doi:10.11887/j.cn.201406027 http://journal.nudt.edu.cn

结合非负张量表示与扩展隐 Dirichlet 分配模型的图像标注*

钱智明,钟 平,王润生
(国防科技大学 电子科学与工程学院,湖南 长沙 410073)

摘 要:由于"语义鸿沟"的存在,自动图像标注是一项极具挑战性的工作。考虑到图像低层视觉特征与高层语义概念的差异,分别从图像表示与语义建模两个方面来实现自动图像标注。在图像表示方面,提出了一种正则化约束下的非负张量表示方法,用以提取符合人眼视觉直观理解的图像高阶结构特征。在语义建模方面,提出了一种三层贝叶斯模型——扩展隐 Dirichlet 分配。该模型利用隐变量来实现图像与标注词的关联,并通过一种基于变分推理的期望最大值方法来估计参数。实验结果表明,ELDA 模型在大规模数据库 NUS-WIDE 上的标注结果相较于现有方法有了显著的提高。

关键词:图像标注;非负张量表示;扩展隐 Dirichlet 分配;变分推理
中图分类号:TP391 **文献标志码**:A **文章编号**:1001-2486(2014)06-152-06

Extended latent Dirichlet allocation for image annotation of nonnegative tensor representation

QIAN Zhiming, ZHONG Ping, WANG Runsheng
(College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: Automatic image annotation is a challenge task due to the well-known semantic gap. Considering the difference between low-level visual features and high-level semantic concepts, the framework of automatic image annotation from the two aspects, image representation and semantic modeling, was constructed. For image representation, a new method of regularized nonnegative tensor representation (RNTP) was presented to abstract the detailed high-order tensor structures according to human's intuitive recognition. A three-level hierarchical Bayesian model, extended latent Dirichlet allocation (ELDA), was developed for semantic modeling. In ELDA, each item of multiple image factors was modeled as a finite mixture over latent variables. Meanwhile, an efficient expectation-maximization algorithm based on variational inference was proposed for parameter estimation. Extensive experimental results are reported on the NUS-WIDE dataset to validate the effectiveness of our proposed solution to the automatic image annotation problem by comparing with other state-of-the-art methods.

Key words: image annotation; nonnegative tensor representation; extended latent Dirichlet allocation; variational inference

图像标注就是根据图像内容以添加标注词的形式对图像进行描述,以便于利用这些标注词来实现快速、准确的图像检索。随着与日俱增的图像资源,自动图像标注在计算机视觉领域得到了广泛应用,如大规模网络图像检索、图像共享社区的个人相册管理与分享等。然而,由于图像标注过程中图像数据视觉差异大、标注词汇量多以及高质量训练数据少等困难的存在,自动图像标注系统需处理好两个难题:一是选择合适的图像表示方法,以能够充分表达图像内容的丰富特征信息;二是要建立合理的标注模型,以实现接近人工标注结果的精确标注。

在图像表示方面,图像一般采用高维特征矢量来表示,如 PCA-SIFT[1],CPAM(Colored Pattern Appearance Model)[2]等。这类表示方法已经在图像分析与理解上起到了重要作用。由于图像标注侧重于描述图像内容,而这些内容通常包含丰富的高阶结构特征,所以图像矢量表示方法在表述复杂图像场景时有一定的局限性。幸而,在图像的张量表示中,高阶张量能够很好地表述图像数据中的结构性信息。因此,一些最新的图像处理方法已经开始关注图像的张量表示,如非相关多重线性主成分分析[3]、基于高阶张量分解的视频压缩表示方法[4]等。

根据图像标注过程中是否涉及图像数据分布的先验信息,图像标注模型可分为生成模型和判

* **收稿日期**:2014-03-31
基金项目:国家自然科学基金资助项目(61271439)
作者简介:钱智明(1986—),男,江苏南通人,博士研究生,E-mail:qianzhiming@nudt.edu.cn;
钟平(通信作者),男,副教授,博士,硕士生导师,E-mail:zhongping@nudt.edu.cn

