

基于多重分形和半监督EM的LPI雷达信号识别

王星[†], 符颖, 陈游, 周一鹏, 芮鹏程

(空军工程大学 航空航天工程学院, 西安 710038)

摘要: 针对先验信息不完整的非合作电子对抗背景下的低截获概率雷达信号识别问题, 提出一种基于多重分形和半监督最大期望(EM)的识别算法. 该算法计算出信号的多重分形谱, 提取出信号的多重分形谱参数特征; 针对EM算法中全部未标记样本集的加入会造成收敛速度缓慢甚至有可能影响到分类精度的缺陷, 引入Self-training思想, 提出一种基于Self-training的半监督EM算法. 该算法通过挑选最为确定的一个或多个未标记样本来更新样本集, 使得未标记样本集不断缩小进而加快分类器的训练速度, 也可有效避免错误的累加, 在一定程度上可提高分类精度. 理论分析和仿真结果表明, 在LPI雷达信号识别问题上, 所提出的算法在不同的信噪比下具有更高的分类识别率和更好的实时性.

关键词: 多重分形; 半监督学习; Self-training; 信号识别; 低截获概率雷达

中图分类号: TN974

文献标志码: A

Radar signal recognition based on multi-fractal and semi-supervised EM algorithm

WANG Xing[†], FU Ying, CHEN You, ZHOU Yi-peng, GUO Peng-cheng

(Aeronautics and Astronautics Engineering College, Air Force Engineering University, Xi'an 710038, China)

Abstract: In order to solve incomplete prior information of the low probability of intercept(LPI) radar in non-cooperative electronic countermeasure environment, the recognition algorithm based on multi-fractal and semi-supervised expectation-maximization(EM) is proposed. Firstly, the multi-fractal spectrum of signal is calculated, and multi-fractal spectrum characteristic parameters are extracted. To solve the problem that the slower convergence speed and lower accuracy rate are caused when all unlabeled samples are joined in samples sets, this paper proposes a semi-supervised EM algorithm based on self-training ideology. By selecting one or some unlabeled samples to update sample sets, the unlabeled sampled sets are reduced, which can speed up the training speed of the classifier, avoid error accumulation, and improve the classification accuracy. The simulated results show that, the proposed algorithm has higher recognition rate and better real-time performance in LPI recognition.

Keywords: multi-fractal; semi-supervised learning; self-training; signal recognition; low probability of intercept radar

0 引言

电子战是取得信息化条件下空战胜利的根本保证, 其中对雷达的分析识别和快速应对直接影响了电子战核心战斗力的生成和转化. 传统的识别算法主要是基于全脉冲描述字(PDW)展开, 包括载频(CF)、到达时间(TOA)、脉宽(PW)、脉冲幅度(PA)、到达角(AOA)以及脉冲重复周期(PRI), 但随着电磁环境的日趋密集复杂以及雷达技术的快速发展, 出现了很多新型体制雷达, 其中低截获概率^[1](Low

probability of intercept, LPI)雷达具有高分辨、抗干扰、时宽带宽积大和低截获等特点^[1-2], 使得传统接收机难以对其进行有效截获与识别, 传统的基于PDW的识别算法已远远不能满足现代电子战的要求, 因此探索出新的雷达信号细微特征^[1-2]尤为重要.

雷达信号识别主要包括两大部分: 特征提取和分类器设计. 特征的好坏决定着识别效果, 因此提取合适的特征至关重要. 文献[3]提出了瞬时特征, 受噪声环境影响较大, 识别率低; 文献[4]提出了时频分析

收稿日期: 2017-06-21; 修回日期: 2017-09-07.

基金项目: 国家自然科学基金项目(61372167); 航空科学基金项目(20152096019).

责任编委: 侯忠生.

作者简介: 王星(1965—), 男, 教授, 博士, 从事电子对抗理论与技术等研究; 符颖(1995—), 女, 硕士生, 从事电子对抗理论与技术的研究.

[†]通讯作者. E-mail: 1571260496@qq.com

特征,虽然稳定性有所提高,但信噪比降低时识别效果较差;文献[5]提取了高阶谱特征,高阶谱特征受噪声的影响较小,但其计算过于复杂,实时性较差.分形特征是表征信号复杂度和不规则度^[6]的物理量,目前在目标检测^[5]、目标分类^[7]、调制模式识别^[8]等领域应用广泛.分形理论包括单分形维数和多重分形,其中单分形维数只从统计意义上对信号进行刻画,不足以将信号中复杂繁多的各种分形结构区分开来,而多重分形能够从全局和局部刻画信号的结构,且计算量较少,因此本文基于多重分形理论,提取出信号的多重分形谱参数特征,作为后续识别的输入.

针对分类器设计部分,目前越来越多的学者致力于机器学习理论的研究,半监督学习就是其中一种比较热门的机器学习算法.半监督学习方法是在少量的有标签样本数据下,利用大量的无标签样本数据来提高训练学习效果^[9],再结合电子对抗工作的非合作特性,考虑将半监督学习用于信号识别中.最大期望(Expectation-maximization, EM)算法是在数据不完整的情况下求解最大似然估计问题的迭代数值计算算法^[10],但在EM算法中全部未标记样本集的加入会造成收敛速度缓慢甚至有可能影响到分类精度,因此本文引入Self-training思想对其进行改进.在更新迭代过程中,通过挑选最为确定的一个或多个未标记样本来更新训练样本集,这样在每次迭代过程中都会促使未标记样本集不断缩小而加快分类器的训练速度,同时通过挑选出最确定的未标记样本可有效避免错误的累加,一定程度上可提高分类精度.

本文选取4种典型的LPI雷达信号作为研究对象,分别为FMCW、Frank、BPSK和FSK/PSK.计算出这4种信号的多重分形谱并提取出信号的多重分形谱参数特征,作为后续分类器的输入;构建基于Self-training的半监督EM算法分类器,从而完成对不同类型LPI雷达信号的识别.

