局部支持向量机的研究进展

尹传环[1] 牟少敏[2] 田盛丰[1] 黄厚宽[1] 朱莹莹[1]
(北京交通大学计算机与信息技术学院 北京 100044)[1]
(山东农业大学信息科学与工程学院 泰安 271018)[2]

摘 要 支持向量机是一种用途广泛的分类器,标准的支持向量机在预测每个样本点的类别时使用了训练集中所有的样本信息(即全局信息),然而这种全局化的方法并不蕴含一致性。局部支持向量机的提出符合"一致性蕴含局部性"的思路。首先回顾局部支持向量机的主要思想,然后阐述各种关于局部支持向量机的改进,并提出基于协同聚类的局部支持向量机用于大规模数据集,最后对局部支持向量机进行总结。

关键词 支持向量机,局部支持向量机,协同聚类
中图法分类号 TP181 **文献标识码** A

Survey of Recent Trends in Local Support Vector Machine

YIN Chuan-huan[1] MU Shao-min[2] TIAN Sheng-feng[1] HUANG Hou-kuan[1] ZHU Ying-ying[1]
(School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)[1]
(School of Computer and Information Engineering, Shandong Agriculture University, Taian 271018, China)[2]

Abstract Support Vector Machine(SVM) is an important and widely used classifier. If a sample wants to be classified, all training data will be used to obtain a hyperplane which is used to determine the label of that sample, that is, the SVM worked in global manner. However, this global behaviour doesn't imply consistency. The design of Local SVM(LSVM) is in accordance with the result of "consistency implies local behaviour". In this paper, we first reviewed the main idea of LSVM, followed by the improvements on LSVM. In the following, we presented an LSVM algorithm which is based on cooperative clustering, reducing the time complexity of LSVM in large scaled dataset. Then we ended this article by the conclusion.

Keywords Support vector machine, Local support vector machine, Cooperative clustering

支持向量机(Support Vector Machine,SVM)是20世纪90年代中期由Vapnik教授根据统计学习理论提出的[1],可用于模式分类和回归估计[2-5]。支持向量机与传统的模式分类、机器学习的方法如人工神经网络等相比有其优越性,如泛化能力强、全局优化等,因此获得了越来越广泛的应用。

1 支持向量机

支持向量机本质上来说是一种二类分类器,能求解各种二类分类问题。给定一个独立同分布的样本集 X:

$$\{(x_i,y_i),i=1,2,\cdots,n\} \tag{1}$$

式中,$x_i\in R^d$,$y_i\in\{-1,1\}$。关于它的分类问题(二类分类)是这样描述的:样本 x_i 若属于第一类,则标记为正($y_i=+1$);若属于第二类,则标记为负($y_i=-1$),将样本集 X 称为训练集。而学习的目标是构造一个决策函数,使得测试集 H:$\{\bar{x}_j,j=1,2,\cdots,l\}$中的每一个样本点 $\bar{x}_j$ 尽可能被正确分类。

根据结构化风险最小化原则,在训练集中构造目标函数[1]:

$$\min \frac{1}{2}||w||^2+C\sum_{i=1}^{n}\xi_i$$
$$s.t.\ y_i(w\cdot\Phi(x_i)+b)\geqslant 1-\xi_i$$
$$\xi_i\geqslant 0,i=1,2,\cdots,n \tag{2}$$

式中,C 是惩罚项常数,ξ_i 是松弛变量(松弛因子)。ξ_i 用来惩罚那些不能被准确分开的样本点,而 C 则用来权衡惩罚的力度,C 越大,错误分类的惩罚就越重。求解上述问题之后,可以利用决策函数

$$f(\bar{x})=\mathrm{sgn}(w\cdot\Phi(\bar{x})+b) \tag{3}$$

来判定测试样本点所属类别,其中 sgn(·)为符号函数。

近年来,关于学习算法的一致性问题受到了一定程度的关注。首先描述二类分类的一致性问题[6,7]。

给定独立同分布 P 的训练样本集 X,X 的样本个数为 n,令 $C(X_n,\bar{x}),\bar{x}\in H$ 为使用分类算法 C 预测服从分布 P 的测

到稿日期:2011-09-21 返修日期:2011-11-24 本文受国家自然科学基金(61105056),中央高校基本科研业务费专项资金,山东农业大学青年科技创新基金项目(200923647),北京交通大学科技基金(2007RC066)资助。

尹传环(1976-),男,博士,讲师,主要研究方向为机器学习、入侵检测等,E-mail:chyin@bjtu.edu.cn;牟少敏(1964-),男,博士,副教授,主要研究方向为机器学习、图像处理、模式识别和信息安全等;田盛丰(1944-),男,教授,博士生导师,主要研究方向为人工智能、模式识别等;黄厚宽(1940-),男,教授,博士生导师,主要研究方向为数据库知识发现、计算机免疫、多Agent系统和人工智能等;朱莹莹(1987-),女,硕士生,主要研究方向为机器学习。

