文章编号：1674-8085(2021)02-0076-07

基于自表示的双图规格化特征选择聚类

*汪宏海[1,2]，吴　樱[1,2]

（1.浙江省文化和旅游发展研究院，浙江，杭州　311231；2.浙江旅游职业学院，浙江，杭州　311231）

摘　要：特征选择得到的识别特征可以用于聚类分析，提高聚类分析的质量。受数据自表示特性和双图规则化学习的启发，提出了一种新的特征选择聚类算法。利用数据和特征的自表示特性，不仅保留了数据的流形信息，而且保留了特征空间的流形信息。此外，为了充分发挥双图模型的作用和鉴别局部聚类的效果，加入局部判别特征选择聚类，大大提高了聚类的有效性和鲁棒性。

关键词：特征选择；自表示；双图规格化；聚类

中图分类号：TP18　　**文献标识码**：A　　DOI:10.3969/j.issn.1674-8085.2021.02.013

FEATURE SELECTION CLUSTERING BASED ON SELF-REPRESENTING DOUBLE GRAPH NORMALIZATION

*WANG Hong-hai[1,2]，WU Ying[1,2]

(1. Zhejiang Institute of Culture and Tourism Development, Hangzhou, Zhejiang 311231, China;

2. Tourism College of Zhejiang, Hangzhou, Zhejiang 311231, China)

Abstract: The recognition features obtained through feature selection can be used in cluster analysis to improve the quality. Inspired by the self-representation characteristics of data and regular learning of double graphs, a new feature selection clustering algorithm was proposed. Utilizing the self-representation characteristics of data and features, not only the manifold information of the data but also the manifold information of the feature space was retained. In addition, in order to make the dual graph model work at full capacity, and identify the effects of local clusters, we added the local discriminative features for clustering, which had greatly improved the effectiveness and robustness of clustering.

Key words: feature selection; self-representation; double graph normalization; clustering

0　引言

大数据中存在着高维数据。这些高维的数据不仅数据维度非常大，冗余的数据也非常多。数据降维可以降低数据的维度，排除一些冗余复杂的噪声数据集[1]。数据降维的方法主要有两种：特征选择和特征提取。特征选择是从原始的众多特征中，选择出具有代表性质的特征，由于是从原始的特征集合中筛选出来的特征集，所以不会破坏原有的数据，并且具有解释性强的优点[2]。本研究主要关注特征选择。特征选择按照使用标签信息的方式，可分为监督、半监督、非监督算法。由于现实中很多数据没有标签信息，因此，非监督特征选择具有更广泛的应用[3-4]。

在特征选择领域中，根据选择思想的不同，选择方法可以分为过滤法(filter)、包裹式(wrapper)、嵌入式(embedding)[5]。Embedding 方法由于结合了机器学习的思想，具有较好的效果[6]。很多非监督特征选择算法用来做聚类，由于在选择后获得表示特征用于聚类，因此，聚类质量得到了加强[7]。在 Embedding 方法的研究中，目前较为成熟且较为广

收稿日期：2020-09-23；修改日期：2020-11-09
作者简介：*汪宏海(1977-)，男，江西樟树人，副教授，硕士，主要从事算法优化、图像处理等方面的研究(E-mail:redsea54@163.com);
吴　樱(1980-)，女，浙江杭州人，讲师，硕士，主要从事酒店人力资源管理、酒店数字化运营等方面的研究(E-mail:652348105@qq.com).

泛使用的方法，基本上都是基于将流形学习与 NMF 思想结合的算法，比如拉普拉斯评分（Laplacianscore：LapScore）[8]、谱特征选择（SPEC：spectral feature selection[9]，多类簇特征选择（MCFS：Feature Selection For Multi-cluster Data）[10]，最小冗余谱特征选择算法(MRSF:Spectral Feature Selection with Minimum Redundancy)[11]，联合嵌入式学习和稀疏回归的特征选择(JELSR:Joint embedding Learning and Sparse)[12]，局部性和相似性保留嵌入特征选择(LSPE：locality and similarity preserving embedding feature selection)[13]。

已有的算法中，LapScore 只保留数据流形的位置，但是没有学习机制和迭代更新机制；SPEC 可以看作是 LapScore 的扩展，但主要用于监督特性选择，并不适用于无监督特征选择；考虑到不同集群之间的关系，MCFS 保留了数据集的多集群结构。JELSR 将嵌入式学习和稀疏回归统一到一个无监督的特征选择框架中，利用学习后的稀疏投影矩阵进行特征选择。MRSF 是基于稀疏多输出模型来最小化冗余特性的。LSPE 将嵌入式学习和特征选择集成在一个联合框架中。总之，上述特征选择算法都是在数据流形空间中进行的，有些保持局部信息，有些保持相似性以提高学习性能，都缺少了对数据表示特征的有效使用。

