

# 一种基于类主题空间的图像场景分类方法

唐颖军<sup>1),2)</sup> 须德<sup>1)</sup> 解文杰<sup>1)</sup> 薄一航<sup>1)</sup>

<sup>1)</sup>(北京交通大学计算机与信息技术学院,北京 100044) <sup>2)</sup>(江西财经大学软件与通信工程学院,南昌 330013)

**摘要** 本文在扩展 LDA(latent dirichlet allocation)的基础上提出了一种新的生成模型——基于类主题空间的潜在狄里克雷分布(CTS-LDA)用来实现自然图像场景分类。该方法不同于以往方法,它在训练时通过将图像场景类别信息引入模型推导过程中,产生各场景类的独立语义主题空间,使得每个场景类都有各自不同的主题空间,图像的最终语义表示采用与其类别相关的类主题集,是一种符合人类认知习惯的方法。以前所用的场景分类方法通常在得到图像主题表示后还需要依赖于其他分类器来完成场景分类,而 CTS-LDA 模型可以在分别计算图像在各类模型中的主题分布时,用最大似然法得出图像的分类信息。此外本文通过分析不同主题数对本模型性能的影响,得出了适用于本模型的最佳主题数。本文分别通过 13,15 等多类场景任务来检验模型的性能,实验证明该模型能够在不需要太多训练的情况下取得较好的性能。

**关键词** 基于类主题空间的潜在狄里克雷分布 潜在狄里克雷分布 概率潜在语义分析 场景分类

中图法分类号: TP18/TP242.6+2 文献标志码: A 文章编号: 1006-8961(2010)07-1067-07

## A Novel Image Scene Classification Method Based on Category Topic Simplex

TANG Yingjun<sup>1),2)</sup>, XU De<sup>1)</sup>, XIE Wenjie<sup>1)</sup>, BO Yihang<sup>1)</sup>

<sup>1)</sup>(Computer and Information Technology School, Beijing Jiaotong University, Beijing 100044)

<sup>2)</sup>(Software and Communication Engineer School, Jiangxi University of Finance and Economics, Nanchang 330032)

**Abstract** The paper presents a novel model named Category Topic Simplex-Latent Dirichlet Allocation(CTS-LDA) based on extending LDA(Latent Dirichlet Allocation), which is used to learn and recognize natural scene category. Unlike previous work, our model can absorb category information in the model inference under learning process to produce its semantic simplex for each category scene. As a result, each category has its own semantic topic simplex, and each image can choose its simplex to denote, which is consistent with people's cognitive pattern. In previous work, recognizing scene category task need to resort to additional classifier after getting images topic representation. Our method can use ML method to recognize image category during the same time of getting topic representation. Furthermore, we also analyze the influence of the topic size in our model, and infer the fittest result to produce the best performance. We investigate the classification performance under changed scene category tasks. The experiments have demonstrated that our model can perform better with less training data than other methods.

**Keywords** Category Topic Simplex-Latent Dirichlet Allocation, Latent Dirichlet Allocation, Probabilistic Latent Semantic Analysis, scene classification

## 0 引言

随着 Internet 和多媒体技术的发展,现代信息

处理的对象与方法均有很大的变化。数字图像作为现代信息传播的重要媒体,其数量的猛增使得对其进行有效管理变成一个紧迫的任务。场景分类其实就是根据图像语义内容进行图像的自动分

基金项目:国家自然科学基金(90820013);江西省教育厅资助科研项目(GJJ09563)

收稿日期:2009-07-15;改回日期:2010-03-10

第一作者简介:唐颖军(1976—),女,讲师。北京交通大学计算机科学技术专业博士研究生。主要研究方向为图像场景分类、图像处理、远程控制等。E-mail:ncx\_tang@tom.com,07112047@bjtu.edu.cn

类,将图像分成特定类别(如海岸、高山、街道、城市、森林等)。20世纪80年代,图像分类主要依靠人工标注的方法。随着计算机和数字化图像技术的发展,图像数据库的规模越来越大,人们逐渐将研究的重点转移到基于语义分析的自动分类研究上<sup>[1]</sup>。因此场景分类已经成为计算机视觉中的一个重要任务。

