

# 多视图有监督的 LDA 模型

李晓旭<sup>1</sup>, 李睿凡<sup>2,3</sup>, 冯方向<sup>2</sup>, 曹 洁<sup>1</sup>, 王小捷<sup>2,3</sup>

(1. 兰州理工大学计算机与通信学院, 甘肃兰州 730050; 2. 北京邮电大学计算机学院, 北京 100876;  
3. 教育部信息网络工程研究中心, 北京 100876)

**摘 要:** 本文主要关注多视图数据的分类问题. 考虑到集成分类方法可组合多个弱分类器构成一个强分类器, 以及主题模型能学习复杂数据的语义表示, 本文试图将集成学习思想引入主题模型中, 以便同时学习多视图数据的分类规则和预测性语义特征. 具体地, 结合概率主题模型 LDA 模型和集成分类方法 Softmax 混合模型, 提出了一个多视图有监督的分类模型. 基于变分 EM 方法, 推导了该模型的参数估计算法. 两个真实图像数据集上的实验结果表明了提出模型有较好的分类性能.

**关键词:** 多视图分类; 概率主题模型; 变分期望最大化

**中图分类号:** TP181

**文献标识码:** A

**文章编号:** 0372-2112 (2014)10-2040-05

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

**DOI:** 10.3969/j.issn.0372-2112.2014.10.26

## Multi-view Supervised Latent Dirichlet Allocation

LI Xiao-xu<sup>1</sup>, LI Rui-fan<sup>2,3</sup>, FENG Fang-xiang<sup>2</sup>, CAO Jie<sup>1</sup>, WANG Xiao-jie<sup>2,3</sup>

(1. College of Computer and Communication, Lanzhou University of Technology, Lanzhou, Gansu 730050, China;

2. School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China;

3. Engineering Research Center of Information Networks, Ministry of Education, Beijing 100876, China)

**Abstract:** In the paper, we mainly focus on classification on multi-view data. Considering that ensemble methods can combine weak classifiers to construct a strong classifier, and topic model can learn latent representations from complex data, we try to introduce ensemble idea to topic model, such that predictive latent representation could be obtained and multi-view classifier could be learned. We propose multi-view supervised latent Dirichlet allocation (multi-view sLDA) model by combining latent Dirichlet allocation model and the mixture of softmax model which is an ensemble classification model. Moreover, we derive a parameter estimation algorithm of the proposed model based on variational expectation maximization (EM) procedure. The experimental results on two real datasets show the effectiveness of the proposed model.

**Key words:** multi-view classification; probabilistic topic model; variational expectation maximization

## 1 引言

在计算机视觉和机器学习中,很多问题都包含多视图数据,例如一段视频可分为声音、图像、文本字幕等数据,一个物体可以分为颜色、形状等特征.目前,多视图数据的分类问题已被广泛关注,目的是希望利用多视图信息提高分类性能.传统的多视图分类方法包括特征水平上的融合<sup>[1]</sup>,即初期融合方法,以及输出水平的融合<sup>[2]</sup>,即后期融合方法.近期的研究逐渐开始关注中期融合方法.在中期融合方法中,一类方法是先使用无监督的特征学习方法学习多视图特征,如 CorLDA<sup>[3]</sup>, CCA<sup>[4]</sup>以及 Memisevic 提出的方法<sup>[5]</sup>,然后使用一般的单

视图分类方法预测其类别.这种两阶段的多视图分类方法,往往不能学到适合分类的特征表示.另一类方法是将特征学习和分类器学习整合为一个模型,以便能学到适合分类的特征表示,如 MMH<sup>[6]</sup>, MC-sLDA<sup>[7]</sup>以及文献[8]给出的多核学习方法.其中文献[8]的方法没有直接从多视图数据中学习特征表示,而是通过学习多视图的核来融合多视图数据.总之,这两种中期融合方法,都是从特征学习的角度来融合多视图信息.

