

灵活自适应的无监督降维

强倩瑶 张斌

(西安交通大学软件学院 西安 710049)

摘要 无标签高维数据在图像处理、自然语言处理、数据挖掘等实际场景中无处不在。人工手动标记工作量大、劳动密集、时间开销高、易受主观因素影响且普适性差。计算机在处理高维数据时,时间复杂度大且对硬件配置要求高。因此,无监督降维成为一项迫切的需求。常规基于图的无监督降维方法,使用提前构建的、固定的相似度图学习得到高维数据的低维表示。然而,使用固定不变的相似度图无法修改噪点、离群值和样本外数据引起的不可靠相似度信息。这对于复杂多样的实际任务要求过于严格。为此,本文提出一种灵活自适应的无监督降维(Flexible and Adaptive Unsupervised Dimensionality Reduction, FAUDR)方法。FAUDR通过引入一个回归项,使严格的线性映射规则得到灵活地松弛,以更好地处理可能会引起不可靠信息的噪点、离群值和样本外数据。在降维过程中,该方法同时依赖原始高维数据和动态变化的低维表示,自适应地学习相似度图。将相似度图的构建和低维表示结合起来。自适应学习的相似度图在原始高维空间以及低维空间都实现了理想的邻居分配。这也促进了最佳低维子空间的探索。此外,本文采用了一种有效的交替迭代优化算法,依次更新目标问题中的所有变量。结束迭代后,同时得到相似度矩阵和低维表示的最优解。最后,本文从理论上分析了该算法的收敛性、计算复杂度和存储复杂度。实验分别在两个合成数据集上和八个基准数据集上进行。合成数据集上的实验直观地展示了FAUDR处理噪点和离群值的能力。基准数据上的实验从降维性能、参数敏感性和收敛性三个方面分别验证了FAUDR的有效性。综合实验结果表明,相比于一些经典方法和当前有代表性的方法,本文所提出方法表现出良好的降维性能。不同维度基准数据集上的实验结果显示,该方法比结果第二好的方法在精度(Accuracy, ACC)、标准化互信息(Normalized Mutual Information, NMI)和纯度(Purity)上分别提升了至少3.25%、0.73%和3.00%。

关键词 无监督学习;降维;相似度图;高维数据;低维表示

中图分类号 TP391 DOI号 10.11897/SP.J.1016.2022.Q2290

Flexible and Adaptive Unsupervised Dimensionality Reduction

QIANG Qian-Yao ZHANG Bin

(School of Software, Xi'an Jiaotong University, Xi'an 710049)

Abstract Label-free high-dimensional data is ubiquitous in practical scenarios such as image processing, natural language processing, and data mining. Manual labeling has the disadvantages of heavy workload, labor-intensive, high time overhead, easy to be affected by subjective factors, and poor universality. When a computer processes high-dimensional data, the time complexity is large, and the hardware configuration requirements are high. In view of the above shortcomings of using unlabeled high-dimensional data, unsupervised dimensionality reduction technology has become an urgent need. The conventional dimensionality reduction method based on graph uses the previous constructed and fixed similarity graphs to learn the low-dimensional representation of high-dimensional data. The dimensionality reduction operation is usually done in two steps, first constructing the similarity matrix and then learning the low-dimensional representation on the

basis of the pre-constructed similarity matrix. However, there is an important problem that the use of a fixed similarity matrix cannot modify the unreliable similarity information caused by noise points, outlier samples and out-of-sample data. Using fixed similarity graph is too strict for practical tasks. Because the actual tasks are complex and diverse, the existence of noise points, outlier samples and out-of-sample data cannot be avoided. It is necessary to heuristically learn the similarity matrix and low-dimensional representation at the same time during the dimensionality reduction procedure, instead of the two-stage dimensionality reduction process. In order to solve the problems mentioned above, a Flexible and Adaptive Unsupervised Dimensionality Reduction (FAUDR) method is proposed. By introducing a regression term, FAUDR flexibly relax the widely adopted strict linear mapping rules to better deal with noise points, outlier samples and out-of-sample data that may cause unreliable information. In the process of dimensionality reduction, the proposed method relies on both the original high-dimensional data and the dynamically changing low-dimensional representations, so as to adaptively learn the similarity graph. It unifies the construction of similarity matrix and the learning of low-dimensional representation. The adaptively learned similarity matrix achieves ideal neighbor assignments in both the original high-dimensional space as well as the low-dimensional subspace. This also facilitates the exploration of low-dimensional subspaces. Furthermore, an efficient alternating iterative optimization algorithm is derived to update all variables in the resultant objective problem one by one. At the termination of the iteration, the optimal solutions of similarity matrix and low-dimensional representation are simultaneously obtained. Finally, the theoretical analysis of the algorithm in terms of convergence analysis, computational complexity and storage complexity is provided. The experiments are conducted on two synthetic datasets and eight benchmark datasets. Experiments on synthetic dataset visually demonstrates the ability of FAUDR to handle noise points and outlier samples. Experiments on benchmark data verify the effectiveness of FAUDR from three aspects including dimensionality reduction performance, parameter sensitivity and convergence. Comprehensive experimental results indicate that compared with some classic methods and current representative methods, the method proposed in this paper has the best superiority. The experimental results on eight benchmark datasets of different dimensionality show that compared with the second best method, FAUDR has improved ACC, NMI and Purity by at least 3.25%, 0.73% and 3.00%, respectively.

Keywords unsupervised learning; dimensionality reduction; similarity graph; high-dimensional data; low-dimensional representation

1 引言

现如今,在计算机视觉、数据挖掘、多媒体应用等领域,高维数据大量存在。众所周知,无关或冗余特征的存在会降低学习质量、增加计算时间并消耗更多内存。一般来说,一个数据集的本征维度要低很多^[1]。降维是处理高维数据的一项至关重要的技术。按照标签的使用情况,可以将其分为有监督、半监督和无监督降维^[2]。真实情景中,有标记数据难以获得,人工手动标记成本高、代价大、效率低。无监督降维可以缩小无标签高维数据的维度规模,学得紧致

的低维特征^[3],使用最广泛。紧致的低维特征可以促进高效的聚类等无监督任务。

基于图的降维方法通过分析相似度图,将高维样本嵌入到低维子空间中,该子空间保留了原始样本之间的相似性^[4]。相似度图由相似度矩阵度量,该矩阵表征数据集的某些统计特征或几何特性。作为一种有效的数据处理手段,基于图的无监督降维在机器学习中作用突出。代表性的方法有:局部线性映射(Locally Linear Embedding, LLE)^[5]、拉普拉斯特征映射(Laplacian Eigenmaps, LE)^[6]、等度量映射(ISometric MAPPING, ISOMAP)^[7]、局部保持投影(Locality Preserving Projections, LPP)^[8]、主成

分分析(Principle Component Analysis, PCA)^[9]等. 其中, LLE、LE、ISOMAP 是非线性的, LPP、PCA 是线性的.

LLE 假设每一个数据点都可以由其周围近邻点的线性加权组合重构得到, 降维后这一线性关系尽可能保持不变. LE 希望相互间有关联的数据点在低维子空间中尽可能地靠近. ISOMAP 通过近似保持数据对之间的测地距离, 控制数据信息的流失, 在低维空间中更加全面地将高维数据表现出来^[10]. PCA 将原始样本投影到一个低维子空间, 在此子空间上, 所有样本共享最大的方差. 由于 LLE、LE 等非线性方法难以合理地处理离群值并对噪声数据较为敏感, 许多线性方法被提出以解决此类问题. 带有附加噪声的局部线性嵌入(Locally Linear Embedding with Additive Noise, LLEAN)^[11] 通过利用局部线性和重构潜力之间的关系恢复无噪数据, 然后使用恢复的无噪数据执行降维. LLEAN 是一种改进了的 LLE. LPP 是众多线性方法中典型的一个, 可以被看作是 LE 的线性扩展, 以较好地映射训练数据和样本外数据. LPP 的正则化最小二乘法^[12] 通过建立 LPP 与最小二乘问题之间的联系, 学习投影矩阵. 图嵌入的广义最小二乘正则(Generalized Least-Squares Regulation in Graph Embedding, GLSRGE)^[13] 方法提出了一种类似 PCA 的思想, 通过考虑数据分布和每个数据点的实例惩罚, 得到可以同时保留数据的固有几何结构和全局结构的低维投影.

传统的基于图的降维方法先构造出一个相似度图, 然后对预先构造的固定的图执行谱分析以探索低维子空间^[6-8]. 尽管已经取得了许多进展, 然而, 实际应用中的数据种类丰富且复杂多样, 噪声数据或离群值普遍存在. 传统方法使用预先构建的固定的相似度图无法修改噪声数据或由离群值引起的不可靠相似度信息, 难以适用于真实应用. 因此, 有必要构造更好的相似度图以提升降维性能. 局部自适应保留投影(Locality Adaptive Preserving Projections, LAPP)^[14] 在最优子空间而不是原始高维空间识别每个样本的真实邻居、确定样本间的相似性关系. 具体来说, LAPP 采用从粗到细的策略更新投影的低维子空间, 优化数据局部结构的识别性. 基于双图的判别投影(Double Graphs-based Discriminant Projections, DGDP)提出基于双图进行判别投影, 通过集成两个设计的全局相似度图发现高维数据的判别结构和几何结构^[15]. 局部敏感判别无监督降维(Locality Sensitive Discriminative Unsupervised Dimensionality

Reduction, LSDUDR)^[16] 通过构造表示原始数据相似性和多样性结构的邻接图, 对拉普拉斯矩阵施加秩约束, 建立新的图学习方法进行降维. 通过自适应图学习和约束进行无监督特征选择的方法 EGCFs^[17], 将图学习纳入优化过程, 利用嵌入图学习和约束来选择但不相关但有区别的特征. EGCFs 以较大的概率连接距离上更接近的样本点. 基于图优化的无监督投影(Unsupervised Projection with Graph Optimization, UPGO)^[18] 将构造相似度图和学习投影矩阵统一到一个通用框架中. 该方法引入对拉普拉斯矩阵的约束, 根据低维表示之间的关系学习包含聚类结构的图相似度矩阵, 并在此基础上学习最优的低维子空间.

在上述方法中, 鉴于 LLE、LE 等非线性方法对噪声数据较为敏感, LLEAN、LPP 和 GLSRGE 基于固定的相似度图、利用线性投影规则处理噪声数据进行降维. 但是, 使用固定的相似度图无法修改因噪声数据或离群值引起的不可靠相似度信息. 因此, LAPP、DGDP、LSDUDR、EGCFs 和 UPGO 等方法设计不同的构图策略构造更可靠的相似度矩阵. 尽管相似度矩阵的质量得到了不同程度的提升, 但这些方法均采用严格的线性投影原则降维. 真实应用中的数据复杂多样, 线性投影规则对于真实数据的要求过于严格. 高维数据与最优低维子空间之间的关系有待进一步探索.

为解决当前降维方法面临的问题: (1) 非线性方法难以合理地处理离群值并对噪声数据较为敏感; (2) 线性投影原则虽然可以处理离群值和噪声数据, 但严格的线性投影对于实际数据情况要求过高; (3) 基于固定的相似度图探索低维子空间会导致学习过程无法修正由噪点和离群值引起的不可靠结果. 本文提出一个灵活自适应的无监督降维(Flexible and Adaptive Unsupervised Dimensionality Reduction, FAUDR)方法. 此方法引入了一个线性回归项, 以便灵活地处理线性和非线性分布的数据. FAUDR 在低维空间的探索过程中, 为原始样本空间中的数据点分配最优近邻点, 自适应地构造相似度矩阵. 也就是说, FAUDR 同时借助原始数据和动态更新的低维表示学习相似度图. 此时, 新数据、噪点、离群值都可以得到更好地处理, 从而学得可以保持原始高维数据内部结构和统计特性的低维表示.

