基于自适应稀疏邻域重构的无监督主动学习算法

吕巨建[1,2] 赵慧民[1,2] 陈荣军[1] 李键红[3]
(广东技术师范学院 广州 510665)[1] (广州数字内容处理及其安全性技术重点实验室 广州 510665)[2]
(广东外语外贸大学语言工程与计算实验室 广州 510006)[3]

摘 要 在很多信息处理任务中,人们容易获得大量的无标签样本,但对样本进行标注是非常费时和费力的。作为机器学习领域中一种重要的学习方法,主动学习通过选择最有信息量的样本进行标注,减少了人工标注的代价。然而,现有的大多数主动学习算法都是基于分类器的监督学习方法,这类算法并不适用于无任何标签信息的样本选择。针对这个问题,借鉴最优实验设计的算法思想,结合自适应稀疏邻域重构理论,提出基于自适应稀疏邻域重构的主动学习算法。该算法可以根据数据集各区域的不同分布自适应地选择邻域规模,同步完成邻域点的搜寻和重构系数的计算,能在无任何标签信息的情况下较好地选择最能代表样本集分布结构的样本。基于人工合成数据集和真实数据集的实验表明,在同等标注代价下,基于自适应稀疏邻域重构的主动学习算法在分类精度和鲁棒性上具有较高的性能。

关键词 主动学习,稀疏重构,优化实验设计,直推式实验设计,局部线性重构
中图法分类号 TP181 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.06.045

Unsupervised Active Learning Based on Adaptive Sparse Neighbors Reconstruction

LV Ju-jian[1,2] ZHAO Hui-min[1,2] CHEN Rong-jun[1] LI Jian-hong[3]
(Guangdong Polytechnic Normal University, Guangzhou 510665, China)[1]
(Key Laboratory of Guangzhou Digital Content Processing and Security Technology, Guangzhou 510665, China)[2]
(Language Engineering and Computing Laboratory, Guangdong University of Foreign Studies, Guangzhou 510006, China)[3]

Abstract In many information processing tasks, individuals are easy to get a lot of unlabeled data, but labeling the unlabeled data is quite time-consuming and usually expensive. As an important learning method in the field of machine learning, active learning reduces the cost of labeling data by selecting the most information data points to label. However, most of the existing active learning algorithms are supervised method based on the classifier, not suitable for the sample selection problem without any label information. Aiming at this problem, a novel unsupervised active learning algorithm was proposed, called active learning based on adaptive sparse neighbors reconstruction, by learning from the optimal experiment design and combining the adaptive sparse neighbors reconstruction. The proposed algorithm adaptively selects the neighborhood scale according to different regional distribution of dataset, searches the sparse neighbors and calculates the reconstruct coefficients simultaneously, and can choose the most representative data points of the distribution structure of dataset without any label information. Empirical results on both synthetic and real-world data sets show that the proposed algorithm has high performance in classification accuracy and robustness under the same labeling cost.

Keywords Active learning, Sparse reconstruction, Optimal experimental design, Transductive experimental design, Local linear reconstruction

1 引言

在许多模式识别如人脸识别、文本分类、手写字体识别和图像分类等实际应用中,无标签样本众多且容易获得,有标签样本稀少甚至没有。面对这种情况,为保证分类器的分类精度和泛化能力,传统的监督学习(即被动学习)需要人工标注大量样本。但对大量样本进行精确标注不仅繁琐乏味、成本昂贵,而且非常困难。因此,主动学习[1]应运而生。主动学习算法通过设计合适的采样策略,选择最有助于提高分类器精度和泛化能力的少量样本进行标注,从而在保证分类器分类

到稿日期:2017-01-11 返修日期:2017-03-18 本文受国家自然科学基金(61672008),广东省自然科学基金重点项目(2016A030311013),广东省普通高校国际合作重大项目(2015KGJHZ021),广东省自然科学基金(2016A030310335)资助。
吕巨建(1984-),男,博士,讲师,主要研究方向为机器学习、信号与信息处理、计算机视觉等,E-mail:jujianlv@163.com(通信作者);**赵慧民**(1966-),男,博士,教授,主要研究方向为信息安全、信号与信息处理等;**陈荣军**(1978-),男,博士,副教授,主要研究方向为物联网技术、智能信息系统等;**李键红**(1981-),男,博士,助理研究员,主要研究方向为语音处理、图像处理、机器学习等。

精度和泛化能力的前提下,有效减少了人工标注量。

现有的主动学习算法中的采样策略大致可以分为两类[2]:基于不确定性的采样算法[2-7]和基于代表性的采样算法[8-11]。基于不确定性的采样算法,也被称为基于分类器的采样算法,是监督主动学习方法。这类算法选择分类器最不能确定的样本进行标注,结构简单、适用性较广,但在实际应用中存在3个问题:1)在初始阶段需要一定数量的有标签样本来训练初始分类器,而在无任何已标注样本的情况下,初始的已标记样本通常是从大量未标注样本中随机选取后交由人工标注者标注得到的,随机选择样本的好坏对算法结果的影响较大;2)由于通常是迭代运行的,因此前面样本的选择会持续影响后面样本的选择和分类器的性能;3)只关注样本的不确定性,忽略了样本的分布信息,容易受到孤立点的影响。因此,基于不确定的主动学习算法并不适用于初期无任何标签信息的样本选择。

基于代表性的主动学习算法则是适用于无任何标签信息的无监督主动学习算法。这类算法根据无标注样本所包含的潜在结构分布信息设计合适的采样策略,来选择部分最能代表样本集结构分布的高价值样本进行人工标注,从而在保证分类器精度和稳定性的前提下,尽可能减少人工标注的工作量和训练分类器所需要的标注样本数量。其中,基于最优实验设计[9]的系列算法成为了近些年的研究热点。最优实验设计源于统计学,算法把每个样本看作一次实验,把样本对应的标签看作一个测量值,而选择最具价值样本的过程则被视为设计一个最优实验。最优实验设计通过最小化测量模型中的某一参数方差来选择最有价值的样本。这类算法仅将优化目标放在相关的参数方差上,并没有直接对样本集数据的预测误差进行优化,因此其效果往往不太令人满意,而且基于所选择样本构建的分类器的泛化能力往往较差。Yu 等人[10]在最优实验设计算法的基础上,将给定数据集上的预测误差作为优化目标,提出了直推式实验设计算法,使得算法性能有了很大的提升。直推式实验设计算法虽然考虑了样本集的全局结构,但是忽视了数据内在的几何结构。Zhang 等人[11]在直推式实验设计算法的基础上,通过引入局部重构理论,提出了一种基于局部线性重构的主动学习算法。此算法假设高维欧氏空间局部邻域的样本点位于一个近似为线性的低维子流形上[11-12]。基于此假设,算法要求任何样本都只可以由其邻域内的样本线性重构。该算法采用最近 k 邻域的方法来搜寻邻域点,并用这些邻域点来构建重构系数,然后通过最小全局重构误差来定位最具代表性的样本点。由于需要用目标样本的最近 k(固定值)邻域点来重构目标样本,因此该算法存在3个问题:1)局部重构要求的所有近邻点的重构系数不稀疏;2)由于利用欧氏距离来寻找最近邻,因此可能会把不是同一线性空间的点选为近邻点,这样有违线性表示的初衷;3)无法根据数据的分布自适应地选择邻域规模,而邻域规模的设定对算法效果的影响很大。确切地说,过小的邻域规模可能导致无法捕捉到足够的流形几何结构信息,尤其是当邻域规模小于流形的本征维数时;而过大的邻域规模则违背了局部线性原则。最近,Xia 等人[13]基于稀疏表示模型和最优实验设计方法,提出了一种基于稀疏线性重构的主动学习算法。该算法通过传统的稀疏表示模型来获得样本与其他样本之间的稀疏重构关系,这种方法虽然避免了处理不同目标数据时邻域规模的设定对算法结果的影响,但无法克服传统稀疏表示模型收敛慢、同权重和忽视数据局部结构信息的缺点;而且由于算法直接在数据原空间中计算重构系数,因此当面对高维数据集时,算法的效率和可靠性都较低。