基于此，受自表示特性和双图规则化学习的启发，提出一种新的特征选择聚类算法。本算法不使用投影矩阵，而是利用数据和特征的自表示特性，在数据空间中使用自表示系数矩阵作为特征自表示构造的重要度量，利用数据空间中的自表示系数矩阵对特征进行排序。不仅保留了数据的流形信息，而且保留了特征空间的流形信息。此外，加入局部判别特征选择聚类，可以大大提高聚类的有效性和鲁棒性。

2 本文算法

本文算法利用数据和特征的自表示，特征空间和数据空间的几何信息可以同时表示的思想，通过数据空间中自表示系数矩阵的 $l_{2,1}-$ 范式约束和局部判别，本算法可以选择最有代表性的特征用来聚类。简单示意如图 1 所示。

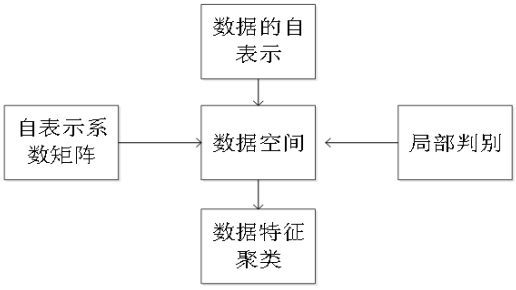

图 1 数据处理过程

Fig.1 Data processing process

2.1 关键技术

本算法中，数据点和特征是通过利用自表示来表示的。对于每个向量 x_i，通过自表示属性，可以由 X 中的所有特征表示，即 x_1; x_2; …; x_n 来表示。可以得出：

$$x_i = \sum_k^n x_k s_{ki} + f_i \tag{1}$$

其中，$s=[s_{ki}]\in R^{n\times n}$ 是特征空间的自表示系数矩阵，f_i 是误差项。公式（1）也可以写为：

$$X=XS+F \tag{2}$$

类似的，在数据空间，可以得到：

$$X^T=X^TP+H \tag{3}$$

其中，$P\in R^{m\times m}$ 是数据空间的自表示系数矩阵，H 为相应的误差矩阵。由于 P 和 S 反映了特征和数据点在自表示过程中的贡献，将 P 和 S 限制为非负的，即 P，$S\geq 0$。

将自表示重构误差最小化，即解决如下问题：

$$\min \| X - XS \|_F^2 + \beta \| X^T - X^T P \|_F^2 \tag{4}$$

其中，参数 β（$\beta\geq 0$）平衡这两个自表示的错误项。

在数据空间和特征空间中构造两个邻域图来保存局部几何信息，即数据图和特征图。在数据空间中构造最近的邻域图 G，图中的每个节点对应一个数据点。如果两个数据点位于 k 个最近的邻域内，则设置一条边。两个数据点之间的相似性作为边缘权重。选择 0-1 加权方案作为权重函数。

0-1 权重方案定义如下：

$$[W^P]_{ij} = \begin{cases} 1, if x_{:,i} \in N(x_{:,j}) or\ x_{:,j} \in N(x_{:,i}) \\ 0, \text{otherwise} \end{cases} \tag{5}$$

其中，$x_{:,j}$ 表示矩阵 X 的第 j 列，$N(x_{:,j})$ 表示 $x_{:,j}$ 的 k 个最近邻居节点集合。

数据图拉普拉斯算子矩阵是$L^P=D^P-W^P$,并且D^P是一个对角矩阵,并且$[D^P]_{lii}=\sum_j[W^P]_{ij}$,同样的,在特征空间中构造特征图,节点对应于特征集$\{X_1^T,\cdots,X_m^T\}$,高斯核函数的定义为:

$$[W^s]_{ij}=\begin{cases}\exp\left(-\frac{\|x_{:,i}-x_{:,j}\|}{2\sigma}\right), if x_{:,i}\in N(x_{:j}) or x_{:,j}\in N(x_{:,i})\\ 0 \quad , \text{otherwise}\end{cases} \tag{6}$$

