



基于双重局部保持的不完整多视角嵌入学习方法

刘彦雯¹, 张金鑫², 张宏杰¹, 经 玲¹

(1. 中国农业大学 理学院, 北京 100083; 2. 中国农业大学 信息与电气工程学院, 北京 100083)

摘 要: 现有的多视角降维方法多数假设数据是完整的, 但该假设在实际应用中难以实现。为解决不完整多视角数据降维问题, 提出一种新的不完整多视角嵌入学习方法。基于多视角数据的一致性与同一视角下样本间的线性相关性学习一组重构系数, 对缺失样本进行线性重构, 通过学习所有视角的公共低维嵌入, 保持原始空间的局部几何结构。在此基础上, 设计一种惩罚参数来度量重构样本的可靠度, 从而权衡缺失样本对学习结果的负面影响。实验结果表明, 该方法在 Yale、ORL 和 COIL-20 数据集上 NMI 值分别达到 65.63%、73.23% 和 78.27%, 较 MVL-IV 算法分别提升 8.37%、16.71% 和 20.24%。

关键词: 多视角学习; 不完整数据; 降维; 嵌入学习; 局部保持

开放科学(资源服务)标志码(OSID):



中文引用格式: 刘彦雯, 张金鑫, 张宏杰, 等. 基于双重局部保持的不完整多视角嵌入学习方法[J]. 计算机工程, 2021, 47(6): 115-122, 141.

英文引用格式: LIU Yanwen, ZHANG Jinxin, ZHANG Hongjie, et al. Incomplete multi-view embedded learning method based on double locality preserving[J]. Computer Engineering, 2021, 47(6): 115-122, 141.

Incomplete Multi-view Embedded Learning Method Based on Double Locality Preserving

LIU Yanwen¹, ZHANG Jinxin², ZHANG Hongjie¹, JING Ling¹

(1. College of Science, China Agriculture University, Beijing 100083, China;

2. College of Information and Electrical Engineering, China Agriculture University, Beijing 100083, China)

[Abstract] Most of the existing multi-view dimensionality reduction methods assume that the data is complete, but it is unrealistic for practical applications. In order to solve the problems in the dimensionality reduction of incomplete multi-view data, this paper proposes a new method for incomplete multi-view embedded learning. Based on the consistency of multi-view data and the linear correlation between samples under the same view, a set of reconstruction coefficients is learnt for the reconstruction of the missing samples. Then the common low-dimensional embedding of all views is learnt to maintain the local geometric structure of the original space. Additionally, a penalty parameter is designed to measure the reliability of the reconstructed samples, so as to balance the negative influence of missing samples on the learning results. Experimental results show that the NMI value of the proposed method reaches 65.63%, 73.23% and 78.27% on the datasets of Yale, ORL and COIL-20, increased by 8.37%, 16.71% and 20.24% compared with the MVL-IV algorithm.

[Key words] multi-view learning; incomplete data; dimensionality reduction; embedded learning; locality preserving

DOI: 10.19678/j.issn.1000-3428.0058129

0 概述

多视角数据是指对于同一个物体, 从不同层面观察或从多个源头得到的特征数据^[1-2]。多视角数据从不同的角度对同一物体进行描述, 比单视角数据更加

全面^[3-4]。多视角学习^[5-7]是根据各个视角间的内在联系及每个视角具有的独特属性对多视角数据进行处理和分析, 进而合理充分地利用多视角数据的一种全新的学习方法。多视角学习已经成为机器学习和数据挖掘领域中备受关注的热点问题^[8-10]。

基金项目: 国家自然科学基金(11671032)。

作者简介: 刘彦雯(1995—), 女, 硕士研究生, 主研方向为机器学习、模式识别; 张金鑫, 博士研究生; 张宏杰, 硕士研究生; 经玲(通信作者), 教授、博士。

收稿日期: 2020-04-21 **修回日期:** 2020-06-01 **E-mail:** jingling@cau.edu.cn

随着科学技术的高速发展,人们收集和处理数据的手段越来越多种多样,在信息爆炸式增长的同时也同样面临着数据维度高和高冗余的问题,增加了实验的计算复杂度并引起维数灾难^[11-12]。降维是克服维数灾难,获取数据本质特征的一个重要途径,根据是否基于样本标签信息,可以分为无监督降维和有监督降维。目前最主流的无监督降维方法有主成分分析(PCA)^[13]、Laplacian 特征映射(LE)^[14]和t-分布随机近邻嵌入(t-SNE)^[15]3种。对于有监督降维,最经典的方法是线性判别分析(LDA)^[16]。近年来,多视角降维方法备受关注^[17-19],然而由于数据收集的困难性、高额成本或设备故障等,高质量无缺失的多视角数据很少会出现在真实的应用中,人们收集的多视角数据常常包含不完整视角^[20-21]。所谓不完整视角是指在该视角下某一样本的部分或全部特征缺失。例如,在对阿兹海默症患者的诊断中,由于脑脊液的获取会对病人产生某种不利影响或攻击性,有些病人会拒绝这项检查,这就导致了该病人脑脊液这一诊断数据的缺失。当处理不完整多视角数据时,传统的多视角降维方法不能实现其良好的性能^[22-23]。因此,不完整多视角数据降维方法的研究已经成为多视角学习中一个重大的挑战。

目前,不完整多视角数据的降维方法主要分为两大类。第一类不考虑缺失数据的重构,旨在利用已有样本直接学习投影矩阵及样本的低维表示。2010年,KIMURA等人^[24]将典型相关分析(CCA)^[25]与主成分分析(PCA)^[13]的广义特征值问题结合,提出一种不完整两视角数据降维方法 Semi-CCA,希望在最大化两视角间成对样本低维表示相关性的同时,保持每个视角数据的全局结构,最终利用已有数据分别求出两个视角的投影矩阵,但仅限于两视角的问题,并且只适用于线性数据。对于非线性数据,TRIVEDI等人在 KCCA(Kernel CCA)^[26]的基础上提出了 MCIV^[27],基于不同视角间的核矩阵一致性的思想重构核矩阵,然后再应用 KCCA 对重构后的多视角数据降维。但是,这种方法要求至少有一个完整视角,这在实际应用中是难以实现的。

