

有限贝塔刘维尔混合模型的变分学习及其应用

赖裕平¹, 丁洪伟², 周亚建¹, 郭玉翠³, 杨义先¹

(1.北京邮电大学计算机学院,北京 100876;2.云南大学信息学院,云南昆明 650091;3.北京邮电大学理学院,北京 100876)

摘 要: 由于贝塔刘维尔分布的共轭先验分布中存在积分表达式,贝叶斯估计有限贝塔刘维尔混合模型参数异常困难.本文提出利用变分贝叶斯学习模型参数,采用 gamma 分布作为近似的先验分布并使用合理的非线性近似技术,得到了后验分布的近似解.与常用的 EM 算法相比,该方法能够同时估计模型参数和确定分量数,且避免了过拟合的问题.在合成数据集及场景分类问题上进行了大量的实验,实验结果验证了本文所提方法的有效性.

关键词: 有限混合模型; 贝塔刘维尔分布; 分解近似; 模型选择; 变分推断

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0372-2112 (2014)07-1347-06

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2014.07.015

Variational Learning for Finite Beta-Liouville Mixture Models and Its Application

LAI Yu-ping¹, DING Hong-wei², ZHOU Ya-jian¹, GUO Yu-cui³, YANG Yi-xian¹

(1. Information Security Center, Beijing University of Posts and Telecommunications, Beijing 100876, China;

2. School of Information Science and Engineering, Yunnan University, Kunming, Yunnan 650091, China;

3. School of Science, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: Since the integration expression is present in the conjugate prior distribution, Bayesian estimation of the parameters in finite Beta-Liouville mixture models (BLM) is analytically intractable. In this paper, an approach based on the variational inference framework is proposed. Adopting gamma distributions to approximate the prior distributions of the parameter in BLM and using some reasonable non-linear approximations; the closed form solution for the posterior distribution of the parameters is obtained. Compared to the conventional expectation maximization (EM) algorithm, the proposed algorithm is able to simultaneously estimate the model parameters and determine the number of components; our method also avoids the problem of overfitting. Extensive experimental results based on the synthetic data sets and scenes classification show that the proposed method is efficient and feasible in terms of parameter estimation and model selection.

Key words: finite mixture models; Beta-Liouville distribution; factorized approximation; model selection; variational inference

1 引言

有限混合模型^[1]是分析复杂数据的一个灵活而强有力的概率建模工具,它提供了一种用简单密度拟合复杂密度的有效的数学方法.有限混合模型广泛应用于模式识别^[2],计算机视觉和机器学习等领域.其中,高斯混合模型凭借其计算模型参数方便,形式简单和描述能力强等特点,成为应用最为广泛的有限混合模型.然而实际数据许多具有非线性、非高斯特性,无法用高斯混合模型准确拟合.非高斯统计建模是目前一个活跃的研究领域,涌现了大量的研究成果.比如文献^[3]用有限逆狄

利克雷混合模型分析正实数,文献^[4]用有限贝塔刘维尔混合模型分析比例数据.

有限混合模型研究的两个核心问题是:混合模型的参数估计和混合分量数的确定.学习算法主要分为两大类:确定性学习算法和非确定的学习算法.其中确定性学习算法以极大似然估计(MLE)的EM算法为主,其核心思想是通过渐进优化模型似然函数获得参数的估计.这类方法存在泛化能力差,高度依赖参数初始值的选择,易于陷入局部最大值和过拟合等问题.非确定的学习算法以MCMC链为代表的Bayesian学习算法为主.Bayesian估计法将待估参数看作符合某种先验概率分布

的随机变量,并通过观测样本将先验概率密度通过贝叶斯规则转化为后验概率密度.通过某种方式结合参数的先验信息和边缘化参数的不确定性, Bayesian 方法避免了确定性方法的不足.然而,它通常存在共轭先验分布难以确定和边际高维积分计算困难的问题,比如贝塔分布的标准共轭先验分布无法直接应用于贝叶斯推断^[10].目前解决高维数积分困难问题的主要方法有马尔科夫蒙特卡罗(MCMC)抽样算法,拉普拉斯近似算法和变分贝叶斯算法. MCMC 存在耗时,收敛缓慢及收敛难以保证等缺点;拉普拉斯近似算法依赖于模型似然函数是单峰的假设,而有限混合模型的真实似然函数是非单峰的,所以这种估计不准确.

