·计算机工程与应用·

新型的图像检索最优实验设计算法

鲁 珂，赵继东，吴 跃

(电子科技大学计算机科学和工程学院 成都 610054)

【摘要】大部分现有的最优实验设计方法是基于线性回归或拉普拉斯正则最小二乘模型(LapRLS)的。提出一种基于二阶Hessian能并具有流形学习能力的主动学习算法，该算法选择那些能使Hessian正则回归模型的参数协方差矩阵最小化的样本作为最优样本，可以克服LapRLS的依赖特定常量及缺乏推算能力等缺点。基于内容的图像检索实验证明了该方法的有效性。

关 键 词 图像检索; LapRLS; 流形学习; 最优实验设计

中图分类号 TP391.4 文献标识码 A doi:10.3969/j.issn.1001-0548.2012.02.019

Novel Optimal Experimental Design Algorithm for Image Retrieval

LU Ke, ZHAO Ji-dong, and WU Yue

(School of Computer Science and Engineering, University of Electronic Science and Techonology of China Chengdu 610054)

Abstract Most of the existing optimal experimental design (OED) methods are based on either linear regression model or Laplacian regularized least square (LapRLS) model. This paper proposes a new active learning algorithm based on the second-order Hessian energy, which has the manifold learning capability. The algorithm selects those optimal samples which minimize the parameter covariance matrix of the Hessian regularized regression model, and overcomes the drawbacks of LapRLS. The experimental results on content-based image retrieval have demonstrated the effectiveness of the proposed approach.

Key words image retrieval; LapRLS; manifold learning; optimal experimental design

目前，由于多媒体数据表示的日益普及，图像数据库的数据量剧增。在图像检索等应用环境中，未标注数据的数量也日渐庞大，使得完全标注这些数据的代价极其昂贵，因而很难实现。针对该问题，一个有效而且可行的方法是主动选取那些信息量较大的数据进行标注，就可以在标注数据量较少的情况下尽量获得更多的数据特征，从而提高机器学习能力，该方法被称为主动学习(active learning)[1]。

在统计学领域，选择最优样本(含有最多分类信息的样本)的方法称为最优实验设计。最优实验设计的目标是通过选择最优样本使模型的输出方差、参数方差、预测误差最小化[2-3]。最优实验设计的性能依赖于统计模型及相关的一个统计标准，不同的统计模型和统计标准会产生不同的处理效果。典型的最优实验设计方法包括A-最优设计、D-最优设计、E-最优设计等。这3种方法均基于最小二乘回归模型并力图使参数协方差矩阵最小化，它们的差别在于使用不同的标准来度量协方差矩阵的大小[2]。近年来，很多研究表明现实世界的数据在很多情况下归属于内嵌在高维欧氏空间里的低维流形[4-6]，而传统的回归模型并不能很好地处理这类数据。文献[7]提出了一种流形正则化处理方法，称为拉普拉斯最小二乘正则方法(LapRLS)，该方法通过构建一个近邻图来建立流形模型，并在最小二乘损失函数中引入拉普拉斯图作为一个正则化因子。LapRLS在如图像检索[8]、查询分类[9]等很多应用中取得了很好的效果，但该方法的主要不足是其依赖于特定的常量而且缺乏推断能力[10]。

为了在发现数据流形特征基础上选择最优样本，本文基于Hessian正则化[10]提出了一种新型的主动学习算法：Hessian最优设计算法。与拉普拉斯正则因子不同，Hessian正则因子可以得到一个对应测地距离线性变化的函数，该属性在图像检索中评判图像与用户请求的相关度时具有重要的作用。在图

收稿日期：2011 – 02 – 18; 修回日期：2011 – 05 – 13

基金项目：国家自然科学基金(60702072)；中央高校基本业务费(ZYGX2009X012)；四川省应用基础研究项目(2010JY0001)

作者简介：鲁珂(1974 –)，男，博士，副教授，主要从事图像识别、网络多媒体技术方面的研究..

