

加权 KNN 分类器在 HRRP 库外目标拒判中的应用

柴 晶, 刘宏伟, 保 铮

(西安电子科技大学雷达信号处理国家重点实验室, 陕西 西安 710071)

摘 要: 针对雷达自动目标识别中的库外目标拒判问题, 提出了一种人工生成库外样本的方法和一种加权 k 最邻近(k nearest neighbors, KNN)分类器。通过人工生成库外高分辨距离像样本, 解决了在训练阶段无法获取库外样本的难题。加权 KNN 分类器同时满足了基于问题和基于数据两大设计要求, 能够很好地处理拒判问题。通过基于接收机工作特性(receiver operating characteristic, ROC)准则和基于损失函数准则的仿真实验, 证明了加权 KNN 分类器具备优良的拒判性能。

关键词: 自动目标识别; 拒判; 分类器; 高分辨距离像; 接收机工作特性; 损失函数

中图分类号: TN 959.1

文献标志码: A

Application of a weighted KNN classifier for HRRP out-of-database target rejection

CHAI Jing, LIU Hong-wei, BAO Zheng

(National Lab of Radar Signal Processing, Xidian Univ., Xi'an 710071, China)

Abstract: To cope with the out-of-database target rejection problem in radar automatic target recognition (ATR), a method that artificially generates out-of-database examples and a weighted k nearest neighbors (KNN) classifier are proposed. By artificially generating out-of-database high-resolution range profiles (HRRPs), the problem of acquiring out-of-database examples during the training step is solved. The weighted KNN classifier is both problem-dependent and data-dependent, and thus it is well suited for the associated rejection problem. Experiments conducted under the receiver operating characteristic (ROC) rule and the loss function rule respectively show that the weighted KNN classifier obtains satisfactory rejection ability.

Keywords: automatic target recognition (ATR); rejection; classifier; high-resolution range profile (HRRP); receiver operating characteristic; loss function

0 引 言

在传统的模式识别问题中, 衡量分类器性能好坏的一个重要准则是正确识别率, 然而在雷达自动目标识别 (automatic target recognition, ATR)^[1-5] 领域仅用识别率来评估分类器性能好坏是不确切的。雷达 ATR 主要针对非合作目标进行识别, 因此需要建立完备的非合作目标的数据库。目标的非合作性决定了建立完备的数据库是极其困难, 甚至无法实现的。由于数据库的非完备性, 如果有新类别的库外目标测试样本出现, 将该样本判定为非完备数据库的任何一类目标都是不合适的^[6]。因此, 给定一个测试样本, 首先需要判断它是否属于库内目标, 如果是, 则继续判定它具体属于库内目标的哪一类; 如果不是, 则它属于新目标 (即库外目标), 即对它进行拒判处理 (拒绝判定它属

于库内目标的哪一类)^[7]。本文涉及的是判定测试样本属于库内目标还是库外目标的问题, 简称为拒判问题。

拒判问题本质上是针对两类 (库内和库外) 目标的分类问题, 而解决一个两类分类问题首先需要针对两类目标进行采样训练, 然后进行测试。所面临的困难在于在训练阶段只有库内目标样本可以利用, 而无法获得库外目标样本。此外, 分类器的选择和设计也是模式识别领域的一个核心问题。一个性能优良的分类器应该是“基于问题的 (problem-dependent, PD)”和“基于数据的 (data-dependent, DD)”, 即该分类器对它所面临的问题和所处理的数据均有很强的针对性^[5]。因此, 如何有针对性地设计分类器是所面临的又一个难题。

本文的工作和创新主要包含下述 3 个方面: (1) 提出了一种人工生成高分辨距离像 (high-resolution range

收稿日期: 2008-11-11; 修回日期: 2009-06-15。

基金项目: 国家自然科学基金 (60772140); 教育部长江学者和创新团队支持计划 (IRT0645); 国防预研项目; 国防预研基金资助课题

作者简介: 柴晶 (1983-), 男, 博士研究生, 主要研究方向为雷达自动目标识别及机器学习理论。E-mail: jingchai@yahoo.cn

profile, HRRP)功率谱库外目标样本的方法,从而解决了库外目标样本的获取问题;(2)针对拒判问题的特殊性以及 HRRP 功率谱数据^[1,8]的分布特性,设计了一种加权 k 最邻近(k nearest neighbors, KNN)分类器来解决该问题。加权 KNN 分类器通过对测试样本的 K 个库内近邻样本赋予不同的权值来获取一个加权近邻样本,然后求出该样本和测试样本的欧式距离并将该距离与设定的阈值进行比较,从而判定测试样本隶属于库内样本还是库外样本;(3)通过仿真实验,用两种不同的评估准则对加权 KNN 分类器以及其他有关算法的拒判性能进行了对比,证明了所设计分类器拒判性能的优越性。

