基于 CNN 和加权贝叶斯的最近邻图像标注方法

王 琳,张素兰*,杨海峰
(太原科技大学 计算机科学与技术学院,山西 太原 030024)

摘 要:图像标注的准确性在很大程度上关系着图像检索的准确性。然而,传统的基于最近邻模型的图像自动标注方法不能有效提取图像底层特征,并且无法有效建立低级视觉特征到高级语义之间的映射关系,使得近邻图像搜索不准确从而影响图像标注的准确性。针对上述问题,提出了一种改进的基于 CNN 和加权贝叶斯的最近邻图像标注方法。首先,利用卷积神经网络(convolutional neural networks,CNN)提取图像特征,并依此特征搜索其近邻图像,构建候选标签集合;然后利用贝叶斯后验概率构建待标注图像的视觉特征与标签之间的映射关系;最后通过设定权重优化概率值并排序,得到最优的候选标签进而实现图像标注。在三个基准数据集 Corel 5K,IAPRTC-12 和 ESP Game 上进行实验,结果表明该方法在准确率、召回率与 F1 值上均取得了较好的效果。

关键词:图像自动标注;最近邻模型;映射关系;卷积神经网络;贝叶斯后验概率
中图分类号:TP391 **文献标识码**:A **文章编号**:1673-629X(2021)10-0063-07
doi:10.3969/j.issn.1673-629X.2021.10.011

A Nearest Neighbor Image Annotation Method Based on CNN and Weighted Bayesian

WANG Lin,ZHANG Su-lan*,YANG Hai-feng
(School of Computer Science and Technology,Taiyuan University of Science and Technology,Taiyuan 030024,China)

Abstract:The accuracy of image annotation is related to the accuracy of image retrieval to a great extent. However,the traditional image automatic annotation method based on the nearest neighbor model cannot effectively extract the features of the image,and the mapping relationship between low-level visual features and high-level semantics cannot be effectively established either. So the accuracy of the nearest neighbor images and labels is low. To solve the above problems,a neighbor image annotation method based CNN and weighted Bayesian is proposed. First,image features are extracted by using CNN,from which the nearest neighbor images are searched and the candidate labels are obtained. Then the mapping relationship between visual features and labels of the unlabeled image based on Bayesian posterior probability is constructed. Finally,the probability value is optimized and sorted according to set the weight,from which the optimal candidate labels are selected to realize image annotation. Experiments on three benchmark datasets Corel 5K, IAPRTC-12 and ESP Game show that the proposed method achieves better results in precision,recall and F1 value.

Key words: automatic image annotation; nearest neighbor model; mapping relationship; convolutional neural networks; Bayesian posterior probability

0 引 言

随着智能科技和互联网的快速发展,图像资源信息迅猛增长,如何对图像进行有效自动标注以提高图像检索的准确性仍是计算机视觉领域的重要研究内容。然而,由于人工图像标注的主观性和不可靠性,使得人们对于同一幅图像有不同的理解,造成图像标注的语义内容和标签不符,影响了图像检索的准确性。而且,人工给海量图像进行标注也很不现实。因此,目前仍有不少研究人员致力于图像语义自动标注(automatic image annotation,AIA)[1-3]模型和方法的研究工作,主要利用人工智能、模式识别和机器学习等方法,对图像内容进行语义解释,从而使计算机可以自动获取图像的语义信息,帮助人们更有效地进行图像检索。

其中,Cheng[4]等将图像自动标注方法主要分为生成模型[5]、判别模型[5]、标签补全[6-7]、深度学习[1,8]

收稿日期:2020-11-20 **修回日期**:2021-03-23
基金项目:国家自然科学基金项目(U1731126)
作者简介:王 琳(1996-),女,硕士研究生,研究方向为机器学习与图像语义标注;通讯作者:张素兰(1971-),女,教授,博士,CCF 会员(66965M),研究方向为数据挖掘、机器学习与计算机视觉。