基于语义分析的场景分类目前主要是通过采用中间语义特征(以下称为视觉词语)来表示图像<sup>[2-4]</sup>,以解决语义鸿沟(semantic gap)。视觉词语表达了图像局部特征的语义信息,每幅图像可以由多个视觉词语表示的。由于这种表示图像的方法借用高维来保留图像所隐藏的特征,其具体维数与所定义的视觉词语个数,即词本尺寸(词汇量)有关。而通常维数太高必然与维数灾难(course of dimension)相联系。因此必须考虑在保留图像语义表示的基础上降低维数并进一步推导其潜在的主题表达。当前场景分类所采用的降维方法主要是最初用于文本分类的概率潜在语义分析(pLSA)和潜在狄里克雷分布(LDA)。Anna等人<sup>[5]</sup>采用pLSA方法实现了8类场景分类。Li等人<sup>[6]</sup>通过对LDA模型的扩展实现了13类场景分类。

采用pLSA模型有两个缺点<sup>[7]</sup>,首先,参数量会随着文件数量线性扩增。因为pLSA将图像特征降维后以 $P(z|d)$ 作为图像 $d$ 在低维主题空间 $z$ 中的二次表示,其低维空间中的二次表示有 $K \times M$ 个参数,其中 $K$ 为主题数量, $M$ 为文件数,可见随着文件数的增加必然导致参数增加,最终影响分类性能;其次,没有直接的方法将概率分配给先前未出现(unseen)文件(即测试图像),只能通过固定 $P(w|z)$ 对最大似然值:

$$\sum_d \sum_w n(d,w) \log P(d,w) \quad (1)$$

进行判断来计算其 $P(z|d)$ ,其中 $n(d,w)$ 为共生矩阵表示图像 $d$ 中对应视觉词语 $w$ 中频率。

LDA是近年来由Blei提出的用于模拟文本的潜在主题分布的模型<sup>[7]</sup>,它解决了pLSA模型中的上述缺点。LDA是一个较完整的生成模型,它将每个文件的概率视为潜在主题中随机出现视觉词语的概率的混合模型。

由于pLSA和LDA都是无监督学习方法,在完成降维后必须与具有判别能力的分类器(如支持向量机、K近邻等)结合才能完成分类判别,如文献

[5]在文中采用了K近邻与pLSA相结合来实现;Quelhas采用支持向量机来对pLSA降维后图像的二次表示实现场景分类<sup>[8]</sup>;Wenbo通过LDA与支持向量机组合实现中文文本分类<sup>[9]</sup>;Elango采用LDA降维后通过K均值(K-means)来进行图像分类<sup>[10]</sup>。由此可见,他们场景分类的最终性能对分类器的选择有一定的依赖性。此外,这两种模型对于各类场景降维表示都是基于同一个主题空间,即各场景类的二次表示采用相同的主题集,区别仅在于比重不同。这与现实生活中人们所理解场景的方法不同。例如就海滩、高山、厨房3类场景而言,也许海滩和高山可能会出现部分相同主题——树,但海滩场景的主要主题构成应是海和沙滩,高山场景的重要构成是山,而对厨房场景而言,海、沙滩、山都不应是其主题。所以pLSA和LDA采用一个主题空间的做法对于多类相异场景是很不适用的。尽管可以通过采用增加主题个数方法来改善,但主题数增多最终有悖于降维的初衷。

因此,需要在保持图像语义特征的基础上构造一个有效模型来学习潜在主题分布。同时摆脱依赖于分类器的局限性,在学习图像主题分布的过程中完成图像的分类工作。本文在LDA模型的基础上提出一种新的学习模型——基于类主题空间的潜在狄里克雷分布(CTS-LDA),将各场景的类别信息加入在主题推导过程中,使每类场景在学习过程中得出各自的潜在主题空间,使得图像的最终语义按人类的认知方式来表示。最重要的是CTS-LDA模型能够在判别过程中通过判别选择相应的类主题空间来完成图像的场景类别的判定,进而得出图像二次表示。这样就可避免依赖于其他分类器所产生的性能影响。

