第 32 卷 第 17 期 计 算 机 工 程 2006 年 9 月
Vol.32 № 17 Computer Engineering September 2006
• 专题论文 • 文章编号：1000—3428(2006)17—0077—03 文献标识码：A 中图分类号：TP18

基于支持向量机的多分类增量学习算法

朱美琳，杨 佩
（南京大学工程管理学院，南京 210093）

摘 要：支持向量机被成功地应用在分类和回归问题中，但是由于其需要求解二次规划，使得支持向量机在求解大规模数据上具有一定的缺陷，尤其是对于多分类问题，现有的支持向量机算法具有太高的算法复杂性。该文提出一种基于支持向量机的增量学习算法，适合多分类问题，并将之用于解决实际问题。

关键词：支持向量机；增量学习；多分类问题

Multi-class Incremental Learning Based on Support Vector Machines

ZHU Meilin, YANG Pei
(School of Management and Engineering, Nanjing University, Nanjing 210093)

【Abstract】 Support vector machines are successfully applied to solve a large number of classification and regression problems. But it may sometimes be preferable to learn incrementally from previous SVM results, as SVMs which involve the solution of a quadratic programming problem suffer from the problem of large memory requirement and CPU time when they are trained in batch mode on large data sets, especially on multi-class problem. An approach for incremental learning based on support vector machines is presented, and is used to solve multi-class real-world problem.

【Key words】 Support vector machines(SVMs); Incremental learning; Multi-class problem

1 概述

在计算机网络迅猛发展的现状下，如何处理爆炸的信息是当前必须解决的问题之一。尤其是企业在线获取的数据具有很强的实时性，会不断将新的数据增加到数据库中，这就要求在对这些新增加的数据进行分析的同时，仍能保留对已有数据的历史分析。

传统的学习方法是无论以前是否进行过学习，都将更新的数据和以前的旧数据放在一起，重新进行训练，得到新的分类准则，这种做法被称之为批量分类方法。相当于忘记以前学习的所有结果，这无疑在空间和时间上都是非常浪费的。

增量学习方法是对不断更新的数据进行学习的更有意义的方法之一，它是在保留以前学习结果的基础上，仅对新增加的数据进行再学习，从而形成一个连续的学习过程。这种方法更适合于像支持向量机（SVMs）这样的分类算法，因为传统的 SVMs 训练时需要求解二次规划，用它来训练大数据集需要比较高的内存，并且收敛速度较慢，增量学习方法可以比较好地解决这些问题。

支持向量机因为其特有的准确性和全局解而成为求解分类问题和回归问题的一个非常有效的工具[1]，被成功地应用到很多方面，例如图像物体的识别[2]和时间序列预测[3]等。

目前已经提出了一些基于SVM的增量学习方法。Syed等人[4]强调了支持向量的作用，将旧数据集中的支持向量保留，加入到新数据集中进行训练，得到新的支持向量和分类函数。Ralaivola[5]等人采用了局部增量学习的方法，当新数据不能被正确分类的时候，就将新数据周围的数据放到新数据集中，从而修正其原有的分类准则。Erdem[6]等人使用投票的方式来确定测试数据的分类，首先对每一批数据都单独进行学习，分别得到分类器，然后将测试数据代入到每一个分类器中，得到其分类标志，标志数最多的那个标志就是其分类。尽管近几年来国际上关于SVMs的研究正如火如荼，并已取得一些成熟和有意义的结果，但有关该领域的增量学习方法的研究还处于初级阶段，仍有许多问题有待研究和突破。SVMs最基本的理论是针对二分类问题的，然而在实际应用中，存在着大量的多分类问题，比如语音识别、字体识别、人脸识别等，因此必须对SVMs进行改进和推广。

本文针对多分类增量学习问题，提出一种改进的 SVMs 方法，并通过实例来验证该算法的有效性，该算法在算法复杂性和数据规模上都占有优势。

2 多分类问题

对于多分类问题，现有的 SVMs 方法主要有下列几种：

一对多方法（one-against-rest），针对不同的 k 个分类，构造 k 个 SVM 分类器，第 m 个分类器是将第 m 类与其余的分类分开，即将第 m 类重新标号为 1，其它类重新标号为-1。完成这个过程需要计算 k 个二次规划，这个方法的不足之处在于容易产生属于多类别的点和没有被分类的点。

一对一方法（one-against-one）[7]，对于任意两个分类，构造一个SVM分类器，仅识别这两个分类，完成这个过程需要$k(k\text{-}1)/2$ 个分类器，计算量是非常庞大的。

