

基于 $L_{2,1}$ 范数的超图正则化非负矩阵分解

文学春

贵州师范大学数学科学学院, 贵州 贵阳

收稿日期: 2022年12月28日; 录用日期: 2023年1月21日; 发布日期: 2023年1月31日

摘要

图学习的方法只考虑了样本间的关系, 忽略了样本间的高维结构, 基于欧氏距离的非负矩阵模型易受到噪声和异常值的影响。为解决以上问题, 给出一种基于 $L_{2,1}$ 范数的超图正则化非负矩阵分解($L_{2,1}$ HNMFL)方法, 并给出了算法的更新规则。通过在COIL20和ORL经典数据集上与其他算法的比较验证了该算法的有效性。

关键词

$L_{2,1}$ 范数, 超图学习, 聚类分析, 非负矩阵分解

Hypergraph Regular Nonnegative Matrix Factorization Based on $L_{2,1}$ Norms

Xuechun Wen

School of Mathematical Sciences, Guizhou Normal University, Guiyang Guizhou

Received: Dec. 28th, 2022; accepted: Jan. 21st, 2023; published: Jan. 31st, 2023

Abstract

The graph learning method only considers the relationship between samples, but ignores the high-dimensional structure between samples. The non-negative matrix model based on Euclidean distance is easily affected by noise and outliers. To solve the above problems, a hypergraph regularization nonnegative matrix factorization ($L_{2,1}$ HNMFL) method based on $L_{2,1}$ norm is presented, and the updating rules of the algorithm are given. The effectiveness of this algorithm is verified by comparing it with other algorithms on classical data sets COIL20 and ORL.

Keywords

$L_{2,1}$ Norm, Hypergraph Learning, Cluster Analysis, Nonnegative Matrix Decomposition

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

聚类是特征学习和计算机视觉中的一项重要但具有挑战性的任务。对于图像聚类任务，最重要的是找到原始数据的有效表示。为了挖掘出原始数据中内在的结构信息，许多研究人员已经做了大量的工作。最流行的特征提取方法包括主成分分析[1] (PCA)、奇异值分解[2] (SVD)、概念分解[3] (CF)非负矩阵分解[4] (NMF)等。1999年，Lee和Seung [4]在《自然》杂志上首次提出了非负矩阵分解的概念。NMF的目标是得到两个非负矩阵，基矩阵和编码矩阵，其乘积尽可能接近原始矩阵。由于NMF的良好性能，在Lee和Seung的工作之后，许多标准NMF的变体被提出，用于从不同的观点改进NMF。

为了提高NMF的稀疏性，Li等人提出了一种加入局部约束的局部NMF [5] (LNMF)。在此之后，Chen等人在NMF中引入了局部坐标约束[6]。Cai等人提出了一个图正则化[7] NMF (GNMF)，试图涉及数据空间的几何信息。GNMF通过构造一个k最近邻(kNN)图来对几何结构进行编码。因为一个简单图中的一条边只能连接两个顶点，GNMF只考虑了两个样本之间的成对关系，而忽略了样本间的高维结构。为了探索样本间的高阶关系，Zeng等人通过构建一个编码更多样本之间关系的超图，开发了一个超图正则化[8]，因为超图可以连接两个以上的顶点，HNMF可以找出数据的高阶关系。

上述方法在某种程度上确实改进了标准的NMF。然而，由于原始数据中所包含的潜在噪声会影响聚类和分类的性能。当数据包含噪声或离群值时，特征和样本的误差都是平方的，因此有一些噪声特征或一些离群值的误差较大将支配目标函数。Kong等人提出了鲁棒非负矩阵分解算法[9]，克服了标准NMF算法对噪声和异常值敏感的缺点。

为了减轻噪声或异常值的影响，处理高维、包含噪声和异常值的数据。整合 $L_{2,1}$ 范数和超图正则化的优点，给出了一种基于 $L_{2,1}$ 范数的超图正则化非负矩阵分解。在COIL20和ORL数据集上的实验结果表明， $L_{2,1}$ HNMF方法在聚类精度和归一化互信息上优于四种具有代表性的方法k-means、NMF、 $L_{2,1}$ NMF和GNMF。在第二节中，简要回顾了标准NMF及其一些具有代表性的变体。在第三节中，详细描述了 $L_{2,1}$ HNMF方法并给出了有效的迭代更新规则。第四节展示了在数据集上的实验结果。最后，在第五节中简要地总结了这篇论文。

