

# 一种集成式Beta过程最大间隔一类分类方法

张维 杜兰\*

(西安电子科技大学雷达信号处理国家重点实验室 西安 710071)

**摘要:** 一类分类是一种将目标类样本和其他所有的非目标类样本区分开的分类方法。传统的一类分类方法针对所有训练样本建立一个分类器, 忽视了数据的内在结构, 在样本分布复杂时, 其分类性能会严重下降。为了提升复杂分布情况下的分类性能, 该文提出一种集成式Beta过程最大间隔一类方法。该方法利用Dirichlet过程混合模型(DPM)对训练样本聚类, 同时在每一个聚类学习一个Beta过程最大间隔一类分类器。通过多个分类器的集成, 可以构造出一个描述能力更强的分类器, 提升复杂分布下的分类效果。DPM聚类模型和Beta过程最大间隔一类分类器在同一个贝叶斯框架下联合优化, 保证了每一个聚类样本的可分性。此外, 在Beta过程最大间隔一类分类器中, 加入了服从Beta过程先验分布的特征选择因子, 从而可以降低特征冗余度以及提升分类效果。基于仿真数据、公共数据集和实测SAR图像数据的实验结果证明了所提方法的有效性。

**关键词:** 雷达信号处理; 一类分类; Dirichlet过程; Beta过程

中图分类号: TN957.51

文献标识码: A

文章编号: 1009-5896(2021)05-1219-09

DOI: 10.11999/JEIT200080

## An Ensembling One-class Classification Method Based on Beta Process Max-margin One-class Classifier

ZHANG Wei DU Lan

(National Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China)

**Abstract:** In the problem of one-class classification, One-Class Classifier (OCC) tries to identify samples of a specific class, called the target class, among samples of all other classes. Traditional one-class classification methods design a classifier using all training samples and ignore the underlying structure of the data, thus their classification performance will be seriously degraded when dealing with complex distributed data. To overcome this problem, an ensembling one-class classification method based on Beta process max-margin one-class classifier is proposed in this paper. In the method, the input data is partitioned into several clusters with the Dirichlet Process Mixture (DPM), and a Beta Process Max-Margin One-Class Classifier (BPMMOCC) is learned in each cluster. With the ensemble of some simple classifiers, the complex nonlinear classification can be implemented to enhance the classification performance. Specifically, the DPM and BPMMOCC are jointly learned in a unified Bayesian frame to guarantee the separability in each cluster. Moreover, in BPMMOCC, a feature selection factor, which obeys the prior distribution of Beta process, is added to reduce feature redundancy and improve classification results. Experimental results based on synthetic data, benchmark datasets and Synthetic Aperture Radar (SAR) real data demonstrate the effectiveness of the proposed method.

**Key words:** Radar signal processing; One-Class Classification (OCC); Dirichlet process; Beta process

### 1 引言

一类分类是一种将感兴趣的目标类样本和其他

所有不感兴趣的非目标样本区分开的分类方法<sup>[1]</sup>。和传统两类分类需要完备的两类训练样本不同, 一类分类器在训练时只需要目标类样本, 因此可以解决在类样本数据不均衡或某一类样本缺失时, 两类分类器无法训练的问题。近年来, 一类分类器广泛用于各种任务中, 如数据异常检测<sup>[2]</sup>、机场异物检测<sup>[3]</sup>、网络流量异常检测<sup>[4]</sup>、异常行为检测<sup>[5]</sup>、虚拟网络切片异常检测<sup>[6]</sup>以及高光谱图像异常检测<sup>[7]</sup>等。

收稿日期: 2020-01-19; 改回日期: 2020-11-12; 网络出版: 2020-11-18

\*通信作者: 杜兰 dulan@mail.xidian.edu.cn

基金项目: 国家自然科学基金(61771362), 高等学校学科创新引智计划(B18039), 陕西省重点科技创新团队计划

Foundation Items: The National Natural Science Foundation of China (61771362), The 111 Project (B18039), Shaanxi Innovation Team Project

一般来说,目前常见的一类分类方法可以划分为5类<sup>[8]</sup>:(1)基于概率的方法;(2)基于距离的方法;(3)基于重构的方法;(4)基于域的方法;(5)基于信息论的方法。基于概率的方法假设目标类样本服从某一种概率分布,并利用训练样本估计目标类样本的概率密度函数。当测试样本的概率密度函数值大于某一阈值时,就认为该测试样本是目标类样本。常见的概率密度函数有高斯分布、混合高斯分布和学生氏分布等。基于距离的方法一般通过定义一个距离度量函数来衡量两个样本之间的相似度,当测试样本和训练样本集的最小距离度量小于某个阈值时,认为该测试样本是目标类样本。均值聚类<sup>[9]</sup>(K-means clustering)就是一种典型的基于距离的一类分类方法。基于重构的方法首先利用训练样本学习一个目标类样本的表示模型,然后将样本在该表示模型下的重构误差作为判断准则,即重构误差小于某一阈值时,样本属于目标类。自编码(Auto-Encoder, AE)模型<sup>[9]</sup>就是一种常用的基于重构的一类分类方法。基于域的方法通过训练目标样本学习了目标类样本的分布边界,即目标类样本的分布范围,当测试样本位于该范围之内时,该样本为目标类,反之则为非目标类。一类支持向量机<sup>[10]</sup>(One-Class Support Vector Machine, OCSVM)和一类支持向量数据描述<sup>[11]</sup>(One-Class Support Vector Data Description, OC-SVDD)就是两种经典的基于域的一类分类方法。基于信息论的方法通过定义一个信息量,如熵、相对熵,来度量两个样本之间的相似度,且信息量越大,两个样本之间的相似度越小。当测试样本与训练目标样本集的最大信息量小于某个阈值时,该样本是目标类样本。

虽然上述的一类分类方法在很多应用中都取得了不错的效果,但是这些方法针对整个训练样本集仅仅建立了一个分类器,在样本分布复杂时,例如样本分布为多模分布,这些分类器的分类性能会不同程度地下降。为了提升复杂分布下的分类性能,本文提出了一种基于Beta过程最大间隔一类分类器的集成式一类分类方法。该方法利用Dirichlet过程混合模型(Dirichlet Process Mixture, DPM)对训练样本聚类,并在每个聚类学习一个Beta过程最大间隔一类分类器。通过多个分类器的集成,可以构造一个描述能力更强的分类器,从而提升复杂分布下的一类分类性能。DPM聚类模型和分类模型在同一个贝叶斯框架下联合优化,保证了每一个聚类的可分性。此外,在Beta过程最大间隔一类分类器中,加入了服从Beta过程先验分布的特征选择因子,从而可以选择可分性最好的特征,降低特征冗

