

LaTeX Author Guidelines for ICCV Proceedings

Anonymous ICCV submission

Paper ID ****

Abstract

我们从Weighted Orthogonal Procrustes Problem出发，设计出一个weighted Frobenious norm作为data term，用weighted sparse coding作为regularization term。这个新的模型用alternative updating方式求解，可以证明其对于每个变量都有闭合解，并且收敛到一个stationary point。这个模型用在了真实去噪问题里，对之前咱提出的guided的方法在60张真实噪声图上的去噪效果更好，PSNR提升了1.1dB左右。

1. 模型Introduction

介绍各种在优化模型的数据项里加权重的方式和含义，目前最常见的是孟老师的pixel-wise的加权和杨师兄的对每个variable加权。

先看孟老师的模型，其主体是：

$$\min_{\mathbf{U}, \mathbf{V}} \|\mathbf{W} \odot (\mathbf{X} - \mathbf{UV}^T)\|_1 \quad (1)$$

这个模型应用在Low-Rank Matrix Factorization with Missing Entries这个问题里， \mathbf{W} 指矩阵的元素是否是missing entry，其元素只有0或1。这几篇文章中，都用Mixture of Gaussians (MoG)模拟噪声，用EM算法估计噪声分布， \mathbf{W}_{ij} 是噪声属于所有Gaussian的期望与高斯噪声方差的比值的和，即 $\mathbf{W}_{ij} = \sum_{k=1}^K \frac{\gamma_{ijk}}{2\sigma_k^2}$ 。这个加权方式不适合用于对真实噪声加权。原因是1，这个权重矩阵变量太多，每个元素对应于一个像素值，但是很难仅仅通过一个像素值就估计出这个像素值里噪声的水平。2，真实噪声里，可能某个像素的噪声是outlier，而EM算法对outlier不鲁棒，导致无法很好地去除这个像素的噪声。3，EM算法估计过程计算量很

大，不适合用于对真实噪声加权。4，单个像素加一个权重忽略了图像的结构信息，所以这个加权方式适用于随机程度比较高的问题，比如矩阵分解，但是对于图像修复或者分类就不太适用。

杨猛模型主体是

$$\min_{\mathbf{c}} \|\mathbf{W}^{\frac{1}{2}}(\mathbf{y} - \mathbf{Dc})\|_2^2 \quad \text{s.t.} \quad \|\mathbf{c}\|_1 < \sigma \quad (2)$$

用权重矩阵左乘residual 向量，其物理含义是对residual 向量的每个variable加权，这是因为这个模型主要用在人脸识别上，当人脸有遮挡时，这个权重可以把遮挡部分的像素值的权重变小，使得模型对于有遮挡的人脸识别依然有效。这个加权方式不适合用于对真实噪声加权。原因是1，在真实噪声去噪问题里，如果把一些相似块堆起来组成一个矩阵，每一行是没有任何空间位置关系的点，要求这些点上的噪声符合高斯分布是没有依据的（相当于在图像中放一个网格，要求网格上的所有交叉点所对应的像素上的噪声符合高斯分布是不合理的）。2，

我们把对角的权重矩阵右乘residual矩阵，注意是矩阵，而不是向量。相当于假设真实噪声图的每个局部的噪声都可以近似于一个高斯，不同的局部的噪声水平不同，对应的权重不同。

2. Weighted Orthogonal Dictionary Learning

现有一个图像块，假设是8x8($d = 64$)的维度，拉成列向量 $\mathbf{y} \in \mathbb{R}^d$ ，那么我们在其周围找 m 个相似块，得到相似块矩阵 $\mathbf{Y} \in \mathbb{R}^{d \times m}$ 。我们在这个数据里学习到一个正交字典，同时利用稀疏表达这个工具。很自然，我想到了K-SVD这个模型，给定数据矩阵 \mathbf{Y} ,KSVD不断迭代更新，从而得到最适合表达数据的字典和稀疏。但是与KSVD不同的是，我希望我的模

型里，字典是正交的并且有闭合解，系数也可以有闭合解。从而，我们想到的模型主体是这样的：

$$\min_{\mathbf{D}, \mathbf{C}} \frac{1}{2} \|(\mathbf{Y} - \mathbf{DC})\mathbf{W}\|_F^2 + \lambda \|\mathbf{S}^{-1}\mathbf{C}\|_1 \quad \text{s.t.} \quad \mathbf{D}^\top \mathbf{D} = \mathbf{I}. \quad (3)$$

其中 \mathbf{S} 是对角正定矩阵(表示sparse coding每一行权值不同)。

权重矩阵 \mathbf{W} 的设计应该遵循如下原理：噪声高，权重就小。噪声低，权重就大。这个原则是从贝叶斯的角度得来的：

$$\hat{\mathbf{c}}_i = \arg \max_{\mathbf{c}_i} \ln P(\mathbf{c}_i | \mathbf{y}_i). \quad (4)$$

由贝叶斯法则，等价于

$$\hat{\mathbf{c}}_i = \arg \max_{\mathbf{c}_i} \{\ln P(\mathbf{y}_i | \mathbf{c}_i) + \ln P(\mathbf{c}_i)\}. \quad (5)$$

其中拟合residual noise的数据项是

$$P(\mathbf{y}_i | \mathbf{c}_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2\sigma^2} \|\mathbf{y}_i - \mathbf{D}\mathbf{c}_i\|_2^2). \quad (6)$$

