



计算机科学

COMPUTER SCIENCE

基于相似度矩阵学习和矩阵校正的无监督多视角特征选择

李斌, 万源

引用本文

李斌, 万源. 基于相似度矩阵学习和矩阵校正的无监督多视角特征选择[J]. 计算机科学, 2022, 49(8): 86-96.

LI Bin, WAN Yuan. [Unsupervised Multi-view Feature Selection Based on Similarity Matrix Learning and Matrix Alignment](#)[J]. Computer Science, 2022, 49(8): 86-96.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[一种用于癌症分类的两阶段深度特征选择提取算法](#)

Two-stage Deep Feature Selection Extraction Algorithm for Cancer Classification

计算机科学, 2022, 49(7): 73-78. <https://doi.org/10.11896/jsjcx.210500092>

[混合改进的花授粉算法与灰狼算法用于特征选择](#)

Hybrid Improved Flower Pollination Algorithm and Gray Wolf Algorithm for Feature Selection

计算机科学, 2022, 49(6A): 125-132. <https://doi.org/10.11896/jsjcx.210600135>

[三维城市场景中的小物体检测](#)

Small Object Detection in 3D Urban Scenes

计算机科学, 2022, 49(6): 238-244. <https://doi.org/10.11896/jsjcx.210400174>

[基于灰狼优化算法的信用评估样本均衡化与特征选择同步处理](#)

Application of Gray Wolf Optimization Algorithm on Synchronous Processing of Sample Equalization and Feature Selection in Credit Evaluation

计算机科学, 2022, 49(4): 134-139. <https://doi.org/10.11896/jsjcx.210300075>

[基于邻域粗糙集和 Relief 的弱标记特征选择方法](#)

Weak Label Feature Selection Method Based on Neighborhood Rough Sets and Relief

计算机科学, 2022, 49(4): 152-160. <https://doi.org/10.11896/jsjcx.210300094>

基于相似度矩阵学习和矩阵校正的无监督多视角特征选择

李 斌 万 源

武汉理工大学理学院 武汉 430070

(2859713954@qq.com)

摘 要 多视角特征选择通过融合多个视角的信息获取具有代表性的特征子集,来提高分类、聚类学习任务效率。然而,描述对象的特征繁杂多样且相互关联,单一地从原始特征中选择特征子空间可以简单地解决维度问题,但无法有效获取数据内部存在的结构信息和特征关联信息,且固定使用相似度矩阵和投影矩阵易损失视角间的相关性。针对以上问题,提出了基于相似度矩阵学习和矩阵校正的无监督多视角特征选择(SMLMA)算法。该算法首先构造所有视角的相似度矩阵,通过流形学习得到一致相似度矩阵以及投影矩阵,最大程度地发现和保留多视角数据的结构信息;其次采用矩阵校正的方法,最大化相似度矩阵和核矩阵之间的相关性,合理利用不同视角之间的关联性,减少特征子集的信息冗余;最后,采用 Armijo 搜索方法快速得到收敛结果。在 4 个实验数据集 Caltech-7, NUS-WIDE-OBJ, Toy Animal 和 MSRC-v1 上的实验结果表明,相比单视角特征选择和部分多视角特征选择方法,所提算法在聚类任务上的准确率平均提高了约 7.54%。其较好地保留了数据的结构信息和多视角之间特征的相关性,捕获了更多高质量的特征。

关键词: 多视角; 无监督; 特征选择; 相似度矩阵; 矩阵校正

中图法分类号 TP181

Unsupervised Multi-view Feature Selection Based on Similarity Matrix Learning and Matrix Alignment

LI Bin and WAN Yuan

School of Science, Wuhan University of Technology, Wuhan 430070, China

Abstract Multi-view feature selection improves the efficiency of classification, clustering and other learning tasks by fusing information from multiple views to obtain representative feature subsets. However, the features of different views that describe objects are complex and interrelated. Simply searching subset of features from original space partly solves the problem of dimension, but it barely obtains the latent structural information and association information among features. Besides, using fixed similarity matrix and projection matrix is prone to lose the correlation between different views. To solve these problems, an unsupervised multi-view feature selection algorithm based on similarity matrix learning and matrix alignment(SMLMA) is proposed. Firstly, the similarity matrix based on all views is constructed, and the consistent similarity matrix and projection matrix are obtained by manifold learning, to explore and reserve the structural information of data to the greatest extent. Then, the matrix alignment method is used to maximize the correlation between the similarity matrix and the kernel matrix, for the purpose of using the correlation between different views and reducing the information redundancy of feature subset. Finally, the Armijo searching method is introduced to obtain the convergence result quickly. Experimental results on four datasets(Caltech-7, NUS-WIDE-OBJ, Toy Animal and MSRC-v1) show that, compared with single view feature selection and some multi-view feature selection methods, the accuracy of SMLMA is averagely improved by about 7.54%. The proposed algorithm well retains the structural information of data and the correlation between multi-view features, and captures more high-quality features.

Keywords Multi-view, Unsupervised, Feature selection, Similarity matrix, Matrix alignment

1 引言

在机器学习、图像处理和生物医学科学领域的许多应用中,

数据往往具有多个视角的特征,这些特征是从不同的数据源获得或者由不同的特征集来共同表示。例如,在图像处理中,每幅图像都可以用不同的视觉描述符来描述,如灰度值、尺度

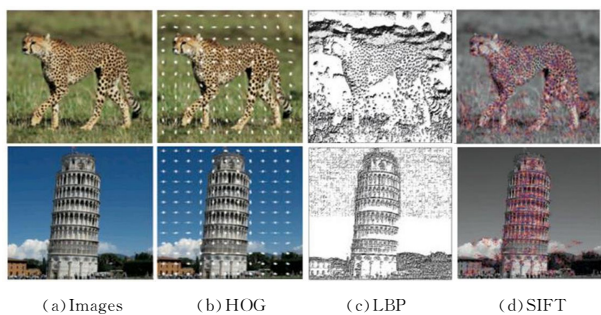
到稿日期:2021-07-13 返修日期:2022-02-27

基金项目:中央高校基本科研业务费专项资金(2021III030JC)

This work was supported by the Fundamental Research Funds for the Central Universities(2021III030JC).

通信作者:万源(wanyuan@whut.edu.cn)

不变特征变换 (Scale-Invariant Feature Transform, SIFT) 特征^[1]、方向梯度直方图 (Histogram of Oriented Gradients, HOG) 特征^[2]、局部二值模式 (Local Binary Patterns, LBP) 特征^[3]等,视觉描述符捕获的特定特征如图 1 所示。在生物医学研究中,一个蛋白质可以用它在不同细胞中的序列和基因表达值来表示每个视角的特征表示在一个特定的特征空间的数据^[4]。多视角数据的各个视角之间通常能够提供互补和相关的信息,旨在利用不同视角之间的多样性和相关性来更精确地描述数据^[5]。考虑到多视角数据的多特征特点,多视角学习已经成为机器学习的一个特定主题^[6-7],包括多视角分类^[8]、多视角聚类^[9]、多视角检索^[9]等。除此之外,多视角数据包含所有特征,其维度较高且包含噪声和无关的特征,直接利用它们会产生巨大的存储成本。因此,多视角特征选择^[10-11]对多视角学习具有重要意义。



注:从最左边原始图像中提取 3 种类型的特征,分别使用 HOG、LBP 和 SIFT 描述符

图 1 多视角图像特征的可视化

Fig. 1 Visualizations of multi-view image features

在各种应用中采集的数据多为无标签数据,而无监督学习在没有标签的前提下对样本进行训练,推动了特征选择研究的发展。无监督的多视角特征选择利用无标签的训练样本学习特征子空间,旨在减少多视角数据的维数,选择一个具有代表性的特征子集,准确地保留数据的内在结构^[12-13]。现有的无监督多视角特征选择方法大致可以分为两类。第一类方法是直接将多个视角的所有特征组合成一个向量,然后采用传统的单视角特征选择方法,如拉普拉斯得分法 (Laplacian Score)^[14]、光谱特征选择 (Spectral Feature Selection, SPEC)^[15]和最小冗余光谱特征选择 (Minimum Redundancy Spectral Feature Selection, MRSF)^[16]。这些方法虽然解决了多视角特征选择中的一些问题,但是没有考虑不同视角之间的相关性。

为了更好地利用原始数据的潜在信息,基于流形学习的方法在近年来的研究中得到了较广泛的应用,经典的流形学习方法有局部保持投影 (Locality Preserving Projection, LPP)^[17]、拉普拉斯嵌入 (Laplacian Embedding, LE)^[18]和局部线性嵌入 (Locally Linear Embedding, LLE)^[19],旨在将数据点之间的邻近关系嵌入低维子空间中,保留数据的原始结构。流形学习在保留数据结构方面具有很大的优势,并被广泛应用于许多特征选择方法。Du 等提出一种基于自适应结构学习的无监督特征选择 (Unsupervised Feature Selection with Adaptive Structure Learning, FSASL)^[20]方法,通过稀疏

重构来保留数据的原始结构,将各个视角的数据信息映射到一个最优稀疏组合权重矩阵,自适应地保持流形结构。Feng 等提出自适应无监督多视角特征选择 (Adaptive Unsupervised Multi-view Feature Selection, AUMFS)^[21]方法,通过利用数据集簇结构、数据相似性和不同视角之间的相关性来进行特征选择,对投影矩阵引入 $l_{2,1}$ -范数惩罚,得到具有稀疏性的特征选择矩阵。Tang 等提出无监督的多视角特征选择 (Unsupervised Multi-view Feature Selection, MVFS)^[22]方法,在 AUMFS 方法的基础上,改进了平衡参数选择和稀疏回归的损失函数。Wang 等提出自适应多视角特征选择 (Adaptive Multi-view Feature Selection, AMFS)^[23]方法,使用局部线性回归模型自适应地学习多个基于不同视角的 Laplacian 图,以保留原始数据的局部几何结构。通过将这些 Laplacian 图与非负权重向量相结合,利用不同视角之间的互补信息,提高了特征子集的质量。Hou 等提出自适应相似度和视角权重的多视角特征选择 (Multi-view Feature Selection with Adaptive Similarity and View Weight, ASVW)^[24]方法,通过学习不同视角的公共相似度矩阵,描述各个视角的公共结构,使用稀疏 $l_{2,1}$ -范数约束来学习具有稀疏性的特征选择矩阵。Shao 等提出在线无监督特征选择算法 (Online Unsupervised Multi-view Feature Selection, OMVFS)^[25],通过非负矩阵分解和稀疏学习将无监督特征选择嵌入聚类算法中。其结合图形正则化保存了数据局部结构的信息,可以帮助选择过程选择区分性特征。

