不完整数据集的MFR辐射源识别方法研究

陈维高*，朱卫纲，唐晓婧，贾 鑫

(航天工程大学 北京 怀柔区 101416)

【摘要】该文提出一种基于随机森林的不完整数据集的多功能雷达(MFR)辐射源识别方法，该方法在MFR辐射源波形单元识别框架基础上，首先对参数缺失的先验知识集进行多重划分，得到多个不含缺失参数的样本子集，然后删减冗余子集并利用随机森林算法对各个子集构建弱分类器，最后根据弱分类器对识别结果贡献率的不同，进行权值设定，得到最终的识别模型。仿真实验证实了提出的MDRF-WA方法能够提高少量先验知识条件下波形单元识别的准确率和鲁棒性，降低计算成本。

关 键 词 不完整数据集； 多功能雷达； 多重划分； 随机森林； 波形单元

中图分类号 TN958.92 **文献标志码** A doi:10.3969/j.issn.1001-0548.2019.01.007

Research on MFR Emitter Identification Method for Incomplete Data

CHEN Wei-gao*, ZHU Wei-gang, TANG Xiao-jing, and JIA Xin

(Aerospace Engineering University Huairou Beijing 101416)

Abstract For the multi-function radar (MFR) emitter identification with incomplete data, aiming at the problems of prior knowledge demand, low accuracy and poor robustness, which exist in the conventional identification methods, a method of waveform unit identification based on incomplete prior knowledge set is proposed. Firstly, based on the MFR waveform unit identification framework, the original prior knowledge with parameter missing is multiply divided, and a number of subsets of samples without parameter missing are obtained. Secondly, the redundant subsets are removed and a weak classifier is constructed for each subset by using the random forest algorithm. Finally, the weight is set according to the contribution rate of each weak classifier to the identification result, and the final identification model is obtained. Simulation results confirm the validity of the MDRF-WA waveform unit identification method proposed, moreover, MDRF-WA method can make full use of known prior knowledge, reduce the computational cost and improve the robustness and accuracy of the waveform unit identification under the condition of small training samples.

Key words incomplete data; MFR; multiple division; random forest (RF); waveform unit

随着雷达技术的飞速发展，以相控阵天线为基础，以数字波束形成、自适应空间滤波、自适应空时处理、空间功率合成以及信号能量管理等技术为支撑[1]，具备搜索、截获、跟踪、距离探测以及制导等多种能力的MFR被广泛地部署和应用。与此同时，为了提供更充分可靠的情报信息，对MFR的识别[2]、状态跟踪及预测技术[3]成为了研究热点。由于MFR的相控阵体制，其波束形状快速捷变、信号调制特征复杂多变，在常规基于统计参数识别模型下，k近邻算法、专家系统、SVM[4]、神经网络[5]等识别方法的研究[6]集中在解决型号的识别问题，不能满足反映、预测MFR功能状态的需求[7]。另外由于干扰、测量误差以及情报渠道等因素的影响，已知辐射源样本数据库中的特征参数存在残缺现象，如何充分地利用这些非完备先验知识对辐射源进行有效识别，是雷达情报分析所面临的一大现实问题。

针对MFR辐射源的识别问题，波形单元作为MFR信号的基本构成，其识别结果不仅可以作为辐射源识别结果，而且识别优劣决定了后续对MFR状态跟踪、预测的准确度，对MFR的威胁和态势感知具有重要意义[8]。针对不完整数据集的识别问题，当前流行的方法主要分为删除法、插值法以及非完备系统处理法3种[9]。其中删除法在数据库样本充足且缺失数据较少时是有效的，一旦雷达数据库中样本数量较少，盲目的删减必然导致先验知识不充分而大大降低识别准确率。插值法中的均值替代、随机替代的思想必然影响样本分布，识别效果不佳。而回归插值方法存在复杂度高、计算成本巨大的问

收稿日期：2017－08－26；修回日期：2018－08－25

作者简介：陈维高(1990－)，男，博士，主要从事认知电子战、雷达信号处理等方面的研究. E-mail: tingxue9278@163.com

题。非完备系统处理法无需对缺失数据进行插补或删除，直接对不完备数据集进行处理，其中基于神经网络集成的非完备系统处理方法通过多个弱分类器构建强分类器，达到高识别准确率的目的，但算法存在着训练样本需求量高、计算量大的问题，并不适用于对时效性要求高的MFR辐射源识别。

综上所述，针对常规处理方法应用于不完整数据集的MFR辐射源识别中存在的识别准确率低、先验知识需求量大、计算成本高的问题，本文提出一种权值修正的基于样本多重划分和随机森林集成的MFR辐射源识别方法。该方法在前期研究的MFR波形单元识别框架基础上[8]，首先对参数缺失的先验知识集进行多重划分，得到多个全参数子集；然后依据各个子集对识别结果贡献率的不同，删减冗余子集并进行权值修正；最后，利用随机森林算法对各个全参数子集构建弱分类器，得到最终识别模型。