第二类降维方法通常基于某些前提假设补全缺失数据,希望能够减轻不完整数据对后续学习任务的影响。2015年,XU等人提出了 MVL-IV^[28],基于低秩假设进行矩阵补全。同时,考虑到多视角数据的一致性^[29-30],希望利用矩阵分解得到所有视角共同的低维表示,但是这种方法只是单纯地将多视角投影到一个公共子空间,没有考虑到数据的结构信息。TAO等人^[31]提出用低秩矩阵近似不完整视角,并通过线性变换学习完整的公共嵌入。此外,为学习到更加合理的低维嵌入,引入了块对角结构先验

正则项。2018年,YANG等人^[32]基于同一视角下样本的线性相关性与不同视角间同一样本对应的低维表示应该相似^[33-35]地假设线性重构缺失样本,同时引入图嵌入项来保持原始空间的局部结构,但是该方法只考虑了完整样本,忽略了重构样本的作用,造成了一定程度的信息损失。ZHANG等人^[36]通过引入不同视角间的一致分布约束进行特征级的缺失数据补全,同时通过构建特征同构子空间来捕捉不同视角间的互补性。

本文提出一种基于双重局部保持的不完整多视角降维方法(DLPEL)。利用不同视角间的局部结构一致性以及同一视角下样本的线性相关性来线性重构缺失样本,并构造所有样本点上的图来学习所有视角的公共低维嵌入。在此基础上,为权衡缺失样本对学习结果的影响,设计一个权重来度量样本的可靠性。

1 相关工作

1.1 局部线性嵌入

局部线性嵌入(Locality Linear Embedding, LLE)^[37]认为数据在局部是线性的,即某个样本可以被它邻域内的样本线性表示,希望降维后能保持高维空间的局部线性结构,是一种无监督非线性降维方法。已知数据集 $X = (x_1, x_2, \dots, x_m) \in \mathbb{R}^{d \times m}$, 其中, m 为样本个数, d 为维度, LLE 希望学习 m 个样本的低维表示 $y_1, y_2, \dots, y_m$ 。

用样本 x_i 的近邻点对 x_i 进行线性重构,计算线性相关系数向量 $w_i = (w_{i1}, w_{i2}, \dots, w_{im}) \in \mathbb{R}^{1 \times m}$ 。优化问题如下:

$$\min_{w_1, w_2, \dots, w_m} \sum_{i=1}^m \left\| x_i - \sum_{j=1}^m w_{ij} x_j \right\|_2^2 \quad (1)$$

$$\text{s.t. } w_{ij} = 0, j \notin Q_i$$

$$\sum_{j=1}^m w_{ij} = 1, i = 1, 2, \dots, m \quad (2)$$

其中, Q_i 为样本 x_i 的近邻点下标集合。LLE 希望低维样本间的局部线性相关性与原始数据一致,即线性相关系数 w_{ij} 保持不变。优化问题如下:

$$\min_{y_1, y_2, \dots, y_m} \sum_{i=1}^m \left\| y_i - \sum_{j=1}^m w_{ij} y_j \right\|_2^2 \quad (3)$$

由式(3)即可求出原始样本的低维表示 $y_1, y_2, \dots, y_m$ 。

1.2 局部保持投影

局部保持投影(Locality Preserving Projection, LPP)^[38]是一种经典的线性降维方法,其目标是学习一个投影矩阵 P ,将原始数据投影到低维子空间中。LPP 希望找到的投影矩阵 P 能够保持原始数据的局部近邻结构,优化问题如下:

$$\min_p \frac{1}{2} \sum_{i,j} W_{ij} \|P^T x_i - P^T x_j\|_2^2 \quad (4)$$

其中, W_{ij} 表示样本 x_i 和 x_j 的相似度, 可以根据不同的要求定义 W_{ij} , 如下:

$$W_{ij} = \begin{cases} e^{-\|x_i - x_j\|_2^2 / \sigma^2}, & \text{若 } x_i \in N_k(x_j), x_j \in N_k(x_i) \\ 0, & \text{否则} \end{cases} \quad (5)$$

或:

$$W_{ij} = \begin{cases} 1, & \text{若 } x_i \in N_k(x_j), x_j \in N_k(x_i) \\ 0, & \text{否则} \end{cases} \quad (6)$$

其中, σ 为热核参数, $N_k(x_i)$ 表示 x_i 的 k 个近邻点的集合, 但是 LPP 只能处理单视角数据, 如何将 LPP 应用于多视角场景仍是一个值得思考的问题。

2 本文模型

对于不完整多视角数据, 本文考虑部分样本在

所有视角下完整, 而其余样本只有部分视角, 即缺失样本的整个特征向量在某些视角下缺失的情况。本文模型希望学习所有视角的公共低维嵌入。

已知 N 个样本, 共有 s 个视角, 第 v 个视角的数据矩阵为 $X^v = (x_1^v, x_2^v, \dots, x_N^v) \in \mathbb{R}^{d^v \times N}$, $v = 1, 2, \dots, s$, 共有 n^v 个缺失样本, I^v 表示第 v 个视角的缺失样本下标集合, $V^i (i = 1, 2, \dots, N)$ 表示第 i 个样本的完整视角标号集合, 即若 $v \in V^i$, 则第 i 个样本在第 v 个视角下是完整的。本文称在所有视角下完整的样本为成对样本。为便于解释, 将前 $p (p < N)$ 个样本设为成对样本, 故数据矩阵可以表示为 $X^v = (X_a^v, x_{p+1}^v, \dots, x_N^v)$, 其中, $X_a^v = (x_1^v, x_2^v, \dots, x_p^v)$ 为第 v 个视角的成对样本矩阵。如图 1 所示, 用不同颜色来区分不同视角, 问号表示缺失样本。前 p 个样本为成对样本, 而 x_{p+1}^2, x_{N-1}^2 和 x_{p+1}^s, x_{N-1}^s 为缺失样本。