其中,$x_{:,i}$表示矩阵X的第i行,$N(x_{:j})$表示特征$x_{:,i}$的k个最近邻居集合。

特征图的拉普拉斯算子矩阵是$L^S=D^S-W^S$,其中D^S是一个对角矩阵,并且

$$[D^P]_{lii}=\sum_j[W^P]_{ij}$$

上式中,W^Pij表示数据点X_i和X_j之间的相似度,并且W^Pij越高,意味着数据点X_i和X_j之间的相似度越高。将式(1)展开:

$$X_i=X_1S_{1i}+X_2S_{2i}+\cdots+X_nS_{ni}$$
$$X_j=X_1S_{1j}+X_2S_{2j}+\cdots+X_nS_{nj}$$

把S的第i列和第j列表示为

$$S_i=[S_{1i},S_{2i},\cdots,S_{ni}]^T$$
$$S_j=[S_{1j},S_{2j},\cdots,S_{nj}]^T$$

如果X_i和X_j具有较高的相似度,由于在邻近点可能具有相似的数据,可以得出这样的结论,W^Pij值越大,意味着S_i和S_j越接近,从而得到表示的光滑性为:

$$\begin{aligned}&\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n\|s_i-s_j\|2W_{ij}^P=\\&\sum_{i=1}^n s_i^Ts_iD_{ii}^P-\sum_{i=1}^n\sum_{j=1}^n s_i^Ts_iW_{ij}^P=\\&Tr(SD^PS^T)-Tr(SW^PS^T)=Tr(SL^PS^T)\end{aligned} \tag{7}$$

同样,考虑到数据空间的自表示矩阵P和相似矩阵W^S,可以得到以下的数据表示平滑度:

$$\begin{aligned}&\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^m\|p_i-p_j\|^2W_{ij}^s=\\&\sum_{i=1}^m p_i^Tp_iD_{ii}^s-\sum_{i=1}^m\sum_{j=1}^m p_i^Tp_iW_{ij}^s=\\&Tr(pD^Sp^T)-Tr(pW^Sp^T)=Tr(pL^Sp^T)\end{aligned} \tag{8}$$

在上述数据图和特征图的基础上,利用自表示特性,同时考虑数据空间和特征空间的多种信息,寻求一个在数据空间和特征空间中都具有的准确表示。因此,本算法解决以下最小化问题:

$$\begin{aligned}&\min\|X-XS\|_F^2+\beta\|X^T-X^TP\|_F^2+\\&\alpha_1Tr(SL^PS^T)+\alpha_2Tr(PL^SP^T)\\&s.t.\quad S,P\geq 0\end{aligned} \tag{9}$$

上式中,$\beta>0,\alpha_1>=0,\ \alpha_2>0$。为了简单和方便,让$\alpha_1=\alpha_1=\alpha$,这样,目标函数就可以改写为:

$$\begin{aligned}&\min\|X-XS\|_F^2+\beta\|X^T-X^TP\|_F^2+\\&\alpha_1[Tr(SL^PS^T)+\alpha_1Tr(PL^SP^T)]\\&s.t.\quad S,P\geq 0\end{aligned} \tag{10}$$

设$P=[P_{1\cdot},\cdots,P_{i\cdot};\cdots,P_{m\cdot}]$,点$P_i$是矩阵$P$的第$i$行,$\|P_{i.}\|_2$代表在自表示过程中对第$i$行的贡献,因此$\|P_{i.}\|_2$可以作为特征权重,对特征进行排序。为了避免过拟合解$P=I_m$(I_m是一个m维单位矩阵)并且确保稀疏性,充分运用矩阵P的$l_{2,1}-$范式,$l_{2,1}-$范式约束确保矩阵P是行稀疏的,所以$\|P_{i.}\|_2$反映了第i特性在整个特征集中的重要性。根据$\|P_{i.}\|_2$的大小,可以选择最重要的特征。

因此,本问题即是:

$$\begin{aligned}&\min\|X-XS\|_F^2+\alpha[Tr(SL^PS^T)+sTr(PL^PP^T)]+\lambda\|P\|_{2,1}\\&s.t.\quad S,P\geq 0\end{aligned} \tag{11}$$

参数λ(λ≥0)平衡最后稀疏的项目。此式子通过迭代的方式进行求解。

2.2 融入LDC(local density clustering)的聚类算法

根据预先设置的聚类数,将聚类分为若干部分,使同一类元素的相似性尽可能大,使不同类元素的相似性尽可能小。在上面的算法中,根据$\|P_i\|_2$的值按降序选取数据集X的前q个特征,从而在特征选取后得到矩阵$X_f=[X_1,X_2,\cdots,X_n]\in q\times n$。然后,对数据集$X_f$进行聚类。将数据集$X_f$划分为聚类簇,得到标记矩阵$M=[m_1,m_2,\cdots,m_n]^T\in\{0,1\}^{n\times c}$。$m_i$的第$j$个元素是$m_{ij}$,如果$X_i\in C_j$,则$m_{ij}=1$,否则为0。

