

基于自适应局部保持投影的无监督特征选择方法

严菲, 王晓栋

(厦门理工学院计算机与信息工程学院, 福建厦门 361024)

摘要: 基于谱图的无监督特征选择方法在原始高维空间构建图, 易受噪声或冗余特征干扰. 为此提出一种基于自适应局部保持投影的无监督特征选择方法, 利用全局线性回归函数建立特征选择模型, 结合自适应局部保持投影提高模型准确度, 引入 $l_{2,1}$ 约束提升特征之间可区分度, 避免噪声干扰. 最后通过实验验证了该方法的有效性.

关键词: 特征选择; 无监督学习; 线性回归; 局部保持投影; $l_{2,1}$ 范数

中图分类号: TP391.4 **文献标识码:** A **doi:** 10.3969/j.issn.0253-2778.2018.04.004

引用格式: 严菲, 王晓栋. 基于自适应局部保持投影的无监督特征选择方法[J]. 中国科学技术大学学报, 2018, 48(4): 290-297.

YAN Fei, WANG Xiaodong. Unsupervised feature selection method based on adaptive locality preserving projection[J]. Journal of University of Science and Technology of China, 2018, 48(4): 290-297.

Unsupervised feature selection method based on adaptive locality preserving projection

YAN Fei, WANG Xiaodong

(College of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, China)

Abstract: The unsupervised feature selection method based on spectrogram is constructed in the original high dimensional data space, which is easily disturbed by noise or redundant features. To overcome these deficiencies, an unsupervised feature selection method based on adaptive locality preserving projection is proposed. Global linear regression function is utilized to construct feature selection model, and the adaptive local preserving projection is adopted to improve model accuracy. Then the $l_{2,1}$ -norm constraint is added to improve the distinguishability of different features and avoid noise interference. A comparison with several state-of-the-art feature selection methods demonstrate the effectiveness of the proposed method.

Key words: feature selection; unsupervised learning; linear regression; locality preserving projection; $l_{2,1}$ -norm

0 引言

目前, 随着互联网技术的发展, 高维数据在模式

识别、机器学习等研究领域呈现出爆发式增长. 如何从高维数据中快速、准确获取有效信息, 成为当前研究热点. 特征选择从原始数据集中提取最优的特征

收稿日期: 2017-05-25; 修回日期: 2017-06-24

基金项目: 国家自然科学基金(61502405), 福建省自然科学基金面上项目(2017J01511), 福建省中青年教育科研项目((JAT160357, JAT170417)资助.

作者简介: 严菲(通讯作者), 女, 硕士/实验师, 研究方向: 模式识别、数据挖掘. E-mail: 275223188@qq.com

子集,进而降低数据维度,得到广泛支持。

按已标记样本的数量分类,特征选择方法可分为监督、半监督和无监督学习。在已知标签信息足够多时,监督特征选择方法^[1-2]利用已知数据集和类别标签训练学习模型,识别效果较好。当已知类别标签较少时,监督类的学习方法性能迅速下降。在该场景下半监督学习方法为一种有效的解决方案。半监督方法结合未标记数据和已知标签样本建立学习模型,如 CSFS (convex semi-supervised multi-label feature selection)^[3]利用全局线性回归函数进行特征选择, SSFS (semi-supervised spectral feature selection)^[4]构建聚类结构和特征选择的联合学习模型进行特征选择。实际应用中,从海量数据中人工获取类别标签非常困难,因此研究无监督特征选择方法尤为重要。

近年来,国内外学者对无监督特征选择方法进行大量的研究。其中,最大方差(maximum variance, MaxVar)^[5]以数据方差为评价机制,并以特征重要性排序来进行筛选,是一种最简单的无监督方法。LS(laplacian score)^[6]在 MaxVar 基础上引入了数据局部结构的分析。这两者仅考虑数据本身的特征,忽略了特征之间相关性,从而导致模型容易陷入局部最优。为解决特征间的相关性问题,有学者借鉴流形学习进行特征选择^[7-13]。其中 UDFS(unsupervised discriminant feature selection)^[7]采用局部类间散度最大化与类内散度最小化的策略以获取最优特征子集。NDFS^[8]在 UDFS 基础上引入非负约束以获得更准确的结果;然而这类方法要求数据样本严格服从相应的函数关系,因而很难准确获得数据样本的流形结构。为提高准确度,研究者利用谱图模型提取数据底层流形结构,如采用谱聚类和 l_1 范数的 USFS(unsupervised spectral feature selection)^[11]。多种实际场景证明,利用谱图构造模型的特征选择有利于提高学习效率。这类方法一般先构造图来标注数据样本间的属性特征,然后在此基础上采用各类方法进行特征选择,因此图的构造直接影响到特征选择的结果。JELSR(joint embedded learning and sparse regression)^[12]和 UFSGO(unsupervised feature selection graph optimization)^[13]对谱图方法进行改进,将图构造嵌入到特征选择过程中,以提高学习效率。总体来说,基于谱图的方法存在以下问题:①该类方法往往需在谱图基础上建立 Laplacian