## 1 本文生成模型

鉴于pLSA和LDA中所存在的上述不足,通过在LDA模型的训练过程进行图像场景类别约束而产生一个改进后的CTS-LDA模型。本文的主要目的是寻找一个生成模型,使得各种图像场景能够按人类认知方式选择在最适合的类主题空间来表示其各视觉词语的分布,并实现图像语义主题表示。

### 1.1 基本定义

为便于后续讨论,以下将对应文献[7]所提出

的 LDA 模型中的术语来定义 CTS-LDA 模型中的术语。

1) 变量  $v$  是视觉词语 (visterm), 表示图像局部特征, 也是码本中的一个码字, 阴影节点表示其是已知变量, 即在图像中可直接观察到的变量;

2) 图像用  $W$  来表示, 是由一系列的  $v$  组成, 即  $W = (v_1, v_2, \dots, v_N)$ 。各图像采用 BoV (bag of visterms) 模型来表示;

3) 图像集合是由多幅图像组成, 表示为  $D = (W_1, W_2, \dots, W_M)$ 。

其中码本 (codebook) 中码字 (即视觉词语) 按 SIFT 特征描述聚类后形成的。各图像用各码字所出现的频率所组成的向量来表示的, 即对图像中的所有视觉词语进行直方图统计 (BoV 模型) 而得出。而图像集合则根据训练和测试任务而变化。在训练期间该集合是由同类场景的图像所组成, 针对多类场景分类任务则存在多个图像集合; 而在测试期间, 所有测试图像不区分类别而组成一个图像集合作为 CTS-LDA 模型的输入数据。

### 1.2 模型结构

本文工作是基于 Li 等人提出的 Theme modell<sup>[6]</sup> (如图 1(b) 所示) 和经典的 LDA 模型<sup>[7]</sup> (图 1(c) 所示) 进行扩展和改进的。将图像场景的类别信息引入模型, 对各个主题进行类别约束, 使图像能够在其对应类主题空间下实现低维表示, 其模型如图 1(a) 所示。

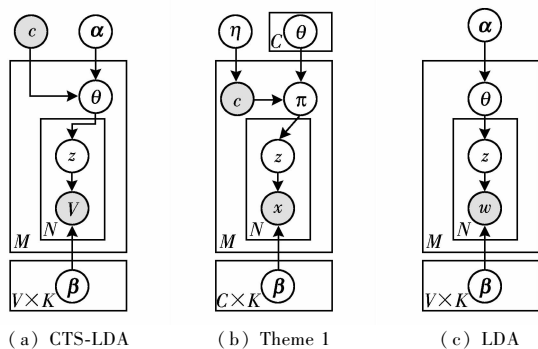


图 1 LDA 模型及扩展模型

Fig. 1 LDA Model and its extending models

由图 1(a) 可看出 CTS-LDA 模型有 3 层, 分别是集合层、图像层和视觉词语层。在这个模型中为每一个场景类分别建立不同的主题空间 (称之为类主题空间), 而不是如上述模型<sup>[6-7]</sup> 中全部场景共享同一个主题空间。CTS-LDA 的这种方法与日常认

知习惯是相符的, 因为它将不同场景分别在不同的主题空间表示。其中参数  $\alpha$  和  $\beta$  是集合层的参数, 对各类图像集合都是不同的;  $\theta$  是图像层参数, 对每个图像都有不同的值;  $z$  和  $v$  则是视觉词语层参数, 其中  $z$  对应当前类别模型中所有的潜在主题,  $v$  对应于图像中每个局部特征块以码本中对应的码本 (视觉词语) 来表示。

由图 1(a) 可知, 对于特定场景类图像, 给定集合参数  $\alpha$  和  $\beta$  就可以得到该图像的边缘概率分布

$$p(W | \alpha, \beta, c) = \int p(\theta | \alpha) \left( \prod_{n=1}^N \sum_{z_n} p(z_n | \theta, c) \times p(v_n | z_n, \beta) \right) d\theta \quad (2)$$

