

联合多流形结构和自表示的特征选择方法



易玉根¹ 李世成¹ 裴洋¹ 陈磊¹ 代江艳²

¹ 江西师范大学软件学院 南昌 330022

² 潍坊学院计算机工程学院 山东 潍坊 261061

(yiyg510@jxnu.edu.cn)

摘要 特征选择是一种通过去除不相关和冗余的特征来降低数据维度和提高后续学习算法效率的数据处理方法。无监督特征选择已经成为维数约简中具有挑战性的问题之一。首先,通过结合特征自表示能力和流形结构,提出了一种联合多流形结构和自表示(Joint Multi-Manifold Structures and Self-Representation, JMSSR)的无监督特征选择方法。不同于现有的方法,为了更准确地刻画特征的流形结构,引入一种自适应加权策略来融合特征的多个流形结构。然后,提出了一种简单且有效的迭代优化算法来求解 JMSSR 方法的目标函数,并利用数值实验验证了优化算法的收敛性。最后,分别在 JAFFE, ORL 和 COIL20 3 个数据集上进行聚类实验,实验结果验证了与现有的无监督特征选择方法相比, JMSSR 方法具有较好的性能。

关键词: 特征选择; 自适应加权; 多流形结构; 自表示

中图法分类号 TP391

Feature Selection Method Combined with Multi-manifold Structures and Self-representation

YI Yu-gen¹, LI Shi-cheng¹, PEI Yang¹, CHEN Lei¹ and DAI Jiang-yan²

¹ School of Software, Jiangxi Normal University, Nanchang 330022, China

² School of Computer Engineering, Weifang University, Weifang, Shandong 261061, China

Abstract Feature selection is to reduce the dimension of data by removing irrelevant and redundant features and improve the efficiency of learning algorithm. Unsupervised feature selection has become one of the challenging problems in dimensionality reduction. Firstly, combining self-representation and manifold structure of features, a Joint Multi-Manifold Structures and Self-Representation (JMSSR) unsupervised feature selection algorithm is proposed. Different from the existing approaches, our approach designs an adaptive weighted strategy to integrate multi-manifold structures to describe the structure of features accurately. Then, a simple and effective iterative updating algorithm is proposed to solve the objective function, and the convergence of the optimization algorithm is also verified by numerical experiments. Finally, experimental results on three datasets (such as JAFFE, ORL and COIL20) show that the proposed approach exhibits better performance than the existing unsupervised feature selection approaches.

Keywords Feature selection, Adaptive weighted, Multi-manifold structures, Self-representation

1 引言

随着信息化和存储技术的飞速发展,人们所采集的数据在样本数量和特征维度上呈现急剧膨胀趋势^[1-3]。高维数据经常包含大量无关或冗余的特征,这无疑增加了数据处理方法的时间复杂度和空间复杂度。此外,不相关和冗余特征也会对后续的聚类或分类任务带来一定的影响。特征选择已经成为解决上述问题的典型方法之一,并受到了研究者的广泛关注^[1-5]。

特征选择方法根据算法是否使用类别标签信息,可大概

分为有监督、半监督和无监督 3 类方法^[1]。有监督特征选择方法在类别信息的引导下选择最具鉴别能力的特征子集,在分类识别任务中获得了优越的性能^[1]。然而,对于某些无法获取数据标签信息的训练数据,有监督方法将无法工作即失效。与有监督特征选择方法相比,无监督特征选择方法通过挖掘高维数据的内在性质或结构信息来选取特征子集,使其更加灵活且实用。因此,设计有效的无监督特征选择方法在数据处理中是非常重要的。

Li 等^[6]介绍了许多无监督特征选择方法。其中,方差得分(Variance Score, VS)是一种经典且简单的无监督特征选择

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:国家自然科学基金(61602221, 61806126);江西省自然科学基金(20171BAB212009);江西省教育厅科技项目(GJJ160315, GJJ170234);山东省高等学校青创科技支持计划(2019KJN012)

This work was supported by the National Natural Science Foundation of China (61602221, 61806126), Natural Science Foundation of Jiangxi Province (20171BAB212009), Science and Technology Research Project of Jiangxi Provincial Department of Education (GJJ160315, GJJ170234) and Youth Innovation Science and Technology Support Plan of Shandong Provincial Institution of Higher Education (2019KJN012).

