基于三支决策的支持向量机增量学习方法

徐久成 刘洋洋 杜丽娜 孙 林
(河南师范大学计算机与信息工程学院 新乡 453007)
(河南省高校计算智能与数据挖掘工程技术研究中心 新乡 453007)

摘 要 针对典型的支持向量机增量学习算法对有用信息的丢失和现有支持向量机增量学习算法单纯追求分类器精准性的客观性,将三支决策损失函数的主观性引入支持向量机增量学习算法中,提出了一种基于三支决策的支持向量机增量学习方法。首先采用特征距离与中心距离的比值来计算三支决策中的条件概率;然后把三支决策中的边界域作为边界向量加入到原支持向量和新增样本中一起训练;最后,通过仿真实验证明,该方法不仅充分利用有用信息提高了分类准确性,而且在一定程度上修正了现有支持向量机增量学习算法的客观性,并解决了三支决策中条件概率的计算问题。

关键词 三支决策,支持向量机,增量学习,条件概率,边界向量
中图法分类号 TP181 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.6.019

Three-way Decisions-based Incremental Learning Method for Support Vector Machine

XU Jiu-cheng LIU Yang-yang DU Li-na SUN Lin
(College of Computer & Information Engineering, Henan Normal University, Xinxiang 453007, China)
(Engineering Technology Research Center for Computing Intelligence & Data Mining of Henan Province, Xinxiang 453007, China)

Abstract Aiming at the problems that the typical incremental learning algorithm for support vector machine (SVM) loses a lot of useful information and the objectivity that existing incremental learning algorithms for SVM aspire to classification accuracy merely, the subjectivity of loss functions of three-way decisions was introduced to incremental learning algorithms for SVM, and a three-way decisions-based incremental learning method for SVM was proposed. Firstly the conditional probability of three-way decisions was denoted by the ratio of feature distances and center distances. Secondly the objects of boundary region of three-way decisions were regarded as boundary vectors to be trained with the original support vectors and the newly added samples. Finally, simulation experiments were done. The results show that the proposed method not only makes full use of the useful information to improve the classification accuracy, but also revises the objectivity of existing incremental learning algorithms for SVM to some extent. Besides, the computation problem of conditional probability of three-way decisions is resolved.

Keywords Three-way decisions, Support vector machine, Incremental learning, Conditional probability, Boundary vectors

1 引言

由 Vapnik 等人提出的支持向量机(SVM)目前已成为解决分类、回归和其他统计学习问题的一种全新的机器学习方法。而增量式算法会使学习机具有在线自适应的能力,能够随着时间而进化[1,2]。文献[3]最早提出了典型的 SVM 增量学习算法,但是该算法仅考虑新样本和原支持向量(SV),而忽略了原先的非支持向量,一些有价值的信息将会丢失,导致得到一个不好的分类器[4]。近年来,针对典型算法的不足,许多学者提出了相应的改进算法。文献[5]提出了一种新的 SVM 对等增量学习算法。但是,随着数据挖掘和机器学习技术在实际问题中的广泛应用,人们越来越多地发现分类问题通常具有代价敏感特性,即误分类代价存在差异性。而现有的 SVM 增量学习算法是一种基于统计学习和机器学习的客观性方法,其单纯地追求分类器的精准性,已不能解决实际问题。

三支决策是决策粗糙集的核心思想之一,它将传统的正域、负域二支决策语义扩展为正域、边界域和负域三支决策语义。与经典粗糙集不同,在三支决策粗糙集模型中,保持决策知识与经验数据的一致性不再是唯一目标,其更倾向于关注

到稿日期:2014-04-25 返修日期:2014-06-02 本文受国家自然科学基金项目(61370169,61402153,60873104),河南省科技攻关重点项目(142102210056),新乡市重点科技攻关计划项目(ZG13004)资助。
徐久成(1964-),男,博士,教授,博士生导师,主要研究方向为粒计算、数据挖掘、生物信息学等;刘洋洋(1992-),女,硕士生,主要研究方向为三支决策、粒计算、机器学习,E-mail:15903863829@163.com;杜丽娜(1989-),女,硕士生,主要研究方向为粒计算、三支决策;孙 林(1979-),男,博士生,讲师,主要研究方向为粒计算、生物信息学等。

决策分类错误带来的风险代价，因此是具有代价敏感性的数据分析工具[6]；此外，三支决策中损失函数的确定需要先验信息和专家知识，这在一定程度上反映了主观能动性，可以修正机器学习的客观性。决策粗糙集尽管在很多领域已取得很多成果，但是也面临着一些挑战，其中包括决策粗糙集模型中条件概率的计算和损失函数的确定问题[7]。

针对上述问题，本文提出了一种基于三支决策的 SVM 增量学习方法。为了有选择地淘汰非 SV，将三支决策的主观性加入 SVM 增量学习中，首先在增量学习时用特征距离与中心距离的比值来计算三支决策中的条件概率，解决三支决策中条件概率的计算问题；然后将三支决策中的边界域作为边界向量加入到原 SV 和新增样本中一起进行训练。最后通过仿真实验验证了本文提出的方法的有效性和合理性。

