

Sparse Tensor Decomposition Based on L_q Norm

Yu Jiang

College of Mathematics and Information Science, Guangxi University, Nanning Guangxi
Email: 1391866790@qq.com

Received: Feb. 25th, 2020; accepted: Mar. 9th, 2020; published: Mar. 16th, 2020

Abstract

Sparsity in tensor decomposition has solved some difficulties encountered in actual application. This paper simply introduces related concepts of tensor and CP (CANDECOMP/PARAFAC) decomposition. Then, we propose the decomposition based on L_q norm ($0 < q < 1$), and solve it by Alternating Direction of Method of Multipliers. This paper also gives an explicit solution when $q = \frac{1}{2}$.

Keywords

Sparsity, Tensor Decomposition, Alternating Direction of Method of Multipliers

基于 L_q 范数的稀疏张量分解问题

姜 玉

广西大学数学与信息科学学院, 广西 南宁
Email: 1391866790@qq.com

收稿日期: 2020年2月25日; 录用日期: 2020年3月9日; 发布日期: 2020年3月16日

摘 要

在张量分解中, 稀疏性的引入解决了实际应用中遇到的一些困难, 本文简单的介绍了张量及张量CP分解的相关概念, 在已有研究的基础上提出基于 L_q 范数($0 < q < 1$)的稀疏张量分解问题, 并使用乘子交替方向法来进行求解, 以及给出 $q = \frac{1}{2}$ 时问题解的形式。

关键词

稀疏性, 张量分解, 乘子交替方向法

Copyright © 2020 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

张量分解可以应用于多维数据的数据压缩、可视化、探索, 这使得它在神经成像、人脸识别、核磁共振成像等领域有着重要的作用。而实际应用的过程中也会遇到一些困难, 如大量的数据需要占据较大的内存空间, 计算复杂度较高。为了解决这样的问题, 学者们提出将稀疏性引入到张量分解的问题中。

在文献[1]中, 作者给出了稀疏性可取的原因: 第一, 稀疏性进一步压缩了多维数据集使得减轻了存储压力; 第二, 因为许多特征并不活跃, 稀疏性的引入为高维张量的特征选取提供了一种自动的工具; 第三, 在高维张量下, 数据的可视化和解释较为困难, 而稀疏性限制了特征的数量从而简化了数据分析结果的可视化和解释。

稀疏性不仅具有良好的理论性质, 而且在文献[1] [2] [3] [4]中通过数值结果表明在稀疏情况下考虑的张量分解有很好的数值性能。

1.1. 张量的符号和运算

为了讨论张量分解以及进一步地去讨论带有稀疏性的张量分解, 我们需要简单地了解张量的符号及相关运算, 文献[5] [6]给出了这些内容。张量作为向量和矩阵在高维空间的推广, 可以容纳较大的多维数据。若张量为二阶, 我们可以将其看做为向量, 三阶的张量则为矩阵。三阶及三阶以上的张量成为高阶张量, 我们所讨论的张量常指高阶张量。

张量常用 $\mathcal{A}, \mathcal{B}, \mathcal{C}, \dots$ 表示, 矩阵用大写字母 A, B, C, \dots 表示, 向量用 a, b, c, \dots 表示。 N 阶张量可以记为 $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, I_i 是它的第 i 个模数, $i = 1, 2, \dots, N$ 。它的元素表示为 $\mathcal{A}_{i_1 i_2 \dots i_N}$, 其中 $i_k = 1, 2, \dots, I_k; k = 1, 2, \dots, N$ 。

张量可以表为若干个向量的外积, 即 $\mathcal{A} = u^{(1)} \circ u^{(2)} \circ \dots \circ u^{(N)}$ 。此时, 张量 \mathcal{A} 为秩为 1 的张量。

张量的 Frobenius 范数定义为 $\|\mathcal{A}\|_F = \langle \mathcal{A}, \mathcal{A} \rangle = \left(\sum_{i_1, \dots, i_N} \mathcal{A}_{i_1 \dots i_N}^2 \right)^{\frac{1}{2}}$ 。

