

基于核保持嵌入的子空间学习



贺文琪^{1,2,3} 刘保龙^{1,3} 孙兆川³ 王磊^{1,2,3} 李丹萍⁴

1 西安电子科技大学青岛计算技术研究院 山东 青岛 266000

2 上海交通大学海洋智能装备与系统教育部重点实验室 上海 200240

3 西安电子科技大学电子工程学院 西安 710071

4 西安电子科技大学通信工程学院 西安 710071

(18852143390@163.com)

摘要 子空间学习是特征提取领域中的一个重要研究方向,其通过一种线性或非线性的变换将原始数据映射到低维子空间中,并在该子空间中尽可能地保留原始数据的几何结构和有用信息。子空间学习的性能提升主要取决于相似性关系的衡量方式和特征嵌入的图构建手段。文中针对子空间学习中的相似性度量与图构建两大问题进行研究,提出了一种基于核保持嵌入的子空间学习算法(Kernel-preserving Embedding based Subspace Learning, KESL),该算法通过自表示技术自适应地学习数据间的相似性信息和基于核保持的构图。首先针对传统降维方法无法挖掘高维非线性数据的内部结构问题,引入核函数并最小化样本的重构误差来约束最优的表示系数,以期挖掘出有利于分类的数据结构关系。然后,针对现有基于图的子空间学习方法大都只考虑类内样本相似性信息的问题,利用学习到的相似性矩阵分别构建类内和类间图,使得在投影子空间中同类样本的核保持关系得到加强,不同类样本间的核保持关系被进一步抑制。最后,通过核保持矩阵与图嵌入的联合优化,动态地求解出最优表示下的子空间投影。在多个数据集上的实验结果表明,所提算法在分类任务中的性能优于主流的子空间学习算法。

关键词:子空间学习;图构建;相似性学习;核保持嵌入

中图分类号 TP391

Kernel-preserving Embedding Based Subspace Learning

HE Wen-qi^{1,2,3}, LIU Bao-long^{1,3}, SUN Zhao-chuan³, WANG Lei^{1,2,3} and LI Dan-ping⁴

1 Xidian University Qingdao Institute of Computing Technology, Qingdao, Shandong 266000, China

2 Key Laboratory of Marine Intelligent Equipment and System of Ministry of Education, Shanghai Jiao Tong University, Shanghai 200240, China

3 School of Electronic Engineering, Xidian University, Xi'an 710071, China

4 School of Telecommunications Engineering, Xidian University, Xi'an 710071, China

Abstract Subspace learning is an important research subject in the field of feature extraction. It maps the original data into a low-dimensional subspace through a linear or nonlinear transformation, and preserves the geometric structure and useful information of the original data as much as possible in this subspace. The performance of subspace learning mainly depends on the design of similarity measure and the graph construction for feature embedding. Aiming at the two issues, a novel kernel-preserving embedding based subspace learning (KESL) method is proposed, which can adaptively learn the similarity information from data and construct the kernel-preserving graph. First, to tackle the problem that the traditional dimension reduction methods cannot preserve the inner structure of high-dimensional nonlinear data, our algorithm introduces the kernel function and minimizes the reconstruction error of samples, which is beneficial for mining the data structural relationship for classification. Then, aiming at the limitation that existing graph-based subspace learning methods mainly concern the similarity information of the samples within a class, our algorithm uses the learned similarity matrices to construct intra-class and inter-class graphs, respectively. Thus, in the projected subspace, the kernel-preserving relationship of the samples in the same class can be strengthened, while the kernel-preserving relationship of the samples from different classes can be largely inhibited. Finally, through the joint optimization of kernel preserving matrix and graph embedding, the desired projection under the optimal representation can be dynamically solved. Expe-

收到日期:2020-09-02 返修日期:2020-11-07 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家重点研发计划(2016YFE0207000);国家自然科学基金(61203137,61401328);陕西省自然科学基金基础研究计划资助项目(2014JQ8306,2015JM6279)

This work was supported by the National Key Research and Development Program of China(2016YFE0207000), National Natural Science Foundation of China(61203137,61401328) and Natural Science Basic Research Program of Shanxi Province of China(2014JQ8306,2015JM6279).

通信作者:李丹萍(dpli@xidian.edu.cn)

perimental results on several datasets show that the proposed algorithm is competitive to the state-of-the-art subspace learning algorithms in various classification tasks.

Keywords Subspace learning, Graph construction, Similarity learning, Kernel-preserving embedding

1 引言

随着计算机技术的发展,每分每秒都有大量的数据产生,在这样高度信息化的时代,数据存储所需的空间越来越大,而大量高维数据的产生使数据处理更为困难。此外,在模式识别、机器学习和计算机视觉等领域,数据通常位于高维空间。因此,通过数据降维^[1]技术从中提取最有用的低维特征不仅有助于缓解“维数灾难”^[2],而且还有助于降低分类或聚类任务的训练成本。子空间投影^[3-4]是数据降维的一种典型手段。在诸多子空间投影算法中,最具代表性的是主成分分析(Principle Component Analysis, PCA)^[5]和线性判别分析(Linear Discriminant Analysis, LDA)^[6]。

尽管基于线性假设的降维方法易于实现,但其忽略了高维非线性数据的内在结构。因此,许多学者提出了非线性的降维方法,目前非线性降维算法主要分为核函数方法和基于流形学习的算法两类。前者针对数据在低维空间中线性不可分的问题,通过非线性映射将低维数据映射到高维空间中,使其在高维空间中的数据线性可分,代表性的算法有核主成分分析(Kernel Principal Component Analysis, KPCA)^[7]和核费舍尔判别分析(Kernel Fisher Discriminant Analysis, KFDA)^[8]。基于流形学习的算法则是利用流形的假设在高维空间挖掘低维流形结构信息,但是由于其泛化能力差,研究者们将这类算法线性化,提出了利用图的形式来表示样本之间的关系。Yan 等于 2007 年提出图嵌入框架^[9],用本征图和惩罚图来表征数据特征,并利用这个框架提出了边界费舍尔分析(Marginal Fisher Analysis, MFA)算法,诸多流形学习算法均可纳入这个图构建框架中。图嵌入框架由于能够很好地保持样本的几何结构而被广泛应用于模式识别和计算机视觉领域^[10-11]。

