

## 贝塔混合模型的变分贝叶斯学习及应用

赖裕平<sup>1</sup>,高 宁<sup>1</sup>,何闻达<sup>1</sup>,平 原<sup>2</sup>,杜春来<sup>1</sup>,王宝成<sup>1</sup>,丁洪伟<sup>3</sup>

(1. 北方工业大学计算机学院,北京 100144;2. 河南许昌学院,河南许昌 461000;

3. 云南大学信息学院,云南昆明 650091)

**摘 要:** 贝塔混合模型(Beta Mixture Model, BMM)是一种重要的非高斯概率模型,常用于有界数据的统计分析. 但是由于其表达式复杂, BMM的参数估计比较困难. 针对该问题, 本文提出一种高效的变分贝叶斯学习方法进行参数估计. 该方法采用形式简单的自由分布, 通过不断最大化初始变分目标函数的下界, 迭代逼近得到真实的贝叶斯后验分布. 在合成数据集与实际数据集上进行实验, 实验结果证明了所提出算法的有效性和可行性.

**关键词:** 贝塔分布; 贝叶斯估计; 模型选择; 变分推理; 目标分类

**中图分类号:** TP391.4 **文献标识码:** A **文章编号:** 0372-2112 (2018)07-1787-06

**电子学报 URL:** <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2018.07.036

## Variational Bayesian Learning for Beta Mixture Model and Its Application

LAI Yu-ping<sup>1</sup>, GAO Ning<sup>1</sup>, HE Wen-da<sup>1</sup>, PING Yuan<sup>2</sup>, DU Chun-lai<sup>1</sup>, WANG Bao-cheng<sup>1</sup>, DING Hong-wei<sup>3</sup>

(1. School of Computer, North China University of Technology, Beijing 100144, China;

2. School of Information Engineering, Xuchang University, Xuchang, Henan 461000, China;

3. School of Information Science and Engineering, Yunnan University, Kunming, Yunnan 650091, China)

**Abstract:** Beta mixture model (BMM) is an important non-Gaussian probability model, which has been widely used in statistical analysis of the bounded data. It is hard to perform parameter estimation for BMM, due to its complex function format. An efficient variational Bayesian learning method has been proposed to deal with this problem. With the variational distribution and by iteratively maximizing the lower bound of the original variational object function, the approximating distribution which is the closest to the true Bayesian posterior distribution is obtained. Both synthetic and real data are experimented to demonstrate the effectiveness and the merits of the proposed approach.

**Key words:** beta distribution; Bayesian estimation; model selection; variational inference; object categorization

### 1 引言

有限混合模型(Finite Mixture Model, FMM)是分析复杂现象的一个灵活而强有力的建模工具,其主要特征就是通过混合多个概率密度函数以代替单个概率密度函数,从而能够很好地逼近任意具有多峰特征的概率密度分布,这是单一概率密度函数所无法企及的.近年来, FMM被广泛应用于模型识别、机器学习、数据挖掘、计算机视觉、生物信息学等不同领域.在这些领域里, FMM被用来完成诸如概率密度函数估计、聚类、分类等任务.由于其具有表达式简单、计算方便等特点,高

斯混合模型(Gaussian Mixture Model, GMM)已成为目前最普遍应用的一种概率混合模型.但是实际数据多具有非线性、非高斯特性,如拖尾数据<sup>[1]</sup>、倾斜数据<sup>[2]</sup>、有界或半有界数据<sup>[3]</sup>, GMM不能准确地描述这些非高斯数据.为了解决该问题,国内外许多学者相继提出了多种非高斯混合模型,如 BMM<sup>[3]</sup>、狄利克雷混合模型<sup>[4]</sup>、贝塔刘维尔混合模型<sup>[5]</sup>、韦伯混合模型<sup>[6]</sup>等.

