

基于EMD和SVD特征提取的通信 辐射源个体识别方法

刘家豪, 郭英, 孟涛, 齐子森, 李红光

(空军工程大学信息与导航学院, 西安, 710077)

摘要 为改善通信辐射源指纹特征提取算法抗噪声及干扰能力差导致的对通信辐射源个体分类识别率低和稳定性差的问题,提出了一种基于经验模态分解和奇异值分解特征提取的方法。通过对信号进行经验模态分解,来克服噪声对指纹特征提取的影响,经希尔伯特-黄变换和奇异值分解实现对通信辐射源信号的指纹特征提取,结合支持向量机算法完成对通信辐射源的个体识别,从而提高了分类识别的正确率,经过对4类辐射源信号的实验验证表明识别效果具有明显提升。

关键词 辐射源个体识别;经验模态分解;指纹特征提取;奇异值分解;支持向量机

DOI 10.3969/j.issn.1009-3516.2021.04.010

中图分类号 TN975 **文献标志码** A **文章编号** 1009-3516(2021)04-0063-07

An Identification of Individual Radiation Sources Based on EMD and SVD Feature Extraction

LIU Jiahao, GUO Ying, MENG Tao, QI Zisen, LI Hongguang

(Information and Navigation College, Air Force Engineering University, Xi'an 710077, China)

Abstract Aimed at the problems that recognition rate is low, and stability of anti jamming and anti noise is poor, with the result that individual classification of communication emitter is poor and interference ability of fingerprint feature extraction algorithm of communication emitter is poor, a method based on empirical mode decomposition and singular value decomposition is proposed. The effect of noise on fingerprint feature extraction is overcome with signal being subjected to the Empirical Mode Decomposition, the fingerprint feature extraction of signal source is realized by the Hilbert-Huang Transform and Singular Value Decomposition in combination with Support Vector Machine (SVM) algorithm to complete individual identification of communication source, thus improving the accuracy of the classification and recognition. The experimental verification of the four types of emitter signals show that the ascension of recognition effect is obvious.

Key words individual radiation source identification; empirical mode decomposition; fingerprint feature extraction; singular value decomposition; support vector machine

收稿日期: 2021-01-25

基金项目: 国家自然科学基金(61871396)

作者简介: 刘家豪(1996—),男,四川绵阳人,硕士生,研究方向:通信信号处理、深度神经网络。E-mail:1391928216@qq.com

通信作者: 郭英(1961—),女,山西临汾人,教授,博士生导师,研究方向:通信信号处理、自适应信号处理。E-mail:yguo163@163.com

引用格式: 刘家豪, 郭英, 孟涛, 等. 基于EMD和SVD特征提取的通信辐射源个体识别方法[J]. 空军工程大学学报(自然科学版), 2021, 22(4): 63-69. LIU Jiahao, GUO Ying, MENG Tao, et al. An Identification of Individual Radiation Sources Based on EMD and SVD Feature Extraction[J]. Journal of Air Force Engineering University (Natural Science Edition), 2021, 22(4): 63-69.

基于传统信号参数估计手段获取信号常规波形参数实现通信辐射源识别的方法,难以对于相同波形参数类型的多个辐射源完成可靠的个体识别。例如同步组网的跳频网台^[1]中,网内每个跳频辐射源发出的信号都具有相同的频率集、跳频速率、码元速率、调制样式等参数特征,仅仅依靠常规的参数化稳态特征分析无法实现网内辐射源的个体识别,特别是在复杂的电磁环境中,提取常规稳态参数的准确性都难以保证。对于通信对抗支援而言,通信辐射源个体识别的可靠性和可信度下降,将直接导致不能提供有效的干扰引导,显然无法满足战场的电磁对抗需求。