2. 预备知识

给定的非负数据矩阵 $X = (x_1, x_2, \dots, x_n)$ ， $x_i \in R^m$ 表示一个数据点。标准NMF定义为：

$$\min_{U, V} \|X - UV\|_F^2, U \geq 0, V \geq 0$$

上述问题通常通过迭代更新算法来解决，其中 U 和 V 的迭代更新方法如下：

$$u_{ik} \leftarrow u_{ik} \frac{(XV^T)_{ik}}{(UVV^T)_{ik}},$$

$$v_{ki} \leftarrow v_{ki} \frac{(U^T X)_{kj}}{(U^T UV)_{kj}}.$$

为了提高 NMF 的鲁棒性, Kong 等人提出 $L_{2,1}$ NMF 算法, 能够降低数据集中噪声和异常值的影响。目标函数定义为:

$$\min_{U,V} \|X - UV\|_{2,1}, U \geq 0, V \geq 0$$

其中 $L_{2,1}$ 范数定义为 $\|Y\|_{2,1} = \sum_{i=1}^m \sqrt{\sum_{j=1}^n Y_{ij}^2} = \sum_{i=1}^m \|Y_{i,:}\|_2$

迭代更新规则为

$$u_{ik} \leftarrow u_{ik} \frac{(XGV^T)_{ik}}{(UVGV^T)_{ik}},$$

$$v_{kj} \leftarrow v_{kj} \frac{(U^T XG)_{kj}}{(U^T UVG)_{kj}}.$$

为了保持数据的局部几何结构, Cai 等人提出了 GNMF, 目标函数为

$$\min_{U \geq 0, V \geq 0} \|X - UV\|_F^2 + \lambda \text{Tr}(VLV^T)$$

其中 $L = D - W$ 是图拉普拉斯矩阵。 W 是图的权重矩阵, D 是对角矩阵, $D_{ii} = \sum_j W_{ij}$ 。

因为一个简单图中的一条边只能连接两个顶点。为了考虑样本之间的高维结构, Zeng 等人提出了超图正则化。下面将简单地介绍超图的一些基本定义[10]:

给定一个超图 $G = (A, E, S)$, 其中 A 表示超图中所有顶点的集合, E 表示所有超边的集合, 任意一个超边 $e \in E$ 是顶点集 A 的一个子集。 S 是超图的权重矩阵, 它是一个对角矩阵, 元素 $s(e)$ 表示超边 e 权重的大小。把一个顶点和一条超边的关系记为关联矩阵 H , 则关联矩阵 H 表示为 $H(a_i, e_j) = \begin{cases} 0 & a_i \notin e_j \\ 1 & a_i \in e_j \end{cases}$ 。

对于任意顶点 $a \in A$, 顶点 a 的度定义为 $d(a) = \sum_{e \in E} s(e)H(a, e)$, 对于超边 $e \in E$, 它的度定义为 $\delta(e) = \sum_{a \in A} H(a, e)$ 。超图的权重矩阵为 S , 其元素 s_{ij} 表示为 $s_{ij} = \sum_{e \in E} \sum_{(i,j) \in e} \frac{s(e)}{\delta(e)}$ 。超图拉普拉斯矩阵 $L_{Hyper} = D_a - B$, $B = HSD_e^{-1}H^T$ 。

3. 基于 $L_{2,1}$ 范数的超图正则化非负矩阵分解

在这一部分中, 给出了基于 $L_{2,1}$ 范数的超图正则化非负矩阵分解算法($L_{2,1}$ HNMFL), 不仅可以克服数据中噪声和异常值的影响, 还可以保持数据集的内在几何结构, 得到更加准确的聚类效果。下面先介绍目标中正则化项。

3.1. 目标函数

首先介绍坐标编码的概念[11]。

定义: 坐标编码是一对 (γ, C) , $C \in R^d$ 是一组锚点, γ 是 $x \in R^d$ 到 $[\gamma_v(x)]_{v \in C} \in R^{|C|}$ 的映射。得到物