式中,  $\alpha$  是  $K$  维向量, 表示在此场景类中  $K$  个主题相对强弱; 视觉词语的概率由参数化的  $K \times V$  矩阵  $\beta$  表示, 其中  $\beta_{ij} = p(w_j = 1 | z_i = 1)$  表示当前主题对应的该视觉词语的概率。主题个数  $K$  和码本容量  $V$  在本模型中视为固定值, 将在第 3 节给出具体的设置方法;  $\theta$  是服从以  $\alpha$  为参数的 Dirichlet 分布;  $z_n$  则服从以  $\theta$  为参数的多项式分布, 是通过类别  $c$  所对应的  $\theta$  计算后验分布而得到。

### 1.3 推导及判别

由于式 (2) 存在两个未知变量  $\theta$  和  $\beta$ , 无法直接来求解。求解的最简单方法是使用 Jensen 不等式来简化模型而得到一个可解的方程下界<sup>[7]</sup>, 即将位于  $\theta$  和  $\beta$  对  $z$  的连线和结点  $v$  去掉, 如图 2 所示, 其中虚线部分为去掉部分。即定义一个任意分布的近似群。

$$q(\theta, z | \gamma, \phi, c) = q(\theta, \gamma) \prod_z q(z | \phi_z, c) \quad (3)$$

对于每个场景类  $c$ , 分别通过优化参数  $\gamma$  和  $\phi$  来计算  $p(\theta, z | W, \alpha, \beta, c)$  和  $q(\theta, z | \gamma, \phi, c)$  之间 KL Divergence 最小化来得到  $p(W | \alpha, \beta, c)$  的近似值如

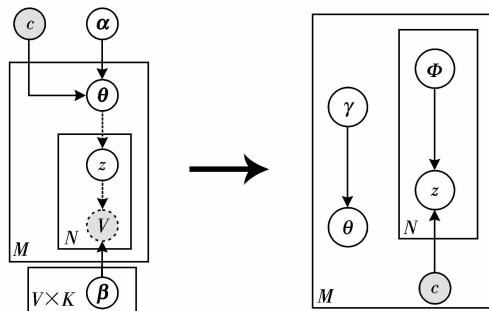


图 2 简化后的模型

Fig. 2 The model after simplification

下式

$$\begin{aligned} \log p(W | \alpha, \beta, c) &= L(\gamma, \phi; \alpha, \beta) + \\ &D(q(\theta, z | \gamma, \phi, c) \| p(\theta, z | W, \alpha, \beta, c)) \approx \\ &L(\gamma, \phi; \alpha_c, \beta_c) = \\ &E_q[\log p(\theta | \alpha_c)] + E_q[\log p(z | \theta)] + \\ &E_q[\log p(W | z, \beta_c)] - E_q[\log q(\theta)] - \\ &E_q[\log q(z)] \end{aligned} \quad (4)$$

采用 EM 算法来对 CTS-LDA 模型中的各个类别模型进行未知参数估计,具体步骤如下:

1) 设置  $\alpha$  和  $\beta$  初值,只作为 EM 算法而调用,对于  $\alpha$  取  $K \times 1$  的列向量,其中各元素值分别为 1 至 2 间的随机数,  $\beta$  则是取  $K \times V$  的矩阵,其各元素值为 0.01 至 1 的随机数,并进行归一化,保证各行值的和为 1。

2) 估计步骤(E 步):根据  $\alpha$  和  $\beta$  值,为每幅图像生成最优的  $\gamma$  和  $\phi$ ;

$$\begin{aligned} \phi_{ni} &= \beta_{i v_n} \exp \{ E_q[\log(\theta_i) | \gamma] \} \\ \gamma_i &= \alpha_i + \sum_{n=1}^N \phi_{ni} \end{aligned} \quad (5)$$

式中,指数部分为 Digamma 函数,  $v_n$  表示在当前图像词包表示中第  $n$  个视觉词语,  $\gamma$  表示图像在相应主题空间下的主题分布(即图像的二次表示)。