层（树）分类方法，这种方法是对一对一方法的改进，将 k 个分类合并为两个大类，每个大类里面再分成两个子类，如此下去，直到最基本的 k 个分类，这样形成不同的层次，每个层次都使用 SVMs 来进行分类。

k-类 SVM 方法，对所有的样本使用同一个二次规划，只

作者简介：朱美琳(1972—)，女，博士、副教授，主研方向：优化理论，人工智能，系统仿真；杨 佩，博士、讲师

收稿日期：2006-06-20 **E-mail：**zhuml@nju.edu.cn

需要一次就可以决定分类，这种方法的局限在于，由于要一次处理所有数据，约束条件急剧增加，进行分类的二次规划相当庞大，即使是转化为线性规划，数据的规模依然受限。

一种球结构的多分类算法是在上述算法的基础之上提出将同一类数据用超球来界定，数据空间就变为由若干个超球组成，在三维上面，就像是很多肥皂泡的集合。

本文针对的是 $k > 2$ 类的识别问题，给出 k 类样本点，构造一个算法来区分各个分类，目的是为了能有效正确地对未知样本点分类。

对于k类问题，数学表述如下：给定k个n维空间的元素集合 A^m， $m=1,\ldots,k$，每一个集合 A^m 包含 l^m 个点 x_i^m，$i=1,\ldots,l^m$，这些点都属于同一分类。对于每一个集合 A^m，$m=1,\ldots,k$，寻找一个球(a^m,R^m)，其中 a^m 是球的中心，R^m 为球的半径平方，并尽可能地达到最小，使得球(a^m,R^m)包含所有（或几乎所有）样本点 x_i^m，$i=1,\ldots,l^m$。由于此种定义对一些非常偏远的点很敏感，因此允许有一些点可以在球的外面。同支持向量机[1]类似，此处引入松弛变量 ξ_i^m，得到如下约束条件：

$$\left\|x_i^m - a^m\right\|^2 \le R^m + \xi_i^m \quad i=1,\ldots,l^m \tag{1}$$

$$\xi_i^m \ge 0 \qquad i=1,\ldots,l^m \tag{2}$$

使松弛变量和半径的平方 R^m 最小，因此有如下目标函数：

$$F(R^m,a^m,\xi_i^m) = R^m + C^m\sum_i \xi_i^m \tag{3}$$

其中 C^m 为某个指定的常数，实际上起到控制对错分样本惩罚程度的作用，实现了球的大小和错分样本之间的折衷。

给定一个测试样本点 x，需判断它属于哪一个分类，首先计算点 x 到各球心的距离的平方，$(x-a^m)\cdot(x-a^m)=(x\cdot x)-2(x\cdot a^m)+(a^m\cdot a^m)$，$m=1,\ldots,k$，与 R^m 进行比较，即计算 $(a^i\cdot a^i)-2(x\cdot a^i)-R^i$，找出最小值所在的分类，就是 x 所属的分类。

通常情况下，即使排除了偏远的样本点，数据依然不会呈现球状分布，为了使本文提出的方法适用于更广泛的领域，并且样本可以被变换到更高维的特征空间，同 SVMs 方法类似，在以上公式中用到内积的部分可以用核函数来代替。

求最小球的二次规划变换为

$$\max L(\alpha_i^m) = \sum_i \alpha_i^m k(x_i^m,x_i^m) - \sum_{i,j}\alpha_i^m\alpha_j^m k(x_i^m,x_j^m) \tag{4}$$

$$s.t.\begin{cases}\sum_i \alpha_i^m = 1\\ 0 \le \alpha_i^m \le C^m\end{cases} \tag{5}$$

判别公式为

$$k(a^i,a^i) - 2k(x,a^i) - R^i \tag{6}$$

不同的核函数，会导致不同的特征空间，因此也会有不同形状的样本分布，这与严格的球状分布相比更加方便及精确。使用到的核函数一般有：

多项式核函数：$K(x,y)=((x\cdot y)+1)^d$

径向基核函数：$K(x,y)=\exp\left(-\|x-y\|^2/(2\sigma^2)\right)$

Sigmoid 核函数：$K(x,y)=\tanh(\kappa(x\cdot y)+\Theta)$

如果使用多项式核函数，参数 d 是多项式的度，随着 d 的增大，样本之间的距离也增大，这会使得原本偏远的样本更偏远，整个样本的分布变得更广、更稀疏。为了限制变换到更大的特征空间，使用径向基核函数更适合球分类方法。

3 多分类增量学习方法

在上述球分类的基础上，再使用增量学习方法，就能更好地对实时数据进行在线分析。本文的方法主要基于Drucker等人提出的增量学习思想[5]，将旧数据集中的支持向量保留，加入到新数据集中进行训练，其基本思想如图 1 所示。