2 SVM 增量学习和三支决策的相关概念

2.1 SVM 增量学习

2.1.1 支持向量机

对于给定的一组样本集$\{x_i,y_i\}$，$i=1,2,\cdots,l$，这里 $y_i=1$ 或 -1，SVM 依据结构风险最小化原则，将其学习过程转化为如下所示的优化问题：

$$\begin{aligned}&\min \frac{1}{2}\| w \|^2+C\sum_{i=1}^{l}\xi_i\\&\text{s. t. } y_i(w^{\mathrm{T}}z_i+b)\geqslant 1-\xi_i\\&\xi_i\geqslant 0,i=1,\cdots,l\end{aligned}\tag{1}$$

其中，训练样本 x_i 被函数 $z_i=\Phi(x_i)$ 映射到高维特征空间，$w\in R^N$ 是超平面的系数向量，$b\in R$ 为阈值，ξ_i 为松弛变量，$C\geqslant 0$ 是一个常数，用于控制对错分样本惩罚的程度。

采用拉格朗日乘子法把上述优化问题转换为其对偶问题：

$$\begin{aligned}&\min(\alpha)=\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j K(x_i,x_j)-\sum_{i=1}^{l}\alpha_i\\&\text{s. t. }\sum_{i=1}^{l}\alpha_i y_i=0,0\leqslant\alpha_i\leqslant C\end{aligned}\tag{2}$$

于是相应的分类决策函数为

$$f(x)=\operatorname{sgn}(\sum_{i=1}^{m}\alpha_i^* y_i K(x_i,x)+b^*)\tag{3}$$

其中，α_i^* 为对应 $\alpha_i\neq 0$ 的向量，称为支持向量，$m(m<l)$ 为支持向量的数目，b^* 为与 α_i^* 对应的阈值，$K(x_i,x)=\Phi(x_i)^{\mathrm{T}}\Phi(x)$ 为满足 Mercer 条件的核函数[8]。

2.1.2 SVM 增量学习算法

文献[3]最早提出的典型的 SVM 增量学习算法的具体步骤如下。

算法 1 典型的 SVM 增量学习算法

输入：训练样本集 X，随机等分为 N 个互不相交的子集，分别为 X_1，X_2，…，X_N

输出：基于 X 的 SVM 分类器 Γ

Step1 取子集 X_1 进行训练，得到支持向量集 SV_1；

Step2 把 SV_1 和子集 X_2 合并，将其作为新的训练样本进行训练，得到支持向量集 SV_2；

Step3 重复 Step2，将得到的新的支持向量 SV_i 和子集 X_{i+1} 合并训练，如此循环，直到 X_N，支持向量集 SV_{N-1} 与 X_N 训练所得到的支持向量机就作为训练整个样本集 X 得到的分类器 Γ，输出 Γ；

Step4 算法结束。

2.2 三支决策

本节简要介绍三支决策的基本理论，更详细的介绍可以参阅文献[9-14]。

决策粗糙集模型是基于贝叶斯决策过程的。基于三支决策的思想，决策粗糙集模型利用 2 个状态集和 3 个行动集描述决策过程。状态集 $\Omega=\{X,\neg X\}$ 分别表示某事件属于 X 和不属于 X，行动集 $A=\{a_P,a_B,a_N\}$ 分别表示接受某事件、延迟决策和拒绝某事件 3 种行动。考虑到采取不同行动会产生不同的损失，用 λ_{PP}、λ_{BP}、λ_{NP} 分别表示当 x 属于 X 时，采取行动 a_P、a_B、a_N 下的损失；用 λ_{PN}、λ_{BN}、λ_{NN} 分别表示当 x 不属于 X 时，采取行动 a_P、a_B、a_N 下的损失。因此采取 a_P、a_B、a_N 3 种行动下的期望损失可分别表示为：

$$\begin{aligned}R(a_P|[x])&=\lambda_{PP}P(X|[x])+\lambda_{PN}P(\neg X|[x])\\R(a_B|[x])&=\lambda_{BP}P(X|[x])+\lambda_{BN}P(\neg X|[x])\\R(a_N|[x])&=\lambda_{NP}P(X|[x])+\lambda_{NN}P(\neg X|[x])\end{aligned}\tag{4}$$

式中，$[x]$ 为样本在属性集下的等价类，$P(X|[x])$ 和 $P(\neg X|[x])$ 分别表示将等价类 $[x]$ 分类为 X 和 $\neg X$ 的概率。根据贝叶斯决策准则，需要选择期望损失最小的行动集作为最佳行动方案，于是可得到如下 3 条决策规则。

(1)若 $R(a_P|[x])\leqslant R(a_B|[x])$ 和 $R(a_P|[x])\leqslant R(a_N|[x])$ 同时成立，那么 $x\in POS(X)$；

(2)若 $R(a_B|[x])\leqslant R(a_P|[x])$ 和 $R(a_B|[x])\leqslant R(a_N|[x])$ 同时成立，那么 $x\in BND(X)$；

(3)若 $R(a_N|[x])\leqslant R(a_P|[x])$ 和 $R(a_N|[x])\leqslant R(a_B|[x])$ 同时成立，那么 $x\in NEG(X)$。