变分贝叶斯是解决高维积分困难问题的一种重要的逼近方法.其核心思想是利用易处理的一类分布族来逼近隐变量的后验分布,通过最大化变分参数的对数似然目标函数的下界获得模型参数的估计值^[5].与 MCMC 相比,变分贝叶斯在较好估计精度的前提下具有收敛速度快,泛化能力强、计算易处理等优势,因此备受青睐.

如何自动确定合适的、准确的混合分量数是有限混合模型学习中的一个具有挑战性的问题.分量个数是有限混合模型的规模度量,分量选择问题也称为模型选择问题.过多的分量混合容易引起模型过适应,导致泛化性能下降;而过少的分量使得模型缺乏柔韧性,不能精确地逼近真实的概率分布.众所周知,MLE 需要结合某种信息准则,比如 AIC^[6] 准则,贝叶斯推理准则(BIC)^[7,8],最小信息长度准则(MML)^[9]等,才能实现模型选择.由于这类算法需要对全部可能的分量数重复进行参数估计,才能获得最优的分量数,因此比较耗时.

鉴于比例数据大量存在于文本、图像中,混合贝塔刘维尔模型是分析这类数据的有效工具,本文研究该模型的学习算法,提出了一个变分推断的框架,能够有效地估计模型参数和分量数,并把该模型应用于场景图像分类.

2 有限贝塔刘维尔混合模型的变分学习

2.1 有限贝塔刘维尔混合模型

有限贝塔刘维尔混合模型(BLM)定义为 M 个贝塔刘维尔分布的线性组合,即

$$p(\mathbf{x}_n | \Theta, \boldsymbol{\pi}) = \sum_{i=1}^M \pi_i BL(\mathbf{x} | \boldsymbol{\theta}_i) \quad (1)$$

其中, $\boldsymbol{\pi} = (\pi_1, \dots, \pi_m)$ 是混合权重且满足 $\pi_i \geq 0$, $\sum_{i=1}^M \pi_i = 1$, $\mathbf{x}_n = (x_{n1}, \dots, x_{nD})$ 是观测数据, $\boldsymbol{\theta}_i = (\theta_{i1}, \dots, \theta_{iD+2})$ 是第 i 个贝塔刘维尔分量的参数,参数 $\Theta = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_M)$, 贝塔刘维尔分布的概率密度函数为:

$$BL(\mathbf{x}_n | \boldsymbol{\theta}_i) = \frac{\Gamma(\sum_{d=1}^D \theta_{id}) \Gamma(\theta_{iD+1} + \theta_{iD+2})}{\prod_{d=1}^{D+2} \Gamma(\theta_{id})}$$

$$\times \prod_{d=1}^D x_{nd}^{\theta_{id}-1} (1 - \sum_{d=1}^D x_{nd}) (\sum_{d=1}^D x_{nd})^{|\alpha_i|} \quad (2)$$

其中: $|\alpha_i| = \theta_{iD+2} - \sum_{d=1}^D \theta_{id}$, $\Gamma(\cdot)$ 是伽马函数. 注意当 $\theta_{iD+2} = \sum_{d=1}^D \theta_{id}$, 式(2)退化狄利克雷分布. 给定服从 BLM 的随机观测变量集 $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_N)$, 并引入其所属分量标签 $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_N)$, $\mathbf{z}_n = (z_{n1}, \dots, z_{nM})$. 其中, 元素 Z_{ni} 取值为 0 或 1, 且 $\sum_{i=1}^M Z_{ni} = 1$. 当 \mathbf{x}_n 来源于第 i 个贝塔刘维尔分量时, $Z_{ni} = 1$, 否则 $Z_{ni} = 0$. 特别注意, 每个样本 \mathbf{x}_n 只能对应一个分量. 给定混合权重 $\boldsymbol{\pi}$, 隐变量 \mathbf{Z} 的条件概率密度函数可以表示为:

$$p(\mathbf{Z} | \boldsymbol{\pi}) = \prod_{n=1}^N \prod_{i=1}^M \pi_i^{z_{ni}} \quad (3)$$

给定类别标签, 则观测变量集 \mathbf{X} 的条件分布为:

$$p(\mathbf{X} | \Theta, \mathbf{Z}) = \prod_{n=1}^N \prod_{i=1}^M BL(\mathbf{x}_n | \boldsymbol{\theta}_i)^{z_{ni}} \quad (4)$$

贝叶斯推理中, 需要预先指定模型参数的先验分布. 因为贝塔刘维尔分布的标准共轭先验不能直接用于贝叶斯推理, 所以我们采用 gamma 分布作为它的近似共轭先验^[10]. 假设 BLM 中各个参数是统计独立的, 则

$$p(\Theta) = \prod_{i=1}^M \prod_{d=1}^{D+2} \frac{\nu_{id}^{u_{id}}}{\Gamma(u_{id})} \theta_{id}^{u_{id}-1} e^{-\nu_{id} \theta_{id}} \quad (5)$$

式中: (u_{id}, ν_{id}) 是超参数且满足 $u_{id} > 0, \nu_{id} > 0$.

