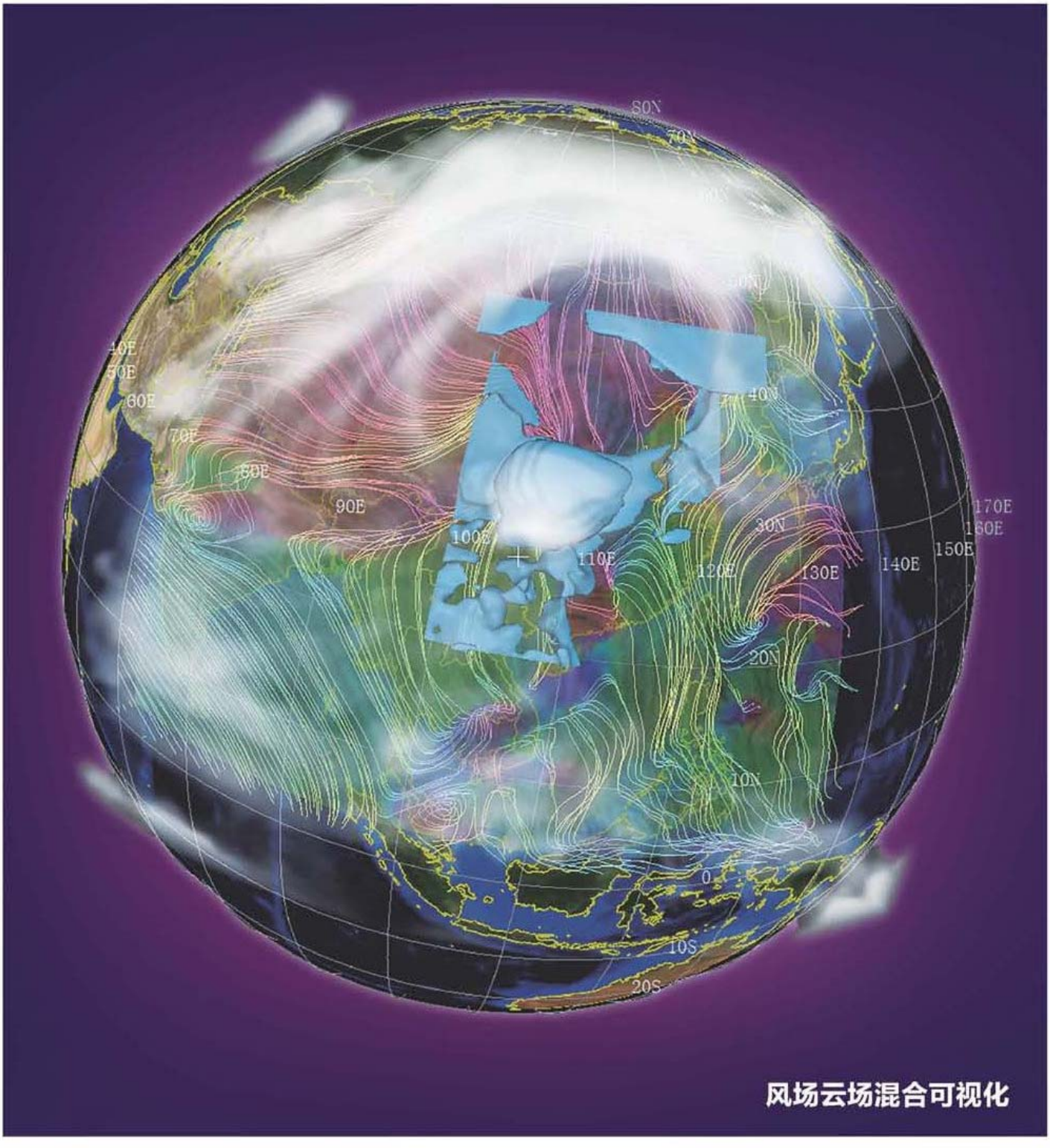
JIG
JOURNAL OF IMAGE AND GRAPHICS
主办：中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2015
05
VOL.20
ISSN1006-8961
CN11-3758/TB

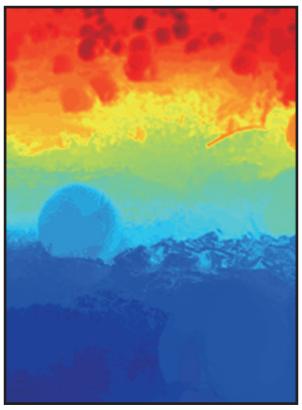


风场云场混合可视化

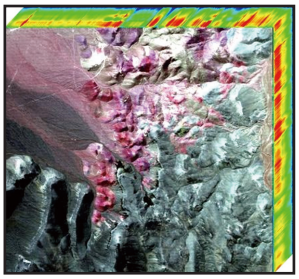
2015年5月 第20卷 第5期（总第229期）
中国图象图形学报
Zhongguo Tuxiang Tuxing Xuebao

目次

面向彩色增强图像的客观质量评价算法(第0643页)



使用柯西分布点扩散函数模型的单幅散焦图像深度恢复(第0708页)



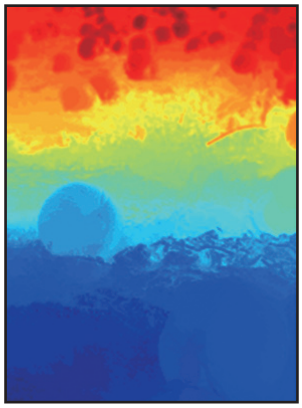
基于蛙跳算法的离散粒子群优化端元提取(第0724页)

封面图片由浙江大学CAD&CG国家重点实验室王章野提供

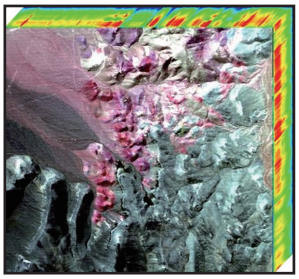
Volume 20, Number 5
Published May 16, 2015

CONTENTS
JOURNAL OF IMAGE AND GRAPHICS

Objective quality assessment for enhanced chromatic images(P0643)



Depth recovery from a single defocused image using a Cauchy-distribution-based point spread function model (P0708)



Discrete particle swarm optimization endmember extraction based on shuffled frog leaping algorithm (P0724)

©中国图象图形学报版权所有

中图法分类号：TP391.4　　**文献标识码**：A　　**文章编号**：1006-8961(2015)05-0687-13

论文引用格式：Li Z X,Shi Z P,Zhang C L,Wang J Y. Hybrid generative/discriminative model for automatic image annotation[J]. Journal of Image and Graphics,2015,20(5):0687-0699.[李志欣,施智平,张灿龙,王金艳. 混合生成式和判别式模型的图像自动标注[J]. 中国图象图形学报,2015,20(5):0687-0699.][DOI:10.11834/jig.20150511]

混合生成式和判别式模型的图像自动标注

李志欣[1]，施智平[2]，张灿龙[1]，王金艳[1]

1. 广西师范大学计算机科学与信息工程学院，桂林　541004；2. 首都师范大学信息工程学院，北京　100048

摘　要：目的　由于图像检索中存在着低层特征和高层语义之间的“语义鸿沟”，图像自动标注成为当前的关键性问题。为缩减语义鸿沟，提出了一种混合生成式和判别式模型的图像自动标注方法。**方法**　在生成式学习阶段，采用连续的概率潜在语义分析模型对图像进行建模，可得到相应的模型参数和每幅图像的主题分布。将这个主题分布作为每幅图像的中间表示向量，那么图像自动标注的问题就转化为一个基于多标记学习的分类问题。在判别式学习阶段，使用构造集群分类器链的方法对图像的中间表示向量进行学习，在建立分类器链的同时也集成了标注关键词之间的上下文信息，因而能够取得更高的标注精度和更好的检索效果。**结果**　在两个基准数据集上进行的实验表明，本文方法在Corel5k数据集上的平均精度、平均召回率分别达到0.28和0.32，在IAPR-TC12数据集上则达到0.29和0.18，其性能优于大多数当前先进的图像自动标注方法。此外，从精度—召回率曲线上看，本文方法也优于几种典型的具有代表性的标注方法。**结论**　提出了一种基于混合学习策略的图像自动标注方法，集成了生成式模型和判别式模型各自的优点，并在图像语义检索的任务中表现出良好的有效性和鲁棒性。本文方法和技术不仅能应用于图像检索和识别的领域，经过适当的改进之后也能在跨媒体检索和数据挖掘领域发挥重要作用。

