

一种融合语义距离的最近邻图像标注方法

吴伟¹ 高光来¹ 聂建云²

(内蒙古大学计算机学院 呼和浩特 010021)¹ (加拿大蒙特利尔大学计算机系 蒙特利尔)²

摘要 传统的基于最近邻的图像标注方法效果不佳,主要原因在于提取图像视觉特征时,损失了很多有价值的信息。提出了一种改进的最近邻分类模型。首先利用距离测度学习方法,引入图像的语义类别信息进行训练,生成新的语义距离;然后利用该距离对每一类图像进行聚类,生成多个类内的聚类中心;最后通过计算图像到各个聚类中心的语义距离来构建最近邻分类模型。在构建最近邻分类模型的整个过程中,都使用训练得到的语义距离来计算,这可以有效减少相同图像类内的变动和不同图像类之间的相似所造成的语义鸿沟。在 ImageCLEF2012 图像标注数据库上进行了实验,将本方法与传统分类模型和最新的方法进行了比较,验证了本方法的有效性。

关键词 图像标注,特征提取,最近邻,距离测度学习,语义距离

中图法分类号 TP391 文献标识码 A DOI 10.11896/j.issn.1002-137X.2015.1.066

Combination of Nearest Neighbor with Semantic Distance for Image Annotation

WU Wei¹ GAO Guang-lai¹ NIE Jian-yun²

(Department of Computer Science, Inner Mongolia University, Huhhot 010021, China)¹

(Department IRO, University of Montreal, Montreal, Canada)²

Abstract Most of the nearest neighbor (NN) based image annotation or classification methods do not achieve desired performances. The main reason is that much valuable information is lost when extracting visual features from image. A novel nearest neighbor method was proposed. Firstly, we obtained a new image semantic distance learned by distance metric learning (DML) using image class information, and then multiple clustering centers were formed based on this learned semantic distance. Finally, we constructed our NN model by calculating the distances between the image and these clusters. Our model can minimize the semantic gap for intra-class variations and inter-class similarities. Experimental results on image annotation task of ImageCLEF2012 confirm that our method is efficient and competitive compared with the traditional and state of the art classifiers.

Keywords Image annotation, Feature extraction, Nearest neighbor, Distance metric learning, Semantic distance

1 概述

图像的标注(或分类)一直以来都是视觉智能领域的研究热点。图像标注的目标是从图像中自动识别出视觉语义概念,包括各种场景(如室内、室外、自然风景等)、对象(汽车、人、动物等)、事件(如工作、旅行等)以及情感(如高兴、沮丧等)等。由于这些待标注的语义概念涵盖的类型极其广泛,且语义相同的图像外观可能差异非常大,不同语义的图像之间外观也可能比较相似,使得图像的自动标注极具挑战性^[1]。目前,全世界有很多的研究人员和组织在从事相关的研究工作,比较著名的组织如 ImageCLEF^[2]、ImageNet^[3]和 Pascal VOC^[4]等,吸引了很多优秀的研究人员参与此项研究。

图像的自动标注问题已经得到了广泛的研究,主要的方法可分为3种:基于分类模型的方法、基于概率模型的方法,以及基于网络关联图像的方法。基于分类的方法把图像标注看成是分类问题,通过构建图像分类器来进行图像文本概念的标注^[5]。基于概率的方法是通过建立图像与待标注语

义的联合概率来推导图像与标注语义之间的相关性^[6-8]。基于网络关联图像的方法利用网络上上下文信息来解决图像标注^[9,10]。此外,还有一些研究者使用多标签学习算法进行图像标注^[11]。

在众多的图像标注方法中,使用最广泛的是基于分类模型的标注方法。分类模型一般使用基于学习的方法,如 SVM^[12,13]、生成模型^[14]等,但是却很少使用基于最近邻的分类模型,主要原因是,通常情况下其性能不如基于学习的方法。但是,我们可能低估了最近邻(nearest neighbor, NN)分类器的能力。Boiman等^[15]指出,引起最近邻方法性能低下的主要原因是其使用的视觉特征并不能很好地反映图像语义信息,特别是目前经常使用的视觉词袋(bag of visual words, BoVW)特征,在生成视觉单词时,会造成很多图像信息损失。因此,在构建图像视觉特征时,如果能减少信息损失,对于最近邻模型来说,就可以提高图像分类或标注性能。

通常,使用聚类算法可得到用于构建 BoVW 特征的码本,但是,在聚类形成视觉单词(即聚类中心)时会引起图像信

到稿日期:2014-02-28 返修日期:2014-05-12 本文受国家自然科学基金项目:面向电话语音的蒙古语关键词检测技术的研究(61263037)资助。

吴伟(1980—),男,博士生,讲师,主要研究方向为人工智能、图像分类与检索, E-mail: cswuwei@imu.edu.cn; 高光来(1964—),男,硕士,教授,主要研究方向为人工智能、图像分类与检索、自然语言处理; 聂建云(1963—),男,博士,教授,主要研究方向为人工智能、信息检索。