2 相关工作

本文工作受到图嵌入和线性投影思想的启发.

下面首先给出文中的符号说明, 然后简单介绍相关工作简介.

2.1 符号说明

本文中, 矩阵和向量分别用大写黑体和小写黑体表示. 对于矩阵 \mathbf{M} , $\text{rank}(\mathbf{M})$ 、 $\text{Tr}(\mathbf{M})$ 、 \mathbf{M}^T 和 \mathbf{M}^{-1} 分别表示 \mathbf{M} 的秩、迹、转置和逆. $\|\cdot\|_2$ 代表 l_2 -范数, $\|\cdot\|_{2,1}$ 代表 $l_{2,1}$ -范数, $\|\cdot\|_F$ 代表 Frobenius 范数. \mathbf{I} 代表单位阵, $\mathbf{1}$ 代表所有元素都为 1 的向量. $\mathbf{M} \geq 0$ 代表 \mathbf{M} 中的所有元素都大于或等于 0.

2.2 相似度图的构建

给定一个样本数量为 n 、维度为 d 的数据集 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$, 其中 $\mathbf{x}_i \in \mathbb{R}^{d \times 1}$ ($1 \leq i \leq n$) 表示第 i 个样本. 根据数据集 \mathbf{X} 构造一个无向加权图 $G = (\mathbf{X}, \mathbf{S})$, 其中, $\mathbf{S} \in \mathbb{R}^{n \times n}$ 表示相似度矩阵, \mathbf{S} 中第 i 行第 j ($1 \leq j \leq n$) 列的元素 s_{ij} 代表样本 \mathbf{x}_i 和 \mathbf{x}_j 之间的相似度. 基于图的方法, 依据矩阵 \mathbf{S} 所编码的样本对之间的相似度进行学习.

k 近邻图(稀疏图)是定义图结构最常用的方法, 因其优秀性能已成功应用于几乎所有基于图的机器学习问题. 在给定的度量(通常是欧几里得距离)中, k 近邻图首先基于完全连接的距离矩阵搜索与 \mathbf{x}_i 相距最近的 k 个邻居样本, 然后根据具体的构图策略定义相似度图. k 最近邻样本的选择可以帮助在计算相似度矩阵时规避离群值或异常值会引入的干扰信息. 但 k 近邻图不考虑数据分布或采样密度, 对所有样本使用固定大小的邻域可能不够合理.

常用构图策略有: 二元构图、余弦构图和高斯核构图. 二元构图策略最简单, 将相似度定义为

$$s_{ij} = \begin{cases} 1, & \mathbf{x}_i \in \text{knn}(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in \text{knn}(\mathbf{x}_i) \\ 0, & \text{其他} \end{cases} \quad (1)$$

其中, $\text{knn}(\mathbf{x}_i)$ 表示 \mathbf{x}_i 的 k 近邻集合. 二元构图中非 0 即 1 的相似度, 可解释性和表示性非常弱, 不利于分析数据结构.

余弦构图策略将相似度定义为

$$s_{ij} = \begin{cases} \frac{\mathbf{x}_i^T \mathbf{x}_j}{\|\mathbf{x}_i\| \|\mathbf{x}_j\|}, & \mathbf{x}_i \in \text{knn}(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in \text{knn}(\mathbf{x}_i) \\ 0, & \text{其他} \end{cases} \quad (2)$$

其中, $\|\cdot\|$ 是向量的模. 余弦相似度只关注方向的差异, 对具体数值的大小不敏感.

高斯核构图策略是定义相似度最流行的方法, 将相似度定义为

$$s_{ij} = \begin{cases} \exp\left(-\frac{d_{ij}^2}{2\sigma^2}\right), & \mathbf{x}_i \in \text{knn}(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in \text{knn}(\mathbf{x}_i) \\ 0, & \text{其他} \end{cases} \quad (3)$$

其中, $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2$, σ 是带宽参数. σ 为适应数据

分布提供了灵活性, 但也引起了参数选择问题. σ 需要被调节以选择适当的值^[19].

2.3 图嵌入

图嵌入作为降维技术的一部分, 包含两个独立过程: (1) 构建相似度矩阵 \mathbf{S} ; (2) 基于 \mathbf{S} 探索低维子空间. 主要思想是, 低维子空间中的样本应尽可能保持原始高维空间中样本之间的相似度. 即较大的相似度 s_{ij} 要求低维表示 \mathbf{f}_i 和 \mathbf{f}_j 之间距离较小, 反之亦然. 因此, 图嵌入解决以下问题:

$$\min_{\mathbf{F}^T \mathbf{Q} \mathbf{F} = \mathbf{I}} \sum_{i,j=1}^n \|\mathbf{f}_i - \mathbf{f}_j\|_2^2 s_{ij} \Rightarrow \min_{\mathbf{F}^T \mathbf{Q} \mathbf{F} = \mathbf{I}} \text{Tr}(\mathbf{F}^T \mathbf{L}_S \mathbf{F}) \quad (4)$$

其中, $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_n]^T \in \mathbb{R}^{n \times m}$ 是低维表示 ($m \ll d$); $\mathbf{Q} \in \mathbb{R}^{n \times n}$ 是第 i 个对角元素为 $d_{ii} = \sum_j s_{ij}$ 的对角度矩阵 \mathbf{D} 或单位阵; \mathbf{L}_S 是图拉普拉斯矩阵, 被定义为 $\mathbf{L}_S = \mathbf{D} - \mathbf{S}$. \mathbf{L}_S 是半正定的, 最小特征值为 0, 0 特征值的数量对应着 \mathbf{L}_S 中连通域的数量. 目标函数

$$\min_{\mathbf{F}^T \mathbf{F} = \mathbf{I}} \text{Tr}(\mathbf{F}^T \mathbf{L}_S \mathbf{F}) = \min_{\mathbf{f}_i^T \mathbf{f}_i = 1} \sum_{i=1}^n \mathbf{f}_i^T \mathbf{L}_S \mathbf{f}_i$$

中的每一个优化子目标是 $\min_{\mathbf{f}_i^T \mathbf{f}_i = 1} \mathbf{f}_i^T \mathbf{L}_S \mathbf{f}_i$, $\min_{\mathbf{f}_i^T \mathbf{f}_i = 1} \mathbf{f}_i^T \mathbf{L}_S \mathbf{f}_i$ 的解在 \mathbf{L}_S 的最小非零特征值对应的特征向量上取得. 当数据集 \mathbf{X} 中存在大量噪声, 相似度矩阵 \mathbf{S} 中会包含大量无关相似性信息, 进而导致 \mathbf{L}_S 的质量低下.

2.4 LE 和 LPP 的工作原理

LE 认为, 若 \mathbf{x}_i 和 \mathbf{x}_j 相似, 则它们的低维表示应尽量接近^[6]. LE 首先构造 k -近邻 \mathbf{S} , 然后求解以下最小化问题得到低维表示 \mathbf{F} :

$$\min_{\mathbf{F}^T \mathbf{F} = \mathbf{I}} \text{Tr}(\mathbf{F}^T \mathbf{L}_S \mathbf{F}) \quad (5)$$

LE 是一个非线性方法, 无法处理离群值和新数据.

LPP 被提出以解决 LE 遇到的问题. LPP 假定 \mathbf{X} 到 \mathbf{F} 的映射是线性的^[8], 对式(5)施加严格的线性约束 $\mathbf{F} = \mathbf{X}^T \mathbf{W}$ ($\mathbf{W} \in \mathbb{R}^{d \times m}$), 求解以下问题:

$$\min_{\mathbf{W}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{W} = \mathbf{I}} \text{Tr}(\mathbf{W}^T \mathbf{X} \mathbf{L}_S \mathbf{X}^T \mathbf{W}) \quad (6)$$

LPP 可以被看作是 LE 的线性近似. 对于训练样本之外的新数据, LPP 是一种成功的方法.

在 LE 和 LPP 中, 相似度矩阵 \mathbf{S} 都是预先构造的, 在低维子空间的探索过程中固定不变. 这种情况下, 因噪点或离群值引起的不可靠相似度也被用做探索低维子空间的依据, 不被修正, 导致最终得到的低维表示的质量无法保证.

2.5 EGCFS 和 UPGO 的工作原理

在近两年的工作中, Zhang 等人^[17]提出的 EGCFS 方法将图学习纳入优化过程, 利用嵌入图学习和约束来选择无关但有区别的特征:

$$\min_{\mathbf{W}, \mathbf{S}} \text{Tr}(\mathbf{W}^T \mathbf{X} \mathbf{L}_s \mathbf{X}^T \mathbf{W}) + \gamma \|\mathbf{S}\|_F^2 - \lambda \text{Tr}(\mathbf{W}^T \mathbf{X} \mathbf{G} (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{X}^T \mathbf{W}) + \alpha \|\mathbf{W}\|_{2,1} \quad (7)$$

其中, \mathbf{G} 表示离散的聚类指示矩阵. 相似度矩阵 \mathbf{S} 在算法优化过程中动态更新.

Dong 等人^[18] 提出的 UPGO 方法将相似度图的构建和投影矩阵的学习统一到一个框架:

$$\min_{\mathbf{W}, \mathbf{S}} \|\mathbf{S} - \mathbf{A}\|_F^2 + \alpha \sum_{i,j=1}^n \|\mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j\|_2^2 s_{ij}$$

$$\text{s. t. } \mathbf{S} \mathbf{1} = \mathbf{1}, \mathbf{S} \geq 0, \text{rank}(\mathbf{L}_s) = n - c \quad (8)$$

UPGO 在降维过程中同时学习相似度矩阵 \mathbf{S} 和投影矩阵 \mathbf{W} .

不同于 LE 和 LPP, EGCFS 和 UPGO 动态学习相似度矩阵. 不可忽视的是, 随着数据收集方法以及数据表达方式的增多, 真实世界中出现的数据多种多样, 数据的分布规则无法一概而论. 显而易见, LPP, EGCFS 和 UPGO 所依赖的线性投影关系对于真实数据要求过于严格. 基于图学习的降维方法有待进一步探索.

3 本文方法

本文方法 FAUDR 的目标是高质量地学习原始高维空间无标签数据的低维表示. FAUDR 希望在低维子空间的探索过程中学习得到一个理想的相似度图. 此相似度图可以更加准确地表示样本对之间的内在统计特性和数据结构. 因此, 不同于当前已有的基于图的降维方法采用“先构图, 然后基于预先构建的固定的图进行子空间探索”两个完全独立的过程. FAUDR 将相似度图构建和低维表示学习结合到同一个模型中. 在降维任务中, 同时优化相似度图和低维表示.