矩阵,并对其进行特征分解,从而耗费大量的时间和空间资源;②该类方法在原始高维数据空间构建谱图,容易受原始数据集噪声和冗余特征影响,进而难以选择最优特征。有学者提出利用局部流形结构信息来提高特征选择的可判别能力,如局部保持投影 LPP(locality preserving projection)^[14],且能节约计算资源,但基于 LPP 的方法依然需在原始高维空间上构建相似矩阵。

本文提出一种新的无监督特征选择方法。首先,利用全局线性回归函数建立特征选择模型,避免基于谱图方法依赖于 Laplacian 特征分解带来的计算复杂度高的问题。其次,引入自适应局部投影的思想,将底层流形结构提取与特征选择过程动态结合,在经特征筛选后的低维特征空间中提取流形结构。同样地,利用所学习到的流形结构进一步约束后续的特征学习过程。综上所述,本文提出一种基于自适应局部保持投影的无监督特征选择方法 UFSA(unsupervised feature selection based on adaptive locality preserving projection),该方法具有如下优点:

(I)不同于传统基于谱图的特征选择方法,本文方法不仅利用数据间的线性关系,还借助数据底层的流形结构有效提取局部特征,建立自适应图以提高学习效率。

(II)为提高特征选择准确度,本文引入 $l_{2,1}$ 范数过滤冗余特征,避免了噪声的干扰。

(III)针对模型涉及 $l_{2,1}$ 范数是非光滑的,无法直接对其求闭合解,本文提出一套快速有效迭代方法求解学习模型。

1 局部保持投影

局部保持投影(LPP)是一种流形学习方法,旨在降维后寻找一种子空间以尽量保持原高维数据之间的局部特性。LPP 利用数据样本的近邻关系,通过构造 Laplacian 矩阵以获取其低维投影矩阵。定义数据集 $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$, 其中 $x_i \in R^d$ 为第 i 组数据, $\mathbf{Q} \in R^{d \times m}$ 为转换矩阵($m \ll d$), LPP 方法目标函数为

$$\min_{\mathbf{Q}^T \mathbf{Q} = \mathbf{I}} \sum_{i,j=1}^n \|\mathbf{Q}^T x_i - \mathbf{Q}^T x_j\|_2^2 s_{ij} \quad (1)$$

式中,正交约束 $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}$ 以确保变换后子空间的数据间的无关性。 \mathbf{S} 为对称相似度矩阵,其元素定义为

$$s_{ij} = \begin{cases} \exp(-\|x_i - x_j\|^2/t), t \in R, x_i \text{ 为 } x_j \text{ 的近邻} \\ 0, \text{其他} \end{cases}$$

从式(1)可看出, LPP 方法的核心思想是寻找矩阵 Q , 将高维数据 X 转换为低维矩阵 $Q^T X$, 同时最大程度地保持原数据集 X 的局部相关性结构. 最小化式(1)可保证, 如果 x_i 和 x_j 相邻, 则 $Q^T x_i$ 和 $Q^T x_j$ 也是相邻的.

2 基于自适应局部保持投影的无监督特征选择方法

传统基于谱图的无监督特征选择方法往往需要借助 Laplacian 矩阵的分解获取全局聚类结构信息, 需要耗费大量的时间和空间, 因此本文将线性回归函数引入特征学习模型中, 逐步训练未标记数据集, 迭代产生置信度最高的特征集进行选择.

给定一个共 n 组数据的数据集 X , 定义 $F = [f_1, f_2, \dots, f_n] \in R^{n \times c}$ 为预测标签矩阵, c 为标签数, 在无监督学习中, F 为未标记训练集且随机初始化, $W \in R^{d \times c}$ 为特征选择分类器, 特征选择模型定义如下:

$$\min_{W, F, 0 \leq f_i \leq 1} \sum_{i=1}^n \text{loss}(g(W, x_i), f_i) + \gamma \Omega(W) \quad (2)$$

式中, $\Omega(\cdot)$ 为正则化模型(一般正则化模型可选择 l_1 范数、 l_2 范数等), 参数 γ 为正则化参数, $g(\cdot)$ 为映射函数, 将 x_i 映射到一个新的空间, $\text{loss}(\cdot)$ 为损失函数, 常见的损失函数包括最小二乘、sigmoid 函数、logistic 函数等. 本文从模型的简单性、高效性角度进行考虑, 选择最小二乘法作为损失函数, 式(2)可表示为

$$\min_{W, F, b, 0 \leq f_i \leq 1} \sum_{i=1}^n \|W^T x_i + b - f_i\|_2^2 + \gamma \Omega(W) \quad (3)$$

式中, $b \in R^{c \times 1}$ 为偏置量.

