Journal of Southeast University (English Edition) Vol. 25, No. 2, pp. 171 – 174 June 2009 ISSN 1003—7985

Modulation classification of MPSK signals based on nonparametric Bayesian inference

Chen Liang[1,2] Cheng Hanwen[1] Wu Lenan[1]

([1] School of Information Science and Engineering, Southeast University, Nanjing 210096, China)
([2] School of Mechanical Engineering, Jiangsu University, Zhenjiang 212013, China)

Abstract: A nonparametric Bayesian method is presented to classify the MPSK (*M*-ary phase shift keying) signals. The MPSK signals with unknown signal noise ratios (SNRs) are modeled as a Gaussian mixture model with unknown means and covariances in the constellation plane, and a clustering method is proposed to estimate the probability density of the MPSK signals. The method is based on the nonparametric Bayesian inference, which introduces the Dirichlet process as the prior probability of the mixture coefficient, and applies a normal inverse Wishart (NIW) distribution as the prior probability of the unknown mean and covariance. Then, according to the received signals, the parameters are adjusted by the Monte Carlo Markov chain (MCMC) random sampling algorithm. By iterations, the density estimation of the MPSK signals can be estimated. Simulation results show that the correct recognition ratio of 2/4/8PSK is greater than 95% under the condition that SNR > 5 dB and 1 600 symbols are used in this method.

Key words: modulation classification; *M*-ary phase shift keying; Dirichlet process; nonparametric Bayesian inference; Monte Carlo Markov chain

The ability of detecting an unknown modulated signal and recognizing its modulation type is an important feature in communication intelligence. This feature is usually referred to as modulation classification and modulation recognition[1]. Much attention has been paid to its key application in adaptive modulation, software defined radio, and cognitive radio[2].

MPSK modulation has been extensively used in shortwave communication, satellite communication and mobile communication. The modulation classification of MPSK signals has important applications in interference recognition and spectrum supervision. Hence, it is worthy of intensive research.

It is robust to utilize the constellation plane as the characteristic for modulation classification[3]. Refs. [4 – 5] exploited the aggregation property of the received signals scattered in the complex plane, and used the *c*-mean clustering method and the blur *c*-mean clustering method to classify the signals respectively. But the premise of the algorithms is that the clustering number is previously known.

In this paper, we use the method of nonparametric Bayesian inference to cluster the MPSK signals, and obtain the classification number *M* according to the cluster results. The main advantage of this algorithm is that it allows the number of clustering components to grow as the size of the signal set grows without assuming a fixed number of components.

Received 2008-10-17.

Biographies: Chen Liang(1977—), male, graduate; Wu Lenan(corresponding author), male, doctor, professor, wuln@seu.edu.cn.

Foundation item: Cultivation Fund of the Key Scientific and Technical Innovation Project of Ministry of Education of China(No. 3104001014).

Citation: Chen Liang, Cheng Hanwen, Wu Lenan. Modulation classification of MPSK signals based on nonparametric Bayesian inference[J]. Journal of Southeast University (English Edition), 2009, 25(2): 171 – 174.

1 Problem Specification

The MPSK equivalent baseband signal $y(n)$ suffers from residual baseband channel effects, timing errors, phase jitters and carrier frequency offset, etc. To facilitate our analysis, in this paper, we assume an ideal working condition with only the presence of white Gaussian noise $\boldsymbol{G}(n)$, where the channel has been adequately equalized, the residual channel effect is negligible, symbol timing and carrier frequency have been successfully estimated. Under these assumptions, the received sequence can be rewritten as

$$y(n) = Ax(n) + g(n) \tag{1}$$

where $x(n)$ is the signal sequence, and A is an unknown amplitude factor.

According to Eq. (1), $\boldsymbol{y}_i$ conforms to a two-dimensional Gaussian distribution in the complex plane, $\boldsymbol{y}_i \sim N(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k)$, in which the mean $\boldsymbol{\mu}_k$ equals a certain constellation center of MPSK $I_l(l = 1, 2, \ldots, M)$, and the covariance $\boldsymbol{\Lambda}_k$ is determined by the SNR. Hence, all the received signals $\boldsymbol{Y} = \{\boldsymbol{y}_{1:n}\}$ can be considered to be generated by a Gaussian mixture model with M components.

