一种基于 Dirichelt 过程隐变量支撑向量机模型的目标识别方法

张学峰 陈 渤* 王鹏辉 刘宏伟
(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：在目标识别中，对于样本数较多且分布复杂的数据，若将所有训练样本用来训练一个单一的分类器，会增加分类器的训练复杂度，且容易忽视样本的内在结构，不利于分类。因此人们提出了混合专家系统(ME)，即将训练样本集划分为多个训练样本子集，并在每个子集上单独训练分类器。但是传统 ME 系统需要人为确定专家个数，并且每个子集的学习独立于后端的任务，如分类。该文提出一种基于 Dirichlet 过程(DP)混合隐变量(LV)支持向量机(SVM)模型(DPLVSVM)的目标识别算法，采用 DP 混合模型自动确定样本聚类个数，同时每个聚类中使用线性隐变量 SVM(LVSVM)进行分类。不同于以往算法，DPLVSVM 将聚类过程和分类器的训练过程联合优化，保证了各个子集中样本的分布上的一致性和可分性，而且可以利用 Gibbs 采样技术对模型参数进行简便有效的估计。基于人工数据集、公共数据集以及雷达实测数据的实验验证了该文方法的有效性。

关键词：目标识别；混合专家系统；Dirichlet 过程混合模型；隐变量支持向量机分类器
中图分类号： TN957.51 **文献标识码**： A **文章编号**：1009-5896(2015)01-0029-08
DOI: 10.11999/JEIT140129

A Target Recognition Method Based on Dirichlet Process Latent Variable Support Vector Machine Model

Zhang Xue-feng Chen Bo Wang Peng-hui Liu Hong-wei
(*National Laboratory of Radar Signal Processing, Xidian University, Xi'an* 710071, *China*)

Abstract: In target recognition community, when dealing with large-scale and complex distributed data, it is very expensive to train a classifier using all input data and the underlying structure of the data is ignored. To overcome these limitations, the Mixture-of-Experts (ME) system is proposed, which partitions the input data into several clusters and learns a classifier for each cluster. However, in the traditional ME system, the number of experts are fixed in advance and clustering procedure and the classification tasks are de-coupled. To deal with these problems, a Dirichlet Process mixture of Latent Variable Support Vector Machine (DPLVSVM) is proposed. In DPLVSVM model, the number of clusters is chosen automatically by DP mixture model, and the linear Latent Variable SVMs (LVSVM) are employed in each cluster. Different from previous algorithms, in DPLVSVM, the clustering procedure and LVSVM are jointly learned to gain infinite discriminative clusters. And the parameters can be inferred simply and effectively via Gibbs sampling technique. Based on the experimental data obtained from the synthesized dataset, Benchmark datasets and measured radar echo data, the effectiveness of proposed method is validated.

Key words: Target recognition; Mixture-of-Experts (ME) system; Dirichlet Process (DP) mixture model; Latent Variable Support Vector Machine (LVSVM) classifier

1 引言

在目标识别中，通常会处理一些样本数较多且分布复杂的数据集，如雷达高分辨距离像数据(High-Resolution Range Profile, HRRP)。由于 HRRP 包含了丰富的目标结构信息而且易于获取和快速处理，其在雷达自动目标识别领域得到了广泛的应用[1−4]。由于目标具有姿态敏感性[1−4]，同一目标的 HRRP 具有多模分布特性，尤其是随着目标库的增大，训练样本个数也会随之增加，数据分布也变得更加复杂。若使用所有样本来训练一个分类器会增加分类器的训练复杂度，而且容易忽视样本的内在结构，不利于分类。为了克服这些问题，人们提出了混合专家(Mixture of Experts, ME)模型[5−6]，这类模型将数据集划分成若干子集然后在各个子集上分别训练简单的分类器来构造全局非线性的复杂分类器，从而避免了设计复杂分类器，可以大大简化分类器的设计。

2014-01-20 收到，2014-05-30 改回
国家自然科学基金(61372132, 61271024, 61322103)，新世纪优秀人才支持计划(NCET-13-0945)，全国优秀博士学位论文作者专项资金(FANEDD-201156)和中央高校基本科研业务费专项资金资助课题
*通信作者：陈渤 bchen@mail.xidian.edu.cn

常见的混合专家模型采用 K-means 等传统聚类方法划分样本子集，然后在每个子集内单独训练分类器，称为有限混合专家模型(finite ME)模型。这类模型存在两个缺点：一是模型选择问题，即如何选择样本子集(聚类)个数；二是样本集的聚类过程是无监督的，独立于后端的分类器任务，因此较难保证每个聚类中数据的可分性，从而影响全局的分类性能。近年来，贝叶斯非参数模型方法如 Dirichlet 过程(Dirichlet Process, DP)混合模型已经成功用于聚类问题(确定聚类个数)和概率密度估计问题(确定多模分布的模态个数)[7,8]。不同于有限模型，它可以根据数据自动确定聚类个数[7]，通常称为无限聚类模型。基于此，文献[9]提出 DPMNL 模型，该模型利用 DP 混合模型将多个广义线性模型 MNL (MultiNomial Logit)模型构建成为一个非线性分类器，取得了良好的分类效果。由于最大间隔分类器支撑向量机(Support Vector Machine, SVM)的良好的分类和推广性能，文献[10]提出无限支撑向量机(infinite SVM, iSVM)模型。iSVM 在用 DP 混合模型对训练样本聚类的同时在每个聚类上训练 SVM 分类器。两种模型都利用 DP 混合模型对样本集进行聚类，并且通过对聚类过程和分类器学习的联合优化，将监督信息引入了聚类过程，增强了分类性能。本文方法与 iSVM 最为相关，然而 iSVM 模型中，DP 混合模型为 Bayes 模型，而 SVM 中的损失函数没有用概率模型进行描述[10,11]，因此该算法的求解过程较为复杂，每次迭代都需要涉及原始的 SVM 优化，从而增加了模型求解难度。