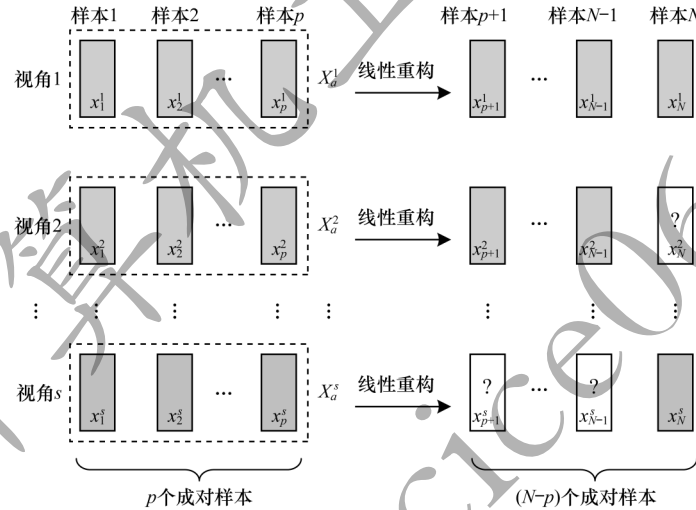


图1 不完整多视角数据的场景

Fig.1 Scenario of incomplete multi-view data

本文方法是一种两阶段方法。第一阶段线性重构缺失样本, 第二阶段求所有视角在 d 维公共子空间中的低维嵌入 $Y = [y_1, y_2, \dots, y_N] \in \mathbb{R}^{d \times N}$ 。

2.1 基于局部线性保持的缺失样本重构

本文认为同一视角下的不同样本不是独立无关的, 它们之间具有一定的线性相关性^[31], 某个样本可以被其余样本线性表示。基于这一假设, 对于缺失样本 $x_i^v (i \in I^v, v = 1, 2, \dots, s)$, 希望利用第 v 个视角下的所有成对样本对 x_i^v 进行线性重构。但由于样本 x_i^v 的缺失, 不能像 LEE 一样直接求出对应的线性相关系数向量。考虑到多视角数据的一致性, 认为不同视角下的样本应该具有相同的局部线性关系。因此, 首先计算其他完整视角中第 i 个样本 $x_i^u (u \in V^i)$ 关于成对样本矩阵 X_a^u 的重构系数向量 $l_i^u \in \mathbb{R}^{p \times 1}$ 。对于任一 $i \in I^v (v = 1, 2, \dots, s), u \in V^i$, 优化问题如下:

$$\min_{l_i^u} \|x_i^u - X_a^u l_i^u\|_2^2 + \lambda \|l_i^u\|_2^2 \quad (7)$$

其中, 第 1 项为重构误差项, $\|l_i^u\|_2^2$ 正则项是为了防止过拟合, λ 是调节重构建误差项与正则项平衡的参数。

通过求解上述优化问题, 获得了第 i 个样本在所有完整视角下关于成对样本的重构系数向量。考虑到视角间的相似程度不同, 引入一个新的权重 $\alpha_u (u \in V^i)$, 可以度量第 u 个视角与第 v 个视角的相似度。结合式 (7), 对任一 $i \in I^v (v = 1, 2, \dots, s)$, 有以下最小化问题:

$$\min_{\{a_u, l_i^u\}} \sum_{u \in V^i} \alpha_u \|x_i^u - X_a^u l_i^u\|_2^2 + \lambda \|l_i^u\|_2^2 \quad (8)$$

$$\text{s.t. } \sum_{u \in V^i} \alpha_u = 1, \alpha_u \geq 0 \quad (9)$$

每个视角都对应一个非负权重 α_u , α_u 越大, 表明第 u 个视角与第 v 个视角的相似度越高。计算出所有完整视角的权重 α_u 及重构系数向量 l_i^u , 就可以用

第 v 个视角下的所有成对样本线性重构缺失样本 x_i^v :

$$x_i^v = \sum_{u \in V^i} \alpha_u X_a^u l_i^u \quad (i \in I^v, v=1, 2, \dots, S) \quad (10)$$

2.2 基于局部结构保持的嵌入学习

在第2阶段,基于重构后的完整多视角数据,希望学习所有视角公共的低维 $Y=[y_1, y_2, \dots, y_N] \in \mathbb{R}^{d \times N}$ 。与LPP相似,希望 Y 能够保持原始空间的局部结构。但LPP不能直接应用于多视角数据,一种自然的方式是对所有 s 个视角求和,优化问题如下:

$$\min_Y \sum_{v=1}^s \sum_{i,j=1}^N S_{ij}^v \|y_i - y_j\|_2^2 \quad (11)$$

$$\text{s.t. } YY^T = I \quad (12)$$

其中, $S^v \in \mathbb{R}^{N \times N}$ 是所有样本的邻接矩阵, S_{ij}^v 定义如下:

$$S_{ij}^v = e^{-\frac{\|x_i^v - x_j^v\|_2^2}{\sigma}} \quad (13)$$

其中, $\|\cdot\|_2$ 是 l_2 -范数, σ 为高斯核参数。权重 S_{ij}^v 越大,对式(11)中目标函数的影响越大,这种方式使得公共低维嵌入 Y 能够保持原始空间的局部几何结构。