由于 $P(X|[x])+P(\neg X|[x])=1$，因此上述规则只与概率 $P(X|[x])$ 和相关的损失函数 λ 有关。此处做一个合理的假设：$0\leqslant\lambda_{PP}\leqslant\lambda_{BP}<\lambda_{NP}$，$0\leqslant\lambda_{NN}\leqslant\lambda_{BN}<\lambda_{PN}$。据此，根据以上 3 条决策规则，令

$$\begin{aligned}\alpha&=\frac{\lambda_{PN}-\lambda_{BN}}{(\lambda_{PN}-\lambda_{BN})+(\lambda_{BP}-\lambda_{PP})}\\\beta&=\frac{\lambda_{BN}-\lambda_{NN}}{(\lambda_{BN}-\lambda_{NN})+(\lambda_{NP}-\lambda_{BP})}\\\gamma&=\frac{\lambda_{PN}-\lambda_{NN}}{(\lambda_{PN}-\lambda_{NN})+(\lambda_{NP}-\lambda_{PP})}\end{aligned}\tag{5}$$

通过引入一对阈值(α,β)，上述 3 条规则可改写为

(1)若 $P(X|[x])\geqslant\alpha$，则 $x\in POS(X)$；

(2)若 $\beta<P(X|[x])<\alpha$，则 $x\in BND(X)$；

(3)若 $P(X|[x])\leqslant\beta$，则 $x\in NEG(X)$。

此处的规则描述了基于决策粗糙集的三支决策语义，给予了一种贝叶斯最小风险下的三支决策语义解释[9]。

3 基于三支决策的支持向量机增量学习

3.1 三支决策中条件概率的构建

为了解决三支决策中条件概率的计算问题，在此采用特征距离与中心距离的比值来定义三支决策中的条件概率。考虑到求解距离的方法具有普遍意义，将所提出的几种距离的求法皆以定理的方式给出。

3.1.1 SVM 线性模式下三支决策条件概率的构建

对于线性模式，设$\{x_1,x_2,\cdots,x_{l1}\}$和$\{x_1',x_2',\cdots,x_{l2}'\}$是来自于不同类别的样本，$m_x=\frac{1}{l_1}\sum_{i=1}^{l_1}x_i$ 和 $m_{x'}=\frac{1}{l_2}\sum_{i=1}^{l_2}x_i'$ 分别为两类样本的中心[15]。

定义 1 对于线性模式，在两状态集合 $\Omega=\{X,\neg X\}$ 的 SVM 线性模式中，样本 $x\in R^N$ 在 X 类别的三支决策条件概率定义为：样本 x 在基于样本类 $\neg X$ 的中心 m_n 到样本类 X 的中心 m_p 的特征方向 $\overrightarrow{m_nm_p}$ 上的特征距离 $s(x)$ 与中心距离 $d(m_p,m_n)$ 的比值，即为：

$$P(X|x)=s(x)/d(m_p,m_n) \tag{6}$$

其中

$$s(x)=d(x,\overrightarrow{m_nm_p})=\frac{(\overrightarrow{m_nx}\cdot\overrightarrow{m_nm_p})}{d(m_n,m_p)\cdot d(m_n,x)}\cdot d(m_n,x)$$

$$=\frac{(\overrightarrow{m_nx}\cdot\overrightarrow{m_nm_p})}{d(m_n,m_p)} \tag{7}$$

$$d(m_n,m_p)=\sqrt{\sum_{i=1}^{N}(m_n^i-m_p^i)^2} \tag{8}$$

$$d(m_n,x)=\sqrt{\sum_{i=1}^{N}(m_n^i-x^i)^2} \tag{9}$$

3.1.2 SVM 非线性可分模式下三支决策条件概率的构建

对于非线性可分的模式，采用非线性映射 Φ 把输入空间映射到某一特征空间 H。

定义 2[15] 已知样本向量组 $\{x_1,x_2,\cdots,x_n\}$，则在特征空间 H 中，样本的中心矢量 m_Φ 为：

$$m_\Phi=\frac{1}{n}\sum_{i=1}^{n}\Phi(x_i) \tag{10}$$

定理 1[16] 已知样本 $x\in R^N$ 和样本集 $\{x_1,x_2,\cdots,x_n\}$，$x_i\in R^N$，$i=1,2,\cdots,n$，则在特征空间 H 中，样本 x 到样本集的中心 m_Φ 的距离为：

$$d^H(x,m_\Phi)=(K(x,x)-\frac{2}{n}\sum_{i=1}^{n}K(x,x_i)+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}K(x_i,x_j))^{\frac{1}{2}} \tag{11}$$

定理 2[16] 已知样本集 $\{x_1,x_2,\cdots,x_{l1}\}$ 和 $\{x_1',x_2',\cdots,x_{l2}'\}$，$x_i\in R^N$，$i=1,\cdots,l_1$，$x_j'\in R^N$，$j=1,\cdots,l_2$，两类样本的中心分别为 $m_{x\Phi}$ 和 $m_{x'\Phi}$，则在特征空间 H 中的中心距离为：

$$d^H(m_{x\Phi},m_{x'\Phi})=(\frac{1}{l_1^2}\sum_{i=1}^{l_1}\sum_{j=1}^{l_1}K(x_i,x_j)-\frac{2}{l_1l_2}\sum_{i=1}^{l_1}\sum_{j=1}^{l_2}K(x_i,x_j')+\frac{1}{l_2^2}\sum_{i=1}^{l_2}\sum_{j=1}^{l_2}K(x'_i,x'_j))^{\frac{1}{2}} \tag{12}$$