为了解决以上问题，本文将隐变量 SVM(Latent Variable SVM, LVSVM)引入 DP 混合模型，提出了 DPLVSVM 模型。LVSVM 是由文献[11]基于数据增广技术提出的，该方法通过引入隐变量将 SVM 用概率模型进行表达。可见 DPLVSVM 是在概率框架下建立的，不同于 iSVM，可以采用 Bayes 估计算法对模型参数进行简单有效的估计[11,12]。本文模型是一种无限混合专家模型，利用 DP 混合模型来自动将数据划分为多个具有简单分布(如高斯分布)的数据子集且不需事先确定样本聚类个数；同时在每个子集上训练一个形式简单的线性 LVSVM 分类器。本文模型通过对数据潜在结构的挖掘，将非线性分类问题分解为多个线性可分的子问题，从而实现对整个数据的非线性分类。基于仿真和实测数据的实验结果表明本文模型可以有效提高目标识别性能并具有良好的拒判性能。

2 DP 及 DP 混合模型

Dirichelet 过程是 1973 年由文献[13]首先提出的，它是一种应用于非参数 Bayes 模型中的随机过程。若 G 服从 DP，记为 $G\sim \mathrm{DP}(G_0,\alpha)$，其中 $\mathrm{Dir}(\bullet)$ 表示 Dirichlet 分布，G_0 称为基础分布，α 为聚集参数。DP 是一种基于分布的分布，G 为对 DP 采样得到的一个随机分布。本文采用 DP 混合模型[7−9,14−16]，根据样本的分布对样本进行聚类。DP 混合模型中假设观测样本 $\boldsymbol{x}_n$ 的分布参数为 $\boldsymbol{\Theta}_n$，且 $\boldsymbol{\Theta}_n$ 服从分布 G。基于 Stick-breaking[14,15]构造的 DP 混合模型为

$$\left.\begin{aligned} &\upsilon_c \mid \alpha\sim \mathrm{Beta}(1,\alpha),\ \boldsymbol{\Theta}_c \mid G_0\sim G\\ &z_n \mid \boldsymbol{\theta}(\boldsymbol{\upsilon})\sim \mathrm{Mult}(\boldsymbol{\theta}(\boldsymbol{\upsilon})),\ \boldsymbol{x}_n \mid z_n=c\\ &\boldsymbol{\Theta}_c\sim p(\boldsymbol{x}\mid\boldsymbol{\Theta}_c)\\ &c=1,2,\cdots,\infty\\ &n=1,2,\cdots,N \end{aligned}\right\} \tag{1}$$

其中 $G=\sum_{c=1}^{\infty}\theta_c(\boldsymbol{\upsilon})\delta_{\boldsymbol{\Theta}_c}$，$\theta_c(\boldsymbol{\upsilon})=\upsilon_c\prod_{j=1}^{c-1}(1-\upsilon_j)\in[0,1]$，$\boldsymbol{\upsilon}=[\upsilon_1\ \upsilon_2\ \cdots\ \upsilon_C]$，$\boldsymbol{\theta}=[\theta_1\ \theta_2\ \cdots\ \theta_C]$；$\alpha$ 为 υ_c 的先验分布参数；z_n 是 $\boldsymbol{x}_n$ 的指示因子，当 $z_n=c$ 时，样本 $\boldsymbol{x}_n$ 属于第 c 个聚类，即 $\boldsymbol{x}_n\sim p(\boldsymbol{x}\mid\boldsymbol{\Theta}_c)$；$\mathrm{Mult}(\bullet)$ 表示多项分布。

由此可见 DP 混合模型中，分布 G 为数据分布参数 $\boldsymbol{\Theta}_c$ 的先验分布。整个参数空间被划分为无限可数的离散点集合 $\{\boldsymbol{\Theta}_c\}_{c=1}^{\infty}$，共享同一个分布参数 $\boldsymbol{\Theta}_c$ 的样本会具有相同分布，自动成为一个聚类。为了实现参数估计，本文采用文献[8]中的截断 Stick-breaking 构造形式，即给定一个较大的值 C 并令 $p(\upsilon_C)=1$，所以对于任何 $c>C$ 有 $\theta_c(\boldsymbol{\upsilon})=0$。可见，聚类个数不会超过 C 个。

3 DP 混合隐变量 SVM 模型

3.1 隐变量 SVM 分类器

给定两类别数据集 $\{(\boldsymbol{x}_n,y_n)\mid\boldsymbol{x}_n=(x_1,x_2,\cdots,x_P)^{\mathrm{T}},y_n\in\{-1,+1\}\}_{n=1}^{N}$，$\boldsymbol{x}_n$ 表示 P 维训练样本，y_n 表示其对应的类标号，N 表示样本个数。SVM 刻画了一个两类别的线性分类器[11,17]，其判别函数为 $\hat{y}_n=\mathrm{sign}(\boldsymbol{w}^{\mathrm{T}}\tilde{\boldsymbol{x}}_n)$，其中 $\mathrm{sign}(\bullet)$ 为符号函数，$\tilde{\boldsymbol{x}}_n=(1,x_1,x_2,\cdots,x_P)^{\mathrm{T}}$ 表示 $\boldsymbol{x}_n$ 的增广向量。若 $\hat{y}_n\ge 0$ 表示样本属于正类(+1)，反之样本属于负类(−1)。为了最大化分类间隔同时降低分类错误率，SVM 的参数可由求解式(2)所示的优化问题来确定[11]。

$$\min_{\boldsymbol{w}} d(\boldsymbol{w})=\frac{1}{2}\|\boldsymbol{w}\|_2^2+\gamma\sum_{n=1}^{N}\max(1-y_n\boldsymbol{w}^{\mathrm{T}}\tilde{\boldsymbol{x}}_n,0) \tag{2}$$

其中，正常数 γ 称为调和参数，$\sum_{n=1}^{N}\max(1-y_n\boldsymbol{w}^{\mathrm{T}}\tilde{\boldsymbol{x}}_n,0)$ 表示损失函数。式(2)是一个凸优化问题，求解过程的计算复杂度为 $O(N^3)$。当样本个数较多时，其求解复杂度增加、计算精度降低，甚至无法计算。

