

基于低秩正则化异构张量分解的子空间聚类算法

张静, 付建鹏, 李新慧*

天津大学电气自动化与信息工程学院, 天津 300072

摘要 张量分解是解决高维数据分析问题的有力工具。传统张量 Tucker 分解模型多采用各项同性假设, 即各个因子矩阵具有相同的约束条件(例如正交、非负等), 但此种假设不适用于异构张量数据分析。本文提出了一种基于低秩正则化的异构张量分解(LRRHTD)算法, 并用于子空间聚类任务。低秩正则化的异构张量分解核心思想是对原始张量寻求一组正交因子矩阵的集合, 将高维张量映射到低维的潜在子空间中, 同时在最后的因子矩阵上引入低秩约束以获得可用于聚类的全局低秩结构表征。此外, 设计了一种基于增广拉格朗日乘子的优化方法对所提算法进行求解。在两个公开数据集上的实验表明, 本文提出的方法不仅可以在较少次数的迭代内达到收敛, 而且与现有的其他聚类方法相比, 取得了较为理想的聚类性能。

关键词 图像处理; 张量; 分解; 低秩; 子空间聚类

中图分类号 TP391

文献标识码 A

doi: 10.3788/LOP55.071003

Low-Rank Regularized Heterogeneous Tensor Decomposition Algorithm for Subspace Clustering

Zhang Jing, Fu Jianpeng, Li Xinhui

School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China

Abstract Tensor decomposition is a powerful computational tool for analyzing multi-dimensional data. The traditional Tucker decomposition models are generally proposed based on the isotropy hypothesis, meaning that the factor matrices are learned in an equivalent way for all modes (such as orthogonal or non-negative constraints), which is not suitable for the heterogeneous tensor data. We propose a low-rank regularized heterogeneous tensor decomposition (LRRHTD) model for subspace clustering. The core idea of LRRHTD is that we seek a set of orthogonal factor matrices for all but the last mode to map the high-dimensional tensor into a low-dimensional latent subspace. In the meantime, we seek the lowest-rank representation of the original tensor by imposing a low-rank constraint on the last mode, in order to reveal the global structure of samples for the purpose of clustering. We also develop an effective optimization algorithm based on augmented Lagrangian multiplier to solve our proposed model. Experiments on two public datasets demonstrate that the proposed method reaches convergence within a small number of iterations and achieves promising clustering results in comparison with state-of-the-art methods.

Key words image processing; tensor; decomposition; low-rank; subspace clustering

OCIS codes 100.2980; 280.3420

1 引言

随着传感器技术的广泛应用和传媒技术与存储技术的发展,数据的类型和数量急剧增长。在这种情况下,从社交网络、通信网络以及计算机视觉等不

同领域收集的信息可以存储为多维或多路特征阵列,这种多维数据被称为张量。

张量分解^[1-2]作为一种分析多维数据的重要工具,可以捕捉大规模高维数据的多重线性关系和多方面结构信息,被广泛应用于维数约简、特征提取,

收稿日期: 2017-11-29; **收到修改稿日期:** 2018-01-03

基金项目: 国家自然科学基金(61572356)

作者简介: 张静(1972—),女,博士,副教授,硕士生导师,主要从事信息安全、张量分解方面的研究。

E-mail: zhangjing@tju.edu.cn

* **通信联系人.** E-mail: 1119345739@qq.com

分类和聚类的任务中。例如,气动光学效应分析^[3]、图像光谱压缩^[4-5]、图像加密^[6]、数字水印^[7]等。在各种张量分解技术中,Tucker^[8]和CANDECOMP/PARAFAC(CP)分解^[9]是两个基本模型。Tucker分解将张量分解为一组因子矩阵以及核张量,而CP分解将张量分解成一阶张量之和。值得注意的是CP分解可以看作是Tucker分解的特例,即其核张量是超对角结构。近年来,越来越多的研究致力于将Tucker分解模型与子空间学习机制结合起来。根据约束条件的不同,Tucker分解大致可以分为三类:正交张量分解、非负张量分解和稀疏张量分解。在正交张量分解中,为了避免中间表示的任意缩放并保持尽可能多的信息,因子矩阵被约束为正交结构,例如Kolda^[10]利用两个不同易得性的定义用于张量的正交分解。通过对因子矩阵增加正交约束,De Lathauwer等^[11-12]提出了高阶奇异值分解(HOSVD)和高阶正交迭代(HOOD)算法。为确保约简后的张量包含输入数据的最大信息,Phan等^[13]提出了一种高维判别分析(HODA)算法用于高维数据的特征提取。

在非负张量分解方法^[14-16]中,利用非负约束将张量数据(如非负时间序列或视频)分解成有意义的潜在非负分量,相应地提高了分解结果的可解释性。几种有效的非负张量分解算法被相继提出,包括非负张量分解(NTF)^[15,17],非负 n 维张量分解(n -NTF)^[18],三因子正交非负张量分解(Tri-ONTD)^[14],具有最大边界约束的监督非负张量分解(SNTFM 2)^[19]和低秩非负张量分解(lraNTF)^[20]。最近,Sun等^[21]提出了一种基于多项式流形(HTD多项式)的异构Tucker分解模型用于张量分解,其结合了正交约束和非负约束的优点。Zhou等^[22]提出了一种低秩近似非负Tucker分解(LRANTD)算法来降低存储和计算的成本。