通信作者:代江艳(daijyan@163.com)

方法^[7]。VS方法首先计算每个特征的方差,然后选择方差较大的特征作为最优特征子集。He等^[8]利用特征的局部保留能力,提出了一种无监督特征选择方法,称为拉普拉斯得分(Laplacian Score,LS)。LS方法可以选择最能保持原始数据流形结构的特征子集。Zhao等^[9]则将谱图理论引入到特征选择中,提出了一种谱特征选择(SPEC)方法。SPEC方法首先基于高维数据构造相似图,然后根据谱图理论选择能够保留其图结构信息的特征子集。然而,VS,LS和SPEC方法都是单独对每个特征进行评估,因此,它们忽略了特征之间的相关性,从而导致所选择的特征子集并不是最优的^[10-11]。为了解决上述问题,Cai等^[10]结合流形结构和 l_1 -范数正则化提出了一种统一框架来选择最优特征子集。之后,Li等^[11]利用判别信息和特征间的相关性来增加特征子集的判别能力,提出一种非负判别特征选择(Non-negative Discriminative Feature Selection,NDFS)方法。此外,为了更好地处理高维数据所包含大量噪声或异常值数据,提高方法的性能,研究学者提出了一系列基于 $l_{2,1}$ -范数的无监督特征选择方法^[12-14]。比如,Nie等^[12]同时对距离度量损失函数和特征选择矩阵加以 $l_{2,1}$ -范数来降低特征算法对异常数据点和噪声点的影响。Qian等^[13]将基于 $l_{2,1}$ -范数的回归模型和非负矩阵分解集成到统一框架中,并提出了一种鲁棒的无监督特征选择(Robust Un-supervised Feature Selection,RUFS)方法。不同于文献^[12-13],Zhu等^[14]提出了一种正则化自表示(Regularized Self-Representation,RSR)方法用于无监督特征选择。RSR方法假设每个特征都可以由其他相关特征线性表示。该方法首先建立特征的回归模型,并分别对损失函数和回归系数加以 $l_{2,1}$ -范数约束,从而选取最具代表性的特征。RSR方法仅仅考虑特征的自表示能力,却忽略了特征的结构保持能力。对此,Liang等^[15]提出了一种流形正则化自表示(Manifold Regularized Self-Representation,MRSR)方法,该方法采用传统的 k -近邻图(KNN图)刻画特征流形结构。此外,为了提高算法对噪声数据的鲁棒性,Tang等^[16]将基于 l_1 -范数的图拉普拉斯正则化引入自表示模型中,以确保特征选择模型更加稳健。尽管上述方法的性能要优于其他方法,但它们的性能在一定程度上依赖于图的构建。

目前,研究学者提出大量的图构建方法^[17],常见方法包括KNN图^[18]和LLE(Locally Linear Embedding)图^[19],然而这两种方法的近邻参数选择与权值计算是相互独立的过程。为了解决这个问题,基于稀疏表示的 l_1 -图^[20]和基于低秩表示(Low-Rank Representation,LRR)的LRR图^[21]被相继提出,但这两种方法具有较高的计算复杂度。为了提高算法的计算效率,Lu等^[22]基于最小二乘方法构建了 l_2 -图,该方法简单且高效,并被广泛应用于基于图的维数约简和谱聚类方法中。然而,具体选择哪种图构建方法仍然是一个开放性问题。

综上所述,本文提出了一种联合多流形结构和自表示的特征选择方法。为了充分利用特征的流形结构,JMMSSR方法引入了一种自适应加权融合策略。与现有的无监督特征选择方法相比,该方法具有如下优点:首先,考虑了特征不同的流形结构;其次,引入自适应加权策略对其进行融合,使之更为有效地刻画特征的本征结构;最后,设计了一种简单且高效的迭代更新算法求解目标函数。通过在3个数据集上进行的大量实验,验证了JMMSSR方法的有效性。

2 联合多流形结构和自表示的特征选择方法

假设样本数据矩阵 $\mathbf{X}=[\mathbf{x}_1;\mathbf{x}_2;\dots;\mathbf{x}_n]\in\mathbb{R}^{n\times d}$,其中, n 为样本的数量,每个样本表示为 d 维的特征向量。矩阵 \mathbf{X} 中的每一行表示一个样本,每一列表示一个特征。对于任意矩阵 $\mathbf{M}\in\mathbb{R}^{m\times n}$, m_{ij} 表示矩阵 \mathbf{M} 中 (i,j) 处的元素, m_i 和 m^j 分别表示矩阵 \mathbf{M} 的第 i 行和第 j 列。 $\text{tr}(\mathbf{M})$ 表示矩阵的迹, \mathbf{M}^T 表示矩阵 \mathbf{M} 的转置。 \mathbf{I}_m 表示大小为 $m\times m$ 的单位矩阵。矩阵 \mathbf{M} 的 $l_{2,1}$ -范数定义为 $\|\mathbf{M}\|_{2,1}=\sum_{i=1}^m\|\mathbf{m}_i\|_2=\sum_{i=1}^m\sqrt{\sum_{j=1}^n m_{ij}^2}$,矩阵 \mathbf{M} 的 F -范数定义为 $\|\mathbf{M}\|_F=\sqrt{\sum_{i=1}^m\sum_{j=1}^n m_{ij}^2}$ 。