FMM研究的一个重要的工作是确定混合模型的分量数,即进行模型选择.过多的分量混合容易引起模型过适应,没有理想的泛化性能;而过少的混合分量容易

收稿日期:2017-12-08;修回日期:2018-04-07;责任编辑:郭游

基金项目:国家自然科学基金(No. 61461053, No. 61461053, No. 61364012);北方工业大学毓优人才支持计划资助;河南省高校科技创新人才计划(No. 18HASTIT022);河南省科技创新人才计划(No. 184100510012);河南省科技攻关计划(No. 182102210123);河南省教育厅科学技术研究重点项目(No. 16A520025, No. 18A520047)

使得模型缺乏柔韧性,无法精确地逼近真实模型概率分布.模型选择问题是一个很复杂而且相当困难的问题.许多学者对其进行了研究并提出了大量的模型选择方法,可以将其分为两类:确定性学习算法和 Bayesian 学习算法.确定性学习算法<sup>[7]</sup>通常为目标函数引入某种信息准则,如最小信息长度,并利用 EM 算法来确定混合分量数.确定性算法对初始参数很敏感、计算耗时、泛化能力较弱,无法应用于大规模数据集的环境. Bayesian 学习算法视模型参数与混合分量数为随机变量,并将参数与分量数的点估计转化为求其后验分布,从而避免了参数学习过程中的过拟合问题.但对于大多数 FMM,求解模型真实后验分布是不可行的.目前多采用近似推理算法来解决该问题.虽然只要有足够多的采样,马尔科夫链蒙特卡洛算法就能以较高的精度逼近真实后验分布,但由于大数据集计算量太大,而且判断一条马尔科夫链何时收敛也是一个难以解决的问题,因此一些快速近似推理算法相继被提出.变分推理用易处理的一簇分布来逼近隐变量的后验分布,最大化变分参数的观测对数似然下界 (Evidence Lower Bound, ELBO) 来求得变分后验分布,从而加快算法参数学习速度.

近年来,变分推理广泛应用于概率混合模型的学习与推理.但对于许多非高斯混合模型,由于其表达式复杂,从而使得观测变量与隐变量的联合分布的期望难以直接计算.因此,利用传统的变分推理对这类非高斯混合模型进行推理时无法得到闭式的后验分布.例如,常用于有界非高斯数据建模的 BMM 就属于这类非高斯混合模型,其定义为若干个多维 Beta 分布的线性组合.由于 BMM 代表的密度函数中存在 Beta 函数,无论利用常用的 EM 算法还是变分贝叶斯学习算法都难以获得模型参数的估计值.因此, BMM 的学习与推理成为一个极具挑战性的课题.为解决该难题, Ma 等<sup>[3]</sup>提出了一种基于多下界的扩展变分推理算法 (简记为 BMM<sub>M</sub>). BMM<sub>M</sub> 算法通过最大化初始 ELBO 的多个不同下界来求得后验分布的解析解,存在收敛性难以得到保障、收敛速度慢、估计精度差等问题.因此,本文对 BMM 提出一种新颖的基于单下界的扩展变分推理算法 (简记为 BMM<sub>S</sub>). 该算法将参数估计与模型选择纳入统一框架,通过最大化初始 ELBO 的唯一下界来求得后验分布的解析解.在仿真数据上的实验结果以及在目标分类问题上的应用均验证了本文所提算法的有效性和可行性.

## 2 贝塔混合模型

设  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]^T$  为独立同分布的随机观测数据集,其中  $\mathbf{x}_n = [x_{n1}, \dots, x_{nD}]^T \in R^D$ , 且  $0 < x_{nd} < 1, d = 1,$

$\dots, D$ . 如果每个  $\mathbf{x}_n$  的概率密度函数表示如下:

$$p(\mathbf{x}_n | \boldsymbol{\Pi}, \mathbf{U}, \mathbf{V}) = \sum_{m=1}^M \pi_m \text{Beta}(\mathbf{x}_n | \mathbf{u}_m, \mathbf{v}_m) \\ = \sum_{m=1}^M \pi_m \prod_{d=1}^D \text{Beta}(x_{nd} | u_{md}, v_{md}) \quad (1)$$

则  $\mathbf{X}$  服从贝塔混合分布,对应的模型为贝塔混合模型<sup>[3]</sup>.

其中,  $M$  为混合分量个数;  $\boldsymbol{\Pi} = [\pi_1, \dots, \pi_M]^T$  为混合系数矢量,  $\pi_m$  为混合系数,即第  $m$  个分量的先验概率,它满足式(2)中定义的限制条件.

