文章编号:1672-3961(2009)06-0013-11

异常检测综述

陈斌[1,2],陈松灿[1*],潘志松[3],李斌[2]
(1. 南京航空航天大学信息科学与技术学院,江苏 南京 210016; 2. 扬州大学信息工程学院,江苏 扬州 225009;
3. 解放军理工大学指挥自动化学院,江苏 南京 210007)

摘要:异常检测旨在检测出不符合期望行为的数据,因而适合应用于故障诊断、入侵和欺诈检测以及数据预处理等多个领域.针对目前众多的专用和通用异常检测方法,本文侧重对基于统计的主流异常检测方法进行了回顾,力图提供一个新的结构化的异常检测方法的认识框架,并依据其监督和无监督学习算法的原理进行了简单分类,特别对部分异常检测方法间的等价性进行了深入探讨.

关键词:异常检测;统计;监督学习;无监督学习
中图分类号:TP181 **文献标志码**:A

Survey of outlier detection technologies

CHEN Bin[1,2], CHEN Song-can[1*], PAN Zhi-song[3], LI Bin[2]
(1. Institute of Information Science and Technology, Nanjing University of Aeronautics & Astronautics, Nanjing 210016, China;
2. College of Information Engineering, Yangzhou University, Yangzhou 225009, China;
3. Institute of Command Automation, PLA University of Science Technology, Nanjing 210007, China)

Abstract: Outlier detection aims to detect those data that significantly deviate from the expected behavior, and thus is widely applied in many fields, such as, machine fault detection, intrusion detection, fraud detection and data preprocessing. Hence, there exist many generic and special algorithms for outlier detection under the unsupervised and supervised learning framework. But up to now, there still has been no clear classification in this aspect. To provide a structural view, the review of the state-of-the-art statistics-based methods for outlier detection was focusedon, and a simple classification was given in this aspect. Moreover, the equivalence between some outlier detectors in depth is particularly discussed.

Key words: outlier detection; statistics; supervised learning; unsupervised learning

0 引言

在机器故障诊断、疾病检测、入侵检测、信用卡(或保险)欺诈检测以及身份辨识等领域[1],一般只存在符合期望(目标类或正常类)行为的数据模式,而不符合期望(异常类或反类)行为的数据模式由于采样代价高昂或者采样非常困难,使得对异常行为所知甚少甚至一无所知,但是异常行为中却蕴含了显著的(通常具有很大危害甚至致命性的)行为信息[2].目前的异常检测,一般均从已知的正常类数据中进行学习,建立正常行为的模型以进行异常检测,从而构建一个假设模型 $h(x)$ 和一个阈值 ρ,当 $h(x)\geqslant\rho$ 时判 x 为正常,否则为异常,而阈值 ρ 的设定则根据训练集上所允许的经验误差 α 进行设定,使得 $P(h(x)\geqslant\rho)\geqslant 1-\alpha$,$P(\cdot)$为分布函数[3].根据该异常检测框架,目前的异常检测方法大多从假设模型的构建方式入手,由此发展出多种方

收稿日期:2009-07-16
基金项目:国家自然科学基金项目(60903130,60603029);江苏省自然科学基金项目(BK2007074)
作者简介:陈斌(1974-),男,江苏泰州人,讲师,博士研究生,研究方向为模式识别和数据分析. E-mail:chb@yzu.edu.cn
*通讯作者:陈松灿(1962-),男,浙江宁波人,博士生导师,教授,研究方向为人工智能、神经网络和模式识别.E-mail: s.chen@nuaa.edu.cn

法实现异常检测.

目前有关异常检测的综述性文章大多探索异常检测的异常产生本质,以及异常检测的应用领域及该领域的专用算法[1-2,4-8].例如在专用算法方面,2002年Bolton等[5]对金融欺诈分析领域的统计方法进行了回顾,探讨了监督学习和无监督学习方法该领域的应用,然而其设计的监督学习方法并未考虑到异常检测中存在的类别分布不平衡性;文献[6]则以网络入侵检测作为应用背景,探讨了异常检测的方法在入侵检测中的研究现状和应用.和前述专用算法不同,2001年Tax[1]根据原理对通用的异常检测器进行了分类,2003年Markou等又进一步从统计学习[7]和神经网络[8]角度对新颖性检测进行综述,阐述两种理论基础上的各种新颖性检测方法.在前人异常检测综述的基础上,文献[9]对基于支撑域的方法从核机理论方面对模型及其改进进行了探讨,因其研究的单类分类方法仅利用目标类样本,仍属于无监督异常检测方法.更为深入的综述还有文献[2],其深度挖掘异常产生的原因,并对异常检测的应用场景进行了广泛分析,但其依然按异常检测原理进行方法分类,而未能在监督和半监督学习方面深入探讨当前进展.本文在前述异常检测综述基础上,从更为广泛的异常检测入手,根据假设模型所利用的样本类别是单类(目标类)和两类(目标类和异常类)将异常检测方法划分为无监督框架和监督框架等两类,从而提供了异常检测的结构化认识框架,并深入研究通用型异常检测算法原理的最新进展.本文首先对无监督异常检测的模型原理进行深入分析和探讨,并采用一种全新的思路进行诠释基于重构的方法,从而将目前基于聚类、曲线拟合、曲面拟合以及流形以及子空间的重构都统一到同一个重构框架下;而对于监督型异常检测方法,因两类分类方法研究已经很深入,故本综述仅从异常类样本的产生方式进行分析和讨论,并探讨了异常检测中专门针对样本类别分布不平衡的异常检测方法.

文章安排如下:第1节回顾了目前无监督型的通用异常检测方法,并根据其原理进行分类,第2节则给出了监督型异常检测算法的一些最新进展,第3节给出了基于支撑域的无监督异常检测方法间的等价性,第4节对全文进行了总结,并探讨了异常检测的研究方向.

1 无监督异常检测方法

因异常检测中异常类数据常常缺乏,不能直接借用监督型学习方法,而仅能利用已知的正常类样本进行学习,故落入无监督学习范畴.根据所用原理,无监督异常检测方法可分为基于密度估计、基于重构和基于支撑域等3类,分别论述如下.

1.1 基于密度估计的方法

在无监督异常检测方法中,由于仅有一类样本可供学习,因此最简单也是最直接的方式就是通过参数化或非参数化方法来估计训练样本的密度模型,并设置密度阈值,小于该阈值的即被认为异常[10].

(1) 参数方法

最简单的密度估计方法就是假设样本服从一元高斯分布[11]:

$$p(x)=\frac{1}{\sqrt{2p\sigma}}\exp-\frac{(x-m)^2}{2\sigma^2}. \tag{1}$$

其中,μ为训练样本的均值,σ为方差.如图1,高斯分布的2σ和3σ准则分别为$P(|x-u|<2\sigma)=95\%$和$P(|x-u|<3\sigma)=99.730\%$.因此如果测试样本与均值距离在2倍或3倍方差以上,则认为该测试样本是不确定的,故可判为异常[12].但该模型太过简单而仅适用于一元单模(mode, modal)数据.

