局部结构自适应的线性投影方法研究

杨兴, 王士同, 胡文军

引用本文

杨兴, 王士同, 胡文军. 局部结构自适应的线性投影方法研究[J]. 计算机科学, 2024, 51(11A): 240100054-7.

YANG Xing, WANG Shitong, HU Wenjun. Study on Linear Projection Method for Local Structure Adaptation [J]. Computer Science, 2024, 51(11A): 240100054-7.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.240100054

局部结构自适应的线性投影方法研究

杨　兴[1,3]　王士同[2]　胡文军[1,3]
1 湖州师范学院信息工程学院　浙江 湖州 313000
2 江南大学人工智能与计算机学院　江苏 无锡 214122
3 浙江省现代农业资源智慧管理与应用研究重点实验室　浙江 湖州 313000
(zzhszzx@163.com)

摘　要　流形学习的核心是通过保持局部结构来捕捉数据中隐藏的几何信息,而局部结构通常利用冗余或干扰的原始数据进行评价,这意味着局部结构是不可靠的,存在局部结构置信度不足的问题。为此,针对性地提出一种局部结构自适应的线性投影方法。该方法的核心在于:一方面,它强制线性投影后的低维表示保持高维空间中的局部结构;另一方面,通过低维空间表示更新高维空间中的局部结构,并通过循环迭代方式实现局部结构的自适应。真实数据集上的实验结果表明,所提方法在各项性能指标上均优于其他对比方法。

关键词:流形学习;局部结构;置信度;自适应;线性投影
中图分类号　TP391

Study on Linear Projection Method for Local Structure Adaptation

YANG Xing[1,3],WANG Shitong[2] and HU Wenjun[1,3]
1 School of Information Engineering,Huzhou University,Huzhou,Zhejiang 313000,China
2 School of Artificial Intelligence and Computer,Jiangnan University,Wuxi,Jiangsu 214122,China
3 Zhejiang Provincial Key Laboratory of Intelligent Management and Application of Modern Agricultural Resources,Huzhou,Zhejiang 313000,China

Abstract　The core of manifold learning lies in capturing hidden geometric information within data by preserving local structures,which are typically assessed using redundant or noisy raw data. This implies that local structures are unreliable,giving rise to issues of insufficient confidence in local structures. To address this issue,a locally adaptive linear projection method is proposed. The essence of this method lies in two aspects:firstly,it enforces that the low-dimensional representation obtained through linear projection preserves local structures in the high-dimensional space;secondly,it updates the local structures in the high-dimensional space through the low-dimensional representation and achieves local structure adaptation through iterative cycles. Experimental results on real datasets demonstrate that the proposed method outperforms other comparative methods across various performance metrics.

Keywords　Manifold learning,Local structure,Confidence,Adaptive,Linear projection

1　引言

随着科学技术的发展,信息量呈爆发式增长,具体体现在数据量越来越大、数据维度越来越高。如何从大量高维度数据中获取有效的信息,成为当今研究的一个热点。同时,在机器学习、模式识别等领域,为了应对高维数据存在的信号干扰和信息冗余等问题,降维方法的研究一直备受关注,其主要包括特征选择和特征提取两类方法。本文主要针对特征提取展开相关研究。

一般地,特征提取可分为线性方法和非线性方法。线性方法主要用于线性可分数据类型,是早期应用的特征提取方法,其代表性方法有线性判别分析(Linear Discriminant Analysis,LDA)[1]和主成分分析(Principal Component Analysis,PCA)[2]等,但是现实中大部分数据是具有非线性结构的[3]。为此,许多学者提出非线性方法,非线性方法主要包括核方法和流形学习方法。核方法将原始数据由数据空间映射到高维特征空间,进而在特征空间进行线性操作,实现数据空间、特征空间之间的非线性变换,其代表性方法有核主成分分析(Kernel Principal Component Analysis,KPCA)[4]和核判别分析(Kernel Discriminant Analysis,KDA)[5]等。假设数据是采样于高维空间中的低维流形,基于流形学习的特征提取方法也被提出用于解决非线性问题。流形学习的目的是找到高维空间中的低维流形,并求出相应的嵌入映射,以实现降维。代表性算法有局部线性嵌入(Locally Linear Embedding,

基金项目:面向医学影像的小样本学习理论、关键技术与实证研究(U20A20228)
This work was supported by Learning Theory,Key Learning Techniques and CASE Studies on Small-sample-sized Medical Imaging Data (U20A20228).
通信作者:胡文军(hoowenjun@foxmail.com)

LLE)[6]等距离度量映射(Isometric Mapping, ISOMAP)[7]、局部保持投影方法(Locality Preserving Projections, LPP)[8]和邻域保持嵌入(Neighborhood Preserving Embedding, NPE)[9]。ISOMAP使用测地线距离来表示两点之间的最短距离[10],并以此保持数据的内部联系,达到降维的目的。LLE假设数据的局部结构满足线性条件,且任一样本可以利用与之近邻的样本进行近似的线性表示[11],而降维后的样本也满足这种线性表示关系。但是LLE没有明确给出对于未知的样本如何提取其低维表示[12],为此,Zhu等提[13]出了LLE的线性化版本NPE。拉普拉斯特征映射(Laplacian Eigenmaps, LE)[13]是一种基于图的降维算法,LE希望在降维后仍保持原有的局部结构信息。Zhu等对LE算法进行了线性化研究并提出了相应的改进算法LPP。