在真实数据中,流形的曲率和密度在不同的区域往往都是不同的[14],因此,自适应地选择每个目标样本的邻域点和邻域规模是非常有必要的。针对这个问题,本文借鉴基于局部线性重构主动学习算法的思想,结合自适应稀疏邻域重构理论,提出基于自适应稀疏邻域重构的主动学习算法。该方法可以根据数据集各区域的不同分布自适应地选择邻域规模,同步完成邻域点的搜寻和重构系数的计算,能在无任何标签信息的情况下较好地选择最能代表样本集分布结构的样本。

2 相关工作

2.1 最优实验设计

在统计学里,主动学习也被称为最优实验设计[9]。最优实验设计的主要目的是从给定的实验数据集 $X=\{x_1,\cdots,x_n\}\in\mathbb{R}^d$ 中找到一个包含最多信息量的子集 $Z=\{z_1,\cdots,z_m\}\subset X$。也就是说,只要这个子集中的实验数据被标注后用作训练样本,就可以设计出一个最佳的实验模型,该模型能最为准确地预测出未知数据的标签值。

经典的实验设计通过测量数据 $y=w^{\mathrm{T}}x+\varepsilon$(其中 $\varepsilon\sim N(0,\delta^2)$为测量误差)来学习得到一个线性函数 $y=w^{\mathrm{T}}x$。为了达到表达上的一致性,通常把实验数据 x 称为样本,相应的测量值 y 称为样本的标签值。假设存在一组带标签的样本 $(z_1,y_1),\cdots,(z_m,y_m)$,则 w 的最大似然估计可以通过求解优化问题

$$\hat{w}=\arg\min_{w}\{J(w)=\sum_{i=1}^{m}(w^{\mathrm{T}}z_i-y_i)^2\} \tag{1}$$

来获得最优解,即

$$\hat{w}=(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}y \tag{2}$$

其中,$Z=[z_1,\cdots,z_m]^{\mathrm{T}}$,$y=[y_1,\cdots,y_m]^{\mathrm{T}}$。

由高斯马尔科夫定理可知,误差 $e=\hat{w}-w$ 的均值为0,协方差矩阵为 σ^2C_w,其中 C_w 为 $J(w)$ 的赫森逆矩阵[9]:

$$C_w=\left(\frac{\partial J(w)}{\partial(w)\partial(w)}\right)^{-1}=(Z^{\mathrm{T}}Z)^{-1}$$

因此,最优化问题变成了最小化 C_w 的某些参数的问题。就一般情况而言,最优化实验设计的优化准则可以分为两种类型[9]。第一种类型是通过选择相应的子集来最小化估计参数 $\hat{w}$ 的置信区域。此类型中最常用的3个准则分别为:

1)A-最优设计:最小化 C_w 的迹,从而最小化置信区域周围封闭空间的维度;

2)D-最优设计:最小化 C_w 的行列式,从而最小化置信区域的体积;

3)E-最优设计:最小化 C_w 的最大特征值,从而最小化置信区域主轴的尺寸。

第二种类型是通过选择相应的子集来最小化某些感兴趣

空间区域内预测值的方差。对于一个给定的测试样本 v，其预测值$\hat{w}^{\mathrm{T}}v$的方差为 $v^{\mathrm{T}}C_w v$。针对此种类型，最常用的两个准则分别为：

1)I-最优设计，即最小化感兴趣空间区域内全部样本预测方差的平均值；

2)G-最优设计，即最小化感兴趣空间区域内全部样本预测方差的最大值。

2.2　直推式实验设计

直推式实验设计是基于 I-最优设计思想的一种主动学习算法[10]。该算法在 I-最优设计的基础上添加了正则化项，改进后的目标函数为：

$$\hat{w}=\arg\min_{w}\{J(w)=\sum_{i=1}^{m}(w^{\mathrm{T}}z_i-y_i)^2+\lambda\|w\|_2^2\} \tag{3}$$

其中，$\lambda>0$ 为规则化参数，$\|\cdot\|$ 为向量的二范数。其最优解为：

$$\hat{w}=(Z^{\mathrm{T}}Z+\lambda I)^{-1}Z^{\mathrm{T}}y \tag{4}$$

其中，I 是一个单位矩阵。最优解 $\hat{w}$ 的协方差矩阵为：

$$\begin{aligned}C_{\hat{w}}&=(Z^{\mathrm{T}}Z+\lambda I)^{-1}Z^{\mathrm{T}}C_yZ(Z^{\mathrm{T}}Z+\lambda I)^{-1}\\&=\sigma^2(Z^{\mathrm{T}}Z+\lambda I)^{-1}Z^{\mathrm{T}}Z(Z^{\mathrm{T}}Z+\lambda I)^{-1}\\&=\sigma^2(Z^{\mathrm{T}}Z+\lambda I)^{-1}(Z^{\mathrm{T}}Z+\lambda I-\lambda I)(Z^{\mathrm{T}}Z+\lambda I)^{-1}\\&=\sigma^2(Z^{\mathrm{T}}Z+\lambda I)^{-1}-\lambda(Z^{\mathrm{T}}Z+\lambda I)^{-1}(Z^{\mathrm{T}}Z+\lambda I)^{-1}\end{aligned}$$

由于通常情况下都设置规则化参数 λ 为很小的值，因此 $C_{\hat{w}}\approx\sigma^2(Z^{\mathrm{T}}Z+\lambda I)^{-1}$。