除了图构建的方式,数据的相似性关系表示也是影响子空间学习性能的一个重要因素,更是谱聚类^[12]、度量学习^[13]、图像分割^[14]、行人重识别^[15]和图像检索^[16]等数据分析任务中必不可少的一部分,在数据降维中同样至关重要。为了寻找一个最优的子空间投影,并且实现在降低运算复杂度的同时还能提升识别效果,需要更好的思路来建模样本之间的相似性关系。自表示学习^[17]是将每一个样本用其他样本的线性组合来表示,两个样本之间越相似,在数据重构时其对应的权值就越大,这一框架在图像处理任务中的应用十分广泛^[18-19]。Wright 等^[20]于 2009 年提出稀疏表示分类算法,其原理是通过 ℓ_1 范数约束,用训练样本的线性组合来表示测试样本,然后使重构误差最小化,进而获得测试样本的分类结果。在此基础上,Qiao 等^[21]提出了稀疏保持投影算法,利用该算法得到的投影矩阵鲁棒性较强,但未使用类标信息。Zhang 等^[22]于 2011 年发现分类任务中起主要作用的是样本间的协作表示而非 ℓ_1 范数的最小化,于是提出协作表示分类(Collaborative Representation Classification, CRC)算法^[17]。

针对子空间学习任务, Ly 等^[23]提出基于图的协作判别分析方法(Collaborative Graph-based Discriminant Analysis, CG-DA),在 CRC 的基础上利用协作表示系数来构造权值矩阵,采用闭式解提高了算法效率,实现了降维。随后,基于自表示关系的图嵌入方法陆续涌现。Peng 等^[24]于 2017 年提出每个样本可以用除自身外的其他所有样本作为字典,再利用 ℓ_2 范数约束相似矩阵和重构误差,该方法的性能要优于 ℓ_1 范数或低秩表示。之后其又提出一种无监督的子空间学习方法——主成分系数嵌入(Principal Coefficients Embedding, PCE)^[25],该算法通过对样本进行奇异值分解得到重构系数矩阵和子空间维数,然后最小化子空间中的重构误差来保持数据的局部结构,计算复杂度低且具备一定的鲁棒性。Wen 等^[26]提出一种基于图正则化约束重构的低秩保持投影算法(Low-Rank Preserving Projection via Graph Regularized Reconstruction, LRPP_GRR),该算法通过最小化样本的重构误差和对重构系数施以低秩约束来保持数据的局部和全局结构,但和众多结构保持算法一样,近邻数的选择仍然是其面临的一大难题。

总体来说,基于表示关系的算法相比传统基于欧氏距离的方法虽然已经取得了很大突破,但是其只考虑数据的线性关系,而实际应用中数据多为非线性结构;此外,基于表示关系的算法大都源于样本间的直接关系建模,由此衍生的子空间学习算法或未能考虑样本的类标信息,或只利用了类内局部关系构图,而忽略了不同类样本之间的表示关系及其在构图中的作用。围绕上述两个问题,本文提出了基于核保持嵌入的相似性度量和子空间学习联合框架,其优点如下:

(1) 针对传统线性降维方法无法充分挖掘高维非线性数据内部结构的问题,将核函数引入样本的自表示学习中,并通过最小化样本的重构误差学习最优的相似性表示。

(2) 针对现有基于表示关系的子空间学习算法未能充分利用类间判别信息的缺点,利用核保持关系学习到的相似性表示分别构建类内图和类间图,增强了对类与类之间不相似性的利用,使得在投影子空间中,同类样本的核保持关系得到加强,不同类样本间的核保持关系被进一步抑制。

(3) 将核保持矩阵和图嵌入整合为一个联合框架,通过二者的迭代优化得到最优表示下的子空间投影。

2 基于核保持嵌入的子空间学习

2.1 算法描述

首先对本文用到的符号进行说明。矩阵用粗体大写字母表示,例如 \mathbf{X} 。向量用粗体小写字母表示,例如 \mathbf{x} 。矩阵的第 i 行第 j 列元素表示为 x_{ij} , \mathbf{I} 表示单位矩阵,1 表示元素均为 1 的矩阵。 $\mathbf{X}^T, \mathbf{X}^{-1}, \text{tr}(\cdot)$ 分别表示求 \mathbf{X} 的转置矩阵、逆矩阵和迹。 $\|\mathbf{X}\|_F^2$ 表示矩阵 \mathbf{X} 的 Frobenius 范数。 $\langle \mathbf{x}_i, \mathbf{x}_j \rangle = \mathbf{x}_i^T \cdot \mathbf{x}_j$ 表示内积操作。

本文提出的基于核保持嵌入的子空间学习算法的思路为:首先引入核函数,将原始数据映射到高维空间,通过映射

同类样本和每类样本的均值来分别构造类内核和类间核;其次使核空间的样本重构误差最小化,得到类内、类间的权值矩阵;最后根据图嵌入框架计算类内和类间图拉普拉斯矩阵,进而求解子空间投影,将原始数据映射到低维空间。

假设原始数据 $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_c\} \in R^{m \times n}$ 是一个含有 c 类、共 n 个样本的训练集。无论是在原始空间还是投影子空间中,都可以通过核保持挖掘数据的结构信息。参照文献[27]中的做法,采用内积作为样本关系的度量方式,将数据映射到高维空间后进行内积运算,然后最小化样本的重构误差以学习数据表示关系。为了解决在高维特征空间中计算的“维数灾难”问题,引入核函数将高维空间的内积运算转化为低维输入空间的计算。核保持问题可表示为:

$$\begin{aligned} \min_{\mathbf{Z}} \|\phi(\mathbf{X})^\top \cdot \phi(\mathbf{X}) - (\phi(\mathbf{X})\mathbf{Z})^\top \cdot (\phi(\mathbf{X})\mathbf{Z})\|_F^2 \\ \Rightarrow \min_{\mathbf{Z}} \|\mathbf{K} - \mathbf{Z}^\top \mathbf{K} \mathbf{Z}\|_F^2 \end{aligned} \quad (1)$$