首先定义一个矩阵Z:

$$Z=[z_1,z_2,\cdots,z_n]^T=M(M^TM)^{-1/2} \tag{12}$$

其中Z_i为数据向量习的聚类赋值向量。由于M^TM是一个对角矩阵,可以很容易地得出结论:

$$Z^TZ=(M^TM)^{-1/2}M^TM(M^TM)^{-1/2}=I_n \tag{13}$$

定义簇间散射矩阵S_b、簇内散射矩阵S_w和总散射矩阵S_t如下:

$$S_b = \tilde{X}_f Z Z^T \tilde{X}_f^T \tag{14}$$

$$S_w = \tilde{X}_f \tilde{X}_f^T - \tilde{X}_f Z Z^T \tilde{X}_f^T \tag{15}$$

$$S_t = S_b + S_w = \tilde{X}_f \tilde{X}_f^T \tag{16}$$

又因为 $\tilde{X}_f = X_f H_n$。定义一个中心矩阵 $H_n = \mathrm{I}_n - \frac{1}{n} 1_n 1_n^T \in \Re^{n\times n}$，其中 I_n 为 n 维单位矩阵，列向量 $1n$ 的所有元素为 1。为了更好地聚类，使不同聚类中的数据尽可能地与同一聚类中的数据尽可能地接近，利用 Fisher 准则，通过最大化以下目标函数，得到最优的聚类分配矩阵 Z_{best}：

$$Z_{\text{best}} = \arg\max_{Z} Tr[Z^T \tilde{X}_f^T (\tilde{X}_f \tilde{X}_f^T + \mu \mathrm{I})^{-1} \tilde{X}_f Z]$$

$$s.\ t. \quad \mu>0 \tag{17}$$

由于式(13)中 $Z^T Z = \mathrm{I}n$，可以通过最小化以下目标得到最佳的聚类分配矩阵 Z_{best} 功能：

$$Z_{\text{best}} = \arg\min_{Z} Tr[Z^T Z - Z^T \tilde{X}_f^{T-1} \tilde{X}_f Z]$$

$$s.\ t. \quad \mu > 0 \tag{18}$$

由于数据的局部流形信息类似于线性，采用由数据点 x_i 的 k 个最近邻组成的局部团 $N_k(x_i)$，其中包含数据点 x_i 本身及其$(k - 1)$个最近邻。因此，建立了一个局部线性判别模型来评价聚类结果。

定义一个矩阵 $X_i = [x_{i1}, x_{i2}, \ldots, x_{ik}]$，其中包含局部分 $N_k(x_i)$中的所有数据点，并将对应的索引集定义为 $Di = \{i_1, i_2, \ldots, i_k\}$。同样，根据局部线性判别模型，定义 $Z_1 = [z_{11}, z_{12}, \ldots, z_{1k}]^T \in k\times c$，由矩阵簇 Z 按照下式计算得到：

$$Z_i = A_i^T Z \tag{19}$$

上式中，A_i 是一个选择矩阵，而且 A_i 不是 1 就是 0。如式(16)，利用费雪准则，得到局部线性判别的最优聚类分配矩阵(Z_i)，通过最大化以下目标函数建立模型：

$$Z_{\text{best}} = \arg\max_{Z} Tr[Z_i^T \tilde{X}_i^T (\tilde{X}_i \tilde{X}_i^T \mu I)^{-1} \tilde{X}_i Z_i]$$

$$s.\ t. \quad \mu > 0 \tag{20}$$

上式中，$\tilde{X}_i = X_i H_k$，如式(17)，可以通过最小化目标函数得到局部线性判别模型的最优聚类分配矩阵(Z_i):

$$(Z_i)_{\text{best}} = \arg\min_{Z_i} Tr[Z_i^T H_k Z_i - Z_i^T \tilde{X}_i^T (\tilde{X}_i \tilde{X}_i^T + \mu I)^{-1} \tilde{X}_i Z_i]$$

$$s.\ t. \quad \mu > 0 \tag{21}$$

因为：

$$Z_i^T \tilde{X}_i^T (\tilde{X}_i \tilde{X}_i^T + \mu I)^{-1} \tilde{X}_i Z_i = Z_i^T [H_k - \mu H_k (\tilde{X}_i^T \tilde{X}_i + \mu I)^{-1} H_k] Z_i \tag{22}$$

