基于自适应流形正则化自表示的无监督特征选择算法

宋 雨,许王琴,李荣鹏,宋学力,肖玉柱
长安大学 理学院,西安 710064

摘 要:针对基于流形正则化自表示(MRSR)的无监督特征选择算法直接从原始的样本空间构造相似矩阵可能会导致重构空间中样本的相似性描述得不够准确的问题,提出了基于自适应流形正则化自表示的无监督特征选择(AMRSR)算法。基于自适应流形正则化自表示的无监督特征选择算法在 MRSR 算法的基础上通过对相似矩阵施加概率最近邻约束将相似矩阵的学习嵌入到优化过程中,在重构空间中自适应地学习样本的相似性,使得在每一次迭代中获取更加精确的样本局部几何流形结构,从而选择具有代表性且保持局部几何流形结构的特征。最后,在四个公开数据集上进行了大量的对比实验,通过将算法的特征选择结果用于 K-means 聚类并采取两种常见的聚类评价指标:聚类精确度和归一化互信息评价聚类效果。实验结果表明,AMRSR 算法与现有的一些算法相比有更高的聚类精确度和归一化互信息,进一步表明该算法特征选择效果更好。

关键词:无监督特征选择;自表示;流形正则化;自适应;相似矩阵
中图分类号:TP181 **文献标识码**:A doi:10.16055/j.issn.1672-058X.2023.0006.006

Unsupervised Feature Selection Algorithm Based on Adaptive Manifold Regularization Self-representation

SONG Yu, XU Wangqin, LI Rongpeng, SONG Xueli, XIAO Yuzhu
School of Science, Chang' an University, Xi' an 710064, China

Abstract: Unsupervised feature selection algorithm based on manifold regularization self-representation (MRSR) directly constructed similarity matrix from the original sample space, which might lead to inaccurate similarity description of samples in the reconstructed space. To solve this problem, an unsupervised feature selection algorithm based on adaptive manifold regularization self-representation (AMRSR) was proposed. On the basis of MRSR algorithm, unsupervised feature selection algorithm based on adaptive manifold regularization self-representation embedded the learning of similar matrix into the optimization process by imposing probabilistic nearest neighbor constraints on similar matrix, and adaptively learned the similarity of samples in the reconstructed space, so that more accurate local geometric manifold structure of samples could be obtained in each iteration, and then representative features with local geometric manifold structure could be selected. Finally, a large number of comparative experiments were carried out on four public datasets. By applying the feature selection results of the algorithms to K-means clustering, two common clustering evaluation indexes were adopted: clustering accuracy and normalized mutual information to evaluate the clustering effect. Experimental results show that the AMRSR algorithm has higher clustering accuracy and normalized mutual information than some existing algorithms, which further indicates that the feature selection effect of this algorithm is better.

Keywords: unsupervised feature selection; self-representation; manifold regularization; adaptation; similar matrix

收稿日期:2022-09-26 **修回日期**:2022-10-26 **文章编号**:1672-058X(2023)06-0044-09
基金项目:长安大学中央高校基本科研业务费专项资金资助项目(310812163504).
作者简介:宋雨(1998—),女,四川资阳人,硕士研究生,从事机器学习、数据挖掘研究.
通讯作者:宋学力(1980—),男,河南开封人,教授,博士,从事深度学习、统计过程控制等研究. Email:xlsung@chd.edu.cn.
引用格式:宋雨,许王琴,李荣鹏,等. 基于自适应流形正则化自表示的无监督特征选择算法[J]. 重庆工商大学学报(自然科学版),2023,40(6):44—52.
SONG Yu, XU Wangqin, LI Rongpeng, et al. Unsupervised feature selection algorithm based on adaptive manifold regularization self-representation[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2023, 40(6): 44—52.

1 引　言

随着人工智能技术的突破和大数据时代的快速发展,各行业领域采集并储存了大量的高维数据。然而,高维数据中存在的大量冗余特征、无关特征和噪声增加了数据处理的复杂程度,带来了“维数灾难”问题,这给传统的数据处理方法带来了前所未有的挑战[1]。作为一种有效的数据降维技术,特征选择一般被用于高维数据的预处理过程中。在确保不丢失原始数据集中重要的特征前提下,特征选择通过相应的评价标准从原始特征集合中选择出一个满足条件的最优特征子集,以此来达到降低数据维数的目的,从而避免了由数据维数过高带来的一系列不利影响[2]。

根据数据集有无标签信息,特征选择方法可以大致分为有监督、半监督和无监督的特征选择方法[3]。在实际生活中收集到的高维数据大多数是没有标签的,而标签是指导模型搜索相关特征的重要信息。因此,无监督特征选择方法的研究相比于有标签信息的特征选择方法更艰难且充满挑战[4]。无监督特征选择方法一般会选择能保留与原始数据样本相似的全局或局部结构和包含判别信息的特征[5]。如何利用原始数据来发掘这些相似性结构和判别信息是研究无监督特征选择方法的关键。

基于自然界中物体的自相似性,即物体的一部分与自身的其他部分相似,有学者提出了基于自表示的方法并且应用到无监督特征选择算法中[6]。具体地,Zhu 等[6]提出了基于正则化自表示的无监督特征选择算法(Unsupervised Feature Selection by Regularized Self-Representation, RSR),通过使用 $L_{2,1}$ 范数分别对表示残差矩阵和表示系数矩阵进行约束,抑制异常值对损失项的影响,有效地删除了冗余特征并选择了较为重要的特征;Zhu 等[7]提出了非凸正则自表示的无监督特征选择算法(Non-convex Regularized Self-Representation for Unsupervised Feature Selection),选择 $L_{2,p}$ 范数($0\leqslant p<1$)对表示系数矩阵进行约束,从而获取更稀疏的自表示系数的解,取得了更有效的特征选择效果。由于 Zhu 等没有考虑样本空间的几何结构,可能会造成模型训练过程中样本的相似性结构发生变化,从而导致特征选择的效果不佳。针对这个问题,Liang 等[3]受到流形学习的启发,提出了基于流形正则化自表示的无监督特征选择算法(Unsupervised Feature Selection by Manifold Regularized Self- Representation, MRSR),在目标函数中加入了对结构学习的流形正则化项,从而来选择出最具代表性且能保持局部结构的特征子集;Li 等[8]提出了结合子空间学习和特征自表示的无监督特征选择算法(Unsupervised Feature Selection by Combining Subspace Learning with Feature Self- Representation),利用数据的性质构造自表示系数矩阵,通过稀疏表示寻找自表示系数矩阵的稀疏结构,并嵌入超图拉普拉斯正则化项来弥补普通图在多关系表示中的不足,从而优化特征选择的结果。上述算法均利用特征自表示的特性将特征的重要性表示出来,但忽略了重构过程中数据内部结构的变化。如果差异过大会进一步影响特征选择的效果。