试集 H 中的某个样本 $\bar{x}$ 的类标号(即 1 或 −1)。令 $f*(\bar{x})=E(y|\bar{x})=2\eta(\bar{x})-1$ 为给定样本 $\bar{x}$ 类标号 y 的期望值,其中 $\eta(\bar{x})=P(y=1|\bar{x})$,即给定样本 $\bar{x}$ 时类标号为 1 的概率。定义

$$C*(\bar{x})=\mathrm{sgn}(f*(\bar{x}))=\mathrm{sgn}(2\eta(\bar{x})-1) \tag{4}$$

如果满足:

$$L_n(C)\equiv E_{X_n,x}|C(X_n,\bar{x})-C*(\bar{x})|\cdot(2\eta(\bar{x})-1)\xrightarrow[n\to\infty]{}0 \tag{5}$$

则称分类算法 C 在分布 P 上一致;如果算法 C 在所有分布上都一致,则称算法 C 是一致(或全局一致)的[6]。

Steinwart[8]证明了支持向量机在给定条件下是全局一致的,但是在一般情况下并不是全局一致的。而 Zakai 与 Ritov[6]证明了全局一致性必然蕴含局部性。因此,如何使支持向量机利用局部信息来满足算法的一致性是局部支持向量机在理论上的驱动。而 Brailovsky 等人[9]、Zhang 等人[10]以及 Blanzieri 与 Melgani[11]提出局部支持向量机主要是受局部学习算法[12,13]的启发。

除了理论研究之外,局部支持向量机的应用也已经得到了一定程度的关注,如局部支持向量机在遥感图像分类[14]、脑电图信号处理[15]、网络流量预测[16]等领域的应用。

下面详细介绍局部支持向量机及其改进策略。

2 局部支持向量机

局部支持向量机的思路最早是由 Brailovsky 等人提出的[9],他们将核函数添加两个乘子,使之具有局部性。定义:

$$h(|x_i-w_r|)=\begin{cases}1, & \text{如果}|x_i-w_r|\leqslant\theta_r\\ 0, & \text{否则}\end{cases} \tag{6}$$

于是,定义一个新的核函数:

$$K*(x_i,x_j)=\sum_{r=1}^{t}K_{w_r}(x_i,x_j) \tag{7}$$

其中

$$K_{w_r}(x_i,x_j)=K(x_i,x_j)h(|x_i-w_r|)h(|x_j-w_r|) \tag{8}$$

$w_1,w_2,\cdots,w_t$ 等 t 个元素是使得 $|x_i-w_r|$ 能够覆盖整个训练集的一系列样本。可以通过设定 θ_r 的取值定义一个局部核函数:

$$K*(x_i,x_j)=\begin{cases}K(x_i,x_j), & \text{如果 } x_i \text{ 在 } x_j \text{ 的 } k\text{-近邻中}\\ 0, & \text{否则}\end{cases} \tag{9}$$

尽管这种思路与以下介绍的 kNNSVM 相似,但是 Brailovsky 等人在计算 k-近邻时,是在样本空间而不是特征空间中进行的[9],这将导致引入的这种局部性不一定适合不同分布的样本,正如他们在实验中表明的那样,这种方法在不同的数据集上性能差异较大。

2.1 SVM-KNN

Zhang 等人将他们提出的局部支持向量机称为 SVM-KNN[10],具体步骤为:

(1)首先找到测试样本 $\bar{x}$ 在训练样本集 X 中的 k 个近邻。

(2)计算这 k-近邻和样本 $\bar{x}$ 两两之间的距离,并得到距离矩阵 A。

(3)将距离矩阵 A 转换为核矩阵 K。

(4)最后在核矩阵 K 上利用 DAGSVM[17]确定 $\bar{x}$ 的类标号。

尽管在 SVM-KNN 中,第(4)步使用的 DAGSVM 是用来解决多类分类问题的,但实际上对于二类分类问题,SVM-KNN 同样适用,只需将第(4)步的 DAGSVM 改为 SVM 即可。

2.2 kNNSVM

Blanzieri 与 Melgani 将他们提出的局部支持向量机称为 kNNSVM。将式(2)中的约束改为:

$$y_{r_{x'}(i)}(w\cdot\Phi(x_{r_{x'}(i)})+b)\geqslant 1-\xi_{r_{x'}(i)},i=1,2,\cdots,k \tag{10}$$

式中,$r_{x'}$ 是一个将训练样本的索引重新排列的函数,其定义域与值域均为 $\{1,2,\cdots,n\}$:

$$\begin{cases}r_{x'}(1)=\underset{i=1,\cdots,n}{\mathrm{argmin}}\|\Phi(x_i)-\Phi(x')\|^2\\ r_{x'}(j)=\underset{i=1,\cdots,n}{\mathrm{argmin}}\|\Phi(x_i)-\Phi(x')\|^2 \text{ with}\\ \quad i\neq r_{x'}(1),\cdots,r_{x'}(j-1) \text{ for } j=2,\cdots,n\end{cases} \tag{11}$$

即 $x_{r_{x'}(1)}$ 是在特征空间中与 x' 距离最近的那个样本,而 $x_{r_{x'}(j)}$ 则是在特征空间中与 x' 距离排名第 j 的那个样本。根据 $r_{x'}$ 的定义,存在性质:

$$j<k\Rightarrow\|\Phi(x_{r_{x'}(j)})-\Phi(x')\|\leqslant\|\Phi(x_{r_{x'}(k)})-\Phi(x')\| \tag{12}$$