因此，公式（21）就可以转化为：

$$(Z_i)_{\text{best}} = \arg\min_{Z_i} Tr[Z_i^T H_k (\tilde{X}_i^T \tilde{X}_i + \mu I)^{-1} H_k Z_i]$$

$$s.\ t. \quad \mu > 0 \tag{23}$$

把上式改为：

$$Z_{\text{best}} = \arg\min_{Z} Tr[Z_i^T L_i Z_i] \tag{24}$$

其中：

$$L_i = H_k (\tilde{X}_i^T \tilde{X}_i + \mu I)^{-1} H_k \tag{25}$$

为了得到全局最优的聚类分配矩阵 Z_{best}，需要对局部线性判别模型的每个最优聚类分配矩阵($\boldsymbol{Z}_i$)进行全局积分。将式(18)代入式(23)得到：

$$Z_{\text{best}} = \sum_{i=1}^{n} (Z_i)_{\text{best}} = \sum_{i=1}^{n} \arg\min_{Z_i} Tr[Z_i^T L_i Z_i] = \arg\min_{Z_i} Tr[Z^T (\sum_{i=1}^{n} A_i L_i A_i^T) Z^T] \tag{26}$$

定义

$$L = \sum_{i=1}^{n} A_i L_i A_i^T \tag{27}$$

那么上式就可以写作：

$$Z_{\text{best}} = \arg\min_{Z_i} Tr[Z^T L Z^T]$$

$$s.\ t. \quad Z^T Z = \mathrm{I}_{n,} = M(M^T M)^{-1/2} \tag{28}$$

利用谱旋转对 Z_{best} 进行离散。但是由于任何旋转矩阵都是正交矩阵，且 $\tilde{Z}_{\text{best}} = Z_{\text{best}} R$ 是式(26)的解，所以目标函数的解不是唯一的。因此，定义一个映射函数来求解最优的二值聚类分配矩阵 M_{best}，而不是最优的聚类分配矩阵 Z_{best}。

$$M_{\text{best}} = f^{-1}(Z_{\text{best}}) = diag(Z_{\text{best}} Z_{\text{best}}^T)^{-1/2} Z_{\text{best}} \tag{29}$$

其中，diag($Z_{\text{best}} Z_{\text{best}}^T$)是由 $Z_{\text{best}} Z_{\text{best}}^T$ 的对角元素组成的对角矩阵。根据[10]中的证据：

$f^{-1}(Z_{\text{best}} R) = M_{\text{best}} R$， $M_{\text{best}} R$ 是式(32)的最优解。通过连续迭代优化，可以同时得到二元值聚类分配矩阵 M 和旋转矩阵 R：

$$\arg\min_{M,R}\left\|M-M_{\text{best}}R\right\|^2$$

$$s.\ t.\quad M1_c=1_{n,}R^TR=I \tag{30}$$

2.3 具体算法流程

本文算法分为以下 10 个步骤：

Step1．加载数据矩阵 X。

Step2. 计算出数据矩阵的特征空间的非负矩阵因子 P，数据空间的非负矩阵因子 S，初始化矩阵 P，S，U。

Step3. 根据迭代更新规则（5），(9)和（10）更新矩阵 P，S 和 U。

Step 4.根据$\|P_i\|_2$的值按降序选取前 q 个特征。

Step 5. 构造每个数据点的局部小圈子。

Step 6. 根据式(24)计算 L_i，通过全局积分求解式(26)中的拉普拉斯矩阵 L。

Step 7. 用（28）中的特征值分解法求解特征向量，得到最优解 $Z_{\text{best}}=[u_1, u_2, \cdots, u_c]$。此外，应该注意删除平凡解 $\lambda_1=0$ 和 $u_1=1_n$。

Step 8. 利用谱旋转对最优解 Z_{best} 进行离散，利用式(29)求解最优二值聚类分配矩阵 M_{best}。

Step 9．通过连续迭代优化得到二值聚类分配矩阵 M。

Step 10．最后，通过评价指标计算出 ACC 和 NMI 输出结果。

3 实验及数据分析

3.1 数据集及参数设置

实验中验证了在 5 个数据集上，本文算法和其他特征选择算法的比较结果，并对结果进行了分析。5 个数据集分别为：UMIST 图像集，包括 20 个人共 575 幅图像，每个人具有不同角度、不同姿态的多幅图像；ORL 图像集包括 40 人，共 400 张面部图像；BC 数据集包括 2 类 569 个数据；Ionosphere 是一个经典的二元分类问题，共有 351 个观测值，34 个输入变量；Dbworld_bodies 包含 2 类 64 个数据。实验比较的算法包括 LapScore[8]、SPEC[9]、MCFS[10]、MRSF[11]、JELSR[12]、LSPE[13]。