像检索中，用户反馈的“相关”或“不相关”被用来训练分类器，然而，很难找到一个函数能够线性连续地将数据库的图像分为“相关”与“不相关”两个区间。为了选择最优样本进行标注，首先得到Hessian正则最小二乘模型的参数协方差矩阵，然后定义最优样本是那些能使参数协方差矩阵的最大特征值最小的样本。

Hessian最优设计最适合的应用场合是结合相关反馈的图像检索[11-13]。当用户初次提出查询请求时，系统根据一个预定义的距离矩阵对图像进行排序，返回排位靠前的图像给用户；用户会被要求就部分返回结果提供相关反馈。大部分系统简单地要求用户标注最靠前的一些结果，但理想的方案是能够选择哪些最具分类信息的结果让用户标注，这样可以更好地优化分类器，从而有效提高检索精度。本文将重点讨论如何用该Hessian最优设计算法来完成这项工作。

1 Hessian最优设计算法

本节将详细介绍主动Hessian最优设计算法。

1.1 Hessian正则化

为了在学习过程中具有流形学习能力，构造一个基于流形的正则化函数，Hessian正则化[10]即是这样一种方法，本文的工作将使用它作为研究基础。

文献[10]提出了Hessian能表达式为：

$$S_{\mathrm{Hess}}(\boldsymbol{f})=\int_M \|\nabla_a\nabla_b \boldsymbol{f}\|^2_{T_X^*M\otimes T_X^*M}\,\mathrm{d}V(x) \tag{1}$$

式中，$\nabla_a\nabla_b \boldsymbol{f}$是$\boldsymbol{f}$的二阶协变导数；d$V(x)$是体积元素。使用标准坐标，可以得到：

$$\|\nabla_a\nabla_b \boldsymbol{f}\|^2_{T_X^*M\otimes T_X^*M}=\sum_{r,s=1}^{m}\left(\frac{\partial^2 \boldsymbol{f}}{\partial x_r\partial x_s}\right)^2 \tag{2}$$

可见，$\boldsymbol{f}$的二阶协变导数恰好是在标准坐标下Hessian正则的Frobenius范式表达。与拉普拉斯正则化依赖常量函数且不具推测能力相比，当测地线函数存在时，Hessian正则化并不依赖常量函数且具有线性的推测能力。在处理半监督回归问题时，当回归函数沿流形平滑甚至是线性变化时，Hessian正则化的这个优点将特别有用。

用$N_k(\mathrm{x}_i)$表示x_i的k近邻，可以得到一个在$N_k(x_i)$上的函数值$f(x_j)$，这样，在x_i上的$\boldsymbol{f}$的Hessian正则化可以近似为：

$$\left.\frac{\partial^2 \boldsymbol{f}}{\partial x_r\partial x_s}\right|_{x_i}\approx\sum_{j=1}^{k}H_{rsj}^{(i)}f(x_j) \tag{3}$$

这样，使用最小二乘法求解一个二阶多项式可以得到$\boldsymbol{H}$算子。设$f_i=f(x_i)$，$\boldsymbol{f}=(f_1, f_2,\cdots,f_k)^{\mathrm{T}}$，在$x_i$上的$\boldsymbol{f}$的Hessian正则化的Frobenius范式的估计可以表示为：

$$\|\nabla_a\nabla_b \boldsymbol{f}\|^2\approx\sum_{r,s=1}^{m}\Big(\sum_{\alpha=1}^{k}H_{rs\alpha}^{(i)}f_\alpha\Big)^2=\sum_{\alpha,\beta=1}^{k}f_\alpha f_\beta B_{\alpha\beta}^{(i)} \tag{4}$$

式中，$B_{\alpha\beta}^{(i)}=\sum_{r,s=1}^{m}H_{rs\alpha}^{i}H_{rs\beta}^{i}$。于是，可以得到Hessian能的最终表示为：

$$\sum_{i=1}^{n}\sum_{r,s=1}^{m}\left(\left.\frac{\partial^2 \boldsymbol{f}}{\partial x_r\partial x_s}\right|_{x_i}\right)^2=\sum_{i=1}^{n}\sum_{\alpha\in N_k(x_i)}\sum_{\beta\in N_k(x_i)}^{k}f_\alpha f_\beta B_{\alpha\beta}^{(i)}=\boldsymbol{f}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{f} \tag{5}$$