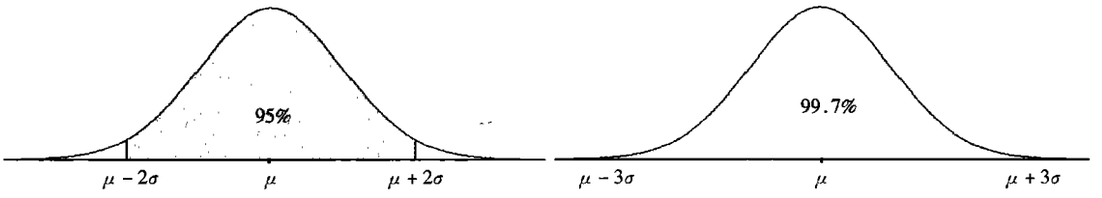

图1 一元高斯分布的2σ和3σ准则

Fig.1 2σ and 3σ rules in univariate Gaussian distribution

针对多元数据,参考文献[13]则假设目标类分布服从多元高斯模型进行异常检测,即

$$p(x)=\frac{1}{\sqrt{(2\pi)^d\det(\boldsymbol{\Sigma})}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(x-\mu)\right). \tag{2}$$

其中,d 为样本维数,$\boldsymbol{\mu}$ 为训练样本的均值,$\boldsymbol{\Sigma}$ 为训练样本的协方差矩阵.因此上两种模型过于简单仅能描述单模数据分布,因此 Sain 等人[14]对一般性的数据分布采用了混合高斯模型进行建模:

$$p(x)=\frac{1}{\sqrt{(2\pi)^d}}\sum_{j=1}^{\gamma}\alpha_j\frac{1}{\sqrt{\det(\boldsymbol{\Sigma}_j)}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_j)^{\mathrm{T}}\boldsymbol{\Sigma}_j^{-1}(\boldsymbol{x}-\mu_j)\right). \tag{3}$$

其中,d 为样本维数,α_j 是混合参数,$\boldsymbol{\mu}_j$ 和 $\boldsymbol{\Sigma}_j$ 是第 j 分量的均值与协方差矩阵,γ 为混合分量的个数.并用期望-最大化(EM)算法优化其参数,继而根据测试样本对目标类的条件概率及阈值进行异常检测.该方法不足之处是混合分量数因数据依赖而难以选择,且其需要更多的样本才能克服维数灾难[11].

(2) 非参数方法

除上述参数方法外,Bishop 等人使用 Parzen 窗方法建模训练集上的概率密度[15-16],然后根据经验风险设置相应的概率密度阈值,分类时将所有概率密度低于该阈值的测试样本判为异常.

$$\hat{p}_n(x)=\frac{1}{n}\sum_{i=1}^{n}K(x,x_i)=\frac{1}{n(\sqrt{2\pi})^d s}\sum_{i=1}^{n}\exp\left(-\frac{\|x-x_i\|^2}{2s^2}\right). \tag{4}$$

其中,n 为训练样本个数,$K(x,x_i)$是高斯核函数,s 为核带宽.

Breunig 等[17]提出了另一种无参数密度估计方法,其设计原理是基于样本点与其局部邻域样本点分隔程度的局部异常度进行异常检测,但其本质仍是利用 KNN 估计样本点的密度.该方法利用样本 o 的第 k 个近邻的距离 $k\text{-}distance(o)$ 及其 k-邻域 $N_k(p)$,定义了样本 p 和 o 之间的可达性距离 $reach\text{-}dist_k(p,o)=\max\{k\text{-}distance(o),\ d(p,o)\}$,进而引申出局部可达性密度(local reachibility density, LRD):

$$lrd_k(p)=1\Big/\left(\frac{\sum_{o\in N_k(p)}reach\text{-}dist_k(p,o)}{|N_k(p)|}\right), \tag{5}$$

根据样本 p 及其 k-邻域 $N_k(p)$的邻居 o 间的平均密度比得到局部异常因子(local outlier factor, LOF)

$$LOF_k(p)=\frac{\sum_{o\in N_k(p)}\frac{lrd_k(o)}{lrd_k(p)}}{|N_k(p)|}. \tag{6}$$

如果 p 的 LDR 低于其 k-邻域 $N_k(p)$样本点的 LRD,则其异常因子越大,相应地属于异常的可能性越大.类似无参数方法,参考文献[18]使用 KNN 图,根据入度少的样本局部密度也小的直觉,设置 KNN 图上样本入度阈值以进行异常检测.

基于密度的方法仅适合目标类训练样本结构简单、维数较低且训练样本较多的情形.但面临高维有限样本时,由于密度估计方法不能再真实地反映模式的特征,很难对目标类数据的稀疏区域做出正确的判断.

针对这种不足,Kanamori 等[19]采用最小二乘法估计样本的重要性并应用于异常检测.所谓样本 x 的重要性,指其在测试样本分布上的概率密度函数 pdf(probability density function, pdf)与训练样本分布的 pdf 间的比值:

$$w(x)=\frac{p_{te}(x)}{p_{tr}(x)}. \tag{7}$$

由于估计密度非常困难,其假设 $w(x)$由一个线性模型 $\hat{w}(x)=\boldsymbol{a}^{\mathrm{T}}(x)$构成,其中系数向量 $\boldsymbol{a}=(a_1,a_2,\cdots,a_b)$,基函数向量 $\boldsymbol{j}(x)=(j_1(x),j_2(x),\cdots,j_b(x))$,并采用最小二乘法最小化该估计 $\hat{w}(x)$和真实 $w(x)$在训练集上的平方误差,借助经验估计将很难的密度估计问题转化为一个显式的凸优化问题.根据该估计,设置重要性阈值即可进行异常检测.

1.2 基于重构的方法

基于重构的方法借助模型及其参数(如中心(或均值、原型)点、平面、曲面或流形、子空间等)的估计,同时假设目标类样本充分符合该模型,而将偏离该模型的样本判为异常.根据模型的维数和形状,可划分为点重构、平面重构以及曲面(子空间)重构等 3 种方法.

(1) 点重构方法

点重构方法主要基于样本充分聚集且可由一系列的中心(或均值、原型)点表示,并设置到这些中心的距离阈值进行异常检测.例如 K-均值($K>1$)[20]、K-中心($K>1$)[21]和 LVQ[22]等都以样本点到最近簇类中心的距离作为重构误差进行异常检测.它们均采用"两步走"的方法设计异常检测器.K-均值单分类器首先进行 K-均值聚类,然后将所有训练样本到最近簇类中心的平均距离作为阈值.与 K-均值算法采用 EM 算法更新簇类中心不同[11],LVQ 方法[13]则采用梯度下降法更新码书(簇类中心)和样本所属簇类.K-中心单分类器则首先进行 K-中心聚类,然后选取所有训练样本到其最近簇类中心的最大距离作为阈值.但是由于 K-均值以及 K-中心算法的自身缺陷,对簇类中心的选择非常敏感,一般只能得到局部最优解,并且其中的 K 值的选取目前仍是一个悬而未决的问题[1,11].而文献[23]则回避了多簇聚类中 K 值的选取问题,直接将 K 设置为1,采用可能性 1-Means 聚类方法,并设置隶属度阈值进行异常检测,因 K 设为1使得该方法可获得全局最优解.

