

# 一种基于 $L_{2,1}$ 范数的 PCA 维数约简算法\*

刘丽敏<sup>a</sup>, 樊晓平<sup>a</sup>, 廖志芳<sup>b</sup>, 刘曼玲<sup>a</sup>

(中南大学 a. 信息科学与工程学院; b. 软件学院, 长沙 410075)

**摘要:** 传统 PCA 存在对异常值和特征噪声敏感等问题, 基于  $L_{2,1}$  范数的 PCA 算法改进了这些缺点。现有的基于  $L_{2,1}$  范数的 PCA 算法是通过降低矩阵的秩来实现维数约简, 而秩的计算复杂。针对这一问题, 提出一种新的维数约简算法。该算法提出利用迹范数代替矩阵的秩来简化  $L_{2,1}$ -PCA 的计算, 提高算法效率; 对于算法的求解提出了基于拉格朗日乘子的方法并将算法应用扩展 Yale B 人脸数据集进行图像去噪。可视化的实验结果表明所提出的算法有效。

**关键词:** 维数约简; 主成分分析;  $L_{2,1}$ -PCA;  $L_{2,1}$  范数; 拉格朗日乘子

**中图分类号:** TP301.6      **文献标志码:** A      **文章编号:** 1001-3695(2013)01-0039-03

**doi:** 10.3969/j.issn.1001-3695.2013.01.008

## Efficient dimension reduction algorithm via $L_{2,1}$ norm PCA

LIU Li-min<sup>a</sup>, FAN Xiao-ping<sup>a</sup>, LIAO Zhi-fang<sup>b</sup>, LIU Man-ling<sup>a</sup>

(a. School of Information Science & Engineering, b. School of Software, Central South University, Changsha 410075, China)

**Abstract:** Traditional PCA is sensitive to outliers and feature noises, PCA based on  $L_{2,1}$ -norm can improve the problems. Whereas present  $L_{2,1}$ -PCA algorithms implement dimension reduction on the rank of the matrix and the rank is complex problem. In order to solve this problem, this paper proposed using trace norm instead of rank, then the calculation of  $L_{2,1}$ -PCA algorithm could simplify and the efficiency could improve. It also put forward an efficient augmented Lagrange multiplier (ALM) algorithm for the solutions. Extensive experiments on extended Yale B face data sets verify the efficiency of the proposed algorithm.

**Key words:** dimension reduction; principal component analysis (PCA);  $L_{2,1}$ -PCA;  $L_{2,1}$  norm; augmented Lagrange multiplier

随着信息化技术的快速发展, 高维数的非结构化数据大量出现, 如人脸图像、生物数据、文本数据等。如何从高维数据中发现数据内在结构信息和本质维数成为机器学习和模式识别研究的热点。维数约减(降维)是处理高维数据的一种常见的方法。维数约减是将高维数据向低维表达空间转换的过程<sup>[1]</sup>。国内外很多学者对此问题进行了大量的研究, 提出了很多算法<sup>[2-9]</sup>。例如, 无监督的主成分分析(PCA)、非负矩阵分解(nonnegative matrix factorization, NMF)等算法; 有监督的线性判别分析(linear discriminant analysis, LDA)、独立分量分析(independent component analysis, ICA)等算法。其中 PCA 是一种常用的维数约减方法, 但是传统的 PCA 是基于 Frobenius 范数, 也就是说它是利用最小二乘法最小化误差的平方, 而这种方法会造成更大的误差, 因此基于 Frobenius 范数的 PCA 对异常值和特征噪声敏感。近些年来, 很多学者为解决此问题做了大量的研究工作。例如, 文献[10]提出基于  $L_1$  范数的 PCA 算法, 并提出利用半正定规划问题来优化算法。文献[11]提出基于核函数的  $L_1$ -PCA 算法, 文献[12]提出了基于 R1-PCA 的算法, 文献[13]提出了基于  $L_{2,1}$ -PCA 的算法。但是现有的  $L_{2,1}$ -PCA 算法通常是通过降低矩阵的秩来实现维数约简, 秩的计算是非凸问题且计算复杂。本文针对这一问题, 提出用迹范数代替矩阵的秩, 简化计算, 从而提高维数约简的效率。对