关键词：图像自动标注；概率潜在语义分析；多标记学习；分类器链；图像检索

Hybrid generative/discriminative model for automatic image annotation

Li Zhixin[1], Shi Zhiping[2], Zhang Canlong[1], Wang Jinyan[1]

1. *College of Computer Science and Information Technology, Guangxi Normal University, Guilin*　541004, *China*;
2. *College of Information Engineering, Capital Normal University, Beijing*　100048, *China*

Abstract: Objective Given the notorious semantic gap between low level features and high level concepts in image retrieval, automatic image annotation has become a crucial issue. To bridge the semantic gap, this paper proposes a hybrid generative/discriminative approach to annotate images automatically. **Method** In the generative learning stage, images are modeled by continuous probabilistic latent semantic analysis model. As a result, we can obtain the corresponding model parameters and the topic distribution of each image. If this topic distribution is taken as an intermediate representation of each image, the image auto-annotation problem could be transformed into a multi-label classification problem. In the discriminative learning stage, we construct ensembles of classifier chains by learning these intermediate representations. At the same time, the contextual information of the annotation words can be integrated into the classifier chains. Therefore, this approach could achieve higher annotation accuracy and better retrieval performance. **Result** Experiments on two baseline data-

收稿日期：2014-09-24；**修回日期**：2015-01-14

基金项目：国家自然科学基金项目(61165009，61262005，61363035，61365009)；国家重点基础研究发展计划(973)项目(2012CB326403)；广西自然科学基金项目(2012GXNSFAA053219，2013GXNSFAA019345，2013GXNSFBA019263，2014GXNSFAA118368)

第一作者简介：李志欣(1971—　)，男，副教授，2010年于中国科学院计算技术研究所获计算机软件与理论专业博士学位，主要研究方向为图像理解、机器学习、多媒体分析与检索。E-mail：lizx@gxnu.edu.cn

Supported by：National Natural Science Foundation of China (61165009，61262005，61363035，61365009)；National Basic Research Program of China (2012CB326403)

© 中国图象图形学报版权所有

sets indicate that the average precision and recall of our approach attained 0.28 and 0.32, respectively, on Corel5k dataset. In addition, these two measures of our approach attained 0.29 and 0.18, respectively, on IAPR-TC12 dataset. The experimental results proved that our approach performed better than most state-of-the-art approaches on many evaluation measures. Furthermore, the precision-recall curve showed the superior performance of our approach over several typical and representative approaches. **Conclusion** On the basis of hybrid learning strategy, this paper presents an image auto-annotation approach, which integrates the advantages of the generative and discriminative models. As a result, the approach exhibits better, more effective, and more robust semantic image retrieval. The methods and techniques of this paper are not only usable in the fields of image retrieval and recognition, but they can play an important role in the fields of cross-media retrieval and data mining after an appropriate adaption.

Key words: automatic image annotation; probabilistic latent semantic analysis; multi-label learning; classifier chain; image retrieval

0 引 言

由于网络和数字技术的快速发展,前所未有的大规模视觉数据得以有效地创建和存储。目前,视觉数据已经和文本数据同样普遍地存在于日常生活之中。过去二十几年,研究人员做了大量关于图像检索的研究,大致可以分为两类不同的检索技术:基于文本的图像检索(TBIR)和基于内容的图像检索(CBIR)[1]。但是,TBIR 的实现需要大量的人工标注,难于推广应用到目前规模越来越大的数据库;而 CBIR 中普遍存在"语义鸿沟"[2]的问题,也就是说视觉特征相似的图像在语义上并不一定一致,这使得 CBIR 的性能受到限制。因此,图像自动标注(AIA)自然成为图像检索领域的重要课题[3-4]。AIA 技术的基本思想是从大量的图像样本中学习语义概念模型,并使用这个模型对未知图像进行标注。一旦为图像标注了相应的语义标签,就能像 TBIR 一样利用关键词进行检索。AIA 的重要特征就是提供了基于图像内容的关键词查询,从而集成了 TBIR 和 CBIR 的优势。

通过合理地设计学习框架,提出一个混合生成式和判别式模型的图像自动标注方法,有效地结合了生成式和判别式模型的学习方法并继承了各自的优势,利用多标记学习技术对图像进行分类,与此同时能够集成图像标注关键词之间的关联。首先采用连续的概率潜在语义分析(PLSA)模型[5-6]进行生成式学习,获取每幅图像的主题分布,然后采用集群分类器链[7]的方法对每幅图像的主题分布进行判别式学习,将置信度最高的若干语义关键词作为图像的语义标注。通过在一个包含 5 000 幅图像的标准 Corel 图像数据集和一个包含约 20 000 幅图像的 IAPR TC12 数据集上进行的实验,采用多种通用的评价指标对本文方法进行了评估。实践证明,本文方法既具备生成式模型能够充分利用训练数据的优点,也能像判别式模型一样能够获取较高的分类精度,其标注和检索性能优于大多数当前先进的图像标注方法。

1 相关工作

根据机器学习方法的特点,现有的图像自动标注方法大致可分为基于生成式模型的标注方法和基于判别式模型的标注方法。

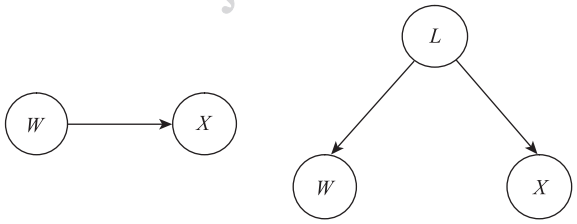
基于生成式模型的标注方法的特点是:先学习图像特征和关键词的联合概率,然后通过贝叶斯公式计算给定图像特征时各个关键词的后验概率,并依据后验概率进行图像标注。这类方法具有可扩展的训练过程,对训练图像集人工标注的质量要求较低。翻译模型[8]是一种代表性的方法,它通过学习联合概率将关键词与图像的区域联系起来,于是标注过程转化一个将区域翻译为关键词的过程。Barnard 等人[9]讨论了几种用于表示量化区域和关键词联合分布的模型,一旦学习获得联合分布,标注问题就转化为一个关联量化区域和关键词的似然问题。类似地,Blei 等人[10]使用关联 LDA(latent Dirichlet allocation)模型在关键词和图像间建立基于语言的关联。此外,Jeon 等人[11]提出跨媒体相关模型(CMRM)标注图像,假设在给定一幅图像的情况下,量化区域和关键词是相互独立的。Lavrenko 等人[12]提出类似的连续空间相关模型(CRM),使用多项分布估计关键词的概率分布,使用无参核密度分布估计区域的特征概率分布。Feng 等人[13]提出多贝努里相关模型(MBRM),使用多贝努里分布

© 中国图象图形学报版权所有

取代 CRM 中的多项分布以生成关键词。随后，Zhang 等人[14]提出一个概率语义框架进行图像的语义标注和多模态检索，该框架通过一个包含语义概念的隐藏层来关联视觉特征和文本关键词，并通过贝叶斯框架进行推断以生成这个关联的置信度。Monay 等人[15]提出 PLSA-WORDS 模型对多模态共现数据进行建模，限制潜在空间的定义使其在语义关键词上保持一致性，同时保持对视觉特征联合建模的能力。在此基础上，文献[16]提出融合文本和视觉两种模态语义主题的标注方法 PLSA-FUSION，在利用 PLSA 建模图像特征的基础上，采用自适应的不对称学习算法学习一个较优的潜在空间，能够更有效地利用训练数据。此后文献[6]又提出建模连续视觉特征的图像自动标注方法 GM-PLSA，使用连续 PLSA 直接建模连续视觉特征，消除了聚类粒度对标注性能的影响。Liu 等人[17]提出基于图学习模型(GLM)进行图像标注。首先，执行图学习算法获取每幅图像的备选标注。为捕获图像数据的复杂分布，提出 NSC(nearest spanning chain)算法来构造以图像为结点的图，该图的边的权值从链的统计信息中获取，而不是从传统的对偶相似度中获取。然后，使用基于词的图学习算法来精炼图像和词之间的关联，并预测每幅图像的最终标注。