表 1 数据集

Table 1 Data set

数据集	维度	大小	类数
UMIST	644	575	20
ORL	1024	400	40
BC	30	569	2
Ionosphere	34	351	2
Dbworld_bodies	4702	64	2

实验中参数的设置如下：LSPE、LapScore、SPEC 和 MCF，邻居数 k 选自{3,5,10,15}，σ 的内核权重选自{1,103,105}。本算法的正则参数 α 选自{0.01,0.1,0.5,1.0,5.0,9.0,13.0,17.0}，β 选自{100, 1000}，λ 选自{300,800,2000,4000,6000,8000}。常规参数 α 选自{0.01,0.1,0.5,1.0,5.0,9.0,13.0,17.0}，β 和 θ 选自{300,800,2000,4000,6000,8000}，μ 选自$\{10^{-8},10^{-6},10^{-4},10^{-2},10^{0},10^{2},10^{4},10^{6},10^{8}\}$。

3.2 评价指标

用两个广泛使用的评价矩阵来评价聚类的性能。聚类精度(ACC)，归一化互信息(NMI)。ACC 和 NMI 值越大，说明性能越好。

给定数据点 $\boldsymbol{x}_i$，c_i 和 g_i 分别为 $\boldsymbol{x}_i$ 的聚类标签和有效真实标签。

ACC 的定义如下：

$$\text{ACC}=\frac{\sum_{i=1}^{m}\delta(g_i,map(c_i))}{n} \tag{31}$$

其中 n 是数据的总数，$\delta(x, y)=\begin{cases}1, x=y\\0, \text{其它}\end{cases}$，其中，map 是使用匈牙利算法对聚类标签和真值标签进行排序的最优映射函数。

NMI 的定义如下：

$$\text{NMI}=\frac{MI(c,c')}{\max(H(c),H(c'))} \tag{32}$$

上式中，其中 c 和 c'分别为聚类标签和有效真实标签。$MI(c,c')$ 是 c 和 c'之间的信息熵，并且

$$MI(c,c')=\sum_{C_i\in C,C_j\in C}p(c_i,c_j)\cdot\log_2\frac{p(c_i,c_j)}{p(c_i)\cdot p(c_j)}$$

其中 $p(c_i)$ 和 $p(c_j)$ 分别表示聚类 c_i 和 c_j 的概率样本。$p(c_i,c_j')$ 是聚类 c_i 和 c_j' 的联合概率。ACC 是基于聚类标签与有效真值标签之间的一对一匹配。

NMI 是评价聚类标签之间的相似性程度的一种外部准则，它评估了聚类标签和基本真实标签之

间的相似程度。ACC 和 NMI 是两个聚类评估标准，它们可能无法同时在一个数据集中获得最佳效果。

3.3 实验结果

实验记录了最优参数的最佳聚类结果。表 2 显示了 5 个数据集的平均 ACC 和标准偏差，用粗体突出显示最佳结果。

3.3.1 本文算法实验结果

表 2 相关算法 ACC

Table 2 ACC of related algorithms

算法	Umist	ORL	Ionosphere	BC	Dbworld_bodies
LapScore	37.30 ± 0.93	44.50 ± 0.73	66.94 ± 2.20	70.17 ± 0.36	73.47 ± 1.16
SPEC	42.56 ± 1.20	49.88 ± 0.23	67.70 ± 2.33	74.00 ± 0.23	77.94 ± 1.85
MCFS	46.55 ± 1.00	49.40 ± 0.93	57.26 ± 3.00	71.00 ± 0.58	91.13 ± 1.04
JELSR	48.90 ± 1.03	50.02 ± 0.56	67.90 ± 2.81	74.20 ± 0.30	90.63 ± 0.00
MRSF	48.38 ± 1.05	49.78 ± 0.69	63.00 ± 2.30	72.9 ± 0.22	85.02 ± 1.59
LSPE	49.26 ± 1.12	50.25 ± 0.80	70.00 ± 2.66	75.86 ± 0.24	93.75 ± 0.00
本文算法	61.21 ± 2.25	60.71 ± 1.51	88.79 ± 2.37	75.55 ± 0.00	95.68 ± 0.014