信息的损失。为了尽可能减少聚类所造成的信息损失,很多研究者提出了一些改进的聚类算法,如 Jurie 和 Trigs^[16]提出了一种基于可变半径长度的聚类方法, Wu 和 Rehg^[17]提出使用直方图交叉核函数来创建码本, Gemert 等人^[18]提出了一种软分配 (soft-assignment) 生成码本模型, 这些改进方法都有效减少了图像信息损失。此外, 也有研究者使用距离测度学习方法 (distance metric learning, DML)^[19-21] 来构建 BoVW 模型, 使其提高了视觉特征的表示能力。Boiman 等^[15]提出了一种最近邻分类方法, 这种方法不使用 BoVW 特征, 而是直接使用图像的原始局部特征, 这样就不存在生成码本时的信息损失, 最终的图像分类性能要优于基于学习的分类模型。

传统的最近邻距离度量方法, 如欧氏距离、Jensen-Shannon divergence (JSD)、 χ^2 -距离、直方图交叉距离、Kullback-Leibler 距离、Earth Movers 距离^[22], 以及几何模糊距离^[19]等, 在大规模图像集下, 并不能很好地度量图像间语义的相似度。近年来, 国内外的研究人员提出了很多改进的最近邻方法。杨立等^[23]利用同一特征属性内取值的语义差异, 提出了一种基于领域本体语义距离的最近邻分类方法; 刘松华等^[24]利用信息能度量方法, 提出了一种核最近邻方法; 郭玉堂^[25]、郑君君等^[26]把最近邻模型转化为属性 k 近邻图模型来解决图像标注与检索问题。

国外的很多研究者把距离测度学习 (DML) 引入到了最近邻分类模型中, 如 Blitzer 等^[27]把学习得到的新的距离测度运用到传统最近邻模型; Wang Z 等^[28]提出了一种基于“图像-类别”的最近邻方法; Wang F 等^[29]把基于 WordNet 的标注词之间的语义关系引入到最近邻模型; Verma 和 Jawahar^[30]提出了一种两阶段最近邻模型。将这些方法运用到图像分类、标注等领域都取得了较好的效果。

本文提出的最近邻分类方法与文献^[15]不同, 仍然可以使用 BoVW 特征模型, 同时引入了图像的类别语义信息, 通过 DML 训练生成新的距离测度, 从而构造出一种新的语义距离, 可以有效减轻大规模图像库中经常出现的类内图像变动及类间图像相似所造成的干扰。所提出的方法与上文提到的基于 DML 的最近邻方法^[27-30]的不同之处在于, 所使用的距离测度学习优化方法不同, 同时, 采用了一种半自动化的训练数据生成方法, 可以有效减少训练数据量, 降低运算复杂度。此外, 提出的最近邻分类器也与上面提到的最近邻方法不同, 这种改进的最近邻模型首先对每类图像进行聚类形成多个子类中心 (即聚类中心), 然后对图像到各个子类的距离求和, 根据求和值判断与图像最接近的类别。在 ImageCLEF2012^[22]图像标注数据库上的实验证实了该最近邻模型的有效性。

本文第 2 节介绍基于语义类别信息的距离测度学习算法; 第 3 节介绍提出的最近邻分类模型; 第 4 节为实验结果及与其它方法的比较分析; 最后为结论。

2 语义距离测度学习算法

在计算图像之间的距离时, 为了减少类内图像的变动或者类间图像的相似性所引起的语义鸿沟, 利用 DML 方法, 引入图像的类别信息来构建新的距离度量函数。DML 的目标是利用给定的训练数据得到最优的距离测度矩阵^[19]。首先,

从训练图像集中提取用于 DML 学习的特征约束集合, 即训练样本集合, 每个特征约束包含两幅图像的类别是否一致的信息, 这些特征约束可以形式化地表示为集合: $\{(f_{i1}, f_{i2}, y_i)\}_{i=1}^N$, 其中, f_{i1} 和 f_{i2} 分别表示两幅图像的特征, N 是训练数据的个数, 如果这两幅图像属于同一图像类, 则 $y_i = 1$, 否则 $y_i = -1$ 。对于给定的图像训练集, 如何选择特征约束训练集合, 会极大地影响最后生成的距离测度的优化效果, 从而影响最近邻分类模型的性能。这里采用一种直观的选择规则: 选择属于相同图像类别。但是视觉外观上差异较大的图像对, 或者在外观上较相似但属于不同类别的图像对, 来采集特征约束训练数据。