本文提出了一种一阶段的中期融合方法,该方法没有特征学习部分而是在分类器学习部分融合多视图信息.其动机是,主题模型适合学习数据的语义特征,集成分类方法能将多个弱分类器组合为一个强分类器,将

其结合以便构建兼具二者优势的多视图分类方法. 本文建立在概率主题模型 LDA<sup>[9]</sup>和集成分类方法 Mixture of Softmax Model (SMM)<sup>[10]</sup>上,通过令 LDA 模型学习出的语义特征作为 SMM 模型的输入,提出了用于分类多视图数据的概率主题模型. 另外,两个模型的结合并不是简单的联合,增加了参数优化中求解 SMM 模型参数的困难. 本文使用一些近似技巧,并基于变分期望最大化 (EM) 方法,推导了提出模型的参数估计算法. 两个真实图像集上的实验结果表明了提出模型有较好的分类性能.

## 2 Mv-sLDA 模型

通过嵌入 LDA 模型到 SMM 中,提出多视图有监督的 LDA 模型——Multi-view sLDA (MV-sLDA) 模型. 该模型条件依赖视图数目  $H$  和视图的主题数目  $K_h, h \in \{1, 2, \dots, H\}$ , 并假设带有类标的多视图数据  $((V_1, V_2, \dots, V_H), c)$  的生成过程如下:

(1) 对于视图

$$V_h = \{v_{h1}, v_{h2}, \dots, v_{hM_h}\}, h \in \{1, 2, \dots, H\}$$

(a) 抽取主题比例  $\theta_h \sim p_{\text{dir}}(\alpha_h)$ .

(b) 对于每个视图词汇  $v_{hm}, m \in \{1, 2, \dots, M_h\}$ :

(i) 抽取主题分派  $z_{hm} | \theta_h \sim p_{\text{mult}}(\theta_h)$ .

(ii) 抽取视图词汇  $v_{hm} | z_{hm} \sim p_{\text{mult}}(\pi_{z_{hm}})$ .

(2) 对于类标  $c$

(a) 抽取“视图”的分派,  $s | y \sim p_{\text{mult}}(y)$ .

(b) 抽取类标  $c | z_h, s \sim p_{\text{softmax}}(\bar{z}_h, s, \mu)$ , 其中  $\bar{z}_h =$

$1/M \sum_{m=1}^{M_h} z_{hm}$  为经验主题频次, Softmax 函数为:

$$p(c | \bar{z}, s, \eta) = \prod_{h=1}^H \left( \frac{\exp(\eta_{hc}^T \bar{z}_h)}{\sum_{l=1}^C \exp(\eta_{hl}^T \bar{z}_h)} \right)^s$$

该模型确定了一个潜变量和观测变量的联合分布:

$$p(E, H | \Omega)$$

$$= \prod_{h=1}^H p(\theta_h | \alpha_h) \prod_{m=1}^{M_h} p(z_{hm} | \theta_h) p(v_{hm} | z_{hm}, \pi) \cdot p(s | y) p(c | \bar{z}, s, \eta) \quad (1)$$

其中,  $E = ((V_1, V_2, \dots, V_H), c)$ ,  $H = \{\theta_1, Z_1, \dots, \theta_H, Z_H, s\}$  和  $\Omega = \{\alpha_1, \pi_1, \dots, \alpha_H, \pi_H, y, \eta\}$ . 其图模型表示见图 1.

在上述生成过程中,步骤(1)描述了视图  $V_h = \{v_{h1}, v_{h2}, \dots, v_{hM_h}\}, h \in \{1, 2, \dots, H\}$  的生成过程,其中每个视图的生成是无序的. 该步骤目的是为了获得每个视图的经验主题频次. 步骤(2)建模了类标的生成过程,采用与 SMM 相似的集成方法,为每个视图构建一个分类器. 首先,选择一个视图的分派,然后根据该视图

上的分类器生成类标. 参数  $y$  记录了每个视图的权重,值越大意味着越重要.