流形正则化项要求原始空间中的相似样本应该在变换后的空间中依然保持接近,因此被广泛用于无监督特征选择中,构造出了一系列基于流形正则化的无监督特征选择算法。大多数现有的算法构造相似矩阵都是通过计算原始的样本空间中的相似性,并且在学习过程中是固定的。然而,样本的相似性矩阵构造不准确,这会导致特征选择的效果不佳。为了解决这一问题,Tang 等[9]提出了基于对偶自表示和流形正则化的鲁棒无监督特征选择算法(Robust Unsupervised Feature Selection via Dual Self-Representation and Manifold Regularization, DSRMR),一方面使用特征自表示项来学习特征表示系数矩阵,以度量不同特征维度的重要性,另一方面使用样本自表示项来自动学习样本相似图,以保持数据的局部几何结构;Du 等[10]提出了基于自适应结构学习的无监督特征选择算法(Unsupervised Feature Selection with Adaptive Structure Learning, FSASL),在模型训练过程中可以同时执行结构学习和特征选择,在保持样本结构的同时选择重要的特征;Zhang 等[11]将相似矩阵的构造嵌入到优化过程中,与最大化类间散度矩阵的思想结合提出了自适应图学习和约束的无监督特征选择算法(Unsupervised Feature Selection via Adaptive Graph Learning and Constraint, EGCFS),从而选择出不相关但有区别的特征。相似矩阵的构造和约束是基于流形正则化的无监督特征选择方法的关键。自适应图学习的方法将相似矩阵的构造嵌入到优化过程中,不仅使图的相似矩阵自适应生成,而且使算法在每一步迭代中获取精确的数据样本的相似性结构,

最终提高了算法的性能。从这些算法的结果中可以了解到,有流形正则化项的无监督特征选择算法的性能普遍比没有流形正则化项的算法要好,因为它可以帮助模型识别和保留了固有的几何流形结构。

MRSR 算法采用文献[12]的方法,即学习的拉普拉斯矩阵在整个优化过程中是固定的,并且是在原始数据集上学习得到的,而在每一次迭代过程中,重构空间一直在发生变化。随着算法的迭代,如果在原始空间中学习到的样本相似性不准确,可能会导致算法的输出结果即选择的特征不理想。本文针对 MRSR 算法中样本的相似矩阵在重构空间中构造不准确的问题,将相似矩阵的学习嵌入到优化过程中,提出了基于自适应流形正则化自表示的无监督特征选择算法(Unsupervised Feature Selection Algorithm based on Adaptive Manifold Regularization Self-Representation, AMRSR)。设算法对相似矩阵施加了一个约束,即以更大的概率连接更近的数据点,不仅可以自适应地学习重构空间中的相似矩阵,而且可以获得相似矩阵的系数的闭式解[11],以较好地保持每一次迭代中重构空间样本间的相似性,从而选择出更具代表性且能保持数据几何结构的特征。

2 问题形式化

2.1 符号说明

给定样本数据集 $\boldsymbol{X}\in R^{n\times m}$,$n$ 和 m 分别表示该数据集的样本数和特征维数,其中 $\boldsymbol{x}_i\in R^m$ 表示第 i 个样本,$\boldsymbol{f}_j\in R^n$ 表示第 j 个特征。$\boldsymbol{w}_i$ 表示 $\boldsymbol{W}=[w_{ij}]\in R^{m\times m}$ 的第 i 行,$\|\boldsymbol{W}\|_{2,1}=\sum_{i=1}^{m}\sqrt{\sum_{j=1}^{m}w_{ij}^2}=\sum_{i=1}^{m}\|\boldsymbol{w}_i\|_2$ 表示 $\boldsymbol{W}$ 的 $L_{2,1}$ 范数,其中 $\|\boldsymbol{w}_i\|_2$ 表示向量 $\boldsymbol{w}_i$ 的 L_2 范数;Tr(·)表示迹的运算符。

2.2 相关工作

特征选择算法通常将特征选择问题转换为一个多输出的回归问题[6],如式(1)所示:

$$\min_{\boldsymbol{W}}\mathrm{L}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{W})+\lambda\mathrm{R}(\boldsymbol{W}) \tag{1}$$

其中,$\boldsymbol{Y}\in R^{n\times m}$ 是响应矩阵,$\boldsymbol{W}\in R^{m\times m}$ 是特征权重矩阵,$\mathrm{L}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{W})$ 是损失(Loss)项,$\mathrm{R}(\boldsymbol{W})$ 是施加在 $\boldsymbol{W}$ 的正则化项,λ 是正则化参数。

由于在无监督特征选择中很难选到一个合适的响应矩阵 $\boldsymbol{Y}$,依据特征的自表示特性,基于自表示的无监督特征选择算法一般使用数据矩阵 $\boldsymbol{X}$ 作为响应矩阵,即 $\boldsymbol{Y}=\boldsymbol{X}$,可以理解为每一个特征都能由相关特征线性表示(包括它自身)。对于每个特征有

$$\boldsymbol{f}_i=\sum_{j=1}^{m}\boldsymbol{f}_j w_{ji}+\boldsymbol{e}_i \tag{2}$$

则对于所有特征有

$$\boldsymbol{X}=\boldsymbol{X}\boldsymbol{W}+\boldsymbol{E} \tag{3}$$

其中,$\boldsymbol{W}$ 是特征自表示系数矩阵,$\boldsymbol{E}$ 是表示残差矩阵。

RSR 算法采用 $L_{2,1}$ 范数约束表示残差矩阵 $\boldsymbol{E}$,避免了 Frobenius 范数对数据异常值敏感的问题。此外,自表示系数矩阵 $\boldsymbol{W}$ 的第 i 行的值对应的是第 i 个特征在表示特征时的系数,那么 $\|\boldsymbol{w}_i\|_2$ 可以衡量第 i 个特征的重要性。如果 $\|\boldsymbol{w}_i\|_2=0$,表示第 i 个特征在重构其他特征时不起作用,则不选择该特征;如果某个特征参与了其他所有特征的重构,表示这个特征很重要,则选择该特征。那么,用 $L_{2,1}$ 范数约束表示系数矩阵 $\boldsymbol{W}$ 作为正则化项,可以获得一个行稀疏的表示系数矩阵 $\boldsymbol{W}$,用来衡量每个特征的重要性,起到了选择特征的效果。因此,RSR 算法的模型如式(4)所示:

$$\min_{\boldsymbol{W}}\|\boldsymbol{X}-\boldsymbol{X}\boldsymbol{W}\|_{2,1}+\lambda_1\|\boldsymbol{W}\|_{2,1} \tag{4}$$

