

A Mathematical Model of Classification and Identification of Communication Radiation Source Signals

Rui Feng¹, Xiaoying Fu², Jiaqi Zhao¹, Rongxun Wang³

¹Shandong University of Science and Technology, Qingdao Shandong

²College of Mathematics and System Science, Shandong University of Science and Technology, Qingdao Shandong

³College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao Shandong

Email: 964413522@qq.com

Received: Mar. 2nd, 2018; accepted: Mar. 21st, 2018; published: Mar. 28th, 2018

Abstract

Under the condition of the rapid development of modern information technology, military forces in various countries are making heavy use of conventional or non-conventional communications equipment in order to win the war. Therefore, this paper attempts to use the Hilbert-Huang transform to extract and analyze the signal characteristics of the same type of communication radiation source, and then use the clustering algorithm to identify the individual signals from the classification and identification of the general communication signals. However, the actual signal is noisy, so this paper applies wavelet threshold denoising method to noise reduction. These are of great value and significance to enhance the capability of military communications in fighting against war under information conditions, to enhance access to and utilization of intelligence information in the coming wars, and to grasp the initiative of the war.

Keywords

Hilbert-Huang Transform, Fuzzy Mean Clustering Method, Minkowski Distance, Wavelet Threshold Denoising

关于通信辐射源信号分类与识别的数学模型

冯睿¹, 付晓莹², 赵佳琪¹, 王荣勋³

¹山东科技大学, 山东 青岛

²山东科技大学, 数学与系统科学学院, 山东 青岛

³山东科技大学, 计算机科学与工程学院, 山东 青岛

Email: 964413522@qq.com

收稿日期: 2018年3月2日; 录用日期: 2018年3月21日; 发布日期: 2018年3月28日

摘要

在现代信息技术高速发展的条件下, 为了取得战争的胜利, 各国军队都在大量使用常规或非常规的通信设备。因此, 本文尝试使用Hilbert-Huang变换对同类通信辐射源信号特征的提取和分析, 再使用聚类算法来实现从一般通信信号的分类识别到个体信号的识别。然而实际信号是有噪声的, 所以本文应用小波阈值降噪方法对信号进行降噪处理。这些对在信息化条件下提高军事通信对抗作战能力, 保证在未来战争中情报信息的获取与利用, 掌握战争的主动权具有重要价值和意义。

关键词

Hilbert-Huang变换, 模糊均值聚类, Minkowski距离, 小波阈值降噪

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

1.1. 问题的产生

在现代战争中, 各国军队对信息战的要求越来越高。为了及时、准确、有效地获取敌方情报, 有针对性地对敌重要通信装备及其载体(如水面或水下目标、空中飞行目标、陆地固定目标等)进行监视、控制、干扰或打击, 这就对通信信号的分析研究手段提出了更高的要求。因此, 通过对同类通信辐射源信号特征的提取和分析, 实现从一般通信信号的分类识别到个体信号的识别, 对在信息化条件下提高军事通信对抗作战能力, 保证在未来战争中情报信息的获取与利用, 掌握战争的主动权具有重要价值和意义。本文尝试研究并解决以下问题:

问题一: 分析并提取通信辐射源信号的特征, 并基于这些特征建立识别聚类模型, 并对第一组数据中 500 个信号样本进行聚类, 区分出五种通信辐射源信号。

问题二: 根据问题一的聚类结果, 对第二组数据所给出的两个降噪后的信号样本进行识别, 将其归类, 并对分类的可靠性进行分析。

问题三: 对第三组数据中四个有噪声影响的实际信号进行识别, 确定这四个信号是否属于问题一的五类, 如果是, 则又属于哪一类。

1.2. 数据重现

第一组数据给出了 5 种通信辐射源的侦测接收信号的 500 个信号样本(如表 1 所示), 其采样频率均为 $f_s = 1.01 \text{ MHz}$; 第二组数据给出了 2 个降噪后的实际通信辐射源信号样本(如表 2 所示); 第三组数据给出了 4 个有噪声影响的实际通信辐射源信号(如表 3 所示)。