从式(3)可以看出, 利用线性回归函数逐渐逼近可找出全局最优, 但其忽略了局部数据之间相关性. 本文结合式(1)的 LPP 方法进行特征选择, 提出以下学习模型:

$$\begin{aligned} & \min_{W, F, b, s_{ij}} \sum_{i,j=1}^n \|W^T x_i - W^T x_j\|_2^2 s_{ij} + \\ & \alpha \sum_{i=1}^n \|W^T x_i + b - f_i\|_2^2 + \gamma \Omega(W) \\ \text{s.t.} \quad & 0 \leq f_i \leq 1, W^T W = I \end{aligned} \quad (4)$$

式中, α 为平衡参数.

在 LPP 方法中, 相似度矩阵 S 一旦建立好, 在后续的特征选择过程中不再更新, 将直接影响最终特征选择的结果; 因此, 利用 LPP 方法要获得较好的特征选择效果, 需提前建立好最佳相似度矩阵. 这在实际应用中是比较困难的. 文献[15]提出建立自适应邻接矩阵的聚类方法(projection clustering with adaptive neighbors, PCAN), 在降维过程中动态重构相似矩阵, 更有利于获取准确的流形结构. 为提高特征选择效果, 本文借鉴 PCAN 方法, 在全局线性回归函数建立全局学习模型基础上, 结合自适应邻接矩阵获取局部流形结构, 提出基于自适应局部保持投影的无监督特征选择方法.

定义 s_{ij} ($0 \leq s_{ij} \leq 1$) 为数据 x_i 与相邻数据的相似程度, 自适应局部保持投影目标函数定义如下:

$$\begin{aligned} & \min_{S, W} \sum_{i,j=1}^n \|W^T x_i - W^T x_j\|_2^2 s_{ij} \\ \text{s.t.} \quad & \forall i, s_i^T \mathbf{1} = 1, 0 \leq s_i \leq 1, W^T W = I \end{aligned} \quad (5)$$

式中, $\|W^T x_i - W^T x_j\|_2^2$ 值越小, 表明 x_i 与 x_j 的邻接可能性越大, 即 s_{ij} 的值越大. 对其进行求最优解时发现, 只有 x_i 的最近邻接点值 s_{ij} 为 1, 而其他相邻数据无法成为其邻接点. 为此在式(5)基础上加上式 $\min_{s_i^T \mathbf{1}=1, 0 \leq s_i \leq 1} \sum_{j=1}^n s_{ij}^2$ 来确保相似矩阵的平滑性, 进而获取有效的函数. 自适应局部保持投影模型为

$$\begin{aligned} & \min_{S, W} \sum_{i,j=1}^n \|W^T x_i - W^T x_j\|_2^2 s_{ij} + \lambda s_{ij}^2 \\ \text{s.t.} \quad & \forall i, s_i^T \mathbf{1} = 1, 0 \leq s_i \leq 1, W^T W = I \end{aligned} \quad (6)$$

式中, λ 为正则化参数.

结合式(4)和式(6), 本文学习模型可表示为

$$\begin{aligned} & \min_{W, S, F, b, s_{ij}} \sum_{i,j=1}^n (\|W^T x_i - W^T x_j\|_2^2 s_{ij} + \lambda s_{ij}^2) + \\ & \alpha \sum_{i=1}^n \|W^T x_i + b - f_i\|_2^2 + \gamma \Omega(W), \end{aligned}$$

s.t. $\forall i, s_i^T \mathbf{1} = 1, 0 \leq s_i \leq 1, 0 \leq f_i \leq 1, W^T W = I$.