文献[11]采用数据增广技术引入隐变量 $\boldsymbol{\lambda}$ 得到 SVM 参数的伪后验分布：

$$\begin{aligned} & p(\boldsymbol{w},\boldsymbol{\lambda}\mid\boldsymbol{y}) \\ & \propto \prod_{n=1}^{N}\lambda_n^{-1/2}\exp\left(-\frac{(\lambda_n - c(1-y_n\boldsymbol{w}^{\mathrm{T}}\tilde{\boldsymbol{x}}_n))^2}{2\lambda_n}\right) \\ & \quad\cdot\exp\left(-\frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}\right) \end{aligned} \tag{3}$$

从而可以采用 Bayes 估计算法求解 SVM 的参数。由于引入隐变量，本文将其称为隐变量 SVM (LVSVM)。

式(3)给出了 LVSVM 参数的后验表达形式，由此可将其与一些传统的 Bayes 方法如 DP 混合模型进行有机结合，如 Gibbs MedLDA 模型[12]。文献[11]给出了采用 MCMC, VB 等 Bayes 估计算法求解参数的步骤，从中可以看出，其算法复杂度为 $O(N)$，受样本个数的限制大大降低。

3.2 DP 混合隐变量 SVM 模型

本文提出的 DP 混合隐变量模型(DPLVSVM)，将大间隔分类器 LVSVM 与 DP 混合模型相结合。该模型将数据聚类和分类器学习联合优化，从而将监督信息引入到聚类过程中，在一定程度上保证了每个聚类的可分性。

DPLVSVM 模型中假设样本服从均值和协方差矩阵均未知的高斯分布，其参数记为 $\boldsymbol{\Theta}=\{\boldsymbol{\mu},\boldsymbol{\Sigma}\}$，考虑到数据特征间的相关特性，模型中基分布 G_0 采用 normal-Wishart 分布 $\mathrm{NW}(\boldsymbol{\mu},\boldsymbol{\Sigma}\mid\boldsymbol{\mu}_0,\boldsymbol{W}_0,\beta_0,v_0)$，其中 $\boldsymbol{\mu}_0,\boldsymbol{W}_0,\beta_0,v_0$ 为给定参数。根据 DP 混合模型式(1)与 LVSVM 式(3)，本文构建了 DPLVSVM 模型的层次化结构式(4)，从而建立了各参数的相关性。

$$\left.\begin{array}{l} v_c\mid\alpha\sim\mathrm{Beta}(1,\alpha),\ \ \boldsymbol{\Theta}_c\mid G_0\sim G_0 \\ z_n\mid\boldsymbol{\theta}(\boldsymbol{\nu})\sim\mathrm{Mult}(\boldsymbol{\theta}(\boldsymbol{\nu})),\ \boldsymbol{x}_n\mid z_n=c \\ \quad\boldsymbol{\Theta}_c\sim p(\boldsymbol{x}_n\mid\boldsymbol{\Theta}_c) \\ \quad\boldsymbol{w}_c,\{\lambda_n\}_c\mid\{\boldsymbol{x}_n,\boldsymbol{y}_n\}_{n=1,z_n=c}^{N}\sim p(\boldsymbol{w}_c,\{\lambda_n\}_c\mid-) \\ \quad c{=}1,2,\cdots,C \\ \quad n{=}1,2,\cdots,N \end{array}\right\} \tag{4}$$

其中，$\boldsymbol{w}_c$ 表示第 c 个聚类中分类器的系数，$\{\lambda_n\}_c$ 表示属于第 c 个聚类样本的增广隐变量。

由模型式(4)，本文可以推导出数据的概率密度函数式(5)和所有参数的联合后验分布式(6)。

$$q\left(\boldsymbol{x}\mid\boldsymbol{\nu},\{\boldsymbol{\mu}_c,\boldsymbol{\Sigma}_c\}_{c=1}^{C}\right)=\sum_{c=1}^{C}\boldsymbol{\pi}_c\boldsymbol{\mathcal{N}}(\boldsymbol{x};\boldsymbol{\mu}_c,\boldsymbol{\Sigma}_c) \tag{5}$$

$$\begin{aligned} & q\left(\{\boldsymbol{\mu}_c,\boldsymbol{\Sigma}_c\}_{c=1}^{C},\boldsymbol{Z},\boldsymbol{\upsilon},\{\boldsymbol{w}_c\}_{c=1}^{C},\boldsymbol{\lambda}\mid\boldsymbol{X},\boldsymbol{y}\right) \\ & \quad\propto p\left(\boldsymbol{X},\{\boldsymbol{\mu}_c,\boldsymbol{\Sigma}_c\}_{c=1}^{C},\boldsymbol{Z},\boldsymbol{\nu},\{\boldsymbol{w}_c\}_{c=1}^{C},\boldsymbol{\Lambda}_\eta,\boldsymbol{\lambda},\boldsymbol{y}\right) \\ & \quad= p\left(\boldsymbol{X}\mid\{\boldsymbol{\mu}_c,\boldsymbol{\Sigma}_c\}_{c=1}^{C},\boldsymbol{Z}\right) \\ & \qquad\cdot\prod_{c=1}^{C}p(\boldsymbol{\mu}_c,\boldsymbol{\Sigma}_c\mid\boldsymbol{\mu}_0,\boldsymbol{W}_0,\beta_0,v_0)p(\boldsymbol{Z}\mid\boldsymbol{\nu}) \\ & \qquad\cdot p(\boldsymbol{\nu}\mid\alpha)\prod_{c=1}^{C}p(\boldsymbol{w}_c,\boldsymbol{\lambda}_c\mid\boldsymbol{y}_c) \end{aligned} \tag{6}$$

其中，数据集 $\boldsymbol{X}=\{\boldsymbol{x}_n\}_{n=1}^{N}$，数据集中样本的聚类标记 $\boldsymbol{Z}=\{z_n\}_{n=1}^{N}$。