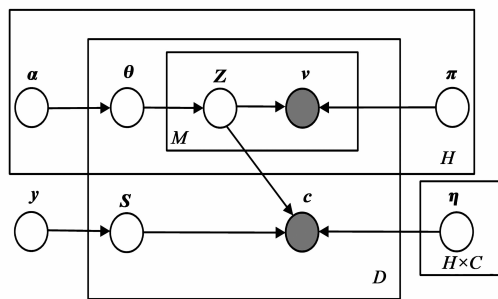


图1 Mv-sLDA模型的概率图模型表示

值得指出的是,提出模型既可分类单视图数据,亦可分类多视图数据,只需要将每个视图的特征表示成词袋表示形式.

## 3 参数估计和预测

### 3.1 参数估计

使用最大似然估计优化提出模型的参数. 考虑到潜变量上的后验概率  $p(H | E)$  很难计算,因而采用变分近似<sup>[10]</sup>来获得近似的后验分布. 给定一个数据  $((V_1, V_2, \dots, V_H), c)$ , 定义一个全分解的潜变量上的变分分布:

$$q(H | \Lambda) = \prod_{h=1}^H q(\theta_h | \gamma_h) \prod_{m=1}^{M_h} q(z_{hm} | f_{hm}) q(s | \lambda) \quad (2)$$

其中,  $\Lambda = \{\gamma_1, \phi_1, \dots, \gamma_H, \phi_H, \lambda\}$ ,  $\gamma_h$  是一个  $K_h$  维的 Dirichlet 参数,  $\phi_{hm}$  是一个  $K_h$  维的多项式参数. 给定模型  $\Omega = \{\alpha_1, \pi_1, \dots, \alpha_H, \pi_H, y, \eta\}$  和变分分布  $q$ , 基于 KL-散度, 可得到  $((V_1, V_2, \dots, V_H), c)$  的似然下限:

$$\log p(E | \Omega) \geq E_q[\log p(E, H | \Omega)] - E_q[\log q(H | \Lambda)] = L(\Lambda; \Omega) \quad (3)$$

则优化目标由原来的最大化  $\log p(E | \Omega)$  变为相对于变分参数  $\Lambda = \{\gamma_1, \phi_1, \dots, \gamma_H, \phi_H, \lambda\}$  和模型参数  $\Omega = \{\alpha_1, \pi_1, \dots, \alpha_H, \pi_H, y, \eta\}$  最大化  $L(\Lambda; \Omega)$ .

(1) 变分 E-step

在 E-step 中, 固定模型参数  $\Omega$ , 相对于变分参数  $\Lambda$  最大化  $L(\Lambda; \Omega)$ , 可得到

$$\gamma_{hi} = \alpha_{hi} + \sum_{m=1}^{M_h} \phi_{hmi} \quad (4)$$

$$\phi_{hmi} \propto \pi_{hmi} \exp\left(\psi(\gamma_{hi}) + \lambda_h \left(\frac{1}{M_h} \eta_{hci} - (b_h^T \phi_{hmi}^{\text{old}})^{-1} b_{hi}\right)\right) \quad (5)$$

其中, 符号 “ $\propto$ ” 代表“正比于”, 并且

$$b_{hi} = \sum_{l=1}^C \left( \exp\left(\frac{1}{M_h} \eta_{hli}\right) \prod_{f \neq m} \left( \sum_{j=1}^{K_h} \phi_{hfi} \exp\left(\frac{1}{M_h} \eta_{hji}\right) \right) \right)$$

和  $\boldsymbol{\varphi}_{hm}^{\text{old}}$  是上一次被更新的值.

$\lambda_h \propto$

$$\exp\left(\boldsymbol{\eta}_{hc}^T \bar{\boldsymbol{\varphi}}_h + \ln(\mathbf{y}_h) - \ln\left(\sum_{l=1}^C \prod_{m=1}^{M_l} \sum_{j=1}^{K_j} \boldsymbol{\varphi}_{mi} \exp\left(\frac{1}{M_h} \boldsymbol{\eta}_{hlj}\right)\right)\right) \quad (6)$$

$\lambda_h$  是当前数据分派给第  $h$  个视图分类器的概率,  $\lambda$  的值越大意味着该视图的特征分辨性越强.