对于辐射源个体来说,由于元器件制造工艺带来的个体误差不可避免,整机装配更存在着细微差别,因此尽管是来自于同一生产线上的同一型号电台,仍然存在着不影响信息传递的可检测、可重现的细微特征差异,我们将这些因为电台辐射源器件之间细微差别而造成的电台辐射源信号的细微特征称之为辐射源的指纹特征^[2-4]。针对辐射源的指纹特征研究表明,具有代表性的指纹特征反映在辐射源波形的暂态特征上,暂态特征^[5]相比于稳态特征,包含着更加丰富的个体特征,因此近年国内外学者针对辐射源的暂态特征进行了广泛的研究。信号暂态特征提取是实现通信辐射源识别的重要一步,常用的信号暂态特征研究主要集中在幅度、频率和相位3个层面,特征处理的数学方法针对性明显,但也存在相应的弊端,其中针对幅度方面,分型特征是进行非平稳信号分析的有力工具,但是基于幅度提取的信号暂态特征,区分度较小,尤其是在低信噪比条件下,识别效果不佳;针对频率方面,传统方法是对非平稳信号进行时频分析,包括不仅限于快速傅里叶变换(fast Fourier transform, FFT)、短时傅里叶变换(short-time Fourier transform, STFT)^[6]、小波变换(wavelet transform, WT)和希尔伯特-黄变换(Hilbert-Huang transform, HHT)^[7-8]等等,其中小波变换应用较为广泛,但是在低信噪比情况下,该方法表现较差,存在小波基的选择问题;针对相位特征方面,常用的方法是把信号暂态部分变换为复信号,之后再提取其相位特征,但是仅仅利用其相位特征来进行个体识别的效果并不理想,需要与信号暂态部分的其他特征相结合进行联合特征分析;此外,文献^[9]中提出了一中基于时频分析方法的辐射源调制识别方法,通过时频分布将信号转换到时间-频率二维图像,通过计算复杂度和特征维数实现特征提取,但这类方法存在最佳核函数的确定以及交叉项干扰等问题,比较复杂。近年来,经验模态分解

(EMD)^[10-12]发展比较迅速,通过EMD分解实现将原信号分解为各个信号分量,有效规避了核函数的选择问题,另外,信号的双谱特性还能有效的抑制高斯噪声,在处理非平稳信号中具有独特优势,可以动态描述信号的时频分布,放大不同辐射源的暂态特征差异,提供了信号暂态部分细微特征分析的新思路,文献^[13]中提出了基于EMD的Hilbert谱对称性参数估计算法,取得了较好的分类效果,但在抗噪声干扰的效果上还有所不足。

综合几类典型暂态个体特征定义及提取特点,本文采用EMD分解和HHT变换提取信号暂态特征,再运用奇异值分解选取重要程度高的部分特征^[14]作为输入数据,并结合机器学习等分类器设计理论,训练支持向量机(SVM)分类器,实现辐射源的个体识别。为了验证本文提出的个体识别方法的有效性,以识别准确率作为评判标准,选取了4类辐射源信号进行特征分析实验,验证了本设计获得了更好的识别效果。

1 特征提取

1.1 信号预处理

本文使用的实验数据是采集到有噪声的4个通信辐射源信号,信号采样率为1.01 MHz,最终获得的每个信号的长度为10 240帧。首先我们将每类信号依照一段数据为512帧划分,获得20个数据,从而得到4类信号共80个数据。

经验模态分解可以将任何一个复杂的序列分解为若干个本征模函数(intrinsic mode function, IMF)^[15]分量之和。采用EMD作为非平稳信号的处理手段,可以依据数据本身特征来进行信号分解,生成有限个IMF,各IMF分量代表了原信号不同的频率分量。通过EMD分解后,得到的IMF分量需满足以下2个条件:①经过分解得到的时间序列,极值点数量与过零点的数量差值不超过1;②在分析时间序列内,经过插值方式拟合的极值包络线的均值为0。

输入一个长度为512帧的通信辐射源信号是 $s(t)$,进行EMD处理的基本流程如下:

步骤1 找到原始信号 $s(t)$ 所有的局部极大值点和局部极小值点;

步骤2 用3次样条曲线拟合出上下极值点的包络 $e_{\max}(t)$ 和 $e_{\min}(t)$;

