

文章编号: 1001-893X(2010)10-0036-05

运用高阶累积量和 SVM 的调制自动识别^{*}

闫朋展, 王振宇

(解放军电子工程学院, 合肥 230037)

摘要:针对数字信号调制模式识别问题,提出了运用高阶累积量和二叉树支持向量机(SVM)进行自动识别的算法。该算法首先使用信号的四阶、六阶、八阶累积量构造了 5 个新的分类特征,然后利用二叉树支持向量机分类器实现了 8 种信号的有效分类。仿真结果表明,该算法优于直接多类分类支持向量机算法,在信噪比大于 5 dB 时,识别率达到 90% 以上。

关键词:数字信号;调制模式识别;高阶累积量;二叉树支持向量机;自动识别

中图分类号: TN911 **文献标识码:** A doi: 10.3969/j.issn.1001-893x.2010.10.007

Automatic Recognition of Digital Modulation Signals by Applying High Order Cumulants and Support Vector Machines

YAN Peng-zhan, WANG Zhen-yu

(PLA Institute of Electronic Engineering, Hefei 230037, China)

Abstract: Applying high order cumulants and support vector machines(SVM), an algorithm is proposed for automatic recognition of digital communication signals. Using the forth, sixth and eighth orders of signals, the new algorithm forms five new classification features at first, then identifies eight types of signals by binary tree-based SVM as the classifier. Computer simulation indicates that the algorithm outperforms the one based on BT-SVM, and the rate of accurate classification is over 90% when SNR is above 5dB.

Key words: digital signal; modulation recognition; high order cumulants; binary tree-based SVM; automatic recognition

1 引言

通信信号调制模式自动识别在军用和民用方面一直发挥着重要作用。在军用方面,调制模式的正确、有效识别是指挥员制定侦察和对抗的前提条件;在民用方面,政府通信部门可以利用它进行信号确认、干扰识别和频谱监测等无线电管理工作,以防止用户对无线频谱的非法利用和干扰,保证合法通信的正常进行。

目前,数字信号调制模式自动识别主要有两种方法:决策论的识别方法和统计模式的识别方法。

由于统计模式识别方法提取特征相对简单,需要的先验知识少,算法灵活,因此,很多学者采用这种方法进行调制模式自动识别研究。

调制模式识别的关键是提取性能优、稳健性好的分类特征。由于高阶累积量具有良好的抗噪性能,因此,使用高阶累积量进行模式识别的研究也越来越多。Swami 利用四阶累积量实现了 2PSK 和 4PSK 信号的分类识别^[1];文献[2]利用四阶、六阶累积量实现了对 2ASK、4ASK、4PSK、8PSK、16QAM 的识别;陆凤波利用信号差分的四阶、八阶累积量实现了 4 种调相信号的识别^[3];文献[4]利用四阶、六阶累积量构造两个分类特征实现了 6 种信号的分类。

^{*} 收稿日期:2010-05-25;修回日期:2010-06-30

通过分析和研究,发现以上文献识别的类型相对较少,在低信噪比下识别率不高。因此,本文采用一种新的基于高阶累积量和支持向量机的识别算法,实现了 2ASK、4ASK、8ASK、4PSK、8PSK、8QAM、16QAM、32QAM 共 8 种信号在低信噪比下的有效分类。

2 高阶累积量特征提取

假设接收到的信号已经过下变频、载波同步和符号同步,此时,接收到的被噪声污染的复基带信号可表示为

$$r(t) = s(t) + n(t) =$$
$$A \sum_n a_n g(t - nT) \exp(j\theta_c) + n(t) \quad (1)$$

式中, A 为载波幅度, a_n 为码元序列, $g(t)$ 为发送码元波形(矩形脉冲), θ_c 为载波初始相位, $n(t)$ 为高斯白噪声信号。

2.1 高阶累积量理论基础

下面给出本文所用到的高阶累积量的定义。在介绍高阶累积量之前,先定义高阶矩。对于一个具

有零均值的复随机过程 $X(t)$, p 阶混合矩定义为

$$M_{pq} = E[X(t)^p - q X^*(t)^q] \quad (2)$$

式中, $*$ 表示复共轭。然后定义各阶累积量^[5]:

$$C_{40} = M_{40} - 3M_{20}^2 \quad (3)$$
$$C_{41} = M_{41} - 3M_{20}M_{21} \quad (4)$$
$$C_{42} = M_{42} - M_{20}M_{22} - 2M_{21}^2 \quad (5)$$
$$C_{60} = M_{60} - 15M_{20}M_{40} + 30M_{20}^3 \quad (6)$$
$$C_{63} = M_{63} - 9C_{42}M_{21} - 6M_{21}^3 \quad (7)$$
$$C_{80} = M_{80} - 28M_{20}M_{60} - 35M_{40}^2 +$$
$$420M_{40}M_{20}^2 - 630M_{20}^4 \quad (8)$$
$$C_{84} = M_{84} - 16C_{63}M_{21} - |C_{40}|^2 -$$
$$18C_{42}^2 - 72C_{42}M_{21}^2 - 24M_{21}^4 \quad (9)$$

2.2 数字调制信号的高阶累积量分析

由于高斯随机过程二阶以上累积量的值为零,所以,接收到的受高斯白噪声污染的复基带信号累积量值即为原复基带信号累积量的值。假设发送的码元服从独立同分布,复基带信号经过零均值预处理和功率归一化后,将其代入到式(3)~(9)中,计算各阶累积量的值。理论计算值如表 1 所示。

表 1 信号的高阶累积量理论值

Table 1 Signals' higher order cumulants theoretic values

信号	C_{40}	C_{41}	C_{42}	C_{60}	C_{63}	C_{80}	C_{84}
2ASK	$-2A^4e^{j4\theta_c}$	$-2A^4e^{j4\theta_c}$	$-2A^4$	$16A^6e^{j6\theta_c}$	$13A^6$	$-272A^8e^{j8\theta_c}$	$-163A^8$
4ASK	$-1.36A^4e^{j4\theta_c}$	$-1.36A^4e^{j2\theta_c}$	$-1.36A^4$	$8.32A^6e^{j6\theta_c}$	$9.16A^6$	$-111.85A^8e^{j8\theta_c}$	$-102.53A^8$
8ASK	$-1.238A^4e^{j4\theta_c}$	$-1.238A^4e^{j2\theta_c}$	$-1.238A^4$	$7.19A^6e^{j6\theta_c}$	$8.76A^6$	$-92.018A^8e^{j8\theta_c}$	$-96.225A^8$
4PSK	$A^4e^{j4\theta_c}$	0	$-A^4$	0	$4A^6$	$-34A^8e^{j8\theta_c}$	$-34A^8$
8PSK	0	0	$-A^4$	0	$4A^6$	$A^8e^{j8\theta_c}$	$-33A^8$
8QAM	$-A^4e^{j4\theta_c}$	$-0.889A^4e^{j2\theta_c}$	$-A^4$	$4.74A^6e^{j6\theta_c}$	$5.33A^6$	$-54.148A^8e^{j8\theta_c}$	$-52.469A^8$
16QAM	$-0.68A^4e^{j4\theta_c}$	0	$-0.68A^4$	0	$2.08A^6$	$-13.98A^8e^{j8\theta_c}$	$-13.98A^8$
32QAM	$-0.19A^4e^{j4\theta_c}$	0	$-0.69A^4$	0	$2.11A^6$	$-1.993A^8e^{j8\theta_c}$	$-13.786A^8$

在实际工程应用中,通常使用信号样本数据来估计高阶累积量的值。首先, p 阶混合矩的样本估计表示式为

$$\hat{M}_{pq} = \frac{1}{N} \sum_{k=1}^N X(k)^p - q X^*(k)^q \quad (10)$$

式中, N 为样本数据点数, $X(k)$ 为信号。然后,把混合矩的估计值代入到累积量公式就能得到各阶累积量的样本估计值。

2.3 构造累积量分类特征

为了实现数字信号调制模式的高效识别,必须运用各阶累积量的组合构造分类特征参数。本文结合待识别信号的情况,构造了 5 个分类特征参数,分别定义如下: $F_1 = \left| \frac{C_{40}}{C_{42}} \right|$, $F_2 = \left| \frac{C_{63}^2}{C_{42}^3} \right|$, $F_3 = \left| \frac{C_{80}}{C_{42}^2} \right|$, $F_4 = \left| \frac{C_{60}}{C_{63}} \right|$, $F_5 = \left| \frac{C_{80}}{C_{84}} \right|$ 。5 个参数的理论值如表 2