因为稀疏约束对于遮挡或噪声具有显著的稳健性^[23],所以张量分解中另一个常用的约束是稀疏性假设。在稀疏张量分解中,基本因子矩阵可以被约束为局部稀疏表示^[24]。例如,Liu等^[25]提出了一种基于逐列坐标下降的稀疏非负张量因子分解(SNTF)的算法。Perros等^[26]提出了一种新的稀疏分层Tucker分解(稀疏H-Tucker)用于稀疏的高阶张量数据。

上述几乎所有的张量分解算法都是基于各向同性假设,也就是说在所有的模式上以相同的方式学习因子矩阵。然而,在实践中这个假设对于许多高

维数据是不合理的,因为张量数据的各个维度代表着不同的信息。低秩特性可以很好地表征存在噪声和损坏的数据。受此启发,低秩正则化已被广泛应用于张量子空间学习方法^[27-29]和张量补全^[30-32]。但是,将低秩正则化用于张量分解的工作非常少。

本文提出低秩正则化异构张量分解(LRRHTD)算法用于子空间聚类。具体而言,本文提出的张量分解模型中,除最后一个模式以外的其他所有模式因子矩阵采用正交约束,即寻求一组正交投影矩阵以将原始张量数据映射到低维公共子空间,最后一个模式因子矩阵采用低秩约束,即通过添加核范数来学习低秩投影矩阵,从而获得样本的全局结构的低秩表征用于聚类。同时,还设计了基于增广拉格朗日乘子(ALM)的有效算法来对其进行优化。

2 基于低秩正则化异构张量分解的子空间聚类算法

2.1 问题描述

聚类是模式识别领域一个重要的研究方向。传统的聚类方法大多是基于向量的方法,而对于高维数据,向量化的过程会损失大量的结构信息,因此基于张量的方法在处理高维数据方面更有优势。一个 M 阶张量被定义为 $\mathbf{X} \in \mathbf{R}^{I_1 \times I_2 \times \dots \times I_M}$,其中 $I_m (1 \leq m \leq M)$ 是模式 m 的维度,张量中元素表示为 $X_{i_1 i_2 \dots i_m}$, i_m 是模式 m 的索引,其中 $1 \leq i_m \leq I_m$ 。

假设给定一个 M 阶张量的集合 $\{\mathbf{X}_i \in \mathbf{R}^{I_1 \times I_2 \times \dots \times I_M}, i = 1, 2, \dots, N\}$, N 是样本总数, $I_1 \times I_2 \times \dots \times I_M$ 是每个样本的维数。沿着一个方向堆叠 N 个 M 阶张量得到 $M+1$ 阶张量 $\mathbf{X} \in \mathbf{R}^{I_1 \times I_2 \times \dots \times I_M \times N}$ 。LRRHTD算法旨在寻求 M 个正交因子矩阵 $\{\mathbf{A}_m \in \mathbf{R}^{I_m \times J_m}\}_{m=1}^M$ [其中 $J_m \leq I_m (m = 1, 2, \dots, M)$],表征样本内部低维结构,以及将映射的低维张量作为字典,获得样本低秩表示 $\mathbf{Z} \in \mathbf{R}^{N \times D}$ 并且 $D \leq N$ 。

堆叠成的张量 \mathbf{X} 的Tucker分解可以表示为

$$\mathbf{X} = \mathbf{G} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \times_3 \dots \times_M \mathbf{A}_M \times_{M+1} \mathbf{Z} + \mathbf{E}, \quad (1)$$

式中 $\mathbf{G} \in \mathbf{R}^{J_1 \times J_2 \times \dots \times J_N \times D}$ 为核张量, \mathbf{E} 为近似误差张量。

潜在因子矩阵 $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_M, \mathbf{Z}$ 可以通过最小化如下代价函数 $f(\mathbf{G}, \mathbf{A}, \mathbf{Z})$ 来获得:

$$f(\mathbf{G}, \mathbf{A}, \mathbf{Z}) = \frac{1}{2} \|\mathbf{X} - \mathbf{G} \times_1 \mathbf{A}_1 \dots \times_M \mathbf{A}_M \times_{M+1} \mathbf{Z}\|_F^2. \quad (2)$$

由于因子矩阵的正交性,可以直接得到核张量:

$$\mathbf{G} = \mathbf{X} \times_1 \mathbf{A}_1^T \times_2 \mathbf{A}_2^T \cdots \times_M \mathbf{A}_M^T \times_{M+1} \mathbf{Z}^{-1}, \quad (3)$$