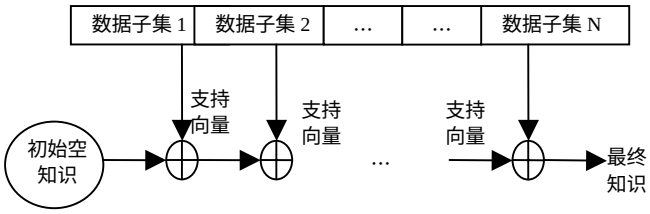

图 1 增量训练过程

这种增量学习方法能够和球分类方法较好地结合起来，因为在球分类中得到的支持向量实际上是每一个类别中边界上的点（比传统 SVMs 超平面上的支持向量数量多，也能更好地反映数据的分布情况），由这些支持向量决定了每个类别球的半径，当该类新加入的数据被球包围的时候，不会改变其类别的支持向量。

令第 i 批的数据子集为 A_i，在 A_i 中第 m 类的数据表示为 A_i^m，具体过程描述如下：

(1)利用球分类方法训练 A_1，对每一类 A_1^m，得到初始的 SV_1^m，$SV=\sum_{m=1}^{k} SV_1^m$

(2)令第 i 批新的数据集为 $A_i^m = A_i^m + SV^m$，$A_i = \cup A_i^m$

(3)再利用球分类方法训练 A_i，对每一类 A_i^m，得到 SV_i^m，$SV=\sum_{m=1}^{k} SV_i^m$，再进行下一批的训练。分类函数公式不发生变化，只是每批数据的中心和半径会随着支持向量的改变而改变。

与各种 k-分类的支持向量机算法进行比较，本文提出的算法有很多优势，主要体现在以下几个方面：

(1)可以适应大规模数据：由于球分类算法中，每一个二次规划只针对本类的样本，因此处理数据的容量大大增强，优于所有的多分类 SVM 算法，尤其优于一对多、k-类 SVM 算法。

(2)算法复杂性小：在各种 k-分类的支持向量机算法和球分类方法中，占计算量最大比例的是对二次规划的计算，一对一算法和层分类算法针对每批数据都需要计算大于 k 次的二次规划，从而增加了计算时间，而球分类增量算法每批只需要计算 k 次，并且约束条件简单，易于推广和改进，并可转化为其它形式的优化问题。

(3)易于扩充：在各种 k-分类的支持向量机算法中，当增加一个新的分类（例如，在人脸库中增加一个新人的人脸图像），原来建立的分类系统就被打破，需要将这个新类与以前的各类别进行比较，重新计算多个二次规划以寻找新的支持向量。而球分类方法并不会干扰以前的分类，以前的计算仍将有效，只需对新的分类本身进行一个二次规划以及相关的简单计算即可，这就使得分类易于操作、没有重复工作、更具备扩展能力、也更适合于增量数据。

4 实验

为了验证本文提出的算法在实际应用中的有效性，将该方法用于某电信行业的客户分析，根据其特征把相似的客户归总到一起，而不同客户组之间的差异最大化。数据主要包括以下内容：电话号码、通话总次数、总时长、主叫次数、主叫时长、被叫次数、被叫时长、(1860，1861）拨打服务次

数、短信次数、忙时次数、忙时时长、闲时次数、闲时时长、3 月平均次数、3 月平均时长等。用户被分为 VIP 用户、一般用户、可能流失客户以及流失客户 4 类。由于电信行业每天都有新数据产生，将某营业厅每半个月新增加的用户作为一个新的数据子集，研究该营业厅的用户情况。程序使用 MatLab 语言编写，并使用留一法进行验证，结果如表 1。

表 1 数据集

数据集	数据量	计算时间（s）	留一法验证结果（错误率）
初始数据集	1000	134	3.7%
数据子集 1	231	7	3.5%
数据子集 2	247	8	2.9%
数据子集 3	216	6	2.8%
数据子集 4	198	6	2.8%

从验证结果可以看出，使用本文提出的基于支持向量机的多分类增量学习算法是非常有效的，不仅能将错误率保持在较低的水平上，而且计算时间非常短。

5 结论

本文主要介绍了针对 SVMs 增量数据的研究，以及在多类别情况下进行识别的一些算法，并在此基础上提出了基于 SVM 的多分类增量学习算法。这种算法可用于比较庞大的多类别识别问题，并且易于扩展。本文详细介绍了该方法的原理，并将之应用到电信的客户分析上，取得了比较好的结果。

参考文献

1 Cristianini N, Shawe-Taylor J. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods[M]. Cambridge University Press, 2000.

2 Blanz V, Schölkopf B, Bültho H, et al. Comparison of View-based Object Recognition Algorithms Using Realistic 3D Models[C]. Proc. of ICANN'96. Berlin: Springer-Verlag, 1996: 251-256.