定理 3 已知样本 $z\in R^N$，样本集 $\{x_1,x_2,\cdots,x_{l1}\}$ 和 $\{x_1',x_2',\cdots,x_{l2}'\}$，$x_i\in R^N$，$i=1,\cdots,l_1$，$x_j'\in R^N$，$j=1,\cdots,l_2$，两类样本的中心分别为 $m_{x\Phi}$ 和 $m_{x'\Phi}$，则在特征空间 H 中，样本 z 在基于 $\overrightarrow{m_{x\Phi}m_{x'\Phi}}$ 特征方向上的投影即特征距离为：

$$s^H(z)=(\frac{1}{l_2}\sum_{i=1}^{l_2}K(z,x_i')-\frac{1}{l_1l_2}\sum_{i=1}^{l_1}\sum_{j=1}^{l_2}K(x_i,x_j')-\frac{1}{l_1}\sum_{i=1}^{l_1}K(z,x_i)+\frac{1}{l_1^2}\sum_{i=1}^{l_1}\sum_{j=1}^{l_1}K(x_i,x_j))/d^H(m_{x\Phi},m_{x'\Phi}) \tag{13}$$

证明：

$$s^H(z)=d^H(z,\overrightarrow{m_{x\Phi}m_{x'\Phi}})$$

$$=\frac{(\overrightarrow{m_{x\Phi}\Phi(z)}\cdot\overrightarrow{m_{x\Phi}m_{x'\Phi}})}{d^H(m_{x\Phi},m_{x'\Phi})\cdot d^H(m_{x\Phi},\Phi(z))}\cdot d^H(m_{x\Phi},\Phi(z))$$

$$=\frac{(\overrightarrow{m_{x\Phi}\Phi(z)}\cdot\overrightarrow{m_{x\Phi}m_{x'\Phi}})}{d^H(m_{x\Phi},m_{x'\Phi})}$$

$$=\frac{(\Phi(z)-m_{x\Phi})\cdot(m_{x'\Phi}-m_{x\Phi})}{d^H(m_{x\Phi},m_{x'\Phi})}$$

$$=\frac{\Phi(z)\cdot m_{x'\Phi}-m_{x\Phi}\cdot m_{x'\Phi}-\Phi(z)\cdot m_{x\Phi}+m_{x\Phi}\cdot m_{x\Phi}}{d^H(m_{x\Phi},m_{x'\Phi})}$$

$$=(\frac{1}{l_2}\sum_{i=1}^{l_2}K(z,x_i')-\frac{1}{l_1l_2}\sum_{i=1}^{l_1}\sum_{j=1}^{l_2}K(x_i,x_j')-\frac{1}{l_1}\sum_{i=1}^{l_1}K(z,x_i)+\frac{1}{l_1^2}\sum_{i=1}^{l_1}\sum_{j=1}^{l_1}K(x_i,x_j))/d^H(m_{x\Phi},m_{x'\Phi}) \tag{14}$$

证毕。

定义 3 在两状态集合 $\Omega=\{X,\neg X\}$ 的 SVM 非线性可分模式中，样本 $x\in R^N$ 在 X 类别的三支决策条件概率定义为：在特征空间 H 中，样本 x 在基于样本类 $\neg X$ 的中心 $m_{n\Phi}$ 到样本类 X 的中心 $m_{p\Phi}$ 的特征方向 $\overrightarrow{m_{n\Phi}m_{p\Phi}}$ 上的特征距离 $s^H(x)$ 与中心距离 $d^H(m_{p\Phi},m_{n\Phi})$ 的比值，即为：

$$P^H(X|x)=s^H(x)/d^H(m_{p\Phi},m_{n\Phi}) \tag{15}$$

性质 1 已知两状态集合 $\Omega=\{X,\neg X\}$ 和样本 $x\in R^N$，则 x 属于类别 X 的三支决策条件概率越大，样本 x 属于类别 X 的可能性就越大。

证明：对于线性模式，设两类样本的中心分别为 m_p,m_n。由式(6)可知，$P(X|x)=s(x)/d(m_p,m_n)$，即在 $d(m_p,m_n)$ 一定时，$P(X|x)$ 越大，则 $s(x)$ 越大；又根据式(7)可知，$s(x)$ 越大，说明样本 x 在 $\overrightarrow{m_nm_p}$ 的特征方向上距离样本类 $\neg X$ 的中心越远，即样本 x 属于类别 X 的可能性就越大。非线性可分模式下的证明与线性模式类似。

3.2 基于三支决策的 SVM 边界向量构建

在三支决策中，三支决策的延迟判断部分即边界域部分表示由于信息不够、认识不足，无法在某一时刻有足够把握做出接受或拒绝的判断，但如果随着信息的增加，决策者有充分把握接受和拒绝时，问题就变为二支决策[17]。基于此，本文用三支决策的边界域来刻画 SVM 中的边界向量(BV)，有选择地淘汰非 SV。

SVM 是一种二分类模型，所以只考虑有两种状态 X 和 $\neg X$ 的二分类三支决策模型。根据 2.2 节的介绍，可计算出三支决策的阈值 α 和 β；由 3.1 节可知 SVM 线性模式和非线性可分模式下样本 x 属于类别 X 的条件概率。下面给出基于三支决策的 SVM 边界向量的定义。

定义 4 给定两状态集合 $\Omega=\{X,\neg X\}$，$X\cup\neg X=U$。在基于三支决策的 SVM 增量学习中，基于三支决策的 SVM 边界向量是被划分到三支决策边界域中的对象，定义为：