式中 -1 为矩阵的 Moore-Penrose 逆。

将(3)式代入(2)式中可得:

$$\begin{aligned} f(\mathbf{G}, \mathbf{A}, \mathbf{Z}) &= \frac{1}{2} \|\mathbf{X} - \mathbf{G} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \cdots \times_M \mathbf{A}_M \times_{M+1} \mathbf{Z}\|_F^2 \\ &= \frac{1}{2} \|\mathbf{X}\|_F^2 - \langle \mathbf{X} \times_1 \mathbf{A}_1^T \cdots \times_M \mathbf{A}_M^T, \mathbf{G} \times_{M+1} \mathbf{Z} \rangle \\ &\quad + \frac{1}{2} \|\mathbf{G} \times_{M+1} \mathbf{Z}\|_F^2 = \\ &= \frac{1}{2} \|\mathbf{X}\|_F^2 - \frac{1}{2} \|\mathbf{X} \times_1 \mathbf{A}_1^T \cdots \times_M \mathbf{A}_M^T\|_F^2 + \\ &\quad \frac{1}{2} \|\mathbf{X} \times_1 \mathbf{A}_1^T \cdots \times_M \mathbf{A}_M^T - \mathbf{G} \times_{M+1} \mathbf{Z}\|_F^2, \quad (4) \end{aligned}$$

式中第一部分是常数项,所以 $f(\mathbf{G}, \mathbf{A}, \mathbf{Z})$ 求最小化函数可以表述如下:

$$\begin{aligned} h(\mathbf{G}, \mathbf{A}, \mathbf{Z}) &= \frac{1}{2} \|\mathbf{X} \times_1 \mathbf{A}_1^T \cdots \times_M \mathbf{A}_M^T - \\ &\quad \mathbf{G} \times_{M+1} \mathbf{Z}\|_F^2 - \frac{1}{2} \|\mathbf{X} \times_1 \mathbf{A}_1^T \cdots \times_M \mathbf{A}_M^T\|_F^2. \quad (5) \end{aligned}$$

实际上, $\mathbf{G} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \cdots \times_M \mathbf{A}_M$ 可以被看作是将原始张量 \mathbf{X} 压缩到矩阵 \mathbf{Z} 的张量“字典”。

矩阵的秩是一种结构性稀疏测度,低秩表征(LRR)通过检测并去除异常值,挖掘数据自身的结构信息,以找到数据的低维本征空间。低秩子空间学习在图像去噪^[33]、数据恢复^[34]以及去除异常值^[35]等方面均有广泛应用。受此启发,考虑对矩阵 \mathbf{Z} 施加一个低秩约束,从而得到内在的低维特征子空间。因此,包含正交和低秩约束的目标函数可以表示为

$$\begin{aligned} \min_{\mathbf{G}, \mathbf{A}_1, \dots, \mathbf{A}_M, \mathbf{Z}} h(\mathbf{G}, \mathbf{A}, \mathbf{Z}) + \gamma \text{Rank}(\mathbf{Z}) \\ \text{s.t. } \mathbf{A}_1^T \mathbf{A}_1 = \mathbf{I}_1, \dots, \mathbf{A}_M^T \mathbf{A}_M = \mathbf{I}_M, \quad (6) \end{aligned}$$

式中 γ 为惩罚参数。

由于(6)式中非凸函数 $\text{Rank}(\cdot)$ 的存在,核范数 $\|\mathbf{Z}\|_*$ 采用 \mathbf{Z} 近似表示。因此,(6)式可以表述为

$$\begin{aligned} \min_{\mathbf{G}, \mathbf{A}_1, \dots, \mathbf{A}_M, \mathbf{Z}} h(\mathbf{G}, \mathbf{A}, \mathbf{Z}) + \gamma \|\mathbf{Z}\|_* \\ \text{s.t. } \mathbf{A}_1^T \mathbf{A}_1 = \mathbf{I}_1, \dots, \mathbf{A}_M^T \mathbf{A}_M = \mathbf{I}_M. \quad (7) \end{aligned}$$

2.2 LRRHTD 算法

采用交替方向乘子法(ADMM)来求解(7)式的优化问题,将复杂问题分解为子问题求解。固定其他变量,迭代更新每个变量,直到收敛或达到最大迭代次数。

1) 固定其他变量,更新 \mathbf{G} :核张量 \mathbf{G} 可以通过(3)式每次迭代更新。

2) 固定 $\mathbf{G}, \mathbf{A}_1, \dots, \mathbf{A}_{m-1}, \mathbf{A}_{m+1}, \dots, \mathbf{A}_M$ 和 \mathbf{Z} ,求 \mathbf{A}_m :张量 \mathbf{X} 的第 m 模式展开矩阵表示可以转换为

$$\begin{aligned} \|\mathbf{X} \times_1 \mathbf{A}_1^T \times_2 \mathbf{A}_2^T \cdots \times_M \mathbf{A}_M^T\|_F^2 = \\ \|\mathbf{A}_m^T \mathbf{X}_{(m)} (\mathbf{A}_M \otimes \cdots \mathbf{A}_{m+1} \otimes \mathbf{A}_{m-1} \cdots \mathbf{A}_1)\|_F^2 = \\ \|\mathbf{A}_m^T \mathbf{X}_{(m)} \mathbf{A}_{(-m)}\|_F^2, \quad (8) \end{aligned}$$

式中 \otimes 为 Kronecker 积算子, $\mathbf{X}_{(m)}$ 为 \mathbf{X} 和 $\mathbf{A}_{(-m)} = \mathbf{A}_N \otimes \mathbf{A}_{N-1} \cdots \mathbf{A}_{m+1} \otimes \mathbf{A}_{m-1} \cdots \mathbf{A}_1$ 的 m 模式展开,相似地:

$$\begin{aligned} \|\mathbf{X} \times_1 \mathbf{A}_1^T \cdots \times_M \mathbf{A}_M^T - \mathbf{G} \times_{M+1} \mathbf{Z}\|_F^2 = \\ \|\mathbf{A}_m^T \mathbf{X}_{(m)} \mathbf{A}_{(-m)} - (\mathbf{G} \times_{M+1} \mathbf{Z})_{(m)}\|_F^2. \quad (9) \end{aligned}$$