3.3 模型参数估计

DPLVSVM 模型采用 LVSVM 分类器，使得模型统一在概率框架下，从而可以通过 VB 或 MCMC 等 Bayes 估计算法进行有效的参数估计。基于 Gibbs 采样的 MCMC 算法则具有很强的可行性，一般不需要作近似处理，得到的结果相对更佳[8,14]。因此本文采用 Gibbs 采样技术进行参数估计。在 Gibbs 采样的每次迭代中，所有参数是从其条件后验分布的采样中获取的。由参数的联合后验分布式(6)可以推导出所有参数的后验分布。

聚类分布参数 $\{\boldsymbol{\mu}_c,\boldsymbol{\Sigma}_c\}$ 条件后验分布为

$$p(\boldsymbol{\mu}_c,\boldsymbol{\Sigma}_c\mid-)\sim\mathrm{NW}(\boldsymbol{\mu}_c,\boldsymbol{\Sigma}_c\mid\boldsymbol{\mu},\boldsymbol{W},\beta,v) \tag{7}$$

其中 $\boldsymbol{\mu}=(N_c\overline{\boldsymbol{x}}_c+\beta_0\boldsymbol{\mu}_0)/\beta$，$\boldsymbol{W}=(\boldsymbol{W}_0^{-1}+N_c\overline{\boldsymbol{\Sigma}}_c+N_c\beta_0(\boldsymbol{\mu}_0-\overline{\boldsymbol{x}}_c)(\boldsymbol{\mu}_0-\overline{\boldsymbol{x}}_c)^{\mathrm{T}}/\beta)^{-1}$，$\beta=\beta_0+N_c$，$v=v_0+N_c$，$N_c$ 表示属于第 c 个聚类的样本个数，$\overline{\boldsymbol{x}}_c$，$\overline{\boldsymbol{\Sigma}}_c$ 分别表示这些样本的均值和协方差的最大似然估计值。

样本聚类标记 z_n 条件后验分布为

$$\begin{aligned} & \boldsymbol{\kappa}_n=[\kappa_{n1},\kappa_{n2},\cdots,\kappa_{nC}], \\ & \kappa_{nc}=p(z_n=c\mid\boldsymbol{\mu}_c,\boldsymbol{\Sigma}_c,\boldsymbol{\nu},\boldsymbol{w}_c,\lambda_n) \\ & \quad\propto\boldsymbol{\mathcal{N}}(\boldsymbol{x}_n;\boldsymbol{\mu}_c,\boldsymbol{\Sigma}_c)\theta_c\phi_{nc},z_n\sim\mathrm{Mult}(\boldsymbol{\kappa}_n) \end{aligned} \tag{8}$$

其中 Mult(•) 表示多项分布。由此可以看出样本的聚类标记不仅与每个聚类的分布有关而且与监督信息有关。

SVM 系数 $\boldsymbol{w}_c$ 条件后验分布为

$$p(\boldsymbol{w}_c\mid\boldsymbol{Z},\boldsymbol{\lambda})\sim\boldsymbol{\mathcal{N}}(\boldsymbol{w}_c;\boldsymbol{\mu}_{\boldsymbol{w}},\boldsymbol{\Lambda}_{\boldsymbol{w}}) \tag{9}$$

其中 $\boldsymbol{\mu}_{\boldsymbol{w}}=\boldsymbol{\Lambda}_{\boldsymbol{w}}\sum_{n=1,z_n=c}^{N}\frac{\gamma y_n(\boldsymbol{\lambda}_n+\gamma l)}{\boldsymbol{\lambda}_n}\tilde{\boldsymbol{x}}_n$，$\boldsymbol{\Lambda}_{\boldsymbol{w}}=\left(\boldsymbol{I}+\sum_{n=1,z_n=c}^{N}\frac{\gamma^2}{\lambda_n}\tilde{\boldsymbol{x}}_n\tilde{\boldsymbol{x}}_n^{\mathrm{T}}\right)^{-1}$。

隐变量 $\boldsymbol{\lambda}_n$ 及 Stick-breaking 参数 $\boldsymbol{\nu}$ 的条件后验分布可分别由文献[11]和文献[8]得到，不再赘述。根据 Gibbs 采样技术，给定所有参数初始值后，对参数的条件后验分布进行循环采样，得到 DPLVSVM

模型参数的估计值。

另外，DPLVSVM 模型中对样本的分布参数进行了估计，由此可以采用混合概率模型对数据进行描述。式(5)给出了数据的概率密度函数，通过比较样本的概率密度函数值与拒判门限，可以判断样本是否为库外样本[18]。

4 识别系统流程

图 1 给出了整个识别系统的流程，可以看出整个系统包括两部分：训练阶段(实线框内)和测试阶段(虚线框内)。其中训练阶段的任务是对 DPLVSVM 模型进行参数估计，测试阶段的任务是根据训练得到参数计算样本所属聚类，然后在该聚类中进行拒判和识别任务。训练阶段和测试阶段的步骤为(略去了特征提取步骤)：

训练阶段：

步骤 1 初始化所有待估计参数，以及预热阶段迭代次数 I 且令 $i=1$；

步骤 2 根据参数的条件后验分布，采样所有模型参数；

步骤 3 判断终止条件：若 $i \le I$，则令 $i=i+1$ 并跳至步骤 2，否则终止预热阶段继续步骤 4；

步骤 4 继续对模型参数进行循环采样，从中抽取并存储 T_0 次采样作为对参数的估计。

测试阶段：

步骤 1 对于测试样本 $\boldsymbol{x}$ 先判断其所属的聚类：在第 t 次采样时，其聚类标记 z^t 条件后验分布为

$$z^t \sim \text{Mult}(\boldsymbol{\kappa}^t),\ \boldsymbol{\kappa}^t=[\kappa_1^t,\kappa_2^t,\cdots,\kappa_C^t],$$
$$\kappa_c^t = p\left(z^t=c \mid \{\boldsymbol{\mu}_c,\boldsymbol{\Sigma}_c,\boldsymbol{\nu}\}^t\right) \propto \pi_c^t \boldsymbol{\mathcal{N}}\left(\boldsymbol{x}_n;\{\boldsymbol{\mu}_c,\boldsymbol{\Sigma}_c\}^t\right) \quad (10)$$
$$t=1,2,\cdots,T_0$$

