基于高斯分布监督学习样本的 Fisher 核构造方法①

黄可望 [1]，方万胜 [2]，朱嘉钢 [3]

[1](无锡职业技术学院 物联网技术学院，无锡 214121)
[2](公安部交通管理科学研究所 RFID 应用组，无锡 214151)
[3](江南大学 物联网工程学院，无锡 214122)

摘 要: 结合实际应用背景，针对各类样本服从高斯分布的监督学习情形，提出了构造 Fisher 核的新方法. 由于利用了样本中的类别信息，该方法用极大似然估计代替 EM 算法估计 GMM 参数，有效降低了 Fisher 核构造的时间复杂度. 结合核 Fisher 分类法，上述方法在标准人脸库上的仿真实验结果显示，用所提方法所构造的 Fisher 核不仅时间复杂度低，且识别率也优于传统的高斯核与多项式核. 本文的研究有利于将 Fisher 核的应用从语音识别领域拓展到图像识别等领域.

关键词: Fisher 核；混合高斯分布；核 Fisher 分类；极大似然估计

Fisher Kernel Construction Method Based on Gaussian Distribution

HUANG Ke-Wang [1], FANG Wan-Sheng [2], ZHU Jia-Gang[3]

[1](School of the Internet of Things Technology, Wuxi Institute of Technology, Wuxi 214121, China)
[2](RFID application group, Traffic Management Research Institute of the Ministry of Public Security, Wuxi 214151, China)
[3](School of the Internet of Things Engineering, Jiangnan University, Wuxi 214122, China)

Abstract: Based on the actual application background and the supervised learning situation in which samples of each class comply with Gaussian distribution, we propose a new method for Fisher kernel construction. With the help of classification information in the sample, this method use maximum likelihood estimation rather than EM algorithm to estimate the GMM parameters, which can effectively reduce the time complexity for Fisher kernel construction. A simulation experiment on standard face database shows that the above-mentioned method combined with Fisher kernel classification can not only reduce the time complexity of fisher kernel construction, but also exceed the traditional Gaussian kernel and polynomial kernel in terms of recognition rate. The study of this method will benefit the application of Fisher kernel from speech recognition to image recognition.

Key words: Fisher kernel; Gaussian mixture classification; kernel Fisher classification; maximum likelihood estimation

引言

图像识别领域已经发展了很多算法，应用最普遍的还是基于统计特性的模式识别方法. 其中核 Fisher 判别(Kernel Fisher Discrimination Analysis, 简称 KFDA)以其对非线性问题的较优处理及良好的分类性能而广泛应用于图像识别中，而核函数是最关键的因素之一 [1].

KFDA 方法的关键在于核函数，核函数的选择不仅包括核函数类型的选择，而且确定核函数类型后还要根据某些规则选择相关的核参数.

核函数的研究主要包括核的构造及其参数选择，而核参数的选择往往更具复杂性，很多学者在SVM中对其进行了大量研究，但在效果方面不是很理想，仍存在一定的缺陷或限制[1].

尽管核函数具有高度复杂性，但科研工作者仍在相关理论方面有了一定的突破：比如核函数的数学特性方面，有 Mercer 定理、内积核理论、再生核理论等；核函数的构造方面，吴涛等人[2]从核函数的组成本质

① 基金项目:江苏省产学研联创项目(BY2013015-40)
收稿时间:2015-04-23;收到修改稿时间:2015-06-08

出发，提出了利用散乱数据插值的方法确定特征空间中感兴趣点的内积值代替传统核函数的一般表达式，从一定程度上解决了核函数的难题，但另一方面这种方法在实际应用中的计算量很大，对实时性要求高的问题可能并不适用. 核参数的选择方面，文献[3]提出了一种基于距离测度的 Gram 矩阵方法，文献[4]提出了一种基于 RM 界的梯度下降算法，这两种都在一定程度上减少了计算量，但训练时间消耗仍然不太理想，且具有相当的复杂性.

在基于核方法的机器学习中，核的构造和选择对于学习性能的影响是很大的. 为了利用训练样本信息以提高核函数的性能，学者们提出了一些基于训练样本的核函数，Fisher 核就是其中之一. Jaakkola 和 Haussler 在 1998 年第一次提出 Fisher 核[5]，是针对序列数据的分类提出的，最早应用在DNA序列的分类中. 该核函数用来描述两个离散序列相似程度，其中离散序列是基于一定的概率模型，通过概率模型参数可将不定长语音特征序列映射到固定长度的样本. 之后，这种方法又被用于语音识别[6]，同样取得不错的效果，但它在图像识别领域中应用很少.