传统的基于图的降维方法如LPP等都是先学习原始空间的局部结构。显然,如果原始数据有误差或噪声,那么这种误差也会传递到所学的数据结构中。考虑到重构后的数据可能是不精确的,会给最终的学习结果带来一定的误差。因此,本文设计了一种权重来减轻重构样本对模型的影响,如(14)所示:

$$S_{ij}^v = e^{-\frac{\|x_i^v - x_j^v\|_2^2}{\sigma - z_{ij}}} \quad (14)$$

其中,惩罚参数 z_{ij} 定义如下:

$$z_{ij} = \begin{cases} 0, & \text{若 } M_{vi} = M_{vj} = 1 \\ 1, & \text{若 } M_{vi} = 0, M_{vj} = 0 \\ \frac{1}{\min(\|M(:, i)\|_1, \|M(:, j)\|_1)}, & \text{若 } M_{vi} = 0, M_{vj} = 1 \end{cases} \quad (15)$$

其中, $\|\cdot\|_1$ 是 l_1 -范数, $M(:, i)$ 表示矩阵 M 的第 i 列,称 $M \in \mathbb{R}^{s \times N}$ 为缺失样本指示矩阵:

$$M_{ij} = \begin{cases} 1, & \text{第 } j \text{ 个样本在第 } i \text{ 个视角下是完整的} \\ 0, & \text{否则} \end{cases} \quad (16)$$

其中,惩罚参数 z_{ij} 可以度量样本对 x_i^v 和 x_j^v 的可靠度,样本对的可靠度越低,权重 S_{ij}^v 越小,进而缺失样本对目标函数式(11)的影响越小。

此外,考虑到多视角数据的互补性,认为每个视角对学习公共嵌入 Y 的贡献都不同。因此,引入一个权重向量 $\beta = (\beta_1, \beta_2, \dots, \beta_s) \in \mathbb{R}^{1 \times s}$, 权重 β_v 越大,表明第 v 个视角的贡献越大。最重要的是,当样本对中包含缺失样本时,对应的权重 S_{ij}^v 就更小,这样可以减轻缺失样本对目标函数的影响。同时,保证了公共子空间能够保持原始空间的局部几何结构,最终的优化问题如下:

$$\min_{Y, \beta} \sum_{v=1}^s \beta_v^2 \sum_{i,j=1}^N S_{ij}^v \|y_i - y_j\|_2^2 \quad (17)$$

$$\text{s.t. } YY^T = I, \sum_{v=1}^s \beta_v = 1, \beta_v \geq 0 \quad (18)$$

为叙述方便,上述优化问题可以化解为如下形式:

$$\min_{Y, \beta} \sum_{v=1}^s \beta_v^2 \text{tr}(YL^v Y^T) \quad (19)$$

$$\text{s.t. } YY^T = I, \sum_{v=1}^s \beta_v = 1, \beta_v \geq 0 \quad (20)$$

其中, $\text{tr}(\cdot)$ 表示矩阵的迹, $L^v = D^v - S^v$ 为图拉普拉斯阵, D^v 是一个对角矩阵, $D_{ii}^v = \sum_j S_{ij}^v$ 。

3 优化方法

3.1 重构阶段

对 $i \in I^v, v=1, 2, \dots, s$, 由于优化式(8)、式(9)是有非线性约束的非凸问题,不能直接计算出全局最优解,本文设计了一种交替迭代优化算法来获得问题的局部最优解。

1) 固定 α_u , 更新 $l_i^u (u \in V^i)$ 。

对第 u 个视角,求解 l_i^u 的子问题如下:

$$\min_{l_i^u} \|x_i^u - X_a^u l_i^u\|_2^2 + \frac{\lambda}{\alpha_u^2} \|l_i^u\|_2^2 \quad (21)$$

计算式(21)中目标函数关于 l_i^u 的一阶偏导,并令其等于0,可得:

$$X_a^{uT} X_a^u l_i^u - X_a^{uT} x_i^u + \frac{\lambda}{\alpha_u^2} l_i^u = 0 \quad (22)$$

可得解析解:

$$l_i^u = \left(X_a^{uT} X_a^u + \frac{\lambda}{\alpha_u^2} I \right)^+ X_a^{uT} x_i^u \quad (23)$$

其中, $I \in \mathbb{R}^{p \times p}$ 是一个单位矩阵, A^+ 表示 A 的广义逆矩阵。

2) 固定 l_i^u , 更新 $\alpha_u (u \in V^i)$ 。

给定 l_i^u , 求解权重 α_u 的子问题如下:

$$\min_{\{\alpha_u\}_{u \in V^i}} \sum_{u \in V^i} \alpha_u^2 \|x_i^u - X_a^u l_i^u\|_2^2 \quad (24)$$

$$\text{s.t. } \sum_{u \in V^i} \alpha_u = 1, \alpha_u \geq 0 \quad (25)$$

为简单起见,式(24)中的目标函数可写为:

$$\begin{aligned} & \sum_{u \in V^i} \alpha_u^2 \|x_i^u - X_a^u l_i^u\|_2^2 = \\ & \sum_{u \in V^i} \alpha_u^2 \left\| X_a^u e_i - X_a^u \begin{pmatrix} I_{p \times p} \\ O_{(n-p) \times p} \end{pmatrix} l_i^u \right\|_2^2 = \\ & \sum_{u \in V^i} \alpha_u^2 \left\| X_a^u \left(e_i - \begin{pmatrix} I_{p \times p} \\ O_{(n-p) \times p} \end{pmatrix} l_i^u \right) \right\|_2^2 = \\ & \sum_{u \in V^i} \alpha_u^2 \text{tr}(X_a^u H^u X_a^{uT}) \end{aligned} \quad (26)$$