迭代式(4),(5)和(6)直到数据的 log 似然(log-likelihood)方程(3)收敛.

## (2) M-step

重复 E-step  $D$  次, 得到所有数据的近似后验分布, 从而简化了 M-step 的计算. 在 M-step 中, 相对于模型参数  $\Omega = \{\boldsymbol{\alpha}_1, \boldsymbol{\pi}_1, \dots, \boldsymbol{\alpha}_H, \boldsymbol{\pi}_H, \mathbf{y}, \boldsymbol{\eta}\}$  最大化  $L(\mathbf{D}) = \sum_{d=1}^D L(\Lambda_d; \Omega)$ . 视图的主题和权重参数更新如下:

$$\boldsymbol{\pi}_{hij} \propto \sum_{d=1}^D \sum_{m=1}^{M_d} \boldsymbol{\varphi}_{dhmi} \mathbf{v}_{dhm}^j \quad (7)$$

$$\mathbf{y}_h \propto \sum_{d=1}^D \lambda_{dh} \quad (8)$$

对于视图的分类规则  $\boldsymbol{\eta}$  的优化, 挑选  $L$  中包含  $\boldsymbol{\eta}$  的项:

$$L[\boldsymbol{\eta}] = \sum_{d=1}^D \sum_{n=1}^H \lambda_{dn} \left[ \boldsymbol{\eta}_{hc}^T \bar{\boldsymbol{\varphi}}_{dh} - \ln\left(\sum_{l=1}^C E_q[\exp(\boldsymbol{\eta}_{hl}^T \bar{\mathbf{z}}_{dh})]\right) \right]$$

关于  $E_q[\exp(\boldsymbol{\eta}_{hl}^T \bar{\mathbf{z}}_{dh})]$  的计算, 采用多变量 delta 方法<sup>[11]</sup>,  $E_f(\mathbf{V}) \simeq f(E\mathbf{V}) + \frac{1}{2} \text{tr}[\partial^2 f(E\mathbf{V}) / \partial \mathbf{V} \partial \mathbf{V}^T \text{cov}(\mathbf{V})]$ . 其中,  $f(\mathbf{V})$  是一个  $K$  维空间的函数. 令  $f(\bar{\mathbf{z}}) = \exp(\boldsymbol{\eta}^T \bar{\mathbf{z}})$ , 则有  $E_q f(\bar{\mathbf{z}}) \simeq \exp(\boldsymbol{\eta}_{hc}^T \bar{\boldsymbol{\varphi}}) \left(1 + \frac{1}{2} \boldsymbol{\eta}_{hl}^T \text{cov}(\bar{\mathbf{z}}) \boldsymbol{\eta}_{hl}\right)$ . 另外, 为了防止过拟合, 采用 Logistic 回归<sup>[10]</sup>等分类器的优化中常用的方法, 在目标函数中加入一个正则化项  $-\tau \|\boldsymbol{\eta}\|^2$ , 则  $L[\boldsymbol{\eta}]$  近似为:

$$L[\boldsymbol{\eta}] \simeq - \sum_{d=1}^D \sum_{n=1}^H \lambda_{dn} \ln\left(\sum_{l=1}^C \exp(\boldsymbol{\eta}_{hl}^T \bar{\boldsymbol{\varphi}}_{dh}) \cdot \left(1 + \frac{1}{2} \boldsymbol{\eta}_{hl}^T \text{cov}(\bar{\mathbf{z}}_{dh}) \boldsymbol{\eta}_{hl}\right)\right) + \sum_{d=1}^D \sum_{n=1}^H \lambda_{dn} \boldsymbol{\eta}_{hc}^T \bar{\boldsymbol{\varphi}}_{dh} - \tau \sum_{h=1}^H \sum_{l=1}^C \boldsymbol{\eta}_{hl} \cdot \boldsymbol{\eta}_{hl} \quad (9)$$