Table 1. The first set of data (excerpt)**表 1.** 第一组数据(节选)

-746	-608	-735	-80	528	191	212	...
-300	-224	-321	176	336	156	153	...
500	736	-473	-48	-720	-228	-247	...
-302	-176	327	-224	144	-279	-100	...
-75	-640	51	144	576	293	325	...
-655	496	-666	144	-464	162	45	...
482	320	474	-240	-208	-288	-235	...
-305	-464	301	64	416	-178	-1	...
729	-16	729	224	112	310	283	...
-305	320	-335	-224	-336	150	-62	...
500	-192	-484	-64	-208	-329	-271	...
...

Table 2. The second set of data (excerpt)**表 2.** 第二组数据(节选)

	64	414
	-64	171
	-64	-405
	208	-172
	-80	417
	-240	196
	192	-424
	160	-175

Table 3. The third set of data (excerpt)**表 3.** 第三组数据(节选)

-198	-476	436	-277
472	-447	-47	-480
334	287	14	209
-554	6	940	138
121	-640	-202	-722
582	427	-300	769
-621	251	531	798
-530	-996	643	-810
690	22	-958	-445
...

1.3. 问题的分析

问题一分析

问题一的关键是对信号的处理,即特征提取和聚类。其中,特征提取是指对信号数据进行处理的过程,特征提取的好与坏对聚类的情况和后期的信号个体的分类识别有着重要的影响。所以,本文需要建立一种具有优异性能的信号特征提取方法。查阅相关资料[1],本文发现 Hilbert-Huang 变换是一种应用广泛的信号特征提取方法,它可以提取信号的时频分布特征,而且局部性良好,对平稳和非平稳信号皆可适用。在后文将介绍 Hilbert-Huang 变换的使用。

另外,本文需要对提取出的信号特征进行聚类分析,聚类分析是将信号特征分组成多个类的过程,使得同一组内的信号特征具有较高的相似度,而不同组中的信号特征相似度低。由于本文的信号特征数据都需要用来聚类,没有大量的学习样本可以使用,所以本文所以本文采用无监督聚类算法进行聚类[2]。

问题二分析

问题二主要是对问题一所建模型的应用,它主要包括:识别归类及其可靠性分析。

基于聚类模型中得出的 5 个类别,本文可以计算新的信号特征与 5 个聚类中心的信号特征的相似程度(Minkowski 距离),显然新的信号应该被归入与其最相似的组内。与此同时,如果新的信号特征与各组的聚类中心信号特征的相似程度较为接近,则认为这种分类较为模糊,可靠性不高;否则则认为这种分类较为明确,有一定的可靠性。

问题三分析

在对问题三的信号进行降噪之后,该问题转化成对以上模型的综合应用,所以问题三的关键在于对信号进行降噪。查阅相关资料得知,受噪声影响的信号的表达式为 $r_n = \phi(x_n) + v_n$,并统计分析了第三组数据中所给的数据发现噪声其实是围绕着信号成正态分布的。因此可以设置一个阈值来对噪声进行过滤,获得一个降噪信号,再对其进行特征提取和分类。

如果发现该降噪信号与 5 个聚类中心的相似程度小于问题一聚类过程中出现的最小相似程度,则认为这个降噪信号不属于这 5 类;否则,归入其中一类。

2. 模型的阐述与求解(总流程如图 1 所示)

2.1. 数据特征提取模型

第一步:进行 EMD 分解

把目标信号的采样值输入 Matlab,做出信号图像,设信号的时间序列形式 $x(t)$,用三次样条插值做出图像的上、下包络线,分别为 $u(t)$ 和 $v(t)$,则上、下包络的平均曲线为

$$m(t) = \frac{1}{2} [u(t) + v(t)].$$

用原信号的公式减去 $m(t)$,得到新的信号函数 $h_1(t) = x(t) - m(t)$,若 $h_1(t)$ 满足下列两个条件:

- 在信号数据序列里, |过零点的数量 - 极值点的数量| ≤ 1 ;
- 在任何时间点,由数据序列的局部极大值与局部极小值确定的包络均值为 0。

则 $h_1(t)$ 为一个 IMF 分量,若不满足上述两个条件,则将 $h_1(t)$ 作为原信号重复上述步骤,直到得到一个 IMF 分量 $c_1(t)$ 。然后,将 c_1 从原始信号 $x(t)$ 中分离出来,得到一个新的信号 $r_1(t)$,即

$$r_1(t) = x(t) - c_1(t),$$

再用 $r_1(t)$ 作为原信号重复上述操作,求取新的 IMF 分量,直到 $r_n(t)$ 成为单调函数,停止循环。 $r_n(t)$ 为残余分量,表示信号的平均趋势。此时已提取出该信号的所有 IMF 分量,原信号可表示为

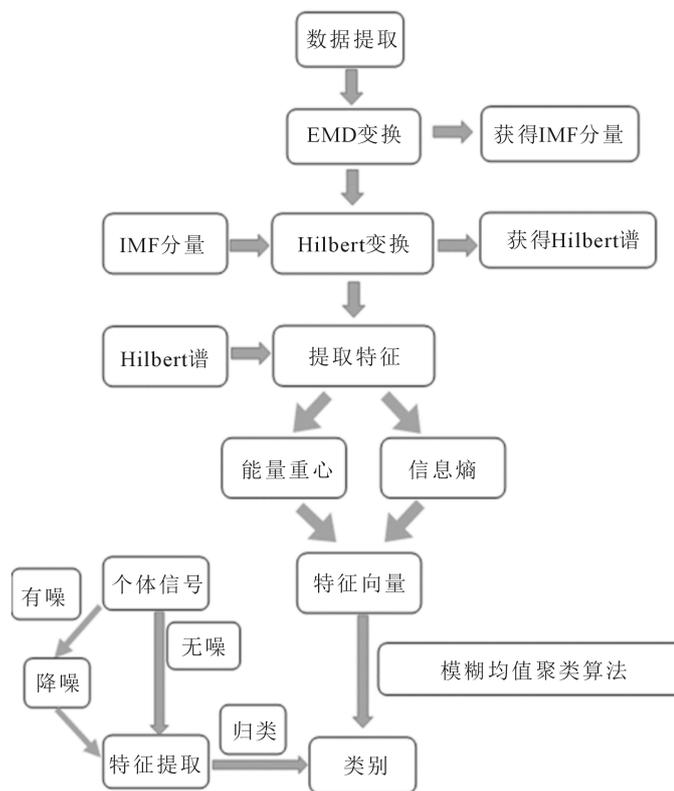


Figure 1. The total flow chart of classification and identification of communication radiation source signals

图 1. 通信辐射源信号的分类与识别的总流程图

$$x(t) = \sum_{i=1}^{n-1} c_i + r_n(t).$$

第二步：进行 Hilbert 变换

首先，对得到的每一个 IMF 分量进行 Hilbert 变换如下，

$$H[c_i(t)] = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{c_i(\tau)}{t-\tau} d\tau,$$

