

一种基于近邻保留的相关反馈图像检索算法

鲁 珂 赵继东 丁正明 吴 跃

(电子科技大学计算机科学与工程学院 成都 610054)

摘 要 图像检索中很多时候会出现相关反馈提供的标注样本数不足,从而导致监督学习方法面临过适应问题的困扰。提出一种能有效使用未标记数据的半监督新型算法:近邻保留回归算法,它通过使已标记数据的观测误差函数最小化,来选择综合性能最好的回归函数,以兼顾图像的语义特征及图像空间的几何结构,并解决过适应问题。实验结果证明,算法能有效提高图像检索系统的性能。

关键词 流形学习,近邻保留,相关反馈,图像检索

中图分类号 TP391.4 **文献标识码** A

Neighborhood Preserving-based Relevance Feedback Algorithm in Image Retrieval

LU Ke ZHAO Ji-dong DING Zheng-ming WU Yue

(School of Computer Science and Engineering, UESTC, Chengdu 610054, China)

Abstract When there are no sufficient feedback samples provided by Relevance feedback, supervised learning methods may suffer from the over-fitting in image retrieval. This paper proposed a novel neighborhood preserving regression algorithm which makes efficient use of unlabeled images. The algorithm selects the function which can minimize the empirical loss on the labeled images, thus, the function can respect both semantic and geometrical structures of the image database. The experimental results show that the algorithm is effective for image retrieval.

Keywords Manifold learning, Neighborhood preserving, Relevance feedback, Image retrieval

1 引言

近年来互联网信息迅速膨胀,特别是多媒体信息急剧增加,使得基于内容的图像检索成为研究热点^[1,2]。相关反馈技术一直被认为是填补低级图像特征与高级语义特征之间的语义鸿沟的重要途径^[3,4]。然而,用户提供的相关反馈是有限的,因此如何更有效地利用这些反馈信息,目前仍然是一个亟待解决的关键问题。

传统的相关反馈图像检索方法分为两种类型。第一种方法依赖于反馈信息在图像空间中学习并获得一个距离矩阵。该方法的性能依赖于表示图像特征的选择及设计一个合适的函数来度量请求图像与数据库中图像的相似性。X. F. He 等人提出一种基于流形学习的相关反馈图像检索方法^[5],其基于用户的反馈信息,通过一个图来模拟图像空间的流形结构,模型中用测地线距离来表示图像之间的距离。其它一些这类算法参见文献^[6,7]。第二类方法通过一个学习到的分类函数完成工作。该方法通过一个监督的分类算法来学习一个分类器,从而将图像分为“相关”或“不相关”。这类算法包括 SVM, Bayesian, Boosting, 决策树等。

目前大部分相关反馈算法均是监督的学习算法。但实际

上,很多未标记数据作为测试集参与学习过程,相关反馈图像检索实际上可以看作一个半监督学习问题。Z. H. Zhou 等人提出了一种基于相关反馈的半监督学习方法^[8],该方法使用两个分类器,分类器由用户请求数据及标注数据学习得到,每个分类器将为另一个分类器标注一些未标记的数据来继续优化对方,最后两个分类器结合为一个。其它一些半监督的相关反馈算法参见文献^[9,10]。

受到近来流形正则化^[11]研究工作的启发,本文提出一种用于图像检索的新型算法:近邻保留回归算法。该算法通过同时使用未标注及已标注数据来试图获得一个能最小化预测误差的分类器。在现实应用中,一个实际的问题是可行的标记样本很少,但图像的特征空间维数却很高,这就使得可能存在很多分类器达到零训练误差(过适应问题)。为解决这一问题,我们引入一个正则项,通过它来选择一个最能保留图像近邻结构的分类器。由于每一个图像数据均可以近似地表示为它的 k 近邻的线性组合,因此其标注也可以表示为它的 k 近邻标注的线性组合。但是,需要注意的是,语义上相关的图像在图像空间中却不一定是近邻。