对于特征选择,这些多视角方法在一定程度上提高了特征选择的效率,但仍然存在一些缺点:1)多视角数据包含所有视角的信息,但不同视角的投影矩阵并不相同。现有的方法大多引入预定义的相似度矩阵,并通过固定它来引导各个视角的特征选择,可能会损失各个视角之间的相关性;2)有些方法直接使用相似度矩阵和特征选择矩阵,可能没有考虑到它们之间的相关性,由此得到的特征子集可能会缺失一些相关的和有利的特征,如 AMFS^[23]和 ASVW^[24]。

针对以上问题,本文提出了一种基于相似度矩阵学习和矩阵校正的无监督多视角特征选择 (Unsupervised Multi-view Feature Selection based on Similarity Matrix Learning and Matrix Alignment, SMLMA)方法,将特征选择嵌入数据流形结构图的学习^[5,26-28]中,两者相互促进。在对相似度矩阵的学习中进行特征选择,既能最大程度地保留原始的数据结构,还能剔除一些不重要的特征,最后再将学习到的相似度矩阵进行校正,使用投影矩阵构造的核矩阵与其进行矩阵差异最小化处理,最大化两者的相关性,将数据结构信息和特征选择进行关联,获得有效且高质量的特征子集。实验结果表明,所提方法可以有效减少信息冗余,较好地保留了特征相关性。

2 相关工作

本节回顾一些无监督多视角特征选择中的代表性工作。

数据大多是由多传感器采集,具有不同的表现形式,使得多视角学习成为一个备受关注的研究课题。当遇到多视角数据时,找到平衡各个视角特征的最佳方法是学习多视角数据的关键。与直接将不同类型的特征串联,使得一个多视角

数据转化为单视角数据的思想不同,利用多视角之间的相关性和互补性信息可以有效地提升聚类性能。

对于自适应全局结构的学习,FSASL^[20]方法采用最优稀疏组合权重矩阵,在重构过程中引入行稀疏特征选择,得到特征选择矩阵 \mathbf{W} 。AUMFS^[21]将数据聚类标签预测和自适应多视角视觉相似图的学习集成到一个统一的框架中,使用 3 种信息,即数据相似性、数据聚类结构和不同视角之间的相关性。假设给定有 V 个视角的多视角数据 $\mathbf{X} = [(\mathbf{X}^{(1)})^T, (\mathbf{X}^{(2)})^T, \dots, (\mathbf{X}^{(V)})^T]^T \in \mathbf{R}^{d \times n}$, $\mathbf{X}^{(v)}$ 表示第 v 个视角的数据,通过对不同视角的拉普拉斯矩阵进行线性组合来保持图的局部结构。它采用每一行向量的 $l_{2,1}$ -范数作为标准来对每个特征的重要性进行排序^[29]。目标函数如下:

$$\begin{aligned} \min_{\mathbf{F}, \boldsymbol{\lambda}, \mathbf{W}} \operatorname{tr}(\mathbf{F}^T \sum_{v=1}^m \lambda_v \mathbf{L}^{(v)} \mathbf{F}) + \alpha \|\mathbf{X}^T \mathbf{W} - \mathbf{F}\|_{2,1} + \beta \|\mathbf{W}\|_{2,1} \\ \text{s. t. } \mathbf{F}^T \mathbf{F} = \mathbf{I}_c, \mathbf{F} \geq 0, \sum_{v=1}^m \lambda_v = 1, \lambda_v \geq 0 \end{aligned} \quad (1)$$

其中, $\mathbf{X} = [(\mathbf{X}^{(1)})^T, (\mathbf{X}^{(2)})^T, \dots, (\mathbf{X}^{(V)})^T]^T \in \mathbf{R}^{d \times n}$, $\mathbf{X}^{(v)}$ 表示第 v 个视角的数据; \mathbf{F} 是伪标签向量; $\boldsymbol{\lambda}$ 是权重参数向量; $\mathbf{L}^{(v)}$ 是第 v 个视角的拉普拉斯矩阵,且 $\mathbf{L}^{(v)} = \mathbf{D}^{(v)} - \mathbf{A}^{(v)}$; $\mathbf{D}^{(v)}$ 是一个对角矩阵,对角线元素 $D_{ii}^{(v)} = \sum_{j=1}^n \mathbf{A}_{ij}^{(v)}$; \mathbf{A} 是相似度矩阵; \mathbf{W} 是投影矩阵; r 是平衡参数; α 和 β 是正则化参数。

ASVW 在 AUMFS 方法的基础上取得了进一步的进展,假设不同视角之间的所有相似度矩阵是一致的,利用参数平衡不同视角的重要性,使用投影矩阵作为嵌入矩阵,并添加稀疏 $l_{2,1}$ -范数约束,使得投影矩阵具有稀疏的结构。为了得到相似度矩阵,引入局部几何描述符来表示第 v 个视角的数据的相似度。然后在局部结构中使用正则化,避免得到平凡解。目标函数可归纳如下:

$$\begin{aligned} \min_{\boldsymbol{\alpha}, \mathbf{W}, \mathbf{S}} \sum_{v=1}^V \sum_{i=1}^n \sum_{j=1}^n (\boldsymbol{\alpha}_v^T \|\mathbf{W}_v^T \mathbf{x}_i^{(v)} - \mathbf{W}_v^T \mathbf{x}_j^{(v)}\|_{2,1} \mathbf{S}_{ij}^{(v)} + \\ \lambda \|\mathbf{S}\|_{2,1}) + \gamma \sum_{v=1}^V \|\mathbf{W}_v\|_{2,1} \\ \text{s. t. } \mathbf{W}_v^T \mathbf{W}_v = \mathbf{I}; \sum_{v=1}^V \boldsymbol{\alpha}_v = 1, \boldsymbol{\alpha}_v \geq 0; \sum_{j=1}^n \mathbf{S}_{ij}^{(v)} = 1, \mathbf{S}_{ij}^{(v)} \geq 0; \\ \|\mathbf{S}_{i \cdot}^{(v)}\|_0 = k \end{aligned} \quad (2)$$

其中, \mathbf{W}_v 是第 v 个视角的权重矩阵; $\boldsymbol{\alpha}_v$ 为第 v 个视角的权重值; $\mathbf{S}_{ij}^{(v)}$ 是数据之间的局部相似性; \mathbf{I} 是一个单位矩阵; λ 和 γ 是正则化参数; r_1 是平衡参数,且 $r_1 > 1$ 。

3 本文方法

ASVW 方法在学习相似度矩阵的过程中嵌入了特征选择,但投影矩阵 \mathbf{W}_v , 即特征选择矩阵,是各个独立视角下的训练结果,它没有考虑各视角投影矩阵的不同,数据结构的信息和选择特征之间的一致性有所不足。该方法得到的特征子集质量不高,无法避免噪声且存在部分冗余的特征^[30]。针对以上问题,本文考虑将得到的相似度矩阵进行校正,最大化相似度矩阵与投影矩阵构造的核矩阵之间的相关性。该过程通过学习多视角数据的一致相似图,并根据图的学习来指导稀疏回归,进行特征选择,因此可保留更多数据内部结构的信息,具体流程如图 2 所示。

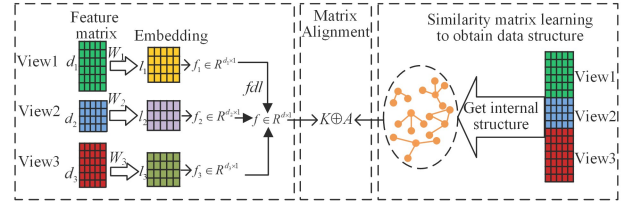


图 2 SMLMA 流程

Fig. 2 Flow chart of SMLMA

3.1 相似度矩阵学习

给定有 M 个视角的多视角数据 $\mathbf{X} = [(\mathbf{X}^{(1)})^T, (\mathbf{X}^{(2)})^T, \dots, (\mathbf{X}^{(M)})^T]^T \in \mathbf{R}^{d \times n}$, $\mathbf{X}^{(m)}$ 表示第 m 个视角的数据,为学习所有视角中数据点的一致相似度矩阵,假设各个视角的相似度矩阵 \mathbf{A} 是一致的。初始化相似度矩阵使用 k 最近邻方法来构造,相似度矩阵元素定义如下:

$$\mathbf{A}_{ij} = \begin{cases} 1, & \text{如果 } x_i \text{ 是 } x_j \text{ 的最近邻或者 } x_j \text{ 是} \\ & x_i \text{ 的最近邻} \\ 0, & \text{其他} \end{cases} \quad (3)$$

为了结合每个视角的重要信息,组合各个视角信息时使用权重值 $\boldsymbol{\alpha}_m$,对该值加上幂 r_1 是为了避免出现平凡解 $\boldsymbol{\alpha}_m = 0$ 。对一致矩阵 \mathbf{A} 加上幂 r_2 同样是为了避免平凡解,但是由于该矩阵最终是在所有数据中训练得到的,一致相似度矩阵 \mathbf{A} 的每一行对应一个数据点^[5,29],因此需要加上以下约束:

$$\sum_{j=1}^n \mathbf{A}_{ij} = 1, \mathbf{A}_{ij} \geq 0 \quad (4)$$

其中, $\mathbf{A}_{i \cdot}$ 表示第 i 行任意列的元素,约束是非负。该约束保证了训练得到的一致相似度矩阵 \mathbf{A} 和初始相似度矩阵的一致性。

