

基于多尺度信息熵的雷达辐射源信号识别

黄颖坤 金炜东* 葛鹏 李冰

(西南交通大学电气工程学院 成都 610031)

摘要: 随着雷达信号的日益复杂,从实数序列中提取特征变得越来越困难,但当它们表示成符号序列时,通常能更容易地挖掘出有效的特征参数。因此,该文提出一种基于多尺度信息熵(MSIE)的雷达信号识别方法。首先通过符号聚合近似(SAX)算法在不同字符集尺度下将雷达信号转换为符号化序列;然后联合各符号序列的信息熵值,组成MSIE特征向量;最后,使用 k 邻近算法(k -NN)作为分类器实现雷达信号的分类识别。通过仿真6种典型的雷达信号进行验证,结果表明该方法在信噪比(SNR)为5 dB时,不同雷达信号的识别正确率大于90%,并且优于传统的基于复杂度特征(盒维数和稀疏性)的识别方法。

关键词: 雷达信号识别;符号聚合近似算法;多尺度信息熵; k 邻近算法

中图分类号: TN95

文献标识码: A

文章编号: 1009-5896(2019)05-1084-08

DOI: 10.11999/JEIT180535

Radar Emitter Signal Identification Based on Multi-scale Information Entropy

HUANG Yingkun JIN Weidong GE Peng LI Bing

(College of Electrical Engineering, Southwest Jiao Tong University, Chengdu 610031, China)

Abstract: With the increasing complexity of radar signals, it is more and more difficult to extract features of the real sequences, but when they are transformed to a symbol sequence, it is usually easier to mine the effective feature parameters. Therefore, a radar signal recognition method based on Multi-Scale Information Entropy (MSIE) is proposed. Firstly, the radar signal is transformed into symbolic sequence by Symbolic Aggregate appRoXimation (SAX) algorithm under different character number scales. Then, the information entropy of each symbol sequence is combined to form the MSIE feature vector. Finally, the k -Nearest Neighbor (k -NN) is used as a classifier to realize the classification and identification of radar signals. The simulation results of 6 typical radar signals show that using the proposed method the correct recognition rate of different radar signals is greater than 90% when Signal to Noise Ratio (SNR) is 5 dB, and better performance can be obtained compared with the traditional identification method based on complexity characteristics (box-dimension and sparseness).

Key words: Radar signal identification; Symbolic Aggregate appRoXimation (SAX) algorithm; Multi-Scale Information Entropy (MSIE); k -Nearest Neighbor (k -NN) algorithm

1 引言

雷达信号识别是雷达威胁告警(Radar Warning Receiver, RWR)、雷达支援侦察(Electronic Support Measurement, ESM)、雷达情报侦察(ELetronic INTelligence, ELINT)的重要内容,是雷达对抗首

先要解决的关键性问题。正确识别雷达信号可以帮助己方对目标雷达类型、载体、用途、威胁等级的识别,进而影响进一步的作战决策^[1]。

在过去十几年里,雷达信号识别受到了非常广泛的研究,纵观这些研究主要集中在筛选(提取)能有效表征辐射源信号特征,构建新的识别特征参数体系上。这些方法如果按处理对象的不同,可以粗略分成两条研究路线:一条是实数序列处理,另一条是符号序列的处理。尽管雷达信号的实数序列处理仍然是最主要的分析手段,并且得到非常多的研究成果^[2-8]。但是,随着雷达信号的日益复杂,一些尝试从符号序列分析的角度进来挖掘雷达数据隐

收稿日期: 2018-05-30; 改回日期: 2019-02-25; 网络出版: 2019-03-04

*通信作者: 金炜东 wdjin@home.swjtu.edu.cn

基金项目: 国家重点研发计划项目(2016YFB1200401-102F), 中央高校基本科研业务费专项资金(2682017CX046)

Foundation Items: The National Key Research and Development Program (2016YFB1200401-102F), The Fundamental Research Funds for the Central Universities (2682017CX046)

含信息的方法也受到关注^[9-11], 如文献[9]是在雷达信号符号化的基础上, 提取了反映序列随机程度的信息熵进行雷达信号识别, 并获得较好的性能; 文献[11]进一步研究了对于相同雷达信号, 不同调制参数的可分性。这些方法的都是在实数序列符号化的基础上进行特征提取, 但是, 所用的符号化方法过于简单, 容易损失信号的大量信息导致提取的特征存在偏差。所以, 选择一种重构误差较小的符号化方法是非常重要的。

纵观当前关于符号化方法的研究中, 符号化聚合近似(Symbolic Aggregate approXimation, SAX)方法被认为是一种简单且非常有效的实数序列符号化的方法^[12,13], 它可以有效地降低序列的维度, 同时能保留原始数据的大部分信息(重构误差小), 并且具有一定的降噪能力, 该方法已经在许多时间序列分析上得到应用。因此, 在本文所提雷达信号识别框架中, 首先使用SAX方法将雷达信号序列转化为符号化序列, 再对符号化序列进行统计学分析。在对符号化序列的统计分析研究上, 本文提出了一种基于信息理论的多尺度信息熵(Multi-Scale Information Entropy, MSIE)特征, 该特征是多个字符集尺度下提取的信息熵集合, 它反映了符号化序列在不同尺度上的复杂度特性。具体的提取方法将在第3节进行详细阐述。