(1)SVM 线性模式下的基于三支决策的 SVM 边界向量：

$$BV=BND(X)=\{x\in U|\beta<P(X|x)<\alpha\} \tag{16}$$

(2)SVM 非线性可分模式下的基于三支决策的 SVM 边界向量：

$$BV=BND(X)=\{x\in U|\beta<P^H(X|x)<\alpha\} \tag{17}$$

性质 2 在提取基于三支决策的 SVM 边界向量时，对于线性模式，如果样本 x 被三支决策粗糙集处理，则 x 满足 $0\leqslant P(X|x)\leqslant 1$；如果 x 满足 $P(X|x)<0$ 或 $P(X|x)>1$，则样本 x 在此不予考虑。非线性可分模式下的性质类似。

证明：对于线性模式，设 SVM 分离超平面为 p，由式(6)可知，如果样本 x 满足 $P(X|x)<0$，则 $s(x)<0$，由式(7)可推断出 $\overrightarrow{m_nx}\cdot\overrightarrow{m_nm_p}<0$，即样本 x 在负类样本 $\neg X$ 一侧，且 $d(x,p)>d(m_n,p)$；如果样本 x 满足 $P(X|x)>1$，由式(6)可知 $s(x)>d(m_p,m_n)$，即样本 x 在正类样本 X 一侧，且 $d(x,p)>d(m_p,p)$。由一个点距离分离超平面的远近可以表示分类预测的确信程度[18]可知，这些样本被分类正确的可能性是很大

的，在此不予考虑。非线性可分模式下的证明与线性模式类似。

线性模式下基于三支决策的 SVM 边界向量提取算法的具体步骤如下。

算法 2 线性模式下基于三支决策的 SVM 边界向量提取算法

输入：初始训练数据集 X_0，分为 X_{0+} 和 X_{0-}

输出：X_0 中基于三支决策的 SVM 边界向量 BV

Step1 在 X_0 上训练 SVM 初始分类器；

Step2 计算 X_{0+} 和 X_{0-} 的样本中心 m_p 和 m_n；

Step3 根据 m_p 和 m_n 计算 X_0 中所有样本对应的条件概率 $P(X_{0+}|x)$；

Step4 选择相应的损失函数，求出阈值 α 和 β；

Step5 对于 Step4 中保留下来的对象，根据阈值 α 和 β 进行三支决策划分，找出对应的边界域 $BND(X_{0+})$；

Step6 令 $BV=BND(X_{0+})$，输出 BV；

Step7 算法结束。

非线性可分模式下的 BV 提取算法在特征空间中进行，具体步骤与线性模式下类似。

3.3 基于三支决策的 SVM 增量学习算法

3.3.1 算法

本文在典型的 SVM 增量学习算法的基础上，引入三支决策的相关理论，提出一种基于三支决策的 SVM 增量学习算法。

基于三支决策的 SVM 增量学习算法的具体步骤如下。

算法 3 基于三支决策的 SVM 增量学习算法

输入：训练样本集 X，随机等分为 N 个互不相交的子集，分别为 X_1，X_2，…，X_N

输出：基于 X 的 SVM 分类器 Γ

Step1 取子集 X_1 进行训练，得到支持向量集 SV_1；

Step2 根据训练结果和算法 2，计算训练集中基于三支决策的 SVM 边界向量 BV_1；

Step3 把 SV_1、BV_1 和子集 X_2 合并，将其作为新的训练样本进行训练，得到支持向量集 SV_2；

Step4 重复 Step2，将得到的新的支持向量 SV_i、BV_i 和子集 X_{i+1} 合并训练，如此循环，直到 X_N，SV_{N-1}、BV_{N-1} 与 X_N 训练所得到的支持向量机就作为训练整个样本集 X 得到的分类器 Γ，输出 Γ；

Step5 算法结束。

3.3.2 算法时间复杂度分析

基于三支决策的 SVM 增量学习算法的时间复杂度主要取决于边界向量的确定和标准 SVM 训练，边界向量的确定主要取决于算法 2 中的 Step3。因此针对线性模式下的边界向量的确定，其时间复杂度为 $O(n^2)$。在非线性可分模式下的边界向量确定过程中，由于其增加了非线性映射 Φ，因此其时间复杂度将大于 $O(n^2)$。对于标准 SVM 训练，设原始训练样本被分为 N 个互不相交的样本子集：X_1，X_2，…，X_N，每一个样本子集的大小为：l_1，l_1，…，l_N。由于标准 SVM 训练的时间复杂度为 $O(n^3)$[16]，因此子集 X_1 的训练时间为 $O(l_1^3)$，因为支持向量和边界向量只占样本子集的很少一部分，因此支持向量、边界向量和样本子集 X_i 一起进行标准 SVM 训练的时间复杂度为 $O(l_i^3)$。综上所述，线性模式下的基于三支决策的 SVM 增量学习算法的时间复杂度为：$O(l_1^3+l_1^2+l_2^3+l_2^2+\cdots+l_N^3+l_N^2)$，非线性可分模式下的时间复杂度将更大。

对于典型的 SVM 增量学习算法，其时间复杂度主要取决于标准 SVM 训练。在每次的循环中，都将前一次训练得到的 SV 和新增的样本放在一起进行再训练。设原始训练样本被分为 N 个互不相交的样本子集：X_1，X_2，…，X_N，每一个样本子集的大小为：l_1，l_1，…，l_N。因 SV 在新增样本中占少数，故典型的 SVM 增量学习算法的时间复杂度为：$O(l_1^3+l_2^3+\cdots+l_N^3)$。