由于相似度矩阵是通过 K -NN 方法构造的图得到的,其保存了数据结构的基本信息^[31],因此,将相似度矩阵 \mathbf{A} 和投影矩阵 \mathbf{W} 的学习同步进行,得到满足需求的两个矩阵 \mathbf{A} 和 \mathbf{W} 。具体目标函数如下:

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{A}, \boldsymbol{\alpha}} \sum_{m=1}^M \sum_{i=1}^n \sum_{j=1}^n \boldsymbol{\alpha}_m^r \|\mathbf{W}_m^T \mathbf{x}_i^{(m)} - \mathbf{W}_m^T \mathbf{x}_j^{(m)}\|_{2,1} (\mathbf{A}_{ij}^{(m)})^{r_2} \\ \text{s. t. } \mathbf{W}_m^T \mathbf{W}_m = \mathbf{I}; \sum_{m=1}^M \boldsymbol{\alpha}_m = 1, \boldsymbol{\alpha}_m \geq 0 \\ \sum_{j=1}^n \mathbf{A}_{ij} = 1, \mathbf{A}_{ij} \geq 0, \|\mathbf{A}_{i \cdot}\|_0 = k; \mathbf{A}_{ij}^{(m)} = \mathbf{A}_{ij}, \\ m = 1, 2, \dots, M \end{aligned} \quad (5)$$

其中, \mathbf{W}_m 是第 m 个视角的投影矩阵; $\mathbf{A} \in \mathbf{R}^{n \times n}$ 和 $\mathbf{A}^{(m)}$ 是一致相似度矩阵; k 表示构造 K -NN 图时的邻域个数; $\|\cdot\|_0$ 表示 l_0 范数,指非 0 元素的个数; $\boldsymbol{\alpha}_m$ 为第 m 个视角的权重参数, $\boldsymbol{\alpha} \in \mathbf{R}^{M \times 1}$ 是权重向量; $r_1 = r_2$ 是平衡参数,且 $r > 1$ 。该项学习一致相似度矩阵,更好地保留了数据的流形结构,使得高维空间的数据结构投影到低维空间时信息损失最小,高维空间中相近的数据点投影到低维空间时仍然相近。

3.2 矩阵校正

考虑融合数据的结构信息和特征信息,对相似度矩阵和核矩阵进行校正,将相似度矩阵和核矩阵之间的相关性最大化,得到具有结构信息和特征选择信息的特征选择向量^[32],进而较大程度地联合数据的结构信息和特征选择信息来筛选出最具代表性的特征,得到高质量的特征子集。

3.2.1 矩阵校正的定义

定义 1(矩阵校正) 对于两个方阵 $\mathbf{A}_1, \mathbf{A}_2 \in \mathbf{R}^{n \times n}$, $\operatorname{tr}(\cdot)$ 表示矩阵的迹,两个矩阵校正的定义为:

$$\rho(\mathbf{A}_1, \mathbf{A}_2) = \frac{\text{tr}(\mathbf{A}_1 \mathbf{A}_2)}{\|\mathbf{A}_1\|_F \cdot \|\mathbf{A}_2\|_F} \quad (6)$$

由于标准化项 $\|\mathbf{A}_1\|_F \cdot \|\mathbf{A}_2\|_F$ 使优化问题更难求解,可采用矩阵校正的非标准化形式,即 $\rho(\mathbf{A}_1, \mathbf{A}_2) = \text{tr}(\mathbf{H}\mathbf{A}_1\mathbf{A}_2)$ 。通常,为减小误差带来的影响,把矩阵 \mathbf{A}_1 和 \mathbf{A}_2 作中心化处理,得到 $\mathbf{H}\mathbf{A}_1\mathbf{H}, \mathbf{H}\mathbf{A}_2\mathbf{H}$, 其中中心化矩阵 $\mathbf{H} = \mathbf{I} - \frac{1}{n} \mathbf{1}\mathbf{1}^T$ 且 $\mathbf{H}\mathbf{H} = \mathbf{H}$ 。

定义 2(中心化矩阵校正) 对于两个方阵 $\mathbf{A}_1, \mathbf{A}_2 \in \mathbf{R}^{n \times n}$, $\text{tr}(\cdot)$ 表示矩阵的迹,且 $\text{tr}(\mathbf{A}\mathbf{B}) = \text{tr}(\mathbf{B}\mathbf{A})$, 则中心化矩阵校正的定义如式(7)所示:

$$\rho(\mathbf{A}_1, \mathbf{A}_2) = \text{tr}(\mathbf{H}\mathbf{A}_1\mathbf{H}\mathbf{H}\mathbf{A}_2\mathbf{H}) = \text{tr}(\mathbf{H}\mathbf{A}_1\mathbf{H}\mathbf{A}_2) \quad (7)$$

两个矩阵校正的结果表示两个矩阵之间的相关性,值越大则相关性越大;反之,相关性越小。

3.2.2 矩阵校正

由 3.1 节训练得到的投影矩阵 \mathbf{W}_m 构造第 m 个视角的初始特征选择向量 $\mathbf{f}^{(m)} \in \{0, 1\}$, 构造过程使用特征评分准则,通过计算 $\|\mathbf{W}_i^{(m)}\|_2$ 选取得分排名前 k 位的特征,并且标记 $\mathbf{f}^{(m)}$ 对应的元素为 1。构造核矩阵元素的表达式如下:

$$\mathbf{K}_{ij}^{(m)} = \exp\left(-\frac{1}{\sigma^2} \|\text{diag}(\mathbf{f}^{(m)})\mathbf{x}_i^{(m)} - \text{diag}(\mathbf{f}^{(m)})\mathbf{x}_j^{(m)}\|_2^2\right) \quad (8)$$

其中, $\mathbf{f}^{(m)} \in \mathbf{R}^{d_m \times 1}$ 为第 m 个视角的特征选择向量; $\mathbf{f}_t^{(m)} \in \{0, 1\}$, $m=1, 2, \dots, M$ 是第 m 个视角的第 t 个特征,满足 $\sum_{t=1}^{d_m} \mathbf{f}_t^{(m)} = s^{(m)}$; $s^{(m)}$ 是第 m 个视角选择的特征数目; $\text{diag}(\mathbf{f}^{(m)}) \in \mathbf{R}^{d_m \times d_m}$ 为对角矩阵; d_m 为第 m 个视角的特征数; $\mathbf{x}_i^{(m)}$ 为第 m 个视角的第 i 个样本数据; σ 是核带宽参数。

使用定义 2 将 3.1 节得到的相似度矩阵 \mathbf{A} 和核矩阵 \mathbf{K} 作中心化矩阵校正处理,可最大化它们之间的相关性,即最小化以下目标函数:

$$\min_{\mathbf{f}^{(m)}} L(\mathbf{f}^{(m)}) = -\text{tr}(\mathbf{H}\mathbf{A}\mathbf{H}\mathbf{K}^{(m)}) \quad (9)$$

其中, $\mathbf{f}^{(m)} \in \{0, 1\}$ 是特征选择向量,且满足约束 $0 \leq \mathbf{f}_t^{(m)} \leq 1$, 因此对 $\mathbf{f}^{(m)}$ 加上 l_1 范数约束,用于指导特征选择。得到目标函数如下:

$$\min_{\mathbf{f}^{(m)}} L(\mathbf{f}^{(m)}, \lambda) = -\text{tr}(\mathbf{H}\mathbf{A}\mathbf{H}\mathbf{K}^{(m)}) + \lambda \|\mathbf{f}^{(m)}\|_1 \quad (10)$$

s. t. $0 \leq \mathbf{f}_t^{(m)} \leq 1, \forall t=1, 2, \dots, d_m$

其中, $\mathbf{K}^{(m)}$ 为第 m 个视角的核矩阵; d_m 是第 m 个视角的特征数; λ 是正则化平衡参数。

通过矩阵校正,以上过程可以得到最优特征选择向量 \mathbf{f} 。得到最优解之后,选取特征选择向量 \mathbf{f} 元素为 1 对应的特征作为特征子集的元素,并对多视角数据进行聚类任务的学习。

4 算法求解

基于相似度矩阵学习和矩阵校正的无监督多视角特征选择(SMLMA)方法的目标函数是多变量优化问题,同时更新这些变量属于非凸问题,故使用交替迭代更新的思想,当优化一个变量时,固定其他变量。

针对相似度矩阵的学习问题,先训练一致相似度矩阵 \mathbf{A} 和投影矩阵 \mathbf{W} , 共有 3 组变量,分别是投影矩阵 \mathbf{W} 、视角权重向量 α 、一致相似矩阵 \mathbf{A} 。优化过程如下。

(1) 固定 $\{\mathbf{W}_m\}_{m=1}^M$ 和 \mathbf{A} , 优化 α 。由于 α 只存在于函数的

第一项,可以设:

$$J_m = \sum_{i=1}^n \sum_{j=1}^n \|\mathbf{W}_m^T \mathbf{x}_i^{(m)} - \mathbf{W}_m^T \mathbf{x}_j^{(m)}\|_2^2 (\mathbf{A}_{ij})^{r_2} \quad (11)$$

所以优化问题式(5)可以表示成:

$$\min_{\alpha} L(\alpha) = \sum_{m=1}^M \alpha_m^{r_1} J_m \quad (12)$$

s. t. $\sum_{m=1}^M \alpha_m = 1, \alpha_m \geq 0$

对式(12)使用拉格朗日乘子法,得到拉格朗日函数为:

$$\sum_{m=1}^M \alpha_m^{r_1} J_m - \gamma (\sum_{m=1}^M \alpha_m - 1) \quad (13)$$

其中, γ 是拉格朗日乘子。对式(13)求在 α_m 上的偏导,且将其偏导设为 0, 可得:

$$\alpha_m = \left(\frac{\gamma}{r_1 J_m}\right)^{\frac{1}{r_1-1}} \quad (14)$$

再把式(14)代入条件 $\sum_{m=1}^M \alpha_m = 1$, 可以求解得到:

$$\sum_{m=1}^M (r_1 J_m)^{1/(1-r_1)} = r_1^{1/(1-r_1)} \sum_{m=1}^M J_m^{1/(1-r_1)} \quad (15)$$