为了简化表示,使 $\mathbf{B}_m = \mathbf{X}_{(m)} \mathbf{A}_{(-m)}$ 、 $\mathbf{C}_m = (\mathbf{G} \times_{M+1} \mathbf{Z})_{(m)}$, \mathbf{A}_m 通过解下面的最优化问题更新:

$$\min_{\mathbf{A}_m^T \mathbf{A}_m = \mathbf{I}} \frac{1}{2} \|\mathbf{A}_m^T \mathbf{B}_m - \mathbf{C}_m\|_F^2 - \frac{1}{2} \|\mathbf{A}_m^T \mathbf{B}_m\|_F^2. \quad (10)$$

以上函数可以简化为

$$\max_{\mathbf{A}_m^T \mathbf{A}_m = \mathbf{I}} \text{tr}(\mathbf{A}_m^T \mathbf{B}_m^T \mathbf{C}_m^T). \quad (11)$$

(11)式具有闭式解。例如, $\mathbf{P}_m = \mathbf{B}_m \mathbf{C}_m^T$ 最优解 \mathbf{A}_m 可以通过下式求得:

$$\mathbf{A}_m = \text{svd}(\mathbf{P}_m, \mathbf{J}_m), \quad (12)$$

式中 $\text{svd}(\mathbf{P}_m, \mathbf{J}_m)$ 即提取 \mathbf{P}_m 的前 \mathbf{J}_m 个左奇异值向量。

3) 固定其他变量,更新 \mathbf{Z} 。固定 $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_M$ 和 \mathbf{G} 得到下面的最优化问题:

$$\begin{aligned} \min_{\mathbf{Z}} \frac{1}{2} \|\mathbf{X} \times_1 \mathbf{A}_1^T \cdots \times_M \mathbf{A}_M^T - \mathbf{G} \times_{M+1} \mathbf{Z}\|_F^2 + \\ \gamma \|\mathbf{Z}\|_*. \quad (13) \end{aligned}$$

经过 $(M+1)$ 模式展开(13)式后得到:

$$\min_{\mathbf{Z}} \frac{1}{2} \|\mathbf{Z} \mathbf{G}_{(M+1)} - \mathbf{D}\|_F^2 + \gamma \|\mathbf{Z}\|_*, \quad (14)$$

式中 $\mathbf{D} = \mathbf{X}_{(M+1)} (\mathbf{A}_M \otimes \mathbf{A}_{M-1} \cdots \otimes \mathbf{A}_2 \otimes \mathbf{A}_1)$ 。

上述方程可以通过增广拉格朗日乘子法来解决。首先引入松弛变量 \mathbf{Q} 松弛原来的问题:

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{Q}} \frac{1}{2} \|\mathbf{Z} \mathbf{G}_{(M+1)} - \mathbf{D}\|_F^2 + \gamma \|\mathbf{Q}\|_* \\ \text{s.t. } \mathbf{Q} = \mathbf{Z}, \quad (15) \end{aligned}$$

然后,定义增广拉格朗日函数为

$$\begin{aligned} \mathcal{Q} = \frac{1}{2} \|\mathbf{Z} \mathbf{G}_{(M+1)} - \mathbf{D}\|_F^2 + \gamma \|\mathbf{Q}\|_* + \\ \langle \mathbf{Y}, \mathbf{Q} - \mathbf{Z} \rangle + \frac{\mu}{2} \|\mathbf{Q} - \mathbf{Z}\|_F^2, \quad (16) \end{aligned}$$

式中 \mathbf{Y} 为拉格朗日乘数, μ 为惩罚参数。

为了更好地解释和理解这个过程,引入一个变量 k 来表示第 k 次迭代,那么可以在 $k+1$ 次迭代中优化变量。

通过下式更新 \mathbf{Q} :

$$\mathbf{Q}^{k+1} = \arg \min_{\mathbf{Q}} \frac{\gamma}{\mu^k} \|\mathbf{Q}\|_* + \frac{1}{2} \|\mathbf{Q} - \mathbf{Z}^k + \mathbf{Y}^k / \mu^k\|_F^2. \quad (17)$$

(17)式可以通过奇异值阈值算法^[36]来求解,以近似求解核范数最小化问题。

通过下式更新 \mathbf{Z} :

$$\mathbf{Z}^{k+1} = \arg \min_{\mathbf{Z}} \frac{1}{\mu^k} \|\mathbf{Z}\mathbf{G}_{(M+1)} - \mathbf{D}\|_F^2 +$$

$$\|\mathbf{H}^{k+1} - \mathbf{Z}\|_F^2, \quad (18)$$

式中 $\mathbf{H}^{k+1} = \mathbf{Q}^{k+1} + \mathbf{Y}^k / \mu^k$ 。

令(18)式的导数为零,可以得到如下 \mathbf{Z} 的更新公式:

$$\mathbf{Z}^{k+1} = \left(\mathbf{H}^{k+1} + \frac{\mathbf{D}\mathbf{G}_{(M+1)}^T}{\mu^k} \right) \left(\frac{\mathbf{G}_{(M+1)}\mathbf{G}_{(M+1)}^T}{\mu^k} + \mathbf{I} \right)^{-1}. \quad (19)$$