其中, $e_i \in \mathbb{R}^{N \times 1}$ 为第 i 个元素为1,其余均为0的列向

量, $I_{p \times p}$ 是一个单位矩阵, $O_{(n-p) \times p}$ 为全0矩阵。因此,关于 α_u 的子问题形式如下:

$$H^u = \left(e_i - \begin{pmatrix} I_{p \times p} \\ O_{(n-p) \times p} \end{pmatrix} I_i^u \right) \left(e_i - \begin{pmatrix} I_{p \times p} \\ O_{(n-p) \times p} \end{pmatrix} I_i^u \right)^T$$

$$\min_{\{\alpha_u\}_{u \in V^i}} \sum_{u \in V^i} \alpha_u^2 \text{tr}(X^u H^u X^{uT}) \quad (27)$$

$$\text{s.t. } \sum_{u \in V^i} \alpha_u = 1, \alpha_u \geq 0 \quad (28)$$

引入拉格朗日乘子 γ , 可得拉格朗日函数:

$$L(\alpha_u, \gamma) = \sum_{u \in V^i} \alpha_u^2 \text{tr}(X^u H^u X^{uT}) - \gamma \left(\sum_{u \in V^i} \alpha_u - 1 \right) \quad (29)$$

令 $L(\alpha_u, \gamma)$ 关于 α_u 和 γ 的一阶偏导等于0, 可得:

$$\alpha_u = \frac{\frac{1}{\text{tr}(X^u H^u X^{uT})}}{\sum_{u \in V^i} \frac{1}{\text{tr}(X^u H^u X^{uT})}} \quad (30)$$

重构阶段的迭代过程如算法1所示。

算法1 重构阶段

输入 不完整多视角数据集 $X^v (v = 1, 2, \dots, s)$, 参数 λ

输出 重构样本 $x_i^v (i \in I^v, v = 1, 2, \dots, s)$

1. for $v = 1, 2, \dots, s$
2. for $u = 1, 2, \dots, s$
3. while $u \in V^i (i \in I^v)$ do
4. 初始化: $\alpha_u = \frac{1}{s-1}$;
5. while 未达到停止准则 do
6. 根据式(23)更新 l_i^u ;
7. 根据式(24)更新 α_u ;
8. end for
9. end for
10. $x_i^v = \sum_{u \in V^i} \alpha_u X_{i+1}^{uT} (i \in I^v, v = 1, 2, \dots, s)$

3.2 降维阶段

与重构阶段类似, 优化问题式(19)、式(20)是有非线性约束的非凸问题, 无法找到其全局最优解。因此, 本文设计了一种交替迭代优化算法来求问题的局部最优解。

1) 固定 β , 更新 Y 。

给定权重向量 $\beta = (\beta_1, \beta_2, \dots, \beta_s)$, 关于 Y 的子优化问题如下:

$$\min_Y \text{tr}(YLY^T) \quad (31)$$

$$\text{s.t. } YY^T = I \quad (32)$$

其中, $L = \sum_{v=1}^s \beta_v^2 L^v$, 故全局最优解 Y 是由 L 的前 d 个最小特征值对应的特征向量组成的矩阵。

2) 固定 Y , 更新 β 。

已知低维嵌入 Y , 关于 β 的子问题形式如下:

$$\min_{\beta} \sum_{v=1}^s \beta_v^2 \text{tr}(YL^v Y^T) \quad (33)$$

$$\text{s.t. } \sum_{v=1}^s \beta_v = 1, \beta_v \geq 0 \quad (34)$$

求解权重向量 β 的方法此处不再赘述, 详情可见3.1节。下面给出 β_v 的形式:

$$\beta_v = \frac{\frac{1}{\text{tr}(YL^v Y^T)}}{\sum_{v=1}^s \frac{1}{\text{tr}(YL^v Y^T)}} \quad (35)$$

降维阶段的迭代过程如算法2所示。

算法2 降维阶段

输入 重构后的多视角数据集 $X^v (v = 1, 2, \dots, s)$, 热核

参数 σ , 子空间维度 d

输出 公共嵌入 Y

1. 计算拉普拉斯阵 $L^v (v = 1, 2, \dots, s)$;
2. 初始化: $\beta_v = \frac{1}{\sqrt{s}} (v = 1, 2, \dots, s)$;
3. while 未达到停止准则 do
4. 求解优化问题式(31)、式(32)更新 Y ;
5. for $v = 1, 2, \dots, s$
6. 根据式(35)更新 β_v ;
7. end for

4 实验

本节将在3个经典的图像数据集上, 通过数值实验与一些相关的不完整多视角降维方法进行比较, 并根据 K-means 聚类结果验证本文模型的有效性。

4.1 数据集

实验中用到的3个图像数据集如下:

- 1) Yale人脸数据集由耶鲁大学计算机视觉与控制中心创建, 包含15个人的165张人脸图像。每个人有11张图像, 分别对应快乐、正常、悲伤、困倦、惊喜、眨眼、左侧光、右侧光、戴眼镜和不戴眼镜等不同的面部表情、光照条件或姿态, 图2为其中一人的11张图像。



图2 Yale人脸数据集中某人的11张图像

Fig.2 11 images of one person in the Yale face dataset

- 2) ORL人脸数据集由剑桥大学 AT&T 实验室采集, 包含40个人的400张人脸图像。实验室在不同的时间以光照条件、面部表情和面部饰物为变量给每个人拍摄了10张照片。

- 3) COIL-20数据集来自哥伦比亚大学图像数据库, 包含20个物体, 如图3所示。每个物体旋转 360° , 每隔 5° 拍摄一张照片, 因此数据集共有1440张照片。



图3 COIL-20数据集中的20个物体

Fig.3 20 objects in the COIL-20 dataset

4.2 评价标准

本文运用3种广泛应用的指标度量聚类结果的好坏,分别是准确率(ACC)、标准化互信息(NMI)和纯度(Purity)。

$$A_{\text{ACC}} = \frac{\sum_{i=1}^n \delta(y_i, \text{map}(c_i))}{n} \quad (36)$$