在流形学习方法中,通常利用原始数据评价局部结构,而原始数据存在冗余或干扰,这导致局部结构是不可靠的,从而对局部结构置信度提出质疑。针对上述问题,本文提出一种局部结构自适应的投影方法,称为局部结构自适应的线性投影(Local Structure Adaptive Linear Projection, LSALP)。该方法不仅能保持高维空间中的局部结构,也能通过循环迭代的方式实现局部结构的自适应。

2 相关工作

2.1 符号定义

数据集 $\boldsymbol{X}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]\in R^{d\times n}$,其中 $\boldsymbol{x}_i$ 是给定的第 i 个 d 维样本;权重矩阵定义为 $\boldsymbol{W}\in R^{n\times n}$;投影矩阵定义为 $\boldsymbol{A}\in R^{d\times m}$;低维嵌入矩阵定义为 $\boldsymbol{Y}=[\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_n]\in R^{m\times n}$;低维嵌入矩阵 $\boldsymbol{Y}$ 通过数据集 $\boldsymbol{X}$ 和投影矩阵 $\boldsymbol{A}$ 表示为 $\boldsymbol{Y}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}$。

2.2 局部保持投影

局部保持投影(LPP)是一种典型的基于流形学习的特征提取算法,该算法旨在寻找一个子空间以尽量保持原始空间中的局部结构关系[14]。对于数据集 $\boldsymbol{X}$,LPP首先通过 K 近邻构建一个近邻图 G,对近邻图 G 中的边分配权重,得到权重矩阵 $\boldsymbol{W}$。构造权重矩阵 $\boldsymbol{W}$ 有不同的方法,常用的是热核函数。

$$\boldsymbol{W}_{ij}=\begin{cases}\mathrm{e}^{-\frac{\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2}{t}}, & \text{if } \boldsymbol{x}_i\in N(\boldsymbol{x}_j) \text{ or } \boldsymbol{x}_j\in N(\boldsymbol{x}_i)\\ 0, & \text{otherwise}\end{cases} \tag{1}$$

其中,t 是热核函数的参数,$N(\boldsymbol{x}_j)$ 表示 $\boldsymbol{x}_j$ 的 K 个近邻样本集合。LPP通过优化如下目标函数得到相应投影矩阵 $\boldsymbol{A}$:

$$\begin{aligned}&\arg\min_{\boldsymbol{A}}\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{W}_{ij}\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2\\&\text{s.t. } \boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{I}\end{aligned} \tag{2}$$

其中,$\boldsymbol{D}$ 为对角矩阵且 $\boldsymbol{D}_{ii}=\sum_{j=1}^{n}\boldsymbol{W}_{ij}$。式(2)中,LPP的核心思想是在低维空间下保持原始空间中的局部结构关系,实现高维数据的有效降维,降低数据的复杂性。

2.3 重加权算法

重加权算法[15-16](Re-weighted Algorithm, RA)用于解决最小化目标函数的优化问题,该算法旨在在满足约束条件的情况下最小化如下目标函数:

$$\begin{aligned}&\arg\min_{x}h(x)+\sum f(g(\boldsymbol{x}))\\&\text{s.t. } \boldsymbol{x}\in C\end{aligned} \tag{3}$$

其中,x 是优化变量,可以是张量、向量或者矩阵,$x\in C$ 表示优化变量 x 满足约束条件集合 C,$h(x)$ 是关于变量 x 不含任何约束的任意函数,$g(x)$ 也可以是张量、向量或者矩阵,$f(g(x))$ 是关于 $g(x)$ 的凹函数。RA具体描述如算法1所示。

算法1 RA算法

输入:变量 $x\in C$

1. 计算式(3)中凹函数 $f(g(x))$ 在 $g(x)$ 处的导数 $f'(g(x))$;
2. 固定 $f'(g(x))$,计算问题 $\min_{x\in C}h(x)+\sum f'(g(x))g(x)$ 的解析解 x;
3. 返回1直到目标函数(3)的值收敛。

3 局部结构自适应的线性投影

3.1 模型提出

由2.2节可以看出,LPP是通过保持局部结构实现的,但是,LPP算法使用的是存在冗余或干扰的原始数据,这导致局部结构是不可靠的。鉴于此,本文提出LSALP,以循环迭代的方式来提高局部结构置信度。令式(3)中的变量 x 为投影矩阵 $\boldsymbol{A}$,$h(x)$ 为0,得到式(4):

$$\begin{aligned}&\arg\min_{\boldsymbol{A}}\sum f(g(\boldsymbol{A}))\\&\text{s.t. } \boldsymbol{A}\in C\end{aligned} \tag{4}$$

其中,$g(A)$ 表示低维子空间中样本间的距离度量,$f(g(\boldsymbol{A}))$ 是关于 $g(\boldsymbol{A})$ 的凹函数,$\boldsymbol{A}\in C$ 表示对于变量 $\boldsymbol{A}$ 的约束条件集合。

式(4)的解析解无法直接获得,为此我们利用算法1将式(4)转换为式(5):

$$\begin{aligned}&\arg\min_{\boldsymbol{A}}\sum f'(g(\boldsymbol{A}))g(\boldsymbol{A})\\&\text{s.t. } \boldsymbol{A}\in C\end{aligned} \tag{5}$$