因此,可以通过计算 $\|\Phi(x)-\Phi(x')\|$ 的大小顺序来确定 $r_{x'}$。同时,根据核函数的定义:

$$K(x_i,x_j)=\Phi(x_i)\cdot\Phi(x_j) \tag{13}$$

可以得到:

$$\begin{aligned}\|\Phi(x)-\Phi(x')\|^2&=\langle\Phi(x),\Phi(x)\rangle+\langle\Phi(x'),\Phi(x')\rangle\\&\quad-2\cdot\langle\Phi(x),\Phi(x')\rangle\\&=K(x,x)+K(x',x')-2\cdot K(x,x')\end{aligned} \tag{14}$$

因此可以通过计算核函数的方式求解 $r_{x'}$。当核函数为 RBF 核函数或者线性核函数时,$r_{x'}$ 实际上可以通过对样本点之间的欧氏距离进行排序得到。但一般情况下,$r_{x'}$ 与欧氏距离的排序是不同的。

测试样本 $\bar{x}$ 可以利用下述决策函数得到分类标号:

$$k\mathrm{NNSVM}(\bar{x})=\mathrm{sgn}\left(\sum_{i=1}^{k}\alpha_{r_{\bar{x}}(i)}y_{r_{\bar{x}}(i)}K(\bar{x}_{r_{\bar{x}}(i)},\bar{x})+b\right) \tag{15}$$

当 $k=n$ 时,kNNSVM 实际上就是标准的 SVM,当 $k=2$ 时,RBF 核或线性核对应于标准的 1NN 分类器。

当测试样本 $\bar{x}$ 的 k-近邻全部是同类的训练样本时,可以直接将这些样本的标号赋予样本 $\bar{x}$。

Blanzieri 和 Melgani 证明了 kNNSVM 比 SVM 具有更好的泛化能力[14]。

SVM-KNN 使用向量空间中的距离寻找 k-近邻,这儿所指的距离可以是多样的,如 χ^2 距离、边界距离或切线距离等。然后利用某些策略将形成的距离矩阵转换为核矩阵。而 kNNSVM 则是直接利用核特征空间中的距离得到 k 近邻,可以看出,kNNSVM 明显比 SVM-KNN 更为直观,而且计算量将更少。

2.3 LSVM

Cheng 等人[18,19]提出了一种新的局部支持向量机,称为 LSVM(Localized Support Vector Machine)。

令 $\sigma(\bar{x},x_i)$ 为测试样本 $\bar{x}$ 与训练样本 x_i 的相似度,为测试集 H 中的每一个样本 $\bar{x}$ 构造一个局部支持向量机:

$$\min\ \frac{1}{2}||w||^2+C\sum_{i=1}^{n}\sigma(\bar{x},x_i)\xi_i$$

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

$$\xi_i \geqslant 0, i=1,2,\cdots,n \tag{16}$$

根据 σ 的定义，有两种 LSVM 的变种：(1)当 σ 为[0,1]之间的实数时，得到的 LSVM 称为 SLSVM(Soft Localized Support Vector Machine)；(2)当 σ 为二值函数，即 $\sigma \in \{0,1\}$ 时，得到的 LSVM 称为 HLSVM(Hard Localized Support Vector Machine)。Cheng 等人认为[19]，如果

$$\sigma(\bar{x},x_i)=\begin{cases}1, & x_i \in \bar{x} \text{ 的 } k\text{-近邻}\\ 0, & \text{其他}\end{cases} \tag{17}$$

得到的 HLSVM 等价于 SVM-KNN，也就是说 SVM-KNN 实际上是 LSVM 的一种特殊形式。但是我们认为，LSVM 仅仅采用线性核，而 SVM-KNN 虽然允许直接计算样本之间的欧氏距离，但还有其他距离可供选择，也就是说，SVM-KNN 可以选择不同的核函数。实际上，采用式(17)为相似度函数的 HLSVM 与采用欧氏距离的 SVM-KNN 是等价的，但是一般情况下，HLSVM 与 SVM-KNN 并不存在包含关系。

从上述 3 种局部支持向量机的描述可以看出，无论是 kNNSVM、SVM-KNN 还是 LSVM，都需要为每个测试样本构造一个模型来判别此测试样本的标号。在测试样本数量比较多的情况下局部支持向量机的时间复杂度相当高，尽管每个模型所使用的样本减少很多，但需要构造的模型个数却增长更多。因此，如何在保留局部支持向量机一致性的同时提高测试效率是近年来局部支持向量机研究的一个主要目标。

3 局部支持向量机的改进

在局部支持向量机中，一个很直观的改进思路就是减少需要训练的支持向量机模型的个数。更具体的改进思路如算法 1 所示：

(1)将训练集中的样本依据某种原则划分为 k 类，并找到 k 个中心；

(2)为每个中心构造一个支持向量机；

(3)为每个测试样本 $\bar{x}$ 找到与之最相近的一个中心；

(4)利用该中心对应的支持向量机对 $\bar{x}$ 进行标号。

实际上现有的大部分改进都是基于上述思路进行的。

3.1 PSVM

PSVM 采用一种聚类的方法将样本集划分成 k 类，Cheng 等人将这种聚类方法命名为 MagKmeans[19]。

首先引入一个矩阵。如果 $\bar{x} \in H$，令

$$\Delta(\bar{x})=[\sigma(\bar{x},x_1),\sigma(\bar{x},x_2),\cdots,\sigma(\bar{x},x_n)]^T \tag{18}$$