类似于 I-优化设计，直推式实验设计的目的就是选择一个可以最小化测试集中全部样本预测方差平均值的子集。为了简化表述，定义测试集为 $X=[x_1,\cdots,x_n]^{\mathrm{T}}$，则预测方差的平均值可以表示为：

$$\begin{aligned}\frac{\sigma^2}{n}\sum_{i=1}^{n}x_i^{\mathrm{T}}C_{\hat{w}}x_i&\approx\frac{\sigma^2}{n}\mathrm{Tr}(XC_{\hat{w}}X^{\mathrm{T}})\\&=\frac{\sigma^2}{n}\mathrm{Tr}\left[X(Z^{\mathrm{T}}Z+\lambda I)^{-1}X^{\mathrm{T}}\right]\\&=\frac{\sigma^2}{\lambda n}\mathrm{Tr}\left[XX^{\mathrm{T}}-XZ^{\mathrm{T}}(ZZ^{\mathrm{T}}+\lambda I)^{-1}ZX^{\mathrm{T}}\right]\end{aligned}$$

其中，σ^2，λ，n 和 $\mathrm{Tr}(XX^{\mathrm{T}})$ 都为常量。因此，直推式实验设计可以公式化为如下的优化问题：

$$\begin{aligned}&\max_{Z}\ \mathrm{Tr}[XZ^{\mathrm{T}}(ZZ^{\mathrm{T}}+\lambda I)^{-1}ZX^{\mathrm{T}}]\\&\text{s.t. }Z\subset X,|Z|=m\end{aligned} \tag{5}$$

经过一些数学变换，优化问题(5)等价于如下问题[10]：

$$\begin{aligned}&\min_{Z,A}\sum_{i=1}^{n}\|x_i-Z^{\mathrm{T}}\alpha_i\|^2+\lambda\|\alpha_i\|^2\\&\text{s.t. }Z\subset X,|Z|=m,A=[\alpha_1,\cdots,\alpha_n]^{\mathrm{T}}\in\mathbb{R}^{n\times m}\end{aligned} \tag{6}$$

2.3　基于局部线性重构的主动学习算法

直推式实验设计只考虑数据集的全局结构，忽视了数据集的内在结构细节。Zhang 等人[11]通过在直推式实验设计的基础上引入局部线性重构理论，提出了一种基于局部线性重构的主动学习算法。局部线性重构，也叫局部线性表达，最早由 Roweis 等人[12]提出并应用于无监督非线性降维。局部线性表达假设在局部邻域内的数据点是线性的，因此邻域内任意一个数据点都可以由其近邻点的线性加权组合得到。这种方法的主要优点是可以用局部线性结构来反映全局的非线性结构。其重构系数矩阵的求解分为两步：1)寻找每个样本点的 k(通常为提前设定的固定值)个近邻点；2)由每个样本点的近邻点计算出该样本点的局部重建权值矩阵。

对于给定的一个样本集 $X=\{x_1,\cdots,x_n\}\in\mathbb{R}^d$，局部线性重构方法通过求解以下最小化问题来获得重构系数矩阵：

$$\begin{aligned}&\min_{W}\sum_{i=1}^{n}\|x_i-\sum_{j=1}^{n}W_{ij}x_j\|_2^2\\&\text{s.t. }\sum_{j=1}^{n}W_{ij}=1,i=1,\cdots,n\\&\qquad W_{ij}=0\text{ if }x_j\notin N_k(x_i)\end{aligned} \tag{7}$$

其中，矩阵 $W\in\mathbb{R}^{n\times n}$ 为重构系数矩阵，矩阵中的元素 W_{ij} 为第 j 个样本对于第 i 个样本的重构系数，$N_k(x_i)$ 为数据点 x_i 的 k 个近邻点所构成的邻域，k 为一个预先给定的值。

为了度量所选择样本的代表性，该算法利用重构系数矩阵 W，构建了如下目标函数[11]：

$$\varepsilon(q_1,\cdots,q_n)=\sum_{i=1}^{m}\|q_{s_i}-x_{s_i}\|_2^2+\mu\sum_{i=1}^{n}\|q_i-\sum_{j=1}^{n}W_{ij}q_j\|_2^2 \tag{8}$$

其中，$Q=[q_1,q_2,\cdots,q_n]^{\mathrm{T}}$ 为重构后的样本，$S=[x_{s_1},x_{s_2},\cdots,x_{s_m}]^{\mathrm{T}}\subset X$ 为选择的样本子集，μ 为一个调节参数。在目标函数中，第一部分是为了确定所选择样本的坐标，第二部分是为了保证重构后的样本集与原样本集具有相同的局部几何结构。

定义一个 n 阶对角矩阵 Λ：

$$\Lambda_{ii}=\begin{cases}1, & \text{if } i\in\{s_1,s_2,\cdots,s_m\}\\0, & \text{otherwise}\end{cases}$$

则式(8)可以转化为如下的矩阵形式：

$$\varepsilon(Q)=\mathrm{Tr}((Q-X)^{\mathrm{T}}\Lambda(Q-X))+\mu\mathrm{Tr}(Q^{\mathrm{T}}MQ) \tag{9}$$

其中，$M=(I-W)^{\mathrm{T}}(I-W)$。式(9)对 Q 求导并置 0 可得到 $Q=(\mu M+\Lambda)^{-1}\Lambda X$。则全局重构误差可以表示为：

$$\begin{aligned}\varepsilon(x_{s_1},\cdots,x_{s_2})&=\|X-Q\|_F^2\\&=\|X-(\mu M+\Lambda)^{-1}\Lambda X\|_F^2\\&=\|X-(\mu M+\Lambda)^{-1}(\Lambda+\mu M-\mu M)X\|_F^2\\&=\|(\mu M+\Lambda)^{-1}\mu MX\|_F^2\end{aligned} \tag{10}$$

其中，$\|\cdot\|_F$ 为矩阵的 Frobenius 范数。由于可以通过最小化全局重构误差来确定最具代表性的样本点的坐标，因此可以把基于局部线性重构的主动学习算法的目标函数简化为[11]：

$$\begin{aligned}&\min\ \|(\mu M+\Lambda)^{-1}\mu MX\|_F^2\\&\text{s.t. }\Lambda\text{ 为对角矩阵；}\Lambda_{ii}\in\{0,1\},i=1,\cdots,n;\sum_{i=1}^{n}\Lambda_{ii}=m\end{aligned} \tag{11}$$