结合 $r_1^{1/(1-r_1)} (\alpha_m^{r_1-1} J_m)^{1/(1-r_1)} = \gamma^{1/(1-r_1)}$ 可得:

$$\alpha_m = \frac{(J_m)^{1/(1-r_1)}}{\sum_{m=1}^M J_m^{1/(1-r_1)}} \quad (16)$$

需要说明的是,上述推导过程中没有考虑 α_m 非负约束,但是由于 $g_m \geq 0$, 所以该条件自动成立。

(2) 固定 $\{\mathbf{W}_m\}_{m=1}^M$ 和 α , 优化 \mathbf{A} 。类似于优化 α_m 问题,由于 \mathbf{A} 存在于第一项,故假设:

$$\varphi_{ij} = \sum_{m=1}^M \alpha_m^{r_1} \|\mathbf{W}_m^T \mathbf{x}_i^{(m)} - \mathbf{W}_m^T \mathbf{x}_j^{(m)}\|_2^2 \quad (17)$$

结合式(17),关于 \mathbf{A} 的优化问题式(5)变为:

$$\min_{\mathbf{A}} L(\mathbf{A}) = \sum_{i=1}^n \sum_{j=1}^n \varphi_{ij} (\mathbf{A}_{ij})^{r_2} \quad (18)$$

s. t. $\sum_{j=1}^n \mathbf{A}_{ij} = 1, \mathbf{A}_{ij} \geq 0, \|\mathbf{A}_i\|_0 = k$

由于对变量 \mathbf{A} 的每一行,有约束 $\|\mathbf{A}_i\|_0 = k$, 因此需要对 \mathbf{A} 的每一行进行更新。加入对第 i 行的更新过程,该过程的目标函数表达式如下:

$$\min_{\mathbf{A}_i} L(\mathbf{A}_i) = \sum_{j=1}^n \varphi_{ij} (\mathbf{A}_{ij})^{r_2} \quad (19)$$

s. t. $\sum_{j=1}^n \mathbf{A}_{ij} = 1, \mathbf{A}_{ij} \geq 0, \|\mathbf{A}_i\|_0 = k$

类似于对 α_m 的优化,在它的基础上加上 $\|\mathbf{A}_i\|_0 = k$ 的约束,所以 \mathbf{A}_i 中只有 k 个元素是非零的,只需要优化 k 个非零元素。 φ_{ij} 越大,目标函数就越大,因此,将 \mathbf{A}_i (对应于 $n-k$ 个最大 φ_{ij}) 的 $n-k$ 个元素设置为 0, 只更新其余 k 个元素。具体来说,假设 φ_{ij} 的最小的 k 个元素(固定 i 和 φ_{ii} 除外)是 $\varphi_{ij_1}, \varphi_{ij_2}, \dots, \varphi_{ij_k}$, 对于式(19)中的问题,求解得到:

$$\mathbf{A}_{ijl} = \begin{cases} \frac{(\varphi_{ijl})^{\frac{1}{1-r_2}}}{\sum_{q=1}^k (\varphi_{ijq})^{\frac{1}{1-r_2}}}, & l=1, 2, \dots, k \\ 0, & \text{其他} \end{cases} \quad (20)$$

(3) 固定 \mathbf{A} 和 α , 优化 $\{\mathbf{W}_m\}_{m=1}^M$ 。由于这部分函数中的相似度矩阵 \mathbf{A} 是非对称矩阵,无法保证它对应的拉普拉斯矩阵是半正定的,于是定义新的对称矩阵 \mathbf{Q} :

$$\mathbf{Q}_i = \mathbf{Q}_i = ((\mathbf{A}_{ij})^{r_2} + (\mathbf{A}_{ji})^{r_2}) / 2 \quad (21)$$

因此问题(5)中第 m 个视角的优化函数可以表述为:

$$\begin{aligned} \min_{\mathbf{W}_m} \sum_{i=1}^n \sum_{j=1}^n \|\mathbf{W}_m^T \mathbf{x}_i^{(m)} - \mathbf{W}_m^T \mathbf{x}_j^{(m)}\|_2^2 (\mathbf{A}_{ij})^{\gamma_2} \\ = \sum_{i=1}^n \sum_{j=1}^n \|\mathbf{W}_m^T \mathbf{x}_i^{(m)} - \mathbf{W}_m^T \mathbf{x}_j^{(m)}\|_2^2 \mathbf{Q}_{ij} \end{aligned} \quad (22)$$

再由对称矩阵 \mathbf{Q} 生成拉普拉斯矩阵 \mathbf{L} , 式(22)可转化为:

$$\min_{\mathbf{W}_m} L(\mathbf{W}_m) = \text{tr}(\mathbf{W}_m^T \mathbf{X}^{(m)} \mathbf{L} (\mathbf{X}^{(m)})^T \mathbf{W}_m) \quad (23)$$

$$\text{s. t. } \mathbf{W}_m^T \mathbf{W}_m = \mathbf{I}$$

令式(23)对 \mathbf{W}_m 的偏导为 0, 得到:

$$((\mathbf{X}^{(m)} \mathbf{L} (\mathbf{X}^{(m)})^T)^T + (\mathbf{X}^{(m)} \mathbf{L} (\mathbf{X}^{(m)})^T)) \mathbf{W}_m = 0 \quad (24)$$

式(24)转化为求半正定矩阵 $(\mathbf{X}^{(m)} \mathbf{L} (\mathbf{X}^{(m)})^T)^T + \mathbf{X}^{(m)} \mathbf{L} (\mathbf{X}^{(m)})^T$ 的特征分解问题, 只需求出该部分最小特征值对应的特征向量即为更新的 \mathbf{W}_m 。

上述优化是对相似度矩阵和投影矩阵的学习, 而矩阵校正过程使用的特征选择向量是 $\{0, 1\}$ 列向量, 因此考虑使用逐步搜索的方法, 可降低优化求解的难度且加快收敛速度^[33]。该过程使用拟牛顿法和 Armijo 搜索算法, 求解过程如下。

(4)由式(10), 对 $f^{(m)}$ 求偏导得:

$$\begin{aligned} \frac{\partial L}{\partial f_i^{(m)}} &= - \sum_{i,j=1}^n \left((\mathbf{HAH})_{ij} \cdot \frac{\partial \mathbf{K}_{ij}^{(m)}}{\partial f_i^{(m)}} \right) + \lambda \\ &= \sum_{i,j=1}^n \left((\mathbf{HAH} \otimes \mathbf{K}^{(m)})_{ij} (f_{it}^{(v)} - f_{jt}^{(m)})^2 \right) \frac{2f_i}{\sigma^2} + \lambda \end{aligned} \quad (25)$$

其中, \otimes 表示矩阵对应位置的元素相乘, 传统的牛顿法在第 t 次迭代时对下列二阶近似进行优化:

$$\begin{aligned} P_t(f) &= L(f_t) + (f - f_t)^T \nabla L(f_t) + \\ &\quad \frac{1}{2} (f - f_t)^T \mathbf{B}_t (f - f_t) \end{aligned} \quad (26)$$

其中, $\mathbf{B}_t = \nabla^2 L(f_t)$ 是 Hessian 矩阵。由于这种计算方法存储量较大, 可以计算下列近似 Hessian 矩阵:

$$\mathbf{B}_{t+1} = \mathbf{B}_t - \frac{\mathbf{B}_t u_t u_t^T \mathbf{B}_t}{u_t^T \mathbf{B}_t u_t} + \frac{y_t y_t^T}{y_t^T u_t} \quad (27)$$

其中, $u_t = f_{t+1} - f_t$, $y_t = \nabla L(f_{t+1}) - \nabla L(f_t)$, 结合式(26), 求解最优 f 的问题(式(10))可转化为下述问题:

$$\begin{aligned} \min p_t(f) \\ \text{s. t. } f \in [0, 1] \end{aligned} \quad (28)$$

最终, 分别得到了 3 个变量 \mathbf{A} , α 和 f 的优化表达式, 具体的更新过程如算法 1 所示。

算法 1 基于相似度矩阵学习和矩阵校正的无监督多视角特征选择(SMLMA)算法

输入: $\mathbf{X}_m = (\mathbf{x}_1^{(m)}, \mathbf{x}_2^{(m)}, \dots, \mathbf{x}_n^{(m)}) \in \mathbf{R}^{d_m \times n}$; 第 m 个视角的特征数 d_m ; $k, r_1, r_2, \sigma, \lambda$

输出: 一致相似度矩阵 \mathbf{A} ; 视角权重 $\{\alpha_m\}_{m=1}^M$; 特征选择向量 f

初始化: 根据式(3)构造 \mathbf{A} ; 初始化 $\alpha_m = \frac{1}{M}$, $m=1, 2, \dots, M$

迭代更新:

For $m=1$ To M (循环 m 个视角)

1. 固定 \mathbf{A} 和 $\{\alpha_m\}_{m=1}^M$, 根据式(24)求解半正定矩阵 $(\mathbf{X}^{(m)} \mathbf{L} (\mathbf{X}^{(m)})^T)^T + \mathbf{X}^{(m)} \mathbf{L} (\mathbf{X}^{(m)})^T$ 对应的特征值和特征向量, 然后更新 \mathbf{W}_m ;

2. 固定 \mathbf{W}_m 和 \mathbf{A} , 根据式(16)更新 α_m ;

3. 固定 \mathbf{W}_m 和 α_m 根据式(20)的准则更新 \mathbf{A} 的每一行。

End for

4. 将得到的投影矩阵 $\{\mathbf{W}_m\}_{m=1}^M$ 依据评分准则转换成特征选择向量 $\{f^{(m)}\}_{m=1}^M$, 初始化核矩阵 $\mathbf{K}^{(m)}$;

For $m=1$ To M (循环 m 个视角)

5. 根据式(25)计算梯度 $\nabla L(f^{(m)})$;

6. 根据式(27)计算 Hessian 矩阵;

7. 依据式(28)求解最优特征选择向量 $\{f^{(m)}\}_{m=1}^M$;

End for

直到收敛

根据特征选择向量 $\{f^{(m)}\}_{m=1}^M$ 选取元素 1 对应的特征作为特征子集的元素。

5 实验结果与分析

本节使用 k -均值聚类方法, 对原始数据进行聚类并对使用不同特征选择算法获取的特征子集进行聚类验证与分析, 同时比较不同特征子集的特征信息冗余率, 给出算法参数的灵敏度分析和收敛性分析。