为了从根本上解决核参数选择的复杂性问题，本文引入 Fisher 核. 本文从另外两个角度观察 Fisher 核的特点并由此引出研究兴趣. 其一是，Fisher 核的构造完全取决于训练样本，这种充分利用了训练样本信息的核函数是否能反映样本之间的相似度，同时省略了参数选择的问题. 其二是，在监督学习应用背景下，如果每类样本分布均服从高斯分布(这一假设在很多应用中均成立)，则相应的 GMM 的建模是否可利用类别信息，而不需要使用 EM 算法. 本文对此作了研究，发现了当每类训练样本均服从高斯分布时，构造 Fisher 核时所用到的 GMM 中的每个成份实际对应一类样本的高斯分布，因此在估计 GMM 参数时，可舍弃需要多次迭代的 EM 算法而用极大似然法估计方法替代之，于是提出了基于高斯分布监督学习样本的 Fisher 核构造新方法，有效简化了 GMM 的建模过程.

将按本文方法构造的 Fisher 核引入到核 Fisher 分类法进行图像识别实验，结果显示使用本文方法构造的 Fisher 核的识别率为 0.8750，高于多项式核的 0.6875 和高斯核的 0.5625. 利用本文的方法可以将 Fisher 核的应用范围从语音识别等领域扩展到图像识别等领域.

1　Fisher核构造的一般方法

1.1 一般概率密度函数的 Fisher 核[5][7-11]

Fisher 核是基于生长模型定义的核函数.

设样本集的概率模型为 $p(x|\theta)$，其中 θ 表示该模型的参数向量. 记 Fisher 分值为

$$g(\theta,x)=\nabla_\theta \ln p(x|\theta)\,, \tag{1}$$

则 Fisher 核的定义为

$$k(x,x')=g(\theta,x)^T F^{-1} g(\theta,x)\,, \tag{2}$$

其中 F 是 Fisher 信息矩阵

$$F=E_x\left[g(\theta,x)g(\theta,x)^T\right]. \tag{3}$$

在实际中，可以用样本均值来简单的替代 Fisher 信息定义的期望值:

$$F\approx\frac{1}{N}\sum_{n=1}^{N}g(\theta,x_n)g(\theta,x_n)^T \tag{4}$$

即 Fisher 核可简化为

$$k(x,x')=g(\theta,x)^T g(\theta,x)\,. \tag{5}$$

1.2 密度函数为混合高斯分布时的 Fisher 核[6]

在语音识别等应用中，式(1)中的 $p(x|\theta)$ 为高斯混合模型. 当 $p(x|\theta)$ 为如下高斯混合模型时，

$$p(x|\theta)=\sum_{k=1}^{c}p_k b_k \tag{6}$$

其中:

$$b_k=\frac{1}{(2\pi)^{R/2}|\Sigma_k|^{1/2}}e^{-\frac{1}{2}(x-u_k)^T\sum_k^{-1}(x-u_k)} \tag{7}$$

有

$$g(\theta,x)=\begin{bmatrix} vec\left(\nabla_{p_k}\ln p(x|\theta)\right)\\ vec\left(\nabla_{\mu_k}\ln p(x|\theta)\right)\\ vec\left(\nabla_{\Sigma_k}\ln p(x|\theta)\right)\end{bmatrix}, \tag{8}$$

其中:

$$\nabla_{p_k}\ln p(x|\theta)=P(k|x)\frac{1}{p_k}\,, \tag{9}$$

$$\nabla_{\mu_k}\ln p(x|\theta)=P(k|x)\Sigma_k^{-1}(x-u_k)\,, \tag{10}$$

$$\nabla_{\Sigma_k}\ln p(x)=\frac{1}{2}P(k|x)(-\Sigma_k{}^{-1}+S_k^T\cdot S_k) \tag{11}$$

$$P(k|x)=\frac{p_k b_k}{\sum_{k=1}^{K}p_k b_k}\,, \tag{12}$$

$$S_k=\left(x-\mu_k\right)^T\Sigma_k^{-1} \tag{13}$$

但为求得 $g(\theta,x)$，先要用EM算法估计此GMM的参数 p_k、u_k 和 Σ_k.

EM 算法(Expectation Maximization Algorithm)即为期望最大化算法，是一个在已知部分相关变量的情况下，估计未知变量的迭代技术. EM 的算法流程如下: (1)初始化分布参数(2)重复直到收敛: E 步骤: 估计未

知参数的期望值, 给出当前的参数估计. M 步骤: 重新估计分布参数, 以使得数据的似然性最大, 给出未知变量的期望估计.