3) 最大化步骤(M 步):最大化似然值  $L(\gamma, \phi; \alpha, \beta)$  来求出新的  $\alpha$  和  $\beta$ ;

$$\begin{aligned} \beta_{ij} &= \sum_{d=1}^M \sum_{n=1}^{N_d} \phi_{dni} v_{dn}^j \\ \alpha_{\text{new}} &= \alpha_{\text{old}} - H(\alpha_{\text{old}})^{-1} g(\alpha_{\text{old}}) \end{aligned} \quad (6)$$

式中,  $v_{dn}$  表示在图像  $W_d$  的词包表示中第  $n$  个视觉词语,  $j$  则为在码本中对应第  $n$  个视觉词语中的序号,  $i$  为主题序号,  $N_d$  为当前图像  $W_d$  所含的视觉词语个数;  $H(\alpha)$  和  $g(\alpha)$  分别是以  $\alpha$  为参数的 Hessian 矩阵和梯度值。

4) 迭代计算前两步,直到似然值变化达到很小的一个阈值则结束。

在训练过程中,完成 EM 计算后,就可以得到 CTS-LDA 模型,它由各个场景类模型组成,分别对应不同的  $\alpha$  和  $\beta$ 。

尽管 CTS-LDA 是一种生成学习模型,但可以在计算图像二次表示  $\gamma$  的同时完成该图像的场景类别,这也是本文的创新点之一。因为  $\gamma$  是图像层的一个 Dirichlet 参数,是图像在主题空间中的二次表达。在测试过程中,分别用各个场景类模型来计算图像近似表示的  $\gamma$ ,通过最大似然估计(ML)来判定图像表示所需的类主题空间,进而来选择最大值所

对应的  $\gamma$ ,即该图像在主题空间中的二次表示;正如式(7)所示,在比较各个场景类模型所产生的  $L(\gamma, \phi; \alpha, \beta)$  时,通过选择最大值所对应的场景类  $c$  来作为图像的场景类别,故而一步完成生成与判别过程。而与此不同的是<sup>[5,7-8]</sup>,它们需要借助其他的分类判别算法来实现场景分类的判别。

$$\gamma = \alpha_c + \sum_{n=1}^N \phi_{nc} \quad (7)$$

$$\alpha_c, \beta_c = \arg \max \{ L(\gamma, \phi; \alpha_j, \beta_j), j \in 1, \dots, C \}$$

$$\phi_{nc} = \beta_{c v_n} \exp \{ E_q[\log(\theta_c) | \gamma] \}$$

## 2 CTS-LDA 场景分类方法

本场景分类过程实际上就是一个实现二次映射的过程,首先对图像提取的特征以聚类方式产生一个多对一的映射,形成码本,各个图像分别按词包模型方法表示,通过 CTS-LDA 模型来再次降维,在各个对应类的主题空间中来表示少量主题对全部图像视觉词语的分布,即是个多对全的映射。其映射关系如图 3 所示。

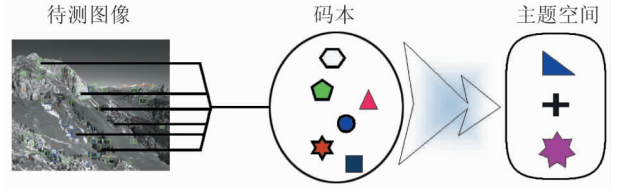


图 3 图像语义映射关系

Fig. 3 Map between image and semantic

与全局特征相比,选择局部图像特征更具鲁棒性,因为它对图像中所存在的遮盖和空间变化更具有不变性。本文选择 DoG (difference of Gaussians) 方法进行局部特征点检测,并用 SIFT (scale invariant feature transform) 进行特征表示。由于当前所用的图像数据集都是灰度图像,采用稀疏的灰度 SIFT 特征来生成码本。训练和测试流程如图 4 所示,图中用黑色虚线所标出的流程为训练流程,而用蓝色实线标出的则为测试流程。