式中，$\boldsymbol{B}=\sum_{i=1}^{n}B_{\alpha\beta}^{(i)}$；$n$是样本的数量。

1.2 目标函数

Hessian正则化回归在直推学习中表现出了很好的效果，它的损失函数为：

$$\min_{\boldsymbol{f}\in R_k}\sum_i(f_i-y_i)^2+\lambda\boldsymbol{f}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{f} \tag{6}$$

式中，y_i为第i个被标注的数据。目前的直推式学习方法还不能得到一个任何场合都适用的目标函数，因而不具有对于未知样本的推测能力。为解决样本外预测问题，先设置一个线性函数$f(\boldsymbol{x}_i)=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}$，设$\boldsymbol{Z}=(z_1,z_2,\cdots,z_k)$表示已标注的样本数据点，$\boldsymbol{X}=(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_m)$表示全部的样本数据点，显然，$\boldsymbol{f}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{w}$，于是损失函数可以重新表示为：

$$\arg\min_{\boldsymbol{w}}\sum_{i=1}^{n}(y_i-\boldsymbol{w}^{\mathrm{T}}z_i)^2+\lambda_1\boldsymbol{w}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{B}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{w}+\lambda_2\|\boldsymbol{w}\|^2 \tag{7}$$

设$\boldsymbol{y}=(y_1,y_2,\cdots,y_k)^{\mathrm{T}}$，$M=\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}}+\lambda_1\boldsymbol{X}\boldsymbol{B}\boldsymbol{X}^{\mathrm{T}}+\lambda_2\boldsymbol{I}$，$\Lambda=\lambda_1\boldsymbol{X}\boldsymbol{B}\boldsymbol{X}^{\mathrm{T}}+\lambda_2\boldsymbol{I}$，通过简单的代数变换，可以得到$\boldsymbol{w}$的表达式为：

$$\hat{\boldsymbol{w}}=M^{-1}\boldsymbol{Z}\boldsymbol{y} \tag{8}$$

于是，参考拉普拉斯最小二乘化正则方法，可以得到$\boldsymbol{w}$的偏离值及协方差分别为：

$$E(\hat{\boldsymbol{w}}-\boldsymbol{w})=-M^{-1}\Lambda\boldsymbol{w} \tag{9}$$

$$\mathrm{Cov}(\hat{\boldsymbol{w}})=\sigma^2(M^{-1}-M^{-1}\Lambda M^{-1}) \tag{10}$$

式中，σ^2是线性模型的方差。如前所述，有很多度量参数协方差矩阵大小的方法，包括A-优化、D-优化、E-优化等。对于D-优化方法，当样本分布沿椭圆曲线靠近坐标轴时，将使得参数估计能力变差，方差也将过大。本文提出了一种基于E-优化的方法，将主动学习问题表示为：

$$\min_{\{z_1,z_2,\cdots,z_k\}\subset\chi}\lambda_{\max}(M^{-1}-M^{-1}\Lambda M^{-1}) \tag{11}$$

式中，$\lambda_{\max}(\boldsymbol{A})$是矩阵$\boldsymbol{A}$的最大特征值。

1.3 优化问题的求解

本节将讨论如何对式(11)表示的优化问题进行求解。实际上，正则化参数λ_1、λ_2的值通常很小，因此，可以假定$M^{-1}-M^{-1}\Lambda M^{-1}\approx M^{-1}$，因此，式(11)可以简化为：

$$\min_{\{z_1,z_2,\cdots,z_k\}\subset\chi}\lambda_{\max}((\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}}+\lambda_1\boldsymbol{X}\boldsymbol{B}\boldsymbol{X}^{\mathrm{T}}+\lambda_2\boldsymbol{I})^{-1}) \tag{12}$$