对比基于三支决策的 SVM 增量学习算法和典型的 SVM 增量学习算法的时间复杂度可知：由于边界向量的确定，基于三支决策的 SVM 算法在算法运行时间上不占优势。

4 实验分析

4.1 数据来源

为了检验本文提出的算法的有效性，采用来自 UCI 数据库的 breast-cancer、heart、diabetes 3 个数据集进行实验。实验中，将每个数据集分为 3 个互不相交的子集，分别作为训练样本集、增量样本集和测试集。实验中使用的 3 个标准数据集的特性如表 1 所列。在实验时选取每个数据集中前两个属性进行实验。

表 1 数据集特性

数据集名称	样本数量	训练样本集	增量样本集	测试集
breast-cancer	683	300	200	183
heart	270	150	70	50
diabetes	768	400	200	168

4.2 数据预处理

对于 3 个数据集，将具有缺失属性值的对象移除；此外，将数据归一化到[−1,1]范围内，具体为：

$$x'=\frac{x-x_{\min}}{x_{\max}-x_{\min}}\times 2-1 \tag{18}$$

4.3 评价指标

为了评价检测性能，本文使用了两个评价指标：检测率和训练时间。

检测率：被正确分类的数据记录在总的测试集中所占的比例。计算方法如下：

$$DR=c/Z \tag{19}$$

DR 表示检测率，c 表示被正确分类的测试数据总数，Z 表示测试数据总数。

4.4 实验结果及分析

本文通过将基于三支决策的 SVM 增量学习算法与文献[3]中典型的 SVM 增量学习算法进行比较，来测试本文算法的检测率。两种方法均采用 10 折交叉验证来获取相关参数，典型的 SVM 增量学习算法的实验结果如表 2 所列。

表 2 典型的 SVM 增量学习算法测试结果

样本	检测率/%	运行时间/s
breast-cancer	85.7923	0.057648
heart	64	0.026512
diabetes	76.7857	0.089381

由于在基于三支决策的 SVM 增量学习算法中，边界向量的选取依赖于阈值参数 α、β 的设定，可根据 2.2 节的介绍由损失函数计算出阈值参数 α、β。在此为了分析阈值参数的设定对最终结果的影响趋势，将分别对每个数据集都采用不同的阈值参数进行实验，阈值参数 α、β 满足 $0\leqslant\beta<\alpha\leqslant 1$，且以

0.1的渐进度在[0,1]的范围内变化。3个数据集在不同阈值参数下的基于三支决策的SVM增量学习的测试结果如图1－图3所示。

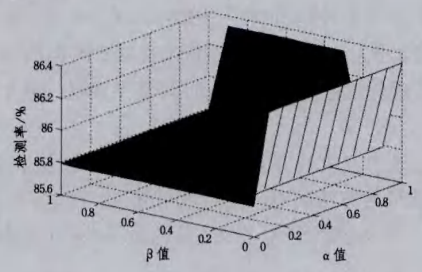

图1　breast-cancer数据集测试结果

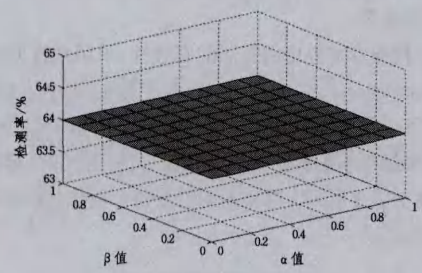

图2　heart数据集测试结果

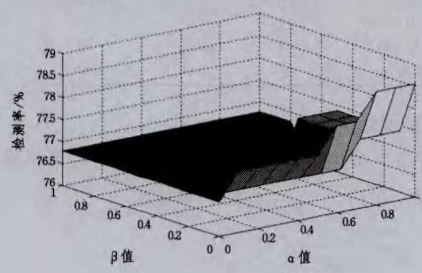

图3　diabetes数据集测试结果

从图1－图3可知：当$\alpha \leqslant \beta$时，由于不存在边界向量，基于三支决策的SVM增量学习算法的检测率和典型的SVM增量学习算法的检测率相等；从图2可知：因为heart数据集在阈值α,β变化时边界向量一直为空，所以其基于三支决策的SVM增量学习算法的检测率一直都与典型的SVM增量学习算法的检测率相等；从图1、图3可知：当$\alpha=1$并且$\beta=0$时，数据集在基于三支决策的SVM增量学习算法下的检测率高于典型SVM增量学习算法下的检测率，且此时的检测率处在全局最高点；当$\alpha=1$或$\beta=0$时，数据集在基于三支决策的SVM增量学习算法下的检测率大部分都高于典型SVM增量学习算法下的检测率。由此可知，当边界向量中的样本靠近正类或负类样本中心时，可以大大提高检测率，边界向量保留了靠近正类或负类样本中心的有用信息，在一定程度上弥补了典型的SVM增量学习算法对历史训练集中SV的过度依赖，可以提高分类的准确率。设定阈值参数时，将阈值设置为$\alpha=1$,$\beta=0$时算法的检测率最好。

