

一种新的直方图核函数及在图像分类中的应用

贾世杰^{*①②} 孔祥维^①

^①(大连理工大学电信学部 大连 116023)

^②(大连交通大学电气信息学院 大连 116028)

摘要: 基于核方法的支持向量机(SVM)以其良好的推广性在图像分类等领域已经得到广泛应用,运用支持向量机的关键是设计有效的核函数。为克服传统核函数较少融合先验知识的弱点,该文提出基于数据驱动的核函数构建方法;并结合词包(BOW)模型,设计了一种基于TF-IDF规则的加权二次卡方(Weighted Quadratic Chi-squared, WQC)距离的直方图核函数;在计算直方图之间距离时充分考虑到不同量化区间的不同区分性能,从而增强核函数对不同类别的区分能力。在 Caltech101/256 等多个经典图像数据集上的分类实验结果验证了该文方法的有效性。

关键词: 图像分类;核函数;直方图;支持向量机

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2011)07-1738-05

DOI: 10.3724/SP.J.1146.2010.01244

A New Histogram-based Kernel Function Designed for Image Classification

Jia Shi-jie^{①②} Kong Xiang-wei^①

^①(Faculty of Electronic Information & Electrical Engineering, Dalian University of Technology, Dalian 116023, China)

^②(College of Electrical & Information, Dalian Jiaotong University, Dalian 116028, China)

Abstract: Kernel-based Support Vector Machine (SVM) is widely used in many fields (e.g. image classification) for its good generalization, in which the key factor is to design effective kernel functions. As there is not much a priori knowledge introduced into traditional kernel functions, the data-driven kernel building method is proposed to construct a new histogram kernel function which is combined with Bag OF Word (BOW) model and based on TF-IDF Weighted Quadratic Chi-squared (WQC) distance. In the process of calculating distances between histograms, the distinct discriminative power of each histogram bin is fully taken into consideration to boost classification performance of kernel functions. Experiments on several classic image data sets (Caltech101/256, etc.) show the better classification performance of the proposed method.

Key words: Image classification; Kernel function; Histogram; Support Vector Machine (SVM)

1 引言

支持向量机(Support Vector Machines, SVM)是在统计学习理论基础上形成的一种模式识别方法,其特征是可以在函数特征空间中有效训练线性学习分类器来处理非线性问题。使用支持向量机进行分类的关键是构建合适的核函数。不同的核函数对分类器的推广能力有很大的影响。虽然目前已经有一些常用的核函数(如线性核、多项式核、高斯核等),但迄今为止并没有提出统一有效的核函数定义方法。实际上,每一个特定的核函数都是一个特征映射问题^[1],核函数的构建应该结合具体的问题背景,能够集成相应的领域先验知识,才能最大程度

地提高分类性能^[2]。从这方面讲,核函数的设计也是一个从训练数据中进行学习的过程。由此,本文提出基于数据驱动构建核函数的方法,并将结合词包模型探讨在基于直方图的图像分类问题中核函数的构建问题。

基于内容的图像分类是根据图像的视觉特征对图像进行语义分类。近几年图像分类研究的焦点是图像场景分类(scene classification)和物体分类(object classification)^[3-6],主要采用有监督学习方法,通过对底层特征建模和中间语义分析来实现分类。图像直方图作为一种简单有效的统计描述方法,在图像处理各领域已经有了广泛的应用^[6,7]。常用的图像直方图形式有颜色直方图、梯度方向直方图等;很多图像特征描述子(如 SIFT^[8], ShapeContext^[9]等)都是基于直方图的表示方法。近年来在图像分类与