由于组合特性(combinatorial nature)的存在，上式很难求解。考虑引入一个权重向量$\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\cdots,\alpha_m)$，$\alpha_i\geqslant 0$表示$x_i$的相对代表性的估计值，因此，式(12)可以变形为：

$$\min_{\boldsymbol{\alpha}}\lambda_{\max}((\sum_{i=1}^{m}\alpha_i\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}}+\lambda_1\boldsymbol{X}\boldsymbol{B}\boldsymbol{X}^{\mathrm{T}}+\lambda_2\boldsymbol{I})^{-1})$$
$$\alpha_i\geqslant 0,\ \sum_i\alpha_i=1 \tag{13}$$

上式表示的目标函数是$\boldsymbol{\alpha}$的凸函数，对于半正定矩阵$\boldsymbol{A}$，$\boldsymbol{A}^{-1}$的最大特征值等于$\boldsymbol{A}$的最小特征值的倒数。因此，式(13)等价于：

$$\max_{\boldsymbol{\alpha}}\lambda_{\min}(\sum_{i=1}^{m}\alpha_i\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}}+\lambda_1\boldsymbol{X}\boldsymbol{B}\boldsymbol{X}^{\mathrm{T}}+\lambda_2\boldsymbol{I})$$
$$\alpha_i\geqslant 0,\ \sum_i\alpha_i=1 \tag{14}$$

式(14)的优化问题也可以表述为如下的半定规划问题(SDP)：

$$\max_{\boldsymbol{\alpha},t}t\sum_{i=1}^{m}\alpha_i\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}}+\lambda_1\boldsymbol{X}\boldsymbol{B}\boldsymbol{X}^{\mathrm{T}}+\lambda_2\boldsymbol{I}\geqslant t\boldsymbol{I}$$
$$\boldsymbol{\alpha}\geqslant 0,\ \mathbf{1}^{\mathrm{T}}\boldsymbol{\alpha}=1 \tag{15}$$

式中，$\boldsymbol{\alpha}\in\mathbb{R}^m$；$t\in\mathbb{R}$；$\boldsymbol{A}\geqslant\boldsymbol{B}$表示$\boldsymbol{A}$–$\boldsymbol{B}$是半正定的。可以得到式(14)与式(15)表示的优化问题是等价的结论。

证明：设α_a^*是式(14)的解，(α_b^*,t)是式(15)的解，为了证明$\alpha_a^*=\alpha_b^*$，定义：

$$g(\boldsymbol{\alpha})=\sum_{i=1}^{m}\alpha_i\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}}+\lambda_1\boldsymbol{X}\boldsymbol{B}\boldsymbol{X}^{\mathrm{T}}+\lambda_2\boldsymbol{I}$$

采用反证法，假设$\alpha_a^*\neq\alpha_b^*$成立，由于α_a^*是最大化优化问题的解，则有：

$$\lambda_{\min}(\boldsymbol{g}(\alpha_a^*))>\lambda_{\min}(\boldsymbol{g}(\alpha_b^*))$$

由于(α_b^*,t)满足式(15)的约束条件，则有$\boldsymbol{g}(\alpha_b^*)\succ t^*\boldsymbol{I}$，说明$\boldsymbol{g}(\alpha_b^*)-t^*\boldsymbol{I}$是半正定的。对于任意一个矩阵$\boldsymbol{A}$，如果$\lambda$是矩阵$\boldsymbol{A}$的特征值，那么$\lambda$–$c$是$\boldsymbol{A}$–$c\boldsymbol{I}$的特征值($\boldsymbol{I}$是单位矩阵)；于是可知，$\lambda_{\min}(\boldsymbol{g}(\alpha_b^*))-t^*$是半正定矩阵$\boldsymbol{g}(\alpha_b^*)-t^*$的最小特征值，因此可得：$\lambda_{\min}(\boldsymbol{g}(\alpha_b^*))\geqslant t^*$。定义$\Delta t=\lambda_{\min}(\boldsymbol{g}(\alpha_a^*))-\lambda_{\min}(\boldsymbol{g}(\alpha_b^*))>0$，于是可得：