1.2. 张量的 CP 分解

CP 分解是张量最为常见的分解方法之一, 它将张量表为一系列秩为 1 的张量的和。对 N 阶张量 $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, 它的 CP 分解可以写为

$$\mathcal{A} = \sum_{i=1}^r u_i^{(1)} \circ u_i^{(2)} \circ \dots \circ u_i^{(N)},$$

其中 $r = \min \left\{ R \mid \mathcal{A} = \sum_{i=1}^R u_i^{(1)} \circ u_i^{(2)} \circ \dots \circ u_i^{(N)} \right\}$ 是张量的秩。张量的 CP 分解保证了张量分解的唯一性。

一般将张量分解的问题转化为优化问题, 于是 N 阶张量的 CP 分解问题等价于优化问题:

$$\min \frac{1}{2} \left\| \mathcal{A} - \sum_{i=1}^r u_i^{(1)} \circ u_i^{(2)} \circ \dots \circ u_i^{(N)} \right\|_F^2. \quad (1)$$

令

$$\begin{aligned} U^{(1)} &= (u_1^{(1)}, u_2^{(1)}, \dots, u_r^{(1)}) \\ U^{(2)} &= (u_1^{(2)}, u_2^{(2)}, \dots, u_r^{(2)}) \\ &\dots \\ U^{(N)} &= (u_1^{(N)}, u_2^{(N)}, \dots, u_r^{(N)}) \end{aligned}$$

则优化问题(1)又可以写成

$$\min \frac{1}{2} \left\| \mathcal{A} - \llbracket U^{(1)}, U^{(2)}, \dots, U^{(N)} \rrbracket \right\|_F^2, \quad (2)$$

其中 $\llbracket U^{(1)}, U^{(2)}, \dots, U^{(N)} \rrbracket = \sum_{i=1}^r u_i^{(1)} \circ u_i^{(2)} \circ \dots \circ u_i^{(N)}$ 。 $U^{(1)}, U^{(2)}, \dots, U^{(N)}$ 称为张量 \mathcal{A} 的因子矩阵，也是我们在张量分解问题中要求得的解。

2. 具有稀疏性的张量分解问题

因目标张量数据过大，我们需要注意到在实际处理中遇到的问题。稀疏性的引入，可以帮我们解决如占据内存过大、不相关不活跃的数据过多而很难从中找出我们所需的特征等类似的问题。

为引入稀疏性，研究者们考虑在张量分解的优化问题中加入因子的正则范数。在已有的文献中，我们可以找到关于在优化问题中加入因子矩阵的 L1 范数的讨论。于是，稀疏的张量 CP 分解问题可以等价于

$$\min \frac{1}{2} \left\| \mathcal{A} - \llbracket U^{(1)}, U^{(2)}, \dots, U^{(N)} \rrbracket \right\|_F^2 + \sum_{i=1}^N \lambda_i \|U^{(i)}\|_1,$$

其中 $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ ， $U^{(i)} \in \mathbb{R}^{I_i \times r}$, $i=1, 2, \dots, N$ 。

为计算方便，我们常以三阶张量为例，更高阶的情况同三阶张量的求解思路及方法一致。

对三阶的张量 $\mathcal{A} \in \mathbb{R}^{m \times p \times q}$ ，加入因子的 L1 范数，它的稀疏分解问题为

$$\min \frac{1}{2} \left\| \mathcal{A} - \sum_{i=1}^r x_i \circ y_i \circ z_i \right\|_F^2 + \lambda_1 \|X\|_1 + \lambda_2 \|Y\|_1 + \lambda_3 \|Z\|_1,$$

其中 $X = (x_1, x_2, \dots, x_r)$, $Y = (y_1, y_2, \dots, y_r)$, $Z = (z_1, z_2, \dots, z_r)$ 分别为 $m \times r$, $p \times r$, $q \times r$ 的矩阵。