表 3 显示了 5 个数据集的 NMI，用粗体突出显示最佳结果。

表 3 相关算法 NMI

Table 3 NMI of related algorithms

算法	Umist	ORL	Ionosphere	BC	Dbworld_bodies
LapScore	56.32 ± 1.52	67.80 ± 1.76	8.16 ± 0.00	16.79 ± 0.00	23.82 ± 1.01
SPEC	57.04 ± 1.24	70.26 ± 1.65	8.33 ± 0.00	18.83 ± 0.00	25.20 ± 1.62
MCFS	69.20 ± 1.31	70.98 ± 1.78	1.01 ± 0.77	17.32 ± 0.00	67.88 ± 1.62
JELSR	70.18 ± 1.64	70.20 ± 1.72	7.84 ± 1.21	18.86 ± 0.00	54.89 ± 0.00
MRSF	66.67 ± 1.43	70.50 ± 1.81	3.82 ± 0.00	17.32 ± 0.00	56.79 ± 2.39
LSPE	70.01 ± 1.50	71.04 ± 1.11	13.10 ± 0.49	18.83 ± 0.00	68.09 ± 0.00
本文算法	70.51 ± 1.52	75.25 ± 1.33	23.12 ± 0.57	30.99 ± 0.00	59.99 ± 1.27

从表 2 和表 3 中，可以看到：本文算法优于其他算法，在几乎所有数据集的聚类精度方面都取得了较好的结果，尤其是在处理面部数据的时候。与其他特征选择算法相比，本算法的主要改进之处在于它利用了特征空间的自表示特性，特征空间中的信息对于聚类是非常重要的。SPEC, MCFS 和 MRSF 是两阶段的特征选择算法，而 JELSR，LSPE 和本文算法同时集成多种策略。JELSR 将嵌入式学习和稀疏回归相结合，LSPE 将嵌入式学习和特征选择相结合，本算法将自表示、流形嵌入和特征选择相结合。总体上，JELSR、LSPE 的聚类质量优于其他算法，说明同时求解多个问题优于按顺序求解。LSPE 是实验中第二好的算法，也说明了将嵌入学习和特征选择结合起来进行特征选择是一种较好的方法。

3.3.2 没有加入局部判别的算法效果

为了更清楚的体现局部判别的对算法结果的影响，在不加入局部判别的情况做了相关实验，表 4 显示了 5 个数据集 ACC 结果，表 5 显示了 5 个数据集 NMI 结果。

表 4 相关算法 ACC（无局部判别）

Table 4 ACC of related algorithms （without local discrimination）

算法	Umist	ORL	Ionosphere	BC	Dbworld_bodies
LapScore	37.30 ± 0.93	44.50 ± 0.73	66.94 ± 2.20	70.17 ± 0.36	73.47±1.16
SPEC	42.56 ± 1.20	49.88 ± 0.23	67.70 ± 2.33	74.00 ± 0.23	77.94±1.85
MCFS	46.55 ± 1.00	49.40 ± 0.93	57.26 ± 3.00	71.00 ± 0.58	91.13±1.04
JELSR	48.90 ± 1.03	50.02 ± 0.56	67.90 ± 2.81	74.20 ± 0.30	90.63±0.00
MRSF	48.38 ± 1.05	49.78 ± 0.69	63.00 ± 2.30	72.9 ± 0.22	85.02±1.59
LSPE	49.26 ± 1.12	50.25 ± 0.80	70.00 ± 2.66	75.86 ± 0.24	93.75±0.00
本文算法	58.68 ± 2.69	57.64 ± 1.84	85.45 ± 2.58	75.24 ± 0.09	94.70±.021

表 5 相关算法 NMI（无局部判别）

Table 5 NMI of related algorithms（without local discrimination）

算法	Umist	ORL	Ionosphere	BC	Dbworld_bodies
LapScore	56.32 ± 1.52	67.80 ± 1.76	8.16 ± 0.00	16.79 ± 0.00	23.82 ± 1.01
SPEC	57.04 ± 1.24	70.26 ± 1.65	8.33 ± 0.00	18.83 ± 0.00	25.20 ± 1.62
MCFS	69.20 ± 1.31	70.98 ± 1.78	1.01 ± 0.77	17.32 ± 0.00	67.88 ± 1.62
JELSR	70.18 ± 1.64	70.20 ± 1.72	7.84 ± 1.21	18.86 ± 0.00	54.89 ± 0.00
MRSF	66.67 ± 1.43	70.50 ± 1.81	3.82 ± 0.00	17.32 ± 0.00	56.79 ± 2.39
LSPE	70.01 ± 1.50	71.04 ± 1.11	13.10 ± 0.49	18.83 ± 0.00	68.09 ± 0.00
本文算法	70.36 ± 1.65	74.14 ± 1.45	21.76 ± 0.73	30.13 ± 0.08	58.24 ± 1.38