一般情况下, 相距更近的数据点之间更有可能具有更大的相似度, 反之亦然. 也就是说相似度 s_{ij} 与样本 \mathbf{x}_i 和 \mathbf{x}_j 之间的距离成反比例关系. 因此, FAUDR 在降维过程中, 结合原始高维空间样本 \mathbf{X}

和学到的低维表示 \mathbf{F} , 自适应地为每个数据点分配 k -最近邻居, 从而得到结构良好的最优相似度矩阵 \mathbf{S} . 问题表述为

$$\min_{\mathbf{S} \mathbf{1} = \mathbf{1}, \mathbf{S} \geq 0, \mathbf{F}} \sum_{i,j=1}^n (\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 s_{ij} + \gamma s_{ij}^2) + \lambda_1 \|\mathbf{f}_i - \mathbf{f}_j\|_2^2 s_{ij} \quad (9)$$

其中, γ 是正则化参数, $\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 s_{ij}$ 用以探索原始数据的相似性结构, 保证相距较远的样本之间相似度较小, 相距较近的样本之间相似度较大; 平方项 γs_{ij}^2 存在的意义是避免平凡解; $\mathbf{S} \mathbf{1} = \mathbf{1}$ 代表 \mathbf{S} 的行和为 1. λ_1 是均衡参数, 用以平衡原始高维样本和降维过程中得到的低维表示对相似度矩阵的贡献. 在式(7)中, \mathbf{S} 不仅依赖于原始高维样本之间的距离 $\|\mathbf{x}_i - \mathbf{x}_j\|_2^2$, 并且依赖于低维表示之间的距离 $\|\mathbf{f}_i - \mathbf{f}_j\|_2^2$. \mathbf{S} 的构建同时考虑了原始样本空间以及低维空间中样本间的邻接关系.

根据拉普拉斯矩阵的性质^[20], 式(9)被重写为

$$\min_{\mathbf{S} \mathbf{1} = \mathbf{1}, \mathbf{S} \geq 0, \mathbf{F}} \sum_{i,j=1}^n (\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 s_{ij} + \gamma s_{ij}^2) + 2\lambda_1 \text{Tr}(\mathbf{F}^T \mathbf{L}_s \mathbf{F}) \quad (10)$$

其中, $\mathbf{L}_s = \mathbf{D} - \mathbf{S}$ 是图拉普拉斯矩阵, $\mathbf{D} \in \mathbb{R}^{n \times n}$ 是第 i 个对角元素为 $d_{ii} = \sum_j s_{ij}$ 的对角度矩阵.

为更好地处理噪点、离群值以及新数据, 一般定义 $\mathbf{F} = \mathbf{X}^T \mathbf{W} + \mathbf{1} \mathbf{b}^T$, 其中, $\mathbf{b} \in \mathbb{R}^{m \times 1}$ 表示偏置向量. 此线性投影关系无法处理非线性分布的数据, 对真实情景要求太过于严格. 为松弛 $\mathbf{F} = \mathbf{X}^T \mathbf{W} + \mathbf{1} \mathbf{b}^T$ 这种从高维空间到低维空间严格的线性投影关系, 本文方法引入一个回归残差 $\mathbf{F}_0 \in \mathbb{R}^{m \times n}$, 并假设 $\mathbf{F}_0 = \mathbf{X}^T \mathbf{W} + \mathbf{1} \mathbf{b}^T - \mathbf{F}$. 此时, 低维表示 \mathbf{F} 和线性投影 $\mathbf{X}^T \mathbf{W} + \mathbf{1} \mathbf{b}^T$ 之间的偏差被很好地建模, 以适应非线性投影数据的失配性. 基于此, 得到了一个回归项:

$$\min_{\mathbf{W}^T \mathbf{W} = \mathbf{I}, \mathbf{b}, \mathbf{F}} \|\mathbf{X}^T \mathbf{W} + \mathbf{1} \mathbf{b}^T - \mathbf{F}\|_F^2 \quad (11)$$

低维表示 \mathbf{F} 和原数据 \mathbf{X} 之间不是线性关系.

FAUDR 旨在同时找到最佳的相似度矩阵 \mathbf{S} 、低维表示 \mathbf{F} 和回归项(如图 1 所示). 因此, 结合式(10)

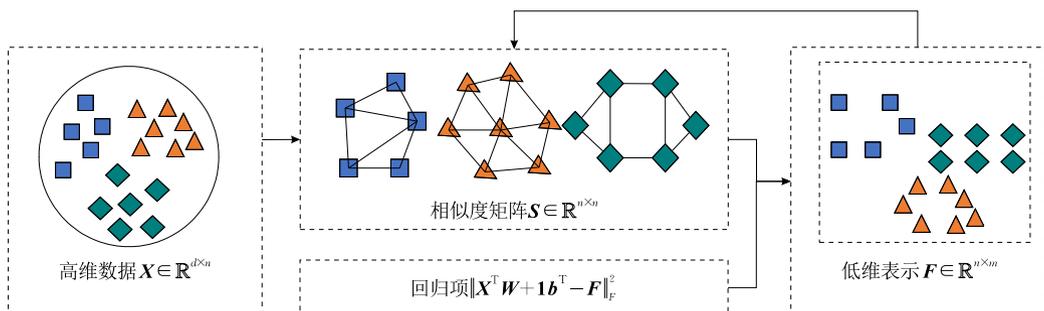


图 1 FAUDR 概述

和式(11), 最终的目标函数为

$$\min_{\substack{\mathbf{S}^T \mathbf{1} = \mathbf{1}, \mathbf{S} \geq 0, \mathbf{F}, \\ \mathbf{W}^T \mathbf{W} = \mathbf{I}, b}} \sum_{i,j=1}^n (\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 s_{ij} + \gamma s_{ij}^2) + 2\lambda_1 \text{Tr}(\mathbf{F}^T \mathbf{L}_S \mathbf{F}) + \lambda_2 \|\mathbf{X}^T \mathbf{W} + \mathbf{1}b^T - \mathbf{F}\|_F^2 \quad (12)$$

其中, λ_1 和 λ_2 是均衡参数.

在式(12)的目标函数中, 相似度矩阵 \mathbf{S} 不再被固定, 而是一个变量. 在降维过程中, \mathbf{S} 依据原始高维样本和低维表示自适应地优化. 低维表示 \mathbf{F} 不是由拉普拉斯矩阵的主特征向量组成, 避免了不容易处理噪声数据的缺点; 同时, \mathbf{F} 与 \mathbf{X} 也不是直接的线性关系, 规避了对数据分布过于严格的线性投影要求. 因此, 增强了处理不同类型或不同分布数据的灵活性.

4 方法优化

式(12)中的目标函数是一个多变量问题. 由于直接求解所有变量非常困难, 因此本文采用交替迭代算法逐一更新所有变量. 初始化 \mathbf{S} 和 \mathbf{F} 之后, 通过以下两个步骤更新所有变量:

4.1 固定 \mathbf{F} , 更新 \mathbf{S}

当 \mathbf{F} 被固定, 求解 \mathbf{S} 的子问题为

$$\min_{\substack{\mathbf{S}_i^T \mathbf{1} = \mathbf{1}, s_{ij} \geq 0}} \sum_{i,j=1}^n ((\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 s_{ij} + \gamma s_{ij}^2) + \lambda_1 \|\mathbf{f}_i - \mathbf{f}_j\|_2^2 s_{ij}) \quad (13)$$

注意到上式中的问题在不同的 i 之间是独立的, 因此, 对于每一个 i , 可以独立求解如下问题:

$$\min_{\substack{\mathbf{S}_i^T \mathbf{1} = \mathbf{1}, s_{ij} \geq 0}} \sum_{i=1}^n ((\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 s_{ij} + \gamma s_{ij}^2) + \lambda_1 \|\mathbf{f}_i - \mathbf{f}_j\|_2^2 s_{ij}) \quad (14)$$

为方便表达, 令 $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2^2 + \lambda_1 \|\mathbf{f}_i - \mathbf{f}_j\|_2^2$, 式(14)为

$$\min_{\substack{\mathbf{S}_i^T \mathbf{1} = \mathbf{1}, s_{ij} \geq 0}} \gamma \sum_{i=1}^n \left(s_{ij}^2 + \frac{1}{\gamma} d_{ij} s_{ij} \right) \quad (15)$$

为式(15)添加一个常数项 $\sum_{i=1}^n \frac{1}{4\gamma^2} d_{ij}^2$, 可以得到

$$\min_{\substack{\mathbf{S}_i^T \mathbf{1} = \mathbf{1}, s_{ij} \geq 0}} \gamma \sum_{i=1}^n \left(s_{ij} + \frac{1}{2\gamma} d_{ij} \right)^2 \quad (16)$$

定义向量 $\mathbf{d}_i \in \mathbb{R}^{n \times 1}$, 其第 j 个元素为 d_{ij} , 则式(16)等价于如下向量形式:

$$\min_{\substack{\mathbf{S}_i^T \mathbf{1} = \mathbf{1}, s_{ij} \geq 0}} \frac{1}{2} \left\| \mathbf{s}_i + \frac{1}{2\gamma} \mathbf{d}_i \right\|_2^2 \quad (17)$$

此问题的拉格朗日函数被定义为

$$L(\mathbf{s}_i, \tau, \boldsymbol{\zeta}) = \frac{1}{2} \left\| \mathbf{s}_i + \frac{1}{2\gamma} \mathbf{d}_i \right\|_2^2 - \tau (\mathbf{s}_i^T \mathbf{1} - 1) - \boldsymbol{\zeta}^T \mathbf{s}_i \quad (18)$$

其中, $\tau > 0$ 和 $\boldsymbol{\zeta} \geq 0$ 为拉格朗日乘子. 将 $L(\mathbf{s}_i, \tau, \boldsymbol{\zeta})$ 关于 \mathbf{s}_i 求导并令其为零, 可得

$$\mathbf{s}_i + \frac{1}{2\gamma} \mathbf{d}_i - \tau \mathbf{1} - \boldsymbol{\zeta} = 0 \quad (19)$$

根据 KKT 条件^[21], 注意到 $s_{ij} \zeta_j = 0$, 因此, 有

$$s_{ij} = \left(-\frac{1}{2\gamma} d_{ij} + \tau \right)_+ \quad (20)$$

本文方法学习稀疏的 \mathbf{s}_i , 即与 \mathbf{x}_i 相距最近的 k 个近邻点才有机会与其相连, 这代表 \mathbf{s}_i 含有 k 个非零元素. 不失一般性, 假设 \mathbf{d}_i 中的元素升序排列, 则 \mathbf{s}_i 含有 k 个非零元素意味着 $s_{ij} > 0 (1 \leq j \leq k)$, $s_{ij} = 0 (k+1 \leq j)$. 结合式(20), 可知

$$\begin{cases} -\frac{1}{2\gamma} d_{ik} + \tau > 0 \\ -\frac{1}{2\gamma} d_{i,k+1} + \tau \leq 0 \end{cases} \quad (21)$$

由式(19)和 $\mathbf{s}_i^T \mathbf{1} = 1$, 可得

$$\tau = \frac{1}{k} \left(1 + \frac{1}{2\gamma} \sum_{j=1}^k d_{ij} \right) \quad (22)$$

结合式(21)和式(22), 进一步有

$$\begin{cases} \gamma > \frac{1}{2} \left(kd_{ik} - \sum_{j=1}^k d_{ij} \right) \\ \gamma \leq \frac{1}{2} \left(kd_{i,k+1} - \sum_{j=1}^k d_{ij} \right) \end{cases} \quad (23)$$

为确保 \mathbf{s}_i 的最优解仅包含 k 个非零元素, γ 取值为

$$\gamma = \frac{1}{2} \left(kd_{i,k+1} - \sum_{j=1}^k d_{ij} \right) \quad (24)$$

结合式(20)、(22)和式(24), 可求解得

$$s_{ij} = \begin{cases} \frac{d_{i,k+1} - d_{ij}}{k}, & j \leq k \\ kd_{i,k+1} - \sum_{j=1}^k d_{ij}, & j = k+1 \\ 0, & j \geq k+1 \end{cases} \quad (25)$$

此时, \mathbf{S} 更新完毕.