在具体采集训练数据时, 使用了一种半自动生成方法。首先, 提取出所有训练图像的某种视觉特征, 并对每一类内图像特征运用 k -均值聚类算法进行聚类, 形成 k 个聚类中心 (可称为图像类内的子类), 可以用这些子类表示相同图像类内的差异, 所以, 可以把这些聚类中心看成是同一图像类中外观差异比较明显的不同“图像”, 对这些“图像”两两配对形成特征约束训练样本: $(f_{i1}, f_{i2}, y_i = 1)$ 。然后, 对每一个“图像”, 从任意其它的训练图像类中选择一个与其距离最近的图像配对, 形成训练样本: $(f_{i1}, f_{i2}, y_i = -1)$ 。这里, 任意选择其它的图像类, 可以人工干预, 以选择一个最恰当的图像类来保证两个待配对的图像类别语义不同, 同时这两类图像在外观上可能比较相似, 如图像类“猫”和“狗”在语义上不同, 但在外观上相对比图像类“飞机”更可能相似。这也是所谓“半自动”而不是“自动”方法的含义。

给定训练数据后, 利用 DML 方法学习得到新的距离测度矩阵 A 来计算图像特征之间的语义距离:

$$d(f_{i1}, f_{i2}) = \sqrt{(f_{i1} - f_{i2})^T A (f_{i1} - f_{i2})} \quad (1)$$

要找到最优的测度矩阵 A , 也就是要使得属于同一图像类别的特征之间的距离最小, 属于不同类别之间的距离最大。基于这一原则, 把距离测度学习形式化为以下优化问题:

$$\begin{aligned} \min_{A, b} \sum_i y_i (\|f_{i1} - f_{i2}\|_A^2 - b) + \frac{\lambda}{2} \text{tr}(A^T A) \\ \text{s. t. } y_i (\|f_{i1} - f_{i2}\|_A - b) \leq 1 \\ A \geq 0, \|A\| = 1/\sqrt{\lambda} \end{aligned} \quad (2)$$

其中, $\|\cdot\|_A$ 是关于测度矩阵 A 的马氏距离 (即式 (1)), 最小化目标函数的第二项是规则项, 作用是防止在最小化过程中出现过拟合现象。第一个不等式约束条件的作用是使得在语义上相同的两幅图像特征之间距离更接近。第二个不等式约束条件的作用是防止测度矩阵 A 被优化为零矩阵。参数 λ 是常数, b 是指定的阈值变量, 用于判断两个图像特征是否相似。这是一个半定规划问题, 可以采用随机梯度搜索算法来实现, 算法过程描述如下:

1. 输入训练数据集 $\{(f_{i1}, f_{i2}, y_i)\}_{i=1}^N$;
2. 参数 λ , 学习因子 γ ;
3. 初始化 $A = I, b = 1$, 迭代参数 $t = 1$;
- Repeat
4. $\lambda = \lambda/t, t = t + 1$
5. $C_t = \{(f_{i1}, f_{i2}, y_i) \mid y_i (\|f_{i1} - f_{i2}\|_A^2 - b) < 1\}$
6. $f(A, C_t) = \sum_{(f_{i1}, f_{i2}, y_i) \in C_t} y_i (\|f_{i1} - f_{i2}\|_A^2 - b) + \frac{\lambda}{2} \text{tr}(A^T A)$
7. 计算函数 f 关于 A 的梯度: $\nabla_A f$

8. 计算函数 f 关于 b 的梯度: $\nabla_b f$

9. 更新 A 和 b : $A = A - \frac{\gamma}{t} \nabla_A f, b = b - \frac{\gamma}{t} \nabla_b f$

10. 约束 A 为半正定矩阵 $A \geq 0: A \leftarrow \sum_i \max(0, \lambda_i) \phi_i \phi_i^T$

11. 满足 A 的限制条件 $\|A\| = 1/\sqrt{\lambda}: A \leftarrow \frac{1}{\sqrt{\lambda}} \frac{A}{\|A\|}$

Until 收敛

输出: 矩阵 A , 阈值 b 。

算法中,第 10 步中的 λ_i 和 ϕ_i 分别表示矩阵 A 的第 i 个特征值与特征向量。该算法是一个迭代的过程,这种半定规划(semi-definite programming, SDP)问题在给定约束条件下,其收敛性很容易得到保证,详细证明可以参考文献[19, 20],而且一般情况下,至多迭代 5 次左右就会达到收敛状态。一般的 DML 算法的复杂度为 $O(n^2 t d^2)$,其中 n 为训练图像的个数, t 为迭代次数, d 为图像特征的维度。仔细分析所提出的算法过程,其时间复杂度为 $O(N t d^2)$, N 是训练样本的个数,即通过上述半自动方法生成的特征约束集合的大小,而普通 DML 方法通常是对任意 2 个训练图像两两配对形成特征约束训练集合(即训练集合大小的量级为 n^2),可见 N 要远小于 n^2 ,降低了运算开销。该算法主要的运算开销在于训练数据集的大小,后面的实验给出了训练数据量的大小与图像标注的性能之间的关系。