别模型两类。生成模型方法结合图像数据分布的先验信息来估计图像与标注词的联合概率分布,并以此对图像进行语义标注;而判别模型则将标注词看作图像类别信息,并通过分类的方法获取图像的语义标注。一般而言,在图像训练数据充足的条件下,判别模型要优于生成模型;而在训练数据相对较少的情况下,判别模型往往很难推广到所有图像数据,因而需要一定的先验信息来构建更加合理的标注模型,这就使得生成模型在这一方面具有一定的优势。本文主要针对大规模图像数据进行标注,其训练数据往往相对较少,因而仅研究生成模型方法。典型的生成模型方法有混合模型方法和主题模型方法。混合模型方法根据图像与标注词的共现概率来构建联合分布函数,进而实现图像的自动标注。其主要方法有交叉媒体相关模型(Cross-Media Relevance Model, CMRM)[5]、多重 Bernoulli 相关模型(Multiple Bernoulli Relevance Model, MBRM)[6]等。主题模型方法则通过引入潜在主题分布来实现图像与标注词的关联,该模型是当前生成模型中应用最广泛的方法之一。2003 年,Monay 和 Gatica-Perez[7]将信息处理中应用较广泛的潜在语义概率分析模型(Probabilistic Latent Semantic Analysis, PLSA)率先引入到图像标注中。然后,Li 等[8]根据图像视觉特征的连续性提出了一种基于高斯多项式分布的 PLSA 模型(Gaussian-Multinomial PLSA, GM-PLSA)。此外, Nikolopoulos 等[9]根据图像数据结构的高阶特性提出了一种基于张量分解的高阶 PLSA 模型。然而,由于 PLSA 模型并没有考虑隐变量的先验信息而使得其模型参数与训练图像样本数成正比,不利于模型的推广。为此,Blei 和 Jordan[10]利用 Dirichlet 分布与多项式分布的共轭特性提出了一种三层贝叶斯模型——隐 Dirichlet 分配(Latent Dirichlet Allocation, LDA)。该模型由于引入多项式分布的先验信息,其模型的标注性能和推广性能有了较大的提高。

将图像的张量表示方法与语义标注模型相结合有着显著的优点,文献[9]对 PLSA 模型进行了相应的研究,但目前还没有研究工作对张量表示基础上的 LDA 模型进行扩展研究。为此,考虑到图像结构特征的稀疏性和非负性,本文采取了一种正则化约束下的非负张量表示方法来提取有利于标注的图像结构特征;同时,结合这种张量表示方法,又提出了一种扩展 LDA(Extended LDA, ELDA)模型来对图像进行语义建模。图 1 展示了本文方法的具体实现框架。

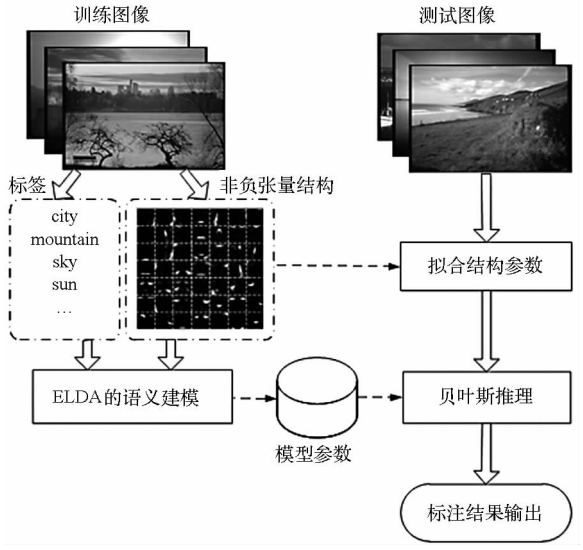


图 1　图像标注方法的实现框架

Fig. 1　Framework of image annotation method

1　图像表示

1.1　符号表示规范与定义

在本文所涉及的所有符号中,矢量由黑斜体小写字母来表示(如矢量 $\boldsymbol{x}$);矩阵由黑斜体大写字母表示(如矩阵 $\boldsymbol{X}$);张量由花体字母表示(如张量 $\mathcal{X}$)。它们的具体元素由小括号引出。张量 $\mathcal{A}\in\Re^{I_1\times\cdots\times I_N}$ 与矩阵 $\boldsymbol{T}\in\Re^{J_n\times I_n}$ 的积可表示为:

$$(\mathcal{A}\times_n\boldsymbol{U})(i_1,\cdots,i_{n-1},j_n,i_{n+1},\cdots,i_N)=\sum_{i_n}\mathcal{A}(i_1,\cdots,i_N)\cdot\boldsymbol{T}(j_n,i_n)\tag{1}$$

由此,可定义张量的 Tucker 分解[11]:

$$\mathcal{A}\approx\mathcal{R}=\mathcal{G}\times_1\boldsymbol{U}^{(1)}\times\cdots\times_N\boldsymbol{U}^{(N)}\tag{2}$$