本文的结构安排如下: 第2节详细介绍SAX符号化方法的理论与方法, 在此基础上, 第3节给出多尺度信息熵的定义和提取方法, 并且分析其抗噪性能; 然后, 在第4节对文中提出的雷达信号识别模型进行详细介绍。最后, 在第5节进行仿真实验的设计和分折。

2 实数序列符号化

面对越来越复杂的雷达信号数据, 一些传统的实数序列处理方法受到了一定的挑战, 因此, 一些学者尝试从另一种途径来处理这些复杂数据的分析问题, 即实数序列符号化。实数序列符号化具有下列两个优势^[14]:

(1) 序列的符号化是一个对原始数据“粗粒化”的过程, 它可以去除数据无关紧要的非本质特征, 留下数据关键的特征信息;

(2) 符号化序列在计算速度上具有很大的优势。

实数序列符号化的关键问题是如何选取合适的符号化策略, 但是目前并没有一个指导性的结论。目前, 普遍认为SAX方法是一种非常有效的符号化方法, 因此, 本文选用该方法对雷达信号序列进行符号化处理。下面对SAX方法进行详细描述。

2.1 分段聚集近似算法

SAX方法是实数序列符号化的一个经典算法,

是由文献[15]在分段聚集近似(Piecewise Aggregate Approximation, PAA)算法的基础上提出的。PAA算法是一种数据的重表达方法, 它能基本保留数据原有形状, 并且算法时间复杂度低, 因此, PAA算法被广泛应用在时间序列分析问题上。经典的PAA算法可以描述如下:

将一个长度为 n 的序列 $Q = \{q_i, i = 1, 2, \dots, n\}$ 映射到 m 维空间 $P = \{p_i, i = 1, 2, \dots, m\}$, 序列 P 的第 i 个元素可以通过式(1)计算

$$p_i = \frac{m}{n} \sum_{j=n/m(i-1)+1}^{n/mi} q_j, \quad 1 \leq i \leq m \quad (1)$$

式(1)可以简单描述为, 序列 Q 被分成 m 个大小相等的“段”, 然后计算每一个“段”的均值来构建新的序列 P 。从式(1)可以发现, n 需要满足 m 的整数倍。然而, 现实情况很难满足数据可分的情况, 所以, 当 n/m 不是整数时可以进行如下操作^[12]:

当 $1 < i < m$ 时

$$p_i = \frac{m}{n} \left((1 - c_{i-1}) q_{t(i-1)} + \sum_{j=t(i-1)+1}^{t(i-1)+b(i)} q_j + c_i q_{t(i-1)+b(i)+1} \right) \quad (2)$$

假设 $F(\bullet)$ 为对数值“ \bullet ”取整数部分的函数, 则式(2)中, $t(i) = t(i-1) + b(i-1) + 1$, 并且令 $t(1) = a + 1$, 其中 $a = F(n/m)$, 另外, $b(i) = F(n/m + c_{i-1} - 1)$ 。所以, 当 $i = 1$ 或 $i = m$ 时

$$p_i = \begin{cases} \frac{m}{n} \left(\sum_{j=1}^a q_j + c_1 q_{a+1} \right), & i = 1 \\ \frac{m}{n} \left(c_m q_{n-a} + \sum_{j=n-a+1}^n q_j \right), & i = m \end{cases} \quad (3)$$

其中, 序列 $\{c_i, i = 1, 2, \dots, m\}$ 是数据点的权值, 并且 $c_1 = c_m = G(n/m)$, $G(\bullet)$ 是对数值“ \bullet ”取小数部分的函数。同时, 式(2)和式(3)满足每一个“段”的数据权值之和等于 n/m 。为了更形象地描述上面的公式, 下面列举一个简单的示例: 假设要将序列 $\{k, k = 1, 2, \dots, 12\}$ 划分成包含5个数值的序列, 记为 $\{s_i, i = 1, 2, \dots, 5\}$, 因为 $G(12/5) = 2/5$, 所以, 当 $i = 1, i = 5$ 时, $s_1 = (1 + 2 + (2/5) \cdot 3) / (5/12) = 1.75$, $s_5 = ((2/5) \cdot 10 + 11 + 12) / (5/12) = 11.25$ 。当 $1 < i < 5$ 时, $s_2 = ((3/5) \cdot 3 + 4 + (4/5) \cdot 5) / (5/12) \approx 4.08$, $s_3 = ((1/5) \cdot 5 + 6 + 7 + (1/5) \cdot 8) / (5/12) = 6.5$, $s_4 = ((4/5) \cdot 8 + 9 + (3/5) \cdot 10) / (5/12) \approx 8.92$ 。

通过上面的计算，可以获得一个新序列 {1.75 4.08 6.5 8.92 11.25}。新序列的每一个元素都包含原始序列 $12/5 = 2.4$ 个数据点，这样就解决了序列不可分的问题。