余度以及提升鉴别效果。仿真数据、公共数据集和实测SAR图像数据的实验结果证明:相比于其他相关的一类分类器,所提方法具有更好的分类效果。

## 2 集成式Beta过程最大间隔一类分类器

### 2.1 Dirichlet过程(DP)和DPM模型

DP是在文献<sup>[12]</sup>中于1973年提出的一种用于非参数贝叶斯模型的随机过程。若 $G$ 服从DP,那么 $G$ 可以表示成

$$G \sim \text{DP}(G_0, \alpha) \quad (1)$$

其中, $G_0$ 是基分布, $\alpha$ 是聚集参数。然而采用DP聚类时,只能将具有相同特征值的样本聚为一类,这样的聚类方式不合理。因此,文献<sup>[13]</sup>提出了DPM模型。式(2)是DPM的截断stick-breaking表示方式,其定义为

$$\left. \begin{aligned} v_c &\sim \text{Beta}(v_c; 1, \alpha), \{\Theta_c\}_{c=1}^C \sim G_0 \\ \pi_c(\mathbf{v}) &= v_c \prod_{j=1}^{c-1} (1 - v_j), z_i \sim \text{Multi}(z_i; \boldsymbol{\pi}) \\ \mathbf{x}_i | z_i, \{\Theta_c\}_{c=1}^C &\sim F(\mathbf{x}_i; \Theta_{z_i}), i = 1, 2, \dots, N \end{aligned} \right\} \quad (2)$$

其中, $\text{Beta}(v_c; 1, \alpha)$ 表示 $v_c$ 是服从参数为1和 $\alpha$ 的Beta分布的随机变量, $\Theta_c$ 是服从基分布 $G_0$ 的随机变量, $z_i$ 是样本 $\mathbf{x}_i$ 的聚类指示因子, $\text{Multi}(z_i; \boldsymbol{\pi})$ 表示 $z_i$ 是服从参数为 $\boldsymbol{\pi}$ 的多项式分布, $F(\mathbf{x}_i; \Theta_{z_i})$ 表示第 $z_i$ 个聚类样本 $\mathbf{x}_i$ 服从参数为 $\Theta_{z_i}$ 的 $F$ 分布, $C$ 表示最大的聚类个数。

### 2.2 Beta过程最大间隔一类分类器

OCSVM作为一种基于域的一类分类方法,寻找的是这样一个分类超平面来确定目标类样本的边界:(1)所有的训练样本都位于超平面的上方;(2)同时原点到超平面的距离最大。原点到超平面的最大距离称为“最大间隔”,因此OCSVM也称为最大间隔一类分类器。OCSVM的优化问题如式(3)

$$\left. \begin{aligned} \min_{\mathbf{w}, \xi_i, \rho} & \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{\eta N} \sum_i \xi_i - \rho \\ \text{s.t.} & (\mathbf{w}^T \phi(\mathbf{x}_i)) \geq \rho - \xi_i, \xi_i \geq 0, \forall i \end{aligned} \right\} \quad (3)$$

其中, $\mathbf{w}$ 是超平面的斜率, $\rho$ 为超平面的截距, $\xi_i$ 为松弛变量, $N$ 为训练样本数, $\eta$ 为超参数, $\phi(\cdot)$ 为高斯核变换,且 $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle = \exp\left(-\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 / 2\sigma^2\right)$ 。

超平面 $f(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) - \rho = 0$ 和 $\hat{f}(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) - 1 = 0$ 唯一的区别就是超平面 $f(\mathbf{x})$ 可以经过原点,而 $\hat{f}(\mathbf{x})$ 不行。但是在OCSVM分类器中,学习的超平面是不能经过原点的,因此可以用 $\hat{f}(\mathbf{x})$ 来代替 $f(\mathbf{x})$ 从而减少需要学习的参数。因此优化问题式(3)可以转为优化问题式(4)

$$\left. \begin{aligned} \min_{\mathbf{w}, \xi_i} & \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{\eta N} \sum_i \xi_i - 1 \\ \text{s.t.} & (\mathbf{w}^T \phi(\mathbf{x}_i)) \geq 1 - \xi_i, \xi_i \geq 0, \forall i \end{aligned} \right\} \quad (4)$$

进一步, 优化问题式(4)可以改写成无约束条件的形式

$$\min_{\mathbf{w}} d(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2 + 2l \sum_i \max(1 - \mathbf{w}^T \phi(\mathbf{x}_i), 0) \quad (5)$$

其中,  $l = 1/\eta N$ 。

根据文献[14]的思想, 采用数据增强技术[15]引入隐变量 $\lambda$ , 那么最优参数 $\mathbf{w}_{\text{opt}}$ 可表示为

$$\mathbf{w}_{\text{opt}} = \int \int p(\mathbf{w}, \lambda | \phi(\mathbf{x})) d\lambda d\mathbf{w} \quad (6)$$

其中,  $p(\mathbf{w}, \lambda | \phi(\mathbf{x})) \propto p(\phi(\mathbf{x}), \lambda | \mathbf{w}) p(\mathbf{w}) \propto \prod_i \exp\{-\frac{(l - l\mathbf{w}^T \phi(\mathbf{x}_i) + \lambda_i)^2}{2\lambda_i}\} / \sqrt{2\pi\lambda_i} \times \exp(-\mathbf{w}^T \mathbf{w}/2)$ 。

式(6)的积分可以采用马尔可夫链蒙特卡罗(Markov Chain Monte Carlo, MCMC)算法结合吉布斯采样[16], 就可以得到最优的参数 $\mathbf{w}_{\text{opt}}$ 。然而, 在式(6)中, 高斯核变换 $\phi(\mathbf{x}_i)$ 没有显式的表达, 导致参数无法求解。为了解决这个问题, 本文提出了一种新的特征变换方式:

$$\phi(\mathbf{x}_i) = \tilde{\mathbf{x}}_i = [\kappa(\mathbf{x}_i, \mathbf{x}_1), \kappa(\mathbf{x}_i, \mathbf{x}_2), \dots, \kappa(\mathbf{x}_i, \mathbf{x}_N)] \quad (7)$$

