IEEE Workshop on Neural Networks for Signal Processing. New York, USA: [s.n.], 1997: 276-285.
- [3] PLATT J. Fast training of support vector machines using sequential minimal optimization[C]//Advances in Kernel Methods: Support Vector Learning. Cambridge, MA: The MIT Press, 1998.
- [4] DAI Liu-ling, HUANG He-yan, CHEN Zhao-xiong. Ternary sequential analytic optimization algorithm for SVM classifier design[J]. Asian Journal of Information Technology, 2005, 4(3): 2-8.

[5]申晓勇,雷英杰,史朝辉,王坚,呼玮.一种 SVM 增量学习淘汰算法[J].计算机工程与应用,2007(06):171-17

[6]刘向. 支持向量机增量学习算法的研究与应用[D].长沙理工大学,2016.

[7] 吴新榕. 基于 SVM 的联机手写分类器设计[D].浙江大学,2012.

[8]高超,许翰林.基于支持向量机的不平衡文本分类方法[J].现代电子技术,2018,41(15):183-186.

[9]李琮,陈利.一种改进的支持向量机文本分类方法[J].计算机技术与发展,2015,25(05):78-82.

[10]廖周宇,王钰婷,谢晓兰,刘建明.基于粒子群优化的支持向量机人脸识别[J].计算机工程,2017,43(12):248-254.

[11]孙珊珊.蚁群算法优化支持向量机的人脸识别[J].现代电子技术,2016,39(21):92-94+98.