2.2 BLM 参数后验分布

结合式(3), (4)和(5)及利用贝叶斯定理, 得到所有模型变量的联合分布的对数形式:

$$\begin{aligned} \ln p(\mathbf{X}, \Theta, \mathbf{Z} | \boldsymbol{\pi}) &= \sum_{n=1}^N \sum_{i=1}^M z_{ni} \{ \ln \pi_i + \mathbf{R}_i + \sum_{d=1}^D (\theta_{id} - 1) \ln x_{nd} \\ &+ \theta_{i, D+1} \ln(1 - \sum_{d=1}^D x_{nd}) + (\theta_{i, D+2} - \sum_{d=1}^D \theta_{id}) \ln(\sum_{d=1}^D x_{nd}) \} \\ &+ \sum_{i=1}^M \sum_{d=1}^{D+2} [u_{id} \ln \nu_{id} - \ln \Gamma(u_{id}) + (u_{id} - 1) \ln \theta_{id} - \nu_{id} \theta_{id}] \end{aligned} \quad (6)$$

式中:

$$\mathbf{R}_i = \ln \frac{\Gamma(\sum_{d=1}^D \theta_{id}) \Gamma(\theta_{iD+1} + \theta_{iD+2})}{\prod_{d=1}^{D+2} \Gamma(\theta_{id})} \quad (7)$$

当 $\theta_{iD+2} = \sum_{d=1}^D \theta_{id}$ 时, 式(7)变成多元 LIB 函数^[10] (MLIB).

根据平均场理论, 近似后验分布可以表示为:

$$q(\Theta, \mathbf{Z}) = q(\Theta) q(\mathbf{Z}) \quad (8)$$

标记 $\langle y \rangle = E_q[y]$, $\Psi(y) = \frac{d}{dy} \ln \Gamma(y)$, $\Psi'(y) = \frac{d^2}{dy^2} \ln \Gamma(y)$.

接下来推导变量 Θ, \mathbf{Z} 近似分布表达式. 利用变分通解表达式(10.7)^[5] 和模型变量的联合分布的对数形

式,得到各优化因子的对数形式:

$$\begin{aligned} \ln q^*(z_{ni}) &= \text{const} + z_{ni} \left[\ln \pi_i + \tilde{\mathbf{R}}_i + \sum_{k=1}^D (\bar{\theta}_{ik} - 1) \ln x_{nk} \right. \\ &\quad \left. + (\bar{\theta}_{iL+2} - \sum_{k=1}^D \bar{\theta}_{ik}) \ln \left(\sum_{k=1}^D x_{nk} \right) + (\bar{\theta}_{iD+1} - 1) \ln \left(1 - \sum_{k=1}^D x_{nk} \right) \right] \end{aligned} \quad (9)$$

式中: $\bar{\theta}_{ik} = \langle \theta_{ik} \rangle$

$$\tilde{\mathbf{R}}_i = \left\langle \ln \frac{\Gamma \left(\sum_{d=1}^D \theta_{id} \right) \Gamma \left(\theta_{iD+1} \right) \Gamma \left(\theta_{iD+2} \right)}{\prod_{d=1}^{D+2} \Gamma \left(\theta_{id} \right)} \right\rangle_{\theta_{i1}, \dots, \theta_{iD+2}} \quad (10)$$

$$\begin{aligned} \ln q^*(\theta_{il}) &= \sum_{n=1}^N \langle z_{ni} \rangle \mathcal{Q}(\theta_{il}) - \theta_{il} \sum_{n=1}^N \langle z_{ni} \rangle \ln \left(\sum_{k=1}^D x_{nk} \right) \\ &\quad - \nu_{il} \theta_{il} + (u_{il} - 1) \ln \theta_{il} + \theta_{il} \sum_{n=1}^N \langle z_{ni} \rangle \ln x_{nl} + \text{const} \end{aligned} \quad (11)$$