$$0 \leq \pi_m \leq 1, \sum_{m=1}^M \pi_m = 1, m = 1, \dots, M \quad (2)$$

$\text{Beta}(\mathbf{x}_n | \mathbf{u}_m, \mathbf{v}_m)$  是给定  $\mathbf{x}_n$  来自于第  $m$  个成分的条件概率密度;  $\mathbf{u}_m$  和  $\mathbf{v}_m$  表示第  $m$  个成分的参数,  $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_M]$ ,  $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_M]$ ,  $\mathbf{u}_m = [u_{m1}, \dots, u_{mD}]^T$ ,  $\mathbf{v}_m = [v_{m1}, \dots, v_{mD}]^T$ ;  $\text{Beta}(\cdot)$  为贝塔分布的概率密度:

$$\text{Beta}(x_{nd} | u_{md}, v_{md}) = \frac{\Gamma(u_{md} + v_{md})}{\Gamma(u_{md})\Gamma(v_{md})} x_{nd}^{u_{md}-1} (1 - x_{nd})^{v_{md}-1} \quad (3)$$

式中,  $0 < x_{nd} < 1$ ,  $u_{md}$  和  $v_{md}$  称为形状参数,且均是正数;  $\Gamma(\cdot)$  为 Gamma 函数,定义为  $\Gamma(s) = \int_0^{\infty} e^{-x} x^{s-1} dx, s > 0$ .

通过对样本建立 Bayesian 框架,采用 Bayesian 估计算法可获得模型参数的估计值.对  $\mathbf{x}_n$  分量引入标记向量  $\mathbf{Z}_n = [z_{n1}, \dots, z_{nM}]^T, z_{nm} \in \{0, 1\}, \sum_{m=1}^M z_{nm} = 1$ .  $z_{nm}$  表示  $\mathbf{x}_n$  来自于第  $m$  个混合分量的概率.观测数据  $\mathbf{X}$  的条件概率密度函数可以写为:

$$p(\mathbf{X} | \boldsymbol{\Theta}) = \prod_{n=1}^N \prod_{m=1}^M \left[ \prod_{d=1}^D \text{Beta}(x_{nd} | u_{md}, v_{md}) \right]^{z_{nm}} \quad (4)$$

式中,  $\boldsymbol{\Theta} = \{\mathbf{Z}, \mathbf{U}, \mathbf{V}\}, \mathbf{Z} = [\mathbf{Z}_1, \dots, \mathbf{Z}_N]^T$ . 给定混合系数  $\boldsymbol{\Pi}$ , 隐变量  $\mathbf{Z}$  的条件概率密度函数可以表示为:

$$p(\mathbf{Z} | \boldsymbol{\Pi}) = \prod_{n=1}^N \prod_{m=1}^M \pi_m^{z_{nm}} \quad (5)$$

假设所有的模型参数  $\mathbf{U}$  和  $\mathbf{V}$  相互独立,选择 Gamma 分布作为其先验分布,即

$$p(\mathbf{U} | \boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{m=1}^M \prod_{d=1}^D \frac{\beta_{md}^{\alpha_{md}}}{\Gamma(\alpha_{md})} u_{md}^{\alpha_{md}-1} e^{-\beta_{md} u_{md}} \quad (6)$$

$$p(\mathbf{V} | \mathbf{S}, \mathbf{T}) = \prod_{m=1}^M \prod_{d=1}^D \frac{t_{md}^{s_{md}}}{\Gamma(s_{md})} v_{md}^{s_{md}-1} e^{-t_{md} v_{md}} \quad (7)$$

式中,  $\boldsymbol{\alpha} = \{\alpha_{md}\}, \boldsymbol{\beta} = \{\beta_{md}\}, \mathbf{S} = \{s_{md}\}, \mathbf{T} = \{t_{md}\}$ .

所有未知变量和观测样本数据的联合分布的概率密度函数表达式为:

$$p(\mathbf{X}, \boldsymbol{\Theta} | \boldsymbol{\Pi}) = p(\mathbf{X} | \boldsymbol{\Theta}) p(\mathbf{Z} | \boldsymbol{\Pi}) p(\mathbf{U} | \boldsymbol{\alpha}, \boldsymbol{\beta}) p(\mathbf{V} | \mathbf{S}, \mathbf{T}) \quad (8)$$

式中包含了  $\mathbf{Z}, \mathbf{U}$  和  $\mathbf{V}$  三个随机变量.由贝叶斯定理可

知,边缘似然函数的计算涉及到多重积分,而且由于贝塔分布的概率密度函数表达式复杂,求解十分困难.下一节将看到扩展变分推理如何解决这个难题.