为排除噪声的干扰, 有效地进行输入数据特征的选择, 在目标函数中引入正则化模型 $l_{2,1}$ 范数, 其定义为 $\|W\|_{2,1} = \sum_{i=1}^d \sqrt{W_{ij}^2}$. 最后, 本文提出的目标函数如下:

$$\min_{W, S, F, b, s_{ij}} \sum_{i,j=1}^n (\|W^T x_i - W^T x_j\|_2^2 s_{ij} + \lambda s_{ij}^2) +$$

$$\begin{aligned} & \alpha \sum_{i=1}^n \|W^T x_i + b - f_i\|_2^2 + \gamma \|W\|_{2,1} \\ \text{s.t. } & \forall_i, s_i^T \mathbf{1} = 1, 0 \leq s_i \leq 1, 0 \leq f_i \leq 1, W^T W = I \end{aligned} \quad (7)$$

3 学习模型求解

由于目标函数式(7)引入的 $l_{2,1}$ 范数是非光滑的,不能直接对其进行求解,因此本文提出一种迭代优化方法来解决.首先定义 $L_s = D_s - \frac{S^T + S}{2}$, $D_s \in R^{n \times n}$ 为对角矩阵,其第 i 个对角元素值为 $\sum_j (s_{ij} + s_{ji})/2$.为方便求解,将目标函数(7)转换为以下矩阵形式:

$$\begin{aligned} & \min_{W, S, F, b} \text{Tr}(W^T X L_s X^T W + \\ & \alpha (X^T W + \mathbf{1} b^T - F)^T (X^T W + \mathbf{1} b^T - F) + \\ & \lambda \|S\|_F^2 + \gamma \|W\|_{2,1} \end{aligned} \quad (8)$$

首先保持 W, S, F 不变,将式(8)对 b 求导,令求导结果为 0,得到

$$b = \frac{1}{n} F^T \mathbf{1} - \frac{1}{n} W^T X \mathbf{1} \quad (9)$$

为简化公式书写,仅将上式推导结果式(9)代入式(8)与 b 相关的部分,得到

$$\begin{aligned} & \min_{W, F} \alpha \left((I - \frac{1}{n} \mathbf{1} \mathbf{1}^T) X^T W - \right. \\ & \left. (I - \frac{1}{n} \mathbf{1} \mathbf{1}^T) F \right)^T \left((I - \frac{1}{n} \mathbf{1} \mathbf{1}^T) X^T W - \right. \\ & \left. (I - \frac{1}{n} \mathbf{1} \mathbf{1}^T) F \right) \end{aligned} \quad (10)$$

式中, $I \in R^{n \times n}$ 为单位矩阵.定义 $H = (I - \frac{1}{n} \mathbf{1} \mathbf{1}^T)$

为中心化后的矩阵,则上述式(10)变成

$$\min_{S, W} (H X^T W - H F)^T (H X^T W - H F) \quad (11)$$

将式(11)替式(8)中相关部分,将式(8)转换为:

$$\begin{aligned} & \min_{W, S, F} \text{Tr}(W^T X L_s X^T W) + \\ & \alpha (H X^T W - H F)^T (H X^T W - H F) + \\ & \lambda \|S\|_F^2 + \gamma \|W\|_{2,1} \end{aligned} \quad (12)$$

保持 S, F 不变,将式(12)对 W 求导并令求导结果为 0,得到

$$X L_s X^T W + \alpha X H X^T W + \gamma D_w W = \alpha X H F \quad (13)$$

式中, D_w 为对角矩阵,可表示为

$$D_w = \begin{bmatrix} \frac{1}{2 \|w_1\|_2} & \circ & \circ \\ \circ & \ddots & \circ \\ \circ & \circ & \frac{1}{2 \|w_d\|_2} \end{bmatrix} \quad (14)$$

由于矩阵 D_w 与 W 相关,无法直接求解上式.因此首先随机初始化 W 以获取矩阵 D_w ,转换式(13),推导 W 为

$$W = \frac{1}{\alpha} (X L_s X^T + \alpha X H X^T + \gamma D_w)^{-1} X H F \quad (15)$$

根据文献[7]描述特征选择矩阵 W 和标签矩阵存在某种线性关系,即 $\tilde{F} = X^T W + \mathbf{1} b$,为实现最小化目标函数,本文按以下公式调整未标记训练集的标签:

$$F_{ij} = \begin{cases} 0, & \tilde{F}_{ij} < 0 \\ \tilde{F}_{ij}, & 0 \leq \tilde{F}_{ij} \leq 1 \\ 1, & \tilde{F}_{ij} > 1 \end{cases} \quad (16)$$

最后求解 S ,首先保持 W, F 不变,将式(7)转换为

$$\begin{aligned} & \min_S \sum_{i,j=1}^n \|W^T x_i - W^T x_j\|_2^2 s_{ij} + \lambda s_{ij}^2 \\ \text{s.t. } & \forall_i, s_i^T \mathbf{1} = 1, 0 \leq s_i \leq 1 \end{aligned} \quad (17)$$