从高斯核函数的表达式可以看出, 如果两个样本 $\mathbf{x}_i$ 和 $\mathbf{x}_j$ 相近, 那么 $\kappa(\mathbf{x}_i, \mathbf{x}_j)$ 的值会接近1, 反之 $\kappa(\mathbf{x}_i, \mathbf{x}_j)$ 会趋近于0。然而, 从式(7)可以看出, 如果训练样本数量较大, 那么变换后样本特征维度会很高。为了降低变换后特征的冗余度以及选择最具可分性的特征, 本文加入了一个二值特征选择因子变量 $\gamma = \{\gamma_1, \gamma_2, \dots, \gamma_N\}$ , 其中 $\gamma_i \sim \text{bern}(\pi_i)$ ,  $\pi_i \sim \text{Beta}(a_0, b_0)$ 。bern-Beta分层先验是Beta过程的一种有限近似, 因此二值特征选择因子变量的先验分布可以近似认为服从Beta过程先验。根据稀疏先验思想, 特征选择因子变量 $\gamma = \{\gamma_1, \gamma_2, \dots, \gamma_N\}$ 中的大部分元素会为0, 从而可以降低特征变换后的特征冗余度。另外, 特征选择因子变量和一类分类器的参数联合优化, 可以保证选出最具可分性的特征。因此, 最终的Beta过程最大间隔一类分类器模型为

$$\tilde{\mathbf{x}}_i, \lambda_i | \mathbf{w}, \gamma \propto \frac{1}{\sqrt{2\pi\lambda_i}} \exp\left\{-\frac{(l - l\mathbf{w}^T (\gamma \odot \tilde{\mathbf{x}}_i) + \lambda_i)^2}{2\lambda_i}\right\} \quad (8)$$

### 2.3 集成式Beta过程最大间隔一类分类器

集成式Beta过程最大间隔一类分类器是DPM

模型和Beta过程最大间隔一类分类器的结合。在集成式Beta过程最大间隔一类分类器中, 训练样本首先通过DPM模型聚类, 然后在每个聚类学习一个Beta过程最大间隔一类分类器。在DPM模型中, 假设每个聚类的数据服从高斯分布 $F(\mathbf{x}_i; \Theta_c) = N(\mathbf{x}_i; \mathbf{u}_c, \Sigma_c)$ , 根据共轭先验的思想, 基分布 $G_0$ 为Norm-Wishart(NW)分布 $\{\mathbf{u}_c, \Sigma_c\} \sim \text{NW}(\mathbf{u}_c, \Sigma_c; \mathbf{u}_0, \Sigma_0, \beta_0, \gamma_0)$ 。因此集成的Beta过程最大间隔一类分类器的模型可以表示为

$$\left. \begin{aligned} v_c & \sim \text{Beta}(v_c; 1, \alpha), \{\mathbf{u}_c, \Sigma_c\}_{c=1}^C \\ & \sim \text{NW}(\mathbf{u}_c, \Sigma_c; \mathbf{u}_0, \Sigma_0, \beta_0, \gamma_0) \\ \pi_c(\mathbf{v}) & = v_c \prod_{j=1}^{c-1} (1 - v_j), z_i \sim \text{Multi}(z_i; \boldsymbol{\pi}) \\ \mathbf{x}_i | z_i, \{\mathbf{u}_c, \Sigma_c\}_{c=1}^C & \sim N(\mathbf{x}_i; \mathbf{u}_{z_i}, \Sigma_{z_i}) \\ \{\mathbf{w}_c\}_{c=1}^C & \sim N(\mathbf{w}_c; 0, \mathbf{I}), \gamma_k \sim \text{bern}(\pi_k), \\ \pi_k & \sim \text{Beta}(a_0, b_0) \\ \tilde{\mathbf{x}}_i, \lambda_i | z_i, \{\mathbf{w}_c\}_{c=1}^C, \\ \gamma & \propto \frac{1}{\sqrt{2\pi\lambda_i}} \exp\left\{-\frac{(l - l\mathbf{w}_{z_i}^T (\gamma \odot \tilde{\mathbf{x}}_i) + \lambda_i)^2}{2\lambda_i}\right\} \end{aligned} \right\} \quad (10)$$

其中, 式(9)是DPM聚类模型, 式(10)是Beta过程最大间隔一类分类器模型。图1给出了集成的Beta过程最大间隔一类分类器的图模型, 其中红色虚线框表示聚类模型, 蓝色虚线框表示贝叶斯最大间隔一类分类器模型。从图1可以看出, 聚类模型和分类模型通过聚类标号联系在一起, 从而实现联合优化。图2给出了集成式Beta过程最大间隔一类分类器流程图。

根据式(9)和式(10), 可以得到全数据的伪后验分布形式为

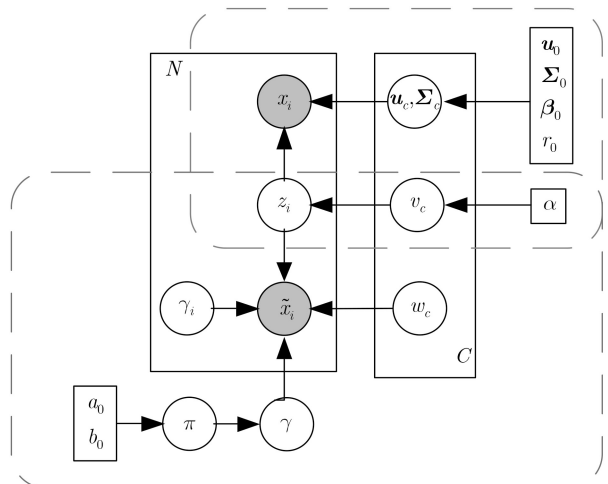


图1 集成式Beta过程最大间隔一类分类器的概率图模型

$$\begin{aligned}
& p\left(\{\mathbf{w}_c\}_{c=1}^C, \{\mathbf{u}_c, \boldsymbol{\Sigma}_c\}_{c=1}^C, \gamma, \pi, \mathbf{v}, \boldsymbol{\lambda}, \mathbf{z} | \mathbf{x}, \tilde{\mathbf{x}}\right) \\
& \propto p\left(\{\mathbf{w}_c\}_{c=1}^C, \{\mathbf{u}_c, \boldsymbol{\Sigma}_c\}_{c=1}^C, \gamma, \pi, \mathbf{v}, \boldsymbol{\lambda}, \mathbf{z}, \mathbf{x}, \tilde{\mathbf{x}}\right) \\
& = p\left(\mathbf{x} | \{\mathbf{u}_c, \boldsymbol{\Sigma}_c\}_{c=1}^C, \mathbf{z}\right) \prod_{c=1}^C p\left(\mathbf{u}_c, \boldsymbol{\Sigma}_c | \mathbf{u}_0, \boldsymbol{\Sigma}_0, \beta_0, \gamma_0\right) \\
& \quad \cdot p\left(\mathbf{z} | \mathbf{v}\right) p\left(\mathbf{v} | \alpha_0\right) \prod_{c=1}^C p\left(\tilde{\mathbf{x}}, \boldsymbol{\lambda} | \mathbf{w}_c, \gamma\right) p\left(\mathbf{w}_c\right) \\
& \quad \cdot p\left(\gamma | \pi\right) p\left(\pi\right) \quad (11)
\end{aligned}$$