和最近邻模型[9-14]等几种。由于大规模网络图像的出现,以及人们通常直观地假设相似图像可能含有共同标签使得基于最近邻模型的图像自动标注方法一直深受研究者的关注。Su 等[9]提出了一种基于图学习的最近邻图像自动标注方法,该方法考虑了标签之间的相关性并将图像到标签之间的距离与基于图学习的分数相结合来获得标签的决策值。虽然该方法在一定程度上提高了图像标注的性能,但其采用全局特征和局部特征进行图像特征提取,过程比较复杂且提取到的图像特征的辨识度不高。Verma 等[10]在 2PKNN[11]的基础上进行了改进,将图像-标签之间的相似性与图像-图像之间的相似性结合起来,并提出了一种度量学习框架,该方法利用了先进的特征提取、编码以及嵌入技术从而提高了标注性能。Jin 等[12]为了弥合“语义鸿沟”,提出了一种基于图像距离度量学习的邻域集(NSIDML)方法,不限制样本是否带标签,充分利用现有资源,进而提高图像自动标注性能。但该方法因没有充分考虑图像视觉特征与标签之间的概率关系,一定程度上影响了图像标注性能。Rad 等[13]利用松弛联合非负矩阵分解(LJNMF)对图像的高维特征进行降维,然后再计算图像特征间的距离,最后依据距离权重进行标签传播实现图像自动标注。柯逍等[14]先基于深度特征从视觉和语义两个方面构建近邻图像,然后根据距离计算标签概率实现图像标注,该方法在一定程度上提高了图像标注的性能,但是没有分析图像视觉特征与标签之间的依赖关系。总之,尽管这些基于最近邻模型的图像自动标注方法取得了一定的效果,但是低效复杂的特征提取方式以及没有充分考虑图像低层视觉特征到高级语义之间的依赖关系使得图像标注性能仍有待提升。

近年来,卷积神经网络(convolutional neural networks,CNN)在图像处理领域得到了很好的应用[15-17],该模型可以直接将原始图像输入到网络中,不需要预先对图像进行复杂处理,并且可以自动提取图像特征,随着训练过程的深入,能够提取出更具有辨识度的图像特征,具有较强的表达能力。随着卷积神经网络的深入研究,越来越多的学者将卷积神经网络应用于图像自动标注中。如高耀东等[18]利用卷积神经网络进行自主学习图像特征,并改进损失函数从而改善输出结果。除了考虑高效的图像特征提取方式,如何有效建立图像与标签之间的某种关系是图像自动标注中需要解决的关键问题,而贝叶斯在不完全信息下对未知状态进行概率估计的理论特性,可以在已知图像的条件下构建图像特征和标签之间的概率分布,从而可以找出图像低层视觉特征与高级语义之间的概率关系,缩小语义鸿沟。如 Verma 等[19]利用贝叶斯后验概率找寻给定样本和标签的 K_1 个最近邻,并根据与邻居的距离计算标签置信度预测标签,有效地提高了图像标注性能。因此,为改善传统的基于最近邻模型图像自动标注方法在图像底层视觉特征提取和视觉特征与标签之间映射关系的不足,文中提出了一种改进的基于 CNN 和加权贝叶斯的最近邻图像标注方法,进一步提高了图像自动标注性能。

1 相关工作

1.1 最近邻模型的图像标注方法

最近邻模型的图像自动标注方法认为若图像有相似的底层视觉特征,则有相似的语义标签。因此,最近邻模型的图像自动标注方法的一般步骤为:(1)构建图像的底层特征;(2)根据图像的底层视觉特征,利用距离度量方法找寻待标注图像的近邻图像;(3)利用合适的标签传播方法,将近邻图像中的标签传播给待标注图像。

现有的基于最近邻模型的图像自动标注方法的改进基本包括三个方面:(1)构造不同的视觉特征用以提高图像标注性能,比如提取图像的 SIFT 特征、HOG 特征、进行特征融合等等;(2)选取不同的距离度量策略,比如欧氏距离、谷歌距离等等;(3)采用优化的标签传播算法,使得图像的标签可以更好地传播。基于最近邻模型的图像标注方法的代表性模型有 JEC[10]、TagProp[10]、GLKNN[9]、2PKNN[11]等等。

1.2 卷积神经网络

卷积神经网络[20]是一种在深度神经网络基础上提出来的多层感知机,相当于一个图像的特征提取器,被广泛应用于计算机视觉领域。CNN 的主要特点是可以在神经元之间进行局部连接和权值共享,并且在一定程度上可以进行图像的平移、旋转、倾斜和尺度不变性等操作,还可以同时完成图像的特征提取以及特征分类,用来提取图像特征十分高效。CNN 的主要结构为输入层、卷积层、池化层、全连接层以及输出层,经过卷积层和池化层操作提取图像的视觉特征图,再通过全连接层将卷积结果与图像全连接,根据权重计算输出结果,以达到提高表达能力的目的。