式中:

$$\mathcal{Q}(\theta_{il}) = \left\langle \ln \frac{\Gamma \left(\theta_{il} + \sum_{d \neq l}^D \theta_{id} \right)}{\Gamma(\theta_{il}) \prod_{d \neq l}^D \Gamma(\theta_{id})} \right\rangle_{\theta \neq \theta_{il}} \quad (12)$$

$$\begin{aligned} \ln q^*(\theta_{il}) &= \text{const} + \sum_{n=1}^N \langle z_{ni} \rangle \langle \mathbf{R}_i \rangle_{\theta \neq \theta_{il}} + \theta_{il} \sum_{n=1}^N \langle z_{ni} \rangle \\ &\quad \times \ln \left(1 - \sum_{k=1}^D x_{nk} \right) + (u_{il} - 1) \ln \theta_{il} - \nu_{il} \theta_{il} \end{aligned} \quad (13)$$

式中:

$$\mathbf{J}(\theta_{il}) = \left\langle \ln \frac{\Gamma(\theta_{iD+1} + \theta_{iD+2})}{\Gamma(\theta_{iD+1}) \Gamma(\theta_{iD+2})} \right\rangle_{\theta \neq \theta_{il}} \quad (14)$$

式中: $l \in [D+1, D+2]$.

由于式(10), (12)和(14)计算困难,不能直接用于变分推断,可使用非线性近似^[10,12]方法求解它们的下限,得到 BLM 参数的近似解:

$$q^*(\mathbf{Z}) = \prod_{n=1}^N \prod_{i=1}^M r_{ni}^{z_{ni}} \quad (15)$$

$$q^*(\Theta) = \prod_{i=1}^M \prod_{d=1}^{D+2} G(u_{id}^*, \nu_{id}^*) \quad (16)$$

式中:

$$r_{ni} = \frac{\rho_{ni}}{\sum_{i=1}^M \rho_{ni}} \quad (17)$$

$$\begin{aligned} \rho_{ni} &= \exp \left[\ln \pi_i + \tilde{\mathbf{R}}_i + \sum_{d=1}^D (\bar{\theta}_{id} - 1) \ln x_{nd} + \ln \left(\sum_{d=1}^D x_{nd} \right) \right. \\ &\quad \left. \times \left(\bar{\theta}_{iD+2} - \sum_{d=1}^D \bar{\theta}_{id} \right) + (\bar{\theta}_{iD+1} - 1) \ln \left(1 - \sum_{d=1}^D x_{nd} \right) \right] \end{aligned} \quad (18)$$

其中: $\tilde{\mathbf{R}}_i = \tilde{\mathbf{R}}_{\text{LIB}} + \tilde{\mathbf{R}}_{\text{MLIB}}$. 其中, $\tilde{\mathbf{R}}_{\text{LIB}}$ 表示 LIB 函数的下限, $\tilde{\mathbf{R}}_{\text{MLIB}}$ 表示 MLIB 函数的下限.

$$\begin{aligned} u_{il}^* &= u_{il} + \sum_{n=1}^N r_{ni} \bar{\theta}_{il} \left[\sum_{k=1, k \neq l}^D (\langle \ln \theta_{ik} \rangle - \ln \bar{\theta}_{ik}) \right. \\ &\quad \left. \times \Psi' \left(\sum_{k=1}^D \theta_{ik} \right) \bar{\theta}_{ik} + \Psi' \left(\sum_{k=1}^D \theta_{ik} \right) - \Psi'(\bar{\theta}_{il}) \right] \end{aligned} \quad (19)$$

$$\begin{aligned} u_{il}^* &= u_{il} + \sum_{n=1}^N r_{ni} \bar{\theta}_{il} \left[\Psi(\bar{\theta}_{iD+1} + \bar{\theta}_{iD+2}) - \Psi(\bar{\theta}_{il}) \right. \\ &\quad \left. + \Psi'(\bar{\theta}_{iD+1} + \bar{\theta}_{iD+2}) (\langle \ln \theta_{il} \rangle - \ln \bar{\theta}_{il}) \bar{\theta}_{il} \right] \end{aligned} \quad (20)$$

$$\nu_{il}^* = \nu_{il} - \sum_{n=1}^N r_{ni} \ln x_{nl} + \sum_{n=1}^N r_{ni} \ln \left(\sum_{k=1}^D x_{nk} \right) \quad (21)$$

$$\nu_{iL+1}^* = \nu_{iL+1} - \sum_{n=1}^N r_{ni} \ln \left(1 - \sum_{k=1}^D x_{nk} \right) \quad (22)$$

$$\nu_{iL+2}^* = \nu_{iL+2} - \sum_{n=1}^N r_{ni} \ln \left(\sum_{k=1}^L x_{nk} \right) \quad (23)$$