2.2 SAX符号化方法

直接根据实数序列的数值特点进行符号划分是最常见的符号化方法，最常用的是静态划分方法，它需要先选定一个大小为 k 的字符集，假设为 $\{1 2 \dots k\}$ ，然后根据式 (4) 可以将时间序列 $Q = \{q_i, i = 1, 2, \dots, n\}$ 符号化为符号序列 $S = \{s_i, i = 1, 2, \dots, w\}$ ，其中 $w \ll n$ ，从而实现数据的降维。

$$s_i = \begin{cases} 1, & -\infty < q_i < \beta_1 \\ 2, & \beta_1 \leq q_i < \beta_2 \\ & \vdots \\ k, & \beta_k < q_i < +\infty \end{cases} \quad (4)$$

其中， $\{\beta_1 \beta_2 \dots \beta_k\}$ 是对原始序列 Q 进行空间划分的断点集合。实数序列符号化的关键在于如何确定这一断点集合。通行的方法有等区间法和等概率法，SAX采用等概率法获得不同的分区，从信息论的角度来说，等概率法有利于保留数据的有效信息。

SAX具体的算法可以简单描述为：在假设时间序列满足正态分布的条件下，(1)对时间序列标准化，将一般正态分布的序列转变为标准正态分布；

(2)对标准化的时间序列用PAA算法进行分段划分，这里需要设置参数 w 大小，表示划分后序列的维数；(3)根据设定的符号集大小(参数 a)，查询等概率断点表(表1给出了参数 a 在3~8之间的断点划分)确定断点集合，最后完成时间序列的符号化。

图1是6类常见雷达信号(包括具有恒定频率的常规信号(CP)、线性调频信号(LFM)、二相编码信号(BPSK)、二进制频率编码信号(2FSK)、非线性调制信号(NLFM)和COSTAS编码信号)的SAX符号化示例。其中，字符集大小 $a=5$ ；符号化后的维度 $w=25$ 。从图1可以看出，不同调制类型的雷达信号，它的符号化序列有明显区别，因此，在这些符号化序列的基础上提取的统计特征具有良好的类别可分性。

表1 参数 a 从3~8的等概率断点查询表^[12]

断点(β_i)	字符集大小(a)					
	3	4	5	6	7	8
β_1	0.43	0.67	0.84	0.97	1.07	1.15
β_2	0.43	0	0.25	0.43	0.57	0.67
β_3	-	0.67	0.25	0	0.18	0.32
β_4	-	-	0.84	0.43	0.18	0
β_5	-	-	-	0.97	0.57	0.32
β_6	-	-	-	-	1.07	0.67
β_7	-	-	-	-	-	1.15