$L[\boldsymbol{\eta}]$  相对于  $\boldsymbol{\eta}_{hc}$  的导数为:

$$\frac{\partial L[\boldsymbol{\eta}]}{\partial \boldsymbol{\eta}_{hc}} \simeq \sum_{d=1}^D \lambda_{dh} \mathbf{c}_d^c \bar{\boldsymbol{\varphi}}_{dh} - 2\tau \cdot \boldsymbol{\eta}_{hc} - \sum_{d=1}^D \lambda_{dh} \mathcal{A} \quad (10)$$

其中  $\mathcal{A} =$

$$\frac{\exp(\boldsymbol{\eta}_{hc}^T \bar{\boldsymbol{\varphi}}_{dh}) \left( \bar{\boldsymbol{\varphi}}_{dh} \left(1 + \frac{1}{2} \boldsymbol{\eta}_{hc}^T \text{cov}(\bar{\mathbf{z}}_{dh}) \boldsymbol{\eta}_{hc}\right) + \boldsymbol{\eta}_{hc}^T \text{cov}(\bar{\mathbf{z}}_{dh}) \right)}{\sum_{l=1}^C \exp(\boldsymbol{\eta}_{hl}^T \bar{\boldsymbol{\varphi}}_{dh}) \left(1 + \frac{1}{2} \boldsymbol{\eta}_{hl}^T \text{cov}(\bar{\mathbf{z}}_{dh}) \boldsymbol{\eta}_{hl}\right)}$$

其中,  $\bar{\boldsymbol{\varphi}}_{dhi} = \sum_{m=1}^{M_{dh}} \bar{\boldsymbol{\varphi}}_{dhmi} / M_{dh}$  并且  $\text{cov}(\bar{\mathbf{z}}_{dh})_{fi} =$

$\sum_{m=1}^{M_d} (\boldsymbol{\varphi}_{dhmi} 1(f=i) - \boldsymbol{\varphi}_{dhmf} \boldsymbol{\varphi}_{dhmi}) / M_{dh}^2$ . 显然得不到  $\boldsymbol{\eta}$  的封闭解, 因而采用共轭梯度法来优化  $\boldsymbol{\eta}$ .

重复调用 EM 步骤, 直到  $L(\mathbf{D}) = \sum_{d=1}^D L(\Lambda_d; \Omega)$  收敛.

## 3.2 预测分类

从提出模型的生成过程看, 类标的生成依赖每个视图的主题频次  $\bar{\mathbf{z}}$  和权重  $\mathbf{y}$ . 对于一个新数据, 需要使用训练好的模型参数  $\boldsymbol{\pi}$  计算每个视图的主题频次. 这里通过 E-setp 计算  $E_q[\bar{\mathbf{z}}] = \bar{\boldsymbol{\varphi}}$  来近似  $\bar{\mathbf{z}}$ . 得到视图的主题频次  $\bar{\mathbf{z}}$  后, 利用训练好的模型参数  $\boldsymbol{\eta}$  (即每个视图的分类规则) 和视图的权重参数  $\mathbf{y}$ , 采用加权求和方式, 最高值的决策将被选择为最终的决策. 具体计算公式如下:

$$C_* = \arg \max_{c \in \{1, 2, \dots, C\}} \sum_{h=1}^H \mathbf{y}_h * \frac{\exp(\boldsymbol{\eta}_{hc}^T \bar{\mathbf{z}}_h)}{\sum_{l=1}^C \exp(\boldsymbol{\eta}_{hl}^T \bar{\mathbf{z}}_h)} \\ = \arg \max_{c \in \{1, 2, \dots, C\}} \sum_{h=1}^H \mathbf{y}_h * \frac{\exp(\boldsymbol{\eta}_{hc}^T \bar{\boldsymbol{\varphi}}_{dh})}{\sum_{l=1}^C \exp(\boldsymbol{\eta}_{hl}^T \bar{\boldsymbol{\varphi}}_{dh})} \quad (11)$$

可见, 最终的决策考虑了每个视图的决策和权重, 这和提出模型的生成过程是一致的, 也符合人的一般的思维方式.

