

基于主题融合和关联规则挖掘的图像标注

张 蕾 蔡 明

(江南大学物联网工程学院 江苏 无锡 214122)

摘 要 为减小“语义鸿沟”,在 LDA 主题模型的基础上,提出了一种主题融合和关联规则挖掘的图像标注方法。首先,针对视觉和文本信息的关联度不高的问题,引入基于向量机的多类别分类得到图像的分类信息。其次,通过文本模态的语义主题分布和类别信息,计算出图像类的文本主题分布。未知图像将其所属类的文本主题分布与其视觉主题分布进行加权融合,并以此概率模型计算初始标签集。最后依据初始标注词概率,利用关联规则挖掘和词间相关性挖掘文本关联度,从而得到精确化语义标注。在 Corel5K 图像数据集上进行对比实验,实验结果证明了方法的有效性。

关键词 图像标注, LDA 主题模型, 加权主题融合, 关联规则挖掘, 词间相关性

中图分类号 TP391 文献标识码 A DOI 10.11896/j.issn.1002-137X.2019.07.037

Image Annotation Based on Topic Fusion and Frequent Patterns Mining

ZHANG Lei CAI Ming

(College of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China)

Abstract In order to reduce the “semantic gap”, based on the LDA topic model, an image annotation approach which uses topics fusion and association rule mining was proposed. First, to solve the problem of low correlation between visual and text information, the vector machine-based multi-category classification is introduced to obtain the category information of the image. Then, the text topic distribution of the image class is calculated by the semantic topic distribution and classification information of the text modality. The unknown image weights the text topic distribution of its class and its visual topic distribution, and calculates the initial label set using this probability model. Finally, based on the probability of initial label words, the association rules mining and inter-word correlation are used to mine the text relevance to obtain precise semantic annotation. The comparative experiments were carried out on the Corel5K image dataset. The experimental results show the effectiveness of the proposed method.

Keywords Image annotation, LDA topic model, Weighted topic fusion, Frequent patterns mining, Correlation of keyword.

1 引言

随着自媒体技术的发展,网络用户上传了海量的图像,如何从大规模的图像库中快速、准确地检索到满足用户需求的图像,成为了图像领域的重要的研究方向。传统手工标注的主观性和低效率问题,影响了基于文本的图像检索结果的稳定性。为此,利用计算机从大量有标注的图像中学习语义概念模型,并使用此模型对未知图像进行标注的图像自动标注^[1-2]应用而生,这是当前最具有挑战性的课题。

图像自动标注的模型包括: Duygulu 等提出的机器翻译模型^[3] (Translation Model, TM)、Jeon 等提出的跨媒体相关模型^[4] (Cross-media Relevance Model, CMRM)、Lavrenko 等提出了连续空间相关模型 (Continuous-space Relevant Model, CRM)^[5]、Feng 等提出的多伯努利相关模型 (Multiple Ber-

noulli Relevance Model, MBRM)^[6]。学者应用概率主题模型^[7],如 Thomas 提出的 PLSA (Probabilistic Latent Semantic Indexing) 模型^[8]、Blei 等提出的 LDA (Latent Dirichlet Allocation) 模型^[9]等在图像标注方面做了大量的工作。近年来,研究人员从不同方面对概率图像标注模型进行了改进,如 Monay 等^[10]提出了 PLSA-WORDS 模型。在此基础上,李志欣等^[11-12]提出了融合文本和视觉模态的标注模型 PLSA-FUSION、混合生成式和判别式模型的图像自动标注方法。赵鹏等^[13]融合了 PLSA 模型与 GMM 模型,并提出了基于概率潜语义的高斯混合模型。孙君顶等^[14-16]提出了基于视觉语义主题的图像自动标注、基于多特征融合与 PLSA-GMM 的图像自动标注和基于视觉语义主题与反馈日志的图像自动标注。田东平等^[17]提出了融合 PLSA 和随机游走模型的自动图像标注。Cao 等^[18-19]提出了改进 corr-LDA 的图像标注概

