

基于类别距离和 Bhattacharyya 距离的 雷达信号特征评价^{*}

邓延丽, 金炜东, 余志斌

(西南交通大学 电气工程学院, 成都 610031)

摘要: 从瞬时自相关法、小波变换法和时频原子法提取的脉内特征比较分析入手, 提出了一种新的最优特征评价准则。以类内距离、类间距离和 Bhattacharyya 距离为基础, 从特征的空间分布和错误识别率的上界等方面对不同算法提取出的脉内特征进行分析, 实现了最优特征的选择。实验的仿真结果表明, 这种最优特征评价准则是有效的, 为雷达辐射源信号的特征评价提供了有意义的参考。

关键词: 类内距离; 类间距离; Bhattacharyya 距离; 特征评价

中图分类号: TN20

文献标志码: A

文章编号: 1001-3695(2012)11-4079-03

doi:10.3969/j.issn.1001-3695.2012.11.019

Feature evaluation of radar signal based on category distance and Bhattacharyya distance

DENG Yan-li, JIN Wei-dong, YU Zhi-bin

(School of Electrical Engineering, Southwest Jiaotong University, Chengdu 610031, China)

Abstract: This paper presented a novel evaluation criterion of optimal feature by means of comparative analysis of the instantaneous autocorrelation method, wavelet transform and time-frequency atom method. It chosen the features extracted in different algorithms by the distance between classes, within-class distance and Bhattacharyya distance with respect to the feature distribution and upper bound of the error rate. The simulation results show that the evaluation criterion of optimal feature is effective. It provides a valuable reference for the feature evaluation of the radar emitter signal.

Key words: within-class distance; the distance between classes; Bhattacharyya distance; feature evaluation

0 引言

面对日益密集、复杂的雷达信号环境, 信号波形的复杂多变以及频率的覆盖范围很宽, 利用传统的 5 参数(到达时间、载频、到达角、脉宽及幅度)进行分选和识别雷达信号已经变得非常困难。随着新体制雷达的出现, 脉内特征分析技术受到普遍的关注^[1~8], 谱相关法^[1]、包络法^[2,3]、相对无模糊相位重构法^[4]、小波变换^[5]、时域瞬时自相关法^[6,7]及时频原子法^[8]等脉内特征提取法相继被提出, 在一定条件下实现了不同调制类型的雷达辐射源识别。

面对当前如此多的雷达辐射源识别参数, 如何选择最优的特征参数是目前最为关注的研究问题。一个直观的想法是把分类的错误率作为可分性测度, 使分类器错误率最小的那组特征应该是最好的特征。但错误率的计算非常复杂且实际问题中不易计算, 这使得直接用错误概率作为分类标准来分析特征的有效性十分困难。因此, 找出实用的分类标准是一种有效的解决方案^[9~11]。在保证分类精度的情况下, 如何从如此多的脉内特征提取算法中选出最佳特征, 则是本文的研究内容。基于文献[12~14]提取的特征, 本文提出了一种新的最优特

征评价准则。该准则利用类内距离、类间距离和 Bhattacharyya 距离可分性测度, 实现了从多种脉内特征提取算法中选出不同调制类型雷达信号的最佳特征。实验结果表明了本文方法的可行性和有效性。

1 特征评价准则

就特征空间分布而言, 希望同一类特征分布越集中越好, 通常利用类内距离表征; 不同类的特征分布越分散越好, 采用类间距离表征。将脉内特征提取算法中第 i 类不同调制类型的雷达信号的特征类内距离 C_{ii} 定义为

$$C_{ii} = \max \left\{ \left[\frac{1}{M_i^q} \sum_{k=1}^{M_i^q} \|x_{ik}^q - E(X_i^q)\|^p \right]^{\frac{1}{p}} \right\} \quad (1)$$

其中: $q=1, 2, \dots, N$, N 是不同调制类型的雷达信号的特征数; M_i^q 是脉内特征提取算法中第 i 类不同调制类型的雷达信号的特征第 q 中特征样本数; x_{ik}^q 是脉内特征提取算法第 i 类不同调制类型的雷达信号的特征第 q 中的第 k 个样本值; $X_i^q = [X_{i1}^q, X_{i2}^q, \dots, X_{iM_i^q}^q]$, $E(X_i^q)$ 是 X_i^q 的期望值; p 是大于 1 的整数。于是, 可得到脉内特征提取算法中的第 j 类不同调制类型的雷达信号的特征类内距离 C_{jj} 。