4.2 固定 \mathbf{S} , 更新 \mathbf{F} 和 \mathbf{W}

当 \mathbf{S} 被固定, 式(12)变成:

$$\min_{\substack{\mathbf{F}, \mathbf{W}^T \mathbf{W} = \mathbf{I}, b}} 2\lambda_1 \text{Tr}(\mathbf{F}^T \mathbf{L}_S \mathbf{F}) + \lambda_2 \|\mathbf{X}^T \mathbf{W} + \mathbf{1}b^T - \mathbf{F}\|_F^2 \quad (26)$$

将式(26)关于 \mathbf{b} 求导并令导数为零, 可得

$$2nb + (\mathbf{X}^T \mathbf{W} - \mathbf{F})^T \mathbf{1} = 0 \quad (27)$$

进一步可以得到

$$\mathbf{b} = \frac{1}{n} (\mathbf{F}^T \mathbf{1} - \mathbf{W}^T \mathbf{X} \mathbf{1}) \quad (28)$$

将式(28)代入式(26), 有

$$\min_{\substack{\mathbf{F}, \mathbf{W}^T \mathbf{W} = \mathbf{I}}} 2\lambda_1 \text{Tr}(\mathbf{F}^T \mathbf{L}_S \mathbf{F}) + \lambda_2 \|\mathbf{H}(\mathbf{X}^T \mathbf{W} - \mathbf{F})\|_F^2 \quad (29)$$

其中, $\mathbf{H} = \mathbf{I} - 1/n \mathbf{1}\mathbf{1}^T$.

将式(29)关于 \mathbf{F} 求导并令导数为零,可得

$$2\lambda_1 \mathbf{L}_S \mathbf{F} + \lambda_2 \mathbf{H} \mathbf{F} - \mathbf{H} \mathbf{X}^T \mathbf{W} = 0 \quad (30)$$

$$\mathbf{F} = \mathbf{L}_0 \mathbf{H} \mathbf{X}^T \mathbf{W} \quad (31)$$

其中, $\mathbf{L}_0 = (2\lambda_1 / \lambda_2 \mathbf{L}_S + \mathbf{H})^{-1}$.

将式(31)代入式(29), \mathbf{W} 的子问题为

$$\min_{\mathbf{W}^T \mathbf{W} = \mathbf{I}} \text{Tr}(\mathbf{W}^T \mathbf{X}(\lambda_1 \mathbf{L}_1 + \lambda_2 \mathbf{L}_2) \mathbf{X}^T \mathbf{W}) \quad (32)$$

其中, $\mathbf{L}_1 = \mathbf{H} \mathbf{L}_0 \mathbf{L}_S \mathbf{L}_0 \mathbf{H}$, $\mathbf{L}_2 = (\mathbf{I} - \mathbf{L}_0 \mathbf{H}) \mathbf{H} (\mathbf{I} - \mathbf{L}_0 \mathbf{H})$. \mathbf{W} 的最优解由 $\mathbf{X}(\lambda_1 \mathbf{L}_1 + \lambda_2 \mathbf{L}_2) \mathbf{X}^T$ 的 m 个最小特征值对应的特征向量构成.

通过以上步骤迭代更新 \mathbf{S} 、 \mathbf{F} 和 \mathbf{W} , 最终可以得到 \mathbf{S} 、 \mathbf{F} 和 \mathbf{W} 的最优解. 观察式(17)和式(31)可知, \mathbf{S} 和 \mathbf{F} 相互作用、共同优化. 算法 1 总结了优化方法.

算法 1. FAUDR 的优化算法.

输入: 数据 \mathbf{X} , 低维子空间维度 m , 参数 λ_1 和 λ_2

输出: 低维表示 \mathbf{F}

1. 通过求解 $\min_{s_i^T \mathbf{1} = 1, s_i \geq 0} \sum_{i,j} (\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 s_{ij} + \gamma s_{ij}^2)$ 初始化相似度矩阵 \mathbf{S} , 用 \mathbf{L}_S 的 m 个最小特征值对应的特征向量初始化低维表示 \mathbf{F} , 并计算目标函数的目标值.
2. 通过求解式(17)按行更新相似度矩阵 \mathbf{S} .
3. 根据 $\mathbf{X}(\lambda_1 \mathbf{L}_1 + \lambda_2 \mathbf{L}_2) \mathbf{X}^T$ 的 m 个最小特征值对应的特征向量更新投影矩阵 \mathbf{W} .
4. 依据式(31)更新低维表示 \mathbf{F} .
5. 计算当前目标函数值. 若所有变量更新后目标值不再变化, 算法结束, 输出低维表示 \mathbf{F} ; 否则, 返回步骤 2.

5 理论分析

5.1 收敛性分析

算法 1 的收敛性由定理 1 给出.

定理 1. 算法 1 中的变量更新规则在每次迭代中减小式(12)的目标函数值, 直到收敛.

证明. 假设第 t 次迭代后的变量为 \mathbf{S}^t 、 \mathbf{W}^t 和 \mathbf{F}^t . 在 $t+1$ 次迭代中, 当 \mathbf{F} 固定为 \mathbf{F}^t , 最优的 \mathbf{S}^{t+1} 通过求解式(13)得到

$$\mathbf{S}^{t+1} = \min_{\mathbf{S}^T \mathbf{1} = \mathbf{1}, \mathbf{S} \geq 0} \sum_{i,j=1}^n (\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 s_{ij} + \gamma s_{ij}^2) + 2\lambda_1 \text{Tr}((\mathbf{F}^t)^T \mathbf{L}_S \mathbf{F}^t) \quad (33)$$

因此,

$$\begin{aligned} & \sum_{i,j=1}^n (\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 (s_{ij})^{t+1} + \gamma (s_{ij})^{t+1}) + \\ & 2\lambda_1 \text{Tr}((\mathbf{F}^t)^T \mathbf{L}_S \mathbf{F}^t) \leq \\ & \sum_{i,j=1}^n (\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 (s_{ij})^t + \gamma (s_{ij})^t) + \\ & 2\lambda_1 \text{Tr}((\mathbf{F}^t)^T \mathbf{L}_S \mathbf{F}^t) \end{aligned} \quad (34)$$

基于 \mathbf{S}^{t+1} , 解下式得到 \mathbf{W}^{t+1} 和 \mathbf{F}^{t+1} :

$$\begin{aligned} \langle \mathbf{W}^{t+1}, \mathbf{F}^{t+1} \rangle = \min_{\mathbf{F}, \mathbf{W}^T \mathbf{W} = \mathbf{I}} & 2\lambda_1 \text{Tr}(\mathbf{F}^T \mathbf{L}_S \mathbf{F}) + \\ & \lambda_2 \|\mathbf{X}^T \mathbf{W} + \mathbf{1} \mathbf{b}^T - \mathbf{F}\|_F^2 \end{aligned} \quad (35)$$

从而, 有

$$\begin{aligned} & 2\lambda_1 \text{Tr}((\mathbf{F}^{t+1})^T \mathbf{L}_S \mathbf{F}^{t+1}) + \\ & \lambda_2 \|\mathbf{X}^T \mathbf{W}^{t+1} + \mathbf{1} \mathbf{b}^T - \mathbf{F}^{t+1}\|_F^2 \leq \\ & 2\lambda_1 \text{Tr}((\mathbf{F}^t)^T \mathbf{L}_S \mathbf{F}^t) + \\ & \lambda_2 \|\mathbf{X}^T \mathbf{W}^t + \mathbf{1} \mathbf{b}^T - \mathbf{F}^t\|_F^2 \end{aligned} \quad (36)$$

组合式(34)和式(36)并消去相同项, 可得

$$\begin{aligned} & \sum_{i,j=1}^n (\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 (s_{ij})^{t+1} + \gamma (s_{ij})^{t+1}) + \\ & 2\lambda_1 \text{Tr}((\mathbf{F}^{t+1})^T \mathbf{L}_S \mathbf{F}^{t+1}) + \\ & \lambda_2 \|\mathbf{X}^T \mathbf{W}^{t+1} + \mathbf{1} \mathbf{b}^T - \mathbf{F}^{t+1}\|_F^2 \leq \\ & \sum_{i,j=1}^n (\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 (s_{ij})^t + \gamma (s_{ij})^t) + \\ & 2\lambda_1 \text{Tr}((\mathbf{F}^t)^T \mathbf{L}_S \mathbf{F}^t) + \\ & \lambda_2 \|\mathbf{X}^T \mathbf{W}^t + \mathbf{1} \mathbf{b}^T - \mathbf{F}^t\|_F^2 \end{aligned} \quad (37)$$

注意到式(12)的目标值下界为 0. 因此, 算法 1 将在每次迭代过程中单调减小目标函数的目标值, 直到收敛.

5.2 计算复杂度

求解式(10)中目标函数的计算复杂度由三部分组成: (1) 更新 \mathbf{S} 的每一行花费 $O(n)$, 因此, 更新 \mathbf{S} 需要花费 $O(n^2)$; (2) 求解 \mathbf{W} 需要计算 $\mathbf{X}(\lambda_1 \mathbf{L}_1 + \lambda_2 \mathbf{L}_2) \mathbf{X}^T$ 的特征向量, 花费 $O(d^3)$; (3) 根据式(29), 求解 \mathbf{F} 的计算量是 $O(n^3)$. 通常, $m \ll n$ 且 $m \ll d$, 因此, 总的计算复杂度约为 $O((n^3 + d^3)t)$, 其中, t 是迭代次数. 此外, 在我们的实验中, 算法 1 在 30 次迭代之内即可收敛. 因此, 当 $d < n$ 时, 计算复杂度可近似为 $O(n^3)$, 否则为 $O(d^3)$.

5.3 存储复杂度

在低维表示的学习过程中, FAUDR 需要存储 \mathbf{S} 、 \mathbf{F} 和 \mathbf{W} . 因此, 算法的存储复杂度为 $O(n^2 + nm + dm)$. 由于 $m \ll n$ 且 $m \ll d$, 因此, 总的存储复杂度近似为 $O(n^2 + dm)$.

6 实验

本文在合成数据和基准数据集上进行了实验, 以综合验证所提出方法 FAUDR 的有效性. 所有实验均在配置为 2.20 GHz Intel Core i7-8750 CPU, 16 GB RAM 和 Matlab R2018a(64 bit) 的 Windows

10 便携式计算机上实施。

6.1 实验设置

为验证 FAUDR 的有效性, 我们将其与以下七种无监督降维方法进行比较。

(1) PCA^[9]: PCA 是使得平均投影代价最小的线性投影。

(2) LE^[6]: LE 是保留了近邻点相似性的一种经典的图嵌入方法。

(3) LPP^[8]: LPP 可以保持数据局部结构, 同时降低空间维度。

(4) 谱回归 (Spectral Regression, SR)^[22]: SR 将学习嵌入函数转换为一个回归框架。

(5) LSDUDR^[16]: LSDUDR 构造相似度图和多样性图来揭示数据的拓扑结构。

(6) GLSRGE^[13]: GLSRGET 最小化图嵌入中的最小二乘重建误差, 保留全局和局部低维子空间。

(7) 基于图优化的无监督投影 (Unsupervised Projection with Graph Optimization, UPGO)^[18]: UPGO 将线性投影和聚类结合, 同时进行降维和聚类的方法。

表 1 汇总了所有方法的计算复杂度。为公平比较, 这些方法的源代码均来自于可靠网站^{①②③④}或作者本人。对于每个比较方法, 我们按照相应论文中的实验设置和参数调整步骤进行实验。UPGO 中 α 使用搜索策略在区间 $\{10^{-5}, 10^{-4.9}, \dots, 10^{-4.1}\}$ 中进行调节。LSDUDR 中的 β 、GLSRGE 中的 α 以及本文方法 FAUDR 中的 λ_1 和 λ_2 使用网格搜索策略在区间 $\{10^{-3}, 10^{-2}, \dots, 10^3\}$ 中进行调节。本文实验记录来自最佳参数配置的结果。当两次迭代中目标函数值的差异小于 10^{-11} 时, 认为优化算法收敛, FAUDR 终止迭代。