其中, n 为样本个数, y_i 和 c_i 分别代表第 i 个样本的真实标签和聚类标签, $\text{map}(\cdot)$ 是一个排列函数, 可以利用匈牙利算法对齐真实标签和聚类标签^[39], δ 是一个指示函数, 若 $y_i = \text{map}(c_i)$, 则为1, 否则为0。

$$N_{\text{NMI}} = \frac{I(T_{\text{TL}}; C_{\text{CL}})}{\sqrt{E(T_{\text{TL}})E(C_{\text{CL}})}} \quad (37)$$

其中, T_{TL} 和 C_{CL} 分别表示真实标签和聚类标签, $I(T_{\text{TL}}; C_{\text{CL}})$ 表示 T_{TL} 和 C_{CL} 间的互信息, 度量聚类结果与真实标签间的相似程度, $E(\cdot)$ 表示变量的熵, 引入分母是为了将互信息的值标准化到 $[0, 1]$ 之内。

$$P_{\text{purity}} = \sum_{i=1}^k \frac{n_i}{n} \max_j (P_{ij}) \quad (38)$$

其中, k 为簇的个数, n_i 为第 i 簇的样本数, $p_{ij} = \frac{n_{ij}}{n_i}$ 为第 i 簇中的样本属于第 j 类的概率, n_{ij} 为第 i 簇中的样本属于第 j 类的个数。

4.3 对比方法与实验设置

本文选择了近几年具有代表性的5种不完整多视角降维方法, 包括 Mean value、Semi-CCA^[24]、MCIV^[27]、MVL-IV^[28]、IML-BDR^[31]。其中 Mean value 利用同一视角下的完整样本平均值补全缺失样本, 使用本文的多视角降维方法对学习到的嵌入进行 K-means 聚类, 其他几种对比方法的更多细节见概述部分。

从数据集的图像中提取出灰度值强度(GSI)、方向梯度直方图(HOG)和局部二元模式(LBP)3个视角的特征, 得到3个完整的多视角数据集。为去除

原始特征的冗余信息, 利用主成分分析(PCA)对数据预处理。通过构造不完整数据集, 随机选取70%的样本作为成对样本, 然后在每个视角剩余的样本中分别选取10%的样本用1填充。由于 MCIV 要求至少有一个完整视角, 因此在第2个、第3个视角中选取15%的样本作为缺失样本, 并用1填充。

4.4 实验结果

在 Yale、ORL 和 COIL-20 数据集上的实验结果如表1~表3所示, 分别对应 ACC、NMI 和 Purity 值, 值越高, 证明模型的性能越好, 其中黑色粗体表示最优结果。

表1 Yale、ORL 和 COIL-20 数据集上的 ACC 结果

方法	Yale	ORL	COIL-20
Mean value	36.24	31.40	62.53
Semi-CCA	38.36	43.15	50.42
MCIV	38.42	41.00	50.00
MVL-IV	41.10	40.60	47.43
IML-BDR	44.61	39.30	38.26
DLPEL	57.67	55.30	64.74

表2 Yale、ORL 和 COIL-20 数据集上的 NMI 结果

方法	Yale	ORL	COIL-20
Mean value	41.28	54.30	74.46
Semi-CCA	45.02	63.53	61.39
MCIV	45.78	62.68	62.09
MVL-IV	57.36	57.94	58.03
IML-BDR	54.19	60.19	53.24
DLPEL	65.63	73.23	78.27

表3 Yale、ORL 和 COIL-20 数据集上的 Purity 结果

方法	Yale	ORL	COIL-20
Mean value	37.82	35.95	65.45
Semi-CCA	40.67	46.55	53.51
MCIV	40.20	44.32	56.21
MVL-IV	44.65	44.65	51.11
IML-BDR	48.36	43.50	42.56
DLPEL	60.84	59.10	68.04

从表1~表3可以看出:

1) 与 DLPEL 方法相比, Mean value 方法的聚类结果更差。这表明直接用均值向量填充缺失样本是不合理的, 并且这种补全方式影响了后续的聚类。本文模型利用了多视角数据的一致性, 更好地实现

了缺失样本重构。

2) MCIV方法要求至少有一个完整视角,该方法可以更好地利用多视角数据的互补性,但仍没有DLPEL方法效果好,表明了本文方法处理一般不完整多视角数据的有效性,可以更广泛地应用到实际问题中。

从以上结果可以看出,本文方法DLPEL在3个数据集上相比其他方法都具有更好的性能。

4.5 参数分析

参数 λ 负责调节重构误差项与正则项之间的平衡,本文应用本文模型DLPEL将样本降到40维后进行K-means聚类,通过在Yale、ORL和COIL-20数据集上的实验结果分析了模型关于参数 λ 的敏感度。

由图4~图6可知,DLPEL关于参数 λ 较为敏感,当 λ 锁定在 $[1, 10^4]$ 的范围内时,可以达到较为理想的聚类效果。在实验过程中发现,高斯核参数 σ 的变化对模型的影响很大,当先固定 $\lambda = 10^3, d = 40$,然后在Yale、ORL、COIL-20数据集上进行实验,寻找最优的参数 σ 。从图7可以直观地看出,随着 σ 的变化,NMI的值波动很大。当 $\sigma = 10^6$ 时,本文模型在3个数据集上都展现了最优越的性能。因此,将所有实验中的参数 σ 均设为 10^6 。