表 1 所示为 LRRHTD 的聚类过程。

表 1 LRRHTD 算法流程

Table 1 Flow chat of LRRHTD algorithm

Algorithm 1 LRRHTD

Input: data tensor $\mathbf{X} \in \mathbf{R}^{I_1 \times I_2 \times \dots \times I_M \times N}$, parameter γ

Initialize: $\mathbf{Q}^0 = 0, \mathbf{Y} = 0, \mu = 10^{-6}, \mu^{\max} = 10^6, \phi = 1.3$

1. Initialize factor matrices $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \dots, \mathbf{A}_M, \mathbf{Z}$
2. Fix the others and update \mathbf{G} by Eq. (3)
3. **for** $m=1, 2, \dots, M$ **do**
4. **for** $t=0, 1, 2, \dots$ **do**
5. Fix the others and update \mathbf{A}_m^{t+1} by Eq. (12)
6. **end for**
7. **end for**
8. **for** $k=0, 1, 2, \dots$ **do**
9. Fix the others and update \mathbf{Q}^{k+1} by Eq. (17)
10. Fix the others and update \mathbf{Z}^{k+1} by Eq. (19)
11. Update the multiplier \mathbf{Y}^{k+1} by $\mathbf{Y}^{k+1} = \mathbf{Y}^k + \mu^k (\mathbf{Q}^{k+1} - \mathbf{Z}^{k+1})$
12. Update the multiplier μ^{k+1} by $\mu^{k+1} = \min(\phi\mu^k, \mu^{\max})$
13. **end for**
14. Iterate steps 2~13 until convergence
15. Perform K-means clustering^[37] on the row of \mathbf{Z}

Output: The cluster assignment of \mathbf{X}

2.3 算法复杂度分析

一个异构张量 $\mathbf{X} \in \mathbf{R}^{I_1 \times I_2 \times \dots \times I_M \times N}$, 其核张量为 $\mathbf{G} \in \mathbf{R}^{J_1 \times J_2 \times \dots \times J_N \times D}$, 其中 N 是样本数据的个数, D 是低秩表征维度, $J_m \leq I_m (m=1, 2, \dots, M)$, I_m 是样本数据的原始维度, J_m 是投影到子空间的维度。算法的时间复杂度主要来源于三个步骤。

1) 步骤 5 矩阵的奇异值分解: 对于大小为 $I_m \times J_m$ 的矩阵 \mathbf{P}_m , 奇异值分解的时间复杂度为 $O(I_m^3)$, 则 M 次奇异值分解总时间复杂度为 $O(\sum_{m=1}^M I_m^3)$ 。

2) 步骤 9 矩阵的奇异值阈值算法: 矩阵 \mathbf{Q} 的维度大小为 $N \times D$, 则 \mathbf{Q} 的奇异值阈值算法时间复杂度为 $O(N^3)$ 。

3) 步骤 10 矩阵的逆: 矩阵 $\mathbf{G}_{(M+1)}\mathbf{G}_{(M+1)}^T$ 大小为 $D \times D$ 维, 则步骤 10 时间复杂度为 $O(D^3)$ 。因此, 整个算法的时间复杂度为

$O(\sum_{m=1}^M I_m^3 + D^3 + N^3)$, 随着样本数量 N 的增加, 算法复杂度呈三次方上升。

3 实验结果与分析

为了验证本文算法的有效性, 在 Extended YaleB 人脸数据集和 MNIST 手写字符数据集上进行实验。采用两种聚类评价指标, 分别是准确度 (AC) 和归一化互信息 (NMI)。

3.1 数据集描述和实验设置

Extended YaleB 数据集由 38 人的 2414 张正面人脸图像组成。每人在实验室控制的照明条件下拍摄约 64 张图像。图像大小调整到 32 pixel \times 32 pixel, 如图 1 所示。MNIST 数据集是一个手写字符数据集, 包含了 0~9 共 10 个类, 如图 2 所示。训练集包含有 60000 个样本, 测试集包含有 10000 个样本。每个数字图像大小为 28 pixel \times 28 pixel。

选取类别 $K = \{3, 4, 5, 6, 7, 8, 9, 10\}$, 来评估算法在不同类别数目下的聚类性能。对于每一类, 随机选取 100 个样本, 运行算法 5 次取平均结果。在所有参数中, 参数 γ 在影响聚类性能方面起着重要的作

用, 可以通过网格搜索的算法选择 γ 。首先执行一个粗糙的网格。先确定理想区域, 然后在这个区域进行更好的网格搜索。对于 Extended YaleB 数据集, 设置 $\gamma = 1$; 而对于 MNIST 数据集, 设置 $\gamma = 2$ 。