$$\lambda_{\min}(\boldsymbol{g}(\alpha_a^*))-(t^*+\Delta t)=$$
$$\lambda_{\min}(\boldsymbol{g}(\alpha_a^*))-(t^*+\lambda_{\min}(\boldsymbol{g}(\alpha_a^*))-\lambda_{\min}(\boldsymbol{g}(\alpha_b^*)))=$$
$$\lambda_{\min}(\boldsymbol{g}(\alpha_b^*))-t^*\geqslant 0$$

这就表示矩阵$\boldsymbol{g}(\alpha_a^*)$的所有特征值均比$t^*+\Delta t$大，即$\boldsymbol{g}(\alpha_a^*)\succ t+\Delta t$，显然$(\alpha_a^*,t^*+\Delta t)$将是一个比$(\alpha_b^*,t)$更优的解，与初始定义矛盾，因而$\alpha_a^*\neq\alpha_b^*$不成立，于是结论得证。

得到优化解$\boldsymbol{\alpha}$后，就可以选择α_i最大的数据点作为最优样本来进行标注。

2 实验结果

近年来，相关反馈技术是图像检索领域的一个研究热点。图像数据一般按照颜色、纹理、形状等特征表示为一个向量；当用户向系统提交一个请求图像后，系统首先按预定的一个距离矩阵计算数据库中图像与请求图像的距离；然后按照距离大小对数据库图像进行排序并返回排位靠前的一些图像；用户这时会被要求对图像进行“相关”或“不相关”的标注，这些标注后的图像会被用作训练集来生成并优化分类器。分类器可以用来预测图像的相关性并对图像进行重新排序，这样的过程可重复进行直到用户满意为止。

大部分常规的检索系统一般均选择最靠前的图像让用户标注，但这些图像并不一定是包含分类信息最多的图像，因此也并不一定是最能提升分类器性能的图像。有时甚至靠前的图像均被标注为“相关”，这时会由于只有一类标注样本，导致分类器不能得到优化。最优实验设计或主动学习方法可以用来解决该问题，通过在返回样本中选择最优样本(包含分类信息最多的样本)进行标注，能更有效地提升分类器的性能。

本文的实验采用颜色柱状图特征及颜色纹理矩(CTM)[14]来表示一幅图像，颜色柱状图采用HSV空间的4×4×4特征，共64维；CTM结合了图像的颜色及纹理特征并压缩为64维向量。因此，本文使用一个128维向量表示图像。本文使用的图像数据来自Corel数据库，共79类，每类100幅图片，参加实验的图片共7 900幅。

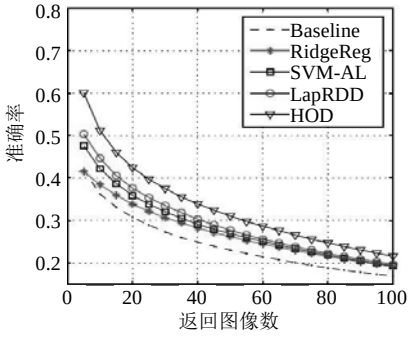
下面对5个算法进行对比测试。

1) 基准算法(Baseline)：不使用相关反馈，图像依据欧式距离排序；2) 岭回归算法(RidgeReg)：用户标记最靠前的图像，标记数据用于训练一个岭回归分类器模型；3) SVM主动学习算法(SVM-AL)[15]：系统选择最靠近分类边界的图像让用户标注，标记