通过查表可进行求解，得到 IMF 分量的解析信号

$$z_i(t) = c_i(t) + jH[c_i(t)] = A_i(t)e^{j\theta_i(t)},$$

其中，信号瞬时振幅为：

$$A_i(t) = \sqrt{c_i(t)^2 + H[c_i(t)]^2},$$

信号瞬时相位为：

$$\theta_i(t) = \arctan(H[c_i(t)]/c_i(t)).$$

去除残余分量后，原信号 $x(t)$ 的 Hilbert 谱为：

$$H(\omega, t) = \sum_{i=1}^n z_i(t)$$

其中, $\omega = \omega_i(t) = \frac{d\theta_i(t)}{dt}$ 为信号的瞬时频率。

最后, 对 Hilbert 谱进行积分, 求取信号 $x(t)$ 的边际谱

$$h(\omega) = \int_0^T H(\omega, t) dt,$$

其中, T 为 $x(t)$ 信号的持续时间, 边际谱反映出在信号持续时间内的幅度分布, 即能量分布。

对第一组数据中的 500 个信号均执行上述两个操作, 完成 Hilbert-Huang 变换, 因数据量太大, 以信号一为例, 共得到 8 个 IMF 分量及其频谱(如图 2 所示)。

第三步: 提取边际谱的特征

在边际谱中, 能量重心和信息熵是衡量频域信息的重要参数, 能量重心表示信号能量在频域中的集中点, 信息熵表示信号能量在频域中分布的均匀程度, 因此提取这两项参数作为信号特征。具体步骤如下:

Step 1 提取边际谱能量重心

用 Matlab 计算出信号的 Hilbert 谱矩阵为 $M(t, f)$, 其中 f 为信号频域包含的频率成分, 计算 Hilbert 谱的总能量为:

$$E = \sum_t \sum_f M(t, f).$$

边际谱的总能量为:

$$h(f) = \sum_t M(t, f).$$

则边际谱能量重心为:

$$W = \frac{\sum_f h(f) * f}{E},$$

即加权积分和与边际谱总能量的比值。

Step2 提取边际谱信息熵

将 Hilbert 谱矩阵中的值转换为 0~1 之间的概率值,

$$P(t, f) = \frac{M(t, f)}{E},$$

计算边际谱中幅度的概率分布

$$P(f) = \sum_t P(t, f),$$

得到边际谱的信息熵

$$S = -\sum_f P(f) \ln P(f),$$

以边际谱能量重心与边际谱信息熵组成辐射原信号样本的特征向量。

2.2. 识别分类模型(模糊均值聚类算法)