由此可以采样得到样本所属聚类的标号；

步骤 2 在所有 T_0 次采样中，根据式(5)计算样本的概率密度，然后与拒判门限比较[18,19]：若目标属于库内则用分类器进行分类，否则认为样本属于库外目标并拒判之；

步骤 3 当样本属于库内样本时通过式(11)判断样本所属类别，计算

$$\hat{y}=\text{sign}\left(\frac{1}{T_0}\sum_{t=1}^{T_0}(\boldsymbol{w}_{z^t}^t)^{\mathrm{T}}\tilde{\boldsymbol{x}}_n\right) \quad (11)$$

其中 $\boldsymbol{w}_{z^t}^t$ 表示第 z^t 个聚类上的分类器系数。

5 仿真实验

本节分别在人工数据集、Benchmark 数据集以及实测雷达 HRRP 数据上进行实验，来验证本文方法的有效性。实验中将本文 DPLVSVM 模型与 4 种分类器的识别性能进行了比较：线性 SVM(LSVM); K-means+线性 SVM(Km+SVM); DP 混合模型聚类+LSVM(DP+SVM)以及 DPMNL[9]。Km+SVM 与 DP+SVM 分别采用 K-means 及 DP 混合模型对样本聚类，然后每个聚类分别训练一个 LSVM，两个过程相互独立。实验中在处理多类问题时采用一对多策略[17]。对所有试验本文均采用常规的超参数设置[20]。DPLVSVM 模型参数设置为：LVSVM 的调和参数 $\gamma=1$；DP 混合模型参数中，分布参数 $\boldsymbol{\mu}_0=\boldsymbol{0}_{P\times 1}$，$\boldsymbol{W}_0=1\text{e}^{-6}\boldsymbol{I}_P$，$\beta_0=0.01$，$v_0=P$，Stick-breaking 尺度参数 $\alpha=0.1$。DP+SVM 模型中采用与之相同的 DP 混合模型参数。DPMNL 模型参数如文献[9]。本文实验环境为：Intel Pentium CPU G630，主频 2.70 GHz，内存 4 G, Matlab 版本为 R2013a。

5.1 人工数据集

为了体现本文提出的DPLVSVM模型的监督聚类以及数据描述的特性，首先在人工数据集 ToyData 上进行了无监督聚类(DP+SVM)与有监督聚类(DPLVSVM)实验对比。图 2(a)给出了 ToyData 的分布，该数据维数为 2，类别数为 2。从图上可以看出，数据呈现多模分布，单一的线性分类器无法将各类数据有效地分开。

图 2(b)给出了 DP+SVM 模型无监督聚类的一次采样结果，图示中不同记号后的数字‘i-j’表示该记号对应的样本属于第 i 个聚类中的第 j 类目标。

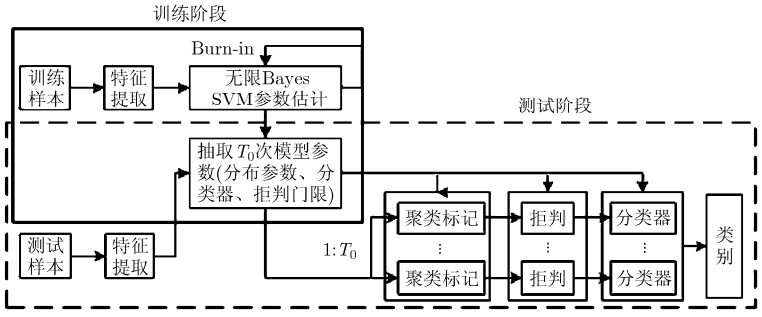


图 1 基于 dpLVSVM 模型的目标识别算法流程图

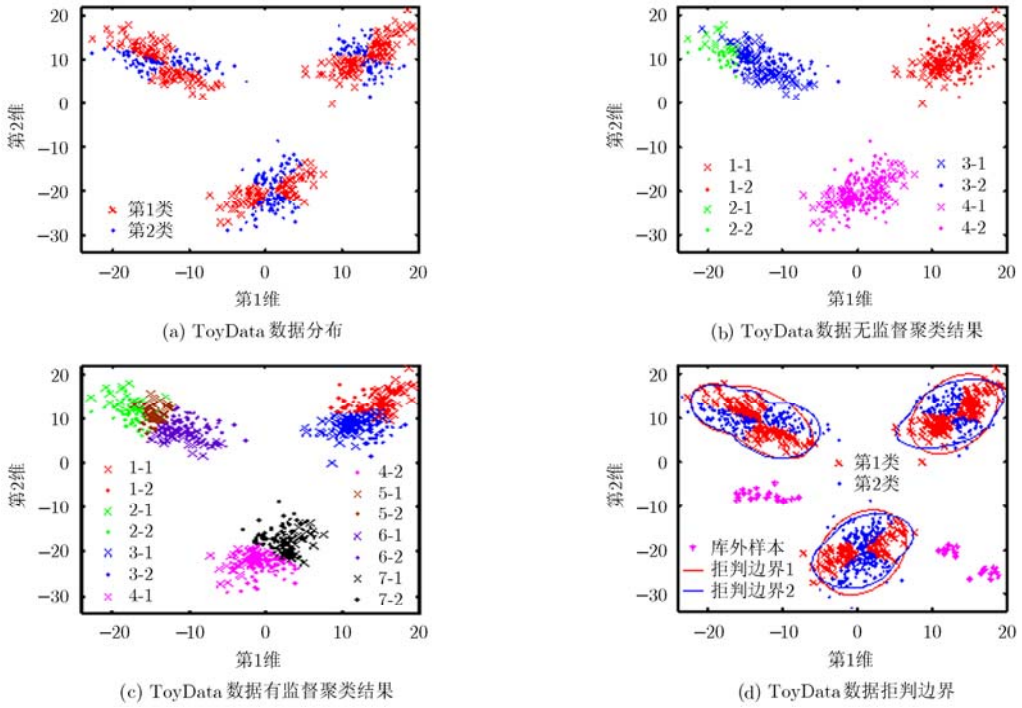
可以看出数据被划分为 4 个聚类，其中除第 2 个聚类外，其余 3 个聚类中的样本均线性不可分。图 2(c)给出了 DPLVSVM 模型有监督聚类的一次采样结果，可以看出数据被划分为 7 个聚类，每个聚类中的样本都具有良好的可分性。由此可见，DPLVSVM 能够保证每个聚类中的样本具有良好的可分性，从而可以提高整体的识别性能。表 1 中第 1 行给出了对不同方法 ToyData 的识别结果。