EM 算法有它的缺陷: “坏”的参数初始值设置可以导致 EM 算法陷进一些局部最优点; EM 算法的收敛速度比较慢; 需要相当的迭代次数. 这样一来会消耗较多的时间去完成样本的训练. 只有在不存在直接解决的算法的情况下, 才应该考虑使用 EM 算法, 它并不是解决限制条件下优化问题的高效方法. 因此, 考虑用新的方法来替代 EM 算法.

2 基于高斯分布监督学习样本的Fisher核构造方法

在很多监督学习的实际应用中, 各类别的样本服从高斯分布. 由于样本中含有类别信息, 则包含各类别样本信息的高斯混合模型中, 每一个高斯成分实际上对应于某一类的样本分布, 这样就可以利用极大似然估计方法估计出每个高斯成分的参数 p_k 、 u_k 和 Σ_k , 使得高斯混合模型的建模大为简化.

以下以类别数为 2 为例, 给出基于高斯分布监督学习样本的 Fisher 核的具体构造方法如下:

输入: 图像样本的类别数 C , 每类样本数 $N_i,(i=1,...C)$. (此处取 $C=2$), 图像样本总数 N ;

输出: Fisher 核 $K(x,x')$ 的表达式;

步骤:

①用极大似然估计计算各高斯成份的参数包括权系数、均值、协方差矩阵, 经推导得:

$$\alpha_i=\frac{N_i}{N} \tag{14}$$

$$\mu_i=\frac{1}{N_i}\sum_{x\in L(i)}x \tag{15}$$

$$\Sigma_i=\frac{1}{N_i}\sum_{x\in L(i)}(x-\mu_i)(x-\mu_i)^T \tag{16}$$

②计算对应的概率密度函数对数的偏导数:

$$\nabla_{p_k}\ln p(x|\theta)=\frac{1}{p_k} \tag{17}$$

$$\nabla_{\mu_k}\ln p(x|\theta)=\Sigma_k^{-1}(x-u_k) \tag{18}$$

$$\nabla_{\Sigma_k}\ln p(x|\theta)=\frac{1}{2}(-\Sigma_k{}^{-1}+S_k^T\cdot S_k) \tag{19}$$

其中,

$$S_k=\left(x-\mu_k\right)^T\Sigma_k^{-1} \tag{20}$$

③利用公式(8)计算 Fisher 分值;

$$g(\theta,x)=\begin{bmatrix} vec\left(\dfrac{1}{p_k}\right)\\ vec\left(\Sigma_k^{-1}(x-u_k)\right)\\ vec\left(\dfrac{1}{2}(-\Sigma_k{}^{-1}+S_k^T\cdot S_k)\right)\end{bmatrix} \tag{21}$$

④利用公式(5)得到 Fisher 核.

$$k(x,x')=\begin{bmatrix} vec\left(\dfrac{1}{p_k}\right)\\ vec\left(\Sigma_k^{-1}(x-u_k)\right)\\ vec\left(\dfrac{1}{2}(-\Sigma_k{}^{-1}+S_k^T\cdot S_k)\right)\end{bmatrix}^T\bullet\begin{bmatrix} vec\left(\dfrac{1}{p_k}\right)\\ vec\left(\Sigma_k^{-1}(x-u_k)\right)\\ vec\left(\dfrac{1}{2}(-\Sigma_k{}^{-1}+S_k^T\cdot S_k)\right)\end{bmatrix} \tag{22}$$

其中, $vec(*)$ 表示对矩阵 * 向量化, 即把一个该矩阵的列向量首尾依次相接, 形成一个总的列向量.

该方法优点是在已知了样本中的类别信息, 用极大似然估计代替 EM 算法, 避免了 EM 算法参数初始值设置不当陷入局部最优, 避免了 EM 算法耗时长的多次迭代. 该方法也有局限性, 即各类别的样本需服从高斯分布, 但这一点在很多监督学习实际应用中, 都能得到满足.

3 实验结果及其分析

3.1 核 Fisher 分类法中 Fisher 核的应用

Fisher 分类法是一种线性分类方法, 其的主要思想是在样本空间中寻找一个投影方向, 使得各类样本投影到该方向时, 其类内距离尽可能小, 同时类间距离尽可能大. 当将核方法引入 Fisher 分类法中, 就产生了核 Fisher 分类法. 核 Fisher 分类法可以解决线性不可分的分类问题.