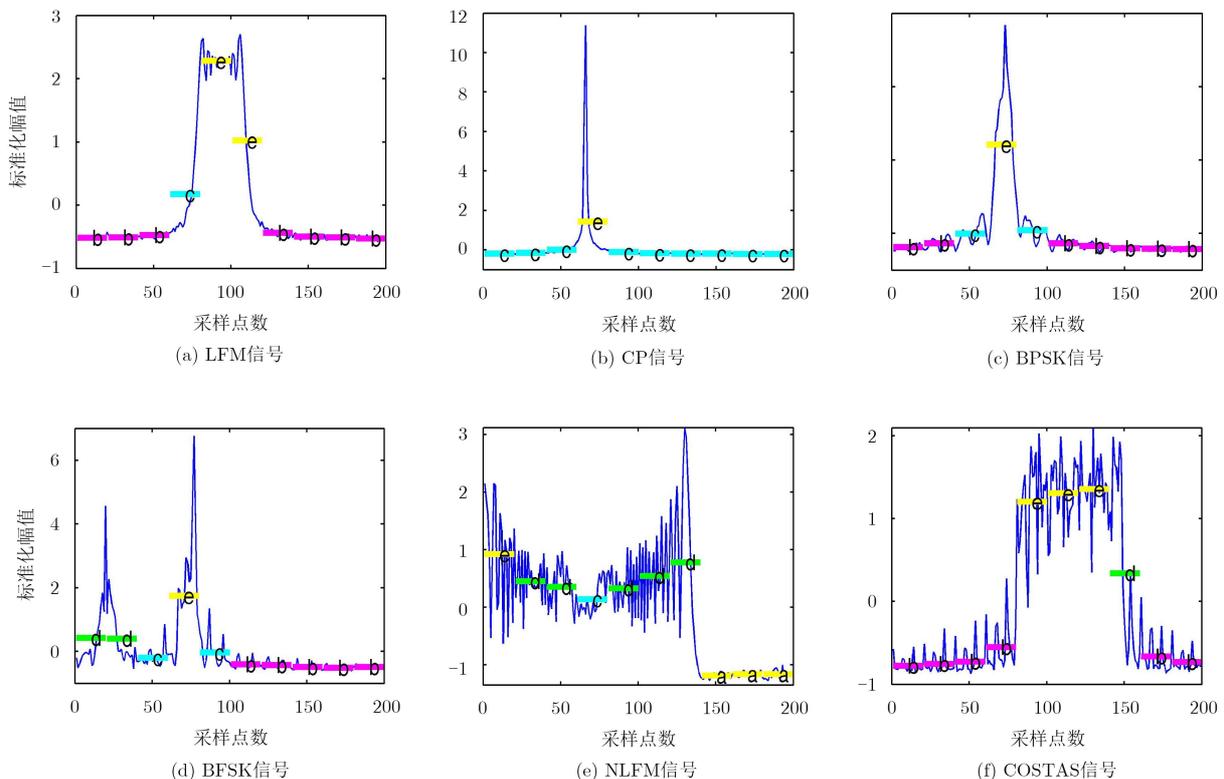


图1 6类雷达信号的SAX符号化序列

3 多尺度信息熵的特征提取

统计不同字符在符号序列中出现的相对频率是分析符号化序列常见方法，一种被广泛应用的频率统计方法是修正的信息熵，它定义为^[9]

$$H = - \frac{1}{\lg(n_s)} \sum_{i=1}^{n_s} p_i \lg(p_i) \quad (5)$$

其中， p_i 是第*i*个字符出现的概率，实际中可以用频率来近似，即 s_i/N (s_i 是第*i*个字符出现的个数； N 是符号化序列的长度)； n_s 是符号化序列中出现的不同字符集大小。修正的信息熵刻画了离散数据的随机程度，对于完全随机的序列，修正信息熵为1，而对于不完全随机的序列，它的修正信息熵在0~1之间。

在修正信息熵的基础上，本文提出了一种多尺度信息熵(Multi-Scale Information Entropy, MSIE)的提取方法，用于描述符号化序列在不同字符集尺度上的无规则程度。因此，字符集大小为*j*的修正信息熵定义为

$$H_j = - \frac{1}{\lg(j)} \sum_{i=1}^{n_j} \varepsilon_{j,i} \lg(\varepsilon_{j,i}), j = a_1, a_2, \dots, a_m \quad (6)$$

式(6)中， $\varepsilon_{j,i}$ 是在该尺度下第*i*个字符出现的概率； a_1 和 a_m 分别是字符集大小的下界和上界， m 为字符集大小区间。

令 \mathbf{T} 为多尺度信息熵向量，选定一个字符集大小区间，假设为 $\{a_1 a_2 \dots a_m\}$ ，则 \mathbf{T} 的计算方法如下：

(1) 对于一个实数序列 $\mathbf{Q} = \{q_i, i = 1, 2, \dots, n\}$ ，采用SAX对其进行符号化，可以得到*m*个符号化序列，记为 $\{\mathbf{S}_t, t = 1, 2, \dots, m\}$ ；

(2) 对每一个符号化序列 \mathbf{S}_t 依据式(6)计算其在第*t*个字符集尺度下的修正信息熵，得到*m*个不同字符集尺度下的修正信息熵，级联这些值所组成的特征向量称为多尺度信息熵，记为 $\mathbf{T} = \{H_{a_1}, H_{a_2}, \dots, H_{a_m}\}$ 。

对于SAX符号化方法，如果对超参数*a*和*w*设置合适的值，符号化后的序列可以很好地保留原序列的结构信息，同时摒弃一些细微的信息。因此，SAX方法本身存在一定的降噪能力，在此基础上，提取的统计量参数(修正的信息熵)也就具有一定的抗噪性能。

为了更好地说明多尺度信息熵的抗噪性能，下面用文献[9]中的仿真信号来分析不同字符集大小下的修正信息熵的抗噪性能，该仿真信号为

$$S_1(t) = 2 \sin(30\pi t) + \sin(150\pi t) + 5 \cos(21.6\pi t) \quad (7)$$

图2信号 S_1 在不同字符集尺度(*a*在3~6之间)下的信息熵与信噪比(Signal to Noise Ratio, SNR)的关系图。其中，SNR在-10~20 dB范围内。从图2中可看出，随着SNR的升高，信息熵的值趋于稳定，在SNR>0 dB时，信息熵的值基本保持不变，所以，信息熵具有较好的抗噪性能。因此，多尺度信息熵也就具有较强的抗噪性能。

4 雷达信号识别模型

综合上面分析和介绍，本文提出了如图3的雷达信号识别方法。其中，数据预处理阶段主要对信号进行消噪处理。考虑到不同调制类型的雷达信号在频率域有不同的幅度谱，同时，在频率域进行特征提取可以削弱载频变化的影响^[16]，本文以雷达信