收稿日期: 2012-04-11; 修回日期: 2012-05-16 基金项目: 国家自然科学基金资助项目(60971103)

作者简介: 邓延丽(1985-), 女, 四川人, 博士研究生, 主要研究方向为雷达辐射源信号处理(dyllovebeijing@163.com); 金炜东(1959-), 男, 安徽人, 教授, 博导, 博士, 主要研究方向为智能信息处理、系统仿真、优化理论与优化控制等; 余志斌(1977-), 男, 湖南长沙人, 博士, 主要研究方向为智能信息处理、雷达信号处理等。

第 i 类和第 j 类不同调制类型的雷达信号特征类间距离可定义为

$$D_{ij} = \min \| E(X_i^q) - E(X_j^q) \| \quad (2)$$

这样,第 i 类与第 j 类不同调制类型的雷达信号的特征分离度可定义为

$$S_{ij} = C_{ii} + C_{jj} - D_{ij} \quad (3)$$

在同一特征空间分布下,若 $S_{ij} > 0$,表明得到的第 i 类与第 j 类不同调制类型的雷达信号特征数据有重叠部分, S_{ij} 越大,说明得到不同调制类型的雷达信号的特征数据重叠越严重;若 $S_{ij} < 0$,说明得到的第 i 类与第 j 类不同调制类型的雷达信号特征数据无重叠部分,即特征数据完全可以分离;若 $S_{ij} = 0$,说明得到的第 i 类与第 j 类不同调制类型的雷达信号特征数据无重叠部分,即特征数据恰好完全可以分离。因此,要从多种脉内特征提取算法中选出不同调制类型的雷达信号的最佳特征,使得 $S_{ij} < 0$,若 S_{ij} 越小,说明两分类特征相隔越远。

2 Bhattacharyya 距离

对于模式分类来说,最理想的准则应是 Bayes 分类误差最小准则,但由于通常情况下数据的概率分布难以确定,导致 Bayes 分类误差难以计算和分析,所以几乎没有采用 Bayes 误差作为准则函数的特征提取方法。

而上述类内距离和类间距离只是反映了不同调制类型的雷达信号特征的空间分布情况,计算方便,但它们没有与分类的错误率有直接的联系,而类别间 Bhattacharyya 距离决定了分类误差率的上界^[9]。所以,在保证分类精度的情况下,为进一步从多种脉内特征提取算法中得到不同调制类型的雷达信号的最佳特征,结合了类别间 Bhattacharyya 距离(简称 B 距离)。

B 距离的定义是:

$$J_B = -\ln \int p(x|\omega_1)p(x|\omega_2) \frac{1}{2} dx \quad (4)$$

它与错误概率的上界有直接关系。因为

$$\begin{aligned} P_e &= P(\omega_1) \int_{\mathfrak{R}} p(x|\omega_1) dx + P(\omega_2) \int_{\mathfrak{R}} p(x|\omega_2) dx = \\ &\int_{-\infty}^{+\infty} \min \{ P(\omega_1)p(x|\omega_1), P(\omega_2)p(x|\omega_2) \} dx \leq \\ &\int_{-\infty}^{+\infty} \{ P(\omega_1)P(\omega_2)p(x|\omega_1)p(x|\omega_2) \}^{\frac{1}{2}} dx = \\ &[P(\omega_1)P(\omega_2)]^{\frac{1}{2}} \int_{-\infty}^{+\infty} p(x|\omega_1)p(x|\omega_2) \frac{1}{2} dx = \\ &[P(\omega_1)P(\omega_2)]^{\frac{1}{2}} \exp \{ -J_B \} \end{aligned} \quad (5)$$

其中: P_e 为错误概率; $\mathfrak{R}$ 为 $P(x|\omega_2) > P(x|\omega_1)$ 的区域,而 $\mathfrak{R}$ 为 $p(x|\omega_1) > p(x|\omega_2)$ 的区域。

对于一维空间中的数据 X ,其 B 距离定义^[9]为

$$J_B = \frac{1}{4} \frac{(\mu_2 - \mu_1)^2}{\sigma_1^2 + \sigma_2^2} + \frac{1}{2} \ln \left(\frac{1}{2} \frac{\sigma_1^2 + \sigma_2^2}{\sigma_1 \cdot \sigma_2} \right) \quad (6)$$