## 4 实验

### 4.1 数据和预处理

为了评估提出模型的性能, 选择了两个真实图像数据集. 一个是带有标注的场景分类数据集: LabelMe 的子集<sup>[12]</sup>. 该数据集包含 8 类自然场景图像, 共 1600 个图像. 另一个是带有标注的事件分类数据集: 8 类的 UIUC-Sport 数据集<sup>[13]</sup>, 共 1791 个图像. 图像和标注文本分别被作为两个不同的单视图特征.

(1) 对于图像视图, 采用文献[7]中相似的方法. 对于 LabelMe 数据集相关参数设置如下: 设置网格的大小为  $5 \times 5$ , 图像区块为  $16 \times 16$ , 码书长度为 240. 对于 UIUC-Sport 的数据, 所有图像一致地抽取 2500 图像区块, 每个区块的大小为  $32 \times 32$ , 码书长度为 240.

(2) 对于文本视图, 使用所有不同的标注词构成标注文本的码书, 则图像在每个标注词维度上的值, 就是该标注词的出现的频次. 在 LabelMe 数据上平均每个图像有 6 个标注词, 对 UIUC-Sport 的数据上平均每个图像有 8 个标注词.

最后, 平分每个类别来生成训练和测试集.

### 4.2 分类性能

为了评估提出模型的性能, 选择如下几个方法进行比较: (1) MC-sLDA, (2) SBMLR<sup>[14]</sup>, (3) SVM-POL<sup>[15]</sup>, (4) SVM-RBF<sup>[15]</sup>, (5) Fu-L<sup>[2]</sup> 和 (6) MCA-sLDA. MC-

sLDA 是有监督的主题模型,该模型嵌入 softmax<sup>[10]</sup>模型到LDA模型中,并且在上述两个数据集上报告了较高的分类性能. SBMLR 是一个带有 Laplace 先验的 softmax 分类器,常常具有很好的泛化能力. 对于两个 SVM 方法,本文使用 libsvm 进行测试.

Fu-L 模型是先分别在图像和文本视图上构建单分类器,然后使用第三个分类器综合前面的结果. MCa-sLDA 模型是可同时做图像分类和标注的概率主题模型,也可看做使用了文本和图像两个视图的分类模型.

表 1 给出了提出模型 Mv-sLDA 在文本和图像两个视图数据上的性能. 为了方便起见,在两个视图上设置相同的主题数目. 实际上,可以设置不同的主题数目. Mv-sLDA 模型中的视图的权重参数  $\mathbf{y}$  是二维数组,第一维表示图像的权重,第二维表示文本的权重. 实验结果表明:两个数据集上文本特征的权重比图像特征都较高,及文本特征较为重要,即有更好的分辨性. 提出模型在 LabelMe 数据上可达到 92.2% 准确率,在 UIUC-Sport 数据上可达到 99.0% 的准确率.

表 1 在 UIUC-sport 和 LabelMe 数据集的 5 个随机训练和测试子集上 Mv-sLDA 模型的平均性能和视图的权重分派. AA-表示平均性能,Weight-表示权重的分派. 二元组的第一个元素表示图像视图的权重,第二个元素表示文本视图的权重

多视图	$K1 = K2 = 5$	$K1 = K2 = 10$	$K1 = K2 = 20$	$K1 = K2 = 30$	$K1 = K2 = 40$
AA-LabelMe	0.796	0.843	0.882	0.905	0.922
AA-UIUC	0.876	0.882	0.990	0.987	0.990
Weight-LabelMe	(0.59, 0.41)	(0.33, 0.67)	(0.23, 0.77)	(0.12, 0.88)	(0.22, 0.78)
Weight-UIUC	(0.03, 0.97)	(0.01, 0.99)	(0.00, 1.00)	(0.00, 1.00)	(0.00, 1.00)