1.2.1 卷积层

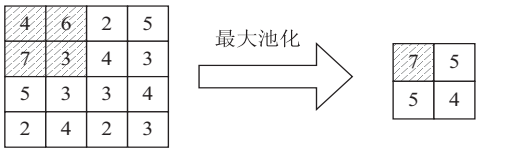
CNN 在卷积层进行特征的局部感知和参数共享,然后通过不同的卷积核和图像像素值进行对应卷积运算得到图像的特征映射,从而提取出图像的视觉特征。这一层也是整个卷积神经网络的核心层,提取出图像特征后,以特征图的形式表示图像特征。其表达式如式(1):

$$a_{i,j}=f\left(\sum_{m=0}^{2}\sum_{n=0}^{2}\omega_{m,n}\times x_{i+m,j+n}+\omega_b\right) \tag{1}$$

其中，$a_{i,j}$ 表示第 i 层的第 j 个卷积核对应的特征值,对卷积核的每个权重进行编号，$\omega_{m,n}$ 表示卷积核的第 m 行第 n 列权重，ω_b 表示卷积核的偏置项，× 表示卷积运算，$f(\cdot)$ 表示激活函数(此处用 Relu 函数)。为了简化操作和复杂数据,Relu 对卷积操作得到的结果进行非线性激活响应,舍弃不相关数据(值小于0 的数据改写为 0)。

1.2.2　池化层

卷积过程中采用多个卷积核进行卷积操作,会使得信息冗余,因此为了减少数据量,降低计算量,减少机器负载,要进行降维也就是池化操作。CNN 的池化层对卷积层的特征向量图进行下采样操作,依据特征图的局部相关原理将卷积层处理图像时产生的冗余信息减少,保留图像的重要信息。现如今常用的池化操作有平均池化和最大池化等,最大池化是将对应区域内神经元的最大值代替该区域进行输出,从而在保留图像特征信息的同时完成数据降维。因此,文中采用大小为 2×2 的池化核进行最大池化,示意图如图 1 所示。



(1) 卷积后的特征图　　(2) 池化后的特征图

图 1　池化核为 2×2 的最大池化示意图

1.2.3　全连接层与输出层

与卷积层的局部连接不同,全连接层采取全连接的思想,将卷积层和池化层的局部信息进行整合,运用 Softmax 分类函数得到每个类别对应的概率值,再传递给输出层,进而最终将特征图映射为特征向量。

2　文中方法

文中给出的改进的基于 CNN 和加权贝叶斯的最近邻图像标注方法的具体思想架构如图 2 所示。

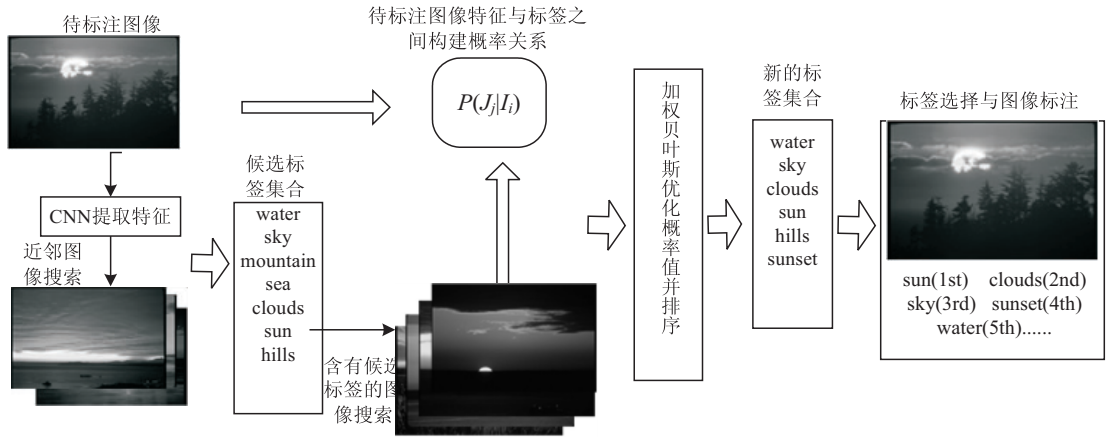


图 2　基于 CNN 和加权贝叶斯的最近邻图像标注方法架构

第一步:利用图像的 CNN 特征找寻待标注图像的近邻图像,并统计近邻图像中所含有的标签以及标签个数,构成候选标签集合。

第二步:筛选含有候选标签的图像得到图像集合,计算其视觉特征矩阵每一维的均值,利用贝叶斯后验概率公式计算候选标签与待标注图像视觉特征之间的概率值,获得候选标签标注给待标注图像的概率。

第三步:选择标签标注图像。考虑到待标注图像的近邻图像所含有的标签的频率不同,设置一个 α 系数表示标签权重。将 α 系数与第二步所得的标签概率相结合,计算新的标签概率,获得新的候选标签,从中选择概率值高的前 5 个进行标注。