训练优化得到测度矩阵 A 之后,将语义距离式(1)用于后续的最近邻分类模型中特征之间的距离计算。下面介绍最近邻分类模型。

3 最近邻分类模型

传统的最近邻分类算法非常简单。如 k 近邻(k -nearest neighbor, k -NN)算法是根据给定样本到 k 个最近邻的训练样本类别,依据 k 个样本所属的类别投票决定该样本的类别,即属于投票数最多的类别。从理论角度来看,当训练样本个数趋于无穷时,最近邻模型性能接近于贝叶斯最优模型。而且,如果待判别图像与训练样本中的某个图像比较相似时,分类效果会很好。但是,实际情况是,训练样本个数不可能无穷大,此外,还存在很多类内图像的变化和类间图像相似的情况,此时传统的最近邻方法就难以胜任了。这里提出的改进最近邻方法不但采用新的语义距离,而且在计算距离时不采用传统的图像到图像的距离计算方法,而是通过图像聚类形成聚类中心,利用图像到聚类中心的距离进行最近邻建模。下面具体介绍提出的最近邻分类模型。

首先提取训练图像的某种视觉特征,然后利用 k 均值聚类(k -means)算法对每一类中的所有图像进行聚类,形成 k 个聚类中心(可以称为子类),每个子类由聚类中心的特征表示。不同于欧氏距离,运行聚类算法时的距离度量函数采用上面学习得到的语义距离(式(1))来计算,并且对所有的图像类,参数 k 取相同的值。这样,每一个图像类就得到 k 个子类: $f_1, f_2, \dots, f_k$ 。对于测试图像来说,就是寻找一个图像类 C ,使得得到各个子类的距离和最小,即计算公式 $\sum_{i=1}^k d(f_{test}, f_i^C)$ 的值使其达到最小,其中函数 $d(\cdot)$ 指的是语义距离(式(1)), f_{test} 是测试图像 $test$ 的特征, C 是图像语义类别的标号。

这种通过图像类内聚类,计算图像到所有子类距离和的方式,有效地调和了类内不同图像外观差异过大所造成的图

像语义的干扰;并且通过引入语义距离的方式,也减轻了类间图像相似的问题。

改进的最近邻算法可以概括如下:

1. 利用 k 均值聚类算法(基于新的语义距离)构建每一个图像类 C 的子类集合: $(f_1^C, f_2^C, \dots, f_k^C)$;

2. 计算测试图像 $test$ 的视觉特征 f_{test} ;

3. 分类结果:

$$\hat{C} = \arg \min_C \sum_{i=1}^k d(f_{test}, f_i^C) \quad (3)$$

当把该最近邻分类模型应用到多标签图像标注任务时,

只需要对每一图像类计算求和公式 $\sum_{i=1}^k d(f_{test}, f_i^C)$ 的值,然后根据该值的大小对所有待标注的图像语义类标签从低到高排序,值越小,说明图像 $test$ 与图像语义类标签 C 越相似,被标注为类标签 C 的可能性越大。

4 实验与讨论

实验数据来源于 ImageCLEF2012 图像自动标注数据库[2],该图像库共包含 94 个语义概念类,这些概念涵盖的范围非常广泛,包括自然景物、天气、人物、动植物、数量、交通、人的感觉印象等,而且是一个多标注分类问题,即同一幅图像可以被标注为多个语义概念。图像集提供了 15000 张图像作为训练集,另 10000 张图像用于测试。标注目标是在 94 个语义概念范围内,对测试集中的每一幅图像判断其包含哪些语义概念,对其进行标注,并按照测试图像与每个语义概念之间的相似度大小对这 94 个语义概念从高到低进行排序。性能评价标准采用 MiAP(Mean interpolated Average Precision),即首先单独计算每一个图像语义类的平均分类精度(Average Precision, AP),然后对所有 94 个图像语义类的 AP 求和再取平均值。这是分类与标注、检索等研究领域经常使用的评测指标。

实验选择了 3 种图像视觉特征进行测试:颜色直方图(Color Histogram)、FCTH(Fuzzy Color and Texture Histogram)[31]以及 SIFT(Scale-invariant Feature transform)局部特征。SIFT 特征提取时,把图像分成 16×16 的图像块,然后分别提取 SIFT 特征,最终形成视觉词袋(BoVW)直方图特征,码本大小即视觉词汇个数通过 k 均值聚类得到,实验中码本大小取值为 500[32,33]。这 3 种特征分别从不同的角度描述图像信息,彼此之间的相关度较低[34],如果组合起来,表示图像信息的能力要更强一些。

实验主要由 3 部分构成,首先测试提出的最近邻方法及与传统最近邻模型比较,然后与其它的模型进行比较测试,最后测试了 DML 算法中训练样本个数对标注性能的影响。

4.1 最近邻模型实验

首先,使用传统的 k -NN 分类器进行对比实验。分别使用上述 3 种图像特征单独进行实验,对于颜色直方图特征和基于 SIFT 的 BoVW 特征,使用 JSD(Jensen-Shannon divergence)[35]距离度量函数,根据实验[35]证明,对于直方图特征来说,使用 JSD 距离要优于欧氏距离。FCTH 特征使用 Tanimoto 系数[31]来计算距离。实验测试了 k -NN 参数 k 取不同值时,使用各个特征的图像标注结果,实验结果如图 1 所示。参数 k 的取值从 1 到 1000,从图 1 中可以看出,当参数 k 取值