本文第 2 节简介线性回归问题;第 3 节介绍提出的近邻保留回归算法;第 4 节提供该算法的几个对比实验结果;最后总结全文。

到稿日期:2011-02-22 返修日期:2011-04-25 本文受国家自然科学基金(60702072),中央高校基本业务费(ZYGX2009X012)资助。

鲁 珂(1974—),男,博士,副教授,主要研究方向为网络多媒体技术,E-mail:kel@uestc.edu.cn;赵继东(1976—),男,博士,副教授,主要研究方向为网络与信息系统;丁正明(1987—),男,研究生,主要研究方向为网络与信息系统。

2 线性回归问题简介

先定义一个二分类问题:给定一个训练集 $X = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$, 其中 $y_i \in \{-1, +1\}$, x_i 是一个训练样本向量。对于线性分类器 $f(x) = a^T x + b$ 来说,它试图找到一个权重想象 a 、一个截距 b 、一个阈值 θ ,使得当标注为 -1 时 $f(x) < \theta$, 当标注为 $+1$ 时 $f(x) \geq \theta$ 。当扩展输入向量 x 为 $[1, x]$ 、权重向量 a 为 $[b, a]$ 时,线性分类器可以表示为 $f(x) = a^T x$ 。

线性回归问题试图找到一个线性函数 $f(x) = a^T x$, 能够最小化估计误差:

$$\hat{a} = \arg \min_a \sum_{i=1}^n (y_i - a^T x_i)^2 \quad (1)$$

该式求解结果为

$$\hat{a} = \left(\sum_{i=1}^n x_i x_i^T \right)^{-1} \sum_{i=1}^n x_i y_i$$

在图像检索中,由于相关反馈的样本数有限,因此矩阵 $\sum_{i=1}^n x_i x_i^T$ 可能是奇异的,这将导致目标函数有多个最小解。为了解决矩阵 $\sum_{i=1}^n x_i x_i^T$ 的奇异性问题,一个办法是引入一个 $L2$ 范式,这种方法称为岭回归:

$$\hat{a} = \arg \min_a \sum_{i=1}^n (y_i - a^T x_i)^2 + \lambda \|a\|^2 \quad (2)$$

式中, $\lambda > 0$ 是一个常量。上式的求解结果如下:

$$\hat{a} = \left(\sum_{i=1}^n x_i x_i^T + \lambda I \right)^{-1} \sum_{i=1}^n x_i y_i$$

岭回归方法与独立分量分析(SVD)和主成分分析(PCA)有较密切的关系,但上述方法均基于欧式距离来分析数据,并不能发现数据空间的流形结构。

3 近邻保留回归算法

前面提到了在图像检索中由于反馈信息较少,分类器的学习会存在过适应问题。对于未正则化的模型来说,分类器学习需要的样本数是与 VC 维数线性相关的。VC 维数定义为能被分类模型分散(shatter)的最大样本点数,一般对应模型参数量线性增长^[12],至少对应输入特征数量线性增长。因此,除非标记样本数量大到与图像特征维数相当(一般是几百维),否则一些附加的方法,如正则化方法等就有必要引入到这种环境下的分类器学习过程^[13]。

本节将介绍我们提出的近邻保留回归算法。本算法基于流形学习理论里的谱图理论^[14]及局部线性嵌入理论^[15,16],设计了一个正则算子来保留图像的近邻结构,使之适用于标注样本较少的相关反馈图像检索环境。

3.1 算法设计

假定一个利用用户反馈标注的数据集 $\{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$, 当图像 x_i 被标注为相关,则 $y_i = 1$; 图像 x_i 标注为不相关,则 $y_i = -1$ 。另外假定数据集 $\{(x_{l+1}, y_{l+1}), \dots, (x_{l+m}, y_{l+m})\}$ 表示数据库中未标注数据,总的图像数据集可表示为 $X = \{x_1, x_2, \dots, x_{l+m}\}$ 。我们的任务是估计 y_j 的值 ($j = l+1, \dots, l+m$)。