2010-11-15 收到, 2011-02-18 改回

国家自然科学基金重大项目(70890083)资助课题

*通信作者: 贾世杰 jiaishijie@dl.cn

识别领域非常成功的词包(Bag Of Word, BOW)模型^[10]就是把图像表示成视觉关键词直方图。文献[7]首次探讨了使用 SVM 进行图像分类的基本方法,提出了 RBF 核函数的推广形式(Laplace kernel, Gaussian kernel, sub-linear kernel),这也是本文构建核函数的基础。文献[11]首次分析了使用局部特征设计核函数的问题,构建了用于图像分类识别的核函数族,并指出使用局部特征构造的核函数能够取得比全局颜色直方图更好的分类性能;文献[12]构造了一种直方图相交(Histogram Intersection, HI)核函数,并给出了满足 Mercer 条件的证明。文献[13]对其进行了改进,提出了一种泛化直方图相交核(Generalize Histogram Intersection, GHI)的核函数构造方法。文献[14]探讨了在图像分类和新颖检测中核函数的构建问题,并指出 Mercer 条件并不是构造有效核函数的必要条件。文献[15]对图像分类中核的设计问题进行了深入研究,文献[16]运用卡方核进行大规模的图像分类,这些方法大大丰富了图像核函数的研究。但这些构建直方图核函数的方法都没有考虑融合先验知识的问题,在采用距离计算时也没有考虑到直方图各量化区间存在的相关性问题。

本文将数据驱动的思想引入到核函数的构建过程中,结合词包模型,在二次卡方(Quadratic Chi-squared, QC)距离^[17]的基础上构建了一种新的广义高斯直方图核函数—基于 TF-IDF 规则的加权二次卡方(Weighted Quadratic Chi-square, WQC)核。

2 基于词包的图像直方图表示

词包(Bag Of Word, BOW)或称特征包(Bag Of Feature, BOF)模型将文本分类技术引入图像分类与检测,将图像表示成一系列视觉关键词(visual words)的统计直方图。视觉关键词直方图的构建分为以下几个步骤:

(1)运用局部特征检测器(如 Harris, Laplace 等)自动检测(或使用密集采样方式)获得图像兴趣点/兴趣区域或局部块;

(2)运用局部特征描述子(如 SIFT, SURF 等)描述局部区域;

(3)将多幅训练图像的局部特征描述符进行 K -均值聚类,形成若干聚类中心(即视觉关键词),聚类中心的个数(即词包的大小)可事先设定;

(4)计算每幅图像中各局部特征与各聚类中心的欧氏距离,距离最小的聚类中心即为此局部特征对应的“视觉关键词”,统计每幅图像中包含各视觉关键词的数目,形成视觉关键词直方图。

3 基于二次卡方距离的广义高斯核

3.1 直方图之间距离

文献[15]指出,使用不同的距离定义方式,可以构造不同形式的广义高斯核函数。卡方核就是一种采用卡方距离构建的一种广义高斯核。设 X, Y 分别表示两个图像的直方图,则基于直方图的广义核函数一般形式为

$$K(X, Y) = \exp(-\rho D(X, Y)) \quad (1)$$

选择不同的距离计算方式会产生不同的核函数。高斯核函数使用欧式距离,拉普拉斯核^[7]则采用 L1 范数。这些距离计算方式只考虑到直方图对应量化区间取值的绝对距离,却忽视了对应区间的取值大小本身带来的影响。实际上对于分类来说,更有意义的是相对距离变化。卡方距离则反映每个对应区间取值的相对变化,因此以此构建的卡方核函数获得了更优的分类性能^[16];然而卡方距离仍旧采用的是传统的区间对应(bin-to-bin)的距离计算方式,并没有考虑到各量化区间本身存在的相关性。实际上,直方图是通过量化方式得到的统计描述,其对应的各个量化区间存在不同程度的相关性,并且每个量化区间对分类性能的影响并不相同。如在基于 SIFT 词包的图像直方图描述中,直方图的每个量化区间(bin)表示经过聚类后的视觉关键词,在这些视觉关键词的分布中,距离越相近的关键词其相关性越大。在形成视觉关键词直方图的过程中采用基于欧氏距离的最近邻投票方式,最终形成的直方图与聚类中心个数及聚类方法密切相关,因而考虑视觉关键词之间的的相关性能够减少量化误差对分类性能带来的影响。

对于 K 个量化区间,构造相似度加权矩阵 \mathbf{A} 。矩阵 \mathbf{A} 的每个元素 A_{ij} 表示区间 i 和区间 j 之间的相似度,相似度的取值在 0 和 1 之间,则两个直方图 X, Y 的二次式距离定义为

$$QF^A(X, Y) = \sqrt{(X - Y)^T \mathbf{A} (X - Y)} \quad (2)$$

如果矩阵 \mathbf{A} 取 X, Y 协方差矩阵的逆,则式(2)就是马氏距离(Mahalanobis distance)。将二次式距离与卡方距离相结合,形成二次卡方(Quadratic Chi-squared, QC)距离^[17]:

$$QC_m^A(X, Y) = \sqrt{\sum_{ij} \left[\frac{(X_i - Y_i)}{\left(\sum_c (X_c + Y_c) A_{ci} \right)^m} \right] \left[\frac{(X_j - Y_j)}{\left(\sum_c (X_c + Y_c) A_{cj} \right)^m} \right] A_{ij}} \quad (3)$$