1 库外样本生成

要判定测试样本属于库内目标还是库外目标,需要分别知道库内样本和库外样本两方面的信息,即需要有一定数量的库内样本和库外样本来训练分类器。库内样本是现成的,而库外样本需要人工生成。理论上,可以假设库外样本是在全空间均匀分布的。但是在实际应用中,要生成在全空间均匀分布的样本是很困难的,因为只能生成有限数量的样本,样本数量的有限性决定了它们只能存在于部分空间。因此,需要首先设定一个有限的范围,然后生成在该范围内均匀分布的库外样本。就 HRRP 功率谱数据而言,由于每个功率谱均为一个 128 维的向量,在生成库外样本时,需要分别针对每一维设定均匀分布的上限值和下限值。这些门限值该如何设定,下面通过一个简单的例子加以说明。

假设库内样本均为 2 维向量,第 1 维服从 $[-1, 1]$ 的均匀分布,第 2 维服从 $[-10, 10]$ 的均匀分布。用两种方法生成库外样本。

方法 1 取上述 2 个均匀分布的分布范围较大值,即 $[-10, 10]$,将其扩大 2 倍,作为库外样本所有维均匀分布的范围,即库外样本的所有维均服从 $[-20, 20]$ 的均匀分布。

方法 2 分别将每一维均匀分布的范围扩大 2 倍,作为该维库外样本均匀分布的范围,即库外样本的第 1 维服从 $[-2, 2]$ 的均匀分布,第 2 维服从 $[-20, 20]$ 的均匀分布。

图 1 为库内样本和库外样本分布,图 1(a)和(b)所示库外样本分别按方法 1 和方法 2 生成,数目均为 500 个,库内样本数目均为 100 个。可以看出,按照方法 2 生成库外样本更加合理一些,它充分利用了库内样本每一维均匀分布的范围信息,既保证了库外样本能够充分完全地覆盖库内样本并留有一定的余量,又能够满足在单位面积内库外样本的分布密度维持在相对较高的水平。对于多维数据且各维的分布范围相差较大的情形,采用与方法 2 类似的方法生成库外样本尤为合适。图 2 给出了一个 HRRP 的功率谱波形。很明显,HRRP 功率谱数据的各维幅度差别很大,应采用与方法 2 类似的方法生成相应的库外样本。

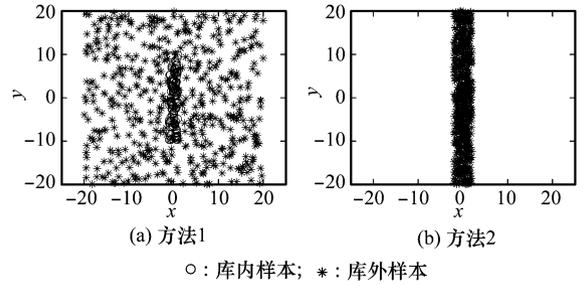


图 1 库内样本和库外样本分布

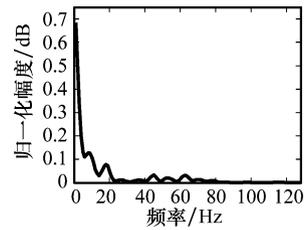


图 2 HRRP 的功率谱波形

HRRP 功率谱数据的库外样本产生方法如下:

步骤 1 假设库内样本的每一维均服从均匀分布,并求取相应的分布范围。例如对第 1 维,假设该维库内样本服从 $[a, b]$ 的均匀分布,分别找出该维样本的最大值 $\max(1)$ 和最小值 $\min(1)$,然后令 $a = \min(1)$, $b = \max(1)$ 。

步骤 2 令 $\mu(1) = (\min(1) + \max(1))/2$, $\epsilon(1) = (\max(1) - \min(1))/2$ 。

步骤 3 设定满足 $p > 1$ 的 p 值,由于功率谱数据每一维均为非负值,令 $c = \min\{0, \mu(1) - p \times \epsilon(1)\}$,然后生成服从 $[c, \mu(1) + p \times \epsilon(1)]$ 均匀分布的数据。