其中,λ_1 是正则化参数。

但此时,该模型忽略了样本之间的相似性结构,因为在重构空间中样本点之间的几何结构可能会发生变化。另外有研究表明,数据的内部结构其实接近于欧几里德环境空间的流形采样。所以在考虑数据结构时,数据的几何流形结构可能更能反映数据内部实际的结构状态[13]。为了能在重构空间中很好地保持原始特征空间中的样本相似性,通过利用拉普拉斯矩阵,将低维流形数据 $\boldsymbol{X}\boldsymbol{W}$ 嵌入到拉普拉斯图中用于保持数据原始的几何结构。MRSR 算法采用文献[12]的方法在原始数据上学习一个鲁棒拉普拉斯矩阵,并且在整个优化过程中一直保持不变。因此,MRSR 算法的模型如式(3)所示:

$$\min_{\boldsymbol{W}}\|\boldsymbol{X}-\boldsymbol{X}\boldsymbol{W}\|_{2,1}+\lambda_1\|\boldsymbol{W}\|_{2,1}+\lambda\mathrm{Tr}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{X}\boldsymbol{W}) \tag{5}$$

其中,$\boldsymbol{L}\in R^{n\times n}$ 是拉普拉斯矩阵,λ 是正则化参数。

3 算法模型优化

3.1 模型建立

MRSR 算法的相似矩阵由原始数据推导出,并在后

续过程中保持不变。然而,真实数据存在的异常值会造成相似矩阵不可靠,进一步影响低维流形数据 $\boldsymbol{XW}$ 的优化,最终导致特征选择效果不佳。

针对 MRSR 算法相似矩阵构造不准确的问题,将相似矩阵的学习嵌入到 MRSR 算法优化过程中,通过对相似矩阵施加概率最近邻约束,相当于以最近邻样本点相连接的概率衡量样本间的相似性,用于保持数据样本的局部几何流形结构,提出 AMRSR 算法,其模型如式(6)所示:

$$\min_{\boldsymbol{W},\boldsymbol{P},\gamma} \|\boldsymbol{X}-\boldsymbol{XW}\|_{2,1}+\lambda_1\|\boldsymbol{W}\|_{2,1}+\lambda(\mathrm{Tr}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{LXW})+\gamma\|\boldsymbol{P}\|_F^2)$$
$$\text{s.t.}\ \sum_{j=1}^{n}p_{ij}=1,p_{ij}\geqslant 0 \tag{6}$$

其中,λ 和 λ_1 是正则化参数,$\boldsymbol{P}\in R^{n\times n}$ 表示样本的相似矩阵,其元素 p_{ij} 表示样本点 $\boldsymbol{x}_i,\boldsymbol{x}_j$ 之间的相似度;根据相似矩阵 $\boldsymbol{P}$ 可以表示出它的度矩阵 $\boldsymbol{D}$,它的第 i 个对角元素为 $\sum_j p_{ij}=1$;拉普拉斯矩阵 $\boldsymbol{L}=\boldsymbol{D}-\boldsymbol{P}=\boldsymbol{I}-\boldsymbol{P}$;$\gamma$ 是一个系数。

AMRSR 算法不仅让重构空间中越接近的样本点以越大的概率相连接,以获得数据样本之间更清晰的相似性结构,而且通过自适应地学习样本的局部相似性结构进一步得到一个更加准确的拉普拉斯矩阵,进而在模型训练过程中保持更加精准的局部结构。系数 γ 可在优化过程中自适应生成,将在下一节详细介绍。

3.2　模型求解

采用交替迭代法对 $\boldsymbol{W}$,$\boldsymbol{P}$ 和 γ 三个变量求解,求解过程具体如下:

(1) 固定 $\boldsymbol{P}$ 和 γ,求解 $\boldsymbol{W}$。当 $\boldsymbol{P}$ 和 γ 固定的时候,变量 $\boldsymbol{W}$ 的解如式所示:

$$\boldsymbol{W}^*=\arg\min_{\boldsymbol{W}}\|\boldsymbol{X}-\boldsymbol{XW}\|_{2,1}+\lambda_1\|\boldsymbol{W}\|_{2,1}+\lambda\mathrm{Tr}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{LXW}) \tag{7}$$

上式采用迭代加权最小二乘算法来求解。给定当前估计值为 $\boldsymbol{W}^t$,定义两个对角权重矩阵 $\boldsymbol{G}_l^t$ 和 $\boldsymbol{G}_r^t$,它们的对角元素分别是 $g_{l,i}^t=\dfrac{1}{2\|\boldsymbol{x}_i-\boldsymbol{x}_i\boldsymbol{W}^t\|_2}$ 和 $g_{r,j}^t=\dfrac{1}{2\|\boldsymbol{w}_j^t\|_2}$,然后通过式来更新 $\boldsymbol{W}^{t+1}$:

$$\boldsymbol{W}^{t+1}=\arg\min_{\boldsymbol{W}}\boldsymbol{Q}(\boldsymbol{W}|\boldsymbol{W}^t)=\arg\min_{\boldsymbol{W}}\mathrm{Tr}((\boldsymbol{X}-\boldsymbol{XW})^{\mathrm{T}}\boldsymbol{G}_l^t(\boldsymbol{X}-\boldsymbol{XW}))+\lambda\mathrm{Tr}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{L}^t\boldsymbol{XW})+\lambda_1\mathrm{Tr}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{G}_r^t\boldsymbol{W}) \tag{8}$$

则 $\boldsymbol{W}^{t+1}$ 的闭式解为

$$\boldsymbol{W}^{t+1}=(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{G}_l^t\boldsymbol{X}+\lambda\boldsymbol{X}^{\mathrm{T}}\boldsymbol{L}^t\boldsymbol{X}+\lambda_1\boldsymbol{G}_r^t)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{G}_l^t\boldsymbol{X} \tag{9}$$