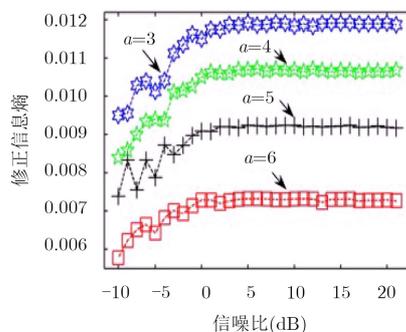


图2 修正的信息熵与SNR的关系图

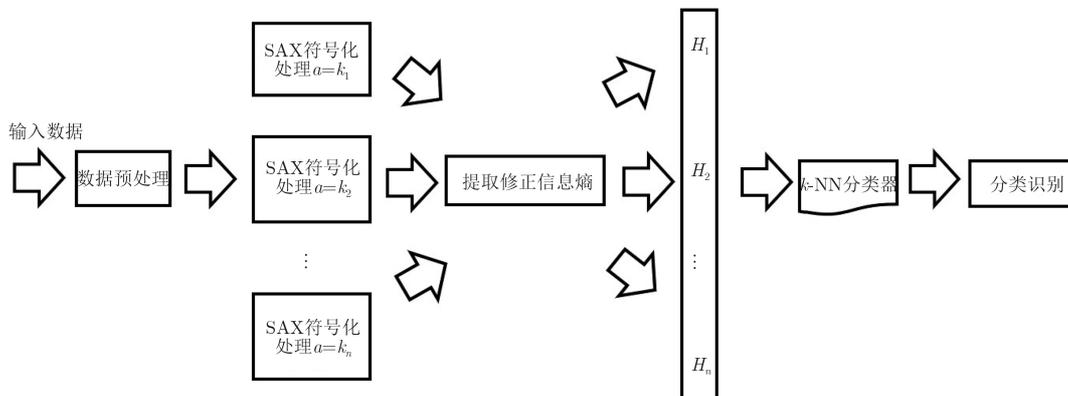


图3 基于多尺度信息熵的雷达信号识别方法

号的频率域序列作为处理对象。在频率域,由于有用信息的能量集中在较窄的频带范围,而噪声能量分布在整个频带上,所以可以按式(8)对信号序列进行消噪处理。令原始雷达信号序列为 $\{r_i, i = 1, 2, \dots, n\}$,并假设消噪后的信号序列为 $\{f_i, i = 1, 2, \dots, n\}$,则移动平均滤波的计算公式为

$$f_i = \frac{r_i + r_{i+1} + \dots + r_{i+d-1}}{d} \quad (8)$$

其中, d 表示移动窗口大小。由于雷达信号的幅度谱是对称的,所以可以只取左边的幅度谱。

在选择分类算法的问题上,考虑到 k -NN算法具有实现简单,易于理解,且对于特征维数较低的多分类问题通常具有较好的性能。因此,选用 k -NN算法作为信号的分类器,该算法的具体实现步骤如下:

步骤 1 给定一个未知的实数序列 Q ,计算 Q 和数据库(训练集)中所有样本的相似性,通常采用欧式距离作为相似性度量方法;

步骤 2 寻找 k 个和 Q 最为相似的样本,作为未知对象 Q 的 k 个近邻;

步骤 3 在这 k 个近邻中出现次数最多的类别即为对象 Q 的类别。

但是, k -NN算法也面临 k 值选择的问题,如果 k 值过大,对于不平衡数据集的分类性能较差;而当 k 值过小时,算法对噪声较为敏感。根据以往的经验,在接下来的仿真实验中,设置超参数 $k=1$ 。

另外,需要注意的是,使用SAX符号化方法需要确定符号集的大小 a 和符号化后的维数 w 的值,然而,目前没有这方面的理论研究,尽管文献[12]给出了参数 a 的经验性选择范围,但是实际应用中还需要根据不同的数据集进行考虑。一般来说,两个超参数都可以通过交叉验证来选择。

5 仿真实验

通过仿真6类常见的辐射源信号来验证本文方法的有效性,包括CP, LFM, BPSK, BFSK, NLFM和COSTAS编码信号。BPSK和BFSK采用13位Barker码, BFSK信号的两个频点分别为10 MHz和12 MHz; LFM的频率变化范围是10~15 MHz; NLFM采用正弦波调制。并且,各辐射源信号的载频为10 MHz,采样频率为60 MHz,脉宽为10 μ s。同时,考虑到实际情况中存在的测量误差,在仿真信号时将载频的均方根误差设置为1 MHz;将脉宽的均方根误差设置为1 μ s。

仿真下列一组数据进行测试:在20~25 dB的信噪比范围内,每种信号样式随机产生20个脉冲信

号,总计120个脉冲信号作为模型的训练样本。同时,分别在5 dB, 10 dB, 15 dB, 20 dB信噪比下各仿真1200个脉冲信号作为测试样本,每一信号样式各200个脉冲。实验中的信号附加的噪声为高斯白噪声。本次实验是在Matlab R2010b平台下完成的。

根据第3节介绍的方法对上面仿真的雷达信号提取多尺度信息熵(设置 a 的范围为3~9; $w=140$),该特征是一个7维的向量,为了便于观察多维数据集的分布情况,这里采用文献[17]提出的t-分布式随机邻域嵌入(t-distributed Stochastic Neighbor Embedding, t-SNE)方法进行数据可视化,将7维的数据映射到2维空间中。图4是在信噪比为5~20 dB范围内6类雷达信号数据的分布图,从图中可以看出,除了BPSK信号和BFSK信号出现小部分重叠无法分开外,多尺度信息熵的类间分离性较好,具有较好的类别可分性,因此,采用 k -NN($k=1$)分类器可以获得良好的性能。