步骤 4 对所有维,重复步骤 1~步骤 3。

步骤 5 把针对每一维生成的均匀数据按照维数从低到高的顺序组合起来,并对其进行 2-norm 归一化处理,得到最终的库外样本数据。

可以看出,上述过程中的 p 值为每一维的库外样本与库内样本均匀分布范围之比, $p > 1$ 保证了库外样本可以完全覆盖库内样本。

2 加权 KNN 分类器

最近邻分类器(NN 分类器)作为一种简便易行的分类工具,在众多模式识别领域均取得了较为成功的应用。它同样可用于解决雷达目标识别中的拒判问题,其使用方法如下:给定测试样本 z ,在库内训练样本中找出其最近邻样本 z_{1NN} ,计算出它们之间的欧式距离 $d_{z, z_{1NN}}^2 = \|z - z_{1NN}\|_2^2$,将该距离与阈值 t 进行比较,若 $d_{z, z_{1NN}}^2 \leq t$,则判定 z 属于库内样本,否则判定 z 属于库外样本。

NN 分类器仅仅利用了一个近邻样本的信息,如果能够利用更多的近邻样本信息,即将近邻数目从 1 扩展到 K ,就有可能获取更优的拒判性能。在此用一种加权 KNN 分类器来处理拒判问题:给定测试样本 z ,在库内训练样本中找出其 K 个近邻样本 $\{z_{1NN}, \dots, z_{KNN}\}$,分别计算出 z 与它

们之间的欧式距离 $\{d_{z,z_{1NN}}^2, \dots, d_{z,z_{KNN}}^2\}$ 和相应的权值 $\{\omega_{z,z_{1NN}}, \dots, \omega_{z,z_{KNN}}\}$

$$d_{z,z_{iNN}}^2 = \|z - z_{iNN}\|_2^2, i = 1, 2, \dots, K \quad (1)$$

$$\omega_{z,z_{iNN}} = \exp(-d_{z,z_{iNN}}^2 / 2\sigma^2), i = 1, 2, \dots, K \quad (2)$$

式中, σ 为核参数, 其大小决定了权值变化趋势的快慢。

求出这 K 个近邻样本的加权平均

$$z_{\text{mean}} = \frac{\sum_{i=1}^K \omega_{z,z_{iNN}} \cdot z_{iNN}}{\sum_{i=1}^K \omega_{z,z_{iNN}}} \quad (3)$$

计算出 z 和 z_{mean} 之间的欧氏距离 $d_{z,z_{\text{mean}}}^2 = \|z - z_{\text{mean}}\|_2^2$, 将该距离与阈值 t 进行比较, 若 $d_{z,z_{\text{mean}}}^2 \leq t$, 则判定 z 属于库内样本, 否则判定 z 属于库外样本。

加权 KNN 分类器对测试样本的 K 个库内近邻样本赋予了不同的权值, 这些权值随着它们和测试样本之间欧式距离的改变而呈现负指数规律变化, 从而使和测试样本距离较近的近邻样本获得较大的权值, 进而计算 z_{mean} 时占据更大的比重。如果令式(3)中 $\omega_{z,z_{1NN}} = \omega_{z,z_{2NN}} = \dots = \omega_{z,z_{KNN}} = 1/K$, 则加权 KNN 分类器就退化为平均 KNN 分类器。平均 KNN 分类器也是一种局部分类器(只利用了 K 个近邻样本的局部信息), 但是加权 KNN 与平均 KNN 相比, 按照距离远近赋予 K 个近邻样本不同的权值, 进一步强调了局部信息的贡献, 是一种局部性更强的分类器。