需要特别注意,在式(16), (18)中, $l \in [1, \dots, D]$, 式(17)中 $l \in [D+1, D+2]$. 上述更新式中的各期望值如下:

$$\langle z_{ni} \rangle = r_{ni}, \bar{\theta}_{ik} = u_{ik} / \nu_{ik}, \langle \ln \theta_{ik} \rangle = \Psi(u_{ik}) - \ln \nu_{ik} \quad (24)$$

$$\langle (\ln \theta_{ik} - \ln \bar{\theta}_{ik})^2 \rangle = [\Psi'(u_{ik}) - \ln \nu_{ik}]^2 + \Psi''(u_{ik}) \quad (25)$$

2.3 模型选择算法

变分贝叶斯推断中,变分下界是一个重要的量,可用来直接判断算法是否收敛.此外,由于变分下界具有在迭代过程中不会减少特点,因此可用来判断算法推导和实现是否存在问题.变分贝塔混合模型的下界为:

$$\begin{aligned} L(q) &= \langle \ln p(\mathbf{X} | \mathbf{Z}, \Theta) \rangle + \langle \ln p(\mathbf{Z} | \boldsymbol{\pi}) \rangle + \langle \ln p(\Theta) \rangle \\ &\quad - \langle \ln q^*(\Theta) \rangle - \langle \ln q^*(\mathbf{Z}) \rangle \end{aligned} \quad (26)$$

$$\begin{aligned} \langle \ln p(\mathbf{X} | \mathbf{Z}, \Theta) \rangle &= \sum_{n=1}^N \sum_{i=1}^M \left\{ \sum_{d=1}^D (\bar{\theta}_{id} - 1) \ln x_{nd} + (\bar{\theta}_{i, D+1} - 1) \right. \\ &\quad \left. \times \ln \left(1 - \sum_{d=1}^D x_{nd} \right) + (\bar{\theta}_{i, D+2} - \sum_{d=1}^D \bar{\theta}_{id}) \ln \left(\sum_{d=1}^D x_{nd} \right) \right\} \end{aligned} \quad (27)$$

$$\langle \ln p(\mathbf{Z} | \boldsymbol{\pi}) \rangle = \sum_{n=1}^N \sum_{i=1}^M r_{ni} \ln \pi_i \quad (28)$$

$$\langle \ln q^*(\mathbf{Z}) \rangle = \sum_{n=1}^N \sum_{i=1}^M r_{ni} \ln r_{ni} \quad (29)$$

$$\begin{aligned} \langle \ln p(\Theta) \rangle &= \sum_{i=1}^M \sum_{d=1}^{D+2} \left\{ u_{id} \ln \nu_{id} - \ln \Gamma(u_{id}) \right. \\ &\quad \left. + (u_{id} - 1) \langle \ln \theta_{id} \rangle - \nu_{id} \bar{\theta}_{id} \right\} \end{aligned} \quad (30)$$

$$\begin{aligned} \langle \ln q^*(\Theta) \rangle &= \sum_{i=1}^M \sum_{d=1}^{D+2} \left\{ u_{id}^* \ln \nu_{id}^* - \ln \Gamma(u_{id}^*) \right. \\ &\quad \left. + (u_{id}^* - 1) \langle \ln \theta_{id} \rangle - \nu_{id}^* \bar{\theta}_{id} \right\} \end{aligned} \quad (31)$$

最大化边缘似然函数 $\ln p(\mathbf{X} | \boldsymbol{\pi})$ 的下界 $L(q)$, 可获得混合权重的估计值. 对 $L(q)$ 关于 $\boldsymbol{\pi}$ 求偏导数并设置导数为零, 得到:

$$\pi_i = \frac{1}{N} \sum_{n=1}^N r_{ni} \quad (32)$$

由于变分后验因子 $q^*(z)$, $q^*(\Theta)$ 和变分下界 $L(q)$ 依

赖于 π , 所以可以利用类似于 EM 的迭代算法来求其解. 算法如下:

- (1) 初始化分量数 M , 超参数 $\{u_{id}, v_{id}\}$, 并利用 K 均值初始化 r_{ni} ;
- (2) 利用式(15), (16) 分别更新 $q^*(\mathbf{Z})$ 、 $q^*(\boldsymbol{\theta})$;
- (3) 关于当前的混合权重估计值, 即式(32), 最大化 $L(q)$;
- (4) 重复步骤(2), (3), 直至收敛;
- (5) 忽略混合权重小于 10^{-5} 的分量, 得到分量数 M ;
- (6) 从新估计式(15), (16) 中的参数.