步骤3 求上下包络的平均值;

$$m(t) = \frac{e_{\max}(t) + e_{\min}(t)}{2} \quad (1)$$

步骤 4 原始信号与上下包络均值相减计算差值:

$$d(t) = s(t) - m(t) \quad (2)$$

步骤 5 判断得到的差值 $d(t)$ 是否满足 IMF 的 2 个条件,若不满足,则用 $d(t)$ 重复前面 4 个步骤,直到 $d_k(t)$ 满足条件为止,此时得到的 $d_k(t)$ 就是第一阶 IMF 分量: $imf_1(t)$;

$$\begin{cases} d_1(t) = s(t) - m_1(t) \\ d_2(t) = d_1(t) - m_2(t) \\ \vdots \\ d_k(t) = d_{k-1}(t) - m_k(t) \end{cases} \quad (3)$$

步骤 6 每得到一个 IMF 分量,就从原始信号 $s(t)$ 中减去,得到剩余分量 $r_1(t)$,再对剩余分量 $r_1(t)$ 重复前 5 个步骤,直到最后得到的剩余分量 $r_i(t)$ 就只是单调序列或者常值序列,此时便得到了信号的各阶模态分量。

$$\begin{cases} r_1(t) = s(t) - imf_1(t) \\ r_2(t) = r_1(t) - imf_2(t) \\ \vdots \\ r_i(t) = r_{i-1}(t) - imf_i(t) \end{cases} \quad (4)$$

经过 EMD 分解后,信号 $s(t)$ 最终分解为 7 个 IMF 分量和 1 个残余分量,如图 1 所示。

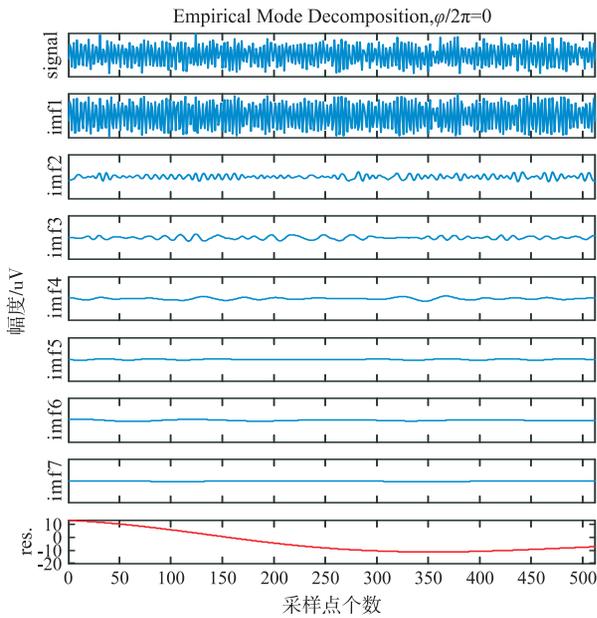


图 1 EMD 分解示意图

本实验共 80 个数据,需要对每个数据都进行 EMD 分解,得到每个数据的各阶 IMF 分量和一个残余分量。残余分量是单调序列或者常值序列,对信号特征提取没有实质影响,因此,舍去残余分量,用 7 个 IMF 分量: $imf_i(t)$, ($i = 1, 2, \dots, 7$) 进行特征提取。

1.2 HHT 变换提取暂态特征

对非平稳信号而言,比较常用的时频分析方法

主要包括 FFT、STFT、WT 和 HHT 等。其中 FFT 对非平稳信号处理效果差,STFT 存在窗函数选择窗长问题直接影响时域和频域的分析不够准确,WT 不存在加窗函数的问题却存在小波基的选择问题,本文选取 HHT 变换作为信号暂态部分的时频分析手段,从而实现辐射源的指纹特征提取,该方法是以前文所提到的经验模态分解方法为基础的一种非线性分析的新方法,可以自适应地利用信号的局部信息,从而获得信号某一时刻的瞬时状态,既可以很好地体现信号的时频变化,又克服了小波变换中需要选择小波基的困难。

对 EMD 分解后的 7 个 IMF 分量: $imf_i(t)$, ($i = 1, 2, \dots, 7$) 进行 HHT 变换,即对每一个 IMF 分量进行 Hilbert 变换:

$$x_i(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{imf_i(\tau)}{t - \tau} d\tau \quad (5)$$