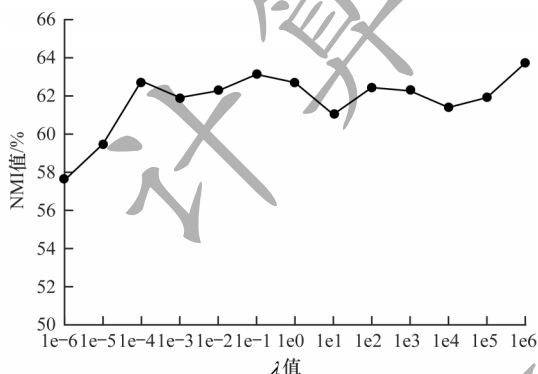


图4 Yale数据集上NMI随 λ 变化的折线

Fig.4 Line chart of NMI changes with λ on Yale dataset

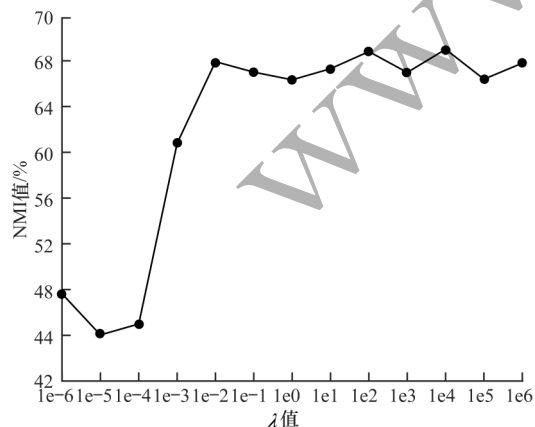


图5 ORL数据集上NMI随 λ 变化的折线

Fig.5 Line chart of NMI changes with λ on ORL dataset

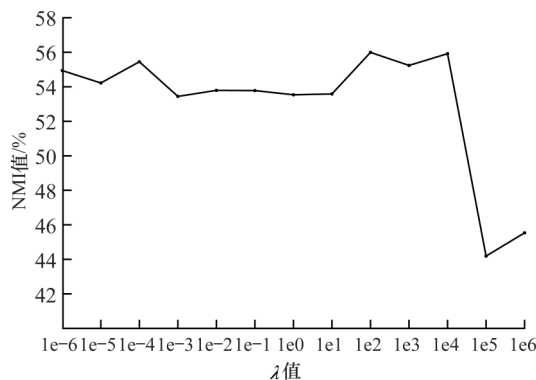


图6 COIL-20数据集上NMI随 λ 变化的折线图

Fig.6 Line chart of NMI changes with λ on COIL-20 dataset

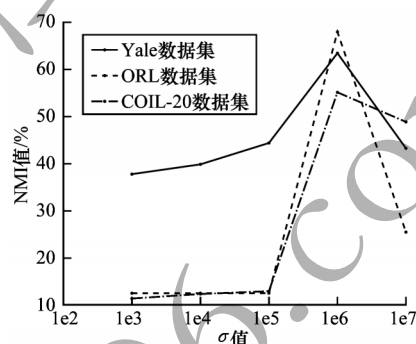


图7 Yale、ORL和COIL-20数据集上NMI随 σ 变化的折线

Fig.7 Line chart of NMI changes with σ on Yale, ORL and COIL-20 datasets

5 结束语

本文结合多视角数据局部结构的一致性与特征空间的线性结构,提出一种基于双重局部保持的不完整多视角降维方法。通过引入一个惩罚参数,减轻了缺失样本对学习结果的影响,得到了保持原始数据局部结构的公共低维嵌入。实验结果表明,与MVL-IV算法相比,该方法可以获得较好的聚类结果,验证了本文模型处理不完整多视角数据时的有效性。本文模型分两阶段考虑了缺失数据的补全和降维,但有可能导致算法的性能退化,下一步将提出一个统一的目标函数,在补全数据的同时获取样本的低维表示,通过两者的相互促进提高算法的性能。

参考文献

- [1] TANG Jingjing, TIAN Yingjie. Overview of multi-view learning[J]. Mathematical Modeling and Its application, 2017, 6(3): 1-15. (in Chinese)
唐静静, 田英杰. 多视角学习综述[J]. 数学建模及其应用, 2017, 6(3): 1-15.
- [2] ZONG Linlin, ZHANG Xianchao, ZHAO Long, et al. Multi-view clustering via multi-manifold regularized non-negative matrix factorization[J]. Neural Networks, 2017, 88: 74-89.