其中: μ_i 和 σ_i ($i=1,2$) 是 ω_i 类别的均值和方差。

由于 B 距离与分类器的错误率上界有关, J_B 的值越大,错误率上界越小,更有利于分类。

3 实验仿真

采用瞬时自相关脉内特征提取算法^[12](以下简称方法 1),从五种(BPSK、QPSK、BFSK、LFM 和 CW)不同调制类型的雷达信号中提取出脉内特征参数。其中相关参数设置如下:采样频率为 100 MHz,载频为 30 MHz,采用 7 位 Barker 码的 BPSK 和采用 16 位 Frank 码的 QPSK 脉宽分别为 1.714 μ s 和

0.5 μ s, LFM 脉宽为 16 μ s,带宽为 5 MHz,采用编码为 7 位 Barker 码的 FSK 脉宽为 1 μ s,终止频率为 28 MHz, CW 的脉冲持续时间为 16 μ s。采用小波变换脉内特征提取算法^[13](以下简称方法 2),从五种(BPSK、QPSK、BFSK、LFM 和 CW)不同调制类型的雷达信号中提取出脉内特征参数,相关参数设置同上。采用时频原子脉内特征提取算法^[14](以下简称方法 3),从五种(BPSK、BFSK、LFM、CW 和 NLFM)不同调制类型的雷达信号中提取出脉内特征参数。相关参数设置如下:采样频率为 100 MHz,载频为 10 MHz,采用 7 位 Barker 码的 BPSK 脉宽为 1 μ s, LFM 脉宽为 10 μ s,带宽为 10 MHz,采用编码为 7 位 Barker 码的 FSK 脉宽为 1 μ s,终止频率为 25 MHz, CW 的脉冲宽度为 10 μ s, NLFM 采用正弦调频,调频率为 1 MHz。从文献[12~14]中分别取出一个特征参数,每种不同调制类型的雷达信号特征样本数为 40,从方法 1~3 中提取特征在特定信噪比下分别为 200。在信噪比为 2~10 dB 下,三种方法的不同调制类型的雷达信号特征总数为 3 000。分别采用式(1)和(2)在信噪比为 2~10 dB 每隔 2 dB 下计算的类内距离和类间距离如图 1~5 所示。