为了测试基于三支决策的SVM增量学习算法的时间效率，利用Matlab的tic和toc命令记录算法的运行时间。为了与典型的SVM增量学习算法进行比较，设定阈值$\alpha=0$,$\beta=0$，使边界向量为空，省去边界向量的标准SVM训练时间。当$\alpha=0$,$\beta=0$时基于三支决策的SVM增量学习算法的测试结果如表3所列。

表3　$\alpha=0$,$\beta=0$时基于三支决策的SVM增量学习算法测试结果

样本	检测率/%	运行时间/s
breast-cancer	85.7923	0.084619
heart	64	0.045464
diabetes	76.7857	0.112223

由表3可知：由于此时的边界向量为空，因此本文所提算法的检测率与典型的SVM增量学习算法的检测率相等；由于算法2中对边界向量的判定，本文提出的算法即使在省去了边界向量的标准SVM训练时间之后，其运行时间也多于典型的SVM增量学习算法的运行时间，因此本文提出的算法在运行时间上不占优势。

结束语　针对典型SVM增量学习算法对有用信息的丢失，现有SVM增量学习算法单纯追求分类器精准性的客观性，以及三支决策中条件概率的计算问题，提出了一种基于三支决策的SVM增量学习方法。该方法具有如下优点：第一，通过三支决策的边界域考虑新增样本集对原始样本集中非SV的影响，保证了不过度依赖历史训练集中的支持向量导致有用信息的过早损失，在一定程度上纠正了典型的方法产生的系统分类错误，提高了分类准确性；第二，在机器学习中引入了三支决策方法，通过三支决策损失函数的主观性来修正现有算法的客观性；第三，采用特征距离与中心距离的比值来计算三支决策中的条件概率，解决了条件概率的计算问题。需要指出的是，本文在算法的运行时间上不占优势，因此如何提高本文算法的速度，以及对所需阈值从SVM学习角度进行研究，将是下一步要研究的工作。

参考文献

[1]　顾彬，郑关胜，王建东. 增量和减量式标准支持向量机的分析[J]. 软件学报，2013，24(7)：1-13
Gu Bin，Zheng Guan-sheng，Wang Jian-dong. Analysis for Incremental and Decremental Standard Support Vector Machine[J]. Journal of Software，2013，24(7)：1-13

[2]　张浩然，汪晓东. 回归最小二乘支持向量机的增量和在线式学习算法[J]. 计算机学报，2006，29(3)：400-406
Zhang Hao-ran，Wang Xiao-dong. Incremental and Online Learning Algorithm for Regression Least Squares Support Vector Machine[J]. Chinese Journal of Computers，2006，29(3)：400-406

[3]　Syed N，Liu H，Sung K. Incremental Learning with Support Vector Mechines[C]//Proceeding of International Joint Conference on Artificial Intelligence. Sweden：Morgan Kaufmann Publishers，1999：352-356

[4]　廖建平，余文利，方建文. 改进的增量式SVM在网络入侵检测中的应用[J]. 计算机工程与应用，2013，49(10)：100-104
Liao Jian-ping，Yu Wen-li，Fang Jian-wen. Improved incremental SVM and application in network intrusion detection[J]. Computer Engineering and Applications，2013，49(10)：100-104

[5]　王晓丹，郑春颖，吴崇明，等. 一种新的SVM对等增量学习算法[J]. 计算机应用，2006，26(10)：2440-2443
Wang Xiao-dan，Zheng Chun-ying，Wu Chong-ming，et al. New algorithm for SVM-Based incremental learnng[J]. Computer Applications，2006，26(10)：2440-2443

[6]　贾修一，商琳，周献忠，等. 三支决策理论与应用[M]. 南京：南京大学出版社，2012：20-36
Jia Xiu-yi，Shang Lin，Zhou Xian-zhong，et al. The three-way decision theory and applications[M]. Nanjing：Nanjing University

Press,2012:20-36

[7] Liu D,Li T R,Liang D C. A new discriminate analysis approach under decision-theoretic rough sets[C]//Proceedings of the 6th International Conference on Rough Sets and Knowledge Technology,Lecture Notes in Computer Science. Berlin Heidelberg: Springer-Verlag,2011:476-485

[8] 韩虎,党建武. 双隶属度模糊粗糙支持向量机[J/OL]. http://www. cnki. net/kcms/doi/10. 3778/j. jssn. 1002-8331. 1311-0260. html

Han Hu, Dang Jian-wu. Fuzzy rough support vector machine with dual membership[J/OL]. http://www. cnki. net/kcms/doi/10. 3778/j. jssn. 1002-8331. 1311-0260. html

[9] Yao Y Y. Three-way decisions with probabilistic rough sets[J]. Information Sciences,2010,180(3):341-353

[10] 刘盾,姚一豫,李天瑞. 三支决策粗糙集[J]. 计算机科学,2011,38(1):246-250

Liu Dun,Yan Yi-yu,Li Tian-rui. Three-way Decision-teheoretic Rough Sets[J]. Computer Science,2011,38(1):246-250

[11] Yao Y Y. Decision-theoretic rough set models[C]//Proceedings of the 2th International Conference on Rough Sets and Knowledge Technology, Lecture Notes in Computer Science. Berlin Heidelberg:Springer-Verlag,2007:1-12

[12] Yao Y Y. Three-way decision:an interpretation of rules in rough set theory[C]//Proceedings of the 4th International Conference on Rough Sets and Knowledge Technology, Lecture Notes in Computer Science. Berlin Heidelberg: Springer-Verlag, 2009: 642-649