模型训练过程中,码本是通过将训练集中各图像 SIFT 特征进行聚类而生成的,因此每个视觉词语就分别对应于图像中某一相似的局部块(也许多个),各个图像分别用词包来表示。此时的特征空间的维数由码本大小来决定,根据文献[6]对码本大小的分析结果,在实验中定为 174。对每类随机选取 100 幅图像则训练图像集的共生矩阵就是

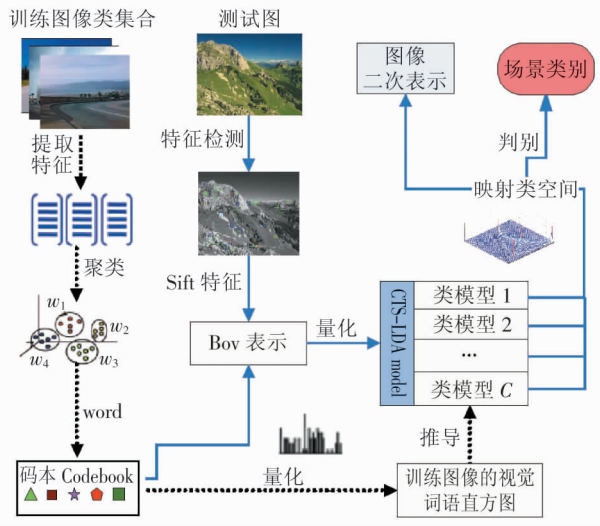


图 4 CTS-LDA 模型训练及测试流程

Fig. 4 The flow chart of CTS-LDA model

174 × 类别数 × 100, 若是 15 类则其共生矩阵就是 174 × 1 500 的高维空间, 因此必须借助降维来实现图像在主题空间中的二次表达, 进而实现分类。本文采用了 CTS-LDA 模型来实现这部分功能, 首先分别用各类场景图像集来训练, 得出各类模型参数  $\alpha$  和  $\beta$ , 对于本实验中的 15 个场景类可以得出 15 个类模型组成的 CTS-LDA 模型。在测试过程中, 将测试图像的 BoV 数据代入各类模型去计算, 根据式 (7) 来选择相应的类模型并完成场景类的判别。

### 3 实验结果与分析

本文数据集是来源于文献 [2] 和 [6], 其中文献 [6] 包括高速公路、市区住宅、高层建筑、街道、郊区住宅、森林、海岸、高山、户外、卧室、厨房、起居室和办公室共 13 类场景。Lazebnik 等人 [2] 在前 13 类图像的基础上增加了厂房和店铺两类场景共 15 类场景。为了验证本模型性能, 将分别与文献 [5] 和 [6] 进行性能比较, 并分别取其中 8 类、4 类进行实验和比较, 其中 8 类场景包括卧室、户外住宅、海岸、森林、高山和高层建筑; 而 4 类包括卧室、森林、高山和高速公路。每个场景类数据, 都有 200 至 400 幅图像, 在训练和测试时分别随机抽取产生训练集和测试集。

通过采用每类 100 幅图像进行训练, 200 幅测试, 本模型对 15 类场景分类的平均正确率是 57.93%。这种平均正确率的定义是借鉴于文献 [6]

中的 the confusion table 定义。即对各类的正确率进行平均而得出。图 5 是采用 CTS-LDA 模型下对 15 类场景进行分类得出的平均正确率, 其中图中右边的颜色条表示由 1 至 200 数量的颜色变化的标尺, 分别对应左边图中相应的颜色, 表示相应类别被正确分类的数量分布, 在对角线中的数值则表示该类图像被正确分类的比例 (即正确率)。理想状态下对角线所对应的颜色应是颜色条中最上面的颜色, 同理在非对角线位置的颜色都应是代表 0 的深蓝色。由图 5 中可看出森林场景得到最高的性能 (达到 181 幅, 正确率 90.5%), 而厨房场景的最低 (仅 54 幅, 正确率 27%), 这可能与森林中存在大量相同纹理的图像特征, 而厨房场景存在较多的特征变化有关。