1 多重分形

1.1 多重分形谱的计算

多重分形又称多标度分形,是由许多标量指数的奇异测度所组成的集合^[11].多重分形以系统的局部特征为切入点,使用广义维数和多重分形谱函数来分析测度集合在不同层次上的标度特征,最终研究得到系统的整体行为.本文从多重分形谱函数的角度进行研究.

为了研究分形集的标度特征,多重分形谱从分形集的测度(归一化分布)的奇异性进行分析.一般而

言,测度的奇异性包含两个指标:奇异指数 $\alpha(q)$ 和对应的奇异谱函数 $f(\alpha(q))$ ^[12].具体计算步骤如下:

1) 概率测度 $P_i(\delta)$.

对于任意一雷达信号序列 $s(t)$,沿时间轴将其划分为尺度为 δ 的一维小盒子,设 $N_i(\delta)$ 表示尺寸为 δ 时第 i 个盒子内所有雷达信号幅度值相加之和,则第 i 个盒子内平均幅度的概率测度为

$$P_i(\delta) = \frac{N_i(\delta)}{\sum N_i(\delta)}, \quad (1)$$

其中 $\sum N_i(\delta)$ 为所有雷达信号幅度之和.

2) 配分函数 $\chi_q(\delta)$.

对概率测度 $P_i(\delta)$ 的 q 次方进行加权求和,计算得出雷达信号的配分函数为

$$\chi_q(\delta) = \sum P_i(\delta)^q. \quad (2)$$

其中: q 为权重因子,表示概率测度 $P_i(\delta)$ 对于配分函数 $\chi_q(\delta)$ 的贡献大小.当 $q > 1$ 时,较大值的概率测度 $P_i(\delta)$ 贡献大;当 $q < 1$ 时,较小值的概率测度 $P_i(\delta)$ 贡献大.

3) 尺度指数 $\tau(q)$.

在无标度区域内,配分函数 $\chi_q(\delta)$ 与尺度 δ 的关系满足

$$\chi_q(\delta) \propto \delta^{\tau(q)}, \quad (3)$$

其中 $\tau(q)$ 为尺度指数.根据上式的幂指关系,利用 $\log_2 \chi_q(\delta)$ 作为函数 $\log_2 \delta$ 的线性衰减计算 $\tau(q)$,即

$$\log_2 \chi_q(\delta) \approx \tau(q) \log_2 \delta + c(q). \quad (4)$$

4) 多重分形谱 $f(\alpha(q))$ 和奇异指数 $\alpha(q)$.

根据Legendre变换建立奇异指数 $\alpha(q)$ 、多重分形谱 $f(\alpha(q))$ 和尺度指数 $\tau(q)$ 这三个物理量之间的关系,有

$$\begin{cases} \alpha(q) = \frac{d\tau(q)}{dq}, \\ f(\alpha) = q\alpha(q) - \tau(q). \end{cases} \quad (5)$$

计算式(5)即可得到多重分形谱 $f(\alpha(q))$ 和奇异指数 $\alpha(q)$ 的具体数值.

1.2 多重分形谱的几何特性

通过上述内容对配分函数、尺度指数进行求解,可以得到雷达信号的多重分形谱与奇异指数(即 $f(\alpha(q)) \sim \alpha(q)$)的关系图,多重分形谱图的几何特性为:

- 1) $f(\alpha(q))$ 恒大于等于0.
- 2) $f(\alpha(q))$ 关于 $\alpha(q)$ 是上凸的.

3) $f(\alpha(q))$ 表示不同的多重分形子集维数的连续谱. 若分形对象是单重的, 则多重分形谱图是一固定值; 反之, 若是多重的, 则多重分形谱图是一条连续的单峰曲线.

4) 理想的多重分形谱图 $f(\alpha(q))$ 与直线 $y = x$ 相切, 在同一坐标系中两个多重分形谱曲线在其 $q > 0$ 部分有且只有一个交点.

2 基于多重分形谱的LPI信号提取

本文研究4种典型的LPI雷达信号, 分别是调频连续波FMCW信号、Costas信号、Frank信号以及基于Costas的FSK/PSK混合信号. 为了从多重分形谱中提取有效的特征参数, 首先计算出各个信号的多重分形谱和奇异指数(即 $f(\alpha(q)) \sim \alpha(q)$)的关系图, 如图1所示.

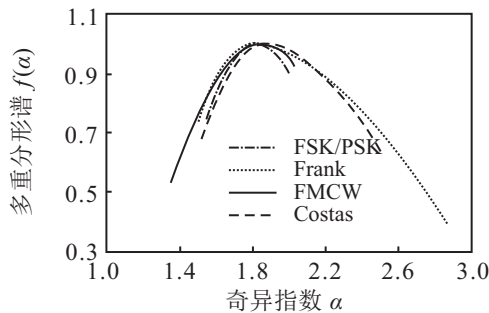


图1 多重分形谱

在图1中: 横轴为奇异指数, 又称 Lipschitz-holder 指数, 表示概率测度的奇异性强度, 其奇异性强度随着奇异指数的增加而减弱; 纵轴为多重分形谱 $f(\alpha(q))$, 又称奇异谱函数, 表示分形体的复杂性、不均匀性. 从图1中可看出, $f(\alpha(q))$ 是单峰曲线, 且是上凸的, 说明LPI雷达信号具有多重分形特性. 另外, 不同信号的多重分形谱具有明显差异, 表明多重分形谱可以作为区分不同信号的依据. 为了更细致地刻画LPI信号多重分形特性, 本文在分析图1的基础上, 提取下面几个多重分形谱图特征参数:

1) 多重分形谱谱宽 $\Delta\alpha$.