表 1 所有方法的计算复杂度

方法	计算复杂度	方法	计算复杂度
PCA	$O(d^2n + d^3)$	LE	$O(n^2d + n^3)$
LPP	$O(n^2d + d^3)$	SR	$O(n^2d)$
LSDUDR	$O(n^3 + d^3)$	GLSRGE	$O(n^3 + d^3 + n^2d)$
UPGO	$O(n^3 + d^3)$	FAUDR	$O(n^3 + d^3)$

在实际进行降维实验时, 与当前已有的大多数无监督降维工作做法相同, 对于每种方法, 我们首先学习得到无标签高维数据的低维表示, 然后对低维表示执行 k -means^[23] 聚类算法来衡量降维性能。聚类结果采用三种流行的评估指标进行评估: 精度 (ACCuracy, ACC)^[24]、标准化互信息 (Normalized

Mutual Information, NMI)^[24] 和纯度 (Purity)^[25]。

ACC: 将 y_i 表示为使用聚类方法为样本 x_i 生成的聚类标签, g_i 表示为样本 x_i 的真实标签, ACC 通过下式计算:

$$ACC = \frac{\sum_i \delta(y_i, \text{map}(g_i))}{n} \quad (38)$$

其中, 若 $x = y$, 则 $\delta(x, y) = 1$, 否则 $\delta(x, y) = 0$, $\text{map}(g_i)$ 是可以将聚类标签 g_i 匹配到真实标签 y_i 最好的映射函数。 n 是样本总数。

NMI: 将 n_l 表示为通过聚类算法生成的第 l 个类 C_l 中的样本数量, n_h 表示为真实情况下第 h 个类 C_h 中的样本数量, n_{lh} 表示为在 C_l 和 C_h 中都出现的样本数量。 NMI 的计算方法如下:

$$NMI = \frac{\sum_l \sum_h n_{lh} \log \frac{n \times n_{lh}}{n_l n_h}}{\sqrt{\left(\sum_l n_l \log \frac{n_l}{n}\right) \left(\sum_h n_h \log \frac{n_h}{n}\right)}} \quad (39)$$

其中, c 是类别总数。

Purity: 除 ACC 和 NMI 外, Purity 是另一种普遍使用的评估指标。 Purity 衡量聚类簇包含主要来自一个类别的样本的程度, 是通过各个聚类簇纯度值的加权总和来观察的。如下所示:

$$Purity = \sum_{i=1}^c \frac{n_i}{n} P(C_i) \quad (40)$$

其中, $P(C_i) = \frac{1}{n_i} \max_j (n_{ij}^i)$, C_i 表示样本数为 n_i 的簇, n_{ij}^i 表示第 i 个簇包含的第 j 类样本的数量。

这三个指标各自具有不同的属性, 综合考虑可以更全面地评估结果。这三个指标的取值区间均为 $[0, 1]$, 值越大代表结果越好。

6.2 合成数据集上的实验

鉴于基准数据结构复杂多样, 且在基准数据集上模拟噪点和异常值比较困难, 为更直观地验证本文方法的性能, 我们使用人工合成的双月数据和瑞士卷数据集来展示本文方法的学习结果。

双月数据用以可视化本文方法学习得到的相似度矩阵。我们将噪声比例设置为 0.12, 分别合成了每类含有 100 个样本点, 且类间紧凑度不同的两个

① PCA. <https://ww2.mathworks.cn/help/stats/pca.html>

② LE. <https://github.com/topics/laplacian-eigenmaps>

③ LPP. <http://www.cad.zju.edu.cn/home/xiaofeihe/LPP.html>

④ SR. <http://www.cad.zju.edu.cn/home/dengcai/Data/SR.html>

原始双月数据,如图 2 中(a)和(f)所示. 图中不同颜色代表不同类别. 图 2 中,(b)和(g)分别是高斯核方法在两个原始数据上构造的相似矩阵.(c)和(h)分别是初始化方法在两个原始数据上构造的相似矩阵. 在这些图中,样本间连接线的粗细程度与对应的相似度大小成正比. 为验证 FAUDR 的相似度学习能力,我们先为最初生成的二维数据增加幅度为 0.1 的噪声;然后使用幅度为 0.5 的八维随机噪声将(a)和(f)中显示的两个二维原始数据分别扩展为两个十维数据;再使用 FAUDR 分别对两个十维数

据集进行降维.(d)和(i)在两个原始数据上分别展示了 FAUDR 最终学到的相似度矩阵.(e)和(j)展示了 FAUDR 最终学到的二维表示及相似度矩阵. 观察图 2 可知,高斯核方法和初始化方法构造的相似度矩阵将所有数据连接在一起,很难区分两个不同的簇. 本文方法最终学习到的相似度矩阵可以清楚地将两个簇分开,同时得到的低维表示有效地保持了原始数据的固有结构. 验证了 FAUDR 可以在噪声存在的情况下,自适应地学得结构良好的相似度矩阵.

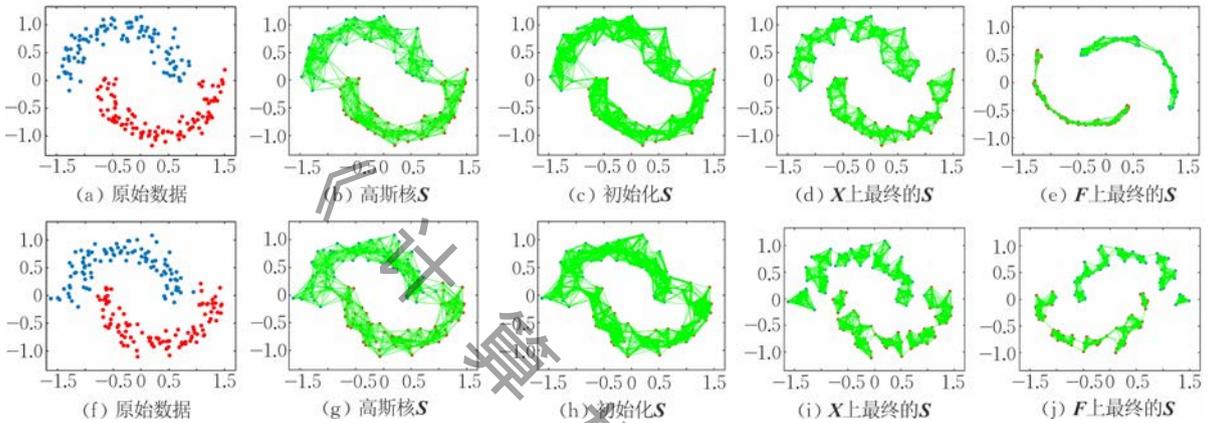


图 2 双月数据集上的实验结果((a)和(f)是两个原始双月数据,(b)和(g)分别展示了高斯核方法在两个原始数据上构造的相似度矩阵,(c)和(h)分别展示了在两个原始数据上初始化方法构造的相似度矩阵,(d)和(i)在两个原始数据上分别展示了 FAUDR 最终学到的相似度矩阵,(e)和(j)分别展示了 FAUDR 最终学到的低维表示和相似度矩阵)

瑞士卷数据集用于验证所有比较方法以及本文方法 FAUDR 对异常数据的处理能力. 原始瑞士卷数据是将噪声比例设置为 0.05 时生成的分为 4 类的 1500 个三维样本,如图 3 所示. 不同颜色的样本点代表不同类别. 我们使用不同幅度的噪声数据将此数据集扩展为两个十维数据集. 第一个是小噪声数据集,其前三个维度是原始的三维瑞士卷数据附加以幅度为 2 的随机噪声(如图 4(a)所示),剩余的七个维度是幅度为 2 的随机噪声. 第二个是大噪声数据集,其构成与第一个数据集相似,不同之处是噪声

幅度为 5,前三个维度如图 5(a)所示. 设置更大幅度噪声的目的是使样本分布更加混乱.

在此实验中,我们将 FAUDR 与前面介绍的 PCA、LE、LPP、SR、GLSRGE、LSDUDR 和 UPGO 七种方法进行比较,学习三维表示. 图 4 和图 5 分别展示了不同方法在小噪声数据集和大噪声数据集上的降维结果. 图 4 和图 5 尽可能以合适的角度清晰地展示三维表示. 图中,相同颜色代表同类数据. 对比观察可发现:当噪声幅度较小时,LE 学到的三维表示保持了类间的判别性,LPP、SR、GLSRGE 和 FAUDR 学到的三维表示都明确地保留了原始数据的结构;当噪声幅度增大时,LPP、SR 和 GLSRGE 的结果差强人意,FAUDR 学习得到的三维表示很好地保留了图 5(a)中原始前三维数据的分布特性和整体结构;无论噪声幅度大小,PCA 的结果均不理想. 此部分实验结果验证了 FAUDR 可以灵活地处理噪声数据,该模型对噪声数据具有较强的鲁棒性.

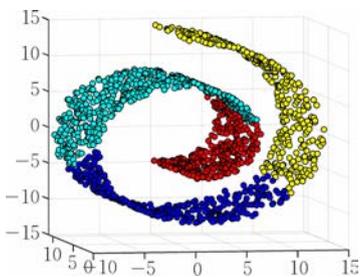


图 3 原始瑞士卷数据

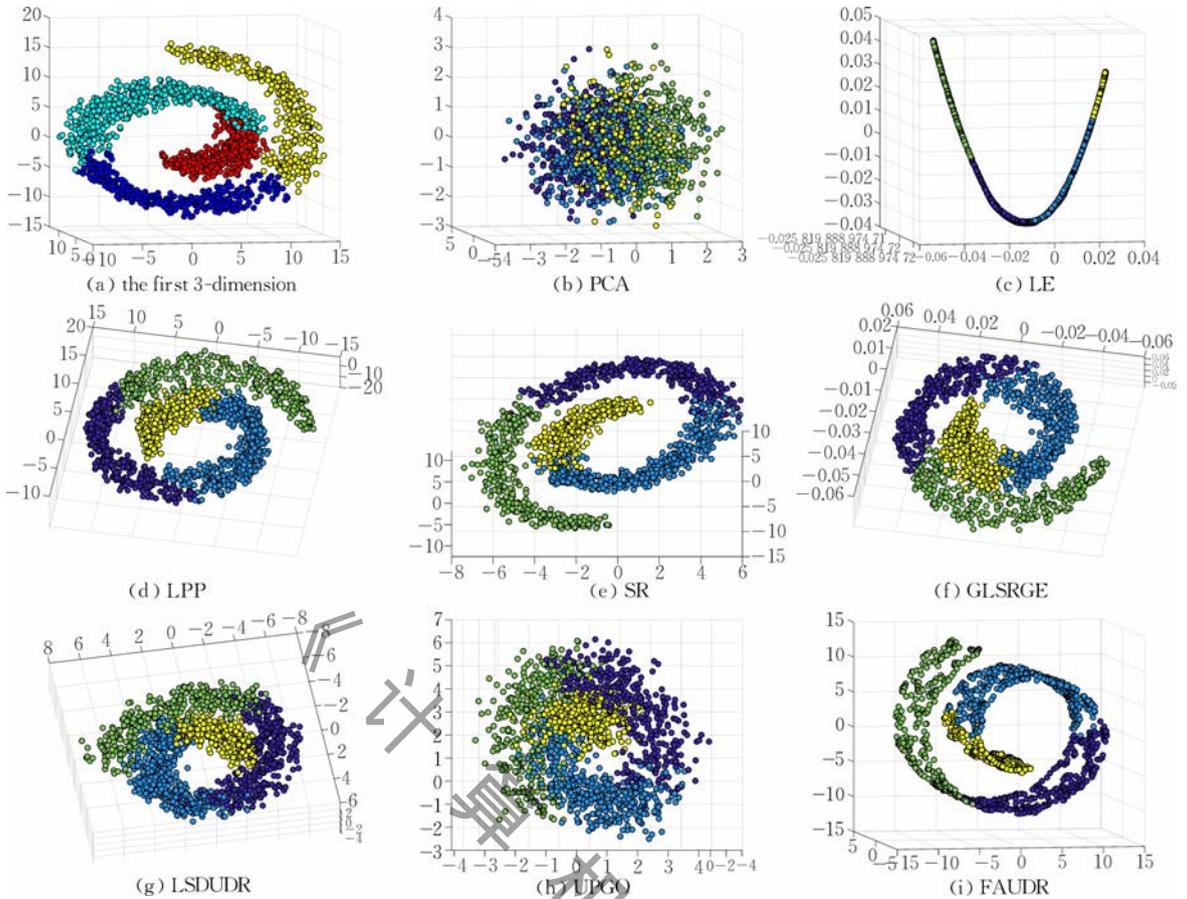


图 4 小噪声瑞士卷数据及各方法的降维结果((a)前三维小噪声瑞士卷数据;(b)~(i)分别展示了不同方法学得的三维表示)

