

基于 IMF 能量熵的目标特征提取与分类方法

张小蓓,张 歆,孙进才

ZHANG Xiao-ji,ZHANG Xin,SUN Jin-cai

西北工业大学 航海学院,西安 710072

College of Marine Engineering,Northwestern Polytechnical University,Xi'an 710072,China

E-mail:nwpuzxj@nwpu.edu.cn

ZHANG Xiao-ji,ZHANG Xin,SUN Jin-cai.Feature extraction and classification experiment based on energy entropy of IMF's.Computer Engineering and Applications,2008,44(4):68-69.

Abstract: A new feature extraction and selection method based on the energy entropy of intrinsic mode functions(IMF's) is presented.Three types of noise signals radiated from the targets are decomposed into their respective IMF's using the Empirical Mode Decomposition procedure,and the energy of the same IMF of three types of signals are different.The energy entropies of the IMF's are calculated by Eq (5) or Eq (6).K-neighbor classifier is used for classification experiments for three types of signals.The results show that the correct identification ratio of experiments based on esq.(6) is above 80%.

Key words: Empirical Mode Decomposition(EMD);Intrinsic Mode Function(IMF);feature extraction;K-neighbor classifier

摘 要:提出了一种基于固有模态函数(IMF)能量熵的特征提取与选择方法。对三类信号进行了经验模态分解(EMD),得到 IMF。对于不同类别的信号,同阶的 IMF 能量有明显的不同。选择 IMF 能量作为特征向量,并选判别熵作为分类判据,同时给出了两种能量熵的计算公式。采用 K-近邻分类器对三类信号进行了分类试验,试验结果表明,基于最佳特征向量选择的分类试验的平均正确识别率达 80%以上。

关键词:经验模态分解;固有模态函数;特征提取;K-近邻分类

文章编号:1002-8331(2008)04-0068-02 **文献标识码:**A **中图分类号:**TN911.7

1 引言

目标特征提取与分类是信号处理系统中的主要环节。由于大多数实际信号是非平稳的随机信号。使得针对非平稳信号处理的时频分析方法得到了广泛应用。Hilbert-Huang 变换是最新发展起来的处理非线性、非平稳信号的时频分析方法^[1-4]。它包括经验模态分解和 Hilbert 变换两步。通过对信号的经验模态分解(EMD)获得一系列表征信号特征时间尺度的固有模态函数(IMF),然后利用 Hilbert 变换和瞬时频率求解获得信号的时频 Hilbert 谱。

由于 EMD 是根据信号本身的尺度特征来对信号进行分解,信号的非线性和非平稳特征可以在各阶 IMF 分量中充分保留下来,因此,本文将研究基于 IMF 分量的特征提取方法,并通过分类试验结果,检验该方法的正确性和有效性。

论文首先分析目标噪声信号的 EMD 分解和 IMF 分量的特点,然后介绍基于 IMF 分量的能量的特征提取与选择方法,最后给出了对实际信号的分类结果及其分析。

2 基于 IMF 的特征提取与选择

考虑三类目标信号 a、b、c,其样本数据的频谱如图 1 所示(数据长度 500)。

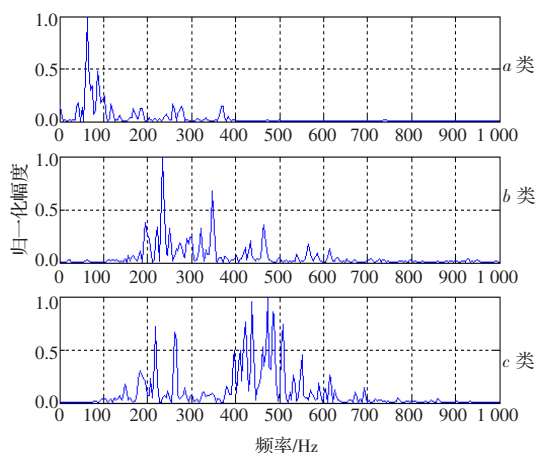


图 1 三类信号的频谱

按照 IMF 的定义,对三类信号进行 N 阶经验模态分解^[2-4],获得 N 个 IMF 分量和剩余信号(趋势项)。原始信号可以表示为 IMF 分量和剩余信号的和,即