于算法的求解过程提出了基于拉格朗日乘子的方法, 最后将算法应用于扩展 Yale B 人脸数据集进行数据维数约简处理。可视化的实验结果表明算法有效。

本文介绍基于迹范数的  $L_{2,1}$ -PCA 的维数约简算法, 详细介绍了基于拉格朗日乘子的算法求解方法以及算法框架, 并进行了可视化的实验。

### 1 基于迹范数的 $L_{2,1}$ -PCA 的维数约简算法

在传统的 PCA 中, 数据的维数约简通常是基于矩阵的秩。假设给定数据  $X = \{x_1, x_2, \dots, x_n\} \in \mathbb{R}^{p \times n}$ :

$$X \approx Z \quad (1)$$

其中:  $Z = FG$ ,  $F$  和  $G$  的解为

$$\min_{F, G} \|X - FG\|_F^2, \text{ 且 } \text{rank}(F) = \text{rank}(G) = k \quad (2)$$

式(2)中的  $\|\cdot\|_F$  是指 Frobenius 范数, 定义为  $\|A\|_F = (\sum_{i=1}^p \sum_{j=1}^n |a_{ij}|^2)^{\frac{1}{2}}$ 。然而正如引言所讲式(2)是利用标准最小二乘法的方法来计算, 这种方法的缺点就是容易产生较大的误差。一种改进的方法是基于  $L_{2,1}$  范数的算法<sup>[10]</sup>, 其数学描述如式(3)所示:

$$\min_{F, G} \|X - FG\|_{2,1}, \text{ 且 } \text{rank}(FG) = k \quad (3)$$

式中  $\|\cdot\|_{2,1}$  中的 2, 1 是指  $L_{2,1}$  范数, 定义为:  $\|A\|_{2,1} =$

**收稿日期:** 2012-05-11; **修回日期:** 2012-06-25      **基金项目:** 国家科技支撑计划资助项目(2012BAH08B00); 国家“863”计划资助项目(2007AA022008)

**作者简介:** 刘丽敏(1976-), 女, 湖南长沙人, 讲师, 博士研究生, 主要研究方向为数据挖掘、计算机视觉、智能信息处理(liulimin@csu.edu.cn); 樊晓平(1961-), 男, 教授, 博士, 主要研究方向为智能信息处理、机器学习; 廖志芳(1968-), 女, 副教授, 博士, 主要研究方向为数据挖掘、推荐系统; 刘曼玲(1960-), 女, 副教授, 主要研究方向为数据挖掘、智能信息处理。

$\sum_{j=1}^n (\sum_{i=1}^p A_{ij}^2)^{\frac{1}{2}}$ 。但是秩的计算是非凸问题且计算复杂,因此本文利用迹范数  $\|Z\|_F$  取代  $\text{rank}(Z)$ 。因此式(3)可以表达为

$$\min_Z \|X - Z\|_{2,1}, \text{ 且 } \|Z\|_F \leq k \quad (4)$$

式(4)可以等价

$$\min_Z \lambda \|X - Z\|_{2,1} + \|Z\|_F \quad (5)$$

其中,迹范数  $\|Z\|_F$  定义为  $\|Z\|_F = \text{Tr}(ZZ^T)^{1/2} = \text{Tr}(Z^T Z)^{1/2} = \sum_{k=1}^{\text{rank}(Z)} \xi_k$ ,  $\xi_k$  为  $Z$  的奇异值,  $\lambda$  是一个正常数。

针对式(5)的求解问题,本文提出利用拉格朗日乘子方法求解。

## 2 算法求解过程

式(5)中,令  $M = X - Z, N = Z$ ,式(5)的求解问题转换为

$$\min_Z \lambda \|M\|_{2,1} + \|N\|_F \quad (6)$$

针对式(6)的求解问题,本文提出利用拉格朗日乘子(augmented lagrange multiplier, ALM)<sup>[14]</sup>来解决。根据 ALM 构造拉格朗日函数,式(6)可以等价

$$\min_{Z,M,N} J = \min_{Z,M,N} (\lambda \|M\|_{2,1} + \langle B, X - Z - M \rangle + \frac{\mu}{2} \|X - Z - M\|_F^2 + \|N\|_F + \langle A, Z - N \rangle + \frac{\mu}{2} \|Z - N\|_F^2) \quad (7)$$

其中: $A, B$  表示拉格朗日乘子; $\mu$  表示惩罚因子; $\langle I, J \rangle$  定义为  $\langle I, J \rangle = \sum_{ij} I_{ij} J_{ij} = \text{Tr} I^T J$ 。