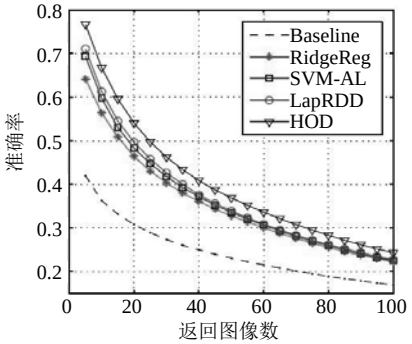
数据用于训练一个SVM分类器模型；4) 拉普拉斯正则化D-优化算法(LapRDD)[3]：选择图像时考虑到图像数据空间的局部几何结构，基于拉普拉斯正则算子来选择图像；5) Hessian优化设计算法(HOD)：本文的基于Hessian正则化来选择图像进行标注的算法(算法中参数λ_1、λ_2取值为0.1)。本文的实验设计与文献[3]类似，每个图像类别被平均分为5个子集，一个子集用作请求图像，其他子集分别组成4个检索数据库，最后按4个库的检索准确率取平均值来作为实验结果。

实验中的基本结果是每次返回结果中的准确率数据；但为了分析算法性能，除了统计在返回结果数量不同时(N=5～100)准确率的变化情况，还要统计随着反馈次数增大(最大到4次)时准确率的变化情况。每一次反馈，用户会被要求标注8个返回结果，RidgeReg算法是标注最靠前的8个结果，而SVM-AL、LapRDD、HOD均是利用主动学习方法选择8个结果让用户标注。

实际应用中，用户一般不愿意多次反馈信息，因此，前两次的反馈效果十分重要。图1显示了5种算法在第1轮(图1a)及第2轮(图1b)反馈后随返回结果数增加(每次加5)的准确率变化曲线。



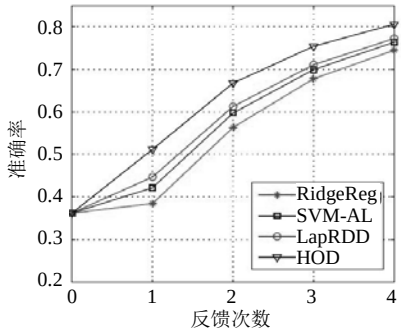
a. 第1轮反馈



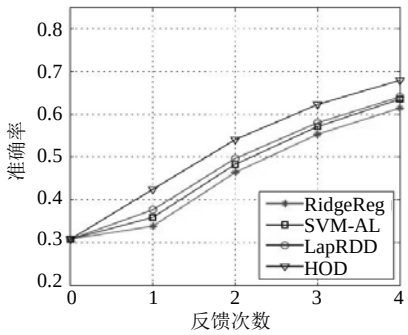
b. 第2轮反馈

图1 不同返回结果数的准确率变化

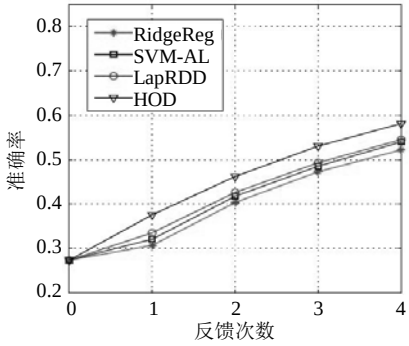
图2显示了5种算法随着0～4轮反馈数增加的准确率变化曲线，图2a～图2d分别表示返回结果数为10、20、30、50时的实验结果。



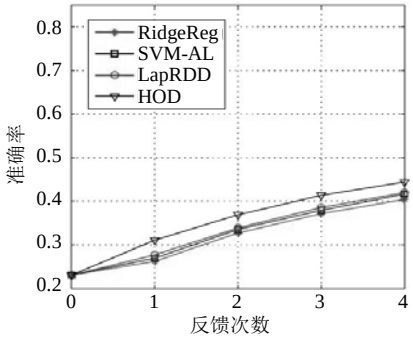
a. 返回结果数为10



b. 返回结果数为20



c. 返回结果数为30



d. 返回结果数为50

图2 不同反馈次数的准确率变化

从实验结果可以看出，在不同情况下，算法性能的排序均稳定为HOD、LapRDD、SVM-AL、RidgeReg。这与目前相关研究文献的研究是基本符合的，一般认为，具有主动学习机制的SVM-AL算法性能优于不具有主动学习能力的RidgeReg方法，