理近似式: $\gamma(x) = \sum_{v \in C} \gamma_v(x)v$ 。

根据上式, 基矩阵 U 的列可以看作是一组锚点, 而数据集中的每个数据点都可以用锚点的线性组合来表示。编码矩阵 V 的列包含了数据点相对于锚点的坐标。为了获得稀疏编码, 每个数据点都应该表示为只有几个附近锚点的线性组合, 即每个数据点应该只足够接近几个锚点。这可以通过引入局部坐标约束来实现, 局部坐标约束可以表述为以下问题:

$$J_1 = \sum_{k=1}^K |v_{ki}| \|u_k - x_i\|^2$$

将上式整合到鲁棒的 NMF 中, 并且增加超图正则项, 目标函数定义为:

$$J_{L_{21}HNMFL} = \|X - UV\|_{2,1} + \alpha \sum_{i=1}^N \sum_{k=1}^K |v_{ki}| \|u_k - x_i\|^2 + \lambda Tr(VL_{Hyper}V^T) \quad (1)$$

$$\text{s.t. } U = [u_1, \dots, u_K] \in R^{M \times K} > 0, V = [v_1, \dots, v_N] \in R^{K \times N} > 0.$$

3.2. 更新规则

下面给出更新规则和推导过程, 经过简单运算, (1)式可以改写为:

$$J_{L_{21}HNMFL} = \|X - UV\|_{2,1} + \alpha \sum_{i=1}^N \left\| (x_i 1^T - U) \Lambda_i^{1/2} \right\|_F^2 + \lambda Tr(VL_{Hyper}V^T)$$

其中 α 和 λ 是大于 0 的常数。目标函数的第二项和第三项分别是局部坐标约束项和超图正则化项。 $\Lambda_i = \text{diag}(|V_i|) \in R^{K \times K}$, 根据迹的知识, 目标函数可以写为:

$$J_{L_{21}HNMFL} = 2Tr \left(X^T IX - 2X^T IUV + V^T U^T IUV + \mu \sum_{i=1}^N (x_i 1^T \Lambda_i 1 x_i^T - 2x_i 1^T \Lambda_i U^T + U \Lambda_i U^T) \right) + \lambda Tr(VL_{Hyper}V^T)$$

其中, $I_{ii} = 1/2 \sqrt{\sum_{j=1}^n (X - UV)_{ij}^2}$, $L_{Hyper} = D_a - B$, $B = HSD_e^{-1} H^T$ 。

设 ψ_{ik} 和 ϕ_{kj} 分别是 u_{ik} 和 v_{kj} 的拉普拉斯乘子。 $\Psi = [\psi_{ik}]$ 和 $\Phi = [\phi_{kj}]$ 是对应的拉普拉斯矩阵, 则拉普拉斯扩展式 L 为:

$$L = \|X - UV\|_{2,1} + \alpha \sum_{i=1}^N \left\| (x_i 1^T - U) \Lambda_i^{1/2} \right\|_F^2 + \lambda Tr(VL_{Hyper}V^T) - Tr(\Psi U^T) - Tr(\Phi V^T)$$

令 $\frac{\partial L}{\partial U} = 0$ 和 $\frac{\partial L}{\partial V} = 0$, 得到:

$$\Psi = -2IXV^T + 2IUVV^T + \mu \sum_{i=1}^N (-2x_i 1^T \Lambda_i + 2U \Lambda_i)$$

$$\Phi = -2U^T IX + 2U^T IUV + \mu (C - 2U^T X + D) + 2\lambda VL_{Hyper}$$

定义列向量 $C = (c, \dots, c)^T$ 。设 $C = (c, \dots, c)^T$ 是一个 $K \times N$ 矩阵, 其行为 c^T 。定义列向量 $p = \text{diag}(U^T U) \in R^K$ 。设 $P = (p, \dots, p)$ 为一个 $K \times N$ 矩阵, 其列为 p 。根据上式, 关于 $L_{2,1}HNMFL$ 算法的迭代公式为:

$$u_{jk} \leftarrow u_{jk} \frac{\left(IXV^T + \mu \sum_{i=1}^N (x_i 1^T \Lambda_i) \right)_{jk}}{\left(IUVV^T + \mu \sum_{i=1}^N U \Lambda_i \right)_{jk}}, \quad (2)$$

$$v_{ki} \leftarrow v_{ki} \frac{\left(2U^T IX + 2\mu U^T X + 2\lambda VHS D_e^{-1} H^T \right)_{ki}}{\left(2U^T IUV + \mu C + \mu D + 2\lambda V D_a \right)_{ki}}. \quad (3)$$