2.1 目标函数

特征自表示被广泛应用于特征选择中,它可以表示为如下模型:

$$\min_{\mathbf{W}}\|\mathbf{X}-\mathbf{XW}\|_{2,1}+\lambda\|\mathbf{W}\|_{2,1} \quad (1)$$

其中,矩阵 \mathbf{W} 表示特征的权值矩阵。为了提高特征选择方法对数据离群点和噪声点的鲁棒性,式(1)中采用基于 $l_{2,1}$ -范数的度量准则计算表示误差,同时通过对权值矩阵 \mathbf{W} 施加稀疏约束实现特征子集选择的目的。尽管基于自表示的方法可以选择最具有代表性的特征,但它忽略了特征的局部流形结构信息。然而,有效保留特征的局部流形结构在特征选择过程中极为重要,因此Zhu等^[14]提出流形正则化自表示模型,其目标函数定义为:

$$\min_{\mathbf{W}}\{\|\mathbf{X}-\mathbf{XW}\|_{2,1}+\alpha\|\mathbf{W}\|_{2,1}+\beta\text{tr}(\mathbf{W}^T\mathbf{X}^T\mathbf{LXW}) \quad (2)$$

其中, $\mathbf{L}=\mathbf{D}-\mathbf{S}$ 表示图的拉普拉斯矩阵, \mathbf{S} 表示图的权值矩阵, \mathbf{D} 表示对角矩阵,其对角元素为 \mathbf{S} 矩阵中每一行或者每一列元素之和(如 $d_{ii}=\sum_{j=1}^n s_{ij}$), α 和 β 为非负平衡参数。式(2)中第三项为流形正则项(拉普拉斯正则化项),主要用于保持特征的局部流形结构信息。

流形学习是一种非常有效地维数约简方法,目前也有大量方法被提出^[17]。流形学习的核心问题是如何有效地构建图来刻画特征流形的本征结构。为了更好地刻画特征的流形结构,本文采用5种常用的图构建方法^[18-22]。文献^[18]提出的基于欧氏距离构建的KNN图方法包括两个步骤,首先对每个样本选择 k 个近邻样本;然后计算其近邻样本之间的权值,如式(3)所示:

$$s_{ij}=\begin{cases} \exp\left(-\frac{\|\mathbf{x}_i-\mathbf{x}_j\|_2^2}{\sigma^2}\right), & \text{if } \mathbf{x}_j\in\Omega_k(\mathbf{x}_i) \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

其中, $\Omega_k(\mathbf{x}_i)$ 表示样本 \mathbf{x}_i 的 k 近邻样本集合, σ 为热核参数, $\exp(\cdot)$ 表示指数函数。

不同于文献^[18],文献^[19]基于局部线性嵌入(Locally Linear Embedding,LLE)的思想提出了LLE图构建方法,其权值通过优化式(4)进行计算。

$$\begin{aligned} \min & \sum_{i=1}^n\|\mathbf{x}_i-\sum_{\mathbf{x}_j\in\Omega_k(\mathbf{x}_i)}s_{ij}\mathbf{x}_j\|_2^2 \\ \text{s. t.} & \sum_{j=1}^n s_{ij}=1, s_{ij}=0, \text{if } \mathbf{x}_j\notin\Omega_k(\mathbf{x}_i) \end{aligned} \quad (4)$$

其中, $\mathbf{S}=[s_{ij}]\in\mathbb{R}^{n\times n}$ 表示权值矩阵,元素 s_{ij} 表示重构系数。

由于KNN图和LLE图的理论简单且容易实现,因此它们成为了目前流行的图构建方法,但是这两种方法中的近邻选择和权值计算是相互独立的。另外,它们都是基于欧氏距

离的,因此对噪声非常敏感。

受启发于稀疏表示被成功应用于信号处理及人脸识别等众多领域,文献[20]提出一种基于稀疏表示构建 l_1 -图的方法,其邻域选择和权值求解可以通过优化式(5)实现。

$$\begin{aligned} \min \quad & \|s_i\|_1 \\ \text{s. t.} \quad & \mathbf{x}_i = \overline{\mathbf{X}}_i s_i \end{aligned} \quad (5)$$