其中 m 是规格化参数, 当 \mathbf{A} 是单位阵, $m=0.5$ 时

$$QC_{0.5}^I(X, Y) = \sqrt{2\chi^2(X, Y)} \quad (4)$$

此时 QC 距离就是卡方距离的形式。当 $m=0$ 时, QC 距离就是二次式距离。

QC 距离具有相似矩阵量化不变性(similarity-matrix-quantization-invariance property)和稀疏不变特性(sparseness-invariance property), 具体证明可参见文献[17]。

3.2 基于 TF-IDF 规则的加权二次卡方距离

二次卡方距离(式(3))采用了跨区间的距离计算方法, 对每个量化区间赋予相同的权重。但实际上, 每个量化区间对分类影响并不相同; 如在词包模型中, 在经过聚类后形成的各个视觉关键词有不同程度的类别“区分特性”^[10] (discriminative characteristics)。本文借用文本分类中常用的 TF-IDF(Term Frequency-Inverse Document Frequency)规则, 对二次卡方距离进行改进, 形成加权二次卡方距离。

TF-IDF 是一种用于文本信息搜索和挖掘的加权技术^[18]。其主要思想是, 如果某个词语在一篇文章中出现的频率高, 并且在其它文章中很少出现, 则认为此词语具有很好的类别区分能力, 适合用来分类。在 BOW 模型中, 依据 TF-IDF 规则区分每个关键词的权重, 如果某个视觉关键词在某类中出现的频率高, 并且在其它类中很少出现, 则认为此关键词“区分性能”较好; 而在多个类中都出现的视觉关键词的“区分性能”较弱。在计算直方图之间距离时考虑到不同量化区间的不同区分性能, 会增强核函数对不同类别的区分能力。

假设有 K 个视觉关键词, 其权重矩阵为:

$$W=[w_1, w_2, \dots, w_i, \dots, w_K], \text{ 其中} \quad (5)$$

$$w_i = \frac{n_{id}}{n_d} \lg \frac{N}{n_i}$$

n_{id} 是第 i 个视觉关键词在类 d 中的出现次数, n_d 是类 d 中所有视觉关键词出现的总次数, N 表示类别总数, 而 n_i 表示出现第 i 个关键词的类别的数目。

构造加权直方图:

$$X^w = X \odot \mathbf{W}, Y^w = Y \odot \mathbf{W} \quad (6)$$

直方图 X, Y 之间的加权二次卡方距离计算公式如下:

$$WQC_m^A(X, Y) = \sqrt{\sum_{ij} \left(\frac{(X_i^w - Y_i^w)}{\left(\sum_c (X_c^w + Y_c^w) A_{ci} \right)^m} \right) \left(\frac{(X_j^w - Y_j^w)}{\left(\sum_c (X_c^w + Y_c^w) A_{cj} \right)^m} \right) A_{ij}} \quad (7)$$

3.3 构建基于加权二次卡方距离的直方图核函数

设 X, Y 分别表示两个图像的直方图, 构建加权二次卡方距离构建直方图核函数如下:

$$K(X, Y) = \exp(-\rho * WQC_m^A(X, Y)) \quad (8)$$

其中控制参数 ρ 的选取可通过多重交叉验证方式获得。按照文献[15]的分析结论, 本文 m 取 0.9, 相似度矩阵 \mathbf{A} 的设置采用如下方法:

$$A_{ij} = \frac{d_{ij}}{\max(d)} \quad (9)$$

其中 d_{ij} 表示量化区间 bin_i 与量化区间 bin_j 之间的卡方距离, $\max(d)$ 表示各个量化区间之间最大的卡方距离。

4 实验结果与分析

4.1 测试图像库

(1)(PI100)5 类服装图片 PI100^[9]是从 MSN 购物网站上收集的商品图像集, 包含 10000 幅商品图像, 图片分辨率约为 100×100 , 本文选取了其中 baby shoe, boxing glove, boot, cap, jacket 等 5 类服装类商品图片进行分类测试实验。