为大约 50 左右时,3 种特征的性能都获得了接近最好的结果。其中, SIFT 特征的结果最好, MiAP 值达到了 0.27, FCTH 次之, MiAP 值为 0.18, Color Histogram 效果最差, MiAP 值为 0.12。

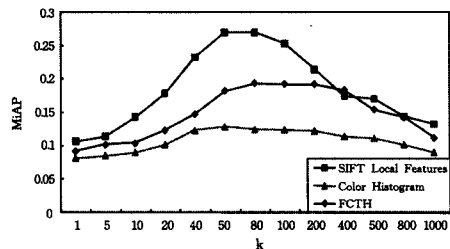


图1 传统 k -NN 分类模型参数 k 的取值对性能的影响

然后,对提出的改进最近邻分类模型进行实验。仍然使用上述 3 种视觉特征进行测试。实验主要测试了引入语义距离后,这种改进的分类模型的性能。此外,作为对比,还测试了不使用语义距离,仍然使用传统的距离度量方法时,对所改进的最近邻分类器的性能影响。实验结果如图 2 所示。

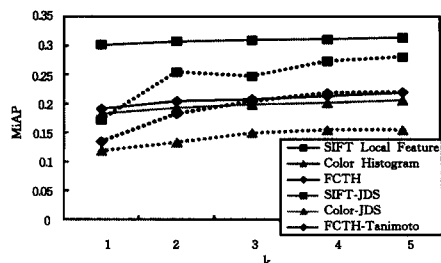


图2 改进的最近邻方法实验结果

图 2 中,横坐标 k 表示每个图像类的子类的个数(即 k 均聚类算法的参数 k),实验中,对所有图像类的子类个数取相同值。实验分别对每一种视觉特征进行了测试,如图 2 所示, SIFT Local Feature、Color Histogram 和 FCTH 分别表示使用相应的特征,在引入语义距离的情况下对改进的最近邻模型(式(3))进行实验的结果。作为对比, SIFT-JSD 和 Color-JSD 指的是不使用语义距离,而是使用 JSD 距离进行测试的结果,同理, FCTH-Tanimoto 是使用 Tanimoto 系数距离公式进行测试,即修改式(3)中的距离度量函数 $d(\cdot)$ 。

比较图 1 和图 2 的 MiAP 指标可以看出,在引入新的语义距离后,这种改进的最近邻方法效果明显优于传统的距离度量方法,除了 FCTH 特征性能变化不大外,其他 2 种特征性能的变化非常明显,如图 2 中基于 SIFT 的特征在参数 k 取值为 5 时, MiAP 值达到了 0.3143,这也是采用单一特征实验所取得的最好结果。而图 1 中 SIFT 特征取得的最好结果是 MiAP 值为 0.27。采用 FCTH 特征的结果稍差一些, Color 直方图特征效果是最差的,但是在引入了语义距离后,性能有了很大的提高,接近于 FCTH 特征的结果, MiAP 值达到了 0.2057。总之,可以看到不论采用哪种视觉特征,基于改进的最近邻方法都要优于传统的 k -NN 分类模型。

再观察图 2,可以看出,同样使用改进的最近邻模型,当引入语义距离后,其效果要好于不使用语义距离的情况。如使用 JSD 距离度量的 SIFT 特征(SIFT-JSD),在参数 k 为 5 时,其 MiAP 值是 0.2806,相比于使用语义距离的 SIFT 特征(MiAP 值为 0.3143)低了大约 3 个百分点,说明引入语义距

离后的确提高了标注性能。同时,与图 1 的结果相比,在不使用语义距离的情况下,这种改进的最近邻模型也要优于传统 k -NN 模型,说明这种改进模型的方法是有效的。

另外,从图 2 的曲线走势来看,当引入语义距离后,随着 k 值的增加,图像标注性能也在逐渐提升,但是变化非常缓慢(曲线走势比较平缓),这意味着参数 k (即类内子类的个数)对标注性能的影响不是很大,也就是说取较小的值就可以达到比较理想的结果,事实上,对 SIFT 特征来说,当 k 取 1 时,分类结果已经比较理想了;而且从计算复杂度来看,与图 1 中的参数 k 相比,图 2 中的 k 值很小,这明显减少了计算开销。

4.2 与其他模型的对比实验

实验使用其他一些分类模型进行了对比测试,如贝叶斯模型(Naïve Bayesian, NB)、一种加权的 k -NN 方法^[36](distance-weighted k -NN, dw- k NN)、SVM,及 Boiman 等^[15]提出的 NBNN 模型。这里 SVM 使用的核函数是直方图交叉核(Histogram Intersection Kernel, HIK)^[17]。其中 k -NN、NB、dw- k NN 以及 SVM 使用相同的 3 种特征: SIFT、FCTH 及 Color 直方图特征,然后采用最大投票法^[36]组成多特征分类模型。NBNN 模型使用单一的 SIFT 局部特征来实验。另外,研究表明^[1,36],结合多特征组合的分类器性能要优于单一特征的分类效果,因此,对上述 3 种特征对应的改进最近邻模型进行线性组合,对所形成的多特征分类模型进行了实验。实验结果如图 3 所示。