表2是本文所提方法的识别结果,从表中可以看出,在信噪比为15 dB时,6类信号的识别正确率都达到90%以上,说明了本文方法的有效性。另外,从图中还可以看出,CP, LFM, BPSK, NLFM和COSTAS信号在信噪比为5 dB时,仍保持98%以上的正确识别率,这是因为这5类雷达信号的幅度谱明显不同,符号化后提取的多尺度信息熵就具有较好的类别可分性。

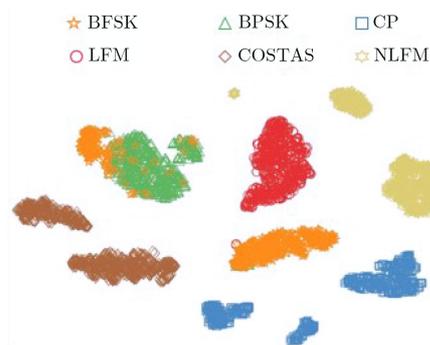


图4 信噪比从5~20 dB时基于信息熵的数据分布图

表2 不同SNR下6种雷达信号的识别率

雷达信号	信噪比SNR (dB)			
	20	15	10	5
LFM	1.000	1.000	1.000	0.985
CP	1.000	1.000	1.000	1.000
BPSK	0.975	0.990	0.990	1.000
BFSK	0.930	0.910	0.800	0.700
NLFM	1.000	1.000	1.000	1.000
COSTAS	1.000	1.000	1.000	1.000

为了圈定本文所提方法的适用范围，需要进一步增加实验进行对比。因此，为上述6类雷达信号增加2种雷达信号，包括四相编码信号(QPSK)和Chirp子脉冲步进频率信号(CSF)。其中，QPSK采用16位Frank码调制。8类信号在20 dB信噪比下仿真800个样本，每类信号包含100个脉冲作为测试集，在20~25 dB的信噪比范围内，每种信号样式随机产生20个脉冲信号，总计180个脉冲信号作为模型的训练样本，实验中用交叉验证选择字符集大小 $a=11$ ，以及符号化后的维度 $w=160$ 。考虑到QPSK信号和BPSK信号的幅度谱较为相似，因此，理论上这两种信号提取的多尺度信息熵存在部分重叠，分类器将难以区分这两类信号，从图5中的混淆矩阵也可以明显看出，QPSK和BPSK的区分度较差。

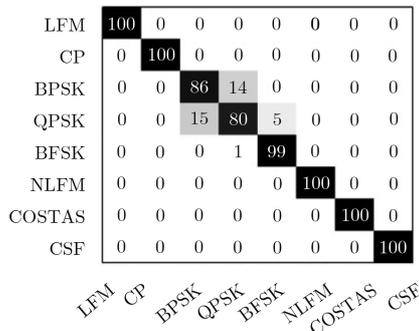


图5 基于多尺度信息熵识别结果的混淆矩阵

传统的方法可以较好地地区分BPSK与QPSK两种信号，但是一般需要将雷达信号转换到时-频域上提取特征，而大多数基于时-频变换方法或多或少的存在问题^[16]，并且受噪声的影响严重，容易导致时-频信号失真，而傅里叶变换能获得稳定的频率域序列。另外，基于时-频变换方法的时间复杂度通常要高于傅里叶变换算法。这里选择一种经典的时-频域特征——小波脊频级联特征(Wavelet Ridge Frequency Cascade-Connection Feature, WRFCCF)^[16]进行实验对比，该特征是一个7维的特征向量，其中每个特征是雷达信号在时-频域上的统计量。首先，对提取WRFCCF和MSIE所耗费的时间进行对比，实验采用上述的100个训练样本进行实验，结果重复10次取平均(如表3所示)。从表中可以看出，提取WRFCCF所耗费的时间要远远高于提取多尺度信息熵的时间。

使用小波脊频级联特征，并选择 k -NN分类器，可以得到如图6的混淆矩阵。从图中可以看出，WRFCCF能有效的区分BPSK, QPSK两种信号，但是对于其它调制类型的雷达信号，如图6中

表3 提取两种特征耗费的时间对比

特征向量	耗费时间(s)
WRFCCF	135.102
MSIE	1.704

的CP、BFSK的区分度较差，而从图5可知，本文所提方法除了BPSK, QPSK外，其它信号的识别率基本上达到了100%(如表4，本文方法获得更好的总体识别率)，这是因为多尺度信息熵是描述雷达信号的幅度谱的结构形态，当雷达信号的幅度谱存在较弱相似性的情况下，可以获得好的识别性能。

多尺度信息熵是一种典型的频率域特征，这里选择两种经典的频率域特征进行识别性能的比较，包括文献[2]提出的盒维数和稀疏性，以及文献[9]提出的信息熵特征。盒维数和稀疏性这两个复杂度特征(complexity characteristics)与信息熵类似，是从不同层面来描述信号的结构。同样采用t-SNE方法对数据进行可视化，图7是在信噪比为5~20 dB范围内6类基于复杂度的雷数据分布图，显然，复杂度特征的类型分离性较差，分布图中有许多数据出现重叠。