### 3 BMM 的扩展变分推理

传统的变分推理的目标函数为<sup>[8]</sup>

$$L(q) = [\ln p(\mathbf{X}, \boldsymbol{\Theta} | \mathbf{I})]_{\boldsymbol{\Theta}} - [\ln q(\boldsymbol{\Theta})]_{\boldsymbol{\Theta}} \quad (9)$$

式中,  $[\cdot]_{\boldsymbol{\Theta}}$  表示对变量  $\boldsymbol{\Theta}$  求期望,  $q(\boldsymbol{\Theta})$  表示是关于变量  $\boldsymbol{\Theta}$  的真实后验分布的一个近似分布. 由于 BMM 中存在 Beta 函数, 使得期望值  $[\ln p(\mathbf{X}, \boldsymbol{\Theta} | \mathbf{I})]_{\boldsymbol{\Theta}}$  难以计算. 在文献[7]中提出的单下界扩展变分推理算法为解决该问题提供了一个有效的解决方案, 关键在于找到一个满足式(9)的辅助函数  $\tilde{p}(\mathbf{X}, \boldsymbol{\Theta} | \mathbf{I})$ .

$$[\ln p(\mathbf{X}, \boldsymbol{\Theta} | \mathbf{I})]_{\boldsymbol{\Theta}} \geq [\ln \tilde{p}(\mathbf{X}, \boldsymbol{\Theta} | \mathbf{I})]_{\boldsymbol{\Theta}} \quad (10)$$

变分下界可以重写为式(11)表示的形式.

$$\tilde{L}(q) = [\ln \tilde{p}(\mathbf{X}, \boldsymbol{\Theta} | \mathbf{I})]_{\boldsymbol{\Theta}} - [\ln q(\boldsymbol{\Theta})]_{\boldsymbol{\Theta}} \quad (11)$$

则每个参数的变分后验分布的通解表达式为:

$$\ln q_m^*(\boldsymbol{\Theta}_m) = [\ln \tilde{p}(\mathbf{X}, \boldsymbol{\Theta} | \mathbf{I})]_{j \neq m} + Const \quad (12)$$

其中  $[\cdot]_{j \neq m}$  代表对关于  $\mathbf{X}$  和  $\boldsymbol{\Theta}$  中除去  $\boldsymbol{\Theta}_m$  外的所有变量求期望,  $Const$  代表常量. 变分分布族假设分解为式(13)所示的形式.

$$q(\boldsymbol{\Theta}) = \prod_{n=1}^N q(z_{nm}) \prod_{m=1}^M \prod_{d=1}^D [q(u_{md})q(v_{md})] \quad (13)$$

利用式(11)可以得到如下变分后验分布:

$$q^*(\mathbf{Z}) = \prod_{n=1}^N \prod_{m=1}^M r_{nm}^{z_{nm}} \quad (14)$$

$$q^*(\mathbf{U}) = \prod_{m=1}^M \prod_{d=1}^D G(u_{md} | \alpha_{md}^*, \beta_{md}^*) \quad (15)$$

$$q^*(\mathbf{V}) = \prod_{m=1}^M \prod_{d=1}^D G(v_{md} | s_{md}^*, t_{md}^*) \quad (16)$$

式(14)~(16)中的参数在下列公式中进行定义:

$$r_{nm} = \frac{\rho_{nm}}{\sum_{k=1}^M \rho_{nk}} \quad (17)$$

$$\ln \rho_{nm} = \ln \pi_m + \sum_{d=1}^D [R_{md} + (\bar{u}_{md} - 1) \ln x_{nd} + (\bar{v}_{md} - 1) \ln(1 - x_{nd})] \quad (18)$$

$$R_{md} = \ln \frac{\Gamma(\bar{u}_{md} + \bar{v}_{md})}{\Gamma(\bar{u}_{md})\Gamma(\bar{v}_{md})} + [\Psi(\bar{u}_{md} + \bar{v}_{md}) - \Psi(\bar{u}_{md})] \cdot (\langle \ln u_{md} \rangle - \ln \bar{u}_{md}) \bar{u}_{md} + [\Psi(\bar{u}_{md} + \bar{v}_{md}) - \Psi(\bar{v}_{md})] \cdot (\langle \ln v_{md} \rangle - \ln \bar{v}_{md}) \bar{v}_{md} \quad (19)$$

$$\alpha_{md}^* = \alpha_{md} + \sum_{n=1}^N r_{nm} [\Psi(\bar{u}_{md} + \bar{v}_{md}) - \Psi(\bar{u}_{md})] \bar{u}_{md} \quad (20)$$

$$\beta_{md}^* = \beta_{md} - \sum_{n=1}^N \langle z_{nm} \rangle \ln x_{nd} \quad (21)$$

$$s_{md}^* = s_{md} + \sum_{n=1}^N r_{nm} [\Psi(\bar{u}_{md} + \bar{v}_{md}) - \Psi(\bar{v}_{md})] \bar{v}_{md} \quad (22)$$

$$t_{md}^* = t_{md} - \sum_{n=1}^N \langle z_{nm} \rangle \ln(1 - x_{nd}) \quad (23)$$