表 2 在 LabelMe 数据集的 5 个随机训练和测试子集上模型的平均性能比较. 图像视图是指当图像特征作为一个单视图,文本视图指文本特征作为单视图,联合特征指文本和图像的联合特征作为单视图,多视图指将文本和图像的分别作为两个不同的视图

视图类型	SVM-POL	SVM-RBF	SBMLR	MC-sLDA	Mca-sLDA	Fu-L	Mv-sLDA
图像视图	0.779	0.811	0.748	0.766	N/A	N/A	0.768
文本视图	0.789	0.879	0.860	0.890	N/A	N/A	0.890
联合视图	0.889	0.837	0.778	0.789	N/A	N/A	0.789
多视图	N/A	N/A	N/A	N/A	0.768	0.822	0.922

表 3 在 UIUC-sport 数据集的 5 个随机训练和测试子集上模型的平均性能比较. 图像视图是指当图像特征作为一个单视图,文本视图指文本特征作为单视图,联合特征指文本和图像的联合特征作为单视图,多视图指将文本和图像的分别作为两个不同的视图

视图类型	SVM-POL	SVM-RBF	SBMLR	MC-sLDA	Mca-sLDA	Fu-L	Mv-sLDA
图像视图	0.647	0.692	0.643	0.640	N/A	N/A	0.639
文本视图	0.976	0.978	0.980	0.981	N/A	N/A	0.983
联合视图	0.493	0.932	0.916	0.660	N/A	N/A	0.662
多视图	N/A	N/A	N/A	N/A	0.670	0.841	0.990

### 4.3 结果分析

对于上述实验结果,可总结为以下三点:(1) 四个单视图方法在联合视图上的性能有时低于在单视图上的性能,即使用联合视图的方法不能很好利用多视图

对于方法(1)~(4),分别测试了在两个数据集的图像视图、文本视图以及将图像和文本特征平行拼接而构成的联合试图上的性能. 对于方法(5)~(6)和本文的方法,测试了在图像和文本两个视图上的性能. 对于 MC-sLDA 模型和 MCa-sLDA 模型,均匀的从 20 到 100 选择 5 组主题数目,并选其最佳性能,见从表 2 可以看出,在 LabelMe 数据集上,当只使用图像视图的时候,单视图方法的最好性能是 SVM-RBF 的 81.1%. 当只使用文本视图的时候,单视图方法的最好性能是 MC-sLDA 的 89.0%. 当使用图像和文本的联合试图时,四个单视图方法的最好性能是 SVM-POL 的 88.9%. 另外,四个单视图方法在联合视图上的性能有时居然比在单视图上的性能还差一些. 在多视图方法中,后期融合方法 Fu-L 的性能是 82.2%,中期融合方法 MCa-sLDA 的性能是 76.8%. 而提出方法 Mv-sLDA 的性能是 92.2%.

在 UIUC-Sport 数据集上,提出模型的性能也有相似的提升,见表 3. 总之,提出的模型在这两个数据集上有较高的分类性能.

信息,而提出的模型却表现的比较好. 原因在于,在上述两个数据集中,文本特征的值相比图像特征的值较小,然而又比较重要. 两个视图的值较大的差距增加找到这些分类器的最优解的难度,即便这些分类器的优

化程序一般能够辨别不同的特征的重要性. 特别在 SBMLR 方法中,  $L1$  约束的使用使得该方法更难辨别两个视图特征的重要性. 而提出的模型属于中期融合方法, 通过构建集成分类器来融合多视图信息, 避免上面提到的使用联合视图的问题.

(2) 提出的模型表现的比后期融合方法 Fu-L 好, 原因也许是: 在 Fu-L 中, 每个视图分类器的训练是独立的. 相反, 在提出的模型中, 分类器的训练是交互的. 另外, 提出的模型能学习适合分类的预测特征, 以致更容易构建适合分类多视图数据的分类器.