根据式(11),可以得到各个参数的条件后验分布,然后利用MCMC算法结合吉布斯采样,就可以得到最终的模型参数。

## 2.4 测试过程

对于一个测试样本 $\mathbf{x}^*$ ,其标签获取的方式如下:首先根据第 $t$ 次采样收集到的参数 $\mathbf{v}^t, \{\mathbf{u}_c^t, \boldsymbol{\Sigma}_c^t\}_{c=1}^C$ ,计算样本的聚类标号 $z_t^*$ ;然后对样本 $\mathbf{x}^*$ 特征变换,得到变换后的测试样本;最后根据第 $t$ 次采样收集到的参数 $\mathbf{w}_{z_t^*}^t$ 和 $\gamma^t$ ,计算最终的类别标签 $y^*$

$$\left. \begin{aligned}
& z_t^* \sim \text{Multi}\left(T_{1,t}^*, T_{2,t}^*, \dots, T_{C,t}^*\right) \\
& T_{c,t}^* = p\left(z_t^* = c | -\right) \propto \pi_c^t\left(\mathbf{v}^t\right) \cdot \text{N}\left(\mathbf{x}^*; \mathbf{u}_c^t, \boldsymbol{\Sigma}_c^t\right) \\
& \tilde{\mathbf{x}}^* = \left[\kappa\left(\mathbf{x}^*, \mathbf{x}_1\right), \kappa\left(\mathbf{x}^*, \mathbf{x}_2\right), \dots, \kappa\left(\mathbf{x}^*, \mathbf{x}_N\right)\right] \\
& y^* = \text{sign}\left(\frac{1}{T - T_{\text{burn-in}}} \sum_{t=T_{\text{burn-in}}+1}^T \right. \\
& \quad \left. \cdot \left(\langle \mathbf{w}_{z_t^*}^t, \left(\gamma^t \odot \tilde{\mathbf{x}}^*\right) \rangle - 1\right)\right)
\end{aligned} \right\} (12)$$

其中, $\mathbf{w}_{z_t^*}^t$ 表示第 $t$ 次采样收集到的第 $z_t^*$ 个聚类中分类器的斜率。至此,本文从理论上对整个集成的Beta过程最大间隔一类分类器的模型建立、模型求解进行了推导,实际应用时,整个分类框架可以总结为训练和测试两个过程,框架如图2所示。训练过程,先对训练样本进行特征提取,再用吉布斯采

样MCMC算法对模型参数进行估计;测试过程,先对测试样本进行特征提取,然后对测试样本的聚类标号进行多次采样,对于每一次采样,确定测试样本所属聚类后,根据该聚类的一类分类器进行分类,最后给出的是多次判决的综合结果。

## 3 实验结果

本节分别在公共数据集和实测SAR图像数据上验证所提方法的有效性。在实验中,吉布斯采样的参数设置为: $T = 6000, T_{\text{burn-in}} = 3000$ ,采样间隔 $T_{\text{space}} = 10$ ,因此能收集到300组采样数据。DPM聚类中的最大聚类个数 $C = 10$ ,Norm-Wishart分布的超参数 $\{\mathbf{u}_c, \boldsymbol{\Sigma}_c\}_{c=1}^C \sim \text{NW}\left(\mathbf{u}_c, \boldsymbol{\Sigma}_c; \mathbf{u}_0, \boldsymbol{\Sigma}_0, \beta_0, \gamma_0\right)$ 设置为 $\text{NW}\left(\mathbf{u}_c, \boldsymbol{\Sigma}_c; \mathbf{u}_x, \boldsymbol{\Sigma}_x, 1, d\right)$ ,其中 $\mathbf{u}_x$ 和 $\boldsymbol{\Sigma}_x$ 为训练样本的均值向量和协方差矩阵, $d$ 为训练样本的原始维度;DPM聚类中Beta分布的超参数设为 $\alpha = 0.5$ 。

### 3.1 仿真toy-data数据集

为了更好地可视化聚类以及分类界面,本文从混合高斯分布中采样了一些2维训练样本。训练样本的分布如图3所示。图4给出了基于DPM模型的聚类结果以及每个聚类的鉴别边界,图中黑色的实线表示每个聚类的鉴别边界。从图4可以看出训练样本自动聚成了6类,这说明DPM模型确实可以自动确定聚类个数且很好地将服从相同分布的样本聚为一类。另外,从6个黑色实线圈可以看出,每个聚类学到的鉴别边界完整地包裹了这个聚类的目标样本。因此,提出模型可以很好地区分属于每个聚类的目标样本和非目标样本,证明了该方法的有效性。