其中, $\overline{\mathbf{X}}_i = [\mathbf{x}_1; \mathbf{x}_2; \dots; \mathbf{x}_{i-1}; \mathbf{x}_{i+1}; \dots; \mathbf{x}_n] \in \mathbb{R}^{d \times (n-1)}$ 表示不包括样本 \mathbf{x}_i 的样本子集, $\|\cdot\|_1$ 表示向量的 l_1 -范数。从式(5)可知, l_1 -图将邻域选择和权值计算同步完成。但求解 l_1 -范数的最优化问题是非常耗时的。此外,在构建 l_1 -图的过程中需要单独对每个样本求解重构系数,因此,所构建的图并不能很好地刻画特征的局部结构。为此,文献[21]提出了一种基于低秩表示的 LRR 图构建方法,其目标函数定义如式(6)所示:

$$\begin{aligned} \min \quad & \|\mathbf{S}\|_* \\ \text{s. t.} \quad & \mathbf{X} = \mathbf{X}\mathbf{S} \end{aligned} \quad (6)$$

其中, $\|\cdot\|_*$ 表示矩阵的核范数,其值等于矩阵奇异值之和。尽管 LRR 图可以很好地挖掘数据的全局结构,但其求解方法仍然比较耗时。于是, Lu 等人^[22] 基于最小二乘法提出 l_2 -图构建方法,目标函数如式(7)所示:

$$\begin{aligned} \min \quad & \|\mathbf{S}\|_F \\ \text{s. t.} \quad & \mathbf{X} = \mathbf{X}\mathbf{S}, \text{diag}(\mathbf{S}) = 0 \end{aligned} \quad (7)$$

其中, $\text{diag}(\cdot)$ 表示矩阵的对角元素操作。

上述 5 种不同的图构建方法也被广泛应用于基于图的维数约简或聚类方法中,但是不同的图构建方法是基于不同的准则,每种图构建方法可能适用于不同的数据集。因此,如何选取合适的图构建方法仍然是开放性课题。为了更好地处理和刻画数据的流形结构,本文采用融合多个图构建方法的策略。假设根据不同的图构建方法可以得到系列图拉普拉斯矩阵,并标记集合为 $\{\mathbf{L}_1, \mathbf{L}_2, \dots, \mathbf{L}_M\}$, 其中 M 表示拉普拉斯矩阵的个数,本文 M 取值为 5。同时,为了充分利用不同拉普拉斯矩阵的互补性,我们引入权值向量 $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_M]$, 则式(2)可以重新定义为:

$$\begin{aligned} \min_{\mathbf{W}, \boldsymbol{\mu}} \quad & \|\mathbf{X} - \mathbf{X}\mathbf{W}\|_{2,1} + \alpha \|\mathbf{W}\|_{2,1} + \\ & \beta \text{tr}(\mathbf{W}^T \mathbf{X}^T (\sum_{m=1}^M \mu_m \mathbf{L}_m) \mathbf{X}\mathbf{W}) \\ \text{s. t.} \quad & \mu_m = 1 / (2 \sqrt{\text{tr}(\mathbf{W}^T \mathbf{X}^T \mathbf{L}_m \mathbf{X}\mathbf{W})}) \end{aligned} \quad (8)$$

其中,参数 α 和 β 用于平衡各项在目标函数中的贡献。从式(8)中可知,根据约束项可自适应求得不同拉普拉斯矩阵的权值,因此体现了图的自适应性融合。

2.2 优化求解

由于目标式(8)中需要同时对变量 \mathbf{W} 和 $\boldsymbol{\mu}$ 进行优化求解,因此无法给出目标函数的全局最优解。于是,本文提出一种基于迭代更新优化策略的求解方法,即固定其中某一个变量,更新另一个变量,交替迭代更新,直到达到收敛条件为止。

2.2.1 固定 μ_m , 求解 \mathbf{W}

从目标式(8)中移除与 \mathbf{W} 无关的项,则有关矩阵 \mathbf{W} 的优化问题可转化为式(9):

$$\min_{\mathbf{W}} \{ \|\mathbf{X} - \mathbf{X}\mathbf{W}\|_{2,1} + \alpha \|\mathbf{W}\|_{2,1} + \beta \text{tr}(\mathbf{W}^T \mathbf{X}^T \boldsymbol{\Omega} \mathbf{X}\mathbf{W}) \} \quad (9)$$