ALM 是一种迭代更新算法,在迭代更新过程中需要解决式(7)中的求解问题和参数迭代问题。

### 2.1 算法求解

对于输入  $A, B$  和  $\mu$ ,算法的关键就是解决式(7)的解问题。很明显,这个问题可以通过  $M, N$  和  $Z$  闭合形式解来解决。

1) 算法固定  $N$  和  $Z$  来求解  $M$ ,式(7)可以等价

$$\min_M J = \min_M \lambda \|M\|_{2,1} + \langle B, X - Z - M \rangle + \frac{\mu}{2} \|X - Z - M\|_F^2 \quad (8)$$

式(8)可以等价

$$\min_M J = \min_M \lambda \|M\|_{2,1} + \frac{\mu}{2} \|M - (X - Z + \frac{B}{\mu})\|_F^2 \quad (9)$$

式(9)的闭合形式解的求解过程如下:

令  $Q = X - Z + \frac{B}{\mu}$ ,式(9)等价

$$\min_M J = \min_M \lambda \|M\|_{2,1} + \frac{\mu}{2} \|M - Q\|_F^2 \quad (10)$$

式(10)的问题可以分解为  $n$  个子问题:

$$\min_M \sum_{i=1}^n \lambda \|m_i\|_{2,1} + \frac{\mu}{2} \|m_i - q_i\|_F^2 \quad (11)$$

根据文献[15],式(11)的解为

$m_i = \text{sign}(q_i) (|q_i| - \frac{\lambda}{\mu})_+$ , 因此,式(9)的闭合形式解为

$$M_{ij} = \text{sign}(Q_{ij}) (|Q_{ij}| - \frac{\lambda}{\mu})_+。$$

2) 算法固定  $M$  和  $Z$  来求解  $N$ ,式(7)可写成:

$$\min_N J = \min_N \|N\|_F + \langle A, Z - N \rangle + \frac{\mu}{2} \|Z - N\|_F^2 \quad (12)$$

式(12)可以等价

$$\min_N J = \min_N \|N\|_F + \frac{\mu}{2} \|N - (Z + \frac{A}{\mu})\|_F^2 \quad (13)$$

式(13)的求解可以通过奇异值分解(singular value decomposition, SVD)获得,也就是,  $\text{SVD}(Z + \frac{A}{\mu}) = P \Sigma Q^T$ , 其中  $P$  是

左奇异向量,  $Q$  是右奇异向量,  $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r)$  且  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$ , 则  $N$  的解为  $N_{ij} = P_{ij} (\Sigma - \frac{1}{\mu}) Q_{ij}^T$ , 其中  $S_+$  定义为  $S_+ = 0.5 (\|S_{ij}\| + S_{ij})$ 。

3) 算法固定  $M$  和  $N$  来求解  $Z$ ,式(7)可以等价

$$\min_Z J = \min_Z (\langle A, Z - N \rangle + \frac{\mu}{2} \|X - Z - M\|_F^2 + \langle B, X - Z - M \rangle + \frac{\mu}{2} \|Z - N\|_F^2) \quad (14)$$

令  $\frac{\partial J}{\partial Z} = 0$ , 可以求得  $Z$  的解, 即为  $Z = \frac{X - (M - N)}{2} - \frac{A - B}{2\mu}$ 。

### 2.2 参数迭代过程

在拉格朗日乘子迭代过程中,每次迭代得到  $M, N$  和  $Z$ , 其中参数  $A, B$  和  $\mu$  的迭代方法如式(15)所示。

$$\begin{aligned} A &\leftarrow A + \mu(Z - N) \\ B &\leftarrow B + \mu(Z - X - M) \\ \mu &\leftarrow \mu\gamma \end{aligned} \quad (15)$$

其中: $\gamma$  是一个常量, 并且  $\gamma > 1$ 。从式(7)中可以看出,公式不仅使用拉格朗日乘子  $A$  和  $B$  来强制约束  $M$  和  $N$ , 而且公式中增加了两项惩罚项  $\frac{\mu}{2} \|Z - N\|_F^2$  和  $\frac{\mu}{2} \|X - Z - M\|_F^2$ , 这就加倍加强约束了  $M$  和  $N$ 。