2.1　基于 CNN 获取初始标签

图像自动标注是将最有可能代表图像的关键词标注给图像,那么,图像越相似,含有相同标签的可能性越大。因此,进行图像自动标注的首要步骤就是寻找待标注图像的相似图像,也就是近邻图像,其中最重要的一步就是提取图像的视觉特征。

卷积神经网络提取图像特征优势明显,叠加卷积层和池化层构建多层网络结构,并利用 Relu 函数对卷积结果进行非线性激活,将图像特征映射为 4 096 维的特征向量。文中采用图像网络大规模视觉识别挑战(ILSVRC)中提出的卷积神经网络方法[21],首先初始化卷积神经网络模型,在 1 000 类分类数据集 ImageNet 上进行预训练,并基于文献[22]的 VGG-16 模型提取 4 096 维的图像特征向量,包括 13 个卷积层,5 个最大池化层(池化核均为 2×2),2 个全连接层。具体过程如下:

(1)使用数据增广。在 256×256 的原始图像中随机选择 224×224 的区域构成输入图像,采用 ImageNet 数据集进行预训练。

(2)进行卷积池化操作。13 个卷积层均使用 3×3 的卷积核,步长设置为 1,第二个卷积层后接一次最大池化第四个卷积层后接一次最大池化,第七个卷积层

后接一次最大池化,第十个卷积层后接一次最大池化,第十三个卷积层后接一次最大池化。

(3)局部响应归一化。在卷积层中使用 LRN 进行局部响应归一化,应用在激活函数和池化函数之后,增大响应较大的值,抑制较小的值。

(4)进行全连接层计算。卷积操作和池化操作完成后,与传统的卷积神经网络接三个全连接层不同,文中接两个全连接层,将图像转化为 1×1×4 096 的输出图像,得到图像的 4 096 维特征向量,并在这两个层内使用 dropout 进行正则化,避免过拟合。

完成图像的特征提取之后,需要根据图像的 CNN 特征向量找寻待标注图像的近邻图像。使用欧氏距离计算图像之间的视觉距离,两幅图像之间的距离值越小,说明两幅图像在视觉上越相似。近邻图像个数的取值将在 3.2 小节进行分析。

2.2 基于贝叶斯构建映射关系

假设训练数据包括图像及其对应的标签集,构成对的关系,即 $\{(J_j,L_i)\}_{i,j=1}^{n}$。给定一幅图像 J,它的底层视觉特征与其所含标签之间的映射关系可以通过条件概率 $P(J\mid l_i)$ 建模[23]。则给出计算图像 J 的标签 l_i 的后验概率 $P(l_i\mid J)$:

$$P(l_i\mid J)=\frac{P(J\mid l_i)P(l_i)}{P(J)} \tag{2}$$

其中,$P(l_i)$ 是标签的先验概率,$P(J)$ 是图像的先验概率。参考文献[19],将 $P(l_i)$ 对所有标签 l_i 设置为 $P(l_i)=1/M$,其中 M 为常数;$P(J)$ 是测试数据中找寻待测图像的概率,将其设置为 1。

针对条件概率 $P(J\mid l_i)$,由于平均数可以表示图像特征矩阵中特征向量值的趋势,而且与图像特征矩阵中的每一个特征向量值都有关系,不会脱离图像特征矩阵,因此利用均值结合高斯密度给定 $P(J\mid l_i)$ 的计算公式:

$$P(J\mid l_i)=\prod_{d=1}^{4\,096}\mathrm{e}^{-(x_d-\mu_{Y_i})^2} \tag{3}$$

其中,x_d 表示待标注图像的每一维的特征向量,Y_i 表示含有共同标签 l_i 的训练数据的子集($Y_i\subseteq J$),μ_{Y_i} 表示通过含有共同标签 l_i 的图像矩阵计算得出的每一维的均值。

通过上述公式给出图像的特征与标签之间的概率关系,得出一幅图像含有某一个标签的概率值为多少,从而初步获得标签属于待标注图像的概率。

2.3 加权贝叶斯优化标签概率并标注

考虑到近邻图像与待标注图像的相似度不同,不同近邻图像含有的标签不一样,那么含有同一个标签的图像个数就不一定,即近邻图像中标签的频率不同。因此,将标签频率看作候选标签的权重,给定系数 α 计算候选标签的权重,系数计算公式为:

$$\alpha=n/G \tag{4}$$

其中,n 表示最近邻图像中标签的频数,G 表示最近邻图像包含的所有标签总数。

结合候选标签在近邻图像中的频率,将候选标签的频率值作为标签权重与 3.2 节得到的概率值相乘得到候选标签的最终概率值,并进行重排,从中选择概率值最高的 k($k=5$)个进行图像标注。