(2)(Catech101)5 类昆虫图片 Caltech 101^[20]包含了 9146 幅图片, 共 101 类, 另外还有一个背景类, 每类至少包含 31 幅图片, 平均分辨率约为 300×200 。本文选取了 butterfly, beetle, dragonfly, mosquito, ant 等 5 类昆虫图片。

(3)(Catech256)5 类野生动物图片 Caltech 256^[21]包含了 30607 幅图片, 共 256 类, 也包含一个背景类, 每类至少包含 80 幅图片。图像分辨率变化范围较大, 平均分辨率约为 450×270 。本文选取了 bear, cameral, chimp, dog, elephant 等 5 类野生动物图片。

4.2 实验设置

训练图片数设为每类 5 幅、15 幅和 30 幅, 测试图片每类 15 幅。提取图像特征采用均匀采样方法, 采样区域为 16×16 , 间隔为 8 个像素, 采用 SIFT 描述子, 词包大小分别设为 200(PI 100), 500(Caltech101), 500(Caltech256)。多类分类策略使用一对一(one versus one)方式, 最终的分类正确率取 5 次分类实验结果的的平均值。所有程序都在 windows XP 操作系统, 2 G 内存, matlab 7.0 环境下运行, SVM 分类器采用台湾大学林智仁开发的 libsvm2.88(<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>)。

4.3 实验结果与分析

实验结果如表 1 所示。从表中可以看出, 随着每类训练样本本数的增加, 5 种核函数的分类性能都

表1 使用不同核函数时的PI100/caltech 101/256平均分类正确率(%)

训练样本数	图像库	核函数类型				
		WQC	卡方核	直方图相交核	线性核	高斯核
5	PI 100	94.6	94.6	92	88	88
	Caltech101	53.3	42.7	45.3	38.7	40
	Caltech256	41.7	37.5	40.3	26.4	23.6
15	PI 100	97.4	97.3	97.1	94.6	94.6
	Caltech101	59.6	54.6	57.3	49.3	54.6
	Caltech256	43.3	42.6	42.6	41.3	37.3
30	PI 100	97.3	97.3	97.3	94.6	94.6
	Caltech101	69.4	64	68.3	54.6	58
	Caltech256	49.3	48	49	41	44

有所提高。从总体上看，线性核与高斯核大体相当，都远低于卡方核与直方图相交核，而后者又不及本文提出的 WQC 核函数。对于不同的图像数据集，各种核的性能也有所不同。对于分辨率较小、无背景干扰的 PI 100 商品图像数据集，WQC 核的性能与卡方核、直方图相交核大体相当，而从分辨率较高、背景较复杂、类内变化较大的 caltech101/256 图像数据集的测试结果看，WQC 核的性能都领先于其它核函数，直方图相交核略好于卡方核，高斯核和线性核分类正确率最低；说明本文所构造的直方图核函数在类内变化较大的数据集上比传统的核函数在分类性能上有一定的优势。从测试结果看，训练样本数越少，WQC 核的相对优势越明显。如使用 caltech101 图像集，当训练样本数分别为 5, 15, 30 时，WQC 核取得的分类正确率比直方图相交核分别高出 8 个、2.3 个和 1.1 个百分点。

5 结束语

本文在研究数据驱动核函数的基础上提出了一种基于加权二次卡方距离的直方图核函数，经过实验验证，在多个常用图像数据集上取得了比传统的直方图核函数更好的分类正确率，说明本文构建的直方图核函数在图像分类识别领域有较好的应用前景。以后的研究方向是(1)探索基于数据驱动的核函数构建的内在规律和一般方法；(2)本文构建的 WQC 核的性能取决于相似度矩阵 \mathbf{A} 的选择，有必要探索对相似度矩阵 \mathbf{A} 的优化方法，比如通过距离学习的方法提高分类性能；(3)卡方核和直方图相交核已经被证明是 Mercer 核函数，WQC 核是不是 Mercer 核目前尚不能确定。尽管根据文献[15]的结论，即使不是 Mercer 核，也可以取得很好的分类性能，但还是有必要从理论上给出 WQC 核是否满足