3 Mattera D, Haykin S. Support Vector Machines for Dynamic Reconstruction of a Chaotic System[M]. Cambridge: MIT Press, 1999: 211-242.

4 Syed N A, Liu H, Sung K. Incremental Learning with Support Vector Machines[C]. Proceedings of the Workshop on Support Vector Machines at the International Joint Conference on Artificial Intelligence (IJCAI-99), Stockholm, Sweden, 1999.

5 Ralaivola L, d' Alch-Buc F. Incremental Support Vector Machine Learning: A Local Approach[C]. Proc. of ICANN'01. Vienna, Austria: Springer, 2001: 322-330.

6 Erdem Z, Polikar R, Gurgen F, et al. Ensemble of SVMs for Incremental Learning[C]. Workshop on Multiple Classifier Systems, 2005: 246-256.

7 Kreßel U. Pairwise Classification And Support Vector Machines[M]. Cambridge: MIT Press, 1999: 255-268.

（上接第 54 页）

实验 2: 与实验 1 雷同，只是每组数据库的记录数是按 20 递增。实验结果表明，随着记录的增加时间开销也在增大，但 EOFS 增大趋势小于 BRSFS(如图 2 所示)。

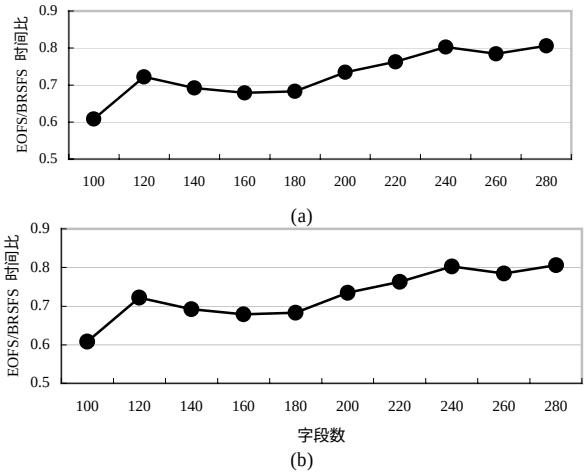

图 2 字段数变、特征值范围和记录数不变

BRSFS 是在求得所有特征子集后，再找出与 EOFS 相同的特征选取。尽管 EOFS 与 BRSFS 多了后两部分的时间开销，但通过启发式信息进行数据空间的快速减小，最终还是提高了特征选取效率。

4 结束语

通过前面分析和实验，高效最小特征选取的关键在于：(1)利用具有当前最大分类对象能力的启发式信息，确保选取的特征子集是最小特征选取；(2)最大限度减小数据空间，使得在后续的特征选取在最小数据空间中进行遍历。尽管在EOFS中存在$Cost2_{Red(Sk)}$时间开销，但所创建的$Red(S_K)$空间开销小于S_K。由于EOFS在特征选取过程中不断减小数据空间直至数据空间为ϕ，因此，不能进行其他特征子集的求解。

参考文献

1 戴东亚, 郑启伦, 胡劲松等. 一种基于粗糙集的混合特征选取方法[C]. 计算机科学, 2001, 28 (5): 95-97.

2 陈　彬, 洪家荣, 王亚东. 最优特征子集选取[J]. 计算机学报. 1997, 20 (2): 133-138.

3 朱　明, 王俊普, 蔡庆生. 一种最优特征集的选取算法[J]. 计算机研究与发展, 1998, 35 (9): 803-805.

4 Kohavi R, Frasca B. Useful Feature Subsets and Rough Set Reducts [C]. The 3th International Workshop on RoughSets and Soft Computing, 1994.

5 李　萌, 魏长华. 一种基于差异矩阵的属性简约算法[C]. 计算机科学, 2002, 29 (9): 403-406.

6 曾黄麟. 粗集理论及其应用[M]. 重庆：重庆大学出版社, 1998: 55-70.

7 Gasca E, SÃÃnchez J S, Alonso R. Eliminating Redundancy and Irrelevance Using a New MLP-based Feature Selection Method[J]. Pattern Recognition, 2006, 39(2): 313-315.

8 Kononenko I, Hong S J. Attribute Selection for Modelling [J]. Future Generation Computer Systems, 1997, 13(2/3): 181-195.

9 Skowron A, Rauszer C. The Discernibility Matrices and Function in Information Systems[C]. Proc. of Intelligent Decision Support-handbook of Application and Advances of Rough Set Theory. Kuwler Academic Publisher, 1992: 331-362.

10 洪家荣. 示例学习的扩张理论[J]. 计算机学报, 1991, 6(6): 401-410.