基于判别式模型的标注方法的特点是：假设图像特征到关键词之间的映射是某种参数化的函数，直接在训练数据上学习此函数的参数，并获得各个语义概念的分类器。这类方法将各个语义概念视为独立的类别，一般来说能取得较高的标注精度，但是不便于利用领域相关的先验知识。Li 等人[18]提出的图像自动语义索引系统使用 2 维多分辨率隐马尔可夫模型捕获不同语义类别视觉特征的空间依赖关系，基于各个语义类别的相似度进行分类。Chang 等人[19]提出一个基于内容的软标注系统，基于 SVM(support vector machine)和 BPM(Bayes point machine)的学习方法建立多类分类器，从而为图像标注语义相近的类标签。Carneiro 等人[20]提出监督多类标注(SML)系统，使用最小错误率的优化准则，将标注任务转化为一个多类分类的问题，并将每一个感兴趣的语义概念都定义为一个图像类，在使用竞争标注机制推导图像包含语义概念的同时，能够产生语义标注的自然排序。Wang 等人[21]提出了一个多标记稀疏编码(MSC)的框架来进行特征提取和分类，并把它应用到图像自动标注中，认为具有部分重叠标记的两张图像之间的语义相似度应该以一种重构的方式而不是一对一的方式来度量。因此，在这个框架中，图像标记向量之间的语义相似度，以及图像特征向量之间的语义相似度，都基于一对多的 l_1 稀疏重构/编码来度量。Makadia 等人[22]引入一种新且简单的基准技术，提出 JEC(joint equal contribution)的方法进行图像标注，将标注问题视为检索问题。JEC 利用全局低层图像特征和基本距离度量的简单结合寻找给定图像的最近邻。然后使用一种贪心的标签传递机制将关键词赋予对应的图像，取得了很好的标注精度和检索性能。

基于生成式模型的方法和基于判别式模型的方法的概率图模型分别如图 1 所示，二者相比较主要有以下几点不同：1)基于判别式模型的方法将图像看做训练数据，各个语义概念看做类别，目的在于将图像分类到各个语义类别中，而基于生成式模型的方法将图像和文本都视为训练数据，其目的是学习图像与文本之间的关联；2)基于判别式模型的方法为每个语义概念训练一个分类器，基于生成式模型的方法只学习一个关联模型并将该模型应用于所有的语义概念；3)独立性假设不同，基于判别式模型的方法假设各个语义类别之间是相互独立的，而基于生成式模型的方法假设在给定隐藏变量的条件下，视觉元素和文本元素是条件独立的。



(a) 基于判别式模型的方法　(b) 基于生成式模型的方法

图 1　两类学习方法的图模型表示

Fig. 1　Graphical model representations of two kinds of learning approaches

在此之前的研究工作中，PLSA-FUSION 方法[16]采用两个 PLSA 模型分别从视觉和文本模态中捕获信息，而 GM-PLSA 方法[6]通过直接建模连续视觉特征提高学习精度。然而，这两种方法使用 PLSA 建模关键词时，每幅图像都对应一个稀疏的直方图，因为只有 4～5 个关键词与一幅图像相关联。因此

© 中国图象图形学报版权所有

PLSA-FUSION 和 GM-PLSA 都采用不对称的学习方法来学习视觉和文本模态之间的关联。虽然不对称学习方法能够获得不错的效果和精度，但从理论上看采用多标记分类的算法学习语义概念应该是一个解决文本词稀疏分布问题的更合理的方法。实验部分可以证明，这种方法在实践上也确实能够获得更高的精度和更好的效果。

2 基于混合学习的图像自动标注框架

在连续 PLSA 和集群分类器链的基础上，本文提出混合生成式和判别式模型的学习方法 HGDM（hybrid generative/discriminative model）。首先采用连续 PLSA 进行生成式学习，能够充分利用训练集的先验知识，并得到每幅图像的主题分布。然后，将这个分布作为每幅图像的中间表示向量，可以将图像标注的问题转化为一个多标记分类的问题，以获取比生成式模型更高的标注精度。HGDM 采用集群分类器链的方法对图像的中间表示向量进行判别式学习，在对图像进行分类的同时也考虑了图像标注之间的关联。在标注阶段，给定一幅未知图像，通过自动提取视觉特征和连续 PLSA 的参数估计算法可获得其主题向量的表示；再使用训练好的集群分类器链对这个主题向量进行分类；最后，将置信度最高的若干语义关键词作为图像的语义标注。其学习和标注的框架如图 2 所示。

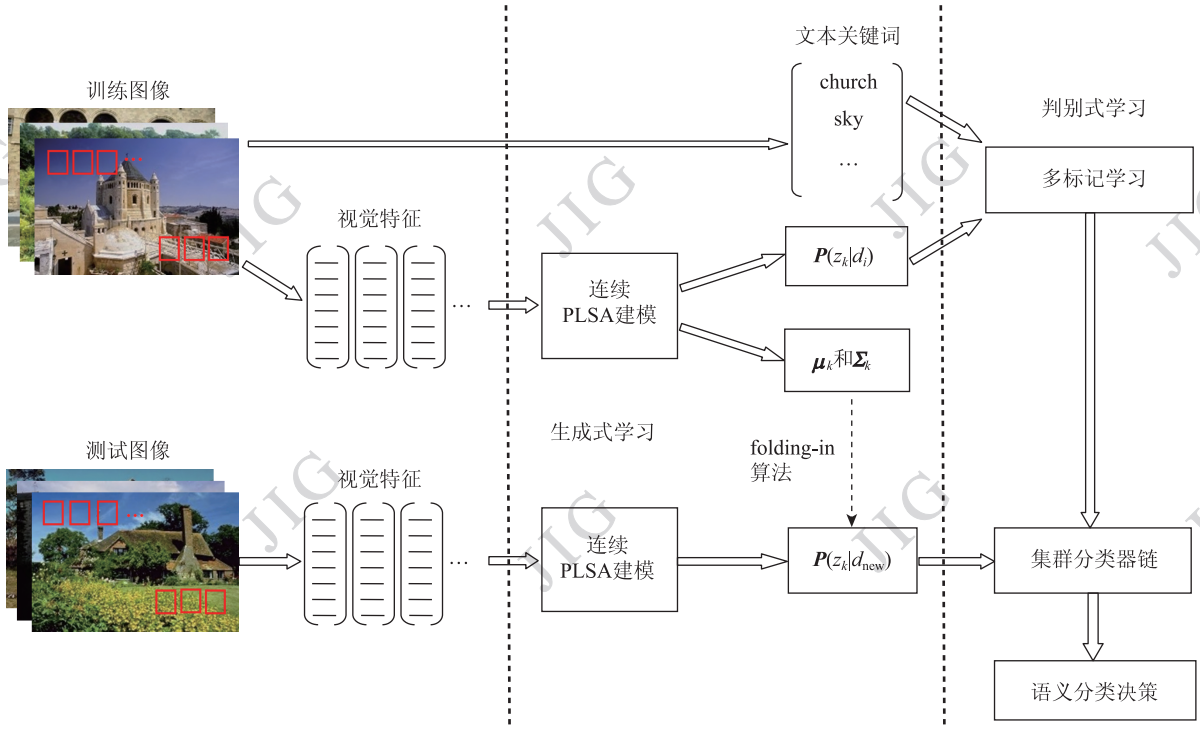


图 2 混合生成式与判别式模型的图像自动标注框架

Fig. 2 Automatic image annotation framework with hybrid generative/discriminative model

由图 2 可见，训练过程分为两步：

1）利用连续 PLSA 建模训练图像的视觉特征，得到给定潜在主题 z_k 下的高斯分布参数 $\boldsymbol{\mu}_k$ 和 $\boldsymbol{\Sigma}_k$，以及给定图像文档 d_i 的主题分布表示 $\boldsymbol{P}(z_k \mid d_i)$，这是一个生成式的学习过程。这里得到的高斯分布参数 $\boldsymbol{\mu}_k$ 和 $\boldsymbol{\Sigma}_k$ 是连续 PLSA 的参数，由连续 PLSA 的独立性假设这些参数对于训练集之外的图像仍然有效，而主题分布表示 $\boldsymbol{P}(z_k \mid d_i)$ 只对应于每幅训练图像本身的性质，不能给测试图像带来任何先验信息。但是，可以利用这个表示将每幅训练图像表示为一个 K 维的主题向量（K 是潜在主题个数），这些向量构成的空间是一个单形（simplex）。