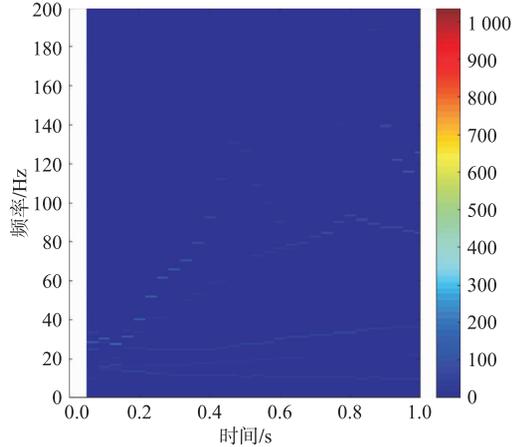
再求相应 IMF 分量的瞬时频率 $\omega_i(t)$:

$$\theta_i(t) = \arctan (d_i(t)/imf_i(t)) \quad (6)$$

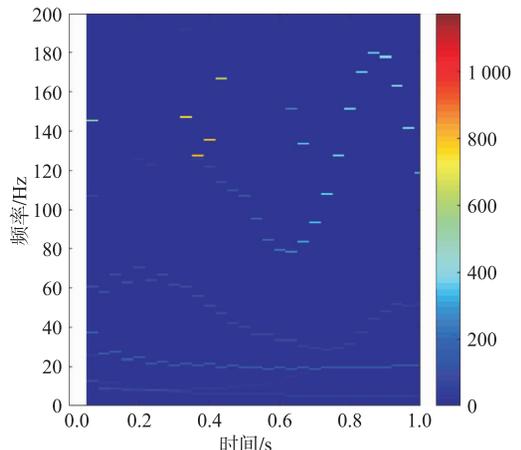
$$\omega_i(t) = \frac{d\theta_i(t)}{dt} \quad (7)$$

根据 $\omega_i(t)$ 和 t 就可以得到时频图。

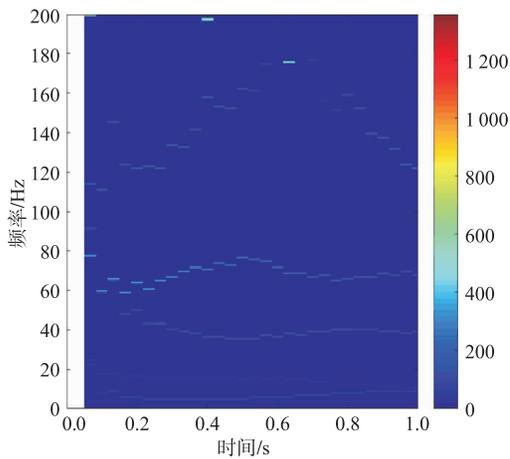
本节选择了 4 类辐射源信号进行实验,对信号进行 EMD 分解后,通过 HHT 实现信号的特征提取,4 类信号的 HHT 谱如图 2 所示。



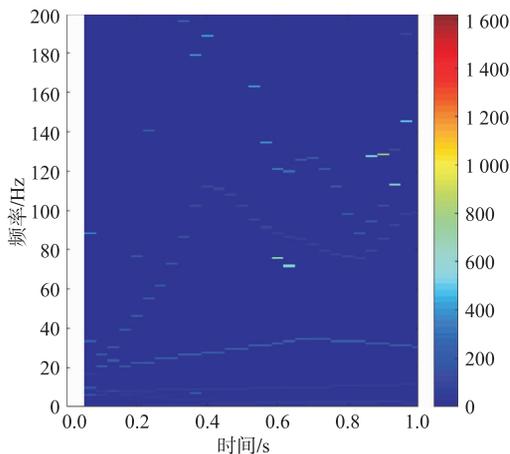
(a) 第 1 个信号 HHT 谱



(b) 第 2 个信号 HHT 谱



(c) 第3个信号 HHT 谱



(d) 第4个信号 HHT 谱

图2 4类信号的 HHT 谱

通过4类信号的HHT谱可以清晰地看出4类辐射源信号的差异所在,由此可以推测,可以通过HHT变换提取信号的指纹特征来实现辐射源的个体识别。

1.3 基于SVD进行特征选择

因为进行经验模态分解之后的信号中,通常少量的数据就包含了信号的主要信息,因此,本文采用奇异值分解(singular value decomposition, SVD)^[16]的方法来进行特征选取。