核 Fisher 分类法的步骤描述如下:

①通过非线性映射将训练样本映射到高维特征空间中, 求出最佳鉴别矢量:

$$W_{opt}=\arg\max\frac{\left|\left(w^{\phi}\right)^T S_B^{\phi}w^{\phi}\right|}{\left|\left(w^{\phi}\right)^T S_W^{\phi}w^{\phi}\right|}; \tag{23}$$

②对所有训练样本在最佳鉴别矢量上作投影,

$$F_{train}=\left(w^{\phi}\cdot\phi\left(x_{train}\right)\right)=\sum_{k=1}^{2}\sum_{i=1}^{n_k}\alpha_{ki}k\left(x_{ki},x_{train}\right), \tag{24}$$

求出两类训练样本中心，

$$m_k = \frac{1}{n_k}\sum_{i=1}^{n_k} F_{train}^k \text{ ,} \tag{25}$$

其中，m_k 为每一类样本中心，$k=1,2$；

③对测试样本在最佳鉴别矢量上作投影，

$$F_{test} = \left(w^{\phi} \cdot \phi\left(x_{test}\right)\right) = \sum_{k=1}^{2}\sum_{i=1}^{n_k} \alpha_{ki} k\left(x_{ki}, x_{test}\right) \text{ ,} \tag{26}$$

计算测试样本特征向量与各类样本中心距离

$$d_k = \left|F_{test} - m_k\right| \text{ ,} \tag{27}$$

若 $d_i = \min\left(d_k\right), k=1,2$，则判定 $x_{test} \in$ 第i类.

实验中核 Fisher 分类法中的核函数采用 Fisher 核，分类类别数为 2. 通常图像样本的各类服从高斯分布，因此以图像识别作为实验内容.

3.2 实验结果分析

本实验在核 Fisher 分类法中分别使用多项式核、高斯核和本文构造的 Fisher 核，对标准图像数据库的样本进行训练和识别，目的是观察在相同的其它条件下 Fisher 核的表现.

训练样本取自标准彩色人脸数据库 FEI，如图 1 所示. 每幅图片大小均为 640*480，需要提取人脸局部图像，提取后的大小为 120*120，再用无迭代双边二维主成分分析(NIB2DPCA)方法[12]进行降维处理，降维后的大小为 3*40.

图 1　自标准彩色人脸数据库 FEI 中部分图片

总共分为 9 组实验，每组随机选取 FEI 的两个类别人脸图像，每个类别训练样本 4 张；测试分为 8 次，每次测试每人 1 张图片，每次测试的人脸样本均不相同，其中第 7 次测试样本的人脸为全侧脸，第 8 次测试样本的图片全光照有变化，实验结果如表 1 所示.

从表 1 的实验数据，可以得到两点结论：

①Fisher 核的总体表现要优于多项式核及高斯核. 在 9 组实验的每组 8 次实验中，本文的 Fisher 核的识别率均值为 0.8263，而同等条件下多项式核及高斯核的识别率均值均为 0.8194；在某些两类训练样本下，本文的 Fisher 核的识别率会比多项式核及高斯核的识别率要高出更多.

②Fisher 核的鲁棒性要优于多项式核及高斯核. 如表 1 中的第八组实验，当图像的光照发生变化时，本文的 Fisher 核的识别率为 0.8750，远高于多项式核的 0.6875 和高斯核的 0.5625. 注意，高斯核在第 3 次测试和第 4 次测试时有很高的错分率，达到了 100%.

以上的结论虽然是针对分类类别数为 2 的情形，但也不难推广到多类的情形.