所示。

表 2 信号的特征参数的理论值

信号	F1	F2	F3	F4	F5
2ASK	1	21.125 0	68	1.230 8	16 687
4ASK	1	33.356 0	60.470 6	0.908 3	1.090 8
8ASK	1	40.436 2	60.028 8	0.820 6	0.956 3
4PSK	1	16	34	0	1
8PSK	0	16	1	0	0.303
8QAM	1	28.444 1	54.148 1	0.888 9	1.032 0
16QAM	1	13.759 4	30.235 3	0	1
32QAM	0.275 4	13.552 4	4.185 3	0	0.144 5

3 二叉树结构的支持向量机分类器

二叉树结构的支持向量机是一种处理多类分类问题的高效方法,其基本思想是:从根节点开始,将该节点所包含的类别划分为两个子类,再将两个子类划分为两个次级子类,如此迭代下去,直到得到一个单独的类别为止,这样就将多类分类问题转换为两类分类问题,其中每两子类之间采用支持向量机作为分类器。通常情况下,二叉树结构的支持向量机分为两种:一种是在每个内节点处,只将一个类别与其它类别分开;另一种是在每个内节点处,将所有类别分为两类。本文采用第一种二叉树结构的支持向量机分类器,并采用一种新的类距离方法生成二叉树结构^[6],其基本思想是把与其它类相隔最远的类优先分离出来。

首先,定义类距离 d_{ij} 为 i 类和 j 类均值向量间的距离减去各自类的平均半径,即:

$$d_{ij} = \| m_i - m_j \|^2 - r_i - r_j \tag{10}$$

$$r_i = \frac{1}{n_i} \sum_{k=1}^{n_i} \| x_k^i - m_i \|^2 \tag{11}$$

式中, $m_i = \frac{1}{n_i} \sum_{k=1}^{n_i} x_k^i$ 表示第 i 类样本集的均值向量, $\| m_i - m_j \|^2$ 为 i 类和 j 类均值向量间的距离, r_i 和 r_j 分别为 i 类和 j 类的类平均半径, n_i 为 i 类中的样本数。

以类均值距离生成二叉树结构支持向量机的步骤如下:

第一步:运用式(10)计算类与类之间的距离;

第二步:对每一类都存在 $k - 1$ 个与其它类的距离值,分别对每个类的这些距离值从小到大排列,并重新编号;

第三步:根据 $d_i^1 (i = 1, 2, 3, \cdots, k, i \neq j)$ 的值对相应的类按照由大到小的顺序进行排序。当存在两个或两个以上的类具有相同的 d_i^1 时,可比较它们的 d_i^2 大小,如此下去,若 $d_i^1, d_i^2, d_i^3, \cdots, d_i^{k-1}$ 都相同,则把类标号小的排在前面,标号大的排在后面;

第四步:根据类标号排序,生成二叉树;

第五步:根据生成的二叉树,利用二分类 SVM 训练算法构造二叉树内各节点的最优超平面。

该方法解决了不可分盲点、传统结构识别率不够优等问题,提高了算法的性能,具有较好的推广性。

4 实验性能分析

本文所研究的待分类调制信号为 2ASK、4ASK、8ASK、4PSK、8PSK、8QAM、16QAM、32QAM; 仿真环境为加性高斯白噪声信道;载波频率 $f_c = 1\,000\text{ Hz}$, 采样频率 $f_s = 8\,000\text{ Hz}$, 码元速率 $f = 1\,000\text{ bit/s}$, 每个样本 2 000 个采样点,用于二叉树结构的支持向量机分类器的样本数为 150 个,其中,50 个样本用于训练,100 个样本用于测试性能,以上参数在后面的仿真和性能测试中均保持不变。

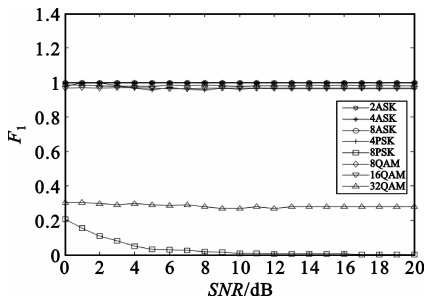
4.1 特征参数随信噪比变化

对 5 个构造的新的分类特征随信噪比变化情况进行仿真。8 种信号,信噪比从 0 ~ 20 dB 以间隔 1 dB 变化,各取 100 个样本,计算特征参数的均值。仿真结果如图 1 所示。