SVD奇异值分解的原理是:假设一个 $m \times n$ 阶的矩阵 \mathbf{M} ,那么必定存在一个 $m \times n$ 阶的对角矩阵 $\mathbf{\Sigma}$,一个 $m \times m$ 阶的正交矩阵 \mathbf{U} 满足 $\mathbf{U}^T \mathbf{U} = \mathbf{I}$,一个 $n \times n$ 阶的正交矩阵 \mathbf{V} 满足 $\mathbf{V}^T \mathbf{V} = \mathbf{I}$,使得矩阵 \mathbf{M} 满足分解:

$$\mathbf{M} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T \quad (8)$$

满足式(8)就称为 \mathbf{M} 的奇异值分解,奇异值是分布在 $\mathbf{\Sigma}$ 对角线上的数值,为 σ_i ,且按大小进行排列。

在很多情况下,前10%甚至更少的奇异值所包含的信息就能够达到90%以上,所以,我们只需要

用最大的 k 个奇异值就能够对矩阵进行描述,即:

$$\mathbf{M}_{m \times n} \approx \mathbf{U}_{m \times k} \mathbf{\Sigma}_{k \times k} \mathbf{V}_{k \times n}^T \quad (9)$$

因此,通过奇异值分解,我们就可以得到较少的包含了大部分特征信息的数据。

2 信号分类算法-SVM 算法

对辐射源进行个体识别的算法通常有聚类算法和分类算法,两类算法是完全不同的。聚类算法所研究的数据是没有标签的,无法通过标签实现数据的分类,其分类依据是通过算法判断各个数据之间的相似程度,通过对比相似程度,相似程度高的数据归为一类,实现了数据的分类,最终归为一类的数据相似程度高,而不同类别之间相似程度低。常见的分类算法如支持向量机(SVM)^[17-18]、神经网络、最邻近分类器都是通过对输入的数据进行训练,按照一定的标准将数据贴上标签,这个标签就是代表着数据的类别,最终通过标签实现对测试集数据的分类。本文选取SVM算法对奇异值分解提取的暂态特征进行分类,实现辐射源的识别。

SVM算法最初是为了二分类问题设计的,通过寻找一个最优的分类器将具有多个属性的数据正确地分为两类,为了达到最大的泛化能力,要求距离分类器最近的不同数据集之间的间隔最大^[19]。分类器的类型取决于数据属性的维数,二维的分类器是直线,三维的分类器是平面,多维的分类器是超平面。使用线性函数构造这个分类器:

$$f(x) = w \cdot x + b \quad (10)$$

式中: w 是权重; $x = (x_1, x_2, \dots, x_n)$ 为训练数组; n 为特征个数; x_i 表示每个 x 在属性 i 上对应的值; b 是偏置; $w \cdot x$ 表示 w 和 x 的内积。对数据进行分类,就需要将数据带入这个函数,通过计算 $f(x)$ 的值来确定所属的类别。当 $f(x) > 0$ 就判定为类1,当 $f(x) < 0$ 就判定为类2。

将数据拓展至任意维,求最优分类器就是求最优超平面。当超平面:

$$w \cdot x + b = 0 \quad (11)$$

满足:①对所有数据分类正确;②分类间隔最大,即 $\|w\|$ 最小,则该超平面为最优超平面。假设分类数据有 m 个 $(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)$, $x_i \in R^n$, $y_i \in \{-1, +1\}$ 表示类别号,实现SVM算法等同于求解一个带有不等式约束的非线性规划问题:

$$\begin{cases} \max \frac{1}{\|\omega\|} \\ \text{s. t. } y_i(\omega \cdot x_i + b) - 1 \geq 0 \\ i = 1, 2, \dots, m, \omega \in R^n, x \in R^n \end{cases} \quad (12)$$

SVM 是一个二分类器,但可以通过训练多个分类器实现多分类,常见的有以下两种方式:

一对一(one-vs-one):假设有 n 类数据,其中每 2 个类就需要训练 1 个分类器,总共需要训练 $n(n-1)/2$ 个分类器。例如有 4 类数据:1,2,3,4,就需要训练 6 个分类器,分别是:1 类和 2 类,1 类和 3 类,1 类和 4 类,2 类和 3 类,2 类和 4 类,3 类和 4 类。输入数据经过训练的所有分类器的预测,最后由投票的方式来决定它最终的类属性。