在前面的模型中本文已将辐射源信号数据转化成和其特征参数数据, 下面本文将对不同信号的特征参数数据进行聚类, 区分出 5 种通信辐射源的信号。

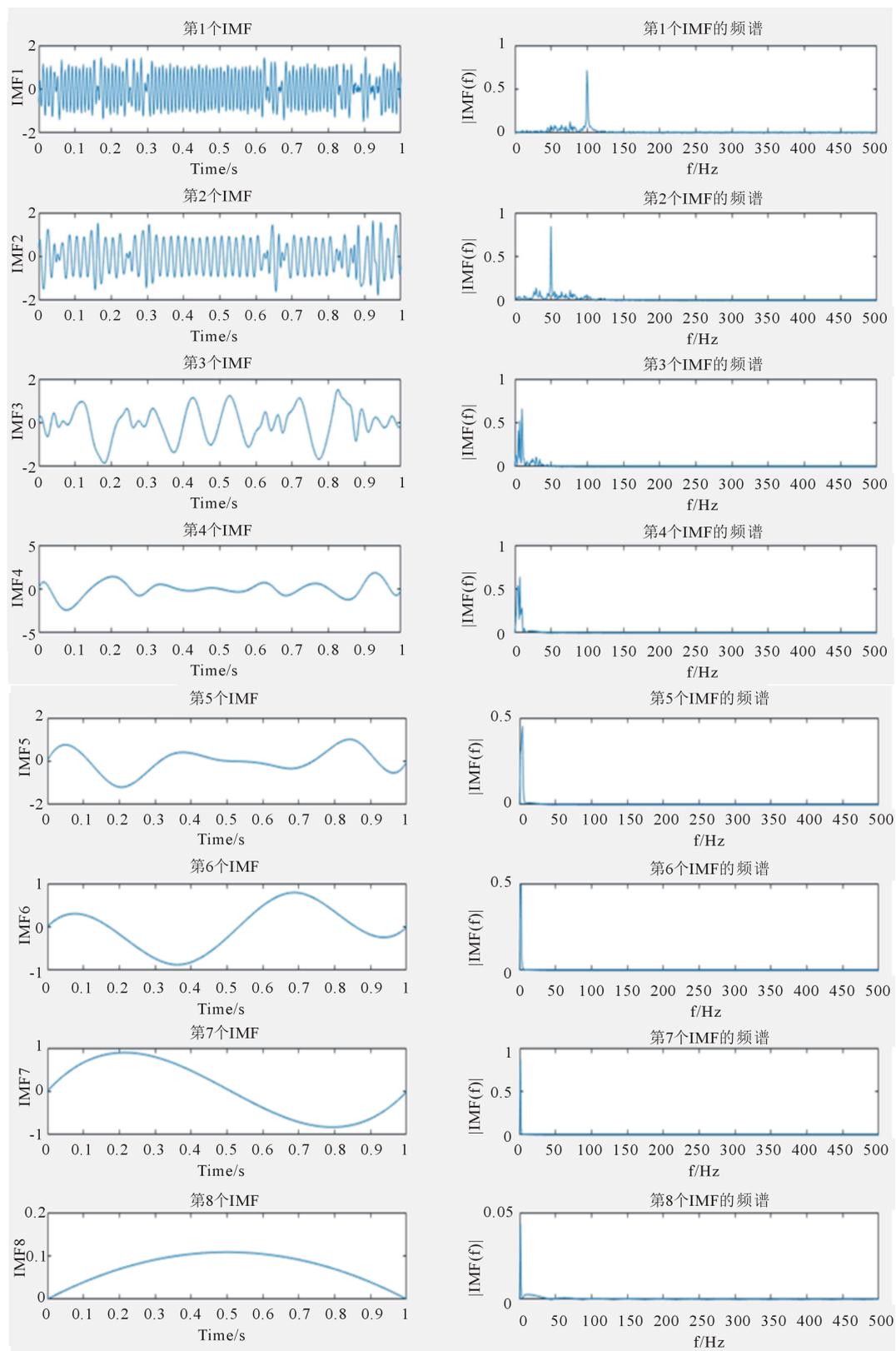


Figure 2. Eight IMF components and their spectra
图 2. 8 个 IMF 分量及其频谱

随着模糊集理论的出现,人们在K-均值算法的基础上提出了模糊均值聚类算法,其核心思想是:设数据集 $Y = \{y_1, y_2, \dots, y_N\}$,其中 y_i 为第 i 个特征提取后的信号向量;它的模糊划分可用隶属度矩阵 $P = \{p_{ij}\} \in R^{CN}$ 表示,矩阵 P 的元素 p_{ij} 表示第 $j(j=1, 2, \dots, N)$ 个特征参数向量属于第 $i(i=1, 2, \dots, N)$ 类的隶属度。 p_{ij} 满足条件如下

$$\sum_{i=1}^C p_{ij} = 1, p_{ij} \in [0, 1], \quad (1)$$

$$\sum_{j=1}^N p_{ij} > 0, \quad (2)$$

$$\sum_{i=1}^C \sum_{j=1}^N p_{ij} = N. \quad (3)$$

聚类目标函数,其一般形式为:

$$J_m = \sum_{i=1}^C \sum_{j=1}^N p_{ij}^m \cdot d(y_j, o_i), \quad (4)$$

其中, o_i 是第 i 个聚类中心,且 $m \in [1, \infty]$ 是一个加权指数。

隶属度矩阵的更新公式为:

$$p_{ij} = \left(\sum_{k=1}^C \left(\frac{d(y_j, o_k)}{d(y_j, o_i)} \right)^{\frac{2}{m-1}} \right)^{-1}. \quad (5)$$