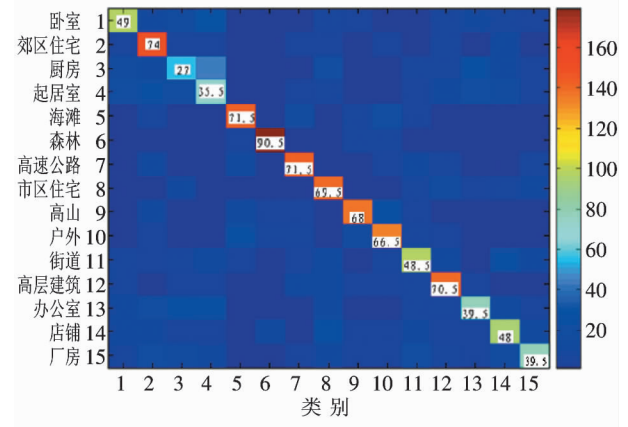


图 5 CTS-LDA 在 15 类场景分类中的性能

Fig. 5 The performance of CTS-LDA for 15 categories task

为进一步地分析实验结果, 图 6 分别针对高正确率和低正确率的场景类列出了其被正确分类和错误分类的例子。其中前 3 列为正确分类的图像及类别, 而在红线右边的最后一列为被错分的图像及其



图 6 15 类场景分类中部分分类结果示例

Fig. 6 Examples of CTS-LDA for 15 categories task

错分后的类别。

由于主题空间中的主题单元数与实际图像的表达密切相关,文献[7]对 LDA 模型的主题数设定为 40,但它是将全部图像作为一个集合来进行训练,必然需要考虑各类图像所存在的相异程度。对于同类图像而言,其用于表示图像的主题应相应减少,因为各图像间存在较多的相似(局部特征相同)。鉴于主题的具体数目只能与实际类别有关系,因此本文只能采用实验的方法寻找一个最佳且合理主题数,进而用各类模型来表示其分布。本文分别按不同的主题数来实验了 CTS-LDA 模型的性能,如图 7 所示。

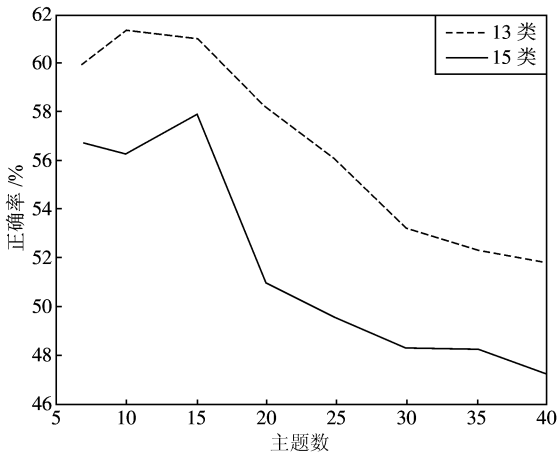


图 7 不同主题下的性能曲线

Fig. 7 The performance curves for changing topic size

根据图 7 结果,将各类主题数设为 10~15,取每类 200 幅图像的测试数据时可以取得较好的性能,其中对于 13 类任务取每类主题数为 10 可达到 61.38%,而 15 类分类任务取每类主题数为 15 可达到 57.93%。

表 1 将 CTS-LDA 模型分别与 pLSA、LDA 以及 Theme1 就 13 类场景分类任务进行平均正确率比较。其中 CTS-LDA 模型每类采用 10 个主题来表示图像的语义二次表达,LDA 和 pLSA 则分别按文献[6-7]中的思想采用 40 个主题不区分类别进行训练和测试。由表 1 可看出,CTS-LDA 模型性能均