### 3.2 公共数据集

本节实验采用的数据集为从UCI Machine Learning Repository中获取的Benchmark数据

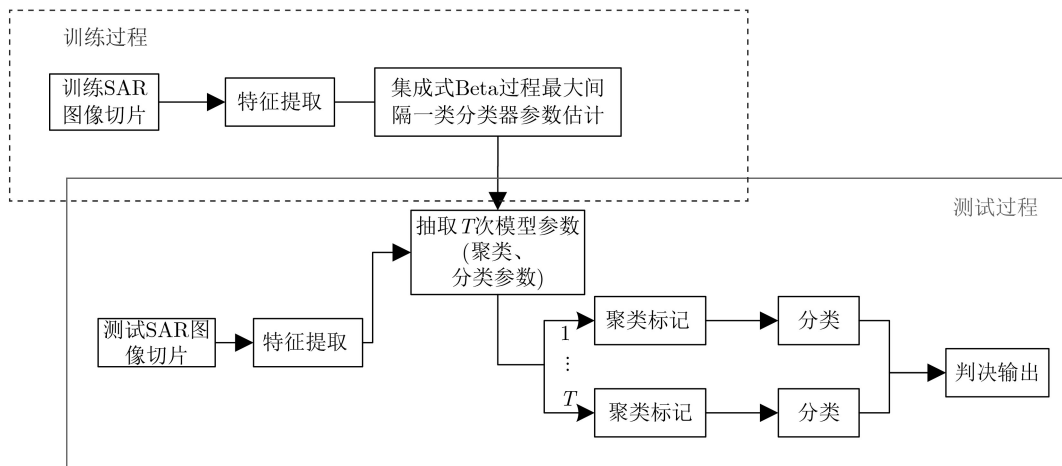


图2 集成式Beta过程最大间隔一类分类器流程图

集，该数据集包含了多种不同特征维数、不同规模的数据。实验所用数据集如表1所示。

实验采用了9种现存的对比方法：主成分分析(PCA)<sup>[17]</sup>，K均值聚类(K-means)<sup>[9]</sup>，线性SVDD(L-SVDD)<sup>[11]</sup>，高斯核形式的一类SVM(KOCSVM)<sup>[10]</sup>，最小生成树(MST)<sup>[18]</sup>，自组织映射(SOM)<sup>[19]</sup>，自编码网络(AE)<sup>[9]</sup>，MPM<sup>[20]</sup>和LPDD<sup>[21]</sup>。另外为了分析提出方法中特征选择因子的影响，我们去掉提出方法中的特征选择因子变量，并将去掉特征选择因子变量后的方法命名为集成式最大间隔一类分类

器，简称En-MMOCC。通过比较En-MMOCC和提出方法，来分析特征选择对实验结果的影响。

本文采用3个指标来描述分类器的性能：鉴别准确率，F1分数和AUC值。通过对训练数据和测试数据进行20次数据划分，然后给出20次随机试验的平均结果。不同方法在5组公共数据集上的实验结果如表2—表5所示。从表2—表5可以看出：在相同特征下，本文提出的分类器在4个数据集上都获得了最高的准确率，F1分数和AUC值，证明所提分类器确实能提高目标鉴别的效果。另外，本文方

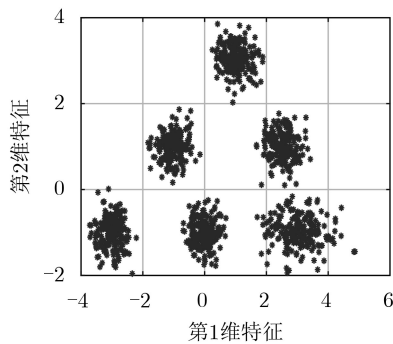


图3 训练样本2维分布图

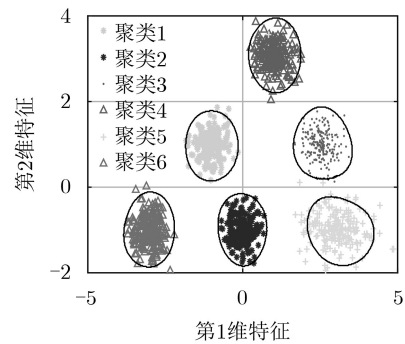


图4 聚类结果和每个聚类鉴别界面可视图

表 1 实验所用公共数据集介绍

数据集	类别数	目标类别	特征维度	训练样本数	测试目标样本数	测试非目标样本数
Australia	2	1	14	215	92	92
Landsat	6	4, 5, 6	36	500	418	1082
Waveform	3	1, 2	21	331	1652	1696
Pageblocks	5	1	10	2456	2457	560

表 2 不同方法在Waveform数据集上的3个指标平均值(%)

指标	L-SVDD	KOCSVM	K-means	PCA	MST	SOM	AE	MPM	LPDD	En-MMOCC	本文方法
准确率	57.05	64.14	66.94	55.26	61.19	67.74	58.00	72.44	74.25	75.92	<b>77.22</b>
F1分数	67.12	59.94	71.92	66.18	70.70	68.30	56.40	65.66	71.02	74.12	<b>75.11</b>
AUC	71.19	69.63	76.71	59.50	73.57	73.83	61.26	83.98	82.01	83.67	<b>85.00</b>

表 3 不同方法在Landsat数据集上的3个指标平均值(%)

指标	L-SVDD	KOCSVM	K-means	PCA	MST	SOM	AE	MPM	LPDD	En-MMOCC	本文方法
准确率	79.72	85.98	82.40	71.28	73.26	80.06	86.01	85.04	86.92	88.06	<b>90.30</b>
F1分数	71.25	73.25	73.76	63.27	65.86	70.29	71.84	69.95	74.54	77.87	<b>81.79</b>
AUC	91.54	90.88	92.72	80.11	87.25	87.76	92.35	89.96	91.69	92.21	<b>94.18</b>

表 4 不同方法在Australia数据集上的3个指标平均值(%)

指标	L-SVDD	KOCSVM	K-means	PCA	MST	SOM	AE	MPM	LPDD	En-MMOCC	本文方法
准确率	55.41	70.76	69.29	66.06	56.41	61.14	78.61	81.90	67.99	81.41	<b>82.21</b>
F1分数	66.67	67.44	74.69	72.13	67.99	67.33	77.65	75.18	67.54	80.36	<b>81.71</b>
AUC	74.42	73.84	82.53	78.32	66.23	67.84	83.14	80.73	71.73	82.57	<b>84.16</b>

表5 不同方法在Pageblocks数据集上的3个指标平均值 (%)

指标	L-SVDD	KOCSVM	K-means	PCA	MST	SOM	AE	MPM	LPDD	En-MMOCC	本文方法
准确率	89.63	90.69	90.06	89.23	90.19	89.79	90.06	87.74	90.09	91.05	<b>92.21</b>
F1分数	93.36	93.91	93.73	93.17	93.95	93.50	93.67	94.17	93.72	94.15	<b>95.22</b>
AUC	95.69	95.99	95.73	94.68	90.02	95.84	95.69	96.00	95.22	94.99	<b>96.03</b>

法可以看作基于聚类的一类分类方法和基于域的一类分类方法的结合, 相比于基于聚类的一类分类方法K-means以及基于域的一类分类方法KOCSVM, 本文方法的分类性能要远远好于上述两种方法, 也说明了这种结合可以很好地提升一类分类性能。最后, 从表2—表5可以明显看出, En-MMOCC的3个分类指标要明显低于本文方法, 特别是在训练样本数较多的数据集上, 说明了特征选择因子在提出方法中的重要性。

### 3.3 实测SAR图像数据

在本节用实测SAR图像数据来验证所提方法的有效性。实测SAR图像来自Sandia MiniSAR数据集。该数据集包含9幅SAR图像。通过CFAR技术, 在9幅SAR图像中一共获得了248个目标切片以及98个杂波切片。图5给出了部分目标切片和杂波切片的示意图, 其中第1行是目标切片, 第2行是杂波切片。在248个目标切片中, 随机选取150个目标切片作为训练样本集, 剩下的98个目标切片和98个杂波切片构造测试样本集。