式(18)~(23)的后验概率的期望表达式如下:

$$\begin{cases} r_{nm} = \langle z_{nm} \rangle, \bar{u}_{md} = \frac{\alpha_{md}^*}{\beta_{md}^*}, \langle \ln u_{md} \rangle = \Psi(\alpha_{md}^*) - \ln \beta_{md}^*; \\ \bar{v}_{md} = \frac{s_{md}^*}{t_{md}^*}, \langle \ln v_{md} \rangle = \Psi(s_{md}^*) - \ln t_{md}^*. \end{cases} \quad (24)$$

式中,  $\Psi(\cdot)$  表示 digamma 函数,  $\langle \cdot \rangle$  表示期望. 通过设置变分目标函数  $\tilde{L}(q)$  的导数为零, 得到混合系数的估计值如下:

$$\pi_m = \frac{1}{N} \sum_{n=1}^N r_{nm} \quad (25)$$

式(14)~(16)代表的解析解并非显式解, 因为求解这三个分布中的任一分布时, 另外两个分布都是未知的, 而且每个分布的计算都需已知其他分布为前提, 即表明这些变量相互耦合. 本文提出的扩展变分推理通过循环迭代计算解决耦合问题. 即任意初始化各个变量, 然后从某些变量为起点, 反复迭代计算式(14)~(16), 直至算法收敛. BMM 的变分贝叶斯推理过程总结如下:

- (1) 设定初始参数  $\{\alpha_{md}, \beta_{md}, s_{md}, t_{md}\}$  与分量数  $M$ ;
- (2) 利用  $K$  均值初始化  $r_{nm}$ ;
- (3) 利用式(14)~(16)分别更新变分后验分布  $q^*(\mathbf{Z})$ 、 $q^*(\mathbf{U})$  和  $q^*(\mathbf{V})$ ;
- (4) 利用式(25)估计混合系数  $\{\pi_m\}$ ;
- (5) 重复步骤(2)至(3)直至算法收敛;
- (6) 选择混合系数大于  $10^{-5}$  的分量, 确定分量数  $M$ .

### 4 实验结果与分析

本节实验分为两部分, 首先进行合成数据集实验, 并对比 BMM<sub>s</sub> 与 BMM<sub>M</sub> 两种算法; 然后将 BMM<sub>s</sub> 算法应用于目标分类, 并将其与 BMM<sub>M</sub>、基于单下界变分推理的 DMM(DMM<sub>s</sub>) 和基于多下界变分推理的 DMM(DMM<sub>M</sub>) 算法进行对比. 各未知参数的初始值设置如下: 1) 超参数集  $\{\alpha_{md}, \beta_{md}, s_{md}, t_{md}\}$  设置为  $\{1, 0.1, 1, 0.1\}$ ; 2) 混合分量数  $M$  设置为 15 且各分量的系数相同.

#### 4.1 仿真实验

为了验证本文提出的 BMM<sub>s</sub> 算法的性能, 实验中用程序实现并产生了可以指定混合分量数的服从混合贝塔分布的四个样本数据集, 且样本容量可变. 为了表示方便, 实验中选择维度  $D=2$ . 下面用该算法进行参数估计与模型选择实验.

表 1 不同合成数据集的参数. 其中,  $N_m$  代表第  $m$  簇内的样本数,  $u_{m1}, v_{m1}, u_{m2}, v_{m2}$  代表真实参数,  $\hat{u}_{m1}, \hat{v}_{m1}, \hat{u}_{m2}, \hat{v}_{m2}$  和  $\hat{\pi}_m$  代表  $BMM_S$  算法的估计结果,  $\tilde{u}_{m1}, \tilde{v}_{m1}, \tilde{u}_{m2}, \tilde{v}_{m2}$  和  $\tilde{\pi}_m$  代表  $BMM_M$  算法的估计结果