在求得上述函数的最优解$\hat{\Lambda}$后，就可以通过其对角线上数值为 1 的元素来定位最有代表性的样本的坐标。

3　基于自适应稀疏邻域重构的主动学习算法

3.1　自适应稀疏邻域重构

3.1.1　稀疏表示

稀疏表示是近年来信号处理领域的一个研究热点[14-17]。

稀疏表示的目的是在给定的超完备字典中用尽可能少的原子来线性近似输入信号。这样的表示可以获得更为简洁的信号表达方式，从而更容易地获取信号中所蕴含的信息，进而更方便对信号进行加工处理。对于给定的一组字典 A 和一个输入信号 y，求解其稀疏表示系数 x 的目标函数为：

$$\min \ \lambda \| x \|_0 \quad \text{s.t. } Ax=y \tag{12}$$

其中，ℓ_0-范数 $\| x \|_0$ 为 x 的稀疏度，表示 x 中非 0 系数的个数。然而，上述目标函数中的 ℓ_0-范数很难优化求解，是一个 NP 难问题。但在大多数大型线性方程组确定的情况下，ℓ_1-范数最小化问题的近似解接近它的稀疏近似解[15]。因此，稀疏表示问题也可以通过优化下述问题来进行求解[17]：

$$\min \lambda \| x \|_1 \quad \text{s.t. } Ax=y \tag{13}$$

其中，ℓ_1-范数 $\| x \|_1$ 表示 x 中所有元素的绝对值之和。考虑到噪音的影响和训练数据的不足，以及一些异常值存在的情况，上述最优化问题还可以表示为另一种形式[17]：

$$\min \ \| x \|_1 \quad \text{s.t. } \| Ax-y \|_2<\varepsilon \tag{14}$$

其中，ε 为一个很小的正数。

3.1.2 自适应稀疏邻域重构

通过式(13)或式(14)求解得到的稀疏系数虽然能保证系数的稀疏度，但忽视了样本的局部结构信息，丧失了数据之间的相互关系；而且面对高维度的数据集，该算法的效率非常低。鉴于此，自适应稀疏邻域重构理论引入了局部距离权重，在保证重构系数稀疏的同时，保持了原始训练样本的局部结构信息。

对于给定的样本集中的每个样本 x_i，定义：

$$X_i=[\frac{x_1-x_i}{\| x_1-x_i \|_2},\cdots,\frac{x_{i-1}-x_i}{\| x_{i-1}-x_i \|_2},\frac{x_{i+1}-x_i}{\| x_{i+1}-x_i \|_2},\cdots,\frac{x_n-x_i}{\| x_n-x_i \|_2}]$$

其中，$X_i\in\mathbb{R}^{d\times n-1}$。自适应稀疏邻域重构理论考虑以下带权重的稀疏优化问题[14]：

$$\min \ \lambda \| C_i s_i \|_1+\frac{1}{2}\| X_i s_i \|_2^2 \quad \text{s.t. } 1^{\mathrm{T}}s_i=1 \tag{15}$$

或者

$$\min \ \| C_i s_i \|_1 \quad \text{s.t. } \| X_i s_i \|_2<\varepsilon, 1^{\mathrm{T}}s_i=1 \tag{16}$$

其中，ℓ_1-范数用于保证函数解的稀疏性；C_i 是一个正定的对角矩阵，用于近邻诱导。矩阵 C_i 的对角元素值的设定遵循这样的规则，即距离样本点 x_i 越近的点的值越小，距离越远的点的值越大。这样的设定有利于将非零系数分配给距离目标样本较近的样本点，而距离较远的样本点的系数则趋向于 0。矩阵 C_i 对应的元素值可以根据实验数据的不同特性进行自由选择，其中两种较为简单的方式为：

$$\frac{\| x_j-x_i \|_2}{\sum\limits_{k\neq i}\| x_k-x_i \|_2}\in(0,1]$$

和

$$\frac{\exp(\frac{\| x_j-x_i \|_2}{\sigma})}{\sum\limits_{k\neq i}\exp(\frac{\| x_k-x_i \|_2}{\sigma})}\in(0,1]$$

求得最优解 $\hat{s}_i$ 后，只需要进行简单的变换就可以得到重构系数 $w_i=[w_{i1},\cdots,w_{in}]^{\mathrm{T}}\in\mathbb{R}^n$，其对应元素为：

$$w_{ii}=0, w_{ij}=\frac{s_{ij}/\| x_j-x_i \|_2}{\sum\limits_{k\neq i}s_{ik}/\| x_k-x_i \|_2}, j\neq i \tag{17}$$

这是因为 $\hat{s}_i$ 满足条件：

$$\sum_{j\neq i}\frac{s_{ij}(x_j-x_i)}{\| x_j-x_i \|_2}\approx 0$$

通过上述变换容易证明，w_i 同样满足 $1^{\mathrm{T}}w_i=1$。在求得所有样本点的重构系数后，就可以得到整体的重构矩阵 $W=[w_1,w_2,\cdots,w_n]^{\mathrm{T}}$。

通过图 1 中的实例可以更加直观地展示自适应稀疏邻域重构模型在搜寻邻域点和邻域规模上的优越性[14]。图 1 中 x_1 为目标样本，样本点 x_4,x_5,x_6 比样本点 x_2 和 x_3 更加靠近目标样本 x_1。对于目标样本 x_1，自适应稀疏邻域重构模型会自适应地选择两个最佳近邻点 x_2 和 x_3 来重构目标样本。而对于局部线性重构模型，以目标样本 x_1 为中心的任何包含样本点 x_2 和 x_3 的邻域必然会包含非同一流形上的点 x_4,x_5 和 x_6。对于传统的稀疏线性重构模型，线性空间上的非零系数样本点既包含样本点的近邻点 x_2,x_3，也会包含非近邻点 x_p。

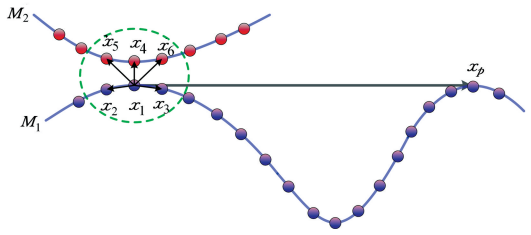

图 1 双流形邻域选择实例

Fig. 1 Double-manifold neighborhood selection instance

3.2 目标函数

本节将自适应稀疏邻域重构理论引入到基于局部线性重构的主动学习算法框架中，提出基于自适应稀疏邻域重构的主动学习算法。为了表述上可区分，定义 W_A 为自适应稀疏重构系数矩阵，$M_A=(I-W_A)^{\mathrm{T}}(I-W_A)$。用 M_A 替代目标函数(11)中的 M，就可以得到基于自适应稀疏邻域重构的主动学习算法的目标函数：

$$\min \ \| (\mu M_A+\Lambda)^{-1}\mu M_A X \|_F^2$$

$$\text{s.t. } \Lambda \text{ 为对角矩阵；} \Lambda_{ii}\in\{0,1\}, i=1,\cdots,n; \sum_{i=1}^{n}\Lambda_{ii}=m \tag{18}$$

在求得上述目标函数的最优解后，可以通过其对角线上数值为 1 的元素来定位最具代表性的样本点的坐标。由于目标函数(18)中包含了离散对角矩阵 Λ 和矩阵的逆函数，因此通过直接对其求导获得梯度来进行优化是非常困难的。为求解上述最小化问题，下面将介绍两种优化方法，即序列优化方法和凸松弛优化方法。

3.3 优化算法

3.3.1 序列优化方法

假设已选取 t 个最具代表性的样本集 $S_t=[x_{s_1},x_{s_2},\cdots,$

$x_{s_t}]^{\mathrm{T}}\subset X$。设 Λ_t 是一个 n 阶对角矩阵，其对角线上的元素为：

$$(\Lambda_t)_{ii}=\begin{cases}1, & \text{if } x_i\in S_t\\ 0, & \text{if } x_i\notin S_t\end{cases}$$