### 2.3 算法框架

本文所提出的算法及其求解过程的框架如下所示:

```
input data:  $X, k$ 
initialize
 $Z = 0.9X, A = 0, B = 0, \mu = \frac{1}{\|X\|_{L_1}}, \rho = 1.1$ 
repeat
 $M = \arg \min \lambda \|M\|_{2,1} + \frac{\mu}{2} \|M - (X - Z + \frac{B}{\mu})\|_F^2$ 
 $N = \arg \min \|N\|_F + \langle A, Z - N \rangle + \frac{\mu}{2} \|Z - N\|_F^2$ 
 $Z = (I + X^T X)^{-1} (N + X^T (X - M + \frac{B}{\mu}) - \frac{A}{\mu})$ 
 $A \leftarrow A + \mu(Z - N)$ 
 $B \leftarrow B + \mu(Z - X - M)$ 
 $\mu = \min(\mu\gamma, 10^{15})$ 
until converge
```

从式(15)可看出,在迭代过程中,  $\mu$  的值越来越大,但是  $\mu$  不能无限大,因此在算法框架中  $\mu$  的取值为  $\min(\mu\gamma, 10^{15})$ 。

算法中的参数  $A, B$  并不是通常意义下的拉格朗日乘子, 实际上是拉格朗日乘子的一个估计, 即当  $t \rightarrow \infty, A^t + \mu^t (Z^t - N^t) \rightarrow A^*$ , 且  $B^t + \mu^t (Z^t - X^t - M^t) \rightarrow B^*$ 。从算法中可以看出, 在每次迭代后,  $M, N, Z$  值的变化越来越小, 即  $M^t \approx M^{t+1}, N^t \approx N^{t+1}, Z^t \approx Z^{t+1}$ , 因此  $M, N$  和  $Z$  是可解的, 所以算法可以收敛。

## 3 可视化实验分析

在计算机视觉、图像处理等应用中,维数约简通常起到降噪的作用。因此,为了评价所提算法的有效性,本文将所提算法应用于扩展 Yale B 人脸数据集进行降噪处理。

扩展 Yale B 人脸数据集包含 38 位志愿者不同光照、表情和姿态变化的 2 414 张图片。本文从扩展 Yale B 人脸数据集中取部分图像,图像大小调整为  $35 \times 35$  像素。

图 1 显示了扩展 Yale B 人脸数据集中四位志愿者不同光照六张原始数据图像和利用本文提出的模型进行数据降噪处理后的对比分析。图 1 中第一行是每个人的原始图像,第二行

是利用本文算法进行降噪后的图像。从图1的对比可以看出,每个人的原始图像光照不同,并且有些图像有很多阴影,这些阴影可以说是图像的遮挡。从利用本文算法进行降噪后的图像即降噪后的第二行图像中可以看出,图像普遍都具有了相近的亮度,而且有些阴影已经从图像中移除,有些全黑的图像也基本恢复,尤其是图像的中部,由此可以说明本文所提出的基于迹范数的  $L_{2,1}$ -PCA 维数约简算法是有效的。



图1 扩展Yale B人脸数据降噪前后比较

#### 4 结束语

本文针对现有  $L_{2,1}$ -PCA 算法存在的计算复杂这一问题,提出用迹范数代替矩阵的秩,对于算法的求解过程提出了基于增强拉格朗日乘子的方法,并将算法应用于扩展 Yale B 人脸数据集进行数据维数约简处理。可视化的实验结果表明算法有效。

#### 参考文献:

[1] 陶剑文,王士同.  $L_1$  范局部线性嵌入[J]. 中国图象图形学报,

2011,16(10):1802-1811.