3 实验结果与分析

为了验证文中所提方法的有效性,在三个基准数据集 Corel 5K、ESP Game 以及 IAPRTC-12 上进行了实验验证。

3.1 数据集与评估指标

Corel 5K 数据集、ESP Game 数据集以及 IAPRTC-12 数据集的具体数据情况如表 1 所示。

表 1 数据集统计

数据集	Corel 5K	ESP Game	IAPRTC-12
图像数	4 999	20 770	19 627
标签数	260	268	291
训练图像数	4 500	18 689	17 495
测试图像数	499	2 081	1 957
平均标签数	3.4	4.7	5.7
标签数中值	4	5	5
最大标签数	5	15	23

实验采用准确率(precision,P)、召回率(recall,R)以及 F_1 值(F_1)三个评估指标度量实验结果,计算公式如下:

$$P=\frac{|L\cap\hat{L}|}{|\hat{L}|},R=\frac{|L\cap\hat{L}|}{|L|},F_1=\frac{2\times P\times R}{P+R} \tag{5}$$

其中,L 表示给定图像的真实标签集合,$\hat{L}$ 表示给定图像的预测标签集合,P 计算每幅图像正确找回的标签数量占实际找回的标签数量的比例;R 计算每幅图像正确找回的标签数量占图像本身标签数量的比例;F_1 值表示查准率和查全率的调和平均数,范围从 0 到 1,数值越大,效果越好。

3.2 参数分析

在文中方法中,参数 N 是待标注图像的近邻图像个数,在本小节对其进行分析。从图 3 ~ 图 5 中可以清晰地看出,在 Corel 5K 和 ESP Game 两个数据集上三个评估指标的值均是有一段先降再升到达峰值,然后再下降。当 $N=30$ 时,Corel 5K 数据集上评估指标的值到达峰值,$N=20$ 时比 $N=30$ 时的评估指标值略有降低,$N=40$ 时比 $N=30$ 时的评估指标值下降明显;当 $N=40$ 时,ESP Game 数据集上评估指标的值到

达峰值,$N=20$ 时比 $N=40$ 时的评估指标值略有降低,$N=30$ 时比 $N=40$ 时的评估指标值下降明显。在 IAPRTC-12 数据集上,三个评估指标值均在 $N=10$ 时达到峰值,然后开始下降;当 $N=20$ 时,评估指标的值比 $N=10$ 时略有降低。因此,综合三个数据集考虑,实验中设置 $N=20$。