那么，选取第 $t+1$ 个样本 $x_{s_{t+1}}$ 的问题就可以转化为求解如下问题：

$$s_{t+1}=\arg\min_{i\notin\{s_1,\cdots,s_t\}}\|(\mu M_A+\Lambda_t+e_ie_i^{\mathrm{T}})^{-1}\mu M_AX\|_F^2$$

其中，e_i 是第 i 个元素为 1、其余元素为 0 的单位向量。为避开矩阵逆函数，根据 Shernab-Morrison-Woodbury 准则，对上述目标函数中的矩阵逆进行相应的变换，可得：

$$\begin{aligned}&\|(\mu M_A+\Lambda_t+e_ie_i^{\mathrm{T}})^{-1}\mu M_AX\|_F^2\\&=\mu^2\mathrm{Tr}(HM_AX^{\mathrm{T}}XM_AH)-\frac{2\mu^2H_{i*}M_AXX^{\mathrm{T}}M_AHH_{*i}}{1+H_{ii}}+\\&\quad\frac{2\mu^2H_{i*}H_{*i}H_{i*}M_AXX^{\mathrm{T}}M_AH_{*i}}{(1+H_{ii})^2}\end{aligned}$$

其中，$H=(\mu M_A+\Lambda_t)^{-1}$，$H_{i*}$ 和 H_{*i} 分别代表 H 的第 i 行和第 i 列。由于 μ，M_A，X 均已知，因此优化问题最终可以改写为：

$$\begin{aligned}s_{m+1}&=\arg\min_{i\notin\{s_1,\cdots,s_t\}}\frac{1}{(1+H_{ii})}\left(\frac{H_{i*}H_{*i}H_{i*}M_AXX^{\mathrm{T}}M_AH_{*i}}{1+H_{ii}}-2H_{i*}M_AXX^{\mathrm{T}}M_AHH_{*i}\right)\\&=\arg\min_{i\notin\{s_1,\cdots,s_t\}}\frac{2}{(1+H_{ii})}H_{i*}\left(M_AXX^{\mathrm{T}}M_A\left(\frac{\|H_{i*}\|_2^2}{2(1+H_{ii})}I-H\right)\right)H_{*i}\end{aligned}$$

在选取完第 $t+1$ 个样本 $x_{s_{t+1}}$ 后，对 H 进行更新：

$$\begin{aligned}H\leftarrow(\mu M_A+\Lambda_{t+1})^{-1}&=(\mu M_A+\Lambda_t+e_ie_i^{\mathrm{T}})^{-1}\\&=H-\frac{H_{*i}H_{i*}}{1+H_{ii}}\end{aligned}$$

不断重复以上过程，直到选取 m 个样本。

算法 1　采用序列优化方法的基于自适应稀疏邻域重构的主动学习算法

输入：样本集 $X=[x_1,x_2,\cdots,x_n]^{\mathrm{T}}$，需要选择的样本个数 m，规则化参数 μ 和 λ

输出：m 个最具代表性的样本的坐标 $Z=\{s_1,s_2,\cdots,s_m\}$

1. 输入样本集数据矩阵 $X\leftarrow[x_1,x_2,\cdots,x_n]^{\mathrm{T}}$；
2. 通过式(15)或式(16)求解每个样本的稀疏表示系数，然后通过式(17)做简单变换得到重构矩阵 W_A；
3. 计算矩阵 $M_A=(I-W_A)^{\mathrm{T}}(I-W_A)$，初始坐标为空，即 $Z\leftarrow\varnothing$；
4. 计算矩阵 $H\leftarrow(\mu M_A)^{-1}$；
5. for t=1 to m
6.　for i=1 to n
7.　if $i\notin I$ then
8.　计算 $h(i)=\frac{2}{(1+H_{ii})}H_{i*}(M_AXX^{\mathrm{T}}M_A(\frac{\|H_{i*}\|_2^2}{2(1+H_{ii})}I-H))H_{*i}$；
9.　end if
10. end for
11. 通过排序对比找到 $s_t=\arg\min_{i\in I}h(i)$，并进行标记；
12. 更新 $Z\leftarrow Z\cup s_t$；
13. 更新 $H\leftarrow H-\frac{H_{*i}H_{i*}}{1+H_{ii}}$；
14. end for
15. 返回 Z.

3.3.2　凸松弛优化算法

对目标函数(8)进行变换可得：

$$\begin{aligned}&\|(\mu M_A+\Lambda)^{-1}\mu M_AX\|_F^2\\&=\mu^2\mathrm{Tr}(X^{\mathrm{T}}M_A)(\mu M_A+\Lambda)^{-2}M_AX)\\&=\mu^2\mathrm{Tr}(X^{\mathrm{T}}M_A)(\mu^2M_A^2+\mu M_A\Lambda+\mu\Lambda M_A+\Lambda)^{-1}M_AX)\end{aligned}$$

设 $\Lambda=\mathrm{diag}(\gamma)$，其中向量 $\gamma=[\gamma_1,\gamma_2,\cdots,\gamma_n]$，$\gamma_1$ 的不同值代表了样本 x_i 是否被选取。定义一个映射函数：

$$h(\gamma)=\mu^2M_A^2+\sum_{i=1}^{m}\gamma_i(\mu M_{*i}^Ae_i^{\mathrm{T}}+\mu e_iM_{i*}^A+e_ie_i^{\mathrm{T}})$$

并将 γ 中的元素松弛为连续值，同时用 ℓ_1-范数来保证 γ 的稀疏性，则优化问题可以松弛为如下的凸优化问题：

$$\begin{aligned}&\min\ \mathrm{Tr}(X^{\mathrm{T}}M_Ah(\gamma)^{-1}M_AX)+\alpha1^{\mathrm{T}}\gamma\\&\text{s.t. }\gamma\geq0\end{aligned}$$

由于存在约束条件 $\gamma\geq0$，因此 $1^{\mathrm{T}}\gamma$ 等价于 $\|\gamma\|_1$。此目标函数是二次连续可微的，因此其最优解可以用标准的凸优化算法来求取。在引入一个辅助变量 $D\in\mathbb{R}^{d\times d}$ 后，上述问题可以变换成一个半正定规划问题：

$$\begin{aligned}&\min\ \mathrm{Tr}(D)+\alpha1^{\mathrm{T}}\gamma\\&\text{s.t. }D\geq_{\mathbb{N}_d^+}X^{\mathrm{T}}Mh(\gamma)^{-1}MX,\gamma\geq0\end{aligned}$$