到稿日期:2018-05-30 返修日期:2018-09-20

张 蕾(1993-),女,硕士生,主要研究方向为计算机视觉与机器学习, E-mail: 289253808@qq.com; 蔡 明(1962-),男,高级工程师,硕士生导师,主要研究方向为网络安全与信息安全、计算机网络, E-mail: mcai@jiangnan.edu.cn (通信作者)。

率模型和融入类别信息的图像标注概率主题模型。孙登第等^[20]提出了关于图标标注的标签丰富算法。

从上述图像标注模型的相关研究可以发现:基于概率模型的标注通常将两个模态的主题视为相互独立的,并没有考虑同一图像集合中不同模态主题的相关性,或者考虑到了视觉和文本模态的关联,但是两者之间的关联度较弱;其次,没有考虑到文本标注词之间的语义相关性或者文本标注词语义关联相对隐式。为此,本文提出基于主题融合和关联规则挖掘的图像标注模型。首先,在训练图像集中用 LDA^[9,18-19]模型分别获得视觉模态和文本模态的语义主题分布,并对视觉模态和文本模态数据进行加权主题融合,使得视觉和文本信息相互关联。针对视觉和文本模态信息关联度不高的问题,本文引入基于 SVM^[21]的多类别分类器进行多类别分类,以得到类别信息。依据类别信息和文本模态的主题分布,计算出图像类的文本模态主题分布,未知图像依据类别信息和图像类的文本模态的主题分布得到其文本主题分布,然后将其文本主题分布与其视觉主题分布进行主题加权融合,从而增强视觉和文本信息的关联度。然后,以主题融合之后的 LDA 模型计算未知图像的初始语义标注集。针对文本间语义关联度不高的问题,在文本信息中使用关联规则挖掘算法^[22]和词间相关性^[20],进一步挖掘文本的潜在语义关联,从而获得未知图像最终精确化的语义标注结果。在数据集上进行实验,实验结果表明本文方法有效。

2 LDA 模型

2.1 LDA 原理

LDA^[9,18-19]模型是一个文档生成模型,其概率图模型如图 1 所示。模型中用到的符号及其意义如表 1 所列。

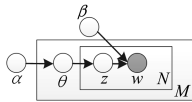


图 1 LDA 模型的概率图模型

Fig. 1 Probabilistic graphical model of LDA model

表 1 图 1 中的符号及其意义

Table 1 Symbols and their meanings in Fig. 1

符号	符号意义	符号	符号意义
M	训练集大小	N	词汇个数
K	主题数	w	单词
z	潜在主题	θ	主题比例
α	模型参数	β	模型参数
γ	α 变分参数	φ	β 变分参数
P_{dir}	狄利克雷分布	$Mult$	多项式分布

具体的生成过程如下:

- (1) 抽取一个主题分派 $\theta \sim P_{dir}(\alpha)$ 。
- (2) 对于文档 d_i 的每个单词 $w_n, n \in (1, \dots, N)$:

1) 抽取主题分派 $z_n | \theta \sim Mult(\theta)$;

2) 生成单词 $w_n | z_n \sim Mult(\beta_{z_n})$ 。

其潜变量和可观测变量的联合分布为:

$$P(w, \theta, z | \alpha, \beta) = P(\theta | \alpha) (\prod A) \quad (1)$$

其中, $A = P(z_n | \theta) P(w_n | z_n, \beta)$ 。

2.2 变分 EM 算法

本文模型采用变分 EM^[11,23]算法求解参数,此算法分为 E, M 两个步骤。为了方便计算引入的变分参数 $\Delta = \{\gamma, \varphi\}$, 变分分布函数为:

$$q(\theta, z | \Delta) = q(\theta | \gamma) \prod_{n=1}^N q(z_n | \varphi_n) \quad (2)$$