1 样本库多重划分

波形单元识别模型构建之后，由于波形单元训练样本库中参数存在缺失，不能直接利用传统识别方法进行处理，因此引入文献[10]中非完备系统处理方法的思路，对先验知识进行多重划分。

1.1 不完整先验知识集分析

在MFR波形单元识别的背景下，不完整先验知识特指由于各种干扰、误差、情报渠道等因素导致已知波形库中存在的样本特征参数值缺失现象。对于已知波形库中的MFR内置波形单元训练集 $M=\{(x_1,y_1,z_1),(x_2,y_2,z_2),\cdots,(x_l,y_l,z_l)\}$，其中 $x_i\in R^n$，$y_i\in\{1,2,\cdots,k\}$，$z_i\in\{1,2,\cdots,m\}$ 分别代表第 i 组训练数据的内嵌脉冲列参数集、内嵌脉冲列标签和波形单元标签，l、n、k、m 分别指代训练样本总数、特征维度、内嵌脉冲列总数以及波形单元总数。样本 $x_i\in R^n$ 具备 n 维特征 $x_i=\{\alpha_1^i,\alpha_2^i,\cdots,\alpha_n^i\}$。令 Φ 代表缺失的特征参数，$\alpha_u^i=\Phi$、$\alpha_v^i=\Phi$ $(u,v\in[1,n])$，则样本表示为 $x_i=\{\alpha_1^i,\alpha_2^i,\cdots,\alpha_{u-1}^i,\Phi,\alpha_{u+1}^i,\cdots,\alpha_{v-1}^i,\Phi,\alpha_{v+1}^i,\cdots,\alpha_n^i\}$。

1.2 样本多重划分方法

为了充分发挥先验知识中未缺失参数对识别模型的贡献，同时避免删除法、插补法造成的参数分布偏差问题，将参数缺失的样本库进行多重划分，得到多个不含缺失参数的数据子集。方法如下：

1) 已知样本参数集 $X=\{x_1,x_2,\cdots,x_l\}$，任意样本 $x_i=\{\alpha_1^i,\alpha_2^i,\cdots,\alpha_n^i\}$ $(i=1,2,\cdots,l)$，具备 n 维特征 $D=\{\alpha_1,\alpha_2,\cdots,\alpha_n\}$。

2) 依据缺失参数的属性对每个样本进行标记，若第 i 个样本的参数 $\alpha_u^i=\Phi$、$\alpha_v^i=\Phi$ $(u,v\in[1,n])$，则该样本的完全特征集和缺失特征集分别为 $D_i'=\{\alpha_1,\alpha_2,\cdots,\alpha_{u-1},\alpha_{u+1},\cdots,\alpha_{v-1},\alpha_{v+1},\cdots,\alpha_n\}$、$\bar{D}_i'=\{\alpha_u,\alpha_v\}$。遍历整个样本集 X，得到整个样本的完全特征集 $D'=\{D_1',D_2',\cdots,D_l'\}$ 和缺失特征集 $\bar{D}'=\{\bar{D}_1',\bar{D}_2',\cdots,\bar{D}_l'\}$，其中 $D'=D-\bar{D}'$。遍历整个缺失特征集 $\bar{D}'$，删除其中重复出现的子集，得到 $\hat{D}'=\text{unique}(\bar{D}')=\{\hat{D}_1',\hat{D}_2',\cdots,\hat{D}_s'\}$ $(1\leqslant s<l)$。

3) 设 $V=\{V_1,V_2,\cdots,V_s\}$ 代表多重划分后生成的特征集，则对于任意子集 V_a $(1\leqslant a\leqslant s)$ 有：

$$V_a=\{\alpha_e,\cdots,\alpha_f|D-\hat{D}_a'\}\quad e,f\in\{1,2,\cdots,n\}\tag{1}$$

式中，n 指样本集的特征总个数。

4) 在多重划分后的特征集 $V=\{V_1,V_2,\cdots,V_s\}$ 中，每个特征子集确定了多重划分后样本子集的特征。利用原始样本集 X 对 V 的各个子集进行投影，删除参数值缺失的样本，即可得多重划分后的样本集，则对于任意子集 V_a 有：

$$x_i^{v_a}=\{\alpha_e^i\,|\,\alpha_e\in V_a\ \&\ \alpha_e^i\neq\Phi\}\tag{2}$$

$$X_{v_a}=\{x_i^{v_a}\,|\,x_i^{v_a}\neq\Phi\}\tag{3}$$

式中，α_e、α_e^i 分别指原始样本集 X 的第 e 个特征和第 i 个样本第 e 个特征的参数值。根据投影的特征子集的不同，重复运算式(2)、式(3)，即可得到多重划分后的样本集 $X_v=\{X_{v_1},X_{v_2},\cdots,X_{v_s}\}$。