在统计学和机器学习中，Lq 范数 ($0 < q < 1$) 中的 q 的取值越小，求得的解的稀疏性越好。我们用 Lq 范数来代替 L1 范数，于是得到本文所提出的基于 Lq 范数的稀疏张量分解。它的优化问题为

$$\min \frac{1}{2} \left\| \mathcal{A} - \sum_{i=1}^r x_i \circ y_i \circ z_i \right\|_F^2 + \lambda_1 \|X\|_q + \lambda_2 \|Y\|_q + \lambda_3 \|Z\|_q,$$

等价地，也可以为

$$\min \frac{1}{2} \left\| \mathcal{A} - \llbracket X, Y, Z \rrbracket \right\|_F^2 + \lambda_1 \|X\|_q + \lambda_2 \|Y\|_q + \lambda_3 \|Z\|_q. \quad (3)$$

3. 使用乘子交替方向法求解目标优化问题

这一节中，我们首先对乘子交替方向法进行简单的描述，随后给出问题(3)在使用乘子交替方向法的求解思路及 $q = \frac{1}{2}$ 时问题解的形式。

3.1. 乘子交替方向法

乘子交替方向法是一种求解优化问题的计算框架，它适用于求解大规模分布式优化问题。通过分解协调过程，将问题分解为多个局部小问题，求解出局部小问题的解从而得到原问题的解。

考虑优化问题

$$\begin{aligned} \min & f(x) + g(y) \\ \text{s.t.} & x - y = 0; \end{aligned}$$

该问题的迭代求解算法为

$$\begin{aligned} x^{k+1} & \in \arg \min L_\beta(x, y^k, \lambda^k), \\ y^{k+1} & \in \arg \min L_\beta(x^{k+1}, y, \lambda^k), \\ \lambda^{k+1} & = \lambda^k - \beta(x^{k+1} - y^{k+1}), \end{aligned}$$

增广拉格朗日函数为 $L_\beta(x, y, \lambda) = f(x) + g(y) - \lambda^T(x - y) + \frac{\beta}{2}\|x - y\|^2$, $\beta > 0$ 是罚参数, λ 是拉格朗日乘子。

3.2. 求解目标问题

我们加入约束条件 $X = L, Y = M, Z = N$, 问题(3)变为

$$\begin{aligned} \min & \frac{1}{2}\|\mathcal{A} - \llbracket X, Y, Z \rrbracket_F^2 + \lambda_1 \|L\|_q + \lambda_2 \|M\|_q + \lambda_3 \|N\|_q, \\ \text{s.t.} & X = L, Y = M, Z = N, \end{aligned}$$

它的增广拉格朗日函数

$$\begin{aligned} & L(X, Y, Z, L, M, N, \Lambda_1, \Lambda_2, \Lambda_3), \\ & = \frac{1}{2}\|\mathcal{A} - \llbracket X, Y, Z \rrbracket_F^2 + \lambda_1 \|L\|_q + \lambda_2 \|M\|_q + \lambda_3 \|N\|_q - \langle \Lambda_1, X - L \rangle \\ & \quad - \langle \Lambda_2, Y - M \rangle - \langle \Lambda_3, Z - N \rangle + \frac{\tau_1}{2}\|X - L\|_F^2 + \frac{\tau_2}{2}\|Y - M\|_F^2 + \frac{\tau_3}{2}\|Z - N\|_F^2 \end{aligned}$$