The classification problem can be specified as follows: we have n measurements $\boldsymbol{Y} = \{\boldsymbol{y}_{1:n}\}$ that are distributed as a mixture density. The probability model can be expressed as

$$\begin{aligned} \boldsymbol{Y} \mid \pi,\ \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M \ &\sim \ \sum_{k=1}^{M} \pi_k N(\boldsymbol{Y} \mid \boldsymbol{\theta}_k) \\ \pi_k \in \{(\pi_1, \pi_2, \ldots, \pi_M) \mid \pi_k \geqslant 0,\ &\sum_{k=1}^{M} \pi_k = 1\} \end{aligned} \tag{2}$$

where $\boldsymbol{\theta}_k \triangleq (\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k)$ is the mean $\boldsymbol{\mu}_k$ and the variance $\boldsymbol{\Lambda}_k$ of the k-th Gaussian component, and π_k is the mixture coefficient. Because M and the SNR of the MPSK signals are unknown, the total number of components M, the parameter $(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k)$ and the mixture proportion π_k are all unknown. However, it is known that each measurement $\boldsymbol{y}_i$ is generated from only one of the components of the mixture. Hence, the problem is to classify or cluster the received signals with regard to the mixture components that generate them.

The Gaussian mixture model was first proposed to model any continuous density involving problems of density estimation[6]. In this paper, we use it to model the MPSK signals with unknown M and the SNR. To cluster the signals and identify the classification number M, the nonparametric Bayesian inference method is applied, which will be further described in section 2.

2 Clustering Method Based on Nonparametric Bayesian Inference

The Bayesian theorem is represented as

$$P(\boldsymbol{\theta}_k \mid \boldsymbol{y}_k) \propto P(\boldsymbol{y}_k \mid \boldsymbol{\theta}_k) P(\boldsymbol{\theta}_k) \tag{3}$$

where $\boldsymbol{\theta}_k$ is the estimated parameter, $P(\boldsymbol{\theta}_k)$ is its prior distribution, $\boldsymbol{y}_k$ is the observed value, and $P(\boldsymbol{y}_k \mid \boldsymbol{\theta}_k)$ is the likelihood of $\boldsymbol{\theta}_k$. In the classification problem of unknown classification numbers, prior distribution $P(\boldsymbol{\theta}_k)$ cannot be determined previously.

The nonparametric estimation method does not assume any particular parametric form for prior distribution $P(\boldsymbol{\theta}_k)$, but instead replaces it with a general distribution on a family of probability distributions. Many methods involve families of computationally tractable distributions on such prior probability distributions, while one important example is the Dirichlet process (DP)[7]. Using the DP as a nonparametric prior distribution for the parameters of a mixture model, the Dirichlet process mixture (DPM) model is induced. If the mixture model is further assumed to be a Gaussian mixture, it is called a Dirichlet process Gaussian mixture (DPGM) model[8]. The DPGM model allows the number of Gaussian mixture components to grow as the size of the data set grows without assuming a fixed number of components underlying the data.

In this section, the clustering method of nonparametric Bayesian inference with the DPGM prior model is presented.

2.1 Dirichlet distribution and Dirichlet process

Given a probability measure $\boldsymbol{G}_0$ on a measurable space (T, A) and a positive real number α, a probability distribution $\boldsymbol{G}$ distributed according to a DP of base distribution $\boldsymbol{G}_0$ and a scale factor α, denoted as $\boldsymbol{G} \sim \mathrm{DP}(\boldsymbol{G}_0, \alpha)$, satisfies, for any partition $A_1, A_2, \ldots, A_k$ of T and any k,

$$(\boldsymbol{G}(A_1), \ldots, \boldsymbol{G}(A_2)) \sim D(\boldsymbol{G}_0(A_1), \ldots, \boldsymbol{G}_0(A_k), \alpha) \tag{4}$$