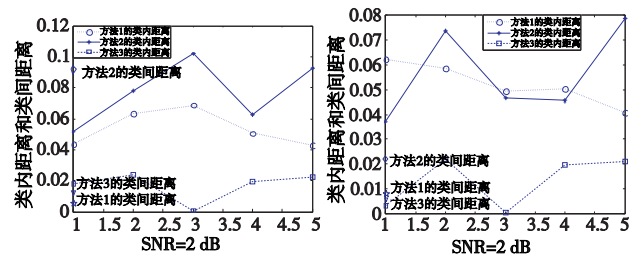


图1 五种调制类型的雷达信号特征在SNR=2 dB下的类内距离和类间距离

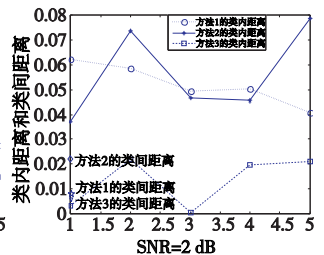


图2 五种调制类型的雷达信号特征在SNR=4 dB下的类内距离和类间距离

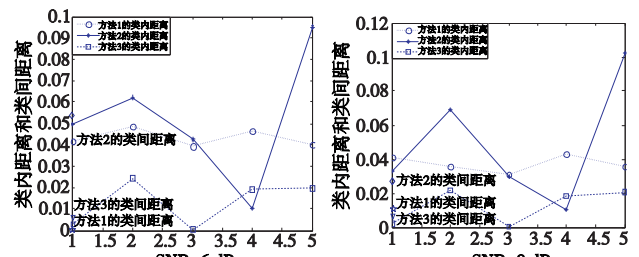


图3 五种调制类型的雷达信号特征在SNR=6 dB下的类内距离和类间距离

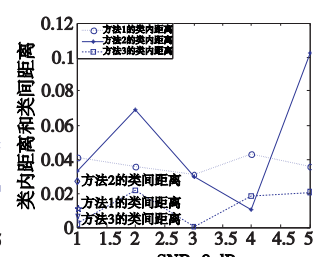


图4 五种调制类型的雷达信号特征在SNR=8 dB下的类内距离和类间距离

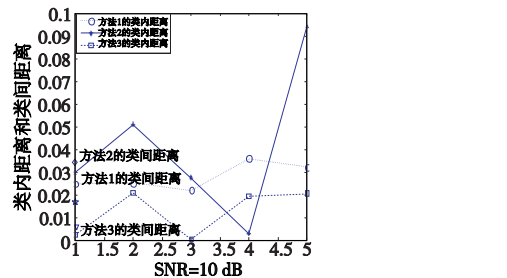


图5 五种调制类型的雷达信号特征在SNR=10 dB下的类内距离和类间距离

从图 1~5 中可知,在保证分类精度的情况下,信噪比为 2 dB 和 4 dB 时,方法 3 提取五种调制类型雷达信号的特征类内距离最小,而方法 2 提取的五种调制类型雷达信号的特征类间距离最大;信噪比为 6~10 dB 时,除了第 4 点(即上述列出的第四种调制类型的雷达信号)雷达信号的特征外,方法 3 提取其余四种调制类型雷达信号的特征类内距离最小,而方法 2 提取的五种调制类型雷达信号的特征类间距离最大。众所周知,

类内距离越小,类间距离越大,说明分类特征越好。方法 3 的类内距离小,而方法 2 的类间距离大,究竟哪种方法提出的脉内特征更好?通过综合分析,计算特征的可分离度值式(3)如表 1 所示。

表 1 方法 2 和 3 的特征分离度值

方法	SNR/dB				
	2	4	6	8	10
2	0.059 5	0.094 9	0.092 3	0.0353 8	0.020 8
3	0.007 6	-0.008	-0.002 8	-0.003 0	-0.002 4

从表 1 中可以看出,信噪比为 2 dB 时,方法 2 和 3 的可分离度值均为正数,说明得到的不同调制类型的雷达辐射源信号特征数据有重叠部分,方法 2 的特征数据分布的重叠部分比方法 3 的重叠部分更为严重;信噪比为 4 dB~10 dB 时,方法 2 的可分离度值都是正数,并依次在减小,说明特征数据分布仍有重叠部分,但随着信噪比的提高,重叠部分在减小。方法 3 的可分离度值都是负数,说明特征数据分布没有重叠部分。

综上所述,从特征空间的分布情况看,方法 3 的分类特征优于方法 2 的分类特征。

为了进一步验证方法 3 的分类特征是否优于方法 2 的分类特征,从错误率的上界角度出发,结合用 B 距离来验证。

在 MATLAB 中利用函数 `lillietest` 对特征数据进行正态分布的检验。函数 `lillietest` 的格式为

$$[H,P,LSTAT,CV] = \text{lillietest}(X)$$

其中: X 为输入向量; H 为测试结果。若 $H=0$, 则认为 X 是服从正态分布的;若 $H=1$, 则否认 X 服从正态分布。 P 为接受假设的概率值, P 越接近于 0, 则可拒绝是正态分布的原假设。 $LSTAT$ 为测试统计量的值, CV 为是否拒绝原假设的临界值。

经检验知,绝大部分特征数据测试的 $H=0$, 即服从正态分布,少数部分数据测的 $H=1$ 。为了使数据满足正态分布,剔除个别奇异点,最终使得数据服从正态分布或近似正态分布。表 2 表征了三种方法提取出不同调制类型雷达信号的特征 B 距离。

表 2 三种方法的特征 B 距离

方法	SNR/dB					
	0	2	4	6	8	10
1	25.43	56.2	108.3	274.43	1 028	3 634.5
2	1 526.1	4 576	22 581	44 352	198 300	267 100
3	93 357	7 300	70 664	163 170	304 460	380 760

从表 2 中可以看出,方法 3 的特征 B 距离最大,则分类错误率的上界最小;方法 1 的特征 B 距离最小,分类错误率的上界最大。如图 6,表征了三种方法的特征归一化 B 距离随信噪比的变化趋势。从图 6 中可以看出,随着信噪比的提高,三种方法的特征 B 距离均依次增大,说明分类错误率的上界越来越小,即分类错误率越来越低。方法 3 的特征 B 距离均远大于其他两种方法的特征 B 距离。图 7 表征了三种方法的特征归一化 B 距离的错误率上界随信噪比的变化趋势。从图 7 中可以看出,随着信噪比的提高,三种方法的分类错误率逐渐降低,且方法 3 提取出的特征分类错误率远低于其他两种方法提取出的特征分类错误率。