假设在迭代过程中得到投影矩阵 \mathbf{P} ,那么可根据投影到子空间中的数据构建类内核和类间核,即:

$$\mathbf{K}_w = \text{diag}\{\mathbf{K}_{w1}, \mathbf{K}_{w2}, \dots, \mathbf{K}_{wc}\} \quad (2)$$

$$\mathbf{K}_b = \{\mathbf{K}_{b1}, \mathbf{K}_{b2}, \dots, \mathbf{K}_{bc}\} \quad (3)$$

其中, $\mathbf{K}_{wi} = \phi(\mathbf{P}^\top \mathbf{X}^i)^\top \cdot \phi(\mathbf{P}^\top \mathbf{X}^i)$, $\mathbf{K}_{bi} = \phi(\mathbf{M}^i)^\top \cdot \phi(\mathbf{M}^i)$, \mathbf{M}^i 为投影后样本每类的均值。增加正则项约束表示系数矩阵,通过下式分别构造类内表示系数矩阵 \mathbf{Z}_w 和类间表示系数矩阵 \mathbf{Z}_b :

$$\|\mathbf{K}_w - \mathbf{Z}_w^\top \mathbf{K}_w \mathbf{Z}_w\|_F^2 + \lambda_1 \|\mathbf{Z}_w\|_F^2 \quad (4)$$

$$\|\mathbf{K}_b - \mathbf{Z}_b^\top \mathbf{K}_b \mathbf{Z}_b\|_F^2 + \lambda_2 \|\mathbf{Z}_b\|_F^2 \quad (5)$$

其中, λ_1 和 λ_2 是正则化参数,防止算法过拟合。为避免表示系数为负,以及样本被自己表示的极端情况,我们进一步约束 $\mathbf{Z}_w, \mathbf{Z}_b$ 非负,并强制其对角元素为 0。同时为了避免平凡解,即一些样本在线性表示中不被选择,我们约束 $\mathbf{Z}_w, \mathbf{Z}_b$ 每一行的和为 1。

再由相似性矩阵 \mathbf{Z}_w 和 \mathbf{Z}_b 分别构造类内图拉普拉斯矩阵 \mathbf{L}_w 和类间图拉普拉斯矩阵 \mathbf{L}_b 。根据图嵌入的降维框架,将原始样本通过投影矩阵 $\mathbf{P} \in R^{m \times d}$ 映射到低维空间,使其在低维空间中同类样本的核保持关系增强,不同类样本的核保持关系被抑制,即:

$$\begin{aligned} \arg \min_{\mathbf{P}} J_1(\mathbf{P}) &= \sum_i \|\mathbf{P}^\top \mathbf{x}_i - \sum_j \mathbf{Z}_w^{(ij)} \mathbf{P}^\top \mathbf{x}_j\|_F^2 \\ &= \text{tr}(\mathbf{P}^\top \mathbf{X} \mathbf{L}_w \mathbf{X}^\top \mathbf{P}) \end{aligned} \quad (6)$$

$$\begin{aligned} \arg \max_{\mathbf{P}} J_2(\mathbf{P}) &= \sum_i \|\mathbf{P}^\top \mathbf{x}_i - \sum_j \mathbf{Z}_b^{(ij)} \mathbf{P}^\top \mathbf{x}_j\|_F^2 \\ &= \text{tr}(\mathbf{P}^\top \mathbf{X} \mathbf{L}_b \mathbf{X}^\top \mathbf{P}) \end{aligned} \quad (7)$$

其中, $\mathbf{L}_w = (\mathbf{I} - \mathbf{Z}_w)^\top (\mathbf{I} - \mathbf{Z}_w)$, $\mathbf{L}_b = (\mathbf{I} - \mathbf{Z}_b)^\top (\mathbf{I} - \mathbf{Z}_b)$ 。

为了同时满足类内和类间样本的核保持关系,以上两式可联立为:

$$\begin{aligned} \arg \min_{\mathbf{P}} J(\mathbf{P}) &= J_1(\mathbf{P}) - \beta J_2(\mathbf{P}) \\ &= \text{tr}(\mathbf{P}^\top \mathbf{X} \mathbf{L}_w \mathbf{X}^\top \mathbf{P}) - \beta \text{tr}(\mathbf{P}^\top \mathbf{M} \mathbf{L}_b \mathbf{M}^\top \mathbf{P}) \\ &= \text{tr}(\mathbf{P}^\top (\mathbf{X} \mathbf{L}_w \mathbf{X}^\top - \beta \mathbf{M} \mathbf{L}_b \mathbf{M}^\top) \mathbf{P}) \end{aligned} \quad (8)$$

将式(4)、式(5)与式(8)结合起来,得到最终的目标函数:

$$\begin{aligned} \min_{\mathbf{P}, \mathbf{K}_w, \mathbf{K}_b, \mathbf{Z}_w, \mathbf{Z}_b} \text{tr}(\mathbf{P}^\top (\mathbf{X} \mathbf{L}_w \mathbf{X}^\top - \beta \mathbf{M} \mathbf{L}_b \mathbf{M}^\top) \mathbf{P}) + \\ \alpha (\|\mathbf{K}_w - \mathbf{Z}_w^\top \mathbf{K}_w \mathbf{Z}_w\|_F^2 + \|\mathbf{K}_b - \mathbf{Z}_b^\top \mathbf{K}_b \mathbf{Z}_b\|_F^2) + \\ \lambda (\|\mathbf{Z}_w\|_F^2 + \|\mathbf{Z}_b\|_F^2) \end{aligned}$$

$$\begin{aligned} \text{s. t. } \mathbf{Z}_w, \mathbf{Z}_b \geq 0, \text{diag}(\mathbf{Z}_w) = 0, \text{diag}(\mathbf{Z}_b) = 0, \\ \mathbf{Z}_w \mathbf{1} = \mathbf{1}, \mathbf{Z}_b \mathbf{1} = \mathbf{1} \end{aligned} \quad (9)$$