- [3] LIU Jialu, WANG Chi, GAO Jing, et al. Multi-view clustering via joint nonnegative matrix factorization[C]// Proceedings of 2013 SIAM International Conference on Data Mining. Austin, USA; [s. n.], 2013; 252-260.
- [4] SUN Shiliang. A survey of multi-view machine learning[J]. Neural Computing and Applications, 2013, 23(7): 2031-2038.
- [5] COLLINS M, SINGER Y. Unsupervised models for named entity classification [C]//Proceedings of 1999 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora. Florham, USA; [s. n.], 1999; 100-110.
- [6] GHANI R. Combining labeled and unlabeled data for text classification with a large number of categories [C]// Proceedings of International Conference on Data Mining. San Jose, USA; IEEE Press, 2001; 597-598.
- [7] MUSLEA I, MINTON S, KNOBLOCK C, et al. Active learning with multiple views [J]. Journal of Artificial Intelligence Research, 2006, 27(1): 203-233.
- [8] BLUM A, MITCHELL T M. Combining labeled and unlabeled data with co-training [C]//Proceedings of the 11th Annual Conference on Learning Theory. Washington D. C. , USA; IEEE Press, 1998; 92-100.
- [9] HECKMANN M, BERTHOMMIER F, KROSCHER K. Noise adaptive stream weighting in audio-visual speech recognition[J]. EURASIP Journal on Advances in Signal Processing, 2002(11): 1-14.
- [10] WU Y, CHANG E Y, CHANG K C C, et al. Optimal multimodal fusion for multimedia data analysis [C]// Proceedings of the 12th Annual ACM International Conference on Multimedia. New York, USA; ACM Press; 2004; 572-579.
- [11] WANG Yang, ZHANG Wenjie, WU Lin, et al. Unsupervised metric fusion over multiview data by graph random walk-based cross-view diffusion [J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 28(1): 57-70.
- [12] INDYK P, MOTWANI R. Approximate nearest neighbors: towards removing the curse of dimensionality [C]// Proceedings of the 30th Annual ACM Symposium on Theory of Computing. New York, USA; ACM Press, 1998; 604-613.
- [13] JOLLIFFE I T. Principal Component Analysis [M]. Berlin, Germany; Springer, 2011.
- [14] BELKIN M, NIYOGI P. Laplacian eigenmaps for dimensionality reduction and data representation [J]. Neural computation, 2003, 15(6): 1373-1396.
- [15] DER M L V, HINTON G E. Visualizing data using t-SNE[J]. Journal of Machine Learning Research, 2008, 31(5): 2579-2605.
- [16] MOORE B A, MCLACHLAN G J. Discriminant analysis and statistical pattern recognition[J]. Biometrical Journal, 1993, 35(7): 784-784.
- [17] XIA Tian, TAO Dacheng, MEI Tao, et al. Multiview spectral embedding[J]. IEEE Transactions on Systems, Man and Cybernetics, 2010, 40(6): 1438-1446.
- [18] KAMALIKA C, SHAM M K. Multi-view clustering via canonical correlation analysis [C]//Proceedings of International Conference on Machine Learnig. Montreal, Canada; [s. n.], 2009; 129-136.
- [19] TAEKYUN K, JOSEF K, ROBERTO C. Discriminative learning and recognition of image set classes using canonical correlations [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 1005-1018.
- [20] HU Menglei, CHEN Songcan. Doubly aligned incomplete multi-view clustering [C]//Proceedings of IEEE International Joint Conference on Artificial Intelligence. Washington D. C. , USA; IEEE Press, 2018; 2262-2268.
- [21] LI Shaoyuan, JIANG Yuan, ZHOU Zhihua. Partial multi-view clustering [C]//Proceedings of AAAI Conference on Artificial Intelligence. [S. l.]; AAAI Press, 2014; 1968-1974.
- [22] CHRISTOPH H L, OLIVER K. Weakly-paired maximum covariance analysis for multimodal dimensionality reduction and transfer learning [C]//Proceedings of European Conference on Computer Vision. Berlin, Germany; Springer, 2010; 566-579.
- [23] CHEN Xiaohong, CHEN Songcan, XUE Hui, et al. A unified dimensionality reduction framework for semi-paired and semi-supervised multi-view data [J]. Pattern Recognition, 2012, 45(5): 2005-2018.
- [24] KIMURA A, KAMEOKA H, SUGIYAMA M, et al. SemiCCA: efficient semi-supervised learning of canonical correlations [C]//Proceedings of the 20th International Conference on Pattern Recognition. Washington D. C. , USA; IEEE Press, 2010; 2933-2936.
- [25] HOTELING H. Relations between two sets of variates [J]. Biometrika, 1936, 28(4): 321-377.
- [26] HARDOON D R, SZEDMAK S, SHAWETYLOR J, et al. Canonical correlation analysis: an over view with application to learning methods [J]. Neural Computation, 2004, 16(12): 2639-2664.
- [27] TRIVEDI A, RAI P, DUVAL S. Multi-view clustering with incomplete views [C]//Proceedings of Workshop on Machine Learning for Social Computing. Whistler, Canada; [s. n.], 2010; 656-579.
- [28] XU Chang, TAO Dacheng, XU Chao. Multi-view learning with incomplete views [J]. IEEE Transactions on Image Processing, 2015, 24(12): 5812-5825.
- [29] ARGYRIOU A, EVGENIOU T, PONYIL M. Multi-task feature learning [C]//Proceedings of Advances in Neural Information Processing Systems. Washington D. C. , USA; IEEE Press, 2007; 19-41.
- [30] WANG Xin, BI Jinbo, YU Shipeng, et al. On multiplicative multitask feature learning [C]//Proceedings of Advances in Neural Information Processing Systems. Washington D. C. , USA; IEEE Press, 2014; 2411-2419.
- [31] TAO Hong, HOU Chenping, YI Dongyun, et al. Joint embedding learning and low-rank approximation: a framework for incomplete multi-view learning [J]. IEEE Transactions on Cybernetics, 2019, 65(5): 1-14.
- [32] YANG Wanqi, SHI Yinghuan, GAO Yang, et al. Incomplete-data oriented multiview dimension reduction via sparse low-rank representation [J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(12): 6276-6291.

(上接第 122 页)

- [33] CHEN Ning, ZHU Jun, SUN Fuchun, et al. Large-margin predictive latent subspace learning for multi-view data analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(12): 2365-2378.
- [34] MARTHA W, ZHANG X, DALE S, et al. Convex multi-view subspace learning [C]//Proceedings of NIPS' 12. Cambridge, USA: MIT Press, 2012: 1673-1681.
- [35] SHARMA A, KUMAR A, DAUME H, et al. Generalized multi-view analysis: a discriminative latent space [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE Computer Society, 2012: 2160-2167.
- [36] ZHANG Lei, ZHAO Yao, ZHU Zhenfeng, et al. Multi-view missing data completion [J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(7): 1296-1309.
- [37] SAM T R, LAWRENCE K S. Nonlinear dimension reduction by locally linear embedding [J]. Science, 2000, 290: 2323-2326.
- [38] HE X F, PARTHA N. Locality preserving projections [C]// Proceedings of NIPS' 02. Cambridge, USA: MIT Press, 2002: 153-160.
- [39] YIN Qiyue, WU Shu, WANG Liang. Multiview clustering via unified and view-specific embeddings learning [J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(11): 5541-5553.

编辑 索书志