综上所述,从特征空间分布角度和错误率的上界角度出发,在满足分类精度的情况下,方法 3 所提取出的不同调制类型的雷达信号特征要优于方法 1 和 2。

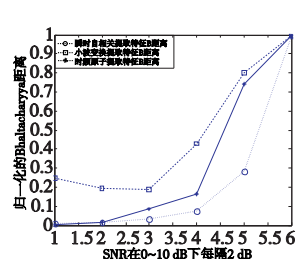


图 6 三种方法提取特征的归一化B距离

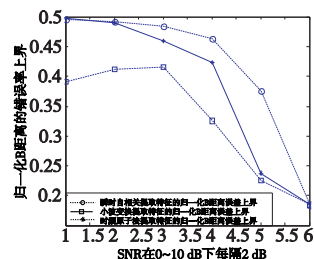


图 7 三种方法提取特征的归一化B距离的错误率上界

4 结束语

通过特征的空间分布和错误识别率的上界等方面对不同算法提取出的脉内特征进行了比较和分析。从仿真结果的分析可以看出,采用时频原子法提取的脉内特征在空间分布和正确识别率上均优于瞬时自相法提取的脉内特征和小波变换法提取的脉内特征。本文在计算 B 距离时,由于少数部分的特征数据不服从正态分布,剔除了个别奇异点,使得数据近似地服从正态分布,可能影响 B 距离有所误差。但从实验结果分析可以看出,并不影响最优特征的选择。通过实验仿真结果证实了该最优特征评价准则的有效性和可行性,这为雷达辐射源信号的特征评价提供了有意义的参考。

参考文献:

- [1] 张庆荣, 单佩均. 雷达信号脉内特征分析的谱相关法[J]. 电子对抗, 1993, 19(4): 1-6.
- [2] LEE J P Y. A multi-channel digital receiver: intra-pulse analysis and direction-finding[C]//Proc of IEEE Pacific Rim Conference on Communications, Computers and Signal Processing. 1999: 589-592.
- [3] ZHOU Y F, LEE J P Y. A MDL approach for determining the number of emitters using intra-pulse information[C]//Proc of IEEE Pacific Rim Conference on Communications, Computers and Signal Processing. 1999: 548-551.
- [4] 黄知涛, 周一宇, 姜文利. 基于相对无模糊相位重构的自动脉内调制特性分析[J]. 通信学报, 2003, 24(4): 153-160.
- [5] MALLAT S. A wavelet tour of signal processing[M]. San Diego, CA: Academic, 1998.
- [6] 毕大平, 董晖, 姜秋喜. 基于瞬时频率的脉内调制识别技术[J]. 电子对抗技术, 2005, 20(2): 6-9.
- [7] 普运伟, 金炜东, 胡来招. 基于瞬时频率二次特征提取的辐射源信号分类[J]. 西南交通大学学报, 2007, 42(3): 373-379.
- [8] 朱明, 普运伟, 金炜东, 等. 基于时频原子方法的雷达辐射源信号特征提取[J]. 电波科学学报, 2007, 22(3): 458-462.
- [9] BIAN Zhao-qi, ZHANG Xue-gong. Pattern recognition[M]. Beijing: Qinghua University Press, 2000.
- [10] ZHAO Jun, WANG Guo-yin, WU Zhong-fu, et al. The study on technologies for feature selection[C]//Proc of the 1st International Conference on Machine Learning and Cybernetics. 2002: 689-693.
- [11] MOLINA L C, BELANCHE L, NEBOT A. Feature selection algorithms: a survey and experimental evaluation[C]//Proc of International Conference on Data Mining. 2002: 306-313.
- [12] 普运伟. 复杂体制雷达辐射源信号分选模型与算法研究[D]. 成都: 西南交通大学, 2007.
- [13] 余志斌. 基于脉内特征的雷达辐射源信号识别研究[D]. 成都: 西南交通大学, 2010.
- [14] 朱明. 复杂体制雷达辐射源信号时频原子特征研究[D]. 成都: 西南交通大学, 2008.
- [15] SKRYPNYK I. Irrelevant features, class separability, and complexity of classification problems[C]//Proc of the 23rd IEEE International Conference on Tools with Artificial Intelligence. 2011: 998-1003.