其中, α, β 和 λ 为正则化参数。

2.2 优化过程

由式(9)可知,目标函数是关于 $\mathbf{Z}_w, \mathbf{Z}_b$ 的四阶函数,因为直接求解存在一定困难,所以需要分别引入辅助变量 \mathbf{J}, \mathbf{W} 和 \mathbf{Q}, \mathbf{N} ,具体表示如下:

$$\begin{aligned} \min \text{tr}(\mathbf{P}^\top (\mathbf{X} \mathbf{L}_w \mathbf{X}^\top - \beta \mathbf{M} \mathbf{L}_b \mathbf{M}^\top) \mathbf{P}) + \alpha (\|\mathbf{K}_w - \mathbf{J}^\top \mathbf{K}_w \mathbf{W}\|_F^2 + \\ \|\mathbf{K}_b - \mathbf{Q}^\top \mathbf{K}_b \mathbf{N}\|_F^2) + \lambda (\|\mathbf{Z}_w\|_F^2 + \|\mathbf{Z}_b\|_F^2) \\ \text{s. t. } \mathbf{Z}_w, \mathbf{Z}_b \geq 0, \text{diag}(\mathbf{Z}_w) = 0, \text{diag}(\mathbf{Z}_b) = 0, \\ \mathbf{Z}_w \mathbf{1} = \mathbf{1}, \mathbf{Z}_b \mathbf{1} = \mathbf{1}, \\ \mathbf{Z}_w = \mathbf{J}, \mathbf{Z}_w = \mathbf{W}, \mathbf{Z}_b = \mathbf{Q}, \mathbf{Z}_b = \mathbf{N} \end{aligned} \quad (10)$$

式(10)表示的问题可以分解成多个子问题,通过交替优化每个子问题来得到整体优化问题的近似闭式解。具体更新过程如下。

(1)固定 \mathbf{Z}_b, \mathbf{P} 和 $\mathbf{K}_w, \mathbf{K}_b$,通过 \mathbf{J} 和 \mathbf{W} 辅助更新 \mathbf{Z}_w ,对应的拉格朗日函数为:

$$\begin{aligned} \ell(\mathbf{Z}_w, \mathbf{J}, \mathbf{W}, \mathbf{Y}_1, \mathbf{Y}_2) \\ = \alpha \|\mathbf{K}_w - \mathbf{J}^\top \mathbf{K}_w \mathbf{W}\|_F^2 + \lambda \|\mathbf{Z}_w\|_F^2 + \\ \frac{\mu}{2} \left(\left\| \mathbf{Z}_w - \mathbf{J} + \frac{\mathbf{Y}_1}{\mu} \right\|_F^2 + \left\| \mathbf{Z}_w - \mathbf{W} + \frac{\mathbf{Y}_2}{\mu} \right\|_F^2 \right) \end{aligned} \quad (11)$$

其中, μ 是惩罚参数, $\mathbf{Y}_1, \mathbf{Y}_2$ 是拉格朗日算子。对式(11)求 \mathbf{J} 的导数,并令导数值为零,得到 \mathbf{J} 的更新规则:

$$\mathbf{J} = (\mu \mathbf{I} + 2\alpha \mathbf{K}_w \mathbf{W} \mathbf{W}^\top \mathbf{K}_w^\top)^{-1} (\mu \mathbf{Z}_w + \mathbf{Y}_1 + 2\alpha \mathbf{K}_w \mathbf{W} \mathbf{K}_w^\top) \quad (12)$$

同理,对式(11)求 \mathbf{W} 的导数并令其为零,可得 \mathbf{W} 的更新规则:

$$\mathbf{W} = (\mu \mathbf{I} + 2\alpha \mathbf{K}_w^\top \mathbf{J} \mathbf{J}^\top \mathbf{K}_w)^{-1} (\mu \mathbf{Z}_w + \mathbf{Y}_2 + 2\alpha \mathbf{K}_w^\top \mathbf{J} \mathbf{K}_w) \quad (13)$$

参考 Wen 等^[28]的方法,再对式(11)关于 \mathbf{Z}_w 求导并令其为零,可得 \mathbf{Z}_w 的近似闭式解:

$$\hat{\mathbf{Z}}_w = (2\lambda \mathbf{I} + 2\mu \mathbf{I})^{-1} (\mu(\mathbf{J} + \mathbf{W}) + \mathbf{Y}_1 + \mathbf{Y}_2) \quad (14)$$

然后根据下式求解目标解 \mathbf{Z}_w :

$$\min_{\mathbf{Z}_w \geq 0, \text{diag}(\mathbf{Z}_w) = 0, \mathbf{Z}_w \mathbf{1} = \mathbf{1}} \|\mathbf{Z}_w - \hat{\mathbf{Z}}_w\|_F^2 \quad (15)$$

对于式(15),可以得到 \mathbf{Z}_w 的每一行为:

$$\mathbf{z}_k = \max(\zeta_k \bar{\mathbf{1}}^\top + \bar{\mathbf{z}}_k, 0) \quad (16)$$