为一个相似度的列向量，其中的每个元素 $\Delta_i(\bar{x})$ 是测试样本 $\bar{x}$ 与训练样本 x_i 的相似度。令

$$\Sigma=[\Delta(\bar{x}_1)\quad \Delta(\bar{x}_2)\quad \cdots\quad \Delta(\bar{x}_m)] \tag{19}$$

为 $n\times m$ 的矩阵，n 是训练样本的个数，而 m 是测试样本的个数。$\Sigma(i,j)$ 则代表训练样本 x_i 和测试样本 $\bar{x}_j$ 的相似度。再令 $Y=(y_1,y_2,\cdots,y_n)^T$ 为训练样本的类标号向量。

MagKmeans 的目标函数是：

$$\min_{C,Z}\sum_{j=1}^{k}\sum_{i=1}^{n}Z_{i,j}\|X_i-C_j\|_2^2+R\sum_{j=1}^{k}\left|\sum_{i=1}^{n}Z_{i,j}y_i\right| \tag{20}$$

式中，X_i 是相似度矩阵 Σ 的第 i 行，C_j 是 $1\times m$ 的行向量，代表第 j 个聚类的中心，R 是一个非负的缩放参数，$Z_{i,j}\in\{0,1\}$ 是聚类隶属度矩阵的元素，如果第 i 个训练样本属于第 j 簇，则 $Z_{i,j}$ 等于 1，否则等于 0。

式(20)中的第一项表示簇的内聚程度，最小化此项能够确保同一簇中的训练样本高度相似。而第二项则表示每一簇中正负类样本的不均衡程度，最小化第二项能够确保每一簇中的正负类样本均衡，如图 1 所示。

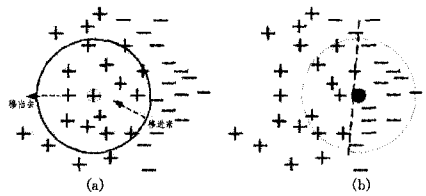

图 1 图(a)所示的样本分布经聚类后得到图(b)所示的簇

MagKmeans 算法的思想来源于 k 均值聚类，但是两种算法的聚类结果不同，图 2 展示了这两种聚类算法的区别。

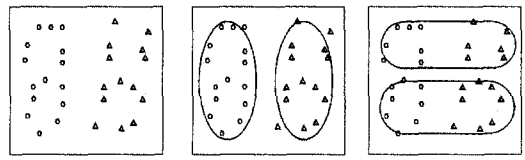

图 2 图(a)中有两类样本点，使用 $k=2$ 时的 k 均值聚类则聚成图(b)，使用 $k=2$ 时的 MagKmeans 则聚成图(c)

Cheng 等人的实验结果表明[19]，HLSVM 的分类精度不如非线性 SVM，而 SLSVM 的分类精度明显高于非线性 SVM；MagKmeans 可以有效减少 HLSVM 和 SLSVM 的运行时间，并且能保持与 SLSVM 相差无几的分类精度，仍然高于非线性 SVM。

然而，Cheng 等人的实验结果还表明 MagKmeans 这种聚类方法耗费了大量的时间，相比较而言，构造多个基于聚类中心的 SVM 所消耗的时间几乎可以忽略不计。

PSVM 尽管取得了比较好的分类精度，但是依然存在几个问题：

(1)LSVM 以及 PSVM 使用的都是线性核，这与局部支持向量机试图获得一致性的理论驱动相悖。

(2)SLSVM 实际上并不是一种采用局部信息的支持向量机，根据相似度函数的不同，它很有可能在预测每个样本的标号时都要用上所有的样本。

(3)在通常情况下，HLSVM 取得比较差的分类精度是可以预料的，因为 HLSVM 仅仅是在样本空间中找到 k 个近邻，并且在支持向量机中采用的是线性核，那么可想而知，只有线性可分的情况下 HLSVM 才有可能取得较好的分类精度。

(4)PSVM 与 HLSVM 相比较而言，实际上最大的改进就是利用 MagKmeans 获得 k 个聚类中心。因此，PSVM 同样存在第(3)点所述的问题。

3.2 Falk-SVM

Segata 和 Blanzieri 同样认识到了 PSVM 存在的问题[20,21]，他们在 kNNSVM 的基础上提出了 Falk-SVM(Fast Local Kernel Support Vector Machine)。

Falk-SVM 的基本思路与算法 1 相似，最大的区别在于选取中心点的过程。

如前所述，PSVM[19] 采用一种聚类的方法来获得中心，这是一种耗时并且提前使用了测试集样本的方法。而 Falk-SVM 则在训练集上采用类似于求最小覆盖集[22] 的方法来获得覆盖所有训练集的中心。