4. 实验

4.1. 数据集的描述

ORL 数据集: ORL 人脸数据库共由总共 400 张人脸图像组成, 为 40 个个体在不同光照条件、有/没有眼镜和不同面部表情的情况下的 10 张灰度人脸图像。所有图像的分辨率都被调整为 32×32 。

COIL20 数据集: 数据集由 20 个物体的 1440 张图像组成, 灰度大小为 32×32 。每个物体有 72 张不同的图像, 所有的图像都是黑色的背景。

4.2. 比较的方法

为了验证算法的有效性, 与四种具有代表性的方法 k-means、NMF、 L_2 NMF 和 GNMF 进行比较。

4.3. 实验设计

在实验中, 使用了两个通常使用的测量方法来评估算法的效果, 分别是聚类精度(AC)和归一化互信息(NMI) [12]。

聚类精度(AC)定义为:

$$AC = \frac{\sum_{i=1}^n \delta(gnd_i, map(r_i))}{n}$$

其中 gnd_i 是数据集给出的标签, r_i 是聚类算法提供的标签, $map(r_i)$ 是将 r_i 映射到等价标签的排列映射函数, $\delta(x, y)$ 是 delta 函数, 如果 $x = y$, 其值为 1, 否则为 0。

互信息(NMI)可以衡量两个集簇之间的相似性, 在聚类应用中得到了广泛的应用。设 C 和 C' 为两组集, 然后将 MI 定义为

$$MI(C, C') = \sum_{c_i \in C, c'_j \in C'} p(c_i, c'_j) \log \frac{P(c_i, c'_j)}{p(c_i) \cdot p(c'_j)}$$

其中, $p(c_i)$ 和 $p(c'_j)$ 分别表示属于集簇 C 和集簇 C' 的样本的概率。 $P(c_i, c'_j)$ 表示一个样本同时属于 C 和 C' 的联合概率。然后将 NMI 定义为:

$$NMI(C, C') = \frac{MI(C, C')}{\max(H(C), H(C'))}$$

其中 $H(C)$ 和 $H(C')$ 分别表示 C 和 C' 的信息熵。

在实验中, 在一定范围内改变了参数, 以寻找所给出的方法的最佳性能, 正则化参数 μ 和 λ 在 $\{0.001, 0.01, 0.1, 1, 10, 100, 1000\}$ 中进行选择。所有方法对上述数据集的聚类结果如表 1~4 所示。对于每个给定的聚类数 k , 从数据集中随机选择的不同类重复 10 次测试运行, 并聚类性能的方差和平均值。

Table 1. Clustering accuracy on COIL20
表 1. COIL20 上的聚类准确度(AC)

k	KM	NMF	$L_{2,1}$ NMF	GNMF	$L_{2,1}$ HNMFL
5	83.75 ± 10.97	82.07 ± 5.63	87.62 ± 7.54	76.45 ± 19.98	93.71 ± 0.47
7	73.33 ± 6.58	73.98 ± 6.01	71.50 ± 6.46	69.58 ± 17.74	82.57 ± 5.92
8	71.46 ± 6.89	71.64 ± 3.38	72.57 ± 4.18	69.13 ± 9.09	75.86 ± 0.73
11	62.44 ± 10.71	68.09 ± 4.41	66.68 ± 5.02	62.64 ± 7.23	74.88 ± 3.21
12	59.02 ± 5.80	63.50 ± 3.83	62.57 ± 3.12	60.29 ± 5.82	68.57 ± 2.58
13	61.93 ± 9.04	62.60 ± 4.80	64.87 ± 4.95	65.87 ± 7.01	72.66 ± 4.68
15	59.83 ± 6.10	62.19 ± 4.50	65.30 ± 5.10	65.22 ± 9.49	74.41 ± 2.72
18	60.25 ± 8.78	61.69 ± 4.21	60.86 ± 3.71	63.02 ± 6.74	66.33 ± 2.21
20	58.23 ± 4.60	61.59 ± 3.24	59.40 ± 4.97	63.09 ± 5.92	64.72 ± 2.95
Avg	65.58	67.48	67.93	66.14	74.86