和3.2节一样, 随机进行20次实验, 然后计算20次实验平均的鉴别准确率、F1分数和AUC值, 实验结果如表6所示。本文方法的鉴别准确率相比于对比方法的最高值大约有2%的提升, F1分数和

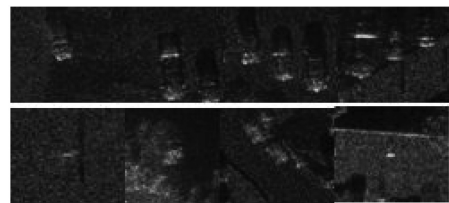


图5 CFAR之后的一些目标和杂波切片示例

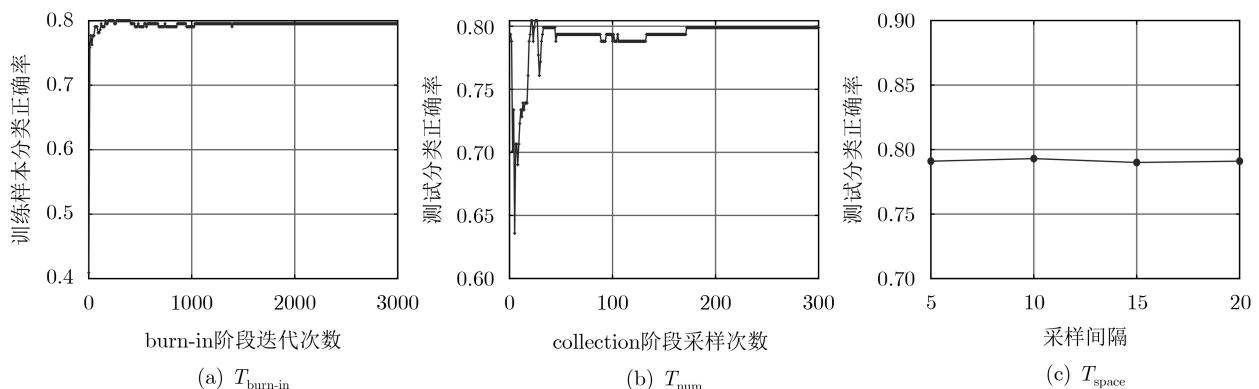
AUC值也是最好的, 说明相比于对比方法, 本文方法不仅有好的鉴别性能, 同时也具有较强的泛化能力。另外, 和UCI数据集上的实验类似, 本文方法的3个指标值也要高于En-MMOCC, 证明了特征选择因子在提出方法中的重要性。

### 3.4 模型超参数分析

(1) 吉布斯采样中的超参数: 吉布斯采样的超参数设置是为了保证算法收敛, 为了验证本文方法中算法的收敛性, 以Australia数据集为例, 在图6中分别给出了分类正确率随超参数 $T_{\text{burn-in}}$ ,  $T_{\text{num}}$ 和 $T_{\text{space}}$ 的变化曲线图。从图6(a)可以看出, 在burn-in阶段的迭代次数 $T_{\text{burn-in}} > 2000$ 后, 训练样本的分类正确率已经稳定, 基本不再变化, 说明模型参数已经稳定。因此, 在实验中将超参数 $T_{\text{burn-in}}$ 设置为3000就能保证在所有数据集上吉布斯采样算法均已收敛。从图6(b)可以看出, 在collection阶段, 当

表6 不同方法在实测SAR数据集上的3个指标平均值 (%)

指标	L-SVDD	KOCSVM	K-means	PCA	MST	SOM	AE	MPM	LPDD	En-MMOCC	本文方法
准确率	82.14	84.72	84.82	73.83	84.13	80.23	83.42	84.67	83.24	85.77	<b>86.12</b>
F1分数	85.05	87.86	86.97	79.11	87.05	82.31	85.98	86.85	85.11	88.00	<b>88.52</b>
AUC	87.30	88.18	88.46	77.99	89.08	89.05	87.65	89.08	88.36	88.35	<b>89.59</b>

图6 Australia数据集上分类正确率随超参数 $T_{\text{burn-in}}$ ,  $T_{\text{num}}$ 和 $T_{\text{space}}$ 的变化曲线图

采样次数 $T_{\text{num}} > 250$ 后，测试分类正确率收敛到一个稳定的值，基本不再变化，说明模型参数已经稳定。因此，当超参数 $T_{\text{num}} > 250$ 时，算法已经收敛，在实验中将超参数 $T_{\text{num}}$ 设置为300。从图6(c)可以看出，随着采样间隔 $T_{\text{space}}$ 的变化，测试分类正确率基本不变。因此，超参数 $T_{\text{space}}$ 对实验结果基本没有影响。采样间隔 $T_{\text{space}}$ 的作用是消除collection阶段相邻两次采样值之间的相关性，一般设置为一个大于1的常数即可。

(2) DPM聚类模型的超参数：根据文献[22]的理论指导，Normal-Wishart分布的超参数 $NW(\mathbf{u}_c, \Sigma_c; \mathbf{u}_0, \Sigma_0, \beta_0, \gamma_0)$ 设置为 $NW(\mathbf{u}_c, \Sigma_c; \mathbf{u}_x, \Sigma_x, 1, d)$ 是最合适的，其中 $\mathbf{u}_x$ 和 $\Sigma_x$ 为训练样本的均值向量和协方差矩阵， $d$ 为训练样本的原始维度。因此，该分布的超参数也是随数据集变化的。对于Beta分布中的超参数 $\alpha$ ，取值范围为 $[0, 1]$ 。由于缺乏关于该参数的先验信息，因此根据相关经验，本文将 $\alpha$ 的取值设置为其对应的中间值，即 $\alpha = 0.5$ 。文献[23]的研究表明，在分层贝叶斯模型中，一些超参数的取值对最终的实验结果影响很小。为了验证该理论，本文以Australia数据集为例，在图7给出了不同 $\alpha$ 取值下的测试分类正确率变化曲线。可以看出，超参数 $\alpha$ 的取值对实验结果基本没有影响，从而验证了文献[23]中理论的正确性。一般来说，DPM中的最大聚类个数 $C$ 可以设置为大于数据