2）利用每幅训练图像的主题向量表示以及它们的原始标注构造分类器，每个类对应于文本词汇表中的语义类别，这是一个判别式的学习过程。因为此时每幅图像都由一个主题向量表示，但却对应于多个关键词标注，与多标记学习的情形一致，故而采用了多标记学习的方法构造多类分类器，同时也

© 中国图象图形学报版权所有

集成了关键词之间的关联信息。

对测试图像的标注过程也分为类似的两步：

1）利用训练阶段得到的模型参数 $\boldsymbol{\mu}_k$ 和 $\boldsymbol{\Sigma}_k$ 以及测试图像的视觉特征，使用 folding-in 算法计算每幅测试图像文档 d_{new} 的主题向量 $\boldsymbol{P}(z_k \mid d_{\text{new}})$。

2）利用训练得到的分类器对该主题向量分类，并将所得的若干置信度最高的语义类别作为该测试图像的语义标注。

HGDM 在生成式学习过程采用连续 PLSA 建模，在判别式学习过程中采用集群分类器链的方法进行多标记分类，具有可扩展的训练过程并能够考虑标记之间的相互关联，从而获得了良好的性能。

2.1　连续 PLSA

由于传统的 PLSA[23] 最初应用于文本分类，因而只能够处理离散量。所以利用传统 PLSA 处理图像需要对图像的视觉特征进行聚类，导致其性能对聚类的粒度比较敏感。HGDM 采用连续 PLSA[5-6] 直接建模图像的连续视觉特征，不会在聚类的过程中丢失重要信息，从而大幅度提高系统性能。

正如传统 PLSA，连续 PLSA 也是一个统计的潜在类模型，它在文档 $d_i(i=1, \cdots, N)$ 的各个元素 x_j $(j=1, \cdots, M)$ 的生成过程中引入隐含变量（潜在主题）$z_k(k=1, \cdots, K)$。但是，给定不可观察的变量 z_k 时，连续 PLSA 假设元素 $\boldsymbol{x}_j$ 服从多元高斯分布采样，而不是像传统 PLSA 那样服从多项分布采样。根据这个定义，连续 PLSA 有下列生成过程[5-6]：

1）以概率 $P(d_i)$ 选择一个文档 d_i；

2）在给定文档 d_i 的条件下，以概率 $P(\boldsymbol{z}_k \mid d_i)$ 采样服从多项分布的一个潜在主题 $\boldsymbol{z}_k$；

3）在给定潜在主题 $\boldsymbol{z}_k$ 的条件下，以概率 $P(\boldsymbol{x}_j \mid \boldsymbol{z}_k)$ 采样服从多元高斯分布 $N(\boldsymbol{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ 的元素 $\boldsymbol{x}_j$。

连续 PLSA 有两个基本假设：

1）观察对 $(d_i, \boldsymbol{x}_j)$ 是独立生成；

2）随机变量对 $(d_i, \boldsymbol{x}_j)$ 在给定潜在主题 z_k 时是条件独立的。所以，通过对潜在主题 z_k 的边缘化，可以得到观察变量的联合概率分布

$$P(d_i, \boldsymbol{x}_j) = P(d_i) \sum_{k=1}^{K} P(\boldsymbol{z}_k \mid d_i) P(\boldsymbol{x}_j \mid \boldsymbol{z}_k) \quad (1)$$

依据图模型表示规则，连续 PLSA 可以用图 3 描述。

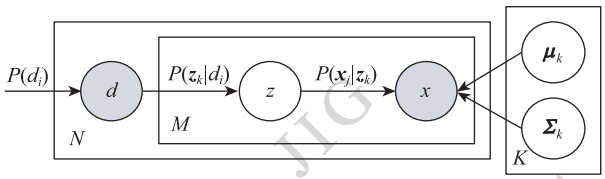


图 3　连续 PLSA 的图模型表示

Fig. 3　Graphical model representation of continuous PLSA

假设高斯混合分布由对应的条件概率密度 $P(\cdot \mid z)$ 构成。换句话说，元素是从 K 个高斯分布中生成，而每个高斯分布对应于一个 $\boldsymbol{z}_k$。对于一个特定的潜在主题 z_k，元素 $\boldsymbol{x}_j$ 的条件概率密度函数为

$$P(\boldsymbol{x}_j \mid z_k) = N(\boldsymbol{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{(2\pi)^{D/2} \mid \boldsymbol{\Sigma}_k \mid^{1/2}} \times \exp\left\{ -\frac{1}{2} (\boldsymbol{x}_j - \boldsymbol{\mu}_k)^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_j - \boldsymbol{\mu}_k) \right\} \quad (2)$$

式中，D 是维数，$\boldsymbol{\mu}_k$ 是一个 D 维均值向量，$\boldsymbol{\Sigma}_k$ 是一个 $D \times D$ 的协方差矩阵。

依据极大似然原则，$P(z_k \mid d_i)$ 和 $P(\boldsymbol{x}_j \mid z_k)$ 可通过最大化以下的对数似然函数来确定，即

$$L = \sum_{i=1}^{N} \sum_{j=1}^{M} n(d_i, \boldsymbol{x}_j) \log P(d_i, \boldsymbol{x}_j) = \sum_{i=1}^{N} \sum_{j=1}^{M} n(d_i, \boldsymbol{x}_j) \left[\log P(d_i) + \log \sum_{k=1}^{K} P(z_k \mid d_i) P(\boldsymbol{x}_j \mid z_k) \right] \quad (3)$$

式中，$n(d_i, \boldsymbol{x}_j)$ 表示文档 d_i 中包含元素 $\boldsymbol{x}_j$ 的个数。

潜在变量模型中极大似然估计的标准过程是期望最大化（EM）算法[24]，其步骤如下[5-6]：

E 步：对式（1）应用贝叶斯法则可得

$$P(z_k \mid d_i, \boldsymbol{x}_j) = \frac{P(z_k \mid d_i) P(\boldsymbol{x}_j \mid z_k)}{\sum_{l=1}^{K} P(z_l \mid d_i) P(\boldsymbol{x}_j \mid z_l)} \quad (4)$$

M 步：利用式（4）更新参数 $P(z_k \mid d_i)$、$\boldsymbol{\mu}_k$ 和 $\boldsymbol{\Sigma}_k$，即

$$P(z_k \mid d_i) = \kappa_k = \frac{\sum_{j=1}^{M} n(d_i, \boldsymbol{x}_j) P(z_k \mid d_i, \boldsymbol{x}_j)}{\sum_{j=1}^{M} n(d_i, \boldsymbol{x}_j)} \quad (5)$$

$$\boldsymbol{\mu}_k = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} n(d_i, \boldsymbol{x}_j) P(z_k \mid d_i, \boldsymbol{x}_j) \boldsymbol{x}_j}{\sum_{i=1}^{N} \sum_{j=1}^{M} n(d_i, \boldsymbol{x}_j) P(z_k \mid d_i, \boldsymbol{x}_j)} \quad (6)$$

$$\boldsymbol{\Sigma}_k = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} n(d_i, \boldsymbol{x}_j) P(z_k \mid d_i, \boldsymbol{x}_j) (\boldsymbol{x}_j - \boldsymbol{\mu}_k) (\boldsymbol{x}_j - \boldsymbol{\mu}_k)^{\mathrm{T}}}{\sum_{i=1}^{N} \sum_{j=1}^{M} n(d_i, \boldsymbol{x}_j) P(z_k \mid d_i, \boldsymbol{x}_j)} \quad (7)$$

© 中国图象图形学报版权所有

EM 算法可以采用早期停止技术或者通过修改收敛条件来终止执行。此外,如果已知参数 $P(z_k \mid d_i)$,可以使用 folding-in 算法快速推导其他参数 $\boldsymbol{\mu}_k$ 和 $\boldsymbol{\Sigma}_k$,反之亦然。folding-in 算法[23]是 EM 算法的不完全版本,该算法在迭代过程中保持已知参数不变,不断更新未知参数使得似然函数极大化。

2.2 集群分类器链

图像自动标注问题可以看做一个多类分类的问题,因为图像标注的最终目标是将图像归入多个语义类别。尽管多类分类器能够解决弱标注等问题并具有很多优点,但它计算复杂度较高,而且没有考虑图像的标注之间的关联,也就没有充分利用现有的可得到的信息。采用多标记学习的方法进行图像的自动标注能够在对图像进行分类的同时集成图像标注关键词之间的关联,是一种适合图像语义标注任务的学习技术。