(3) 提出的模型表现的比多视图中期融合方法 MCa-sLDA 好, 是因为本文使用的两个数据集上文本特征值比较小, 并且高频词对主题模型 MCa-sLDA 学习语义特征的影响大, 以致文本特征对语义特征学习有较小的作用. 而提出的模型集合了特征学习和集成分类器的学习, 并选择在视图分类器的训练阶段利用多视图信息, 从而避免该问题.

## 5 结论

本文提出了一个分类多视图数据的概率主题模型 Mv-sLDA, 并基于变分 EM 推导了参数估计算法. 两个真实图像数据集上的实验结果表明了提出模型有较好的分类性能, 也表明了该模型充分利用了多视图信息. 下一步工作, 将在本文算法的基础上, 考虑如何确定主题数目, 以及当给定一组特征时, 如何确定特征的视图分组等问题.

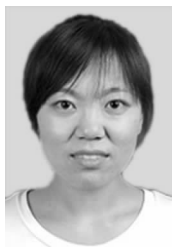
## 参考文献

- [1] Wu L, et al. Multimodal integration—A statistical view[J]. IEEE Transactions on Multimedia, 1999, 1(4): 334–341.
- [2] Wang G, et al. Building text features for object image classification[A]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR-09)[C]. IEEE, 2009. 1367–1374.
- [3] Blei D M and Jordan M I. Modeling annotated data[A]. Proc of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval[C]. ACM,

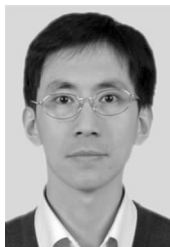
2003. 127–134.

- [4] Frawley W J, et al. Knowledge discovery in databases: An overview[J]. AI Magazine, 1992, 13(3): 57.
- [5] Memisevic, R. On multi-view feature learning[A]. Proceedings of the 29th International Conference on Machine Learning (ICML-12)[C]. New York, USA: Omnipress, 2012. 161–168.
- [6] Chen N, et al. Predictive subspace learning for multi-view data: A large margin approach[A]. Advances in Neural Information Processing Systems[C]. Vancouver: Curran Associates, 2010. 361–369.
- [7] Wang C, et al. Simultaneous image classification and annotation [A]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR-09)[C]. IEEE, 2009. 1903–1910.
- [8] Sonnenburg S, et al. Large scale multiple kernel learning[J]. The Journal of Machine Learning Research, 2006, 7: 1531–1565.
- [9] Blei D M, et al. Latent dirichlet let allocation[J]. The Journal of Machine Learning Research, 2003, 3: 993–1022.
- [10] Bishop C M, et al. Pattern Recognition and Machine Learning [M]. Springer New York, 2006. 461–674.
- [11] Braun M, McAuliffe J. Variational inference for large-scale models of discrete choice[J]. Journal of the American Statistical Association, 2010, 105(489): 324–335.
- [12] Russell B C, et al. Labelme: A database and web-based tool for image annotation[J]. International Journal of Computer Vision, 2008, 77(1): 157–173.
- [13] Li L J, Fei-Fei L. What, where and who? classifying events by scene and object recognition [A]. IEEE 11th International Conference on Computer Vision (ICCV 2007) [C]. IEEE, 2007. 1–8.
- [14] Cawley G C, et al. Sparse multinomial logistic regression via bayesian  $L1$  regularisation[A]. Advances in neural information processing systems [C]. Cambridge MA USA: MIT Press, 2007. 209–216.
- [15] Chang CC, Lin C J. Libsvm: A library for support vector machines [J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2011, 2(3): 27.

## 作者简介



李晓旭 女, 1982 年生于吉林白城. 兰州理工大学计算机与通信学院讲师. 研究方向为计算机视觉、机器学习.  
E-mail: xiaoxulit@gmail.com



李睿凡 男, 1975 年生于河北完县. 北京邮电大学计算机学院讲师. 主要研究兴趣为多模态智能信息处理.  
E-mail: rli@bupt.edu.cn