欧氏距离是最常用的距离度量方式,为此本文采用欧氏距离度量低维子空间中样本间的距离,即 $g(\boldsymbol{A})=\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2$,则式(5)转换为式(6):

$$\begin{aligned}&\arg\min_{\boldsymbol{A}}\sum_{i=1}^{n}\sum_{j=1}^{n}f'(\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2)\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2\\&\text{s.t. } \boldsymbol{A}\in C\end{aligned} \tag{6}$$

其中,$f'(\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2)$ 表示局部结构关系,$\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2$ 表示低维子空间中的距离度量。

3.2 定理证明

定理1 若 $f(\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2)=(\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2)^p$ 且 $0<p<1$,则 $f(\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2)$ 为凹函数,$f'(\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2)$ 可以作为局部关系的度量。

证明:为保证 $f(\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2)$ 为凹函数,即 $f''(\|\boldsymbol{A}^{\mathrm{T}}x_i-\boldsymbol{A}^{\mathrm{T}}x_j\|^2)<0$,引入参数 p:

$$f(\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2)=(\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2)^p \tag{7}$$

同时,对式(7)中等式两边的 $\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2$ 求二阶导函数,得到式(8):

$$f''(\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2)=p(p-1)(\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2)^{p-2} \tag{8}$$

因为 $\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2>0$,且正数的任何实数次幂都是正数,所以 $(\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2)^{p-2}>0$,则需满足条件 $p(p-1)<0$。综上所述,当参数 p 的取值范围为 $0<p<1$ 时,$f(\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2)$ 为凹函数。同时,由凹函数的性质可知,$f(\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2)$ 关于 $\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2$ 的一阶导函数 $f'(\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2)$ 为单调递减函数,则 $\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2$

和 $f'(\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2)$成反比关系，而 $\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2$ 表示低维子空间中样本间的距离度量，所以可将 $f'(\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2)$定义为局部结构关系。

3.3　模型求解

为了简单，将 $f'(\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2)$记为 $\boldsymbol{W}_{ij}$，由式(6)可得：

$$\begin{aligned}&\arg\min_{\boldsymbol{A}} J=\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{W}_{ij}\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2\\&\text{s.t. }\boldsymbol{A}\in C\end{aligned}\tag{9}$$

因式(9)是一个包含投影矩阵 $\boldsymbol{A}$ 和权重矩阵 $\boldsymbol{W}$ 的多变量问题，故可采用交替迭代的方法分别求解变量 $\boldsymbol{A}$ 和 $\boldsymbol{W}$。

3.3.1　固定 $\boldsymbol{W}$，求解 $\boldsymbol{A}$

为了消除任意的缩放因素，加入了约束条件 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{I}$，其中，$\boldsymbol{D}$ 为对角矩阵且 $\boldsymbol{D}_{ii}=\sum_{j=1}^{n}\boldsymbol{W}_{ij}$，则式(9)转换为：

$$\begin{aligned}&\arg\min_{\boldsymbol{A}}\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{W}_{ij}\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2\\&\text{s.t. }\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{I}\end{aligned}\tag{10}$$

对式(10)进行展开，可得到式(11)：

$$\begin{aligned}&\sum_{i=1}^{n}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i\boldsymbol{D}_{ii}\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{A}-\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\boldsymbol{W}_{ij}\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{A}\\&=\mathrm{Tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A})-\mathrm{Tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{W}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A})\\&=\mathrm{Tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A})\end{aligned}\tag{11}$$

其中，$\boldsymbol{L}$ 为拉普拉斯矩阵，且 $\boldsymbol{L}=\boldsymbol{D}-\boldsymbol{W}$，则式(10)转化为：

$$\begin{aligned}&\arg\min_{\boldsymbol{A}}\mathrm{Tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A})\\&\text{s.t. }\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{I}\end{aligned}\tag{12}$$

对式(12)构造拉格朗日函数：

$$L(\boldsymbol{A},\lambda)=\mathrm{Tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A})-\lambda(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}-\boldsymbol{I})\tag{13}$$

其中，λ 为拉格朗日乘子。对投影矩阵 $\boldsymbol{A}$ 求偏导并置为 0，可得：

$$\frac{\partial L}{\partial A}=2\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}-2\lambda\boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}=0\tag{14}$$

即：

$$\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}=\lambda\boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\tag{15}$$

其中，将 m 个最小的非零特征值对应的特征向量组成投影矩阵 $\boldsymbol{A}$，则低维嵌入矩阵 $\boldsymbol{Y}$ 表示为 $\boldsymbol{Y}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}$。

3.3.2　固定 $\boldsymbol{A}$，求解 $\boldsymbol{W}$

由于 $f'(\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2)=\boldsymbol{W}_{ij}$，因此对式(7)中等式右边的 $\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2$ 求一阶导函数，即：

$$\boldsymbol{W}_{ij}=p(\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^2)^{p-1}\tag{16}$$

对式(16)化简，得到式(17)：

$$\boldsymbol{W}_{ij}=p\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_j\|^{2p-2}\tag{17}$$

通过式(17)可以基于低维子空间的表示更新原始高维空间中的局部结构，再通过式(15)求解投影矩阵 $\boldsymbol{A}$，然后基于新的投影矩阵 $\boldsymbol{A}$ 更新高维空间中的局部结构，以循环迭代的方式实现局部结构的自适应。