5.1 实验设置

5.1.1 实验数据集

实验中使用了 4 个图像数据集来测试基于相似度矩阵学习和矩阵校正的多视角特征选择方法的性能。具体的数据集包括物体分类图像数据 Caltech-101、web 标签图像数据 NUS-WIDE-OBJ、玩具动物图像数据 Toy Animal 和用于对象识别的图像数据 MSRC-v1, 这几种数据是在许多实际应用中产生的, 常用于多视角的学习^[21,30,34]。表 1 列出了几种数据集的具体信息。其中同一图像从不同角度观察到的特征不同, 故从各图像中提取了几种具体的特征类型, 类别数是根据实验对不同数据集的需要选择。

表 1 样本数据集信息

Table 1 Information of sample datasets

View	Caltech-7	NUS-WIDE-OBJ	Toy Animal	MSRC-v1
#1	LBP(256)	CH(64)	CQ(2688)	LBP(256)
#2	HOG(100)	CMT(225)	SURF(2000)	HOG(100)
#3	GIST(512)	CORR(144)	LSS(200)	GIST(512)
#4	CMT(48)	EDH(73)	RGSIFT(2000)	CMT(48)
#5	CENT(1302)	WT(128)	—	CENT(1302)
#6	SIFT(441)	—	—	DGSIFT(210)
Total	2659	634	6888	2428
Class	7	20	20	7
Size	441	9133	14122	210

5.1.2 对比实验方法

基于 5.1.1 节的 4 个数据集, 分别对拉普拉斯得分法(Laplacian score, LS)^[14]、光谱特征选择方法(SPEC)^[15]、自适应无监督的多视角特征选择方法(AUMFS)^[21]、自适应视角权重和相似度的多视角特征选择方法(ASVW)^[24]、基于跨视角局部结构的多样性保持和共识学习的多视角无监督特征选择方法(CRV-DCL)^[35]、基于非负结构图学习的多视角特征选择方法(NSGL)^[36]、基于相似度矩阵学习和矩阵校正的无监督多视角特征选择(SMLMA)方法等经典的特征选择算法进行实验对比, 各方法的具体内容如下。

(1)LS^[14]方法: 基于拉普拉斯特征映射的思想, 得到每个特征的拉普拉斯得分, 得分越高, 对样本的贡献越大; 反之, 得分越低, 该特征越没有代表性。

(2)SPEC^[15]方法: 是拉普拉斯得分法的一种扩展, 对噪声更具鲁棒性。

(3)AUMFS^[21]方法: 联合利用 3 种重要信息, 其中包括数据聚类结构、数据相似性和原始数据中的不同视角之间的

相关性来学习视觉相似图,在低维流形中保留数据的局部几何结构。

(4)ASVW^[24]方法:利用学习机制自适应地刻画公共结构,在不同视角上用公共拉普拉斯图构造目标函数,并设计用于特征选择的稀疏 $l_{2,p}$ -范数约束。

(5)CRV-DCL^[35]方法:将数据的每个视角投影到一个共同的语义标签空间中,捕获公共信息和跨不同视角的区分知识。此外,分别构造每对成对视角和单个视角的视角内相似度图,以保持不同视角中的数据的局部结构和同一视角中的不同样本的局部结构。

(6)NSGL^[36]方法:通过施加秩约束从原始特征中学习结构化图,同时利用多视角特征的互补性进行自适应特征选择。此外,引入伪标签学习提取非监督场景下的有区别语义信息,从而引导图学习过程。

(7)SMLMA方法:通过数据学习相似度矩阵,保留数据局部几何结构的信息,并将相似度矩阵和核矩阵进行校正,最大化相似度矩阵和核矩阵的相关性,有效降低特征选择的冗余率,提高特征选择的质量^[32]。

5.1.3 评价指标

使用聚类准确率(ACC)指标来评估使用不同特征选择之后进行聚类的性能,可以评估出样本被正确分类的准确程度^[37]。具体定义如下:

$$ACC = \frac{1}{n} \sum_{i=1}^n \delta(\mathbf{g}_i, \text{map}(c_i)) \quad (29)$$

其中, n 是样本总数; \mathbf{g}_i 是实际的样本标签; c_i 是聚类的样本标签; $\delta(\cdot)$ 是一个判别函数,如果 $\mathbf{g}_i = \text{map}(c_i)$,则值为1,否则为0; $\text{map}(\cdot)$ 是一个映射函数,将聚类之后得到的最佳标签映射到实际标签空间中。

使用标准化互信息(NMI)来评估聚类结果与真实数据

标签的一致性。给定两个变量 \mathbf{P} 和 \mathbf{Q} ,NMI的定义如下:

$$NMI(\mathbf{P}, \mathbf{Q}) = \frac{I(\mathbf{P}, \mathbf{Q})}{\sqrt{H(\mathbf{P})H(\mathbf{Q})}} \quad (30)$$

其中, $H(\mathbf{P})$ 和 $H(\mathbf{Q})$ 分别是变量 \mathbf{P} 和 \mathbf{Q} 的熵, $I(\mathbf{P}, \mathbf{Q})$ 是变量 \mathbf{P} 和 \mathbf{Q} 之间的互信息;对于聚类任务, \mathbf{P} 和 \mathbf{Q} 分别表示聚类结果和真实标签。

使用冗余率(RED)来评估所选择的特征质量。假设 $\mathbf{F} = \{f_i\}$ 是选择出来的特征, \mathbf{X}_F 表示在 \mathbf{F} 中的特征所表示的数据^[38],冗余率定义为:

$$RED(\mathbf{F}) = \frac{1}{|\mathbf{F}|(|\mathbf{F}|-1)} \sum_{f_i, f_j \in \mathbf{F}, i > j} corr_{i,j} \quad (31)$$

5.1.4 参数设置

对于参数的选择,在两个数据集上进行实验时,单视角特征选择算法中需要确定构造最近邻图的邻域大小,设置尺寸为5。同样,初始化 \mathbf{A} 时,设置尺寸 $k=5$ 。对于所涉及的常数参数,设置 $r_1=r_2$,并从 $\{2,4,6,8,10\}$ 中选择,参数 $\sigma=0.1$,另外一个平衡参数 λ 从 $\{0.01,0.1,1,10,100\}$ 中选择。特征选择的数量在 $\{50,100,150,200,250,300,350,400,450,500\}$ 中选择。因 k -均值聚类算法依赖初始的聚类中心,每次聚类的结果之间会存在差异,故取5次实验结果的平均值作为最后的结果。

5.2 实验结果与分析

5.2.1 不同聚类方法的聚类结果

对所有的数据使用 k -均值聚类方法进行聚类,并将得到的聚类结果作为基准。使用LS^[14]方法、SPEC^[15]方法进行单视角的特征选择,使用AUMFS^[21],ASVW^[24],CRV-DCL^[35],NSGL^[36]和SMLMA特征选择方法进行多视角的特征选择,再通过 k -均值聚类得到聚类结果并进行比较。7种特征选择算法的聚类准确率结果如图3所示。

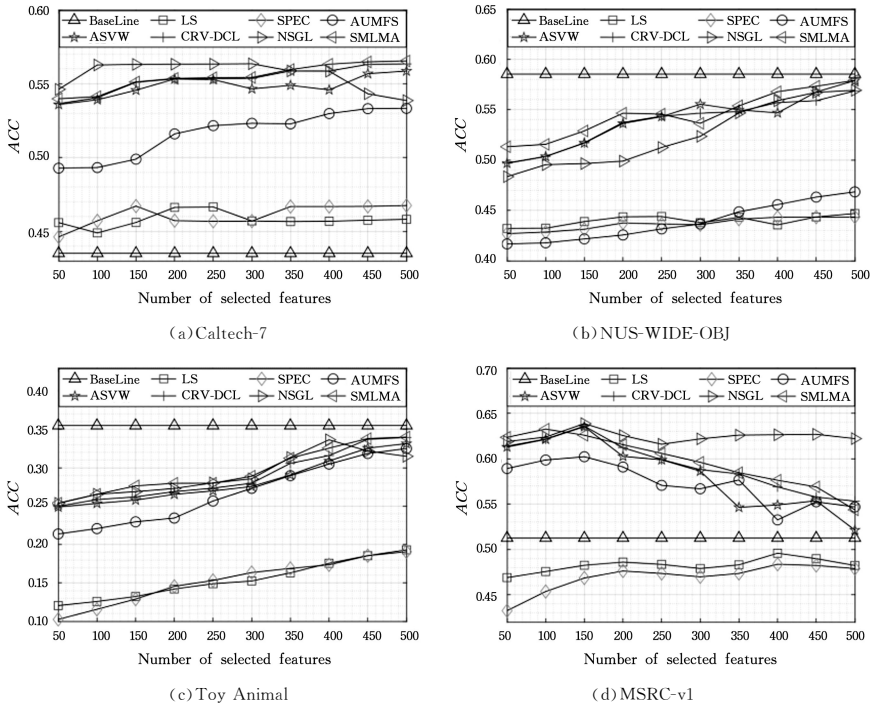


图3 在不同的数据集上采用不同方法选择特征的聚类准确率(ACC)

Fig.3 Clustering accuracy(ACC) of feature selection by different methods on different datasets