如前所述,一个关键的问题是相关反馈使得该问题成为一个小样本问题。因此,如果不进行特别处理,将导致过适应

问题。过适应问题的一般解决方法是引入一个正则算子。回归问题引入正则算子后的损失函数一般可以表示为:

$$\ell(f) = \sum_{i=1}^l (y_i - f(x_i))^2 + \lambda J(f) \quad (3)$$

式中, $J(f)$ 用来度量假设模型的复杂性,而 λ 用来实现模型复杂性与观测误差间的平衡。

对于上面假定的图像数据集 X , 首先构建一个邻接图,尽管图像空间可能是一个非线性的流形,但图像的局部近邻结构可以假定是线性的,这样每一个图像都可以由它的近邻通过线性变换来表示,这种方法的重构误差为

$$\phi(W) = \sum_{i=1}^{l+m} \|x_i - \sum_{j=1}^{l+m} W_{ij} x_j\|^2$$

上式将所有图像与它的重构图像间的距离平方求和。需要说明的是,如果图像间距离较远,则 W_{ij} 的值就较小,甚至为 0。现在我们通过 W_{ij} , 利用一个图像的近邻标注的线性变换来估计它自己的标注值。对于设计一个好的分类器来说,一个合理的正则算子如下:

$$J(f) = \sum_{i=1}^{l+m} (f(x_i) - \sum_{j=1}^{l+m} W_{ij} f(x_j))^2 \quad (4)$$

该算子在进行图像的流形空间学习时表现优异^[15], 式中 W_{ij} 为固定的权重值。X. F. He 等人发现在特定情况下 f 是一个线性函数^[16], 即 $f(x) = a^T x$, 于是有

$$J(a) = \sum_{i=1}^{l+m} (a^T x_i - \sum_{j=1}^{l+m} W_{ij} a^T x_j)^2$$

为了简化 $J(a)$, 定义

$$u_i = a^T x_i - \sum_{j=1}^{l+m} W_{ij} a^T x_j = (x_i - \sum_{j=1}^{l+m} W_{ij} x_j)^T a$$

于是,可以得到

$$u = (u_1, \dots, u_{l+m})^T = (X^T - WX^T)a = (I - W)X^T a \text{ 这样, 通过一些代数变换,可以得到 } J(a) \text{ 的简化形式:}$$

$$J(a) = \sum_{i=1}^{l+m} (u_i)^2 = u^T u = a^T X(I - W)^T (I - W)X^T a = a^T XMX^T a$$

式中, $M = (I - W)^T (I - W)$, 显然它是对称的半正定矩阵。设 $y_i = f(x_i)$, $y = (y_1, \dots, y_l)^T$, $Z = (x_1, \dots, x_l)$, 则已标注数据的最小方差可以计算如下:

$$\sum_{i=1}^l (y_i - a^T x_i)^2 = (y - Z^T a)^T (y - Z^T a) = y^T y - 2a^T Z y + a^T Z Z^T a$$

由于 $y^T y$ 是恒定的, 则式(3)表示的损失函数可以转换为:

$$\ell(a) = -2a^T Z y + a^T Z Z^T a + \lambda a^T X M X^T a \quad (5)$$

对 $\ell(a)$ 求导数极值, 可得:

$$\begin{aligned} \frac{\partial \ell}{\partial a} = 0 &\Rightarrow -2a^T Z y + a^T Z Z^T a + \lambda a^T X M X^T a = 0 \\ &\Rightarrow a = (Z Z^T + \lambda X M X^T)^{-1} Z y \end{aligned} \quad (6)$$

一旦设计好回归函数 f , 图像 x_i 的标签就可以通过 $f(x_i) = a^T x_i$ 来估计。于是, 可以将数据库中的未标注图像通过 $f(x_i)$ 进行标注并排序。最后, 将靠前的 N 个图像作为检索结果返回。