where D is a standard Dirichlet distribution[7]. Ref. [9] established that the realizations of a DP are discrete with probability one and admit stick-breaking representation:

$$\boldsymbol{G}(\cdot) = \sum_{k=1}^{\infty} \pi_k \delta_{\boldsymbol{\theta}_k} \tag{5}$$

with $\boldsymbol{\theta}_k \sim \boldsymbol{G}_0, \beta_k \sim \mathrm{B}(1, \alpha)$ and $\pi_k = \beta_k \prod_{j=1}^{k-1} (1 - \beta_j)$, where $\delta_{\boldsymbol{\theta}_k}$ denotes the Dirac delta measure located in $\boldsymbol{\theta}_k$.

The main advantage of the DP is the simplicity of the posterior update using its conjugate property. Let $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n$ be n random samples from $\boldsymbol{G}$, and $\boldsymbol{G} \sim \mathrm{DP}(\boldsymbol{G}_0, \alpha)$, then the posterior distribution of $\boldsymbol{G} \mid \boldsymbol{\theta}_{1:n}$ is also DP:

$$\boldsymbol{G} \mid \boldsymbol{\theta}_{1:n} \sim \mathrm{DP}\left(\frac{\alpha}{\alpha + n}\boldsymbol{G}_0 + \frac{1}{\alpha + n}\sum_{k=1}^{n} \delta_{\boldsymbol{\theta}_k}, \alpha + n\right) \tag{6}$$

Moreover, by integration out $\boldsymbol{G}$, the predictive distributions admit the following Polya urn representation[10]:

$$\boldsymbol{\theta}_{n+1} \mid \boldsymbol{\theta}_n \sim \frac{1}{\alpha + n}\sum_{k=1}^{n} \delta_{\boldsymbol{\theta}_k} + \frac{\alpha}{\alpha + n}\boldsymbol{G}_0 \tag{7}$$

2.2 Normal-inverse-Wishart analysis

If a multivariate Gaussian mean and covariance are both uncertain, the normal-inverse-Wishart (NIW) provides an appropriate conjugate prior. The covariance matrix is assigned an inverse-Wishart prior $\boldsymbol{\Lambda} \sim \mathrm{IW}(\nu, \boldsymbol{\Delta})$. Conditioned on covariance matrix $\boldsymbol{\Lambda}$, the mean $\boldsymbol{\mu}/\boldsymbol{\Lambda} \sim N(\boldsymbol{\vartheta}, \boldsymbol{\Lambda}/\kappa)$. Here, $\boldsymbol{\vartheta}$ is the expected mean, for which we have the quasi observation κ on the scale of observations $\boldsymbol{y}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Lambda})$. The joint prior distribution, denoted by $\mathrm{NIW}(\kappa, \boldsymbol{\vartheta}, \nu, \boldsymbol{\Delta})$, takes the following form[11]:

$$p(\boldsymbol{\mu}, \boldsymbol{\Lambda} \mid \kappa, \boldsymbol{\vartheta}, \nu, \boldsymbol{\Delta}) \propto |\boldsymbol{\Lambda}|^{-(\frac{\nu+d}{2}+1)} \cdot \exp\left\{-\frac{1}{2}\mathrm{tr}(\nu\boldsymbol{\Delta}\boldsymbol{\Lambda}^{-1}) - \frac{\kappa}{2}(\boldsymbol{\mu} - \boldsymbol{\vartheta})^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}(\boldsymbol{\mu} - \boldsymbol{\vartheta})\right\} \tag{8}$$

where the parameter d is the dimension of the NIW; in this paper, $d = 2$, since modulation classification is based on the signal constellation in the complex plane.