它的迭代算法为

$$\begin{aligned} L^{k+1} & = \arg \min \frac{\tau_1}{2} \left\| L - X^k - \frac{1}{\tau_1} \Lambda_1^k \right\|_F^2 + \lambda_1 \|L\|_q, \\ X^{k+1} & = \arg \min \frac{1}{2} \left\| \mathcal{A} - \sum_{i=1}^r x_i \circ y_i^k \circ z_i^k \right\|_F^2 + \frac{\tau_1}{2} \left\| L^{k+1} - X - \frac{1}{\tau_1} \Lambda_1^k \right\|_F^2, \\ M^{k+1} & = \arg \min \frac{\tau_2}{2} \left\| M - Y^k - \frac{1}{\tau_2} \Lambda_2^k \right\|_F^2 + \lambda_2 \|M\|_q, \\ Y^{k+1} & = \arg \min \frac{1}{2} \left\| \mathcal{A} - \sum_{i=1}^r x_i^{k+1} \circ y_i \circ z_i^k \right\|_F^2 + \frac{\tau_2}{2} \left\| M^{k+1} - Y - \frac{1}{\tau_2} \Lambda_2^k \right\|_F^2, \\ N^{k+1} & = \arg \min \frac{\tau_3}{2} \left\| N - Z^k - \frac{1}{\tau_3} \Lambda_3^k \right\|_F^2 + \lambda_3 \|N\|_q, \\ Z^{k+1} & = \arg \min \frac{1}{2} \left\| \mathcal{A} - \sum_{i=1}^r x_i^{k+1} \circ y_i^{k+1} \circ z_i \right\|_F^2 + \frac{\tau_3}{2} \left\| N^{k+1} - Z - \frac{1}{\tau_3} \Lambda_3^k \right\|_F^2, \\ \Lambda_1^{k+1} & = \Lambda_1^k - \lambda_1 (L^{k+1} - X^{k+1}), \\ \Lambda_2^{k+1} & = \Lambda_2^k - \lambda_2 (M^{k+1} - Y^{k+1}), \end{aligned}$$

$$\Lambda_3^{k+1} = \Lambda_3^k - \lambda_3 (N^{k+1} - Z^{k+1});$$

方法的思路是对 $L, X, M, Y, N, Z, \Lambda_1, \Lambda_2, \Lambda_3$ 依次进行迭代, 直至问题收敛。求得的矩阵 X, Y, Z 为张量的因子矩阵。

当 $q = \frac{1}{2}$ 时, L, M, N 的子优化问题具有显示解。于是, 令 $q = \frac{1}{2}$, L 的优化问题变为

$$L^{k+1} = \arg \min \frac{\tau_1}{2} \left\| L - X^k - \frac{1}{\tau_1} \Lambda_1^k \right\|_F^2 + \lambda_1 \|L\|_{\frac{1}{2}}$$

根据文献[7]中的证明, 推出相应的阈值算子为

$$H_{\frac{2\lambda_1}{\tau_1}}(x) = \begin{cases} \frac{2}{3}x \left(1 + \cos \left(\frac{2\pi}{3} - \frac{2\varphi(x)}{3} \right) \right), & |x| > \frac{\sqrt[3]{54}}{4} \left(\frac{2\lambda_1}{\tau_1} \right)^{\frac{2}{3}} \\ 0, & \text{其他} \end{cases}$$

其中 $\varphi(x) = \arccos \left(\frac{\lambda_1}{4\tau_1} \left(\frac{|x|}{3} \right)^{\frac{3}{2}} \right)$, x 对应矩阵 $X^k + \frac{1}{\tau_1} \Lambda_1^k$ 的每一个元素。 M 和 N 的子问题的解可以参照 L 得出。

X 的子优化问题

$$X^{k+1} = \arg \min \left\| \frac{1}{2} \mathcal{A} - \sum_{i=1}^r x_i \circ y_i^k \circ z_i^k \right\|_F^2 + \frac{\tau_1}{2} \left\| L^{k+1} - X - \frac{1}{\tau_1} \Lambda_1^k \right\|_F^2 \quad (4)$$