E 步骤:利用初始化的模型参数求解最优的变分参数 γ 和 φ , 其中 γ_i 表示指定文档的第 i 个主题的概率。

$$\gamma_i = \alpha_i + \sum_{n=1}^N \varphi_{n,i} \quad (3)$$

$$\varphi_{n,i} \propto \beta_{i,w_n} \exp(\psi(\gamma_i)) \quad (4)$$

M 步:以 E 步骤获得的变分参数 γ 和 φ 来求模型参数 α 和 β 。固定变分参数 $w_d = \{\gamma, \varphi\}, d \in (1, \dots, M)$, 来最大化集合 D 上的 log 似然函数。

$$L(D) = \sum_{d=1}^M \log P(w_d | \alpha, \beta) \quad (5)$$

$$\beta_{i,j} \propto \sum_{d=1}^M \sum_{n=1}^N \varphi_{d,n,i} w_{d,n}^j \quad (6)$$

参数 $\beta_{i,j}$ 表示第 i 个主题下第 j 个词的概率。对于模型参数 α , 本文没有优化 α , 通过多次实验直接将 α 设置为 1, 此时模型的性能最好。重复迭代 E 步骤、 M 步骤, 直到 $L(D)$ 收敛。

3 图像语义标注建模与实现

3.1 改进加权主题融合

LDA^[9,18-19]主题模型建立了相互独立的视觉模态和文本模态的数据,在训练集中进行视觉模态和文本模态的主题融合加强了视觉词与文本词的联系。但是, LDA 概率模型计算未知图像的标注词时,是以视觉词的主题分布作为图像的主题分布,因此图像的视觉和文本的信息联系较弱。为此,将未知图像所属图像类的文本主题分布作为其文本模态的主题分布,然后对其进行加权主题融合。相同类别中图像的文本主题分布相似,依据图像的分类信息对属于同一图像类的所有图像的文本主题分布进行融合,可以得到图像类的文本主题分布。

假设视觉模态和文本模态的主题在同一主题空间,为了增强两个模态的关联度,对未知图像进行加权主题融合。改进后融合主题分布的主题数与视觉、文本模态的主题数相等,即所有主题分布的主题数都是 K 。这里 S 和 T 分别表示视觉模态、文本模态的主题, Z 表示加权融合后的主题, D 表示训练图像集,则视觉模态的主题分布为 $P_v(S_k | D)$, 文本模态的主题分布为 $P_w(T_k | D)$, 融合后的主题分布为 $P(Z_k | D)$ 。融合主题分布的公式为:

$$P(Z_k | d_i) = \tau P_v(S_k | d_i) + (1 - \tau) P_w(T_k | d_i) \quad (d_i \in D, k \in [1, K]) \quad (7)$$

其中, τ 表示图像 d_i 中视觉模态主题分布的权重, τ 由式(8)计算:

$$\tau = \begin{cases} 1, & H(v(d_i)) \leq N \\ \exp(N - H(v(d_i))), & H(v(d_i)) > N \end{cases} \quad (8)$$

其中, $H(v(d_i))$ 表示图像 d_i 视觉词分布的信息熵, N 表示 $H(v(d_i))$ 的上界, 由多次实验得到。

3.2 关联规则算法和词间相关性

基于 LDA 概率主题模型对未知图像进行标注, 选择概率最大的多个标注词作为其标注结果, 这种方法没有涉及文本标签的语义关联, 使得最终的标注结果不太准确。本文使用关联规则挖掘算法^[22]和词间相关性^[20]计算文本关联度, 以此对初始标注集进行优化来得到待标注图像的精确化标注集合。

本文采用 FP-Growth 算法^[22]对已标注的图像的文本词进行挖掘, 以此得到标注词的频繁项集和关联规则, 并利用它对初始标注词进行扩展以得到候选标注词集。在训练图像标注集中, 如果两个或者多个语义标注词在多张图像上共同出现, 那么这些标注词是高度相关的, 即共同出现的标注词在语义上存在关联。FP-Growth 算法对初始标注进行扩展得到候选标注词集合, 但它们不能全部作为未知图像集的标注结果, 需要根据词间相关性来计算候选标签集中标注词的概率, 并将其与初始标注词集的概率相融合, 从而得到更好的标注结果。计算词间相关性^[20]的计算公式为:

$$C_{w_u, w_v} = \frac{Num_{w_u, w_v}}{Num_{w_u}} \quad (9)$$

其中, Num_{w_u, w_v} 表示标注词 w_u 和标注词 w_v 共同出现的次数, Num_{w_u} 表示标注词 w_u 在训练标注集中出现的次数。根据式(9)可以建立一个以标注词个数为行和列的矩阵, 若矩阵中某一个值越大, 那么行和列分别对应的两个标注词在同一个图像中的可能性越大。扩展标注集的词概率的计算公式为:

$$P(w_{new}) = C_{w_u, w_v} P(w) \quad (10)$$

其中, $P(w)$ 是初始标注的标注词 w 在图像出现的概率, w_{new} 为候选标注集中的标注词。根据式(10)计算图像候选标注词的概率, 并将候选标注集的标注词进行重新排序, 选择概率最大的若干个标注词作为图像的最终标注。

3.3 算法描述

下面给出基于 LDA 建模的算法的描述。假设图像和标注词的训练集为 $D = \{(d_1, c_1), \dots, (d_M, c_M)\}$, 其中训练图像集为 $T_D = \{d_1, \dots, d_M\}$, 图像 $d_i \in T_D$, 标注词汇表为 $L = \{w_1, \dots, w_L\}$, c_i 表示图像的标注词集合, 其中 $c_i \subset L, i \in \{1, \dots, N\}$ 。测试图像集用 T_T 表示, 而且 $T_D \cap T_T = \emptyset$ 。

训练算法描述如下:

(1) 处理训练图像, 对于训练图像 $d_i \in T_D$, 提取视觉特征并进行聚类得到视觉词, 然后使用 BOW^[11]模型量化图像并得到其视觉词 $v(d_i)$; 处理图像 d_i 的标注词集合 c_i , 得到文本词 $w(d_i)$ 。

(2) 基于视觉词和文本词, 分别使用 LDA 概率主题模型建模得到两个模态的模型参数, 即视觉模态的主题分布 $P_v(S|d_i)$ 和主题词分布 $P_v(v|S)$, 以及文本模态的主题分布 $P_w(T|d_i)$ 和主题词分布 $P_w(w|T)$ 。

(3) 进行主题融合, 引入参数 τ 来表示图像 d_i 中视觉模

态主题分布的权重, 使用式(7)融合视觉模态的主题分布和文本模态的主题分布, 得到新的主题分布 $P(Z|d_i)$ 。

(4) 根据融合后的主题分布 $P(Z|d_i)$ 和 EM 变分算法得到主题融合后的主题视觉词分布 $P(v|Z)$ 和 $P(w|Z)$ 。

(5) 根据训练图像集的视觉词对 SVM 分类器进行训练, 得到类别信息; 由已标注图像集类别信息和文本主题分布计算图像类的文本模态的主题分布, 即对同一图像类的训练图像的文本模态的主题分布进行融合。

完成了标注模型和 SVM 分类器的训练后, 其获得的参数和分类器对训练集之外的图像同样有效。记未知图像 $d_i \notin T_D$, 未知图像集 D_{new} , 则初始标注算法为:

(1) 对于未知图像集 D_{new} 进行处理, 其步骤与训练算法步骤(1)一样, 从而得到视觉词 $v(d_{new})$ 。

(2) 根据训练算法得到的 LDA 模型、视觉词 $v(d_{new})$ 和主题融合后的主题视觉词分布 $P_v(v|Z)$, 计算出未知图像视觉模态的主题分布 $P_v(S|d_{new})$ 。

(3) 根据训练算法步骤(5)得到的训练好的分类器对未知图像进行分类。依据未知图像的分类结果和图像类的文本模态的主题分布, 可以推测得到未知图像的文本模态的主题分布 $P_w(T|d_{new})$ 。

(4) 对未知图像集 D_{new} 进行主题融合, 使用式(7)计算未知图像集的融合主题分布 $P_w(Z|d_{new})$ 。

(5) 计算未知图像中所有标注词 L 的后验概率, 然后选取其中概率的一组标注词作为未知图像的初始标注, 其计算公式为:

$$P(w|d_{new}) = \sum_{k=1}^K P(w|Z_k) P(Z_k|d_{new}) \quad (11)$$

优化算法将未知图像的初始标注词、训练图像的标注词集合等作为输入, 具体步骤为:

(1) 对训练图像 d_i 的标注词集合进行处理, 得到文本词 $w(d_i)$;

(2) 使用式(9)对训练集的全部文本词进行计算, 得到词间相关性;

(3) 根据训练集的标注词集合和 FP-Growth 算法找出频繁项集, 扩展待标注图像的初始标注词以得到候选标注集;

(4) 依据初始标注词的后验概率和词间相关性, 使用式(11)计算所有候选标注集的词概率, 选出概率最大的 5 个标注词, 从而得到待标注图像的精确标注词。

4 实验结果分析

本文采用的数据集是 Corel5k^[11] 图像集, 该图像集包含 5000 张图像, 共有 50 个语义类别, 每个语义类别包含 100 张相似的图像, 每幅图像有 1~5 个标注词。Corel5k 图像集共有 371 个标注词, 将标注图像较少的标注词剔除, 其他的 260 个单词构成标注词汇表。实验选取 Corel5k 中的 4500 幅图像作为训练图像集, 将剩下的 500 幅图像作为测试图像集。实验提取的底层视觉特征^[24-25]包括 128 维的 SIFT^[11]特征向量; 将图像划分为 16×16 的规则方块, 并为每个方块提取

256 维的 HSV 颜色直方图特征和 256 维的 LBP 纹理特征,将颜色特征和纹理特征融合形成 256 维的方块特征。最后对视觉特征进行 K-means 聚类算法得到对应的视觉词,把图像表示成视觉词的词袋。

4.1 标注模型参数

在本实验中将 SIFT^[11]特征和方块特征的视觉词都设置为 1000,即将一幅图像表示为 2000 维的直方图;影响图像标注效果的模型参数有两个,分别是潜在主题个数和主题融合信息熵上界 N 。设置标注模型中视觉模态和文本模态的潜在主题个数为 K , K 取值从 60 到 140,步长为 20,在不同 K 值下的平均召回率、平均准确率如图 2 所示。从图 2 可以看出:当视觉模态和文本模态的潜在主题 K 为 120 时,性能指标达到最优。视觉词分布的信息熵上界 N 分为训练集 N_{train} 和测试集 N_{test} ,对 N_{train} 进行实验时,将 N_{test} 固定为 1,然后设置 N_{train} 从 1 增加到 10,步长为 1;对 N_{test} 进行实验时,将 N_{train} 固定为 3,然后设置 N_{test} 从 1 增加到 10,步长为 1。训练集和测试集在视觉词信息熵上界的联合交叉验证下平均召回率如图 3 所示。从图 3 中可以看出,当 $N_{\text{train}}=3, N_{\text{test}}=1$ 时,数据集的召回率达到最优,实验效果最佳。

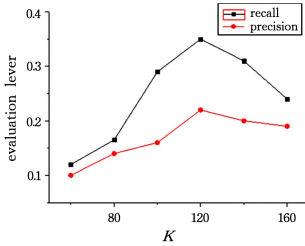


图 2 不同 K 值下的评价曲线

Fig. 2 Curves of recall with different value of K

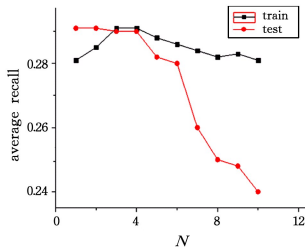


图 3 不同 N 值下的召回率曲线

Fig. 3 Curves of recall with different value of N

4.2 性能评价指标

实验采用召回率、准确率和综合评价指标 F 值作为评价图像自动标注模型性能的评价指标。对于给定的标注词 w ,其准确率为 $precision(w)$ 、召回率为 $recall(w)$ 、 F 值为 $F(w)$ 。这 3 种评价指标的计算公式为:

$$precision(w) = \frac{B}{A} \quad (12)$$

$$recall(w) = \frac{B}{C} \quad (13)$$

$$F(w) = \frac{2 * recall(w) * precision(w)}{recall(w) + precision(w)} \quad (14)$$