LSALP 优化算法的伪代码如算法 2 所示。

算法 2　LSALP 算法

输入：数据集 $\mathbf{X}$，近邻数 K，目标维数 m，参数 p，最大迭代次数 τ，容差参数 $\theta=10^{-6}$

输出：低维嵌入矩阵 $\mathbf{Y}=\mathbf{A}^{\mathrm{T}}\mathbf{X}$

1. 根据 K 最近邻方法，构建近邻图 G；
2. 根据近邻图 G 和式(1)计算权重矩阵 $\mathbf{W}$；
3. 初始化迭代次数 t=0；
4. 计算度矩阵 $\mathbf{D}$ 与拉普拉斯矩阵 $\mathbf{L}=\mathbf{D}-\mathbf{W}$；
5. 利用式(15)求解投影矩阵 $\mathbf{A}$；
6. 根据式(17)更新权重矩阵 $\mathbf{W}$；
7. t=t+1；
8. 计算目标函数值 J^t；
9. 若 $t=\tau$ 或 $|J^t-J^{t-1}|\leqslant\theta$ 时终止迭代，否则返回步骤 4。

3.4　复杂度

假设原始高维数据的维数为 d，目标维数是 m，最近邻数是 K，数据集包含的样本个数为 n，整个算法的迭代次数是 T。其中，LPP 算法的时间复杂度主要是由寻找 K 近邻、计算权重以及特征分解 3 个部分决定。首先在寻找 K 近邻阶段，需要的时间复杂度为 $O(dn^2)$。其次在计算权重的阶段，需要的时间复杂度为 $O(dnK)$。最后的特征分解阶段，求解一个维度是 $d\times d$ 矩阵的特征值，需要的时间复杂度为 $O(d^3)$。因此，LPP 算法总体的时间复杂度为 $O(dn^2+dnK+d^3)$[17]。本文提出的 LSALP 算法提出了自适应的思想，这就会增加求解的复杂性，在算法 2 的步骤 5 中，需要花费$O(dmn^2)$的时间复杂度去计算权重矩阵，则总体的时间复杂度为 $O(dn^2+dnK+T(d^3+dmn^2))$。表 1 列出了所有算法的时间复杂度。

表 1　所有算法的时间复杂度

Table 1　Time complexity of all algorithms

算法	时间复杂度
LSALP	$O(dn^2+dnK+T(d^3+dmn^2))$
LPP	$O(dn^2+dnK+d^3)$
PCA	$O(nd^2+d^3)$
ISOMAP	$O(dn^2+n^3)$
NPE	$O(dn^2+(d+K)K^2n+d^3)$

4　实验与分析

为验证 LSALP 算法的有效性，所有实验在 CPU 2.5 GHz、内存为 8GB 以及操作系为 Windows10 64 位的环境下完成，编译环境为 MATLAB R2020a。本实验利用 LPP，PCA，ISOMAP，NPE 和 LSALP 对数据进行降维，并添加未经过降维的 All-Feature 方法作为基准，通过分类实验，对比分析 5 种算法的低维表示能力。

4.1　实验数据集及实验设计

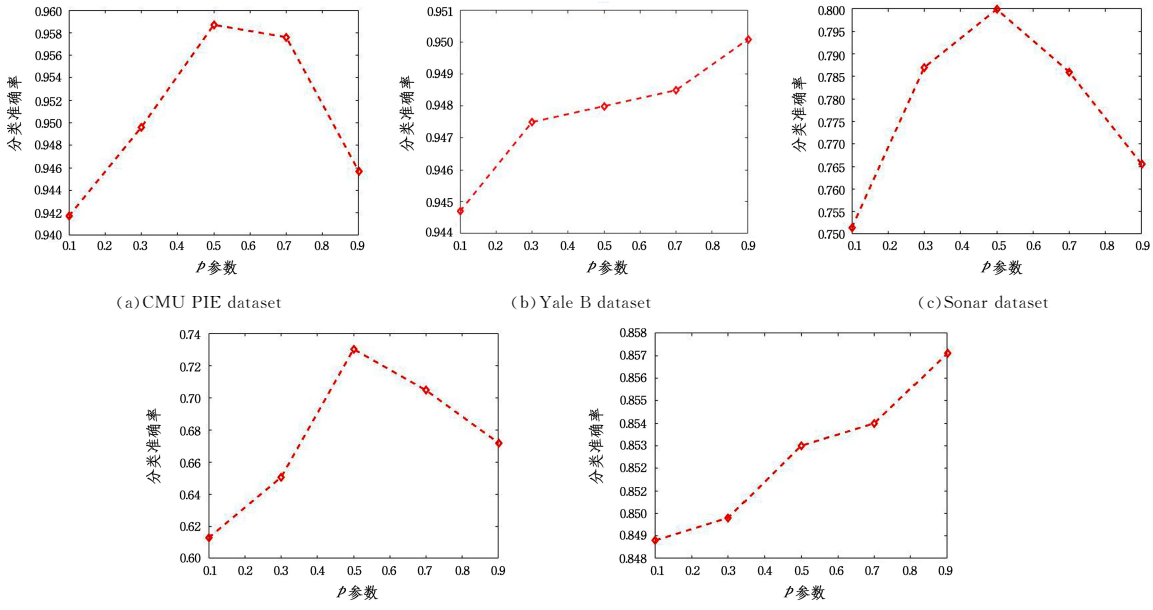
将各方法应用到多种数据集，包括 CMU PIE，Yale B 以及 3 个 UCI 数据集，分别为 Sonar，Vowel 和 Australian。CMU PIE 人脸数据集含有 41 368 张图像和 68 个类，每个人有 13 种姿态条件、43 种光照条件和 4 种表情，其中的姿态和光照变化图像也是在严格控制的条件下采集的。本实验从中选取了 1 166 张图像和 53 个类，数据集部分图像如图 1 和图 2 所示，图像像素处理为 1 024。Yale B 数据集包含 2 414 张正面人脸图像，该数据集中的人脸含有不同的光照变换，共含有 38 个类，每个类含有 59～64 张图像，人脸数据集部分图片如图 2 和图 3 所示。此外，使用 3 个 UCI 数据集，分别是 Sonar，Vowel 和 Australian。其中，Sonar 数据集包含 2 个类，分别有 97 个样本和 111 个样本，总共 208 个样本，每个样本有 60 维的特征数；Vowel 数据集为元音数据，包含 528 个数据，11 个类别，每个数据有 10 维的特征数；Australian 包含 690 个样本数，包括 2 个类别，每个数据有 14 维的特征数。处理后的数据集信息如表 2 所列。本实验中 p 参数统一设置为 0.5，关于 p 参数对算法的影响将在 4.5 节进行针对性的研究。