(2) 固定 $\boldsymbol{W}$ 和 γ,求解 $\boldsymbol{P}$。当 $\boldsymbol{W}$ 和 γ 固定的时候,根据拉普拉斯矩阵的性质,如式(10)所示:

$$\|\boldsymbol{f}_i-\boldsymbol{f}_j\|_2^2p_{ij}=2\mathrm{Tr}(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{LF}) \tag{10}$$

其中,$\boldsymbol{F}\in R^{n\times m}$ 并且 $\boldsymbol{f}$ 是矩阵 $\boldsymbol{F}$ 的列向量。那么,变量 $\boldsymbol{P}$ 的解可以写为

$$\boldsymbol{P}^*=\arg\min_{\boldsymbol{P}}\mathrm{Tr}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{LXW})+\gamma\|P\|_F^2=\arg\min_{\boldsymbol{P}}\sum_{i,j=1}^{n}(\|\boldsymbol{W}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{W}^{\mathrm{T}}\boldsymbol{x}_j\|_2^2p_{ij}+\gamma p_{ij}^2)$$
$$\text{s.t.}\ \sum_{j=1}^{n}p_{ij}=1,p_{ij}\geqslant 0 \tag{11}$$

定义 $d_{ij}=\|W^{\mathrm{T}}x_i-W^Tx_j\|_2^2$,注意到上述问题独立于不同的 j,所以把式(11)转换为向量形式

$$\min_{\boldsymbol{p}_i^{\mathrm{T}}\boldsymbol{1}=1,\boldsymbol{p}_i\geqslant 0}\left\|\boldsymbol{p}_i+\frac{1}{2\gamma}\boldsymbol{d}_i\right\|_2^2 \tag{12}$$

上式的拉格朗日函数可以表示为

$$L(\boldsymbol{p}_i,\eta,\boldsymbol{\beta}_i)=\frac{1}{2}\left\|\boldsymbol{p}_i+\frac{1}{2\gamma}\boldsymbol{d}_i\right\|_2^2-\eta(\boldsymbol{p}_i^{\mathrm{T}}\boldsymbol{1}-1)-\boldsymbol{\beta}_i^{\mathrm{T}}\boldsymbol{p}_i \tag{13}$$

其中,$\eta\geqslant 0$ 和 $\boldsymbol{\beta}_i\geqslant 0$ 是拉格朗日乘子。根据 Karush-Kuhn-Tucker 条件和互补松弛条件,最优解 p_{ij} 如式所示:

$$p_{ij}=\left(-\frac{d_{ij}}{2\gamma_i}+\eta\right)_+ \tag{14}$$

其中,$(a)_+=\begin{cases}a,a\geqslant 0\\0,a<0\end{cases}$,任意 $a\in R$。

(3) 固定 $\boldsymbol{W}$ 和 $\boldsymbol{P}$,求解 γ。由于在无监督特征选择算法中,保持数据的局部几何流形结构比保持全局结构效果更好,所以只考虑 k 个近邻点来构造相似矩阵[14]。γ 的最优解可以表示为所有 γ_i 的平均值。不失一般性,假设 $d_{i,1}\leqslant d_{i,2}\leqslant\cdots\leqslant d_{i,n}$,因为 $\boldsymbol{p}_i$ 满足 $p_{i,k}>0\geqslant p_{i,k+1}$,所以有

$$\begin{cases}p_{i,k}>0\Rightarrow-\dfrac{d_{i,k}}{2\gamma_i}+\eta>0\\p_{i,k+1}\leqslant 0\Rightarrow-\dfrac{d_{i,k+1}}{2\gamma_i}+\eta\leqslant 0\end{cases} \tag{15}$$

又根据式(14)和约束条件 $\boldsymbol{p}_i^{\mathrm{T}}\boldsymbol{1}=1$,有

$$\sum_{j=1}^{k}\left(-\frac{d_{ij}}{2\gamma_i}+\eta\right)=1\Rightarrow\eta=\frac{1}{k}+\frac{1}{2k\gamma_i}\sum_{j=1}^{k}d_{ij} \tag{16}$$

将式(16)代入式(15),有

$$\frac{k}{2}d_{i,k}-\frac{1}{2}\sum_{j=1}^{k}d_{ij}<\gamma_i\leqslant\frac{k}{2}d_{i,k+1}-\frac{1}{2}\sum_{j=1}^{k}d_{ij} \quad (17)$$

最终将 γ 的值设置为所有 γ_i 的均值,给定当前估计值为 $\boldsymbol{W}^t$,则

$$\gamma^{t+1}=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{k}{2}d^t_{i,k+1}-\frac{1}{2}\sum_{j=1}^{k}d^t_{ij}\right) \quad (18)$$

将式(18)代入式(14),则

$$p^{t+1}_{ij}=\left(\frac{d^t_{i,k+1}-d^t_{ij}}{kd^t_{i,k+1}-\sum_{j=1}^{k}d^t_{ij}}\right) \quad (19)$$

AMRSR 算法的详细描述如下:

AMRSR 算法

输入:原始数据 $\boldsymbol{X}\in R^{n\times m}$,参数 λ,λ_1,选择的特征数 m_1,选择的近邻数 k

初始化:$t=0$,$\boldsymbol{G}_l^t$,$\boldsymbol{G}_r^t$,$\boldsymbol{L}^t$,γ^t

重复以下步骤:

(1)更新 $\boldsymbol{W}^{t+1}=(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{G}_l^t\boldsymbol{X}+\lambda\boldsymbol{X}^{\mathrm{T}}\boldsymbol{L}^t\boldsymbol{X}+\lambda_1\boldsymbol{G}_r^t)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{G}_l^t\boldsymbol{X}$

(2)更新 $\boldsymbol{G}_l^{t+1}$,$\boldsymbol{G}_r^{t+1}$

(3)按式(19)更新 $\boldsymbol{L}^{t+1}=\boldsymbol{D}^{t+1}-\boldsymbol{P}^{t+1}=\boldsymbol{I}-\boldsymbol{P}^{t+1}$

(4)按式(18)更新 γ^{t+1}

直到收敛

(5)计算特征权重 $\|\boldsymbol{w}_i\|_2(i=1,2,\cdots,m)$ 并且按降序排列

输出:前 m_1 个特征作为特征选择的结果

4 实验验证与分析

为了验证 AMRSR 算法的效果是否有提升,将其与其他几个经典算法在四个数据集上进行对比实验。通过将特征选择结果应用于 K-means 聚类对各算法的聚类结果(聚类精确度和归一化互信息)进行比较,所有算法的精确度和归一化互信息结果以及 AMRSR 算法的参数敏感性用图表展示。