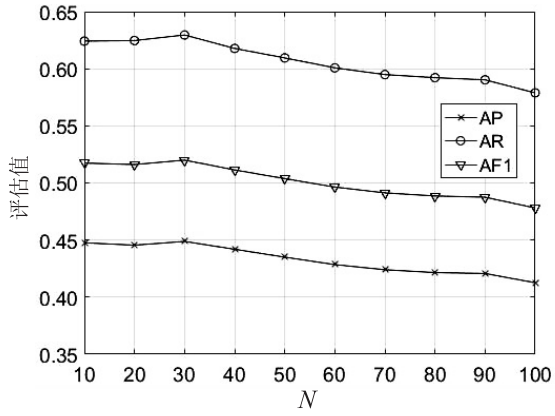


图 3　Corel 5K 数据集上评估指标的值随参数 N 的变化

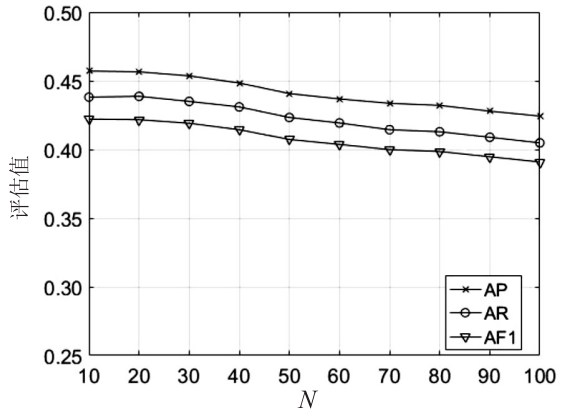


图 4　IAPRTC-12 数据集上评估指标的值随参数 N 的变化

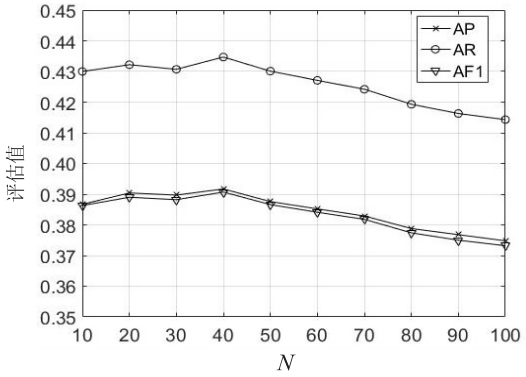


图 5　ESP Game 数据集上评估指标的值随参数 N 的变化

3.3　对比实验

由于文中采用 CNN 提取图像特征,因此分别与采用 CNN 提取特征以及采用其他方法提取特征的现有的比较好的标注方法进行对比:GLKNN[9]、CCA-2PKNN[10]、NSIDML[12]、IDFRW[24]、NL-ADA[25] 以及 OPSL[26]。

3.3.1　与采用 CNN 提取特征的方法比较

表 2 表示一些采用 CNN 提取特征的基于最近邻模型的图像自动标注的先进方法与文中所提方法的实验结果对比,其中 C 表示提取特征为 CNN 特征。CCA-2PKNN 通过使用典型相关分析(CCA)将不同特征组合,包括卷积神经网络特征、编码计算特征等等,嵌入到公共子空间从而最大化视觉内容和文本之间的相关性。IDFRW 通过集成图形的深层特征和标签相关性构建图像特征与图像语义之间的映射关系,从而提高标注性能。GLKNN 将图学习方法和最近邻方法相结合,利用图学习方法传播图上对应于测试图像的 K 最近邻标签,进一步提高标注性能。通过在 Corel 5K、IAPRTC-12 和 ESP Game 三个数据集上与传统的图像自动标注算法进行比较,比较结果如表 2 所示。

表 2　在三个数据集上的实验结果评估比较

数据集	Corel 5K			IAPRTC-12			ESP Game		
	P	R	F_1	P	R	F_1	P	R	F_1
JEC(C)[10]	26	21	23	20	09	12	38	18	24
TagProp-SD(C)[10]	21	14	17	48	21	29	52	22	31
TagProp-σSD(C)[10]	27	22	24	46	25	32	47	25	33
CCA-2PKNN(C)[10]	37	32	34	50	14	22	50	25	33
GLKNN[9]	36	47	41	41	36	38	31	34	32
IDFRW[24]	38	49	43	49	31	38	–	–	–
文中方法	45	62	52	46	44	42	39	43	39

从表 2 可以看出,在 Corel 5K 数据集上,文中方法与实验结果相对较好的 IDFRW 相比,P 提高了 7%,R 提高了 13%,F_1 提高了 9%。这是因为文中在获取近邻图像之后进行图像标签与特征的概率构建,避免了在更多数据下进行关系映射,并且近邻图像的标签与待标注图像的联系更紧密。在 IAPRTC-12 数据集

上,文中方法与查准率最高的 CCA-2PKNN 相比, P 降低了 4%, R 提高了 30%, F_1 提高了 20%;与整体实验结果较好的 GLKNN 相比, P 提高了 5%, R 提高了 8%, F_1 提高了 4%。在 ESP Game 数据集上,文中方法与查准率最高的 TagProp-SD 相比, P 降低了 13%, R 提高了 21%, F_1 提高了 8%;与整体实验结果较好的 CCA-2PKNN 相比, P 降低了 11%, R 提高了 18%, F_1 提高了 6%;与 GLKNN 相比, P 提高了 8%, R 提高了 9%, F_1 提高了 7%。这是由于在进行图像标注时,文中在近邻图像的基础上考虑了图像的视觉特征与语义之间的映射关系,进一步提高了标注性能,而以上方法均忽略了图像视觉特征与标签之间的关系。虽然查准率有所降低,但是查全率与 F_1 值均有大幅提高。

3.3.2 与采用其他方法提取特征的方法比较

图 6 ~ 图 8 表示一些采用其他方法提取特征的图像自动标注的先进方法与文中所提方法的结果对比。NL-ADA 是 Ke 等人提出的属性判别标注框架,基于未知图像构造平衡数据集,并判别图像的高频低频属性,然后标注图像。OPSL 是 Xue 等人提出的通过最优预测子空间学习的方法去除特征空间的冗余信息,更好地进行图像表示和图像标注。NSIDML 是 Jin 等人提出的基于图像距离度量学习和邻域集的图像标注方法,目的是弥合图像之间的语义鸿沟,进而提高图像标注性能。通过在 Corel 5K、IAPRTC-12 和 ESP Game 三个数据集上与传统的图像自动标注算法进行比较,比较结果分别如图 6 ~ 图 8 所示。