The posterior distribution $p(\boldsymbol{\mu}, \boldsymbol{\Lambda} \mid \boldsymbol{y}_1, \ldots \boldsymbol{y}_i, \kappa, \boldsymbol{\vartheta}, \nu, \boldsymbol{\Delta})$ is also NIW, via conjugacy, and thus compactly described by a set of updated hyperparameters NIW $(\bar{\kappa}, \bar{\boldsymbol{\vartheta}}, \bar{\nu}, \bar{\boldsymbol{\Delta}})$. These posterior hyperparameters are equal to[11]

$$\bar{\kappa}\bar{\boldsymbol{\vartheta}} = \kappa\boldsymbol{\vartheta} + \sum_{i=1}^{n} \boldsymbol{y}_i \qquad \bar{\kappa} = \kappa + n \tag{9}$$

$$\bar{\nu}\bar{\boldsymbol{\Delta}} = \nu\boldsymbol{\Delta} + \sum_{i=1}^{N} \boldsymbol{y}_i\boldsymbol{y}_i^{\mathrm{T}} + \kappa\boldsymbol{\vartheta}\boldsymbol{\vartheta}^{\mathrm{T}} - \bar{\kappa}\bar{\boldsymbol{\vartheta}}\,\bar{\boldsymbol{\vartheta}}^{\mathrm{T}} \qquad \bar{\nu} = \nu + n \tag{10}$$

Integrating the parameters of the NIW posterior distribution, the predictive likelihood of a new observation $\boldsymbol{y}_{n+1}$ can be approximated by[11]

$$p(\boldsymbol{y}_{n+1} \mid \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n, \kappa, \boldsymbol{\vartheta}, \nu, \boldsymbol{\Delta}) \approx N\left(\boldsymbol{y}_{n+1}; \bar{\boldsymbol{\vartheta}}, \frac{(\bar{\kappa}+1)\bar{\nu}}{\bar{\kappa}(\bar{\nu} - d - 1)}\bar{\boldsymbol{\Delta}}\right) \tag{11}$$

3 Clustering Algorithm Description

The Gaussian mixture model combined with the DP prior is called the DPGM model. Because the number of Gaussian components and their mean and covariance are all unknown. From the analysis above, the base distribution of the mixture model $\boldsymbol{G}_0$ can be defined as NIW. The hierarchical DPGM model can be described as

$$\boldsymbol{G} \sim \mathrm{DP}(\boldsymbol{G}_0, \alpha), \quad \boldsymbol{\theta}_k \mid \boldsymbol{G} \sim \boldsymbol{G}, \quad \boldsymbol{y}_k \mid \boldsymbol{\theta}_k \sim N(\cdot \mid \boldsymbol{\theta}_k) \tag{12}$$

where $\boldsymbol{G}_0 = \mathrm{NIW}(\kappa_0, \boldsymbol{\vartheta}_0, \nu_0, \boldsymbol{\Delta}_0)$.

The calculation of the posterior probability $p(\boldsymbol{\theta}_{1:n} \mid \boldsymbol{y}_{1:n})$ in DPGMM is too complex to obtain a close-form solution. The MCMC method using the Gibbs sampler can provide numerical approximation[12]. The Gibbs sampler needs to sample from $P(\boldsymbol{\theta}_i \mid \boldsymbol{\theta}^{(i-)}, \boldsymbol{y}_i)$, where $\boldsymbol{\theta}^{(i-)} \triangleq \boldsymbol{\theta}_{1:n\backslash i}$. According to Bayesian equation:

$$P(\boldsymbol{\theta}_i \mid \boldsymbol{\theta}^{(i-)}, \boldsymbol{y}_i) \propto P(\boldsymbol{y}_i \mid \boldsymbol{\theta}_i) P(\boldsymbol{\theta}_i \mid \boldsymbol{\theta}^{(i-)}) \tag{13}$$

Following Eq. (6), prior probability $\boldsymbol{\theta}_i$ equals

$$P(\boldsymbol{\theta}_i = \boldsymbol{\theta} \mid \boldsymbol{\theta}^{(i-)}) = \frac{1}{\alpha + n - 1}\sum_{\substack{j=1 \\ j \neq i}}^{n} \delta_{\boldsymbol{\theta}_j = \boldsymbol{\theta}} + \frac{\alpha}{\alpha + n - 1}\boldsymbol{G}_0 \tag{14}$$

The likelihood function $f(\boldsymbol{y}_i \mid \boldsymbol{\theta}_i)$ is presented by Eq. (11).