定义一个集合 A 的最小距离 $d(A)$ 和最大距离 $D(A)$：

$$d(A)=\min\|x-x'\|,x,x'\in A,\text{且 } x\neq x'$$
$$D(A)=\max\|x-x'\|,x,x'\in A \tag{21}$$

因此对于训练集 X，最小距离和最大距离分别为 $d(X)$ 和 $D(X)$。

首先构造一系列 X 的子集 S_i，使得 $S_i\subseteq X$，并且满足：

$$\begin{cases}\cdots\subset S_{i+1}\subset S_i\subset S_{i-1}\subset\cdots\\ \cdots>d(S_{i+1})>d(S_i)>d(S_{i-1})>\cdots\end{cases} \tag{22}$$

在构造 S_i 时可以更具体地引入一个大于 1 的数值 b，使得 $d(S_i)>b^i$。将 S_i 中的最大索引 i 称为 top，最小的则称为 bot。S_i 可以通过递归得到：

$$S_{top}=\{\text{choose}(X)\}$$
$$S_i=S_{i+1}\bigcup\underset{S\in\bar{X}\setminus S_{i+1}}{\text{argmax}}(|S|\ s.t.\quad d(S_{i+1}\bigcup S)>b^i) \tag{23}$$
$$i=top-1,\cdots,bot$$

式中，choose(A)是从非空集合 A 中选择一个元素的函数。可以简单地将 choose(A)定义为取得 A 中下标最小的元素。该递归的真正含义是：S_i 除了包含 S_{i+1} 中所有的元素之外，还将尽可能多的不在 S_{i+1} 中且距离 S_{i+1} 够远的那些样本包含进来。可以看出，通过式(23)构造出来的一系列 S_i 满足式(22)。

接下来从一系列 S_i 中选取中心，使得这些中心的 k'-近邻能够覆盖所有的训练样本。选取中心也是通过一个递归公式：

$$\begin{cases}c_1=\text{choose}(S_{top})\\ c_j=\text{choose}(S_l)\ s.t.\ l=\max(m\in N|S_m\setminus X_{c_{j-1}}\neq\phi)\end{cases} \tag{24}$$

式中，$X_{c_{j-1}}=\bigcup_{i=1}^{j-1}\{x_{r_{c_i}}(h)|h=1,\cdots,k'\}$ 是已经被包含在 C 中的中心的近邻的并集。

具体的选取方法描述如下：第一个中心 c_1 就是 S_{top} 中的那个样本。下一个中心 c_2 从非空的 S_l 集中选取，其中 S_l 是 S_i 去除了第一个中心 c_1 以及它的 k'-近邻后得到的集合。因为 S_i 有多个，所以这样的 S_l 往往会有多个，通常从使得 l 最大且 S_l 非空的集合中选取。其余 c_j，$j=2,3,\cdots$ 的选取过程类似，唯一的区别在于是从 S_l 中去除所有中心 c_t 及其近邻，其中 $t<j$。直到最后所有的训练样本都在某个(或者多个)中心的 k'-近邻中，递归算法结束。

注意，选取中心时，上述算法确保的是所有中心的 k'-近邻能够覆盖所有训练样本，而不是它们的 k-近邻覆盖所有样本，$k'\leqslant k$。而如前所述，基于某中心构造支持向量机时，是将该中心的 k-近邻选作训练集，因此有一部分训练样本将被重复使用。实际上 Segata 和 Blanzieri 在他们的工作中设置 $k'=\frac{k}{2}$，这是他们在实验中通过参数调优得到的一个经验。尽管这样会导致需要训练更多的支持向量机，但是却能取得几乎最好的分类精度。

还需要注意的是，Segata 和 Blanzieri 并不是为每个测试样本寻找距离最近的中心，而是判断测试样本在所有中心的 k'-近邻中的排序，选择该测试样本排序序号最小(离中心最近)的那个中心，然后再利用该中心对应的支持向量机分类。

他们的实验结果表明[20]：

(1)当使用线性核时，kNNSVM 和 Falk-SVM 的分类精度明显高于 SVM，而使用 RBF 核时差距并不明显，这是因为 RBF 核中包含了一定的局部信息；

(2) Falk-SVM 的分类精度不如 kNNSVM，是因为在 Falk-SVM 中，对于一个测试样本而言，由找到的那个中心的 k-近邻构造的支持向量机一般来说不如直接根据该测试样本的 k-近邻构造的支持向量机合适；

(3)Falk-SVM 能够降低 SVM 以及其他一些近似求解器的时间复杂度，同时取得与它们相仿甚至更好的分类精度。

Yang 等人[23]将局部支持向量机的思想应用到 Proximal support vector machine 中，而 Ye 等人[24]则在该工作的启发下，提出了 Localized Twin SVM。但他们的工作都是将局部性引入支持向量机的变种，而不是对局部支持向量机的改进。

上述改进很好地促进了局部支持向量机的理论研究及其在各个领域中的应用发展。而我们在前期研究的基础之上也提出一种提高局部支持向量机训练速度的方法，称之为基于协同聚类[25]的局部支持向量机(Cooperative Clustering based LSVM，简称为 C2LSVM)。