4.1 数据集描述

实验共有 4 个数据集,包括 control、JAFFE、ATT40 和 Yale。表 1 中详细介绍了这些数据集。

表 1 4 个公开数据集

Table 1 Four public datasets

数据集	样本数	特征数	类别数
control	600	60	6
JAFFE	213	256	10
ATT40	400	1024	40
Yale	165	1024	15

4.2 对比算法

MRSR 算法:通过使用 $L_{2,1}$ 范数约束表示残差矩阵和系数矩阵,并且对重构样本进行流形正则化以保持重构空间中原始空间的样本相似性,以此来选择最具代表性且保持局部结构的特征子集[3]。

EGCFS 算法:将相似矩阵的学习嵌入到优化过程中,使图的结构自适应,并且将最大化类间散度矩阵的思想集成到一个统一的框架中,以保持优异的结构,从而选择出更有效的特征子集[11]。

RNE 算法:该算法基于 L_2 范数作为损失项,首先通过局部线性嵌入算法获得特征权值矩阵,再用 L_1 范数描述重构误差最小化,通过交替迭代乘子法来获取每个特征的权重,由此来得到最优的特征子集[15]。

LS 算法:该算法用拉普拉斯分来评价一个特征的重要性,再根据分数选取最优特征构成特征子集,是一种经典的过滤式方法[16]。

MCFS 算法:考虑不同特征之间可能存在的相关性,该算法通过聚类分析中的谱嵌入方法来尽量更好地保留数据的多簇结构。通过 L_1 范数正则化来稀疏系数,从而能有效地进行特征选择[17]。

JELSR 算法:通过局部线性逼近利用权值构造图,将嵌入学习和稀疏回归结合起来进行特征选择。通过添加 $L_{2,1}$ 范数正则化,学习一个行稀疏系数矩阵来进行特征排序[18]。

K-means 算法:算法按照样本之间的距离大小来评价其相似度,即两个样本的距离越近,其相似度就越大,更有可能在一个类簇。在此过程中,不断更新聚类中心,最终把迭代求解出的紧凑且独立的簇作为聚类结果[19]。

4.3 参数设置

对于两个正则化参数 λ 和 λ_1,采用网格搜索法在 $\{10^{-2},10^{-1},1,10,10^2\}$ 范围内设置,并记录所有算法的最佳结果。由于 K-means 聚类算法的初始类簇中心是随机生成的,所以将其执行 10 次,并且记录平均值,而其中的参数 K 为使用数据集的真实类别数[3]。control 数据集所选特征数为 $\{16,20,\cdots,48\}$,JAFFE、ATT40 数据集所选特征数为 $\{20,40,\cdots,180\}$,Yale 数据集所选特征数为 $\{20,30,\cdots,100\}$。并且使用选择所有特征并执行 K-means 聚类作为基线。

4.4 评价指标

对特征选择结果用于 K-means 聚类,对聚类结果采用两种评价指标,即精确度(Accuracy,f_{ACC})和归一化互信息(Normalized Mutual Information,f_{NMI}),它们的值越大,说明算法效果越好。

f_{ACC} 的定义如下:

$$f_{\mathrm{ACC}}=\frac{\sum_{i=1}^{N}\delta(y_i,c_i)}{N} \quad (20)$$

其中,N 是数据集的样本总数,y_i 和 c_i 分别是样本点 x_i

对应的真实类别标签和预测类别标签,$\delta(y_i,c_i)$ 是一个示性函数,如果 $y_i=c_i$,则等于 1,如果不相等,则等于 0[20]。f_{ACC} 越高,说明特征选择效果越好。

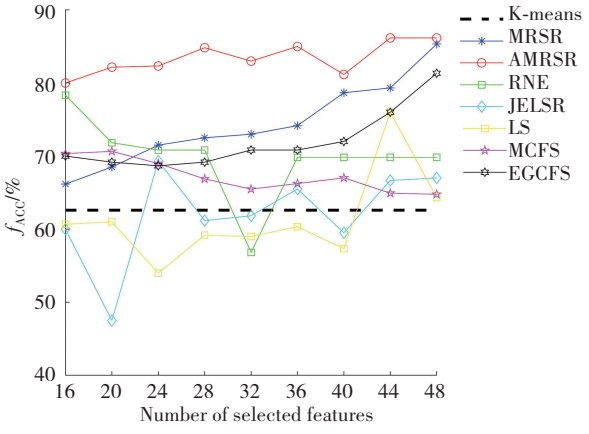
给定两个变量 P 和 Q,f_{NMI} 的定义如下:

$$f_{\mathrm{NMI}}=\frac{I(P,Q)}{\sqrt{H(P)H(Q)}} \tag{21}$$

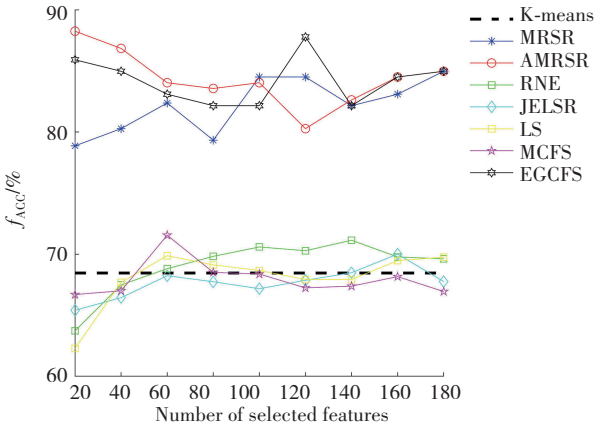
其中,$H(P)$ 和 $H(Q)$ 分别代表 P 和 Q 的信息熵,表示的是 P 和 Q 之间的互信息,则 f_{NMI} 反映了这两个变量之间的相近程度[20]。对于本文算法,P 和 Q 分别是聚类结果即伪标签和数据集的真实标签,f_{NMI} 越高,说明特征选择效果越好。