实际聚类个数的任意值。不同的数据集的实际聚类个数是不同的，因此对于不同数据，可以根据经验设置不同的最大聚类个数 $C$ 。

(3) Beta过程最大间隔一类分类器的超参数：在该部分，需要设置的超参数是变量 $\pi_k$ 的先验分布 $\pi_k \sim \text{Beta}(a_0, b_0)$ 中的 $a_0$ 和 $b_0$ 。理论表明： $a_0$ 的值越小， $b_0$ 的值越大，对应的特征选择因子变量 $\gamma$ 中0的个数越多，因此选择的特征越少；反之， $a_0$ 的值越大， $b_0$ 的值越小，对应的特征选择因子变量 $\gamma$ 中1的个数越多，即选择的特征越多。若选择的特征太稀疏，会去掉一些可分性强的特征，降低分类性能；若保留的特征太多，会造成特征冗余，导致“维度灾难”。因此，我们需要设置合适的 $a_0$ 和 $b_0$ 值，既能保留可分性的特征，又能去除冗余的特征。为了分析超参数 $a_0$ 和 $b_0$ 对实验结果的影响，本文以Australia数据集为例，在图8分别给出了 $a_0$ 和 $b_0$ 不同取值下测试分类正确率的曲线图。从图8可以看出，当 $a_0$ 和 $b_0$ 过小或者过大时，都会造成识别率的下降。在实验中，通过交叉验证来确定这两个模型超参数。

### 3.5 时间复杂度分析

本节从训练和测试两个方面比较不同一类分类方法所需的计算时间。以Australia数据集和Landsat数据集为例，表7给出了不同一类分类方法在这两个数据集上所需的计算时间。整体来说，相比于Australia数据集，Landsat数据集训练样本数更多，样本维度也更高，因此，所有一类分类器在Landsat数据集上的计算时间要大于在Australia数据集上的。在训练阶段，L-SVDD, KOCSVM, PCA, K-means, MST, MPM和LPDD的计算时间要小于AE, SOM和本文方法。换句话说，本文方法的计算复杂度和AE以及SOM在同一数量级上。在测试阶段，本文方法和AE的计算复杂度也要大于其他对比方法。由于本文使用的是吉布斯采样来估计后验分布的参数，因此，本文方法的计算负担要大于一些对比方法，这是本文方法的一个不足之处。

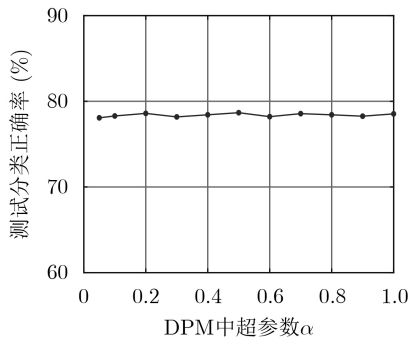
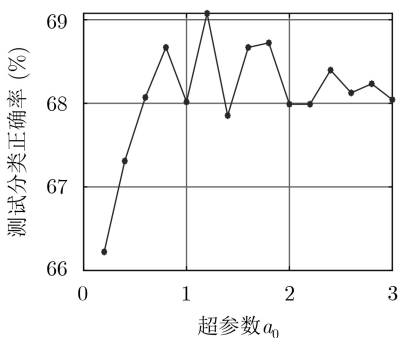
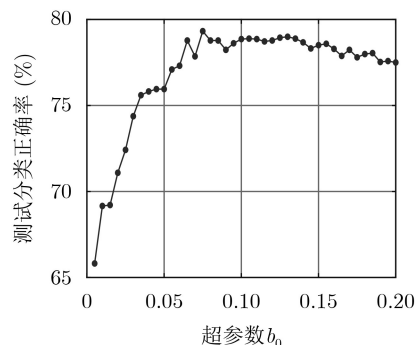


图7 不同 $\alpha$ 取值下的测试分类正确率变化曲线



(a) 不同 $a_0$ 取值



(b) 不同 $b_0$ 取值

图8 不同 $a_0$ 和 $b_0$ 取值下，本文方法在Australia数据集的测试分类正确率曲线

表7 不同一类分类方法所需的计算时间(s)

数据集		L-SVDD	KOCSVM	K-means	PCA	MST	SOM	AE	MPM	LPDD	本文方法
Australia	训练	0.510	0.828	<b>0.290</b>	0.448	0.489	23.086	21.011	0.347	1.122	37.432
	测试	0.097	0.252	0.024	<b>0.013</b>	0.049	0.023	0.277	0.026	0.187	0.535
Landsat	训练	1.572	1.822	<b>0.363</b>	0.512	2.934	122.611	530.921	1.176	10.177	531.735
	测试	0.411	0.513	0.120	<b>0.100</b>	3.349	0.109	22.362	0.232	7.452	5.323

## 4 结束语

为了提升复杂分布下的一类分类性能, 本文提出了一种集成的Beta过程最大间隔一类分类器。所提方法通过DPM模型对训练样本聚类, 同时在每个聚类学习一个Beta过程最大间隔一类分类器。通过多个分类器的集成, 可以构造出一个描述能力更强的强分类器, 从而提升复杂分布下的一类分类性能。DPM模型和Beta过程最大间隔一类分类器在同一个贝叶斯框架下联合优化, 保证每一个聚类的可分性。在一类分类器中, 加入了1个服从Beta过程先验分布的特征选择因子, 选择出最具有可分性的特征, 降低特征冗余度以及提升鉴别效果。仿真数据、公共数据集和实测雷达数据的实验结果表明了本文方法的有效性。