其中,$\mathcal{G}\in\Re^{J_1\times\cdots\times J_N}$,$\boldsymbol{U}^{(n)}\in\Re^{I_n\times J_n}$。式(2)可通过矩阵化表示为:

$$\boldsymbol{A}_{(n)}\approx\boldsymbol{U}^{(n)}\boldsymbol{G}_{(n)}(\boldsymbol{U}^{(N)}\otimes\cdots\otimes\boldsymbol{U}^{(n+1)}\otimes\boldsymbol{U}^{(n-1)}\otimes\cdots\otimes\boldsymbol{U}^{(1)})^{\mathrm{T}}\tag{3}$$

其中,"⊗"表示 Kronecker 积,$\boldsymbol{A}_{(n)}$ 表示对张量 $\mathcal{A}$ 的矩阵化。

1.2　正则化约束下的非负张量表示

对于一幅二维灰度图像 I,其对应的张量表示形式为 $\mathcal{X}\in\Re^{I_1\times I_2\times I_3}$,其中 $\{I_1,I_2\}$ 表示图像大小,I_3 表示图像中每个像素点的特征维数。这里,引入高斯差分(Gaussian Difference, GD)和高斯偏微分(Difference of Offset Gaussian, DOOG)[12]:

$$\begin{cases}E_{\sigma,\theta}(\boldsymbol{I}(x,y))=|\boldsymbol{I}(x,y)*GD_{\sigma,\theta}(x,y)|\\F_{\sigma,\theta}(\boldsymbol{I}(x,y))=|\boldsymbol{I}(x,y)*DOOG_{\sigma,\theta}(x,y)|\end{cases}\tag{4}$$

其中,(x,y) 表示像素坐标,"$*$"表示 Hadamard 积。由此,图像张量可进一步表示为:

$$\mathcal{Y}=[\boldsymbol{Y}\boldsymbol{C}_b\boldsymbol{C}_r E_{\sigma,0}(\boldsymbol{Y})E_{\sigma,\pi/4}(\boldsymbol{Y})E_{\sigma,\pi/2}(\boldsymbol{Y})E_{\sigma,3\pi/4}(\boldsymbol{Y})F_{\sigma,0}(\boldsymbol{Y})F_{\sigma,\pi/4}(\boldsymbol{Y})F_{\sigma,\pi/2}(\boldsymbol{Y})F_{\sigma,3\pi/4}(\boldsymbol{Y})] \tag{5}$$

其中，$\boldsymbol{Y},\boldsymbol{C}_b,\boldsymbol{C}_r$ 为 YCbCr 颜色空间各通道值。这种表示方式通常会导致大量的信息冗余。因此，需要对所获得的图像张量数据进行降维。假设 $\{\mathcal{Y}_i \mid \mathcal{Y}_i \in \Re^{I_1\times I_2\times I_3}, i=1,\cdots,N\}$ 为图像样本数据集，降维问题可表示为：

$$\Psi(\boldsymbol{U}^{(1)},\boldsymbol{U}^{(2)},\boldsymbol{U}^{(3)}) = \min_{\boldsymbol{U}^{(1)},\boldsymbol{U}^{(2)},\boldsymbol{U}^{(3)}} \sum_i \|\mathcal{Y}_i - \mathcal{G}_i \times_1 \boldsymbol{U}^{(1)} \times_2 \boldsymbol{U}^{(2)} \times_3 \boldsymbol{U}^{(3)}\|_F^2$$

$$\text{s.t. } \boldsymbol{U}^{(n)} \geqslant 0,\ \sum_{i_n=1}^{I_n} \boldsymbol{U}^{(n)}(i_n;j_n)=1,\quad n=1,2,3;\ j_n=1,\cdots,J_n$$

$$\mathcal{G}_i = \mathcal{Y}_i \times_1 (\boldsymbol{U}^{(1)})^{\mathrm{T}} \times_2 (\boldsymbol{U}^{(2)})^{\mathrm{T}} \times_3 (\boldsymbol{U}^{(3)})^{\mathrm{T}} \tag{6}$$

其中，$\{\boldsymbol{U}^{(n)} \in \Re^{I_n\times J_n}, n=1,2,3,\ J_n < I_n\}$ 为非负投影矩阵，其张量积即为所求的非负张量结构。定义 $\boldsymbol{Z}_{i(n)} = \boldsymbol{G}_{i(n)}(\boldsymbol{U}^{(3)}\otimes\cdots\otimes\boldsymbol{U}^{(n+1)}\otimes\boldsymbol{U}^{(n-1)}\otimes\cdots\otimes\boldsymbol{U}^{(1)})^{\mathrm{T}}$，则上述降维问题可改写为：

$$\Psi(\boldsymbol{U}^{(1)},\boldsymbol{U}^{(2)},\boldsymbol{U}^{(3)}) = \min_{\boldsymbol{U}^{(1)},\boldsymbol{U}^{(2)},\boldsymbol{U}^{(3)}} \sum_i \|\boldsymbol{X}_{i(n)} - \boldsymbol{U}^{(n)}\boldsymbol{Z}_{i(n)}\|_F^2 \tag{7}$$