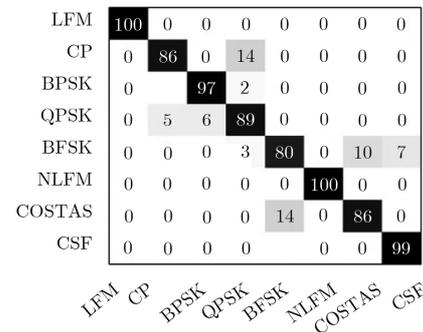


图6 基于小波脊频级联特征识别结果的混淆矩阵

表4 两种方法的总体识别正确率比较(%)

识别方法	总体识别率
WRFCCF+ k -NN	92.13
MSIE+ k -NN	95.63

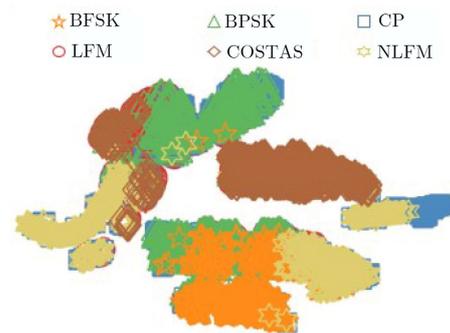


图7 信噪比从5~20 dB时基于复杂度的数据分布图

在测试特征的识别性能之前, 首先对特征提取算法的计算量进行对比。提取盒维数特征时, 需要计算序列的绝对值差分, 所以算法的执行次数 $f_b(n) = N + (N + 1)/2$, 其中, N 是序列的长度; 提取稀疏性特征时, 首先需要先对序列进行二进制符号化, 将序列划分为 L 个长度为 N_0 的子序列, 统计窗内序列中“1”元素的占比, 最后求和。所以, 该算法的执行次数 $f_s(n) = N + LN_0$ 。假设样本集的大小为 M , 则复杂度特征提取算法的时间复杂度为

$$\begin{aligned} T(n) &= O(M(f_b(n) + f_s(n))) \\ &= O(M(N + (N + 1)/2 + N + LN_0)) \\ &= O(MN). \end{aligned} \quad (9)$$

SAX是在PAA的基础上进行符号化, 而且容易知道PAA算法的时间复杂度为 $O(L_1)$, 其中 L_1 是算法划分的段数。另外, 修正信息熵是统计不同字符出现的相对频率, 因此, 它的时间复杂度也为 $O(L_1)$ 。所以, 多尺度信息熵的时间复杂度为 $T(n) = O(cML_1)$, 其中, c 是字符集尺度的取值范围, 实际情况中通常较小, 可以认为 $cL_1 \approx N$, 则多尺度信息熵的时间复杂度约等于 $O(MN)$ 。另外, 由文献[9]中关于信息熵的提取算法, 可知其时间复杂度为 $O(MN)$ 。通过上面的分析可知, 3种特征具有相同的时间复杂度, 能在合理的时间内提取特征。

表5分别是3种方法在5 dB, 10 dB, 15 dB, 20 dB信噪比下的总体识别正确率。为了便于描述, 这里将本文所提方法表述为“MSIE+ k -NN”; 将文献[2]提出的方法表述为“CC+ k -NN”; 将文献[9]提出的方法表述为“SIE+ k -NN”。3种方法的总体识别率如表所示, 从表中可以容易看出, 本文方法相比两种经典的频率域特征具有更好的识别正确率。

6 结论

本文提出了一种基于多尺度信息熵的雷达信号识别方法, 能够较好地对雷达信号进行识别。本文方法将信号从实数域映射到符号域, 去除信号中的细微特征的同时保留其结构特征, 从而提取具有较好抗噪性能的特征。通过仿真实验表明, 本文方法在雷达信号的幅度谱存在较弱相似性的情况下, 具