2 基于随机森林的识别方法

通过多重划分方法得到不含参数缺失的样本子集 $X_{v_1},X_{v_2},\cdots,X_{v_s}$ 后，即可借助集成学习思想，利用常规识别方法构建多个弱分类器，进而投票得到识别结果。然而通过分析发现，该思路存在两个问题：1) 各个样本子集间存在样本、特征重复利用的现象，倘若两个样本子集相似度较高，则构建的弱分类器性能相近，存在冗余；2) 由于各个样本子集的特征个数、样本个数、特征重要性以及训练准确率都不相同，对识别结果的贡献有大小之分。针对上述两方面的问题，为了降低计算成本，提高识别准确率，需要依据各个子集的相似程度剔除冗余，进行样本子集的约简，并对约简后各个子集构成的弱分类器设定权值。

2.1 样本子集约简

通过分析多重划分后各个样本子集的特点，从子集重复利用的样本和特征的角度出发，找寻高相

似度的样本子集，提出约简方法。

1) 已知多重划分后的样本集 $X_v=\{X_{v_1},X_{v_2},\cdots,X_{v_s}\}$，及其相应的特征集 $V=\{V_1,V_2,\cdots,V_s\}$。

2) 设 $o=1,2,\cdots,s$，$p=1,2,\cdots,s$，且 $o\neq p$，如果样本子集 X_o 和 X_p 满足条件：$V_p\subseteq V_o$，$f_{\mathrm{FN}}(V_o)-f_{\mathrm{FN}}(V_p)=1$，$\dfrac{f_{\mathrm{SN}}(X_o\bigcap X_p)}{f_{\mathrm{SN}}(X_p)}\geqslant \mathrm{srr}$。则判定子集 X_o 与 X_p 具备高的相似度，且 X_o 拥有较好的样本表征能力，删除子集 X_p。其中，$f_{\mathrm{FN}}(V)$ 指计算集合V所包含的特征个数，$f_{\mathrm{SN}}(X)$ 指计算集合X所包含的样本个数，$X_o\bigcap X_p$ 指 X_o 与 X_p 重复利用样本的集合，$\mathrm{srr}\in(0,1)$ 指样本重复利用率，一般取值较大。

3) 按照步骤2)，遍历整个子集 X_v，即可得到约简后的样本集 $X'_{v'}$、特征集 V'，其中 $X'_{v'}\subseteq X_v$，$V'\subseteq V$。

2.2　分类器构建

对于约简后的样本子集，需要针对各个子集构建弱分类器，并集成各个弱分类器的结果得到强分类器。相对于其他识别算法，基于决策树组合的随机森林(random forests, RF)算法，通过对样本和特征的随机选择，较好地克服了决策树过学习、泛化能力差的缺点，具备结构简单、运算速度快和鲁棒性高等优势[11]。利用RF构建分类器的基本原理如下：

1) 设约简后的第 b ($b=1,2,\cdots,c$，c 为约简后子集总个数)个样本参数子集为 $X'_{v'_b}$，与之对应的标签集为 $Y'_{v'_b}$、特征集为 V'_b，则输入样本可表示为 $(X'_{v'_b},Y'_{v'_b})=(x'_{b1},y'_{b1}),(x'_{b2},y'_{b2}),\cdots,(x'_{bN},y'_{bN})$。

2) 令 η 为组成RF的决策树个数，对特征集 V'_b 进行 η 次随机取样，每次取样的特征数为 $\mathrm{sqrt}(|V'_b|)$ ($|V'_b|$ 指集合 V'_b 的特征个数)，得到特征集 V'_b 的采样子集 $V''_{b1},V''_{b2},\cdots,V''_{b\eta}$。相似地，对输入的样本集 $(X'_{v'_b},Y'_{v'_b})$ 进行 η 次随机的有放回取样，每次取样个数与样本总数一致，得到输入样本子集 $(X''_{v''_{b1}},Y''_{v''_{b1}}),(X''_{v''_{b2}},Y''_{v''_{b2}}),\cdots,(X''_{v''_{b\eta}},Y''_{v''_{b\eta}})$。

3) 利用采样后的各个样本子集 $(X''_{v''_{bd}},Y''_{v''_{bd}})$ ($d=1,2,\cdots,\eta$)，训练与之对应的RF子决策树分类器 $\mathrm{C}_d(X''_{v''_{bd}})$。训练过程中，选用CART算法作为RF的子决策树，利用Gini系数和特征子集 V''_{bd} 对树节点进行分裂。

4) 各个子决策树经过训练后，对于某个输入数据 x，通过投票法得到第 b 个RF弱分类器的类别判定结果：$y^*_b=\arg\max\limits_{\tilde{y}}\sum\limits_{d=1}^{\eta}I(\mathrm{C}_d(x)=\tilde{y})$。其中，$I(\bullet)=\{0,1\}$ 为示性函数，$\tilde{y}=1,2,\cdots,\max(y)$ 为样本的类别标签。