而结合了流形学习的LapRDD主动学习算法优于SVM-AL算法。注意到，在LapRDD、SVM-AL、RidgeReg的检索准确度差异不大(一般不超过2%)的情况下，本文的HOD算法较排第二位的LapRDD提升准确度一般都超过约3%以上。显然，这是由于本文的算法基于Hessian正则化来选择图像，使主动学习具有了推断能力，可以利用数据库中大量的未标注数据，因而可以最大限度地获得分类信息，从而有效提升分类器性能。

3　总　结

本文提出了一种基于Hessian正则化的主动学习算法：Hessian最优设计算法，与拉普拉斯正则化相比，Hessian正则化的主要优点在于其函数是在流形空间中基于测地线距离线性变化的，这个特点使得Hessian能表达式在设计一个回归模型时非常有效。基于Hessian正则化回归模型，使回归模型的参数协方差矩阵最小的数据选择为最优样本。将Hessian最优设计算法用于具有相关反馈机制的图像检索系统，基于Corel数据库的实验结果显示，与当前表现最好的几种主动学习方法相比，Hessian最优设计算法具有更好的性能。

参　考　文　献

[1] ZHANG Q, SUN S. Multiple-view multiple-learner active learning[J]. Pattern Recognition, 2010, 43(9): 3113-3119.

[2] ATKINSON A C, DONEV A N. Optimum experimental designs[M]. Oxford: Oxford University Press, 2007.

[3] HE X. Laplacian regularized d-optimal design for active learning and its application to image retrieval[J]. IEEE Transactions on Image Processing, 2010, 19(1): 254-263.

[4] TENENBAUM J, de SILVA V, LANGFORD J. A global geometric framework for nonlinear dimensionality reduction[J]. Science, 2000, 290(5500): 2319-2323.

[5] HE X, JI M, BAO H. Graph embedding with constraints[C]//Proceedings of the International Joint Conference on Artificial Intelligence. Pasadena, CA: [s.n.], 2009.

[6] LI X, LIN S, YAN S, et al. Discriminant locally linear embedding with highorder tensor data[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B, 2008, 38(2): 342-352.

[7] 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.

[8] CAI D, HE X, HAN J. Regularized regression on image manifold for retrieval[C]//Proceedings of the Ninth ACM SIGMM International Workshop on Multimedia Information Retrieval. Augsburg, Germany: [s.n.], 2007.

[9] HE X, JHALA P. Regularized query classification using search click information[J]. Pattern Recognition, 2008, 41 (7): 2289-2297.

[10] KIM K I, STEINKE F, HEIN M. Semi-supervised regression using hessian energy with an application to semi-supervised dimensionality reduction[C]//Advances in Neural Information Processing Systems. Vancouver, Canada: [s.n.], 2009.

[11] HOI C, LYU M R, JIN R. A unified log-based relevance feedback scheme for image retrieval[J]. IEEE Transactions on Knowledge and Data Engineering, 2006, 18(4): 509-524.

[12] 鲁珂, 赵继东, 吴跃, 等. 基于保局投影的相关反馈算法[J]. 计算机辅助设计与图形学学报, 2007, 19(1): 20-24.
LU Ke, ZHAO Ji-dong, WU Yue, et al. Relevance feedbacks algorithm based on locality preserving projections[J]. Journal of Computer-aided Design & Computer, 2007, 19(1): 20-24.

[13] TAO D, LI X, MAYBANK S. Negative samples analysis in relevance feedback[J]. IEEE Transactions on Knowledge and Data Engineering, 2007, 19(4): 568-580.

[14] YU H, LI M, ZHANG H J, et al. Color texture moments for content-based image retrieval[C]//International Conference on Image Processing. [S.l.]: [s.n.], 2002: 24-28.

[15] TONG S, CHANG E. Support vector machine active learning for image retrieval[C]//Proceedings of the Ninth ACM International Conference on Multimedia. Ottawa, Canada：ACM, 2001: 107-118.

编　辑　漆　蓉