多标记分类的方法可以分为两类[25]:一类是"问题转化";另一类是"算法改进"。第 1 类方法独立于算法,它将学习任务转化为一个或多个单标记分类任务,而完成单标记分类的任务有很多算法可供选择。第 2 类方法直接对特定算法进行改进以处理多标记数据。二值相关(BR)方法是应用得最广泛的问题转化方法之一,它学习多个二值分类器,每一个分类器对应于标记集合 L 中的一个标记。

分类器链(CC)[7]与 BR 方法一样,包含 $|L|$ 个二类分类器,每个分类器处理一个标记 l_j 的二值相关问题。但是与 BR 方法不同的是,这些二值分类器都通过一个链连接起来,其中每个结点的特征空间都与前面结点的类标记有关。

分类器链的训练算法如算法 1 所示,这里训练样本表示为 $(\boldsymbol{x}, \boldsymbol{S})$,$\boldsymbol{S} \subseteq \boldsymbol{L}$ 可以用二值向量 $(l_1, l_2, \cdots, l_{|L|})$ 表示,而 $\boldsymbol{x}$ 是特征向量。算法 1 中,按照指定的标记顺序,每次循环学习关联一个标记的二值分类器,更重要的是,每次循环特征空间都要加上已学习的分类器对应的标记信息,因而特征信息不断地得到增强,最后,可以构造一个二值分类器链,链上的每一个分类器 C_j 负责与标记 l_j 相关的学习和预测。

算法 1:分类器链的训练算法

输入:样本集 $\boldsymbol{D} = \{(\boldsymbol{x}_1, \boldsymbol{S}_1), (\boldsymbol{x}_2, \boldsymbol{S}_2), \cdots, (\boldsymbol{x}_n, \boldsymbol{S}_n)\}$。

输出:分类器链 $\{C_1, C_2, \cdots, C_{|L|}\}$。

for $j \in 1, 2, \cdots, |\boldsymbol{L}|$
  do 单标记的转化和训练
    $\boldsymbol{D}' \leftarrow \{\}$
    for $(\boldsymbol{x}, \boldsymbol{S}) \in \boldsymbol{D}$
      do $\boldsymbol{D}' \leftarrow \boldsymbol{D}' \cup ((\boldsymbol{x}, l_1, l_2, \cdots, l_{j-1}), l_j)$
    end for
    训练与 l_j 相关的二值分类器 C_j
    $C_j: \boldsymbol{D}' \rightarrow l_j \in \{0, 1\}$
end for

分类器链的分类过程从分类器 C_1 开始,再不断地向后传播。分类器 C_1 确定标记 l_1 的分类结果 $Pr(l_1 \mid \boldsymbol{x})$,然后将这个结果以二值的方式加入测试样本的特征,后续的分类器则确定标记 l_j 的分类结果 $Pr(l_j \mid x_i, l_1, l_2, \cdots, l_{j-1})$。这个过程可以用算法 2 表示。

算法 2:分类器链的分类算法

输入:测试样本 $\boldsymbol{x}$。

输出:样本 $\boldsymbol{x}$ 在所有分类器上的分类结果 $\boldsymbol{Y} = (C_1, C_2, \cdots, C_{|L|})$。

$\boldsymbol{Y} \leftarrow \{\}$
for $j \in 1, 2, \cdots, |\boldsymbol{L}|$
  do $\boldsymbol{Y} \leftarrow \boldsymbol{Y} \cup (l_j \leftarrow C_j: (\boldsymbol{x}, l_1, l_2, \cdots, l_{j-1}))$
end for
return $(\boldsymbol{x}, \boldsymbol{Y})$

使用链的方法可以在分类器间传递标记信息,同时考虑标记之间的关联信息,因而能克服 BR 方法中的标记独立问题。而且,分类器链仍然保持 BR 方法的优势,包括存储需求低和运行效率高等。

虽然对于每个示例平均要增加 $|\boldsymbol{L}|/2$ 维的特征数据量,但由于在实践中 $|\boldsymbol{L}|$ 一般是一个有限的值,因而由这个原因引起的计算复杂度问题几乎可以忽略不计。分类器链的计算复杂度与 BR 方法非常接近,取决于标记的个数与基本的二值分类器的复杂度。BR 方法的复杂度是 $O(|\boldsymbol{L}| \times f(|\boldsymbol{X}|, |\boldsymbol{D}|))$,其中 $f(|\boldsymbol{X}|, |\boldsymbol{D}|)$ 是基本的二值分类器的复杂度。分类器链的复杂度则为 $O(|\boldsymbol{L}| \times f(|\boldsymbol{X}| + |\boldsymbol{L}|, |\boldsymbol{D}|))$,也就是多了 $|\boldsymbol{L}|$ 维附加的特征值。而 HGDM 采用线性分类器 SVM 作为基本的二值分类器,所以分类器链的复杂度可以简化为 $O(|\boldsymbol{L}| \times |\boldsymbol{X}| \times |\boldsymbol{D}| + |\boldsymbol{L}| \times |\boldsymbol{L}| \times |\boldsymbol{D}|)$。可以看到,只要 $|\boldsymbol{L}| < |\boldsymbol{X}|$,第一项就会起主要作用。这样分类器链的计算复杂度则为 $O(|\boldsymbol{L}| \times |\boldsymbol{X}| \times |\boldsymbol{D}|)$,与 BR 方法的计算复杂度相同。而只有当 $|\boldsymbol{L}| > |\boldsymbol{X}|$ 时,分类

© 中国图象图形学报版权所有

器链的计算复杂度才会高于 BR 方法。

此外，虽然链式的过程意味着分类器链不能够并行化，但它能够串行化，也就是在任何时候内存中只需要保留一个二值分类器，对比别的方法这是一个明显的优势。

分类器链的顺序显然会影响它的精度。虽然有一些启发式算法来确定链的顺序，但还是采用集群式框架来解决这个问题。采用集群式方法能够提高整体精度、避免过拟合，也能够实现并行化。注意二值分类的方法也常常称为集群式方法，因为包含多个二值分类模型。但是，这些模型都不能完成多标记任务。本文所说的集群是指多标记方法的集群，也就是分类器链的集群。

集群分类器链训练 m 个分类器链 C_1，C_2，…，C_m，其中每个分类器链都由一个随机的链顺序和训练集的一个随机子集训练得到。因此每一个模型 C_k 都是互不相同的而且能够给出不同的多标记分类结果。将这些分类结果按照标记作求和计算，那么每一个标记都会得到若干投票。使用一个阈值选择票数最高的标记可以构成一个多标记集合，并以此作为最终的分类结果。

设第 k 个单独的模型的预测结果为向量 $\boldsymbol{y}_k = (l_1, l_2, \cdots, l_{|L|}) \in \{0,1\}^{|L|}$。对所有模型求和可得向量 $\boldsymbol{W} = (\lambda_1, \lambda_2, \cdots, \lambda_{|L|}) \in \mathbf{R}^{|L|}$，这里 $\lambda_j = \sum_{k=1}^{m} l_j \in y_k$。因此每一个 $\lambda_j \in \boldsymbol{W}$ 都代表了对标记 $l_j \in \boldsymbol{L}$ 投票的结果。对向量 $\boldsymbol{W}$ 作归一化得到 $\boldsymbol{W}^{\text{norm}}$，就能够得到每个标记在[0,1]上的一个分布。做完阈值的判定之后，可以根据这个分布作一个排序。与别的图像自动标注模型类似，HGDM 取前 5 个置信度最高的关键词标记作为图像的语义标注。

综上所述，HGDM 在学习过程中混合了生成式模型和判别式模型，对于输入图像视觉特征的学习采用生成式模型，而对于图像的语义学习过程采用判别式模型，因而具有以下几个特点：

1）在生成式学习阶段，采用连续 PLSA 直接建模图像视觉特征，不需要进行视觉特征的量化过程，因而不会丢失重要的视觉信息。

2）连续 PLSA 将图像从特征集合的表示变换为一个 K 维的主题向量表示，这也可以视为一个降维的过程。而这个主题向量的表示也集成了图像视觉内容的隐含语义信息，对于图像的语义检索具有重大意义。