[13] Liu D, Li T R, Liang D C. Incorporating logistic regression to decision-theoretic rough sets for classifications[J]. International Journal of Approximate Reasoning,2013:55(1):197-210

[14] Yao Y Y. The superiority of three-way decisions in probabilistic rough set models[J]. Information Sciences,2011,181(6):1080-1096

[15] 焦李成,张莉,周伟达. 支撑矢量预选取的中心距离比值法[J]. 电子学报,2001,29(3):383-386

Jiao Li-cheng, Zhang Li, Zhou Wei-da. Pre-extracting Support Vectors for Support Vector Machine[J]. Acta Electronica Sinica,2001,29(3):383-386

[16] 杨静,于旭,谢志强. 改进向量投影的支持向量预选取方法[J]. 计算机学报,2012,35(5):1002-1010

Yang Jing, Yu Xu, Xie Zhi-qiang. Support Vectors Pre-extracting Method Based on Improved Vector Projection[J]. Chinese Journal of Computers,2012,35(5):1002-1010

[17] 刘盾,李天瑞,李华雄. 粗糙集理论:基于三枝决策视角[J]. 南京大学学报:自然科学版,2013,49(5):574-581

Liu Dun,Li Tian-rui,Li Hua-xiong. Rough set theory:A three-way decisions perspective[J]. Journal of Nanjing University: Natural Sciences,2013,49(5):574-581

[18] 李航. 统计学习方法[M]. 北京:清华大学出版社,2012:95-130

Li Hang. Statistics learning method[M]. Beijing: Tsinghua University Press,2012:95-130

(上接第 78 页)

参 考 文 献

[1] 阎石. 数字电子技术基础(第五版)[M]. 北京:高等教育出版社,2006

Yan shi. Fundamentals of digital electronics(5th Edition)[M]. Beijing:Higher Education Press,2006

[2] 刘宝琴. 数字电路与系统(第二版)[M]. 北京:清华大学出版社,2007

Liu Bao-qin. Digital circuits and systems(2nd Edition)[M]. Beijing:Tsinghua University Press,2007

[3] 刘淑琴,虞烈,张颖红,等. 化简逻辑函数的新方法[J]. 西安交通大学学报,1998,32(8):48-51

Liu Shu-qin,Yu Lie,Zhang Ying-hong,et al. Method for simplifying logic functions[J]. Journal of Xi'an Jiaotong University,1989,32(8):48-51

[4] 朱幼莲. 计算机化简逻辑函数的算法研究[J]. 南京理工大学学报:自然科学版,2003,27(4):405-408

Zhu You-lian. Research on simplification algorithm of logical funictions with computer[J]. Journal of Nanjing University of Science and Technology:natural science,2003,27(4):405-408

[5] 王波. 关于实质本源蕴涵项的识别问题[J]. 计算机研究与发展,1995,32(12):40-44

Wang Bo. On the identification of essential prime implicants[J]. Computer Research and Developmen,1995,32(12):40-44

[6] 张义清,管致锦,吕彦明. 基于粗糙集的组合逻辑优化算法[J]. 兰州理工大学学报,2007,33(1):88-91

Zhang Yi-qing, Guan Zhi-jun, Lv Yan-ming. An algorithm of combinatory logic optimization based on rough set[J]. Journal of Lanzhou University of Techbnology,2007,33(1):88-91

[7] 张义清,吕彦鸣,李洵. 基于粗糙集多输出逻辑函数优化算法的研究[J]. 南通大学学报:自然科学版,2006,4(4):59-61

Zhang Yi-qing,Lv Yan-ming,Li Xun. An optimal algorithm of multiple output logic function based on rough set[J]. Journal of Nantong University:natural science,2006,4(4):59-61

[8] Zadeh L A. Some reflections on soft computing,granular computing and their roles in the conception,design and utilization of information/intelligent systems[J]. Soft Computing, 1998, 2(1):23-25

[9] Lin T Y. Granular Computing:Practices, Theories, and Future Directions[M]//Encyclopedia of Complexity and Systems Science. 2009:4339-4355

[10] 苗夺谦,王国胤,刘清. 粒计算:过去,现在与展望[M]. 北京:科学出版社,2007

Miao Duo-qian,Wang Guo-yin,Liu Qing. Granular computing: past,present and future[M]. Beijing:Science Press,2007

[11] 刘清. Rough 集及 Rough 推理[M]. 北京:科学出版社,2001

Liu Qing. Rought set and rough resoning[M]. Beijing: Science Press,2001

[12] Nosrati M, Hariri M. An Algorithm for Minimizing of Boolean Functions Based on Graph DS[J]. World Applied Programming, 2011,1(3):209-214

[13] Eigen D. Minimizing Boolean Sum of Products Functions[C]// CCSIT 2012. 2012:36-48

[14] Duşa A. A mathematical approach to the boolean minimization problem[J]. Quality & Quantity,2010,44(1):99-113

[15] 陈泽华,曹长青,谢刚. 基于粒矩阵的多变量真值表快速约简算法[J]. 模式识别与人工智能,2013,26(8):745-750

Chen Ze-hua,Cao Chang-qing,Xie Gang. Granular matrix based rapid reduction algorithm for multivariable truth table[J]. Pattern Recognition and Artificial Intelligence,2013,26(8):745-750