算法 2　采用凸松弛优化方法的基于自适应稀疏邻域重构的主动学习算法

输入：样本集 $X=[x_1,x_2,\cdots,x_n]^{\mathrm{T}}$，需要选择的样本个数 m，规则化参数 μ，λ，α

输出：m 个最具代表性的样本的坐标 $Z=\{s_1,s_2,\cdots,s_m\}$

1. 输入样本集数据矩阵 $X\leftarrow[x_1,x_2,\cdots,x_n]^{\mathrm{T}}$；
2. 通过式(15)或式(16)求解每个样本的稀疏表示系数，然后通过式(17)做简单变换后得到重构矩阵 W_A；
3. 计算矩阵 $M_A=(I-W_A)^{\mathrm{T}}(I-W_A)$，设初始 γ 为单位矩阵，初始坐标为空，即 $Z\leftarrow\varnothing$；
4. 计算 $h(\gamma)=\mu^2M_A^2+\sum_{i=1}^{m}\gamma_i(\mu M_{*i}^Ae_i^{\mathrm{T}}+\mu e_iM_{i*}^A+e_ie_i^{\mathrm{T}})$
5. 求解下述 SDP 问题：

$$\begin{aligned}&\min\ \mathrm{Tr}(D)+\alpha1^{\mathrm{T}}\gamma\\&\text{s.t. }\begin{bmatrix}h(\gamma) & MX\\ X^{\mathrm{T}}M & D\end{bmatrix}\geq_{\mathbb{N}_{d+m}^+}0,\gamma\geq0\end{aligned}$$

6. for t=1 to m
7.　查找 $s_t=\arg\max_{i\notin I}\gamma_i$；
8.　更新 $Z\leftarrow Z\cup s_j$；
9. end for
10. 返回 Z.

其中，$\geq_{\mathbb{N}_d^+}$ 表示矩阵不等式，如 $A\geq_{\mathbb{N}_d^+}B$ 表示 $A-B$ 是一个半正定的 $d\times d$ 矩阵。利用 Schur 补定理[18]进一步变换，可得如下形式：

$$\begin{aligned}&\min\ \mathrm{Tr}(D)+\alpha1^{\mathrm{T}}\gamma\\&\text{s.t. }\begin{bmatrix}h(\gamma) & MX\\ X^{\mathrm{T}}M & D\end{bmatrix}\geq_{\mathbb{N}_{d+m}^+}0,\gamma\geq0\end{aligned}$$

利用常用的凸优化工具，如 CVX[19]，可以求解上述半正定规划问题。所得最优解 $\hat{\gamma}$ 中包含了如何选取样本的信息：对角元素值越大，其对应的样本就越具有代表性，即这些样本就越应该被选取。

4　实验结果及分析

基于自适应稀疏邻域重构的主动学习算法(ASNR)，是

一种可以在无任何标签样本的情况下选择最能代表样本集实际分布样本的主动学习算法，即其样本选择策略完全独立于标签信息。因此，除随机采样算法外，本文选择了另外 3 种样本选择策略同样独立于标签信息的主动学习算法来做对比实验：A-最优设计算法[9]（AOD），直推式实验设计算法[10]（TED），基于局部线性重构的主动学习算法[11]（LLR）。

实验分别在人工合成数据集和真实数据集上进行。人工合成数据集是由计算机生成的两个同心圆，如图 2 所示，主要用于直观地比较 LLR 算法和 ASNR 算法的样本选择结果。设置不同大小的邻域规模，通过直观的实验结果来证实 LLR 算法的效果受邻域规模大小的影响较大，而 ASNR 具备解决这一问题的能力。通过在人工合成的二维数据集上运行这两个算法，可以较为直观地反映出算法性能的好坏。

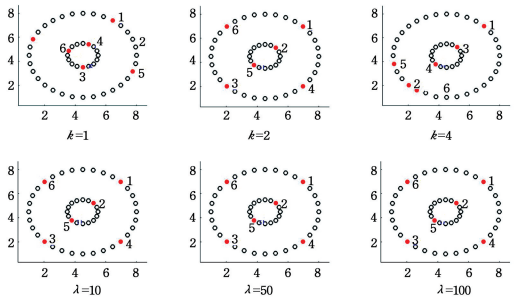

图 2 LLR 和 ASNR 算法在人工数据的不同参数设置下的结果对比

Fig. 2 Comparison results of artifical dataset for LLR and ASNR algorithms under different parameters

在真实数据集上的实验主要采用两个较为常用的图像数据集，分别为 Yale 人脸数据集和 USPS 手写字体数据集。实验的设置和具体过程为：首先，将各个主动学习算法选取的样本作为训练集来训练分类器；然后，用分类器对余下的样本进行预测，通过与真实标签信息进行对比，得出分类准确率，并以此作为评判各个算法性能好坏的依据。因为实验针对的是多类别分类问题，文中采用一对所有的（One-Versus-All，OVA）的方案来训练分类器，即如果样本集含有 c 个类，OVA 方案就训练 c 个二元分类器，每个分类器将其对应的类归为正类，而其他类则视为负类；然后用这 c 个二元分类器分别测试每个样本，并将输出值最大的视为其所在的类别。实验中采用支持向量机（SVM）作为基本的二元分类器，且所有的共有参数设置相同；对于一些非共同参数，则按照相应参考文献的建议进行设置。LLR 和 ASNR 的共有参数 $\mu=0.01$，LLR 的邻域规模按文献[11]中的建议设置 $k=10$，ASNR 的参数 $\lambda=20$。实验结果中的最优值均用黑体标出。

4.1 基于人工数据集的实验结果

人工合成数据集是由计算机生成的两个同心圆（如图 2），其中大圆有 32 个点，小圆有 16 个点，每个点相当于一个二维的样本数据。主动学习算法选择的样本点用实心点标注，旁边的数字为对应点选择的顺序标号。图 2 中第一行为 LLR 算法设置不同邻域规模 k 的结果，第二行为本文提出的 ASNR 算法设置不同参数 λ 的结果。通过对比观察可以发现，当邻域规模 k 设置为 1 时，LLR 算法的效果很差，这也印证了 LLR 算法在邻域规模小于数据本征维度时，局部线性重就不能很好地捕捉到数据的流形结构；同时，当邻域规模 k 设置为 4 时，LLR 算法的效果也很差，说明邻域规模过大时，LLR 算法对局部结构的描述就变得非常不准确。综合考虑，只有当 k 设置为 2 时，LLR 算法的结果才比较好。而本文提出的算法在 λ 设置为 10，50 或 100 时的效果都与 LLR 算法的最佳效果一样。由此可见，本文提出的基于自适应稀疏邻域重构的主动学习算法能够更好地捕捉到数据集的真实流形结构，而且具有很好的鲁棒性。

4.2 基于公开数据集的实验结果

4.2.1 Yale 数据库人脸识别

在人脸识别上的实验采用的是 Yale 人脸数据库[20]。Yale 数据库是由耶鲁大学计算视觉与控制中心创建的人脸数据库，其中包含 15 位志愿者的 165 张 32×32 像素的灰度人脸图片。图片内容包含光照角度、表情、是否佩戴眼镜和姿态的不同变化。图 3 给出了其中的部分人脸图片。