$\Delta\alpha$ 是雷达信号最大概率与最小概率所对应的奇异性强度差值, 数学表达式为

$$\Delta\alpha = \alpha_{\max} - \alpha_{\min},$$

它可反映出信号的幅值波动, 表示信号分形体奇异性分布范围的大小, 即雷达信号概率测度不均匀性的衡量. $\Delta\alpha$ 越大, 信号幅值波动越大, 时间序列幅值分布越不均匀. 从图1中可以看出, 不同雷达信号的多重分形谱谱宽 $\Delta\alpha$ 具有明显的不同, 可作为信号识别的特征之一.

2) 多重分形谱值差 Δf .

多重分形谱值差 Δf 是指最小奇异强度与最大奇异强度所对应的多重分形谱值的差值, 即 $\Delta f = f(\alpha_{\min}) - f(\alpha_{\max})$, 描述的是信号时间轴处于峰值与谷值数目的比例. 当 Δf 大于0时, 最小奇异指数所对应的奇异谱值大于最大奇异指数所对应的奇异谱值, 说明 $f(\alpha(q))$ 此时呈左偏移状态, 此时信号幅度达到波峰的频次较高; 反之, Δf 小于0, 说明 $f(\alpha(q))$ 呈右偏移状态, 此时信号更多地停留在波谷^[13];

3) 多重分形谱偏移度 W .

定义多重分形谱偏移度

$$W = \frac{\alpha_m - \alpha_{\min}}{\alpha_{\max} - \alpha_m}, \quad (6)$$

其中 α_m 为 $f(\alpha(q))$ 取最大值时的奇异值. 参数 W 描述的是多重分形谱的对称性, 当 $k > 1$ 时, 多重分形谱图顶点偏左, 此时信号奇异性较强; 当 $k < 1$ 时, 信号奇异性较弱.

将提取的上述多重分形谱参数作为特征向量, 用于后续分类器的输入, 完成对LPI信号的识别过程.

3 基于Self-training的半监督EM算法

3.1 EM算法

EM算法, 即最大期望算法, 是一种最大似然估计的迭代数值计算算法, 根据已存在的数据, 借助未知变量, 通过期望值之间的迭代估计出似然函数, 使之收敛于某个最优解, 是一种简单的经典学习算法^[14].

给定有标记样本数据集 $X = \{(x_1, l_1), (x_2, l_2), \dots, (x_L, l_L)\}$, 假定这些有标记样本数据是由特定的 $p(x/\theta)$ 独立同分布采样获得的, θ 是定义分布的参数, 则对数似然函数为

$$\log L(\theta/X) = \log p(X/\theta) = \sum_{i=1}^L \log p(x_i/\theta). \quad (7)$$

最大似然问题就是寻求最优参数(分类器) θ , 使之对应的似然函数值最大, 即

$$\hat{\theta}_{ml} = \arg \max \log L(\theta/X). \quad (8)$$

在观测到的有标记样本集 X 中, 引入未知变量(无标记样本集) Y , 形成完备数据集

$$D = \{X, Y\} = \{(x_1, l_1), (x_2, l_2), \dots, (x_L, l_L), x_{L+1}, \dots, x_{L+U}\},$$

则完备整数据集 D 所对应的似然函数为

$$\log L(\theta/D) = \log P(D/\theta) = \log P(X, Y/\theta) = \log(P(Y/X, \theta)P(X/\theta)). \quad (9)$$

定义期望值 $Q(\theta/\theta^{(t)})$: 表示在有标记样本集 X

基础上完备数据集 D 的对数似然函数的期望值,其中 $\theta^{(t)} (t = 0, 1, \dots)$ 是迭代过程中第 t 次似然估计得到的最优值,则有

$$\begin{aligned} Q(\theta/\theta^{(t)}) &= E\{\log L(\theta/D)/X, \theta^{(t)}\} = \\ &E\{\log P(D/\theta)/X, \theta^{(t)}\} = \\ &\int [\log P(D/\theta)]P(Y/X, \theta^{(t)})dy. \end{aligned} \quad (10)$$

定义隐藏变量 z_{ij} : 表示未标记样本 Y 缺失的标签信息,则有

$$z_{ij} = \begin{cases} 1, & l_i = c_j; \\ 0, & l_i \neq c_j; \end{cases} \quad j = 1, 2, \dots, C. \quad (11)$$

其中 $C = \{c_1, c_2, \dots, c_c\}$ 为样本的标签(类别)集合.

通过以上分析可以得出,EM算法主要包括两个求解步骤:

1) E步(求解期望): 初始化参数 $\theta^{(0)}$, 在有标记样本集 X 的基础上, 求解未知变量(无标记样本集) Y 的条件数学期望 $Q(\theta/\theta^{(t)})$, 即

$$z^{(t+1)} = Q(\theta/\theta^{(t)}). \quad (12)$$

2) M步(求解参数最大化): 求解最优参数(分类器) $\theta^{(t)}$, 使得条件数学期望 $Q(\theta/\theta^{(t)})$ 的极大似然估计值最大, 假定此时最大化的最优值为 $\theta^{(t+1)}$, 则更新后的 $\theta^{(t)}$ 最优参数为

$$Q(\theta^{(t+1)}/\theta^{(t)}, X) = \max Q(\theta/\theta^{(t)}, X). \quad (13)$$

返回E步, 重复上述步骤, 直到满足收敛条件时停止.

EM算法通过计算参数 $\theta^{(t)}$ 来促使未知变量(无标记样本)的极大似然估计概率达到最大, 有效利用了无标记样本集中隐藏的有用信息, 但在更新迭代过程中, EM算法并未对未标记样本作出有效判定就加入至下一轮求解参数中, 这就使得中间分类器更新速度缓慢, 从而影响收敛速度. 此外, 当对未标记样本的计算过程出现错误时, 错误会进行累加甚至会影响到分类精度.