图 1 CMU PIE 人脸数据集

Fig. 1 CMU PIE face dataset

图 2 Yale B 人脸数据集

Fig. 2 Yale B face dataset



(a) CMU PIE dataset (b) Yale B dataset (c) Sonar dataset

(d) Vowel dataset (e) Australian dataset

图 3 不同 p 参数在 5 个数据集上分类准确率

Fig. 3 Classification accuracy of different parameters p on five datasets

表 2 实验数据集

Table 2 Experimental dataset

数据集	样本数	维数	类别数
CMU PIE	1166	1024	53
Yale B	2414	1024	38
Sonar	208	60	2
Vowel	528	10	11
Australian	690	14	2

4.2 分类算法

K 近邻分类(K-NN)[18]是一种常用的分类算法,该算法原理是给定测试样本,基于某种距离度量方法找出与其最靠近的 K 个训练样本,在这 K 个训练样本中,以少数服从多数的原则,将出现次数最多的类别标记作为预测结果。K-NN 算法的核心是样本间的距离度量方法,确定使用何种度量方法对分类的准确度起到关键作用,样本间的距离可以定义如下:

$$d(\boldsymbol{x}_i,\boldsymbol{x}_j)=\left(\sum_{l=1}^{d}|\boldsymbol{x}_i^{(l)}-\boldsymbol{x}_j^{(l)}|^q\right)^{\frac{1}{q}} \tag{18}$$

其中,$\boldsymbol{x}_i^{(l)}$ 表示第 i 个样本的第 l 个属性值,q 代表具体的距离度量方式。当 $q=1$,$q=2$ 以及 $q=\infty$时,分别代表曼哈顿距离、欧氏距离和各坐标距离的最大值。本实验采用欧氏距离,即 $q=2$,并且采用 K-NN 算法对降维后的数据进行分类实验。由于近邻值 K 的不同取值会影响特征提取的效果,为了避免该影响,涉及的近邻值 K 统一设为 10。

4.3 评价指标

为了保证分类效果的有效性,采用十折交叉验证的方法,利用准确率均值(Mean of Accuracy,MA)、方差均值(Mean of Variance,MV)和宏 F1 均值(Mean of Macro-F1,MM-F1)3 个指标来评价不同特征提取算法的分类性能。

假设测试集的正确标签向量为 $\boldsymbol{u},\boldsymbol{u}_i\in\boldsymbol{u}$,且 $i=1,2,\cdots,N$,测试集分类后的标签向量为 $\boldsymbol{v},\boldsymbol{v}_i\in\boldsymbol{v}$ 且 $i=1,2,\cdots,N$,则准确率 AR 计算式如下:

$$AR=\frac{\sum_{i=1}^{N}\delta(\boldsymbol{u}_i,\boldsymbol{v}_i)}{N} \tag{19}$$

其中,当 $\boldsymbol{u}_i=\boldsymbol{v}_i$ 时 $\delta(\boldsymbol{u}_i,\boldsymbol{v}_i)$函数为 1,否则为 0,$N$ 为测试集样本数。利用式(19)可以得到所有准确率的值,进而得到准确率均值和方差均值。准确率均值越大表明分类的精度越高,而方差均值的值越小表明分类的"稳定性"越好[19]。

宏 F1(Macro-F1)是一种用来衡量多分类问题的评价指标,该指标兼顾了查准率(Precision)和召回率(Recall)[20]。令 macro-$\bar{P}$ 和 macro-$\bar{R}$ 分别为测试集中所有类别的 Precision 均值和 Recall 均值,则 Macro-F1 定义如下:

$$Macro\text{-}F1=\frac{2\times macro\text{-}\bar{P}\times macro\text{-}\bar{R}}{macro\text{-}\bar{P}+macro\text{-}\bar{R}} \tag{20}$$