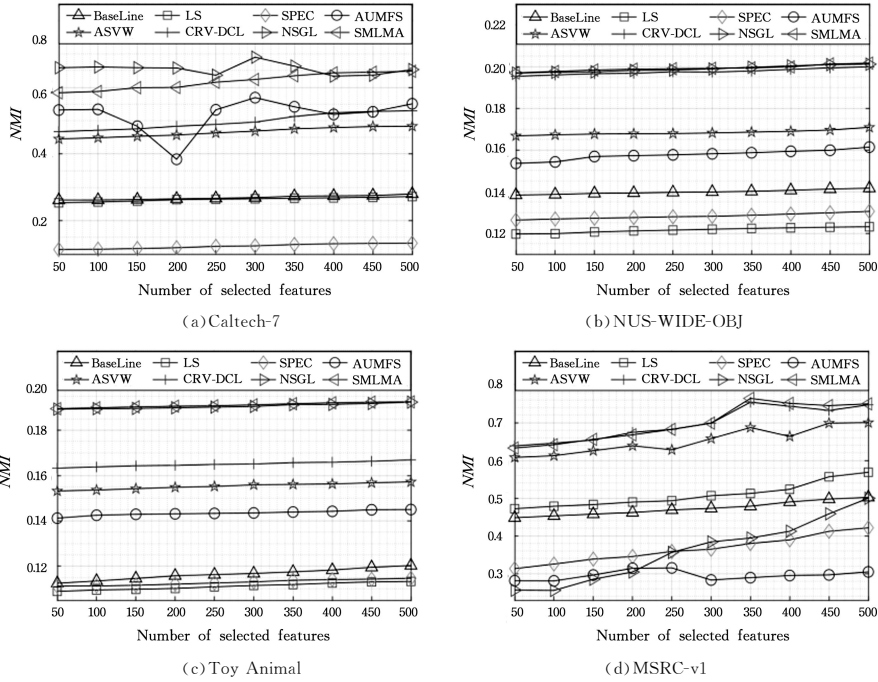


图4 在不同的数据集上采用不同方法选择特征的归一化互信息

Fig. 4 Normalized mutual information(NMI) of feature selection by different methods on different datasets

由图3和图4可以得出以下结论。

(1)从图3(a)和图3(c)可以看出, k -均值方法在数据集 NUS-WIDE-OBJ 和 Toy Animal 上的聚类结果比其他特征选择方法的结果更好,原因可能是数据集原始特征之间足够紧凑,不进行特征选择也能较好地体现特征空间的所有信息。而对于其他3个数据集,进行特征降维之后的聚类结果明显更好,说明特征选择可以有效地选择较少的特征去表征原始的特征空间,且保留大部分信息。

(2)由图3可知,相比单视角特征选择方法,用多视角特征选择方法处理多视角数据具有更好的性能。因为单视角特征选择方法在处理多视角数据时不能合理利用不同视角之间的关联,导致信息损失,特征选择结果较差。

(3)相比其他方法,SMLMA 方法的性能较好。在4个数据集上的实验结果表明,相比不使用特征选择方法的聚类结果,SMLMA 方法的准确率都有提高。

(4)LS 方法和 SPEC 方法有相近的聚类结果,这说明两种方法本质上是相近的。正如前文所述,SPEC 是在 LS 的基础上拓展的,因此特征选择结果的质量相近,导致聚类性能也相似。

(5)相比 ASVW 方法,SMLMA 方法虽然对聚类精确率

的提高较小,但是总体性能较好,特别是在数据集 MSRC-v1 上的聚类结果体现较为明显。

(6)相比 NSGL 方法,SMLMA 方法在 Caltech-7 数据集上的性能表现较差,可能是 NSGL 在利用自适应图引导特征选择的过程中获取了比较充分的结构信息,而 SMLMA 方法对图结构的考虑有所欠缺。

表2和表3列出了7种特征选择方法的聚类准确率结果,表4和表5列出了7种特征选择方法的聚类标准化互信息,其中最优结果用加粗的字体表示,次优结果加下划线。可见,SMLMA 方法在数据集 NUS-WIDE-OBJ 和 Toy Animal 上的结果基本上是最好的,说明 SMLMA 方法在处理相似度矩阵和投影矩阵的过程中得到了较优的特征子集,捕获到了更多数据结构的信息以及得到了更多不同视角之间的关联信息。而在数据集 NUS-WIDE-OBJ 上进行实验且特征数为300,在 MSRC-v1 上进行实验且特征数为150时,SMLMA 方法出现次优结果,ASVW 方法最优,说明 SMLMA 方法仍然存在不足,这可能是受到相似度矩阵训练过程中的特征选择向量构造的影响,使得部分具有代表性的特征被筛选过滤。但从总体结果的对比中可看出,SMLMA 方法能有效地提高聚类的性能。

表2 使用不同方法在 Caltech-7 和 NUS-WIDE-OBJ 数据集上对选择的特征进行聚类的 ACC 比较

Table 2 ACC comparison of features selected by different methods on Caltech-7 and NUS-WIDE-OBJ datasets

Dataset d	Caltech-7									
	50	100	150	200	250	300	350	400	450	500
SMLMA	<u>0.5398</u>	<u>0.5413</u>	<u>0.5512</u>	<u>0.5536</u>	<u>0.5541</u>	0.5543	0.5596	0.5632	0.5648	0.5654
NSGL	0.5462	0.5627	0.5631	0.5632	0.5633	0.5635	<u>0.5587</u>	0.5585	0.5432	0.5386
CRV-DCL	0.5365	0.5405	0.5510	0.5533	0.5535	0.5536	<u>0.5587</u>	<u>0.5586</u>	<u>0.5631</u>	<u>0.5634</u>
ASVW	0.5359	0.5391	0.5456	0.5532	0.5528	0.5465	0.5489	0.5458	0.5568	0.5586
AUMFS	0.4928	0.4932	0.4989	0.5160	0.5216	0.5232	0.5228	0.5298	0.5331	0.5333
SPEC	0.4562	0.4571	0.4571	0.4572	0.4568	0.4569	0.4568	0.4568	0.4569	0.4575
LS	0.4559	0.4562	0.4560	0.4562	0.4565	0.4570	0.4566	0.4568	0.4575	0.4581

(续表)

Dataset	NUS-WIDE-OBJ										
	d	50	100	150	200	250	300	350	400	450	500
SMLMA		0.5132	0.5156	0.5289	0.5465	0.5456	0.5365	0.5538	0.5678	0.5733	0.5791
NSGL		0.4835	0.4954	0.4965	0.4989	0.5125	0.5235	0.5468	<u>0.5589</u>	0.5672	0.5692
CRV-DCL		<u>0.4969</u>	<u>0.5036</u>	<u>0.5171</u>	<u>0.5373</u>	<u>0.5435</u>	0.5465	0.5481	0.5571	0.5589	0.5686
ASVW		0.4968	0.5035	0.5169	0.5363	0.5431	<u>0.5554</u>	<u>0.5498</u>	0.5468	<u>0.5675</u>	<u>0.5789</u>
AUMFS		0.4167	0.4178	0.4217	0.4256	0.4316	0.4365	0.4486	0.4557	0.4632	0.4684
SPEC		0.4269	0.4285	0.4313	0.4375	0.4365	0.4358	0.4413	0.4432	0.4433	0.4435
LS		0.4321	0.4323	0.4389	0.4436	0.4439	0.4378	0.4431	0.4356	0.4435	0.4469

表 3 使用不同方法在 Toy Animal 和 MSRC-v1 数据集上对选择的特征进行聚类的 ACC 比较

Table 3 ACC comparison of features selected by different methods on Toy Animal and MSRC-v1 datasets

Dataset	Toy Animal										
	d	50	100	150	200	250	300	350	400	450	500
SMLMA		0.2538	0.2654	0.2764	0.2798	0.2799	0.2890	<u>0.3125</u>	<u>0.3256</u>	0.3380	0.3400
NSGL		<u>0.2537</u>	<u>0.2655</u>	<u>0.2689</u>	<u>0.2731</u>	<u>0.2803</u>	<u>0.2856</u>	0.3134	0.3365	0.3218	0.3149
CRV-DCL		0.2498	0.2586	0.2621	0.2689	0.2735	0.2799	0.3058	0.3159	<u>0.3371</u>	<u>0.3395</u>
ASVW		0.2488	0.2536	0.2579	0.2654	0.2699	0.2756	0.2906	0.3089	0.3256	0.3315
AUMFS		0.2138	0.2208	0.2296	0.2345	0.2567	0.2732	0.2897	0.3048	0.3186	0.3251
SPEC		0.1025	0.1156	0.1289	0.1456	0.1532	0.1635	0.1689	0.1736	0.1855	0.1906
LS		0.1205	0.1258	0.1323	0.1421	0.1489	0.1525	0.1631	0.1756	0.1855	0.1931

Dataset	MSRC-v1										
	d	50	100	150	200	250	300	350	400	450	500
SMLMA		0.6235	0.6325	0.6258	<u>0.6154</u>	<u>0.6058</u>	<u>0.5960</u>	<u>0.5845</u>	<u>0.5762</u>	<u>0.5689</u>	0.5436
NSGL		<u>0.6189</u>	<u>0.6238</u>	0.6386	0.6254	0.6159	0.6218	0.6259	0.6263	0.6268	0.6221
CRV-DCL		0.6138	0.6216	<u>0.6358</u>	0.6121	0.5989	0.5879	0.5831	0.5689	0.5576	<u>0.5532</u>
ASVW		0.6128	0.6213	0.6352	0.6025	0.5987	0.5862	0.5463	0.5489	0.5536	0.5213
AUMFS		0.5891	0.5985	0.6021	0.5908	0.5706	0.5668	0.5765	0.5325	0.5524	0.5468
SPEC		0.4325	0.4536	0.4685	0.4762	0.4735	0.4698	0.4735	0.4836	0.4823	0.4789
LS		0.4756	0.4825	0.4860	0.4835	0.4756	0.4789	0.4830	0.4958	0.4898	0.4823

表 4 使用不同方法在 Caltech-7 和 NUS-WIDE-OBJ 数据集上对选择的特征进行聚类的 NMI 比较

Table 4 NMI comparison of features selected by different methods on Caltech-7 and NUS-WIDE-OBJ datasets

Dataset	Caltech-7										
	d	50	100	150	200	250	300	350	400	450	500
SMLMA		<u>0.5836</u>	<u>0.5875</u>	<u>0.5989</u>	<u>0.5996</u>	0.6153	0.6235	0.6346	0.6432	0.6453	<u>0.6482</u>
NSGL		0.6587	0.6609	0.6588	0.6579	0.6362	0.6903	0.6635	<u>0.6336</u>	<u>0.6363</u>	0.6532
CRV-DCL		0.4672	0.4712	0.4756	0.4832	0.4889	0.4956	0.5125	0.5236	0.5276	0.5295
ASVW		0.4451	0.4486	0.4531	0.4568	0.4632	0.4687	0.4754	0.4793	0.4823	0.4832
AUMFS		0.5318	0.5339	0.4832	0.3842	0.5325	0.5690	0.5416	0.5187	0.5265	0.5505
SPEC		0.1136	0.1145	0.1165	0.1189	0.1232	0.1246	0.1286	0.1309	0.1319	0.1328
LS		0.2538	0.2562	0.2589	0.2631	0.2646	0.2658	0.2675	0.2686	0.2701	0.2724