采用多元梯度下降法[13]，可得：

$$\boldsymbol{U}^{(n)} \leftarrow \boldsymbol{U}^{(n)} * \left(\frac{\sum_i \boldsymbol{X}_{i(n)}\boldsymbol{Z}_{i(n)}^{\mathrm{T}}}{\sum_i \boldsymbol{X}_{i(n)}\boldsymbol{Z}_{i(n)}\boldsymbol{Z}_{i(n)}^{\mathrm{T}}}\right) \tag{8}$$

通过对非负投影矩阵进行交替迭代就可以得到最终结果。在迭代过程中，$\{\boldsymbol{U}^{(n)}, n=1,2,3\}$ 需归一化。

为了更有效地对图像 $\mathcal{Y}_i$ 进行非负张量表示，引入 l_1 正则化来实现张量结构特征的稀疏表示，其目标函数为：

$$\Phi(\mathcal{G}) = \min_{\mathcal{G}} \|\mathcal{Y}_i - \mathcal{G}_i \times_1 \boldsymbol{U}^{(1)} \times_2 \boldsymbol{U}^{(2)} \times_3 \boldsymbol{U}^{(3)}\|_F^2 + \lambda \|vec(\mathcal{G}_i)\|_1 \tag{9}$$

其中 λ 控制 $\mathcal{G}_i$ 的稀疏度，$vec(\mathcal{G}_i)$ 表示张量 $\mathcal{G}_i$ 的矢量化。对于这个非平滑优化问题，可求得最终的解如式(10)所示：

$$\mathcal{G}_i(j_1,j_2,j_3) = \max\{\mathcal{Y}_i \times_1 [\boldsymbol{U}^{(1)}(:,j_1)]^{\mathrm{T}} \times_2 [\boldsymbol{U}^{(2)}(:,j_2)]^{\mathrm{T}} \times_3 [\boldsymbol{U}^{(3)}(:,j_3)]^{\mathrm{T}} - \lambda/2, 0\} \tag{10}$$

2 语义建模

2.1 ELDA 模型

如图 2 所示，本文所提的 ELDA 模型是一个三层贝叶斯模型，其主要流程如下：

1）选取一个 Dirichlet 随机变量 $\boldsymbol{\theta} \sim Dir(\boldsymbol{\alpha})$；

2）对于每一标注词 $w_m \in \{1,\cdots,V\}$：

a）选取一主题 $\boldsymbol{z}_m \in \mathcal{R}^K \sim Multinomial(\boldsymbol{\theta})$；

b）根据隐变量 $\boldsymbol{z}_m$ 上的多项式分布 $p(w_m \mid \boldsymbol{z}_m, \boldsymbol{\beta})$，选取标注词 w_m；

3）对于每一张量结构，选取主题 $\boldsymbol{h}_d \in \mathcal{R}^K \sim Multinomial(\boldsymbol{\theta})$；

4）根据 $\boldsymbol{h}_d$ 上的高斯分布 $p(\mathcal{Y} \mid \boldsymbol{h}_d, \boldsymbol{\mu}, \boldsymbol{\sigma})$ 对图像 $\mathcal{Y}$ 进行采样。

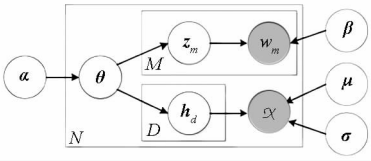


图 2 ELDA 模型示意图

Fig. 2 Illustration of ELDA

给定参数 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}$ 和 $\boldsymbol{\mu}$，$\boldsymbol{\sigma}$，根据贝叶斯规则，可得联合概率密度：

$$p(\boldsymbol{\theta},\boldsymbol{z},\boldsymbol{h},\boldsymbol{w},\mathcal{Y} \mid \boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\mu},\boldsymbol{\sigma}) = p(\boldsymbol{\theta} \mid \boldsymbol{\alpha})p(\boldsymbol{z} \mid \boldsymbol{\theta})p(w \mid \boldsymbol{z},\boldsymbol{\beta})p(\boldsymbol{h} \mid \boldsymbol{\theta})p(\mathcal{Y} \mid \boldsymbol{h},\boldsymbol{\mu},\boldsymbol{\sigma}) \tag{11}$$