聚类中心的更新公式为:

$$o_i = \frac{\sum_{j=1}^N p_{ij}^m y_j}{\sum_{j=1}^N p_{ij}^m}. \quad (6)$$

模糊均值聚类算法通过反复迭代优化目标函数,即可完成对数据集 $Y = \{y_1, y_2, \dots, y_N\}$ 的聚类,具体步骤如下:

- 1) 初始化聚类中心 $O = \{o_1, o_2, \dots, o_c\}$;
- 2) 用 $[0, 1]$ 之间的随机数初始化隶属度矩阵 P ,使其满足(1)、(2)、(3)的约束条件;
- 3) 通过式(6)更新聚类中心;
- 4) 通过式(4)计算价值函数,并选取合适的距离形式和阈值 ε 。比较 J_m^k 和 J_m^{k+1} ,若 $|J_m^{k+1} - J_m^k| < \varepsilon$,则停止迭代;否则用式(5)重新计算 P 值,返回步骤(3)。

模糊均值聚类算法需要两个参数:一个是聚类数目 C ,另一个是参数 m 。一般来说, C 要远远小于聚类样本个数,同时要保证 $C > 1$ 。对于 m ,它是一个控制算法柔性参数,从相关资料中得知,当 m 过大,则聚类效果会很差,而如果 m 过小,则会接近于非模糊的均值聚类算法。因此在一般情况下,加权指数 $m = 2$ [3]。输出的结果是 C 个聚类中心向量 y_i^c 和 $C \times N$ 的一个模糊划分矩阵,这个矩阵表示的是每个向量对于各类的隶属度,其中隶属度的最大值即为该向量属于的类别,而 C 个中心向量则可视作该类的代表向量。

2.3. 归类模型

对于一个新的信号本文同样可以获取其特征参数向量 y_i ,计算该特征向量 y_i 与上文推导得出的 C 个

中心向量 y_k^c 的距离, 选取距离的最小值

$$\min d(y_i, y_k^c)$$

其对应中心向量 y_k^c 所在组即为新信号的组别。

第二组数据中两个信号的分类结果如下表 4 所示。

3. 可靠性分析

定义一个的信号特征 y 与各组的聚类中心信号特征 o_i 的距离和为:

$$k = \sum_{i=1}^c d(y, o_i),$$

其中, k 称为识别的可靠性, k 的标准值取 5 个聚类中心向量之间距离的最大值 $k_{\max}^c = 2$ 。

当 $k > k_{\max}^c$ 时, 说明这次识别具有可靠性, 反之则没有。 k 值越大, 说明这次识别的可靠性越好。

对于问题二中的识别可靠性如下表 5 所示。

4. 小波阈值降噪模型

在实际通信信号的传输过程中, 信号会受到噪声的影响, 干扰信号的识别, 无法直接用模型一进行分类, 因此, 需对信号进行降噪处理。资料表明, 小波阈值降噪是一种常用的信号降噪方法, 它具有去相关性、基的选择的多样性等特点, 能刻画信号的非平稳特性且具有良好的时频特性, 是一种理想的信号降噪方法。

小波阈值降噪通过小波的多分辨分析特性, 将信号在不同尺度下进行分解, 使交织在一起的各种不同频率组成的混合信号分解成不同频段的子信号, 将各频段采用不同的阈值去噪后再进行重构, 从而达到去噪的目的。共包括四个部分:

- 1) 对信号进行小波分析, 将原始信号分解为一系列的近似分量和细节分量, 信号的噪声主要集中在信号的细节分量上。
- 2) 选取阈值函数, 即选择对超过和低于阈值的小波系数的模的处理方法。
- 3) 确定阈值, 确定阈值即可确定噪声频段的范围, 进行去除。
- 4) 小波重构, 上述处理后可获得去噪后的小波系数和尺度函数, 将其重构后可以得到去噪后的信号。