3 实验结果及分析

为了验证本文所提出的变分贝叶斯算法的性能 (记变分 BLM 为 varBLM), 在合成数据集及真实数据集上, 我们进行了大量的实验. 实验中各参数的初始值选择如下: 初始分量数为 15, 初始混合权重等系数, 超参数 $(u_{id}, v_{id}) = (1, 0.01)$.

3.1 合成数据

实验中用程序产生了可以指定维数、分量数的服从 BLM 分布的四个合成数据集, 样本容量不同. 进行以下实验:

第一, 对比变分贝叶斯算法与 MML-EM 学习算法^[4] (记 MML-EM 学习的 BLM 为 DBLM) 的估计精度. 算法运行 15 次, 结果取统计平均值. 表 1 量化地给出了变分贝叶斯和 MML-EM 两种学习方法对于四个合成数据

集的参数学习结果. 通过对比两种算法发现, 利用 var-BLM 得到的估计值与真实值基本吻合, 其估计精度总体优于 MML-EM. 图 1 给出了变分贝叶斯算法对于四个合成数据集学习得到的概率密度图. 由图可以见, BLM 的概率密函数既可以是对称的, 也可以是非对称的, 因此它对比例数据有很强的描述能力;

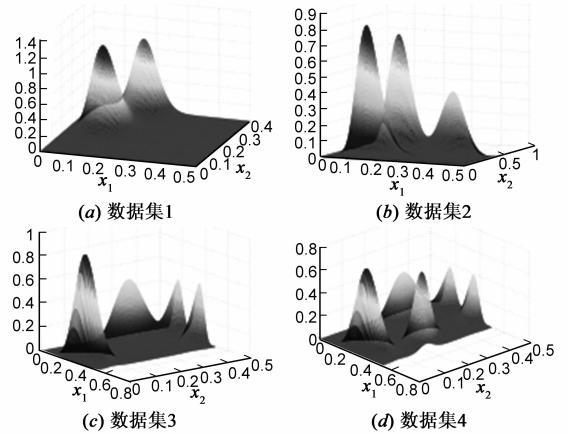


图1 合成数据集变分学习后的混合密度

第二, 算法初始化时给定的贝塔刘维尔分量个数为 15. 由于变分贝叶斯算法设置了隐变量, 每次迭代循环中更新完超参数后它都会重新计算各个分量的权重. 当隐变量所指示分量在加权求和时权重过小, 该分量是冗余分量, 可忽略它的存在, 即认为 $z_{ni} = 0$. 变分算法对于四个合成数据集的有效分量学习结果如图 2 所示. 由此可见, 该算法最终收敛后通过模型选择.

表 1 不同合成数据集的参数. 其中, N_i 代表第 i 簇内的样本数, θ_{id}, π_i 代表真实参数, θ_{id}^V, π_i^V 代表变分贝叶斯算法的估计结果, θ_{id}^D, π_i^D 代表 DBLM 算法的估计结果.

	N_i	i	θ_{i1}	θ_{i2}	θ_{i3}	θ_{i4}	π_i	θ_{i1}^V	θ_{i2}^V	θ_{i3}^V	θ_{i4}^V	π_i^V	θ_{i1}^D	θ_{i2}^D	θ_{i3}^D	θ_{i4}^D	π_i^D
数据集 1	160	1	5	16	35	26	0.5	5.23	16.80	36.19	27.05	0.498	4.56	16.92	36.26	25.37	0.481
	160	2	15	50	18	40	0.5	14.81	49.79	17.71	39.29	0.502	14.82	50.34	18.43	41.31	0.519
数据集 2	200	1	5	16	35	26	0.4	4.97	16.24	33.83	25.26	0.401	5.76	15.31	37.12	27.28	0.341
	200	2	15	50	18	40	0.4	14.59	48.82	17.31	38.50	0.399	15.73	53.32	16.27	38.25	0.432
	100	3	28	45	10	58	0.2	27.24	43.83	10.41	60.82	0.200	28.42	47.64	9.02	61.21	0.227
数据集 3	160	1	5	16	35	26	0.25	5.16	16.34	36.83	27.35	0.249	4.73	16.67	33.38	25.73	0.236
	160	2	15	50	18	40	0.25	15.91	52.93	17.12	37.35	0.252	15.23	54.27	18.13	36.13	0.261
	160	3	28	45	10	58	0.25	27.11	43.29	10.38	61.42	0.249	29.18	46.23	10.83	60.28	0.238
	160	4	44	10	25	14	0.25	44.31	10.00	25.43	14.29	0.250	43.27	10.49	25.31	14.11	0.265
数据集 4	160	1	5	16	35	26	0.2	4.79	15.23	35.92	27.00	0.200	5.31	17.12	36.73	25.14	0.226
	160	2	15	50	18	40	0.2	13.74	46.61	20.06	44.36	0.201	17.15	48.17	16.48	43.21	0.178
	160	3	28	45	10	58	0.2	28.19	45.22	10.23	59.78	0.200	2637	42.18	10.23	55.16	0.184
	160	4	44	10	25	14	0.2	42.72	9.79	25.86	14.44	0.199	45.58	9.47	26.12	13.74	0.235
	160	5	50	30	40	70	0.2	49.72	29.89	40.87	71.440	0.200	51.24	31.27	38.32	73.68	0.177