3 分类器拒判性能评估准则

3.1 ROC 准则

衡量分类器拒判性能的通用评估准则是接收机工作特性(receiver operating characteristic, ROC)曲线。ROC 曲线最初是用来评估接收机检测性能的, 其横轴为虚警率, 纵轴为检测率。在此用 ROC 曲线来评估分类器的拒判性能。如果把库内样本视为正样本(positive examples, PE), 把库外样本视为负样本(negative examples, NE), 则 ROC 曲线的横轴表示负样本被错分为正目标的比率(false positive rate, FPR), 用 P_{f+} 表示; 纵轴表示正样本被正确分为正目标的比率(true positive rate, TPR), 用 P_{t+} 表示, 定义如下^[9-10]

$$P_{f+} = \frac{\text{库外(负)样本被错分为库内(正)目标的数目}}{\text{库外(负)样本总数}} \quad (4)$$

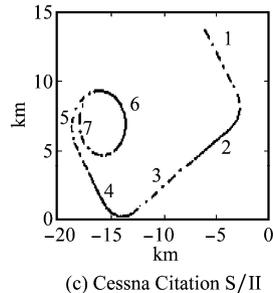
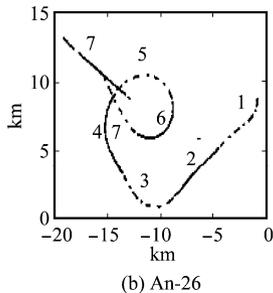
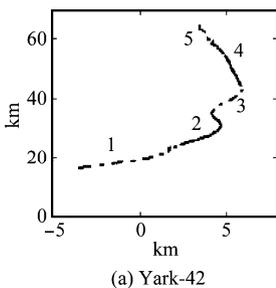


图 3 飞行轨迹在地平面的投影

库内样本的训练数据和测试数据在不同的数据段内选取, 其中“Yark-42”的第 2、5 段, “An-26”的第 5、6 段,

$$P_{t+} = \frac{\text{库内(正)样本被正确分为库内(正)目标的数目}}{\text{库内(正)样本总数}} \quad (5)$$

以加权 KNN 分类器为例(NN 分类器和平均 KNN 分类器同样适用), 可以通过使阈值 t 在一定范围内变化来得到不同的工作点, 进而得到 ROC 曲线。显而易见, ROC 曲线是一个二维曲线, ROC 曲线越靠近二维平面的左上方, 分类器的拒判性能越好。因此, 一个分类器 ROC 曲线下方的积分面积越大, 该分类器的拒判性能越好, 其面积大小可以作为一个准则来定量评估分类器的拒判性能好坏。

3.2 损失函数准则

在具体的工程应用背景下, 可以根据实际需要不同的错分行为设定不同的损失值。假设库外样本被错分为库内样本带来的损失为 $loss1$, 库内样本被错分为库外样本带来的损失为 $loss2$, 则总损失为

$$\text{总损失} = loss1 \times P_{f+} + loss2 \times (1 - P_{t+}) \quad (6)$$

为简便起见, 令 $\alpha = loss2/loss1$, 总损失也可表述为

$$\text{总损失} = P_{f+} + \alpha \times (1 - P_{t+}) \quad (7)$$

ROC 准则是一种比较通用的准则, 它可以在一般意义下衡量分类器的拒判性能。但是, 在具体的工程应用中, 要寻找的往往不是一条 ROC 曲线, 而是 ROC 曲线上一个具体的工作点。因此, 可以根据实际需要设定具体的 α 值, 然后寻找使式(7)的总损失最小的点, 并将其作为所需的工作点。

4 实验结果

选用国内某研究所的 3 类 ISAR 飞机实测数据作为库内样本, 然后人工生成库外样本, 进行仿真实验。实测数据雷达和飞机的参数如表 1、表 2 所示。

表 1 ISAR 实验雷达参数

参数	数据
中心频率/MHz	5 520
信号带宽/MHz	400

表 2 ISAR 实验飞机参数

飞机类型	机长/m	机宽/m	机高/m
Yark-42	36.38	34.88	9.83
An-26	23.80	29.20	9.83
Cessna Citation S/II	14.40	15.90	4.57

3 类飞机的飞行轨迹在地平面上的投影如图 3 所示, 可以据此大致估计飞机的方位角。

“Cessna Citation S/II”的第 6、7 段数据采样后作为训练数据, 其他各段数据采样后作为测试数据。考虑到 HRRP 数

据的平移敏感性和幅度敏感性,提取 2-norm 归一化的功率谱作为平移不变特征来进行识别和拒判,其中训练功率谱共 550 个,测试功率谱共 1 200 个。

设定 $p=1.05$,人工生成 20 000 个功率谱数据作为库外样本,其中 10 000 个用来训练,另外 10 000 个用来测试。

下面比较 SVDD (support vector data description)、NN、平均 KNN 和加权 KNN 四种分类器的拒判性能。SVDD 由 Tax 和 Duin 提出^[11],用来处理单类分类问题。假设 $\{x_i | i=1, 2, \dots, l\}$ 表示库内样本,寻求包含这些样本的一个半径为 R 的超球体,同时使半径 R 最小,以期获取包含库内样本的一个紧凑的球形边界。其数学表示形式为