3.2 基于Self-training的半监督EM算法

针对EM算法中未标记样本集的加入会造成收敛速度缓慢甚至有可能影响到分类精度的缺陷, 考虑将Self-training思想引入EM算法中.

半监督学习是机器学习领域的一个分支, 是指同时利用少量的有标记样本和大量的无标记样本进行训练学习分类器的过程, 通常在很少的有标记样本下就能得到理想的效果. Self-training思想是半监督理论中很重要的一部分, 是基于自身的预测结果来不断

对自身进行更新的过程. 假定存在有标记样本集 X , 无标记样本集 Y , Self-training思想具体步骤为: 首先, 利用有限的有标记样本 X 进行训练学习, 构造初始分类器 f_0 ; 然后, 使用初始分类器 f_0 对大量的无标记样本进行标记, 得到类标签, 从中挑选最为确定的无标记样本, 将挑选出的无标记样本连同其所对应的类标签添加至有标记样本集 X 中, 得到更新后的样本集 $X^{(1)}$, 再利用 $X^{(1)}$ 训练学习分类器, 重复此过程, 直至所有的无标记样本都被标记或者达到收敛条件即可.

本文将Self-training思想引入至EM算法中, 在E步更新迭代的计算过程中, 首先挑选出所有隐藏变量 z_{ij} 中最大的 $z_{i^*j^*}$, 将其对应的样本 x_i^* 从集合 Y 中剔除掉, 然后将 (d_i^*, l_j^*) 添加至集合 X 中更新样本集, 并且将 $z_{i^*j^*}$ 置1而 z_{i^*j} ($j \neq j^*$) 置0. 具体步骤如表1所示.

表1 基于Self-training的EM算法

Input X, Y ;
Set $t = 0$;
Initialize $\theta^{(0)} = \arg \max_{\theta} P(\theta/X)$;
While $Y \neq \emptyset$ do:
E Step: Set $z^{(t+1)} = E[z/D; \theta^{(t)}]$;
Set $(i^*, j^*) = \arg \max_{(i,j)} \{z_{ij}/d_i \mid d_i \in Y\}$;
Set $X = X \cup \{(d_i^*, l_j^*)\}$;
Set $U = U \setminus \{d_i^*\}$;
For $j = 1$ to $ C $:
Set $z_{ij}^* = 0$;
Set $z_{i^*j^*}^* = 1$;
M Step: Set $\theta^{(t+1)} = \arg \max_{\theta} P(\theta/D; z^{(t+1)})$;
Set $t = t + 1$;
Output $\theta^{(t)}$.

基于Self-training的半监督EM算法在迭代过程中, 每次挑选出E步中最确定的一个或多个未标记样本, 将他们从未标记样本集 Y 中剔除并添加至有标记样本集 X 中, 然后根据更新后的样本集在M步中进行分类器的训练学习, 每次迭代过程中都会使得未标记样本集 Y 不断缩小, 从而加快分类器的训练速度, 同时通过挑选出最确定的未标记样本可有效避免错误的累加, 一定程度上可提高分类精度.

4 基于多重分形和半监督EM算法的LPI雷达信号识别

结合以上分析, 给出基于多重分形特征提取和Self-training的半监督EM算法的LPI雷达信号识别流程, 如图2所示.

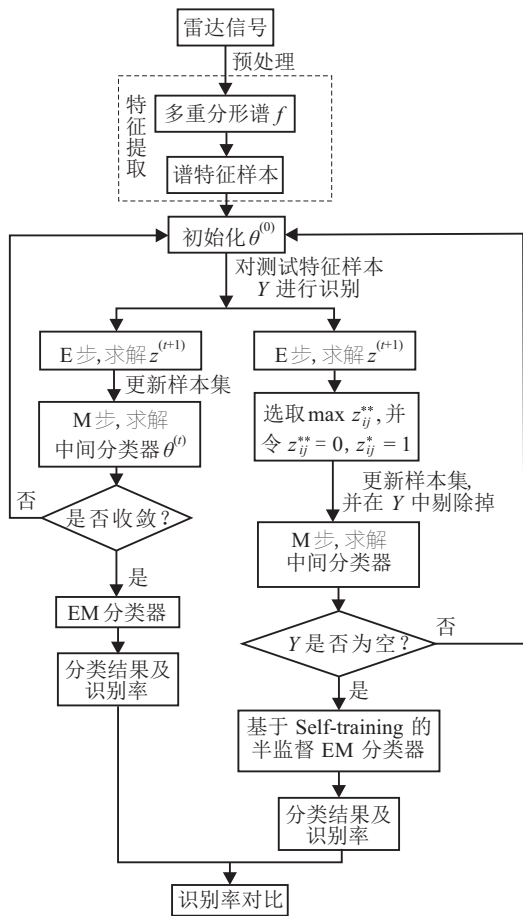


图2 基于多重分形和半监督EM的LPI雷达信号识别流程

基于多重分形和半监督EM的LPI雷达信号识别主要包括3大部分:第1部分为LPI雷达信号预处理阶段.将待训练的雷达信号进行参数测量,分选预处理.第2部分为LPI雷达信号特征提取阶段.分别计算出雷达信号的多重分形谱 $f(\alpha(q))$,提取谱参数特征作为特征向量.第3部分为LPI雷达信号分类器设计阶段.在0~20 dB的信噪比范围内,随机选取样本集的一部分作为已标记样本(本文中选取样本集的5%、10%、15%、20%、25%作为已标识样本),剩余的作为未标记样本(测试样本).将有标识样本与无标识样本组成训练集合,按照如图2所示步骤分别训练学习构造EM分类器以及Self-training的EM分类器.将测试特征样本(无标识样本)作为已经训练好的EM分类器以及Self-training的EM分类器的输入,从而完成对不同类型的LPI雷达信号的识别.