- [2] WRIGHT J, PENG Yi-gang, MA Yi. Robust principal component analysis: exact recovery of corrupted low-rank matrices by convex optimization [C]//Advance in Neural Information Processing Systems. 2009:2080-2088.
- [3] WU Xiao-jun, JOSEF K, YANG Jing-yu, et al. A new direct LDA (D-LDA) algorithm for feature extraction in face recognition [C]//Proc of the 17th International Conference on Pattern Recognition. 2004:545-548.
- [4] HYVÄRINEN A, OJA E. Independent component analysis: algorithms and applications [J]. Neural Networks, 2000, 13(4-5):411-430.
- [5] ZHAO Ying, ATKESON C G. Implementing projection pursuit learning [J]. IEEE Trans on Neural Networks, 1996, 7(2):362-373.
- [6] NICOLAS G, FRANCOIS G. A multilevel approach for nonnegative matrix factorization [J]. Journal of Computational and Applied Mathematics, 2012, 236(7):1708-1723.
- [7] BURGESS C J C. Dimension reduction: a guided tour [J]. Foundations and Trends® in Machine Learning, 2009, 2(4):275-365.
- [8] AILON N, CHAZELLE B. Faster dimension reduction [J]. Communications of the ACM, 2010, 53(2):97-104.
- [9] 陈诗国, 张道强. 半监督降维方法的实验比较 [J]. 软件学报, 2011, 22(1):28-43.
- [10] KE Qi-fa, KANADE T. Robust  $L_1$  norm factorization in the presence of outliers and missing data by alternative convex programming [C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2004:739-746.
- [11] 李勇, 梁志贞, 夏士雄. 基于核函数的 PCA- $L_1$  算法 [J]. 计算机工程, 2011, 37(22):174-175, 178.
- [12] DING C, ZHOU Ding, HE Xiao-feng.  $R_1$ -PCA: rotational invariant  $L_1$ -norm principal component analysis for robust subspace factorization [C]//Proc of the 23rd International Conference on Machine Learning. 2006:281-288.
- [13] NIE Fei-ping, HUANG H, XIAO Cai, et al. Efficient and robust feature selection via joint  $L_{2,1}$ -norms minimization [C]//Proc of the 24th Annual Conference on Neural Information Processing Systems. 2010:1813-1821.
- [14] LIN Zhou-chen, CHEN Min-ming, MA Yi. The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices [EB/OL]. (2012-04-24) [2012-06-19]. <http://arxiv.org/abs/1009.5055>.
- [15] TIBSHIRANI R. Regression shrinkage and selection via the LASSO [J]. Journal of the Royal Statistical Society, Series B, 1994, 58(1):267-288.

(上接第38页)

考虑连通可靠度的不断变化,可以增加在实际路网路径选择中考虑行程时间波动和正点到达目的地的概率的合理性和准确性,以提供更准确的决策辅助信息,从而实现安全、可靠、高效的运营。

#### 参考文献:

- [1] MINE H, KAWAI H. Mathematics for reliability analysis [M]. Tokyo: Asakura-shoten, 1982:12-14.
- [2] 饭田恭敬. 交通工程学 [M]. 北京:人民交通出版社, 1994:171-180.
- [3] BELL M G H, HIDA Y. Transport network analysis [M]. New York: Wiley, 1997:179-191.
- [4] LIDA Y. Basic concepts and future directions of road network reliability analysis [J]. Journal of Advanced Transportation, 1999, 33(2):125-134.
- [5] DU Zhen-ping, NICHOLSON A. Degradable transportation system: sensitivity and reliability analysis [J]. Trans on Research B, 1997, 31(3):225-237.
- [6] NICHOLSON A, DU Zhen-ping. Degradable transportation systems; an

integrated equilibrium model [J]. Trans on Research Part B, 1997, 31(3):209-223.

- [7] SANSO B, MILOT L. Performability of a congested urban transportation network when accident information is available [J]. Trans on Science, 1999, 33(1):68-79.
- [8] 刘海旭. 基于关键路段的路网可靠度的改进 [J]. 西南交通大学学报, 2006, 141(5):645-648.
- [9] 朱顺应, 王炜, 邓卫, 等. 交通网络可靠度及其通路算法研究 [J]. 中国公路学报, 2000, 13(1):91-94.
- [10] 陈艳艳, 刘小明. 可靠度在交通系统规划与管理中的应用 [M]. 北京:人民交通出版社, 2006.
- [11] 侯立文. 城市道路网络可靠度的研究 [J]. 系统工程, 2000, 18(5):46-48.
- [12] 陈晓明, 邵春福, 熊志华. 混合交通信号交叉口的通行能力可靠度 [J]. 中国公路学报, 2008, 21(4):99-104.
- [13] 吕斌, 牛惠民. 信号控制交叉口可靠度建模与仿真 [J]. 交通运输系统工程与信息, 2011, 11(6):45-50.
- [14] 朱毅麟. 网络可靠度的简捷计算法 [J]. 中国空间科学技术, 1984, 4(4):1-6.