4.5　实验结果分析

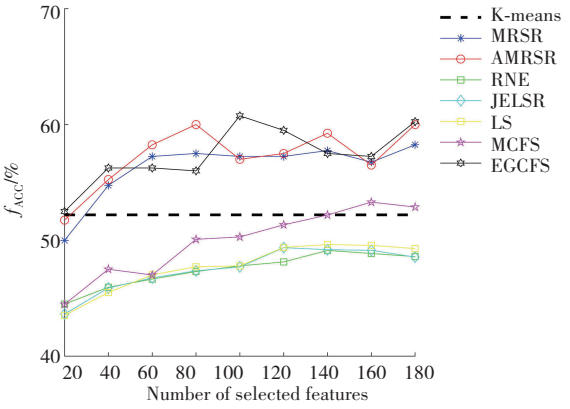
图 1 和图 2 分别展示了表 1 数据集上 AMRSR 算法和其余六种算法选择不同特征数时对应的最佳评价指标结果。图中横坐标表示的是特征子集的维数,纵坐标代表的是各算法的聚类精确度和归一化互信息。从图 1 和图 2 中观察到,当特征数目增加时,算法的聚类效果大多都是呈上升趋势的;并且在大多数情况下,AMRSR 算法性能是有提升的,因为自适应流形正则化方法使得相似矩阵能够更好地反映重构空间数据结构,从而达到更好的性能。



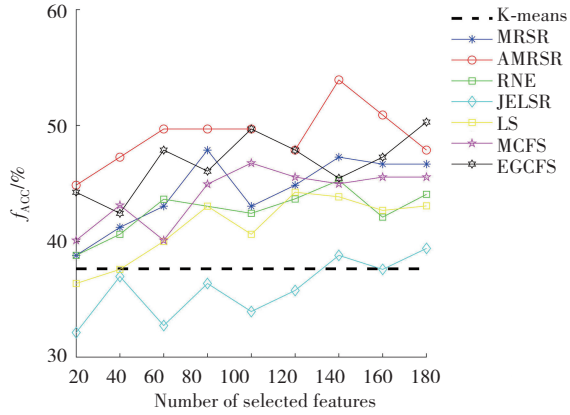
(a) Control



(b) JAFFE



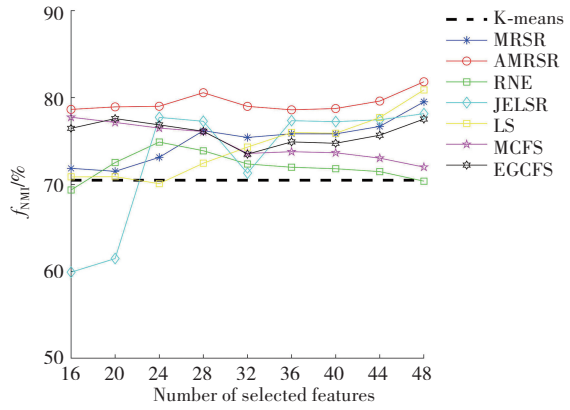
(c) ATT40



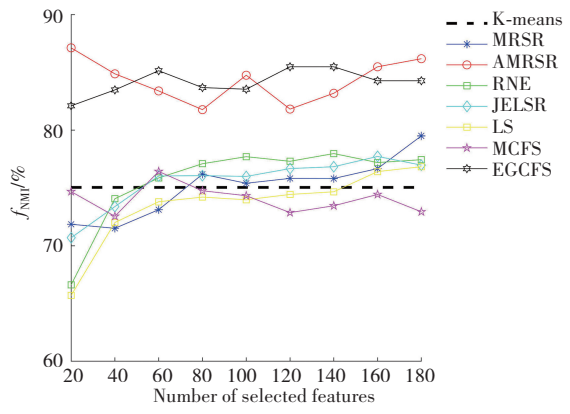
(d) Yale

图 1　在各数据集上选定不同数量特征时各算法的性能(f_{ACC})

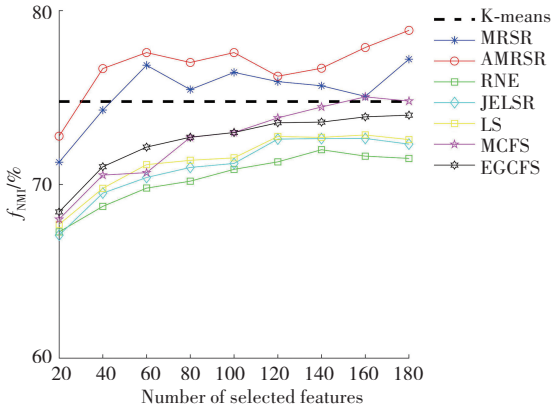
Fig. 1　f_{ACC} of each algorithm with different number of selected features on all datasets



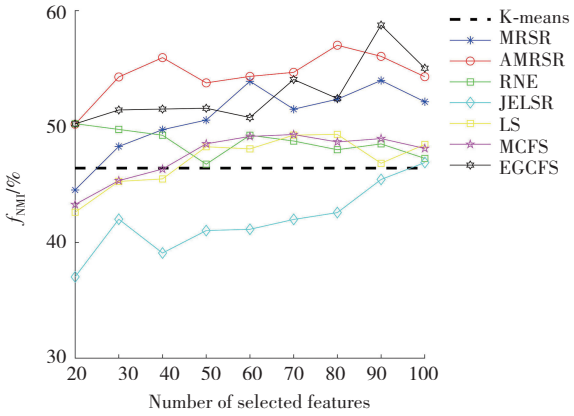
(a) Control



(b) JAFFE



(c) ATT40



(d) Yale

图 2 在各数据集上选定不同数量特征时各算法的性能(NMI)

Fig. 2 NMI of each algorithm with different number of selected features on all datasets

表 2 和表 3 分别记录了各算法在数据集上对应的最佳评价结果,其中最优结果和次优结果分别由加粗项和点式下划线突出表示。与用原始数据集(所有特征)作 K-means 聚类结果(表格中的最后一行)相比,这些无监督特征选择算法不仅缓解了维数灾难,而且降低了数据处理的复杂度,更进一步提高了聚类性能。值得注意的是,该算法相比于 MRSR 算法,在聚类精确度上大致有 1%~6%的提升,在归一化互信息上大致有 2%~9%的提升。

表 2 各算法在每个数据集上的最佳精确度

Table 2 The best accuracy of each algorithm on each dataset

算 法	control	JAFFE	ATT40	Yale_
AMRSR	**0.861 7**	**0.882 6**	0.600 0	**0.539 4**
MRSR	0.853 3	0.849 8	0.582 5	0.478 8
EGCFS	0.813 3	0.877 9	**0.607 5**	0.503 0
LS	0.760 0	0.698 7	0.496 4	0.442 4
RNE	0.783 3	0.711 4	0.491 3	0.452 7
MCFS	0.706 5	0.715 5	0.533 0	0.467 6
JELSR	0.693 3	0.700 0	0.493 7	0.393 9
K-means	0.623 6	0.690 2	0.521 1	0.382 2

表 3 各算法在每个数据集上的最佳归一化互信息

Table 3 The best of normalized mutual information of each algorithm on each dataset