5 实例仿真

5.1 预处理

实验测试在PC上进行,实验环境为:CPU主频为3.30 GHz,内存为4 GB,Win7,32位操作系统.本文使用的LPI雷达信号均通过Low probability of intercept

toolbox工具箱^[15]产生,具体每种信号参数设置如下:

FMCW信号:载频为300 MHz,采样频率为700 MHz,调制带宽 ΔF 范围为30 MHz,信号长度为500,信号幅度为1,采用三角波波形;

Costas信号:采样频率为300 MHz,频率序列为[3, 2, 6, 4, 5, 1] MHz和[2, 4, 8, 5, 10, 9, 7, 3, 6, 1] MHz,信号长度为500,信号幅度为1;

Frank信号:载频为300 MHz,采样频率为700 MHz,步进频率为8个;

FSK/PSK信号:采用基于Costas序列跳频,跳频序列为[5, 4, 6, 2, 3, 1] MHz,信号长度为500,信号幅度为1,采样频率为700 MHz.

多重分形谱参数 q 取值区间为 $[-2, 2]$,步进为0.1,对于每种LPI雷达信号,在0~20 dB信噪比范围每隔1 dB产生100个特征样本,样本总数为2 100个.

在多重分形谱参数的特征提取过程中,噪声有可能会对整个过程中造成影响,图3~图6给出了去噪前多重分形谱与奇异指数之间的关系谱图.

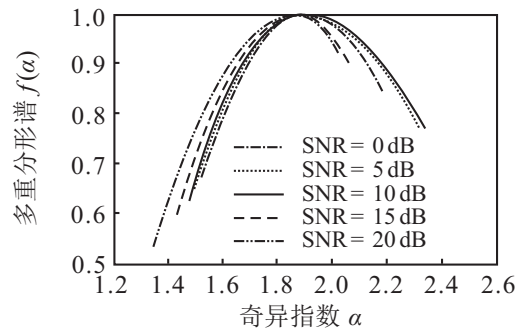


图3 去噪前信号多重分形谱(FMCW)

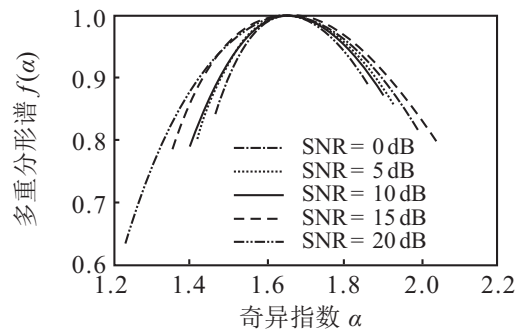


图4 去噪后信号多重分形谱(Costas)

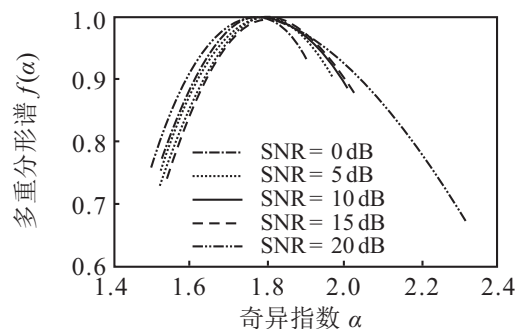


图5 去噪后信号多重分形谱(Frank)

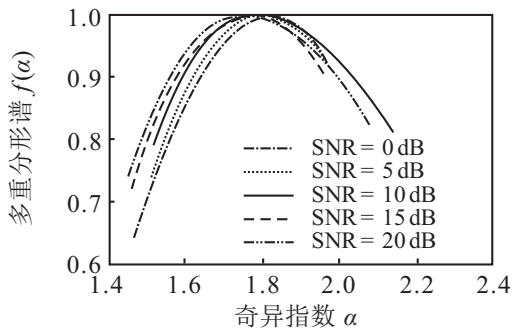


图6 去噪前信号多重分形谱(FSK/PSK)

从图3~图6可以看出,不同信噪比下多重分形谱曲线呈现出发散的状态,且曲线之间差异较大,说明噪声对谱参数的提取产生了一定影响,若不经处理,则提取到的谱参数特征受噪声影响过大,提取的特征不稳定,进而影响后续的认识过程.因此,本文在特征提取过程中,利用小波变换的去噪功能对信号进行Daubechies小波变换^[16],得到的多重分形谱如图7~图10所示.

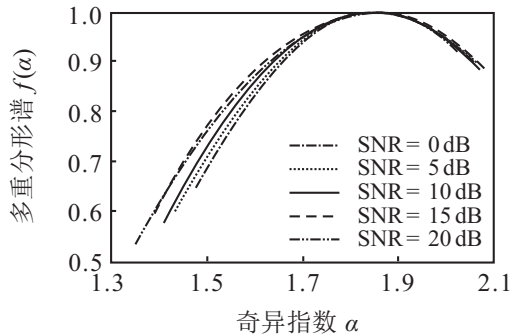


图7 去噪后信号多重分形谱(FMCW)

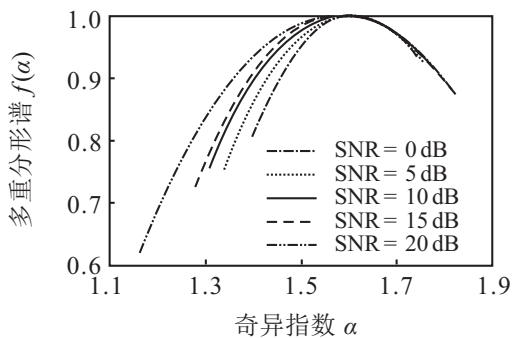


图8 去噪后信号多重分形谱(Costas)