Dataset	NUS-WIDE-OBJ										
	d	50	100	150	200	250	300	350	400	450	500
SMLMA		0.1969	0.1976	0.1983	0.1987	0.1989	0.1991	<u>0.1993</u>	<u>0.1998</u>	0.2012	0.2016
NSGL		<u>0.1968</u>	<u>0.1972</u>	<u>0.1974</u>	<u>0.1981</u>	<u>0.1983</u>	<u>0.1988</u>	0.1996	0.2003	<u>0.2009</u>	<u>0.2011</u>
CRV-DCL		0.1953	0.1959	0.1965	0.1967	0.1974	0.1973	0.1978	0.1986	0.1993	0.1998
ASVW		0.1667	0.1672	0.1676	0.1677	0.1678	0.1681	0.1685	0.1689	0.1695	0.1708
AUMFS		0.1535	0.1542	0.1568	0.1572	0.1576	0.1581	0.1586	0.1593	0.1598	0.1613
SPEC		0.1263	0.1268	0.1272	0.1275	0.1279	0.1281	0.1286	0.1292	0.1298	0.1305
LS		0.1196	0.1198	0.1206	0.1211	0.1215	0.1219	0.1223	0.1226	0.1229	0.1231

表 5 使用不同方法在 Toy Animal 和 MSRC-v1 数据集上对选择的特征进行聚类的 NMI 比较

Table 5 NMI comparison of features selected by different methods on Toy Animal and MSRC-v1 datasets

Dataset	Toy Animal										
	d	50	100	150	200	250	300	350	400	450	500
SMLMA		0.1895	0.1899	0.1903	0.1906	0.1908	0.1912	0.1917	0.1921	0.1924	0.1925
NSGL		<u>0.1893</u>	<u>0.1895</u>	<u>0.1896</u>	<u>0.1899</u>	<u>0.1902</u>	<u>0.1905</u>	<u>0.1913</u>	<u>0.1914</u>	<u>0.1918</u>	<u>0.1924</u>
CRV-DCL		0.1633	0.1638	0.1643	0.1645	0.1649	0.1651	0.1657	0.1659	0.1663	0.1669
ASVW		0.1532	0.1536	0.1542	0.1548	0.1552	0.1559	0.1561	0.1563	0.1568	0.1572
AUMFS		0.1412	0.1425	0.1429	0.1431	0.1433	0.1435	0.1439	0.1442	0.1449	0.1450
SPEC		0.1112	0.1113	0.1116	0.1121	0.1126	0.1131	0.1138	0.1141	0.1143	0.1147
LS		0.1089	0.1095	0.1098	0.1102	0.1109	0.1115	0.1119	0.1126	0.1131	0.1133

(续表)

Dataset	MSRC-v1										
	d	50	100	150	200	250	300	350	400	450	500
SMLMA		0.6328	<u>0.6415</u>	0.6559	<u>0.6678</u>	0.6823	0.6989	0.7643	0.7512	0.7452	0.7498
NSGL		0.2563	0.2557	0.2854	0.3046	0.3564	0.3843	0.3945	0.4125	0.4587	0.4999
CRV-DCL		0.6386	0.6457	<u>0.6539</u>	0.6745	0.6823	<u>0.6987</u>	<u>0.7542</u>	<u>0.7436</u>	<u>0.7324</u>	<u>0.7472</u>
ASVW		0.6079	0.6125	0.6258	0.6387	0.6281	0.6578	0.6871	0.6635	0.6987	0.7001
AUMFS		0.2815	0.2810	0.2965	0.3143	0.3154	0.2833	0.2896	0.2952	0.2968	0.3048
SPEC		0.3133	0.3256	0.3387	0.3456	0.3589	0.3648	0.3799	0.3897	0.4123	0.4215
LS		0.4721	0.4786	0.4831	0.4897	0.4932	0.5064	0.5125	0.5236	0.5568	0.5686

5.2.2 冗余率分析

图 5 给出了 5 种算法的冗余率。从图中可以看出算法 SMLMA 出现了较小的冗余率,而 AUMFS 方法在前 3 个数据集上实验的结果总是出现冗余率较高的情况,说明 AUMFS 方法得到的特征子集质量较低,容易出现冗余的信息。ASVW 方法在整体上也表现出较低的冗余率,因为

ASVW 方法利用视角权重值来平衡各个视角的重要程度,筛选相关以及互补的特征,一定程度上过滤掉了重复的特征信息,提高了特征子集的质量。而 SMLMA 方法将相似度矩阵和核矩阵进行校正,较大程度地剔除了不相关与不具代表性的特征,在捕获数据结构的同时也减少了信息的冗余,提高了特征子集的纯度,进而提高聚类学习的性能。

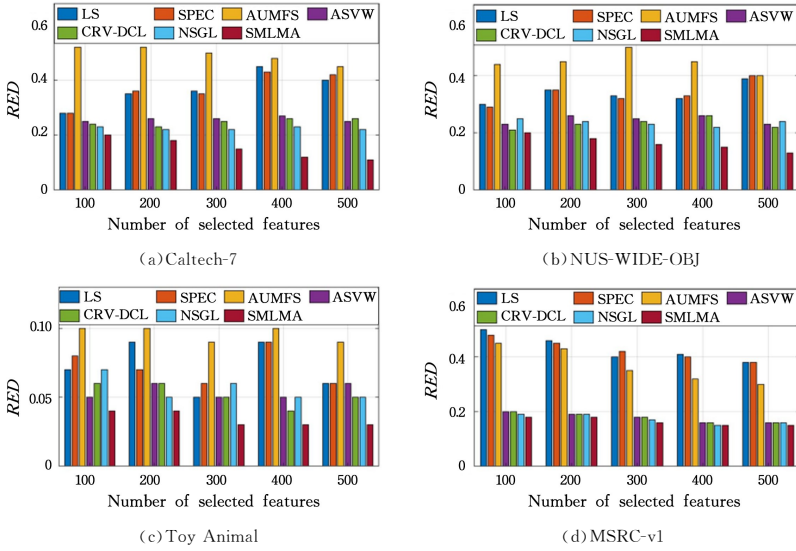


图 5 不同特征数下不同方法的 RED

Fig. 5 RED of different methods with different number of features

CRV-DCL 方法与 ASVW 的冗余率比较接近,但都高于 SMLMA 方法,这是因为 CRV-DCL 方法在 ASVW 的基础上做了全局和局部结构的保持,但是对所获取的局部结构之间的相关信息考虑有所欠缺。NSGL 方法的冗余率仅在 MSRC-v1 数据集上与 SMLMA 方法接近,在 Caltech-7, NUS-WIDE-OBJ 和 Toy Animal 数据集上均较高,说明 NSGL 方法虽然对数据结构的學習有较优的性能,但是剔除冗余信息的能力不足。

5.2.3 参数灵敏度分析

SMLMA 做了多次实验来测试参数的灵敏度,结果如图 6 所示。对于参数 $r_1 = r_2$ 以及 λ ,当有不同的组合时,聚类准确率出现了不同的变化。由图 6 可以看出,当参数 r 一定时,聚类准确率随 λ 值的变化波动比较明显,特别是 λ 值为 0.1 时, SMLMA 方法的准确率较高,说明此时该方法具有较好的性能,得到的结果较优;而当 λ 值固定时,聚类准确率的值稳定在一定范围内,说明 SMLMA 方法对参数 r 不敏感,而相对依赖于 λ 值的变化。

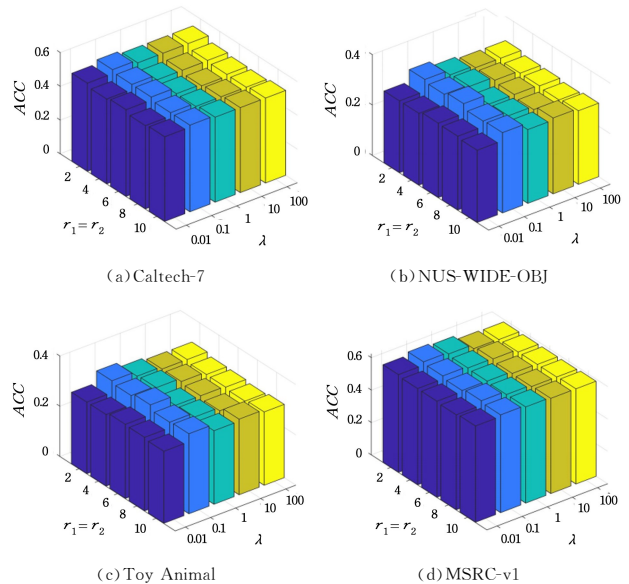


图 6 SMLMA 在不同参数组合下的 ACC

Fig. 6 ACC of SMLMA with different parameter combinations

5.2.4 收敛性分析

SMLMA 方法在数据集 Caltech-7 和 NUS-WIDE-OBJ 上的收敛曲线如图 7 所示。通过计算相似度矩阵学习以及矩阵校正过程中的目标函数总值来度量该方法的收敛速度,从图 7(a)可看出,SMLMA 方法在 Caltech-7 数据集上实验时,目标函数值随迭代次数增加而减小,并在迭代 9 次后开始收敛;从图(b)可看出,SMLMA 方法在 NUS-WIDE-OBJ 数据集上实验时,目标函数值也随迭代次数增加而逐渐减小,在第 7 次迭代之后开始收敛。这表明 SMLMA 方法在有限的迭代次数内能以较快的速度达到收敛。