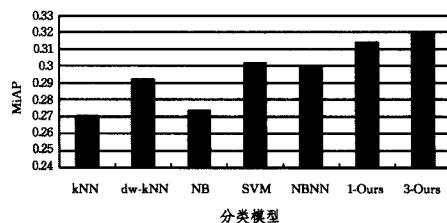


图3 不同分类模型的结果比较

图 3 显示了不同模型的标注结果,1-Ours 指的是利用提出的改进最近邻模型,采用 SIFT 单一特征的实验结果;3-Ours 是采用 3 种特征的最近邻模型进行线性组合的实验结果。可以看出,相比于单一特征模型,多特征组合模型确实改进了图像标注性能, MiAP 值从 0.3143 提高到了 0.3201,并且与其他模型相比,效果也是最好的。观察其他的方法, SVM 取得了相对最好的结果, MiAP 值为 0.3015, dw- k NN 次之, k -NN 效果最差, NBNN 的结果非常接近于 SVM, MiAP 值为 0.3004。每一种分类模型的详细实验结果如表 1 所列。

表1 不同分类模型的分类结果

分类模型	特征类型	MiAP
k-NN	多特征	0.2702
distance-weighted k-NN	多特征	0.2917
Naïve Bayesian	多特征	0.2732
SVM	多特征	0.3015
NBNN	单特征	0.3004
1-Ours	单特征	0.3143
3-Ours	多特征	0.3201
ImageCLEF2012 (使用视觉特征得到的最好结果)	多特征	0.3481

表 1 中最后一行是目前 ImageCLEF2012 图像自动标注

数据库上所公布的最好结果(仅使用图像视觉特征的情况下)^[2], MiAP 值达到了 0.3481。这一结果比 3-Ours 方法高大约 2 个百分点。因此本文所提出的方法有一定的竞争力。

4.3 语义距离学习实验

在距离测度学习(DML)算法中,提取的训练数据大小对最终标注的结果影响很大。实验验证了训练数据量大小与标注性能的关系。实验时,首先对 94 类图像的每一类进行聚类,使得每一类图像形成 N 个子类(聚类中心),把它们看成是同一图像类内的 N 幅“图像”,然后对这 N 幅“图像”两两配对提取特征约束训练数据: $\{f_{i1}, f_{i2}, y_i = 1\}$, 共有 $N \times (N - 1) / 2$ 个。然后,对每一幅“图像”,从其它 93 种图像类中随机选择 5 种不同的图像类,再分别从这 5 种图像类的每一类图像中选择与该“图像”距离最近的图像配对,提取特征约束训练数据: $\{f_{i1}, f_{i2}, y_i = -1\}$, 共有 $5 \times N$ 个。这样,对于 94 类图像,总共提取的特征约束个数为 $94 \times (N \times (N - 1) / 2 + 5 \times N)$ 。实验以 N 为参数,即每类图像选择的聚类中心个数,从 2 到 10 进行测试,对应的总的训练数据大小为从 1034 到 8930。以 SIFT 局部特征为例,实验结果如图 4 所示。

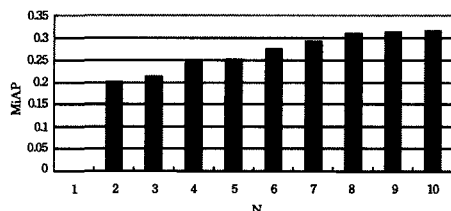


图 4 特征约束对的个数对性能的影响

从图中可以看到,随着训练数据量的增加,标注性能也逐渐提高。但是,当 N 值超过 8 时,提高的幅度趋于平缓,因此考虑到计算开销与性能之间的平衡,把子类的个数固定为 $N = 10$,即总的训练数据大小为 8930,来对 DML 算法进行训练。其他 2 种视觉特征 FCTH 和 Color 直方图也采用了相同的训练数据量进行训练。训练得到的新的距离测度函数用于上述改进的最近邻分类模型实验。

结束语 本文提出了一种基于语义距离的改进最近邻分类方法,通过距离测度学习算法训练得到新的语义距离。通过在 ImageCLEF2012 图像自动标注数据库上的实验及与其它方法的比较得知,这种方法实验结果较好,比较适合于大规模图像的标注任务。值得一提的是,尽管该方法实验取得了较满意的结果,但是在 DML 训练过程中,训练数据的提取还需要部分人工操作,而且依赖于经验。如果能够开发出完全自动的训练样本抽取方法,并增加训练数据,应该可以进一步提高图像标注的性能。

参考文献

- [1] Huiskes M, Thomee B, Lew M. New trends and ideas in visual concept detection[C]//Proceedings of the 11th ACM Conference on Multimedia Information Retrieval. Philadelphia, PA, USA, 2010; 527-536
- [2] Thomee B, Popescu A. Overview of the ImageCLEF 2012 Flickr Photo Annotation and Retrieval Task[C]//CLEF 2012 Working notes. Rome, Italy, 2012
- [3] Deng J, Berg A, Sstheesh S, et al. ImageNet Large Scale Visual