对于该模型，做出以下假设。首先，标注词个数、张量特征维数和 Dirichlet 参数维数是确定的，依次为 T,D,K。其次，$\boldsymbol{\beta}(k,w_m) = p(w_m \mid \boldsymbol{z}_m(k)=1)$ 为 $K\times D$ 的待估计参数矩阵。根据以上假设，ELDA 模型可根据最大化边缘概率求得：

$$p(w,\mathcal{Y} \mid \boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\mu},\boldsymbol{\sigma}) = \frac{\Gamma[\sum_k \boldsymbol{\alpha}(k)]}{\prod_k \Gamma[\boldsymbol{\alpha}(k)]} \prod_{n=1}^{N} \int_{\boldsymbol{\theta}_n} [\prod_k \boldsymbol{\theta}_n(k)^{\boldsymbol{\alpha}(k)-1}] [\prod_{m=1}^{M} \sum_k \boldsymbol{\theta}_n(k)\boldsymbol{\beta}(k,w_{mn})] \left(\sum_k \boldsymbol{\theta}_n(k) \prod_{d=1}^{D} \frac{\exp\{-[\boldsymbol{\varepsilon}_{nd} - \boldsymbol{\mu}(k,d)]^2/2\boldsymbol{\sigma}^2(k,d)\}}{2\pi\boldsymbol{\sigma}(k,d)}\right) \mathrm{d}\boldsymbol{\theta}_n \tag{12}$$

该模型的参数结构包括：整体数据层（$\boldsymbol{\alpha}$，$\boldsymbol{\beta}$）、图像层（$\boldsymbol{\theta}_n$ 和 $\mathcal{Y}_n$）或标注层（$\boldsymbol{\beta}$，$\boldsymbol{z}_{nm}$ 和 $\boldsymbol{w}_{nm}$）、图像结构层（$\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{h}_d$）。

2.2 模型参数估计

为了求解式(12)所表示的优化问题，提出一种基于变分推理的 EM 算法。考虑隐变量上的变分分布：

$$q(\boldsymbol{\theta},z,\boldsymbol{h}\mid\boldsymbol{\eta},\boldsymbol{\varphi},\boldsymbol{\zeta})=\prod_{n=1}^{N}q(\boldsymbol{\theta}_n\mid\boldsymbol{\eta}_n)\Big(\prod_{m=1}^{M_n}q(z_{nm}\mid\boldsymbol{\varphi}_{nm})\Big)\Big(\prod_{d=1}^{D}q(\boldsymbol{h}_{nd}\mid\boldsymbol{\zeta}_{nd})\Big)\tag{13}$$

则变分参数$\{\boldsymbol{\eta},\boldsymbol{\varphi},\boldsymbol{\zeta}\}$可由实际分布与变分分布的 KL 差分进行最小值计算:

$$\{\boldsymbol{\eta}^*,\boldsymbol{\varphi}^*,\boldsymbol{\zeta}^*\}=\arg\min_{\{\boldsymbol{\eta},\boldsymbol{\varphi},\boldsymbol{\zeta}\}}D[q(\boldsymbol{\theta},z,\boldsymbol{h}\mid\boldsymbol{\eta},\boldsymbol{\varphi},\boldsymbol{\zeta})\,\|\,p(\boldsymbol{\theta},z,\boldsymbol{h}\mid w,\mathcal{Y},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\mu},\boldsymbol{\sigma})]\tag{14}$$

根据梯度下降法[14]对式(14)进行优化求解可得:

$$\begin{cases}\boldsymbol{\varphi}_{nm}(k)\propto\boldsymbol{\beta}(k,w_{nm})\exp\{\Psi[\boldsymbol{\eta}_n(k)]-\Psi[\sum_{j=1}^{K}\boldsymbol{\eta}_n(j)]\}\\ \boldsymbol{\zeta}_{nd}(k)\propto\exp\Big(-\dfrac{[\varepsilon_{nd}-\boldsymbol{\mu}(k,d)]^2}{2\boldsymbol{\sigma}^2(k,d)}+\Psi[\boldsymbol{\eta}_n(k)]-\Psi[\sum_{j=1}^{K}\boldsymbol{\eta}_n(j)]\Big)\\ \boldsymbol{\eta}_n(k)=\boldsymbol{\alpha}(k)+\sum_{m=1}^{M}\boldsymbol{\varphi}_{nm}(k)+\sum_{d=1}^{D}\boldsymbol{\zeta}_{nd}(k)\end{cases}\tag{15}$$

其中,ε_{nd}为图像 $\mathcal{Y}_n$ 在图像结构 $\mathcal{V}_d$ 上的投影变量,$\Psi(\cdot)$为 log$\Gamma(\cdot)$函数的一阶导数。