算法可以归纳如下。

(1) 构建一个近邻图。

对于所有的 $l+m$ 个图像,对于每一个图像 x_i ,找到它的 k 近邻,然后在图像 x_i 和它的近邻之间做一个有向边连接。

(2)选取权重 W_{ij} 。

当图像 i, j 之间没有连接边时,定义 $W_{ij}=0$ 。当有连接边时,权重的取值可以通过最小化下面的目标函数来得到:

$$\min_W \sum_{i=1}^{l+m} \left\| x_i - \sum_{j=1}^{l+m} W_{ij} x_j \right\|^2$$

上式约定条件为

$$\sum_{j=1}^{l+m} W_{ij} = 1, i=1, 2, \dots, l+m$$

上述问题可以通过求解一个最小方差得到一个优化解,具体见文献[15]。

(3)回归函数的计算。

可以通过如下计算得到线性分类器:

$$f(x) = a^T x$$

先定义

$$M = (I - W)^T (I - W), I = \text{diag}(1, \dots, 1)$$

求解下面的优化问题:

$$\min_a \sum_{i=1}^l (y_i - a^T x_i)^2 + \lambda a^T X M X^T a$$

结果为:

$$a = (Z Z^T + \lambda X M X^T)^{-1} Z y$$

式中, $y = (y_1, \dots, y_l)^T, Z = (x_1, \dots, x_l)$ 。

(4)在数据库中标注图像。

对于数据库中的未标注图像 ($i = l+1, \dots, l+m$),通过 $f(x_i) = a^T x_i$ 来进行标注并排序,将靠前的 N 个图像作为检索结果返回。

3.2 模型设计

在我们的算法中,有两个参数:一个是近邻数 k ,另一个是正则参数 λ 。 k 定义局部的程度,而 λ 实现局部几何结构与观测误差间的平衡。当 $\lambda=0$ 时,算法就成为纯粹的线性回归算法。当 $\lambda \rightarrow \infty$ 时,算法则忽略了观测误差,而仅仅得到一个最好保留局部几何结构的函数。因此,一个适当的 λ 取值需要在二者间得到一个平衡。

一般而言,模型的设计可以采用基于分析的方法或基于重采样的方法^[13]。基于分析的方法包括 Akaike 信息准则、Bayes 信息准则、最小描述长度准则等。基于重采样的方法则包括交叉确认、Bootstrap 算法等。由于并没有一个确切的概率估计,因此将使用交叉确认的方法来确定参数 λ 的取值。

对于参数 k ,实验中发现 k 的取值较小(小于 15)时实验效果最好。而且,实验效果的变化对于 k 的取值变化是很敏感的。

4 实验分析

4.1 实验设计

将对以下几种算法进行对比实验:

- (1)岭回归算法(简称 RR);
- (2)基于支持向量机的算法(简称 SVM);
- (3)保局投影算法(简称 LPP);
- (4)扩展相关嵌入算法(简称 ARE);

(5)我们提出的近邻保留回归算法(简称 NPR)。

上面提到的几种对比算法中,SVM 算法的原理是试图使相关图像与不相关图像间的分隔距离最大化,本实验使用的 SVM 算法采用了一个线性核。RR 算法是标准线性回归算法结合 L2 范式后的扩展方法,它通过反馈信息学习一个分类函数 $f(x)$,当数据库中的 x_i 使得 $f(x_i)$ 更大时,表示 x_i 与用户请求图像的相关性更大。LPP 算法是一种经典的线性流形学习算法,它试图找到一个线性映射 $y_i = A x_i$,将高维图像映射到一个能保留图像局部几何结构的低维子空间中,子空间的样本间距离定义为

$$\begin{aligned} \text{dist}(y_i, y_j) &= \sqrt{(y_i - y_j)^T (y_i - y_j)} \\ &= \sqrt{(x_i - x_j)^T A^T A (x_i - x_j)} \end{aligned}$$

