

扩展阅读

线性核 SVM 迄今仍是文本分类的首选技术。一个重要原因可能是：若将每个单词作为文本数据的一个属性，则该属性空间维数很高，冗余度很大，其描述能力足以将不同文档“打散”。关于打散，参见 12.4 节。

m 是样本个数。

支持向量机于 1995 年正式发表 [Cortes and Vapnik, 1995]，由于在文本分类任务中显示出卓越性能 [Joachims, 1998]，很快成为机器学习的主流技术，并直接掀起了“统计学习” (statistical learning) 在 2000 年前后的高潮。但实际上，支持向量的概念早在二十世纪六十年代就已出现，统计学习理论在七十年代就已成型。对核函数的研究更早，Mercer 定理 [Cristianini and Shawe-Taylor, 2000] 可追溯到 1909 年，RKHS 则在四十年代就已被研究，但在统计学习兴起后，核技巧才真正成为机器学习的通用基本技术。关于支持向量机和核方法有很多专门书籍和介绍性文章 [Cristianini and Shawe-Taylor, 2000; Burges, 1998; 邓乃扬与田英杰, 2009; Schölkopf et al., 1999; Schölkopf and Smola, 2002]，统计学习理论则可参阅 [Vapnik, 1995, 1998, 1999]。

支持向量机的求解通常是借助于凸优化技术 [Boyd and Vandenberghe, 2004]。如何提高效率，使 SVM 能适用于大规模数据一直是研究重点。对线性核 SVM 已有很多成果，例如基于割平面法 (cutting plane algorithm) 的 SVM^{perf} 具有线性复杂度 [Joachims, 2006]，基于随机梯度下降的 Pegasos 速度甚至更快 [Shalev-Shwartz et al., 2011]，而坐标下降法则在稀疏数据上有很高的效率 [Hsieh et al., 2008]。非线性核 SVM 的时间复杂度在理论上不可能低于 $O(m^2)$ ，因此研究重点是设计快速近似算法，如基于采样的 CVM [Tsang et al., 2006]、基于低秩逼近的 Nyström 方法 [Williams and Seeger, 2001]、基于随机傅里叶特征的方法 [Rahimi and Recht, 2007] 等。最近有研究显示，当核矩阵特征值有很大差别时，Nyström 方法往往优于随机傅里叶特征方法 [Yang et al., 2012]。

支持向量机是针对二分类任务设计的，对多分类任务要进行专门的推广 [Hsu and Lin, 2002]，对带结构输出的任务也已有相应的算法 [Tsochantaridis

et al., 2005]. 支持向量回归的研究始于 [Drucker et al., 1997], [Smola and Schölkopf, 2004] 给出了一个较为全面的介绍.

核函数直接决定了支持向量机与核方法的最终性能, 但遗憾的是, 核函数的选择是一个未决问题. 多核学习(multiple kernel learning) 使用多个核函数并通过学习获得其最优凸组合作为最终的核函数 [Lanckriet et al., 2004; Bach et al., 2004], 这实际上是在借助集成学习机制.

替代损失函数在机器学习中被广泛使用. 但是, 通过求解替代损失函数得到的是否仍是原问题的解? 这在理论上称为替代损失的“一致性”(consistency)问题. [Vapnik and Chervonenkis, 1991] 给出了基于替代损失进行经验风险最小化的一致性充要条件, [Zhang, 2004] 证明了几种常见凸替代损失函数的一致性.

SVM 已有很多软件包, 比较著名的有 LIBSVM [Chang and Lin, 2011] 和 LIBLINEAR [Fan et al., 2008] 等.

集成学习参见第 8 章.

一致性亦称“相合性”.