其中, $\bar{\mathbf{z}}_k = [\bar{z}_{k1}, \dots, \bar{z}_{kk}, \dots, \bar{z}_{kn}]$ 是 $\hat{\mathbf{Z}}_w$ 的第 k 行且 \bar{z}_{kk} 被设置为 0; $\bar{\mathbf{1}}$ 是第 k 个元素为 0、其余全为 1 的列向量; ζ_k 是拉格朗日乘子,可根据下式计算:

$$\zeta_k = (1 + \bar{\mathbf{z}}_k \mathbf{1}) / (n - 1) \quad (17)$$

由此可得 \mathbf{Z}_w 的更新规则。

(2)固定 $\mathbf{Z}_w, \mathbf{P}, \mathbf{K}_w$ 和 \mathbf{K}_b ,通过 \mathbf{Q} 和 \mathbf{N} 辅助更新 \mathbf{Z}_b ,对应的拉格朗日函数为:

$$\begin{aligned} \ell(\mathbf{Z}_b, \mathbf{Q}, \mathbf{N}, \mathbf{Y}_3, \mathbf{Y}_4) \\ = \alpha \|\mathbf{K}_b - \mathbf{Q}^\top \mathbf{K}_b \mathbf{N}\|_F^2 + \lambda \|\mathbf{Z}_b\|_F^2 + \frac{\mu}{2} \left(\left\| \mathbf{Z}_b - \mathbf{Q} + \frac{\mathbf{Y}_3}{\mu} \right\|_F^2 + \right. \\ \left. \left\| \mathbf{Z}_b - \mathbf{N} + \frac{\mathbf{Y}_4}{\mu} \right\|_F^2 \right) \end{aligned} \quad (18)$$

其中, μ 是惩罚参数, $\mathbf{Y}_3, \mathbf{Y}_4$ 是拉格朗日算子。对式(18)求 \mathbf{Q} 的导数,并令导数值为零,得到 \mathbf{Q} 的更新规则:

$$\mathbf{Q} = (\mu \mathbf{I} + 2\alpha \mathbf{K}_b \mathbf{N} \mathbf{N}^\top \mathbf{K}_b^\top)^{-1} (\mu \mathbf{Z}_b + \mathbf{Y}_3 + 2\alpha \mathbf{K}_b \mathbf{N} \mathbf{K}_b^\top) \quad (19)$$

同理对式(18)求 N 的导数并令其为零,可得 N 的更新规则:

$$N = (\mu I + 2\alpha K_b^T Q Q^T K_b)^{-1} (\mu Z_b + Y_4 + 2\alpha K_b^T Q K_b) \quad (20)$$

同理,参考 Wen 等^[28]的方法,对式(18)关于 Z_b 求导并令其为零,可得 Z_b 的近似闭式解:

$$\hat{Z}_b = (2\lambda I + 2\mu I)^{-1} (\mu(Q + N) + Y_3 + Y_4) \quad (21)$$

然后根据下式求解目标解 Z_b :

$$\min_{z_b \geq 0, \text{diag}(Z_b) = 0, z_{b1} = 1} \|Z_b - \hat{Z}_b\|_F^2 \quad (22)$$

对于式(22),可以得到 Z_b 的每一行为:

$$z_r = \max(\zeta_r \bar{1}^T + \bar{z}_r, 0) \quad (23)$$

其中, $\bar{z}_r = [\bar{z}_{r1}, \dots, \bar{z}_{rr}, \dots, \bar{z}_{rc}]$ 是 \hat{Z}_b 的第 r 行且 \bar{z}_{rr} 被设置为 0; $\bar{1}$ 是第 r 个元素为 0、其余全为 1 的列向量; ζ_r 是拉格朗日乘子,可根据下式计算:

$$\zeta_r = (1 + \bar{z}_r) / (n - 1) \quad (24)$$

由此可得 Z_b 的更新规则。

(3) 固定 Z_w , Z_b , K_w 和 K_b , 更新 P 。根据式(8),对 $(X L_w X^T - \beta M L_b M^T)$ 进行奇异值分解,得到前 d 个最小特征值对应的特征向量,构成投影矩阵:

$$P = [p_1, p_2, \dots, p_d] \in R^{m \times d} \quad (25)$$

(4) 固定 Z_w , Z_b 和 P , 更新 K_w , K_b 。对投影到低维子空间后的样本根据式(2)和式(3)进行更新。

在得到最优子空间的投影后,将测试样本投影到低维空间,并利用最近邻分类准则预测其类别。

KESL 的算法流程如算法 1 所示。

算法 1 KESL

输入:含有 C 类样本的训练集 $X \in R^{m \times n}$, 测试集 X_{test} , 正则化参数 λ ,

α, β , 迭代次数 t_{max}

输出:权值矩阵 Z_w 和 Z_b , 投影矩阵 $P \in R^{k \times d}$

初始化:单位阵 $P, Z_w, Z_b, Y_1 = Y_2 = Y_3 = Y_4 = 0$, 参数 $\mu = 0.1, \rho =$

1. $\mu_{\text{max}} = 10^8$, 通过式(2)和式(3)计算 K_w, K_b 。

训练:

1. 通过 PCA 算法对样本进行预处理,该投影矩阵记为 $P_{\text{PCA}} \in R^{m \times k}$;

2. 根据式(12)更新 J ;

3. 根据式(13)更新 W ;

4. 根据式(16)更新 Z_w ;

5. 根据式(19)更新 Q ;

6. 根据式(20)更新 N ;

7. 根据式(23)更新 Z_b ;

8. 根据式(25)更新 P ;

9. 根据式(2)和式(3)更新 K_w, K_b ;

10. 更新拉格朗日算子和参数 μ :

$$Y_1 = Y_1 + \mu(Z_w - J),$$

$$Y_2 = Y_2 + \mu(Z_w - W)$$

$$Y_3 = Y_3 + \mu(Z_b - Q)$$

$$Y_4 = Y_4 + \mu(Z_b - N)$$

$$\mu = \min(\rho\mu, \mu_{\text{max}})$$

迭代次数 $t = t + 1$;

11. 如果式(9)达到收敛条件或者 $t = t_{\text{max}}$, 训练结束,否则跳转到步骤

(2)继续执行。

测试:

$$1. X_{\text{newtrain}} = P^T P_{\text{PCA}}^T X \text{ 和 } X_{\text{newtest}} = P^T P_{\text{PCA}}^T X_{\text{test}};$$