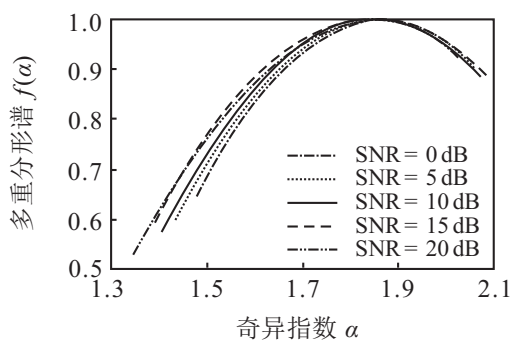


图9 去噪后信号多重分形谱(Frank)

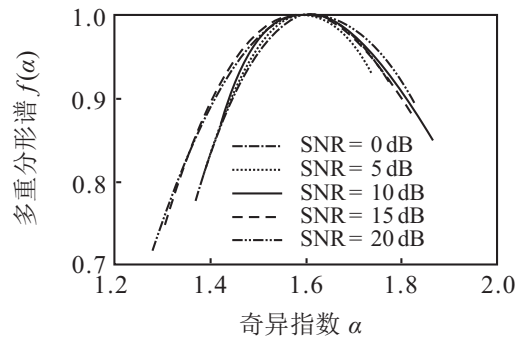


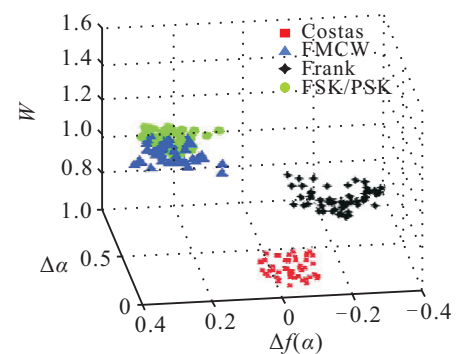
图10 去噪后信号多重分形谱(FSK/PSK)

将图3~图6与图7~图10对比可以看出,经过Daubechies小波变换去噪后,信号的多重分形谱曲线较去噪前更为收敛,此时提取的特征更为稳定,信号识别过程的噪声容限也有所提高.

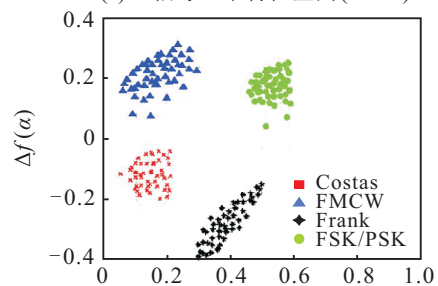
5.2 谱参数特征提取

以多重分形谱为研究对象,在上文分析的基础上,提取出4种信号的多重分形谱曲线 $f(\alpha(q))$ 的谱宽 $\Delta\alpha$ 、谱值差 Δf 和谱偏移度 W .为了更直观地体现提取的特征包含不同信号之间的差别信息,在信噪比分别为0 dB和10 dB的情况下,将提取到的每种信号的100个特征作如图11和图12所示的空间特征分布图.

从图11和图12中可以看出,在不同的信噪比条件下提取到的特征都能对4种信号进行有效区分,在信噪比为10 dB时,不同的信号之间区分度较明显,分布的差距较大且信号自身具有一定的聚集性.随着信噪比的降低,信号的区分度稍有降低,但在0 dB噪声条件下,各个信号自身仍具有一定的聚集性,说明



(a) 信号三维特征空间(10 dB)



(b) 信号特征空间x-y面投影(10 dB)

图11 10 dB下信号的空间特征分布图

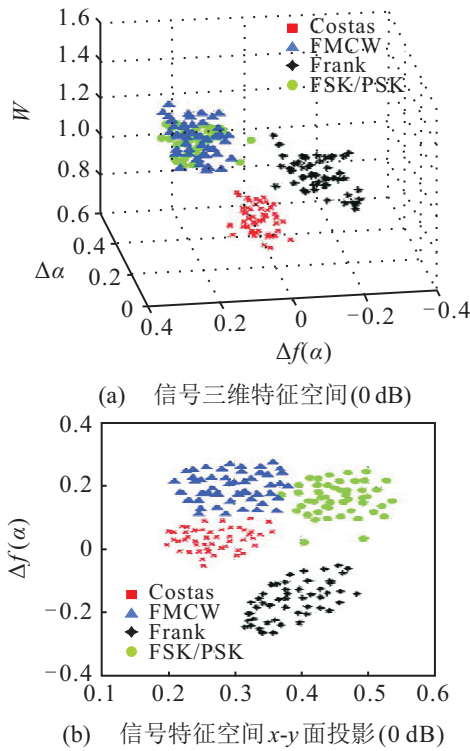


图 12 0 dB 下信号的空间特征分布

提取到的多重分形谱参数特征能够对不同LPI雷达信号进行区分,因此可作为后续半监督EM算法的输入特征,进而完成识别过程.

5.3 算法识别效果

1) 识别率分析.

对于每一种LPI雷达信号,在0~20 dB信噪比范围内,随机选取提取到的多重分形谱特征样本的5%、10%、15%、20%、25%作为已标识样本,剩下的作为未标记样本集.对比EM算法和Self-training的EM算法识别效果,对比结果如表2所示,其中基于Self-training的EM算法用SEM简化表示.

分析表2数据结果可以得出:在有标记样本比例分别占5%、10%、15%、20%、25%的情况下,对于每一种LPI雷达信号,基于Self-training的半监督EM算法识别率大部分都高于EM算法,说明Self-training的半监督EM具有更好的分类性能.但也有部分样本的识别率略低于EM算法,这是因为在有标记样本比例较少的情况下,产生的中间参数分类器的准确率较差,初期对测试信号的误判率较高,错误标记的样本在Self-training过程中进行训练学习导致错误被累积,从而影响到识别率.但当有标记样本比例增加时,中间参数分类器的准确性提高之后,分类性能会有很大的提升.两种算法在不同样本规模下的平均识别率如图13所示.