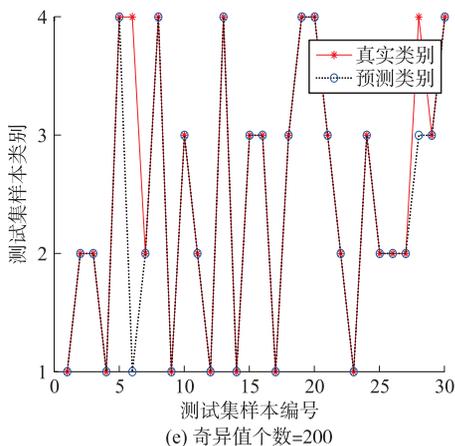
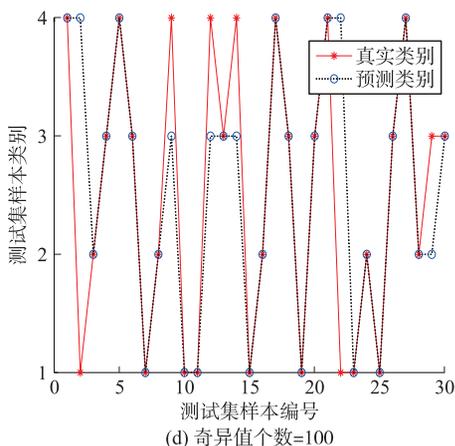
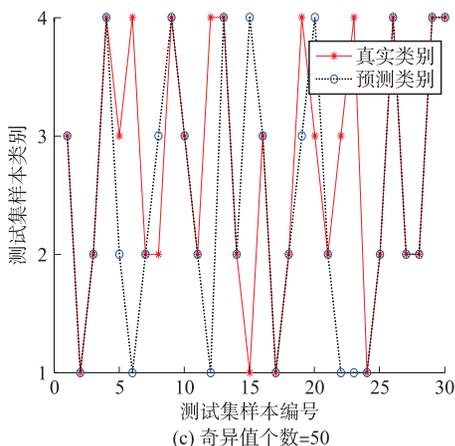
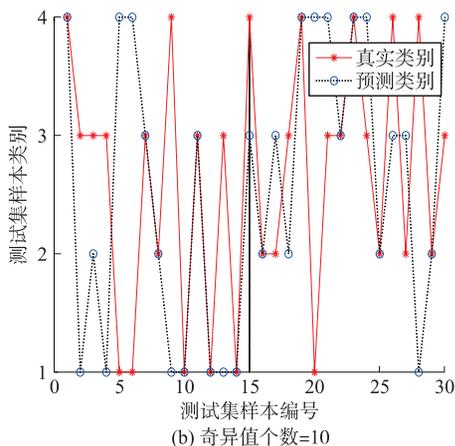
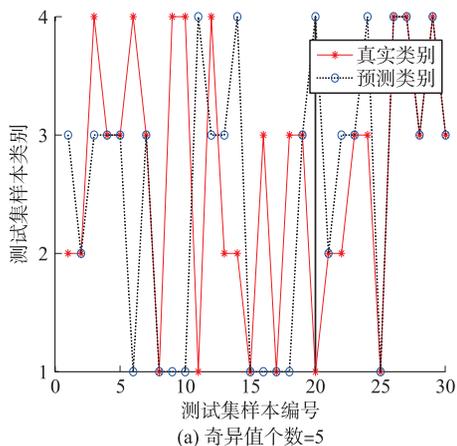
一对多(one-vs-rest):假设有 n 类数据,需要训练 n 个分类器。例如有 3 类数据: a, b, c ,要分出 a 就将 a 类数据设为正集,其它 2 类同设为负集,同样操作区分 b 类和 c 类,一需要训练 3 个分类器。输入数据经过训练的所有分类器的预测,再通过比较置信度的大小,选择置信度最大的类别标记为最终的分类结果。