2. 使用最近邻分类器对投影后的低维测试样本进行分类。

3 实验结果与分析

本节对所提 KESL 算法在 3 个数据集 (ORL^[29]、UMIST^[30] 和雷达辐射源数据集^[19]) 上进行分类性能测试,具体数据集信息如表 1 所列。在数据预处理阶段,前两个数据集保留 99% 的能量,雷达辐射源数据集保留 100% 的能量,将样本映射到 PCA 子空间中。所选取的对比算法有 CRC^[17]、PCA^[5]、MFA^[9]、CGDA^[23]、KPCA^[7]、KFDA^[8]、L2-Graph^[24]、PCE^[25] 和 LRPP_GRR^[26], 其中,因为 CRC 没有实现降维,所以将其性能作为基准;PCA, KPCA, KFDA, L2-Graph, PCE 和 LRPP_GRR 未考虑标签信息,属于无监督算法, MFA 和 CGDA 是有监督算法。为了保证对比的公平性,所有算法的最优参数组合均在 $\{10^{-3}, 10^{-2}, \dots, 10^3\}$ 区间内采用网格搜索获得。所有实验设置最大训练次数为 20。为降低随机选取样本造成的实验偏差,对数据进行 10 次随机划分,汇报各方法在最优参数下的平均正确率及投影空间维度。所有实验均在软件环境为 MATLAB R2019b, 硬件环境为 8.00 GB RAM, 3.00 GHz CPU, 64 位 Windows 操作系统的计算机上进行。

表 1 实验采用的数据集信息

Table 1 Information of used datasets

Dataset	Samples	Dim	Classes
ORL	400	1024	40
UMIST	564	2576	20
雷达辐射源数据	665	512	10

实验 1 不同训练数目下各算法的识别率比较

本实验对比各算法的整体识别性能。对于 ORL 人脸数据集,每类分别随机选择 4, 5, 6 个样本作为训练集,其余样本作为测试集。对于 UMIST 人脸数据集,每类随机选择 4, 5, 6 个样本图像作为训练集,其余样本作为测试集。而对于自建的雷达辐射源数据集,提取信号的 512 维 welch 功率谱作为特征,每类随机选择 5, 10, 15 个样本作为训练集,其余样本作为测试集。表 2 列出了各算法在各数据集上不同训练样本数目情况下 10 次实验的平均分类正确率,括号内为取得最高识别率的投影空间维度。图 1 给出了各算法在 ORL 数据集上分类正确率随投影维度变化的曲线图。由图 1 可得出以下结论。

(1) 不论是人脸数据集还是雷达辐射源数据集,随着样本数量的增加, KESL 算法的分类正确率都逐渐提升,且优于对比算法,这种优势在低维时更加明显,说明所提算法能更好地挖掘出样本的结构信息,保持了数据间的核表示关系。

(2) 对比方法中的无监督算法,或没有构图(如 PCA 和 KPCA),或只利用样本间的直接关系建模(如 L2-Graph 和 LRPP_GRR)。所提算法 KESL 不仅充分利用了类标信息,而且引入了核函数,通过最小化重构误差挖掘数据的非线性结构来得到最优的表示系数。

(3)有监督算法 CGDA 只利用类内局部关系构图而忽略了类间判别信息,KESL 则利用核保持关系学习到的相似性

表示分别构建类内图和类间图,增强了对类与类之间不相似性的利用,表现出更好的分类性能。

表 2 不同训练数目下各算法在 3 个数据集上的分类正确率

Table 2 Classification accuracy of different algorithms on 3 datasets with different number of training samples

(单位:%)

Algorithms	ORL			UMIST			雷达辐射源数据		
	4-Train	5-Train	6-Train	4-Train	5-Train	6-Train	5-Train	10-Train	15-Train
CRC	79.00(80)	81.30(100)	84.50(100)	74.01(60)	81.59(70)	83.92(80)	88.89(40)	89.86(40)	94.03(40)
PCA	80.79(34)	85.85(88)	88.56(82)	74.83(60)	82.09(56)	86.10(76)	85.13(40)	88.21(40)	92.68(38)
MFA	91.21(50)	94.10(62)	96.06(84)	84.90(42)	90.78(36)	92.79(38)	84.90(4)	94.45(8)	95.65(6)
CGDA	89.25(44)	92.45(62)	96.00(68)	78.39(60)	86.55(60)	89.46(62)	92.98(10)	92.78(10)	93.73(12)
KPCA	82.33(48)	87.70(58)	89.50(64)	74.74(28)	81.16(62)	86.71(32)	94.18(6)	95.01(8)	95.96(6)
KFDA	88.42(40)	92.70(40)	94.37(40)	78.88(18)	84.09(20)	87.75(20)	88.88(6)	92.07(6)	93.01(6)
L2-Graph	82.17(74)	88.72(72)	90.88(84)	74.55(48)	80.43(70)	84.41(80)	82.57(10)	78.65(16)	73.28(16)
PCE	82.33(78)	88.30(94)	90.00(98)	75.74(54)	81.03(52)	85.99(80)	92.62(22)	95.50(22)	96.13(32)
LRPP_GRR	84.08(50)	89.70(70)	91.75(60)	78.47(60)	83.62(50)	88.78(54)	93.50(10)	94.30(10)	95.61(16)
KESL	93.33(30)	97.00(46)	98.00(34)	86.98(26)	90.60(8)	94.64(8)	94.86(10)	95.75(10)	96.83(20)