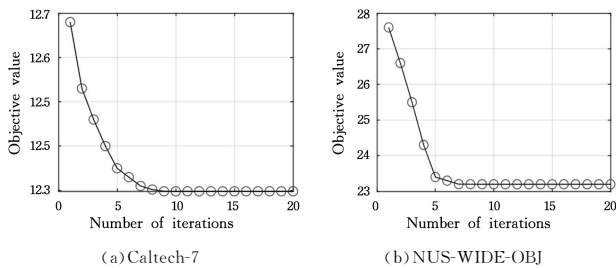


图 7 不同迭代次数下 SMLMA 的目标函数值

Fig. 7 Objective function values of SMLMA with different iterations

5.3 算法分析

针对第 4 节的算法求解和算法 1 进行分析可知,SMLMA 算法使用交替迭代优化的方法,需进行两个内循环,由于在更新核矩阵时需要计算 n^2 个元素,而其他过程均为 $O(1)$ 的时间,并且需要对 M 个视角执行循环,故时间复杂度是 $O(Mn^2)$ 。

SMLMA 算法在时间上有一定的优势,但是算法过程有两个内循环,分别进行不同过程的迭代,可能在投影矩阵转换为特征选择向量的过程中丢失部分信息,导致选择的特征影响聚类的性能,这需要在后续研究工作中对模型的优化提出改进。

结束语 本文提出基于相似度矩阵学习和矩阵校正的无监督多视角特征选择(SMLMA)方法,通过学习相似度矩阵和投影矩阵,嵌入式地学习特征子空间,捕获数据的内部结构;将特征选择向量构造为核心矩阵,保留具有代表性的特征信息与特征空间的结构信息;融合式地将相似度矩阵和核矩阵进行校正,使得这两个矩阵的相关性达到最大,进而较大程度地联合数据的结构信息和特征选择信息来筛选出最具代表性的特征,得到高质量的特征子集。经过实验对比发现,SMLMA 方法在聚类学习的过程中体现出了较好的性能,说明 SMLMA 方法选择的特征较 CRV-DCL, ASVW, AUMFS 等方法更优,能有效地剔除重复特征,较好地保留数据的结构信息,提高聚类任务的效率。本文根据投影矩阵与特征选择向量构造核矩阵,可能会将少部分有用的特征过滤掉,从而影响特征子集的代表性,出现聚类精确率低于 ASVW 和 NSGL 方法的结果。这说明 SMLMA 方法在未来的研究过程中需要进一步考虑如何有效保留原始特征空间中有用的特征,在减少冗余的基础上学习具有代表性特征的特征子空间。本文的 SMLMA 方法虽然相比侧重学习相似度矩阵的 CRV-DCL

和 ASVW 方法具有更好的性能,但是与侧重学习图结构的 NSGL 方法相比仍有不足。在未来的研究中,将考虑使用图结构和伪标签学习相结合、全局和局部结构学习相结合的方法,选择高质量和高效的特征子集;同时,为减少信息损失,还需提出更加简洁和高效的算法步骤。

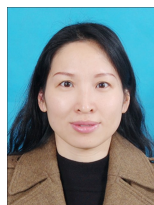
参 考 文 献

- [1] JAMIESON K, BALAKRISHNAN H, TAY Y C. Sift: A MAC Protocol for Event-Driven Wireless Sensor Networks [C]// European Workshop on Wireless Sensor Networks. Berlin: Springer, 2006: 260-275.
- [2] WANG X, HAN T X, YAN S. An HOG-LBP human detector with partial occlusion handling [C]// 2009 IEEE 12th International Conference on Computer Vision. IEEE, 2009: 32-39.
- [3] TAN X, TRIGGS B. Fusing Gabor and LBP Feature Sets for Kernel-Based Face Recognition [C]// International Workshop on Analysis and Modeling of Faces and Gestures. Berlin: Springer, 2007: 235-249.
- [4] LI L, CAI M. Drug target prediction by multi-view low rank embedding[J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2017, 16(5): 1712-1721.
- [5] ZHANG H, WU D, NIE F, et al. Multilevel projections with adaptive neighbor graph for unsupervised multi-view feature selection [J]. Information Fusion, 2020, 70(3): 129-140.
- [6] SUN S. A survey of multi-view machine learning [J]. Neural Computing and Applications, 2013, 23(7): 2031-2038.
- [7] ZHAO J, XIE X, XU X, et al. Multi-view learning overview: Recent progress and new challenges [J]. Information Fusion, 2017, 38: 43-54.
- [8] XIE X. Regularized multi-view least squares twin support vector machines [J]. Applied Intelligence, 2018, 48(9): 3108-3115.
- [9] LI X, ZHANG H, WANG R, et al. Multi-view Clustering: A Scalable and Parameter-free Bipartite Graph Fusion Method [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 44(1): 330-344.
- [10] XIE X, SUN S. General multi-view learning with maximum entropy discrimination [J]. Neurocomputing, 2019, 332: 184-192.
- [11] YIN J, SUN S. Multiview Uncorrelated Locality Preserving Projection [J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(9): 3442-3455.
- [12] LIN G F, ZHU H, FAN C X, et al. Multi-cluster Feature Selection Based on Grassmann Manifold [J]. Computer Engineering, 2012, 16: 3511-3518.
- [13] WAN Y, CHEN X, ZHANG J. Global and intrinsic geometric structure embedding for unsupervised feature selection [J]. Expert Systems with Applications, 2018, 93(March): 134-142.
- [14] HE X, CAI D, NIYOGI P. Laplacian Score for Feature Selection [C]// Advances in Neural Information Processing Systems. 2005: 1-8.
- [15] ZHAO Z, LIU H. Spectral feature selection for supervised and unsupervised learning [C]// Proceedings of the 24th International Conference on Machine learning. Association for Computing Machinery, 2007: 1151-1157.
- [16] ZHAO Z, WANG L, LIU H. Efficient spectral feature selection with minimum redundancy [C]// Proceedings of the AAAI Con-

- ference on Artificial Intelligence, 2010;1-6.
- [17] HE X. Locality preserving projections [J]. *Advances in Neural Information Processing Systems*, 2003, 16(1): 186-197.
- [18] BELKIN M, NIYOGI P. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation [J]. *Neural Computation*, 2003, 15(6): 1373-1396.
- [19] 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*, 2013, 25(6): 1083-1095.
- [20] DU L, SHEN Y D. Unsupervised feature selection with adaptive structure learning [C]// *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015; 209-218.
- [21] FENG Y, XIAO J, ZHUANG Y, et al. Adaptive Unsupervised Multi-view Feature Selection for Visual Concept Recognition [C]// *Asian Conference on Computer Vision*. Berlin: Springer, 2012; 343-357.
- [22] LI J, HU X, TANG J, et al. Unsupervised streaming feature selection in social media [C]// *Proceedings of the 24th ACM International Conference on Information and Knowledge Management*, 2015; 1041-1050.
- [23] WANG Z, FENG Y F, TIAN Q, et al. Adaptive multi-view feature selection for human motion retrieval [J]. *Signal Processing*, 2016, 100(120): 691-701.
- [24] HOU C, NIE F, TAO H, et al. Multi-View Unsupervised Feature Selection with Adaptive Similarity and View Weight [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2017, 29(9): 1998-2011.
- [25] SHAO W, HE L, LU C T, et al. Online Unsupervised Multi-view Feature Selection [J]. 2016 *IEEE 16th International Conference on Data Mining (ICDM)*. IEEE, 2016; 1203-1208.
- [26] DING Z, FU Y. Low-rank common subspace for multi-view learning [C]// *2014 IEEE International Conference on Data Mining*. IEEE, 2014; 110-119.
- [27] WAN Y, SUN S Z, ZENG C, et al. Adaptive Similarity Embedding For Unsupervised Multi-View Feature Selection [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2020, 33(10): 3338-3350.
- [28] XU C, TAO D, XU C, et al. Multi-View Intact Space Learning [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(12): 2531-2544.
- [29] TANG C, CHEN J, LIU X, et al. Consensus Learning Guided Multi-view Unsupervised Feature Selection [J]. *Knowledge-Based Systems*, 2018, 160: 49-60.
- [30] DONG X, ZHU L, SONG X, et al. Adaptive collaborative similarity learning for unsupervised multi-view feature selection [C]// *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2018; 2064-2070.
- [31] SHI C, GU Z, DUAN C, et al. Multi-view Adaptive Semi-supervised Feature Selection with the self-paced learning [J]. *Signal Processing*, 2020, 168: 107332.
- [32] ZHENG X, CHEN J, TANG C, et al. Single-Cell RNA-Sequencing Data Clustering via Locality Preserving Kernel Matrix Alignment [J]. *IEEE Access*, 2020, 8: 201577-201594.
- [33] HUANG F, WU Z Z. Analysis and comparison of several conjugate gradient methods based on Armijo search step [J]. *Journal of Chengdu Institute of Information Engineering*, 2019, 34(2): 209-215.
- [34] ZHANG R, NIE F, LI X, et al. Feature selection with multi-view data: A survey [J]. *Information Fusion*, 2019, 50: 158-167.
- [35] TANG C, ZHU X, LIU X, et al. Cross-View Local Structure Preserved Diversity and Consensus Learning for Multi-View Unsupervised Feature Selection [C]// *The AAAI Conference on Artificial Intelligence*, 2019; 5101-5108.
- [36] XB A, LEI Z A, CHENG L A, et al. Multi-view feature selection via Nonnegative Structured Graph Learning [J]. *Neurocomputing*, 2020, 387: 110-122.
- [37] STREHL A, GHOSH J, CARDIE C, et al. Cluster Ensembles: A Knowledge Reuse Framework for Combining Multiple Partitions [J]. *Journal of Machine Learning Research*, 2002, 3(3): 583-617.
- [38] YE X Y, YE X Y, ZHOU H. Feature selection based on influence community detection and ant colony algorithm [J]. *Computer Engineering and Design*, 2019, 40(9): 2684-2691.



LI Bin, born in 1997, postgraduate. His main research interests include machine learning, pattern recognition and dimension reduction of high dimensional image features.



WAN Yuan, born in 1976, Ph.D, professor, is a member of China Computer Federation. Her main research interests include machine learning, image processing and pattern recognition.

(责任编辑:杨雪敏)