通过2.1小节的样本约简，得到了 c 个样本子集，每个子集训练一个RF弱分类器，由于各个子集缺失的特征和样本不同，对分类器效果的影响也有大小之分，因此利用各子集的重要程度 w 作为权重来组合 c 个RF弱分类器，即可得到最终识别结果：$y^*=\arg\max\limits_{\tilde{y}}\sum\limits_{b=1}^{c}w_bI(y^*_b=\tilde{y})$。

2.3　权重设定

通过对约简后不同样本子集的分析发现，影响最终识别结果的主要因素有：样本子集的样本个数 $N_{x'}$、样本子集所含的特征个数 $N_{v'}$、RF弱分类器的训练准确率 Acc、样本子集的信息增益 $\mathrm{MI}_{x'}$。

在权重的选取上，如果单独利用 $N_{x'}$ 或 $N_{v'}$ 作为权重，仅能够片面地侧重样本个数或特征个数的重要性，并不能完整反映样本子集的重要程度。如果利用 Acc 作为权重，由于其反映的仅是分类器对训练样本的识别效果，当测试样本与训练样本的差异性较大时，Acc 的参考意义将显著下降。如果将 $\mathrm{MI}_{x'}$ 作为权重，一定意义上的确能够反映样本子集的重要程度，然而信息增益的计算成本远远超过了其他因素，得不偿失。综上所述，为了全面反映样本子集的贡献，同时兼顾计算成本，将参与训练分类器的数据项占总数据项的比例作为权重，即 $w_b=\dfrac{N_{x'_b}\times N_{v'_b}}{ln}$，其中 l 和 n 分别代表原始样本集的样本总数和特征总数。

3　面对缺失数据MFR辐射源识别流程

面对缺失数据的MFR波形单元识别方法流程如图1所示，主要分为样本初始化、多重划分、分类器构建3大部分。

1) 样本初始化。参照文献[8]波形单元识别模型，利用波形单元对MFR训练波形库和未知波形库进行描述，生成原始训练、测试样本集。

2) 多重划分。为剔除缺失参数，同时充分利用未缺失参数，参考前面对训练样本库进行多重划分。

3) 分类器构建。在识别模型构建过程中，为了降低计算成本，首先需要对多重划分后冗余的训练样本进行约简，得到约简后的样本子集 $(X'_{v'_b},Y'_{v'_b})$ ($b=1,2,\cdots,c$)；然后针对每个训练样本子集

$(X'_{V'_b}, Y'_{V'_b})$，构建RF弱分类器 RF_b，计算各个 RF_b 的权值，并通过组合各 RF_b 得到最终的识别模型。在对未知波形单元识别过程中，无需多重划分和样本约简，仅需将未知样本对训练样本约简后的特征集 V'_b 投影，得到多个未知样本子集，进而通过已优化的识别模型即可得到最终识别结果。

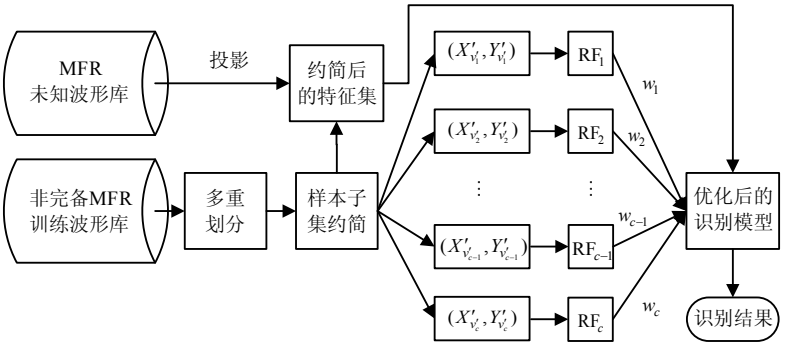

图1 波形单元识别流程

4 实 验

为构建非完备先验知识的实验环境，模拟生成MFR波形单元样本集，包含*A*、*B*、*C*、*D*、*E*这5部MFR信号，各部MFR拥有互不相同的波形单元和特征参数各异的内嵌脉冲列，共计15个波形单元、26个内嵌脉冲列。为模拟真实信号，各内嵌脉冲列具备脉内调制*F*，计CW、LFM、NLFM、BPSK、QPSK以及LFM-BPSK这6种脉内调制类型。

由于MFR具备复杂多变的信号调制样式，仅靠常规的PRI、CF和PW脉间特征非但不能全面地描述信号，甚至还存在交叠问题，因此增加相像系数Cr和小波包特征Wpt2来描述MFR的脉内调制特征[12]，其中Cr指以三角形序列为基准提取的与各信号频谱的相像系数；Wpt2指利用小波包对信号进行3层分解后，提取的第2个频带的特征信息。则样本的特征集为{PRI, CF, PW, Cr, Wpt2}。MFR波形单元样本集的特征信息如表1所示(所涉及MFR的调制类型及参数在文献[13]的基础上仿真生成)。