数据集	$N_m$	$m$	$u_{m1}$	$v_{m1}$	$u_{m2}$	$v_{m2}$	$\pi_m$	$\hat{u}_{m1}$	$\hat{v}_{m1}$	$\hat{u}_{m2}$	$\hat{v}_{m2}$	$\hat{\pi}_m$	$\tilde{u}_{m1}$	$\tilde{v}_{m1}$	$\tilde{u}_{m2}$	$\tilde{v}_{m2}$	$\tilde{\pi}_m$
1	120	1	8	12	18	4	0.40	7.89	12.32	17.59	3.84	0.40	8.42	11.27	17.23	4.28	0.39
	180	2	12	24	6	12	0.60	12.29	23.91	6.07	11.79	0.60	12.72	24.36	5.83	12.34	0.61
2	100	1	8	44	18	24	0.20	8.15	44.21	18.61	25.01	0.21	8.21	43.08	17.47	24.65	0.18
	150	2	21	35	16	5	0.30	21.76	37.22	15.64	5.52	0.30	22.03	37.84	17.32	4.29	0.31
	250	3	18	24	8	35	0.50	17.81	22.63	8.09	33.62	0.49	19.04	25.26	7.38	32.66	0.51
3	200	1	32	12	40	24	0.20	30.93	12.22	38.77	24.83	0.21	31.42	13.25	42.31	24.68	0.18
	250	2	12	38	36	24	0.25	11.89	36.64	37.21	23.69	0.25	12.34	39.03	34.27	22.74	0.26
	300	3	18	28	38	12	0.30	18.21	29.91	36.43	12.84	0.31	16.27	27.44	39.88	11.35	0.32
	250	4	32	8	8	20	0.25	32.91	8.26	7.93	21.45	0.23	34.25	8.45	9.23	18.62	0.24
4	150	1	12	32	14	84	0.15	12.22	31.36	14.26	82.38	0.14	13.44	31.67	16.03	81.31	0.17
	200	2	32	18	24	15	0.20	33.02	17.89	24.42	15.37	0.21	29.73	16.80	23.24	14.23	0.22
	300	3	24	18	48	12	0.30	24.34	18.65	49.11	11.15	0.29	25.01	18.82	46.18	12.89	0.29
	200	4	12	18	16	32	0.20	11.88	18.21	16.24	32.70	0.20	12.86	17.64	17.34	34.76	0.18
	150	5	12	54	24	18	0.15	12.04	56.29	22.95	18.63	0.16	11.27	52.25	24.68	18.53	0.14

表 1 给出了真实参数和估计参数. 表 2 给出了  $BMM_S$  与  $BMM_M$  两种算法的迭代时间和迭代次数. 由表 1 和表 2 可见,  $BMM_S$  算法比  $BMM_M$  算法的估计结果更准确、收敛速度更快. 图 1 描述了  $BMM_S$  算法的变分目标函数的变化趋势. 由图 1 可以看出, 随着迭代次数不断增加, 目标函数逐渐递增直至收敛. 图 2 给出了  $BMM_S$  算法收敛后混合分量的分布情况. 由此可见,  $BMM_S$  算法能有效地自动确定混合分量数.

表 2 两种算法的性能指标对比

数据集	$BMM_S$		$BMM_M$	
	迭代时间(s)	迭代次数	迭代时间(s)	迭代次数
1	0.17	95	0.23	137
2	0.23	157	0.38	235
3	0.35	218	0.51	389
4	0.39	232	0.65	434

## 4.2 目标分类

目标分类是计算机视觉、模式识别、机器学习和人工智能等领域内最具挑战性的研究课题之一<sup>[8,9]</sup>, 其主要任务是判断某个目标图形是否属于某个类别. 目标分类在智能监控、图像自动标注、自动图片过滤、图像检索等领域具有广泛的应用前景. 但在实际应用中, 环境的复杂性、局部遮挡、光照变化、形状变化、尺度变化、颜色变化和姿态变化等综合因素的影响给目标分类技术的发展带来了巨大的挑战. 目前大多数目标分类系统都采用局部特征描述子表示图像目标, 常用的局部特征主要有 SIFT<sup>[10]</sup> 和 HOG<sup>[11]</sup>, 本节实验采用文献