$$\min R^2 + C \sum_{i=1}^l \xi_i$$

$$\text{s. t. } \|x_i - c\|^2 \leq R^2 + \xi_i, i = 1, 2, \dots, l \quad (8)$$

式中, c 和 R 分别表示超球体的球心和半径; ξ_i 为松弛变量; C 为惩罚因子。

引入核函数,将训练样本映射到核特征空间,即 $x_i \rightarrow \phi(x_i)$,式(8)在核空间可表示为

$$\min R^2 + C \sum_{i=1}^l \xi_i$$

$$\text{s. t. } \|\phi(x_i) - c\|^2 \leq R^2 + \xi_i, i = 1, 2, \dots, l \quad (9)$$

文献[11]中证明了高斯核比多项式核等其他常用核函数具有更加优良的性能,因此在本文的 SVDD 算法中,采用高斯核函数,即

$$K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle = \exp(-\|x_i - x_j\|^2 / 2\sigma^2)$$

式中, σ 为核参数。

表 3 不同分类器 ROC 曲线下的积分面积

SVDD	平均 2NN	平均 3NN	平均 4NN	平均 5NN	平均 6NN	平均 7NN	平均 8NN
0.939 40	0.998 32	0.998 13	0.997 91	0.997 72	0.997 52	0.997 27	0.997 03
NN	加权 2NN	加权 3NN	加权 4NN	加权 5NN	加权 6NN	加权 7NN	加权 8NN
0.998 48	0.999 00	0.999 50	0.999 58	0.999 55	0.999 59	0.999 57	0.999 62

4.2 基于损失函数准则的实验结果

下面用损失函数准则评估不同分类器的拒判性能。计算损失函数首先需要设定 α 值,在此令 $\alpha=1$ 。对 SVDD 而言,在训练阶段,根据总损失值最小准则选定 C 和 α 的值,然后固定这些值进行测试,得到测试阶段的总损失值;对加

4.1 基于 ROC 准则的实验结果

用 ROC 准则评估上述分类器的拒判性能。对 SVDD 而言,通过设定不同的 C 值,然后让 σ 值在一定范围内变化,可以得到不同的 ROC 曲线。因此,在训练阶段,根据 ROC 曲线下积分面积最大准则选定一个最优的 C 值,然后固定 C 值搜索 σ 值进行测试,从而得到所需的测试 ROC 曲线;对加权 KNN 分类器而言,通过设定不同的 σ 值,然后让阈值 t 在一定范围内变化,可以得到不同的 ROC 曲线。在训练阶段,根据同样的准则选定一个最优的 σ 值,然后固定 σ 值搜索 t 值进行测试,从而得到测试 ROC 曲线;对 NN 分类器和平均 KNN 分类器而言,不存在影响 ROC 曲线的参数,让阈值 t 在一定范围内变化即可得到 ROC 曲线,因此可以越过训练阶段直接搜索 t 值进行测试,从而得到测试 ROC 曲线。

图 4 给出了不同分类器的测试 ROC 曲线,其中平均 KNN 分类器和加权 KNN 分类器的 K 值分别从 2 递增至 8,图 4(a1)~(g1) 表示完整的原始 ROC 曲线,图 4(a2)~(g2) 表示沿纵轴截取的部分 ROC 曲线。

下面对图 4 中各分类器的 ROC 曲线进行定量分析,表 3 给出了各分类器 ROC 曲线下的积分面积。通过求取 ROC 曲线下的最大积分面积所对应的 K 值,可以得到参数 K 的最优值 K_{opt} 。从表 3 可得,对于平均 KNN 分类器, $K_{opt}=2$;对于加权 KNN 分类器, $K_{opt}=8$ 。实际上,当 K 取 3~8 之间的任意值时,加权 KNN 分类器总能使相应 ROC 曲线下的积分面积不小于 0.999 50,这说明在基于 ROC 曲线的评判准则下,加权 KNN 分类器具有较强的鲁棒性(对 K 的取值不是十分敏感)。