表 2 两种方法在不同标记样本下的识别效果比较

	5%		10%		15%		20%		25%	
	EM	SEM	EM	SEM	EM	SEM	EM	SEM	EM	SEM
FMCW	78.58	80.25	80.25	82.29	80.75	84.86	82.46	87.83	81.76	87.42
Costas	77.69	79.42	82.33	80.51	80.28	84.54	81.82	86.79	82.80	87.52
Frank	78.36	80.39	79.20	83.30	80.09	84.92	82.57	87.28	80.91	86.09
FSK/PSK	79.98	82.44	80.19	84.82	83.90	82.61	82.65	86.97	82.78	86.63
平均识别率/%	78.65	80.63	80.32	82.73	81.26	84.23	82.38	87.22	82.06	86.92

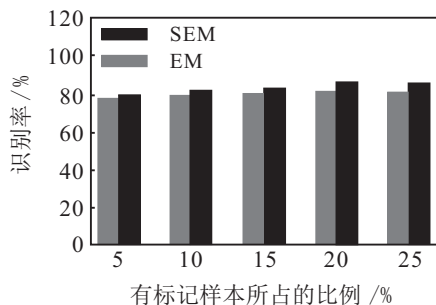


图 13 不同规模下平均识别率对比

从图13中可以看出,将Self-training思想引入到EM算法中,分类精度有了很大的改善.特别地,当有标记样本比例达到20%时,识别率达到最高,说明充足的训练样本构造的中间分类器的准确性最佳,此时分类效果最理想且所需的样本数据量也较少.基于Self-training的EM算法在更新样本集的过程中,选取

最为确定的一个或多个样本,将其从无标记样本中删除,利用更新的样本集进行训练学习构造分类器,这样每次无标记样本集会不断缩小从而加快训练速度,同时也可以避免错误的累积,提高分类精度和分类性能.

2) 分类器实时性分析.

为进一步对比两种算法构建的分类器的实时性,以FMCW信号为例,图14给出了在有标记样本比例为20%时,不同规模样本集下两种算法的迭代次数和运行时间,此时EM算法的收敛条件采用文献[17]给出的收敛条件.

实验表明:基于Self-training的EM算法的迭代次数始终少于EM算法,且随着样本特征集数量的增加,基于Self-training的EM算法相比EM算法需要更少

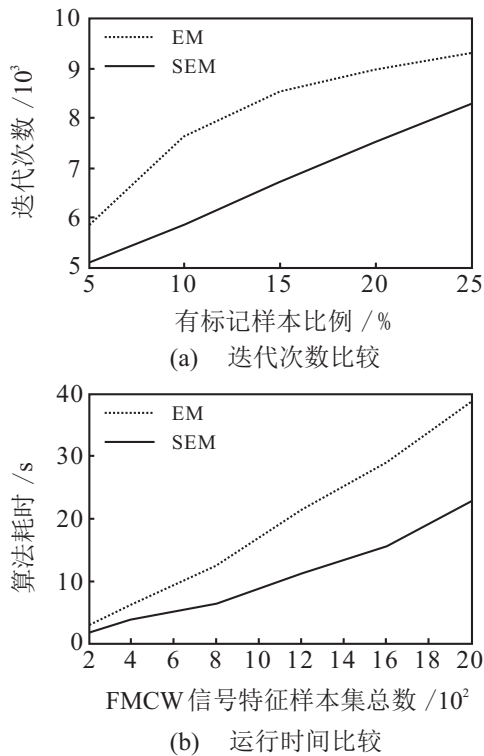


图14 两种算法实时性分析

的运行时间.这是由于基于Self-training的EM算法对未标记样本的准确性进行了预判,选取的都是最为准确的未标记样本,这样中间分类器的迭代次数会明显减少,更新速度也能有所提高,所以实时性更好.

3) 不同算法识别效果对比

为了验证本文算法的有效性,将本文提出的算法(本文算法用SEM表示)与EM算法、文献[18]提取的单分形维数(盒维数和信息维数)、文献[10]提出的有监督NB算法、无监督K-means算法进行对比.其中有监督NB算法和无监督K-means算法的输入特征都为本文的多重分形谱参数.对比结果如表3所示.

表3 不同算法识别率对比

信噪比/dB	0	5	10	15	20
SEM	80.3	82.1	84.0	87.2	88.9
EM	74.4	76.3	78.5	80.9	83.7
文献[18]	69.8	72.6	75.5	80.1	82.3
文献[10]	60.6	69.6	72.7	76.5	80.1
K-means	52.1	59.4	67.3	72.8	79.6

分析表3可知,在不同信噪比下EM算法识别率低于基于Self-training的EM算法,原因是EM算法在更新样本集时未能挑选出较为确定的未标记样本,一旦出现错误便会导致错误的累加,同时还会造成分类器更新速度缓慢;文献[18]通过提取单分形维数即盒维数和信息维数进行识别,单分形维数只从平均意义上对信号进行刻画,未能完整表述出信号中细微的分

形结构,故识别率低于本文提取到的多重分形特征进行的识别;文献[10]算法和K-means算法在分类器设计方面分别采用传统的有监督NB算法和无监督算法,有监督学习充分利用到有标记样本中的先验信息进行训练学习,但实际信号中有先验信息的有标记样本特别少,而无监督学习使用未标记样本训练学习构造分类器,未能结合到信号中已有的先验信息,故两种算法训练出的分类器都不够准确,识别率不高;本文算法结合了多重分形特征和半监督EM算法,提取的多重分形特征能从整体和局部刻画出信号的结构,同时在更新样本集过程中,通过选取最为确定的未标记样本避免错误样本的累加,所以能有效地对信号进行识别,得到更好的识别率.

6 结论

本文针对LPI雷达信号识别,在结合多重分形和基于Self-training的EM算法的基础上,提出一种识别方法.

1) 在特征提取过程中,分形特征可表征信号的复杂度和不规则度,本文提取多重分形谱参数作为特征向量,极大地减少了计算量且在低信噪比下也能取得很好的识别效果.