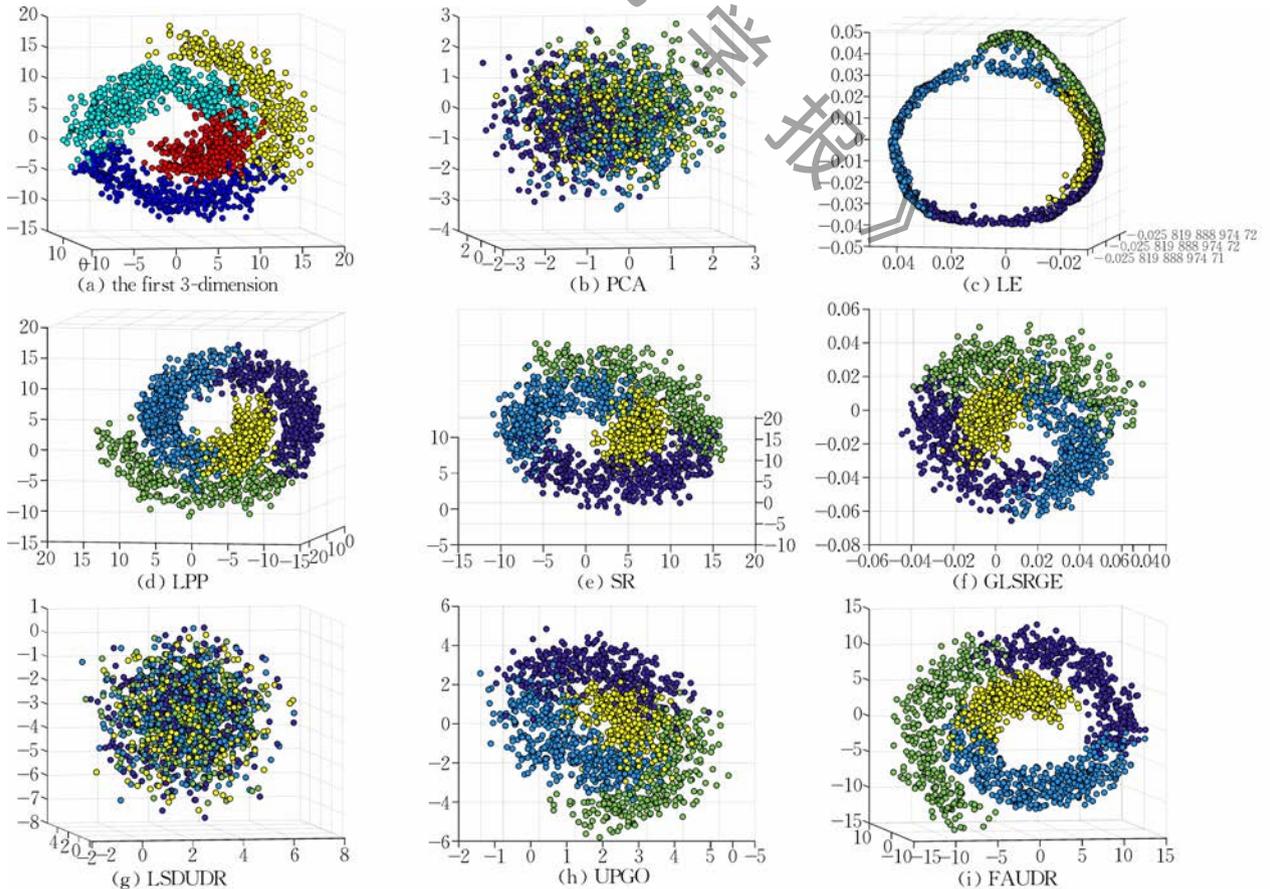


图 5 大噪声瑞士卷数据集及各方法的降维结果((a)前三维大噪声瑞士卷数据;(b)~(i)分别展示了不同方法学得的三维表示)

6.3 基准数据集上的实验

我们选择了样本个数从 1400 到 15 000、维度从 256 到 11 547、类别数从 2 到 100 不等的八个基准数据集,表 2 总结了这些数据集.其中,Mpeg7 和 Palm-Data25 的类别较多,Mpeg7、text1 和 Reuters 是大维度数据集.本部分从降维性能和参数敏感度分析两个方面分别进行实验.

表 2 基准数据集概述

数据集	样本数	类别数	维度
Mpeg7	1400	70	6000
MSRA25	1799	12	256
text1	1946	2	7155
PalmData25	2000	100	256
mnistdata_05	3495	10	748
20news	3970	20	8041
usps	9298	10	256
Reuters	15000	6	11547

6.3.1 降维性能实验

在比较各方法的降维性能实验中,为方便起见,我们将子空间的维度均设置为 $m=20$.在每个数据

集上,我们选用最佳参数配置对所有方法都重复进行十次实验以记录平均结果.

表 3、表 4 和表 5 分别记录了所有方法在所有数据集上分别使用精度、标准化互信息和纯度三个评估指标测得的均值和方差.其中,最优结果用粗体标注,次优结果用下划线标注以增强区分性.观察分析结果可知,与其他方法相比,本文方法 FAUDR 性能最佳.很明显,在维度为 1400 的 Mpeg7 数据集上,相比于第二好的结果,精度、标准化互信息和纯度分别提高了 4.52%、2.23% 和 5.06%;在维度为 1799 的 MSRA25 数据集上,FAUDR 的精度、标准化互信息和纯度分别比第二好的结果提高了 4.57%、2.22% 和 4.24%;在维度为 11 647 的 Reuters 数据集上,相比于次优结果,精度、标准化互信息和纯度分别提高了 6.49%、0.73% 和 5.20%.与此同时,FAUDR 降维结果所对应的方差也是所有方法中较小的.表 6 记录了所有方法在各数据集上的运行时

表 3 所有方法在基准数据集上的精度(平均精度(标准差))

数据集	PCA	LE	LPP	SR	GLSRGE	LSDUDR	UPGO	FAUDR
Mpeg7	0.5037 (0.0195)	0.4314 (0.0203)	0.4514 (0.0210)	0.4387 (0.0142)	0.5078 (0.0088)	0.5006 (0.0102)	0.3933 (0.0132)	0.5530 (0.0036)
MSRA25	0.4966 (0.0328)	0.4760 (0.0252)	0.5220 (0.0350)	0.4531 (0.0387)	0.5176 (0.0364)	0.5118 (0.0086)	0.5283 (0.0252)	0.5740 (0.0056)
text1	0.5038 (0.0018)	0.5398 (0.0062)	0.5401 (0.0523)	0.5159 (0.0192)	0.5770 (0.0919)	0.5057 (0.0042)	0.5141 (0.0064)	0.8990 (0.0018)
PalmData25	0.6583 (0.0233)	0.5353 (0.0167)	0.7685 (0.0250)	0.3664 (0.0117)	0.7351 (0.0144)	0.7197 (0.0220)	0.6136 (0.0240)	0.8010 (0.0117)
mnistdata_05	0.4612 (0.0241)	0.4962 (0.0681)	0.4817 (0.0237)	0.4980 (0.0233)	0.5486 (0.0239)	0.5549 (0.0453)	0.5119 (0.0177)	0.6320 (0.0450)
20news	0.2530 (0.0017)	0.2555 (0.0017)	0.2545 (0.0012)	0.2556 (0.0012)	0.2546 (0.0035)	0.2622 (0.0000)	0.2631 (0.0041)	0.3643 (0.0045)
usps	0.5851 (0.0150)	0.5262 (0.0249)	0.6355 (0.0499)	0.6736 (0.0396)	0.6748 (0.0426)	0.6728 (0.0493)	0.6577 (0.0209)	0.7280 (0.0493)
Reuters	0.3639 (0.0316)	0.3179 (0.0246)	0.3267 (0.0105)	0.3298 (0.0123)	0.3780 (0.0323)	0.3060 (0.0182)	0.3049 (0.0034)	0.4429 (0.0089)

表 4 所有方法在基准数据集上的标准化互信息(平均互信息(标准差))

数据集	PCA	LE	LPP	SR	GLSRGE	LSDUDR	UPGO	FAUDR
Mpeg7	0.6976 (0.0116)	0.6128 (0.0159)	0.6307 (0.0153)	0.6254 (0.0173)	0.7104 (0.0062)	0.6707 (0.0290)	0.5523 (0.0115)	0.7327 (0.0039)
MSRA25	0.5965 (0.0309)	0.6025 (0.0389)	0.6889 (0.0282)	0.5456 (0.0619)	0.5584 (0.0344)	0.5549 (0.0194)	0.5990 (0.0141)	0.7111 (0.0099)
text1	0.0040 (0.0028)	0.0465 (0.0078)	0.0466 (0.0078)	0.0286 (0.0232)	0.0648 (0.0095)	0.0005 (0.0000)	0.0008 (0.0007)	0.5275 (0.0057)
PalmData25	0.8683 (0.0099)	0.7818 (0.0215)	0.9327 (0.0074)	0.5989 (0.0072)	0.9022 (0.0053)	0.9123 (0.0052)	0.7879 (0.0125)	0.9458 (0.0029)
mnistdata_05	0.3999 (0.0166)	0.5147 (0.0499)	0.4636 (0.0157)	0.4731 (0.0178)	0.4902 (0.0268)	0.5799 (0.0145)	0.5032 (0.0230)	0.6857 (0.0274)
20news	0.0027 (0.0014)	0.0078 (0.0025)	0.0068 (0.0028)	0.0073 (0.0023)	0.0045 (0.0033)	0.0013 (0.0000)	0.0012 (0.0005)	0.1150 (0.0036)
usps	0.5551 (0.0177)	0.5719 (0.0331)	0.6622 (0.0206)	0.7244 (0.0170)	0.6518 (0.0236)	0.6478 (0.0054)	0.6372 (0.0238)	0.7780 (0.0150)
Reuters	0.1226 (0.0237)	0.0464 (0.0118)	0.0655 (0.0133)	0.0659 (0.0145)	0.1990 (0.0289)	0.0131 (0.0056)	0.0134 (0.0061)	0.2063 (0.0043)

表 5 所有方法在基准数据集上的纯度(平均纯度(标准差))