权 KNN 而言,根据同样的准则在训练阶段选定 α 和 t 的值,然后固定它们进行测试,得到测试总损失值;对 NN 和平均 KNN 而言,根据同样的准则在训练阶段选定 t 值,然后固定 t 值进行测试,得到测试总损失值。

表 4 给出了不同分类器在测试阶段的总损失值。

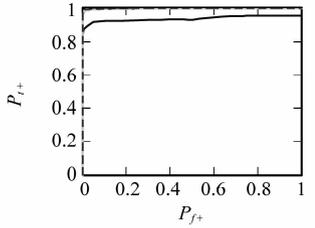
表 4 不同分类器的测试总损失值

SVDD	平均 2NN	平均 3NN	平均 4NN	平均 5NN	平均 6NN	平均 7NN	平均 8NN
0.123 9	0.024 6	0.029 2	0.032 5	0.035 4	0.038 3	0.040 1	0.041 8
NN	加权 2NN	加权 3NN	加权 4NN	加权 5NN	加权 6NN	加权 7NN	加权 8NN
0.022 0	0.015 6	0.015 3	0.011 8	0.012 4	0.012 4	0.013 1	0.012 3

通过求取分类器总损失值最小时所对应的 K 值,可以得到参数 K 的最优值 K_{opt} 。从表 4 可得,对于平均 KNN 分类器, $K_{opt}=2$;对于加权 KNN 分类器, $K_{opt}=4$ 。实际上,当 K 取 4~8 之间的任意值时,加权 KNN 分类器总能使相应的总损失值不大于 0.013 1,这说明在基于损失函数的评判准则下,加权 KNN 分类器具有较强的鲁棒性(对 K 的取值不是十分敏感)。

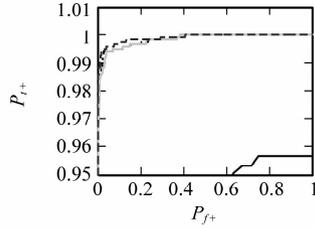
4.3 计算复杂度分析

下面比较上述各算法的计算复杂度。对 SVDD 而言,由于 SVDD 最终可以表述为一个凸二次规划 (quadratic programming, QP) 问题,其计算量主要集中于对 QP 问题的求解,计算复杂度为 $O(l^3)$,其中 l 为库内训练样本数目。对 NN、平均 KNN 和加权 KNN 三种算法而言,其计算量主要集中于搜索近邻样本,计算复杂度为 $O(l^2 d)$,其中 d 表



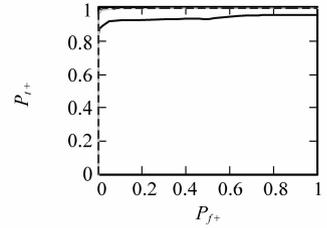
—: SVDD; - - -: NN;
 -●-: 平均2NN; - - -: 加权2NN

(a1) $K=2$ 时完整的ROC曲线



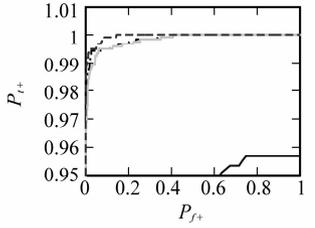
—: SVDD; - - -: NN;
 -●-: 平均2NN; - - -: 加权2NN

(a2) $K=2$ 时截取的部分ROC曲线



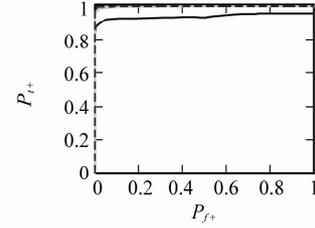
—: SVDD; - - -: NN;
 -●-: 平均3NN; - - -: 加权3NN

(b1) $K=3$ 时完整的ROC曲线



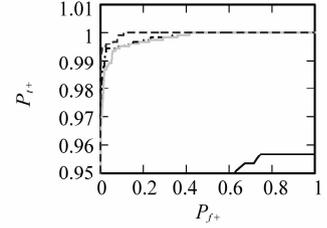
—: SVDD; - - -: NN;
 -●-: 平均3NN; - - -: 加权3NN