利用式(20)可以得到每一折测试数据的 Macro-F1,而宏 F1 均值(MM-F1)[21]的计算式如式(21)所示:

$$MM\text{-}F1=\frac{\sum_{i=1}^{c}(Macro\text{-}F1)_i}{c} \tag{21}$$

其中,$c=10$ 为交叉验证次数。MM-F1 值的范围被约束至 0～1 之间,其值越大说明分类效果越好[22]。

4.4 分类结果

表 3 和表 4 分别列出了 CMU PIE,Yale B,Sonar,Vowel 和 Australian 5 个数据集的分类准确率均值、方差均值和宏 F1 均值,其中 All-Feature 是对未经过特征提取的数据进行分类,加粗字体表示最佳数值。

表 3　不同算法的分类实验结果对比

Table 3　Comparison of experimental results of different classification algorithms

数据集	All-Feature	LPP	PCA	ISOMAP	NPE	LSALP
CMU PIE	84.61±0.14	90.58±0.08(300)	86.94±0.09(200)	84.05±0.07(500)	93.21±0.06(200)	**95.87±0.22(200)**
Yale B	65.10±0.07	93.24±0.03(100)	61.04±0.16(500)	59.48±0.07(500)	81.43±0.07(200)	**95.01±0.27(400)**
Sonar	67.38±1.05	72.22±0.79(10)	70.26±2.09(10)	69.68±0.81(30)	72.53±0.73(20)	**79.99±0.31(30)**
Vowel	65.21±0.67	64.26±0.53(6)	66.23±0.24(6)	65.48±0.27(4)	65.21±0.41(8)	**67.43±0.51(6)**
Australian	69.60±0.25	80.49±0.13(12)	70.25±0.56(12)	70.51±0.39(10)	84.17±0.11(12)	**85.71±0.41(12)**

注：表中数据为“准确率均值±方差均值”，括号中数据为最佳子空间维度。

表 4　不同算法的宏 F1 均值对比

Table 4　Comparison of macro-F1 mean values of different algorithms

数据集	All-Feature	LPP	PCA	ISOMAP	NPE	LSALP
CMU PIE	0.8305	0.9310(300)	0.8722(200)	0.8982(500)	0.9494(200)	**0.9671(200)**
Yale B	0.6435	0.9346(100)	0.6618(500)	0.6829(500)	0.8186(200)	**0.9479(400)**
Sonar	0.6538	0.7004(10)	0.6891(10)	0.6733(30)	0.7080(10)	**0.7932(30)**
Vowel	0.6324	0.6433(6)	0.6485(6)	0.6308(4)	0.6322(8)	**0.6675(6)**
Australian	0.6819	0.7947(12)	0.6847(12)	0.6888(10)	0.8359(12)	**0.8527(12)**

注：表中数据为“宏 F1 均值”，括号中数据为最佳子空间维度。

从表 3 和表 4 的实验结果可以发现，与其他方法相比，本文提出的 LSALP 方法在人脸分类和目标分类任务上都获得了最优的效果。此外，对比表中所列的结果还能发现：

(1)在本实验中使用到的流形学习算法中，ISOMAP 的效果相对来说较差，甚至比未经过降维的数据集的分类效果更差，这说明所使用的数据集并不适合通过保持测地线距离来进行降维，而更适合通过保留局部结构信息的方式来降维。使用其他流形学习算法的分类效果大部分都优于 PCA 的分类效果，这说明使用的数据集可能含有非线性的流形结构，而基于流形学习的特征提取方法在这些数据集上的表现更好。

(2)在本实验使用到的 5 个数据集上，NPE，LPP 和 LSALP 获得了优于 PCA 和 ISOMAP 的性能，与 PCA 和 ISOMAP 相比，NPE，LPP 和 LSALP 都利用了数据的局部结构进行降维，而 PCA 和 ISOMAP 是利用数据的全局结构进行降维，这说明通过捕捉数据的局部结构能够更好地提取有用特征，从而更有利于后续的分类实验。

4.5　p 参数分析

在式(17)中，存在一个参数 $p(0<p<1)$，为了评估参数 p 是如何影响最终的分类效果并且找到一个合适的参数 p，使用网格搜索法，并把 p 设定在[0.1，0.3，0.5，0.7，0.9]的范围内，得到的关于参数 p 对分类准确率的影响结果如图 3 所示。从图中可以观察出，当 p 被设为不同值时，不同数据集得到的分类准确率是不同的，并且，在 CMU PIE，Sonar 和 Vowel 数据集上，当 p 设为 0.5 时，得到的分类准确率是最高的；而在 Yale B 和 Australian 数据集上，当 p 设为 0.9，得到的分类准确率是最高的。在这 5 个数据集中，大部分情况下当 p 为 0.5 时得到的分类效果是较好的，因此将 p 设为 0.5 与其他对比的算法进行比较是合理的。