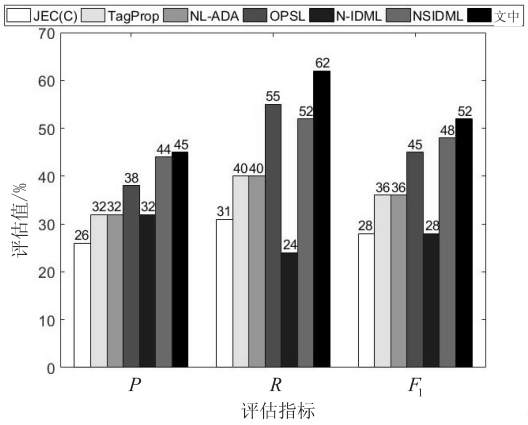


图 6 在 Corel 5K 数据集上的实验结果评估比较

从图 6 可以看出,在 Corel 5K 数据集上,文中方法与实验结果较好的 NSIDML 相比, P 提高了 1%, R 提高了 10%, F_1 提高了 4%;从图 7 可以看出,在 IAPRTC-12 数据集上,文中方法与实验结果较好的 NSIDML 相比, P 降低了 11%, R 提高了 7%, F_1 降低了 3%;从图 8 可以看出,在 ESP Game 数据集上,文中方法与实验结果较好的 NSIDML 相比, P 降低了 11%, R 提高了 13%, F_1 提高了 2%。文中方法取得了一定的改进效果,是因为文中采取卷积神经网络方法,经过卷积神经网络卷积层和池化层的作用,图像的特征从基础的颜色、纹理等特征转换成更适用于图像识别的特征,能更有效地进行待标注图像的近邻图像搜索,从而提高图像标注性能。总体来说,文中方法在 Corel 5K 数据集、IAPRTC-12 数据集和 ESP Game 数据集上的实验体现出了比较好的效果。

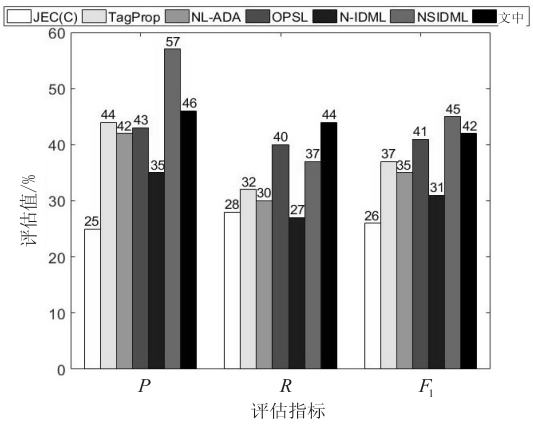


图 7 在 IAPRTC-12 数据集上的实验结果评估比较

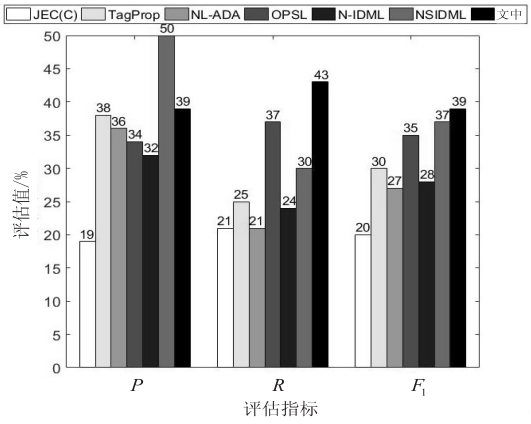


图 8 在 ESP Game 数据集上的实验结果评估比较

从上述分析中可以看出,基于 CNN 和贝叶斯的图像自动标注是有效的。从结果中可以看出,在 Corel 5K 数据集上的实验,文中方法的评估指标值均优于所比方法,在 IAPRTC-12 数据集和 ESP Game 数据集上的实验,文中方法整体上优于所比方法,查准率有一定的降低,但查全率和 F_1 值均高于所比方法,这是因为待标注图像的近邻图像可以为待标注图像提供更准确的标签,在此基础上再考虑图像低层特征与标签之间的映射关系,丰富标签信息,可以进一步提高图像标注性能。

4 结束语

文中提出了一种改进的基于 CNN 和加权贝叶斯后验概率的最近邻图像标注方法,利用 CNN 模型提取图像特征以获得表达能力更强的图像特征,并根据此特征找寻更准确的待标注图像的近邻图像,从而得到

更准确的标签。再通过贝叶斯构建图像低层特征和语义之间的关系,选择合适的标签为待标注图像进行标注。分别在三个基准数据集 Corel 5K、IAPRTC-12 和 ESP Game 上进行实验分析,结果表明该方法可以有效提高标注性能。

参考文献:

[1] MARKATOPOULOU F, MEZARIS V, MEMBER S, et al. Implicit and explicit concept relations in deep neural networks for multi-label video/image annotation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 29(6): 1631-1644.