DP 混合模型可以估计出每个聚类的分布参数，由此可以构造出混合高斯分布来描述数据的整体分布。通过设定某个概率密度值作为拒判门限可以实现对库外目标的拒判[18]。图 2(d)给出了两种方法在对测试样本检测率为 95%时得到的拒判边界，其中拒判边界 1, 2 分别对应 DP+SVM 和 DPLVSVM。可以看出，采用 DP 混合模型能够对数据进行很好的描述，能够实现对库外目标的拒判。本文在第 5.3 节实测 HRRP 数据上进行了拒判实验及分析。

5.2 Benchmark 数据集

本节实验采用的数据集为从 UCI Machine Learning Repository 中获取的 Benchmark 数据集，该数据集包含了多种不同特征维数、不同规模的数据。本文从中选取具有多模分布或者样本数较多的 Banana, German, Image, Twonorm, Waveform 5 个数据集。实验中采用原始维度数据，共重复 20 次，每次试验中随机地划分训练样本集和测试样本集且样本个数保持不变。表 1 为 5 种不同方法分类结果。

从表 1 中的结果可以看出，相比于单个线性分类器(LSVM)以及传统的有限混合专家模型(Km+SVM)，无限混合专家模型取得了更好的识别性能。DP+SVM 模型虽然采用了 DP 混合模型对训练样



(a) ToyData 数据分布　(b) ToyData 数据无监督聚类结果

(c) ToyData 数据有监督聚类结果　(d) ToyData 数据拒判边界

图 2 人工数据集实验结果

表 1　Bechmark 数据介绍及识别结果

数据名称	类别数	维数	训练样本个数	测试样本个数	识别率				
					LSVM	Km+SVM	DP+SVM	DPMNL	DPLVSVM
ToyData	2	2	1200	1500	0.597	0.809	0.666	**0.939**	**0.937**
Banana	2	2	400	4900	0.525	0.665	0.606	0.808	**0.835**
German	2	20	700	300	0.762	0.755	0.800	0.757	**0.804**
Image	2	18	1200	1010	0.826	0.926	0.827	0.921	**0.937**
Twonorm	2	20	400	7000	0.965	0.965	0.968	**0.970**	**0.968**
Waveform	2	21	400	4600	0.865	0.881	0.888	**0.897**	**0.895**

本进行聚类，但是聚类过程是无监督的，较难保证每个聚类的可分性，因此其识别效果要略低于DPMNL和本文提出的DPLVSVM。比较DPLVSVM和DPMNL的结果，可以看出DPLVSVM可以得到与DPMNL相当甚至比其更好的识别结果。

5.3 实测雷达 HRRP 数据

前面实验中数据较为简单，为了进一步评估所提模型的性能，本节实验采用维数较高且分布相对复杂的实测雷达HRRP数据。该数据为某院的C波段雷达实测飞机的1维HRRP数据[1−4,19]。其中雷达载频为5.2 GHz，信号带宽为400 MHz。数据中包含3类飞机目标(雅-42、奖状、安-26)，三者航迹在地面上的投影如图 3 所示。从图中可以看出，3类飞机的HRRP数据均被划分成了若干段。为了检验本文方法的推广性能，分别选取不同的数据段作为训练和测试数据，其中选择“雅-42”的第 2, 5 段，“奖状”的第 6, 7 段以及“安-26”的第 5, 6 段共600个样本作为训练数据集，选择其余段中2400个样本作为测试数据集。

雷达 HRRP 数据具有幅度敏感性和平移敏感性[1−4,19]，这些特性对于识别是不利的。本文采用幅度 2 范数归一的方法只保留信号形状信息，消除幅度敏感性。为了消除平移敏感性，本文提取了HRRP的功率谱特征[1−3]。为了提高计算效率，文中采用PCA算法对数据进行降维，并比较了不同维数下各个分类器的识别性能。

图 4 给出了不同方法在不同特征维数下的识别结果。从图中可以看出，带有监督信息的无限混合专家模型DPMNL和DPLVSVM相比于LSVM识别性能大大提高。未考虑监督信息两个模型：Km+SVM识别性能低于LSVM; DP+SVM的识别性能相对于LSVM在大部分特征维度下有所提高，但在20维时下降较多。实验结果进一步证明了将监督信息引入聚类过程，可以大大提高整体的识别性能；而无监督情况下较难保证全局识别性能。DPLVSVM模型在大部分特征维度下都取得了最好的识别性能，特别是当特征维数为10时平均正确识别率达到了最高的0.930。比较DPLVSVM与DPMNL可以看出，前者取得了更好的识别效果。其原因一是DPLVSVM模型假设样本各维相关，而DPMNL模型中为了降低模型参数估计时对样本数量的需求，假设样本的各维之间相互独立，从而导致其模型不够准确；二是DPLVSVM模型采用的最大间隔分类器相比于DPMNL采用的MNL模型有更好的泛化能力。图 4 同时表明维数对识别率产生一定的影响：当特征维数较小时由于损失了较多的信息，识别率较低；特征维数较大时，特征中会包含一些的冗余信息，对识别有一定干扰作用，识别率有所降低。