其中, $\boldsymbol{\Omega} = \sum_{m=1}^M \mu_m \mathbf{L}_m$ 。根据 l_{21} -范数的定义,式(9)可以进一步改写为式(10):

$$\begin{aligned} \min_{\mathbf{W}} \{ & \text{tr}(\mathbf{X} - \mathbf{X}\mathbf{W})^T \mathbf{R} (\mathbf{X} - \mathbf{X}\mathbf{W}) + \alpha \text{tr}(\mathbf{W}^T \mathbf{G}\mathbf{W}) + \\ & \beta \text{tr}(\mathbf{W}^T \mathbf{X}^T \boldsymbol{\Omega} \mathbf{X}\mathbf{W}) \} \end{aligned} \quad (10)$$

其中, \mathbf{R} 和 \mathbf{G} 分别表示对角矩阵,其对角元素分别定义为

$$r_{ii} = \frac{1}{2 \|\mathbf{x}_i - \mathbf{x}_i \mathbf{W}\|_2} \text{ 和 } g_{ii} = \frac{1}{2 \|\mathbf{w}_i\|_2}。$$

对式(10)进行简单运算后,再对矩阵 \mathbf{W} 求偏导,并令其导数为零,则有:

$$\frac{\partial \varphi(\mathbf{W})}{\partial \mathbf{W}} = \{ \mathbf{X}^T \mathbf{R} \mathbf{X} \mathbf{W} - \mathbf{X}^T \mathbf{R} \mathbf{X} + \alpha \mathbf{G} \mathbf{W} + \beta \mathbf{X}^T \boldsymbol{\Omega} \mathbf{X} \mathbf{W} \} = 0 \quad (11)$$

通过式(11),可以得到矩阵 \mathbf{W} 的更新规则如下:

$$\mathbf{W} = (\mathbf{X}^T \mathbf{R} \mathbf{X} + \alpha \mathbf{G} + \beta \mathbf{X}^T \boldsymbol{\Omega} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{R} \mathbf{X}) \quad (12)$$

2.2.2 固定 \mathbf{W} , 求解 $\boldsymbol{\mu}$

为了使每个拉普拉斯矩阵可以自适应分配权值,在每次迭代过程中,都需要对 $\boldsymbol{\mu}$ 更新,其更新规则如式(13)所示:

$$\mu_m = 1 / (2 \sqrt{\text{tr}(\mathbf{W}^T \mathbf{X}^T \mathbf{L}_m \mathbf{X}\mathbf{W})}) \quad (13)$$

从式(13)可知,参数 μ_m 的取值依赖变量 \mathbf{W} ,一旦根据式(12)计算出矩阵 \mathbf{W} 后,参数 μ_m 就可以根据式(13)进行更新。因此,可以通过计算式(12)和式(13),不断更新变量 \mathbf{W} 和 $\boldsymbol{\mu}$,直到目标函数到达收敛为止。

2.3 算法流程

本文特征选择方法的流程如算法 1 所示。

算法 1 JMMSSR

输入: 样本集 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$, 参数 α 和 β , 迭代次数 MaxIter

输出: 前 p 个特征排序索引

- 步骤 1 根据式(4)一(8)求解各个图模型的权值矩阵,并计算相应的拉普拉斯矩阵。
- 步骤 2 通过式(12)更新 \mathbf{W} 矩阵。
- 步骤 3 通过式(13)更新 $\boldsymbol{\mu}$ 矩阵。
- 步骤 4 判断是否达到最大迭代次数,若是,则执行步骤 5,否则循环执行步骤 2—步骤 4。
- 步骤 5 计算矩阵 \mathbf{W} 中所有行的 l_2 范数,并将其值按降序排序,确定特征顺序。

2.4 计算复杂性分析

本小节主要分析算法的计算复杂度。根据算法 1 可知,首先需要构建 5 个图,它们的复杂度如表示 1 所列,其中 k 为近邻数, t 为相应优化算法的迭代次数。其次,需要执行优化算法,其计算复杂度为 $O(T(d^3 + dn^2))$,其中 T 为算法迭代次数。