系数正则项是

$$P(\mathbf{c}_i) = \prod_{j=1}^d \lambda \exp(-\lambda |\mathbf{c}_{ij}|). \quad (7)$$

$$\hat{\mathbf{c}}_i = \arg \min_{\mathbf{c}_i} \left\| \frac{\mathbf{y}_i - \mathbf{D}\mathbf{c}_i}{\sigma} \right\|_2^2 + \sum_{j=1}^d \lambda |\mathbf{c}_{ij}|. \quad (8)$$

$$\mathbf{W}_{ii} = \frac{1}{\sigma} \quad (9)$$

3. 初始化

- $\mathbf{W}^{(0)}$ 是对角矩阵，初始化为恒等矩阵 $\mathbf{I}_{m \times m}$ (原因是一开始把所有块平等对待，视为受到相同的高斯噪声的影响)。或者初始化为一个对角矩阵，对角线上的每个元素对应于 \mathbf{Y} 的每一列(图像块)上的噪声水平 σ ，这个噪声水平可以用现有估计高斯噪声水平的方法估计出来)；
- 正交字典 $\mathbf{D}^{(0)} = \mathbf{V}\mathbf{U}^\top$ 由 $(\mathbf{Y}\mathbf{W}^{(0)})(\mathbf{Y}\mathbf{W}^{(0)})^\top = \mathbf{U}\Sigma^{(0)}\mathbf{V}^\top$ 做SVD分解得到；
- 稀疏系数 \mathbf{C} 的行方向上的权值 $\mathbf{s}^{(0)} = \text{diag}(\Sigma^{(0)})$ 是weighted sparse coding的权重。 \mathbf{C} 每一行的权值不同，意味着正交字典 \mathbf{D} 的每一列对 \mathbf{C} 的重要性不同；

4. 反复迭代求解 \mathbf{D}, \mathbf{C} 直到收敛,更新 \mathbf{W}

4.1. 反复迭代求解 \mathbf{D}, \mathbf{C} 直到收敛

$$k = 0, 1, 2, \dots:$$

a. update \mathbf{C}

$$\mathbf{C}^{(k+1)} = \arg \min_{\mathbf{C}} \frac{1}{2} \|(\mathbf{Y} - \mathbf{D}^{(k)}\mathbf{C})\mathbf{W}^{(k)}\|_F^2 + \lambda \|\mathbf{C}\|_{1, \mathbf{s}}. \quad (10)$$

有闭合解，每一列单独求解：

$$\mathbf{c}_i^{(k+1)} = \arg \min_{\mathbf{c}_i} \frac{1}{2} \|(\mathbf{y}_i - \mathbf{D}^{(k)}\mathbf{c}_i)\mathbf{W}_{ii}^{(k)}\|_2^2 + \lambda \|\mathbf{s}^T \mathbf{c}_i\|_1. \quad (11)$$

闭合解为：

$$\mathbf{c}_i^{(k+1)} = \text{sgn}(\mathbf{D}^\top \mathbf{y}) \odot \max(|\mathbf{D}^\top \mathbf{y}| - \frac{\lambda}{(\mathbf{W}_{ii}^{(k)})^2} \mathbf{s}, 0), \quad (12)$$

b. update \mathbf{D}

$$\mathbf{D}^{(k+1)} = \arg \min_{\mathbf{D}} \frac{1}{2} \|(\mathbf{Y} - \mathbf{D}\mathbf{C}^{(k+1)})\mathbf{W}^{(k)}\|_F^2 \quad \text{s.t.} \quad \mathbf{D}^\top \mathbf{D} = \mathbf{I}. \quad (13)$$

闭合解为：

$$\mathbf{D}^{(k+1)} = \mathbf{U}\mathbf{V}^\top \quad (14)$$

其中， \mathbf{U} 和 \mathbf{V} 由 $\mathbf{C}^{(k+1)}(\mathbf{W}^{(k)})^2\mathbf{Y}^\top = \mathbf{U}\Sigma^{(k+1)}\mathbf{V}^\top$ 做SVD分解得到。 $(\mathbf{W}^{(k)})^2$ 是指 \mathbf{W} 的对角线上的元素分别平方后得到的对角矩阵。

c. update \mathbf{s}

$$\mathbf{s}^{k+1} = \Sigma^{(k+1)}$$

4.2. (3),(6)收敛后，更新 \mathbf{W}

d. update \mathbf{W}

\mathbf{W} 是对角矩阵，只需要更新迭代对角元即可：

$$\mathbf{W}_{ii}^{new} = \exp(-\lambda_2 \|\mathbf{y}_i - \mathbf{D}^{(k+1)}\mathbf{c}_i^{(k+1)}\|_2) \quad (15)$$

λ_2 是参数；

或

$$\mathbf{W}_{ii}^{new} = \lambda_2 (\sigma_i - \|\mathbf{y}_i - \mathbf{D}^{(k+1)}\mathbf{c}_i^{(k+1)}\|_2) \quad (16)$$

或者其它可能性。

待说明问题：模型(1)或(2)的收敛性，收敛速率等问题，收敛值的上下界等问题。

实际中，我采用模型(2)，没有多次迭代（因为多次迭代速度太慢），权重的设计采用(8)，权重的物理意义，设计与效果值得探讨。