为方便求解,对于式(17)中不同的 i 值,假设求解 s_i 为相互独立^[15].因此本文将式(17)进行转换,得到

$$\begin{aligned} & \min_{s_i} \sum_{j=1}^n \|W^T x_i - W^T x_j\|_2^2 s_{ij} + \lambda s_{ij}^2 \\ \text{s.t. } & \forall_i, s_i^T \mathbf{1} = 1, 0 \leq s_i \leq 1 \end{aligned} \quad (18)$$

令 $q_{ij}^w = \|W^T x_i - W^T x_j\|_2^2$, $q_i^w \in R^{n \times 1}$, q_{ij}^w 为其第 j 列元素.上述公式(18)可改为

$$\min_{s_i^T \mathbf{1} = 1, 0 \leq s_i \leq 1} \left\| s_i + \frac{1}{2\lambda} q_i^w \right\|_2^2 \quad (19)$$

基于以上推导过程,本文提出一种迭代算法对学习模型求解,算法的具体描述如下:

算法 3.1

输入:数据集 $X \in R^{d \times n}$,参数 α, β, γ

输出:特征选择矩阵 W ,预测标签矩阵 F

1 初始化矩阵 S, F

2 $t=0$,随机初始化特征选择矩阵 W^0

3 repeat

4 更新 $L_s = D_s - \frac{S^T + S}{2}$, $D_s \in R^{n \times n}$ 为对角矩阵,其第 i 个元素值为 $\sum_j (s_{ij} + s_{ji})/2$

5 根据式(14)计算对角矩阵 D_w .

6 计算 $W^{t+1} = \frac{1}{\alpha} (X L_s X^T + \alpha X H X^T + \gamma D_w) - 1 X H F^T$

7 计算 $b_{t+1} = \frac{1}{n} (F^T)^T 1 - \frac{1}{n} (W^T)^T X 1$

8 计算 $\hat{F}_{t+1} = X^T W^T + 1 (b^T)^T$, 按式(16)调整 F^t

9 for $i=1$ to n

10 按式(19)求解 s_i

11 $i=i+1$

12 end

13 $t=t+1$

14 until 收敛

15 在得到特征选择矩阵 W 后, 按照 $\|w_i\|_2 (i=1, \dots, d)$ 的值, 对输入数据的特征进行降序排列, 其中最前面的 p 个特征即为所选择的特征.

4 实验分析

4.1 对比方法及实验数据

为了验证本文方法的有效性, 所有实验在 CPU 3.7 GHz、内存 16 GB 以及操作系统为 Windows10 64 位的环境下完成. 本文选取了目前流行的几种无监督特征选择方法进行对比分析, 相关对比方法描述如下:

(I) All-feature. 数据未采用特征选择, 本文选择该方法的结果作为基准线.

(II) MaxVar. 经典的无监督选择方法, 逐个选择最大方差选择特征.

(III) 多簇特征选择方法 MCFS^[10]. 采用对 Laplacian 矩阵的谱分析获取数据特征间的相关关系, 结合 l_1 正则回归选择特征.

(IV) 无监督判别特征选择 UDFS^[7]. 采用以局部类内最大化和类间最小化为约束, 结合 $l_{2,1}$ 范数进行特征选择.

(V) 联合嵌入式学习与稀疏回归方法 JELSR^[12]. 采用局部线性近似估计获取流形结构进行特征选择.

同时将各方法应用到多种开源数据库, 包括 BREAST、WINE、VOTE、UMIST、JAFPE^[16], 表 1 列出了所选数据集的相关描述.

表 1 实验数据集相关描述

Tab.1 The description of datasets

数据集大小(K)	维度	类别	特征数
BREAST	699	10	2 {1,2,3,4,...,7,8,9}
WINE	178	13	3 {1,2,3,4,...,6,7}
VOTE	435	16	2 {1,2,3,4,...,9,10}
UMIST	575	644	20 {100,200,300,...,500,600}
JAFPE	213	1 024	10 {100,200,300,..., 700,800,900}

4.2 聚类准确度分析

定义 p_i 和 q_i 为给定样本 x_i 聚类结果和样本自带类标签, 聚类准确度 ACC (accuracy) 定义公式如下:

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

式中, n 为样本集总数, $\delta(x, y)$ 为比较函数, 当 $x = y$ 时, 其值为 1, 否则为 0. $\text{map}(q_i)$ 为正交函数, 采用最佳匹配算法 KM (Kuhn-Munkres) 将聚类结果 p_i 映射成与样本自带的类标签等价的类标签.