(2) 平面重构方法

上述点重构方法等同于围绕簇类中心的球形描述,显得过于松散而难以刻画平面簇形的数据.此时基于 K-平面聚类[24]的方法能较好地适应这种场景,并以到最近超平面的距离作为重构误差进行异常检测.

K-平面聚类采用 EM 算法优化目标

$$J=\sum_{k=1}^{K}\sum_{i=1}^{n}p_{ik}\parallel w_k\cdot x_i+g_k\parallel^2 \tag{8}$$

来发现 K 个簇类超平面$(w_k,g_k)_{k=1}^{K}$,其中样本 x_i 对所有簇类超平面的隶属度 p_{ik} 之和为1,且

$$p_{ik}=\begin{cases}1, & \text{if } k=\arg\min_C \parallel w_C\cdot x_i+g_C\parallel^2,\\ 0, & \text{otherwise}.\end{cases} \tag{9}$$

参考前述 K-均值($K>1$)的阈值设置方法,使用样本点到所属簇类超平面的平均距离为重构误差阈值,即构建了 K 个带状簇类数据描述实现异常检测.

(3) 曲线或曲面、子空间重构

主分量分析(principal component analysis, PCA)[11,25]是一种线性降维方法,其通过数据特征的线性组合获得最小均方意义下数据表示的一组正交特征向量,以捕捉数据的最大变化方向,其中大特征值对应的特征向量即为其主分量.假设最大的 M 个特征向量组成的降维矩阵为 $\boldsymbol{W}^{d*M}$,针对异常检测应用,设置重构误差

$$E_{PCA}(\boldsymbol{x})=\parallel \boldsymbol{x}-(\boldsymbol{W}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W})^{-1}\boldsymbol{W}^{\mathrm{T}})\boldsymbol{x}\parallel^2=\parallel \boldsymbol{x}-\boldsymbol{W}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{x}\parallel^2 \tag{10}$$

的阈值即可执行异常检测.

PCA 由于只能捕捉单椭球形数据的方差,缺乏对多簇、曲面和流形等内在数据结构的刻画能力.而将线性问题推广到"非线性"问题是目前统计学习的研究主流[26].其中两种主要路线分别是:其一,使用核基理论,将数据映射到高维甚至无穷维特征空间,利用 Cover 定理使非线性问题近乎以概率1转化为线性问题[27],并在高维特征空间中计算主分量,例如 KPCA(kernel PCA)[28];其二,充分挖掘数据本身的结构,使用曲线或曲面以及流形等描述数据[26],例如主曲线[29]、主流形[30]以及最小生成树[31]等.

KPCA 是 PCA 的一种非线性扩展.KPCA 将数据映射到高维特征空间,即 $\phi: x_i\rightarrow\phi(x_i)$,并在特征空间中对样本 $\phi(x_i)$进行中心化:$\tilde{\phi}(x_i)=\phi(x_i)-\sum_{r=1}^{n}\phi(x_r)$,在特征空间中利用 PCA 方法抽取数据分布的若干大特征值所对应的特征向量构成降维矩阵 $\boldsymbol{W}$ 后,根据样本到 KPCA 抽取的子空间 $\boldsymbol{W}$ 的平方距离重构误差 $p(\tilde{\phi})=\tilde{\phi}\cdot\tilde{\phi}-(\boldsymbol{W}\tilde{\phi}\cdot\boldsymbol{W}\tilde{\phi})$进行异常检测.

PCA 的另一类非线性拓展是主曲线[29],它是 PCA 中第一主分量的非线性推广,其用光滑曲线代替线性主成分,得到从数据中间穿过的光滑曲线,使得数据点到该曲线的正交距离最小[26].根据每个点到主曲线的正交距离即重构误差以及所允许的经验误差的约束,设定正交距离的阈值,形成以主曲线为中心的柱形数据描述[31],参见图2.多维情形则在一维主流形即主曲线基础之上,假设数据中存在一个多维的(几何)流形结构[30],并采用一个光滑的嵌入逼近内在的流形结构,同时在最小二乘意义下使该嵌入逼近内在的流形结构,例如 ISOMAP[32]、LLE[33]等,设置各个样本上的经验逼近误差阈值进行异常检测.和上述两种方法采用一个光滑函数逼近数据中存在的主流形或主曲线不同,最小生成树数据描述[31]则寻找样本数据的拓扑结构中的主骨干形成最小生成树,进而形成围绕最小生成树上所有边的一个柱状数据描述,其中柱状描述的半径由所有训练样本到其最近邻的平均距离确定.

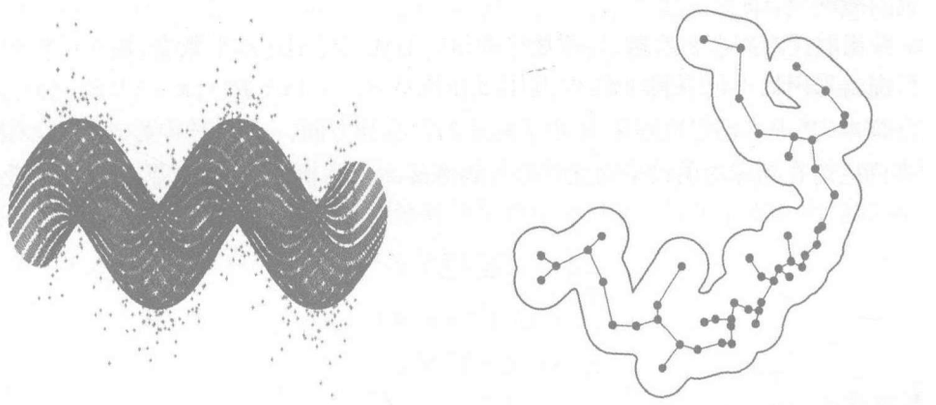

图2　主曲线和最小生成树数据描述

Fig.2　Data descriptions by principal curves and minimum spanning tree

1.3　基于支撑域的方法

该方法预先假设正常数据的描述形状,在符合给定经验误差的基础上,通过最小化体积以保证给定正常样本的正确接收率,同时最小化异常样本的误接收率,从而使得分类器的分类性能ROC曲线向上移动,提高分类器异常检测的性能.