$$x(t) = \sum_{i=1}^N imf_i(t) + r(t) \quad (1)$$

式中, $imf_i, i=1, 2, \dots, N$, 表示 IMF 分量, $r(t)$ 为剩余信号。对图 1

作者简介:张小蓓(1958-),男,副教授,硕士,主要研究方向为信息信号检测与处理技术;张歆(1963-),女,副教授,博士,主要研究方向为通信信号检测与处理技术;孙进才(1937-),男,教授,博导,主要研究方向为阵列信号处理。

收稿日期:2007-05-29 **修回日期:**2007-08-06

所示的三类信号的样本经 $N=8$ 阶分解后得到的前 4 阶 IMF 分量信号如图 2 所示。

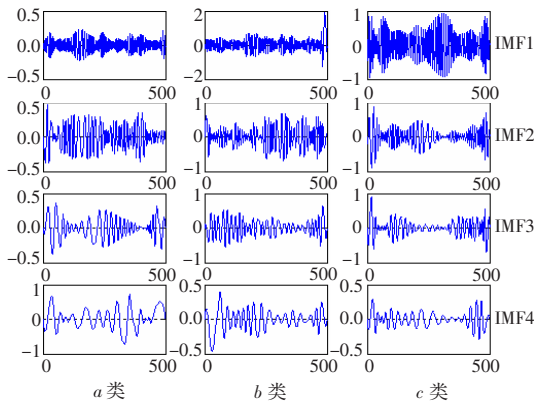


图 2 三类信号的前 4 阶 IMF 分量

由图 2 可见,每阶 IMF 分量含有不同频率成分,IMF 的阶数愈低,其所含高频成份愈多。在同一阶 IMF 分量上,不同类别的信号有着不同的频率成分,而且 IMF 的幅度相差很大。因此,可以考虑采用 IMF 的能量作为特征进行分类。

为了便于特征提取,选择 IMF 能量的判别熵来表征不同类别 IMF 特征的差别。

设以 IMF 能量为特征的特征向量表示为: $X^{(i)}=(x_1^{(i)}, x_2^{(i)}, \dots, x_N^{(i)})$, $i=1, 2, \dots, c$, c 为类别数, N 为特征向量的维数。设第 l 个特征向量的均值为 $\mu_l^{(i)}$, $l=1, 2, \dots, N$, 归一化均值 $P_{el}^{(i)}$ 表示为

$$P_{el}^{(i)} = \frac{\mu_l^{(i)}}{\sum_{i=1}^c \mu_l^{(i)}} \quad (2)$$

对所有的 $i=1, 2, \dots, c$, 有 $\sum p_{el}^{(i)}=1$ 。对于任意两类 s_i 和 s_j , $i, j=1, 2, \dots, c$, 第 l 个分量的判别熵定义为:

$$W_l(s_i, s_j) = V(p_{el}^{(i)}, p_{el}^{(j)}) + V(p_{el}^{(j)}, p_{el}^{(i)}) \quad (3)$$

式中, $V(p_{el}^{(i)}, p_{el}^{(j)})$ 表示相对熵, 定义为

$$V(p_{el}^{(i)}, p_{el}^{(j)}) = -\sum p_{el}^{(i)} \log[p_{el}^{(i)} p_{el}^{(j)}] \leq 0 \quad (4)$$

可以选择所有尺度的 IMF 能量的判别熵之和

$$J_c = \sum_{l=1}^c W_l(s_1, s_2) \quad (5)$$

作为类别可分性的熵度量。但在分类试验时发现,由于 EMD 将数据向量分解成各个不同频率和幅度的细节分量,它对数据中细微的频率分量或幅度的差异都能体现出来。因此,用式(5)计算同类信号不同样本之间的判别熵时,也可能出现较小的值,导致分类能力下降。