## 参考文献

- [1] TAX D M J. One-class classification: Concept-learning in the absence of counter-examples[D]. [Ph. D. dissertation], Delft University of Technology, 2001: 13–19.
- [2] 王布宏, 罗鹏, 李腾耀, 等. 基于粒子群优化多核支持向量数据描述的广播式自动相关监视异常数据检测模型[J]. 电子与信息学报, 2020, 42(11): 2727–2734. doi: [10.11999/JEIT190767](https://doi.org/10.11999/JEIT190767). WANG Buhong, LUO Peng, LI Tengyao, et al. ADS-B anomalous data detection model based on PSO-MKSVDD[J]. *Journal of Electronics & Information Technology*, 2020, 42(11): 2727–2734. doi: [10.11999/JEIT190767](https://doi.org/10.11999/JEIT190767).
- [3] 王宝帅, 兰竹, 李正杰, 等. 毫米波雷达机场跑道异物分层检测算法[J]. 电子与信息学报, 2018, 40(11): 2676–2683. doi: [10.11999/JEIT180200](https://doi.org/10.11999/JEIT180200). WANG Baoshuai, LAN Zhu, LI Zhengjie, et al. A hierarchical foreign object debris detection method using millimeter wave radar[J]. *Journal of Electronics & Information Technology*, 2018, 40(11): 2676–2683. doi: [10.11999/JEIT180200](https://doi.org/10.11999/JEIT180200).
- [4] 董书琴, 张斌. 基于深度特征学习的网络流量异常检测方法[J]. 电子与信息学报, 2020, 42(3): 695–703. doi: [10.11999/JEIT190266](https://doi.org/10.11999/JEIT190266). DONG Shuqin and ZHANG Bin. Network traffic anomaly detection method based on deep features learning[J]. *Journal of Electronics & Information Technology*, 2020, 42(3): 695–703. doi: [10.11999/JEIT190266](https://doi.org/10.11999/JEIT190266).
- [5] 陈莹, 何丹丹. 基于贝叶斯融合的时空流异常行为检测模型[J]. 电子与信息学报, 2019, 41(5): 1137–1144. doi: [10.11999/JEIT180429](https://doi.org/10.11999/JEIT180429). CHEN Ying, HE Dandan. Spatial-temporal stream anomaly detection based on Bayesian fusion[J]. *Journal of Electronics & Information Technology*, 2019, 41(5): 1137–1144. doi: [10.11999/JEIT180429](https://doi.org/10.11999/JEIT180429).
- [6] 王威丽, 陈前斌, 唐伦. 虚拟网络切片中的在线异常检测算法研究[J]. 电子与信息学报, 2020, 42(6): 1460–1467. doi: [10.11999/JEIT190531](https://doi.org/10.11999/JEIT190531). WANG Weili, CHEN Qianbin, and TANG Lun. Online anomaly detection for virtualized network slicing[J]. *Journal of Electronics & Information Technology*, 2020, 42(6): 1460–1467. doi: [10.11999/JEIT190531](https://doi.org/10.11999/JEIT190531).
- [7] 孟令博, 耿修瑞. 基于协峭度张量的高光谱图像异常检测[J]. 电子与信息学报, 2019, 41(1): 150–155. doi: [10.11999/JEIT180280](https://doi.org/10.11999/JEIT180280). MENG Lingbo and GENG Xiurui. A hyperspectral imagery anomaly detection algorithm based on cokurtosis tensor[J]. *Journal of Electronics & Information Technology*, 2019, 41(1): 150–155. doi: [10.11999/JEIT180280](https://doi.org/10.11999/JEIT180280).
- [8] PIMENTEL M A F, CLIFTON D A, CLIFTON L, et al. A review of novelty detection[J]. *Signal Processing*, 2014, 99: 215–249. doi: [10.1016/j.sigpro.2013.12.026](https://doi.org/10.1016/j.sigpro.2013.12.026).
- [9] LAMPARIELLO F and SCIANDRONE M. Efficient training of RBF neural networks for pattern recognition[J]. *IEEE Transactions on Neural Networks*, 2001, 12(5): 1235–1242. doi: [10.1109/72.950152](https://doi.org/10.1109/72.950152).
- [10] SCHOLKOPF B and SMOLA A J. Learning with Kernels[M]. Cambridge: MIT Press, 2001: 469–513.
- [11] TAX D M J and DUIN R P W. Support vector domain description[J]. *Pattern Recognition Letters*, 1999, 20(11/13): 1191–1199. doi: [10.1016/s0167-8655\(99\)00087-2](https://doi.org/10.1016/s0167-8655(99)00087-2).
- [12] FERGUSON T S. A Bayesian analysis of some nonparametric problems[J]. *Annals of Statistics*, 1973, 1(2): 209–230. doi: [10.1214/aos/1176342360](https://doi.org/10.1214/aos/1176342360).
- [13] BLEI D M and JORDAN M I. Variational inference for Dirichlet process mixtures[J]. *Bayesian Analysis*, 2006, 1(1): 121–143. doi: [10.1214/06-BA104](https://doi.org/10.1214/06-BA104).
- [14] SOLLICH P. Bayesian methods for support vector machines: Evidence and predictive class probabilities[J]. *Machine Learning*, 2002, 46(1–3): 21–52. doi: [10.1023/](https://doi.org/10.1023/)



- a:1012489924661.
- [15] POLSON N G and SCOTT S L. Data augmentation for support vector machines[J]. *Bayesian Analysis*, 2011, 6(1): 1–23. doi: [10.1214/11-BA601](https://doi.org/10.1214/11-BA601).
- [16] GEMAN S and GEMAN D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1984, PAMI-6(6): 721–741. doi: [10.1109/TPAMI.1984.4767596](https://doi.org/10.1109/TPAMI.1984.4767596).
- [17] WOLD S, ESBENSEN K, and GELADI P. Principal component analysis[J]. *Chemometrics and Intelligent Laboratory Systems*, 1987, 2(1/3): 37–52. doi: [10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9).
- [18] JUSZCZAK P, TAX D M J, PEKALSKA E, *et al.* Minimum spanning tree based one-class classifier[J]. *Neurocomputing*, 2009, 72(7/9): 1859–1869. doi: [10.1016/j.neucom.2008.05.003](https://doi.org/10.1016/j.neucom.2008.05.003).
- [19] KOHONEN T and SOMERVUO P. Self-organizing maps of symbol strings[J]. *Neurocomputing*, 1998, 21(1/3): 19–30. doi: [10.1016/s0925-2312\(98\)00031-9](https://doi.org/10.1016/s0925-2312(98)00031-9).
- [20] LANCKRIET G R G, EL GHAOU L, and JORDAN M I. Robust novelty detection with single-class MPM[C]. The 15th International Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, 2002: 929–936.
- [21] PEKALSKA E, TAX D M J, and DUIN R P W. One-class LP classifier for dissimilarity representations[C]. Proceedings of the 15th International Conference on Neural Information Processing Systems, British Columbia, Canada, 2002: 777–784.
- [22] DUNSON D B and PARK J H. Kernel stick-breaking processes[J]. *Biometrika*, 2008, 95(2): 307–323. doi: [10.1093/biomet/asn012](https://doi.org/10.1093/biomet/asn012).
- [23] KRUSCHKE J K. Bayesian data analysis[J]. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2010, 1(5): 658–676. doi: [10.1002/wcs.72](https://doi.org/10.1002/wcs.72).
- 张 维: 男, 1992年生, 博士生, 研究方向为机器学习及其在雷达目标识别方面的应用.
- 杜 兰: 女, 1980年生, 教授, 博士生导师, 研究方向为统计信号处理、雷达信号处理、机器学习及其在雷达目标检测与识别方面的应用.

责任编辑: 马秀强