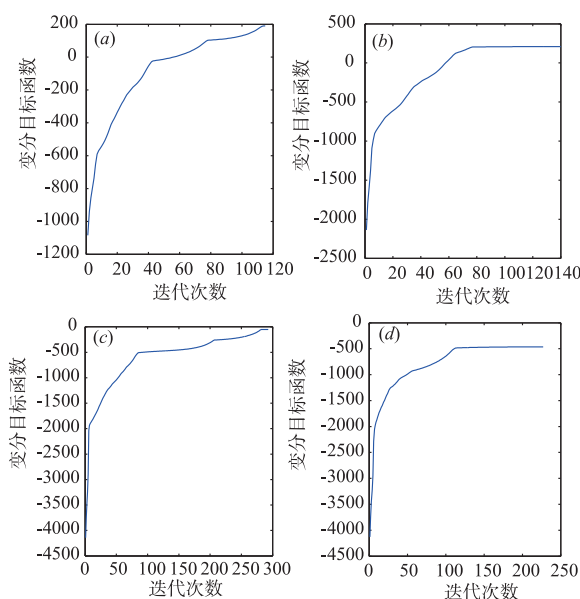


图 1 变分目标函数与迭代次数的关系

[12] 中所提出的 RHOG 特征. 检测窗口 (window) 和直方图 bin 的个数均设置为 3, 则每幅图像目标样本可以用一个 81 维的特征矢量表示.

本实验采用两个公共图像数据集验证本文所提出算法的有效性. 第一个图像数据集是 Caltech4 目标数据集<sup>[13]</sup>, 由飞机 (1074 幅)、汽车 (526 幅)、人脸 (450 幅) 和摩托车 (826 幅) 四类目标图像组成. 第二个图像数据集是 ETH-80 目标数据集<sup>[14]</sup>, 该数据集收集了 8 类目

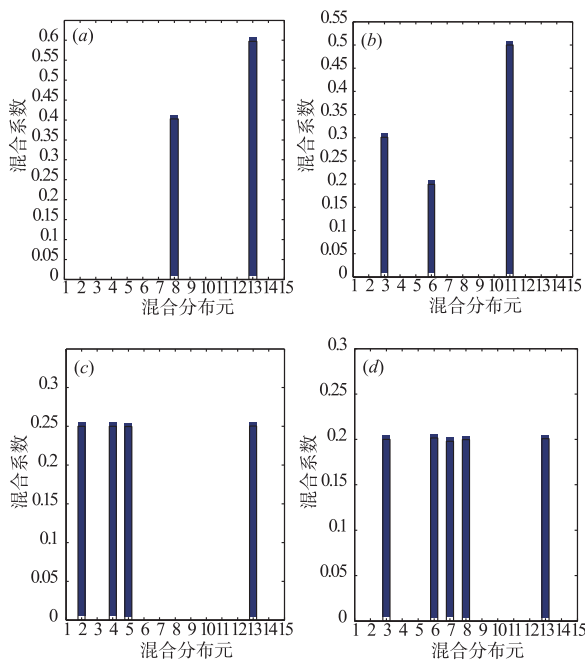


图2 模型选择的结果

标对象,每类目标包含 410 幅图片,共 3280 幅图像. 实验中对每类目标,选取一半图像作为训练样本,另外一半图像作为测试样本,整个分类过程一共进行 10 次实验,取 10 次分类平均值作为最后的测试结果. 表 3 比较了四种算法对应的识别率均值,图 3 给出了四种算法对应的识别率分布的箱形图. 从表 3 和图 3 可见:(1) 采用单下界变分推理算法的分类精度明显好于采用多下界的扩展变分推理算法;(2) 相比于 DMM, BMM 具有更强的建模能力.

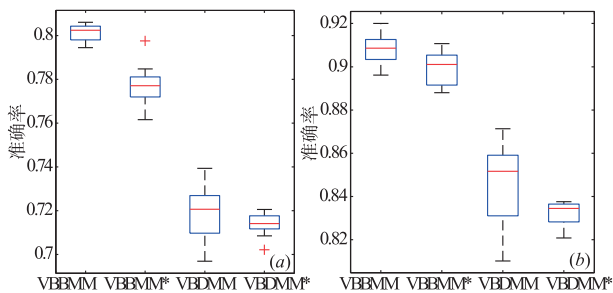


图3 目标分类准确率分布图. (a) Caltech4; (b) ETH-80.

表 3 各种算法分类准确率的比较

图像集	BMM <sub>s</sub>	BMM <sub>M</sub>	DMM <sub>s</sub>	DMM <sub>M</sub>
Caltech4	90.82%	89.96%	84.41%	83.18%
ETH-80	80.12%	77.61%	71.95%	71.38%

### 5 总结

针对 BMM 的贝叶斯参数估计问题,本文提出了一

种单下界的扩展变分推理算法,并进行了详细的推导. 大量的仿真数据集和真实数据集的实验结果表明,该算法可以准确地进行参数估计和模型选择. 后续的研究可以在算法的实时性、准确性与特征选择等方面进行深入的讨论.