(1) 单类支持向量机(one-class support vector machine, One-class SVM)

支持向量机(support vector machine,SVM)[34]是一种主流的两类分类方法,其通过寻求一个超平面将两类样本以最大间隔分开.在异常检测领域,由于一般仅有一类样本,故原有SVM不再适用.Scholkopf等[35]提出了One-class SVM用于异常检测,并形成一个半空间的数据描述.One-class, SVM假设原点为唯一异常点,寻求一个超平面将目标类样本和原点之间以最大间隔分开,其中硬间隔One-class SVM要求所有目标类样本都落在超平面正半侧,而软间隔One-class SVM不严格要求所有目标类样本都落在超平面的正半侧,但对落在超平面左半侧的样本点根据其和超平面的距离进行惩罚,并对经验误差和超平面和原点之间的间隔$1/\|\boldsymbol{w}\|$通过正则化因子C进行折中,相应的超平面$(\boldsymbol{w},b)$则通过以下问题求解获得($\boldsymbol{w}$为法向量,b为截距):

$$\begin{aligned}&\min \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{n}\xi_i - b\\ &\text{s.t. } \boldsymbol{w}\cdot\boldsymbol{x}_i \geqslant b - \xi_i,\\ &\qquad \xi_i \geqslant 0,\ \forall\, i.\end{aligned} \tag{11}$$

(2) 支撑向量数据描述(support vector data description, SVDD)

由于半空间的数据描述有可能太松,最小包围球[36](minimal enclosing ball, MEB)则寻找一个球形描述,其包含所有的正常类训练样本,并最小化超球$(R,\boldsymbol{a})$的体积,因其是一种硬间隔方法,易受野值点影响.为避免训练集中野值点带来的危害,SVDD采用了软间隔的思想[1, 37-38],即不严格要求训练样本到球心$\boldsymbol{a}$的平方距离小于等于R^2,但是对大于R^2的平方距离进行惩罚,并引入松持变量ξ_i放宽所有训练样本都应在球内的约束而允许部分训练样本落在球外,软间隔SVDD优化如式(12)以发现包围大多数训练样本的超球(半径为R):

$$\begin{aligned}&\min R^2 + C\sum_{i=1}^{n}\xi_i\\ &\text{s.t. } \|\boldsymbol{x}_i - \boldsymbol{a}\|^2 \leqslant R^2 + \xi_i,\ \xi_i \geqslant 0,\ \text{for all } i.\end{aligned} \tag{12}$$

这里$\boldsymbol{a}$表示超球中心,C为正则化因子.采用Lagrange乘子法,可获得上述问题的对偶问题,并将上述极小极大化问题转化为凸二次规划问题,且由其对偶$\boldsymbol{a} = \sum_{i=1}^{n}\alpha_i\boldsymbol{x}_i$表示导致球心的稀疏,提高了其测试性能.借助于核技巧,SVDD亦能获得对非线性问题灵活的数据描述.

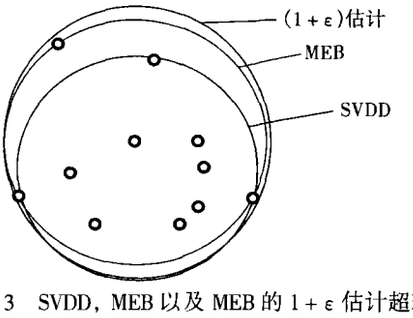

图3　SVDD, MEB以及MEB的$1+\varepsilon$估计超球

Fig.3　SVDD, MEB and its $1+\varepsilon$ approximation

为提高MEB的训练效率,文献[39]采用核心向量机(core vector machine, CVM)原理,得到了MEB的一个$1+\varepsilon$估计($(1+\varepsilon)R,\boldsymbol{a}$),其训练复杂性仅线性依赖于训练样本数,因而极大地提高了MEB的训练效率.

(3) 最小包围椭球 MVCE

由于 SVDD 采用欧氏距离求解超球,导致最终的超球描述仅能描述球形数据,缺乏对椭球形数据的描述能力.Juszczak[40]提出采用最小包围椭球描述,其用 d 维椭球 $\boldsymbol{E}_{E,c}=\{\boldsymbol{x}\in\mathbf{R}^d\mid(\boldsymbol{x}-c)^{\mathrm{T}}\boldsymbol{E}^{-1}(\boldsymbol{x}-c)\leqslant 1\}$ 来包含所有目标类训练样本,其中半正定矩阵 $\boldsymbol{E}$ 确定了椭球的性状和方向,c 是椭球中心.同样引入松驰变量 ξ_i 放宽所有训练样本都应落在椭球内的约束而允许部分训练样本落在椭球外,构成了软间隔的最小体积包围椭球:

$$\begin{aligned}&\min_{E}-\sqrt{\det(\boldsymbol{E})}+C\sum_{i=1}^{n}\xi_i\\&\text{s.t. }(\boldsymbol{x}_i-c)^{\mathrm{T}}\boldsymbol{E}^{-1}(\boldsymbol{x}_i-c)\leqslant 1+\xi_i,\\&\xi_i\geqslant 0,\ E\geqslant 0,\ \forall i.\end{aligned}\tag{13}$$

(4) 条带数据描述 Slab

在 One-class SVM 基础上,Scholkopf 等[41]提出了 Slab SVM,其在求解 One-class SVM 的超平面时,在数据两侧各用一个超平面将目标类样本进行约束,从而使目标类样本落在一个条带中.Slab SVM 主问题如下:

$$\begin{aligned}&\min\frac{1}{2}\|\boldsymbol{w}\|^2+C\sum_{i=1}^{n}(\xi_i+\xi_i^*)-b\\&\text{s.t. }\delta-\xi_i\leqslant\boldsymbol{w}\cdot\boldsymbol{x}_i-b\leqslant\delta^*+\xi_i^*,\\&\xi_i\geqslant 0,\xi_i^*\geqslant 0,\forall i.\end{aligned}\tag{14}$$

其中 δ 和 δ^* 是两个固定参数,用于确定条形带的宽度,ξ_i 和 ξ_i^* 是松弛变量,如图 4.

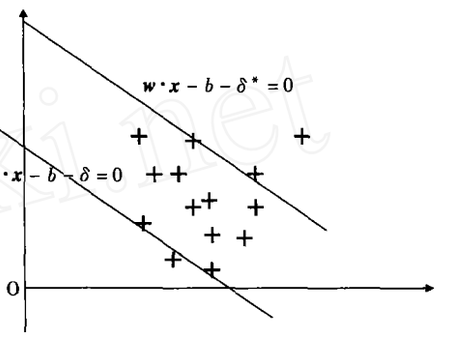

图 4 Slab SVM 原理图

Fig.4 Principles of Slab SVM

根据 Slab SVM 可得到相应的条形带数据描述,即 $\delta\leqslant\boldsymbol{w}\cdot\boldsymbol{x}-b\leqslant\delta^*$.和 Slab SVM 采用二范数和二次规划方法不同,陶等[42]又在 Slab SVM 基础上,采用一范数和 Boosting 方法[11]实现了一个线性规划的条形带数据描述.