Recognition Competition 2012 (ILSVRC2012)[OL]//http://www.image-net.org/challenges/LSVRC/2012/

- [4] Everingham M, Gool L, Williams C, et al. The pascal visual object classes (voc) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338
- [5] Carneiro G, Chan A, Moreno P, 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
- [6] Jeon J, Lavrenko V, Manmatha R. Automatic Image Annotation and Retrieval Using Cross-media Relevance Models[C]//Proceedings of SIGIR. ACM, 2003
- [7] 刘峥, 马军. 一种基于图划分和图像搜索引擎的图像标注改善算法[J]. 计算机研究与发展, 2011, 48(7): 1246-1254
- [8] 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
- [9] Wang Chang-hu, Zhang Lei, Zhang Hong-jang. Learning to Reduce the Semantic Gap in Web Image Retrieval and Annotation [C]//Proceedings of SIGIR. ACM, 2008; 355-362
- [10] 许红涛, 周向东, 向宇, 等. 一种自适应的 Web 图像语义自动标注方法[J]. 软件学报, 2010, 21(9): 2183-2195
- [11] Kang F, Jin R, Sukthankar R. Correlated label propagation with application to multi-label learning[C]//Proceedings of Computer Vision and Pattern Recognition. IEEE, 2006; 1719-1726
- [12] Lin Y, Lv F, Zhu S, et al. Large-scale image classification; fast feature extraction and svm training[C]//Proceedings of Computer Vision and Pattern Recognition. IEEE, 2011; 1689-1696
- [13] Zhang H, Berg A, Maire M, et al. SVM-KNN: Discriminative nearest neighbor classification for visual category recognition [C]//Proceedings of Computer Vision and Pattern Recognition. IEEE, 2006; 2126-2136
- [14] Wang Gang, Zhang Ye, Li Fei-fei. Using dependent regions for object categorization in a generative framework[C]//Proceedings of Computer Vision and Pattern Recognition. IEEE, 2006; 1597-1604
- [15] Boiman O, Shechtman E, Irani M. In defense of nearest-neighbor based image classification[C]//Proceedings of Computer Vision and Pattern Recognition. IEEE, 2008; 1-8
- [16] Jurie F, Trigs B. Creating efficient codebooks for visual recognition[C]//Proceedings of International Conference on Computer Vision. IEEE, 2005; 604-610
- [17] Wu J, Rehg J. Beyond the euclidean distance; Creating effective visual codebooks using the histogram intersection kernel[C]//Proceedings of International Conference on Computer Vision. IEEE, 2009; 630-637
- [18] Gemert J C, Veenman C J, Smeulders A W, et al. Visual Word Ambiguity[J]. IEEE Trans. Pattern Analysis and Machine Intelligence, 2010, 32(7): 1271-1283
- [19] Varma M, Ray D. Learning the discriminative power-invariance trade-off[C]//Proceedings of International Conference on Computer Vision. IEEE, 2007
- [20] Hazan E, Agarwal A, Kale S. Logarithmic regret algorithms for online convex optimization[J]. Machine Learning, 2007, 69(2/3): 169-192
- [21] Wang S, Jiang S, Huang Q, et al. Multi-feature metric learning