ARE 算法是 LPP 算法的半监督扩展,它能同时使用标注及未标注数据来学习一个嵌入函数,因而综合性能优于 LPP 算法。当所有的反馈样本均标注为“相关”时,该算法将简单地计算反馈样本的质心,并根据距质心的距离来进行数据分类。

与 RR 算法及 SVM 算法相比,LPP 算法的主要缺点是不能有效使用用户标注数据,而且对图像相关性特征不能很好地感知。RR 算法及 SVM 算法能有效感知图像的相关性分类特征,但不能使用未标注数据。我们提出的算法可以通过使用标注后的数据来最小化观测误差,同时也可以使用未标注数据来发现局部的流形结构。在对比实验中,为了减少运行时间,所有算法均使用线性形式。当然,如果通过核函数变换,这些算法同样也适应非线性情况。

实验使用的图像数据库是从 Corel 图库选出的包括 79 个语义类的 7900 个图像,它们构成一个大型多种类的图集。在系统中使用了 3 种颜色特征和 3 种纹理特征,特征空间的维度是 435,如表 1 所列。

表 1 图像检索实验中使用的图像特征

1# 颜色特征	HSV 颜色空间量化到 256 维颜色直方图
2# 颜色特征	LAB 空间的一阶矩和二阶矩
3# 颜色特征	LUV 颜色空间量化到 64 维的一致特征
1# 纹理特征	Tamura 稀疏度特征
2# 纹理特征	Tamura 方向性特征
3# 纹理特征	金字塔结构小波变换纹理特征

采用文献[17]使用的自动反馈方案来模拟真实用户操作下的检索过程。在每一次循环中,系统将标注前 10 个返回结果为“相关”或“不相关”的样本,再结合数据库中的未标注样本,一起参与分类器的学习。上一轮循环标注的样例将排除在下一轮选择之外。为了评价算法的性能,定义了检索准确度:

$$\text{检索准确度} = (\text{头 } N \text{ 个检索结果中相关图像的个数}) / N$$

4.2 实验结果

进行了近邻保留回归算法与其它 4 种算法的比较实验。对于 SVM 算法,使用了 5-分类交叉确认方法来估计正则化参数^[18]。对于 LPP 及 ARE 算法,设置子空间的维数为数据集的图像类别数(79)。对于近邻保留回归算法,取参数 k 和 λ 的值为 7 和 0.1。下一节还将介绍参数选择对实验结果的

影响。随机在每一类中选择 10 幅图像作为用户请求。图 1 显示的是随反馈次数增大, 头 20 个返回结果的准确度变化情况。图 2 显示的是在第 1、2 轮反馈后不同测量范围(前 10~100 个返回结果)中的准确度。

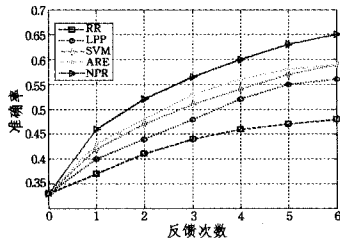
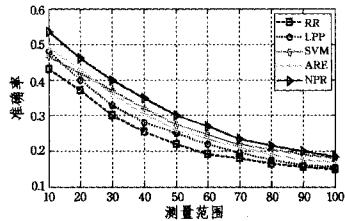
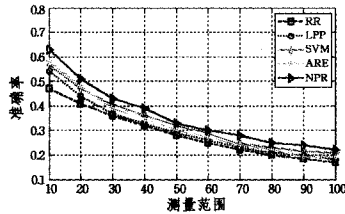