根据变分参数$\{\boldsymbol{\eta},\boldsymbol{\varphi},\boldsymbol{\zeta}\}$,可以通过优化式(16)对数似然函数来求解参数$\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\mu},\boldsymbol{\sigma}$:

$$L(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\mu},\boldsymbol{\sigma})=\ln p(w,\mathcal{Y}\mid\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\mu},\boldsymbol{\sigma})\tag{16}$$

其中,参数$\boldsymbol{\alpha}$可由牛顿迭代法来近似求解,其迭代过程如式(17)所示:

$$\boldsymbol{\alpha}^{(i+1)}(k)=\boldsymbol{\alpha}^{(i)}(k)-\frac{1}{(N\Psi'(\boldsymbol{\alpha}^{(i)}(k)))}\left(\frac{\partial L}{\partial\boldsymbol{\alpha}^{(i)}(k)}-\frac{\sum_{j=1}^{K}\dfrac{\partial L}{\partial\boldsymbol{\alpha}^{(i)}(j)}\Big/(N\Psi'(\boldsymbol{\alpha}^{(i)}(j)))}{1/\big(N\Psi'\big(\sum_{j=1}^{K}\boldsymbol{\alpha}^{(i)}(j)\big)\big)+\sum_{j=1}^{K}1/\big(N\Psi'(\boldsymbol{\alpha}^{(i)}(j))\big)}\right)\tag{17}$$

令$\partial L/\partial\boldsymbol{\beta}=0$,参数$\boldsymbol{\beta}$为:

$$\boldsymbol{\beta}(k,t)\propto\sum_{n=1}^{N}\sum_{m=1}^{M}\boldsymbol{\varphi}_{nm}(k)\delta(w_{nm},t)\tag{18}$$

同理可得:

$$\begin{cases}\boldsymbol{\mu}(k,d)=\sum_{n}\boldsymbol{\zeta}_{nd}(k)\varepsilon_{nd}/\sum_{n}\boldsymbol{\zeta}_{nd}(k)\\ \boldsymbol{\sigma}^2(k,d)=\sum_{n}\boldsymbol{\zeta}_{nd}(k)[\varepsilon_{nd}-\boldsymbol{\mu}(k,d)]^2/\sum_{n}\boldsymbol{\zeta}_{nd}(k)\end{cases}\tag{19}$$

综上,参数估计过程如下:

1)(E 步)针对每一幅训练图像,计算其变分参数$\{\boldsymbol{\eta},\boldsymbol{\varphi},\boldsymbol{\zeta}\}$;

2)(M 步)最大化对数似然函数 $L(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\mu},\boldsymbol{\sigma})$,求解模型参数$\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\mu},\boldsymbol{\sigma}$。

根据这些模型参数,可在给定图像的情况下求得标注词的条件概率为:

$$\begin{aligned}p(w\mid\mathcal{Y},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\mu},\boldsymbol{\sigma})&=\frac{p(w,\mathcal{Y}\mid\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\mu},\boldsymbol{\sigma})}{p(\mathcal{Y}\mid\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\mu},\boldsymbol{\sigma})}\\&=\frac{\int_{\theta}p(\boldsymbol{\theta}\mid\boldsymbol{\alpha})\Big[\sum_{t}p(z\mid\boldsymbol{\theta})p(w\mid z,\boldsymbol{\beta})\Big]\Big[\sum_{h}p(\boldsymbol{h}\mid\boldsymbol{\theta})p(\mathcal{Y}\mid\boldsymbol{h},\boldsymbol{\mu},\boldsymbol{\sigma})\Big]\mathrm{d}\boldsymbol{\theta}}{\int_{\theta}p(\boldsymbol{\theta}\mid\boldsymbol{\alpha})\Big[\sum_{h}p(\boldsymbol{h}\mid\boldsymbol{\theta})p(\mathcal{Y}\mid\boldsymbol{h},\boldsymbol{\mu},\boldsymbol{\sigma})\Big]\mathrm{d}\boldsymbol{\theta}}\end{aligned}\tag{20}$$

2.3　模型复杂度分析

在上述 EM 算法中,E 步中变分参数的计算是根据式(15)交替迭代而得,其迭代次数一般与标注词和张量结构的数目线性相关。故 E 步的计算复杂度为 $O(N(\bar{M}+D)^2K)$,其中 $\bar{M}$ 为平均每幅图像的标注词个数。在 M 步中,求解 $\boldsymbol{\alpha}$ 的计算复杂度为$O(NK^2)$,而估计$\boldsymbol{\beta},\boldsymbol{\mu},\boldsymbol{\sigma}$的计算复杂度为 $O(NK(T+D))$,故 M 步的总体计算复杂度为 $O(NK(T+D))$。综上所述,EM 算法的计算复杂度为 $O(NK(K+T+N(\bar{M}+D)))$。