3）基于多标记学习方法构建分类器，在对图像进行分类的同时集成了图像标注关键词之间的关联。能够很好地解决弱标注问题，对训练集规模具有可扩展性。

4）判别式学习阶段采用集群分类器链进行图像的语义分类。分类器链中的每个二值分类器都基于 SVM 构建，所以运行效率和分类精度都较高，且具有可接受的计算复杂性。

5）从整体上看，由于集群分类器链的计算复杂性和效率都接近 BR 方法，HGDM 的计算复杂性主要集中在利用连续 PLSA 建模图像数据集所运行的 EM 算法。由于没有对视觉特征进行量化，连续 PLSA的 EM 算法在迭代过程中需要处理从图像文档中提取得到的所有特征向量，会付出较大的时间代价。但是，由于图像数据集的建模是离线完成，建模之后的图像可以用维数低得多的主题向量表示。而在线进行的算法只是利用训练好的集群分类器链对测试图像的主题向量进行多标记分类。因此，连续 PLSA 建模所耗费的时间并不影响在线的标注和检索算法的效率。

3　实验结果分析

在实验部分开发的原型系统中，实现了 PLSA-WORDS、PLSA-FUSION、GM-PLSA 和本文提出的 HGDM。原型系统中基于连续 PLSA 的模型拟合和分类器训练的过程离线执行，而图像的标注和检索结果则在线展示。

3.1　数据集

为了测试 HGDM 的有效性与精度并与其他典型的图像自动标注方法进行比较，采用两个图像标注领域的基准数据集进行实验，即 Corel5k 和 IAPR TC12。

Corel5k 最初用于文献[8]，是在图像标注领域用于与近期的研究工作相比较的基本数据集。该数据集包含 5 000 幅图像，来自 50 个 Corel 库存图像光盘，每个光盘包含 100 幅同样主题的图像。将这个数据集分成 3 个部分：一个包含 4 000 幅图像的训练集，一个包含 500 幅图像的验证集和一个包含 500 幅图像的测试集，其中验证集用于确定系统参数。而在参数确定之后，将 4 000 幅图像的训练集

© 中国图象图形学报版权所有

和500幅图像的验证集合并为一个新的训练集,与文献[8]中使用的一个包含4 500幅图像的训练集和一个包含500幅图像的测试集一致。此外,只有至少标注了8幅图像的关键词(总共260个)才能入选词汇表。作为基准数据集,Corel5k数据集可以对大多数图像标注方法进行比较和评估,但是该数据集常被挑剔为图像数量不足且标注的关键词不够自然。

相比Corel5k数据集,IAPR TC12数据集[22]是一个更具挑战性的数据集。该数据集包含约20 000幅自然场景图像,这个规模更适合于测试图像标注方法的可扩展性。由于该数据集中图像对应的标注是自然语句,需要通过预处理提取语句中的名词作为关键词。经过预处理的数据集共包含19 805幅图像和291个关键词,平均每幅图像标有4.7个关键词。可以选取数据集中的17 825幅图像用于训练,1 980幅图像用于测试,使得测试集和训练集的数量比大致与Corel5k相同。

3.2 评估措施

图像标注的性能评估是通过比较对测试集自动生成的标注与原始标注来实现的。类似于文献[12-13],定义后验概率最大的5个语义关键词作为图像的自动标注,并计算测试集中每个关键词的召回率和精度。给定一个语义关键词,精度 $precision = B/A$,召回率 $recall = B/C$,这里 A 表示自动生成的标注包含给定关键词的图像个数,B 表示正确标注了给定关键词的图像个数,C 表示原始标注中包含了给定关键词的图像个数。关键词检索得到的精度和召回率的平均值能够大体上代表系统性能。最后,还考虑了非零召回率(即 $B>0$ 的关键词)的关键词个数,这个数字能够指示系统能够有效地学习多少个关键词。

此外,语义检索的性能也能通过检测精度和召回率来衡量。然而这些措施对于综合评估检索性能还不够充分,于是使用了另一个称为 mAP(mean average precision)的重要度量标准。mAP 在多年以前就已成为文本检索领域标准的度量措施,它能够以一种有意义的方式总结检索性能。为计算 mAP,首先将每个查询 q 的 AP(average precision)定义为在各个正确的检索图像相应的位置 i 的精度 $pre(i)$ 之和,除以本次查询的相关图像个数 $rel(q)$,即

$$AP(q) = \frac{\sum_{i \in \boldsymbol{R}} pre(i)}{rel(q)} \tag{8}$$

式中,$\boldsymbol{R}$ 表示本次查询的所有相关图像在检索结果中的位置集合。

因而查询的 AP 对于文档的整体排位是敏感的。于是对 N_q 次查询的 AP 取平均值,即能将检索系统的性能归结为 mAP 一个值,有

$$mAP = \frac{\sum_{q=1}^{N_q} AP(q)}{N_q} \tag{9}$$

3.3 参数设置

在实验过程中,连续PLSA的主题个数设置很重要,因为主题个数决定了图像的中间表示的维数。这个数目过大则会降低系统的效率,过小则会丢失图像信息。由于连续PLSA的拟合比较费时,选取了5个主题个数(分别为90、120、150、180与210)进行实验,验证集的实验结果表明,当主题个数为180时,系统性能比使用其他主题个数时好,所以最终确定的主题个数为180。

在Corel5k数据集中,本文方法构造一个包含90个分类器链的集群,每个分类器链随机的选取一个包含500幅图像的子集进行训练。而在数据集IAPR-TC12上进行实验时,使用150个分类器链的集群,每个分类器链随机的选取一个包含1 000幅图像的子集进行训练。不排除选择更多的分类器链或者选择更大的训练子集时效果可能会更好,但是需要耗费更多的时间代价。此外,分类器链中的各个结点所代表的二值分类器使用LIBSVM软件包实现,选用RBF核函数 $K(\boldsymbol{x}, \boldsymbol{x}') = \exp(-\gamma \| \boldsymbol{x} - \boldsymbol{x}' \|^2)$,相应的参数通过网格搜索法确定为 $(C, \gamma) = (2^7, 2^1)$,其中 C 为误差惩罚因子,γ 为核函数参数。

由于本文方法独立于所采用的视觉特征,故采用与文献[13]类似的视觉特征,以便与其他模型进行性能比较。首先将数据集中的每幅图像划分为规则方块(方块大小由验证集确定为16×16),然后为每个方块提取一个36维的特征向量,包含24维的颜色特征和12维的纹理特征,颜色特征是在8个量化颜色和3个街区距离上计算的颜色自相关图[26],纹理特征是在3个尺度和4个方向上计算的Gabor能量系数[27]。于是,每个方块就可表示为一个36维的特征向量,从而每幅图像就可表示为一个“特征袋”,也就是若干个36维的视觉特征向量的集合。这些预处理步骤为进一步使用连续PLSA进行建模提供了一致的接口。

© 中国图象图形学报版权所有

3.4　Corel5k 上的实验结果

3.4.1　图像自动标注结果比较

将 HGDM 的性能与许多代表性模型进行比较，包括翻译模型[8]、CMRM[11]、CRM[12]、MBRM[13]、PLSA-WORDS[15]、TGLM[17]、SML[20]、MSC[21] 和 JEC[22]。图像标注的性能评估通过比较自动生成的标注与原始标注来实现。类似于文献[12-13]，计算测试集中每个关键词的召回率和精度，使用其平均值来总结系统性能。

所得标注结果在两个关键词集合上报告：性能最佳的 49 个关键词集合和训练集中所有 260 个关键词集合。系统评估结果如表 1 所示。从表 1 中可以看到 HGDM 的性能要优于其他大多数模型，它的性能明显高于 MSC，与 JEC 的性能几乎相同。使用混合框架学习语义概念是得到这个结果的主要原因，因为混合学习方法能够在某种程度上解决文本关键词表示稀疏引起的问题。表 2 给出了几个自动标注的结果实例，包括 PLSA-WORDS 和 HGDM 的标注结果，可以看到大多数情况下 HGDM 的标注比 PLSA-WORDS 更合理。