第三,对比变分贝叶斯与 MML-EM 的收敛速度,图 2 给出了两种算法对于四个合成数据集学习的收敛时间.由表可见,变分贝叶斯算法具有更快的收敛速度.

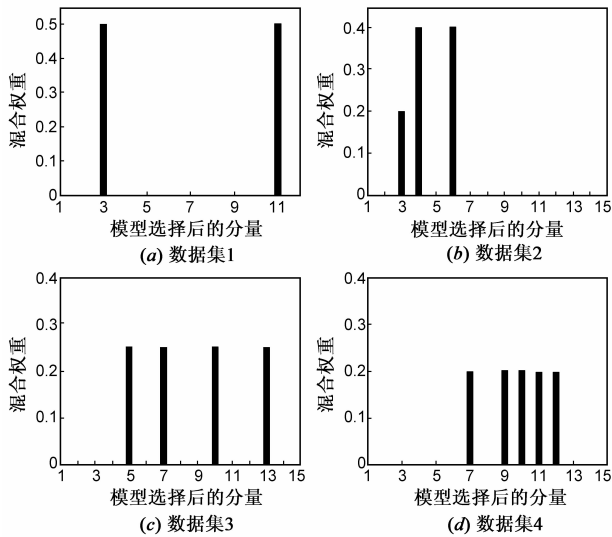


图2 模型选择后的有效分量

表 2 算法的收敛时间(s)

方法	varBLM	DBLM
数据集 1	3.84	8.24
数据集 2	3.21	7.76
数据集 3	3.04	6.57
数据集 4	3.54	7.13

3.2 场景图像分类

场景分类是计算机视觉中一个富有挑战性的课题,被广泛应用于图像库的管理和基于内容的图像检索.图像特征描述子提取是场景分类中关键的一步.近年来,基于视觉词袋模型(Bag-of-Visual-Word)^[13]的方法在场景分类问题中有出色的表现^[14].因此本文采用基于词袋模型的方法提取特征.

以下实验采用文献[15]中提到的图像集,由 8 类图像组成,共 2668 张图片,可以划分为人工图像(4 类)和自然图像集(4 类).实验中把这些图片集随机地分成两半,其中一半为训练集,另一半为测试集.考虑三种分类方案:所有图像分类,人工图像分类和自然图像分类.图像分类实现过程如下:

(1)提取 SIFT 图像特征描述子,每一副图片由 128 维的特征矢量集来表示;

(2)从每一类图片中,随机选择 35 幅图片的 SIFT 特征,利用 K 均值算法构建视觉单词数可指定的视觉词汇表 W ,实验中发现最佳的视觉单词数为 800;

(3)利用视觉词汇表 W ,统计词频;

(4)指定主题数,用 LDA(Latent Dirichlet Allocation)

训练词频统计数据.实验过程中发现最佳的主题数为 35,最后每一副图片由一个 35 维的比例矢量表示.

(5)对这些比例数据进行分类,实验运行 10 次,分类精度取统计平均值.其中, varDMM 表示的是变分有限狄利克雷混合模型^[16], SVM-P 表示多项式核的支持向量机.表 3 给出了不同分类算法的分类精度.由表可见,与其它分类算法相比, varBLM 具有更好的分类性能.