3　实验

3.1　实验数据

为了测试本文方法的有效性,选取 NUS-WIDE 图像标注库[15]作为实验数据。该数据库来源于 Flickr 网站,包含了 269 648 张图像和 425 059个标签。其中标注过 100 次以上的标签有 9 325 个,通过剔除一些不规范的标签,共剩 5 018个标签。

3.2　实验设置与评价准则

对于每一幅图像 $\mathcal{Y}_n$,将其大小固定为 256×256,并取前 1024 维张量结构系数$\{\varepsilon_{nd},d=1,\cdots,D\}$所构成的矢量作为图像特征。在实验中,随机选择 NUS-WIDE 数据库中的 10 000 张图像用于训练,并将其余图像用于测试。同时,设定稀疏参数 $\lambda=0.1$,Dirichlet 参数维数 $K=500$。此外,实验中所涉及的比较方法有:GM-PLSA[8]模型,

GM-LDA[10] 模型以及具有较好标注性能的 TagProp[1-6] 模型。

在实验中,选取前 5 个标注词对图像进行标注,并通过计算每个标注词的查全率和查准率[2]来评估标注结果。

3.3 图像标注结果比较

表 1 自动图像标注结果比较

Tab. 1 Comparison of the annotation results for different annotation methods

标注方法	平均查准率	平均查全率
GM – PLSA	0.073	0.046
GM – LDA	0.112	0.091
TagProp	0.132	0.089
ELDA	0.151	0.133

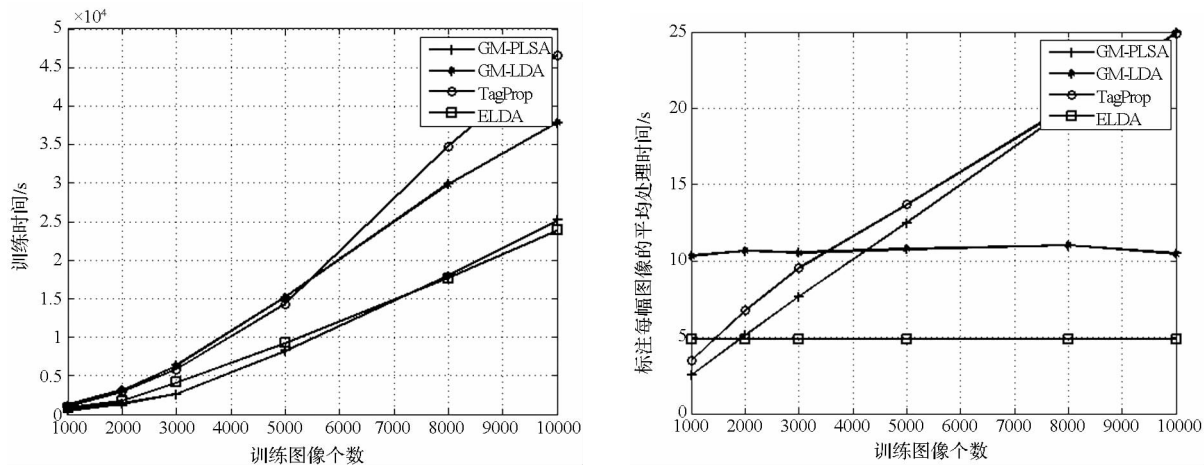
表 1 给出了本文方法和几种常用标注方法的结果比较。结果显示,本文所提方法在平均查准率和平均查全率的评价指标上均有了较为显著的提高。

图 3 给出了 NUS-WIDE 数据库中几幅示例图像的标注结果。可以看出,GM-PLSA 模型由于建模能力不足而更容易出现高频词汇(如“explore”);TagProp 模型由于缺失先验分布信息而导致生成一些视觉差异性较大的标注词(如第三图像中的“camel”);GM-LDA 模型与 ELDA 模型比较接近但所用表示方法不同,GM – LDA 模型所生成的 25 个标注词中有 9 个标注词是正确的,而 ELDA 模型则有 13 个是正确的,这进一步说明用非负张量表示有利于提高自动图像标注的准确性。