从表 4 和表 5 的结果可以看出，本文提出的算法即使在没有局部判别信息的情况下，总体效果也优于相关算法。但与加入局部判别信息的本最终算法比较，效果略微差一些，方差变大。因此，说明了加入局部判别，可以提高聚类的有效性和鲁棒性。

4 结论

本研究提出了一种自表示的双图规格化特征选择聚类算法，不仅保留了数据的流形信息，而且保留了特征空间的流形信息。此外，加入局部判别特征选择聚类，可以大大提高聚类的有效性和鲁棒性。

参考文献：

[1] 机器学习中的特征选择方法研究及展望[J].北京邮电大学学报, 2018(1):1-12.

[2] 谢娟英,丁丽娟,王明钊.基于谱聚类的无监督特征选择算法[J].软件学报,2020,56(1):20-32.

[3] Meng Y,Shang R,Jiao L,et al.Feature selection based dual-graph sparse non-negative matrix factorization for local discriminative clustering[J].Neurocomputing,2018，32(3)：1783-1796.

[4] Shang R,Wang W,Stolkin R,et al.Non-Negative Spectral Learning and Sparse Regression-Based Dual-Graph Regularized Feature Selection[J]. IEEE Transactions on Cybernetics,2017,45(6):1-14.

[5] Meng Y,Shang R,Jiao L,et al.Dual-graph regularized non-negative matrix factorization with sparse and orthogonal constraints[J].Engineering Applications of Artificial Intelligence,2018, 69:24-35.

[6] Hou C, Nie F, Li X, et al. Wu,Joint embedding learning and sparse regression:a framework for unsupervised feature selection[J]. IEEE Trans on Cybernetics, 2014,44 (6):793-804.

[7] Zhu X Y. Improved Algorithm Based on Non-negative Low Rank and Sparse Graph for Semi-supervised Learning[J].Journal of Electronics & Information Technology,2017,10(4):1123-1128.

[8] Tian S,Zhang Y,Yan Y.$\ell_{1/2}$-norm regularized nonnegative low-rank and sparse affinity graph for remote sensing image segmentation[J].Journal of Applied Remote Sensing,2016,10(4):042008.

[9] Zhao Z, Wang L, Liu H. Efficient spectral feature selection with minimum redundancy[C]. Proceedings of 24th AAAI Conference on Artificial Intelligence, 2010:673-678.

[10] Zhao Z, Liu H.Spectral feature selection for supervised and unsupervised learning[C]. Machine Learning, Proceedings of the Twenty-Fourth International Conference (ICML 2007),2007.

[11] Cai D, Zhang C, He X. Unsupervised feature selection for multi-cluster data[C]. Acm Sigkdd International Conference on Knowledge Discovery & Data Mining, ACM, 2010.

[12] Li Z,Tang J.Unsupervised Feature Selection via Nonnegative Spectral Analysis and Redundancy Control[J].IEEE Transactions on Image Processing,2015, 24(12):5343-5355.

[13] Fang Y,Xu X,Li Z.Locality and similarity preserving embedding for feature selection[J]. Neurocomputing, 2014,128:304-315.

[14] Zhu X Y, Wang Y, Li Y B. A new unsupervised feature selection algorithm using similarity-based feature clustering.[J] Computational Intelligence,2019, 47(1):161-173.

[15] 方佳艳,刘峤.具有同步化特征选择的迭代紧凑非平行支持向量聚类算法[J].电子学报,2020，56(1):20-32.

[16] Yang Y, Xu D, Nie F, et al. Image clustering using local discriminant models and global integration[J].IEEE Trans. Image Process. 2010,19(10): 2761-2773.

[17] Wu S S, Hou M X, Feng C M. LJELSR: A Strengthened Version of JELSR for Feature Selection and Clustering[J]. International journal of molecular sciences,2018, 22(8):888-905.

[18] Ahmad AliyuUsman,Starkey Andrew.Application of feature selection methods for automated clustering analysis:a review on synthetic datasets[J].Neural computing and applications,2018,56(3):115-119.