其中, A 表示所有自动标注中有 w 的图像个数; B 表示正确标注 w 的图像个数,即图像的原始标注和自动标注中都包含有标注词 w ; C 表示图像原始标注中包含 w 的图像个数。此外,还考虑了召回率大于 0 的标注词的个数,即表示有效学习的标注词的个数。

4.3 标注结果

为了验证本文提出的标注模型的有效性,将本文模型与传统的图像自动标注方法以及近年常用的一些图像标注方法如 CMRM^[4], CRM^[5], PLSA-FUSION^[11], PLSA-GMM^[13], PLSA-GMMVS^[14] 进行对比。表 2 列出了上述所有标注方法在图像集 Corel5k 中具有最佳性能的 49 个标注词构成的子集和由标注词表中 260 个标注词构成的集合上,标注性能的评价结果。从表 2 的 260 个标注词构成的子集数据中可以看出,本文的标注方法在平均召回率、平均准确率和 F -值等性能指标上比其他模型的最优者分别高出了 12%, 2%, 5.5%。



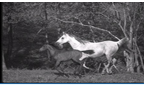

表 3 列出了 4 幅图像的自动标注结果与图像的原始标注词、PLSA-FUSION 模型的对比实例结果。本文的自动标注结果是从优化的精确化标注词中选取概率最大的 5 个关键词,对于视觉词比较稀疏的图像(如第二幅图像),本文的标注模型要优于 PLSA-FUSION,即自动标注的标注词不仅包含大部分原始标注词,也包含其他没有出现在原始标注的单词,在某种意义上也符合图像内容的标注(如第二幅图像的标注词 birds 和 waved)。对于视觉词比较复杂的图像,本文的标注模型的优势并不明显,如第四幅图中的“boat”。分析原因,在本文主题模型下每个主题包含了多个相关语义的标注词,精确化标注时容易将与关键词紧密相关的标注词插入包含此关键词的所有图像中,造成部分标注不准确,但整体效果较好。

表 2 图像标注的性能比较

Table 2 Performance comparison of image annotation

Model	CMRM	CRM	PLSA-FUSION	PLSA-GMM	PLSA-GMMVS	本文模型
# words with recall>0	66	107	112	104	108	131
Result on 49 best words						
Mean per-word Recall	0.48	0.70	0.76	0.67	0.72	0.98
Mean per-word Precision	0.40	0.59	0.65	0.52	0.55	0.71
F -value	0.436	0.64	0.70	0.58	0.62	0.82
Result on all 260 words						
Mean per-word Recall	0.09	0.19	0.22	0.18	0.21	0.35
Mean per-word Precision	0.10	0.16	0.19	0.15	0.20	0.22
F -value	0.095	0.174	0.204	0.163	0.204	0.270

表3 标注结果比较
Table 3 Annotation results comparison

Image	Ground truth	PLSA-FUSION	本文模型
	bear, polar, snow, tundra	polar, bear tundra, snow, ice	snow, tundragrizzly, polarbear
	sky, flight, albatross	city, flight ceremony, pond swallow-tailed	birds, sky, flight, waved albatross
	field, horses, mare, foals	tree, garden, house, mare, foal	foals, horsesmare, fence, tree
	mountain, water, sky, clouds	mountain, clouds, boat, coast, hut	water, reflection, boat cloud, mountain