(b2) $K=3$ 时截取的部分ROC曲线



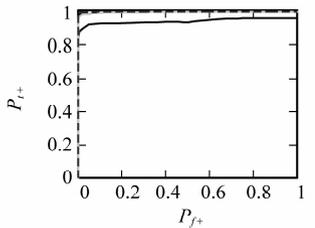
—: SVDD; - - -: NN;
 -●-: 平均4NN; - - -: 加权4NN

(c1) $K=4$ 时完整的ROC曲线



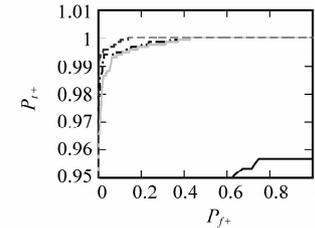
—: SVDD; - - -: NN;
 -●-: 平均4NN; - - -: 加权4NN

(c2) $K=4$ 时截取的部分ROC曲线



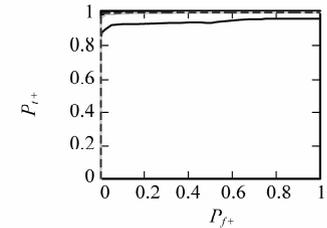
—: SVDD; - - -: NN;
 -●-: 平均5NN; - - -: 加权5NN

(d1) $K=5$ 时完整的ROC曲线



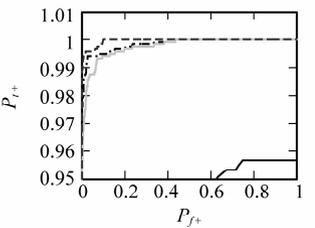
—: SVDD; - - -: NN;
 -●-: 平均5NN; - - -: 加权5NN

(d2) $K=5$ 时截取的部分ROC曲线



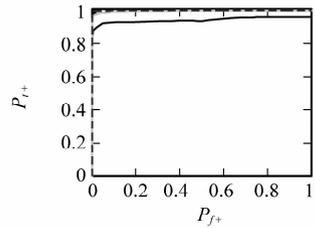
—: SVDD; - - -: NN;
 -●-: 平均6NN; - - -: 加权6NN

(e1) $K=6$ 时完整的ROC曲线



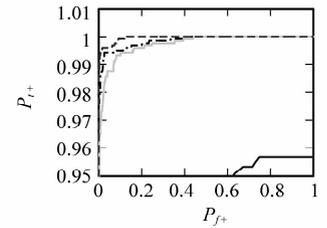
—: SVDD; - - -: NN;
 -●-: 平均6NN; - - -: 加权6NN

(e2) $K=6$ 时截取的部分ROC曲线



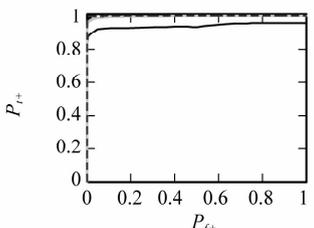
—: SVDD; - - -: NN;
 -●-: 平均7NN; - - -: 加权7NN

(f1) $K=7$ 时完整的ROC曲线



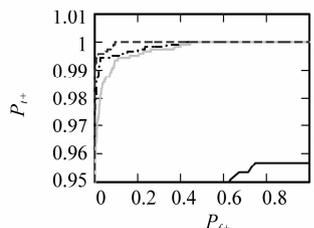
—: SVDD; - - -: NN;
 -●-: 平均7NN; - - -: 加权7NN

(f2) $K=7$ 时截取的部分ROC曲线



—: SVDD; - - -: NN;
 -●-: 平均8NN; - - -: 加权8NN

(g1) $K=8$ 时完整的ROC曲线



—: SVDD; - - -: NN;
 -●-: 平均8NN; - - -: 加权8NN

(g2) $K=8$ 时截取的部分ROC曲线

图 4 不同分类器的测试 ROC 曲线

示样本维数。对 HRRP 功率谱数据而言, $d=128, l=550$, 因此 NN、平均 KNN 和加权 KNN 三种算法的计算复杂度低于 SVDD(约为 SVDD 的 $1/3$)。