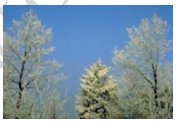
表 1　Corel5k 上图像自动标注模型的性能比较

Table 1　Performance comparison of automatic image annotation models on Corel5k

模型	召回率 >0 的关键词个数	49 个性能最佳的词集上的结果		全部 260 个词集上的结果	
		平均精度	平均召回率	平均精度	平均召回率
TM	49	0.20	0.34	0.06	0.04
CMRM	66	0.40	0.48	0.10	0.09
CRM	107	0.59	0.70	0.16	0.19
MBRM	122	0.74	0.78	0.24	0.25
PLSA-WORDS	105	0.56	0.71	0.14	0.20
TGLM	131	—	—	0.25	0.29
SML	137	—	—	0.23	0.29
MSC	136	0.76	0.82	0.25	0.32
JEC	139	—	—	0.27	0.32
HGDM	137	0.78	0.83	0.28	0.32

表 2　Corel5k 上 PLSA-WORDS 和 HGDM 生成标注的比较

Table 2　Comparison of annotations made by PLSA-WORDS and HGDM on Corel5k

图像	原始标注	PLSA-WORDS 的自动标注	HGDM 的自动标注
	blue-footed, booby, rock, bird	bird, nest, close-up, booby, branch	booby, bird, trees, rock, sky
	trees, sky, frost, ice	grass, desert, ice, path, sculpture	sky, trees, branch, ice, grass
	sand, water, people, sky	beach, iceburg, snow, ice, water	water, sky, beach, people, snow
	building, courtyard, sky, trees	sky, building, wall, sand, temple	building, sky, wall, trees, courtyard

© 中国图象图形学报版权所有

图 4 展示了 PLSA-WORDS、SML、MSC 与 HGDM 在 Corel5k 数据集上的精度—召回率曲线，标示了当为每幅图像自动标注关键词个数从 2 变化到 10 时，所有关键词的平均精度和平均召回率结果。正如图 4 中曲线所示，HGDM 的精度和召回率均持续

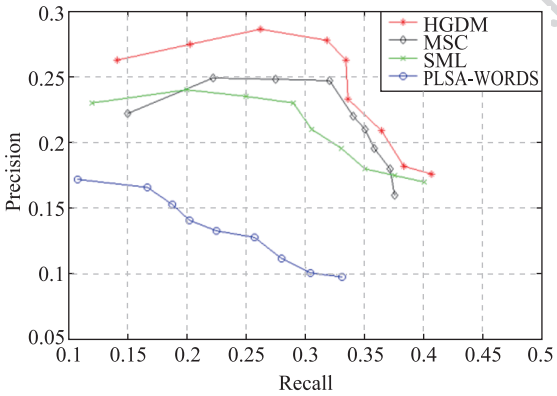


图 4　Corel5k 上几个图像自动标注模型的精度—召回率曲线

Fig. 4　Precision-recall curves of several models for automatic image annotation on Corel5k

的高于 SML 和 PLSA-WORDS，而与 MSC 相比较召回率较接近而精度稍高。

3.4.2　图像语义检索结果比较

标注的结果忽略排位顺序。但是，用户通常都乐意对检索的图像进行排位，并希望相关图像排位在前。实际上，大多数用户都不愿意在一次查询中浏览即使是 10 幅或 20 幅的图像。给定一个查询词，系统将返回自动标注了该查询词的所有图像，并按照该词对应的后验概率对图像进行排位。

表 3 给出了几种典型的标注模型的 *mAP*，第 2 列是在词汇表中所有关键词集合上计算得到的 *mAP*，第 3 列是在召回率大于等于零的关键词集合（即在测试集的原始标注中出现过的关键词集合）上计算得到的 *mAP*，第 4 列是在召回率大于零的关键词集合（即在表 1 中的第 2 列数据）上计算得到的 *mAP*。由表 3 中的数据可见，HGDM 的检索性能要高于除 MSC 之外的其他方法。

表 3　HGDM 与其他标注模型在 Corel5k 数据集上的 *mAP* 值比较

Table 3　Comparison of *mAP*s of HGDM and other annotation models on Corel5k

模型	全部 260 词	召回率≥0 的词	召回率 >0 的词
CMRM	0.17	0.20	—
CRM	0.24	0.27	—
MBRM	0.30	0.35	—
PLSA-WORDS	0.22	0.26	0.55
SML	0.31	—	0.49
TGLM	0.29	—	0.52
MSC	0.42	—	0.79
JEC	0.33	—	0.52
HGDM	0.35	0.41	0.67

图 5 展示了对 3 个具有挑战性的概念进行单个关键词查询的排位检索结果。图 5 中每行给出一个查询中排位最高的 5 幅匹配图像，由上至下的查询关键词分别为“fox”、“house”和“sunset”。返回检索结果的多样性可以表明，HGDM 具备良好的学习和泛化能力。

3.5　IAPR TC12 上的实验结果

IAPR TC12 数据集最先用于文献[22]中的图像自动标注和检索任务，是一个相对较新的数据集。因此只与那些可以直接从文献中获取实验数据的方法进行比较，如 JEC 与 MBRM。此外，也报告了在原型系统中已实现的方法（如 PLSA-WORDS）所得到的相关实验数据。

表 4 给出了 IAPR TC12 数据集上的实验结果，包括召回率大于零的关键词个数、平均精度、平均召回率、所有 291 个关键词的 *mAP* 值、召回率大于零的关键词的 *mAP* 值 5 个性能指标。其中前 3 个主要用于评估自动标注的性能，后两个用于评估语义检索的性能。从表 4 中数据可见，相比其他先进的标注模型，HGDM 具有最高的性能，这表明该模型具备很好的稳定性和可扩展性。

表 5 给出了一些由 HGDM 和 PLSA-WORDS 得出的标注实例。对比由 HGDM 得到的标注与原始标注可见，即使某个 HGDM 标注的关键词没有出现

© 中国图象图形学报版权所有

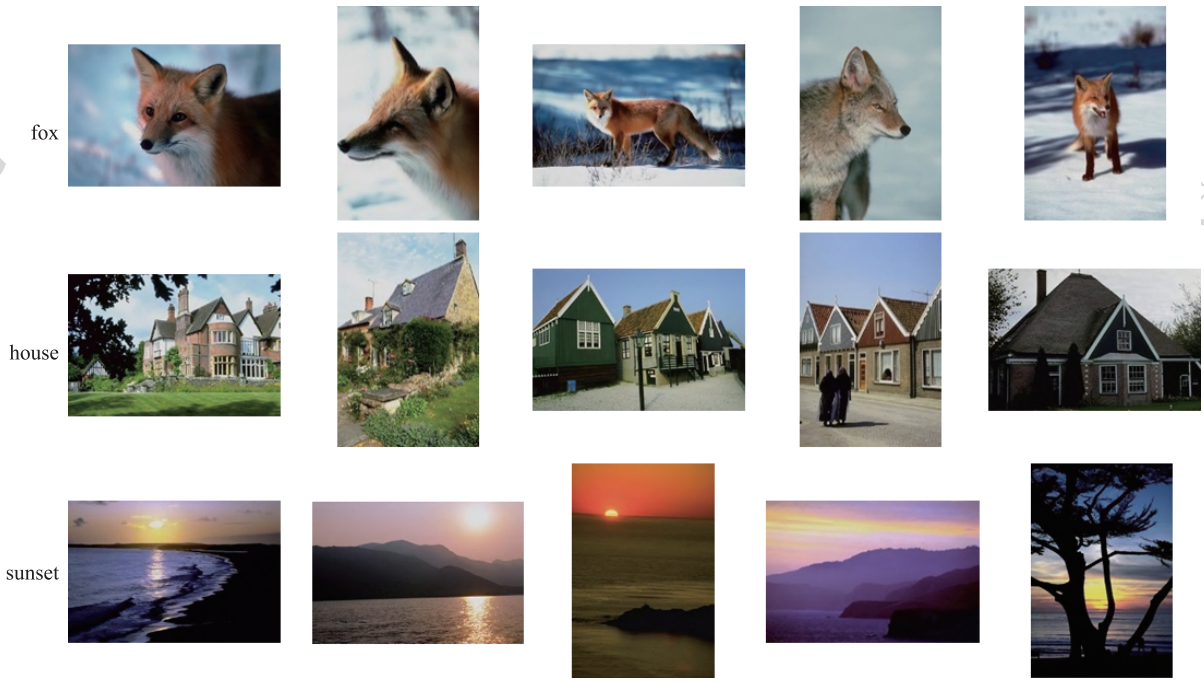


图 5　HGDM 在 Corel5k 上的 3 个语义检索实例

Fig. 5　Three semantic retrieval results of HGDM on Corel5k

表 4　IAPR TC12 上图像自动标注模型的性能比较

Table 4　Performance comparison of automatic image annotation models on IAPR TC12

模型	平均精度	平均召回率	召回率 >0 的关键词个数	所有 291 个关键词的 *mAP* 值	召回率 >0 的关键词的 *mAP* 值
MBRM	0. 21	0. 14	186	0. 23	0. 36
PLSA-WORDS	0. 18	0. 12	177	0. 19	0. 30
JEC	0. 25	0. 16	196	0. 27	0. 41
HGDM	0. 29	0. 18	194	0. 30	0. 45