表 1　三种核的两类样本正确识别数及识别率实验结果

	测试数据	测1	测2	测3	测4	测5	测6	测7	测8	识别率
多项式核	1 组	2	2	2	2	2	2	2	1	0.9375
	2 组	2	1	2	1	2	2	1	1	0.7500
	3 组	2	1	2	1	2	2	1	1	0.7500
	4 组	2	2	2	2	2	2	2	1	0.9375
	5 组	2	2	1	2	2	1	1	1	0.7500
	6 组	2	2	2	1	2	1	1	1	0.7500
	7 组	2	2	2	1	2	2	2	1	0.8750
	8 组	2	1	1	1	2	2	1	1	0.6875
	9 组	2	2	2	2	2	2	2	1	0.9375
高斯核	1 组	2	2	2	1	2	2	1	1	0.8125
	2 组	2	2	2	2	2	2	2	1	0.9375
	3 组	2	2	2	2	2	2	2	1	0.9375
	4 组	2	2	2	2	2	2	2	1	0.9375
	5 组	2	2	2	2	2	2	2	0	0.8750
	6 组	2	2	2	2	2	1	1	2	0.8750
	7 组	2	1	1	1	2	2	1	1	0.6875
	8 组	2	1	0	0	2	2	1	1	0.5625
	9 组	2	1	1	1	2	2	2	1	0.7500
本文 Fisher 核	1 组	2	2	2	2	2	1	2	2	0.9375
	2 组	1	1	2	1	2	1	2	1	0.6875
	3 组	2	1	2	1	2	2	1	2	0.8125
	4 组	2	2	2	2	2	2	2	1	0.9375
	5 组	2	1	2	1	2	2	2	1	0.8125
	6 组	2	2	2	2	2	1	1	2	0.8750
	7 组	2	1	1	1	2	2	1	1	0.6875
	8 组	2	2	1	2	2	2	2	1	0.8750
	9 组	2	2	1	1	2	2	2	1	0.8125

4　结语

本文针对各类样本服从高斯分布的监督学习情形，提出了构造 Fisher 核的新方法. 由于利用了样本中的

类别信息, 该方法用极大似然估计代替 EM 算法估计 GMM 参数, 避免 EM 算法大量的迭代运算, 有效降低了 Fisher 核构造的时间复杂度. 实验表明用所提方法所构造的 Fisher 核不仅时间复杂度低, 且识别率也优于传统的高斯核与多项式核. 今后的研究工作主要是针对多类别多尺寸的图像样本, 进一步进行 Fisher 核的改进, 使 Fisher 核更加深入地应用于在图像识别等各类样本服从高斯分布的监督学习场合.

参考文献

1 方万胜.核 Fisher 方法及其组件在图像识别中的应用研究[学位论文].无锡:江南大学,2013.

2 吴涛,贺汉根,贺明科.基于插值的核函数构造.计算机学报,2003,26(8):990–996.

3 李晓宇,张新峰,沈兰荪.一种确定径向基核函数参数的方法.电子学报,2005,(33):2459–2463.

4 颜根廷,马广富.一种混合核函数支持向量机算法.哈尔冰工业大学学报,2007,39(11):1704–1706.

5 Jaakkola, Haussler. Exploiting generative models in discriminative classifiers. Advances in Neural Information Processing Systems, 1998, (11): 487–493.

6 高毅.基于 FISHER 品质的中文姓名语音识别技术[学位论文].成都:电子科技大学,2006.

7 Perronnin F, Rodriguez-Serrano JA. Fisher kernels for handwritten word-spotting. The 10th International Conference on Document Analysis and Recognition . Beijing. IEEE. 2009.

8 Travieso CM, Briceno JC, Ferrer MA, et al. Using Fisher kernel on 2D-shape identification.Computer Aided Systems Theory-EUROCAST 2007. LNCS 4739. Berlin. Springer. 2007.

9 Dick U, Kersting K. Fisher kernels for relational data. Machine Learning. ECML 2006. LNCS4212. Berlin. Springer. 2006.

10 Ryo I, Saddaaki M. Nonparametric fisher kernel using fuzzy clustering. Knowledge-Based Intelligent Information and Engineering Systems. LNCS 4252. Berlin. Springer-Verlag. 2006.

11 Holub AD, Welling M, Perona P. Combining generative models and Fisher kernels for object recognition. The 10th IEEE International Conference on Computer Vision. Barcelona. IEEE. 2005.

12 Guan YP. Robust video foreground segmentation and face recognition. Journal of Shanghai University(English Edition), 2009, 13(4): 311–315.

13 阳春华,王觉.一种混合核函数 SVM 建模方法及其应用.控制工程,2010,17(4):524–526.

14 Zhao MY, Ren J, Ji LP, et al. Parameter selection of support vector machines and genetic algorithm based on change area search. Neural Computing and Applications, 2012, 21(1): 1–8.

15 尹君梅,杨明,万建武.一种面向不平衡数据集的核 Fisher 线性判别分析方法.模式识别与人工智能,2010,23(3): 415–420.

16 Reshma KC, Jayadeva SC. Optimal kernel selection in twin support vector machines. Optimization Letters, 2009, 3(1): 77– 88.

17 周欣,吴瑛.核 Fisher 判别分析在数字信号分类中的应用.北京邮电大学学报,2011,(34):35–39.