Table 2. Normalized Mutual Information on COIL20
表 2. COIL20 上的归一化互信息(NMI)

k	KM	NMF	$L_{2,1}$ NMF	GNMF	$L_{2,1}$ HNMFL
5	84.66 ± 7.53	83.84 ± 5.71	83.52 ± 7.30	83.95 ± 14.65	88.91 ± 0.75
7	80.89 ± 2.84	78.85 ± 4.59	80.67 ± 3.97	81.11 ± 10.67	83.69 ± 5.39
8	75.61 ± 5.14	78.02 ± 1.70	76.04 ± 3.12	76.86 ± 5.28	79.56 ± 0.72
11	72.28 ± 6.17	74.07 ± 2.05	74.68 ± 2.29	75.39 ± 4.27	78.10 ± 2.17
12	66.74 ± 2.89	69.74 ± 2.49	69.00 ± 2.52	70.22 ± 3.17	74.08 ± 2.00
13	72.73 ± 5.80	74.63 ± 2.77	73.79 ± 3.01	74.36 ± 3.60	77.87 ± 3.19
15	73.37 ± 3.14	75.70 ± 2.38	75.50 ± 2.57	75.45 ± 5.44	79.56 ± 1.85
18	73.75 ± 4.44	72.44 ± 2.16	74.44 ± 1.81	73.86 ± 3.17	75.51 ± 1.10
20	73.53 ± 1.94	73.26 ± 1.67	74.18 ± 2.56	74.12 ± 2.73	75.80 ± 1.63
Avg	74.84	75.62	75.76	76.15	79.23

Table 3. Clustering accuracy on ORL
表 3. ORL 上的聚类准确度(AC)

k	KM	NMF	$L_{2,1}$ NMF	GNMF	$L_{2,1}$ HNMFL
5	73.80 ± 13.74	82.20 ± 8.60	78.05 ± 7.72	79.05 ± 10.42	86.50 ± 2.63
10	59.60 ± 6.77	60.57 ± 6.83	62.58 ± 6.66	62.93 ± 7.35	71.50 ± 3.17
15	51.20 ± 4.17	57.93 ± 5.33	60.17 ± 5.32	61.82 ± 5.75	62.13 ± 4.00

Continued

20	55.80 ± 4.78	59.47 ± 3.52	64.33 ± 6.47	61.94 ± 5.01	69.21 ± 3.23
25	54.16 ± 4.00	60.11 ± 3.46	62.74 ± 4.08	61.62 ± 4.27	67.89 ± 3.17
30	51.33 ± 3.64	55.54 ± 4.72	57.22 ± 3.71	59.23 ± 4.09	62.78 ± 3.00
35	51.74 ± 2.66	55.16 ± 3.56	58.23 ± 2.71	57.57 ± 3.30	60.15 ± 2.77
40	52.72 ± 2.44	54.20 ± 2.79	55.44 ± 3.67	56.84 ± 3.68	59.55 ± 2.70
Avg	56.29	60.65	62.33	62.63	67.46

Table 4. Normalized Mutual Information on ORL**表 4.** ORL 上的归一化互信息(NMI)

k	KM	NMF	$L_{2,1}$ NMF	GNMF	$L_{2,1}$ HNMFL
5	73.41 ± 11.76	80.60 ± 7.68	75.59 ± 6.15	80.33 ± 8.18	82.59 ± 4.09
10	68.95 ± 4.70	68.26 ± 4.84	70.42 ± 5.08	69.94 ± 2.30	77.21 ± 2.10
15	62.88 ± 3.68	67.99 ± 4.04	70.56 ± 3.88	71.92 ± 3.85	71.08 ± 2.62
20	70.39 ± 3.66	73.78 ± 2.44	77.22 ± 3.30	74.93 ± 2.95	79.06 ± 2.01
25	71.67 ± 2.52	75.61 ± 2.00	77.40 ± 2.29	77.17 ± 2.38	79.94 ± 1.60
30	69.96 ± 1.61	72.87 ± 2.69	73.66 ± 2.21	75.36 ± 2.27	76.33 ± 1.51
35	70.78 ± 1.74	72.84 ± 1.93	74.89 ± 1.45	75.02 ± 1.75	75.83 ± 1.72
40	72.47 ± 1.69	73.43 ± 1.60	74.00 ± 2.02	75.55 ± 1.67	76.63 ± 1.51
Avg	70.06	73.17	74.22	75.03	77.33