为此,考虑选择部分 IMF 能量组成特征向量,即选择那些在不同类别的样本之间有较小判别熵,同时在同类的样本之间有较大判别熵的 IMF 能量分量进行分类器设计。

例如,当分解阶数 $N=8$ 时,第 1、2、4 阶 IMF 分量在同类样本之间有较大的判别熵,而在不同类别样本之间的判别熵较小。由此可选

$$J_c = W_1 + W_2 + W_4 \quad (6)$$

作为熵度量,式中,下标表示 IMF 分量的阶数,即选择第 1、2、4 阶的 IMF 能量的判别熵进行分类器的设计。

3 基于 IMF 能量熵的分类试验

基于 IMF 能量的特征提取及分类过程的流程图如图 3 所示。输入信号首先进行时域预处理,如对信号幅度进行归一化、去直流分量等。然后对信号进行经验模态分解,得到 IMF 分量。计算 IMF 分量的判别熵,并从中选择特征向量,进行分类器设计。对各类信号的样本进行分类试验,最后得到分类结果。

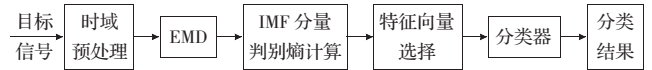


图 3 基于 IMF 能量的特征提取及识别过程

用 a, b, c 类信号进行分类试验。每类信号有 101 个样本,每个样本的数据长度为 500。从 101 个样本中随机选出 30 个作为标准样本,再从剩余的 71 个样本中随机抽取 45 个进行测试。试验时,对样本数据进行 $N=8$ 阶经验模态分解,按照式(5)或式(6)计算样本之间的判别熵矩阵。采用 K -近邻分类器, K 设计为 9。

对标准样本和测试样本进行分类试验,采用式(5)计算判别熵。对测试样本进行多次分类试验,得到的平均分类结果如表 1 所示。采用式(6)计算判别熵,进行分类试验,得到分类结果如表 2 所示。

表 1 标准样本和测试样本的分类试验结果

样本集	标准样本			测试样本		
	a 类	b 类	c 类	a 类	b 类	c 类
平均识别率/%	86.67	86.67	93.33	88.00	85.56	80.56

表 2 采用式(6)计算判别熵得到的分类结果

类别	测试样本数	分为该类的数目			平均正确识别率/%	平均识别率/%
		a 类	b 类	c 类		
a 类	45	38	7	0	84.44	
b 类	45	0	41	4	91.11	89.63
c 类	45	1	42	2	93.33	

对比表 1 和表 2,选择式(6)计算判别熵,对 b 类和 c 类信号的识别率有显著提高,但对 a 类识别率反而下降。通过频谱分析发现, a 类各样本信号的频谱之间有较大差异。这导致由式(6)计算的熵度量没能取得 a 类信号的最佳特征向量。

4 小结

本文对基于 IMF 能量的特征提取方法及分类试验进行了介绍。从分类试验的过程可以看出,基于 IMF 能量的特征提取方法算法非常简单,计算量小,对分解过程、试验样本的选取都无特殊要求,试验中的标准样本和测试样本都是随机选取的。从分类试验的结果看,该方法对实际的三类信号的分类有良好的效果。

参考文献:

- [1] Huang N E. The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis[C]//Proc R Soc, Lond, 1998, 454(12): 903-995.
- [2] 张小莉,张歆,孙进才.基于经验模态分解的目标特征提取与选择[J].西北工业大学学报, 2006, 24(4): 453-456.
- [3] 谭善文,秦树人,汤宝平. Hilbert-Huang 变换的滤波特性及其应用[J].重庆大学学报, 2004, 27(2): 9-12.
- [4] 钟佑明,秦树人. 希尔伯特-黄变换的统一理论依据研究[J]. 振动与冲击, 2006, 25(3): 40-44.