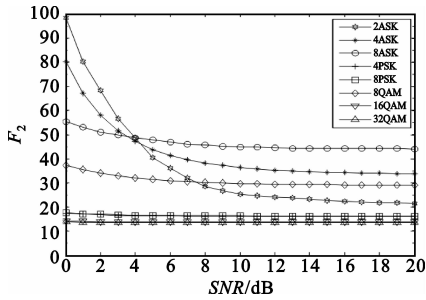
从图 1 可看出,变化曲线之间能够较好地区分,即说明 8 种信号能够通过 5 个新的分类特征进行识别,为采用支持向量机进行识别提供了依据。

4.2 识别性能

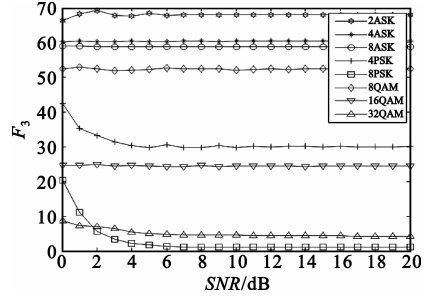
本文采用 Monte Carlo 仿真实验的方法来分析算法的性能。通过两个仿真实验的对比来验证算法的有效性,即基于高阶累积量与二叉树结构的支持向量机识别算法和基于高阶累积量与直接多类分类支持向量机的算法。仿真实验中所用的参数均一致,信噪比从 0 ~ 20 dB 以间隔 5 dB 变化;SVM 采用径向基核函数,核函数参数 $\delta^2 = 0.001\,5$, 惩罚因子 $C = 650$ 。



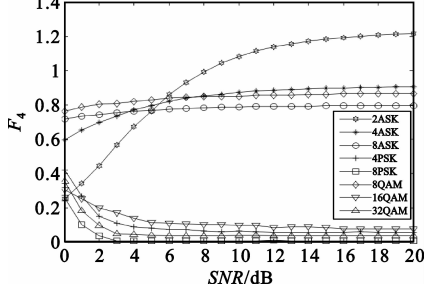
(a)参数 F_1 随信噪比变化曲线



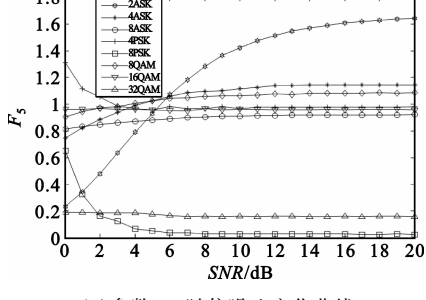
(b)参数 F_2 随信噪比变化曲线



(c)参数 F_3 随信噪比变化曲线



(d)参数 F_4 随信噪比变化曲线



(e)参数 F_5 随信噪比变化曲线

图 1 特征参数随信噪比变化曲线

仿真实验 1: 基于高阶累积量与直接多类分类支持向量机算法仿真。仿真结果如表 3 所示。

表 3 直接多类分类 SVM 识别率仿真结果

Table 3 The recognition rate based on direct SVM

信号	识别率/(%)				
	0 dB	5 dB	10 dB	15 dB	20 dB
2ASK	74	100	100	100	100
4ASK	41	94	100	100	100
8ASK	85	96	96	97	98
4PSK	52	100	99	100	99
8PSK	54	98	100	100	100
8QAM	84	100	100	100	100
16QAM	75	78	82	85	84
32QAM	75	78	82	83	82

仿真实验 2: 本文算法仿真, 即高阶累积量与二叉树结构的支持向量机算法仿真。根据本文所研究识别信号的特点和构造的 5 个特征参数随信噪比变化情况, 可得到该算法的识别流程图, 见图 2。

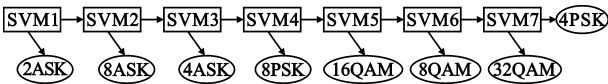


图 2 信号识别流程图

Fig.2 The identification flowchart of signals

依据图 2 流程进行仿真, 仿真结果见表 4。

表 4 基于二叉树结构的 SVM 识别率仿真结果

Table 4 The recognition rate based on BT-SVM

信号	识别率/(%)				
	0 dB	5 dB	10 dB	15 dB	20 dB
2ASK	72	100	100	100	100
4ASK	45	99	100	100	100
8ASK	97	97	98	98	99
4PSK	34	99	100	100	100
8PSK	63	98	100	100	100
8QAM	100	100	100	100	100
16QAM	84	90	90	95	92
32QAM	92	93	93	94	93