### 参考文献

- [1] Kong L, Zhang H, Zheng Y, et al. Image segmentation using a hierarchical student-t mixture model[J]. IET Image Processing, 2017, 11(11): 1094 - 1102.
- [2] Matza A, Bistriz Y. Skew Gaussian mixture models for speaker recognition [J]. IET Signal Processing, 2014, 8(8): 860 - 868.
- [3] Ma Z, Leijon A. Bayesian estimation of beta mixture models with variationally inference [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(11): 2160 - 2173.
- [4] Bouguila N, Ziou D, Vaillancourt J. Unsupervised learning of a finite mixture model based on the Dirichlet distribution and its application [J]. IEEE Transactions on Image Processing, 2004, 13(11): 1533 - 1543.
- [5] Bouguila N. Hybrid generative/discriminative approaches for proportional data modeling and classification [J]. IEEE Transactions on Knowledge and Data Engineering, 2012, 24(12): 2184 - 2202.
- [6] Wu T, Luo J, Fang J, et al. Unsupervised object-based change detection via a Weibull mixture model-based binarization for high-resolution remote sensing images [J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(1): 63 - 67.
- [7] Ma Z, Rana P K, Taghia J, et al. Bayesian estimation of Dirichlet mixture model with variational inference [J]. Pattern Recognition, 2014, 47(9): 3143 - 3157.
- [8] 潘泓, 朱亚平, 夏思宇, 金立左. 基于上下文信息和核熵成分分析的目标分类算法 [J]. 电子学报, 2016, 44(3): 580 - 586.  
PAN Hong, ZHU Ya-ping, XIA Shi-yu, JIN Li-luo. Object classification using context cue and kernel entropy component analysis [J]. Acta Electronica Sinica, 2016, 44(3): 580 - 586. (in Chinese)
- [9] 赵永威, 周苑, 李弼程, 柯圣财. 基于近义词自适应软分配和卡方模型的图像目标分类方法 [J]. 电子学报, 2016, 44(9): 2081 - 2188.  
ZHAO Yong-wei, ZHOU Yuan, LI Bi-cheng, KE Sheng-cai. Image object classification method with homion-ym based on adaptive soft-assignment and Chi-Square model [J]. Acta Electornica Sinica, 2016, 44(9): 2081 - 2188. (in Chinese)
- [10] Lowe D. Distinctive image features from scale-invariant

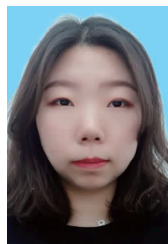
- keypoints[J]. International Journal of Computer Vision, 2004, 60(2):91 – 110.
- [11] Dalal N, Triggs B. Histograms of oriented gradients for human detection[A]. Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition [C]. San Diego, CA, USA: IEEE Press, 2005. 886 – 893 vol. 1.
- [12] Ludwig O, Delgado D, Goncalves V, et al. Trainable classifier-fusion schemes: An application to pedestrian detection[A]. Proceedings of the 12th IEEE International Conference on Intelligent Transportation Systems [C]. St. Louis, MO, USA: IEEE Press, 2009. 1 – 6.
- [13] Li F F, Fergus R, Perona P. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories[A]. Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition [C]. Washington, DC, USA: IEEE Press, 2004. 178 – 178.
- [14] Leibe B, Schiele B. Analyzing appearance and contour based methods for object categorization[A]. Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) [C]. Madison, WI, USA: IEEE Press, 2003. II-409-15 vol. 2.

## 作者简介



赖裕平 男, 1983 年 10 月生于江西瑞金, 北方工业大学计算机学院讲师, 主要研究方向为模式识别、机器学习、非高斯统计模型、贝叶斯网络、图像分类.

E-mail: laiyp2009@126.com



高宁 女, 1994 年 6 月生于辽宁鞍山, 北方工业大学硕士研究生, 主要研究方向为模式识别、机器学习、图像分类.

E-mail: gaoning\_ust@126.com



丁洪伟 (通信作者) 男, 1964 年 5 月生于云南省景洪市, 云南大学信息学院教授, 主要研究方向为轮询系统、优化算法研究、人工智能、机器学习.

E-mail: dhw1964@163.com