4 C²LSVM

局部支持向量机只需要在某个样本的 k-近邻范围内构造支持向量机，尽管这样大大减少了每次构造支持向量机的训练样本数量，但是对于大规模数据集而言，这样的 k-近邻数据量依然比较庞大，将会导致单个支持向量机的训练速度较低，从而大大增加训练多个支持向量机的时间复杂度。而如果选择一个较小的 k 值，需要构造的支持向量机个数将会大大增加。

在 Segata 和 Blanzieri 的实验中，他们将 k 最大设置为 8000，这意味着每次构造支持向量机时样本数量最多为 8000，而这样的支持向量机的构造将会重复多次。因此，非常有必要利用某种技术将每次构造支持向量机时的样本数量大量减少，从而降低整个局部支持向量机的时间复杂度。而协同聚类[25]已经验证了它在减少支持向量工作上的有效性，因此我们将协同聚类引入局部支持向量机。

协同聚类的基本思想是：对两类样本点分别采用 k 均值聚类[26]，并寻找两类之间距离最近的样本点组成类中心对，之后使其向类边界方向相互靠拢。这些类中心对可以近似作为支持向量，这种策略能够尽量保留分类超平面附近的样本点，从而保留具有丰富分类信息的有效样本点。协同聚类具体的主要思想描述如下[27]：

假设有一个两类数据集，其相应的正类与负类数据分别是 X^+ 和 X^-，首先在两类中分别运行 k 均值聚类，各自得到 c 个中心点，令 $v^+=\{v_1^+,\cdots,v_c^+\}$ 是正类的类中心，$v^-=\{v_1^-,\cdots,v_c^-\}$ 是负类的类中心，定义 A 是距离矩阵，其元素 a_{ij} 表示 v^+ 中的第 i 个类中心到 v^- 中的第 j 个类中心的距离，可以通过式(25)求得：

$$a_{ij}=||v_i^+-v_j^-||^2,i,j=1,2,\cdots,c \tag{25}$$

在矩阵 A 中，可以求得满足条件 v_p^+ 到 v_q^- 最小距离的聚类中心对$\langle v_p^+,v_q^-\rangle$。令 r_p^+ 为正类第 p 个子类的平均半径，r_q^- 为负类第 q 个子类的平均半径：

$$r_p^+=\frac{1}{n_p^+}\sum_{x_k\in X_p^+}\|x-v_p^+\|$$
$$r_q^-=\frac{1}{n_q^-}\sum_{x_k\in X_q^-}\|x-v_q^-\| \tag{26}$$

式中，n_p^+ 为正类第 p 个子类的样本数，n_q^- 为负类第 q 个子类的样本数，X_p^+ 为正类第 p 个子类的集合，X_q^- 为负类第 q 个

子类的集合。

每一对聚类中心$\langle v_p^+, v_q^- \rangle$可以迭代修改：

$$v_p^+ = v_p^+ + \lambda \frac{r_p^+}{r_p^+ + r_q^-}(v_q^- - v_p^+)$$
$$v_q^- = v_q^- + \lambda \frac{r_q^-}{r_p^+ + r_q^-}(v_p^+ - v_q^-) \quad (27)$$

式中，$\lambda \in (0,1)$是控制v_p^+和v_q^-之间距离的参数。

迭代持续进行，直到聚类中心对不再发生变化，此时将c对聚类中心作为训练样本集，利用式(2)构造分类超平面。

基于 Falk-SVM 的 C²LSVM 算法的训练步骤描述如算法 2 所示：

(1)利用 Falk-SVM 算法，找到各个中心；

(2)为每个中心选取 k 个近邻，加入到训练集；

(3)利用协同聚类将(2)获得的 k 个近邻简化为 $2c$ 个样本；

(4)对于每个中心，利用(3)生成的 $2c$ 个样本构造支持向量机。

基于 Falk-SVM 的 C²LSVM 算法的测试步骤与 Falk-SVM 类似。

同样，基于 kNNSVM 或者 PSVM 的 C²LSVM 的训练步骤与算法 2 类似，只需将第(1)步进行相应的修改即可。

结束语 标准支持向量机及其变种在训练时使用的是全局数据，即每次训练都使用所有的训练样本，实际上这并不符合“一致性蕴含局部性”的思路，因此，研究者在局部学习的启发下，提出了局部支持向量机的思路，正好契合了“一致性蕴含局部性”的理论基础。

构造局部支持向量机有多种方法：定义与样本之间的欧氏距离相关的核函数；将样本之间的各种距离转换为核矩阵；直接选取 k-近邻中的样本构造支持向量机。其中直接选取 k-近邻构造支持向量机(称为 kNNBSVM，kNN Based SVM)的思路比较直接，并且简单有效，如 kNNSVM 和 HLSVM。

然而，在 kNNBSVM 中，需要为每个测试样本构造一个支持向量机，这将导致在测试样本较多的实际应用中时间复杂度过高。因此需要采用某种策略来降低 kNNBSVM 的时间复杂度。PSVM 与 Falk-SVM 都是降低 kNNBSVM 的时间复杂度提出的算法，前者应用了一种特殊的聚类方法获得聚类中心，而后者则是采用一种启发式的方法获得多个中心，它们的目的都是减少需要构造的支持向量机模型的数量。但是 PSVM 存在几个缺陷：在训练时使用了测试样本；为了快速计算核函数，采用的是线性核；为了平衡正负数据，聚类算法时间复杂度较高。而 Falk-SVM 从理论上和实验中验证了 Falk-SVM 的快速和准确。实际上 Falk-SVM 是目前为止各种局部支持向量机中，训练和测试速度以及分类精度最好的一种局部支持向量机。