图像					
人工基准	bird nature sky wildlife avian	mountain castle Italy explore sky	road tree desert land sky	mountain view moisture tree	people India girl face pretty
GM – PLSA	explore plant leaf sky flower	explore sky sea water tree	explore sky sea tree mountain	explore plant tree sky nature	explore people face life film
GM-LDA	sky sea water animal bravo	mountain sky sea water tree	land sky tree water sea	mountain tree sky nature sea	black love negro face eyes
TagProp	tree sky flower sea explore	building sky sea water tree	desert night sky camel view	mountain land night dark tree	people face animal girl eyes
ELDA	bird animal avian sky bravo	building sky sea tree mountain	desert night sky water tree	mountain tree nature sky view	people face negro eyes girl

图 3 图像标注结果示例

Fig. 3 Samples of image annotation results



(a) (b)

图 4 各模型计算复杂度比较

Fig. 4 Comparison of the computation complexities for different annotation methods

3.4　模型计算复杂度比较

本节分别从模型训练和图像标注两个部分对本文方法进行计算复杂度分析与比较。

由图 4(a)可知,ELDA 模型的训练时间是渐近线性的,这与 2.3 小节的计算复杂度分析是一致的。当训练样本数达到 10 000 时,ELDA 模型所用的训练时间最少。由图 4(b)可知,GM-PLSA 与 TagProp 的平均标注时间随着训练样本的增加而增加,而 GM-LDA 与 ELDA 的平均标注时间则相对稳定。

4　结论

结合非负张量表示和扩展 LDA 模型对自动图像标注方法展开研究。实验结果表明了本文所提方法在标注大规模图像数据上具有良好的标注性能。然而,本文方法因模型所涉及的标注词汇量较大,其平均查准率和平均查全率都相对较低。在今后的研究工作中,将通过层次化分类方法来细化标注词汇,以求进一步提高自动图像标注的查全率和查准率。

参考文献(References)

[1] Ke Y, Sukthankar R. PCA-SIFT: a more distinctive representation for local image descriptors[C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004, 2: 506 – 513.

[2] Zhou N, Cheung W K, Qiu G P, et al. A hybrid probabilistic model for unified collaborative and content-based image tagging[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(7):1281 – 1294.

[3] Lu H P, Plataniotis K N, Venetsanopoulos A N. Uncorrelated multilinear principal component analysis for unsupervised multilinear subspace learning [J]. IEEE Transactions on Neural Networks, 2009, 20(11): 1820 – 1836.

[4] Zhou B Y, Zhang F, Peng L Z. Compact representation for dynamic texture video coding using tensor method[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2013, 23(2): 280 – 288.

[5] Jeon J, Lavrenko V, Manmatha R. Automatic image annotation and retrieval using cross-media relevance models [C]// Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval, 2003: 119 – 126.

[6] Feng S L, Manmatha R, Lavrenko V. Multiple Bernoulli relevance models for image and video annotation [C]// Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004, 2: 1002 – 1009.

[7] Monay F, Gatica-Perez D. On image auto-annotation with latent space models[C]//Proceedings of ACM International Conference on Multimedia, 2003:275 – 278.

[8] Li Z X, Shi Z P, Liu X, et al. Modeling continuous visual features for semantic image annotation and retrieval [J]. Pattern Recognition Letters, 2011, 32(3):516 – 523.

[9] Nikolopoulos S, Zafeiriou S, Patras I, et al. High order pLSA for indexing tagged images[J]. Signal Processing, 2013, 93(8): 2212 – 2228.

[10] Blei D M, Jordan M I. Modeling annotated data [C]// Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval, 2003:127 – 134.

[11] Lu H P, Plataniotis K N, Venetsanopoulos A N. A survey of multilinear subspace learning for tensor data [J]. Pattern Recognition, 2011, 44(7): 1540 – 1551.

[12] Ma W Y, Manjunath B S. EdgeFlow: A technique for boundary detection and image segmentation [J]. IEEE Transactions on Image Processing, 2000, 9(8): 1375 – 1388.

[13] Févotte C, Bertin N, Durrieu J L. Nonnegative matrix factorization with the Itakura-Saito divergence: with application to music analysis[J]. Neural Computation, 2009, 21(3): 793 – 830.

[14] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet allocation[J]. Journal of Machine Learning Research, 2003, 3:993 – 1022.

[15] Chua T S, Tang J H, Hong R C, et al. US-WIDE: A real-world web image database from National University of Singapore[C]//Proceedings of ACM International Conference on Image Video Retrieval, 2009:48 – 56.

[16] Guillaumin M, Mensink T, Verbeek J, et al. TagProp: discriminative metric learning in nearest neighbor models for image auto-annotation [C]//Proceedings of International Conference on Computer Vision, 2009: 309 – 316.