表 1 不同图构建方法的计算复杂度

Table 1 Computation complexity of different graph

construction methods	
方法	计算复杂度
KNN 图	$O(n^2 d)$
LLE 图	$O(n^2 d + dnk^3)$
l_1 -图	$O(td^2 n^2 + tdn^3)$
LRR 图	$O(tnd^2 + tn^3)$
l_2 -图	$O(n^3)$

3 实验及分析

在实验中,本文采用了 3 个公开的数据库,分别为 JAFFE^[23], ORL^[24] 和 COIL20^[25],并将本文方法与 LS^[8], SPEC^[9], UDFS^[11], RUFSS^[13], RSR^[14], MRSR^[15] 和

L1UFS^[16]方法进行对比。数据库的详细描述如表2所列,数据库中的部分样本图像如图1所示。本实验采用两个聚类准则(准确率和归一化互信息)评价算法的性能^[26]。给定样本 x_i 的聚类结果为 q_i ,而 p_i 表示样本的真实标记,则聚类准确率(Accuracy, ACC)定义为:

$$ACC = \frac{\sum_{i=1}^n \delta(p_i, \text{map}(q_i))}{N} \quad (14)$$

其中,当 $x=y$ 时, $\delta(x,y)=1$;否则 $\delta(x,y)=0$ 。 $\text{map}(\cdot)$ 为真实标签与聚类标签的一个映射函数。给定两个变量 P 和 Q ,归一化互信息(Normalized Mutual Information, NMI)定义为:

$$NMI(P, Q) = \frac{I(P, Q)}{\sqrt{H(P)H(Q)}} \quad (15)$$

其中, $H(P)$ 和 $H(Q)$ 分别表示变量 P 和 Q 的熵。 $I(P, Q)$ 表示两个变量的互信息。NMI的值反映了聚类标签与真实标签的一致性。

表2 数据库的详细信息

Table 2 Details of databases

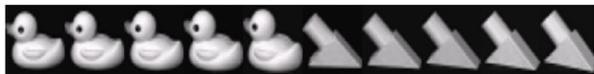
数据库名	图像大小	样本数	类别数
JAFFE	32×32	213	10
ORL	32×32	400	40
COIL20	32×32	1440	20



(a) JAFFE



(b) ORL



(c) COIL20

图1 数据库中部分实例图

Fig.1 Samples from three databases

所有对比方法和本文方法都是采用 Matlab 2012a 编程语言实现,实验平台为 Inter(R) Core(TM) i7-4970 CPU@3.60GHz 和 8GB RAM 的台式机,操作系统为 64 位 Windows10 系统。本文方法融合了 5 种常用的图构建方法,所有方法的参数均采用网格搜索方式寻找其最优值,并利用 k -means 聚类算法对选取的特征进行聚类。由于 k -means 聚类算法的性能很大程度上依赖于初始化,实验中对聚类过程随机执行 50 次,并给出平均结果和标准差。不同方法在 3 个数据库上的实验结果如表 3 和表 4 所列。

表3 不同方法在 3 个数据库上的准确率

Table 3 ACC of different methods on three databases

方法	JAFFE	ORL	COIL20
Baseline	0.7873±0.0228	0.7526±0.0439	0.5527±0.0271
LS	0.8343±0.0630	0.7850±0.0310	0.5984±0.0246
SPEC	0.8521±0.0708	0.8030±0.0756	0.6128±0.0476
UDFS	0.8742±0.0823	0.8340±0.0447	0.6325±0.0289
RUFS	0.8864±0.0781	0.8300±0.0542	0.6408±0.0484
RSR	0.8728±0.0518	0.8310±0.0378	0.6486±0.0272
MRSR	0.8874±0.0456	0.8462±0.0434	0.6520±0.0251
L1UFS	0.8994±0.0609	0.8622±0.0589	0.6657±0.0354
JMMSSR	0.9123±0.0403	0.8745±0.0339	0.6792±0.0238

表4 不同方法在 3 个数据库上的归一化互信息

Table 4 NMI of different methods on three databases

方法	JAFFE	ORL	COIL20
Baseline	0.8213±0.0140	0.7964±0.0310	0.7035±0.0131
LS	0.8756±0.0301	0.8138±0.0455	0.7068±0.0122
SPEC	0.8828±0.0614	0.8318±0.0228	0.7245±0.0128
UDFS	0.8927±0.0357	0.8583±0.0238	0.7244±0.0167
RUFS	0.8914±0.0448	0.8588±0.0316	0.7486±0.0096
RSR	0.8952±0.0276	0.8600±0.0298	0.7463±0.0163
MRSR	0.9009±0.0243	0.8736±0.0225	0.7500±0.0228
L1UFS	0.9074±0.0401	0.8987±0.0254	0.7592±0.0202
JMMSSR	0.9193±0.0237	0.9116±0.0210	0.7684±0.0159