表1 MFR波形单元信息表

MFR	波形单元	内嵌脉冲列	PRI/μs		CF/MHz		PW/μs		*F*	Cr	Wpt2	脉冲个数
			类型	取值	类型	取值	类型	取值				
A		ep_{11}^{A}	5	830/880/930/980/1 030	2	[3 890,3 943]	3	9.6±1.5	2	[0.881 1,0.910 7]	[0.089 7,0.179 6]	21
	ω_1^A	ep_{12}^{A}	4	880±5%	1	3 925	3	9.6±1.5	1	[0.031 6,0.145 2]	[0.226 8,0.590 3]	16
		ep_{13}^{A}	1	990	1	3 880	1	8.6	3	[0.694 1,0.694 3]	[0.352 0,0.352 5]	8
	ω_2^A	ep_{21}^{A}	1	1 772	4	1 650/1 665/1 680	1	20.5	4	[0.182 3,0.221 0]	[0.377 0,0.413 2]	23
	ω_3^A	ep_{31}^{A}	1	1 547	1	3 868	1	12.6	3	[0.678 8,0.679 0]	[0.181 6,0.181 9]	11
		ep_{32}^{A}	1	1 765	1	2 961	1	12.6	1	[0.120 8,0.121 1]	[0.409 0,0.409 3]	8
B	ω_1^B	ep_{11}^{B}	4	2 300±5%	2	[3 362,3 448]	1	101.2	5	[0.380 0,0.404 1]	[0.377 3,0.441 7]	23
	ω_2^B	ep_{21}^{B}	2	337/373/409/428	1	5 420	1	2.5	1	[0.054 7,0.055 3]	[0.213 8,0.214 4]	17
		ep_{22}^{B}	1	474	1	5 420	1	2.2	4	[0.405 1,0.406 1]	[0.393 5,0.394 4]	9
	ω_3^B	ep_{31}^{B}	4	1 300±5%	2	[3 611,3 658]	3	15.6±1.5	3	[0.662 6,0.673 1]	[0.121 1,0.175 2]	20
		ep_{32}^{B}	1	1 342	1	3 628	1	16.1	5	[0.264 1,0.264 4]	[0.407 4,0.407 7]	10
C	ω_1^C	ep_{11}^{C}	2	651/678/705	1	4 820	1	3.5	6	[0.813 1,0.813 5]	[0.131 7,0.132 3]	21
	ω_2^C	ep_{21}^{C}	4	860±5%	1	5 430	1	2.3	2	[0.878 0,0.878 3]	[0.047 0,0.047 9]	13
		ep_{22}^{C}	1	790	4	5 456/5 478/5 499	1	2.3	1	[0.162 2,0.229 4]	[0.381 9,0.547 4]	9
	ω_3^C	ep_{31}^{C}	1	2 876	3	3 452/3 508	1	10.9	4	[0.378 5,0.393 3]	[0.408 4,0.424 3]	11
		ep_{32}^{C}	1	2 942	3	3 452/3 508	1	10.9	3	[0.666 4,0.667 6]	[0.129 7,0.157 9]	11
	ω_4^C	ep_{41}^{C}	1	2 276	1	2 852	1	20.3	6	[0.805 5,0.805 7]	[0.284 4,0.284 7]	5

（续表）

MFR	波形单元	内嵌脉冲列	PRI/μs		CF/MHz		PW/μs		F	Cr	Wpt2	脉冲个数
			类型	取值	类型	取值	类型	取值				
D	ω_1^D	ep_{11}^D	3	[2 762,2 782,2 842]	1	2 457	3	20.1±1.5	2	[0.871 1,0.873 2]	[0.018 5,0.031 3]	13
		ep_{12}^D	3	[2 135,2 155,2 185]	1	2 216	3	20.1±1.5	2	[0.869 5,0.870 8]	[0.018 0,0.024 2]	13
	ω_2^D	ep_{21}^D	1	2 136	3	1 341/1 441/1 541	2	130.7/86.4/42.1	4	[0.077 0,0.126 3]	[0.385 2,0.419 2]	19
		ep_{22}^D	1	2 432	1	2 642	1	15.5	1	[0.025 7,0.025 9]	[0.311 4,0.311 5]	23
E	ω_1^E	ep_{11}^E	1	2 838	3	1 628/1 649/1 681	2	100.2/57.6/35.7	5	[0.122 8,0.185 4]	[0.395 6,0.413 2]	19
	ω_2^E	ep_{21}^E	3	[1 630,1 660,1 720]	1	1 337	3	52.1±1.5	3	[0.658 7,0.698 1]	[0.107 9,0.216 8]	10
		ep_{22}^E	3	[1 020,1 060,1 100]	1	1 343	3	52.1±1.5	3	[0.659 3,0.682 3]	[0.085 1,0.180 6]	10
		ep_{23}^E	3	[1 270,1 330,1 390]	1	1 392	3	52.1±1.5	3	[0.661 4,0.679 8]	[0.099 1,0.186 6]	10
	ω_3^E	ep_{31}^E	5	725/755/785/815	4	4 381/4 426	1	3.3	6	[0.777 6,0.809 8]	[0.146 6,0.157 4]	9