4.4 模型时间复杂度分析

利用 LDA 概率主题模型计算标注词所需要的时间为 $T(n) = N_{iter} \cdot M \cdot K \cdot l_{v+w}$, 其中 N_{iter} 为模型迭代次数, M 为训练图像的数量, K 为潜在主题数, l_{v+w} 为训练图像的视觉单词数和文本单词数的和。由于文本单词数远远小于视觉单词数, 因此 l_{v+w} 由图像的视觉单词数 l_v 替代, 而 $N_{iter} \ll M \cdot l_v$, $K \ll M \cdot l_v$, 故计算标注词生成概率的复杂度为 $O(M \cdot l_v)$ 。基于支持向量机的多分类方法中, 计算图像类别信息的时间为 $T(n) = N_c \cdot M \cdot l_v$, 其中 N_c 为支持向量机的个数, 即图像类的个数, 由于 $N_c \ll M \cdot l_v$, 因此本文模型结合支持向量机的图像标注方法的复杂度为 $O(M \cdot l_v + M \cdot l_v)$ 。在基于关联规则挖掘和词间相关性的图像标注中, 计算所有文本词关联度的时间为 $T(l_w^2)$, 迭代累积计算前 k 个标注词关联度的时间为 $T(k \cdot l_w^2)$, 计算关联规则挖掘多的时间为 $T(M \cdot l_w)$ 。基于主题融合和关联规则挖掘的图像自动标注方法所需要的总时间为 $T(M \cdot l_v + M \cdot l_v + l_w^2 + k l_w^2 + M l_w)$, 由于 $k \ll l_w^2$, 因此该方法的时间复杂度为 $O(M l_v + l_w^2 + M l_w)$ 。最后, 相比于 LDA 概率主题模型, 本文提出的方法的时间复杂度增加了 $O(l_w^2 + M l_w)$ 。

结束语 本文提出基于主题融合和关联规则挖掘的图像标注模型, 首先构建两个 LDA 模型来分别学习视觉模态和文本模态的数据, 为了加强视觉和文本信息关联, 对训练集和测试集的两个模态主题分布都进行主题融合, 再使用主题融合后的训练模型来计算未知图像的初始标注。基于关联规则挖掘和词间相关性进一步挖掘初始标注词之间的关系, 以获得未知图像的精确标注。通过图像集 Corel5k 上的对比实验可以看出, 本文模型的标注性能和效果比大多数经典模型都优秀。

本文通过对图像进行量化得到了离散的视觉词, 在此过程中丢失了大量的视觉信息, 影响了图像的标注性能。其次, 优化初始标注词是建立在测试集的原始标注的关联之上的, 但从单一的数据集获得的词间关系具有局限性, 有时候可能会得到错误的关联, 从而降低标注性能。此外, 本文的标注模型虽然在图像标注的性能上有所提升, 但其时间复杂度也有所上升。下一步的工作准备对模型进行改进, 使之可以通过

连续建模来避免视觉信息丢失, 另一方面引入 WordNet 来计算词间关系, 以获得更好的标注性能, 同时也要关注时间复杂度的优化问题。

参考文献

- [1] GU Y, XUE H Y, YANG J. Cross-modal saliency correlation for image annotation[J]. *Neural Processing Letters*, 2017, 45(3): 777-789.
- [2] HE C, CHEN Z X, LIU C Y. Bottom-up image saliency target detection via bottomup[J]. *Journal of Optoelectronic · Laser*, 2016, 27(2): 886-892. (in Chinese)
贺超, 陈振学, 刘学云. 自底向上的图像显著目标检测研究[J]. *光电子·激光*, 2016, 27(2): 886-892.
- [3] DUYGULU P, BARNARD K, DE FREITAS J, et al. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary[C]// *Proceedings of the 7th European Conference of Computer Vision*. Copenhagen, Kongeriget, Danmark: ECCV, 2002: 97-112.
- [4] 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*. 2003: 119-126.
- [5] 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.
- [6] FENG S L, MANMATHA R, LAVRENKO V. Multiple Bernoulli relevance models for image and video annotation [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Washington: IEEE Computer Society, 2004: 1002-1009.
- [7] MORALES-GONZALEZ A, GARCIA-REYES E, SUCAR L E. Unsupervised segmentation evaluation for image annotation[C]// *10th International Conference on Computer Vision Theory and Application*. Berlin, Germany, 2015.