数据集	PCA	LE	LPP	SR	GLSRGE	LSDUDR	UPGO	FAUDR
Mpeg7	0.4787 (0.0153)	0.4679 (0.0150)	0.4907 (0.0163)	0.4864 (0.0103)	<u>0.5465</u> (0.0066)	0.5135 (0.0056)	0.4387 (0.0144)	0.5971 (0.0033)
MSRA25	0.5309 (0.0312)	0.5520 (0.0248)	0.5575 (0.0285)	0.4714 (0.0443)	0.5376 (0.0346)	0.5129 (0.0071)	<u>0.5668</u> (0.0199)	0.6092 (0.0073)
text1	0.5062 (0.0026)	0.5283 (0.0635)	0.5103 (0.0758)	0.5164 (0.0230)	<u>0.5770</u> (0.0092)	0.5062 (0.0000)	0.5145 (0.0058)	0.8983 (0.0017)
PalmData25	0.7015 (0.0125)	0.6095 (0.0148)	<u>0.8290</u> (0.0175)	0.4320 (0.0067)	0.7839 (0.0122)	0.7287 (0.0130)	0.6711 (0.0189)	0.8590 (0.0076)
mnistdata_05	0.5250 (0.0187)	<u>0.5994</u> (0.0469)	0.5207 (0.0184)	0.5682 (0.0206)	0.5788 (0.0321)	0.5954 (0.0237)	0.5671 (0.0266)	0.6558 (0.0197)
20news	0.2542 (0.0059)	0.2586 (0.0027)	0.2549 (0.0024)	0.2559 (0.0035)	0.2554 (0.0039)	<u>0.2627</u> (0.0000)	0.2643 (0.0040)	0.3756 (0.0038)
usps	0.6364 (0.0168)	0.5960 (0.0295)	0.6785 (0.0305)	<u>0.7302</u> (0.0216)	0.6980 (0.0344)	0.6821 (0.0082)	0.6523 (0.0323)	0.7691 (0.0212)
Reuters	0.3727 (0.0431)	0.3300 (0.0152)	0.3527 (0.0116)	0.3151 (0.0124)	<u>0.4620</u> (0.0299)	0.3153 (0.0178)	0.3034 (0.0120)	0.5140 (0.0031)

表 6 所有方法在基准数据集上的运行时间

(单位:s)

数据集	PCA	LE	LPP	SR	GLSRGE	LSDUDR	UPGO	FAUDR
Mpeg7	3.4981	3.5986	3.1837	1.4564	95.7750	2796.5524	81.4077	63.2307
MSRA25	2.2476	7.4237	2.5763	2.9418	3.8723	116.4175	6.6218	9.3061
text1	4.1372	2.1190	3.8703	2.5907	117.6029	3855.4282	110.6278	142.7056
PalmData25	1.4324	9.2785	1.8895	2.2525	3.5849	137.8905	4.5642	6.6039
mnistdata_05	2.8265	51.9408	2.5685	2.3541	13.3552	645.9738	19.5738	26.1045
20news	38.1946	71.5982	75.2340	7.2401	291.0818	5734.9136	220.8719	245.2751
usps	10.8856	814.5948	21.9702	20.5378	224.0438	7660.3317	235.5217	539.1236
Reuters	480.8326	924.0093	428.6214	428.8859	1748.0672	22789.8696	1516.2690	2581.6721

间. 观察可知,与其他方法相比,在取得了显著的性能提升之外,FAUDR 并没有大幅度增加运行时间. 证明了 FAUDR 的有效性. 综上所述,在真实情景中,本文方法是一个值得考虑的选择.

与此同时,我们还进行了维度评估:将每个数据集嵌入到维度从 10 到 100 的低维子空间中,步长为 10. 图 6 显示了所有方法在不同数据集上学习到的

不同维度低维表示时的聚类精度. 从图中可以看到,不同方法的降维结果随低维子空间维度变化的程度不一. 总的来说,在 MSRA25、PalmData25、mnistdata_05 和 usps 数据集上,LE 方法的降维结果变化较大;在 MSRA25、PalmData25 和 mnistdata_05 数据集上,SR 方法的降维结果不稳定;除 text1 和 20news 外,本文方法 FAUDR 的结果总体稳定,并

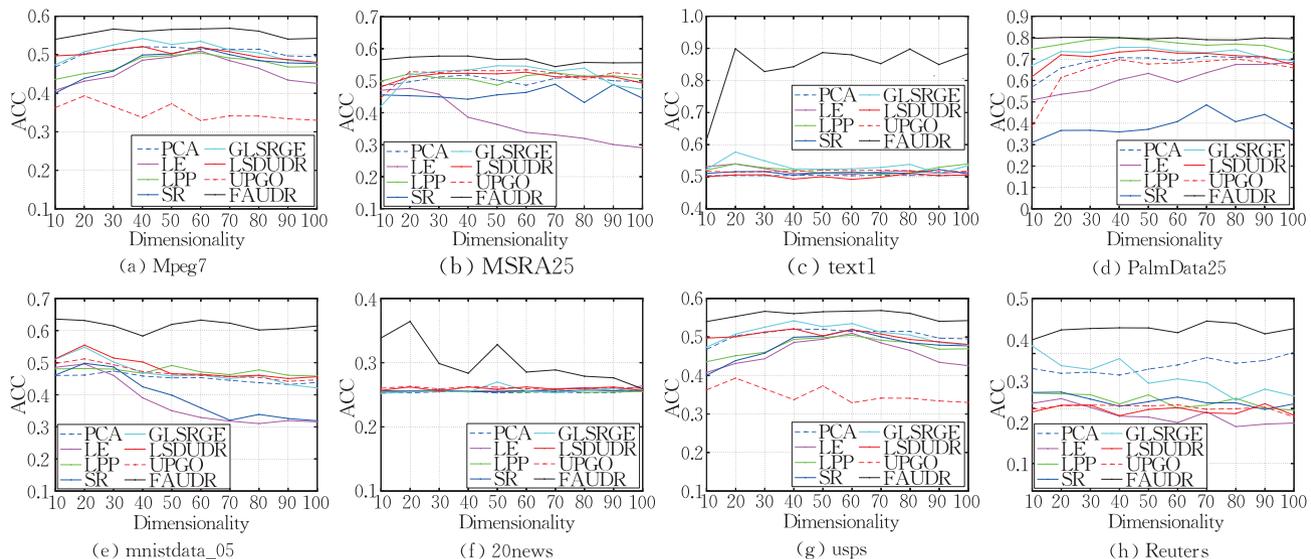


图 6 基准数据集上所有方法关于不同低维子空间的精度

且,在 text1 和 Reuters 数据集的不同子空间维度上性能提升明显.因此,本文方法可以获得相对令人满意的结果.

此外,我们将本文方法 FAUDR 与基于自编码器(autoencoder)的深度特征降维方法^[26]进比较.本部分实验使用大小为 28×28 的 70 000 张手写数字图像数据集 MNIST 进行降维,并可视化降维结果. autoencoder 的 matlab 代码和数据集均来源于作者提供的网站^①.我们使用由网络层数为 784-1000-500-250-2 的编码器和一个对称的解码器组成的 autoencoder 为 MNIST 训练集中的所有样本提取 2

维特征.除了 2 个线性单元外,其他所有单元均为 logistic.为清楚地展示实验结果,对所有 60 000 张训练图像进行训练之后,我们在剩余的 10 000 张图像的前五类样本中分别随机选择了 100 张,总共 500 张图像进行测试. FAUDR 直接学习这 500 张图像的 2 维表示.图 7 显示了实验结果.其中,图 7(a)和图 7(b)分别为 autoencoder 和 FAUDR 生成的 2 维表示.图中不同显色的点代表不同类别的 2 维样本.观察发现,图 7(a)中不同类别的样本分布混乱;图 7(b)中同类样本聚集在一起,不同的类簇区分明显.因此,FAUDR 产生了更好的降维结果.

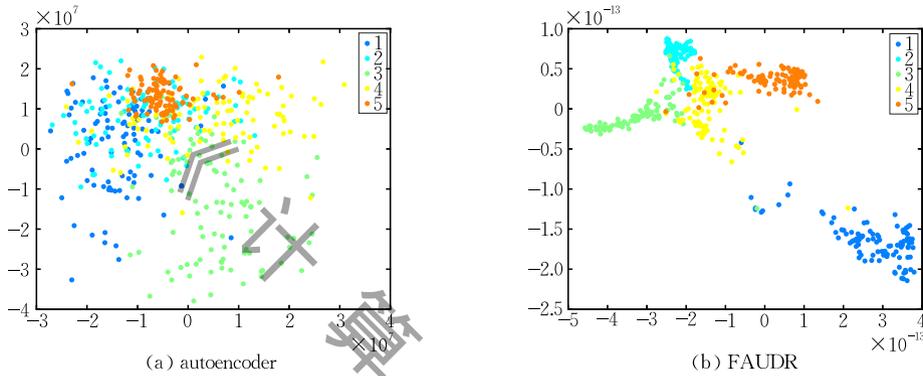


图 7 (a) autoencoder 生成的 500 个数字的 2 维表示, (b) FAUDR 学习得到的 500 个数字的 2 维表示

6.3.2 参数敏感性实验

FAUDR 的目标函数中有两个均衡参数 λ_1 和 λ_2 需要调节.参数敏感性分析实验中,我们在所有数据集上使用前文提到的网格调参策略来验证 λ_1 和 λ_2 分别取 $\{10^{-3}, 10^{-2}, \dots, 10^3\}$ 时的敏感性.图 8 显示了本文方法 FAUDR 在不同数据集上的不同参数配置下学得的低维表示的聚类精度.

观察分析图 8 可知,在不同数据集上,低维表示的聚类精度随着参数配置的不同而变化.在 Mpeg7、MSRA25 和 PalmData25 数据集上,聚类精度随着参数取值变化较小、整体稳定.在 text1、mnistdata_05 和 usps 数据集上,聚类精度随着参数取值变化较大;在 text1 数据集上, λ_1 取值为 10^{-1} 、 λ_2 取值为 10^1 时,聚类精度最高;在 20news 数据集上, λ_1 取值为

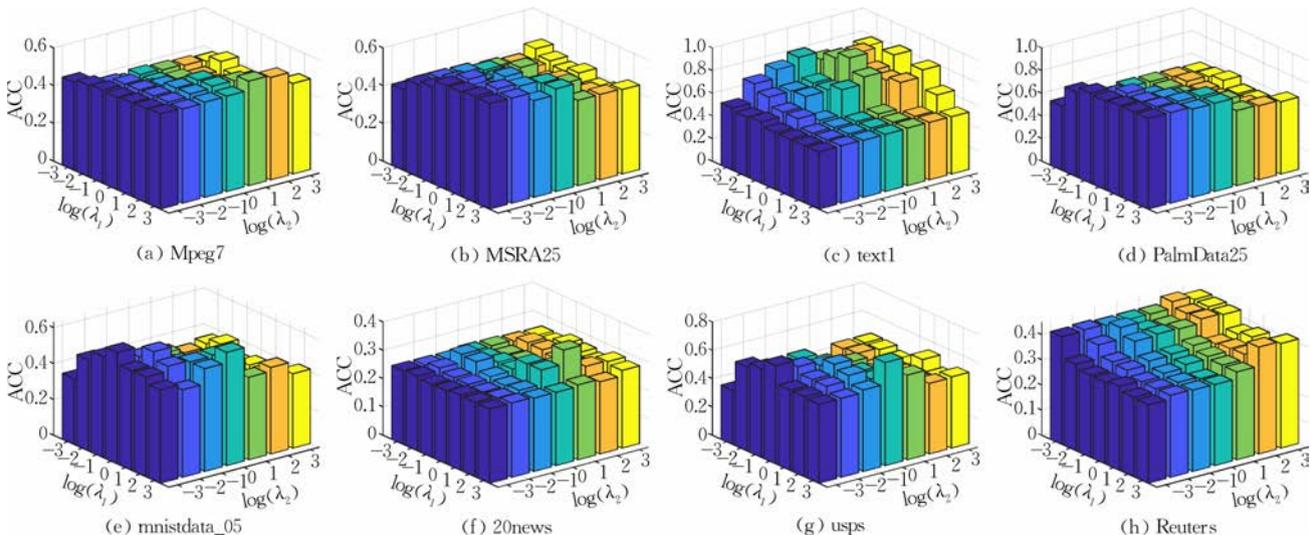


图 8 FAUDR 在不同 λ_1 和 λ_2 值下学得的低维表示的聚类精度(x 、 y 和 z 轴分别代表 $\log(\lambda_1)$ 、 $\log(\lambda_2)$ 和精度)

① <http://www.cs.toronto.edu/~hinton/MatlabForSciencePaper.html>

10^{-2} 、 λ_2 取值为 10^1 时, 聚类精度最高; 在 usps 数据集上, λ_1 取值为 10^3 、 λ_2 取值为 10^0 时, 聚类精度最高; 在 Reuters 数据集上, λ_1 取值在 $\{10^{-3}, 10^{-2}, 10^{-1}\}$ 内时, 聚类精度整体偏高. 以上分析表明, 在实际应用中工作时, 需要为具体任务选择最佳的参数配置.