对于每种算法所有涉及的参数, 范围设定为 $\{10^{-9}, 10^{-6}, 10^{-3}, 10^0, 10^3, 10^6, 10^9\}$, 将涉及的邻近值 k 设为 15. 本次实验中, 首先采用特征选择方法挑选出最具代表性的特征子集, 再利用 k -means 进行聚类, 每种方法 k -means 运行 20 次. 每组实验数据重复 5 次, 最后计算 5 次结果的平均值和标准方差. 各类方法的平均聚类准确度如表 2 所示.

表 2 各种方法平均聚类准确度对比及特征选择值

Tab.2 Mean clustering accuracy comparison of different feature selection methods and value of feature selection %

数据集	All-feature	MaxVar	MCFS	UDFS	JELSR	UFSA
BREAST	60.09±0.00	60.09±0.00(1)	95.42±0.00(7)	63.81±0.00(1)	95.42±0.00(7)	95.48±0.13(7)
WINE	70.22±0.00	70.22±0.00(2)	72.38±0.48(7)	84.27±0.00(4)	89.33±0.00(4)	90.34±0.25(4)
VOTE	83.45±0.00	71.49±0.00(10)	72.20±2.10(10)	93.56±0.00(2)	91.72±0.00(5)	95.59±0.10(2)
UMIST	41.23±0.41	46.99±0.32(100)	45.55±0.30(100)	44.92±0.32(300)	46.07±0.25(100)	47.94±1.42(100)
JAFPE	72.82±1.35	75.02±1.77(800)	70.82±0.92(700)	95.40±1.42(900)	97.74±0.21(400)	96.71±0.00(900)

表 2 列出了各类方法在不同数据集上的聚类准确度比较及最终选择的特征值,由于 All-feature 方法没有进行特征选择,因此表 2 中未给出相应数据.从实验结果中可以看出:

(I) 在所有数据集中测试中,大部分特征选择方法要优于未采用特征选择的 ALL-feature 方法,由此证明特征选择方法有助于提高分类性能.其中,在 VOTE、UMIST 和 JAFFE 数据集中,虽然某些方法准确度稍低于 All-feature,但其特征维度有所降低,从而能有效地节省后续分类或聚类时间.

(II) 本文提出的 UFSA 方法大多数数据集中优于其他方法.其中 JAFFE 数据集由于存在光照等因素,造成部分局部相关性受影响.因此,本文方法聚类准确度低于基于全局流形学习的 JELSR 方法.相对于基于局部流形结构 UDFS 方法,本文方法略显优势,由此可证明,利用自适应局部投影动态获取局部流形结构有助于提升性能.

4.3 时间复杂度分析

从算法 3.1 描述中可看出,UFSA 最主要的操作集中在相似度矩阵计算和求解特征矩阵 W 过程

中的矩阵逆运算,其复杂度分别对应为 $O(n\log(n) + d^3)$,其中 n 代表训练样本数, d 代表样本特征维度, m 代表已选择的特征数.实验中所对比方法的具体时间复杂度如表 3 所示.

实验中,各种方法所涉及的参数(如果有的话)固定为 1,并选择平均特征选择数进行测试,其测试过程与 4.2 小节中的过程相同.每个方法运行 20 次,最后结果取其平均值.具体运行结果如表 4 所示.

表 3 各种方法时间复杂度对比
Tab.3 Time complexity comparison of different feature selection methods

方法	时间复杂度
All-feature	$O(dn)$
MaxVar	$O(dn)$
MCFS	$O(n^2 + d^3)$
UDFS	$O(n^2 + d^3)$
JELSR	$O(d^2n + n^3)$
UFSA	$O(n\log(n) + d^3)$

表 4 各种方法平均运行时间对比(单位:秒)

Tab.4 Mean run-time comparison of different feature selection methods

数据集	All-feature	MaxVar	MCFS	UDFS	JELSR	UFSA
BREAST	0.069 5	0.050 6	0.113 53	0.267 21	1.249 2	0.364 75
WINE	0.064 1	0.063 0	0.081 0	0.097 6	0.156 29	0.125 2
VOTE	0.065 25	0.054 2	0.080 5	0.123 9	0.690 29	0.423 8
UMIST	6.329 1	3.069 3	4.247	6.175 6	4.438 1	5.340 7
JAFFE	1.329	0.612 66	3.727 6	7.007 6	2.253 6	1.917 8

从表 4 可以看出:① All-feature 没有涉及特征选择,时间优于其他特征选择算法.随着数据集的维度增大,其运行时间显著增大(如 UMIST 数据集);② 虽然 MaxVar 和 All-feature 具有相同的时间复杂度,但其在后续的聚类时,去除了大量冗余噪声,其运算速度得到了有效的提升(如 UMIST 和 JAFFE);③ UDFS、MCFS 和 UFSA 都具有 $O(d^3)$ 复杂度,但前两种方法涉及较多的特征分解运算,当数据维度增大时,其复杂度要高于 UFSA;④ UFSA 相较于 JELSR 具有更高的运算效率,尤其是当训练样本数较大时,UFSA 具有更多的优势.