- [8] HOFMANN T. Unsupervised learning by probabilistic latent semantic analysis[J]. *Machine Learning*, 2001, 42(1-2): 177-196.
- [9] BLEI D M, NG A Y, JORDAN M I. Latent Dirichlet allocation [J]. *Journal of Machine Learning Research*, 2003, 3: 993-1022.
- [10] 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.
- [11] 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. (in Chinese)
李志欣, 施志平, 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.
- [12] LI Z X, SHI Z P, ZHANG C L, et al. Hybrid generative/discriminative model for automatic image annotation [J]. *Journal of Image and Graphics*, 2015, 20(5): 687-699. (in Chinese)
李志欣, 施志平, 张灿龙, 等. 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): 687-699.
- [13] ZHAO P, WANG W B, ZHU W W, et al. Automatic image annotation by combining aspects and visual semantics[J]. *Journal of Computer Aided Design & Computer Graphics*, 2013, 25(11): 1709-1714. (in Chinese)
赵鹏, 王文彬, 朱伟伟, 等. 融合主题和视觉语义的图像自动标注方法[J]. *计算机辅助设计与图形学学报*, 2013, 25(11): 1709-1714.
- [14] SUN J D, LI H H, JIN J L, et al. Image automatic annotation based on visual semantic topics [J]. *Measurement & Control Technology*, 2016, 35(12): 11-15. (in Chinese)
孙君顶, 李海华, 靳姣林, 等. 基于视觉语义主题的图像自动标注[J]. *测控技术*, 2016, 35(12): 11-15.
- [15] SUN J D, LI H H, JIN J L, et al. Image Automatic annotation based on multi-feature fusion and PLSA-GMM [J]. *Measurement & Control Technology*, 2017, 36(4): 31-35. (in Chinese)
孙君顶, 李海华, 靳姣林, 等. 基于多特征融合与 PLSA-GMM 的图像自动标注[J]. *测控技术*, 2017, 36(4): 31-35.
- [16] SUN J D, LI H H, JIN J L. Image automatic annotation based on the visual semantic topics and feedback log[J]. *Journal of Optoelectronics. Laser*, 2017, 28(4): 441-450. (in Chinese)
孙君顶, 李海华, 靳姣林, 等. 基于视觉语义主题与反馈日志的图像自动标注[J]. *光电子·激光*, 2017, 28(4): 441-450.
- [17] TIAN D P. Integrating PLSA and random walk model for automatic image annotation [J]. *Journal of Chinese Computer Systems*, 2017, 38(8): 1899-1907. (in Chinese)
田东平. 融合 PLSA 和随机游走模型的自动图像标注[J]. *小型微型计算机系统*, 2017, 38(8): 1899-1907.
- [18] CAO J, LOU J X, LI X X. Image annotation probabilistic topic model improving corr-LDA model [J]. *Journal of Chinese Computer Systems*, 2017, 38(3): 615-619.
- [19] CAO J, LOU J X, LI X X. Image annotation probabilistic topic model fusing class information [J]. *Computer Engineering and Applications*, 2017, 53(10): 187-192.
- [20] SUN D D, GE M L, DING Z L, et al. Tagging enrichment algorithm based on tag semantic and image visual [J]. *Journal of Chinese Computer Systems*, 2017, 38(4): 886-890. (in Chinese)
孙登第, 葛美玲, 丁转莲, 等. 基于标注词语义与图像视觉的标签丰富算法[J]. *小型微型计算机系统*, 2017, 38(4): 886-890.
- [21] CHANG C C, LIN C J. LIBSVM: A library for support vector machines [J]. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2011, 2(3): 27.
- [22] HAN J, PEI J, YIN Y. Mining frequent patterns without candidate generation [C] // *Proceedings of the ACM SIGMOD International Conference on Management of Data*. USA, 2000: 1-12.
- [23] GAO Y Y, YIN Y X, UOZUMI T. A hierarchical image annotation method based on SVM and semi-supervised EM [J]. *Acta Automatica Sinica*, 2010, 36(7): 960-967.
- [24] ZHOU S J, MENG J, HUANG Z P, et al. A method for discrimination of processed ginger based on image color feature and support vector machine model [J]. *Analytical Methods*, 2016, 8: 2201-2206.
- [25] PONTI M, NAZARE T S, THUME G S. Image quantization as a dimensionality reduction procedure in color and texture feature extraction [J]. *Neurocomputing*, 2016, 173: 385-396.