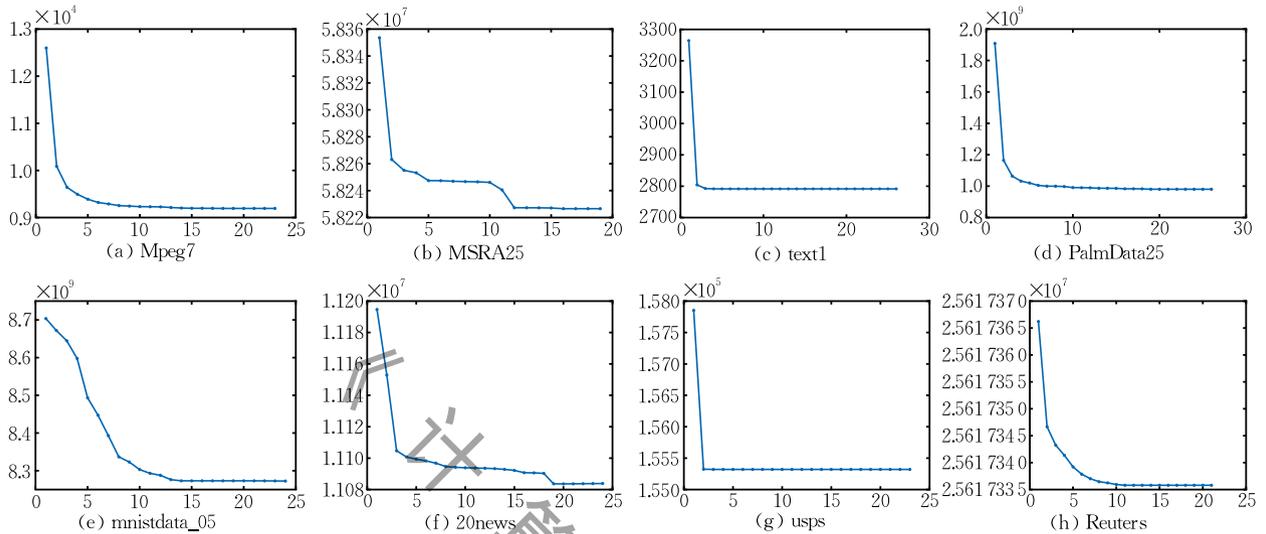


图 9 FAUDR 的目标值收敛曲线 (x 和 y 轴分别代表迭代次数和目标值)

7 结 论

为了更有效地将未标记的高维数据压缩到具有代表性的低维子空间中, 本文提出了一种灵活自适应的无监督降维方法 FAUDR. 引入的松弛回归项有助于 FAUDR 在实际应用中更灵活地处理不同种类的高维数据. 在探索低维子空间的过程中, 同时在考虑保持原始样本空间的数据结构和动态变化的低维子空间中数据点之间的最近邻关系的基础上, 自适应地学习相似度矩阵. 此相似度矩阵有助于提升降维性能. 在合成数据上的实验结果表明, 本文所提出的方法可以学习得到结构良好的相似度矩阵, 并且在噪声影响的情况下也可以获得保持了原始数据结构的低维表示. 基准数据集上的实验结果显示, 在对维度为千量级和万量级的数据分别进行降维时, FAUDR 所得到的降维结果的聚类精度比七个对比方法至少提升了 4.52 个百分点. 实验结果清楚地证明了本文方法优于其他比较方法.

参 考 文 献

[1] Ben-Bassat M. Pattern recognition and reduction of dimen-

6.3.3 收敛性实验

算法 1 交替迭代地更新所有变量. 此处将目标函数的值可视化, 以实验证明 FAUDR 的收敛性. 不同数据集上目标值的收敛曲线如图 9 所示. 从图 9 可以清楚地看到, 在所有数据集上, FAUDR 均在 30 次迭代内收敛. 验证了本文方法可以快速收敛.

sionality. Handbook of Statistics, 1982, 2(1982): 773-910

- [2] Espadoto M, Martins R M, Kerren A, et al. Towards a quantitative survey of dimension reduction techniques. IEEE Transactions on Visualization and Computer Graphics, 2019, 27(3): 2153-2173
- [3] Song L, Ma H B, Wu M, et al. A brief survey of dimension reduction. Proceedings of the International Conference on Intelligent Science and Big Data Engineering. Cham, Swiss: Springer, 2018: 189-200
- [4] Goyal P, Ferrara E. Graph embedding techniques, applications, and performance: A survey. Knowledge-Based Systems, 2018, 151: 78-94
- [5] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding. Science, 2000, 290(5500): 2323-2326
- [6] Belkin M, Niyogi P. Laplacian eigenmaps and spectral techniques for embedding and clustering. Proceedings of the 14th International Conference on Neural Information Processing Systems. Vancouver, Canada, 2001: 585-591
- [7] Li H F, Jiang T, Zhang K S. Efficient and robust feature extraction by maximum margin criterion. IEEE Transactions on Neural Networks, 2006, 17(1): 157-165
- [8] He X F, Niyogi P. Locality preserving projections. Advances in Neural Information Processing Systems, 2004, 16(16): 153-160
- [9] Jolliffe I T. Principal component analysis. Technometrics, 2003, 45(3): 276

- [10] Yan S C, Xu D, Zhang B Y, et al. Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 29(1): 40-51
- [11] Wang J, Wong R K W, Lee T C M. Locally linear embedding with additive noise. *Pattern Recognition Letters*, 2019, 123: 47-52
- [12] Wei W, Dai H, Liang W. Regularized least squares locality preserving projections with applications to image recognition. *Neural Networks*, 2020, 128: 322-330
- [13] Shen X J, Liu S X, Bao B K, et al. A generalized least-squares approach regularized with graph embedding for dimensionality reduction. *Pattern Recognition*, 2020, 98: 107023
- [14] Wang A, Zhao S, Liu J, et al. Locality adaptive preserving projections for linear dimensionality reduction. *Expert Systems with Applications*, 2020, 151: 113352
- [15] Gou J, Xue Y, Ma H, et al. Double graphs-based discriminant projections for dimensionality reduction. *Neural Computing and Applications*, 2020, 32(23): 17533-17550
- [16] Gao Y L, Luo S Z, Wang Z H, et al. Locality sensitive discriminative unsupervised dimensionality reduction. *Symmetry*, 2019, 11(8): 1036
- [17] Zhang R, Zhang Y, Li X. Unsupervised feature selection via adaptive graph learning and constraint. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 33(3): 1355-1362
- [18] Nie F P, Dong X, Li X L. Unsupervised and semisupervised projection with graph optimization. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 32(4): 1547-1559
- [19] Qiao L S, Zhang L M, Chen S C, Shen D G. Data-driven graph construction and graph learning: A review. *Neurocomputing*, 2018, 312: 336-351
- [20] Von Luxburg U. A tutorial on spectral clustering. *Statistics and Computing*, 2007, 17(4): 395-416
- [21] Boyd S, Boyd S P, Vandenberghe L. *Convex optimization*. *IEEE Transactions on Automatic Control*, 2006, 51(11): 1859-1859
- [22] Cai D, He X F, Han J W. Spectral regression: A unified subspace learning framework for content-based image retrieval // *Proceedings of the 15th ACM International Conference on Multimedia*. Augsburg, Germany, 2007: 403-412
- [23] Hartigan J A, Wong M A. A k -means clustering algorithm. *Applied Statistics*, 1979, 28(1): 100-108
- [24] Cai D, He X F, Han J W. Document clustering using locality preserving indexing. *IEEE Transactions on Knowledge and Data Engineering*, 2005, 17(12): 1624-1637
- [25] Han J W, Xiong K, Nie F P, Li X L. Structured graph reconstruction for scalable clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2021, 33(5): 2252-2265
- [26] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks. *Science*, 2006, 313(5786): 504-507



QIANG Qian-Yao, Ph. D. candidate.

Her research interests include machine learning, data mining and computer vision.

ZHANG Bin, Ph. D., associate professor. His research interests include image processing, computer vision, and VLSI architecture design for embedded vision system.

Background

High-dimensional data widely exists in real-life applications. As we all know, using high-dimensional data is time-consuming and storage-intensive. Dimensionality reduction is a vital technique to dispose high-dimensional data. This technique can be divided into supervised, semi-supervised and unsupervised methods based on the use of label information. Since labeled data is difficult to obtain in real scenarios, and manual labeling is expensive and inefficient, unsupervised method is the most used among these three types. As an effective fundamental data preprocessing mean, graph-based

dimensionality reduction plays an important role in the fields of machine learning, data mining and computer vision. Generally, conventional graph-based methods construct similarity graphs at first, and then explore low-dimensional subspaces non-linearly or linearly on the pre-constructed fixed graph. However, there exist some drawbacks. First, data in the real world always contains noise and outliers, learning on fixed similarity graph may lead to unreliable results that are rarely modified during the learning process. Second, non-linear methods cannot handle outliers and new-coming data. Third, although

the linear method solves the problem encountered by the nonlinear method, the strict linear projection is too strict for actual situations.

In this study, we propose a method named FAUDR, to overcome the aforementioned drawbacks in graph-based unsupervised dimensionality reduction. By introducing a relaxation regression term, FAUDR can flexibly handle a variety of data including noise and outliers. In contrast to conventional works learn on fixed graph, we learn the similarity graph by adaptively assigning the optimal neighbors for each sample, and combining with the dynamically changing low-dimensionality during the dimensionality reduction process. The learned similarity graph can better explore the data structure and serve for dimensionality reduction. It is worth

noting that we incorporate adaptive construction of graph and flexible relaxation regression in our algorithm. The proposed framework can learn an ideal low-dimensional representation for unlabeled high-dimensional data, and is significantly superior to several classical methods and current representative methods in dimensionality reduction ability.

We have published articles in major journals and conferences include IEEE TIP, IEEE TKDE, AAI and so on.

This work was supported by the Key Research and Development Program of Shaanxi (No. 2022GY-075), the National Major Science Technology Projects of China (No. 2019ZX01008103) and the Fundamental Research Funds for the Central Universities.

计算机学报