图 3 Yale 人脸数据库中的部分图片

Fig. 3 Partial images of Yale face dataset

对于 Yale 人脸图像数据集的实验，实验的设置和过程为：首先，随机从每个类中选取 10 张图片生成一个含 150 张图片的数据子集，重复这种操作 10 次，得到 10 个含 150 张图片的测试集；然后，在 10 个测试集上运行各主动学习算法，每种算法分别从中选取 5，10，…，50 个样本作为训练样本，余下未选取的样本作为测试数据。以最终的平均准确率作为评判算法性能的依据，平均准确率越高，说明算法的性能越好，反之性能越差。实验结果如表 1 所列，实验效果中的最优值均用黑体标出。

表 1 Yale 数据集上的分类结果（均值和标准偏差的百分比）

Table 1 Recognition accuracy of Yale dataset (percentage of mean and standard deviation)

样本数目	分类准确率				
	Random	AOD	TED	LLR	ASNR
5	15.3±3.6	16.8±2.1	12.9±2.0	16.1±2.4	**16.7±1.6**
10	23.8±4.1	19.1±1.9	27.5±4.2	**30.0±3.1**	29.8±2.1
15	30.9±4.8	27.6±2.5	37.9±2.4	**38.2±3.6**	38.0±2.5
20	35.9±5.7	31.7±3.0	42.6±2.8	**44.1±4.1**	43.9±3.1
25	41.9±6.1	33.9±3.6	47.4±2.9	51.9±3.5	**52.3±2.5**
30	46.4±5.7	38.3±3.7	52.5±3.2	58.1±4.1	**58.7±3.1**
35	50.6±5.6	43.6±2.8	56.6±4.6	63.9±3.7	**64.2±2.6**
40	54.5±5.3	45.1±2.9	61.5±4.7	66.5±3.3	**67.6±3.2**
45	57.4±5.7	46.7±2.9	62.8±4.3	69.3±3.6	**69.5±2.6**
50	60.0±5.5	48.2±3.1	63.7±4.5	70.4±3.0	**71.2±3.2**

从表 1 中可以看出,当选择样本的数目少于 10 时,各个算法的效果都很差,这是因为选择样本的数目 10 小于数据集本身的类别个数 15,也即某些类别的人脸图片没被选到,从而导致标注后的样本数目较少,训练分类器可参考的类别信息较少。随着样本数目的不断增多,标注后样本的数目和信息含量都有所增加,各个算法的效果也变得越来越好,原因在于分类器性能的好坏依赖于已标注样本的个数和质量。通过对比可以发现,TED 算法的效果比 AOD 算法和随机采样算法的效果好,这也说明了利用数据的结构分布信息有助于系统选择那些最有代表性的样本。而 AOD 算法的样本选择效果甚至不及被动随机采样方法的效果,原因在于 AOD 算法容易出现过拟合。此外,从表 1 中可以很清晰地看到,本文提出的 ASNR 算法和 LLR 算法明显优于其他算法,这也充分说明了利用数据的流形几何结构可以改善样本的选择效果。而 ASNR 算法的效果要优于 LLR 算法,说明自适应稀疏邻域重构模型在描述数据集流形几何结构上具有优越性,证实了基于稀疏邻域重构的主动学习算法的有效性。

为了验证邻域规模对 LLR 算法的影响,在完整的 Yale 人脸数据集上,通过设定多个邻域规模进行多次实验,最后得到统计结果。同样地,对于 ASNR 算法,也进行了类似的实验,只是设置了不同的参数。在 LLR 算法的实验中,按照 k 等于 5,10,15,20 设定邻域规模,然后得出平均值和标准偏差。在 ASNR 算法的实验中,参数 λ 按照 10,20,40,80 进行设定,然后得出平均值和标准偏差。实验比对结果如表 2 所列。从表 2 可以看出,邻域规模的变化对 LLR 算法的影响比较大,标准偏差最小值为 4.7,最大值为 7.2。而通过本文所提算法的实验效果可以看出,参数的变化对其影响较小,标准偏差最小为 1.4,最大仅为 2.9,而且参数可调范围很广,这在实际应用中是非常有意义的。

表 2　LLR 与 ASNR 在 Yale 数据集上的分类结果对比(均值和标准偏差的百分比)

Table 2　Recognition accuracy of Yale dataset for LLR and ASNR algorithms(percentage of mean and standard deviation)

样本数目	5	10	15	20	25	30	35	40	45	50
LLR	13.5±5.4	28.6±4.7	33.4±4.7	39.2±4.9	46.8±5.2	53.4±5.4	58.2±6.4	61.2±6.9	62.3±7.0	63.2±7.2
ASNR	**15.2±1.9**	**29.5±2.3**	**35.7±2.4**	**41.3±2.5**	**48.4±2.7**	**55.1±2.9**	**60.5±2.4**	**64.6±1.4**	**67.2±1.7**	**68.4±1.4**

4.2.2　USPS 数据库手写字体识别

在关于手写字体识别的实验中,采用的是 USPS 手写字体数据库[21],即美国邮政服务手写数字识别库,库中有从数字 0 到 9 的 16×16 像素的手写数字灰度图像。图 4 给出了其中的部分图片。USPS 手写字体数据库上的实验设置和过程为:首先,从整个数据集中进行 10 次随机采样,每次随机从每个类中选取 100 张图片,生成 10 个含 1000 张图片的数据子集作为实验数据集;然后,在这些实验数据集上运行各主动学习算法,每种算法都分别从中选取 10,20,…,100 个样本作为构建分类器的训练样本,并将余下未选取的未标注样本作为实验测试数据。利用各自构建的分类器对测试数据进行预测,并将预测值与真实值的平均准确率作为评判算法效果好坏的标准。平均准确率越高,说明该算法的性能越好;反之,则说明此算法的性能越差。

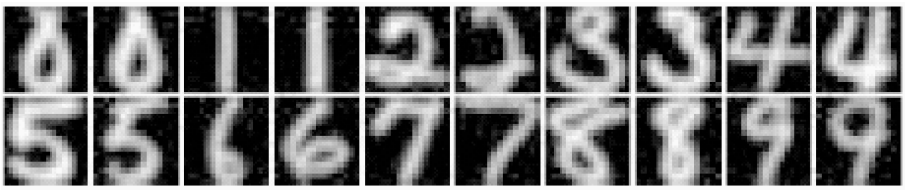

图 4　USPS 手写字体数据库的部分图片

Fig. 4　Partial sample images of USPS handwritten font dataset

从表 3 中可以看出,与 Yale 人脸数据集上的实验结果类似,当所选择的样本数很少时,各个算法的效果都较差;但随着样本数目的不断增多,各个算法的效果变得越来越好。在此次实验中,AOD 算法的效果依然不如被动的随机采样方法,这也说明 AOD 算法很容易出现过拟合。同样地,通过对比可以发现 TED 算法的效果比 AOD 算法和随机采样算法的效果要好,这也再次证明了利用数据的结构分布信息有助于系统选择那些最具代表性的高价值样本。此外,本文提出的 ASNR 算法和 LLR 算法明显优于其他算法,说明利用数据的流形几何结构有助于改善样本的选择效果。同样地,ASNR 算法的效果要优于 LLR 算法,从而证实了基于稀疏邻域重构的主动学习算法的有效性,说明了自适应稀疏邻域重构模型在描述数据集流形几何结构时的优越性。