当观测目标不属于模板库内的任一目标类别时，需要能够对该库外目标进行拒判。分类器的拒判性能则通常用接收机工作特性(Receiver Operating Characteristic, ROC)曲线来衡量[18,19]。ROC 曲线下的面积 AUC(Area Under an ROC Curve)越大，说明分类器的拒判性能越好。文献[18]给出了基于聚类算法(K-means)以及基于数据分布的拒判方法。根据文献[18]，Km+SVM, DP+SVM, DPMNL 和 DPLVSVM 均可以实现对库外样本拒判。此外，实验中同时引入了识别领域常用的支撑向量域描述(Support Vector Domain Description, SVDD)算法[18]。为了验证模型的拒判性能，选取了4 类其它飞机目标作为库外目标，每个目标等间隔抽取200个样本(共800个样本)作为库外目标样本。经过实验以上 5 种拒判方法所得到的 ROC 曲线如图 5 所示，图 5 中同时标注了各个方法的AUC值。从中可以看出采用 DP 混合模型的分类模型(DP+SVM, DPMNL 以及 DPLVSVM)均可以对数据的分布进行较好的描述，具有较好的拒判效果。综合之前识别和拒判结果可见，DPLVSVM模型既能提高分类性能又有良好的拒判性能。

5.4 模型时间复杂度对比

本节比较了不同分类方法的时间复杂度。表 2 中列出了 5 种分类方法分别在 Bechmark 数据集以

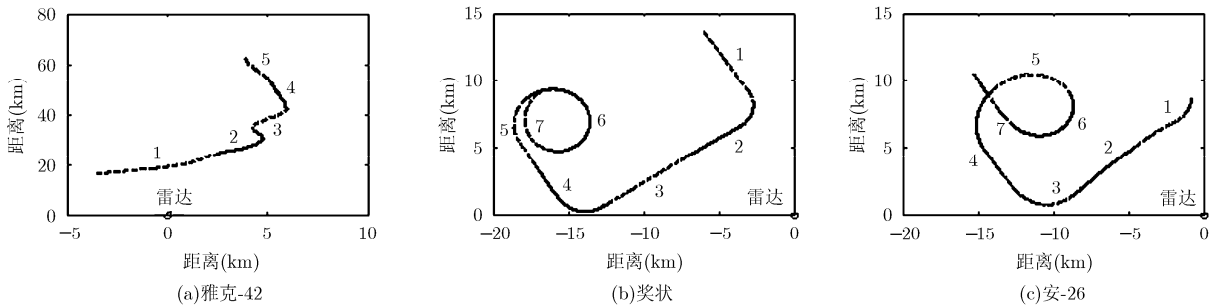


图 3 3 类飞机的航迹在地平面上投影图

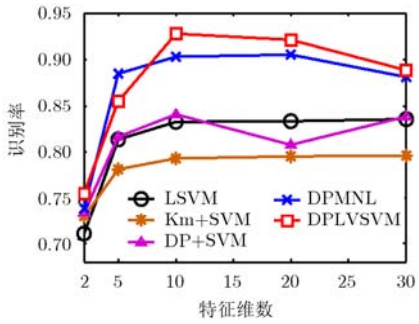


图 4 不同特征维数下 HRRP 识别结果

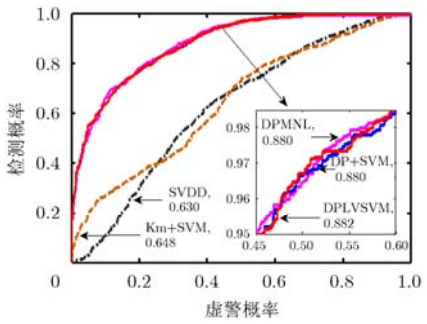


图 5 不同方法的 ROC 曲线及 AUC 值

表 2　5 种不同分类方法的所消耗的 CPU 时间(s)

数据集	训练时间					测试时间				
	LSVM	Km+SVM	DP+SVM	DPMNL	DPLVSVM	LSVM	Km+SVM	DP+SVM	DPMNL	DPLVSVM
Banana	0.063	*0.080*	21.477	**410.358**	21.917	0.0020	*0.0027*	*0.2113*	*0.1795*	***0.2125***
German	0.087	*0.124*	36.900	**758.903**	37.753	0.0010	*0.0008*	*0.0134*	***0.0148***	*0.0136*
Image	0.071	*0.249*	64.675	**1525.627**	66.117	0.0002	*0.0015*	*0.0447*	***0.0467***	*0.0453*
Twonorm	0.054	*0.165*	23.045	**294.698**	23.240	0.0025	*0.0125*	*0.3081*	*0.2498*	***0.3084***
Waveform	0.045	*0.118*	23.124	**431.794**	23.695	0.0009	*0.0047*	*0.2008*	*0.1848*	***0.2032***
HRRP10	0.102	*0.123*	36.634	**436.909**	37.364	0.0004	*0.0015*	*0.1070*	*0.0880*	***0.1071***

注：粗体数字表示最长耗时，斜体数字表示存在并行结构时所需的单次运行时间。

及 HRRP 数据集(特征维度为 10)上的训练时间和对所有测试样本的测试时间。在训练阶段中，Km+SVM 需要采用交叉验证的方法选择聚类个数，且需要多次聚类避免初值敏感性，消耗了较多时间，然而此操作可以采用并行处理结构，因此只考虑单次训练的时间。同样，Km+SVM 模型在测试阶段也可以进行相应的并行处理。另外，在 DP+SVM, DPMNL 与 DPLVSVM 中，测试样本根据存储的 T_0 组模型参数分别进行分类然后根据式(11)进行判决，也可以做并行处理(见图 1)。

采用 DP 混合模型的 3 个模型，DP+SVM, DPMNL 与 DPLVSVM，由于要采用 Gibbs 采样算法迭代求解模型参数，所以较前两种方法在训练和测试上消耗了更多的时间，特别是在训练阶段。由于识别问题中，训练阶段往往是离线操作的，对训练时间要求不高。测试阶段，采用并行处理可以获得可接受的时间复杂度。

6　结束语

为了处理目标识别算法中的大规模、多模分布数据，本文提出了 DPLVSVM 模型。不同于传统混合专家模型，该模型能够自动确定聚类的个数，并同时在各个聚类中训练一个大间隔的概率模型分类器(LVSVM)用于该聚类中样本的分类，从而将监督信息引入聚类过程。DPLVSVM 模型将 DP 混合模型聚类以及 LVSVM 分类器统一在概率框架下联合优化，不仅在一定程度上保证了各个聚类中的数据的可分性和分布上的一致性，而且通过 Gibbs 采样技术可以对模型的参数进行简单且有效的估计。通过在人工数据集、公共数据集以及雷达实测数据上的实验表明DPLVSVM模型提高了识别性能而且有较好的拒判性能。

参 考 文 献

[1] Du L, Liu H W, Wang P H, *et al*.. Noise robust radar HRRP target recognition based on multitask factor analysis with small training data size[J]. *IEEE Transactions on Signal Processing*, 2012, 60(7): 3546–3559.