有很好的识别率, 并且优于传统的基于复杂度特征的识别方法, 说明了本文方法具有一定的参考价值。

参考文献

- [1] WILEY R G. ELINT: The Interception and Analysis of Radar Signals[M]. Norwood, USA: Artech House, 2006: 1–15.
- [2] 韩俊, 何明浩, 朱振波, 等. 基于复杂度特征的未知雷达辐射源信号分选[J]. 电子与信息学报, 2009, 31(11): 2552–2556. HAN Jun, HE Minghao, ZHU Zhenbo, *et al.* Sorting unknown radar emitter signal based on the complexity characteristics[J]. *Journal of Electronics & Information Technology*, 2009, 31(11): 2552–2556.
- [3] 曲志昱, 毛校洁, 侯长波. 基于奇异值熵和分形维数的雷达信号识别[J]. 系统工程与电子技术, 2018, 40(2): 303–307. doi: 10.3969/j.issn.1001-506X.2018.02.10. QU Zhiyu, MAO Xiaojie, and HOU Changbo. Radar signal recognition based on singular value entropy and fractal dimension[J]. *Systems Engineering and Electronics*, 2018, 40(2): 303–307. doi: 10.3969/j.issn.1001-506X.2018.02.10.
- [4] LI Jingchao and YING Yulong. Radar signal recognition algorithm based on entropy theory[C]. Proceedings of the 2nd International Conference on Systems and Informatics, Shanghai, China, 2014: 718–723.
- [5] GUO Yuanyuan and ZHANG Xudong. Radar signal classification based on cascade of STFT, PCA and naïve Bayes[C]. Proceedings of the 7th International Conference on Intelligent Systems, Modelling and Simulation, Bangkok, Thailand, 2016: 191–196.
- [6] DUDCZYK J. A method of feature selection in the aspect of specific identification of radar signals[J]. *Bulletin of the Polish Academy of Sciences Technical Sciences*, 2017, 65(1): 113–119. doi: 10.1515/bpasts-2017-0014.
- [7] GUO Qiang, NAN Polong, and WAN Jian. Signal classification method based on data mining for multi-mode radar[J]. *Journal of Systems Engineering and Electronics*, 2016, 27(5): 1010–1017. doi: 10.21629/JSEE.2016.05.09.
- [8] KONOPKO K, GRISHIN Y P, and JAŃCZAK D. Radar signal recognition based on time-frequency representations and multidimensional probability density function estimator[C]. Proceedings of 2015 Signal Processing Symposium, Dębe, Poland, 2015: 1–6.
- [9] 陈韬伟, 金炜东. 雷达辐射源信号符号化脉内特征提取方法[J]. 数据采集与处理, 2008, 23(5): 521–526. doi: 10.3969/j.issn.1004-9037.2008.05.004. CHEN Taowei and JIN Weidong. Intra-pulse feature extraction of radar emitter signals based on symbolization method[J]. *Journal of Data Acquisition & Processing*, 2008,

表5 3种方法的总体识别正确率比较(%)

识别方法	信噪比SNR(dB)			
	20	15	10	5
MSIE+ k -NN	98.42	97.25	94.25	91.25
CC+ k -NN	80.25	73.08	54.33	<50
SIE+ k -NN	81.42	79.08	71.25	62.92

- 23(5): 521–526. doi: [10.3969/j.issn.1004-9037.2008.05.004](https://doi.org/10.3969/j.issn.1004-9037.2008.05.004).
- [10] CHEN Taowei and JIN Weidong. Feature extraction of radar emitter signals based on symbolic time series analysis[C]. Proceedings of 2007 International Conference on Wavelet Analysis and Pattern Recognition, Beijing, China, 2007: 1277–1282.
- [11] CHEN Taowei, LIU Zugen, LI Jie, *et al.* Symbolic time series analysis for measuring complexity in radar emitter signals[C]. Proceedings of the 7th International Congress on Image and Signal Processing, Dalian, China, 2014: 918–922.
- [12] LIN J, KEOGH E, WEI Li, *et al.* Experiencing SAX: A novel symbolic representation of time series[J]. *Data Mining and Knowledge Discovery*, 2007, 15(2): 107–144. doi: [10.1007/s10618-007-0064-z](https://doi.org/10.1007/s10618-007-0064-z).
- [13] SONG Wei, WANG Zhiguang, ZHANG Fan, *et al.* Empirical study of symbolic aggregate approximation for time series classification[J]. *Intelligent Data Analysis*, 2017, 21(1): 135–150. doi: [10.3233/IDA-150351](https://doi.org/10.3233/IDA-150351).
- [14] 向馗, 蒋静坪. 时间序列的符号化方法研究[J]. 模式识别与人工智能, 2007, 20(2): 154–161. doi: [10.3969/j.issn.1003-6059.2007.02.003](https://doi.org/10.3969/j.issn.1003-6059.2007.02.003).
- XIANG Kui and JIANG Jingping. Study on symbolization analysis of time series[J]. *Pattern Recognition and Artificial Intelligence*, 2007, 20(2): 154–161. doi: [10.3969/j.issn.1003-6059.2007.02.003](https://doi.org/10.3969/j.issn.1003-6059.2007.02.003).
- [15] KEOGH E, CHAKRABARTI K, PAZZANI M, *et al.* Dimensionality reduction for fast similarity search in large time series databases[J]. *Knowledge and Information Systems*, 2001, 3(3): 263–286. doi: [10.1007/PL00011669](https://doi.org/10.1007/PL00011669).
- [16] 余志斌. 基于脉内特征的雷达辐射源信号识别研究[D]. [博士学位论文], 西南交通大学, 2010: 1–56.
- YU Zhibin. Study on radar emitter signal identification based on intra-pulse features[D]. [Ph.D. dissertation], Southwest Jiaotong University, 2010: 1–56.
- [17] VAN DER MAATEN A and HINTON G. Visualizing data using t-SNE[J]. *Journal of Machine Learning Research*, 2008(9): 2579–2605.
- 黄颖坤: 男, 1989年生, 博士生, 研究方向为雷达信号处理, 机器学习.
- 金炜东: 男, 1959年生, 教授, 博士生导师, 研究方向为智能信息处理、系统仿真与优化方法.
- 葛 鹏: 男, 1986年生, 讲师, 研究方向为雷达信号处理, 电子对抗.
- 李 冰: 女, 1988年生, 讲师, 研究方向为电磁场与电磁波, 微波成像.