表 1 13 类场景分类中各模型性能比较

Tab. 1 The performance comparison among models for 13 categories task

模型	CTS-LDA	Theme 1	LDA	pLSA
平均正确率/%	61.38	52.5	37	47.19

高于其他 3 类模型,由此可知在图像进行语义降维中引入类别约束是合理且有效的。

为了进一步检验 CTS-LDA 模型的性能,通过改变场景类别数将 CTS-LDA 与 Theme1<sup>[6]</sup>和 pLSA 的性能进行了比较,其结果如表 2 所示。对于 8 类和 4 类的比较是按文献[6]中的方法从 15 类数据集词包表示数据中抽取其中相应 8 类和 4 类,调用 CTS-LDA 时以主题数为 15 计算后的测试结果。由表 2 可看出 CTS-LDA 在分类性能上均高于 pLSA 和 Theme1 模型,且在类别较少的任务中性能尤其显著。此外文献[5]采用稠密 SIFT 特征表示图像,通过 pLSA 与 KNN 结合的方法对 4 类场景分类,其结果是 88.56%,也远远低于 CTS-LDA 的分类结果(92.63%)。

表 2 变化场景类别数任务中的性能比较

Tab. 2 The performance comparison with varied categories task

模型	测试图像	不同场景类别任务下的性能/%			
		15	13	8	4
CTS-LDA	100	44.00	49.92	65.88	86.50
	200	57.93	61.38	77.25	92.63
pLSA	100	41.73	51.15	59.00	72.00
	200	44.27	47.19	60.19	71.75
Theme1 <sup>[6]</sup>	> 100	—	52.5	—	76

## 4 结 论

提出了一种新的生成模型来学习和识别自然场景类别,该模型通过对各场景类别分别建立各自的主题空间,使得不同场景在不同的主题空间表示视觉词语分布,实现该图像所对应的主题空间中的二次表示,使图像的最终语义按人类的认知方式来表示。与文献[5-8]不同,CTS-LDA 能够在测试图像增加的情况下仍然能够取得较好的分类效果。此外本模型的一个最大特点在于不需借助其他分类器来实现判别,改变了以往必须结合生成模式和判别模式进行场景分类的做法。本文分别采用 15 类和 13 类图像数据集来验证模型的性能,并通过实验不同的主题数来验证本模型的图像表示方法以符合人类的认知思维方式。但本文对本模型判别图像类别所得出的图像二次表示(主题分布)暂时未作进一步的应用,因此,如何将图像的主题分

布与具体应用相结合(如目标识别任务)将成为今后的工作重点。

### 参考文献 (References)

- [ 1 ] Bosch A, Zisserman A. Scene classification using a hybrid generative/discriminative approach [ J ]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(4): 712-727.
- [ 2 ] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories [ C ]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE Press, 2006, 2: 2169-2178.
- [ 3 ] Gökalp D, Aksoy S. Scene classification using bag-of-regions representations [ C ]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, New York, USA: IEEE Press, 2007, 6: 1-8.
- [ 4 ] Nowak E. Sampling strategies for bag of features image classification [ C ]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer-Verlag, 2006, 3954: 490-503.
- [ 5 ] Anna B, Zisserman A. Scene classification via pLSA [ C ]//Proceedings of European Conference on Computer Vision. Berlin, Germany, Springer-Verlag, 2006, 3954: 517-530.
- [ 6 ] Li F F, Perona P. A bayesian hierarchical model for learning natural scene categories [ C ]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE Press, 2005, 2: 524-531.
- [ 7 ] Blei D, Andrew Y. Latent dirichlet allocation [ J ]. Journal of Machine Learning Research, 2003, 3(1): 993-1020.
- [ 8 ] Quelhas P. A thousand words in a scene [ J ]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(9): 1575-1589.
- [ 9 ] Li Wenbo, Sun Le. Text classification based on Label-LDA model [ J ]. Chinese Journal of Computers, 2009, 31(4): 620-627. [ 李文波, 孙乐. 基于 Labeled-LDA 模型的文本分类新算法 [ J ]. 计算机学报, 2008, 31(4): 620-627. ]
- [ 10 ] Elango P K. Clustering Images Using the Latent Dirichlet Allocation Model [ EB/OL ]. (2005-12-08) [ 2009-11-07 ]. <http://pages.cs.wisc.edu/~pradheep/Clust-LDA.pdf>.