4.4 特征数对聚类准确度的影响分析

本次实验挑选了 WINE、VOTE 和 UMIST 数

据集,分析各方法在选择不同特征数时的最大聚类准确度对比效果,如图 1 所示.从图 1 可以看出,对于所有的数据,最高聚类准确度值并不是在选择所有特征值时,这意味着选择最大特征数并不一定能产生最高的聚类准确度.另外,本文所提出的 UFSA 方法的聚类准确度要普遍高于其他方法.

4.5 参数对数据集的敏感性分析

本次实验将对学习模型中涉及的参数进行具体分析,其中参数 λ 可参考文献[15]进行自动调节.为节省篇幅,仅挑选 VOTE、UMIST 和 JAFFE 三组数据集进行分析.首先,固定 α 的值为 1,即参数调节范围的中位数,调整 γ 和特征数量值进行数据聚类 and 标准差分析,结果如图 2(a) 所示.同样,固定参数

γ 的值为 1, 对 α 和特征数量的变化进行分析, 具体如图 2(b) 所示. 从图 2 可以看出, 参数 α 、 γ 的选择

依赖于所选数据集, 不同的数据集需设置不同的参数.

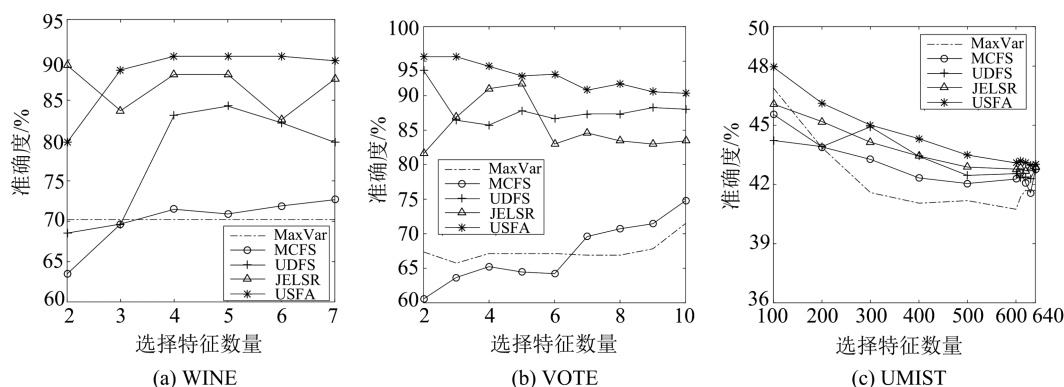


图 1 聚类准确度与选择的特征数量关系

Fig.1 Relation between clustering accuracy and the number of selected features

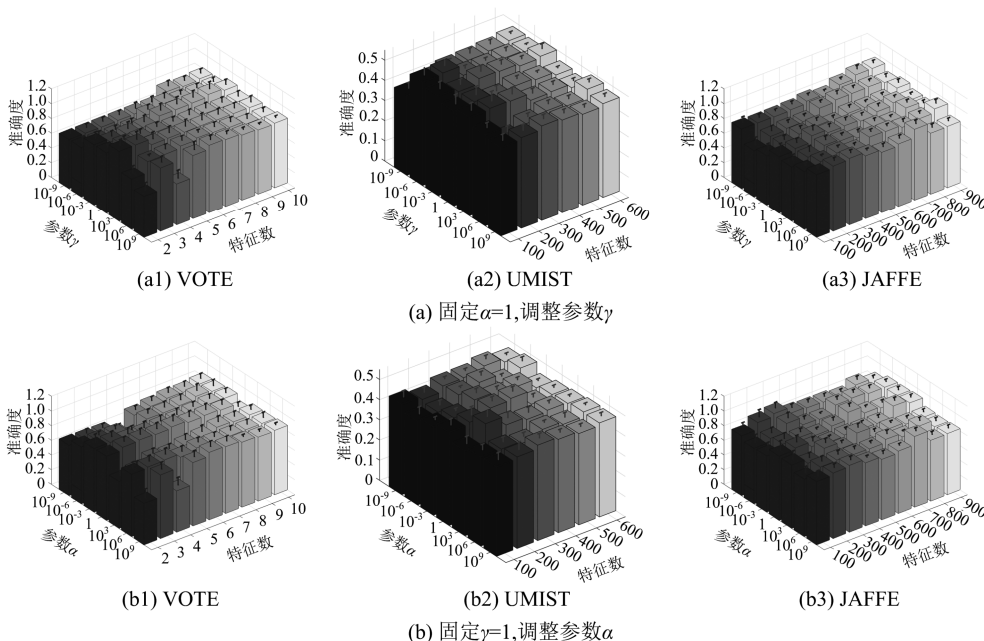


图 2 本文方法的参数敏感性分析

Fig.2 Parameter sensitivity analysis of the proposed method

5 结论

本文提出基于自适应局部保持投影的无监督特征选择方法 UFSA. 该方法利用线性回归建立全局模型, 并结合自适应局部保持投影获取底层流形结构, 以提高特征选择效率. 通过在多个开源数据集上的实验, 表明本文方法是一种高效的无监督特征选择方法.

参考文献 (References)

[1] NIE F, CAI X, HUANG H, et al. Efficient and robust feature selection via joint $l_{2,1}$ -norms minimization

[C]// Advances in Neural Information Processing Systems. Chicago: IEEE Press, 2010:1813-1821.

[2] YANG Y, MA Z, HAUPTMANN A G, et al. Feature selection for multimedia analysis by sharing information among multiple tasks [J]. IEEE Transactions on Multimedia, 2013, 15(3): 661-669.

[3] CHANG X, NIE F, YANG Y, et al. A convex formulation for semi-supervised multi-label feature selection [C]// 28th AAAI Conference on Artificial Intelligence. Québec, Canada: AAAI Press, 2014: 1171-1177.

[4] 王晓栋, 严菲, 谢勇, 等. 基于稀疏图表示的特征选择方法研究 [J]. 计算机工程与科学, 2015, 37(12): 2372-2378.

- WANG X D, YAN F, XIE Y, et al. A feature selection method based on sparse graph representation [J]. *Computer Engineering & Science*, 2015, 37(12): 2372-2378.
- [5] REN Y, ZHANG G, YU G, et al. Local and global structure preserving based feature selection [J]. *Neurocomputing*, 2012, 89(10):147-157.