图 1 前 20 个返回结果中的准确率



(a) 第一轮反馈后



(b) 第二轮反馈后

图 2 前 20 个返回结果中的准确率

从比较实验结果中可以得到以下结论。

(1) RR 算法表现最差, 可能是因为它不能发现数据集的内在结构; LPP 算法优于 RR 算法, 可能是因为它通过降维减少了数据的复杂度; SVM 优于 LPP 算法, 可能是它通过最大化分类距离降低了数据复杂度的同时还具有更强的分类特征感知能力; ARE 算法在 LPP 基础上增加了利用未标注数据的能力, 因而表现更好; 我们提出的 NPR 算法可以同时使用已标注及未标注数据, 又通过引入一个正则化因子使算法具有流形学习能力, 因而表现最好。

(2) 随着相关反馈次数的增多, 5 种算法的准确度均明显提升, 表明了相关反馈技术在 CBIR 系统中的重要作用。

结束语 具有相关反馈机制的 CBIR 检索过程实际上是一个半监督学习过程。由于大多数时候反馈信息量远小于数据维数, 这会导致过适应问题的出现。提出一种新的近邻保留回归算法来解决这一问题。通过引入一个正则项, 本算法可以同时使用已标注及未标注数据, 并能获得一个能保留图像近邻结构的分类器。实验结果显示, 近邻保留回归算法优于岭回归算法及基于 SVM 及 LPP 的类似算法。

参考文献

[1] 王守觉, 孙华, 柳培忠, 等. 基于仿生形象思维方法的图像检索算法[J]. 电子学报, 2010, 38(5): 993-997

[2] He X. Laplacian regularized d-optimal design for active learning and its application to image retrieval[J]. IEEE Transactions on Image Processing, 2010, 19

[3] Rui Y, Huang T S, Ortega M, et al. Relevance feedback: a power tool for interactive content based image retrieval [J]. IEEE Transactions on Circuit and Systems for Video Technology, 1998, 8: 644-655

[4] 许相莉, 张利彪, 刘向东, 等. 基于粒子群的图像检索相关反馈算法[J]. 电子学报, 2010, 38(8)

[5] He X, Ma W-Y, Zhang H-J. Learning an image manifold for retrieval[C]//Proceedings of the ACM Conference on Multimedia. New York, 2004: 17-23

[6] Si L, Jin R, Hoi S C H. Collaborative image retrieval via regularized metric learning[J]. Multimedia Systems, 2006, 12(1): 34-44

[7] 周刚, 邬义杰, 宋德玉, 等. 基于 MMP 三角曲面测地线算法研究[J]. 中国图象图形学报, 2010, 15(8): 1260-1268

[8] Zhou Z-H, Chen K-J, Dai H-B. Enhancing relevance feedback in image retrieval using unlabeled data[C]//ACM Transactions on Information Systems. 2006, 24: 219-244

[9] 魏莱, 王守觉. 基于流形距离的半监督判别分析[J]. 软件学报, 2010, 21(10): 2445-2453

[10] He X, Ji M, Bao H. A unified active and semi-supervised learning framework for image compression [C] // Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). Miami, FL, 2009

[11] Belkin M, Niyogi P, Sindhwani V. Manifold regularization: a geometric framework for learning from examples[J]. Journal of Machine Learning Research, 2006, 7: 2399-2434

[12] Vapnik V. Estimation of Dependences Based on Empirical Data [M]. Springer-Verlag, 1982

[13] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction [M]. New York: Springer-Verlag, 2001

[14] Chung F R K. Spectral Graph Theory [C] // Regional Conference Series in Mathematics. vol. 92, 1997

[15] Roweis S, Saul L. Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000, 290: 2323-2326

[16] He X, Cai D, Yan S, et al. Neighborhood preserving embedding [C] // IEEE International Conference on Computer Vision. Beijing, China, 1208-1213

[17] He X. Incremental semi-supervised subspace learning for image retrieval [C] // Proceedings of the ACM Conference on Multimedia. New York, 2-8

[18] Vapnik V N. The Nature of Statistical Learning Theory [M]. Springer-Verlag, 1995