- with knowledge transfer among semantics and social tagging[C]// Proceedings of Computer Vision and Pattern Recognition. IEEE, 2012; 2240-2247
- [22] Deselaers T. Features for image retrieval[R]. Rheinisch-Westfälische Technische Hochschule, Technical Report, Aachen, 2003
- [23] 杨立, 左春, 王裕国. 基于语义距离的 K-最近邻分类方法[J]. 软件学报, 2005, 16(12): 2054-2062
- [24] 刘松华, 张军英, 许进, 等. Kernel-kNN: 基于信息能度量的核 k-最近邻算法[J]. 自动化学报, 2010, 36(12): 1681-1688
- [25] 郭玉堂. 基于互 K 近邻图的自动图像标注与快速求解算法[J]. 计算机科学, 2011, 38(2): 277-280
- [26] 郑君君, 李新光, 祝一薇, 等. 海量图像集中 K 近邻求解的高效算法[C]//第十五届全国图象图形学学术会议论文集. 中国广东广州, 2010: 417-421
- [27] Blitzer J, Weinberger K Q, Saul L K. Distance metric learning for large margin nearest neighbor classification[J]. Advances in neural information processing systems, 2006, 18: 1473-1480
- [28] Wang Zheng-xiang, Hu Yi-qun, Chia Liang-tien. Image-to-class distance metric learning for image classification[C]// Computer Vision-ECCV 2010. Springer Berlin Heidelberg, 2010: 706-719
- [29] Wang F, Jiang S, Herranz L, et al. Improving image distance metric learning by embedding semantic relations[C]// Advances in Multimedia Information Processing-PCM 2012. Springer Berlin Heidelberg, 2012: 424-434
- [30] Verma Y, Jawahar C V. Image annotation using metric learning in semantic neighbourhoods [C] // Computer Vision-ECCV 2012. Springer Berlin Heidelberg, 2012: 836-849
- [31] Chatzichristofis S A, Boutalis Y S. FCTH: Fuzzy Color and Texture Histogram-a Low Level Feature for Accurate Image Retrieval[C]// Proceedings of IEEE 9th International Workshop on Image Analysis for Multimedia Interactive Services. Klagenfurt, Austria, 2008
- [32] Yang J, Yu K, Gong Y. Linear spatial pyramid matching using sparse coding for image classification[C]// Proceedings of Computer Vision and Pattern Recognition. IEEE, 2009: 1794-1801
- [33] Jia Y, Huang C, Darrell T. Beyond Spatial Pyramids: Receptive Field Learning for Pooled Image Features[C]// Proceedings of Computer Vision and Pattern Recognition. IEEE, 2012: 3370-3377
- [34] Zeng Z, Pantic M, Roisman G, et al. A survey of affect recognition methods: audio, visual and spontaneous expressions [J]. IEEE Transactions on pattern analysis and machine intelligence, 2009, 31(1): 39-58
- [35] Thomas D, Daniel K, Hermann N. Features for image retrieval: an experimental comparison[J]. Information Retrieval, 2008, 11(2): 77-107
- [36] Zhang Li, Zhou Wei-da. Sparse ensembles using weighted combination methods based on linear programming[J]. Pattern Recognition, 2011, 44(1): 97-106

(上接第 292 页)

观可行, 且具有有效的聚类分类性能, 通过仿真实验验证了理论的正确性和算法的有效性。

参 考 文 献

- [1] Jain A K, Murty M N, Flynn P J. Data clustering: A Review[J]. ACM Computing Surveys, 1999, 31(3): 264-323
- [2] Krishnapuram R, Keller J M. A possibilistic approach to clustering[J]. IEEE Trans Fuzzy Systems, 1993, 1(2): 98-110
- [3] Bahrapour S, Moshiri B, Salahshoor K. Weighted and constrained possibilistic C-means clustering for online fault detection and isolation[J]. Applied Intelligence, 2011, 35(2): 269-284
- [4] Liao Y Y, Jia K X, He Z S. Similarity measure based robust possibilistic C-means clustering algorithms[J]. Journal of Convergence Information Technology, 2011, 6(12): 129-136
- [5] Quintanilla-Dominguez J, Ojeda-Magañ B, Cortina-Januchs M G, et al. Image segmentation by fuzzy and possibilistic clustering algorithms for the identification of microcalcifications[J]. Scientia Iranica, 2011, 18(3): 580-589
- [6] Krishnapuram R, Keller J M. The possibilistic c-means algorithm: insights and recommendations[J]. IEEE Trans Fuzzy Systems, 1996, 4(3): 385-393
- [7] Duda R O, Hart P E, Stork D G. Pattern classification (2nd ed) [M]. New York: Wiley, 2001
- [8] Zhang J S, Leung Y W. Improved possibilistic c-means clustering algorithms[J]. IEEE Transactions on Fuzzy Systems, 2004, 12(12): 209-217
- [9] Pal N R, Pal K, Bezdek J C. A possibilistic fuzzy c-means clustering algorithm[J]. IEEE Transactions on Fuzzy Systems, 2005, 13(4): 517-530
- [10] 武小红, 周建江, 李海林, 等. 基于非欧式距离的可能性 C 均值聚类[J]. 南京航空航天大学学报, 2006, 38(6): 702-705
- [11] 刘兵, 夏士雄, 周勇, 等. 基于样本加权的可能性模糊聚类算法[J]. 电子学报, 2012, 40(2): 371-375
- [12] 陈东辉, 刘志镜, 王纵虎. 一种基于粒子群优化的可能性 c 均值聚类改进方法[J]. 计算机科学, 2012, 39(11): 122-126
- [13] Mohanad A, Mohammad M, Abdullah R. Using GA for Optimization of the fuzzy C-means clustering algorithm[J]. Research Journal of Applied Sciences, Engineering and Technology, 2013, 5(3): 695-701
- [14] Niu Q, Huang X J. An improved fuzzy C-means clustering algorithm based on PSO[J]. Journal of Software, 2011, 6(5): 873-879
- [15] Xu C, Zhang P L, Li B, et al. Vague C-means clustering algorithm[J]. Pattern Recognition Letters, 2013, 34(5): 505-510
- [16] Zhang Y, Huang D, Ji M, et al. Image segmentation using PSO and PCM with Mahalanobis distance[J]. Expert Systems with Applications, 2011, 38(7): 9036-9040
- [17] Anitha M, Selvy P T, Palanisamy V. WML detection of brain images using fuzzy and possibilistic approach in feature space [J]. WSEAS Transactions on Computers, 2012, 11(6): 180-189