表中，ω^A 表示第A部MFR弱分类器，ep^A 表示第A部MFR波形单元的内嵌脉冲序列。利用数字来代表PRI、CF、PW以及脉内调制F的类型，PRI：1-固定，2-参差，3-组变，4-抖动，5-滑变；CF：1-固定，2-捷变，3-组变，4-跳变；PW：1-固定，2-多脉宽组合，3-抖动；F：1-CW，2-LFM，3-NLFM，4-BPSK，5-QPSK，6-LFM-BPSK。此外，对于表中PRI、CF调制类型为组变、PW调制类型为多脉宽组合的内嵌脉冲列：ep_{31}^C、ep_{32}^C 的CF参数值每5个脉冲变化一次；ep_{21}^D、ep_{11}^E 的RF、PW为联合变化调制，每6个脉冲共同变化一次；、ep_2^E 的PRI每3个脉冲变一次。

在表1所示的MFR波形单元参数信息的基础上，构建训练、测试样本集。对于训练样本集，为满足非完备先验知识的条件，需要对参数进行随机丢失，令$\lambda\in[5,50]\%$为参数丢失率，则有：

$$\lambda=\frac{\sum_{i=1}^{l}\sum_{j=1}^{n}I(\alpha_j^i=\Phi)}{ln}\times100\% \tag{4}$$

对于测试样本集，为满足验证算法鲁棒性的需求，需加入一定的噪声，设信噪比(signal-to-noise ration, SNR)$\in[2,22]$ dB，EDL为PRI、CF、PW的参数偏离误差，EDL随SNR的增大逐渐减小，且当SNR=22 dB时，EDL=0 %(训练样本的条件)；当SNR=2 dB时，EDL=50 %。其中EDL=$\left|\xi_j^i/\alpha_j^i\right|\times100\%$，$\xi_j^i$ 为误差值，α_j^i 为参数真实值。

在表1参数信息的基础上构建两部分数据：1) 将每个内嵌脉冲列扩充为100个脉冲样本，共计2 600个，对这些样本随机排序，为训练样本提供数据，记为数据集A；2) 相似地，对每个内嵌脉冲列扩充为100个脉冲样本，共计2 600个，随机排序后为测试样本提供数据，记为数据集B。为验证提出的MDRF-WA方法在先验知识部分参数丢失条件下的识别有效性，从训练样本个数、训练样本参数丢失率以及测试样本信噪比3个方面进行实验。

4.1　训练样本数目对识别性能的影响

为分析训练样本数目对MDRF-WA识别效果的影响，验证在少量训练样本下识别的有效性，分别利用BP(back propagation)、CART_del、CART_mod、bagging_del、bagging_mod这5种经典方法对相同的训练、测试数据进行识别运算，对比分析实验结果。其中BP方法特指在训练样本多重划分的基础上，利用BP算法代替RF构建弱分类器，然后通过投票法得到识别结果(不包含训练样本约简、权值设定过程)，设置BP网络包含1个隐含层，隐含层节点数为10，优化算法为Adam，学习速率为0.5。CART_del和CART_mod是在CART决策树算法基础上，结合删除法(剔除参数丢失的样本)、众数插值法(利用丢失参数所属特征的众数替代丢失的参数值)来处理非完备先验知识识别问题，CART(classification and regression trees)以基尼不纯度(gini impurity)为节点分裂指标，分裂门限为10^{-7}，不设树的深度，节点分裂的最小样本个数为2。bagging_del和bagging_ mod是在bagging算法基础上，分别结合删除法和众数插值法构成，设置bagging由10个CART弱分类器构成，最终结果由弱分类器投票决定。

设置参数λ=20%、SNR=20 dB、srr=0.9，选取加入噪声的数据集B作为测试样本集。依次从数据集A中选取100,150,…,1 000样本，分别进行参数随机丢失处理并作为训练样本集，得到19个样本数目各异的训练样本集。利用上述6种方法分别对这些训练、测试样本进行运算，得到识别结果。100次Monte Carlo实验后，统计得到各方法识别结果如图2所示。