表 5　IAPR TC12 上 PLSA-WORDS 和 HGDM 生成标注的比较

Table 5　Comparison of annotations made by PLSA-WORDS and HGDM on IAPR TC12

图像	原始标注	PLSA-WORDS 的自动标注	HGDM 的自动标注
	building, cupola, column, tree	sky, tree, snow, wall, group	building, tree, sky, house, column
	tourist, river, ravine, forest, waterfall	tree, grass, water, forest, cloud	forest, river, snow, waterfall, ravine
	people, meadow, fence, tree, sky	grass, sky, tree, stone, building	meadow, tree, sky,people, house
	girl, skirt, waistcoat, hill, house	sky, sand, wall, desert, landscape	girl, sky, people, cloud, hill

© 中国图象图形学报版权所有

在原始标注中,它往往也能恰当的表示图像的内容(如第 1 幅图像标注的“sky”和第 4 幅图像标注的“people”)。

在 Corel5k 和 IAPR TC12 数据集上的实验结果表明,HGDM 的自动标注质量和语义检索效果要高于大多数当前先进的图像标注方法,与当前最具代表性的方法 JEC 的性能相当,这表明了混合生成式模型和判别式模型的学习方法是可行的和有效的。同时,HGDM 的性能相对于参数设置比较稳定,具备很好的鲁棒性和可扩展性。

4 结 论

本文提出混合生成式和判别式模型的图像自动标注方法 HGDM。通过连续 PLSA 的生成式学习,不仅可以获得相应的模型参数,还可以将每幅图像都表示为在各个主题上的分布。于是,图像自动标注的问题就可以转化为一个多标记学习的问题。随后,HGDM 采用构造集群分类器链的方法进行多标记学习,在对图像进行分类的同时也考虑了图像标注之间的关联信息。在学习过程中,HGDM 对于图像视觉特征的学习采用生成式模型,而对于图像的语义学习过程采用判别式模型,因而既具有生成式模型的可扩展性,也具有判别式模型精度较高的优点。在 Corel5k 和 IAPR TC12 数据集上的实验结果表明,对比当前大多数先进的图像自动标注模型,HGDM 具有更高的精度和更好的效果。

由于本文方法独立于媒体对象的低层特征,所以经过适当的改进之后该方法也能在跨媒体检索和数据挖掘领域发挥重要作用。此外,如果图像自动标注的方法经过提升能够获得更高的性能,就能够进一步应用于一些实际的课题,如数字图书馆、视频检测和过滤、医学辅助治疗等。下一步的工作将考虑引入半监督学习或深度学习的技术以提升图像自动标注和检索的性能。

参考文献(References)

[1] Li Z X, Shi Z P, Li Z Q, et al. A survey of semantic mapping in image retrieval [J]. Journal of Computer-Aided Design and Computer Graphics, 2008, 20(8): 1085-1096. [李志欣, 施智平, 李志清, 等. 图像检索中语义映射方法综述[J]. 计算机辅助设计与图形学学报, 2008, 20(8): 1085-1096.]

[2] Smeulders A W M, Worring M, Santini S, et al. Content-based image retrieval at the end of the early years [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(12): 1349-1380.

[3] Datta R, Joshi D, Li J, et al. Image retrieval: ideas, influences, and trends of the new age[J]. ACM Computing Surveys, 2008, 40(2): 1-60.

[4] Zhang D S, Islam M M, Lu G J. A review on automatic image annotation techniques [J]. Pattern Recognition, 2012, 45(1): 346-362.

[5] Li Z X, Shi Z P, Liu X, et al. Automatic image annotation with continuous PLSA [C]// Proceedings of the 35th IEEE International Conference on Acoustics, Speech and Signal Processing. Los Alamitos: IEEE Computer Society, 2010: 806-809.

[6] Li Z X, Shi Z P, Liu X, et al. Semantic image annotation by modeling continuous visual features [J]. Journal of Computer-Aided Design and Computer Graphics, 2010, 22(8): 1412-1420. [李志欣, 施智平, 刘曦, 等. 建模连续视觉特征的图像语义标注方法[J]. 计算机辅助设计与图形学学报, 2010, 22(8): 1412-1420.]

[7] Read J, Pfahringer B, Holmes G, et al. Classifier chains for multi-label classification [J]. Machine Learning, 2011, 85(3): 333-359.

[8] Duygulu P, Barnard K, de Freitas J F G, et al. Object recognition as machine translation: learning a lexicon for a fixed image vocabulary [C] // Lecture Notes in Computer Science, Berlin: Springer, 2002, 2353: 97-112.

[9] Barnard K, Duygulu P, Forsyth D, et al. Matching words and pictures [J]. Journal of Machine Learning Research, 2003, 3(2): 1107-1135.

[10] Blei D M, Jordan M I. Modeling annotated data [C] // Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2003: 127-134.

[11] Jeon J, Lavrenko V, Manmatha R. Automatic image annotation and retrieval using cross-media relevance models[C] // Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2003: 119-126.

[12] Lavrenko V, Manmatha R, Jeon J. A model for learning the semantics of pictures[C] // Advances in Neural Information Processing Systems 16. Cambridge: MIT Press, 2004: 553-560.

[13] 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. Los Alamitos: IEEE Computer Society Press, 2004: 1002-1009.

[14] Zhang R F, Zhang Z F, Li M J, et al. A probabilistic semantic

© 中国图象图形学报版权所有

model for image annotation and multi-modal image retrieval [C] // Proceedings of the 10th IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2005: 846-851.

[15] Monay F, Gatica-Perez D. Modeling semantic aspects for cross-media image indexing [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(10): 1802-1817.

[16] Li Z X, Shi Z P, Li Z Q, et al. Automatic image annotation by fusing semantic topics [J]. Journal of Software, 2011, 22(4): 801-812. [李志欣, 施智平, 李志清, 等. 融合语义主题的图像自动标注[J]. 软件学报, 2011, 22(4): 801-812.]

[17] Liu J, Li M J, Liu Q S, et al. Image annotation via graph learning[J]. Pattern Recognition, 2009, 42(2): 218-228.

[18] Li J, Wang J Z. Automatic linguistic indexing of pictures by a statistical modeling approach [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(9): 1075-1088.

[19] Chang E, Goh K, Sychay G, et al. CBSA: content-based soft annotation for multimodal image retrieval using Bayes point machines[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2003, 13(1): 26-38.

[20] Carneiro G, Chan A B, Moreno P J, et al. Supervised learning of semantic classes for image annotation and retrieval [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(3): 394-410.

[21] Wang C H, Yan S C, Zhang L, et al. Multi-label sparse coding for automatic image annotation [C] // Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2009: 1643-1650.

[22] Makadia A, Pavlovic V, Kumar S. Baselines for image annotation [J]. International Journal of Computer Vision, 2010, 90(1): 88-105.

[23] Hofmann T. Unsupervised learning by probabilistic latent semantic analysis[J]. Machine Learning, 2001, 42(1-2): 177-196.

[24] Bishop C M. Pattern Recognition and Machine Learning [M]. Berlin: Springer, 2006: 423-455.

[25] Li Z X, Zhuo Y Q, Zhang C L, et al. Survey on multi-label learning [J]. Application Research of Computers, 2014, 31(6): 1601-1605. [李志欣, 卓亚琦, 张灿龙, 等. 多标记学习研究综述[J]. 计算机应用研究, 2014, 31(6): 1601-1605.]

[26] Huang J, Kumar S R, Mitra M, et al. Spatial color indexing and applications [J]. International Journal of Computer Vision, 1999, 35(3): 245-268.

[27] Manjunath B S, Ma W Y. Texture features for browsing and retrieval of image data [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 18(8): 837-842.

© 中国图象图形学报版权所有