算 法	control	JAFFE	ATT40	Yale
AMRSR	**0.818 2**	**0.871 3**	**0.788 7**	0.570 2
MRSR	0.795 1	0.795 1	0.768 6	0.539 8
EGCFS	0.775 8	0.854 9	0.739 9	**0.587 6**
LS	0.808 8	0.768 4	0.728 5	0.493 2
RNE	0.748 8	0.779 8	0.720 0	0.502 5
MCFS	0.777 4	0.764 3	0.750 4	0.493 1
JELSR	0.781 5	0.777 3	0.726 5	0.468 9
K-means	0.700 2	0.741 6	0.745 7	0.466 0

另外,在 ATT40 数据集上随机选取两个不同类别的样本来测试 AMRSR 算法是否能找到相关的有代表性且有区别性的特征,在它们的特征集合中分别选出 20、40、…、180 个特征,然后执行此算法。如图 3 所示,第一幅图是原图,然后依次为选择 20、40、…、180 个特征对应的人脸图,AMRSR 算法选择的特征设置为白色。从图中可以看出,AMRSR 算法倾向于捕捉人脸上有判别性的部位,如额头、眼睛、鼻子、嘴和脸颊等,这也进一步说明了 AMRSR 算法选出了更具代表性且利于分类的特征子集。

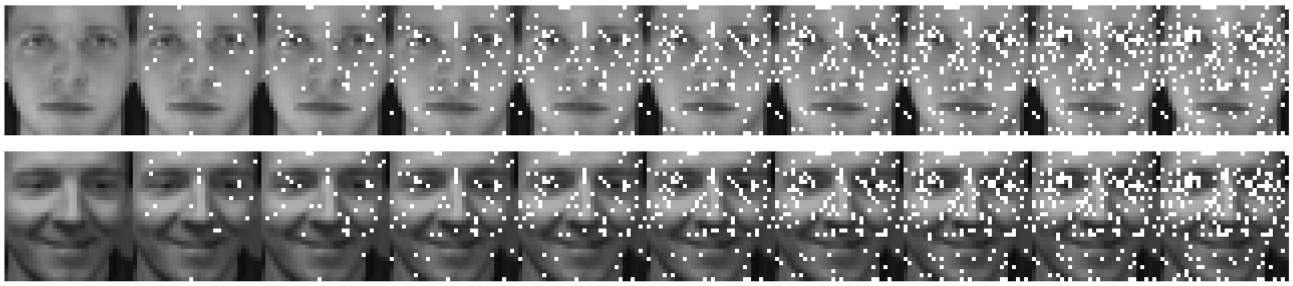
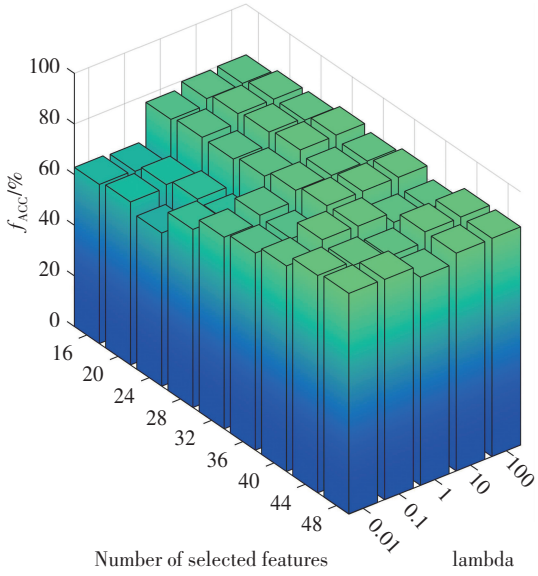


图 3 AMRSR 算法在 ATT40 数据集上的两个效果示例图

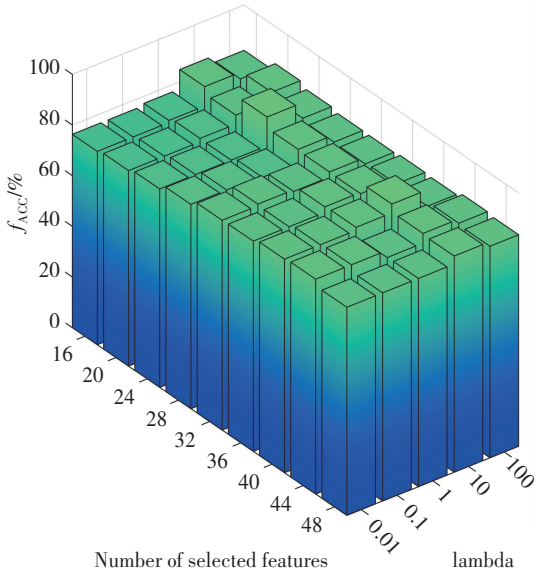
Fig. 3 Two sample graphs of effects of AMRSR algorithm on ATT40 dataset

4.6　参数敏感性分析

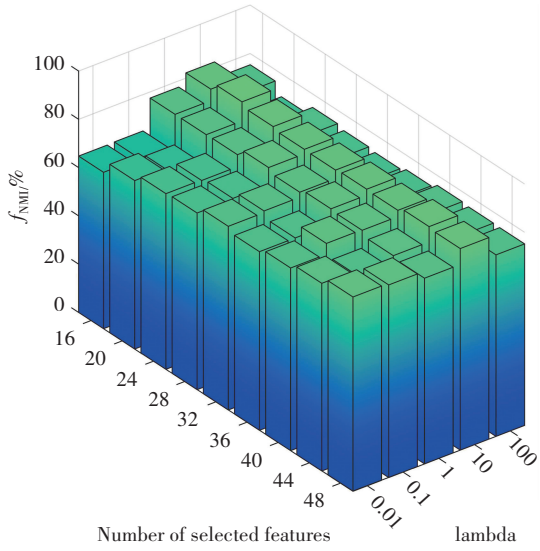
对于 AMRSR 算法,在实验中需要调整两个正则化参数 λ 和 λ_1。以 control 数据集为例,分别讨论了不同参数对结果的影响。当其中一个参数在设置的范围内变化的时候,另一个参数设置为 1。图 4 展示了在该数据集上选择不同特征数时不同参数值对 AMRSR 算法的影响,其中横轴表示选择的特征数,纵轴表示参数的值。从实验结果可以看出,在 control 数据集上 AMRSR 算法对参数 λ 更敏感:参数 λ 值越大,结构学习的精度越高,选择的特征越能保持数据样本之间的结构,而参数 λ_1 对该方法的聚类效果影响不大。