[2] LIU Y, WEN K, GAO Q, et al. SVM based multi-label learning with missing labels for image annotation[J]. Pattern Recognition, 2018, 78(C): 307-317.

[3] 张素兰,郭　平,张继福,等. 图像语义自动标注及其粒度分析方法[J]. 自动化学报,2012,38(5):688-697.

[4] CHENG Q M, ZHANG Q, FU P, et al. A survey and analysis on automatic image annotation[J]. Pattern Recognition, 2018, 79: 242-259.

[5] JI P, GAO X, HU X. Automatic image annotation by combining generative and discriminant models[J]. Neurocomputing, 2017, 236: 48-55.

[6] HOU Y Q, LIN Z C. Image tag completion and refinement by subspace clustering and matrix completion[C]//2015 visual communications and image processing (VCIP). Singapore: IEEE, 2015: 1-4.

[7] WU B, LYU S, HU B, et al. Multi-label learning with missing labels for image annotation and facial action unit recognition[J]. Pattern Recognition, 2015, 48(7): 2279-2289.

[8] WANG R, XIE Y, YANG J, et al. Large scale automatic image annotation based on convolutional neural network[J]. Journal of Visual Communication & Image Representation, 2017, 49: 213-224.

[9] SU F, XUE L K. Graph learning on k nearest neighbours for automatic image annotation[C]//International conference on multimedia retrieval. Shanghai, China: Association for Computing Machinery, 2015: 403-410.

[10] VERMA Y, JAWAHAR C V. Image annotation by propagating labels from semantic neighbourhoods[J]. International Journal of Computer Vision, 2017, 121(1): 126-148.

[11] VERMA Y, JAWAHAR C V. Image annotation using metric learning in semantic neighbourhoods[C]//European conference on computer vision. Berlin: Springer-Verlag, 2012: 836-849.

[12] JIN C, JIN S W. Image distance metric learning based on neighborhood sets for automatic image annotation[J]. Journal of Visual Communication & Image Representation, 2016, 34: 167-175.

[13] RAD R, JAMZAD M. Automatic image annotation by a loosely joint non-negative matrix factorisation[J]. IET Computer Vision, 2015, 9(6): 806-813.

[14] 柯　逍,周铭柯,牛玉贞. 融合深度特征和语义邻域的自动图像标注[J]. 模式识别与人工智能,2017,30(3):193-203.

[15] ZHANG M, LI W, DU Q. Diverse region-based CNN for hyperspectral image classification[J]. IEEE Transactions on Image Processing, 2018, 27(6): 2623-2634.

[16] 常　亮,邓小明,周明全,等. 图像理解中的卷积神经网络[J]. 自动化学报,2016,42(9):1300-1312.

[17] 白　琮,黄　玲,陈佳楠,等. 面向大规模图像分类的深度卷积神经网络优化[J]. 软件学报,2018,29(4):1029-1038.

[18] 高耀东,侯凌燕,杨大利. 基于多标签学习的卷积神经网络的图像标注方法[J]. 计算机应用,2017,37(1):228-232.

[19] VERMA Y. Diverse image annotation with missing labels[J]. Pattern Recognition, 2019, 93: 470-484.

[20] 周飞燕,金林鹏,董　军. 卷积神经网络研究综述[J]. 计算机学报,2017,40(6):1229-1251.

[21] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.

[22] CHATFIELD K, SIMONYAN K, VEDALDI A, et al. Return of the devil in the details: delving deep into convolutional nets[C]//BMVC 2014-Proceedings of the British machine vision conference. UK: BMVA Press, 2014.

[23] 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.

[24] NING Z L, ZHOU G H, CHEN Z K, et al. Integration of image feature and word relevance: toward automatic image annotation in cyber-physical-social systems[J]. IEEE Access, 2018, 6: 44190-44198.

[25] KE X, ZHOU M, NIU Y, et al. Data equilibrium based automatic image annotation by fusing deep model and semantic propagation[J]. Pattern Recognition, 2017, 71: 60-77.

[26] XUE Z, LI G R, HUANG Q M. Joint multi-view representation and image annotation via optimal predictive subspace learning[J]. Information Sciences, 2018, 451-452: 180-194.