Table 4. Classification results for two signals

表 4. 两个信号的分类结果

信号	类别
1	第一类
2	第五类

注: 信号标号分别为两个信号在第二组数据中的列号。

Table 5. Reliability of recognition

表 5. 识别的可靠性

k	大于标准值 k_{\max}^c 百分比
2.9654	48.27%
2.9416	47.08%

应用小波阈值降噪方法对第三组数据中给出的信号进行降噪处理, 得到 4 个降噪信号, 代入前述模型进行特征提取及识别分类, 从而求解该信号是否属于问题一中得到的五种类别。

通过小波阈值降噪对信号进行降噪处理的步骤如下:

第一步: 对信号进行小波分析

信号 $x(t)$ 的小波变换公式为:

$$\varphi_{a,b}(t) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t) \left(\frac{t-b}{a} \right) dt.$$

定义尺度子空间 V_j 和小波子空间 W_j , $\phi \in V_0 \subset V_1$. 小波基函数 $\psi(t)$ 可表示为:

$$\psi(t) = \sqrt{2} \sum_{k \in \mathbb{Z}} g_k \phi(2t-k)$$

其中, ϕ 为相应于多分辨率分析 $\{V_j\} \in \mathbb{Z}$ 的尺度函数。小波分解公式为

$$f_k^{j+1} = \sum_{l \in \mathbb{Z}} \bar{h}_{l-2k} f_l^j$$

$$d_k^{j+1} = \sum_{l \in \mathbb{Z}} \bar{g}_{l-2k} f_l^j$$

经过上述正交分解步骤, 原信号被分解为尺度函数和小波函数, 对小波函数选取阈值进行处理, 可去除噪声。

第二步: 选取阈值函数

阈值函数决定如何对超过和低于阈值的小波系数的模进行处理, 即对含有噪声的波段进行处理的方法。由于软阈值处理后得到的结果较平滑, 本文中选取软阈值函数进行去噪。其函数表达式如下:

$$T_s(Y, \lambda) = \begin{cases} 0, & |Y| < \lambda \\ Y - \lambda \text{sign}(Y), & |Y| \geq \lambda \end{cases},$$

其中, Y 是信号 x 的小波系数, λ 为阈值, $T_s(Y, \lambda)$ 表示软阈值化后的小波系数。

第三步: 确定阈值

阈值的选取是否适宜直接关系到去噪效果, 若阈值过大, 会去除有用信息, 阈值过小则去噪不完全。阈值选取的方法很多, 本文中选取规则较保守的 SURE 阈值, 以防有用信息被去除。SURE 阈值的数学表达如下:

$$T_{SURE} = \arg \min_{0 \leq t \leq t_{\min}} \left[\sigma^2 N + \sum_{j=0}^{N-1} \{w_j \Lambda t^2\} - 2\sigma^2 \sum_{j=0}^{N-1} I\{|w_j| \leq t\} \right],$$

其中, σ 为噪声方差, w_j 为第 j 个小波系数, Λ 为两数取小, I 是示性函数, N 是原信号经小波分解后得到的 N 个小波系数。

第四步: 小波的重构

完成小波系数的处理后, 信号的噪声已去除, 此时需要把去噪后的小波系数与尺度函数重新组合成一个信号, 即得到去噪后的信号。

对任意整数 j, k , 由系数序列 $\{f_l^{j+1}\}, \{d_l^{j+1}\}$ 表示系数, 则重构公式为:

$$f_k^j = \sum_{l \in \mathbb{Z}} \bar{h}_{k-2l} f_l^{j+1} + \sum_{l \in \mathbb{Z}} (-1)^k \bar{h}_{2l-k+1} d_l^{j+1} = \sum_{l \in \mathbb{Z}} \bar{h}_{k-2l} f_l^{j+1} + \sum_{l \in \mathbb{Z}} \bar{g}_{k-2l} d_l^{j+1}$$