4.6　维数对分类准确率的影响实验分析

LSALP，LPP，PCA，ISOMAP 和 NPE 特征提取算法在不同特征提取维数下的分类准确率如图 4 所示。

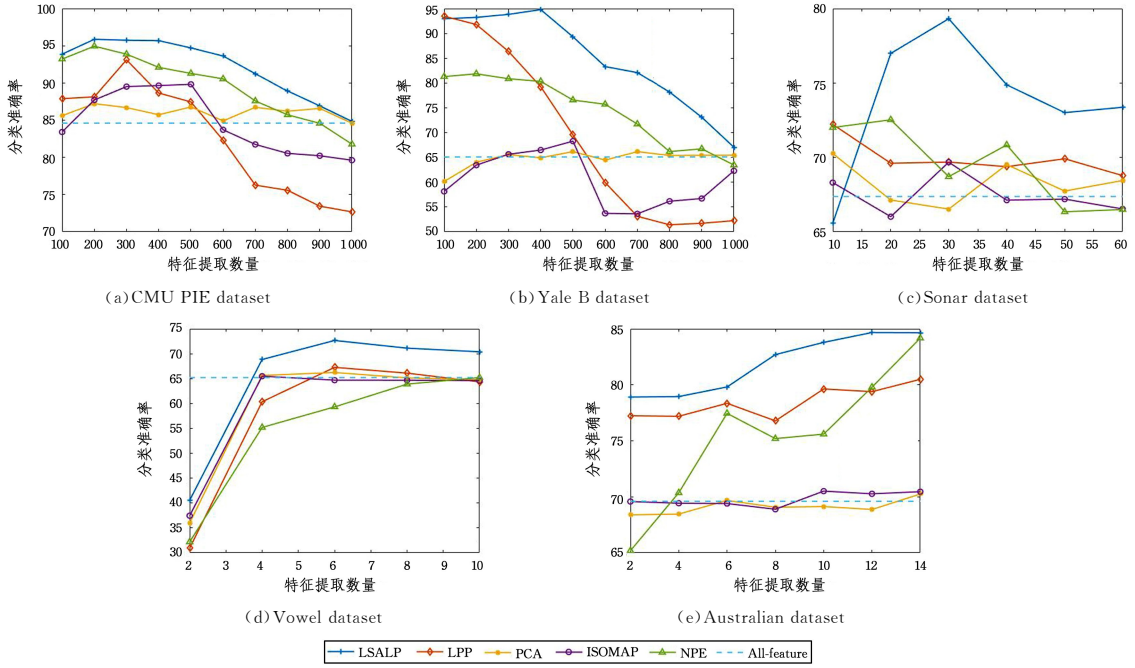


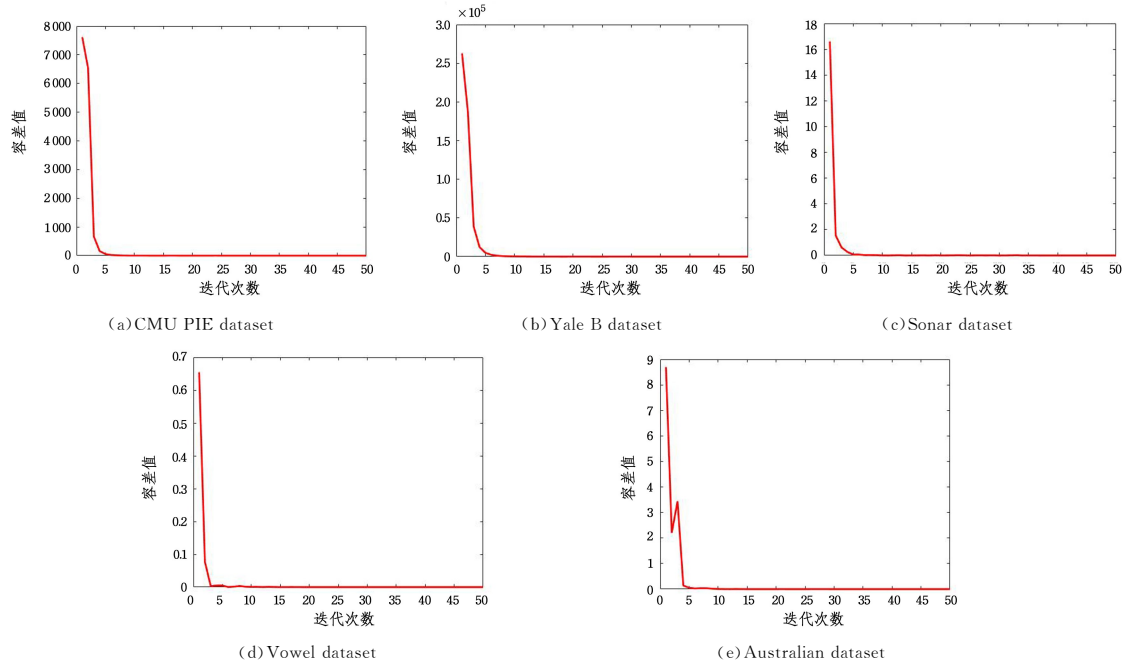
图 4　分类准确率与特征提取维数的关系

Fig. 4　Relationship between classification accuracy and feature extraction dimension

在 CMU PIE 和 Yale B 上维度选取范围为[100,1 000],步长为 100,在 Sonar 上维度选取范围为[10,60],步长为 10,在 Australian 上维度选取范围为[2,14],步长为 2,在 vowel 上维度选取范围为[2,10],步长为 2。从图 4 可以看出,在 5 个数据集上,最高分类准确率并不是在选择所有特征时出现,这意味着选择最大特征数并不一定能产生最高的分类准确率,也意味着数据集中可能存在噪声或冗余的特征,进而影响分类准确率。此外,从图 4 中还可以看出,本文提出的 LSALP 方法的分类准确率要普遍高于其他方法。

4.7 收敛性实验分析

为了证明 LSALP 的收敛性,将 LSALP 目标函数的值可视化,不同数据集上目标函数的收敛曲线如图 5 所示。可以看出,LSALP 的目标函数值在这些数据集上均在 10 次迭代内逐渐收敛,表明 LSALP 方法是可收敛的。