从表 3 和表 4 的实验结果中可得出如下结论:

1) 所有特征选择方法都优于基方法(Baseline),这表明特征选择对提高聚类性能起着重要作用。

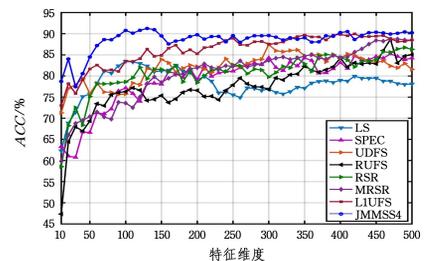
2) LS 方法和 SPEC 方法的性能要低于其他特征选择方法。主要原因是 LS 和 SPEC 方法都是以独立的方式选择特征,忽略了特征之间的相关性。

3) 由于 RSR 方法考虑了特征的自表示能力,因此它的性能要优于 UDFS 和 RUFS 方法的性能。

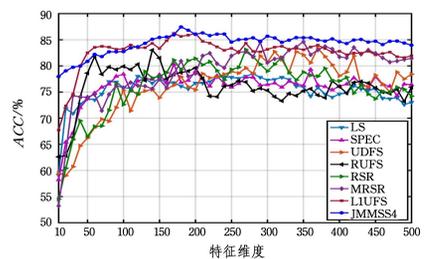
4) 与 RSR 方法不同, MRSR 和 L1UFS 方法在特征选择过程中考虑了数据局部流形结构,它们的性能要优于 UDFS, RUFS 和 RSR 方法。

5) JMMSSR 方法不仅利用了特征的自表示能力,而且挖掘了特征的局部流形结构,通过自适应加权融合多个不同的流形结构,因此其性能要优于其他无监督特征选择方法。

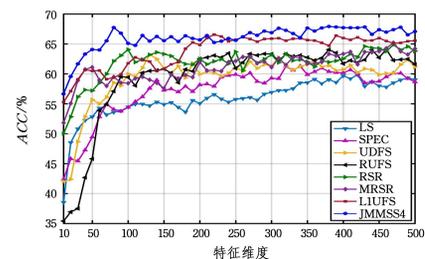
图 2 给出了不同方法的聚类准确率(ACC)与特征维度变化曲线图。



(a) JAFFE



(b) ORL



(c) COIL20

图2 不同方法的 ACC 与特征维度曲线图

Fig.2 Curves of ACC vs feature dimension of different methods

从图 2 中可以看出,随着特征维度的增加,所有特征选择

方法的聚类性能上升。然而,在达到最优性能后,算法的性能开始趋于稳定或者下降。另外,所有方法的最佳特征维度如表5所列。从表5中可知,JMMSSR方法的最优特征维度低于大多数方法。由此可见,JMMSSR方法能够选择最具有判别能力的特征。

表5 不同方法在3个数据库中的最佳特征维数

Table 5 Best feature dimension of different methods on three

databases			
方法	JAFFE	ORL	COIL20
Baseline	1024	1024	1024
LS	110	280	400
SPEC	390	170	340
UDFS	140	370	130
RUFS	500	140	360
RSR	500	270	470
MRSR	480	350	470
L1UFS	400	170	220
JMMSSR	130	180	380

最后,为了验证本文所提出的优化算法的收敛性,图3给出了JMMSSR算法在3个数据库上的目标曲线,其中横坐标和纵坐标分别表示算法迭代数和目标函数值。从图3中可以看出,JMMSSR算法可以在较少的迭代次数情况下能达到收敛。