图 1 Extended YaleB 数据集

Fig. 1 Extended YaleB dataset

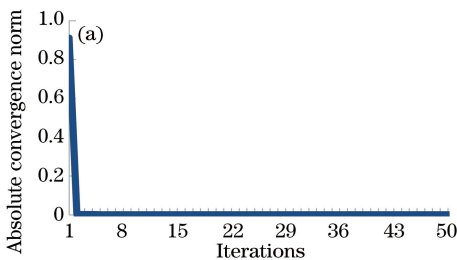
```

3 6 8 1 7 9 6 6 9 1
6 7 5 7 8 6 3 4 8 5
2 1 7 9 7 1 2 3 1 5
4 8 1 9 0 1 8 8 9 4
7 6 1 8 6 4 1 5 6 0
7 5 9 2 6 5 8 1 9 7
2 2 2 2 2 3 4 4 8 0
0 2 3 8 0 7 3 8 5 7
0 1 4 6 4 6 0 2 4 3
7 1 2 8 1 6 9 8 6 1

```

图 2 MNIST 数据集

Fig. 2 MNIST dataset



3.2 算法收敛性验证

为了验证本文算法的收敛性, 随机选择两个数据集上的实验结果进行展示。在张量分解模型中, 选取核张量 \mathbf{G} 两次更新的差值来判断模型是否达到收敛。由图 3 可知, 本文算法得到的 \mathbf{G} 的差值随着迭代次数的增加而迅速下降, 经过大约 10 次迭代后收敛。因此, 应用基于交替方向乘子法的更新方法, 保证了本文算法的收敛性。

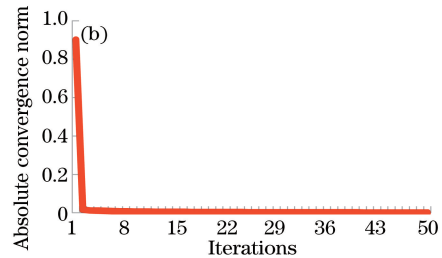


图 3 (a) Extended YaleB 和 (b) MNIST 数据集的收敛曲线

Fig. 3 Convergence curves of (a) the Extended YaleB and (b) MNIST datasets

3.3 聚类算法的比较

为了评估 LRRHTD 方法的有效性, 将本文算法与非负矩阵分解的多重可压缩 (MM) 方法^[38]、贝叶斯非负矩阵分解 (B-NFM)^[39] 方法、非负矩阵分解的投影梯度 (PGM)^[40] 方法、双正则协同聚类 (DRCC)^[41] 方法、高阶正交迭代 (HOOI)^[10] 方法、非负 Tucker 分解 (NTD)^[20] 方法, 以及 HTD-Multinomial^[21] 方法等进行比较。这些方法通过网格搜索来获得参数以保证最佳性能。

表 2 所示为不同算法在 Extended YaleB 数据集上的聚类结果。可以看出, 由于该数据集光照变化比较大, 所有的算法表现都不佳。与基于矩阵的方法 (如 MM、B-NFM 和 DRCC 方法) 相比, 基于张量的方法 (HOOI、NTD、HTD-Multinomial、LRRHTD 方法) 在 Extended YaleB 数据集上表现

较好, 表明在此数据集中张量建模的方法比基于矩阵的方法更具优势。而本文提出的 LRRHTD 算法在 AC 和 NMI 两个评价指标上都优于其他张量分解方法 (HOOI、NTD、HTD-Multinomial 方法), 表明了本文算法用于聚类的有效性。

表 2 不同算法在 Extended YaleB 数据集上的聚类结果

Table 2 Clustering performance with different algorithms for the Extended YaleB dataset

Method	AC	NMI
MM	0.2175	0.3534
B-NFM	0.2104	0.3508
PGM	0.2084	0.3477
DRCC	0.1002	0.1455
HOOI	0.2243	0.4113
NTD	0.2335	0.4238
HTD-Multinomial	0.3293	0.4772
LRRHTD	0.4225	0.5601

表 3 所示为不同算法在 MNIST 数据集上的聚类表现。从表 3 可以看出,除了 $K=7$ 外,本文提出的 LRRHTD 算法的表现优于其他算法,而当 $K=7$ 时,DRCC 算法取得了最好的表现。LRRHTD 和

HTD-Multinomial 方法都是异构张量分解算法,但是注意到 LRRHTD 方法比 HTD Multinomial 方法获得了更高的精度,这表明通过揭示数据中固有的低秩结构可以提高聚类性能。

表 3 不同算法在 MNIST 数据集上的聚类结果

Table 3 Clustering performance with different algorithms for the MNIST dataset

No.	MM method		B-NFM method		PGM method		DRCC method		HTD-Multinomial method		LRRHTD method	
	AC	NMI	AC	NMI	AC	NMI	AC	NMI	AC	NMI	AC	NMI
3	0.9173	0.7351	0.8820	0.6806	0.9160	0.7307	0.9013	0.7115	0.9114	0.7252	0.9333	0.7700
4	0.7104	0.4654	0.6780	0.4446	0.7195	0.4680	0.6575	0.4071	0.6800	0.4360	0.8350	0.6663
5	0.6216	0.5201	0.6104	0.4992	0.6764	0.5373	0.6632	0.5198	0.7196	0.5547	0.8440	0.6730
6	0.5370	0.4007	0.5156	0.3980	0.5466	0.3962	0.4983	0.3670	0.5540	0.4083	0.6633	0.5609
7	0.5988	0.5382	0.5762	0.4972	0.5940	0.5342	0.6585	0.5629	0.6231	0.5401	0.6042	0.5103
8	0.6455	0.5427	0.593	0.4979	0.5942	0.5152	0.6270	0.5033	0.6485	0.5254	0.6662	0.5507
9	0.5211	0.4621	0.5137	0.4343	0.5302	0.4720	0.5240	0.4493	0.5756	0.4774	0.5955	0.5004
10	0.4980	0.4470	0.4474	0.4089	0.4958	0.4516	0.4668	0.4053	0.5126	0.4461	0.5140	0.4953