Mercer 条件的证明。

参考文献

- [1] Camps-Valls G, Rojo-Alvarez J, and Martinez-Ramon M. Kernel Methods in Bioengineering, Signal and Image Processing[M]. London, England: Idea Group Publishing, 2007, Chapter I, 2-3.
- [2] Decoste D and Schölkopf B. Training invariant support vector machines[J]. *Machine Learning*, 2002, 46(1): 161-190.
- [3] 贾世杰, 孔祥维, 付海燕, 金光. 基于互补特征与类描述的商品图像自动分类[J]. *电子与信息学报*, 2010, 32(10): 2294-2300.
Jia Shi-jie, Kong Xiang-wei, Fu Hai-yan, and Jin Guang. Auto classification of product images based on complementary features and class descriptor[J]. *Journal of Electronics & Information Technology*, 2010, 32(10): 2294-2300.
- [4] Jia Shi-jie, Kong Xiang-wei, Fu Hai-yan, and Jin Guang. Product images classification with multiple features combination[C]. The 2010 International Conference on E-Business Intelligence (ICEBI2010), Kunming, Yunnan, China, December 19-21, 2010: 469-475.
- [5] Jia Shi-jie, Kong Xiang-wei, and Jin Guang. Automatic fast classification of product images with class-specific descriptor[J]. *Journal of Electronics(China)*, 2010, 27(6): 808-814.
- [6] Dalal N and Triggs B. Histograms of oriented gradients for human detection[C]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, Vol. 1: 886-893.
- [7] Chapelle O, Haffner P, and Vapnik V. SVMs for histogram-based image classification[J]. *IEEE Transactions on Neural Networks*, 1999, 10(5): 1055-1064.

- [8] Lowe D. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [9] Belongie S, Malik J, and Puzicha J. Matching shapes[C]. Proceedings of the Eighth IEEE International Conference on Computer Vision, Vancouver, Canada, July 7-14, 2001, Vol. 1: 454-461.
- [10] Sivic J and Zisserman A. Video google: a text retrieval approach to object matching in videos[C]. Proceedings of the Ninth IEEE International Conference on Computer Vision, Nice, France, Oct. 13-16, 2003, Vol. 2: 1470-1477.
- [11] Wallraven C, Caputo B, and Graf A. Recognition with local features: the kernel recipe[C]. Proceedings of the Ninth IEEE International Conference on Computer Vision, Nice, France, Oct. 13-16, 2003, Vol. 1: 257-264.
- [12] Barla A, Odone F, and Verri A. Histogram intersection kernel for image classification[C]. Proceedings of the International Conference on Image Processing, Barcelona, Catalonia, Spain, Sept. 14-17, 2003, Vol. 2: 513-516.
- [13] Boughorbel S, Tarel J, and Boujemaa N. Generalized histogram intersection kernel for image recognition[C]. Proceedings of the International Conference on Image Processing, Image Processing, Genoa, Italy, Sept. 11-14, 2005: 161-164.
- [14] Barla A, Franceschi E, Odone F, and Verri A. Image kernels[C]. International Workshop on Pattern Recognition with Support Vector Machines, Niagara, Canada, Oct.10, 2002: 617-628.
- [15] Jianguo Z, Marszalek M, Lazebnik S, and Schmid C. Local features and kernels for classification of texture and object categories: a comprehensive study. Proceedings of the IEEE Computer Society Conference of Computer Vision and Pattern Recognition (CVPR'06), New York, USA, June 17-22, 2006, Vol. 2: 13.
- [16] Bosch A. Image classification for a large number of object categories[D]. [Ph.D. dissertation], University of Girona, 2007.
- [17] Pele O and Werman M. The quadratic-chi histogram distance family[C]. The 11th European Conference on Computer Vision (ECCV 2010), Crete, Greece, Sept. 5-11, 2010: 749-762.
- [18] Wikipedia, tf-idf, <http://en.wikipedia.org/wiki/Tf%E2%80%93idf>, 2010.11.
- [19] Microsoft Research: Product image categorization data set (PI 100), <http://research.microsoft.com/en-us/people/xingx/pi100.aspx>, 2010.11.
- [20] The Caltech-101 object categories, <http://www.vision.caltech.edu/feifeili/Datasets.htm>, 2010.11.
- [21] Caltech256, http://www.vision.caltech.edu/Image_Datasets/Caltech256, 2010.11.
- 贾世杰: 男, 1969 年生, 博士生, 副教授, 研究方向为图像分类与检索.
- 孔祥维: 女, 1963 年生, 教授, 博士生导师, 研究方向为多媒体信息安全、统计图像处理和识别等.