即使这样，在大规模数据集中，构造每一个支持向量机都涉及到众多的样本数量，因此，本文提出一种基于协同聚类的局部支持向量机算法(C²LSVM)，试图大幅度减少参与训练的样本数量，从而加快局部支持向量机的训练速度。我们将来的工作主要集中于对 Falk-SVM 选取中心的改进。

参 考 文 献

[1] Vapnik V N. Statistical Learning Theory[M]. New York: John Wiley and Sons, 1998

[2] Burges C J C. A tutorial on support vector machines for pattern recognition[J]. Data Mining and Knowledge Discovery, 1998, 2(2): 121-167

[3] Scholkopf B, Smola A. Learning with Kernels[M]. Cambridge, MA: MIT Press, 2001

[4] Smola A, Schölkopf B. A tutorial on support vector regression [J]. Statistics and Computing, 2004, 14(3): 199-222

[5] Vapnik V. The Nature of Statistical Learning Theory[M]. Berlin: Springer Verlag, 2000

[6] Zakai A, Ritov Y. Consistency and localizability [J]. Journal of Machine Learning Research, 2009(10): 827-856

[7] Zhang T. Statistical behavior and consistency of classification methods based on convex risk minimization[J]. Annals of Statistics, 2004, 32(1): 56-134

[8] Steinwart I. Support vector machines are universally consistent [J]. Journal of Complexity, 2002, 18(3): 768-791

[9] Brailovsky V L, Barzilay O, Shahave R. On global, local, mixed and neighborhood kernels for support vector machines[J]. Pattern Recognition Letters, 1999, 20(11-13): 1183-1190

[10] Zhang H, Berg A C, Maire M, et al. SVM-KNN: Discriminative nearest neighbor classification for visual category recognition [C]// Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2006. New York, NY, 2006: 2126-2136

[11] Blanzieri E, Melgani F. An adaptive SVM nearest neighbor classifier for remotely sensed imagery[C]// Proc. of IEEE International Conference on Geoscience and Remote Sensing Symposium 2006. Dever, Colorado, 2006: 3931-3934

[12] Bottou L, Vapnik V N. Local learning algorithms[J]. Neural Computation, 1992, 4(6): 888-900

[13] Vapnik V N, Bottou L. Local algorithms for pattern recognition and dependencies estimation[J]. Neural Computation, 1993, 5(6): 893-909

[14] Blanzieri E, Melgani F. Nearest neighbor classification of remote sensing images with the maximal margin principle[J]. IEEE Transactions on Geoscience and Remote Sensing, 2008, 46(6): 1804-1811

[15] Shen Min-fen, Chen Jia-liang, Lin Chun-hao. Modeling of Nonlinear Medical Signal Based on Local Support Vector Machine [C]// Proc. of International Instrumentation and Measurement Technology Conference 2009. Singapore, 2009: 675-679

[16] Meng Qing-fang, Chen Yue-hui, Peng Yu-hua. Small-time scale network traffic prediction based on a local support vector machine regression model[J]. Chinese Physics B, 2009, 18(6): 2194-2199

[17] Platt J, Cristianini N, Shawe-Taylor J. Large margin DAGSVM's for multiclass classification[C]// Proc. of the 12th Advances in Neural Information Processing Systems. Denver, Colorado, 2000: 547-553

[18] Cheng H, Tan P-N, Jin R. Localized support vector machine and its efficient algorithm[C]// Proc. of SIAM International Conference on Data Mining 2007. Minneapolis, Minnesota, 2007: 461-466

[19] Cheng H, Tan P-N, Jin R. Efficient Algorithm for Localized Support Vector Machine[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 20(4): 537-549

(下转第 189 页)

为了验证算法的有效性，从 UCI 数据库获得了 18 个数据集进行实验，将 L1-BNC 与数据挖掘中 9 种常用分类算法在分类准确度上进行了比较。实验中将选取的 9 种分类算法分成两组进行比较，即贝叶斯分类算法和其它分类算法。实验表明，L1-BNC 在分类精度上优于已有的贝叶斯分类器分类算法。为了更进一步验证 L1-BNC 的有效性，将 L1-BNC 与其它分类算法（如 SVM，KNN，J48）进行了比较。实验表明，在大部分数据集上，L1-BNC 优于这些算法。从实验结果可以看出，L1-BNC 是一种简单有效的分类算法。

参 考 文 献

[1] Cheng J, Greiner R, Kelly J, et al. Learning Bayesian networks from data: An information-theory based approach[J]. Artificial Intelligence, 2002, 137: 43-90

[2] Kojima K, Perrier E, Imoto S, Satoru Miyano Optimal Search on Clustered Structural Constraint for Learning Bayesian Network Structure[J]. Journal of Machine Learning Research, 2010, 11: 285-310