得到去噪后的信号。对于第三组数据中第 1 列的信号降噪后的信号图如下(图 3)。

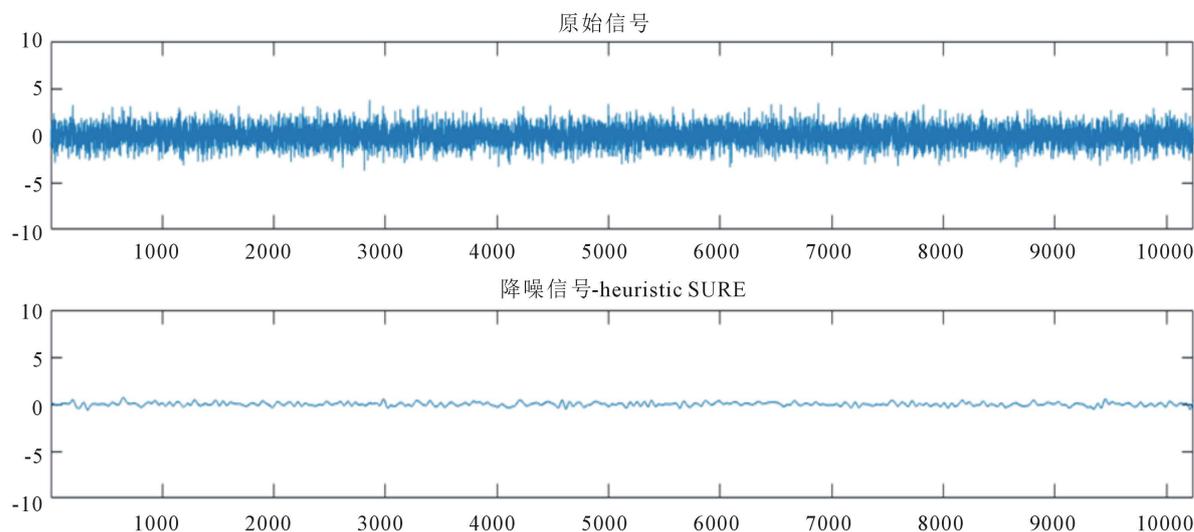


Figure 3. Noise reduction example for the first signal

图 3. 信号一的降噪示例

Table 6. Classification results for the third group of data

表 6. 第三组数据的分类结果

信号	类别
1	第二类
2	第三类
3	第二类
4	第五类

注：信号标号分别为四个信号在第三组数据中的列号。

第五步：将降噪信号代入上述模型，进行分类

第三组数据中的 4 个信号的分类结果如下表 6 所示。

5. 改进与推广

对于小波阈值降噪模型，小波分析中可选用的函数太多，本文只通过阅读文献确定合适问题三中信号特点的函数，因此本模型中在阈值确定及阈值函数步骤中采取的函数不能保证使附件三中所给信号在小波分析降噪的情况下效果达到最好，可多进行不同函数组合的搭配，从而增强降噪效果。

本模型着重解决通信辐射源信号特征的提取和分析，实现从一般通信信号的分类识别到个体信号的识别，从而确定信号来源及位置，在军事方面具有重要的现实意义。

一般噪声类型为加性白噪声，该噪声假设功率谱密度具有无限宽的带宽，是一种便于进行数学分析的理想化模型，应考虑在实际传输中形成的噪声的各种类型，由于模型中使用的方法均从时频方面考虑，有较强的稳定性，可尝试使用该模型，使其对其他类型平稳或非平稳噪声进行识别分类。

参考文献

- [1] 徐超. 基于杂散特征的辐射源个体识别研究[D]: [硕士学位论文]. 西安: 西安电子科技大学, 2014.

- [2] 利稷夫. 无监督聚类算法在辐射源信号分析中的应用[D]: [硕士学位论文].成都: 西南交通大学, 2010.
- [3] 高新波, 李洁, 谢维信. 模糊 C 均值聚类算法中参数 m 的优选[J]. 模式识别与人工智能, 2000, 13(1): 7-11.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2324-7991, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>
期刊邮箱: aam@hanspub.org