- [6] HE X, CAI D, NIYOGI P. Laplacian score for feature selection [J]. *Advances in Neural Information Processing Systems*, 2005, 18: 507-514.
- [7] YANG Y, SHEN H T, MA Z, et al. $l_{2,1}$ -norm regularized discriminative feature selection for unsupervised learning [C]// *International Joint Conference on Artificial Intelligence*. Barcelona, Spain: AAAI Press, 2011: 1589-1594.
- [8] LI Z, YANG Y, LIU J, et al. Unsupervised feature selection using nonnegative spectral analysis[C]// *26th AAAI Conference on Artificial Intelligence*. New York: AAAI Press, 2012:1026-1032.
- [9] ZHAO Z, LIU H. Spectral feature selection for supervised and unsupervised learning [C]// *Proceedings of the 24th International Conference on Machine Learning*. Corvallis, USA: ACM Press, 2007: 1151-1157.
- [10] CAI D, ZHANG C, HE X. Unsupervised feature selection for multi-cluster data[C]// *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Washington: ACM Press, 2010: 333-342.
- [11] WANG X D, ZHANG X, ZENG Z, et al. Unsupervised spectral feature selection with l_2 -norm graph[J]. *Neurocomputing*, 2016, 200(C): 47-54.
- [12] HOU C, NIE F, LI X, et al. Joint embedding learning and sparse regression: A framework for unsupervised feature selection [J]. *IEEE Transactions on Cybernetics*, 2014, 44(6): 793-804.
- [13] ZHANG Z, BAI L, LIANG Y, et al. Unsupervised Feature Selection by Graph Optimization[M]// *Image Analysis and Processing - ICIAP*, Springer, 2015.
- [14] HE X, NIYOGI P. Locality preserving projections (LPP)[J]. *Advances in Neural Information Processing Systems*, 2002, 16(1): 186-197.
- [15] NIE F, WANG X, HUANG H. Clustering and projected clustering with adaptive neighbors [C]// *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Québec, Canada: ACM Press, 2014: 977-986.
- [16] LYONS M J, BUDYNEK J, AKAMATSU S. Automatic classification of single facial images [J]. *IEEE Transactions on Pattern Analysis*, 1999, 21(12): 1357-1362.