The whole algorithm is described as follows:

Algorithm 1 Calculating posterior density $p(\boldsymbol{\theta}_{1:n} \mid \boldsymbol{y}_{1:n})$

1) Initialization

Set DP's scale factor α;

Set $(\kappa_0, \boldsymbol{\vartheta}_0, \nu_0, \boldsymbol{\Delta}_0)$ to define the basic distribution $\boldsymbol{G}_0$; set initial cluster and number N in the cluster.

2) Iteration

for $n = 2, 3, \ldots, N_{\text{iter}}$

Take $\boldsymbol{y}_i$ from N observations sequentially and the number in the cluster which $\boldsymbol{y}_i$ belongs to minus one;

Add a new cluster value $\boldsymbol{\theta}_i$ from sampling $\boldsymbol{G}_0$;

Calculate the value in all clusters: $\bar{\kappa}\bar{\boldsymbol{\vartheta}}$(Eq. (9)), $\bar{\nu}\bar{\boldsymbol{\Delta}}$(Eq. (10)) and variance $\dfrac{(\bar{\kappa}+1)\bar{\nu}}{\bar{\kappa}(\bar{\nu}-d-1)}\bar{\boldsymbol{\Delta}}$;

Calculate $P(\boldsymbol{y}_i \mid \boldsymbol{\theta}_i)$ of all clusters(Eq. (11));

Calculate the prior $P(\boldsymbol{\theta}_i \mid \boldsymbol{\theta}^{(i-)})$ (Eq. (14));

Calculate posterior probability $P(\boldsymbol{\theta}_i \mid \boldsymbol{\theta}^{(i-)}, \boldsymbol{y}_i)$ (Eq. (13));

Sample the posterior probability distribution, determine the cluster which $\boldsymbol{y}_i$ belongs to, update $\bar{\kappa}\bar{\boldsymbol{\vartheta}}, \bar{\nu}\bar{\boldsymbol{\Delta}}$ of this cluster and update the number of observations in this cluster;

End.

4 Simulation Results and Analysis

In this section, simulation results are given and analyzed. The SNR changes from 0 to 15 dB. Initial hyperparameters are set at small values to reflect the non-informative prior distribution: $\alpha = 4$, $\kappa_0 = 0.1$, $\nu_0 = 6$. We first assume that all the received signals belong to one cluster and $\boldsymbol{\vartheta}, \boldsymbol{\Delta}_0$ are set as the overall mean and variance in this cluster. The initial cluster number $M = 1$, the simulation symbol number $N = 1\,600$, and the iteration number $N_{\text{iter}} = 200$.

Figs. 1 to 3 show the clustering results. One cluster is marked with the same symbol. It can be seen from the results that, although signals in the overlapped region are classified into wrong clusters, the final clustering number M is not effected. So the correct clustering number can be obtained under such SNR conditions, and the PSK signals are correctly identified.

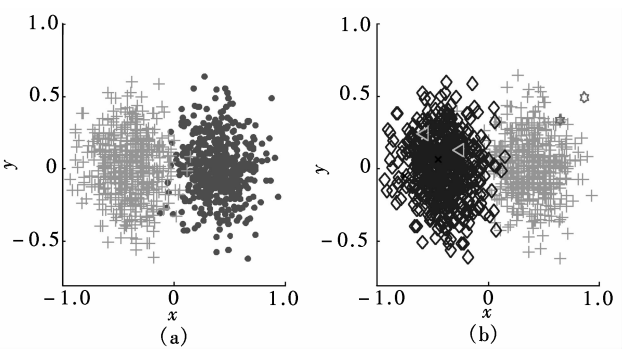

Fig. 1 Clustering results at 3 dB, $M = 2$. (a) BPSK generative signals; (b) BPSK clustering results