2 有监督的异常检测方法

上述的异常检测方法都是基于无监督学习框架,仅利用已有目标类样本获得样本的数据描述.然而目前监督学习方法研究非常成熟,因而不少方法利用两类分类方法进行异常检测.其中一些方法采用人工生成异常类样本并使用现有的两类分类方法进行异常检测,而另一些方法则利用现存的少量珍贵的异常类样本构成极端不平衡问题后,设计一些能够处理不平衡问题的专用算法进行异常检测,下面分别进行介绍.

2.1 人工生成异常样本的异常检测方法

人工异常类样本的生成方法主要有 3 种:(一)目标类样本的原点对称生成;(二)样本空间内的均匀生成;(三)根据已知目标类样本的边界生成.

(1) 目标类样本的原点对称生成

核 Fisher 判别分析(kernel Fisher discriminant analysis,KFDA)[43]是线性判别分析(Fisher discriminant analysis, FDA)[11]的核化(非线性)版本,FDA 寻求一个分类面,将两类尽可能分开,并最小化类内散度、最大化类间散度,是一种具有优良性能和直觉的两类分类方法.Roth[44-45]利用 KFDA 方法进行异常检测,其首先采用原点对称法得到 $\boldsymbol{X}$ 的原点对称集 $-\boldsymbol{X}$ 为负类,即异常类样本,然后对两类样本 $\boldsymbol{X}$ 和 $-\boldsymbol{X}$ 使用 KFDA 方法获得最佳分类面,然后根据该分类面进行异常检测.

(2) 均匀分布的异常样本生成

Steinwart 等[3]则将异常检测问题解释为一个二类分类问题,其采用目标类样本空间上的均匀分布生成异常类样本,然后用这些人工产生的异常类样本和已有目标类样本构建出一个 SVM 进行异常检测.然而异常类样本服从目标类样本空间上的均匀分布的假设太过困难和苛刻,未必符合真正的异常类样本分布,故而变形出如下的异常类样本产生方法.

(3) 边界异常样本的生成

Banhalmi 等[46]设计了通过增加负类(或反类)样本的方法,扩展了无监督异常检测方法中仅有的目标类训练样本集 $\boldsymbol{X}$,进而采用标准的 ν- SVM[34]训练分类器进行异常检测.和均匀分布生成反类样本不同,其用输入目标类样本的局部边界点生成反类样本点.寻找输入目标类样本的局部边界点时利用输入样本中的每个点与其 k 近邻样本点之间能否存在一个超平面将它们完全分开(采用线性硬间隔 SVM),如存在,则该点被判为边界点,否则判为内部点,同时记录下该分隔超平面的法方向 $\boldsymbol{w}$.生成反类样本时,对每个输入目标样本点 $\boldsymbol{x}$,寻找其最近的边界点 $\boldsymbol{x}_b$,然后根据 $\boldsymbol{x}_b$ 的分离超平面法方向 $\boldsymbol{w}_b$,与两个常数间隔距离因子 $dist$ 和弯曲因子 $curv$,沿 $\boldsymbol{x}_b - \boldsymbol{x}$ 方向向外进行迁移生成样本点 $\boldsymbol{y}$:

$$\boldsymbol{y} = \boldsymbol{x}_b + (\boldsymbol{x}_b - \boldsymbol{x})\frac{T(\boldsymbol{x}, \boldsymbol{x}_b, \boldsymbol{X})}{\| \boldsymbol{x}_b - \boldsymbol{x} \|}. \tag{15}$$

其中 $T(\boldsymbol{x}, \boldsymbol{x}_b, \boldsymbol{X}) = \dfrac{dist}{dist * curv + \cos angle(\boldsymbol{x}, \boldsymbol{x}_b, \boldsymbol{X})}$, $\cos angle(\boldsymbol{x}, \boldsymbol{x}_b, \boldsymbol{X}) = \dfrac{\boldsymbol{w}_b^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}_b)}{\| \boldsymbol{w}_b \| \| \boldsymbol{x} - \boldsymbol{x}_b \|}$.如果 $\boldsymbol{y}$ 属于内部点,则将 $\boldsymbol{x}_b$ 从边界点中删除,$\boldsymbol{y}$ 亦不作为反类样本点,否则将 $\boldsymbol{y}$ 添加到生成的反类样本集中.

根据上述方法产生的反类样本集和已有的目标类样本集,采用标准的 SVM 方法求得分离超平面,并应用于异常检测.

2.2　利用现有异常样本的异常检测方法

在没有异常类样本时,采用人工方式生成异常类样本后采用标准两类分类器进行异常检测;有时异常检测任务中会存在极少数珍贵的异常类样本,此时现有的两类方法由于对样本类别分布不平衡的适应性不佳,因而不少研究者提出了如下的一些能够较好地适应该种情形的异常检测方法.

(1) 带反类的支持向量数据描述(SVDD with negative examples, NSVDD)

TAX 等[38]在提出 SVDD 后,为进一步提高 SVDD 的灵活性,提出了如下的带反类样本的 SVDD,其假设目标类样本尽可能位于超球内,而异常类样本尽可能位于超球外:

$$\begin{aligned} &\min R^2 + C_1\sum_{i=1}^{n}\xi_i + C_2\sum_{l=1}^{m}\xi_l \\ \text{s.t. } &\| \boldsymbol{x}_i - \boldsymbol{a} \|^2 \leqslant R^2 + \xi_i, \xi_i \geqslant 0, \text{ for all } i, \\ &\| \boldsymbol{x}_l - \boldsymbol{a} \|^2 \geqslant R^2 - \xi_l, \xi_l \geqslant 0, \text{ for all } l. \end{aligned} \tag{16}$$

其中 $\boldsymbol{x}_i$ 为目标类样本,$\boldsymbol{x}_l$ 为异常类样本,n 是目标类样本数,m 是异常类样本数,C_1、C_2 是平衡两类经验风险和超球体积的正则化因子.将上述形式简单变化后,得到如下问题:

$$\begin{aligned} &\min R^2 + C\sum_{i=1}^{n}\xi_i \\ \text{s.t. } &y_i(R^2 - \| \boldsymbol{x}_i - \boldsymbol{a} \|^2) \geqslant 0 - \xi_i, \\ &\xi_i \geqslant 0, \text{ for all } i. \end{aligned} \tag{17}$$

其中 $y_i \in \{+1, -1\}$ 是训练样本 $\boldsymbol{x}_i$ 的标号,目标类标号是 +1,异常类标号是 -1.由上述形式可知,NSVDD 构成的两类分类间隔是 0,因而其推广能力欠佳.

(2) 最小包围最大排斥机(minimum enclosing maximum excluding machines, MEMEM)

针对 NSVDD 的不足,MEMEM[47]为进一步提高其推广能力,其在 NSVDD 基础上应用最大间隔理论[48],优化包围目标类样本的超球$(R_1, \boldsymbol{a})$体积的同时,优化包围球和排斥异常类样本的超球$(R_2, \boldsymbol{a})$间的间隔,实现最小化包围球和最大排斥球的求解,即最大化两者的间隔.