4.4. 实验结果与分析

通过以上两个数据集上的实验结果进行分析，可以得出：

1) k-means 的聚类性能不如其他方法。这表明，通过其他方法获得的高维样本的低维表示在数据聚类中是有用的，并可以提高聚类性能。2) 在大多数情况下，GNMF 的性能优于 NMF，这验证了几何结构在矩阵分解算法中的重要作用。3) 由于 $L_{2,1}$ HNMFL 算法是利用局部约束来保持几何结构，并利用 $L_{2,1}$ 范数来保持稀疏性，故 $L_{2,1}$ HNMFL 算法优于所有其他算法。4) 与其他方法相比，给出的 $L_{2,1}$ HNMFL 算法在测试的 AC 和 NMI 值上都获得最好的聚类性能。这表明，通过将超图的构造和局部约束合并到一个统一的框架中， $L_{2,1}$ HNMFL 算法比其他方法具有更强的判别能力。

5. 结语

新方法给出了一个基于超图正则化和局部约束的聚类模型，可以有效地处理高维、稀疏和有噪声的数据。在 COIL20 数据集和 ORL 数据集上，聚类精度分别提高了 6.90% 和 4.83%，归一化互信息分别提高了 3.08% 和 2.30%。

参考文献

[1] Fukunada, K. (1990) Introduction to Statistical Pattern Recognition. In: Rheinboldt, W. and Siewiorek, D., Eds., *Com-*

-
- puter Science and Scientific Computing*, Academic Press, Cambridge.
- [2] Kalman, D. (1996) A Singularly Valuable Decomposition: The SVD of a Matrix. *The College Mathematics Journal*, **27**, 2-23. <https://doi.org/10.1080/07468342.1996.11973744>
 - [3] Xu, W. and Gong, Y.H. (2004) Document Clustering by Concept Factorization. *Research and Development in Information Retrieval*.
 - [4] Lee, D.D. and Seung, H.S. (1999) Learning the Parts of Objects by Non-Negative Matrix Factorization. *Nature*, **401**, 788-791. <https://doi.org/10.1038/44565>
 - [5] Li, S.Z., Hou, X.W., Zhang, H.J., et al. (2001) Learning Spatially Localized, Parts-Based Representation. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, **1**, 1-1.
 - [6] Chen, Y., Zhang, J., Cai, D., et al. (2012) Nonnegative Local Coordinate Factorization for Image Representation. *IEEE Transactions on Image Processing*, **22**, 969-979. <https://doi.org/10.1109/TIP.2012.2224357>
 - [7] Deng, C., et al. (2010) Graph Regularized Nonnegative Matrix Factorization for Data Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **33**, 1548-1560. <https://doi.org/10.1109/TPAMI.2010.231>
 - [8] Zeng, K., Yu, J., Li, C., et al. (2014) Image Clustering by Hyper-Graph Regularized Non-Negative Matrix Factorization. *Neurocomputing*, **138**, 209-217. <https://doi.org/10.1016/j.neucom.2014.01.043>
 - [9] Kong, D.G., Ding, C. and Huang, H. (2011) Robust Nonnegative Matrix Factorization Using L21-Norm. *Information and Knowledge Management*.
 - [10] Zhou, D., Huang, J. and Schölkopf, B. (2006) Learning with Hypergraphs: Clustering, Classification, and Embedding. *Advances in Neural Information Processing Systems*, **19**.
 - [11] Yu, K., Zhang, T. and Gong, Y. (2009) Nonlinear Learning Using Local Coordinate Coding. *Advances in Neural Information Processing Systems*, **22**.
 - [12] Xu, W., Liu, X. and Gong, Y.H. (2003) Document Clustering Based on Non-Negative Matrix Factorization. *Research and Development in Information Retrieval*.