本文运用一对多的方式训练多个分类器,对奇异值分解后得到辐射源特征进行分类。

3 实验验证

将辐射源信号进行 EMD 分解以及 HHT 变换提取暂态特征,并通过 SVD 奇异值分解得到的数据量少、包含了主要信息的信号特征作为 SVM 分类器的输入。我们把数据随机划分为训练集和测试集,其中训练集划分 50 个数据,测试集划分 30 个数据。运用训练集的数据训练 4 个分类器,再用测试集对训练好的分类器进行测试,观察识别准确率,实验结果如下所示:

当奇异值个数分别为 5、10、50、100、200、240 时,测试集与预测结果对比图见图 3。



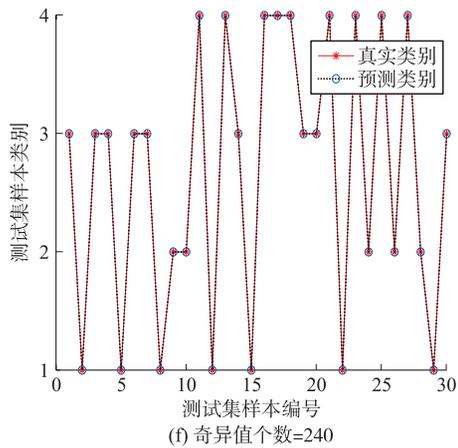


图3 奇异值个数取不同值时的对比图

通过表格能够更加直观地显示分类结果及识别率,如表1~6所示。

表1 奇异值个数=5时的分类结果

信号类别编号	识别结果/个				识别率/%	平均识别率/%
	1类	2类	3类	4类		
1	4	0	0	2	66.7	53.3
2	0	2	3	1	33.3	
3	2	0	7	1	70.0	
4	3	0	2	3	37.5	

表2 奇异值个数=10时的分类结果

信号类别编号	识别结果/个				识别率/%	平均识别率/%
	1类	2类	3类	4类		
1	3	0	0	3	50.0	43.3
2	0	4	2	0	66.7	
3	3	2	3	2	27.3	
4	2	0	2	3	42.9	

表3 奇异值个数=50时的分类结果

信号类别编号	识别结果/个				识别率/%	平均识别率/%
	1类	2类	3类	4类		
1	3	0	0	1	75	70
2	0	9	1	0	90	
3	1	1	3	1	50	
4	3	0	1	6	60	

表4 奇异值个数=100时的分类结果

信号类别编号	识别结果/个				识别率/%	平均识别率/%
	1类	2类	3类	4类		
1	7	0	0	2	77.8	80
2	0	5	0	0	100.0	
3	0	1	7	0	87.5	
4	3	0	2	5	37.5	

表5 奇异值个数=200时的分类结果

信号类别编号	识别结果/个				识别率/%	平均识别率/%
	1类	2类	3类	4类		
1	7	0	0	0	100	93.3
2	0	8	0	0	100	
3	0	0	7	0	100	
4	1	0	1	6	75	

表6 奇异值个数=240时的分类结果

信号类别编号	识别结果/个				识别率/%	平均识别率/%
	1类	2类	3类	4类		
1	7	0	0	0	100	100
2	0	5	0	0	100	
3	0	0	9	0	100	
4	0	0	0	9	100	

因为本实验中训练集和测试集是随机分配,所以每一次实验,测试集中每一类信号的个数是不固定的。

从表格中可以看出,当奇异值个数较少时,识别准确率较低,识别效果不佳;当奇异值个数较多时,随着个数的增加,识别准确率不断上升,当奇异值个数达到240个时,识别率为100%。此时,作为信号特征参与信号识别的奇异值个数240占信号总奇异值个数510的47.1%,占比较大,运算速度略有下降,以较小的代价获得了更高的识别率。实验结果表明,本文所设计的分类器能够较好地辐射源个体进行识别,识别率高。其优点在于处理较大数据时可以保证较高的效率。

4 对比验证

通过小波变换WT提取暂态特征^[20-21],并结合支持向量机进行分类,得到的识别准确率与本文方法得到的识别准确率对比如表7所示。

表7 不同特征提取方法的分类识别结果

特征提取方法	辐射源1识别率/%	辐射源2识别率/%	辐射源3识别率/%	辐射源4识别率/%	平均识别率/%
	EMD+SVD	100	100	100	
WT	87.5	100	100	85.7	93.3