设分类面超球$(R, \boldsymbol{a})$的半径满足 $R^2 = (R_1^2 + R_2^2)/2$,故两类间隔为 $\Delta R^2 = (R_2^2 - R_1^2)/2$,因而最小包围目标类超球$(R_1, \boldsymbol{a})$的体积和最大化排斥间隔 $\Delta R^2 = R_2^2 - R_1^2$,即 MEMEM 优化最小化 $R^2/3 - \Delta R^2$,为增加灵活性,将常数 1/3 用一个可变系数 γ 代替,即优化如下问题:

$$\begin{aligned} &\min_{\boldsymbol{a}, R^2, \Delta R^2} \gamma R^2 - \Delta R^2 \\ \text{s.t. } &\| \boldsymbol{x}_i - \boldsymbol{a} \|^2 \leqslant R^2 - \Delta R^2, \text{ for } y_i = 1; \\ &\| \boldsymbol{x}_i - \boldsymbol{a} \|^2 \geqslant R^2 + \Delta R^2, \text{ for } y_i = -1. \end{aligned} \tag{18}$$

其中 γ 为正则化因子以平衡超球体积和两类分类间隔.放宽目标类必须在内球内及异常均在外球之外的约束,问题(18)的软松弛形式为:

$$\min_{a,R^2,\Delta R^2} \gamma R^2 - \Delta R^2 + C\sum_{i=1}^{n}\xi_i$$
$$\text{s.t. } y_i(R^2 - \| x_i - a \|^2) \geqslant \Delta R^2 - \xi_i, \Delta R^2 \geqslant 0,\ \xi_i \geqslant 0. \tag{19}$$

其中 C 为平衡经验误差和 MEMEM 优化目标的正则化因子.比较 MEMEM(19)和 NSVDD(17),可知 MEMEM 的两类分类间隔 ΔR^2 一般大于 0,具有更小的 VC 维,故其推广能力更好.

3 基于支撑域的方法间的等价性关系

3.1 SVDD 和 One-class SVM 的等价性

在文献[35]和[38]中,作者分别阐述了 RBF 核[25] $K(x,z)=\exp(-\| x-z \|^2/(2s^2))$($s$ 为核带宽)下的 SVDD 和 One-class SVM 的等价性.由于采用 RBF 核,$K(x_i,x_i)$为常数,故 SVDD 的对偶问题等价与 One-class SVM 的对偶问题,故两者获得相同的 Lagrange 系数 α_i 和对应支持向量 x_i($\alpha_i \geqslant 0$),形式不同的决策函数也因相同的支持向量而导致一致的决策函数,即分类面也相同.

图 5 中展示了两者等价性的几何解释,当样本映射到 RBF 核空间后,均位于该空间的第一象限,则 SVDD 获得了和 One-class SVM 相同的支持向量,且 One-class SVM 的判别超平面将绝大多数目标类样本以最大间隔和原点分隔开,而 SVDD 则从样本中心以一个最小超球包含绝大多数的目标类样本,获得等价的判别函数[49].

3.2 One-class SVM 与 Slab SVM 的等价性

根据(14),核化后的软松弛 Slab SVM 主问题如下:

$$\min \frac{1}{2}\| w \|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) - \rho$$
$$\text{s.t. } \delta - \xi_i \leqslant w\cdot\phi(x_i) - \rho \leqslant \delta^* + \xi_i^*,$$
$$\xi_i \geqslant 0, \xi_i^* \geqslant 0, \forall i. \tag{20}$$

其中 δ 和 δ^* 是用于确定条形带的宽度的两个固定参数,C 为平衡经验误差和 Slab SVM 优化目标的正则化因子.其与 One-class SVM 的区别主要在于其存在双边约束,为证明 Slab SVM 和 One-class SVM 的等价性,主要针对其间约束的差异进行处理,以去除其中的右约束,并使其左约束和 One-class SVM 相同.

令 $\delta=0$,则其左约束 $\delta-\xi_i \leqslant w\cdot\phi(x_i)-b$ 即等价于 One-class SVM 的约束 $w\cdot\phi(x_i)\geqslant b-\xi_i$.任意原空间中的样本 x 在 RBF 核诱导的核空间中对应的映射 $\phi(x)$的模 $\| \phi(x) \|$ 为一个常量.假设 One-class SVM 求得的超平面为(w,b),因而 $w\cdot\phi(x)-b$ 在 $\phi(x)=cw$(c 为一常数)时取得一最大值,再令 $\delta^* \geqslant c\| w \|^2 - b$ 为一个足够大的常数使得所有的 $w\cdot\phi(x_i)-b\leqslant\delta^*$,即松弛变量 ξ_i^* 均为 0,因而右约束成为一个空约束,从而式(20)转换为式(11)的核化形式:

$$\min_{b,\xi,w}\frac{1}{2}\| w \|^2 + C\sum_{i=1}^{n}\xi_i - b$$
$$\text{s.t. } w\cdot\phi(x_i)\geqslant b-\xi_i,$$
$$\xi_i\geqslant 0,\ \forall i. \tag{21}$$

由于 One-class SVM 和 SVDD 在 RBF 核的特征空间中存在等价关系,又因 Slab SVM 能够在特殊的条形带宽参数 δ 和 δ^* 的设置下,转化为 One-class SVM,由等价关系的传递性知,三者在 RBF 核诱导的特征空间中相互等价.图 5 在文献[49]基础上进一步展现了在 RBF 诱导的核空间中三者的等价性,即三者具有相同的判别函数及支持向量.

4 结束语

本文对异常检测方法作了一个综述性回顾,并根据对异常类的使用与否将其分为无监督和有监督的异

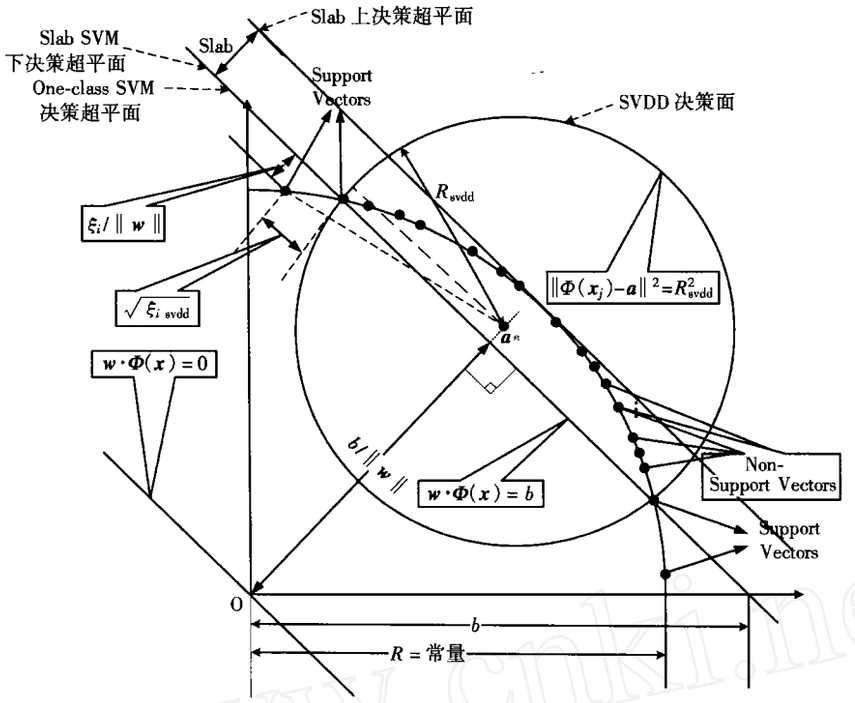

图 5　RBF 核空间中 SVDD 和 One-class SVM 以及 Slab SVM 的等价性示意图

Fig.5　Equivalence between SVDD, One-class SVM and Slab SVM in the feature space induced by RBF kernel