4 结 论

提出了一种新颖的基于低秩正则化异构张量分解的子空间聚类算法。该算法对于原始的异构张量数据,寻求一组正交因子矩阵的集合,将高维张量映射到低维的潜在子空间中,同时在最后的因子矩阵上引入低秩约束以获得可用于聚类的全局低秩结构表征。此外,设计了基于增广拉格朗日乘子的优化方法对本文算法进行优化求解。在两个公共数据集上的实验结果验证了本文算法的可行性与聚类性能优越性。

参 考 文 献

- [1] Kolda T G, Bader B W. Tensor decompositions and applications[J]. SIAM Review, 2009, 51(3): 455-500.
- [2] Vu X T, Maire S, Chau C, *et al.* A new stochastic optimization algorithm to decompose large nonnegative tensors [J]. IEEE Signal Processing Letters, 2015, 22(10): 1713-1717.
- [3] Xu D, Liu H W, Wu L, *et al.* High-order singular value decomposition applied in aero-optical effects analysis[J]. Acta Optica Sinica, 2010, 30(12): 3367-3372.
许东, 刘浩伟, 吴琳, 等. 高阶奇异值分解在气动光学效应分析中的应用[J]. 光学学报, 2010, 30(12): 3367-3372.
- [4] Du L M, Li J, Jin G, *et al.* Compression of interference hyperspectral image based on FHALS-NTD[J]. Spectroscopy and Spectral Analysis, 2012, 32(11): 3155-3160.

杜丽敏, 李进, 金光, 等. 基于快速层次交替最小二乘非负张量 Tucker 分解的干涉高光谱图像光谱信息压缩方法[J]. 光谱学与光谱分析, 2012, 32(11): 3155-3160.

- [5] Li J, Han S L, Lü Z M, *et al.* Multi-spectral image compression based on Tucker decomposition in integer wavelet domain [J]. Journal of Optoelectronics • Laser, 2012, 23(12): 2418-2425.
李进, 韩双丽, 吕增明, 等. 整数小波域张量 Tucker 分解的多光谱图像压缩[J]. 光电子 • 激光, 2012, 23(12): 2418-2425.
- [6] Li Y, Xun X C, Wang Q Z. Image encryption based on higher-order singular value decomposition [J]. Infrared and Laser Engineering, 2014, 43(S1): 243-247.
李勇, 荀显超, 王青竹. 基于高维张量奇异值分解的图像加密[J]. 红外与激光工程, 2014, 43(S1): 243-247.
- [7] Xia J P, Zhou J L, He K, *et al.* Color image watermark based on DWT and Tucker decomposition [J]. Application Research of Computers, 2010, 27(2): 798-800.
夏建平, 周激流, 何坤, 等. 基于小波变换和 Tucker 分解的彩色图像数字水印[J]. 计算机应用研究, 2010, 27(2): 798-800.
- [8] Tucker L R. Some mathematical notes on three-mode factor analysis [J]. Psychometrika, 1966, 31(3): 279-311.
- [9] Kiers H A L. Towards a standardized notation and terminology in multiway analysis [J]. Journal of Chemometrics, 2015, 14(3): 105-122.
- [10] Kolda T G. Orthogonal tensor decompositions [J].