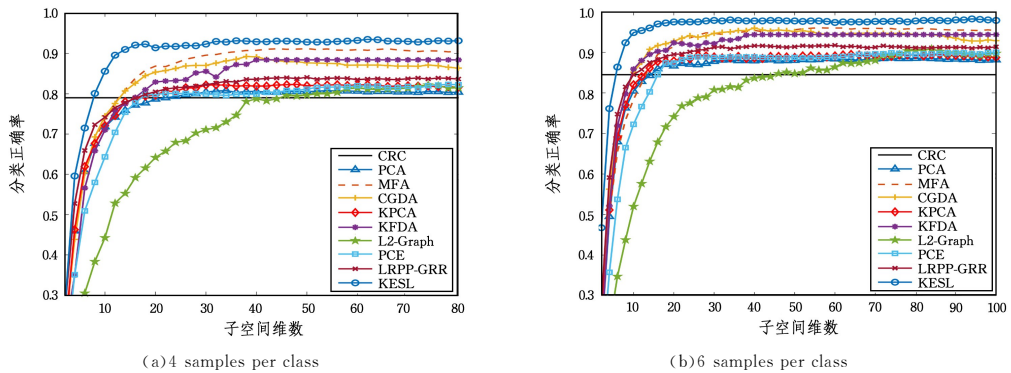


图 1 ORL 数据集上各方法分类正确率随维度变化的曲线

Fig. 1 Classification accuracy of different algorithms with different dimensions on ORL dataset

实验 2 参数敏感性分析

由式(9)可知,本算法有 3 个参数需要调节。为了衡量不同的参数对算法分类性能的影响,对算法在 ORL 数据集上在以下 3 种情况下进行实验:1)固定 α ,调整 β 和 λ ;2)固定 β ,调整 α 和 λ ;3)固定 λ ,调整 α 和 β 。实验在每类样本中随机选取

4 个样本构成训练集,其余样本作为测试集。

图 2 给出了 ORL 数据集上 KESL 分类正确率随参数变化的三维柱状图。由图 2 可知,所提算法的识别率随参数变化的幅度都很小,这说明所提算法对参数变化均不敏感,参数选择并不困难。

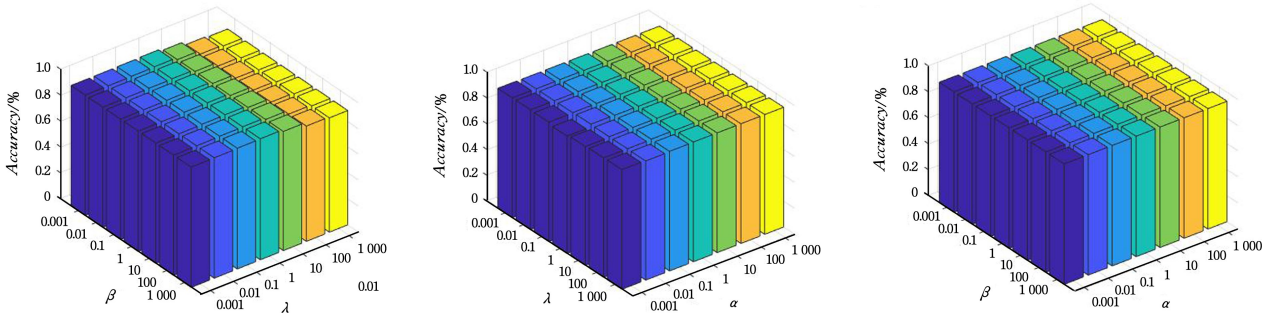


图 2 ORL 数据集上 KESL 分类正确率随参数变化的三维柱状图

Fig. 2 Parameter sensitivity analysis of KESL under different parameters on ORL dataset

实验 3 算法收敛性

为了考查算法是否具有良好的收敛性,以 ORL 数据集为例,本文给出了所提算法在整个训练过程中目标函数值的变化曲线。如图 3 所示,函数值在第 2、第 3 次迭代后就趋于平稳,可见 KESL 具有良好的收敛特性。

实验 4 算法训练时间的比较

除分类性能以外,算法训练阶段的时间消耗也是衡量算法有效性的重要指标。本实验比较了所提算法与对比算法在上述 3 个数据集上的训练时间,结果如表 3 所列。由表 3 可知,所提算法的耗时与 MFA,CGDA 等经典图嵌入方

法在同等水平上;其相比 LRPP_GRR 具有更低的时间代价。因此,所提算法有着较高的识别率且没有高昂的时间开销,这进一步证明了其有效性。

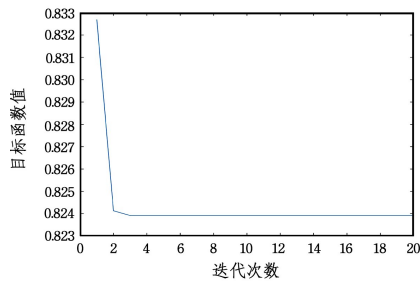


图3 KESL在ORL数据集上的收敛曲线

Fig. 3 Convergence curve of KESL on ORL dataset

表3 各算法在不同数据集上的训练时间

Table 3 Training time of each algorithm on different datasets

Algorithms	(单位:s)		
	ORL	UMIST	雷达辐射源数据
CRC	0.479	0.593	0.484
PCA	0.369	0.634	0.499
MFA	0.481	0.648	1.511
CGDA	0.666	0.695	0.786
KPCA	1.765	4.001	3.859
KFDA	1.496	2.811	2.882
L2-Graph	0.517	1.834	1.756
PCE	1.793	2.100	1.780
LRPP_GRR	11.727	67.83	11.275
KESL	3.200	4.652	4.053

结束语 本文提出一种新的基于核保持嵌入的子空间学习算法。该算法具有以下特点:1)引入核函数并最小化样本的重构误差来约束最优的表示系数,挖掘出有利于分类的数据结构关系;2)同时考虑类内样本相似性信息和类间样本的判别性信息,利用学习到的相似性矩阵分别构建类内和类间图,使得在投影子空间中同类样本的核保持关系得到加强而不同类样本的核保持关系被有效抑制;3)通过核保持矩阵与图嵌入的联合优化,动态地求解出最优表示下的子空间投影。在3个数据集上的实验结果充分验证了KESL较之传统子空间学习方法的优点,其实现了数据非线性结构和判别信息的有效挖掘。需要说明的是,尽管引入了核函数,在对比实验中,本文算法仅采用了线性核,因此如何设计更优的核函数以进一步提升其性能是未来值得研究的课题。

参考文献

[1] CHEN X J, YE Y M, XU X F, et al. A feature group weighting method for subspace clustering of high-dimensional data[J]. Pattern Recognition, 2012, 45(1): 434-446.