常检测方法.无监督的异常检测方法重点探讨了基于重构的方法,该模型将目前的基于聚类、曲线拟合、曲面拟合以及流形以及子空间的重构都统一到同一个重构框架下.而有监督的异常检测方法则重点探讨了人工样本的生成方式,以及利用现存的少量珍贵异常类样本的处理不平衡问题的两类分类方法.其后本文针对目前流行的基于支撑域的异常检测方法,重点探讨了其中几种方法之间的等价关系.

目前异常检测方法的原理已基本上涵盖机器学习的各个方面,单纯从新的原理方法进行异常检测已经相对比较困难,下一步异常检测的研究将主要集中在挖掘数据本身的结构信息,从而为后续异常检测提供一定的先验信息.另外,针对大规模数据以及联机异常检测也将成为未来的研究和应用焦点.

参考文献:

[1] TAX D. One-class classification-concept-learning in the absence of counter-examples[D]. Netherlands: Universiteit Delft, 2001.

[2] CHANDOLA V, BANERJEE A, KUMAR V. Anomaly detection: a survey[J]. ACM Computing Surveys, 2009, 41(3):1-58.

[3] STEINWART I, HUSH D, SCOVEL C. A classification framework for anomaly detection[J]. Journal of Machine Learning Research, 2005, 6:211-232.

[4] HODGE V, AUSTIN J. A survey of outlier detection methodologies[J]. Artificial Intelligence Review, 2004, 22(2):85-126.

[5] BOLTON R J, HAND D J. Statisitcal fraud detection: a review[J]. Statistical Science, 2002, 17(3):235-255.

[6] BURBECK K. Current research and use of anomaly detection[C]// WETICE'05: Proceedings of the 14th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprise Brussels, June 13-15, 2005. Washington, DC, USA: IEEE Computer Society, 2005.

[7] MARKOS M, SAMEER S. Novelty detection: a review-part I: statistical approaches[J]. Signal Processing, 2003, 83(12):2481-2497.

[8] MARKOS M, SAMEER S. Novelty detection: a review-part II: neural networks based approaches[J]. Signal Processing, 2003, 83(12): 2499-2521.

[9] 冯爱民, 陈松灿. 基于核的单类分类器研究[J]. 南京师范大学学报:工程技术版, 2008, 8(4):1-5.
FENG Aimin, CHEN Songcan. Study on one-class classifiers based on kernel method[J]. Journal of Nanjing Normal University: Engineering and Technology, 2008, 8(4):1-5.

[10] TARASSENKO L, HAYTON P, BRADY M. Novelty detection for the identification of masses in mammograms[C]// Fourth International Conference on Artificial Neural Networks, London, June 26-28, 1995. London: University of Cambridge, 1995.

[11] DUDA R, HART P, STORK D. Pattern classification[M]. 2nd ed. New York: John Wiley & Sons, 2001.

[12] PUKELSHEIM F. The three sigma rule[J]. The American Statistician, 1994, 48(2):88-91.

[13] BISHOP C. Neural networks for pattern recognition[M]. London: Oxford University Press, 1995.

[14] SAIN S R, GRAY H L, WOODWARD W A, et al. Outlier detection from a mixture distribution when training data are unlabeled[J]. Bulletin of the Seismological Society of America, 1999, 89(1):294-304.

[15] BISHOP C. Novelty detection and neural network validation [J]. IEE Proceedings on Vision, Image and Signal Processing, 1994, 141(4):217-222.

[16] YEUNG D, CHOW C. Parzen-window network intrusion detectors[C]// Proceedings of the 16th International Conference on Pattern Recognition (ICPR'02), Quebec, Canada, August 11-15, 2002. Washington, DC, USA: IEEE Computer Society, 2002.

[17] BREUNIG M M, KRIEGEL H P, NG R T, et al. LOF: identifying density-based local outliers[J]. SIGMOD Record, 2000, 29(2): 93-104.

[18] HAUTAM V, K I, FR P. Outlier detection using k-nearest neighbor graph[C]// Proceedings of the 17th International Conference on Pattern Recognition, Cambridge, UK, Aug 23-26, 2004. Washington, DC, USA: IEEE Computer Society, 2004.

[19] KANAMORI T, HIDO S, SUGIYAMA M. Efficient direct density ratio estimation for non-stationary adaptation and outlier detection [C]// Advances in Neural Information Processing Systems, Vancouver, Dec 8-13, 2008. Massachusetts: MIT Press, 2008.

[20] JIANG M F, TSENG S S, SU C M. Two-phase clustering process for outliers detection[J]. Pattern Recognition Letters, 2001, 22(6-7):691-700.

[21] HOCHBAUM D, SHMOYS D. A best possible heuristic for the *K*-center problem[J]. Mathematics of Operations Research, 1985, 10(2):180-184.

[22] LEE H, CHO S. Application of LVQ to novelty detection using outlier training data[J]. Pattern Recognition Letters, 2006, 27(13): 1572-1579.

[23] 陈斌, 冯爱民, 陈松灿,等. 基于单簇类聚类的数据描述[J]. 计算机学报, 2007, 30(8):1325-1332.
CHEN Bin, FENG Aimin, CHEN Songcan, et al. One-cluster clustering based data description[J]. Chinese Journal of Computers, 2007, 30(8):1325-1332.

[24] BRADLEY P S, MANGASARIAN O L. *k*-plane clustering[J]. Journal of Global Optimization, 2000, 16(1):22-32.

[25] 边肇祺, 张学工. 模式识别[M]. 第2版. 北京:清华大学出版社, 2000.
BIAN Zhaoqi, ZHANG Xuegong. Pattern recognition[M]. 2nd ed. Beijing: Tsinghua University Press, 2000.

[26] 张军平, 王珏. 主曲线研究综述[J]. 计算机学报, 2003, 26(2):129-146.
ZHANG Junping, WANG Jue. An overview of principal curve[J]. Chinese Journal of Computers, 2003, 26(2):129-146.

[27] COVER T. Geometrical and statistical properties of systems of linear inequalities in pattern recognition[J]. IEEE Transaction of Electronic Computers, 1965, EC-14:326-334.