(a) CMU PIE dataset (b) Yale B dataset (c) Sonar dataset

(d) Vowel dataset (e) Australian dataset

图 5 LSALP 收敛性分析

Fig. 5 LSALP convergence analysis

结束语 因为原始数据存在冗余或干扰,导致局部结构是不可靠的,所以本文提出一种局部结构自适应的线性投影方法 LSALP,LSALP 通过循环迭代的方式实现局部结构自适应。LSALP 在两种人脸数据集和三种 UCI 数据集上的分类效果比其他特征提取算法表现更好,同时目标函数在迭代过程中达到收敛,验证了 LSALP 的可行性和有效性。目前提出的方法主要应用于基于线性投影的降维方法,在下一步工作中,将进行在其他流形学习的降维方法中实现局部结构自适应的探索。

参 考 文 献

[1] ZHU F, GAO J, YANG J, et al. Neighborhood Linear Discriminant Analysis[J]. Pattern Recognition, 2022, 12 3: 108422.

[2] TANG T M, ALLEN G I. Integrated Principal Components Analysis[J]. The Journal of Machine Learning Research, 2021, 22(1): 8953-9023.

[3] HASAN B M S, ABDULAZEEZ A M. A Review of Principal Component Analysis Algorithm for Dimensionality Reduction [J]. Journal of Soft Computing and Data Mining, 2021, 2(1): 20-30.

[4] LI P, ZHANG W, LU C, et al. Robust Kernel Principal Component Analysis with Optimal Mean[J]. Neural Networks, 2022, 152: 347-352.

[5] LI S, ZHANG H, MA R, et al. Linear Discriminant Analysis with Generalized Kernel Constraint for Robust Image Classification[J]. Pattern Recognition, 2023, 136: 109196.

[6] AYESHA S, HANIF M K, TALIB R. Overview and Comparative Study of Dimensionality Reduction Techniques for High Dimensional Data[J]. Information Fusion, 2020, 59: 44-58.

[7] WANG Y, ZHANG Z, LIN Y. Multi-cluster Feature Selection Based on Isometric Mapping[J]. IEEE/CAA Journal of Automatica Sinica, 2021, 9(3): 570-572.

[8] LU X, LONG J, WEN J, et al. Locality Preserving Projection with Symmetric Graph Embedding for Unsupervised Dimensionality Reduction[J]. Pattern Recognition, 2022, 131: 108844.

[9] LOU X, YAN D Q, WAN B L, et al. An Improved Neighborhood Preserving Embedding Algorithm[J]. Computer Science, 2018, 45: 255-258, 278.

[10] DING S, KEAL C A, ZHAO L, et al. Dimensionality Reduction and Classification for Hyperspectral Image Based on Robust Supervised ISOMAP[J]. Journal of Industrial and Production Engineering, 2022, 39(1): 19-29.

[11] MIAO J, YANG T, SUN L, et al. Graph Regularized Locally Linear Embedding for Unsupervised Feature Selection[J]. Pattern Recognition, 2022, 122: 108299.

[12] ANOWAR F, SADAOUI S, SELIM B. Conceptual and Empirical Comparison of Dimensionality Reduction Algorithms(PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, T-SNE)[J]. Computer Science Review, 2021, 40: 100378.

[13] ZHU H, SUN K, KONIUSZ P. Contrastive Laplacian Eigenmaps[J]. Advances in Neural Information Processing Systems,

2021,34:5682-5695.

[14] LIU N,LAI Z,LI X,et al. Locality Preserving Robust Regression for Jointly Sparse Subspace Learning[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 31(6):2274-2287.

[15] JIANG W,NIE F,HUANG H. Robust Dictionary Learning with Capped L1-norm[C]// Twenty-Fourth International Joint Conference on Artificial Intelligence. 2015.

[16] NIE F,YUAN J,HUANG H. Optimal Mean Robust Principal Component Analysis[C]// International Conference on Machine Learning. PMLR,2014:1062-1070.

[17] CHANG W,NIE F,WANG Z,et al. Self-weighted Learning Framework for Adaptive Locality Discriminant Analysis [J]. Pattern Recognition,2022,129:108778.

[18] ZHAN Y H,CHANG Z N. Research on Feature Filtering Preprocessing Based on KNN Algorithm[J]. Modern Information Technology,2022,6(4):126-128.

[19] DAI W,CHAI J,LIU Y J. Semi-supervised Learning Algorithm Based on Maximum Margin and Manifold Hypothesis[J]. Computer Science,2024,51(2):259-267.

[20] GUO H,ZOU H,TAN J. Semi-supervised Dimensionality Reduction via Sparse Locality Preserving Projection[J]. Applied Intelligence,2020,50:1222-1232.

[21] GAMBELLA C,GHADDAR B,NAOUM-SAWAYA J. Optimization Problems for Machine Learning:A Survey[J]. European Journal of Operational Research,2021,290(3):807-828.

[22] HU X F,CHEN S P. Review of Small Sample Learning Based on Machine Learning[J]. Intelligent Computers and Applications, 2021,11(7):191-195,201.

YANG Xing, born in 1998, postgraduate, is a member of CCF (No. N0566G). His main research interests is manifold learning.

HU Wenjun, born in 1977, Ph.D, professor. His main research interests include machine learning and pattern recognition.