4.4 实验结果分析

综合图 4、表 3 和表 4 的实验结果,可以发现加权 KNN 分类器具备更优的拒判性能。SVDD 旨在寻求一个包含库内样本的最小超球体,虽然在训练阶段选择参数时同时使用了库内和库外两类样本,但它描述的仅仅是库内样本的分布边界,而不是两类样本的分类边界,而拒判问题实质上是关于库内和库外两类样本的分类问题,因此就拒判问题而言,SVDD 不是“基于问题的”,其拒判性能较差。NN、平均 KNN 和加权 KNN 三种分类器寻求的都是库内样本和库外样本的分类边界,它们都是“基于问题的”,因此无论是按照 ROC 准则还是按照损失函数准则来衡量,它们的拒判性能均优于 SVDD。平均 KNN 分类器比 NN 分类器利用了更多的近邻样本信息,其性能略逊于 NN,但是这不能说明利用一个近邻样本的信息比利用 K 个近邻样本的信息更适用于 HRRP 功率谱数据。加权 KNN 分类器同样利用了 K 个近邻样本的信息,但是性能却优于 NN 分类器,这说明平均 KNN 分类器的性能之所以差是因为它采用的平均准则不适合 HRRP 功率谱数据。相对于平均 KNN 分类器来说,加权 KNN 分类器更加注重局部信息的贡献,其局部性要强于平均 KNN。HRRP 功率谱样本可以视为典型的多模分布数据,对于这种数据采用局部性更强的分类器往往能够获得更优良的性能,因此相比 NN 分类器和平均 KNN 分类器而言,加权 KNN 分类器更加是“基于数据的”,其拒判性能要优于前两者。

5 结束语

本文提出了一种新的适用于处理 HRRP 功率谱目标识别中拒判问题的分类器——加权 KNN 分类器。该分类器对测试样本的 K 个库内近邻样本按照距离不同赋予了不同权值,更加注重近邻样本中局部信息的贡献,是一种局部性很强的分类器。该分类器比 SVDD 更加贴近问题模型(寻求分类边界,而不是仅仅描述库内样本的分布边界),比 NN 分类器利用了更多的近邻样本信息,比平均 KNN 分类器局部性更强,更加适用于多模分布的 HRRP 功率谱数

据。通过基于 ROC 准则和基于损失函数准则的仿真实验,证明了本文所提出的加权 KNN 分类器比上述其他分类器具备更优的拒判性能。

参考文献:

- [1] Du Lan, Liu Hongwei, Bao Zheng, et al. Radar HRRP target recognition based on higher order spectra[J]. *IEEE Trans. on Signal Processing*, 2005, 53(7): 2359 - 2368.
- [2] Du Lan, Liu Hongwei, Bao Zheng, et al. A two-distribution compounded statistical model for radar HRRP target recognition [J]. *IEEE Trans. on Signal Processing*, 2006, 54 (6): 2226 - 2238.
- [3] Chen Bo, Yuan Li, Liu Hongwei, et al. Kernel subclass discriminant analysis [J]. *Neurocomputing*, 2007, 71 (1 - 3): 455 - 458.
- [4] Chen Bo, Liu Hongwei, Bao Zheng. A kernel optimization method based on the localized kernel Fisher criterion[J]. *Pattern Recognition*, 2008, 41(3): 1098 - 1109.
- [5] Chen Bo, Liu Hongwei, Bao Zheng. Optimizing the data-dependent kernel under a unified kernel optimization framework [J]. *Pattern Recognition*, 2008, 41(6): 2107 - 2119.
- [6] Landgrebe Thomas C W, Tax David M J, Paclik Pavel, et al. The interaction between classification and reject performance for distance-based reject-option classifiers[J]. *Pattern Recognition Letters*, 2006, 27(8): 908 - 917.
- [7] 柴晶, 刘宏伟, 保铮. 一种提高雷达 HRRP 识别和拒判性能的新方法[J]. *西安电子科技大学学报:自然科学版*, 2009, 36(2): 233 - 239.
- [8] 保铮, 邢孟道, 王彤. 雷达成像技术[M]. 北京:电子工业出版社, 2006.
- [9] Landgrebe Thomas C W, Duin Robert P W. Approximating the multiclass ROC by pairwise analysis [J]. *Pattern Recognition Letters*, 2007, 28(13): 1747 - 1758.
- [10] Landgrebe Thomas C W, Duin Robert P W. Efficient multiclass ROC approximation by decomposition via confusion matrix perturbation analysis [J]. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2008, 30(5): 810 - 822.
- [11] Tax David M J, Duin Robert P W. Support vector data description[J]. *Machine Learning*, 2004, 54(1): 45 - 66.