[28] HOFFMANN H. Kernel PCA for novelty detection[J]. Pattern Recognition, 2007, 40(3):863-874.

[29] HASTIE T, STUETZLE W. Principal curves[J]. Journal of the American Statistical Association, 1988, 84(406):502-516.

[30] SMOLA A J, MIKA S, SCHOLKOPF B, et al. Regularized principal manifolds[J]. Journal of Machine Learning Research, 2001, 1(6):179-209.

[31] JUSZCZAK P, TAX D M J, PEKALSKA E, et al. Minimum spanning tree based one-class classifier[J]. Neurocomputing, 2009, 72(7-9):1859-1869.

[32] TENENBAUM J B, SILVA V, LANGFORD J. A global geometric framework for nonlinear dimensionality reduction [J]. Science, 2000, 290(5500):2319-2323.

[33] ROWEIS S T, SAUL L K. Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000, 290(5500):2323-2326.

[34] 邓乃扬, 田英杰. 数据挖掘中的新方法——支持向量机[M]. 北京:科学出版社, 2004.
DENG Naiyang, TIAN Yingjie. A new method of data mining: support vector machines[M]. Beijing: Science Press, 2004.

[35] SCHOLKOPF B, PLATT J, SHAWE-TAYLOR J, et al. Estimating the support of high-dimensional distribution[J]. Neural Computation, 2001, 13(7):1443-1471.

[36] SCHOLKOPF B, BURGES C, VAPNIK V. Extracting support data for a given task[C]// First International Conference on Knowledge Discovery & Data Mining, Montreal, August 20-21, 1995. Menlo Park: AAAI Press, 1995.

[37] TAX D, DUIN R. Support vector domain description[J]. Pattern Recognition Letters, 1999, 20(11-13):1191-1199.

[38] TAX D, DUIN R. Support vector data description[J]. Machine Learning, 2004, 54(1):45-66.

[39] TSANG I W, KWOK J T, CHEUNG P. Core vector machines: fast SVM training on very large data sets[J]. Journal of Machine Learning Research, 2005, 6(5):363-392.

[40] JUSZCZAK P. Learning to recognize-a study on one-class classification and active learning[D]. Netherlands: Universiteit Delft, 2006.

[41] SCHOLKOPF B, GIESEN J, SPALINGER S. Kernel methods for implicit surface modeling[C]// Advances in Neural Information Processing Systems, Vancouver, Dec 13-18, 2004. Cambridge, MA: MIT Press, 2005.

[42] TAO Q, WU G, WANG J. A new maximum margin algorithm for one-class problems and its boosting implementation[J]. Pattern Recognition, 2005, 38(10):1071-1077.

[43] SHAW-TAYLOR J, CRISTIANNI N. Kernel methods for pattern analysis[M]. London: Cambridge University Press, 2004.

[44] ROTH V. Kernel fisher discriminants for outlier detection[J]. Neural Computation, 2006, 18(4):942-960.

[45] ROTH V. Outlier detection with one-class kernel fisher discriminants[C]// Advances in Neural Information Processing Systems, Vancouver, Dec 13-18, 2004. Cambridge, MA: MIT Press, 2005.

[46] BANHALMI A, KOCSOR A, BUSA-FEKETE R. Counter-example generation-based one-class classification[C]// The 18th European Conference on Machine Learning (ECML), Warsaw, Sep 17-21, 2007. Berlin: Springer-Verlag/Heidelberg, 2007.

[47] LIU Y, ZHENG Y F. Minimum enclosing and maximum excluding machine for pattern description and discrimination[C]// The 18th International Conference on Pattern Recognition (ICPR2006), Hong Kong, Aug 20-24, 2006. New York: IEEE Press, 2006.

[48] VAPNIK V N. The nature of statistical learning theory[M]. 2nd ed. New York: Springer-Verlag, 1999.

[49] CHEN Bin, LI Bin, FENG Aimin, et al. Essential relationship between domain-based one-class classifiers and density estimation[J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2008, 25(4):275-281.

(编辑:许力琴)

(上接第7页)

[12] PLAAT A, SCHAEFFER J, PIJLS W. An algorithm faster than NegaScout and SSS* in practice[J]. Computing Science in the Netherlands, 1995: 182-193.

[13] FISHBURN J. Analysis of speedup in distributed algorithms [D]. Madison: University of Wisconsin Madison, 1981.

[14] BAUDET G. The design and analysis of algorithms for asynchronous multiprocessors[D]. Pittsburgh PA: Department of Computer Science, Carnegie-Mellon University, 1978.

[15] BRUDNO A L. Bounds and valuations for shortening the scanning of variations[J]. Problems of Cybernetics, 1963, 10: 225-241.

[16] GILLOGY J. Performance analysis of the technology chess program [D]. Pittsburgh, PA: Carnegie-Mellon University, 1978.

[17] PLAAT A, SCHAEFFER J, PIJLS W. Exploiting graph properties of game trees [C]//Proceedings of AAAI'96.[S.l.]: [s.n.], 1996: 234-239.

[18] REINEFELD A. An improvement of the Scout tree-search algorithm [J]. J Int Computer Chess Assoc, 1983, 6(4):4-14.

[19] SCHAEFFER J. The history heuristic and alpha-beta search enhancements in practice[J]. IEEE Transations on Pattern Analysis and Machine Intelligence, 1989, 11(1): 1203-1212.

[20] MARSLAND T A. A review of game-tree pruning+[J]. ICCA Journal, 1986, 9(1): 3-19.

[21] STOCKMAN G C. A minimax algorithm better than alpha-beta? [J]. Artificial Intelligence, 1979, 12(2):179-196.

[22] CAMPBELL M S, MARSLAND T A. A comparison of minimax tree search[J]. Artificial Intelligence, 1983, 20(4): 347-367.

(编辑:许力琴)

(上接第12页)

[10] SHISANU T, PRABHAS C. Parallel genetic algorithm with parameter adaptation [J]. Information Processing Letters, 2002, 82(1):47-54.

[11] 张葛祥,金炜东.量子遗传算法的改进及其应用[J].西南交通大学学报,2003,38(6):712-722.
ZHANG Gexiang, JIN Weidong. Improvement of quantum genetic algorithm and its application[J]. Journal Of Southwest Jiaotong University, 2003, 38(6):712-722.

[12] de CASTRO L N, von ZUBEN F J. Learning and mization using the clonal selection principle[J]. IEEE Transactions on Evolutionary Computation, 2002, 6(3):239-251.

[13] 黄友锐.智能优化算法及其应用[M].北京:国防工业出版社,2008:54-57.
HUANG Yourui. Intelligent optimization algorithm and its application [M]. Beijing: National Defense Industry Press, 2008:54-57.

[14] CHEN Ling, ZHANG Chunfang. Adaptive exchanging strategies in parallel ant colony algorithm[J]. Journal of Software, 2007, 18(3):617-624.

(编辑:许力琴)