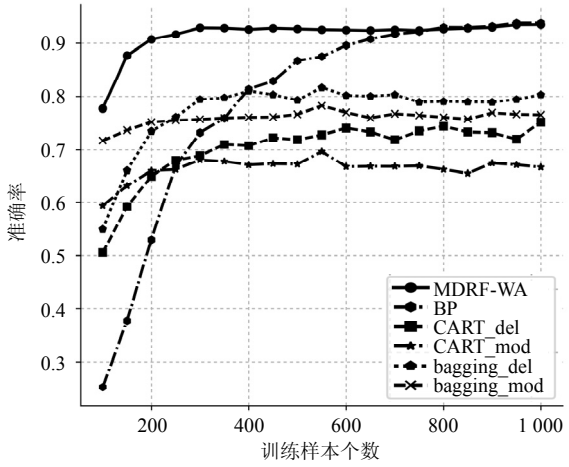

图2 不同训练样本下识别结果对比

从图2可以看出，随着训练样本数目的增加，6种方法的识别准确率都有所增长，并逐渐趋于平稳，其中基于删减和插值的4种方法识别准确率较低，一般小于80%。当训练样本数目从200增加到800时，MDRF-WA始终具备最高的识别准确率，而BP算法的识别准确率随样本数目迅速提升；当样本数目从800增加到1 000时，BP算法的识别准确率与MDRF-WA持平或具备微弱优势。然而在实验中通过对比运算时间发现，BP算法随训练样本的增加，计算成本近乎直线增加，而MDRF-WA的计算成本受训练样本数目的影响较小。因而综合识别准确率和计算成本两方面，MDRF-WA方法较其他算法具有一定优势，适合解决少量先验知识的识别问题。

4.2 训练样本参数丢失率对识别性能的影响

为分析先验知识参数丢失率对MDRF-WA识别效果的影响，令SNR=20 dB、srr=0.9，选取加入噪声的数据集 B 作为测试样本集。在数据集 A 中选取200个样本，参数随机丢失处理后作为训练样本集，令参数丢失率依次增加 λ=5%,10%,…,50%，共得到10个不同参数丢失率的训练样本集。利用上述6种方法分别对这些训练、测试样本进行运算，得到识别结果，各算法参数设置与前面相同。100次Monte Carlo实验后，统计得到平均识别准确率如图3所示。

从图3可以看出，随着先验知识参数丢失率的增加，所有方法的识别准确率都逐渐下降，其中BP的识别准确率受丢失率的影响最大；删除法CART_del、bagging_del，插值法CART_mod、bagging_mod受丢失率的影响较小；MDRF-WA受丢失率的影响最小，即使在丢失率为50%时，依旧能够取得0.844的识别准确率。因此MDRF-WA对先验知识丢失率的敏感度较低，且具备较高的识别准确率，进一步证实了提出方法的有效性。

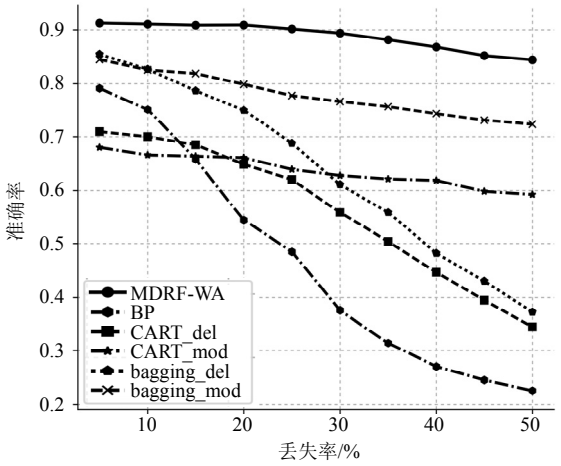

图3 不同参数丢失率下识别结果对比

4.3 测试样本信噪比对识别性能的影响

为分析MDRF-WA对测试样本信噪比(参数误差)的鲁棒性，令 λ=20%、srr=0.9，在数据集 A 中选取200个样本，参数随机丢失处理后作为训练样本集。选取加入噪声的数据集 B 作为测试样本集，令信噪比依次为SNR=2,4,…,22，得到11个不同信噪比的测试样本集。分别利用上述6种方法对这些训练、测试样本进行运算，得到识别结果。100次Monte Carlo实验后，统计得到平均识别准确率如图4所示。

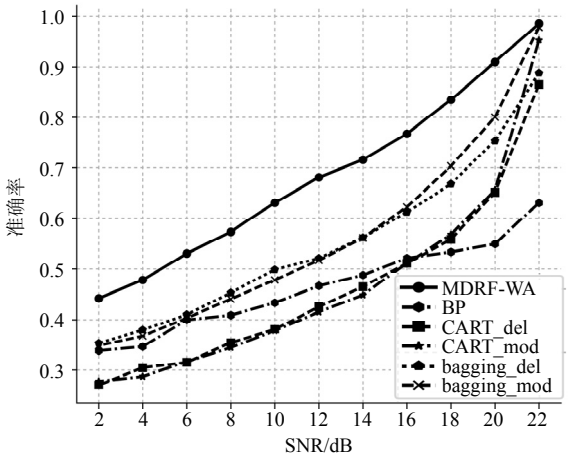

图4 不同信噪比下识别结果对比

从图4可以看出，随着测试样本信噪比的增加，所有方法的识别准确率都逐渐提高。其中在相同测试样本信噪比的条件下，MDRF-WA方法的识别准确率始终优于其他算法；BP方法先验知识需求量大，因此即使信噪比很大时，依旧不能取得较高的识别准确率；基于删减或插值的4种方法，在信噪比SNR=22能够取得较好的识别准确率，但随着误差的增大，信噪比降低，识别准确率迅速下降。由此可见，相较其他方法，MDRF-WA受信噪比的影响