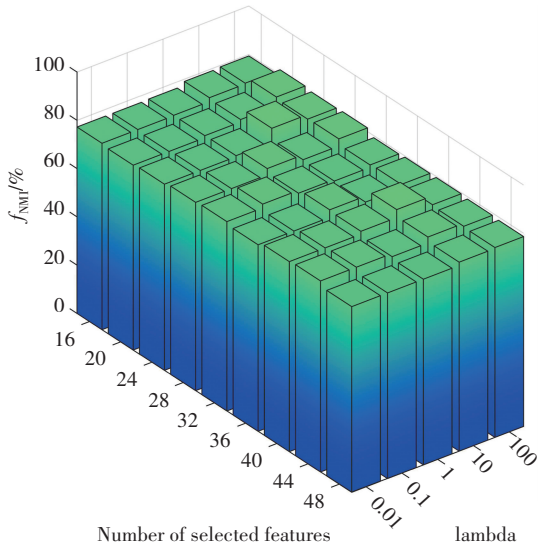
(a) 参数 λ 对聚类精确度的影响



(b) 参数 λ_1 对聚类精确度的影响



(c) 参数 λ 对归一化互信息的影响



(d) 参数 λ_1 对归一化互信息的影响

图 4　control 数据集上各参数对聚类精确度和归一化互信息的影响

Fig. 4　Effect of parameters on clustering accuracy and normalized mutual information on control dataset

5　结　论

在基于流形正则化自表示的无监督特征选择算法的基础上,将相似矩阵的学习嵌入到模型训练过程中,在每次迭代中自适应地更新相似矩阵,以便于获取到更好的重构空间数据结构,进一步提高保持数据结构的准确性。重构表示系数矩阵即特征选择矩阵与相似矩阵相互优化,进而选择更具代表性且更能保持数据局部结构的特征子集。通过大量对比实验验证了基于

自适应流形正则化自表示的无监督特征选择算法在聚类效果上优于现有的一些算法,这也进一步说明了该算法能选出具有更高精确度和归一化互信息的代表性特征子集。由于基于流形正则化的无监督特征选择方法关键点在于构造的拉普拉斯矩阵的质量,这直接取决于相似矩阵的构造。能否更加准确构造出数据样本的相似矩阵,是进一步探索的方向。

参考文献(References):

[1] 汪志远. 无监督特征选择方法研究[D]. 太原: 太原理工大学, 2020.
WANG Zhi-yuan. Researches on unsupervised feature selection methods[D]. Taiyuan: Taiyuan University of Technology, 2020.

[2] 周传华, 柳智才, 丁敬安, 等. 基于 filter + wrapper 模式的特征选择算法[J]. 计算机应用研究, 2019, 36(7): 1975—1979+2010.
ZHOU Chuan-hua, LIU Zhi-cai, DING Jin-an, et al. Feature selection algorithm based on filter+wrapper pattern[J]. Computer Application Research, 2019, 36(7): 1975—1979+2010.

[3] LIANG S, XU Q, ZHU P, et al. Unsupervised feature selection by manifold regularized self-representation [C]// 2017 IEEE International Conference on Image Processing. IEEE, 2017: 2398—2402.

[4] 任超宏. 基于特征级图学习的无监督特征选择算法研究[D]. 太原: 山西大学, 2020.
REN Chao-hong. Researches on unsupervised feature selection algorithm based on feature level graph learning[D]. Taiyuan: Shanxi University, 2020.

[5] 刘波, 何希平. 高维数据的特征选择: 理论与算法[M]. 北京: 科学出版社, 2016.
LIU Bo, HE Xi-ping. Feature selection of high-dimensional data: theory and algorithm[M]. Beijing: Science Press, 2016.

[6] ZHU P, ZUO W, ZHANG L, et al. Unsupervised feature selection by regularized self-representation[J]. Pattern Recognition, 2015, 48(2): 438—446.

[7] ZHU P, ZHU W, WANG W, et al. Non-convex regularized self-representation for unsupervised feature selection[J]. Image and Vision Computing, 2017, 60(5): 22—29.

[8] LI Y, LEI C, FANG Y, et al. Unsupervised feature selection by combining subspace learning with feature self-representation[J]. Pattern Recognition Letters, 2018, 109: 35—43.

[9] TANG C, LIU X, LI M, et al. Robust unsupervised feature selection via dual self-representation and manifold regularization[J]. Knowledge-Based Systems, 2018, 145(2): 109—120.

[10] DU L, SHEN Y D. Unsupervised feature selection with adaptive structure learning[C]//Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015: 209—218.

[11] ZHANG R, ZHANG Y, LI X. Unsupervised feature selection via adaptive graph learning and constraint[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 1—8.

[12] BELKIN M, NIYOGI P, Sindhwani V. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples[J]. Journal of Machine Learning Research, 2006, 7(11): 2400—2430.

[13] 徐彬. 基于图的无监督特征选择算法研究[D]. 上海: 华东交通大学, 2021.
XU Bin. Research on unsupervised feature selection algorithm based on graph[D]. Shanghai: East China Jiaotong University, 2021.

[14] LIU X, WANG L, ZHANG J, et al. Global and local structure preservation for feature selection [J]. IEEE Transactions on Neural Networks and Learning Systems. IEEE, 2013, 25(6): 1083—1095.

[15] LIU Y, YE D, LI W, et al. Robust neighborhood embedding for unsupervised feature selection[J]. Knowledge-Based Systems, 2020, 193(6): 105462.

[16] HE X, CAI D, NIYOGI P. Laplacian score for feature selection[C]//Advances in Neural Information Processing Systems. IEEE, 2006: 507—514.

[17] CAI D, ZHANG C, HE X. Unsupervised feature selection for multi-cluster data[C]//Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2010: 333—342.

[18] HOU C, NIE F, YI D. Feature selection via joint embedding learning and sparse regression[C]//Twenty-Second International Joint Conference on Artificial Intelligence. DBLP, 2011, 22: 1324—1329.

[19] 丛思安, 王星星. K-means 算法研究综述[J]. 电子技术与软件工程, 2018(17): 155—156.
CONG Si-an, WANG Xing-xing. Review of K-means algorithm research [J]. Electronic Technology and Software Engineering, 2018(17): 155—156.

[20] NIE F, WANG X, HUANG H. Clustering and projected clustering with adaptive neighbors[M]. NewYork: ACM, 2014.

责任编辑:陈 芳