[3] Friedman N, Geiger D, Goldszmidt M. Bayesian Network Classifiers[J]. Machine Learning, 1997, 29(2/3): 131-163

[4] Domingos P, Pazzani M. Beyond Independence: Conditions for the Optimality of the Simple Bayesian Classifier[C]// Proceedings of the Thirteenth International Conference on Machine Learning. San Francisco, CA: Morgan Kaufmann Publishers, Inc., 1996: 105-112

[5] Webb J I. Not so Naïve: Aggregating one Dependence Estimator [J]. Machine Learning, 2005, 58: 5-24

[6] Zhang H, Jiang L, Su J. Hidden Naive Bayes[C]// Proceedings of the Twentieth National Conference on Artificial Intelligence (AAAI-05). AAAI Press, 2005: 919-924

[7] Zhang H, Jiang L, Su J. Augmenting Naive Bayes for Ranking [C]// Proceedings of the 22nd International Conference on Machine Learning(ICML). ACM Press, 2005: 1025-1032

[8] Geiger D, Heckerman D. Knowledge representation and inference in similarity networks and Bayesian multinets[J]. Artificial Intelligence, 1996, 82: 45-74

[9] Cheng J, Greiner R. Comparing Bayesian network classifiers[C]// UAI-99. 1999

[10] Cheng Jie, Greiner R. Learning Bayesian Belief Network Classifiers: Algorithms and System[C]// AI 2001, LNAI 2056. 2001: 141-151

[11] Madden M G. On the classification performance of TAN and general Bayesian networks [J]. Knowledge-Based Systems, 2009, 22: 489-495

[12] Hruschka E R J, Ebecken N F F. Towards efficient variable ordering for Bayesian networks classifier[J]. Data & Knowledge Engineering, 2007: 258-269

[13] Pernkopf F, Bilmes J. Order-based Discriminative Structure Learning for Bayesian Network Classifiers[C]// ISAIM. 2008

[14] Liu Feng, Tian Feng-zhan, Zhu Qi-liang. A Novel Ordering-based Greedy Bayesian Network Learning Algorithm on Limited Data[C]// AI 2007, LNAI 4830. 2007: 80-89

[15] Efron B, Hastie T, Johnstone I, et al. Least Angle Regression [J]. Annals of Statistics(with discussion), 2004, 32(2): 407-499

[16] Tibshirani R. Regression shrinkage and selection via the lasso [J]. J. Royal. Statist. Soc B., 1996, 58(1): 267-288

[17] Vidaurre D, Bielza C, Larrañaga P. Learning an L1-regularized Gaussian Bayesian Network in the Equivalence Class Space[J]. IEEE Transacions of Systems, 2010: 1083-4419

[18] Asuncion A, Newman D J. UCI Machine Learning Repository [EB/OL]. http://www.ics.uci.edu/~mlearn/MLRepository.html, 2007

[19] Cooper G, Herskovitz E. A Bayesian method for the induction of probabilistic networks from data[J]. Machine Learning, 1992, 9: 309-347

[20] idaurre D, Bielza C, Larrañaga P. Learning an L1-regularized Gaussian Bayesian Networkin the Equivalence Class Space[J]. IEEE Transactions on Systems, 2010: 1231-1242

[21] Wu X, Kumar V, Ghosh R J, et al. Top 10 algorithms in data mining[J]. Knowledge and Information Systems, 2008, 14(1): 1-37

[22] BNT-SLP. BNT Structure Learning Package[EB/OL]. http://bnt.insa-rouen.fr/SL.html, 2010

[23] Weka Wiki. http://weak.sourceforge.net/wiki/. 40

[24] Aliferis C F, Tsamardinos I, Stanikov A R, et al. Causal Explorer: A Causal Probabilistic Network Learning Toolkit for Biomedical Discovery[C]// Proceedings of the International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences. 2003: 371-376

（上接第 174 页）

[20] Segata N, Blanzieri E. Fast and Scalable Local Kernel Machines [J]. Journal of Machine Learning Research, 2010, 11 (2010): 1883-1926

[21] Segata N. Local Approaches for Fast, Scalable and Accurate Learning with Kernels[D]. Trento, University of Trento, 2009

[22] Chen L. New analysis of the sphere covering problems and optimal polytope approximation of convex bodies[J]. Journal of Approximation Theory, 2005, 133(1): 134-145

[23] Yang Xu-bing, Chen Song-can, Chen Bin, et al. Proximal support vector machine using local information [J]. Neurocomputing, 2009, 73(1-3): 1227-1234

[24] Ye Qiaolin, Zhao Chunxia, YeNing, et al. Localized twin SVM via convex minimization[J]. Neurocomputing, 2011, 74(4): 580-587

[25] Tian Sheng-feng, Mu Shao-min, Yin Chuan-huan. Cooperative clustering for training SVMs[C]// Proc. of 3rd International Symposium on Neural Networks. LNCS 3971, Chengdu, 2006: 962-967

[26] MacQueen J B. Some Methods for classification and Analysis of Multivariate Observations[C]// Proc. of the 5th Berkeley Symposium on Mathematical Statistics and Probability. Berkeley, University of California Press, 1967: 281-297

[27] 牟少敏. 核方法的研究及其应用[D]. 北京：北京交通大学，2008