2) 在分类器设计部分,本文将半监督思想与EM算法结合起来并针对其不足进行改进,改进后的算法克服了传统算法的缺陷,有效解决了传统识别方法分类精度低、分类性能不稳定等缺点.

3) 本文提出的算法有效提高了分类精度,同时改善了分类性能,文章通过理论证明了该方法的可行性,并将改进的算法与经典算法进行对比,通过仿真实验验证了算法的有效性.

参考文献(References)

- [1] 李娜, 王珂, 李保珠. 低截获概率雷达信号检测方法的优化及应用[J]. 光学精密工程, 2014, 22(11): 3122-3128.
(Li N, Wang K, Li B Z. Optimization and application of LPI radar signal detection method[J]. Optics and Precision Engineering, 2014, 22(11): 3122-3128.)
- [2] Yong J L, Peng X, Hong C W, et al. LPI radar signal detection based on radial integration of Choi-Williams time-frequency image[J]. J of Systems Engineering and Electronics, 2015, 26(5): 973-981.
- [3] Headley W C, Da Silva C R. Asynchronous classification of digital amplitude-phase modulated signals in flat-fading channels[J]. IEEE Trans on Communications, 2011, 59(1): 7-12.
- [4] 杨宏飞, 和正日. 时频分析在雷达信号识别中的应

- 用[J]. 电子科技, 2016, 29(8): 58-61.
(Yang H F, He Z R. Time-frequency analysis in radar signal recognition[J]. Electronics Science and Technology, 2016, 29(8): 58-61.)
- [5] 王福友, 倪箐, 桂佑林. 基于分形特征的高频雷达目标检测技术[J]. 现代雷达, 2011, 33(12): 36-42.
(Wang F Y, Ni J, Gui Y L. High frequency radar targets detection based on fractal character[J]. Modern Radar, 2011, 33(12): 36-42.)
- [6] 陈红, 蔡晓霞, 徐云, 等. 基于多重分形特征的通信调制方式识别研究[J]. 电子与信息学报, 2016, 33(4): 863-869.
(Chen H, Cai X X, Xu Y, et al. Communication modulation recognition based on multi-fractal dimension characteristics[J]. J of Electronic & Information Technology, 2016, 33(4): 863-869.)
- [7] 李秋生, 谢维信. 基于多重分形特征的防空雷达目标分类方法[J]. 计算机应用研究, 2013, 30(2): 405-409.
(Li Q S, Xie W X. Target classification by surveillance radar based on multifractal features[J]. Application Research of Computers, 2013, 30(2): 405-409.)
- [8] 秦立龙, 王振宇. 边际谱和多重分形在调制模式识别中的应用[J]. 智能系统学报, 2014, 9(6): 756-762.
(Qin L L, Wang Z Y. Marginal spectrum and multifractal theory and its application in modulation recognition[J]. CAAI Trans on Intelligent Systems, 2014, 9(6): 756-762.)
- [9] Berve F, Zhao L, Quiles M, et al. Particle competition and cooperation in networks for semi-supervised learning[J]. IEEE Trans on Knowledge and Data Engineering, 2012, 24(9): 1686-1698.
- [10] 张博峰, 白冰, 苏金树. 基于自训练的EM算法的半监督文本分类[J]. 国防科技大学学报, 2007, 29(6): 65-69.
(Zhang B F, Bai B, Su J S. Semi-supervised text classification based of self-training EM algorithm[J]. J of National University of Defense Technology, 2007, 29(6): 65-69.)
- [11] 孙永奎, 余志斌, 金炜东. 基于多重分形与SVM的高速列车运行状态识别方法[J]. 西南交通大学学报, 2015, 50(1): 7-12.
(Sun Y K, Yu Z B, Jin W D. Recognizing Running state of high-speed trains based on multifractal theory and SVM[J]. J of Southwest Jiaotong University, 2015, 50(1): 7-12.)
- [12] 李兆飞, 柴毅, 李华峰. 多重分形的振动信号故障特征提取方法[J]. 数据采集与处理, 2013, 28(1): 34-40.
(Li Z F, Chai Y, Li H F. Fault feature extractin method of vibration signals based on multi-fractal[J]. J of Data Acquisition and Processing, 2013, 28(1): 34-40.)
- [13] 苏岭东, 翟明岳. 宽带电力通信信号多重分形特性研究[J]. 中国电机工程学报, 2014, 34(25): 4430-4436.
(Su L D, Zhai M Y. Research of multi-fractal characteristics of broadband power line commuication signal[J]. Processing of the CSEE, 2014, 34(25): 4430-4436.)
- [14] Zhou X C, Zhao Z X, Zhou K J, et al. A hierarchical imalge annotation method based on SVM and semi-supervised EM[J]. Acat Automatica Sinica, 2010, 36(7): 960-967.
- [15] Philip E Pace. Detecting and classifying low probability of intercept radar[M]. 2nd ed. Aretch House: Norwood, 2009: 51-129.
- [16] 吴建飞, 袁学华. 基于小波多重分形的通信信号调制识别[J]. 航天电子对抗, 2014, 30(3): 27-30.
(Wu J F, Yuan X H. Communication signal modulation recognition based on wavelet and multifractal[J]. Space Electronic Countermeasure, 2014, 30(3): 27-30.)
- [17] Nigam K. Using unlabeled data to improve text classification[D]. Lawton: Computer Science Department, Camegie Mellon Unversity, 2001: 30-49.
- [18] 刘文涛, 陈红, 蔡晓霞, 等. 基于分形维数的数字调制信号识别[J]. 火力与指挥控制, 2014, 39(10): 1766-1772.
(Liu W T, Chen H, Cai X X, et al. Didital modulation signal identification based on fractal dimension[J]. Fire Control and Command Control, 2014, 39(10): 1766-1772.)

(责任编辑: 闫妍)