表 3　USPS 数据集上的分类结果(均值和标准偏差的百分比)

Table 3　Recognition accuracy of USPS handwritten font dataset (percentage of mean and standard deviation)

样本数目	分类准确率				
	Random	AOD	TED	LLR	ASNR
10	31.9±4.1	32.0±5.1	37.8±3.3	39.3±3.6	**40.3±2.8**
20	43.7±4.2	42.3±4.8	51.1±4.1	**53.9±3.2**	52.8±3.2
30	50.8±3.9	45.1±5.4	57.7±4.2	**62.1±1.9**	61.9±2.9
40	57.5±3.7	48.2±4.8	62.5±3.9	**67.2±2.2**	66.9±2.4
50	61.9±3.8	50.4±5.1	65.3±3.3	70.5±2.4	**71.0±2.6**
60	65.0±3.4	53.2±5.0	68.4±3.2	73.2±1.9	**74.1±2.0**
70	67.7±3.2	54.7±4.5	70.3±2.5	74.6±1.8	**75.1±1.9**
80	69.8±3.0	55.8±4.6	71.8±2.6	76.8±1.6	**77.0±1.8**
90	71.5±2.8	58.3±4.2	74.4±2.5	77.9±1.8	**78.2±1.6**
100	72.9±2.5	59.6±4.1	75.3±2.1	78.8±1.3	**79.2±1.4**

为了验证邻域规模对 LLR 算法的影响,在完整的数据集

上，通过设定多个邻域规模进行多次实验，最后取统计结果。对提出的 ASNR 算法做同样的实验，只是参数设置不同。LLR 算法按 5，10，15，20 设定邻域规模，然后得出平均值和标准偏差。ASNR 算法的参数 λ 按 10，20，40，80 设定，然后得出平均值和标准偏差。实验对比结果如表 4 所列。从表 4 同样可以看出，邻域规模的变化对 LLR 算法的影响比较大，标准偏差最小为 4.8，最大为 6.2；而对于本文提出的 ASNR 算法，参数的影响较小，标准偏差最小为 1.6，最大仅为 3.5。

表 4 USPS 数据集上的 LLR 与 ASNR 的分类结果对比(均值和标准偏差的百分比)

Table 4 Recognition accuracy of LLR and ASNR algorithms on USPS handwritten font dataset (percentage of mean andstandard deviation)

样本数目	10	20	30	40	50	60	70	80	90	100
LLR	34.5±5.2	49.6±4.8	57.4±4.9	62.3±4.8	64.2±5.1	67.4±5.2	68.7±5.4	70.8±5.9	71.6±6.0	72.6±6.2
ASNR	**36.2±3.5**	**50.1±3.3**	**58.7±2.4**	**64.4±2.2**	**66.6±2.7**	**68.1±2.8**	**70.5±2.3**	**73.6±1.8**	**75.2±1.7**	**77.4±1.6**

结束语 本文提出了一种基于自适应稀疏邻域重构的主动学习算法，该算法能同步完成邻域点的搜寻和重构系数的计算，通过最小化全局重构误差选择最能代表样本集分布结构的样本。算法能根据数据集各区域的不同分布自适应地选择邻域规模，克服了局部线性重构受邻域规模影响较大的缺陷，提高了算法的鲁棒性和实用价值。在人工合成数据集、Yale 人脸图像数据集和 USPS 手写字体数据集上的实验结果证明，在同等标注代价的情况下，ASNR 算法在分类精度和鲁棒性上具有比同类算法更高的性能。

参 考 文 献

[1] ANGLUIN D. Queries and concept learning[J]. Machine Learning, 1988, 2(4): 319-342.

[2] SETTLES B. Active learning literature survey: Computer Sciences Technical Report 1648[R]. University of Wisconsin-Madison, 2010.

[3] LEWIS D, CATLETT J. Heterogeneous uncertainty sampling for supervised learning[C]// International Conference on Machine Learning(ICML). 1994: 148-156.

[4] FUJII A, TOKUNAGA T, INUI K, et al. Selective sampling for example based word sense disambiguation[J]. Computational Linguistics, 1998, 24(4): 573-597.

[5] TONG S, KOLLER D. Support vector machine active learning with applications to text classification[C]// International Conference on Machine Learning(ICML). 2000: 999-1006.

[6] LINDENBAUM M, MARKOVITCH S, RUSAKOV D. Selective sampling for nearest neighbor classifiers[J]. Machine Learning, 2004, 54(2): 125-152.

[7] YANG Y, MA Z, NIE F. et al. Multi-Class Active Learning by Uncertainty Sampling with Diversity Maximization[J]. International Journal of Computer Vision, 2015, 113(2): 113-127.

[8] NGUYEN H T, SMEULDERS A. Active learning using pre-clustering[C]// International Conference on Machine Learning (ICML). 2004: 79-86.

[9] ATKINSON A, DONEV A, TOBIAS R. Optimum Experimental Designs[M]. New York: SAS Oxford University Press, 2007.

[10] YU K, BI J, TRESP V. Active Learning via transductive experimental design[C]// International Conference on Machine Learning(ICML). 2006: 1081-1088.

[11] ZHANG L, CHEN C, BU J. Active learning based on locally linear reconstruction[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(10): 2026-2038.

[12] ROWEIS S T, SAUL L K. Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000, 290(5500): 2323-2326.

[13] XIA J M, YANG J A, CHEN G. Active learning based on sparse linear reconstruction[J]. Pattern Recognition and Artificial Intelligence, 2013, 26(12): 1121-1129. (in Chinese)
夏建明，杨俊安，陈功. 基于稀疏线性重构的主动学习算法[J]. 模式识别与人工智能，2013，26(12)：1121-1129.

[14] ELHAMIFAR E. Sparse manifold clustering and embedding [C]// International Conference on Neural Information Processing Systems. 2011: 55-63.

[15] DONOHO D. For most large underdetermined systems of linear equations the minimal L1-norm solution is also the sparsest solution[J]. Communications on Pure and Applied Mathematics, 2006, 59(6): 797-829.

[16] WRIGHT J, YANG A, GANESH A, et al. Robust face recognition via sparse representation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(2): 210-227.

[17] ZHANG Z, XU Y, LI X, et al. A Survey of Sparse Representation: Algorithms and Applications[J]. IEEE Access, 2017, 3: 49-530.

[18] BOYD S, VANDENBERGHE L. Convex Optimization[M]. Cambridgeshire Cambridge University Press, 2004.

[19] GRANT M, BOYD S. CVX: Matlab Software for Disciplined Convex Programming (Version1. 21) [EB/OL]. http://cvxr.com/cvx.

[20] GEORGHIADES A, BELHUMEURAND P, KRIEGMAN D. From few to many: Illumination cone models for face recognition under variable lighting and pose[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(6): 643-660.

[21] ROWEIS S. USPS Handwritten Digits [EB/OL]. http://www.cs.nyu.edu/~roweis/data.html.