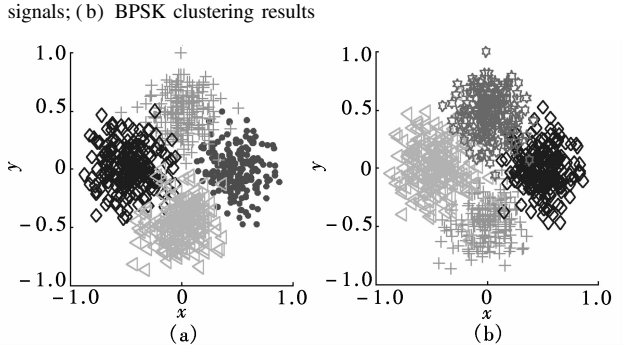

Fig. 2 Clustering results at 6 dB, $M = 4$. (a) 4PSK generative signals; (b) 4PSK clustering results

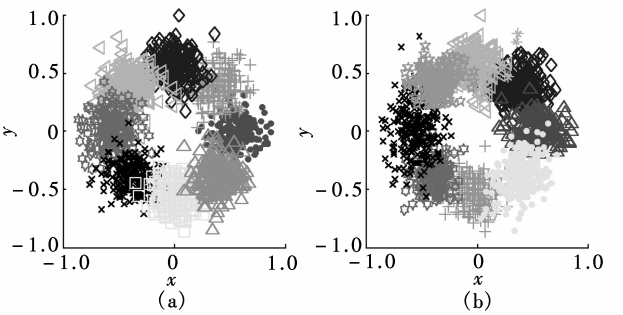

Fig. 3 Clustering results at 10 dB, $M = 8$. (a) 8PSK generative signals; (b) 8PSK clustering results

In order to evaluate the identification probability, 100 Monte Carlo experiments are independently done for every 2/4/8PSK signals under the SNR ranging from 0 to 15 dB. The decision rule is as follows: the clustering number $N_{\text{cluster}} \leqslant 3$, 2PSK is decided; $N_{\text{cluster}} \in [4, 7]$, 4PSK is decided; $N_{\text{cluster}} \geqslant 8$, 8PSK is decided. The result is shown in Fig. 4. It can be seen that the identification probability of these three signals is above 95% when the SNR > 5 dB.

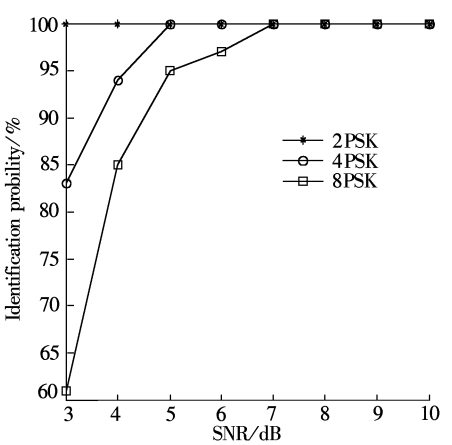

Fig. 4 Identification probability vs. SNR for 2/4/8PSK

We further compare the average identification probability of the proposed method with that of the method in Ref. [13]. In Ref. [12], clustering loss functions are calculated for $M = 2$ and 4, respectively. And the classification number M is decided according to the inflexion of the loss function. This method is inflexible because it must calculate all the loss functions of different M when M is large. Additionally, it is difficult to determine the value of M because the function inflexion is blurry in low SNR. In the proposed algorithm, the signals are clustered by the nonparametric Bayesian inference, which are not assumed to be a fixed number of components, and the clustering value M is automatically involved by the inference results according to the observed values. From Fig. 5, it is clear that the proposed method improves the average identification probability of 2/4PSK, compared with the method in Ref. [13].

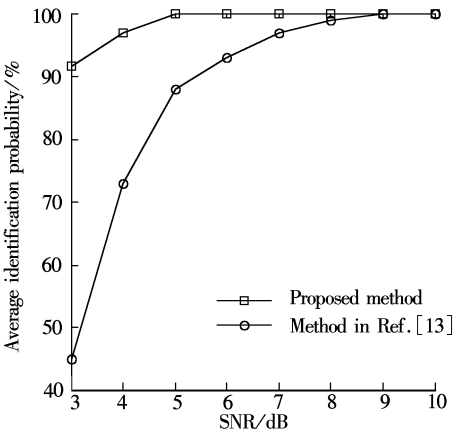

Fig. 5 Average identification comparison for 2/4PSK signals

5 Conclusion

In this paper, the MPSK signals with unknown SNR are modeled as a Gaussian mixture signal set with unknown means and covariances. A clustering method is proposed to classify MPSK signals, which is based on the DPGM model, inferred by the nonparametric Bayesian method, and calculated by the Gibbs sampling algorithm. This method is an automatic learning algorithm and the classification number is adaptively involved according to observation values. Simulation results show that the correct recognition ratio of 2/4/8PSK is greater than 95% under the condition that SNR > 5 dB and 1 600 symbols are used.