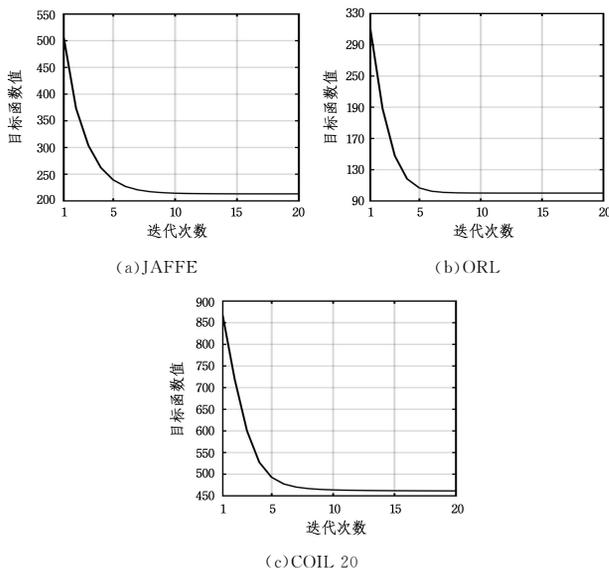


图3 不同数据库上的目标曲线

Fig. 3 Convergence curves of our method on three databases

结束语 为了在特征选择过程中更好地挖掘特征的流形结构,通过自适应融合多种不同流形结构,提出了一种联合多流形结构和自表示的无监督特征选择方法。该方法不仅利用了特征的自表示特性,还充分利用了特征的流形结构。在JAFFE,ORL和COIL20 3个数据库上的实验结果验证了本文方法的性能要优于LS,SPEC,UDFS,RUFS,RSR,MRSR和L1UFS等方法。此外,本文提出了一种迭代交替更新的优化算法对目标函数进行求解,并通过数值实验验证了优化算法的收敛性。

虽然本文方法中考虑了不同的图构建方法,但是这些图的构建与后续的特征选择仍是相互独立的,即事先预定义的图结构未必有利于后续的特征选择,因此,如何在特征选择过程中自适应地构建图,即同时实现特征选择和图构建过程是下一步需要研究的内容。此外,将传统的特征选择方法与深

度学习结合也是未来的重点研究方向。

参考文献

- [1] LI Y, LI T, LIU H. Recent advances in feature selection and its applications [J]. Knowledge and Information Systems, 2017, 53(3): 551-577.
- [2] LI J, LIU H. Challenges of feature selection for big data analytics [J]. IEEE Intelligent Systems, 2017, 32(2): 9-15.
- [3] GUI J, SUN Z, JI S, et al. Feature selection based on structured sparsity: A comprehensive study [J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 28(7): 1490-1507.
- [4] LI Z Q, DU J Q, NIE B, et al. Summary of Feature Selection Methods. CEA, 2019, 55(24): 10-19.
- [5] FANG B, CHEN H M, WANG S W. Feature Selection Algorithm Based on Rough Sets and Fruit Fly Optimization [J]. Computer Science, 2019, 46(7): 157-164.
- [6] LI J, CHENG K, WANG S, et al. Feature selection: A data perspective [J]. ACM Computing Surveys (CSUR), 2018, 50(6): 94.
- [7] RIPLEY B D, HJORT N L. Pattern recognition and neural networks [M]. Cambridge University press, 1996.
- [8] HE X, CAI D, NIYOGI P. Laplacian score for feature selection [C] // Advances in Neural Information Processing Systems, 2006: 507-514.
- [9] ZHAO Z, LIU H. Spectral feature selection for supervised and unsupervised learning [C] // Proceedings of the 24th International Conference on Machine Learning. ACM, 2007: 1151-1157.
- [10] 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.
- [11] LI Z, YANG Y, LIU J, et al. Unsupervised feature selection using nonnegative spectral analysis [C] // Twenty-Sixth AAAI Conference on Artificial Intelligence, 2012.
- [12] NIE F, HUANG H, CAI X, et al. Efficient and robust feature selection via joint $\ell_2, 1$ -norms minimization [C] // Advances in Neural Information Processing Systems, 2010: 1813-1821.
- [13] QIAN M, ZHAI C. Robust unsupervised feature selection [C] // Twenty-Third International Joint Conference on Artificial Intelligence, 2013.
- [14] ZHU P, ZUO W, ZHANG L, et al. Unsupervised feature selection by regularized self-representation [J]. Pattern Recognition, 2015, 48(2): 438-446.
- [15] LIANG S, XU Q, ZHU P, et al. Unsupervised feature selection by manifold regularized self-representation [C] // 2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2017: 2398-2402.
- [16] TANG C, ZHU X, CHEN J, et al. Robust graph regularized unsupervised feature selection [J]. Expert Systems with Applications, 2018, 96: 64-76.
- [17] QIAO L, ZHANG L, CHEN S, et al. Data-driven graph construction and graph learning: A review [J]. Neurocomputing, 2018, 312: 336-351.
- [18] BELKIN M, NIYOGI P. Laplacian eigenmaps and spectral techniques for embedding and clustering [C] // Advances in Neural Information Processing Systems, 2002: 585-591.