表 3 不同算法的平均分类准确率

	总图像集	人工图像	自然图像
varDMM	72.73%	75.37%	73.52%
DBLM	73.63%	75.71%	73.93%
varBLM	74.37%	76.65%	74.81%
SVM-P	71.18%	74.72%	72.85%
KNN	71.07%	74.51%	72.72%

4 总结

针对贝塔刘维尔混合模型的学习问题,本文提出了一种极具吸引力的变分推断框架.在这个变分框架内能够有效地估计混合模型参数及分量数.大量的仿真实验验证了该方法能够有效地估计模型参数同时确定混合分量数.此外,通过实验对比变分贝叶斯与 MML-EM 两种学习算法,前者在收敛速度、估计精度具有一定的优势,在实际的场景分类问题上有更好的性能.因此该算法是有效的和可行的. BLM 可用于比例数据统计建模,无监督学习和自然语言处理等领域.无限贝塔刘维尔混合模型及模型选择是下一步工作的重点.

参考文献

- [1] Everitt B S, Hand D J. Finite Mixture Distributions [M]. London: Chapman & Hall, 1981.
- [2] McLachlan G, Peel D. Finite Mixture Models [M]. New York: John Wiley, 2000.
- [3] Taoufik Bdiri, Nizar Bouguila. Positive vectors clustering using interted Dirichlet finite mixture Models [J]. Expert Systems with Applications, 2012, 39(2): 1869 – 1882.
- [4] Nizar Bouguila. Hybrid generative/discriminative approaches for proportional data modeling and classification [J]. IEEE Transactions on Knowledge and Data Engineering, 2012, 24(12): 2184 – 2202.
- [5] Bishop C M. Pattern Recognition and Machine Learning [M]. New York: Springer-Verlag, 2006.
- [6] Akaike H. A new look at the statistical model identification [J]. IEEE Transactions on Automatic Control, 1974, 19(6): 716 – 723.
- [7] Schwarz G. Estimating the dimension of a model [J]. The An-

nals of Statistics, 1976, 6(2): 461 – 464.

- [8] 刘伟峰, 杨爱兰. 基于 BIC 准则和 Gibbs 采样的有限混合模型无监督学习算法[J]. 电子学报, 2011, 39(3A): 134 – 139.
- LIU Wei-feng, YANG Ai-lan. Unsupervised learning for finite mixture models based on BIC Criterion and Gibbs Sampling [J]. Acta Electronica Sinica, 2011, 39(3A): 134 – 1139. (in Chinese)
- [9] Rissanen J. Modeling by shortest data description[J]. Automatica, 1978, 14(5): 456 – 471.
- [10] Ma Zhan-yu, Arne Leijon. Bayesian estimation of beta mixture models with variational inference [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(33): 2160 – 2173.
- [11] Corduneanu A, Bishop C. Variational bayesian model selection for mixture distributions [A]. In Proceeding of the 8th International Conference on Artificial Intell-igece and Statistics [C]. USA: IEEE, 2001. 27 – 34.
- [12] FanWen-tao, Nizar Bouguila, Djemel Ziou. Variaion-al learning for finite Dirichlet mixture models and applications[J]. IEEE Transaction on Neural Net-works and Learning Systems, 2012, 23(5): 762 – 774.
- [13] JosefSivic, Andrew Zisserman. Video google: A text retrieval approach to object matching in videos [A]. Proceedings of Ninth IEEE International Conference on Computer Vision [C]. 2003. USA: IEEE, 2003. 1470 – 177.
- [14] Jiang Yu-Gang, Ngo Chong-wah, Yang Jun. Towards optimal bag-of-features for object categorization and semantic video retrieval [A]. Proceedings of the 6th ACM International Conference on Image and Video Retrieval [C]. USA: ACM, 2007. 494 – 501.
- [15] Aude Oliva, Antonio Torralba. Modeling the shape of the scene: a holistic representation of the spatial envelope [J]. International Journal of Computer Vision, 2011, 42(3): 145 – 175.

- [16] David M Blei, Andrew Y Ng, Michael I Jordan. Latent dirichlet allocation [J]. Journal of Machine Learning Research, 2003, 3(33): 993 – 1022.

作者简介



赖裕平 男, 1983 年 10 月生于江西瑞金, 北京邮电大学博士研究生, 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



丁洪伟 男, 1964 年 5 月生于云南景洪, 云南大学信息学院副教授, 博士研究生, 主要研究方向为轮询系统、随机多址通信系统、优化算法研究。

周亚建 男, 1971 年生于陕西镇安. 北京邮电大学副教授、硕士生导师, 主要研究方向为移动通信、无线传感网络、信息安全等。

郭玉翠 女, 1962 年 5 月生, 北京邮电大学理学院教授. 研究方向为微分方程边值问题的分析与求解。

杨义先 男, 1961 年 3 月生, 北京邮电大学计算机学院教授、博士生导师, 长江学者, 主要研究方向为现代密码基础理论, 信息隐愿技术与理论。