[2] HUANG S, ELGAMMAL A M, YANG D. Learning Speed Invariant Gait Template via Thin Plate Spline Kernel Manifold Fitting[C]//British Machine Vision Conference 2013. 2013.

[3] LAI Z H, BAO J Q, KONG H, et al. Discriminative low-rank projection for robust subspace learning[J]. International Journal

of Machine Learning and Cybernetics, 2020, 11(5): 2247-2260.

[4] LI J X, ZHAO Z G, LI Q, et al. Improved Locality and Similarity Preserving Feature Selection Algorithm[J]. Computer Science, 2020, 47(S1): 480-484.

[5] WOLD S, ESSENSEN K H, GELADI P, et al. Principal Component Analysis[J]. Chemometrics and Intelligent Laboratory Systems, 1987, 2(1/2/3): 37-52.

[6] FISHER R A. The use of multiple measurements in taxonomic problems[J]. Annals of Human Genetics, 1936, 7(2): 179-188.

[7] KIM K I, JUNG K, KIM H J, et al. Face recognition using kernel principal component analysis[J]. IEEE Signal Processing Letters, 2002, 9(2): 40-42.

[8] MIKA S, RATSCHE G, WESTON J, et al. Fisher discriminant analysis with kernels[C]//Neural Networks for Signal Processing IX, 1999. Proceedings of the IEEE Signal Processing Society Workshop. IEEE, 1999: 41-48.

[9] YAN S C, XU D, ZHANG B Y, et al. Graph Embedding and Extensions: A General Framework for Dimensionality Reduction [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(1): 40-51.

[10] XUE X Y, MA X H. Double Adjacency Graphs Based Orthogonal Neighborhood Preserving Projections for Face Recognition [J]. Computer Science, 2017, 44(8): 31-35.

[11] YIN M, GAO J, LIN Z, et al. Laplacian Regularized Low-Rank Representation and Its Applications[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(3): 504-517.

[12] NG A Y, JORDAN M I, WEISS Y, et al. On Spectral Clustering: Analysis and an algorithm[C]//Neural Information Processing Systems. 2001: 849-856.

[13] WEINBERGER K Q, SAUL L K. Distance Metric Learning for Large Margin Nearest Neighbor Classification[J]. Journal of Machine Learning Research, 2009, 10(1): 207-244.

[14] JIAO X, CHEN Y G, DONG R. An Unsupervised Image Segmentation Method Combining Graph Clustering and High-Level Feature Representation [J]. Neurocomputing, 2020, 409(7): 83-92.

[15] HIRZER M, ROTH P M, KÖSTINGER M, et al. Relaxed pairwise learned metric for person re-identification[C]//European Conference on Computer Vision. Berlin, Heidelberg: Springer, 2012: 780-793.

[16] HOI S C, LIU W, CHANG S, et al. Semi-supervised distance metric learning for Collaborative Image Retrieval[C]//Computer Vision and Pattern Recognition. 2008: 1-7.

[17] ZHANG L, YANG M, FENG X, et al. Collaborative Representation based Classification for Face Recognition[J]. arXiv: 1204. 2358, 2012.

[18] TAO Z, LIU H, LI S, et al. From Ensemble Clustering to Multi-View Clustering[C]//International Joint Conference on Artificial Intelligence. 2017: 2843-2849.

[19] WANG L, LI M, JI H, et al. When collaborative representation

- meets subspace projection: A novel supervised framework of graph construction augmented by anti-collaborative representation[J]. *Neurocomputing*, 2019, 328: 157-170.
- [20] WRIGHT J, YANG A Y, GANESH A, et al. Robust Face Recognition via Sparse Representation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(2): 210-227.
- [21] QIAO L S, CHEN S C, TAN X Y, et al. Sparsity preserving projections with applications to face recognition[J]. *Pattern Recognition*, 2010, 43(1): 331-341.
- [22] ZHANG L, YANG M, FENG X C, et al. Sparse representation or collaborative representation: Which helps face recognition? [C] // *International Conference on Computervision*. 2011: 471-478.
- [23] LY N H, DU Q, FOWLER J E. Collaborative graph-based discriminant analysis for hyperspectral imagery[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2014, 7(6): 2688-2696.
- [24] PENG X, YU Z, YI Z, et al. Constructing the L2-Graph for Robust Subspace Learning and Subspace Clustering [J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 2017, 47(4): 1053-1066.
- [25] PENG X, LU J, YI Z, et al. Automatic Subspace Learning via Principal Coefficients Embedding [J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 2017, 47(11): 3583-3596.
- [26] WEN J, HAN N, FANG X, et al. Low-Rank Preserving Projection Via Graph Regularized Reconstruction [J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 2019, 49(4): 1279-1291.
- [27] KANG Z, LU Y, SU Y, et al. Similarity Learning via Kernel Preserving Embedding [C] // *National Conference on Artificial Intelligence*. 2019: 4057-4064.
- [28] WEN J, ZHANG B, XU Y, et al. Adaptive weighted nonnegative low-rank representation[J]. *Pattern Recognition*, 2018, 81: 326-340.
- [29] SAMARIA F S, HARTER A. Parameterisation of a stochastic model for human face identification [C] // *Workshop on Applications of Computer Vision*. 1994: 138-142.
- [30] GRAHAM D B, ALLINSON N M. Characterising virtual eigen-signatures for general purpose face recognition[M]. Berlin, Heidelberg: Springer, 1998: 446-456.



HE Wen-qi, born in 1996, postgraduate, is a member of China Computer Federation. Her main research interests include subspace learning and multi-view learning.



LI Dan-ping, born in 1981, Ph.D, lecturer. Her main research interests include wireless signal processing and machine learning.