的求解涉及到了张量的计算, 为降低计算困难, 我们可以将张量展开成矩阵的形式。张量的展开是将张量沿某一个模的展开, 张量展开成的矩阵包括张量中的每一个元素, 且张量沿不同的模展开得到的矩阵也不一定相同。张量按它的第一个模展开所得的矩阵记为 $A_{(1)}$, 它是一个 $m \times pq$ 的矩阵, 我们在 X 的优化问题中选用张量沿第一个模展开的矩阵来计算。问题(4)转化为

$$X^{k+1} = \arg \min \frac{1}{2} \left\| A_{(1)} - X (Z^k \circ Y^k)^T \right\|_F^2 + \frac{\tau_1}{2} \left\| L^{k+1} - X - \frac{1}{\tau_1} \Lambda_1^k \right\|_F^2, \quad (5)$$

\circ 表示矩阵的 Khatri-Rao 积, 等式(5)的右边为最小二乘问题。对(5)求导,

$$0 = (Z^k \circ Y^k)^T (A_{(1)}^T - (Z^k \circ Y^k) X^T) - \tau_1 (L^{k+1T} - X^T) - \Lambda_1^{kT},$$

整理后得第 $k+1$ 步矩阵 X 的解

$$X^T = \left((Z^k \circ Y^k)^T (Z^k \circ Y^k) - \tau_1 I \right)^{-1} \left((Z^k \circ Y^k)^T A_{(1)}^T - \tau_1 L^{k+1T} - \Lambda_1^{kT} \right).$$

Y, Z 的优化问题的求解过程类似, 不同的是, Y, Z 的求解所用的展开矩阵分别为沿张量的第二个模和第三个模展开的矩阵。

4. 总结

具有稀疏性的张量分解因其理论性质及在应用中良好的体现, 逐渐得到学者们的关注。加入因子 L_1 范数来引入稀疏性在一些文献中得到了验证, 而 L_q 范数 ($0 < q < 1$) 的情况目前还未有学者提出。本文提出了基于 L_q 范数的稀疏张量分解问题, 同时给出了使用乘子交替方向法求解张量分解问题的思路和过程

以及 $q = \frac{1}{2}$ 时问题所具有的显式解的形式。在接下来的研究中, 我们希望能够进行数据实验, 和普通的张量分解以及引入 L1 范数的稀疏张量分解的结果进行比较, 证明基于 L_q 范数的分解问题的稀疏性以及它具有更好的稀疏性。

参考文献

- [1] Kim, H.J., Ollila, E. and Koivunen, V. (2013) Sparse Regularization of Tensor Decompositions. *IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, 26-31 May 2013. <https://doi.org/10.1109/ICASSP.2013.6638376>
- [2] Allen, G. (2012) Sparse Higher Order Principal Components Analysis. <http://proceedings.mlr.press/v22/allen12/allen12.pdf>
- [3] Liu, J., Liu, J., Wonka, P., et al. (2012) Sparse Non-Negative Tensor Factorization Using Columnwise Coordinate Descent. *Pattern Recognition*, **45**, 649-656. <https://doi.org/10.1016/j.patcog.2011.05.015>
- [4] Sun, W.W., Li, J.W., Liu, H., et al. (2017) Provable Sparse Tensor Decomposition. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **79**, 899-916. <https://doi.org/10.1111/rssb.12190>
- [5] Kolda, T.G. and Bader, B.W. (2009) Tensor Decompositions and Applications. *SIAM Review*, **51**, 455-500. <https://doi.org/10.1137/07070111X>
- [6] Qi, L., Sun, W. and Wang, Y. (2007) Numerical Multilinear Algebra and Its Applications. *Frontiers of Mathematics in China*, **2**, 501-526. <https://doi.org/10.1007/s11464-007-0031-4>
- [7] Xu, Z., Chang, X., Xu, F. and Zhang, H. (2012) L-1/2 Regularization: A Thresholding Representation Theory and a Fast Solver. *IEEE Transactions on Neural Networks & Learning Systems*, **23**, 1013-1027. <https://doi.org/10.1109/TNNLS.2012.2197412>