[2] 潘勉, 王鹏辉, 杜兰, 等. 基于TSB-HMM模型的雷达高分辨距离像目标识别算法[J]. 电子与信息学报, 2013, 35(7): 1547–1554.
Pan Mian, Wang Peng-hui, Du Lan, *et al*.. Radar HRRP target recognition based on truncated stick-breaking hidden Markov model[J]. *Journal of Electronics & Information Technology*, 2013, 35(7): 1547–1554.

[3] 张玉玺, 王晓丹, 姚旭, 等. 基于Bagging-SVM动态集成的多极化HRRP识别[J]. 系统工程与电子技术, 2012, 34(7): 1366–1371.
Zhang Yu-xi, Wang Xiao-dan, Yao Xu, *et al*.. HRRP

recognition for polarization radar based on Bagging-SVM dynamic ensemble[J]. *Systems Engineering and Electronics*, 2012, 34(7): 1366–1371.

[4] 崔姗姗, 周建江, 朱劼昊. 基于半参数化SLC的雷达目标识别[J]. 雷达学报, 2012, 1(4): 414–419.
Cui Shan-shan, Zhou Jian-jiang, and Zhu Jie-hao. Radar target recognition based on semiparametric density estimation of SLC[J]. *Journal of Radars*, 2012, 1(4): 414–419.

[5] Collober R, Bengio S, and Bengio Y. A parallel mixture of SVMs for very large scale problems[J]. *Neural Computation*, 2002, 14(5): 1105–1114.

[6] Fu Z, Robles-Kelly A, and Zhou J. Mixing linear SVMs for nonlinear classification[J]. *IEEE Transactions on Neural Networks*, 2010, 21(12): 1963–1975.

[7] Anoniak C E. Mixtures of Dirichlet process with applications to Bayesian nonparametric problems[J]. *Annals of Statistics*, 1974, 2(6): 1152–1174.

[8] Blei D M and Jordan M I. Variational inference for Dirichlet process mixtures[J]. *Bayesian Analysis*, 2006, 1(1): 121–144.

[9] Shahbaba B and Neal R. Nonlinear models using Dirichlet process mixtures[J]. *The Journal of Machine Learning Research*, 2009, 10(4): 1829–1850.

[10] Zhu J, Chen N, and Xing E P. Infinite SVM: a Dirichlet process mixture of large-margin kernel machines[C]. The 28th International Conference on Machine Learning (ICML-11), Belevue, WA, USA, 2011: 617–624.

[11] Polson N G and Scott S L. Data augmentation for support vector machines[J]. *Bayesian Analysis*, 2011, 6(1): 1–24.

[12] Zhu J, Chen N, Perkins H, *et al.*. Gibbs max-margin topic models with fast sampling algorithms[C]. The 30th International Conference on Machine Learning (ICML-13), Atlanta, USA, 2013: 124–132.

[13] Ferguson T S. A Bayesian analysis of some nonparametric problems[J]. *The Annals of Statistics*, 1973, 1(2): 209–230.

[14] 周建英, 王跃飞, 曾大军. 分层Dirichlet过程及其应用综述[J]. 自动化学报, 2011, 37(4): 389–407.
Zhou Jian-ying, Wang Yue-fei, and Zeng Da-jun. Hierarchical Dirichlet process and their applications: a survey[J]. *Acta Automatica Sinica*, 2011, 37(4): 389–407.

[15] Sethuraman J. A constructive definition of Dirichlet priors[J]. *Statistica Siniac*, 1994, 4(2): 639–650.

[16] Fan W and Bouguila N. Online learning of a Dirichlet process mixture of Beta-Liouville distributions via variational inference[J]. *IEEE Transactions on Neural Networks and Learning System*, 2013, 24(11): 1850–1862.

[17] Jordan M, Kleinberg J, and Scholkopf B. Pattern Recognition and Machine Learning[M]. New York: Springer Scinence+ Business Media, 2008: 291–359.

[18] Tax D M J. One-class classification[D]. [Ph.D. dissertation], Delft University of Technology, The Netherland, 2001.

[19] 王鹏辉, 刘宏伟, 杜兰, 等. 基于线性动态模型的雷达高分辨距离像小样本目标识别方法[J]. 电子与信息学报, 2012, 34(2): 305–311.
Wang Peng-hui, Liu Hong-wei, Du Lan, *et al.*. Linear dynamic model based radar HRRP target recognition under small training set conditions[J]. *Journal of Electronics & Information Technology*, 2012, 34(2): 305–311.

[20] Chen B, Polatkan G, Sapiro G, *et al.*. Deep learning with hierarchical convolutional factor analysis[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(8): 1887–1901.

张学峰： 男，1987 年生，博士生，研究方向为雷达自动目标识别.

陈 渤： 男，1979 年生，博士，教授，研究方向为雷达目标识别、统计信号处理、统计机器学习、深度学习网以及大规模数据处理等.

王鹏辉： 男，1984 年生，博士，讲师，研究方向为雷达自动目标识别以及统计机器学习理论等.

刘宏伟： 男，1971 年生，博士，教授，博士生导师，研究方向为雷达信号处理、雷达自动目标识别、认知雷达、协同探测等.