References

[1] Mobasseri B G. Constellation shape as a robust signature for digital modulation recognition [C]//*Proceedings of IEEE Military Communications Conference*. Atlantic City, NJ, USA, 1999: 442 - 446.

[2] Wong M L D, Nandi A K. Semi-blind algorithms for automatic classification of digital modulation schemes[J]. *Digital Signal Processing*, 2008, **18**(2): 209 - 227.

[3] Dobre O A, Abdi A, Bar-Ness Y, et al. Survey of automatic modulation classification techniques: classical approaches and new trends [J]. *IET Communications*, 2007, **1**(2): 137 - 156.

[4] Schreyogg C. Identification of voice band data signal constellations using a divisive cluster algorithm [C]//*Proceedings of IEEE Digital Processing Workshop*. Loen, Norway, 1996: 474 - 477.

[5] Mobasseri B G. Digital modulation classification using constellation shape [J]. *Signal Processing*, 2000, **80**(2): 251 - 277.

[6] McLachlan G J, Basford K E. *Mixture models*: *inference and applications to clustering* [M]. New York: Marcel Dekker, 1988.

[7] Ferguson T S. A Bayesian analysis of some non-parametric problems [J]. *The Annals of Statistics*, 1973, **1**(2): 209 - 230.

[8] Antoniak C E. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems [J]. *The Annals of Statistics*, 1974, **2**(6): 1152 - 1174.

[9] Sethuraman J. A constructive definition of Dirichlet priors [J]. *Statistica Sinica*, 1994, **4**(2): 639 - 650.

[10] Blackwell D, MacQueen J B. Ferguson distributions via Polya urn schemes [J]. *The Annals of Statistics*, 1973, **1**(2): 353 - 355.

[11] Gelman A, Carlin J B, Stern H S, et al. *Bayesian data analysis* [M]. London: Chapman & Hall, 2004.

[12] Neal R M. Markov chain sampling methods for Dirichlet process mixture models [J]. *Journal of Computational and Graphical Statistics*, 2000, **9**(2): 249 - 265.

[13] Ruo Ming, Yang Shaoquan. Modulation classification of MPSK signals based on dynamic clustering [J]. *Journal of Circuit and Systems*, 2005, **10**(2): 83 - 87. (in Chinese)

基于非参数贝叶斯推断的 MPSK 信号调制识别

陈 亮[1,2] 程汉文[1] 吴乐南[1]

([1] 东南大学信息科学与工程学院,南京 210096)
([2] 江苏大学机械工程学院,镇江 212013)

摘要:依据星座图采用非参数贝叶斯方法对多元相移键控(MPSK)信号进行调制识别. 将未知信噪比(SNR)水平的 MPSK 信号看成复平面内多个未知均值和方差的高斯分布依照一定的比例混合而成,利用非参数贝叶斯推断方法进行密度估计,实现对 MPSK 信号分类目的. 推断过程中,引入 Dirichlet 过程作为混合比例因子的先验分布,结合正态逆 Wishart(NIW)分布作为均值和方差的先验分布,根据接收信号,利用 Gibbs 采样的 MCMC (Monte Carlo Markov chain)随机采样算法,不断调整混合比例因子、均值和方差. 通过多次迭代,得到对调制信号的密度估计. 仿真表明,在 SNR > 5 dB,码元数目大于 1 600 时,2/4/8PSK 的识别率超过了 95%.

关键词:调制分类;多元相移键控;Dirichlet 过程;非参数贝叶斯推断;Monte Carlo Markov chain
中图分类号:TN911