- SIAM Journal on Matrix Analysis and Applications, 2000, 23(1): 243-255.
- [11] De Lathauwer L, De Moor B, Vandewalle J. On the best rank-1 and rank-(R_1, R_2, \dots, R_N) approximation of higher-order tensors [J]. SIAM Journal on Matrix Analysis and Applications, 2000, 21(4): 2262-2271.
- [12] De Lathauwer L, Moor B D, Vandewalle J. A multilinear singular value decomposition [J]. SIAM Journal on Matrix Analysis and Applications, 2000, 21(4): 1253-1278.
- [13] Phan A H, Cichocki A. Tensor decompositions for feature extraction and classification of high dimensional datasets [J]. Nonlinear Theory and Its Applications, 2010, 1(1): 37-68.
- [14] Zhang Z Y, Li T, Ding C. Non-negative tri-factor tensor decomposition with applications [J]. Knowledge and Information Systems, 2013, 34(2): 243-265.
- [15] Zhou G X, Cichocki A, Zhao Q B, *et al.* Nonnegative matrix and tensor factorizations: an algorithmic perspective [J]. IEEE Signal Processing Magazine, 2014, 31(3): 54-65.
- [16] Cohen J, Farias R C, Comon P. Fast decomposition of large nonnegative tensors [J]. IEEE Signal Processing Letters, 2015, 22(7): 862-866.
- [17] Lee D D, Seung H S. Learning the parts of objects by non-negative matrix factorization [J]. Nature, 1999, 401: 788-791.
- [18] Shashua A, Hazan T. Non-negative tensor factorization with applications to statistics and computer vision [C] // Proceedings of the 22nd International Conference on Machine, 2005: 792-799.
- [19] Wu F, Tan X, Yang Y, *et al.* Supervised nonnegative tensor factorization with maximum-margin constraint [C] // Proceedings of the 27th AAAI Conference on Artificial Intelligence, 2013: 962-968.
- [20] Zhou G X, Cichocki A, Xie S L. Fast nonnegative matrix/tensor factorization based on low-rank approximation [J]. IEEE Transactions on Signal Processing, 2012, 60(6): 2928-2940.
- [21] Sun Y F, Gao J B, Hong X, *et al.* Heterogeneous tensor decomposition for clustering via manifold optimization [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(3): 476-489.
- [22] Zhou G X, Cichocki A, Zhao Q B, *et al.* Efficient nonnegative Tucker decompositions: algorithms and uniqueness [J]. IEEE Transactions on Image Processing, 2015, 24(12): 4990-5003.
- [23] Lai Z H, Xu Y, Yang J, *et al.* Sparse tensor discriminant analysis [J]. IEEE Transactions on Image Processing, 2013, 22(10): 3904-3915.
- [24] Heiler M, Schnörr C. Controlling sparseness in non-negative tensor factorization [C] // Proceedings of European Conference on Computer Vision, 2006: 56-67.
- [25] Liu J, Liu J, Wonka P, *et al.* Sparse non-negative tensor factorization using columnwise coordinate descent [J]. Pattern Recognition, 2012, 45(1): 649-656.
- [26] Perros I, Chen R, Vuduc R, *et al.* Sparse Hierarchical Tucker factorization and its application to healthcare [C] // Proceedings of IEEE International Conference on Data Mining, 2015: 943-948.
- [27] Zhang C Q, Fu H Z, Liu S, *et al.* Low-rank tensor constrained multiview subspace clustering [C] // Proceedings of IEEE International Conference on Computer Vision, 2015: 1582-1590.
- [28] Jia C C, Fu Y. Low-rank tensor subspace learning for RGB-D action recognition [J]. IEEE Transactions on Image Processing, 2016, 25(10): 4641-4652.
- [29] Fu Y F, Gao J B, Tien D, *et al.* Tensor LRR and sparse coding-based subspace clustering [J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 27(10): 2120-2133.
- [30] Liu J, Musialski P, Wonka P, *et al.* Tensor completion for estimating missing values in visual data [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 208-220.
- [31] Wang H, Nie F P, Huang H. Low-rank tensor completion with spatio-temporal consistency [C] // Proceedings of the 28th AAAI Conference on Artificial Intelligence, 2014: 2846-2852.
- [32] Yang Y N, Feng Y L, Suykens J A K. A rank-one tensor updating algorithm for tensor completion [J]. IEEE Signal Processing Letters, 2015, 22(10): 1633-1637.
- [33] Luo L, Feng X C, Zhang X D, *et al.* An image denoising method based on non-local two-side random projection and low rank approximation [J]. Journal of Electronics & Information Technology, 2013, 35(1): 99-105.
- 罗亮, 冯象初, 张选德, 等. 基于非局部双边随机投影低秩逼近图像去噪算法 [J]. 电子与信息学报, 2013, 35(1): 99-105.

- [34] Peng Y G, Suo J L, Dai Q H, *et al.* From compressed sensing to low-rank matrix recovery: theory and applications[J]. *Acta Automatica Sinica*, 2013, 39(7): 981-994.
彭义刚, 索津莉, 戴琼海, 等. 从压缩传感到低秩矩阵恢复:理论与应用[J]. *自动化学报*, 2013, 39(7): 981-994.
- [35] Qiu L D, Fu P, Lin N, *et al.* Discriminative low-rank projection dictionary pair learning[J]. *Laser & Optoelectronics Progress*, 2016, 53(11): 111003.
邱立达, 傅平, 林南, 等. 具有判别能力的低秩投影字典对学习[J]. *激光与光电子学进展*, 2016, 53(11): 111003.
- [36] Cai J F, Candès E J, Shen Z W. A singular value thresholding algorithm for matrix completion [J]. *SIAM Journal on Optimization*, 2010, 20(4): 1956-1982.
- [37] Zhou W Z, Chen G L, Du S S, *et al.* An improved iterative closest point algorithm using clustering[J]. *Laser & Optoelectronics Progress*, 2016, 53(5): 051202.
周文振, 陈国良, 杜珊珊, 等. 一种聚类改进的迭代最近点配准算法[J]. *激光与光电子学进展*, 2016, 53(5): 051202.
- [38] Lee D D, Seung H S. Algorithms for non-negative matrix factorization [C] // *Proceedings of the 13th International Conference on Neural Information Processing Systems*, 2000: 535-541.
- [39] Schmidt M N, Winther O, Hansen L K. Bayesian non-negative matrix factorization[C] // *Proceedings of the 8th International Conference on Independent Component Analysis and Signal Separation*, 2009: 540-547.
- [40] Lin C J. Projected gradient methods for nonnegative matrix factorization[J]. *Neural Computation*, 2007, 19(10): 2756-2779.
- [41] Gu Q Q, Zhou J. Co-clustering on manifolds [C] // *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2009: 359-368.