较小，且识别准确率具有明显优势，证实了该方法的良好鲁棒性。

5 结 束 语

本文通过对非完备先验知识集的多重划分，将原始训练样本转化为多个不含缺失参数的样本子集，使得分类器能够充分利用未缺失的样本参数。通过样本子集约简和权值设定，一方面剔除了冗余样本，降低了计算成本，另一方面提高了某些重要弱分类器的权重，间接提高了识别准确率。

本文通过引入RF算法对约简后的样本子集构建集成分类器，能够在少量先验知识、大参数丢失率的条件下，取得优异的识别效果，具备识别准确率高，鲁棒性、泛化能力强的优势。

参 考 文 献

[1] 张光义. 相控阵雷达原理[M]. 北京: 国防工业出版社, 2009.
ZHANG Guang-yi. Principles of phased array radar[M]. Beijing: National Defense Industry Press, 2009.

[2] WANG A, KRISHNAMURTHY V. Signal interpretation of multifunction radars: Modeling and statistical signal processing with stochastic context free Grammar[J]. IEEE Transactions on Signal Processing, 2008, 56(3): 1106-1119.

[3] ARASARATNAM I, HAYKIN S, KIRUBARAJAN T, et al. Tracking the mode of operation of multi-function radars[C]// IEEE Conference on Rader. [S.l.]: IEEE, 2006: 10.1109/radar.2006.1631804.

[4] PING Y, CHANG Y F, ZHOU Y, et al. Fast and scalable support vector clustering for large-scale data analysis[J]. Knowledge and Information Systems, 2015,43(2):281-310.

[5] JIA W, ZHAO D, SHEN T, et al. An optimized classification algorithm by BP neural network based on PLS and HCA[J]. Applied Intelligence, 2015, 43(1): 1-16.

[6] 陈昌孝, 何明浩, 徐璟, 等. 雷达辐射源识别技术研究进展[J]. 空军预警学院学报, 2014(1): 1-5.
CHEN Chang-xiao, HE Ming-hao, XU Jing, et al. Progress of study on recognition technology of radar emitter[J]. Journal of Air Force Early Warning Academy, 2014(1): 1-5.

[7] 陈维高, 贾鑫, 朱卫纲, 等. 基于机动特征辅助的MFR状态预测方法[J]. 电子学报, 2018, 46(6): 1404-1409.
CHEN Wei-gao, JIA Xin, ZHU Wei-gang, et al. MFR state prediction method based on aircraft maneuvering features assistance[J]. Chinese Journal of Electronics, 2018, 46(6): 1404-1409.

[8] 陈维高, 贾鑫, 朱卫纲, 等. 基于HMM的雷达状态转移估计方法[J]. 北京航空航天大学学报, 2017, 43(10): 2171-2180.
CHEN Wei-gao, JIA Xin, ZHU Wei-gang, et al. The radar state transfer estimation algorithm based on HMM model[J]. Journal of Beijing University of Aeronautics and Astronautics, 2017, 43(10): 2171-2180.

[9] 刘海军. 雷达辐射源识别关键技术研究[D]. 长沙: 国防科学技术大学, 2010.
LIU Hai-jun. Researches on identification key technology for radar emitter[D]. Changsha: National University of Defense Technology, 2010.

[10] 严远亭. 不完整数据集的多视角集成分类研究[D]. 合肥: 安徽大学, 2016.
YAN Yuan-ting. Research on multi-view ensemble classification for incomplete data[D]. Hefei: AnHui University, 2016.

[11] 胡峻峰. 基于机器视觉的实木地板分选技术研究[D]. 哈尔滨: 东北林业大学, 2015.
HU Jun-feng. Research on technology of Solid wood floor sorting based on machine vision[D]. Harbin: Northeast Forestry University, 2015.

[12] 张葛祥. 雷达辐射源信号智能识别方法研究[D]. 成都: 西南交通大学, 2005.
ZHANG Ge-xiang. Intelligent recognition methods for radar emitter signals[D]. Chengdu: Southwest Jiaotong University, 2005.

[13] 马爽. 多功能雷达电子情报信号处理关键技术研究[D]. 长沙: 国防科学技术大学, 2013.
MA Shuang. Research on ELINT signal processing key technologies for multifunction radar[D]. Changsha: National University of Defense Technology, 2013.

编　辑　叶　芳