从表 3 和表 4 的仿真结果来看, 基于高阶累积量与直接多类分类支持向量机算法在信噪比大于 5 dB 时, 总体识别率 78% 以上, 当信噪比为 0 dB 时, 8 种信号识别率较低; 本文算法在信噪比大于 5 dB 时, 总体识别率在 90% 以上, 除 16QAM 和 32QAM 外, 其它信号识别率接近 100%。

5 结 论

本文通过对高阶累积量理论的研究和待识别信号的分析,提出了 5 个新的分类特征,这 5 个特征对信号的尺度大小和相位变化具有很好的稳健性,再结合二叉树支持向量机对所研究信号进行识别,通过仿真验证,证明了本文算法的有效性。但本文算法在信噪比低于 5 dB 时,对 2ASK、4ASK、4PSK、8PSK 4 种信号识别率较低,比较难区分,在后续研究中要深入分析进行解决。且本文算法仅是在高斯白噪声信道下进行验证的,而实际信道环境比较复杂,如何将该算法推广、应用到实际复杂的信道环境下,也是今后研究的方向。

参考文献:

[1] Swami A, Sadler B M. Hierarchical Digital Modulation Classification Using Cumulants[J]. IEEE Transactions on Communication, 2000, 48(3): 416 - 429.

[2] 韩钢, 张文红, 李建东, 等. 基于高阶累积量和支撑矢量机的调制识别研究[J]. 系统工程与电子技术, 2003, 25(8): 1008 - 1011.

HAN Gang, ZHANG Wen - hong, LI Jian - dong, et al. Study on the recognition based on higher order cumulants and SVM[J]. Journal of Systems Engineering and Electronics. 2003, 25(8): 1008 - 1011. (in Chinese)

[3] 陆凤波, 黄知涛, 易辉荣, 等. 一种基于高阶累积量的数字调相信号识别方法[J]. 系统工程与电子技术, 2008, 30(9): 1612 - 1615.

LU Feng - bo, HUANG Zhi - tao, YI Hui - rong, et al. Recognition algorithm of phase shift keying signals by higher

order cumulants[J]. Journal of Systems Engineering and Electronics, 2008, 30(9): 1612 - 1615. (in Chinese)

[4] WANG Lan - xun, REN Yu - jing, ZHANG Rui - hua. Algorithm of digital modulation recognition based on support vector machines[C]//Proceedings of 2009 International Conference on Machine Learning and Cybernetics. Baoding: IEEE, 2009: 980 - 983.

[5] 陈筱倩, 王宏远. 基于联合特征向量的自动数字调制识别算法[J]. 计算机应用研究, 2009, 26(7): 2478 - 2480.

CHEN Xiao - qian, WANG Hong - yuan. Automatic digital modulation classification algorithm based on novel combined feature vector[J]. Application Research of Computers, 2009, 26(7): 2478 - 2480. (in Chinese)

[6] 吕晓丽, 李雷, 曹未丰. 基于二叉树的 SVM 多类分类算法[J]. 信息技术, 2008(4): 1 - 3.

LV Xiao - li, LI Lei, CAO Wei - feng. SVM multi - class classification based on binary tree[J]. Journal of Information Technology, 2008(4): 1 - 3. (in Chinese)

作者简介:

闫朋展(1983 -), 男, 陕西户县人, 硕士研究生, 主要研究方向为数字信号调制模式识别;

YAN Peng - zhan was born in Huxian, Shaanxi Province, in 1983. He is a now graduate student. His research direction is digital signals recognition.

Email: yanpengzhan08@163.com

王振宇(1956 -), 男, 四川蓬溪人, 副教授, 主要研究方向为信号处理和嵌入式系统。

WANG Zhen - yu was born in Pengxi, Sichuan Province, in 1956. He is now a professor. His research interests include signals processing and embedded system.



欢迎订阅全国中文核心期刊《电讯技术》

邮发代号: 62 - 39

全国各地邮局均可订阅!