从分类识别结果的表格中可以看出,本文提出的通信辐射源识别方法效果更好。

5 结语

通过实验验证,本文采用的EMD算法和SVM

算法,通过不断调整提高奇异值分解个数来增加指纹特征的数据量,最终实现了对四类辐射源数据的个体识别,在增加奇异值分解个数的同时,运算速度并没有因此而大幅降低,保证了较高的运算效率。通过EMD分解克服了以往对信号进行特征提取的经典时频分析方法的抗噪性能差的弊端,对辐射源信号的特征提取手段进行了一定的优化,最终通过SVM算法也获得了较好的分类结果。

参考文献

- [1] 徐倩. 跳频同步组网技术研究[D]. 西安:西安电子科技大学, 2007.
- [2] 睦萍, 郭英, 张坤峰, 等. 半监督条件下的CRC跳频电台指纹特征识别[J]. 系统工程与电子技术, 2019, 41(1):192-198.
- [3] 顾晨辉, 王伦文. 基于频域瞬时特征的跳频电台个体识别方法[J]. 计算机工程与应用, 2013, 49(22): 223-226.
- [4] 宋春云, 徐建敏. 基于混沌理论的跳频电台细微特征提取[J]. 通信对抗, 2009(4):10-13, 17.
- [5] 周斌, 王秀敏, 果然, 等. 辐射源个体特征提取技术综述[J]. 电讯技术, 2011, 51(6):162-168.
- [6] CHKKERUR S, CARTWRIGHT A N, GOVINDARAJU V. Fingerprint Enhancement Using STFT Analysis[J]. Pattern Recognition, 2007, 40(1): 198-211.
- [7] 陈娟. Hilbert-Huang变换及其在信号处理中的应用[D]. 大连:大连理工大学, 2006.
- [8] 秦毅, 秦树人, 毛永芳. 改进的Hilbert-Huang变换在信号瞬态特征提取中的应用[J]. 振动与冲击, 2008, 27(11):129-133.
- [9] 高敬鹏, 孔维宇, 刘佳琪, 等. 基于时频分析的自适应PCA辐射源调制识别[J]. 应用科技, 2018, 45(5): 33-37.
- [10] 桂云川, 杨俊安, 吕季杰, 等. 基于经验模态分解的通信辐射源分形特征提取算法[J]. 探索与控制学报, 2016, 38(1):104-108.
- [11] HUANG N E, SHEN Z, LONG S R, et al. The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis[J]. Proceedings Mathematical Physical & Engineering, 1998, 454(1971):903-995.
- [12] 梁江海. 基于经验模态分解的通信信号细微特征分析[D]. 长沙:国防科学技术大学, 2013.
- [13] 周成燕. 基于暂态分析的无线通信设备信号指纹识别[D]. 绵阳:西南科技大学, 2018.
- [14] 陆满君. 通信辐射源个体识别与参数估计[D]. 哈尔滨:哈尔滨工程大学, 2010.
- [15] 赵园青, 池龙, 马赛, 等. 基于EMD算法的空间自旋目标平动补偿与微动特征提取[J]. 空军工程大学学报(自然科学版), 2013, 14(5):40-43.
- [16] 杨宇, 于德介, 程军圣. 基于EMD的奇异值分解技术在滚动轴承故障诊断中的应用[J]. 振动与冲击, 2005, 24(2):12-15.
- [17] HAN J, TANG X, TANG Z, et al. An Evaluation Method on Radar Emitter Signal Recognition Effect Based on SVM[C]//2018 5th International Conference on Systems and Informatics (ICSAD). Nanjing, China: IEEE, 2018: 700-704.
- [18] FRIEDRICH S, IGEL C. Evolutionary Tuning of Multiple SVM Parameters[J]. Neurocomputing, 2005, 64:107-117.
- [19] 杨立明, 哈章, 杨晓蓉, 等. 辐射源个体识别中分类器应用[J]. 空军工程大学学报(自然科学版), 2012, 13(1): 28-32.
- [20] 周斌. 信号细微特征提取及识别技术研究[D]. 哈尔滨:哈尔滨工业大学.
- [21] 冯浩, 周维维, 刘毅. 基于小波变换和二叉树结构的电能质量扰动分类[J]. 低压电器, 2011(5):35-38, 50.

(编辑:徐楠楠)